

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ УКРАИНЫ

ХАРЬКОВСКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ
РАДИОЭЛЕКТРОНИКИ

ISSN 0555-2656

БИОНИКА ИНТЕЛЛЕКТА

ИНФОРМАЦИЯ, ЯЗЫК, ИНТЕЛЛЕКТ

№ 1 (66)

2007

НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Основан в 1967 г.

Свидетельство о государственной регистрации КВ № 9148 от 10.09.2004

Журнал включен в список специальных изданий ВАК Украины
по техническим наукам

(приложение к постановлению ВАК Украины № 1-05/7 от 04.07.2006)

СОДЕРЖАНИЕ

<i>Путятин Е. П.</i> Нормализация изображений: достижения и нерешенные задачи	5
<i>Каграманян А. Г., Маисталир В. П., Шляхов В. В.</i> Метрические свойства грануляции информации	17
<i>Широков В. А.</i> Очерк основных принципов квантовой лингвистики	25
<i>Бондаренко М. Ф., Шабанов-Кушнаренко С. Ю., Шабанов-Кушнаренко Ю. П.</i> Модель сглаживания в зрении	33
<i>Четвериков Г. Г.</i> Алгебро-логические средства моделирования естественного языка. Сообщение 3	48
<i>Бодянский Е. В., Тесленко Н. А.</i> Автоассоциативная нейросетевая память на основе нечетких базисных функций ..	53
<i>Руденко О. Г., Островерхий А. В.</i> Нейросетевая аппроксимация многомерных функций при наличии помех измерений	58
<i>Кучеренко Е. И., Краснокутский Д. Е.</i> Гибридные математические модели на основе расширений нечетких сетей Петри	64
<i>Валик А. В., Герасин С. Н., Ланга С. И.</i> Метод определения взаимосвязи элементов стохастической нейронной сети	68
<i>Кургаев А. Ф.</i> Характеристика архитектуры компьютерной системы, приближенной к идеальной	73
<i>Dudar Z. V., Vorochek O. G.</i> About the theoretical foundation for a Meta-Context Interchange technology	81
<i>Асеев Г. Г.</i> Математическая модель определения экономической эффективности внедрения корпоративных систем электронного документооборота	88
<i>Авраменко В. П., Калачева В. В.</i> Интеллектуальные технологии повышения эффективности бизнес-процессов	95
<i>Гребенник И. В., Баранов А. В.</i> Экстремальные свойства функций на классах композиционных образов комбинаторных множеств	98
<i>Дюбка Г. Ф., Преснякова Д. В.</i> Модель поверхностного смысла естественного языка на базе семантических функций	103
<i>Ерохин А. Л.</i> О визуальном кодировании информации на основе идентификации когнитивных функций человека-оператора. Сообщение 2	107
<i>Петров К. Э., Евсеева И. В.</i> Синтез формальной модели порядковой ординальной классификации	113
<i>Бых А. И., Жемчужкина Т. В., Носова Т. В.</i> Поиск информативных количественных показателей электромиографического сигнала. Сообщение 1	118
<i>Кораблев Н. М., Овчаренко И. В., Токарев В. В.</i> Применение искусственных иммунных систем в задачах восстановления и оптимизации нелинейных зависимостей	126
<i>Филатов В. А., Семенец Р. В.</i> Идентификация пользователей в системе дистанционного образования на основе программных агентов	130
<i>Колесник А. Б., Колесник Л. В.</i> Интеллектуализация процедур определения координат мобильных терминалов в беспроводных сетях	134
<i>Ребезюк Л. Н., Свичкарь Е. М.</i> Метод сравнения ситуаций при ситуационно-прецедентном принятии решений ...	139
<i>Топал А. С., Зайцев В. Е., Шостак И. В., Белов С. Б.</i> Автоматизация поддержки принятия решений при организации агрегатной сборки в авиастроении на основе специализированной экспертной системы	145
<i>Любченко В. А., Передрий Е. О.</i> Нахождение и нормализация объектов на изображениях, искаженных проективными преобразованиями на примере номерного знака	153
<i>Шатовская Т. Б., Меницкий С. П.</i> Современные информационные распределенные системы баз знаний	157
<i>Репка В. Б., Ключник Е. В., Козопольская А. А., Миняйло И. А.</i> Интеллектуальный агент тестового контроля в дистанционном обучении	162
Рефераты	166
Об авторах	173
Правила оформления рукописей для авторов научно-технического журнала «Біоніка інтелекту»	175
Правила оформления рукописей для авторов научно-технического журнала «Бионика интеллекта»	176
Instructions for authors of manuscripts of the scientific journal «Bionics of Intelligence»	177
Зразок оформлення статті	178
Информационное сообщение	179

НАШИ ЮБИЛЯРЫ

Научно-техническому журналу «Бионика интеллекта» в этом году исполняется 40 лет. По ходатайству министра высшего и среднего специального образования УССР в 1967 году в Харьковском институте радиоэлектроники был основан Всеукраинский научно-технический сборник «Проблемы бионики», впоследствии переименованный в журнал «Бионика интеллекта». Инициаторами его создания были кафедра математического моделирования (зав. кафедрой, доц., канд. техн. наук Юрий Петрович Шабанов-Кушнаренко), кафедра физики (зав. кафедрой, доц., канд. биол. наук Василий Григорьевич Червов – проректор по учебной работе) и кафедра прикладной математики (зав. кафедрой, проф., докт. физ.-мат. наук Владимир Логвинович Рвачев – впоследствии академик АН УССР и чл.-корр. АН СССР).

Первым главным редактором журнала стал В.Л. Рвачев. Затем его сменил чл.-корр. АН СССР, проф., докт. техн. наук Борис Степанович Сотсков – председатель секции бионики Совета по кибернетике при Президиуме АН СССР. Ответственным редактором журнала со дня его основания и до настоящего времени является Ю.П. Шабанов-Кушнаренко. Журнал «Бионика интеллекта» имеет статус ВАКовского издания, публикует статьи на трех языках (украинском, русском и английском).

Термин «бионика» образован из первой части слова «биология» и второй части слова «техника». Он выражает идею взаимосвязи живых систем с искусственными. Бионикой интеллекта называется физико-математическое учение о естественном интеллекте, в первую очередь человеческом, нацеленное на искусственное воспроизведение интеллекта техническими средствами. Бионика интеллекта изучает разум человека и животных, механизмы эволюции живых организмов, строение мозга, работу нейронов и образованных из них нейронных сетей, механизмы зрения, слуха, восприятия, речи, понимания, узнавания, принятия решений, оценивания и т. п. В практическом плане бионика интеллекта подводит научную базу для создания мозгоподобных ЭВМ.

В более далекой перспективе бионика интеллекта готовит почву для киборгизации, которая, используя достижения техники, стремится подчинить разуму и воле человека процессы, наблюдаемые в живой природе. Человечество, опираясь на науку и технику, приближается к тому, чтобы взять в свои руки судьбу всего живого на Земле (в том числе –

и свою собственную). Этому способствуют достижения генетики, молекулярной биологии, геномной инженерии, нанотехники, клонирования, искусственного интеллекта и многих других научных и технических новаций.

Юбилей журнала совпал с 75-летием его ответственного редактора, ветерана войны и труда, заслуженного деятеля науки и техники УССР, профессора, доктора технических наук Ю.П. Шабанова-Кушнаренко, который является моим учителем. Юрий Петрович пришел в наш вуз (в то время он назывался Харьковским горным институтом) в 1949 году, закончил его с отличием по специальности «горная электромеханика». В 1959 году защитил под руководством зав. кафедрой горной механики чл.-корр. АН УССР проф., докт. техн. наук Павла Петровича Нестерова кандидатскую диссертацию, в которой разработал высокопроизводительную систему безуровневого многоканатного рудничного подъема, впоследствии использовавшуюся для глубоких шахт Советского Союза и других стран.

В 1950–62 годах Юрий Петрович, опираясь на труды родоначальника современной науки Рене Декарта, в качестве любителя знакомится с возникшей в то время бионикой интеллекта и пытается проводить в этой области научные исследования. В конце 1962 года наш вуз изменил свой профиль, превратившись в Харьковский институт горного машиностроения, автоматики и вычислительной техники, и Юрий Петрович с марта 1963 года переходит на работу в должности старшего научного сотрудника на только что организованную кафедру промэлектроники и вычислительной техники, которую возглавил кандидат технических наук, доцент Александр Андреевич Волков. С июня того же года эта кафедра делится, и из нее выделяется кафедра вычислительной техники, которую возглавляет В.Л. Рвачев, а Юрий Петрович становится доцентом этой кафедры. В то время он был в числе тех немногих ученых-подвижников нашего вуза, на плечи которых легла непростая задача налаживания



подготовки кадров в области компьютеризации и информатизации.

Кафедра снова делится, и в 1966 году Юрий Петрович становится заведующим кафедрой математического моделирования. В этой должности он работает до 1974 года. Кафедра продолжает делиться, и в дальнейшем Юрий Петрович работает в должности профессора кафедры программного обеспечения ЭВМ. В период 1963—68 годов он под руководством В.Л. Рвачева готовит докторскую диссертацию в области бионики интеллекта под названием «Математическое моделирование некоторых функций человеческого зрения», которую защищает в 1969 году в Институте кибернетики АН УССР по специальности «техническая кибернетика».

За время работы в нашем вузе Юрий Петрович подготовил ряд курсов лекций. К их числу относятся «Теория автоматического регулирования и электронная автоматика», «Основы вычислительной техники», «Электрическое моделирование», «Аналоговые вычислительные машины», «Цифровые вычислительные машины», «Арифметические и логические основы цифровых машин», «Цифровые модели», «Математическое моделирование», «Математическая логика», «Инженерная логика», «Дискретная математика», «Теория интеллекта», «Основы теории искусственного интеллекта», «Логический анализ», «Бионика интеллекта», «Основы научных исследований».

Начиная с 1963 года, Юрий Петрович участвует в подготовке научных кадров. Им создана научная школа, в состав которой сегодня входят свыше 20 докторов и 200 кандидатов наук. Среди них —

его научные «внуки» и «правнуки». Некоторые из его учеников уже создали собственные научные школы.

Научные разработки в области бионики интеллекта начинают давать осязаемые практические результаты. На повестке дня стоит вопрос о промышленных испытаниях и подготовке серийного производства мозгоподобных ЭВМ, превосходящих по производительности компьютеры с программным управлением. Практически используются предложенные бионикой интеллекта методы построения математических моделей механизмов интеллекта, а также методы преобразования их в логические и электронные сети, выполняющие роль процессора мозгоподобной ЭВМ. Освоение мозгоподобных ЭВМ создает предпосылки для проведения практических разработок в области киборгизации. Результаты в области алгебраизации логики, полученные в ходе бионических исследований, открывают заманчивые перспективы создания математических и электронных моделей механизмов, обнаруживаемых биологами в нейронных сетях мозга, которые могут быть использованы при протезировании органов чувств и движения, структур нервной системы человека.

М. БОНДАРЕНКО,
*ректор Харьковского национального
университета радиозлектроники,
академик Академии высшей школы,
заслуженный деятель науки и техники Украины,
доктор технических наук, профессор,
главный редактор журнала
«Бионика интеллекта».*

УДК 681.3.07

НОРМАЛИЗАЦИЯ ИЗОБРАЖЕНИЙ: ДОСТИЖЕНИЯ И НЕРЕШЕННЫЕ ЗАДАЧИ

Е. П. Путятин

ХНУРЭ, г. Харьков, Украина

Приведены новые результаты по нормализации аффинных, проективных и нелинейных преобразований. Обсуждаются проблемные вопросы нормализации общих преобразований двумерных изображений. В сравнительном аспекте рассматриваются основные математические методы нормализации: метод моментов, интегральной и одномерной последовательной корреляции. Описан подход к нормализации сложных групп, включая проективную, путем синтеза частичных нормализаторов, соответствующих подгруппам разложения.

КОМПЬЮТЕРНОЕ ЗРЕНИЕ, ИНТЕЛЛЕКТ, ИЗОБРАЖЕНИЕ, НОРМАЛИЗАЦИЯ, РАСПОЗНАВАНИЕ, ПРЕОБРАЗОВАНИЕ, АФФИННЫЕ, НЕЛИНЕЙНЫЕ, ЯРКОСТЬ, КОРРЕЛЯЦИЯ, МОМЕНТЫ.

Введение

Проблеме построения систем компьютерного зрения (СКЗ) уделяется все возрастающее внимание. Одним из основных направлений совершенствования таких систем является придание им свойств человеческого зрения*. Бионика зрительно-интеллекта как наука лежит в основе СКЗ транспортных и технологических роботов различного назначения. В широком смысле интеллектуальные СКЗ решают две взаимно дополняющие задачи: правильное восприятие образа и его анализ, понимание и распознавание. Задача нормализации заключается в обеспечении правильного восприятия, при котором создается некоторый стандартный вид изображения: его размеры, ориентация, распределение яркости. В принципе, стандарты или эталоны могут быть разными для глаза человека и робота, хотя в целом требования часто совпадают. Нормализация имеет важное практическое значение не только для распознавания изображений, облегчая сравнение с эталонами и увеличивая резерв времени на принятие решений на заключительном этапе. Существуют и другие области, непосредственно не связанные с распознаванием. Например, сопровождение подвижных объектов (охранная сигнализация, компьютерные игры и т. д.). Важное значение имеет «выравнивание» поля зрения при наблюдении реальных сцен из замкнутого пространства, например, из кабины пилота.

Ограничимся одиночными плоскими изображениями, заданными на однородном фоне, т. е. исключим задачу сегментации отдельных изображений из многоплановых сцен. Несмотря на это существенное упрощение, проблема нормализации, как мы

увидим ниже, остается сложной. Это связано с тем, что входные изображения отличаются от эталонных геометрическими преобразованиями: аффинными, проективными, нелинейными, преобразованиями яркости. Несмотря на такие отличия, человеческое зрение уверенно справляется со многими задачами нормализации изображений, чего еще нельзя сказать о современном компьютерном зрении.

Рассмотрим проблему последовательно. Вначале остановимся на моделях входных сигналов, затем исследуем возможности их нормализации.

1. Математические модели входных сигналов

В табл. 1 приведена классификация типов преобразований, с которыми сталкивается человек в повседневной жизни. Преобразования условно разделены на пять классов. Буквенные обозначения приведены в таблице справа, номера преобразований — слева. Преобразования из класса I являются базовым для других классов. Они образуют группы, причем преобразования 1–3 и 8–10 образуют аддитивные группы со значением единичного элемента, равным нулю, остальные (4–7) — мультипликативные группы с параметром единица.

Многопараметрические преобразования из класса II также образуют группы, элементами которых являются матрицы, составленные из параметров преобразований, так что единичному элементу соответствует единичная матрица. Удобно использовать матрицы (3×3) , так что аффинные преобразования 21 характеризуются матрицей вида A , множество которых образуют группу G_a .

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{pmatrix}, \det |A| \neq 0, \quad (1)$$

где a_j — параметры преобразования ($i=1,2, j=\overline{1,3}$).

* Исследования по математическому моделированию зрения, в том числе нормализации начаты в университете радиоэлектроники в 1963 г. [1] по инициативе моего учителя проф. Ю. П. Шабанова-Кушнарченко, которому посвящается данная работа.

Математические модели входных сигналов интеллектуальных СКЗ

I. Однопараметрические преобразования			Обозначение
1.	$x \rightarrow x + a \cos \gamma, y \rightarrow y + a \sin \gamma$, где: a — параметр	Смещения вдоль направления $(\cos \gamma, \sin \gamma)$	$G_{c,\gamma}$
2.	$x \rightarrow x + \alpha, y \rightarrow y$, где: α — параметр	Горизонтальные смещения	$G_{c,x}$
3.	$x \rightarrow x, y \rightarrow y + \beta$, где: β — параметр	Вертикальные смещения	$G_{c,y}$
4.	$x \rightarrow kx, y \rightarrow ky$, где: $k \neq 0$ — параметр	Подобия изменения масштаба	G_m
5.	$x \rightarrow kx, y \rightarrow y$, где: $k \neq 0$ — параметр	Горизонтальные растяжения	$G_{m,x}$
6.	$x \rightarrow x, y \rightarrow ky$, где: $k \neq 0$ — параметр	Вертикальные растяжения	$G_{m,y}$
7.	$x \rightarrow qx, y \rightarrow q^{-1}y$, где: $q \neq 0$ — параметр	Гиперболические растяжения	G_q
8.	$x \rightarrow x \cos \gamma + y \sin \gamma,$ $y \rightarrow -x \sin \gamma + y \cos \gamma$, где: γ — параметр	Повороты	G_u
9.	$x \rightarrow x + hy, y \rightarrow y$, где: h — параметр сдвига	Косые сдвиги (вдоль оси OX)	$G_{h,x}$
10.	$x \rightarrow x, y \rightarrow y + hx$, где: h — параметр сдвига	Косые сдвиги (вдоль оси OY)	$G_{h,y}$
II. Многопараметрические преобразования			
11.	$x \rightarrow x + \alpha, y \rightarrow y + \beta$, где: α, β — параметры смещения	Двумерные смещения	G_c
12.	$x \rightarrow kx + \alpha, y \rightarrow y$, где: k, α — параметры, $k \neq 0$	Одномерные смещения и растяжение по оси OX	$G_{c,m,x}$
13.	$x \rightarrow x, y \rightarrow ky + \beta$, где: k, β — параметры, $k \neq 0$	Одномерные смещения и растяжение по оси OY	$G_{c,m,y}$
14.	$x \rightarrow kx + \alpha, y \rightarrow ky + \beta$, где: k, α, β — параметры, $k \neq 0$	Смещения и изменения масштаба	$G_{c,m}$
15.	$x \rightarrow k_1x, y \rightarrow k_2y$, где: k_1, k_2 — параметры, $k_1 \neq 0, k_2 \neq 0$	Неоднородные изменения масштаба	G_d
16.	$x \rightarrow k_1x + \alpha, y \rightarrow k_2y + \beta$, где: k_1, k_2, α, β — параметры, $k_1 \neq 0, k_2 \neq 0$	Смещения и неоднородные изменения масштаба	$G_{c,d}$
17.	$x \rightarrow x \cos \gamma + y \sin \gamma + \alpha,$ $y \rightarrow -x \sin \gamma + y \cos \gamma + \beta,$ где: γ, α, β — параметры	Смещения и повороты (группа движений, или Евклидова группа)	$G_{c,u}$
18.	$x \rightarrow kx \cos \gamma + ky \sin \gamma,$ $y \rightarrow -kx \sin \gamma + ky \cos \gamma,$ где: γ, k — параметры	Изменения масштаба и повороты	$G_{m,u}$
19.	$x \rightarrow kx \cos \gamma + ky \sin \gamma + \alpha,$ $y \rightarrow -kx \sin \gamma + ky \cos \gamma + \beta,$ где: γ, k, α, β — параметры	Смещения, изменения масштаба, повороты (метрическая группа)	$G_{m,c,u}$
20.	$x \rightarrow a_{11}x + a_{12}y, y \rightarrow a_{21}x + a_{22}y,$ где: $a_{11}, a_{12}, a_{21}, a_{22}$ — параметры $\det \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0$	Центроаффинная группа (неоднородные изменения масштаба, косые сдвиги, повороты и отражения)	G_a^c
21.	$x \rightarrow a_{11}x + a_{12}y + \alpha, y \rightarrow a_{21}x + a_{22}y + \beta,$ где: $a_{11}, a_{12}, a_{21}, a_{22}$ — параметры $\det \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0$	Группа аффинных преобразований (неоднородные изменения масштаба, косые сдвиги, повороты, отражения и смещения)	G_a
III. Нелинейные преобразования			
22.	$x \rightarrow \frac{x}{(\lambda-1)y+1}, y \rightarrow \frac{\lambda(y)}{(\lambda-1)y+1},$ где: λ — параметр преобразования	Перспектива (гиперболическая гомология)	G_r
23.	$x \rightarrow \frac{x}{\lambda y+1}, y \rightarrow \frac{y}{\lambda y+1},$ где: λ — параметр преобразования	Перспектива (параболическая гомология)	G_p

24.	$x \rightarrow \frac{a_{11}x + a_{12}y + a_{13}}{a_{31}x + a_{32}y + a_{33}}, y \rightarrow \frac{a_{21}x + a_{22}y + a_{23}}{a_{31}x + a_{32}y + a_{33}}$, где: a_{ij} — параметры ($i, j = 1, 2, 3$)	Проективные преобразования (общий случай)	G
25.	$x \rightarrow e^{ay}x, y \rightarrow y; x \rightarrow x, y \rightarrow e^{ax}y$, где: a — параметр преобразования	Преобразование масштаба с переменным коэффициентом	G_s
26.	$x \rightarrow a_1x^2 + a_2y^2 + a_3, y \rightarrow a_4x^2 + a_5y^2 + a_6$, где: a_i — параметры ($i = \overline{1,6}$)	Квадратичные преобразования (дисторсия линз)	K
27.	$x \rightarrow a_1x^2 + a_2xy + a_3y^2 + a_4x + a_5y + a_6$, $y \rightarrow a_7x^2 + a_8xy + a_9y^2 + a_{10}x + a_{11}y + a_{12}$, где: a_i — параметры ($i = \overline{1,12}$)	Полные квадратичные преобразования (кривизна Земли)	Z
IV. Непараметрические нелинейные преобразования			
28.	$x \rightarrow x, y \rightarrow y; x \rightarrow -x, y \rightarrow -y$; $x \rightarrow x, y \rightarrow -y; x \rightarrow -x, y \rightarrow y$	Зеркальные отражения	G_s
29.	$x \rightarrow \pm \frac{x}{y}, y \rightarrow \pm \frac{1}{y}; x \rightarrow \pm \frac{y}{x}, y \rightarrow \pm \frac{1}{x}$; $x \rightarrow \pm \frac{1}{x}, y \rightarrow \pm \frac{y}{x}; x \rightarrow \pm \frac{1}{y}, y \rightarrow \pm \frac{x}{y}$	Ортогональные проективные преобразования	G_o
30.	$x \rightarrow f_1(x), y \rightarrow f_2(y); (x \rightarrow x^2, y \rightarrow y^2)$; $x \rightarrow f_1(x, y), y \rightarrow f_2(x, y)$	Общий вид преобразования координат	
V. Преобразования яркости			
31.	$B(x, y) = kB_0(x, y) + c$, где $B(x, y)$ — функция яркости входного изображения, $B_0(x, y)$ — функция яркости эталона, k, c — параметры группы	Постоянный фильтр и подсветка	G_k, G_c, G_{kc}
32.	$B(x, y) = K(x, y)B_0(x, y) + C(x, y)$, где: $K(x, y), C(x, y)$ — мультипликативная и аддитивная составляющие искажений	Общий случай преобразований яркости	
33.	$B(x, y) = K(x, y)B_0[T(x, y)] + C(x, y)$, где: $T(x, y)$ — геометрические преобразования	Комбинированный случай (эквализация гистограмм, борьба с помехами в условиях геометрической нормализации)	

Преобразованиям 11–20 соответствуют различные значения параметров матрицы (1), например, для смещений 11: $a_{13} = \alpha, a_{23} = \beta, a_{11} = a_{22} = 1, a_{21} = a_{12} = 0$; для косоуго сдвига 9: $a_{12} = h, a_{11} = a_{22} = 1, a_{21} = a_{13} = a_{23} = 0$. Зеркальные отражения 28 (симметрия) образуют дискретную группу G_s , которая естественным образом входит в базовые преобразования аффинной группы в предположении, что параметры a_{11}, a_{22} могут иметь любой знак (+, -).

Наиболее сложными преобразованиями в рамках групповой модели G являются проективные преобразования 24, матрица которых имеет вид:

$$P = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \det(P) \neq 0, a_{ij} (i, j = \overline{1,3}). \quad (2)$$

Составляющими проективной группы являются аффинные, ортогональные проективные преобразования 29, 30 и преобразования перспективы 22, 23.

Человек может распознавать в определенных пределах изображения, спроектированные на нелинейную поверхность, например, Землю, компенсировать нелинейные искажения, вызванные дисторсией линз, и т. д. В какой-то степени они моделируются преобразованиями 25–27 (из них только преобразования 25 образуют группу). Нелинейные изменения пространства восприятия сцены описываются преобразованием 32. Простейшие преобразования яркости 31 хорошо описывают изменение яркости объекта, когда Солнце зашло за облако, а также наличие однородных фильтров и подсветки в СКЗ. Такие преобразования также образуют группы [2].

Последняя графа таблицы 1 описывает отличия входного изображения $B(x, y)$ от эталонного $B_0(x, y)$, включая шумы и локальные помехи $C(x, y)$, неравномерную чувствительность фотокамеры или сенсоров по пространству $K(x, y)$.

С геометрической точки зрения аффинные преобразования первого и второго классов не требуют

особых комментариев. На рис. 1. приведена иллюстрация косо́го сдвига вдоль оси абсцисс. Проективные преобразования G (рис. 2) возникают в результате поворота объекта относительно оси, перпендикулярной оси наблюдения. Перспектива G_p обычно ассоциируется с уходящими вдаль рельсами железной дороги (рис. 3). Понятно, что численное моделирование проективной группы будет не совсем адекватно теоретическим моделям ввиду невозможности реализации точек на бесконечности. Проблема восприятия и нормализации возникает также при анализе ортогональной подгруппы G_0 проективной группы G . На рис. 4, 5 показано, когда изменяется положение точек при таких преобразованиях до неузнаваемости. С этой группой связаны дополнительные ограничения при нормализации. На рис. 6 проиллюстрировано преобразование 25 нелинейной группы G_3 .



Рис. 1. Косой сдвиг $B_0(x, y) = B(x + y \operatorname{tg}(\alpha), y)$, где α — угол сдвига

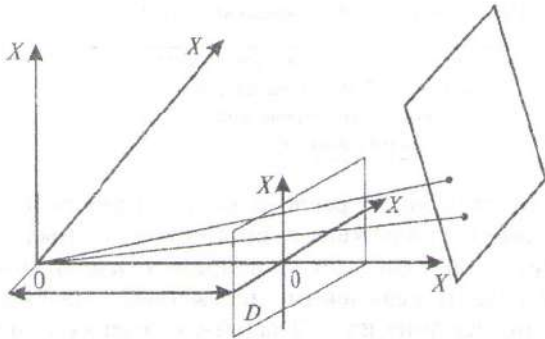


Рис. 2. Возникновение проективных преобразований

При нормализации важное значение имеет изучение свойств преобразований, причем внимание уделяется не только известным геометрическим свойствам типа инвариантов, скажем, для аффинной группы и ее подгрупп. Новым является изучение возможности применения конкретного преобразования в составе суперпозиции, что важно при синтезе нормализаторов. С этой точки зрения преобразования могут быть коммутативными в данной суперпозиции, перестановочными или неперестановочными. Для перестановочных, в отличие от коммутативных, допускается изменение величины параметра, но не самого вида преобразования.

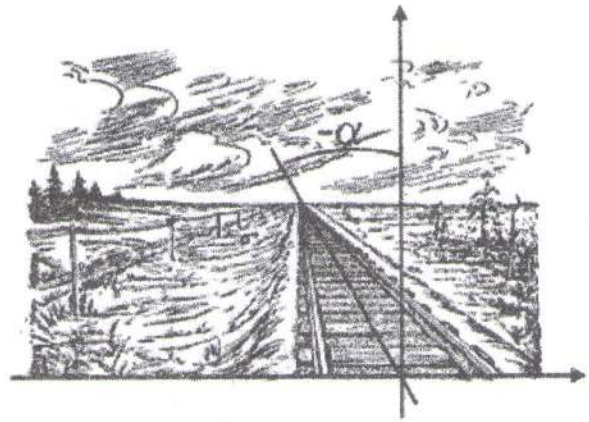


Рис. 3. Пример преобразования перспективы вдоль оси с углом наклона α

$$B_0(x, y) = B \left(\frac{x}{(\lambda \sin(-\alpha) + \lambda \cos(-\alpha))y + 1}, \frac{y}{(\lambda \sin(-\alpha) + \lambda \cos(-\alpha))y + 1} \right),$$

где λ — параметр перспективы;
 B, B_0 — входное и эталонное изображения

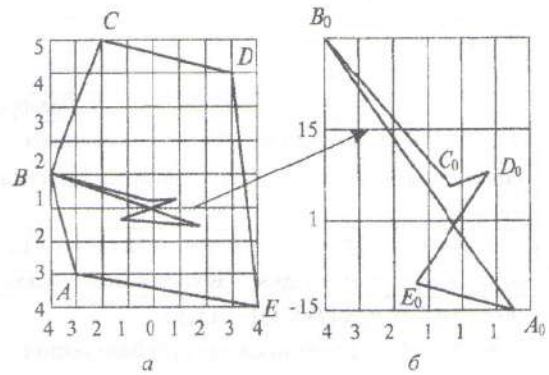


Рис. 4

a — пример искажения фигуры $ABCDE$, внутри которой находится начало координат, преобразованием $X = x/y, Y = 1/y$ в фигуру $A_0B_0C_0D_0E_0$;
 b — фигура $A_0B_0C_0D_0E_0$, увеличенная в 5 раз

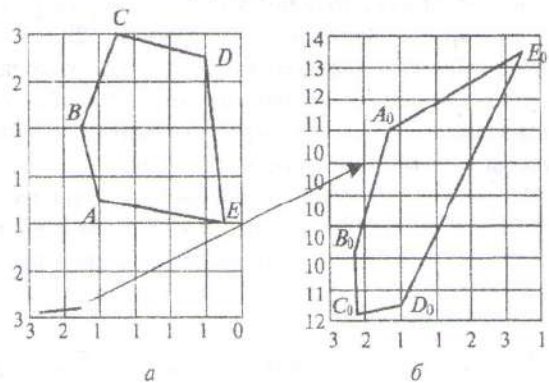


Рис. 5

a — пример искажения фигуры $ABCDE$, расположенной в первой четверти, преобразованием $X = x/y, Y = 1/y$ в фигуру $A_0B_0C_0D_0E_0$;
 b — та же фигура, увеличенная в 70 раз

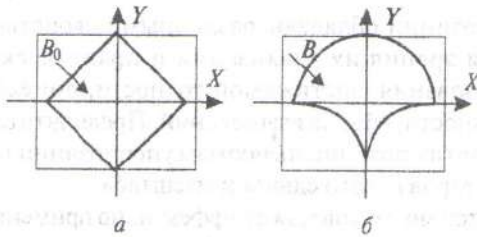


Рис. 6. Нелинейное масштабирование

В нижних рядах таблицы классификации приведены входные сигналы общего вида, соответствующие реальным сценам, для которых нормализация — это получение стандартного фона и контраста (группы G_k, G_c, G_{kc}). Более того, процедуры подавления помех, выравнивания гистограмм и т. д. кажутся вполне естественными нормализаторами.

2. Математические модели нормализации

Прежде всего обратим внимание на интегральный подход, основанный на преобразовании всего поля изображения. Например, модуль преобразования Фурье обеспечивает автоматическое центрирование, т. е. компенсацию параметров α, β группы G_c , преобразование Меллина компенсирует преобразование масштаба.

Однако этот подход имеет ограниченную область применения, т. к. не вычисляет величины параметров отличия эталонных от входных изображений. Кроме того, цифровая реализация преобразований Меллина весьма затруднительна.

Другой подход к построению нормализаторов основан на непосредственном вычислении параметров и действия обратных преобразований, построенных на их базе. Несмотря на то, что обратные преобразования формально существуют для всех моделей входных сигналов из табл. 1, полностью обоснованные модели нормализации возможны только для преобразований, образующих группы.

В самом деле, каким бы ни было сложным преобразование входного изображения, если множество этих преобразований образует группу, то всегда есть тождественное преобразование, характеризующее эталон, и существует обратное преобразование, с помощью которого можно получить этот эталон. Результат последовательного выполнения двух преобразований также принадлежит той же группе. Такими свойствами обладают все преобразования из табл. 1, помеченные справа буквой G (с индексом и без).

Насколько обоснованными являются эти рассуждения, видно из такого примера. Преобразования G_d с неодинаковым изменением масштаба по осям и преобразования поворота G_u сами по себе образуют группы, однако их совместное действие группы не образует. Следовательно, эталон, опре-

деленный единицами групп ($\gamma=0, k_1=k_2=1$) не существует, так как последовательное выполнение нескольких таких преобразований выведет итоговое изображение за пределы трех параметров как основной характеристики исходных групп. Понадобится еще один параметр (например, косой сдвиг или другой поворот), но тогда получается группа уже с четырьмя параметрами (центроаффинная G_a^c). Заметим, что суперпозиция поворота и одинакового масштаба образует группу, и можно говорить об эталоне и нормализации.

Аппарат теории групп позволяет с единых позиций рассмотреть задачу нормализации входных сигналов, приведенных в табл. 1.

Пусть M — множество входных изображений B , M_0 — множество эталонов B_0 , G — группа преобразований, $g \in G$ — элемент группы, т. е. конкретное преобразование. Рассматривается действие группы G на множестве входных изображений.

Оператором нормализации F относительно действия группы G называется отображение $F: M \rightarrow M_0$, для которого существует $g \in G$ такое, что $F(B) = B_0$ для любого $B \in M$.

При этом учитывается, что различные изображения $B \in M$ образуют классы эквивалентности, если $B = gB_0, g \in G$. Общий вид оператора нормализации обоснован в [2, 4] и имеет вид:

$$F(B) = \Phi(B)B, \quad \Phi: M \rightarrow G, \Phi(B) = g^{-1}, g \in G, B \in M. \quad (3)$$

Отображение $\Phi: M \rightarrow G$ представляет собой вещественный функционал для однопараметрических групп и вектор-функцию для многопараметрических.

Например, для параллельных смещений (группа G_c), когда входной сигнал определяется классом эквивалентности

$$B(x, y) = gB_0(x, y) = B_0(x - \alpha, y - \beta),$$

имеем:

$$F(B) = \Phi(B)B = B(x + \Phi_1(B), y + \Phi_2(B)), \quad (4)$$

где $\Phi: M \rightarrow G_c$ определяется двумя вещественными функционалами $\Phi_1(B)$ и $\Phi_2(B)$.

Физический смысл нормализатора (4) заключается в компенсации смещений α, β , т. е. центрировании. При этом $\Phi_1(B) = \alpha^{-1}, \Phi_2(B) = \beta^{-1}$.

Для преобразований перспективы G_p (параболическая гомология) входной сигнал определяется классом эквивалентности

$$B(x, y) = gB_0(x, y) = B_0\left(\frac{x}{\lambda y + 1}, \frac{y}{\lambda y + 1}\right),$$

а нормализатор имеет вид:

$$F(B) = \Phi(B)B = B \left(\frac{x}{\Phi(B)y+1}, \frac{y}{\Phi(B)y+1} \right), \quad (5)$$

где Φ — функционал, вычисляющий параметр λ .

Аналогично можно выписать модели нормализаторов и для всех других типов входных сигналов, образующих группы, однако понятно, что основная трудность ложится на поиск подходящих отображений $\Phi(B)$.

3. Аффинная нормализация

В принципе существует два подхода к построению вектор-функции $\Phi(B)$: параллельный и последовательный. Параллельный подход основан на использовании интегральных моментов в форме линейных и нелинейных функционалов. Зависимость между моментами для эталонных изображений и преобразованных (искаженных) выражается в виде системы уравнений, решение которой позволяет определить параметры a_{ij} матрицы (1). Параллельная нормализация хорошо зарекомендовала себя в режиме определения малых приращений параметров с целью управления, например, рамкой слежения в системах динамического видеонаблюдения [5].

Последовательный подход обеспечивает вычисление параметров преобразований, а не их приращений, и лежит в основе целого ряда технических устройств, именуемых координаторами положения объектов в пространстве. Теоретической предпосылкой вычисления параметров являются найденные различные способы представления аффинной группы в виде суперпозиции подгрупп. Если ранее из классической геометрии для центроаффинной группы следовало только разложение: поворот (U_1) — преобразование масштаба (D) — поворот (U_2), то есть $G_a^c = G_u G_d G_{u_2}$, то теперь найден ряд других возможных разложений центроаффинной группы:

$$\begin{aligned} &G_u G_d G_{h,y}, G_u G_d G_{h,x}, G_d G_u G_{h,x}, \\ &G_d G_u G_{h,y}, G_d G_{h,x} G_u, G_d G_{h,y} G_u, \quad (6) \\ &G_d G_u G_{m,x}, G_d G_u G_{m,y}, G_d G_{h,x} G_{h,y}, \end{aligned}$$

$$G_d G_{h,y} G_{h,x}, G_{h,x} G_d G_{h,y}, G_{h,x} G_d G_{h,y}, G_{h,y} G_d G_{h,x},$$

где обозначения подгрупп соответствуют обозначениям, приведенным в табл. 1.

В основе последовательной нормализации полной аффинной группы лежит синтез частичных нормализаторов, соответствующих подгруппам разложений (6) и нормализатора смещений F_c , согласно (4). Как сами частичные нормализаторы, так и их

суперпозиции обладают различными свойствами с точки зрения их реализации и практического использования: достижимой точности, помехозащищенности, объема вычислений. Предпочтительными с этих позиций являются суперпозиции нормализаторов косоугольного сдвига и масштаба.

Метод моментов может эффективно применяться не только для параллельной, но и для последовательной нормализации, однако его сравнительно низкая помехозащищенность вынуждает искать другие подходы.

Разработана модификация метода частных корреляций, именно корреляция вдоль прямой, образованной пересечением полутонового изображения и перпендикулярной к нему плоскости. В этом случае корреляция по плоскому изображению заменяется корреляцией вдоль прямой. Задача состоит в поиске максимального сходства распределения яркости вдоль прямых на эталонном и входном изображениях путем перебора с некоторым шагом всех положений прямых [6].

Для центрированных изображений, когда начало координат совпадает с обобщенным центром тяжести, поиск параметров аффинных преобразований сводится к последовательной корреляции двух пар прямых эталонного и входного изображений по масштабу.

Сравнительный анализ экспериментальных данных центроаффинной нормализации позволил сделать следующие выводы по трем методам: интегральной корреляции (ИК) (по всем четырем параметрам), одномерной корреляции (ОК) и применении моментов. Так, метод одномерной корреляции превосходит по быстродействию метод интегральной корреляции на три порядка, но на порядок уступает методу моментов. По помехозащищенности правильное распознавание объектов с вероятностью 0,96 достигалось для случайного шума методом ИК при $\mu > 3$, методом ОК при $\mu > 7$, методом моментов при $\mu > 10$, где μ — отношение сигнал/шум \bar{B}/δ , \bar{B} — средняя яркость изображения, δ — среднеквадратическое отклонение.

Иная картина наблюдается при действии локальных помех. Пусть $\mu_l = \frac{S_n}{S_c} \cdot 100$, где S_n — площадь помехи на изображении, S_c — площадь полезного сигнала. Метод ОК позволяет производить распознавание с вероятностью 0,96 при $\mu_l \leq 24\%$, метод ИК работает только до 9%, а метод моментов — до 5%. Экспериментальные данные подтверждают целесообразность применения метода моментов в системах нормализации реального времени при слежении за объектами. В то же время метод одномерных корреляций применим для решения задач распознавания изображений в условиях аффинных

преобразований и действия случайных и даже локальных помех.

Большой интерес представляют методы математической морфологии для определения параметров преобразований в задачах нормализации.

На данный момент успешным оказался опыт применения морфологии для аффинных преобразований. В основу положена возможность представления полутоновых изображений сечениями, причем с позиций функционального анализа сечения можно конструировать по диапазонам яркости с любой точностью приближения к исходному изображению и даже с переменным порогом. Сечения — бинарные образы, к ним применимы операции алгебры Минковского. С помощью таких операций разработаны методы выделения особых точек: крайних точек и скелетов для силуэтных изображений [7]. На рис. 7, 8 показаны примеры работы программ на реальных изображениях.

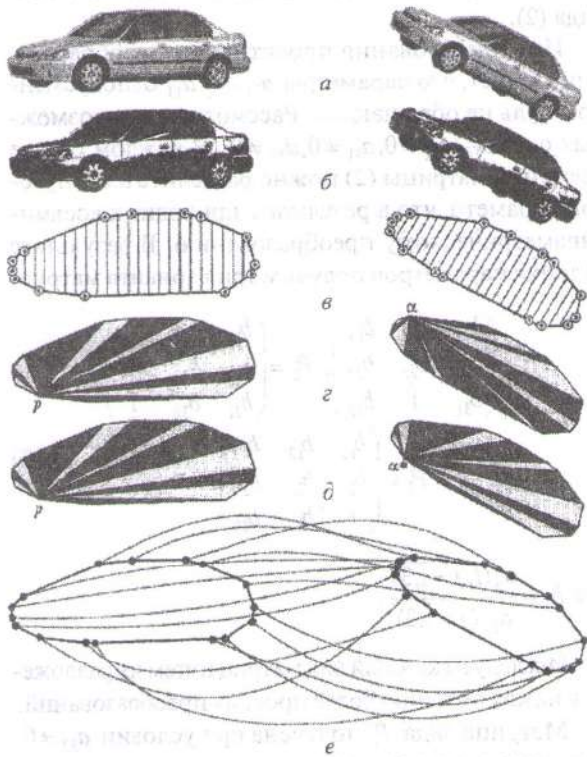


Рис. 7. Метод крайних точек:
 а — эталон и исходное изображение; б — их бинарные сечения; в — выделенные крайние точки;
 г, д — иллюстрация поиска соответствующих точек;
 е — результат поиска

После выделения особых точек (крайних на рис. 7 и точек скелетона на рис. 8) возникает задача поиска соответствий особых точек эталона и входного изображения. Решается она путем анализа возможных подстановок из множеств особых точек и с

учетом известного из геометрии инварианта постоянства площади объекта при аффинных преобразованиях. В конечном счете, переопределенная система линейных уравнений относительно параметров аффинных преобразований позволяет с высокой точностью (методом наименьших квадратов) находить параметры аффинных преобразований.

Еще предстоит отыскать возможности применения методов морфологии для нормализации проективных преобразований.

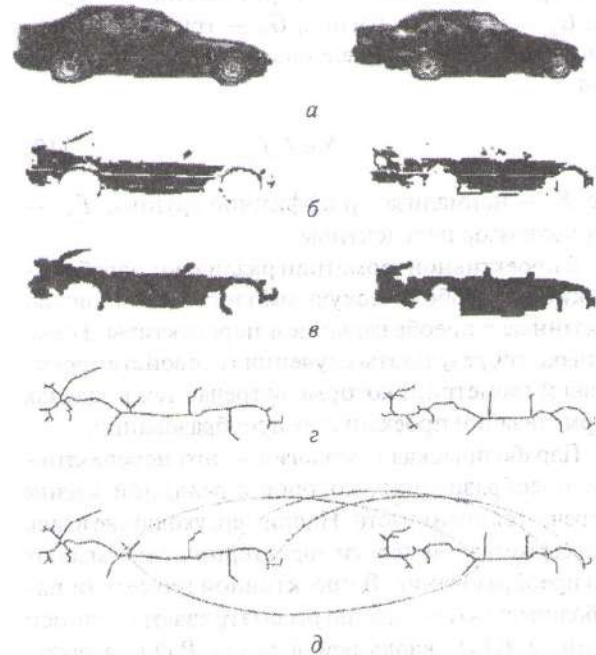


Рис. 8. Метод скелетов:
 а — эталон и исходное изображение; б — сечения;
 в — улучшенные сечения; г — скелеты;
 д — результат нахождения соответствующих точек

4. Нормализация проективных преобразований

Проективную группу G также можно представить в виде композиции более простых, основными из которых являются аффинная, ортогональная и группа перспективы. Класс эквивалентности входных сигналов задается соотношением:

$$B(x, y) = B_0 \left(\frac{a_{11}x + a_{12}y + a_{13}}{a_{31}x + a_{32}y + a_{33}}, \frac{a_{21}x + a_{22}y + a_{23}}{a_{31}x + a_{32}y + a_{33}} \right). \quad (7)$$

Частным случаем (подгруппой) проективной группы является центропроективная группа G_0 , определяемая классом эквивалентности:

$$B(x, y) = B_0 \left(\frac{a_{11}x + a_{12}y}{a_{31}x + a_{32}y + 1}, \frac{a_{21}x + a_{22}y}{a_{31}x + a_{32}y + 1} \right). \quad (8)$$

Существуют два подхода к синтезу нормализатора полной проективной группы [8]. Один из них основан на разложении $G = \bar{G}C'$, где \bar{G} — центро-проективная группа, C' — группа центрирования ($a_{13} = 0, a_{23} = 0$). Соответствующий нормализатор имеет вид:

$$F = F_0 F'_C, \quad (9)$$

где F_0 — нормализатор группы \bar{G} , F'_C — нормализатор проективного центрирования.

Другой подход использует разложение $G = G_a G_p$, где G_a — аффинная группа, G_p — группа перспективы. Нормализатор представляется как суперпозиция:

$$F = F_a F_p, \quad (10)$$

где F_a — нормализатор аффинной группы, F_p — нормализатор перспективы.

В проективной геометрии различают параболическую и гиперболическую гомологии как понятия, связанные с преобразованием перспективы. Ниже приводятся результаты изучения тех свойств проективной геометрии, которые встречаются в задачах нормализации проективных преобразований.

Параболическая гомология — это перспективное преобразование, которое в реальной жизни встречается очень часто. Например, уходящие вдаль рельсы железной дороги характерно представляют это преобразование. В проективной геометрии параболическую гомологию рассматривают вдоль оси ординат $P_y(\lambda)$, вдоль оси абсцисс $P_x(\lambda)$, а также вдоль прямой с угловым коэффициентом $P_\alpha(\lambda)$ (рис. 3). Очевидно, что

$$P_x(\lambda) = U(90)P_y(\lambda),$$

где $U(90)$ — поворот на угол 90 градусов.

Между гиперболической и параболической гомологиями существует прямая зависимость по параметрам сжатия $G_a(1, \lambda)$. Сказанное позволяет рассматривать только параболическую гомологию, которую будем называть просто перспективой.

Исследования перспективных преобразований позволили сформулировать критерий присутствия проективных преобразований на изображении.

Если энергия изображения изменяется пропорционально изменению площади, то на изображении присутствуют преобразования, отличные от аффинных. Если эти преобразования вписываются в проективную модель, то перераспределение энергии происходит вдоль оси преобразования перспективы.

На основании этого критерия функционал Φ , определяющий параметр перспективы, для нормализатора (5) можно определить в виде:

$$\lambda = \Phi = \frac{E_B}{S} - \frac{E_{B_0}}{S_0} = 0, \quad (11)$$

где $E_B = \iint B(x, y) dx dy$ — энергия изображения $B(x, y)$, S — площадь изображения $B(x, y)$,

$E_{B_0} = \iint B_0(x, y) dx dy$ — энергия изображения $B_0(x, y)$, S_0 — площадь изображения $B_0(x, y)$.

В (11) параметр λ задан неявным образом. Процедура нормализации заключается в многократном преобразовании перспективы, и при этом после каждого шага вычисляется функционал Φ . Преобразования продолжают до тех пор, пока функционал не будет равен нулю. Этот принцип осуществления нормализации соответствует построению нормализатора следящего типа [2].

Вернемся теперь к общей постановке вопроса, а именно, разработке математической модели нормализации в рамках модели восприятия, с матрицей вида (2).

Из существования проективных преобразований следует, что параметры a_{31}, a_{32}, a_{33} одновременно в нуль не обращаются. Рассмотрим три возможных случая: $a_{32} \neq 0, a_{31} \neq 0, a_{33} \neq 0$. В каждом случае элементы матрицы (2) можно разделить на ненулевой параметр, что в результате приводит к восьми-параметрическому преобразованию. В результате деления параметров получим три варианта матриц:

$$P_1 = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & 1 & b_{33} \end{pmatrix}, P_2 = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & 1 \end{pmatrix}, \quad (12)$$

$$P_3 = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ 1 & b_{32} & b_{33} \end{pmatrix},$$

где $b_{ij} = \frac{a_{ij}(i, j = \overline{1, 3})}{a_{3j}(j = \overline{1, 3})}$.

Исследуем каждый вид матриц путем их разложения на композицию более простых преобразований.

Матрица вида P_1 получена при условии $a_{32} \neq 0$. Она допускает разложение:

$$P_1 = ANK, \quad (13)$$

где $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{pmatrix}, K = \begin{pmatrix} 1 & 0 & 0 \\ a_{31} & 1 & a_{33} \\ 0 & 0 & 1 \end{pmatrix}$ — матри-

цы аффинных преобразований; $N = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$ —

матрица ортогонального непараметрического преобразования вида $B(x, y) = B_0 \left(\frac{x}{y}, \frac{1}{y} \right)$ (группа G_0 табл. 1).

Исследования показали, что это преобразование вносит столь значительные искажения в изображение, что нормализация практически невозможна. Подобное разложение можно привести и для матрицы P_3 , которое также будет содержать ортогональное проективное преобразование. В итоге — те же выводы, что и для матрицы P_1 .

Обратим внимание на отличительные свойства вычислительной проективной геометрии. Проективные преобразования в задачах нормализации чаще всего рассматриваются в некоторой замкнутой области. Это накладывает свою специфику на свойства проективных преобразований, отличных от классических. Так, из проективной геометрии известно, что любые собственные точки могут перейти при преобразовании в несобственные. В задачах же нормализации изучают в основном конечномерные объекты. Следовательно, изучаемый объект не должен выходить из поля зрения.

Рассмотрим координатное представление полной проективной группы. В зависимости от выбора начала координат восьмипараметрическое представление проективного искажения будет иметь один из трех видов (12). Из определения преобразований P_1 и P_3 следует, что они не существуют в точке $(0, 0)$, откуда вытекает следующее утверждение: для того, чтобы проективное преобразование в некоторой замкнутой области зрения имело вид P_2 , достаточно, чтобы начало координат находилось внутри этой области. Далее обратим внимание на то, как искажаются параллельные и пересекающиеся прямые реальных объектов. Так, параллельные рельсы железной дороги при восприятии глазом человека кажутся пересекающимися вдали, то есть теряют свою параллельность. Однако в машинном поле зрения они не будут иметь общей точки.

Другой случай: пусть исходные прямые не параллельны, но и не пересекаются в реальном поле зрения. При проективных преобразованиях они могут пересекаться, и фигуры, вообще говоря, переходят в другие фигуры, что сильно усложняет нормализацию и распознавание. В связи с этим необходимо вводить ограничения на проективные преобразования внутри поля зрения. Так, увеличение размеров поля зрения D и объекта B на порядок приводит к тому, что интервал возможных значений параметра перспективы уменьшается на порядок. Например, для поля зрения D размера $2a \times 2a$ и объекта B размера $2b \times 2b$ интервал допустимых значений параметра перспективы λ при изменении поля зрения D

и объекта B (чтобы ни одна точка объекта не вышла за пределы поля зрения) должен лежать в пределах:

- 1) $a = 10, b = 5, \quad \lambda \in [-10^{-1}, 10^{-1}]$,
- 2) $a = 100, b = 50, \quad \lambda \in [-10^{-2}, 10^{-2}]$,
- 3) $a = 1000, b = 500, \quad \lambda \in [-10^{-3}, 10^{-3}]$.

Обратимся к матрице P_2 . Нами получены два возможных разложения матрицы:

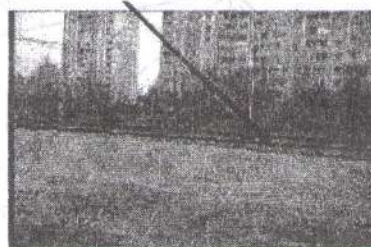
$$\begin{aligned} P_2 &= AP_x(h_1)P_y(h_2), \\ P_2 &= AP_\alpha(h), \end{aligned} \quad (14)$$

где A — аффинная матрица; $P_x(h_1), P_y(h_2), P_\alpha(h)$ — преобразования перспективы с параметрами h_1, h_2, h .

Именно эти разложения позволяют представить нормализатор проективной группы в виде (9) и (10).

Нормализатор вида (9) целесообразно применять, когда есть возможность центрировать входное и эталонное изображения, а нормализатор (10) — когда возможно установить ось перспективного преобразования. На рис. 9 приведен пример реальной ситуации, когда эффективен нормализатор (10).

Направление оси преобразования перспективы



Объект, за которым ведется наблюдение

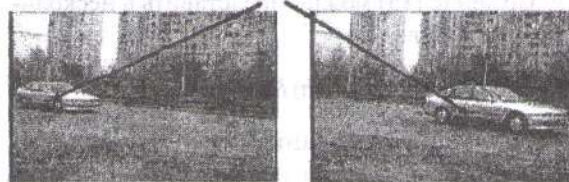


Рис. 9. Пример проективных преобразований

Рассмотрим нормализацию группы центропроекттивных преобразований \bar{G} , т. е. преобразований, оставляющих неподвижной точку, расположенную в начале координат. Очевидно, что центропроекттивное преобразование относится к виду P_2 .

Совокупность центропроекттивных преобразований образует группу относительно операции умножения. Матрица \bar{P}_2 этой группы представима в виде:

$$\bar{P}_2 = D(m, n)P_y(t)P_x(k)H_x(\beta)U(\alpha), \quad (15)$$

где $D_m(m, n)$, $P_y(t)$, $P_x(k)$, $H_x(\beta)$, $U(\alpha)$ — матрицы преобразований масштаба, перспективы вдоль осей x , y косоугольного сдвига вдоль оси абсцисс и поворота (со своими параметрами).

Рассмотрим геометрическую интерпретацию разложения (15).

Пусть векторы ℓ и η (рис. 10) совпадают по направлению с координатными осями и при центропроективном преобразовании переходят, соответственно, в ℓ' и η' . Последовательная нормализация векторов ℓ' и η' представляет собой геометрическую интерпретацию разложения (15), где α — угол между ℓ и ℓ' ; β — угол между η и η' ; \vec{n} — вектор нормали к прямой ℓ' ; m, t — соответственно, параметры сжатия и перспективы вдоль вектора ℓ' ; n, k — параметры сжатия и перспективы вдоль вектора η' ; т. е. $\ell' = \frac{m\ell}{t\ell+1}$ и $\eta' = \frac{n\eta}{k\eta+1}$.

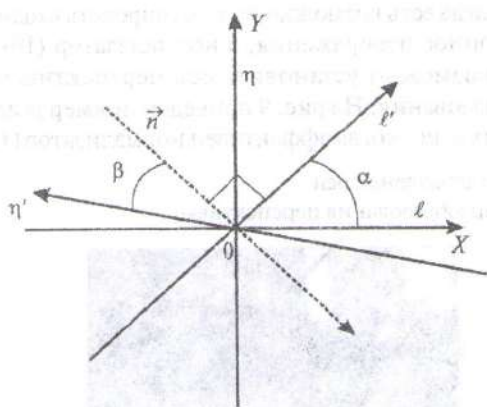


Рис. 10. Геометрическая интерпретация разложения

$$\ell_0 = D(m, n)P_y(t)P_x(k)H_x(\beta)U(\alpha)$$

Разложение (15) можно представить в несколько упрощенном виде:

$$P_0 = D(m, n)P_\alpha(h)H_x(\beta)U(\alpha), \quad (16)$$

где $P_\alpha(h)$ — преобразование перспективы вдоль прямой с углом наклона α и параметром h .

На основе разложений (15), (16) нормализатор (9) проективной группы представляется в следующем виде:

$$F = F_0 F_C',$$

где

$$F_0 = F_D F_{P_\alpha} F_{H_x} F_U$$

или

$$F_0 = F_D F_{P_y} F_{P_x} F_{H_x} F_U.$$

Для нормализатора (9) важным моментом является центрирование изображений. Легко показать, что функционалы обобщенного центра тяжести, пригодные для аффинных преобразований, не под-

ходят для проективных. Конечно, эта задача может быть легко решена, если заранее заданы соответствующие метки (вешки) на эталонном и входном изображениях, что часто делается при аэрофотосъемке. Возникает вопрос, существует ли общий метод нахождения соответствующих точек для произвольных полутоновых изображений. Эта задача, в принципе, решена методом последовательных одномерных корреляций относительно двух пар точек, заданных произвольно на эталоне и входном изображении. При этом затраты машинного времени весьма велики, поскольку метод вначале предполагает поиск соответствующих линий на эталоне и входном изображении. Метод универсален и не предполагает наличия хотя бы одной пары заранее заданных соответствующих точек.

На рис. 11 приведены примеры нормализации проективных изображений двух типов транспортных средств.

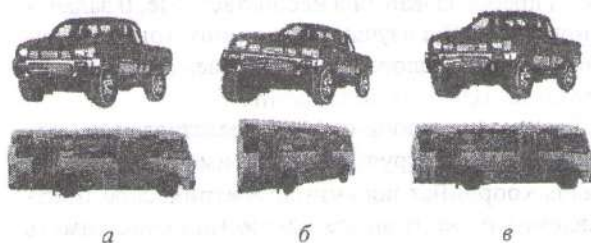


Рис. 11. Примеры нормализации проективных изображений: а — эталонные изображения; б — входные изображения, проективно искаженные; в — нормализованные изображения

5. Нелинейные преобразования

Преобразования классов III и IV можно разделить дополнительно на параметрические и непараметрические. Для параметрических преобразований эталон естественно определяется заданием значений параметров (единица или нуль в зависимости от того, мультипликативную или аддитивную функцию они выполняют).

Примеры таких преобразований приведены в табл. 1 под номерами 25, 26, 27. Это очень широкий класс преобразований, включающий некоторые группы Ли, за исключением аффинных и проективных, а также нелинейные преобразования типа квадратичных форм, которые вообще не образуют групп. Нормализация целого ряда групп Ли представляется сложной и пока нерешенной задачей, хотя попытки использовать их для распознавания изображений предпринимались [15].

В работе [9] рассматривалась возможность нормализации квадратичных преобразований, связывающих эталонное и входное изображения, начиная с одномерных вида $B(x) = B(a_1 x^2 + a_2)$ и заканчивая достаточно сложными двумерными:

$$B(x) = B(a_1x^2 + a_2y^2 + a_3, a_4x^2 + a_5y^2 + a_6).$$

В основу исследований положен подход, основанный на подборе моментов для эталонных и входных изображений типа:

$$\iint_D B(x, y) K_i(x, y) dx dy, \quad i = 1, 2, \dots, n.$$

Например, если связь между эталонным и преобразованным изображением задается в виде $B(x) = B(a_1x^2 + a_2y^2, a_3x^2 + a_4y^2)$, то нормализатор выглядит следующим образом:

$$\begin{aligned} F &= F_2 F_1 [B(x, y)] = F_2 F_1 \left[B_0 \left(A \begin{pmatrix} x^2 \\ y^2 \end{pmatrix} \right) \right] = \\ &= F_2 F_1 \left[B_0 \left(A^{-1} A \begin{pmatrix} x^2 \\ y^2 \end{pmatrix} \right) \right] = \\ &= F_2 [B(x^2, y^2)] \Rightarrow B_0(x, y), \quad A = \begin{vmatrix} a_1 & a_2 \\ a_3 & a_4 \end{vmatrix} \neq 0. \end{aligned}$$

Для нахождения неизвестных $a_i, i = \overline{1, 4}$ возможно применение пяти моментов с функциями $K_i(x, y) = (xu, xy^3, x^3y, xy^5, x^5y)$. При этом моменты для эталонных изображений имеют следующие функции: $K_i^0(x, y) = (1, y, x, xy, y^2, x^2)$. Нелинейная система пяти уравнений, связывающая моменты входных изображений и эталона, решается однозначно при условии $x, y > 0$.

Можно строго и конструктивно определить задачу нормализации в форме (3) для нелинейных экспоненциальных шкал преобразований, поскольку они образуют группу G_3 .

Несмотря на определенные перечисленные успехи в нахождении параметров нормализации, многие вопросы остаются нерешенными. Это прежде всего стабильность эталона, неоднозначность результатов нормализации. Особенно это наглядно проявляется в задачах автоматического масштабирования шкал на осях координат.

Например, нелинейные преобразования 25 из табл. 1 образуют, как нетрудно показать, группу. Тем не менее вопросы автоматической нормализации подобных преобразований (включая, например, и логарифмические шкалы с параметрами) еще не решены.

Возможности нормализации нелинейных преобразований при отсутствии параметров выглядят вообще неясными, особенно с точки зрения дистанционного зондирования объектов через их изображения. Речь идет о преобразованиях 29, 30 из табл. 1. Правда, зеркальные отражения (как дискретная группа с четырьмя преобразованиями G_4) состав-

ляют исключение и могут однозначно автоматически приводиться к эталону нормализатором (3) с функционалами, описанными в [2].

Обратим внимание на тот факт, что нелинейные преобразования 26, 27 после нормализации их параметров приводят входные изображения не к эталону $B_0(x, y)$, а к нелинейной зависимости типа $B(x, y) = B_0(x^2, y^2)$. Разумеется, если заранее известен вид функциональных преобразований координат, то нормализация возможна применением необходимых обратных преобразователей. Это осуществимо не только для преобразований вида $x^2 \rightarrow x, y^2 \rightarrow y$, но, в принципе, и для преобразований G_6 . Вопрос о том, как отыскать функционалы $\Phi(B)$ из (3), которые бы позволили идентифицировать каждое из 16 типов преобразований 29 табл. 1, например, по образцу зеркальных отражений, пока не ясен.

Разумеется, остается открытым вопрос о распознавании типа преобразований, которым подвергается входное изображение. Если соотношение (11) позволяет отличить аффинное преобразование от неаффинного, то отличить, скажем, проективное преобразование от преобразований типа квадратичных форм пока невозможно.

Очевидно, что задачи улучшения и реставрирования изображений уместно рассматривать с позиций общей теории нормализации в совокупности с преобразованиями носителя. Успешные попытки в этом направлении уже сделаны для преобразований 31 [2]. Возможности изучения комбинированного случая (преобразования 33) явно не исчерпаны.

Заключение

Более чем сорокалетний период развития теории и практики нормализации изображений, включая, например, прогресс в области создания систем слежения за подвижными объектами, подтверждает высокую интеллектуальную сложность проблемы, сравнимую с задачами распознавания образов. Несмотря на успехи первых исследований в этой области [10–15], в дальнейшем серьезные результаты стали появляться реже. Связано это как с математическими трудностями решения задач «обратной» геометрии, так и с проблемой восприятия сцен реального мира как человеком, так и компьютерным зрением. На первый взгляд, может представиться нецелесообразной постановка задачи о нормализации ряда преобразований, например, нелинейных или ортогональных (номера 25–30 табл. 1), поскольку человеческое зрение уже не может правильно воспринимать изображения с такими сильными искажениями. Однако нам представляется, что интеллектуальное компьютерное зрение не должно ограничиваться возможностями зрения биологических объектов, включая человека.

Если компьютер считает быстрее человека, а «космическое зрение» спутников уже работает в пяти диапазонах электромагнитных волн (вместо трех человеческих), то, естественно, возникает вопрос о возможном превосходстве в будущем бионических систем нормализации и распознавания образов. Успешное решение поставленных в настоящей работе задач и практическая реализация полученных результатов будут способствовать более широкому пониманию проблемы зрительного восприятия внешнего мира.

Список литературы: 1. Шабанов—Кушнаренко Ю.П., Рвачев В.Л., Мурашко А.Г. Математичні моделі зору. — Київ: Наукова Думка, 1967. — 250 с. 2. Путьятин Е.П., Аверин С.И. Обработка изображений в робототехнике. — М.: Машиностроение, 1990. — 320 с. 3. Путьятин Е.П., Шульгин И.В., Юрченко В.П. Построение инвариантов смещения и поворота зрительных картин // Биологическая и медицинская кибернетика и бионика. — Киев: ИК АН УССР, 1970. — Вып. 3. — С. 57–67. 4. Путьятин Е.П. Теоретические предпосылки нормализации изображений // Проблемы бионики. — Харьков: Вища школа, 1973. — Вып. 10. — С. 82–89. 5. Липанов В.А., Путьятин Е.П. Построение алгоритмов нормализации следящей рамки // Ракетно-космична техніка. — Харків: ХВУ, 1999. — С. 189–193.

6. Путьятин Е.П., Яковлева Е.В. Исследование свойств центральных прямых для нормализации аффинных преобразований // АСУ и приборы автоматики. — Харьков: ХНУРЭ, 2002. — Вып. 121. — С.35–42. 7. Маиталир С.В., Мегель Ю.Е., Путьятин Е.П. Анализ изображений биологических объектов по бинарным сечениям // Радиоэлектроника и информатика. — 2004. — №2. — С. 116–121. 8. Любченко В.А., Путьятин Е.П. Математические модели зрительного восприятия и нормализации изображений // Проблемы бионики. — 2002. — № 56. — С. 44–48. 9. Путьятин Е.П., Долженкова Т.Г. Вопросы нормализации нелинейных преобразований. Сообщение 1 и 2. — Сб. Проблемы бионики. — Харьков, 1980. — Вып. 24. — С. 111–120. 10. Ху М.К. Pattern recognition by moment invariants. Proc. IRE, v. 49, 1428, 1961. 11. Васильев В.И., Реуцкий В.Ю. Инвариантные описания плоских зрительных образов. «Автоматика», 1966. — № 4. 12. Amari S. Theory of normalization of pattern signals in feature spaces. Electronic and Communications in Japan, vol. 49, № 7, 1966. 13. Махонин В.А. Об аффинном опознании плоских фигур. // Изв. АН СССР, сер. Техническая кибернетика, 1963, — №1; 14. Шмидт А.А., Якубович В.А. Алгоритмы математической фильтрации и их применение в задачах распознавания сложных изображений // Автоматика и телемеханика. — 1978. — № 2. — С. 161–176; 15. Файн В.С. Опознавание изображений. — М.: Наука, 1970. — 296 с.

Поступила в редколлегию 12.01.07

МЕТРИЧЕСКИЕ СВОЙСТВА ГРАНУЛЯЦИИ ИНФОРМАЦИИ

А. Г. Каграманян¹, В. П. Машталир², В. В. Шляхов³

¹ ХНУ, г. Харьков, Украина

² ХНУРЭ, г. Харьков, Украина

³ ХНУРЭ, г. Харьков, Украина

Для задач грануляции — представления информации с различной степенью ее детализации или обобщения — исследуются метрические свойства конечных разбиений произвольных измеримых множеств. Вводится и обосновывается метрика, для которой указываются равносильные формы представления, предлагаются ее вероятностные интерпретации. Обсуждается гипотеза об общем виде функционала, являющегося метрикой на разбиениях измеримых множеств.

ГРАНУЛЯЦИЯ, МНОЖЕСТВО, РАЗБИЕНИЕ, МЕТРИКА.

Введение

Granular Computing — грануляция информации, включающая в себя огрубление–детализацию данных, их стратифицированную структуризацию и причинно–следственные отношения в процессе регистрации, трансформации и интерпретации [1], оперирует, как правило, с разбиениями исходного (или индуцированного признакового) пространства [2]. Применение классов эквивалентностей, выступающих в качестве, пожалуй, основного объекта изучения и манипуляций, приводит к необходимости решения целого класса задач. В первую очередь следует указать поиск и анализ внутренних эквивалентностей [3, 4], установление правомочности (согласованности) использования совокупности эквивалентностей [5, 6], в том числе внешних, введение которых диктуется предметной областью, наконец, разумное определение контекстно–зависимых операций на классах эквивалентностей [7, 8]. Достаточно общим вопросом для задач указанного типа является анализ сходства (различия) между ансамблями группировок данных. Однако метрические свойства, вообще говоря, хорошо изучены лишь для анализа в форматах «точка–точка», «точка–множество», «множество–множество».

Наибольшего развития сравнение разбиений достигло при исследовании эффективности кластеризации [9, 10] и сегментации изображений [11–13] в основном в рамках «ground truth» подхода. Иначе говоря, когда результаты автоматической грануляции сравниваются с результатами, полученными человеком. Как правило, самые распространенные коэффициенты сходства (различия) не удовлетворяют неравенству треугольника, что ограничивает диапазон их использования. Вместе с тем известен и ряд метрик на разбиениях. Прежде всего — это метрика, имеющая структурное сходство с метрикой Хаусдорфа, в основу которой положены суммы максимальных мощностей пересечений каждого класса эквивалентностей со всеми элементами разбиений второго множества [14]. Достаточно широ-

кое распространение получила метрика энтропийного типа, названная вариацией информации, которая базируется на понятиях теории информации [15]. При использовании метрики MiCRoM (Minimum Cost Region Matching) поиск расстояния между разбиениями сводится к решению задачи математического программирования, а именно — транспортной задачи [16].

На интуитивном уровне понятно, что метрика на разбиениях должна интегрально учитывать характеристики и сходства, и различия классов эквивалентностей. В этом плане можно выделить простейший инструмент — операции пересечения и симметрической разности, соответственно. Конечномерный случай построения метрики такого типа обоснован в [17], некоторые метрические свойства для разбиений произвольных множеств проанализированы в [18]. Цель работы — обоснование, анализ и интерпретация инструментальных средств «множественного» сравнения, т. е. сравнения фактор–множеств.

1. Постановка задачи

Пусть некоторое измеримое множество Ω с мерой $\mu(\cdot)$ характеризует информацию, подлежащую в процессе анализа грануляции. Под мерой понимается количество точек (в конечном случае), длина, площадь, объем, распределение масс, вероятностей и т. п. Будем полагать, что мера конечна, т. е. $\mu(\Omega) < \infty$ и F_Ω — множество всех измеримых подмножеств множества Ω . Рассмотрим Π_Ω — множество конечных (по числу элементов) разбиений множества Ω , т. е. $\alpha \in \Pi_\Omega \Leftrightarrow \alpha = \{A_i\}_{i=1}^n$:

$$\begin{cases} A_i \in F_\Omega \quad \forall i \in \{1, 2, \dots, n\}; \\ A_i \cap A_{i'} = \emptyset, \quad \forall i \neq i' \in \{1, 2, \dots, n\}; \\ A_1 \cup \dots \cup A_n = \Omega. \end{cases} \quad (1)$$

Пусть даны произвольные разбиения $\alpha, \beta, \gamma, \dots \in \Pi_\Omega$ и $\alpha = \{A_i\}_{i=1}^n, \beta = \{B_j\}_{j=1}^m, \gamma = \{C_k\}_{k=1}^l, \dots$. Задача заключается в установлении, изучении и интерпретации соотношений между $\alpha, \beta, \gamma, \dots$

2. Метрические свойства разбиений

Введем на $\Pi_\Omega \times \Pi_\Omega$ функционал

$$\rho(\alpha, \beta) = \sum_{i=1}^n \sum_{j=1}^m \mu(A_i \Delta B_j) \mu(A_i \cap B_j), \quad (2)$$

где « Δ » — симметрическая разность множеств, и покажем, что он представляет собой метрику на произвольных разбиениях измеримых множеств. Предварительно укажем несколько полезных свойств разбиений и найдем эквивалентную форму функционала (2).

Свойство 1. Для любого множества $D \subseteq \Omega$ и произвольного разбиения $\alpha \in \Pi_\Omega$ справедливо равенство $\mu(D) = \sum_{i=1}^n \mu(D \cap A_i)$.

Действительно, разбиение $\alpha \in \Pi_\Omega$ дробит множество D на непересекающиеся подмножества $D \cap A_i$, и тогда из свойства аддитивности меры следует требуемое равенство.

Свойство 2. Пересечение $\alpha \cap \beta = \{A_i \cap B_j\}_{i=1, n, j=1, m}$ пары $\alpha, \beta \in \Pi_\Omega$ произвольных разбиений также является разбиением.

Действительно, если выбрать произвольный элемент $\omega \in \Omega$, то, поскольку α и β являются разбиениями, найдутся номера $i \in \{1, \dots, n\}$ и $j \in \{1, \dots, m\}$, для которых $\omega \in A_i$ и $\omega \in B_j$, т. е. $\omega \in A_i \cap B_j$. Таким образом,

$$\bigcup_{i=1}^n \bigcup_{j=1}^m \{A_i \cap B_j\} = \Omega. \quad (3)$$

С другой стороны, в силу ассоциативности и коммутативности операции пересечения множеств можно записать:

$$(A_i \cap B_j) \cap (A_{i'} \cap B_{j'}) = (A_i \cap A_{i'}) \cap (B_j \cap B_{j'}).$$

Но поскольку пары (i, j) и (i', j') не равны, то либо $i \neq i'$, либо $j \neq j'$, либо эти неравенства выполняются одновременно, что означает: одно из множеств $A_i \cap A_{i'}$ или $B_j \cap B_{j'}$ равно \emptyset , т. к. они принадлежат разбиениям α и β , соответственно. Таким образом, в итоге имеем

$$(A_i \cap B_j) \cap (A_{i'} \cap B_{j'}) = \emptyset$$

для любых пар $(i, j), (i', j') \in \{1, \dots, n\} \times \{1, \dots, m\}$. В совокупности с равенством (3) получаем выполнение соотношений (1), т. е. справедливость свойства 2.

Рассмотрим пересечения трех произвольных разбиений $\alpha, \beta, \gamma \in \Pi_\Omega$ и обозначим $\theta_{ijk} = \mu(A_i \cap B_j \cap C_k)$, $i = \overline{1, n}$; $j = \overline{1, m}$; $k = \overline{1, l}$.

Свойство 3. Для пересечений трех произвольных разбиений $\alpha, \beta, \gamma \in \Pi_\Omega$ верны соотношения:

$$\mu(C_k) = \sum_{i=1}^n \sum_{j=1}^m \theta_{ijk}, \quad (4)$$

$$\mu(A_i \cap B_j) = \sum_{k=1}^l \theta_{ijk}, \quad (5)$$

$$\mu(A_i \cap C_k) = \sum_{j=1}^m \theta_{ijk}, \quad (6)$$

$$\mu(B_j \cap C_k) = \sum_{i=1}^n \theta_{ijk}. \quad (7)$$

Действительно, если в соответствии со свойством (2) рассмотреть разбиение $\xi = \alpha \cap \beta$ и применить свойство 1, полагая, что D — это множество C_k , то

$$\mu(C_k) = \sum_{i=1}^n \sum_{j=1}^m \mu(C_k \cap (A_i \cap B_j)) = \sum_{i=1}^n \sum_{j=1}^m \theta_{ijk},$$

т. е. получаем равенство (4). Рассуждая совершенно аналогично и используя в качестве D поочередно множества $A_i \cap B_j, A_i \cap C_k, B_j \cap C_k$, а в качестве соответствующего разбиения — γ, β и α , получаем справедливость равенств (5) — (7).

Лемма. Функционал $\rho(\alpha, \beta)$ имеет эквивалентную форму

$$\rho(\alpha, \beta) = \sum_{i=1}^n [\mu(A_i)]^2 + \sum_{j=1}^m [\mu(B_j)]^2 - 2 \sum_{i=1}^n \sum_{j=1}^m [\mu(A_i \cap B_j)]^2.$$

Доказательство. Для любых измеримых множеств E и F имеет место

$$\mu(E \Delta F) = \mu(E) + \mu(F) - 2\mu(E \cap F). \quad (8)$$

Действительно, учитывая равенства

$$E = (E \setminus F) \cup (E \cap F) \text{ и } F = (F \setminus E) \cup (E \cap F),$$

причем $E \setminus F$ и $E \cap F$ не пересекаются, из свойства аддитивности меры получаем:

$$\begin{aligned} \mu(E) &= \mu(E \setminus F) + \mu(E \cap F), \\ \mu(F) &= \mu(F \setminus E) + \mu(E \cap F), \end{aligned}$$

складывая которые с учетом $E \Delta F = (E \setminus F) \cup (F \setminus E)$, имеем:

$$\begin{aligned} \mu(E) + \mu(F) &= \mu(E \setminus F) + \mu(F \setminus E) + 2\mu(E \cap F) = \\ &= \mu(E \Delta F) + 2\mu(E \cap F), \end{aligned}$$

что равносильно (8).

Применяя (8) к A_i и B_j , представим (2) в виде:

$$\begin{aligned} \rho(\alpha, \beta) &= \sum_{i=1}^n \sum_{j=1}^m \mu(A_i \Delta B_j) \mu(A_i \cap B_j) = \\ &= \sum_{i=1}^n \sum_{j=1}^m \mu(A_i \cap B_j) [\mu(A_i) + \mu(B_j) - 2\mu(A_i \cap B_j)] = \\ &= \sum_{i=1}^n \sum_{j=1}^m \mu(A_i) \mu(A_i \cap B_j) + \\ &+ \sum_{i=1}^n \sum_{j=1}^m \mu(B_j) \mu(A_i \cap B_j) - 2 \sum_{i=1}^n \sum_{j=1}^m [\mu(A_i \cap B_j)]^2 = \\ &= \sum_{i=1}^n \mu(A_i) \sum_{j=1}^m \mu(A_i \cap B_j) + \\ &+ \sum_{j=1}^m \mu(B_j) \sum_{i=1}^n \mu(B_j \cap A_i) - 2 \sum_{i=1}^n \sum_{j=1}^m [\mu(A_i \cap B_j)]^2. \end{aligned}$$

В цепочке этих равенств выделим суммы вида

$$\sum_{j=1}^m \mu(A_i \cap B_j), \sum_{i=1}^n \mu(B_j \cap A_i)$$

и, применяя к ним свойство 1, подразумевая под D A_i и B_j , соответственно, получим:

$$\sum_{j=1}^m \mu(A_i \cap B_j) = \mu(A_i), \sum_{i=1}^n \mu(B_j \cap A_i) = \mu(B_j)$$

для любых i и j , где $i = \overline{1, n}$, $j = \overline{1, m}$. Таким образом,

$$\rho(\alpha, \beta) = \sum_{i=1}^n [\mu(A_i)]^2 + \sum_{j=1}^m [\mu(B_j)]^2 - 2 \sum_{i=1}^n \sum_{j=1}^m [\mu(A_i \cap B_j)]^2,$$

что и требовалось доказать.

Теорема 1. Функционал (2), определенный на произвольных измеримых разбиениях α, β измеримого множества Ω , является метрикой.

Доказательство сводится к проверке выполнения аксиом рефлексивности, симметричности и неравенства треугольника.

Симметричность непосредственно следует из вида функционала (2). Докажем рефлексивность, а именно: $\rho(\alpha, \beta) = 0 \Leftrightarrow \alpha = \beta$. Рассмотрим необходимость. Для этого представим функционал (2) в виде:

$$\rho(\alpha, \alpha) = \sum_{i=1}^n \sum_{j=1}^n \mu(A_i \Delta A_j) \mu(A_i \cap A_j) = \sum_{i=1}^n \mu(A_i \Delta A_i) \mu(A_i \cap A_i) + 2 \sum_{\substack{i, j=1 \\ i < j}}^n \mu(A_i \Delta A_j) \mu(A_i \cap A_j),$$

т. е. $\rho(\alpha, \alpha)$ состоит из n слагаемых с совпадающими номерами, а значит, и с совпадающими элементами разбиения, и из $n^2 - n$ слагаемых с различными элементами разбиения. Первая группа слагаемых состоит из нулей, т. к. $\mu(A_i \Delta A_i) = 0$, для $i = \overline{1, n}$, а вторая группа дает нулевые слагаемые, т. к. $\mu(A_i \cap A_j) = 0$ для $\forall i, j = \overline{1, n}$ и $i \neq j$. Таким образом, в одну сторону рефлексивность доказана.

Покажем достаточность, т. е. пусть для двух, в общем случае, различных разбиений $\alpha, \beta \in \Pi_\Omega$ имеет место $\rho(\alpha, \beta) = 0$. Поскольку все слагаемые функционала (2) положительны, то он равен 0 только тогда, когда каждое из них равно 0. Выберем некоторый элемент разбиения $A' \in \alpha$. В исходном функционале он входит в набор «нулевых» слагаемых $\mu(A' \Delta B_j) \mu(A' \cap B_j) = 0$, где $j = \overline{1, m}$. Если предположить, что A' не является элементом разбиения β , то для всех $B_j \in \beta$ выполняется $\mu(A' \Delta B_j) \neq 0$. Но тогда для всех номеров j должно выполняться

$\mu(A' \cap B_j) = 0$. Однако это возможно только тогда, когда $A' \neq \emptyset$, т. к. $A' \in \Omega$, а весь набор $\beta = \{B_j\}_{j=1}^m$ покрывает множество Ω . В то же время A' — элемент покрытия β этого же множества Ω и, естественно, $A' \neq \emptyset$. Тогда найдутся элементы разбиения β , покрывающие A' и имеющие с ним непустые пересечения. Но тогда для этих элементов выполняется неравенство $\mu(A' \cap B_j) \neq 0$, т. е. получаем противоречие. В итоге можно утверждать, что произвольный элемент $A' \in \alpha$ является элементом β , или $\alpha \subset \beta$. В силу симметрии абсолютно аналогично можно показать, что $\beta \subset \alpha$, т. е. $\alpha = \beta$. Таким образом, рефлексивность доказана.

Приступим к доказательству неравенства треугольника. Допустим, имеется три произвольных разбиения $\alpha, \beta, \gamma \in \Pi_\Omega$ множества Ω . Необходимо доказать, что

$$\rho(\alpha, \gamma) + \rho(\gamma, \beta) \geq \rho(\alpha, \beta). \quad (9)$$

Используя лемму 1, получим:

$$\begin{aligned} \rho(\alpha, \gamma) + \rho(\gamma, \beta) &= \sum_{i=1}^n [\mu(A_i)]^2 + \sum_{j=1}^m [\mu(B_j)]^2 + \\ &+ 2 \sum_{k=1}^l [\mu(C_k)]^2 - 2 \sum_{i=1}^n \sum_{k=1}^l [\mu(A_i \cap C_k)]^2 - \\ &- 2 \sum_{k=1}^l \sum_{j=1}^m [\mu(C_k \cap B_j)]^2. \end{aligned}$$

Отсюда:

$$\begin{aligned} \rho(\alpha, \gamma) + \rho(\gamma, \beta) - \rho(\alpha, \beta) &= \\ &= \sum_{k=1}^l [\mu(C_k)]^2 - \sum_{i=1}^n \sum_{k=1}^l [\mu(A_i \cap C_k)]^2 - \\ &- \sum_{k=1}^l \sum_{j=1}^m [\mu(B_j \cap C_k)]^2 + \sum_{i=1}^n \sum_{j=1}^m [\mu(A_i \cap B_j)]^2. \end{aligned}$$

Тем самым, неравенство (9), используя свойство 3, можно представить в виде:

$$\begin{aligned} \sum_{k=1}^l \left(\sum_{i=1}^n \sum_{j=1}^m \theta_{ijk} \right)^2 + \sum_{i=1}^n \sum_{j=1}^m \left(\sum_{k=1}^l \theta_{ijk} \right)^2 \geq \\ \geq \sum_{i=1}^n \sum_{k=1}^l \left(\sum_{j=1}^m \theta_{ijk} \right)^2 + \sum_{j=1}^m \sum_{k=1}^l \left(\sum_{i=1}^n \theta_{ijk} \right)^2. \end{aligned} \quad (10)$$

Зафиксируем $l = 1$, тогда, обозначая $g_{ij} = \theta_{ij1}$, для любых $g_{ij} \geq 0$, $i = \overline{1, n}$, $j = \overline{1, m}$, имеем:

$$\left(\sum_{i=1}^n \sum_{j=1}^m g_{ij} \right)^2 + \sum_{i=1}^n \sum_{j=1}^m (g_{ij})^2 \geq \sum_{i=1}^n \left(\sum_{j=1}^m g_{ij} \right)^2 + \sum_{j=1}^m \left(\sum_{i=1}^n g_{ij} \right)^2. \quad (11)$$

Заметим, если интерпретировать набор чисел g_{ij} как $(n \times m)$ матрицу G , то в правой части этого неравенства в скобках, соответственно, стоят сумма элементов строк и сумма элементов столбцов. Структура левой части неравенства имеет вид:

$$2 \sum_{i,j} g_{ij}^2 + \sum_{(i,j) \neq (i',j')} g_{ij} g_{i'j'}, \quad (12)$$

а правой:

$$2 \sum_{i,j} g_{ij}^2 + \sum_i \sum_{j \neq j'} g_{ij} g_{ij'} + \sum_j \sum_{i \neq i'} g_{ij} g_{i'j}. \quad (13)$$

Сравнивая (12) и (13), отметим: у них есть общая часть (удвоенная сумма квадратов), в (12) входят все попарные произведения элементов матрицы G с неравными индексами, а в (13) — все попарные произведения с неравными индексами, но из одной строки или из одного столбца (т. е. отсутствующих попарные произведения, например, диагональных элементов). Таким образом, в (13) число слагаемых меньше, чем в выражении (12), и при этом все слагаемые выражения (13) содержатся в (12), т. е. неравенство (11) доказано (выполняется (10) при $l = 1$). Рассмотрим теперь числа θ_{ijk} как набор из l матриц:

$$G_k = (g_{ij}^k)_{j=1, \dots, n}, \quad g_{ij}^k = \theta_{ijk}, \quad k = \overline{1, l}.$$

Тем самым, неравенство (10) трансформируется к виду:

$$\sum_k \left(\sum_{i,j} g_{ij}^k \right)^2 + \sum_{i,j} \left(\sum_k g_{ij}^k \right)^2 \geq \sum_{i,k} \left(\sum_j g_{ij}^k \right)^2 + \sum_{j,k} \left(\sum_i g_{ij}^k \right)^2$$

или

$$\begin{aligned} & \sum_k \left[\left(\sum_{i,j} g_{ij}^k \right)^2 + \left(\sum_{i,j} g_{ij}^k \right)^2 \right] + S \geq \\ & \geq \sum_k \left[\left(\sum_i \left(\sum_j g_{ij}^k \right)^2 \right) + \left(\sum_j \left(\sum_i g_{ij}^k \right)^2 \right) \right], \end{aligned}$$

где $S = \sum_{i,j} \left(\sum_k g_{ij}^k \right)^2 - \sum_{i,j,k} \left(g_{ij}^k \right)^2 \geq 0$,

поскольку состоит только из двойных произведений элементов матриц G_k , а все $g_{ij}^k \geq 0$. С другой стороны, применяя (11) к каждой из матриц G_k , суммируя по k и добавляя в левую часть S , получаем последнее неравенство, а следовательно, справедливость (10), что завершает доказательство неравенства треугольника и теоремы в целом.

Таким образом, показано, что введенный нами функционал (2) на конечных разбиениях на измеримые подмножества измеримого множества Ω является метрикой.

Теорема 2. Для произвольной области $\Omega \subseteq \mathbb{R}^s$ и почти всюду (за исключением множеств меры ноль) положительной функции $f(x)$, для которой справедливо $\int_{\Omega} f(x) dx < \infty$, и пары произвольных разбиений $\alpha, \beta \in \Pi_{\Omega}$, имеет место

$$\begin{aligned} \left(\int_{\Omega} f(x) dx \right)^2 & \geq \sum_{i=1}^n \sum_{j=1}^m \left(\int_{A_i \cap B_j} f(x) dx \right)^2 \\ & + \sum_{i=1}^n \sum_{i'=i-1}^n \int_{A_i} f(x) dx \int_{A_{i'}} f(x) dx + \\ & + \sum_{j=1}^m \sum_{j'=j+1}^m \int_{B_j} f(x) dx \int_{B_{j'}} f(x) dx. \end{aligned} \quad (14)$$

Доказательство. Поскольку функции $f(x)$ индуцируют меру на измеримых подмножествах $A \subseteq \Omega$ в виде интеграла $\mu(A) = \int_A f(x) dx$, тогда для двух конечных разбиений α, β будем иметь:

$$d(\alpha, \beta) = \sum_{i=1}^n \sum_{j=1}^m \int_{A_i \Delta B_j} f(x) dx \int_{A_i \cap B_j} f(x) dx.$$

В соответствии с леммой, принимая во внимание неотрицательность функции $d(\alpha, \beta)$, получаем интегральное неравенство

$$\sum_{i=1}^n \left[\int_{A_i} f(x) dx \right]^2 + \sum_{j=1}^m \left[\int_{B_j} f(x) dx \right]^2 \geq 2 \sum_{i=1}^n \sum_{j=1}^m \left[\int_{A_i \cap B_j} f(x) dx \right]^2,$$

дополняя в котором левую часть до сумм полных квадратов, находим

$$\begin{aligned} & \left[\sum_{i=1}^n \int_{A_i} f(x) dx \right]^2 + \left[\sum_{j=1}^m \int_{B_j} f(x) dx \right]^2 \geq 2 \sum_{i=1}^n \sum_{j=1}^m \int_{A_i \cap B_j} f(x) dx \int_{A_i \cap B_j} f(x) dx + \\ & + 2 \sum_{i=1}^n \sum_{i'=i+1}^n \int_{A_i} f(x) dx \int_{A_{i'}} f(x) dx + 2 \sum_{j=1}^m \sum_{j'=j+1}^m \int_{B_j} f(x) dx \int_{B_{j'}} f(x) dx. \end{aligned}$$

Учитывая, что α, β — конечные (по числу классов эквивалентностей) разбиения множества Ω , имеем:

$$\sum_{i=1}^n \int_{A_i} f(x) dx = \int_{\Omega} f(x) dx, \quad \sum_{j=1}^m \int_{B_j} f(x) dx = \int_{\Omega} f(x) dx,$$

откуда сразу же следует неравенство (14), что и требовалось доказать.

На множестве разбиений Π_{Ω} введем функционал

$$S(\alpha) = \sum_{i=1}^n [\mu(A_i)]^2 \quad (15)$$

и обозначим $\alpha\beta = \alpha \cap \beta$, тогда справедлива

Теорема 3. Для любых трех разбиений $\alpha, \beta, \gamma \in \Pi_{\Omega}$ множества Ω справедливо неравенство

$$S(\gamma) \geq S(\alpha\gamma) + S(\beta\gamma) - S(\alpha\beta). \quad (16)$$

Доказательство. Во-первых, метрика на разбиениях $\rho(\alpha, \beta)$ с учетом леммы 1 может быть выражена через функционал (15):

$$\rho(\alpha, \beta) = S(\alpha) + S(\beta) - 2S(\alpha\beta).$$

Во-вторых, из неравенства треугольника для метрики (2) и любых трех разбиений $\alpha, \beta, \gamma \in \Pi_\Omega$ получаем:

$$S(\alpha) + S(\gamma) - 2S(\alpha\gamma) + S(\beta) + S(\gamma) - 2S(\beta\gamma) \geq S(\alpha) + S(\beta) - 2S(\alpha\beta),$$

откуда следует требуемое неравенство (16). Таким образом, теорема 3 доказана.

Из теорем 2 и 3 вытекает

Следствие 1. Для произвольной области $\Omega \subseteq \mathbb{R}^s$ и функции $f(x)$, удовлетворяющей условиям теоремы 2 и произвольных разбиений $\alpha, \beta, \gamma \in \Pi_\Omega$, имеет место

$$\left(\int_{\Omega} f(x) dx \right)^2 \geq \sum_{i=1}^n \sum_{k=1}^l \left(\int_{A_i \cap C_k} f(x) dx \right)^2 + \sum_{j=1}^m \sum_{k=1}^l \left(\int_{B_j \cap C_k} f(x) dx \right)^2 + 2 \sum_{k=1}^l \sum_{i=1}^n \int_{C_k} f(x) dx \int_{A_i} f(x) dx - \sum_{i=1}^n \sum_{j=1}^m \left(\int_{A_i \cap B_j} f(x) dx \right)^2.$$

Доказательство. Для произвольного конечного разбиения $\alpha \in \Pi_\Omega$ функционал (12) примет вид:

$$S(\alpha) = \sum_{i=1}^n \left[\int_{A_i} f(x) dx \right]^2.$$

Неравенство (16) в этой ситуации станет интегральным и будет иметь форму:

$$\sum_{k=1}^l \left(\int_{C_k} f(x) dx \right)^2 \geq \sum_{i=1}^n \sum_{k=1}^l \left(\int_{A_i \cap C_k} f(x) dx \right)^2 + \sum_{j=1}^m \sum_{k=1}^l \left(\int_{B_j \cap C_k} f(x) dx \right)^2 - \sum_{i=1}^n \sum_{j=1}^m \left(\int_{A_i \cap B_j} f(x) dx \right)^2.$$

Если к левой и правой части этого неравенства прибавить слагаемое

$$2 \sum_{k=1}^l \sum_{k'=k+1}^l \int_{C_k} f(x) dx \int_{C_{k'}} f(x) dx,$$

то его правая часть сразу примет требуемый вид, а левая часть с учетом того, что $\gamma = \{C_k\}_{k=1}^l$ — разбиение области Ω , будет представлена в форме:

$$\sum_{k=1}^l \left(\int_{C_k} f(x) dx \right)^2 + 2 \sum_{k=1}^l \sum_{k'=k+1}^l \int_{C_k} f(x) dx \int_{C_{k'}} f(x) dx = \left(\sum_{k=1}^l \int_{C_k} f(x) dx \right)^2 = \left(\int_{\Omega} f(x) dx \right)^2,$$

что и требовалось доказать.

3. Вероятностная интерпретация метрики

Из свойства 1 следует, что для любого $D \subseteq \Omega$ при $\mu(A_i) \neq 0$

$$\frac{\mu(D)}{\mu(\Omega)} = \sum_{i=1}^n \frac{\mu(D \cap A_i)}{\mu(\Omega)} = \sum_{i=1}^n \frac{\mu(D \cap A_i)}{\mu(A_i)} \frac{\mu(A_i)}{\mu(\Omega)}, \quad (17)$$

тогда, прибегая к терминологии теории вероятностей, можно говорить о наборе $\{A_i\}_{i=1}^n$ как о наборе гипотез, и соотношение (17) представляет собой формулу полной вероятности

$$p(D) = \sum_{i=1}^n p(A_i) p(D/A_i),$$

где $p(D) = \mu(D)/\mu(\Omega)$, $p(A_i) = \mu(A_i)/\mu(\Omega)$ — вероятности события D и гипотез A_i , а

$$p(D/A_i) = \frac{\mu(D \cap A_i)}{\mu(\Omega)} \frac{\mu(\Omega)}{\mu(A_i)} = \frac{\mu(D \cap A_i)/\mu(\Omega)}{\mu(A_i)/\mu(\Omega)} = \frac{p(D \cap A_i)}{p(A_i)}$$

является условной вероятностью события D при условии выполнения гипотезы A_i . Таким образом, можно заключить: свойство 1 может быть получено и как следствие формулы полной вероятности. Но тогда возможна и более тесная связь метрических свойств грануляции информации с иными понятиями теории вероятностей. Остановимся на этом подробнее.

Нетрудно заметить, что если $\Omega \subset \mathbb{R}^s$ — некоторое измеримое множество и имеет место схема геометрической вероятности, тогда вероятность любого события D равна

$$p(D) = \frac{\mu(D)}{\mu(\Omega)},$$

где μ — мера, соответствующая \mathbb{R}^s ($s=1$ — длина, $s=2$ — площадь и т. д.). С другой стороны, если $\rho(\alpha, \beta)$ нормировать, а именно разделить на $\mu^2(\Omega)$, то получим

$$\bar{\rho}(\alpha, \beta) = \frac{\rho(\alpha, \beta)}{\mu^2(\Omega)} = \sum_{i=1}^n \sum_{j=1}^m \frac{\mu(A_i \Delta B_j)}{\mu(\Omega)} \frac{\mu(A_i \cap B_j)}{\mu(\Omega)} = \sum_{i=1}^n \sum_{j=1}^m p(A_i \Delta B_j) p(A_i \cap B_j), \quad (18)$$

где $p(A_i \Delta B_j)$ и $p(A_i \cap B_j)$ — вероятности событий $A_i \Delta B_j$ и $A_i \cap B_j$, соответственно. Для метрики (18), использовался функционал (2). В случае, если мы будем пользоваться равенством из леммы, то

$$\bar{\rho}(\alpha, \beta) = \sum_{i=1}^n [p(A_i)]^2 + \sum_{j=1}^m [p(B_j)]^2 - 2 \sum_{i=1}^n \sum_{j=1}^m [p(A_i \cap B_j)]^2 \quad (19)$$

Если рассматривать множество Ω как пространства элементарных исходов, являющееся элементом вероятностного пространства $\langle \Omega, \mathcal{F}_\Omega, p \rangle$, где Ω — в общем случае произвольное множество,

\mathcal{F}_Ω — δ -алгебра его подмножеств, включающая в себя пустое множество \emptyset и Ω , а p — вероятностная мера, относительно которой все элементы \mathcal{F}_Ω измеримы, т. е. для любого $D \in \mathcal{F}_\Omega$ существует число $0 \leq p(D) \leq 1$, удовлетворяющее аксиоматике теории вероятностей А. Н. Колмогорова, то на исходном вероятностном пространстве можно рассматривать различные полные группы событий или наборы гипотез. Более точно, учитывая, что для произвольного набора гипотез $\{S_i\}_{i=1}^n$ имеет место: $S_i \in \mathcal{F}_\Omega$, $i = \overline{1, n}$, $\bigcup_{i=1}^n S_i = \Omega$, $S_i \cap S_{i'} = \emptyset \quad \forall i, i' : i \neq i' \in \{1, \dots, n\}$, то метрику (2) и ее равносильную форму из леммы для двух наборов гипотез $K_1 = \{S_i\}_{i=1}^n$ и $K_2 = \{Q_j\}_{j=1}^m$ можно представить в виде:

$$\begin{aligned} \bar{\rho}(K_1, K_2) &= \sum_{i=1}^n \sum_{j=1}^m p(S_i \Delta Q_j) p(S_i \cap Q_j) = \\ &= \sum_{i=1}^n [p(S_i)]^2 + \sum_{j=1}^m [p(Q_j)]^2 - 2 \sum_{i=1}^n \sum_{j=1}^m [p(S_i \cap Q_j)]^2. \end{aligned}$$

С другой стороны, можно рассматривать две дискретные случайные величины ξ_1 и ξ_2 , заданные конечным числом значений, т. е.

ξ_1	x_1	...	x_i	...	x_n
p	u_1	...	u_i	...	u_n
ξ_2	y_1	...	y_j	...	y_m
p	v_1	...	v_j	...	v_m

Здесь $u_i = p(\xi_1 = x_i)$, $i = \overline{1, n}$; $v_j = p(\xi_2 = y_j)$, $j = \overline{1, m}$, а $\{x_i\}_{i=1}^n, \{y_j\}_{j=1}^m$ — наборы различных действительных (комплексных) чисел. Используя равенство (19), можно ввести метрику на множестве этих случайных величин:

$$\begin{aligned} \bar{\rho}(\xi_1, \xi_2) &= \sum_{i=1}^n [p(\xi_1 = x_i)]^2 + \sum_{j=1}^m [p(\xi_2 = y_j)]^2 - \\ &- 2 \sum_{i=1}^n \sum_{j=1}^m [p(\xi_1 = x_i, \xi_2 = y_j)]^2. \end{aligned}$$

Если ξ_1 и ξ_2 независимы, то

$$p(\xi_1 = x_i, \xi_2 = y_j) = p(\xi_1 = x_i) p(\xi_2 = y_j), \quad (3.34)$$

тогда, подставляя (3.32) в равенство (3.31), получим:

$$\begin{aligned} \bar{\rho}(\xi_1, \xi_2) &= \sum_{i=1}^n [p(\xi_1 = x_i)]^2 + \sum_{j=1}^m [p(\xi_2 = y_j)]^2 - \\ &- 2 \sum_{i=1}^n [p(\xi_1 = x_i)]^2 \sum_{j=1}^m [p(\xi_2 = y_j)]^2 = \\ &= \sum_{i=1}^n [p(\xi_1 = x_i)]^2 \left\{ 1 - \sum_{j=1}^m [p(\xi_2 = y_j)]^2 \right\} + \\ &+ \sum_{j=1}^m [p(\xi_2 = y_j)]^2 \left\{ 1 - \sum_{i=1}^n [p(\xi_1 = x_i)]^2 \right\}. \end{aligned}$$

Если предположить, что распределения случайных величин ξ_1 и ξ_2 равномерны, т. е. все их значения равновероятны: $p(\xi_1 = x_i) = 1/n$, $p(\xi_2 = y_j) = 1/m$ при любых $i = \overline{1, n}$ и $j = \overline{1, m}$, то

$$\bar{\rho}(\xi_1, \xi_2) = \frac{n}{n^2} \left(1 - \frac{m}{m^2}\right) + \frac{m}{m^2} \left(1 - \frac{n}{n^2}\right) = \frac{1}{n} + \frac{1}{m} - \frac{2}{nm}.$$

Таким образом, фактически доказана

Теорема 4. Расстояние между двумя независимыми конечными равномерными распределениями в смысле равенства (2) равно

$$\bar{\rho}(\xi_1, \xi_2) = \frac{1}{n} + \frac{1}{m} - \frac{2}{nm}. \quad (20)$$

Отметим, что если говорить о независимых случайных величинах, то нельзя говорить об их равенстве. В связи с этим свойство рефлексивности в данном случае отсутствует, что видно и из равенства (20).

Рассмотрим вероятностные интерпретации теорем предыдущего раздела. Из теоремы 2 вытекают три любопытных следствия.

Следствие 2 (для гипотез). Если в некотором вероятностном пространстве имеется два набора гипотез $K_1 = \{S_i\}_{i=1}^n$ и $K_2 = \{Q_j\}_{j=1}^m$, то

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^m [p(S_i \cap Q_j)]^2 + \sum_{i=1}^n \sum_{i'=i+1}^n p(S_i) p(S_{i'}) + \\ + \sum_{j=1}^m \sum_{j'=j+1}^m p(Q_j) p(Q_{j'}) \leq 1. \end{aligned}$$

Доказательство с учетом неотрицательности функционала $\bar{\rho}(K_1, K_2)$ совершенно аналогично доказательству теоремы 2.

Следствие 3. Если в некотором вероятностном пространстве произвольное событие D разбито на два набора попарно непересекающихся подсобытий $\{E_i\}_{i=1}^n$ и $\{\Phi_j\}_{j=1}^m$, образующих относительно D полную группу, т. е. $D = \sum_{i=1}^n E_i = \sum_{j=1}^m \Phi_j$, то

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^m [p(E_i \cap \Phi_j)]^2 + \sum_{i=1}^n \sum_{i'=i+1}^n p(E_i) p(E_{i'}) + \\ + \sum_{j=1}^m \sum_{j'=j+1}^m p(\Phi_j) p(\Phi_{j'}) \leq [p(D)]^2. \end{aligned}$$

Следствие 4 (для дискретных случайных величин).

Пусть на одном вероятностном пространстве заданы две дискретные случайные величины ξ_1 и ξ_2 с конечным числом значений, т. е.

$$u_i = p(\xi_1 = x_i), i = \overline{1, n}; v_j = p(\xi_2 = y_j), j = \overline{1, m},$$

а $\{x_i\}_{i=1}^n, \{y_j\}_{j=1}^m$ — наборы различных действительных (комплексных) чисел, то

$$\sum_{i=1}^n \sum_{j=1}^m [p(\xi_1 = x_i, \xi_2 = y_j)]^2 + \sum_{i=1}^n \sum_{i'=i+1}^n u_i u_{i'} + \sum_{j=1}^m \sum_{j'=j+1}^m v_j v_{j'} \leq 1.$$

Если эти случайные величины ξ_1 и ξ_2 еще и независимы, т. е.

$$p(\xi_1 = x_i, \xi_2 = y_j) = p(\xi_1 = x_i) p(\xi_2 = y_j) = u_i v_j$$

для любых $i = \overline{1, n}$ и $j = \overline{1, m}$, то

$$\sum_{i=1}^n u_i^2 \sum_{j=1}^m v_j^2 + \sum_{i=1}^n \sum_{l'=i+1}^n u_l u_{l'} + \sum_{j=1}^m \sum_{j'=j+1}^m v_j v_{j'} \leq 1.$$

Следствие 4 для своего обоснования использует рассуждения, полностью аналогичные рассуждениям, доказывающим теорему 3, неотрицательность функционала $\bar{p}(\xi_1, \xi_2)$ и равенства

$$\sum_{i=1}^n u_i = \sum_{j=1}^m v_j = 1.$$

Рассмотрим теперь следствия из теоремы 3.

