

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ УКРАИНЫ

ХАРЬКОВСКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ  
РАДИОЭЛЕКТРОНИКИ

ISSN 0555-2656

# БИОНИКА ИНТЕЛЛЕКТА

ИНФОРМАЦИЯ, ЯЗЫК, ИНТЕЛЛЕКТ

---

---

№ 1 (68)

2008

---

---

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>Бондаренко М.Ф., Шабанов-Кушнарченко С.Ю., Шабанов-Кушнарченко Ю.П.</i> Модель инерции зрения .....	3
<i>Бондаренко М.Ф., Шабанов-Кушнарченко С.Ю., Шабанов-Кушнарченко Ю.П.</i> Модель иррадиации зрения .....	13
<i>Бондаренко М.Ф., Шабанов-Кушнарченко С.Ю., Шабанов-Кушнарченко Ю.П.</i> Модель краевого контраста зрения .....	23
<i>Бурцев Вал. Н., Бурцев Влад. Н., Ерохин А. Л.</i> Световые устройства визуализации для управления дорожным движением .....	30
<i>Тевяшев А.Д., Набатова С.Н.</i> Интеллектуальная система Оптимизации режимов работы компрессорной станции магистрального газопровода .....	34
<i>Левыкин В.М., Чалый С.Ф.</i> Модели представления знаний для бизнес-процессов с изменяемой структурой .....	40
<i>Авраменко В.П., Тулупов В.В.</i> Эвристическое многоальтернативное проектирование компьютерных систем обработки данных.....	45
<i>Любченко Т.П.</i> Моделирование морфологии естественного флективного языка.....	52
<i>Прасол И.В., Нечипоренко А.С.</i> Методика оценки уровня громкости звукового сигнала с учётом особенностей звуковосприятия.....	65
<i>Процай Н.Т.</i> Кванторная алгебра предикатных операций .....	69
<i>Кобылин О.А.</i> Использование вейвлет-разложения изображения для компенсации геометрических искажений .....	74
<i>Гребенник И. В., Романова Т. Е., Шеховцов С. Б.</i> Принятие решений в информационных системах решения задач геометрического проектирования .....	79
<i>Yaskov G. N.</i> Packing spheres into a parallelepiped by means of the decremental neighbourhood method .....	84
<i>Гороховатский В. А.</i> Применение отношений на множестве характерных признаков изображений при распознавании на основе голосования .....	87
<i>Бритик В.И., Егорова Е.А.</i> Выделение информативных признаков в задачах распознавания образов .....	94
<i>Машталир С.В., Сверхдюк Н.Л.</i> Метод нахождения характерных признаков формы ладони по скелетону изображения.....	101
<i>Кораблёв Н.М.</i> Формализация нечеткой информации при оценивании качественных признаков в вербальных шкалах .....	106
<i>Бых А. И., Жемчужкина Т. В., Носова Т. В.</i> Поиск информативных количественных показателей электромиографического сигнала. Сообщение 2. ....	111
<i>Шигимага В.А.</i> Графоаналитические методы определения параметров необратимого импульсного пробоя мембраны клетки (часть 2).....	117
<i>Самынина М.Г.</i> Способ построения измерительной системы для регистрации малых изменений температуры.....	123
<i>Ситник Л.Г., Шабанов-Кушнарченко С.Ю.</i> Решение систем дизъюнктивных уравнений методом логических определителей .....	128
<i>Удовенко С.Г., Перепелица В.И.</i> Рекуррентное байесовское оценивание состояний управляемых стохастических процессов.....	134
<i>Тимофеев В. А., Самер Лага, Непочатова В. Д.</i> Модифицированный рекуррентный метод наименьших квадратов с критическими свойствами.....	139
<i>Лесная Н.С., Репка В.Б., Ивченко О.В., Шерстнюк А.В.</i> Экспериментальные исследования эффективности карт кохонена для визуализации закономерностей.....	143
<i>Шатовская Т.Б., Каменева И.В., Тарасов Ю.А.</i> Комбинированный подход к классификации текстов .....	149
<i>Четвериков Г.Г., Лесовец Т.В., Касала О.В.</i> Искусственные нейронные сети как метод интеллектуального анализа данных в финансовых системах.....	153
<i>Kirichenko L. O., Radivilova T.A.</i> Analysis of network performance under selfsimilar system loading by computer simulation .....	158
<i>Олейник А.А., Субботин С.А.</i> Синтез моделей сложных объектов и систем на основе метода пчелиной колонии .....	161
<i>Штанько В.И., Добровольская Е.В.</i> Философские основания формальной онтологии (феноменологический и аналитический аспекты) .....	171
<i>Герасин С.Н., Калинин О.В., Лоцман В.П., Троцило А.С.</i> О способе кодирования сигналов различной природы .....	177
Об авторах.....	181
Правила оформления рукописей для авторов научно-технического журнала «Бионика интеллекта» .....	183

УДК 519.7



## МОДЕЛЬ ИНЕРЦИИ ЗРЕНИЯ

М.Ф. Бондаренко<sup>1</sup>, С.Ю. Шабанов-Кушнарченко<sup>2</sup>,  
Ю.П. Шабанов-Кушнарченко<sup>3</sup>

<sup>1, 2, 3</sup> ХНУРЭ, г. Харьков, Украина

Исследуется модель инерции зрения и ее динамические реакции на различные виды входных сигналов. Представлены диаграммы, дающие представление о характере изменения яркости и светлоты зрительного ощущения во времени. Обобщен закон Плато. Сформулировано условие, обеспечивающее равенство критических частот для периодических мельканий произвольной формы.

### МОДЕЛЬ ИНЕРЦИИ ЗРЕНИЯ, КРИТИЧЕСКАЯ ЧАСТОТА МЕЛЬКАНИЙ, ЗАКОН ПЛАТО

#### Введение

В статье рассматриваются некоторые аспекты экспериментальной проверки математической модели инерции зрения Аллара — Луизова:

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

вытекающей как частный случай из модели инерции и иррадиации, описанной в работе [1]. Луизов [2, с. 108] проверял эту модель на фактах пороговой видимости прямоугольной вспышки и получил хорошее согласование теории с опытом.

Он рассмотрел входные сигналы вида:

$$B(t) = \begin{cases} B_0, & \text{если } t \leq 0, \\ B_0 + \Delta B, & \text{если } 0 < t \leq \Delta t, \\ B_0, & \text{если } t > \Delta t \end{cases} \quad (2)$$

и получил следующую реакцию модели:

$$S(t) = \begin{cases} kB_0, & \text{если } t \leq 0, \\ kB_0 + k\Delta B(1 - e^{-\frac{t}{a}}), & \text{если } 0 < t \leq \Delta t, \\ kB_0 + k\Delta B(e^{\frac{\Delta t}{a}} - 1)e^{-\frac{t}{a}}, & \text{если } t > \Delta t. \end{cases} \quad (3)$$

На рис. 1 а и б показан ход изменения во времени входного и выходного сигналов модели.

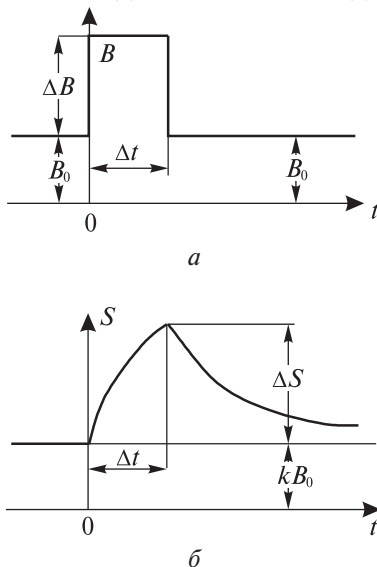


Рис. 1

Основываясь на этих данных, Луизов вывел зависимость, связывающую яркость вспышки  $\Delta B$  с ее пороговой длительностью  $\Delta t_n$ ,

$$\Delta t_n = a \ln \frac{1}{1 - \frac{\Delta B_n}{\Delta B}}, \quad (4)$$

которая допускает экспериментальную проверку. В формуле (4)  $\Delta B_n$  означает статический порог различения яркости,  $a$  — постоянная инерции зрения.

Луизов также выполнил эксперименты по проверке этой зависимости и получил хорошее согласование теории с опытом [2]. На рис. 2 дано сравнение вычислений по формуле (4) с результатами одного из его опытов. При построении теоретической диаграммы постоянная инерции зрения принята равной  $a = 0,0595$  сек.

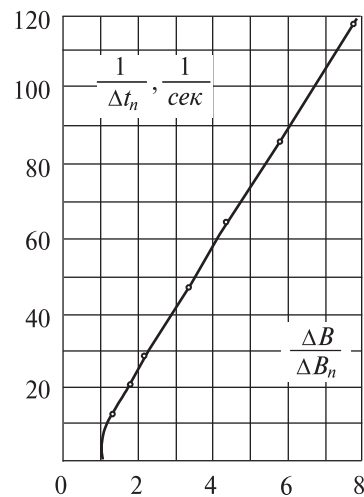


Рис. 2

В этой статье мы продолжим работу в указанном направлении, а именно сравним реакции модели инерции зрения с некоторыми фактами, связанными с критической частотой слияния периодических мельканий [3, 4]. Здесь будет показано, что модель инерции Аллара — Луизова также хорошо согласуется и с этим кругом фактов.

### 1. Исследование реакций модели инерции на прямоугольные мелькания

Изучим реакцию модели инерции на прямоугольные мелькания с произвольным соотношением длительности светлой и темной фаз. Пример такого рода мельканий изображен на рис. 3 а. Яркость зрительной картины принимает два попеременно сменяющих друг друга уровня  $B_1$  и  $B_2$ .

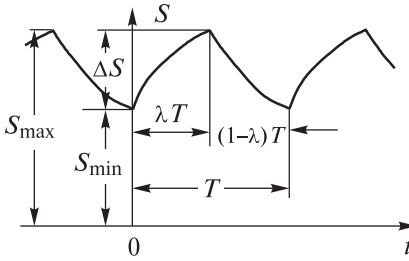
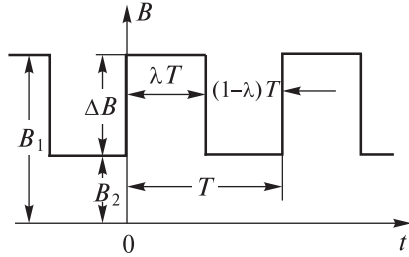


Рис. 3

Математически зависимость яркости  $B$  от времени  $t$  запишется в виде:

$$B(t) = \begin{cases} B_1, & \text{если } nT < t \leq (n+\lambda)T, \\ B_2, & \text{если } (n+\lambda)T < t \leq (n+1)T. \end{cases} \quad (5)$$

Для определенности будем полагать, что  $B_1 > B_2$ . Параметр  $\lambda$  (назовем его параметром скважности мельканий) определяет длительность первой фазы колебания яркости, равную  $\lambda T$ ; длительность второй фазы равна  $(1-\lambda)T$ . Величина  $\lambda$  для мельканий различной формы может изменяться в пределах от 0 до 1.

Используя формулу (1) для условий (5), получим следующее выражение для определения светлоты зрительного ощущения:

$$S(t) = \begin{cases} k B_1 - k \Delta B e^{-\frac{t-nT}{a}} \cdot \frac{1 - e^{-\frac{1-\lambda}{a}T}}{1 - e^{-\frac{T}{a}}}, & \text{если } nT < t \leq (n+\lambda)T, \\ k B_2 - k \Delta B e^{-\frac{t-nT}{a}} \cdot \frac{e^{\frac{\lambda}{a}} - 1}{1 - e^{-\frac{T}{a}}}, & \text{если } (n+\lambda)T < t \leq (n+1)T. \end{cases} \quad (6)$$

В выражении (6) величина  $\Delta B$  обозначает амплитуду колебания яркости при мельканиях:

$$\Delta B = B_1 - B_2. \quad (7)$$

На рис. 3 б представлена диаграмма изменения светлоты зрительного ощущения во времени, построенная по формуле (6). Из диаграммы видно, что светлота колеблется между минимальным  $S_{\min}$  и максимальным  $S_{\max}$  значениями, достигаемыми соответственно в моменты времени  $nT$  и  $(n+\lambda)T$ .

По формуле (6) находим:

$$S_{\min} = k \frac{B_1 (e^{-\frac{1-\lambda}{a}T} - e^{-\frac{T}{a}}) + B_2 (1 - e^{-\frac{1-\lambda}{a}T})}{1 - e^{-\frac{T}{a}}}, \quad (8)$$

$$S_{\max} = k \frac{B_1 (1 - e^{-\frac{\lambda}{a}T}) + B_2 (e^{-\frac{\lambda}{a}T} - e^{-\frac{T}{a}})}{1 - e^{-\frac{T}{a}}}.$$

Величина колебания светлоты равна:

$$\Delta S = S_{\max} - S_{\min} = 2k\Delta B \frac{\operatorname{sh} \frac{\lambda}{2a} \cdot \operatorname{sh} \frac{1-\lambda}{2a}}{\operatorname{sh} \frac{T}{2a}}. \quad (9)$$

Формула (9) показывает, что величина колебания светлоты  $\Delta S$  зависит от значения параметра  $\lambda$ . Характер этой зависимости показан на диаграмме (рис. 4). Как видно из диаграммы, колебание светлоты достигает максимального значения  $\Delta S_{\max}$  при  $\lambda = \frac{1}{2}$ ; то есть когда обе фазы мельканий имеют одинаковую длительность, равную половине периода.

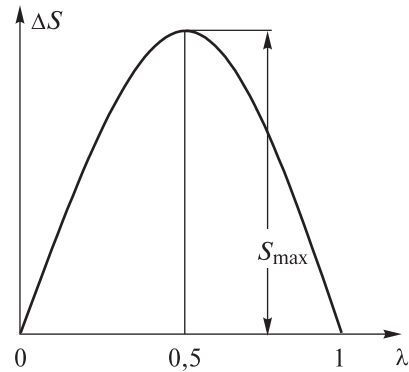


Рис. 4

Согласно формуле (9):

$$\Delta S_{\max} = k\Delta B \frac{T}{4a}. \quad (10)$$

Как видно из формулы (10), величина колебания светлоты  $\Delta S_{\max}$  зависит от времени периода мельканий  $T$ . На диаграмме (рис. 5) показан характер этой зависимости. При неограниченном росте периода мельканий  $T$  колебание видимой яркости стремится к значению  $\Delta S = k\Delta B$ . При уменьшении периода мельканий  $T$  до нуля величина  $\Delta S_{\max}$  также стремится к нулю.

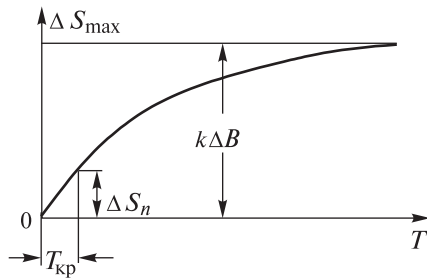


Рис. 5

Из диаграммы на рис. 5 видно, что колебание светлоты зрительного ощущения при уменьшении периода мельканий уменьшается, однако, оно полностью не прекращается при любом, сколь угодно малом периоде мельканий. Может показаться, что практический опыт наблюдения мельканий противоречит этому заключению. Известно, что при некоторой конечной частоте, называемой критической, мелькания совершенно исчезают. Величина критической частоты мельканий зависит от многих факторов и обычно имеет величину порядка 40÷50 Гц. Из этого факта мы не склонны делать вывод о ложности модели инерции зрения.

Дело в том, что о колебании светлоты зрительного ощущения мы узнаем лишь потому, что оно обнаруживается сознанием наблюдателя, выступающего в данном случае в роли измерительного прибора. Именно показания этого прибора мы осознаем как колебание светлоты зрительного ощущения и записываем их в протоколы испытаний. Таким образом, мы приходим к схеме, изображенной на рис. 6.

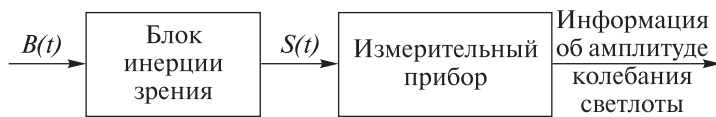


Рис. 6

Однако, пользуясь аналогией с техническими измерительными приборами, мы неизбежно приходим к выводу о том, что никакой реальный измерительный прибор не может иметь неограниченно высокую чувствительность. Точность измерительного прибора всегда ограничена его зоной нечувствительности. Должно обладать зоной нечувствительности и сознание наблюдателя. Следовательно, при достаточно высокой частоте мельканий, когда амплитуда колебания светлоты снизится настолько, что сравняется с зоной нечувствительности сознания, последнее перестанет сигнализировать о мельканиях, и мы не сможем воспринимать мелькания светлоты. Таким образом, из того факта, что мы не обнаруживаем мельканий светлоты зрительного ощущения, еще не следует то, что этих мельканий вовсе нет. Эти мелькания просто могут быть малы и не обнаруживаться сознанием — прибор-

ром, анализирующим эти мелькания. Возможно, что противоречие между теорией и опытом только лишь кажущееся.

Факт наличия зоны нечувствительности сознания можно привлечь для экспериментальной проверки теоретических выводов из модели инерции зрения. Пороговому колебанию светлоты  $\Delta S_n$  соответствует критическая длительность периода мельканий  $T_{кр}$  (рис. 5), связанная с критической частотой слияния мельканий формулой  $n_{кр} = \frac{1}{T_{кр}}$ .

Величину  $\Delta S_n$  будем считать постоянной. К вопросу о том, в какой мере это предположение соответствует фактическому положению дела, мы еще возвратимся.

Найдем выражение для определения критической частоты мельканий  $n_{кр}$  в зависимости от величины колебания яркости мельканий для случая, когда  $\lambda = \frac{1}{2}$ . С этой целью подставим в формулу (10) вместо  $\Delta S_{max}$  величину  $\Delta S_n = k\Delta B_n$ , а вместо  $T$  — величину  $\frac{1}{n_{кр}}$ . В результате получим:

$$n_{кр} = \frac{1}{4a \operatorname{Arth} \frac{\Delta B_n}{\Delta B}} \quad (11)$$

На рис. 7 показано изменение критической частоты мельканий  $n_{кр}$  в зависимости от колебания яркости  $\Delta B$ .

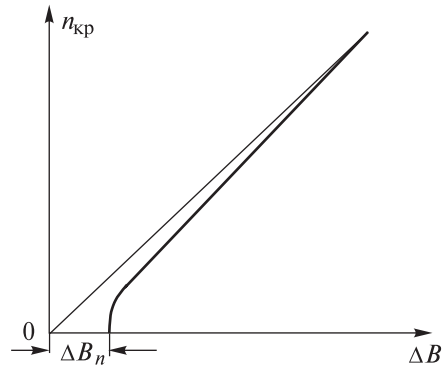


Рис. 7

Как видим, с увеличением колебания яркости критическая частота мельканий растет. Значению  $n_{кр}=0$  соответствует величина колебания яркости  $\Delta B_n$ . Таким образом,  $\Delta B_n$  означает пороговое колебание яркости при очень медленных мельканиях. Формула (11) допускает экспериментальную проверку. При достаточно больших по сравнению с  $\Delta B_n$  значениях  $\Delta B$ , зависимость (11), как это видно из диаграммы на рис. 7, приобретает практически линейный характер:

$$n_{кр} \approx \frac{1}{4a} \cdot \frac{\Delta B}{\Delta B_n} \quad (12)$$

Рассмотрим теперь зависимость критической частоты мельканий  $n_{кр}$  от параметра  $\lambda$ , то есть со-

отношения длительности отдельных фаз в периоде мельканий. С этой целью определим из соотношения (9) величину  $\lambda$ , предварительно заменив в нем  $\Delta S$  на  $\Delta S_n = k\Delta B_n$  и  $T$  на  $\frac{1}{n_{кр}}$ . В результате получим:

$$\lambda = \frac{1}{2} \pm an_{кр} \operatorname{Arch}\left( \operatorname{ch}\frac{1}{2an_{кр}} - \frac{\Delta B_n}{\Delta B} \cdot \operatorname{sh}\frac{1}{2an_{кр}} \right). \quad (13)$$

На рис. 8 в виде диаграммы показан характер зависимости  $n_{кр}$  от  $\lambda$ , даваемой формулой (13). Формула (13) допускает экспериментальную проверку. При достаточно больших по сравнению с  $\Delta B_n$  значениях  $\Delta B$  зависимость (13) приобретает более простой вид:

$$n_{кр} \approx \frac{\lambda(1-\lambda)}{a} \cdot \frac{\Delta B}{\Delta B_n}. \quad (14)$$

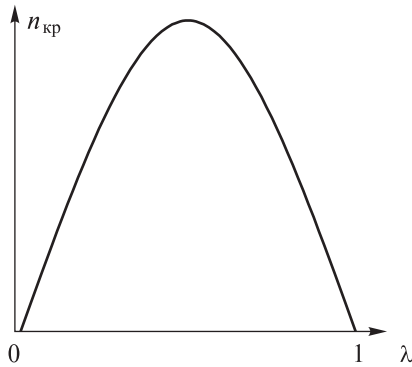


Рис. 8

В последующих разделах этой статьи будет рассмотрен вопрос о частичной экспериментальной проверке формул (11)-(14).

## 2. Сознание как прибор, анализирующий колебания светлоты

В предыдущем разделе мы получили, исходя из модели инерции зрения Аллара-Луизова, несколько зависимостей для описания критической частоты мельканий, допускающих непосредственную опытную проверку. Теперь мы приступаем к частичной проверке этой зависимости на фактах. Прежде всего, из формул (11)-(14) вытекает, что критическая частота совершенно не должна зависеть от общего уровня яркости мельканий. Критическая частота в то же время зависит от значения колебания яркости  $\Delta B$ . Однако, обращаясь к опыту, мы видим, что это не так. Пусть, к примеру, яркость экрана периодически изменяется так, что мелькания заметны для испытуемого. Налагая на эти мелькания свет немигающей яркости и постепенно повышая уровень яркости этого света, мы рано или поздно придем к такому положению, когда мелькания не будут обнаруживаться испытуемым, несмотря на то, что величина колебания яркости осталась прежней.

Несовпадение критических частот мельканий при неодинаковой их средней яркости можно продемонстрировать также и с помощью вертушки. Для этого возьмем кружок белой бумаги, раскрашенный черной тушью так, как показано на рис. 9.

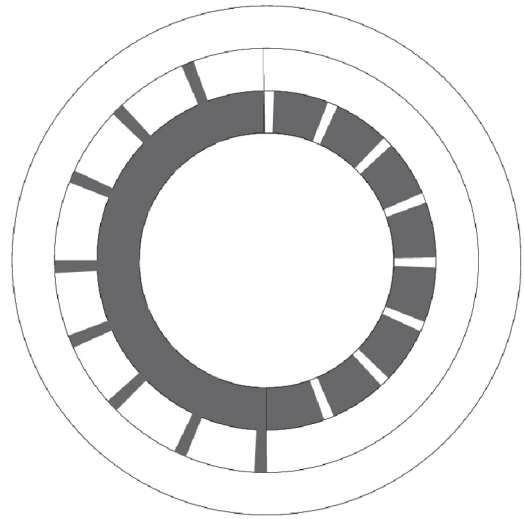


Рис. 9

Дозировка яркости отдельных фаз мельканий осуществлена комбинированием белых и черных коротких дуг, соотношение длин которых определяется на основе закона Талбота. Такие вспомогательные мелькания сольются значительно раньше, чем те, которые нас интересуют, и дадут требуемый уровень яркости отдельных фаз мельканий.

Кружку, изображенному на рис. 9, соответствуют диаграммы мелькающей яркости, представленные на рис. 10. Символы  $B_б$  и  $B_ч$  обозначают яркости белого и черного полей.

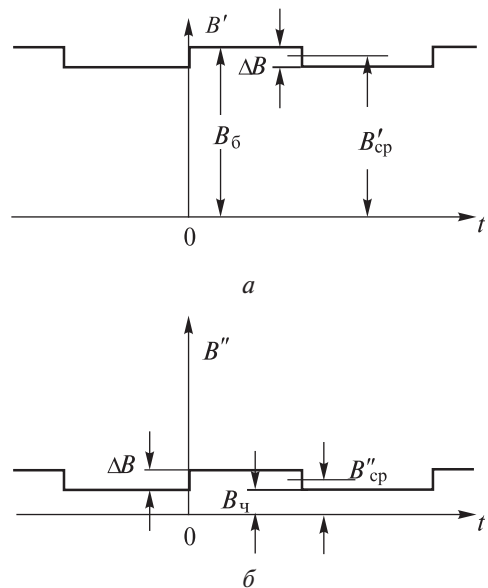


Рис. 10

Как видно из диаграмм, мелькания отличаются лишь средним значением яркости. На первой диаграмме (рис. 10 а)  $B'_{cp} = B_ч + 0,95(B_б - B_ч)$ , на вто-

рой диаграмме (рис. 10 б)  $B''_{cp} = B_{\text{ч}} + 0,05(B_6 - B_{\text{ч}})$ . Амплитуда колебаний в обоих случаях одна и та же и равна  $\Delta B = 0,1(B_6 - B_{\text{ч}})$ .

Опыт показывает, что хотя критические частоты таких мельканий и близки, но в точности они не совпадают. Мелькания на черном фоне сливаются при более высокой частоте, чем мелькания на белом фоне. Очевидно, что такого рода отклонения в функционировании зрения человека от реакций модели носят целесообразный характер. Ведь ясно, что при более низком уровне освещения и амплитуда мельканий будет, как правило, пропорционально меньше. Поэтому повышение чувствительности органа зрения к мельканиям при уменьшении средней яркости выглядит вполне целесообразным. Не считаться с такого рода отклонениями нельзя.

Можно ли согласовать такого рода факты с моделью Аллара — Луизова? Нам представляется это возможным. Снова вспомним, что сознание наблюдателя, воспринимающее мелькания светлоты, выполняет, по существу, функции измерительного прибора. Попытаемся хотя бы эскизно представить, как функционирует этот прибор. С этой целью проведем аналогию сознания, работающего в режиме анализа мельканий, с каким-либо техническим измерительным прибором, например, электронным осциллографом.

Представим, что создана действующая модель Аллара-Луизова, воспринимающая световые мелькания  $B(t)$ , на выходе которой возникают колебания напряжения  $S(t)$ , являющиеся аналогом колебания светлоты. Представим далее, что мы располагаем электронным осциллографом, и нам дано задание определить амплитуду колебания светлоты. Очевидно, что прежде, чем производить определение амплитуды, целесообразно отстроиться с помощью вертикального смещения луча от постоянной составляющей мельканий. Далее, если амплитуда колебания яркости окажется чрезмерно большой, то мы уменьшим усиление, а если малой, увеличим.

Итак, еще до того, как будет осуществлено измерение величины колебания светлоты, представляется целесообразным выполнить следующие две операции: 1) отфильтровать постоянную составляющую мельканий; 2) установить подходящее усиление. Трудно представить, чтобы такой совершенный прибор, как человеческое сознание, не располагал возможностями такого рода. Поскольку в случае сознания о ручном выполнении этих операций говорить не приходится, необходимо полагать, что они выполняются автоматически.

Как уже говорилось, амплитуда мельканий в естественных условиях наблюдения, как правило, будет тем большей, чем больше общий уровень освещенности. В связи с этим естественно допустить,

что величина коэффициента усиления мельканий регулируется в зависимости от общего уровня яркости зрительной картины. Итак мы приходим к следующей модели сознания, работающего в режиме анализа мельканий (рис. 11).

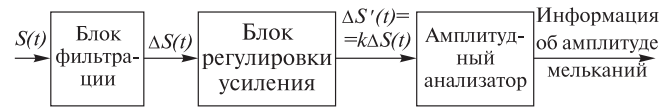


Рис. 11

Мелькания светлоты  $S(t)$  поступают на блок, отфильтровывающий постоянную составляющую мельканий. Оставшееся колебание светлоты  $\Delta S(t)$  поступает на блок регулировки усиления. На этот же блок поступает информация об общем уровне яркости зрительной картины  $B_{cp}$ . Коэффициент усиления блока является функцией  $B_{cp}$ , то есть  $k = k(B_{cp})$ . Если принять такую модель работы сознания, анализирующего мелькания, то нетрудно прийти к заключению о том, что фактическая зависимость критической частоты от уровня яркости мельканий вполне может быть следствием работы блока регулировки усиления. В то же время этот факт может не иметь никакого отношения к вопросу об истинности или ложности модели инерции зрения.

Если эксперимент подтвердит справедливость такого взгляда на роль сознания в анализе мельканий, то никаких изменений в модель инерции зрения вносить не придется. Нужно будет лишь иметь в виду, что коэффициент усиления  $k$ , фигурирующий в модели, может принимать различные значения в зависимости от общего уровня яркости зрительной картины. При таком понимании дела становится ясным, что опыты с определением критических частот, выполненные при различном общем уровне яркости зрительной картины, нельзя сравнивать между собой, если при этом не вносить поправку на изменение коэффициента усиления.

### 3. Модель инерции и закон Плато

Продолжим проверку зависимостей для описания критической частоты слияния мельканий. Пусть имеются два периодических прямоугольных мелькания  $B'(t)$  и  $B''(t)$ , у которых амплитуды колебания яркости одинаковы

$$\Delta B' = \Delta B'' = \Delta B, \quad (15)$$

а параметры скважности  $\lambda'$  и  $\lambda''$  связаны соотношением:

$$\lambda' = 1 - \lambda'' = \lambda. \quad (16)$$

Как следует из формулы (9), колебания светлоты этих мельканий  $\Delta S'$  и  $\Delta S''$  должны быть равны между собой:

$$\Delta S' = \Delta S'' = \Delta S. \quad (17)$$

Отсюда следует, что для таких мельканий должны быть равными также и их критические частоты:

$$n'_{кр} = n''_{кр} \cdot \tag{18}$$

Формула (13) и диаграмма на рис. 8 иллюстрируют это положение.

Известен экспериментальный факт, который носит название закона Плато, имеющий к этому выводу весьма близкое отношение [5]. Плато рассматривал пары мельканий, для которых соблюдаются не только условия (15) и (16), но дополнительно к этому равны между собой яркости отдельных фаз мелькания:

$$\begin{aligned} B'_1 = B''_1 = B_1, \\ B'_2 = B''_2 = B_2. \end{aligned} \tag{19}$$

Пример такой пары мельканий показан на диаграммах а и б рис. 12.

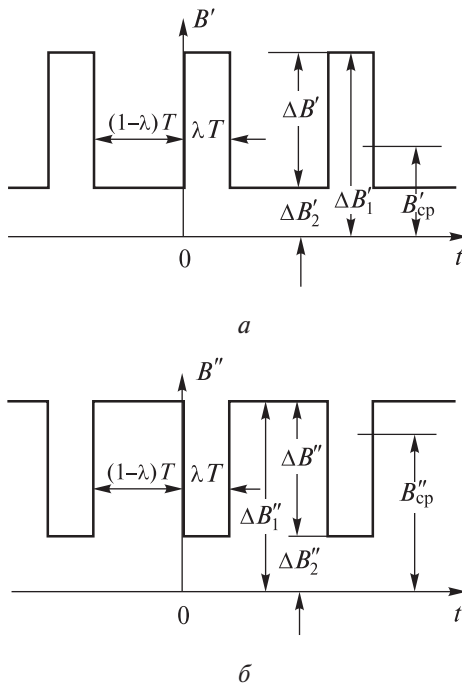


Рис. 12

Пары мельканий такого вида легко получаются на вертушке с помощью так называемых «дополнительных дисков». Пример пары дополнительных дисков представлен на рис. 13. Для перехода от одного диска к другому нужно всего лишь поменять местами яркости секторов. Плато обнаружил, что критические частоты, измеренные по двум дополнительным дискам, равны между собой.

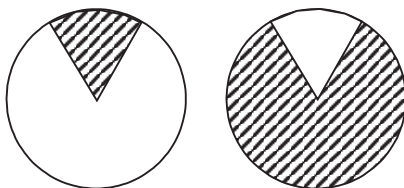


Рис. 13

Можно было бы считать, что закон Плато подтверждает только, что полученный из модели вывод о равенстве критических частот, если бы не одно обстоятельство. Дело в том, что средний уровень яркости на дополнительных дисках в момент слияния мельканий резко различен:  $B'_{cp} \neq B''_{cp}$ . Поэтому нет никаких оснований общий уровень яркости этих двух зрительных картин полагать одинаковым. Следовательно, в соответствии с тем, что было сказано в предыдущем разделе, результаты опытов на дополнительных дисках не могут быть сравниваемы друг с другом, пока мы не введем поправку в коэффициенты усиления модели инерции  $k'$  и  $k''$ , учитывающие различия в общем уровне яркости обеих зрительных картин. Мы не знаем точно, какова зависимость коэффициента усиления от уровня яркости зрительной картины, однако, наличие этой связи несомненно.

Если же мы хотя бы качественно учтем эту связь, то немедленно приходим к выводу: из модели инерции зрения вытекает, что критические частоты, измеренные по дополнительным дискам, не должны быть равны между собой. Подобного рода соображения привели Луизова к выводу о том, что модель инерции (1) противоречит закону Плато [2, с. 47]. Тем не менее, мы не склонны отказываться от принятой модели инерции зрения при объяснении закономерностей, связанных с понятием критической частоты мельканий.

Прежде всего, попытаемся ответить на вопрос, что произойдет, если выполнить условия (15) и (16), а вместо условия (19) потребовать равенство средних яркостей в момент слияния мельканий, то есть:

$$B'_{cp} = B''_{cp} \cdot \tag{20}$$

Теперь средние яркости зрительных картин одинаковы и следовательно, согласно модели, коэффициенты усиления в обоих случаях должны быть равными. Вместе с тем, должны быть, согласно модели, одинаковыми и критические частоты. Однако такой вывод представляется невероятным. Если критические частоты, согласно закону Плато, оказываются равными при различных средних уровнях яркости, то при одинаковых средних уровнях яркости мелькания с одинаковыми колебаниями яркости должны сливаться при заведомо различных частотах.

Несмотря на такое негативное априорное заключение, мы все же выполнили опыты по проверке этого следствия из модели. Опыты выполнялись на вертушке. Кружок вертушки имел два поля. Поля вертушки раскрашивались таким образом, чтобы выполнялись условия (15), (16) и (20). Параметр скважности  $\lambda$  в опытах изменялся в широких пределах от 1/32 до 31/32. Заданный закон изменения яркости во времени на полях сравнения фор-

мировался на основе использования обобщенного закона Галбота с помощью быстрой смены черных и белых полей, длительность которых находится в заданном соотношении. Пример раскраски полей вертушки показан на рис. 14.

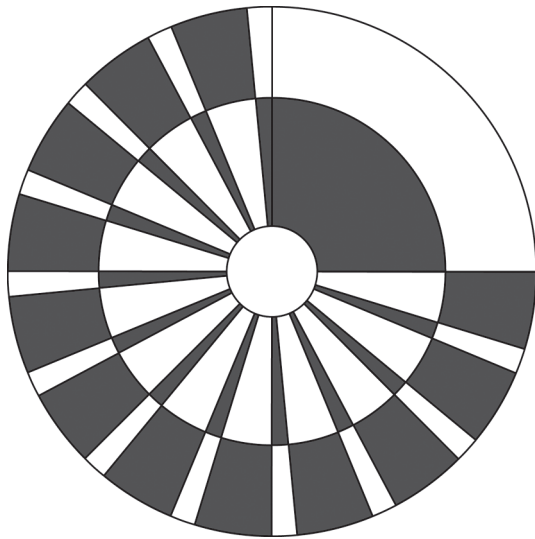


Рис. 14

Такому кружку соответствуют диаграммы изменения яркостей  $B'$  и  $B''$  внешнего и внутреннего полей, изображенные на рис. 15. В рассматриваемом примере принято:

$$B'_1 = B_6; B'_2 = \frac{1}{3}B_6 + \frac{2}{3}B_4; B'_3 = \frac{2}{3}B_6 + \frac{1}{3}B_4;$$

$$B''_2 = B_4; B''_{cp} = B'_{cp} = \frac{1}{2}B_6 + \frac{1}{2}B_4.$$

Символами  $B_6$  и  $B_4$  обозначены яркости белого и черного полей.

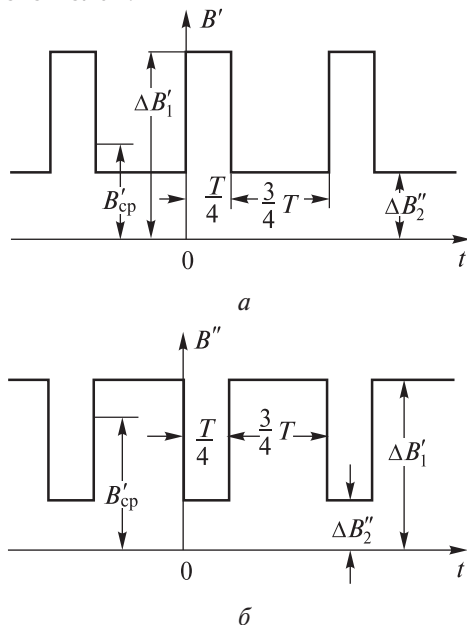


Рис. 15

Дискретность раскраски кружка не может существенно повлиять на результаты опыта. При разбивке окружности на 16 частей, как это дела-

лось при раскраске, вспомогательные быстрые мелькания будут сливаться уже при 2÷3 оборотах вертушки в секунду. Слияние же основных мельканий наступает при скорости вращения диска порядка 20÷50 об/сек. Проведенные опыты дали поразительный результат: критические частоты на обоих полях вертушки во всех опытах оказались строго одинаковыми. Вопреки ожиданиям, теоретический вывод из модели полностью подтвердился на опыте.

Однако, если это так, то отсюда логически вытекает невозможность строгого соблюдения закона Плато в его первоначальной формулировке. Критические частоты двух дополнительных дисков при  $\lambda \neq 1/2$ , строго говоря, должны отличаться между собой. И действительно, обращаясь к работе Айвса [4], мы находим у него вывод о том, что критические частоты для дополнительных дисков чрезвычайно близки друг другу. Не точно равны, но лишь чрезвычайно близки между собой! Различия критических частот для дополнительных дисков легко обнаружить, обращаясь непосредственно к экспериментам Айвса. Эти различия тем больше, чем больше параметр скважности мельканий отличается от значения 1/2. Так, в одном из опытов Айвса при  $\lambda' = 0,1$ ;  $\lambda'' = 0,9$  имеем:  $n'_{кр} = 43$  об/сек;  $n''_{кр} = 42$  об/сек. Как видим, критические частоты чрезвычайно близки, но, строго говоря, они различны, и это различие обнаруживается в эксперименте.

Различие критических частот дополнительных дисков можно легко продемонстрировать с помощью следующего опыта. Возьмем кружок с двумя соприкасающимися кольцами, раскраска которых соответствует дополнительным дискам при  $\lambda' = 7/8$  (внешнее кольцо),  $\lambda'' = 1/8$  (внутреннее кольцо). Схема такого кружка изображена на рис. 16.

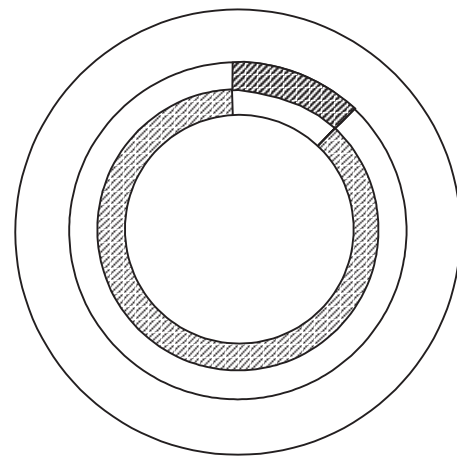


Рис. 16

Плавно меняя скорость вращения диска, можно легко подобрать такое ее значение, когда на внешнем кольце мелькания уже не замечаются, на внутреннем же кольце они еще отчетливо видны.

Итак, мы приходим к следующему выводу: первоначальная формулировка закона Плато неверна. Строго говоря, критические частоты дополнительных дисков не совпадают. В то же время имеется точное совпадение критических частот прямоугольных мельканий при формулировке закона Плато в виде условий (15), (16) и (20).

Попробуем теперь обобщить закон Плато. Рассмотрим две периодические функции яркости  $B'(t)$  и  $B''(t)$ , средние значения  $B_{cp}$  которых совпадают, и выполняется условие:

$$B'(t) - B_{cp} = B_{cp} - B''(t) \quad (21)$$

Пары функций  $B'(t)$  и  $B''(t)$ , удовлетворяющие этим условиям, назовем взаимно симметричными. Примером взаимно симметричных функций могут служить функции  $B'(t)$  и  $B''(t)$ , изображенные на рис. 17. Прямоугольные мелькания, удовлетворяющие условиям (15), (16) и (20), являются частным случаем взаимно симметричных функций.

Исходя из модели инерции зрения (1), нетрудно показать, что мелькания яркости в виде взаимно симметричных функций должны иметь одинаковые критические частоты. Этот вывод проверялся нами в опытах с вертушкой. На полях вертушки с помощью соответствующей раскраски формировались взаимно симметричные функции яркости произвольной формы. Во всех без исключения случаях наблюдалось совершенно точное совпадение критических частот мельканий полей вертушки.

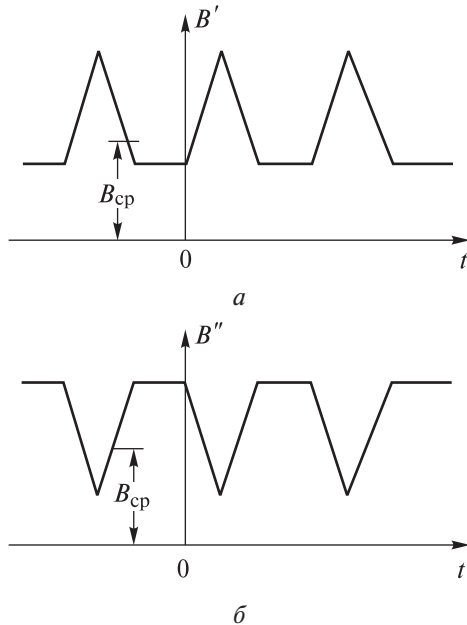


Рис. 17

Найденную закономерность можно рассматривать как обобщение закона Плато.

#### 4. Условие равенства критических частот для мельканий произвольной формы

Выше было установлено, что мелькания яркости в виде взаимно симметричных функций имеют

одинаковые критические частоты. В настоящем разделе, исходя из модели инерции зрения, мы получим несколько более общее условие равенства критических частот. Рассмотрим пару периодических мельканий произвольной формы  $B'(t)$  и  $B''(t)$ , средние значения яркости которых одинаковы. Поставим перед собой задачу установить условие, обеспечивающее равенство критических частот этих мельканий. Вначале изучим этот вопрос на примере прямоугольных мельканий. Рассмотрим пару прямоугольных мельканий, имеющих одинаковую среднюю яркость. Пусть величины колебания яркости и параметры скважности для этих мельканий соответственно равны  $\Delta B'$ ,  $\Delta B''$ , и  $\lambda'$ ,  $\lambda''$ . Формула (14) дает нам следующее условие равенства критических частот для этих мельканий:

$$\lambda'(1-\lambda')\Delta B' = \lambda''(1-\lambda'')\Delta B'' \quad (22)$$

Если модель инерции зрения (1) верна, то при соблюдении условия (22) критические частоты мельканий должны совпадать. Оговоримся, однако, что, поскольку формула (14) приближенная, совпадение частот будет достаточно точным лишь в случае, когда колебания яркостей  $\Delta B'$  и  $\Delta B''$  во много раз превосходят пороговое значение  $\Delta B_n$ . Заметим, что условие (22) весьма хорошо согласуется с частным случаем эмпирической формулы Портера — Айвса для прямоугольных мельканий, получаемым при условии равенства критических частот.

С целью проверки справедливости формулы (22) были проведены специальные опыты на диске Максвелла. Прямоугольные мелькания различной амплитуды и скважности, удовлетворяющие условию (22), формировались с помощью специальной раскраски на полях сравнения вертушки. Средние яркости мельканий принимались одинаковыми. На рис. 18 показан пример раскраски кружка при  $\lambda' = 1/2$ ,  $\lambda'' = 1/4$ ,  $\Delta B' = 3/4 \cdot \Delta B''$ .

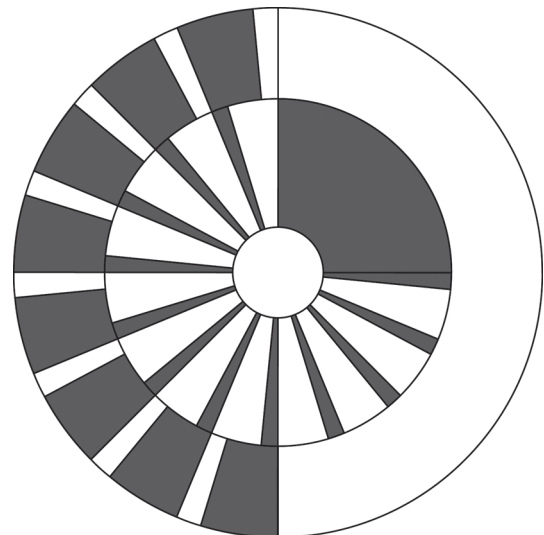


Рис. 18

Соответствующие этой раскраске диаграммы изменения яркостей полей сравнения даны на рис. 19. Эксперименты проводились при самых различных значениях параметров  $\lambda'$  и  $\lambda''$ , варьирувавшихся в пределах от 1/16 до 15/16. Во всех без исключения случаях наблюдалось довольно точное совпадение критических частот (отличие не более 1 – 2 гц).

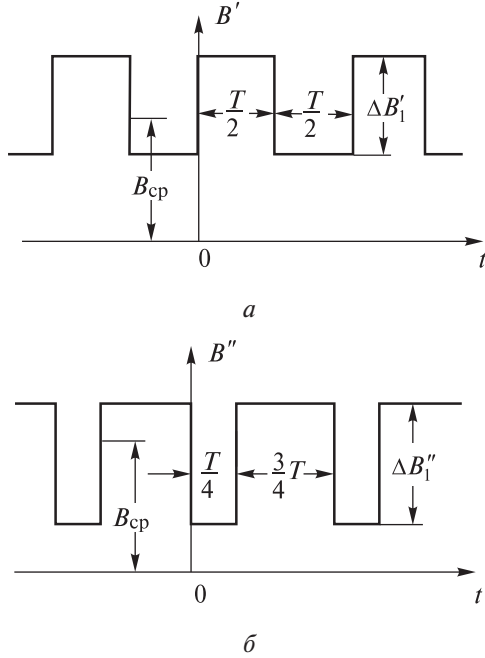


Рис. 19

Перейдем теперь к выводу условия равенства критических частот мельканий произвольной формы. Пусть имеется пара периодических мельканий произвольной формы  $B'(t)$  и  $B''(t)$ . Согласно формуле (59) из статьи [1] светлоты зрительных ощущений этих мельканий выразятся в виде:

$$S'(t) = \frac{k}{a} \cdot \frac{e^{-\frac{t}{a}}}{1 - e^{-\frac{T}{a}}} \cdot \int_{t-T}^t B'(\tau) e^{\frac{\tau}{a}} d\tau, \quad (23)$$

$$S''(t) = \frac{k}{a} \cdot \frac{e^{-\frac{t}{a}}}{1 - e^{-\frac{T}{a}}} \cdot \int_{t-T}^t B''(\tau) e^{\frac{\tau}{a}} d\tau.$$

Амплитуды колебаний светлоты  $\Delta S'$  и  $\Delta S''$  определяются формулами:

$$\Delta S' = \frac{k}{a} \cdot \frac{e^{-\frac{t}{a}}}{1 - e^{-\frac{T}{a}}} \left[ \max \int_{t-T}^t B'(\tau) e^{\frac{\tau}{a}} d\tau - \min \int_{t-T}^t B'(\tau) e^{\frac{\tau}{a}} d\tau \right],$$

$$\Delta S'' = \frac{k}{a} \cdot \frac{e^{-\frac{t}{a}}}{1 - e^{-\frac{T}{a}}} \left[ \max \int_{t-T}^t B''(\tau) e^{\frac{\tau}{a}} d\tau - \min \int_{t-T}^t B''(\tau) e^{\frac{\tau}{a}} d\tau \right]. \quad (24)$$

Приравнивая колебания светлоты  $\Delta S'$  и  $\Delta S''$ , получаем следующее условие, обеспечивающее равенство критических частот мельканий:

$$\max \int_{t-T}^t B'(\tau) e^{\frac{\tau}{a}} d\tau - \min \int_{t-T}^t B'(\tau) e^{\frac{\tau}{a}} d\tau =$$

$$= \max \int_{t-T}^t B''(\tau) e^{\frac{\tau}{a}} d\tau - \min \int_{t-T}^t B''(\tau) e^{\frac{\tau}{a}} d\tau. \quad (25)$$

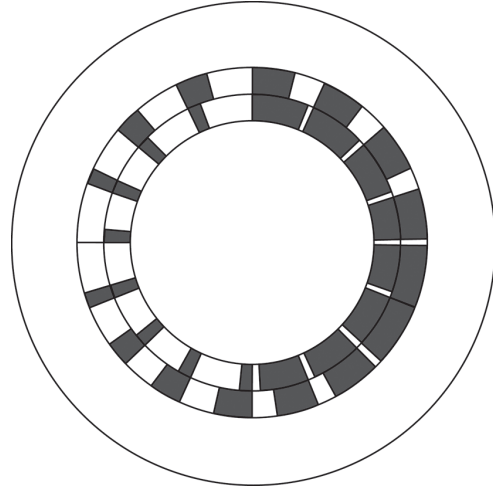


Рис. 20

Использование этой формулы в случае пары прямоугольных ( $\lambda=1/2$ ) и синусоидальных мельканий приводит к следующему условию, обеспечивающему равенство критических частот:

$$\Delta B_{\text{пр}} = \frac{2}{\pi} \Delta B_{\text{син}}, \quad (26)$$

где  $\Delta B_{\text{син}}$  и  $\Delta B_{\text{пр}}$  – амплитуды соответственно синусоидальных и прямоугольных мельканий яркости. При выводе формулы (26) принято допущение, что амплитуда мельканий во много раз превосходит ее пороговое значение. Справедливость формулы (26) проверялась в эксперименте с помощью диска Максвелла. Пример раскраски кружка дан на рис. 20. Соответствующие диаграммы колебания яркости мельканий представлены на рис. 21.

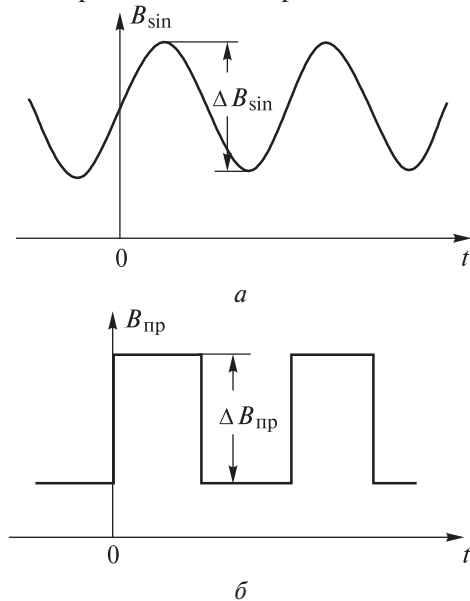


Рис. 21

Подобные условия были выведены также для различных, образованных треугольными, пилообразными, трапецевидными и другими видами мельканий. Эти условия проверялись в эксперименте. Во всех случаях получено довольно точное совпадение критических частот мельканий (отличие в критических частотах не превышало 5 гц).

### Выводы

На основе математической модели инерции зрения Аллара-Луизова исследованы реакции на прямоугольные периодические мелькания яркости и выведены формулы для определения критической частоты слияния мельканий.

С помощью модели установлено, что в своей первоначальной формулировке закон Плато строго не выполняется на опыте. Предложена видоизмененная формулировка закона Плато. Выполнены эксперименты, демонстрирующие точное выполнение в опыте видоизмененной формулировки закона Плато. Предложено обобщение закона Плато на случай произвольных периодических мельканий. Поставлены опыты, подтверждающие точное выполнение обобщенно закона Плато в психофизическом эксперименте.

Выведено условие равенства критических частот для прямоугольных мельканий различной скважности. Выполнены эксперименты, демонстрирующие довольно точное совпадение критических частот, полученных в эксперименте и с помощью расчета. Исходя из модели, выведено условие равенства критических частот для мельканий произвольной формы. Опытная проверка этого условия, выполненная для различных пар прямоугольных, синусоидальных, треугольных и других видов мельканий, подтвердила довольно точное согласование эксперимента и теории.

**Список литературы:** 1. Бондаренко М.Ф., Шабанов-Кушнаренко С.Ю., Шабанов-Кушнаренко Ю.П. Модель сглаживания в зрении // Бионика интеллекта: науч.-техн. журнал. — 2007. — № 1. — С. 3–14. 2. Луизов А.В. Инерция зрения. — М.: Оборонгиз, 1961. 3. Porter T.C. Contributions to the study of «flicker» // Proc. Roy. Soc., 1902, v. 70. 4. Jves H.E. A theory of intermittent vision // JOSA, 1922, № 6. 5. Plateau M. Sur une loi de la persistance des impression dans l'oeil // Bullet de l'Academie Royale de Belgique, serie 2, 1878, № 46.

*Поступила в редколлегию 17.01.2008*

УДК 519.7



## МОДЕЛЬ ИРРАДИАЦИИ ЗРЕНИЯ

М.Ф. Бондаренко<sup>1</sup>, С.Ю. Шабанов-Кушнарченко<sup>2</sup>,  
Ю.П. Шабанов-Кушнарченко<sup>3</sup>

<sup>1, 2, 3</sup> ХНУРЭ, г. Харьков, Украина

Исследуется модель иррадиации зрения и ее реакции на различные виды картин. Представлены диаграммы, характеризующие распределение яркости и светлоты зрительного ощущения в поле зрения. Сформулировано условие, обеспечивающее равенство критической густоты периодических полос произвольной формы.

### МОДЕЛЬ ИРРАДИАЦИИ ЗРЕНИЯ, КРИТИЧЕСКАЯ ГУСТОТА ПОЛОС, НУЛЕВОЙ ПРИБОР

#### Введение

В настоящей статье мы рассмотрим некоторые следствия, вытекающие из математической модели иррадиации зрения

$$S(x, y) = \frac{k}{2\pi b^2} \int \int_{-\infty}^{\infty} B(\xi, \eta) K_0 \left( \frac{\sqrt{(\xi-x)^2 + (\eta-y)^2}}{b} \right) d\xi d\eta. \quad (1)$$

предложенной в работе [1]. Следствия эти будут подвергнуты экспериментальной проверке. Определим, исходя из модели, как будет выглядеть вертикальная полоса, имеющая ширину  $\Delta x$ . Будем считать, что ось полосы проходит через точку фиксации. Яркость фона равна  $B_0$ , яркость полосы отличается от яркости фона на величину  $\Delta B$ . Яркость описанной зрительной картины может быть выражена в следующем виде:

$$B(x) = \begin{cases} B_0, & \text{если } x \leq -\frac{\Delta x}{2}, \\ B_0 + \Delta B, & \text{если } -\frac{\Delta x}{2} < x \leq \frac{\Delta x}{2}, \\ B_0, & \text{если } x > \frac{\Delta x}{2}. \end{cases} \quad (2)$$

На рис. 1 а показан ход изменения яркости зрительной картины в зависимости от координаты  $x$ . Произведя в правой части формулы (1) интегрирование с учетом выражения (2), получим следующее равенство, определяющее светлоту полосы в функции координаты  $x$ :

$$S(x) = \begin{cases} k B_0 + k \Delta B \operatorname{sh} \frac{\Delta x}{2b} e^{\frac{x}{b}}, & \text{если } x \leq -\frac{\Delta x}{2}; \\ k B_0 + k \Delta B (1 - e^{-\frac{\Delta x}{2b} \operatorname{ch} \frac{x}{b}}), & \text{если } -\frac{\Delta x}{2} < x \leq \frac{\Delta x}{2}; \\ k B_0 + k \Delta B \operatorname{sh} \frac{\Delta x}{2b} e^{-\frac{x}{b}}, & \text{если } x > \frac{\Delta x}{2}. \end{cases} \quad (3)$$

Диаграмма изменения светлоты полосы представлена на рис. 1. Из диаграммы видно, что уровень светлоты достигает максимального значения при  $x = 0$ , то есть на оси полосы. При удалении от

оси полосы в ту или иную сторону, светлота постепенно снижается, стремясь к значению  $B_0$ .

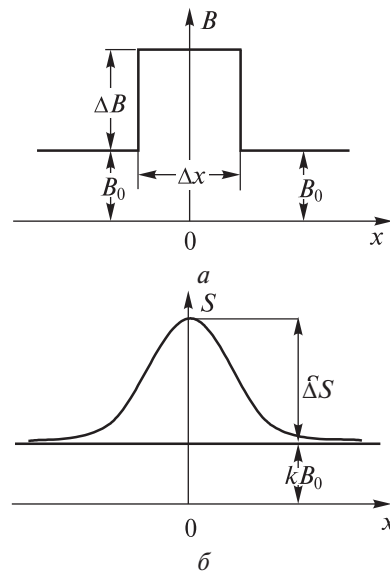


Рис. 1

Определим максимальное приращение светлоты зрительного ощущения полосы, для чего подставим во второе из уравнений (3) значение  $x=0$ :

$$\Delta S = k \Delta B (1 - e^{-\frac{\Delta x}{2b}}). \quad (4)$$

Формула (4) показывает, что максимальное приращение светлоты  $\Delta S$  зрительного ощущения полосы зависит от ширины  $\Delta x$  полосы, возрастающая при ее увеличении. На диаграмме (рис. 2) показан характер этой зависимости.

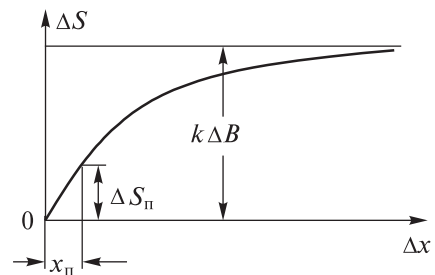


Рис. 2

Из диаграммы видно, что при неограниченном росте ширины полосы максимальное приращение

светлоты зрительного ощущения полосы стремится к нулю. Уменьшая ширину полосы, всегда можно прийти к такому положению, когда максимальное приращение светлоты ощущения полосы сравняется с пороговым значением  $\Delta S_{\text{п}}$ . Ширину полосы, при которой максимальное приращение светлоты имеет пороговое значение, назовем пороговой шириной полосы и обозначим ее через  $\Delta x_{\text{п}}$ . Определим пороговую ширину полосы по формуле (4), подставляя в нее вместо  $\Delta S$  величину  $\Delta S_{\text{п}} = k\Delta B_{\text{п}}$ , а вместо  $\Delta x$  — величину  $\Delta x_{\text{п}}$ :

$$\Delta x_{\text{п}} = 2b \ln \frac{1}{1 - \frac{\Delta B_{\text{п}}}{\Delta B}}. \quad (5)$$

Полученное соотношение показывает, что пороговая ширина полосы  $\Delta x_{\text{п}}$  зависит от величины яркости полосы  $\Delta x_{\text{п}}$ . Характер этой зависимости показан на диаграмме (рис. 3). Из диаграммы видно, что при повышении яркости полосы пороговая ширина полосы  $\Delta x_{\text{п}}$  снижается и при неограниченном росте  $\Delta B$  стремится к нулю. При уменьшении яркости полосы ее пороговая ширина растет и при значении  $\Delta B = \Delta B_{\text{п}}$  обращается в бесконечность. Величина  $\Delta B_{\text{п}}$  имеет смысл яркостного порога различения.

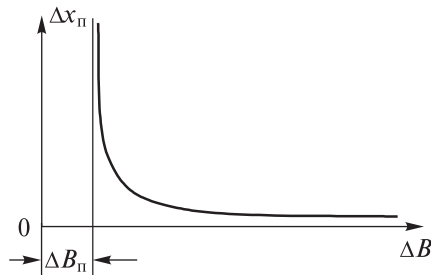


Рис. 3

Заметим, что зависимость (5) допускает опытную проверку.

Определим теперь, исходя из модели, как будет выглядеть кружок, имеющий диаметр  $D$ . Будем считать, что центр кружка совпадает с точкой фиксации. Яркость фона равна  $B_0$ , яркость кружка отличается от яркости фона на величину  $\Delta B$ . В этом случае яркость зрительной картины может быть выражена в виде следующих условий:

$$B(r) = \begin{cases} B_0 + \Delta B, & \text{если } r \leq \frac{D}{2}, \\ B_0, & \text{если } r > \frac{D}{2}, \end{cases} \quad (6)$$

где  $r$  — радиус полярной системы координат. Решая уравнение (1) при условии (6), получим следующее выражение для светлоты зрительного ощущения кружка:

$$S(r) = \begin{cases} kB_0 + k\Delta B \left[ 1 - \frac{D}{2b} K_1 \left( \frac{D}{2b} \right) J_0 \left( \frac{r}{b} \right) \right], & \text{если } r \leq \frac{D}{2}, \\ kB_0 + k\Delta B \frac{D}{2b} J_1 \left( \frac{D}{2b} \right) K_0 \left( \frac{r}{b} \right), & \text{если } r > \frac{D}{2}, \end{cases} \quad (7)$$

где  $K_0, K_1, J_0, J_1$  — бесселевы функции.

На рис. 4 а и б показаны диаграммы изменения яркости кружка и светлоты его зрительного ощущения в зависимости от радиуса  $r$ , построенные по формулам (6) и (7). Из диаграмм видно, что уровень светлоты достигает максимального значения при  $r = 0$ , то есть в центре кружка. При удалении от центра кружка светлота постепенно снижается, стремясь к значению  $B_0$ .

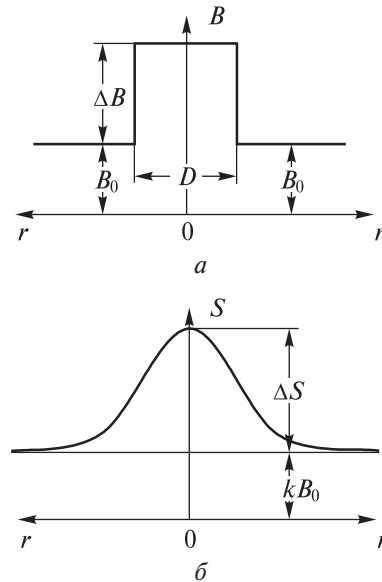


Рис. 4

Определим максимальное приращение светлоты кружка, для чего подставим в первое из уравнений (7) значение  $r = 0$ :

$$\Delta S = k\Delta B \left[ 1 - \frac{D}{2b} K_1 \left( \frac{D}{2b} \right) \right]. \quad (8)$$

Формула (8) показывает, что максимальное приращение светлоты  $\Delta S$  зависит от диаметра кружка  $D$ , возрастаая при его увеличении. На диаграмме (рис. 5) показан характер этой зависимости.

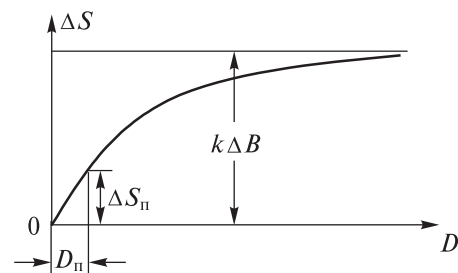


Рис. 5

Из диаграммы видно, что, уменьшая диаметр кружка, можно снизить максимальное приращение светлоты  $\Delta S$  к пороговому значению  $\Delta S_{\text{п}}$ . Диаметр кружка, при котором максимальное приращение светлоты кружка имеет пороговое значение, назовем пороговым диаметром кружка и обозначим его через  $D_{\text{п}}$ . Пороговый диаметр кружка можно определить из выражения

$$\frac{\Delta B_{\text{п}}}{\Delta B} = 1 - \frac{D_{\text{п}}}{2b} K_1 \left( \frac{D_{\text{п}}}{2b} \right), \quad (9)$$

вытекающего из формулы (8) при подстановке в нее вместо  $\Delta S$  величины  $k\Delta B_{\Pi}$  и вместо  $D$  – величины  $D_{\Pi}$ . Полученная зависимость допускает опытную проверку. Опытная проверка формул (5) и (9) описывается в разделе 3.

**1. Реакция модели иррадиации на серию прямоугольных полос**

Рассмотрим теперь реакцию модели на серию вертикальных черно-белых полос. Белые полосы имеют яркость  $B_1$ , черные –  $B_2$ . Яркость  $B(x)$  для этого случая запишется в виде:

$$B(x) = \begin{cases} B_1, & \text{если } nX < x \leq (n+\lambda)X, \\ B_2, & \text{если } (n+\lambda)X < x \leq (n+1)X. \end{cases} \quad (10)$$

На диаграмме (рис. 6 а) показан характер изменения яркости зрительной картины в зависимости от координаты  $x$  при ее задании условиями (10). Параметр  $\lambda$  определяет ширину белой полосы, равную  $\lambda X$ ; ширина черной полосы равна  $(1-\lambda)X$ . Величина  $\lambda$  назначается в пределах от 0 до 1.

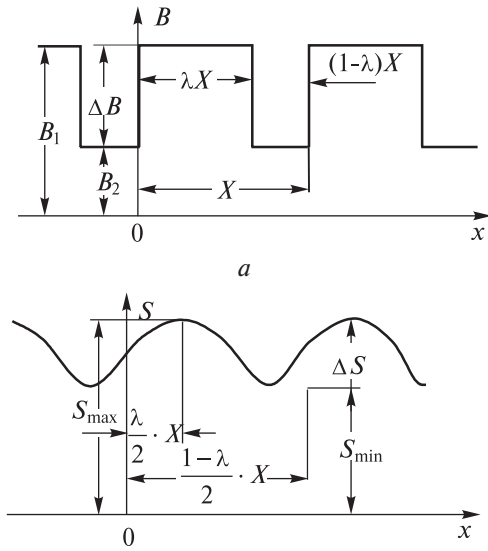


Рис. 6

Пользуясь формулой (1) для условий (10), можно получить следующее выражение для определения светлоты ощущения картины в виде серии полос

$$S(x) = \begin{cases} k B_1 - \frac{sh \frac{X}{2b} (1-\lambda)}{sh \frac{X}{2b}} \cdot ch \frac{2(x-nX) - \lambda X}{2b} k \Delta B, & \text{если } nX < x \leq (n+\lambda)X, \\ k B_2 + \frac{sh \frac{X}{2b} \lambda}{sh \frac{X}{2b}} \cdot ch \frac{2(x-nX) - (1-\lambda)X}{2b} k \Delta B, & \text{если } (n+\lambda)X < x \leq (n+1)X, \end{cases} \quad (11)$$

где

$$\Delta B = B_1 - B_2. \quad (12)$$

На рис. 6 б представлена диаграмма изменения светлоты серии полос в зависимости от координаты  $x$ , построенная по формулам (11). Из диаграммы видно, что светлота колеблется между минимальным  $S_{\min}$  и максимальным  $S_{\max}$  значениями, достигаемыми соответственно на оси белой и черной полосы при значении координаты  $x$  равном  $(n+\lambda/2)X$  и  $(n+(1+\lambda)/2)X$ . По формулам (11) находим:

$$S_{\max} = k B_1 - \frac{sh \frac{X}{2b} (1-\lambda)}{sh \frac{X}{2b}} k \Delta B, \quad (13)$$

$$S_{\min} = k B_2 + \frac{sh \frac{X}{2b} \lambda}{sh \frac{X}{2b}} k \Delta B.$$

Величина колебания светлоты равна:

$$\Delta S = S_{\max} - S_{\min} = 2k \Delta B \frac{sh \frac{\lambda}{4b} \cdot sh \frac{1-\lambda}{4b} X}{ch \frac{X}{4b}}. \quad (14)$$

Формула (14) показывает, что величина колебания светлоты  $\Delta S$  картины в виде серии полос зависит от значения параметра  $\lambda$ . Характер этой зависимости показан на диаграмме (рис. 7).

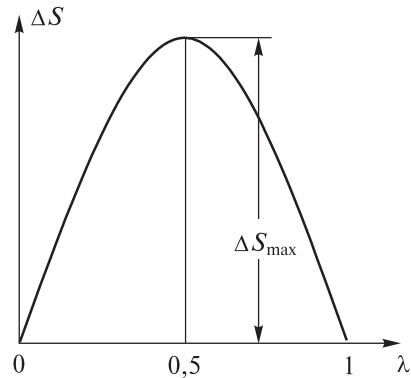


Рис. 7

Как видно из диаграммы, колебание светлоты достигает максимального значения  $\Delta S_{\max}$  при  $\lambda=1/2$ , то есть когда белая и черная полосы имеют одинаковую ширину:

$$\Delta S_{\max} = k \Delta B \cdot th \frac{X}{4b} \cdot th \frac{X}{8b}. \quad (15)$$

Согласно формуле (15) величина колебания светлоты  $\Delta S_{\max}$  зависит от ширины периода  $X$ . На диаграмме рис. 8 показан характер этой зависимости. Из диаграммы видно, что при неограниченном росте ширины периода  $X$  колебание светлоты зрительной картины стремится к значению  $\Delta S=k\Delta B$ . При уменьшении ширины периода  $X$  до нуля величина  $\Delta S$  также стремится к нулю.

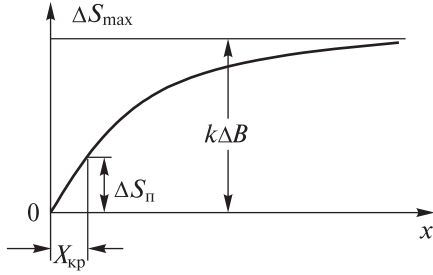


Рис. 8

Уменьшая ширину периода, то есть увеличивая густоту полос, можно снизить величину колебания светлоты  $\Delta S$  до порога различения  $\Delta S_{\text{п}}$ . Ширину периода, при которой колебание светлоты имеет пороговое значение, назовем критической шириной периода и обозначим ее через  $X_{\text{кр}}$ . Будем также пользоваться понятием критической густоты полос, равной  $m_{\text{кр}}=1/X_{\text{кр}}$ . Колебания яркости полос не будут обнаруживаться человеком, если их густота превысит критическую величину, при этом произойдет слияние полос. Найдем выражение для определения критической густоты полос  $m_{\text{кр}}$  в зависимости от величины колебания яркости полос  $\Delta B$  для случая, когда  $\lambda=1/2$ . С этой целью подставим в формулу (15) вместо  $\Delta S_{\text{max}}$  величину  $\Delta S_{\text{п}}=k\Delta B$ , а вместо  $X$  – величину  $1/m_{\text{кр}}$ . В результате получим:

$$m_{\text{кр}} = \frac{1}{4b \left| \text{Arch} \frac{1}{1 - \frac{\Delta B_{\text{п}}}{\Delta B}} \right|}. \quad (16)$$

На рис. 9 показано, как изменяется критическая густота полос  $m_{\text{кр}}$  в зависимости от колебания яркости  $\Delta B$ . Значению  $m_{\text{кр}}=0$  соответствует величина колебания яркости  $\Delta B_{\text{п}}$ . Формула (16) допускает экспериментальную проверку. При достаточно больших по сравнению с  $\Delta B_{\text{п}}$  значениях  $\Delta B$ , зависимость (16), как это видно из диаграммы (рис. 9), приобретает практически параболический характер:

$$m_{\text{кр}} = \frac{\sqrt{\frac{\Delta B}{\Delta B_{\text{п}}}}}{4b\sqrt{2}}. \quad (17)$$

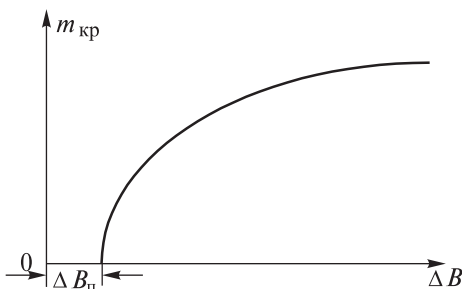


Рис. 9

Рассмотрим теперь зависимость критической густоты полос  $m_{\text{кр}}$  от параметра  $\lambda$ , то есть от соотношения ширины белой и черной полос. С этой целью определим из уравнения (14) величину  $\lambda$ , предварительно заменив в нем  $\Delta S$  на  $\Delta S_{\text{п}}=k\Delta B_{\text{п}}$  и  $X$  на  $1/m_{\text{кр}}$ . В результате получим:

$$\lambda = \frac{1}{2} + 2bm_{\text{кр}} \text{Arch} \left[ \left( 1 - \frac{\Delta B_{\text{п}}}{\Delta B} \right) \text{ch} \frac{1}{4bm_{\text{кр}}} \right]. \quad (18)$$

На рис. 10 в виде диаграммы показан характер зависимости  $m_{\text{кр}}$  от  $\lambda$ , определяемой формулой (18). Формула (18) допускает экспериментальную проверку.

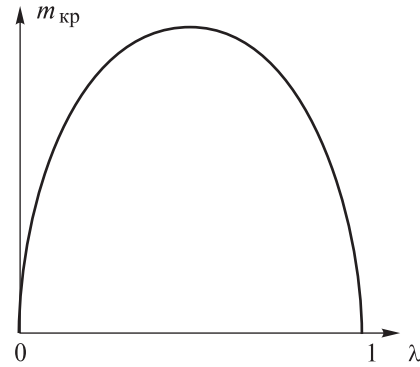


Рис. 10

При достаточно больших по сравнению  $\Delta B_{\text{п}}$  значениях  $\Delta B$  зависимость (18) принимает более простой вид:

$$m_{\text{кр}} \approx \frac{\sqrt{2\lambda(1-\lambda) \frac{\Delta B}{\Delta B_{\text{п}}}}}{4b}. \quad (19)$$

## 2. Эксперименты по проверке модели иррадиации

В двух предыдущих разделах были получены аналитические зависимости для описания условий пороговой видимости одиночной полосы, кружка и серии полос. Здесь описываются эксперименты, подтверждающие справедливость этих зависимостей при условии, что яркость полосы или кружка отличается от яркости фона на небольшую величину (до 10 пороговых значений). Первый опыт состоял в определении условий пороговой видимости узкой серой полосы на белом фоне. Ширина полосы  $\Delta x$ , выраженная в угловых единицах, и разность яркости фона и полосы  $\Delta B$  в процессе опыта регулировались. Цель опыта состояла в определении зависимости между пороговой шириной полосы  $\Delta x_{\text{п}}$  и разностью яркости фона и полосы  $\Delta B$ :

$$\Delta x_{\text{п}} = f(\Delta B). \quad (20)$$

Ниже описываются особенности методики проведения первого опыта. Эти особенности в равной степени относятся и к последующим опытам, из-

ложенным в этом разделе. Эксперимент осуществлялся с помощью диска Максвелла (вертушки), на котором устанавливались два белых бумажных кружка (из ватмана) диаметром 50 мм. Кружки имеют прорезы и вставлены друг в друга, как показано на рис. 11. На кружок 1 черной тушью наносится полоса в виде дуги окружности со средним радиусом 15 мм, толщина полосы  $\Delta x$  в опытах изменялась в пределах от 0,06 до 5 мм. Кружок 2 оставался белым. Поворотом кружка 1 относительно кружка 2 можно точно дозировать размер дуги видимого участка полосы. Измерение дуги осуществляется с помощью лимба 3, имеющего шкалу на 60 делений. Одно деление составляет 1:1280 часть окружности. Размер дуги устанавливается с точностью до одного деления лимба. В опытах размер дуги измерялся от 6 до 60 делений.

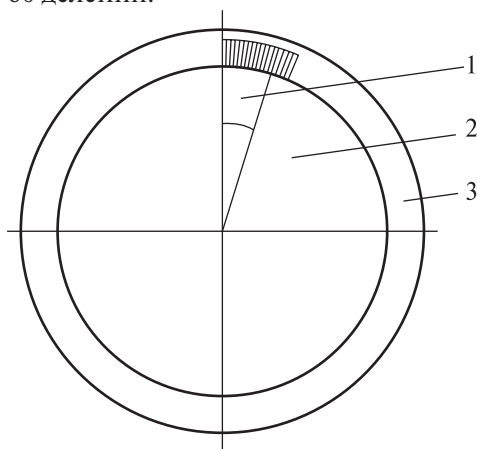


Рис. 11

При вращении вертушки с большой скоростью (порядка 50 об/сек) мы увидим, благодаря инерции зрения, вместо движущегося отрезка черной дуги серую линию в виде непрерывной окружности той же толщины  $\Delta x$ . Согласно обобщенному закону Талбота такая искусственным путем созданная зрительная картина тождественна по своему действию на глаз неподвижной серой окружности, вычерченной на белом фоне, яркость  $B$  которой равна:

$$B = \frac{\varphi B_ч + (2\pi - \varphi) B_б}{2\pi}, \quad (21)$$

где  $B_ч$  и  $B_б$  – яркость соответственно черной и белой поверхности;  $\varphi$  – угловой размер дуги, установленный на вертушке. Использование формулы (21) позволяет точно дозировать яркость полосы во время проведения опытов. Это дозирование осуществляется выбором углового размера дуги  $\varphi$ . Разность яркости белого фона и яркости полосы окружности равна:

$$\Delta B = B_б - B. \quad (22)$$

В связи с тем, что в опытах яркость полосы лишь незначительно отличалась от яркости белого фона,

существенной помехой для точного дозирования яркости полосы явилось просвечивание прикрываемого участка черной дуги сквозь белый кружок. Для борьбы с просвечиванием дуги пришлось увеличить толщину белого кружка, делая его двухслойным. Опыт не может считаться удовлетворительным, если при полностью прикрываемой дуге ( $\varphi=0$ ) глаз все же ее обнаруживает вследствие просвечивания сквозь белый кружок. Во время проведения опытов диск освещался лампой накапливания с вольфрамовой нитью напряжением 220 В мощностью 150 Вт с расстояния 3 м под углом  $45^\circ$ . Освещенность составила 130 лк. Для коэффициентов отражения белого и черного оттенков, использованных в опытах, получены значения соответственно 0,83 и 0,02. По этим данным с помощью формулы (21) определялись искомые коэффициенты отражения серой полосы. В опытах наибольшее отношение яркости белого фона  $B_б$  к яркости полосы  $B$  составляло менее 1,05, то есть яркость фона превышала яркость полосы на величину не более 5 %.

Яркость полос, использованных в опытах, сравнивалась с яркостью эталонного сплошного черного поля под микроскопом при стократном увеличении. К испытанию допускались лишь те полосы, для которых глаз не замечал отклонений яркости по сравнению с эталонным черным полем. Полосы, имеющие толщину 1 мм и более, сличались по яркости с эталонным полем непосредственно, без увеличения. Толщина полос также определялась с помощью микроскопа с точностью 0,005 мм, причем главным ограничением для точности замера являлось наличие неровностей краев у полосы. Поскольку толщину полосы при вычерчивании точно дозировать не представилось возможным, полоса требуемой толщины выбиралась из большого числа вычерченных полос разной толщины. Для полос с толщиной до 0,5 мм отклонения фактического размера толщины полосы от указанного в приведенных ниже таблицах не превышают 0,005 мм, для полос с толщиной свыше 0,5 мм – 0,01 мм.

Важным моментом при постановке опытов явились меры по тщательному центрированию кружка с нанесенной на нем дугой, так как малейший эксцентриситет или перекося дуги приводит к увеличению ее видимой ширины и искажению яркости. Другим важным фактором, от которого также зависел успех опыта, является правильный выбор пределов расстояния, с которого рассматривается полоса. Эти пределы должны выбираться с таким расчетом, чтобы не могли сказаться оптические несовершенства глаза (близорукость или дальновидность). В описываемых опытах расстояние наблюдения в первом опыте составляло постоянную величину, равную 500 мм, а в последующих опытах изменялось в пределах 210÷540 мм.

Опыты осуществлялись на одном наблюдателе, замеры многократно повторялись в течение месяца в различное время суток. Определялись пороги как на появление, так и на исчезновение полосы. В нижеприведенных таблицах указаны средние значения из 8-12 замеров. Разброс результатов в отдельных замерах, как правило, не выходил за пределы 10%. Результаты опытов представлены в табл. 1.

Таблица 1

Пороговая ширина полосы $h$ , мм	0,06	0,07	0,09	0,12	0,16	0,25	0,5	0,16	0,9	1,8	5,0
Число делений по шкале $n$	57	49	39	30	23	16	10	23	7	6	6

В первой строке таблицы указана ширина полосы, предъявляемой для рассматривания, во второй строке указано число делений по шкале, соответствующее размеру дуги, при котором имеет место пороговая видимость полосы.

Согласно формулам (21) и (22) разность яркости белого фона и полосы определится зависимостью:

$$\Delta B = \frac{n}{1280} (B_6 - B_4). \quad (23)$$

Введем пороговую разность яркостей  $\Delta B_{пн}$ , соответствующую 6 делениям по шкале  $\Delta B = \frac{n}{1280} (B_6 - B_4)$ , и отношение  $\frac{\Delta B}{\Delta B_{пн}} = \frac{n}{6}$ . Пороговая ширина полосы  $\Delta x_{пн}$  в угловых единицах (радианах) определится зависимостью  $\Delta x_{пн} = h/H$ , где  $h$  – пороговая ширина полосы в линейных единицах,  $H=500$  мм, расстояние, с которого велось наблюдение полосы.

На рис. 12 в координатах  $\Delta B/\Delta B_{пн}$  и  $1/\Delta x_{пн}$  в виде точек представлены результаты опытов по данным табл. 1. На этой же диаграмме нанесена кривая, построенная по формуле (5):

$$\Delta x_{пн} = 2b \ln \frac{1}{1 - \frac{\Delta B_{пн}}{\Delta B}}$$

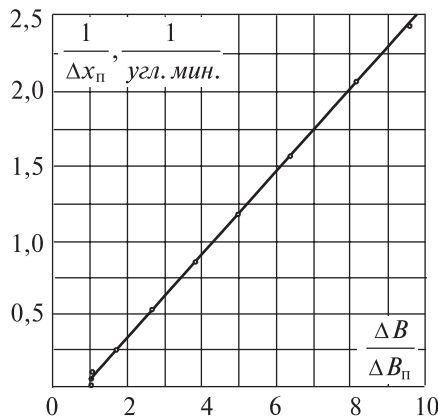


Рис. 12

При построении теоретической кривой по формуле (5) постоянная иррадиации зрения принята равной  $B=1,85'$ . Как видим, теоретическая кривая хорошо соответствует экспериментальным точкам. Отклонения экспериментальных данных от теоретических находятся в пределах точности постановки опытов. Необходимо заметить, что учет сравнительно небольшой кривизны линии, имевшей место в опыте, при теоретическом расчете не вносит заметной разницы в результаты вычислений по сравнению с формулой (5), при выводе которой полоса предполагалась прямой.

Второй опыт заключался в определении условий слияния серии из 12 серых концентрических полос. Яркость полос, так же как и в первом опыте, дозировалась с помощью вертушки изменением углового размера видимого участка серии полос. Ширина белого промежутка между соседними полосами равнялась ширине одной серой полосы. Все полосы одинаковы по ширине. Результаты опытов представлены в табл. 2.

Таблица 2

Пороговая ширина периода (белой и серой полос) $h$ , мм	0,55	0,55	0,55	0,55	0,55	1,2	1,2	1,2	2,0
Число делений по шкале $n$	60	51	42	33	24	16	10	7	6
Расстояние наблюдения $H$ , мм	540	500	450	390	320	540	350	340	500

На рис. 13 по данным табл. 2 точками нанесена экспериментальная зависимость между критической плотностью полос  $m_{кр} = H/h$  и величиной  $\Delta B/\Delta B_{пн}$ , равной, как и в предыдущем опыте,  $\Delta B/\Delta B_{пн} = n/6$ . На точки наложена теоретическая кривая, построенная по формуле (16):

$$m_{кр} = \frac{1}{4b \left| \text{Arch} \frac{1}{1 - \frac{\Delta B_{пн}}{\Delta B}} \right|}$$

Постоянная иррадиации зрения принята равной  $B=1,85'$ , то есть такой же, как и в случае с одиночной полосой. Сравнение теоретических и опытных данных указывает на хорошее согласование теории с опытом. Следует заметить, что при выводе формулы (16) предполагалось, что полосы прямые, а их число бесконечно велико. Фактически же в опытах полосы несколько искривлены, а их число конечно. Однако, как показывают расчеты, теоретический учет этих факторов не вносит сколько-нибудь заметных поправок в ход теоретической кривой.

Третий опыт отличался от второго тем, что при сохранении постоянной ширины периода в размере 0,55 мм ширина серой полосы изменялась от 0,03 до 0,52 мм. Число делений по шкале во всех опытах было постоянным и равным 60. В табл. 3 указаны результаты опытов.

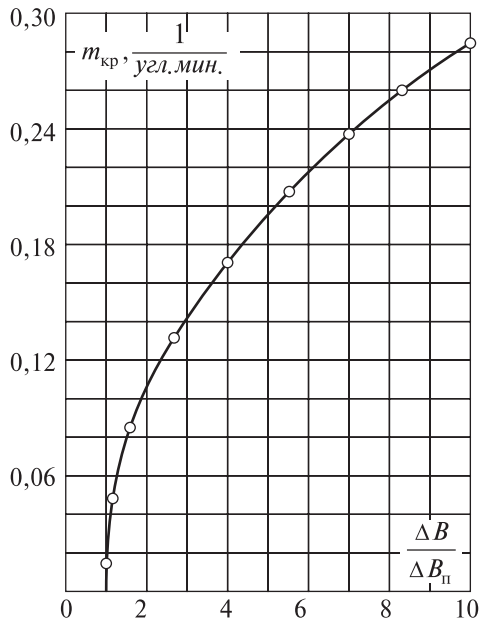


Рис. 13

Таблица 3

Пороговая ширина серой полосы $h$ , мм	0,03	0,06	0,10	0,16	0,27	0,39	0,45	0,49	0,52
Расстояние наблюдения $H$ , мм	210	310	420	500	540	490	410	330	200

На диаграмме (рис. 14) результаты опыта изображены графически в координатах  $m_{кр} = H/0,55$  и  $\lambda = h/0,55$ , где  $H$  и  $h$  необходимо подставлять в мм. Теоретическая диаграмма строилась по формуле (18):

$$\lambda = \frac{1}{2} + 2bm_{кр} \text{Arch} \left[ \left( 1 - \frac{\Delta B_{п}}{\Delta B} \right) ch \frac{1}{4bm_{кр}} \right].$$

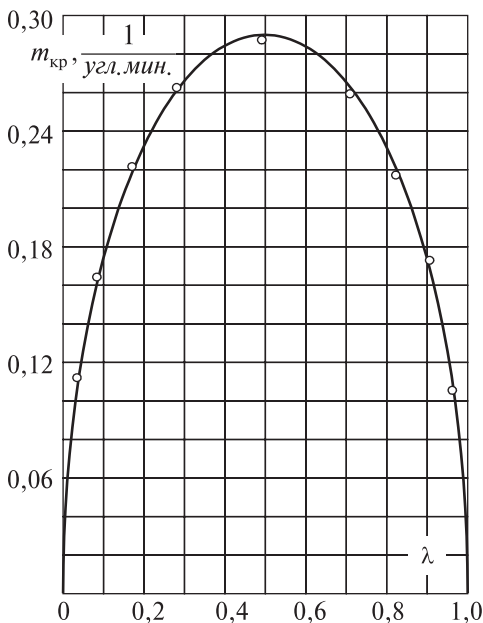


Рис. 14

При построении кривой принято прежнее значение постоянной иррадиации зрения  $b=1,85'$ .

Кроме того, в согласии с предыдущими опытами принято  $\Delta B/\Delta B_{п}=10$ , что соответствует 60 делениям по шкале вертушки. Как видим, и в этом случае наблюдается удовлетворительное согласование теории с опытом. Итак, в трех разнородных опытах получено хорошее согласование с теоретически выведенными модельными реакциями. Это обстоятельство служит веским доводом в пользу справедливости предложенной нами математической модели иррадиации зрения.

В литературе описаны опыты А.А. Смирнова по определению условий пороговой видимости светлого кружка на более темном фоне [2]. На рис. 15 в виде серии точек показаны результаты одного из этих опытов.

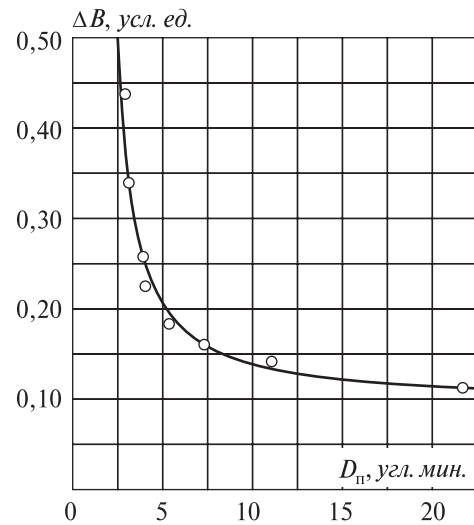


Рис. 15

На экспериментальные точки наложена теоретическая кривая, построенная по формулам (9), при значении постоянной иррадиации зрения  $b = 1,9'$ , то есть почти такой же, как и в наших опытах ( $b = 1,85'$ ). Как видим, теоретическая кривая хорошо соответствует опытным точкам.

### 3. Исследование совместного действия моделей инерции и иррадиации зрения

До сих пор рассматривались реакции модели инерции и иррадиации в тех случаях, когда зрительная картина являлась либо только функцией времени, либо только функцией координат поля зрения. В настоящем параграфе будут изучены некоторые реакции модели инерции и иррадиации зрения при подаче на ее вход таких зрительных картин, которые зависят сразу и от времени и от одной из координат поля зрения (координаты  $x$ ), то есть  $B=B(x, t)$ . В этом случае реакции модели будут описываться следующим интегральным выражением, являющимся частным случаем формулы (22) из работы [1]:

$$S(x, t) = \frac{k}{2b\sqrt{\pi}} \int_{-\infty}^t \frac{e^{-\frac{t-\tau}{a}}}{\sqrt{t-\tau}} dt \int_{-\infty}^{\infty} B(\xi, \tau) e^{-\frac{(\xi-x)^2}{4b^2 \frac{t-\tau}{a}}} d\xi. \quad (24)$$

Рассмотрим, как, согласно модели, будет выглядеть движущаяся в поле зрения полоса. Пусть в поле зрения на фоне с яркостью  $B_0$  с равномерной скоростью  $v$  движется вертикальная полоса. Яркость полосы отличается от яркости фона на величину  $\Delta B$ . Полоса имеет ширину  $\Delta x$ . В момент времени  $t=0$  ось полосы проходит через точку фиксации ( $x=0$ ). Функция яркости в этом случае запишется в виде:

$$B(x,t) = \begin{cases} B_0, & \text{если } x \leq vt - \Delta x/2, \\ B_0 + \Delta B, & \text{если } vt - \Delta x/2 < x \leq vt + \Delta x/2, \\ B_0, & \text{если } x > vt + \Delta x/2. \end{cases} \quad (25)$$

На рис. 16 а показано положение полосы в поле зрения в момент времени  $t$ . Рис. 16 б изображает изменение яркости  $B$  в функции координаты  $x$  в тот же момент времени.

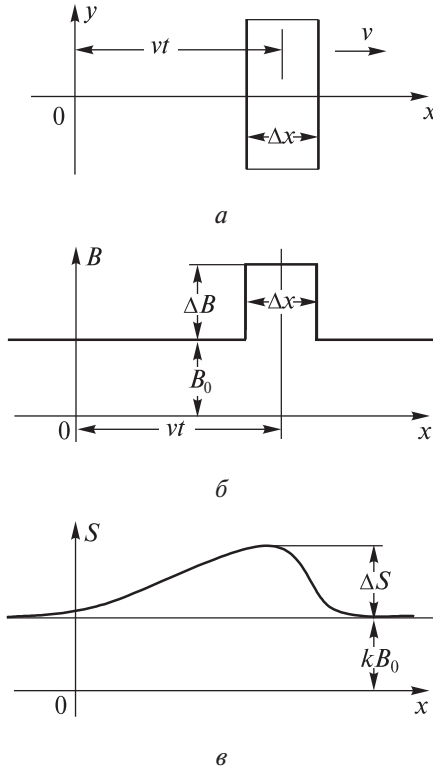


Рис. 16

Форму задания зрительной картины можно упростить, если перейти к системе координат, движущейся вместе с полосой:

$$x^* = x - vt. \quad (26)$$

В новой системе координат яркость  $B^*$  зрительной картины запишете в виде:

$$B^* = \begin{cases} B_0, & \text{если } x^* \leq -\frac{\Delta x}{2}, \\ B_0 + \Delta B, & \text{если } -\frac{\Delta x}{2} < x^* \leq \frac{\Delta x}{2}, \\ B_0, & \text{если } x^* > \frac{\Delta x}{2}. \end{cases} \quad (27)$$

Решая уравнение (24) при условии (27), получим следующее выражение для определения светлоты ощущения полосы в движущейся системе координат:

$$S^*(x^*) = \begin{cases} kB_0 - k\Delta B \frac{2r_1}{r_2 - r_1} sh \frac{r_2 \Delta x}{2} e^{r_2 x^*}, & \text{если } x^* \leq -\frac{\Delta x}{2}, \\ kB_0 + k\Delta B - k\Delta B \frac{r_2}{r_2 - r_1} e^{\frac{r_1 \Delta x}{2}} e^{r_1 x^*} + \\ + k\Delta B \frac{r_1}{r_2 - r_1} e^{-\frac{r_2 \Delta x}{2}} e^{r_2 x^*}, & \text{если } -\frac{\Delta x}{2} < x^* \leq \frac{\Delta x}{2}, \\ kB_0 - k\Delta B \frac{2r_2}{r_2 - r_1} sh \frac{r_1 \Delta x}{2} e^{r_1 x^*}, & \text{если } x^* > \frac{\Delta x}{2}, \end{cases} \quad (28)$$

где

$$r_{1,2} = \frac{-av \pm \sqrt{a^2 v^2 + 4b^2}}{2b^2}. \quad (29)$$

Определение светлоты ощущения в неподвижной системе координат может быть выполнено по формуле:

$$S(x,t) = S^*(x - vt). \quad (30)$$

Диаграмма изменения светлоты ощущения полосы в зависимости от координаты  $x$ , построенная по формулам (28)–(30) для некоторого момента времени, представлена на рис. 16 в. Из диаграммы видно, что полоса видится размытой, ее светлота не остается постоянной и имеет максимум в некоторой точке, сдвинутой от оси в сторону, противоположную направлению движения полосы. Координату  $x_0^*$  точки с максимальной яркостью в подвижной системе координат найдем, дифференцируя по  $x^*$  второе из равенств (28) и приравнявая производную  $dS^*/dx^*$  нулю. В результате получим:

$$x_0^* = -\frac{r_1 - r_2}{r_2 - r_1} \cdot \frac{\Delta x}{2}. \quad (31)$$

Определим максимальное приращение светлоты  $\Delta S$  ощущения движущейся полосы. Для этого подставим во второе из равенств (28) вместо текущей координаты  $x^*$  ее значение  $x_0^*$ , при котором светлота достигнет максимума. В результате получим:

$$\Delta S = k\Delta B \left( 1 - e^{-\frac{\Delta x}{\sqrt{a^2 v^2 + 4b^2}}} \right). \quad (32)$$

Формула (32) показывает, что максимальное приращение светлоты  $\Delta S$  полосы зависит от скорости движения полосы  $v$ . Диаграмма на рис. 17 показывает характер этой зависимости. Из диаграммы видно, что при увеличении скорости движения полосы максимальное приращения светлоты  $\Delta S$  уменьшается, стремясь к нулю. При некоторой критической скорости  $v_{кр}$  движения полосы светлота достигает порогового значения  $\Delta S_{п}$  и полоса вовсе не будет обнаруживаться наблюдателем.