Следствие 5 (для гипотез). Для любых трех наборов гипотез, $K_1 = \{S_i\}_{i=1}^n$, $K_2 = \{Q_j\}_{j=1}^m$, $K_3 = \{R_k\}_{k=1}^l$, заданных на одном пространстве элементарных исходов Ω , вероятности, связанные с этими гипотезами, удовлетворяют неравенству

$$\sum_{k=1}^l [p(R_k)]^2 \geq \sum_{i=1}^n \sum_{k=1}^l [p(S_i \cap R_k)]^2 + \sum_{j=1}^m \sum_{k=1}^l [p(Q_j \cap R_k)]^2 - \sum_{i=1}^n \sum_{j=1}^m [p(S_i \cap Q_j)]^2.$$

Следствие 6 (для дискретных случайных величин).

Пусть на одном пространстве элементарных исходов заданы три дискретные случайные величины с конечным числом значений, т. е. $u_i = p(\xi_1 = x_i)$, $i = \overline{1, n}$; $v_j = p(\xi_2 = y_j)$, $j = \overline{1, m}$; $t_k = p(\xi_3 = z_k)$, $k = \overline{1, l}$, где $\{x_i\}_{i=1}^n$, $\{y_j\}_{j=1}^m$, $\{z_k\}_{k=1}^l$ — наборы действительных (комплексных) чисел, тогда:

$$\begin{aligned} \sum_{k=1}^l t_k^2 &\geq \sum_{i=1}^n \sum_{k=1}^l [p(\xi_1 = x_i, \xi_3 = z_k)]^2 + \\ &+ \sum_{j=1}^m \sum_{k=1}^l [p(\xi_2 = y_j, \xi_3 = z_k)]^2 - \\ &- \sum_{i=1}^n \sum_{j=1}^m [p(\xi_1 = x_i, \xi_2 = y_j)]^2. \end{aligned}$$

Если эти случайные величины независимы в совокупности, т. е.

$$\begin{cases} p(\xi_1 = x_i, \xi_2 = y_j) = u_i v_j, & i = \overline{1, n}, j = \overline{1, m}; \\ p(\xi_1 = x_i, \xi_3 = z_k) = u_i t_k, & i = \overline{1, n}, k = \overline{1, l}; \\ p(\xi_2 = y_j, \xi_3 = z_k) = v_j t_k, & j = \overline{1, m}, k = \overline{1, l}, \end{cases}$$

то вероятности их распределений будут удовлетворять неравенству

$$\sum_{k=1}^l t_k^2 \geq \sum_{i=1}^n \sum_{k=1}^l u_i^2 t_k^2 + \sum_{j=1}^m \sum_{k=1}^l v_j^2 t_k^2 - \sum_{i=1}^n \sum_{j=1}^m u_i^2 v_j^2.$$

Приведенные вероятностные интерпретации могут найти широкое применение в мульталгебраических системах [6, 8] и методах грануляции информации [1–5].

Результаты и перспективы исследований

Необходимо подчеркнуть, что проведенное изучение метрических свойств разбиений направлено не столько на получение инструментария для оценки эффективности алгоритмов продуцирования классов эквивалентностей (в задачах кластеризации данных, сегментации изображений и т. п.), сколько на создание объективных предпосылок для синтеза методов интеллектуальной интерпретации фактор-множеств. В качестве типичного примера можно указать поиск компромисса между чрезмерной и недостаточной сегментацией изображений. В этом плане (точнее говоря, при стратифицированной грануляции информации) возникает необходимость исследования метрических свойств частично упорядоченных фактор-множеств — вложенных разбиений. Определенную перспективу составляет и поиск на разбиениях так называемых подобных метрик первого типа ($\rho'(a, \beta) = \text{ср}''(f(a), f(\beta))$, $\text{с} = \text{const}$) и второго типа ($\rho'(f(a), f(\beta)) = \rho''(f(\gamma), f(\delta)) \Leftrightarrow \rho''(a, \beta) = \rho''(\gamma, \delta)$, $a, \beta, \gamma, \delta \in \Pi_\Omega$). Поясним подробнее на примере.

Поскольку под энтропией понимается функционал вида $H(q) = -\sum_{i=1}^n q_i \ln q_i$, где $\bar{q} = (q_1, \dots, q_n)$ — вообще говоря, некоторое дискретное распределение: $\sum_{i=1}^n q_i = 1$, $0 \leq q_i \leq 1$, $i = \overline{1, n}$, то, если даны два любых распределения $\bar{u} = (u_1, \dots, u_n)$ и $\bar{v} = (v_1, \dots, v_m)$, или, что равносильно, в нашем случае (см. раздел 4) $u_i = p(A_i)$, $i = \overline{1, n}$, $v_j = p(B_j)$, $j = \overline{1, m}$, тогда метрика энтропийного типа примет вид [15]:

$$\begin{aligned} \bar{h}(\bar{u}, \bar{v}) = h(\alpha, \beta) &= -\sum_{i=1}^n p(A_i) \ln p(A_i) - \sum_{j=1}^m p(B_j) \ln p(B_j) + \\ &+ 2 \sum_{i=1}^n \sum_{j=1}^m p(A_i \cap B_j) \ln p(A_i \cap B_j) = \\ &= H(\alpha) + H(\beta) - 2H(\alpha \cap \beta). \end{aligned}$$

Сравнивая полученное равенство с (19), можно отметить явное сходство между ними. Напрашивается вывод об общей формуле:

$$\rho(\alpha, \beta) = \sum_{i=1}^n f(p(A_i)) + \sum_{j=1}^m f(p(B_j)) - 2 \sum_{i=1}^n \sum_{j=1}^m f(p(A_i \cap B_j)).$$

В нашем случае используется $f(x) = x^2$. В энтропийной метрике — $f(x) = -x \ln x$. Возможно, допустимо использовать другие функции $f(x)$ для получения новых метрик на множестве конечных разбиений измеримого множества Ω . Ответы на вопросы, каковы эти функции и каким свойствам они должны удовлетворять, остаются открытыми.

Список литературы: 1. *Bargiela A., Pedrycz W.* Granular computing: an introduction. Boston, Kluwer Academic Publishers, The Kluwer International Series in Engineering and Computer Science. Vol. 717. 2002. 478 P. 2. *Yao Y.Y.* A partition model of granular computing // Transactions on Rough Sets I / J.F. Peters, et al. (Eds.). — Lecture Notes in Computer Science. — Berlin Heidelberg: Springer-Verlag. — Vol. 3100. — 2004 — P. 232-253. 3. *Yao J.T., Yao Y.Y.* Induction of classification rules by granular computing // Rough Sets and Current Trends in Computing / J.J. Alpigini, et al. (Eds.). — Lecture Notes in Artificial Intelligence. — Berlin Heidelberg: Springer-Verlag. — Vol. 2475. — 2002. — P. 331-338. 4. *Yao Y.Y.* Perspectives of granular computing // Proceedings of IEEE International Conference on Granular Computing. — Vol. 1. — 2005. — P. 85-90. 5. *Doherty P., Lukaszewicz W., Szalas A.* Information granules for intelligent knowledge structures // Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing / G. Wang, et al. (Eds.). — Lecture Notes in Artificial Intelligence. — Berlin Heidelberg: Springer-Verlag. — Vol. 2639. — 2003. — P. 405-412. 6. *Машталир В.П., Шляхов В.В.* Индуцированная согласованность отношений в задачах грануляции информации // Бионика интеллекта. — №1 (64). — 2006. — С. 19-26. 7. *Lin T.Y.* Granular computing (Structures, representations, and applications) // Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing / G. Wang, et al. (Eds.). — Lecture Notes in Artificial Intelligence. — Berlin Heidelberg: Springer-Verlag. — Vol. 2639. — 2003. — P. 16-24. 8. *Машталир В.П., Шляхов В.В.* Свойства мультиталгебраических систем в задачах компаративного распознавания // Кибернетика и системный анализ. — 2003. — №6. — С. 12-32. 9. *Jain A.K., Murty M.N., Flynn P.J.* Data clustering: a review // ACM Computing Surveys. — Vol. 31, No. 3. — 1999. — P. 264-323. 10. *Jiang X., Marti C., Irniger C., Bunke H.* Image segmentation evaluation by techniques of comparing clusterings // Image

Analysis and Processing / F. Roliand, S. Vitulano (Eds.). — Lecture Notes in Computer Science. — Berlin Heidelberg: Springer-Verlag. — Vol. 3617. — 2005. — P. 344-351. 11. *Jiang X.* Performance evaluation of image segmentation algorithms. // Handbook of Pattern Recognition and Computer Vision / C.H. Chen, P.S.P. Wang (Eds.). — Singapore: World Scientific. — 2005. — P. 525-542. 12. *Jiang X., Marti C., Irniger C., Bunke H.* Distance measures for image segmentation evaluation // EURASIP Journal on Applied Signal Processing. — Vol. 2006. — Article ID 35909. — 10 P. 13. *Jiang X., Marti C., Irniger C., Bunke H.* Image segmentation evaluation by techniques of comparing clusterings. // Image Analysis and Processing / F. Roliand, S. Vitulano (Eds.). — Lecture Notes in Computer Science. — Berlin Heidelberg: Springer-Verlag. — Vol. 3617. — 2005. P. 344-351. 14. *Dongen S.* Performance criteria for graph clustering and Markov cluster experiments // Technical Report: INS-R0012. — Amsterdam: CWI (Centre for Mathematics and Computer Science) — 2000. 15. *Meila M.* Comparing clusterings by the variation of information // Learning Theory and Kernel Machines. — Lecture Notes in Computer Science. — Berlin Heidelberg: Springer-Verlag. — Vol. 2777. — 2003. — P. 173-187. 16. *Stehling R.O., Nascimento M.A., Falcao A.X.* MiCRoM: A metric to compare segmented images // VISUAL 2002 / S.-K. Chang, Z. Chen, S.-Y. Lee. (Eds.). — Lecture Notes in Computer Science. — Berlin Heidelberg: Springer-Verlag. — Vol. 2314. — 2002. P. 12-23. 17. *Mashtalir V., Mikhnova E., Shlyakhov V., Yegorova E.* A Novel metric on partitions for image segmentation. // Proceedings of IEEE International Conference on Advanced Video and Signal Based Surveillance. *avss*. — 2006. — P. 18. 18. *Kinoshenko D., Mashtalir V., Shlyakhov V.* A partition metric for clustering features analysis // International Journal «Information Theories and Applications» — Vol. 14, No 3. — 2007. — P. 230-236.

Поступила в редколлегию 18.01.07

УДК 681.32

ОЧЕРК ОСНОВНЫХ ПРИНЦИПОВ КВАНТОВОЙ ЛИНГВИСТИКИ**В. А. Широков**

Украинский языково-информационный фонд Национальной академии наук Украины

Статья посвящена определению возможных границ формального подхода к описанию языка. Выявлено, что концептуальное описание квантовомеханических и мыслеречевых систем базируется на общих подходах. Введено понятие семантического состояния языковой единицы, приведено их описание средством теории нечетких множеств. Приведены также формальные аспекты отношения синонимии.

СУБСТАНЦИЯ ЯЗЫКОВАЯ, ПРИНЦИПЫ КВАНТОВЫЕ, СОСТОЯНИЕ СЕМАНТИЧЕСКОЕ, МНОЖЕСТВО НЕЧЕТКОЕ, ОТНОШЕНИЕ СИНОНИМИИ.

«Кто говорит, тот кроме имен, взятых с предметов видимых, ничем иным не может слушающим изобразить невидимого».

Св. Ефрем Сирий

«Вера же есть осуществление ожидаемого и уверенность в невидимом».

*Св. Ап. Павел, К Евреям, 11, 1.***Введение**

В языкознании время от времени возникают различные течения и движения, направленные на создание формальных схем, призванных, как полагают их адепты, продвинуть внедрение тех или иных математических методов в дело описания языковой субстанции. Такими, в частности, представляются попытки изобразить языковые феномены с помощью случайных процессов, используя для этого аппарат теории вероятности и математической статистики, попытки описать явления языка (в первую очередь, синтаксические) с помощью формальных грамматик, «натянуть» языковую феноменологию на некие логистические исчисления типа логики предикатов и т. п. Одной из последних попыток в этом ряду явилось изобретение нами теории лексикографических систем и лексикографических исчислений [1].

Следует, однако, отметить, что после первых успехов применения названных (а также других) формальных схем у их энтузиастов, как правило, возникает некое разочарование, связанное с тем, что разработанный или освоенный ими формальный аппарат не способен учесть множество существенных деталей даже тех ограниченных областей языковой системы, для которых он с самого начала, как им казалось, был специализирован.

Автор, переживший периоды и энтузиазма, и разочарования в такого рода деятельности, наконец, успокоился, составив определенное отношение к попыткам формализации вещей, имеющих черты (и соответственно, поведение) субстанции, включая и языковую субстанцию. Сразу же оговоримся, что мы не владем формальным и замкну-

тым определением понятия субстанции, полагаясь на интуитивное его понимание — в конце концов, так же поступают и представители других наук (физики, химии, биологии...) по отношению к субстанциям, изучаемым в данных науках.

Несмотря на не очень впечатляющие достижения в области внедрения формальных методов в лингвистику, автор придерживается мнения, что попытки изобретения все новых способов и средств формализации языковой системы оставлять не следует — ввиду хотя бы того, что они помогают яснее очертить возможные границы формального подхода к описанию языка, а в процессе их реализации время от времени возникают полезные идеи, иногда заимствованные из других наук, которые способствуют расширению методологии научного исследования в языкознании.

Примерно в таком логическом ключе следует воспринимать и попытку, предпринятую в данной работе.

Одной из идей, «заимствованных» из других наук, мы считаем попытку применения идеологии квантовой механики к описанию языковой субстанции. Отмеченная экстраполяция находит свое оправдание в неких общих чертах, присущих микромиру, с одной стороны, и миру языка, с другой, которые мы видим в следующем.

1. Квантовые принципы

Как известно, квантовая механика была изобретена с целью описания микрочастиц материи в той области взаимодействия, где уже нельзя пренебречь влиянием прибора (или наблюдателя) на наблюдаемый объект. В языковой же области влияние наблюдателя на наблюдаемый объект (а именно: влияние субъекта мыслеречевого процесса на семантические состояния языковых единиц) является, на наш взгляд, основным признаком и основным принципом самого существования языковой системы, поскольку оно представляет квинтэссенцию процесса понимания речевого акта. В этом аспекте квантовомеханические и мыслеречевые системы

действительно конгениальны. Логично предположить, что и концептуальное их описание должно содержать некие общие черты и базироваться на общих подходах. Вообще взаимодействие субъекта и объекта, одним из проявлений которого есть влияние наблюдателя на наблюдаемое, представляется нам настолько фундаментальным фактом, что оно, несомненно, заслуживает того, чтобы стать одним из оснований практически любой рациональной схемы описания в науке.

Вследствие отмеченной особенности квантовомеханических систем возникла необходимость формулировки иной, нежели классическая, парадигмы описания квантовых объектов и процессов. Пришлось отказаться от понятия непосредственно и детерминистски воспринимаемой (или наблюдаемой) траектории частицы, представляющей исчерпывающее описание системы, заменив ее неким «суррогатом» — так называемым *состоянием квантовой системы*. Понятие состояния оказалось, однако, чрезвычайно плодотворным как в эпистемологическом, так и в техническом отношении, поскольку его применение позволило не только создать эффективный математический аппарат, способный достаточно точно описывать явления микромира, но и сформулировать новую философию естествознания, что в конце концов привело и к созданию новой — квантовой научной картины мира.

В плане технического удалось построить формальную теорию квантовомеханических состояний, согласно которой они образуют довольно простой математический объект — линейное бесконечномерное метрическое (гильбертово) пространство — пространство решений так называемого уравнения Шредингера — основного уравнения квантовой механики. Значения наблюдаемых величин при этом представляются через матричные элементы неких (эрмитовых) операторов по данным состояниям, которые служат формально-математическими репрезентантами наблюдаемых. Для самих состояний (в квантовой механике их еще называют волновыми функциями) была со временем предложена прозрачная вероятностная интерпретация. Таким образом, теория состояний квантовомеханических систем предложила — уже на новом, неклассическом уровне — формальную основу для описания соответствующего класса явлений.

В философском плане квантовая теория представила целую гамму новых для науки идей. Во-первых, она естественным образом примирила две фундаментальные физические теории: классическую механику и электродинамику, дав корректное объяснение устойчивости атомных систем, носившей до этого характер парадокса, и определив, кроме того, пределы применимости отмеченных теорий. Точно так же естественно она объяснила дискретно-непрерывный характер материи и корпускуляр-

но-волновой дуализм, дала своеобразное разрешение проблемы тождества и, наконец, предоставила в лице принципа дополнительности Бора новую гносеологическую парадигму.

Одним из фундаментальных следствий квантовой механики стало осознание (и наблюдение на эксперименте) так называемых смешанных состояний, а именно таких, в которых квантовая система обладает одновременно не одним определенным значением наблюдаемой величины, а, например, сразу двумя. Это, в частности, означает, что квантовая система в один и тот же момент может пребывать в двух различных состояниях; допуская некоторую вольность речи, можно утверждать, что с вероятностью p она находится в состоянии 1 и с вероятностью $(1 - p)$ в состоянии 2. Такое комбинированное состояние представляется суперпозицией (линейной комбинацией) состояний 1 и 2 — их своеобразной (когерентной) «смесью». Такое положение можно интерпретировать в том смысле, что объект одновременно, в один и тот же момент времени представляет собой и *«нечто»*, и *«нечто иное»*. Принцип суперпозиции состояний сыграл роль основного системообразующего принципа квантовой теории. Один из основоположников квантовой теории П. А. М. Дирак отмечал [2]: «Если состояние образовано путем суперпозиции двух других состояний, то оно будет иметь свойства, которые в некотором, несколько неопределенном смысле слова являются промежуточными между свойствами обоих исходных состояний и которые в большей или меньшей степени приближаются к свойствам одного из них, в зависимости от того, с большим или меньшим «весом» это состояние вошло в суперпозицию».

Следует отметить, что упомянутые феномены в эпоху создания квантовой теории выглядели весьма непривычно, они, казалось, противоречили и обыденному опыту с его наглядными представлениями, и так называемому здравому смыслу, и классической аристотелевой логике. В этой связи приведем еще одно рассуждение П. А. М. Дирака [2]: «...главная задача физической науки заключается не в том, чтобы снабжать нас наглядными картинками, а в том, чтобы формулировать законы, управляющие явлениями, и использовать эти законы для открытия новых явлений. Если наглядная картина существует, то тем лучше; однако существует она или нет — это лишь второстепенный вопрос. В случае атомных явлений нельзя ожидать, что существует наглядная картина в обычном смысле этого слова, в котором под «наглядной» понимается модель, действующая в основном по классическим принципам. Можно, однако, расширить смысл слова «картина» так, чтобы включить в него *любой способ рассмотрения основных законов, при котором их согласованность становится очевидной*».

Заметим, что данное утверждение имеет значение для науки вообще, а не только для физики. И хотя объекты языка могут кому-то представляться весьма наглядными (написанные на бумаге, они могут быть нами осмотрены), но наглядность эта весьма обманчива, поскольку наблюдаемая («наглядная») часть языкового состояния далеко не исчерпывает его сущности. Эффекты такого рода весьма обычны для квантовых систем.

Таким образом, цель данной работы — продемонстрировать ту глубокую аналогию, которая, как нам представляется, существует между квантовой механикой и лингвистикой и простирается так далеко, что позволяет использовать не только понятийный аппарат первой, но и даже ряд конкретных идей квантовомеханических математических конструкций для формального описания языковых феноменов. По ходу дела мы предполагаем сформулировать в терминах состояний интегральный подход к описанию языковой системы, в определенном смысле объединяющий и обобщающий традиционные грамматические и лексикографические подходы.

Сознавая, что аналогии имеют один общий недостаток, а именно — все они неверны, мы тем не менее попытаемся изложить наши соображения по поводу данного соответствия с надеждой, что более проницательные исследователи смогут, вероятно, сделать и более глубокие выводы.

2. Языковая субстанция и ее описание

Остановимся на некоторых общеметодологических замечаниях относительно принципов моделирования языковой системы.

Первым вопросом, который возникает при формулировке этих принципов, является вопрос об объектах моделирования, а именно: что представляют собой объекты языка и что, собственно, мы предполагаем моделировать. В качестве исходной посылки мы выбираем положение о том, что собственными объектами языка являются определенные психофизические состояния и процессы, происходящие в мыслеречевом аппарате человека, а устная и письменная ее формы служат элементами инфраструктуры языкового процесса. Описанию и моделированию подлежат как собственно упомянутые психофизические состояния и процессы, так и их инфраструктура. Исходя из данного положения, выясним их роль в процессах моделирования языка.

Очевидно, что мыслеречевой процесс сам по себе является интегрированным, то есть содержащим как языковую, так и ментальную компоненты. В мыслеречевом аппарате он находит выражение в виде динамической системы взаимосвязанных рефлексов, содержание и характер которых исследуется, например, в книге В. М. Бехтерева [3], которая до сих пор не утратила своей актуальности. Согласно теории В. М. Бехтерева, одним из сочетательных

рефлексов, имеющих место в мозге человека, является естественный язык. Учитывая изложенное, отделение языковых процессов от ментальных, принятое многими лингвистами, а также попытки изучения языка «самого по себе» нам представляются неоправданным и методологически некорректным упрощением. Языковую систему следует рассматривать как открытую, что предполагает существенное расширение и феноменологической базы языка, и соответствующего концептуального аппарата. Вспомним в этой связи слова В. Гумбольдта из его письма Ф. Вольфу (1804 г.): «Мне удалось открыть — и эта мысль все более захватывает меня, — что через посредство языка можно окинуть взглядом самые высокие и глубокие сферы и все многообразие мира». При определенном взгляде на вещи язык приобретает черты универсальной субстанции космического масштаба («Вначале было Слово. И Слово было у Бога. И Слово было Бог»).

Устная и письменная формы языка в этом смысле играют роль моделей мыслеречевых процессов и одновременно — коммуникативной для них среды (как отмечалось — инфраструктуры). При такой факторизации можно утверждать, что они представляют языковую периферию.

Сделаем, однако, предостережение относительно возможной недооценки инфраструктурных составляющих языка, которая может возникнуть ввиду того, что «основными», первичными языковыми объектами нами признаются психофизические состояния и процессы мыслеречевого аппарата. Дело в том, что современные данные свидетельствуют, что язык не является врожденным свойством человека. Врожденной является всего лишь *способность* к языку. А процесс «инсталляции» языка у человека с необходимостью предполагает наличие таких инфраструктурных элементов, как так называемые «внешний» язык и «эгоцентрический» язык, которые функционируют уже на ранних стадиях филогенетического развития языковой системы у ребенка и заканчиваются формированием у него «внутреннего» языка, венчающего процесс создания полноценного языкового аппарата [4]. Таким образом, языковая периферия является неотъемлемым элементом языковой системы. Кроме того, она обеспечивает открытость языковой системы (если, разумеется, не принимать во внимание возможность телепатии). Языковая периферия играет также важную роль «классического» объекта, благодаря которому, как отмечал Нильс Бор, только и возможно «квантовое» описание мира.

Следует отметить, что психофизические состояния и процессы (а среди них и мыслеречевые), как правило, не в полной мере достижимы для непосредственного наблюдения, а тем паче — объективной фиксации. Следовательно, устная и письмен-

ная формы языка фактически служат репрезентантами наблюдаемых компонентов состояний языковых объектов и процессов, происходящих в мыслеречевом аппарате. В качестве таковых они и используются в роли основных объектов при концептуальном моделировании языка. Отмеченное обстоятельство еще более подчеркивает значение языковой периферии в изучении системы языка.

Идея наблюдаемости, впервые в истории науки подробно разработанная при обосновании квантовой механики, вообще является весьма продуктивной в методологическом отношении применительно к описанию сложных процессов. Как отмечал А. Эйнштейн в своей дискуссии с В. Гейзенбергом в 1926 г. при обсуждении оснований квантовой механики [5], построить теорию только из одних наблюдаемых величин нельзя (хотя без них никакая научная теория вообще немыслима). В конце концов лишь сама теория должна определять, какие из ее величин являются непосредственно наблюдаемыми, а какие — нет. В частности, в квантовой механике состояния, вообще говоря, не являются непосредственно наблюдаемыми величинами. Последним, как отмечалось выше, соответствуют определенному типу операторы, действующие в пространстве состояний, и собственным значениям которых отвечают значения наблюдаемых.

Таким образом, в мире, а значит, и в познании существует «невидимое», то есть то, что не может быть наблюдаемо непосредственно. Это дает нам основания для признания роли веры в когнитивно-интеллектуальных процессах, которая, по выражению апостола Павла, есть уверенность в невидимом.

Следует, однако, признать, что более глубокие эпистемологические причины наблюдаемости не раскрываются и в квантовой теории. По нашему мнению, эти причины кроются в феноменологии сложности проявленного бытия. Любая реальность, будучи объектом наблюдения (иными словами, вступая в отношения «субъект — объект»), проявляет свою сущность в неких ограниченных формах, достижимых для восприятия субъектом. Проявленная сложность отмеченных форм (потенциально она бесконечна — ведь материя не имеет пределов делимости, так сказать, «вглубь» и не допускает контроля всех ее взаимосвязей и опосредствований) оказывается компенсированной, согласованной с «перцептивно-сенсорным» аппаратом воспринимающего субъекта или прибора (что, в принципе, одно и то же). Механизм самокомпенсации сложности представляется нам настолько универсальной «универсалией» отношения «субъект — объект», что его, возможно, стоит отнести к основополагающим принципам системного анализа. Отметим, что сложность здесь мы понимаем как позитивную научную категорию, для которой могут быть сформу-

лированы точные и даже формальные определения типа тех, что имеют место в теории А. Н. Колмогорова [6] и в обширных развитиях этой теории, где понятие сложности оказывается непосредственно связанным с понятием информации.

3. Очерк теории семантических состояний

3.1. Понятие семантического состояния языковой единицы

Насколько нам известно, это понятие также впервые было использовано А. Н. Колмогоровым — он глубоко интересовался и лингвистическими вопросами — при попытке формального определения понятия падежа [7]. Применение понятия семантического состояния получило дальнейшее развитие в работах [8] — [14], согласно которым любое слово (вообще говоря — любая единица языка) в контексте или в языковом потоке находится в определенном семантическом состоянии. Для единиц лексического уровня это состояние представляет собой некую сумму признаков грамматической и лексической семантики. Собственно, сам процесс понимания языка с этой позиции выглядит как редукция априорного распределения лексем по тем признакам грамматической и лексической семантики, которые присущи субъективному лексикону реципиента, к одному определенному грамматическому и лексическому значению, характерному именно для того контекста, который находится в поле внимания реципиента и подлежит в данный момент процессу его индивидуальной языковой обработки.

Несмотря на то, что и до сих пор не создан формальный аппарат для определения системы всех возможных семантических состояний любой лексики, естественно считать, что достаточно адекватной моделью данной системы может служить комплекс грамматических и лексических значений — так, как он представлен, скажем, в большом толковом словаре того или иного языка.

Таким образом, при рассмотрении формальных аспектов семантики мы будем исходить из существования соответствия между языковой единицей и ее семантическим состоянием:

$$g : X \rightarrow g(X), \quad (1)$$

где X — некоторая единица языка; g — соответствие между X и $g(X)$ — формальным объектом, представляющим семантическое состояние единицы X , который имеет своими детерминантами элементы средств материального выражения семантики. Для любой единицы X семантические состояния образуют определенное множество — будем считать его конечным, но не ограниченным и обозначим символом $\{g(X)\}$. Класс определенных единиц языка L обозначим символом $W(L)$ или просто W , если речь идет лишь об одном конкретном языке; принадлеж-

ность X к W обозначим: $X \in W$; множество всех семантических состояний для всех $X \in W$ обозначим через $\{g(X); X \in W\} \equiv G$.

Допустим, что существует оператор F , действие которого определено на множестве семантических состояний G и который будем интерпретировать как оператор значения некой семантической категории (ими могут быть, кстати, и категории грамматической семантики — если, например, F является оператором части речи, то его значениями являются: f_1 — «имя существительное», f_2 — «глагол», f_3 — «имя прилагательное» и т. д.). Это означает, что F представляет собой некий интеллектуальный механизм, который, анализируя семантическое состояние $g(X)$, в котором находится единица X , идентифицирует определенное значение конкретной семантической категории, отвечающей именно данному семантическому состоянию. Понятно, что установление системы операторов типа F предполагает наличие некой семантической теории, формализованной надлежащим образом.

На математическом языке действие оператора F может быть выражено следующим образом:

$$Fg_i(X) = f_i g_i(X), \quad (2)$$

где f_i — некоторое значение семантической категории; функции $g_i(X)$, $i = 1, 2, \dots$, представляют семантические состояния единицы X , маркирующие ее принадлежность значению f_i категории F . Значения f_i , $i = 1, 2, \dots$, будем называть *собственными значениями оператора F* , отвечающими семантическим состояниям $g_i(X)$, $i = 1, 2, \dots$. Множество всех семантических состояний, отвечающих собственному значению f_i , будем называть множеством *частных семантических состояний* и обозначим символом $G(f_i)$:

$$G(f_i) := \{g: Fg = f_i g_i\}. \quad (3)$$

По определению множество $G(f_i)$ состоит только из тех семантических состояний, которые характеризуются определенным значением семантической категории F , а именно — значением f_i .

Уравнение (2) в тех случаях, когда множество, которое пробегает индекс i , состоит более чем из одного элемента, представляет формальное выражение явления семантической многозначности. Для единиц лексического уровня — это полисемия и омонимия (как лексическая, так и грамматическая). В частности, в случае омонимии это отображает ситуацию, когда словоформа X в одном контексте может находиться в состоянии, скажем, существительного, а в другом — в состоянии, например, глагола. Такая разнovidность омонимии обычно называется *межчастеречной омонимией*.

Возможны случаи, когда в уравнении (2) определенному значению f_i соответствует не одно семантическое состояние $g_i(X)$, а несколько: $g_j^i(X)$, $j = 1, 2, \dots$.

Такие состояния будем называть *вырожденными*; о вырожденности сигнализирует наличие верхнего индекса возле символа семантического состояния $g_j^i(X)$. Например, в украинском языке словоформа «МАТИ» со значением категории «часть речи» «имя существительное» имеет два грамматических состояния:

$$g_{\text{іменник}}^{\text{жін. рід, одна, наз. відм.}}$$

где слово «МАТИ» имеет лексическое значение «Жінка стосовно дитини, яку вона народила», и:

$$g_{\text{іменник}}^{\text{чол. рід, множина, наз. відм.}}$$

где слово «МАТИ» имеет лексическое значение «Спортивні матраци». Этот пример иллюстрирует явление *внутреннечастеречной омонимии*.

Количество собственных семантических состояний единицы X , соответствующих определенному значению категории F , будем называть *кратностью вырождения этого семантического состояния*. Так, в рассмотренном примере состояние $g(\text{МАТИ})$ имеет кратность вырождения 2:

$$g_{\text{іменник}}^{\text{жін. рід, одна, наз. відм.}} \text{ та } g_{\text{іменник}}^{\text{чол. рід, множина, наз. відм.}}$$

Семантические состояния, в которых оператор F имеет определенное значение, будем называть *чистыми*. Однако а priori нельзя запрещать существование семантических состояний, для которых оператор F не приобретает одного определенного значения, но может характеризоваться, например, двумя. Формально такую ситуацию можно изобразить с помощью соотношения:

$$Fg(X) = f_1 \alpha_1(g_1) g_1(X) + f_2 \alpha_2(g_2) g_2(X), \quad (4)$$

где семантическое состояние g языковой единицы X при действии на него оператора F расщепляется на два, а именно: $g_1(X)$ и $g_2(X)$, где $g_1(X)$ соответствует значению семантической категории f_1 , а $g_2(X)$ — значению семантической категории f_2 ; лингвистический смысл функций $\alpha_1(g_1)$ и $\alpha_2(g_2)$ будет выяснен ниже.

Состояния, для которых оператор F , действуя на функцию семантического состояния, изображается комбинацией определенного числа чистых семантических состояний, соответствующих различным собственным значениям этого оператора, будем называть *смешанными семантическими состояниями*.

Таким образом, уравнение (4) определяет семантическое состояние, в котором находится единица X и которое представляет собой своеобразное явление суперпозиции («смеси») чистых семантических состояний $g_1(X)$ и $g_2(X)$, отвечающих собственным значениям f_1 и f_2 , соответственно. Лингвистическая интерпретация уравнения (4) состоит в том, что единица X имеет семантические признаки одновременно и f_1 и f_2 . Соответствующие показатели —

идентификаторы отмеченной принадлежности содержатся в выражениях для семантических состояний $g_1(X)$ и $g_2(X)$.

Такая ситуация довольно типична в языке. Например, в украинских и русских причастиях соединены свойства глагола и прилагательного. Рассмотрим, например, русские причастные лексемы «ведущий» и «ведомый». Они склоняются по словоизменительной парадигме прилагательного (по шесть падежей в мужском, женском и среднем родах и в множественном числе), имея одновременно в своей структуре глагольный морфологический признак активности/пассивности — он приобретает здесь материальное выражение в виде суффиксов *-ущ* и *-ом*, соответственно. Отмеченный морфологический признак не выглядит сильным с точки зрения принадлежности к глаголу, где он не приобрел статуса словоизменительного; кроме того, он характерен для полной словоизменительной парадигмы, а не только для отдельных ее членов — это дает основание для выделения класса слов с такими свойствами в самостоятельную часть речи — причастие. В украинском же переводе причастие звучит как «дієприкметник», и в этом названии нашли свое отражение оба начала — глагола и прилагательного.

3.2. Семантические состояния в формализме нечетких множеств

Подчеркнем принципиальное отличие между описанной ситуацией и обычной омонимией. Явление омонимии также можно описать уравнением типа (4): $Fg(X) = f_1\alpha_1(g_1)g_1(X) + f_2\alpha_2(g_2)g_2(X) + \dots$, где разные члены правой части соответствуют различным омонимическим состояниям слова X — если ему присуща та или иная разновидность омонимии. Но в процессе языковой обработки, когда происходит снятие омонимии, правая часть этого уравнения редуцируется к одному члену, который, собственно, и представляет чистое грамматическое состояние анализируемого слова X в конкретном контексте. Совсем иная ситуация, когда X находится в смешанном состоянии — в этом случае никакая языковая обработка не способна уменьшить количество членов уравнения (4), которое и представляет ее конечный результат. Такое положение сигнализирует о существовании единиц, которые в контексте функционируют в смешанных семантических состояниях.

Вывод о существовании единиц, для которых язык допускает функционирование только в смешанных семантических состояниях, выглядит весьма интересным свидетельством невозможности полной формализации языковой системы, точнее, свидетельством пределов этой формализации. Это, в частности, означает, что даже наиболее формальная область теории языка, каковой является грамматика, имеет черты нечеткости, что входит в опре-

деленное противоречие с грамматическим детерминизмом и стимулирует выработку специального языка описания многозначных грамматических ситуаций.

Концептуальную базу такого языка, по нашему мнению, предоставляет теория нечетких множеств Л. Заде [15]. Продемонстрируем, каким образом это достигается.

Обратимся к определению множества G семантических состояний и множества $G(f_i) = \{g: fg = fg_i\}$ частичных семантических состояний. Если бы существовала принципиальная возможность редукции в процессе языковой обработки любого семантического состояния к чистому, то множество G можно было бы представить в виде объединения множеств $G(f_i)$, которые между собой не пересекаются, то есть справедливой была бы формула:

$$G = \cup_i G(f_i); G(f_i) \cap G(f_j) = \emptyset \text{ при } i \neq j. \quad (5)$$

Ситуация совсем иная, если мы учтем описанную возможность существования единиц, одновременно характеризующихся, например, двумя (или более) значениями определенной семантической категории. Семантическое состояние $g(X)$ такой единицы X уже не будет принадлежать лишь одному из подмножеств $G(f_i)$, но одновременно двум, а возможно, даже и больше.

Формальный механизм описания подобных явлений осуществляется следующим образом. Определим на множестве $G = \cup_i G(f_i)$ структуру нечеткого в смысле Заде множества. Для этого на каждом из подмножеств $G(f_i)$ определим функцию принадлежности $\alpha_i(g)$, которая для каждого $g \in G(f_i)$ приобретает определенное числовое значение из интервала $[0, 1]$:

$$\alpha_i(g) \in [0, 1]. \quad (6)$$

При этом будем полагать, что если $\alpha_i(g) = 1$, то состояние g является чистым, а если $\alpha_i(g) < 1$, соответствующий g является компонентом смешанного состояния, второй компонент которого g' (для двухкомпонентных состояний) принадлежит определенному подмножеству $G(f_j)$, $i \neq j$, со значением функции принадлежности $\alpha_j(g')$, также меньшим 1, но таким, чтобы выполнялось условие:

$$\alpha_i(g) + \alpha_j(g') = 1. \quad (7)$$

Тогда каждое из подмножеств $G(f_i)$ превращается в нечеткое множество с функцией принадлежности $\alpha_i(g)$:

$$G(f_i) \rightarrow \{G(f_i), \alpha_i(g), g \in G(f_i)\}. \quad (8)$$

Элементами определенного таким образом нечеткого множества служат пары $(g(f_i), \alpha_i(g))$. Одновременно структура нечеткого множества индуцируется и на всем множестве G как объединении нечетких множеств (8) при выполнении условия (7).

Функция принадлежности $\alpha_i(g)$ при этом получает интерпретацию как мера приобретения семантическим состоянием g свойств значения f_i категории F . Максимальное значение, равное 1, функция принадлежности приобретает на чистых состояниях, характеризующих слова с однозначно определенным показателем принадлежности к определенной семантической категории.

Для слов со смешанными семантическими состояниями вида:

$$g(X) = \alpha_1(g_1)g_1(X) + \alpha_2(g_2)g_2(X) \quad (9)$$

величина $\alpha_1(g_1)$ демонстрирует степень, в которой языковой объект X проявляет свойства значения f_1 , а $\alpha_2(g_2)$, соответственно, — степень, в которой X проявляет свойства значения f_2 . Условие (7), а именно: $\alpha_1(g_1) + \alpha_2(g_2) = 1$, обеспечивает полноту семантических свойств рассматриваемого объекта и замкнутость семантического описания.

3.3. Формальные аспекты отношения синонимии и аналогия с определением переводных эквивалентов

В этом пункте рассмотрим некоторые формальные аспекты отношения лексической синонимии, поскольку данное отношение играет весьма важную роль в языке, в частности, в процессах межъязыковой коммуникации.

Будем считать, что лексемы x и y находятся в отношении синонимии S , то есть xSy , если существуют семантические состояния (состояния лексической семантики) $c(x)$ и $c(y)$, которые являются близкими: $c(x) \sim c(y)$. Это означает, что:

$$|c(x) - c(y)| < \varepsilon, \quad (10)$$

где ε — некоторая достаточно малая величина.

Понятно, что для оценки величины $|c(x) - c(y)|$ необходимо иметь такие формальные определения для $c(x)$ и $c(y)$, которые позволяют вычислить разность $c(x) - c(y)$ и оценить ее абсолютную величину. Это довольно сложная задача, которая обычно решается путем экспертной оценки, выполняемой лингвистом, который, анализируя значения (семантические состояния) $c(x)$ и $c(y)$ и полагаясь на свою лингвистическую компетенцию, сам устанавливает их близость.

Не останавливаясь на чисто лингвистических вопросах, отметим некоторые формальные свойства отношения синонимии.

Очевидно, что рефлексивно и симметрично, то есть: xSx и $xSy \Rightarrow ySx$. Однако отношение синонимии не обязательно является транзитивным, то есть из фактов xSy и ySz , вообще говоря, не следует xSz . Это означает, что разности $|c(x) - c(y)|$ и $|c(y) - c(z)|$ могут накапливаться так, что $|c(x) - c(z)| > \varepsilon$. Мы рассматриваем, однако, лишь такие синонимические ряды (синсеты) SX , что если $X = \{x_1, x_2, \dots, x_n\}$ — множество лексем, составляющих некоторый синсет SX , то $x_i S x_j$ для всех $i, j = 1, 2, \dots, n$. Это дает возмож-

ность обозначить множество значений синсета SX через $C(SX)$. Следовательно, для любых $c(x) \in C(SX)$ и $c(y) \in C(SX)$ $|c(x) - c(y)| < \varepsilon$.

Отметим принципиальную разницу между описанием лексической семантики в толковых и синонимических словарях. А именно, толковый словарь стремится представить, так сказать, абсолютную семантику, то есть для каждой лексемы x определить и как можно полнее описать множество ее семантических состояний:

$$C(x) = \{c_i(x) \mid i = 1, 2, \dots, k\}, \quad (11)$$

причем таким образом, чтобы для любого контекста $M(x)$, содержащего форму x , нашелся бы элемент $c_M(x) \in C(x)$, представляющий значение лексемы x и в этом контексте. Понятно, что множество $C(x)$ содержит как чистые, так и смешанные семантические состояния.

В то же время семантика синонимии относительно и в формализме семантических состояний представляется всего лишь разностью $|c(x) - c(y)|$ семантических состояний $c(x)$ и $c(y)$, удовлетворяющих условию близости в смысле, изложенном выше.

На практике семантика толковых словарей, как правило и к сожалению, не согласована с семантикой синонимических, то есть определение семантических состояний при синонимии, зафиксированное в формуле (10), не всегда согласуется с определениями состояний лексической семантики, представленными в формуле (11). Отмеченный факт создает значительные проблемы при разработке семантических анализаторов и двуязычных переводных словарей и систем лексикографического обеспечения машинного перевода.

Действительно, между определением синонимии и нахождением переводных эквивалентов существует глубокая аналогия. Она базируется на том обстоятельстве, что, если нахождение синонима сводится к поиску близких значений лексем одного языка, то нахождение переводного эквивалента есть тот же поиск близких значений лексем, но разных языков. Введем верхний индекс возле языковой единицы и ее семантического состояния для обозначения языка, в котором эта единица (или состояние) представлены. Тогда синонимия отображается соотношением: $|c^1(x) - c^1(y)| < \varepsilon$.

В то же время отношение «переводной эквивалент»: $x^1 T x^2$ изображается соотношением $|c^1(x^1) - c^2(x^2)| < \varepsilon$. Понятно, что для формального определения последней процедуры необходимо привести величины $c^1(x^1)$ и $c^2(x^2)$ к какому-то единому представлению, которое может предоставить, например, язык-посредник — обозначим ее верхним индексом 0. Таким образом, сначала имеем отображение $c^1(x^1) \rightarrow c^0(x)$ и $c^2(x^2) \rightarrow c^0(y)$. Тогда получаем эквивалентность:

УДК 519.7

МОДЕЛЬ СГЛАЖИВАНИЯ В ЗРЕНИИ

М.Ф. Бондаренко¹, С.Ю. Шабанов-Кушнарченко², Ю.П. Шабанов-Кушнарченко³

¹ ХНУРЭ, г. Харьков, Украина

² ХНУРЭ, г. Харьков, Украина, sergabaev@mail.ru

³ ХНУРЭ, г. Харьков, Украина

Исследуются задачи построения тепловой модели, модели инерции и иррадиации. Выведен и обобщен закон Талбота. Доказана равносильность модели инерции и иррадиации зрения и обобщенного закона Талбота для случая нестационарных однородных зрительных картин при однокомпонентном зрении. Приведена экспериментальная проверка обобщенного закона Талбота.

МОДЕЛЬ ИНЕРЦИИ И ИРРАДИАЦИИ, МОДЕЛЬ ТЕПЛОВАЯ, ЗАКОН ТАЛБОТА, КАРТИНА ЗРИТЕЛЬНАЯ, ОПЕРАТОР ПРЕОБРАЗОВАНИЯ.

Введение

В статье [1] было показано, что оператор зрения можно представить в виде двухкаскадной модели, состоящей из последовательно соединенных блоков 1 и 2. Там же было отмечено, что оператор F , описывающий работу блока 2 этой модели, нельзя признать взаимно однозначным. Такую неоднозначность вносят инерция и иррадиация зрения. Существуют совершенно различные зрительные картины (например, быстрые периодические световые мелькания и немигающий свет), которые дают одинаковые зрительные ощущения. В настоящей работе ставится задача: разделить блок 2 модели на два последовательно соединенных блока, первый из которых должен вобрать в себя неоднозначность в преобразовании сигналов, второй же блок должен осуществлять взаимно однозначное преобразование сигналов.

Решение поставленной задачи оказывается возможным на основе использования закона Талбота [2]. Получаемая в результате модель имеет вид схемы, изображенной на рис. 1.

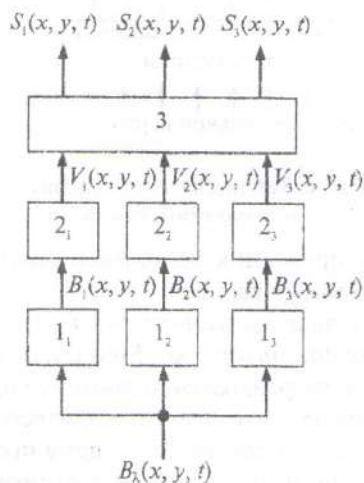


Рис. 1. Модель подавления избытка информации в органе зрения человека

Блоки 1_i ($i = 1, 2, 3$), вместе взятые, образуют блок 1 прежней модели. Блок 2 прежней модели теперь расчленен на два последовательно соединенных яруса, а именно: нижний ярус преобразования в виде трех блоков 2_i ($i = 1, 2, 3$) и верхний ярус преобразования в виде блока 3.

Работу блоков 2_i описываем линейным оператором следующего вида:

$$V_i(x, y, t) = \frac{k_i}{4\pi b_i^2} \int_{-\infty}^t \int \int B_i(\xi, \eta, \tau) \frac{e^{-\frac{t-\tau}{a_i}}}{t-\tau} e^{-\frac{(\xi-x)^2 + (\eta-y)^2}{4b_i^2 \frac{t-\tau}{a_i}}} d\xi d\eta d\tau, \quad (1)$$

где a_i, b_i, k_i — некоторые константы зрения. Функционирование блока 3 описывается оператором:

$$\bar{S}(x, y, t) = \Phi(\bar{V}(x, y, t)), \quad (2)$$

где Φ — произвольно выбираемый взаимно однозначный непрерывный оператор; $\bar{V}(x, y, t)$ — вектор-функция с компонентами $V_1(x, y, t), V_2(x, y, t), V_3(x, y, t)$; $\bar{S}(x, y, t)$ — вектор-функция зрительного ощущения с компонентами $S_1(x, y, t), S_2(x, y, t), S_3(x, y, t)$.

Следует отметить, что в частном случае при однородных зрительных картинах в предположении, что Φ — тождественный оператор, формула (1) запишется в виде:

$$V_i(t) = \frac{k_i}{a_i} \int_{-\infty}^t B_i(\tau) e^{-\frac{t-\tau}{a_i}} d\tau, \quad (3)$$

что совпадает с формулой Луизова, предложенной им для описания инерции ахроматического зрения [3, с. 17, 31]. Все дальнейшее изложение в настоящей статье будет направлено на уточнение и обоснование предлагаемой схемы преобразования сигналов в органе зрения.

1. Содержательная постановка задачи

В чем же состоит значение введенной схемы преобразования сигналов? Дело в том, что инерцию и иррадиацию зрения можно трактовать как своего рода фильтр, предохраняющий мозг от поступления в него количества информации большего, чем он в состоянии переработать. Именно благодаря инерции и иррадиации зрения мы многие существенно различные входные сигналы не различаем. Инерция и иррадиация зрения выполняют, в сущности, функцию, аналогичную той, которую в схеме на рис. 1 осуществляют блоки 1₁. Однако там подавление избыточной информации ведется по длине волны λ . Инерция же осуществляет подавление избытка информации по времени t , а иррадиация — по полю зрения x, y . Эти три механизма, вместе взятые, можно рассматривать как полную систему подавления избытка информации, имеющуюся на входе органа зрения человека. Все вышележащие механизмы органа зрения обрабатывают поток зрительной информации таким образом, что дальнейшей потери и подавления информации не происходит.

Построенная модель, представленная на рис. 1, интересна в том отношении, что позволяет выделить из общего преобразования информации в органе зрения все механизмы, осуществляющие подавление избытка информации. Эти механизмы представлены на схеме рис. 1 блоками 1 и 2. Они, и только они, осуществляют согласование объема информации, поступающей в мозг через орган зрения, с пропускной способностью мозга. В связи с этим такую модель можно рассматривать как модель подавления избытка информации, поступающей в мозг через орган зрения.

Блок 3 содержит в себе в нерасшифрованном виде все остальные механизмы обработки информации органа зрения человека, осуществляющие преобразование потока информации без дальнейшего его сокращения. Так как работа блока 3 описывается взаимно однозначным оператором, то различным сигналам на входе этого блока всегда будут соответствовать различные сигналы на его выходе. Понимая инерцию и иррадиацию зрения лишь как фильтр, подавляющий избыток информации во времени и в поле зрения, мы можем считать, что блоки 2₁ и только они, представляют собой оператор инерции и иррадиации зрения. В настоящей работе показывается, что предлагаемая модель, с одной стороны, и орган зрения человека, с другой, равносильны в роли фильтра, подавляющего избыток информации.

2. Тепловая модель

Для того чтобы естественным образом подойти к построению и обоснованию модели подавления

избытка информации в органе зрения в том виде, как она была представлена на рис. 1, мы построим вначале некоторую гипотетическую модель инерции и иррадиации зрения, воспользовавшись аналогией между процессами инерции и иррадиации зрения, с одной стороны, и тепловыми процессами, с другой. На плодотворность такой аналогии при исследовании инерции зрения указывал еще Аллар [4]. Построенная таким образом модель будет удобным инструментом при решении поставленной задачи.

Уподобим орган зрения некоторой искусственной системе (тепловой модели), представленной на рис. 2. Свет от зрительной картины $b_\lambda(x, y, t)$ проецируется на двумерное поле бесконечно малых фотоприемников, осуществляющих преобразование сигналов по формуле:

$$B(x, y, t) = \int_{\lambda_1}^{\lambda_2} b_\lambda(x, y, t) k(\lambda) d\lambda. \quad (4)$$

«Зрение» тепловой модели полагаем однокомпонентным. Под $k(\lambda)$ будем для определенности понимать кривую видности, являющуюся одной из функций сложения фотоприемников глаза. Тогда сигнал $B(x, y, t)$, передаваемый по проводнику, отходящему от каждого фотоприемника, будет представлять собой фотометрическую яркость зрительной картины в различных ее точках в различные моменты времени t .

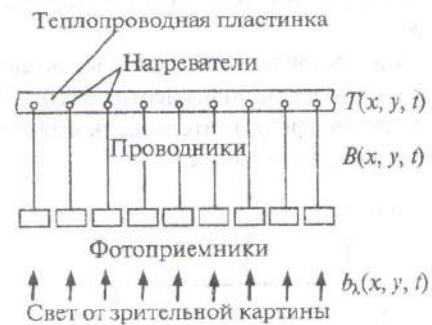


Рис. 2. Тепловая модель инерции и иррадиации зрения

Каждый проводник оканчивается нагревателем бесконечно малых размеров. Эти нагреватели расположены в виде двумерного поля внутри тонкой теплопроводной пластинки. Мощность, развиваемую каждым нагревателем, принимаем пропорциональной численному значению сигнала $B(x, y, t)$, передаваемого по соответствующему проводнику. Тепловое поле, возникающее в пластинке в результате работы нагревателей, уподобим зрительному ощущению, причем температуру $T(x, y, t)$ пластинки

ки будем считать аналогом светлоты зрительного ощущения.

Функционирование тепловой модели имеет много общего с явлениями инерции и иррадиации зрения. При выключении света зрительное ощущение не исчезает мгновенно, а его светлота уменьшается до нуля постепенно. Аналогичный процесс произойдет и в тепловой модели. Если прекратить освещение фотоприемников, произойдет постепенное остывание теплопроводной пластинки и плавное падение ее температуры. При включении света светлота зрительного ощущения нарастает постепенно и через определенное время устанавливается на некотором постоянном уровне. Аналогично этому при внезапном освещении фотоприемников температура теплопроводной пластинки будет нарастать постепенно и, в конце концов, установится на постоянном уровне.

Пусть теперь зрению человека предъявляется зрительная картина, фотометрическая яркость которой периодически изменяется во времени (мелькания). Если частота мельканий невелика, то зрительное ощущение будет испытывать подобное колебание светлоты. Однако при повышении частоты амплитуда колебаний светлоты ощущения будет постепенно уменьшаться, и при некоторой (критической) частоте колебания светлоты вовсе не будут обнаруживаться сознанием наблюдателя. Подобным образом будет реагировать также и тепловая модель. При повышении частоты мельканий амплитуда колебания температуры пластинки будет уменьшаться и, в конце концов, достигнет настолько малого значения, что оно не будет обнаруживаться приборами, измеряющими температуру пластинки, что обусловлено их конечной чувствительностью.

Рассматривая границу между черным и белым, человек видит ее несколько размытой. Аналогичное явление будет наблюдаться и в тепловой модели, если одну часть поля фотоприемников осветить, а другую оставить затемненной. На линии, отделяющей яркую часть зрительной картины от темной, наблюдается скачкообразное изменение яркости. Однако на соответствующей линии, проведенной на теплопроводной пластинке, температура будет меняться плавно. Развивая аналогию в том же направлении, можно воспроизвести на тепловой модели явления слияния густо расположенных ярких и темных полос, образования следа за ярким быстро движущимся объектом, приближения светлоты зрительного ощущения объекта при уменьшении его размеров к светлоте фона и другие явления, связанные с инерцией и иррадиацией зрения [5, гл. 7].

Опишем математически процессы, происходящие в тепловой модели, и определим зависимость температуры $T(x, y, t)$ пластинки от фотометриче-

ской яркости $B(x, y, t)$ зрительной картины. Мощность q , развиваемую нагревателями на единице площадки пластинки, определим по формуле:

$$q = k_1 B, \quad (5)$$

где k_1 — коэффициент пропорциональности. Введем на пластинке декартову систему координат и выделим на ней элементарную прямоугольную площадку с размерами dx, dy (рис. 3).

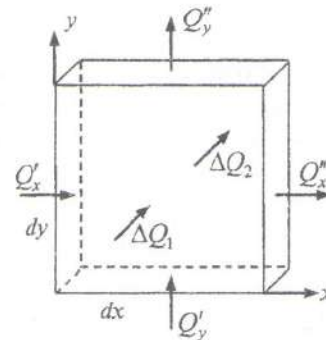


Рис. 3. Декартова система координат в тепловой модели

Количество тепла ΔQ_1 , поступающее за время dt на эту площадку за счет работы нагревателей, равно:

$$\Delta Q_1 = q dx dy dt. \quad (6)$$

Количество тепла ΔQ_2 , отводимое за время dt стенками пластинки в пределах элементарной площадки, определим, основываясь на законе Ньютона-Рихмана [6]:

$$\Delta Q_2 = k_2 T dx dy dt, \quad (7)$$

где k_2 — коэффициент пропорциональности. Количество тепла Q'_x, Q''_x, Q'_y, Q''_y , проходящего за время dt через сечения элементарной площадки, будем определять, исходя из закона Био-Фурье [7]:

$$Q'_x = k_3 \frac{\partial T}{\partial x} dy dt, \quad (8)$$

$$Q''_x = k_3 \frac{\partial}{\partial x} \left(T + \frac{\partial T}{\partial x} dx \right) dy dt, \quad (9)$$

$$Q'_y = -k_3 \frac{\partial T}{\partial y} dx dt, \quad (10)$$

$$Q''_y = -k_3 \frac{\partial}{\partial y} \left(T + \frac{\partial T}{\partial y} dy \right) dx dt, \quad (11)$$

где k_3 — коэффициент пропорциональности.

За время dt количество тепла, содержащееся в элементе пластинки, изменится на величину ΔQ , равную:

$$\Delta Q = \Delta Q_1 - \Delta Q_2 + Q'_x - Q''_x + Q'_y - Q''_y. \quad (12)$$

Подставляя в полученное равенство значения входящих в него величин по формулам (5)–(11), получим:

$$\Delta Q = (k_1 B - k_2 T + k_3 (\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2})) dx dy dt. \quad (13)$$

Полагая, что изменение теплосодержания элемента вызывает пропорциональное изменение его температуры (закон постоянства теплоемкости [8, стр. 392]), можем записать следующее равенство:

$$\Delta Q = k_4 \frac{\partial T}{\partial t} dx dy dt, \quad (14)$$

где k_4 — коэффициент пропорциональности.

Приравнивая правые части формул (13) и (14), будем иметь:

$$a \frac{\partial T}{\partial t} - b^2 (\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}) + T = k B. \quad (15)$$

В уравнении (15) приняты следующие новые обозначения:

$$a = \frac{k_4}{k_3}, \quad b = \sqrt{\frac{k_3}{k_2}}, \quad k = \frac{k_1}{k_2}. \quad (16)$$

Формула (15) представляет собой уравнение тепловой модели. Это уравнение дает нам в дифференциальной форме искомую связь температуры пластинки T с фотометрической яркостью B зрительной картины.

3. Модель инерции и иррадиации

Руководствуясь аналогией между зрением и процессами, происходящими в тепловой модели, будем предполагать, что связь между фотометрической яркостью B зрительной картины и светлотой S зрительного ощущения описывается уравнением, аналогичным уравнению (15) тепловой модели. Тогда мы придем к следующему уравнению, устанавливающему в дифференциальной форме связь между фотометрической яркостью $B(x, y, t)$ зрительной картины и светлотой $S(x, y, t)$ зрительного ощущения, возникающего в сознании наблюдателя при восприятии этой картины:

$$a \frac{\partial S}{\partial t} - b^2 (\frac{\partial^2 S}{\partial x^2} + \frac{\partial^2 S}{\partial y^2}) + S = k B. \quad (17)$$

Функции $B(x, y, t)$ и $S(x, y, t)$ будем считать определенными и ограниченными при любых значениях x, y, t .

Необходимо отметить, что в частном случае, когда зрительная картина однородна и зависит лишь от времени $B=B(t)$, мы приходим к уравнению, предложенному для описания инерции зрения Алларом [4]:

$$a \frac{dS}{dt} + S = k B. \quad (18)$$

С помощью подстановки

$$G(x, y, t) = e^{\frac{t}{a}} S(x, y, t) \quad (19)$$

выражение (17) преобразуется в уравнение теплопроводности [9]:

$$a \frac{\partial G}{\partial t} - b^2 (\frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2}) = k e^{\frac{t}{a}} B, \quad (20)$$

ограниченное решение которого имеет вид:

$$G(x, y, t) = \frac{k}{4\pi b^2} \int_{-\infty}^t \int_{-\infty}^{\infty} \int B(\xi, \eta, \tau) \frac{e^{\frac{\tau}{a}}}{t-\tau} e^{-\frac{(\xi-x)^2 + (\eta-y)^2}{4b^2 \frac{t-\tau}{a}}} d\xi d\eta d\tau. \quad (21)$$

Переходя к функции $S(x, y, t)$, получаем решение уравнения (17):

$$S(x, y, t) = \frac{k}{4\pi b^2} \int_{-\infty}^t \int_{-\infty}^{\infty} \int B(\xi, \eta, \tau) \frac{e^{-\frac{\tau}{a}}}{t-\tau} e^{-\frac{(\xi-x)^2 + (\eta-y)^2}{4b^2 \frac{t-\tau}{a}}} d\xi d\eta d\tau. \quad (22)$$

Выражения (17) и (22) можно рассматривать в качестве дифференциальной и интегральной форм математической модели, воспроизводящей, по нашему предположению, явления инерции и иррадиации зрения. В частном случае, когда зрительная картина однородна и нестационарна, то есть когда яркость $B=B(t)$ изменяется лишь со временем и не зависит от координат x, y поля зрения (равномерно освещенный экран), выражение (22) может быть записано в виде:

$$S(t) = \frac{k}{4\pi b^2} \int_{-\infty}^t B(\tau) \frac{e^{-\frac{\tau}{a}}}{t-\tau} d\tau \int_{-\infty}^{\infty} \int e^{-\frac{(\xi-x)^2 + (\eta-y)^2}{4b^2 \frac{t-\tau}{a}}} d\xi d\eta. \quad (23)$$

Переходя в интеграле

$$J = \int_{-\infty}^{\infty} \int e^{-\frac{(\xi-x)^2 + (\eta-y)^2}{4b^2 \frac{t-\tau}{a}}} d\xi d\eta \quad (24)$$

к новым переменным интегрирования:

$$\xi^* = \frac{\xi - x}{2b\sqrt{\frac{t-\tau}{a}}}, \quad (25)$$

$$\eta^* = \frac{\eta - y}{2b\sqrt{\frac{t-\tau}{a}}}, \quad (26)$$

получим:

$$J = 4b^2 \frac{t-\tau}{a} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\xi^{*2} - \eta^{*2}} d\xi^* d\eta^* \quad (27)$$

Вычисление интеграла J ведем в полярных координатах:

$$\begin{aligned} J &= 4b^2 \frac{t-\tau}{a} \int_0^{\infty} e^{-\rho^2} \rho d\rho \int_0^{2\pi} d\varphi = \\ &= 8\pi b^2 \frac{t-\tau}{a} \int_0^{\infty} e^{-\rho^2} \rho d\rho = 4\pi b^2 \frac{t-\tau}{a}. \end{aligned} \quad (28)$$

Используя найденное значение интеграла J , получаем формулу для определения светлоты зрительного ощущения:

$$S(t) = \frac{k}{a} \int_{-\infty}^t B(\tau) e^{-\frac{t-\tau}{a}} d\tau. \quad (29)$$

Эта формула совпадает с формулой, предложенной Луизовым для описания инерции зрения [3].

Рассмотрим теперь функционирование модели в случае стационарной неоднородной зрительной картины $B(x, y)$. Этот случай соответствует рассмотрению наблюдателем зрительной картины произвольного вида, яркость в каждой точке которой не меняется со временем. Выражение (22) может быть записано в виде:

$$\begin{aligned} S(x, y) &= \\ &= \frac{k}{4\pi b^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} B(\xi, \eta) d\xi d\eta \int_{-\infty}^t \frac{e^{-\frac{t-\tau}{a}}}{t-\tau} e^{-\frac{(\xi-x)^2 + (\eta-y)^2}{4b^2 \frac{t-\tau}{a}}} d\tau. \end{aligned} \quad (30)$$

Переходя в интеграле

$$J_1 = \int_{-\infty}^t \frac{e^{-\frac{t-\tau}{a}}}{t-\tau} e^{-\frac{(\xi-x)^2 + (\eta-y)^2}{4b^2 \frac{t-\tau}{a}}} d\tau \quad (31)$$

к новой переменной интегрирования

$$\tau^* = \frac{t-\tau}{a}, \quad (32)$$

получим [10, стр. 973]:

$$J = \int_0^{\infty} \frac{e^{-\tau^* - \frac{\gamma^2}{4\tau^*}}}{\tau^*} d\tau^* = 2k_0(2\gamma), \quad (33)$$

где

$$\gamma = \frac{\sqrt{(\xi-x)^2 + (\eta-y)^2}}{2b}, \quad (34)$$

k_0 — функция Макдональда с нулевым индексом [11]. Используя найденное значение J_1 , получаем формулу для определения светлоты $S(x, y)$ зрительного ощущения:

$$\begin{aligned} S(x, y) &= \\ &= \frac{k}{2\pi b^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} B(\xi, \eta) k_0\left(\frac{\sqrt{(\xi-x)^2 + (\eta-y)^2}}{b}\right) d\xi d\eta. \end{aligned} \quad (35)$$

Наконец, когда яркость зрительной картины зависит лишь от координаты x , формула (22) примет вид:

$$\begin{aligned} S(x) &= \\ &= \frac{k}{4\pi b^2} \int_{-\infty}^{\infty} B(\xi) d\xi \int_{-\infty}^t \frac{e^{-\frac{t-\tau}{a}}}{t-\tau} e^{-\frac{(\xi-x)^2}{4b^2 \frac{t-\tau}{a}}} d\tau \int_{-\infty}^{\infty} \frac{e^{-\frac{(\eta-y)^2}{4b^2 \frac{t-\tau}{a}}}}{4b^2 \frac{t-\tau}{a}} d\eta. \end{aligned} \quad (36)$$

Используя значение интеграла [12, стр. 407]

$$\int_{-\infty}^{\infty} \frac{e^{-\frac{(\eta-y)^2}{4b^2 \frac{t-\tau}{a}}}}{4b^2 \frac{t-\tau}{a}} d\eta = 2\sqrt{\pi b} \sqrt{\frac{t-\tau}{a}}, \quad (37)$$

будем иметь:

$$S(x) = \frac{k}{2\sqrt{\pi} b} \int_{-\infty}^{\infty} B(\xi) d\xi \int_{-\infty}^t \frac{1}{a} \frac{e^{-\frac{t-\tau}{a}}}{\sqrt{t-\tau}} e^{-\frac{(\xi-x)^2}{4b^2 \frac{t-\tau}{a}}} d\tau. \quad (38)$$

Производя замену переменной интегрирования по формуле

$$\tau_1^2 = \frac{t-\tau}{a}, \quad (39)$$

находим значение интеграла [13]:

$$\int_{-\infty}^{\infty} \frac{1}{a} \frac{e^{-\frac{t-\tau}{a}}}{\sqrt{t-\tau}} e^{-\frac{(\xi-x)^2}{4b^2 \frac{t-\tau}{a}}} d\tau = 2 \int_0^{\infty} e^{-\tau_1^2 - \frac{(\xi-x)^2}{4b^2 \tau_1^2}} d\tau_1 = \sqrt{\pi} e^{-\frac{|\xi-x|}{b}}. \quad (40)$$

Окончательное выражение для определения светлоты зрительного ощущения имеет вид:

$$S(x) = \frac{k}{2b} \int_{-\infty}^{\infty} B(\xi) e^{-\frac{|\xi-x|}{b}} d\xi. \quad (41)$$

В построенную математическую модель инерции и иррадиации зрения входят два параметра a и b . Выясним функциональный смысл каждого из них.

Для выяснения смысла параметра a рассмотрим функционирование модели в случае, когда при $t=0$ яркость зрительной картины равна нулю (затемненный экран), а при $t > 0$ яркость скачкообразно изменяется и во всех точках зрительной картины становится равной некоторому значению B_0 (освещенный экран). Математически изменение яркости во времени запишется в виде следующей системы условий:

$$B(t) = \begin{cases} 0 & \text{при } t \leq 0, \\ B_0 & \text{при } t > 0. \end{cases} \quad (42)$$

На рис. 4а показан ход изменения яркости B зрительной картины с течением времени t .

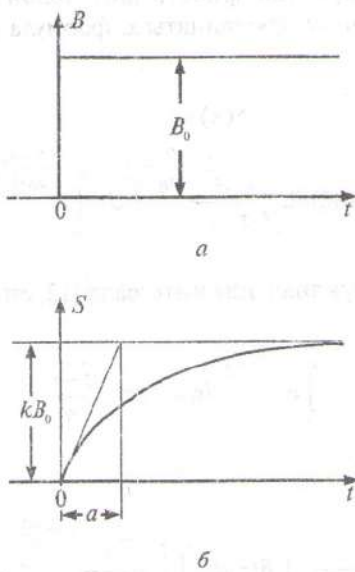


Рис. 4. Изменение яркости во времени

Определим с помощью модели закон изменения во времени светлоты зрительного ощущения $S(t)$. Согласно формуле (29), имеем при $t \leq 0$:

$$S(t) = \frac{1}{a} \int_{-\infty}^t 0 \cdot e^{-\frac{t-\tau}{a}} d\tau = 0, \quad (43)$$

при $t > 0$:

$$S(t) = \frac{1}{a} \int_{-\infty}^0 0 \cdot e^{-\frac{t-\tau}{a}} d\tau + \frac{1}{a} \int_0^t B_0 e^{-\frac{t-\tau}{a}} d\tau = B_0 (1 - e^{-\frac{t}{a}}). \quad (44)$$

Окончательно выражение для определения светлоты запишется в виде:

$$S(t) = \begin{cases} 0 & \text{при } t \leq 0, \\ k B_0 (1 - e^{-\frac{t}{a}}) & \text{при } t > 0. \end{cases} \quad (45)$$

На рис. 4б представлена диаграмма изменения светлоты во времени, построенная по формулам (45). Как видим, при внезапном освещении, согласно математической модели, светлота S должна нарастать постепенно, что и наблюдается в действительности. Параметр a , изображенный на диаграмме в виде отрезка, представляет собой время, за которое светлота достигла бы установившегося значения B_0 , если бы она нарастала со скоростью, имевшей место в момент включения света. Величина параметра a характеризует инерцию зрения: чем больше его значение, тем более инерционным будет зрение человека. Параметр a будем называть постоянной инерции зрения. Необходимо заметить, что постоянную a вводит также Луизов, называя ее временем инерции [3, стр. 26].

Перейдем теперь к выяснению функционального смысла параметра b . Для этого рассмотрим функционирование модели в следующем случае. Поле зрения разделено вертикальной границей $x = 0$ на две половины. Левая половина поля зрения затемнена ($B = 0$), правая — освещена. Яркость правой половины поля зрения во всех точках одинакова, не изменяется со временем и равна некоторой величине B_0 . Математически такая зрительная картина описывается следующим образом:

$$B(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ B_0 & \text{при } x > 0. \end{cases} \quad (46)$$

На рис. 5а показан ход изменения яркости зрительной картины в зависимости от значения координаты x .

В этом случае светлота зрительного ощущения $S(x)$, согласно модели, определится следующим образом. По формуле (41) имеем при $x \leq 0$:

$$S(x) = \frac{1}{2b} \int_{-\infty}^0 0 \cdot e^{-\frac{|\xi-x|}{b}} d\xi + \frac{1}{2b} \int_0^{\infty} B_0 e^{-\frac{\xi-x}{b}} d\xi = \frac{1}{2} B_0 e^{\frac{x}{b}}, \quad (47)$$

при $x > 0$:

$$S(x) = \frac{1}{2b} \int_{-\infty}^0 0 \cdot e^{-\frac{|\xi-x|}{b}} d\xi + \frac{1}{2b} \int_0^x B_0 e^{-\frac{\xi-x}{b}} d\xi + \frac{1}{2b} \int_x^{\infty} B_0 e^{-\frac{\xi-x}{b}} d\xi = B_0 (1 - \frac{1}{2} e^{-\frac{x}{b}}). \quad (48)$$

Окончательно выражение для определения светлоты запишется в виде:

$$S(x) = \begin{cases} \frac{1}{2}kB_0e^{\frac{x}{b}} & \text{при } x \leq 0, \\ kB_0(1 - \frac{1}{2}e^{-\frac{x}{b}}) & \text{при } x > 0. \end{cases} \quad (49)$$

На рис. 5б представлена диаграмма изменения светлоты зрительного ощущения в зависимости от координаты x поля зрения. Как видим, граница между темным и светлым участками выглядит размытой. Это наблюдается и в действительности.

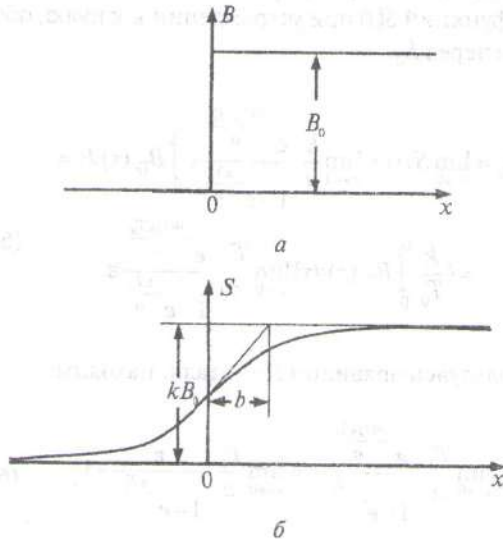


Рис. 5. Зависимость светлоты зрительного ощущения b от координаты x поля зрения

Параметр b , изображенный на диаграмме в виде отрезка, представляет собой угол, в пределах которого видимая яркость достигла бы установившегося значения B_0 , если бы она нарастала с той же скоростью, что и на границе между темным и светлым участками поля зрения. Величина коэффициента B характеризует иррадиацию зрения: чем больше значение этого коэффициента, тем более размытыми, иррадиирующими будут контуры видимых объектов. Параметр b будем называть постоянной иррадиации зрения.

4. Вывод закона Талбота из модели

Приступим теперь к экспериментальной проверке построенной в предыдущем разделе математической модели инерции и иррадиации зрения. Известно, что зрение человека подчиняется психофизическому закону, впервые сформулированному Талботом и названному его именем [2]. Различают временной и пространственный законы Талбота. Предъявим наблюдателю на одном поле сравнения

однородную зрительную картину, яркость которой $B_f(t)$ меняется во времени по произвольному периодическому закону:

$$B_f(t) = B_f(t - nT), \quad (50)$$

где T — период колебания яркости, n — произвольное целое число. Если период T колебания яркости взять достаточно малым, то благодаря инерции зрения мелькания сольются, и наблюдатель увидит немигающее поле.

На втором поле сравнения сформируем однородную и стационарную зрительную картину, характеризующуюся яркостью B_0 . Временной закон Талбота утверждает, что для получения одинаковых зрительных ощущений на полях сравнения необходимо взять уровень яркости B_0 на втором поле сравнения равным среднему значению мелькающей яркости $B_f(t)$, то есть:

$$B_0 = \frac{1}{T_0} \int_0^{T_0} B_{T_0}(t) dt, \quad (51)$$

где T_0 — некоторый произвольно выбранный фиксированный период мельканий.

Для демонстрации пространственного закона Талбота на одном поле сравнения сформируем зрительную картину в виде серии вертикальных полос, яркость $B_x(x)$ которой является периодической функцией координаты x :

$$B_x(x) = B_x(x - nX), \quad (52)$$

где X — период колебания яркости. При достаточно густом расположении полос они благодаря иррадиации зрения сольются в однородный фон. Пространственный закон Талбота гласит, что для достижения одинаковости зрительных ощущений на другом поле сравнения необходимо сформировать стационарную и однородную зрительную картину, яркость B_0 которой должна быть равна среднему значению изменяющейся в поле зрения яркости $B_x(x)$, т. е.:

$$B = \frac{1}{X} \int_0^{X_0} B_{X_0}(x) dx. \quad (53)$$

Модель инерции и иррадиации зрения удовлетворяет как временному, так и пространственному законам Талбота.

Докажем сначала выводимость из модели временного закона Талбота. Найдем выражение для светлоты зрительного ощущения $S(t)$, порождаемого зрительной картиной, яркость которой $B_f(t)$ изменяется во времени по произвольному периодическому закону. Используя зависимость (29), имеем:

Таким образом:

$$S(t) = \frac{1}{a} \int_{-\infty}^t B_T(\tau) e^{-\frac{t-\tau}{a}} d\tau = \frac{1}{a} e^{-\frac{t}{a}} \sum_{n=0}^{\infty} \int_{t-(n+1)T}^{t-nT} B_T(\tau) e^{\frac{\tau}{a}} d\tau. \quad (54)$$

Переходя к новой переменной интегрирования

$$\tau_0 = \tau + nT \quad (55)$$

и используя соотношение (50), получим:

$$\int_{t-(n+1)T}^{t-nT} B_T(\tau) e^{\frac{\tau}{a}} d\tau = \int_{t-T}^t B_T(\tau - nT) e^{\frac{\tau_0 - nT}{a}} d\tau = e^{-\frac{nT}{a}} \int_{t-T}^t B_T(\tau_0) e^{\frac{\tau_0}{a}} d\tau_0 = e^{-\frac{nT}{a}} \int_{t-T}^t B_T(\tau) e^{\frac{\tau}{a}} d\tau. \quad (56)$$

Следовательно:

$$S(t) = \frac{1}{2} e^{-\frac{t}{a}} \left(\sum_{n=0}^{\infty} e^{-\frac{nT}{a}} \right) \int_{t-T}^t B_T(\tau) e^{\frac{\tau}{a}} d\tau. \quad (57)$$

Определяем сумму бесконечно убывающей геометрической прогрессии:

$$\sum_{n=0}^{\infty} e^{-\frac{nT}{a}} = \frac{1}{1 - e^{-\frac{T}{a}}}. \quad (58)$$

Окончательно формула для определения светлоты в случае периодически изменяющейся во времени яркости запишется в виде:

$$S(t) = \frac{k}{a} \cdot \frac{e^{-\frac{t}{a}}}{1 - e^{-\frac{T}{a}}} \int_{t-T}^t B_T(\tau) e^{\frac{\tau}{a}} d\tau. \quad (59)$$

Воспользовавшись обобщенной теоремой о среднем [14], можем записать:

$$S(t) = \frac{k}{a} \cdot \frac{e^{-\frac{t}{a}} e^{-\frac{\theta(t)T}{a}}}{1 - e^{-\frac{T}{a}}} \int_{t-T}^t B_T(\tau) d\tau, \quad (60)$$

где $\theta(t)$ — некоторая функция времени, принимающая значения на интервале $[0, 1]$.

Поскольку $B_T(t)$ — функция периодическая, следовательно:

$$\int_{t-T}^t B_T(\tau) d\tau = \int_0^T B_T(\tau) d\tau. \quad (61)$$

$$S(t) = \frac{k}{a} \cdot \frac{e^{-\frac{\theta(t)T}{a}}}{1 - e^{-\frac{T}{a}}} \int_0^T B_T(\tau) d\tau. \quad (62)$$

Обозначим через T_0 некоторый произвольно выбранный фиксированный период мельканий и определим, во что превратится функция $S(t)$ при неограниченном уменьшении периода мельканий T , равного:

$$T = \varepsilon T_0, \quad (63)$$

то есть при $\varepsilon \rightarrow 0$.

Предел, к которому стремится последовательность функций $S(t)$ при устремлении ε к нулю, обозначим через S_0 :

$$S_0 = \lim_{\varepsilon \rightarrow 0} S(t) = \lim_{\varepsilon \rightarrow 0} \frac{k}{a} \cdot \frac{e^{-\frac{\theta(t)\varepsilon T_0}{a}}}{1 - e^{-\frac{\varepsilon T_0}{a}}} \varepsilon \int_0^{T_0} B_{T_0}(\tau) d\tau = \left(\frac{k}{T_0} \int_0^{T_0} B_{T_0}(\tau) d\tau \right) \lim_{\varepsilon \rightarrow 0} \frac{T_0}{a} \cdot \frac{e^{-\frac{\theta(t)\varepsilon T_0}{a}}}{1 - e^{-\frac{\varepsilon T_0}{a}}} \varepsilon. \quad (64)$$

Пользуясь правилом Лопиталья, находим:

$$\lim_{\varepsilon \rightarrow 0} \frac{T_0}{a} \cdot \frac{e^{-\frac{\theta(t)\varepsilon T_0}{a}}}{1 - e^{-\frac{\varepsilon T_0}{a}}} \varepsilon = \lim_{\varepsilon \rightarrow 0} \frac{T_0}{a} \cdot \frac{\varepsilon}{1 - e^{-\frac{\varepsilon T_0}{a}}} = 1. \quad (65)$$

Таким образом:

$$S_0 = \frac{k}{T_0} \int_0^{T_0} B_{T_0}(\tau) d\tau. \quad (66)$$

Зрительное ощущение такой же светлоты можно получить при наблюдении однородной и стационарной зрительной картины яркости B_0 , равной

$$B_0 = \frac{1}{k} S_0. \quad (67)$$

Объединяя равенства (66) и (67), приходим к формуле (51), что доказывает согласие модели с временной формулировкой закона Талбота. Необходимо отметить, что факт выводимости закона Талбота из формулы (29) для прямоугольных мельканий был доказан Луизовым [3, стр. 43].

Переходим теперь к выводу из модели пространственного закона Талбота. Найдем выражение для светлоты зрительного ощущения $S(t)$, порождаемого зрительной картиной, яркость которой $B_X(x)$ будет изменяться в поле зрения вдоль оси абсцисс

по произвольному периодическому закону. Используя зависимость (41), имеем:

$$\begin{aligned}
 S(x) &= \frac{k}{2b} \int_{-\infty}^{\infty} B_X(\xi) e^{\frac{|\xi-x|}{b}} d\xi = \\
 &= \frac{k}{2b} \int_{-\infty}^x B_X(\xi) e^{\frac{\xi-x}{b}} d\xi + \frac{k}{2b} \int_x^{\infty} B_X(\xi) e^{\frac{\xi-x}{b}} d\xi = \\
 &= \frac{k}{2b} e^{\frac{x}{b}} \sum_{n=0}^{\infty} \int_{x-(n+1)X}^{x-nX} B_X(\xi) e^{\frac{\xi}{b}} d\xi + \\
 &+ \frac{k}{2b} e^{\frac{x}{b}} \sum_{n=0}^{\infty} \int_{x+nX}^{x+(n+1)X} B_X(\xi) e^{\frac{\xi}{b}} d\xi. \quad (68)
 \end{aligned}$$

Переходя к новым переменным интегрирования

$$\xi_1 = \xi + nX, \quad \xi_2 = \xi - nX \quad (69)$$

и используя соотношение (52), получим:

$$\begin{aligned}
 \int_{x-(n+1)X}^{x-nX} B_X(\xi) e^{\frac{\xi}{b}} d\xi &= \int_{x-X}^x B_X(\xi_1 - nX) e^{\frac{\xi_1 - nX}{b}} d\xi_1 = \\
 &= e^{\frac{nX}{b}} \int_{x-X}^x B_X(\xi_1) e^{\frac{\xi_1}{b}} d\xi_1 = e^{\frac{nX}{b}} \int_{x-X}^x B_X(\xi) e^{\frac{\xi}{b}} d\xi; \quad (70)
 \end{aligned}$$

$$\begin{aligned}
 \int_{x+nX}^{x+(n+1)X} B_X(\xi) e^{\frac{\xi}{b}} d\xi &= \int_x^{x+X} B_X(\xi_2 + nX) e^{\frac{\xi_2 + nX}{b}} d\xi_2 = \\
 &= e^{\frac{nX}{b}} \int_x^{x+X} B_X(\xi_2) e^{\frac{\xi_2}{b}} d\xi_2 = e^{\frac{nX}{b}} \int_x^{x+X} B_X(\xi) e^{\frac{\xi}{b}} d\xi. \quad (71)
 \end{aligned}$$

Таким образом:

$$\begin{aligned}
 S(x) &= \frac{k}{2b} e^{\frac{x}{b}} \left(\sum_{n=0}^{\infty} e^{\frac{nX}{b}} \right) \int_{x-X}^x B_X(\xi) e^{\frac{\xi}{b}} d\xi + \\
 &+ \frac{k}{2b} e^{\frac{x}{b}} \left(\sum_{n=0}^{\infty} e^{\frac{nX}{b}} \right) \int_x^{x+X} B_X(\xi) e^{\frac{\xi}{b}} d\xi = \\
 &= \frac{k}{2b} \left(\sum_{n=0}^{\infty} e^{\frac{nX}{b}} \right) \int_{x-X}^x B_X(\xi) e^{\frac{|\xi-x|}{b}} d\xi + \\
 &+ \frac{k}{2b} \left(\sum_{n=0}^{\infty} e^{\frac{nX}{b}} \right) \int_x^{x+X} B_X(\xi) e^{\frac{|\xi-x|}{b}} d\xi. \quad (72)
 \end{aligned}$$

При выводе последней формулы учтено, что для первого слагаемого суммы $x \geq \xi$, для второго слагаемого $x \leq \xi$:

Учитывая, что

$$\sum_{n=0}^{\infty} e^{\frac{nX}{b}} = \frac{1}{1 - e^{\frac{X}{b}}}, \quad (73)$$

получим следующую формулу для определения светлоты в случае периодически изменяющейся вдоль оси Х яркости зрительной картины:

$$S(x) = \frac{k}{2b} \cdot \frac{1}{1 - e^{\frac{X}{b}}} \int_{x-X}^{x+X} B_X(\xi) e^{\frac{|\xi-x|}{b}} d\xi. \quad (74)$$

Воспользовавшись обобщенной теоремой о среднем, запишем:

$$\begin{aligned}
 S(x) &= \frac{k}{2b} \cdot \frac{1}{1 - e^{\frac{X}{b}}} \int_{x-X}^{x+X} B_X(\xi) e^{\frac{|\xi-x|}{b}} d\xi = \\
 &= \frac{k}{2b} \cdot \frac{1}{1 - e^{\frac{X}{b}}} \int_{x-X}^x B_X(\xi) e^{\frac{\xi-x}{b}} d\xi + \\
 &+ \frac{k}{2b} \cdot \frac{1}{1 - e^{\frac{X}{b}}} \int_x^{x+X} B_X(\xi) e^{\frac{\xi-x}{b}} d\xi = \\
 &= \frac{k}{2b} \cdot \frac{e^{\frac{x}{b}} e^{-\frac{x-X_1(x)X}{b}}}{1 - e^{\frac{X}{b}}} \int_{x-X}^x B_X(\xi) d\xi + \\
 &+ \frac{k}{2b} \cdot \frac{e^{\frac{x}{b}} e^{\frac{x+X_2(x)X}{b}}}{1 - e^{\frac{X}{b}}} \int_x^{x+X} B_X(\xi) d\xi = \\
 &= \frac{k}{2b} \cdot \frac{e^{-\frac{X_1(x)X}{b}}}{1 - e^{\frac{X}{b}}} \int_{x-X}^x B_X(\xi) d\xi + \\
 &+ \frac{k}{2b} \cdot \frac{e^{\frac{X_2(x)X}{b}}}{1 - e^{\frac{X}{b}}} \int_x^{x+X} B_X(\xi) d\xi, \quad (75)
 \end{aligned}$$

где x_1 и x_2 — некоторые функции координаты x , принимающие значения на интервале $[0, 1]$.

Функция $B(x)$ периодическая, следовательно:

$$\int_{x-X}^x B(\xi) d\xi = \int_0^X B(\xi) d\xi, \quad \int_x^{x+X} B(\xi) d\xi = \int_0^X B(\xi) d\xi. \quad (76)$$

С учетом этих соотношений формулу для определения $S(x)$ запишем в виде:

$$S(x) = \frac{k}{2b} \cdot \frac{e^{-\frac{X_1(x)X}{b}} + e^{\frac{X_2(x)X}{b}}}{1 - e^{\frac{X}{b}}} \int_0^X B_X(\xi) d\xi. \quad (77)$$

Обозначим через X_0 некоторый произвольно выбранный период и определим, во что превратится функция $S(x)$ при неограниченном уменьшении периода X , равного

$$X = \mu X_0, \tag{78}$$

то есть при $\mu \rightarrow 0$.

Предел, к которому стремится последовательность функций $S(x)$ при устремлении μ к нулю, обозначим через S_0 . Имеем:

$$\begin{aligned} S_0 &= \lim_{\mu \rightarrow 0} S(x) = \\ &= \lim_{\mu \rightarrow 0} \frac{k}{2b} \cdot \frac{e^{\frac{X_1(x)\mu X_0}{b}} + e^{\frac{X_2(x)\mu X_0}{b}}}{1 - e^{\frac{\mu X_0}{b}}} \cdot \mu \int_0^{X_0} B_{X_0}(\xi) d\xi = \\ &= \left[\frac{k}{X_0} \int_0^{X_0} B(\xi) d\xi \right] \lim_{\mu \rightarrow 0} \frac{X_0}{2b} \cdot \frac{e^{\frac{X_1(x)\mu X_0}{b}} + e^{\frac{X_2(x)\mu X_0}{b}}}{1 - e^{\frac{\mu X_0}{b}}} \mu. \end{aligned} \tag{79}$$

Пользуясь правилом Лопиталья, находим:

$$\begin{aligned} \lim_{\mu \rightarrow 0} \frac{X_0}{2b} \cdot \frac{e^{\frac{X_1(x)\mu X_0}{b}} + e^{\frac{X_2(x)\mu X_0}{b}}}{1 - e^{\frac{\mu X_0}{b}}} \mu &= \\ &= \lim_{\mu \rightarrow 0} \frac{X_0}{2b} \cdot \frac{2\mu}{1 - e^{\frac{\mu X_0}{b}}} \mu = 1. \end{aligned} \tag{80}$$

Таким образом:

$$S_0 = \frac{k}{X_0} \int_0^{X_0} B(\xi) d\xi, \tag{81}$$

что с учетом (67) приводит к формуле (53). Следовательно, модель инерции и иррадиации зрения (22) удовлетворяет также и пространственной формулировке закона Талбота.

5. Обобщение закона Талбота

Руководствуясь сформулированной в п. 3 методикой исследования, поставим перед собой теперь следующую задачу: построить такую модель инерции и иррадиации зрения, которая была бы в логическом отношении равносильна закону Талбота. Эта задача сводится к отысканию такого оператора преобразования функции $B(x, y, t)$ в функцию $S(x, y, t)$, чтобы, во-первых, из закона Талбота вытекала справедливость этого оператора, и наоборот, из справедливости такого оператора вытекала справедливость закона Талбота. Ясно, что в результате решения этой задачи мы должны прийти либо к оператору (22), либо к некоторому более общему оператору, включающему оператор (22) как частный случай.

Приступая к решению этой задачи, заметим, что закон Талбота свидетельствует о том, что некоторые

существенно различные входные сигналы преобразуются органом зрения в одинаковые выходные сигналы. Зададимся вопросом: возможно ли построить какие-либо другие, не предусмотренные законом Талбота, множества входных сигналов, преобразуемых глазом в одинаковые выходные сигналы? Из последующего изложения станет ясно, что в той формулировке, которая была дана выше, закон Талбота не исчерпывает всех возможностей. Наша ближайшая задача будет состоять в таком обобщении закона Талбота, чтобы с его помощью можно было для двух любых входных сигналов органа зрения установить, одинаковы или различны будут соответствующие им выходные сигналы.

Остановимся сначала на временной формулировке закона Талбота. Рассмотрим более подробно, что представляют собой входные сигналы органа зрения $B(t)$. Функции $B(t)$ удобно считать заданными на всей оси времени $(-\infty, \infty)$, поскольку инерционные процессы зрения весьма кратковременны и протекают в течение долей секунды, что несоизмеримо со сроком существования глаза. На любом конечном отрезке времени $[t_1, t_2]$ функции $B(t)$ локально суммируемы [15], поскольку интеграл

$$\int_{t_1}^{t_2} B(t) dt \tag{82}$$

пропорционален энергии излучения и, следовательно, имеет конечную величину.

Однако, кроме всевозможных локально суммируемых функций, глаз может воспринимать также некоторые иные функции. Одной из таких функций может служить функция, фигурирующая в формулировке временного закона Талбота. Она получается из периодической функции $B_T(t) = B_T(t-nT)$ при неограниченном уменьшении периода T колебаний яркости. Таким образом, входными функциями глаза могут быть также некоторые предельные функции, определяемые бесконечной последовательностью локально суммируемых функций. Такого рода входные функции для нас весьма важны, поскольку в терминах именно этих функций формулируется закон Талбота.

Приступим теперь к обобщению формулировки закона Талбота. Сформируем на одном поле сравнения зрительную картину, описываемую некоторой произвольным образом выбранной локально суммируемой функцией $B(t)$, порождающую зрительное ощущение $S(t)$. На другом поле сравнения будем поочередно формировать ряд зрительных картин, каждая из которых описывается функцией из бесконечной последовательности $\{B_1(t), B_2(t), \dots, B_N(t), \dots\}$ локально суммируемых функций. Пусть последовательность этих зрительных картин порождает последовательность зрительных ощущений $\{S_1(t), S_2(t), \dots,$

$S_N(t), \dots$. Опыт показывает, что можно подобрать такие последовательности зрительных картин, чтобы последовательности порождаемых ими зрительных ощущений сходились к зрительному ощущению $S(t)$ второго поля, т. е. чтобы $\lim_{N \rightarrow 0} S_N(t) = S(t)$.

Как определить, какая последовательность зрительных картин дает в пределе зрительное ощущение $S(t)$, а какая нет? Исчерпывающий ответ на этот вопрос дает закон Талбота в обобщенной формулировке: формула

$$\lim_{N \rightarrow 0} S_N(t) = S(t) \quad (83)$$

справедлива в том, и только том случае, если для любого конечного интервала времени $[t_1, t_2]$ будет иметь место равенство:

$$\lim_{N \rightarrow 0} \int_{t_1}^{t_2} B_N(t) dt = \int_{t_1}^{t_2} B(t) dt. \quad (84)$$

В выражении (84) сходимость имеет следующий смысл: каковы бы ни были t_1 и t_2 (но $|t_2 - t_1| \leq T_0$, где T_0 — некоторое фиксированное число) для любого N найдется число $\epsilon_N > 0$, такое что:

$$\left| \int_{t_1}^{t_2} (B_N(t) - B(t)) dt \right| \leq \epsilon_N, \quad (85)$$

причем при $N \rightarrow \infty$, $\epsilon_N \rightarrow 0$.

Таким образом, в обобщенной формулировке временной закон Талбота означает следующее. Пусть имеются две различные зрительные картины, одна из которых достаточно быстро изменяется во времени по яркости. Если зрительные картины на любом отрезке времени в среднем совпадают по яркости, то порождаемые ими зрительные ощущения будут одинаковыми. При невыполнении этого условия зрительные ощущения будут различными. Заметим, что прежняя формулировка временного закона Талбота является частным случаем новой, поскольку для периодических функций условие (84) переходит в условие (51).

Закон Талбота можно обобщить также и на случай неоднородных и нестационарных зрительных картин, когда яркость является не только функцией времени t , но и функцией координат поля зрения x, y . Формулировка закона Талбота теперь будет такой.

Если, и только если:

$$\begin{aligned} \lim_{N \rightarrow 0} \int_{x_1}^{x_2} \int_{y_1}^{y_2} \int_{t_1}^{t_2} B_N(x, y, t) dx dy dt = \\ = \int_{x_1}^{x_2} \int_{y_1}^{y_2} \int_{t_1}^{t_2} B(x, y, t) dx dy dt, \end{aligned} \quad (86)$$

то

$$\lim_{N \rightarrow 0} S_N(x, y, t) = S(x, y, t). \quad (87)$$

В обобщенной формулировке пространственно-временной закон Талбота означает, по существу, следующее. Пусть имеются две зрительные картины, одна из которых достаточно быстро изменяется по яркости во времени и в поле зрения. Порождаемые этими картинами зрительные ощущения будут одинаковыми в том, и только в том случае, если на любом участке поля зрения и на любом отрезке времени зрительные картины имеют одинаковую среднюю яркость. До сих пор мы оперировали лишь с одной компонентой зрительного ощущения — фотометрической яркостью $B(x, y, t)$, отвлекаясь от факта трехкомпонентности зрения. С учетом трехкомпонентности зрения обобщенный закон Талбота может быть сформулирован в следующем виде.

Зрительные ощущения совпадают в том, и только в том случае, если совпадают средние значения координат цвета соответствующих зрительных картин в произвольном пространственно-временном параллелепипеде $[x_1, x_2; y_1, y_2; t_1, t_2]$. Иными словами, если, и только если:

$$\begin{aligned} \lim_{N \rightarrow 0} \int_{x_1}^{x_2} \int_{y_1}^{y_2} \int_{t_1}^{t_2} B_{iN}(x, y, t) dx dy dt = \\ = \int_{x_1}^{x_2} \int_{y_1}^{y_2} \int_{t_1}^{t_2} B_i(x, y, t) dx dy dt, \end{aligned} \quad (88)$$

$$\lim_{N \rightarrow 0} S_{iN}(x, y, t) = S_i(x, y, t), \quad (i=1, 2, 3). \quad (89)$$

В такой форме закон Талбота дает исчерпывающий ответ на поставленный выше вопрос; с помощью этого закона можно для любого входного сигнала органа зрения $b_\lambda(x, y, t)$ указать всевозможные последовательности входных сигналов $\{b_{\lambda N}(x, y, t)\}$, дающие в пределе одно и то же зрительное ощущение. Заметим, что вовсе не обязательно полагать, что на одном из полей сравнения должна быть сформирована некоторая фиксированная зрительная картина $b_\lambda(x, y, t)$. Можно рассматривать также и тот случай, когда на обоих полях сравнения формируются последовательности зрительных картин $\{b'_{\lambda N}(x, y, t)\}$ и $\{b''_{\lambda N}(x, y, t)\}$. В этом случае закон Талбота приобретает наиболее общий вид и может быть сформулирован таким образом.

Обобщенный закон Талбота.

Если, и только если:

$$\begin{aligned} \lim_{N \rightarrow 0} \int_{x_1}^{x_2} \int_{y_1}^{y_2} \int_{t_1}^{t_2} B'_{iN}(x, y, t) dx dy dt = \\ = \lim_{N \rightarrow 0} \int_{x_1}^{x_2} \int_{y_1}^{y_2} \int_{t_1}^{t_2} B''_{iN}(x, y, t) dx dy dt, \end{aligned} \quad (90)$$

то

$$\lim_{N \rightarrow \infty} S'_{iN}(x, y, t) = S''_{iN}(x, y, t), \quad (i=1, 2, 3). \quad (91)$$

Именно в этой формулировке закон Талбота будет нами использоваться в дальнейшем.

6. Равносильность модели инерции и иррадиации зрения и обобщенного закона Талбота

В предыдущем разделе мы сформулировали обобщенный закон Талбота. Теперь мы докажем равносильность обобщенного закона Талбота и математической модели инерции и иррадиации зрения, предложенной в п. 1 в виде формул (1) и (2). Располагая таким доказательством, мы будем вправе считать, что математическая модель инерции и иррадиации зрения достоверна в той же мере, в какой справедлив обобщенный закон Талбота. Вместе с тем имеются все основания полагать, что обобщенный закон Талбота находится в полном согласии с данными опыта (этот вопрос будет обсужден в следующем разделе). Чтобы не загромождать работу, доказательство равносильности обобщенного закона Талбота и математической модели инерции и иррадиации зрения выполним для частного случая нестационарных однородных зрительных картин при однокомпонентном зрении. В общем случае при нестационарных и неоднородных зрительных картинах и трехкомпонентном зрении доказательство совершенно аналогично, хотя значительно более громоздко.

В рассматриваемом случае обобщенный закон Талбота может быть сформулирован следующим образом.

Если, и только если:

$$\lim_{N \rightarrow \infty} \int_{t_1}^{t_2} (B'_N(t) - B''_N(t)) dt = 0, \quad (92a)$$

то

$$\lim_{N \rightarrow \infty} (S'_N(t) - S''_N(t)) = 0. \quad (92b)$$

Математическая модель инерции и иррадиации зрения в данном случае запишется:

$$S(t) = \Phi \left(\lim_{N \rightarrow \infty} \frac{k}{a} \int_{-\infty}^t B_N(\tau) e^{-\frac{t-\tau}{a}} d\tau \right). \quad (93)$$

Для доказательства равносильности утверждений (92) и (93) мы должны установить, что: 1) из справедливости утверждения (92) вытекает справедливость утверждения (93); 2) из справедливости утверждения (93) вытекает справедливость утверждения (92). Докажем сначала, что из справедливости утверждения (92) вытекает справедливость утверждения (93). Доказательство состоит в установлении взаимной однозначности оператора Φ . Докажем сперва, что при совпадении функций

$$S'(t) = \lim_{N \rightarrow \infty} S'_N(t), \quad S''(t) = \lim_{N \rightarrow \infty} S''_N(t) \quad (94)$$

совпадут также функции

$$\begin{aligned} V'(t) &= \lim_{N \rightarrow \infty} \frac{k}{a} \int_{-\infty}^t B'_N(\tau) e^{-\frac{t-\tau}{a}} d\tau, \\ V''(t) &= \lim_{N \rightarrow \infty} \frac{k}{a} \int_{-\infty}^t B''_N(\tau) e^{-\frac{t-\tau}{a}} d\tau. \end{aligned} \quad (95)$$

Как уже упоминалось выше, по-существу, формула (92a) означает следующее:

$$\left| \int_{t_1}^{t_2} (B'_N(\tau) - B''_N(\tau)) d\tau \right| \leq E_N, \quad (96)$$

где $t_2 - t_1 \leq T_0$ и $E_N \rightarrow 0$ при $N \rightarrow \infty$. Выберем произвольно некоторый фиксированный интервал времени так, чтобы $0 < T < T_0$. Тогда:

$$\begin{aligned} &\frac{k}{a} \int_{-\infty}^t (B'_N(\tau) - B''_N(\tau)) e^{-\frac{t-\tau}{a}} d\tau = \\ &= \frac{k}{a} e^{-\frac{t}{a}} \sum_{m=0}^{\infty} \int_{(m+1)T}^{mT} (B'_N(\tau) - B''_N(\tau)) e^{\frac{t-mT}{a}} d\tau. \end{aligned} \quad (97)$$

Согласно формуле Бонне [16, с. 325], с учетом (96) имеем:

$$\begin{aligned} &\left| \int_{(m+1)T}^{mT} (B'_N(\tau) - B''_N(\tau)) e^{\frac{t-\tau}{a}} d\tau \right| = \\ &= e^{-\frac{(m+1)T}{a}} \left| \int_{(m+1)T}^{\xi} (B'_N(\tau) - B''_N(\tau)) d\tau \right| + \\ &+ e^{-\frac{t-mT}{a}} \left| \int_{\xi}^{mT} (B'_N(\tau) - B''_N(\tau)) d\tau \right| \leq \\ &\leq e^{-\frac{t-mT}{a}} \left(\left| \int_{(m+1)T}^{\xi} (B'_N(\tau) - B''_N(\tau)) d\tau \right| + \right. \\ &\left. + \left| \int_{\xi}^{mT} (B'_N(\tau) - B''_N(\tau)) d\tau \right| \right) \leq 2E_N e^{-\frac{t-mT}{a}}, \end{aligned} \quad (98)$$

где $t - (m+1)T \leq \xi \leq t - mT$.

Следовательно:

$$\begin{aligned} &\left| \frac{k}{a} \int_{-\infty}^t (B'_N(\tau) - B''_N(\tau)) e^{-\frac{t-\tau}{a}} d\tau \right| \leq \\ &\leq 2E_N \frac{k}{a} \sum_{m=0}^{\infty} e^{-\frac{mT}{a}} = 2 \frac{k}{a} \frac{E_N}{1 - e^{-\frac{T}{a}}}. \end{aligned} \quad (99)$$

При $N \rightarrow \infty$ величина

$$2 \frac{k}{a} \frac{E_N}{1 - e^{-\frac{1}{a}}} \rightarrow 0 \quad (100)$$

при любом t . Это значит, что

$$\lim_{N \rightarrow \infty} \frac{k}{a} \int_{-\infty}^t (B'_N(\tau) - B''_N(\tau)) e^{-\frac{t-\tau}{a}} d\tau = V'(t) - V''(t) = 0 \quad (101)$$

в смысле равномерной сходимости относительно t . Итак, мы доказали, что при совпадении функций $S'(t)$ и $S''(t)$ функции $V'(t)$ и $V''(t)$ также совпадают.

Докажем теперь обратное, а именно, что при совпадении функций $V'(t)$ и $V''(t)$ совпадают также функции $S'(t)$ и $S''(t)$. Действительно, пусть

$$V'(t) = V''(t). \quad (102)$$

В этом случае при любом t имеет место равенство:

$$\lim_{N \rightarrow \infty} \frac{k}{a} \int_{-\infty}^t (B'_N(\tau) - B''_N(\tau)) e^{-\frac{t-\tau}{a}} d\tau = 0 \quad (103)$$

в смысле равномерной сходимости относительно t . Равенство (103) означает следующее:

$$\left| \frac{k}{a} \int_{-\infty}^t (B'_N(\tau) - B''_N(\tau)) e^{-\frac{t-\tau}{a}} d\tau \right| \leq \eta_N, \quad (104)$$

где $\eta_N \rightarrow 0$ при $N \rightarrow \infty$. Рассмотрим уравнение:

$$a \frac{dg(t)}{dt} + g(t) = K(B'_N(t) - B''_N(t)). \quad (105)$$

На всей оси t это уравнение имеет единственное решение (см. п. 3):

$$g(t) = \frac{k}{a} \int_{-\infty}^t (B'_N(\tau) - B''_N(\tau)) e^{-\frac{t-\tau}{a}} d\tau. \quad (106)$$

Подставляя (106) в (105), получим тождество:

$$\begin{aligned} B'_N(t) - B''_N(t) &= \\ &= \frac{a}{k} \frac{d}{dt} \left(\frac{k}{a} \int_{-\infty}^t (B'_N(\tau) - B''_N(\tau)) e^{-\frac{t-\tau}{a}} d\tau \right) + \\ &+ \frac{1}{k} \left(\frac{k}{a} \int_{-\infty}^t (B'_N(\tau) - B''_N(\tau)) e^{-\frac{t-\tau}{a}} d\tau \right). \end{aligned} \quad (107)$$

Интегрируя (107) в пределах от t_1 до t_2 , имеем:

$$\begin{aligned} &\int_{t_1}^{t_2} (B'_N(\tau) - B''_N(\tau)) d\tau = \\ &= \frac{a}{k} \left(\frac{k}{a} \int_{-\infty}^{t_2} (B'_N(\tau) - B''_N(\tau)) e^{-\frac{t_2-\tau}{a}} d\tau - \right. \\ &\left. - \frac{a}{k} \left(\frac{k}{a} \int_{-\infty}^{t_1} (B'_N(\tau) - B''_N(\tau)) e^{-\frac{t_1-\tau}{a}} d\tau \right) + \right. \\ &\left. + \frac{1}{k} \int_{t_1}^{t_2} \left(\frac{k}{a} \int_{-\infty}^t (B'_N(\tau) - B''_N(\tau)) e^{-\frac{t-\tau}{a}} d\tau \right) dt. \right. \end{aligned} \quad (108)$$

Используя условие (104), получим:

$$\begin{aligned} &\left| \int_{t_1}^{t_2} (B'_N(\tau) - B''_N(\tau)) d\tau \right| \leq \frac{a}{k} \eta_N + \frac{a}{k} \eta_N + \frac{1}{k} \int_{t_1}^{t_2} \eta_N dt = \\ &= 2 \frac{a}{k} \eta_N + \frac{1}{k} (t_2 - t_1) \eta_N \leq 2 \frac{a}{k} \eta_N + \frac{T_0}{k} \eta_N. \end{aligned} \quad (109)$$

Полагая $\frac{1}{k} (2a + T_0) \eta_N = E_N$, приходим к условию (96). Это значит, что:

$$\lim_{N \rightarrow \infty} \int_{t_1}^{t_2} (B'_N(\tau) - B''_N(\tau)) d\tau = 0, \quad (110)$$

откуда в силу (92) следует равенство функций $S'(t)$ и $S''(t)$. Итак, мы доказали, что при совпадении функций $V'(t)$ и $V''(t)$ функции $S'(t)$ и $S''(t)$ также совпадают. Следовательно, оператор Φ взаимно однозначный. Это значит, что из справедливости утверждения (92) вытекает справедливость утверждения (93).

Докажем теперь, что из справедливости утверждения (93) вытекает справедливость утверждения (92). Положим вначале, что выполняется условие (92а), и докажем, что при этом выполняется также условие (92б), т.е., что совпадают функции $S'(t)$ и $S''(t)$. В силу взаимной однозначности оператора Φ нам достаточно доказать совпадение функций $V'(t)$ и $V''(t)$. Следовательно, требуется из условия (92а) вывести условие (103). Однако такой вывод был только что выполнен.

Теперь положим, что функции $S'(t)$ и $S''(t)$ совпадают, т.е. что выполняется условие (92б), и докажем, что при этом выполняется также условие (92а). В силу взаимной однозначности оператора Φ в этом случае совпадают также функции $V'(t)$ и $V''(t)$, следовательно, выполняется условие (103). Однако только что было доказано, что из условия (103) следует зависимость (92а). Итак, мы доказали, что из справедливости утверждения (93) вытекает справедли-

вость утверждения (92). Этим завершается доказательство равносильности обобщенного закона Талбота и математической модели инерции и иррадиации зрения для случая нестационарных однородных зрительных картин при однокомпонентном зрении.

Доказательство этого же положения в общем случае при нестационарных и неоднородных зрительных картинах и трехкомпонентном зрении совершенно аналогично и не требует привлечения каких-либо новых идей. Чтобы не загромождать работу, мы его здесь не приводим. В заключение приведем строгую формулировку общей математической модели инерции и иррадиации зрения, равносильной обобщенному закону Талбота. Модель имеет следующий вид:

$$S_i(x, y, t) = \Phi_i \left(\lim_{N \rightarrow \infty} \frac{k_i}{4\pi b_i^2} \int_{-\infty}^t \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} B_{iN}(\xi, \eta, \tau) \frac{e^{-\frac{t-\tau}{a_i}}}{t-\tau} \times \right. \\ \left. \times e^{-\frac{(\xi-x)^2 + (\eta-y)^2}{4b_i^2 \times (t-\tau)/a_i}} d\eta d\xi d\tau \right), i = (1, 2, 3). \quad (111)$$

Важно отметить, что гипотетическая модель инерции и иррадиации зрения в виде формулы (22), полученная на базе тепловой модели и использованная нами в качестве инструмента исследования, является частным случаем модели (111), выведенной из обобщенного закона Талбота.

Выводы

Приведем далее экспериментальную проверку обобщенного закона Талбота. В предыдущем разделе мы построили математическую модель инерции и иррадиации зрения в виде формулы (111) и доказали ее равносильность закону Талбота. Эта модель верна в той же степени, в которой верен закон Талбота. Теперь остановимся на вопросе, в какой мере закон Талбота выполняется на опыте. Что касается временного закона Талбота, то он проверялся экспериментально многими авторами [17–20]. В широком диапазоне изменения яркости, соответствующем практически произвольным условиям видения, не обнаружено каких-либо отклонений от закона Талбота. Некоторые незначительные отклонения обнаружены в крайних, нерабочих режимах функционирования глаза: при очень низком уровне яркости, близком к абсолютному порогу чувствительности зрения, и при очень высоком уровне яркости, соответствующем уровню невыносимости глаза. Если не считаться с этими отклонениями, которые, очевидно, связаны с неточностями работы глаза в предельных режимах, то можно констатировать идеальное согласование опытных данных с законом Талбота.

Опыты с цветными мельканиями показали, что и в условиях трехкомпонентного зрения закон Талбота не нарушается [21]. Пространственный закон Талбота в случае ахроматического зрения выполняется безукоризненно, в случае цветного зрения также известны работы, подтверждающие справедливость этого закона, хотя имеются работы, ставящие под сомнение справедливость этого закона [21, 22]. Что же касается формулировки закона Талбота в обобщенном виде, то литературные данные по его экспериментальной проверке отсутствуют. Однако и нет никаких данных, которые могли бы поставить под сомнение его справедливость.

Нами были поставлены специальные опыты по проверке закона Талбота в его обобщенной формулировке. Существо опытов сводилось к следующему. До сих пор эксперименты по проверке закона Талбота ставились таким образом, что после слияния мельканий получалось равномерное и немигающее зрительное ощущение. Однако, исходя из обобщенной формулировки закона Талбота, можно прийти к выводу о возможности получения с помощью различных зрительных картин одинаковых нестационарных ощущений. В опытах были взяты примеры именно таких зрительных картин.

В одном из опытов наблюдателю предъявлялись для сравнения две ахроматические картины $B_1(t)$ и $B_2(t)$, яркость которых изменялась во времени так, как показано на диаграммах рис. 6а и 6б. На диаграммах B_6 обозначают уровни яркости, соответствующие белому и черному цвету с коэффициентами отражения 83 и 2%. Зрительная картина, соответствующая диаграмме 6а, формировалась путем чередования во времени двух серых полей, имеющих яркости B' и B'' , соответствующие коэффициентам отражения 63 и 22%. Длительность T предъявления полей одинакова. Диаграмме 6б соответствует зрительная картина, получаемая с помощью значительно более быстрых мельканий белых и черных полей. Соотношение в длительности фаз этих быстрых мельканий выбиралось с таким расчетом, чтобы средняя яркость быстрых мельканий равнялась яркости медленных мельканий на том же интервале времени.

Согласно обобщенному закону Талбота, обе зрительные картины должны выглядеть одинаковыми при условии слияния быстрых мельканий.

Опыты выполнялись с помощью диска Максвелла. Схема кружка, использовавшегося в опыте, изображена на рис. 7. На кружке имеются два соприкасающихся кольца. На внутреннем кольце сформирована зрительная картина, соответствующая диаграмме 6а, на внешнем — зрительная картина, соответствующая диаграмме 6б. Частота смены яркости быстрых мельканий принята в 16 раз более высокой, чем медленных.

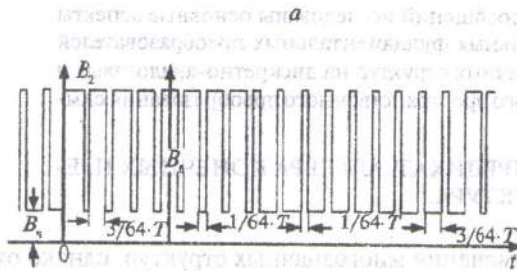
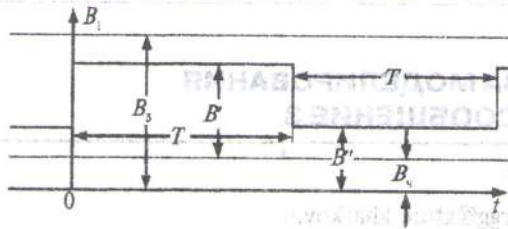


Рис. 6. Диаграммы проверки обобщенного закона Талбота

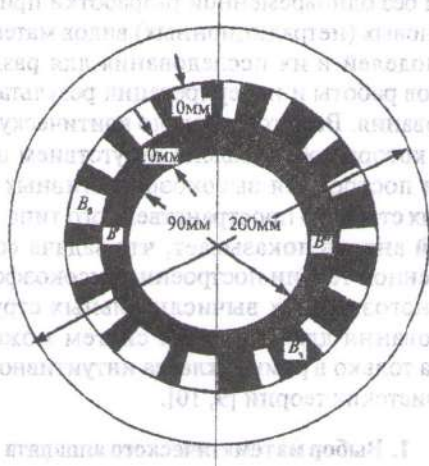


Рис. 7. Диск Максвелла

При постепенном раскручивании диска-вертушки вначале на внешнем кольце заметны не только медленные, но также и быстрые мелькания яркости. При этом заметно различие зрительных картин, наблюдаемых на соседних кольцах. Однако, как только быстрые мелькания яркости на внешнем кольце перестают замечаться (это происходит при скорости вращения диска около 3 об/сек), обе зрительные картины становятся неотличимыми, в полном соответствии с обобщенным законом Талбота. При этом наблюдатель видит на обеих дорож-

ках синфазные колебания светлоты. Дальнейшее повышение скорости вращения диска приводит к изменению характера колебаний светлоты на обоих полях, однако между соседними зрительными ощущениями все время сохраняется тождество. Так продолжается вплоть до полного слияния медленных мельканий при 40 + 50 об/сек. Во время проведения опытов освещенность диска вертушки изменялась в широких пределах (от 10 до 1000 лк).

Подобного рода опыты выполнялись также при изменении яркости медленных мельканий во времени в виде синусоиды, пил и др. Производились также опыты с цветными мельканиями. Во всех этих опытах не было обнаружено каких-либо отклонений от обобщенного закона Талбота.

Список литературы. 1. Бондаренко М.Ф., Шабанов-Кушнаренико Ю.П. Модели цветового зрения. // Бионика интеллекта: Научн.-техн. журнал. — 2006. — № 1 (64). — С. 27—41. 2. Talbot H.F. Experiments on light. // Phil. Mag., 1834, № 5. 3. Луизов А.В. Инерция зрения. — М.: Оборонгиз, 1961. 4. Allard M.E. Memoire sur l'intensite et la portee des phares, Paris, 1876. 5. Кравков С.В. Глаз и его работа. — М.-Л.: Изд. АН СССР, 1950. 6. Кирпичев М.В., Михеев М.А., Моделирование тепловых устройств. — М.-Л.: Изд. АН СССР, 1936. 7. Михеев М.А. Основы теплопередачи. — М.-Л.: Госэнергоиздат, 1949. 8. Справочник машиностроителя. — Т. 1. — М.: Машгиз, 1951. 9. Тихонов А.Н., Самарский Л.Л. Уравнения математической физики. — М.: Гостехтеориздат. Изд. 2-е, 1933. 10. Градштейн И.С., Рыжик И.М. Таблицы интегралов, сумм, рядов и произведений. Изд. 4-е. — Физматгиз, 1962. 11. Ватсон Д.Н. Теория бесселевых функций. Пер. с англ. — М.: ИЛ, 1949. 12. Бронштейн И.Н., Семендяев К.А. Справочник по математике. Изд. 10-е. — М.: Наука, 1965. 13. Девит Г.Б. Таблицы интегралов и другие математические формулы. Пер. с англ. Изд. 2-е. — М.: Наука, 1966. 14. Фихтенгольц Г.М. Курс дифференциального и интегрального исчисления. Изд. 6-е. — Т. 2. — М.: Наука, 1966. 15. Натансон И.П. Конструктивная теория функций. — М.-Л.: Гостехтеориздат, 1949. 16. Ильин В.А., Позняк Э.Г. Основы математического анализа. — М.: Наука, 1967. 17. Hyde E., Talbot's law as applied to rotating sectored disk, Bul. Bureau of Stand., 1906, v. 2. 18. Piéron H., Excitation lumineuse intermittente et excitation alterne caracteristiques et lois, L'année psychologique, 1927, v. 28. 19. Exner S. Studien auf dem Grenzgebiete lokalisierten Sehens, "Pflugers Archiv", 1898, Bd. 73. 20. Marbe K. Tatsachen und Theorien des Talbot'schen Gesetzes, «Pflugers Archiv», 1903, Bd. 97. 21. Lempicka W., Paumliche Farbmischung auf der Netzhaut, «Zeitschr. F. Sinnesphysiol», 1919, №50. 22. Теплов Б.М., Яковлева С.П. О законах пространственного и временного смешения цветов. // Сб. «Зрительные ощущения и восприятия». — Т. 2. — М.: Союзгиз, 1935.

Поступила в редколлегию 02.02.07

УДК 519.62

**АЛГЕБРО-ЛОГИЧЕСКИЕ СРЕДСТВА МОДЕЛИРОВАНИЯ
ЕСТЕСТВЕННОГО ЯЗЫКА. СООБЩЕНИЕ 3**

Г. Г. Четвериков

ХНУРЭ, г. Харьков, Украина, chetvergg@kture.kharkov.ua

На базе теоретических предпосылок двух предыдущих сообщений исследованы основные аспекты математического моделирования АКП-структур универсальных фундаментальных преобразователей пространственного типа на основе декомпозиции многозначных структур на дискретно-аналоговые и цифровые субблоки, в частности, в рамках их промежуточного пространственного преобразования символической информации.

ЕСТЕСТВЕННЫЙ ЯЗЫК, ЛОГИКА, КОНЕЧНЫЙ ПРЕДИКАТ, АЛГЕБРА КОНЕЧНЫХ ПРЕДИКАТОВ, МНОГОЗНАЧНАЯ СТРУКТУРА, АКП-СТРУКТУРА.

Введение

Развитие и совершенствование средств вычислительной техники лежит в основе процесса автоматизации умственной деятельности, что явилось отправной точкой возникновения понятия «искусственный интеллект». Однако успехи в области интеллектуализации вычислительной техники незначительны, особенно, если сравнивать достигнутое с ожидаемыми результатами и прогнозами. Ориентация на достижение качественно новых технологий обработки информации проявляется в попытках реализовать на фон-Неймановских компьютерах системы искусственного интеллекта (ИИ). Следовательно, новые требования к технологии обработки информации обусловлены необходимостью решать проблемы, которые плохо формализуются, и наличием пользователя, не являющегося профессиональным программистом. Таким образом, приходим к осознанию одного из вариантов на пути создания систем ИИ — это путь анализа, моделирования и синтеза естественно-языкового интеллектуального интерфейса средствами многозначной логической системы, в частности, алгебры конечных предикатов, а также теории многозначных структур и кодирования.

С момента возникновения вычислительной техники ведутся исследования и осуществляется реализация на уровне инженерных решений многозначных структур и кодирования в связи с высокой информационной насыщенностью их сигналов. Многозначными называют структуры средств обработки данных, построенные на базе многозначных логических элементов и модулей с соответствующими связями. К таким структурам относятся все объекты, которые описываются конечным структурным алфавитом: элементы, модули, структуры и системы вычислительной, измерительной, управляющей техники и технические средства обработки информации на естественном языке.

В настоящее время существует большое количество разрозненных подходов и методов построения

и применения многозначных структур, однако отсутствуют их систематизация и классификация (в каком-либо виде упорядоченная система средств реализации). В то же время оптимальное проектирование и техническая реализация вычислительных устройств на базе многозначных структур невозможны без одновременной разработки принципиально новых (нетрадиционных) видов математических моделей и их исследования для различных режимов работы и интерпретации результатов моделирования. Все это и создало критическую ситуацию, которая обусловлена отсутствием цельной теории построения высокоэффективных многозначных структур пространственного типа. Проведенный анализ показывает, что задача создания обобщенной теории построения высокоэффективных многозначных вычислительных структур и кодирования для языковых систем может быть решена только в рамках класса интуитивно-конструктивистских теорий [9, 10].

1. Выбор математического аппарата

Наличие алгебры конечных предикатов (АКП) дает интересную возможность осуществления перехода от алгоритмического описания информационных процессов к их описанию в виде уравнений на языке данной алгебры, а уравнения задают отношения между его переменными [11, 12]. Все переменные в уравнении равноправны, и любые из них могут выступать в роли как независимых, так и в роли зависимых. Наличие уравнений и их преимущество перед алгоритмами состоит в том, что появляется возможность рассчитать реакцию системы и в случае неполной определенности исходной информации, в то время как не полностью разработанный алгоритм является неработоспособным. Показано, что изменение знаний об объекте, представленное системой уравнений АКП и реализованное в виде соответствующей структуры, всегда готово к использованию, а алгоритм часто требует существенного его изменения. Отметим, что с помощью формул АКП строим АКП-структуры, которые реализуют

соответствующие конечные предикаты. Данный подход аналогичен процессу построения комбинационных схем по формулам алгебры логики. В зависимости от уровня функционально-структурной реализации имеем АКП-структуры первого, второго и третьего рода [13].

В качестве математического аппарата исследования используется алгебра конечных предикатов [12]. Последнюю трактуем как алгебру, носителем которой является множество M всех предикатов U^m , где U — непустое множество каких-либо переменных, которое называют универсумом, т.е. $U = \{x_1, x_2, \dots, x_m\}$. Здесь x_1, x_2, \dots, x_m — какие-либо места предметов. Поэтому их иначе называют предметными переменными. Заметим, если предмет a находится на месте x_i $i = (\overline{1, m})$, то говорят, что переменная x_i принимает значение a и при этом имеет такую запись $x_i = a$. Если $a_1, a_2, \dots, a_m \in U$ и $x_1 = a_1, x_2 = a_2, \dots, x_m = a_m$, то справедлива запись $(a_1, a_2, \dots, a_m) \in U^m$, и говорят, что предметный вектор (набор) принадлежит предметному пространству U^m . Число m называют размерностью пространства U^m . Любое подмножество T пространства U^m называют m -местным отношением, заданным на U^m . Для формульной (аналитической) записи таких отношений используют явный способ задания конечного алфавитного оператора (КАО), который служит отправным моментом для аппаратурного метода решения уравнений алгебры конечных предикатов [7, с. 123].

Пусть T — множество всех отношений на U^m , Q — множество всех предикатов на U^m . Отношение T и предикат Q называют соответствующими друг другу, если для любых x_1, x_2, \dots, x_m имеем:

$$Q(x_1, x_2, \dots, x_m) = \begin{cases} 0, & \text{если } x_1, x_2, \dots, x_m \notin T \\ 1, & \text{если } x_1, x_2, \dots, x_m \in T. \end{cases} \quad (1)$$

В соответствии с (1) возможен переход от произвольного отношения T к соответствующему ему предикату Q . Предикат Q , который находят по выражению (1), называют характеристической функцией отношения T .

Предикатом распознавания предмета $a \in U$ переменной x_i $i = (\overline{1, m})$ называют условия вида:

$$a(x_i) = x_i^a = \begin{cases} 0, & \text{если } a \neq x_i \\ 1, & \text{если } a = x_i. \end{cases} \quad (2)$$

Предикат $a(x_i)$ следует рассматривать как предикат $a(x_1, x_2, \dots, x_i, \dots, x_m)$ из $P \subseteq Q$, все аргументы которого, кроме x_i , несущественны. Выражение в виде $a(x_i)$, где $i = (\overline{1, m})$, $a \in U$, заменим на x_i^a (здесь a называют показателем для переменной x_i). Таким образом, алгеброй конечных предикатов над

M называют множество T с базисными элементами x_i^a ($i = \overline{1, m}, a \in U$) и базисными операциями: дизъюнкцией, конъюнкцией и отрицанием. Исключение из базиса данной алгебры операции отрицания позволяет получить так называемую дизъюнктивно-конъюнктивную алгебру предикатов (ДКАП). Доказана ее полнота [14, 15]. Таким образом, данная алгебра рассматривается как инструмент исследования, а не как его предмет.

2. Формализация концепции унификации многозначных структур пространственного типа

На базе проведенных архитектурных исследований, изложенных в работах [16–18], а также исходя из необходимости структуризации задач при разработке интуитивно-конструктивистской теории построения многозначных структур пространственного типа, для языковых систем предложена концепция унификации (приведение к единообразию и неразрывному взаимодействию) двузначно-многозначных аналого-цифровых средств обработки соответствующих (символьных) данных, заданных на естественном языке [18]. Настоящий подход базируется на единой методологической и целевой основе путем применения предложенных методов теории интеллекта [12] для математического описания и соответствующей формализации концепции унификации входных/выходных данных [13] и их промежуточных преобразований [19] соответствующей АКП-структурой третьего рода (рис. 1) [20–21].



Рис. 1. Двухвходовой многозначный УФЭ (АКП-структура третьего рода)

АКП-структура третьего рода на основе двухвходового универсального многозначного функционального преобразователя [20] включает такие компоненты: элемент распознавания k -значной переменной, который образован параллельным аналого-цифровым преобразователем вместе с пространственным дешифратором, матричный селектор и коммутатор, блок управления, параллельный цифро-аналоговый преобразователь (ключевой коммутатор). Описав логику работы этих компонент

соответствующими уравнениями алгебры конечных предикатов, получим их математические модели. Использование концепции унификации и данной алгебры обеспечит граничный параллелизм и однородность структуры в целом. Получение аналитических соотношений входных/выходных переменных составляющих компонент позволит формализовать и автоматизировать процедуры синтеза многозначных структур пространственного типа [13–18].

Элемент распознавания k -значной переменной реализует предикат распознавания (2) — базисного элемента алгебры конечных предикатов. АКП задается на множестве M всех n -местных k -значных предикатов, то есть функций вида $y = f(x_1, x_2, \dots, x_n)$, где x_1, x_2, \dots, x_n — буквенные переменные, заданные на алфавите букв $A = \{a_1, a_2, \dots, a_k\}$, $y \in \{0, 1\}$ — логическая переменная, т. е.

$$x_i^a = \begin{cases} 0, & \text{если } a_j \neq x_i \\ 1, & \text{если } a_j = x_i \end{cases}, \quad (i = \overline{1, n}, j = \overline{1, k}). \quad (3)$$

Поскольку в элементе распознавания k уровней ($0 + (k - 1)$), то и выходных сигналов распознавания для пространственного набора элемента распознавания будет k : a_1, a_2, \dots, a_{k-1} . Таким образом, уже на входе АКП-структуры обеспечивается распараллеливание процесса на уровне граничного быстрогодействия соответствующих преобразований.

Логику работы дешифраторов, в частности, дешифратора 2 (ДШ2), описывает система уравнений АКП следующего вида:

$$\begin{cases} b_{1,2,3}^0 = \bar{y}_1; \\ b_{1,2,3}^1 = y_1 \& \bar{y}_2; \\ \dots \\ b_{1,2,3}^{k-1} = y_{k-1}. \end{cases} \quad (4)$$

Логику работы матричного селектора (МС3) описывает система уравнений АКП следующего вида:

$$\begin{cases} g_{00} = b_1^0 \& b_2^0; g_{01} = b_1^0 \& \bar{b}_2^0; g_{0(k-1)} = b_1^0 \& b_2^{k-1}; \\ g_{10} = b_1^1 \& b_2^0; g_{11} = b_1^1 \& \bar{b}_2^0; g_{1(k-1)} = b_1^1 \& b_2^{k-1}; \\ \dots \\ g_{(k-1)0} = b_1^{(k-1)} \& b_2^0; g_{(k-1)1} = b_1^{(k-1)} \& \bar{b}_2^0; g_{(k-1)(k-1)} = b_1^{(k-1)} \& b_2^{k-1}, \end{cases} \quad (5)$$

где g_{ij} — выходные сигналы матричного селектора, которые принимают значения из множества $E_2 \in \{0, 1\}$, $i, j \in \{0, 1, \dots, k-1\}$.

Логика работы пространственного коммутатора в АКП-структуре первого рода описывается следующей системой АКП:

$$\begin{cases} g_{00}^0 \vee g_{01}^1 \vee \dots \vee g_{0(k-1)}^{(k-1)} = q_0, \\ g_{10}^0 \vee g_{11}^1 \vee \dots \vee g_{1(k-1)}^{(k-1)} = q_1, \\ \dots \\ g_{(k-1)0}^0 \vee g_{(k-1)1}^1 \vee \dots \vee g_{(k-1)(k-1)}^{(k-1)} = q_{(k-1)}. \end{cases} \quad (6)$$

Соответственно, входные сигналы коммутатора образуются за счет проводного ИЛИ для сигналов с одинаковыми индексами, но с разных уровней коммутации:

$$\begin{cases} z_0 = r_0 \vee s_0 \vee q_0, \\ z_1 = r_1 \vee s_1 \vee q_1, \\ \dots \\ z_{(k-1)} = r_{(k-1)} \vee s_{(k-1)} \vee q_{(k-1)}. \end{cases} \quad (7)$$

Окончательный результат универсального преобразования формально может быть представлен в виде следующего операторного изображения:

$$F(z_i) = \max_{i=0}^{k-1} \left(\min_{t=0}^{k-1} z_i t_i \right), \quad (8)$$

где $i = \overline{1, k-1}$, $(t_0, t_1, \dots, t_{k-1})$ — наборы сигналов настройки (выбора) выходных функций универсальной АКП — структуры третьего рода [21]. Таким образом, цель данного подхода достигается путем декомпиляции многозначных аппаратных средств (АКП-структур третьего рода) на многозначно-двухзначные дискретно-аналоговые и цифровые субблоки, особенно в части их промежуточного пространственного преобразования информации.

Исследования показали, что применение традиционных методов комбинационного синтеза в функционально полных базисах в виде совершенных дизъюнктивных (конъюнктивных) нормальных форм для многозначных структур пространственного типа малоэффективно с точки зрения сохранения свойств однородности и параллелизма структурных построений [7, 17]. Необходим поиск наиболее естественных и приближенных к внутренней логике функционирования объектов исследования, в частности, естественного языка, разновидности алгебро-логических средств моделирования и новых методов синтеза на основе [19–21].

В задачах обработки информации на естественном языке (например, славянском) уже на уровне фонетики необходимо распознавать и обрабатывать не менее 33 букв. Следовательно, начиная с задач фонетического уровня обработки языковой информации, увеличение значности является неизбежной задачей дальнейших исследований. Очевидно, что рассмотрение такого подхода к исследованию проблем создания и построения АКП-структур третье-

го рода на основе универсальных многозначных функциональных преобразователей необходимо расширять с точки зрения нарастания значности структурного алфавита ($k \geq 3$) [21].

Поскольку преобразователь обладает свойством универсальности, то мощность множества функций, которые могут быть реализованы одноходовой АКП-структурой третьего рода, составляет величину $N = k^k$. Нарастание значности структурного алфавита (количества параллельных каскадов АЦП из линейки компараторов с совместным опорным делителем напряжения) с 3 до 10 позволяет повысить функциональные возможности относительно реализации множества логических функций одной переменной в $10^{10}/10^3$ раз. Таким образом, применение двухходовых АКП-структур третьего рода, а также построение пространственного дешифратора, матричного селектора и коммутатора на элементах конъюнкции позволило обеспечить однотипность и однородность ее внутренней структуры, а также повысить быстродействие за счет граничного параллелизма структуры. В АКП-структуре используются логические, а не вычислительные методы промежуточных преобразований с применением концепции унификации двух- и k -значного кодирования, что обеспечивает упрощение структуры промежуточных субблоков матричного селектора и коммутатора (рис. 2).

Результаты и перспективы исследований

Таким образом, приведенные выше результаты позволяют сделать следующий важный вывод: использование новых алгебро-логических средств моделирования естественно-языковых конструкций в виде системы уравнений на языке алгебры конечных предикатов и явного способа задания конечного алфавитного оператора, который лежит в основе аппаратного метода решения этих уравнений, обеспечивает реализацию свойства обратимости АКП-структур и широкое распараллеливание обработки символической информации. Фундаментальные исследования алгебро-логической структуры естественного языка, а также алгебро-логических средств его моделирования в виде соответствующих АКП-структур первого, второго и третьего рода позволяют вплотную подойти к решению важной научной проблемы: достижению качественно новых технологий обработки символической информации на основе концепции унификации и методов синтеза обратимых пространственных многозначных структур языковых систем.

Список литературы: 1. Будущее искусственного интеллекта. — М.: Наука, 1991. — 302 с. 2. Лачинов В.М., Поляков А.О. Информодинамика, или Путь к Миру открытых систем. — СПб.: Изд-во СПбГТУ, 1999. — 432 с. 3. Загоруйко Н. Г. Искусственная мудрость // Proc. of the X-th International Conf. «Knowledge-Dialog-Solution» (KDS-

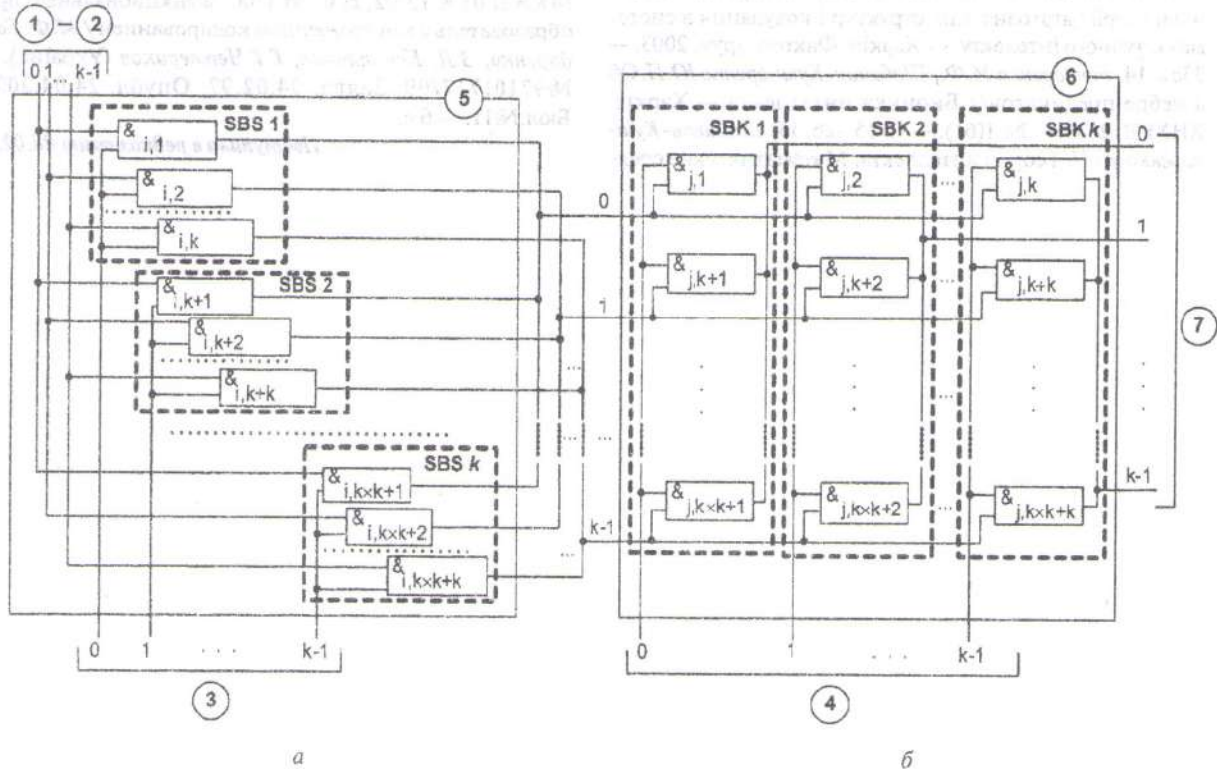


Рис. 2. Структура и логика реализации комбинационно-матричных схем: а — селектора; б — коммутатора

го рода на основе универсальных многозначных функциональных преобразователей необходимо расширять с точки зрения нарастания значности структурного алфавита ($k \geq 3$) [21].

Поскольку преобразователь обладает свойством универсальности, то мощность множества функций, которые могут быть реализованы одноходовой АКП-структурой третьего рода, составляет величину $N = k^k$. Нарастание значности структурного алфавита (количества параллельных каскадов АЦП из линейки компараторов с совместным опорным делителем напряжения) с 3 до 10 позволяет повысить функциональные возможности относительно реализации множества логических функций одной переменной в $10^{10}/10^3$ раз. Таким образом, применение двухходовых АКП-структур третьего рода, а также построение пространственного дешифратора, матричного селектора и коммутатора на элементах конъюнкции позволило обеспечить однотипность и однородность ее внутренней структуры, а также повысить быстродействие за счет граничного параллелизма структуры. В АКП-структуре используются логические, а не вычислительные методы промежуточных преобразований с применением концепции унификации двух- и k -значного кодирования, что обеспечивает упрощение структуры промежуточных субблоков матричного селектора и коммутатора (рис. 2).

Результаты и перспективы исследований

Таким образом, приведенные выше результаты позволяют сделать следующий важный вывод: использование новых алгебро-логических средств моделирования естественно-языковых конструкций в виде системы уравнений на языке алгебры конечных предикатов и явного способа задания конечного алфавитного оператора, который лежит в основе аппаратного метода решения этих уравнений, обеспечивает реализацию свойства обратимости АКП-структур и широкое распараллеливание обработки символической информации. Фундаментальные исследования алгебро-логической структуры естественного языка, а также алгебро-логических средств его моделирования в виде соответствующих АКП-структур первого, второго и третьего рода позволяют вплотную подойти к решению важной научной проблемы: достижению качественно новых технологий обработки символической информации на основе концепции унификации и методов синтеза обратимых пространственных многозначных структур языковых систем.

Список литературы: 1. Будущее искусственного интеллекта. — М.: Наука, 1991. — 302 с. 2. Лачинов В.М., Поляков А.О. Информодинамика, или Путь к Миру открытых систем. — СПб.: Изд-во СПбГТУ, 1999. — 432 с. 3. Загоруйко Н. Г. Искусственная мудрость // Proc. of the X-th International Conf. «Knowledge-Dialog-Solution» (KDS-

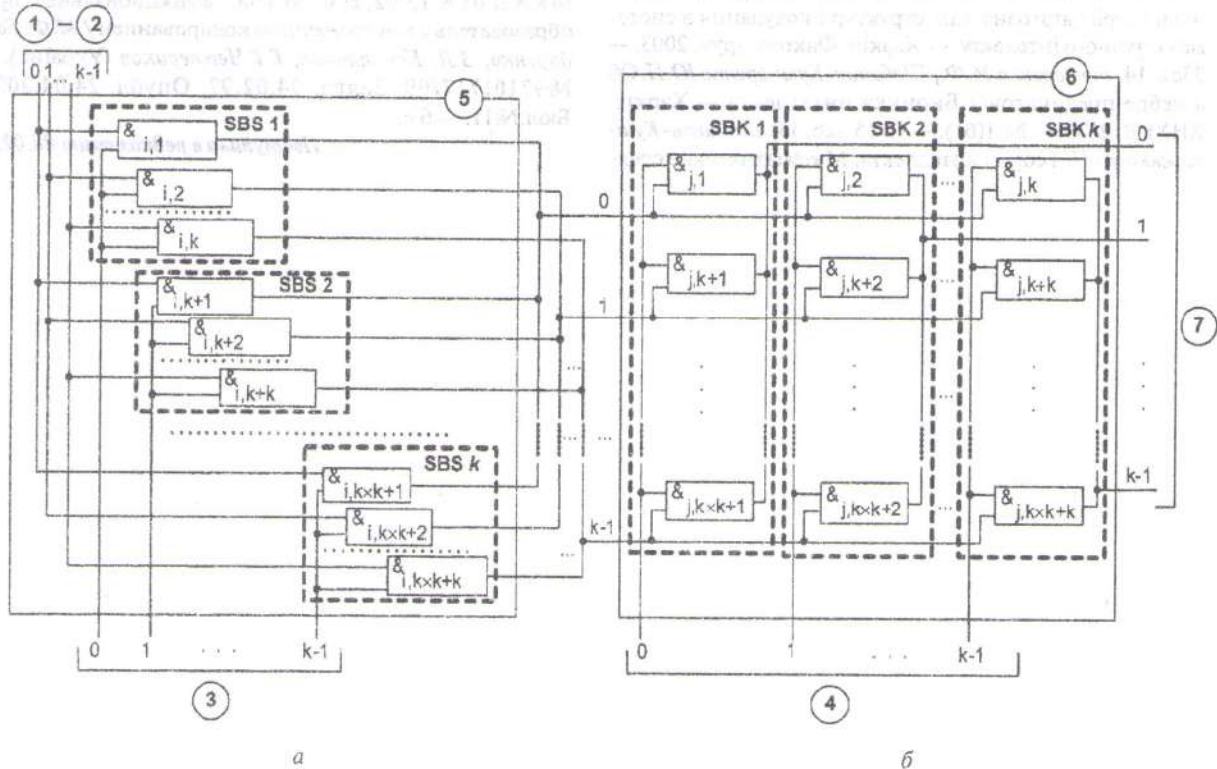


Рис. 2. Структура и логика реализации комбинационно-матричных схем: а — селектора; б — коммутатора

- 2003). — Varna (Bulgaria). — 2003. — С. 3–15. 4. Коноплянко З.Д., Четвериков Г.Г. Проблеми інтелектуалізації та українізації цифрових систем та мереж телекомунікацій // Труды УНИИРТ. — 1995. — №4. — С.72–81. 5. Bondarenko M., Chetverikov G., Karpuhin A. Structural Synthesis of Universal Multiple-Valued Structures of Artificial Intelligence Systems // Proc. of the 9th World Multi-Conference in Systemics, Cybernetics and Informatics (WMSCI 2005). — Orlando, Florida (USA). — 2005. — Vol. VII. — P. 127–130. 6. Bondarenko M., Karpuhin A., Chetverikov G., Leshchinsky V. Synthesis Methods of Multiple-valued Structures of Biological Networks // Proc. of the 12-th International Conf. «Mixed design of integrated circuits and systems» (MIXDES 2005), Krakow (Polska). — 2005. — P. 201–204. 7. Bondarenko M., Chetverikov G., Karpuhin A., Roshka S., Deyneko Zh. Synthesis Methods of Multiple-valued Structures of Language Systems // Proc. of the Third International Conf. «Information Research, Applications and Education» (i.TECH 2005). — Varna (Bulgaria). — 2005. — P. 102–107. 8. Реализация многозначных структур автоматизи. Под ред. М.А.Ракова. — К.: Наукова думка, 1976. — 350 с. 9. Самофалов К.Г., Корнейчук В.И., Романкевич А.М., Тарасенко В.П. Цифровые многозначные элементы и структуры. — К.: Вища школа, 1974. — 168 с. 10. Г.Г. Четвериков. Многозначные структуры (анализ, сравнение, синтез, обобщение). Ч. I. — К.: ИСМО, 1997. — 192 с. 11. Коноплянко З.Д., Чаплыга В.М., Чаплыга М. В. Багатозначні структури та кодування систем економічної кібернетики. — Львів: ЛБІ НБУ, 2004. — 314 с. 12. Бондаренко М.Ф., Коноплянко З.Д., Четвериков Г.Г. Основи теорії синтезу надшвидкодійних структур мовних систем штучного інтелекту — К.: ІЗМН, 1997. — 264 с. 13. Бондаренко М.Ф., Коноплянко З.Д., Четвериков Г.Г. Основи теорії багатозначних структур і кодування в системах штучного інтелекту. — Харків: Фактор-друк, 2003. — 336 с. 14. Бондаренко М.Ф., Шабанов-Кушнарченко Ю.П. Об алгебре предикатов // Бионика интеллекта — Харків: ХНУРЭ, 2004. — № 1(61). — С. 15–26. 15. Шабанов-Кушнарченко Ю.П. Теория интеллекта. Математические средства. — Харків: Вища шк., 1984. — 144с. 16. Четвериков Г.Г. Формалізація принципів побудови універсальних k -значних структур мовних систем штучного інтелекту // Доповіді НАН України. — 2001. — №1 (41). — С. 76–79. 17. Четвериков Г.Г. Алгебро-логічні засоби моделювання природної мови. Повідомлення 1 // Бионика интеллекта. — 2004. — №1(61). — С. 42–45. 18. Четвериков Г.Г. Алгебро-логічні засоби моделювання природної мови. Повідомлення 2 // Бионика интеллекта. — 2005. — № 2 (62). — С. 12–19. 19. Бондаренко М.Ф., Четвериков Г.Г. Особливості архітектурних побудов просторових k -значних структур мовних систем штучного інтелекту // Проблеми бионики. — Харьков: ХНУРЭ. — 2002. — Вып. 56. — С. 3–9. 20. Бондаренко М.Ф., Коноплянко З.Д., Четвериков Г.Г. Методи завадостійкого k -значного кодування та захисту інформації в україномовних інтерфейсах систем штучного інтелекту // Відбір і обробка інформації. — Львів: ФМІ НАН України. — 1998. — Вип. 11 (87). — С. 16–19. 21. Четвериков Г.Г. Концепция построения сверхбыстродействующих многозначных структур языковых систем искусственного интеллекта // Автоматизация производственных процессов. — 1996. — №2. — С. 89–94. 22. Четвериков Г.Г., Бондаренко М.Ф., Шабанов-Кушнарченко Ю.П. Модуль для преобразования кодов, заданных логическими уравнениями: А.с. 1241483 СССР, МКИН03М 7/00 / (СССР). — №3844735/24; Заявлено 11.12.84; Опубл. 30.06.86, Бюл.№324.—5 с.ил. 23. Пат. 20462 А. Україна, МКВ Н 03 К 19/02. Двовходовий багатозначний логічний елемент / М.Ф.Бондаренко, З.Д. Коноплянко, Г.Г. Четвериков (Україна). — №97031289; Дата подання 20.03.97; Опубл. 15.07.97, Бюл. № 3. — 4 с. 24. Пат. 2147789 РФ, МКВ Н 03 К 19/02, Н 03 М 1/00. Функциональный преобразователь с многозначным кодированием / М.Ф. Бондаренко, З.Д. Коноплянко, Г.Г. Четвериков (Україна). — №97101717/09; Заявл. 04.02.97; Опубл. 24.04.2000, Бюл.№11. — 6 с.

Поступила в редколлегию 08.02.07

УДК 519.7:004.8

АВТОАССОЦИАТИВНАЯ НЕЙРОСЕТЕВАЯ ПАМЯТЬ НА ОСНОВЕ НЕЧЕТКИХ БАЗИСНЫХ ФУНКЦИЙ

Е. В. Бодянский¹, Н. А. Тесленко²¹ХНУРЭ, г. Харьков, Украина, bodya@kture.kharkov.ua;²ХНУРЭ, г. Харьков, Украина, ntnp@ukr.net.

Предложена нейро-фаззи архитектура автоассоциативной нейросетевой памяти, обучение которой основано на нечетких базисных функциях. Определены процессы накопления и восстановления информации в фундаментальной памяти. Использование гибридной нейро-фаззи системы позволяет повысить емкость ассоциативной памяти без существенного усложнения её конструкции.

АССОЦИАТИВНАЯ ПАМЯТЬ, НЕЧЕТКИЕ БАЗИСНЫЕ ФУНКЦИИ, ГИБРИДНЫЕ НЕЙРО-ФАЗЗИ СИСТЕМЫ, ФУНКЦИИ ПРИНАДЛЕЖНОСТИ, МНОЖЕСТВО ОБРАЗОВ, ФУНДАМЕНТАЛЬНАЯ ПАМЯТЬ, ЕМКОСТЬ ПАМЯТИ.

Введение

Одним из важных свойств биологического мозга является способность к запоминанию информации и её последующему восстановлению с помощью системы ассоциаций, при этом увиденный когда-то образ может быть опознан даже спустя длительное время, несмотря на произошедшие с ним изменения. Эту функцию мозга моделирует достаточно широкий класс искусственных нейронных сетей, называемых нейронными сетями ассоциативной памяти [1–7].

Такая искусственная память представляет собой, как правило, нейронную сеть либо с прямой передачей информации (статическая ассоциативная память), либо рекуррентную (динамическая ассоциативная память), которая в режиме обучения накапливает предъявляемые ей образы (фаза записи), а в режиме (фазе) восстановления вызывает их из памяти, ассоциируя вновь предъявленные данные с ранее увиденными. При этом весь набор предъявленных и запомненных образов формирует множество фундаментальной памяти.

В отличие от обычных блоков памяти ЭВМ, в нейросетевой памяти информация распределяется по всему множеству нейронов, формирующих сеть, что повышает надежность ее хранения, при этом поиск информации производится не по адресу, а по содержанию, и, кроме того, обеспечивается возможность обработки искаженной или неполной информации, что позволяет правильно опознавать по «зашумленному» фрагменту образ в целом.

Основное отличие нейросетевой ассоциативной памяти от аппроксимирующих нейронных сетей состоит в том, что, если последние реализуют однозначное нелинейное отображение

$$y(k) = F(x(k)), \quad y(k) \in R^m, \quad x(k) \in R^n,$$

то ассоциативная память отображает в $y(k)$ все возможные векторы входов x , принадлежащие неко-

торой окрестности $x(k)$ такой, что $\|x - x(k)\| < \epsilon$. Здесь $y(k)$, $x(k) - (m \times 1)$, $(n \times 1)$ — запоминаемые векторы фундаментальной памяти, $k = 1, 2, \dots, l$ — номер конкретного образа в множестве фундаментальной памяти, ϵ — положительный параметр.

Среди множества структур нейросетевой ассоциативной памяти широкое распространение получили автоассоциативные памяти, реализующие отображение

$$x(k) = F(x(k))$$

для всех x , принадлежащих окрестности, определяемой параметром ϵ , при этом основной функцией этих нейросетей является восстановление неполной или искаженной информации, например, узнавание образа по обрывку старой фотографии или диагностика по данным с пропусками и ошибками.

Среди автоассоциативных нейросетевых памяти наиболее популярны [6] корреляционная матрица-память, автомат собственных векторов, нейронная сеть Хопфилда, автоассоциативный многослойный перцептрон, двунаправленная ассоциативная память, «Brain-State-in-a-Box» нейромодель, общим недостатком которых является малая емкость, определяемая отношением числа запоминаемых образов l к их размерности n . Для перечисленных памяти значение l/n , как правило, не превышает единицу.

По сравнению с вышеперечисленными конструкциями гораздо большей емкостью обладают решетчатые ассоциативные нейросетевые памяти (РАНП) [8, 9], весьма близкие по архитектуре к широко распространенным радиально-базисным нейронным сетям (РБНС). Выходной сигнал этих нейросетей формируется на основе линейной комбинации взаимно перекрывающихся радиально-базисных колоколообразных функций, равномерно распределенных в n -мерном пространстве входов. Основным отличием РАНП от РБНС является то,

что многомерные радиально-базисные функции формируются из одномерных нечетких (фаззи) базисных функций, центры которых равноудалены друг от друга на одномерной оси, соответствующей i -й ($i = 1, 2, \dots, n$) компоненте многомерного входного сигнала. Основным недостатком этих нейросетей является проклятие размерности, ведущее к экспоненциальному росту нейронов скрытого слоя с увеличением размерности входного пространства.

Повышение емкости ассоциативных памяти без существенного усложнения конструкции может быть обеспечено с помощью гибридных нейро-фаззи систем [11, 12], сочетающих в себе достоинства искусственных нейронных сетей и систем нечеткого вывода.

1. Обучение автоассоциативной памяти на основе нечетких базисных функций

Введем в рассмотрение нейро-фаззи архитектуру, приведенную на рис. 1 и предназначенную для накопления множества образов фундаментальной памяти $x^*(k)$, $k = 1, 2, \dots, l$, $x^*(k) \in R^n$, образующих $(n \times l)$ — матрицу X^* . Данная архитектура содержит четыре слоя обработки информации и достаточно близка к нечеткой сети Л. Ванга-Дж. Менделя [10, 13], хотя решает при этом иные задачи. Первый

слой сети образован nl узлами, каждый из которых реализует функцию принадлежности $\mu_{\tilde{x}_i^*(k)}$, соответствующую i -й координате k -го запоминаемого образа. Данные функции определены на интервале $[0, 1]$, что требует предварительного кодирования исходных данных с помощью соотношения

$$\tilde{x}_i^* = \frac{x_i^* - x_i^{*\min}}{x_i^{*\max} - x_i^{*\min}}, \quad (1)$$

где $x_i^{*\max}$ и $x_i^{*\min}$ — максимальный и минимальный элементы i -й строки матрицы X^* , соответственно. Обратное (1) преобразование имеет вид:

$$x_i^* = x_i^{*\max} \tilde{x}_i^* - x_i^{*\min} (\tilde{x}_i^* - 1),$$

при этом, если $x_i^{*\min} \leq x_i^* \leq x_i^{*\max}$, то $0 \leq \tilde{x}_i^* \leq 1$.

В простейшем случае функции принадлежности могут быть приняты треугольными так, как это показано на рис. 2, а их аналитическое выражение имеет вид:

$$\mu_{1i}(\tilde{x}_i) = \frac{c_{2i} - \tilde{x}_i}{c_{2i}}, \text{ если } \tilde{x}_i \in [0, c_{2i}], \quad (2)$$

$$\mu_{2i}(\tilde{x}_i) = \frac{\tilde{x}_i - c_{1-1,i}}{1 - c_{1-1,i}}, \text{ если } \tilde{x}_i \in [c_{1-1,i}, 1], \quad (3)$$

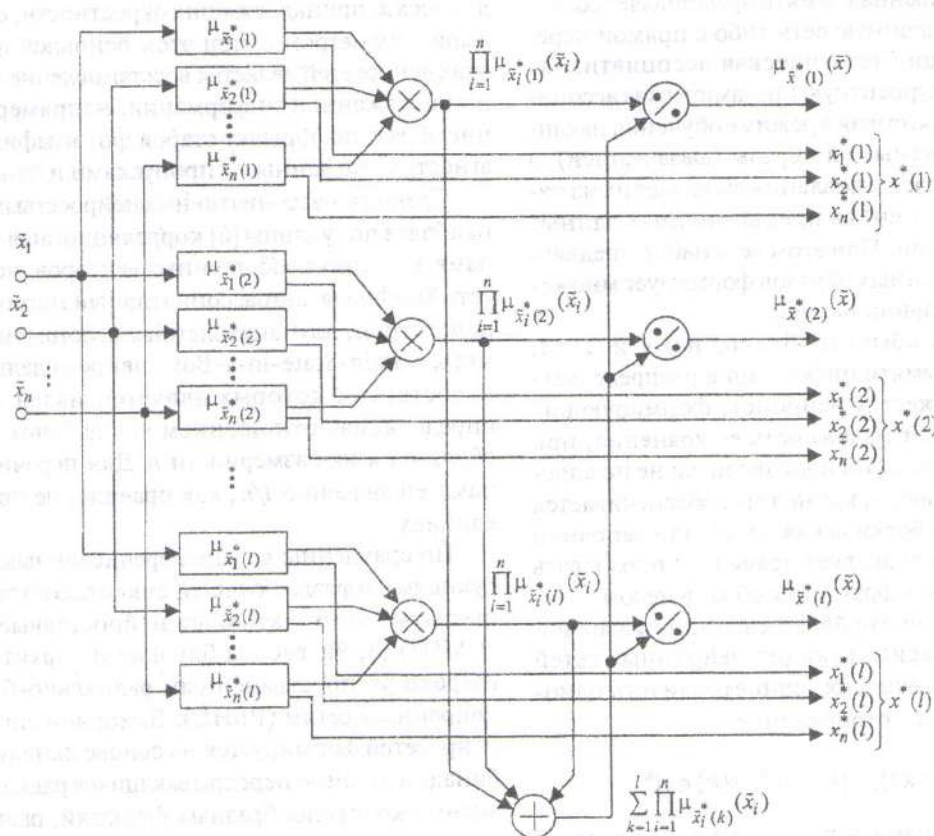


Рис. 1. Подсистема записи сети ассоциативной памяти на основе нечетких базисных функций

$$\mu_{ji}(\bar{x}_i) = \begin{cases} \bar{x}_i - c_{j-1,i}, & \text{если } \bar{x}_i \in [c_{j-1,i}, c_{ji}], \\ c_{ji} - c_{j-1,i}, & \\ c_{j+1,i} - \bar{x}_i, & \text{если } \bar{x}_i \in [c_{ji}, c_{j+1,i}], \\ c_{j+1,i} - c_{ji}, & \\ 0, & \text{в противном случае,} \\ j = 2, \dots, l-1, \end{cases} \quad (4)$$

где c_{ji} — параметры центра функций принадлежности, определяемые в процессе обучения.

Такой способ задания функций принадлежности автоматически обеспечивает разбиение Руспини, т. е. выполнение условия

$$\sum_{j=1}^l \mu_{ji}(\bar{x}_i) = 1, \quad (5)$$

хотя, конечно, возможно использование функций иного вида.

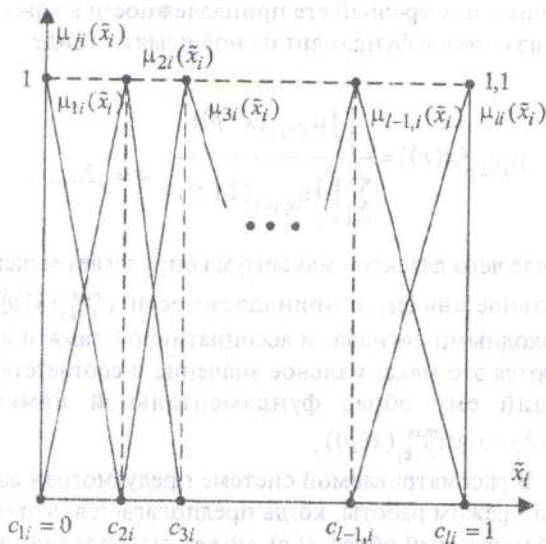


Рис. 2. Функции принадлежности подсистемы записи

Функции принадлежности первого слоя объединены в l блоков по n функций, каждый из которых соответствует одному конкретному образу фундаментальной памяти $\bar{x}^*(k)$. Второй скрытый слой образован l блоками умножения и выполняет агрегирование значений активации для каждого запоминаемого образа так, что в блоках умножения вычисляются значения $\prod_{i=1}^n \mu_{\bar{x}_i^*(k)}(\bar{x}_i)$, которые складываются в сумматоре третьего скрытого слоя, вычисляющего величину $\sum_{k=1}^l \prod_{i=1}^n \mu_{\bar{x}_i^*(k)}(\bar{x}_i)$. Выходной слой осуществляет нормализацию выходного сигнала так, что при подаче на вход системы произвольного n -мерного образа \bar{x} на выходе появляется l сигналов:

$$\mu_{\bar{x}^*(k)}(\bar{x}) = \frac{\prod_{i=1}^n \mu_{\bar{x}_i^*(k)}(\bar{x}_i)}{\sum_{k=1}^l \prod_{i=1}^n \mu_{\bar{x}_i^*(k)}(\bar{x}_i)}, \quad k = 1, 2, \dots, l.$$

Процесс обучения данной архитектуры подобен обучению вероятностных нейронных сетей [14] и обобщенных регрессионных нейронных сетей [15] и реализуется по принципу «нейроны в точках данных» [16] так, что центры функций принадлежности первого слоя совпадают с координатами проекций запоминаемых образов.

Данный процесс иллюстрируется примером на рис. 3, где в верхней части показано три двумерных образа фундаментальной памяти $x^*(1), x^*(2), x^*(3)$, подлежащих запоминанию. Их проекции на оси x_1 и x_2 расположены так, что $x_1^*(1) < x_1^*(3) < x_1^*(2)$ и $x_2^*(3) < x_2^*(2) < x_2^*(1)$. В нижней части рисунка эти же образы, подвергнутые преобразованию (1), обозначены $\bar{x}^*(1), \bar{x}^*(2), \bar{x}^*(3)$ и расположены либо в вершинах, либо на сторонах единичного квадрата (гиперкуба в многомерном случае). В данной конкретной ситуации архитектура, приведенная на рис. 1, имеет два входа и шесть функций принадлежности: $\mu_{\bar{x}_1^*(1)}, \mu_{\bar{x}_2^*(1)}, \mu_{\bar{x}_1^*(2)}, \mu_{\bar{x}_2^*(2)}, \mu_{\bar{x}_1^*(3)}, \mu_{\bar{x}_2^*(3)}$. С проекциями образов $\bar{x}^*(k)$ на оси $\bar{x}_i^*(k)$ связываются центры функций принадлежности так, что $c_{11} = \bar{x}_1^*(1) = 0$, $c_{21} = \bar{x}_1^*(3)$, $c_{31} = \bar{x}_1^*(2) = 1$ и $c_{12} = \bar{x}_2^*(3) = 0$, $c_{22} = \bar{x}_2^*(2)$, $c_{32} = \bar{x}_2^*(1) = 1$. С каждым из центров c_{ji} связываются свои функции принадлежности μ_{ji} , $j = 1, 2, 3$; $i = 1, 2$, также приведенные на рис. 3 и описываемые соотношениями (2)–(4). Таким образом, в первом слое в процессе обучения определяются центры функций принадлежностей и устанавливается однозначное соответствие между $\bar{x}_i^*(k)$ и c_{ji} и между $\mu_{\bar{x}_i^*(k)}$ и μ_{ji} .

В результате процесса обучения формируется архитектура подсистемы запоминания ассоциативной памяти, определяемая количеством образов фундаментальной памяти, и рассчитываются функции принадлежности, определяемые взаимным расположением этих образов.

2. Восстановление образов в автоассоциативной памяти на основе нечетких базисных функций

После того как в автоассоциативную память записано l образов фундаментальной памяти $\bar{x}^*(1), \bar{x}^*(2), \dots, \bar{x}^*(l)$, система готова к работе в режиме восстановления, для чего записывающая подсистема, приведенная на рис. 1, должна быть дополнена $l(n+1)$ -входовыми блоками умножения, суммато-

$$\mu_{ji}(\bar{x}_i) = \begin{cases} \bar{x}_i - c_{j-1,i}, & \text{если } \bar{x}_i \in [c_{j-1,i}, c_{ji}], \\ c_{ji} - c_{j-1,i}, & \\ c_{j+1,i} - \bar{x}_i, & \text{если } \bar{x}_i \in [c_{ji}, c_{j+1,i}], \\ c_{j+1,i} - c_{ji}, & \\ 0, & \text{в противном случае,} \\ j = 2, \dots, l-1, \end{cases} \quad (4)$$

где c_{ji} — параметры центра функций принадлежности, определяемые в процессе обучения.

Такой способ задания функций принадлежности автоматически обеспечивает разбиение Руспини, т. е. выполнение условия

$$\sum_{j=1}^l \mu_{ji}(\bar{x}_i) = 1, \quad (5)$$

хотя, конечно, возможно использование функций иного вида.

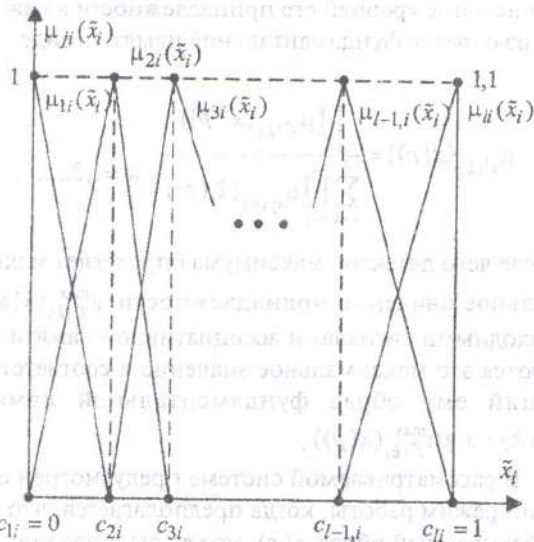


Рис. 2. Функции принадлежности подсистемы записи

Функции принадлежности первого слоя объединены в l блоков по n функций, каждый из которых соответствует одному конкретному образу фундаментальной памяти $\bar{x}^*(k)$. Второй скрытый слой образован l блоками умножения и выполняет агрегирование значений активации для каждого запоминаемого образа так, что в блоках умножения вычисляются значения $\prod_{i=1}^n \mu_{\bar{x}_i^*(k)}(\bar{x}_i)$, которые складываются в сумматоре третьего скрытого слоя, вычисляющего величину $\sum_{k=1}^l \prod_{i=1}^n \mu_{\bar{x}_i^*(k)}(\bar{x}_i)$. Выходной слой осуществляет нормализацию выходного сигнала так, что при подаче на вход системы произвольного n -мерного образа \bar{x} на выходе появляется l сигналов:

$$\mu_{\bar{x}^*(k)}(\bar{x}) = \frac{\prod_{i=1}^n \mu_{\bar{x}_i^*(k)}(\bar{x}_i)}{\sum_{k=1}^l \prod_{i=1}^n \mu_{\bar{x}_i^*(k)}(\bar{x}_i)}, \quad k = 1, 2, \dots, l.$$

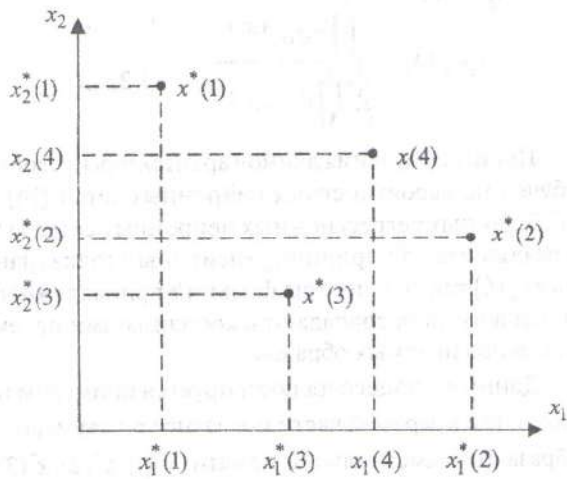
Процесс обучения данной архитектуры подобен обучению вероятностных нейронных сетей [14] и обобщенных регрессионных нейронных сетей [15] и реализуется по принципу «нейроны в точках данных» [16] так, что центры функций принадлежности первого слоя совпадают с координатами проекций запоминаемых образов.

Данный процесс иллюстрируется примером на рис. 3, где в верхней части показано три двумерных образа фундаментальной памяти $x^*(1), x^*(2), x^*(3)$, подлежащих запоминанию. Их проекции на оси x_1 и x_2 расположены так, что $x_1^*(1) < x_1^*(3) < x_1^*(2)$ и $x_2^*(3) < x_2^*(2) < x_2^*(1)$. В нижней части рисунка эти же образы, подвергнутые преобразованию (1), обозначены $\bar{x}^*(1), \bar{x}^*(2), \bar{x}^*(3)$ и расположены либо в вершинах, либо на сторонах единичного квадрата (гиперкуба в многомерном случае). В данной конкретной ситуации архитектура, приведенная на рис. 1, имеет два входа и шесть функций принадлежности: $\mu_{\bar{x}_1^*(1)}, \mu_{\bar{x}_2^*(1)}, \mu_{\bar{x}_1^*(2)}, \mu_{\bar{x}_2^*(2)}, \mu_{\bar{x}_1^*(3)}, \mu_{\bar{x}_2^*(3)}$. С проекциями образов $\bar{x}^*(k)$ на оси $\bar{x}_i^*(k)$ связываются центры функций принадлежности так, что $c_{11} = \bar{x}_1^*(1) = 0$, $c_{21} = \bar{x}_1^*(3)$, $c_{31} = \bar{x}_1^*(2) = 1$ и $c_{12} = \bar{x}_2^*(3) = 0$, $c_{22} = \bar{x}_2^*(2)$, $c_{32} = \bar{x}_2^*(1) = 1$. С каждым из центров c_{ji} связываются свои функции принадлежности μ_{ji} , $j = 1, 2, 3$; $i = 1, 2$, также приведенные на рис. 3 и описываемые соотношениями (2)–(4). Таким образом, в первом слое в процессе обучения определяются центры функций принадлежностей и устанавливается однозначное соответствие между $\bar{x}_i^*(k)$ и c_{ji} и между $\mu_{\bar{x}_i^*(k)}$ и μ_{ji} .

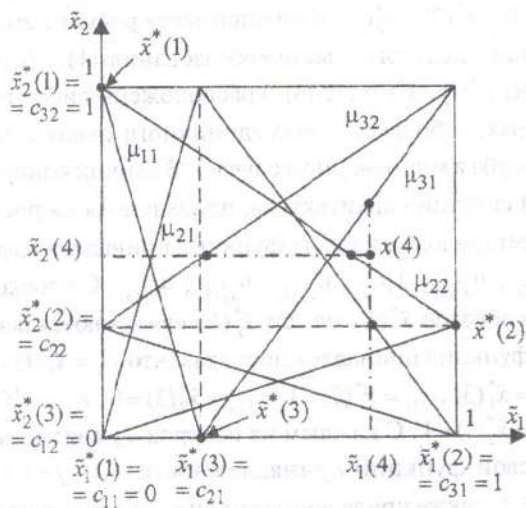
В результате процесса обучения формируется архитектура подсистемы запоминания ассоциативной памяти, определяемая количеством образов фундаментальной памяти, и рассчитываются функции принадлежности, определяемые взаимным расположением этих образов.

2. Восстановление образов в автоассоциативной памяти на основе нечетких базисных функций

После того как в автоассоциативную память записано l образов фундаментальной памяти $\bar{x}^*(1), \bar{x}^*(2), \dots, \bar{x}^*(l)$, система готова к работе в режиме восстановления, для чего записывающая подсистема, приведенная на рис. 1, должна быть дополнена $l(n+1)$ -входовыми блоками умножения, суммато-



а



б

Рис. 3. Обучение центров функций принадлежности

ром с $n!$ входами и детектором максимума, в качестве которого может выступать, например, нейронная сеть MAXNET [17]. Поступающие в систему образы $x(p)$, $p = l+1, l+2, \dots$ также должны быть закодированы согласно соотношению (1), при этом, однако, возможны ситуации, когда $\tilde{x}_i(p)$ принимает либо отрицательные, либо превосходящие единицу значения. В связи с этим должно быть введено дополнительное преобразование поступающих на обработку сигналов в виде

$$\tilde{x}_i(p) = \begin{cases} 0, & \text{если } \tilde{x}_i(p) < 0, \\ \tilde{x}_i(p), & \text{если } 0 \leq \tilde{x}_i(p) \leq 1, \\ 1, & \text{если } \tilde{x}_i(p) > 1, \end{cases}$$

т. е. фактически производится проецирование на единичный отрезок $[0, 1]$.

При подаче на вход системы, приведенной на рис. 4, произвольного образа $\tilde{x}(p)$ производится вычисление уровней его принадлежности к каждому из образов фундаментальной памяти в виде

$$\mu_{\tilde{x}^*(k)}(\tilde{x}(p)) = \frac{\prod_{i=1}^n \mu_{\tilde{x}_i^*(k)}(\tilde{x}_i(p))}{\sum_{k=1}^l \prod_{i=1}^n \mu_{\tilde{x}_i^*(k)}(\tilde{x}_i(p))}, \quad k = 1, 2, \dots, l,$$

после чего детектор максимума определяет максимальное значение принадлежности $\mu_{\tilde{x}^*(k)}^{\max}(\tilde{x}(p))$. Выходными сигналами ассоциативной памяти являются это максимальное значение и соответствующий ему образ фундаментальной памяти $\tilde{x}^*(k) = \arg \mu_{\tilde{x}^*(k)}^{\max}(\tilde{x}(p))$.

В рассматриваемой системе предусмотрен еще один режим работы, когда предполагается, что обрабатываемый образ $\tilde{x}(p)$ может быть ассоцииро-

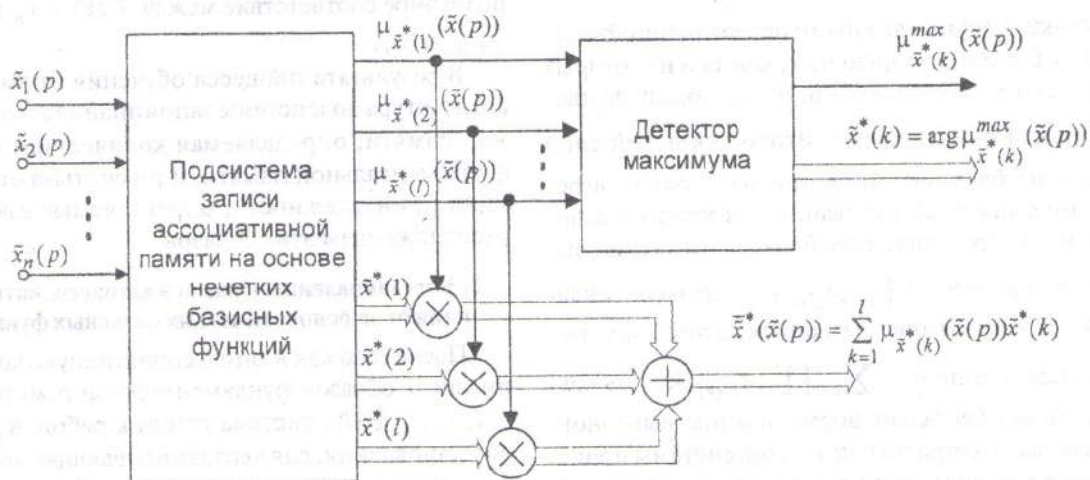


Рис. 4. Восстановление образов в ассоциативной нейросетевой памяти на основе нечетких базисных функций

ван не с единственным образом фундаментальной памяти $\tilde{x}^*(k)$, а представлен в виде комбинации группы образов, с которыми $\tilde{x}(p)$ связан ненулевой принадлежностью. При этом на выходе системы появляется «гибридный» образ фундаментальной памяти

$$\tilde{x}^*(\tilde{x}(p)) = \sum_{k=1}^l \mu_{\tilde{x}^*(k)}(\tilde{x}(p)) \tilde{x}^*(k). \quad (6)$$

Благодаря условию (5) оценка (6) не требует дополнительной нормализации.

Процесс восстановления также может быть проиллюстрирован на рис. 3, где образ, подаваемый в систему после ее обучения, обозначен $x(4)$ на рис. 3а и $\tilde{x}^*(4)$ на рис. 3б. Несложно видеть, что с максимальным уровнем принадлежности $x(4)$ связан с $\tilde{x}^*(2)$, который и будет выделен детектором максимума. С другой стороны, как опять-таки видно на рис. 3б, образ $x(4)$ является линейной комбинацией образов фундаментальной памяти $\tilde{x}^*(2)$ и $\tilde{x}^*(3)$ и никак не связан в $\tilde{x}^*(1)$.

Заключение

Предложены архитектура и алгоритм обучения автоассоциативной нейросетевой памяти на основе нечетких базисных функций, обладающей повышенной емкостью и характеризующейся простотой реализации. Введенная нейро-фаззи ассоциативная память обладает расширенными функциональными возможностями и позволяет производить восстановление в форме взвешенной комбинации образов фундаментальной памяти.

Список литературы: 1. *Specht D.F.* Probabilistic neural networks for classification, mapping, or associative memory// Proc. IEEE Int. Conf. Neural Networks. — June 1988. — V.1.

— P. 525—432. 2. *Michel A.N., Farrel J.A.* Associative memories via artificial neural networks// IEEE Control System Magazine. — 1990. — 10. — № 3. — P.6-17. 3. *Hassoun M.H.* Fundamentals of Artificial Neural Networks.- Cambridge, MA: MIT Press, 1995. 4. *Rojas R.* Neural Networks. A Systematic Introduction.-Berlin: Springer-Verlag, 1996. — 502p. 5. *Hagan M.T., Demuth H.B., Beale M.* Neural Networks Design.- Boston: PWS Publishing Company, 1996. — 729p. 6. *Hassoun M.H., Watta P.B.* Associative memory networks/ In «Handbook of Neural Computation». — Oxford: IOP Publishing Ltd and Oxford University Press, 1997. — C1.3:1-C1.3:14. 7. *Haykin S.* Neural Networks. A Comprehensive Foundation. — Upper Saddle River, N.J.: Prentice Hall, Inc., 1999. — 842p. 8. *Brown M., Harris C.J.* Neural networks for modeling and control/ In: Eds. by C.J. Harris «Advances in Intellectual Control». — London: Taylor and Francis, 1994. — P.17-55. 9. *Wang H., Liu G.P., Harris G.J., Brown M.* Advanced Adaptive Control. — Oxford: Pergamon, 1995. — 262p. 10. *Wang L.-X., Mendel J.M.* Fuzzy basis functions, universal approximation, and orthogonal least squares learning//IEEE Trans. on Neural Networks. — 1992. — 3. — P.807-814. 11. *Jang J.-S. R., Sun G.-T.* Neuro-fuzzy modeling and control// Proc. IEEE. — 1995. — 83. — P. 378-406. 12. *Cios K.J., Pedrycz W.* Neuro-fuzzy algorithms// In «Handbook on Neural Computation» — Oxford: IOP Publishing Ltd and Oxford University Press, 1997. — D1.3:1-D1.3:7. 13. *Wang L.X.* Adaptive Fuzzy Systems and Control. Design and Stability Analysis. - Englewood Cliffs, N.J.: Pentice Hall, 1994. 14. *Specht D.F.* Probabilistic neural networks// Neural Networks. — 1990. — 3. — P.109-118. 15. *Specht D.F.* A general regression neural network// IEEE Trans. on Neural Networks. — 1991. — 2. — P. 568-576. 16. *Zahirniak D.R., Chapman R., Rogers S.K., Suter B.W., Kabrisky M., Pyati V.* Pattern recognition using radial basis function network// Proc. 6-th Ann. Aerospace Application of AI Conf. — Dayton, OH, 1990. — P. 249-260. 17. *Lippmann R.P.* An introduction to computing with neural nets// IEEE Acoustic, Speech, and Signal Processing Magazine. — 1987. — April. — P. 4022.

Поступила в редколлегию 27.02.07

УДК 519.71

НЕЙРОСЕТЕВАЯ АППРОКСИМАЦИЯ МНОГОМЕРНЫХ ФУНКЦИЙ ПРИ НАЛИЧИИ ПОМЕХ ИЗМЕРЕНИЙ

О. Г. Руденко¹, А. В. Островерхий²¹ХНУРЭ, г. Харьков, Украина, rudenko@kture.kharkov.ua²ХНУРЭ, г. Харьков, Украина, aostrich@gmail.com

В статье приводится сравнительный анализ восстановления зашумленных многомерных функций нейронными сетями СМАС различной архитектуры и РБС. Рассматривается проблема выбора базисных функций сети СМАС и параметров РБС. Показано, что для решения данной задачи весьма эффективными являются иерархическая и линейная архитектуры СМАС при выборе в качестве базисной параболической функции, а применение РБС обеспечивает заданную точность восстановления, требуя меньшего объема памяти, но значительно больших вычислительных затрат.

АППРОКСИМАЦИЯ, БАЗИСНЫЕ ФУНКЦИИ, ВОССТАНОВЛЕНИЕ, НЕЙРОННАЯ СЕТЬ СМАС, МНОГОМЕРНАЯ ФУНКЦИЯ, РАДИАЛЬНО-БАЗИСНАЯ СЕТЬ.

Введение

Решение широкого круга задач науки, техники, экономики, например, идентификация, фильтрация, сглаживание, прогнозирование и т. д. связано с аппроксимацией некоторых нелинейных функций вида

$$y(k) = f[x(k)] + \xi(k), \quad (1)$$

где $f[\cdot]$ — непрерывная нелинейная функция; $\xi(k)$ — помеха измерения; $k = 1, 2, \dots$ — дискретное время.

Отсутствие информации о виде функции $f[\cdot]$ зачастую делает традиционные методы аппроксимации неэффективными, а в ряде случаев — неприменимыми. Альтернативой традиционным методам являются нейросетевые технологии.

Являясь универсальными аппроксиматорами, некоторые типы искусственных нейронных сетей (ИНС) позволяют восстановить с любой заданной точностью любую сколь угодно сложную непрерывную нелинейную функцию. Наибольшее распространение при решении данной задачи получили многослойный персептрон (МП) [1], радиально-базисные сети (РБС) [2] и сеть СМАС (Cerebellar Model Articulation Controller) [3]. Все эти сети используют представление нелинейного оператора некоторой системой базисных функций, реализуемой нейронами, сводя задачу аппроксимации к обучению сети, т. е. настройке параметров нейронов на основе предъявления обучающих пар. Такими обучающими парами служат значения аргументов $x(k)$ и функции $y(k)$.

Увеличение размерности решаемой задачи существенно усложняет задачу аппроксимации, а наличие в измерениях помех не только приводит к увеличению времени обучения ИНС, но и выдвигает определенные требования к используемому алгоритму обучения.

Обучение МП, содержащего несколько (чаще всего не больше двух) скрытых слоев, осуществляется обычно с помощью алгоритма обратного распространения ошибки (ОРО), реализация которого связана с существенными вычислительными трудностями. Поэтому более предпочтительными представляются РБС, содержащие лишь один скрытый слой нейронов, для обучения которых применяются рекуррентный метод наименьших квадратов или фильтр Калмана, и сеть СМАС, не имеющая скрытых слоев и использующая простой, но обеспечивающий высокую скорость обучения алгоритм настройки параметров.

Целью данной работы является сравнительный анализ сетей СМАС и РБС в задаче аппроксимации нелинейных функций большой размерности при наличии помех измерений.

1. Сеть СМАС

Сеть СМАС в общем случае осуществляет следующие преобразования:

$$S: X \Rightarrow A, \quad (2)$$

$$P: A \Rightarrow y, \quad (3)$$

где X — N -мерное пространство непрерывных входных сигналов; A — n -мерное пространство ассоциаций; y — вектор выходных сигналов.

Преобразование (2) соответствует кодированию информации

$$a = S(x), \quad (4)$$

а (3) — вычислению выходного сигнала.

Входной слой сети состоит из нейронов, имеющих, как правило, одинаковые базисные (активационные) функции (БФ).

Традиционная сеть СМАС использует БФ прямоугольной формы, однако в некоторых задачах такой выбор является неэффективным, а в ряде слу-

чаев и неприемлемым. Если в сети используются нейроны с БФ, отличными от прямоугольных, то преобразование, осуществляемое сетью СМАС, принимает вид:

$$\hat{y} = a^T \Phi(x) w, \quad (5)$$

где $\Phi(x)$ — диагональная матрица с элементами $\Phi_i(x) = \prod_{j=1}^N \phi_{ij}(x_j)$, $\phi_{ij}(x_j)$ — значение выбранной базисной функции в точке x_j .

Одним из наиболее простых и наиболее эффективных является выбор в качестве БФ тригонометрических функций, например, косинусоидальной [4]:

$$\Phi(x) = \begin{cases} \cos\left(\pi \frac{x-m}{\lambda}\right); & \text{при } x \in \left[m - \frac{\lambda}{2}; m + \frac{\lambda}{2}\right] \\ 0; & \text{в противном случае} \end{cases} \quad (6)$$

и параболической [5]:

$$\Phi(x) = \begin{cases} 1 - \left(2 \frac{x-m}{\lambda}\right)^2; & \text{при } x \in \left[m - \frac{\lambda}{2}; m + \frac{\lambda}{2}\right] \\ 0; & \text{в противном случае,} \end{cases} \quad (7)$$

которая не только имеет форму, близкую к тригонометрическим, но и требует значительно меньших вычислительных затрат при ее реализации. В (6), (7) приняты обозначения: m — центр гиперкуба, λ — длина гиперкуба.

Обучение сети СМАС, как практически и всех других ИНС, заключается в настройке вектора ее весовых параметров w размерности $n \times 1$.

При выборе БФ, отличных от прямоугольных, алгоритм обучения имеет вид:

$$w(k+1) = w(k) + \gamma(k) \times \left(\frac{y(k) - a^T(k) \Phi(x) w(k)}{\|\Phi(x) a(k)\|^2} \Phi(x) a(k) \right), \quad (8)$$

где $\gamma(k)$ — некоторый в общем случае переменный параметр.

Свойства алгоритма (8) в значительной степени зависят от выбора $\gamma(k)$. Несложно показать, что оптимальное значение этого параметра, обеспечивающее максимальную скорость обучения при отсутствии помех ξ , будет равно единице. Для обеспечения же сходимости алгоритма (8) при наличии помех измерений параметр $\gamma(k)$ должен удовлетворять условиям Дворецкого.

Следует, однако, отметить, что с ростом размерности пространства входных переменных N объем

памяти, требуемый для хранения информации о весах сети СМАС, растет экспоненциально. Кроме того, ограниченные размеры памяти сужают сферу применения данной ИНС в реальных приложениях, а увеличение размерности N приводит к возрастанию сложности кодирования информации.

2. Модифицированные архитектуры СМАС

В качестве эффективного способа уменьшения объема требуемой памяти в сети СМАС при работе с многомерными объектами в [8] предлагается построение иерархической структуры СМАС (Hierarchical СМАС — НСМАС), состоящей из нескольких более простых модулей, например, двумерных СМАС. На рис. 1 приведена топология НСМАС, учитывающая, что каждая СМАС содержит два входа, и выходной сигнал СМАС первого слоя является входным сигналом для СМАС второго слоя и т. д. Архитектура НСМАС может быть, соответственно, расширена с использованием данной топологии бинарного дерева.

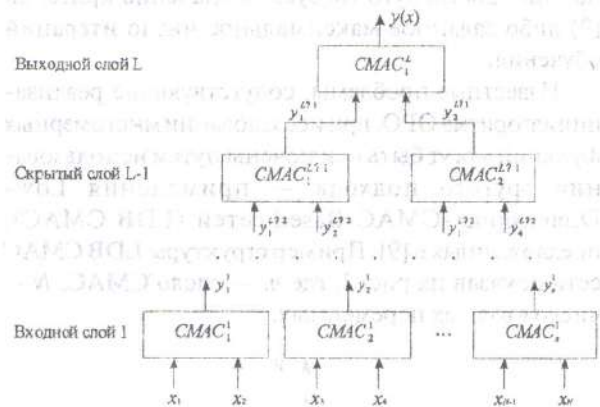


Рис. 1. Топологическая структура нейронной сети НСМАС

На рис. 1 использованы следующие обозначения: x_j ($i = 1, 2, \dots, N$) — i -ый вход нейронной сети НСМАС; y_j^l ($j = 1, 2, \dots, n$; $l = 1, 2, \dots, L$) — выход j -ого СМАС слоя l ; $y(x)$ — выходной сигнал НСМАС для входного сигнала x .

При выборе дифференцируемых БФ [4–7] для обучения данной сети (настройки её параметров) может быть применён алгоритм ОРО. Если в качестве минимизируемого выбран квадратичный функционал ошибки

$$E = \frac{1}{2} (\hat{y}(x) - y(x))^2, \quad (9)$$

где $y(x)$, $\hat{y}(x)$ — требуемый и реальный выходные сигналы НСМАС для входного сигнала x , соответственно, то обобщенная процедура обучения сети может быть представлена так.

1) *Настройка СМАС выходного слоя* (практически ничем не отличается от настройки обычной двухвходовой СМАС, например, по алгоритму (8), за исключением того, что входными для данного слоя являются выходные сигналы предыдущего слоя).

2) *Настройка сетей СМАС скрытых слоев* по правилу:

$$w_h^i(k+1) = w_h^i(k) + \gamma \cdot \left(\hat{y}_{[h/2]}^{i+1}(y') - y_{[h/2]}^{i+1}(y') \right) \times \frac{\partial y_{[h/2]}^{i+1}(x)}{\partial y_h^i} \cdot \Phi(y') a(k); \quad (10)$$

$$\frac{\partial y_{[h/2]}^{i+1}(x)}{\partial y_h^i} = \sum_{i=1}^p a_i(y') \cdot w_{hi}^i \cdot \left[\prod_{j=2}^{[2]^{[h/2]}+1} \Phi_j(y_j') \right] \cdot \frac{\partial \Phi_i(y_h^i)}{\partial y_h^i}, \quad (11)$$

где $h = 1, 2, \dots, n$; w_{hi}^i — значение веса в i -ой ячейке памяти СМАС $_h$ слоя l ; $\lceil \cdot \rceil$ — означает округление в сторону ближайшего большего целого числа.

Данная процедура повторяется до тех пор, пока не будет достигнуто требуемое значение критерия (9) либо заданное максимальное число итераций обучения.

Известные проблемы, сопутствующие реализации алгоритма ОРО, при исследовании многомерных функций могут быть исключены путем использования другого подхода — применения Low-Dimensional-СМАС-Based сетей (LDB СМАС), предложенных в [9]. Пример структуры LDB СМАС сети показан на рис. 2, где m — число СМАС, N — число входных переменных.

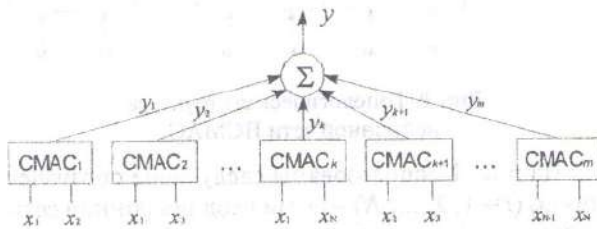


Рис. 2. Структура сети LDB СМАС

Данную структуру образуют множество малоразмерных (базовых) сетей СМАС (например, двумерных), на которые подаются все возможные парные комбинации входных сигналов. Взвешенные весами z_i ($i = 1, 2, \dots, m$) выходы этих СМАС формируют общий выход сети

$$y(x) = \sum_{i=1}^m y_i(x) z_i, \quad (12)$$

где $y_i(x)$ — выходное значение i -ой сети СМАС; z_i — вес соответствующего выхода $y_i(x)$.

Данная архитектура позволяет создавать также и неполные структуры, учитывающие не все воз-

можные комбинации пар входных переменных, а лишь часть их.

Процедура обучения сети LDB СМАС состоит в следующем:

1) *Настройка весов z_i* , например, по алгоритму

$$z(k+1) = z(k) + \frac{y(x) - \hat{y}(x)}{\|Y(k)\|^2} Y(k), \quad (13)$$

где $Y(k) = (y_1(k), y_2(k), \dots, y_m(k))$ — вектор выходов базовых сетей СМАС.

2) *Настройка параметров базовых сетей СМАС $_i$* по правилу (8).

Следует отметить, что данная структура позволяет применять градиентный метод обучения при выборе БФ любой формы, включая прямоугольную.

3. Радиально-базисная сеть

Аппроксимация нелинейной функции (1) радиально-базисными функциями

$$\Phi_i(x) = \exp \left\{ -\frac{\|x - \mu_i\|^2}{\sigma_i^2} \right\}, \quad (14)$$

где μ_i, σ_i — центры и радиусы базисных функций, соответственно; $\|\cdot\|$ — евклидова норма; $\Phi_0(x) = 1$; приводит к нейросетевой модели

$$\hat{y}(k) = \sum_{i=0}^N c_i \Phi_i(x), \quad (15)$$

где c_i — весовые коэффициенты.

Изменение структуры сети осуществляется ее постепенным усложнением с добавлением новых нейронов, проводимым каждый раз, когда при появлении очередного входного сигнала возникает ошибка аппроксимации $e = y - \hat{y}$, превышающая допустимую. В этом случае, если в l -й момент времени сеть содержала N нейронов, а появление сигнала $x(l)$ привело к появлению ошибки на l -м выходе $e_l(l) > e_{l \text{ доп}} = \alpha_l$, в сеть вводится новый, $(N+1)$ -й, нейрон, а центр базисной функции, её вес и радиус принимаются равными, соответственно, $\mu_{N+1} = x(l)$, $c_{N+1} = e(l)$, $\sigma_{N+1} = \|x(l) - \mu_m(l)\|$, где $\mu_m(l)$ — центр базисной функции для l -го входа сигнала. Таким образом, условием введения нового нейрона является выполнение неравенств

$$e_l(l) > \alpha_l, \quad (16)$$

$$\|x(l) - \mu_m(l)\| > \beta, \quad (17)$$

где α_l и β — априорно устанавливаемые предельно допустимые значения ошибки реакции сети и отклонения обобщенного сигнала $x(l)$ от ближайшего к данному входу центра.

Обучение сети состоит в определении её параметров μ_i , σ_i и c_i и сводится к минимизации обычно квадратичного функционала (9).

В настоящее время существует множество методов настройки параметров сети, среди которых достаточно широко используется рекуррентный алгоритм метода наименьших квадратов (РМНК) с экспоненциальным взвешиванием информации, согласно которому вектор оценок настраиваемых параметров

$$w(k) = (c_0^T(k), c_1^T(k), \mu_1^T(k), \sigma_1(k), \dots, c_N^T(k), \mu_N^T(k), \sigma_N(k)) \quad (18)$$

корректируется следующим образом:

$$\begin{aligned} w(k) &= w(k-1) + K(k)e(k); \\ K(k) &= P(k-1)\nabla_w \hat{y}(k) \times \\ &\times [\lambda I + \nabla_w^T \hat{y}(k)P(k-1)\nabla_w \hat{y}(k)]^{-1}; \\ P(k) &= \lambda^{-1} (P(k-1) + K(k)\nabla_w^T \hat{y}(k)P(k-1)), \end{aligned} \quad (19)$$

где

$$\begin{aligned} \nabla_w \hat{y}(k) &= [I, \Phi_1(x(k)), \Phi_1(x(k))\frac{2c_1}{\sigma_1^2}(x(k)-\mu_1)^T, \\ &\Phi_1(x(k))\frac{2c_1}{\sigma_1^2}\|x(k)-\mu_1\|^2, \dots, \Phi_N(x(k)), \\ &\Phi_N(x(k))\frac{2c_N}{\sigma_N^2}(x(k)-\mu_N)^T, \\ &\Phi_N(x(k))\frac{2c_N}{\sigma_N^2}\|x(k)-\mu_N\|^2]^T; \end{aligned}$$

I — единичная матрица; $\lambda \in [0, 1]$.

Несмотря на огромное число работ, в которых используется алгоритм (19), общих рекомендаций по выбору оптимального значения коэффициента λ в настоящее время, к сожалению, не существует, поэтому при решении практических задач чаще всего ограничиваются выбором $\lambda = 0,995 \div 0,999$.

4. Экспериментальные исследования

Сравнительный анализ аппроксимирующих свойств модифицированных СМАС и РБС производился на ПК с процессором Intel Pentium 4 3.2 GHz в среде MatLab 7.0 под ОС Linux 2.6.20.

Для оценки эффективности работы сетей использовались:

1. Время вычислений в секундах — T ;
2. Средняя квадратичная ошибка — MSE .

$$MSE = \sqrt{\frac{\sum_{i=1}^M (y_i(x) - \hat{y}_i(x))^2}{M}}$$

где M — количество экспериментов.

Рассматривалась задача восстановления зашумленной нелинейной четырехмерной функции $F(x_1, x_2, x_3, x_4)$:

$$F(x_1, x_2, x_3, x_4) = x_1 + \sin(\pi x_1) \cdot \cos(\pi x_2) \times \sin(\pi x_3) \cdot (\sin(\pi x_4)^2 - 1). \quad (20)$$

Сечение данной функции при $x_2 = x_3 = 0,25$, т. е. $F(x_1, 0,25, 0,25, x_4)$, показано на рис. 3.

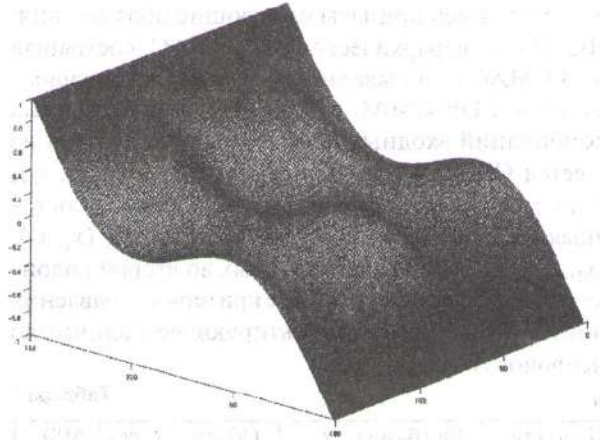


Рис. 3. Вид исходной функции $F(x_1, 0,25, 0,25, x_4)$

На данную функцию был наложен равномерно распределенный шум с амплитудой 0,1, в результате чего она приобрела вид, приведенный на рис. 4.

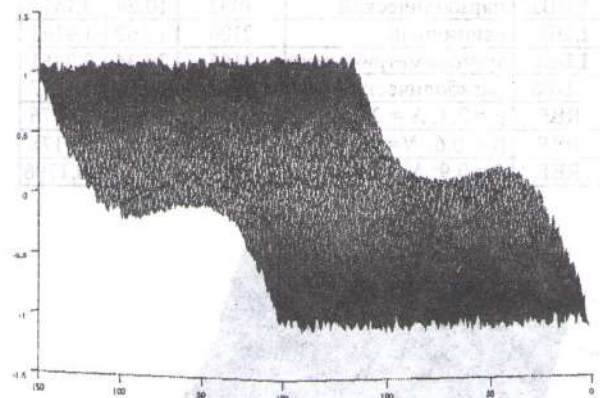


Рис. 4. Вид зашумленной функции $F(x_1, 0,25, 0,25, x_4)$

Обучение сетей осуществлялось на выборке из 20000 случайных точек, равномерно распределенных в интервале $[-1; 1]$. Целью эксперимента было исследование влияния различных характеристик сетей на их восстанавливающие свойства. В качестве моделируемой использовалась окрестность зашумленной функции $F(x_1, 0,25, 0,25, x_4)$ $F(x_1, 0 \dots 0,5, 0 \dots 0,5, x_4)$.

Все исследуемые базовые сети СМАС использовали $R = 50$ (уровней квантования), распределен-

ных по $\rho = 10$ ступеням квантования. Базовая двумерная сеть СМАС с такими характеристиками требует 350 ячеек памяти. Для восстановления данной функции обычной четырехмерной СМАС потребовалось бы 12119 ячеек памяти.

При исследовании РБС базисные функции выбирались гауссовскими (14), параметры сети настраивались по алгоритму (19) с $\lambda = 0,99$. Критерии введения нового нейрона имели вид (16), (17) с $\alpha_i = 0,01$ и различными значениями $\beta = 0,3; 0,6; 0,9$.

Результаты вычислений приведены в табл. 1 и на рис. 5–9. Здесь приняты следующие обозначения: НСМАС — иерархическая сеть СМАС, состоящая из 3 СМАС (1 в выходном и 2 в скрытом слоях); LDB6 — LDB СМАС с полным набором парных комбинаций входных переменных, состоящая из 6 сетей СМАС с входами $(x_1, x_2), (x_1, x_3), (x_1, x_4), (x_2, x_3), (x_2, x_4), (x_3, x_4)$; LDB3 — LDB СМАС, состоящая из 3 сетей СМАС с входами $(x_1, x_2), (x_1, x_4), (x_3, x_4)$; RBF — РБС, для которых во второй колонке указано заданное значение критерия добавления нового нейрона β и результирующее количество нейронов N .

Таблица 1

Тип сети	Вид базисных функций	Объем памяти, ячеек	T, сек	MSE, $\cdot 10^{-3}$
НСМАС	тригонометрический	1050	9,88	2,4034
НСМАС	параболический	1050	9,42	2,5923
LDB3	единичный	1053	10,12	4,8339
LDB3	тригонометрический	1053	10,43	3,9295
LDB3	параболический	1053	10,24	3,1513
LDB6	единичный	2106	19,62	1,9143
LDB6	тригонометрический	2106	20,34	2,2461
LDB6	параболический	2106	20,11	2,1517
RBF	$\beta = 0,3; N = 221$	1326	368,48	2,4376
RBF	$\beta = 0,6; N = 46$	276	85,65	7,0176
RBF	$\beta = 0,9; N = 16$	96	45,94	10,1796

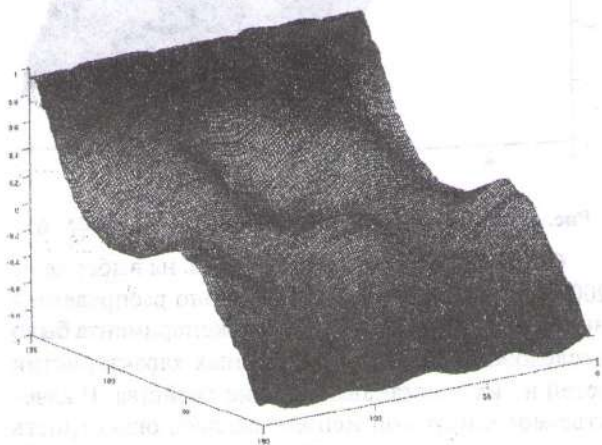


Рис. 5. Восстановление НСМАС с тригонометрическими БФ

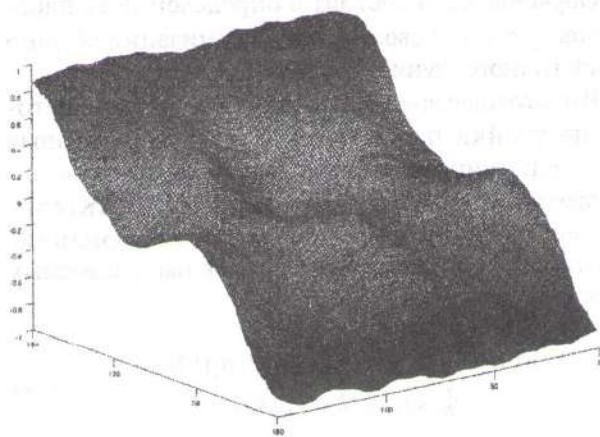


Рис. 6. Восстановление LDB3 СМАС с параболическими БФ

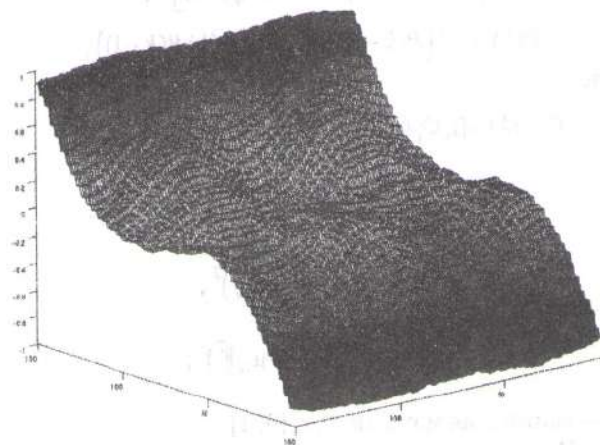


Рис. 7. Восстановление LDB6 СМАС с единичными БФ

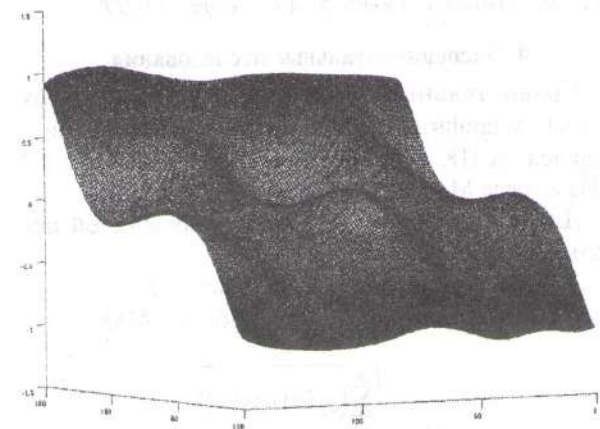


Рис. 8. Восстановление РБС с $\beta = 0,3 (N = 221)$

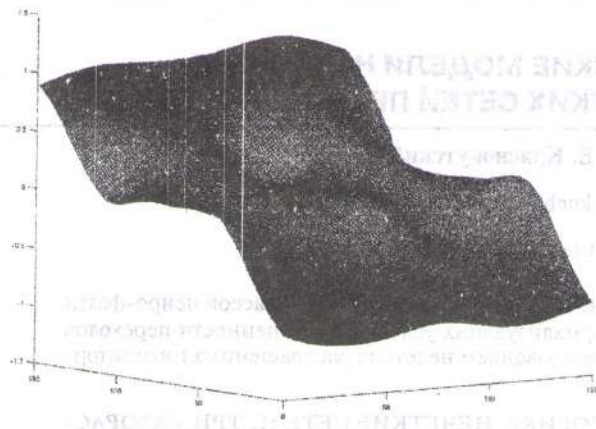


Рис. 9. Восстановление РБС с $\beta = 0,6$ ($N = 46$)

Выводы

Результаты исследований свидетельствуют о том, что восстановление многомерных функций может быть достаточно эффективно осуществлено путем применения сетей СМАС модифицированной архитектуры (иерархической НСМАС или линейной LDB СМАС, образованной сетями малой размерности) либо РБС.

В случае применения НСМАС используемые БФ обязательно должны быть дифференцируемыми, в то время как LDB СМАС позволяет применять произвольные БФ. Среди дифференцируемых наиболее простой и в то же время весьма эффективной является параболическая функция. Однако во всех случаях удается существенно сократить объем требуемой памяти сети СМАС и добиться приемлемого качества восстановления многомерной нелиней-

ной функции. Для рассмотренного примера объем требуемой СМАС памяти удалось сократить примерно в 5–11 раз.

Применение РБС обеспечивает заданную точность восстановления, требуя меньшего объема памяти, но значительно больших вычислительных затрат. Повышение точности восстановления, достигаемое путем введения новых нейронов, сопровождается резким возрастанием времени обучения.

Список литературы: 1. Хайкин С. Нейронные сети: полный курс, 2-е издание. — М.: Изд. дом «Вильямс», 2006. — 1104 с. 2. Руденко О. Г., Бодянский Е. В. Искусственные нейронные сети. — Харьков: Компания «СМИТ», 2005. — 408 с. 3. *Albus J.S.* A new approach to manipulator control: the cerebellar model articulation controller (CMAC) // *ASME Trans., J. Dynamic Systems, Measurement and Control.* — 1975. — 97. — №3. — P. 220–227. 4. Руденко О. Г., Бессонов А. А. О выборе базисных функций в нейронной сети СМАС // *Проблемы управления и информатики.* — 2004. — № 2. — С. 143–155. 5. Руденко О. Г., Островерхий А. В., Островерхая Н. Н. Аппроксимация многомерных функций с помощью нейронной сети СМАС // *Бионика интеллекта: Научн.-техн. журнал.* — 2006. — № 2 (65). — С. 8–13. 6. *Chiang Ch.-T., Lin Ch.-Sh.* CMAC with General Basis Functions // *Neural Networks.* — 1996. — 9. — №7. — P. 1199–1211. 7. *Lane S.H., Handelman D.A., Gelfand J.J.* Theory and development of higher-order CMAC neural networks // *IEEE Control Systems.* — 1992. — 12. — №2. — P. 23–30. 8. *Lee H.-M., Chen Ch.-M., Lu Yu.-F.* A Self-organizing HCMAC Neural Network Classifier // *IEEE Transactions on Neural Networks.* — 2003. — 14. — № 1. — P. 15–26. 9. *Lin Ch.-Sh., Li Ch.-K.* A Low-Dimensional-CMAC-Based Neural Network // *IEEE Int. Conf. on Neural Networks.* — 1996. — P. 1297–1302.