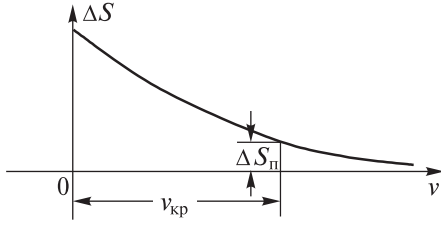


Рис. 17

Определим критическую скорость по формуле (32), подставляя в нее вместо  $\Delta S$  величину  $\Delta S = k\Delta B_{\text{п}}$ , а вместо  $v$  — величину  $v_{\text{кр}}$ :

$$v_{\text{кр}} = \frac{1}{a} \sqrt{\frac{\Delta x^2}{\ln^2\left(1 - \frac{\Delta B_{\text{п}}}{\Delta B}\right)} - 4b^2}. \quad (33)$$

Как видно из соотношения (33), критическая скорость увеличивается при увеличении ширины полосы  $\Delta x$  и уменьшается, стремясь к нулю, при уменьшении  $\Delta x$  до величины  $\Delta x_{\text{п}} = 2B \ln(1 - \Delta B_{\text{п}}/\Delta B)$ . Диаграмма зависимости  $v_{\text{кр}}$  от  $x$  показана на рис. 18 а. Зависимость  $v_{\text{кр}}$  от величины  $\Delta B$  носит аналогичный характер. При увеличении  $\Delta B$  величина  $v_{\text{кр}}$  растет, стремясь к бесконечности, а при уменьшении  $\Delta B$  — к величине  $\Delta B_{\text{п}}^* = \frac{\Delta B_{\text{п}}}{1 - e^{-\frac{\Delta x}{2b}}}$ ,  $v_{\text{кр}}$  убывает до

нуля (рис. 18 б). Формула (33) допускает экспериментальную проверку.

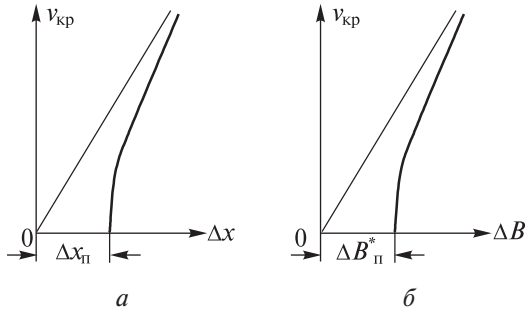


Рис. 18

Рассмотрим теперь реакцию модели на зрительную картину в виде серии движущихся полос. Пусть в поле зрения на фоне с яркостью  $B_1$  движется бесконечная серия полос. Яркость каждой полосы  $B_2$  отличается от яркости фона на величину  $\Delta B$ . Ширина каждой полосы равна  $\lambda X$ , расстояние (просвет) между соседними полосами  $(1-\lambda)X$ . Параметр  $\lambda$  может колебаться в пределах от 0 до 1. Функция яркости  $B^*(x, t)$  в этом случае запишется в виде (в движущейся системе координат):

$$B^*(x^*) = \begin{cases} B_1, & \text{если } X(n - \frac{\lambda}{2}) < x^* \leq X(n + \frac{\lambda}{2}), \\ B_2, & \text{если } X(n + \frac{\lambda}{2}) < x^* \leq X(n + 1 - \frac{\lambda}{2}). \end{cases} \quad (34)$$

На рис. 19 а представлена диаграмма изменения яркости  $B^*(x^*)$  зрительной картины, определяемой условиями (34).

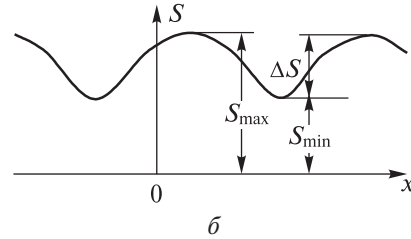
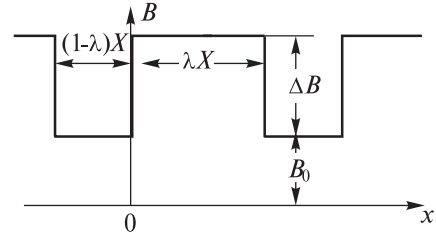


Рис. 19

Воспользовавшись линейностью уравнения (24), мы можем получить его решение в данном случае как бесконечную сумму слагаемых вида (28). После преобразований эту сумму можно представить в следующем виде:

$$S^*(x^*) = \begin{cases} kB_1 - k\Delta B \frac{r_2}{r_2 - r_1} e^{\eta_1(x^* - nX)} \cdot \frac{\text{sh} \frac{r_1(1-\lambda)X}{2}}{\text{sh} \frac{r_1X}{2}} + \\ + k\Delta B \frac{r_1}{r_2 - r_1} e^{\eta_2(x^* - nX)} \cdot \frac{\text{sh} \frac{r_2(1-\lambda)X}{2}}{\text{sh} \frac{r_2X}{2}}, \\ \text{если } X(n - \frac{\lambda}{2}) < x^* \leq X(n + \frac{\lambda}{2}), \\ kB_2 - k\Delta B \frac{r_2}{r_2 - r_1} e^{\eta_1[x^* - (n + \frac{\lambda}{2})X]} \cdot \frac{\text{sh} \frac{r_1\lambda X}{2}}{\text{sh} \frac{r_1X}{2}} - \\ - k\Delta B \frac{r_1}{r_2 - r_1} e^{\eta_2[x^* - (n + \frac{\lambda}{2})X]} \cdot \frac{\text{sh} \frac{r_2\lambda X}{2}}{\text{sh} \frac{r_2X}{2}}, \\ \text{если } X(n + \frac{\lambda}{2}) < x^* \leq X(n + 1 - \frac{\lambda}{2}). \end{cases} \quad (35)$$

Характер изменения светлоты  $S^*(x^*)$  серии движущихся полос показан на диаграмме (рис. 19 б). Дифференцируя соотношение (35) и приравнявая производные к нулю, найдем максимальное  $S_{\text{max}}$  и минимальное  $S_{\text{min}}$  значения светлоты:

$$S_{\text{max}} = kB_1 - k\Delta B \left( \frac{\text{sh} \frac{r_1(1-\lambda)X}{2}}{\text{sh} \frac{r_1X}{2}} \right)^{\frac{r_2}{r_2 - r_1}} \times \\ \times \left( \frac{\text{sh} \frac{r_2(1-\lambda)X}{2}}{\text{sh} \frac{r_2X}{2}} \right)^{\frac{\eta_1}{r_2 - r_1}}, \quad (36)$$

**Выводы**

Исследованы реакции модели иррадиации зрения на одиночную полосу и кружок. Получены формулы, определяющие условия порогового видения полосы и кружка, допускающие прямую экспериментальную проверку. Выполнены психофизические эксперименты по проверке найденных зависимостей. Получено удовлетворительное согласие теории с опытом. Исследованы реакции модели на серию прямоугольных полос различной скважности, получены зависимости для определения критической густоты слияния серии полос, осуществлена их экспериментальная проверка, получено согласие теории с опытом. Найдено численное значение постоянной иррадиации для человеческого зрения, равное  $b = 1,85'$ . Исследованы реакции модели инерции и иррадиации зрения на одиночную движущуюся полосу и на серию движущихся полос, получены зависимости, допускающие их прямую опытную проверку.

**Список литературы:** 1. Бондаренко М.Ф., Шабанов-Кушнаренко С.Ю., Шабанов-Кушнаренко Ю.П. Модель сглаживания в зрении // Бионика интеллекта: науч.-техн. журнал. — 2007. — № 1. — С. 33-47. 2. Смирнов А.А. Зависимость различительной чувствительности глаза от величины объектов // В сб. «Зрительные ощущения и восприятия». — М.-Л.: Соцэкгиз, 1935. Т. 2.

*Поступила в редколлегию 22.01.2008*

$$S_{\min} = kB_2 + k\Delta B \left( \frac{sh \frac{r_1 \lambda X}{2}}{sh \frac{r_1 X}{2}} \right)^{\frac{r_2}{r_2-r_1}} \cdot \left( \frac{sh \frac{r_2 \lambda X}{2}}{sh \frac{r_2 X}{2}} \right)^{-\frac{r_1}{r_2-r_1}}. \quad (37)$$

Полагая  $S_{\max} - S_{\min} = \Delta B_{\Pi}$ , найдем соотношение между критической скоростью  $v_{кр}$  движения полос, пороговой шириной  $\Delta x_{\Pi}$  и параметром  $\lambda$ :

$$\frac{\Delta B_{\Pi}}{\Delta B} = 1 - \left( \frac{sh \frac{r_1 (1-\lambda) X}{2}}{sh \frac{r_1 X}{2}} \right)^{\frac{r_2}{r_2-r_1}} \cdot \left( \frac{sh \frac{r_2 (1-\lambda) X}{2}}{sh \frac{r_2 X}{2}} \right)^{-\frac{r_1}{r_2-r_1}} - \left( \frac{sh \frac{r_1 \lambda X}{2}}{sh \frac{r_1 X}{2}} \right)^{\frac{r_2}{r_2-r_1}} \cdot \left( \frac{sh \frac{r_2 \lambda X}{2}}{sh \frac{r_2 X}{2}} \right)^{-\frac{r_1}{r_2-r_1}}. \quad (38)$$

Характер зависимости величины  $\Delta B$  от  $v_{кр}$ ,  $\lambda$  и  $x$  соответственно представлен на рис. 20а, б, в. Эти зависимости допускают опытную проверку.

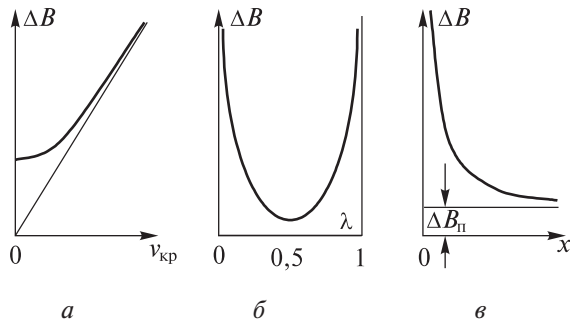


Рис. 20

УДК 519.7



## МОДЕЛЬ КРАЕВОГО КОНТРАСТА ЗРЕНИЯ

М.Ф. Бондаренко<sup>1</sup>, С.Ю. Шабанов-Кушнарченко<sup>2</sup>,  
Ю.П. Шабанов-Кушнарченко<sup>3</sup>

<sup>1, 2, 3</sup> ХНУРЭ, г. Харьков, Украина

Исследуется модель краевого контраста зрения и ее реакции на различные виды входных сигналов. Представлены диаграммы, характеризующие изменения яркости и светлоты зрительного ощущения в поле зрения. Экспериментально определены численные значения параметров модели краевого контраста зрения человека. Достигнуто удовлетворительное согласование теории с опытом.

### МОДЕЛИРОВАНИЕ, НУЛЕВОЙ ПРИБОР, КРАЕВОЙ КОНТРАСТ

#### 1. Постановка задачи

Известно, что орган зрения человека осуществляет подчеркивание границ, содержащихся в изображении, которое обнаруживается в виде явления краевого контраста [1, 2]. Это свойство зрения, несомненно, носит целесообразный характер, поскольку, благодаря наличию краевого контраста, облегчается обнаружение и последующий анализ контура изображения. В настоящей статье предлагается математическая модель обработки зрительной информации, лежащей в основе явления краевого контраста. Ограничиваясь ахроматическими зрительными картинками, рассматриваем их как функции яркости  $B = B(x, y)$  координат  $x, y$  поля зрения. Зрительное ощущение характеризуем светлотой  $S = S(x, y)$ .

Модель преобразования информации в органе зрения человека представим в виде схемы, изображенной на рис. 1. Функции  $P(x, y)$ ,  $Q(x, y)$  и  $T(x, y)$  обозначают некоторые промежуточные сигналы преобразования.

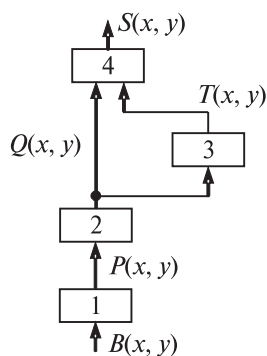


Рис. 1

Блок 1 осуществляет иррадиационное преобразование зрительной информации согласно уравнению:

$$P(x, y) = \frac{k}{2\pi b^2} \int \int_{-\infty}^{\infty} B(\xi, \eta) K_0 \left( \frac{\sqrt{(\xi-x)^2 + (\eta-y)^2}}{b} \right) d\xi d\eta. \quad (1)$$

Функционирование этого блока было рассмотрено в работах [5, 10].

Поскольку явления краевого контраста разыгрываются на значительно более крупных по площади участках поля зрения, чем явления иррадиации, в настоящей статье для упрощения задачи иррадиационные процессы вовсе не учитываются, а блок 1 рассматривается в качестве тождественного преобразователя:

$$P(x, y) = kB(x, y). \quad (2)$$

Блок 2 осуществляет логарифмирование сигнала  $P$ :

$$Q = c \lg dP, \quad (3)$$

где  $c$  и  $d$  – константы, численные значения и размерность которых зависят от выбора единицы измерения соответственно для светлоты и яркости.

Преобразование сигналов в блоке 3 описывается уравнением:

$$-n^2 \left( \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right) + T = k_k Q, \quad (4)$$

сходным с уравнением иррадиации зрения. Константу  $k_k$  назовем постоянной интенсивности краевого контраста, константу  $n$  – постоянной затухания краевого контраста. Блок 3 осуществляет вычитание сигналов:

$$S = Q - T. \quad (5)$$

Решение дифференциального уравнения (4) для случая безграничной двумерной зрительной картины имеет вид:

$$T(x, y) = \frac{k_k}{2\pi n^2} \int \int_{-\infty}^{\infty} Q(\xi, \eta) K_0 \left( \frac{\sqrt{(\xi-x)^2 + (\eta-y)^2}}{n} \right) d\xi d\eta, \quad (6)$$

где  $k_0$  – функция Макдональда [3];  $\xi$  и  $\eta$  – переменные интегрирования. В частном случае, когда зрительная картина неоднородна только в направлении координатной оси  $x$ , решение уравнения (4) запишется в виде:

$$T(x) = \frac{k_k}{2n} \int_{-\infty}^{\infty} Q(\xi) e^{-\frac{|\xi-x|}{n}} d\xi. \quad (7)$$

Для выяснения функционального смысла параметров  $k_k$  и  $n$  модели рассмотрим зрительную картину  $B(x)$ , содержащую вертикальную границу  $x=0$ . Левая половина поля зрения имеет яркость  $B_1$ , правая –  $B_2$ , то есть:

$$B(x) = \begin{cases} B_1, & \text{если } x \leq 0, \\ B_2, & \text{если } x > 0. \end{cases} \quad (8)$$

В этом случае светлота зрительного ощущения, определяемая с помощью (2), (3), (5) и (7), запишется в виде:

$$S(x) = \begin{cases} S_1 - \Delta S e^{\frac{x}{n}}, & \text{если } x \leq 0, \\ S_2 + \Delta S e^{-\frac{x}{n}}, & \text{если } x > 0, \end{cases} \quad (9)$$

где

$$\begin{aligned} S_1 &= (1 - k_k) c \lg dB_1, \\ S_2 &= (1 - k_k) c \lg dB_2, \\ \Delta S &= \frac{1}{2} k_k c \lg \frac{B_2}{B_1}. \end{aligned} \quad (10)$$

На рис. 2 а и б представлены диаграммы изменения яркости зрительной картины и светлоты зрительного ощущения в функции координаты  $x$ . Как видим, на уровне зрительного ощущения граница выглядит подчеркнутой. Подобное подчеркивание границы наблюдается также и в психофизическом эксперименте в виде краевого контраста.

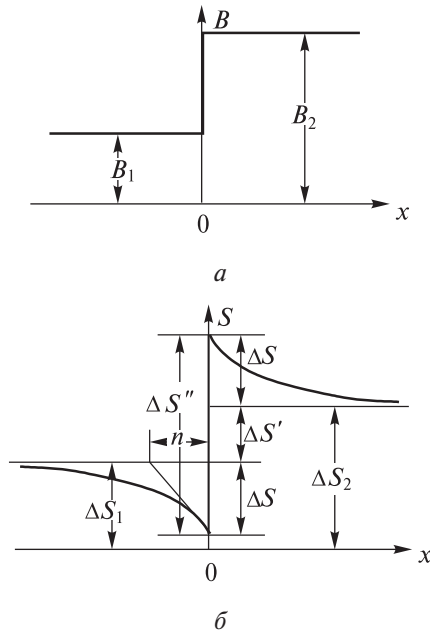


Рис. 2

Определим, во сколько раз увеличивается перепад светлоты на границе, благодаря действию краевого контраста. При отсутствии действия краевого контраста перепад светлоты на границе  $\Delta S$  был бы равен:

$$\Delta S' = S_2 - S_1 = (1 - k_k) c \lg \frac{B_2}{B_1}. \quad (11)$$

Перепад светлоты с учетом краевого контраста равен:

$$\Delta S'' = 2\Delta S + \Delta S' = c \lg \frac{B_2}{B_1}. \quad (12)$$

Следовательно, за счет действия краевого контраста перепад светлоты на границе увеличивается в

$$\frac{\Delta S''}{\Delta S'} = \frac{1}{1 - k_k} \quad (13)$$

раз.

Формула (13) показывает, что интенсивность подчеркивания контура определяется исключительно значением постоянной модели  $k_k$ , чем оправдывается введенное для нее выше название. Из формулы (13) также следует, что постоянная  $k_k$  заключена в пределах от 0 до 1. Рост величины параметра  $k_k$  приводит к усилению подчеркивания границы. Из диаграммы (рис. 2 б) можно видеть, что при удалении от границы приращение светлоты  $\Delta S$ , обусловленное краевым контрастом, плавно снижается, затухая до нуля. Темп этого затухания полностью определяется значением константы  $n$ , изображенной на диаграмме в виде отрезка прямой.

## 2. Исследование ступенчатой реакции модели краевого контраста

Теперь, когда математическая модель краевого контраста построена, приступим к ее экспериментальной проверке. Уже простое наблюдение видимой границы между полями различной яркости свидетельствует о факте подчеркивания границы, интенсивность которого постепенно угасает при удалении от границы. Для строгого количественного сравнения реакций глаза и модели необходимо было бы определить фактический закон изменения светлоты при удалении от границы. Однако, к сожалению, сознание наблюдателя, используемое в качестве записывающего прибора, работает очень неточно, и, кроме грубых оценок, здесь ничего другого не удастся извлечь. Однако известно, что сознание наблюдателя можно использовать как весьма точный нулевой прибор, фиксирующий равенство или неравенство светлоты на различных участках зрительного ощущения.

Этим свойством сознания можно воспользоваться при решении обратной задачи. До сих пор мы по заданной зрительной картине определяли, каким будет зрительное ощущение. Теперь ставится обратная задача: задано зрительное ощущение, необходимо подыскать такую зрительную картину, которая порождает именно это зрительное ощущение. Решив эту задачу, мы получим возможность формировать такие простые зрительные ощущения, для анализа которых будет достаточно использования сознания наблюдателя лишь в качестве нулевого прибора. Для возбуждения такого зрительного ощущения придется взять достаточно сложную зрительную картину. Сравнивая теоретически предсказанную зрительную картину со

зрительной картиной, фактически найденной в эксперименте, можно будет осуществить количественную проверку справедливости модели.

Пусть задано произвольное зрительное ощущение  $S(x, y)$ . Зрительная картина  $B(x, y)$ , порождающая это ощущение, может быть определена по формулам:

$$B = \frac{1}{d} 10^{\frac{1}{c} Q}, \quad (14)$$

$$Q(x, y) = S(x, y) + \frac{k_k}{2\pi n^2} \iint_{-\infty}^{\infty} Q(\xi, \eta) K_0 \left( \frac{\sqrt{(\xi-x)^2 + (\eta-y)^2}}{\frac{n}{\sqrt{1-k_k}}} \right) d\xi d\eta. \quad (15)$$

В случае, когда зрительное ощущение однородно вдоль оси  $y$ , формула (15) запишется в более простом виде:

$$Q(x) = S(x) + \frac{k_k}{2\sqrt{1-k_k}} \int_{-\infty}^{\infty} S(\xi) e^{-\frac{|\xi-x|}{n\sqrt{1-k_k}}} d\xi. \quad (16)$$

Зададимся зрительным ощущением, описываемым следующей системой условий:

$$S(x) = \begin{cases} S_1, & \text{если } x \leq 0, \\ S_2, & \text{если } x > 0. \end{cases} \quad (17)$$

Такое зрительное ощущение разделено вертикальной границей на две области, левая область имеет светлоту  $S_1$ , правая – светлоту  $S_2$ . Характер изменения светлоты зрительного ощущения показан на диаграмме (рис. 3 а).

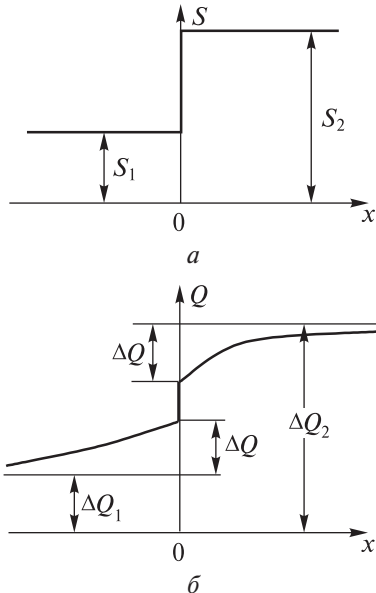


Рис. 3

Зрительная картина, порождающая такое ощущение, определится зависимостью:

$$Q(x) = \begin{cases} Q_1 + \Delta Q e^{\frac{x}{n}\sqrt{1-k_k}}, & \text{если } x \leq 0, \\ Q_2 - \Delta Q e^{-\frac{x}{n}\sqrt{1-k_k}}, & \text{если } x > 0. \end{cases} \quad (18)$$

где

$$\begin{aligned} Q_1 &= c \lg dB_1, \\ Q_2 &= c \lg dB_2, \\ \Delta Q &= \frac{1}{2} k_k c \lg \frac{B_2}{B_1}. \end{aligned} \quad (19)$$

Переход от переменной  $Q(x)$  к яркости зрительной картины  $B(x)$  осуществляется по формуле (14). Диаграмма зависимости функции  $Q$ , пропорциональной логарифму яркости зрительной картины, от координаты  $x$  дана на рис. 3 б.

С целью опытной проверки зависимости (18) на шести испытуемых были выполнены эксперименты, описание которых приведено ниже. С помощью вертушки (диска Максвелла) формировалась специальная зрительная картина, яркость которой изменяется вдоль радиуса по заданному закону, определяемому формулами (14), (18) и (19). Для достижения этого коэффициент отражения  $\rho$  диска вертушки должен меняться вдоль радиуса  $r$  согласно следующей зависимости:

$$\lg \rho = \begin{cases} \lg \rho_1 + \frac{1}{2} k_k \lg \frac{\rho_1}{\rho_2} e^{\frac{r-r_0}{hn}\sqrt{1-k_k}}, & \text{если } r \leq r_0, \\ \lg \rho_2 - \frac{1}{2} k_k \lg \frac{\rho_2}{\rho_1} e^{-\frac{r-r_0}{hn}\sqrt{1-k_k}}, & \text{если } r > r_0. \end{cases} \quad (20)$$

Здесь:  $r_0=60$  мм – радиус на диске вертушки, при котором происходит скачкообразное изменение светлоты;  $\rho_1$  и  $\rho_2$  – коэффициенты отражения, соответствующие яркостям  $B_1$  и  $B_2$ ;  $h=200$  мм – расстояние наблюдения.

В одном из опытов было принято:  $\rho_1=73\%$ ,  $\rho_2=83\%$ ,  $k_k=0,9$ ,  $n=20'=0,0058$  рад. При этих данных соответствующая фигура, вычерченная черной тушью на кружке из белой бумаги, имеет вид, изображенный на рис. 4 а.

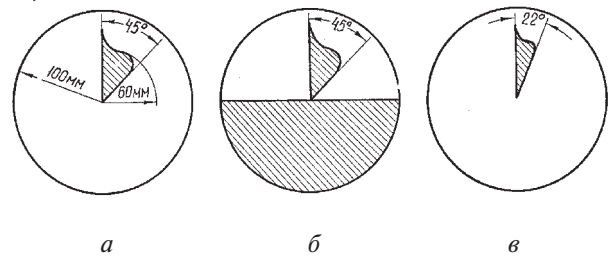


Рис. 4

Рассматривая быстро вращающийся диск Максвелла с указанным выше распределением яркости, испытуемый видит четкую границу без каких-либо признаков краевого контраста или размыва границы. Изменение параметра  $k_k$  в пределах  $0,85 \div 0,95$  не приводит к заметным искажениям зрительного ощущения. При увеличении параметра  $k_k$  до значения  $0,95$  и более испытуемый отмечает размыв границы, при уменьшении  $k_k$  до  $0,85$  и ниже становится заметным краевой контраст. Зрительное ощущение заметно не искажается также при изме-

нении параметра  $n$  в пределах  $15' \div 30'$ . Если взять  $n < 15'$ , то становится заметной неоднородность светлоты зрительного ощущения вблизи границы, в первую очередь на более темном поле. Если же  $n > 30'$ , то становится заметным краевой контраст.

Следует отметить, что изменение освещенности диска вертушки в широких пределах (от 10 до 1000 нит) в полном соответствии с теорией не вносит существенных изменений в характер зрительного ощущения, которое с обеих сторон границы продолжает выглядеть равномерным. При значениях параметров  $k_k=0,90$  и  $n=20'$  были также сформированы зрительные картины для других пар коэффициентов отражения:  $\rho_1=32\%$ ,  $\rho_2=42\%$ ;  $\rho_1=78\%$ ,  $\rho_2=83\%$ . Для этих картин порождающие их фигуры имеют вид, изображенный на рис. 4 б и в. Опыты с этими образцами также не дали существенных отклонений от теоретически предсказываемого результата. Выполненные опыты свидетельствуют в пользу справедливости предложенной модели краевого контраста зрения.

### 3. Исследование краевого контраста при предъявлении полосы

Возьмем в качестве входного сигнала модели зрительную картину в виде вертикальной полосы на однородном фоне. Обозначим яркость фона через  $B_0$ , прибавку яркости полосы по сравнению с фоном – через  $\Delta B$ , ширину полосы в угловых единицах – через  $\Delta x$ . Полагая, что ось полосы проходит через точку фиксации, можем записать следующие выражения для описания яркости зрительной картины:

$$B(x) = \begin{cases} B_0, & \text{если } x \leq -\Delta x/2, \\ B_0 + \Delta B, & \text{если } -\Delta x/2 < x \leq \Delta x/2, \\ B_0, & \text{если } x > \Delta x/2. \end{cases} \quad (21)$$

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

Ход изменения светлоты зрительного ощущения  $S(x)$ , вычисляемой по формулам (2)-(5) при условии (21), определится формулами:

$$S(x) = \begin{cases} Q_0 - kQ_0 - k\Delta Qsh \frac{\Delta x}{2n} e^{\frac{x}{n}}, & \text{если } x \leq -\frac{\Delta x}{2}, \\ Q_0 + \Delta Q - k\Delta Q \left( 1 - e^{-\frac{\Delta x}{2n} ch \frac{x}{n}} \right), & \text{если } -\frac{\Delta x}{2} < x \leq \frac{\Delta x}{2}, \\ Q_0 - kQ_0 - k\Delta Qsh \frac{\Delta x}{2n} e^{-\frac{x}{n}}, & \text{если } x > \frac{\Delta x}{2}, \end{cases} \quad (22)$$

где:

$$\begin{aligned} Q_0 &= c \lg dB_0, \\ \Delta Q &= c \lg \frac{B_0 + \Delta B}{B_0}. \end{aligned} \quad (23)$$

Диаграмма изменения светлоты зрительно-го ощущения показана на рис. 5 б. Из диаграммы видно, что контуры полосы должны выглядеть

подчеркнутыми с двух сторон. Средняя часть полосы должна быть по светлоте ближе к фону, чем ее края. Простой психофизический эксперимент демонстрирует, что зрительное ощущение, порождаемое полосой, по своему характеру именно таково.

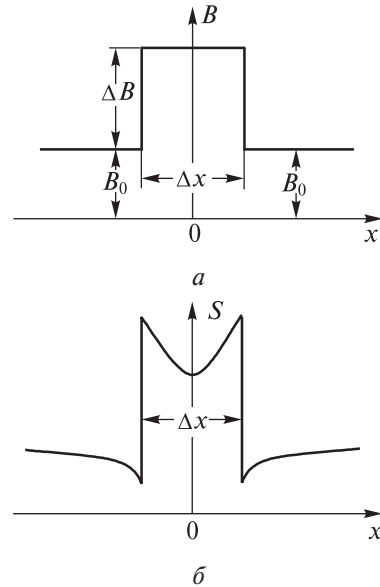


Рис. 5

Воспользуемся диском Максвелла (вертушкой) и вычертим для него на белом бумажном кружке сектор кольца с размерами, указанными на рис. 6. Закручивая вертушку со сверхкритической скоростью (порядка 50 об/сек), мы получим, согласно закону Галбота, зрительное ощущение такое же, как от зрительной картины в виде полосы (21). Небольшое искривление полосы в данном случае не имеет существенного значения. Наблюдая полосу, сформированную на диске вертушки, мы увидим распределение светлоты, соответствующее по своему характеру тому, которое было предсказано на основе анализа модельных реакций (рис. 5 б). Внутри кольца вблизи границ имеет место потемнение, интенсивность которого наиболее велика в непосредственной близости у границ. Вне кольца вблизи границ заметно посветление фона, оно наиболее велико у самих границ. Средняя часть полосы между двумя границами заметно темнее периферических ее частей.

Описанный эксперимент, подтверждая теорию в качественном отношении, почти ничего, однако, не дает для количественной ее проверки, поскольку на глаз очень трудно оценивать уровень светлоты на различных участках зрительного ощущения. Совершенно иначе будет обстоять дело при такой постановке эксперимента, когда испытуемый должен лишь устанавливать равенство или неравенство уровня светлоты на различных участках поля зрения. В этом случае можно осуществить довольно точный количественный анализ зрительного ощущения, поскольку сознание наблюдателя хорошо справляется с ролью нулевого прибора.

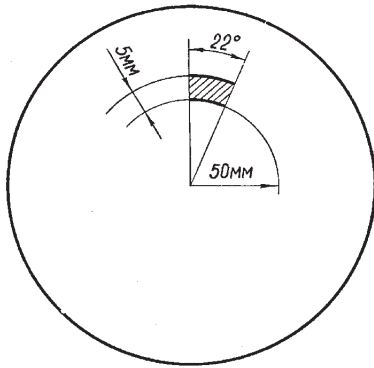


Рис. 6

Чтобы подойти к постановке подобного рода эксперимента, рассмотрим такой случай функционирования модели, когда на ее выходе возникает зрительное ощущение  $S(x)$  в виде полосы постоянной светлоты на однородном фоне. Математическое представление такого зрительного ощущения запишем в виде:

$$S(x) = \begin{cases} S_0, & \text{если } x \leq -\frac{\Delta x}{2}, \\ S_0 + \Delta S, & \text{если } -\frac{\Delta x}{2} < x \leq \frac{\Delta x}{2}, \\ S_0, & \text{если } x > \frac{\Delta x}{2}. \end{cases} \quad (24)$$

Диаграмма изменения светлоты зрительного ощущения показана на рис. 7 а.

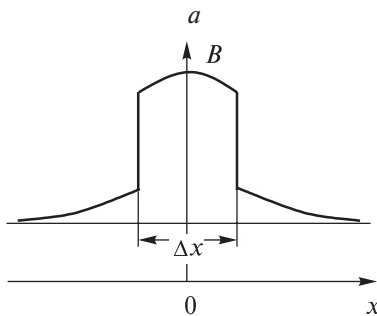
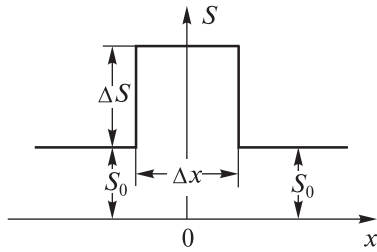


Рис. 7

Чтобы получить такого вида зрительное ощущение, необходимо на вход модели подать некоторую, вполне определенную зрительную картину. Формулы, характеризующие такую зрительную картину, можно получить, решая систему уравнений (14) и (16) при условии (24). Они имеют вид:

$$B = \frac{1}{d} 10^{\frac{1}{c} Q}, \quad (25)$$

$$Q(x) = \begin{cases} Q_0 + k_k \Delta Q sh \frac{\Delta x}{2n} \sqrt{1-k_k} \cdot e^{\frac{x}{n} \sqrt{1-k_k}}, & \text{если } x \leq -\frac{\Delta x}{2}, \\ Q_0 - k_k \Delta Q e^{-\frac{\Delta x}{2n} \sqrt{1-k_k}} \cdot ch \frac{x}{n} \sqrt{1-k_k}, & \text{если } -\frac{\Delta x}{2} < x \leq \frac{\Delta x}{2}, \\ Q_0 + k_k \Delta Q sh \frac{\Delta x}{2n} \sqrt{1-k_k} \cdot e^{-\frac{x}{n} \sqrt{1-k_k}}, & \text{если } x > \frac{\Delta x}{2}. \end{cases} \quad (26)$$

Диаграмма изменения яркости зрительной картины в поле зрения, построенная по формулам (25) и (26), имеет вид, указанный на рис. 7 б. Если теория верна, то при предъявлении зрительной картины, имеющей распределение яркости такое, как на рис. 7 б, наблюдатель должен увидеть полосу совершенно равномерной яркости. Фон вокруг полосы также должен видаться совершенно равномерным. Если зрительная картина подобрана неправильно, то это тотчас же обнаружится в эксперименте и проявится в виде неравномерностей яркости в поле зрения. Сознание наблюдателя легко обнаружит эти неравномерности, выполняя при этом функции весьма точного нулевого прибора. Эксперимент выполнялся на шести испытуемых с помощью той же вертушки, дозировка и распределение яркости зрительной картины регулировались соотношением угловых размеров белого и черного секторов вертушки. Этим достигалось желаемое изменение коэффициента отражения вдоль радиуса диска вертушки.

Чтобы зрительная картина, сформированная на диске вертушки, соответствовала формулам (25) и (26), коэффициент отражения  $\rho$  поверхности диска должен изменяться вдоль радиуса  $r$  по следующему закону:

$$\lg \rho = \begin{cases} \lg \rho_0 + k_k \lg \left( 1 + \frac{\Delta \rho}{\rho_0} \right) sh \frac{\Delta r}{2hn} \sqrt{1-k_k} \cdot e^{\frac{r-r_0}{hn} \sqrt{1-k_k}}, & \text{если } r - r_0 \leq -\frac{\Delta r}{2}, \\ \lg(\rho_0 + \Delta \rho) - k_k \lg \left( 1 + \frac{\Delta \rho}{\rho_0} \right) e^{-\frac{\Delta r}{2hn} \sqrt{1-k_k}} ch \frac{r-r_0}{hn} \sqrt{1-k_k}, & \text{если } -\frac{\Delta r}{2} < r - r_0 \leq \frac{\Delta r}{2}, \\ \lg \rho_0 + k_k \lg \left( 1 + \frac{\Delta \rho}{\rho_0} \right) sh \frac{\Delta r}{2hn} \sqrt{1-k_k} \cdot e^{-\frac{r-r_0}{hn} \sqrt{1-k_k}}, & \text{если } r - r_0 > \frac{\Delta r}{2}. \end{cases} \quad (27)$$

В формулах (27)  $\Delta r = 5\text{мм}$  означает ширину полосы в линейных единицах,  $r_0 = 52,5\text{мм}$  – радиус оси

полосы,  $h=500\text{мм}$  – расстояние, с которого ведется наблюдение полосы. Коэффициенты отражения  $\rho_0$  и  $\rho_0+\Delta\rho$  соответствуют светлотам  $S_0$  и  $S_0+\Delta S$  однородных полей. В эксперименте принято  $\rho_0=83\%$ ,  $\rho_0+\Delta\rho=78,2\%$ .

Раскраска кружков велась при различных значениях параметров  $k_k$  и  $n$ . Оказалось, что при значениях  $k_k=0,9$  и  $n=20\text{мм}$  получаем зрительное ощущение без каких-либо следов неравномерности по светлоте как внутри полосы, так и вне ее. В то же время сравнительно небольшие отклонения в значениях этих параметров приводят к появлению неравномерности светлоты зрительного ощущения. Характер отклонений примерно такой же, как это было описано в предыдущем параграфе. Раскраска кружка, дающая равномерное зрительное ощущение, изображена на рис. 8. Подобные опыты проводились также при ширине полосы 10мм и дали аналогичные результаты.

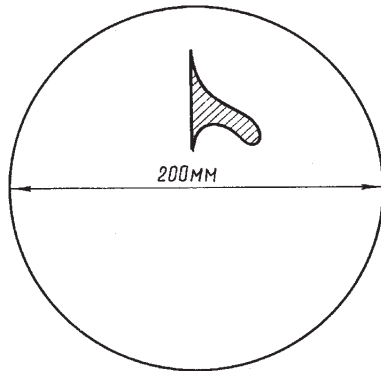


Рис. 8

Такого же рода опыты и теоретические расчеты выполнялись и при других разновидностях зрительных картин и зрительных ощущений – двойном скачке яркости и светлоты, а также “лесенке”. В этих случаях также получено удовлетворительное согласие теории и опыта.

**Выводы**

Предложена модель краевого контраста зрения, предназначенная для математического описания неоднородных процессов, протекающих в органе зрения человека при предъявлении ему крупных зрительных картин. Из моделей выведены теоретические зависимости, представляющие собой реакции различного вида зрительных картин. Выполненные в работе эксперименты, указывают на удовлетворительное согласование этих зависимостей с данными опыта. Из опыта найдены также численные значения двух введенных в модели констант, характеризующих способность человеческого зрения к подчеркиванию контуров изображения: постоянной интенсивности контраста  $k_k=0,9$  и постоянной затухания контраста  $n=20$ .

Сформулированная в этой и в предыдущих статьях этой серии [4-9] задача математического

моделирования зрения человека при зрительных картинах произвольного вида еще далеко не исчерпана, а полученные в них результаты следует рассматривать лишь как первые шаги на пути ее решения. Одним из ограничений в настоящей работе явилось то, что в экспериментах по проверке построенных моделей испытуемому предъявлялись зрительные картины, яркость которых не всегда варьировалась в достаточно широких пределах. В частности, в опытах не использовались светящиеся поверхности. Другое ограничение состояло в том, что значительная часть рассмотренных моделей относится лишь к однокомпонентному ахроматическому зрению. Мы предприняли попытку достроить математическую модель зрительного анализатора человека с тем, чтобы освободить ее от этих и иных ограничений. Опытных данных для прочного обоснования такой модели пока явно недостаточно, поэтому такая модель по необходимости выглядит, в основном, как умозрительная. Модель представляет собой многоярусную конструкцию, ее схема изображена на рис. 9. Передача сигналов в ней производится по трем каналам ( $i=1,2,3$ ).

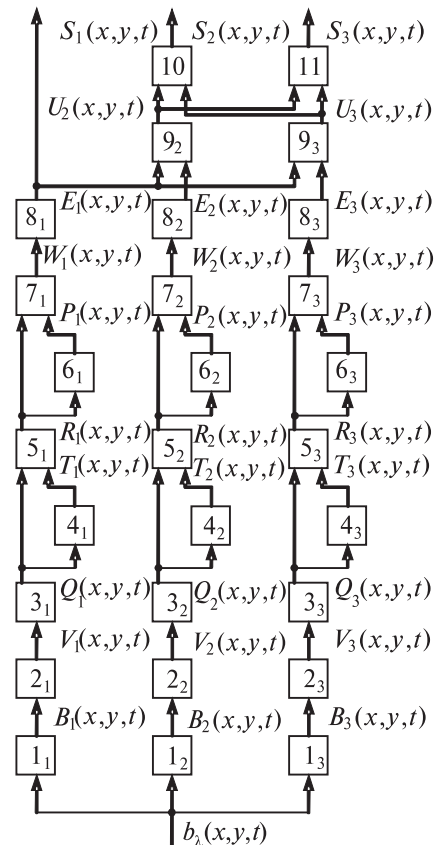


Рис. 9

Ниже приводятся уравнения, описывающие работу отдельных блоков:

Блок 1<sub>i</sub>:

$$B_i(x, y, t) = \int_{\lambda_1}^{\lambda_2} b_\lambda(x, y, t) m_i(\lambda) d\lambda. \quad (28)$$

Блок 2<sub>i</sub>:

$$a_i \frac{\partial V_i(x, y, t)}{\partial t} - b_i^2 \left( \frac{\partial^2 V_i(x, y, t)}{\partial x^2} + \frac{\partial^2 V_i(x, y, t)}{\partial y^2} \right) + V_i = B_i(x, y, t). \quad (29)$$

Блок 3<sub>i</sub>:

$$Q_i(x, y, t) = c_i \lg d_i V_i(x, y, t). \quad (30)$$

Блок 4<sub>i</sub>:

$$l_i \frac{\partial T_i(x, y, t)}{\partial t} - n_i^2 \left( \frac{\partial^2 T_i(x, y, t)}{\partial x^2} + \frac{\partial^2 T_i(x, y, t)}{\partial y^2} \right) + T_i = k_{ki} Q_i(x, y, t). \quad (31)$$

Блок 5<sub>i</sub>:

$$R_i(x, y, t) = Q_i(x, y, t) - T_i(x, y, t). \quad (32)$$

Блок 6<sub>i</sub>:

$$m_i \frac{\partial P_i(x, y, t)}{\partial t} - p_i^2 \left( \frac{\partial^2 P_i(x, y, t)}{\partial x^2} + \frac{\partial^2 P_i(x, y, t)}{\partial y^2} \right) + P_i = k_{ai} R_i(x, y, t). \quad (33)$$

Блок 7<sub>i</sub>:

$$W_i(x, y, t) = R_i(x, y, t) - P_i(x, y, t). \quad (34)$$

Блок 8<sub>i</sub>:

$$E_i(x, y, t) = r_i 10^{W_i(x, y, t)} + s_i. \quad (35)$$

Блок 9<sub>i</sub>:

$$U_i(x, y, t) = E_i(x, y, t) - E_1(x, y, t), \quad (i = 2, 3). \quad (36)$$

Блок 10:

$$S_2(x, y, t) = \sqrt{U_2^2(x, y, t) + U_3^2(x, y, t)}. \quad (37)$$

Блок 11:

$$S_3(x, y, t) = \arctg \frac{U_2(x, y, t)}{U_3(x, y, t)}. \quad (38)$$

Модель реагирует на зрительные картины  $b_\lambda(x, y, t)$  общего вида и формирует вполне определенный сигнал  $\bar{S}(x, y, t)$ , моделирующий зрительное ощущение. По промежуточным каналам модели передаются сигналы  $B_i, E_i, P_i, Q_i, R_i, T_i, U_i, V_i, W_i$ ,

являющиеся, подобно входным и выходным сигналам модели, функциями координат поля зрения  $x, y$  и времени  $t$ . Величины  $m_i(\lambda), a_i, b_i, c_i, d_i, k_{ki}, k_{ai}, l_i, m_i, n_i, p_i, r_i, s_i$  являются параметрами модели. Первый канал моделирует ахроматическое зрение, второй и третий – цветное зрение. Каждый ярус модели представляет одну из функций зрения. Так, первый ярус представляет спектральную чувствительность зрения, второй – инерцию зрения, ярусы от третьего по пятый – быструю адаптацию зрения и краевой контраст, ярусы от шестого по восьмой – медленную адаптацию и индукцию зрения. Последние два яруса осуществляют формирование цветового фона и насыщенности зрительного ощущения. Все рассмотренные в настоящей работе модели являются частными случаями общей модели зрения. Из 28 блоков, содержащихся в общей модели, в настоящей работе рассматривались только 12. Хотя в большей своей части эта модель еще недостаточно обоснована экспериментом, однако, она, по нашему мнению, может служить отправным пунктом для постановки новых экспериментальных и теоретических исследований.

**Список литературы:** 1. *Helmholtz H.*, Handbuch der physiologischen Optic, Hamburg u. Leipzig, 1909-1911. 2. *Hering E.*, Die Lehre vom Lichtsinn, 1874. 3. *Ватсон Д.Н.* Теория бесселевых функций: Пер. с англ. ИЛ, М., 1949. 4. *Бондаренко М.Ф., Шабанов-Кушнарченко Ю.П.* Модели цветового зрения // Бионика интеллекта: науч.-техн. журнал. – 2006. – № 1. – С. 27-41. 5. *Бондаренко М.Ф., Шабанов-Кушнарченко С.Ю., Шабанов-Кушнарченко Ю.П.* Модель сглаживания в зрении // Бионика интеллекта: науч.-техн. журнал. – 2007. – № 1. – С. 33-47. 6. *Бондаренко М.Ф., Шабанов-Кушнарченко Ю.П.* Модель статики адаптации органов чувств // Бионика интеллекта: науч.-техн. журнал. – 2007. – № 2. – С. 3-14. 7. *Бондаренко М.Ф., Шабанов-Кушнарченко Ю.П.* Модель динамики адаптации зрения // Бионика интеллекта: науч.-техн. журнал. – 2007. – № 2. – С. 15-20. 8. *Бондаренко М.Ф., Шабанов-Кушнарченко С.Ю., Шабанов-Кушнарченко Ю.П.* Модель инерции зрения // Бионика интеллекта: науч.-техн. журнал. – 2008. – № 1. – С. 3-12. 9. *Бондаренко М.Ф., Шабанов-Кушнарченко С.Ю., Шабанов-Кушнарченко Ю.П.* Модель иррадиации зрения // Бионика интеллекта: науч.-техн. журнал. – 2008. – № 1. – С. 13-22.

Поступила в редколлегию 25.01.2008

УДК 681.518:004:912



## СВЕТОВЫЕ УСТРОЙСТВА ВИЗУАЛИЗАЦИИ ДЛЯ УПРАВЛЕНИЯ ДОРОЖНЫМ ДВИЖЕНИЕМ

Вал. Н. Бурцев<sup>1</sup>, Влад. Н. Бурцев<sup>2</sup>, А. Л. Ерохин<sup>3</sup>

<sup>1,2</sup> НПФ «Цветодинамика», г. Харьков, Украина, new\_idea@kr.net

<sup>3</sup> Харьковский национальный университет внутренних дел, г. Харьков, Украина, ayerokhin@ukr.net

Рассматривается решение задачи создания световых информационных экранов на основе визуального кодирования информации о дорожной обстановке с учетом моделей восприятия информации человеком. Принципы работы устройств основаны на использовании принципов световой проводимости в оптически прозрачных материалах и программном управлении засветкой информационных слоев.

### УСТРОЙСТВА ВИЗУАЛИЗАЦИИ, КОГНИТИВНЫЕ ФУНКЦИИ ЧЕЛОВЕКА, УПРАВЛЕНИЕ ДОРОЖНЫМ ДВИЖЕНИЕМ

#### Введение

Одним из важных направлений в разработке систем и средств искусственного интеллекта является создание интеллектуальных систем синтеза зрительной информации на основе моделей интеллектуальной деятельности человека. Современное дорожное движение можно представить в виде человеко-машинной системы, в которой водитель является лицом, принимающим решение.

При этом важнейшим параметром такой системы является скорость и частота принятия решения человеком-оператором, поскольку от этого зависит безопасность дорожного движения. Рост грузоперевозок и постоянное увеличение числа автомобильного транспорта ставит задачу обеспечения высокого уровня безопасности дорожного движения, устранения заторов на дорогах с одновременным обеспечением оптимальной скорости передвижения. На ближайшие десятилетия управление дорожным движением будет оставаться прерогативой систем управления на основе световых знаковых систем, которые со времени появления их первых образцов [1, 2] кардинально не изменились.

#### 1. Постановка задачи

Решение проблем управления большими грузопотоками с помощью световых систем можно решать в двух направлениях.

Первое из них — разработка простых и надежных световых систем, преобразователей световых сигналов по принципу цветового кодирования (коды запрещения, внимания и запрета), но основанных на новых принципах формирования указанных световых кодов.

Второе направление — разработка интеллектуальной системы поддержки принятия решений с несколькими уровнями иерархии.

#### 2. Разработка модели светового информационного устройства

Для построения эффективных световых устройств визуализации параметров дорожного дви-

жения вначале необходимо исследовать модели восприятия зрительным анализатором человека актуальной динамической информации. Как известно [3], конечным результатом зрительного восприятия любого объекта человеком является возникновение зрительного образа этого объекта. При этом под зрительным образом понимается состояние выхода зрительной системы, вызываемое множеством различных, но сходных по каким-то признакам изображений. В теории информации распознавание рассматривается как получение наиболее сжатого и экономного описания визуального образа, получаемого, как правило, в результате статистического кодирования.

К интеллектуальным задачам водителя, как лица, принимающего решение, относятся задачи двух групп:

- 1) выделение и распознавание визуального объекта из фона (распознавание участников дорожного движения, сигналов устройств управления дорожным движением);
- 2) синтез решения в условиях дефицита времени.

Вторая задача достаточно хорошо исследована, и предложено множество методов, моделей и средств синтеза решений. Первой задаче также посвящено достаточно исследований, однако, в них недостаточно внимания уделено так называемой проблеме моделирования механизмов адаптации зрительного анализатора человека [4]. Под адаптацией понимают изменение чувствительности в зависимости от интенсивности стимуляции. Механизмы зрительной адаптации у человека-оператора проявляются в снижении чувствительности зрительного анализатора как в условиях гиперосвещенности, так и в условиях недостаточной освещенности и часто являются причинами ошибок при решении задач первой группы (невозможность адекватного выделения визуального объекта из фона), и как следствие, второй группы (синтез неверного решения). Различают световую и темновую адаптацию зрительного анализатора. При-

мерами стимулов для запуска темновой адаптации зрительного анализатора человека могут служить как ослепление водителя встречным автомобилем, так и повышенная яркость свечения светофора в темное время суток. На рис. 1 приведен примерный график темновой адаптации человека в зависимости от яркости вспышек, который показывает, что полная адаптация после серии вспышек наступает через 20-30 минут [4].

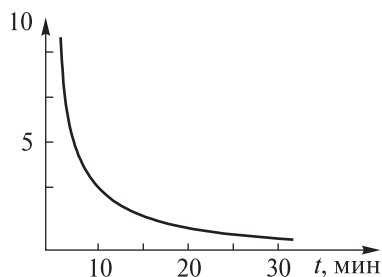


Рис. 1. График темновой адаптации зрительного анализатора

В качестве примера стимула световой адаптации можно привести недостаточную видимость сигналов светофора в яркую солнечную погоду. Процессы световой адаптации человека занимают в среднем от 3 до 5 минут [4], при том, что время на принятие решение у человека в дорожной обстановке ограничено долями секунды.

Поэтому создание световых устройств визуализации, учитывающих модели интеллектуальной деятельности человека-оператора при восприятии визуальной информации и принятии решения, является актуальным направлением.

В последние годы в световых системах управления дорожным движением произошла замена ламп накаливания на более надежные и экономичные источники света — твердотельные светодиоды в виде светодиодных матриц [4]. Такая замена дала возможность избавиться от некоторых традиционных недостатков в светофорах, но породила новые.

Наиболее существенными недостатками можно считать наличие в таких светофорах растровой структуры светодиодной матрицы, узкая направленность излучения высокоинтенсивных светодиодов и ограниченность площади светящихся иллюминаторов. Растровая структура, состоящая из счетного множества светодиодов (горящих точек) в матрице, воспринимается как целое изображение только на расстоянии не менее 10 метров, тем самым снижается уровень психофизиологического восприятия любой визуальной информации. Это ограничение связано с пределом углового разрешения глаза человека, равного 1 угловой минуте. К тому же узкая направленность светового излучения высокоинтенсивных светодиодов вызывает ослепляющий эффект и ограничивает угол визирования

световой информации при превышении водителем углового положения критического угла выходной апертуры светодиода. Последний недостаток светофоров унаследован от классической конструкции светофоров с лампами накаливания.

### 3. Разработка светового информационного устройства для управления дорожным движением

Альтернативой классическому светофору является световое информационное устройство, принципы работы которого основаны на использовании принципов световой проводимости в оптически прозрачных материалах [5]. Общий фронтальный вид устройства, состоящего из одного светопроводящего светящегося экрана, и его поперечное сечение представлены на рис. 2.

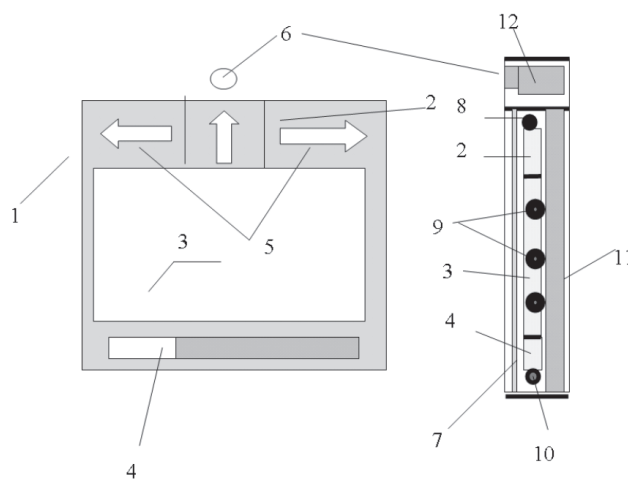


Рис. 2. Световое информационное устройство

Конструкция светового информационного устройства состоит из корпуса 1, охватывающего экран, разделенный на три информационные зоны — верхнюю 2, среднюю 3 и нижнюю 4. Верхняя зона 2 предназначена для размещения в ней светящихся знаков, например, стрелки-указатели 5. Средняя зона 3 служит непосредственно для визуального предъявления цветовых кодов (зеленый, желтый и красный цвета). В нижней зоне 4 размещается цветодинамический индикатор обратного отсчета времени, посредством которого визуально представляется время работы того или иного цветового кода. В верхней части корпуса 1 предусмотрено входное окно 6 фотоприемника, измеряющего уровень освещенности или цифровой малогабаритной видеокамеры на приборах с зарядовой связью.

Вся экранная часть устройства снабжена защитным акриловым стеклом 7. Информационные зоны 2-4 экрана устройства выполнены из оптически прозрачного акрилового стекла, торцевые поверхности которых оптически согласованы с наборами светодиодов 8-10, которые электрически связаны с электронным блоком программного

управления. Для возврата света, диффузно отраженного от информационных светопроводящих зон 2–4, имеется светоотражающий экран 11, выполненный из вспененного поливинилхлорида. Фотоприемник (видеокамера) 12 установлен в пространстве верхней части корпуса 1 совместно с электронным блоком измерения уровня освещенности и/или определения количества транспортных средств на перекрестке дорог. Электронный блок 13 программного управления светодиодами размещен внутри корпуса 1. Информационные зоны 2–4 экрана освещаются светодиодами 8–10 торцевой подсветки. За счет световой проводимости этих зон и использования специальных слоев, отражающих свет диффузным образом, достигается равномерная освещенность зон 2–4, что, кроме визуального предъявления цветовых кодов управления, дает возможность разместить на этих зонах дополнительную символно-знаковую информацию. Предъявление цветовых кодов осуществляется светодиодами 9, которые сгруппированы в триады красного, желтого и зеленого (RYG) излучений, работающих по программе управления.

Новое конструктивное исполнение светового информационного устройства основано на использовании эффекта световой проводимости, что обеспечивает возможность визуального предъявления на одном информационном экране последовательности из трех цветовых кодов управления с дополнительной символно-знаковой информацией. В зависимости от уровня внешней освещенности возможно изменять освещенность информационного экрана устройства. Совместно с такими многослойными световыми информационными устройствами достаточно использовать программное обеспечение автоматизированной системы управления дорожным движением на первом ее уровне иерархии.

#### 4. Алгоритм работы устройства для управления дорожным движением

Так как по правилам организации дорожного движения каждый перекресток снабжается двумя парами устройств управления, то указанные пары имеют между собой коммутацию обеспечивающих очередность предъявления цветовых кодов на каждой из дорог (полос движения), разрешающих или запрещающих движение.

Такая коммутация осуществляется, например, программным образом или в автоматизированном режиме. На рис. 3 представлена схема алгоритма работы устройства, которая включает в себя блоки: документирования, микропрограммного управления, коррекции временных интервалов и яркости свечения, замера скорости транспортных средств, электронного управления интенсивностью свечения и подсчет плотности потока транспорта.

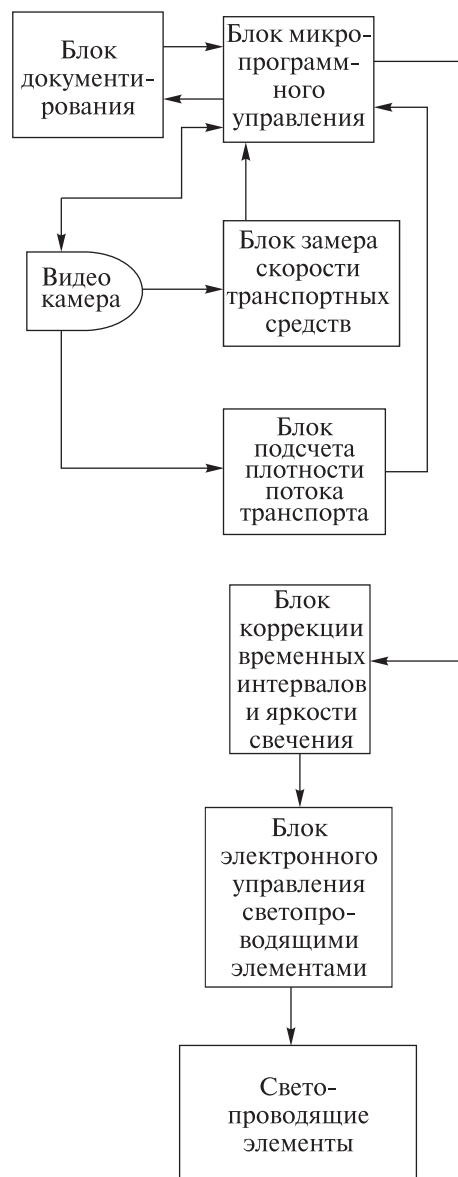


Рис. 3. Схема алгоритма работы устройства

Приведенный алгоритм предназначен для решения ряда проблем, возникающих на пересечениях главных дорог в зависимости от изменения плотности транспортных средств, например, в час пик.

Основной задачей этой системы является оценка транспортной ситуации на пересечении дорог и принятие решения по коррекции времени работы разрешающего (запрещающего) сигнала управления по направлениям движения. Наиболее важным является блок микропрограммного управления, который посредством прямых и обратных связей согласован с блоками замера скорости транспортных средств, подсчета плотности потока транспорта и документирования. Исходная информация в реальном времени поступает на эти блоки от малогабаритной видеокамеры. Управляющие сигналы от блока микропрограммного управления поступают в блоки коррекции временных интервалов и яркости свечения и электронного управления светопроводящими элементами.

Основными функциями блока микропрограммного управления являются:

- анализ дорожной ситуации по заторам;
- анализ дорожной ситуации по средней скорости движения через перекресток с помощью замеров скорости;
- вычисление временных интервалов переключения в зависимости от плотности транспорта и средней скорости движения;
- вычисление уровней внешней освещенности и коррекция яркости свечения светопроводящих элементов.

Блок подсчета плотности потока транспорта выполняет следующие функции:

- сбор информации с видеокамеры о подвижных и неподвижных объектах на перекрестке;
- вычисление оценок плотности транспорта за период времени  $\tau$ , определяемый программой в блоке микропрограммного управления:

$$\tau = f(V_{\text{ср}}, N),$$

где  $V_{\text{ср}}$  – средняя скорость потока;  $N$  – количество единиц транспорта в потоке.

Тогда плотность транспорта представима как функция

$$P_i = N / \tau,$$

где  $i$  – номер светового указателя направления движения

Блок коррекции временных интервалов и яркости свечения предназначен для вычисления времени переключения светопроводящих слоев в сторону увеличения (уменьшения) временного интервала работ разрешающего (запрещающего) знака в направлении пропуска транспорта с наибольшей (наименьшей) плотностью. Для корректировки времени используются данные плотности потока  $P_i$

$$t_i = f(P_i, \tau).$$

Коррекция яркости свечения светопроводящих слоев в сторону уменьшения (увеличения) осуществляется на основании данных уровня внешней освещенности, получаемых с видеокамеры через блок микропрограммного управления, в котором производится анализ пороговых значений внешней освещенности, и при превышении этого

значения увеличивается интегральная освещенность экрана устройства.

Блок документирования выполняет следующие функции:

- запись в файл мгновенных изображений транспортных средств, которые превысили установленную скорость;
- запись в видеофайл нештатной ситуации (в случае резкой смены дорожной обстановки по резкому изменению уровней яркости видеокамеры).

### Выводы

Разработанные принципы работы новых световых информационных устройств обеспечивают следующие преимущества:

- 1) возможность визуально предъявлять на одном информационном экране последовательно три цветовых кода управления с дополнительной символично-знаковой информацией;
- 2) изменение в автоматическом режиме освещенности информационного экрана устройства в зависимости от уровня внешней освещенности;
- 3) эффективное использование простого программного управления для автоматизированной системы управления дорожным движением.

Предложенное световое информационное устройство для управления дорожным движением является перспективной реализацией новых технических возможностей проектирования новых светооптических систем визуального представления знаковой, графической и иной информации для замены, например, светофоров или других световых информационных устройств.

**Список литературы:** 1. *G.Morgan*. Traffic Signal. Pat. USA №1475024, 27.02.1923. 2. *Combination Traffic Guide and Traffic-Regulating Signal*. Pat. USA №1447659, 07.08.1923. 3. *Броника*. Биологические аспекты / Под общей редакцией Л.В.Решодько. – К.: Вища школа, 1978. – 304 с. 4. *Светофоры* одно или двухсекционные производства ООО Саранская Электронная компания. (Режим доступа: <http://www.svetled.ru/content.view/62/72/>). 5. *Бондаренко М.Ф., Бурцев Вал.М., Бурцев Вол.М., Єрохін А.Л.* Світловий інформаційний екран диспетчерського пульту моніторингу складних систем / Заявка на винахід № a200611597. МПК G09F 9/00, G09F 13/00. Пріоритет від 03.11.2006. – 19 с.

*Поступила в редакцию 5.02.2008*

УДК 621.643:62-192



## ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ОПТИМИЗАЦИИ РЕЖИМОВ РАБОТЫ КОМПРЕССОРНОЙ СТАНЦИИ МАГИСТРАЛЬНОГО ГАЗОПРОВОДА

А.Д. Тевяшев<sup>1</sup>, С.Н. Набатова<sup>2</sup><sup>1</sup>ХНУРЭ, г. Харьков, Украина, tad45@mail.ru<sup>2</sup>ХНУРЭ, г. Харьков, Украина, svet\_nik@ukr.net

В статье сформулирован новый класс задач оптимизации плановых режимов работы компрессорной станции магистрального газопровода по векторному критерию, включающему показатели надежности работы технологического оборудования и обеспечивающему максимальную эффективность работы КС МГ. В качестве оценки эффективности работы КС МГ использовано математическое ожидание суммарных затрат топливного газа на компримирование заданного количества природного газа на планируемом интервале.

СТРУКТУРНАЯ И ПАРАМЕТРИЧЕСКАЯ ОПТИМИЗАЦИИ, НАДЕЖНОСТЬ, ЭФФЕКТИВНОСТЬ, КОМПРЕССОРНАЯ СТАНЦИЯ МАГИСТРАЛЬНОГО ГАЗОПРОВОДА, ТЕХНОЛОГИЧЕСКОЕ ОБОРУДОВАНИЕ

### Введение

Газовая промышленность относится к числу важнейших отраслей топливно-энергетического комплекса Украины. При управлении развитием газотранспортной системы (ГТС) необходимо взаимосвязанное решение двух проблем. Первая проблема состоит в развитии структуры управляемой производственно-транспортной системы, то есть определении оптимального состава элементов и их взаимосвязей, распределении плановых заданий по элементам и тому подобное. Вторая – вопросы развития структуры системы управления, включающие выбор иерархии управления, распределение выполняемых функций управления между уровнями и узлами системы.

Компрессорные цеха (КЦ) как части компрессорных станций (КС) являются основным управляемым элементом в комплексе технологических объектов магистрального газопровода (МГ). Управление КЦ позволяет изменять режимы работы ГТС, компенсировать колебания потребления и подачи газа, а также максимально использовать аккумулирующую способность газопровода.

Существующие методы оптимизации режимов работы КС МГ в большинстве своем направлены на:

– обеспечение надежной работы КС МГ за счет оптимального управления технологическими процессами;

– достижение высоких технико-экономических показателей за счет снижения удельных затрат энергоресурсов, эффективного использования установленных мощностей.

Однако эти методы, в лучшем случае, учитывают техническое состояние газоперекачивающих агрегатов (ГПА) путем расчета производительности каждого конкретного ГПА и выбора оптимальных частот вращения вала центробежного нагнетателя

(ЦБН), передаваемых в систему автоматического управления (САУ) ГПА в качестве уставок. Но они не учитывают такой значительный факт как техническое состояние аппаратов воздушного охлаждения (АВО) и пылеуловителей (ПУ), являющихся также основным технологическим оборудованием КС МГ. Учет этой информации возможен за счет включения в состав информационно-аналитической системы управления режимами транспорта и распределения природного газа в ГТС модуля интеллектуальной системы оптимизации режимов работы КС МГ (ИСОРРКСМГ). Основная задача ИСОРРКСМГ – оптимизация режимов работы КС МГ с учетом технического состояния всего основного технологического оборудования, что делает разработку такой системы чрезвычайно актуальной.

Целью данной работы является рассмотрение одного из возможных методов решения проблемы оптимизации плановых режимов работы КС МГ с учетом технического состояния технологического оборудования. Достижение поставленной цели можно осуществить путем решения следующих задач:

– проведение структурной оптимизации по критерию надежности работы компрессорной станции магистрального газопровода, результатом которой является набор структур КС МГ, обеспечивающий заданную степень надежности при известном техническом состоянии технологического оборудования;

– проведение параметрической оптимизации по критерию минимума затрат топливного газа при поддержании заданного режима работы КС МГ для структур с надежностью выше заданной;

– выбор оптимального планового режима работы КС МГ, обеспечивающего максимальную эффективность и надежность работы КС МГ.

## 1. Расчет структурной надежности КС МГ

Газотранспортная система представляет собой систему магистральных газопроводов (МГ) с многониточными линейными участками (ЛУ) и многоцеховыми КС, объединенными между собой многониточными перемышками [1], что позволяет говорить о ГТС как о сложной технической системе (ТС), состоящей из совокупности технологического оборудования (элементов), предназначенных для выполнения определенных функций.

С формальной точки зрения, если ЛУ МГ и расположенную за ним КС МГ представить в виде локальной подсистемы (ЛП), то весь МГ можно представить в виде цепочки последовательно соединенных ЛП [1]. Построенная таким образом модель ЛП, являющейся компонентом общей модели ГТС, позволяет исследовать надежность ЛУ и КС как локальных объектов. Учитывая, что исследования надежности ЛУ приводят к более простым задачам, чем задачи исследования надежности КС, а вероятности состояний ЛП непосредственно находятся через вероятности состояний ЛУ и КС, подробнее остановимся на задаче анализа надежности КС МГ.

КС МГ, являющаяся элементом ГТС, в свою очередь также представляет сложную ТС, основное назначение которой – компримирование природного газа (ПГ), транспортируемого по ГТС. Элементами системы в данном случае выступают установки очистки ПГ (комплекс пылеуловителей (ПУ)), установки охлаждения (комплекс аппаратов воздушного охлаждения (АВО)) и компрессорные цеха (газоперекачивающие агрегаты (ГПА)).

В терминах теории надежности компрессорная станция магистрального газопровода – это резервированная техническая система с восстанавливаемыми элементами [2]. Для исследования надежности таких систем хорошо применим аппарат общей теории надежности.

Для КС МГ основным показателем надежности является безотказность ее функционирования, то есть обеспечение на интервале управления  $[0, T]$  требуемого давления ПГ на выходе КС, при заданной температуре ПГ на входе и выходе КС, а также расходе коммерческого газа через станцию, при условии гарантированной подачи ПГ в требуемом объеме и с нужными параметрами на входе системы.

Расчет надежности системы подразумевает определение показателей надежности выполнения заданных функций технологическим оборудованием (элементами) системы (например вероятность того, что ПУ будет работать заданное время, в заданных режимах эксплуатации с сохранением всех необходимых параметров по показателям очистки ПГ).

Рассмотрим в качестве исследуемой системы одноцеховую КС (ОКС).

Примем допущения, позволяющие представить реальную структуру ОКС при оценке ее надежности в виде эквивалентной структурной схемы:

1) элемент системы (ГПА, АВО, ПУ) может находиться только в двух состояниях: работоспособном и отказавшем;

2) случайные интервалы времени между отказами элементов и длительности нахождения их в отказавшем состоянии не противоречат экспоненциальным функциям распределения вероятностей;

3) интервалы времени между отказами (поток отказов) элементов и длительность их восстановлений (поток восстановлений) взаимно независимы;

4) предполагается, что после каждого отказа элемента и его последующего ремонта, работоспособность данного элемента полностью восстанавливается;

5) одновременно может ремонтироваться только один элемент.

Необходимо отметить, что не всегда структурная схема надежности аналогична конструктивной схеме расположения элементов.

Принятые допущения позволяют представить технологическое оборудование ОКС в виде элементов, характеризующихся следующими параметрами:

1) параметр потока отказов КЦ, вызванных внутренними причинами (механические повреждения элементов ГПА, дефект проточной части или других частей ГПА и так далее) –  $\lambda_1$ , 1/год;

2) параметр потока отказов АВО, вызванных внутренними причинами (механические повреждения элементов АВО, коробление или провисание теплообменных труб, отказ вентиляторов и так далее) –  $\lambda_2$ , 1/год;

3) параметр потока отказов ПУ, вызванных внутренними причинами (механические повреждения элементов ПУ, дефект изготовления или монтажа и так далее) –  $\lambda_3$ , 1/год;

4) параметр восстановлений КЦ, АВО, ПУ –  $\mu$ , 1/год (принимается стратегия ограниченного обслуживания).

При этом параметры отказов потоков обладают следующими свойствами:

1) для любого момента времени параметр безотказной работы больше, чем частота отказов;

2) при длительной эксплуатации ремонтируемого оборудования поток его отказов независимо от закона распределения времени безотказной работы становится стационарным;

3) параметр потока отказов системы не равен сумме параметров потока отказов элементов.

Параметры потоков отказов элементов системы считаются в зависимости от выбранной струк-

турной схемы (последовательной, параллельной, последовательно-параллельной). Для анализа надежности сложных систем чаще всего применяются последовательно-параллельные структурные схемы [3]. Однако, если ТС не сводятся к параллельному, последовательному или последовательно-параллельному типу, то можно воспользоваться также методом логических схем с применением алгебры логики (булевой алгебры).

Учитывая, что система является резервируемой, то есть при отказе элемента в работу вступает один из резервных (запасных) элементов, примем допущения о том, что:

- 1) отказ системы происходит, если откажут все элементы, включая резервные;
- 2) вероятность отказа каждого элемента не зависит от состояния остальных элементов (отказы статистически независимы);
- 3) отказывать может только оборудование, находящееся в работе, а запасное оборудование не может выходить из строя до того, как оно будет включено в работу;
- 4) переключающие устройства считаются абсолютно надежными;
- 5) резервные элементы имеют характеристики как новые.

Не нарушая общности, при принятых допущениях проведем анализ надежности структур ОКС  $S = (s_1, s_2, \dots, s_n)$  ( $n$  — количество различных структур) на интервале управления  $[0, T]$ , состоящей из трех неодинаковых элементов (ПУ, АВО, КЦ), для которых характерны отказы.

Для каждой  $s_i \in S$  проведем оценку вероятности нахождения системы в различных состояниях.

Введем следующие обозначения состояния системы:

$P_0(t)$  — система исправна в момент времени  $t \in [0, T]$ ;

$P_1(t)$  — в системе за момент времени  $t \in [0, T]$  элемент 1 (ПУ) вышел из строя, а элементы 2 (КЦ) и 3 (АВО) функционируют;

$P_2(t)$  — в системе за момент времени  $t \in [0, T]$  элемент 2 вышел из строя, а элементы 1 и 3 функционируют;

$P_3(t)$  — в системе за момент времени  $t \in [0, T]$  элемент 3 вышел из строя, а элементы 1 и 2 функционируют;

$P_4(t)$  — в системе за момент времени  $t \in [0, T]$  элементы 1 и 2 вышли из строя, а элемент 3 функционирует;

$P_5(t)$  — в системе за момент времени  $t \in [0, T]$  элементы 2 и 3 вышли из строя, а элемент 1 функционирует;

$P_6(t)$  — в системе за момент времени  $t \in [0, T]$  элементы 1 и 3 вышли из строя, а элемент 2 функционирует;

$P_7(t)$  — в системе за момент времени  $t \in [0, T]$  элементы 1, 2 и 3 вышли из строя.

Составим граф вероятностей перехода из одного состояния в другое для перечисленных 8 возможных состояний (рис. 1).

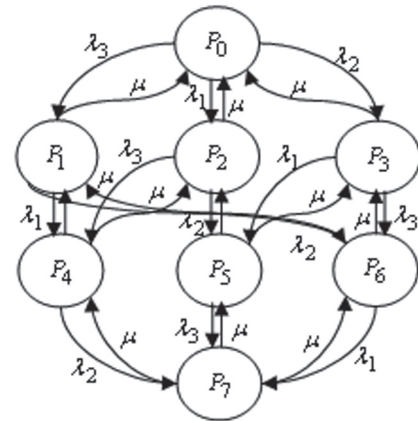


Рис. 1. Граф интенсивности перехода ОКС из одного состояния в другое

Эти состояния могут быть объединены в группы, каждая из которых объединяет состояния, приводящие к одним и тем же типам отказов.

Стоит отметить, что рассматривается система с ограниченным восстановлением и, в случае восстановления (ремонта) элемента ОКС, система может перейти в исправное состояние только из смежных состояний с интенсивностью  $\mu$ .

Учитывая, что при отказе КЦ (или отказе КЦ совместно с любым другим элементом) в момент времени  $t \in [0, T]$  система приходит в состояние функционального отказа, а при отказе АВО или ПУ (или их совместного отказа) система приходит в состояние параметрического отказа, для рассматриваемой системы условно можно выделить три группы состояний системы:

0:  $\{P_0(t)\}$  — система исправна и полностью работоспособна;

1:  $\{P_1(t), P_3(t), P_6(t)\}$  — система находится в состоянии параметрического отказа;

2:  $\{P_2(t), P_4(t), P_5(t), P_7(t)\}$  — система находится в состоянии функционального отказа.

Определение вероятностей пребывания системы в каждом из состояний производится методами теории массового обслуживания с использованием вероятной схемы «гибели и размножения» [4].

Вероятности переходов из одного состояния в другое вычисляются для каждой конкретной задачи в зависимости от того, какая структура ОКС анализируется.

Для упрощения поставленной задачи можно ограничиться вычислением только вероятностей групп состояний, тем более что именно они представляют основной интерес. В этом случае придется иметь дело с тремя состояниями системы. Граф перехода такой упрощенной системы из одной группы состояний в другую приведен на рис. 2.

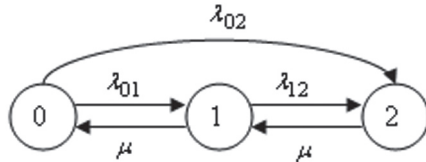


Рис. 2. Граф интенсивности перехода ОКС из одной группы состояний в другую

Математическая модель системы, представленной на рис. 2, представляет собой систему уравнений Эрланга следующего вида:

$$\begin{aligned} -\lambda_{01} \cdot P(0) + \mu \cdot P(1) &= 0; \\ \lambda_{01} \cdot P(0) - (\lambda_{12} + \mu) \cdot P(1) + \mu \cdot P(2) &= 0; \\ \lambda_{12} \cdot P(1) - \mu \cdot P(2) + \lambda_{02} \cdot P(3) &= 0; \\ P(0) + P(1) + P(2) + P(3) + P(4) &= 1, \end{aligned}$$

где  $P(0), P(1), P(2), P(3), P(4)$  — вероятности нахождения системы в соответствующем состоянии;  $\lambda_{01}, \lambda_{12}, \lambda_{02}$  — вероятности переходов системы из одной группы состояний в другую, с учетом исходных вероятностей переходов, входящих в рассматриваемую группу.

Решая построенную систему линейных уравнений, получаем вероятности нахождения системы в каждой из групп состояний (0,1,2). Далее, при необходимости, рассчитываются вероятности нахождения системы в каждом из выделенных состояний.

Проведенный таким образом анализ надежности структур  $S = (s_1, s_2, \dots, s_n)$  КС МГ дает возможность выделить некоторый набор  $S' = (s_1, s_2, \dots, s_m)$ ,  $S' \subset S$ , обеспечивающих безотказную работу системы  $P_0^{s_i}(t)$  на интервале управления  $[0, T]$  с заданной степенью надежности при известном техническом состоянии технологического оборудования, входящего в каждую из структур.

## 2. Распределение нагрузки между ГПА по критерию минимума затрат топливного газа

Параметрическую оптимизацию режимов работы КС МГ проведем по критерию минимума затрат топливного газа для каждой из структур  $S' = (s_1, s_2, \dots, s_m)$ , у которых  $P_0^{s_i}(t) \geq c, i = 1 \dots m$ , где  $c$  — заданная степень надежности анализируемой системы.

Опишем модель стационарного неизотермического транспорта газа на КС МГ [5]:

$$P_{hi}^2 - P_{ki}^2 = \varphi_1(q_i, T_{hi}, T_{ki}), \quad i \in M; \quad (1)$$

$$P_{hi}^2 - P_{ki}^2 = -\varphi_2(q_i, T_{hi}, T_{ki}, n_i), \quad i \in L; \quad (2)$$

$$P_{hi}^2 - P_{ki}^2 = \varphi_3(q_i, i), \quad i \in K; \quad (3)$$

$$P_{hi}^2 - P_{ki}^2 = \varphi_4(q_i, i), \quad i \in Z; \quad (4)$$

$$P_{hi}^2 - P_{ki}^2 = \varphi_5(q_i, T_{hi}, T_{ki}), \quad i \in P; \quad (5)$$

$$P_{hi}^2 - P_{ki}^2 = \varphi_6(q_i, T_{hi}, T_{ki}, u_i), \quad i \in A; \quad (6)$$

$$T_{ki} = \varphi_7(q_i, T_{hi}, P_{hi}, P_{ki}), \quad i \in M; \quad (7)$$

$$T_{ki} = \varphi_8(T_{hi}, P_{hi}, P_{ki}), \quad i \in L; \quad (8)$$

$$T_{ki} = \varphi_9(q_i, T_{hi}, P_{hi}, P_{ki}, u), \quad i \in A; \quad (9)$$

$$\sum_{i \in G_j^+} q_i - \sum_{i \in G_j^-} q_i = 0, \quad j \in V; \quad (10)$$

$$T_j \sum_{i \in G_j^+} q_i - \sum_{i \in G_j^-} q_i T_{jk} = 0, \quad j \in V; \quad (11)$$

$$X_{j,l \in G_j^-} \sum_{i \in G_j^+} \sum_{l \in G_j^-} X_{i,l} m_{i,l} - \sum_{i \in G_j^+} m_{i,l} = 0, \quad j \in V, \quad (12)$$

где  $M$  — множество участков трубопровода (УТ),  $|M| = m$ ;  $L$  — множество ГПА,  $|L| = l$ ;  $K$  — множество кранов,  $|K| = k$ ;  $Z$  — множество регулируемых кранов,  $|Z| = z$ ;  $P$  — множество ПУ,  $|P| = p$ ;  $A$  — множество АВО,  $|A| = a$ ;  $F$  — множество фиктивных дуг графа,  $G(V, E)$ ,  $|F| = f$ .

Вычисление  $\varphi(\cdot)$  связано с решением систем нелинейных уравнений, описывающих математическую модель конкретного экземпляра технологического оборудования.

Рассмотрим задачу распределение нагрузки между ГПА по критерию минимума затрат топливного газа  $Q_{\text{тп}}$  (тыс. м<sup>3</sup>/ч) для одноцеховой КС (ОКС).

При заданных структуре  $s_i \in S, i = 1 \dots m$  ОКС, значениях давлений  $P_{2E}(\omega), P_{2KE}$  (МПа) и температур  $T_{\text{вх}}(\omega), T_{\text{вых}}(\omega)$  (К) природного газа на входе и выходе ОКС, а также коммерческой продуктивности  $q(\omega)$  (млн. м<sup>3</sup>/сут) ОКС, данная задача может быть сведена к решению задачи нелинейного стохастического программирования в жесткой постановке следующего вида:

$$\min_{q \in \Omega_q \cap \Omega} Q_{\text{тп}}(q, \omega), \quad (13)$$

$$\Omega_q : Q_i^-(q, \omega) \leq q_i(q, \omega) \leq Q_i^+(q, \omega), \quad i \in L_i, \quad (14)$$

где область ограничений  $\Omega$  описывается уравнениями (1-12), дополненными уравнениями связи в виде уравнений состояния газа, уравнений, которые связывают компонентный состав транспортируемого природного газа с его физическими свойствами и так далее;  $L_i$  —  $i$ -й ГПА;  $Q_i^-(q, \omega)$  и  $Q_i^+(q, \omega)$  — верхнее и нижнее предельно допустимые значения коммерческой производительности  $i$ -ого ГПА, расчет которых производится по следующим формулам [6]:

$$\begin{aligned} Q_i^-(q, \omega) &= n_{\text{тп}}^-(P_{\text{н}}(\omega), P_{\text{к}}, T_{\text{н}}(\omega), T_{\text{к}}(\omega)) \times \\ &\times Q_{\text{тп}}^-(q, \omega) \frac{60 \cdot 10^3 P_{\text{н}}(\omega)}{\rho_0 \sqrt{Z_{\text{н}}(\omega) R T_{\text{н}}(\omega)} \sqrt{Z_{\text{тп}} R_{\text{тп}} T_{\text{тп}}}}, \end{aligned}$$

$$Q_i^+(q, \omega) = n_{\text{пр}}^+ (P_{\text{н}}(\omega), P_{\text{к}}, T_{\text{н}}(\omega), T_{\text{к}}(\omega)) \times \\ \times Q_{\text{пр}}^+(q, \omega) \frac{60 \cdot 10^3 P_{\text{н}}(\omega)}{\rho_0 \sqrt{Z_{\text{н}}(\omega) R T_{\text{н}}(\omega)} \sqrt{Z_{\text{пр}} R_{\text{пр}} T_{\text{пр}}}}$$

где  $P_{\text{н}}(\omega), P_{\text{к}}$  (МПа) – давление газа на входе и выходе ЦБН соответственно;  $T_{\text{н}}(\omega), T_{\text{к}}(\omega)$  (К) – температура газа на входе и выходе ЦБН;  $Q_{\text{пр}}(q, \omega)$  (м<sup>3</sup>/мин) – объемная приведенная производительность;  $n_{\text{пр}}$  (об/мин) – приведенная относительная частота вращения вала;  $\rho_0$  (кг/м<sup>3</sup>) – плотность газа в нормальных условиях (293,15 К и 0,101325 МПа);  $Z_{\text{н}}(\omega)$  (-),  $R(\omega)$  (Дж/кг·К) – параметры газового потока (коэффициент сжимаемости и газовая постоянная) на входе ЦБН, а  $Z_{\text{пр}}, R_{\text{пр}}, T_{\text{пр}}$  – условия приведения.

Функция  $Q_{\text{тг}}(q, \omega) = \sum_{i=1}^{l_i} Q_{\text{тг}i}(q, \omega)$  вычисляется согласно [6]:

$$Q_{\text{тг}i}(q, \omega) = q_{\text{тг}i}^{\text{н}}(q, \omega) \times \\ \times \left( 0,76 \cdot \frac{N_{\text{цбн}i}^{\text{факт}}(q, \omega)}{N_e^{\text{н}}} + 0,25 \cdot \sqrt{\frac{T_3}{T_3^{\text{н}}}} \cdot \frac{P_a}{0,1013} \right) \frac{Q_p^{\text{н}}}{Q_p(\omega)}, \quad (15)$$

$$i \in L_i$$

где  $Q_{\text{тг}i}(q, \omega)$  – затраты топливного газа на  $i$ -ом ГПА;  $q_{\text{тг}i}^{\text{н}}(q, \omega)$  – номинальный расход топливного газа с учетом поправки на допуски и техническое состояние  $i$ -ого ГПА;  $N_{\text{цбн}i}^{\text{факт}}(q, \omega)$  – потребляемая мощность  $i$ -м ЦБН, полученная в результате расчета параметров нагнетателя;  $Q_p(\omega)$  – низшая теплота сгорания топливного газа при стандартных условиях, (кДж/м<sup>3</sup>); величины  $q_{\text{тг}i}^{\text{н}}(q, \omega), N_e^{\text{н}}, T_3^{\text{н}}, T_3, P_a$  определяются согласно [7]. Величина  $Q_p^{\text{н}} = 34500$  (кДж/м<sup>3</sup>), а  $Q_p(\omega)$  – по исходным данным для расчетного состава транспортируемого природного газа.

Результатом решения поставленной задачи будет набор структур  $S = (s_1, s_2, \dots, s_m)$  с соответствующими им значениями производительности ГПА и количеством затрачиваемого топливного газа.

### 3. Выбор оптимального режима работы КС МГ

Для выбора оптимального режима работы КС МГ на интервале управления  $[0, T]$  (в нашем случае, как и ранее ОКС) следует провести оценку эффективности функционирования КС МГ для выбранных структур с учетом построенных для них функций надежности.

В качестве оценки эффективности ОКС будем использовать математическое ожидание суммарных затрат топливного газа на компримирование заданного количества природного газа на интервале  $[0, T]$  [8]:

$$I_0(T, s, u) = M_{\omega} \left\{ \int_0^T Q_{\text{тг}}(q, \omega, t) dt \right\},$$

где  $u$  – вектор управляемых переменных.

При заданной стоимости топливного газа, потребляемого ОКС, в качестве оценки эффективности функционирования ОКС целесообразно использовать математическое ожидание суммарной стоимости топливного газа на интервале управления [8]:

$$I_1(T, s, u) = p \cdot I_0(T),$$

где  $p$  – действующий тариф на топливный газ.

Для сравнительной оценки эффективности работы различных структур ОКС целесообразно использовать удельные затраты топливного газа на компримирование тысячи метров кубических природного газа:

$$I_2(T, s, u) = I_0(T, s, u) / Q_{\text{ОКС}}(T, u),$$

где  $Q_{\text{ОКС}}(T)$  – математическое ожидание суммарного количества компримируемого природного газа за весь интервал управления  $[0, T]$ .

Таким образом, с учетом вышеизложенного, задачу оптимизации режимов работы компрессорной станции магистрального газопровода с учетом технического состояния технологического оборудования на интервале управления  $[0, T]$  можно представить в виде многокритериальной задачи оптимизации с системой нелинейных ограничений (включающих систему уравнений (1)-(12) дополненных уравнениями связи в виде уравнений состояния газа, уравнений, которые связывают компонентный состав транспортируемого природного газа с его физическими свойствами и так далее, систему уравнения (13) и уравнения для расчета верхнего и нижнего предельно допустимых значений коммерческой производительности  $i$ -ого ГПА):

$$I_2(T, s, u) \rightarrow \min_{S, u \in \Omega'}, \quad (16)$$

$$\Omega' : \{ P_0^{s_i}(t) \geq c, i = 1 \dots m; \Omega_q \cap \Omega \} \quad (17)$$

где  $c$  – требуемая степень надежности системы, определяемая при проведении структурной оптимизации.

### Выводы

В статье рассмотрена проблема оптимизации плановых режимов работы компрессорной станции магистрального газопровода с учетом технического состояния технологического оборудования. Приведен метод расчета надежности КС МГ по показателю безотказности работы рассматриваемой системы при известном техническом состоянии технологического оборудования, входящего в каждую из структур. Рассмотрен метод параметрической оптимизации по критерию минимума затрат топливного газа при поддержании заданного режима работы КС МГ.

Научная новизна полученных результатов состоит в том, что впервые предложен метод оптими-

зации планового режима работы КС МГ, отличающийся от известных непосредственным учетом технического состояния (параметров потоков отказов) технологического оборудования (ПУ, АВО, ГПА), использование которого позволяет получать оптимальное по векторному критерию решение, устойчивое к прогнозируемому уровню внешних и внутренних возмущений.

**Список литературы:** 1. *Тевяшев А.Д., Набатова С.Н., Фролов В.А.* Системный анализ проблемы повышения надежности поставок природного газа потребителям газотранспортных систем // Восточно-европейский журнал передовых технологий: науч.-техн. журнал. – Харьков: Технологический центр, 2006. – Вып. 6/2 (24). – С. 34-39. 2. *Сухарев М.Г., Ставровский Е.Р.*, Резервирование систем. Магистральные трубопроводы. – М.: Недра, 1987. – 167 с. 3. *Герцбах И.* Теория надежности с приложениями к

профилактическому обслуживанию: Монография / Под ред. В.В. Рыкова / Пер. с англ. М.Г. Сухарева. – М.: Нефть и газ, 2003. – 263 с. 4. *Вентцель Е.С.* Теория вероятностей. – М.: Наука, 1969. – 576 с. 5. *Евдокимов А.Г., Тевяшев А.Д., Дубровский В.В.* Моделирование и оптимизация потоко-распределения в инженерных сетях. – М.: Стройиздат, 1990. – 365 с. 6. *Розгонюк В.В., Хачикян Л.А. та ін.* Експлуатаційникові газонафтового комплексу. Довідник. – К.: Росток, 1998. – 430 с. 7. *ОНТП 51-1-85* Магистральные трубопроводы. 8. *Набатова С.Н.* Системы интегральных показателей качества и эффективности функционирования газотранспортных систем // Восточно-европейский журнал передовых технологий: науч.-техн. журнал. – Харьков: Технологический центр, 2007. – Вып. 4 (28). – С. 3-12.

*Поступила в редколлегию 12.02.2008*

УДК 007.5



## МОДЕЛИ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ ДЛЯ БИЗНЕС-ПРОЦЕССОВ С ИЗМЕНЯЕМОЙ СТРУКТУРОЙ

В.М.Левыкин<sup>1</sup>, С.Ф.Чалый<sup>2</sup><sup>1</sup>ХНУРЭ, г.Харьков, Украина, levykin@kture.kharkov.ua<sup>2</sup>ХНУРЭ, г.Харьков, Украина, chaliy@datasvit.net

Предложены логические модели представления знаний в форме бизнес-правил, ориентированные на формализацию требований к бизнес-процессам. Использование предложенных правил формирует условия для выполнения реорганизации бизнес-процессов с изменяемой структурой.

БИЗНЕС-ПРАВИЛА, БИЗНЕС-ПРОЦЕССЫ С ИЗМЕНЯЕМОЙ СТРУКТУРОЙ, ЛОГИЧЕСКАЯ МОДЕЛЬ, ФУНКЦИОНАЛЬНЫЕ ТРЕБОВАНИЯ

### Введение

Бизнес-правила традиционно рассматриваются как средство построения баз данных с учетом действий и событий, связанных с сущностями данных. В более широком смысле бизнес-правило – это инструкция, которая определяет или ограничивает некоторый аспект деятельности организации. Соответственно, бизнес-правила формализуют знания, которые необходимы для управления функционированием организации [1].

Таким образом, знания о конкретной организации либо ее бизнес-процессах могут быть представлены в форме бизнес-правил, что позволяет на их основе сформировать требования к бизнес-процессам (БП). Такие требования подразделяются на функциональные и нефункциональные [2].

Функциональные требования определяют возможности и поведение процессов, которые должны быть реализованы, например «расследование нарушения должно заканчиваться в течение 24 часов после сообщения о нарушении». Нефункциональные требования определяют ограничения на поведение процессов, например, «отчет о нарушении должен быть доступен для корректировки специалистом через интранет». Бизнес-правила реализуют функциональные требования.

Таким образом, бизнес-правила *Bpr* предназначены для управления функционированием предприятия, содержат знания о бизнес-процессах, реализуют функциональные требования к БП, определяют возможности построения бизнес-процессов, а также осуществления их реорганизации.

### 1. Формулировка задачи

Реализация процессного подхода к управлению предприятием направлена на повышение эффективности его функционирования, базируется на совокупности мероприятий по совершенствованию системы управления и технологий деятельности предприятия и ориентирована на стратегию его развития. Вследствие изменения организационной структуры и целей деятельности предприятия, применения новых технологий, появления

новых процессов, изменения внешней среды выполняется реорганизация бизнес-процессов.

К такой реорганизации, как правило, приводят динамические изменения в функционировании предприятия, требующие регулярной адаптации бизнес-процессов согласно изменившимся целям предприятия, а также вследствие неконтролируемых внешних воздействий. Соответственно, реализация процессного подхода к управлению предприятием должна основываться на бизнес-процессах с изменяемой структурой (БПИС). Для управления такими бизнес-процессами предлагается использовать бизнес-правила.

Необходимым условием реализации данного подхода является решение задачи формализации представления знаний о функционировании БПИС в форме бизнес-правил.

Исходными данными задачи являются: многокомпонентное представление бизнес-процессов с изменяемой структурой; существующие типы бизнес-правил в рамках функционального подхода к управлению организацией.

Требуется получить модели представления знаний в форме бизнес-правил различных типов с учетом особенностей представления бизнес-процессов.

### 2. Логические модели представления знаний в форме бизнес-правил

Бизнес-правила *Bpr* в широком смысле слова характеризуют стратегию и ограничения деятельности предприятия, условия его функционирования и выражают знания, определяющие поведение конкретных бизнес-процессов [2]. Бизнес-правила определяют функциональные требования к модели предприятия в целом и бизнес-процессам в частности:

$$\forall Bpr_l \in Bpr, l = \overline{1, L},$$

$$Bpr_l = \begin{cases} true, & \text{если выполняется} \\ l\text{-требование,} \\ false & \text{в противном случае.} \end{cases} \quad (1)$$

Функциональные требования определяют описание БП и могут быть выражены с помощью следующих средств:

- управляющие последовательности (например, в форме диаграммы событий и состояний);
- объектно-ориентированная технология (с использованием классов объектов и описанием взаимодействия объектов);
- структуры данных и знаний.