Поступила в редколлегию 28.02.07

УДК 519.71

ГИБРИДНЫЕ МАТЕМАТИЧЕСКИЕ МОДЕЛИ НА ОСНОВЕ РАСШИРЕНИЙ НЕЧЕТКИХ СЕТЕЙ ПЕТРИ

Е. И. Кучеренко¹, Д. Е. Краснокутский²¹ХНУРЭ, Харьков, Украина, kucherenko@kture.kharkov.ua²ХНУРЭ, Харьков, Украина, kuzzman@gmail.com

Сформулирована и решена задача расширения нечетких моделей на основе классов нейро-фаззи сетей Петри. Предложен комплекс утверждений, формализующих условия разрешенности переходов в условиях синхронизации нечетких процессов с использованием нечетких раскрашенных ингибиторных дуг сетей Петри.

МАТЕМАТИЧЕСКИЕ МОДЕЛИ, НЕЧЕТКАЯ ЛОГИКА, НЕЧЕТКИЕ СЕТИ ПЕТРИ, ГАЗОРАС-ПРЕДЕЛИТЕЛЬНЫЕ СИСТЕМЫ.

Введение

Ряд важных объектов промышленного назначения функционирует в условиях априорной неопределенности, характеризуется нечетким пространством состояний, что требует новых интеллектуальных подходов для повышения достоверности принимаемых решений. Следует также отметить, что современные объекты часто характеризуются функциональной и территориальной распределенностью [1], сложной иерархией взаимодействия их компонент [2]. Это ставит достаточно высокие требования к математическому аппарату и эффективности методов моделирования и анализа процессов предметной области.

Существующие подходы к построению и анализу систем управления и обработки данных в таких условиях [3, 4] обычно малоэффективны из-за их функциональной ограниченности, а в ряде случаев они не дают практических подходов к решению поставленных задач. Перспективным аппаратом построения систем является применение нечеткой логики Л. Заде, а для моделирования и анализа сложных процессов при решении комплекса поставленных задач целесообразным является использование аппарата сетей Петри (СП) [5] и его расширений [6, 7, 8].

Однако ординарные СП, возможные их расширения при решении прикладных задач обычно не учитывают или учитывают не в полной мере такие важные характеристики процессов, как нечеткость представления и обработки данных [5], сложное взаимодействие и динамику процессов в реальных системах [6]. Требуется также учет ряда дополнительных факторов для адекватного отображения предметной области. К ним, в первую очередь, следует отнести параметры времени, сложности, надежности, стоимости и другие, отображающие конкретные предметные области.

Существующие решения на основе интегрированных (СП (ИСП) [7] являются наиболее совершенными, но и они требуют новых решений. Поэтому работа является актуальной. В связи с этим

целесообразно рассмотреть возможность создания эффективных моделей, построенных на идеях и принципах нечеткой логики, теориях распределенных и иерархических систем и моделей.

Целью данной работы является повышение эффективности систем вычислительного интеллекта на основе разработки и внедрения математических моделей анализа и модификации нечетких динамических процессов, отображающих предметные области.

1. Постановка задачи

Пусть существует множество нечетких процессов $\{PR_i^{(S)}\}$, $i \in I$. (1)

Процессы (1) при целенаправленном взаимодействии на множестве декартова произведения

$$\{PR_k^{(S)}\} \times \{PR_l^{(S)}\} \times \dots \times \{PR_u^{(S)}\},$$

$$k \in K, l \in L, u \in U, K \subset I, L \subset I, U \subset I, \quad (2)$$

причем в общем случае для (2) справедливо

$$K \cap L \cap U = \emptyset,$$

реализуют некоторый комплекс взаимодействующих задач T_s . Задачи T_s определены согласно логики функционирования сложного объекта.

Процессы (1) и их взаимодействие (2) могут быть синхронизированы, развиваться во времени t , могут быть определены в детерминированном, вероятностном, нечетком пространствах состояний.

Необходимо предложить подходы к построению и реализации таких моделей, адекватно отображающих процессы (1) и взаимодействие (2) реальных объектов и систем.

2. Построение модифицированных математических моделей с использованием нечеткой логики

Как показал предварительный анализ, перспективным и целесообразным для решения поставленной задачи является применение нечетких СП и их модификаций.

Учитывая изложенное выше, существенные ограничения и недостатки известных моделей, в исследованиях [7] предложены новые решения и подходы к дальнейшему развитию идей по построению математических моделей. Рассмотрены классы нечетких сетевых моделей (НСМ), которые интегрируют преимущества моделей на основе СП и нейро-фаззи сетей (НФС). Они во многом свободны от недостатков и ограничений образующих их моделей.

Математическую НСМ представим как:

$$\tilde{S}(f) = \langle \tilde{P}, \tilde{T}, \tilde{F}(f), \tilde{M}(f)_0, L \rangle, \quad (3)$$

где $\tilde{P} = \{\tilde{p}_j : \mu_{\tilde{p}_j}(k)\}$ — конечное множество нечетких позиций \tilde{p}_j , $\mu_{\tilde{p}_j}(k)$ — функция принадлежности j -й нечеткой позиции множеству \tilde{P} , k — некоторая переменная, определяющая аргумент функции $\mu_{\tilde{p}_j}(k)$, $j = \overline{1, m}$, $\tilde{P} \neq \emptyset$, $|\tilde{P}| = m$; $\tilde{T} = \{\tilde{t}_i : \mu_{\tilde{t}_i}(k)\}$ — конечное множество нечетких переходов \tilde{t}_i , $i = \overline{1, n}$, $\tilde{T} \neq \emptyset$, $|\tilde{T}| = n$; $\mu_{\tilde{t}_i}(k)$ — функция принадлежности i -го нечеткого перехода, $\tilde{F}(f) : (\tilde{P} \times \tilde{T}) \cup (\tilde{T} \times \tilde{P}) \rightarrow \{x_{ij}(k), y_{ij}(k)\}$ — нечеткая функция инцидентностей \tilde{P} и \tilde{T} , $x_{ij}(k)$, $y_{ij}(k)$ — функции принадлежности входной и выходной инцидентностей некоторых нечетких позиций $\tilde{p}_j \in \tilde{P}$ и нечетких переходов $\tilde{t}_i \in \tilde{T}$.

Начальное пространство состояний определяется вектором нечеткого начального маркирования $\tilde{M}(f)_0$ нечетких позиций \tilde{P} модели. Тогда

$$\tilde{M}(f)_0 = \{\tilde{M}(F) : z_{\tilde{p}_j}(k),$$

$\tilde{M}(\tilde{p}_j) \rightarrow [0, 1]$ — нечеткое маркирование нечеткой позиции $\tilde{p}_j \in \tilde{P}$ НСМ, $z_{\tilde{p}_j}(k)$ — функция принадлежности маркирования j -ой нечеткой позиции $\tilde{p}_j \in \tilde{P}$; L — некоторый предикат от множества переменных $\{x_u\}$, $u \in U$.

С увеличением размерности моделируемой системы эффективность подходов построения НСМ уменьшается из-за большой размерности модели и необходимости учета множества дополнительных параметров, характеристик, условий $\{x_u\}$, $u \in U$.

Действенным средством снижения размерности сетевых моделей можно считать раскрашенные нечеткие сети Петри.

В работе [7] введено понятие сетей, которые представлены следующим образом:

$$\tilde{S}_C(f) = \langle \tilde{P}, \tilde{T}, \tilde{F}_C(f), \tilde{M}_{oc}(f), \tilde{M}_C(f), L\{x_u\}, \tilde{C}, \tilde{V}, \tilde{K} \rangle, \quad (4)$$

где \tilde{P} — множество нечетких позиций; \tilde{T} — множество нечетких переходов; $L\{x_u\}$, $u \in U$ — некоторый предикат, отнесенный на модели к множеству позиций, переходов, функции инцидентности в

пространстве состояний нечетких взаимодействующих процессов и определяющий дополнительные условия выполнения переходов;

$$\tilde{F}_C(f) = (\tilde{P} \times \tilde{T}) \cup (\tilde{T} \times \tilde{P}) \quad (5)$$

нечеткая функция инцидентностей сети $\tilde{S}_C(f)$; \tilde{C} — функция цвета маркера, определяющая в данном случае цвет C каждого из маркеров $\tilde{M}(\tilde{p}_j)$ для позиций сети; \tilde{V} — условия срабатывания переходов в зависимости от цвета маркера; \tilde{K} — емкость маркеров в позициях с учетом \tilde{C} ; $\tilde{M}_C(f)_0$ — вектор начальной маркировки; $\tilde{M}_C(f)$ — вектор текущей маркировки.

Как следует из (4), сеть интегрирует сеть (3) и преимущества раскрашенных СП [5]. Ввод предиката $L\{x_u\}$, $u \in U$ в модель (4), а также ввод свойств \tilde{C} , \tilde{V} , \tilde{K} существенно увеличивает возможности модели по сравнению с существующими подходами. Все это позволяет нам рассматривать (4) как существенное расширение существующих нечетких моделей.

Однако в сетях (4) отсутствуют средства синхронизации выполнения переходов из \tilde{T} , их применение затруднено при моделировании процессов, для которых актуальна их синхронизация. В связи с этим введем в функцию инцидентности сети раскрашенные ингибиторные дуги. Существующие подходы к применению ингибиторных дуг [5] основаны на множестве $\{0, 1\}$, что в реальных системах в нечетком пространстве состояний плохо применимо.

Утверждение 1. Если задана сеть ИСП₂, то для синхронизации выполнения некоторых нечетких переходов

$$\exists t_i(sn) \in \tilde{T} \quad (6)$$

во входную компоненту $In = \tilde{P} \times \tilde{T}$ ее функции инцидентности $\tilde{F}(f) = (\tilde{P} \times \tilde{T}) \cup (\tilde{T} \times \tilde{P})$ достаточно ввести нечеткие ингибиторные (запрещающие) дуги

$$Arc(sn)_\alpha \in \{Arc_p\}, \alpha \in A, \beta \in B, A \subseteq B. \quad (7)$$

Справедливость утверждения 1 непосредственно следует из определения ингибиторной дуги [5]. Введение же свойств цвета и нечеткости расширяет функцию ингибиторной дуги (7).

Пусть существует некоторый разрешенный переход (6) сети (4) [7]. Определим для него дополнительные условия разрешенности с учетом ввода дуги (7).

Утверждение 2. Некоторый переход (6) разрешен, если для него дополнительно существует входная дуга (7) от инцидентной входной позиции $p_j \in \{p_j(in)\}$ и справедливо

$$\begin{aligned} \exists t_i(sn) \in \tilde{T} | (((\tilde{M}_C(f)_{p_j \in \tilde{P}_i(m)}) = 0) \& \\ \&(x_{y_j}(k) > x_{y_j}(k)^*)) \& (\tilde{V}_C(sn) = \text{true})), \end{aligned} \quad (8)$$

где $\tilde{V}_C(sn)$ — условия срабатывания переходов в зависимости от цвета маркера.

Справедливость утверждения 2 следует из условий функционирования входной ингибиторной дуги данного перехода с учетом ввода дополнительно функции цвета и допустимого нижнего порогового значения функции принадлежности $x_{y_j}(k)^*$.

Утверждение 3. Если существует модель (4) и для некоторых переходов (6) введены дополнительно нечеткие ингибиторные дуги (6), то ее функция инцидентности может быть представлена, как

$$\tilde{F}_C^M = \tilde{F}_C(f) \cup \tilde{F}_C(Arc), \quad (9)$$

где $\tilde{F}_C(Arc)$ — входная инцидентность некоторого перехода (6), определяемая ингибиторной дугой (7).

Справедливость утверждения 3 определяется сущностью функции инцидентности (5) и предлагаемой в работе конструкции на основе нечетких ингибиторных дуг (7).

Тогда предлагаемая в работе модель может быть представлена в виде:

$$\begin{aligned} \tilde{S}_C^{(M)}(f) = \langle \tilde{P}, \tilde{T}, \tilde{F}_C^{(M)}(f), \tilde{M}_{0C}(f), \\ \tilde{M}_C(f), L(x_u), \tilde{C}, \tilde{V}, \tilde{K} \rangle. \end{aligned} \quad (10)$$

Для отображения и формализации процессов на модели в работе [7] сформулированы следующие правила интерпретации компонент модели:

- множество нечетких переходов $\tilde{t}_i \in \tilde{T}$ НСМ $\tilde{S}(f)$ интерпретирует множество нечетких действий $\{\tilde{d}_r\}, r \in R$ моделируемых нечетких процессов;
- множество нечетких позиций $\tilde{p}_j \in \tilde{P}$ НСМ $\tilde{S}(f)$ интерпретирует множество нечетких условий $\tilde{U}_l, l \in L$ выполнения множества нечетких действий $\{\tilde{d}_r\}$;
- динамика моделируемых процессов интерпретируется перемещением нечетких меток на множестве нечетких позиций $\{\tilde{p}_j\}$ через множество разрешенных нечетких им инцидентных переходов;
- функция цвета маркера \tilde{C} , определяющая в данном случае цвет c каждого из маркеров $\tilde{M}(\tilde{p}_j)$ для позиций сети, интерпретирует множество признаков маркера, отображающего некоторую физическую сущность объекта моделирования;
- пространство состояний $\{\tilde{A}_j\}, j \in J$ динамических взаимодействующих нечетких процессов интерпретируется множеством векторов маркирования $\{\tilde{M}(f)_s\}, s \in S$ множества позиций $\{\tilde{p}_j\}$ в пространстве состояний НСМ.

Замечание. НСМ, построенные на расширении нечетких СП $\tilde{S}_C^{(M)}(f)$, обладая значительными функциональными возможностями моделирования, не всегда позволяют эффективно исследовать также сложные процессы типа временных рядов, решать задачи идентификации, прогнозирования для непрерывных и дискретных процессов. Этими свойствами обладают искусственные нейронные сети и их расширения — нейро-фаззи сети (НФС) [7, 9, 10].

В связи с этим целесообразно рассмотреть возможность применения НФС в составе расширений НСМ.

3. Развитие НСМ на основе интеграции расширений нечетких сетей Петри и нейро-фаззи сетей

В задачах моделирования сложных интеллектуальных систем большой размерности моделями на основе сетей (10) может также возникнуть проблема, связанная с недопустимо большой размерностью модели. Это снижает ее преимущества и делает слабо адаптируемой к объекту моделирования.

Таким образом, как уже отмечалось выше, для снижения размерности модели удобно применить модели на основе интеграции сетей класса (10) и НФС. Такой подход концентрирует преимущества и достоинства образующих моделей, а также снижает размерность и расширяет их возможности [7].

Утверждение 4. Пусть существует расширенная модель $\tilde{S}_C^{(M)}(f)$ и модель на основе НФС $\tilde{S}_N(f)$, и интеграция этих моделей

$$\tilde{S}^{(Int)}(f) = \tilde{S}_C^{(M)}(f) \cup \tilde{S}_N(f) \quad (11)$$

может быть представлена в виде реализации следующих действий:

— для некоторых переходов $\tilde{t}_i \in \tilde{T}$ введены дополнительно процедуры

$$\tilde{P}r_{ki} \in \{\tilde{P}r_k\}, k \in K, i \in I; \quad (12)$$

— для некоторых позиций $\tilde{p}_j \in \tilde{P}$ введены дополнительно процедуры

$$\tilde{P}r_{kj} \in \{\tilde{P}r_k\}, k \in K, j \in J; \quad (13)$$

— для некоторых маркировок позиций $\tilde{M}_{C\alpha}(\tilde{p}_j) \in \tilde{M}_C(f)$ с учетом цвета маркера введены дополнительно процедуры

$$\tilde{P}r_{k\tilde{M}(\tilde{p}_j)} \in \{\tilde{P}r_k\}, k \in K, j \in J, \quad (14)$$

причем процедуры (12), (13), (14) реализуются моделями на основе сетей $\tilde{S}_N(f)$.

Справедливость утверждения 4 непосредственно следует из определения моделей на основе сетей $\tilde{S}_C^{(M)}(f)$ и возможности реализации процедур $\tilde{P}r_k, k \in K$ моделями на основе сетей $\tilde{S}_N(f)$.

Результаты выполнения процедур (12), (13), (14) определяют дополнительные условия разрешенности (8) переходов $\tilde{l}_i \in \tilde{T}$ и маркирования позиций $\tilde{p}_j \in \tilde{P}$ с учетом цвета маркирования $\tilde{M}_{C_\alpha}(\tilde{p}_j) \in \tilde{M}_C(f)$ при моделировании сложных процессов моделями на основе сетей $\tilde{S}_N(f)$. Обозначив разрешенность некоторого перехода сети $\tilde{S}_C^{(M)}(f)$ с учетом (8) как $\tilde{R}_C^{(M)}(f)$, сформулируем следующее утверждение.

Утверждение 5. Если существует сеть $\tilde{S}^{(Int)}(f)$ (11), то разрешенность ее переходов может быть представлена в виде:

$$\tilde{R}_C^{(Int)}(f) = (R_C^{(M)}(f) = \text{true}) \& \& (\tilde{P}r_k \in \{\tilde{P}r_k\} | \tilde{P}r_k = \text{true}). \quad (15)$$

Справедливость (15) непосредственно следует из условия разрешенности переходов [7], а также выражений (8), (11) и условия истинности выполнения процедур (12), (13), (14).

4. Приложения научных результатов

Важным направлением повышения уровня конкурентоспособности производственных объектов является контроль и оценка технического состояния ресурсов как в текущее время, так и в прогнозном будущем на основе интеллектуальных средств принятия решений.

Особенно это актуально в системах газотранспортных и газораспределительных магистралей регионов Украины. При этом необходимо учитывать как объективные, так и субъективные факторы, способные влиять на объект контроля. Важными составляющими являются также влияние метеорологических факторов, физико-химической составляющей среды функционирования трубопроводов, состав объекта транспортировки. Большинство существующих факторов плохо формализуемы, и их влияние может быть выполнено на основе нечетких оценок экспертов.

Для решения актуальных задач оценки состояния объекта в нечетких условиях авторами предложены подходы с использованием научных результатов данной работы. В практических реализациях реального объекта внедрение позволило, по оценкам экспертов, существенно повысить достоверность принимаемых решений по оценке состояния объекта, что подтверждено внедрениями.

Выводы

1. Сформулирована и решена задача расширения модели на основе классов нейро-фаззи сетей Петри, которая отличается от существующих подходов наличием средств синхронизации нечетких процессов с использованием впервые введенных нечетких раскрашенных ингибиторных дуг сетей Петри.

2. Предложен комплекс утверждений, формализующих условия разрешенности переходов в условиях синхронизации нечетких процессов на основе нечетких раскрашенных ингибиторных дуг сетей Петри, что позволяет адекватно отображать и моделировать реальные процессы предметных областей.

3. Внедрение научных результатов в задачах оперативного контроля состояния газотранспортных и газораспределительных магистралей региона подтвердило эффективность и достоверность предложенных подходов.

4. Перспективным направлением дальнейших исследований является адаптация разработок на иерархические и распределенные плохо формализуемые объекты различного назначения.

Список литературы: 1. Вейцман К. Распределенные системы мини- и микроЭВМ: Пер. с англ. / Под ред. Васильева Г.П. — М.: Финансы и статистика, 1983. — 362 с. 2. Месарович М., Мако Д., Такахара И. Теория иерархических многоуровневых систем. — М.: Мир, 1976. — 344 с. 3. Дмитриев А.К., Мальцев П.А. Основы теории построения и контроля сложных систем. — Л.: Энергоатомиздат. Ленингр. отд-ние, 1988. — 192 с. 4. Глушков В.М., Иванов В.В., Яненко В.М. Моделирование развивающихся систем. — М.: Наука. Главная редакция физ.-мат. литературы, 1983. — 350 с. 5. Murata T. Temporal uncertainty and fuzzy-timing high-level Petri nets // Proc. 17th Int. Conf. of Application and Theory of PNs, Osaka, Japan, June 26 / IEEE Computer Society Press. — Los Alamitos, CA, 1996. — P. 11–28. 6. Pedrycz W., Gomide F. A generalized fuzzy Petri net model // IEEE Trans. on Fuzzy System. — 1994. — 2. — № 4. — P. 295–301. 7. Бодянский Е.В., Кучеренко Е.И., Михалев А.И. Нейро-фаззи сети Петри в задачах моделирования сложных систем / Монография — наукове видання. — Дніпропетровськ: Системні технології. — 311 с. 8. Koriem S.M. A fuzzy Petri net tool for modeling and verification of knowledge — Based Systems // The Computer Journal. — 2000. — 43. — № 3. — P. 206–223. 9. Руденко О.Г., Бодянский Е.В. Основы теории искусственных нейронных сетей. — Харьков: Телетех, 2002. — 317 с. 10. Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы: Пер. с польск. И.Д. Русинского. — М.: Горячая линия — Телеком, 2004. — 452 с.

Поступила в редколлегию 02.03.07

УДК 004.93:007.52

МЕТОД ОПРЕДЕЛЕНИЯ ВЗАИМОСВЯЗИ ЭЛЕМЕНТОВ СТОХАСТИЧЕСКОЙ НЕЙРОННОЙ СЕТИ

А.В. Волик¹, С.Н. Герасин², С.И. Лапта³¹ ХНУРЭ, г. Харьков, Украина² ХНУРЭ, г. Харьков, Украина, sgerasin@mail.ru³ ХНУРЭ, г. Харьков, Украина, sergey_lapta@ukr.net

Предложен метод для определения взаимосвязи между элементами нейронной сети. Он допускает нестационарность входного потока и позволяет учитывать нелинейный характер связей между элементами нейронной сети. Метод может быть использован при анализе активности сети и дает возможность оценивать динамические характеристики нейронной сети.

**СЕТЬ НЕЙРОННАЯ СТОХАСТИЧЕСКАЯ, ДИНАМИКА АКТИВНОСТИ, ВЗАИМОСВЯЗЬ
МЕЖДУ ЭЛЕМЕНТАМИ СЕТИ.**

Введение

В настоящее время происходит существенное повышение интереса к искусственному интеллекту, вызванного как развитием технических средств, так и потребностью рынка программного обеспечения в качестве нового продукта. На фоне этого процесса, а вернее, как неотъемлемая часть его, производятся многочисленные попытки применения тех или иных моделей нейронных сетей для решения различных задач. Математический аппарат, позволяющий моделировать работу нейронных сетей, был разработан достаточно давно, но широкое практическое применение его для решения прикладных задач началось сравнительно недавно. В то же время исследование динамики нейронной сети как единого целого и присущие ей эффекты изучены не в полном объеме. Целью статьи является анализ возможных подходов к решению данной задачи с позиций случайного анализа.

1. Анализ взаимосвязи элементов стохастической нейронной сети на основе разложения в ряд Вольтерра

Рассмотрим основные положения анализа нелинейных систем на основе вольтерровского разложения. При этом мы будем говорить об идентификации нелинейной системы в терминах принятого разложения. Обычная постановка задачи, которая при этом рассматривается, состоит в следующем. Имеется нелинейная система T , принцип устройства которой неизвестен, и потому можно представлять себе эту систему в виде «черного ящика». Система имеет вход (один или несколько), на который подается некоторый случайный процесс $x(t)$, и выход $y(t)$, который также является случайным процессом, зависящим от входа $y(t) = T[x(t)]$. Задача идентификации — определить детерминированный закон $T[x(t)]$ на основе известных процессов на

входе и выходе. Для этого неизвестный функционал T представляется в виде ряда, в который входят неизвестные функции (ядра). При некоторых предположениях на рассматриваемую систему можно определить эти функции и тем самым идентифицировать неизвестную структуру системы. При дальнейшем анализе можно предположить некоторую функциональную аппроксимацию (модель) и оценивать неизвестные параметры модели.

Метод практического вычисления ядер разложения Вольтерра (или Винера) был разработан М. Шетценом [1], он показал, что в случае, когда входной поток является белым гауссовским шумом, ядра разложения равны корреляционным функциям произведения входных функций и выходной функции.

Однако при идентификации биологических систем возникает необходимость рассматривать и другие входные процессы, отличные от белого шума. В работах [2–5] рассматриваются различные обобщения метода винеровской идентификации на случай различных входных процессов. Для приложений наиболее интересным является случай входного сигнала в форме точечного процесса, в частности, пуассоновский поток событий на входе системы. Причем метод, предложенный в работах [3, 4], пригоден как для непрерывного, так и для точечного выходного процесса, и легко обобщается на случай более общих точечных процессов на входе, например, можно рассматривать входную марковскую цепь событий.

При рассмотрении взаимодействия элементов нейронной сети с помощью метода разложения в ряд Вольтерра обычно рассматривают одну из трех задач.

1) Подавая на вход нейрона случайный поток стимулов и регистрируя импульсный поток на выходе, описать деятельность нейрона в терминах

ядер разложения. Решение этой задачи позволяет предсказывать поведение нейронов при воздействии на них различных входных потоков. Эти результаты можно использовать, например, для классификации нейронов.

2) Предположим, что входной поток на изучаемый нейрон поступал с выхода другого элемента сети. Тогда изучение зависимости входного и выходного потоков позволяет анализировать взаимодействие элементов сети. Разлагая функцию условной кроссинтенсивности в ряд Вольтерра и определяя ядра разложения, можно полностью охарактеризовать взаимосвязь между элементами. Такой подход является более общим, чем кросскорреляционный, поскольку он позволяет учитывать нелинейные эффекты взаимодействия.

3) Для описания нейрона выбирается простая модель, удобная для работы с разложением в ряд. Тогда результат анализа зависимости между входным и выходным потоками, формулируемый в терминах разложения, можно интерпретировать в терминах выбранной модели и тем самым получить идентификацию таких параметров, как эффективный порог, функция затухания и т. д.

Подчеркнем, что метод разложения в ряд в задаче оценки зависимости элементов сети является обобщением метода вычисления кросскорреляционной функции на случай нелинейных систем. Кросскорреляция входного и выходного потоков пропорциональна ядру первого порядка в разложении Вольтерра и в этом смысле является линейной мерой зависимости. Вычисление ядер второго, третьего, более высоких порядков уточняет структуру взаимодействия элементов и позволяет правильно описывать нелинейные эффекты. При построении оценок ядра первого порядка (пропорционального кросскорреляции) рекомендуется проводить вычисления в частотной области (перейти к преобразованию Фурье), так как при этом удается оценивать статистическую значимость получаемых оценок [5].

Следует заметить, что применение метода разложения требует выполнимости целого ряда условий и накладывает ограничения на систему, рассматриваемые потоки, выбираемую модель и т. д. Некритическое отношение к этим требованиям может приводить к неверным результатам [6]. Кроме того, подчеркнем, что метод изложения позволяет определить структуру неизвестной системы лишь в терминах ядер разложения. При этом остаются неизвестными биофизические механизмы, лежащие в основе функционирования системы. При оценивании ядер второго и более высоких порядков требуются на порядок более длинные выборки, при этом тесты значимости отсутствуют.

2. Метод определения зависимости между элементами нейронной сети, основанный на статистическом анализе зависимости точечных процессов

Известные статистические методы по тем или иным причинам непригодны для решения сформулированных задач. Поэтому в статье разработан статистический метод для определения связей между элементами нейронной сети. Метод специально приспособлен для анализа нейронной импульсной активности и учитывает специфику генерации импульсов элементами сети. Этот метод является, в некотором математическом смысле, наилучшим из возможных методов, поскольку статистические оценки этого метода получаются при максимизации условной функции правдоподобия. Он свободен от большинства недостатков, присущих другим методам, в частности, не предполагается стационарность исследуемых потоков, имеются возможности для учета нелинейности связей, легко обобщается на случай трех и более потоков.

В основе данного метода лежит статистический анализ зависимости точечных процессов, предложенный Д. Коксом [7]. Рассмотрим точечный процесс, подверженный внешнему воздействию. Основное предположение Кокса состоит в том, что риск срабатывания точечного процесса является произведением собственного риска (риска срабатывания при отсутствии воздействия) и функции, зависящей от воздействия:

$$\varphi(t) = \lambda_0(t) \exp(\beta z(t)).$$

Оцениваемый параметр β характеризует степень воздействия, а функция $z(t)$ отражает структуру воздействия, $\lambda_0(t)$ — собственный риск. В частности, $\beta = 0$ означает отсутствие воздействия. Д. Кокс [8] предложил способ оценивания параметра β (вообще говоря, многомерного в случае многих воздействий), исходя из максимума условной функции правдоподобия, что позволяет получить хорошие асимптотические свойства оценки.

Пусть M — точечный процесс, вообще говоря, нестационарный и пусть в момент t^* произошло событие в этом процессе. Функция $\varphi^{**}(t)$ называется функцией риска процесса M , если

$$\varphi^{**}(t) = \lim_{\Delta t \rightarrow 0} \frac{P_r \left\{ M(t, t + \Delta t) = 1 \left| \frac{M\{t^*\} = 1}{M(t^*, t) = 0} \right. \right\}}{\Delta t},$$

то есть $\varphi^{**}(t)$ — условная плотность вероятности срабатывания при условии, что в момент t^* было событие и от t^* до t срабатываний не происходило.

Очевидно, что для пуассоновского процесса

$$\varphi^{**}(t) = \varphi(t) = P_M \beta,$$

где P_M — интенсивность пуассоновского процесса, $u(t)$ — обратное время возвращения.

Аналогично, для процесса восстановления риск в момент t зависит лишь от обратного времени возвращения:

$$\varphi^{**}(t) = \varphi(u(t)) = \lambda_0(u(t)).$$

Предположим, что для нестационарного точечного процесса M функция риска $\varphi^{**}(t)$ имеет следующий вид:

$$\varphi^{**}(t) = \lambda_0(u(t))e^{\beta_1 Z_1(t) + \beta_2 Z_2(t) + \dots + \beta_k Z_k(t)}, \quad (1)$$

где $Z_1(t), Z_2(t), \dots, Z_k(t)$ — заданные функции; $\beta_1, \beta_2, \dots, \beta_k$ — неизвестные параметры.

Процесс M , удовлетворяющий этому условию, называют модулированным процессом восстановления, поскольку собственный риск процесса $\lambda_0(u(t))$ как бы модулируется воздействующим вектором $Z_1(t), Z_2(t), \dots, Z_k(t)$. В частном случае $\lambda_0(u(t)) = P_M$ процесс M называется модулированным пуассоновским процессом.

Пример 1. Тренд. Пусть $K=1$ и $Z_1(t)=t$, тогда:

$$\varphi^{**}(t) = \lambda_0(u(t))e^{\beta t}.$$

При малых t риск срабатывания процесса примерно равен собственному риску $\lambda_0(u(t))e^{\beta t}$ и экспоненциально возрастает с постом t (т. е. срабатывания процесса становятся все более частыми при увеличении t).

Пример 2. Синусоидальные изменения. Пусть $k=1$, $Z_1(t) = \sin(\omega_0 t)$, тогда:

$$\varphi^{**} = \lambda_0(u(t))e^{\beta \sin(\omega_0 t)}.$$

В этом примере риск процесса изменяется синусоидально с известной частотой ω_0 .

Пример 3. Зависимость двух точечных процессов. Рассмотрим двумерный точечный процесс (A, B) . Обозначим $u_B(t)$ — обратное время возвращения для процесса B . Пусть $k=1$, $Z_1(t) = Y(u_B(t))$, тогда функция риска для процесса A будет иметь вид:

$$\varphi^{**}(t) = \lambda_0(u(t))e^{\beta Y(u_B(t))}.$$

Этот пример иллюстрирует простейшую возможность введения зависимости двух точечных процессов. Риск процесса A зависит от обратного времени возвращения в процессе B , которое предполагается известным и неслучайным. Можно показать, что для модельной нейронной сети риск срабатывания элемента A зависит от обратного времени возвращения $u_B(t)$ в процессе, генерируемом воздействующим на A элементом B , причем

$$Y(u_B(t)) = e^{-\frac{1}{2}u_B(t)}.$$

Значения неизвестных параметров β_i характеризуют зависимость основного процесса A от модулирующих воздействий. Если все β_i равны нулю, то процесс не зависит от воздействующей на него модуляции, и риск срабатывания равен собственному риску; если же величина β_i отлична от нуля, то она характеризует степень воздействия модулирующего процесса.

Перейдем к построению оценок для параметров β_i методом условной функции правдоподобия [7]. Рассмотрим сначала случай модулированного пуассоновского процесса, то есть $\varphi(t) = pe^{\beta Z(t)}$, где для простоты возьмем случай $K=1$.

Пусть на интервале наблюдения $(0, T]$ события процесса появились в моменты t_1, t_2, \dots, t_n . Тогда функция правдоподобия имеет вид:

$$L(p, \beta, t_1, t_2, \dots, t_n) = f^0(t_1)f^{t_1}(t_2 - t_1) \dots \\ \dots f^{t_{n-1}}(t_n - t_{n-1})(1 - F^{t_n}(T - t_n)),$$

где $F^{t^*}(x)$ — функция распределения длины интервала до очередного срабатывания (t^* — момент предыдущего срабатывания); $f^{t^*}(x)$ — плотность этого распределения: $f(x) = \frac{\partial}{\partial x} F^{t^*}(x)$.

Поскольку:

$$\varphi^{**}(t^* + x) = \frac{f^{t^*}(x)}{1 - F^{t^*}(x)}, \text{ то}$$

$$\frac{\partial}{\partial x} [\ln(1 - F^{t^*}(x))] = -\varphi^{**}(t^* + x),$$

и, решая это уравнение, находим:

$$F^{t^*}(x) = 1 - e^{-S_{t^*+x} \varphi^{**}(u) du},$$

$$f^{t^*}(x) = \frac{\partial}{\partial x} F^{t^*}(x) = \varphi^{**}(t^* + x) e^{-S_{t^*+x} \varphi^{**}(u) du}.$$

Учитывая предположение о виде функции $\varphi^{**}(t^* + x)$ и подставляя полученные соотношения в формулу для функции правдоподобия, получим:

$$\alpha(p, \beta, t_1, t_2, \dots, t_n) = p^n e^{\beta(Z(t_1) + Z(t_2) + \dots + Z(t_n))} \times \\ \times e^{-p \int_0^T e^{\beta Z(u)} du}.$$

В функцию правдоподобия, кроме параметра, который мы хотим оценить, входит неизвестный мешающий параметр p . Чтобы исключить его, воспользуемся тем, что число срабатываний процесса — n на интервале наблюдения $(0, T]$ является достаточной статистикой для p , и перейдем к условной функции правдоподобия:

$$L(p, \beta; t_1, t_2, \dots, t_n | n) = \frac{\alpha(p, \beta; t_1, t_2, \dots, t_n)}{P_r\{A(0, T] = n\}}.$$

Известно, что для стационарного пуассоновского процесса число событий на интервале $(0, T)$ имеет распределение Пуассона [4], поэтому:

$$P_r \{A(0, T) = n\} = \frac{p^n \left[\int_0^T e^{\beta Z(u)} du \right]^{nT} e^{-p \int_0^T e^{\beta Z(u)} du}}{n!}$$

Подставляя полученное выражение, логарифмируя и отбрасывая постоянные сомножители, получим:

$$L_p(\beta) = \beta S - n \ln \left[\int_0^T e^{\beta Z(t)} dt \right],$$

где $S = \sum_{i=1}^n Z(t_i)$.

Обозначим

$$U_p(\beta) = \frac{\partial L_p(\beta)}{\partial \beta} = S - n \frac{\int_0^T Z(t) e^{\beta Z(t)} dt}{\int_0^T e^{\beta Z(t)} dt} = S - n A_p(Z, \beta),$$

$$I_p(\beta) = -\frac{\partial^2 L_p(\beta)}{\partial \beta^2} =$$

$$= n \frac{\left(\int_0^T Z^2(t) e^{\beta Z(t)} dt \right) \left(\int_0^T e^{\beta Z(t)} dt \right) - \left(\int_0^T Z(t) e^{\beta Z(t)} dt \right)^2}{\left(\int_0^T e^{\beta Z(t)} dt \right)^2} =$$

$$= n \left[A_p(Z^2, \beta) - A_p^2(Z, \beta) \right],$$

где

$$A_p(Z, \beta) = \frac{\int_0^T Z(t) e^{\beta Z(t)} dt}{\int_0^T e^{\beta Z(t)} dt}.$$

Известно, что $U_p(\beta)$ имеет асимптотически нормальное распределение со средним ноль и дисперсией $I_p(\beta)$ [7]. Отсюда находим условия для вычисления $(1-\alpha)$ доверительного интервала параметра β :

$$\left\{ \beta : \left| S - n A_p(Z, \beta) \right| < K_{\alpha/2} \sqrt{I_p(\beta)} \right\},$$

где $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-k_\alpha} e^{-\frac{u^2}{2}} du = \alpha$.

С вычислительной точки зрения это условие не является простым. Можно рассчитать требуемую статистику $U_p(\beta)$ при различных β и, интерполируя, найти такие β , при которых выполняется неравенство.

Рассмотрим один частный случай, для которого полученная формула имеет простой вид. Предположим, что $Z(t) = u_p(t)$ и найдем критическое соотношение для проверки гипотезы о равенстве нулю параметра β . Для этого случая

$$A_p(Z, 0) = \frac{1}{2} \cdot \frac{\sum_{j=1}^n (x_j^B)^2}{T} = \frac{1}{2} m_2^B,$$

$$A_p(Z^2, 0) = \frac{1}{3} \cdot \frac{\sum_{j=1}^n (x_j^B)^3}{T} = \frac{1}{3} m_3^B,$$

и величина

$$\xi = \frac{\sum_{i=1}^n u_B(t_i) - \frac{1}{2} n m_2^B}{\sqrt{n \left[\frac{1}{3} m_3^B - \frac{1}{4} (m_2^B)^2 \right]}}$$

имеет нормальное распределение со средним 0 и дисперсией 1. Здесь мы обозначили через $x_1^B, x_2^B, \dots, x_n^B$ интервалы между моментами появления событий в процессе B , кроме того, для простоты предполагается, что интервал $(0, T)$ начинается и заканчивается событием процесса B .

Предполагаем, что эти моменты являются фиксированными, а не случайными, и анализируем зависимость процесса A от процесса B при условии известных зафиксированных моментов появления событий процесса B на интервале наблюдения $(0, T)$.

Рассмотрим теперь случай модулированного процесса восстановления A , для которого функция риска имеет вид:

$$\varphi^{**}(t) = \lambda_0(u(t)) e^{\beta Z(t)}.$$

Здесь предполагается, что $Z(t)$ — модулирующее воздействие — фиксированная (не случайная) функция текущего времени t , если $Z(t)$ связана с реализацией другого случайного процесса, то всю процедуру оценивания проводим при условии, что задана реализация воздействующего процесса и обрабатываемся с $Z(t)$ как с неслучайной величиной; $\lambda_0(u(t))$ — неизвестный собственный риск, который является мешающей функцией и должен быть исключен из окончательных оценок; β — неизвестный параметр, который требуется оценить (для простоты рассматривается одномерный случай).

Для оценивания β используется не безусловная функция правдоподобия, содержащая неизвестную мешающую функцию $\lambda_0(\cdot)$, а условная функция правдоподобия, при условии, что известны длины межимпульсных интервалов x_1, x_2, \dots, x_n в процессе A . Это условие позволяет исключить мешающую функцию и получить оценку для параметра β , обладающую хорошими асимптотическими свойствами.

Длины межимпульсных интервалов, упорядоченные по возрастанию, обозначим $x_{(1)}, x_{(2)}, \dots, x_{(n)}$, так что $x_{(1)} < x_{(2)} < \dots < x_{(n)}$, причем предполагается, что все интервалы различны. Обозначим Z_{ij} следующую величину: пусть $i \geq j$, тогда интервал $x(j) \leq x(i)$, величина Z_{ij} зависит от интервала $x_{(j)}$, но вычисляется в момент $x_{(i)}$, от начала интервала $x_{(j)}$. Другими словами: к началу большего интервала $x_{(i)}$ прикладываем меньший интервал $x_{(j)}$ и в момент времени, соответствующий концу интервала $x_{(j)}$, вычисляем величину Z_{ij} . Величина Z_{ij} вычисляется в момент, соответствующий концу интервала $x_{(j)}$.

Аналогично тому, как это было сделано для модулированного пуассоновского процесса, можно получить функцию правдоподобия для модулированного процесса восстановления:

$$\begin{aligned} & \alpha(\lambda_0(u(t)), \beta; x_1, x_2, \dots, x_n) = \\ & = \lambda_0(x_{(1)}) e^{\beta Z(x_{(1)})} e^{-\int_0^{x_{(1)}} \lambda_0(u) e^{\beta Z(u)} du} \times \\ & \times \lambda_0(x_{(2)}) e^{\beta Z(x_{(1)}+x_{(2)})} e^{-\int_0^{x_{(2)}} \lambda_0(u) e^{\beta Z(x_{(1)}+u)} du} \times \dots \\ & \dots \times \lambda_0(x_{(n)}) e^{\beta Z(x_{(1)}+x_{(2)}+\dots+x_{(n)})} e^{-\int_0^{x_{(n)}} \lambda_0(u) e^{\beta Z(x_{(1)}+\dots+x_{(n-1)}+u)} du} \end{aligned}$$

Для того, чтобы написать условное правдоподобие, нужно безусловную функцию правдоподобия разделить на вероятность условия:

$$\begin{aligned} & P_r \left\{ \begin{array}{l} \text{длины интервалов} \\ \text{равны } x_{(1)}, x_{(2)}, \dots, x_{(n)} \end{array} \right\} = P_r \left\{ \begin{array}{l} \text{минимальный интервал} \\ \text{равен } x_{(1)} \end{array} \right\} \times \\ & \times P_r \left\{ \begin{array}{l} \text{второй по величине} \\ \text{интервал равен } x_{(2)} \end{array} \middle| \begin{array}{l} \text{минимальный} \\ \text{равен } x_{(1)} \end{array} \right\} \times \dots \times \\ & \times P_r \left\{ \begin{array}{l} \text{максимальный по} \\ \text{величине интервал} \\ \text{равен } x_{(n)} \end{array} \middle| \begin{array}{l} \text{минимальный равен } x_{(1)}, \\ \text{второй равен } x_{(2)}, \dots, \\ \text{(n-1)-й по величине} \\ \text{равен } x_{(n-1)} \end{array} \right\}. \end{aligned}$$

Самый короткий интервал $x_{(1)}$ дает следующий вклад в эту вероятность:

$$\lambda_0(x_{(1)}) e^{-\int_0^{x_{(1)}} \lambda_0(u) e^{\beta Z(u)} du} (e^{\beta Z_{11}} + e^{\beta Z_{21}} + \dots + e^{\beta Z_{n1}}).$$

Вклад от следующего по величине интервала $x_{(2)} > x_{(1)}$ при условии, что минимальный интервал $x_{(1)}$ известен, будет:

$$\lambda_0(x_{(2)}) e^{-\int_0^{x_{(2)}} \lambda_0(u) e^{\beta Z(u)} du} (e^{\beta Z_{22}} + e^{\beta Z_{32}} + \dots + e^{\beta Z_{n2}})$$

и т. д.

При выводе формул используется соображение, что информация о β не может быть получена кроме как из наблюдаемых интервалов между появлениями событий, то есть x_1, x_2, \dots, x_n ; поэтому для всех других значений аргумента функция $\lambda_0(\cdot)$ тождественно равна нулю. Подставляя полученные выражения в формулу для условной функции правдоподобия, получаем:

$$L(\beta; x_1, x_2, \dots, x_n) = \frac{e^{\beta \sum_{i=1}^n Z_{ii}}}{Z_{i=1}^n e^{\beta Z_{i2}} \times \dots \times \sum_{i=1}^n e^{\beta Z_{in}}}$$

Логарифмируя, получим:

$$L_R(\beta) = \beta S - \sum_{i=1}^n \log \left(\sum_{i=1}^n e^{\beta Z_{ii}} \right),$$

где S — то же самое, что и в пуассоновском случае.

Обозначим:

$$U_R(\beta) = \frac{\partial L_R(\beta)}{\partial \beta} = S - n A_R(Z, \beta),$$

$$I_R(\beta) = -\frac{\partial^2 L_R(\beta)}{\partial \beta^2} = n \left[A_R(Z^2, \beta) - \frac{1}{n} \sum_{i=1}^n \frac{(\sum_{i=1}^n Z_{ii} e^{\beta Z_{ii}})^2}{(\sum_{i=1}^n e^{\beta Z_{ii}})^2} \right],$$

$$\text{где } A_R(Z, \beta) = \frac{1}{n} \sum_{i=1}^n \frac{\sum_{i=1}^n Z_{ii} e^{\beta Z_{ii}}}{\sum_{i=1}^n e^{\beta Z_{ii}}}.$$

И совершенно аналогично тому, как это делается в пуассоновском случае, строится доверительный интервал для параметра β и выписывается критическое отношение для проверки гипотезы $\beta = 0$. В случае, когда β есть вектор размерности $k (k \geq 2)$, все формулы получаются аналогично. Для проверки гипотезы $\beta = 0$ рассматривается величина

$$\eta = U_R^T(0) [I_R(0)]^{-1} U_R(0),$$

где U_R^T — транспонированный вектор первых производных условной функции правдоподобия; $[I_R(0)]^{-1}$ — матрица, обратная к матрице вторых производных.

Величина η имеет распределение χ^2 с k степенями свободы. Поэтому для построения границы доверительной области нужно найти такие значения параметров β , для которых $\eta = \chi_{\text{крит}}^2$, где $\chi_{\text{крит}}^2$ — критическое значение, соответствующее выбранной доверительной вероятности.

Выводы

В статье изучены различные методы анализа взаимодействия нелинейных нейронных сетей в предположении, что на вход сети поступает случайный процесс общего вида.

Предложен статистический метод определения взаимосвязи между элементами нейронной сети. Он позволяет получать эффективные оценки параметров, которые характеризуют взаимосвязи между элементами сети, и позволяет производить оценку ее динамических характеристик.

Список литературы: 1. Schetzen M. The Volterra and Wiener theories of nonlinear systems. — N.-Y.: John Wiley, 1980. — 531 p. 2. Пятигорский Б. Я., Чинаров В. А. Винеровская идентификация простых нейронных систем. / Сб. Взаимодействующие марковские процессы и их применение к математическому моделированию биологических систем. — Пушкино, 1982. — С. 144–152. 3. Kroeker J. P. Wiener analysis of nonlinear systems using Poisson — Charlier crosscorrelation // Biol. Cybern., 1977. — V. 27. — P. 221–227. 4. Kroeker J. P. Wiener analysis of functionals of a Markov chain: application to neural transformation of random signals // Biol. Cybern., 1980. — V. 36. — P. 243–248. 5. Хайкин С. Нейронные сети: Полный курс, 2-е изд. — М.: Изд. Дом «Вильямс», 2006. — 1104 с. 6. Круглов В.В., Борисов В.В. Искусственные нейронные сети. — М.: Горячая линия — Телеком, 2001. — С. 324–325. 7. Кокс Д., Льюис П. Статистический анализ последовательностей событий. — М.: Мир, 1969. — 310 с. 8. Cox D. R. The statistical analysis of dependencies in point processes // Stochastic point processes. — N.-Y.: Wiley, 1972. — P. 55–66.

Поступила в редколлегию 04.03.07

УДК 681.32

ХАРАКТЕРИСТИКА АРХИТЕКТУРЫ КОМПЬЮТЕРНОЙ СИСТЕМЫ, ПРИБЛИЖЕННОЙ К ИДЕАЛЬНОЙ

А. Ф. Кургаев

Институт кибернетики НАНУ, г. Киев, Украина

Для решения выявленных проблем развития архитектуры современных компьютеров впервые предложено принять структуру систем обработки знаний (СОЗ) за образец архитектуры компьютерных систем (КС), а в качестве образца структуры СОЗ — идеал структуры научной теории, поскольку именно в ней явно и наиболее концентрированно выражена форма мышления людей при решении проблем. Сформулированы наиболее существенные признаки идеальной архитектуры КС, непротиворечиво сочетающие универсальность со специализацией КС на решение произвольных проблем.

АРХИТЕКТУРА, МОДУЛЬ СЕМАНТИЧЕСКИЙ, МЕТАЯЗЫК, СИСТЕМА КОМПЬЮТЕРНАЯ.

Постановка задачи

Все выявленные в глубине веков прототипы не имеют отношения к современным компьютерам, их нельзя квалифицировать иначе, чем примитивное воплощение эвристической мечты об искусственной сущности с неограниченными возможностями.

Создание современных компьютерных систем (КС) ведет начало от теоретических моделей-открытий К. Геделя, А. Черча, А. Тьюринга, Э. Поста и др., полученных при решении проблем разрешимости в конструктивной логике, и тезиса Черча-Клини, по сути, предлагающего считать эти модели разными вариантами формального определения интуитивного смысла терминов «разрешимый предикат» и «эффективно вычислимая функция» [1–5]. В последующем было доказано точное совпадение классов функций, вычисляемых с использованием этих моделей. Однако, будучи эквивалентными в этом, они существенно различны в отношении используемых выразительных средств.

Модели А. Тьюринга и Э. Поста (далее Т-машины) оказались весьма близкими *явными* формализациями понятия алгоритма и *неявными* — эффективно вычисляемых функций (или предикатов) через возможность построения для каждой из них соответствующей программы (Т-машины).

Две другие модели представляются разными формальными системами:

- основанная на результатах Д. Гильберта теория рекурсивных функций К. Геделя задает в *явной* форме правила определения *декларативной* семантики вычисляемых функций (и разрешимых предикатов) и в *неявной* форме *процедурную* семантику — правила вывода;

- функциональное λ -исчисление А. Черча задает в *явной* рекурсивной форме правила вывода (*процедурную* семантику) и *неявно* — *декларативную* семантику вычисляемых функций (и разрешимых предикатов).

Для Т-машин фон Нейман обосновал конечность программ каждой из них, возможность коди-

ровать целыми числами, хранить в памяти программы вместе с данными и использовать эту пару (программы и данные всякой Т-машины) в качестве исходных данных для универсальной «исполнительной» программы (U-машины Тьюринга [6]). Это решение определило двухуровневую архитектуру первых компьютеров и дало возможность их реализации даже примитивными средствами электроники.

К особенностям КС с архитектурой Тьюринга-фон Неймана (ТН-архитектурой) обычно относят [7–9]:

- линейную память, в которой программы хранятся вместе с данными;
- последовательное выполнение программы;
- отсутствие различий между командами и данными и в семантике данных;
- чрезмерную универсальность.

Следствиями этих особенностей являются проблемы [7–9]:

- неизбежность «кризиса программирования», связанного с высокой трудоемкостью и низкой эффективностью программирования, непрерывно возрастающей сложностью, ассортиментом и низкой надежностью программного продукта;
- сложность достижения параллелизма в решении задач;
- низкая эффективность использования физических ресурсов (памяти и быстродействия) современных КС и технологии микроэлектроники.

В настоящее время перечисленные особенности и проблемы общепринято [10] объяснять процедурным характером исходной вычислительной модели КС с ТН-архитектурой и ее последующим воплощением в форме императивных языков программирования современных компьютеров.

Совершенствование показателей качества КС в непрерывно меняющихся условиях их создания, производства и применения в конечном итоге до-

стигается за счет эволюции их архитектуры. Используя термин «архитектура» (процессора, компьютера, системы), концентрируют внимание на семантической модели специфичной компоненты соответствующего изделия:

- процессах обработки данных (со многими АЛУ, многопроцессорные, конвейерные, теговые, архитектуры потока данных, RISC-архитектуры, матричные и т. п.);
- функциональной структуре (процессора, памяти, других устройств), определяющей взаимосвязь управления с обработкой данных; общей функциональной структуре КС и др.

В этой области накоплено большое количество методологически значимых и технических решений, достойных внимания архитекторов новых КС. Среди нововведений обычно различают КС с архитектурой, ориентированной на язык высокого или сверхвысокого уровня, группу языков программирования или высокого уровня (КС с гибкой архитектурой).

Целая серия методологически значимых архитектурных нововведений была реализована коллективом под руководством В.М. Глушкова в компьютерах семейства «МИР», объективно получивших непревзойденную в течение многих лет популярность в СССР среди исследователей и других категорий пользователей. Среди нововведений отметим: диалоговый характер взаимодействия с пользователями в процессе постановки и решения проблем, ставший доминирующим в современных КС, и микропрограммную интерпретацию языка, семантически эквивалентного текстовому.

Ряд методологически значимых архитектурных нововведений реализован в микропроцессорах семейства «Электроника С5», созданных Институтом кибернетики АН УССР совместно ЛКТБ ЛОЭП «Светлана» в 70-х годах прошлого века на основе опыта разработки мини-ЭВМ М-180 — одной из первых в СССР с микропрограммным управлением. Среди этих нововведений отметим мультиоперационную обработку данных, реализованную уже в первых моделях микропроцессоров этого семейства, и микропрограммную эмуляцию, реализованную в первом в СССР однокристалльном 16-разрядном микропроцессоре компьютера «Электроника С5-21» [11].

Целью настоящей статьи является определение характеристик архитектуры КС, приближенной к идеальной.

1. Требования к архитектуре КС

С целью создания «более благоприятной среды для программистов и корректных программ» были выдвинуты «шесть основных требований, которым должна удовлетворять новая система» [9, с. 292]:

- обнаружение семантических ошибок в программах;
- ограничение последствий возможных ошибок программирования;
- усовершенствование процесса программирования;
- упрощение процесса программирования;
- рост эффективности тестирования и отладки за счет специальных средств;
- упрощение операционных систем и компиляторов.

В разработках современных КС перечисленные требования постоянно учитываются, обеспечивая непрерывный прогресс КС, но их архитектура еще далека от совершенства:

1. Большой семантический разрыв между языками программирования и языком описания проблем, используемым людьми в науке и практике.
2. Столь же велик семантический разрыв между технологией решения проблем в науке и практике и технологией программирования.
3. Рост коэффициента вторичного использования программного продукта (в свое время по расчетам А.П. Ершова он составлял не более 1,5 [12]) достигается преимущественно за счет тиражирования и комплексирования фирменного системного программного обеспечения.
4. Структура аппаратной компоненты КС статична, не изменяется естественно, в ходе решения задачи. Всякое изменение структуры требует программирования.

2. Основные положения для определения характеристик архитектуры КС, приближенной к идеальной

Основная идея достижения цели состоит в принятии структуры системы обработки знаний (СОЗ) за образец архитектуры КС, а в качестве образца структуры СОЗ — идеал структуры научной теории, поскольку именно в ней явно и наиболее концентрированно выражена форма мышления людей при решении проблем.

Обоснованные современными достижениями логико-гносеологических и лингвистических исследований в [13–18] выполнены анализ и синтез наиболее существенных признаков идеала логической структуры системы научных теорий, непротиворечиво сочетающие формальную строгость с содержательностью знаний научных теорий. Эти результаты обобщаются в форме следующих утверждений.

Утверждение 1. Структура системы научных теорий в развернутой (в пространстве или во времени) форме состоит из четырех блоков идентичной структуры — знаков, синтаксиса языков, семантик и функций научных теорий, взаимосвязанных меж-

ду собой, с источниками-приемниками знаковых образов и с реальным миром.

Утверждение 2. Структура каждого из блоков представляется иерархической сетью модулей идентичной структуры, каждый из которых включает словарь и интерпретаторы концептуальных и эмпирических моделей синтаксиса, семантики и прагматики некоторой подтеории.

Утверждение 3. В свернутой в пространстве (но способной к развертыванию во времени) форме структура научных теорий представима структурой половины модуля (в силу симметрии его узлов), содержащей память словаря, интерпретатор концептуальных моделей с памятьми, содержащими концептуальные модели синтаксиса, семантики и прагматики разных подтеорий, и интерпретатор эмпирических моделей.

Утверждение 4. Функционирование всякого модуля определяется решением задач анализа и синтеза структуры некоторого представления информации. Всякая задача всякой подтеории научной теории представляется итерационно-рекурсивной структурой.

Утверждение 5. Множество функций блока функций включает дескриптивно-прескриптивные и аргументативные функции, под управлением которых осуществляются все деятельностные процессы становления, эволюции и использования всякой научной теории.

3. Наиболее существенные признаки идеальной архитектуры КС

Наиболее существенные признаки идеальной архитектуры КС определим в согласии с требованиями устранения недостатками.

1. Первый недостаток устраним предоставлением пользователям всех категорий метаязыка, выразительные возможности которого адекватны представлению знаний произвольной прикладной теории. Тем самым проблема семантического разрыва сводится к проблеме разработки семантического метаязыка с неограниченными выразительными возможностями.

2. Второй недостаток обуславливает проблему операционной системы, предоставляющей пользователям технологию решения проблем, адекватную

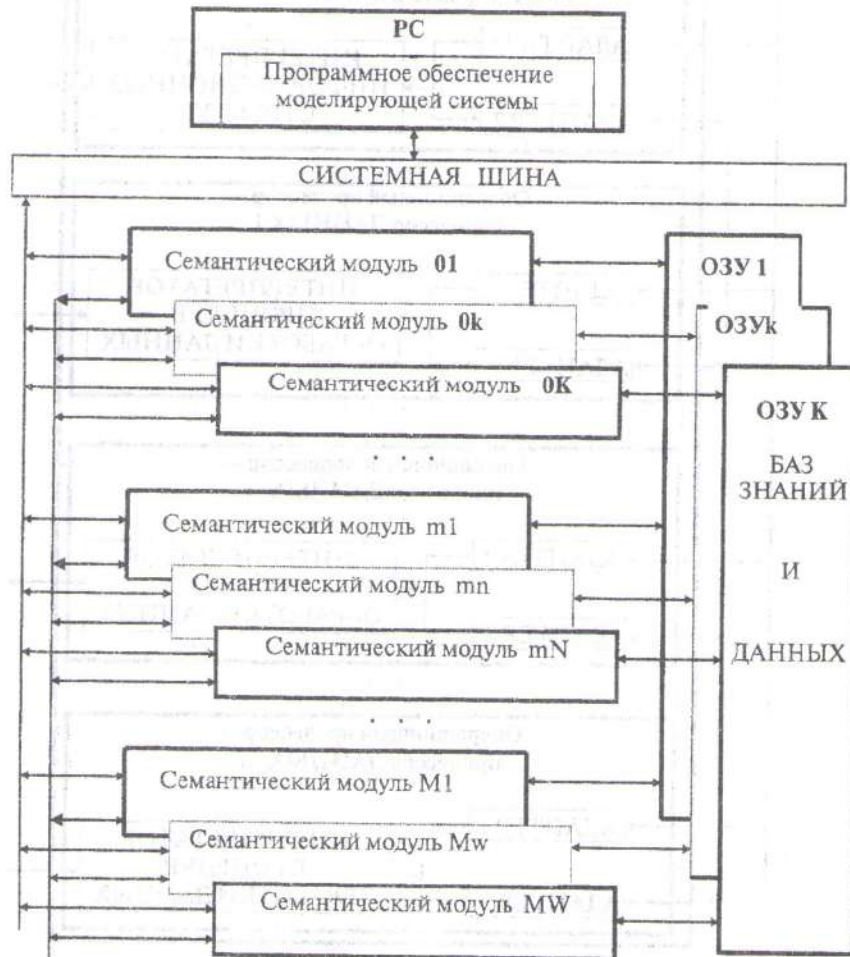


Рис. 1. Моделирующая система с распределенными функциями

используемой в науке и практике, т. е. сводится к проблеме реализации моделей всех функций научных теорий.

3. Третий недостаток устраним реализацией в операционной среде функций накопления знаний для многократного использования (включая разные приложения), модификации и пополнения знаний прикладной теории.

4. Структура аппаратной компоненты КС при решении всякой прикладной проблемы должна быть тождественна структуре соответствующей прикладной теории. Наиболее адекватным решением этой проблемы являются архитектуры КС с распределенными функциями.

4. Архитектура КС, приближенная к идеальной

4.1. Архитектура моделирующей КС содержит (рис. 1): управляющий компьютер РС с программным обеспечением моделирующей системы и сис-

темной шиной, к которой подключены многоярусная ($m = 0 + M$), сеть семантических модулей из $k = 1 + K$ семантических модулей яруса 0, $n = 1 + N$ семантических модулей яруса m и $w = 1 + W$ семантических модулей яруса M , взаимосвязанных между собой и с $k = 1 + K$ блоками памяти.

Все семантические модули сети имеют тождественную структуру (рис. 2): управляющий процессор (процессор баз знаний, ПБЗ) состоит из интерпретатора информационных структур, подключенного к блоку памяти баз знаний, и двух адаптеров, один из которых соединен с системной шиной управляющего компьютера РС, а второй — с шиной управления операционными процессорами семантического модуля и с ПБЗ других семантических модулей; всякий ($j = 1 + J$) операционный процессор (процессор данных, ПД) состоит из интерпретатора процедур обработки данных, подключенного к блоку памяти данных, и двух адаптеров, один из

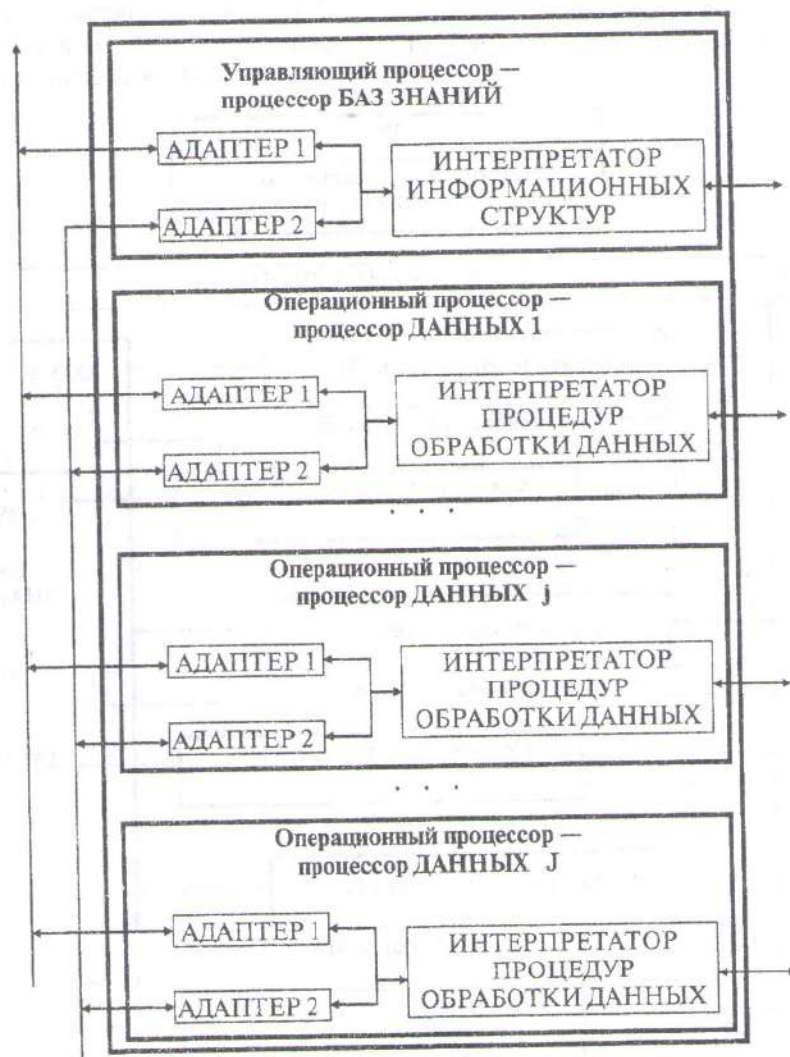


Рис. 2. Структура семантического модуля

которых соединен с системной шиной управляющего компьютера РС, а второй — с шиной управления.

В этой системе управляющий компьютер РС выполняет настройку системы и под управлением сети семантических модулей в процессе постановки и решения всякой прикладной проблемы реализует алгоритмы взаимодействия с пользователем. Функционирование моделирующей системы при решении всякой прикладной проблемы состоит в следующей последовательности работ: под управлением РС установка структуры сети семантических модулей согласно структуре решаемой проблемы; под управлением РС загрузка баз знаний (концептуальных моделей теорий) в ПБЗ, в ПД — операционных программ (операциональных определений эмпирических моделей теорий) и в соответствующие ОЗУ — баз данных (остенсивных определений предметов онтологической картины мира); решение проблемы в процессе взаимодействия ПБЗ с ПД в каждом из семантических модулей и семантических модулей между собой.

В каждом семантическом модуле ПБЗ (во взаимодействии с ПД и с ПБЗ других семантических модулей) интерпретирует концептуальную модель, а каждый ПД выполняет семантические программы эмпирической модели некоторой теории.

Обе структуры (рис. 1, 2) представляют структуру КС в развернутой в пространстве форме. Функционирование этой системы возможно и в КС минимальной конфигурации (рис. 3), содержащей, кроме управляющего РС, сеть в свернутой форме — из одного семантического модуля, содержащего ПБЗ, взаимосвязанный с РС, выполняющим функции ПД [19].

4.2. Системное программное обеспечение моделирующей КС должно выполнять в полном объеме модели всех функций научных теорий, лишь частично реализованных в современных КС, СОЗ, экспертных и интеллектуальных программных системах. В частности, дескриптивная функция научных теорий в современных КС реализована в форме языков программирования (высокого и сверхвысокого уровня — ЛИСП, Пролог, объектно-ориентированных, продукционных и др.) с трансляторами полного описания задачи в машинные исполнимые модели, преспиктивная — как интерпретация машинных моделей решения проблем.

Для осуществления в полном объеме дескриптивной функции требуются: универсальный метаязык, выразительные возможности которого необходимы и достаточны для описания произвольных концептуальных моделей, инкрементный транслятор метапрограмм, реализующий функцию накопления машинных исполнимых моделей, и система поддержки проектирования трансляторов (или интерпретаторов) в машинные исполнимые модели с разных языков прикладных теорий, вплоть до естественных языковых описаний проблем. В свою очередь, система поддержки проектирования трансляторов (или интерпретаторов) должна моделировать процесс интерактивного информационного взаимодействия, включая процедуры ввода, накопления, ретрансляции, редактирования описаний.

Для осуществления в полном объеме преспиктивной функции необходима аппаратно-микропрограммная реализация интерпретатора машинных моделей, семантически эквивалентных и адекватных текстовым.



Рис. 3. Минимальная структура моделирующей системы — машина баз знаний

Осуществление *аргументативных* функций в развитой форме требует:

- для обоснования (при накоплении знаний) всякого нового описания (определения), развивающего или модифицирующего концептуальную модель произвольной эмпирической теории, необходимо реализовать модель процедуры синтеза структуры дедуктивной связи терминов нового описания с терминами имеющегося описания и/или терминов имеющегося описания с терминами нового описания;
- для объяснения теоретических терминов концептуальной модели необходимо реализовать модель процедуры ретрансляции структуры дедуктивной связи данного термина с другими;
- для объяснения результата решения задачи необходимо реализовать модель процедуры демонстрации структуры дедуктивного вывода результата;
- для предсказания фактов, допустимых теорией, необходимо придать интерпретатору машинных моделей функции синтеза фактов в процессе дедуктивного вывода результата решения задачи;
- для понимания высказываний необходимо придать интерпретатору машинных моделей функции дедуктивного вывода высказываний из машинных моделей произвольных эмпирических теорий.

Осуществление функций становления и эволюции теорий необходимо требует разработки и реализации моделей всех процедур синтеза компонент теории (см. [15, 18]).

Все эти системные средства и модели должны быть объединены в интерактивную СОЗ разрабатываемого типа с построением моделей. Кроме того, эта СОЗ должна включать две дополнительные компоненты (их прототипы используют наиболее совершенные современные программные системы), существенно облегчающие процессы использования КС, — *трассировщика* процесса решения задач, необходимого для отладки и верификации знаний в процессе критики, попытках фальсификации теории, а также подсистему *обучения* технологии решения задач и проблем.

Следующим этапом развития этой СОЗ является ее дополнение системой автоматизации проектирования структуры и реализации эмпирических моделей теорий, т.е. специализированных ПД.

Процесс создания такой СОЗ — методом «раскрутки», т.е. начиная с реализации аппаратной компоненты моделирующей КС и метаязыковой модели инкрементного транслятора метапрограмм, реализующего функцию накопления метапрограмм. При этом наиболее эффективна реализация каждой из компонент СОЗ по типу структуры научной теории, а всей СОЗ — как структуры системы научных теорий.

5. Эффективность моделирующей КС

Применение моделирующей КС меняет характер взаимодействия человека с техникой. Система, владеющая знаниями в конкретной прикладной области, ее теорией, для решения всякой проблемы должна быть обеспечена только постановкой проблемы, а в случае неполноты исходной информации для получения дополнительной информации система сможет вести разумный диалог с людьми и/или с техническими системами.

Предоставление пользователю метаязыка для описания прикладных проблем и совершенной технологии постановки и решения задач обеспечивает предельно высокий рост надежности «мягкого» продукта и эффективности программирования. При аппаратной реализации ПБЗ пользователю в форме метаязыка предоставляются все возможности аппаратного уровня, т.е. обеспечиваются предельно достижимые возможности проблемной ориентации архитектуры КС и производительности решения задач. В этом случае предельно простыми и, следовательно, эффективными оказываются реализации операционной системы КС, транслятора метаязыка и прикладных систем.

Если оценивать экономический эффект, то прежде всего следует отметить эффект от изменения качества продукции, изготавливаемой с помощью новой техники. В качестве первичной продукции, создаваемой с помощью моделирующей КС, выступают результаты решения задач — новая информация, принимаемые решения. За счет повышения выразительных возможностей языка, большей естественности технологии решения задач обеспечивается более полная и точная постановка задач, что в сочетании с адекватностью архитектуры КС структуре научной теории повышает качество принимаемых решений, их близость к оптимальным. В свою очередь, улучшение качества принимаемых решений при их внедрении обеспечивает экономию общественных ресурсов, вид которых зависит от приложения, от решаемых задач. Хотя численное значение эффекта от изменения качества решений определить в общем виде невозможно, можно утверждать, что для многих приложений он может достигать значительной доли общего эффекта.

Социальный эффект образуется за счет изменения структуры работ в новой (I) информационной технологии в сравнении с традиционной (II) (рис. 4), где 1 — знания, накопленные в предшествующий период; 2 — новые знания; 3 — программирование; 4 — решение задачи; T — время.

Процесс решения задачи в новой технологии, в отличие от традиционной, опирается на накопленные знания и начинается с анализа (в диалоге человека с КС) прикладной области с целью уточнения,

модификации и пополнения имеющихся знаний (концептуальной модели) до нужного уровня и выделения множества эмпирических терминов, соответствующих прикладной проблеме. Следующий шаг состоит в программировании операциональных определений эмпирических терминов и выполняется теми же средствами и включает те же работы, что и традиционное программирование. Отличие лишь в том, что операциональные определения существенно меньше по объему и структурно элементарно просты в сравнении с обычными программами. Существенной особенностью процесса постановки и решения задач в новой технологии является, как правило, использование значительной доли ранее накопленного знания прикладной теории, которая постоянно растет по мере обучения КС с решением всякой новой проблемы.

Социальный эффект состоит из двух компонент, первая из которых обеспечена принципиально новым характером использования труда, вложенного в постановку и решение проблем с помощью новой техники. Современные компьютеры не допускают вторичного использования знаний, заложенных в прикладных программах для решения задач, постановка которых отличается от предусмотренной этими программами, и тем самым по существу сводит программы и труд, затраченный на их создание, к роли малоценных, быстро изнашивающихся средств труда. Новая же КС оснащается специальными средствами накопления знаний (инкрементным транслятором). Знания, введенные и верифицированные, могут использоваться многократно вне зависимости от постановки конкретных задач до того момента, пока не изменится реальный мир или наши представления о нем. Эти изменения потребуют от пользователя только корректировки, уточнения знаний прикладной теории, объем и сложность которых определяются исходными изменениями. Труд, затраченный на создание знаний, накапливается и переходит в категорию капитальных средств, а характер обновления труда совпадает

с характером обновления капитальных средств. Эта компонента социального эффекта может быть рассчитана как суммарная стоимость прошлого труда, вложенного в постановку и решение задач на новой технике, за вычетом стоимости труда, затраченного на корректировку, уточнение, дополнение базы знаний.

Вторая компонента социального эффекта определяется экономией общественного труда на постановку проблем на новой технике по сравнению с традиционной технологией программирования за счет адекватности процессу решения проблем людьми путем создания и использования научных теорий.

Выводы

Для решения выявленных проблем развития архитектуры современных компьютеров впервые предложено принять структуру СОЗ за образец архитектуры КС, а в качестве образца структуры СОЗ — идеал структуры научной теории, поскольку именно в ней явно и наиболее концентрированно выражена форма мышления людей при решении проблем.

Наиболее существенными признаками идеальной архитектуры КС, непротиворечиво сочетающимися универсальность со специализацией КС на решение произвольных проблем, являются: семантический метаязык представления знаний, системное программное обеспечение, выполняющее модели всех функций научных теорий, и структура моделирующей КС по типу КС с распределенными функциями.

Аппаратная реализация метаязыка, метаязыковая реализация инкрементного транслятора метапрограмм, других компонент системного программного обеспечения и прикладных систем обеспечивают предельно высокую эффективность применения моделирующей КС.

Для достижения глобальной цели — максимальной помощи в решении каждой из бесконечного

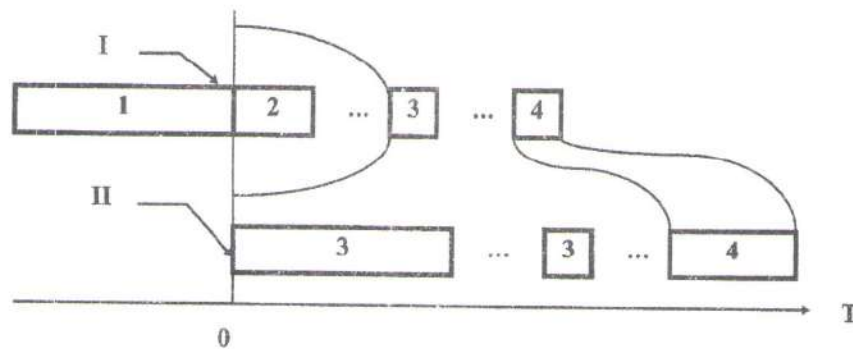


Рис. 4. Сравнение двух технологий работ

множества проблем эволюции общества — идеальная КС должна исполнять модели всех функций научных теорий, объединенные в концептуально единую систему.

Список литературы: 1. *Войшвилло Е.К., Дегтярев М.Г.* Логика как часть теории познания и научной методологии (фундаментальный курс). — М.: Наука, 1994. — Кн. II. — 332 с. 2. *Клини С.К.* Конечная аксиоматизируемость теорий в исчислении предикатов с помощью дополнительных предикатных символов // Математическая теория логического вывода / Сб. пер. под ред. А.В. Идельсона, Г.Е. Минца. — М.: Наука, 1967. — С. 237–284. 3. *Balzer W., Moulines C.U., Sneed J.D.* An architectonic for science: The structuralist program. — Dordrecht: Reidel, 1987. — 431 p. 4. *Мальцев А.И.* Алгоритмы и рекурсивные функции. — М.: Наука, 1986. — 368 с. 5. *Ершов А.П.* Интервью члена редакционной коллегии // Кибернетика. — 1981. — № 4. — С. 9–12. 6. *Turing A.M.* On computable numbers with an application to the Entscheidungsproblem. — Proc. London Math. Soc., Ser. 2, 1936. — 42. — P. 230–265. 7. *Амамия М., Танака Ю.* Архитектура ЭВМ и искусственный интеллект / Пер. с яп. — М.: Мир, 1993. — 398 с. 8. *Симонс Дж.* ЭВМ пятого поколения: компьютеры 90-х годов / Пер. с англ.; Под ред. Б.И. Шитикова. — М.: Финансы и статистика, 1985. — 174 с. 9. *Майерс Г.* Архитектура современных ЭВМ: В 2 кн. / Пер. с англ. — М.: Мир, 1985. — Кн. 1. —

364 с. 10. *Осуга С.* Обработка знаний / Пер. с яп. — М.: Мир, 1989. — 292 с. 11. *Палагин А.В., Кургаев А.Ф.* Проблемная ориентация в развитии компьютерных архитектур // Кибернетика и системный анализ. — 2003. — № 4. — С. 167–180. 12. *Ершов А.П.* Предварительные соображения о лексиконе программирования // Кибернетика и вычислительная техника / Под ред. В.А. Мельникова. — 1985. — Вып. 1. — С. 199–210. 13. *Кургаев А.Ф.* Анализ развития идеала структуры научной теории // Кибернетика и вычислительная техника. — 2003. — Вып. 139. — С. 50–63. 14. *Кургаев А.Ф.* Синтез-обоснование идеала структуры научных теорий: Препр. / НАН Украины. Ин-т кибернетики им. В.М. Глушкова; 2006-1. — К., 2006. — Ч. 1. — 45 с. 15. *Кургаев А.Ф.* Синтез-обоснование идеала структуры научных теорий: Препр. / НАН Украины. Ин-т кибернетики им. В.М. Глушкова; 2006-2. — К., 2006. — Ч. 2. — 24 с. 16. *Кургаев А.Ф.* Синтез-обоснование идеала структуры семантики научных теорий // Кибернетика и вычислительная техника. — 2005. — Вып. 147. — С. 22–32. 17. *Кургаев А.Ф.* Синтез-обоснование идеала структуры языка научной теории // Математичні машини і системи. — 2006. — № 1. — С. 99–112. 18. *Кургаев А.Ф.* Модели функций структуры научной теории // Электронное моделирование. — 2006. — 28, № 3. — С. 19–34. 19. *Кургаев А.Ф.* Исследование архитектуры машины баз знаний // УСИМ. — 2000. — №1. — С. 74–91.

Поступила в редколлегию 10.03.07



UDK 658.512

ABOUT THE THEORETICAL FOUNDATION FOR A META-CONTEXT INTERCHANGE TECHNOLOGY

Z. V. Dudar¹, O. G. Vorochek²

¹ KhNURE, Kharkiv, Ukraine, dudar@kture.kharkov.ua;

² KhNURE, Kharkiv, Ukraine, relf@kture.kharkov.ua

The proposed research is devoted to investigation Context Interchange technology [1] as a new approach for achieving semantic interoperability among autonomous and heterogeneous databases. This technology is intended for the organization of logical connectivity among sources and consumers of large-scale semantically heterogeneous systems. The Context Interchange technology assumes comparison of contexts associated with any data source and consumer of semantically heterogeneous databases and elimination of semantic conflicts that are detected [2]. This comparison uses ontologies as specifications of concepts, their properties and relationships between them in a knowledge domain. This work is concentrated on building, integrating and using ontologies in the Context Interchange technology.

ONTOLOGY, CONTEXT, INTEROPERABILITY, ONTOLOGICAL GRAPH, META-CONTEXT INTERCHANGE TECHNOLOGY.

Introduction

The wide proliferation of Internet and Web-based technologies has caused a large amount of data that is often stored in databases with fundamentally different structures and semantics. Attempts to integrate such data from the diverse databases frequently cause semantic conflicts that inhibit data exchange. There are some approaches for resolving these conflicts [1]. One of the new approaches for integrating large-scale semantically heterogeneous databases is Context Interchange [1, 2]. These works are devoted to development of Context Interchange technology that is intended for organizing logical connectivity among sources and consumers of the diverse databases. Logical connectivity means ability to exchange meaningful information. The main component of this technology is the Context Mediator, which is an intelligent agent that performs data conversions between the source and consumer. It performs the conversions comparing contexts associated with any two databases involved in data exchange and eliminating semantic conflicts that are detected [2]. The Context Interchange technology is implemented in the Context Interchange system. The top-level structure of the Context Interchange system is developed [1,2].

The preceding works [10,11] to this work are devoted to investigation and adaptation of multilevel knowledge representation models for formalizing the semantics of the data in semantically heterogeneous database, and implementation of these models in the Context Interchange technology. Namely, these works were devoted to modeling multilevel database contexts and their comparison for elimination of semantic conflicts.

The comparison of contexts, detection and elimination of semantic conflicts in the Context Interchange technology is performed by the Context Mediator using Common Ontology, a context-independent area in the knowledge base of the Context Mediator. The Common

Ontology is a formal specification of concepts and relationships between them, which are common (or understood) for any source or consumer involved in data exchange. Gruber [3] defines ontology as «an explicit specification of a conceptualization». The problem of applying the Context Interchange technology to large-scale databases is the components of such databases can belong to different domains of knowledge. That is why it is necessary to integrate domain ontologies of the systems involved in data exchange with the goal to build the area of Common Ontology. This work is devoted to the integration of ontologies as the process of building new ontologies from the existing in terms of some methodology.

The definitions and basic principles and criteria for the design of ontologies are introduced in [3]. The classification of ontologies is described in [5, 6]. Within these classifications and the task of context mediation the most interesting types of ontologies are: terminological ontologies to specify the terms that are used to represent knowledges in a domain, information ontologies that specify the record structure of databases, knowledge modeling ontologies that specify conceptualization of the knowledge, and domain ontologies to express conceptualizations that are specific for particular domains [5]. Some existing methodologies for building ontologies are introduced in [4, 7, 8, 9]. Within the task of building Common Ontology for the context mediation the most interesting are methodologies by Uschold and King [7, 8], and by Gruninger and Fox [8, 9]. The Skeletal Methodology by Uschold and King provides general guidelines for developing ontologies. This methodology consists of four stages: identification of purpose, building the ontology, evaluation and documentation. The stage 2 of this methodology is very reasonable for building the Common Ontology for the Context Interchange technology because it fits in the previously used methodolo-

gy for building contexts of databases [10, 11]. The methodology by Gruninger and Fox involves building a logical model of the knowledge. It assumes informal description of specifications with its further formalization. This methodology is also considered for building the Common Ontology as a formal approach to ontology design.

In terms of this the Skeletal Methodology the ontology to be built is the Common Ontology for the Context Interchange technology. The process of building the Common Ontology assumes three steps [6]: ontology capture, coding ontology using a formal language, integrating existing ontologies. This work is concentrated on the second and third steps. The ontological graph is proposed for formalizing ontologies, the genetic search method is proposed for providing integration of ontologies.

Because the role of ontologies is to be base for «understanding» data contexts, the Context Mediator, where process of integrating ontologies is implemented, becomes Meta-Context Mediator. In this case the process of data conversion between the source and consumer, which involves integrating ontologies, is called meta-context mediation [12].

1. Formalizing ontology

The basic knowledge representation model for formalizing ontology in this work is *ontological graph (ontograph)*. Formally, ontograph is the following:

$$G_o = (V_o, E_o), \tag{1}$$

where V_o – a set of nodes formalizing concepts of the ontology; E_o – a set of arcs formalizing relationships between the concepts.

The following adjacency matrix is used for the description of the ontograph:

$$A = (a_{ij}), \text{ dimension } |V_o| * |V_o|, \tag{2}$$

$$a_{ij} = \begin{cases} S_{E_{oij}} \times K_{doij}, & \text{if } (i, j) \in E_o \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

where $S_{E_{oij}}$ – semantic of a relationship, K_{doij} – data quality coefficient.

The ontological graph is multilevel, because during the acquisition and forming ontological knowledge the installation of steady associations is happening. A steady association is a set of ontological concepts and relationships among them unambiguously describing some knowledge. The steady association is considered as a new concept on a higher level of the ontograph. Because the ontograph is multilevel it is necessary to link its levels. The following transition matrix is used for formalizing the transitions between levels:

$$P = (p_{lk}), \text{ dimension } |\mathcal{L}_{S_o}| * |V_o'|, \tag{4}$$

$$p_{lk} = \begin{cases} 1, & \text{if } (l, k) = \text{true} \\ 0, & \text{otherwise,} \end{cases} \tag{5}$$

where \mathcal{L}_{S_o} – set of steady associations; V_o' – set of associated concepts.

A new adjacency matrix can describe the steady associations:

$$As = (As_{mn}), \text{ dimension } |\mathcal{L}_{S_o}| \times |\mathcal{L}_{S_o}|, \tag{6}$$

$$As_{mn} = \begin{cases} a_{ij}, & \text{if } (i, j) = \text{true} \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

The task of searching semantics of concepts is laid in the searching chains of semantic links between the corresponding concepts of integrated ontologies.

2. Principle of forming context-meaningful ontological chains

A context-meaningful ontological Chain (CMOC) – is a sequence of nodes and arcs of ontological graph, which meaningfully describes properties of an on object in terms of the import and export contexts. There are 2 types of CMOC depending on use: static CMOC is used if the import and export contexts are constant and can be reused for a long time; dynamic CMOC is used if the import and export contexts are variable and belongs to different (but adjacent) domains. For achieving interoperability in the static case the search of a semantic link between any two objects is leading using methods and algorithms of the graph theory for the searching of the shortest path. In the case of dynamic CMOC the goal is finding the shortest path for a limited period of time while this path is the most qualified, that is why the only classic methods of search are not appropriate. A genetic search method is proposed in this work. In terms of the meta-context interoperability the main task is the search of CMOC that minimize a number of context conversions between the import and export contexts.

A scheme (fig. 2) represents semantic links in case of meta-context mediation for integration of two databases. There are the following possible ways of the process of integration:

- 1) DB1 – Import Context – Mediator – Export Context – DB2;
- 2) DB1 – Ontology1 – Mediator – Ontology2 – DB2;
- 3) DB1 – Import Context – Ontology1 – Import Context – Mediator – Export Context – Ontology2 – Export Context – DB2;
- 4) DB1 – Ontology1 – Import Context - Ontology1 – Mediator – Ontology2 – Export Context – Ontology2 – DB2.

The first case does not consider ontological knowledge about domains of the databases, the second – context. That is why these schemes are not appropriate. Both #3 and #4 ways are reasonable, but in the case #4 context conversions are necessary the only to form

context representation of data, and the rest of integration process is based on fuzzy ontological links between domain ontologies. However this way requires defining context-ontological relationships and integrating ontologies. These two steps are described below.

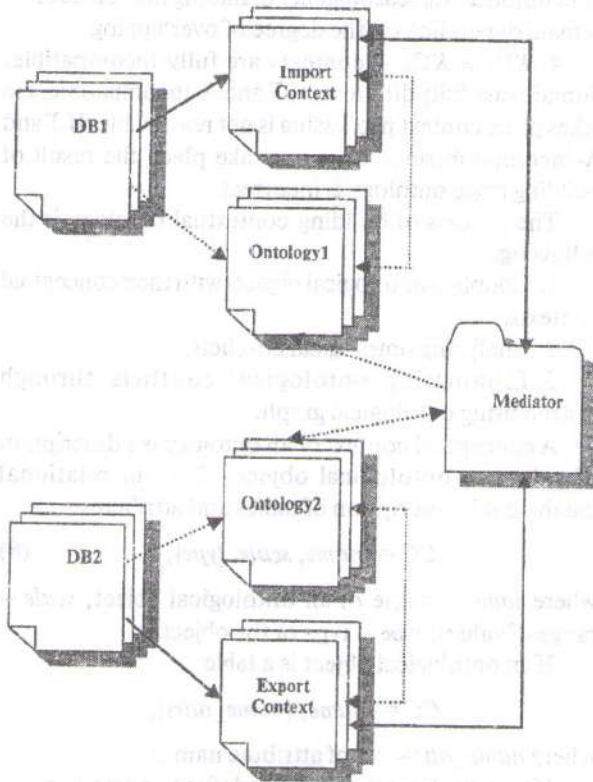


Fig. 1. Data flows in meta-context mediation

3. Forming context-ontological relationships in the knowledge base of the Meta-Context Mediator

In terms of the meta-context mediation domain ontology — is a set of object and relationships between them, which are described by conceptual model of a database. Context describes relationships between particular data and its semantics (concepts) in the particular databases. There are the following context and ontological concepts useful for the meta-context approach:

- basic ontology (BO) — is a basic ontological graph formed for any source or consumer;
- integrated ontology (IO) — is a result of integration of basic ontologies;
- steady associated ontology (AO) — is a resulting ontology with steady associations; this is the higher-level ontology;
- contextual ontology (CO) — is an ontology coupled with a context of any source of consumer involved in data exchange;
- key context (KC) — a set of key concepts and relationships between them for a domain of sources and consumers involved in data exchange;

• conceptual context (CC) — is defined for a fragment of a database (f.e. for only source’s database, for a table, for an attribute).

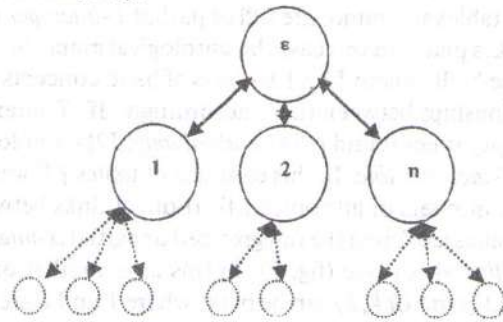


Fig. 2. Context hierarchy

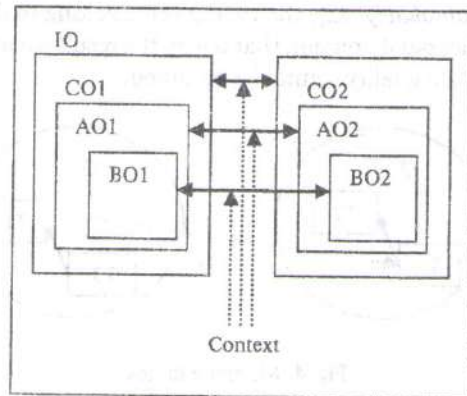


Fig. 3. Ontological hierarchy

Both contexts and ontologies are organized as hierarchy. In the hierarchy of contexts (fig. 2) root ε — is a key context describing an external representation of a domain, branches and leaves — are conceptual contexts describing detail properties of database fragments. Basing the hierarchy of ontologies (fig. 4) it is possible to describe a process of forming an integrated ontology:

$$IO = CO + KC = CO1 + CO2 + \sum CC = AO1 * CC1 + AO2 * CC2 + \sum CC$$

Thus, for finding an integrated ontology it is necessary to form basic ontologies of systems involved in data exchange, to find steady associations and to resolve ontological conflicts between contextual ontologies.

4. Building basic ontologies

An integration of ontologies of two systems involved in data exchange requests the following steps:

- 1) ontology capture and forming ontology graph,
- 2) resolving ontological conflicts and genetic search of ontological chains for interoperability.

The first step is based on the mapping of ontology and data of the first system involved in data exchange into according ontology and data of the second system.

For example, for two relational databases, where sets of predefined ontologies O_1 and O_2 are identical to sets

of tables: $O1 \equiv \{T1\}$, $O2 \equiv \{T2\}$, it is necessary to map $\{T1\} \rightarrow \{T2\}$ and install links between tables.

If direct (or clear) links between semantically equivalent tables are found, the full or partial *T-interopability* takes place. In this case the ontological graph $G(V, E)$ can be built, where V and E – sets of basic concepts and relationships between them accordingly. If *T-interopability* is not found ($\{T1\} \neq absolutely \{T2\}$) ontologies are *T-incompatible*. In this case sets of tables $\{T\}$ are divided into sets of attributes $\{A\}$. If direct links between attributes are found the full grouped or partial *A-interopability* takes place (fig. 7). In this case another ontological graph $G(V, E)$ can be built, where V and E are sets of attribute-level basic concepts and relationships between them. If *A-interopability* is not found ($\{A1\} \neq absolutely \{A2\}$) the two systems belong to different conceptual domain, that is why the result wit appropriate data quality cannot be obtained.

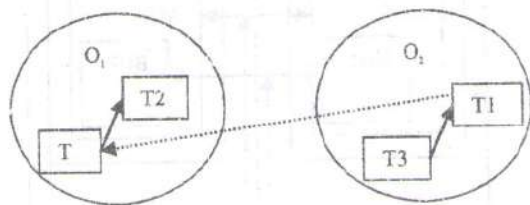


Fig. 4. Mapping tables

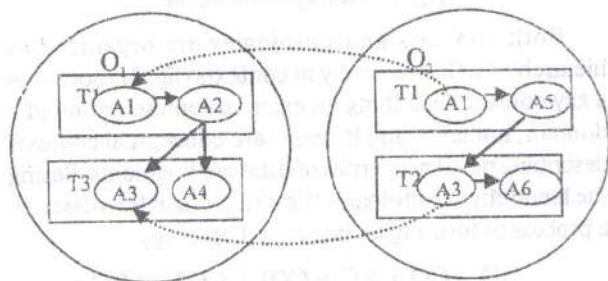


Fig. 5. Mapping attributes

5. Building contextual and steady associated ontologies

The process of building contextual ontology consists of finding hidden relationships, analyzing clear relationships and coupling them. Within this process key context (KC), which defines key concepts of a domain, can be a criteria for possibility of integrating potentially incompatible systems. The analysis of key contexts of two systems involved in data exchange (KC1 and KC2) is the first step of building contextual ontology. There are some types of relationships between two key contexts:

1. $KC1 \equiv KC2$ – integrated objects belong to the same domain. All concepts of the first system are unambiguously mapped into concepts of the second system. Find-

ing interoperability is easy, and assumes the only classic search of context-ontological links between objects.

2. $KC1 \in KC2$ – domains are nested. Finding interoperability is the same as in (1).

3. $KC1 \approx KC2$ – domains are partly overlapped. A conclusion about reasonability of finding interoperability is made depending on the degree of overlapping.

4. $KC1 \neq KC2$ – contexts are fully incompatible, domains are fully different. If T and A-incompatibility takes place context processing is not reasonable. If T and A-incompatibility does not take place the result of building basic ontology is incorrect.

The process of building contextual ontology is the following:

1. Coupling ontological objects with their conceptual contexts.
2. Analyzing ontological conflicts.
3. Eliminating ontological conflicts through restructuring ontological graph.

A conceptual context of an ontology is a description of values of ontological objects. F.e. for relational database it is description of tables and attributes:

$$CC = \{name, scale, type\}, \tag{8}$$

where *name* – name of an ontological object; *scale* – range of values; *type* – type of the object.

If an ontological object is a table:

$$CC = \{name, \{name_attr\}\},$$

where *name_attr* – set of attribute names.

If the ontological object needs further detailing:

$$CC = \{name, \{name_attr, scale, type\}\}. \tag{9}$$

After contextual ontology is built it is necessary to resolve ontological conflicts to build steady associated ontology.

Assume two sets of objects are built: $O1 = \{o_{1i}\}$ – describes a set of ontological objects of the first system involved in data exchange, $O2 = \{o_{2j}\}$ – the second. Assume that relationships between objects are defined, and links for interoperability are found.

The first step of resolving ontological conflict is eliminating homonyms. Assume we have two potentially interoperable objects $o_{1i} \in O1$ и $o_{2j} \in O2$. If conceptual contexts $CC1 \equiv CC2$ the probability of homonymy is very low, that is why the link between these objects is correct. If $CC1 \neq CC2$ then homonymy exists.

F.e., we have two ontological fragments (fig. 6) with potential interoperability, equal conceptual contexts for the objects #3 and different conceptual contexts for the object #1. Thus we have the following set of conceptual contexts:

$$\begin{aligned} CC1_1 &= \{name = 1, scale1, type1\} \\ CC1_2 &= \{name = 2, scale2, type2\} \\ CC1_3 &= \{name = 3, scale3, type3\} \end{aligned}$$

$$CC2_1 = \{name = 1, \{name_attr1\}\}$$

$$CC2_2 = \{name = 4, \{name_attr2\}\}$$

$$CC2_3 = \{name = 3, scale3, type3\}$$

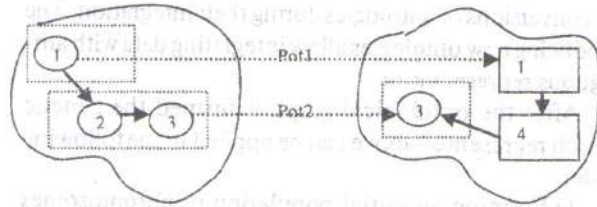


Fig. 6. Homonymy

There are two potential links:

$$Pot1: O1_1 \rightarrow O2_1, CC1_1 \neq CC2_1$$

$$Pot2: O1_3 \rightarrow O2_3, CC1_3 = CC2_3$$

In this case objects $O1_1$ and $O2_1$ have the same name but different conceptual contexts, that is why the link Pot1 is excluded and one of the objects is renamed to 1_hom (this renaming is applied only for ontological object but not for physical data structure). This renaming allows processing homonymic objects independently. The result of eliminating homonymy is represented on a figure 7.

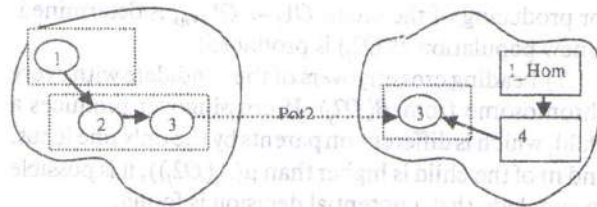


Fig. 7. Eliminated homonymy

Analysis of synonymy is more complicated because synonymy assumes full coincidence of conceptual contexts and links to other objects for two objects in different ontologies. The coincidence of conceptual contexts only assumes potential synonymy; that is why analyzing synonymy is leading by the following way:

- 1) Extraction of an integrated object of the first ontology;
- 2) Forming a set of links from this object to other objects. Thus we obtain a mask of interoperability;
- 3) Comparing this mask with every object of the second ontology. Coincidence means synonymy – this is ideal case because interoperability is achieved, or steady association is installed (fig. 8, object #2 and #6). Otherwise we have partial synonymy – it means ontological conflict exists (fig. 9) and should be eliminated.

In the case of partial synonymy (fig. 9) key object #5 of the second ontology (part of its link is overlapped by the mask) is split on two objects: with overlapped (#5)

and not overlapped (#5_{link}) links. Thus, a quasi-steady association is installed between the integrated objects #2 of the first ontology and the object #5 of the second one. A link of “neighborhood” is installed between objects #5 and #5_{link}. The next step is search meaningful ontological chains between integrated ontology for achieving interoperability.

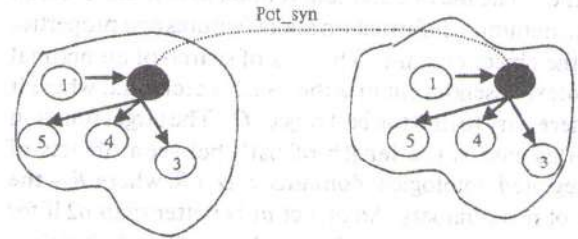


Fig. 8. Synonymy

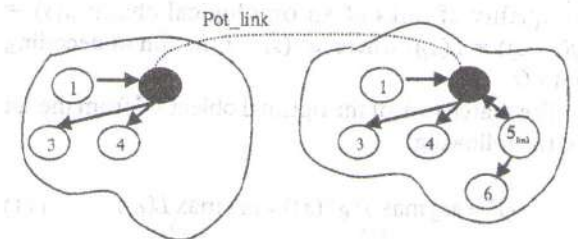


Fig. 9. Quasi-steady link

6. Search of ontological chains for achieving interoperability

After installation of steady associations, basing on these sets of ontologies, it is necessary to search ontological chains for interoperability. If integrated objects are compatible links between them are clear and interoperability is guaranteed. If integrated objects are incompatible it is necessary to use indirect methods to provide interoperability. In this case the task of searching compatibility between two objects $O1_i \rightarrow O2_j$ is the task of searching an ontological chain $O1_i \rightarrow O1_{i+n}$, where $O1_{i+n} \approx O2_{j+n}$, and the length of the chain $L(O1_i \rightarrow O1_{i+n})$ must provide quality as closer as possible to the required quality of the result. This length of chain is assumed as minimum length for integration of the objects. In this case the following equation should be true:

$$L(O2_{j+n} \rightarrow O2_j) \geq Ll(O1_i \rightarrow O1_{i+n}), \quad (10)$$

where $O2_{j+n} \approx O1_{i+n}$ and $O2_j \approx O1_i$.

The second model, which is used by the meta-context mediator, is intended for the genetic search of context-meaningful ontological chains for interoperability. The goal of the search of the optimal context-meaningful chain for the first system involved in data exchange is the finding ontology of the second system participating in data exchange. This ontology should correspond to

ontology of the fist system with the predetermined probability. There are two domains in this model: domain of ontological objects O and domain of statements S , where: $O \neq S$.

Representation is mapping from the O to the S . In our case it is encoding function $e: O \rightarrow S$.

For any object o from the O and a statement s from the S the formula $s=e(o)$ means that the s is the statement of the o . The use of statements allows us to make a search with minimum information about features and properties of the object domain. The task of search of an optimal context-essential chain is the usual search task, where it is necessary to find the best object O^* . The target function in this case is the length of path between objects of integrated ontological domains: $f: O \rightarrow R$, where R – the set of real numbers. An object $o1$ is better than $o2$ if for any two objects $o1$ and $o2$ from the set $O: f(o1) > f(o2)$.

The fitness-function is necessary for evaluation of statements: $m: S \rightarrow R$, where R – the set of real numbers, μ – quality of links of an ontological chain: $\mu(s) = f(e^{-1}(s)) = L(o_i)$, where $e^{-1}(s)$ – function of decoding S into O .

The search task of the optimal object O^* from the set O is the following:

$$O^* = \arg \max_{s \in S} f(e^{-1}(s)) = \arg \max_{o_i \in O} L(o_i). \quad (11)$$

The resolving this task is happening through the search of the optimal representation s^* in the domain S :

$$s^* = \arg \max_{s \in S} \mu(s). \quad (12)$$

The adjacency matrixes, which describe ontology, are used in the proposed genetic search. These matrixes consider steady associations. All elements of an adjacency matrix, which describes mutual correspondences between objects, must be represented as single directed correspondence with the computation of data quality indicator. According to a query the sub-domain of statements from the adjacency matrix is defined.

The obtained domain has the $n \times m$ dimension, where n – the total number of ontological objects, m – the number of ontological objects linked to the object O^* .

Use of the genetic model in the meta-context mediation supposes redefinition of some classical concept of this formalism.

Mutation operator changes the particular selected hen with the value 1 rather than random hen. The mutated hen is changed to 0 and determines a collection of locus (considering the object O), which state is changed to 1. This mechanism allows considering possible links among the objects of ontology.

Inversion destroys the sequence of fragments of chromosome from parent to child. It means inclusion or exclusion of a possible property of ontology.

Crossingover produces a new chromosome with the properties from the both of parents. It allows changing a recessive fragment of chromosome to a dominant one that can give a better result of the search.

The proposed genetic approach to the knowledge modeling allows resolving a problem of modeling semantic conversions of ontologies during their integration. The producing new ontologies allow integrating data with ambiguous representation.

After the ontological graph is formed the genetic search represented above can be applied in the following order:

- 1) Forming an initial population of chromosomes $B(O1_i) = (A_1(O1_i) \dots A_n(O1_i))$;
 - 2) Finding hens, which are set to 1 and their locus correspond to the integrated object;
 - 3) Mutation of the $B(O1_i)$ and finding a hen of interoperability for the new population $B'(O1_i) = (A_1'(O1_i) \dots A_{i-1}'(O1_i), A_{i+1}'(O1_i), \dots A_n'(O1_i))$;
 - 4) Repeat 2 and 3 while $B^{(m)}(O1_i) \neq (0)$ or while it is reasonable. A chromosome can be included into the population of candidates only if $\mu(A_{i+1}(O1_i)) \geq \mu(A_i(O1_i))$;
 - 5) Selecting a chromosome from the population $B_k(O1_i)$ with the maximum $\mu(A_i(O1_i))$. The hen of interoperability of this chromosome defines an object $O1_{i+n} \approx O2_{j+n}$.
 - 6) The number of mutations N_i , which are necessary for producing of the chain $O1_i \rightarrow O1_{i+n}$, is determined. A new population $B(O2_j)$ is produced;
 - 7) Leading crossingovers of the candidate with every chromosome from $B(O2_j)$. If crossingover produces a child, which is different on parents by the only one locus, and m of the child is higher than $\mu(A_i(O2_j))$, it is possible to conclude that a potential decision is found;
 - 8) The produced child, the parent are decoded.
- Thus a fragment of ontological chain is found. Iterative call of this algorithm allows finding almost all ontological chains for providing interoperability for integration of ontologies.