Выявление, проверка правильности и верификация функциональных требований представляет собой достаточно сложную проблему. При выявлении требований используются различные технологии – интервьюирование, построение прототипов и тому подобное. Требования представляются в виде текстовых документов, диаграмм, комментариев либо в виде формальных моделей.

После формирования требований возникает проблема проверки их целостности и непротиворечивости, соответствия требований и целей, контроля функциональных требований в течение всего жизненного цикла бизнес-процессов предприятия.

Решение рассмотренной проблемы на основе предлагаемого подхода связано с выделением, проверкой и верификацией бизнес-правил как основы для функциональных требований.

Таким образом, выявление бизнес-правил является важной частью этапа формирования требований. Ошибочные, неполные, неоднозначные и противоречивые требования в значительной степени основываются на отсутствии или противоречивости бизнес-правил.

Следует отметить, что наряду с рассмотренными ранее особенностями, бизнес-правила выражают логику функционирования БП и бизнес-процедур на основе знаний. Например подпроцесс «определить продукт» или бизнес-процедура «предоставить скидку» подразумевают выполнение некоторых действий. Выбор тех или иных действий требуют дополнительных знаний в форме бизнес-правил.

Бизнес-правила определяют жизненный цикл продуктов, услуг и соответствующей инфраструктуры, а также бизнес-процессов. В соответствии с указанными правилами выполняются процессы закупок, продаж, приема на работу, производства, исследований, отчетности и планирования.

Использование бизнес-правил предоставляет ряд преимуществ, связанных с более точным и формальным выявлением требований к бизнес-процессам организации: независимость от используемой технической базы; непосредственная реализация правил в рамках автоматизации прикладных задач; обеспечение формализованного представления требований к качеству; уменьшение затрат при внесении изменений в бизнес-процес-

сы; обеспечение баланса между гибкостью и централизованным контролем в системе управления качеством.

После выявления бизнес-правил выполняется построение отображения бизнес-функций на БП. Такое отображение реализуется на основе суперпозиции  $Bpr$ , является способом проверки правильности бизнес-правил и позволяет перейти от сформулированных требований к реализации бизнес-функций и, в дальнейшем – к реализации бизнес-процессов.

Таким образом, применение суперпозиции бизнес-правил в соответствии с предлагаемым подходом эквивалентно применению оператора, обеспечивающего нахождение элементов БП (бизнес-объектов, процедур) на основе целей и бизнес-функций.

**Определение 1.** Оператором суперпозиции правил ( $P$ -оператором) называется процедура нахождения элементов бизнес-процессов  $E(BP_i)$  и их состояния по заданным бизнес-правилам  $Bpr^*$  на основе имеющейся бизнес-функции  $Bf_k$  и (или) предварительного описания бизнес-процесса  $BP_i$ :

$$E(BP_i) = P[Bf_k, BP_i, Bpr^*], \quad (2)$$

$$Bpr_i \subset Bpr, Bf_k \in Bf.$$

При первоначальном извлечении  $Bpr$  путем анализа документации и опроса менеджмента организации они обычно формулируются нестрого и неоднозначно. Поэтому на практике является целесообразной их декомпозиция на множество базовых  $Bpr$ , каждое из которых выражает единую мысль. Базовое бизнес-правило обладает следующими общими свойствами [1, 2]: имеет декларативную форму; выражает одну закономерность; обычно выражено в виде текста на естественном языке; ориентировано на потребности предприятия, а не на технологию реализации правила; выражает особенности функционирования предприятия.

При реализации функционального подхода к управлению предприятием бизнес-правила подразделяются на 6 категорий [3] в зависимости от их назначения: определение терминологии предметной области; определение взаимосвязи между объектами; формирование ограничений на деятельность предприятия; разрешение выполнения процедур; выделение закономерностей предметной области; формирование результирующих данных.

Реализация процессного подхода к управлению организацией потребовала разработки формальных моделей представления бизнес-правил. Такие модели отражают особенности процессного представления организации, особенности управления бизнес-процессами и направлены на переопределение и изменение объектов процесса во время его функционирования.

Проведенный анализ БП показал, что их ключевыми элементами являются: объекты, с которыми оперирует бизнес-процесс, а также процедуры (последовательности процедур). Следовательно, решение задачи управления и реорганизации БП требует формирования правил, переопределяющих указанные элементы бизнес-процесса и их взаимодействие. Иными словами, реализация отображения бизнес-функций на БП требует формирования моделей представления правил следующих четырех основных типов:

- определение объектов  $b_n$  бизнес-процесса;
- классификация объектов бизнес-процесса;
- управление бизнес-процедурами;
- определение структурных взаимосвязей между элементами БП.

Рассмотрим детальнее определение указанных типов правил.

**Определение 2.** Правилем определения объектов  $Bpr$  будем называть бизнес-правило, которое характеризуется следующими особенностями:

- аргументом правила является объект (или объекты) бизнес-процесса;
- результатом применения к правилу Р–оператора является определение нового объекта БП:

$$\begin{aligned} & \text{Пусть } \exists Bpr_l \in Bpr, \exists BP_i \in BP, \exists b_k, \\ & b_n \in B, b_k \neq b_n, l = \overline{1, L}, k = \overline{1, K}, n = \overline{1, N}, \\ & i = \overline{1, I}. \\ & \text{If } P[BP_i, Bpr_l(\{b_k\})] = b_n, \\ & \text{then } Bpr_l \in Bpr_{OO}, \end{aligned} \quad (3)$$

где  $BP_i$  – текущее описание бизнес-процесса;  $B$  – множество объектов  $BP$ .

Как видно из (3), каждый новый объект  $b_n$  бизнес-процесса определяется путем применения правила  $Bpr_l$  на основе множества уже используемых объектов  $\{b_k\}$ .

Практическое применение правил данного типа связано с возможностью дополнять перечень бизнес-объектов как при определении БП, так и во время его функционирования, что создает условия для формализации управления бизнес-процессами с изменяемой структурой.

Отметим, что предложенные правила определения объектов БП при процессном подходе к управлению соответствуют правилам определения терминов либо объектов, имеющих специфическое значение для деятельности предприятия при функциональном управлении. Текстовое представление таких правил имеет вид: «термин (объект)  $b_n$  определяется как термины (объекты)  $\{b_k\}$ ».

При процессном подходе к управлению предприятием взаимосвязи между объектами и характеристиками объекта определяются в рамках общей модели бизнес-процесса и обычно исполь-

зуются при реализации бизнес-процедур. Иными словами, характер обработки бизнес-объекта и используемая процедура в значительной степени определяются его характеристиками, что требует проведения классификации объектов БП.

**Определение 3.** Правилем классификации объектов бизнес-процесса  $Bpr_{KO}$  будем называть бизнес-правило, которое характеризуется следующими особенностями:

- аргументом правила является объект бизнес-процесса;
- результатом применения к правилу Р–оператора является подмножество объектов БП:

$$\begin{aligned} & \text{Пусть } \exists Bpr_l \in Bpr, \exists BP_i \in BP, \exists B_n \subset B, \\ & \exists b_k \in B, l = \overline{1, L}, k = \overline{1, K}, n = \overline{1, N}, i = \overline{1, I}. \\ & \text{If } P[BP_i, Bpr_l(b_k)] = B_n, \text{ then } Bpr_l \in Bpr_{KO}, \end{aligned} \quad (4)$$

где  $B_n$  – подмножество объектов бизнес-процесса  $BP_i$ .

Как видно из (4), правило классификации  $Bpr_l$  позволяет определить принадлежность некоторого объекта  $b_k$  к подмножеству  $B_n$  множества всех объектов бизнес-процесса  $B$ . Данный тип правил обеспечивает возможность стандартизации бизнес-процедур, оперирующих с объектами подмножеств  $B_n$ , что позволяет вводить и обрабатывать новые объекты БП во время его выполнения.

Вторым преимуществом правил данного типа является возможность выделения подмножеств объектов, находящихся в нештатном состоянии, и обработки их стандартным способом. Также правила данного типа отображают ассоциативные связи между объектами и позволяют построить иерархию объектов для рассматриваемых бизнес-процессов.

Таким образом, практическое применение правил классификации объектов направлено на использование унифицированных бизнес-процедур при обработке новых объектов БП. Классификация позволяет сопоставить признаки новых и уже существующих объектов и выбрать для их обработки соответствующие бизнес-процедуры. Это упрощает реорганизацию и управление БП с изменяемой структурой за счет использования типовых бизнес-процедур.

Например в результате применения правила указанного типа к объектам «продавец» и «поставщик» устанавливаются соответствия между этими объектами – то есть в рамках БП продавец играет также роль поставщика.

В текстовой форме такие правила имеют вид:

«объект(ы)  $B_n$  включают объект  $b_k$ »;

«объект  $b_k$  связан с объектами  $B_n$ ».

**Определение 4.** Правилем управления бизнес-процедурами  $Bpr_{yП}$  будем называть бизнес-прави-

ло, которое характеризуется следующими особенностями:

- аргументом правила является бизнес-процедура;
- результатом применения к правилу Р-оператора является запуск или останов соответствующей процедуры.

В соответствии с приведенным определением,  $Vpr_{уп}$  подразделяются на правила запуска и правила останова процедур:  $Vpr_{уп} = Vpr_{зп} \cup Vpr_{оп}$ , где  $Vpr_{зп}$ ,  $Vpr_{оп}$  – соответственно правила запуска и останова процедуры бизнес-процесса.

Текущее состояние процедуры определяется предикатом  $Executed(Br_j)$ , который принимает значение *true* для запущенной и *false* остановленной процедуры соответственно. Тогда правила запуска и останова процедур  $Br_j$  бизнес-процесса  $BP_i$  определяются следующим образом:

$$\begin{aligned} & \text{Пусть } \exists Vpr_l \in Vpr, \exists Br_j \in Br, \\ & \exists BP_i \in BP, Br \subset BP_i, \\ & l = \overline{1, L}, j = \overline{1, J}, i = \overline{1, I}. \\ & \text{If } Executed(Br_j) = false \wedge \\ & Executed(P[BP_i, Vpr_l(Br_j)]) = true \quad (5) \\ & \text{then } Vpr_l \in Vpr_{зп}. \\ & \text{If } Executed(Br_j) = true \wedge \\ & Executed(P[BP_i, Vpr_l(Br_j)]) = false \\ & \text{then } Vpr_l \in Vpr_{оп}, \end{aligned}$$

где  $Br$  – множество бизнес-процедур БП;  $Br_j$  – бизнес-процедура, входящая в состав рассматриваемого бизнес-процесса;  $Executed(Br_j)$  предикат, который принимает истинное значение в том случае, если процедура  $Br_j$  выполняется.

Рассмотренные правила  $Vpr_{зп}$  реализуют ограничения по управлению бизнес-процессами, а также описывают условия прерывания операций, процедур, транзакций, принятия решений и тому подобное. Например: «каждый заказ должен быть оформлен только одним покупателем».

Текстовое представление правил  $Vpr_{зп}$  имеет вид:

«если <УСЛОВИЕ>, то исполнителю  $Rl$  прервать выполнение процедуры  $Br_j$ »;

«выполнить процедуру  $Br_j$   $N$  раз».

Отметим, что при описании правил – ограничений в текстах соответствующих документов часто используются слова «должен, только тогда, при условии» и выражения вида «должен быть выполнен как минимум один раз».

Правила  $Vpr_{оп}$  реализуют разрешение некоторых действий, а также определяют условия выполнения заданных действий на предприятии. Простейшими примерами правил типа  $Vpr_{оп}$ , выраженного в текстовой форме, являются:

– «Если продавец поставляет материалы в течение более чем 6 месяцев, то необходимо установить специальные условия оплаты для данного продавца».

– «Если оборудование отработало межсервисный интервал, то необходимо провести сервисное обслуживание».

Типичные правила разрешения действия в текстовом виде формулируются следующим образом:

– «исполнитель  $Rl$  может выполнить процедуру  $Br_j$ »;

– при <СРОК > выполнить процедуру  $Br_j$ ;

– если <УСЛОВИЕ>, то выполнить процедуру  $Br_j$ .

С практической точки зрения правила рассмотренного типа обеспечивают возможность корректировки структуры бизнес-процесса путем добавления новых либо запрета выполнения существующих процедур БП. При этом корректировка БП возможна и во время его функционирования.

**Определение 5.** Правилom  $Vpr_{сбп}$  определения структурных взаимосвязей между элементами БП будем называть бизнес-правило, которое характеризуется следующими особенностями:

– аргументом правила является объект, бизнес-процедура либо последовательность процедур ;

– результатом применения к правилу Р-оператора является элемент бизнес-процесса включающий аргумент правила в качестве своей составной части.

Формально представление правил определения структурных взаимосвязей имеет вид:

$$\begin{aligned} & \text{Пусть } \exists Vpr_l \in Vpr, \exists Br_j \in Br, \exists b_k \in B, \\ & \exists BP_i \in BP, Wf'(Br_j), Wf(Br_j) \in BP_i, \quad (6) \\ & l = \overline{1, L}, j = \overline{1, J}, k = \overline{1, K}, i = \overline{1, I}. \end{aligned}$$

$$\begin{aligned} & \text{If } P[BP_i, (Vpr_l(Br_j))] = Wf(Br_j), \\ & \text{then } Vpr_l \in Vpr_{сбп}. \quad (7) \end{aligned}$$

$$\begin{aligned} & \text{If } P[BP_i, Vpr_l(Wf'(Br_j))] = Wf(Br_j), \\ & \text{then } Vpr_l \in Vpr_{сбп}, \quad (8) \end{aligned}$$

где  $Wf'(Br_j)$ ,  $Wf(Br_j)$  – последовательности выполнения процедур  $Br_j$  процесса  $BP_i$ .

Исходя из представления правил (6) – (8), можно сделать вывод, что  $Vpr_{сбп}$  определяют иерархию элементов БП. Действительно, правила вида (6) устанавливают взаимосвязи между бизнес-объектом  $b_k$  и обрабатывающей его процедурой  $Br_j$ . Правила вида (7) определяют принадлежность процедуры  $Br_j$  последовательности процедур бизнес-процесса  $Wf(Br_j)$ , что задает выполнение  $Br_j$  при реализации данного БП. Правила вида (8) определяют иерархию последовательностей процедур для

заданного процесса. В частности последовательность  $Wf'(Br_j)$  входит в состав последовательности  $Wf(Br_j)$ .

Отметим, что правила вида (8) позволяют реализовать выделение подпроцессов рассматриваемого БП. Это дает возможность согласовать процессное описание предприятия с его иерархической организационной структурой на основе выделения иерархии подпроцессов.

С практической точки зрения бизнес-правила рассматриваемого типа позволяют определить структуру бизнес-процесса. Изменение правил приводит к изменению БП. Следовательно,  $Vpr_{СБП}$  позволяют сформировать гибкую реконфигурируемую структуру процесса, что обеспечивает возможность автоматизированного управления бизнес-процессом с изменяемой структурой и его реконфигурирование во время функционирования.

Примером правила типа  $Vpr_{СБП}$  является правило вида: «Если поставщик поставил материала более, чем на указанную сумму, его необходимо перевести в категорию постоянных поставщиков».

Текстовое представление бизнес-правил определения структурных взаимосвязей между элементами БП имеет следующий вид:

- «обработка процедурой  $Br_j$  выполняется на основе данных  $b_k$ »;
- «объект  $b_k$  определяется при условии выполнения процедуры  $Br_j$ »;
- «если выполнена процедура  $Br_j$ , то необходимо выполнить  $Br_l$ ».

Таким образом, в соответствии с разработанным подходом к представлению бизнес-правил, последние определяют функциональные требования к БП, особенности реализации, усовершенствования и реинжиниринга бизнес-процессов. Бизнес-правила состоят из правил определения, классификации объектов, управления бизнес-процедурами и определения структурных взаимосвязей между элементами БП:  $Vpr = Vpr_{ОО} \cup Vpr_{КО} \cup Vpr_{УП} \cup Vpr_{СБП}$ .

### Выводы

Разработанные модели представления знаний в форме бизнес-правил формализуют определение и классификацию объектов бизнес-процесса, управление бизнес-процедурами, а также определение структурных взаимосвязей между элементами бизнес-процесса. Полученные модели представления знаний обеспечивают формализацию функциональных требований и, соответственно, создают условия для целенаправленного изменения компонентов бизнес-процесса с изменяемой структурой, а также взаимосвязей между этими компонентами.

**Список литературы:** 1. Guide Business Rules Project, Final Report, 11/95. <http://www.businessrulesgroup.org/brgactv.htm>. 2. Gottesdiener E. Turning rules into requirements // Application Development Trends, 1999. – [http://www.adtmag.com/section.asp?section=tools/Turning rules into requirements.htm](http://www.adtmag.com/section.asp?section=tools/Turning%20rules%20into%20requirements.htm). 3. Gottesdiener E. Business RULES Show, Power, Promise., Application Development Trends, Vol 4, nro. 3 (March 1997). P. 36.

*Поступила в редколлегию 14.02.2008*

УДК 681.324: 519.713



## ЭВРИСТИЧЕСКОЕ МНОГОАЛЬТЕРНАТИВНОЕ ПРОЕКТИРОВАНИЕ КОМПЬЮТЕРНЫХ СИСТЕМ ОБРАБОТКИ ДАННЫХ

В.П. Авраменко<sup>1</sup>, В.В. Тулупов<sup>2</sup><sup>1</sup> ХНУРЭ, г. Харьков, Украина<sup>2</sup> Харьковский национальный университет внутренних дел, г. Харьков, Украина

Исследовано применение информационных технологий многоальтернативного проектирования компьютерных систем обработки данных (КСОД) в условиях неопределенности. Показано, что фактическое проектирование КСОД ведется, как правило, по совокупности количественных и качественных технико-экономических показателей, выступающих в качестве локальных критериев эффективности. Полученные результаты демонстрируются на конкретном тестовом примере выбора рационального аппаратно-программного компьютерного интерфейса КСОД.

СИСТЕМА ОБРАБОТКИ ДАННЫХ, МНОГОАЛЬТЕРНАТИВНОЕ МОДЕЛИРОВАНИЕ, МЕТОД АНАЛИЗА ИЕРАРХИЙ, КОЭФФИЦИЕНТ ПРЕВОСХОДСТВА

### Введение

**Постановка проблемы.** Существующая практика проектирования компьютерных систем обработки данных часто (КСОД) сталкивается с фактом, когда материализованная совокупность заданных качественных и количественных показателей, представляющая множество заданных требований, часто превышает общую совокупность выделенных ресурсов. Проектирование в этом случае сводится к нахождению альтернативных компромиссных решений, позволяющих сбалансировать множество заданных требований (их более чем достаточно) с совокупностью выделенных ресурсов (они всегда ограничены).

На начальных этапах проектирования КСОД допускается применение эвристических методов многоальтернативного выбора структуры и компонентного состава создаваемого изделия – компьютерного интерфейса. Упрощенные многоальтернативные методы оптимизации позволяют выбрать предпочтительные проектные решения по совокупности количественных и качественных показателей, выступающих в качестве локальных целевых функций. Решение такого рода задач требует разработки конструктивных моделей и инструментальных средств, ориентированных на многоальтернативные технологии проектирования КСОД.

**Анализ состояния проблемы** проектирования различных информационных объектов и телекоммуникационных систем свидетельствует об актуальности многоальтернативного подхода к принятию проектных и управленческих решений [1–3]. В известных источниках информации практически отсутствуют методологии решения задач эвристического многоальтернативного проектирования.

**Целью работы** является исследование многоальтернативных процедур принятия проектных решений, разработка альтернативных моделей и технологий проектирования КСОД, оценивание

человеческого и технико-экономических факторов в принятии проектных решений.

### 1. Постановка и решение задачи исследования

**Принцип многоальтернативности моделирования.** Формализация задач проектирования предопределяется принципами абстрагирования и многоальтернативности, набором инструментальных средств для построения моделей и наличием вычислительных методов для решения задач рассматриваемого класса. Согласно принципу абстрагирования в модель включаются те компоненты проектируемой системы, которые имеют непосредственное отношение к выполнению системой своих функций или своего целевого назначения. При этом все второстепенные детали опускаются, чтобы чрезмерно не усложнять процесс анализа и синтеза полученной модели.

Принцип многоальтернативности моделирования компьютерных систем состоит в том, что никакая единственная модель не может в достаточной степени адекватно описать все аспекты функционирования системы. В достаточной степени адекватное математическое описание системы можно получить путем построения множества моделей, каждая из которых составлена под определенным углом зрения и отражает некоторый аспект поведения или структуры системы.

Многоальтернативность моделирования можно реализовать различными методами. Одним из них является принцип иерархичности, который предусматривает построение иерархической совокупности моделей. При этом исходная модель системы имеет наиболее общее представление. Такая метамодель строится на начальном этапе проектирования и может не содержать многих деталей и аспектов поведения или структуры системы. Каждый следующий уровень иерархии содержит более конкретные модели концептуального, логического

и физического представления проектируемой кибернетической системы.

*Модель задачи многоальтернативного проектирования* в условиях неопределенности может быть представлена следующим выражением:

$\langle \text{МП, МК, МО, МН, ММ}_1, \text{ММ}_2, \text{МА, ПВ} \rangle,$

где МП – множество постановок решаемой задачи; МК – множество критериев решаемой задачи; МО – множество ограничений решаемой задачи; МН – множество видов проявляемых неопределенностей; ММ<sub>1</sub> – множество моделей решаемой задачи; ММ<sub>2</sub> – множество вычислительных методов, алгоритмов и процедур для решения сформулированных задач; МА – множество альтернативных решений; ПВ – процедуры выработки предпочтительных решений из множества допустимых решений.

При выработке и принятии многоальтернативных решений используются эвристические процедуры, методы линейного и целочисленного программирования, методы многоальтернативной и многокритериальной оптимизации. К числу сходных задач относятся задачи о ранце и наименьшем покрытии. Однако поскольку эти задачи относятся к NP-полным, то сложность их решения методом ветвей и границ соизмерима со сложностью решения задачи эвристического многоальтернативного проектирования всей КСОД.

Метод многоальтернативной оптимизации предусматривает декомпозицию исходной задачи на отдельные составляющие, обеспечивая структурирование и упрощение ее с выделением иерархии, содержащей различные подцели, частные критерии и альтернативные мероприятия, подлежащие оценке. Относительная значимость различных количественных и качественных критериев определяется в отдельности для каждого вышестоящего элемента путем сопоставления и анализа пар альтернативных вариантов. Таким же образом в модель интегрируется фактор выгодности мероприятий.

*Метод анализа иерархий* [1–5] предназначен для принятия многоальтернативных проектных решений в условиях слабой структурированности КСОД и неопределенности исходной информации, заданной набором количественных и качественных зависимостей. Причинами неопределенности являются неполнота знаний эксперта о свойствах объектов; недостаточная уверенность лица, принимающего решение, в правильности своих оценок; противоречивость знаний; нечеткость представления информации. Наличие неопределенности приводит к возникновению ошибок в экспертных оценках, несогласованности данных и нарушению основных свойств суждений, таких как связность и транзитивность.

Обоснованность и достоверность принимаемых решений зависят от согласованности экспертных мнений, которые выражаются через свойства связности и транзитивности между экспертными оценками исходного факторного пространства. Отсутствие этих свойств в системе предпочтений не позволяет осуществить однозначный выбор на множествах критериев и альтернатив принимаемых решений. В практических задачах для повышения качества и обоснованности решений следует сначала восстановить указанные свойства, а затем согласовывать мнения экспертов.

В основу метода анализа иерархий заложен принцип декомпозиции исходной проблемы совокупностью более простых составляющих, называемых иерархическими уровнями или иерархиями. Составляющие проблемы в зависимости от системного назначения делятся на объекты-критерии и объекты-альтернативы принимаемых решений. Из объектов-критериев организуется иерархическая структура, содержащая уровни целей, подцелей, целевых функций, а из объектов-альтернатив создается иерархическая структура, отражающая соподчиненность иерархий принимаемых решений. В результате декомпозиции образуется архитектура проблемы, отражающая относительную степень взаимосвязи объектов иерархии.

Многоальтернативное решение задач проектирования компьютерных систем реализуется как процесс поэтапного установления приоритетов. На первом этапе выявляются наиболее важные составляющие задачи; на втором этапе выбирается лучший способ проверки наблюдений, испытания и оценки элементов; на третьем этапе осуществляется выработка и принятие многоальтернативного решения с оценкой его качества. Процесс подвергается многократному переосмыслению до тех пор, пока не будет уверенности, что он охватил все основные характеристики, необходимые для представления и решения задачи. Процесс решения задачи является итерационным над последовательностью иерархий, то есть результаты предыдущей итерации используются в качестве входных данных при поиске следующей иерархии.

*Алгоритм метода анализа иерархий*, используемый в задачах многокритериальной оптимизации, содержит следующие этапы:

1. Содержательная постановка задачи принятия многокритериальных проектных решений в условиях неопределенности.

2. Математическая постановка задачи принятия многокритериальных проектных решений в условиях неопределенности включает в себя формирование иерархической структуры обобщенного критерия эффективности в виде соподчиненных уровней целей, подцелей и целевых функций; математическое описание функциональных зави-

симостей и параметрических ограничений задачи принятия многокритериальных проектных решений в условиях неопределенности; формирование иерархической структуры взаимосвязи альтернатив принимаемых решений.

3. Ранжирование конечного множества объектов-критериев и объектов-альтернатив принимаемых решений

$$p = \{p_1, \dots, p_i, \dots, p_m\}$$

по важности путем задания вектора весовых коэффициентов

$$\alpha = \{\alpha_1, \dots, \alpha_i, \dots, \alpha_m\},$$

значения которых удовлетворяют ограничениям

$$\sum_{i=1}^m \alpha_i = 1, \quad \alpha_i \geq 0.$$

В многоуровневой иерархической системе принятия решений ранжирование по важности каждого  $k$ -го уровня множества объектов-критериев и множества объектов-альтернатив

$$p^k = \{p^k_1, \dots, p^k_i, \dots, p^k_m\}$$

производится путем задания  $k$  векторов весовых коэффициентов

$$\alpha^k = \{\alpha^k_1, \dots, \alpha^k_i, \dots, \alpha^k_{m_k}\}, \quad k = \overline{1, K},$$

где  $K$  – количество уровней иерархической структуры объектов-критериев и структуры объектов-альтернатив;  $m_k$  – количество объектов на  $k$ -м уровне критериев и  $k$ -м уровне альтернатив.

Задача ранжирования объектов по важности в пределах каждого  $k$ -го уровня иерархии состоит в том, чтобы на основании опроса экспертов и методов обработки экспертных данных установить множество соотношений

$$p^k_i \rightarrow \alpha^k_i$$

для всех уровней иерархической структуры критериев и иерархической структуры альтернатив. Эти суждения позволяют перевести качественные характеристики в количественные или числовые зависимости.

4. Формирование матрицы парных сравнений  $[S^k_{p \times m}]$  для каждого  $k$ -го уровня множества объектов (рис. 1).

$$[S^k_{p \times m}] = \begin{matrix} & p^k_1 & \dots & p^k_j & \dots & p^k_m \\ \begin{matrix} p^k_1 \\ \dots \\ p^k_i \\ \dots \\ p^k_m \end{matrix} & \left| \begin{matrix} \alpha^k_{11} & \dots & \alpha^k_{1j} = \alpha^k_1 / \alpha^k_j & \dots & \alpha^k_{1m} \\ \dots & \dots & \dots & \dots & \dots \\ \alpha^k_{i1} & \dots & \alpha^k_{ij} = \alpha^k_i / \alpha^k_j & \dots & \alpha^k_{im} \\ \dots & \dots & \dots & \dots & \dots \\ \alpha^k_{m1} & \dots & \alpha^k_{mj} = \alpha^k_m / \alpha^k_j & \dots & \alpha^k_{mm} \end{matrix} \right. \end{matrix}$$

Рис. 1. Структура исходной матрицы парных сравнений

Матрицы парных сравнений  $[S^k_{p \times m}]$  являются основным хранилищем информации, необходимой для принятия многоальтернативных решений. Каждая матрица  $[S^k_{p \times m}]$  составляется по таким правилам:

- Мнение каждого эксперта записывается в виде строки матрицы парных сравнений  $[S^k_{p \times m}]$ .

- Эксперт должен уметь быстро отвечать на вопросы: во сколько раз весовой коэффициент  $\alpha_i$  больше весового коэффициента  $\alpha_j$  или во сколько раз весовой коэффициент  $\alpha_j$  меньше коэффициента  $\alpha_i$ .

- Каждый элемент  $\alpha^k_{ij}$  матрицы парных сравнений  $[S^k_{p \times m}]$  определяется выражением  $\alpha^k_{ij} = \alpha^k_i / \alpha^k_j$ , где  $\alpha^k_i$  и  $\alpha^k_j$  – весовые коэффициенты приоритетности объектов парной связности  $k$ -го уровня иерархии:

$$p^k_i \rightarrow \alpha^k_i, \quad p^k_j \rightarrow \alpha^k_j; \quad \alpha^k_i, \alpha^k_j \rightarrow \alpha^k_{ij} = \alpha^k_i / \alpha^k_j.$$

Размерности парируемых коэффициентов  $\alpha^k_i$  и  $\alpha^k_j$  должны быть одинаковыми, а значения этих коэффициентов не допускают деление на нуль. Если отношение  $\alpha^k_i / \alpha^k_j > 1$ , то объект  $p_i$  считается важнее объекта  $p_j$ . Полученные значения весовых коэффициентов являются оценками в шкале отношений.

5. Поиск решения задачи многокритериальной оптимизации осуществляется путем поэтапного установления приоритетов. На первом этапе выявляются наиболее важные объекты решаемой проблемы, на втором изыскивается наилучший способ проверки наблюдений, испытания и оценки объектов. На последующих этапах осуществляется выработка рационального решения и оценивание его качества.

**Проверка согласованности экспертных мнений** является исходной предпосылкой метода анализа иерархий. Для определения согласованности экспертных мнений используется исходная матрица парных сравнений, полученная путем опроса экспертов методом парных сравнений в шкале Саати [1, 2]. В качестве меры согласованности чаще всего используется индекс согласованности и отношение согласованности [2, 3]. Согласованность обратно симметричной исходной матрицы парных сравнений эквивалентна требованию равенства ее максимального собственного значения  $\lambda_{\max}$  числу сравниваемых объектов  $n$ , то есть  $\lambda_{\max} = n$ .

В качестве меры рассогласования (несогласованности) принято рассматривать нормированное отклонение  $\lambda_{\max}$  от  $n$ , называемое индексом согласованности:

$$ИС = \frac{\lambda_{\max} - n}{n - 1}.$$

Чтобы оценить степень согласованность экспертных мнений, индекс согласованности (ИС)

сравнивают со случайным индексом (СИ). Случайным индексом называют индекс согласованности, рассчитанный для квадратной  $n$ -мерной положительной обратнo симметричной матрицы, элементы которой сгенерированы датчиком случайных чисел для интервала значений от 1 до 9.

Для матрицы с фиксированным значением индекс рассчитывается как среднее значение для выборки  $N = 100$ . В табл. 1 представлены средне-квадратичные значения согласованности для случайных матриц порядка от 1 до 10.

Таблица 1

Величина случайной согласованности

Размер матрицы	1	2	3	4	5	6	7	8	9	10
Случайная согласованность	0	0	0,58	0,9	1,12	1,24	1,32	1,41	1,45	1,49

Получив индекс согласованности и выбрав из табл. 1 случайный индекс для заданного порядка матрицы, определяют отношение согласованности (ОС):

$$ОС = ИС / СИ.$$

Если величина  $ОС \leq 0,1$ , то степень согласованности экспертных данных считается приемлемой. В противном случае (если  $ОС > 0,1$ ) эксперту рекомендуется пересмотреть свои суждения. Для этого необходимо выявить те позиции в матрице суждений, которые вносят максимальный вклад в величину отношения согласованности, и попытаться изменить меру несогласованности в меньшую сторону на основе более глубокого анализа вопроса.

**Модель принятия многоальтернативного проектного решения** может быть представлена совокупностью целевых функций  $f_i, i=1, m$  и набором альтернатив принимаемых решений  $x = \{x_j\} \subset X, j=1, n$  в виде [4, с. 126 - 133]:

$$f_i(x_j) \rightarrow \max_{x \in X}, \quad i=1, m, \quad j=1, n,$$

где  $m$  – количество целевых функций;  $x = \{x_1, x_2, \dots, x_n\} \subset X$  – конечное множество альтернатив при-

нимаемых решений, содержащее  $n$  элементов  $x_j$ . Значения чисел  $m$  и  $n$  должны быть относительно невелики, поскольку именно они в МАИ определяют трудоемкость диалоговых процедур реального масштаба времени по извлечению дополнительной информации о задаче.

После реализации логико-семантической процедуры определения частных критериев используется метод линейной свертки

$$J(x_j) = \sum_{i=1}^m \alpha_i f_i(x_j), \quad j=1, n$$

для получения исследуемых на оптимальность альтернатив принимаемых решений  $x_1 = A_1, x_2 = A_2, x_3 = A_3$ .

## 2. Пример многоальтернативного проектирования компьютерных систем обработки информации (КСОИ)

**Постановка задачи.** Требуется спроектировать наиболее предпочтительный альтернативный вариант КСОИ по совокупности заданных технико-экономических показателей [5], приведенных в табл. 2.

В табл. 2 приведены характеристики альтернатив принимаемых решений:  $A_1$  – система «Атлант»,  $A_2$  – система «Ай-Ти»,  $A_3$  – система «Мир-АСР»,  $A_4$  – система «Орёл-М»,  $A_5$  – система «Энран». В качестве частных критериев выступают технико-экономические показатели:  $f_1(x)$  – гибкость,  $f_2(x)$  – производительность,  $f_3(x)$  – аппаратная платформа,  $f_4(x)$  – надёжность,  $f_5(x)$  – экономическая эффективность.

Глобальный показатель качества, характеризующий правильность выбора структуры и компонентного состава системы, обычно связывается с ее функциональными, техническими и экономическими характеристиками. Такой показатель является достаточно расплывчатым и неопределенным, поэтому можно воспользоваться набором критериев-заместителей, которые позволяют заменить исходную слабо формализованную задачу другой хорошо структурированной многоальтернативной задачей.

Таблица 2

Сравнительный анализ альтернативных вариантов КСОИ

Частные критерии	Альтернативные варианты КСОИ				
	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$
$F_1(x)$	Легкая адаптация	Сравнительно гибкая	Действительно гибкая	Сравнительно легкая	Сравнительно гибкая
$F_2(x)$	До 2 млн. абонентов	До 1 млн. абонентов	Без ограничений	500 – 500000 абонентов	Без ограничений
$F_3(x)$	ОС Windows NT, Linux	ОС Windows NT, UNIX	ОС Windows NT	Windows NT	Windows NT, UNIX
$F_4(x)$	Очень высокая	Сравнит. высокая	Высокая надежность	Высокая надежность	Сравнительно высокая
$F_5(x)$	Высокий	Не высокий	Достаточно высокий	Не высокий	Сравнительно высокий

Исходный глобальный критерий эффективности решаемой задачи аппроксимирован набором локальных технико-экономических показателей качества, каждый из которых подлежит максимизации:

- $f_1(x)$  – гибкость системы, которая представляет собой способность быстро приспосабливаться к изменяющимся обстоятельствам, потребностям любого оператора связи, возможности использования системы как в самостоятельном режиме, так и в составе эксплуатируемой КСОИ;

- $f_2(x)$  – производительность, масштабируемость и открытость системы, состоящие в том, что при росте числа абонентов и дополнительных услуг не должна возникать необходимость доработки программного обеспечения, то есть должна существовать возможность расширения конфигурации КСОИ;

- $f_3(x)$  – аппаратная платформа, представляющая собой набор согласованных между собой технических средств, обеспечивающих решение заданных функциональных задач с требуемыми качественными показателями;

- $f_4(x)$  – надёжность, характеризующая интегрированное свойство инструментальных аппаратно-программных средств системы сохранять работоспособность в процессе их эксплуатации;

- $f_5(x)$  – экономическая эффективность (экономичность) от внедрения выбранного варианта аппаратно-программных средств проектируемой КСОИ.

**Алгоритм решения задачи** многоальтернативно-го проектирования КСОИ методом анализа иерархий содержит следующие этапы:

1. Формирование иерархической структуры частных критериев задачи многоальтернативного проектирования КСОИ, которая включает следующие этапы: 1) установить симметрично по центру фокус иерархии; 2) уровнем ниже расположить частные критерии и подкритерии; 3) ниже последнего уровня подкритериев разместить уровень альтернатив принимаемых решений.

2. Формирование матрицы парных сравнений  $[S_{f_{m \times m}}]$  частных критериев задачи многоальтернативного проектирования, которое включает следующие этапы:

- 1) задать размерность матрицы парных сравнений  $[S_{f_{m \times m}}]$ , равную  $5 \times 5$  (по числу критериев);

- 2) назвать строки и столбцы матрицы  $[S_{f_{m \times m}}]$  именами частных критериев;

- 3) присвоить каждому диагональному элементу матрицы  $[S_{f_{5 \times 5}}]$  значение, равное единице;

- 4) назначить относительную важность элементов матрицы  $[S_{f_{5 \times 5}}]$  по шкале предпочтений Саати;

- 5) принять значения недиагональных элементов матрицы  $[S_{f_{5 \times 5}}]$  по результатам экспертных опросов ( $\alpha_{12} - ?$ ,  $\alpha_{23} - ?$ ,  $\alpha_{34} - ?$ ,  $\alpha_{45} - ?$ ) с последующим

вычислением всех недостающих коэффициентов по формулам:

$$\begin{aligned} \alpha_{ij} &= 1/\alpha_{ji}, \quad \alpha_{ij} = \alpha_i/\alpha_j, \quad \alpha_{ij} = \\ &= \alpha_{ik} \times \alpha_{kj}, \quad \frac{\alpha_i \cdot \alpha_k}{\alpha_k \cdot \alpha_j} = \frac{\alpha_i}{\alpha_j}; \end{aligned}$$

- б) вычислить элементы  $v_{f_{ij}}$  нормированной матрицы парных сравнений частных критериев  $[N_{f_{m \times m}}]$  следующим образом: соответствующий элемент  $\alpha_{f_{ij}}$  исходной матрицы парных сравнений частных критериев  $[S_{f_{m \times m}}]$  разделить на сумму элементов  $j$ -го столбца, то есть элементы первого столбца матрицы  $[S_{f_{m \times m}}]$  необходимо разделить на сумму элементов первого столбца, элементы второго столбца – на сумму элементов второго столбца и так далее по формуле:

$$v_{f_{ij}} = \alpha_{f_{ij}} / \sum_{i=1}^m \alpha_{f_{ij}}; \quad j = \overline{1, m};$$

- 7) вычислить относительные значения весовых коэффициентов  $v_{f_1} - v_{f_5}$  частных критериев  $f_1(x) - f_5(x)$  как средние значения элементов соответствующих строк нормированной матрицы парных сравнений  $[N_{f_{5 \times 5}}]$  по формуле:

$$v_{f_i} = \frac{1}{m} \sum_{j=1}^m v_{f_{ij}}; \quad i = \overline{1, m};$$

- 8) результаты вычисления относительных значений весовых коэффициентов  $v_{f_i}$  и усредненных значений весовых коэффициентов  $v_{f_i}$  частных критериев занести в нормированную матрицу парных сравнений  $[N_{f_{5 \times 5}}]$  и в столбец  $v_{f_i}$ , присоединенный справа к этой матрице.

**Определение коэффициентов превосходства альтернатив** производится по схеме, аналогичной определению коэффициентов превосходства частных критериев. Для каждого частного критерия ( $f_1$  – гибкость,  $f_2$  – производительность,  $f_3$  – аппаратная платформа,  $f_4$  – надежность,  $f_5$  – экономичность) строится матрица парных сравнений  $[S_{f_i_{n \times n}}]$ . Для построения такой матрицы попарно сравнивается альтернатива строки с альтернативой столбца по отношению к одному исследуемому частному критерию. Матрица  $[S_{f_i_{n \times n}}]$  имеет размерность  $5 \times 5$  (по числу альтернатив).

Для оценки фактора производительности строится матрица  $[S_{f_2_{5 \times 5}}]$ . При этом попарно сравнивается альтернатива строки с альтернативой столбца по отношению к производительности каждой из систем. Никакие другие критерии при этом не учитываются. Значения из шкалы относительной важности вписываются в ячейки, образованные пересечением соответствующей строки и столбца. Диагональ этой матрицы заполняется значением «1», а ячейки, лежащие ниже диагонали – обратными значениями. Нормализация матрицы парных сравнений альтернатив по критерию произво-

длительности [ $Sf_{2 \times 5}$ ] выполняется по приведенной выше процедуре.

Результаты вычисления относительных элементов заносятся в нормированную матрицу парных сравнений [ $Nf_{2 \times 5}$ ]. Вычисленные значения весовых коэффициентов частных критериев заносятся в столбец  $vf_{2A_j}$ , который присоединяется справа к матрице [ $Nf_{2 \times 5}$ ].

Значения весовых коэффициентов критерия производительности (столбец  $vf_{2A_j}$ ) свидетельствуют, что наиболее предпочтительными альтернативами по критерию производительности являются системы «Атлант» (вариант  $A_1$ , весовой коэффициент  $vf_{2A_1} = 0.439$ ) и «Ай-Ти» (вариант  $A_2$ , весовой коэффициент  $vf_{2A_2} = 0.206$ ).

Значение коэффициента относительной согласованности исходной матрицы [ $Sf_{2 \times 5}$ ] равно 8,15 %, что меньше граничного допустимого значения 10%, за пределами которого требуется пересматривать суждения экспертов или менять их состав.

Для оценки фактора экономичности системы построены матрицы исходных парных сравнений [ $Sf_{5 \times 5}$ ] и нормированных парных сравнений. Значения весовых коэффициентов критерия экономичности (столбец  $vf_{5A_j}$ ) свидетельствуют о том, что наиболее предпочтительными альтернативами по критерию экономичности является система «Мир-АСР» (вариант  $A_3$ , весовой коэффициент  $vf_{5A_3} = 0,472$ ). Она значительно выгоднее всех остальных систем. Значение относительной согласованности матрицы [ $Sf_{5 \times 5}$ ] равно 1,50 %, что меньше граничной величины 10%, не позволяющей считать, что мнения экспертов согласованы.

**Формирование набора предпочтительных весовых коэффициентов** превосходства альтернатив принимаемых решений осуществляется на основании информации о значениях весовых коэффициентов частных критериев  $vf_i$  и весовых коэффициентов альтернатив принимаемых решений (по характеристикам КСОД) относительно каждого из частных критериев (гибкость, производительность, экономичность и другие)  $vf_i(A_j)$ , которые приведены в табл. 3.

Таблица 3

Результирующие значения глобального приоритета

Альтернативы	Веса частных критериев $vf_i(A_j)$					Вес $v^*A_j$
	$vf_1$	$vf_2$	$vf_3$	$vf_4$	$vf_5$	
$A_1$	0,414	0,439	0,303	0,448	0,059	0,3886
$A_2$	0,135	0,206	0,064	0,211	0,179	0,1817
$A_3$	0,074	0,157	0,164	0,064	0,472	0,1452
$A_4$	0,241	0,037	0,164	0,064	0,194	0,1107
$A_5$	0,135	0,157	0,303	0,211	0,095	0,1717
$vf_i$	0,201	0,321	0,061	0,311	0,106	

Значение предпочтительного весового коэффициента каждой отдельно взятой альтернативы принимаемого решения  $v^*A_j$  определяется как сумма произведений значений компонентов вектора приоритета частных критериев  $vf_i$  на значения компонентов вектора локального приоритета  $vf_i(A_j)$  рассматриваемой альтернативы  $A_j$  в отношении данного критерия  $f_i$ , а именно

$$v^*A_j = \sum_{i=1}^m vf_i \times vf_{i(A_j)}, \quad j = \overline{1, n}.$$

Например, вычисление предпочтительного весового коэффициента превосходства альтернатив  $v^*A_1$  для альтернативы  $A_1$  (КСОД «Атлант») выполняется следующим образом:

$$\begin{aligned} v^*A_1 &= vf_1 \times vf_{1A_1} + vf_2 \times vf_{2A_1} + vf_3 \times vf_{3A_1} + \\ &+ vf_4 \times vf_{4A_1} + vf_5 \times vf_{5A_1} = 0,201 \times 0,414 + \\ &+ 0,321 \times 0,439 + 0,061 \times 0,303 + 0,311 \times 0,448 + \\ &+ 0,106 \times 0,059 = 0,3886. \end{aligned}$$

**Многоальтернативный выбор КСОД с приоритетным локальным критерием «Экономичность».** Результаты парных сравнений по критерию «Экономичность» строится матрица [ $S_{\Sigma f_i 5 \times 5}$ ]. Индекс согласованности этой матрицы как отношение согласованности равен (ОС) = 9,26 % < 10%.

Значения приоритета каждой из альтернатив приведены в табл. 4.

Таблица 4

Результирующие значения глобального приоритета при приоритете критерия «Экономичность»

Альтернативы	Веса частных критериев $vf_i(A_j)$					Вес $v^*A_j$
	$vf_1$	$vf_2$	$vf_3$	$vf_4$	$vf_5$	
$A_1$	0,414	0,439	0,303	0,448	0,059	0,1797
$A_2$	0,135	0,206	0,064	0,211	0,179	0,1789
$A_3$	0,074	0,157	0,164	0,064	0,472	0,3519
$A_4$	0,241	0,037	0,164	0,064	0,194	0,1652
$A_5$	0,135	0,157	0,303	0,211	0,095	0,1241
$v_{\Sigma f_i}$	0,078	0,115	0,032	0,102	0,670	

Наилучшей средневзвешенной альтернативой проекта КСОИ оказалась альтернатива «Мир-АСР» с максимальным значением глобального приоритета. Однако на практике более предпочтительной альтернативой, рекомендуемой к выбору, оказалась КСОИ «Атлант» с максимальным значением глобального приоритета, несмотря на ее самую высокую стоимость. Если у заказчика не хватает финансовых средств на приобретение КСОИ «Атлант», то выбор более дешевого варианта системы осуществляется путем пересчета всех таблиц с учетом привлечения дополнительной информации о новых требованиях к разрабатываемому проекту и согласованности экспертных мнений.

### Выводы

Многоальтернативное проектирование находит широкое применение в различных сферах человеческой деятельности, таких как разработка компьютерных интерфейсов обработки данных, инжиниринг крупномасштабных корпоративных систем в условиях риска и неопределенности, оптимизация инфотелекоммуникационных систем по совокупности технико-экономических показателей качества, создание Web-ориентированных систем подготовки и переподготовки кадров, имитационное моделирование процесса контроля знаний в системе дистанционного обучения.

Если нет оснований считать множество достижимости рассматриваемой многокритериальной задачи выпуклым, то вместо линейной свертки в качестве обобщенного критерия целесообразно использовать свертку Джоффриона, основанную на комбинации линейной и максиминной свертки.

Помимо исследованного в работе метода анализа иерархий на практике для принятия многоаль-

тернативных решений в условиях неопределенности часто применяются методы теории полезности и теории ценности со многими признаками, что имеет место при проектировании компьютерных распределенных корпоративных систем. Эти методы исходят из эвристических предположений и естественной способности людей думать логически и творчески, определять события и устанавливать отношения между ними.

**Список литературы:** 1. Саати Т. Принятие решений. Метод анализа иерархий. — М.: Радио и связь. 1989. — 316 с. 2. Саати Т., Кернс К. Аналитическое планирование. Организация систем. — М.: Радио и связь. 1991. — 224 с. 3. Таха Х. Введение в исследование операций. — М.: Вильямс, 2001. — 912 с. 4. Черноуцкий И.Г. Методы принятия решений. — СПб.: БХВ-Петербург, 2005. — 416 с. 5. Авраменко В.П., Колесников О.А., Калачева В.В., Юрьева Н.Н. Технология многоальтернативного выбора информационных систем на основе метода анализа иерархий // Нові технології. Вісник Інституту економіки та нових технологій. — 2002. — Вип. 1. — С. 32-38.

*Поступила в редколлегию 12.03.2008*

УДК 658.012.56



## МОДЕЛИРОВАНИЕ МОРФОЛОГИИ ЕСТЕСТВЕННОГО ФЛЕКТИВНОГО ЯЗЫКА

Т.П. Любченко

Украинский языково-информационный фонд НАН Украины, г. Киев, Украина, ltp@i.com.ua

Построена формальная модель словоизменения флективного языка. При моделировании процессов словоизменения использован формализм мультимножеств. Изложенная модель составляет концептуальную основу для компьютерного моделирования и реализации функции парадигматических отношений для определенного класса флективных языков. Практическая реализация модели выполнена на материале русского языка.

ЯЗЫК ЕСТЕСТВЕННЫЙ ФЛЕКТИВНОГО СТРОЯ, СЛОВОИЗМЕНЕНИЕ, ПАРАДИГМАТИЧЕСКИЙ ТИП, ГРАММАТИЧЕСКИЙ КЛАСС, ПАРАДИГМАТИЧЕСКИЙ КЛАСС, ГРАММАТИЧЕСКОЕ ЗНАЧЕНИЕ, СЛОВОИЗМЕНТЕЛЬНАЯ ПАРАДИГМА, ОТНОШЕНИЕ ПАРАДИГМАТИЗАЦИИ, ОПЕРАТОР ПАРАДИГМАТИЗАЦИИ, ГРАММАТИЧЕСКАЯ ЛЕКСИКОГРАФИЧЕСКАЯ СИСТЕМА

### Введение

Неотъемлемой составляющей автоматизированной системы обработки флективного естественного языка является подсистема морфологического анализа (морфологический анализатор), в результате работы которой для каждой текстовой словоформы определяется ее начальная форма, а также необходимые значения грамматических параметров, которые идентифицируют грамматический статус словоформы в тексте. Эффективность систем обработки в целом определяется, прежде всего, результатом работы морфологического анализатора. Например аннотирование текстов лингвистических корпусов без использования морфологического анализатора вообще немыслимо.

Существуют разные схемы построения морфологических анализаторов. Среди большого количества работ, посвященных этой теме, можно выделить три основных подхода: подход, основывающийся на использовании больших словарей (так называемая четкая морфология) [1, 4, 5, 7, 8, 11], подход на основе правил (нечеткая морфология) [14, 15] и вероятностный подход [16]. Первый из них позволяет получать точные результаты морфологического разбора (но лишь при условии, если анализируемое слово имеется в словаре). Второй позволяет строить гипотезы относительно возможного грамматического статуса слов, «незнакомых» системе. Третий подход эффективен для языков с четким порядком слов в предложении. Сочетаемость первого и второго подходов разрешает получать точный и полный морфологический анализ [14, 18]. Таким образом, морфологические анализаторы целесообразно основывать на использовании больших морфологических или грамматических словарей, что обеспечивает морфологический анализ с высокой точностью (см. например, [1], [4], [8]).

В Украинском языково-информационном фонде НАН Украины (УЯИФ) алгоритмы морфо-

логического анализа ориентированы на использование грамматических словарей [1, 11]. Электронные грамматические словари разработаны для нескольких языков (украинского, русского, английского, немецкого, испанского, французского и турецкого). Словари разрабатывались на основе теории лексикографических систем [2, 3]. Они ориентированы на письменные варианты языков и предназначены, прежде всего, для использования их в системах автоматической обработки текстов. Помимо этого, предусмотрено предоставление пользователю возможности доступа к словарю как к справочно-информационной системе (поиск слов, предоставление информации относительно словоизменения определенных реестровых единиц, их грамматической параметризации и тому подобное).

При создании грамматической лексикографической системы флективного языка определяющим является наличие формализованной модели словоизменения лексики, которая, в свою очередь, означает построение определенной классификации (то есть разбиения лексического состава по определенным параметрам на классы слов, внутри которых словоизменение происходит согласно единому правилу). Заметим, что модели и действующие автоматизированные системы словоизменения для русского [4–8, 14–19] и украинского языков [12, 13] построены уже довольно давно. Они изложены, в частности, и в наших работах [1, 3, 20, 21]. Тем не менее, корректной формальной модели словоизменения нам до сих пор не попадалось. Подчеркнем, что речь идет именно о формальной модели, которая требует для своего формулирования применения стандартных математических объектов, например, теоретико-множественных. В то же время, применение теоретико-множественных объектов, так сказать, в лоб натывается на определенные формальные преграды. Ведь первым шагом при формулировании модели есть оп-

ределение некоего «множества слов», что на самом деле есть и первым шагом математического «грехопадения», поскольку, как правило, совокупности рассматриваемых слов, строго говоря, не являются множествами — они содержат повторяющиеся элементы, что является следствием грамматической омонимии. Поэтому мы сочли необходимым публикацию работы, в которой наведен математический порядок и соблюдены необходимые условия корректности.

### 1. Построение формальной модели

Построение формальной модели словоизменения флективного языка требует установления и формализации лингвистических критериев, согласно которым лексический состав языка разбивается на определенные непересекающиеся совокупности, внутри каждой из которых словоизменение происходит в соответствии с одинаковыми правилами. Совокупности слов с такими свойствами назовем словоизменительными (или парадигматическими) классами. При формализации процессов словоизменения мы будем использовать формализмом мультимножеств<sup>1</sup>.

Моделирование распределения языкового лексикона на парадигматические классы осуществляется в несколько этапов. На первом определяется понятие парадигматического типа; при этом принципиальную роль играют понятия грамматической категории, грамматического значения и грамматической формы [9].

Пусть  $L$  — некоторый фиксированный язык (для примера будем рассматривать русский);

$A$  — множество<sup>2</sup> слов языка  $L$ , которое является порождающим множеством для мультимножества (ММ)  $W$ , представляющего лексикон рассматриваемого языка;

$\Omega$  — множество грамматических значений (значения грамматических категорий таких как род, число, падеж, залог, наклонение, время, лицо);

$\Omega(T_i)$  — множество грамматических значений, которые отвечают типу  $T_i$ .

Рассмотрим множество грамматических значений детальнее.

Элементами множества грамматических значений (значений грамматических категорий) русского языка  $\Omega = \{\Omega^1, \Omega^2, \Omega^3, \Omega^4, \Omega^5, \Omega^6, \Omega^7, \Omega^8\}$  являются следующие подмножества:

<sup>1</sup> Мультимножество или множество с повторяющимися элементами как новое математическое понятие, насколько нам известно, впервые встречалось еще в работах Д. Кнута [22, С. 498]. В последние годы появилась серия работ А.Б. Петровского [23–29], посвященных развитию теории мультимножеств и проблемам применения данной теории для принятия решений в условиях нечеткой исходной информации, в кластерном анализе многопризнаковых объектов, объектов с противоречивыми свойствами, в сетях Петри и тому подобное.

<sup>2</sup> Множество (в данном случае совокупность слов языка), состоящее из элементов (слов), среди которых нет повторяющихся. Это множество порождает мультимножество слов языка (или множество с повторяемыми элементами).

$\Omega^1 = \{\omega_1^1, \omega_2^1, \omega_3^1\} \equiv \{м.р., ж.р., с.р.\}$  — множество значений грамматической категории *род*;

$\Omega^2 = \{\omega_1^2, \omega_2^2\} \equiv \{\text{единственное число, множественное число}\}$  — множество значений грамматической категории *число*;

$\Omega^3 = \{\omega_1^3, \omega_2^3, \omega_3^3, \omega_4^3, \omega_5^3, \omega_6^3\} \equiv \{\text{именительный, родительный, дательный, винительный, творительный, предложный}\}$  — множество значений грамматической категории *падеж*;

$\Omega^4 = \{\omega_1^4, \omega_2^4\} \equiv \{\text{действительный, страдательный}\}$  — множество значений грамматической категории *залог*;

$\Omega^5 = \{\omega_1^5, \omega_2^5, \omega_3^5\} \equiv \{\text{изъявительное, сослагательное, повелительное}\}$  — множество значений грамматической категории *наклонение*;

$\Omega^6 = \{\omega_1^6, \omega_2^6, \omega_3^6\} \equiv \{\text{настоящее, прошлое, будущее}\}$  — множество значений грамматической категории *время*;

$\Omega^7 = \{\omega_1^7, \omega_2^7, \omega_3^7\} \equiv \{1 \text{ лицо, } 2 \text{ лицо, } 3 \text{ лицо}\}$  — множество значений грамматической категории *лицо*.

$\Omega^8 = \{\omega_1^8, \omega_2^8, \omega_3^8\} \equiv \{\text{совершенный вид, несовершенный вид, совершенный и несовершенный вид}\}$  — множество значений грамматической категории *вид*. Или в обозначениях А. Зализняка [10]:  $\Omega^8 = \{\omega_1^8, \omega_2^8, \omega_3^8\} = \{св, нсв, св-нсв\}$ .

Множество грамматических значений может быть представлено в виде графа (рис. 1).

Определенные комбинации определенных грамматических значений (элементов множеств  $\Omega^i$ ) характеризуют словоизменительную форму слова. Например словоизменительные формы лексемы «дирижёр» (которая является существительным) характеризуются комплексом грамматических значений, состоящим из пар элементов множеств значений грамматических категорий *число* и *падеж*:  $\{\langle \omega_i^2, \omega_j^3 \rangle\}$ ,  $i = 1, 2$ ;  $j = 1, 2, \dots, 6$ . Каждая пара значений  $\langle \text{число, падеж} \rangle$  отвечает определенной форме лексемы из ее словоизменительной парадигмы:  $\langle \omega_1^2, \omega_1^3 \rangle$  — *дирижёр*,  $\langle \omega_1^2, \omega_2^3 \rangle$  — *дирижёра*, ...,  $\langle \omega_1^2, \omega_5^3 \rangle$  — *дирижёром*, ...,  $\langle \omega_2^2, \omega_6^3 \rangle$  — *дирижёрах*. Грамматические формы глагола характеризуются другим комплексом грамматических значений, который состоит из кортежей (из  $n$ -ок, где  $n$  может принимать значение от 1 до 6) элементов множеств грамматических значений категорий *залог, наклонение, время, лицо, число, род*. Например словоизменительная парадигма глагола «*произойти*» состоит из 15 словоформ<sup>3</sup>, которые определяются таким комплексом грамматических значений:

$$\{\omega_1^V, \omega_2^V, \omega_3^V, \dots, \omega_{15}^V\} = \{\langle \omega_i^4, \omega_j^5, \omega_k^6, \omega_l^2, \omega_m^7, \omega_n^1 \rangle_k\},$$

$$i = 1, 2; j = 1, 2, 3; k = 1, 2, \dots, 15.$$

<sup>3</sup> Потенциально глагольная парадигма в общем случае может состоять из большего количества форм — в нашей модели максимально возможное количество разных форм — 45 (см. [1, 21]).

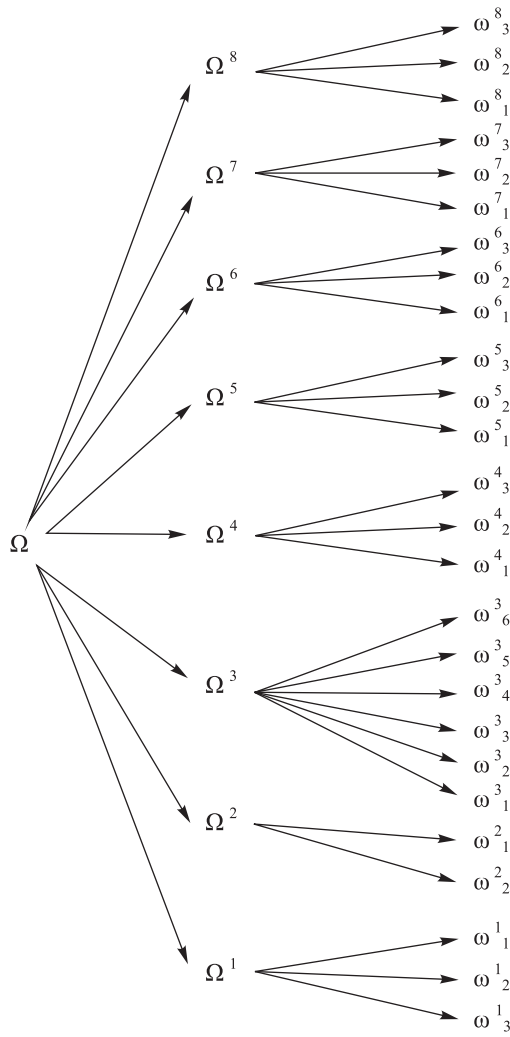


Рис. 1. Множество грамматических значений

Причем значение категории рода не определено для  $k = 2, 3, 4, 5, 6, 7, 12, 13, 14$ , значение категории числа не определено для  $k = 14$ , значение категорий лица и наклонения не определены для  $k = 14, 15$ , а для инфинитивной формы не определенными являются все грамматические значения.

Каждой из  $n$ -ок грамматических значений отвечает определенная парадигматическая форма рассматриваемой лексемы:  $\omega_1^V$  — *произойти*,  $\omega_2^V = \langle \omega_1^4, \omega_1^5, \omega_3^6, \omega_1^2, \omega_1^7 \rangle$  — *произойду*, ...,  $\omega_8^V = \langle \omega_1^4, \omega_1^5, \omega_2^6, \omega_1^2, \omega_{1,2,3}^7, \omega_2^1 \rangle$  — *произошла*,  $\omega_{11}^V = \langle \omega_1^4, \omega_1^5, \omega_2^6, \omega_1^2, \omega_{1,2,3}^7, \omega_{1,2,3}^1 \rangle$  — *произошли*, ...,  $\omega_{15}^V = \langle \omega_1^4, \omega_2^6, \omega_1^2 \rangle$  — *происшедший*.<sup>4</sup>

Как видим, словоизменительная парадигма слов из разных классов определяется своим индивидуальным набором словоизменительных категорий.

<sup>4</sup> Нижние индексы в записи  $\omega_{1,2,3}^7$  означают, что у соответствующей парадигматической словоформы грамматическая категория лицо может принимать какое-либо из трёх возможных значений — первое, второе или третье лицо (я, ты, она — пришла; я, ты, он — пришёл; я, ты, оно — пришло; мы, вы, они — пришли). В записи  $\omega_{1,2,3}^1$  нижние индексы означают возможность какого-либо (одного или комбинации нескольких) из значений грамматической категории род.

В соответствии со словоизменительными категориями, определяющими словоизменительную парадигму конкретных слов, вводятся парадигматические типы (ПТ). Пусть  $T_i$  ( $i=1, 2, \dots, N$ ) — парадигматические типы,  $N$  — количество парадигматических типов;  $W(T_i)$  — слова языка  $L$ , имеющие парадигматический тип  $T_i$ . Так, что  $W = \bigcup_{i=1}^N W(T_i)$ . Заметим, что  $W(T_i)$ , как, собственно, и весь лексический состав языка, является мультимножеством (ММ), или множеством с повторяющимися элементами. То есть, на самом деле,  $W(T_i) = \{(n_{T_i}(w_j) \bullet w_j) \mid w_j \in A, n_{T_i}(w_j) \in \mathbb{Z}^+\}$ , где  $n_{T_i}(w_j)$  называется функцией числа экземпляров элемента  $w_j$  в ММ  $W(T_i)$  и определяет, сколько раз элемент  $w_j \in A$  входит в ММ  $W(T_i)$ ; символ  $\bullet$  обозначает кратность вхождения элемента в соответствующую компоненту ММ. Элементы  $w_j$  мультимножества  $W(T_i)$ , для которых  $n_{T_i}(w_j) > 1$ , это слова, тождественные графемно, но имеющие определенные различия в своих грамматических и семантических свойствах.

Подробное описание парадигматических типов русского языка представлено в наших публикациях [1, с. 128-156, с. 218-223], [21]. Для изменяемых слов русского языка определяется четыре парадигматических типа: субстантивный, адъективный, глагольный и тип количественных числительных, а для неизменяемых слов — так называемый «нулевой парадигматический тип».

Каждый парадигматический тип характеризуется определенным набором грамматических значений.

Субстантивный парадигматический тип характеризуется комплексом значений грамматических категорий *число* ( $\Omega^2$ ) и *падеж* ( $\Omega^3$ ):

$$\Omega(T^S) \subseteq \Omega^2 \times \Omega^3 = \{\langle \omega_i^2, \omega_j^3 \rangle\}, \quad (1)$$

$$\omega_i^2 \in \Omega^2, i = 1, 2; \omega_j^3 \in \Omega^3, j = 1, 2, \dots, 6$$

Категория рода не включается в число дифференцирующих словоизменительных категорий.

Субстантивный парадигматический тип может рассматриваться как характеризующийся мультимножеством, полученным (порожденным) из множеств  $\Omega^2$  и  $\Omega^3$  значений грамматических категорий число и падеж:

$$\Omega(T^S) \equiv \Omega^S$$

$$= \{2 \bullet \omega^2, 6 \bullet \omega^3 \mid \omega^2 = \{\omega_1^2, \omega_2^2\}, \omega^3 = \{\omega_1^3, \omega_2^3, \dots, \omega_6^3\}\} =$$

$$= \{6 \bullet \omega_1^2, 6 \bullet \omega_2^2, 2 \bullet \omega_1^3, 2 \bullet \omega_2^3, 2 \bullet \omega_3^3, 2 \bullet \omega_4^3, 2 \bullet \omega_5^3, 2 \bullet \omega_6^3\}.$$

Графическое изображение ММ грамматических значений субстантивного парадигматического типа продемонстрировано на рис. 2.

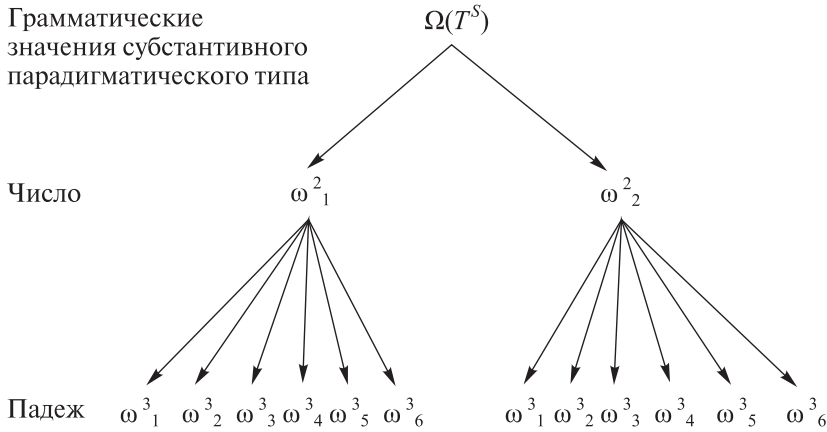


Рис. 2. Отношение между грамматическими значениями субстантивного парадигматического типа

По субстантивному парадигматическому типу происходит словоизменение у существительных и местоимений-существительных.

Адъективный парадигматический тип характеризуется комплексом грамматических категорий *род* ( $\Omega^1$ ), *число* ( $\Omega^2$ ) и *падеж* ( $\Omega^3$ ):

$$\Omega(T^A) \subseteq \Omega^1 \times \Omega^2 \times \Omega^3 = \{ \{ \omega_i^1, \omega_j^2, \omega_k^3 \} \},$$

где  $\omega_i^1 \in \Omega^1, i = 1, 2, 3; \omega_j^2 \in \Omega^2, j = 1, 2; \omega_k^3 \in \Omega^3, k = 1, 2, \dots, 6.$  (2)

Графическое представление отношений между грамматическими значениями адъективного парадигматического типа изображено на рис. 3.

По адъективному парадигматическому типу происходит словоизменение у прилагательных, причастий, местоимений-прилагательных и порядковых числительных.

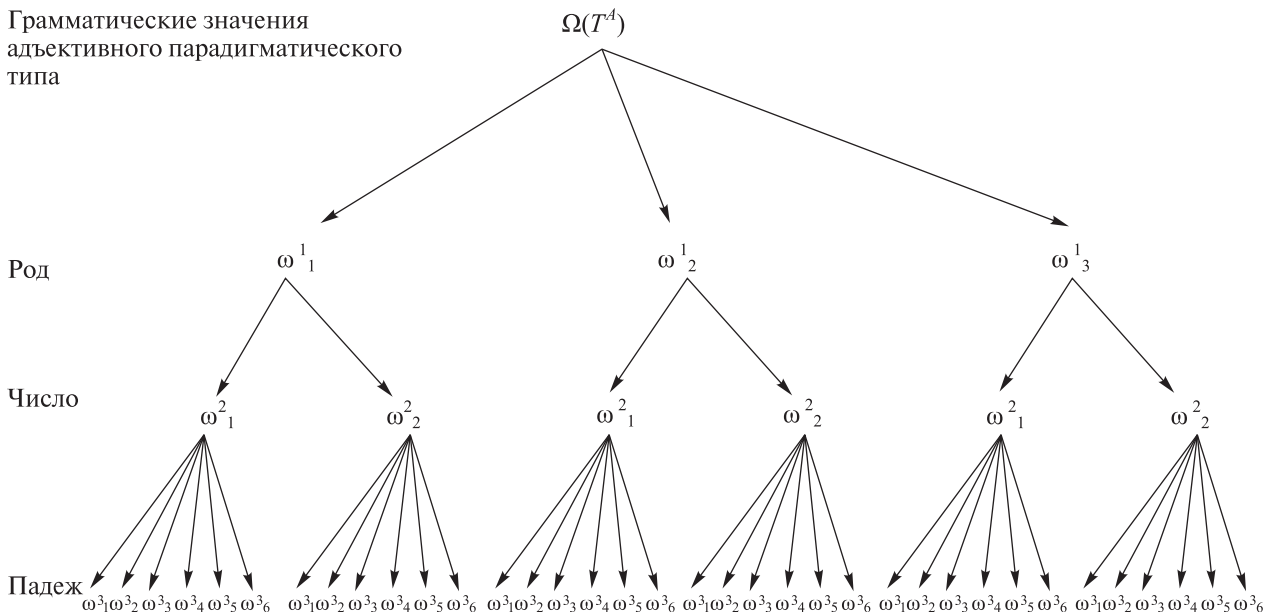


Рис. 3. Отношение между грамматическими значениями адъективного парадигматического типа

Грамматические значения глагольного парадигматического типа определяются значениями грамматических категорий *залог* ( $\Omega^4$ ), *наклонение* ( $\Omega^5$ ), *время* ( $\Omega^6$ ), *лицо* ( $\Omega^7$ ), *число* ( $\Omega^2$ ), *род* ( $\Omega^1$ ):

$$\Omega(T^V) \subseteq \Omega^4 \times \Omega^5 \times \Omega^6 \times \Omega^7 \times \Omega^2 \times \Omega^1. \quad (3)$$

Графическое представление отношений между грамматическими значениями глагольного парадигматического типа дано на рис. 4–6.

Парадигматический тип количественных числительных характеризуется грамматическими значениями категории *падеж* ( $\Omega^3$ ):

$$\Omega(T^C) \subseteq \Omega^3. \quad (4)$$

Графическое представление отношений между грамматическими значениями парадигматического типа количественных числительных изображено на рис. 7.

Обозначим  $P_j, (j = 1, 2, \dots, p)$  – лексико-грамматические классы, где  $p$  – количество лексико-грамматических классов для данного языка;  $W(P_j)$  – слова языка  $L$ , принадлежащие лексико-грамматическому классу  $P_j$ .

По признаку принадлежности к определенной части речи и по дополнительным признакам, являющимся классифицирующими (не словоизменяемыми) в пределах определенной части речи, ММ слов  $W$  распределяем на подмультимножества, которые будем называть *грамматическими классами* (обозначим их  $P_j$ ), следующим образом.

Грамматические значения глагольного парадигматического типа

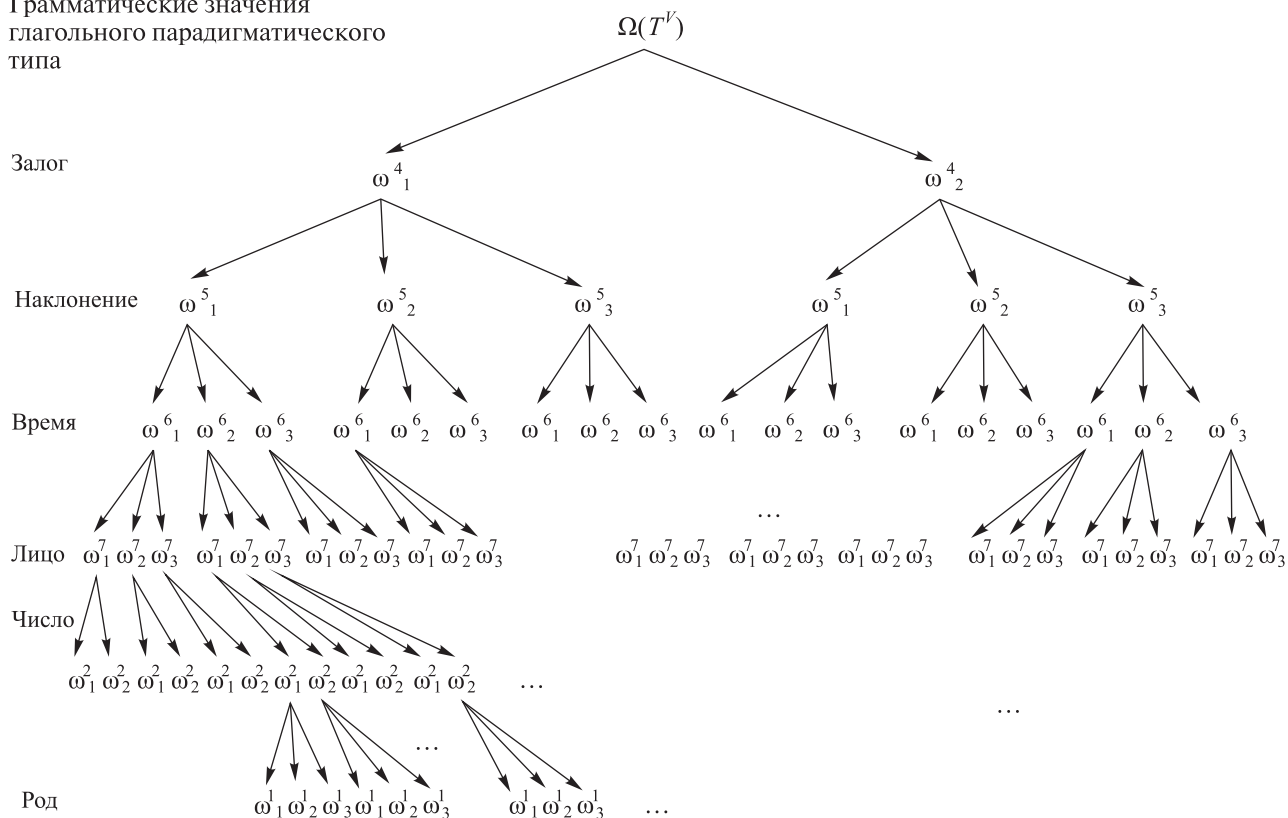


Рис. 4. Отношение между грамматическими значениями глагольного парадигматического типа

Грамматические значения глагольного парадигматического типа

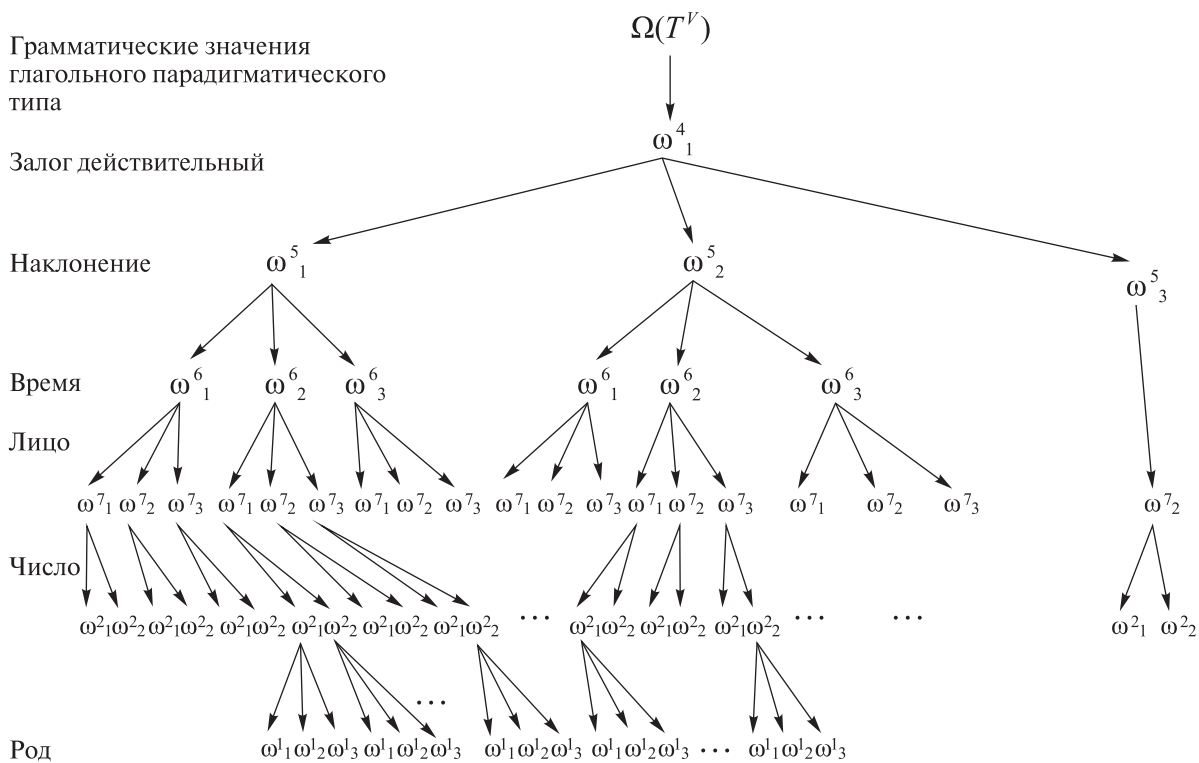


Рис. 5. Отношение между грамматическими значениями глагольного парадигматического типа (действительный залог)

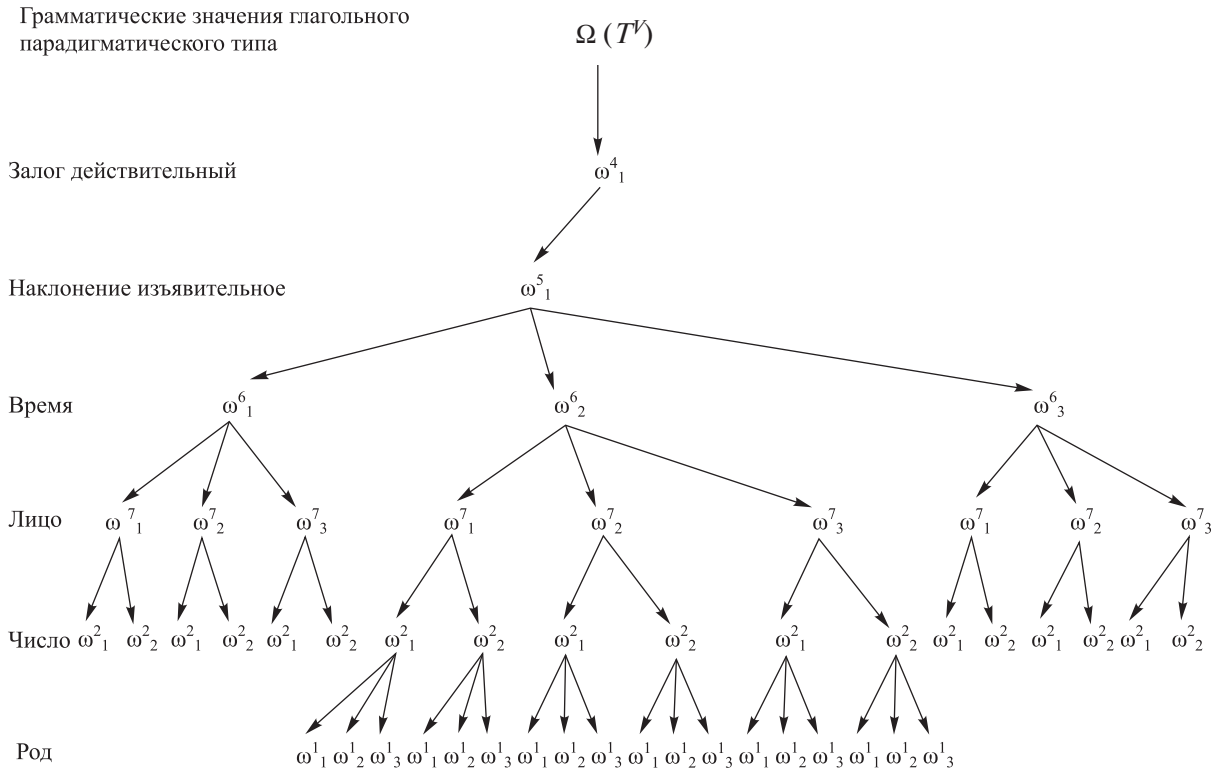


Рис. 6. Отношение между грамматическими значениями глагольного парадигматического типа (действительный залог, изъявительное наклонение)

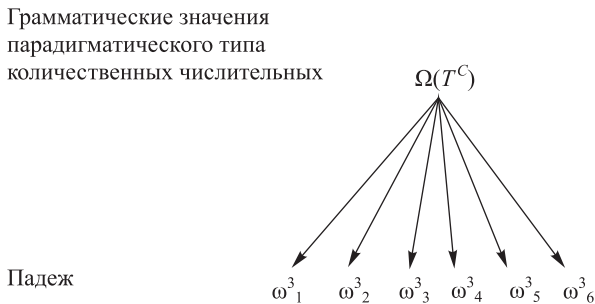


Рис. 7. Отношение между грамматическими значениями парадигматического типа количественных числительных

Существительные по значению грамматической категории «род» (которая в пределах этой части речи является классифицирующим признаком) [10] распределяются на 3 грамматических класса: существительные мужского рода, существительные женского и существительные среднего рода; множественные существительные (существительные pluralia tantum) образуют отдельный грамматический класс. Таким образом, существительные составляют 4 грамматических класса, которые в последующем изложении будем обозначать соответственно  $P_1, P_2, P_3$  и  $P_4$ .

Глаголы по значению грамматической категории «вид» (которая рассматривается нами как

классифицирующая) [10] распределяются на такие три грамматических класса: глаголы совершенного вида (св), глаголы несовершенного вида (нсв) и двувидовые глаголы (св-нсв).

Местоимения разделяются на два грамматических класса: местоимения-прилагательные и местоимения-существительные.

Грамматический класс адъективов составляют прилагательные и порядковые числительные (то есть этот грамматический класс образован путем объединения).

Все другие слова из ММ  $W$  слов языка отнесены к своим грамматическим классам по признаку принадлежности к конкретной части речи (то есть в этом случае понятие «грамматический класс» совпадает с понятием «часть речи»).

Таким образом, в русском языке нами выделены грамматические классы, которые мы представляем в табл. 1.

$$W = \bigcup_{j=1}^{19} W(P_j).$$

Определенный парадигматический тип может быть присущ нескольким грамматическим классам, внутри каждого из которых выделяются парадигматические классы.

Схема отношений между парадигматическими типами, грамматическими классами и парадигматическими классами представлена на рис. 8.

Таблица 1  
Грамматические классы русского языка

№	Грамматический класс	Символ ГК
1	Существительные мужского рода	$P_1$
2	Существительные женского рода	$P_2$
3	Существительные среднего рода	$P_3$
4	Множественные существительные	$P_4$
5	Адъективы (прилагательные+порядковые числительные)	$P_5$
6	Глаголы совершенного вида	$P_6$
7	Глаголы несовершенного вида	$P_7$
8	Двувидовые глаголы	$P_8$
9	Причастия	$P_9$
10	Местоимения (местоимения-существительные)	$P_{10}$
11	Местоимения-прилагательные	$P_{11}$
12	Числительные количественные	$P_{12}$
13	Наречия	$P_{13}$
14	Междометия	$P_{14}$
15	Союзы	$P_{15}$
16	Частицы	$P_{16}$
17	Предлоги	$P_{17}$
18	Предикативы	$P_{18}$
19	Сокращения	$P_{19}$

Как видно из схемы, согласно принципу распределения:

$$W(T_i) = \bigcup_{j=1}^{p_i} W(P_j), \quad (5)$$

где  $P_j \subseteq T_i$ ,  $p_i$  – количество грамматических классов, в которых словоизменение имеет парадигматический тип  $T_i$ .

$$W(P_j) = \bigcup_{k=1}^{n_j} W(\Pi_k), \quad (6)$$

где  $\Pi_k \subseteq P_j \subseteq T_i$ ;  $n_j$  – количество парадигматических классов грамматического класса  $P_j$ , в которых словоизменение имеет парадигматический тип  $T_i$ .

В соответствии с принципами, изложенными в наших предыдущих работах [1, 11, 20, 21], внутри грамматических классов выделяются парадигматические классы, строятся отношения парадигматизации и оператор парадигматизации, а каждая лексема  $x$  приобретает представление в виде комбинации ее квазиосновы и квазифлексии:

$$x = c(x) * f(x), \quad (7)$$

где  $c(x)$  – часть лексемы  $x$ , которая в процессе словоизменения остается неизменной (квазиоснова);  $f(x)$  – ее изменяемая составляющая (квазифлексия);  $*$  – конкатенация, а словоизменительная парадигма представляется в виде:

$$\pi(x) = c(x) * \{f_j(x)\}, \quad (8)$$

где  $f_j(x)$ ;  $j=0,1,2,\dots,n(T_i)$  – квазифлексии в соответствующих грамматических формах. Например лексема  $x = \text{'лѐд'}$  представляется в виде  $x = \text{'л'} * \text{'ѐд'}$ , где  $\text{'л'}$  – квазиоснова, а  $\text{'ѐд'}$  квазифлексия начальной формы лексемы. Ее словоизменительная парадигма:

$$\pi(x) = c(x) * \{f_j(x)\} = \text{'л'} * \{\text{'ѐд'}; \text{'ьда'}, \text{'ьду'}; \text{'ьду'}; \text{'ѐд'}; \text{'ьдом'}; \text{'ьду'}, \text{'ьде'}; \text{'ьды'}; \text{'ьдов'}; \text{'ьдам'}; \text{'ьды'}; \text{'ьдами'}; \text{'ьдах'}\}^5$$

Отношения парадигматизации  $\pi_i$  строятся на декартовых произведениях  $P_i \times P_i$ ,  $i = 1, 2, \dots, 19$ , (то есть построение оператора парадигматизации выполняется на мультимножестве лексем каждого грамматического класса отдельно). Как было указано выше, в нашей системе имеется 19 грамматических классов, а, следовательно, соответственно столько же отношений парадигматизации  $\pi_i$  вида:

$$\forall x^1, x^2 \in P_i \quad x^1 \pi_i x^2 : x^1 = c(x^1) * f^k, \quad (9)$$

$$x^2 = c(x^2) * f^k, f^k \in [F]^k,$$

где  $[F]^k$  состоит из квазифлексий слов, имеющих во всех своих грамматических значениях  $\omega_1, \omega_2, \dots, \omega_{n(T_i)}$  (парадигматический тип  $T_i$ ) одинаковые изменяемые составляющие. Обозначим  $C = \{n(c_i) \bullet c_i\}$  – ММ квазиоснов,  $i = \overline{1, n_{co}}$  – индекс на обозначение уникального номера квазиосновы,  $n_{co}$  – количество квазиоснов в словаре;  $F = \{n(f_j) \bullet f_j\}$  – ММ квазифлексий,  $j = \overline{1, n_f}$  – индекс на обозначение уникального номера квазифлексии,  $n_f$  – количество квазифлексий в словаре.

Отношение парадигматизации является отношением эквивалентности, поскольку оно, очевидно, рефлексивно, симметрично и транзитивно. Фактор-множество  $P_i / \pi_i$  составляет множество парадигматических классов грамматического класса  $P_i$  (парадигматического типа  $T_i$ ):  $P_i = \bigcup_{j=1}^n \Pi_j$ . Код-

ному парадигматическому классу относятся слова, имеющие одинаковые наборы квазифлексий для всех грамматических форм, а отличаются один от другого лишь неизменяемой составляющей  $c(x)$ .

В каждом из грамматических классов (парадигматических типов  $T_i$ ) построим разбиение на парадигматические классы и рассмотрим все возможные отношения между элементами указанных выше множеств  $\Omega, T, P, C, F, \Pi$ .

Отношение между грамматическими значениями определенных парадигматических типов описаны выше (1) – (4). Далее приведем другие отношения.

<sup>5</sup> Квазифлексии в разных грамматических формах записаны через точку с запятой; если для грамматической формы существует несколько вариантов словоформ, то соответствующие квазифлексии поданы через запятую.

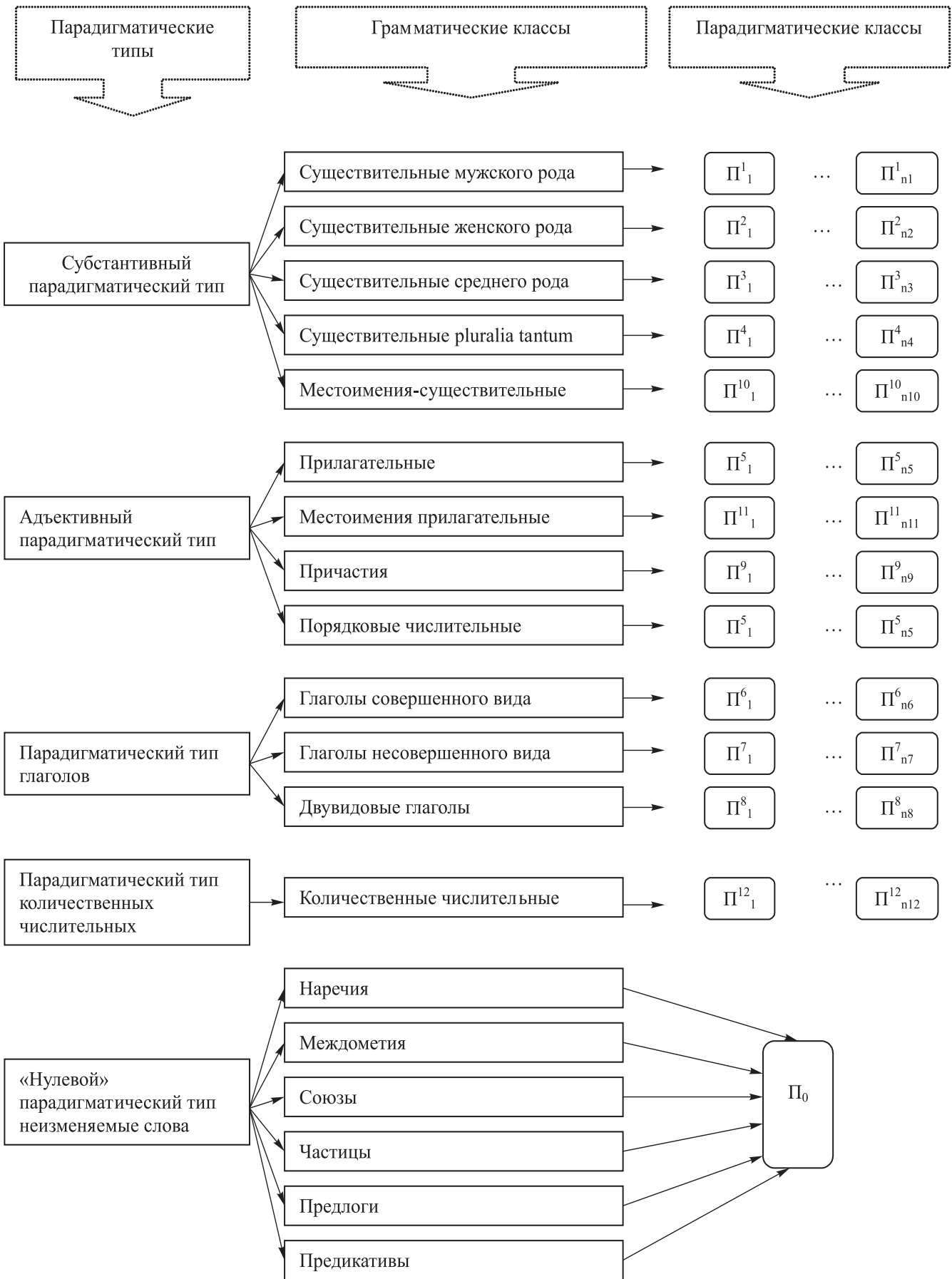


Рис. 8. Схема отношений между парадигматическими типами, грамматическими классами и парадигматическими классами

Отношение «парадигматический тип»:

$$T \subseteq T^S \cup T^A \cup T^C \cup T^0 \quad (10)$$

Отношение «грамматическое значение»:

$$\Omega(T) \subseteq \Omega(T^S) \cup \Omega(T^A) \cup \Omega(T^C). \quad (11)$$

Отношение «парадигматический тип – грамматический класс»:

$$R_{TP} \subseteq T \times P : \{ \langle T_i, P_j \rangle \}, \quad (12)$$

где  $i = 1, 2, 3, 4; j = 1, 2, \dots, 19$ ;

$$T_i = \{ T_0 = T^0, T_1 = T^S, T_2 = T^A, T_3 = T^V, T_4 = T^C \}.$$

Соответствие между парадигматическими типами и грамматическими классами хорошо видно на рис. 8.

$$R_{TP} \subseteq T \times P : \{ \langle T_1, P_1 \rangle, \langle T_1, P_2 \rangle, \langle T_1, P_3 \rangle, \langle T_1, P_4 \rangle, \langle T_1, P_{10} \rangle, \langle T_2, P_5 \rangle, \langle T_2, P_9 \rangle, \langle T_2, P_{11} \rangle, \langle T_3, P_6 \rangle, \langle T_3, P_7 \rangle, \langle T_3, P_8 \rangle, \langle T_4, P_{12} \rangle, \langle T_0, P_{13} \rangle, \langle T_0, P_{14} \rangle, \langle T_0, P_{15} \rangle, \langle T_0, P_{16} \rangle, \langle T_0, P_{19} \rangle \} \quad (13)$$

Отношение «грамматический класс – грамматическое значение»:

$$R_{P\Omega(T)} \subseteq P \times \Omega(T) : \{ \langle P_j, \Omega(T_i) \rangle \}, \quad (14)$$

где  $j=1, 2, \dots, 19; T_0 = T^0; T_1 = T^S; T_2 = T^A; T_3 = T^V; T_4 = T^C$ .

$$R_{P\Omega(T)} \subseteq P \times \Omega(T) : \{ \langle P_1, \Omega(T_1) \rangle, \langle P_2, \Omega(T_1) \rangle, \langle P_3, \Omega(T_1) \rangle, \langle P_4, \Omega(T_1) \rangle, \langle P_{10}, \Omega(T_1) \rangle, \langle P_5, \Omega(T_2) \rangle, \langle P_9, \Omega(T_2) \rangle, \langle P_{11}, \Omega(T_2) \rangle, \langle P_6, \Omega(T_3) \rangle, \langle P_7, \Omega(T_3) \rangle, \langle P_8, \Omega(T_3) \rangle, \langle P_{12}, \Omega(T_4) \rangle, \langle P_{13}, \Omega(T_0) \rangle, \langle P_{14}, \Omega(T_0) \rangle, \langle P_{15}, \Omega(T_0) \rangle, \langle P_{16}, \Omega(T_0) \rangle, \langle P_{17}, \Omega(T_0) \rangle, \langle P_{18}, \Omega(T_0) \rangle, \langle P_{19}, \Omega(T_0) \rangle \} \quad (15)$$

Отношение «лексема – грамматический класс»:

$$R_{WP} \subseteq W \times P. \quad (16)$$

Например 'стол'  $\in P_1$ , 'вишня'  $\in P_2$ , 'слово'  $\in P_3$ , 'ножницы'  $\in P_4$ , 'горячий'  $\in P_5$ , 'выучить'  $\in P_6$ , 'учить'  $\in P_7$ , 'автоматизировать'  $\in P_8$ , ...