Conclusion

This paper describes two models, which are useful for integration of ontologies in the Context Interchange technology. The first model is ontological graph, which is used for modeling ontologies. The second model is genetic search of ontological chains on the ontological graph, which is intended to find interoperability between integrated ontologies. The integration of ontologies is targeted to form the area of Common Ontology, which is used in the Context Interchange technology as the base for comparison of contexts and elimination of semantic conflicts.

Literature: 1. Goh, C. H., Madnick, S. E., and Siegel, M. D. Context interchange: overcoming the challenges of large scale interoperable database systems in a dynamic environment. In

Proceedings of the Third International Conference on Information and Knowledge Management (Gaithersburg, MD, Nov 29—Dec 1 1994), pp. 337—346. 2. *C. Goh*, «Representing and reasoning about semantic conflicts in heterogeneous information systems», Ph.D. Thesis, MIT Sloan School of Management, 1996. 3. *Gruber T.*, (1995), Towards Principles for the Design of Ontologies Used for Knowledge Sharing, International Journal of Human-Computer studies, 43 (5/6): 907 — 928. 4. *M. Uschold and M. Gruninger*. Ontologies: principles, methods and applications. The Knowledge Engineering Review, 11(2):93—136, November 1996. 5. *G. van Heijst, A. T. Schreiber, and B. J. Wielinga*. Using explicit ontologies in KBS development// International Journal of Human-Computer Studies, 46(2/3).- 1997.- 183—292. 6. *Guarino N.* Some ontological principles for designing upper level lexical resources. In Proceedings of the First International Conference on Language Resources and Evaluation, Granada, 1998. 7. *M. Fernandez Lopez*, Overview of Methodologies for Building Ontologies. Workshop on Ontologies and Problem-Solving Methods: Lessons Learned and Future Trends, IJCAI-99. Stockholm, Sweden. August,

1999. 8. *M. Uschold and M. King*. Towards a methodology for building ontologies. In Proceedings of the IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, Canada, 1995. 9. *M. Gruninger and M. Fox*. Methodology for the design and evaluation of ontologies. In Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing held in conjunction with IJCAI-95, Montreal, Canada, 1995. 10. *Ye. V. Biletskiy, O. O. Biletska, M. F. Bondarenko and Z. V. Dudar*. The Context Interchange Technology Based on the Predicate Knowledge Representation Model // The 2002 International Conference on Artificial Intelligence (IC-AI'02: Las Vegas, USA, June 24-27, 2002). 11. *M. F. Bondarenko, Z. V. Dudar, Ye. V. Biletskiy, O. G. Baygozina*. The Context Interchange System Framework // Problems of Bionics (Problemi Bioniki), KTURE, Kharkiv, 1999, No 50. P.18-24. 12. *Baygozina O. G.* «Using ontology in the context mediator system»// 6-th International Conference «Theory and technique of information receiving, handling and sending», Kharkiv, 2000.

Поступила до редколегії 15.03.07

... (faint, mostly illegible text, likely bleed-through from the reverse side of the page)

... (faint, mostly illegible text, likely bleed-through from the reverse side of the page)

УДК [005.912:004]:519.862

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ОПРЕДЕЛЕНИЯ ЭКОНОМИЧЕСКОЙ ЭФФЕКТИВНОСТИ ВНЕДРЕНИЯ КОРПОРАТИВНЫХ СИСТЕМ ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА

Г. Г. Асеев

ХГАК, г. Харьков, Украина, aseev@ic.ac.kharkov.ua

Предлагается математическая модель эффективности корпоративных систем электронного документооборота, позволяющая оценить прямой и косвенный эффекты от внедрения системы вне зависимости от типа и размера организации. Результатом расчета является определение общих затрат на внедрение системы и срока окупаемости системы. Приводится пример расчета эффективности внедрения системы.

ЭКОНОМИЧЕСКАЯ ЭФФЕКТИВНОСТЬ, ЭЛЕКТРОННЫЙ ДОКУМЕНТООБОРОТ, ИНФОРМАЦИОННЫЙ ПОТОК, БИЗНЕС-ПРОЦЕСС, СРОК ОКУПАЕМОСТИ КАПИТАЛЬНЫХ ВЛОЖЕНИЙ, ТРУДОЗАТРАТЫ, СЕБЕСТОИМОСТЬ, ВНЕДРЕНИЕ.

Введение

В предыдущей работе [1] была приведена методика определения экономической эффективности корпоративных систем электронного документооборота (КСЭД). В связи с проводимыми реальными расчетами определения экономической эффективности КСЭД были выявлены определенные недостатки в методике, изложенной в [1], и в нижеизложенной математической модели они были учтены.

В настоящее время за рубежом и в Украине происходят процессы замены традиционного документооборота электронным. В большинстве организаций сложилось мнение, что использование КСЭД — это вопрос престижа, который повысит качество работы и сделает организацию более привлекательной для инвестиций. Однако в результате исследований, проведенных П. Страсманом на нескольких тысячах коммерческих предприятий в полусотне стран, были получены следующие данные [2]:

- расходы на КСЭД возрастают, если организация в течение прогнозируемого периода времени устойчиво получает прибыль, и сокращаются, если прибыль отсутствует или снижается;
- чем больше квалифицированных сотрудников, тем выше расходы на аппаратное и программное обеспечение КСЭД;
- затраты на компьютерные ресурсы больше всего увеличиваются в расчете на сотрудников организации, которые не заняты в непосредственном получении прибыли;
- у большинства организаций уровень расходов на КСЭД превышает величину собственного капитала.

Для организаций, находящихся на государственном финансировании (в частности, для органов государственной власти и управления), оценка прибыли от внедрения КСЭД является более сложной задачей, требующей учета большого числа субъек-

тивных факторов, во многом зависящих от специфики деятельности каждой конкретной организации.

Таким образом, опыт внедрения КСЭД показывает, что величина расходов на систему не связана напрямую с получением прибыли. При этом принято считать, что эффективность КСЭД определяются [3]:

- сокращение затрат на создание бумажных документов и ведение традиционного документооборота;
- сокращение непроизводственных затрат рабочего времени сотрудников (например, на поиск документов, создание отчетов, ожидание в приемных, повторные согласования и т. д.);
- увеличение скорости распространения информации между сотрудниками и подразделениями организации;
- повышение качества работы организации в целом (повышение качества принимаемых управленческих решений, прозрачность управления, раскрытие творческого потенциала сотрудников, возможность более качественных деловых коммуникаций между сотрудниками и подразделениями и т. д.).

Сокращение разного рода затрат и ускорение информационных потоков относят к прямому эффекту от внедрения КСЭД, а повышение качества работы организации — к косвенному эффекту.

Методы оценки прямого эффекта от внедрения КСЭД рассматриваются в работах [3, 4]. Они основываются на некоторых усредненных показателях, например, на характеристиках суммарных затрат на выполнение типовых операций с документами до и после внедрения КСЭД [3]. Однако в большинстве случаев учет только прямого эффекта не окупает суммарных затрат на КСЭД, особенно в организациях управляющего типа.

Расчет косвенного эффекта затруднен из-за трудности формализации таких характеристик, как

творческая инициатива сотрудников, качество принимаемых решений и т. д. Полученные в результате такого расчета оценки являются экспертными и в большинстве случаев субъективными. В [4] стоимость сбереженного рабочего времени рассчитывается довольно приблизительно, процент его экономии имеет большой разброс (например, экспертная оценка экономии рабочего времени за счет ускорения бизнес-процессов имеет диапазон 10–30%), определяемый типом и спецификой организации. Такой разброс может давать существенную погрешность при оценке окупаемости системы.

Таким образом, целью данной работы является разработка математической модели оценки эффективности, позволяющей оценить затраты и прямой и косвенный эффект от внедрения КСЭД, не зависящего от направления деятельности и масштаба организации.

1. Постановка задачи

Разрабатываемая математическая модель оценки эффективности КСЭД должна быть независимой от типа и размера организации и учитывать следующие характеристики:

- существующие на момент внедрения в организации аппаратные и программные средства, осуществляющие реализацию электронного документооборота (ЭД);
- экономию от снижения эксплуатационных расходов и повышения качества управленческих решений в КСЭД;
- предпроизводственные затраты на внедрение КСЭД.

Результатами расчета являются общая стоимость работ по внедрению системы и срок окупаемости капитальных вложений на КСЭД.

За основу расчета берутся годовой объем документооборота на предприятии (H_2) (поток входящих и исходящих документов, объем внутреннего документооборота, в том числе и технологического и пр.) и процент (A_i) реализаций ЭД, выполняющихся на собственной вычислительной технике до внедрения КСЭД.

2. Математическая модель определения экономической эффективности внедрения КСЭД

Годовая экономическая эффективность ($\mathcal{E}_{\text{год}}$, тыс. у. е./год) рассчитывается по формуле:

$$\mathcal{E}_{\text{год}} = \Delta C + \mathcal{E}_k - \left(\frac{\Delta K}{E_A} + K_{2j} E_V \right), \quad (1)$$

где ΔC — снижение себестоимости реализаций ЭД в расчетном году после внедрения КСЭД (тыс. у. е./год); \mathcal{E}_k — годовая экономия от снижения экс-

плуатационных расходов и повышения качества решений в КСЭД (тыс. у. е./год) за счет:

1) экономии усилий путем повторного использования существующих документов;

2) экономии времени на поиск нужных документов за счет систематизации в хранении документов и эффективных средств поиска информации;

3) ускорения всех бизнес-процессов за счет их автоматизации, формализации и контроля исполнительской дисциплины и т. д.;

ΔK — величина общего изменения капитальных затрат (капитальных дополнительных затрат) при внедрении КСЭД (тыс. у. е./год); E_A — срок окупаемости основных фондов (лет) (по отдельному расчету, но для ускоренной оценки можно принимать 3 года); K_{2j} — предпроизводственные затраты на внедрение КСЭД (тыс. у. е./год); E_V — норма амортизационных отчислений капитальных вложений в конфигурацию вычислительной техники и программное обеспечение (1/год) (обычно $E_V = 0,25$).

Изменение (снижение) себестоимости выполнения реализаций документов в КСЭД (тыс. у. е./год) определяется по соотношению:

$$\Delta C = 10^{-3}(C_i - C_j), \quad (2)$$

где C_i и C_j , соответственно, себестоимость выполнения реализаций электронных документов (ЭД) до внедрения и после внедрения КСЭД (у. е./реал.).

Себестоимость выполнения реализаций всех документов до внедрения КСЭД (у. е./реал.) определяется по формуле:

$$C_i = \varphi_i (Q_{1i} Z_{1i} + Q_{2i} Z_{2i}), \quad (3)$$

где φ_i — коэффициент накладных расходов и прочих прямых затрат (общепроизводственные плюс административные) до внедрения КСЭД; Q_{1i} и Q_{2i} — трудозатраты на выполнение реализаций ЭД на компьютерной технике и документов традиционным способом, соответственно, до внедрения КСЭД (чел./час.); Z_{1i} и Z_{2i} — среднечасовая заработная плата исполнителя, работающего на компьютерной технике и традиционным способом, соответственно, до внедрения КСЭД (у. е. чел./час.).

Трудозатраты на выполнение реализаций ЭД на компьютерной технике в (3) до внедрения КСЭД (чел./час.) находятся по формуле:

$$Q_{1i} = H_{1i} D_{1i}, \quad (4)$$

где H_{1i} — количество выполнения реализаций ЭД на компьютерной технике до внедрения КСЭД; D_{1i} — среднее количество часов, за которое выполняется 1 ЭД в организации до внедрения КСЭД. Значение

D_{1i} определяется, исходя по факту, или из экспертных оценок (при незначительной автоматизации $D_{1i} \approx (3 \div 10 \text{ дней}) \cdot 8 \approx 24 \div 80 \text{ часов}$), или по соотношению (6).

В свою очередь, количество выполнения реализаций ЭД на компьютерной технике до внедрения КСЭД определяется следующим образом:

$$H_{1i} = H_2 \cdot A_i / 100, \quad (5)$$

где H_2 — годовой объем документооборота на предприятии; A_i — процент реализаций ЭД, выполняющихся на собственной вычислительной технике до внедрения КСЭД.

Среднее количество часов в (4), за которое выполняется 1 ЭД в организации до внедрения КСЭД, можно также найти из соотношения:

$$D_{1i} = \frac{K_{1i} t}{H_{1i}}, \quad (6)$$

где K_{1i} — количество компьютерных рабочих мест в организации, имеющихся до внедрения КСЭД; t — количество рабочих часов в расчетном году (час/год). Определяется прямым счетом и для односменной работы $t = 2080 \text{ час/год}$.

Если подставить (6) в (4), то получим, что трудозатраты на выполнение реализаций ЭД на компьютерной технике в (3) до внедрения КСЭД (чел./час.) будут находиться по формуле:

$$Q_{1i} = K_{1i} t. \quad (7)$$

Далее, в (3) трудозатраты на выполнение реализаций документов традиционным способом (чел./час.) до внедрения КСЭД находятся так:

$$Q_{2i} = H_{2i} D_{2i}, \quad (8)$$

где H_{2i} — количество выполнения реализаций документов традиционным способом; D_{2i} — среднее количество часов, за которое выполняется 1 документ традиционным способом до внедрения КСЭД. Определяется по факту в каждой организации до внедрения КСЭД или по выражению (10).

В свою очередь, количество выполнения реализаций документов традиционным способом до внедрения КСЭД получается из соотношения:

$$H_{2i} = H_2 - H_{1i}. \quad (9)$$

Среднее количество часов, за которое выполняется 1 документ традиционным способом до внедрения КСЭД в (8), можно получить по такому выражению:

$$D_{2i} = \frac{N_{2i} t}{H_{2i}}, \quad (10)$$

где N_{2i} — среднесписочная численность сотрудников, выполняющих реализацию документов традиционным способом до внедрения КСЭД.

Если подставить (10) в (8), то получим, что трудозатраты на выполнение реализаций документов традиционным способом (чел./час.) до внедрения КСЭД будут находиться следующим образом:

$$Q_{2i} = N_{2i} t. \quad (11)$$

Если подставить (7) и (11) в (3), то получим, что себестоимость выполнения реализаций электронных документов до внедрения КСЭД будет определяться по формуле:

$$C_i = \varphi_i t (K_{1i} Z_{1i} + N_{2i} Z_{2i}). \quad (12)$$

Себестоимость выполнения реализаций ЭД после внедрения КСЭД (у. е./реал.) рассчитывается по формуле:

$$C_j = \varphi_j Q_j Z_j, \quad (14)$$

где φ_j — коэффициент накладных расходов и прочих прямых затрат (общепроизводственные плюс административные после внедрения КСЭД); Q_j — трудозатраты на выполнение реализаций ЭД после внедрения КСЭД (чел./час.); Z_j — среднечасовая заработная плата исполнителя после внедрения КСЭД (у. е. чел./час.).

Трудозатраты на выполнение реализаций ЭД после внедрения КСЭД:

$$Q_j = H_2 D_2, \quad (15)$$

где D_2 — среднее количество часов, за которое выполняется 1 ЭД после внедрения КСЭД. Значение D_2 определяется, исходя из фактических данных или из экспертных оценок ($D_2 \approx 1 \div 6 \text{ дней}$). Можно использовать данные экспертных оценок, приведенных в [5, с. 109], которые говорят, что «на 20–25% возрастает производительность труда персонала при использовании электронного документооборота». В таком случае, используя (10), будем иметь:

$$D_2 = (0,75 \div 0,80) D_{2i}. \quad (16)$$

Возвращаясь к (1), рассчитаем годовую экономию от снижения эксплуатационных расходов и повышения качества решений в КСЭД (тыс. у. е./год):

$$\Theta_k = \frac{10^{-3} C_j D_{3j}}{\left(\frac{1}{T_j} + E_v \right)}, \quad (17)$$

где D_{3j} — параметр снижения экономии трудозатрат, и по экспертным оценкам, приведенным в [4], равен $10 \div 30\%$; T_j — срок службы эксплуатации конфигурации вычислительной техники (лет).

Величина общего изменения капитальных затрат (капитальных дополнительных затрат) при внедрении КСЭД в (1) определяется по формуле:

$$\Delta K = K_i - K_j, \quad (18)$$

где K_i — капитальные удельные затраты на выполнение реализаций документов до внедрения КСЭД (тыс. у. е.); K_j — капитальные удельные затраты на выполнение реализаций ЭД в КСЭД (тыс. у. е.).

Капитальные удельные затраты на выполнение реализаций документов до внедрения КСЭД рассчитываются следующим образом:

$$K_i = \frac{10^{-3} Q_{ii} Z_{ii}}{q} \cdot \frac{\Phi_i}{\Phi_j} + \frac{N_{2i}}{q N_i} (\Phi_{ii} - \Phi_i), \quad (19)$$

где q — продолжительность расчетного периода ($q = 1$ год); K_{ii} — цена конфигурации компьютерной техники до внедрения КСЭД (тыс. у. е.); Φ_i — стоимость основных фондов конфигурации компьютерной техники до внедрения КСЭД (тыс. у. е.); Φ_j — стоимость основных фондов, в которых эксплуатируется КСЭД (тыс. у. е.). Зачастую внедрение системы происходит без увеличения основных фондов (стоимость активной части) Φ_j . Тогда принимается, что $\Phi_i = \Phi_j$. Φ_{ii} — общая стоимость основных фондов организации (тыс. у. е.); N_i — общая численность сотрудников, выполняющих реализацию документов до внедрения КСЭД.

Капитальные удельные затраты на выполнение реализаций ЭД в КСЭД определяются по формуле:

$$K_j = \Phi_j \frac{10^{-3} C_j}{C_{2j} q}, \quad (20)$$

где Φ_j — общая стоимость основных фондов, построенных после внедрения КСЭД (стоимость активной части) (тыс. у. е.); C_{2j} — общая себестоимость годового объема работ, выполняемых средствами КСЭД (тыс. у. е./год). Если расчет производится для органов государственной власти, управляющих компаний холдингов, общественной организации, благотворительного фонда и пр., то расчет по (20) не производится.

Далее, рассчитываем общее число компьютерных рабочих мест (P_j), необходимых для внедрения КСЭД:

$$P_j = \frac{H_2}{D_2}, \quad (21)$$

и, наконец, число компьютерных рабочих мест (P), необходимых для приобретения:

$$P = P_j - P_i, \quad (22)$$

где P_i — количество компьютерных рабочих мест в организации до внедрения КСЭД.

Предпроизводственные затраты на внедрение КСЭД (K_{2j} , тыс. у. е.) определяются так:

$$K_{2j} = SP_j + D + R + Z_k M_k + G, \quad (23)$$

где S — стоимость одной лицензии КСЭД (тыс. у. е.). Обычно эта цифра находится в диапазоне от 100 до 1000 у. е. на рабочее место и в существенной степени зависит от количества закупаемых лицензий, то есть от количества рабочих мест в КСЭД; D — стоимость дополнительной техники, необходимой для внедрения КСЭД (компьютеры, принтеры, сканеры и т. д., а также, если необходимо, стоимость монтажа ЛВС, выход в Интернет и пр.) (тыс. у. е.) (см. (24)). Величина, в большинстве случаев практически пропорциональная N_j (среднесписочная численность сотрудников, выполняющих реализацию ЭД после внедрения системы). Такая пропорциональность соблюдается, если не требуется закупка дорогой специализированной техники, например, промышленных сканеров или крупных роботизированных библиотек для электронного архива; R — стоимость работ по внедрению системы (тыс. у. е.). Эта величина не пропорциональна размеру организации, но, естественно, находится в определенной зависимости от него (см. (25)). Стоимость внедрения в существенной степени зависит от того, как будет организовано внедрение, чьими силами оно будет проводиться, во сколько этапов; Z_k — стоимость консультанта в зависимости от условий предоставления услуг (тыс. у. е.) (от 2000 до 8000 у. е. в месяц); M_k — количество месяцев работы консультанта; G — годовая техническая поддержка по сопровождению и обслуживанию системы (тыс. у. е.) (исходя из экспертных оценок, в среднем можно считать 20% от стоимости лицензий).

Стоимость дополнительной техники, необходимой для внедрения КСЭД, найдем по выражению:

$$D = 10^{-3} \left(\begin{aligned} &PS_{2j} + K_{sj} S_j + K_{pj} S_{pj} + S_{Hj} + \\ &+ Q_k S_k + K_{kon} S_{kon} + P_j S_p + \\ &+ K_{2s} S_{2s} + K_{Mgb} S_{Mgb} + K_{2s} S_2 \end{aligned} \right), \quad (24)$$

где S_{2j} — стоимость 1 компьютера (у. е.). На время написания работы ~ 600 у. е.; K_{sj} — количество сканеров, необходимых для внедрения КСЭД; S_j — стоимость сканера; K_{pj} — количество принтеров; S_{pj} — стоимость принтера; S_{Hj} — стоимость концентраторов (для выхода в Интернет один стоимостью ~ 1000 у. е. и несколько, в зависимости от числа приобретаемых компьютеров стоимостью ~ 150 у. е.); Q_k — длина кабеля локальной сети (в зависимости от расположения основных фондов, в которых эксплуатируется КСЭД ~ 10–25 м на один компьютер); S_k — стоимость 1 м кабеля (~ 0,3 у. е.);

$K_{\text{кон}}$ — количество коннекторов в локальной сети (в среднем ~ на 5–8 м кабеля 1 коннектор); $S_{\text{кон}}$ — стоимость 1 коннектора (~ 0,2 у. е. каждый); $S_{\text{р}}$ — стоимость подключения одного рабочего места в локальной сети (~ 1,2 у. е.); K_{2S} — количество Интернет-серверов; S_{2S} — стоимость одного Интернет-сервера (у. е.); $K_{\text{Мгб}}$ — количество потребляемой информации из Интернета в Мгб/год; $S_{\text{Мгб}}$ — стоимость 1 Мгб информации из Интернета (~ 0,05 у. е.); S_2 — стоимость подключения к Интернету по выделенной линии на один сервер (~ 150–200 у. е.).

Стоимость работ по внедрению системы (R) можно оценить следующим образом:

$$R = MV(1 + 0,15 \lg P_j), \quad (25)$$

где M — это количество месяцев, нужное для осуществления работ по внедрению (обычно от 2 до 10 месяцев); V — средняя стоимость 1 месяца работ (включающая работы по проработке требований, созданию проекта внедрения, доработке ПО, инсталляции, обучению и пр.). В зависимости от типа КСЭД $\approx 1200 + 2000$ у. е./мес. Второе слагаемое в скобках — результат обработки зависимости поправочного коэффициента от объема организации, приведенного в [4].

В (23) годовую техническую поддержку по сопровождению и обслуживанию системы (тыс. у. е.) определим по соотношению:

$$G = 0,2SP_j. \quad (26)$$

Далее определяется расчетный коэффициент экономической эффективности, E_p (1/год):

$$E_p = \frac{\Delta C + \mathcal{E}_k}{\Delta K + K_{2j}}. \quad (27)$$

Срок окупаемости капитальных вложений $T_{\text{ок}}$ (лет), необходимый для внедрения КСЭД, находится из соотношения:

$$T_{\text{ок}} = \frac{1}{E_p}. \quad (28)$$

3. Предварительные исходные данные для расчета экономической эффективности внедрения КСЭД

Для упрощения расчета принимаем, что и $N_i = N_j$, и $Z_{1i} = Z_{2i} = Z_i = Z_j$, $\Phi_i = \Phi_j$, что практически не оказывает влияния на получаемые результаты. Принята работа по реализации ЭД в одну смену.

1. Срок окупаемости основных фондов (лет):

$$E_A = 3.$$

2. Норма амортизационных отчислений капитальных вложений в конфигурацию вычислительной техники и программное обеспечение (1/год):

$$E_{\text{р}} = 0,25.$$

3. Коэффициент накладных расходов и прочих прямых затрат (общепроизводственные плюс административные):

$$\Phi_i = \Phi_j = 1,5.$$

4. Среднечасовая заработная плата исполнителя, работающего на компьютерной технике и традиционным способом (у. е. чел./час):

$$Z_{1i} = Z_{2i} = Z_i = Z_j = 1,0.$$

5. Годовой объем документооборота на предприятии (док/год):

$$H_2 = 2000.$$

6. Процент реализаций ЭД, выполняющихся на собственной вычислительной технике до внедрения КСЭД:

$$A_i = 20 \%.$$

7. Количество компьютерных рабочих мест в организации, имеющих до внедрения КСЭД:

$$K_{1i} = 13.$$

8. Количество рабочих часов в расчетном году (час/год):

$$t = 2080.$$

9. Среднесписочная численность сотрудников, выполняющих реализацию документов традиционным способом до внедрения КСЭД:

$$N_{2i} = 60.$$

10. Параметр снижения экономии трудозатрат:

$$D_{\text{э}j} = 0,2.$$

11. Срок службы эксплуатации конфигурации вычислительной техники (лет):

$$T_j = 4.$$

12. Цена конфигурации компьютерной техники до внедрения КСЭД (тыс. у. е.):

$$K_{1i} = 8.$$

13. Стоимость основных фондов конфигурации компьютерной техники до внедрения КСЭД (тыс. у. е.):

$$\Phi_i = 90.$$

14. Общая стоимость основных фондов организации (тыс. у. е.):

$$\Phi_{1i} = 380.$$

15. Общая численность сотрудников, выполняющих реализацию документов до внедрения КСЭД:

$$N_i = 75.$$

16. Общая себестоимость годового объема работ, выполняемых средствами КСЭД (тыс. у. е./год):

$$C_{2j} = 400.$$

В этом объеме подготовка исходных данных для расчета экономической эффективности пока заканчивается. Нам нужно рассчитать количество

компьютерных рабочих мест, необходимых для приобретения, и затем возможно будет применить формулы (23)–(28).

4. Алгоритм предварительного расчета

1. По формуле (5) найдем количество выполнения реализаций ЭД на компьютерной технике до внедрения КСЭД:

$$H_{1i} = 2000 \cdot 20/100 = 400.$$

2. Количество выполнения реализаций документов традиционным способом до внедрения КСЭД получим из соотношения (9):

$$H_{2i} = 2000 - 400 = 1600.$$

3. Среднее количество часов, за которое выполняется 1 документ традиционным способом до внедрения КСЭД, получаем по выражению (10):

$$D_{2i} = 60 \cdot 2080 / 1600 = 78.$$

4. Среднее количество часов, за которое выполняется 1 ЭД после внедрения КСЭД, определяем по (16):

$$D_2 = 0,75 \cdot 78 = 58,5.$$

5. Рассчитываем общее количество компьютерных рабочих мест по (21), необходимых для внедрения КСЭД:

$$P_j = 2000 / 58,5 = 34.$$

Далее мы можем сформировать дополнительные исходные данные для окончательного расчета экономической эффективности внедрения КСЭД.

5. Окончательные исходные данные для расчета экономической эффективности внедрения КСЭД

17. Стоимость основных фондов, в которых эксплуатируется КСЭД (стоимость активной части) (тыс. у. е.):

$$\Phi_j = 185.$$

18. Стоимость одной лицензии КСЭД (тыс. у. е.):

$$S = 0,5.$$

19. Количество месяцев работы консультанта:

$$M_k = 5.$$

20. Стоимость консультанта (тыс. у. е. в месяц):

$$Z_k = 5,0.$$

21. Стоимость 1 компьютера (у. е.) при внедрении КСЭД. На момент написания работы:

$$S_{2j} = 600.$$

22. Количество сканеров, необходимых для внедрения КСЭД:

$$K_{sy} = 15.$$

23. Стоимость сканера (у. е.):

$$S_j = 60.$$

24. Количество принтеров:

$$K_{py} = 3.$$

25. Стоимость принтера (у. е.):

$$S_{py} = 100.$$

26. Стоимость концентраторов (у. е.):

$$S_{HI} = 1000 + 3 \cdot 150 = 1550.$$

27. Длина кабеля локальной сети:

$$Q_K = 15 \cdot 34 = 510.$$

28. Количество коннекторов:

$$K_{кон} = 153 / 6 = 26.$$

29. Количество Интернет-серверов:

$$K_{2S} = 2.$$

30. Стоимость одного Интернет-сервера (у. е.):

$$S_{2S} = 1000.$$

31. Количество потребляемой информации из Интернета в Мгб/год:

$$K_{Mgb} = 140000.$$

32. Количество месяцев, нужное для осуществления работ по внедрению:

$$M = 6.$$

33. Средняя стоимость 1 месяца работ (тыс. у. е.):

$$V = 1,5.$$

34. Стоимость подключения к Интернету по выделенной линии на один сервер (у. е.):

$$S_2 = 200.$$

6. Окончательный алгоритм расчета

6. Количество компьютерных рабочих мест, необходимых для приобретения, определим по (22):

$$P = 34 - 13 = 21.$$

7. По (24) рассчитаем затраты на внедрение КСЭД (тыс. у. е.):

$$D = 10^{-3} (21 \cdot 600 + 15 \cdot 60 + 3 \cdot 100 + 1550 + 510 \cdot 0,3 + 26 \cdot 0,2 + 34 \cdot 1,2 + 2 \cdot 1000 + 140000 \cdot 0,05 + 2 \cdot 200) = 24,95.$$

8. Стоимость работ по внедрению системы (тыс. у. е.) оцениваем по (25):

$$R = 6 \cdot 1,5(1 + \lg 34) = 22,8.$$

9. Годовую техническую поддержку по сопровождению и обслуживанию системы (тыс. у. е.) определяем по соотношению (26):

$$G = 0,2 \cdot 0,5 \cdot 34 = 3,4.$$

10. Предпроизводственные затраты на внедрение КСЭД (тыс. у. е.) определяем по (23):

$$K_{2j} = 0,5 \cdot 34 + 24,95 + 22,8 + 5 \cdot 5 + 3,4 = 93,15.$$

11. Трудозатраты на выполнение реализаций ЭД после внедрения КСЭД (чел./час.) находим по (15):

$$Q_j = 2000 \cdot 58,5 = 117000.$$

12. Себестоимость выполнения реализаций ЭД после внедрения КСЭД (у. е./реал.) рассчитаем по формуле (14):

$$C_j = 1,5 \cdot 117000 \cdot 1 = 175500.$$

13. По (17) рассчитаем годовую экономию от снижения эксплуатационных расходов и повышения качества решений в КСЭД (тыс. у. е./год):

$$\Delta_k = 10^{-3} \cdot 175500 \cdot 0,2 / (1/4 + 0,25) = 70,20.$$

14. Среднее количество часов, за которое выполняется 1 ЭД в организации до внедрения КСЭД, также находим из соотношения (6):

$$D_{1i} = 13 \cdot 2080 / 400 = 67,6.$$

15. Трудозатраты на выполнение реализаций ЭД на компьютерной технике до внедрения КСЭД (чел./час.) находим по формуле (7):

$$Q_{1i} = 13 \cdot 2080 = 27040.$$

16. Себестоимость выполнения реализаций всех документов до внедрения КСЭД (у. е./реал.) определим по формуле (12):

$$C_i = 1,5 \cdot 2080(13 \cdot 1 + 60 \cdot 1) = 227760.$$

17. Изменение (снижение) себестоимости выполнения реализаций документов в КСЭД (тыс. у. е./год) определим по соотношению (2):

$$\Delta C = 10^{-3}(227760 - 175500) = 52,26.$$

18. Капитальные удельные затраты на выполнение реализаций документов до внедрения КСЭД (тыс. у. е.) рассчитаем по формуле (19):

$$K_i = 10^{-3} \cdot 27040 \cdot 1 \cdot 90 / 185 + 60 \cdot (380 - 90) / 75 = 245,155.$$

19. Капитальные удельные затраты на выполнение реализаций ЭД в КСЭД определим по формуле (20):

$$K_j = 185 \cdot 10^{-3} \cdot 227760 / (400 \cdot 1) = 105,340.$$

20. Величину общего изменения капитальных затрат (капитальных дополнительных затрат) при внедрении КСЭД определим по соотношению (18):

$$\Delta K = 245,155 - 105,340 = 139,815.$$

21. Снижение себестоимости реализаций ЭД в расчетном году после внедрения КСЭД (тыс. у. е./год) определим по выражению (2):

$$\Delta C = 10^{-3} \cdot (227760 - 175500) = 52,260.$$

22. Годовую экономическую эффективность (тыс. у. е./год) рассчитаем по формуле (1):

$$\Theta_{\text{год}} = 52,26 + 70,2 - (139,815/3 + 93,15 \cdot 0,25) = 52,46.$$

23. Определяем расчетный коэффициент экономической эффективности по формуле (27):

$$E_p = (52,260 + 70,2) / (139,815 + 93,15) = 0,526.$$

24. Срок окупаемости капитальных вложений (лет), необходимый для внедрения КСЭД, находим из соотношения (28):

$$T_{\text{ок}} = 1 / 0,526 = 1,901.$$

Заключение

Научная новизна предложенного метода заключается в возможности учета при оценке эффективности КСЭД уже существующих до внедрения реализаций ЭД, а также экономии от повышения качества принимаемых решений.

Практическое значение заключается в возможности оценки общих затрат на внедрение и окупаемость КСЭД вне зависимости от направления деятельности и масштаба организации. Таким образом, использование приведенного метода расчета экономической эффективности на этапе проектирования КСЭД позволит оценить результативность различных вариантов построения системы, а также отдать предпочтение системе, наиболее полно реализующей необходимые функции с минимальными затратами и сроком окупаемости.

Список литературы: 1. Асеев Г.Г., Белова Т.Г. Оценка экономической эффективности внедрения корпоративных систем электронного документооборота. // АСУ и приборы автоматики. — 2005. — Вып. 130. — С. 32–40. 2. Козаченко В.Е. Управление общей стоимостью КИС [Электронный ресурс]. Режим доступа: <http://www.cs.comizdat.com>. Загл. с экрана. 3. Романов Д.А., Ильина Т.Н., Логинова А.Ю. Правда об электронном документообороте. — М.: ДМК Пресс, 2002. — 224 с. 4. Пацханян А. Внедрение систем электронного документооборота: проблемы и решения [Электронный ресурс] Директор ИС. 2002. № 01. Режим доступа: <http://www.osp.ru/cio/2002/01/028.htm>. Загл. с экрана. 5. Асеев Г.Г. Электронный документооборот. Учебн. — К.: Кондор, 2007. — 500 с.

Поступила в редколлегию 17.03.07

УДК 681.5.015

ИНТЕЛЛЕКТУАЛЬНЫЕ ТЕХНОЛОГИИ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ БИЗНЕС-ПРОЦЕССОВ

В. П. Авраменко¹, В. В. Калачева²

¹Харьковский национальный университет радиоэлектроники

²Харьковский университет Воздушных Сил им. И. Кожедуба

Исследовано применение интеллектуальных технологий для поддержки принятия управленческих решений по повышению эффективности бизнес-процессов. Разработаны математические модели и регуляризованные процедуры эффективного управления бизнес-процессами.

ИНТЕЛЛЕКТУАЛЬНЫЕ ТЕХНОЛОГИИ. ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ, БИЗНЕС-ПРОЦЕССЫ, ПРИНЯТИЕ ОПТИМАЛЬНЫХ РЕШЕНИЙ.

Введение

Важная роль в повышении эффективности бизнес-процессов отводится средствам искусственного интеллекта, которые позволяют повысить качество принимаемых решений в условиях слабой структурированности задачи принятия решений, а также неопределенности цели, условий функционирования и проявления внешней среды.

Источниками слабой структурированности и неопределенности задач повышения эффективности бизнес-процессов являются такие факторы [1–3]: недостаточное количество информации при выборе решения, несогласованность и противоречивость ограничений, дрейф характеристик, отклонение исходных предпосылок от фактических, слабая формализуемость процессов, плохая обусловленность системы, варьирование исходных данных и т. п.

Решения, принимаемые в условиях неопределенности исходных данных, и проявления внешней среды приводят к худшей эффективности бизнес-процессов, чем при полной определенности. В этом случае принимаются квазиоптимальные решения, т. е. не безусловно лучшие решения, а лучшие в некотором смысле, например, в среднем, или близкие к некоторому предпочтительному решению с точки зрения конкретного лица, принимающего решения.

При формализации задачи моделирования и оптимизации бизнес-процессов предпринимаются попытки снизить меру неопределенности путем привлечения дополнительной информации. Фактически осуществляется переход от полной неопределенности к частичной неопределенности, что приводит к множеству допустимых квазиоптимальных решений.

Целью данных научных исследований является анализ состояния и выбор перспективных интеллектуальных технологий (моделей, методов и процедур) для поддержки принятия допустимых предпочтительных квазиоптимальных решений, направленных на повышение эффективности функционирования бизнес-процессов.

1. Постановка задачи повышения эффективности бизнес-процессов

Анализ эффективности бизнес-процессов начинается с выяснения вопроса, к какому типу моделей принадлежит описание бизнеса. На сегодняшний день известны три основных способа описания бизнеса: функциональный, процессный и объектный. Функциональное описание является традиционным и широко применяется при иерархическом управлении производством. Им руководствуются уже более ста лет, однако в последнее десятилетие у него появилась конкурентоспособная альтернатива в виде процессного подхода.

Процессный подход считается более эффективным по сравнению с функциональным подходом с точки зрения конкурентоспособности в условиях рыночной экономики. В основу процессного подхода положена идея непрерывного эволюционного и скачкообразного революционного повышения качества всех аспектов деятельности организации. Время объектного описания бизнеса еще не пришло, хотя основы объектного программного обеспечения уже разработаны и успешно применяются.

В реальных ситуациях присутствуют все три составляющие описания бизнеса. Чаще других используется описание в пространстве «функции — процессы». Функции «разрезают» организацию на полосы сверху донизу, а процессы «разрезают» организацию поперек, пересекая границы функциональных подразделений. Преобладание процессного подхода существенно меняет логику организации и механизм повышения эффективности.

При функциональном подходе каждое подразделение пытается оптимизировать уровень собственных технико-экономических показателей, что ведет к конфликту между целями и конфликту принимаемых решений. Применение процессного подхода разрушает барьеры между подразделениями, которые являются одним из главных «врагов» совершенствования бизнеса.

Для повышения эффективности бизнес-процессов предлагаются различные интеллектуальные тех-

нологии, такие как эволюционное и адаптивное моделирование, методы внешнего дополнения и регуляризации процедур решения многокритериальных логически связанных задач. За организацией в целом остается интеллектуализация синергизма бизнес-процессов, состоящая в том, чтобы результаты каждой составляющей процесса способствовали улучшению результатов всех остальных составляющих бизнес-процесса.

2. Результаты выполненных исследований

Интеллектуальные технологии принятия эффективных решений. Идея любой интеллектуальной технологии поддержки принятия решений заключается в том, чтобы отыскать «наилучшее решение» как альтернативу, обладающую максимальной полезностью, которая окажется действительно наилучшей при реализации конкретного бизнес-процесса.

С точки зрения исследования операций «наилучших альтернатив» может оказаться несколько, если максимум функции полезности достигается в нескольких элементах множества допустимых решений. Выбор конкретной «наилучшей альтернативы» из множества «допустимых лучших» осуществляется лицом, принимающим решение (ЛПР), с привлечением предпочтительных интеллектуальных технологий.

В компетенцию ЛПР входит проведение анализа показателей эффективности бизнес-процессов, принятие надлежащих решений о совершенствовании бизнеса, осуществление всевозможных согласований, выдача руководящих указаний основным исполнителям и личный контроль выполняемой работы. Только ЛПР имеет право/обязанность в форс-мажорных ситуациях немедленно принимать решения / отдавать распоряжения на задействование резервов и оказание помощи исполнителям.

Все крупные бизнес-процессы возглавляет владелец, который наделяется полномочиями по управлению всеми ресурсами и всеми видами активности процесса. Это позволяет сформулировать оперативные, тактические и стратегические цели ведения бизнеса. При формулировании целей бизнеса используются интеллектуальные процедуры бенчмаркинга (постоянного измерения и сравнения отдельных бизнес-процессов с эталонным процессом ведущей организации с целью их непрерывного совершенствования).

Возможности использования интеллектуальных технологий повышения эффективности функционирования бизнес-процессов гораздо шире, чем может показаться на первый взгляд. Дело в том, что всякий бизнес-процесс обладает свойством «фрактальности». Это свойство заключается в том, что

почти любая часть процесса может рассматриваться как процесс. «Фрактальностью» можно воспользоваться для работы с процессами любого масштаба, от вселенских глобальных до индивидуальных.

Интеллектуальные технологии принятия оптимальных решений. Задача повышения эффективности функционирования бизнес-систем как совокупности логически связанных бизнес-процессов в общем случае принадлежит к классу задач нелинейного стохастического программирования

$$S^* = \arg \operatorname{extr}_{S \in S_D} E \left\{ \sum_{i=1}^k \lambda_i [Q_i(S) - Q_i^*] / Q_i^* \right\}, \quad (1)$$

где S_D — область допустимых решений, удовлетворяющая уравнениям материального баланса и последовательности преобразования информации; Q_1, Q_2, Q_3 — показатели качества выполняемых функций; λ_i — весовые коэффициенты критериев, причем $\sum_i \lambda_i = 1$; $E\{\cdot\}$ — оператор стохастического усреднения.

Задачи моделирования бизнес-процессов часто сводятся к тому, чтобы найти такое значение вектора $x^* \in \Omega_x \subseteq R^n$, которое доставляет экстремум (например, максимум) частным критериям $y_i = \varphi_i(x)$, $i = \overline{1, k}$, удерживая остальные y_j , $j = \overline{k+1, m}$ на некотором уровне B_j , $j = \overline{k+1, m}$:

$$Q(x) = \{\varphi_1(x), \dots, \varphi_k(x)\} \rightarrow \max_{x \in R^n}, \quad (2)$$

$$R_j(x) = \varphi_j(x) \geq B_j, \quad j = \overline{k+1, m}, \quad (3)$$

$$a_j(x_j) \leq x_j \leq b_j(x_j), \quad x_j \geq 0, \quad j = \overline{1, n}, \quad (4)$$

$$\varphi_j(x) \geq B_j, \quad \varphi_\ell \leq B_\ell. \quad (5)$$

Задача повышения эффективности бизнес-процессов в условиях неопределенности может быть представлена моделью:

$$Q(x, \gamma_1) = \{\varphi_1(x), \dots, \varphi_k(x); \gamma_1\} \rightarrow \max_{x \in D}, \quad (6)$$

$$D(x, \gamma_2, \gamma_3) = \{x \in R^n \mid R_j = \varphi_j(x, \alpha; \gamma_2) \geq B_j, \\ a_j(x_j, \gamma_3) \leq x_j \leq b_j(x_j, \gamma_3), \quad x_j \geq 0, \quad j = \overline{1, n}\}, \quad (7)$$

где γ_1 — компонента, отражающая неопределенность от свертки критериев; γ_2 — случайная компонента, отражающая варьирование ограничений; γ_3 — детерминированная компонента от изменения исходных данных.

Модель распределения ресурсов (6) — (7) принадлежит к классу задач стохастического программирования, которую при решении желательно

редуцировать совокупностью детерминированных задач линейного и нелинейного программирования при фиксированных значениях случайного вектора $\gamma = \{\gamma^\ell\}$, где $\ell = 1, 2, \dots$. В результате редукции задачи (6) — (7) получается совокупность детерминированных задач:

$$Q(x, \gamma_1^k) = \{\varphi_1(x), \dots, \varphi_k(x); \gamma_1^k\} \rightarrow \max_{x \in D}, \quad (8)$$

$$D(x; \gamma_2^k, \gamma_3^k) = \{x \in R^n \mid \varphi_i(x, a, \gamma_2^k) \geq B_i, \\ a_j(x, \gamma_3^k) \leq x_j \leq b_j(x, \gamma_3^k), x_j \geq 0, j = \overline{1, n}\}, \quad (9)$$

решения которых образуют совокупность $\{x^\ell\}$. Выбор предпочтительного решения из совокупности квазиоптимальных осуществляется с помощью интеллектуальных процедур лицом, принимающим решение.

При варьировании параметров модели в некотором диапазоне задача распределения ресурсов в классе линейного программирования

$$\min_{x \in X} \{c, x\}, Ax \geq b, x \in E_+^n \quad (10)$$

заменяется задачей параметрического программирования, близкой к исходной:

$$\min_{x \in X} \{c_\delta, x\}, A_\delta x \geq b_\delta, x \in E_+^n, \quad (11)$$

где $c_\delta, A_\delta, b_\delta$ — возмущенные значения параметров, близкие к истинным c, A, b в некоторой метрике, например, в евклидовой:

$$\|c_\delta - c\| < \delta, \|A_\delta - A\| < \delta, \|b_\delta - b\| < \delta \quad (12)$$

(величина погрешности δ может быть принята одинаковой для всех параметров, что не влияет на общность получаемых результатов).

Получаемые локально оптимальные эффективные решения оцениваются по всей совокупности локальных критериев, и из этого множества выделяются решения, оптимальные по Парето.

Регуляризирующие процедуры принятия оптимальных решений. Большинство задач повышения эффективности бизнес-процессов в лучшем случае являются слабо корректными в силу слабой структурированности. В них отсутствует достоверная информация о непротиворечивости ограничений, характере возмущающих воздействий и погрешностях вычислений.

В качестве регуляризованного решения задачи можно принять нормальное решение x_n , наименее уклоняющееся от некоторого заданного вектора x_0 , представляющего собой показатели производствен-

ной программы. Чем ближе искомое решение x_n к показателям производственной программы, тем эффективнее полученное решение. Мету уклонения нового решения от старого можно задать квадратом нормы

$$\Omega[x_n - x_0] = \|x_n - x_0\|_n^2. \quad (13)$$

Предположим, что задан некоторый вектор $x_0 \in X$. Искомый вектор x_n представляет нормальное решение задачи линейного программирования (по отношению к x_0), если справедливо соотношение

$$\|x_n - x_0\|_n^2 = \min_{x^* \in X} \|x^* - x_0\|_n^2, \quad (14)$$

где x^* — любое решение этой задачи.

Из совокупности квазиоптимальных решений с помощью интеллектуальной технологии выделяется нормальное решение, наилучшее в смысле выбранной функции уклонения.

Заключение

Многокритериальные интеллектуальные технологии повышения эффективности бизнес-процессов принято относить к классу многократно некорректных задач. Их некорректность возникает, во-первых, из-за некорректности задач локальной оптимизации, во-вторых, из-за процедур принятия многокритериальных решений, в основу которых положен принцип неединственности. Множественность эффективных решений бизнес-процессов является скорее достоинством, а не недостатком, поскольку «жесткие» схемы получения единственного решения неадекватны сущности задач многокритериальной оптимизации, а имеющаяся интеллектуальная «свобода» выбора предпочтительного решения из множества эффективных позволяет учесть неопределенность целей и критериев [4–6].

Список литературы: 1. Авраменко В. П. Концепции интеллектуализации процедур принятия управленческих решений в условиях неопределенности // АСУ и приборы автоматизации. — Вып. 108. — 1998. — С. 42–58. 2. Авраменко В. П., Калачев И. Д., Калачева В. В. Интеллектуальные процедуры принятия решений в условиях неопределенности // Проблемы бионики. — Вып. 48. — 1998. — С. 124–128. 3. Андерсен Б. Бизнес-процессы. Инструменты совершенствования. — М.: Стандарты и качество, 2004. — 272 с. 4. Андрейчиков А. В., Андрейчикова О. Н. Интеллектуальные информационные системы. — М.: Финансы и статистика, 2004. — 424 с. 5. Тельнов Ю. Ф. Интеллектуальные информационные системы в экономике. — М.: СИНТЕГ, 1999. — 216 с. 6. Усков А. А., Кузьмин А. В. Интеллектуальные технологии управления. Искусственные нейронные сети и нечеткая логика. — М.: Горячая линия-Телеком, 2004. — 143 с.

Поступила в редколлегию 20.03.07

ЭКСТРЕМАЛЬНЫЕ СВОЙСТВА ФУНКЦИЙ НА КЛАССАХ КОМПОЗИЦИОННЫХ ОБРАЗОВ КОМБИНАТОРНЫХ МНОЖЕСТВ

И. В. Гребенник¹, А. В. Баранов²

¹ ХНУРЭ, г. Харьков, Украина, stdep@kture.kharkov.ua

² ХНУРЭ, г. Харьков, Украина, Aleksey.Baranov@gmail.com

Исследуются задачи оптимизации выпуклых и сильно выпуклых функций на классах комбинаторных множеств. Формулируются оценки и достаточные условия минимума функций на множествах перестановок кортежей и композиции перестановок. Обсуждается применение результатов при разработке интеллектуальных систем поддержки принятия решений различного назначения.

ПРИНЯТИЕ РЕШЕНИЙ, КОМБИНАТОРНАЯ ОПТИМИЗАЦИЯ, КОМБИНАТОРНОЕ МНОЖЕСТВО, ВЫПУКЛАЯ ФУНКЦИЯ, ОЦЕНКА МИНИМУМА.

Введение

При разработке интеллектуальных систем поддержки принятия решений различного назначения важную роль играют математические модели объектов, составляющих предметную область, и методы их анализа. Применение математических моделей и методов позволяет повысить эффективность решений, принимаемых с помощью интеллектуальных систем. В настоящее время системы поддержки принятия решений создаются в различных областях, одной из которых является геометрическое проектирование [1].

Многие задачи проектирования, управления, контроля описываются моделями комбинаторной оптимизации [2]. Области допустимых решений этих задач часто представляются классическими комбинаторными множествами [3]. Комбинаторные свойства новых задач таковы, что классические комбинаторные множества не всегда позволяют адекватно описывать соответствующие математические модели задач. Значит, необходимо вводить новые комбинаторные множества как средства построения математических моделей задач указанных классов. Это справедливо, в частности, при решении многих комбинаторных оптимизационных задач геометрического проектирования [1, 4].

Для построения математических моделей задач указанного класса в [5] вводится новый класс комбинаторных множеств — композиционные k -образы комбинаторных множеств. В связи с этим актуальной задачей является построение и анализ различных классов оптимизационных моделей на композиционных k -образах комбинаторных множеств.

Целью настоящей работы является исследование свойств задач оптимизации выпуклых функций на классах композиционных k -образов комбинаторных множеств.

1. Определение множеств перестановок кортежей и композиции перестановок [6, 7]

Композиционный образ комбинаторных множеств $P_{nk}, T_1, T_2, \dots, T_n$, порожденный множествами

$\{z_1^1, z_2^1, \dots, z_m^1\}, \{z_1^2, z_2^2, \dots, z_m^2\}, \dots, \{z_1^n, z_2^n, \dots, z_m^n\}$. Здесь $T_i = \{(z_1^i, z_2^i, \dots, z_m^i)\}$ — кортеж, составленный из элементов множества $\{z_1^i, z_2^i, \dots, z_m^i\}$, $z_j^i \in R$, $i \in J_n = \{1, 2, \dots, n\}$, $j \in J_m$. При этом среди n множеств T_i являются различными. Обозначим это множество через $PT_{nk}(T_1, T_2, \dots, T_n)$ или PT_{nk}^m и назовем *множеством перестановок кортежей*. Множество PT_{nk}^m представляет собой множество перестановок кортежей $z^i = (z_1^i, z_2^i, \dots, z_m^i)$, то есть упорядоченных наборов вида $w \in PT_{nk}^m$, $w = (z^i, z^j, \dots, z^h) = (z_1^i, z_1^j, \dots, z_1^h, z_2^i, z_2^j, \dots, z_2^h, \dots, z_m^i, z_m^j, \dots, z_m^h)$, где $i_s, j_s \in J_n$, $i_s \neq j_s$, $s \in J_n$. Элементы множества PT_{nk}^m отличаются друг от друга только порядком следования кортежей z^i в наборах.

Композиционный образ комбинаторных множеств $P_{nk}, P_{m_1 k_1}, P_{m_2 k_2}, \dots, P_{m_n k_n}$ порожденный множествами $\{a_1^1, a_2^1, \dots, a_{m_1}^1\}, \{a_1^2, a_2^2, \dots, a_{m_2}^2\}, \dots, \{a_1^n, a_2^n, \dots, a_{m_n}^n\}$. Здесь P_{nk} — множество перестановок из n элементов, из которых различными являются $a_j^i \in R^1$, $i \in J_m, j \in J_n$. Такое множество назовем *композицией перестановок* и обозначим PW_N . Множество PW_N состоит из элементов вида (e_1, e_2, \dots, e_n) , где $(i_1, i_2, \dots, i_n) \in L_n$, $e_i = (a_{s_1}^i, a_{s_2}^i, \dots, a_{s_{m_i}}^i)$, $i \in J_n$. В наборе (e_1, e_2, \dots, e_n) k элементов являются различными, среди элементов $a_{s_1}^j, a_{s_2}^j, \dots, a_{s_{m_j}}^j$ ровно k_j различных. Последовательность индексов $(s_1, s_2, \dots, s_{m_j}) \in L_{m_j}$, где через L_k обозначим множество всевозможных перестановок элементов индексного множества J_k .

2. Постановка задачи

Пусть X — композиционный образ комбинаторных множеств [5], $X \in \{PT_{nk}^m, PW_N\}$. Рассмотрим задачу оптимизации

$$\Phi(h) \rightarrow \min, h \in X, \quad (1)$$

где $\Phi: X \rightarrow R^1$ — некоторый функционал.

В результате погружения f множества X в R^N [1, 4] сформулируем задачу оптимизации функции $\varphi(x)$, эквивалентную (1):

$$\varphi(x) \rightarrow \min, \quad x \in E, \quad (2)$$

где $\varphi: E \rightarrow R^1$ — функция N переменных, определенная на множестве $E \subset R^N$, $E = f(X)$ — образ множества X в пространстве R^N , $\varphi(x) = \Phi(h)$ при $x = f(h)$, $\forall h \in H$.

После погружения координаты точек множества $ET_{nk}^m = f(PT_{nk}^m)$ принимают значения всевозможных перестановок кортежей $z^i = (z_1^i, z_2^i, \dots, z_m^i)$, $i \in J_n$. Множество $EW_N = f(PW_N)$ представляет собой множество векторов в R^N , координаты которых принимают значения всевозможных перестановок векторов $e_i = (e_{s_1}^i, e_{s_2}^i, \dots, e_{s_m}^i) = (a_{s_1}^i, a_{s_2}^i, \dots, a_{s_m}^i)$, где $(a_{s_1}^i, a_{s_2}^i, \dots, a_{s_m}^i) \in E_{m; k_i}$, а множество перестановок $E_{m; k_i} = f(P_{m; k_i})$ порождено элементами $a_1^i \leq a_2^i \leq \dots \leq a_m^i$ [4].

В дальнейших построениях будем считать, что для множеств PT_{nk}^m , PW_N справедливо равенство $m_1 = m_2 = \dots = m_n = m$.

Исследуем случай, когда в задаче (2) $\varphi(x)$ — выпуклая (сильно выпуклая с параметром $\rho > 0$) на выпуклом множестве $V \supset \text{conv} E$ функция, а в качестве множества E выступают образы множеств перестановок кортежей PT_{nk}^m , композиции перестановок PW_N в пространстве R^N — соответственно, множества ET_{nk}^m и EW_N .

Вопросы исследования и решения задач оптимизации выпуклых и сильно выпуклых функций на евклидовых комбинаторных множествах рассматривались во многих публикациях. Так, работы [8–12] посвящены описанию теории и конструктивных методов построения выпуклых и сильно выпуклых дифференцируемых продолжений для классов функций, заданных на различных комбинаторных множествах. В работах [4, 8, 13–17] исследуются декомпозиционные методы, использующие оценки минимума выпуклых целевых функций на евклидовых комбинаторных множествах и их подмножествах.

Распространим некоторые результаты, приведенные в этих работах, на множества ET_{nk}^m и EW_N .

3. Оценки минимума выпуклых функций на множествах ET_{nk}^m , EW_N

В работах [4, 8, 9, 13–17] излагается общий подход к построению оценок минимума выпуклых и сильно выпуклых функций на евклидовых комбинаторных множествах. В рамках этого подхода получены оценки минимума выпуклых и сильно выпуклых продолжений функций, заданных на евклидовых комбинаторных множествах перестановок, размещений, сочетаний и других. Эти резуль-

таты используются при реализации декомпозиционных методов решения задач оптимизации на указанных классах комбинаторных множеств. Применяя данный подход, построим оценки и достаточные условия минимума выпуклых и сильно выпуклых функций на множествах перестановок кортежей и композиции перестановок. При этом будем опираться на утверждения, доказанные в [17].

Лемма 1 [17]. Пусть функция $\varphi(x)$ выпукла и дифференцируема на выпуклом множестве V , где $E \subset V \subset R^N$. Тогда для любого $x \in V$:

$$\min_{y \in E} \varphi(y) \geq \varphi(x) - (\nabla \varphi(x), x) + \min_{y \in E} \sum_{i=1}^N \frac{\partial \varphi(x)}{\partial x_i} y_i. \quad (3)$$

Теорема 1. Пусть функция $\varphi(x)$ выпукла и дифференцируема на выпуклом множестве $V \supset ET_{nk}^m$, где множество $ET_{nk}^m \subset R^N$, $N = mn$ порождено множествами $T_i = \{(z_1^i, z_2^i, \dots, z_m^i)\}$, $i \in J_n$. Тогда для любого $x \in V$:

$$\min_{y \in ET_{nk}^m} \varphi(y) \geq \varphi(x) - (\nabla \varphi(x), x) + \sum_{j=1}^w \frac{\partial \varphi(x)}{\partial x_j} y_j^*, \quad (4)$$

где

$$y^* = (y_1^*, y_2^*, \dots, y_N^*) \in ET_{nk}^m, \quad (5)$$

здесь $y_{(j-1)m+t}^* = z_t^{i_j}$, $i_j \neq i_s$ при $s \neq j$, $i_j, i_s \in J_n$, $j \in J_n$, $s \in J_n$, $t \in J_m$, а последовательность $\{i_1, i_2, \dots, i_n\}$ такова, что

$$z_1^{i_1} \prec_c z_2^{i_2} \prec_c \dots \prec_c z_m^{i_n} \quad (6)$$

при $c = \nabla \varphi(x)$.

Доказательство. Оценку (4) получим из соотношения (3) при $E = ET_{nk}^m$, $N = mn$. Тогда задача оптимизации в правой части неравенства (3) может быть решена с помощью утверждения, доказанного в [6], на основе предложенного отношения линейного порядка \prec_c :

$$\begin{aligned} & ((z_1^{i_j}, z_2^{i_j}, \dots, z_m^{i_j}) \prec_c (z_1^{i_k}, z_2^{i_k}, \dots, z_m^{i_k})) \Leftrightarrow \\ & \Leftrightarrow \left(\sum_{t=1}^m (c_{(j-1)m+t} - c_{(k-1)m+t}) (z_t^{i_j} - z_t^{i_k}) \leq 0 \right). \end{aligned}$$

При этом минимизируемая функция является

линейной с коэффициентами $c_i = \frac{\partial \varphi(x)}{\partial x_i}$, $i \in J_N$. Упорядочивая кортежи z^j , $j \in J_n$ в соответствии с соотношением (6) при $c = \nabla \varphi(x)$, приходим к справедливости утверждения теоремы.

Теорема 2. Пусть функция $\varphi(x)$ выпукла и дифференцируема на выпуклом множестве $V \supset EW_N$, где множество $EW_N \subset R^N$, $N = nm$ порождено множествами $\{a_1^i, a_2^i, \dots, a_m^i\}$, $i \in J_n$. Тогда для любого $x \in V$:

$$\min_{y \in EW_N} \varphi(y) \geq \varphi(x) - (\nabla \varphi(x), x) + \sum_{j=1}^N \frac{\partial \varphi(x)}{\partial x_j} y_j^*, \quad (7)$$

где

$$y^* = (y_1^*, y_2^*, \dots, y_N^*) \in EW_N. \quad (8)$$

$y_{(j-1)m+r_j}^* = e_{s_j}^{i_j}$, $i \in J_m$, $j \in J_n$, $\{r_1, r_2, \dots, r_m\}$ и $\{s_1, s_2, \dots, s_m\}$ таковы, что $c_{(j-1)m+s_1} \geq c_{(j-1)m+s_2} \geq \dots \geq c_{(j-1)m+s_m}$ и $e_{r_1}^{i_1} \leq e_{r_2}^{i_2} \leq \dots \leq e_{r_m}^{i_m}$, а последовательность $\{i_1, i_2, \dots, i_m\}$ удовлетворяет условию

$$e_{i_1} \prec_c e_{i_2} \prec_c \dots \prec_c e_{i_m} \quad (9)$$

при $c = \nabla\varphi(x)$.

Доказательство. Оценку (7) получим из соотношения (3) при $E = EW_N$, $N = mn$. Задача оптимизации в правой части неравенства (3) может быть решена с помощью теоремы о минимуме $EW_N \subset R^N$ [18] на основе предложенного отношения линейного порядка \prec_c , поскольку функция цели является линейной с коэффициентами

$$c_i = \frac{\partial\varphi(x)}{\partial x_i}, \quad i \in J_N.$$

Здесь

$$\begin{aligned} e_{i_j} \prec_c e_{i_k} &\Leftrightarrow (e_1^{i_j}, e_2^{i_j}, \dots, e_m^{i_j}) \prec_c (e_1^{i_k}, e_2^{i_k}, \dots, e_m^{i_k}) \Leftrightarrow \\ &\Leftrightarrow \left(\sum_{i=1}^m c_{(j-1)m+s_i} e_{p_i}^{i_j} + \sum_{i=1}^m c_{(k-1)m+r_i} e_{q_i}^{i_k} - \right. \\ &\left. - \sum_{i=1}^m c_{(j-1)m+\alpha_i} e_{r_i}^{i_k} - \sum_{i=1}^m c_{(k-1)m+\beta_i} e_{\delta_i}^{i_j} \right) \leq 0, \quad (10) \end{aligned}$$

где $i \in J_n$, а последовательность индексов удовлетворяет условиям:

$$\begin{aligned} \{s_1, s_2, \dots, s_m\} &: c_{(j-1)m+s_1} \geq c_{(j-1)m+s_2} \geq \dots \geq c_{(j-1)m+s_m}, \\ \{p_1, p_2, \dots, p_m\} &: e_{p_1}^{i_j} \leq e_{p_2}^{i_j} \leq \dots \leq e_{p_m}^{i_j}, \\ \{q_1, q_2, \dots, q_m\} &: e_{q_1}^{i_k} \leq e_{q_2}^{i_k} \leq \dots \leq e_{q_m}^{i_k}, \\ \{r_1, r_2, \dots, r_m\} &: c_{(k-1)m+r_1} \geq c_{(k-1)m+r_2} \geq \dots \geq c_{(k-1)m+r_m}, \\ \{\alpha_1, \alpha_2, \dots, \alpha_m\} &: c_{(j-1)m-\alpha_1} \geq c_{(j-1)m+\alpha_2} \geq \dots \geq c_{(j-1)m+\alpha_m}, \\ \{\gamma_1, \gamma_2, \dots, \gamma_m\} &: e_{\gamma_1}^{i_j} \leq e_{\gamma_2}^{i_j} \leq \dots \leq e_{\gamma_m}^{i_j}, \\ \{\delta_1, \delta_2, \dots, \delta_m\} &: e_{\delta_1}^{i_k} \leq e_{\delta_2}^{i_k} \leq \dots \leq e_{\delta_m}^{i_k}, \\ \{\beta_1, \beta_2, \dots, \beta_m\} &: c_{(k-1)m+\beta_1} \geq c_{(k-1)m+\beta_2} \geq \dots \geq c_{(k-1)m+\beta_m}. \end{aligned}$$

Лемма 2 [17]. Для того, чтобы точка $x^* = (x_1^*, x_2^*, \dots, x_N^*) \in E$ была точкой минимума на множестве E выпуклой дифференцируемой на выпуклом множестве V функции $\varphi(x)$, где $E \subset V \subset R^N$, достаточно, чтобы

$$\min_{y \in E} \sum_{i=1}^N \frac{\partial\varphi(x^*)}{\partial x_i} y_i - (\nabla\varphi(x^*), x^*) = 0. \quad (11)$$

Теорема 3. Пусть функция $\varphi(x)$ выпукла и дифференцируема на выпуклом множестве $V \supset ET_{nk}^m$, где множество $ET_{nk}^m \subset R^N$, $N = mn$ порождено множествами $T_i = \{(z_1^i, z_2^i, \dots, z_m^i)\}$, $i \in J_n$. Для того, чтобы точка $x^* \in ET_{nk}^m$ была точкой минимума $\varphi(x)$ на ET_{nk}^m , достаточно, чтобы

$$\sum_{i=1}^N \frac{\partial\varphi(x^*)}{\partial x_i} y_i^* - (\nabla\varphi(x^*), x^*) = 0, \quad (12)$$

где $y^* \in ET_{nk}^m$ определяется соотношениями (5).

Доказательство достаточного условия получим на основании леммы 2 при $E = ET_{nk}^m$, $N = nm$. Задача минимизации в левой части равенства (11) может быть решена на основе теоремы о минимуме линейной функции на множестве ET_{nk}^m [6] с помощью подхода, аналогичного примененному при доказательстве теоремы 1. В результате приходим к справедливости равенства (12) и утверждения теоремы.

Теорема 4. Пусть функция $\varphi(x)$ выпукла и дифференцируема на выпуклом множестве $V \supset EW_N$, где множество $EW_N \subset R^N$, $N = nm$ порождено множествами $\{a_1^i, a_2^i, \dots, a_m^i\}$, $i \in J_n$. Для того, чтобы точка $x^* \in EI_{nk}^m$ была точкой минимума $\varphi(x)$ на EW_N , достаточно, чтобы

$$\sum_{i=1}^N \frac{\partial\varphi(x^*)}{\partial x_i} y_i^* - (\nabla\varphi(x^*), x^*) = 0, \quad (13)$$

где $y^* \in EW_N$ определяется соотношениями (8).

Доказательство утверждения проведем на основании леммы 2 при $E = EW_N$, $N = nm$. Задача минимизации в левой части равенства (11) может быть решена на основе теоремы о минимуме линейной функции на множестве EW_N [18] с помощью подхода, аналогичного примененному при доказательстве теоремы 2. В результате приходим к справедливости равенства (13) и утверждения теоремы.

Рассмотрим теперь случай, когда функция $\varphi(x)$ является сильно выпуклой с параметром $\rho > 0$ на выпуклом замкнутом множестве $V \subset R^N$, $E \subset V$. Обозначим

$$y^* = \arg \min_{y \in V} \varphi(y). \quad (14)$$

Лемма 3 [17]. Пусть функция $\varphi(x)$ сильно выпукла с параметром $\rho > 0$ на выпуклом замкнутом множестве V , где $E \subset V \subset R^N$. Тогда

$$\min_{x \in E} \varphi(x) \geq \varphi(y^*) + \rho \cdot \min_{y \in E} \|x - y^*\|^2, \quad (15)$$

где y^* определяется соотношением (14).

Теорема 5. Пусть функция $\varphi(x)$ сильно выпукла с параметром $\rho > 0$ на выпуклом замкнутом множестве $V \supset ET_{nk}^m$, где множество $ET_{nk}^m \subset R^N$, $N = nm$

порождено множествами $T_i = \{(z_1^i, z_2^i, \dots, z_m^i)\}$, $i \in J_n$. Тогда

$$\min_{x \in ET_{nk}^m} \varphi(x) \geq \varphi(y^*) + \rho \left(\sum_{i=1}^n \sum_{j=1}^m (z_j^i)^2 + \sum_{i=1}^N (y_i^*)^2 + 2 \sum_{i=1}^N c_i^* x_i^0 \right), \quad (16)$$

где $c_i^* = -y_i^*$, $i \in J_w$, $x^0 \in ET_{nk}^m$, $x_{(j-1)m+t}^0 = z_t^j$, $i_j \neq i_s$ при $s \neq j$, $i_j, i_s \in J_n$, $j \in J_n$, $s \in J_n$, $t \in J_m$, а последовательность $\{i_1, i_2, \dots, i_n\}$ такова, что

$$z_1^{i_1} \prec_c^* z_2^{i_2} \prec_c^* \dots \prec_c^* z_n^{i_n}. \quad (17)$$

Доказательство. Оценку (16) получим на основании леммы 3 при $E = ET_{nk}^m$, $N = nm$. Задачу об определении минимума в правой части неравенства (15) решим с помощью соотношения, полученного в [6]. Непосредственная подстановка минимального значения нормы $\|x - y^*\|^2$ на множестве ET_{nk}^m в (15) приведет к справедливости оценки (16).

Теорема 6. Пусть функция $\varphi(x)$ сильно выпукла с параметром $\rho > 0$ на выпуклом замкнутом множестве $V \subset EW_N$, где множество $EW_N \subset R^N$, $N = nm$ порождено множествами $\{a_1^i, a_2^i, \dots, a_m^i\}$, $i \in J_n$. Тогда

$$\min_{x \in EW_N} \varphi(x) \geq \varphi(y^*) + \rho \left(\sum_{i=1}^n \sum_{j=1}^m (a_j^i)^2 + \sum_{i=1}^N y_i^{*2} + 2 \sum_{i=1}^N d_i^* y_i^0 \right), \quad (18)$$

где $d_i^* = -y_i^*$, $i \in J_N$, а $y^0 \in EW_N$ определяется как $y_{(j-1)m+t}^0 = e_{s_j}^j$, $i_j \neq i_s$ при $s \neq j$, $i_j, i_s \in J_n$, $j \in J_n$, $s \in J_n$, $t \in J_m$, $\{r_1, r_2, \dots, r_m\}$ и $\{s_1, s_2, \dots, s_m\}$ таковы, что $c_{(j-1)m+s_1} \geq c_{(j-1)m+s_2} \geq \dots \geq c_{(j-1)m+s_m}$ и $e_{r_1}^j \leq e_{r_2}^j \leq \dots \leq e_{r_m}^j$, а последовательность $\{i_1, i_2, \dots, i_n\}$ удовлетворяет соотношению (10) при $c = d^*$, то есть $e_{i_1} \prec_{d^*} e_{i_2} \prec_{d^*} \dots \prec_{d^*} e_{i_n}$.

Доказательство проведем на основании леммы 3 при $E = EW_N$, $N = nm$. Задача оптимизации в правой части неравенства (15) может быть решена с помощью соотношения, полученного в [6]. Непосредственная подстановка минимального значения нормы $\|x - y^*\|^2$ на множестве EW_N в (15) приведет к справедливости оценки (18).

Лемма 4 [17]. Если функция $\varphi(x)$ сильно выпукла с параметром $\rho > 0$ и дифференцируема на выпуклом множестве V , где $E \subset V \subset R^N$, то для любого $x \in V$:

$$\min_{y \in E} \varphi(y) \geq \varphi(x) - \frac{1}{4\rho} \|\nabla \varphi(x)\|^2 + \rho \min_{y \in E} \left\| y - x + \frac{1}{2\rho} \nabla \varphi(x) \right\|^2. \quad (19)$$

Теорема 7. Пусть функция $\varphi(x)$ сильно выпукла с параметром $\rho > 0$ и дифференцируема на выпуклом множестве $V \subset ET_{nk}^m$, где множество $ET_{nk}^m \subset R^N$, $N = nm$ порождено множествами $T_i = \{(z_1^i, z_2^i, \dots, z_m^i)\}$, $i \in J_n$. Тогда для любого $x \in V$.

$$\min_{y \in ET_{nk}^m} \varphi(y) \geq \varphi(x) - \frac{1}{4\rho} \|\nabla \varphi(x)\|^2 + \rho \left(\sum_{i=1}^n \sum_{j=1}^m (z_j^i)^2 + \sum_{i=1}^N \left(x_i - \frac{1}{2\rho} \frac{\partial \varphi(x)}{\partial x_i} \right)^2 + 2 \sum_{i=1}^N c_i^* x_i^0 \right),$$

где $c_i^* = -x_i + \frac{1}{2\rho} \frac{\partial \varphi(x)}{\partial x_i}$, $i \in J_N$, $x^0 \in ET_{nk}^m$ такова, что $x_{(j-1)m+t}^0 = z_t^j$, $i_j \neq i_s$ при $s \neq j$, $i_j, i_s \in J_n$, $j \in J_n$, $s \in J_n$, $t \in J_m$, а $\{i_1, i_2, \dots, i_n\}$ удовлетворяет (17).

Доказательство проведем на основании леммы 4 при $E = ET_{nk}^m$, $N = nm$. Задачу об определении минимума в правой части неравенства (19) решим с помощью соотношения о минимуме нормы разности [6] способом, аналогичным примененному при доказательстве теоремы 5.

Теорема 8. Пусть функция $\varphi(x)$ сильно выпукла с параметром $\rho > 0$ и дифференцируема на выпуклом множестве $V \subset EW_N$, где множество $EW_N \subset R^N$, $N = nm$ порождено множествами $\{a_1^i, a_2^i, \dots, a_m^i\}$, $i \in J_n$. Тогда для любого $x \in V$:

$$\min_{y \in EW_N} \varphi(y) \geq \varphi(x) - \frac{1}{4\rho} \|\nabla \varphi(x)\|^2 + \rho \left(\sum_{i=1}^n \sum_{j=1}^m (a_j^i)^2 + \sum_{i=1}^N \left(x_i - \frac{1}{2\rho} \frac{\partial \varphi(x)}{\partial x_i} \right)^2 + 2 \sum_{i=1}^N d_i^* x_i^0 \right), \quad (20)$$

где $d_i^* = -x_i + \frac{1}{2\rho} \frac{\partial \varphi(x)}{\partial x_i}$, $i \in J_N$, $x^0 \in EW_N$ такова, что $x_{(j-1)m+t}^0 = e_{s_j}^j$, $i_j \neq i_s$ при $s \neq j$, $i_j, i_s \in J_n$, $j \in J_n$, $s \in J_n$, $t \in J_m$, $\{r_1, r_2, \dots, r_m\}$ и $\{s_1, s_2, \dots, s_m\}$ таковы, что $c_{(j-1)m+s_1} \geq c_{(j-1)m+s_2} \geq \dots \geq c_{(j-1)m+s_m}$ и $e_{r_1}^j \leq e_{r_2}^j \leq \dots \leq e_{r_m}^j$, а $\{i_1, i_2, \dots, i_n\}$ удовлетворяет

$$e_{i_1} \prec_c e_{i_2} \prec_c \dots \prec_c e_{i_n}. \quad (21)$$

Доказательство проведем на основании леммы 4 при $E = EW_N$, $N = nm$. Задачу об определении минимума в правой части неравенства (33) решим с помощью соотношения о минимуме нормы разности [18] способом, аналогичным примененному при доказательстве теоремы 6.

Лемма 5 [17]. Для того чтобы точка $x^* = (x_1^*, x_2^*, \dots, x_N^*) \in E$ была точкой минимума на множестве E сильно выпуклой с параметром $\rho > 0$ дифференцируемой на выпуклом множестве V функции $\varphi(x)$, где $E \subset V \subset R^N$, достаточно, чтобы

$$\|\nabla \varphi(x^*)\|^2 = 4\rho^2 \min_{y \in E} \left\| y - x^* + \frac{1}{2\rho} \nabla \varphi(x^*) \right\|^2. \quad (22)$$

Теорема 9. Пусть функция $\varphi(x)$ сильно выпукла с параметром $\rho > 0$ и дифференцируема на выпуклом множестве $V \supset ET_{nk}^m$, где множество $ET_{nk}^m \subset R^N$, $N = nm$ порождено множествами $T_i = \{(z_1^i, z_2^i, \dots, z_m^i)\}$, $i \in J_n$. Для того, чтобы точка $x^* \in ET_{nk}^m$ была точкой минимума $\varphi(x)$ на ET_{nk}^m , достаточно, чтобы

$$\|\nabla\varphi(x^*)\|^2 = 4\rho^2 \left(\sum_{i=1}^n \sum_{j=1}^m (z_j^i)^2 + \sum_{i=1}^n \left(x_i^* - \frac{1}{2\rho} \frac{\partial\varphi(x^*)}{\partial x_i} \right)^2 + 2 \sum_{i=1}^n c_i^* x_i^0 \right), \quad (23)$$

где $c_i^* = -x_i + \frac{1}{2\rho} \frac{\partial\varphi(x^*)}{\partial x_i}$, $i \in J_n$, $x^0 \in ET_{nk}^m$ такова, что $x_{(j-1)m+i}^0 = z_j^i$, $i_j \neq i_s$ при $s \neq j$, $i_j, i_s \in J_n$, $j \in J_n$, $s \in J_n$, $t \in J_m$, а последовательность $\{i_1, i_2, \dots, i_n\}$ удовлетворяет (17).

Доказательство проведем на основании леммы 5 при $E = ET_{nk}^m$, $N = nm$. Задача минимизации в правой части равенства (22) может быть решена с помощью формулы, доказанной в [6]. При этом используем подход, аналогичный примененному при доказательстве теоремы 7. В результате приходим к справедливости равенства (23) и теоремы 9.

Теорема 10. Пусть функция $\varphi(x)$ сильно выпукла с параметром $\rho > 0$ и дифференцируема на выпуклом множестве $V \supset EW_N$, где множество $EW_N \subset R^N$, $N = nm$ порождено множествами $\{a_1^i, a_2^i, \dots, a_m^i\}$, $i \in J_n$. Для того, чтобы точка $x^* \in EW_N$ была точкой минимума $\varphi(x)$ на EW_N , достаточно, чтобы

$$\|\nabla\varphi(x^*)\|^2 = 4\rho^2 \left\| y^0 - x^* + \frac{1}{2\rho} \nabla\varphi(x^*) \right\|^2, \quad (24)$$

где $y^0 \in EW_N$ удовлетворяет соотношению (21) при $c^* = x^* - \frac{1}{2\rho} \nabla\varphi(x^*)$.

Доказательство утверждения проведем на основании леммы 5 при $E = EW_N$, $N = nm$. Задача минимизации в левой части равенства (22) может быть решена на основе подхода, примененного при доказательстве теоремы 8. В результате приходим к справедливости равенства (24) и утверждения теоремы.

Выводы

Таким образом, в статье получены новые теоретические результаты, касающиеся экстремальных свойств функций на композиционных образах комбинаторных множеств.

В работе сформулированы новые оценки и достаточные условия минимума выпуклых функций на классах композиционных образов комбинаторных множеств — перестановок кортежей и композиций перестановок.

Полученные результаты могут послужить основой для построения оптимизационного ядра интеллектуальных систем принятия решений, ориенти-

рованных на решение оптимизационных задач со сложной комбинаторной структурой в различных областях, чем определяется их научная ценность и практическая значимость.

Дальнейшие исследования в данном направлении могут быть связаны с изучением экстремальных свойств функций на новых классах композиционных k -образов комбинаторных множеств и с разработкой на основе описанных результатов эффективных методов комбинаторной оптимизации.

Список литературы: 1. Стоян Ю.Г., Яковлев С.В. Математические модели и оптимизационные методы геометрического проектирования. — К.: Наук. думка, 1986. — 268 с. 2. Серуенко И.В. Математические модели и методы решения задач дискретной оптимизации. — К.: Наук. думка, 1988. — 472 с. 3. Айгнер М. Комбинаторная теория. — М.: Мир, 1982. — 558 с. 4. Стоян Ю.Г., Емец О.О. Теория и методы евклидовой комбинаторной оптимизации. — К.: Институт системних досліджень освіти, 1993. — 188 с. 5. Стоян Ю.Г., Гребенник И.В. Композиционные образы комбинаторных множеств и некоторые их свойства // Пробл. машиностроения. — 2005. — Т. 8. — № 3. — С. 56–62. 6. Гребенник И.В. Комбинаторное множество перестановок кортежей и его свойства // Радиоэлектроника. Информатика. Управление. — 2005. — № 1. — С. 92–98. 7. Гребенник И.В. Классы композиционных образов комбинаторных множеств в математических моделях задач геометрического проектирования // Радиоэлектроника и информатика. — 2005. — № 3. — С. 69–73. 8. Яковлев С.В., Гребенник И.В. О некоторых классах задач оптимизации на множествах размещений и их свойствах // Изв. вузов. Математика. — 1991. — № 11. — С. 74–86. 9. Стоян Ю.Г., Яковлев С.В. Построение выпуклых и вогнутых функций на перестановочном многограннике // ДАН УССР. — Сер. А. — 1988. — № 5. — С. 68–70. 10. Яковлев С.В. Теория выпуклых продолжений функции на вершинах выпуклых многогранников // Журн. вычислит. матем. и матем. физики. — 1994. — Т. 34. — № 7. — С. 1112–1119. 11. Стоян Ю.Г., Яковлев С.В., Емец О.А., Валуйская О.А. Построение выпуклых продолжений для функций, заданных на гиперсфере // Кибернетика и системный анализ. — 1998. — № 2. — С. 27–36. 12. Валуйская О.А., Емец О.А., Романова Н.Г. Выпуклое продолжение многогранов, заданных на полиперестановках, модифицированным методом Стояна-Яковлева // Журн. вычислит. матем. и матем. физики. — 2002. — Т. 42, № 4. — С. 591–596. 13. Стоян Ю.Г., Яковлев С.В. Свойства выпуклых функций на перестановочном многограннике // ДАН УССР. — Сер. А. — 1988. — № 3. — С. 238–240. 14. Емец О.А. Множество сочетаний с повторениями, отображенное в R^k , и свойства задач оптимизации на нём // ДАН УССР. — Сер. А. — 1991. — № 4. — С. 69–72. 15. Гребенник И.В., Лапко Д.А. Исследование оценок минимума выпуклых продолжений функций, заданных на евклидовых комбинаторных множествах // Радиоэлектроника и информатика. — 2002. — № 1. — С. 109–113. 16. Гребенник И.В. Экстремальные свойства функций на композиционных образах комбинаторных множеств // Радиоэлектроника и информатика. — 2005. — № 2. — С. 36–44. 17. Стоян Ю.Г., Яковлев С.В., Гребенник И.В. Экстремальные задачи на множестве размещений. — Х., 1991. — 35 с. — (Препринт АН УССР/Ин-т пробл. машиностроения, 347). 18. Гребенник И.В., Баранов А.В. Оптимизация линейных функций на множестве композиций перестановок // Компьютерное моделирование и интеллектуальные системы. Сборник научных трудов. — Запорожье. — 2007. — С. 116–121.

Поступила в редколлегию 22.03.07

УДК 519.62

МОДЕЛЬ ПОВЕРХНОСТНОГО СМЫСЛА ЕСТЕСТВЕННОГО ЯЗЫКА НА БАЗЕ СЕМАНТИЧЕСКИХ ФУНКЦИЙ

Г. Ф. Дюбко¹, Д. В. Преснякова²¹ Харьковский национальный университет радиоэлектроники² Харьковский национальный университет радиоэлектроники

В статье предлагается подход к формализации семантики словосочетаний естественного языка, который декомпозирует семантику на два уровня: поверхностный и глубинный. Рассматривается поверхностный уровень, который представлен семантическими функциями. Приведены определение семантической функции, принцип её использования при доказательстве смысловой эквивалентности словосочетаний, метод конструирования семантических функций из словосочетаний.

ЕСТЕСТВЕННЫЙ ЯЗЫК, СЕМАНТИЧЕСКАЯ ФУНКЦИЯ, ПОВЕРХНОСТНАЯ СЕМАНТИКА, ФОРМАЛИЗАЦИЯ СЕМАНТИКИ.

Введение

Проблема интеллектуального анализа и понимания текстов на естественном языке (ЕЯ) появилась одновременно с созданием компьютеров. Однако до настоящего времени достичь полного успеха в её решении не удалось. Методы для решения частных задач разработаны в большом количестве, но количество так и не перешло в качество. Полное понимание языка на вычислительной (формальной) базе на данный момент остаётся далеко за пределами современных возможностей, и многие фундаментальные проблемы в области автоматической обработки ЕЯ (Natural Language\Processing-NLP) ещё не решены, но развитие современных информационных технологий требует решения частных задач. Так, например, полнотекстовые базы данных, поисковые системы, интерфейсы экспертных систем, машинный перевод требуют анализировать ЕЯ на основе формальных моделей смысла.

Исследователи в области ЕЯ различают общую и прикладную NLP. Задачей общей NLP является разработка моделей обработки языка человеком, а модели при этом должны быть компьютерно эффективными. В основе таких моделей лежит концепция общего понимания текстов и их смысла, как это делается в работах Чарнака, Куллиана, Шенка, Филмора [1, 2, 3, 4].

Прикладная NLP занимается обычно не моделированием, а непосредственной реализацией моделей в некоторой прикладной области. В этом случае не так важно, как реализуемая естественно-языковая конструкция (ЕЯК) будет понята с точки зрения знаний о реальном мире. Здесь важно извлечение информации из ЕЯК о том, чем и как ЭВМ может быть полезной пользователю.

Основной проблемой NLP является языковая неоднозначность: смысловая, падежная, предпочтительная, литерация и т. д. Как в общей NLP, так и в прикладной, разрешение неоднозначностей производится с помощью перевода внешнего представления на ЕЯ в некоторую внутреннюю структуру на основе знаний. Для общей NLP такая трансформа-

ция требует набора знаний о реальном мире, тогда как для прикладной только об узкой предметной области.

Целью данной работы является построение формальной модели поверхностного смысла на основе функциональной зависимости.

1. Постановка задачи

Смысл ЕЯК является фрагментом знаний, которые содержатся в некоторой форме в соответствующей базе знаний, что требует формальной модели смысла.

Формализация смысла берёт своё начало от работ Н. Хомского. Для отображения смысла используется структура непосредственно составляющих (НС). Преобразование ЕЯК в структуру НС производится посредством порождающей грамматики. Причём в процессе преобразования ЕЯК в НС не используется никакая семантическая информация. В дальнейшем выяснилось, что порождающая грамматика не в состоянии адекватно «языковому миру» выдавать смысл анализируемых конструкций. Одинаковые структуры НС могут приводить к разным смыслам. В то же время в порождающей грамматике не отражено, что разные структуры могут иметь один и тот же смысл. Если дополнить порождающую грамматику моделями, отражающими смысл составляющих ЕЯК, и использовать их в трансформациях, можно достичь большей адекватности модели реалиям «языкового мира».

В данной работе получает дальнейшее развитие модель смысла в виде семантических функций [5], т. е. суперпозиции функций, отражающих смысл словосочетаний языков с фразовой структурой. Суперпозиция функций позволяет компоновать смыслы больших ЕЯК из меньших, допуская при этом формальные преобразования конструкций, представляющих смысл. Эта модель должна позволить формальный вывод из ЕЯК, поставив необходимую информацию о подсмыслах в выводящую систему.

Модель смысла в виде семантических функций должна согласовываться с уже апробированными моделями, предлагая новые возможности (выразительные или вычислительные) и допуская формальное доказательство эквивалентности смыслов языковых конструкций, а также осмысленности анализируемой ЕЯК.

Представление смысла в виде определённой формулы требует вычисления этой формулы посредством программной системы (анализатора ЕЯК), структура которой определяется заданной моделью смысла.

2. Анализатор ЕЯК

Структура предлагаемого анализатора и потоки данных вычислительных блоков представлены на рис. 1.



Рис. 1. Анализатор ЕЯК

Краткая расшифровка функционирования и структур данных представленной схемы приводится ниже.

Блок РИТЛЭ воспринимает ЕЯК и превращает её во внутреннее представление символов, устраняя пробелы и не производя контроля входной информации. Реализовать такой блок можно на уровне конечных автоматов или широко известных алгоритмов поиска подстрок с линейными характеристиками временной сложности алгоритма.

Блок МА превращает поток ЛЭ в последовательность МЕ, т. е. выполняет, по сути, морфологический анализ. Отличим от большинства подходов (особенно в славянских языках) реализации морфологического анализа, где МЕ вычисляется, является декларативный подход, где каноническая форма слова находится в морфологическом словаре. Морфологический словарь в этом случае должен представляться в специфической форме: задаются все словоформы слова и их морфологические признаки. Каноническая словоформа слова связана с такой же словоформой в семантическом словаре. Авторами разработана структура морфологического словаря и эффективный (с вычислительной точки зрения) доступ к канонической форме. Отметим, что элемент МЕ на уровне морфологии может быть неоднозначным. На этом же уровне анализа осуществляется контроль правильности слов во входном потоке.

Блок СА построен на базе использования формальных грамматик и формально организованного семантического словаря. Отличим от традиционного представления атрибутивных продукций служит факт включения в продукцию правил применения продукций. Это делает алгоритм анализа более быстроедействующим. Обсуждение принципа построения семантического словаря выходит за рамки рассматриваемой здесь работы. Заметим лишь, что словарь построен на основе семантических функций и является лингвистической базой знаний (в отличие от предметной) анализируемого языка. За счёт этой базы знаний возможно существенно сократить нечёткость получаемых знаний. Дальнейшее ограничение нечёткости может быть достигнуто использованием предметной базы знаний.

Блок ФСМ несёт техническую нагрузку, формируя выход анализатора в форме семантических функций из уже сформированных фрагментов.

3. Формальные модели смысла

Большинство работ по формализации семантики языка касаются разработки формальных моделей и языков их представления, которые должны позволять одинаковое формальное каноническое представление для различных ЕЯК, имеющих одинаковый смысл. Формализация семантики ЕЯ уменьшает проблемы, связанные с многозначностью. Само представление в процессе построения канонической формы охватывает много ЕЯ-форм. Однако эквивалентным по смыслу ЕЯ-формам соответствует единственная каноническая форма. Как отмечают исследователи в области языка и математики, про-

цесс преобразования языка в каноническую форму вряд ли можно полностью алгоритмизировать. Как доказано, преобразование в каноническую форму нельзя реализовать даже для моноидов алгебраических групп, которые значительно проще, чем ЕЯ. С точки зрения моделирования языкового поведения человека нет уверенности, что человек хранит свои знания в какой-нибудь канонической форме. В рассматриваемой работе не предлагается канонических форм для представления эквивалентных ЕЯ-конструкций, а доказательство эквивалентности предусмотрено на уровне эвристик.

Анализ литературных источников показал, что формализация семантики сводится преимущественно к построению баз знаний для предметных областей. В этом подходе предложен ряд моделей, которые более или менее сходны с логикой предикатов. Сама логика предикатов является моделью, довольно абстрактно представляющей смысл. К упомянутым моделям относятся семантические сети, концептуальные графы, фреймы, онтологии.

Трансформация ЕЯ-конструкций в смысловые представления на основе баз знаний предметной области встречает определённые трудности вычислительного плана, которые можно уменьшить, декомпозируя трансформацию в два этапа: этап поверхностной семантики и этап глубинной семантики (предметная база знаний). Поверхностный уровень семантики основывается на базе знаний, подобной толковым словарям, используемым человеком.

Как показывают экспериментальные исследования, применение поверхностной семантики позволяет существенно сократить неоднозначность представлений, оставляя возможность установить соответствие между поверхностным и глубинным уровнями.

Предлагаемая в работе модель семантики в виде семантических функций (СФ) относится к поверхностному уровню.

4. Семантические функции

Естественный язык представляется словосочетаниями. Под словосочетанием понимается текст, фраза, сложное предложение, простое предложение, прилагательное плюс существительное, существительное плюс существительное в родительном падеже и т. д. Базовым словосочетанием языка является простое повествовательное предложение. Таким образом, словосочетание — это последовательность слов, составляющих некоторую синтаксическую группу. Каждому словосочетанию ставится в соответствие его смысл. Например, словосочетанию «красный шар» соответствует смысл (речь идёт о поверхностном уровне смысла) «цвет». Это соответствие устанавливается с помощью толкового словаря, где поверхностный смысл описывается в терминах ЕЯ. При этом поверхностный смысл каждого слова задаётся соответствующей словарной статьёй

этого слова в толковом словаре, а обрабатывает имеющуюся информацию человек-носитель языка.

Чтобы сделать возможной аналогичную человеческую обработку языка вычислительными устройствами, необходимо формализовать толковый словарь и представление смысла словосочетаний. Такая формализация может быть достигнута посредством семантических функций. Дадим определение семантической функции.

Обозначим через $L(\text{ЕЯ})$ множество конструкций ЕЯ, т. е. множество словосочетаний, а через $L(\text{СФ})$ — множество конструкций языка, в котором представляется смысл. Заметим, что $L(\text{ЕЯ})$ — неформальный язык, а $L(\text{СФ})$ — формальный. Пусть Σ_T терминальный алфавит для представления цепочек в $L(\text{ЕЯ})$, Σ_{PT} — часть Σ_T , используемая только для представления слов $L(\text{ЕЯ})$. Через $\alpha \in \Sigma_T^*$ обозначим языковую конструкцию α , состоящую из символов алфавита Σ_T (включая пустое слово ϵ). Тогда $\Sigma_T^+ = \Sigma_T^* - \{\epsilon\}$. N — алфавит для представления конструкций языка $L(\text{СФ})$, и $N = \{V, F_{\text{и}}, (,), '\} \cup \Sigma_{\text{PT}}$.

Пусть $f(\alpha_1, \dots, \alpha_n)$ — функция, соответствующая словосочетанию, состоящему из слов $\alpha_1, \dots, \alpha_n \in \Sigma_T^+$. Значением этой функции является смысл, который может быть вычислен посредством процедур, приведенных на рис. 1, т. е. $f(\alpha_1, \dots, \alpha_n) = \beta$, где $\beta \in N^+$. Языковая конструкция β представляет семантическую функцию

$$F_{\text{и}}(X_1, \dots, X_m), \quad (1)$$

где $X_j = V_i(\alpha_k)$ или $X_j = F'_{\text{и}}$ ($1 \leq j \leq m$, $1 \leq i \leq n$, k — номер словарной статьи в семантическом словаре); и — название функции (словосочетания); $F'_{\text{и}}$ — семантическая функция, вложенная в $F_{\text{и}}$; α_k — слово из словосочетания.

Следовательно, можно записать:

$$f(\alpha_1, \dots, \alpha_n) = F_{\text{и}}(X_1, \dots, X_m), \quad (2)$$

где правая часть равенства определена в (1).

Приведём примеры семантических функций в украинском языке. Пусть рассматривается словосочетание «ціна товару в упаковці». Ему соответствует семантическая функция

$$F_{\text{родового відмінку}}(V_j(\text{ціна}), F_{\text{прийменник}}(V_r(b), V_e(\text{товару}), V_s(\text{упаковці}))).$$

Для словосочетания «ціна товару в гривнях» семантическая функция имеет вид:

$$F_{\text{родового відмінку}}(F_{\text{прийменник}}(V_l(b), V_j(\text{ціна}), V_k(\text{гривнях}) V_e(\text{товару}))).$$

В обоих примерах индексы i, j, k, e, r, s обозначают номер словарной статьи в семантическом словаре.

Выбранные примеры демонстрируют факт, что в славянских языках слова словосочетания могут идти в $L(\text{ЕЯ})$ не подряд, а быть разорванными. Семантическая же функция представляет все слова словосочетания в единой конструкции.

Функция $f(\alpha_1, \dots, \alpha_n)$ из (2) имеет областью своего определения множество слов, а областью значений — множество строк (термов), являющихся семантической функцией. В свою очередь, семантическая функция может интерпретироваться как функция, аргументами которой являются другие функции, и которая имеет свою область значений. Задавая правила вычисления семантических функций, можно получать их значения. С этой точки зрения равенство (2) можно переписать в виде:

$$f(\alpha_1, \dots, \alpha_n) = P^{m+1} F_{\text{и}}(X_1, \dots, X_m), \quad (3)$$

где $P^{m+1} - (m+1)$ — местная функция с известными правилами вычисления (вычисляются значения X_1, \dots, X_m , а затем — значение функции $F_{\text{и}}$).

Областью значений $F_{\text{и}}$ является множество семантических функций. Возникает вопрос, в чём разница между $F_{\text{и}}$ и множеством её значений, поскольку значением семантической функции является другая семантическая функция. В ЕЯ смысл некоторой языковой конструкции может отражать конкретику этой конструкции, а может быть обобщённым. Конкретика отражается функцией $F_{\text{и}}$, а обобщённый смысл значением $F_{\text{и}}$. Чтобы продемонстрировать это, вернёмся к примеру «красный шар»:

$$f(\text{красный, шар}) = F_{\text{прилагательное}}(V_i(\text{красный}), V_k(\text{шар})) = F_{\text{родительного падежа}}(V_k(\text{цвет}), V_c(\text{объекта})).$$

Здесь $F_{\text{родительного падежа}}$ является обобщающим смыслом для $F_{\text{прилагательное}}$. Заметим, что каждая переменная семантической функции играет определённую роль. Эта роль может фиксироваться либо местом этой переменной, либо ролевым дополнением семантической функции.

Посредством семантических функций можно выразить поверхностную осмысленность ЕЯ-конструкции. Если привлечение семантического словаря не позволяет построить семантическую функцию, то конструкция не обладает смыслом.

Различные семантические функции могут представлять один и тот же смысл. Доказательство этого факта можно формализовать. Рассмотрим это на примере.

Пусть имеются семантические функции

$$F_{\text{глагол}}(X_1, X_2, X_3), \quad (2)$$

где X_1 — глагол, обозначающий действие; X_2 — объект, который оказывает действие; X_3 — объект, на который оказывается действие;

$$F_{\text{родительного падежа}}(Y_1, Y_2), \quad (3)$$

где Y_1 — существительное; Y_2 — существительное в родительном падеже или семантическая функция $F_{\text{и}}$, где существительное в родительном падеже (Y_2) является главным словом.

Функция $F_{\text{и}}$ есть $F_{\text{творительного падежа}}$, т. е. $F_{\text{творительного падежа}}(Y_2, Y_3)$.

Формулы (2) и (3) представляют одинаковый смысл, если слово Y_1 (существительное) является отглагольным от глагола X_1 , а Y_2 является семанти-

ческой функцией $F_{\text{творительного падежа}}$ с $Y_2 = X_2$, $Y_3 = X_3$. ЕЯ-конструкции «программа стирает память» и «стирание памяти программой» с семантическими функциями:

$F_{\text{глагол}}(I(\text{стирать}), I(\text{программа}), I(\text{память})),$
 $F_{\text{родительного падежа}}(I(\text{стирание}),$
 $F_{\text{творительного падежа}}(I(\text{память}), I(\text{программа}))),$

соответственно, являются эквивалентными по смыслу, что легко проверить по семантическому словарю [6].

С помощью подобных эвристических правил устанавливается эквивалентность для всех возможных словосочетаний.

Выводы

В данном исследовании находит своё отражение подход к формализации смысла словосочетаний естественного языка в виде семантических функций. Семантические функции предлагают формальный аппарат для машинной обработки смыслов, позволяют устанавливать осмысленность и эквивалентность словосочетаний ЕЯ. Точность вычислений регламентируется качеством семантического словаря.

Предложена вычислительная схема для конструирования семантических функций, отражающих смысл конкретных словосочетаний, уточнено определение семантической функции.

Научная новизна работы состоит в том, что получают дальнейшее развитие: понятие семантической функции, метод установления эквивалентности словосочетаний на базе семантических функций, метод построения семантических функций.

Практическая направленность исследований: естественно-языковые интерфейсы, поиск в полнотекстовых базах данных, автоматическое аннотирование и реферирование, машинный перевод.

Дальнейшее направление исследований: выделение всего множества семантических функций в разных естественных языках; определение полного множества операций установления эквивалентности семантических функций; формальные методы представления смысла в семантическом словаре; приложение семантических функций в машинном переводе.

Список литературы: 1. Charnak E. and Mc Dermot D. Introduction to Artificial Intelligence. Reading, MA: Addison-Wesley, 1985. — 360 с. 2. Quillian M.R. Word concepts. A theory and simulation of some basic semantic capabilities. In Brachman and Levesque. 1967. — 325 с. 3. Schank R.C. and Rieger C.J. Inference and the computer understanding of natural language. Artificial Intelligence 5(4): 373–412. 1974. — 512 с. 4. Filmor C.J. The case for case. In Bach and Harms, 1968. — 420 с. 5. Дюбко Г.Ф., Валенда Н.А. Использование формального вывода для семантического анализа конструкций естественного языка // Сборник научных трудов по материалам 6-й международной конференции «Теория и техника передачи, приема и обработки информации». — Х.: ХТУРЭ, 2000. — С. 251–253. 6. Ожегов С.И. Словарь русского языка: Ок. 57 тыс. слов / Подред. чл.-корр. АН СССР Н.Ю. Шведовой. — 17-е изд., стереотип. — М.: Рус. яз., 1985. — 797 с.

Поступила в редакцию 25.03.07

УДК 681.518:004:912

О ВИЗУАЛЬНОМ КОДИРОВАНИИ ИНФОРМАЦИИ НА ОСНОВЕ ИДЕНТИФИКАЦИИ КОГНИТИВНЫХ ФУНКЦИЙ ЧЕЛОВЕКА-ОПЕРАТОРА. СООБЩЕНИЕ 2

А. Л. Ерохин

ХНУВД, г. Харьков, Украина, auegokhin@ukr.net

Рассматривается решение задачи визуального кодирования информации для моделирования психофизиологических состояний человека-оператора. Предложен метод невербального неинвазивного воздействия на систему психофизиологического состояния человека-оператора.

КОГНИТИВНЫЕ ФУНКЦИИ ЧЕЛОВЕКА, ЗРИТЕЛЬНЫЕ ОБРАЗЫ, ЦВЕТОДИНАМИЧЕСКИЕ ВОЗДЕЙСТВИЯ.

Введение

В [1] рассматривается постановка задачи идентификации когнитивных функций человека на основе моделирования и адаптации психофизиологических состояний (ПФС) человека-оператора как одного из важных компонентов системы управления сложной системой. Статья посвящена исследованию возможности построения технических средств адаптации ПФС человека-оператора во время стрессорных воздействий при выполнении им операторских функций на основе анализа принципов работы мозга как моделирующей установки [2]. Разработан метод невербального неинвазивного воздействия на ПФС человека-оператора на основе взаимодействия цветодинамических изображений с эндогенными системорегулирующими процессами [3] в системе когнитивных функций человека-оператора.

1. Анализ принципов работы мозга с точки зрения моделей восприятия зрительных образов

Как известно, образование моделей образов происходит в результате перекодирования информации из низшего кода в высший по вероятностному принципу. Например, для речевого образа (по Н.М. Амосову): код звукового сообщения — код слов — код фраз — код смысла. Так, психофизиологи установили следующие основные принципы формирования моделей образов: принцип активного изоморфизма (образование модели происходит по законам изоморфного отображения); принцип компарации (процесс сличения или установления меры сходства, идентификации); принцип функционального кольца (образование новых моделей происходит при сравнении врожденных или приобретенных в ходе индивидуального развития моделей со вновь возникающими при поступлении сигналов в мозг, моделирование — это результат циркуляции по кольцу: центр — периферия — кора — подкорковые образования); принцип энтропийности (создание моделей в мозге ведет к уменьшению энтропии); принцип вероятностного прогнозирования (наличие обратной связи — канала, по которо-

му в мозг идет сигнализация о том, что делается на периферии, имеет смысл тогда, когда в мозгу происходит сравнение того, что сделано, и того, что должно быть в результате действия); принцип минимизации отрицательного (организм избирает такую тактику, которая сулит ему удовлетворение жизненных потребностей и позволяет избежать неблагоприятных ситуаций и отрицательных эмоций, то есть принцип максимума выигрыша при минимальных затратах) [4]. Эти принципы будем использовать при компьютерном моделировании человека-оператора при выполнении им операторских функций во время стрессорных воздействий.

Основной целью функционирования оператора является обеспечение выполнения им своих основных операторских функций в течение заданного времени (время дежурства) на заданном уровне, независимо от внешних проявлений негативных факторов. Поэтому важной задачей является совершенствование интерфейсов «человек-машина» путем добавления в систему специальных устройств, обеспечивающих адаптацию ПФС человека-оператора на заданном уровне.