Отношение «лексема – грамматическое значение»:

$$R_{W\Omega(T)} \subseteq W \times \Omega(T). \quad (17)$$

Например для лексемы 'слово', принадлежащего к классу существительных среднего рода ('слово'  $\in P_3$ ) и характеризуется набором грамматических значений  $\Omega(T_1)$ , отношение  $R_{W\Omega(T)}$  имеет вид:

$$\forall w \in P_3 : R_{W\Omega(T)} \subseteq [w] \times \Omega(T_1) = [w] \times \Omega_2 \times \Omega_3 :$$

$$\{ \langle 'слово', \langle \text{ед.}, \text{им.} \rangle \rangle, \langle 'слово', \langle \text{ед.ч.}, \text{род.} \rangle \rangle, \langle 'слову', \langle \text{ед.ч.}, \text{дат.} \rangle \rangle, \langle 'слово', \langle \text{ед.ч.}, \text{вин.} \rangle \rangle, \langle 'словом', \langle \text{ед.ч.}, \text{твор.} \rangle \rangle, \langle 'слове', \langle \text{ед.ч.}, \text{предл.} \rangle \rangle, \langle 'слово', \langle \text{мн.ч.}, \text{им.} \rangle \rangle, \langle 'слов', \langle \text{мн.ч.}, \text{род.} \rangle \rangle, \langle 'словам', \langle \text{мн.ч.}, \text{дат.} \rangle \rangle, \langle 'слова', \langle \text{мн.ч.}, \text{вин.} \rangle \rangle, \langle 'словами', \langle \text{мн.ч.}, \text{твор.} \rangle \rangle, \langle 'словах', \langle \text{мн.ч.}, \text{предл.} \rangle \rangle \}.$$

$\langle \text{мн.ч.}, \text{твор.} \rangle \rangle, \langle 'словах', \langle \text{мн.ч.}, \text{предл.} \rangle \rangle \}.$

Проекция отношения  $R_{W\Omega(T)}$  по первому компоненту  $\pi_W(R_{W\Omega(T)})$  есть мультимножество (в словоизменительной парадигме имеются словоформы, которые графемно совпадают, но имеют различающиеся грамматические характеристики. Одинаковая форма слова в именительном и винительном падежах единственного числа – 'слово', а форма 'слова' в парадигме фигурирует трижды – в родительном падеже единственного числа, в именительном и винительном падежах множественного числа:

$$\pi_W(R_{W\Omega(T)}) = \{ 2 \cdot \langle 'слово' \rangle, \langle 'слова' \rangle, \langle 'слову' \rangle, \langle 'словом' \rangle, \langle 'слове' \rangle, 3 \cdot \langle 'слова' \rangle, \langle 'слов' \rangle, \langle 'словам' \rangle, \langle 'словами' \rangle, \langle 'словах' \rangle \}.$$

Отношение «лексема – грамматический класс – грамматическое значение»:

$$R_{WP\Omega(T)} \subseteq W \times P \times \Omega(T). \quad (18)$$

Учитывая формулу (7), согласно которой лексема представляется в виде комбинации неизменяемой и изменяемой составляющих (квазиосновы и квазифлексии), рассмотрим отношение «квазиоснова – квазифлексия – грамматический класс – грамматическое значение»:

$$R_{CFP\Omega(T)} \subseteq C \times F \times P \times \Omega(T). \quad (19)$$

Распишем его детальнее:

$$\left. \begin{array}{l} \langle c^{j_{P_i}}, f_0^{j_{P_i}}, P_i, \omega_0^S \rangle \\ \langle c^{j_{P_i}}, f_1^{j_{P_i}}, P_i, \omega_1^S \rangle \\ \dots \\ \langle c^{j_{P_i}}, f_{n(T^S)}^{j_{P_i}}, P_i, \omega_{n(T^S)}^S \rangle \end{array} \right\} i = 1, 2, 3, 4, 10; j_{P_i} = \overline{1, n_S},$$

$$n_S = \sum_{i=1}^4 n(P_i) + n(P_{10}), \quad (20)$$

где  $n(P_i)$  – количество слов грамматического класса  $P_i$ ;  $n_S$  – количество существительных в реестре;  $n(T^S)$  – количество словоизменительных форм в парадигме субстантивного типа. Общепринятым для русского языка:  $n(T^S) = 12; \omega_m^S = \langle \omega_i^2, \omega_j^3 \rangle \in \Omega(T^S), \omega_i^2 \in \Omega^2, i = 1, 2; \omega_j^3 \in \Omega^3, j = 1, 2, \dots, 6;$

$$\left. \begin{array}{l} \langle c^{j_{P_i}}, f_0^{j_{P_i}}, P_i, \omega_0^A \rangle \\ \langle c^{j_{P_i}}, f_1^{j_{P_i}}, P_i, \omega_1^A \rangle \\ \dots \\ \langle c^{j_{P_i}}, f_{n(T^A)}^{j_{P_i}}, P_i, \omega_{n(T^A)}^A \rangle \end{array} \right\}, \quad (21)$$

$i = 5, 9, 11.$

$$j_{P_i} = \overline{1, n_A}, n_A = n(P_5) + n(P_9) + n(P_{11}),$$

где  $n_A$  – количество адъективов (прилагательные + порядковые числительные + причастия + местоиме-

6 Имеются и иные мнения относительно количества словоизменительных форм в парадигме существительного, а именно: 14. См., например, [27; 28].

ния-прилагательные) в реестре;  $n(T^A)$  – количество словоизменительных форм в парадигме адъективного типа:  $n(T^A) = 28$ ;  $\omega_m^A = \langle \omega_l^1, \omega_i^2, \omega_j^3 \rangle \in \Omega(T^A)$ ;  $\omega_l^1 \in \Omega^1, l=1,2,3; \omega_i^2 \in \Omega^2, i=1,2; \omega_j^3 \in \Omega^3, j=1,2,\dots,6$ ;  $m = \overline{1,28}$ .

$$\left. \begin{array}{l} \langle c^{j_{P_i}}, f_0^{j_{P_i}}, P_i, \omega_0^V \rangle \\ \langle c^{j_{P_i}}, f_1^{j_{P_i}}, P_i, \omega_1^V \rangle \\ \dots \\ \langle c^{j_{P_i}}, f_{n(T^V)}^{j_{P_i}}, P_i, \omega_{n(T^V)}^V \rangle \end{array} \right\} \quad (22)$$

$$i = \overline{6,7,8},$$

$$j_{P_i} = \overline{1, n_V}, n_V = \sum_{i=6}^8 n(P_i),$$

где  $n_V$  – количество глаголов в реестре;  $n(T^V)$  – количество словоизменительных форм в парадигме глагольного типа:  $n(T^V) = 49$ ;  $\omega_m^V = \langle \omega_i^4, \omega_j^5, \omega_j^7, \omega_j^2, \omega_j^1 \rangle \in \Omega(T^V)$ ,  $m = \overline{1,49}$ ;  $i = 1, 2$ ;  $j = 1, 2, 3$ ;  $\omega_i^4 \in \Omega^4, \omega_j^5 \in \Omega^5, \omega_j^6 \in \Omega^6, \omega_j^7 \in \Omega^7, \omega_i^2 \in \Omega^2, \omega_j^1 \in \Omega^1$ .

$$\left. \begin{array}{l} \langle c^{j_{P_i}}, f_0^{j_{P_i}}, P_i, \omega_0^C \rangle \\ \langle c^{j_{P_i}}, f_1^{j_{P_i}}, P_i, \omega_1^C \rangle \\ \dots \\ \langle c^{j_{P_i}}, f_{n(T^C)}^{j_{P_i}}, P_i, \omega_{n(T^C)}^C \rangle \end{array} \right\} \quad (23)$$

$$i = 12,$$

$$j_{P_i} = \overline{1, n_C}, n_C = n(P_{12}),$$

где  $n_C$  – количество числительных в реестре;  $n(T^C)$  – количество словоизменительных форм в парадигме количественного числительного:  $n(T^C) = 6$ ;  $\omega_m^C = \langle \omega_j^3 \rangle \in \Omega(T^C)$ ,  $m = \overline{1, n(T^C)}$ .

Построим селекцию  $R_{CFP\Omega(T)}$  по условию  $P = P_i$  ( $i = \overline{1,2,\dots,12}$ )<sup>7</sup>  $\sigma_{P=P_i}(R_{CFP\Omega(T)})$ .

$$\bigcup_{i=1}^4 \sigma_{P=P_i}(R_{CFP\Omega(T)}) \cup \sigma_{P=P_{10}}(R_{CFP\Omega(T)}) = R_{CFP\Omega^S} \quad (24)$$

задает отношение «квазиоснова – квазифлексия – грамматический класс – грамматическое значение» для лексем парадигматического типа  $T^S$ .

Проекцией отношения  $R_{CFP\Omega^S}$  по второму столбцу (по атрибуту F) является мультимножество квазифлексий  $F(T^S)$  для лексем парадигматического типа  $T^S$ :

$$\pi_F(R_{CFP\Omega^S}) = \{\{f_0^{j_{P_i}}\}, \{f_1^{j_{P_i}}\}, \dots, \{f_{n(T^S)}^{j_{P_i}}\}\}, j_{P_i} = \overline{1, n_S}. \quad (25)$$

$$F_0^{T^S} = \{f_0^{j_{P_i}}\}, F_1^{T^S} = \{f_1^{j_{P_i}}\}, \dots,$$

$$F_{n(T^S)}^{T^S} = \{f_{n(T^S)}^{j_{P_i}}\}. \quad (26)$$

Построим на  $F_i^{T^S}, i = \overline{0, n(T^S)}$  отношение

$$R_{F(T^S)} \subseteq F_0^{T^S} \times F_1^{T^S} \times \dots \times F_{n(T^S)}^{T^S}. \quad (27)$$

Отношение  $R_{F(T^S)}$  по построению является отношением эквивалентности  $a$ , следовательно, разбивает множество  $F^{T^S} \subset F$  на классы. Кортежи  $\{f_0^j, f_1^j, \dots, f_{n(T^S)}^j\}$  задают классы  $\Pi_j(P, T^S)$ . Нижний индекс при  $f$  отвечает номеру грамматического значения в парадигме, верхний индекс – номер парадигматического класса. Например кортеж квазифлексий  $\{\langle 'b', \omega_1^S \rangle, \langle 'я', \omega_2^S \rangle, \langle 'ю', \omega_3^S \rangle, \langle 'я', \omega_4^S \rangle, \langle 'ем', \omega_5^S \rangle, \langle 'е', \omega_6^S \rangle, \langle 'ья', \omega_7^S \rangle, \langle 'ей', \omega_8^S \rangle, \langle 'ьям', \omega_9^S \rangle, \langle 'ей', \omega_{10}^S \rangle, \langle 'ьями', \omega_{11}^S \rangle, \langle 'ьях', \omega_{12}^S \rangle\}$  задает парадигматический класс существительных, представителем которого есть слово 'князь'.

Аналогично,

$$\sigma_{P=P_3}(R_{CFP\Omega(T)}) \cup \sigma_{P=P_9}(R_{CFP\Omega(T)}) \cup \sigma_{P=P_{11}}(R_{CFP\Omega(T)}) = R_{CFP\Omega^A} \quad (28)$$

задает отношение «квазиоснова – квазифлексия – грамматический класс – грамматическое значение» для лексем парадигматического типа  $T^A$ .

Проекция отношения  $R_{CFP\Omega^A}$  по второму столбцу (по атрибуту F) является множеством квазифлексий  $F(T^A)$  для лексем парадигматического типа  $T^A$ :

$$\pi_F(R_{CFP\Omega^A}) = \{\{f_0^{j_{P_i}}\}, \{f_1^{j_{P_i}}\}, \dots, \{f_{n(T^A)}^{j_{P_i}}\}\}, j_{P_i} = \overline{1, n_A}, \quad (29)$$

где  $F_0^{T^A} = \{f_0^{j_{P_i}}\}, F_1^{T^A} = \{f_1^{j_{P_i}}\}, \dots, F_{n(T^A)}^{T^A} = \{f_{n(T^A)}^{j_{P_i}}\}$ .  
Отношение

$$R_{F(T^A)} \subseteq F_0^{T^A} \times F_1^{T^A} \times \dots \times F_{n(T^A)}^{T^A}, \quad (31)$$

построенное на декартовом произведении  $F_i^{T^A}, i = \overline{0, n(T^A)}$ , разбивает множество  $F^{T^A} \subset F$  на классы эквивалентности  $\Pi_j(P, T^A)$ .

Кортежи  $\{f_0^j, f_1^j, \dots, f_{n(T^A)}^j\}$  задают классы  $\Pi_j(P, T^A)$ . Например кортеж  $\{\langle 'льв', \omega_1^A \rangle, \langle 'лого', \omega_2^A \rangle, \langle 'лому', \omega_3^A \rangle, \langle 'льв', \omega_4^A \rangle, \langle 'льв', \omega_5^A \rangle, \langle 'лом', \omega_6^A \rangle, \langle 'лая', \omega_7^A \rangle, \langle 'лой', \omega_8^A \rangle, \langle 'лой', \omega_9^A \rangle, \langle 'люю', \omega_{10}^A \rangle, \langle 'лой', \omega_{11}^A \rangle, \langle 'лой', \omega_{12}^A \rangle, \langle 'лое', \omega_{13}^A \rangle, \langle 'лого', \omega_{14}^A \rangle, \langle 'лому', \omega_{15}^A \rangle, \langle 'лое', \omega_{16}^A \rangle, \langle 'льв', \omega_{17}^A \rangle, \langle 'лом', \omega_{18}^A \rangle, \langle 'лье', \omega_{19}^A \rangle, \langle 'льх', \omega_{20}^A \rangle, \langle 'льв', \omega_{21}^A \rangle, \langle 'лье', \omega_{22}^A \rangle, \langle 'льми', \omega_{23}^A \rangle, \langle 'льх', \omega_{24}^A \rangle, \langle 'ел', \omega_{25}^A \rangle, \langle 'ла', \omega_{26}^A \rangle, \langle 'ло', \omega_{27}^A \rangle, \langle 'ль', \omega_{28}^A \rangle\}$  задает парадигматический класс адъективов, в состав которого входит прилагательное 'светлый'.

$$\bigcup_{i=6}^8 \sigma_{P=P_i}(R_{CFP\Omega(T)}) = R_{CFP\Omega^V} \quad (32)$$

<sup>7</sup> Рассматриваем лишь словоизменительные классы

задает отношение «квазиоснова – квазифлексия – грамматический класс – грамматическое значение» для лексем грамматического типа  $T^V$ .

Проекция отношения  $R_{CFP\Omega^V}$  по второму столбцу (по атрибуту  $F$ ):

$$\pi_F(R_{CFP\Omega^V}) = \{\{f_0^{j_{P_i}}\}, \{f_1^{j_{P_i}}\}, \dots, \{f_{n(T^V)}^{j_{P_i}}\}\}, j_{P_i} = \overline{1, n_V}, \quad (33)$$

где  $F_0^{T^V} = \{f_0^{j_{P_i}}\}, F_1^{T^V} = \{f_1^{j_{P_i}}\}, \dots,$

$$F_{n(T^V)}^{T^V} = \{f_{n(T^V)}^{j_{P_i}}\}. \quad (34)$$

На декартовом произведении элементов  $F_i^{T^V}, i = 0, n(T^V)$  строим отношение эквивалентности

$$R_{F(T^V)} \subseteq F_0^{T^V} \times F_1^{T^V} \times \dots \times F_{n(T^V)}^{T^V}, \quad (35)$$

которое разбивает множество  $F^{T^V} \subset F$  на классы эквивалентности  $\Pi_j(P, T^V)$ .

Кортежи  $\{f_0^j, f_1^j, \dots, f_{n(T^V)}^j\}$  задают классы  $\Pi_j(P, T^V)$ . Например парадигматический класс, представителем которого является глагол совершенного вида ‘*произойти*’, задается кортежем

$$\{\langle flex, i \mid i = 0, \dots, n(T^V) \rangle = \{\langle 'зойти', \omega_1^V \rangle, \langle 'зойду', \omega_2^V \rangle, \langle 'зойдешь', \omega_3^V \rangle, \langle 'зойдём', \omega_4^V \rangle, \langle 'зойдём', \omega_5^V \rangle, \langle 'зойдете', \omega_6^V \rangle, \langle 'зойдут', \omega_7^V \rangle, \langle 'зошёл', \omega_8^V \rangle, \langle 'зошла', \omega_9^V \rangle, \langle 'зошло', \omega_{10}^V \rangle, \langle 'зошли', \omega_{11}^V \rangle, \langle 'зойди', \omega_{12}^V \rangle, \langle 'зойдите', \omega_{13}^V \rangle, \langle 'зойдя', \omega_{14}^V \rangle, \langle 'зошедший', \omega_{15}^V \rangle, \langle 'сшедший', \omega_{15}^V \rangle\},$$

где  $\omega_i^V (i = 1, 2, \dots, 15)$  – комплексы грамматических значений  $\langle зало́г, на́клонение, вре́мя, чи́сло, ли́цо, ро́д \rangle$ , которые для данного примера парадигматического класса имеют такое содержание:

$\omega_1^V$  –  $\langle$  неопределенная форма/ инфинитив  $\rangle$ ;  
 $\omega_2^V$  –  $\langle$  действительный залог, изъявительное наклонение, будущее время, единственное число, первое лицо  $\rangle$ ;

$\omega_3^V$  –  $\langle$  действительный залог, изъявительное наклонение, будущее время, единственное число, второе лицо  $\rangle$ ;

$\omega_4^V$  –  $\langle$  действительный залог, изъявительное наклонение, будущее время, единственное число, третье лицо  $\rangle$ ;

$\omega_5^V$  –  $\langle$  действительный залог, изъявительное наклонение, будущее время, множественное число, первое лицо  $\rangle$ ;

$\omega_6^V$  –  $\langle$  действительный залог, изъявительное наклонение, будущее время, множественное число, второе лицо  $\rangle$ ;

$\omega_7^V$  –  $\langle$  действительный залог, изъявительное наклонение, будущее время, множественное число, третье лицо  $\rangle$ ;

$\omega_8^V$  –  $\langle$  действительный залог, изъявительное наклонение, прошедшее время, единственное число, третье лицо, мужской род  $\rangle$ ;

$\omega_9^V$  –  $\langle$  действительный залог, изъявительное наклонение, прошедшее время, единственное число, третье лицо, женский род  $\rangle$ ;

$\omega_{10}^V$  –  $\langle$  действительный залог, изъявительное наклонение, прошедшее время, единственное число, третье лицо, средний род  $\rangle$ ;

$\omega_{11}^V$  –  $\langle$  действительный залог, изъявительное наклонение, прошедшее время, множественное число, третье лицо  $\rangle$ ;

$\omega_{12}^V$  –  $\langle$  действительный залог, повелительное наклонение, будущее время, единственное число, второе лицо  $\rangle$ ;

$\omega_{13}^V$  –  $\langle$  действительный залог, повелительное наклонение, будущее время, множественное число, второе лицо  $\rangle$ ;

$\omega_{14}^V$  –  $\langle$  действительный залог, прошедшее время  $\rangle$ ;

$\omega_{15}^V$  –  $\langle$  действительный залог, прошедшее время, единственное число  $\rangle$ .

$$\sigma_{P=P_2}(R_{CFP\Omega(T)}) = R_{CFP\Omega^C}. \quad (36)$$

$$\pi_F(R_{CFP\Omega^C}) = \{\{f_0^{j_{P_i}}\}, \{f_1^{j_{P_i}}\}, \dots, \{f_{n(T^C)}^{j_{P_i}}\}\}, j_{P_i} = \overline{1, n_C}, \quad (37)$$

где  $F_0^{T^C} = \{f_0^{j_{P_i}}\}, F_1^{T^C} = \{f_1^{j_{P_i}}\}, \dots,$

$$F_{n(T^C)}^{T^C} = \{f_{n(T^C)}^{j_{P_i}}\}. \quad (38)$$

Отношение

$$R_{F(T^C)} \subseteq F_0^{T^C} \times F_1^{T^C} \times \dots \times F_{n(T^C)}^{T^C} \quad (39)$$

разбивает множество  $F^{T^C} \subset F$  на классы эквивалентности  $\Pi_j(P, T^C)$ . Кортежи  $\{f_0^j, f_1^j, \dots, f_{n(T^C)}^j\}$  задают классы  $\Pi_j(P, T^C)$ .

Отношение «квазифлексия – парадигматический класс»

$$R_{F\Pi} \subseteq R_{F(T)} \times N, \text{ где } N = \{1, 2, 3, \dots\}. \quad (40)$$

Объединение отношения  $R_{CFP\Omega(T)}$  и  $R_{F\Pi}$  по атрибуту  $F$ :

$$\pi_{C, R_{F\Pi}, F, P, \Omega(T), \Pi}(\sigma_{R_{CFP\Omega(T)}, F=R_{F\Pi}, F}(R_{CFP\Omega(T)} \times R_{F\Pi})) \quad (41)$$

дает разбиение множества  $W(P, T) \subset W_P \subset W_T$  на классы (парадигматические классы).

### 3. Оператор парадигматизации

Для автоматического построения полной парадигмы по исходной (начальной) форме  $x_0$  определяется оператор парадигматизации:

$$N: x_0 \rightarrow [x] = c(x) * \{f_0(x), f_1(x), \dots, f_n(x)\} \equiv \{c(x) * f_0(x), c(x) * f_1(x), \dots, c(x) * f_n(x)\}, \quad (42)$$

действие которого определяется отношением  $\pi(x^1, x^2)$ .

Оператор полной парадигматизации (который действует на ММ лексем  $W$ ) определяется по формуле

$$H = \sum_{i=1}^N H_i \cdot \delta(x; T_i), \quad (43)$$

где

$$\delta(x; T_i) = \begin{cases} 1, & x \in W(T_i), \\ 0, & x \notin W(T_i), \end{cases} \quad (44)$$

$N$  – количество парадигматических типов,  $H_i$  – оператор парадигматизации, действующий на ММ лексем соответствующего парадигматического типа  $W(T_i)$ <sup>8</sup>:

$$H_i : x_0 \rightarrow [x] \forall x \in W(T_i). \quad (45)$$

Поскольку имеет место (2), справедлива формула:

$$H_i = \sum_{j=1}^p H_i^j \cdot \delta(x; P_j), \quad (46)$$

где

$$\delta(x; P_j) = \begin{cases} 1, & x \in W(P_j), \\ 0, & x \notin W(P_j), \end{cases} \quad (47)$$

$i = 1, 2, \dots, N$ ;  $p$  – количество грамматических классов в множестве  $W(T_i)$ .

Оператор  $H_i^j$  действует на множестве лексем в пределах грамматического класса  $P_j$  парадигматического типа  $T_i$ :

$$H_i^j : x_0 \rightarrow [x] \quad \forall x \in W(P_j) \subseteq W(T_i) \quad (48)$$

В свою очередь, поскольку  $P_j$  является объединением парадигматических классов ( $P_j = \bigcup_{k=1}^n \Pi_k$ ), можно записать, что

$$\forall x \in W(\Pi_k) \subseteq W(P_j) \subseteq W(T_i),$$

$$H_i^j : x_0 \rightarrow c(x) * [F]_y^k \quad H_i^j = \sum_{k=1}^{n_j} H_i^{j,k} \cdot \delta(x, \Pi_k), \quad (49)$$

где  $H_i^{j,k}$  – оператор парадигматизации, который действует в пределах парадигматического класса  $\Pi_k$ ;  $i = 1, 2, \dots, N$ ;  $j = 1, 2, \dots, n_j$ ;  $n_j$  – количество парадигматических классов в ММ  $W(P_j) \subseteq W(T_i)$ ;  $[F]_y^k$  – множество наборов квазифлексий слов грамматического класса  $P_j$  парадигматического типа  $T_i$ ; функция

$$\delta(x; \Pi_k) = \begin{cases} 1, & x \in W(\Pi_k), \\ 0, & x \notin W(\Pi_k). \end{cases} \quad (50)$$

Таким образом, для каждого из парадигматических типов  $W(T_i)$  оператор парадигматизации определяется независимо.

<sup>8</sup> на ММ лексем каждого из парадигматических типов действует свой оператор парадигматизации. Прежде всего потому, что каждый из парадигматических типов характеризуется своим комплексом значений грамматических категорий, который отличен для каждого ПТ.

Оператор  $H$  отображает лексему  $x$  на ее полную парадигму  $[x]$ . Его реализация для лексики соответствующего флективного языка осуществляется при помощи словаря квазифлексий и набора алгоритмов построения полных словоизменительных парадигм. С помощью парадигматического словаря любой лексеме приписывается ее словоизменительный тип, далее по алгоритму морфологического анализа (с использованием набора алгоритмов построения полных словоизменительных парадигм) осуществляется грамматическая идентификация лексемы  $x$ . После этого лексема приобретает представление (4).

Алгоритмическая реализация оператора  $H^{-1}$  осуществляет процесс лемматизации, то есть сведение произвольной словоформы к ее исходной канонической форме.

Изложенная морфологическая модель составляет концептуальную основу для компьютерного моделирования и реализации функции парадигматических отношений для определенного класса флективных языков. Согласно изложенной концепции строение парадигматической лексикографической системы для флективного языка показано на рис. 9.

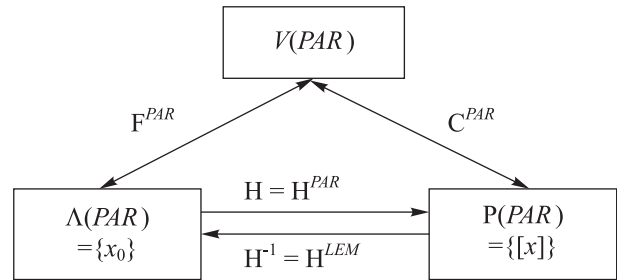


Рис. 9. Структура парадигматической Л-Системы

В приведенной схеме элементы имеют такую интерпретацию:

$V(PAR)$  – множество словарных статей грамматической Л-Системы;

$$\Lambda(PAR) = F^{PAR}V(PAR) = \{x_0\};$$

$$P(PAR) = C^{PAR}V(PAR) = \{[x]\};$$

$H = H^{PAR}$  – оператор парадигматизации:

$$H = H_{x_0}^{PAR} = [x];$$

$H^{-1} = H^{LEM}$  – оператор лемматизации:

$H^{LEM} \circ \chi(x) = x_0$ , где  $\chi(x)$  – любой элемент парадигмы  $[x]$ .

### Выводы

На основе описанного формализма нами была построена грамматическая лексикографическая система для русского языка, в основу которой положена классификация, базирующаяся на выделении в лексическом составе языка парадигмати-

ческих (словоизменительных) типов и подтипов, грамматических классов. Внутри грамматических классов выделялись парадигматические классы, каждый из которых включает лексемы с одинаковыми правилами образования определенных словоизменительных форм. Построен и программно реализован оператор парадигматизации, в результате чего каждому слову был поставлен в соответствие его парадигматический класс. Это позволяет автоматически получать полную словоизменительную парадигму любого слова. Реализация и верификация описанной модели морфологии русского языка выполнена на объеме лексики около 170 тыс. единиц. Действующая модель системы представлена в свободном доступе в Интернете на сайте Украинского лингвистического портала ([www.ulif.org.ua](http://www.ulif.org.ua), Словарь русского словоизменения).

**Список литературы:** 1. *Корпусна лінгвістика: Монографія / В.А. Широков, О.В. Бугаков, Т.О. Грязнухіна, Т.П. Любченко, О.Г. Рабулець, О.О. Сидоренко, Н.М. Сидорчук, И.В. Шевченко, О.О. Шипнівська, К.М. Якименко; Український мовно-інформаційний фонд НАН України.* — К.: Довіра, 2005. — 472 с. 2. *Широков В.А. Інформаційна теорія лексикографічних систем.* — К.: Довіра, 1998. — 331 с. 3. *Широков В.А. Феноменологія лексикографічних систем.* — К.: Наукова думка, 2004. — 327 с. 4. *Белоногов Г.Г., Зеленков Ю.Г. Алгоритм морфологического анализа русских слов // Вопросы информационной теории и практики.* М.: ВИНТИ, 1985, № 53. 5. *Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л. и др. Лингвистическое обеспечение системы автоматического перевода Этап-2.* — М.: Наука, 1989. — 296 с. 6. *Бидер И.Г., Большаков И.А., Еськова Н.А. Формальная модель русской морфологии.* — ИРЯ АН СССР, Проблемная группа по экспериментальной и прикладной лингвистике. — Вып. 112. — М.: 1978. — I — 60 с., II — 59 с. 7. *Хорошилов А.А. Автоматическая нормализация слов в системах обработки научно-технической информации: Автореф. дис. ... канд. техн. наук.* — М., 1987. — 18 с. 8. *Гельбух А.Ф., Сидоров Г.О. К вопросу об автоматическом морфологическом анализе флективных языков // Компьютерная лингвистика и интеллектуальные технологии: Тр. междунар. конф. "Диалог'2005"* — <http://www.dialog-21.ru/Archive/2005/Gelbukh%20Sidorov/GelbukhA.htm> 9. *Лингвистический энциклопедический словарь / Гл. ред. В.Н. Ярцева.* М.: Советская энциклопедия, 1990. — 688 с. 10. *Зализняк А.А. Грамматический словарь русского языка: Словоизменение.* — М.: Русский язык, 1978. — 878 с. (4-е изд., испр. и

доп.: 2003). 11. *Грязнухина Т.А., Любченко Т.П., Рабулець А.Г. Электронная версия грамматического словаря русского языка (А.А. Зализняк) как инструмент автоматического морфологического анализа русского текста // Докл. науч. конф. «Корпусная лингвистика и лингвистические базы данных».* — Санкт-Петербург, март 2002. — С. 63-70. 12. *Шевченко І.В. Алгоритмічна словозмінна класифікація української лексики // Мовознавство.* — 1996. — № 4-5. — С. 40-44. 13. *Шевченко І.В. Моделі та алгоритмічно-програмне забезпечення лексикографічних систем: Автореф. дис. ... канд. техн. наук.* — К., 2000. — С. 20. 14. *Гарант-парк-інтернет. Технології аналізу і пошуку текстової інформації / URL: <http://research.metric.ru>* 15. *Ножов І.М. Процесор автоматизованого морфологического анализа без словаря. Деревья и корреляция // Диалог'2000. Труды конференции — Протвино, 2000. — Т.2. — С. 284-290.* 16. *SRILM — The SRI Language Modeling Toolkit / URL: <http://www.speech.sri.com/projects/srilm>* 17. *Селезнев К. Обработка текстов на естественном языке // Журнал "Открытые системы", #12, 2003 ч // Изд. "Открытые системы" / URL: <http://www.osp.ru/os/2003/12/048.htm>* 18. *Сокирко А. В. Морфологические модули на сайте [www.aot.ru](http://www.aot.ru) // Диалог-2005.* 19. *Коваленко А. Вероятностный морфологический анализатор (стеммер) // Системный администратор.* — №1. — Октябрь, 2002. / URL: <http://linguist.nm.ru/ling/> 20. *Любченко Т.П. Програмно-технологічні аспекти створення граматичних лексикографічних систем // Проблеми програмування.* — 2007. — № 3. — С. 61-75. 21. *Любченко Т.П. Морфологічна модель словозміни флективної мови та електронний граматичний словник // Біоніка інтелекту: наук.-техн. журнал.* — 2006. — №1 (64). — С. 72-77. 22. *Кнут Д.Э. Искусство программирования.* — М.: Мир, 1977. — т.2.: Получисленные алгоритмы. 23. *Петровский А.Б. Метрические пространства множеств // Доклады Академии наук.* — 1995. — Т.344, № 2. — С. 175-177. 24. *Петровский А.Б. Комбинаторика множеств // Доклады Академии наук.* — 2000. — Т.370, №6. — С. 750-753. 25. *Петровский А.Б. Пространства множеств и множеств. — М.: Едиториал УРСС, 2003.* 26. *Петровский А.Б. Многокритериальное принятие решений по противоречивым данным: подход теории множеств // Информационные технологии и вычислительные системы.* — 2004. — №2. — С. 56-66. 27. *Коваль С.А. К вопросу о числе падежей русского существительного (Решения для компьютерной лингвистики) // <http://dialog-21.ru/Archive/2004/Koval.htm>* 28. *Успенский В.А. К определению падежа по А.Н. Колмогорову // Бюллетень Объединения по проблемам машинного перевода.* — № 5. — М.: [ I МГПИИЯ], 1957. — С. 11-18.

*Поступила в редколлегию 24.03.2008*

УДК 534.773



## МЕТОДИКА ОЦЕНКИ УРОВНЯ ГРОМКОСТИ ЗВУКОВОГО СИГНАЛА С УЧЁТОМ ОСОБЕННОСТЕЙ ЗВУКОВОСПРИЯТИЯ

И.В. Прасол<sup>1</sup>, А.С. Нечипоренко<sup>2</sup><sup>1</sup>ХНУРЭ, г. Харьков, Украина<sup>2</sup>ХНУРЭ, г. Харьков, Украина, alinanechiporenko@gmail.com

Статья посвящена проблеме регулирования уровня громкости в слуховых аппаратах. Рассмотрены особенности восприятия громкости звукового сигнала больными тугоухостью, в частности пациентами с диагнозом нейросенсорная тугоухость. Проведен обзор существующих технологий регулирования громкости в слуховых аппаратах. Предложена методика оценки уровня громкости, учитывающая особенности восприятия звуков больными тугоухостью, в следствие чего позволяющая оптимально оценить громкость и различимость речи.

НЕЙРОСЕНСОРНАЯ ТУГОУХОСТЬ, УРОВЕНЬ ГРОМКОСТИ, ЭКВИВАЛЕНТНЫЙ НЕПРЕРЫВНЫЙ УРОВЕНЬ ЗВУКОВОГО ДАВЛЕНИЯ, ОЦЕНКА ГРОМКОСТИ

### Введение

В настоящее время число людей, страдающих различными нарушениями слуха, непрерывно растёт. Отметим, что среди патологий слуховой системы преобладает нейросенсорная тугоухость, которая вызвана повреждением механизмов звуковосприятия улитки внутреннего уха. Лечение данного заболевания медикаментозными средствами на сегодняшний день малоэффективно, поэтому такое снижение слуха может быть компенсировано с помощью слухопротезирования. Слуховой аппарат должен скорректировать индивидуальные особенности слуха, характерные для определённого вида патологии. Современные тенденции таковы, что к слуховому аппарату предъявляются требования выполнения функции непрерывного анализа звуковой ситуации и адаптации к ней с одновременно максимально возможной разборчивостью речи и сохранением приятного и комфортного звучания. Ввиду вышесказанного серьёзной проблемой при слухопротезировании является разница восприятия уровня громкости больными тугоухостью в зависимости от смены акустической обстановки. Поэтому данная работа посвящена проблеме компенсации нарушения восприятия громкости больными тугоухостью.

### 1. Особенности восприятия громкости

Восприятие громкости основано на способности человеческого уха различать изменение звукового давления. Термин «громкость» достаточно трудно формализуем. Громкость является субъективной физиологической характеристикой звука и характеризует уровень слухового ощущения. В акустике звуковое давление обычно измеряется в децибелах (дБ) относительно порога слышимости. По определению, величина порога принята равной  $p_t = 0,00002 \text{ Па} = 20 \text{ мкПа}$ . Порог слышимости принимается за 0 дБ, а громкость вычисляется согласно формуле

$$L = 20 \log \left( \frac{p}{p_t} \right), \quad (1)$$

где  $L$  – громкость;  $p$  – звуковое давление, Па;  $p_t$  – порог слышимости, Па. Существует также и громкость записи, которая не является звуковым давлением. Громкость цифрового сигнала обычно измеряется относительно максимального цифрового уровня, принятого за 0 дБ, следовательно, выражается отрицательными величинами. Коррекция восприятия уровня громкости основана на таком понятии как субъективная громкость. Таким образом, звук одного и того же уровня громкости может восприниматься как громкий или тихий в зависимости от окружающих условий.

Восприятие различной громкости звуковых сигналов человеком с нормальным слухом обусловлено адаптацией слухового анализатора к акустической ситуации в конкретный момент времени. Однако разница в восприятии громкости пациентами, страдающими тугоухостью, является серьёзной проблемой, поскольку механизмы естественной адаптации нарушены. В результате повреждения наружных волосковых клеток в улитке внутреннего уха теряется способность к восприятию звуков низкой интенсивности. Звуки среднего уровня громкости воспринимаются как тихие, а громкие — как звуки со средним уровнем громкости. У некоторых больных нейросенсорной тугоухостью наблюдается феномен ускоренного нарастания громкости (ФУНГ), когда незначительное повышение интенсивности звука вызывает непропорциональное усиление ощущения громкости. ФУНГ характеризуется отсутствием слуховых ощущений у больного до определённого времени, пока уровень интенсивности звука не станет достаточно большим, после чего даже незначительное повышение интенсивности воспринимается на слух как слишком громкий звук [1], в некоторых случаях достигая порога дискомфорта. При проектировании слуховых аппаратов одной из основных

задач является предотвращение дискомфортных перегрузок для таких пациентов. Однако наличие феномена ускоренного нарастания громкости значительно усложняет этот процесс, поскольку у таких пациентов слишком узкие границы между порогами слышимости и дискомфорта.

## 2. Обзор технологий регулирования громкости в слуховых аппаратах

Для регулирования уровня громкости в слуховых аппаратах используют различные технологии. Как известно, по способу усиления слуховые аппараты делятся на линейные и нелинейные. Суть линейного усиления состоит в том, что все входящие звуки усиливаются с одинаковым коэффициентом усиления, независимо от их уровня громкости. Недостатком такого подхода является чрезмерное усиление громких звуков, которые являются причиной дискомфорта. Поэтому для устранения отдельных пиков используется клиппирование, которое, однако, вызывает искажение выходного сигнала. Существует более эффективный способ регулирования громкости, базирующийся на технологии компрессии и применяемый в слуховых аппаратах с нелинейным усилением — автоматическая регулировка усиления (АРУ), (лимитирование сигнала). Суть его состоит в том, чтобы автоматически понижать уровень сигнала таким образом, чтобы сигнал не превышал заданный максимальный уровень. Коэффициент усиления нелинейных слуховых аппаратов, имеющих функцию АРУ, зависит от интенсивности входного сигнала. До тех пор, пока уровень входного сигнала не достигнет определенной величины, называемой порогом срабатывания АРУ, коэффициент усиления остается постоянным, как у линейного аппарата. При превышении входным сигналом порога срабатывания АРУ, который устанавливается в процессе слухопротезирования в соответствии с индивидуальной потерей слуха, коэффициент усиления слухового аппарата снижается, что очень важно для протезирования нейросенсорной тугоухости с выраженным явлением ФУНГ [2]. В итоге удаётся избежать появления отдельных пиков, вызывающих ощущение дискомфорта у больного. Однако в результате лимитирования все звуки, превышающие установленный порог, имеют идентичную громкость, в то время как изначально они имели разную громкость.

Для оптимального частотно-зависимого отображения широкого динамического диапазона входных сигналов в суженный динамический диапазон остаточного слуха больного используется технология компрессии широкого динамического диапазона (*WDRC*). В частности метод *WDRC* применяется в многоканальных слуховых аппаратах в одной или нескольких частотных полосах, где па-

раметры компрессии настраиваются отдельно для каждого канала и регулируют интенсивность излишне громких звуков. Входящие сигналы низкого уровня получают значительное усиление, что делает их слышимыми, в то время как звуки высокого уровня усиливаются слабо, что препятствует появлению слишком громких звуков. Недостаток метода — чрезмерное усиление громких звуков, появление отдельных пиков, вызывающих ощущение дискомфорта. Однако механизм их возникновения отличается от скачков громкости, возникающих при линейном усилении, а их характер — иной, они проявляются в основном при очень сильном усилении тихих звуков [3], а не при обрезании громких, как при обычном усилении. Также к недостаткам компрессии широкого динамического диапазона можно отнести неестественное качество звучания, когда звуки высокого уровня становятся излишне приглушёнными и недостаточно громкими. В некоторых случаях не удаётся отобразить временные и частотные характеристики сигнала, важные для понимания речи в сложных акустических ситуациях. Чрезмерный уровень компрессии приводит к уплощению звуковой картины — все звуки стремятся к одинаковой громкости и невыразительности [3].

Если говорить о проблеме регулирования громкости в целом, вышеописанные методы позволяют выполнить грубое выравнивание уровней громкости поскольку в силу того, что динамический диапазон узок и поддается контролю, выравнивание пиковых значений даёт возможность приблизительно выровнять среднюю громкость сигнала. Однако исследования показали, что даже в этом случае существует значительный разброс в субъективном восприятии громкости звуковых сигналов при смене акустической обстановки [4]. Таким образом, применение технологии компрессии и лимитирования в слуховом аппарате не решает проблемы регулирования громкости. Для решения данной проблемы предлагается использовать методику оценки уровня громкости, основанную на стандартизированном методе измерения *LeqA*, учитывающую особенности восприятия звуков больными тугоухостью. Основной задачей данной методики является сохранение естественного соотношения различных уровней громкости выходного сигнала.

## 3. Описание методики оценки громкости

Известны методы определения уровня громкости, используемые в практике телевизионного вещания. Базируясь на данных, полученных в результате измерения среднего (*VU*) и пикового (*PPM*) уровней громкости, производится нормализация уровня сигнала. Однако следует отметить, что в данных методах не учитываются особенности

восприятия звуков слуховым аппаратом слушателя. В результате регулирование уровня громкости осуществляется на основании абсолютных уровней сигналов, тогда как восприятие звукового сигнала человеком зависит от амплитудно — частотной характеристики уха и среднего уровня сигнала. Для расчета оптимальной громкости восприятия пациентом звуковой картины целесообразно использовать фильтры, адаптирующие звук к АЧХ слуховой системы человека. Поэтому для оценки уровня громкости в слуховом аппарате предлагается использовать метод *LeqA*, метод измерения среднего значения уровня громкости звукового сигнала, учитывающий кривую взвешивания *A*, которая соответствует АЧХ слуховой системы человека при комфортном уровне звука. Данная кривая оптимально подходит для оценки громкости и различимости речи [5].

$$A_{wsp} = 10 \log \frac{p_A^2}{p_0^2}, \quad (2)$$

где  $A_{wsp}$  — уровень звукового давления;  $p_0$  — базовый уровень давления, (20μПа);  $p_A$  — текущий уровень давления, Па.

Как правило, амплитуда измеряемого звукового сигнала изменяется во времени. Поэтому изменение уровня звукового давления сигнала рассчитываем в пределах заданного промежутка времени согласно формуле:

$$L_{eq} = 10 \log \left( \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \frac{p_A^2}{p_0^2} dt \right), \quad (3)$$

где  $L_{eq}$  — эквивалентный непрерывный уровень звукового давления, Дб; ( $p_0$  — базовый уровень давления (20μПа);  $p_A$  — текущий уровень давления, Па;  $t_1$  — начало времени измерения;  $t_2$  — конец времени измерений).

Изменение амплитуды звукового сигнала во времени интерпретируется как изменение уровня звукового давления сигнала и представляет собой некоторую флуктуацию уровня громкости. Кривая взвешивания *A* является наиболее часто используемой из семейства кривых, первоначально определенная в Международном стандарте измерения звука IEC123 [6]. Данная кривая является характеристикой, используемой для «взвешивания» измеренных значений переменных в соответствии с их значениями по отношению к некоторым эталонным результатам.

*A*-взвешенное значение в Дб может быть определено как следующая функция:

$$W_A = 10 \log \left[ \frac{1,562339 f^4}{(f^2 + 107,65265^2)(f^2 + 737,86223^2)} \right] + 10 \log \left[ \frac{2,242881 \cdot 10^{16} f^4}{(f^2 + 20,598997^2)(f^2 + 12194,22^2)} \right], \quad (4)$$

где  $W_A$  — вес, который должен быть применен, Дб;  $f$  — частота, Гц.

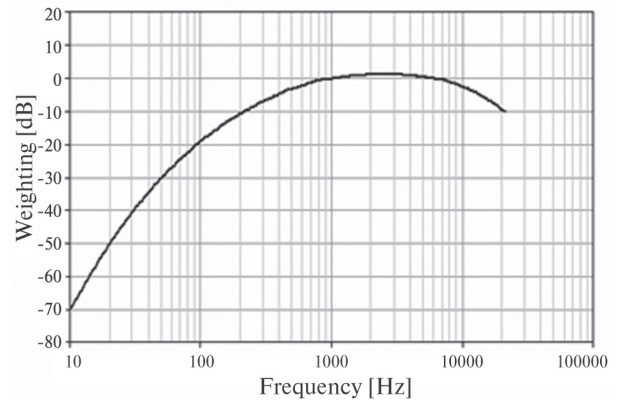


Рис. 1. Частотная характеристика измерения уровня звука

Для определения комфортного уровня громкости сигнала для конкретного пациента результаты его аудиологического обследования [7], в частности кривую тональной пороговой аудиометрии, предлагается соотносить с кривой взвешивания *A*, соответствующей АЧХ слуховой системы человека с нормальным слухом.

В качестве примера расчёта уровня громкости, рассмотрим три звуковых сигнала, имеющих одинаковый временной диапазон изменения амплитуды (рис. 2–4). Среднее значение амплитуды сигналов, приведенных на рис. 2, 3, равно 40 Дб. В результате компрессии динамического диапазона среднее значение амплитуды сигнала, изображённого на рис. 4, в два раза меньше и равно 20 Дб. Для оценки громкости сигнала в течение заданного промежутка времени производим расчёт эквивалентного непрерывного уровня звукового давления согласно (3). В результате получаем величину, характеризующую среднее значение уровня громкости на заданном временном интервале.

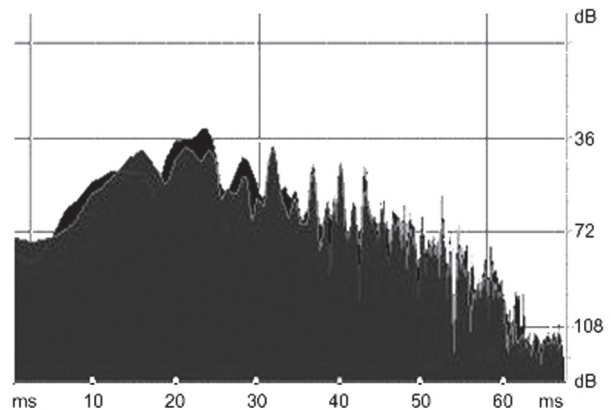


Рис. 2. Фрагмент звукового сигнала

Используя полученные данные, построим диаграмму уровней громкости (рис. 5) звуковых сигналов, изображенных на рис. 2–4. Из рис. 5 видно,

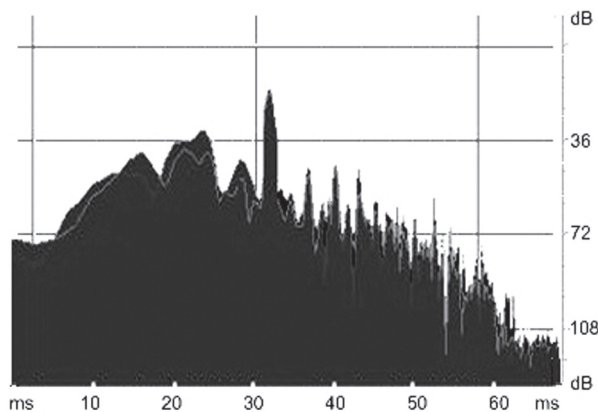


Рис. 3. Звуковой сигнал с отдельным пиком

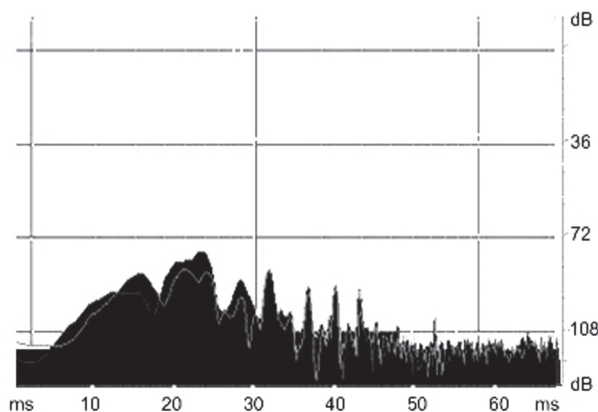


Рис. 4. Компрессированный звуковой сигнал

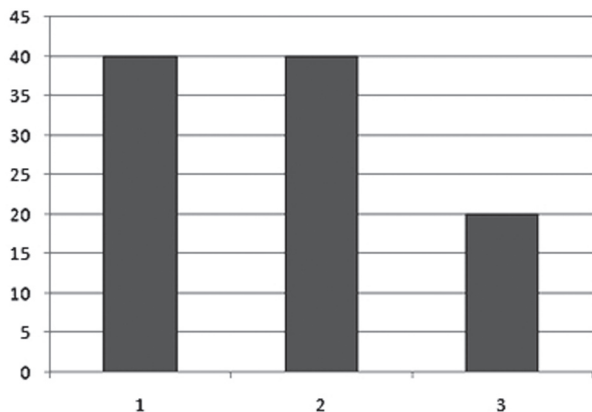


Рис. 5. Диаграмма уровней громкости

что уровень громкости сигналов, изображённых на рис. 2, 3, остаётся неизменным, несмотря на появление отдельного пика большой амплитуды. Таким образом, нивелируется влияние значений амплитуды отдельно взятых пиков, которые не должны оказывать влияние на изменение общей громкости фрагмента сигнала. Уровень громкости сигнала, приведенного на рис. 4, ниже, поскольку среднее значение амплитуды сигнала ниже, чем у предыдущих двух сигналов. В результате расчёта субъективной громкости звукового сигнала с помощью метода, учитывающего особенности восприятия

слуховой системой человека, а также используя данные измерения для регулирования уровня громкости звука в зависимости от акустической обстановки, можно рассчитать оптимальный уровень громкости восприятия для конкретного пациента, а также устранить лишние вариации громкости в пределах динамического диапазона остаточного слуха пациента. Соотношение различных уровней громкости выходного сигнала в слуховом аппарате предлагается регулировать относительно полученного среднего значения амплитуды.

### Выводы

В статье описана методика оценки уровня громкости, которую предлагается использовать в слуховом аппарате. Данная методика базируется на стандартизированном методе измерения LeqA, а также учитывает особенности восприятия звуков слуховой системой человека. Обосновано, что оценка уровня громкости звукового сигнала, поступающего в слуховой аппарат, необходима для определения соответствующей стратегии компрессии, а также для сглаживания динамических пиков и выбросов сигнала, которые приводят к возникновению ощущения дискомфорта у пациентов с тугоухостью. Умеренное применение компрессии позволяет сохранить естественное качество звучания и обеспечить передачу максимального количества временных и спектральных характеристик для понимания речи в сложных акустических ситуациях. Также, благодаря предварительному оцениванию громкости удаётся избежать высокого уровня искажений на малых уровнях громкости выходного сигнала. По результатам исследований сделан вывод, что применение данной методики в слуховых аппаратах даёт возможность регулирования различных уровней громкости выходного сигнала таким образом, чтобы сохранялось их первоначальное, естественное соотношение, а также обеспечивает непрерывный контроль уровня громкости звукового сигнала в режиме реального времени, что является необходимой базой для создания алгоритма измерения громкости в слуховом аппарате.

**Список литературы:** 1. Нейман Л.В., Богомильский М.Р. Анатомия, физиология и патология органов слуха и речи. — «Владос», — 2001. — 224 с. 2. О.М. Лисенко, Сучасні методи дослідження слуху людини // Монографія. — «КВІЦ». — 2002. — С. 21-22. 3. David Salomon Data compression, the complete reference. 3<sup>rd</sup> edition, Springer, 2004, — 54 с. 4. Alen I.Are Movies Too Loud?, conference SMPTE, 03.22.1997. 5. Fletcher, Munson, Journal of the Acoustic Society of America, vol.4, № 2, 1993, pp. 23. 6. United States Advanced Television Systems Committee. Standards Document A/54, pp. 49-52.

Поступила в редакцию 26.03.08

УДК 519.62



## КВАНТОРНАЯ АЛГЕБРА ПРЕДИКАТНЫХ ОПЕРАЦИЙ

Н.Т. Процай

ХНУРЭ, г. Харьков, Украина

В статье представлена кванторная алгебра предикатных операций. Доказана теорема о равносильности фундаментальной и кванторной алгебр предикатных операций. Сформулирована теорема о полноте кванторной алгебры предикатных операций. Приведены система основных тождеств и аксиоматическое построение кванторной алгебры.

КВАНТОРНАЯ АЛГЕБРА, АЛГЕБРА ПРЕДИКАТНЫХ ОПЕРАЦИЙ, ПРИКЛАДНАЯ АЛГЕБРА, ЛИНЕЙНЫЙ ЛОГИЧЕСКИЙ ОПЕРАТОР

### Введение

Первая научно-техническая революция содействовала усилению физических возможностей человека за счет создания механизмов и машин. Она потребовала интенсивного изучения внешнего по отношению к человеку физического мира и развития *числовой математики*. Вторая научно-техническая революция призвана содействовать усилению интеллектуальных возможностей людей за счет совершенствования искусственного интеллекта. Она требует объективного изучения внутреннего психологического мира человека и развития *математики интеллекта*, в роли которой выступает *логическая математика*. Для развития логической математики и ее приложений в искусственном интеллекте особую роль играет *алгебро-логический анализ естественного языка и мышления человека* [1]. Так, Джордж Буль, изучая в середине 19-го столетия логическую структуру сложных предложений естественного языка, открыл алгебру логики, которая впоследствии стала теоретической базой построения схем первых ЭВМ и программ для них. В конце того же века Готлоб Фреге, алгебраизируя естественный язык математики, извлек из него кванторы, играющие в современной логической математике важную роль. В 70-х годах 20-го столетия попытки формального описания морфологии естественного языка привели к открытию *алгебры предикатов*. Наконец, работы 90-х годов по абстрактному описанию смысла текстов естественного языка содействовали отысканию целого семейства *алгебр предикатных операций*. Фундаментальная алгебра предикатных операций, описанная в работе [2], – первая полная алгебра из этого семейства. Кроме того, для фундаментальной алгебры известна полная система законов. Эти качества фундаментальной алгебры сделали ее незаменимой для теоретических изысканий в области логической математики. Вместе с тем, для практических применений фундаментальная алгебра не всегда удобна, поскольку ее формулы и тождества не обладают достаточной компактностью и изяществом. В связи с этим в [3] разрабатывают алгебру предикатных операций, которая более удобна

для тех или иных приложений, чем фундаментальная алгебра – прикладную алгебру или, как ее еще называют, алгебру подстановочных операций. Прикладная алгебра изучена меньше, чем фундаментальная алгебра. В данной области имеются только отдельные разрозненные результаты. [3,4,5] Один из таких результатов связан с идеей введения так называемой *кванторной алгебры*, являющейся, как выяснилось, логическим аналогом интегрального исчисления и функционального анализа [1]. В прикладной алгебре кванторы вводятся в порядке ее консервативного расширения [3], кванторы вводятся просто по определению, выражаются через базис прикладной алгебры. Однако роль кванторов в классической математике, естественном языке, базах данных, информатике чрезвычайно велика. И сами кванторы обладают достаточной силой, чтобы на них основать особую алгебру предикатных операций.

*Цель статьи* – разработать кванторную алгебру предикатных операций и обосновать целесообразность ее применения для различных практических приложений. Для достижения поставленной цели, следует решить следующие *задачи*: 1) на основании базиса прикладной алгебры построить несократимый базис кванторной алгебры предикатных операций; 2) определить систему законов кванторной алгебры; 3) дать аксиоматическое описание кванторной алгебры предикатных операций.

### 1. Кванторная алгебра предикатных операций

В результате, консервативно расширив базис прикладной алгебры [3] кванторами общности и существования, получим какую-то новую алгебру предикатных операций с *базисом операций*, состоящим из дизъюнкции  $X \vee Y$ , отрицания  $\neg X$ , подстановок  $x_i / a(X)$ , ( $i = \overline{1, m}$ ,  $a \in U$ ), кванторов существования  $\exists x_i \in A_i(X)$  и общности  $\forall x_i \in A_i(X)$ , и с базисом элементов, образованным из предикатов равенства вида  $D(x_1, x_i) = \bigvee_{a \in M} x_1^a x_i^a$ , ( $i = \overline{2, m}$ ) и переменных  $X_j$  ( $X, Y \in N$ ;  $i, k = \overline{1, m}$ ;  $j = \overline{1, n}$ ;  $a \in A_i$ ). Базис этой алгебры избыточен, попытаемся его сократить. Один из кванторов можно исключить из базиса. Исключаем квантор общности:

$$\begin{aligned} \forall x_i(X) &= \overline{\overline{\forall x_i(X)}} = \overline{\bigwedge_{a \in U} x_i/a(X)} = \overline{\bigvee_{a \in U} \overline{x_i/a(X)}} = \\ &= \overline{\bigvee_{a \in U_i} x_i/a(\overline{X})} = \exists x_i(\overline{X}). \end{aligned}$$

Оказывается, можно исключить из базиса новой алгебры и операцию подстановки, выразив ее через квантор существования:  $x_i/a(X) = \exists x_i \in A_i(x_i^a \wedge X)$ . Операцию конъюнкции выразим посредством операций дизъюнкции и отрицания, имеющих в базисе прикладной алгебры, применяя законы де Моргана. Для выражения подстановки нам потребовались предикаты узнавания предмета  $x_i^a, (i = \overline{1, m}, a \in U)$ , а их нет в базисе прикладной алгебры. Вводим их в базис новой алгебры. Располагая предикатами узнавания предмета, дизъюнкцией и конъюнкцией, можем исключить из базиса новой алгебры предикаты равенства:  $D(x_1, x_i) = \bigvee_{a \in U} x_1^a x_i^a, (i = \overline{2, m})$ . Итак, мы получили новую алгебру предикатных операций, которая называется кванторной. Для ее получения мы исключили из базиса прикладной алгебры операции подстановки и предикаты равенства, а добавили кванторы существования  $\exists x_i(X), (i = \overline{1, m})$  и предикаты узнавания предмета  $x_i^a, (i = \overline{1, m}, a \in U)$ . Таким образом, приходим к следующему определению кванторной алгебры.

*Кванторной алгеброй* называется алгебра предикатных операций с базисом операций, образованным из квантора существования  $\exists x_i \in A_i(X), i = \overline{1, m}$ , отрицания  $\neg X$  и дизъюнкции  $X \vee Y$  всевозможных предикатных операций, и с базисом элементов, образованным из предикатов узнавания предметов  $x_i^a$  и предикатных переменных  $X_j, (X, Y \in N; i, k = \overline{1, m}; j = \overline{1, n}; a \in A_i)$ . Таким образом, несократимый базис кванторной алгебры состоит из операций дизъюнкции, отрицания и

$m$  кванторов существования, из  $\sum_{i=1}^m k_i$  предикатов узнавания предмета, где  $k_i$  – число предметов в множестве  $A_i (i = \overline{1, m})$  и из  $n$  предикатных переменных. Дизъюнктивно-конъюнктивная алгебра предикатов является подалгеброй кванторной алгебры, поскольку в базисе последней имеются дизъюнкция, конъюнкция (выражается посредством операций дизъюнкции и отрицания) и предикаты узнавания предмета, а также предикаты 0 ( $0 = X \wedge \overline{X} = \overline{X \wedge \overline{X}} = \overline{X \vee \overline{X}}$ ) и 1 (используем закон истинности для предикатов узнавания предмета:  $\bigvee_{b \in U} x_i^b = 1, i = \overline{1, m}$ ).

**Теорема:** Фундаментальная и кванторная алгебры предикатных операций равносильны.

*Доказательство.* Равносильность кванторной и фундаментальной алгебр будет доказана, если удастся базисные элементы и операции фундаментальной алгебры выразить через базисные элементы и операции кванторной алгебры, и наоборот.

1) Выражаем базис фундаментальной алгебры через базис кванторной. Базисными операциями в фундаментальной алгебре служат дизъюнкция и конъюнкция. Дизъюнкция присутствует уже в базисе кванторной алгебры. Конъюнкцию выражаем через дизъюнкцию и отрицание:

$$X \wedge Y = \overline{\overline{X} \vee \overline{Y}}. \quad (1)$$

Базисными элементами в ней являются элемент 0, предикаты узнавания предмета  $x_i^a$  и предиката  $X_j^P$ . Предикаты узнавания предмета  $x_i^a$  в базисе кванторной алгебры уже имеются. Осталось выразить элементы  $X_j^P$  и 0. Это можно сделать посредством зависимостей:  $0 = X_j \wedge \overline{X_j}$ , где  $X_j$  – предикатная переменная, а операция конъюнкции выражена выше (1). В свою очередь предикат узнавания предиката выражается с помощью постоянных предикатов  $P$  и предикатных переменных  $X_j, j = \overline{1, n}$  [6]:

$$\begin{aligned} X_j^P &= \forall x_1 \in A_1 (\forall x_2 \in A_2 \dots (\forall x_m \in A_m (X_j(x_1, x_2, \dots, x_m) \sim \\ &\sim P(x_1, x_2, \dots, x_m) \dots)); X \sim P = \overline{\overline{X} P} \vee X P. \end{aligned}$$

Постоянные предикаты  $P$  выражаем на языке дизъюнктивно-конъюнктивной алгебры предикатов, которая, как было сказано выше, является подалгеброй кванторной алгебры. Любой фиксированный предикат  $P$  выражается формулой дизъюнктивно-конъюнктивной алгебры предикатов с помощью операций  $\vee$  и  $\wedge$ , примененных к предикатам узнавания предмета  $x_i^a, (i = \overline{1, m}, a \in U)$ . Весь базис фундаментальной алгебры выражен через базис кванторной алгебры.

2) Выражаем базис кванторной алгебры через базис фундаментальной. Операциями в несократимом базисе кванторной алгебры, как было сказано выше, служат дизъюнкция, которая присутствует также и в базисе фундаментальной алгебры, а также операция отрицания и квантор существования. Операцию отрицания в кванторной алгебре выражаем посредством *законов отрицания* для предикатов узнавания предмета  $x_i^a = \bigvee_{\substack{b \in A_i \\ b \neq a}} x_i^b, (i = \overline{1, m})$  и узнавания предиката  $\overline{X_j^P} = \bigvee_{\substack{Q \in B_j \\ Q \neq P}} X_j^Q, j = \overline{1, n}$ . Квантор существования выражается через дизъюнкцию и подстановки по формуле:

$$\exists x_i \in A_i(X) = \bigvee_{a \in A_i} x_i/a(X).$$

Операция дизъюнкции есть в базисе фундаментальной алгебры. Операции подстановки  $x_i/a (i = \overline{1, m})$  перемещаются непосредственно на предикаты 0 и  $x_i^b (i = \overline{1, m}, b \in A_k)$  и на предикатные переменные  $X_j (j = \overline{1, n})$ , а затем исключаются вовсе с помощью зависимостей [6]:

$$x_i/a(0) = 0;$$

$$x_i/a(x_k^b) = x_k^b, \text{ если } i \neq k ;$$

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

$$x_i/a(X_j) =$$

$$= \bigvee_{P \in B_j} P(x_1, x_2, \dots, a, \dots, x_m) X_j^{P(x_1, x_2, \dots, x_i, \dots, x_m)};$$

Предикат 0 и предикаты узнавания предмета в базисе фундаментальной алгебры уже есть. Предикат 1 получаем, используя законы истинности для предикатов узнавания предмета и предикатов узнавания предиката:

$$\overline{x_i^a} = \bigvee_{\substack{b \in A_i \\ b \neq a}} x_i^b, (i = \overline{1, m}); \quad \bigvee_{P \in B_j} X_j^P = 1, \quad j = \overline{1, n}.$$

Постоянные предикаты  $P$  выражаем на языке дизъюнктивно-конъюнктивной алгебры предикатов, которая является подалгеброй фундаментальной алгебры. Предикаты узнавания предиката в фундаментальной алгебре уже есть.

Элементами в несократимом базисе произвольной кванторной алгебры служат предикаты  $x_i^a$  ( $i = \overline{1, m}$ ), которые присутствуют также и в базисе фундаментальной алгебры, а также предикатные переменные  $X_j$  ( $j = \overline{1, n}$ ). Последние выражаются по формуле  $X_j = \bigvee_{P \in B_j} P(x_1, x_2, \dots, x_m) X_j^{P(x_1, x_2, \dots, x_m)}$  через дизъюнкцию, конъюнкцию, предикаты узнавания предиката и постоянные предикаты. Дизъюнкция, конъюнкция, предикаты узнавания предиката в фундаментальной алгебре имеются, постоянные же предикаты на языке фундаментальной алгебры, как уже упоминалось, выражаются. Весь базис кванторной алгебры выражен через базис фундаментальной алгебры. Итак, фундаментальная и кванторная алгебры равносильны. Теорема доказана.

Из данной теоремы и теоремы о полноте фундаментальной алгебры [2] непосредственно следует теорема о полноте кванторной алгебры.

**Теорема о полноте кванторной алгебры:** Кванторная алгебра полна при любых  $U$ ,  $m$  и  $n$ .

Кванторная алгебра с несократимым базисом, состоящим из операций дизъюнкции и конъюнкции, из  $m$  кванторов существования и  $m$  кванторов общности, из предикатов узнавания предмета, где  $k_i$  – число предметов в множестве  $A_i, i = \overline{1, m}$ , из  $n$  предикатных переменных и константы 0, представлена в работах [6,7]. Также в этих же работах доказана теорема о равносильности фундаментальной и кванторной алгебр предикатных операций с вышеуказанным базисом.

## 2. Законы кванторной алгебры

Формулами кванторной алгебры служат кванторные выражения. Кванторные выражения используются в двух ролях: для записи предикатных

операций и для записи отношений второго порядка. Во второй роли они используются в классической математике для записи ее утверждений. В первой роли – в информатике для преобразования отношений первого порядка. Кванторные выражения в числовой математике используются как логический инструментарий математики. Лишь логическая математика поставила вопрос о предании им статуса операций некой алгебры, а именно – алгебры предикатных операций, сделала их предметом специального изучения.

Кванторная алгебра имеет такую же описательную силу, как и язык. Однако наряду с наличием полного базиса кванторной алгебры, у нас нет полной системы законов данной алгебры. Законы однозначно определяют алгебру, если их набор полон, мы можем образовать все ее свойства. Тожеств в кванторной алгебре очень много. Здесь вопрос лишь в том, чтобы отобрать в полную систему законов законы из числа уже известных. Приведем ниже известные законы для кванторов, которые являются базой для поиска полной системы законов кванторной алгебры.

*Законы исключения и введения кванторов:*

$$\forall y P(x) = P(x); \quad \exists y P(x) = P(x).$$

Эти равенства справедливы для любого предиката  $P$ . Квантор можно опускать, если предикат, на который он действует, не зависит от переменной, по которой берется этот квантор. И обратно – в этом случае квантор можно вводить.

*Законы переброски кванторов:*

$$\forall y (P(x) Q(y)) = P(x) \forall y Q(y);$$

$$\forall y (P(x) \vee Q(y)) = P(x) \vee \forall y Q(y);$$

$$\exists y (P(x) Q(y)) = P(x) \exists y Q(y);$$

$$\exists y (P(x) \vee Q(y)) = P(x) \vee \exists y Q(y).$$

Это – обобщение законов исключения и введения кванторов. Квантор по переменной  $y$  можно перебрасывать через любой предикат, не зависящий от переменной  $y$ .

*Законы замены связанной переменной:*

$$\forall x P(x) = \forall y P(y); \quad \exists x P(x) = \exists y P(y).$$

Связанную квантором переменную можно менять на любую другую, но только в тех случаях, когда это не приводит к коллизии предметных переменных. Например неверно, что  $\forall x \forall y P(x, y) = \forall x \forall x P(x, x)$ . Связанную переменную нельзя заменять на такую другую, которая уже имеется в том же кванторном выражении. Но можно заменить предметную переменную  $y$  в той же формуле на переменную  $z$ , так как она не встречается в этой формуле.

*Частные законы дистрибутивности для кванторов:*

$$\forall xP(x) \wedge \forall xQ(x) = \forall x(P(x) \wedge Q(x)),$$

$$\exists xP(x) \vee \exists xQ(x) = \exists x(P(x) \vee Q(x)).$$

Эти равенства верны для любых предикатов  $P$  и  $Q$ . Это — дистрибутивные свойства кванторов, регулирующие раскрытие скобок и вынос кванторов за скобки. Под символом  $x$  можно понимать также и целый набор переменных. После раскрытия скобок первое условие становится более сильным, а второе — более слабым. Поэтому имеют место такие следования:

$$\forall xP(x) \vee \forall xQ(x) \Rightarrow \forall x(P(x) \vee Q(x)),$$

$$\exists xP(x) \wedge \exists xQ(x) \Rightarrow \exists x(P(x) \wedge Q(x)).$$

Обратные следования неверны.

*Общие законы дистрибутивности для кванторов:*

Можно так скорректировать частные законы дистрибутивности, приведенные выше, несколько усложнив их, что они будут выполняться в любой комбинации знаков  $\forall, \wedge; \forall, \vee; \exists, \wedge; \exists, \vee$ :

$$\begin{aligned} \forall xP(x) \vee \forall xQ(x) &= \forall x(P(x) \vee \forall yQ(y)) = \\ &= \forall x\forall y(P(x) \vee Q(y)) \end{aligned}$$

$$\begin{aligned} \exists xP(x) \wedge \exists xQ(x) &= \exists x(P(x) \wedge \exists yQ(y)) = \\ &= \exists x\exists y(P(x) \wedge Q(y)). \end{aligned}$$

Эти законы верны и при прежних комбинациях знаков:

$$\forall xP(x) \wedge \forall xQ(x) = \forall x\forall y(P(x) \wedge Q(y)),$$

$$\exists xP(x) \vee \exists xQ(x) = \exists x\exists y(P(x) \vee Q(y)).$$

Однако, как правило, при комбинациях знаков  $\forall, \wedge$  и  $\exists, \vee$  переходят на частные дистрибутивные законы, так что последние два закона используются редко (например, когда надо приводить кванторные выражения к стандартной форме, так называемой пренексной стандартной форме Сколема). При переводе этих законов мы воспользовались законами замены связанной переменной и законами переброски кванторов, которые были приведены выше. Имеется важный частный случай, когда можно раскрывать скобки в кванторных выражениях при комбинациях знаков  $\forall, \vee$  и  $\exists, \wedge$ :

$$\forall y(P(x) \vee Q(x, y)) = P(x) \vee \forall yQ(x, y),$$

$$\exists y(P(x) \wedge Q(x, y)) = P(x) \wedge \exists yQ(x, y).$$

Здесь один из компонентов логической суммы (произведения) не зависит от переменной, по которой берется квантор. Для перехода от левой части равенств к правой используются законы переброски кванторов.

*Коммутативные законы для кванторов:*

$$\forall x\forall yP(x, y) = \forall y\forall xP(x, y),$$

$$\exists x\exists yP(x, y) = \exists y\exists xP(x, y).$$

Однако необходимо отметить, что коммутативные законы справедливы только для одноименных кванторов. Разноименные кванторы менять местами нельзя. Тем не менее, в данном выражении левое условие сильнее правого, поэтому имеет место следование:

$$\exists x\forall yP(x, y) = \forall y\exists xP(x, y).$$

Обратное следование, вообще говоря, неверно.

*Кванторные законы де Моргана:*

$$\overline{\forall xP(x)} = \exists x\overline{P(x)}; \overline{\exists xP(x)} = \forall x\overline{P(x)}.$$

При переброске отрицания через знак квантора он заменяется на противоположный. С помощью этих законов и закона двойного отрицания получают зависимости, выражающие один квантор через другой.

*Законы замены кванторов:*

$$\overline{\forall xP(x)} = \exists x\overline{P(x)},$$

$$\overline{\forall x, yP(x, y)} = \exists x\forall y\overline{P(x, y)},$$

$$\overline{\exists x, yP(x, y)} = \forall y\forall x\overline{P(x, y)}.$$

Последние два правила непосредственно следуют из первого правила замены кванторов и коммутативных законов для кванторов.

Это только наиболее известные законы. Подобных тождеств — необозримое множество. Актуальная задача — разобраться в тождествах кванторной алгебры; найти полные и несократимые системы законов, удобные для человека и машины.

### 3. Аксиоматическое построение кванторной алгебры предикатных операций

Аксиоматическое построение кванторной алгебры — это важнейшее направление дальнейшего развития в теории интеллекта. Задание понятий признаками является мощным инструментом, позволяющим определить данное понятие по задающим его свойствам. Попробуем аксиоматически задать кванторную алгебру.

Дизъюнкцию и отрицание определяем законами булевой алгебры. Предикаты узнавания предмета и предикаты узнавания предиката определяем законами отрицания, истинности и ложности для предикатов узнавания предмета и предикатов узнавания предиката соответственно [6]. Остается аксиоматически определить квантор существования. Квантор будет иметь вид:

$$\exists x \in MP(x)K(x, y) = Q(y).$$

Общее решение этой системы логических уравнений:  $F(P) = \exists x \in AP(x)K(x, y)$ .  $F$  — произвольная аддитивная и однородная предикатная операция;  $K(x, y)$  — произвольно выбираемый бинарный предикат на  $A \times B$ .  $P(x)$  — унарный предикат на  $A$ ,  $Q(y)$  — унарный предикат на  $B$ .  $F(P) = Q$ .

$$\exists x \in AP(x) K(x, y) = Q(y) \quad (2)$$

Равенство (2) дает общий вид линейного логического оператора. Таким образом, общий вид квантора существования можно задать свойствами аддитивности и однородности:  $F(P \vee Q) = F(P) \vee F(Q)$ ;  $F(\lambda P) = \lambda F(P)$ ,  $\lambda \in \Sigma$ . Вместо однородности достаточно использовать свойство нуля:  $F(0) = 0$ . Возможно, такая система тождеств кванторной алгебры будет полной. Это надо проверять и исследовать. То, что здесь изложено — это только заготовка для дальнейшего исследования.

### Выводы

Основным результатом работы является разработка кванторной алгебры предикатных операций с базисом элементов, состоящим из предикатов узла предмета и из предикатных переменных, и базисом операций, состоящим из операций дизъюнкции, отрицания и кванторов существования. Алгебры предикатных операций полезно использовать при проектировании систем обработки информации, различных информационных структур и их электронных схем. При изучении механизмов интеллекта человека алгебры предикатных операций целесообразно использовать для формульной записи свойств (законов) интеллектуального поведения испытуемого. Само же это поведение удачно описывается на языке алгебры предикатов [2]. С помощью формул алгебр предикатных операций изящно и просто выражается смысловая структура предложений и текстов естественного языка [8].

В статье также была предпринята попытка аксиоматического описания кванторной алгебры. Однако данная алгебраическая система изучена еще слабее, чем прикладная алгебра. Главным ее недостатком является то, что неизвестна полная система тождеств. В работе приведены наиболее известные законы. Подобных тождеств — необозримое множество. Актуальная задача — разобраться в тождествах кванторной алгебры; найти полные и несократимые системы законов, удобные для человека и машины; разработать способы тождественного преобразования формул; найти стандартные формы предикатных операций. Если бы мы имели удобные (и единственные для каждой предикатной операции) стандартные формы и умели приводить к ним любые кванторные выражения, то отпала бы необходимость в доказательствах. Но до этого пока еще очень далеко. Необходимы дополнительные исследования, чтобы до конца разработать эту алгебраическую систему.

Совершенно неизученна и область решения кванторных уравнений: не существует регулярного метода решения систем логических уравнений. А к решению логических уравнений (в том числе и кванторных) сводится задача формирования машиной ответов на запросы. Любой вопрос — это требование найти решение некоторого логического уравнения, системы логических уравнений, обычно кванторных. Отвечая на вопросы, мы, сами того не сознавая, решаем «в уме» логические уравнения. В интеллектуальных системах природа на 95 % использует линейные логические преобразования [7], а линейные логические преобразования, как было показано выше — это ничто иное, как кванторы.

Имеется еще одна проблема, которая пока до конца не решена, но ее чрезвычайно важно решить. Мы сейчас находимся на втором этаже абстракции (алгебры предикатных операций), а есть еще бесконечно много этажей алгебр. Хотелось бы иметь не бесконечную систему алгебр, а одну — единственную алгебру, так называемую универсальную алгебру предикатов, которая бы заменила эту бесконечную иерархию алгебр.

**Список литературы:** 1. Шабанов-Кушнарченко Ю.П. Алгебраизация логики как катализатор информатизации // Радиоэлектроника и информатика. — Харьков: ХНУРЭ. — 2001. № 4. — С. 10-11. 2. Дударь З.В., Кравец Н.С., Шабанов-Кушнарченко Ю.П. О фундаментальной алгебре предикатных операций // АСУ и приборы автоматизации. — Харьков: ХНУРЭ. — 1998. — Вып. 49. — С. 68-77. 3. Дударь З.В., Кравец Н.С., Шабанов-Кушнарченко Ю.П. О прикладной алгебре предикатных операций // Проблемы бионики: наук.-техн. журнал. — 1998. — Вып. 49. — С. 78-87. 4. Баталин А.В. Модификация алгебры подстановочных операций // Сб. материалов 8 Международного молодежного форума «Радиоэлектроника и молодежь в XXI веке». — Харьков: ХНУРЭ, 2004. — Ч.2. — С. 115. 5. Баталин А.В. О прикладных вариантах алгебры предикатных операций // Сб. материалов 10 Международного конф. «Теория и техника передачи, приема и обработки информации». — Харьков-Туапсе, 2004. — Ч.2. — С. 36-39. 6. Бондаренко М.Ф., Дударь З.В., Процай Н.Т., Черкашин В.В., Чикина В.А., Шабанов-Кушнарченко Ю.П. Алгебра предикатов и предикатных операций // Радиоэлектроника и информатика. — Харьков: ХНУРЭ. — 2000. — №3. — С.15-23. 7. Лещинский В.А. Модели бинарных логических сетей и их применение в искусственном интеллекте. Дисс. ... канд. техн. наук: 05.13.23. — Харьков: ХНУРЭ, 2006 — 160 с. 8. Дударь З.В., Гвоздинская Н.А., Шабанов-Кушнарченко Ю.П. О математическом описании смысла текстов естественного языка // Проблемы бионики: наук.-техн. журнал. — Харьков: ХНУРЭ. — 1998. Вып 48. — С. 141-149.

Поступила в редколлегию 31.03.2008

УДК 681.3.07



## ИСПОЛЬЗОВАНИЕ ВЕЙВЛЕТ-РАЗЛОЖЕНИЯ ИЗОБРАЖЕНИЯ ДЛЯ КОМПЕНСАЦИИ ГЕОМЕТРИЧЕСКИХ ИСКАЖЕНИЙ

О.А. Кобылин

ХНУРЭ, м. Харьков, Украина, kblin@kture.kharkov.ua

Использование уменьшенных копий изображения, полученных в результате применения вейвлет-разложения, позволяет эффективно проводить компенсацию геометрических искажений, возникающих в системах технического зрения. Рассмотрены базовые модели геометрических искажений. Использование уменьшенных копий изображения позволяет сократить время методов его нормализации.

НОРМАЛИЗАЦИЯ ИЗОБРАЖЕНИЙ, ВЕЙВЛЕТ-АНАЛИЗ, ГРУППА ПРЕОБРАЗОВАНИЙ, СИСТЕМА ТЕХНИЧЕСКОГО ЗРЕНИЯ, ЭТАЛОН

### Введение

В системах технического зрения (СТЗ) используется множество различных методов обработки изображения [1, 4]. Однако не все они достаточно эффективны для различных типов изображений. Эффективность методов обработки изображений зависит от способа их представления. Именно от способа представления изображения зависит результативность его анализа. Существуют различные способы представления, такие как – частотная область Фурье, область Гильбертова представления и другие. Далеко не все способы представления информации приемлемы для использования в современных СТЗ. Поэтому на современном этапе развития СТЗ наблюдается большой интерес к разработке новых способов представления информации, в частности, используются различные алгоритмы быстрого преобразования Фурье, варианты косинусного и синусного преобразования, алгоритмы быстрого преобразования Хартли и другие.

В современных СТЗ необходимо в реальном масштабе времени проводить геометрическую компенсацию изображений для последующего сравнения с эталоном. Одним из методов, используемым в СТЗ, является нормализация изображений.

Нормализация изображения заключается в компенсации геометрических преобразований, полученных в результате отклонения входного изображения от эталонного. В зависимости от группы преобразования вычисляются параметры, позволяющие компенсировать геометрические преобразования. В зависимости от решаемой задачи допускается, что объект подвергнут искажениям и находится в заданной группе преобразований. Определение вида преобразования позволяет существенно сократить время на определение параметров для компенсации геометрических преобразований.

Цель работы – используя вейвлет – разложение изображения, провести геометрическую компенсацию искажений, возникающих в СТЗ в заданной группе преобразований.

Вейвлет-преобразование является одним из видов спектрального анализа.

Вейвлет-преобразование (wavelet transform) в настоящее время получило распространение во многих сферах применения обработки изображений: в медицине, в астрономии, в промышленности, системах компьютерной графики и других отраслях.

Главная идея вейвлетного преобразования – частотно-временное представление сигнала [2, 3].

Под вейвлет-преобразованием (wavelet transform) понимают разложение с помощью функций, каждая из которых является сдвинутой и масштабированной копией одной функции – материнского вейвлета [3].

Вейвлет-преобразование функции  $f(t)$  определяется выражением:

$$W[f(t)] = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \varphi\left(\frac{t-b}{a}\right) dt,$$

где  $a$  – масштаб;  $b$  – центр временной локализации; функция  $\varphi(t)$  является материнским вейвлетом и удовлетворяет условию  $\int_{-\infty}^{+\infty} \varphi(t) dt = 0$ .

Если изображение представлено в виде  $B(x, y) \in L_2(R)$  (рис. 1), то разложение вычисляется по формулам:



Рис. 1. Пример изображения

$$vv_j(k, l) = \frac{1}{2} \sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} h'(2k-n) h'(2l-m) vv_{j-1}(n, m),$$

$$wv_j(k, l) = \frac{1}{2} \sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} g'(2k-n) h'(2l-m) vv_{j-1}(n, m),$$

$$vw_j(k, l) = \frac{1}{2} \sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} h'(2k-n) g'(2l-m) vv_{j-1}(n, m),$$

$$ww_j(k,l) = \frac{1}{2} \sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} g'(2k-n)g'(2l-m)vv_{j-1}(n,m),$$

где  $j$  – степень разложения;

$$m,n=0\dots 2^j-1, k,l=0\dots 2^j-1;$$

$$j=0..J, L=2^j-1, h'(n)=h(L-n-1),$$

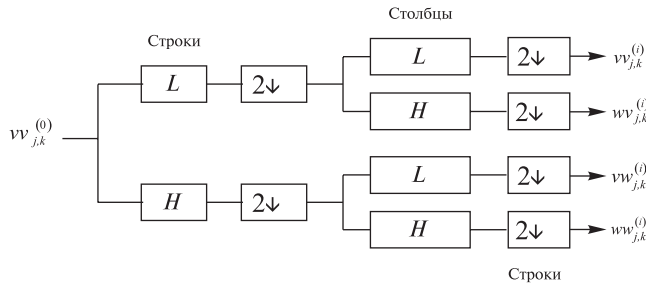
$$g'(n)=g(L-n-1);$$

$n$  и  $m$  – координаты пикселей изображения  $vv_{j,k}^{(i)}$ ;  $k$  и  $l$  – координаты пикселей изображения  $ww_{j,k}^{(i)}$ ;  $vv_{j,k}^{(i)}$ ,  $ww_{j,k}^{(i)}$ ;  $h(t)u g(t)$  – импульсные характеристики низкочастотного и высокочастотного фильтра.

Рассмотрим последовательность изображений, полученных в результате разложения. Схема разложения и пример разложения тестового изображения на 3 уровня представлен на рис. 2 и 3-5 соответственно.

В результате разложения получим по три матрицы горизонтальных, вертикальных и диагональных составляющих с одинаковым разрешением, принадлежащих каждому уровню разложения. При этом, размер уровней декомпозиции изображения будет равен: для первого уровня  $128 \times 128$ , второго  $64 \times 64$  и третьего  $32 \times 32$ .

Из рисунка ясно видно структуру изображения, состоящую из ярко выраженных горизонтальных, вертикальных и диагональных составляющих, которые характеризуют изображение (рис. 3-5).



( $L$  – низкочастотный фильтр,  
 $H$  – высокочастотный фильтр)

Рис. 2. Двухмерное дискретное ВП



Рис. 3. Результат разложения по уровням декомпозиции 1-й уровень

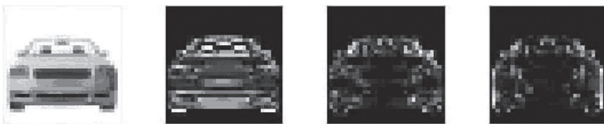


Рис. 4. Результат разложения по уровням декомпозиции 2-й уровень

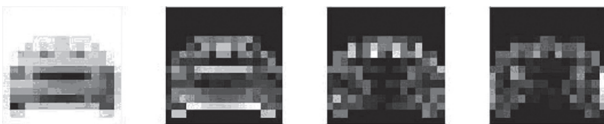


Рис. 5. Результат разложения по уровням декомпозиции 3-й уровень

В результате разложения получим набор изображений, представляющих уменьшенные копии изображения

$$WB^{(0)}(x,y) \supset WB^{(1)}(x,y) \supset \dots \supset WB^{(k)}(x,y),$$

где  $WB^{(0)}(x,y)$  – исходное изображение;  $WB^{(1)}(x,y) \supset \dots \supset WB^{(k)}(x,y)$  – его уменьшенные копии, полученные с применением декомпозирующего фильтра  $h$ .

Полученные уменьшенные копии изображения  $WB^{(1)}(x,y) \supset \dots \supset WB^{(k)}(x,y)$  оптимально подходят для выполнения процедур нормализации изображения.

### Компенсация геометрических искажений для различных моделей смещений

Проведем анализ различных моделей смещений, возникающих в СТЗ.

Произведем смещение изображения  $B(x,y)$  вдоль оси  $Y$ . Данное смещение математически имеет следующий вид:

$$B(x,y) = B_1(x,y-m), \quad (1)$$

где  $B, B_1$  – изображение;  $m$  – параметр смещения вдоль оси  $Y$ .

В этом виде преобразования объект не меняет своих геометрических размеров и ориентации, смещается вдоль относительно  $Y$ .

Так как изображение не изменило формы, а произошло его смещение вдоль оси  $Y$ , то можно утверждать, что данное преобразование относится к группе преобразования параллельных переносов вдоль координатной оси  $Y$  и принадлежит группе  $G_{c,y}$ . Группа  $G_{c,y}$  – однопараметрическая.

Для группы  $G_{c,y}$  нормализация смещения будет иметь вид:

$$F_C(B) = B(x,y+\Phi_2), \quad (2)$$

где  $\Phi_2$  – функционал, определяющий параметры  $m$  в зависимости (1):

$$\Phi_2 = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} m v v_0^0(n,m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v_0^0(n,m)} - \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} m v v^0(n,m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v^0(n,m)}.$$

Произведем смещение изображения  $B(x,y)$  вдоль оси  $X$ . Данное смещение математически имеет следующий вид:

$$B(x,y) = B_1(x-n,y), \quad (3)$$

где  $B, B_1$  – изображение;  $n$  – параметры смещения вдоль оси  $X$ .

В этом виде преобразования объект не меняет своих геометрических размеров и ориентации, смещается вдоль относительно оси  $X$ .

Так как изображение не изменило формы, а произошло его смещение вдоль оси  $X$ , то можно

утверждать, что данное преобразование относится к группе преобразования параллельных переносов вдоль координатной оси  $X$  и принадлежит группе  $G_{c,x}$ . Группа  $G_{c,x}$  – однопараметрическая.

Нормализация группы смещения  $G_{c,x}$  будет иметь вид:

$$F_C(B) = B(x + \Phi_1, y), \quad (4)$$

где  $\Phi_1$  – функционал, определяющий параметры  $n$  в зависимости:

$$\Phi_1 = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} n v v_0^0(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v_0^0(n, m)} - \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} n v v^0(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v^0(n, m)}.$$

Произведем смещение изображения  $B(x, y)$  вдоль оси  $X$  и  $Y$ . Данное смещение математически имеет следующий вид:

$$B(x, y) = B(x-n, y-m), \quad (5)$$

где  $B$  – изображение;  $n, m$  – параметры смещения вдоль координатных осей.

В этом виде преобразования объект не меняет своих геометрических размеров и ориентации, смещается вдоль относительно осей  $X$  и  $Y$ .

Для группы смещений  $G_{c,x,y}$  нормализация будет иметь вид:

$$F_C(B) = B(x + \Phi_1, y + \Phi_2), \quad (6)$$

где  $\Phi_1, \Phi_2$  – функционалы, определяющие соответственно параметры  $n$  и  $m$  в зависимости ():

$$\Phi_1 = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} n v v_0^0(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v_0^0(n, m)} - \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} n v v^0(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v^0(n, m)};$$

$$\Phi_2 = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} m v v_0^i(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v_0^i(n, m)} - \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} m v v^i(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v^i(n, m)}.$$

Рассмотрим смещения «косой сдвиг». Группы  $G_{h,x}, G_{h,y}$  – «косой сдвиг» – являются однопараметрическим преобразованием. Произведем преобразование с помощью смещения «косой сдвиг» изображения  $B(x, y)$  вдоль оси  $X$ .

Данное смещение математически имеет следующий вид:

$$B_0(x, y) = B(x + h y, y), \quad (7)$$

где  $h$  – параметр «косого сдвига» вдоль оси абсцисс.

Параметр  $h$  задают в виде  $h = tg\alpha$ . В этом случае  $\alpha$  задает угол сдвига.

При сдвиге вдоль оси ординат группы  $G_{h,y}$ :

$$B_0(x, y) = B(x, y + h x), \quad (8)$$

Преобразование при этом виде смещения происходит вдоль оси ординат, вдоль оси абсцисс, а также вдоль прямой с некоторым угловым коэффициентом.

В случае смещения «косой сдвиг» вид нормализатора для модели (7) и (8) будет иметь вид:

$$F_{H_x}(B) = B(x + \Phi y, y), \quad (9)$$

$$F_{H_y}(B) = B(x, y + \Phi x), \quad (10)$$

где  $\Phi$ , – функционал вида:

$$\Phi = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} n m v v (n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} m^2 v v (n, m)}.$$

Пусть объект  $O$  вращается вокруг оси. В этом случае объект  $O$  не удаляется от камеры и не меняет своих геометрических размеров. Все получаемые изображения рассматриваются в одной плоскости  $\beta$ . Изображения  $B_1$  и  $B_2$  объекта  $O$  отличаются лишь углом поворота  $\theta$ .

Угол поворота  $\theta$  является единственным параметром преобразования. В декартовой системе координат эталонное и входное изображения связаны зависимостью:

$$B(x, y) = B_0(x \cos(\theta) + y \sin(\theta) - x \sin(\theta) + y \cos(\theta)). \quad (11)$$

Нормализатор группы  $G_u$  имеет вид:

$$F_U(B) = B(x \sin \Phi + y \cos \Phi, -x \sin \Phi + y \cos \Phi), \quad (12)$$

где  $\Phi$  – функционал вида:

$$\Phi = \frac{1}{2} \operatorname{arctg} \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} 2 v v (n, m) (n^3 m + n m^3)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v (n, m) (n^4 + m^4)}.$$

Подвергнем изображение изменению масштаба. При данном изменении все точки объекта удаляются на одинаковое расстояние от камеры. Математическая модель преобразования масштаба соответствует выражению вида:

$$B(x, y) = B(\kappa x, \kappa y), \quad (13)$$

где  $B$  – эталонное изображение;  $\kappa$  – параметры сжатия.

В случае одномерного изменения масштаба нормализатор будет иметь вид:

$$F_d(B) = B(x \Phi, y \Phi), \quad (14)$$

где  $\Phi$  – функционал вида:

$$\Phi = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v (n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v_0 (n, m)}.$$

В случае неравномерного изменения масштаба нормализатор имеет следующий вид:

$$F_D(B) = B(x\Phi_1, y\Phi_2), \quad (15)$$

где  $\Phi_1$  – функционал, определяющий параметр сжатия вдоль оси абсцисс;  $\Phi_2$  – функционал, определяющий параметр сжатия вдоль оси ординат.

В этом случае функционал для нормализатора (15) можно построить, базируясь на свойствах эллипса энергии изображения:

$$\Phi_1 = \sqrt{\frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} n^2 vv(n,m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} vv(n,m)}}, \quad \Phi_2 = \sqrt{\frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} m^2 vv(n,m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} vv(n,m)}}.$$

Математическая модель перспективы имеет вид:

$$B_\lambda(x, y) = B\left(\frac{x}{\lambda y + 1}, \frac{y}{\lambda y + 1}\right), \quad (16)$$

где  $\lambda$  – параметр гомологии вдоль оси ординат.