Для построения такой модели вначале необходимо решить задачу формальной идентификации ПФС человека-оператора. Важной остается проблема влияния стресса монотонии на человека-оператора, включенного в интеллектуальную систему. Стресс монотонии является неспецифической реакцией человека-оператора на выполнение им однообразных длительных действий, и основным механизмом развития стресса является способность центральной нервной системы отфильтровывать длительные и монотонные стимулы [5–7]. Указанная фильтрация выводит эти стимулы из области внимания, переводя сенсомоторные функции в область подсознательного автоматизма, а само сознание — в измененное состояние. Человек-оператор, находясь в состоянии монотонии, начинает допускать ошибки управления и принятия адекватного ситуации решения.

Представляет интерес разработка специальных методов, позволяющих адаптировать систему пси-

хофизиологического состояния человека-оператора к негативному воздействию монотонии, для чего необходимо рассмотреть два класса указанных систем ПФС по степени включения его когнитивных функций. Степень включения этих функций определяет открытое множество психических структур-процессов, которые, исходя из синергетических воззрений [8], расположены между полем сознания и подсознанием, то есть взаимно переходящих с уровня перцепции через уровень восприятия к уровню представлений.

Первый класс систем ПФС человека-оператора — это многообразие ансамбля психических структур-процессов, образующих спектр состояний между процессами перцепции и восприятия [9]. При этом к самым нижним психическим процессам можно отнести процессы «субъективных состояний», локализованных ближе всех к подсознательному. Относительная близость систем ПФС этого слоя к областям подсознательного определяет возможность воздействия на систему ПФС, расположенных ближе всех к подсознательному, посредством какого-либо внешнего (экзогенного) воздействия. Такое эндогенное невербальное воздействие способно достичь решения, по крайней мере, двух задач.

Первая из них — помочь системе гомеостаза вернуть систему ПФС из неустойчивого состояния в исходное устойчивое состояние (в исходный аттрактор), вторая — перевести систему ПФС в новое устойчивое состояние (в новый аттрактор). Благодаря открытости и нелинейности семантического пространства психических процессов переходы систем ПФС с уровня на уровень осуществляются скачком, образуя спектр переходных состояний. Примером таких переходов может являться процесс лабиринтного развития ассоциативной цепи вне поля внимания [10]. Рождение ассоциации возникает при минимальном экзогенном воздействии, однако такое же экзогенное воздействие с минимальной интенсивностью, но уже с другой топологией и архитектурой воздействия на систему ПФС становится достаточным для вытеснения предыдущей ассоциации и замены ее на новую.

Ко второму классу систем ПФС человека-оператора относятся психические структуры-процессы, рождающиеся и развивающиеся на «поле мозга и сознания» [9]. Системы ПФС этого класса отличает активное включение познавательных и креативных процессов, образующих открытое многоуровневое вербальное пространство. Переход структур-процессов с нижних уровней систем ПФС на более высокие подчинен механизмам последовательной обработки информации. Переход системы ПФС на более высокий уровень сознательного происходит

по принципу эмерджентности, по мере поступательного процесса движения мысли к решению.

Общими принципами эволюции первого и второго классов системы ПФС от подсознательного к сознательному являются: топология, время и интенсивность экзогенных воздействий. Топологически структурированные, темпорально правильные воздействия резонансно взаимодействуют с системой ПФС человека-оператора и классифицируются как резонансные хронотопы [8]. Правильно подобранные низкоинтенсивные резонансные хронотопы могут рассматриваться как факторы адаптации систем ПФС человека-оператора к стрессорам.

Формально системы ПФС человека-оператора можно представить как функцию множества динамично изменяемых параметров $P_{SF} = [\bigcup(P_S), \bigcup(P_F)]$, где $(P_S), (P_F)$ — психологические и физиологические параметры. Для метастабильной фазы во временном интервале Δt_c психологические и физиологические параметры $P(S_j), (P_{Fi}) = C \quad \forall S_j, F_i, \Delta t$. Системы ПФС человека-оператора могут рассматриваться как неизменные.

Применим для структур-процессов первого класса, включающих механизмы переходов ощущений с уровня пространственного поля на уровень действий, нейропсихологический механизм взаимодействий систем ПФС человека-оператора с резонансными хронотопами. В качестве таких воздействий на человека-оператора используем последовательность визуальных стимулов с динамическим изменением цветовых образов, или цветодинамические воздействия (ЦДВ). При этом ЦДВ обладают дискретной топологией, дающей возможность генерировать на дискретной фазовой плоскости детерминированный цветовой хаос, и частотно-временными параметрами, позволяющими подобрать наиболее оптимальный режим резонансного взаимодействия ЦДВ и СПФС человека-оператора.

Нейропсихологический механизм резонансно-го взаимодействия описывается следующей последовательностью:

а) ЦДВ первично кодируется в зрительном анализаторе, после чего закодированные сигналы поступают в соответствующие проекционные зоны поля мозга [11–13];

б) нервные импульсы возбуждения спускаются по колонкам нейронов с верхних слоев проекционных зон мозга к нижним, на которых формируются медленные импульсы торможения, направленные встречно [14];

в) спектры нервных импульсов и медленных волн посредством интермодальных переходов коллатералей аксонов нейронов поступают в проекционные зоны других модальностей, осуществляя взаимодействие на различных уровнях поля мозга;

в) резонансное усиление или ослабление нервных импульсов или медленных волн осуществляется через взаимосвязь проекционных зон зрительного анализатора с зоной гипокампа, лимбической системой и областей гипоталамуса [14].

ЦДВ представляют собой спектр частот модулируемых цветовых стимулов в виде абстрактных, изменяемых в каждое мгновение времени образов и траекторий перемещения в диапазоне частот модуляции от 2 до 12 Гц. Максимальная частота цветового импульса, посредством которого модулируются ЦДВ, составляет 24–30 Гц [15]. Цветовые стимулы, предъявляемые человеку-оператору в виде ЦДВ, препятствуют развитию стресса монотонии, поскольку топология их образов и частотно-временные параметры способны обеспечить в проекционных зонах мозга постоянный уровень возбуждения, временно отключая при этом механизмы фильтрации циклически повторяющихся стимулов. При отключении механизма фильтрации монотонных стимулов включается механизм выделения спектров ведущих ритмов (или частот) заинтересованными структурами-функциями сознания-подсознания. Благодаря разветвленной системе аксонов мозга в нем происходят разнообразные когерентные взаимодействия эндогенных ритмов с экзогенными ритмами. Ввиду того, что регулирующие центры интрацептивных уровней управления физиологическими процессами находятся в глубинных слоях подкорки, когерентно-резонансное усиление (ослабление) амплитуд эндогенных ритмов способно косвенно воздействовать на эти центры управления.

Такие внешне наведенные цветовые ритмы избирательно воздействуют на те структуры сознания-подсознания, которые в данный момент времени становятся топологически и темпорально эквивалентными одному из эндогенных ритмов. Так, для активации когнитивной функции внимания переменные поля эндогенного воздействия красного цвета должны быть предъявлены человеку-оператору с частотой 10–12 Гц, так как эта частота соответствует частотному диапазону бета-волн электрической активности мозга [16]. Кроме известных альфа-, бета-, дельта- и тета-волн электрической активности мозга, существует ряд других эндогенных ритмов, таких, как колебания биологических процессов в трехсекундном ритме [16], частота сердечных сокращений и дыхания, посредством которых возможно модулировать ЦДВ и оказывать адаптивное влияние на систему ПФС человека-оператора.

Таким образом, гипотеза взаимодействия системы ПФС человека-оператора и ЦДВ основывается на механизмах амплитудного усиления или ослабления эндогенных ритмов и, в первую очередь, ам-

плитуд волн электрической активности головного мозга.

Для когнитивных процессов на уровне концептуальных структур предлагается гипотеза взаимодействия резонансного хронотопа ЦДВ и системы ПФС человека-оператора, основанная на механизмах циклически изменяемых процессов ассоциации — концентрации [17, 18].

Волна ассоциаций устанавливается именно в режиме концентрации, для которой необходим определенный интервал Δt времени для синтеза или извлечения ассоциаций. Абстрактный и динамический фрагмент цветового образа, находящийся в поле внимания человека-оператора, сравнивается с содержимым памяти, на которое затем переключается внимание. Ассоциированный объект внимания, в свою очередь, является основой последующей ассоциации, что порождает цепочку связанных ассоциаций в виде последовательного хода мысли. При дефиците времени $t < \Delta t$ происходит обрыв цепи связанных ассоциаций, выводящий последнюю ассоциацию из поля внимания, переключая внимание на следующую ассоциацию. Из этого следует, что изменяемые ЦДВ обладают высоким значением важности [18].

На основании указанных исследований определим минимальные значения времени Δt , которое необходимо для восприятия субъектом образа во время ЦДВ. Известно, что информация в кратковременной памяти (КВП) кодируется в виде акустических, визуальных и семантических кодов [11–14], при этом часть времени информация в КВП кодируется зрительно в иконической памяти [19].

Обработка цветовой информации в КВП после ее восприятия кодируется несколькими различными системами и с различной продолжительностью. Так, для цветовой информации различают три типа кодов: физический код цвета, код названия цвета и концептуальный (ассоциативный) код. Экспериментально установлено, что код ассоциации возникает через 1500 мс. Из этого следует, что для эволюции цепи ассоциаций к образу человек должен иметь временной интервал $\Delta T_a \gg (500 - 1500)$ мс.

2. Разработка метода воздействия на систему ПФС человека-оператора

Для формирования ЦДВ с указанными характеристиками будем использовать метод комбинаторно-топологического преобразования двумерных изображений [20], который дает возможность разработать новые способы цветодинамического воздействия на психофизиологическое состояние человека и устройств для использования в практической психологии для обеспечения психологической адаптации человека к стрессорным нагрузкам. Спе-

циальные волоконно-оптические устройства, преобразующие изображения с искажением структуры исходного изображения для кодирования информации, описаны в [21–23]. Каждый элементарный канал (ЭК) такого устройства имеет два сопряженных поперечных сечения, которые фиксируются с двух сторон, образуя с одной стороны приемную поверхность, или вход, а с другой стороны — выходную поверхность (выход). В зависимости от взаимного расположения указанных сопряженных сечений различают волоконно-оптические элементы (ВОЭ) с регулярным (упорядоченным) расположением, которые используются для передачи изображения без искажения его структуры. При нарушении такой упорядоченности, например, со стороны выхода структура изображения искажается.

Рассмотрим способ преобразования двумерной информации — комбинаторно-топологическое преобразование (КТП) [20]. Для реализации способа используются специальные волоконно-оптические преобразователи (ВОП) изображений [15, 20]. В способе КТП внутри ВОП случайным образом вводятся комбинаторные перестановки ЭК, нарушающие их упорядоченность. При этом положения входных и выходных сечений понимаются как декартовы произведения упорядоченных пар относительно их базовых систем координат. В результате таких перестановок на выходе ВОП формируется совокупность некоторых цветных, абстрактных фигур. При линейных и угловых перемещениях относительно указанного входа ВОП формирует на своем выходе последовательность сменяемых друг друга цветowych, хаотических по форме и траектории перемещений изображений, так называемых цветодинамических изображений (ЦДИ). Визуально ЦДИ складывается из множества дискретных, разноокрашенных фрагментов, которые формально могут рассматриваться как детерминированные хаотические процессы. Исследования [13–6] доказали, что ЦДВ обладают эффективным психореабилитирующим и психокорректирующим воздействием на человека.

Для формирования ЦДВ будем использовать модель волоконно-оптического системного канала (СК) [20], состоящего из множества оптических ЭК. Для представления формальной модели обозначим СК как (W) , а совокупность поперечных сечений ЭК, образующих СК, с одной стороны рассматриваем как элементы (a_i) его входа $[A]$. Тогда все сопряженные сечения (b_j) определим как элементы выхода $[B]$ для (W) . Топологические свойства множеств (a_i) , $(b_j) \in (W)$ определяются следующим образом:

1) элементы (a_i) и (b_j) на входе $[A]$ и выходе $[B]$ в (W) являются дискретными, с геометрически правильными формами, преимущественно круг-

лого сечения диаметра d , свойство определяет тип топологического пространства как дискретное;

2) любой ЭК, далее (w_i) , обладает сопряженностью своих входных и выходных сечений, которое обосновывает изоморфизм отображений на элементах входа-выхода (a_i) и (b_j) :

$$(a_i) \leftrightarrow (b_j); \quad (1)$$

3) любой (w_i) обладает тождественностью передачи информации по ЭК со входа на выход $(a_i) \equiv (b_j)$ как следствие инварианта Штраубеля [24], что означает отсутствие оптических искажений в самом ЭК $(w_i) \in (W)$;

4) все $(w_i) \in (W)$ являются слабоструктурированными, поскольку ЭК, образующие СК, не имеют между собой детерминированных связей, и вследствие этого любой из ЭК внутри СК может занять любое положение.

Свойство 4 не выполняется в зонах компактно-го объединения элементов $(a_i) \in [A], (b_j) \in [B]$ на входе-выходе СК в виде двумерных структур. Слабая структурированность ЭК в модели дает возможность выполнить любую последовательность комбинаторных преобразований (перестановок) элементов выхода $(b_j) \in [B]$ относительно их сопряжений на входе. За счет деформаций сжатия, кручения и изгиба подмножеств ЭК внутри СК элементы выхода образуют совокупность топологически подобных кругу фигур или гомеоморфизмов $\bigcup (b_j) \in [B]$. Введение в СК хаотичности положений ЭК за счет комбинаторных перестановок и формообразование подмножеств ЭК в виде гомеоморфизмов при топологических преобразованиях характеризуют КТП информации. Любые из этих параметров КТП могут быть зафиксированы двумерными структурами входа-выхода СК. В общем виде процесс формального комбинаторного и топологического преобразований состоит из следующих этапов:

1) формирование образа исходного двумерного изображения проективным непрерывным преобразованием [17] исходного изображения $F(x, y, l)$ на входе $[A]$ в виде множества дискретных его фрагментов $\bigcup f'(a_{ij})$, отображенных на множестве $[A] = \bigcup (a_{ij})$;

2) выполнение произвольных перестановок ЭК внутри СК с частичным или полным нарушением упорядоченности координат на выходе, которые задаются таблицей отношений между координатами входа и выхода:

$$\begin{aligned} (a_{ij}) &\leftrightarrow (b_{kl}) \quad \forall (i, j) \neq (k, l), \\ (a_{pr}) &\leftrightarrow (b_{pr}) \quad \forall (p, r) = (p, r), \end{aligned} \quad (2)$$

и формирование гомеоморфизмов подмножеств $(b_{pr}) \cup (b_{kl}) \in [B_i]$, где $i = 1, 2, \dots$;

3) формирование фрагментов прообраза исходной информации на выходе в виде объединения гомеоморфизмов $[B] = \bigcup_{i \in M^2} [B_i]$ с комбинаторным нарушением упорядоченности элементов внутри каждого из $[B_i]$.

Неупорядоченность в СК на этапе 2 можно ввести с помощью коэффициента нерегулярности K , который в теории волоконной оптики характеризует величину фазовых искажений пространственных частот, и для него есть два ограничения.

$K = 0$ соответствует СК с сохранением полной упорядоченности взаимных координат ЭК на входе-выходе.

При $K = 1$ все ЭК расположены в СК хаотически, с полным нарушением упорядоченности. Если упорядоченность входа-выхода ввести посредством матричного представления множеств, то вышеприведенные этапы преобразования исходной информации могут быть представлены в виде системы:

$$\left\{ \begin{array}{l} \sum_{i=1}^m \sum_{j=1}^n (S_{ij}) p \xrightarrow{K} \sum_{i=1}^m \sum_{j=1}^n (S_{ij}^*) p, p = 1, 2, \dots, M, \\ 0,3 \leq K \leq 0,7, \\ \bigcup_{p=1}^M [A_i] \xrightarrow{T, K_1} \bigcup_{p=1}^M [B_i], i \geq 2, 3, \dots, \\ 0 \leq K \leq 1, 0, \end{array} \right. \quad (3)$$

где K — оператор комбинаторного преобразования элементов (b_{kl}) внутри гомеоморфизмов $[B_i]$ относительно сопряженных им входов $[A_i]$, определяющий степень хаотичности в виде коэффициента $k = \sum \sum S^* / (m \times n)$, где S^* — число неупорядоченных элементов в СК, соотношенное к общему числу элементов; T — оператор топологических преобразований входов в объединении гомеоморфизмов на выходе; K_1 — оператор неупорядоченности гомеоморфизмов выхода относительно сопряженных им входов, который также определяет фазовые искажения в СК, когда число входов $i > 2$.

Если двумерное изображение $F(x, y, I)$, заданное в континуальном пространстве координатами (x, y) и интенсивностью I и отображенное на входе СК, линейно перемещается относительно него, то его отображение есть образ последовательностей перемещаемых цветовых потоков с геометрически правильными формами. На этапе 1 в каждый момент времени первичные изображения преобразуются во множество фрагментов, образующих в со-

вокупности образ, эквивалентный исходному изображению:

$$F(x, y, I) \approx [F'(i, j, \sum I)] = \sum_{i=1}^m \sum_{j=1}^n (a_{ij} \otimes f'_{ij}) \subset [A], \quad (4)$$

где $(a_{ij} \otimes f'_{ij})$ — дискретные фрагменты (f'_{ij}) образа изображения, отображенные на элементах (a_{ij}) множества входа; $\sum I_{ij}$ — значение обобщенной интенсивности, проинтегрированной по всей площади элемента (a_{ij}) с диаметром d ;

4) дискретные фрагменты образа изображения формируются на элементах $(b_{ij}) \in [B]$ выхода в виде прообраза $[F^*(k, l, \sum I)]$, который уже может быть неадекватным образу изображения, при этом их неадекватность пропорциональна степени неупорядоченности:

$$[F^*(k, l, \sum I)] = \sum_{k=1}^p \sum_{l=1}^s (b_{kl} \otimes f^*_{kl}) \subset [B], \quad (5)$$

где $(b_{kl} \otimes f^*_{kl})$ — дискретные фрагменты (f^*_{kl}) прообраза изображения, отображенные на элементах (b_{kl}) выхода.

Таким образом, разработан метод, реализующий формирование ЦДВ, которые пригодны для психологической реабилитации операторских функций человека. Результатом взаимодействия ЦДВ и систем ПФС человека-оператора является адаптация к стрессу монотонии, обеспечивающая оптимальный уровень когнитивных функций и уменьшение вероятности принятия неадекватного решения [23].

3. Практическая реализация метода

Экспериментально [18, 20] установлено, что оптимальными значениями коэффициента k , при котором преобразованные изображения воспринимаются человеком как ЦДВ, является интервал $0,3 \leq k \leq 0,7$.

Для формирования ЦДВ применяются специальные проекционные установки, состоящие из проекционного источника света, вращающейся цветопрограммной матрицы, которая представляет собой набор цветных светофильтров, устройства для формирования изображений и проекционного объектива. Светофильтры устанавливаются в непосредственной близости от входной поверхности устройства, и тогда первичными изображениями являются ограниченные их геометрией цветовые потоки от каждого фильтра. Принципы, заложенные в КТП, дают возможность программировать абстрактные формы ЦДВ.

Например, для активации когнитивной функции внимания посредством специального преобра-

зователя изображений моделируют на экране топологические структуры, дискретные фрагменты которых создают визуальную иллюзию цветового пространства, сходящегося в точку. Для этого из множества элементарных каналов последовательностью деформаций формируют объединения гомеоморфизмов, топологически подобные кольцевым или спиралевидным и вложенным друг в друга. При движении исходного цветового изображения, например, слева направо на выходе преобразователя изображений формируются ЦДВ в виде сходящихся в точку концентрических или спиралевидных цветовых волн. При реверсе направления движения исходного изображения происходит изменение траектории цветовых волн на противоположное («расходящееся» цветовое пространство).

Указанные проекционные установки выпускаются мелкими сериями на научно-производственном предприятии «Цветодинамика» [21, 22]. Результатом взаимодействия ЦДВ и систем ПФС человека-оператора является адаптация к стрессу монотонии, обеспечивающая оптимальный уровень когнитивных функций и уменьшение вероятности принятия неадекватного решения.

Выводы

Обосновано использование цветодинамических преобразователей изображений для визуального моделирования хаотических процессов, отображаемых на двумерных фазовых дискретных плоскостях с детерминированными коэффициентами нерегулярности и гомеоморфизмами, образующими группу устойчивых аттракторов.

Предложено использовать цветодинамические воздействия в качестве экзогенного резонансного хронотопа, взаимодействующего с эндогенными, системорегулирующими процессами в системе ПФС человека-оператора.

Разработан метод воздействия на ПФС человека-оператора в виде экзогенного резонансного хронотопа, взаимодействующего с эндогенными системорегулирующими процессами в системе ПФС человека-оператора.

Список литературы: 1. Ерохин А.Л. О визуальном кодировании информации на основе идентификации когнитивных функций человека-оператора. Сообщение 1 // Бионика интеллекта: научн.-техн. журнал. — 2006. — № 2 (65). — С. 19–26. 2. Винер Н. Кибернетика или управление и связь в животном и машине. — М.: Сов. Радио, 1968. — 325 с. 3. Федоров Л.И. Эндогенные ритмы организма как факторы модуляции параметров стимуляции // Биофизика. — 1996. — Т. 41. — Вып. 3. — С. 718–724. 4. Васильев Ю. М., Гельфанд И. М., Губерман Ш. А., Шик М.Л. Взаимодействие в биологических системах // Природа. — 1969. — № 6. —

С. 13–21; №7. — С. 24–33. 5. Селье Г. Стресс без дистресса. — Рига: Виеда, 1992. — 109 с. 6. Суворов Н.Б., Меницкий Д.Н., Булгакова О.С., Гусева Н.Л., Фролова Н.Л. Психофизиологическая подготовка оператора в биотехнической системе // Системный анализ и управление в биомедицинских системах. — 2003. — Т. 2. — № 1. — С. 4–12. 7. Абдусаматов Р.М., Беркенблит М.Б., Фельдман А.Г., Чернавский А.В. Моторика и интеллект // Интеллектуальные процессы и их моделирование. — М.: Наука, 1987. — С. 13–39. 8. Князева Е. Н., Курдюмов С.П. Интуиция как самодостраивание // Вопросы философии. — № 2. — 1994. 9. Величковский Б.М., Капица М.С. Психологические проблемы изучения интеллекта // Интеллектуальные процессы и их моделирование. — М.: Наука, 1987. — С. 120–141. 10. Дорфман Я.Г., Сергеев В.М. Нейроморфогенез и модели мира в сетях нейронных процессов // Интеллектуальные процессы и их моделирование. — М.: Наука, 1987. — С. 39–65. 11. Ruggieri V., Morelli J. Chromatic perception in relation to an hypothesized cerebral dominance // Percept and Mot. Skills. — 1985. — V. 60. — P. 583–589. 12. Solso R.L. Cognition and Visual Arts, Cambridge, MA: The MIT Press, 1994. 13. Shepard R.N. The perceptual organization of color: An adaptation to the regularities of the terrestrial world // The adapted mind: Evaluator psychology and the generation of culture New York: Oxford University Press, 1992. — P. 495–532. 14. Прибрам К. Языки мозга. — М.: Прогресс, 1975. — 453 с. 15. Бондаренко М.Ф., Бурцев В.Н., Бурцев Вл.Н., Ерохин А.Л. Моделирование стохастических процессов и их применение в практической психологии // Системный анализ, управление и информационные технологии. Вестник ХГПУ: Сборник научных трудов. — Вып. 99. — Харьков: ХГПУ, 2000. — С. 11–14. 16. Зенкин Г.М., Петров А.П. Функциональная организация зрительного процесса и принцип гелыштата // Интеллектуальные процессы и их моделирование. — М.: Наука, 1987. — С. 265–293. 17. Бурцев Вал.Н., Бурцев Вл.Н., Ерохин А.Л. Обеспечение устойчивости системы психофизиологического состояния ЛПР в системах поддержки принятия решений // Проблемы бионики. Всеукр. межвед. науч.-техн. сб. — Харьков: Изд-во ХНУРЭ, 2002. — Вып. 57. — С. 91–94. 18. Исследования и разработка рекомендаций по улучшению условий труда и совершенствованию профессиональной подготовки пожарных ОАЭ и промышленности. Отчет о НИР / ВНИИПО МВД РФ. Рук. Марьин М.И. №ГР01900057262. — М., 1990. — 279 с. 19. Tversky A., Hutchinson J.W. Nearest neighbor analysis of psychological spaces // Psychological Review. — 1986. — Vol. 9, № 3. — P. 3–22. 20. Ерохин А.Л., Бурцев Вал.Н., Бурцев Влад.Н. Исследование стохастических процессов комбинаторно-топологического кодирования информации. Сообщение 1. // Радиоэлектроника и информатика. — 2000. — № 4 (13). — С. 44–48. 21. <http://www.color-dynamic.com/science.htm>. 22. <http://www.colordyn.com.ua/>. 23. Корнелиук В. А., Давыдов Д.М. Использование цветодинамических приборов для комнат психологической разгрузки персонала // Новые промышленные технологии. Минатом. — Вып. 4 (279). — 1997. — С. 82–90. 24. Вейнберг В. Б., Самтаров Д.К. Оптика световодов. — М.: Машиностроение, 1977. — 320 с.

Поступила в редакцию 25.03.07

УДК 519.08

СИНТЕЗ ФОРМАЛЬНОЙ МОДЕЛИ ПОРЯДКОВОЙ ОРДИНАЛЬНОЙ КЛАССИФИКАЦИИ

К. Э. Петров¹, И. В. Евсева²

¹ ХНУРЭ, г. Харьков, Украина

² ХНУРЭ, г. Харьков, Украина, E-mail: irynayev@yahoo.com

Предложен новый подход к формальной идентификации интеллектуальной модели ordinalной классификации, основанный на принципах теории компараторной идентификации. Приведено доказательство того, что решение данной задачи связано с определением модели оптимальной сложности в классе полинома Колмогорова-Габора. Приведены развернутые результаты моделирования, которые подтверждают эффективность описанного метода.

КЛАССИФИКАЦИЯ ОРДИНАЛЬНАЯ, ИДЕНТИФИКАЦИЯ КОМПАРАТОРНАЯ, ИДЕНТИФИКАЦИЯ ПАРАМЕТРИЧЕСКАЯ, КРИТЕРИЙ.

Введение

Задача порядковой ordinalной классификации является частным случаем общей проблемы классификации. Суть задачи заключается в следующем. Задано множество объектов $X = \{x_i\}$, $i = \overline{1, n}$, подлежащих классификации. Каждый объект $x_i \in X$ характеризуется кортежем показателей (частных критериев) $K(x_i) = \langle k_j(x_i) \rangle$, $j = \overline{1, m}$. Структура кортежа $K(x_i)$ одинакова для всех объектов $x_i \in X$, а частные характеристики разнородны по семантике, размерности, типам шкал, в которых они измерены. В частности, характеристики могут быть измерены в количественных или ранговых качественных шкалах. Последнее означает, что на значениях шкалы установлены естественное или искусственное отношения порядка и определено направление доминирования, т. е. направление возрастания качества измеряемого свойства. Кроме перечисленного, предполагается известным количество классов $B = \{b_l\}$, $l = \overline{1, L}$, по которым должны быть распределены объекты классификации $x_i \in X$. При этом на классах установлено отношение порядка и направление доминирования.

Количество классов, отношение порядка и направление доминирования определяют эвристически на основе системного анализа особенностей предметной области, целей классификаций и т. п.

Типичным примером сформулированной задачи является оценка знаний учащихся. В этом случае количество классов равно пяти (плохо b_1 , неудовлетворительно b_2 , удовлетворительно b_3 , хорошо b_4 , отлично b_5). На классах установлено отношение порядка и направление доминирования $b_5 > b_4 > b_3 > b_2 > b_1$.

Каждый учащийся характеризуется кортежем показателей, измеренных в количественных шкалах (например, количество правильно решенных задач, затраченное на выполнение задания время) и качественных шкалах (например, объем фактографиче-

ских знаний, умение устанавливать функциональную зависимость между фактами, логичность мышления и изложения материала и т. п.). Приведенный перечень является иллюстративным и не претендует на полноту. В результате учащийся должен быть отнесен к одному из классов $B = \{b_1, b_2, b_3, b_4, b_5\}$.

Спектр подобных задач очень широк. К ним относятся задачи классификации учащихся на однородные группы при индивидуализации обучения; принятия редакцией решения о возможности публикации научной статьи (классы: принять, принять после доработки, отклонить); принятия решения банком о выдаче кредита, классификации продукции по качеству и т. п.

В настоящее время такого рода задачи решаются эвристическими методами, т. е. на основе экспертных оценок. При этом сложились и развиваются два подхода.

Первый ориентирован на синтез некоторого формально-логического правила классификации. Его реализация связана со структуризацией эвристической процедуры классификации, построением на этой основе некоторого формального правила и его параметрической идентификации. Последнее связано с необходимостью экспертного количественного оценивания параметров сходства, различия, зоны нечувствительности, коэффициентов важности характеристик и т. д. [1, 2]. Экспертные оценки, особенно количественные, имеют принципиально интервальный характер. В процессе многократных последовательных экспертиз накапливается неопределенность, которая во многих случаях делает исходную модель неконструктивной.

Альтернативный подход основан на гипотезе о том, что эксперт более уверенно и точно принимает сложное решение. Например, опытный преподаватель достаточно уверенно и воспроизводимо оценивает знания учащегося в целом, но испытывает затруднения при структуризации факторов, оп-

ределяющих оценку, и количественной оценке вклада каждого из них.

При реализации таких методов [3, 4] собственно классификацию и определение границ классов производит эксперт. Основные усилия направлены на минимизацию количества обращений к эксперту, обеспечение адекватности (транзитивности) и т. п. Предполагается, что после того, как произведена эталонная классификация, распределение новых объектов по классам не вызывает затруднений. Однако это справедливо только на множестве согласованных характеристик [5]. Если характеристики противоречивы (принадлежат к области компромиссов) [6, 7], то для классификации необходимо снова привлекать эксперта. Причем если эксперт другой, то его предпочтения могут несколько отличаться от предпочтений предыдущего. Это может породить невозпроизводимость результатов (отсутствие преемственности).

Независимо от метода классификации единственным носителем и возможным источником информации является эксперт. Это связано с тем, что классификация — интеллектуальный процесс, характеристики которого не поддаются непосредственному измерению. Общая проблема идентификации моделей интеллектуальной деятельности заключается в разработке альтернативных прямым интроспективным экспертным оценкам методов извлечения знаний.

Один из возможных подходов основан на идеях компараторной идентификации.

Целью настоящей статьи является разработка метода формальной идентификации интеллектуальной модели ординальной классификации, базирующейся на идеях компараторной идентификации, по результатам уже принятых ЛПР решений.

1. Постановка задачи

Пусть задано некоторое множество объектов $X = \{x_1, x_2, \dots, x_n\}$, подлежащих ординальной классификации. Определено количество возможных классов $B = \{b_l\}$, $l = \overline{1, L}$.

Каждый объект $x_i \in X$ характеризуется кортежем показателей $K(x_i) = \langle k_j(x_i) \rangle$, $j = \overline{1, m}$. В результате применения некоторой эвристической процедуры классификации, например, ORCLASS или КЛАРА [4, 8, 9], реализована классификация множества X , т. е. получено его разбиение на классы:

$$b_1 = \{x_1^{b_1}, x_2^{b_1}, \dots, x_{n_1}^{b_1}\}, b_2 = \{x_1^{b_2}, x_2^{b_2}, \dots, x_{n_2}^{b_2}\}, \dots, \\ b_L = \{x_1^{b_L}, x_2^{b_L}, \dots, x_{n_L}^{b_L}\},$$

где $\bigcup_{l=1}^L b_l = X$, а $\bigcap_{l=1}^L b_l = \emptyset$. Основываясь на этих результатах, необходимо идентифицировать процедуру (модель) ординальной классификации.

2. Метод формальной идентификации модели ординальной классификации

Согласно теории полезности [6, 9], любому объекту (решению) $x_i \in X$, $i = \overline{1, n}$ можно поставить в соответствие некоторую многофакторную оценку $P(x_i) = F[K(x_i)]$, для которой выполняется условие: если $x_i, x_r \in X$ и $x_i \succ x_r$, то $P(x_i) > P(x_r)$.

Учитывая, что ординальная классификация предполагает наличие отношения порядка на классах, т. е., например,

$$b_1 \prec b_2 \prec \dots \prec b_L, \tag{1}$$

можно записать:

$$P(x_{i_1}^{b_1}) < P(x_{i_2}^{b_2}) < \dots < P(x_{i_L}^{b_L}), \tag{2}$$

$$\forall x_{i_1}^{b_1} \in b_1, i_1 \in I_{n_1}; \forall x_{i_2}^{b_2} \in b_2,$$

$$i_2 \in I_{n_2}; \dots; \forall x_{i_L}^{b_L} \in b_L, i_L \in I_{n_L}.$$

Отношение порядка для элементов внутри любого класса неизвестно, но по постановке задачи известны граничные объекты каждого класса. На основании этого можно записать:

$$P(x_{n_1}^{b_1}) < P(x_{i_1}^{b_1}) < P(x_{i_2}^{b_2}) < P(x_{i_1}^{b_2}) < P(x_{i_2}^{b_2}) < \dots \\ \dots < P(x_{i_L}^{b_L}) < P(x_{i_L}^{b_L}) < P(x_{i_L}^{b_L}), \tag{3}$$

$$i_1 = \overline{1, n_1}, i_2 = \overline{1, n_2}, i_L = \overline{1, n_L};$$

$$\forall x_{i_l}^{b_l} \neq x_{n_l}^{b_l}, \forall x_{i_l}^{b_l} \neq x_{i_l}^{b_l}, l = \overline{1, L},$$

где $x_{n_l}^{b_l}$, $x_{i_l}^{b_l}$ — соответственно, нижний и верхний граничные элементы классов b_l , $l = \overline{1, L}$.

Для удобства представим (3) в виде системы неравенств:

$$P(x_{n_1}^{b_1}) P(x_{i_1}^{b_1}) < 0;$$

$$P(x_{n_1}^{b_1}) P(x_{i_1}^{b_1}) < 0, i_1 = \overline{1, n_1};$$

$$P(x_{i_1}^{b_1}) P(x_{i_2}^{b_2}) < 0, i_1 = \overline{1, n_1};$$

$$P(x_{i_2}^{b_2}) P(x_{i_1}^{b_2}) < 0;$$

$$P(x_{n_2}^{b_2}) P(x_{i_2}^{b_2}) < 0; \tag{4}$$

$$P(x_{n_2}^{b_2}) P(x_{i_1}^{b_2}) < 0, i_2 = \overline{1, n_2};$$

$$P(x_{i_1}^{b_2}) P(x_{i_2}^{b_2}) < 0, i_2 = \overline{1, n_2};$$

$$P(x_{i_2}^{b_2}) P(x_{i_1}^{b_2}) < 0;$$

.....

$$P(x_{i_L}^{b_L}) P(x_{i_1}^{b_L}) < 0;$$

$$\forall x_{i_l}^{b_l} \neq x_{n_l}^{b_l}, \forall x_{i_l}^{b_l} \neq x_{i_l}^{b_l}, l = \overline{1, L}.$$

Определим структуру скалярной многокритериальной оценки $P(x_i)$, $x_i \in X$. В общем случае ее можно представить в виде некоторого фрагмента полинома Колмогорова-Габор. Этот полином, с учетом принятых выше обозначений, имеет вид:

$$P(x_i) = a_0 + \sum_{j=1}^m a_j k_j(x_i) + \sum_{j=1}^m \sum_{q=1}^m a_{jq} k_j(x_i) k_q(x_i) + \dots + \sum_{j=1}^m \sum_{q=1}^m \sum_{r=1}^m a_{jqr} k_j(x_i) k_q(x_i) k_r(x_i) + \dots \quad (5)$$

Полином оптимальной сложности в рамках (5) можно определить, используя генетические алгоритмы [10] или метод группового учета аргументов (МГУА) [11, 12]. При этом для каждого варианта структуры фрагмента (5) необходимо решать задачи параметрической идентификации коэффициентов a_j, a_{jq}, \dots . Независимо от конкретного вида полинома, описывающего модель оптимальной сложности, его, путем расширения пространства переменных [13], т. е. рассматривая нелинейные комбинации как новые переменные, можно представить в виде линейной функции

$$P(x_i) = \sum_{j=1}^s a_j k_j^H(x_i), \quad i = \overline{1, n}, \quad (6)$$

где a_j — относительные безразмерные весовые коэффициенты, для которых выполняются следующие ограничения:

$$0 \leq a_j \leq 1, \quad \sum_{j=1}^s a_j = 1, \quad (7)$$

а $k_j^H(x_i)$ — это нормализованные, т. е. приведенные к изоморфному виду частные критерии.

Нормализация производится по формуле:

$$k_j^H(x_i) = \left(\frac{k_j(x_i) - k_j^{\text{нх}}(x_i)}{k_j^{\text{нл}}(x_i) - k_j^{\text{нх}}(x_i)} \right)^{\alpha_i}, \quad (8)$$

где $k_j(x_i)$ — значение частного критерия; $k_j^{\text{нл}}(x_i)$, $k_j^{\text{нх}}(x_i)$ — соответственно, его наилучшее и наихудшее значения, которые он принимает на множестве допустимых значений $x_i \in X$, $i = \overline{1, n}$; α_i — параметр нелинейности i -ой характеристики.

В зависимости от направления доминирования значений характеристик объектов классификации $k_j^{\text{нл}}(x_i)$ и $k_j^{\text{нх}}(x_i)$ определяются следующим образом:

$$k_j^{\text{нл}}(x_i) = \begin{cases} \max_{x_j \in X} k_j(x_i), & \text{если } k_j(x_i) \rightarrow \max, \\ \min_{x_j \in X} k_j(x_i), & \text{если } k_j(x_i) \rightarrow \min; \end{cases} \quad (9)$$

$$k_j^{\text{нх}}(x_i) = \begin{cases} \max_{x_j \in X} k_j(x_i), & \text{если } k_j(x_i) \rightarrow \min, \\ \min_{x_j \in X} k_j(x_i), & \text{если } k_j(x_i) \rightarrow \max. \end{cases}$$

Определение $k_j^H(x_i)$ не вызывает затруднений для характеристик, измеренных в количественных шкалах. Если же характеристики измерены в ранговых шкалах с заданным направлением доминирования, то производится взаимно-однозначное преобразование в числовые значения. Например, для шкалы измерения знаний (плохо, неудовлетворительно, удовлетворительно, хорошо, отлично) присваиваются числовые «имена» (соответственно, 1, 2, 3, 4, 5). Тогда $k_j^{\text{нх}}(x_i)$ будет равно единице, а $k_j^{\text{нл}}(x_i)$ — пяти. Для учета возможной нелинейности характеристик $k_j^H(x_i)$ используется показатель α_i , который определяется методом наименьших квадратов по нескольким экспертным оценкам.

После подстановки нормализованных значений $k_j^H(x_i)$ в полином (5) получим систему неравенств вида (4), на основе которой при фиксированной структуре скалярной многокритериальной оценки $P(x_i)$ можно решить задачу ее параметрической идентификации.

Задача параметрической идентификации по модели (4) является некорректной по Адамару, так как не имеет единственного решения. Для ее регуляризации используется схема, основанная на определении чебышевской, или средней точки. В силу линейности по параметрам ограничений (4) задача определения параметров является стандартной задачей линейного программирования [14].

Решение перечисленных выше задач позволяет идентифицировать формальную модель классификации. Принцип классификации заключается в следующем. В соответствии с полученной в ходе структурно-параметрической идентификации скалярной многофакторной оценкой $P(x_i)$ вычисляется ее значение для любого объекта, подлежащего классификации. Объект $x_i \in X$ относится к классу b_l , $l = \overline{1, L}$, если выполняется следующее условие:

$$P(x_i^{b_l}) \leq P(x_i) \leq P(x_i^{b_r}), \quad (10)$$

3. Иллюстративный пример

Для оценки адекватности и точности предложенного метода синтеза формальной модели ordinalной классификации использована общая методика оценки качества моделей многофакторного оценивания [15], основанная на принципе внешне-го дополнения.

Для ее реализации с помощью датчика случайных чисел было сгенерировано множество объектов $X = \{x_i\}$, $i = \overline{1, 30}$, каждый из которых характеризуется кортежем нормированных характеристик $K^H(x_i) = \langle k_j^H(x_i) \rangle$, $j = \overline{1, 7}$. Эти данные приведены в табл. 1.

Таблица 1

Исходные данные тестового примера

	k_1^H	k_2^H	k_3^H	k_4^H	k_5^H	k_6^H	k_7^H	$P(x_i)$
a_i	0,19	0,11	0,17	0,06	0,08	0,14	0,25	---
x_1	0,79	0,07	0,92	0,57	0,26	0,58	0,00	0,4504
x_2	0,24	1,00	0,66	0,76	0,45	0,11	0,12	0,3948
x_3	0,30	0,25	1,00	0,67	0,93	0,51	0,84	0,6505
x_4	0,63	0,75	0,66	0,88	0,02	0,60	1,00	0,7028
x_5	0,73	0,69	0,12	0,92	0,15	0,13	0,58	0,4654
x_6	0,52	0,80	0,39	0,68	0,54	0,60	0,24	0,4811
x_7	0,86	0,04	0,79	0,00	0,05	0,12	0,87	0,5404
x_8	0,93	0,77	0,14	0,60	0,68	0,00	0,06	0,3906
x_9	0,31	0,85	0,00	0,58	0,79	0,42	0,87	0,5267
x_{10}	0,61	0,92	0,50	0,82	0,00	1,00	0,14	0,5263
x_{11}	0,17	0,97	0,15	0,26	0,78	0,61	0,45	0,4404
x_{12}	0,57	0,16	0,66	0,55	0,56	0,98	0,55	0,5906
x_{13}	0,19	0,15	0,81	0,09	0,62	0,08	0,56	0,3965
x_{14}	1,00	0,00	1,00	1,00	0,11	0,64	0,05	0,5309
x_{15}	0,00	0,62	0,49	0,21	1,00	0,43	0,86	0,5193
x_{16}	0,18	0,71	0,64	0,32	0,25	0,51	0,24	0,3917
x_{17}	0,49	0,00	0,70	0,15	0,79	0,63	1,00	0,6225
x_{18}	0,39	0,60	0,62	0,58	0,48	0,46	0,83	0,5906
x_{19}	0,13	0,84	0,00	0,73	0,73	0,42	0,51	0,4056
x_{20}	0,99	0,88	0,00	0,18	0,11	0,81	0,15	0,4554
x_{21}	0,83	0,00	0,29	0,10	0,55	0,89	0,33	0,4641
x_{22}	0,79	0,45	0,21	0,42	0,64	0,57	0,52	0,5215
x_{23}	0,13	0,60	0,87	0,06	0,32	0,81	0,56	0,5212
x_{24}	0,81	0,54	0,66	0,57	0,61	0,54	0,07	0,5016
x_{25}	0,00	0,88	0,13	0,55	0,63	0,46	0,25	0,3292
x_{26}	0,53	0,94	0,70	0,53	0,52	1,00	0,00	0,5365
x_{27}	0,53	0,48	0,33	1,00	0,88	0,00	0,88	0,5600
x_{28}	0,66	0,72	0,53	0,59	1,00	0,54	0,43	0,5932
x_{29}	0,48	0,34	0,84	0,05	0,38	0,60	1,00	0,6388
x_{30}	0,42	0,00	0,52	0,47	0,00	0,30	0,41	0,3409

Кроме того, был сформирован кортеж весовых коэффициентов, удовлетворяющих условиям (7) (первая строка табл. 1). На основе этих исходных данных по формуле (6) были вычислены значения оценок всех объектов $P(x_i)$, $i = \overline{1, 30}$ (столбец $P(x_i)$ табл. 1). Все объекты были ранжированы в порядке возрастания значений их оценок и случайным образом разделены на четыре класса с граничными объектами $b_1[x_{25}, x_{19}]$, $b_1[x_{11}, x_{22}]$, $b_1[x_{10}, x_{13}]$, $b_1[x_{28}, x_4]$ (табл. 2). Перечисленные выше данные определили тестовую эталонную ситуацию.

Для проверки предложенного метода в качестве исходных данных из тестовой ситуации были использованы:

- 1) множество объектов $X = \{x_i\}$, $i = \overline{1, 30}$, подлежащих классификации, и кортежи их характеристик $K^H(x_i) = \langle k_j^H(x_i) \rangle$, $j = \overline{1, 7}$ (табл. 1);
- 2) количество классов $B = \{b_l\}$, $l = \overline{1, 4}$ и объекты, определяющие границы классов.

Таблица 2

Сравнительные данные результатов классификации

Объекты	$P(x_i)$	Объекты	$P^M(x_i)$
x_{25}	0,3292	x_{25}	0,33484
x_{30}	0,3409	x_{30}	0,34102
x_8	0,3906	x_{16}	0,38887
x_{16}	0,3917	x_{13}	0,39284
x_2	0,3948	x_8	0,40651
x_{13}	0,3965	x_2	0,40927
x_{19}	0,4056	x_{19}	0,41473
x_{11}	0,4404	x_{11}	0,43773
x_1	0,4504	x_{20}	0,44903
x_{20}	0,4554	x_1	0,44928
x_{21}	0,4641	x_{21}	0,44999
x_5	0,4654	x_5	0,48108
x_6	0,4811	x_6	0,48586
x_{24}	0,5016	x_{24}	0,50445
x_{15}	0,5193	x_{23}	0,50546
x_{23}	0,5212	x_{15}	0,5142
x_{22}	0,5215	x_{22}	0,52022
x_{10}	0,5263	x_{10}	0,52647
x_9	0,5267	x_9	0,53152
x_{14}	0,5309	x_7	0,53197
x_{26}	0,5365	x_{26}	0,5321
x_7	0,5404	x_{14}	0,53599
x_{27}	0,56	x_{27}	0,57749
x_{12}	0,5906	x_{12}	0,58133
x_{18}	0,5906	x_{18}	0,59092
x_{28}	0,5932	x_{28}	0,5965
x_{17}	0,6225	x_{17}	0,60891
x_{29}	0,6388	x_{29}	0,62379
x_3	0,6505	x_3	0,64952
x_4	0,7028	x_4	0,70535

На основе только этой информации была составлена система компараторных неравенств вида (4) и решена задача параметрической идентификации кортежа весовых коэффициентов $\langle a_j^M \rangle$, $j = \overline{1, 7}$ методом, основанным на определении чебышевской точки. Результаты приведены в табл. 3.

С учетом этих значений для каждого объекта $x_i \in X$, $i = \overline{1, 30}$ были вычислены по формуле (6) модельные значения оценок $P^M(x_i)$ (табл. 2),

и на их основе по правилу (11) проведена классификация.

Таблица 3

Модельные значения весовых коэффициентов

	k_1^H	k_2^H	k_3^H	k_4^H	k_5^H	k_6^H	k_7^H
a_i^M	0,191	0,114	0,166	0,079	0,082	0,123	0,245

Анализ результатов показывает полное совпадение состава классов, хотя внутри классов имеется инверсия объектов.

Заключение

Статья посвящена дальнейшему развитию формальных методов структурно-параметрической идентификации моделей интеллектуальной деятельности, изложенных в работах [6, 16] и основанных на идеях компараторной идентификации. Совершенствование и развитие метода открывает широкие перспективы конструктивной интеллектуализации автоматизированных информационно-аналитических и управляющих систем.

Список литературы: 1. Belacel N. Multicriteria assignment method PROAFTN: Metodology and medical applications. European Journal of Operation Research. 2000. №125. P. 175–183. 2. Zopounidis C., Doumpos M. Multicriteria classification and sorting methods: A literature review. European Journal of Operation Research. 2002. № 138. P. 229–246. 3. Ларичев О.И., Мошкович Е.М. Качественные методы принятия решений. — М.: Физматлит, 1996. — 275 с. 4. Ларичев О.И., Мечетов А.И., Мошкович Е.М., Фуремс Е.М. Выявление экспертных знаний. — М.: Наука, 1989. — 171 с. 5. Петров К.Э. Постановка и методы решения общей задачи скалярного многофакторного

оценивания // Вестник Херсонского государственного технического университета. — 2004. — № 1 (19). — С. 7–11. 6. Овезгельдыев А.О., Петров Э.Г., Петров К.Э. Синтез и идентификация моделей многофакторного оценивания и оптимизации. — К.: Наук. думка, 2002. — 164 с. 7. Евсеева И.В. Комплексный подход к классификации обучаемых // Вестник Херсонского государственного технического университета. — 2004. — № 1 (19). — С. 490–494. 8. Ларичев О.И. Теория и методы принятия решений, а также хроника событий в волшебной стране. — М.: Логос, 2000. — 294 с. 9. Райфа Г. Анализ решений. — М.: Наука, 1977. — 408 с. 10. Петров Э.Г., Булавин Д.А. Петров К.Э. Использование генетических алгоритмов для решения задачи структурно-параметрической идентификации модели индивидуального многофакторного оценивания // Бионика интеллекта. — 2004. — № 60. — С. 17–27. 11. Петров Э.Г., Булавин Д.А., Петров К.Э. Решение задачи структурно-параметрической идентификации модели индивидуального многофакторного оценивания методом группового учета аргументов // АСУ и приборы автоматки. — 2004. — Вып. 129. — С. 4–13. 12. Овезгельдыев А.О., Петров К.Э. Структурно-параметрическая идентификация модели индивидуального многофакторного оценивания методом группового учета аргументов // Кибернетика и системный анализ. — 2007. — № 1. — С. 151–160. 13. Cover T.M. Geometrical and statistical of systems of linear inequalities with applications in pattern recognition // IEEE Trans. On Electronic Computers. 1965. № 14. P. 326–334. 14. Зуховицкий С.И., Авдеева Л.И. Линейное и выпуклое программирование. — М.: Наука, 1967. — 460 с. 15. Петров К.Э., Колесник Л.В. Методика верификации достоверности и точности моделей определения предпочтений лиц, принимающих решения // Радиоэлектроника и информатика. — 2005. — № 3. — С. 62–69. 16. Петров К.Э. Проблема формализации интеллектуальной деятельности человека // Бионика интеллекта. — 2004. — № 61. — С. 86–90.

Поступила в редколлегию 26.03.07

УДК 615.47:617-089

ПОИСК ИНФОРМАТИВНЫХ КОЛИЧЕСТВЕННЫХ ПОКАЗАТЕЛЕЙ ЭЛЕКТРОМИОГРАФИЧЕСКОГО СИГНАЛА. Сообщение 1

А. И. Бых¹, Т. В. Жемчужкина², Т. В. Носова³¹ ХНУРЭ, г. Харьков, Украина² ХНУРЭ, г. Харьков, Украина³ ХНУРЭ, г. Харьков, Украина, bykh@kture.kharkov.ua

Работа посвящена поиску информативных количественных показателей ЭМГ-сигналов для дифференциальной диагностики состояния нервно-мышечной системы. Для исследования применяются известные методы анализа медико-биологических сигналов, в частности, методы количественного анализа вариабельности ритма сердца. В данной работе проводились исследования методом временного статистического анализа биомедицинских сигналов.

ЭЛЕКТРОМИОГРАММА, АНАЛИЗ СТАТИСТИЧЕСКИЙ, ДИАГНОСТИКА ДИФФЕРЕНЦИАЛЬНАЯ, НЕЙРОПАТИЯ, НЕЙРОНОПАТИЯ.

Введение

Анализ микроструктуры интерференционной электромиограммы (ЭМГ), отведенной накожными электродами, позволяет выявлять патологические изменения в периферической части двигательного аппарата — мышце, а также в нервных проводниках и передних рогах спинного мозга. Специфические изменения величины электрической активности и ее распределение отражают функциональные нарушения опорно-двигательного аппарата. Это обусловило применение электромиографии при обследовании больных с различными двигательными нарушениями для диагностики, определения степени и распространенности поражения, оценки эффективности лечения.

1. Физиологические аспекты электромиографии

При использовании накожного отведения электромиографического сигнала выявляют некоторые патологические формы электрической активности мышц — фасцикуляции, урежение ЭМГ, повышение тонуса мышцы и т. д.

Фасцикуляциями называют подергивания, которые обусловлены возбуждением группы мышечных волокон. Они регистрируются как спонтанные колебания потенциала, сходные с потенциалами действия двигательных единиц (в отличие от фибрилляций, представляющих собой возбуждение отдельного мышечного волокна и сопровождающихся регистрацией коротких потенциалов фибрилляций). Фасцикуляции могут иметь вид единичных импульсов или ряда редких иглообразных колебаний, частота следования которых 5–20 импульсов в секунду, амплитуда 100–300 мкВ и более. Они характерны для переднероговых поражений спинного мозга (боковой амиотрофической склероз, невральная амиотрофия, сирингомиелия, спинальные опухоли и др.) [1].

При парезах, вызванных поражением мотонейронов, в произвольно напряженной мышце регистрируют уреженную в разной степени ЭМГ; при крайних степенях урежения ЭМГ приобретает «частотную» форму. Вторым фактором, играющим существенную роль в формировании редких ЭМГ, является одновременность возбуждения большого количества мышечных волокон. Это может быть связано, с одной стороны, с образованием крупных двигательных единиц в результате компенсаторной реиннервации волокон погибших двигательных единиц интактными аксонами, с другой — с синхронным разрядом функционирующих мотонейронов, весьма характерным для переднероговых поражений.

При ряде заболеваний нервной системы наблюдается повышение тонуса, проявляющееся в появлении выраженной электрической активности мышцы в состоянии произвольного ее расслабления. Отводить электрическую активность тонуса целесообразно с помощью накожных электродов (которые в этом случае более чувствительны, чем игольчатые, так как суммируют потенциалы с большей площади мышцы). Повышение тонической электрической активности — типичное электромиографическое проявление спастичности у больных с пирамидными поражениями. Электрическая активность при синергическом повышении тонуса может иметь амплитуду 250 мкВ и более. Повышение тонической электрической активности наблюдается также при ригидности, характерной для паркинсонического синдрома. В этом случае повышение электрической активности «покоя» большей частью имеет форму тремора. Повышенная, по сравнению с нормой, электрическая активность «покоя» может наблюдаться и в отсутствие специфических нарушений тонуса в результате боли, эмоционального возбуждения, а также у больных с неврозами. При хорее, атетозе и т. п. на ЭМГ различных мышц

регистрируются беспорядочные залпы активности разной длительности. ЭМГ в этих случаях отражает распространенность и пассивность гиперкинеза, особенности координационных отражений разных мышц в насильственных движениях [2, 3].

При патологическом треморе ЭМГ имеет характерный прерывистый вид. Залпы активности перемежаются с периодами отсутствия активности с частотой около 5–10 колебаний в 1 с. Однако необходимо учитывать, что электрическая активность произвольного напряжения может иметь прерывистый вид, напоминая ЭМГ тремора, в связи со слабостью мышцы любого происхождения.

2. Цель и постановка задачи

Все вышеперечисленные патологические характеристики ЭМГ определяются врачом по внешним признакам, то есть носят субъективный характер (человеческий фактор). Таким образом, постановка диагноза зависит от квалификации, опыта, интуиции и физиологического и эмоционального состояния врача. На настоящий момент не существует набора количественных показателей, однозначно дифференцирующих патологии нервно-мышечной системы. Более подробное изучение и анализ ЭМГ позволит, на наш взгляд, обеспечить не только дифференцирование патологий, но и обнаруживать эти патологии на более ранних стадиях, а также сократить время диагностирования.

Таким образом, несмотря на широкое применение анализа ЭМГ в клинической медицине, остается актуальным получение информативных показателей ЭМГ-сигнала для улучшения диагностирования патологий нервно-мышечной системы.

Целью исследования является определение информативных количественных показателей ЭМГ-сигнала для дифференцирования состояний нервно-мышечной системы.

Для достижения поставленной цели необходимо исследовать различные количественные характеристики ЭМГ-сигнала и выбрать наиболее значимые.

Применяемыми методами в данной работе являются: статистический анализ временных и амплитудных характеристик ЭМГ-сигнала.

3. Количественный анализ ЭМГ

В данной работе проводится обработка ЭМГ-сигналов известными методами, применяемыми для анализа медико-биологических сигналов. В частности, в основу анализа были положены методы количественного анализа variability ритма сердца по ритмограмме. Все многообразие этих методов можно разбить на две основные группы: методы анализа во временной области и методы анализа в частотной области.

Среди методов анализа во временной области можно выделить два направления: статистические

методы, основанные на оценке различных статистических характеристик, и геометрические методы, заключающиеся в оценке формы и параметров гистограммы распределения. На первой стадии исследования остановимся на одном из направлений временных методов анализа биомедицинских сигналов — статистическом.

Динамический ряд кардиоинтервалов (ритмограмма) может быть отнесен к числу стационарных или нестационарных. Стационарными называют случайные процессы, протекающие приблизительно однородно и имеющие вид непрерывных колебаний вокруг некоторого среднего значения. Стационарные процессы характеризуются эргодичностью, то есть усреднение по времени соответствует усреднению по множеству реализаций. Иными словами, на любом участке времени мы должны получать одни и те же характеристики. Нестационарные (или переходные) процессы имеют определенную тенденцию развития во времени, и их характеристики зависят от начала отсчета. Практически в каждой кардиоинтервалограмме содержатся элементы нестационарности (фрактальные компоненты).

ЭМГ-сигнал является случайным нестационарным сигналом. Из-за временного дефицита на практике характеристики миографического процесса принято определять по данным небольшого числа реализаций. Известно, что при длительности реализации миографических процессов больше 160 с среднеквадратичное отклонение мало изменяется. Поэтому указанное значение длительности интервала реализации можно считать достаточным для вычисления по одной реализации статистически достоверных характеристик, соответствующих истинным характеристикам процесса, определенным по множеству реализаций [4].

Исходя из вышеизложенного, очевидно, что ЭМГ-сигнал и ритмограмма относятся к одному типу сигналов; следовательно, мы можем применить известные методы анализа ритмограмм для анализа ЭМГ-сигнала.

Для решения поставленной задачи были исследованы ЭМГ-сигналы, полученные с мышц нижних конечностей при различной степени нейропатологии и в норме (более 80 пациентов). На рис. 1 приведен один из исследуемых ЭМГ-сигналов.

Для получения достоверных статистических характеристик предлагается ЭМГ-сигнал пронормировать, то есть разделить все значения данного сигнала на его максимальное значение и исследовать процентное соотношение амплитуд колебаний, составляющих данный сигнал. На рис. 2 приведен фрагмент пронормированного ЭМГ-сигнала. ЭМГ-сигнал представляет собой последовательность случайных колебаний с амплитудами A_n и длительностями T_n (рис. 3).

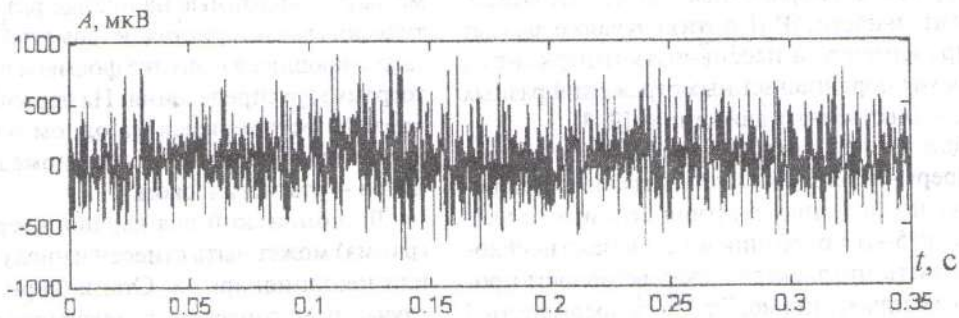


Рис. 1. Фрагмент ЭМГ-сигнала

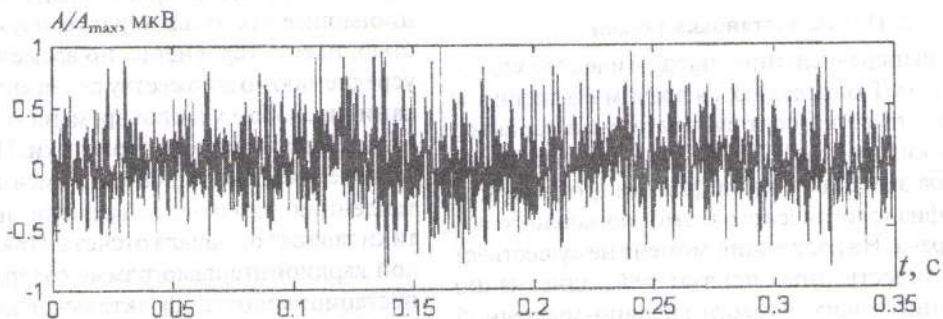


Рис. 2. Фрагмент нормированного ЭМГ-сигнала

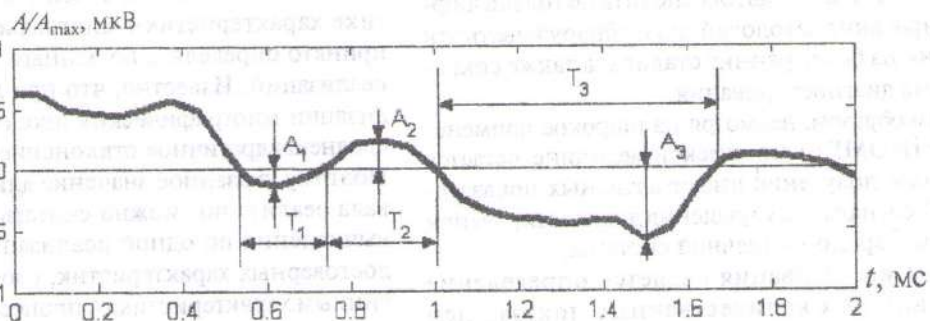


Рис. 3. Пример определения параметров ЭМГ-сигнала

Простейшими характеристиками ЭМГ являются средние значения (математическое ожидание) амплитуд M_A и длительностей M_T колебаний ЭМГ и их среднеквадратические отклонения SD_A и SD_T :

$$M_A = \frac{1}{N} \sum_{i=1}^N A_i, \quad M_T = \frac{1}{N} \sum_{i=1}^N T_i,$$

$$SD_A = \sqrt{\frac{\sum_{i=1}^N (A_i - M_A)^2}{N}}, \quad SD_T = \sqrt{\frac{\sum_{i=1}^N (T_i - M_T)^2}{N}},$$

где N — количество колебаний ЭМГ.

Для анализа ЭМГ-сигнала вычислялись также статистические характеристики более высоких порядков: асимметрия амплитуд As_A и длительностей As_T ; эксцесс амплитуд Ex_A и длительностей Ex_T :

$$As_A = \frac{\sum_{i=1}^N (A_i - M_A)^3}{N \cdot SD_A^3}, \quad As_T = \frac{\sum_{i=1}^N (T_i - M_T)^3}{N \cdot SD_T^3},$$

$$Ex_A = \frac{\sum_{i=1}^N (A_i - M_A)^4}{N \cdot SD_A^4}, \quad Ex_T = \frac{\sum_{i=1}^N (T_i - M_T)^4}{N \cdot SD_T^4}.$$

Некоторые результаты расчета основных статистических характеристик представлены в табл. 1.

В настоящее время для количественного дифференцирования состояний нервно-мышечной ткани (норма, нейропатия, нейроэнпатия и др.) в [4] предложено использовать среднеквадратическое отклонение (СКО) ЭМГ-сигнала. Из рисунка 4 видно, что при рассмотрении СКО (SD_A) значения для некоторых пациентов «выпадают» из общей картины (наличие резких «пиков» на кривой), что подтверждает изменение СКО при возникновении патоло-

Таблица 1

№	M_A	$SD_A, \%$	As_A	Ex_A	M_T	$SD_T, \%$	As_T	Ex_T
1	0,04	512,1	0,58	4,06	7,79	173,47	10,93	199,37
2	0,09	234,95	-0,05	3,86	7,74	888,88	-26,14	711,76
3	0,01	1597,98	-1,04	21,37	7,78	1396,59	-17,84	346,53
4	0,02	688,35	1,24	7,4	7,78	406,98	-24,4	650,92
5	0,03	955,29	0,8	7,61	7,72	602,69	-17,58	335,25
6	0,03	406,48	0,59	6,9	7,47	116,7	4,22	34,85
7	0,08	257,58	0,3	3,52	7,45	789,86	-25,48	700,02
8	0,02	961,83	0,24	10,75	7,46	876,24	-20,74	486,56
9	0,03	773,16	0,17	3,98	7,37	892,33	-20,57	476,58
10	0,05	401,98	0,37	3,64	7,45	887,12	-20,2	465,52
11	0,01	1127,86	0,82	6,05	7,47	221,81	-20,01	511,83
12	0,03	531,24	0,54	8,05	7,4	101,82	1,96	8,47
13	0,01	4197,74	-0,11	6,77	7,41	769	-26,61	741,58
14	0,01	1682,21	-0,11	6,61	7,4	847,33	-21,9	527,77
15	0,02	964,04	-0,67	5,71	7,41	405,34	-24,87	679,37
16	0,03	669,21	0,58	8,17	7,41	409,51	-24,15	652,13
17	0,03	703,74	-0,25	4,19	7,41	1098,83	-23,56	597,99
18	0,13	173,65	0,36	3,44	7,41	660	-13,89	281,31
19	0,02	763,47	0,41	7,03	7,34	530,64	-18,23	361,14
20	0,02	992,09	0,45	4,56	7,39	920,07	-18,76	393,72

гических отклонений в ЭМГ, так как в исследуемой выборке сигналов присутствовали ЭМГ в норме и с различной степенью патологии. Такие же феномены мы наблюдаем и на графиках (рис. 4–5) для математического ожидания (M_A , M_T), СКО (SD_T), асимметрии (As_A , As_T) и, особенно, эксцесса (Ex_A , Ex_T), что дает возможность для предположения статистической связи между значениями этих характеристик и состоянием нервно-мышечной ткани. Дальнейшие исследования в этом направлении необходимы для получения количественных значений интервалов изменения каждой из этих характеристик для каждого из возможных патологических состояний нервно-мышечной ткани, т. е. позволят проводить более точную дифференциальную диагностику.

4. Анализ распределения статистических показателей

На рис. 6–7 представлены распределения каждой из исследуемых статистических характеристик, по которым можно определить диапазоны измене-

ния данных характеристик и интервальное изменение этих характеристик для исследуемой выборки пациентов.

Все исследуемые электромиографические сигналы по результатам статистической обработки были поделены с использованием диагностики по СКО на две группы: ЭМГ-сигналы с $SD_A > 1000\%$ (различная степень патологии нервно-мышечной ткани) и ЭМГ-сигналы с $SD_A < 1000\%$ (условная норма). На рис. 8–9 представлены значения статистических характеристик следующим образом: в первом столбце треугольными маркерами обозначены значения характеристик для пациентов с $SD_A > 1000\%$, во втором столбце круглыми маркерами — значения характеристик для пациентов с $SD_A < 1000\%$. Из способа разделения исходных сигналов очевидно, что треугольные и круглые маркеры для SD_A группируются в два отдельных класса (рис. 8). Маркеры для других статистических характеристик группируются иначе. Изучив результаты данного исследования, можно дифференцировать ЭМГ-сигналы в отдельные группы по плотности, не

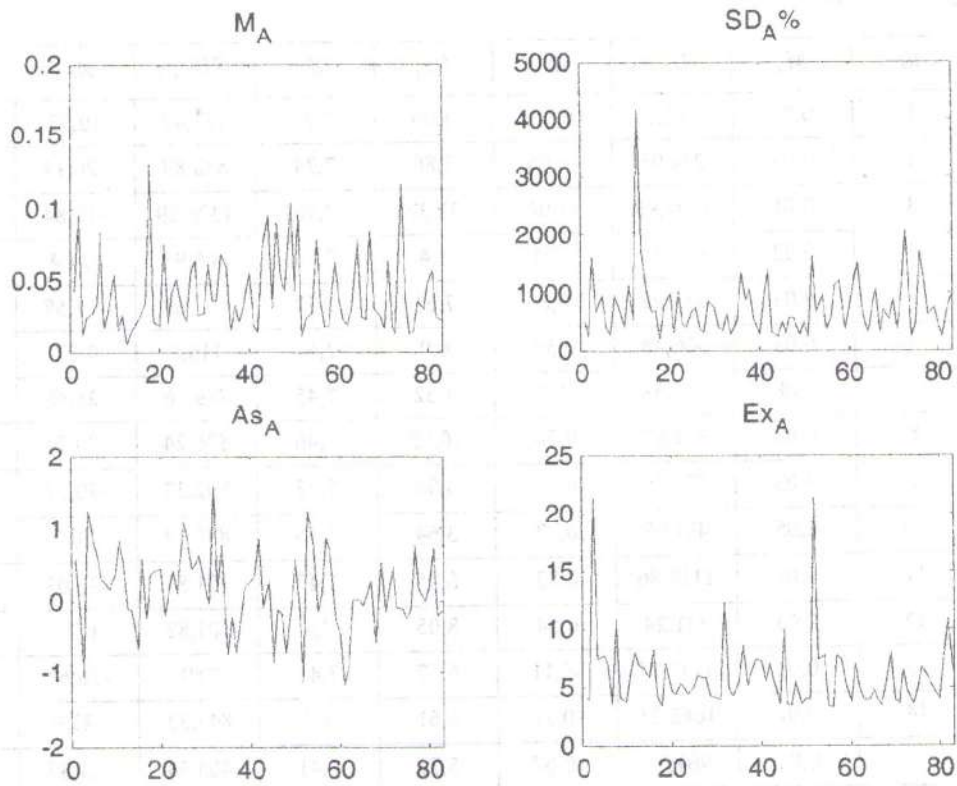


Рис. 4. Значения амплитудных статистических характеристик миографических сигналов для 83 пациентов

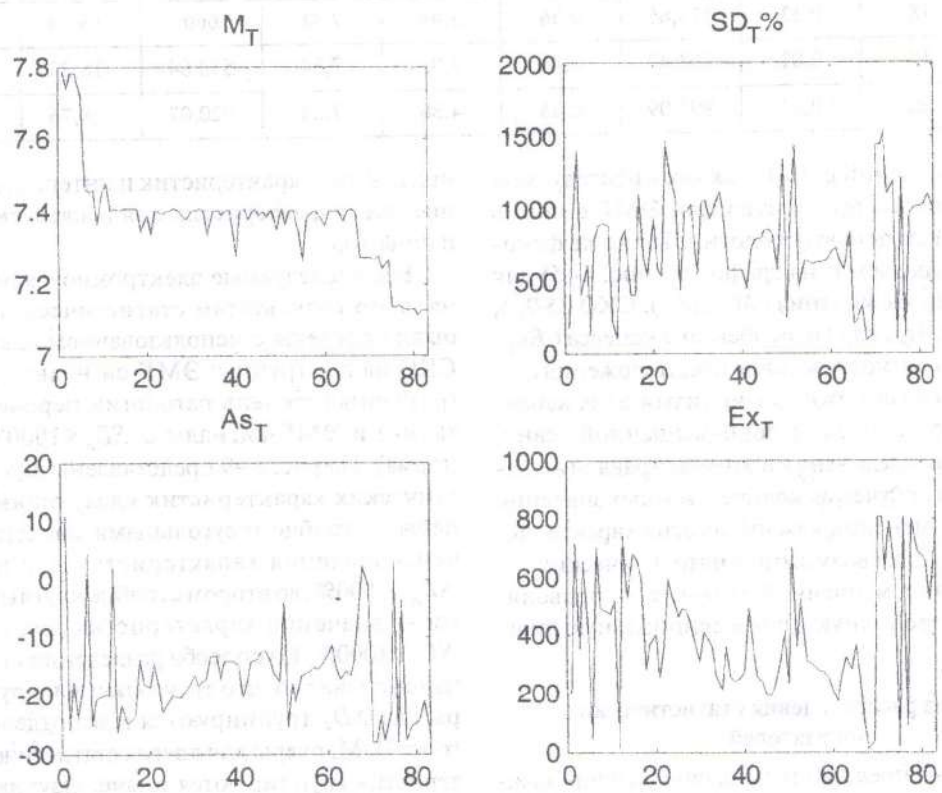


Рис. 5. Значения временных статистических характеристик миографических сигналов для 83 пациентов

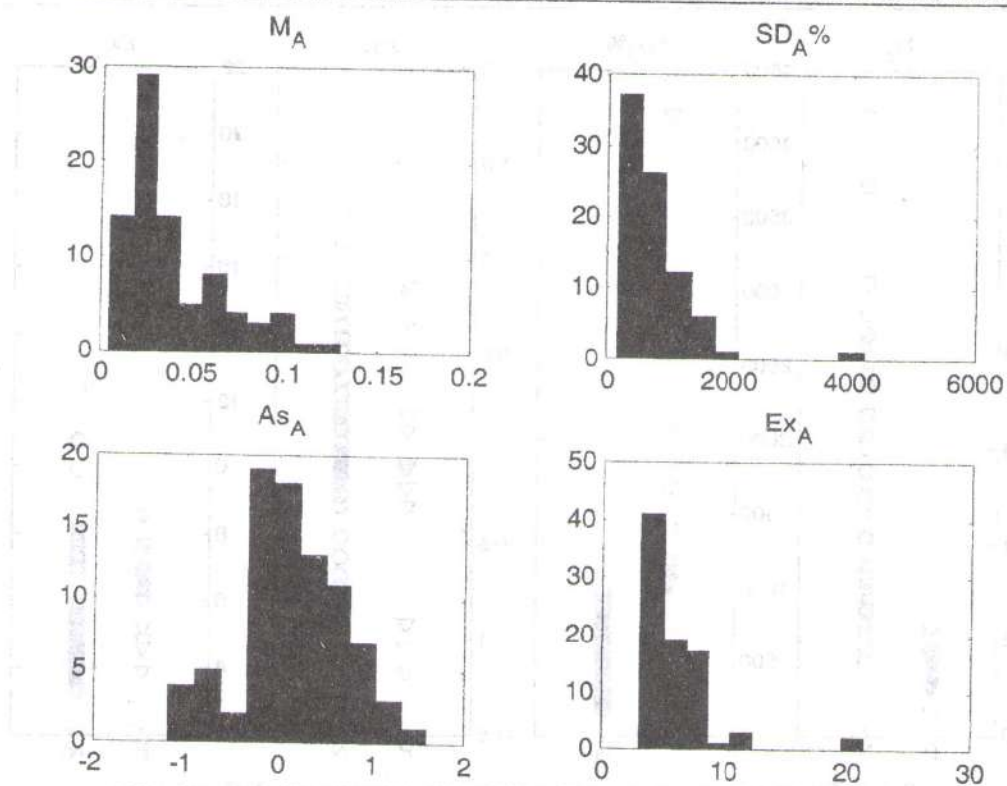


Рис. 6. Распределения амплитудных статистических характеристик

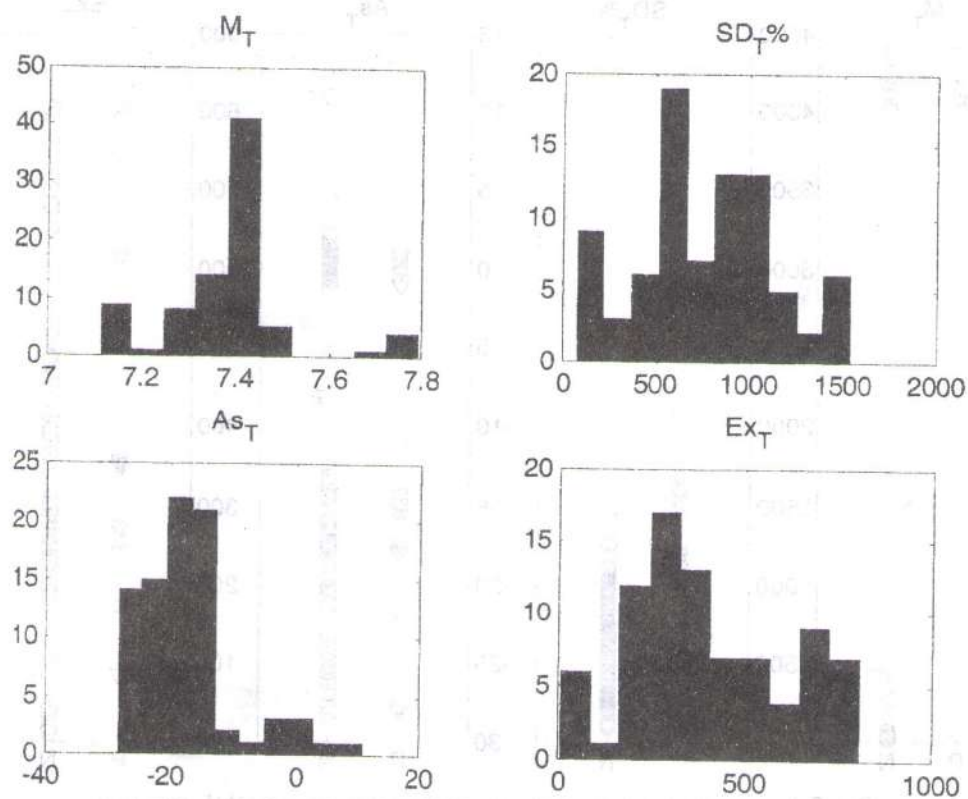


Рис. 7. Распределения временных статистических характеристик

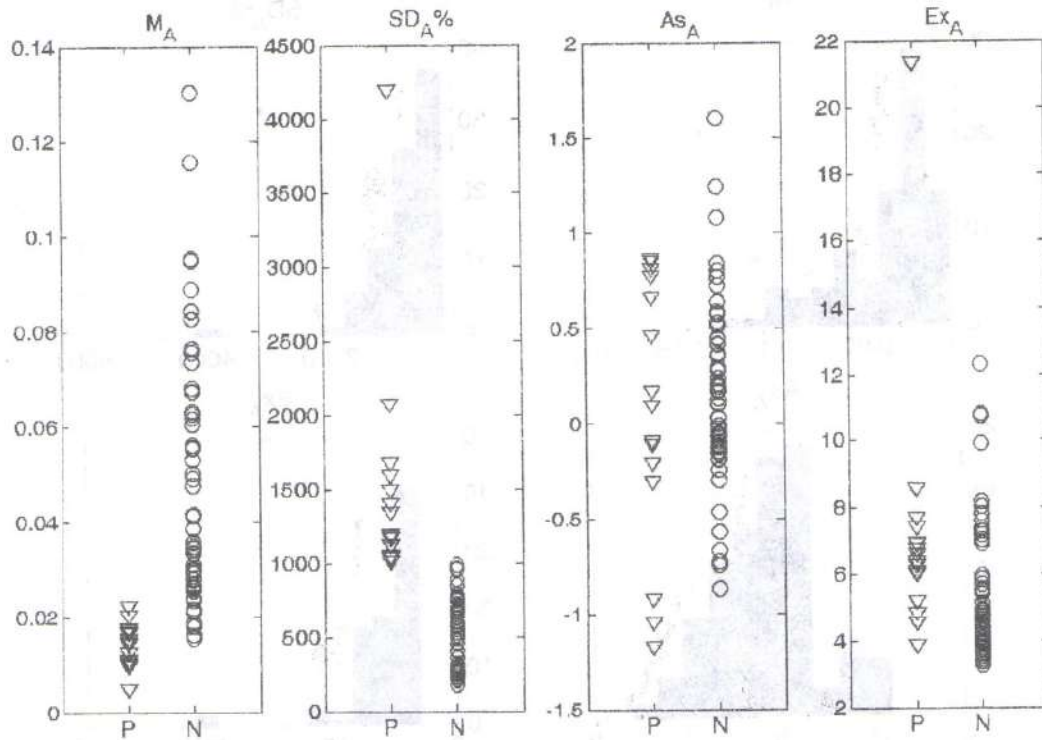


Рис. 8. Группы амплитудных статистических характеристик ЭМГ-сигналов (треугольный маркер — значения характеристик для пациентов с патологией по СКО; круглый маркер — значения характеристик для пациентов с нормой)

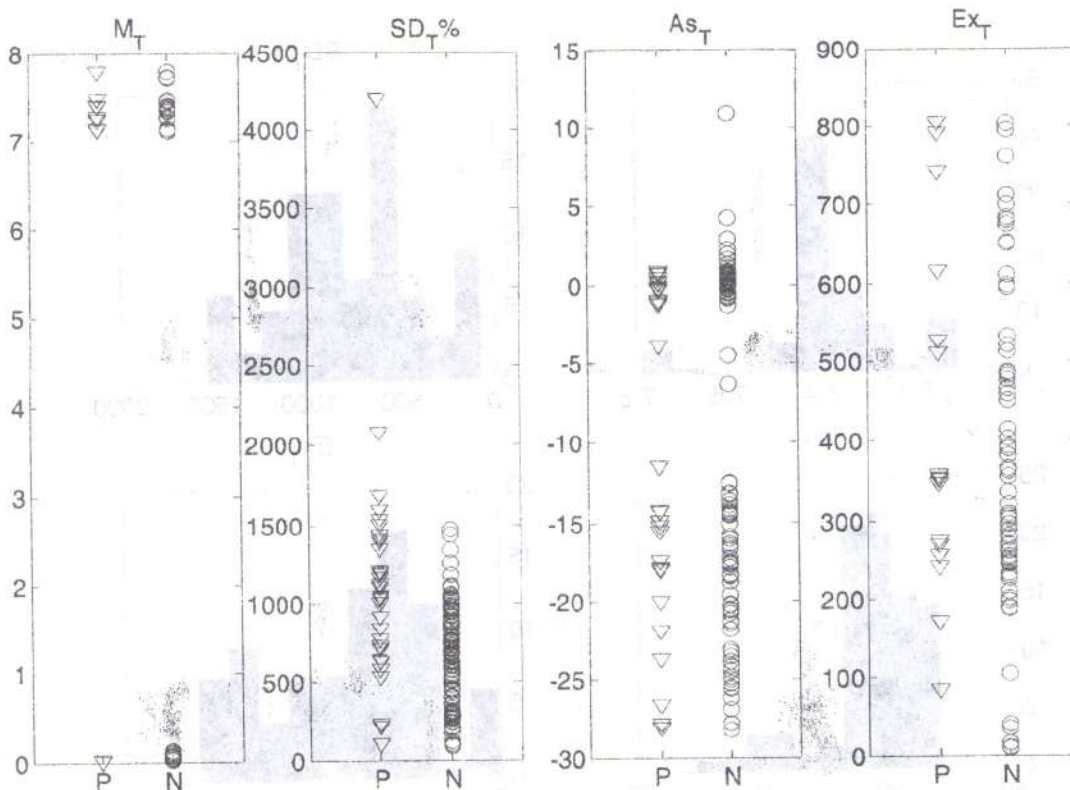


Рис. 9. Группы временных статистических характеристик ЭМГ-сигналов (треугольный маркер — значения характеристик для пациентов с патологией по СКО; круглый маркер — значения характеристик для пациентов с нормой)

только по СКО, но и по другим статистическим характеристикам. Данные исследования позволяют утверждать, что статистическая связь между этими характеристиками отсутствует, поскольку маркеры для каждой из характеристик сгруппировались особым образом, то есть они не являются взаимозаменяемыми, а могут быть взаимодополняющими и нести новую информацию о состоянии нервно-мышечной ткани пациента, что явно видно из рисунков 8–9, поскольку значения различных характеристик группируются различным образом.

Выводы

В данной работе был проведен поиск новых информативных количественных показателей ЭМГ-сигнала, которые могут быть использованы для дифференциальной диагностики состояния нервно-мышечной ткани. Предложены характеристики: математическое ожидание амплитуд M_A и длитель-

ностей M_T , их среднеквадратические отклонения SD_A и SD_T , асимметрия амплитуд As_A и длительностей As_T , эксцесс амплитуд Ex_A и длительностей Ex_T . Показано, что данные характеристики являются статистически независимыми и несут дополнительную диагностическую информацию о состоянии нервно-мышечной ткани пациента.

Список литературы: 1. Зенков Л.Р., Ронкин М.А. Функциональная диагностика нервных болезней. — М.: Медицина, 1991. — 623 с. 2. Румянцева М.Ф., Лосева Т.Н., Бунина Т.П. Руководство к практическим занятиям по физиологии с основами анатомии человека. — М.: Медицина, 1986. — 301 с. 3. Гаусманова-Петрусевич Я. Мышечные заболевания. — Варшава: Польское гос. мед. изд-во, 1971. — 440 с. 4. Носова Т.В., Радченко В.И., Семенец В.В. Способ обработки миографической информации // Збірник наукових праць Харківського університету Повітряних сил. — 2005. — Випуск 5 (5). — С. 78–83.

Поступила в редколлегию 02.04.07

Вісник Харківського національного університету Повітряних Сил і космонавтики, випуск 5 (5), 2005. С. 78–83. **Поступила в редколлегию 02.04.07**

Вісник Харківського національного університету Повітряних Сил і космонавтики, випуск 5 (5), 2005. С. 78–83.

Вісник Харківського національного університету Повітряних Сил і космонавтики, випуск 5 (5), 2005. С. 78–83.

УДК 004.89

ПРИМЕНЕНИЕ ИСКУССТВЕННЫХ ИММУННЫХ СИСТЕМ В ЗАДАЧАХ ВОССТАНОВЛЕНИЯ И ОПТИМИЗАЦИИ НЕЛИНЕЙНЫХ ЗАВИСИМОСТЕЙ

Н. М. Кораблев¹, И. В. Овчаренко², В. В. Токарев³

¹ ХНУРЭ, г. Харьков, Украина, korablev@kture.kharkov.ua

² ХНУРЭ, г. Харьков, Украина, i.ovcharenko@inbox.ru

³ ХНУРЭ, г. Харьков, Украина, tvv.v@mail.ru

Искусственные иммунные системы (ИИС) — направление искусственного интеллекта, которое эмулирует биологическую иммунную систему человека. В работе рассматривается применение ИИС для решения задач восстановления и оптимизации нелинейных функций. Восстановление и оптимизация выполнены на примере тестовых функций на основе алгоритмов клонального отбора и иммунной сети.

ИСКУССТВЕННЫЕ ИММУННЫЕ СИСТЕМЫ, ИММУННАЯ СЕТЬ, КЛОНАЛЬНЫЙ ОТБОР, АФФИННОСТЬ, ОПТИМИЗАЦИЯ.

Введение

Для решения задач идентификации, принятия решений, прогнозирования и управления в сложных системах, когда для исследуемой системы не существует адекватной математической модели, значительный научный и практический интерес представляет использование методов интеллектуальной обработки информации, к которым относятся искусственные нейронные сети, модели на основе нечеткой логики, эволюционное программирование, искусственные иммунные системы (ИИС) и др. Интенсивно развивается теория гибридных сетей, объединяющая в себе технологии нейронных сетей, нечеткой логики и эволюционных вычислений. Если использование искусственных нейронных сетей, эволюционных вычислений и нечеткой логики для решения различных практических задач достаточно хорошо проработано, то искусственные иммунные системы — это новая парадигма мягких вычислений, которая может интегрироваться с другими методиками. В связи с этим задача оптимизации с помощью искусственных иммунных систем является актуальной.

1. Постановка задачи

Классические методы многомерной статистики, используемые в задачах принятия решений, прогнозирования и управления, основаны на определенных предположениях о структуре исходных данных, и, как следствие, попытки их применения приводили к значительным ошибкам при условии несоответствия обучающих наборов этим предположениям. Наиболее перспективным направлением при решении данной задачи является использование универсальных аппроксиматоров широкого класса многомерных нелинейных функций — адаптивных моделей нечеткого логического вывода, которые представляют собой дальнейшее развитие нечетких экспертных систем. Основная трудность их исполь-

зования заключается в структурной и параметрической адаптации — поиске оптимального набора параметров термов лингвистических переменных, структуры множества правил логического вывода. Параметры адаптивной модели нечеткого логического вывода формируются путем их оптимизации в смысле некоторого критерия, формируемого по данным из обучающей выборки. Эта задача относится к классу так называемой «существенно» безусловной задаче глобальной оптимизации. Нахождение глобального экстремума является сложной, а часто нерешаемой задачей. Для этих целей могут использоваться специальные методы, в частности, эволюционные вычисления, случайный поиск и др. Недостатком этих методов является необходимость вычисления градиента функции, зависимость от начального приближения и, как следствие, большая вероятность «застревания» алгоритма в локальном экстремуме.

Широко распространена оптимизация на основе случайного поиска, когда градиент функции заменяется случайным вектором [1]. Генетические алгоритмы как часть парадигмы эволюционных вычислений позволяют решать многие проблемы классических методов оптимизации [2], но далеко не все. С увеличением размерности вектора переменных сильно увеличивается размер области поиска. Поэтому в работе ставится задача восстановления и оптимизации нелинейных унимодальных и мультимодальных функций с использованием искусственных иммунных систем на основе синтеза алгоритмов клонального отбора и иммунной сети.

2. Модели искусственных иммунных систем

Естественная иммунная система человека может рассматриваться как сложная сетевая структура, которая в состоянии реагировать на практически неограниченное множество внешних воздействий, таких как вирусы и бактерии. Следовательно, это

параллельная распределенная адаптивная система, которая имеет огромный потенциал для многих интеллектуальных вычислительных приложений [3].

Иммунная система представляет вычислительный интерес благодаря тому, что ей присущи следующие свойства: распознавание, выделение признаков, разнообразие, обучение, память, распределенное обнаружение, саморегулирование, метадинамика, сетевая организация. ИИС используют вычислительные модели обработки информации в иммунологических взаимодействиях с практическими применениями для решения многих проблем, таких как распознавание образов, поиск данных, компьютерная безопасность, обнаружение ошибок, классификация, оптимизация и др. [4].

В ИИС используется способность естественной иммунной системы вырабатывать новые типы антител и отбирать наиболее подходящие из них для взаимодействия с попавшими в организм антигенами. Методом проб и ошибок иммунная система вырабатывает большое количество антител против множества неизвестных антигенов. При этом антиген рассматривается как задача, которую необходимо решить, а антитело — как вектор, наилучшим образом соответствующий ее решению.

Образование антител основывается на теории клонального отбора при распознавании антигенов. Если антитела распознают антиген, они стимулируются, после чего клонируются и синтезируют новые антитела (осуществляется клональный отбор).

Важной особенностью иммунной системы является то, что она динамически меняется, и иммунный ответ базируется не только на взаимодействии антител и антигенов, но и на взаимодействии антител с другими антителами. При этом обеспечивается как эффект стимуляции антител, так и эффект их супрессии (сжатия), т.е. происходит производство и связывание антител, что имеет существенное значение при проектировании ИИС.

В целом ИИС могут быть описаны как своего рода смесь генетических алгоритмов и нейронных сетей. Преимущество ИИС перед генетическими алгоритмами состоит в том, что в ИИС используются эффекты сети, а преимущества перед нейронными сетями — популяция решений, эволюционный отбор и мутация.

При решении практических задач на основе использования ИИС необходимо решать вопросы выбора подходящего кодирования антител и антигенов, а также выбор соответствующих функций близости/различия (функций аффинности).

Антитела и антигены кодируются одинаково. Для большинства задач самым очевидным представлением является символьная строка чисел или характеристик фиксированной длины, где длина рав-

на числу переменных, а величина — действительное значение самой переменной (binary или real). Но используются и другие, более сложные способы кодирования [5].

В зависимости от принятого способа представления антител и антигенов используются те или иные критерии аффинности. Аффинность — это скалярная величина оценки, показывающая близость результата к оптимальному значению и позволяющая осуществлять отбор антител для клонирования или супрессии [6, 7]. Чаще всего в качестве критерия аффинности используется расстояние Хэмминга или Евклидово расстояние. В зависимости от того, решается задача максимизации или минимизации критерия аффинности, лучшим по популяции считается антитело с самым высоким или низким значением аффинности, соответственно, которое может клонироваться и видоизменяться (мутировать). Антитела с плохой аффинностью, в соответствии с установленным пороговым значением, могут быть удалены из сети. Тем самым реализуется процедура адаптации, которая управляет динамикой системы.

Мутация, используемая в ИИС, подобна мутации, используемой в генетических алгоритмах. Случайные изменения вносятся в гены, ответственные за взаимодействие антител и антигенов, что может привести к улучшению аффинности антител. Мутация антител соответствует поддержанию разнообразия направлений поиска. Однако видоизменение данных имеет смысл не для всех задач, где не очевидно, как видоизменить данные таким образом, чтобы вновь полученные искусственные объекты представляли более вероятные данные.

3. Восстановление нелинейных функций

Для решения задачи восстановления нелинейных функций двух переменных в качестве популяции антигенов ИИС (обучающей выборки) используется набор данных, которые нужно распознать. Антитела представлены в виде строки фиксированной длины $S = \langle s_1, s_2, s_3 \rangle$, состоящей из двоичных символов и представляющей вещественные значения вектора координат антитела в трехмерном пространстве. Исходный вектор координат генерируется случайным образом путем заполнения разрядов в бинарном коде. Для декодирования значений координат используется код Грея. Следовательно, ИИС представляется сетью в виде графа, состоящего из узлов (антител) и взвешенных ребер, которые определяют степень аффинности между соответствующими узлами.

В процессе восстановления нелинейной функции необходимо находить как аффинность связи «антиген-антитело» (степень различия), так и аф-

финность связи «антитело-антитело» (степень подобия). Аффинность вычисляется по соответствующим координатам антитела и антигена после декодирования вектора. В качестве меры аффинности используется расстояние Хэмминга.

Отобранные антитела формируют популяцию клонов, которые затем подвергаются мутации, что ведет к улучшению аффинности антител. Лучшие в популяции антитела формируют клетки памяти, которые восстанавливают нелинейную зависимость.

Решение задачи нахождения глобального экстремума является частным случаем задачи восстановления, при этом в роли антигена выступает целевая функция, значение которой необходимо оптимизировать.

4. Экспериментальные исследования

Для исследования возможности нахождения глобальных экстремумов и восстановления нелинейных функций с помощью ИИС были использованы следующие тестовые функции:

$$F_1(x_1, x_2) = x_1^2 + x_2^2 \quad (1)$$

$$x_1, x_2 \in [-1, 1]; F_1(x_1, x_2) \in [0; 2];$$

$$F_2(x_1, x_2) = 3(1 - x_1)^2 \exp(-x_1^2 - (x_2 + 1)^2) - 10\left(\frac{x_1}{5} - x_1^3 - x_2^5\right) \exp(-x_1^2 - x_2^2) - \frac{1}{3} \exp(-(x_1 + 1)^2 - x_2^2); \quad (2)$$

$$x_1, x_2 \in [-1, 1]; F_2(x_1, x_2) \in [-6.56; 8.11];$$

$$F_3(x_1, x_2) = x_1^2 + x_2^2 - \cos(18x_1) - \cos(18x_2) \quad (3)$$

$$x_1, x_2 \in [-1, 1]; F_3(x_1, x_2) \in [-2; 3.54].$$

4.1. Восстановление тестовых нелинейных функций

Для восстановления функций проведены эксперименты с использованием алгоритмов клонального отбора и теории иммунной сети. В ходе вычислений использовались следующие значения параметров иммунной сети: размер начальной популяции — 121; вероятность мутации — 0.1; разрядность элементов вектора антитела — 22.

Результаты восстановления унимодальной функции F_1 на интервале значений $[-1; 1]$ приведены на рис. 1 для 100 поколений (а) и 700 поколений (б). Видно, что после обучения на этапе 100 поколений размер образовавшейся популяции (клеток памяти) недостаточен для восстановления функции F_1 .

Для оценки качества восстановления функции используется точечная оценка дисперсии. Для 100 поколений оценка дисперсии равна 0,0012; для 700

поколений — 0,0011. Использование эффекта супрессии на основе вычисления аффинности «антитело-антитело» позволяет сократить размер популяции с 1281 до 1075 для 700 поколений.

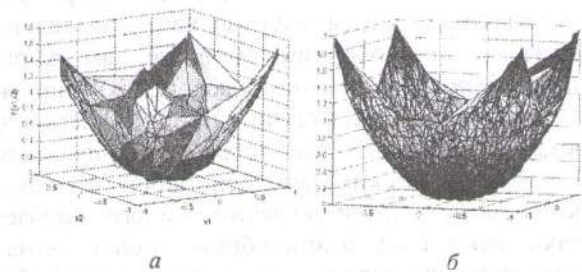


Рис. 1. Результаты восстановления функции F_1

Результаты восстановления мультимодальной функции F_2 на интервале значений $[-3; 3]$ приведены на рис. 2 для 450 поколений (а) и 700 поколений (б). Для 450 поколений оценка дисперсии равна 0,0035; для 700 поколений — 0,0033. Использование эффекта супрессии на основе вычисления аффинности «антитело-антитело» позволяет сократить размер популяции с 691 до 671 для 700 поколений.

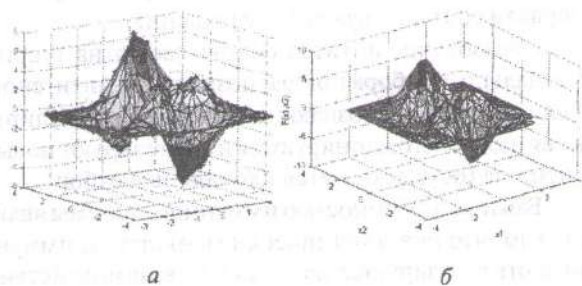


Рис. 2. Результаты восстановления функции F_2

Результаты восстановления функции F_3 приведены для 700 поколений (рис. 3), т.к. на этапах менее 700 поколений образовалось недостаточно клеток памяти для восстановления функции. Оценка дисперсии для 700 поколений равна 0,0051. В результате сжатия сети размер популяции уменьшился с 1442 до 1244 для 700 поколений.

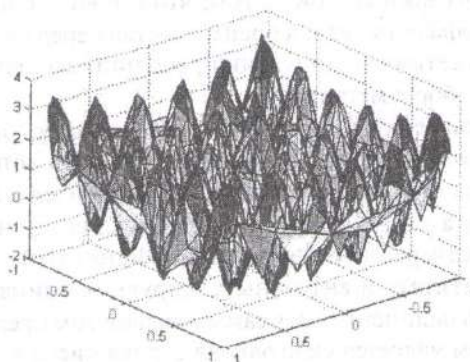


Рис. 3. Результаты восстановления функции F_3

4.2. Нахождение глобального экстремума тестовых нелинейных функций

Иммунный алгоритм, предназначенный для задач оптимизации, упрощен за счет отсутствия популяции антигенов, вместо нее в роли антигена выступает целевая функция оптимизации. При нахождении глобального экстремума вычисление аффинности соответствует вычислению функции $F(x_1, x_2)$ после декодирования x_1, x_2 . В ходе выполнения иммунного алгоритма большинство антител сосредотачиваются в одном пике (рис. 4 — тестовая функция F_2).

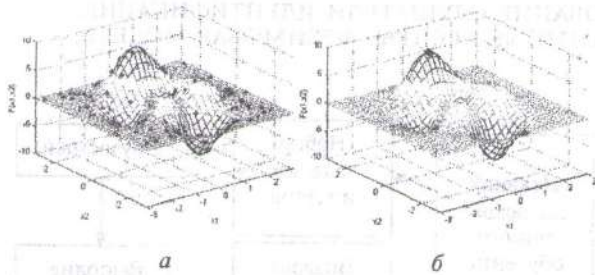


Рис. 4. Оптимизация функции F_2

Графики изменения лучшей и средней по популяции функции цели при нахождении глобального максимума мультимодальной функции F_2 приведены на рис. 5.

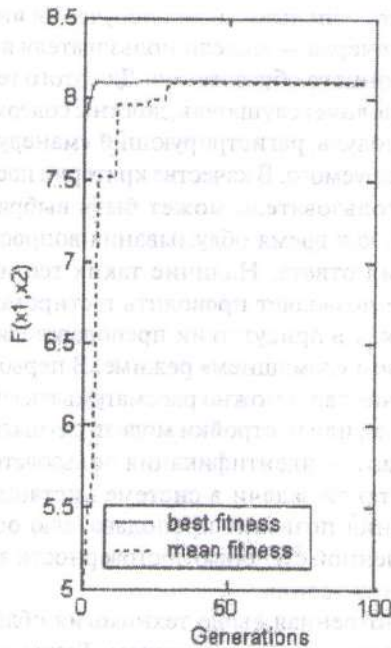


Рис. 5. Результаты оптимизации мультимодальной функции F_2

Графики изменения лучшей и средней по популяции функции цели при нахождении глобального минимума функций F_1 и F_3 приведены на рис. 6а и 6б, соответственно.

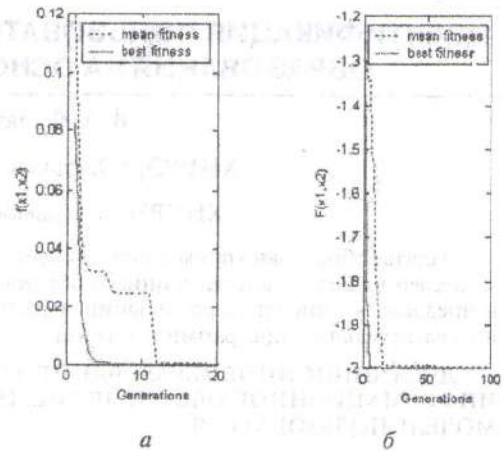


Рис. 6. Результаты оптимизации функций F_1 и F_3

Результаты и перспективы исследований

Научная новизна работы заключается в том, что использование методов теории ИИС для восстановления и оптимизации нелинейных функций на основе синтеза алгоритмов клонального отбора и иммунной сети получило дальнейшее развитие.

Практическая ценность работы состоит в том, что разработаны новые методики восстановления и оптимизации нелинейных зависимостей, которые основаны на способности самообучаться и могут быть использованы в различных практических задачах, в частности, для структурной и параметрической адаптации правил нечеткого вывода.

Список литературы: 1. Жигельский А.А., Жилинкас А.Г. Методы поиска глобального экстремума. — М.: Наука, 1991. — 248 с. 2. Herrera F, Lozano M. Fuzzy genetic algorithms: issues and models // Technical report DECSAI-98116, Dept. of Computer science and A.I., University of Granada, June 1998. — 32 p. 3. Искусственные иммунные системы и их применение / Под ред. Д. Дасгупты. Пер. с англ. под ред. Романюхи А.А. — М.: ФИЗМАТЛИТ, 2006. — 344 с. 4. Timmis J.I., Knight T., De Castro L.N., Hart E. An Overview of Artificial Immune Systems // Computation in Cells and Tissues: Perspectives and Tools for Thought, Natural Computation Series, Springer, 2004 — p. 51-86. 5. Литвиненко В.И., Бидюк П.И., Фелелов А.А., Баклан И.В. Гибридная иммунная сеть для решения задач структурной идентификации // Интеллектуальні системи прийняття рішень та прикладні аспекти інформаційних технологій: Матеріали науково-практичної конференції. — Том 5. — Херсон: Видавництво Херсонського морського інституту, 2005. — С. 41-48. 6. De Castro L. N., Timmis, J. I. Artificial Immune Systems: A Novel Paradigm to Pattern Recognition // In Artificial Neural Networks in Pattern Recognition, SOCO-2002, University of Paisley, UK, 2002 — p. 67-84. 7. De Castro L. N., Von Zuben F. J. Ошибка! Залка не определена. // In Proceedings of GECCO'00, Workshop on Artificial Immune Systems and Their Applications, 2000. — p. 36-37.

Поступила в редколлегию 03.04.07

УДК 519. 72

ИДЕНТИФИКАЦИЯ ПОЛЬЗОВАТЕЛЕЙ В СИСТЕМЕ ДИСТАНЦИОННОГО ОБРАЗОВАНИЯ НА ОСНОВЕ ПРОГРАММНЫХ АГЕНТОВ

В. А. Филатов¹, Р. В. Семенец²¹ ХНУРЭ, г. Харьков, Украина, Filatov_val@ukr.net² ХНУРЭ, г. Харьков, Украина, Ars_sem@ukr.net

В статье обосновано применение агентных технологий для решения задачи идентификации пользователей в системе дистанционного образования, построена математическая модель пользователя и предложен вариант ее реализации на основе агентного подхода, приведена логическая схема интеллектуального программного агента.

ДИСТАНЦИОННОЕ ОБРАЗОВАНИЕ, ТЕСТИРОВАНИЕ, СЛУШАТЕЛИ, ИДЕНТИФИКАЦИЯ, ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ, ПРОГРАММНЫЕ АГЕНТЫ, ФРЕЙМОВАЯ МОДЕЛЬ, МОДЕЛЬ ПОЛЬЗОВАТЕЛЯ.

Введение

Информационные технологии стали неотъемлемой частью существования современного общества. Сейчас трудно представить такую область человеческой деятельности, где не применялись бы персональные компьютеры с комплексом программных средств информационной поддержки. К одной из них — наиболее динамично развивающейся — относится сфера образования. За последние годы качественно изменились все составляющие технологии образования: от методики преподавания до технических средств обучения. Являясь следствием объективного процесса информатизации, интенсивно развивается принципиально новое направление в обучении — дистанционное образование.

Важно отметить, что это направление получило развитие не только в рамках национальных систем образования, но и в коммерческих компаниях с преимущественной ориентацией на подготовку в области бизнеса. Частные корпоративные образовательные сети созданы такими компаниями, как IBM, General Motors, Ford и др. Многие из этих образовательных систем намного опережают по сложности и объему системы, созданные в университетах.

1. Постановка проблемы

В технологии дистанционного образования используется следующая схема обучения и контроля знаний. Обучаемый получает комплекс методических материалов по теме: лекционный материал, набор практических заданий, лабораторный практикум и набор тестов как элемент контроля усвоения материала. Тесты с результатами возвращаются преподавателю для оценки степени усвоения материала.

Таким образом, в деканате дистанционного обучения на каждого студента формируется база данных с результатами тестовых задач по разным дисциплинам, представленная на рис. 1.

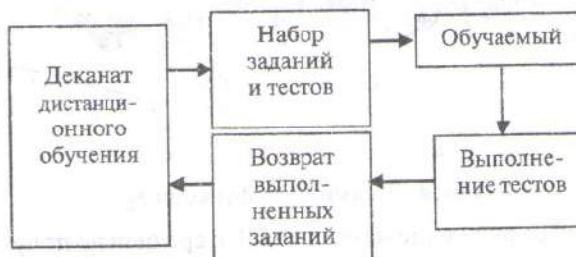


Рис. 1. Схема организации процесса обучения в системе дистанционного образования

Набор выполненных студентом тестовых задач может быть использован для получения индивидуального почерка — модели пользователя в системе дистанционного образования. Для этого тесты, которые выполняет слушатель, должны содержать служебный модуль, регистрирующий «манеру поведения» тестируемого. В качестве критерия построения модели пользователя может быть выбран такой параметр, как время обдумывания вопроса и формирования ответа. Наличие таких тестирующих программ позволяет проводить тестирование слушателей как в присутствии преподавателя, так и в автономном «домашнем» режиме. В первом случае выполнение теста можно рассматривать как режим «обучения», или настройки модели слушателя. Вторым вариантом — идентификация пользователя. Реализация такой задачи в системе дистанционного образования позволит преподавателю оценивать с определенной степенью достоверности авторство «домашних» тестов.

Рассмотренная выше технология обладает одним существенным недостатком. Тесты, выполняемые одним слушателем по разным предметам, разработаны разными преподавателями, структура и технология тестирования в которых, в общем случае, может быть совершенно разной. Поэтому критерий «время обдумывания одного вопроса» в системе идентификации слушателя будет неэф-

фективен. Сохранить привлекательность метода идентификации пользователя на основе временных показателей можно, применив модульный подход. Для этого разрабатывается одна универсальная оболочка, к которой может быть подключен любой тестирующий программный модуль. Однако, чтобы «войти» в специализированный тест по дисциплине, необходимо пройти «входной» контроль или «входное тестирование». Структура такой системы представлена на рис. 2.

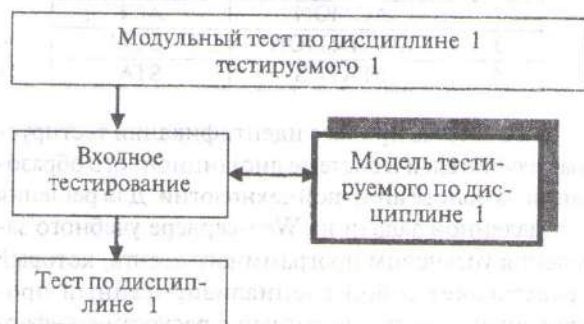


Рис. 2. Схема универсального модульного теста

Наличие модуля «Входное тестирование» обязывает всех слушателей проходить универсальный тест при выполнении любого теста по изучаемой дисциплине. Таким образом, модульный тест будет сохранять в себе результаты как теста по дисциплине, так и индивидуальный почерк на основе входного теста. Затем необходимо обработать все модульные тесты одного слушателя на предмет сравнения результатов входных тестов. На рис. 3 изображена схема анализа модульных тестов.

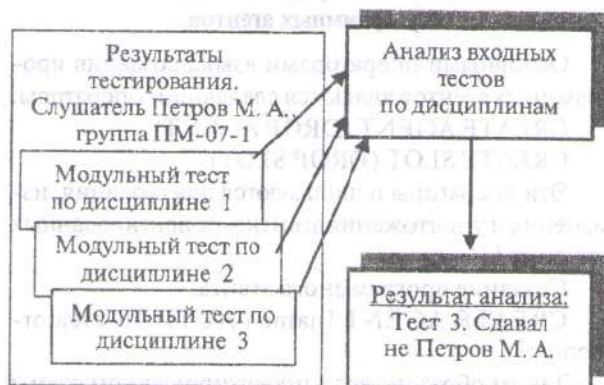


Рис. 3. Подсистема анализа модульных тестов

В рассмотренной технологии дистанционного образования в качестве универсального входного теста может быть предложен «тест на клавиатуре». Слушателю последовательно выводятся на экран группы слов, которые он должен правильно набрать на клавиатуре. В таком тесте эффективно может быть использован критерий «время выполнения за-

дания» и реализован метод идентификации на основе функции подобия моделей поведения тестируемых.

Цель исследования. Целью проводимых исследований является разработка математической модели идентификации пользователей в системе дистанционного образования и ее реализация на основе технологии программных агентов.

2. Разработка математической модели

Для разработки математической модели пользователей на первом этапе необходимо предусмотреть тестирование в присутствии преподавателя.

Пусть $U = \{u_k\}, k=1, K$ — множество тестов по дисциплинам, которые необходимо сдать студенту в присутствии преподавателя, $D = \{d_k\}, i=1, L$ — множество заданий во входном тесте.

Множество D является полным множеством всех заданий входного теста. Отношение слушателя, выполняющего тестирование по первой дисциплине, к полному множеству заданий входного теста D представляется в виде бинарного вектора, содержащего единичные элементы в той позиции, которой соответствует ответ в отведенный регламент времени на это задание, и нулевые в противном случае. Интервалы времени переходов из 1 в 0 по каждому заданию теста определяются разработчиком теста. В результате тестирования формируется бинарная матрица $B = \|b_{kj}\|$, проиндексированная по двум множествам: тестируемых дисциплин $U = \{u_k\}$ и полного множества заданий во входном тесте $D = \{d_k\}$.

Далее на основании анализа матриц $B_l = \|b_{kj}\|, l=1, L$, где L — количество студентов, одним из методов кластерного анализа производится настройка моделей слушателей в системе дистанционного образования.

Студенты могут выполнять тесты по изучаемым дисциплинам в домашних условиях без преподавателя. Присылаемые результаты тестов по изучаемым дисциплинам будут содержать в себе служебную информацию — данные входных тестов.

Следующий этап — сравнение данных входного теста с моделью слушателя для оценки степени самостоятельности выполнения заданий.

Для получения количественной характеристики степени общности модели 1-го слушателя с данными очередного входного теста воспользуемся понятием меры подобия $0 \leq S_l \leq 1$. Меру подобия предлагается вычислять при помощи выражения:

$$S_l = (p_1) / (p_1 + p_{10} + p_{01}), \quad (1)$$

где p_1 — количество общих элементов модели и входного теста; p_{10} — количество элементов, принадлежащих модели, но отсутствующих во входном тесте;

p_{01} — количество элементов, отсутствующих в модели, но принадлежащих модели входного теста.

Одним из способов программной реализации рассмотренной выше математической модели идентификации пользователей в системе дистанционного образования может быть подход на основании технологии программных агентов.

3. Обоснование модели мультиагентной системы

Агентные технологии имеют большое количество преимуществ по сравнению с традиционными технологиями для решения специфических проблем. Адаптированные в наиболее общем виде к объектно-ориентированным технологиям, включая и технологию распределенных объектов, они представляют собой мощное и эффективное инструментальное средство для решения специфических задач информационной поддержки. Перспективным направлением является интеграция агентных технологий с традиционными информационными технологиями. Такие свойства агентных систем позволяют расширить область их применения и на сферу дистанционного образования [1].

Мультиагентной системой назовем многокомпонентную систему, состоящую из программных агентов, выполняющих локальные задачи информационной поддержки для достижения общей цели, — управление распределенной информацией.

Основные составляющие мультиагентной системы:

Агент — приложение, функционирующее на локальном компьютере и используемое в общем случае для реализации доступа и администрирования информационных ресурсов.

Система управления агентами (менеджер агентов) — программа-администратор мультиагентного пространства, осуществляющая следующие функции:

- контроль работы системы;
- дистанционное управление поведением мультиагентной системы;
- модификация и настройка функций программных агентов на удаленных компьютерах;
- накопление и анализ информации, полученной от агентов в процессе их функционирования.

Конструктор агентов — подсистема, обеспечивающая логическое и физическое проектирование агентов, их генерацию и модификацию [2, 3].

Модель интеллектуального агента построена на основе фрейма. В общем случае модель фрейма имеет вид:

$$FR = [(R_1, A_1), (R_2, A_2), \dots, (R_n, A_n)], \quad (2)$$

где FR — имя фрейма; R — имя слота; A — значение слота.

На основании модели фрейма (2) может быть представлена концептуальная модель программного агента в терминах <объекты>, <условия>, <действия>, <приоритет>. Каждый слот будет формироваться из четырех атрибутов базовых сущностей, представленных в табл. 1.

Таблица 1

Типы слотов

№	Сущности	Имя
1	OBJECT	OBG
2	ACTION	ACT
3	CONDITION	CON
4	STATUS	STA

Рассмотрим процесс идентификации тестируемых слушателей в системе дистанционного образования на основе агентной технологии. Для решения поставленной задачи на Web-сервере учебного заведения установим программного агента, который представляет собой специализированный программный модуль, связанный с ресурсами информационной системы, — базой данных зарегистрированных слушателей. Его первой задачей является проверка соответствия вводимого логина и пароля записям базы данных. Второй задачей программного агента является осуществление входного тестирования слушателя и программная реализация математической модели на основе анализа клавиатурного почерка. При этом один программный агент может решать несколько задач по управлению доступом к ресурсам данного персонального компьютера.

4. Языковые средства создания программных агентов

Основными операторами языка создания программных агентов являются следующие операторы:

CREATE AGENT (DROP AGENT)
CREATE SLOT (DROP SLOT)

Эти операторы используются для создания, изменения и уничтожения агентно-ориентированных структур [4].

Создание программного агента.

CREATE AGENT [name | AUTHOR creator-identifier]

Таким образом, если проектировщиком схемы под именем test будет пользователь Admin, то данный оператор будет выглядеть следующим образом:

CREATE AGENT test AUTHOR Admin

Схема удаляется с помощью оператора DROP AGENT, который имеет следующий формат:

DROP AGENT [RESTR | CASCAD]

Если указано ключевое слово RESTR (именно оно принимается по умолчанию), схема взаимосвязей программных агентов должна быть пустой, ина-

че выполнение операции будет отменено. Ключевое слово CASCAD определяет автоматическое удаление всех связанных с данным агентом других программных агентов.

После создания структуры программного агента можно приступить к созданию функциональных слотов-отношений, входящих в состав агента. Для этой цели используется оператор CREATE SLOT, имеющий следующий общий формат:

```
CREATE SLOT slot-name {{OBG}, {CON}, {ACT}, {STA}, {PROC}}
```

В результате выполнения этого оператора будет создан слот-отношение, имя которого определяется параметром slot-name. Набор типовых атрибутов [OBG], [CON], [ACT], [STA], [PROC] определяет функциональное назначение слота-отношения. Формат типовых атрибутов:

OBG [character], <определяет путь к нахождению объекта>;

CON [character], <содержит условие выполнения операции, указанной во фразе ACT, над объектом, путь к которому содержится в OBG>;

ACT [character], <определяет тип операции над объектом OBG>;

STA [character, 1|0|IS NULL], <содержит признак завершения операции над объектом OBG>;

1 — операция завершена успешно;

0 — операция завершена аварийно;

IS NULL — операция находится в стадии выполнения>;

PROC [character], <расширение опции CON, определения условий выполнения операций над объектами >; slot-name, <ключевой атрибут отношения AGENT [name]>.

С течением времени структура программного агента может изменяться, для удаления слота-отношения можно применять оператор, имеющий следующий формат:

```
DROP SLOT slot-name [RESTR|CASCAD]
```

Ключевое слово RESTR (принимается по умолчанию), при этом связи слотов-отношений со слотами-отношениями других программных агентов должны быть отключены, иначе выполнение операции будет отменено. Ключевое слово CASCAD определяет автоматическое удаление всех слотов-отношений других программных агентов, связанных с данным слотом-отношением.

5. Пример

Рассмотрим пример логической модели программного агента с именем CONTROL, который выполняет две задачи и, соответственно, состоит из двух слотов.

Слот A1 содержит четыре атрибута. Первый атрибут определяет объект действия — модуль провер-

ки вводимого логина и пароля password.exe, который находится в каталоге c:\MP. Второй атрибут определяет условие, при котором должно состояться выполнение операции, — если текущее время удовлетворяет ограничениям Time > 8-00 and Time < 18-00. Третий атрибут определяет вид операции или вид действия — запуск приложения password.exe. Четвертый атрибут определяет уровень приоритета — очень важно.

Слот A2 также содержит четыре атрибута и выполняет задачу активизации модуля model.exe, в котором программно реализованы все функции по построению математической модели почерка пользователя. Эта задача выполняется только после успешного завершения предыдущей задачи A1. Структура программного агента, выполняющего рассмотренные выше функции, приведена в табл. 2.

Таблица 2
Логическая структура программного агента CONTROL

Имя слота	Объект	Условие	Действие	Приоритет
A1	OBG: c:\MP\ password. exe	CON: IF 00<Time< 18-00	ACT: RUN c:\MP\ password. exe	STA: VERY IMP
A2	OBG: c:\MP\ model. exe	CON: AFTER A1	ACT: RUN c:\MP\ model.exe	STA: IMP

Реализация предлагаемого способа идентификации слушателей в системе дистанционного образования позволит повысить эффективность мер авторского контроля обучения.

Выводы

В статье рассмотрены вопросы построения модели пользователя в системе дистанционного образования. Предложен современный подход к идентификации пользователей на основе технологии программных агентов, комплекс языковых средств создания программного агента, разработана логическая схема интеллектуального программного агента, приведен пример логической схемы агента.

Список литературы: 1. Simon A. R. Strategic Database Technology: Management for the Year 2000. — Morgan Kaufmann Publishers, 1995 ISBN 1-55860-264-X, pp. 100–250. 2. Wooldridge M. and N. Jennings (1995) Intelligent agents: theory and practice. The Knowledge Engineering Review, 10(2), pp. 115–152. 3. Intelligent Agents II / Ed. M. Wooldridge, J. P. Mueller, and M. Tambe. — Springer-Verlag, 1996. — Lecture Notes in Artificial Intelligence, — V. 1037, p. 356. 4. Пономаренко Л. А., Філатов В. О. Програмні агентні технології в адмініструванні баз даних. // Вісник Київського торговельно-економічного університету. — К. — Вип. 3/2001. — С. 68–73.

Поступила в редколлегию 15.04.07

УДК 519.7

ИНТЕЛЛЕКТУАЛИЗАЦИЯ ПРОЦЕДУР ОПРЕДЕЛЕНИЯ КООРДИНАТ МОБИЛЬНЫХ ТЕРМИНАЛОВ В БЕСПРОВОДНЫХ СЕТЯХ

А. Б. Колесник¹, Л. В. Колесник²¹ХНУРЭ, г. Харьков, Украина, gunter@arko.ua²ХНУРЭ, г. Харьков, Украина

Проведен анализ текущего состояния развития систем определения координат и параметров состояния мобильных объектов в беспроводных сетях. Предложен метод определения координат беспроводных терминалов, не использующий системы глобального позиционирования.

ПОЗИЦИОНИРОВАНИЕ МОБИЛЬНЫХ ОБЪЕКТОВ, БЕСПРОВОДНЫЕ СЕТИ, УЗЕЛ СЕТИ, ТОЧКА ДОСТУПА.

Введение

Для эффективного управления беспроводными сетями немаловажную роль играет процедура своевременного и непрерывного обновления информации о перемещениях мобильных терминалов внутри сети и возможной миграции терминалов из соты в соту и из одной подсети в другую. Решение данной проблемы при помощи систем, использующих методы, основанные на GPS и GIS (система глобального позиционирования и система глобального информирования, соответственно), далеко не всегда экономически и технически оправдано. Связано это не столько с относительной дороговизной абонентской платы за такие услуги, но и со сравнительно низкой точностью определения физических координат мобильных объектов. Сама технология спутникового позиционирования позволяет определять физические координаты с точностью до нескольких метров, однако министерство обороны США, которой принадлежит система GPS, искусственно завышает погрешность определения местоположения объектов до 30 метров, что может быть неприемлемо для адекватного отслеживания миграций мобильных объектов в условиях сравнительно небольших по площади беспроводных сетей.

Целью настоящей статьи является разработка метода определения координат мобильных терминалов без привязки к спутниковым системам глобального позиционирования мобильных объектов с заданной точностью.

1. Системы самопозиционирования

Для определения (вычисления) координат мобильных терминалов, как физических, так и виртуальных, обратимся к усовершенствованному алгоритму «самопозиционирования», не использующему систему глобального позиционирования, применение которой может быть неоправданно дорого для небольших беспроводных вычислительных сетей. Данный алгоритм является распределенным, использующим для вычисления координат мобиль-

ных терминалов расстояния между точками доступа и мобильными терминалами. Метод основывается на построении набора локальных систем координат для каждого из сегментов сети и исследовании их по отношению к общей системе координат, которая является единой для всей беспроводной ЛВС. Основная роль мобильных агентов в данном алгоритме заключается в определении мобильных терминалов по их идентификаторам и последовательном, повторяющемся во времени вычислении их координат как виртуальных, так и физических, раскрывая, таким образом, топологию всей беспроводной сети. Для вычисления координат предлагается [1] использовать «Метод времени прибытия», который вычисляет данные о расстояниях между мобильными терминалами, расстояния между мобильными терминалами и точками доступа, а также угловые пропорции между узлами сети. Для определения расстояний между мобильными устройствами и точками доступа используется метод, базирующийся на том, что скорость света, а следовательно, и скорость распространения электромагнитных волн в воздушной среде конечна, и разница во времени между временем старта передачи данных из терминала-источника передачи данных и временем получения данных на терминале-приемнике информации хоть и невелика, но существует.

2. Построение локальной системы координат мобильных объектов

Изначально необходимо для каждого сегмента сети (точки доступа и набора мобильных терминалов, находящихся в зоне покрытия этой точки доступа) построить локальную систему координат. Пример такой системы локальных координат представлен на рис. 1.

Точку доступа целесообразно использовать как начало системы координат, координаты которой принимают значение (0, 0). Координаты всех мобильных терминалов, входящих в данный сегмент, вычисляются относительно точки доступа. Более

того, все мобильные терминалы являются «соседями» для точки доступа. Из спецификации вычислительных сетей: два узла сети i и j являются соседями, если передача данных между ними осуществляется напрямую — за один раз.

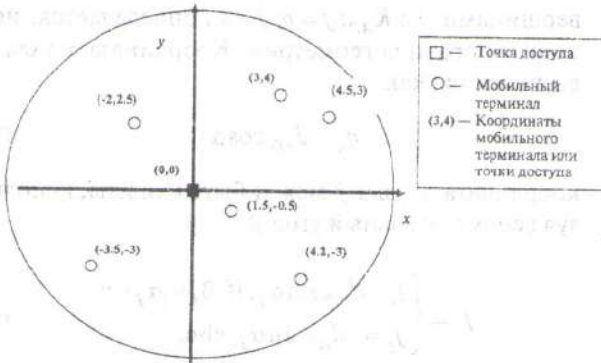


Рис. 1. Локальная система координат для отдельно взятого кластера беспроводной сети

Для проводных сетей при выборе очередного промежуточного сетевого узла во время сеанса передачи данных важную роль играет понятие так называемых терминалов-соседей. Они представляют собой множество S , $S \in N$, то есть множество терминалов, между которыми нет промежуточных терминалов. В подобных сетях, вследствие неподвижности сетевых терминалов, множество S выделить сравнительно просто. Более того, эта процедура проводится один раз, еще на этапах идентификации сетевой топологии.

На рис. 2 представлен пример проводной сетевой топологии, где у терминала T множество соседей состоит из трех терминалов, обозначенных как C .

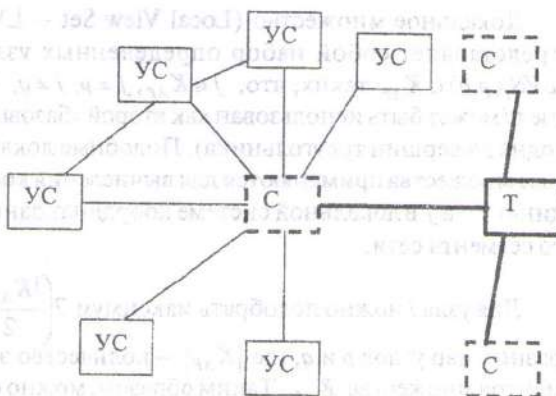


Рис. 2. Пример множества S

Исходя из специфики мобильных беспроводных сетей (мобильность сетевых узлов и отсутствие проводных соединений), необходимо обусловить единое правило формирования множества терминалов соседей S . Теоретически для мобильных сетей мно-

жество терминалов-соседей состоит из всей совокупности сетевых терминалов, с которыми возможна связь, но, исходя из соображений поддержания высокого уровня помехоустойчивости и надежности сетевых соединений, это множество искусственно уменьшают — $S^o \in S$, $S^o \leq S$ для всех терминалов, где $d_{TC} \leq d_{max}$.

На рис. 3 представлен вариант беспроводной сетевой топологии, где множество терминалов-соседей обозначены C , а терминалы, не попавшие в это множество, — UC . Штрихпунктирная линия задает потенциальный радиус, определяющий размер множества соседей, а пунктирная — реальное множество. Для удобства представления множество терминалов-соседей изображается на графе соединенными линиями связи.

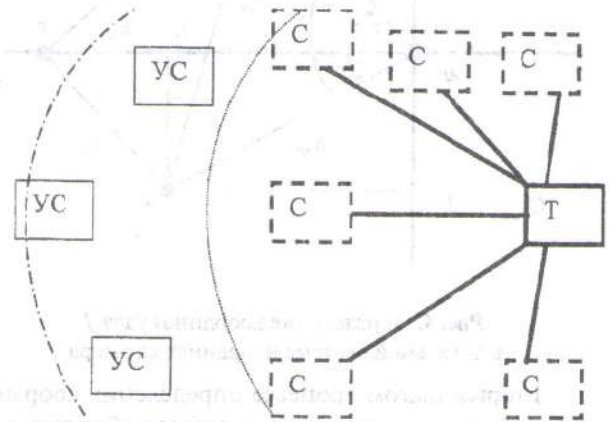


Рис. 3. Пример определения множества терминалов-соседей для беспроводных сетей

Для описания структуры сегмента сети используются следующие данные:

- 1) N — набор всех узлов в сети;
- 2) для любого $j \in N$ существует набор узлов K_{AP} такой, что для каждого $j \in K_{AP}$ j является «соседом» для точки доступа данного кластера;
- 3) K_{AP} — набор всех мобильных терминалов внутри данного кластера беспроводной ЛВС;
- 4) $\forall j \in N$ существует набор узлов D_{AP} , который представляет собой перечень приблизительных расстояний между данной точкой доступа и ее внутренними беспроводными терминалами.

Соседи кластера (группа мобильных терминалов, находящихся в зоне покрытия данной точки доступа) определяются по отклику (маячку). Если в течение определенного количества опросов мобильных терминалов в данном кластере ответа от определенного терминала нет, такой терминал вычеркивается из набора терминалов-соседей текущего сегмента сети. Как уже упоминалось выше, для определения расстояний между объектами внутри сегмента применяются методы времени прибытия.

3. Вычисление координат узла сети

Для вычисления координат узлов сети следует провести следующие процедуры:

- 1) обнаружить все беспроводные терминалы внутри кластера (сформировать множество K_{AP});
- 2) сформировать множество D_{AP} .

После, обладая такими данными, непосредственно вычисляются координаты мобильных объектов внутри каждого сегмента сети – рис. 4.

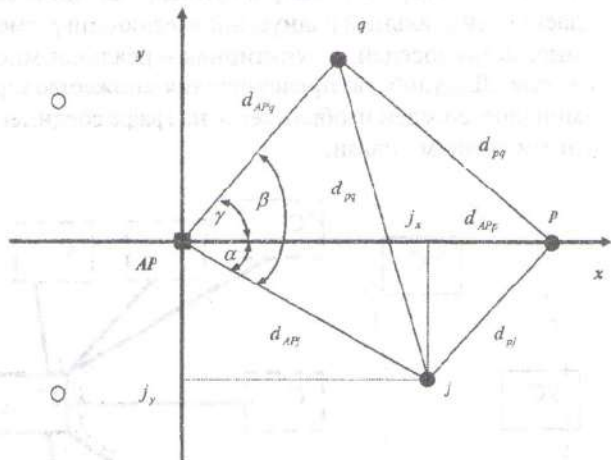


Рис. 4. Определение координат узла j в локальной системе координат кластера

Вторым шагом процесса определения координат мобильных терминалов является выбор двух узлов $p, q \in K_{AP}$ таких, что:

- 1) расстояние между p, q и AP определено;
- 2) расстояния между p, q и AP не равны нулю ($d_{pq} > 0; d_{APq} > 0; d_{APp} > 0; p \neq q$);
- 3) AP, p и q не располагаются на одной линии;
- 4) локальную систему координат для каждого определенного кластера можно безболезненно вращать;
- 5) узел q должен иметь положительное значение по оси y ($q_y > 0$).

После выполнения вышеуказанных ограничений можно определить локальную систему координат [2] для текущего сегмента сети, которая будет задана уникально функциями отношений между AP, p и q . Координаты центра системы координат, а также вершин p и q определены как:

$$AP_x = 0; AP_y = 0; \quad (1)$$

$$p_x = d_{APp}; p_y = 0; \quad (2)$$

$$q_x = d_{APq} \cos \gamma; q_y = d_{APq} \sin \gamma, \quad (3)$$

где γ – угол $\langle p, AP, q \rangle$ в треугольнике (p, AP, q) (рис. 4) вычисляется, используя правило косинуса:

$$\gamma = \arccos \left[\frac{d_{APq}^2 + d_{APp}^2 - d_{pq}^2}{2d_{APp}d_{APq}} \right]. \quad (4)$$

Локальная система координат, представленная вершинами $j \in K_{AP}; j \neq p, j \neq q$, определяется, используя правила геометрии. Координата x узла j вычисляется как:

$$q_x = d_{APj} \cos \alpha_j; \quad (5)$$

координата y узла j может быть найдена, используя вспомогательный угол β :

$$j_y = \begin{cases} j_y = d_{APj} \sin \alpha_j, & \text{if } \beta_j = |\alpha_j - \gamma| \\ j_y = -d_{APj} \sin \alpha_j, & \text{else,} \end{cases} \quad (6)$$

где α_j – угол $\langle p, AP, j \rangle$ в треугольнике $\Delta(p, AP, j)$, β_j – угол $\langle j, AP, q \rangle$ в треугольнике $\Delta(j, AP, q)$ и γ – угол $\langle p, AP, q \rangle$ в треугольнике $\Delta(p, AP, q)$. Величины углов α_j и β_j определяются, используя правило косинусов:

$$\alpha = \arccos \left[\frac{d_{APj}^2 + d_{APp}^2 - d_{pj}^2}{2d_{APj}d_{APp}} \right], \quad (7)$$

$$\beta = \arccos \left[\frac{d_{APq}^2 + d_{APj}^2 - d_{jq}^2}{2d_{APq}d_{APj}} \right]. \quad (8)$$

Координаты любого узла $k \in K_{AP}, k \neq p, k \neq q$ могут быть вычислены, используя координаты центрального узла сегмента и двух других узлов, которые уже вычислены.

4. Определение локального множества

Локальное множество (Local View Set – LVS) представляет собой набор определенных узлов $LVS_i(p, q) \in K_{AP}$ таких, что, $j \in K_{AP}, j \neq p, j \neq q$, где узел i может быть использован как второй «базовый» (одна из вершин треугольника). Подобные локальные множества применяются для вычисления координат узла j в локальной системе координат данного сегмента сети.

Для узла i можно подобрать максимум $2 \left\lfloor \frac{|K_{AP}|}{2} \right\rfloor$ разных пар узлов p и q , где $|K_{AP}|$ – количество элементов множества K_{AP} . Таким образом, можно определить множество C_i , в которое входят все возможные комбинации p и q , для узла i :

$$C_i = \{p, q \in K_{AP}, \text{ для } p \in K_{AP}\}, \quad (9)$$

$$0 \leq |C_i| \leq 2 \left\lfloor \frac{|K_{AP}|}{2} \right\rfloor. \quad (10)$$

Выбор разных p и q должен максимизировать количество узлов, для которых удобно вычислять координаты, используя данный LVS:

$$p, q = \arg \max_{(p_k, q_k) \in C_i} |LVS_i(p_k, q_k)|. \quad (11)$$

5. Построение глобальной системы координат

После построения во всех кластерах локальных систем координат необходимо построить общую глобальную систему координат, исходя из следующих соображений:

- Различные локальные системы координат могут иметь разные шкалы.
- Различные локальные системы координат, как правило, имеют разную направленность осей абсцисс и ординат.
- Необходимо преобразовывать «виртуальные» координаты в физические координаты, которые привязаны к реальному миру, реальным мерам расстояний и скоростей.

Для достижения единого направления осей координат локальных систем координат необходимо «доворачивать» их по отношению к глобальной системе координат.

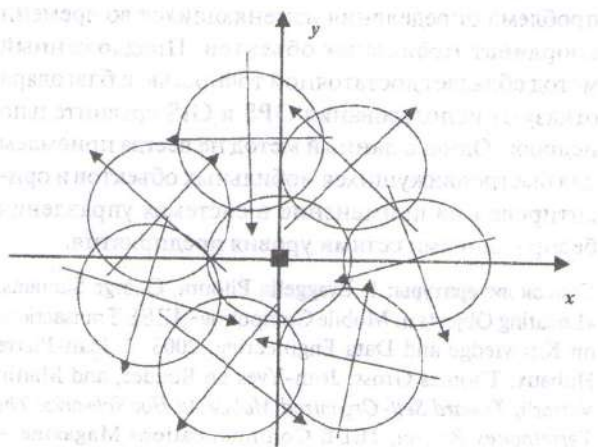


Рис. 5. Определение координат узла j в локальной системе координат кластера

6. Ориентация осей локальных систем координат по отношению к осям глобальной системы координат

Рассмотрим две точки доступа AP: i и k . Для ориентации направлений осей координат AP k относительно локальной системы координат AP i AP k необходимо «довернуть», или сделать зеркальное отражение «mirror» своей системы координат.

Для вычисления угла поворота должны выполняться два следующих условия:

$$i \in LVS_k \text{ и } k \in LVS_i;$$

$$\exists j \neq i, \exists j \neq k, \text{ такие что } j \in LVS_k \text{ и } j \in LVS_i.$$

Возможно возникновение трех ситуаций (рис. 6):

- направления осей координат i и k абсолютно одинаковы («невозможный» сценарий) (рис. 6 а);
- направления осей координат i и k таковы, что необходимо довернуть одну из этих систем координат на определенный угол (рис. 6 б);
- направления осей координат i и k таковы, что целесообразно не просто довернуть одну из этих систем координат на определенный угол, но и сделать зеркальное отражение (рис. 6 в).

В первом сценарии корректировка не нужна.

Угол коррекции для AP k во втором сценарии — $\beta_i - \alpha_k + \pi$.

В третьем сценарии угол коррекции для AP k равен $\beta_i + \alpha_k$, и необходимо выполнить зеркальное отражение для оси y (рис. 6).

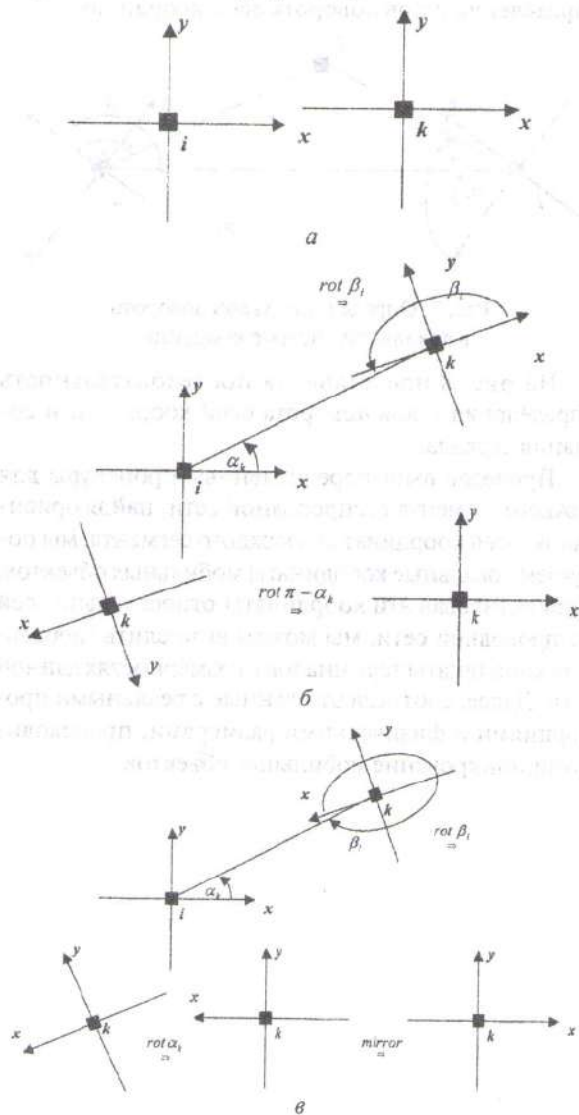


Рис. 6. Возможные сценарии коррекции направления осей локальной системы координат: здесь, α_k — угол вектора ik в системе координат AP i и β_i — угол вектора ki в системе координат узла k

Перед стартом процедуры корректировки для АР k предлагается использовать следующую схему для определения подходящего сценария:

если $\alpha_j - \alpha_k < \pi$ и $\beta_j - \beta_k < \pi$, или $\alpha_j - \alpha_k > \pi$ и $\beta_j - \beta_k > \pi$, то необходимо создать зеркало; угол поправки = $\beta_i + \alpha_k$

если $\alpha_j - \alpha_k < \pi$ и $\beta_j - \beta_k > \pi$, или $\alpha_j - \alpha_k > \pi$ и $\beta_j - \beta_k < \pi$, то зеркало не создается; угол поправки = $\beta_j - \alpha_k + \pi$

На рис. 7 представлена последовательность определения углов поворота осей координат.

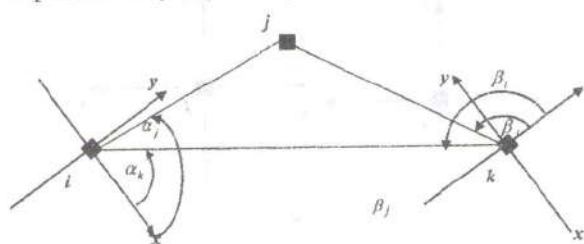


Рис. 7. Определение углов поворота в локальной системе координат

На рис. 8 представлена последовательность определения углов поворота осей координат и создания зеркала.

Пролетав вышеперечисленные процедуры для каждого сегмента беспроводной сети, найдя ориентацию осей координат для каждого сегмента, мы получаем локальные координаты мобильных объектов. Сориентировав эти координаты относительно всей беспроводной сети, мы можем вычислить глобальные координаты терминалов в размерностях данной сети. Далее, соотнеся эти данные с реальными пропорциями и физическими размерами, происходит позиционирование мобильных объектов.

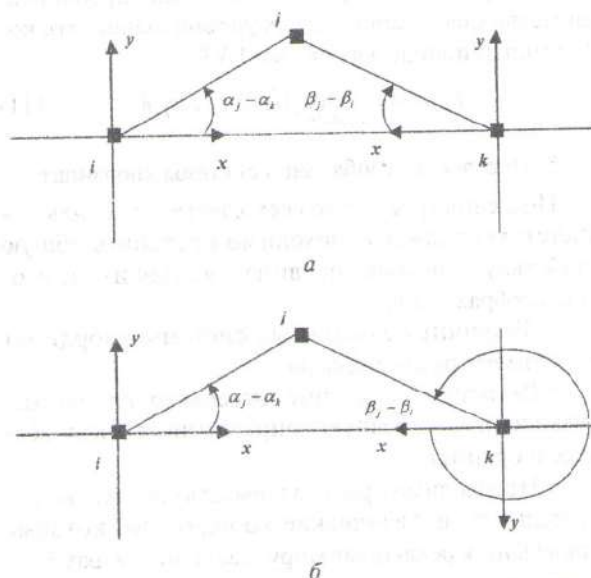


Рис. 8. Поворот и создание зеркала для локальной системы координат сегмента беспроводной сети

Выводы

В предложенной статье рассмотрена актуальная проблема определения изменяющихся во времени координат мобильных объектов. Предложенный метод обладает достаточной точностью и благодаря отказу от использования GPS и GIS сравнительно недорог. Однако данный метод не всегда приемлем для быстро движущихся мобильных объектов и ориентирован на применение в системах управления беспроводными сетями уровня предприятия.

Список литературы: 1. Evaggelia Pitoura, George Samaras, «Locating Objects in Mobile Computing», IEEE Transactions on Knowledge and Data Engineering, 2003. 2. Jean-Pierre Hubaux, Thomas Gross, Jean-Yves Le Boudec, and Martin Vetterli, *Toward Self-Organized Mobile Ad Hoc Networks: The Terminodes Project*, IEEE Communications Magazine • January 2001.

Поступила в редакцию 20.04.07

УДК 519.8

МЕТОД СРАВНЕНИЯ СИТУАЦИЙ ПРИ СИТУАЦИОННО-ПРЕЦЕДЕНТНОМ ПРИНЯТИИ РЕШЕНИЙ

Л. Н. Ребезюк¹, Е. М. Свичкарь²

¹ХНУРЭ, г. Харьков, Украина, rebezyuk@kture.kharkov.ua

²ХНУРЭ, г. Харьков, Украина, eugene19851985@list.ru

Исследуются задачи, составляющие методику ситуационно-прецедентного принятия решений. Представлены варианты решения этих задач. Разрабатывается и исследуется метод количественного сравнения ситуаций, имеющих сложную многоуровневую структуру.

ПРИНЯТИЕ РЕШЕНИЙ, СИТУАЦИЯ, ПРЕЦЕДЕНТ, МЕРА ПОДОБИЯ, ЯЗЫК ОПИСАНИЯ СИТУАЦИЙ.

Введение

Задачи принятия решений возникают и решаются практически во всех сферах человеческой деятельности. Данные задачи имеют следующие характерные черты: наличие единственной конечной цели, существование множества путей достижения цели, необходимость формирования меры, позволяющей сравнивать решения между собой, необходимость осуществления выбора из возможного множества экстремального (т.е. наилучшего по качеству) единственного решения [1].

Применение классических подходов к решению задач принятия решений [2] часто связано со значительными затратами времени и денежных средств, а также они чаще всего не учитывают опыт принятия решений в той или иной предметной области в прошлом. В связи с этим перспективным направлением является применение ситуационно-прецедентного подхода при принятии решений.

1. Основные принципы ситуационно-прецедентного метода

Ситуационно-прецедентный метод подразумевает применение накопленного опыта к выработке управленческих решений. Он базируется на двух основных принципах:

– представление проблемы, требующей решения, в виде ситуации (которая описывает состояние

системы во время постановки цели и которая строится из понятий);

– применение прецедентов для выработки решений.

Представление проблемы в виде ситуации требует разработки языковых средств описания ситуаций. Языковые средства описания ситуаций можно разделить на два типа: модель (язык) представления ситуаций и язык описания ситуаций. Язык представления ситуаций представляет собой средство формальной постановки задачи (формального представления ситуации), а язык описания ситуаций позволяет описывать ситуации, возникающие при решении задач принятия решений в форме, удобной для человека.

Механизм поиска решений с использованием ситуационного подхода (распознаватель ситуаций) можно представить схемой, которая показана на рис. 1. Здесь введены следующие аббревиатуры: ЯОС — язык описания ситуаций; БЗ — база знаний (хранит множество прецедентов, смысл которых представлен ниже); ЛПР — лицо, принимающее решение.

Второй принцип, на котором построен ситуационно-прецедентный метод, — это принятие решений (рассуждение) на основе прецедентов (СВР — Case-Based Reasoning) [3].

Прецедент определяется как случай, имевший место ранее и служащий примером или оправдани-

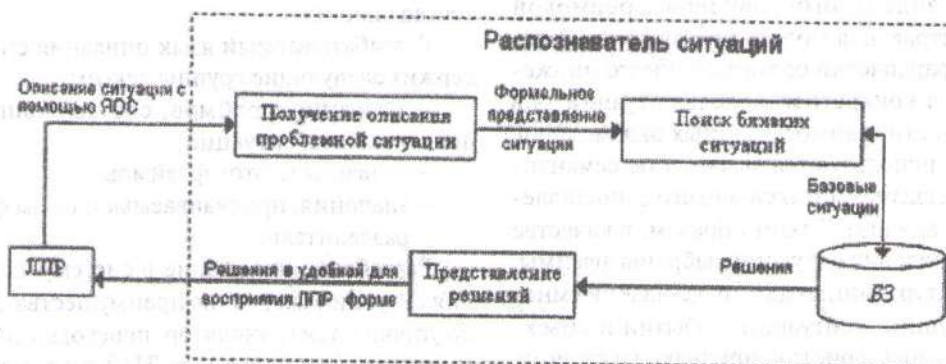


Рис. 1. Механизм поиска проектных решений с применением ситуационного подхода

ем для последующих случаев подобного рода. В рамках ситуационно-прецедентного метода прецедент представляется кортежем <Ситуация, Решения>, где Ситуация — формальное описание базовой ситуации с помощью языка представления ситуаций, Решения — множество решений, имеющих когда-либо место для данной ситуации. Структура элементов множества решений зависит от предметной области, в которой применяется ситуационно-прецедентный метод. Таким образом, прецедент представляет собой информационный блок, включающий в себя базовую ситуацию и соответствующие ей решения; в процессе профессиональной деятельности в некоторой области формируются проблемно ориентированные прецеденты, которые накапливаются в хранилище, в качестве которого могут выступать традиционные СУБД, специализированные серверы знаний, многомерные базы данных и т.д.

Для сравнения ситуаций между собой разрабатывается мера подобия. Мера подобия ситуаций применяется на этапе сравнения текущей ситуации с базовыми ситуациями из БЗ и представляет собой количественную меру, выражающую степень сходства ситуаций.

Особенности реализации, каждой из процедур, составляющих методику применения ситуационно-прецедентного метода, представлены ниже.

2. Языковые средства описания ситуаций

К языку (модели) представления ситуаций предъявляются следующие основные требования [4]:

- наглядность и понятность для заказчика (архитектора системы и её потенциального пользователя);

- нормативность и информативность (языковые средства должны диктовать состав описаний, необходимых для анализа и синтеза системы);

- языковые средства должны обладать чертами непроцедурного языка программирования.

Такими свойствами обладают языки представления знаний общего назначения. Они характеризуются следующими особенностями: данные представляются в виде многоуровневой фреймовой структуры, абстрактная модель предметной области — в виде иерархически организованного множества понятий, а конкретная модель ситуации — в виде совокупности взаимосвязанных экземпляров этих понятий; используются механизмы семантического (как развитие синтаксического сопоставления образцов) анализа. Таким образом, в качестве модели представления ситуаций выбраны фреймы. Фрейм — структура данных для представления множества стереотипных ситуаций, событий и объектов, а также их характеристик, признаков и свойств. Эта информация (о характеристиках, признаках и

свойствах) хранится в слотах фрейма [5]. Представление ситуаций с помощью фреймов при решении задач принятия решений оправдано в связи со следующими соображениями: фреймы по сравнению с другими формализмами легко реализуются программно, так как имеют много общего с классами, составляющими основу объектно-ориентированного программирования; фреймы удобны для представления сложных многоуровневых структур, с помощью которых целесообразно представлять ситуации, возникающие при разработке и принятии решений.

В работах, посвященных общим концепциям ситуационного поиска управленческих решений [4, 6], языковые средства описания ситуаций отождествляются с повествовательным текстом, используя который ЛПР (лицо, принимающее решение) может описать (ввести в систему) ситуацию в проблемной области, требующую решения. Но этот текст не может быть хаотичным. Он должен быть построен с учетом следующих составляющих:

- словарь (тезаурус), объединяющий все нормативные списки и правила, определяющие вхождение элементов лексики в текст, их употребление, значение, написание;

- правила объединения лексических элементов (синтаксис) в структуры более высокого уровня (предложение, абзац, статья);

- семантика как аппарат обеспечения контекстно-зависимого общения, обеспечивающего предсказуемость сложной системы [4].

Для перехода от описания ситуации в виде текста к формальному описанию разрабатывается соответствующий транслятор.

В данной работе предложено отказаться от описания ситуации в виде повествовательного текста и разработать для этой цели специальный язык — язык описания ситуаций, учитывающий специфику решаемой с использованием ситуационно-прецедентного метода задачи и специфику предметной области. Следует отметить, что данный язык также строится с использованием перечисленных выше составляющих.

Разрабатываемый язык описания ситуаций содержит следующие группы лексем:

- названия фреймов, составляющих модель представления ситуаций;

- названия слотов фреймов;

- значения, присваиваемые слотам фреймов;

- разделители.

Разработка и введение в систему языка описания ситуаций дает такие преимущества: значительно упрощается транслятор, переводящий описание ситуации, представленное ЛПР, в формальное описание с помощью языка представления ситуаций;

уменьшается количество ошибок, вносимых пользователем системы в описание ситуации, так как описание ситуации с помощью разработанного языка строго регламентировано его синтаксисом.

При этом описание ситуации в некоторой мере «отдаляется» от человека и «приближается» к формальному её представлению. Но удобный и интуитивно понятный синтаксис языка описания ситуаций позволяет свести до минимума данное явление.

3. Применение прецедентов

Рассуждение на основе прецедентов является подходом, позволяющим решить новую (неизвестную) задачу, используя или адаптируя решение уже известной задачи. Как правило, методы рассуждения на основе прецедентов включают в себя четыре основных этапа, образующих так называемый СВР-цикл:

- извлечение наиболее соответствующего (подобного) прецедента (или прецедентов) для сложившейся ситуации из базы прецедентов;
- повторное использование извлеченного прецедента для попытки решения текущей проблемы (задачи);
- пересмотр и адаптация в случае необходимости полученного решения в соответствии с текущей проблемой (задачей);
- сохранение вновь принятого решения как части нового прецедента.

Прецедент является основной составляющей единицей базы знаний проектируемой системы автоматизации принятия решений на основе накопленного опыта. Извлечение наиболее подобного прецедента осуществляется на основе сравнения текущей ситуации (требующей решения) с базовыми ситуациями каждого прецедента БЗ. Извлекается тот прецедент, базовая ситуация которого является наиболее близкой по отношению к текущей.

4. Разработка меры подобия ситуаций

Мера подобия должна удовлетворять следующим требованиям:

- возможность сравнения любых ситуаций, возникающих в данной предметной области;
- обеспечение достаточно адекватной оценки степени схожести ситуаций (для абсолютно одинаковых ситуаций мера подобия должна принимать предельное значение, показывающее, что ситуации являются идентичными);
- учет относительной важности каждой из составляющих ситуации.

В работах, посвященных ситуационно-прецедентному методу [4], была представлена мера подобия, позволяющая сравнивать ситуации, которые представляются набором атрибутов, каждый из ко-

торых принимает количественное значение. Данная мера подобия определяется следующим образом (применяя взвешенную меру близости (1)):

$$d_{pq}^{(w)} = \left(\sum_{i=1}^n w_i^2 \cdot (x_{pi} - x_{qi})^2 \right)^{\frac{1}{2}}, \quad (1)$$

где $d_{pq}^{(w)}$ — взвешенная мера близости; w_i — вес i -го признака (атрибута) ситуации ($w_i \in [0,1]$); x_{pi} — значение i -го признака ситуации p ; x_{qi} — значение i -го признака ситуации q ; n — количество признаков в ситуации.

Учитывая вычисленную взвешенную меру близости, меру подобия можно вычислить, используя выражение:

$$SM_{pq}^{(w)} = \frac{1}{1 + \alpha \cdot d_{pq}^{(w)}}, \quad (2)$$

где $SM_{pq}^{(w)}$ — мера подобия ситуаций; α — ненулевой положительный коэффициент, влияющий на гладкость функции (2).

Представленная выше мера подобия (2) имеет следующие особенности:

- она принимает значения в диапазоне (0, 1], причем значение 1 достигается при условии, что ситуации идентичны: для них $d_{pq}^{(w)} = 0$; чем ближе мера подобия к нулю, тем более различны сравниваемые ситуации (при этом значение 0 никогда не достигается, т.к. $d_{pq}^{(w)}$ всегда имеет конечное положительное значение);

– коэффициент α позволяет изменять гладкость функции (2) (рис. 2).

– значения x_{pi} и x_{qi} должны быть пронормированы (т.е. они должны быть в диапазоне [0, 1]).

Мера подобия (2) проста для вычислений и понимания. Но данная мера не позволяет сравнивать ситуации, имеющие сложную многоуровневую структуру, и, кроме того, не позволяет учесть отсутствие одного или нескольких признаков в сравниваемых ситуациях, а также вес отсутствующих признаков. Отсутствие перечисленных возможностей при применении меры (2) порождает задачу синтеза меры подобия ситуаций другого вида, которая бы предоставляла эти возможности.

Таким образом, помимо уже сформулированных требований к разрабатываемой мере подобия добавляются еще следующие требования:

- возможность сравнения ситуаций, имеющих сложную многоуровневую структуру;
- учет отсутствия в сравниваемых ситуациях некоторых признаков на любом из уровней структуры ситуации;
- учет весов как присутствующих, так и отсутствующих частей ситуаций.

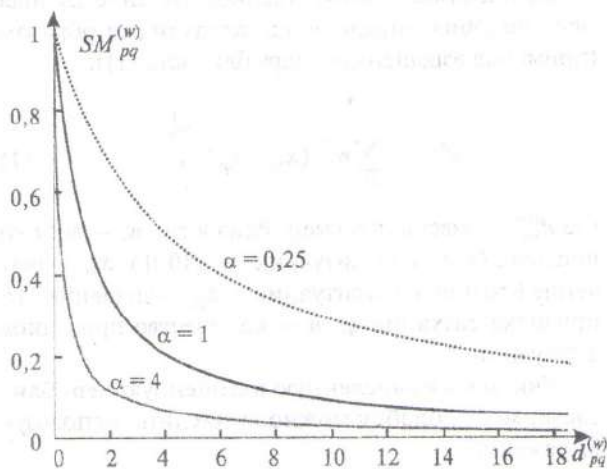


Рис. 2. Вид зависимости меры подобия от взвешенной меры различия при различных значениях коэффициента α

Прежде чем приступить к собственно разработке меры подобия, удовлетворяющей данным требованиям, рассмотрим, что представляет собой ситуация с многоуровневой структурой. Для этого как пример рассмотрим структуру, которая была разработана для представления ситуаций, возникающих при решении задачи выбора компонентов СУА. Структура данной ситуации является четырехуровневой и имеет вид, представленный на рис. 3. Рассмотрим, какие элементы составляют эту структуру.

На первом уровне (всегда) находится один агрегирующий элемент — «Ситуация». Он содержит элементы следующего типа:

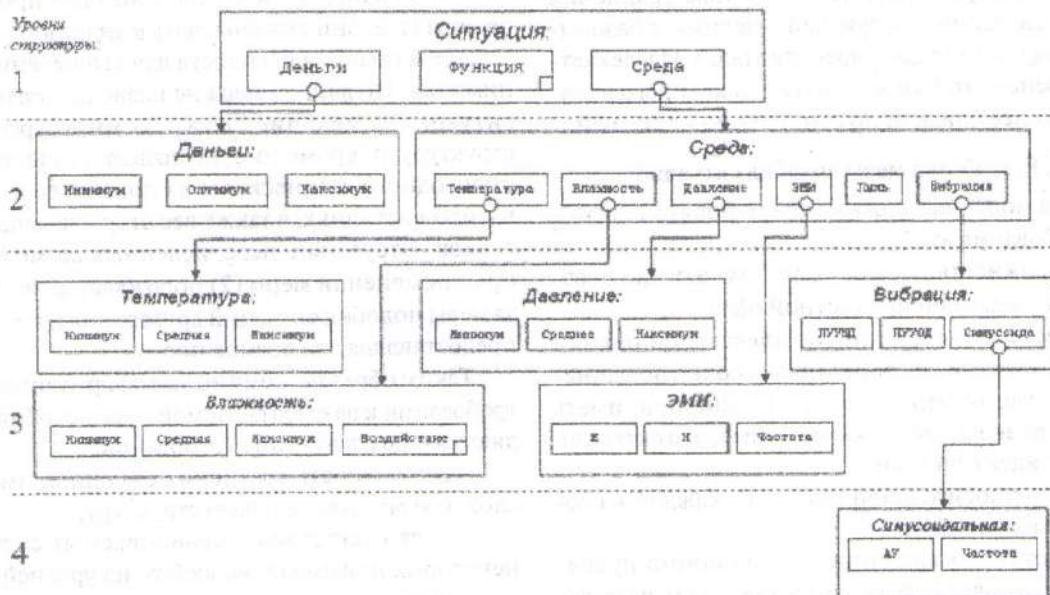


Рис. 3. Структура ситуации: ЭМИ — электромагнитные излучения; ПУУМД — пиковое ударное ускорение (удар многократного действия); ПУУОД — пиковое ударное ускорение (удар одиночного действия); Е — напряженность электрического поля; Н — напряженность магнитного поля; АУ — амплитуда ускорения

— качественный признак перечислимого типа; обозначен прямоугольником с маленьким квадратом в правом нижнем углу и с названием признака внутри прямоугольника;

— признак ссылочного типа; содержит ссылку на агрегирующий элемент нижнего уровня; обозначается прямоугольником с кружком на нижней стороне и названием признака внутри прямоугольника.

Кроме этих двух элементов, ситуации могут иметь еще один, количественный признак — признак, имеющий определенное числовое значение в некотором диапазоне; обозначается прямоугольником, который внутри содержит название элемента.

Каждый агрегирующий элемент может содержать любой из перечисленных выше элементов.

Таким образом, применяя перечисленные выше элементы, можно построить многоуровневую структуру ситуации в любой предметной области. Можно заметить, что получаемая таким образом структура является древовидной.

Следует отметить, что мера подобия (2) позволяет сравнивать ситуации, структура которых состоит из одного уровня, на котором отсутствуют признаки ссылочного типа. Кроме того, ни один из элементов не может отсутствовать.

При разработке меры подобия для сравнения ситуаций, имеющих сложную многоуровневую структуру, возникают следующие проблемы:

— переход от качественных характеристик к их количественным значениям (при этом должно выполняться важное для сравнения ситуаций условие:

чем более близки качественные характеристики, тем меньше разница между числами, к которым они преобразованы);

— формализация сравнения в случае, когда в структуре одной из сравниваемых ситуаций элемент присутствует, а в структуре второй — отсутствует. Решение первой проблемы в значительной степени зависит от предметной области. При разработке меры подобия будем считать, что данная проблема решена. Тогда элемент «качественный признак перечислимого типа» преобразуется в элемент «количественный признак». В таком случае каждый агрегирующий элемент структуры будет содержать два множества признаков: множество количественных признаков и множество признаков ссылочного типа.

Решение второй проблемы представим следующим образом. Для каждого элемента полной структуры ситуации необходимо определить два весовых коэффициента: один применяется при сравнении двух элементов в случае, когда они присутствуют в структуре каждой из сравниваемых ситуаций; второй весовой коэффициент используется в случае, когда в структуре одной из ситуаций сравниваемый элемент отсутствует (при этом в структуре другой ситуации он должен обязательно присутствовать). Если в структурах каждой из сравниваемых ситуаций некоторые элементы отсутствуют, то по этим элементам сравнение осуществляться не будет.

Прежде чем представить методику сравнения ситуаций, имеющих сложную многоуровневую структуру, рассмотрим, какие при этом возможны случаи:

— один и тот же количественный признак присутствует в структуре каждого из сравниваемых агрегирующих элементов (ПК);

— количественный признак присутствует в структуре одного и отсутствует в структуре другого агрегирующего элемента (ОК);

— количественный признак отсутствует в структуре каждого из сравниваемых элементов;

— один и тот же внутренний агрегирующий элемент присутствует в структуре каждого из сравниваемых агрегирующих элементов (ПА);

— внутренний агрегирующий элемент присутствует в структуре одного и отсутствует в структуре другого элемента;

— внутренний агрегирующий элемент отсутствует в структуре каждого из сравниваемых элементов.

Для получения количественной меры подобия агрегирующих элементов (а ситуация — это агрегирующий элемент самого верхнего уровня) предлагается следующий алгоритм.

Шаг 1. Посчитать оценку степени близости сравниваемых элементов. Для этого необходимо посчитать и просуммировать следующие величины:

— степень близости по количественным признакам, присутствующим в каждом из сравниваемых элементов, вычисляемая по формуле:

$$d_{pq}^{\text{ПК}} = \sum_{i=1}^{N_{\text{ПК}}} \alpha_i^{\text{П}} \cdot |x_{pi} - x_{qi}|,$$

где $N_{\text{ПК}}$ — число количественных признаков, присутствующих в каждом из сравниваемых элементов, $\alpha_i^{\text{П}}$ — относительный вес i -го признака для случая, когда этот признак имеется в каждом из сравниваемых элементов, x_{pi} — значение i -го признака в элементе p , x_{qi} — значение i -го признака в элементе q ;

— степень различия по количественным признакам, присутствующим только в одном из агрегирующих элементов, требующих сравнения; определяется выражением:

$$d_{pq}^{\text{ОК}} = \sum_{j=1}^{N_{\text{ОК}}} \alpha_j^{\text{О}} \cdot |x_{bj} - 0.5|,$$

где $N_{\text{ОК}}$ — число количественных признаков, присутствующих только в одном из сравниваемых элементов, $\alpha_j^{\text{О}}$ — относительный вес j -го признака для случая, когда этот признак имеется только в одном из сравниваемых элементов, x_{bj} — значение j -го признака в том элементе, в структуре которого этот признак присутствует. При этом:

$$x_{bj} = \begin{cases} x_{bj}, & \text{если } j\text{-й признак отсутствует в элементе } q, \\ x_{qj}, & \text{в противном случае;} \end{cases}$$

— степень близости по внутренним агрегирующим элементам, присутствующим в каждом из сравниваемых элементов, определяется:

$$d_{pq}^{\text{ПА}} = \sum_{k=1}^{N_{\text{ПА}}} (d'_{pq})_k,$$

где $N_{\text{ПА}}$ — количество внутренних агрегирующих элементов, присутствующих в структуре каждого из сравниваемых элементов, $(d'_{pq})_k$ — степень различия между k -ми агрегирующими элементами сравниваемых элементов p и q , определяющаяся путем рекурсивного применения шага 1 данного алгоритма к k -м внутренним агрегирующим элементам, имеющимся в составе элементов p и q ;

— степень различия по внутренним агрегирующим элементам, которые содержатся только в одном из сравниваемых элементов; вычисляется по формуле:

$$d_{pq}^{\text{ОА}} = \sum_{m=1}^{N_{\text{ОА}}} (d'_p)_m,$$

где $N_{\text{ОА}}$ — количество внутренних агрегирующих элементов, присутствующих в структуре только одного из сравниваемых элементов, $(d'_p)_m$ — оценка

степени различия между m -м внутренним агрегирующим элементом, присутствующим в одном из сравниваемых элементов, и фиктивным внутренним агрегирующим элементом, который имеет структуру, аналогичную присутствующему агрегирующему элементу, и строится следующим образом:

а) определить, какие из агрегирующих элементов присутствуют в существующем агрегирующем элементе;

б) в создаваемый фиктивный элемент добавить элементы, соответствующие выявленным в пункте а), создав их с помощью рекурсивного применения данной процедуры (т. е. алгоритма создания фиктивного элемента) к агрегирующим элементам нижнего уровня;

в) важный момент — в фиктивном элементе ни на одном из уровней его структуры не создаются количественные признаки; это необходимо, чтобы при сравнении фиктивного элемента с существующим применялись весовые коэффициенты, предназначенные специально для сравнения признаков, присутствующих в одном элементе и отсутствующих во втором.

Шаг 2. Используя полученную взвешенную меру близости, вычислить меру подобия, которая описывается в этом случае выражением (2).

Представленная выше мера подобия удовлетворяет всем требованиям, предъявленным к мере подобия ситуаций, имеющих сложную многоуровневую структуру. Однако разработанная мера подобия имеет свои недостатки:

— замена отсутствующего количественного признака значением 0,5 не всегда может быть оправдана (поскольку значение 0,5 может даже не соответствовать среднему арифметическому значению признака по всем ситуациям из базы прецедентов);

— при повышении сложности структуры ситуаций (увеличение количества уровней и количества элементов на каждом из уровней) значительно усложняется вычислительная сложность данной процедуры (за счет рекурсивного применения процедуры сравнения и возникающих при этом рекурсивных процедур построения фиктивных элементов).

Но все же, несмотря на перечисленные недостатки, главным достоинством разработанной меры подобия является то, что она позволяет сравнивать ситуации, представленные сложной многоуровневой структурой, удовлетворяя при этом основным требованиям, которые к ней предъявляются.

Выводы

В результате исследований были выявлены основные проблемы (задачи), составляющие методику реализации ситуационно-прецедентного принятия решений. Суть каждой проблемы (задачи)

подробно изложена. Представлены процедуры их решения.

Особое внимание уделено разработке меры подобия ситуаций, поскольку мера подобия является «узким местом» ситуационно-прецедентного метода. Метод может давать совершенно неудовлетворительные результаты, если мера подобия ситуаций не будет адекватно отображать степень сходства ситуаций и учитывать особенности предметной области. При этом нивелируется положительное воздействие на результат таких факторов, как значительное увеличение количества прецедентов в базе знаний или усовершенствование модели представленных ситуаций.

В результате исследований был предложен метод количественного сравнения ситуаций как мера подобия, использующаяся в практике сравнения ситуаций.

Таким образом, в работе представлены все процедуры, составляющие методику применения ситуационно-прецедентного метода принятия решений. Используя данную методику, разрабатывается программное и информационное обеспечение для реализации рассматриваемого метода. Полученная в результате программной реализации ситуационно-прецедентного метода система позволяет принимать решения в той предметной области, для которой метод адаптирован, на основе накопленного опыта в этой предметной области. При этом система не исключает применения оптимизационных методов — она лишь сохраняет ситуации и принятые для них решения, а также извлекает решения, в наибольшей степени соответствующие текущей ситуации, из базы прецедентов и позволяет на их основе осуществить принятие нового решения.

Система позволяет получать адекватные решения, только когда база знаний будет заполнена значительным количеством прецедентов.

Список литературы: 1. Петров Е.Г., Новожилова М.В., Гребенник І.В. Методи та засоби прийняття рішень у соціально-економічних системах: Навч. посіб./За ред. Е.Г.Петрова. — К.: Техніка, 2004. — 256 с. 2. Воробійов С.А., Мар'їн С.О., Пономаренко О.С. Теорія прийняття рішень. Класичні підходи: Навч. посібник. — Харків: ХТУРЕ, 2000. — 196 с. 3. Варшавский П.Р. Применение метода аналогий в рассуждении на основе прецедентов для интеллектуальных систем поддержки принятия решений. // Девятая Национальная конференция по искусственному интеллекту с международным участием КИИ-2004. — М.: Физматлит, 2004. — Т.1. — С. 218–226. 4. Сорочан М.В., Куземин А.Я. Понятийное представление ситуации при поиске и классификации проектных решений // Прикладная радиоэлектроника. — 2004. — Т. 3. — № 3. — С. 60–66. 5. Мински М. Фреймы для представления знаний. — М.: Энергия, 1979. — 152 с. 6. Поспелов Д.А. Ситуационное управление: теория и практика. — М.: Наука, 1986. — 288 с.

Поступила в редколлегию 20.04.07

УДК 004.891 : 629.735.33.002.72

АВТОМАТИЗАЦИЯ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПРИ ОРГАНИЗАЦИИ АГРЕГАТНОЙ СБОРКИ В АВИАСТРОЕНИИ НА ОСНОВЕ СПЕЦИАЛИЗИРОВАННОЙ ЭКСПЕРТНОЙ СИСТЕМЫ

А. С. Топал¹, В. Е. Зайцев², И. В. Шостак³, С. Б. Белов⁴¹ НАУ им. Н.Е. Жуковского «ХАИ», г. Харьков, Украина, alexa_u@ukr.net² НАУ им. Н.Е. Жуковского «ХАИ», г. Харьков, Украина, vitaliy.zaytsev@khai.edu³ НАУ им. Н.Е. Жуковского «ХАИ», г. Харьков, Украина, iv_shostak@rambler.ru⁴ НАУ им. Н.Е. Жуковского «ХАИ», г. Харьков, Украина, alexa_u@ukr.net

Проведен анализ типовых решений и разработана классификация решений по корректировке директивной технологии сборки планера самолета на авиационном предприятии. Описано построение сетевой модели представления знаний, учитывающей специфику агрегатно-сборочного производства. Выполнен переход от сетевой модели к соответствующей продукционной, основанной на многосортной логике предикатов со временем. Приведен двухэтапный метод формирования решений при организации агрегатно-сборочного производства, включающий постоянный мониторинг текущего состояния производства и непосредственное формирование решений в случае отклонений хода производства от плана.

АГРЕГАТНО-СБОРОЧНОЕ ПРОИЗВОДСТВО, МОНИТОРИНГ, СЕТЕВАЯ МОДЕЛЬ, ДИРЕКТИВНАЯ ТЕХНОЛОГИЯ, МНОГОСОРТНАЯ ЛОГИКА ПРЕДИКАТОВ СО ВРЕМЕНЕМ.

Введение

На текущем этапе развития промышленности предприятия внедряют в структуру управления производством автоматизированные системы класса ERP/MRP, которые решают задачи эффективного управления сбытом, закупками, финансами, бухгалтерией и кадрами. Внедрение таких систем не решает задач управления сложными производственными процессами, для этого необходимо использовать системы класса MES, учитывающие технологическую специфику предприятия и специально разработанные для каждой конкретной области.

Существующие MES-системы ориентированы на интеграцию со SCADA-системами, установленными на уровне производственных процессов. Однако характерные особенности авиастроительного производства — большая доля ручного труда, мелкосерийный выпуск продукции, многотысячная номенклатура деталей, узлов и агрегатов (от 60 тыс. до 250 тыс.), преобладание механосборочных и ручных сборочных операций при производстве изделий, а также отсутствие или частичное применение АСУТП — усложняют внедрение существующих MES-систем на таких предприятиях.

К настоящему времени не разработано MES-систем, учитывающих характерные особенности авиационного производства, в частности, агрегатно-сборочного производства, являющегося одним из самых сложных и ответственных процессов самолетостроения, эффективная организация которого позволяет значительно повысить эффективность предприятия в целом. В агрегатно-сборочном производстве очень затруднено использование SCADA-

систем из-за большого количества сборочных операций, которые выполняются рабочими вручную. Эта характерная для авиастроения черта делает невозможным автоматизацию получения актуальной информации о текущем состоянии производства, что усложняет принятие решений при организации агрегатно-сборочного производства на предприятии.

Процесс организации агрегатно-сборочного производства в самолетостроении характеризуется довольно высоким уровнем неопределенности, которая обуславливается недостаточным объемом информации о текущем состоянии производства, поэтому этот процесс не может быть эффективно реализован на основе традиционных, аналитических методов обработки информации, которые применяются в современных автоматизированных системах.

Использование методов искусственного интеллекта, в частности, технологии экспертных систем, предоставит возможность существенно повысить адекватность производственных решений за счет того, что при формировании решений будет учтено текущее состояние объектов производства. Именно это является основной функцией динамических экспертных систем (ДЭС), которая обеспечит поддержку принятия решений на основе достоверной информации, полученной со всех уровней производственной иерархии.

Современные оболочки создания ДЭС [1], такие как G2, RTworks и др., ориентированы на сбор текущих данных с помощью развернутой на производстве сети SCADA-систем, применение которых

слишком затруднено в агрегатно-сборочном производстве на авиастроительном предприятии в силу его особенностей. Вместе с тем, современные ДЭС, которые базируются преимущественно на продукционных моделях представления знаний, не позволяют адекватно отображать фактор времени, который выступает на передний план при решении задач организации агрегатно-сборочного производства. Представление временных зависимостей позволит четко прогнозировать развитие процессов агрегатно-сборочного производства, что повысит качество принимаемых решений для эффективного устранения последствий отклонений от плана в случае их возникновения.

Таким образом, к настоящему времени не решена задача создания экспертной системы для эффективной поддержки принятия решений при организации агрегатно-сборочного производства на авиастроительном предприятии.

Цель статьи состоит в изложении подхода к синтезу ядра экспертной системы на основе усовершенствованных сетевых моделей знаний для формирования решений при организации агрегатно-сборочного производства на авиастроительном предприятии.

1. Анализ предметной области и постановка задач исследования

Технология сборки и монтажа самолетных конструкций, включая выбор схем сборки, метода сборки, технологической последовательности выполнения операций, состава и количества сборочной и контрольно-испытательной оснастки, оборудования и инструмента, в значительной степени зависит от конструкции собираемых изделий и ее особенностей.

Многообразие практических схем сборки зависит от конструктивно-технологических особенностей изделий, а возможность их использования в конкретных условиях — от степени обеспечения агрегатно-сборочного цеха ресурсами, оснащением, своевременностью поставки агрегатов, узлов, сборок, подборок и деталей из цехов-поставщиков.

Таким образом, стоит задача описать варианты сборочной технологии и процесс формирования решений в зависимости от текущего состояния производственных процессов.

2. Сетевая модель формирования решений при реализации директивной технологии агрегатной сборки

На основании классификатора сборочных единиц, основанного на конструкторско-технологическом членении самолета [2], а также особенностей технологических схем сборки разработана классификация возможных решений (рис. 1), которые

принимаются специалистами агрегатно-сборочного цеха в случае появления коллизий (запаздывание агрегатов, узлов и сборок на критический срок). Эти решения согласованы со стандартами предприятия и спецификацией типа самолета.

В ходе разработки приведенной классификации последовательно применялись следующие признаки:

- 1) характер принимаемых решений;
- 2) участки агрегатно-сборочного цеха, на которых возможно принятие решений по корректировке директивной технологии сборки;
- 3) сборочные единицы, отсутствие которых приводит к необходимости принятия решений при реализации директивной технологии.

На решения технологического характера выпускаются «Технологические указания», которые утверждаются главным технологом завода.

На рис. 1. представлены типовые решения (листья дерева), принимаемые в агрегатно-сборочном цехе в ходе производства в зависимости от текущего его состояния.

Представим модель формирования решений при реализации директивной технологии агрегатно-сборочного производства сетевой моделью следующего вида:

$$AT = \{(P, B, R), T, I(T), O(T), V, Q, M_0\}, \quad (2.1)$$

где $P = \{p_i\}$ — конечное непустое множество позиций сети (их графическое изображение совпадает с изображением позиций в сети Петри [6] — кружки), моделирующее условия реализации или следствия событий; $B = \{b_k\} \subset P$ — множество позиций сети (их графическое изображение также совпадает с изображением позиций в сети Петри — кружки), моделирующее условия (b_k) и интервалы (int_k) их продолжительности. Причем существует биективное отображение $BI: b_k \rightarrow int_k$, задающее продолжительность справедливости условий позиции; $R = \{r_q\}$ — множество решающих позиций, с помощью которых выполняется выбор маркирования позиций после срабатывания инцидентного перехода в зависимости от текущего маркирования. Такие позиции в сети графически изображаются шестиугольниками; $T = \{t_n, (s_j, time_j), f_m\}$ — множество переходов сети, которые моделируют события, происходящие в процессе производства; t_n — переходы, которые происходят без учета временных зависимостей; s_j — переходы, которые характеризуются моментами времени $time_j$, т. е. существует биективное отображение $ST: s_j \rightarrow time_j$, задающее момент времени выполнения перехода; f_m — переходы, которые имеют смежные по входу решающие позиции. Поведение такого перехода (т. е. маркирование смежных по выходу позиций) будет зависеть

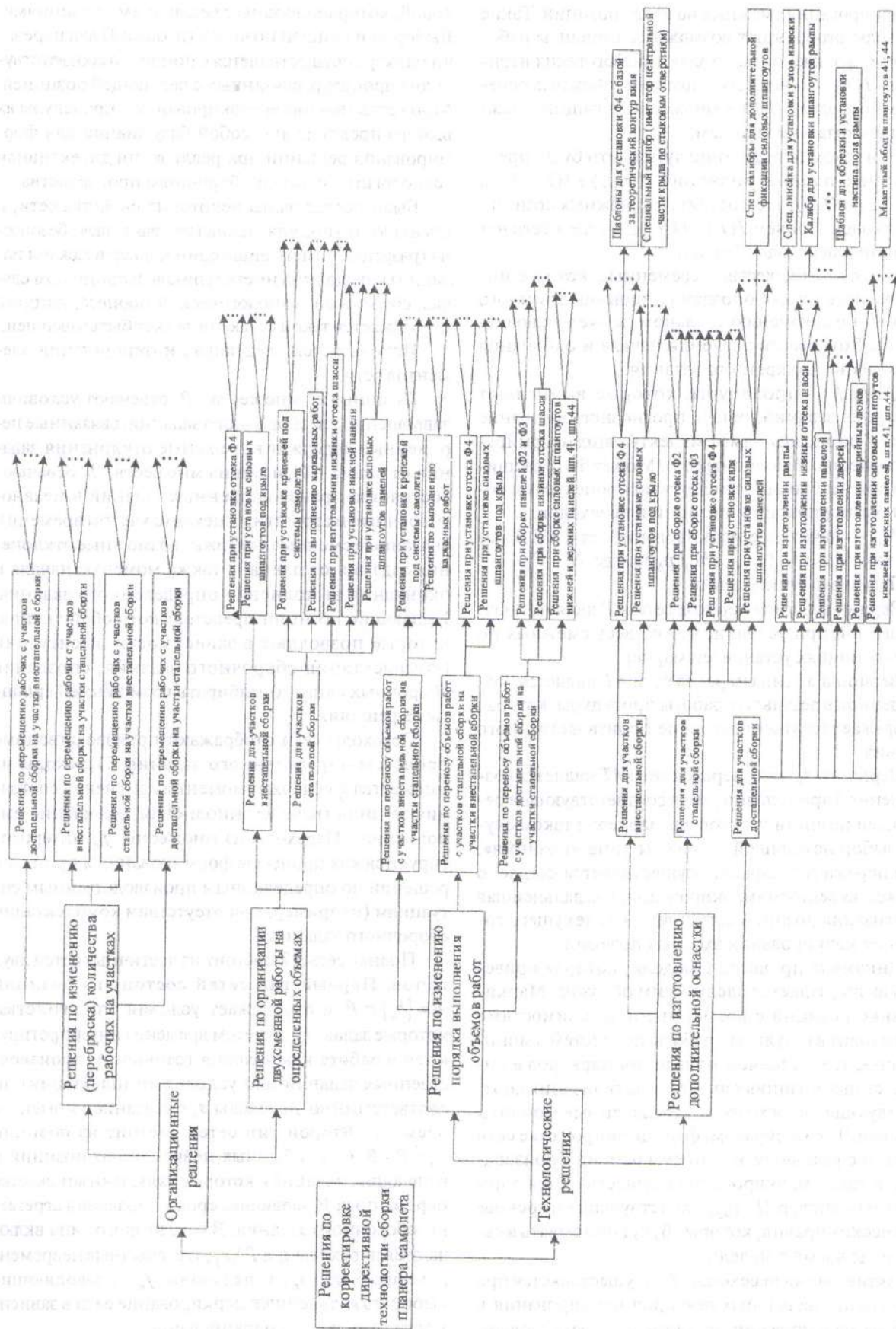


Рис. 1. Классификация решений по корректировке директивной технологии агрегатной сборки в авиастроении

от маркирования смежных по входу позиций. Такие переходы отображают возможные варианты работы сети, а также осуществляют выбор таких вариантов; $I(T)$ — множество позиций, смежных переходам по входам; $O(T)$ — множество позиций, смежных переходам по выходам.

При таком определении дуги в сети будут представлены парами значений $(p_i, t_n) \in I(T) \times T$ и $(t_n, p_i) \in T \times O(T)$, т. е. состоять из смежных позиций и переходов. Причем $I(T) \cup O(T) \subset P$, т. е. в сети нет позиций и переходов без дуг.

$V = \{v_i\}$ — множество переменных, которое интерпретируется как опоздание производственного задания по сравнению с планом, может включать при необходимости значения начала и окончания выполнения конкретного задания.

$Q = \{q_c\}$ — процедуры, которые вычисляют опоздание заданий, общее прогнозируемое опоздание агрегата по опозданиям комплектующих (формируют множество переменных V). Может быть расширено при необходимости другими процедурами согласно интерпретации позиций и переходов.

M_0 — начальное маркирование сети AT . Задается как $M_0 = (M_0(p_1), \dots, M_0(p_N))$, где $N = |P|$.

Дадим ряд определений.

Вершина t_n типа переход сети AT является возбужденной при условии, что во всех смежных по входу позициях установлен маркер.

Вершина s_j типа переход сети AT является возбужденной в результате работы процедуры, которая отображает текущее состояние производственного задания.

Вершина f_m типа переход сети AT является возбужденной при условии, что в соответствующей решающей позиции установлен маркер; в таком случае выбор исходящей вершины типа «позиция», куда передается маркер, осуществляется согласно множеству векторов маркирования, т. е. дальнейшая активизация позиций сети зависит от текущего состояния маркирования входных позиций.

Динамика процесса в модели, которая приведена выше, задается следующим образом. Маркирование позиций с временными зависимостями происходит в результате работы подсистемы мониторинга, т. е. условием появления маркеров в определенных позициях является активизация соответствующих переходов в определенные моменты времени. Таким образом, функционирование сети связано с реальным течением времени производства, и такое маркирование осуществляется с помощью процедур $Q = \{q_c\}$, действующих на основе логических правил, которые будут описывать в явном виде время в модели.

Активизация переходов f_m осуществляется при совпадении начальных векторов маркирования и вектора маркирования из множества векторов-ус-

ловий, которые связаны с решающими позициями. Выбор дальнейшей позиции (позиций) для передачи маркера осуществляется с помощью соответствующих процедур, связанных с решающей позицией. Множество векторов маркирования и процедуры их выбора представляют собой базу знаний для формирования решений при реализации директивной технологии агрегатно-сборочного производства.

Были исследованы некоторые свойства сети, и показано, что представленная сетевая модель безопасна (разрешает иметь лишь один маркер в каждой позиции) и не должна иметь тупиков, в противном случае, сборочный технологический процесс, который отображается такой сетью, не может быть завершен.

Используется следующая интерпретация элементов сети.

Позиции из множества P отвечают условиям завершенности сборочного задания, связанные переменные содержат возможные отклонения задания от плана. Позиции из множества B отвечают состоянию производственных заданий в механо-сборочных и агрегатных цехах (с учетом времени), связанные сменные содержат возможные отклонения задания от плана, а также моменты начала и окончания выполнения определенного задания. Решающие позиции представляют собой знания, которые позволяют в зависимости от ситуации (комплектации сборочного задания, опоздания сборочных единиц) выбирать возможные варианты ее решения.

Переходы сети отображают производственные процессы определенного задания. Переходы из множества s_j содержат моменты изменения состояния задания (начало выполнения, окончание выполнения). Переходы из множества f_m интерпретируются как процессы формирования и принятия решений по определенным производственным ситуациям (например, при отсутствии комплектации сборочного задания).

Полная сеть AT состоит из частичных сетей двух типов. Первый тип сетей состоит из позиций $B = \{b_k\} \subset P$ и отображает условия или свойства, которые задаются с учетом времени (интерпретируются в работе как условия готовности производственных заданий или условия комплектации), и, соответственно, переходы s_j с указанием моментов времени. Второй тип сетей состоит из позиций $p_i \in P \setminus B$ (т. е. обычных невременных позиций и решающих позиций), которые включают множество переменных V , задающее сроки опоздания агрегатно-сборочного задания. В сеть второго типа включаются переходы $t_i \in T \setminus s_j$, т. е. обычные невременные переходы t_n , и переходы f_m , позволяющие выбирать дальнейшее маркирование сети в зависимости от текущего маркирования.

Для построения полной сети, которая будет отображать весь ход производства, задается биективное отображение множества свободных исходящих позиций частичной сети первого типа (тех позиций, которые не являются смежными по входу или не имеют исходящих дуг, т. е. позиций из множества $VO(T) \setminus VI(T)$, $VI(T) \cup VO(T) \subset B$) на свободные входящие позиции сети второго типа (тех позиций, которые не являются смежными по выходу или не имеют входных дуг, т. е. позиций из множества $PI(T) \setminus PO(T)$, $PI(T) \cup PO(T) \subset P \setminus B$).

Смысл такого отображения заключается в том, что им задается передача информации о состоянии выполнения производственных заданий из уровня механических и механосборочных цехов (отображается сетью первого типа) на уровень агрегатно-сборочных цехов (отображается сетью второго типа).

Сети первого типа (названные в исследовании «сетями мониторинга») служат для обеспечения мониторинга текущего состояния производства и получения информации о готовности сборочных единиц с учетом их отклонения от плана. На рис. 2 изображены фрагменты сетей первого и второго типов. Позиции и переходы имеют следующие значения: b_{k1} — задание не выдано, b_{k2} — задание в работе, b_{k3} — задание завершено и сдано, s_{j1} — начало выполнения, s_{j2} — конец выполнения, p_1-p_i — условия наличия сборочных единиц, $p_{i0}-p_{i0}$ — дальнейшие варианты процесса сборки.

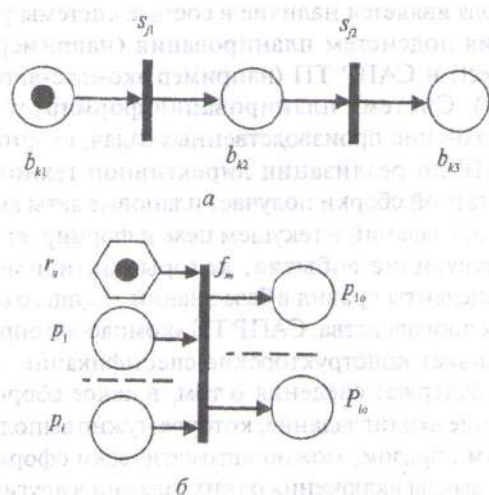


Рис. 2. Фрагменты сети мониторинга (а) и сети формирования решений (б)

Сеть второго типа (названная в исследовании «сетью формирования решений») отображает директивную технологию сборки агрегатов и обеспечивает принятие решений по результатам мониторинга на основе данных о задержках сборочных единиц и знаний о возможных вариантах директивного технологического процесса агрегатно-сборочного про-

изводства. Исходящие позиции сетей мониторинга совпадают с входящими позициями сети формирования решений при организации агрегатной сборки на авиастроительном предприятии (такое совпадение может отображать перемещение готовых деталей между механическими и агрегатными цехами).

Как известно, сеть может быть однозначно определена правилами продукции с использованием многосортного продукционного языка (правила логики предикатов первого порядка). Таким образом, можно записать продукционные правила, которые будут отображать построенную сеть, т. е. построить соответствующую логическую модель.

3. Отражение времени в сетевой модели

Для отображения времени были проанализированы существующие логические модели времени и обоснована необходимость синтеза модели представления каузально-темпоральных зависимостей. Проанализированы подходы к созданию моделей знаний с учетом временных зависимостей и выбран подход, который позволяет синтезировать модель представления каузально-темпоральных зависимостей с учетом специфики предметной области [5].

Для описания процесса агрегатно-сборочного производства задано 3 сорта временных переменных, которые будут введены в исчисления предикатов первого порядка: *Time* — сорт моментов времени, *Interval* — сорт интервалов времени, *Duration* — сорт длительностей, которые формируются с помощью трех непересекающихся множеств символов.

На основании базовых примитивов задаются базовые отношения между ними (отношение упорядоченности моментов времени, отношение соответствия момента началу и концу интервала времени, качественные отношения на моментах времени; качественные отношения на интервалах по Аллену; качественные отношения на моментах и интервалах времени). Для задания атомарных отношений между интервалами времени используется запись следующего вида: $\langle \text{отношение} \rangle (i_1, i_2)$, где $i_1, i_2 \in \text{Interval}$, $\langle \text{отношение} \rangle$ — соответствующее интервальное отношение. Например, $meets(i_1, i_2)$ — совпадение конца одного и начала другого интервала.

Для связывания логических утверждений со временем обосновано использование метода конкретизации [4]. В общем виде правильно построенная формула с помощью метода конкретизации выглядит таким образом:

Истинностный предикат

(логическое утверждение первичного языка, утверждение конкретизации)

Метод конкретизации дает возможность представлять отношения между терминами и их временны-

ми зависимостями с высокой степенью общности, которая позволяет формализовать общие свойства временных утверждений, несовместимые факты, причинно-следственные (каузальные) зависимости, которые являются важнейшими для описания агрегатно-сборочного производства.

Для свойств, справедливых в моменты и на интервалах времени, а также для выделения событий использованы следующие предикаты временных утверждений:

$Holds_on(f, i)$ — это означает, что свойство $f \in Fluent$ имеет место (истинно) на интервале времени $i \in Interval$;

$Occurs(ep)$ — это означает, что событие $e \in Event$ имеет место (истинно) в момент или на интервале времени $p \in Time \cup Interval$.

Используя предикаты временных утверждений, были построены правила для синтеза процедур вычисления отклонений производственных задач от плана. Например, факт опоздания сдачи задания запишется так:

$$\forall S_i \text{ в задание, } \exists i, j, k \in Interval, \exists t \in Time$$

$$\begin{aligned} & Occurs(\text{Плановый_конец_выполнения}(S_i), t) \wedge \\ & Holds_on(\text{status}(S_i, \text{ в работе}), i) \wedge \\ & Holds_on(\text{Timeliness}(S_i, \text{ своевременно}), j) \\ & \wedge \text{end}(j) = t \Rightarrow \\ & Holds_on(\text{Timeliness}(S_i, \text{ опаздывает}), k) \wedge \text{meets}(j, k) \end{aligned}$$

4. Метод формирования решений при реализации директивной технологии агрегатной сборки

Представим двухэтапный метод формирования решений на основе созданной модели представления знаний. Метод состоит из двух этапов: первый основан на логическом выводе на множестве каузально-темпоральных правил и использовании процедур сети мониторинга для выявления отклонений заданий от плана, второй — на логическом выводе на множестве классических производственных правил и использовании процедур сети формирования решений (используя глубинные знания экспертов) для устранения последствий отклонений от плана.

Множество S всех заданий, которые нужно выполнить в течение месяца, состоит из множеств φ_n — заданий, которые не выданы, φ_w — заданий, которые находятся в работе, и φ_r — задания, которые были сданы; таким образом, $S = \{\varphi_n \cup \varphi_w \cup \varphi_r\}$, причем $\varphi_n \cap \varphi_w \cap \varphi_r = \emptyset$. Задания при нормальном ходе производственного процесса переходят из множества φ_n в множество φ_w , далее в множество φ_r . Введем признак Status, который характеризует принадлежность задания S к φ_n , φ_w или φ_r . Однако различные причины вызывают задержки при пере-

ходе заданий из одного множества в другое. Для отслеживания своевременности выдачи и сдачи заданий был введен признак Timeliness, который характеризует принадлежность задания S к $\bar{\omega}$ или $\bar{\omega}$. Таким образом, множество всех заданий состоит из двух множеств $S = \{\bar{\omega} \cup \bar{\omega}\}$, причем $\bar{\omega} \cap \bar{\omega} = \emptyset$, где $\bar{\omega}$ — множество выданных или выполненных заданий без отклонения от плана, а $\bar{\omega}$ — множество заданий, время выполнения или выдачи которых было задержано по некоторым причинам. Задания разделяются на комплекующие и сборочные $S = \{S_k \cup S_a\}$, причем $S_k \cap S_a = \emptyset$, где S_k — задание на изготовление комплекующих, S_a — сборочные задачи.

Существует отображение $S \rightarrow P$, которое задает соответствие заданий позициям сети, причем существует биективное отображение $S_k \rightarrow B$ и $S_a \rightarrow (P \setminus B) \setminus R$.

Введем множество E всех событий, которые изменяют признаки заданий в моменты времени $T = \{T_p \cup T_f\}$, причем $T_p \cap T_f = \emptyset$, где T_p — множество моментов времени, которые уже наблюдались в прошлом, T_f — множество моментов времени, которые состоятся в будущем. Существует отображение некоторых событий во временные переходы сети, т. е. переходы сети могут иметь интерпретацию процесса выполнения задания $E \rightarrow s_j$.

Необходимыми условиями для реализации метода является наличие в составе системы управления подсистем планирования (например, MS Project) и САПР ТП (например «компас-автопроект»). Система планирования формирует план выполнения производственных задач, из которого СППР по реализации директивной технологии агрегатной сборки получает плановые даты выдачи и сдачи заданий в текущем цехе и формирует соответствующие события, которые активизируют antecedенты правил в базе знаний текущего состояния производства. САПР ТП «компас-автопроект» сохраняет конструкторские спецификации, которые содержат сведения о том, в какое сборочное задание входит задание, которое нужно выполнить, таким образом, можно автоматически сформировать факты включения одних заданий в другие.

Исходя из вышесказанного, двухэтапный метод вывода на знаниях при принятии решений по организации агрегатно-сборочного производства может быть представлен следующей последовательностью шагов:

Начальная ситуация:

$$\forall S_i \in \bar{\omega} \wedge \forall S_i \in \varphi_n, \exists S_p : S_k \rightarrow S_a.$$

1. В случае возникновения событий $E_j^f \in E$ фактических изменений состояния задания, где

$j \in \{\text{нач}, \text{оконч}\}$, *нач* — начало выполнения задания, *оконч* — окончание выполнения задания.

$$\forall S_i | S_i \in \bar{\omega} \wedge S_i \in \varphi_n$$

- событие $E_{\text{нач}S_i}^{\phi}$ — (пере)формирование

$$\varphi_w \setminus \setminus \varphi_n = \{S_i | S_i \in \varphi_w \wedge S_i \notin \varphi_n\};$$

- событие $E_{\text{оконч}S_i}^{\phi}$ — формирование

$$\varphi_r \setminus \setminus \varphi_w \wedge \varphi_r \setminus \setminus \varphi_n = \{S_i | S_i \in \varphi_r \wedge S_i \notin \varphi_n \wedge S_i \notin \varphi_w\}.$$

2. В случае возникновения событий $E_j^{pl} \in E$ плановых изменений состояния задания, где $j \in \{\text{нач}, \text{оконч}\}$, *нач* — начало выполнения задания, *оконч* — окончание выполнения задания.

- $\forall S_i | S_i \in \bar{\omega} \wedge S_i \in \varphi_n \wedge t_{\text{нач}S_i}^{\phi} \in T_f \wedge t_{\text{нач}S_i}^{pl} \in T_p$ событие $E_{\text{нач}S_i}^{pl}$ — переформирование

$$\bar{\omega} \setminus \setminus \bar{\omega} = \{S_i | S_i \in \bar{\omega} \wedge S_i \notin \bar{\omega}\}.$$

- $\forall S_i | S_i \in \bar{\omega} \wedge \forall S_i \in \varphi_w \wedge t_{\text{оконч}S_i}^{\phi} \in T_f \wedge t_{\text{оконч}S_i}^{pl} \in T_p$ событие $E_{\text{оконч}S_i}^{pl}$ — переформирование

$$\bar{\omega} \setminus \setminus \bar{\omega} = \{S_i | S_i \in \bar{\omega} \wedge S_i \notin \bar{\omega}\};$$

- $\forall S_i | S_i \in \bar{\omega} \wedge S_i \in \varphi_r \wedge t_{\text{оконч}S_i}^{\phi} \leq t_{\text{оконч}S_i}^{pl}$ событие $E_{\text{оконч}S_i}^{pl}$ — переформирование

$$\bar{\omega} \setminus \setminus \bar{\omega} = \{S_i | S_i \in \bar{\omega} \wedge S_i \notin \bar{\omega}\}.$$

3. В случае возникновения событий $E_k^{ch} \in E$ проверки выполнения задания, где $k \in \{\varphi, \omega, b\}$, φ — проверка состояния задания, ω — проверка своевременности выполнения задания, b — проверка опоздания выдачи задания.

- $\forall S_i | S_i \in \varphi_w \wedge t_{\text{оконч}S_i}^{\phi} \in T_f \wedge t_{\text{пров}S_i}$ событие E_{φ}^{ch} — вычисление продолжительности выполнения задания $dur_{S_i}^w = t_{\text{пров}S_i} - t_{\text{нач}S_i}^{\phi}$

- $\forall S_i | S_i \in \varphi_w \wedge S_i \in \bar{\omega} \wedge t_{\text{пров}S_i}$ событие E_{ω}^{ch} — вычисление продолжительности опоздания сдачи задания $dur_{S_i}^r = t_{\text{пров}S_i} - t_{\text{оконч}S_i}^{pl}$

- $\forall S_i | S_i \in \varphi_r \wedge S_i \in \bar{\omega} \wedge t_{\text{оконч}S_i}^{pl} \leq t_{\text{оконч}S_i}^{\phi}$ событие E_b^{ch} — переформирование

$$\bar{\omega} \setminus \setminus \bar{\omega} = \{S_i | S_i \in \bar{\omega} \wedge S_i \notin \bar{\omega}\}.$$

4. Для каждого сборочного задания $S_i \in S_a$, которое выполняется в цехе, осуществление проверки опозданий Z_{S_i} — входящих в него заданий, и определение такого опоздания dur_{max} , что $Z_m = \max(dur_{S_i}^r)$ и утверждение, что dur_{max} и будет опозданием всего задания.

5. Определение цеха-потребителя для задания $S_i \in S$ путем анализа маршрутного технологического процесса и передача опоздания dur_{max} (выполняется процедура из множества Q сети мониторинга) задания S_i в цех-потребитель (который не является текущим цехом). На этом шаге осуществляется связь сети мониторинга и сети принятия решений (передача информации об отклонении от плана).

6. Установление наличия критической ситуации. Процедура в решающей позиции из множества B путем сравнения продолжительности опоздания dur_{max} задания $S_i \in S$ с критическим опозданием проверяет совпадение вектора маркирования и связанных параметров со списком векторов, которые описывают определенные ситуации.

7. В случае выявления коллизии — осуществление вывода по правилам сети формирования решений. Эти правила определяют часть базы знаний, которая соответствует одному из этапов агрегатно-сборочного производства (доставляемая, стапельная, внестапельная сборка). Осуществляется выбор исходящей позиции $I(O)$, которая получит маркер и определит дальнейшие шаги процесса сборки, т. е. выход из критической ситуации. Знания в базе основаны на знаниях экспертов по организации агрегатно-сборочного производства и формированию технологических указаний диспетчеру агрегатно-сборочного цеха.

Первый этап приведенного метода (шаги 1–4) осуществляется параллельно, асинхронно и непрерывно на частичных сетях мониторинга в разных звеньях производства, что дает возможность в процессе поддержки принятия решений учитывать текущее состояние сборочных единиц, которые изготавливаются во всех подразделениях авиационного предприятия.

Второй этап метода (шаги 5–7) позволяет сформировать решения по конкретной производственной ситуации, которые представляют собой конечное множество и могут быть описаны в виде условий и соответствующих вариантов решения.

Выводы

В статье было описано построение сетевой модели представления знаний, учитывающей специфику агрегатно-сборочного производства на авиационном предприятии. Проведен анализ типовых решений, которые могут быть приняты для корректирования директивной технологии агрегатной сборки в авиационном производстве в случае возникновения отклонений от плана, которые не позволяют придерживаться принятого на предприятии варианта сборки. Выделены критерии и разработана классификация решений по корректировке директивной технологии сборки планера самолета на авиационном предприятии.

Выделены и описаны позиции и переходы в сети, обоснованы необходимые виды позиций и переходов для отображения процессов производства с учетом времени, а также особые позиции и переходы для формирования решений на основе информации об опоздании сборочных единиц. Был осуществлен переход от сетевой модели к соответствующей предикатной, а также обоснован выбор многосортной логики предикатов для расширения временем. Были введены дополнительные сорта для моментов и интервалов времени, а также сорта для временных утверждений — свойства и события. Эта логическая модель встраивается в производственную модель представления знаний, которыми естественно представляется сетевая модель.

Таким образом, путем модификации этапа представления знаний стало возможным учитывать временные зависимости между событиями в ходе производственного процесса, что позволит вычислить отклонения от плана изготовления узлов и агрегатов, поступающих в агрегатно-сборочный цех. Так-

же приведен двухэтапный метод формирования решений по организации агрегатно-сборочного производства, включающий постоянный мониторинг текущего состояния производства и непосредственное формирование решений в случае отклонений хода производства от плана.

Список литературы: 1. Попов Э.В. Динамические интеллектуальные системы в управлении и моделировании. — М.: МИФИ, 1996. — 360 с. 2. Бабушкин А.И. Методы сборки самолетных конструкций. — М.: Машиностроение, 1985. — 248 с. 3. Vila L. A survey on temporal reasoning in artificial intelligence // AI Communications. — 1994. — Vol. 7, no. 1. — Pp. 4-28. 4. Allen J. F. Towards a general theory of action and time // Artificial Intelligence. 1984. — Vol. 23. — Pp. 123-154. 5. Еремеев А.П., Троицкий В.В. Представление временных и причинно-следственных зависимостей в системах поддержки принятия решений реального времени семиотического типа // Перспективные информационные технологии и интеллектуальные системы. — 2000. — №3. — С. 25-31. 6. Питерсон Дж. Теория сетей Петри и моделирование систем: Пер. с англ. — М.: Мир, 1984. — 264 с.

Поступила в редколлегию 15.05.07

УДК 681.3.07

НАХОЖДЕНИЕ И НОРМАЛИЗАЦИЯ ОБЪЕКТОВ НА ИЗОБРАЖЕНИЯХ, ИСКАЖЕННЫХ ПРОЕКТИВНЫМИ ПРЕОБРАЗОВАНИЯМИ НА ПРИМЕРЕ НОМЕРНОГО ЗНАКА

В. А. Любченко¹, Е. О. Передрий²¹ХНУРЭ, г. Харьков, Украина, lubchenko_v@mail.ru²ХНУРЭ, г. Харьков, Украина, elena_peredriy@ukr.net

В задаче автоматической идентификации автомобиля наиболее сложным считается этап выделения на изображении области, содержащей номерной знак при различных яркостных и геометрических искажениях. В данной статье описан подход к выделению номерного знака автомобиля на растровых изображениях и приведены результаты работы компьютерной модели, разработанной для тестирования предложенного метода. Он включает такие этапы, как контрастирование и эквализация гистограммы, фильтры для подавления шумов, бинаризация изображений, выделение контуров областей, аппроксимация выделенных областей четырехугольниками. Для коррекции геометрических искажений номерной пластины на заключительном этапе применяется нормализация проективных преобразований по четырем точкам.

ОБРАБОТКА ИЗОБРАЖЕНИЙ, НАХОЖДЕНИЕ НОМЕРНОГО ЗНАКА, ПРОЕКТИВНЫЕ ПРЕОБРАЗОВАНИЯ, НОРМАЛИЗАЦИЯ, КОНТУР, АППРОКСИМАЦИЯ.

Введение

В последнее время в связи со стремительным развитием вычислительной техники все больший интерес стал проявляться к задачам обработки и распознавания изображений. При решении данных задач часто появляется необходимость разбить исходное изображение на некоторое множество связанных областей, пиксели которых близки по некоторым признакам (однородности цветов, соседства, морфологическим и т. д.), с последующим выделением областей, отвечающих заданному набору признаков. Например, в задаче идентификации транспортного средства при выделении номерного знака необходимо на изображении выделить все подозрительные четырехугольные области с геометрическими характеристиками, подобными номерному знаку автомобиля. Но при решении реальных задач на изображениях часто присутствуют геометрические искажения [1], что усложняет поиск необходимой области или объекта. Поскольку проективная модель наиболее полно описывает восприятие зрительных образов человеком, то целесообразно разрабатывать метод, инвариантный к проективным преобразованиям.

Задача идентификации автомобиля стала интересна с началом автоматизации систем технического зрения. Ее практическое решение имеет широкое применение, например, для решения задач контроля движения транспортных средств на пропускных пунктах предприятий, охраняемых территориях, стоянках, паркингах и т. п. Несмотря на широкое использование и эксплуатацию подобных систем в других странах, они остаются несовершенными, постоянно дорабатываются и изменяются, поэтому данная задача остается актуальной на сегодняшний день как в Украине, так и за рубежом.

В области выделения и распознавания номерного знака автомобилей на сегодняшний день проведено много исследований. Например, для выделения областей символов могут использоваться такие методы, как нейронные сети и сравнение по шаблону для распознавания [3], нечеткая логика и нейронные сети [5], преобразование Хафа [6], сегментация цветных изображений [4] и т. п.

Разработанный метод выделения области номерного знака автомобиля на изображениях, искаженных проективными преобразованиями, может использоваться в других задачах для обнаружения на изображении объектов заданной геометрической формы (например, деталей на конвейере, контроля качества печатных плат и др.).

1. Описание разработанного метода

Изображение, полученное с цифровой камеры высокого разрешения, преобразуется в 8-битное (в оттенках серого) и передается системе выделения номерного знака.

Разработанный алгоритм выделения номерной пластины можно представить в виде схемы, изображенной на рис. 1.

Рассмотрим подробнее каждый из этапов алгоритма.

2. Предобработка

На данном этапе входное изображение преобразовывается для устранения возможных шумов, связанных с выделением и передачей изображения (например, используя медианную фильтрацию или низкочастотные фильтры), влияний, связанных с плохими погодными условиями и освещенностью, улучшения визуального восприятия (эквализация гистограммы) и повышения контраста между сим-



Рис. 1. Схема алгоритма

волами и фоном номерной пластины (используя высокочастотные фильтры).

3. Бинаризация

Поскольку изображения содержат большой объем информации, важную роль начинают играть вопросы ее представления. Черно-белые (бинарные) изображения легче получать, хранить и обрабатывать, чем изображения, в которых имеется много уровней яркости. Пороговую бинаризацию удобно использовать, это достаточно простой и быстрый алгоритм, к тому же для бинарных изображений разработан целый ряд методов их анализа и обработки.