Общий вид нормализатора для этой модели представим в виде:

$$F_P(B) = B\left(\frac{x}{\Phi y + 1}, \frac{y}{\Phi y + 1}\right), \quad (17)$$

где – функционал  $\Phi$  определяет параметр параболической гомологии вида:

$$\Phi = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} \frac{vv(n,m)}{n^4}}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} \frac{vv(n,m)}{n^3}}.$$

В результате анализа первого уровня разложения, с помощью формул вычисления средней разности, максимальной разности и среднеквадратической погрешности можно утверждать, что для компенсации геометрических искажений можно использовать и последующие уровни разложения изображения. Поэтому нормализаторы и функционалы для базовых групп для различных уровней разложения можно записать в следующей форме:

1. Для модели (3) группы  $G_{c,x}$  нормализатор примет вид:

$$WF_C(B) = WB(x + \Phi w^i_1, y),$$

где

$$\Phi w^i_1 = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} n v v_0^i(n,m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v_0^i(n,m)} - \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} n v v^i(n,m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v^i(n,m)}.$$

2. Для модели (2) группы  $G_{c,y}$  нормализатор примет вид:

$$WF_C(B) = WB(x, y + \Phi w^i_2),$$

где

$$\Phi w^i_2 = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} m v v_0^i(n,m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v_0^i(n,m)} - \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} m v v^i(n,m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v^i(n,m)}.$$

3. Для модели (5) группы  $G_{c,x,y}$  нормализатор примет вид:

$$WF_C(B) = WB(x + \Phi w^i_1, y + \Phi w^i_2),$$

где

$$\Phi w^i_1 = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} n v v_0^i(n,m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v_0^i(n,m)} - \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} n v v^i(n,m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v^i(n,m)},$$

$$\Phi w^i_2 = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} m v v_0^i(n,m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v_0^i(n,m)} - \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} m v v^i(n,m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v^i(n,m)}.$$

4. Для модели (13) группы  $G_m$  нормализатор в случае равномерного и неравномерного изменения масштаба имеет вид:

$$WF_d(B) = WB(x\Phi w^i, y\Phi w^i),$$

$$WF_d(B) = WB(x\Phi_1 w^i, y\Phi_2 w^i),$$

где

$$\Phi w^i = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v^i(n,m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v_0^i(n,m)},$$

$$\Phi_1 w^i = \sqrt{\frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} n^2 v v^i(n,m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v^i(n,m)}},$$

$$\Phi_2 w^i = \sqrt{\frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} m^2 v v^i(n,m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v^i(n,m)}}.$$

5. Для модели (7) группы  $G_{h,x}$  нормализатор примет вид:

$$WF_{Hx}(B) = WB(x + \Phi w^i, y),$$

где

$$\Phi w^i = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} n m v v^i(n,m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} m^2 v v^i(n,m)}.$$

6. Для модели (8) группы  $G_{h,y}$  нормализатор примет вид:

$$WF_{Hy}(B) = WB(x, y + \Phi w^i x),$$

где

$$\Phi w^i = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} nm v v^i(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} m^2 v v^i(n, m)}.$$

7. Для модели (11) группы  $G_u$  нормализатор имеет вид:

$$F_U(B) = B(x \sin \Phi w^i + y \cos \Phi w^i, -x \sin \Phi w^i + y \cos \Phi w^i),$$

где

$$\Phi w^i = \frac{1}{2} \arctg \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} 2 v v^i(n, m) (n^3 m + n m^3)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v^i(n, m) (n^4 + m^4)}.$$

8. Для модели (16) группы  $G_p$  нормализатор перспективы имеет вид:

$$F_p(B) = B\left(\frac{x}{\Phi w^i y + 1}, \frac{y}{\Phi w^i y + 1}\right),$$

где

$$\Phi w^i = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} \frac{v v^i(n, m)}{n^4}}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} \frac{v v^i(n, m)}{n^3}}.$$

### Выводы

В статье предложено использовать уменьшенные копии изображений, полученные в результате вейвлет - разложения, для компенсации геометрических искажений, возникающих вследствие взаимного расположения видеокамеры и наблюдаемого объекта. Геометрические искажения рассматривались в заданной группе преобразования. Использование вейвлет-разложения изображений позволяет сократить время на геометрические преобразования (нормализации) в системах технического зрения с целью дальнейшего распознавания наблюдаемого объекта.

**Список литературы:** 1. Путьтин Е.П., Аверин С.И. Обработка изображений в робототехнике. – М.: Машиностроение, 1990. – 320 с. 2. Петухов А.П. Введение в теорию базисов всплесков. – СПб.: Изд-во СПбГТУ, 1999. – 132 с. 3. Малла С. Вэйвлеты в обработке сигналов. – М.: Мир, 2005. – 671 с. 4. Форсайт Д., Понс Ж. Компьютерное зрение. Современный подход: Пер. с англ. – М.: Изд. дом «Вильямс», 2004. – 928 с.

*Поступила в редколлегию 1.04.2008*

УДК 519.859



## ПРИНЯТИЕ РЕШЕНИЙ В ИНФОРМАЦИОННЫХ СИСТЕМАХ РЕШЕНИЯ ЗАДАЧ ГЕОМЕТРИЧЕСКОГО ПРОЕКТИРОВАНИЯ

И. В. Гребенник<sup>1</sup>, Т. Е. Романова<sup>2</sup>, С. Б. Шеховцов<sup>3</sup>

<sup>1</sup>ХНУРЭ, г. Харьков, Украина, grebennik@onet.com.ua

<sup>2</sup>ИПМаш НАНУ, г. Харьков, Украина, sherom@kharkov.ua

<sup>3</sup>ХНУВД, г. Харьков, Украина, tarom7@yahoo.com

Статья является продолжением исследований задач геометрического проектирования. В ней содержится материал, лежащий в основе обобщения знаний, необходимых для построения многофакторных оценок при выборе наилучшего решения в информационных системах решения оптимизационных задач размещения, основанных на знаниях о математических моделях и современных методах локальной и глобальной оптимизации.

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ, ЗАДАЧИ РАЗМЕЩЕНИЯ, ПРИНЯТИЕ РЕШЕНИЙ, МНОГОФАКТОРНЫЕ ОЦЕНКИ

### Введение

Теория геометрического проектирования изучает круг фундаментальных и прикладных проблем, направленных на реализацию идеи математического моделирования процесса размещения реальных объектов и создание эффективных методов оптимизации этого процесса в соответствии с заданным критерием оптимальности.

Обзор зарубежных публикаций [1-4], посвященных решению оптимизационных задач размещения (packing&cutting), дает возможность сделать следующие выводы.

При решении конкретной оптимизационной задачи размещения геометрических объектов строятся различные математические модели или приводятся некоторые алгоритмические представления ее постановки в зависимости от области науки, отрасли промышленности или непосредственно от исследователя. Вследствие этого, для решения задач рассматриваемого класса, как правило, используют эвристические методы. Кроме того, высокий уровень организации архитектуры баз знаний [5] интеллектуальных систем решения задач размещения геометрических объектов, к сожалению, базируется на узкой классификации пространственных форм геометрических объектов, ограничений и функций цели.

В [6, 7] предлагается математическая модель системы решения задач оптимального размещения геометрических объектов, основанной на знаниях о единой конструктивной математической модели рассматриваемого класса задач [8] и современных методах оптимизации.

Целью статьи является формирование многофакторных оценок при выборе наилучшего решения на основе знаний о математических моделях и методах в информационных системах решения оптимизационных задач размещения.

Полагаем, что множество исходных данных  $x_0 \in X_0$  порождает множество реализаций  $X_1$  мате-

матической модели задач размещения, которые отличаются некоторыми математическими характеристиками.

Действительно, в общем случае реализация  $x_1 \in X_1$  математической модели задач размещения характеризуется размерностью пространства; видом  $\Phi$ -функций, описывающих множество допустимых решений  $D$ ; числом неравенств, формирующих  $D$ ; непрерывностью или дискретностью переменных; линейностью или нелинейностью функции цели.

Это означает, что в общем случае элементы  $x_1 \in X_1$  требуют различных методов для их эффективного решения. Обозначим через  $X_2$  множество всех методов, предназначенных для решения задач размещения.

Из приведенных выше рассуждений следует, что множество  $X_0$  порождает множество реализаций  $X_1$  математической модели, для решения которых используется множество методов  $X_2$ , алгоритмов  $X_3$ , программ  $X_4$ .

### 1. Постановка задачи

Запишем векторное представление [7] элементов как  $x_0 \in X_0$ ,  $x_1 \in X_1$ ,  $x_2 \in X_2$ ,  $x_3 \in X_3$ ,  $x_4 \in X_4$ . Тогда принятие решений в информационных системах оптимизации размещений сводится к решению следующей оптимизационной задачи [6]:

$$f(x_0, x_R^*) = \underset{x_R \in X_R}{extr} f(x_0, x_R), \quad (1)$$

где  $x_0 \in X_0$ ;  $x_R = (x_1, x_2, x_3, x_4) \in X_1 \times X_2 \times X_3 \times X_4 = X_R \subset R^l$ .

Функционал  $f(x_0, x_R)$  является векторным, то есть  $f(x_0, x_R) = (\alpha_1(x_0, x_R), \alpha_2(x_0, x_R), \dots, \alpha_k(x_0, x_R))$ , где  $\alpha_1, \dots, \alpha_k$  — критерии, определяющие эффективность решения задач размещения и затраты вычислительных ресурсов на его получение.

Из поставленной задачи (1) очевидна необходимость формирования функционала  $f(x_0, x_R)$ .

**З а м е ч а н и е 1.** Полагаем, что вычислительные ресурсы позволяют решить любую задачу размещения с исходными данными  $x_0 \in X_0$ .

**З а м е ч а н и е 2.** Под методом решения  $x_2 \in X_2$  понимается либо метод, предназначенный для решения только оптимизационных задач геометрического проектирования, либо способ, в основе которого лежит некоторый базисный метод оптимизации (метод ветвей и границ, градиентный метод, симплекс-метод и т. д.), модифицированный для решения класса задач геометрического проектирования [9-11].

Следует отметить, что на основе одного и того же базисного метода может быть несколько модификаций методов решения задач геометрического проектирования, а также различные композиции нескольких методов решения данного класса задач, которые в данном исследовании представлены как различные методы.

В общем случае каждому методу  $x_2$  может соответствовать более чем один алгоритм  $x_3$ , а каждому алгоритму  $x_3$  – более чем одна программа  $x_4$ . Однако для удобства анализа модели и простоты изложения материала полагаем, что между элементами множеств  $X_2, X_3$ , а также  $X_3, X_4$  задано инъективное отображение, то есть каждому элементу  $x_2 \in X_2$  соответствует один и только один элемент  $x_3 \in X_3$ , а каждому элементу  $x_3 \in X_3$  соответствует один и только один элемент  $x_4 \in X_4$ .

Таким образом, задача (1) сводится к решению следующей задачи:

$$f(x_0, x_p^*) = \text{extr}_{x_p \in X_p} f(x_0, x_p), \quad (2)$$

где  $x_0 \in X_0$ ;  $x_p = (x_1, x_2) \in X_1 \times X_2 = X_p$ .

В данной работе в качестве критериев, оценивающих решение задачи размещения понимается точность  $\alpha_1$  (допустимое решение, локальный экстремум, приближение к глобальному экстремуму, глобальный экстремум) решения задачи и быстроедействие (время решения)  $\alpha_2$ , то есть:

$$f(x_0, x_p) = (\alpha_1(x_0, x_p), \alpha_2(x_0, x_p)). \quad (3)$$

Так как функционал  $f(x_0, x_p)$  является векторным, то следует начать с построения критериев  $\alpha_1, \alpha_2$ .

## 2. Критерий точности

Поскольку в предлагаемом исследовании не учитываются погрешности алгоритмов  $X_3$  и программ  $X_4$ , связанные с точностью вычислительных процессов, то точность  $\alpha_1$  решения задачи объективно зависит от погрешности задания исходных данных, от исходных данных, влияющих на выбор реализации  $x_1 \in X_1$  математической модели задачи и выбор метода  $x_2 \in X_2$ , а также непосредственно от точности метода  $x_2 \in X_2$ .

**З а м е ч а н и е 3.** В общем случае в классе задач размещения любому элементу  $x_0 \in X_0$  соответствует более чем один элемент. Каждый элемент

$x_1^i, i = 1, \dots, k$  должен строиться в соответствии с исходными данными  $x_0$  так, чтобы все реализации математической модели задачи были эквивалентны с точки зрения их погрешностей.

Рассмотрим понятие точности метода  $x_2 \in X_2$ .

**З а м е ч а н и е 4.** В общем случае любому элементу  $x_1 \in X_1$  соответствует более чем один элемент  $x_2^j \in X_2, j = 1, \dots, l$ .

Метод  $x_2^j$  осуществляет решение конкретной задачи с определенной точностью  $\alpha_{12}$ . Точность  $\alpha_{12}$  метода решения задач размещения зависит от исходных данных  $x_0$ , построенной реализации  $x_1^i$  математической модели и непосредственно от возможностей метода  $x_2^j \in X_2$ , связанных с достигаемой точностью данного экстремума.

Таким образом, в общем виде точность  $\alpha_1$  решения конкретной задачи совпадает с точностью  $\alpha_{12}$  метода  $x_2 \in X_2$  и имеет вид  $\alpha_1 = \alpha_{12} = \alpha_{12}(x_0, x_1, x_2)$ . Следовательно, решить задачу (2) по критерию точности означает определить

$$\alpha_1(x_0, x_p^*) = \alpha_{12}(x_0, x_p^*) = \text{extr}_{x_p \in X_p} \alpha_{12}(x_0, x_p). \quad (4)$$

В силу дискретности и небольшой мощности множеств  $X_1, X_2$  ниже предлагается подход к решению задачи (4), основные положения которого заключаются в следующем.

1. Выделяется подмножество  $X_1'$  реализаций  $x_1^i, i = 1, \dots, k$ , математической модели, построенных соответственно с учетом допустимых погрешностей исходных данных  $x_0 \in X_0$ .

2. Каждому элементу  $x_1^i \in X_1', i \in \{1, \dots, k\}$  ставится в соответствие подмножество  $X_2^i \subset X_2 = \{x_2^j, j = 1, \dots, n_i\}$  методов, которые могут осуществить решение конкретной реализации математической модели  $x_1^i, i \in \{1, \dots, k\}$ .

3. Элементы  $x_p^r, r = 1, \dots, \sum_{i=1}^k n_i$  множества  $X_p$  упорядочиваются по критерию точности.

**З а м е ч а н и е 5.** Для реализации п. 1 требуются фундаментальные исследования, связанные с введением отношения эквивалентности с точки зрения погрешности построения реализаций  $x_1^i, i = 1, \dots, k$ , по исходным данным  $x_0 \in X_0$ .

**З а м е ч а н и е 6.** П. 2 предполагается осуществлять на основании векторного представления реализации математической модели  $x_1$  и метода  $x_2$ , используя функциональную предикатную связь между элементами вектора математической модели и вектора метода решения задач геометрического проектирования [7].

Пусть реализация  $x_1$  математической модели задачи описывается вектором  $a = (a_1, a_2, \dots, a_n)$ , где  $a_i, i = 1, \dots, n$ , – признаки математической модели задачи, влияющие на выбор метода  $x_2$  из множества  $X_2$ . Пусть также элемент  $x_2 \in X_2$  задается вектором  $b = (b_1, b_2, \dots, b_m)$ , где  $b_j, j = 1, \dots, m$ , – признаки метода ре-

шения, значение которых непосредственно определяет элемент  $x_2 \in X_2$ .

Каждый элемент  $a_i$  является в общем случае некоторой предикатной функцией  $\psi_i$  от элементов вектора  $x_0$ , то есть  $a_i = \psi_i(x_0)$ .

Каждый элемент  $b_j$  является в общем случае некоторой предикатной функцией  $\psi_j$  от элементов вектора  $a$ , то есть  $b_j = \psi_j(a)$ .

Таким образом, строится множество  $X_2^i = \{x_2^1, x_2^2, \dots, x_2^{k_i}\}$  методов решения для конкретной  $i$ -й реализации математической модели  $x_1^i, i \in \{1, \dots, k\}$ .

Предположим, что множество  $X_p$  построено, то есть все элементы  $x_p^r (r=1, \dots, l, l = \sum_{i=1}^k n_i)$  множества  $X_p$  определены.

Каждый элемент  $x_p^r$  из множества  $X_p$  имеет некоторую погрешность  $\alpha_1^r$ . Следовательно, с одной стороны, имеется множество  $X_p$  пар  $(x_1, x_2)^r = x_p^r$ , с другой – множество  $\alpha_p$  погрешностей элементов  $x_p^r, r = 1, 2, \dots, l$ . Тогда выполнение п. 3 сводится к решению следующей задачи. Требуется определить  $x_p^* \in X_p$ , для которого

$$\alpha_1^* = \min\{\alpha_1^1, \dots, \alpha_1^l\}, \quad (5)$$

где  $\alpha_1^* = \alpha_1(x_0, x_p^*)$ .

Для сравнения элементов  $\alpha_1^i, \alpha_1^j$  множества  $\alpha_1^r$  используется классификация методов по критерию точности в рассматриваемом классе задач.

При решении задач размещения используются точные и приближенные методы глобальной и локальной оптимизации, которые, в свою очередь, определяются множеством методов поиска глобального, локального экстремумов и локального экстремума с последующим перебором. Таким образом, по возможностям метода  $x_2$  в смысле точности достижения экстремума формируются классы  $[X_2^s]$  методов с соответствующим порядком  $s$  приоритета по точности.

Пусть  $k$  – число реализаций математической модели задачи,  $n_i$  – число методов для решения каждой из  $x_1^i, i \in \{1, \dots, k\}$ . Если  $k > 1, n_i > 1$ , то необходимо сравнение элементов по точности решения, которое предполагается осуществить следующим образом.

Сформируем подмножества  $\hat{X}_2^i, i=1, \dots, s$  методов  $x_2^{ij}$  по принадлежности к классам  $[X_2^s]^i$ . Каждое из подмножеств  $\hat{X}_2^i \subset [X_2^i]$  может быть либо пустым множеством, либо иметь мощность, не превышающую  $l = \sum_{n=1}^k n_i$ . Из всех множеств  $\hat{X}_r^s$  рассматривается только то  $X_r^{*S}$ , которое имеет приоритет по точности класса, к которому это множество принадлежит.

Все элементы  $x_2 \in X_r^{*S}$ , где  $S \leq l$ , эквивалентны с точки зрения точности решения. Поэтому эле-

менты, соответствующие методам  $x_2, r=S$ , имеют точность  $\alpha_1^*$  и применимы для решения конкретной задачи по критерию точности для исходных данных  $x_0$ .

Возможны следующие частные случаи определения наилучшей пары  $(x_1^i, x_2) = x_p^r$  при решении задачи (5).

1. Пусть для исходных данных  $x_0$  существует только одна реализация  $x_1 \in X_1$ , а элементу  $x_1$  соответствует только один метод  $x_2 \in X_2$ . Тогда, естественно, пара  $x_p = (x_1, x_2)$  есть решение задачи (5).

2. Если реализация  $x_1 \in X_1$  единственна, но ей соответствует более чем один метод  $x_2^j, j=1, \dots, n_0$ , то выбирается тот элемент  $x_2^j, j \in \{1, \dots, n_1\}$ , который имеет приоритет по принадлежности к классам  $[X_2^s]$ .

3. Если  $k > 1, a n_i = 1, i = 1, \dots, k$ , то выбирается такая пара  $(x_1^i, x_2^j)$ , для которой  $x_2^j$  принадлежит к классу методов, имеющему приоритет по точности.

4. Если  $k > 1$  и все они могут быть реализованы одним и тем же методом  $x_2$ , то все пары  $(x_1^i, x_2), i = 1, \dots, k$  эквивалентны в смысле точности решения задачи.

### 3. Критерий быстродействия

Перейдем к рассмотрению критерия  $\alpha_2$  быстродействия решения задач размещения. Аналогично замечанию о погрешностях алгоритмов и программ при исследовании критерия  $\alpha_1$  точности решения задачи в данной работе не учитывается и быстродействие алгоритмов  $X_3$  и программ  $X_4$ , связанных с быстродействием вычислительных процессов. Поэтому быстродействие решения задач размещения объективно зависит от исходных данных, влияющих на выбор реализации  $x_1 \in X_1$  и выбор метода  $x_2 \in X_2$ .

Согласно замечанию 3 исходным данным  $x_0 \in X_0$  соответствует множество  $\{X_1^*\} = \{x_1^1, \dots, x_1^k\} \subset X_1$  реализаций математической модели задачи.

Полагаем, что каждому элементу  $x_1^i, i=1, \dots, k$  соответствует в общем случае не пустое множество  $X_2^i$  методов  $x_2^{ij}, j=1, \dots, n_j$ . Каждый элемент  $x_2^{ij}, j \in \{1, \dots, n_j\}$  реализует решение  $x_1^i$  с точки зрения критерия быстродействия за определенное время в зависимости от значения признаков элемента  $x_1 \in X_1$ , влияющих непосредственно на быстродействие метода  $x_2^{ij}$ .

В общем случае на быстродействие метода влияют следующие характеристики реализации  $x_1$  математической модели задачи задач размещения: размерность пространства  $R^l$ , вид  $\Phi$ -функций [12, 13], формирующих область допустимых решений; число неравенств (линейных, нелинейных), формирующих область допустимых решений; характер поведения переменных (дискретный, непре-

рывный, дискретно-непрерывный); присутствие технологических ограничений; вид функции цели (линейность, нелинейность, характер нелинейности).

Таким образом, формируется некоторый вектор характеристик реализации  $x_1 \in X_1$  математической модели задачи, которые могут повлиять на быстродействие метода  $x_2^j \in X_2$ , реализующего  $x_1^i$ , то есть  $Ax_1 = (d_1, d_2, \dots, d_m)$ , где  $d_k = \psi_k(x_0)$ ,  $k = 1, \dots, m$ .

Для каждого метода  $x_2^j \in X_2$  объективно существует оценка его быстродействия  $v^j$ , которая в общем виде представляется некоторыми функциями

$$v^j = \Omega_j(Ax_1^i(x_0)). \quad (6)$$

Следовательно, для одной и той же реализации  $x_1^i$  можно по оценкам  $v^j$  сравнивать методы  $x_2^j, j = 1, \dots, n_j$  на множестве  $v^j$ . В общем виде, согласно (6),  $\alpha_2 = v^j$ .

Итак, решить задачу по критерию быстродействия  $\alpha_2$  — это значит определить

$$\begin{aligned} \alpha_2(x_0, x_p^*) &= \min_{i,j} \alpha_{22}(x_1^{*i}, x_2^{*ij}) = \\ &= \min_{i,j} v^{*ij} = \min_i \min_j \{v^{i1}, v^{i2}, \dots, v^{ik_i}\}. \end{aligned} \quad (7)$$

Аналогично предыдущим рассуждениям предлагается алгоритмический подход к решению задачи (7).

1. Для исходных данных  $x_0 \in X_0$  выделяется подмножество  $X_1^i$  реализаций  $x_1^i, i = 1, \dots, k$  математической модели задачи, построенных соответственно с учетом допустимых погрешностей исходных данных  $x_0 \in X_0$ .

2. Каждому элементу  $x_1^i \in X_1^i, i \in \{1, \dots, k\}$  ставится в соответствие подмножество  $x_2^j \subset X_2$  методов  $x_2^j, j = 1, \dots, n_j$ , которые могут осуществить решение реализации математической модели  $x_1^i, i \in \{1, \dots, k\}$ .

3. Элементы  $x_p^j, j = 1, \dots, n_j$  упорядочиваются по критерию быстродействия  $\alpha_2^j$  методов  $x_2^j, j = 1, \dots, n_j$  на подмножествах  $x_2^j \subset X_2, i = 1, \dots, k$ .

Учитывая замечания 5 и 6, предположим, что п. 1 и 2 реализованы. Выполнение п. 3 сводится к решению следующей задачи.

Требуется определить  $x_p^* \in X_p$ , для которого

$$\alpha_2^{p*} = \min\{\alpha_2^1, \dots, \alpha_2^k\}, l = \sum_{i=1}^k n_i, \quad (8)$$

где  $\alpha_2^{p*} = \alpha_{22}(x_0, Ax_1, v^*)$ .

Найдем решение задачи (8). Пусть  $k > 1, n_i > 1$ . Тогда необходимо сравнение элементов по быстродействию  $\alpha_2^i$ , которое предполагается осуществить следующим образом.

- Формирование подмножества  $[X_2]^i, i = 1, \dots, k$ .
- Определение подмножества  $\{v^j\}$  для каждого  $i \in \{1, \dots, k\}$ .
- Упорядочивание элементов множеств  $\{v^j\}$  по возрастанию согласно (6).
- Сравнение по быстродействию  $\alpha_2^j$  всех первых элементов множеств  $\{v^j\}, i = 1, \dots, n, j \in \{1, \dots, k_j\}$ .

### Многофакторные оценки

Решение задачи (2) при  $\alpha_1 \neq 0, \alpha_2 \neq 0$  предполагает использование методов многокритериальной оптимизации [14, 15]. Рассмотрим применение некоторых из них.

Используем понятие компромиссной (приемлемой) для всего множества функций цели многокритериальной задачи альтернативы [14]. Здесь под приемлемой понимается такая альтернатива  $x_p \in X_p$ , для которой величина отклонений от оптимальных значений по каждой функции цели достигает наименьшего значения. Поскольку наименьшее значение всех отклонений не достигается на одной альтернативе, то необходимо использование дополнительной информации.

Введем  $A = (A_1, A_2)$ , где  $A_1 = \alpha_1^*$ ,  $A_2 = \alpha_2^{p*}$  — решения задач оптимизации (5), (8). Можно рассматривать  $A$  как нижнюю оценку решения задачи оптимизации (2).

Рассмотрим следующую схему решения задачи оптимизации (2). Будем считать решением задачи такой элемент  $x_p^* \in X_p$ , для которого сумма квадратов отклонений  $\alpha_1(x_p^*)$  от  $A_1$  и  $\alpha_2(x_p^*)$  от  $A_2$  была бы минимальной. Другими словами, минимизируем критерий:

$$G(x_p) = (\alpha_1(x_p) - A_1)^2 + (\alpha_2(x_p) - A_2)^2 \quad (9)$$

на множестве допустимых решений  $X_p$ . Решению этой задачи будет соответствовать такая точка  $x_p \in X_p$ , которая доставляет минимальное значение  $\rho(f(x_p), A)$ , где  $f(x_p) = (\alpha_1(x_p), \alpha_2(x_p))$ , а  $\rho(x, y)$  — метрика в пространстве  $R^2$ .

Таким образом, решение двухкритериальной задачи  $x_p^* \in X_p$  является приемлемой альтернативой в смысле определения, введенного в [14], и обеспечивает минимальное расстояние по значению функции между допустимым решением и его нижней оценкой в задаче оптимизации (2).

Отметим, что в ряде случаев при построении моделей практических задач критерии  $\alpha_1, \alpha_2$  и их отклонения от  $A_1$  и  $A_2$  не равнозначны. Для учета значимости точности решения задачи и быстродействия в схему решения двухкритериальной задачи введем весовые коэффициенты. Тогда критерий (9) преобразуем к виду:

$$G_\delta(x_p) = \delta(\alpha_1(x_p) - A_1)^2 + (1 - \delta)(\alpha_2(x_p) - A_2)^2, \quad (10)$$

где значение  $\delta \in [0, 1]$  устанавливается экспертным путем из содержательных соображений.

Другой подход к упорядочению элементов множества  $X_p$  по двум критериям  $\alpha_1$  и  $\alpha_2$  может быть реализован путем построения многофакторных оценок полезности элементов  $x_p \in X_p$  [15]. Получение таких многофакторных оценок связано с предварительной нормализацией критериев на основе применения функций полезности час-

тных критериев. Для каждого критерия  $\alpha_i(x_p)$ ,  $i=1,2$  вычислим функцию полезности вида:

$$p_i(\alpha_i(x_p)) = (\Delta\alpha_i \cdot (\alpha_i(x_p) - \alpha_i^-))^{\lambda_i},$$

где  $\Delta\alpha_i = \frac{1}{\alpha_i^+ - \alpha_i^-}$ ;  $\alpha_i^-$ ;  $\alpha_i^+$  — соответственно наилучшее и наихудшее значение критерия  $\alpha_i(x_p)$ ,  $i=1,2$ ;  $\lambda_i$  — параметр, определяющий нелинейность зависимости локальной полезности от нормализованного значения критерия  $\alpha_i(x_p)$ .

В зависимости от наличия информации об относительной важности критериев рассмотрим применение следующих видов многофакторных оценок полезности альтернатив  $x_p \in X_p$ .

1. Когда известны весовые коэффициенты относительной важности критериев точности решения и быстродействия и их функций полезности применяется аддитивная оценка вида:

$$P(x_p) = \delta \cdot p_1(\alpha_1(x_p)) + (1 - \delta) \cdot p_2(\alpha_2(x_p)) \rightarrow \max_{x_p \in X_p},$$

$$x_p^* = \arg \max_{x_p \in X_p} P(x_p).$$

2. Для случая, когда информация об относительной важности критериев отсутствует, применяется следующая оценка:

$$P(x_p) = \max_{x_p \in X_p} \min_{i=1,2} p_i(\alpha_i(x_p)),$$

$$x_p^* = \arg \max_{x_p \in X_p} \min_{i=1,2} p_i(\alpha_i(x_p)).$$

Отметим, что рассмотренные подходы к решению многокритериальной задачи (2) могут быть легко обобщены на случай произвольного числа критериев  $\alpha_i(x_p)$ .

### Выводы

Предложенный в данной работе подход к формированию многофакторных оценок при выборе наилучшего решения оптимизационных задач размещения может быть широко использован в системах, основанных на знаниях о математических моделях и современных методах локальной и глобальной оптимизации, предназначенных для решения задач упаковки, компоновки и раскроя в различных предметных областях науки и техники.

**Список литературы.** 1. Dyckhoff H., Scheithauer G., Terno J. Cutting and packing// in: M. Dell'Amico, F. Maffioli and S. Martello Eds. Annotated Bibliographies in Combinatorial Optimization, John Wiley & Sons, Chichester.- 1997.- P. 393-412. 2. Bennell, J.A. and J.F. Oliveira, A tutorial in nesting problem: the geometry. European Journal of Operational Research, Invited review, 184, 397-415, 2008. 3. Burke E. A new bottom-left-fill heuristic algorithm for the two-dimensional irregular packing problem/ Burke E., Heller R., Kendall G., Whitwell G. //Operational research, 2004, Volume 53, N3. 4. Liu Huyao, He Yuanjun and Julia A Bennell, The irregular nesting problem – A new approach for nofit polygon calculation, Journal of the Operational Research Society, 58, 1235-1245. 5. Dagli C.H. Knowledge-based systems for cutting stock problems// Europ.Journ. of Oper. Res. – 1990. – Vol. 44. – P. 160 – 166. 6. Стоян Ю.Г., Романова *ТО ЕСТЬ* Концептуальные основы построения системы решения R-задач: Препр./ АН Украины. Ин-т проблем машиностроения; 366.- Харьков:1992. – 14 с. 7. Романова *ТО ЕСТЬ* Представление знаний в системе решения задач геометрического проектирования // Кибернетика. – 1991. – № 5. – С. 36-42. 8. Стоян Ю.Г., Яковлев С.В. Математические модели и оптимизационные методы геометрического проектирования. – Киев: Наук. думка, 1986.- 266с. 9. Yu. G. Stoyan, G. Yaskov. A mathematical model and a solution method for the problem of placing various-sized circles into a strip//European Journal of Operational Research, 2004, Vol. 156, pp. 590-600. 10. Стоян Ю.Г. Размещение кругов и невыпуклых многоугольников с поворотами в прямоугольнике минимальной длины/ Стоян Ю.Г., Злотник М.В.// Докл. НАН Украины – 2007. – N2. – С. 37-42. 11. Стоян Ю.Г. Оптимизация упаковки одинаковых кругов в многосвязную область/Стоян Ю.Г., Чугай А. // Доклады НАН Украины. – 2004. – N12. – С. 64-68. 12. Y. Stoyan, J. Terno, G. Scheithauer, N. Gil, T. Romanova Phi-function for 2D primary objects//Studia Informatica Universalis, 2002, Vol. 2, No 1, pp. 1-32. 13. Yuriy Stoyan, Guntram Scheithauer, Nikolay Gil, Tatiana Romanova. Phi-functions for complex 2D-objects//4OR Quarterly Journal of the Belgian, French and Italian Operations Research Societies. Vol. 2, No 1, 2004 – pp. 69 – 84. 14. Михалевич В.С., Волкович В.Л. Вычислительные методы исследования и проектирования сложных систем.–М.: Наука, 1982.–286 с. 15. Овезгельдыев А.О., Петров Э.Г., Петров К.Э. Синтез и идентификация моделей многофакторного оценивания и оптимизации. – Киев: Наук. думка, 2002. – 164 с.

Поступила в редколлегию 2.04.2008

UDC 519.859



## PACKING SPHERES INTO A PARALLELEPIPED BY MEANS OF THE DECREMENTAL NEIGHBOURHOOD METHOD

G. N. Yaskov

A.M. Pidgorny Institute for mechanical engineering problems of NASU, Kharkov, yaskov@ukr.net

The paper deals with the optimisation problem of packing various solid spheres into a parallelepiped with minimal height. Based on the mathematical model the peculiarities of the problem are given. Regarding these peculiarities a solution method is suggested. The method is a new modification of the decremental neighbourhood method. Numerical results are discussed.

PACKING, SPHERE, PARALLELEPIPED, OPTIMISATION, DECREMENTAL NEIGHBORHOOD METHOD

### Introduction

3D packing problems are of importance in different branches of human activity, e.g., in CAD, medicine (radiosurgery treatment planning), the filtration problem, granular medium modeling, powder metallurgy packing catalysts in chemistry, etc.

Packing solid spheres into a parallelepiped is of special value in research of 3D packing problems due to simplicity of description of interactions between spheres.

T.C. Hales [1] has proved the Kepler conjecture that the packing density of identical spheres cannot exceed value  $\pi/\sqrt{18} \approx 0.74$ .

The problem of unequal sphere packing in a 3D polytope is analyzed in [2] as a non-convex optimization problem with quadratic constraints and a linear objective function. Algorithms which improved the existing branch-and-bound algorithm for the general non-convex quadratic program are proposed.

G.E. Mueller [3] for packing identical spheres in a cylinder uses the sequential addition technique which is based on the dimensionless packing parameter. The identical sphere packing in high Euclidean dimensions by means of their random sequential addition is considered in [4]. The technique is the optimization algorithm by groups of variables [3], i.e., as a matter of fact, it is one of greedy algorithms [5].

Paper [6] takes in a wide range of circular and spherical packing problems. Original mathematical models for these problems are offered. Local optimal solutions are calculated starting from several random initial points. The best local solution obtained is supposed to be an approximation to a global solution. A huge number of numerical examples for the range of problems are given.

Paper [7] deals with packing of different radii spheres into a parallelepiped of minimal height. A modification of the decremental neighbourhood method (DNS) in combination with local optimization based on the reduced gradient method, the Newton method and the active set strategy have been used. DNS to search for an approximation to the global minimum is adopted to the problem under consideration.

This paper is a continuation of research of DNS and suggests a new selection criterion of neighbourhoods to obtain better objective function values.

### 1. Mathematical model

Let there be spheres  $S_i \subset R^3 \quad i \in I = \{1, 2, \dots, n\}$  with radii  $r_i, \quad i \in I = \{1, 2, \dots, n\}$ :

$$S = \{(x, y, z) \in R^3 : x^2 + y^2 + z^2 - r^2 \leq 0\}$$

and a rectangular parallelepiped (cuboid)

$$P = \{(x, y, z) \in R^3 : 0 \leq x \leq a, 0 \leq y \leq b, 0 \leq z \leq h\}$$

where  $R^3$  is the Euclidean 3D arithmetic space. Length  $a$  and width  $b$  of  $P$  are constant, while height  $h$  is a variable. The translation of sphere  $S_i$  by vector  $v_i = (x_i, y_i, z_i)$  is denoted by  $S_i(v_i), \quad i \in I$ . Translation vector  $v_i$  is also called the placement parameter vector of  $S_i$ .

**Problem.** Find a placement parameter vector

$$v = (v_1, v_2, \dots, v_n) \in R^{3n}$$

so that all spheres  $S_i(v_i)$  are contained completely within parallelepiped  $P$  without mutual intersections and height  $h$  of  $P$  attains a minimal value.

Thus,

$$u = (v, h) \in R^m$$

where  $m = 3n + 1$ , is a vector of all variables of the problem.

A mathematical model of the identical sphere packing problem is represented as follows.

$$F(u^*) = \min F(u), \text{ s.t. } u \in G \quad (1)$$

where

$$G = \{u \in R^m : \varphi_k(w_k) \geq 0, k = 1, 2, \dots, \theta = 6n + \frac{n^2 - n}{2}\} \quad (2)$$

$$\varphi_k(w_k) =$$

$$\begin{cases} x_i - r, \quad i = 1, 2, \dots, n & \text{if } k = 1, 2, \dots, n, \\ y_i - r, \quad i = 1, 2, \dots, n & \text{if } k = n + 1, n + 2, \dots, 2n, \\ z_i - r, \quad i = 1, 2, \dots, n & \text{if } k = 2n + 1, 2n + 2, \dots, 3n, \\ a - x_i - r, \quad i = 1, 2, \dots, n & \text{if } k = 3n + 1, 3n + 2, \dots, 4n, \\ b - y_i - r, \quad i = 1, 2, \dots, n & \text{if } k = 4n + 1, 4n + 2, \dots, 5n, \\ h - z_i - r, \quad i = 1, 2, \dots, n & \text{if } k = 5n + 1, 5n + 2, \dots, 6n, \\ (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2 - 4r^2, \quad i < j = 1, 2, \dots, n & \\ & \text{if } k = 6n + 1, 6n + 2, \dots, 6n + \frac{n(n-1)}{2}, \end{cases}$$

$$u = (v_1, v_2, \dots, v_n, h) \in R^m, \quad m = 3n + 1, \quad F(u) = h;$$

$$w_k \in \{x_i, y_i, z_i, (v_i, v_j), \quad i \neq j, \quad i, j \in \{1, 2, \dots, n\}\}.$$

Note that the inequalities (2) give the containment of the spheres within parallelepiped  $P$  and ensure non-overlapping of spheres.

Some characteristics of problem (1-2) can be pointed out.

(i)  $F(u)$  is linear, restrictions are linear and non-linear.

(ii) The frontier of  $G$  is formed by  $6n$  linear and  $\frac{n(n-1)}{2}$  surfaces of the second order.

(iii)  $G$  is in general disconnected, i. e.

$$G = \bigcup_{i=1}^n G_i, \quad G_i \cap G_j = \emptyset, \quad i \neq j,$$

This can occur if

$$(a-2r)^2 + (b-2r)^2 < 4r^2.$$

(iv) Each connected component of  $G$  is multiply connected.

(v) The problem is multi-extremal in general and  $NP$ -hard.

(vi) Any extreme point  $u \in G$  is specified by a system of  $m = 3n + 1$  equations of type  $\varphi_k(w_k) = 0$ .

(vii) To any local minimum  $u^*$  there always corresponds at least one extreme point  $\tilde{u}^* \in G$  such that  $F(\tilde{u}^*) = F(u^*)$ .

## 2. Solution method

In paper [7] to find an approximation of a global minimum a computational procedure was proposed which consists of the following steps:

(i) According to randomly generated sequences of the spheres a number of extreme points  $u^i \in G$ ,  $i = 1, 2, \dots, \lambda$  is computed.

(ii) Among these  $\lambda$  extreme points a best one, say  $\bar{u}^{-1}$ , is chosen which determines the centre point of a neighbourhood.

(iii) According to a modification of DNS a number of better solutions (extreme points)  $\bar{u}^j$ ,  $j = 1, 2, \dots, \tau$  are computed iteratively where  $\bar{u}^j$  denotes the best solution found within the current neighbourhood  $U_j$ . It is used as centre of the next neighbourhood.

(iv) Then, taking several of the best extreme points found in the last neighbourhood  $U_j$ , say  $\tilde{u}^j$ ,  $j = 1, 2, \dots, \kappa$ , as starting points, local minima  $u^{*j}$ ,  $j = 1, 2, \dots, \kappa$ , of the problem (1-2) are calculated.

(v) Finally, the point

$$u^* = \arg \min \{F(u^{*j}) : j = 1, 2, \dots, \kappa\}$$

is chosen as an approximation of a global minimum.

In this paper a new modification of DNS is proposed.

Let

$$\pi_j = \{S_{j_1}, S_{j_2}, \dots, S_{j_n}\}$$

be some random sequence of spheres. It means, the sequence may be considered as some permutation consisting of spheres  $S_j$ ,  $j = 1, 2, \dots, n$ . If all spheres are

different then the number of all sequences is equal to  $n!$ . Thus, all sequences of kind

$$\pi_j = \{S_{j_1}, S_{j_2}, \dots, S_{j_n}\}$$

form a discrete set  $\Pi$ .

The definition of a neighbourhood within the permutation set  $\Pi$  is realized by a one-to-one correspondence of a sequences

$$\pi_j = \{S_{j_1}, S_{j_2}, \dots, S_{j_n}\} \in \Pi$$

of spheres and a points

$$(r_{j_1}, r_{j_2}, \dots, r_{j_n}) \in \Pi_R$$

of the Euclidean space  $R^n$ , i. e. a one-to-one correspondence between sets  $\Pi$  and  $\Pi_R$  is established. By means of neighbourhood in  $\Pi_R$  a corresponding subset of sphere permutations in  $\Pi$  is determined which can be viewed also as a neighbourhood in  $\Pi$ , i.e. sequence

$$\rho_j = \{r_{j_1}, r_{j_2}, \dots, r_{j_n}\}$$

is considered as a point in the Euclidean space  $R^n$ .

Let

$$B = \left\{ (r_{j_1}, r_{j_2}, \dots, r_{j_n})^T \in R^n : (j_1, j_2, \dots, j_n) : \Pi(1, \dots, n) \right\}$$

denote the (finite) image of the set  $\Pi_R$  within  $R^n$  ( $\Pi(1, \dots, n)$  is a set of permutations of  $1, \dots, n$ ). As is known, the set  $B$  consists of vertices of a convex hyper-polytope inscribed into a  $(n-1)$ -dimensional hyper-sphere i.e. the set  $B$  is contained in an  $(n-1)$ -dimensional hyper-sphere. The diameter of this hyper-sphere is equal to

$$\beta = d(X^1, X^{n!}) = \sqrt{(r_1 - r_n)^2 + (r_2 - r_{n-1})^2 + \dots + (r_n - r_1)^2}$$

where

$$X^1 = (r_1, r_2, \dots, r_n), \quad X^{n!} = (r_n, r_{n-1}, \dots, r_1)$$

and  $r_1 \leq r_2 \leq \dots \leq r_n$ , i.e. the distance between any two points of  $B$  is equal to or smaller than  $\beta$ .

The approach proposed in this paper is desired to packing problems with various different radii. In case of only identical radii  $\beta$  becomes zero, and therefore the approach is not appropriate.

First, a set  $\Pi_1 \subset \Pi$  with  $\lambda$  elements ( $50 \leq \lambda \leq 100$  was defined empirically) is chosen randomly of set  $\Pi$ . For any  $\pi^q \in \Pi_1$  an extreme point of  $G$  is computed according to the permutation which corresponds to sequence  $\pi^q$ . Thus, one can obtain a set  $U_1 \subset G$  of extreme points  $u^{1\xi}$ ,  $\xi = 1, 2, \dots, \lambda$ , that are generated by set  $\Pi_1$ . After that, taking extreme points of set  $U_1$  as starting points, appropriate set  $U_1^*$  consisting of  $\lambda$  local minima is found. Then three points  $u^{*1\xi_1}, u^{*1\xi_2}, u^{*1\xi_3} \in U_1^*$  such that

$$F(u^{*1\xi_1}) < F(u^{*1\xi_2}) < F(u^{*1\xi_3}) \leq F(u^{*1\xi}),$$

$\xi = 1, 2, \dots, \lambda$ , are selected. To these local minima there correspond starting points  $u^{1\xi_1}, u^{1\xi_2}, u^{1\xi_3}$ . To points  $u^{1\xi_1}, u^{1\xi_2}, u^{1\xi_3}$  there correspond points  $\pi^{1\xi_1}, \pi^{1\xi_2}, \pi^{1\xi_3} \in \Pi_1$  in turn. Whence, to points  $u^{*1\xi_1}, u^{*1\xi_2}, u^{*1\xi_3} \in U_1^*$ , there correspond points  $\pi^{1\xi_1}, \pi^{1\xi_2}, \pi^{1\xi_3} \in \Pi_1$ , which are taken as centres of neighbourhoods:

$N_j(\pi^{\xi_j}) = \{ \pi \in \Pi : \rho(\pi^{\xi_j}, \pi) = \|\pi^{\xi_j} - \pi\| \leq \alpha\mu \}, j = 1, 2, 3,$   
 where  $0 < \mu < 1$  is suitable chosen (see below).

Then, samples

$$\Pi_{2j} \subset N_j(\pi^{2\xi_j}) \subset \Pi$$

with  $\lambda$  points are chosen in the same manner as above, and appropriate sets  $U_{1j} \subset G$  of extreme points  $u^{2j\xi}$  and sets  $U_{1j}^* \subset G$  consisting of  $\lambda$  local minima,  $j = 1, 2, 3,$   $\xi = 1, 2, \dots, \lambda,$  are constructed. From each sample of  $\Pi_{2j}$  points

$$\pi^{2j\xi_1}, \pi^{2j\xi_2}, \pi^{2j\xi_3} \in \Pi_{2j}, j = 1, 2, 3,$$

corresponding to local minima  $u^{*2\xi_1}, u^{*2\xi_2}, u^{*2\xi_3}$  such that

$$F(u^{*2j\xi_1}) < F(u^{*2j\xi_2}) < F(u^{*2j\xi_3}) \leq F(u^{*2j\xi}),$$

$j = 1, 2, 3, \xi = 1, 2, \dots, \lambda,$  are extracted.

Since a promising criterion is the probability of obtaining better objective function values, it is evident that the criterion depends both on mathematical expectation  $m_{2j}$  and dispersion  $\sigma_{2j}$  of a random sample of  $\Pi_{2j}$ . For example, for different neighbourhoods with equal mathematical expectations and with normal, logarithmically normal or Weibul-Gnidenko probability distributions, a better value of  $F$  will be obtained for a neighbourhood with a larger dispersion [8] and [9]. In this investigation, the criterion is a quantity  $\chi_{2j} = m_{2j} - \tau\sigma_{2j}$ , where parameter  $\tau = 1.5$  is defined empirically.

The next neighbourhood centres are chosen as follows:

1.  $\pi^{21}$  is a point that corresponds to starting point  $u^{21}$  from which the best local minimum  $u^{*2}$  among all local minima is obtained.
2.  $\pi^{22} \in \{\pi^{2j\xi_1}, \pi^{2j\xi_2}, \pi^{2j\xi_3}\}$  is a point that corresponds to the centre of the sample having  $\chi_2 = \min\{\chi_{21}, \chi_{22}, \chi_{23}\}$ .
3.  $\pi^{23}$  is a point that corresponds to a starting point from which the best local minimum among all local minima of the sample with factor  $\chi_2$  is obtained.

Points  $\pi^{2j}, j = 1, 2, 3,$  define the next promising neighbourhoods

$$N_j(\pi^{2j}) = \{ \pi \in \Pi : \rho(\pi^{2j}, \pi) \leq \alpha\mu^2 \}, j = 1, 2, 3,$$

and so on.

If a current local minimum  $u^*$  is obtained such that  $F(u^*) < F(u^{*t})$ , then radius  $\alpha\mu^t$  of neighbourhoods on the next iterations is multiplied by  $1/\mu$ .

Since due to a large number of inconsistent systems of low levels of the search tree when forming samples in neighbourhoods of little radii, the probability of repetitions of identical starting points increases dramatically, then the solution process is repeated until either the radius of neighbourhoods becomes less than  $\beta$  or the number of identical points in a neighbourhood becomes greater than 20. The best local minimum obtained is an approximation of the global minimum.

The value of parameter  $\mu$  influences the convergence of the process. A too small value of  $\mu$  yields a bad solution. Computational experience has shown that

$\mu \in [0.9, 0.95]$  is most suitable to take into account both the runtime and the quality of solutions.

If computational capacities do not allow to search for local minima for each starting point, the following modification of the DNS is offered. The solution process is the same as above but searching for the local minima is realized only for two extreme points  $u^{1\xi_j}, u^{2\xi_j} \in U_{\xi_j},$   $\xi = 1, 2, 3, j = 1, 2, \dots,$  such that

$$F(u^{1\xi_j}) < F(u^{2\xi_j})$$

and

$$F(u^{1\xi_j}) = \min F(u), \text{ s.t. } u \in U_{\xi_j};$$

$$F(u^{2\xi_j}) = \min F(u), \text{ s.t. } u \in U_{\xi_j} \setminus \{u^{1\xi_j}\}$$

for each new neighbourhood.

The best local minimum is taken as some approximation of the global minimum.

8 numerical instances are computed [7]. Using approximately the same amount of computer time in comparison to the proposed algorithm, the modification of DNS proposed in [7] is carried out. Percentage of improvement of the proposed algorithm against the modification from [7] was about 1%.

### Conclusions

A modification of DNS is proposed to solve three-dimensional problem of packing various sized spheres into a parallelepiped of minimal height.

Some details remain open, namely, refining parameters individually for each neighbourhood.

It is obvious that set  $U$  does not contain all extreme points of  $G$  in general, so that the approach cannot find a proved global minimum.

The calculation of proved local minima can also be applied in combination with other heuristics or meta-heuristic approaches.

**References:** 1. Hales, T.C. A Proof of the Kepler Conjecture // Ann. Math., 2005. — T. 162(1). — P.065-1185. 2. Sutou, A., Dai, Y. Global optimization approach to unequal sphere packing problems in 3D // Journal of Optimization Theory and Applications, 2002. — T. 114 (3). — P. 671–694. 3. Mueller, G.E. Numerically packing spheres in cylinders // Powder Technology, 2006. — T. 159. — P. 105-110. 4. Torquato, S. Random sequential addition of hard spheres in high Euclidean dimensions // Physical Review, 2006.— E74.— P.061-308. 5. Huang, W.Q., Li, Y., Akeb, H. and Li, C.M. Greedy algorithms for packing unequal circles into a rectangular container // Journal of the Operational Research Society, 2004.— P.1-10. 6. Birgin, E.G., Sobral, F.N.C. Minimizing the object dimensions in circle and sphere packing problems // Computers and Operations Research, 2007.— T. 35.— P. 2357-2375. 7. Stoyan, Y., Yaskov, G., Scheithauer, G. Packing of various radii solid spheres into a parallelepiped // Central European Journal of Operations Research, 2003. — T. 11(4).— P. 389-407. 8. Stoyan, Yu. G. Mathematical methods for geometric design. Proceedings of PROLAMAT 82, Leningrad, 1983.— P. 67-86. 9. Стоян Ю.Г., Соколовский В.З. Решение некоторых многоэкстремальных задач с помощью метода сужающихся окрестностей. — К. Наукова думка, 1980. — 365 с.

Надійшла до редколегії 2.04.2008

УДК 681.3.07



## ПРИМЕНЕНИЕ ОТНОШЕНИЙ НА МНОЖЕСТВЕ ХАРАКТЕРНЫХ ПРИЗНАКОВ ИЗОБРАЖЕНИЙ ПРИ РАСПОЗНАВАНИИ НА ОСНОВЕ ГОЛОСОВАНИЯ

В. А. Гороховатский

ХНУРЭ, г. Харьков, Украина

Приведены новые результаты в теории построения систем распознавания изображений на основе голосования множеств характерных признаков. Основная идея состоит в применении отношений на множестве признаков, что в целом улучшает качественные характеристики распознавания. Обсуждаются вопросы теоретико-множественного представления отношений в процедурах голосования, обоснования вероятностной модели принятия решений на основе функции соответствия признаков, анализа способов использования отношений, оценки эффективности подхода.

КОМПЬЮТЕРНОЕ ЗРЕНИЕ, ИЗОБРАЖЕНИЕ, ХАРАКТЕРНЫЕ ПРИЗНАКИ, РАСПОЗНАВАНИЕ, ГОЛОСОВАНИЕ, ОТНОШЕНИЯ, ФУНКЦИЯ СООТВЕТСТВИЯ КЛАССУ, ВЕРОЯТНОСТНЫЕ МОДЕЛИ

### Введение

Эффективным способом установления соответствия распознаваемых объектов с эталонами является метод голосования множеств характерных признаков (точек интереса), которыми представлены изображения [1-6]. Наряду с такими преимуществами как устойчивость и гибкость принятия решения при влиянии локальных помех и фона методы голосования по сравнению с другими подходами, например, с методами согласованной маркировки, обладают таким важным достоинством как простота построения, что при достаточно высокой надежности делает их конкурентоспособными среди известных подходов.

Использование различного типа отношений, построенных на множестве характерных признаков (ХП), дает возможность усовершенствования и улучшения качества работы систем голосования путем дополнения их информационных ресурсов новой высокоуровневой системой признаков.

Цель исследования — формализация и конкретизация аппарата отношений из теории множеств применительно к задаче распознавания визуальных объектов на основе голосования.

Задачи исследования — теоретико-множественное представление процедур голосования, построение отношений на множестве ХП, обоснование вероятностной модели принятия решений на основе функции соответствия признаков, анализ путей конкретного применения отношений, обоснование эффективности обсуждаемого подхода.

### 1. Формализация процедур голосования

Распознавание на основе множества ХП рассматривается как отображение заданных структурных компонентов модели на компоненты анализируемого изображения [2] и реализуется путем решения задачи оптимизации вида:

$$j^* = \arg \max_{j \in J} \psi(Z, Z(j)), \quad (1)$$

где  $J$  — множество классов ( $j^* \in J$ );  $Z, Z(j)$  — соответственно множества характерных признаков распознаваемого изображения и эталона  $j$ -го класса;  $\psi$  — мера сходства для множеств. Множество классов  $J$  можно считать конечным множеством целых чисел.

В теоретико-множественном представлении решение (1) путем голосования сводится к построению на множестве  $Z$  разбиения  $Z = \cup Z_j$  на непересекающиеся между собой подмножества  $Z_i \cap Z_j = \emptyset$ , состоящие из элементов, каждый из которых получает метку класса  $j$  путем реализации однозначного отображения  $\Lambda: Z \rightarrow J$ . Отображение  $\Lambda$  реализует локальное решение на множестве ХП. Поиск оптимума в (1) представляется как выбор из сформированных подмножеств  $Z_j$  элемента максимальной мощности, то есть

$$j^* = \arg \max_{j \in J} \# Z_j, \quad (2)$$

где символ  $\#$  означает мощность (количество элементов) множества. Разбиением  $Z = \cup Z_j$  на множестве признаков  $Z$  фактически устанавливается отношение эквивалентности. Признаки, отнесенные к одному и тому же  $Z_j$ , считаются одинаковыми, так как принадлежат классу  $j$ . При этом в представлении  $Z = \cup Z_j$  не обязательно присутствуют все номера эталонов из  $J$ , зато в реальных задачах состав разбиения часто дополняется множеством  $Z_0$ , что соответствует случаю неопределенности или отказу от локального решения в пользу одного из эталонов. Рисунок демонстрирует схему разбиения множества признаков  $Z$  при реализации однозначного отображения  $\Lambda$ .

Наряду с этим, при построении голосования возможно также применение множественного (неоднозначного) отображения  $\Lambda$ , когда каждый из признаков  $z \in Z$  формирует не один, а несколько элементов  $j_1, \dots, j_q; j_i \in J$  из множества классов  $J$ . В реальности это происходит по причине близости

значений признаков разных классов. Представление  $Z = \cup Z_j$  в такой постановке не подходит под классическое определение разбиения [7], так как элементы могут повторяться в разных  $Z_j$ . Множества  $Z_j$  в рамках описанной модели трансформируются так, что могут иметь общие элементы, то есть  $Z_i \cap Z_j \neq \emptyset$ .

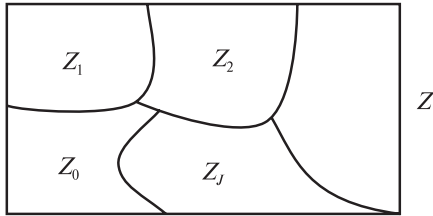


Рис. Схема разбиения множества ХП

Для того чтобы при реализации такого решения не усиливать влияния множественных признаков, их общий голос нужно представить в виде суммы  $\sum_{i=1}^q \alpha_i = 1$ , где  $\alpha_i$  — вес отдельного ХП, а  $q$  — количество повторных значений классов для конкретного признака. Для упрощения можно считать  $\alpha_i = 1/q$ . При такой нормировке число голосов остается равным мощности элементов множества  $Z$ . Правило (2) для этой модификации остается в силе, а мощность множества вычисляется по формуле

$$\# Z_j = \sum_i \beta_i, \beta_i = \begin{cases} 1, & \text{для одиночного голоса,} \\ \alpha_i, & \text{для множественного} \end{cases}, \quad (3)$$

и при необходимости округляется до ближайшего целого значения.

В основу функционирования систем голосования ХП могут быть положены также мажоритарные структуры [8], которые часто связаны с достаточно сложными схемами голосования. В зависимости от используемой схемы возможны парадоксальные ситуации, когда даже в условиях полного равенства элементов сколь угодно маленькая их доля оказывается решающей коалицией [8]. Не углубляясь в подробности построения сложных схем голосования, будем использовать для наших целей лишь наиболее простые из них. При прямом голосовании решение принимается большинством, превышающим половину голосующих элементов. Прямое, но не равное, голосование состоит в том, что элементы имеют разные веса, сумма которых равна единице:  $\sum_{z \in Z} \gamma(z) = 1$ . Такая нормировка требуется в целях принятия решения путем управления значением величины, характеризующей решающее число голосов и называемой коалицией. Решение принимается коалицией  $\tilde{Z} \subset Z$  в виде  $\sum_{z \in \tilde{Z}} \gamma(z) \geq \epsilon$ , где обычно считают  $0,5 < \epsilon \leq 1$  — порог голосования [8]. Коалиция для нашей задачи — это максимальное по мощности из множеств  $Z_j$ .

При снятии ограничения на правило «больше половины» значение  $\epsilon$  может быть выбрано из условия  $\epsilon < 0,5$ . При этом заранее предполагается возможное решение на основе относительно небольшого числа элементов [3]. Такое решение не вписывается в рамки традиционных статистических решений. Его надежность должна быть заранее оговорена и обоснована, так как речь идет о том, что в процессе голосования объект может быть представлен незначительной частью своих элементов.

Важным моментом, усиливающим эффективность распознающих систем на базе голосования, является использование комбинаций локальных признаков, основанных на построении отношений между ними [1,4-6]. Построение и использование отношений в целом дает возможность на основе имеющегося множества ХП получить устойчивые признаки более высокого уровня. Применение отношений между признаками естественным образом заложено в самой природе структурного представления как путь получения целостной информации о распознаваемом объекте. В конечном итоге любой визуальный объект можно формально описать путем построения отношений на множестве его частей [1].

С другой стороны, очевидно, что для визуального объекта в зависимости от условий его восприятия системой распознавания существует оптимальное в смысле степени глобальности представление, позволяющее осуществить его распознавание с достаточной надежностью. Это следует из опыта работы зрительного анализатора человека. Вся проблема состоит в том, чтобы успешно адаптироваться к этим условиям. Если помехи и фон вообще отсутствуют, то лучше применить методы интегрального типа, не предполагающие разбиения на компоненты. Структурный анализ в этой ситуации есть смысл применять только в случаях сильной схожести распознаваемых объектов. Основное предназначение структурного представления в виде множества характерных признаков — надежное распознавание в условиях частичных искажений объектов [3].

В теории системного анализа рассматривают два основных способа изучения структуры некоторого объекта [8]. Один из них состоит в том, чтобы «препарировать» его внутреннее содержание, установить состав и структуру частей, составляющих этот объект. Другой способ состоит в «проектировании» объекта на некоторую совокупность близких объектов. По свойствам проекций выносится суждение о внутреннем содержании изучаемого объекта. Считается, что для сложно организованного объекта второй способ более приемлем. При анализе и распознавании объектов в системах компьютерного зрения нашли применение оба способа [1-6].

## 2. Отношения на множестве ХП

Отношение (relation) — это абстрактное математическое понятие. Наибольшую известность в анализе изображений получили топологические отношения, которые учитывают взаимное расположение, а также метрические, оценивающие расстояние между элементами. Такие топологические отношения, как связность, смежность, инцидентность обладают устойчивостью к распространенным дефектам изображений, приводящим к нарушению метрических отношений. Примером метрических отношений есть расстояние, которое определено для любых двух элементов изображения. Расстояние в координатном понимании инвариантно относительно поворотов и переносов, но не относительно масштабирования.

Сопоставление с применением топологических и метрических отношений используют в целях расширения множества элементов, для которых установлено соответствие [1].

Пусть  $A = \{a^i\}$  — некоторое множество элементов, которыми могут быть, например, ХП. Совокупность упорядоченных наборов (кортежей) из  $k$  элементов множества  $A$  вида  $\langle a^1, a^2, \dots, a^k \rangle$ ,  $a^i \in A$  называется  $k$ -местным отношением  $R_A$  на множестве  $A$ . Иначе говоря, отношение  $R_A$  — это подмножество декартова произведения  $R_A \subseteq A \times A \times \dots \times A$ , в котором множество  $A$  участвует  $k$  раз. Величина  $k$  — это число «мест» или «арность» отношения. Более формально, отношение — это упорядоченная пара  $\langle R_A, A \rangle$ . Множество с заданным набором отношений получило в математике название «модель» [8]. Таким образом, кортеж  $\mathcal{A} = \langle A, \mathcal{R} \rangle$ , где  $\mathcal{R} = \{R_A^1, \dots, R_A^n\}$ , представляет модель с  $n$  отношениями, в терминах которой описывается распознаваемый визуальный объект.

Различают унарные, бинарные или  $k$ -арные ( $k > 2$ ) отношения в зависимости от количества участвующих в них элементов [7]. Унарные отношения характеризуют индивидуальные признаки элемента (ХП). Бинарные отношения наиболее популярны из-за простоты их анализа и построения. Кроме того, имеются простые и атрибутивные типы отношений. В атрибутивных отношениях кортежу из  $k$  элементов сопоставляется некоторый атрибут (число), представляющий, например, степень выполнимости отношения.

В задачах анализа изображений часто используют бинарные отношения, связанные с пространственной близостью элементов (в смысле координат), а также отношения близости в признаковом пространстве, фиксирующие эквивалентность значений признаков [4]. Примерами выражений, задающих указанные отношения, являются

$$\begin{aligned} |x_1 - x_2| + |y_1 - y_2| &\leq \delta, \\ \sum_{i=1}^n |z_i^1 - z_i^2| &\leq \varepsilon, \end{aligned}$$

где  $x, y$  — координаты;  $z_i^k$  — компоненты вектора признаков;  $\delta, \varepsilon$  — пороги.

Именно на отношении пространственной близости точек построено описание локальной окрестности, пиксели которой используются при формировании ХП. На отношениях близости значений в пространстве признаков так или иначе реализуются все подходы, связанные с принятием решений о типе признака или классе объекта. Другой популярный тип — отношения порядка, когда для пары элементов установлено отношение вида « $\langle \rangle$ » или « $\leq$ ».

Будем для определенности рассматривать бинарные отношения. Тот факт, что элементы множества  $a^i, a^k \in A$  находятся в отношении  $R_A$ , записывается в виде  $(a^i, a^k) \in R_A$  или  $a^i R_A a^k$ . Множество бинарных отношений имеет вид  $R_A \subseteq A \times A$ . Для унарных отношений  $a^i \in R_A$ , при этом  $R_A \subseteq A$ . Атрибутивное отношение записывается как  $d(a^i, a^k) \in R_A$ , где  $d$  — число или вектор, отображающий степень выполнимости отображения  $R_A$  на этой паре. Для простых отношений принято  $d = 1$ . Отношения называются симметричными, если из условия  $(a^i, a^k) \in R_A$  вытекает  $(a^k, a^i) \in R_A$ . Бинарные отношения могут быть заданы матрицей или графом.

На базе множества  $\mathcal{R} = \{R_A^1, \dots, R_A^n\}$  отношений элементов из  $A$  можно получить отношения более высокого уровня. Примером такой иерархии отношений может быть пирамидальное или вейвлетное представление. Теория отношений лежит в основе теории выбора и принятия решений [7, 8].

Очевидно, что максимальная мощность множества пар элементов, удовлетворяющих симметричному бинарному отношению  $R_A$ , определяется числом комбинаций  $m_R = C_m^2$ , где  $m$  — мощность множества  $A$ ,  $C_m^2$  — число комбинаций. В результате мощность множества бинарных отношений  $R_A$  можно оценить величиной  $m_R = m(m-1)/2$ , что в  $(m-1)/2$  раза больше числа элементов исходного множества  $A$ .

По причине роста количества признаков, построенных на множестве отношений, перед их применением необходимо каким-либо образом структурировать информацию, чтобы добиться сокращения числа анализируемых элементов. Кроме того, часто так выбирают содержимое отношений, чтобы уменьшить количество признаков в сформированной системе [1, 7].

Пусть  $V = \{v^i\}$  — множество значений ХП изображения,  $C = \{c^i\}$  — множество их координат, а  $Z = V \cup C$  — их объединение,  $z^i = \{(v^i, c^i)\}$ . Между элементами множеств  $V$  и  $C$  имеется однозначное соответствие вида  $C \rightarrow V$ . Будем рассматривать три типа отношений:  $R_V$  как отношения на множестве  $V$ ,  $R_C$  — отношения на множестве  $C$ , а также комбинированные отношения  $R_Z = R_{VC}$ ,

заданные на множестве пар элементов признак — координаты, то есть  $([v^i, c^i], [v^k, c^k]) \in R_{VC}$ .

Заметим, что отношения, как и признаки, могут обладать свойством инвариантности к геометрическим преобразованиям. Этот факт можно использовать в процессе применения отношений.

Пусть имеется некоторая база данных  $E = \{E_1, \dots, E_J\}$ , содержащая  $J$  моделей (эталонов) объектов. Модель объекта  $E_j$ , в свою очередь, включает множество значений  $Z(j)$  и множество  $R_{Z(j)}$  отношений признаков, то есть  $E_j = \{Z(j), R_{Z(j)}\}$ . При этом вид отношений  $R_Z$  будем считать известным и фиксированным для базы данных и для распознаваемого изображения. Унарные отношения находят свое отражение непосредственно в значениях признаков  $Z(j)$ . Если рассматривать  $R_{Z(j)}$  как множество бинарных отношений, то  $E_j$  будет представлять собой совокупность пар  $z_i, z_v$ , характеризующихся значениями конечного множества атрибутов  $d_1, \dots, d_\theta$ , где  $\theta$  — количество отношений, то есть  $E_j = \{z_i, z_v, d_1, \dots, d_\theta\}$ ,  $z_i, z_v \in Z(j)$ . Эталонная информация, на основе которой строится голосование, представима в виде следующей таблицы.

$z_i$	$z_v$	$d_1$	...	$d_\theta$
-------	-------	-------	-----	------------

Аналогичная таблица строится и для всех пар ХП из множества  $Z$ .

При внесении описанных изменений схема распознавания (1), (2) не изменится, меняются лишь вид и число элементов, участвующих в голосовании, а также мера для сопоставления этих элементов. Член голосования с учетом отношений представляет собой вектор  $r_{iv}[d] = (z_i, z_v, d_1, \dots, d_\theta)$ ,  $d_i \in R_Z$ .

Одним из вариантов голосования является представление, когда отношения между признаками задаются неявным образом, то есть без вычисления значений  $(d_1, \dots, d_\theta)$ . Примером такого задания отношений есть алгоритмический (процедурный) путь, когда признаки из множества  $Z$  при формировании новой системы признаков  $R_Z$  выбираются посредством некоторой процедуры.

Теоретико-множественная формулировка задачи распознавания с учетом отношений сводится к получению представления  $R_Z = \bigcup R_{Z(j)}$  на множестве отношений  $R_Z$  множества  $Z$  и решения задачи оптимизации (2) в виде

$$j^* = \arg \max_{j \in J} \# R_{Z(j)}. \quad (4)$$

В связи с появлением дополнительной информации в виде значений атрибутов или процедурных отношений расширяется разнообразие возможных схем принятия как локального  $\Lambda$ , так и глобального решений. При этом перспективно совместное применение правил (2) и (4), когда одновременно используются отношения и значения признаков.

Еще один вариант комбинированного построения состоит в поэтапном использовании признаков и их отношений. Например отношения строятся и проверяются лишь в случае, если принято одинаковое локальное решение относительно входящих в них признаков.

Теория отношений не ограничивается применением лишь к структурным элементам распознаваемых объектов. Учитывая, что отношения можно построить также для элементов разных множеств, саму процедуру голосования и подсчета голосов можно трактовать как реализацию некоторого бинарного отношения.

Формально бинарное отношение  $Q(j)$  для двух разных множеств  $Z$  и  $Z(j)$  опишем как тройку  $\langle Q(j), Z, Z(j) \rangle$ , где  $Q(j) \subseteq Z \times Z(j)$ . Во множество  $Q(j)$  могут быть включены, например, пары равных между собой в некотором смысле элементов множеств  $Z, Z(j)$ . Заметим, что во множества признаков  $Z, Z(j)$ , в свою очередь, могут включаться внутренние отношения элементов этих множеств. При такой формулировке решение (2) или (4) выглядит как поиск среди  $\{Q(j)\}$  множества максимальной мощности и определение в соответствии с этим значения класса. При этом трактовка (2) фактически основана на построении пересечения множеств  $Z, Z(j)$ , а (4) — на пересечении множеств отношений  $R_Z, R_{Z(j)}$ . При этом считается, что множества совпадают, если любой объект первого множества содержится во втором и наоборот, каждый элемент второго множества входит в первое.

### 3. Примеры построения отношений ХП

Рассмотрим вначале применение отношений на примере метода частных корреляций [2, 3]. В его классическом представлении локальное решение принимается, если сходство двух соответствующих фрагментов объекта и эталона будет выше некоторого порога  $\delta_1$ . Глобальное же решение принимается, если количество положительных локальных решений превышает порог  $n = \delta_2 m$ , где  $m$  — общее количество фрагментов, соответствующее максимально возможному числу локальных решений. Здесь имеем  $n$ -арное отношение на множестве значений локальных сходств, на выполнении которого основано построение глобального решения. Функционально это отношение фиксируется местоположением фрагментов (множеством номеров) в построенной системе. Словесно данное отношение можно сформулировать так: «величины сходства  $n$  любых фрагментов из общего количества  $m$  превышают порог  $\delta_1$ ».

Схема независимого голосования (без применения отношений) состоит в том, что каждый фрагмент принимает самостоятельное решение, а далее на их основе реализуется глобальное решение путем

поиска максимума среди подмножеств локальных решений для отдельных классов. Сравнивая схемы, можем рассчитывать на большую надежность схемы, основанной на отношениях по сравнению с независимым голосованием. Это предположение основано на использовании несколько большего объема информации об объектах, связанной с учетом структуры системы фрагментов.

В рассмотренной схеме метода частных корреляций на глобальном и локальном уровнях используется отношение типа  $R_V$ , хотя на локальном уровне неявно при вычислении сходства применено также отношение типа  $R_C$ , строго устанавливающее соответствие координат сравниваемых фрагментов в структуре объекта.

Другой пример использования отношений на множестве ХП описан в [5]. Здесь отношения строятся путем формирования новой системы высокоуровневых признаков, которые включают: значение (или голос) центрального для некоторой окрестности ХП; условие, что более половины из имеющегося множества ХП в данной окрестности имеют то же значение; условие совпадения углов между тройками согласованных ХП с соответствующими углами для объекта. Эти свойства названы «полулокальными условиями». Такие условия, по мнению авторов, существенно увеличивают селективную силу исходной системы голосования, что явно усиливает максимум в гистограмме голосов, подавляя локальные максимумы [6]. Здесь имеем комбинированный вариант  $R_{VC}$  отношений для ХП, где параллельно используются их координаты и значения. Описанный тип отношений можно характеризовать как «поддержку» локальных решений, когда окрестность ХП подтверждает или опровергает фиксированное локальное решение.

Следующим примером эффективного использования отношений может быть способ, основанный на сравнении координат групп признаков точек, представленных в двумерных аффинных базисах [1]. Этот метод получил название «геометрическое хеширование». Важной особенностью подхода есть инвариантность полученных характеристик локальных признаков (аффинных координат) при действии группы аффинных преобразований. Достоинства метода состоят в достаточно высоком быстродействии на этапе распознавания путем построения объемной хеш-таблицы признаков на этапе предварительной обработки. Здесь каждая признаковая точка на основе ее координат представляется в базисе трех других неколлинеарных между собой точек посредством аффинных координат — пары чисел  $(\vartheta_1, \vartheta_2)$ , по значениям которых создается хеш-таблица. Этим фактически устанавливается и затем сравнивается на множестве эталонов некоторое пространственное отношение каждой четверки особенных точек объекта.

Итоговое голосование осуществляется на основе значений величин  $(\vartheta_1, \vartheta_2)$ , вычисленных для точек объекта. Метод построен на использовании только координатной информации (отношение типа  $R_C$ ), однако его можно усилить и обобщить за счет применения значений признаков из  $R_V$ .

#### 4. Меры отношений и вероятностные модели

Рассмотрим теперь вероятностную модель системы распознавания, построенной на голосовании отношений для ХП. Пусть содержательный смысл отношения состоит в том, что два или более признака голосуют за один и тот же эталон. Это отношение типа  $R_V$ .

Введем в рассмотрение полную группу событий  $\Omega = \{\omega_0, \omega_1, \dots, \omega_J\}$ , где  $J$  — количество классов распознаваемого объекта,  $\omega_0$  соответствует классу неопределенности («не знаю»). Событие  $\omega_j(z)$  состоит в том, что признак  $z$  относится к классу  $j$ :  $\omega_j(z) = \{z \in j\}$ . Считаем события  $\omega_j$  несовместными для разных  $j$ , тогда полная группа событий  $\Omega = \sum_{j=0}^J \omega_j$ .

В качестве меры выполнимости события  $\omega_j$  будем использовать значение функции  $\xi(z, j)$  соответствия признака  $z$  классу  $j$ . Один из способов получения значения  $\xi(z, j)$  состоит в вычислении расстояния от точки  $z$  до множества  $Z(j)$  эталонных признаков класса  $j$ . Его можно определить через величину расстояния  $\rho$  в пространстве значений признака. Одним из вариантов есть вычисление расстояния  $\rho$  между признаком  $z$  и его ближайшим соседом  $z_j^{NN}$  в классе  $j$  [4]. Другой вариант основан на расстоянии до среднего значения признака класса, если величины средних для классов имеют значимые отличия [6].

Пусть по правилу ближайшего соседа  $\xi(z, j) = \rho(z, z_j^{NN})$ , где  $z_j^{NN} = \arg \min_{k=1, K(j)} \rho(z, z_j^k)$ ;  $K(j)$  — количество элементов во множестве  $Z(j)$ ;  $z_j^k \in Z(j)$ . Вид расстояния  $\rho$  определяется множеством значений и типом признака  $z$ . Для этого может быть использовано расстояние в векторном пространстве.

Вычисление функции соответствия, в свою очередь, устанавливает некоторое отношение между элементами множеств признаков  $Z$  и  $Z(j)$ . Атрибут этого отношения задается значением функции  $\xi(z, j)$ .

Учитывая, что значение меры соответствия естественным образом должно возрастать с уменьшением расстояния до эталона, в качестве  $\xi(z, j)$  целесообразно использовать функции вида:

$$\xi(z, j) = \exp\{-\lambda \rho(z, z_j^{NN})\}, \quad (5)$$

где экспоненциальное преобразование классического расстояния  $\rho$  кроме получения необходимых свойств позволяет нормировать величину  $\xi(z, j)$

в пределах отрезка  $[0,1]$ , что упрощает анализ значений  $\xi(z, j)$ . Величина  $\lambda$  отражает степень чувствительности построенной меры к изменению расстояния и выбирается в пределах фиксированного отрезка [4]. Появление события  $\omega_0$  будем характеризовать условием  $\xi(z, j) \leq \varepsilon_0$ , где  $\varepsilon_0$  — порог для величины незначительного сходства, которое не может быть засчитано ни одному из классов.

Вероятность  $p(z, j)$  того, что признак  $z$  относится к классу  $j$  (метка признака  $z$  равна  $j$ ), можно вычислить через соотношение значений мер, выражающих число случаев, благоприятных этому событию, и общего числа случаев в виде

$$p(z, j) = \frac{\xi(z, j)}{\sum_{i=0}^J \xi(z, i)}. \quad (6)$$

Заметим, что здесь выполняется условие

$$\sum_{j=0}^J p(z, j) = 1,$$

так как в результате признак будет обязательно отнесен к одному из классов, то есть выполнится одно из событий  $\omega_j$ .

Рассмотрим теперь событие, состоящее в том, что два признака  $z_t, z_v$  будут отнесены к классу  $j$ . Оно состоит в одновременном выполнении событий  $\omega_j(z_t)$  и  $\omega_j(z_v)$ , то есть в реализации бинарного отношения для признаков  $z_t, z_v$ . Вероятность того, что оба признака принадлежат классу  $j$ , определяется как произведение вероятностей двух независимых событий в виде:

$$p(z_t, j | z_v, j) = p(z_t, j) p(z_v, j) = \frac{\xi(z_t, j) \xi(z_v, j)}{\sum_{i=0}^J \xi(z_t, i) \sum_{i=0}^J \xi(z_v, i)}. \quad (7)$$

Вероятность (7) можно получить аналогично (6) как соотношение значений мер, соответствующих положительному исходу и общему числу вариантов для пар признаков. По аналогии с (7) вычисляются вероятности для трех и более признаков, реализующих  $k$ -арные отношения.

Для вероятностей вида (7) из-за отсутствия пересечения событий  $\omega_j$  выполняется условие  $\sum_{j=0}^J \sum_{i=0}^J p(z_t, j | z_v, i) = 1$ .

Анализируя соотношения (6) и (7), можно заметить, что с увеличением количества признаков, входящих в объединяющее их отношение  $R_Z$ , вероятность их отнесения к одному и тому же классу уменьшается, так как произведение двух вероятностей всегда меньше или равно максимуму среди них. Этот факт можно трактовать следующим образом. Имеем, что для произвольной пары признаков и их меры соответствия вероятность отнесения обоих признаков к одному и тому же классу не больше, чем для каждого из признаков в отдельности. Это означает, что надежность распознава-

ния, основанного на паре признаков, будет выше, чем для одного признака.

Заметим, что значения (6), (7) являются функциями от нескольких параметров: вида функции соответствия  $\xi$ , способа формирования и значений признаков эталонов, величин признаков распознаваемого объекта. Основываясь на этом факте, величины (6), (7) можно использовать для оценки качества системы распознавания, построенной на сопоставлении множеств локальных признаков. Для конкретного набора эталонов путем вычисления, например, значения (7), можно оценить вероятность «совпадения» между собой пар признаков внутри множества эталонов. Если эта вероятность велика, значит, необходимо менять либо эталоны, либо способ вычисления признаков, либо вид функции соответствия, либо увеличить количество признаков, формирующих отношение. Фактически таким способом определяются предельные возможности системы распознавания. На этой основе для конкретного множества (базы) эталонов можно подобрать необходимое для голосования количество и вид признаков, используя фиксированную функцию  $\xi$ .

На основе выражения (6) можно подсчитать вероятность того, что ровно  $m$  признаков из общего количества  $s$  будут отнесены к классу  $j$ . Это событие, которое обозначим как  $H_m$ , соответствует  $m$ -арному отношению на множестве признаков. Учитывая, что вероятности  $p_k = p(z, j)$ , кроме номера класса, зависят от значения признака  $z$  (эта зависимость отражена индексом  $k$ ), вычисление вероятности события  $H_m$  отличается от схемы Бернулли и определяется более сложным распределением. Если обозначить через  $(i[1], \dots, i[m])$  выборку объема  $m$  из последовательности чисел от 1 до  $s$ , то вероятность события  $H_m$  равна

$$P(H_m) = \sum_{(i[1], \dots, i[m])} p_{i[1]} \dots p_{i[m]} q_{r[1]} \dots q_{r[s-m]}, \quad (8)$$

где  $p_{i[v]}$  — значения  $p_k$  с соответствующими номерами;  $(r[1], \dots, r[s-m])$  — дополнение выборки  $(i[1], \dots, i[m])$  до  $s$  элементов;  $q_{i[v]} = 1 - p_{i[v]}$ . Тогда вероятность отнесения  $m$  и более признаков к классу  $j$  исходя из (8) равна

$$Q(p_k, m) = \sum_{t=m}^s P(H_t). \quad (9)$$

Если же предположить, что значение вероятности  $p_k$  одинаково для всех признаков из  $Z$ , то есть выполнено условие  $\forall k p_k = p$ , то вероятность  $Q$  в соответствии со схемой Бернулли равна

$$Q(p, m) = \sum_{t=m}^s C_s^t p^t q^{s-t}, \quad (10)$$

где  $q = 1 - p$ . Учитывая сложность вычисления  $Q(p_k, m)$  в виде (9), связанную с перебором всевозможных кортежей из  $m$  элементов признакового пространства, в качестве верхней оценки искомой

вероятности вместо (9) можно использовать значение  $Q(p, m)$  из (10), положив, например,  $p = \max_k p_k$ . Другой вариант оценки можно построить на усреднении значений  $p_k$ .

Вероятности для отношений признаков, полученные в соответствии с (6)-(10), могут быть положены в основу работы системы распознавания путем голосования [4]. На базе (6)-(10) вычисляется апостериорная вероятность отнесения множества ХП к одному из классов. Решение принимается по максимуму апостериорной вероятности.

### 5. Некоторые практические аспекты

В плане особенностей практической реализации предложенного подхода отметим следующее.

Исходные данные для системы распознавания можно представить в виде совокупности матриц эталонов и распознаваемого объекта

$$Z(j) = \begin{bmatrix} z_{11} & \dots & z_{1n} \\ \dots & \dots & \dots \\ z_{K[j]1} & \dots & z_{K[j]n} \end{bmatrix}, \quad j = \overline{1, J}, \quad Z = \begin{bmatrix} z_{11} & \dots & z_{1n} \\ \dots & \dots & \dots \\ z_{s1} & \dots & z_{sn} \end{bmatrix}.$$

Матрицы в общем случае имеют разное количество строк, а каждая строка состоит из фиксированного числа  $n$  элементов, соответствующих признаку  $z$ . Отношения устанавливаются между одинаковыми по числу подмножествами строк. Например, бинарные отношения предполагают установление соответствия для пар строк. Распознавание сводится к сопоставлению строк матриц или их отношений и установлению матрицы из  $\{Z(j)\}$ , наиболее близкой в смысле критерия распознавания к матрице  $Z$ . Степень соответствия отдельных строк выражается значением функции  $\xi(z, j)$ . Вероятностные характеристики распознавания могут быть получены из (6)-(10).

При реализации схемы распознавания для сходства комбинаций строк должны быть установлены пороги, равные половине количества голосующих строк или их комбинаций (максимальное правдоподобие), либо один общий порог для всех эталонов.

Учитывая, что сравнение осуществляется на основе полного перебора подмножеств строк, при сопоставлении необходимо использовать эффективные способы поиска, например, хеш-структуры.

### Выводы

Отношения, построенные на множестве ХП распознаваемого объекта и эталонов, являются мощным источником информации, позволяющим сформировать признаки более высокого уровня. Применение этой новой высокоуровневой системы признаков в целом позволяет повысить надежность функционирования и расширить возможности распознающих систем, основанных на голосовании.

Научная новизна проведенных исследований состоит в формализации процесса применения отношений при голосовании признаков и построении вероятностных моделей, позволяющих оценить качество распознавания.

Практическая значимость работы заключается в построении конкретных схем распознавания, обладающих повышенными характеристиками надежности и селективных свойств по сравнению со схемами независимого голосования признаков.

Дальнейшее развитие подхода может быть связано с выбором оптимального для конкретных условий распознавания количества характерных признаков, составляющих отношение.

**Список литературы:** 1. Шапиро Л., Стокман Дж. Компьютерное зрение: Пер. с англ. — М.: Бином, 2006. — 752 с. 2. Путятин Е.П., Гороховатский В.О., Матам О.О. Методы та алгоритми комп'ютерного зору: Навч. посібник. — Харків: ТОВ «Компанія СМІТ», 2006. — 236 с. 3. Путятин Е.П., Гороховатский В.А., Кузьмин С.В. Распознавание изображений в пространстве инвариантных локальных признаков // Радиоэлектроника и информатика. — 2006, №1(32). — С. 69–73. 4. Kim S., Kweon I.-S. Biologically motivated perceptual feature: Generalized robust invariant feature. — In Asian Conference of Computer Vision (ACCV-06), 2006, P. 305-314. 5. Schmid C., Mohr R. Local Grayvalue Invariants for Image Retrieval. — IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997 — 19(5): 530-534 p. 6. Фортсайт Д., Понс Ж. Компьютерное зрение. Современный подход: Пер. с англ. — М.: Изд. дом «Вильямс», 2004. — 928 с. 7. Нечеткие множества и теория возможностей. Последние достижения / Под ред. Р. Ягера. — М.: Радио и связь, 1986. — 408 с. 8. Шрейдер Ю.А., Шаров А.А. Системы и модели. — М.: Радио и связь, 1982. — 152 с.

Поступила в редколлегию 4.04.2008

УДК 519.6



## ВЫДЕЛЕНИЕ ИНФОРМАТИВНЫХ ПРИЗНАКОВ В ЗАДАЧАХ РАСПОЗНАВАНИЯ ОБРАЗОВ

В.И. Бритик<sup>1</sup>, Е.А. Егорова<sup>2</sup>

<sup>1</sup>ХНУРЭ, г. Харьков, Украина, valbri@inbox.ru

<sup>2</sup>ХНУРЭ, г. Харьков, Украина, evg.egorova@gmail.com

В статье представлены результаты исследований проблемы выделения информативных признаков в задачах распознавания образов, которые позволили выделить актуальность дальнейших исследований относительно возможности повышения эффективности выделения информативных признаков в задачах кластеризации посредством использования критериев качества кластеризации.

**ЗАДАЧИ РАСПОЗНАВАНИЯ ОБРАЗОВ, ПОДЗАДАЧА ВЫДЕЛЕНИЯ ПРИЗНАКОВ РАСПОЗНАВАНИЯ, ИНФОРМАТИВНЫЕ ПРИЗНАКИ, МЕТОДЫ ВЫДЕЛЕНИЯ ИНФОРМАТИВНЫХ ПРИЗНАКОВ**

### Введение

Выделение информативных признаков (признаков распознавания) исследуемых объектов рассматривается многими исследователями как один из самых важных и сложных этапов решения задач распознавания [1, 2, 3, 4].

Философские и общеметодологические концепции, составляющие теоретический базис области распознавания образов, определяют и подтверждают значимость данной подзадачи в решении задач распознавания образов.

Сложность подзадачи выбора признаков распознавания обуславливается ее переборным характером. Достаточно большая размерность данных, исследуемых в задачах распознавания, и проблема комбинаторного взрыва<sup>1</sup> ограничивают возможности применения точных математических методов, реализующих общий подход к решению данной подзадачи – полный перебор [1].

Данная ситуация привела к появлению многочисленных математических методов выделения признаков, направленных на сокращение полного перебора. Однако для исследователей в области распознавания образов выделение признаков распознавания в многомерных задачах распознавания образов остается актуальной проблемой.

Целью данной работы является исследование проблемы выделения признаков распознавания в задачах распознавания образов.

Достижение поставленной цели предполагает решение следующих задач:

- исследовать типы и основные этапы решения задач распознавания образов;
- отследить взаимосвязь подзадачи выделения признаков с остальными основными этапами решения задач распознавания;
- провести анализ существующих методов отбора признаков распознавания.

<sup>1</sup> Уже при наличии порядка 20 потенциальных признаков прямой перебор всех комбинаций становится невозможным, поскольку приводит к комбинаторному взрыву.

### 1. Типы задач распознавания и основные этапы их решения

В зависимости от условий, в которых происходит процесс распознавания, и преследуемой цели в теории распознавания образов выделяются следующие типы задач распознавания<sup>2</sup>.

1) Задача распознавания – отнесение предъявленного (нового) объекта по его описанию к одному из заданных классов.

2) Задача автоматической классификации (задача таксономии, кластеризации) – разбиение множества объектов по их описаниям на систему непересекающихся классов.

3) Динамическое распознавание и динамическая классификация – задачи 1 и 2 в постановке исследования объектов в динамике.

4) Задача прогнозирования – это задачи 3, в которых решение должно относиться к некоторому моменту в будущем [5].

Несмотря на существенные различия, которые прослеживаются в постановках приведенных выше задач распознавания, общие этапы их решения представляются в виде следующих подзадач [3]:

– первая подзадача связана с измерением характеристик, свойств объектов, подлежащих распознаванию, и представлением полученных в результате измерения числовых значений этих характеристик в некотором виде. Данная подзадача определяется как задача измерения и представления исходных данных;

– вторая подзадача связана с выделением характерных свойств (признаков распознавания) из полученных исходных данных и снижением размерности векторов измерений. Эту подзадачу часто определяют как задачу предварительной обработки и выбора признаков;

– третья подзадача состоит в отыскании решающих процедур, правил, необходимых для распознавания объектов.

<sup>2</sup> Следует отметить, что приведенные ниже задачи имеют также вероятностные постановки.

## 2. Задача представления и измерения исходных данных

Решение любой задачи распознавания одного из вышеприведенных типов начинается с анализа предметной области, которая заключается в формировании выборки объектов  $X = \{x_1, x_2, \dots, x_L\}$ , представителей этой предметной области, а также в исследовании и измерении их характеристик.

Любой объект материального мира  $x$  может быть описан с помощью бесконечного количества каких-либо характеристик или свойств  $G = \{g_1, g_2, \dots, g_s, \dots\}$ , то есть  $(g_1(x), g_2(x), \dots, g_s(x), \dots)$ . В идеале, на этапе анализа выборки объектов исследуемой предметной области набор этих характеристик должен содержать всю информацию об объектах, которая поддается измерению. На практике это условие невыполнимо и в процессе исследований приходится ограничиваться определенным объемом информации, обосновывая достаточность некоторого набора характеристик  $G' = \{g_1, g_2, \dots, g_s\}$ ,  $G' \subset G$ , для решения поставленной задачи распознавания.

Далее решается задача измерения характеристик или свойств из сформированного набора  $G' = \{g_1, g_2, \dots, g_s\}$ , которая предполагает выбор соответствующих шкал измерения.

Существует пять основных типов шкал измерений: номинальная (наименований), порядковая, интервальная, относительная и шкала разностей. Измерение количественных характеристик, как правило, производится с помощью шкал интервалов, отношений или шкалы разностей [5]. Качественные характеристики объектов описываются либо номинальной шкалой, либо подвергаются предварительным преобразованиям и ранжируются порядковой или бальной шкалами.

Нередко тип шкал определяют сами значения свойств и характеристик исследуемых объектов: например, такие свойства физического происхождения как рост, вес, температура, значения интенсивности пикселей на полутоновых изображениях и прочие. Но подзадача выбора типа шкал перестает быть тривиальной в случаях, когда приходится работать с признаками, получаемыми искусственным путем, например, в процессе синтеза методами факторного или корреляционного анализов [5]. В таких случаях задача выбора шкал требует дополнительных исследований, поскольку любые преобразования, в том числе переход от одной шкалы к другой, вносят погрешности в измеряемые значения свойств объектов.

Следует отметить, что задача измерения не ограничивается выбором типа шкал и самого процесса измерения. Характерным свойством шкал измерения является такая величина как различимость. Различимость шкалы  $\rho$  — это нормированное расстояние между градациями шкалы, рассчитываемое по формуле:

$$\rho = \frac{1}{n-1}, \quad (1)$$

где  $n$  — количество градаций.

Свойство различимости не зависит от типа шкалы и, как следует из формулы (1), с ростом числа градаций на шкале  $\rho$  стремится к нулю. Условно множество значений различимости можно разделить на два класса. Шкалы в первом классе, различимость внутри которого достаточно велика, удобны для человека, поскольку измеряемые ими значения существенно отличаются друг от друга. Вторым классом характеризуется малыми значениями различимости. Шкалами этого класса человек в повседневной жизни практически не пользуется. Однако к ним приходится обращаться в исследованиях естественных или искусственных микрообъектов, и выбор значения  $\rho$  может значительно сказаться на результатах исследований, особенно если это значение превышает порог различимости человека [6]. Таким образом, учет свойства различимости шкал на этапе анализа объектов задачи распознавания является важным моментом: «закладываются» надежность и (или) разрешающая способность разрабатываемой системы распознавания.

Завершающей задачей анализа предметной области является формализация представления исходных данных. Как правило, исходные данные представляются посредством математических моделей, аналогичных моделям описания объектов (образов), которые используются системами распознавания в процессе их функционирования: вектор измерений (или вектор образа), матрица «объект — свойство», древовидные структуры или графы отношений и прочее. [3, 7, 8].

## 3. Задача предварительной обработки и выбор признаков

Дальнейшие исследования заключаются в выборе признаков распознавания из множества характеристик и упрощения (окончательного формирования) модели описания образов.

Согласно [1, 2, 3, 4] «признаки распознавания» представляют собой характеристики свойств объектов распознавания, которые отражают существенные особенности этих объектов и позволяют устанавливать их сходство или различия с другими объектами.

Таким образом, признаками распознавания является некоторое подмножество  $F = \{f_1, f_2, \dots, f_N\}$  множества  $G' = \{g_1, g_2, \dots, g_S\}$  ( $N \leq S$ ) характеристик или свойств объектов, подлежащих распознаванию. Количественный и качественный состав подмножества  $F$ , называемый множеством признаков распознавания, обуславливается поставленной задачей распознавания. Это объясняется тем, что именно условия и преследуемая цель ре-

шаемой задачи «задают» существенность характеристикам объектов распознавания.

Учитывая факты бесконечного количества характеристик объектов и зависимости множества признаков от постановки задачи распознавания, а также разнообразие самих задач распознавания, можно утверждать о неограниченном числе потенциальных признаков распознавания в рамках общей постановки задачи распознавания.

Исходя из приведенных выше определения «признаков распознавания» и математических моделей описания образов, признаки распознавания являются средством описания объектов исследования, получившим в теории распознавания образов название «признакового описания». Специфика описания объектов признаками распознавания и их информационные особенности прослеживаются в их классификации по следующим критериям:

- по способу описания объектов признаки делятся на: количественные, качественные и логические;
- по характеру доставляемой информации выделяют: вероятностные, детерминированные и квазидетерминированные признаки;
- по источнику доставляемой информации выделяют: физические, математические и структурные признаки;
- по выполняемым функциям признаки делятся на: объединительные и разделительные.

Следует отметить, что одним из важных требований к признакам распознавания является их независимость от условий распознавания. Данное требование выражается в необходимости доказательства инвариантности значений используемых признаков к условиям распознавания.

Универсального формализованного метода формирования набора признаков распознавания  $F$  не существует.

Один из подходов к выбору признаков распознавания заключается в использовании знаний экспертов исследуемой предметной области. Данный подход предусматривает применение таких методов извлечения экспертных знаний как интервью, диалог, круглый стол и прочие. [9]. Но экспертный подход минимизирует или исключает возможность автоматизации процесса выбора признаков. Кроме того, выбор признаков экспертным путем поднимает проблему различимости шкал, описанную выше. Решение данной проблемы заключается в определении пороговых величин градаций шкал признаков, превышение которых свидетельствует о целесообразности отказа от мнения экспертов.

Второй подход формирования набора признаков предполагает применение различных математических методов и подходов: корреляционный; кластерный и факторный анализ; информационный подход; подход к выделению признаков; основанный на введении функционалов качества распознавания и прочие.

Нередко применение математических методов требует проведения предварительной обработки исходных данных – их нормализации, поскольку характеристики объектов имеют различные измерения и не сравнимы между собой.