Изображение, получаемое в результате порогового преобразования, определяется следующим образом [7]:

$$B_2(x, y) = \begin{cases} 1, & \text{если } B_1(x, y) \geq T_0; \\ 0, & \text{если } B_1(x, y) < T_0, \end{cases}$$

где B_1 — исходное полутоновое изображение; B_2 — бинарное изображение.

Таким образом, пиксели, которым присвоено значение 1, соответствуют объекту, а пиксели со значением 0 соответствуют фону (или наоборот) (рис. 2б). При выполнении пороговой обработки основная проблема состоит в выборе порога T_0 . В нашем случае порог подбирается экспериментально, основываясь на том факте, что цвет фона номерного знака известен.

После бинаризации изображения на данном этапе проводится «фильтрация» областей, т. е. исключение тех областей, которые не отвечают установленным геометрическим характеристикам (на данном этапе — это площадь областей). Области, имеющие слишком малую или слишком большую



a



б

Рис. 2. Бинаризация изображения: а — входное изображение; б — преобразованное изображение

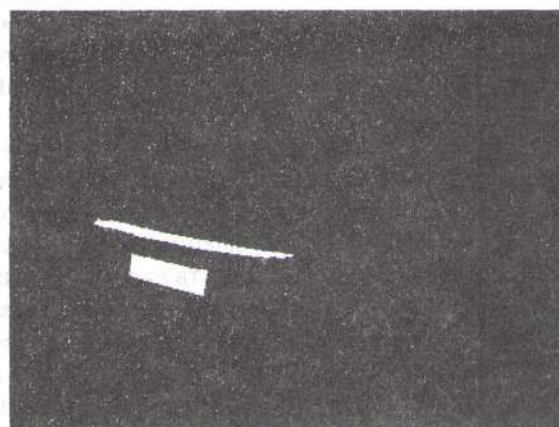


Рис. 3. «Фильтрация» областей

площадь для номерной пластины, при проверке отбрасываются, а также заливаются все области, находящиеся на границе изображения, т.к. номер не может находиться частично в поле зрения (рис. 3).

При бинаризации на изображении объекты могут соприкасаться.

Для разделения объектов можно использовать морфологические операции. Сначала изображение подвергают эрозии до тех пор, пока не исчезнет «слипание» объектов. При этом объекты не должны в результате эрозии совсем исчезнуть. Затем производится наращивание объектов.

4. Выделение границ областей и аппроксимация их прямыми

Алгоритмы выделения контуров можно условно разбить на две группы: отслеживающие и сканирующие.

Отслеживающие алгоритмы основаны на том, что на изображении отыскивается объект (первая встретившаяся точка объекта) и контур объекта отслеживается и векторизуется. Достоинством данных алгоритмов является их простота, к недостаткам можно отнести их последовательную реализацию и некоторую сложность при поиске и обработке внутренних контуров. Пример отслеживающего алгоритма — «алгоритм жука». «Жук» начинает движение с белой области по направлению к черной. Как только он попадает на черный элемент, он поворачивает налево и переходит к следующему элементу. Если этот элемент белый, то «жук» поворачивается направо, иначе — налево. Процедура повторяется до тех пор, пока «жук» не вернется в исходную точку. Координаты точек перехода с черного на белое и с белого на черное и описывают границу объекта.

Сканирующие алгоритмы основаны на просмотре (сканировании) всего изображения и выделении контурных точек без отслеживания контура объекта.

В данном проекте был реализован отслеживающий алгоритм для получения упорядоченных множеств граничных пикселей каждой области (рис. 5а).

Далее полученный контур области мы аппроксимируем заданной фигурой. Под аппроксимацией будем понимать максимально точное описание объекта прямыми. Критерием качества аппроксимации служит среднеквадратичное отклонение. На каждом шаге проверяется выполнение неравенства

$\sum (x_i - x_i')^2 < N$, где N задается программно на основе экспериментальных исследований и желаемой точности (рис. 4).

На следующем этапе выделяются области, аппроксимированные четырехугольником (рис. 5б). Далее проверяется изменение яркости. Выполняется сканирование вдоль горизонтальной линии и строится яркостная проекция, т.е. отбрасываются все области однородной яркости.

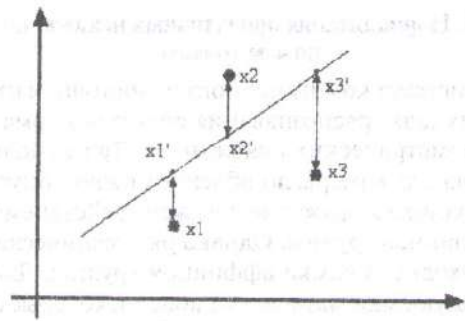
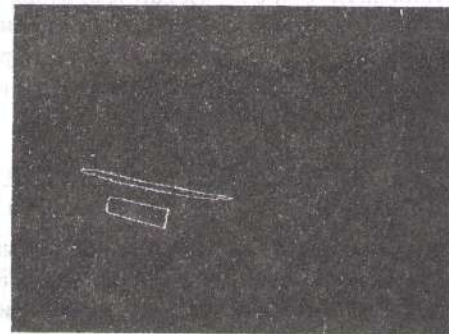
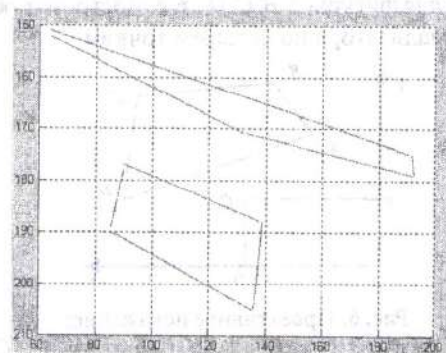


Рис. 4. Расчет расстояния до прямой в методе наименьших квадратов



а



б



в

Рис. 5. Оконтуривание объектов: а — выделение контура; б — аппроксимация четырехугольником; в — выделенная область

5. Нормализация проективных искажений по 4-м точкам

В системах компьютерного зрения одной из важнейших задач распознавания является нормализация геометрических искажений. При небольших расстояниях камеры до объекта можно допустить, что входное изображение искажено действием только аффинной группы. Однако ряд технических задач выходит за рамки аффинной группы. Для них используют математические модели, которые основываются на проективных преобразованиях.

Допустим, что у нас есть два изображения: эталонное и входное, полученное из эталонного путем воздействия на него преобразованием, входящим в проективную группу. Обозначим изображения, соответственно, $B_0(x, y)$ и $B(x, y)$. Таким образом, изображения связаны между собой математической моделью, представленной в виде:

$$B_0(x, y) = B \left(\frac{b_{11}x + b_{12}y + b_{13}}{b_{31}x + b_{32}y + b_{33}}, \frac{b_{21}x + b_{22}y + b_{23}}{b_{31}x + b_{32}y + b_{33}} \right)$$

Предположим, что мы имеем стандартные размеры изображения объекта $ABCD$, фигура $A'B'C'D'$ получена в результате проективного искажения фигуры $ABCD$ (рис. 6). Пусть координаты расположения фигур известны. Необходимо получить нормализатор для фигуры $A'B'C'D'$, т. е. построить общий вид нормализатора по четырем точкам.

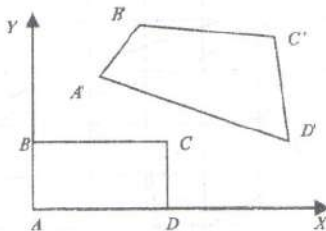


Рис. 6. Проективное искажение: исходная фигура — $ABCD$, искаженная — $A'B'C'D'$

Параметры проективного преобразования в этом случае для фигур $ABCD$ и $A'B'C'D'$ определяются по формулам, описанным в [2].

По определенным параметрам проективного преобразования для фигуры $A'B'C'D'$ с помощью следующих формул можно построить обратное преобразование g^{-1} , с помощью которого привести ее к фигуре $ABCD$:

$$\begin{cases} x = \frac{(b_{13} - x')(y'b_{32} - b_{22}) - (b_{23} - y')(x'b_{32} - b_{12})}{(x'b_{31} - b_{11})(y'b_{32} - b_{22}) - (y'b_{31} - b_{21})(x'b_{32} - b_{12})} \\ y = \frac{(b_{13} - x')(y'b_{31} - b_{21}) - (b_{23} - y')(x'b_{31} - b_{11})}{(x'b_{32} - b_{12})(y'b_{31} - b_{21}) - (y'b_{32} - b_{22})(x'b_{31} - b_{11})} \end{cases}, (1)$$

где x, y — координаты эталона; x', y' — координаты искаженного изображения [2].

Область, описываемая четырьмя точками и предположительно содержащая номерной знак,

может быть представлена фигурой $A'B'C'D'$ и по формуле (1) приведена к стандартному виду — шаблону $ABCD$ (рис. 7), начальные координаты и размеры которого задаются программно. Т. е. область нормализована и максимально подготовлена к этапу выделения и распознавания символов.

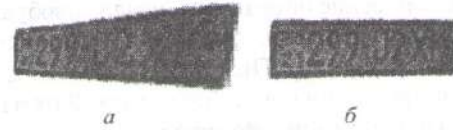


Рис. 7. Нормализация проективных искажений: a — исходная область; b — область, приведенная к стандартному виду

Заключение

Разработан метод нахождения контуров областей, которые потенциально могут содержать номерной знак. Для тестирования разработанного метода построен алгоритм, который был реализован в виде компьютерной модели. Компьютерная модель позволяет осуществлять бинаризацию изображений, выделять контуры областей, осуществлять аппроксимацию выделенных четырехугольниками областей, а также проводить нормализацию проективных искажений по четырем точкам. Эти методы являются составной частью предлагаемого метода.

Тестирование предлагаемого метода выделения номерного знака проводилось на реальных изображениях автомобилей, которые искажены различными проективными преобразованиями.

Научная новизна данной работы состоит в том, что предложенный алгоритм учитывает возможность наличия на изображениях проективных преобразований, в то время как большинство существующих алгоритмов не учитывают такой возможности и для успешной работы системы требуют отсутствия ярко выраженных геометрических искажений номерного знака автомобиля.

Список литературы: 1. Пулятин Е.П., Аверин С.И. Обработка изображений в робототехнике. — М.: Машиностроение, 1990. — 320 с. 2. Генкин В.Л., Ерош И.Л., Москалев Э.С. Системы распознавания автоматизированных производств. — Л.: Машиностроение, 1988. — 246 с. 3. E.R. Lee, P.K. Kim, H.J. Kim Automatic Recognition of a Car License Plate Using Color Image Processing // Proceedings of the International Conference on Image Processing, 1994. 4. H.J. Kim, D.W. Kim, S.K. Kim, J.V. Lee, J.K. Lee Automatic Recognition of Car License Plates Using Color Image Processing // Engineering Design & Automation, 3(2), pp. 215–225, 1997. 5. J.A.G. Nijhuis, M.H. Ter Brugge, K.A. Helmholt, J.P.W. Pluim, L. Spaanenburg, R.S. Venema, M.A. Westenberg Car License Plate Recognition with Neural Networks and Fuzzy Logic // IEEE International Conference on Neural Networks, 1995. 6. H.S. Kim, et al. Recognition of a Car Number Plate by a Neural Network // Proceedings of the Korea Information Science Society Fall Conference, Vol. 18, pp. 259–262, 1991. 7. Гонсалес Р., Вудс Р. Цифровая обработка изображений. — М.: Техносфера, 2005. — 1070 с.

Поступила в редколлегию 16.05.07

УДК 519.62

**СОВРЕМЕННЫЕ ИНФОРМАЦИОННЫЕ РАСПРЕДЕЛЕННЫЕ СИСТЕМЫ
БАЗ ЗНАНИЙ**Т. Б. Шатовская¹, С. П. Менинник²¹ ХНУРЭ, г. Харьков, Украина;² ХНУРЭ, г. Харьков, Украина, stas_sam@rambler.ru;

Проведен разноуровневый анализ информационных систем баз знаний. Проанализированы первоначальные предпосылки и концепции создания с дальнейшим анализом динамики и истории развития до современного состояния. Описание и аналитика взаимного соотношения поисковых систем Интернет на рынке на текущее время.

БАЗА ЗНАНИЙ, ИНТЕРНЕТ, ПОИСК, БИЛЛИНГ.**Введение**

На современном этапе развития нового постиндустриального общества, в условиях постоянно и стремительно изменяющейся внешней среды особое значение приобретает развитие компьютерных технологий, неотъемлемой частью которых являются информационные системы, давно и прочно вошедшие в нашу повседневную жизнь. Трудно представить себе какую-либо сферу, в которую не проникли бы информационные технологии. Они применяются в сфере транспорта, культуры, медицины, образования и, в частности, в библиотечарском деле. Определяющим фактором эффективности систем обработки и взаимодействия необходимой информации является использование информационных баз данных с целью облегчения повседневного труда человека, частичной замены или переквалификации в оператора персонального компьютера, что приведет к минимизации времени, затрачиваемого им для обработки, поиска и сортировки данных.

Исходя из вышеизложенного, тема применения в области документооборота информационных систем является в наше время невероятно актуальной. Действительно, для работы с любой литературой требуется, в первую очередь, чтобы человек обладал отличной памятью для запоминания местонахождения той или иной литературы, фиксации хотя бы относительного содержания для повышения скорости и эффективности рабочего процесса. Понятно, что сохранение на бумажном носителе не является таким надежным, как сохранение данных в информационной системе, а эффективность зависит в основном от скорости предоставления информации. Поэтому очевидна назревшая необходимость в переходе общества к компьютерным информационным системам. Таким образом, в представляемой информационной системе показываются преимущества, связанные с хранением документов (информации), обеспечением возможности легкого быстрого доступа, рассматриваются вопросы эффективности хранения и обработки информации в распределен-

ных системах баз знаний. Также стоит отметить важность в современном мире применения более инновационных технологий и степень обширности поставленной задачи. Ввиду универсальности системы в ней могут и должны использоваться лингвистические технологии, математическая психология, методы информатизации, технологии поиска, на которых отдельно остановимся и рассмотрим на более детальном примере [4].

Цель данной статьи — разъяснить проблематику и методику формирования эффективных информационных систем баз данных как всеобъемлющего механизма с множеством возможностей в современном мире.

Задачей, решаемой в данной статье, является разносторонний анализ современных информационных систем и аналитические предсказания значимости, темпов развития, эффективности их в будущем. Также рассматриваются некоторые концепции защиты информации, поддержка биллинг-системы.

С целью наглядного демонстраирования некоторых аспектов вопроса часто используется унифицированный язык моделирования UML. Он позволяет проектировать варианты использования системы, диаграммы для визуализации функциональных возможностей системы в целом и отдельные модули [5].

1. Анализ предметной области

С развитием науки, компьютерных и информационных технологий возникает необходимость в получении все большего количества информации. В современном мире человеку нужно получать как можно больше информации за короткий промежуток времени, поэтому возникает необходимость постоянного присутствия источников информации у себя «под рукой», а также возможность быстрого поиска информации. Долгие годы развития технологий, не утихающие споры о безопасности для здоровья человека, споры о моральном облике нового мира — вот тот небольшой список проблем из истории, которые предшествовали появлению на свет

современного электронного документооборота. Печатные источники не могут дать таких возможностей, поэтому на смену им пришли электронные издания. Электронные издания могут распространяться посредством электронных носителей информации и компьютерных сетей, количество экземпляров при этом не является ограниченным. Также электронные издания не подвержены повреждениям, поскольку на них не влияет физический фактор, а при необходимости их можно вывести на бумагу. Бесспорным преимуществом электронных изданий является и то, что для их хранения не требуется огромного пространства, так как это происходит в случае традиционных библиотек. Актуальность системы является неоспоримой, поскольку развитие электронных изданий имеет больше преимуществ по сравнению с печатными изданиями, но, к сожалению, не может полностью заменить их.

Данная система призвана помочь многим предприятиям и другим организациям, а также просто отдельным людям в документообороте, получении важной информации. Эффективный поиск реальной информации, хранение большой базы ссылок на внешние и внутренние источники данных — все это дает безусловные преимущества и очевидные выгоды в современном мире электронных технологий.

Существуют системы различного уровня сложности и покрытия, глобальной системой на сегодняшний день является сеть Интернет, о которой и пойдет речь в статье с точки зрения истории, причин возникновения и эффективного использования в качестве универсального обменника полезной информацией. Интернет многообразен и обладает множеством как отрицательных, так положительных свойств, на которых и будем акцентировать внимание.

Основными проблемами данной предметной области являются практическая ненадежность печатной продукции, большой объем для хранения, потребность в посреднике для доступа к книгам. В данном случае возникает необходимость быстрого и эффективного доступа и поиска, возможность сортировки по многим параметрам.

Предметной областью этой статьи являются физическая деятельность, документооборот в какой-либо его форме, распределенные системы баз знаний, а именно: предоставление прямых ссылок на издания или информацию сколь угодно различной тематики, их сортировка и систематизация для более удобного доступа. Анализируя предметную область, оговорим некоторые ограничения и условности модели для рассматриваемой системы. Любое информативное издание, любую информацию будем условно называть книгой — термин, в который

вкладывается смысл любого вида текстового, графического и других видов содержания. Книга может фактически относиться к многим разделам науки, и конкретные направления имеют множество книг по своей тематике [4].

При анализе предметной области были рассмотрены основные проблемы, связанные с особенностями инвентаризации литературы в каталоги разных разделов, а также возможности эффективной автоматизации получения требуемой информации, возможность и разновидности доступа к информации с разных физических носителей и серверов, что привело к потребности в биллинг-системе при доступе к данным непосредственно с самого сервера [7].

Анализ предметной области позволил выделить основные понятия (книга, автор, издательство, раздел) и смысловые связи между ними в виде простейшей концептуальной модели предметной области, приведенной на рис. 1. Под книгой, как было оговорено выше, на данном рисунке стоит понимать любую информацию в электронном виде.

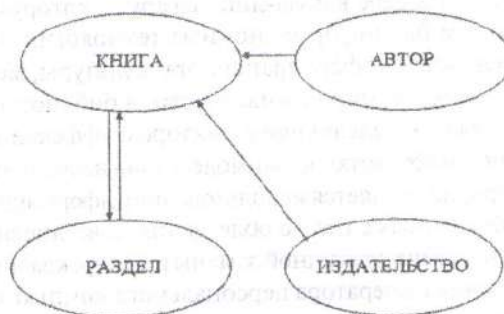


Рис. 1. Простейшая концептуальная модель предметной области

2. Анализ существующей глобальной системы

Для того, чтобы понять темпы развития, а главное, темпы ускорения развития современных баз знаний, приведу некоторые факты из истории создания глобальной информационной распределенной системы Интернет. Данные факты являются общедоступными и исторически логичными в стремлении человека к информатизации общества.

1967 год. Ларри Робертс (Larry Roberts), практик, воплощающий в жизнь теоретические идеи Ликлидера, предлагает связать между собой компьютеры ARPA. Начинается работа над созданием ARPANET.

1969 год. ARPANET заработал. К нему подключаются компьютеры ведущих, в том числе и невоенных, лабораторий и исследовательских центров США.

1971 год. Рэй Томлисон (Ray Tomlison), программист из компьютерной фирмы Bolt Beranek and

Newman, разрабатывает систему электронной почты и предлагает использовать значок @ («собака»).

1974 год. Открыта первая коммерческая версия ARPANET — сеть Telenet.

1976 год. Роберт Меткалф (Robert Metcalfe), сотрудник исследовательской лаборатории компании Xerox, создает Ethernet — первую локальную компьютерную сеть.

1977 год. Число хостов достигло одной сотни.

1980 год. Писатель и политический аналитик Алвин Тоффлер (Alvin Toffler) опубликовал книгу «Третья Волна» (The Third Wave), в которой описал постиндустриальный мир, где «первую скрипку» играют информационные технологии. Тоффлер, в частности, сумел оценить перспективы развития компьютерных сетей и предположил, что однажды такая сеть сможет объединить весь мир, наподобие того, как все обладатели телевизоров могут смотреть одну и ту же передачу. При этом компьютерная сеть, по прогнозу Тоффлера, даст людям несравненно больше возможностей, чем обычное ТВ.

1982 год. Рождение современного Интернета. ARPA создала единый сетевой язык TCP/IP.

1984 год. Число хостов превысило тысячу. И это всего навсего за 15 лет. Сравните темпы роста. За первые 8 лет рост составил 100 хостов, тогда как за последующие 7 лет — уже 900 хостов, что явно указывает на огромную динамику развития отрасли.

1986 год. Национальный Фонд Науки США (The National Science Foundation) создал NSFNET, связавшую центры с «суперкомпьютерами». Эта сеть доступна лишь для зарегистрированных пользователей, в основном, для университетов.

1989 год. Число хостов превысило 10 тысяч. Т. е. за 5 лет рост составил 9000, сохраняя и увеличивая темпы.

1991 год. Европейская физическая лаборатория CERN создала известный всем протокол — www — World Wide Web. Эта разработка была сделана, прежде всего, для обмена информацией среди физиков. Появляются первые компьютерные вирусы, распространяемые через Интернет.

1993 год. Создан первый интернет-браузер Mosaic, созданный Марком Андреесеном (Marc Andreessen) в Университете штата Иллинойс (University of Illinois). Число интернет-хостов превысило 2 млн., в Сети действует 600 сайтов.

1996 год. Началось соревнование между браузерами Netscape, созданным под руководством Марка Андреесона, и Internet Explorer, разработанным компанией Microsoft. В мире существует 12,8 млн. хостов и 500 тыс. сайтов.

1999 год. Впервые предпринята попытка цензуры Интернета (популярен принцип: «Интернет никому не принадлежит»). В ряде стран (Китай, Сау-

довская Аравия, Иран, Египет, страны бывшего СССР) государственными органами предприняты серьезные усилия, чтобы технически блокировать доступ пользователей к определенным серверам и сайтам политического, религиозного или порнографического характера.

2002 год. Сеть Интернет связывает 689 млн. человек и 172 млн. хостов. Разрабатываются новые технологии Интернета, которые должны заменить «старый Интернет», расширить его функции или создать национальные компьютерные сети [1].

Таблица 1

Период, гг.	Количество хостов в конце периода	Темп роста в среднем за год
1969	0	—
1969–1977	100	—
1977–1984	1000	128 %
1984–1989	10000	180 %
1989–1993	2 млн	4500 %
1993–2002	172 млн	860 %

Как можно заметить, где-то в период 1989–1993 гг. в мире разразился информационный бум с соответствующими последствиями для истории и всей современной жизни. В целом сегодня темпы роста немного замедляются ввиду того, что система фактически охватила большинство сфер жизни в развитых странах, и единственным мощным источником развития и роста сети Интернет как мощнейшей информационной распределенной системы баз знаний являются слаборазвитые страны. Также стоит отметить, что на базе Интернет существует возможность создания полнофункционального искусственного интеллекта из-за слабой вычислительной способности одной единицы вычислительной техники, но весьма внушительной мощности миллиардной армии вычислительных машин по всему миру.

Существует огромная проблематика засоренности и высокого процента бесполезности информации из всемирной глобальной сети, изначально созданной для научно-исследовательских потребностей. Анализируя современное состояние сети Интернет, можно с уверенностью сказать, что она перестала быть научным инструментом в той мере, в которой планировалось, и больше похожа на зону коммерческих войн. Также следует развеять миф о том, что можно найти абсолютно все. По-настоящему интересной и нужной информации скорее всего вы не найдете по множеству причин. Одна из них — эффективность поиска, что составляет огромную проблематику в современном мире. Проблема именно в том, чтобы быстро найти требуемые данные, по пути отсеивая всю рекламу, мусор и другие,

возможно, не имеющие фактического отношения к тематике или разделу, но маскирующиеся под них с целью получения более высоких рейтингов. Вся глобальная система состоит из множества электронных документов как статических, так и динамических.

Интернет — это не только собрание технологий, но и собрание сообществ. Успехи Интернет в значительной степени объясняются как способностью удовлетворить основные социальные потребности, так и умением эффективно использовать общественность для развития инфраструктуры.

Дух коллективизма, содружества в Интернет имеет глубокие корни, как было описано выше, он зародился еще в начале работ над ARPANET. И, развившись до наших дней, дух коллективизма сформировал современное информационное сообщество. В связи с чем еще одной особенностью является то, что глобальные информационные системы превращаются в универсальные коммуникаторы современного общества по отношению его к себе самому и новым технологиям [2].

2. Биллинг-системы

Однако сама система Интернет не является самодостаточной и опирается на материальную основу. Фактически вся глобальная сеть — это связанные между собой различные вычислительные устройства, которые могут ломаться, потреблять электроэнергию и другие расходные материалы, что приводит к требованиям финансовой безубыточности. Даже простая передача простого сигнала по кабелю уже заложена в стоимость услуг системы. С большим развитием информационной инфраструктуры данные затраты явно теряют свою актуальность с приносимой пользой и, как результат, чей-то конечный доход. В случае, когда система слабо развита, существует необходимость оплаты расходов на физическое поддержание. Одним из таких механизмов является биллинговая система.

Биллинговая система — программный комплекс, осуществляющий учет объема потребляемых абонентами услуг, расчет и списание денежных средств в соответствии с тарифами компании.

Исходя из задач и запросов бизнеса, можно набросать схему системы. Чтобы не обсуждать какого-то абстрактного сферического коня в вакууме, будем рассматривать типовой пример оператора связи, продающего трафик абонентам.

Услуги могут быть разными (например — VPN-доступ, dial-up пул, обычный неинкапсулированный трафик, Proxy, VoIP etc), надо обеспечить доставку ядру в единообразном виде информации о том, какой тип услуги, какой абонент, в каком объеме и в какое время потребил. В худшем случае для каждого из типов услуг придется разрабатывать

свой коллектор, но, если повезет, что-то удастся унифицировать. Технологии, которые могут помочь при создании коллекторов, — SNMP, Radius, NetFlow.

Принятие решения о блокировке абонента при окончании средств на его счету на практике происходит не мгновенно, этот факт тоже надо учитывать. Например, если блокировка срабатывает раз в минуту — при скорости соединения 1 Мбит/с абонент может скачать лишних 7,5 мегабайт в худшем случае. Данная система учета финансовых средств пока что актуальна для Украины в частности, но медленно сдает свои позиции в пользу безучетных услуг передачи информации за единый установленный платеж, который уже косвенно ложится как бремя на конечного пользователя информационной системы баз данных в виде отдельного платного контента или в иной форме. Также существуют варианты с физической за счет пожертвований и на общественных началах, которые изначально не несут в себе серьезных финансовых ресурсов [6, 7].

3. Представители прогрессивных поисковых систем Интернет

Для начала скажем, что современные поисковые системы базируются на финансовой основе, применяют самые современные методики и технологии. Та же лингвистическая технология давно используется в поисковых машинах. Применение новаторских идей отдельных людей в свое время привело таких гигантов, как Google и другие, к вершине сегодня. Самые прогрессивные являются самыми удачными, успешными и как результат — доходными.

В настоящее время рынок поиска уже достаточно развит, и мы можем смело выделить основных игроков (а их не так уж и много). Львиная доля всего мирового поиска приходится на такие поисковые системы: Google Search, Yahoo! Search и Microsoft Live Search. Эти три американские компании конкурируют между собой в борьбе за главный приз, то есть пользователей, которые пришли за помощью в Интернет.

Чтобы оценить доли присутствия этих и других компаний, было бы неплохо, если бы поисковые системы сами давали подробные отчеты о своей работе в свободный доступ, однако это маловероятно, так как подобная информация является коммерческой тайной.

Но все-таки существуют компании, которые как раз и занимаются сбором статистики по подобным вопросам. Наиболее авторитетной в этом плане является компания NetRatings, Inc. Именно она постоянно информирует Интернет-сообщество о результатах своих исследований в областях интернет-коммерции, маркетинга, онлайн-рекламы и поисковых технологий.

По данным последнего опубликованного отчета об объемах поисковых запросов за декабрь 2006 года, на американском рынке мы можем наблюдать ситуацию, отображенную в табл. 2.

Таблица 2

Google Search	50,80%
Yahoo! Search	23,60%
MSN/Windows Live Search	8,40%
AOL Search	6,10%
My Way Search	2,40%
Ask.com Search	2,10%
EarthLink Search	0,50%
Others	6,10%

Судя по этой статистике, сейчас более 80% американского рынка поиска делят между собой те самые три компании, о которых мы говорили ранее. Оставшиеся 20% распределились между другими менее популярными поисковыми системами и порталами. Некоторые из них используют поисковые технологии первых трех компаний. Даже с учетом того, что, казалось бы, 20% рынка — это тоже очень много, какого-то серьезного игрока, который смог бы отобрать себе этот кусок, на сегодняшний день пока не наблюдается.

Поисковые системы Рунета — так называемой области Интернет русскоязычного населения мира, имеют свое место.

Назовем наиболее популярные из них: Yandex, Mail.ru (на данный момент транслируют выдачу Yandex), Rambler, Google, Meta.ua, Bigmir.net (на данный момент транслируют выдачу Yandex).

Можно назвать еще некоторых представителей, таких как Webalta или Aport. Но о Webalta пока нет смысла говорить, так как эта поисковая система лишь на заре своей деятельности, и ее будущее доселе неизвестно. Вторая же уже угасает [3].

Оценить эффективность мы можем по количеству передаваемого поискового трафика (по России). Поисковые системы можно расположить так, как указано на рис. 2.

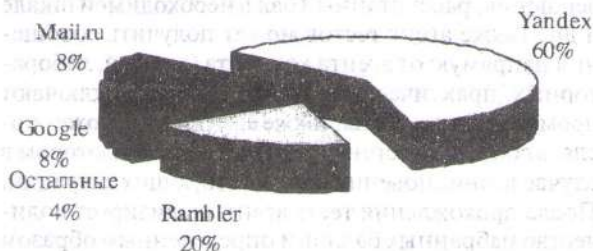


Рис. 2. Количество передаваемого поискового трафика

В целом, как видно из рисунка, лидирующее место занимает система Yandex. Стоит отметить, что она имеет свою собственную уникальную методику поиска и ранжирования необходимой информации.

Статистика по Украине показана на рис. 3.

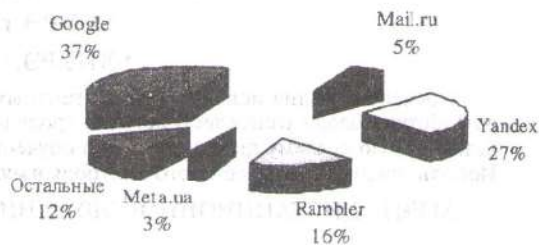


Рис. 3. Количество передаваемого поискового трафика по Украине

Выводы

Самой современной информационной распределенной системой баз знаний является сеть Интернет, которая уже далеко ушла и от первоначальной задачи документооборота, и от первоначальной концепции в качестве простого электронного хранилища данных, перейдя от простой сети немногочисленных вычислительных машин, созданной в научных целях, до обширной системы с невероятно широкой областью применения. Это система, растущая и развивающаяся огромными темпами, которые не замедляются даже в современном мире. Это современная система, она постоянно эволюционирует и вбирает в себя все новые технологии в области лингвистики, статистики, искусственного интеллекта и др. По результатам моих исследований информационная система баз знаний Интернет, обросшая мощными инструментами эффективного поиска запрашиваемой информации, займет главенствующие позиции во всех сферах жизни.

Список литературы: 1. Гусев В. С. «Освоение Internet. Краткое руководство». — Диалектика, 2005. — 288 с. 2. Гусев В. С. «Internet: учеба, работа, полезные ресурсы. Краткое руководство». — Диалектика, 2005. — 256 с. 3. Гусев В. С. «Google: эффективный поиск информации в Интернет. Краткое руководство». — Диалектика, 2005. — 240 с. 4. Клушанов С. В., Ламотько Д. В. «Базы данных» — 2002. 5. Rational Unified Process Made Easy: A Practitioner's Guide to the RUP, The By Per Kroll, Philippe Kruchten Publisher Addison Wesley, Pub Date April 11, 2003, ISBN 0-321-16609-4, Pages 464. 6. Tom Kait «Oracle expert one on one». 7. Turganov A. G. Some Problems of Corporative Information Systems Design // Proc. of the 6th International Workshop on Computer Science and Information Technologies CSIT'2005, Ufa, Russia, 2005, pp. Pages 275.

Поступила в редколлегию 16.05.07

УДК 004.78:37.01

ИНТЕЛЛЕКТУАЛЬНЫЙ АГЕНТ ТЕСТОВОГО КОНТРОЛЯ В ДИСТАНЦИОННОМ ОБУЧЕНИИ

В. Б. Репка¹, Е. В. Ключник², А. А. Козополянская³, И. А. Миняйло⁴

¹ ХНУРЭ, г. Харьков, Украина, victoriya_repka@kture.kharkov.ua

² ХНУРЭ, г. Харьков, Украина, RedCat_87@mail.ru

³ ХНУРЭ, г. Харьков, Украина

⁴ ХНУРЭ, г. Харьков, Украина

Проведен анализ использования агентных технологий в дистанционном обучении. Рассмотрена разработка модели агента тестового контроля в системе Jadex, которая позволяет получить платформо-независимую систему дистанционного обучения. Модель данного агента наследует BDI-архитектуру. Использование агента тестового контроля в мультиагентной системе дистанционного обучения.

АГЕНТ, ДИСТАНЦИОННОЕ ОБУЧЕНИЕ, ТЕСТИРОВАНИЕ, BDI, JADEX.

Введение. Агентные технологии в дистанционном обучении

В последние годы сделан еще один шаг к усовершенствованию систем компьютерной поддержки дистанционного обучения — реализовано множество моделей дистанционного обучения с учетом новейших тенденций создания распределенных систем и использования агентных технологий [1, 2]. Этот факт способствует смещению акцентов в образовании, переводя пользователей систем ДО из ранга пассивного получателя знаний в активных участников процесса обучения.

В классическом понимании, агент — программа, которая после получения задания способна поставить себя на место пользователя. Если агент попадает в тупик, он может задать пользователю вопрос, чтобы определить, каким образом ему необходимо действовать дальше. Агенты рассматриваются как активные объекты, которые, в отличие от обычных (пассивных) компьютерных программ, не «засыпают» до получения следующего сообщения (от пользователя или из внешней среды) и его выполнения, а постоянно функционируют, решая порученные им задачи. Таким образом, главное их отличие состоит в том, что они сами являются инициаторами действий по изменению своего поведения в окружающей среде.

Под интеллектуальным агентом понимается агент, который обладает рядом знаний о себе и окружающем мире и поведение которого определяется этими знаниями. Если простой объект определяется как «данные + методы», то интеллектуальный агент — это уже «данные + методы + знания», причем методы в последнем случае включают функции работы с данными, знаниями, а также методы взаимодействия с окружающей средой и с другими агентами.

В данной работе используется технология Jadex, которая позволяет моделировать агентов для разработки моделей агентов как объектов системы дистанционного обучения. Каждый объект системы дистанционного обучения представлен BDI-агентом с наборами фактов, целей, планов и др., кото-

рые используются для обеспечения поддержки персонализированного обучения. Прежде всего, вся система ДО взаимодействует с соответствующими агентами. Агенты способны обмениваться сообщениями, чем формируют сеть обучающей методики и помогают студенту разрабатывать его собственную программу обучения.

Тестирование как форма контроля знаний обучаемых на сегодняшний день занимает значимое место. С его помощью можно адекватно оценить общий уровень знаний обучаемого, его знания в конкретной предметной области, частной тематике и т. д. Тестирование также может помочь преподавателю оценить личные качества обучаемого: интеллект, внимание, память и мышление.

Покажем на примере разработки агента тестового контроля идею создания совокупности интеллектуальных агентов, сопровождающих процесс дистанционного обучения и позволяющих изучать материал адаптировано к уже имеющимся знаниям обучаемого и его предпочтениям к логике прохождения курса. Данный агент позволяет обеспечить проверку уровня знаний студентов, выставить им оценку, исходя из ответов на вопросы теста, привести ее к различным шкалам оценивания, контролировать время прохождения теста, а также передавать управление всем остальным агентам системы. Данный агент взаимодействует с агентом-координатором, получая сообщение об активации определенного теста и передавая ему сообщения о факте завершения, рассчитанный балл в необходимой шкале и др. Также агент тестов может получить сообщения напрямую от агента контента (лекций, лабораторных, практических работ), если они включают промежуточные тесты, также агент тестов может послать сообщение агенту чата для связи с тьютором в случае возникновения соответствующих ситуаций. После прохождения теста агент анализирует количество набранных баллов и определенным образом вырабатывает решение: перезапускает данный тест, чтобы студент прошел его заново, или выводит на экран результаты теста и посылает сообщение агенту-координатору о том, чтобы он переслал резуль-

таты преподавателю. Агент работает с разными типами вопросов и разнообразными общепринятыми шкалами оценивания.

1. Модель агента тестового контроля в системе дистанционного обучения

Рассмотрим разработку модели агента тестового контроля в системе Jadex, которая позволяет получить платформу-независимую систему дистанционного обучения. Для определения агента в Jadex необходимо сформулировать его факты, цели и планы [3–5]. Чтобы работать с платформой Jadex, нужно создать два типа файлов: файлы ADF (Agent Definition File) — XML файл, определяющий агента и классы Java для выполнения планов агента.

Выделим для агента тестового контроля его составляющие:

а) цели: запустить тест по нужной теме, вывести вопросы теста, контроль времени прохождения теста, подсчет набранных баллов по тесту, формировать оценку, анализ результата;

б) факты: наличие ответа на вопрос, завершение теста, начало теста, сохранение скриншота экрана, соответствие шкалы оценивания, временной лимит на прохождение теста, анализ результатов тестирования;

в) планы: в зависимости от начала теста — тема, в зависимости от результата ответа — правильно/неправильно, оценка есть/нет, в зависимости от наличия ответа на вопрос — тест продолжается/останавливается, в зависимости от вида теста, в зависимости от количества набранных баллов, в зависимо-

сти от времени, в зависимости от сложности вопроса, в зависимости от законченности теста, в зависимости от времени, потраченного для ответа на вопрос.

На рис. 1 представлена схема работы агента тестового контроля и взаимодействия с другими агентами системы.

В соответствии со схемой (рис. 1) на вход данному агенту от агента-координатора подаются данные о тесте, который хочет/должен пройти студент до/после изучения определенной порции материала.

Предлагаем общий алгоритм работы агента тестового контроля в системе ДО:

— агент тестов реагирует на запрос агента-координатора и выводит пользователю на экран монитора требуемый тест;

— вывод вопросов по тесту. На каждый вопрос накладывается временной лимит, и, если студент не укладывается в него, то ответ на вопрос приравнивается к нулю, и студент переходит к следующему вопросу;

— по истечении данного теста происходит подсчет общего балла;

— анализ полученных результатов: если количество баллов, полученных обучаемым, меньше требуемого порога, агент передает управление агенту-координатору для связи с преподавателем или перезапускает тест для повторного прохождения; если же тест завершен с допустимым количеством набранных баллов, агент выводит результат тестирования, переводит оценку в необходимую шкалу, и ответ отсылается преподавателю.

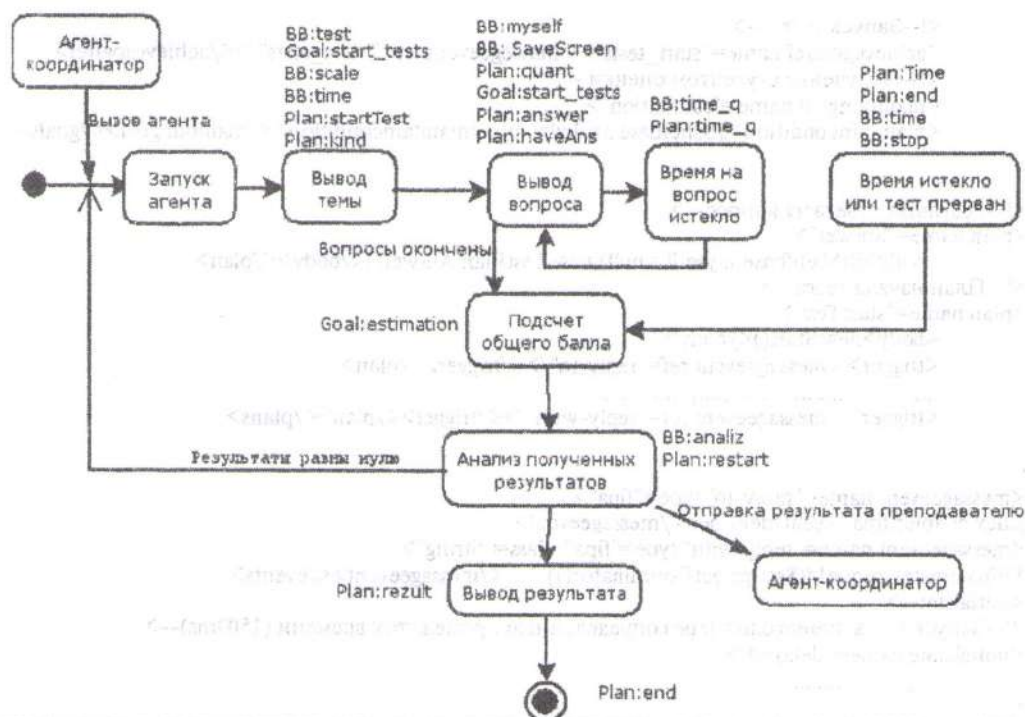


Рис. 1. Схема работы агента тестового контроля

В тесте присутствует таймер, который по истечении заданного промежутка времени прерывает тест, все результаты на оставшиеся вопросы приравниваются нулю, то же происходит при остановке теста испытуемым.

В соответствии с предложенными компонентами модели агента тестового контроля и алгоритмом его работы разработана XML-модель для проведения тестирования (рис. 2).

Разработанная XML-модель использует BDI-архитектуру и позволяет смоделировать агента тестового контроля, который позволяет выводить тест и вместе с этим реагировать на действия студента. Он дает возможность испытуемому получить оцен-

ку за тест, выполняя минимальное количество действий. В данной модели взвешены шкала, время и количество вопросов теста, что позволяет адекватно оценить знания студента.

2. Мультиагентная система дистанционного обучения

Мультиагентные системы — это системы, в которых несколько взаимодействующих интеллектуальных агентов имеют некоторый набор целей или исполняют некоторый набор задач.

При разработке мультиагентной системы ДО в каждый момент времени студент имеет доступ только к одному агенту ДО и следует за командами,

```

<agent xmlns="http://jadex.sourceforge.net/jadex"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://jadex.sourceforge.net/jadex
    http://jadex.sourceforge.net/jadex-0.94.xsd"
  name="StartTest"
  package="test">
  <!--Импортируемые данные-->
  <imports>
    <import>jadex.planlib.*</import></imports>

  <beliefs>
    <!--Параметры теста -->
    <belief name="test" class="Info">
      <fact> new Info($massegeevent)</fact></belief>
    <!--Наличие ответа на вопрос -->
    <belief name="myself" class="Test">
      <fact>new Test.HaveAns()</fact></belief>
      .....
      <fact>new Test.Analiz()</fact></belief></beliefs>

  <goals>
    <!--Запуск теста -->
    <achievegoalref name="start_tests">< massegeevent ref=" start_tests"/></achievegoalref>
    <!--Получение студентом оценки -->
    <maintaingoal name="estimation">
      <maintaincondition>$beliefbase.analiz!=null</maintaincondition></maintaingoal></goals>

  <plans>
    <!--Результат ответа на вопрос-->
    <plan name="answer">
      <body>if($beliefbase.myself!= null) new TestPlan.Answer()</body></plan>
    <!--План начала теста-->
    <plan name="startTest">
      <body>new Start()</body>
      <trigger> <messageevent ref="reply-to"/></trigger></plan>
      .....
      <trigger> <messageevent ref="reply-with"/></trigger></plan></plans>

  <events>
    <messageevent name="reply-to" type="fipa">
      jadex.adapter.fipa. AgentIdentifier </messageevent>
    <messageevent name="reply-with" type="fipa" class="String">
      SFipa.create UniqueId($scope.getCoordinator()) </messageevent></events>
    <initialstates>
    <!--Запуск теста происходит через определенный промежуток времени (1500ms)-->
    <initialstate name="delayed">
      .....
  </agent>
  
```

Рис. 2. XML-модель для проведения тестирования

сгенерированными агентом из его фактов, целей и планов. Когда определенные условия достигнуты, агент посылает сообщения другому агенту с различными целями, например, сделать запрос для справки или переадресовать студента к следующей части образовательной программы. Переходы и связи одного агента с другими формируют студенту его предпочтительную траекторию обучения. Это помогает студенту избежать статической последовательности, которой подчиняется преподаватель. Последовательность переходов управляется агентом-координатором. Основная цель агента состоит в том, чтобы достигнуть лучшего и более тесного взаимодействия между агентами и указать студенту на относительно оптимальные пути.

Курс ДО может иметь в своей структуре лекции, лабораторные работы, практические занятия, глоссарий терминов, тестовые задания, другое. Все это может вызывать различные действия пользователя. Агент ДО для теоретических компонентов (например, лекций) может рассматриваться как простой агент. Это означает, что он имеет очень простые цели и модель быстрого функционирования, основанную на правилах обучения. Пользователь может выполнять некоторые действия, которые генерируют условия для активации правил агента. Наиболее сложный случай такого функционирования агента — когда цикл доставки, отображения, объяснения, повторения оценки достигает удовлетворительного уровня знаний пользователей. Агент контроля может оценивать знания студента путем тестирования.

При выполнении лабораторных работ студенты получают некоторые практические навыки и усиливают теоретические знания. Агент лабораторных работ в системе ДО может помогать студенту. В то время как обучаемый следует командам лабораторных заданий, агент оценивает действия каждого отдельного студента. Если студенту будет необходима помощь, агенты контента могут предложить одну из сохраненных стандартных команд, которые помогут преодолеть проблему или посредством агента чата послать уведомление с запросом преподавателю. Также посредством агента предоставления контекстной помощи обучаемый может, не задействовав преподавателя, получить дополнительную информацию об интересующем понятии или явлении из дополнительных источников или даже глоссария.

Заключение

Агентные технологии все активней применяются в области дистанционного обучения, являясь весьма важным инструментом, облегчающим работу преподавателя и позволяющим сделать обучение более качественным и персонализированным. Они ориентируются на человеческие ощущения и позволяют студенту строить свою программу обучения.

Отметим, что применяемая технология интеллектуальных агентов для разработки системы дистанционного обучения через Интернет позволит не только упростить и качественно улучшить процесс получения человеком знаний и информации, но и даст возможность агенту автономно решать задачи, поставленные перед ним, приобретать и систематизировать знания. Что позволит вывести подобные системы на качественно другой уровень, сделав агентов незаменимыми помощниками в процессе обучения.

Взаимодействуя друг с другом, интеллектуальные агенты позволяют обеспечить процесс активного дистанционного обучения. Это означает, что обучаемый не привязан к конкретной структуре и последовательности прохождения курса, установленной преподавателем, а получает возможность изучения материала в соответствии с его персональными предпочтениями, способностями, уровнем усвоения материала, предыдущим опытом использования систем ДО, а также базового уровня знаний.

При этом мультиагентная система самообучается в соответствии с реакцией пользователя на те или иные события при прохождении курса, вырабатывает свои рекомендации и подсказки и таким образом максимально упрощает работу преподавателя, переводя его в ранг тьютора. В такой системе от преподавателя требуется только качественно подготовленный учебный материал, весь процесс сопровождения обучения выполняют интеллектуальные агенты.

В данной работе приведен пример разработки агента тестового контроля, модель которого наследует BDI-архитектуру. В соответствии с этим для агента разработан набор планов, фактов, целей, событий, которые обеспечивают его функционирование в мультиагентной системе активного дистанционного обучения. Благодаря такой модели разработанный агент тестового контроля может быть использован и как автономное средство обеспечения проведения разного рода тестов при проверке знаний обучаемого.

Список литературы: 1. Глибовец Н. Н. Использование JADE (Java Agent Development Environment) для разработки компьютерных систем поддержки дистанционного обучения агентного типа. // *Educational Technology & Society* 8(3) (ISSN 1436-4522) — 2005, pp. 325-345. 2. I. Keleberda, N. Lesna, V. Repka, Using the Multi-Agent Ontological Approach to Creation of the Distant Learning Distributed Systems. *Educational Technology & Society* 7(2) (ISSN 1436-4522) — 2004, 190-205 pp. 3. A. Pokahr, L. Braubach, and W. Lamersdorf. *Jadex: A BDI Reasoning Engine*. R. Bordini, M. Dastani, J. Dix, and A. El Fallah Seghrouchni. *Programming Multi-Agent Systems*. Kluwer Academic Publishers. 2005. pp. 149-174. 4. A. Pokahr, L. Braubach, and A. Walczak. *Jadex User Guide*. 2005. 5. L. Braubach, A. Pokahr, and A. Walczak. *Jadex Tutorial*. 2005.

Поступила в редколлегию 16.05.07

УДК 681.3.07

Нормалізація зображень: досягнення та невирішені задачі / Є. П. Путятін // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 5–16.

Досліджені можливості нормалізації вхідних зображень при геометричних перетвореннях: афінних, різних нелінійних, у тому числі проєктивних.

У порівняльному аспекті розглянуті три основних математичних методи: метод моментів, загальної та одновимірної кореляції. Звернуто увагу на невирішені задачі нелінійної та комбінованої нормалізації носія та перетворень яскравості.

Лл. 10. Бібліогр.: 15 найм.

UDC 681.3.07

Image normalization: achievement and unsolved problems / Ye. P. Putyatin // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 5–16.

The possibility of normalization of input images subject to the geometric transformations was researched. The considered transformations are affine, different non-linear including projective.

Three different mathematical techniques were considered and compared: these are the method of moments, general and one-dimensional correlation methods. The particular attention was paid to the unsolved problems of non-linear and combined normalization of coordinates and transformation of brightness.

Fig. 10. Ref.: 15 items.

УДК 004.93'1

Метричні властивості грануляції інформації / О. Г. Каграманян, В. П. Машталір, В. В. Шляхов // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 17–24.

Обговорюються метричні властивості розбиттів довільних вимірювальних множин. Визначається та обґрунтовується метрика на кінцевих розбиттях. Пропонуються ймовірнісні інтерпретації різноманітних метричних співвідношень на класах еквівалентностей. Формулюються задачі, які потребують розв'язку щодо метричних властивостей розбиттів вимірювальних множин.

Бібліогр.: 18 найм.

UDC 004.93'1

Metrical properties of information granulation / A. G. Kagramanyan, V. P. Mashtalir, V. V. Shlyakhov // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 17–24.

Metrical properties of arbitrary measurable set partitions are discussed. A metric on finite partitions is introduced and proved. Probabilistic interpretations of various metric relations on equivalence classes are proposed. Unsolved problems concerning metrical properties partitions of measurable sets are formulated.

Ref.: 18 items.

УДК 681.32

Нарис основних принципів квантової лінгвістики / В. А. Широков // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 25–32.

У статті зроблені спроби чіткіше окреслити можливі границі формального підходу до опису мови. Наведені основні квантові принципи, визначена подібність між квантовомеханічними та мисленнєво-мовними системами. Введено поняття семантичного стану, досліджено його формальні аспекти.

Бібліогр.: 17 найм.

UDC 681.32

About main principles of quantum linguistics / V. A. Shirokov // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 25–32.

An attempt to determine formalism of language description is made in this article. The quantum principles is investigated. Similarity between quantum-mechanical and thought-language systems. The concept of semantic state is investigated and formalized.

Ref.: 17 items.

УДК 519.7

Модель згладжування в баченні / М. Ф. Бондаренко, С. Ю. Шабанов-Кушнарєнко, Ю. П. Шабанов-Кушнарєнко // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 33–47.

У статті розглянута задача моделювання процесів бачення. Досліджені теплова модель, модель інерції та іррадіації. Отримано узагальнення закону Талбота. Наведена експериментальна перевірка узагальненого закону Талбота.

Лл.: 7. Бібліогр.: 22 найм.

UDC 519.7

The model of smoothing vision. / M.F. Bondarenko, S. Yu. Shabanov-Kushnarenko, Yu.P. Shabanov-Kushnarenko // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 33–47.

Modelling task of vision process is considered in this article. The heat model, inertial and irradiation model are studied. The generalization of the Talbot's law is given and experimentally tested.

Fig.: 7. Ref.: 22 items.

УДК 519.62

Алгебро-логічні засоби моделювання природної мови. Повідомлення 1 / Г. Г. Четвериков // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 48–52.

У статті розглянуто перспективні напрямки розвитку сучасних цифрових пристроїв, мереж і систем. Стверджується, що розвиток засобів обчислювальної техніки є базою автоматизації розумової діяльності людини. Проте успіхи в галузі інтелектуалізації обчислювальної техніки незначні. Наведено, що технічна реалізація концепції штучного інтелекту потребує значного підвищення продуктивності ЕОМ, а саме: розроблення паралельних (просторових) принципів та методів обробки інформації на рівні інтерфейсу для сприйняття і видачі інформації природною мовою.

Лл.: 2. Бібліогр.: 24 найм.

UDK 519.62

Algebra-logical tools of modeling natural language / G. G. Chetverikov // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 48–52.

In article the perspective directions of modern digital devices, networks and systems development are considered. The carried out analysis shows means of computer facilities development is a baseline of automation of the man intellectual activity. However, the successes in the field of computer facilities intellectualization are insignificant. It asserts and proves that the technical realization of the concept of artificial intelligence requires essential increase of the computer productivity, namely: creation of parallel (spatial) principles and methods of the information processing at the interface level for perception and delivery of the information in natural language.

Fig.: 2. Ref.: 24 items.

УДК 519.7:004.8

Автоасоціативна нейромережна пам'ять на основі нечітких базисних функцій / Є. В. Бодянский, Н. О. Тесленко // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 53–57.

У статті запропоновано автоасоціативну нейромережну пам'ять, яка заснована на нечітких базисних функціях. Описано алгоритм визначення центрів функцій належності, розглянуто процеси запам'ятовування образів фундаментальної пам'яті та їхнього відновлення.

Лл.4. Бібліогр.: 17 найм.

UDC 519.7:004.8

Autoassociative neural memory based on fuzzy basis functions / Ye. V. Bodyanskiy, N. O. Teslenko // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 53–57.

Autoassociative neural memory network, which is based on fuzzy basis functions is proposed in this article. The algorithm of membership functions centers determination is described, processes of fundamental memory patterns learning and their reconstruction are considered.

Fig. 4. Ref.: 17 items.

УДК 519.71

Нейромережна апроксимація багатовимірних функцій за умов наявності завад вимірів / О. Г. Руденко, О. В. Островерхий // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 58–63.

У статті надається порівняльний аналіз відновлення зашумлених багатовимірних функцій нейронними мережами СМАС різної архітектури і РБМ. Розглядається проблема вибору базисних функцій мережі СМАС і параметрів РБМ. Показано, що для вирішення даної задачі досить ефективними є ієрархічна та лінійна архітектури СМАС при виборі як базисної параболічної функції, а застосування РБМ забезпечує задану точність відновлення, потребуючи меншого об'єму пам'яті, але значно більших обчислювальних витрат.

Лл.: 9. Табл.: 1. Бібліогр.: 9 найм.

UDC 519.71

Neuronetwork approximation of multidimensional functions with measurement interferences / O. G. Rudenko, O. V. Ostroverhij // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 58–63.

In the article it is given the comparative analysis of recovering noise-distorted multidimensional functions by RBF and CMAC neural networks of different architectures. The problem of effective selection of the CMAC neural networks basis functions and RBF network parameters is considered. It is shown that hierarchical and linear low-dimensional-CMAC-based architectures with the parabolic basis functions are very effective in the solution of this problem. Application of RBF networks provides the estimated accuracy requiring less memory but considerably higher computational costs.

Fig.: 9. Tab.: 1. Ref.: 9.

УДК 519.71

Гібридні математичні моделі на основі розширень нечітких мереж Петрі / Є. І. Кучеренко, Д. Є. Краснокутський // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 64–67.

Запропоновано розширення нечітких моделей на основі класів нейро-фаззі мереж Петрі. Сформульовано комплекс тверджень, що формалізують умови дозволених переходів за умов синхронізації нечітких процесів на основі нечітких кольорових інгібіторних дуг мереж Петрі. Запропоновано рекомендації щодо впровадження наукових результатів.

Бібліогр.: 10 найм.

UDC 519.71

Hybrid mathematical models based on Petri fuzzy networks widening / Ye. I. Kucherenko, D. Ye. Krasnokutskiy // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 64–67.

It is proposed to expand fuzzy models based on Petri neuro-fuzzy networks classes. The statements what formalizes permissibility of the transition conditions are formulated. Also recommendations on the scientific results implementation are proposed.

Ref.: 10 items.

УДК 004.93:007.52

Метод визначення взаємозв'язку елементів стохастичної нейронної мережі / А. В. Волик, С. М. Герасін, С. І. Лапта // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 68–72.

Запропонований метод визначення взаємозв'язку між елементами стохастичної нейронної мережі. Зв'язки між елементами передбачаються нелінійними, вхідний потік — нестационарний. Метод може бути застосований при аналізі активності мережі і дозволяє проводити оцінку динамічних характеристик нейронної мережі.

Бібліогр.: 8 найм.

UDC 004.93:007.52

Method of determination of intercommunication of elements of stochastic neural network / A. V. Volik, S. N. Gerasin, S. I. Lapta // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 68–72.

The method of determination of intercommunication between the elements of stochastic neuron network is offered. Connections between elements are assumed nonlinear, input stream — non-stationary. The method can be used for the analysis of activity of network and allows to make the estimation of dynamic descriptions of neural network.

Ref.: 8 items.

УДК 681.32

Характеристика архітектури комп'ютерної системи, наближеної до ідеальної / О. П. Кургаєв // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 73–80.

Для вирішення виявлених проблем розвитку архітектури сучасних комп'ютерів вперше запропоновано прийняти структуру систем обробки знань (СОЗ) за зразок архітектури комп'ютерних систем (КС), а як зразок структури СОЗ — ідеал структури наукової теорії, оскільки саме в ній яскраво й найбільш концентровано виражено форму мислення людей при вирішенні проблем. Сформульовані найсуттєвіші ознаки ідеальної архітектури КС, що поєднують універсальність зі спеціалізацією КС на вирішення довільних проблем.

Л.: 4. Бібліогр.: 19 найм.

UDC 681.32

The characteristic of architecture of the computer system approached to ideal / A. P. Kurgaev // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 73–80.

For the decision of revealed problems of development of architecture of modern computers for the first time it is offered to accept structure of systems of processing of knowledge (SPK) for the sample of architecture of computer systems (CS), and as the sample of structure SPK — an ideal of structure of the scientific theory as in it obviously and most konsentrirovanno the form of thinking of people is expressed at the decision of problems. The most essential attributes of ideal architecture CS consistently combining universality with specialization CS on the decision of any problems are formulated.

Fig.: 4. Ref.: 19 items.

УДК 658.512

Про теоретичні основи технології метаконтекстного обміну / З. В. Дудар, О. Г. Ворочек // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 81–87.

У даній роботі технологія метаконтекстного обміну даними розглядається як один з можливих підходів до досягнення семантичної інтероперабельності автономних різномірних баз даних.

Л. 9. Бібліогр.: 12 найм.

УДК 658.512

О теоретических основах технологии метаконтекстного обмена / З. В. Дударь, О. Г. Ворочек // Бионика интеллекта: Научн.-техн. журнал. — 2007. — № 1 (66). — С. 81–87.

В данной работе технология метаконтекстного обмена данными рассматривается как один из возможных подходов к достижению семантической интероперабельности автономных разнородных баз данных.

Ил. 9. Библиогр.: 12 назв.

УДК [005.912:004]:519.862

Математична модель визначення економічної ефективності впровадження корпоративних систем електронного документообігу / Г. Г. Асеев // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 88–94.

Пропонується математична модель ефективності корпоративних систем електронного документообігу, що дозволяє оцінити прямий і непрямий ефекти від впровадження системи незалежно від типу і розміру організації. Результатом розрахунку є визначення загальних витрат на впровадження системи і строку окупності системи. Наводиться приклад розрахунку ефективності впровадження системи.

Бібліогр.: 5 найм.

UDC [005.912:004]:519.862

A mathematical model for determination of economic effectiveness of application of corporate systems of electronic documents circulation / G. Aseyev // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 88–94.

A mathematical model of effectiveness of corporate systems of electronic documents circulation is proposed which makes it possible to evaluate direct and oblique effects of application of a system regardless of the type and dimensions of the organization. The result of calculation is determination of general costs on the system application and the term of the system payback. An example of calculation of effectiveness of the system application is given.

Ref.: 5.

UDK 681.5.015

Інтелектуальні технології підвищення ефективності бізнес-процесів / В. П. Авраменко, В. В. Калачова // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 95–97.

Досліджено застосування інтелектуальних технологій для підтримки прийняття рішень задля підвищення ефективності бізнес-процесів. Розроблено математичні моделі та регуляризовані комп'ютерні процедури ефективного управління бізнес-процесами.

Бібліогр.: 6 найм.

UDC 681.5.015

Intellectual technologies for increasing effectiveness of business process / V. P. Avramenko, V. V. Kalachova // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 95–97.

The intellectual technologies of acceptance support decisions for increasing effectiveness of business process is researched. The mathematical models and the regularization computer procedures for increasing effectiveness of business process were put into operation.

Ref.: 6 items.

UDK 519.85

Екстремальні властивості функцій на класах композиційних образів комбінаторних множин / І. В. Гребеннік, О. В. Баранов // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 98–102.

Досліджено екстремальні властивості опуклих і сильно опуклих функцій на класах комбінаторних множин. Отримано оцінки і достатні умови мінімуму функцій на комбінаторних множинах перестановок кортежів і композицій перестановок. Обговорюється застосування результатів під час розробки методів комбінаторної оптимізації і формування математичного забезпечення інтелектуальних систем прийняття рішень.

Бібліогр.: 18 найм.

UDK 519.85

Extremal properties of functions on classes of composition images of combinatorial sets / I. V. Grebennik, A. V. Baranov // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 98–102.

Extremal properties of convex and strongly convex functions on classes of combinatorial sets are investigated. Estimates and sufficient conditions of minimum of functions on combinatorial sets of tuples permutations and permutations compositions are obtained. Application of the results in development of methods of combinatorial optimization and making up of intelligent decision-making systems software is discussed.

Ref.: 18 items.

UDK 519.62

Модель поверхового рівня натуральної мови на основі семантичних функцій / Г. Ф. Дюбко, Д. В. Преснякова // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 103–106.

Проведено аналіз словосполучень природної мови. Підкреслені словосполучення російської мови. Для кожного словосполучення описується сенс у функціональній формі. Формалізація семантики словосполучень здійснюється семантичними функціями. Запропонована схема обчислення семантичних функцій на основі формалізованого семантичного словника.

Л. 1. Бібліогр.: 6 найм.

UDC 519.62

Model of the surface sense of the natural language on base semantic function / G. F. Dyubko, D. V. Presnjakova // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 103–106.

The Organized analysis of the word-combinations of the natural language. The Chosen word-combinations of the russian language. Sense is described For each word-combination in functional form. The Formalization of the semantics of the word-combinations introduces the semantic function. The Offered model of calculation semantic function on base of the formalized semantic dictionary.

Fig. 1. Ref.: 6 items.

UDK 681.518:004.912

Про візуальне кодування інформації на основі ідентифікації когнітивних функцій людини-оператора. Повідомлення 2 / А. Л. Єрохін // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 107–112.

Розглядається рішення задачі візуального кодування інформації для моделювання психофізіологічних станів людини-оператора. Запропоновано метод невербального неінвазивного впливу на систему психофізіологічного стану людини-оператора.

Бібліогр.: 24 найм.

UDC 681.518:004.912

About visual information coding on base of cognitive function identification of human-operator. Report 2 / A. L. Yerokhin // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 107–112.

The decision of a problem of visual coding information for modelling of psychophysiological states of human-operator is considered. The method of non-verbal non-invasive intrusion to human-operator psychophysiological states is offered.
Ref.: 24 items.

УДК 519.08

Синтез формальної моделі порядкової ординальної класифікації / К. Е. Петров, І. В. Євсєєва // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 113–117.

Запропоновано метод формальної ідентифікації інтелектуальної моделі ординальної класифікації, заснований на ідеях теорії компараторної ідентифікації. Показано, що вирішення даної задачі пов'язане з визначенням моделі оптимальної складності в класі полінома Колмогорова-Габора. Наведені результати чисельного моделювання, які підтверджують ефективність описаного методу.

Табл. 3. Бібліогр.: 16 найм.

UDC 519.08

Synthesis of Formal Model of the Serial Ordinal Classifications / K. E. Petrov, I. V. Evseeva // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 113–117.

The method of formal identification of intellectual model of the ordinal classifications, based on ideas of the theory comparative identification is offered. It is shown, that the decision of the given problem is connected to definition of model of optimum complexity in a class of a Kolmogorov-Gabors polynom. The results of numerical modeling confirming efficiency of the described method are described.

Tab. 3. Ref.: 16 items.

УДК 615.47:617-089

Пошук інформативних кількісних показників електроміографічного сигналу. Повідомлення 1 / А. І. Бих, Т. В. Жемчужкіна, Т. В. Носова // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 118–125.

Робота присвячена визначенню кількісних інформативних показників ЕМГ-сигналу для подальшого диференціювання стану нервово-м'язової системи людини. Основою пошуку є методи кількісного аналізу варіабельності ритму серця. Досліджуються кількісні статистичні часові показники ЕМГ-сигналу. Отримані показники виявилися статистично незалежними і такими, що мають додаткову діагностичну інформацію про стан нервово-м'язової системи людини.

Табл. 1. Лл. 9. Бібліогр.: 4 найм.

UDC 615.47:617-089

Search of informative quantitative characteristics of electromyography signal. Message 1 / A. I. Bykh, T. V. Zhemchuzhkina, T. V. Nosova // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 118–125.

This work is dedicated to determination of informative quantitative characteristics for EMG-signal for further differentiation of the neural-muscle system state. The basis of the analysis is methods of quantitative analysis of heart rhythm variability. In this work quantitative statistical temporal characteristics of EMG-signal are researched. Obtained characteristics are statistically independent and have additional diagnostic information about the neural-muscle system state.

Tab. 1. Fig. 9. Ref.: 4 items.

УДК 004.89

Застосування штучних імунних систем в задачах відновлення й оптимізації нелінійних залежностей / М. М. Кораблев, І. В. Овчаренко, В. В. Токарев // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 126–129.

Штучні імунні системи (ШИС) — напрямок штучного інтелекту, який емулює біологічну імунну систему людини. В роботі розглянуто застосування ШИС для розв'язання задач відновлення й оптимізації нелінійних функцій. Відновлення й оптимізація виконані на прикладі тестових функцій на базі алгоритмів клонального відбору та імунної мережі.

Лл.: 6. Бібліогр.: 7 найм.

UDC 004.89

Artificial immune systems in the problems of reconstruction and optimization of nonlinear functions / N. M. Korablev, I. V. Ovcharenko, V. V. Tokarev // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 126–129.

Artificial immune systems is a direction of an artificial intelligence which emulates biological immune system. The given work consider the use of artificial immune systems in the problems of optimization and reconstruction of nonlinear functions. The problem of reconstruction and optimization is solved for the test functions using the clonal selection algorithm and immune network.

Fig.: 6. Ref.: 7 items

УДК 519.72

Ідентифікація користувачів в системі дистанційної освіти на основі програмних агентів / В. О. Філатов, Р. В. Семєнь // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 130–133.

У статті розглянуто питання застосування сучасних інформаційних технологій в галузі дистанційної освіти. Авторами запропоновано процедуру визначення авторства під час тестування: спершу використовується вхідний тест на клавіатурі для побудови математичної моделі користувача, а потім спеціалізовані тести за дисциплінами. Запропоновано реалізувати підхід за технологією програмних агентів.

Табл. 2. Лл. 3. Бібліогр.: 4 найм.

UDC 519.72

User's identification in the distance education system based on the program agents / V. O. Filatov, R. V. Semenec // *Bionics of Intelligence: Sci. Mag.* — 2007. — № 1 (66). — P. 130–133.

The questions of modern information technologies adaptation in a field of distance education are considered in given paper. The authors offer following procedure of authorship definition participating in testing: at first to use the entrance test on the keyboard for construction the user's mathematical model, and then specialized subject tests. The offered approach is carrying-out with program agent's technology.

Tab. 2. Fig. 3. Ref.: 4 items.

УДК 519.7

Інтелектуалізація процедур визначення координат мобільних терміналів у безпроводних мережах / О. Б. Колесник, Л. В. Колесник // *Біоніка інтелекту: Наук.-техн. журнал.* — 2007. — № 1 (66). — С. 134–138.

У статті розглянуто метод вирішення проблеми набуття фізичних координат мережних терміналів, що пересуваються з часом. Запропонований метод відрізняється порівняною дешевизною і точністю.

Л. 8. Бібліогр.: 2 найм.

UDC 519.7

An intelligent procedure for position's computing of mobile terminals in wireless networks / A. B. Kolesnyk, L. V. Kolesnyk // *Bionics of Intelligence: Sci. Mag.* — 2007. — № 1 (66). — P. 134–138.

Current article is devoted to a new strategy of mobile network management. A method for mobile terminals Local/Global coordinate obtaining system building for wireless networks are considered as an appropriate way to solve managerial problems.

Fig. 8. Ref.: 2 items.

УДК 519.8

Метод порівняння ситуацій при ситуаційно-прецедентному прийнятті рішень / Л. М. Ребезюк, Є. М. Свічкарь // *Біоніка інтелекту: Наук.-техн. журнал.* — 2007. — № 1 (66). — С. 139–144.

Розглядаються задачі, що виникають при розробці процедури застосування ситуаційно-прецедентного методу прийняття рішень. Наводяться основні принципи, на яких базується даний метод. Пропонується метод для порівняння ситуацій, що мають складну багаторівневу структуру. Досліджуються властивості розробленої міри подібності ситуацій.

Л. 3. Бібліогр.: 6 найм.

UDC 519.8

Method of comparison of situations at situational-case decision-making / L. M. Rebezyuk, E. M. Svichkar // *Bionics of Intelligence: Sci. Mag.* — 2007. — № 1 (66). — P. 139–144.

The tasks arising by development of procedure of application of a situational — case method of decision-making are considered. Main principles on which this method is based are resulted. The method of comparison of the situations having complex multilevel structure is offered. Properties of the developed measure of similarity are investigated.

Fig. 3. Ref.: 6 items.

UDK 004.891 : 629.735.33.002.72

Автоматизація підтримки прийняття рішень під час організації агрегатного складання в авіабудуванні на базі спеціалізованої експертної системи / О. С. Топал, В. Є. Зайцев, І. В. Шостак, С. Б. Белов // *Біоніка інтелекту: Наук.-техн. журнал.* — 2007. — № 1 (66). — С. 145–152.

У статті розглянуто мережну модель формування рішень під час реалізації директивної технології агрегатного складання в авіабудуванні. Описано перехід від мережної моделі до відповідної продукційної, в яку вбудовано багатосортну логіку предикатів першого порядку з часом. Наводиться двоетапний метод підтримки прийняття рішень під час організації агрегатно-складального виробництва, який включає постійний моніторинг поточного стану і безпосереднє формування рішень у випадку відхилень ходу виробництва.

Л.: 2. Бібліогр.: 6 найм.

UDK 004.891 : 629.735.33.002.72

Automation of decision-making support during organization of the aggregate assembling in aircraft building based on the dedicated expert system / A. S. Topal, V. Y. Zaitsev, I. V. Shostak, S. B. Belov // *Bionics of Intelligence: Sci. Mag.* — 2007. — № 1 (66). — P. 145–152.

In the article the network model of decisions forming is examined during organization of directive technology of the aggregate assembling in Aircraft building. A transition from network models to the proper product models is described, in which many-kinds first order logic with time is built-in. A twostage method of decision-making support during organization of aggregate-assembling production is listed, including the permanent monitoring of current state and forming of decisions in case of rejections of manufacturing process.

Fig.: 2. Ref.: 6 items.

УДК 681.3.07

Знаходження та нормалізація об'єктів на зображеннях, спотворених проєктивними перетвореннями на прикладі номерного знаку / В. А. Любченко, О. О. Передрій // *Біоніка інтелекту: Наук.-техн. журнал.* — 2007. — № 1 (66). — С. 153–156.

Запропоновано метод виділення областей на зображеннях, спотворених проєктивними перетвореннями.

Послідовно описано та наведено результати роботи запропонованого алгоритму на прикладі задачі виділення номерного знаку автомобіля. Запропоновано метод корегування проєктивних перетворень області зображення.

Л. 8. Бібліогр.: 7 найм.

UDK 681.3.07

Finding and normalization of the objects on the images distorted by the projective transformation demonstrated with an example of the number plate / V. Lubchenko, E. Peredriy // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 153–156.

The method of choosing region in image with projective transformation is offered.

Results that were obtained by applying proposed algorithm for the car license plate recognition are described and reduced in series. Also, the method of the domain projective transformation normalization is offered in the present work.

Fig. 8. Ref.: 7 items.

УДК 519.62

Сучасні інформаційні розподілені системи баз знань / Т. Б. Шатовська, С. П. Меніннік // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 157–161.

Розглядається проблематика сучасних інформаційних систем, первинна концепція, деякі методи проектування. На прикладі всесвітньої мережі Інтернет проаналізовані історичні етапи формування глобальної системи баз знань, створення новітніх технологій в галузі теорії статистики й інших. Описані принципи й основи фізичної всесвітньої мережі як всесвітнього сховища. Наведені приклади надбудови пошуковими системами, їхня частка на ринках США, Росії й України.

Л.: 3. Табл.: 2. Бібліогр.: 7 найм.

UDC 519.62

Modern information systems of bases of knowledge / T. B. Shatovska, S. P. Meninnik // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 157–161.

The problematics of modern information systems, the primary concept, some methods of designing is considered. By the example of a worldnet the Internet historical stages of formation of global system of bases of knowledge, accretion of the newest technologies from area of the theory of statistics and others are analysed. Principles and bases of a physical worldnet as world storehouse are described. Examples of a superstructure search systems, their share in the markets the USA, Russia and Ukraine are resulted.

Fig.: 3. Tab.: 2. Ref.: 7.

УДК 004.78:37.01

Інтелектуальний агент тестового контролю в дистанційному навчанні / В. Б. Репка, К. В. Ключник, Г. О. Козопольська, І. О. Міняйло // Біоніка інтелекту: Наук.-техн. журнал. — 2007. — № 1 (66). — С. 162–165.

Сьогодні все активніше в дистанційне навчання впроваджуються агентні технології, які дозволяють зробити процес навчання пристосованим до потреб користувача. В роботі пропонується застосування агентних технологій для побудови мультиагентної системи ДН. На прикладі агента тестового контролю показано принцип побудови моделі інтелектуального агента. Агент тестового контролю призначений для забезпечення контролю знань студента в системі ДН. Для агента розроблений набір планів, фактів, цілей, подій відповідно до BDI-архітектури.

Л. 2. Бібліогр.: 5 найм.

UDC 004.78:37.01

Intellectual agent of test control in distance education / V. B. Repka, K. K. Klyushnik, A. A. Kozopolyanska, I. A. Minyailo // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 162–165.

Nowadays progressive agent technology is more active used in a distant learning which allows to make process of training adapted to needs of the user. In this paper application of agent technologies for construction of multi-agent systems for DL is offered. By the example of the agent of the test control the principle of construction of model of the intellectual agent is shown. The agent of the test control is intended for maintenance of the control of knowledge of the student in system DL. For the agent the set of plans, the facts, the purposes, events is developed according to BDI architecture.

Fig. 2. Ref.: 5 items.

ОБ АВТОРАХ

Авраменко Валерий Павлович	95	д-р техн. наук, профессор кафедры информационных управляющих систем ХНУРЭ
Асеев Георгий Георгиевич	88	д-р техн. наук, профессор, зав. кафедрой информационных технологий ХДАК
Баранов Алексей Васильевич	98	магистрант ХНУРЭ
Белов Станислав Борисович	145	студент группы 555, специальность «Специализированные компьютерные системы» Национального аэрокосмического университета им. Н. Е. Жуковского «ХАИ»
Бодянский Евгений Владимирович	53	д-р техн. наук, профессор кафедры искусственного интеллекта ХНУРЭ
Бондаренко Михаил Федорович	33	д-р техн. наук, профессор, ректор ХНУРЭ
Бых Анатолий Иванович	118	д-р физ.-мат. наук, профессор, заведующий кафедрой биомедицинских электронных устройств и систем ХНУРЭ
Волик Андрей Валентинович	68	аспирант кафедры высшей математики ХНУРЭ
Ворочек Ольга Григорьевна	81	канд. техн. наук, доцент кафедры программного обеспечения ЭВМ ХНУРЭ
Герасин Сергей Николаевич	68	д-р техн. наук, профессор кафедры высшей математики ХНУРЭ
Гребенник Игорь Валериевич	98	д-р техн. наук, профессор кафедры системотехники ХНУРЭ
Дударь Зоя Владимировна	81	канд. техн. наук, профессор, и. о. заведующего кафедрой программного обеспечения ЭВМ ХНУРЭ
Дюбко Геннадий Федорович	103	канд. техн. наук, профессор кафедры программного обеспечения ЭВМ ХНУРЭ
Евсеева Ирина Викторовна	113	соискатель кафедры системотехники ХНУРЭ
Ерохин Андрей Леонидович	107	д-р техн. наук, начальник кафедры информатики Харьковского национального университета внутренних дел
Жемчужкина Татьяна Владимировна	118	канд. техн. наук, старший преподаватель кафедры биомедицинских электронных устройств и систем ХНУРЭ
Зайцев Виталий Егорьевич	145	канд. техн. наук, доцент кафедры технологии производства летательных аппаратов Национального аэрокосмического университета им. Н. Е. Жуковского «ХАИ»
Каграманян Александр Георгиевич	17	заместитель директора Центра международного образования Харьковского национального университета им. В. Н. Каразина
Калачева Вероника Валериевна	95	канд. техн. наук, старший научный сотрудник, доцент кафедры компьютерных сетей Харьковского университета Воздушных Сил им. И. Кожедуба
Клюшник Екатерина Владимировна	162	студентка факультета компьютерных наук ХНУРЭ
Козопольская Анна Александровна	162	студентка факультета компьютерных наук ХНУРЭ
Колесник Алексей Брониславович	134	соискатель кафедры системотехники ХНУРЭ
Колесник Людмила Владимировна	134	канд. техн. наук, доцент кафедры системотехники ХНУРЭ
Кораблев Николай Михайлович	126	канд. техн. наук, доцент кафедры ЭВМ ХНУРЭ
Краснокутский Дмитрий Евгеньевич	64	аспирант кафедры искусственного интеллекта ХНУРЭ
Кургаев Александр Филиппович	73	ведущий научный сотрудник Института Кибернетики НАН Украины, г. Киев
Кучеренко Евгений Иванович	64	д-р техн. наук, профессор кафедры искусственного интеллекта ХНУРЭ
Лапта Сергей Иванович	68	д-р техн. наук, профессор кафедры высшей математики ХНУРЭ
Любченко Валентин Анатольевич	153	канд. техн. наук, доцент кафедры информатики ХНУРЭ
Мангалир Владимир Петрович	17	д-р техн. наук, профессор, декан факультета компьютерных наук ХНУРЭ
Менинник Станислав Петрович	157	студент факультета компьютерных наук ХНУРЭ
Миняйло Илья Александрович	162	студент факультета компьютерных наук ХНУРЭ
Носова Татьяна Витальевна	118	канд. техн. наук, младший научный сотрудник кафедры биомедицинских электронных устройств и систем ХНУРЭ
Овчаренко Ирина Витальевна	126	аспирантка кафедры ЭВМ ХНУРЭ

Островерхий Алексей Викторович	58	аспирант кафедры ЭВМ ХНУРЭ
Передрий Елена Олеговна	153	студентка факультета прикладной математики и менеджмента ХНУРЭ
Петров Константин Эдуардович	113	канд. техн. наук, старший научный сотрудник ПНИЛ АСУ ХНУРЭ
Преснякова Дарья Владимировна	103	аспирант кафедры программного обеспечения ЭВМ ХНУРЭ
Путятин Евгений Петрович	5	д-р техн. наук, профессор, зав. кафедрой информатики ХНУРЭ
Ребезюк Леонид Николаевич	139	канд. техн. наук, доцент кафедры системотехники ХНУРЭ
Репка Виктория Борисовна	162	канд. техн. наук, доцент кафедры программного обеспечения ЭВМ ХНУРЭ
Руденко Олег Григорьевич	58	д-р техн. наук, профессор, зав. кафедрой ЭВМ ХНУРЭ
Свичкарь Евгений Михайлович	139	магистрант ХНУРЭ
Семенец Руслан Валерьевич	130	инженер Научно-исследовательской лаборатории системных технологий ХНУРЭ
Тесленко Наталья Александровна	53	аспирант кафедры искусственного интеллекта ХНУРЭ
Токарев Владимир Владимирович	126	канд. техн. наук, доцент кафедры ЭВМ ХНУРЭ
Топал Алексей Сергеевич	145	канд. техн. наук, ассистент кафедры программного обеспечения компьютерных систем Национального аэрокосмического университета им. Н. Е. Жуковского «ХАИ»
Филатов Валентин Александрович	130	д-р техн. наук, профессор кафедры искусственного интеллекта ХНУРЭ
Четвериков Григорий Григорьевич	48	д-р техн. наук, профессор кафедры программного обеспечения ЭВМ ХНУРЭ
Шабанов-Кушнаренко Сергей Юрьевич	33	д-р техн. наук, профессор кафедры программного обеспечения ЭВМ ХНУРЭ
Шабанов-Кушнаренко Юрий Петрович	33	д-р техн. наук, профессор кафедры программного обеспечения ЭВМ ХНУРЭ
Шатовская Татьяна Борисовна	157	канд. техн. наук, доцент кафедры программного обеспечения ЭВМ ХНУРЭ
Широков Владимир Анатольевич	25	д-р техн. наук, профессор, директор Украинского языково-информационного фонда НАН Украины, г. Киев
Шляхов Владислав Викторович	17	канд. техн. наук, доцент, ведущий научный сотрудник кафедры информатики ХНУРЭ
Шостак Игорь Владимирович	145	канд. техн. наук, доцент кафедры программного обеспечения компьютерных систем Национального аэрокосмического университета им. Н. Е. Жуковского «ХАИ»

ПРАВИЛА ОФОРМЛЕННЯ РУКОПИСІВ для авторів науково-технічного журналу «БІОНІКА ІНТЕЛЕКТУ»

Науково-технічний журнал «Біоніка інтелекту» приймає до друку написані спеціально для нього оригінальні рукописи, які раніше ніде не друкувались. Структура рукопису повинна бути такою: індекс УДК, заголовок, відомості про авторів, анотація, ключові слова, вступ, основний текст статті, висновки, список використаної літератури.

Відповідно до Постанови ВАК України від 15.01.2003 №7-05/1 (Бюлетень ВАК, №1, 2003, с. 2), стаття повинна мати такі необхідні елементи: постановка проблеми в загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями; аналіз останніх досліджень і публікацій із зазначенням досі не розв'язаних у даній галузі проблем; формулювання цілей та завдань дослідження; виклад основного матеріалу досліджень з повним обґрунтуванням отриманих наукових результатів; висновки з даного дослідження та перспективи подальших досліджень у даному напрямку.

Статті повинні бути виконані в редакторі Microsoft Word. Формат сторінки — А4 (210 × 297 мм), поля: верхнє — 25 мм, нижнє — 20 мм, лівє, правє — 17 мм. Кількість колонок — 2, з інтервалом між ними 5 мм, основний шрифт Times New Roman, кегль основного тексту — 10 пунктів, міжрядковий інтервал — множник (1,1), абзацний відступ — 6 мм. Обсяг рукопису — від 4 до 12 аркушів (мови: російська, українська, англійська).

УДК друкується з першого рядка, без відступів, вирівнювання по лівому краю.

Назва статті друкується прописними літерами; шрифт прямий, напівжирний, кегль 12. **Назви розділів** нумерують арабськими цифрами, виділяють жирним шрифтом. Відступи для назви статті, ініціалів і прізвищ авторів, відомостей про авторів, назв розділів, вступу та висновків, списку літератури: зверху — 6 пт, знизу — 3 пт.

Анотацію (мовою статті, абзац 4–10 рядків, кегль 9) розміщують на початку статті, вона повинна містити інформацію про результати описаних досліджень.

Ключові слова (4–10 слів з тексту статті, які з точки зору інформаційного пошуку несуть змістовне навантаження) наводять мовою рукопису, через кому, в називному відмінку, кегль 9.

Малюнки та таблиці (чорно-білі, контрастні) розміщуються в тексті після першого посилання у вигляді окремих об'єктів і нумеруються арабськими цифрами наскрізною нумерацією за наявності більше ніж одного об'єкта. Невеликі схеми, що складаються з 3–4 елементів рисують, використовуючи вставку об'єкта Рисунок Microsoft Word. Складніші виконують у графічних редакторах у вигляді чорно-білих графічних файлів форматів .tiff, .jpg, .wmf, .cdr із розділенням 300 dpi. Рисунки мають міститися у текстовому файлі, а також обов'язково подаватися окремим файлом з відповідною назвою (наприклад, ris_1.cdr).

Усі елементи малюнка, включаючи написи, повинні бути згруповані. Всі написи в малюнках і таблицях мають бути виконані шрифтом Times New Roman, кегль у малюнках — 10, у таблицях — 9.

Малюнок повинен мати центрований підпис (поза малюнком), шрифт 9, відступи зверху і знизу по 6 пт. Ширина малюнка має відповідати ширині колонки (або ширині сторінки).

Формули, символи, змінні повинні бути набрані в редакторі формул MathType або Microsoft Equation. Формули розміщують посередині рядка й нумерують за наявності посилань на них у рукописі. Шрифт — Times New Roman. Висота змінної — 10 пунктів, великих і малих індексів — 8 пт, основний математичний символ — 12 (10) пт. Змінні, позначені латинськими літерами, набирають курсивом, грецькі літери, скорочення російських слів і цифри — прямим написанням. Змінні, які є в тексті, також набирають у редакторі формул.

Список літератури вміщує опубліковані джерела, на які є посилання в тексті, укладені в квадратні дужки, друкують без абзацного відступу, кегль 9 пт, відступ зверху — 6 пт.

Після списку літератури з відступом зверху 6 пт зазначають дату подання статті до редколегії. Число тамісяць задають двозначними числами через крапку. Розмір шрифта — 9 пт, курсив, вирівнювання по правому краю.

Реферати (Times New Roman, кегль — 9 пунктів, 3–4 речення) подаються російською та англійською мовами. Текст анотації не повинен дублювати реферат.

Разом із рукописом (на аркушах білого паперу формату А4 щільністю 80–90 г/м², надрукований на лазерному принтері, у 2-х примірниках) необхідно подати такі документи:

1. Заяву за наведеним зразком, яку повинні підписати всі автори:

«Прошу прийняти статтю ... (зазначити прізвища авторів, назву статті) обсягом ... сторінок (вказати кількість сторінок) для опублікування в науково-технічному журналі «Біоніка інтелекту». Оплату гарантуємо.

Відомості про авторів: (вказати прізвище, ім'я та по батькові, місце роботи, вчений ступінь, наукове звання, контактний телефон, поштову та електронну адресу).

Підписи авторів».

2. Акт експертизи про можливість опублікування матеріалів у відкритому друку.

3. Рецензію, підписану доктором наук.

4. Відомості про авторів.

5. Електронний варіант (на дискеті 3,5" або електронною поштою) рукопису, реферату та відомостей про авторів.

6. Оплата за публікацію.

ПРАВИЛА ОФОРМЛЕНИЯ РУКОПИСЕЙ ДЛЯ АВТОРОВ научно-технического журнала «БИОНИКА ИНТЕЛЛЕКТА»

Научно-технический журнал «Бионика интеллекта» принимает для опубликования написанные специально для него оригинальные рукописи, которые нигде раньше не издавались. Структура рукописи должна быть следующей: индекс УДК, заглавие, сведения об авторах, аннотация, ключевые слова, введение, основной текст статьи, выводы, список использованной литературы.

Согласно Постановлению ВАК Украины от 15.01.2003 №7-05/1 (Бюллетень ВАК, №1, 2003, с. 2), статья должна иметь следующие необходимые элементы: постановка 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.

Статьи должны быть выполнены в редакторе Microsoft Word. Формат страницы — А4 (210 × 297 мм), поля: верхнее — 25 мм, нижнее — 20 мм, левое, правое — 17 мм. Количество колонок — 2, с интервалом 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).

УДК печатается с первой строки, без отступа, выравнивание по левому краю.

Название рукописи печатается прописными буквами; шрифт прямой, полужирный, кегль 12.

Название разделов нумеруют арабскими цифрами, выделяют жирным шрифтом. Отступы для названия рукописи, инициалов и фамилий авторов, сведений об авторах, названий разделов, введения и выводов, списка литературы сверху — 6 пт, снизу — 3 пт.

Аннотация (на языке рукописи, абзац 4–10 строк, кегль 9) помещается в начале статьи и содержит информацию о результатах описанных исследований.

Ключевые слова (4–10 слов из текста статьи, которые с точки зрения информационного поиска несут смысловую нагрузку, на языке рукописи, через запятую в именительном падеже, кегль 9).

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. Простые схемы, состоящие из 3–4 элементов, рисуют, используя вставку объекта Рисунок Microsoft Word. Более сложные выполняют в графических редакторах в виде черно-белых графических файлов форматов .tiff, .jpg, .wmf, .cdr с разрешением 300 dpi. Рисунки должны быть включены в текстовый файл, а также обязательно представлены отдельным файлом с соответствующим названием (например, fig_1.cdr).

Все элементы рисунка, включая надписи, должны быть сгруппированы. Все надписи в рисунках и таблицах должны быть выполнены шрифтом Times New Roman, кегль в рисунках — 10, в таблицах — 9. Табличный заголовок располагается справа над таблицей (кегель 9 пт).

Рисунок содержит подрисуночную центрированную подпись (вне рисунка), кегль 9, по центру, отступы сверху и снизу по 6 пт. Ширина рисунка должна соответствовать ширине колонки (или ширине страницы).

Формулы, символы, переменные должны быть набраны в редакторе формул MathType или Microsoft 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 математический символ — 12 (10) пт. Переменные, обозначенные латинскими буквами, набирают курсивом, греческие буквы, сокращения русских слов и цифры — прямым начертанием. Переменные, содержащиеся в тексте, также набирают в редакторе формул.

Небольшие формулы можно размещать не в отдельной строке, а прямо по тексту.

Список литературы включает опубликованные источники, на которые имеются ссылки в тексте, заключенные в квадратные скобки, печатается без абзацного отступа, кегль 9 пунктов, отступ сверху — 6 пт.

После списка литературы с отступами сверху — 6 пт указывается дата поступления статьи в редколлегия. Число и месяц задаются двузначными числами через точку. Размер шрифта — 9 пт, курсив, выравнивание по правому краю.

Рефераты (Times New Roman, кегль — 9 пунктов, 3–4 предложения) на украинском и английском языках. Текст аннотации не должен дублировать реферат.

Вместе с рукописью (на листах белой бумаги формата А4 плотностью 80–90 г/м², напечатанной на лазерном принтере, в 2-х экземплярах) необходимо подать следующие документы:

1. Заявку утвержденного образца, подписанную всеми авторами:

«Прошу принять статью ... (указать фамилии авторов, название статьи) объемом ... страниц (указать количество страниц) для опубликования в научно-техническом журнале «Бионика интеллекта». Оплату гарантируем.

Сведения об авторах: (указать фамилию, имя, отчество каждого автора, место работы, ученую степень, научное звание, контактный телефон, почтовый и электронный адреса).

Подписи авторов».

2. Акт экспертизы о возможности опубликования материалов в открытой печати.

3. Рецензию, подписанную доктором наук.

4. Сведения об авторах.

5. Электронный вариант (на дискете 3,5" или электронной почтой) рукописи, реферата и сведений об авторах.

6. Оплата за публикацию.

INSTRUCTIONS FOR AUTHORS OF MANUSCRIPTS of the scientific journal «BIONICS of INTELLIGENCE»

The scientific journal «Bionics of intelligence» accepts for publication original manuscripts which have not been published earlier. The manuscript structure should be as follows: Universal Decimal Classification (UDC) title, authors' initials and surname (in alphabetical order), abstract, key words, introduction, main text, conclusions, references.

According to the Editorial board resolution, based on the Presidium Convention of Ukraine's Supreme Attestation Committee of 15.01.2003 №7-05/1 (Bulletin of Supreme Attestation Committee, №1, 2003, p.2) manuscripts must have the following required elements: introduction (general statement of a problem and its relation to important scientific and practical tasks; analysis of recent research, publications and highlighting of unsolved parts of the general problem in the given field); formulating aims and tasks of research; presentations of the main research material with full substantiation of scientific results obtained; conclusions and perspectives of further research in the given field.

Manuscripts should be submitted in Microsoft Word. Page format — A4 (210 × 297mm), margins: top — 25mm, bottom — 20mm; left, right — 17mm. Double column format with 5mm spacing, font — Times New Roman, font size — 10 points, line spacing — multiplier (1,1), indentation — 6mm. The manuscript should be from 4 to 12 pages (languages: Russian, Ukrainian, English).

The *UDC* is published from the first line, without indentation, the alignment is by a left edge. The *title* is in capital letters; the type is medium bold-faced Roman; type size 12. *The names of sections* are of extra bold type and numbered in Arabic figures. There are indentions for the names of manuscripts, initials and surnames of authors, information about authors, the names of sections, introduction and conclusions, references: top — 6pt; bottom — 3pt.

An abstract (in the language of a manuscript, an indention is made up of 4–10 lines; type 9) is in the beginning of an article and contains information about the results of described studies.

Key words (4–10 words from the text of an article, which from the point of view of information search bear sense in the language of a manuscript, by way of a comma in nominative case, type 9).

Figures and tables (black-and-white, sharp and of good quality) should be in a text after a first reference in the form of embedded item and numbered separately by Arabic numerals in case of more than one item. All legends of figures and tables, including inscriptions, must be grouped. All inscriptions in figures and tables must be in Times New Roman, font size in figures — 10, in tables — 9. A table title is to the right above the

table (font size — 9). The figure should contain a centered figure legend (outside a figure), font size 9, in the centre, top and bottom indentions — 6pt. The figure width must agree with the column width (or page width).

Equations, symbols, variables should be submitted in Math Type (Equation). Equations are centered and numbered in case of references in the text. The font — Times New Roman. The size of variable — 10 points, superscript and subscript characters — 8pt, a main math. symbol — 12 (10) pt. Variables, designated by Latin letters, should be italicized; Greek letters, abbreviations of Russians words and figures should be set in Roman type. Variables which are in the text are also submitted in Math Type (Equation).

References, submitted to the state standards, include published sources that are referred to in the main text in square brackets, without an indentions, 9pt., top indention — 6pt.

The date of receiving an article by the Editorial board is designated after the references with top indentions — 6pt. Date and month should be given in numbers by way of a full stop. The font size — 9pt, italic type, alignment should be done on the right edge.

Abstracts should be submitted in two languages: Ukrainian and Russian (Times New Roman, 9pt, 3–4 sentences). The text of a resume must not duplicate an abstract.

The following documents must be submitted together with a manuscript on A4 format white color sheets of 80–90gr/m² density typed on a laser printer:

1. An application of the following form signed by all the authors:

«You are kindly requested to accept the paper (authors' full names and the name of an paper should be indicated) in ... pages (the number of pages should be indicated) for publication in the scientific journal «Bionics of intelligence». We guarantee the payment.

Information about the authors (surname, first name and patronimic of each authors, place of work, degree, academic status, contact telephone, mailing and electronic addresses should be indicated).

Signatures of authors».

2. A certificate of expertise about a possibility of having the materials published in the press.

3. A review signed by a doctor of sciences.

4. Information about the authors.

5. An electronic variant of a manuscript, an abstract and information about the authors (on a 3.5" diskette or by electronic mail).

6. A receipt of payment for publication.

АЛГЕБРО-ЛОГІЧНІ ЗАСОБИ МОДЕЛЮВАННЯ ПРИРОДНОЇ МОВИ

Г. Г. Четвериков¹, І. Д. Вечірська²¹ ХНУРЕ, м. Харків, Україна, chetvergg@kture.kharkov.ua;² ХНУРЕ, м. Харків, Україна, ira_se@list.ru

Проведено аналіз алгебро-логічної структури природної мови. Розглянуто концептуально-методологічний підхід до мови людини, що дозволяє сприймати її як деяку алгебру, а її тексти — як формули цієї алгебри.

МОВА ПРИРОДНА, АЛГЕБРА ПРЕДИКАТИВ, ВІДНОШЕННЯ, АЛГЕБРА ПРЕДИКАТНИХ ОПЕРАЦІЙ.

Вступ

Формальним моделям семантико-синтаксичних структур мови відводиться вирішальна роль у сучасній проблематиці комп'ютерної лінгвістики та системах штучного інтелекту (ШІ). Це пов'язано з необхідністю створення програмно-апаратного комплексу генерації та аналізу речень природної мови (ПМ).

1. Дослідження алгебро-логічної структури природної мови

У роботі використовується апарат алгебри предикатів [1]. Множина U може бути як скінченною, так і нескінченною. У першому випадку простір U^m називатимемо скінченним, а в іншому — нескінченним.

$$P(x_1, x_2, \dots, x_n) = \begin{cases} 0, & \text{якщо } (x_1, x_2, \dots, x_n) \notin T, \\ 1, & \text{якщо } (x_1, x_2, \dots, x_n) \in T. \end{cases} \quad (1)$$

Згідно з (1), можливий перехід від будь-якого відношення T до відповідного йому предикату P . Предикат P , що знаходимо по (1), називатимемо характеристичною функцією відношення T .

2. Шляхи автоматизації обробки мовної інформації

Сьогодні в системах штучного інтелекту машинний словник та комплекс програм (тезауруси) використовуються, як правило, для виконання будь-якої однієї функції.

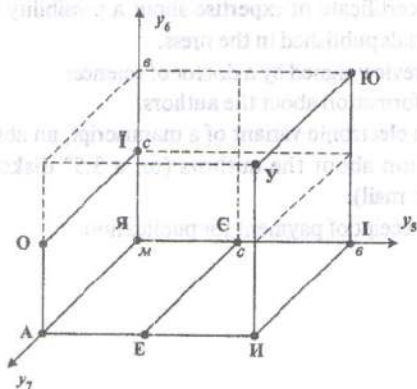


Рис. 1. Тривимірний простір ознак для голосних фонем

Висновки

У статті запропонована узагальнена структура інтелектуальної системи, яка відповідає новій інформаційній технології вирішення задач на ЕОМ, орієнтованих на досягнення високорівневої технології обробки мовної інформації (отримання нової якості). Істотно новим в роботі є розширення алгебри скінченних предикатів (АСП). Тепер вона охоплює не тільки скінченні предикати, а також — нескінченні. Область її рекомендованого застосування розширена та охоплює довільні відношення, які надалі будемо описувати за допомогою ДКАП.

Список літератури: 1. Бондаренко М. Ф., Шабанов-Кушнarenко Ю. П. О бионике интеллекта // Бионика интеллекта. — Харьков: ХНУРЕ, 2004. — Вып. 1(61). — С. 3–14. 2. Бондаренко М. Ф., Коноплянко З. Д., Четвериков Г. Г. Основы теории синтеза надшвидкодiючих структур мовних систем штучного інтелекту. — К.: ІЗМН, 1997. — 264 с.

Поступила до редколегії 15.02.2007

УДК 519.62

Алгебро-логические средства моделирования естественного языка / Г. Г. Четвериков, И. Д. Вечирская // Бионика интеллекта: научн.-техн. журнал. — 2007. — №1 (66). — С. 00–00.

В статье рассматриваются перспективные направления развития современных цифровых устройств, сетей и систем. Утверждается, что развитие средств вычислительной техники является основой автоматизации умственной деятельности человека.

Ил. 5. Библиогр.: 7 назв.

UDK 519.7

Algebra-logical tools of modeling natural language / G. G. Chetverikov, I. D. Vechirskaya // Bionics of Intelligence: Sci. Mag. — 2007. — № 1 (66). — P. 00–00.

In article the perspective directions of modern digital devices, networks and systems development are considered. The carried out analysis shows means of computer facilities development is a baseline of automation of the man intellectual activity.

Fig. 5. Ref.: 7 items.

Просимо надати короткі біографічні дані (до 20 рядків) та скановані фотографії кожного з авторів.

Видавництво здійснює остаточне форматування тексту відповідно до вимог друку.

Адреса редакції:

Україна, 61166, м. Харків, пр. Леніна, 14

ХНУРЕ, к. 285, тел. 702-14-46, факс 702-10-13

e-mail: bionics@kture.kharkov.ua

ИНФОРМАЦИОННОЕ СООБЩЕНИЕ

Факультет компьютерных наук ХНУРЭ
Ежемесячный научный семинар Q.E.D.



Если в рамках научных направлений факультета КН Вы

хотите обсудить в «полновесном» формате (до полутора часов):

- «Новый результат в ...»
- «Состояние и перспективы развития ...»
- «Сравнительный анализ подходов (методов...) ...»

сообщить в формате «лягушатника» (до 20 минут):

- обоснование темы диссертации
- цели и задачи диссертационных исследований
- теоретическую значимость и практическую ценность результатов
- полученные теоретические результаты и их новизну
- представить монографию, статью, доклад ...

приглашаем принять участие в работе семинара Q.E.D.

Постоянные члены семинара:

проф. Авраменко В.П.
проф. Бескоровайный В.В.
проф. Бодянский Е.В.
проф. Бондаренко М.Ф.
проф. Борячок М.Д.
проф. Гребенник И.В.
проф. Дударь З.В.
проф. Кучеренко Е.И.
проф. Левыкин В.М.
проф. Лесная Н.С.
проф. Машталир В.П. (зам. председателя)
проф. Петров Э.Г. (председатель)
доц. Рябова Н.В.
проф. Филатов В.А.
доц. Чалая Л.Э. (ученый секретарь)
проф. Четвериков Г.Г.
проф. Шабанов-Кушнаренко Ю.П.

***Семинар проводится
каждый второй четверг
в 15⁰⁵, ауд. 259***

СЛЕДИТЕ ЗА ОБЪЯВЛЕНИЯМИ!

Запись для выступлений на семинаре
у секретаря (комн. 370)

ПРИГЛАШАЮТСЯ ВСЕ ЖЕЛАЮЩИЕ

Харьковский национальный университет
радиоэлектроники, г. Харьков, просп. Ленина, 14

Наукове видання

БІОНІКА ІНТЕЛЛЕКТУ
інформація, мова, інтелект

Науково-технічний журнал

№ 1 (66)
2007

Головний редактор — *М. Ф. БОНДАРЕНКО*
Відповідальний редактор — *Ю. П. Шабанов-Кушнарєнко*
Заступник відповідального редактора — *Г. Г. Четвериков*
Відповідальний секретар — *І. Д. Вечірська*

Коректор — *Н. В. Єрик*
Комп'ютерна верстка — *О. Б. Ісаєва*

Рекомендовано Вченою Радою
Харківського національного університету радіоелектроніки
(протокол № 33 від 25.05.2007)

Адреса редакції:
Україна, 61166, Харків-166, просп. Леніна, 14,
Харківський національний університет радіоелектроніки, к. 285,
тел. 702-14-46, факс 702-10-13,
e-mail: bionics@kture.kharkov.ua

Підписано до друку 25.05.2007. Формат 60 × 84 1/8. Друк ризографічний.
Папір офсетний. Гарнітура Newton. Умов. друк. арк. 20,1. Обл.-вид. арк. 23,4.
Тираж 300 прим. Зам. № 2-484.

Надруковано в навчально-науковому видавничо-поліграфічному центрі ХНУРЕ
61166, Харків-166, просп. Леніна, 14

покупка на 25/05 2008?