Наиболее распространенные способы нормирования следующие [7, 10]:

$$z^1 = \frac{x - \bar{x}}{\sigma}, z^2 = \frac{x}{x'}, z^3 = \frac{x}{x_{\max}},$$

$$z^4 = \frac{x}{x_{\max}}, z^5 = \frac{x - \bar{x}}{x_{\max} - x_{\min}},$$

где  $x$  – исходное значение;  $z$  – нормированное значение;  $\bar{x}, \sigma$  – соответственно среднее и среднее квадратичное отклонение  $x$ ;  $x'$  – некоторое эталонное (нормативное) значение  $x$ ;  $x_{\max}, x_{\min}$  – наибольшее и наименьшее значения  $x$ .  $z^1$  и  $z^5$  определены для шкал отношений и интервалов, остальные способы – только для шкал отношений. Существуют и другие способы нормирования, производные от вышеперечисленных.

Предварительная обработка, приводя исходные данные к виду, удобному для распознавания, способствует уменьшению информативности этих данных, и важно, чтобы это уменьшение не сказывалось на качестве обрабатываемой информации.

При выборе признаков распознавания лучшим вариантом считается совместное использование методов извлечения экспертных знаний и математических методов обработки многомерных данных. Таким образом, формирование набора признаков распознавания, который учитывает трудности, связанные с реализацией процессов выделения (синтеза) или выбора признаков, и обеспечивает в то же время необходимое качество распознавания, представляет собой одну из наиболее трудных подзадач при построении систем распознавания.

#### 4. Задача поиска решающих правил

Завершающим этапом решения задач распознавания является поиск решающих процедур или правил для распознавания объектов (образов). Универсального подхода нахождения вышеуказанных правил распознавания не существует. Выделяются следующие концепции их поиска: принцип сравнения с эталоном, принцип общности свойств, принцип кластеризации [3].

Использование концепции сравнения с эталоном требует хранения всех объектов, входящих в класс. Кроме этого, процедура сопоставления с эталоном не допускает существенных вариаций характеристик отдельных образов.

Основное допущение при использовании принципа общности свойств заключается в том, что образы, принадлежащие одному и тому же классу, обладают рядом общих свойств или призна-



Применение корреляционного анализа к исходным данным позволяет определить взаимосвязь характеристик объектов и выявить дублирующую информацию, использование которой способствует только усложнению алгоритмической схемы распознавания.

Установить наличие связей внутри совокупности характеристик исследуемых объектов и попытаться внести в нее структуру также с целью дальнейшего снижения размерности данной совокупности можно, применив методы кластерного анализа в пространстве измерений. Наибольшее распространение получили следующие методы кластерного анализа: метод ближайшего соседа, полносвязывающий метод, метод Кинга, метод  $k$  средних, ISODATA и пр. [8, 10].

Данные подходы отбора признаков применяются в детерминистских задачах распознавания образов как при известной, так и неизвестной правильной классификации исследуемого множества объектов. Следует также отметить, что кластеризацию признаков рекомендуется производить в качестве предварительной обработки в других методах выделения признаков, ссылаясь на следующие недостатки этих подходов [1]:

1) Информации о попарном сходстве между признаками в общем случае недостаточно для выделения оптимального набора признаков. В частности, набор попарно некоррелированных признаков может оказаться линейно зависимым.

2) Могут существовать кластеры, полностью состоящие из неинформативных признаков. В таких случаях, выбрав по одному из представителей от каждого кластера, все равно придется решать задачу отбора признаков.

Для выделения признаков в детерминистских задачах распознавания образов с известной правильной классификацией можно использовать подход, основанный на введении различных критериев качества распознавания, например функционал ошибки обобщения, критерий скользящего контроля, критерии непротиворечивости и регуляризации и прочие. С произвольными функционалами качества распознавания совместимы следующие методы отбора признаков: полный перебор; последовательное добавление признаков; поочередное добавление и удаление признаков; поиск в глубину (метод ветвей и границ); поиск в ширину (МГУА); генетические алгоритмы; случайный поиск с адаптацией [1].

Метод полного перебора характеризуется достаточно простой вычислительной схемой. Алгоритм, реализующий данный метод, осуществляет полный перебор наборов признаков в порядке возрастания сложности этих наборов (числа признаков в наборе). Для каждого значения сложности рассматриваются все возможные наборы. На втором

шаге из них выбирается лучший набор. На шаге 3 запоминается значение сложности, при котором был получен самый лучший набор. Если результат не улучшается на протяжении  $d$  итераций, считается, что минимум внешнего критерия пройден, и алгоритм заканчивает работу, выдавая наилучший набор. Таким образом, сокращение полного перебора происходит за счет того, что наборы с высокой сложностью вообще не рассматриваются.

Стратегию «жадного» наискорейшего спуска реализует метод последовательного добавления признаков, добавляя к набору по одному признаку, который приводит к наибольшему уменьшению внешнего критерия. Данный метод сокращает трудоемкость перебора с  $O(2^n)$  операций до  $O(n^2)$ , но включает «лишние» признаки, поскольку однажды включенный признак остается в наборе до конца работы алгоритма.

В тех случаях, когда заранее известно, что информативных признаков намного больше шумовых, целесообразнее использовать метод последовательного удаления признаков. Работа алгоритма данного метода начинается с набора  $F_0 = G'$ , где множество выборочных характеристик и/или свойств объектов, подлежащих распознаванию, и далее из набора последовательно исключается по одному признаку так, чтобы значение критерия убывало как можно быстрее. Метод последовательного удаления признаков работает медленнее метода последовательного добавления, поскольку на начальных итерациях приходится обрабатывать почти все признаки.

Стратегии двух жадных алгоритмов, описанных выше, совмещает в себе метод поочередного добавления и удаления признаков. Процессы последовательных добавлений и удалений чередуются до тех пор, пока значение критерия качества распознавания в точках минимума функционала не перестанет уменьшаться, или пока состав признаков в оптимальном наборе не стабилизируется. Данный метод характеризуется достаточно большой трудоемкостью и сложностью реализации.

*Метод ветвей и границ* позволяет перейти к сокращенному перебору за счет следующих эвристик:

1) Оценивать перспективность ветви дерева признаков и отказаться от ее наращивания, если уже имеется лучшая ветвь. Набор не наращивается, если значение критерия качества распознавания оказывается хуже, чем на самом лучшем из уже проверенных наборов меньшей сложности.

2) Желательно как можно раньше построить наиболее удачную ветвь дерева признаков. Тогда значения критерия на самых лучших наборах сложности  $j$ ,  $j = 1, S$  будут близки к наилучшему значению критерия для набора сложности  $j$ , и первая эвристика будет отсекал большинство ветвей в самом начале алгоритма.

Качество решения задачи выбора признаков данным методом во многом определяется выбором значений его параметров и, следовательно, требует проведения дополнительных экспериментов.

Достаточно широкое применение для формирования набора признаков распознавания получили метод группового учета аргументов и генетические алгоритмы.

На каждой итерации алгоритма, реализующего метод группового учета аргументов (МГУА) [1, 2], строится не один набор, а множество из  $B_j$  наборов, называемое  $j$ -м рядом:  $R_j = \{F_j^1, \dots, F_j^{B_j}\}$ ,  $F_j^b \subseteq G'$ ,  $|F_j^b| = j$ ,  $b = \overline{1, B_j}$ . Для перехода от текущего ряда  $R_j$  к следующему  $R_{j+1}$  от каждого набора  $F \in R_j$  порождается  $n-j$  новых наборов путем присоединения одного из признаков, не принадлежащих набору  $F$ . Из порожденных  $B_j(n-j)$  наборов в следующий ряд отбирается не более  $B$  наборов, лучших по внешнему критерию. Таким образом, на каждой итерации сложность всех наборов увеличивается на единицу. Число  $B$  является параметром алгоритма и называется шириной поиска.

Генетический алгоритм (ГА) [1, 11] осуществляет поиск наилучшего набора признаков по принципам дарвиновской эволюции. Первое поколение наборов генерируется случайным образом. К этим наборам применяются операции скрещивания и мутации для порождения большого числа новых наборов. Затем производится искусственный отбор или селекция: во второе поколение отбираются только  $B$  наборов, лучших по заданному внешнему критерию  $Q$ . Ко второму поколению также применяются операции скрещивания, мутации и селекции, и порождается третье поколение. Эволюционный процесс переходит от поколения к поколению до тех пор, пока не наступит стагнация, то есть качество лучшего набора в поколении перестанет улучшаться. Анализ показывает, что генетический алгоритм от МГУА отличается, главным образом, правилом порождения следующей популяции, остальные различия – в терминологии.

Применение генетического алгоритма позволяет вводить различные эвристики. Можно по-разному вводить операции скрещивания и мутации, применять совокупности критериев качества, разрешать скрещиваться только лучшим индивидуумам, переносить или не переносить лучших индивидуумов в следующее поколение, увеличивать вероятность мутаций при наступлении стагнации, параллельно выращивать изолированные популяции и так далее.

Недостатком генетического алгоритма является его относительно медленная сходимость. Кроме этого, наряду с параметрами размера популяции  $B$ , максимального числа поколений и вероятнос-

ти мутации, ГА имеет и другие параметры, подбор которых является искусством и зависит от особенностей задачи.

Если в генетическом алгоритме отказаться от операции скрещивания, получается метод случайного поиска. В  $t$ -м поколении популяции из  $B$  наборов модифицируется каждый из них  $R$  раз случайным образом. Из полученных  $BR$  наборов отбираются  $B$  лучших по заданному внешнему критерию  $Q$ , и из них формируется  $(t+1)$ -е поколение. Недостатком данного метода (аналогично ГА) является довольно медленная сходимость.

На ускорение сходимости направлен алгоритм случайного поиска с адаптацией [1]. Идея адаптации заключается в том, чтобы, генерируя наборы признаков случайным образом, увеличивать вероятность включения в них тех признаков, которые чаще входят в наилучшие наборы, а для признаков, входящих в наихудшие наборы, уменьшать вероятность их появления.

Трудоёмкость алгоритма СПА составляет  $O(Rr(j^* + d))$  операций. К достоинствам данного метода относят простоту реализации и существенно меньшее, по сравнению с ГА, число параметров. СПА сходится гораздо быстрее, чем случайный поиск без адаптации, и во многих экспериментах находит решения лучшего качества, чем алгоритм поочередного добавления и удаления признаков.

Приведенные выше методы являются эвристиками и не могут гарантировать оптимальность получаемого результата, поэтому можно говорить лишь об удовлетворяющем качестве сформированного набора признаков. Недостатком данного подхода является значительная трудоёмкость его реализации.

Альтернативным средством выделения признаков в детерминистских задачах распознавания образов с известной классификацией является информационный подход, в рамках которого можно рассчитать количество информации в событии класса по отношению к каждому из событий признака или оценить информативность каждого признака отдельно с помощью мер Хартли, Шеннона, Макинтоша и прочих [3, 2]. Формирование набора признаков распознавания в рамках информационного подхода производится на основании принципов учета признаков, которые дают наибольшую информацию об исследуемых объектах, и пренебрежении признаками, обладающими малой информативностью. Информационный подход используется также для выделения информативных признаков в детерминистских задачах распознавания при неизвестной правильной классификации и в вероятностных задачах распознавания.

В рамках вероятностной задачи распознавания, когда известны существующие вероятностные связи между признаками объектов и классами, к ко-

торым они принадлежат, выбор информативных признаков распознавания можно также произвести, используя подходы минимизации энтропии или максимизации дивергенции. В [3] отмечается, что кроме вычислительных трудностей возможность применения на практике этих подходов ограничивает также то обстоятельство, что количество реальных задач распознавания, имеющих полную или частичную вероятностную информацию: плотности распределения, характеризующие совокупности объектов и/или условные плотности распределения значений признаков – невелика.

Известно, что в процессе решения конкретной задачи распознавания в рамках структурного и структурно-лингвистического подходов выбор признаков объектов обуславливается результатами многочисленных экспериментов, опытом и интуицией исследователя, спецификой данных, поставленными условиями и ограничениями решаемой задачи, а также общими требованиями к примитивам [8]. Но в то же время частичное уменьшение доли субъективизма в выборе примитивов объектов распознавания возможно посредством применения математических методов и подходов к отбору признаков.

### Выводы

Проведенные исследования показали, что теоретический «багаж» области распознавания образов обладает широким набором средств для решения подзадачи выделения признаков распознавания. Тем не менее, для решения данной подзадачи в задачах автоматической классификации выбор достаточно ограничен: наибольшее применение получили кластерный, корреляционный и информационный подходы, в рамках которых выделение информативных признаков является «изолированным» процессом с точки зрения отсутствия обратных связей с этапом кластеризации [1].

Данная ситуация обуславливает актуальность дальнейших исследований относительно возможности повышения эффективности выделения информативных признаков в задачах автоматической классификации за счет внесения указанных выше обратных связей посредством критериев качества кластеризации. Потенциальным преимуществом данного подхода по сравнению с существующими альтернативными методами и подходами выделения признаков является учет кластеризации, что наиболее близко отображает сущность понятия «признак» («признак распознавания», «информативный признак», «значимый признак») в рамках теории распознавания образов.

**Список литературы:** 1. *Воронцов К.В.* Лекции по методам отбора и синтеза признаков и моделей. <http://ccas.ru/voron/download/FeaturesTransform.pdf> 2. *Ивахненко А.Г.* Самообучающиеся системы распознавания и автоматического управления. – К.: Техника, 1969. – 392 с. 3. *Дж. Ту, Р. Гонсалес.* Принципы распознавания образов. – М.: Мир, 1978. – 390 с. 4. *Белозерский Л.А.* Базовые понятия теории и практики современного распознавания // Искусственный интеллект. 2005. №2. С. 4–15. 5. *Афонин В.Л., Макушкин В.А.* Интеллектуальные робототехнические системы. Курс лекций. <http://www.intuit.ru/department/human/isrob/4/>. 6. *Крислов В.А., Юдин С.А.* Естественная и искусственная таксономия // Искусственный интеллект. – 2005. – №1. – С. 74–85. 7. *Браверман Э.М., Мучник И.Б.* Структурные методы обработки эмпирических данных. – М.: Наука, 1983. – 464 с. 8. *Классификация и кластер / Под ред. Дж. Райзина.* 1900. 9. *Андрейчиков А.В., Андрейчикова О.Н.* Интеллектуальные информационные системы. – М.: Финансы и статистика, 2004. – 424 с. 10. *Мандель И.Д.* Кластерный анализ. – М.: Финансы и статистика, 1988. – 176 с. 11. *Рутковская Д., Пилинский М., Рутковский Л.* Нейронные сети, генетические алгоритмы и нечеткие системы. – М.: Горячая линия, 2007. – 452 с.

*Поступила в редакцию 7.04.2008*

УДК 621.397



## МЕТОД НАХОЖДЕНИЯ ХАРАКТЕРНЫХ ПРИЗНАКОВ ФОРМЫ ЛАДОНИ ПО СКЕЛЕТОНУ ИЗОБРАЖЕНИЯ

С.В. Машталир<sup>1</sup>, Н.Л. Свередюк<sup>2</sup><sup>1</sup> ХНУРЭ, г. Харьков, Украина, mashtalir\_s@kture.kharkov.ua<sup>2</sup> ХНУРЭ, г. Харьков, Украина, nadia\_esperansa@mail.ru

Рассмотрены методы построения скелетонов изображения. Исследованы подходы к построению скелетона ладони. Проведен анализ известных признаков формы руки. На основании проведенного анализа выбран вектор характерных признаков, идентифицирующий человека.

ГЕОМЕТРИЯ ФОРМЫ РУКИ, БИОМЕТРИЯ, МАТЕМАТИЧЕСКАЯ МОРФОЛОГИЯ, СКЕЛЕТОН ИЗОБРАЖЕНИЯ, ИНТЕРПОЛЯЦИЯ СПЛАЙНАМИ

### Введение

В настоящее время всё более широкое распространение получают биометрические системы идентификации человека. На сегодняшний день такие системы являются самыми надежными. Они используются на различных секретных объектах, для защиты важной коммерческой информации и т. п. Биометрические системы основываются на уникальных биологических характеристиках человека, которые трудно подделать и которые однозначно определяют конкретного человека.

Биометрия представляет собой методику распознавания и идентификации людей на основе их индивидуальных и уникальных психологических или поведенческих характеристик. Биометрические свойства включают в себя отпечатки пальцев, форму лица, рисунок радужной оболочки глаза, рисунок сетчатки, геометрию руки, речь, почерк, особенности печати на клавиатуре и даже узор вен на запястье.

Биометрический контроль доступа – автоматизированный метод, с помощью которого путем исследования уникальных физиологических особенностей или поведенческих характеристик человека осуществляется идентификация личности.

В статье рассматривается метод идентификации личности по форме руки. Этот метод по своей технологической структуре и уровню надежности вполне сопоставим с методом идентификации личности по отпечатку пальца.

Человеческая ладонь имеет устойчивые признаки собственной формы, которые позволяют различать ладони разных людей. Основными геометрическими признаками являются: ширина ладони, радиус вписанной в ладонь окружности, длина пальцев, ширина пальцев.

Распознавание формы силуэта ладони при биометрической идентификации личности сводится к задаче сравнения формы объектов, заданных в виде бинарных изображений. Цветное изображение ладони, полученное с помощью камеры или фотоаппарата, преобразуется в монохромное изображение, в котором объект (ладонь) представлен

чёрными пикселями, а фон – белыми. Полученное бинарное изображение довольно точно описывает форму руки.

Среди подходов к сравнению формы объектов известен подход на основе выравнивания контуров границы объектов и подход, основанный на построении скелетонов. При этом второй подход является более надежным. Это связано с тем, что скелетон дает возможность более детально описать топологическую структуру формы и его использование позволяет избежать влияния дискретизации, особенно ярко проявляющегося при преобразовании поворота.

Целью данной работы является построение метода нахождения характерных признаков формы ладони, а также выделение вектора признаков, который однозначно характеризует человека. Данный метод может быть использован для построения системы идентификации личности.

### 1. Анализ существующих методов

Практически всё о конкретном человеке можно прочитать по его руке.

С руки можно собрать до 90 информационных знаков, часть из которых не используется в биометрике. Например уникальный узор на ладони.

Основными геометрическими признаками являются (см. рис. 1): ширина ладони, радиус вписанной в ладонь окружности, длины пальцев, ширина пальцев, высота кисти руки в трёх местах ( $a$ ,  $b$ ,  $c$  на рис. 1). Все эти признаки объединяются в так называемый вектор значений.

Для проведения идентификации с пользователя снимают несколько силуэтов его руки. Для каждого из этих силуэтов формируется свой вектор значений. На основе нескольких векторов значений создается специальный класс. Далее все признаки в классе усредняются, и получают признаки эталонного образа (или, говорят, находится центр класса). Сравнить между собой два образа можно по нескольким критериям. Наиболее очевидный из них – наименьшее расстояние от характеристик исследуемого образа до характеристик эталона.

Более сложный метод — снимать четыре характеристики, три из которых — характерные размеры, а четвертая — полутоновое изображение складок кожи на сгибе между фалангами. Однако для получения последней характеристики необходим специальный прибор, использующий инфракрасную подсветку и регистрацию данных специальной CCD-телекамерой. Информацию о толщине и габаритах кисти руки можно получить за счет боковых зеркал, которые попадают в обзор телекамеры. Такой метод сильно затрудняет обмен прибора, но при этом является довольно дорогостоящим. Также в этом методе используется трехразовое сканирование руки для получения различных проекций.

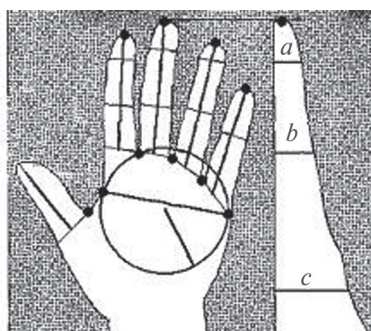


Рис. 1. Биометрические признаки ладони

В связи с этими недостатками предлагается метод нахождения признаков по скелетону изображения, который находится при помощи математической морфологии и метода интерполяции сплайнами.

## 2. Построение скелетона изображения

Существует несколько подходов к построению скелетона изображения. Первый из них — это построение скелетона с помощью методов математической морфологии. Другой подход основан на построении скелетона методом интерполяции сплайнами.

Скелетон — модель изображения, основанная на совокупности некоторых усредненных внутренних линий — должен представляться максимально тонкими линиями. Связность скелетона должна соответствовать связности исходного изображения, результат должен быть максимально центрирован и, наконец, он должен быть обратимым, то есть необходимо, чтобы по результату восстанавливалась исходная форма объекта [1].

Рассмотрим метод построения скелетона с помощью методов математической морфологии.

Морфология (гр. *morphè* — форма + *logos* — учение) в технических приложениях означает изучение формы объектов по информации о них, как правило, в визуальной форме [1, 2, 3]. Это направление базируется на операциях алгебры Минковского [4]. Операции бинарной морфологии являются реализацией операций алгебры Минковского применительно к обработке изображений [3, 5].

Любое изображение в компьютерной графике обычно представляется также в виде набора пикселей, поэтому операции математической морфологии могут быть применены и к изображению для исследования некоторых свойств его формы и структуры.

Мы будем рассматривать морфологию двоичных изображений. Двоичное изображение представляется в виде упорядоченного набора (упорядоченного множества) черно-белых точек (пикселей) или 0 и 1. Под областью (*region*) изображения обычно понимается некоторое подмножество точек изображения. Каждая операция двоичной морфологии является некоторым преобразованием этого множества. В качестве исходных данных принимаются двоичное изображение  $B$  и некоторый структурный элемент  $S$ . Результатом операции также является двоичное изображение.

Структурный элемент — также некоторое двоичное изображение (геометрическая форма — *shape*). Он может быть произвольного размера и произвольной структуры. Чаще всего используются симметричные элементы, такие как прямоугольник фиксированного размера или круг некоторого диаметра. В каждом элементе выделяется особая точка, называемая *начальной* (*origin*). Она может быть расположена в любом месте элемента, хотя в симметричных множествах это обычно центральный пиксель. Примеры структурных элементов для традиционных видов обработки бинарных изображений приведены на рис. 2. Структурные элементы  $a - g$  применяются для выделения границ, элементы  $d - m$  используются при параллельном переносе изображения.

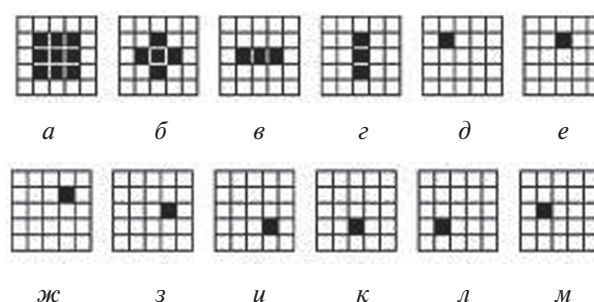


Рис. 2. Примеры структурного элемента

Вначале результирующая поверхность заполняется 0, образуя полностью черное изображение. Затем осуществляется сканирование исходного изображения структурным элементом. Для сканирования каждого пикселя на изображение «накладывается» структурный элемент так, чтобы совместились зондируемая и начальные точки. Затем проверяется некоторое условие на соответствие пикселей структурного элемента и точек изображения «под ним». Если условие выполняется, то на результирующем изображении в соответствующем месте ставится 1 (в некоторых случаях будет добав-

ляться не один единичный пиксель, а все единички из структурного элемента).

Скомбинировав несколько операций математической морфологии, можно получить производную, позволяющую выделять из объекта его скелет, и эта операция получила название «скелетонизации» [6, 7, 8].

$n$ -м элементом скелета  $S$  изображения  $X$  по структурному элементу  $Q$  называется

$$S_n = (X - nQ) \setminus [(X - nQ) \circ Q], n = 0, 10, \dots, N,$$

где  $N = \max\{n: X - nQ \neq \emptyset\}$ ;  $\emptyset$  – пустое множество;  $\setminus$  – теоретико-множественно вычитание.

Тогда скелетом  $S$  изображения  $X$  по структурному элементу  $Q$  назовем объединение

$$S = \bigcup_{k \leq n \leq N} S_n.$$

Метод математической морфологии выделения скелетона удобен тем, что с помощью применения операции расширения по тому же структурному элементу к скелетону мы сможем восстановить исходное изображение. Также к его свойствам относятся инвариантность относительно преобразований смещения, масштаба и поворота.

На рис. 3, а представлен пример бинарного изображения и скелетон (рис. 3, б), полученный при использовании операций математической морфологии.

К достоинствам данного метода можно отнести его простоту и высокую скорость работы.

Рассмотрим второй метод построения скелетона – метод, основанный на применении интерполяции сплайнами.

Построение скелетона этим методом происходит в два этапа. Первый этап – это преобработка.

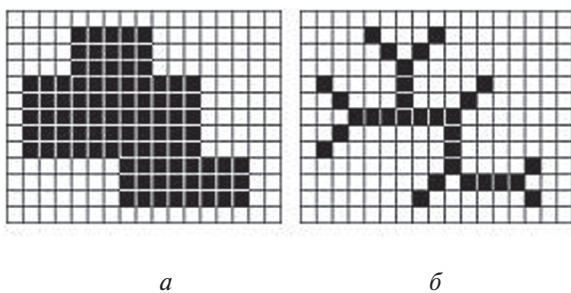


Рис. 3. Пример бинарного изображения и соответствующего скелетона

В случае входного изображения в градациях серого:

- нормализация изображения (находятся самый светлый и самый темный оттенки, самый светлый объявляется белым, самый темный объявляется черным; промежуточные оттенки рассчитываются из соответствующей пропорции);

- 4-х проходное (горизонтальное, вертикальное и два диагональных) сканирование исходного об-

разца и сложение получившихся 4-х изображений в одно.

С помощью этого мы приводим изображение к бинарному.

В случае входного цветного изображения на основании информации о цвете выбирается порог, после чего изображение преобразуется к бинарному.

Пример преобработки изображения представлен на рис. 4.

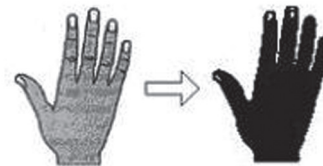


Рис. 4. Пример преобразования изображения в бинарное

Затем идет второй этап – построение скелетона. Построение осуществляется на основе вычисления точек перегиба на переходах белое/черное и черное/белое и последующего вычисления середин этих отрезков с помощью сплайновой интерполяции. Точки, поставленные в этих серединах и соединенные между собой, и есть изображение скелетона. Второй этап проводится в два прохода: горизонтальный и вертикальный. В результате получается скелетон исходного образца (см. рис. 5), по которому, в случае необходимости, мы можем восстановить сам образец.

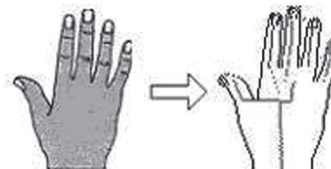


Рис. 5. Построение скелетона ладони

Для того чтобы очистить изображение от случайных точек, являющихся шумом, к уже построенному скелетону возможно применение критерия «8 соседей» (рис. 6). Он заключается в подсчете количества закрасенных соседних точек для каждой данной точки.

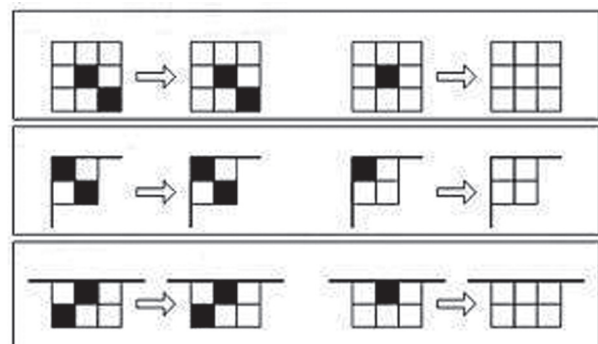


Рис. 6. Удаление одиночных точек

В случае наличия таковых, данная точка оставляется нетронутой. В противном случае данная точка удаляется. Очевидно, что для краевых точек надо производить проверку только для пяти соседей, а для угловых — только для трех.

К достоинствам данного метода можно отнести способность алгоритма довольно сильно очищать исходный образец.

### 3. Выбор характеристик ладони

В предлагаемом методе идентификации человека используется вектор характерных признаков, получаемый по скелетону изображения. С учетом характерных признаков ладони человека выбран вектор из десяти признаков:

$$A = (a_1, a_2, \dots, a_{10}),$$

где  $a_1$  — ширина кисти;  $a_2$  — ширина ладони;  $a_3, a_4, a_5, a_6, a_7$  — длины пяти пальцев;  $a_8$  — отношение длины указательного пальца к длине среднего;  $a_9$  — отношение длины безымянного пальца к длине среднего;  $a_{10}$  — отношение длины большого пальца к длине среднего.

Такие характеристики можно получить без использования специальных приборов с использованием инфракрасной подсветки и системы зеркал. Используемые в работе характеристики ладони можно получить с изображения ладони, что значительно упрощает технические требования к прибору для сканирования руки, так как необходима лишь фотокамера.

Для выделения признаков на изображении ладони необходимо сначала преобразовать изображение в бинарное, далее, используя описанные выше методы, построить скелетон ладони. По построенному скелетону выделяются характерные признаки ладони.

Признак  $a_1$  определяется на уровне нижней точки скелетона (точка  $A$ , см. рис. 7) как расстояние от точки перегиба на переходах белое/черное до точки перегиба черное/белое.

Значение параметра  $a_2$  вычисляется на уровне точки  $B$ , от которой расходятся линии пальцев. Так же как и ширина кисти, она определяется как расстояние от точки перегиба на переходах белое/черное до точки перегиба черное/белое.

Длины пальцев (признаки  $a_3 - a_7$ ) — достаточно уникальные характеристики ладони человека. Эти признаки используются во многих современных биометрических системах идентификации человека и достаточно хорошо характеризуют форму ладони. В данной работе, однако, под длиной пальца понимается длина соответствующей линии скелетона от конца пальца до точки соединения этой линии с линией ладони — точки  $B$ .

Однако использование в качестве признаков только длин пальцев не достаточно уникально ха-

рактеризует ладонь и могут возникнуть ошибки при идентификации. Общеизвестно, что длины указательного и безымянного пальцев у разных людей достаточно сильно различаются. Кроме того, у одних людей указательный палец длиннее, чем безымянный, у других — короче. Разность длин этих пальцев у разных людей также отличается. Эта особенность была учтена при выборе таких характеристик как отношение длин указательного и безымянного пальцев к длине среднего пальца (параметры  $a_8, a_9$ ). Большой палец также несет в себе большой объем информации о руке. Отношение длины большого пальца к длине среднего (параметр  $a_{10}$ ) также внесено в вектор характеристик ладони.

Значение величин признаков определяется следующим образом. По скелетону изображения определяется длина линии, соответствующей определяемому признаку, в пикселях. Далее, используя информацию о разрешении фотокамеры, значение признака приводится к стандартным единицам измерения (пиксели — к дюймам, дюймы — к сантиметрам).

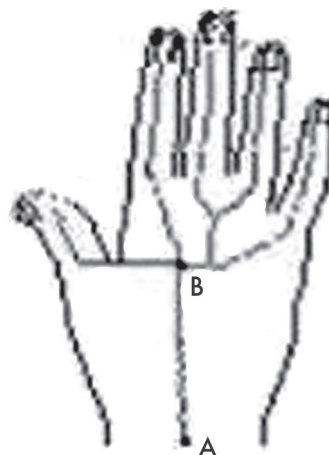


Рис. 7. Нахождение характерных признаков ладони по скелетону

Выбор таких характерных признаков формы руки основан на особенностях строения человеческой ладони. Все характеристики, использованные в данном методе, получены из изображения ладони на основании построенного скелетона изображения. Кроме того, полученные признаки инвариантны к преобразованиям смещения, масштабирования и поворотам на небольшие углы, что следует из инвариантности к указанным преобразованиям скелетона изображения.

Далее производится идентификация человека. С руки человека, допущенного на объект, делаются снимки, с которых определяют эталонный вектор изображений, который заносится в базу данных. При идентификации получаем изображение ладони, для которого выделяем вышеуказанным методом характерные признаки. По полученно-

му вектору признаков происходит поиск в базе данных. Сравнение ведется по признакам, а не по самому изображению, что позволяет избежать многих ошибок, так как при сканировании ладони человека могут возникать искажения, связанные с различным расположением камеры и объекта. При сравнении по характерным признакам они остаются неизменными независимо от положения ладони.

### Выводы

В статье рассмотрен метод идентификации человека, основанный на биометрических характеристиках. Предложен метод распознавания, основанный на использовании геометрии ладони. Характерные признаки формы ладони определяются по скелетону изображения. Рассмотрены методы построения скелетона изображения. Проанализированы информационные знаки ладони и выбран вектор характеристик, использующийся для распознавания.

Метод идентификации человека, основанный на предложенном методе нахождения характерных признаков ладони, позволяет избежать ошибок при распознавании, которые могут возникнуть при сравнении изображений.

Следует отметить, что серьезным преимуществом предложенного метода является то, что использование скелетона позволяет не задавать ограничений на положение руки идентифицируемого человека, а также метод позволяет проводить анализ не по непосредственно изображению ладони, а по вектору характеристик, которые получены на

основании этого изображения. В результате это позволяет уменьшить объем обрабатываемой информации, что позволяет хранить большой объем записей и быстро осуществлять поиск. Кроме того, предлагаемый метод существенно упрощает процесс получения характеристик, что приводит к увеличению скорости идентификации и уменьшению расходов на оборудование.

**Список литературы:** 1. Maragos P. A., Schafer R.W. Morphological skeleton representation and coding of binary images // IEEE Trans. on Acoustics, Speech, and Signal Processing. – Vol. 34, №9. – 1986. – P. 1228–1244. 2. Heijmans H.J.A.M., Ronse C. The algebraic basis of mathematical morphology: I. Dilations and erosions // Computer Vision, Graphics, and Image Processing. – Vol. 50, №3. – 1990. – P. 245–295. 3. Serra J. Image analysis and mathematical morphology. – Vol. I. – London: Academic Press, 1982. – 610 p. 4. Хадвигер Г. Лекции об объеме, площади поверхности и изопериметрии. – М.: Наука, 1966. – 416 с. 5. Verwer B.J.H., van Vliet L.J., Verbeek P.W. Binary and grey-value skeletons: Metrics and algorithms // International Journal of Pattern Recognition and Artificial Intelligence. – Vol. 7, №5. – 1993. – P. 1287–1308. 6. Болдак А.А., Мохамед Хашим Р. Метод скелетизации изображений // Вісник нац. технічного ун-ту України «КПІ», Інформатика, управління та обчислювальна техніка. – Київ: КПІ, 2000. – №36. – С. 65–80. 7. Choi W.P., Lam K.M., Siu W.C. Extraction of the Euclidean skeleton based on a connectivity criterion // Pattern Recognition. – Vol. 36, №3. – 2003. – P. 721–729. 8. Fan K.C., Chen D.F., Wen M.G. Skeletonization of binary images with nonuniform width via block decomposition and contour vector matching // Pattern Recognition. – Vol. 31, №7. – 1998. – P. 823–838.

*Поступила в редколлегию 8.04.2008*

УДК 519.816



## ФОРМАЛИЗАЦИЯ НЕЧЕТКОЙ ИНФОРМАЦИИ ПРИ ОЦЕНИВАНИИ КАЧЕСТВЕННЫХ ПРИЗНАКОВ В ВЕРБАЛЬНЫХ ШКАЛАХ

Н.М. Кораблёв

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

Рассматривается построение моделей экспертного оценивания качественных признаков на основе семантических пространств в вербальных шкалах, значениями которых являются слова, выражающие степень интенсивности проявления признаков. Эти модели могут применяться для формализации информации в рамках как качественных, так и количественных признаков.

ФУНКЦИЯ ПРИНАДЛЕЖНОСТИ, ВЕРБАЛЬНАЯ ШКАЛА, ТЕРМ-МНОЖЕСТВО, СЕМАНТИЧЕСКОЕ ПРОСТРАНСТВО, ПРИЗНАК

### Введение

Важным этапом обработки информации является этап ее формализации, то есть этап представления этой информации в виде, позволяющем на следующих этапах ее обработки применять аппараты известных математических теорий. Известно, что нечеткая экспертная информация трудно формализуема при использовании традиционных подходов [1, 2]. Модельный подход на основе аппарата теории нечетких множеств позволил устранить недостатки традиционных формализаций нечеткой экспертной информации [1–4]. С точки зрения этого подхода моделями экспертного оценивания признаков служат семантические пространства (СП), термы которых соответствуют уровням вербальных шкал, используемых для оценивания признаков [5, 6]. Однако не все модели, построенные на основе СП, обладают свойствами, обеспечивающими успешность решения практических задач на основе этих моделей. Одним из таких свойств является свойство полноты, которое состоит в возможности описания каждого из элементов универсального множества в лингвистических термах этого пространства [4, 7, 8].

В [9] рассмотрены вопросы формализации нечеткой экспертной информации на основе СП путем опроса одного эксперта о типичных значениях термов и о разбиении универсального множества признака. Вместе с тем для оценивания качественных признаков и для описания количественных признаков используют вербальные шкалы, значениями которых являются слова, выражающие степень интенсивности проявления признаков [10]. При формализации значений количественных признаков с помощью вербальных шкал имеет место недостаток, состоящий в том, что при описании объектов с пограничными значениями показателя эксперт испытывает трудности в связи со скачкообразным переходом от одного значения к другому. Устранить этот недостаток позволяет аппарат теории нечетких множеств, с позиций которого вербальным уровням количественного признака в

соответствие ставятся не четкие интервалы значений, а нечеткие множества.

При построении лингвистической шкалы для качественных признаков нельзя однозначно определить для них универсальные множества, как для количественных признаков. Поэтому разработка методов формализации нечеткой экспертной информации на основе СП в рамках вербальных шкал является актуальной задачей.

### 1. Постановка задачи

Пусть  $X$  – некоторое множество  $x$  с функцией принадлежности (ФП)  $\mu_{\hat{A}}(x)$  к некоторому множеству  $\hat{A}$  с терм-множеством  $T(x) = \{X_1, X_2, \dots, X_m\}$ , где  $X_l$  – терм, соответствующий минимальной интенсивности проявления признака,  $X_m$  – терм, соответствующий максимальной интенсивности проявления признака на универсальном множестве  $U = [a, b]$ . Предположим, что экспертом определены типичные для термов  $X_l, l = \overline{1, m}$  интервалы  $(x_l^1, x_l^2), l = \overline{1, m}$ , то есть интервалы, для всех точек которых ФП соответствующих термов равны единице.

Необходимо построить ФП терм-множеств СП на основе апостериорной информации, полученной в результате оценивания экспертом качественного признака  $X$  у совокупности объектов, которые удовлетворяли бы требованиям к ФП  $\mu_l(x), l = \overline{1, m}$  их терм-множеств [4]:

1. Существует  $\hat{U}_l \neq \emptyset$ , где  $\hat{U}_l = \{x \in U : \mu_l(x) = 1\}$  есть точка или отрезок.

2. Если  $\hat{U}_l = \{x \in U : \mu_l(x) = 1\}$ , тогда  $\mu_l(x), l = \overline{1, m}$  не убывает слева от  $\hat{U}_l$  и не возрастает справа от  $\hat{U}_l$ .

3. ФП  $\mu_l(x), l = \overline{1, m}$  имеют не более двух точек разрыва первого рода.

4. Для каждого  $x \in U$  существует  $l, l = \overline{1, m} : \mu_l(x) \neq 0$ .

5. Для каждого  $x \in U$   $\sum_{l=1}^m \mu_l(x) = 1$ .

ФП, которые удовлетворяют этим требованиям, обладают свойством полноты [4].

Математической основой построения методов обработки и анализа нечеткой информации с ис-

пользованием нечетких множеств является алгебра нечетких чисел. Поэтому рассмотрим сначала построение совокупности нечетких чисел, используемых для формализации лингвистических значений признаков.

## 2. Построение совокупности нечетких чисел для формализации лингвистических значений признаков

Известно [5], что нечеткое число  $\tilde{A}$  с ФП  $\mu_{\tilde{A}}(x)$  называется нормальным, если  $\max_x \mu_{\tilde{A}}(x) = 1, x \in R$ .

Нечеткое число  $\tilde{A}$  с ФП  $\mu_{\tilde{A}}(x)$  называется унимодальным, если существует единственная точка  $x \in R$ , для которой выполняется равенство  $\mu_{\tilde{A}}(x) = 1$ . Нечеткое число  $\tilde{A}$  с ФП  $\mu_{\tilde{A}}(x)$  называется многомодальным, если точка  $x \in R: \mu_{\tilde{A}}(x) = 1$  не является единственной, и толерантным, если существуют интервалы, для всех точек которых выполняется равенство  $\mu_{\tilde{A}}(x) = 1$ . Этот интервал называется интервалом толерантности нечеткого числа  $\tilde{A}$ .

Рассмотрим толерантные и унимодальные  $(L-R)$ -числа с ФП [10]

$$\mu_{\tilde{A}}(x) = \begin{cases} L\left(\frac{a_1 - x}{a_L}\right), & 0 \leq \frac{a_1 - x}{a_L} \leq 1, a_L > 0; \\ R\left(\frac{x - a_2}{a_R}\right), & 0 \leq \frac{x - a_2}{a_R} \leq 1, a_R > 0; \\ 1, & \frac{a_1 - x}{a_L} < 0 \cap \frac{x - a_2}{a_R} < 0; \\ 0, & \frac{a_1 - x}{a_L} > 1 \cup \frac{x - a_2}{a_R} > 1 \end{cases} \quad (1)$$

и следующими условиями на функции  $L$  и  $R$ :

1)  $L(0) = R(0) = 1, L(1) = R(1) = 0,$

2)  $L(x)$  и  $R(x)$  – монотонно убывающие функции на множестве  $[0, 1]$ .

Нечеткое число  $\tilde{A}$  записывается в виде  $\tilde{A} \equiv (a_1, a_2, a_L, a_R)$  (или  $\mu_{\tilde{A}}(x) \equiv (a_1, a_2, a_L, a_R)$ ), где  $a_1, a_2, a_L, a_R$  являются параметрами толерантного  $(L-R)$ -числа  $\tilde{A}$ . Отрезок  $[a_1, a_2]$  называется интервалом толерантности, а  $a_L$  и  $a_R$  – соответственно левым и правым коэффициентами нечеткости.

Функция  $L\left(\frac{a_1 - x}{a_L}\right)$  является левой границей ФП толерантного  $(L-R)$ -числа, а функция  $R\left(\frac{x - a_2}{a_R}\right)$

является правой границей ФП толерантного  $(L-R)$ -числа. При  $a_L = 0$  предполагается, что  $L\left(\frac{a_1 - x}{a_L}\right) = 0$ ,

при  $a_R = 0$  предполагается, что  $R\left(\frac{x - a_2}{a_R}\right) = 0$ .

Унимодальное  $(L-R)$ -число  $\tilde{A}$  имеет ФП толерантного  $(L-R)$ -числа при условии  $a_1 = a_2$ . Симво-

лично унимодальное  $(L-R)$ -число  $\tilde{A}$  записывается в виде  $\tilde{A} \equiv (a_1, a_L, a_R)$ .

Обозначим через  $\Lambda$  совокупность всех толерантных и унимодальных чисел с условиями на функции  $L$  и  $R$ . Элементы совокупности  $\Lambda$  будем называть  $\Lambda$ -числами, которые используем для построения ФП качественных признаков в вербальных шкалах.

## 3. Построение функций принадлежности качественных признаков в вербальных шкалах

Построим ФП терм-множеств СП на основе апостериорной информации, полученной в результате оценивания экспертом качественного признака  $X$  у совокупности объектов. Предполагается, что оценивание признака осуществлялось в рамках вербальной шкалы с уровнями  $X_l, l = \overline{1, m}, m \geq 2$ , упорядоченными по возрастанию интенсивности проявления признака. Уровни используемой вербальной шкалы однозначно задают термножество –  $T(x) = \{X_1, X_2, \dots, X_m\}$ . В качестве универсального множества СП признака  $X$  выбрано множество  $U = [0, 1]$ . Точка  $x = 0$  соответствует полному отсутствию проявления признака  $X$  и поэтому считается типичной точкой терма  $X_1$ , точка  $x = 1$  соответствует полному присутствию проявления признака  $X$  и поэтому считается типичной точкой терма  $X_m$ .

Обозначим относительные частоты появления объектов, у которых интенсивность признака  $X$  оценена уровнями  $X_l, l = \overline{1, m}$ , соответственно через  $a_l, l = \overline{1, m}, \sum_{l=1}^m a_l = 1$ . Будем предполагать, что нечеткие числа, соответствующие термам  $X_l, l = \overline{1, m}$  с ФП  $\mu_l(x), l = \overline{1, m}$ , принадлежат совокупности  $\Lambda$  и удовлетворяют дополнительному условию: если  $L(x), R(x)$  – нелинейные, то они имеют центральную симметрию относительно точки перегиба.

Построение ФП терм-множеств СП будет осуществляться таким образом, чтобы обеспечивалось выполнение требований, предъявляемых к этим функциям в рамках определения СП, и чтобы площади фигур, ограниченных графиками функций  $\mu_l(x), l = \overline{1, m}$  и осью абсцисс, равнялись  $a_l, l = \overline{1, m}$ . Построение необходимо ограничить логичными требованиями на области нечеткости между соседними термами (или параметры нечеткости нечетких чисел, соответствующих термам). С одной стороны, эту область желательно сделать как можно меньше, тогда соответственно будет меньше нечеткость построенной в виде СП модели. С другой стороны, необходимо опираться на содержательный смысл области нечеткости, поэтому предлагается мощность (длину) этой области для крайних термов вычислять как  $\min(a_1, a_2)$  или соответственно  $\min(a_{m-1}, a_m)$ , а для средних термов вычислять, исходя из соотношений между числами

$a_{l-1}, a_l, a_{l+1}, l = \overline{3, m-2}$ . Графики построенных ФП будут представлять собой криволинейные трапеции со средними линиями, равными  $a_l, l = \overline{1, m}$ .

Построим функцию принадлежности терма  $X_m$ :

1. Если  $a_m \leq a_{m-1}$ , то

$$\mu_m(x) = \begin{cases} 0, & 0 \leq x \leq 1 - \frac{3a_m}{2}; \\ L\left(\frac{1 - \frac{a_m}{2} - x}{a_m}\right), & 1 - \frac{3a_m}{2} < x \leq 1 - \frac{a_m}{2}; \\ 1, & 1 - \frac{a_m}{2} < x \leq 1. \end{cases} \quad (2)$$

На рис. 1 и 2 изображены функции принадлежности терма  $X_m$  для частного случая:  $L(x) = 1 - x, 0 \leq x \leq 1$  — функции принадлежности  $T$ -чисел.

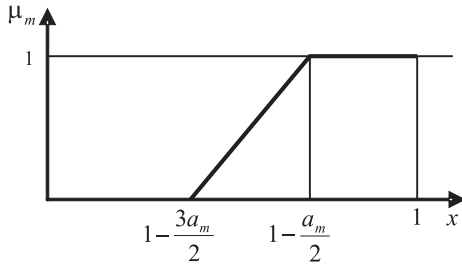


Рис. 1. Функция принадлежности (2)

2. Если  $a_m > a_{m-1}$ , то

$$\mu_m(x) = \begin{cases} 0, & 0 \leq x \leq 1 - a_m - \frac{a_{m-1}}{2}; \\ L\left(\frac{1 - a_m + \frac{a_{m-1}}{2} - x}{a_{m-1}}\right), & 1 - a_m - \frac{a_{m-1}}{2} < x \leq 1 - a_m + \frac{a_{m-1}}{2}; \\ 1, & 1 - a_m + \frac{a_{m-1}}{2} < x \leq 1. \end{cases} \quad (3)$$

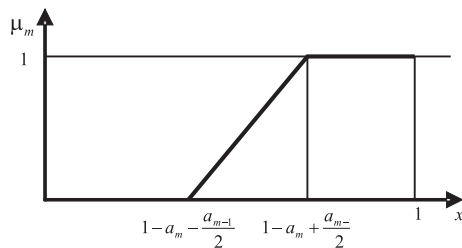


Рис. 2. Функция принадлежности (3)

Построим функцию принадлежности терма  $X_{m-1}$ :

1. Если  $a_{m-1} \geq \max(a_m, a_{m-2})$ , то

$$\mu_{m-1}(x) = \begin{cases} 0, & 0 \leq x \leq \sum_{l=1}^{m-3} a_l + \frac{a_{m-2}}{2}; \\ R\left(\frac{\sum_{l=1}^{m-2} a_l + \frac{a_{m-2}}{2} - x}{a_{m-2}}\right), & \sum_{l=1}^{m-3} a_l + \frac{a_{m-2}}{2} < x \leq \sum_{l=1}^{m-2} a_l + \frac{a_{m-2}}{2}; \\ 1, & \sum_{l=1}^{m-2} a_l + \frac{a_{m-2}}{2} < x \leq 1 - \frac{3a_m}{2}; \\ L\left(\frac{x - 1 + \frac{3a_m}{2}}{a_m}\right), & 1 - \frac{3a_m}{2} < x \leq 1 - \frac{a_m}{2}; \\ 0, & 1 - \frac{a_m}{2} < x \leq 1. \end{cases} \quad (4)$$

На рис. 3–6 изображены функции принадлежности терма  $X_{m-1}$  для частного случая:  $L(x) = 1 - x, R(x) = 1 - x, 0 \leq x \leq 1$ .

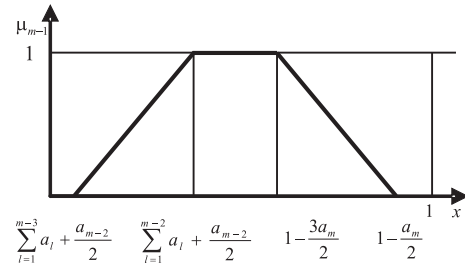


Рис. 3. Функция принадлежности (4)

2. Если  $a_m < a_{m-1} < a_{m-2}$ , то

$$\mu_{m-1}(x) = \begin{cases} 0, & 0 \leq x \leq 1 - a_m - \frac{3a_{m-1}}{2}; \\ R\left(\frac{1 - a_m - \frac{a_{m-1}}{2} - x}{a_{m-1}}\right), & 1 - a_m - \frac{3a_{m-1}}{2} < x \leq 1 - a_m - \frac{a_{m-1}}{2}; \\ 1, & 1 - a_m - \frac{a_{m-1}}{2} < x \leq 1 - \frac{3a_m}{2}; \\ L\left(\frac{x - 1 + \frac{3a_m}{2}}{a_m}\right), & 1 - \frac{3a_m}{2} < x \leq 1 - \frac{a_m}{2}; \\ 0, & 1 - \frac{a_m}{2} < x \leq 1. \end{cases} \quad (5)$$

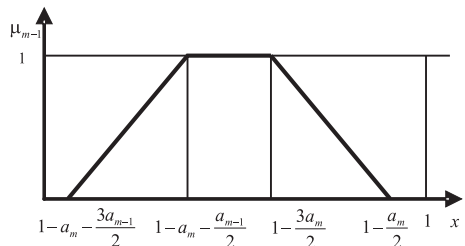


Рис. 4. Функция принадлежности (5)

3. Если  $a_{m-2} < a_{m-1} < a_m$ , то

$$\mu_{m-1}(x) = \begin{cases} 0, & 0 \leq x \leq \sum_{l=1}^{m-3} a_l + \frac{a_{m-2}}{2}; \\ R \left( \frac{\sum_{l=1}^{m-2} a_l + \frac{a_{m-2}}{2} - x}{a_{m-2}} \right), & \sum_{l=1}^{m-3} a_l + \frac{a_{m-2}}{2} < x \leq \sum_{l=1}^{m-2} a_l + \frac{a_{m-2}}{2}; \\ 1, & \sum_{l=1}^{m-2} a_l + \frac{a_{m-2}}{2} < x \leq 1 - a_m - \frac{a_{m-1}}{2}; \\ L \left( \frac{x - 1 + a_m + \frac{a_{m-1}}{2}}{a_{m-1}} \right), & 1 - a_m - \frac{a_{m-1}}{2} < x \leq 1 - a_m + \frac{a_{m-1}}{2}; \\ 0, & 1 - a_m + \frac{a_{m-1}}{2} < x \leq 1. \end{cases} \quad (6)$$

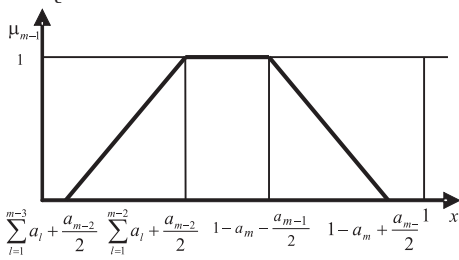


Рис. 5. Функция принадлежности (6)

4. Если  $a_{m-1} \leq \min(a_m, a_{m-2})$ , то

$$\mu_{m-1}(x) = \begin{cases} 0, & 0 \leq x \leq 1 - a_m - \frac{3a_{m-1}}{2}; \\ R \left( \frac{1 - a_m - \frac{a_{m-1}}{2} - x}{a_{m-1}} \right), & 1 - a_m - \frac{3a_{m-1}}{2} < x \leq 1 - a_m - \frac{a_{m-1}}{2}; \\ L \left( \frac{x - 1 + a_m + \frac{a_{m-1}}{2}}{a_{m-1}} \right), & 1 - a_m - \frac{a_{m-1}}{2} < x \leq 1 - a_m + \frac{a_{m-1}}{2}; \\ 0, & 1 - a_m + \frac{a_{m-1}}{2} < x \leq 1. \end{cases} \quad (7)$$

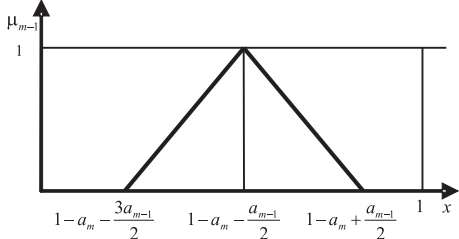


Рис. 6. Функция принадлежности (7)

Аналогично ФП  $\mu_{m-1}(x)$  строятся ФП  $\mu_l(x), l = 2, m-2$ .

Построим ФП для терма  $X_l$  при условии четного числа термов:

1. Если  $a_1 \leq a_2$ , то

$$\mu_1(x) = \begin{cases} 1, & 0 \leq x \leq \frac{a_1}{2}; \\ L \left( \frac{x - \frac{a_1}{2}}{a_1} \right), & \frac{a_1}{2} < x \leq \frac{3a_1}{2}; \\ 0, & \frac{3a_1}{2} < x \leq 1. \end{cases} \quad (8)$$

На рис. 7 и 8 изображены ФП для частного случая:  $L(x) = 1 - x, 0 \leq x \leq 1$ .

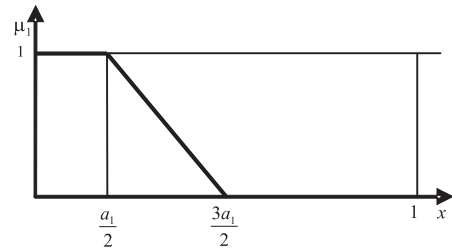


Рис. 7. Функция принадлежности (8)

2. Если  $a_1 > a_2$ , то

$$\mu_1(x) = \begin{cases} 1, & 0 \leq x \leq a_1 - \frac{a_2}{2}; \\ L \left( \frac{x - a_1 + \frac{a_2}{2}}{a_2} \right), & a_1 - \frac{a_2}{2} < x \leq a_1 + \frac{a_2}{2}; \\ 0, & a_1 + \frac{a_2}{2} < x \leq 1. \end{cases} \quad (9)$$

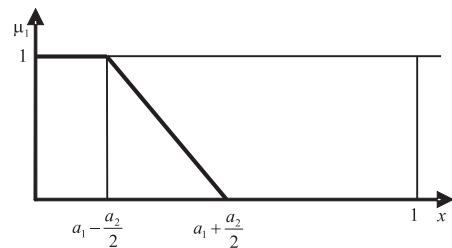


Рис. 8. Функция принадлежности (9)

При нечетком числе термов получаем:

1. Если  $a_1 \leq a_2$ , то

$$\mu_1(x) = \begin{cases} 1, & 0 \leq x \leq \frac{a_1}{2}; \\ R \left( \frac{x - \frac{a_1}{2}}{a_1} \right), & \frac{a_1}{2} < x \leq \frac{3a_1}{2}; \\ 0, & \frac{3a_1}{2} < x \leq 1. \end{cases} \quad (10)$$

2. Если  $a_1 > a_2$ , то

$$\mu_1(x) = \begin{cases} 1, & 0 \leq x \leq a_1 - \frac{a_1}{2}; \\ R\left(\frac{x - a_1 + \frac{a_2}{2}}{a_2}\right), & a_1 - \frac{a_2}{2} < x \leq a_1 + \frac{a_2}{2}; \\ 0, & a_1 + \frac{a_2}{2} < x \leq 1. \end{cases} \quad (11)$$

Вид ФП (10) и (11) совпадает соответственно с видом ФП (8) и (9), изображенных на рис. 7 и 8.

Примеры некоторых нелинейных функций  $L(x)$  и  $R(x)$ , которые могут использоваться при построении ФП СП рассмотренным методом, имеют вид:

$$1. L(x) = \begin{cases} 2(x - \frac{1}{2})^2 + \frac{1}{2}, & 0 \leq x \leq \frac{1}{2}; \\ -2(x - \frac{1}{2})^2 + \frac{1}{2}, & \frac{1}{2} < x \leq 1. \end{cases}$$

$$2. R(x) = \begin{cases} \sqrt{\frac{1}{4} - \frac{x}{2}} + \frac{1}{2}, & 0 \leq x \leq \frac{1}{2}; \\ \sqrt{\frac{x}{2} - \frac{1}{4}} + \frac{1}{2}, & \frac{1}{2} < x \leq 1. \end{cases}$$

Предложенный метод построения ФП термножеств СП можно применять не только в условиях апостериорной информации, представленной к обработке в настоящее время. Эксперт может строить СП в отсутствие такой информации в настоящий момент, пользуясь информацией, которой он обладал раньше на основании своего опыта. Этот метод может применяться для построения ФП термножеств СП, формализующего информацию по оцениванию некоторого объекта несколькими экспертами. Задачу в такой постановке можно считать двойственной задачей по отношению к задаче оценивания одним экспертом некоторой совокупности объектов. ФП термов СП по данному методу могут быть получены даже в том случае, когда невозможно привлечение экспертов для проведения стандартных процедур парных сравнений результатов [2] с целью получения значений этих функций.

## Выводы

Разработан метод формализации нечеткой информации, полученной в результате оценивания качественных признаков в вербальных шкалах. Метод имеет существенное достоинство, состоящее в том, что построение ФП формализованной информации может осуществляться в условиях большого числа оцениваемых объектов, и для этого построения не требуется никакой дополнительной информации (полученной от экспертов, проводящих стандартные парные сравнения объектов друг с другом). Разработанный метод может применяться для построения ФП моделей экспертного оценивания не только в рамках информации, полученной непосредственно после проведения оценочных процедур, но также на основе информации из предыдущего опыта их проведения.

**Список литературы:** 1. *Андрейчиков А.В., Андрейчикова О.Н.* Интеллектуальные информационные системы. – М.: Финансы и статистика, 2004. – 424 с. 2. *Борисов А.Н., Крумберг О.А., Федоров И.П.* Принятие решений на основе нечетких моделей: Примеры использования. – Рига.: Зинатне, 1990. – 184 с. 3. *Прикладные нечеткие системы / Под ред. Т. Тэрано, К. Асни, М. Сугэно:* Пер. с япон. – М.: Мир, 1993. – 368 с. 4. *Рыжов А.П.* Элементы теории нечетких множеств и измерения нечеткости – М.: Диалог-МГУ, 1998. – 116 с. 5. *Аверкин А.Н., Батыршин И.З., Блишун А.Ф. и др.* Нечеткие множества в моделях управления и искусственного интеллекта. – М.: Наука. Гл. ред. физ.-мат. лит., 1986. – 312 с. 6. *Ежкова И.В.* Семантически – инвариантная формализация лингвистических оценок // В кн. Семантические аспекты формализации интеллектуальной деятельности. – М.: МДНТП, 1983. – С.48-51. 7. *Полещук О.М.* Методы представления экспертной информации в виде совокупности термножеств полных ортогональных семантических пространств // Вестник МГУЛ. – 2002. – № 5 (25). – С. 198–216. 8. *Полещук О.М.* О развитии систем обработки нечеткой информации на базе полных ортогональных семантических пространств // Вестник МГУЛ. – 2003. №1 (26). – С. 112–117. 9. *Кораблев Н.М.* Формализация нечеткой информации на основе опроса одного эксперта // Системы обработки информации. – 2007. – Вып. 9 (67). – С. 20–23. 10. *Литвак Б.Г.* Экспертные оценки и принятие решений. – М.: Патент, 1996. – 271 с.

Поступила в редколлегию 10.04.2008

УДК 615.47:617-089



## ПОИСК ИНФОРМАТИВНЫХ КОЛИЧЕСТВЕННЫХ ПОКАЗАТЕЛЕЙ ЭЛЕКТРОМИОГРАФИЧЕСКОГО СИГНАЛА.

### Сообщение 2

А. И. Бых<sup>1</sup>, Т. В. Жемчужкина<sup>2</sup>, Т. В. Носова<sup>3</sup>

<sup>1, 2, 3</sup> 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 геометрическими методами 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.

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. В 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 [1,2].

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.

В сообщении 1 [3] нами были рассмотрены количественные показатели ЭМГ-сигнала, полученные методом временного статистического исследования биомедицинских сигналов. Данные показатели, по нашему мнению, могут быть использованы для дифференциальной диагностики состояния

нервно-мышечной ткани. Было показано, что предложенные нами характеристики (математическое ожидание амплитуд  $M_A$  и длительностей  $M_T$ , их среднеквадратические отклонения  $SD_A$  и  $SD_T$ , асимметрия амплитуд  $As_A$  и длительностей  $As_T$ , эксцесс амплитуд  $Ex_A$  и длительностей  $Ex_T$ ) являются статистически независимыми и несут дополнительную диагностическую информацию о состоянии нервно-мышечной ткани пациента.

#### 1. Цель и постановка задачи

Целью настоящего 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: геометрические методы исследования функции плотности распределения различных статистических характеристик ЭМГ-сигнала.

#### 2. Количественный анализ ЭМГ

Под геометрическими методами статистического анализа будем понимать оценку формы и параметров функции распределения плотности вероятности амплитуд и интервалов ЭМГ-сигнала. (то есть амплитудной и интервальной гистограмм ЭМГ)

Распределение вероятностей амплитуд интерференционной ЭМГ описывается функцией пуассоновского типа [4]. Из аналитического выражения для плотности распределения вероятности  $p(x)$  следует, что при  $x=0$  имеет место пик в распределении, который соответствует ненулевой вероятности нулевых значений интерференционной ЭМГ [4]. В частном случае, когда случайными являются только амплитуды элементарных импульсов ЭМГ-сигнала  $A$ , причем они распределены

нормально с нулевым средним и дисперсией  $\sigma_A$ , функция  $p(x)$  содержит два слагаемых, соответствующих нормальному распределению при  $x \neq 0$  и пику в форме  $\delta$ -функции при  $x=0$  (рис. 1). По своей форме теоретическая функция близка к экспериментально наблюдаемому закону распределения ЭМГ некоторых биологических структур (рис. 2).

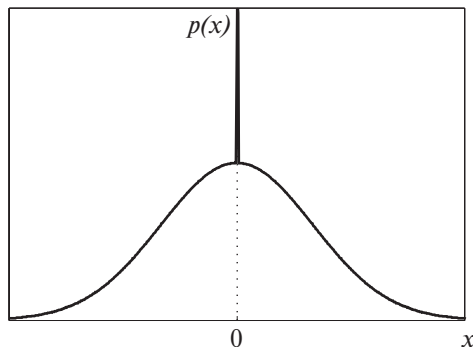


Рис. 1. Функция распределения мгновенных значений ЭМГ

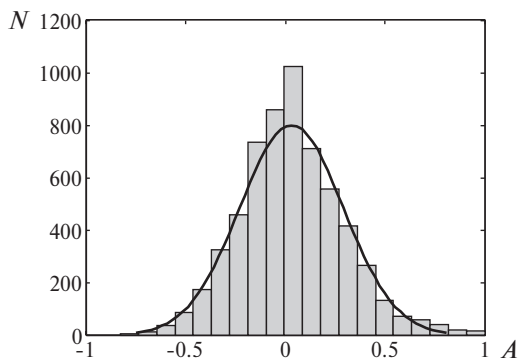


Рис. 2. Экспериментальная гистограмма распределения амплитуд ЭМГ, аппроксимированная нормальным законом распределения

Для решения поставленной задачи были исследованы ЭМГ-сигналы, полученные с мышц нижних конечностей при различной степени нейрорпатологии и в норме (более 80 пациентов). Для получения достоверных статистических характеристик предлагается ЭМГ-сигнал пронормировать относительно максимального значения, исследовать процентное соотношение амплитуд колебаний, составляющих данный сигнал. На рис. 3 приведен фрагмент пронормированного ЭМГ-сигнала. ЭМГ-сигнал представляет собой последовательность случайных колебаний с амплитудами  $A_n$  и длительностями  $T_n$  (рис. 4).

Для построения гистограммы распределения амплитуд ЭМГ-сигнала разбиваем весь диапазон изменения амплитуд нормированной ЭМГ на 20 непересекающихся интервалов, каждому из которых ставим в соответствие количество амплитуд, попавших в этот интервал. По оси абсцисс откладываем значения амплитуд ( $A$ ) для рассчитанных

интервалов, по оси ординат – количество амплитуд в интервале ( $N$ ).

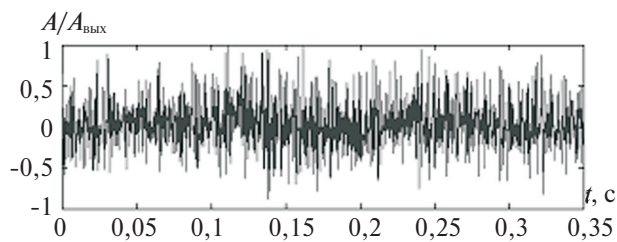


Рис. 3. Фрагмент нормированного ЭМГ-сигнала

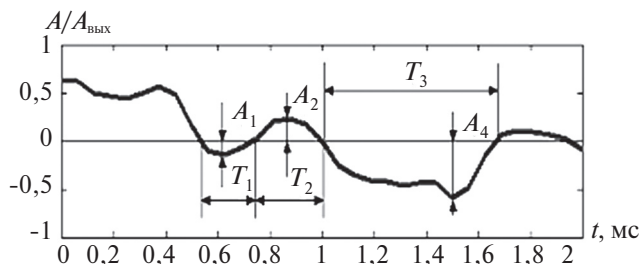


Рис. 4. Пример определения параметров ЭМГ-сигнала

Для оценки формы и параметров функции распределения используем следующие стандартные количественные показатели:

- мода ( $Mo$ ) – наиболее часто встречающееся значение амплитуды, соответствует максимуму гистограммы;
- амплитуда моды ( $AMo$ ) – количество амплитуд, соответствующих значению моды, и  $AMo\%$  –  $AMo$  в процентном отношении к общему их количеству;
- вариационный размах ( $\Delta$ ) – ширина основания гистограммы.

Зарубежные исследователи используют аппроксимацию кривой распределения кардиоинтервалов треугольником и вычисляют так называемый триангулярный индекс – интеграл плотности распределения (общее количество кардиоинтервалов), отнесенный к максимуму плотности распределения [5]. Мы также используем параметры, полученные из аппроксимации гистограммы треугольником:

- триангулярный индекс ( $ТИ$ ) – отношение общего количества амплитуд ЭМГ-сигнала к амплитуде моды ( $AMo$ ), то есть отношение площади треугольника к его высоте;
- параметр наклона ( $H$ ), вычисляемый по следующей формуле:

$$H = \frac{Mo - Min}{Max - Mo},$$

где  $Min$  – минимальное значение амплитуды ЭМГ-сигнала;  $Max$  – максимальное значение амплитуды ЭМГ-сигнала.

Параметр наклона гистограммы ( $H$ ) характеризует направление и степень наклона гистограммы (рис. 5).

Для определения соответствия экспериментальной гистограммы распределения амплитуд ЭМГ-

сигнала нормальному закону вычислим параметр отклонения экспериментальной гистограммы от теоретической  $\Delta S_{\text{норм}}$  (%):

$$\Delta S_{\text{норм}} = \frac{\sum (H_A(X_A) - N(X_A))^2}{\sum N(X_A)^2} \cdot 100\%,$$

где  $H_A$  — значения экспериментальной гистограммы распределения амплитуд ЭМГ сигнала;  $X_A$  — значения амплитуд ЭМГ в интервалах, рассчитанных для экспериментальной гистограммы;  $N$  — значения теоретической нормальной гистограммы.

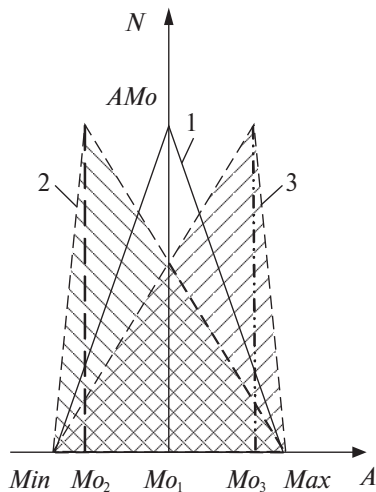


Рис. 5. Аппроксимация гистограммы распределения амплитуд треугольником:

- 1 — параметр наклона  $H = 1$ ;
- 2 — параметр наклона  $H < 1$ ;
- 3 — параметр наклона  $H > 1$

На рис. 6–9 представлены результаты аппроксимации гистограмм распределения амплитуд ЭМГ-сигнала нормальным законом с указанием значения параметра отклонения экспериментальной гистограммы распределения амплитуд ЭМГ-сигнала от нормального закона  $\Delta S_{\text{норм}}$ . На рис. 10 — 13 представлены результаты аппроксимации гистограмм распределения амплитуд ЭМГ-сигнала треугольником с указанием параметра наклона  $H$ .

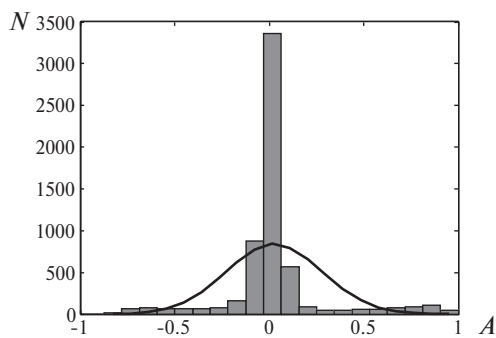


Рис. 6. Аппроксимация нормальным законом гистограммы распределения амплитуд ЭМГ-сигнала с патологией ( $\Delta S_{\text{норм}} = 218,9\%$ )

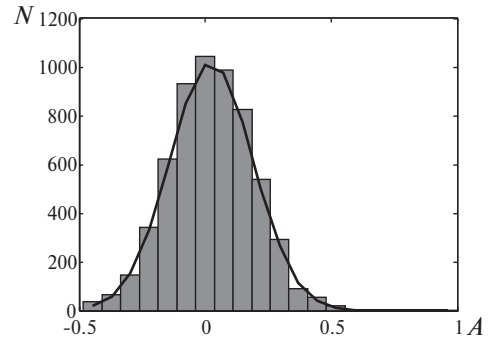


Рис. 7. Аппроксимация нормальным законом гистограммы распределения амплитуд низкочастотного ЭМГ-сигнала ( $\Delta S_{\text{норм}} = 0,39\%$ )

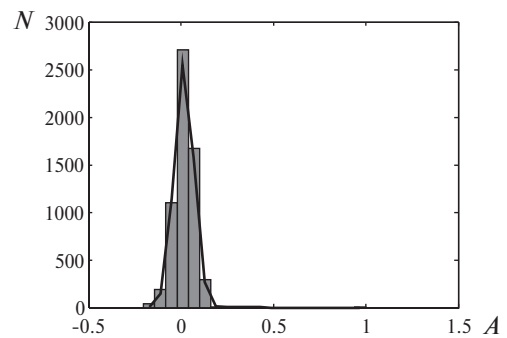


Рис. 8. Аппроксимация нормальным законом гистограммы распределения амплитуд низкочастотного ЭМГ-сигнала ( $\Delta S_{\text{норм}} = 0,38\%$ )

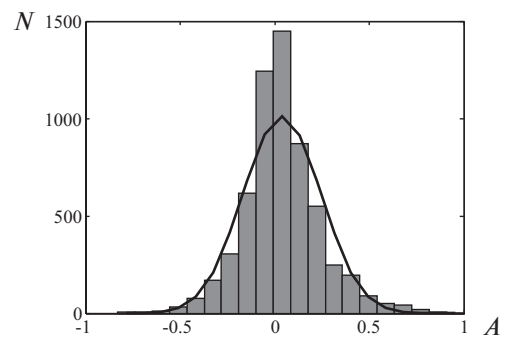


Рис. 9. Аппроксимация нормальным законом гистограммы распределения амплитуд высокочастотного ЭМГ-сигнала ( $\Delta S_{\text{норм}} = 9,01\%$ )

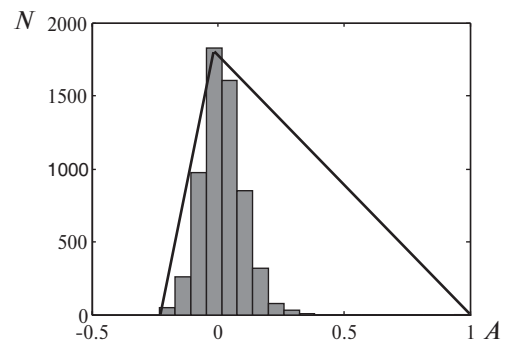


Рис. 10. Аппроксимация треугольником гистограммы распределения амплитуд низкочастотного ЭМГ-сигнала ( $H = 0,21$ )

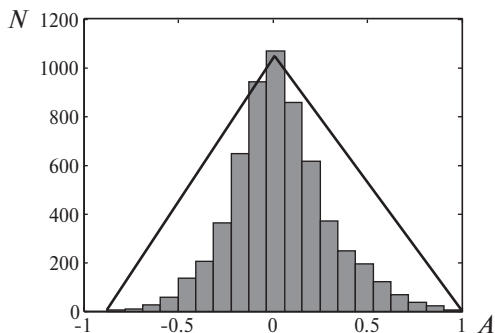


Рис. 11. Аппроксимация треугольником гистограммы распределения амплитуд низкочастотного ЭМГ-сигнала ( $H = 0,9$ )

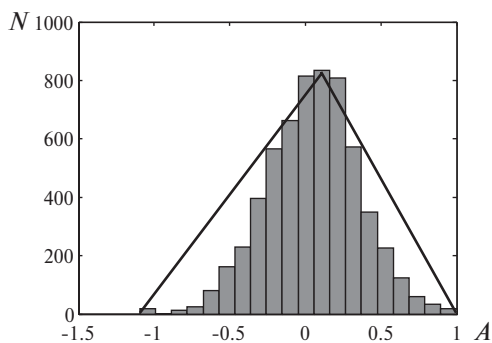


Рис. 12. Аппроксимация треугольником гистограммы распределения амплитуд низкочастотного ЭМГ-сигнала ( $H = 1,35$ )

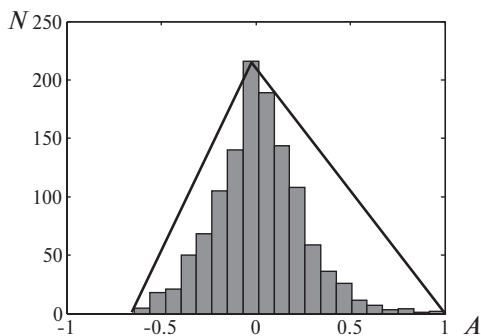


Рис. 13. Аппроксимация треугольником гистограммы распределения амплитуд высокочастотного ЭМГ-сигнала ( $H = 0,6$ )

Для построения интервальной гистограммы, то есть гистограммы распределения длительностей полувольт ЭМГ-сигнала (см. рис. 4), разбиваем весь диапазон изменения длительностей нормированной ЭМГ на 20 непересекающихся интервалов, каждому из которых ставим в соответствие количество длительностей, попавших в этот интервал. По оси абсцисс откладываем значения длительностей ( $T$ ) для рассчитанных интервалов, по оси ординат — количество длительностей в интервале ( $N$ ). Экспериментальный характер гистограммы предлагается аппроксимировать теоретической функцией плотности вероятности для экспоненциального распределения с тем же значением математического ожидания, что и в эксперименте. Для определения отклонения экспериментальной

гистограммы распределения от экспоненциального закона вычислим параметр  $\Delta S_{\text{эксп}}$  (%):

$$\Delta S_{\text{эксп}} = \frac{\sum (H_T(X_T) - E(X_T))^2}{\sum E(X_T)^2} \cdot 100\%,$$

где  $H_T$  — значения экспериментальной гистограммы распределения длительностей полувольт ЭМГ-сигнала;  $X_T$  — значения длительностей полувольт ЭМГ в интервалах, рассчитанных для экспериментальной гистограммы;  $E$  — значения теоретической экспоненциальной гистограммы.

На рис. 14–17 представлены результаты аппроксимации гистограмм распределения длительностей полувольт ЭМГ-сигнала экспоненциальным законом с указанием значения параметра отклонения экспериментальной гистограммы распределения длительностей полувольт ЭМГ-сигнала от экспоненциального закона  $\Delta S_{\text{эксп}}$ .

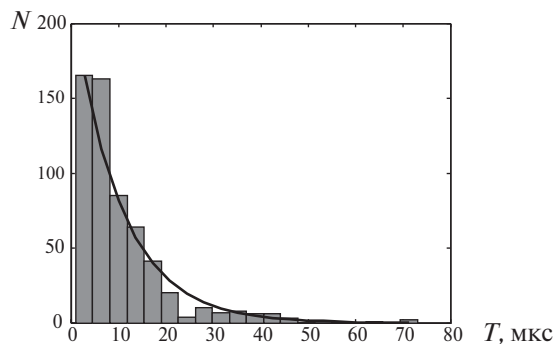


Рис. 14. Аппроксимация экспоненциальным законом распределения гистограммы распределения длительностей полувольт ЭМГ-сигнала с патологией ( $\Delta S_{\text{эксп}} = 4,9\%$ )

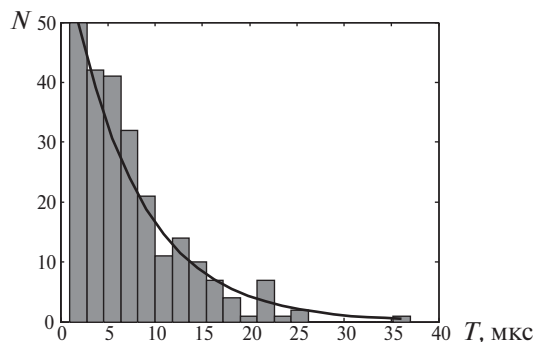


Рис. 15. Аппроксимация экспоненциальным законом распределения гистограммы распределения длительностей полувольт высокочастотного ЭМГ-сигнала ( $\Delta S_{\text{эксп}} = 3,74\%$ )

Некоторые результаты расчета геометрических статистических характеристик представлены в табл. 1. Все исследуемые электромиографические сигналы по результатам статистической обработки были разделены с использованием диагностики по среднеквадратическому отклонению (СКО) на две группы: различная степень патологии нервно-мышечной ткани и условная норма [3].

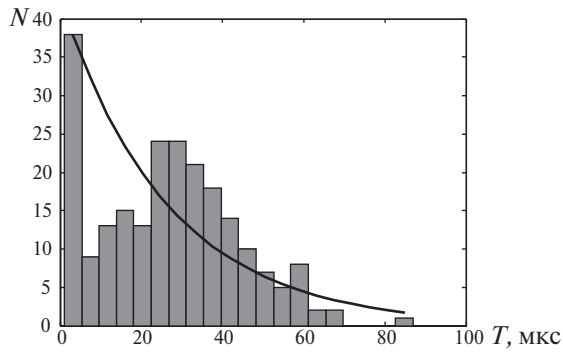


Рис. 16. Аппроксимация экспоненциальным законом распределения гистограммы распределения длительностей полуволн низкочастотного ЭМГ-сигнала ( $\Delta S_{\text{экср}} = 23,38\%$ )

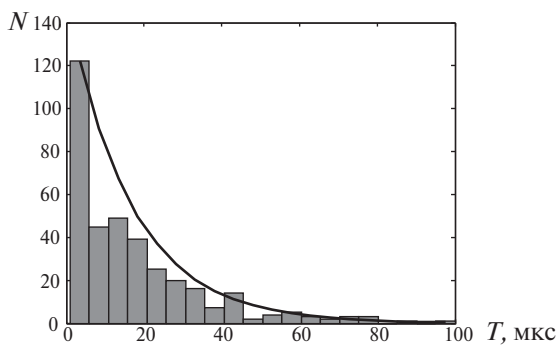


Рис. 17. Аппроксимация экспоненциальным законом распределения гистограммы распределения длительностей полуволн низкочастотного ЭМГ-сигнала ( $\Delta S_{\text{экср}} = 8,64\%$ )

На рис. 18 представлены значения геометрических статистических характеристик следующим образом: в первом столбце маркерами «звездочка» обозначены значения характеристик для пациентов с патологией (*P*), во втором столбце треугольными маркерами — значения характеристик для пациентов с условной нормой (*N*). Из способа разделения исходных сигналов очевидно, что звездочки и треугольники для всех характеристик не группируются в два отдельных класса, как это было для СКО [3]. Изучив результаты данного исследования, можно дифференцировать ЭМГ-сигналы в отдельные группы по плотности не 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 рис. 18, поскольку значения различных характеристик группируются различным образом.

### 3. Перспективы работы

Перспективой работы авторы видят дальнейшее проведение исследований по выявлению из приведенного набора количественных показателей ЭМГ-сигнала информативных параметров для дифференцирования различных патологических состояний нервно-мышечной ткани. Такие пара-

Таблица 1

№	$\Delta$	$Mo$	$AMo$	$AMo\%$	$TI$	$H$	$\Delta S_{\text{норм}}$	$\Delta S_{\text{экср}}$
1	1,61	-0,005	1095	18,25	5,48	0,6	3,96	0,52
2	1,78	0,065	1044	17,4	5,75	0,9	1,78	20,36
3	2,02	0,042	4262	71,03	1,41	1,11	181,48	16,99
4	1,56	0,027	1518	25,3	3,95	0,6	14,09	12,96
5	1,87	0,016	3335	55,58	1,8	0,9	214,86	6,22
6	1,54	0,04	1618	26,97	3,71	0,6	5,05	2,004
7	1,56	0,026	939	15,65	6,39	0,6	1,37	2,44
8	1,88	0,011	3317	55,28	1,81	0,9	70,69	0,33
9	1,86	0,024	1293	21,55	4,64	0,9	3,58	6,92
10	1,68	0,033	1058	17,63	5,67	0,74	3,73	16,86
11	1,61	-0,004	1966	32,77	3,05	0,6	23,15	3,27
12	1,73	0,003	2456	40,93	2,44	0,74	20,75	5,55
13	2,1	0,005	2609	43,48	2,3	1,11	75,24	3,56
14	1,9	0,001	2688	44,8	2,23	0,9	68,28	6,96
15	2,18	0,075	1711	28,52	3,51	1,35	11,84	2,429
16	1,97	-0,036	2160	36	2,78	0,9	32,41	11,76
17	2,14	0,089	1170	19,5	5,13	1,35	2,42	11,06
18	1,59	0,083	1003	16,72	5,98	0,74	4,24	0,156
19	1,82	0,046	2196	36,6	2,73	0,9	25,21	0,16
20	1,51	-0,023	1276	21,27	4,7	0,48	4,03	8,50

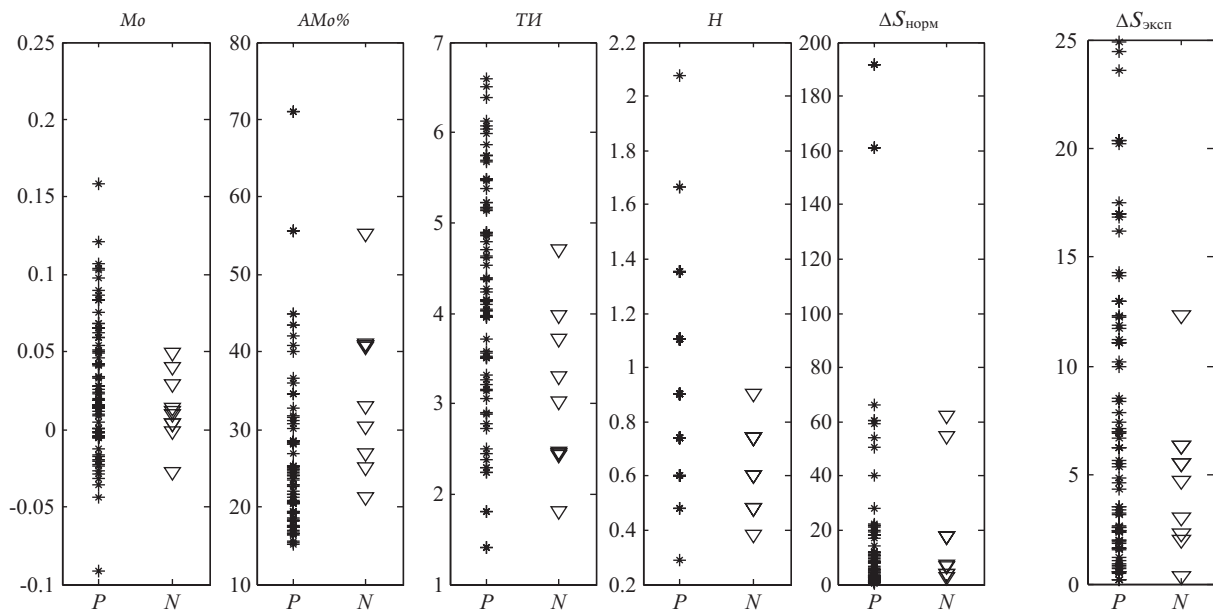


Рис. 18. Группы геометрических статистических характеристик ЭМГ-сигналов (маркер «звездочка» — значения характеристик для пациентов с патологией по СКО (P), треугольный маркер — значения характеристик для пациентов с нормой (N))

метры могут использоваться для создания баз данных пациентов при скрининговых исследованиях и выявления патологических изменений в нервно-мышечной ткани на ранних стадиях.

### 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 — мода  $Mo$ , амплитуда моды  $AMo$  и амплитуда моды в процентном отношении к общему количеству амплитуд  $AMo\%$ , вариационный размах  $\Delta$ , триангулярный индекс  $TI$ , параметр наклона  $H$ , параметр отклонения экспериментальной гистограммы распределения амплитуд ЭМГ-сигнала от нормального закона  $\Delta S_{норм}$ , параметр отклонения экспериментальной гистограммы распре-

ления длительностей полувольт ЭМГ-сигнала от экспоненциального закона  $\Delta S_{эксп}$ . Показано, что данные характеристики являются статистически независимыми и несут дополнительную диагностическую информацию о состоянии нервно-мышечной ткани пациента (табл. 1).

**Список литературы:** 1. *Зенков Л.Р., Ронкин М.А.* Функциональная диагностика нервных болезней. — М.: Медицина, 1991. — 623 с. 2. *Гаусманова-Петрусевич Я.* Мышечные заболевания. — Варшава: Польское гос. мед. изд-во, 1971. — 440 с. 3. *Бых А.И., Жемчужкина Т.В., Носова Т.В.* Поиск информативных количественных показателей электромиографического сигнала. Сообщение 1 // Бионика интеллекта. — 2007. — 1(66). — С. 118-125. 4. *Бакалов В.П.* Методы биотелеметрии. — Л.: Наука, 1983. — 176 с. 5. Вариабельность сердечного ритма. Стандарты измерения, физиологической интерпретации и клинического использования // Вестник аритмологии. — 1999. — № 11. — С. 53-78.

Поступила в редколлегию 10.04.2008

УДК 57.08:632.082



## ГРАФОАНАЛИТИЧЕСКИЕ МЕТОДЫ ОПРЕДЕЛЕНИЯ ПАРАМЕТРОВ НЕОБРАТИМОГО ИМПУЛЬСНОГО ПРОБОЯ МЕМБРАНЫ КЛЕТКИ (часть 2)

В.А. Шигимага

Институт животноводства УААН, Харьков

Электропорация мембран живых клеток импульсным электрическим полем предполагает выбор определенных режимов их обработки. Превышение критической величины поля приводит к электропробое мембраны клетки, что определяется по ее проводимости. Аппроксимация экспериментальной зависимости проводимости клетки от напряженности поля, приложенного к мембране, выполнена полиномом нечетной степени. Значения параметров пробоя мембраны клетки получены графоаналитически как координаты точки максимальной кривизны функции полинома.

КЛЕТКА, МЕМБРАНА, ИМПУЛЬС, ПРОВОДИМОСТЬ, ПОЛИНОМ, ПАРАМЕТРЫ ЭЛЕКТРОПРОБОЯ

### Введение

Электропорация как метод искусственного повышения проницаемости мембраны клетки уже давно и успешно применяется в различных областях науки и технологии: репродуктивной биологии, биофизике, медицине, животноводстве, растениеводстве и традиционно является предметом биотехнологии.

Реакция живых клеток на приложенное извне электрическое поле вызывает большой интерес исследователей уже длительное время [1–6]. В первой части статьи отмечалось, что применение этого поля требует осторожности для выбора правильных электрических режимов обработки клеток. Для этого нужно знать критические и безопасные параметры воздействия на них. В частности, наложение интенсивного электрического поля поперек мембраны клетки, как полагают, ответственно за мембранный разрыв или обратимые мембранные процессы для некоторых типов повреждений клеток ткани в случаях электрической травмы, которые не могут быть отнесены к джоулеву нагреву. Удлиненные клетки, подобные нервным, особенно уязвимы к такому повреждению [5]. В этой работе исследована простая модель цилиндрической клетки, соответствующей нервной или мышечной. Получена система дифференциальных уравнений первого порядка для мембранного электрического поля, которая объединяет произвольную временную зависимость во внешнем поле и нелинейный мембранный отклик (неомическая проводимость). Поведение этой модели исследовано для различных полей в линейных и сильно нелинейных режимах. Показано, что пиковые поля, предсказанные нелинейной моделью, приблизительно вдвое интенсивнее в условиях травмы с разрывом мембраны, чем таковые в линейном случае.

Многokратная электрическая импульсная электропорация (стимуляция) используется также при изучении нервных волокон [7]. Было исследова-

но повреждение (пробой) мембраны как функции параметров импульса и геометрии электрода, используя флуоресцентную метку на хориоаллантоидной оболочке цыпленка *in vivo* и сетчатке *in vitro*. Данные по этой модели были проверены также на свиной сетчатке *in vitro*. Для некоторых размеров электрода плотность тока при повреждении мембраны была независима от них. Порог повреждения плавно уменьшался с ростом числа импульсов, а на больших электродах при плотности тока, вычисленной по половине длительности импульса, явился характеристикой электропорации [7].

Фундаментальные исследования динамики нейронных сетей – необходимость для понимания, как работает нервная система [8]. Формирование и развитие памяти связаны с изменениями в электрической деятельности нейронов. Чтобы понять суть этих изменений, необходимо провести изучение *in vitro* на отдельных нейронах. Для этого разрабатываются специальные системы микроэлектродов, пространственно выстраивающих упорядоченные нейроны путем генерации градиентного поля в диэлектрофоретических ловушках [8, 9]. Необходимо отметить, что параметры поля при этом, безусловно, должны быть значительно ниже пробоя мембраны. Однако этот порог точно не определялся. Тем не менее, была зарегистрирована внеклеточная электрическая активность отдельного нейрона и снят ее спектр с использованием быстрого Фурье-преобразования, чтобы определить нормы всплеска деятельности индивидуальных нейронов [8].

Для нормальной функции центральной нервной системы критическое значение имеет поддержание объема мозговой клетки [10]. В обзоре рассматривается, как проходит регулирование объема клетки, прежде всего, в ответ на изменение осмолярности жидкости тела. Мозговые клетки противостоят тенденции набухать или сжиматься путем соответствующей настройки их внутреннего

осмотического потенциала [10]. Каким же образом клетка регулирует свой объем? Это достигается за счет сброса или ввода неорганических ионов и низкомолекулярных органических веществ, что почти наверняка происходит за счет управляемого гуморально изменения проницаемости мембраны. Отсюда следует замечательная возможность внешнего управления проницаемостью нервных клеток, а значит и некоторыми функциями мозга, посредством обратимой электропорации.

Известно, что высокая напряженность внешнего поля может индуцировать также высокий потенциал снаружи мембраны клетки [1, 2]. При электропорации потенциалы мембраны снаружи и внутри выравниваются, мембранный потенциал становится постоянным из-за перехода ее от состояния изолирования к высокой проводимости и проницаемости [11]. При этом, как полагают, создаются поры, через которые различные молекулы могут транспортироваться внутрь и наружу клетки [6, 11]. За расширением пор или созданием многих мелких пор поперек некоторой части мембранной поверхности может последовать разрыв мембраны. В неоднородных электрических полях трудно предсказать электропорлируемую мембранную область [11]. Поэтому в этой работе индуцированный мембранный потенциал и мембранная область, где этот потенциал превышает пробойный уровень, исследованы путем моделирования. Результаты экспериментов, в которых был обнаружен коллапс нейронных клеток, были сопоставлены с вычисленной напряженностью поля, чтобы исследовать мембранный пробой или разрыв. Обнаружено, что в неоднородном поле разрыв мембраны зависит от ее положения, особенно при более высоких пробойных уровнях. Это указывает на то, что размер мембранного участка, который подвергается электропорации, определяет его разрыв [11].

Ранее предпринимались попытки оценить требуемое напряжение между электродами для адекватного воздействия на клетки мозга, например, при электрошоковой терапии [3]. Показано, что это напряжение в 2–3 раза больше, чем потенциалы мембраны, при которых происходит их электрический пробой. В работе теоретически обоснована связь напряжения между электродами с мембранным потенциалом отдельных нейронов. Оказалось, что для некоторых нейронов вычисленные значения превышают уровень пробоя. Исследование показало, что электрический пробой действительно имеет место при электрошоке. Предположено, что электрический пробой мог бы отчасти объяснить сам электрошок и некоторые его побочные эффекты [3].

Проводились также исследования повреждающего действия электрического поля на клетки через нетепловые механизмы [4]. Предположено, что

для многих видов электрической травмы местное электрическое поле имеет достаточную величину, чтобы вызвать электрический пробой мембран и лизис клетки. Для иллюстрации значения размера клетки и ее ориентации относительно силовых линий поля проанализирована геометрически простая модель вытянутой клетки. Теоретически показано, что большие (удлиненные) клетки, например, нервные, более уязвимы к электрическому пробую [4].

Измерение сопротивления (проводимости) — важный метод для оценки электрических контактов между культивируемыми клетками и подложкой с внедренными микроэлектродами [12]. Это подходит для любого типа клеток, будь то нейроны или нет. Однако правила перевода измеренного сопротивления в параметры успешной стимуляции, то есть установление потенциала действия, не известны. Поэтому предложен метод для обнаружения активных мембранных токов по внеклеточному току стимуляции. Метод базируется на предсказании линейной части отклика на приложенный стимулирующий импульс тока, используя модель импеданса нейрон-электродного контакта. Требуемые параметры модели импеданса получены путем его спектроскопии или непосредственно из измеренных откликов. Обнаружены активные мембранные токи в нелинейном отклике, полученные вычитанием предсказанного линейного отклика из измеренного [12]. Здесь также, безусловно, внешняя импульсная стимуляция должна быть значительно ниже порога пробоя мембраны, что легко предварительно определить по измеряемому току, но в этой работе такое измерение, по видимому, не проводилось.

Интересно применение электропорации в качестве метода загрузки нейронов чувствительным к кальцию красителем с целью изучения миграции этого иона в нейронной сети [13]. Описано использование метода на изолированном спинном мозге новорожденной мыши *in vitro*. Эта техника приводит к быстрой и полной маркировке нейронов для наблюдения за их деятельностью во всех глубинах спинного мозга и определяется положением и размером электродов. Электропорация прерывает местные рефлекс и функцию сети ненадолго (приблизительно на 1 ч), после чего они восстанавливаются [13]. Возможно, было бы корректнее сначала определить более безопасные параметры электропорации для ускорения восстановления жизненных функций.

Вообще говоря, жизнеспособность клетки, подвергнутой действию внешнего поля, должна быть обязательно всесторонне исследована. Мембранный потенциал, индуцированный этим полем, играет критическую роль в сохранении жизнеспособности. Мембрана может быть фактически пред-

ставлена емкостью с параллельно подключенным большим сопротивлением. При высоких индуцированных потенциалах это представление для мембраны больше не действительно. В этой точке пробой мембраны обычно и происходит [9].

Создание мелких отверстий, называемых порами — необходимая предпосылка для истечения медиатора из синаптических пузырьков. Известно, что высокая интенсивность электрического поля может создать поры в пузырьках за счет электропорации [14]. Благодаря скоплению заряженных молекул фосфатидилсерина на внутренней части мембраны, могло бы возникнуть достаточно сильное электрическое поле, способное вызвать самоэлектропорацию синаптического пузырька. В предположении плоской геометрии мембраны показано, что поля, достаточные, чтобы вызвать электропорацию, могут возникнуть в межмембранных зазорах меньше, чем 3 нм. Несмотря на то, что мембрана в целом не плоска, но в том месте, где в пузырьке возникает отверстие, она близка к этому. Обнаружено также, что поля, превышающие порог электропорации, возникают на расстоянии 2 нм или меньше, подтверждая заявление, что самоэлектропорация могла бы быть возможным механизмом для формирования поры [14].

Как следует из анализа доступной литературы, по этому направлению все еще мало аналитических методов определения параметров пробоя мембраны, поэтому и в аспектах научных исследований и в практических приложениях эта задача по-прежнему вызывает интерес. Тем более, что на сегодня пока не существует полного понимания природы электропорации и обусловленной ею дополнительной проницаемости мембраны [6], особенно в неуправляемых режимах пробоя с необратимыми последствиями для мембраны и клетки в целом [9]. Тем не менее, этот запредельный режим с успехом используется в медицине, например, для терапии опухолей мозга [15] или подавления эпилепсии [3].

В то же время есть возможность аналитического подхода к экспериментальным исследованиям электропорации от начальных этапов до необратимого пробоя мембраны по измеренным и вычисленным параметрам проводимости с целью дальнейшего их использования, например, в технологии управления нейрорепарацией или нейротерапии.

Несмотря на краткость этого, далеко неполного обзора, становится очевидной необходимость разработки объективных методов определения параметров электропробоя мембраны клетки.

### 1. Цель и задачи исследования

В качестве ближайшего аналога (прототипа) предлагаемого метода, который описан ниже, выбран метод, описанный в первой части статьи [16].

Напомним вкратце его суть. В капле среды располагают исследуемую клетку между микроэлектродами, подают на них импульсы напряжения с возрастающей от нуля амплитудой и определяют зависимость проводимости клетки от напряженности поля. Полученную зависимость проводимости от напряженности поля аналитически аппроксимируют двумя прямыми и вычисляют координаты точки их пересечения, которые считают значениями параметров электропробоя мембраны клетки.

Однако этот метод не обеспечивает достаточной точности вычисления параметров электропробоя мембраны. Это объясняется особенностью линейной аппроксимации двумя прямыми такой существенно нелинейной функции, как проводимость клетки в возрастающем импульсном электрическом поле. Линейная аппроксимация двумя пересекающимися прямыми двух ветвей этой функции (до и после пробоя) дает несколько смещенные координаты точки предполагаемого электропробоя, а именно, заниженное значение проводимости и завышенное — напряженности. Это происходит потому, что точка пересечения прямых при такой аппроксимации ложится ниже и правее той части кривой проводимости, где она начинает резко расти, то есть там, где ее кривизна максимальна. При малых скоростях роста проводимости клетки в области электропробоя этим смещением координат можно практически пренебречь, но при больших скоростях роста оно становится особенно заметным и уже требует учета. Кроме того, в последнем случае, вследствие отсутствия фиксации, клетка часто начинает либо поворачиваться, либо ее выталкивают из межэлектродного пространства реактивные силы, возникающие вследствие истечения компонентов цитоплазмы сквозь сильно электропорированную мембрану, то есть клетка теряет неподвижность. Тем самым нарушается геометрия эксперимента и его продолжение становится просто некорректным. Поэтому и возникла необходимость разработки более продвинутого метода.

*Цель статьи* — представить второй из разработанных графоаналитических методов определения параметров необратимого электропробоя мембраны клетки, реализованный с помощью более сложной математической модели, которая позволяет использовать также известные программные продукты, но получаемые при этом результаты оказываются точнее и информативнее.

В первой части статьи был рассмотрен простой метод для определения параметров электропробоя и только. В этой части предлагается более сложный метод обработки данных, который позволяет определить не только эти параметры, но и может обеспечить вычисление промежуточных допробойных параметров электропорации. Эти последние как

раз и необходимы для реализации различных практических приложений электропорации, например, загрузки клеток различными веществами с тем, чтобы они остались живыми и полностью функциональными.

## 2. Методика

Проводимость клеток в линейно возрастающем импульсном поле определена с помощью аппаратуры и по методике, которые подробно описаны ранее [17, 18].

Удобной моделью исследования снова послужили довольно крупные и легко доступные одиночные клетки — ооциты (яйцеклетки) мыши. Клетка размещалась в капле среды между микроэлектродами на предметном стекле микроскопа. Микроэлектроды выполнены из золотой проволоки, запаянной в стеклянный капилляр, закреплены в манипуляторах и юстированы соосно в поле зрения инвертированного микроскопа.

Построение графиков по экспериментальным точкам выполнено с помощью программного пакета Microsoft Excel 2000 v.9.0. Подгонка по ним аппроксимирующих уравнений полиномов выполнена с помощью прикладной программы TableCurve v.1.10. Координаты точки, отвечающей необратимому пробое мембраны, получены визуально на графике подогнанного полинома по максимуму кривизны в области резкого роста функции. Ниже в качестве примеров приведены рассчитанные по этому алгоритму параметры электропробоя мембран ооцитов мыши в диэлектрических растворах маннита.

## 3. Результаты и обсуждение

В основу разработанного метода поставлен усовершенствованный способ определения параметров электропробоя мембраны клетки [19]. Суть усовершенствования состоит в нелинейном анализе экспериментальной зависимости ее проводимости от напряженности поля, который проводят путем аппроксимации проводимости полиномом нечетной степени не ниже 5, а затем вычисляют его максимальную кривизну в точке, координаты которой и считают значениями параметров электропробоя [20].

Сущность метода поясняется графиком. На нем изображены экспериментальные точки зависимости проводимости ооцитов мыши от напряженности импульсного поля в 0,3М и 0,4М водных растворах маннита при температуре 25 °С.

На обеих зависимостях приведены также соответствующие кривые аппроксимирующих полиномов, которые в обозначенных на графике точках их максимальной кривизны дают координаты параметров электропробоя мембраны клетки при данной молярности раствора маннита.

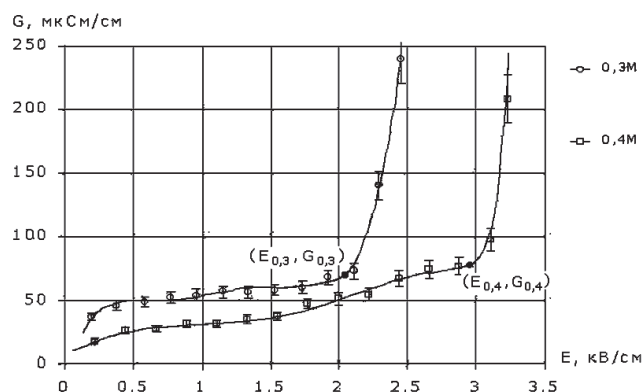


Рис. График зависимости проводимости ооцитов мыши от напряженности импульсного поля в 0,3 и 0,4М манните. Экспериментальные точки сглажены полиномами 7 и 5 степени соответственно

Метод реализуется следующим образом. Между микроэлектродами, погруженными в среду, помещают исследуемую клетку — ооцит. Микроэлектроды слегка поджимают, тем самым фиксируя клетку, и подают на них известное линейно возрастающее импульсное напряжение. По измеренным значениям напряжения и тока через клетку рассчитывают ее проводимость и строят график зависимости проводимости от напряженности. Затем, используя, например, пакет Excel, проводят аппроксимацию полученной зависимости полиномом нечетной степени не ниже 5:

$$G = a + bE + cE^2 + dE^3 + eE^4 + fE^5, \quad (1)$$

где  $G$  и  $E$  — текущие переменные — проводимость клетки и напряженность поля соответственно;  $a, b, c, d, e$  и  $f$  — свободный член и коэффициенты при степенях  $E$  соответственно.

Возможна также аппроксимация полиномом степени 7 и выше в зависимости от требуемого критерия согласия модельной функции с экспериментальными точками. Полиномы четной степени не используют, поскольку они не имеют физического смысла по причине резкого роста вблизи нуля. Далее по известной формуле дифференциальной геометрии определяют кривизну полученного полинома как функцию напряженности поля:

$$K = \frac{G''}{[1 + (G')^2]^{3/2}} = f(E), \quad (2)$$

где  $K$  — текущая кривизна полинома;  $G'$  и  $G''$  — первая и вторая производные полинома соответственно.

Затем проводят анализ функции кривизны полинома  $K = f(E)$  на наличие максимума, то есть выполняют необходимое условие, которое требует равенство нулю первой производной этой функции:

$$K' = f'(E) = 0. \quad (3)$$

Решая соответствующее степенное уравнение, получают несколько корней, из которых выбирают наибольший — абсциссу точки максимальной кривизны полинома, то есть напряженность поля, при которой случается необратимый пробой мембраны клетки. Подставляя полученное значение напряженности пробоя в исходное уравнение полинома, определяют ординату — значение проводимости клетки при пробое мембраны. Вычисленные значения координат максимума кривизны считают началом необратимого разрыва мембраны, то есть параметрами ее электропробоя. Решение уравнения (3) можно получить, например, с помощью программы MathCad, однако, проще и с не худшей точностью можно визуальным образом найти точку максимума кривизны по графику  $K = f(E)$ . Еще проще сразу визуальным образом найти эту точку и непосредственно по графику подогнанного полинома  $G(E)$ , как описано ниже, находясь в среде TableCurve и оправдывая название метода — графоаналитический.

**Пример 1.** Получена экспериментальная зависимость проводимости ооцита мыши от напряженности поля в 0,4М манните. Проведена аппроксимация этой зависимости по уравнению (1), но обратным полиномом 5 степени:

$$G^{-1} = 0,094 - 0,226E + 0,313E^2 - 0,204E^3 + 0,061E^4 - 0,007E^5,$$

где  $G$  и  $E$  — текущие переменные — проводимость клетки и напряженность поля соответственно.

Критерием согласия аппроксимирующего полинома с экспериментом взят коэффициент детерминации, составивший  $R^2 = 0,994$ .

Далее по графику полинома, построенному программой TableCurve, путем перемещения указателя текущих координат визуальным образом подобрана точка максимальной кривизны полинома, координаты которой автоматически отображаются указателем. Таким образом, определенные графоаналитическим методом параметры электропробоя мембраны ооцита мыши в 0,4М манните составили: напряженность поля  $E_{0,4} = 2,92$  кВ/см, проводимость в момент разрыва мембраны  $G_{0,4} = 76,34$  мкСм /см.

**Пример 2.** Получена экспериментальная зависимость проводимости ооцита мыши от напряженности поля в 0,3М манните. Проведена аппроксимация этой зависимости полиномом 7 степени:

$$G = -47 + 888E - 3348E^2 + 6445E^3 - 6792E^4 + 3965E^5 - 1202E^6 + 148E^7,$$

где  $G$  и  $E$  — имеют тот же смысл, что в примере 1.

Коэффициент детерминации эксперимент — модель для этого уравнения составил  $R^2 = 0,997$ .

Далее проведено аналогичное определение координат ( $E_{0,3}, G_{0,3}$ ) точки максимальной кривизны полинома, которое выполнено по такому же при-

нципу, как в примере 1. Получены следующие значения параметров электропробоя мембраны ооцита мыши в 0,3М манните: напряженность поля  $E_{0,3} = 2,03$ кВ/см, проводимость в момент разрыва мембраны  $G_{0,3} = 67,94$ мкСм /см.

По определенным значениям параметров электропробоя хорошо видно влияние молярности среды на электрическую прочность мембраны ооцита, причем, значительно сильнее меняется напряженность пробоя мембраны клетки, чем ее проводимость.

Таким образом, эти примеры показывают, что предложенный метод позволяет корректнее, в соответствии с выбранной моделью полиномиальной аппроксимации, вычислять параметры электропробоя мембраны клетки. Кроме того, метод позволяет определять также и координаты точек промежуточных экстремумов кривизны по графику полинома, которые являются параметрами обратной электропорации.

Преимущество представленного метода состоит в том, что в отличие от прототипа он не требует предварительной условной разбивки экспериментальной нелинейной зависимости на две ветви по характеру роста (слабый и резкий рост проводимости), для того чтобы провести аппроксимацию. Применение предложенного метода возможно непосредственно к таблице экспериментальных данных без каких-либо предварительных предположений о местонахождении точки необратимого электропробоя (вслепую). К тому же, метод обеспечивает надежную фиксацию клетки путем легкого сжатия между микроэлектродами, что оказывается достаточным для компенсации выталкивающих реактивных сил, возникающих при истечении цитоплазмы в результате сильной электропорации мембраны в области пробоя. Данный метод является, по-видимому, универсальным подходом к анализу экспериментальных нелинейных кривых на предмет поиска координат точек резкого изменения скорости роста (падения) функции при равномерном росте аргумента.

При наличии современной измерительной аппаратуры, ПК и соответствующего программного обеспечения электропорацию мембраны вплоть до пробоя можно автоматически контролировать в реальном времени путем непрерывного вычисления кривизны функции  $G(E)$  с ростом напряженности поля. При этом достижение и превышение критического значения напряженности, определенное по максимуму кривизны, определяет необратимый электропробой, а все предыдущие ее значения — разную степень электропорации мембраны, то есть разную плотность пробойных отверстий в ней [2, 6, 11]. Это открывает возможность интеллектуального автоматизированного управления электропорацией мембраны клетки в зависимости от цели

эксперимента — загрузка в клетку веществ разной природы и размеров, ДНК-технологии, клеточная или тканевая терапия с высвобождением части или всей цитоплазмы (лизис) клетки и тому подобное.

### Выводы

1. Предложен метод определения параметров электропробоя мембраны путем аппроксимации полиномом нечетной степени зависимости проводимости клетки от напряженности возрастающего импульсного поля и графоаналитического поиска координат точки максимальной кривизны полинома в области пробоя.

2. Метод позволяет исследовать влияние различных факторов на электрическую прочность мембраны клетки и не требует предварительной оценки массива экспериментальных точек по характеру роста проводимости.

**Список литературы:** 1. *Zimmermann U., Neil G.A.* Electromanipulation of cells. N.Y.: CRC Press, 1996. — 375 p. 2. *Chang D.C., Chassy B.M., Saunders J.A., Sowers A.E.* Guide to Electroporation and Electrofusion. -San Diego.: Academic Press. — 1992. — 581 p. 3. *Johansson B.* Electrical membrane breakdown: a possible mediator of the actions of electroconvulsive therapy // *Med. Hypotheses.* — 1987. — V. 24(4). — P. 313-324. 4. *Lee RC, Kolodney MS.* Electrical injury mechanisms: electrical breakdown of cell membranes // *Plast. Reconstr. Surg.* — 1987. — V. 80(5). — P. 672-679. 5. *Bardos DC, Thompson CJ, Yang YS, Joyner KH.* Nonlinear cell response to strong electric fields // *Phys. Med. Biol.* — 2000. — V. 45(7). — P. 1965-1988. 6. *Teissie J, Golzio M, Rols MP.* Mechanisms of cell membrane electropermeabilization: a minireview of our present (lack of?) knowledge // *Biochim. Biophys. Acta.* — 2005. — V. 1724(3). — P. 270-280. 7. *Butterwick A, Vankov A, Huie P, Freyvert Y, Palanker D.* Tissue damage by pulsed electrical stimulation // *IEEE Trans. Bio-med. Eng.* — 2007. — V. 54(12). — P. 2261-2267. 8. *Prasad S, Zhang X, Yang M, Ni Y, Parpura V, Ozkan CS, Ozkan M.* Separation of individual

neurons using dielectrophoretic alternative current fields // *J. Neurosci. Methods.* — 2004. — 135 (1-2). — P. 79-88. 9. *Heida T.* Electric field-induced effects on neuronal cell biology accompanying dielectrophoretic trapping // *Adv. Anat. Em-bryol. Cell Biol.* — 2003. — V. 173(III-IX). — P. 1-77. 10. *Law R. O.* Regulation of mammalian brain cell volume // *J. Exp. Zoology.* — 2005. — V. 268. — Is2. — P.90-96. 11. *Heida T, Wagenaar JB, Rutten WL, Marani E.* Investigating membrane breakdown of neuronal cells exposed to nonuniform electric fields by finite-element modeling and experiments // *IEEE Trans. Biomed. Eng.* — 2002. — V. 9(10). — P.1195-1203. 12. *Buitenweg JR, Rutten WL, Marani E, Polman SK, Ursum J.* Extracellular detection of active membrane currents in the neuron-electrode interface // *J. Neurosci. Methods.* — 2002. — V. 115 (2). — P. 211-221. 13. *Bonnot A, Mentis GZ, Skoch J, O'Donovan MJ* Electroporation loading of calcium-sensitive dyes into the CNS // *J. Neurophysiol.* — 2005. — V. 93(3). — P. 793-1808. 14. *Luitel P, Schroeter DF, Powell JW.* Self-electroporation as a model for fusion pore formation // *J. Biomol. Struct. Dyn.* — 2007. — V. 24(5). — P. 495-503. 15. *Davalos RV, Mir IL, Rubinsky B.* Tissue ablation with irreversible electroporation // *Ann Biomed Eng.* — 2005. — V. 33(2). — P. 223-231. 16. *Шигимага В.А.* Графоаналитические методы определения параметров необратимого импульсного пробоя мембраны клетки (ч. 1) // *Бионика интеллекта: науч.-техн. журнал.* — Харьков. — 2007. — Вып. 2 (67). — С. 84-87. 17. *Шигимага В.А.* Определение проводимости эмбриональных клеток животных // *Проблемы бионики: Всеукр. науч.-техн. сб.* — Харьков. — 2003. — Вып. 59. — С. 60-64. 18. *Шигимага В.О.* Апаратура для електроліття та вивчення провідності клітин // *Вісник Харків. держ. техн. ун-ту с/г-ва.* — Харків. — 2001. — Вып. 6. — С. 386-389. 19. Пат. кор. мод. № 24210, Україна, G01N 33/483, G01R 27/22. Спосіб визначення параметрів електропробою клітинної мембрани / Шигимага В.О. — 2007. — Бюл. № 9. 20. Пат. кор. мод. № 30272, Україна, G01N 33/483, G01R 27/22. Спосіб визначення параметрів електропробою мембрани клітини за кривизною провідності / Шигимага В.О. — 2008. — Бюл. № 4.

Поступила в редколлегию 11.04.2008

УДК 591.1:681.2



## СПОСОБ ПОСТРОЕНИЯ ИЗМЕРИТЕЛЬНОЙ СИСТЕМЫ ДЛЯ РЕГИСТРАЦИИ МАЛЫХ ИЗМЕНЕНИЙ ТЕМПЕРАТУРЫ

М.Г. Самынина

Институт животноводства УААН, г. Харьков, Украина

Рассмотрена возможность повышения чувствительности измерительной системы для регистрации малых изменений температуры. Для измерения температуры в диапазоне 36–42 °С предложен способ построения схемы термопреобразователя с чувствительностью до сотых долей градуса на базе полупроводниковых термосопротивлений с номиналом около 2 кОм, при котором учитывается максимальная погрешность, возникающая при линеаризации градуировочной кривой, и диапазон измерения сопротивления – от 0 до 2 кОм на измерительном приборе.

ТЕМПЕРАТУРА, ТЕРМОСОПРОТИВЛЕНИЕ, ЧУВСТВИТЕЛЬНОСТЬ, ПОГРЕШНОСТЬ

### Введение

Для разработки кибернетических систем, в частности, искусственного интеллекта, определенную ценность представляет изучение принципов функционирования живых организмов. В зависимости от их реакции на температурные условия окружающей среды различают гомойо- и пойкилотермные организмы. Главный признак гомойотермии – характерный уровень средней температуры тела в пределах приблизительно 35–42 °С. Этот уровень поддерживается у человека и ряда животных в адекватных условиях существования. Наиболее постоянна температура центральной нервной системы. Многие исследователи привыкли считать температурный гомеостаз абсолютным. Тем не менее, хорошо известно, что любая регулирующая система (в частности, техническая) неизбежно порождает колебания регулируемой величины в процессе регуляции. Величина и период этих колебаний характеризуют способ регулирования, его точность, чувствительность и некоторые особенности регулируемой системы в целом. У человека, например, допустимые отклонения температуры мозга в зоне температурного комфорта составляют всего несколько сотых градуса, и изменение температуры на определенную величину осуществляется в соответствии с физиологическими процессами. Превышение некоторой средней температуры мозга всего на 0,06–0,08 °С может вызвать тепловой дисбаланс и ощущение перегрева [1].

За последние 30–40 лет в разных лабораториях мира подробно изучены и описаны свойства центральных и периферических терморцепторов и специфическая сигнализация от них [2–9]. Однако точный количественный смысл этой сигнализации, физиологические цели и принципы ее обработки в центре терморегуляции пока еще не имеют однозначного толкования. В связи с этим большой научный и практический интерес представляет изучение динамики изменения температуры тела у млекопитающих.

Учитывая, что имеющаяся информация недостаточна для исчерпывающего представления об основных количественных принципах деятельности центра терморегуляции в гипоталамусе и всего температурного контроля, делаются попытки решить проблему в целом с помощью математического моделирования. Например Бенцингер [10] теоретически вычислил, что у человека отклонение температуры гипоталамуса в сторону повышения всего на 0,01 °С уже может вызвать активную реакцию выведения тепла. Позже экспериментально удалось установить у человека порог сосудистой реакции на кистях рук в пределах нескольких сотых градуса (0,04–0,05 °С) [11], а также изучить эти реакции с помощью специальной прецизионной термометрии наружного слухового прохода (температура которого отражает колебания температуры барабанной перепонки и, следовательно, температуры мозга) и термометрии пальцев рук. Известное предположение о том, что в конечном счете суммация температурных сигналов в центре терморегуляции организована так, что позволяет «измерять» изменения средней температуры тела, остается лишь предположением на теоретическом уровне. Ввиду сложности системы терморегуляции Блай [12] считает, что принципиальные вопросы работы этой системы следует решать способом «черного ящика», то есть подавая на вход системы количественно идентифицированное воздействие, измерять достаточно точно количественный ответ системы на данное воздействие.

В некоторых случаях изучение взаимоотношений, возникающих в отдельных органах и системах в связи с их сопряженной деятельностью, требует высокой чувствительности измерения температуры. Регистрация малых изменений температур затруднена при значениях, сопоставимых с порогом чувствительности измерительной системы, и как следствие, невозможно наблюдать небольшие, но характерные для данного процесса колебания, в результате чего может возникнуть ложное пред-

ставление о высокой стабильности температуры в данном процессе. В связи с такой высокой точностью работы системы терморегуляции следует обеспечить высокую чувствительность измерительного тракта в целях получения необходимого разрешения по амплитуде измеряемой разности температур.

Цель работы – установить возможность повышения чувствительности измерительной системы на базе полупроводниковых термосопротивлений с отрицательным температурным коэффициентом сопротивления (ТКС) для регистрации малых изменений температуры.

### 1. Особенности измерения температуры с помощью термометров сопротивления

Высокой чувствительностью к изменениям температуры обладают полупроводниковые термосопротивления, преимуществом которых перед другими типами термопреобразователей является большой отрицательный ТКС (в среднем  $40 \cdot 10^{-3} \text{ }^\circ\text{C}^{-1}$ ). Однако зависимость сопротивления термисторов от температуры имеет нелинейный характер. Температуру можно рассчитать по известной формуле [13], что весьма трудоемко. Поэтому целесообразно линеаризировать градуировочную характеристику. Наиболее простыми являются схемы линеаризации с помощью параллельного или последовательного подключения пассивных резисторов с сопротивлениями, практически не зависящими от температуры [13]. Существенный недостаток таких схем заключается в том, что крутизна градуировочной характеристики существенно снижается, а, следовательно, в несколько раз снижается чувствительность схемы и зачастую приближается к чувствительности проволочных термометров сопротивления.

При использовании термисторов определенному диапазону измеряемой температуры соответствует диапазон измерения прибора, измеряющего их сопротивление. Следовательно, номиналы термометров сопротивления и термнезависимых линеаризующих резисторов необходимо подбирать так, чтобы не переключать диапазон измерения сопротивления.

Известно, что при последовательном включении термопар чувствительность измерительного преобразователя увеличивается [14]. В доступной нам литературе нет информации о том, что такой же подход применяется для построения термопреобразователей на базе термосопротивлений. Очевидно, это связано с тем, что при увеличении количества термисторов необходимо переключать диапазон измерения на измерительном приборе, вследствие этого чувствительность схемы уменьшается на порядок.

### 2. Методика исследований

Температурная характеристика терморезистора с отрицательным ТКС описывается уравнением

$$R_T = R_{20} \exp \left[ B \left( \frac{1}{T} - \frac{1}{T_{20}} \right) \right], \quad (1)$$

где  $T$  и  $T_{20}$  – соответственно текущая температура и  $20 \text{ }^\circ\text{C}$  (в градусах Кельвина);  $R_T$  и  $R_{20}$  – соответственно сопротивление при температурах  $T$  и  $T_{20}$ ;  $B$  – константа терморезистора, имеющая размерность К [13].

Для наиболее эффективной линеаризации [15] сопротивление линеаризующего резистора, подключенного параллельно термосопротивлению, определяется с учетом средней температуры заданного диапазона температур по формуле:

$$R_l = R_{TM} \frac{B - 2T_M}{B + 2T_M}, \quad (2)$$

где  $R_l$  – сопротивление линеаризации, не зависящее от температуры, Ом;  $R_{TM}$  – сопротивление схемы терморезисторов, соответствующее среднему значению температуры в заданном диапазоне, Ом;  $T_M$  – средняя температура установленного диапазона, К.

Чувствительность схемы  $E$  соединения терморезисторов можно оценить по изменению сопротивления  $\Delta R$  в диапазоне температур  $\Delta T$ :

$$E = \frac{\Delta R}{\Delta T}.$$

Соответственно, разрешающая способность схемы определяется как величина, обратная чувствительности.

Для анализа различных схем соединения терморезисторов были использованы точечные термосопротивления СТЗ-14В номиналом  $2,2 \text{ кОм} \pm \pm 20\%$  и выбран диапазон измеряемых температур  $36 \div 42 \text{ }^\circ\text{C}$ .

### 3. Результаты исследования

Подобно способу соединения термопар с целью увеличения чувствительности измерительного преобразователя нами рассмотрена возможность подключения нескольких терморезисторов одинакового типа. Аналог такому подходу легко найти в технике и биологии: чем больше сенсоров, тем выше чувствительность. Для построения измерительной системы за основу принято параллельное подключение сопротивления линеаризации и полупроводникового термосопротивления по общепринятой схеме (рис. 1) и определение температуры путем измерения сопротивления измерительным прибором. Для терморезистора СТЗ-14В номиналом  $2,2 \text{ кОм} \pm 20\%$  при расчете по (1) крутизна температурной характеристики в заданном диапазоне температур составляет приблизительно  $40 \pm 5 \text{ Ом}/^\circ\text{C}$ , а при линеаризации сопротивлением,

рассчитанным по (2),  $-4 \pm 1 \text{ Ом}/^\circ\text{C}$ , и при этом минимальное изменение температуры, регистрируемое как изменение сопротивления на  $1 \text{ Ом}$ , составляет  $0,25 \pm 0,05 \text{ }^\circ\text{C}$ .

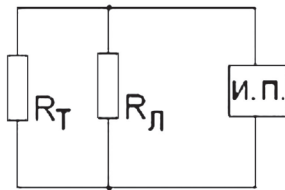


Рис. 1. Схема линейаризации температурной характеристики полупроводниковых терморезисторов с отрицательным ТКС

В связи с тем, что при параллельном подключении термосопротивлений такого же типа чувствительность схемы уменьшается, для анализа возможности повышения чувствительности схемы нами исследовались схемы комбинированного (последовательно-параллельного) и последовательного соединения терморезисторов (рис. 2), температурные характеристики которых представлены на рис. 3.

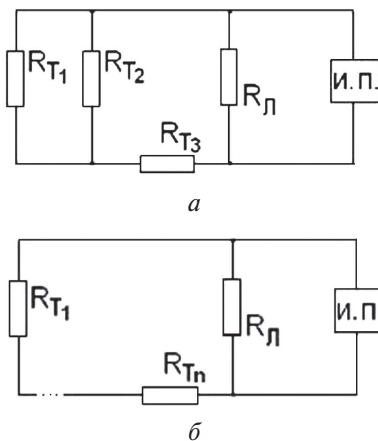


Рис. 2. Схемы соединения терморезисторов: а – параллельно-последовательное соединение; б – последовательное соединение

Чувствительность схемы комбинированного соединения термисторов (рис. 2 а) выше, чем чувствительность схемы, представленной на рис. 1, но ниже, чем у схемы соединения двух терморезисторов по принципу, изображенному на рис. 2 б. Поэтому в дальнейшем исследовании последовательно-параллельное соединение не рассматривалось. Сравнительный анализ температурных характеристик с линейризацией, рассчитанных для одного и нескольких термисторов (рис. 1 и 2 б) сопротивлением, величина которого определялась по (2), показал, что крутизна линейризованной характеристики увеличивается с увеличением количества термосопротивлений практически прямо пропорционально (рис. 3). Для четырех последовательно соединенных терморезисторов она больше при-

близительно в 4 раза. Следует отметить, что при таком способе линейризации для данного диапазона измеряемых температур с увеличением количества последовательно соединенных терморезисторов нет необходимости переключать диапазон на измерительном приборе. Однако для пяти последовательно соединенных термосопротивлений по схеме рис. 2 б температурная характеристика выходит за установленный диапазон, поэтому для последующего анализа было выбрано количество терморезисторов не более 4.

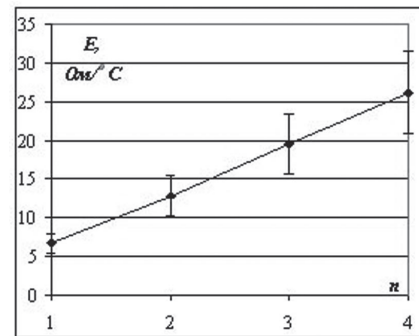
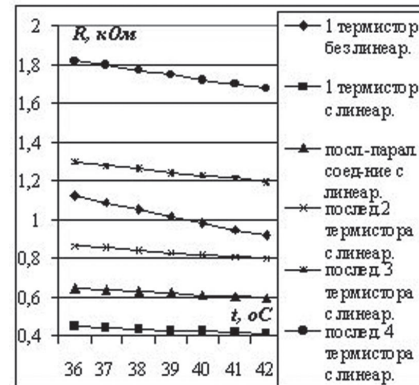


Рис. 3. Температурные характеристики анализируемых схем соединения терморезисторов

Известно, что лишь до некоторой степени можно линейризовать температурную характеристику одного термосопротивления путем удачного подбора термонеинзависимого сопротивления, так как полученная в результате линейризации характеристика имеет некоторую точку перегиба [13]. При оценке погрешности линейризации, возникающей вследствие отличия градуировочной характеристики от прямой, установлено, что ее максимальное значение для схем терморезисторов (рис. 2) в заданном диапазоне измеряемых температур существенно не отличается, достаточно мало и им можно пренебречь (рис. 4). Таким образом определение значения линейризирующего сопротивления по (2) возможно и для схем последовательного соединения нескольких термисторов.

Следует подчеркнуть, что номинал линейризирующего резистора определяет чувствительность схемы, поэтому для ее увеличения нами рассмотрена возможность выбора значения термонеинзависи-

мого сопротивления большего, чем было определено по (2). Отметим, что с увеличением крутизны градуировочной кривой схемы (рис. 5) одновременно увеличивается ее кривизна, и вследствие этого погрешность линеаризации может оказаться существенной. Например для схемы из трех последовательно соединенных терморезисторов при величине линеаризующего сопротивления, определенной по (2),  $R_l = 2,467$  кОм, чувствительность схемы составила  $17 \text{ Ом}/^\circ\text{C}$ , а максимальное значение погрешности линеаризации  $\Delta t = 0,001^\circ\text{C}$ , то есть в 50 раз меньше разрешающей способности схемы. При  $R_l = 3,1$  кОм чувствительность схемы увеличилась до  $25 \text{ Ом}/^\circ\text{C}$ , а  $\Delta t$  — до  $0,017^\circ\text{C}$  (приблизительно в 2,4 раза меньше разрешающей способности схемы). Следовательно, соотношение максимального значения погрешности линеаризации и минимальное изменение температуры, которое можно различить с помощью измерительного прибора, является определяющим для выбора максимально допустимого значения линеаризующего сопротивления. Приняв, что погрешность линеаризации должна составлять не более десятой части разрешающей способности схемы, нами установлено, что с помощью схемы четырех последовательно соединенных терморезисторов и сопротивления линеаризации  $3,064$  кОм можно контролировать изменение температуры на  $0,04^\circ\text{C}$  в диапазоне  $36 \div 42^\circ\text{C}$ .

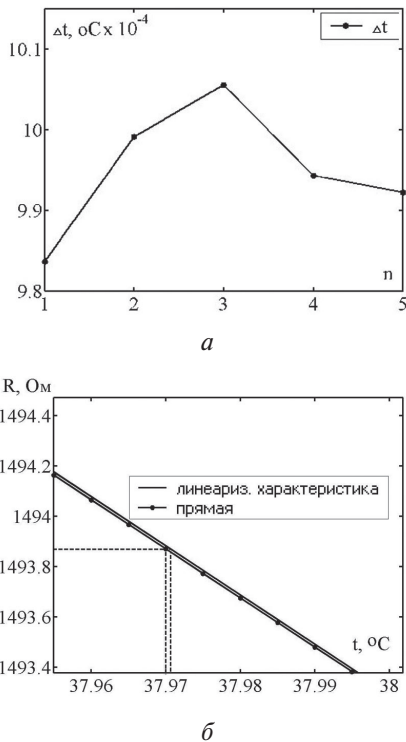


Рис. 4. Максимальная погрешность линеаризации температурной характеристики схем, состоящих из разного количества термосопротивлений с отрицательным ТКС: *a* —  $n$  соответствует количеству терморезисторов; *б* — для трех последовательно соединенных терморезисторов

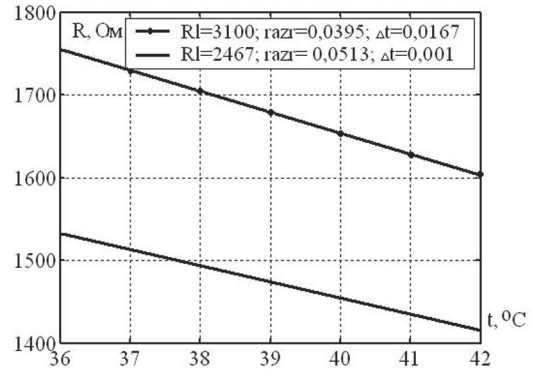


Рис. 5. Линеаризованные температурные характеристики трех последовательно соединенных термисторов при значениях линеаризующего сопротивления 2,467 и 3,1 кОм

Экспериментальная проверка сделанных выводов показала соответствие результатам проведенных расчетов для схемы четырех последовательно соединенных термисторов (рис. 6) при выборе значения линеаризующего сопротивления большего, чем было определено по (2).

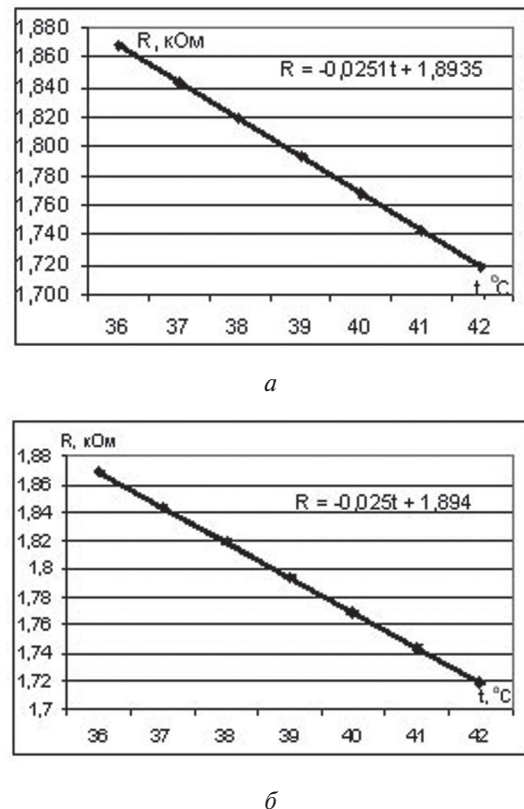


Рис. 6. Температурные характеристики схемы четырех последовательно соединенных термисторов, полученные: *a* — расчетным путем; *б* — экспериментально

Предложенный способ построения измерительной системы для регистрации малых изменений температуры может быть использован для построения высокочувствительных термоприемников на основе полупроводниковых термометров сопротивления, возможно даже в интегральном

исполнении. Подобные схемы имеют важное преимущество — не нужен дополнительный источник питания кроме того, что имеется в измерительном приборе. С другой стороны, такой подход может иметь научное значение и найти применение для моделирования на базе ЭВМ чувственного восприятия и работы системы терморегуляции у млекопитающих.

### Выводы

Таким образом, нами установлено, что для измерения температуры в диапазоне  $36 \div 42$  °С существует способ построения измерительной системы с чувствительностью до сотых долей градуса на базе полупроводниковых термосопротивлений СТЗ-14В номиналом  $2,2 \text{ кОм} \pm 20\%$ . Кроме того, допустимое соотношение максимальной погрешности линеаризации к значению разрешающей способности этой схемы может учитываться как фактор, определяющий ее максимальную чувствительность. Для диапазона измерения —  $0 \div 2$  кОм на измерительном приборе с помощью четырех последовательно соединенных терморезисторов данного типа нами достигнута минимально различимая величина изменения температуры, составляющая  $0,04$  °С и менее.

**Список литературы:** 1. *Иванов К.П.* Основы энергетики организма: Теоретические и практические аспекты. Т. 1.

Общая энергетика, теплообмен и терморегуляция. — Л.: Наука, — 1990. — 307 с. 2. *Слоним А. Д.* Эволюция терморегуляции. — Л., 1986. 3. *Иванов К. П.* Физиология терморегуляции: Руководство по физиологии. — Л.: Наука, 1984. — 470 с. 4. *Stevenson P.* Thermoregulation. — Medical Sciences A&P, 1998. — 296 p. 5. *Бачеригов А. Н., Кузьминов В. Н., Ткаченко Т. В., Назарчук А. Г.* Современные представления о системе терморегуляции // Вісник психіатрії та психофармакотерапії. — 2006. — № 1. — С. 178-182. 6. *Blatteis C. M.* Physiology and pathophysiology of temperature regulation. — 1998. — 128 p. 7. *Hensel H.* Neural processes in thermoregulation // Physiology Review. — 1973. — Vol. 53, № 4. — P. 948-1017. 8. *Houdas Y., Ring E. F.* Human body temperature. Its measurement and regulation. — New York: Plenum Press, 1982. — 224 p. 9. *Simon E. et al.* Central and peripheral thermal control // Physiology Review. — 1986. — Vol. 66. — P. 235-300. 10. *Benzinger T.H.* Heat regulation: Homeostasis of central temperature in man // Physiol. Rev. 1969. Vol. 49, N 4, P. 671-757. 11. *Иванов К.П., Слечук Н.А.* О чувствительности и точности работы физиологического термостата человека // ДАН СССР. 1985. Т. 281, № 3. С. 753—757. 12. *Bligh J.* Cells, cell-talk and mammalian homeothermy // Thermoreception and temperature regulation / Ed. J. Bligh, K. Voigt. Berlin; New York; London. 1990. P. 163-175. 13. *Виглеб Г.* Датчики: Пер. с нем. — М.: Мир, 1989. — 196 с. 14. *Герашенко О.А., Федоров В.Г.* Тепловые и температурные измерения / Справочное руководство. — К.: «Наукова думка», — 1965. — 304 с. 15. *Мэклин Э.Д.* Терморезисторы: Пер. с англ. — М. 1983.

Поступила в редколлегию 14.04.2008



ях будут принимать единичные значения. Вторая группа компонент такова, что (при том же  $B$ ) во всех решениях они будут сохранять нулевые значения. Наконец, компоненты третьей группы могут принимать нулевые значения, единичные или как нулевые, так и единичные значения, но не произвольно, о чем речь будет идти в последующих параграфах. Компоненты этой группы будем называть неоднозначными.

Из всех возможных решений системы (1) выделим два, именуя их 0-решением и I-решением. Первое характеризуется тем, что всем неоднозначным переменным присваиваются нулевые значения, а во втором, наоборот, присваиваются единичные значения. Получить эти решения можно методом подстановок, используя два правила. Правило правого поглощения состоит в том, что выражение вида  $x = ax \vee b$  заменяется на  $x=b$ , а правило правого исключения – в том, что выражение вида  $x = ax \vee b$  позволяет заменить на  $x = a \vee b$ .

Использование только правила правого поглощения приводит к 0-решению. В рассматриваемом примере это будет  $x_1=b_1, x_2=a_{21}b_1$ . Наоборот, используя только правило правого исключения, приходим к I-решению. В рассматриваемом примере это будет:

$$x_1 = a_{12}a_{21} \vee b_1, x_2 = a_{12}a_{21} \vee a_{21}b_1.$$

Используя подстановки в (1) сначала значений вектора  $B$ , а затем последовательно определяемых значений некоторых переменных, в конечном итоге можно прийти к конкретному 0-решению. Нас же будут интересовать 0-решения и 1-решения в виде их алгебраических выражений как функций от значений коэффициентов  $a_{ij}$  и вектора  $B$ . Для этого потребуется исследование системы (1) при  $b_i=0, i=1, n$ , которую мы будем называть дизъюнктивной однородной.

Далее будем пользоваться обозначениями:

$\tilde{X}$  – частное решение системы (1) при фиксированном векторе  $B$ ,

$X^0$  – 0-решение системы (1) при том же  $B$ ,

$X^1$  – I-решение системы (1) при том же  $B$ ,

$\tilde{X}^0$  – частное решение однородной системы,

$X_0^1$  – I-решение однородной системы.

Матрицу  $A$  сопоставим булев определитель ( $B$ -определитель)  $|A|$ , вычисляемый по той же схеме, что и при решении алгебраических линейных систем уравнений, но вместо операций сложения и умножения будем использовать операции дизъюнкции и конъюнкции соответственно, не приписывая при этом, как и в перманенте, знаков получаемым членам. Например:

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} \vee a_{21}a_{12}.$$

В общем виде  $B$  – определитель находится так:

$$|A| = \bigcup_{i,j,\dots,q} a_{1i}a_{2j}\dots a_{nq}. \quad (2)$$

В этом выражении «суммирование» производится по всем перестановкам  $i, j, \dots, q$  индексов столбцов  $1, 2, \dots, n$  матрицы  $A$ . Дополнением или адьюнктом элемента  $a_{ij}$  матрицы  $A$  будем называть определитель  $|A_{ij}|$  матрицы  $A_{ij}$ , получаемой из матрицы  $A$  путем вычеркивания  $i$ -ой строки и  $j$ -го столбца. В матрице  $A$  все элементы главной диагонали заменим на единицы. Полученную матрицу будем обозначать  $A^{(1)}$ . Соответственно, адьюнкт элемента  $a_{ij}^{(1)}$  матрицы  $A^{(1)}$  будем обозначать через  $|A_{ij}^{(1)}|$ .

Сформулируем и докажем следующую теорему.

**Теорема 1.** Система дизъюнктивных уравнений (1) имеет единственное 0-решение  $X^0$ :

$$x_i^0 = \bigcup_{j=1}^n b_j |A_{ij}^{(1)}| = |A_i^{(1)}|, i = \overline{1, n}. \quad (3)$$

Доказательство. Покажем, что формула (3) дает решение системы (1), например, для первого ее уравнения. Очевидно, что при  $b_1=1$  из (1) следует и  $x_1=1$ . Это же следует и из (3). Поэтому проверить (3) достаточно при  $b_1=0$ . Тогда, подставив (3) в систему (1), получим:

$$x_1^0 = \bigcup_{j=1}^n a_{1j} \left( \bigcup_{i=1}^n b_i |A_{ij}^{(1)}| \right) = \bigcup_{i=1}^n b_i \left( \bigcup_{j=1}^n a_{ij} |A_{ij}^{(1)}| \right).$$

Но в силу пятого свойства булева определителя

$$\bigcup_{j=1}^n a_{ij} |A_{ij}^{(1)}| = |A_{i1}^{(1)}|,$$

откуда и получим ожидаемое равенство  $x_1^0 = x_1^0$ .

То, что (3) есть 0 – решение, показано ниже. Единственность 0-решения вытекает из его определения. Решение (3) является частным, ибо из приведенного доказательства не следует, что других решений нет.

Из (3) следует метод решения дизъюнктивной системы уравнений почти по правилу Крамера: для нахождения  $i$ -ой компоненты 0-решения в матрицу  $A^{(1)}$ , полученную из матрицы  $A$ , следует подставить вместо  $i$ -го столбца вектор  $B$  и вычислить определитель образовавшейся матрицы  $A_i^{(1)}$ , определитель  $|A_i^{(1)}|$  в дальнейшем будем воспринимать либо как формулу для вычисления компоненты  $x_i^0$ , либо как ее значение.

**Пример 1.** Пусть

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, B = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix},$$

тогда

$$x_1^0 = \begin{vmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{vmatrix} = 0, x_2^0 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{vmatrix} = 1,$$

и

$$x_3^0 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{vmatrix} = 1, x_4^0 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{vmatrix} = 1$$

0-решением будет вектор  $X^0 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}$ .

## 2. Графовая интерпретация систем логических уравнений

Сопоставим каждой переменной  $x_i$  системы (1) вершину полного ориентированного графа  $G_n$ , наделив ее тем же именем, а всем коэффициентам  $a_{ij}$  сопоставим дуги с именами  $a_{ij}$ , заходящие в  $x_i$  из  $x_j$ . Независимым переменным  $b_i$  также сопоставим вершины-истоки с теми же именами, соединив их с вершинами  $x_i$  графа  $G_n$ , исходящими из  $b_i$  дугами с единичными значениями. Граф  $G$ , получаемый из построенного исключением всех дуг  $a_{ij}$ , соответствующих нулевым значениям соответствующих коэффициентов  $a_{ij}$ , назовем графом, интерпретирующим систему (1).

Согласно (3) вектор  $X^0$  представим столбцом формул вида  $x_i^0 = F_i(A^{(1)}, B), i = \overline{1, n}$ , выражающих его компоненты  $x_i^0$  через имена компонент вектора  $B$  и имена элементов матрицы  $A^{(1)}$ . Интерпретируем формулы  $F_i$  и  $G$ .

Сначала дадим интерпретацию определителя  $|A^{(1)}|$  на  $G_n$ . Будем рассматривать члены этого определителя, зафиксировав в (2) первые индексы коэффициентов в порядке  $1, 2, \dots, n$ . Тогда вторые индексы суть все перестановки на  $1, 2, \dots, n$ . А это значит, что каждый член определителя будет представлен одним или несколькими циклами в перестановке, быть может, включающей единичные циклы, соответствующие единицам главной диагонали в матрице  $A^{(1)}$ . Лишь один член определителя (назовем его диагональным) будет представлен только единичными циклами.

Для каждого члена  $\alpha$ , содержащего в индексах более одного неединичного цикла, найдется член  $\beta$ , содержащий кроме единичных лишь один из этих циклов. Будем говорить, что член  $\beta$  покрывает (в смысле  $\beta \vee \alpha = \beta$ ) член  $\alpha$ . Тогда можно считать, что  $|A^{(1)}|$  представим одним диагональным членом и членами, содержащими в индексах по одному неединичному циклу. Хотя диагональный член покрывает все остальные, в дальнейших рассуждениях нам будет удобно пользоваться именно таким представлением  $|A^{(1)}|$ .

Каждому неединичному циклу, содержащемуся в индексах какого-либо члена  $|A^{(1)}|$ , в графе  $G_n$  соответствует замкнутый непересекающийся путь (также именуемый циклом), быть может, содержащий дуги с нулевыми значениями, а каждому единичному циклу – петля. Объединение всех таких циклов и петель и есть интерпретация  $|A^{(1)}|$  на  $G_n$ .

Воспользуемся построенной конструкцией (объединением циклов и петель) и трактовкой  $|A^{(1)}|$  на  $G_n$  для интерпретации  $|A_i^{(1)}|$  на  $G$ . Рассуж-

дения проведем для  $i=1$  и распространим их на все остальные значения индексов  $i$ . Заменяем, как это и требуется по определению  $|A_i^{(1)}|$ , первый столбец в матрице  $A^{(1)}$  на вектор  $B$ .  $B$  – определитель полученной матрицы  $A_i^{(1)}$ , будет отличаться от  $|A^{(1)}|$  тем, что в составляющих его членах коэффициенты  $a_{j1}$  заменятся на  $b_j$ . При этом диагональный член можно будет представить как  $b_1$ , который покроет все члены, не содержащие  $b_j, j \neq 1$ , а поэтому их можно исключить из рассмотрения. На графе  $G_n$  рассмотренным изменениям в  $|A^{(1)}|$  соответствуют, во-первых, исключение всех дуг и петель, исходящих из вершины  $x_1$  (исключен 1-й столбец), во-вторых, добавление к  $G_n$  вершин-истоков  $b_j, j = \overline{1, n}$  (добавлен столбец  $B$ ) связываемые единичными дугами, соответственно с вершинами  $x_j$ , и в-третьих, исключение всех замкнутых путей (циклов), не проходящих через вершину  $x_1$  в силу покрытия соответствующих членов диагональных членов компонентой  $b_1$  вектора  $B$ . А это значит, что все упоминавшиеся при интерпретации  $|A^{(1)}|$  замкнутые пути на  $G_n$ , проходящие через вершину  $x_1$ , преобразуются в разомкнутые пути, исходящие из вершин  $b_j$  и заканчивающиеся вершиной  $x_1$ . Исключив из этого множества разомкнутых путей те, которые содержат хотя бы одну дугу с нулевым значением, или начинаются истоком с нулевым значением, получим интерпретацию  $F_1$  на  $G$ , а множество истоков, вошедших в объединение оставшихся путей, обозначим как  $B_1$ .

Аналогичные построения проведем и для остальных значений индекса  $i$ . Так как  $F_i = |A_i^{(1)}|$ , то решение системы (1) можно представить в виде

$$x_i^0 = \bigcup_{b_j \in B_i} b_j, B_i \subset B, i = \overline{1, n}, \quad (4)$$

где  $B_i$  – множество единичных истоков, из которых пути в  $x_i$  на графе  $G$ ,  $B$  – множество всех истоков и  $x_i=0$ , если  $B_i = \emptyset$ .

Таким образом, интерпретацией  $F_i$  на  $G$  является его подграф, состоящий из всех ненулевых путей, ведущих от ненулевых истоков в вершину  $x_i$ . Объединение всех таких подграфов для  $i = \overline{1, n}$  образует подграф  $G^0$ , интерпретирующий решение (3) на графе  $G$ .

Если с ненулевыми истоками связать окраску в единичный цвет, то решение (3) можно трактовать как окрашенную часть графа  $G$ , в котором эту окраску принимают все пути (вместе с вершинами), исходящие из окрашенных истоков и заканчивающиеся в вершинах без исходящих дуг. Это значит, что выделенное подмножество окрашенных вершин при заданном  $B$  будет окрашено в единичный цвет независимо от того, как будут окрашиваться остальные вершины, и во всех остальных решениях  $\tilde{x}$ . А это значит, что (3) есть 0-решение по определению.

Изложенное позволяет сформулировать следующее

*Свойство 1.* Среди всех решений системы (1) 0-решение содержит максимальное число нулевых компонент.

Это свойство эквивалентно определению 0-решения.

### 3. Общее решение дизъюнктивной системы

Ранее уже отмечалось, что система (1) может иметь несколько частных решений  $\tilde{X}$ . В силу свойства 1 различие в частных решениях будет определяться компонентами вектора, принявшими в 0-решении нулевые значения.

Для получения общего решения системы дизъюнктивных уравнений построим специальный алгоритм, позволяющий получить любое частное решение. Одновременно такой алгоритм позволяет выявить и структуру общего решения системы (1). Для этого воспользуемся графовой интерпретацией системы (1). Для выявления структуры общего решения можно поступить двояко. Первый путь состоит в исключении из интерпретирующего графа  $G$  подграфа  $G^0$ , интерпретирующего 0-решение, с последующим исследованием оставшейся части графа. Второй путь состоит в исключении из  $G$  только вершин-истоков  $b_i$ ,  $i = \overline{1, n}$  вместе с исходящими дугами с последующим изучением оставшегося графа  $G_0$ , соответствующего однородной системе (1). Пойдем по второму пути.

Если первопричиной присвоения единичных значений переменным в 0-решении были истоки с ненулевыми значениями, то таковых нет в графе  $G_0$ . Поэтому источниками «единиц» в любом решении однородной системы могут быть только петли или замкнутые пути в графе  $G_0$ . Действительно, наличие петли в вершине  $x_i$  означает, что  $i$ -е уравнение в однородной системе (1) имеет вид  $x_i = x_i \vee f_i(x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ , из которого следует, что в нем и во всех остальных уравнениях можно положить  $x_i = 1$ , и это не приведет к противоречию ни в  $i$ -ом уравнении, ни во всех остальных. Единичные значения можно присвоить всем вершинам (переменным) какого-либо замкнутого пути в графе  $G_0$ , и это также не приведет к противоречиям ни в одном из уравнений. Заметим, что в замкнутом пути либо всем переменным следует присвоить единичные значения, либо ни одной.

Пересекающиеся замкнутые пути (включая петли) образуют одну из сильносвязанных компонент в графе  $G_0$ , а это значит, что все вершины этой компоненты в смысле решения однородной системы образуют некоторый класс эквивалентности (назовем его  $K$ -классом). Такие вершины можно стянуть в одну, наделив ее петлей, дабы иметь возможность непротиворечиво наделять ее единичным значением. Заменив все указанные сильносвязанные компоненты графа  $G_0$  на вершинах

с петлями, получим ациклический и, быть может, несвязный так называемый конденсированный граф  $G_c$ . Все вершины графа  $G_c$ , имеющие петли, назовем свободными в  $G_c$ . Вершины, вошедшие в соответствующие  $K$ -классы, будем также называть свободными, но в  $G$ . Этим вершинам независимым образом можно присваивать единичные значения, и это не будет приводить к противоречиям в однородной системе (1).

В графе  $G_c$  все или некоторые вершины могут находиться в отношении следования (частичного порядка), причем, если какая-либо из вершин примет единичное значение, то и все ее последователи также должны принять единичные значения. А так как единичные значения (окраску) можно присваивать независимо любым свободным его вершинам, то всякой начальной окраске свободных вершин будет соответствовать окрашивание некоторого подграфа в  $G_c$ . Присвоив единичные значения всем переменным, сопоставленным окрашенным вершинам, «раскрыв»  $K$ -классы, при условии, что все остальные переменные наделены нулевыми значениями, сформируем вектор  $\tilde{X}_0$ , соответствующий частному решению однородной системы.

Множество найденных таким путем всех частных решений однородной системы, включая нулевое, составляет ее общее решение. Других решений однородная система не имеет.

Действительно, найденные частные решения можно различать по единичным компонентам или, что одно и то же, по окрашенной части графа  $G_0$ , интерпретирующей  $\tilde{X}_0$ . Отличительной чертой этих решений является наличие у любой окрашенной вершины петли или заходящей дуги из уже окрашенной вершины. В силу свободы присвоения единичных значений вершинам с петлями в графе  $G_c$ , перечислены все варианты таких решений. Если же имеются другие решения однородной системы, то они должны содержать либо изолированную вершину без петли и наделенную единичным значением, либо вершину с единичным значением, но связанную заходящими дугами только с вершинами, наделенными нулевыми значениями. Но то и другое противоречит однородной системе (1).

Очевидно, что объединение 0-решения системы (1) и любого частного решения однородной системы (1) будет решением неоднородной системы (1). Интерпретация 0-решения и решения однородной системы (1) на графе  $G_0$ , а также аналогично строящаяся интерпретация любого частного решения  $\tilde{X}$  на  $G$  позволяет сформулировать общее для них.

*Свойство 2.* Если  $\tilde{X}$  – решение системы (1), то любая вершина (компонента вектора  $\tilde{X}$ ), принявшая единичное значение, должна:

1 – либо иметь петлю;

2 – либо находиться в замкнутом пути, где всем вершинам присвоены единичные значения;

3 – либо быть связанной заходящим путем с вершиной, которая уже наделена единичным значением в силу 1 или 2;

4 – либо быть связанной заходящим путем с какой-либо вершиной-истоком, наделенной единичным значением.

Верно и обратное.

Установленные свойства позволяют доказать следующую теорему.

**Теорема 2.** Любое частное решение системы (1) может быть представлено покомпонентной дизъюнкцией ее 0-решений и одного из решений соответствующей однородной системы (1).

Если в формулировке теоремы 2 упоминаемое 0-решение заменить на любое другое фиксированное частное решение системы (1), то аналогичное утверждение не будет верным. В этом состоит одно из отличий систем логических уравнений от линейных алгебраических систем.

**Доказательство.** Пусть  $X$  – решение неоднородной системы (1), не совпадающее с 0-решением. Тогда оно должно отличаться от 0-решения хотя бы одной единичной компонентой (вершиной) вне подграфа  $G^0$ . Обозначим множество всех таких вершин через  $W$ . Это множество удовлетворяет свойству 2, так как вершины из  $W$  связаны с  $G^0$  только исходящими дугами, а поэтому в  $W$  отсутствуют вершины, связанные с вершинами-истоками. Применяя к вершинам из  $W$  свойство 2 (сначала прямое утверждение для неоднородной системы, а затем обратное – для однородной системы), получим, что  $W$  порождает конкретное решение однородной системы  $\tilde{X}$ . Это решение строится из  $W$  добавлением к нему вершин из подграфа  $G^0$ , связанных с ним заходящими путями, которые также наделяются единичными значениями. В силу свойства 2 в 0-решении и в решении  $\tilde{X}$  все вершины подграфа  $G^0$  имеют единичные значения, а поэтому

$$\tilde{X} = X^0 \vee \tilde{X}_0.$$

Теорема доказана.

**Пример 1 (продолжение)**

Интерпретирующий граф  $G_0$  представлен на рис. 1, в котором  $x_1$  и пара  $(x_3, x_4)$  – свободные переменные. У  $x_1$  и у пары  $(x_3, x_4)$  последователем является  $x_2$ . Следовательно, частными решениями однородной системы будут

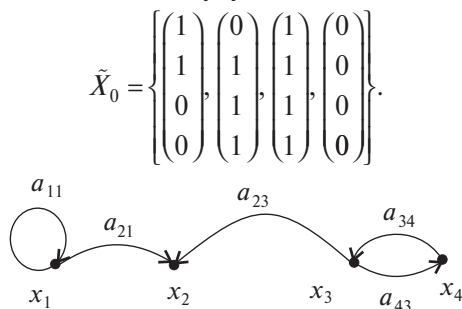


Рис. 1. Интерпретирующий граф из примера 1

$$\tilde{X}_0 = \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right\}.$$

Рассмотрим алгебраический способ отыскания всех частных решений однородной системы.

Переменная  $x_l$  входит в состав свободных, если через соответствующую ей вершину в графе  $G_0$  проходит петля или замкнутый путь. Установить этот факт можно следующим образом. Заменяя  $l$ -й столбец в матрице  $A^{(1)}$  на  $l$ -й столбец из матрицы  $A$ , в результате получим матрицу  $\tilde{A}_l^1$ , которой воспользуемся для вычисления  $x_l$  по формуле, аналогичной (3):

$$x_l = \bigcup_{j=1}^n a_{jl} |A_{jl}^{(1)}| = |\tilde{A}_l^1|, l = \overline{1, n}. \tag{5}$$

Вектор  $x_l^1$  с компонентами  $|\tilde{A}_l^1|$  из (5) назовем вектором автоединиц. Тогда лишь при  $|\tilde{A}_l^1| = 1$  переменная  $x_l$  свободна.

Справедливость такого утверждения подтверждается тем же методом, который был использован при интерпретации 0-решения на графе  $G$ . Действительно, (2.9) интерпретируется на графе  $G$  так же, как и 0-решение, но вместо вектора  $B$  здесь выступает  $l$ -й столбец матрицы  $A$ . То есть, единичное значение определителя  $|\tilde{A}_l^1|$  указывает на существование хотя бы одного единичного пути от введенных вместо  $B$  (ненулевых) вершин-истоков  $a_{jl}, j = \overline{1, n}$  в вершину  $x_l$ , или, что то же, на наличие петли или цикла в графе  $G_0$ , проходящих через  $x_l$ .

Исходя из этого, формулу (5) можно использовать вместо (3) при  $b_j = 0, j \neq l$  и  $b_l = |\tilde{A}_l^1|$  для нахождения того частного решения  $X_0^{(l)}$  однородной системы, которое связано с присвоением единичных значений всем переменным из  $K$ -класса, куда входит  $x_l$ :

$$x_{0i}^{(l)} = |\tilde{A}_i^1| |A_{il}^{(1)}|, i = \overline{1, n}.$$

Такое решение можно найти при всех  $l = \overline{1, n}$ , включая и случаи с  $|\tilde{A}_l^1| = 0$ . Назовем такие решения базисными, принимая всегда допустимое нулевое решение также за базисное. Тогда любое частное решение однородной системы  $\tilde{X}_0$  представимо покомпонентной дизъюнкцией любых наборов базисных решений.

**Пример 1 (продолжение).** Вычислим базисные решения. Сначала по (5) определим свободные переменные:

$$x_1 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{vmatrix} = 1, x_2 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{vmatrix} = 0,$$

$$x_3 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{vmatrix} = 1, x_4 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{vmatrix} = 1.$$

Таковыми оказались  $x_1, x_3$  и  $x_4$ . Вычислим базисное решение:

$$x_{0_1}^1 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{vmatrix} = 1, x_{0_2}^1 = \begin{vmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{vmatrix} = 1,$$

$$x_{0_3}^1 = \begin{vmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 0, x_{0_5}^1 = \begin{vmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{vmatrix} = 0,$$

а это совпадает с первым частным решением, найденным выше. Аналогично находятся и остальные два, а нулевое приписывается.

#### 4. Частное 1-решение дизъюнктивной системы

Особое место занимает 1-решение однородной системы (1), так как позволяет легко найти все те компоненты вектора  $\tilde{X}$ , которые могут принимать неоднозначные значения. Один из способов построения 1-решения однородной системы состоит в определении и суммировании всех базисных решений.

**Теорема 3.** Однородная система, соответствующая системе дизъюнктивных уравнений (1), имеет единственное 1-решение  $X_0^1$ :

$$x_{0_i}^1 = \bigcup_{l=1}^n |\hat{A}_l^1| |A_{ii}^{(1)}|, i = \overline{1, n}. \quad (6)$$

**Доказательство.** Единственность 1-решения очевидна, а единственность (5) как формулы определяется с точностью до эквивалентных форм представления булевых функций.

**Следствие 1.** Интерпретирующий систему (1) граф  $G$  ацикличен, если, и только если,  $x_0^1 = 0$ .

**Следствие 2.** Если интерпретирующий систему (1) граф ацикличен, то она имеет единственное решение, совпадающее с 0-решением (1-решением).

Для выявления ацикличности графа достаточно исследовать  $A$  матрицу по (5). Поэтому имеет место следствие 3.

**Следствие 3.** Интерпретирующий систему (1) граф  $G$  ацикличен, если и только если все компоненты вектора автоединиц нулевые:

$$|\hat{A}_l^1| = 0, l = \overline{1, n}.$$

В соответствии с теоремой 2 1-решение однородной системы (1) может быть найдено как  $X^1 = X^0 \vee X_0^1$ .

**Пример 2.** Найдем 1-решение однородной системы, матрица  $A$  которой взята из примера 1.

Вычислим вектор автоединиц  $X_0^1$ , используя (5):

$$x_1 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{vmatrix} = 1, x_2 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{vmatrix} = 0,$$

$$x_3 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{vmatrix} = 1, x_4 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{vmatrix} = 1,$$

и воспользуемся им для вычисления  $X_0^1$  по (5):

$$x_{0_1}^1 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{vmatrix} = 1, x_{0_2}^1 = \begin{vmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{vmatrix} = 1,$$

$$x_{0_3}^1 = \begin{vmatrix} 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{vmatrix} = 1, x_{0_5}^1 = \begin{vmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{vmatrix} = 1,$$

то есть 1-решением будет:

$$X_0^1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

#### Выводы

В статье развит и обоснован метод решения квадратных дизъюнктивных систем логических уравнений с помощью аппарата логических определителей. Формулы для нахождения корней подобны известным формулам Крамера для решения систем линейных уравнений. Выделены особые типы частных решений, и найден общий вид представления решения для системы дизъюнктивных уравнений. Предложена графовая интерпретация указанных решений. Развитый в статье метод можно распространить на системы конъюнктивного и смешанного типов. Единственным ограничением при этом будет требование квадратности главной матрицы системы логических уравнений.

**Список литературы:** 1. Бондаренко М.Ф., Коноплянко З.Д., Четвериков Г.Г. Основы теории синтеза надшвидкодуючих структур мовних систем штучного інтелекту: Монографія. – К.: ІЗМН, 1997. – 264 с. 2. Закревский А.Д. Матричный аппарат логического синтеза дискретных устройств / ДАН БССР, 1977. – Т. 21, № 11. – С. 991-994. 3. Чистов В.П. Аналитические решения логических уравнений // Изв. АН СССР, Техн. кибернетика, № 2, 1994. – С. 219-224. 4. Бондаренко М.Ф., Шабанов-Кушнарченко Ю.П. Теория интеллекта. Учебник. – Х.: Изд-во СМІТ, 2006. – 571 с.

Поступила в редколлегию 15.04.2008

УДК 681.513



## РЕКУРРЕНТНОЕ БАЙЕСОВСКОЕ ОЦЕНИВАНИЕ СОСТОЯНИЙ УПРАВЛЯЕМЫХ СТОХАСТИЧЕСКИХ ПРОЦЕССОВ

С.Г. Удовенко<sup>1</sup>, В.И. Перепелица<sup>2</sup>

<sup>1, 2</sup> ХНУРЭ, г. Харьков, Украина, udovenko@kture.kharkov.ua

Приведена байесовская модель стохастического процесса в пространстве состояний. Предложена рекуррентная процедура оценивания состояний, позволяющая осуществить прогнозирование выходных параметров процесса. Приведены экспериментальные результаты прогнозирования гидрофизических параметров по разработанной модели.

СТОХАСТИЧЕСКИЙ ПРОЦЕСС, ПРОСТРАНСТВО СОСТОЯНИЙ, БАЙЕСОВСКАЯ МОДЕЛЬ, РЕКУРРЕНТНЫЙ МЕТОД

### Введение

В начале 90-х годов начала формироваться новая комплексная дисциплина, известная в настоящее время под названием «вычислительный интеллект» (ВИ). ВИ базируется на применении мягких вычислений, главным принципом которых является терпимость к неточности и частичной истинности для достижения интерпретируемости, гибкости и низкой стоимости принимаемых решений. К мягким вычислениям относятся, в частности, методы теории хаоса и методы байесовского оценивания. В последнее время расширяется сфера приложений этих методов, связанных с алгоритмизацией анализа временных рядов, включающего восстановление аттрактора в псевдофазовом пространстве, нелинейное предсказание и редукцию шумов [1].

В работах [2,3] показано, что прогнозирование эволюции процесса  $x(k)$ , описываемого ARMAX или RARMAX – моделью в пространстве состояний, требует реализации оперативного оценивания условных функций, с помощью которых описывается неопределенность состояния  $x(k)$ . Представляется целесообразным рассмотреть концептуальное решение задачи такого оценивания, позволяющее определить алгоритмические возможности прогнозирования состояний системы. Кроме того, при дрейфе параметров объекта возникает задача определения условий возможности раздельного или одновременного оценивания параметров и состояний управляемых процессов.

### 1. Постановка задачи

Рассмотрим стохастическую систему, данные о которой наблюдаются в дискретные моменты времени  $k = 1, 2, \dots$ . В соответствии с общей схемой непрямого адаптивного управления, эти данные содержат управляющий вход  $u(k)$ , управляемый выход  $y(k)$ , внешнее измеряемое возмущение  $v(k)$ , вектор задающих воздействий  $y_0(k)$ . Совокупность всей информации, доступной в момент  $k$ , обозначим как

$$d(1, k) = (u(1), y(1), v(1), y_0(1), \dots, u(k), y(k), v(k), y_0(k)). \quad (1)$$

Предположим, что существует стохастическая модель, позволяющая определять распределение вероятностей вектора данных  $d(1, k)$  и его отдельных составляющих для последующих тактов идентификации и управления по результатам текущих наблюдений.

В случае обычного регулирования или программного управления процессом сигналы  $y_0(k)$ ,  $k > 0$  для всех тактов принятия решений априори известны и заданы. В более общем случае (для следящего управления) будущий задающий сигнал, как правило, однозначно не определен.

Применяя последовательно байесовское цепное правило, можно представить совместную функцию условного распределения для прогнозирования будущих данных следующим образом:

$$p(u(1, k), y(1, k), v(1, k) | y_0(1, k)) = \prod_{i=1}^k p(u(i) | u(1, i-1), y(1, i-1), v(1, i-1), y_0(1, i)) \times p(y(k) | y(1, k-1), u(1, k), v(1, k), y_0(1, k)) p(v(k) | v(1, k-1)) p(y_0(i) | y_0(1, k-1)) \quad (2)$$

Рассмотрим смысл отдельных сомножителей произведения (2).

Стратегия управления функцией условных распределений вида

$$f_u = p(u(k) | u(1, k-1), y(1, k-1), v(1, k-1), y_0(1, k)) = p(u(k) | d(1, k-1), y_0(1, k)) \quad (3)$$

представляет собой стохастическое преобразование, в соответствии с которым управление  $u(k)$  на каждом такте генерируется на основании доступных данных.

Эволюция внешних возмущений определяется функцией вида

$$f_v = p(v(k) | v(1, k-1)). \quad (4)$$

Очевидно, что изменение сигнала  $v(i)$  можно считать автономным процессом, определяемым внешней средой.

Если предположить, что не существует скрытых связей в контуре управления, которые могут

влиять на изменение задающего сигнала  $y_0(k)$ , то в общем случае можно прогнозировать его эволюцию с помощью функции вида

$$f_{y_0} = p(y_0(k)|y_0(1, k-1)). \quad (5)$$

Важнейшей составляющей произведения в правой части (2) является функция, определяющая вероятностную зависимость выхода управляемой системы от предыстории процесса и текущих данных, то есть

$$\begin{aligned} f_y &= p(y(k)|y(1, k-1), u(1, k), v(k), y_0(1, k)) = \\ &= p(y(k)|d(1, k-1), u(k), v(k), y_0(1, k)). \end{aligned} \quad (6)$$

Очевидно предположить, что обусловленность выхода  $y(k)$  от  $y_0(1, k)$  является избыточной, т.к. эволюция задающего сигнала может влиять на  $y(k)$  только посредством данных  $d(k-1)$ , которые уже включены в условие. Таким образом, управляемый процесс характеризуется следующим стохастическим преобразованием:

$$f_y = p(y(k)|d(1, k-1), u(k), v(k)). \quad (7)$$

Семейство функций условных распределений (7) для всех  $k$  задает математическую модель процесса. Целью моделирования с использованием байесовских оценок можно считать определение параметров этих распределений, позволяющих проводить выбор необходимых стратегий управления, основанных на пошаговом вычислении функции (3).

Решим задачу представления байесовской модели в пространстве состояний, а также последующего синтеза процедуры рекуррентного оценивания состояний стохастической системы.

## 2. Решение задачи

**Модель процесса.** Рассмотрим зависимость (7). Предположим, что в соответствии с зафиксированной ранее совокупностью данных  $d(1, k-1)$  выбирается некоторое управление  $u(k)$ , которому соответствует оценка прогнозируемого выхода  $\hat{y}(k)$ . Этот прогноз может оказаться и абсолютно точным на отдельных тактах контроля и управления, но в общем случае действительное значение выхода будет равно

$$y(k) = \hat{y}(k) + e(k), \quad (8)$$

где величина  $e(k)$  отражает рассогласование между действительным значением  $y(k)$  и его оценкой и может быть как положительной, так и отрицательной. В момент времени  $k$  регистрируется новый набор данных  $\{u(k), y(k), v(k)\}$  и выбирается новое управление  $u(k+1)$  в соответствии с некоторой стратегией. По этим данным опять можно осуществить прогноз значения выхода  $\hat{y}(k+1)$ , действительное значение, которого составит

$$y(k+1) = \hat{y}(k+1) + e(k+1). \quad (9)$$

Естественно предположить, что выбираемые оценки  $\hat{y}(k-1)$ ,  $\hat{y}(k)$ ,  $\hat{y}(k+1)$  являются средними значениями действительных величин

$$\begin{aligned} \hat{y}(k) &= M\{y(k)|d(1, k-1), u(k), v(k)\} = \\ &= \int y(k) p(y(k)|d(1, k-1), u(k), v(k)) dy(k). \end{aligned} \quad (10)$$

Выразим оценку  $\hat{y}(k)$  с помощью некоторой детерминированной функции

$$\hat{y}(k) = f(d(1, k-1), u(k), v(k)). \quad (11)$$

В этом случае выходную величину можно представить в виде

$$y(k) = f(d(1, k-1), u(k), v(k)) + e(k). \quad (12)$$

Таким образом, модель процесса определяется совокупностью детерминированной части и случайной составляющей  $e(k)$ .

Для большинства реальных технических систем допустимо считать временной ряд  $\{e(k), k=1, 2, \dots\}$  дискретным белым шумом, то есть последовательностью некоррелированных величин с нулевым средним значением. Отметим, что в детерминированной части зависимости (12) учитывается полная совокупность данных от начального момента наблюдения до момента  $(k-1)$ . Однако известно, что в реальных системах на текущий выход не влияют существенно данные, измеренные много тактов наблюдения назад. Зачастую можно предположить, что существует такое целое число  $n$ , для которого справедливо

$$\begin{aligned} p(y(k)|d(1, k-1), u(k), v(k)) = \\ = p(y(k)|d(k-n, k-1), u(k), v(k)) \end{aligned} \quad (13)$$

где  $n$  определяет учитываемую часть предыстории. Если оценивать качество модели по критерию среднеквадратичной ошибки прогноза, то для целей управления, как правило, достаточно использовать модель порядка  $n \leq 3$ .

Очевидно, что в случае линейности функции (11) прогнозирующая модель процесса будет также линейной.

Модель стохастического процесса является линейной, если математическое ожидание  $y(k)$  (среднее значение функции (13)) может быть представлено линейной функцией данных, причем ковариация  $y(k)$  не зависит от этих данных, то есть

$$\begin{aligned} M\{y(k)|d(k-n, k-1), u(k), v(k)\} = \hat{y}(k) = \\ = -\sum_{i=1}^n A'_i(k) y(k-i) + \sum_{i=0}^n B'_i(k) u(k-i) + \\ + \sum_{i=0}^n C'_i(k) v(k-i) + g'_y(k); \end{aligned} \quad (14)$$

$$\text{Cov}\{y(k)|d(k-n, k-1), u(k), v(k)\} = R_y(k). \quad (15)$$

Подставив (14) в (8), получаем

$$A'_0(k)y(k) = -\sum_{i=1}^n A'_i(k)y(k-i) + \sum_{i=0}^n B'_i(k)u(k-i) + \sum_{i=0}^n C'_i(k)v(k-i) + g'_y(k) + e(k), \quad (16)$$

где  $A'_0(k) = I$ .

Допустим, что матрица ковариации (15) может быть представлена в следующем факторизованном виде:

$$R_y(k) = L_y(k)D_e(k)L_y^T(k), \quad (17)$$

где  $L_y(k)$  – моническая нижняя треугольная матрица,  $D_e(k)$  – диагональная матрица с неотрицательными диагональными элементами. Если умножить на  $L_y^{-1}(k)$  обе части уравнения (16), то  $A_0(k)$  преобразуется в нижнюю треугольную матрицу, то есть

$$A_0(k) = L_y^{-1}(k). \quad (18)$$

Очевидно, что стохастическая составляющая  $e(k)$  также переопределится

$$e(k) = L_y^{-1}(k)(y(k) - \hat{y}(k)), \quad (19)$$

причем теперь  $e(k)$  имеет не только нулевое математическое ожидание, но и некоррелированные компоненты

$$Cov\{e(k)\} = M\{e(k)e^T(k)\} = D_e(k) = D_e = const. \quad (20)$$

Таким образом, модель (16) без изменения структуры может быть легко представлена в форме, характеризующейся важным свойством (20).

Реальные стохастические процессы зачастую характеризуются транспортным запаздыванием. Их можно разделить на процессы с транспортным запаздыванием, которые можно учесть повышением порядка модели и на процессы, для которых временная задержка должна быть отражена в соответствующих векторах данных зависимости (16):

$$A(k)y(k) = -\sum_{i=1}^n A_i(k)y(k-i) + \sum_{i=0}^n B_i(k)u(k-i-\rho) + \sum_{i=0}^n C_i(k)v(k-i-\mu) + g_y(k)e(k) \quad (21)$$

где  $\rho, \mu$  – векторные величины, элементы которых соответствуют числу периодов транспортного запаздывания по отдельным каналам « $y(k) - u(k)$ », « $y(k) - v(k)$ ».

В этой модели условное среднее значение линейно зависит от  $n$  предыдущих входов. Если удастся удержать процесс в окрестности некоторой рабочей точки, то можно предположить, что имеет место действительно линейная зависимость.

Насколько большой может быть эта окрестность, зависит от свойств конкретной системы. Случайные отклонения от линейности можно отнести к составляющей  $e(k)$ .

Зависимость (21) представляет собой многомерную регрессионную модель стохастического объекта, часто именуемую авторегрессионной моделью со скользящим средним (ARMAX: Autoregressive Moving Average Exogenous). Очевидно, что форма линейной модели типа «вход-выход» не всегда удобна для непосредственного практического использования. Так как любая математическая модель может служить лишь упрощенным отражением объективной реальности, то один и тот же процесс может быть описан с помощью различных ее модификаций, отражающих некоторые специфические допущения.

Введем понятие состояния многомерного управляемого стохастического процесса, используя условную вероятностную зависимость (6).

Конечное векторное множество  $x(k)$  фиксированной размерности, характеризующее текущую динамику объекта управления и соответствующее зависимости

$$p(y(k), x(k) | y(1, k-1), u(1, k), v(1, k), x(k-1)) = p(y(k), x(k) | u(k), v(k), x(k-1)), \quad (22)$$

назовем состоянием управляемого стохастического процесса.

Очевидно, что из общей зависимости (22) можно выделить локальные условные зависимости  $p(x(k))$  и  $p(y(k))$  от предыстории процесса и предыдущего состояния  $x(k-1) = z(k) \setminus u(k)$ .

При этом первая из формируемых таким способом функций, то есть

$$p(x(k) | u(k), v(k), x(k-1)) \quad (23)$$

непосредственно описывает эволюцию состояний, а функция

$$p(y(k) | u(k), v(k), x(k-1)) \quad (24)$$

соответствует эволюции выхода для модели процесса в пространстве состояний.

Введем обозначение

$$x_1(k-1) = \sum_{j=1}^n [-A_j y(k-j) + B_j u(k-\rho-j) + C_j v(k-\mu-j) + h_j e(k-j) + g_e].$$

Тогда ARMAX – модель типа (16) можно представить в виде:

$$A_0 y(k) = B_0 u(k-\rho) + C_0 v(k-\mu) + e(k) + x_1(k-1). \quad (25)$$

Очевидно, что сдвиг индекса времени на один такт вперед преобразует (25) следующим образом:

$$A_1 y(k) = B_1 u(k-\rho) + C_1 v(k-\mu) + h_1 e(k) + x_2(k-1) - x_1(k), \quad (26)$$

где

$$x_2(k-1) = \sum_{j=2}^n [-A_j y(k+1-j) + B_j u(k+1-\rho-j) + C_j v(k+1-\mu-j) + h_j e(k+1-j)] + g_e.$$

В общем случае для  $i < n$  можно получить зависимость

$$A_i y(k) = B_i u(k-\rho) + C_i v(k-\mu) + h_i e(k) + x_{i+1}(k-1) - x_i(k), \quad (27)$$

где

$$x_i(k-1) = \sum_{j=i+1}^n [-A_j y(k+i-j) + B_j u(k+i-\rho-j) + C_j v(k+i-\mu-j) + h_j e(k+i-j)] + g_e. \quad (28)$$

При  $i = n$  формируется ARMAX – модель в виде:

$$A_n y(k) = B_n u(k-\rho) + C_n v(k-\mu) + h_n e(k) + g_e - x_n(k). \quad (29)$$

Используя зависимости (26)–(29), можно представить матричную ARMAX – модель в пространстве состояний:

$$A x_c(k) = B u(k-\rho) + C v(k-\mu) + H e(k) + E x(k-1) + g_e, \quad (30)$$

где

$$x_c^T(k) = [y^T(k), x^T(k)], x^T(k) = [x_1^T(k), x_2^T(k), \dots, x_n^T(k)],$$

$$E = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, A = \begin{bmatrix} a_0 & 0 & \cdot & 0 \\ a_1 & 1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ a_n & 0 & \cdot & 1 \end{bmatrix},$$

$$B^T = [b_0^T, \dots, b_n^T], C^T = [c_0^T, \dots, c_n^T], g_c^T = [0, \dots, g_e],$$

$I$  – единичная матрица размерности  $m$ .

Очевидно, что  $A$  является монической нижней треугольной матрицей размерности  $[(n+1)m_y \times (n+1)m_y]$ .

Для упрощения изложения положим, что временные задержки  $\rho = \mu = 0$ . Тогда условное математическое ожидание и условная матрица ковариации переменной  $x_c(k)$  примут вид

$$M\{x_c(k) | u(k), v(k), x(k-1)\} = A^{-1}(Bu(k) + Cv(k) + Ex(k-1) + g_e); \quad (31)$$

$$Cov\{x_c(k) | u(k), v(k), x(k-1)\} = A^{-1} H D_e H^T (A^{-1})^T.$$

**Оценивание состояний.** Предположим, что известна предыстория стохастического процесса  $d(1, k-1)$ , то есть зафиксированы соответствующие значения входных и выходных переменных в моменты времени  $t = iT$ ,  $i = 1 \div (k-1)$ .

Тогда проблему байесовского оценивания состояний можно сформулировать следующим обра-

зом: по заданной модели процесса в пространстве состояний, определяемой условной функцией вида (22), найти прогнозирующую функцию условного распределения вероятностей для очередного значения выхода и после его наблюдения определить функцию условного распределения вероятностей для состояния  $x(k)$  с целью подготовки следующего шага рекурсии.

Решение этой проблемы можно представить в виде трехэтапной процедуры. На первом этапе определяется совместное распределение вероятностей для  $y(k)$  и  $x(k)$  по результатам наблюдения  $d(1, k-1)$  и расчетного значения управления  $u(k)$ . Такое распределение, полученное на основе элементарных операций байесовской статистики, будет иметь вид:

$$p(y(k), x(k) | d(1, k-1), u(k)) = \int p(y(k), x(k), x(k-1) | d(1, k-1), u(k)) dx(k-1) = \int p(y(k), x(k) | d(1, k-1), x(k-1), u(k)) p(x(k-1) | d(1, k-1), u(k)) dx(k-1).$$

В соответствии с (22) получаем:

$$p(y(k), x(k) | d(1, k-1), x(k-1), u(k)) = p(y(k), x(k) | x(k-1), u(k)). \quad (32)$$

Очевидно, что можно сформулировать естественное условие независимости текущих значений состояния  $x(k)$  от текущих значений управления  $u(k)$ :

$$p(x(k-1) | d(1, k-1), u(k)) = p(x(k-1) | d(1, k-1)). \quad (33)$$

С учетом (31) и (32) имеем:

$$p(y(k), x(k) | d(1, k-1), u(k)) = \int p(y(k), x(k) | d(1, k-1), u(k)) \times p(x(k-1) | d(1, k-1)) dx(k-1). \quad (34)$$

На втором этапе определяем функцию условного распределения для прогнозирования выхода процесса:

$$p(y(k) | d(1, k-1), u(k)) = \int p(y(k), x(k) | d(1, k-1), u(k)) dx(k). \quad (35)$$

На третьем этапе завершаем рекурсию, определяя условное распределение вероятностей с учетом распределения (31) и результатов прогнозирования (34):

$$p(x(k) | d(1, k)) = \frac{p(y(k), x(k) | d(1, k-1), u(k))}{p(y(k) | d(1, k-1), u(k))}. \quad (36)$$

Очевидно, что если вероятность  $p(y(k), x(k) | x(k-1), u(k))$  подчинена нормальному закону распределения и соответствующая модель процесса является линейной, то прогнозирующие функции (4), (5) и (6) также распределены по

нормальному закону. Это означает, что если для линейной ARMAX– модели, представленной в пространстве состояний, задать априорную неопределенность начального состояния  $x(0)$  с помощью нормально распределенной функции  $p(x(0))$ , то задача байесовского оценивания может быть решена путем рассмотрения лишь первого и второго моментов распределения (35).

При распределении вероятностей, используемых в рекурсиях, но отличных от нормальных законов, решение задачи оценивания состояний сопряжено со значительными вычислительными трудностями.

### 3. Экспериментальные результаты

Полученные теоретические результаты были частично использованы применительно к задачам прогнозирования и оперативной регистрации опасности возникновения волн цунами по данным реальных гидрофизических измерений [4].

Исследуемый процесс представляет собой дискретную последовательность регистраций уровня с десятиминутным периодом квантования. По условиям эксперимента в ситуациях, когда абсолютная величина прогнозирования на текущем такте вычислений превышала пороговую величину, принималась гипотеза о потере устойчивости процесса вследствие возникновения волны цунами. Качество прогнозирования в целом оценивалось по среднему квадрату ошибки прогноза на всем интервале моделирования.

В качестве прогнозирующей зависимости по данным ретроспективного анализа была использована обобщенная модель в пространстве состояний, а для настройки ее параметров использовался алгоритм байесовского оценивания с экспоненциальным дисконтированием. Параллельно для сравнения качества прогнозирования оценки параметров модели вычислялись также по рекуррентному методу наименьших квадратов (РМНК). Начальные значения во всех случаях принимались нулевыми.

Процедуры байесовского оценивания во всех рассмотренных ситуациях обеспечивали более высокое качество прогнозирования по сравнению с РМНК. При этом наблюдается тенденция улучшения качества прогнозирования при уменьшении величины «скользящего окна» на участке квазистационарности процесса ( $105 \leq k \leq 204$ ). На услов-

но стационарных интервалах качество прогнозов было приблизительно одинаковым для различных величин скользящего окна  $S_{ок}$ . Относительная ошибка прогноза на стационарных участках для всех  $S_{ок}$  не превышает 0,09. На участках нестационарности максимальная относительная ошибка для  $S_{ок} = 5$  составила 0,35; для  $S_{ок} = 10$  составила 0,65; для  $S_{ок} = 15$  составила 0,67.

Анализ отдельных реализаций, зарегистрированных в процессе измерений, показал их хаотичность, не позволяющую осуществлять прогнозирование и принимать решения по приведенным в настоящей работе алгоритмам. В то же время расширение возможностей прогнозирования становится реальным при дополнении вычислительных процедур методами теории динамического хаоса, на основании которых выделяется регулярная составляющая анализируемого процесса либо делается вывод о его шумовой природе.

### Выводы

Приведено концептуальное решение одной из задач вычислительного интеллекта – байесовского оценивания состояний, позволяющего расширить алгоритмические возможности прогнозирования состояний в динамических системах. В случае хаотичности анализируемых последовательностей целесообразным является дополнение предложенной модели прогнозирования тестами теории динамического хаоса. Результаты экспериментального моделирования подтверждают работоспособность предложенной рекуррентной процедуры. Прогнозирование с помощью байесовских моделей может быть использовано при решении задач динамического предсказания ветвлений и переходов в компьютерах новых поколений.

**Список литературы.** 1. *Е.В. Бодянский, С.Г. Удовенко, А.Е. Ачкасов, Г.К. Вороновский.* Субоптимальное управление стохастическими процессами – Харьков: Основа, 1997. – 140 с. 2. *P. Congdon.* Bayesian Statistical Modeling. – New York.: Wiley, 2001. – 327 p. 3. *Zyng L.* Recursive technique for identifying dynamic system // Proc. Of the Annual Control Conference. – Indiana, 1985. – P. 1–11. 4. *Калоша В.А., Кирьяк Р.Д., Удовенко С.Г.* Алгоритм робастного управления стохастическими объектами с дрейфом параметров // Вестник Харьковского государственного политехнического университета. – 2000. – Вып. 99. – С. 70–73.

*Поступила в редколлегию 18.04.2008*

УДК 681.3



## МОДИФИЦИРОВАННЫЙ РЕКУРРЕНТНЫЙ МЕТОД НАИМЕНЬШИХ КВАДРАТОВ С КРИТИЧЕСКИМИ СВОЙСТВАМИ

В. А. Тимофеев<sup>1</sup>, Самир Лара<sup>2</sup>, В. Д. Непочатова<sup>3</sup>

<sup>1, 2, 3</sup> ХНУРЭ, г. Харьков, Украина, timofeev2001@yahoo.com

Рассматривается задача идентификации динамического объекта в предположении, что известен лишь уровень помехи. Исследованы существующие алгоритмы. Разработан рекуррентный алгоритм идентификации, имеющий критические свойства и свойства МНК-оценок, определены условия его сходимости. Преимуществом разработанного алгоритма является простота его использования в задачах контроля и управления.

КОНТРОЛЬ, АДАПТАЦИЯ, КРИТИЧЕСКАЯ ИДЕНТИФИКАЦИЯ, ПОМЕХА, СХОДИМОСТЬ

### Введение

Принятие решений в экологических, технических, экономических системах невозможно без учета информации о текущем состоянии объекта. При этом объект может быть нелинейным, нестационарным, его параметры могут изменяться во времени. Таким образом, решение задач контроля и принятия решений в условиях нестационарности объекта и динамического проявления внешней среды возможно лишь на основе использования математического аппарата, предполагающего изменение параметров модели во времени. Данная задача может быть решена с использованием теории адаптивных систем. К настоящему времени сформировался ряд относительно независимых направлений в теории адаптивных систем [1-6]. Однако во всех классических подходах предполагается, что возмущения, действующие в системе, имеют стохастическую природу, причем это, как правило, белый шум с нулевым математическим ожиданием и ограниченной дисперсией. В практических ситуациях статистические предпосылки являются надуманными, в связи с чем гораздо более реальными представляются допущения лишь об ограниченности шума или его разностей по амплитуде. В этих условиях использование методов идентификации, основанных на квадратичных критериях, и, прежде всего, рекуррентного метода наименьших квадратов явно неэффективно. Возникающие затруднения частично могут быть преодолены в рамках адаптивных робастных систем управления [7-10], в которых, тем не менее, все равно «спрятаны» определенные статистические предпосылки.

В связи с этим представляет интерес разработка математических моделей, предполагающих синтез теории адаптивного и критического управления, что приведет к созданию адаптивных критических методов контроля и идентификации динамических объектов, функционирующих в условиях существенной неопределенности о характеристиках объекта и окружающей среды.

Целью настоящей работы и является разработка рекуррентного метода идентификации, обеспечивающего получение оценок, обладающих суп-

ремальными свойствами, которые не зависят от статистических характеристик сигналов и помех, и свойствами МНК-оценок.

### 1. Постановка задачи

Рассмотрим динамический объект, функционирующий в замкнутой системе управления  $S_D(P, C)$ , описываемый разностным уравнением

$$A(q)y(k) = q^{-d}B(q)u(k) + \xi(k), \quad (1)$$

где полиномы  $A(q) \in R[q, n]$  с  $a_0 = 1$ ,  $B(q) \in R[q, m]$ ;  $d$  — время чистого запаздывания  $d \in N^+$ ;  $y, u$  и  $\xi$  — выходной, управляющий и возмущающий сигналы соответственно.

Относительно возмущений предполагается ограниченность их первых разностей.

В случае, если параметры объекта неизвестны, можно воспользоваться тем или иным методом идентификации. Как правило, в качестве процедур идентификации применяются те или иные модификации рекуррентного метода наименьших квадратов либо проекционные алгоритмы, так или иначе связанные с квадратичными критериями. При использовании критериев, отличных от квадратичных, например, модульных, хотя и получают робастные процедуры, статистический смысл задачи идентификации тем не менее сохраняется. Естественно, что такие алгоритмы идентификации не могут быть использованы в критических системах управления.

В связи с этим возникает необходимость синтеза адаптивных алгоритмов идентификации, не связанных ни с какими статистическими предпосылками, обладающих высокой скоростью сходимости, вычислительной простотой и пригодных для работы в реальном времени в контуре критической системы управления динамическим объектом.

### 2. Алгоритмы идентификации, применяемые в критических системах

Рассмотрим полином

$$G(q) = 1 - \Delta A(q), \quad (2)$$

где  $G(q) = g_1 q^{-1} + g_2 q^{-2} + \dots + q^{-n-1}$ ,

и перепишем уравнение объекта (1) в виде

$$y(k) = \Theta^T \psi(k-1) + \Delta \xi(k), \quad (3)$$

где  $\Theta = (g_1, g_2, \dots, g_{n+1}, b_0, b_1, \dots, b_m)^T$ ;

$$\psi(k-1) = (y(k-1), y(k-2), \dots, y(k-n-1),$$

$$\Delta u(k-d), \Delta u(k-d-1), \dots, \Delta u(k-d-m))^T;$$

$$\Delta u(k) = u(k) - u(k-1);$$

$$\Delta \xi(k) = \xi(k) - \xi(k-1).$$

Тогда задача идентификации с позиции теории критических систем сводится к нахождению оценок неизвестного вектора параметров  $\Theta$  таких, что

$$\Omega(\hat{\Theta}) = \{\hat{\Theta} : |y(k) - \hat{\Theta}^T \psi(k-1)| \leq \delta, \forall k \in \mathbb{N}\}. \quad (4)$$

Здесь  $\hat{\Theta}$  — оценка параметра  $\Theta$ .

К настоящему времени сложился ряд подходов к задаче идентификации, связанной с неравенством (4). Это, прежде всего, подход Фогеля-Хуанга [11], в основе которого лежат некоторые геометрические построения. Известна также процедура Лозано-Лиля-Ортеги [12], синтезированная как на геометрических предпосылках, так и исходя из условий устойчивости процесса сходимости. Нельзя не отметить также алгоритм Канудас де Вита-Каррильо [13], являющийся некоторой модификацией экспоненциально взвешенного рекуррентного МНК. Несмотря на эффективность этих процедур, их использование в критических системах наталкивается на серьезные затруднения.

Так, оптимальный алгоритм Фогеля-Хуанга настолько сложен с вычислительной точки зрения, что не может быть и речи о его использовании в режиме реального времени. Эта сложность обусловлена, прежде всего, необходимостью отыскания на каждой итерации глобального минимума многоэкстремальной функции  $n+m+2$  переменных, что само по себе является достаточно сложной проблемой.

В алгоритме Лозано-Лиля-Ортеги, имеющем вид

$$\hat{\Theta}(k) = \hat{\Theta}(k-1) + \frac{\alpha P(k-1) \psi(k-1)}{1 + \psi^T(k-1) P(k-1) \psi(k-1)} \times \times (|e(k)| - \delta_1 \text{sign } e(k), \alpha \in (0,1); \quad (5)$$

$$P^{-1}(k) = \begin{cases} P^{-1}(k-1) + \\ + \frac{\alpha P(k-1) \psi(k-1)}{(1 + \psi^T(k-1) P(k-1) \psi(k-1)) e(k)} \times \\ \times (|e(k)| - S_1) \text{sign } e(k), & |e(k)| > \delta_1; \\ P^{-1}(k-1), & |e(k)| \leq \delta_1; \end{cases} \quad (6)$$

$$S_1 = \sqrt{1 + \alpha} \delta; \quad (7)$$

$$e(k) = y(k) - \hat{\Theta}^T(k-1) \psi(k-1), \quad (8)$$

априори предполагается ограниченность значения  $\psi^T(k-1) P(k-1) \psi(k-1)$ , из которого следует условие сходимости

$$\lim_{k \rightarrow \infty} |e(k)| = \sqrt{1 + \alpha} \delta, \alpha \in (0,1),$$

то есть ошибка идентификации  $e(k)$  никогда не может быть по модулю меньше заданных ограничений  $\delta$ .

В алгоритме Канудас де Вита-Каррильо

$$\hat{\Theta}(k) = \hat{\Theta}(k-1) + \frac{\alpha(k) P(k-1) \psi(k-1)}{\psi^T(k-1) P(k-1) \psi(k-1)} (|e(k)| - \delta) \text{sign } e(k); \quad (9)$$

$$P(k) = \lambda^{-1} (P(k-1) - \frac{\alpha(k) P(k-1) \psi(k-1) \psi^T(k-1) P(k-1)}{\psi^T(k-1) P(k-1) \psi(k-1)} (1 - \frac{\delta}{|e(k)|})), \quad (10)$$

$$\lambda \in (0,1];$$

$$\alpha(k) =$$

$$= \begin{cases} 1, & |e(k)| > \delta \text{ или } \psi^T(k-1) P(k-1) \psi(k-1) = 0; \\ 0, & |e(k)| \leq \delta, \end{cases} \quad (11)$$

где  $e(k)$  определяется соотношением (8),

$$e(k) = y(k) - \hat{\Theta}^T(k-1) \psi(k-1).$$

В ситуации, когда  $|e(k)| > \delta$  и значение  $\psi^T(k-1) P(k-1) \psi(k-1)$  близко к нулю, возникает режим неустойчивости, поскольку компоненты вектора

$$P(k-1) \psi(k-1) (\psi^T(k-1) P(k-1) \psi(k-1))^{-1}$$

могут неограниченно возрастать. Кроме того, в случае, когда  $\alpha(k) = 0$ , невозможно гарантировать выполнение условия  $|e(k)| \leq \delta$  в предположении, что  $\psi^T(k-1) P(k-1) \psi(k-1)$  ограничено.

### 3. Модифицированный алгоритм идентификации и оценивание его сходимости

Объединяя достоинства рассмотренных процедур, введем комбинированный алгоритм, являющийся своеобразной комбинацией рекуррентного МНК и процедур (5)-(7) и (9)-(11).

Рассмотрим алгоритм вида

$$\hat{\Theta}(k) = \hat{\Theta}(k-1) + \frac{\alpha(k) P(k-1) \psi(k-1)}{1 + \psi^T(k-1) P(k-1) \psi(k-1)} \times \times (|e(k)| - \delta) \text{sign } e(k), \quad (12)$$

$$P(k) = P(k-1) - \frac{\alpha(k) P(k-1) \psi(k-1) \psi^T(k-1) P(k-1)}{|e(k)| + (2|e(k) - \delta) \psi^T(k) P(k-1) \psi(k-1)} \times \times (|e(k)| - \delta), \quad (13)$$

$$\alpha(k) = \begin{cases} 1, & \text{если } |e(k)| > \delta, \\ 0, & \text{если } |e(k)| \leq \delta, \end{cases} \quad (14)$$

где  $e(k)$  определяется в соответствии с (8).

#### Анализ сходимости предложенного алгоритма

Рассмотрим вектор уклонений оценок от истинных значений параметров

$$\tilde{\Theta}(k) = \Theta - \hat{\Theta}(k)$$

и функцию Ляпунова

$$V(k) = \tilde{\Theta}^T(k)P^{-1}(k)\tilde{\Theta}(k).$$

Объединяя (3) с (12)-(14), получаем

$$\begin{aligned} V(k) &= V(k=1) + \\ &+ \frac{\alpha(k)(|e(k)|-\delta)}{(1+\psi^T(k-1)P(k-1)\psi(k-1))|e(k)|} \times \\ &\quad \times (\Delta\xi^2(k) - \\ &\quad - \frac{(1+\psi^T(k-1)P(k-1)\psi(k-1))|e(k)|^3}{|e(k)|+(2|e(k)|-\delta)\psi^T(k-1)P(k-1)\psi(k-1)}). \end{aligned} \quad (15)$$

С учетом ранее введенного условия  $\xi \in D(0, \delta)$  несложно переписать (15) в виде неравенства

$$\begin{aligned} V(k) &\leq V(k-1) + \frac{\alpha(k)(|e(k)|-\delta)}{(1+\psi^T(k-1)P(k-1)\psi(k-1))|e(k)|} \times \\ &\times (\delta^2 - \frac{(1+\psi^T(k-1)P(k-1)\psi(k-1))|e(k)|^3}{|e(k)|+(2|e(k)|-\delta)\psi^T(k-1)P(k-1)\psi(k-1)}) = \\ &= V(k-1) - \frac{\alpha(k)(|e(k)|-\delta)}{(1+\psi^T(k-1)P(k-1)\psi(k-1))|e(k)|} \times \\ &\times \frac{(|e(k)|^3 - 2|e(k)|\delta^2 + \delta^3)(1+\psi^T(k-1)P(k-1)\psi(k-1)) + (|e(k)|-\delta)\delta^2}{|e(k)|+(2|e(k)|-\delta)\psi^T(k-1)P(k-1)\psi(k-1)} \leq \\ &\leq V(k-1) - \frac{\alpha(k)(|e(k)|-\delta)(|e(k)|^3 - 2|e(k)|\delta^2 + \delta^3)}{(|e(k)|+(2|e(k)|-\delta)\psi^T(k-1)P(k-1)\psi(k-1))|e(k)|}, \end{aligned}$$

которое справедливо в случае  $|e(k)| \geq \delta$ . Кроме того, поскольку в этом случае

$$e^2(k) - 2\delta^2 + \frac{\delta^3}{|e(k)|} \geq |e(k)|(|e(k)|-\delta),$$

то

$$V(k) \leq V(k-1) - \frac{\alpha(k)(|e(k)|(|e(k)|-\delta)^2)}{2(1+\psi^T(k-1)P(k-1)\psi(k-1))|e(k)|}, \quad (16)$$

откуда

$$\lim_{k \rightarrow \infty} \frac{\alpha(k)(|e(k)|-\delta)^2}{1+\psi^T(k-1)P(k-1)\psi(k-1)} = 0, \quad (17)$$

что свидетельствует о критериальной сходимости алгоритма (12)-(14).

Перенеся в левую часть (12)  $\hat{\Theta}(k-1)$  и возведя обе части полученного выражения в квадрат, получаем

$$\begin{aligned} \left\| \hat{\Theta}(k) - \hat{\Theta}(k-1) \right\|_{V(k)}^2 &= \\ &= \frac{\alpha(k)\psi^T(k-1)P^2(k-1)\psi(k-1)}{(1+\psi^T(k-1)P(k-1)\psi(k-1))^2} (|e(k)|-\delta^2) \leq \\ &\leq \frac{\alpha(k)\lambda_{\max}(P(k-1))}{1+\psi^T(k-1)P(k-1)\psi(k-1)} (|e(k)|-\delta^2), \end{aligned} \quad (18)$$

где  $\lambda_{\max}(P(k-1))$  — максимальное собственное значение матрицы  $P(k-1)$ .

Из выражения (13) следует условие

$$\lambda_{\max}(P(k)) \leq \lambda_{\max}(P(k-1)) \leq \dots \leq \lambda_{\max}(P(0)),$$

позволяющее переписать (18) в виде

$$\begin{aligned} \left\| \hat{\Theta}(k) - \hat{\Theta}(k-1) \right\|_{V(k)}^2 &\leq \\ &\leq \alpha(k)\lambda_{\max}(P(0)) \frac{(|e(k)|-\delta)^2}{1+\psi^T(k-1)P(k-1)\psi(k-1)}, \end{aligned}$$

который вместе с выражением (17) свидетельствует об аргументной сходимости алгоритма.

Далее, используя лемму об обращении матриц, запишем выражение

$$\begin{aligned} P^{-1}(k) &= P^{-1}(k-1) + \\ &+ \frac{\alpha(k)\psi(k-1)\psi^T(k-1)}{1+\psi^T(k-1)P(k-1)\psi(k-1)} \left(1 - \frac{\delta}{|e(k)|}\right), \end{aligned}$$

откуда следует неравенство

$$\lambda_{\min}(P^{-1}(k)) \geq \lambda_{\min}(P^{-1}(k-1)) \geq \dots \geq \lambda_{\min}(P^{-1}(0)),$$

где  $\lambda_{\min}(P^{-1}(k))$  — минимальное собственное значение матрицы.

Это неравенство совместно с (16) приводит к тому, что

$$V(k) \leq V(0)$$

и

$$\lambda_{\max}(P^{-1}(0)) \left\| \tilde{\Theta}(k) \right\|_{V(k)}^2 \leq \lambda_{\max}(P^{-1}(k)) \left\| \tilde{\Theta}(0) \right\|_{V(k)}^2,$$

откуда можно записать выражение

$$\left\| \tilde{\Theta}(k) \right\|_{V(k)}^2 \leq \frac{\lambda_{\max}(P^{-1}(k))}{\lambda_{\min}(P^{-1}(0))} \left\| \tilde{\Theta}(0) \right\|_{V(k)}^2,$$

определяющее скорость сходимости введенного алгоритма.

#### Выводы

В работе предложена модификация рекуррентного МНК, обладающая критическими свойствами. Так как основой данного алгоритма является рекуррентный МНК, трудностей с его практической реализацией не возникает. Полученная оценка скорости сходимости предложенного алгоритма свидетельствует о том, что эта скорость в значи-

тельной мере определяется свойствами ковариационной матрицы наблюдений  $P^{-1}(k)$  (соотношением ее максимального и минимального собственных чисел). Кроме того, входящая в алгоритм величина  $\delta$  зачастую известна лишь приближенно, поэтому необходимо в процессе идентификации осуществлять оценивание (уточнение) этой величины и подставлять полученные оценки в алгоритм идентификации.

**Список литературы:** 1. Clarke D. W., Gawthrop P. J. Self-tuning controller // Proc. IEE. — 1975. — 122. — P. 929-934. 2. Wellstead P. E., Edmunds M. J., Prager D., Zanker P. Self-tuning pole/zero assignment regulators // Int. J. Contr. — 1979. — 30. — №1. — P. 1-26. 3. Wellstead P. E., Prager D., Zanker P. Pole assignment self-tuning regulator // Proc. DEE. — 1979. — 126. — D. — P. 781-787. 4. Astrom K. J., Wittenmark B. Self-tuning controllers based on pole-zero placement // Proc. IEE. — 1980. — 127. — D. — P. 120-130. 5. Kreisselmeier G., Narendra K. S. Stable MRAC in the presence of bounded disturbances // IEEE Trans. on Autom. Contr. — 1982. — 27. —

P. 1169-1175. 6. Clarke D. W., Mohtadi C., Tuffs P. S. Generalized predictive control. Tue basis algorithm // Automatica. — 1987. — 22. — № 2. — P. 137-148. 7. Samson C. Stability analysis of adaptively controlled systems subject to bounded disturbances // Automatica. — 1983. — 19. — P. 81-89. 8. Narendra K. S., Annaswamy A. M. Robust adaptive control in the presence of bounded disturbances // IEEE Trans. on Autom. Contr. — 1986. — 31. — № 4. — P. 306-315. 9. Ortega R., Lozano-Leal R., A note on direct adaptive control of systems with bounded disturbances // Automatica. — 1987.-23. — № 2. — P. 253-254. 10. Fogel E., Huang Y. F. On the value of information in system identification — bounded noise case // Automatica. — 1982. — 18. — № 2. — P. 229-238. 11. Lozano-Leal R., Ortega R. Reformulation of the parameter identification problem for system with bounded disturbances // Automatica. — 1987. — 23. — № 2. — P. 245-257. 12. Canudas de Wit C. C., Carrilo J. A modified EW — RLS algorithm for systems with bounded disturbances // Automatica. — 1990. — 26. — P. 599-606. 13. Фомин В. Н. Математическая теория обучаемых опознающих систем. — Л.: Изд. ЛГУ, 1976. — 235 с.

*Поступила в редколлегию 18.04.2008*

УДК 004.032.26(043)



## ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ ЭФФЕКТИВНОСТИ КАРТ КОХОНЕНА ДЛЯ ВИЗУАЛИЗАЦИИ ЗАКОНОМЕРНОСТЕЙ

Н.С. Лесная<sup>1</sup>, В.Б. Репка<sup>2</sup>, О.В. Ивченко<sup>3</sup>, А.В. Шерстнюк<sup>4</sup>

<sup>1</sup>ХНУРЭ, г. Харьков, Украина, lmd@kture.kharkov.ua

<sup>2</sup>ХНУРЭ, г. Харьков, Украина, victoria\_repka@kture.kharkov.ua

<sup>3</sup>ХНУРЭ, г. Харьков, Украина, ivchenko.o@mail.ru

<sup>4</sup>ХНУРЭ, г. Харьков, Украина, troll\_andrey@mail.ru

Проведён анализ способов визуализации карт Кохонена для выявления скрытых структур в данных низкой и высокой размерности. Оценена эффективность нескольких способов визуализации карт Кохонена. Доработаны и программно реализованы способы визуализации карт Кохонена. Исследована возможность применения нейронной сети Кохонена и построенных по результатам её работы карт для сегментации двумерных и многомерных входных векторов. На основании анализа экспериментов сделан вывод о целесообразности использования определённых способов визуализации при решении задачи сегментации.

НЕЙРОННАЯ СЕТЬ, КАРТА КОХОНЕНА, ВИЗУАЛИЗАЦИЯ, МАТРИЦА РАССТОЯНИЙ, ДИАГРАММА ХИНТОНА, КЛАСТЕРИЗАЦИЯ

### Введение

Визуализация данных — задача, с которой сталкивается в своей работе любой исследователь. К задаче визуализации данных сводится проблема представления в наглядной форме данных эксперимента или результатов теоретического исследования. Традиционные инструменты в этой области — графики и диаграммы — плохо справляются с задачей визуализации, когда возникает необходимость изобразить более трёх взаимосвязанных величин [1].

На сегодняшний день одним из самых используемых способов визуализации являются самоорганизующиеся карты Кохонена, с помощью которых возможно эффективно визуально отобразить скрытые закономерности в исследуемых данных. Существует множество способов визуализации карт Кохонена, таких как унифицированная матрица расстояний, проекция Саммона, матрица плотности попадания, матрица кластеров, матрица ошибок квантования и другие, каждый из которых позволяет наблюдать те или иные свойства анализируемых данных.

Из-за большого разнообразия способов визуализации становится актуальной задача экспериментального исследования эффективности их работы на разных наборах данных, с разной размерностью и с разными структурами. Таким образом, ставится задача исследования различных методов визуализации карты Кохонена при анализе картой различных выборок, для определения степени применимости конкретных способов визуализации в конкретной рассматриваемой задаче.

### 1. Особенности нейронных сетей Кохонена (двумерный случай)

Нейронная сеть Кохонена (самоорганизующаяся карта/Self-organizing map/SOM) относится к самоорганизующимся сетям, которые при пос-

туплении входных сигналов, в отличие от сетей, использующих обучение с учителем, не получают информации о желаемом выходном сигнале. В связи с этим невозможно сформировать критерий настройки, основанный на рассогласовании реальных и требуемых выходных сигналов ИНС. Все предъявляемые входные сигналы из заданного обучающего множества самоорганизующаяся сеть в процессе обучения разделяет на классы, строя так называемые топологические карты [3].

Существует несколько алгоритмов обучения сети Кохонена, отличающиеся способом определения выигравших нейронов, а также различные модификации этих алгоритмов. В работе использован алгоритм обучения, описанный в [4]. После окончания процесса обучения карта Кохонена классифицирует входные примеры по группам схожести, что можно отобразить цветами на карте.

Для визуализации карт Кохонена могут быть использованы 1-, 2- и 3-мерные пространства, но обычно практически ограничиваются отображением с помощью 2-мерных поверхностей, так как именно в таком виде человек воспринимает геометрические структуры наиболее естественно, и отношения между объектами выглядят наиболее наглядно.

Под визуализацией данных картой Кохонена понимается такой способ представления многомерного распределения данных на двумерной плоскости, при котором качественно отражены основные закономерности, присущие исходному распределению: его кластерная структура, внутренние зависимости между признаками, информация о расположении данных в исходном пространстве и другие. Карта признаков, полученная алгоритмом SOM, является топологически упорядоченной в том смысле, что пространственное положение нейронов в решётке соответствует конкретной

области или признаку входного образа [4]. Две точки, близко лежащие на карте Кохонена, будут близки и в  $N$ -мерном входном пространстве, но не наоборот.

В результате обучения сети Кохонена строится совокупность карт, каждая из которых представляет двумерную сетку узлов, размещенных в многомерном пространстве. Самый простой вариант — использование градаций серого цвета. Также для раскраски можно использовать любую иную градиентную палитру.

Рассмотрим наиболее наглядные варианты визуализации карт Кохонена на примере с 2-мя входными характеристиками и многомерный случай. Для проведения экспериментов была использована выборка «Spirals» (рис. 1), которая представляет собой две спирали (два линейно неразделимых класса) и состоит из 2-мерного вектора входных данных и 194-х базовых точек. Набор данных «Spirals» был выбран для описания визуализации карт Кохонена, так как его отображение на картах даёт наглядные и простые для понимания результаты.

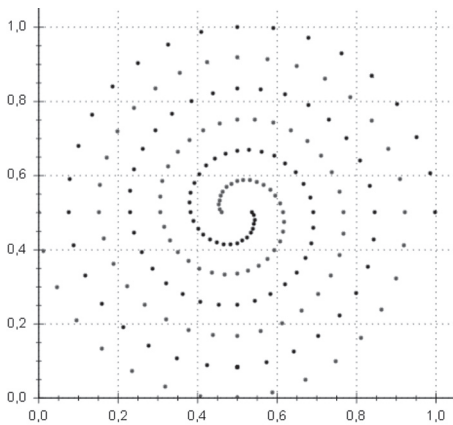


Рис. 1. Выборка «Spirals»

На данной выборке рассмотрим применение карт Кохонена для решения задачи кластеризации, то есть представим, что число классов заранее неизвестно. Приведем результаты, полученные за 100 проходов нейронной сети Кохонена. В качестве функции соседства выбрана функция Гаусса, радиус в начале обучения — 3, норма обучения — 0,5.

## 2. Способ визуализации посредством унифицированной матрицы расстояний

Унифицированная матрица расстояний (U-matrix) — это представление SOM, визуализирующее расстояние между нейронами. Рассчитывается расстояние между соседними нейронами и представляется в виде различной раскраски между соседними вершинами. Темная раскраска соответствует большему расстоянию между нейронами и соответственно — промежуткам между величинами кодовой книги (codebook) в пространстве входных векторов. Светлая раскраска между нейронами оз-

начает, что векторы кодовой книги (codebook vectors) расположены близко друг к другу в пространстве входных векторов [5]. Светлые области могут считаться кластерами, а темные — разделителями между ними. Это может быть полезным представлением при попытке найти кластеры во входных данных без наличия априорной информации о кластерах.

Для построения матрицы расстояний необходимо определить расстояние между весовыми коэффициентами нейрона и его ближайшими соседями, как показано на рис. 2.

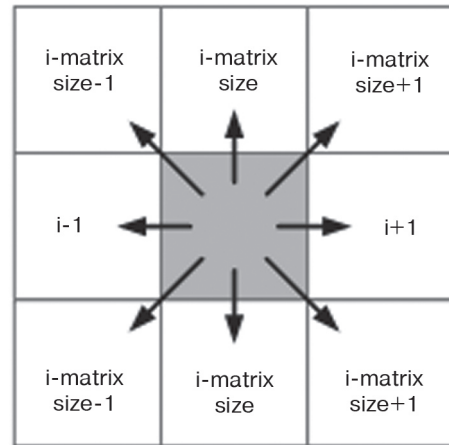


Рис.2. Ближайшие соседи  $i$ -го нейрона

В качестве меры близости был выбран квадрат евклидова расстояния (1):

$$d_{pq} = \sum_i^n (w_{pi} - w_{qi})^2, \quad (1)$$

где  $d_{pq}$  — квадрат евклидова расстояния между нейронами  $p$  и  $q$ ;  $w_{pi}$  —  $i$ -й весовой коэффициент нейрона  $p$ ;  $w_{qi}$  —  $i$ -й весовой коэффициент нейрона  $q$ ;  $n$  — число весовых коэффициентов нейрона.

Оттенок ячейки, расположенной между двумя узлами, отражает расстояние между узлами в исходном пространстве. Оттенок самого узла вычисляется с помощью усреднения. Оттенки раскраски карты задаются в соответствии с формулами 2-4:

$$R' = 255 \cdot \left( \frac{R - R_{\min}}{R_{\max} - R_{\min}} \right), \quad (2)$$

$$G' = 255 \cdot \left( \frac{G - G_{\min}}{G_{\max} - G_{\min}} \right), \quad (3)$$

$$B' = 255 \cdot \left( \frac{B - B_{\min}}{B_{\max} - B_{\min}} \right), \quad (4)$$

где  $R, G, B$  — составляющие цветовой модели, а  $\min$  и  $\max$  — минимальное и максимальное расстояния на карте Кохонена.

На рис. 3 приведена матрица расстояний, отображенная в оттенках серого, которая соответствует карте размером  $15 \times 15$  (225 нейронов). SOM полностью покрывает все точки выборки, и 194 узла

на карте соответствуют расположениям данных точек. Светлые участки соответствуют точкам, находящимся близко друг от друга; в центре изображения чётко видны две спирали, по мере удаления от центра, спирали становятся нечёткими, так как увеличивается расстояние между точками, и ближайшими соседями для самых дальних точек будут ближайшие точки другой спирали. Можно сделать вывод, что входная выборка представляет собой два класса, границы которых чётко видны в центре изображения.

Карта, матрица расстояний для которой приведена на рис. 3, построена за 500 проходов алгоритма обучения, приемлемый результат достигается уже на 100-м проходе – видны спирали, но изображение не ровное.

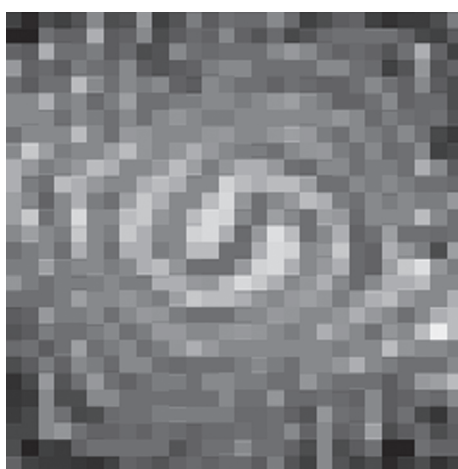


Рис. 3. Матрица расстояний

На рис. 4 приведены матрицы расстояний для карт разного размера:  $7 \times 7$ ,  $15 \times 15$  и  $25 \times 25$ . Похоже выглядят U-матрицы для карт размером  $15 \times 15$ ,  $30 \times 30$  и  $50 \times 50$ , если из их алгоритма раскраски убрать ячейки, расположенные между узлами, которые отображают расстояния между ними. Если карта больше выборки, ячейки карты будут заполнены неактивными нейронами, которые будут заменять эти ячейки с подобными оттенками, если меньше – то оттенки узлов будут отображать расстояния. Визуализация, реализованная подобным образом, является более чёткой и может быть использована в случаях, когда входная выборка приемлемого размера и увеличение размера карты в 2 раза незначительно скажется на скорости обучения нейронной сети.

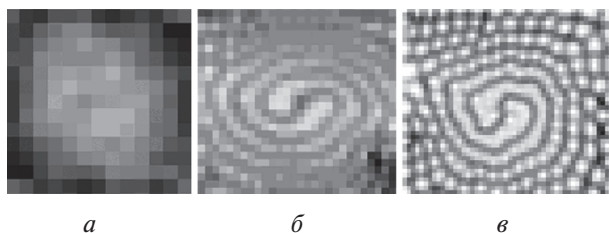


Рис. 4. Матрицы расстояний для карт Кохонена разного размера:  $a - 7 \times 7$ ;  $b - 15 \times 15$ ;  $v - 25 \times 25$

Точность визуализации и дальнейшего анализа зависят от размера карты. Это обусловлено тем, что во время переноса точек из пространства на карту каждой точке сопоставляется ближайший узел карты. Это приводит к тому, что средний квадрат расстояний от точки до её проекции на карте сильно зависит от количества узлов в сетке. На первой карте (рис. 4а), сетка которой состоит из наименьшего числа узлов, видно, что в центре изображения точки расположены близко друг другу; чем дальше от центра – расположение становится более разреженным, что также характерно и для входного вектора. Размер последней карты (рис. 4в) слишком велик, и в отдельные кластеры начинают отделяться дальние точки спиралей, что обусловлено в данной ситуации избытком неактивных нейронов.

Одной из проблем построения карт Кохонена является топологический дефект. Если изначально параметр расстояния  $\sigma$  выбран малым или очень быстро уменьшается, то далеко расположенные нейроны не могут влиять друг на друга [2]. Результатом этого является топологический дефект карты Кохонена, приведённый на рис. 5а.

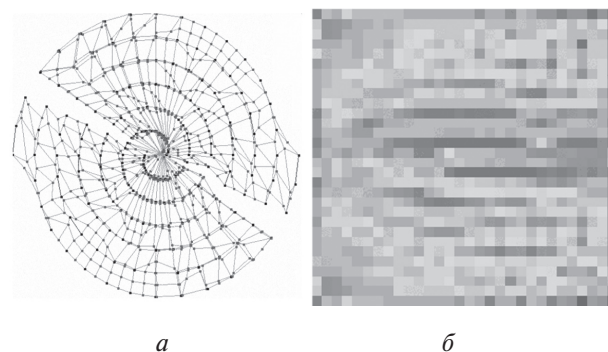


Рис. 5. Топологический дефект карты Кохонена:  
 $a$  – дефект карты Кохонена;  
 $b$  – влияние дефекта на U-Matrix

Как видно на рис. 5а, обе части карты настроились правильно, но на ней явно виден дефект, который отрицательно сказывается на матрице расстояний, построенной по данной карте (рис. 5б). Такая карта, в зависимости от степени дефекта, не может быть интерпретирована корректно.

### 3. Визуализация диаграммой Хинтона

Популярным способом визуализации является диаграмма Хинтона. На каждом узле сетки изображается квадрат, размер которого пропорционален числу точек, ближайших к данному узлу, а оттенок соответствует значению соответствующего отображаемого признака [1].

На рис. 6 приведена диаграмма Хинтона для карты размером  $7 \times 7$ . Данная сетка представляет собой карту, подходящую для отображения двумерного и многомерного пространства данных.

Чем больше точек попадает в узел, тем больше соответствующий размер квадрата. Оттенок квадрата соответствует значению цвета на матрице расстояний для данного нейрона.

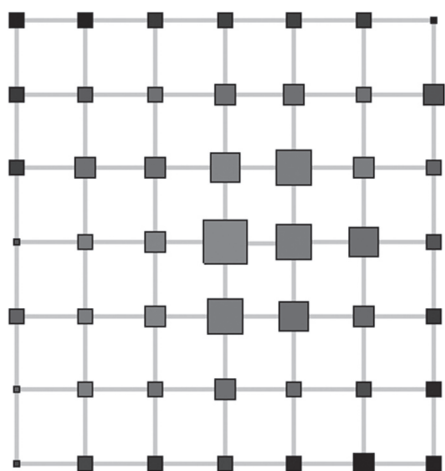


Рис. 6. Диаграмма Хинтона для «Spirals»

На рис. 7 приведена сетка карты Кохонена, соответствующая диаграмме Хинтона, и исходный набор данных, который она покрывала, то есть точки исходного набора были распределены по узлам карты.

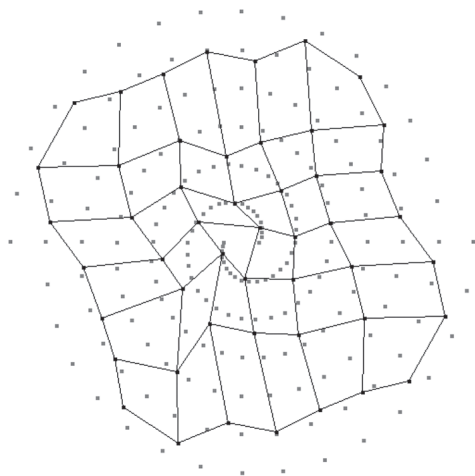


Рис. 7. Карта Кохонена 7×7

В выборке «Spirals», приведённой на рис. 1, в центре наблюдается сгущение точек, поэтому в центре диаграммы Хинтона находятся узлы большего размера, а по краям – меньшего.

Также частоту попадания точек в узел можно отобразить на каждой эпохе, как показано на рис. 8.

Карты, показанные на рис. 8, соответствуют картам частот. Они построены в соответствии с количеством реагирований нейронов-победителей на данные из обучающего множества на разных эпохах. На данных картах размер круга соответствует числу реагирований нейрона за эпоху.

Диаграмма Хинтона может соответствовать матрице плотности попадания.

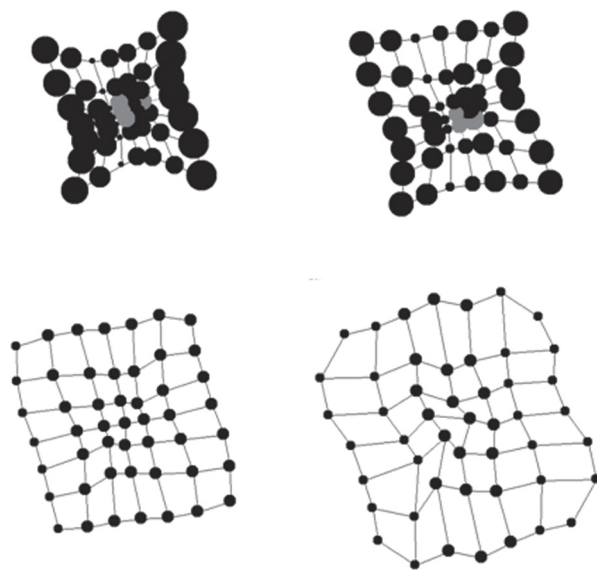


Рис. 8. Карты Кохонена 7×7 на разных эпохах

#### 4. Применение карт Кохонена для многомерной кластеризации

Рассмотрим применение карт Кохонена для выявления скрытых закономерностей в данных высокой размерности на примере решения задачи кластеризации кредитных историй и информации о заемщиках.

Набор исходных данных состоит из 150 образов. Каждый образ представляет собой информацию о заемщиках и выданных кредитах и состоит из 31 характеристики (сумма кредита, срок кредита, цель кредитования, возраст, пол, образование, квартира и другое).

Рассмотрим визуализацию карт Кохонена и определение кластеров на выборке из 146 образов, после чего классифицируем 4 образа с помощью обученной НС.

На рис. 9 и рис. 10 показаны матрицы расстояний и соответствующие им диаграммы Хинтона.

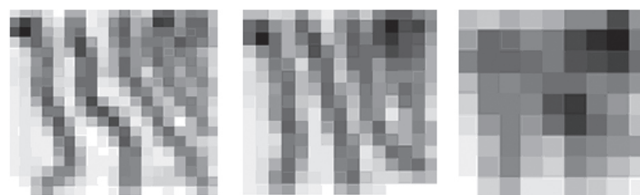


Рис. 9. Матрицы расстояний для карт разного размера

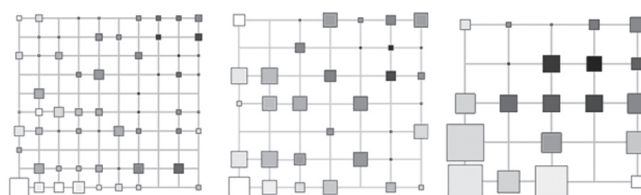


Рис. 10. Диаграммы Хинтона для карт разного размера

По картам на рис. 9 и 10 можно сделать следующие выводы. На картах большего размера видны 3-4 области сгущения точек, отображённые светлым цветом, каждая из которых плавно переходит в другую, они соответствуют областям в  $N$ -мерном входном пространстве.

На всех диаграммах, приведённых на рис. 10, есть нейроны, которые ни разу не среагировали на данные из обучающей выборки, в то же время, другие нейроны среагировали несколько раз, что говорит о том, что в исходном обучающем множестве есть очень схожие образы. В табл. 1 приведены результаты реагирования нейронов на обучающую выборку для карт разного размера.

Таблица 1

Размер карты	% среагировавших нейронов	% не среагировавших нейронов
10 × 10	60	40
7 × 7	69,4	30,6
5 × 5	84	16

По результатам, приведенным в табл. 1, видно, что число не среагировавших нейронов с уменьшением размеров карты уменьшается незначительно, пока карта не достигает слишком малого размера, например, 25 нейронов.

Узлы диаграммы Хинтона также можно представить в виде кластеров, тогда образы, попавшие в один узел, будут соответствовать одному классу. Ни рис. 11 приведена диаграмма Хинтона, представляющая собой 4 узла, то есть отображающая 4 класса. Оттенки узлов соответствуют степени удаленности их друг от друга, а также сгущению точек, то есть самый удалённый узел карты также соответствует и наименьшему числу образов, попавшим в него.

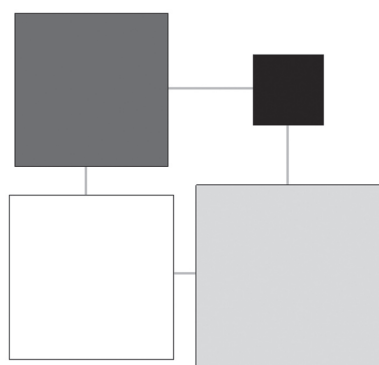


Рис. 11. Диаграмма Хинтона для карты Кохонена с 4-мя нейронами

В табл. 2 приведено число попаданий в каждый узел для диаграммы, приведённой на рис. 11.

В табл. 3 приведены наиболее весомые усреднённые значения по результатам сегментации нейронной сетью Кохонена информации о заемщиках в банках. По рис. 11 и табл. 2 и 3 можно сделать

следующие выводы: самым небольшим по числу принятых образов является четвёртый сегмент, он удалён от трёх других кластеров и точки, находящиеся в нём, разрежены. Кроме того, в нём находится наименьшее количество образов. По результатам табл. 3 видна аналогичная ситуация: это лица в возрасте свыше 60 лет – пенсионеры, что вполне соответствует небольшому числу таких заемщиков, у которых есть множество расхождений в характеристиках, и эти характеристики почти не пересекаются с характеристиками других классов.

Таблица 2

Номер узла	Цвет	Число отнесённых образов
1	Тёмно-серый	39
2	Светло-серый	47
3	Белый	42
4	Чёрный	18

Таблица 3

Характеристики	Усреднённые значения свойств сегментов			
	Сегмент 1	Сегмент 2	Сегмент 3	Сегмент 4
Возраст	до 22 лет	от 22 до 40	от 40 до 60	свыше 60
Семейное положение	холост (не замужем)	женат (замужем)	женат (замужем)	разведён(а), вдовец(а), женат (замужем)
Количество иждивенцев	0	2-3	3-5	до 2
Социальный статус	студент	не руководящий работник, руководящий работник	руководящий работник	пенсионер
Образование	среднее, неоконченное высшее	высшее, специальное	высшее	среднее, высшее
Занятость	нет	да	да	нет
Квартира	нет	да	да	да
Машина	нет	нет	да	да, нет
Среднемесячный доход	до 500	от 1500	от 2000	до 500
Среднемесячный расход	до 400	от 1000	от 1000	до 400
Основное направление расходов	образование, аренда недвижимости, оплата услуг	покупка товаров длительного пользования	покупка и ремонт недвиж, покупка товаров длител. пользования	оплата услуг
Цель кредита	образование, потребительские товары, иное	потребительские товары, транспорт, недвижимость	транспорт, недвижимость	потребительские товары, иное

Наиболее приближен к нему первый – тёмно серый сегмент – лица в возрасте до 22 лет. Приближение данных сегментов обусловлено значи-

тельным отдалением от других сегментов, также у них есть схожие характеристики, отличные от двух других сегментов, такие как: небольшой средне-месячный доход и расход, незанятость. Данные сегменты не очень схожи друг с другом, их приближение обусловлено значительным отдалением от других сегментов. К первому сегменту менее всего приближен третий сегмент, характеристики этих сегментов значительно отличаются.

Самые крупные по числу отнесённых к ним образов второй – светло серый и третий – белый сегменты, расположенные наиболее близко друг к другу, это лица в возрасте от 22 до 40 лет и от 40 до 60 соответственно. Также по результатам, приведенным в табл. 3, эти сегменты содержат образы с наиболее схожими характеристиками, такими как семейное положение, количество иждивенцев, цели кредита, среднемесячный доход и расход и другие. Расхождения в характеристиках третьего сегмента меньше, чем в трёх других сегментах.

Результаты табл. 3 также соответствуют построенным матрицам расстояний, приведённым на рис. 9, то есть плавные переходы цветов от одного класса к другому соответствуют общим характеристикам для данных сегментов.

Как показали результаты проведенных исследований, диаграмма Хинтона хорошо отображает разделение исходного множества на известное число сегментов, в данной ситуации – 4, что подтверждают результаты, приведенные в табл. 3. Карты Кохонена как аппарат интеллектуального анализа могут применяться в банковских системах для сегментации клиентов и многих других областях человеческой деятельности, где возникают задачи кластеризации и классификации, и необходима визуализация более чем двумерного входного пространства.

### Выводы

В работе рассмотрены различные способы визуализации нейросетевых карт Кохонена для выявления скрытых структур в данных низкой и высокой размерности. Проведены экспериментальные исследования построения карт Кохонена для двумерных данных на примере набора данных с координатами двух спиралей. Применены способы визуализации посредством унифицированной

матрицы расстояний и визуализация диаграммой Хинтона.

Экспериментальные исследования, проведённые для оценки качества визуализации, с помощью матрицы расстояний показали, что более эффективным является построение карты большего размера, чем добавление ячеек для отображения расстояния между нейронами при раскраске карты. Полученные результаты позволили сделать заключение о том, что при увеличении размера топологической карты свободные ячейки заполняются неактивными нейронами, которые фактически заполняют ячейки, визуализирующие расстояние, или то же отображают активные нейроны в зависимости от размера карты и распределения образов по узлам.

Также проведен эксперимент по построению и визуализации карт Кохонена для многомерного набора данных из области банковского кредитования. Данные о клиентах были разделены сетью на 4 группы и отображены посредством диаграммы Хинтона.

Полученные результаты свидетельствуют об эффективности применения аппарата нейронных сетей Кохонена для визуализации данных различной размерности, с различными структурами.

Перспективы и дальнейшие исследования направлены на применение средств визуализации к другим нейронным сетям, решающим аналогичные задачи. Будет рассмотрена возможность применения данной нейронной сети в качестве модуля аналитической системы для принятия решений при кредитовании в банке, а также разработаны другие карты для визуализации работы нейронной сети Кохонена.

**Список литературы:** 1. *Зиновьев А.Ю.* Визуализация многомерных данных. — Красноярск: Изд-во КГТУ, 2000. — 168 с. 2. *Бодянский Е.В., Руденко О.Г.* Искусственные нейронные сети: Уч. пособие. — Харьков: ООО «Компания СМИТ», 2005. — 408 с. 3. *Калан Р.* Основные концепции нейронных сетей.: Пер. с англ. — М.: Изд. дом «Вильямс», 2003. — 288 с. 4. *Хайкин С.* Нейронные сети: полный курс, 2-е изд. — М.: ООО «И.Д. Вильямс», 2006. — 1104 с. 5. *Kohonen T.* Self-organizing maps / Teuvo Kohonen. — 3. ed. — Berlin: Springer-Verlag Berlin Heidelberg, 2001. — 500 p.

*Поступила в редколлегию 18.04.2008*

УДК 519.7:007.52



## КОМБИНИРОВАННЫЙ ПОДХОД К КЛАССИФИКАЦИИ ТЕКСТОВ

Т.Б. Шатовская<sup>1</sup>, И.В. Каменева<sup>2</sup>, Ю.А. Тарасов<sup>3</sup>

<sup>1</sup>ХНУРЭ, г. Харьков, Украина, tanita\_uk@mail.ru

<sup>2</sup>ХНУРЭ, г. Харьков, Украина, iriska@vk.kh.ua

<sup>3</sup>ХНУРЭ, г. Харьков, Украина, iceman.yt@gmail.com

В данной статье мы представляем сравнительный анализ методов и интегрированный иерархический подход классификации текста основанный на интегрированном подходе используемых дендрограммы и *k*-средних кластеризации. Этот подход позволяет нам представить интегрированный новый метод иерархической кластеризации, который может классифицировать данные без предварительного задания количества классов.

КЛАСТЕРИЗАЦИЯ, ТЕКСТОВАЯ КЛАССИФИКАЦИЯ, АЛГОРИТМ *K*-СРЕДНИХ, МЕТОД ДЕНДРОГРАММЫ, МЕРА СХОЖЕСТИ, ВЕКТОРНАЯ МОДЕЛЬ, СТЕММИНГ, СТОП-СЛОВА

### Введение

В настоящее время классификация текста является одной из актуальных научно – исследовательских проблем. Методы классификации текстов применяются в фильтрации документов, распознавании спама, автоматическом аннотировании, снятии неоднозначности (автоматические переводчики), составлении Интернет-каталогов, классификации новостей, распределении рекламы, в персональных новостях. В большинстве систем классификации используется традиционный интегрированный подход на основе двух методов: *k*-средних и байесовского классификатора. Эти методы работают с учителем (supervised learning) и требуют непосредственного участия экспертов в процессе решения задачи классификации.

### 1. Постановка задачи

На сегодняшний день одной из наиболее важных социальных проблем в Украине является формальная безработица среди молодых людей. Даже после получения высшего образования молодые профессионалы достаточно редко могут найти работу, которая совпадает с их специальностью. В частности, сложно найти подходящее место даже после окончания технического вуза. В Харьковском национальном университете радиоэлектроники был создан информационный «Центр-Карьера», который повышает внутренние и внешние уровни информации, оказывая помощь молодым людям найти высококвалифицированную работу после окончания вуза. Была создана Информационная система, которая помогает образовательной инфраструктуре поддерживать информационную взаимосвязь между университетом и индустрией (<http://rabota.kture.kharkov.ua>). Ежедневно в данную информационную систему поступает большое количество информации от компаний. Такие как: вакансии, новости, объявления и так далее. К сожалению, большая часть полученной информации

не структурирована, и вакансии, высланные на e-mail — свободного содержания. Высокая популярность этой системы среди компаний позволяет в обязательном порядке ежедневно непосредственно принимать большое количество текстовых документов. Более того, часто составленные описания вакансий не структурированы, и документы приходится относить к определенным классам инженера, а это достаточно сложно.

Например: «Инженер-электроник со знаниями MSSQL и Ассемблера». В данной системе в основе стандартные методы классификации текста относят набор вакансий к двум или более главам.

Мы предлагаем интегрированный метод текстовой классификации, используя экспертную оценку (рис. 1).



Рис. 1. Набор вакансий

## 2. Сравнение подходов text Mining

Существует огромное количество классификаторов. Чаще всего применяются Naive Bayesian метод классификации [1, 2], Expectation Maximization (EM) алгоритм [2], Support Vector Machines (SVM) [3] и другие. Все эти подходы обучаются с учителем. В методе обучения с учителем обучающие выборки маркированных документов используются для обучения алгоритма, чтобы построить классификатор. В конечном итоге текстовая классификация определяет документы в одну или более предопределённых категорий. В классификации текста новый текстовый документ назначается в один из уже существующих наборов документов класса.

Текстовая классификация — это лингвистическая и реляционная технология, используемая с целью анализа доступных документов в результате набора полученных ключевых термов из статей и абстрактов. Результаты могут быть представлены графически с картой, которая обеспечивает краткий обзор кластеров и указание взаимосвязи среди них. Главная идея — найти документы, которые имеют много общих слов, и отнести документы с большим количеством общих слов в схожие группы. Текстовая классификация — это автоматическая организация документов в пределах кластера, которые имеют высокое сходство в сравнении друг с другом, но очень несходны с документами в других кластерах [4].

С помощью автоматической классификации мы даем возможность машине решить, к какой из встроенных категории принадлежит текст. В кластеризации машина решает, как предоставленный текст должен быть разделен. Классификация применяется тогда, когда новые тексты классифицируются согласно известной классификации, а кластеризация применяется тогда, когда обнаруживаются новые заранее неизвестные структуры.

### 3. Иерархическая кластеризация и неиерархическая кластеризация

Иерархическая кластеризация [6] — процесс организации данных в древовидную структуру, основанной на их сходстве. Этот метод очень мощный и полезный для анализа больших наборов данных. Основная идея — создать набор элементов в дереве. Дерево имеет много ветвей. Если элементы подобны друг другу, к ним присоединяются короткие ветви, и наоборот, если их схожесть уменьшается, тогда увеличиваются ветви.

Задача кластеризации текстов состоит в следующем. Имеется некоторое количество текстов. Необходимо сгруппировать эти тексты в соответствии со схожестью их стилей. Такая группировка может быть как одноуровневой («плоской», с выделением таких кластеров, что каждый объект в них является одним из текстов, представленных в наборе, клас-

теризацию), так и иерархической, когда кластеры, полученные в результате объединения наиболее похожих текстов, сами могут объединяться в кластеры, а кластеры кластеров — в другие кластеры и так далее. Принадлежность текста к некоторому кластеру на определенном уровне иерархической кластеризации может быть однозначной (каждый рассматриваемый текст принадлежит только одному кластеру), или неоднозначной (каждый рассматриваемый текст может принадлежать нескольким кластерам).

Кластеризация документов была использована, чтобы автоматически генерировать иерархические кластеры документов [7].

Результат иерархической агломеративной кластеризации может быть графически представлен как дерево под названием дендрограмма. Алгоритм агломеративной кластеризации [8] порождает кластеры, последовательно соединяя операции. Процесс агломерации начинается с инициализации каждого вектора данных как его собственного кластера. Два кластера соединяются на каждом шаге, и процесс повторяется до тех пор, пока не будет получено желаемое количество кластеров. Если мы рассматриваем общую проблему кластеризации, то существует много различных правил слияния. Различные правила слияния предоставляют различные кластерные решения, и таким образом стратегии принимают различные кластерные формы. Метод single linkage (SL) определяет, что кластерная пара соединяется, основываясь на двух самых близких векторах. Метод complete linkage (CL) определяет, что кластер, который соединен, основан на двух ближайших векторах. Вообще, как single-, так и complete-link подходы не очень хорошо работают, потому что они также основываются на собственных решениях, а следовательно и на ограниченном количестве информации (single-link), или они предполагают, что все документы в кластере подобны друг другу (complete-link). Агломеративные алгоритмы всегда детерминированно генерируют схожую кластерную иерархию (таблица).

Сравнение HAC алгоритмов

Method	Combination similarity	Time compl	Optimal	Comment
Single-link	Max sim of any two docs	$O(N^2)$	Yes	Chaining effect
Complete-link	Min sim of any two docs	$O(N^2 \log N)$	No	Sensitive to outliers

Неиерархическая кластеризация — это процесс монотонно возрастающего ранжирования, поскольку становятся членами больших кластеров. Эти кластерные методы не обладают структурами, подобными дереву, и новые кластеры формируются в последовательной кластеризации или слиянием и разделением кластеров.

Одним из неиерархических методов кластеризации являются методы разбиения [4]. Рассматривается набор кластеров как объектных методов и методов разбиения объектов, чтобы получить требуемые кластеры. В отличие от иерархического метода, этот метод разбиения позволяет объектам изменять группу с помощью процесса кластерного образования. Метод разбиения обычно начинается с начального решения, после которого перераспределение происходит согласно некоторому оптимальному критерию.

Параметр  $k$  определяет пользователь, поэтому лучше запустить алгоритмы несколько раз, чтобы выбрать самый лучший параметр  $k$ . Также возможно генерировать значение  $k$  автоматически, а затем выбирать лучший по проверенным критериям. Более популярный метод разбиения — алгоритм  $k$ -средних. Алгоритм  $k$ -средних — один из самых простых методов обучающих алгоритмов, который обучается без учителя. Алгоритм  $k$ -средних имеет входной параметр,  $k$ , и делит набор  $n$  объектов в  $k$  кластерах таким образом, что результирующее внутрикластерное сходство высоко, но межкластерное сходство низкое. Сначала метод случайным образом выбирает  $k$  объектов, каждый из которых изначально представляет центроиды. Каждый оставшийся объект приписывается к кластеру, к которому он является самым ближним, основываясь на расстоянии между объектом и центроидом. Затем считается новый центроид для каждого кластера. Этот процесс повторяют, пока оценочная функция не сводится в одну точку [5].

#### 4. Методы текстовой кластеризации

На сегодняшний день векторная модель является широко используемой моделью представления данных для классификации и кластеризации документов. Общая структура этой модели данных начинается с представления любого документа как вектор слов, которые появляются в документах набора данных. Вес (обычно частоты термов) слов также содержится в каждом векторе. Схожесть между двумя документами считается на основании

двух соответствующих по свойствам векторов, например, Jaccard measure, и Euclidean distance [9]. Мы использовали cosine measure. Когда описание вакансий получено нашей вэб системой, мы используем метод предварительной обработки. Предварительная обработка — это сокращение текста для более точной классификации. С обработкой методов различные документы могут быть созданы как структурированные представления документа [10]. Обычно задачи предварительной обработки действий включают стандартизацию документа, токенизацию, лематизацию и стемминг [8]. Технология этого процесса рассмотрена на рис. 2.

Мы используем stop-list для исключения стоп-слов. Stop-list — это словарь стоп-слов, который имеет низкую частоту (a, the, and, и так далее) для документов. Стоп-слова всегда удаляют из документов перед преобразованием к векторной модели. Мы используем готовый stop-list. В следующем шаге мы осуществляем stemming алгоритм. Стемминг — это алгоритм, который помогает нам определить значимые части слов. Стемминг удаляет из слов суффиксы рекурсивно. Процесс имеет две цели. В терминах эффективности стемминг сокращает число уникальных слов в индексе, который в свою очередь сокращает пространство памяти, требуемое для индекса, и развивает скорость поискового процесса. В терминах эффективности стемминг улучшается, сводя слова к базовой форме. В нашей работе мы используем стемминг-алгоритм Портера и готовый словарь. Это помогает сократить словарь так, чтобы увеличить скорость нашего метода и улучшить качество алгоритма. На третьем шаге каждый документ использует новую векторную модель tf.idf-exprt. В стандартной векторной модели [10], полагается, что каждый документ является вектором в пространстве термов. В его стандартной форме каждый документ представляется вектором частоты (TF)  $df = (tf_1, tf_2, \dots, tf_n)$ , где  $tf_i$  — частота  $i$ -й строки в документе. Обратная частота (IDF) документа в частоте документов  $i$  на  $\log(n/df_i)$ , где  $n$  — полное число документов в выборке, и  $df_i$  — число документов, которые содержат

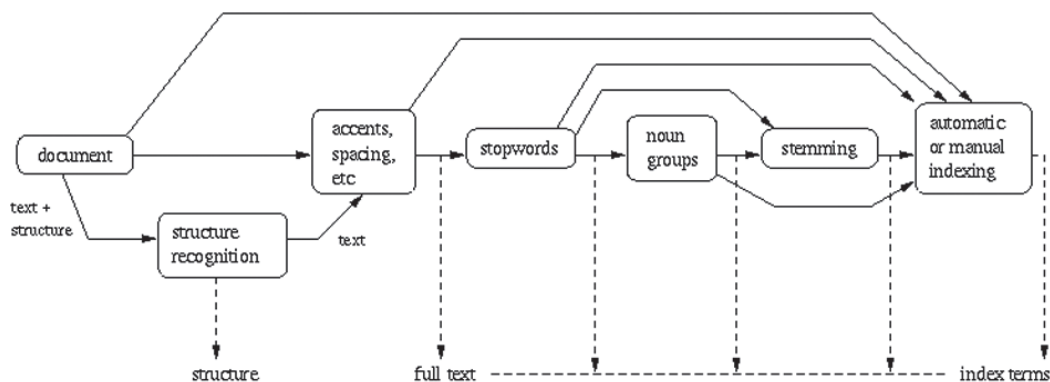


Рис. 2. Последовательность предварительной обработки документа

$i$ -ый терм (то есть, частота документа). Наконец, чтобы посчитать длину каждого документа, необходимо каждый вектор документа нормализовать от 0 до 1, то есть  $\|d_{fidf}\|_2 = 1$ . Мы также добавляем экспертную оценку для каждого термина, потому что для вакансий очень важно оценить значение каждого термина в документе. Мы использовали  $tf \times idf \times \exp(-f)$  как

$$d_{fidf} = tf_1 \log(n/df_1) \cdot f, tf_2 \log(n/df_2) \cdot f, \dots, tf_n \log(n/df_n) \cdot f.$$

Все параметры остаются без изменения, но  $f$  — это экспертная оценка, где  $f = 0 \dots 1$ . Для подсчета схожести между документами часто применяется cosine measure, которое определяется как

$$\text{cosine}(d_i, d_j) = \frac{\langle d_i, d_j \rangle}{\|d_i\|_2 \times \|d_j\|_2} = \frac{\sum_{i=1}^t d_i \times d_j}{\sqrt{\sum_{i=1}^t d_i^2} \times \sqrt{\sum_{i=1}^t d_j^2}},$$

где  $d_i$  и  $d_j$  — компоненты векторных документов;  $t$  — размерность вектора.

Расстояние между векторами  $d_1$  и  $d_2$  представляется как косинус угла между ними (рис. 3).

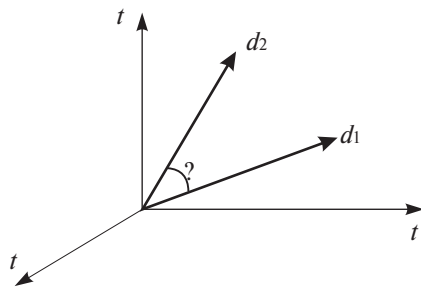


Рис. 3. Схожесть между  $d_1$  и  $d_2$

Мера расстояния для векторов удовлетворяется следующими свойствами:

- если два вектора совпадают целиком их сходство должно быть максимально, то есть равняться 1;
- если два вектора не имеют никаких ключевых общих слов, то есть, если вектор запроса не имел положительных весов документа, вектор имеет вес 0, и наоборот — или другими словами, если векторы ортогональны — сходство должно быть минимально, то есть равняться 0.

Во всех других случаях схожесть должна быть от 1 до 0. Документы, которые соединены друг с другом в векторном пространстве показывают схожесть между ними, где  $x$  и  $y$  это cosine measure (схожесть) между документами (рис. 4).

На четвертой стадии кластеризации, на базе Ward's linkage правила и cosine measure, мы конструируем дендрограмму, чтобы получить начальное разделение входных документов. Так как мы не знаем предварительного количества классов в наших данных, первая дендрограмма показывает глобальное число классов.

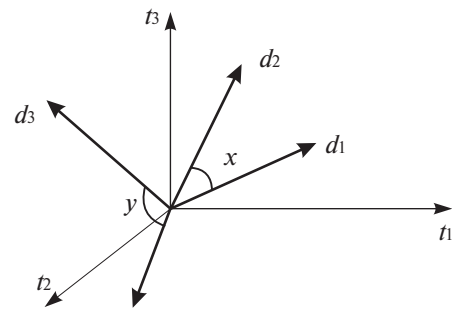


Рис. 4. Схожесть между документами

После этого мы применяем метод кластеризации  $K$ -средних к каждому классу дендрограммы. Для оптимальных результатов мы использовали меру компактности кластеров. После этого кластеризация дендрограммы будет применена к каждому подкластеру и общее количество будет оценено. Как заключительный результат мы получаем кластеры и все их подкластеры. Этот подход применяется непосредственно к вакансиям, группирующимся среди 20 категорий, и частота появления ошибок составила 8% из-за некорректного группирования.

### Заключение

Комбинированный подход использования итеративной кластеризации и агломеративных методов используются для кластеризации вакансий веб-сайта. Векторная модель использовалась для классификации документа и кластеризации. Главными шагами предоставленного подхода являются: кластерная дендрограмма для всего текстового репозитория, затем итеративная кластеризация, которая используется для разделения на подкластеры внутри каждого класса. и заключительным шагом является кластерная дендрограмма для получения иерархической структуры целого набора данных.

**Список литературы:** [http://en.wikipedia.org/wiki/Naive\\_Bayes\\_classifier](http://en.wikipedia.org/wiki/Naive_Bayes_classifier) 2. [http://apex.sjtu.edu.cn/apex\\_wiki/dwyak?action=AttachFile&do=get&target=aaai07.pdf3](http://apex.sjtu.edu.cn/apex_wiki/dwyak?action=AttachFile&do=get&target=aaai07.pdf3). <http://jmlr.csail.mit.edu/papers/volume2/tong01a/tong01a.pdf4>. <http://www.cs.sfu.ca/~ester/papers/Encyclopedia.pdf5>. <http://www.nada.kth.se/~rosell/undervisning/sprakt/irinro060801.pdf6>. <http://www.ims.uni-stuttgart.de/lehre/teaching/2007-SS/ir/hier/hier.pdf7>. Daphe Koller and Mehran Sahami, Hierarchically classifying documents using very few words, Proceedings of the 14th International Conference on Machine Learning (ML), Nashville, Tennessee, July 1997, Pages 170-178. 8. J. Boberg and T. Salakoski. General formulation and evaluation of agglomerative clustering methods with metric and non-metric distances. Pattern Recognition, 26(9):1395-1406, September 1993. 9. [http://www.ee.ucr.edu/~barth/EE242/clustering\\_survey.pdf](http://www.ee.ucr.edu/~barth/EE242/clustering_survey.pdf) 10. [Salton et al., 1975] G. Salton, A. Wong, and C.S. Yang. A vector space model for information retrieval. Journal of the American Society for Information Science, 18(11):613-620, Nov.1975.

Поступила в редколлегию 28.04.2008

УДК 519.62



## ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ КАК МЕТОД ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В ФИНАНСОВЫХ СИСТЕМАХ

Г.Г. Четвериков<sup>1</sup>, Т.В. Лесовец<sup>2</sup>, О.В. Касала<sup>3</sup><sup>1</sup> ХНУРЭ, г. Харьков, Украина, chetvergg@kture.kharkov.ua<sup>2,3</sup> ХНУРЭ, г. Харьков, Украина, mailolya@ukr.net

В данной работе рассмотрено применение нейронных сетей в решении проблем определения жизнеспособных и склонных к банкротству предприятий и предсказания курса акций. Приведена архитектура искусственных нейронных сетей и алгоритмы их обучения. Показаны результаты работы нейронных сетей на конкретных примерах. В статье также определены направления дальнейших исследований.

НЕЙРОННАЯ СЕТЬ, ВЕКТОР КВАНТОВАНИЯ, РЕГРЕССИЯ, ПРОГНОЗИРОВАНИЕ, КЛАССИФИКАЦИЯ

### Введение

Становление теории искусственных нейронных сетей произошло под влиянием исследования функционирования и строения человеческого мозга. Стало возможным создание искусственных нейронных сетей на основе естественных.

Такие результаты были продиктованы как теоретическими, так и прикладными достижениями. Неожиданно открылись возможности использования вычислений в сферах, до этого относящихся лишь к области человеческого интеллекта, возможности создания машин, способность которых учиться и запоминать напоминает мыслительные процессы человека [1].

Лучшее понимание функционирования нейрона и картины его связей позволило исследователям создать математические модели для проверки своих теорий. В первых же работах выяснилось, что эти модели не только повторяют функции мозга, но и способны выполнять функции, имеющие свою собственную ценность. Поэтому возникли и остаются в настоящее время две взаимно обогащающие друг друга цели нейронного моделирования: первая – понять функционирование нервной системы человека на уровне физиологии и психологии и вторая – создать вычислительные системы (искусственные нейронные сети), выполняющие функции, сходные с функциями мозга. Именно эта последняя цель и находится в центре внимания этой статьи [1].

Исследования в области естественного и искусственного интеллекта позволили разработать множество моделей искусственных сетей, способных решать многие задачи. Наиболее используемыми являются: распознавание зрительных и слуховых образов; ассоциативный поиск информации и создание ассоциативных моделей; синтез речи; формирование моделей различных нелинейных и трудно описываемых математических систем, прогнозирование развития этих систем во времени;

системы управления и регулирования с предсказанием; управление работами; принятие решений и диагностика; исключаящий логический вывод [2].

### 1. Нейронные сети в прогнозировании банкротства

Риск присущ любой форме человеческой деятельности, что связано со множеством условий и факторов, влияющих на положительный исход принимаемых людьми решений. Статистика утверждает, что большая часть вновь созданных компаний разоряется в течение первых трех лет существования. Тем не менее, аналогичное случается и с компаниями, благополучно просуществовавшими и даже удерживавшими лидирующие позиции на протяжении нескольких десятков лет [3].

Современное состояние науки позволяет использовать нейронные сети в решении данной проблемы. Поскольку прогнозирование банкротства является бинарным процессом принятия решений, то целесообразно считать его задачей классификации. Причем, исходя из вышесказанного, все множество организаций разбивается на два класса: банкрот и не банкрот.

Задача классификации представляет собой задачу отнесения образца к одному из нескольких попарно не пересекающихся множеств. При решении задач классификации необходимо отнести имеющиеся статические образцы (в данной задаче это коэффициенты Альтмана) к определенным классам. Возможно несколько способов представления данных. Наиболее распространенным является способ, при котором образец представляется вектором. Компоненты этого вектора представляют собой различные характеристики образца, которые влияют на принятие решения о том, к какому классу можно отнести данный образец [4].

Нейронные сети являются универсальным средством аппроксимации функций, что позволяет использовать их в решении задач классификации. Как правило, нейронные сети оказываются наиболее

лее эффективным способом классификации, потому что генерируют фактически большое число регрессионных моделей (которые используются в решении задач классификации статистическими методами). Нейронная сеть представляет собой совокупность нейроподобных элементов, определенным образом соединенных друг с другом и с внешней средой с помощью связей, определяемых весовыми коэффициентами. В сложных практических задачах обученная нейросеть выступает как эксперт, обладающий большим опытом и способный дать ответ на трудный вопрос. Нейросетевой подход особенно эффективен в задачах экспертной оценки по той причине, что он сочетает в себе способность компьютера к обработке чисел и способность мозга к обобщению и распознаванию.

## 2. Оптимизированная нейронная сеть веторного квантования

Многие задачи связаны с необходимостью хранения, обработки и передачи больших массивов данных, что требует значительных вычислительных ресурсов и временных затрат. Если массивы данных отличаются незначительно, то существенной экономии как вычислительных средств, так и необходимого для решения задачи времени возможно достичь, используя сжатие данных или специальные методы их кодирования. При этом данные могут быть, например, разгруппированы в некоторые классы, которым присваивается свой код или индекс. Достаточно хорошо разработанные методы сжатия данных требуют для своей реализации наличие статистической информации (плотности распределения) об исследуемых процессах. Применение эффективных методов возможно, если известна частота появления образов. В условиях, когда такая информация отсутствует, наиболее эффективными оказываются искусственные нейронные сети векторного квантования [5].

Под векторным квантованием подразумевается процесс преобразования некоторого вектора  $X$  из множества  $A \in R^N$  в вектор  $W$  из множества  $B \in R^N$ , где  $M < N$ .

Векторное квантование осуществляется по методу «ближайшего соседа», причем под «ближайшим» понимается вектор, удовлетворяющий различным требованиям. Если в качестве такового выбирается вектор, находящийся от данного на минимальном евклидовом расстоянии, имеем классификатор, называемый в литературе Voronoi-классификатор. В сети векторного квантования реализован принцип «победитель получает все».

Существуют различные методы обучения сети векторного квантования (Learning Vector Quantization — LVQ): неконтролируемое обучение сети, контролируемое обучение сети (LVQ 1, LVQ 2.1, LVQ 3), оптимизированный OLVQ1. Основным

недостатком метода неконтролируемого обучения сети является то, что если начальные распределения векторов весов и входных образов не являются примерно одинаковыми, то может возникнуть ситуация, когда некоторые из нейронов никогда не станут победителями, то есть их векторы весов не будут изменяться. Основным преимуществом контролируемого обучения сети векторного квантования от рассмотренного выше является использование для каждого входного образа  $X$  желаемого соответствующего выходного сигнала [5]. Таким образом, вектор весов нейрона-победителя, который ближе всего расположен к предъявляемому входному вектору, смещается в направлении последнего, если входной вектор относится к одному с ним классу, и удаляется от него в противном случае [5]. Отличие LVQ 2.1 от LVQ 1 состоит в способе коррекции весов: вектор  $w_i$  и следующий  $w_j$  относятся к разным классам; вектор  $x$  принадлежит либо тому классу, к которому относится  $w_i$ , либо тому, к которому принадлежит  $w_j$ ; вектор  $x$  находится в некотором «окне» относительно перпендикуляра между этими двумя классами.

LVQ 3 является модификацией LVQ 1, позволяющей изменить не только веса векторов, но и границу между классами, если оба вектора принадлежат тому же классу, что и  $x$ . Метод OLVQ1 улучшает методы контролируемого обучения за счет ускорения скорости сходимости путем адаптации векторов весов и настраивания параметра обучения  $\alpha$ .

## 3. Архитектура нейронной сети

Архитектура оптимизированной нейронной сети векторного квантования (OLVQ1) представлена на рис. 1

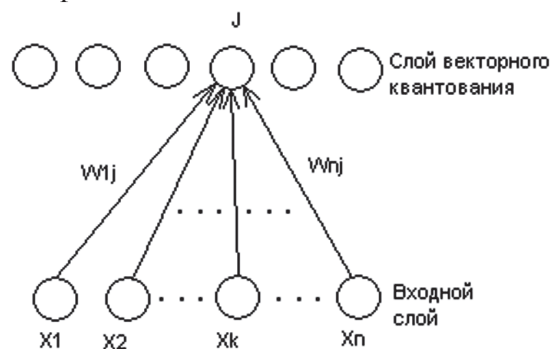


Рис. 1. Топология сети векторного квантования

Сеть состоит из двух слоев: входного слоя и слоя векторного квантования. Каждый нейрон входного слоя соединен со всеми нейронами слоя векторного квантования. Сила связи определяется соответствующим весом.

## 4. Алгоритм обучения сети

На первом этапе определяется нейрон-победитель путем вычисления евклидова расстояния, как на показано в формуле

$$\|x - w_c\|^2 = \min_j (\|x - w_j\|^2), \quad (1)$$

где  $x$  — вектор входов;  $w_c$  — вектор весов;  $w_j$  — вектор весов  $j$  нейрона Кохонена.

Веса выигравшего нейрона изменяются по следующему правилу в представленной ниже формуле [5]:

$$w_c(k+1) = \begin{cases} w_c(k) + \alpha_c(k) * [x(k) - w_c(k)], \\ w_c(k) - \alpha_c(k) * [x(k) - w_c(k)], \end{cases} \quad (2)$$

где  $w_c(k+1)$  — следующее значение веса;  $w_c(k)$  — предыдущее значение веса;  $x(k)$  — значение входа,  $\alpha_c(k)$  — коэффициент обучения. Первый вариант используется, если  $x$  и  $w_c$  принадлежат одному классу, в противном случае используется второй вариант

Параметр обучения  $\alpha_c(k)$  вычисляется по формуле:

$$\alpha_c(k) = \frac{\alpha_c(k-1)}{1 + \alpha_c(k-1) * S(k)}, \quad (3)$$

где  $\alpha_c(k)$  — следующее значение параметра обучения;  $\alpha_c(k-1)$  — предыдущее значение параметра; а  $S(k)$  вычисляется по формуле 3:

$$S(k) = \begin{cases} +1, \text{ если } w \text{ и } x \text{ принадлежат одному классу} \\ -1, \text{ в противном случае} \end{cases} \quad (4)$$

### 5. Применение нейронной сети векторного квантования на примере

На основании данных бухгалтерской отчетности по 5 фирмам решим задачу классификации при помощи оптимизированной нейронной сети векторного квантования.

Представленный пример выступает скорее в роли демонстрационного, поскольку построен на небольшом количестве данных.

Таблица 1  
Обучающая выборка

Фирмы	Коэффициенты Альтмана					Классы
X1	0,22	0,2	0,28	1,49	1,37	1
X2	-0,05	0,14	0,06	0,42	0,74	2
X3	0,00	0,04	0,02	0,54	0,34	2
X4	0,08	0,01	0,08	3,58	0,52	1
X5	0,08	0,03	0,08	1,13	0,4	2

В столбце «Классы» класс 1 определяет отнесение фирмы к не банкроту, а класс 2 — к банкроту. Мы разделяем поданную выборку на две части: коэффициенты фирмы X1 и X2 используем в качестве начальных весовых коэффициентов. Таким образом, архитектура сети для решения данной задачи представляется в следующем виде: слой входных векторов состоит из трех нейронов (X3, X4, X5) и

слоя векторного квантования, состоящего из двух нейронов, представляющих собой два класса — банкрот и не банкрот. Параметр обучения сети на начальном этапе принимается равным 0,3.

Применив формулы (1)–(5), мы увидели, что сеть обучилась за один такт. Полученные векторы весов позволяют правильно классифицировать фирмы:  $w1 = (0,2523; 0,2439; 0,3262; 1,0076; 1,5662)$ ,  $w2 = (-0,0193; 0,095; 0,054; 0,5823; 0,5787)$ .

### 6. Нейронные сети в прогнозировании курса акций

Прогнозирование — одна из самых востребованных, но при этом одна из самых сложных задач интеллектуального анализа данных. Проблемы прогнозирования связаны с недостаточным качеством и количеством исходных данных, изменениями среды, в которой протекает процесс, воздействием субъективных факторов.

Начальным этапом в решении данной проблемы является исследование предметной области, что позволяет выделить значимые информационные ресурсы. Следующим этапом является выбор наиболее подходящего метода прогнозирования курса акций и в конечном итоге его программная реализация.

В данной статье приводится обоснование выбора лучшего из методов — прогнозирование с помощью нейронной сети GRNN. Результаты проведенной работы могут быть использованы для написания специализированного программного продукта с целью дальнейшего его использования на рынках ценных бумаг.

### 7. Сравнительная характеристика нейронных сетей

Задача прогнозирования курса акций может быть решена при помощи различных нейронных сетей. Наиболее подходящими для этой цели являются такие сети, как сеть обратного распространения (Backpropagation neural network), радиально-базисная сеть (RBFN), обобщенно-регрессионная сеть (GRNN), модульная сеть.

Сеть обратного распространения является самой первой сетью, в которой количество скрытых слоев было больше единицы. Эта сеть способна к успешной аппроксимации любой нелинейной функции преобразования. Она применима к данным с нелинейной зависимостью таким, как: хаотические временные ряды, зашумленные данные, финансовые временные ряды и регрессионные модели. Нежелательно использовать эту сеть при нестационарных данных или для данных, которые скрывают в себе влияние различных факторов, как в случае прогнозирования курса акций. В таком случае лучше применять GRNN, RBFN или модульные сети [6].

Преимуществами радиально-базисной сети является быстрая обучаемость и способность определять границы лучше, чем сеть обратного распространения при решении задачи классификации. К недостаткам можно отнести более медленную реакцию в фазе отклика. В начальной фазе обучения некоторая выборочная информация может быть утеряна. Также сложно определить число нейронов.

Сравнение с сетью обратного распространения показало аналогичную работу модульной сети, но дало лучшие результаты при вычислении ошибки. Принцип работы основан на том, что каждый модуль «участвует» в соревновании, и это позволяет ускорить работу сети. Главным недостатком является сложная архитектура сети, что ведет к возникновению трудностей при программной реализации.

### 8. Характеристика обобщенно-регрессионной нейронной сети

Обобщенно-регрессионная нейронная сеть (англ. General Regression Neural Network) – разновидность радиально-базисных сетей, предназначена для решения задач регрессии с использованием ядерной аппроксимации. Обобщенно-регрессионная искусственная нейронная сеть в отличие от классического регрессионного анализа не требует задания функциональной зависимости, связывающей входные и выходные переменные. Все входные образы каким-либо образом разбиваются на определенное число кластеров. Если число образов невелико, каждый образ может образовать свой кластер [7].

GRNN состоит из четырех слоев: входные модули, которые образуют полную связь со слоем образов; слой образов; слой суммирования; выходной слой. Функцией активации служит Гауссова функция. ОРНС обучается с учителем.

Архитектура сети представлена на рис. 2 [7].

На входной слой подается обучающая выборка. При этом весовым коэффициентам присваиваются значения выборки и подаются на слой образов, в слое образов весовые коэффициенты проходят через функцию активации, которая имеет следующий вид:

$$\sum_{i=1}^n \exp\left(-\frac{D_i^2}{2\sigma^2}\right), \quad (5)$$

где  $D_i^2$  — квадрат Евклидова расстояния;

$$D_i^2 = (X - X^i)^T (X - X^i); \quad (6)$$

$\sigma$  — параметр сглаживания, для наиболее результативного значения сети выбирается в промежутке от 2 до 6. Преобразованные на предыдущем шаге значения подаются на слой суммирования, состоящий из двух нейронов: S-нейрон накапливает сумму взвешенных выходов слоя образов.

$$\sum_{i=1}^n Y^i \exp\left(-\frac{D_i^2}{2\sigma^2}\right), \quad (7)$$

D-нейрон вычисляет сумму невзвешенных выходов слоя образов

$$\sum_{i=1}^n \exp\left(-\frac{D_i^2}{2\sigma^2}\right), \quad (8)$$

В выходном слое рассчитывается взвешенное среднее по формуле (9):

$$\hat{Y}(X) = \frac{\sum_{i=1}^n Y^i \exp\left(-\frac{D_i^2}{2\sigma^2}\right)}{\sum_{i=1}^n \exp\left(-\frac{D_i^2}{2\sigma^2}\right)}. \quad (9)$$

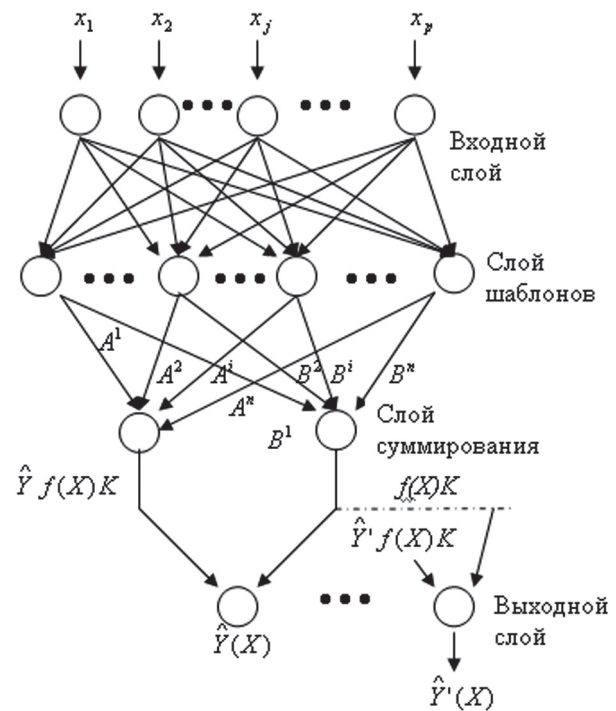


Рис.2. Архитектура сети GRNN

### 9. Применение обобщенно-регрессионной нейронной сети на примере

Входными данными к сети является выборка курсов акций за предыдущие дни, полученная с [finance.yahoo.com](http://finance.yahoo.com). Выходом является прогнозируемый курс на следующий день.

В табл. 2 представлены значения обучающей выборки за 30 дней.

Таблица 2

Входная выборка					
108,57	107,24	104,82	102,7	104,26	103,1
101,41	102,59	104,34	104,51	103,24	106,12
102,73	101,89	104,92	105,6	113,56	113,49
115,11	114,42	115,5	114,75	114,5	113,9
113	113,32	114,2	113,78	110,97	113,9

Настраиваемым параметром сети является параметр сглаживания  $\sigma$ . В таблице 3 представлены результаты работы сети при различных значениях параметра сглаживания. Учителем сети является значение 113,9.

Таблица 3

Результаты работы сети

$\sigma = 2$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$
113,69	113,78	114,08	114,4

Результаты показывают, что при  $\sigma = 3$  прогнозируемое значение наиболее приближено к учителю.

#### Выводы и направление дальнейших исследований

В данной статье было предложено рассмотрение использования OLVQ1 для решения задачи классификации, состоящей в определении жизнеспособных и склонных к банкротству предприятий. Был проведен сравнительный анализ алгоритмов обучения для нейронной сети векторного квантования и выбран наилучший метод — оптимизированная нейронная сеть векторного квантования.

Была изложена общая теория сети векторного квантования, на основании которой был приведен небольшой пример, демонстрирующий возможность практического применения этой сети для решения задач классификации. Представленные результаты дают основания считать сеть векторного квантования достаточно хорошим методом интеллектуального анализа данных.

Также была рассмотрена проблема прогнозирования курса акций и предложено ее решение на основе обобщенно-регрессионной сети.

Направление будущих исследований заключается в изучении самоорганизующихся карт Кохонена, использовании различных моделей конкурентного обучения (правило жесткой конкуренции, справедливой конкуренции, мягкой конкуренции), практическом применении для решения задачи определения жизнеспособных и склонных к банкротству предприятий. Планируется сравнить полученные в ходе эксперимента результаты на примере сети Кохонена и сети вектора квантования. Результаты позволят сравнить две парадигмы обучения нейронных сетей: с учителем для сети OLVQ1 и самоорганизующихся карт Кохонена.

Для задачи прогнозирования курса акций планируется изучение и применение сети радиально-базисных функций с различными функциями активации и сравнение ее работы с работой обобщенно-регрессионной нейронной сети.

**Список литературы:** 1. *З.Ф. Уоссермен* Нейрокомпьютерная техника: Теория и практика, 1992. — 118 с. 2. *Заенцев И.В.* Нейронные сети основные модели. — Воронеж 1999. — 150 с. 3. *G. Peter Zhang.* Neural networks in business forecasting, October 2004, 287 pages. 4. *Amir F. Atiya.* Bankruptcy prediction for credit risk using neural networks: a survey and new results. // IEEE transactions on neural networks, Vol. 12, No 4, July 2001. 5. *Руденко О. Г., Бодянский Е.В.* Основы искусственных нейронных сетей. — Харьков. — 2003. — 323 с. 6. *Keun-Rong Hsieh, Wen-Tsuen Chen* A Neural Network Model which Combines Unsupervised and Supervised Learning // IEEE Trans. on Neural Networks. — 1993. — Vol.4, №2. 7. *Donald F. Specht* A general regression neural network // IEEE Transactions on neural networks. — November 1991. — Vol. 2, № 6

*Поступила в редколлегию 28.04.2008*

УДК 004.7:519.2



## ANALYSIS OF NETWORK PERFORMANCE UNDER SELF-SIMILAR SYSTEM LOADING BY COMPUTER SIMULATION

L.O. Kirichenko<sup>1</sup>, T.A. Radivilova<sup>2</sup><sup>1</sup>KNURE, Kharkov, Ukraine, ludmila@kture.kharkov.ua<sup>2</sup>KNURE, Kharkov, Ukraine, lmd@kture.kharkov.ua

The simulation have shown that management of self-similar traffic allows to improve quality of network service and avoid overflow of the buffer memory.

SELF-SIMILAR DATA, TRAFFIC MODEL, SIMULATING COMPUTER NETWORKS.

### Introduction

Numerous researches of processes in a network have shown that statistical characteristics of the traffic have property of time scale invariance (self-similarity). The reasons of such effect are features of distribution of files on servers, their sizes, and typical behavior of users. There were found, that data flows that initially not showing self-similarity properties, having passed processing on main servers and active network elements, obtains attributes of self-similarity. The self-similar traffic has the special structure kept on many scales. There are always a number of extremely large surges at rather small average level of the traffic. These surges cause significant delays and losses of packages, even when the total loading of all streams are more less than maximal values. In a classical case for Poisson stream buffers of an average size will be enough. The queue can be formed in short-term prospect, but for the long period buffer will be cleared. However in case of self-similar traffic queues have more greater length [1-3].

The traffic in computer networks with high servicing factor shows properties of self-similarity. Because of it the fast overload of devices' buffers is possible even with small servicing factor. Especially it happens if the buffer's size has been calculated for loading with Poisson streams' distributions.

For the majority of networks actual that incoming load can exceed one which can be served even at optimum routing. Thus without restrictions of the incoming traffic, queue on the most loaded lines will grow without limit, and eventually will exceed the sizes of buffers in corresponding units. This can cause the situation, when incoming packets will be ignored and thus will have to be transmitted again, that leads to irrational expenditure of network resources.

So, increasing load of network will lead to channel capacity decrease and information delays growing. It is obvious, that number of lost packets should be reduced as much as possible to improve network quality. The simplest, but also the most expensive method is channel bandwidth widening. More cheap, but also quite effective results can be obtained by using methods of switching, routing and information streams management. Also some special methods of queuing are applicable for this problem. These methods can be broken into three

major groups: special queuing strategies, traffic shaping and speed limitation [1].

**Queuing strategies.** Queues are usually formed only in case of occupied interface. In opposite case packets are retransmitted without any additional processing. All standard queues use FIFO (first in first out) principle: packet that came earlier will be transmitted first. If queue is overloaded and new packets are incoming, the tail drop is take place. More complicated way is to use several queues. Packets are classified according to user needs and then are sort over corresponding queues. When interface is able to transmit packet, special algorithms are used to select queue.

**Traffic shaping.** While shaping count of traffic for the interface is take place. Shaping can be applied to the whole traffic or only to those packets, which matches some criteria. This happens both in case of free interface and in case of full queue. When traffic exceeds some user-defined value all incoming packets are pushed into queue and delayed. Thus, used network channel capacity is limited by some value.

**Speed limitation.** This method is quite similar to shaping. The difference is that abuse traffic is processed separate from usual, by user-defined rules. The most widely used method is deletion of excessive traffic, while some other methods exist, such as priority decreasing in IP-header.

### 1. Switching of information flows

There're several methods of flow switching: circuit switching that can be done both on logical and physical layers, message switching, packet switching [1, 6]. Circuit switching is the persisting of channel that connects some net abonent on overall connection time. The main disadvantages of circuit switching method are: inability of speed and code transformation, that causes necessity of similar hardware usage; organization of multicast and circular transmitting can become very tricky; the lost of requests for connection in case of free channel absence; low channel loading in cases, when the probability of packet lost must be decreased greatly.

Network with message switching method implemented are almost free of these lacks. Here the incoming message is moved to memory and then transmitted into outgoing channel. Such network, especially with

high channel usage, needs significant number of memorizing devices in switching nodes.

Packet switching method has the same advantages as message switching, but requires much less memory, because packet size is significantly smaller than message size.

## 2. Routing and intensity management

There are different criteria of routing methods classification which used in information systems. One of it is the classification by management centralization: centralized, distributed, zone. [1, 4].

In centralized methods route choosing is held in central control point. Distributed methods cause route to be chosen by each joint. Joints can interact with each other to tell some system information.

Centralized routing methods' main idea can be stated as – each joint transmits its state information to the central control point, that builds complete network state. This information is used then to choose optimal routes.

In order to combine advantages of these methods some hybrid methods were created. Here central control point observes the global state, and other joints can react to local traffic and component changes. Each joint can decide by itself what route to choose with respect to local infrastructure.

Other classification of routing methods is based on route-changing property. In static methods route that is defined by pair “sender”–“receiver” is fixed and doesn't depend on traffic fluctuation. It can be changed only in case of some hardware breakage (i.e. network topology change). Such routing method can be recommended only for simplest networks, or when network effectiveness is not a priority, because it can't provide enough channel capacity in cases of incoming traffic variations. Some intermediate place is occupied by quasi-static methods. In opposite to static methods (that have hard-coded routes) and dynamic routes (routes are defined by analyzing current network information) it uses the number of local route tables that are not fixed and is changed (but quite slow) when traffic intensity varies.

More to say, all routing method can be split into two groups by such criteria: some mathematical background is used in method for taking optimal route by performance criteria; heuristic methods.

The concrete routing method is usually implemented in network level protocol that manages packets transmission over network to the destination host. Flow control can be necessary on data transmitting between abonent and incoming network node or between two network nodes.

But, none of discussed methods of overload network are prevention provides significant quality of service. Furthermore, they don't use parameters of incoming traffic that actually cause queue creation.

## 3. The simulation

Simulation Modeling is becoming an increasingly popular method for network performance analysis. Generally, there are two forms of network simulation: analytical modeling and computer simulation. The first is by mathematical analysis that characterizes a network as a set of equations.

The main disadvantage is its over simplistic view of the network and inability to simulate the dynamic nature of a network. Thus, the study of a complex system always requires a discrete event simulation package, which can compute the time that would be associated with real events in a real-life situation. Software simulator is a valuable tool especially for today's network with complex architectures and topologies. Designers can test their new ideas and carry out performance related studies, therefore freed from the burden of the “trial and error” hardware implementations.

One of well-known network simulators is OPNET Modeler. OPNET (Optimized Network Engineering Tool) provides a comprehensive development environment for the specification, simulation and performance analysis of communication networks. A large range of communication systems from a single LAN to global satellite networks can be supported. Discrete event simulations are used as the means of analyzing system performance and their behavior. [5].

The key features of OPNET are summarized here as:

- modeling and Simulation Cycle. OPNET provides powerful tools to assist user to go through three out of the five phases in a design circle (i.e. the building of models, the execution of a simulation and the analysis of the output data);
- hierarchical Modeling. OPNET employs a hierarchical structure to modeling. Each level of the hierarchy describes different aspects of the complete model being simulated;
- specialized in communication networks. Detailed library models provide support for existing protocols and allow researchers and developers to either modify these existing models or develop new models of their own;
- automatic simulation generation. OPNET models can be compiled into executable code. An executable discrete-event simulation can be debugged or simply executed, resulting in output data.

In order to construct adequate model the information about modeling objects is needed. The network from the physical point of view is the union of some number of computers, hubs and switches, that are connected via communication lines. Let's see main units of the network:

- computer – source and destination of a signal, with management and data buffering abilities;
- connection lines (wires, hubs) – the delay element;

- switch, modem, server – delay element with management and data caching abilities;
- router – element that manages data flow with data buffering abilities.

Buffering here means the ability to store some data inside device. The main part of system (“executive mechanism”) is represented by server and router. From the physical point of view this is buffer for incoming data – capacitance element, packet switching element and outgoing buffers. Figure 1 shows the example of computer network. This network consists of four workstations, server, cisco router, and hubs, connected via wire.

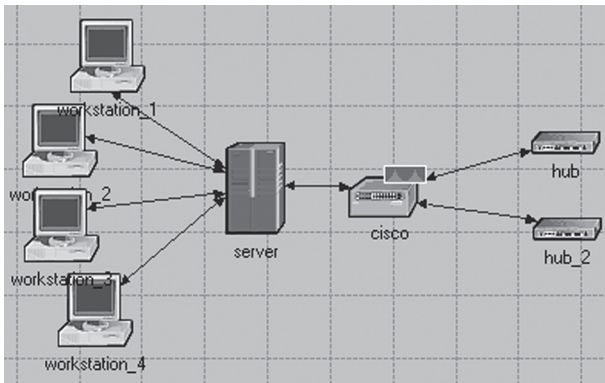


Fig. 1. Example of computer network

While there exists lots of software, that allows to build and simulate different computer networks, there no unified model of selfsimilar traffic, that will provide a way to simulate such traffic with given characteristics.

In this paper the mathematical model of traffic is provided [6]. The input traffic let us treat a certain random process with arbitrary law of distribution. An our task is find a model of random process such that when it passes through the communication channel the queues and probabilities of losses have values similar to those of the real traffic.

In this work we show that the length of a queue is defined by three key parameters of the input traffic: intensity, the Fano parameter  $F$  and Hurst’s parameter. The Fano parameter  $F$  is defined as the ratio of the dispersion of the number of events on the given time interval  $T$  to the average of this number:  $F(T) = \frac{D[N(T)]}{M[N(T)]}$ , where

$N(T)$  is the random variable  $N$  defines the number of events of the given stream on the interval  $T$ . Big values of the Fano parameter correspond to a wide scatter of values in the input stream, which provides queues even at small intensity. Selfsimilarity and long-term dependence of the traffic is indicated traditionally by Hurst’s parameter. Hurst’s parameter value  $H > 0.5$  characterizes long-term dependence of the process. Particularly, it means that high values of the process intensity will be most probably followed with the same high. That does not allow the buffer to be released quickly enough.

The model of traffic presented in this work is self-similar random process with discrete time, based on fractal Gaussian noise. It correspond to the certain value of Hurst’s parameter. Traffic model is an exponent of Gaussian noise:  $Y(t) = b \cdot e^{k \cdot X(t)}$  where  $X$  is the realization of fractal Gaussian noise, obtained by random summation method.  $B, k$ , - parameters that depends on intensity and Fano parameter.

Obtained mathematical model of selfsimilar traffic allows to make simulation with given buffer size and link channel capacity. The result of research states, that for 80-90% system load buffer size should be greater that traffic intensity in hundreds times to minimize the losses. Numerical experiment results provide a way to compute the average size of buffer, needed to normal transmission (not more that 7% losses) of traffic with given channel capacity  $C$  and obtained parameters of incoming data flow  $Buf = f(C, \bar{X}, F, H)$  [7].

If computed buffer size if greater than existing size, the data will be transmitted to reserve channel or router with additional buffer (fig. 2).

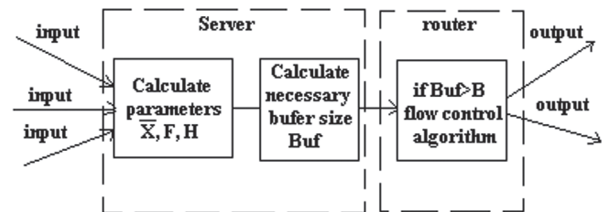


Fig. 2. Traffic management in modeled network

## Conclusion

This paper describes a new traffic model we implemented in OPNET Modeler. The simulation shows a way to manage data flow with goal to network node for processing. The management of data allows to improve quality of network service and avoid overflow of the buffer memory in network node. This method is based on finding the maximum of system load by monitoring of incoming traffic.

**References:** 1. *W. Stollings*. High-speed networks and Internets. Performance and quality of service. New Jersey, 2002. 2. *Vern Paxson and Sally Floyd*. Wide-Area Traffic: The Failure of Poisson Modeling IEEE/ACM Transactions on Networking, vol. 3, № 3, pp. 226-244, June 1995. 3. *Leland W.E., Taqqu M.S., Willinger W., and Wilson D.V.* On the selfsimilarnature of Ethernet traffic, IEEE/ACM Transactions of Networking, 2(1), 1994. pp.1-15. 4. *М. Кульгин*. Практика построения компьютерных сетей. Для профессионалов. 2001, СПб., “Питер”. 5. *J. Theunis, P. Leys, J. Potemans, Bart Van den Broeck, E. Van Lil and A. Van de Capelle*. “Advanced Networking Training for Master Students Through OPNET Projects”, OPNETWORK 2003, Washington D.C., USA, August 2003. 6. *Кириченко Л.О., Радивилова Т.А.* Исследование влияния самоподобия трафика при проектировании фрагмента сети // Нові технології. – 2007. – № 1-2. – С.124-129. 7. *Кириченко Л.О., Радивилова Т.А.* Управление параметрами сети на основе мониторинга входной нагрузки // Материалы 2-й Международной научной конференции «Современные информационные системы. Проблемы и тенденции развития», Харьков-Туапсе. – 2007. С. 89-90.

Поступила в редколлегию 07.05.2008

УДК 519.6:004.93



## СИНТЕЗ МОДЕЛЕЙ СЛОЖНЫХ ОБЪЕКТОВ И СИСТЕМ НА ОСНОВЕ МЕТОДА ПЧЕЛИНОЙ КОЛОНИИ

А.А. Олейник<sup>1</sup>, С.А. Субботин<sup>2</sup>

<sup>1</sup> Запорожский национальный технический университет,  
г. Запорожье, Украина, olejnik@zntu.edu.ua

<sup>2</sup> Запорожский национальный технический университет,  
г. Запорожье, Украина, subbotin@zntu.edu.ua

Рассматриваются мультиагентные системы. Приводится анализ метода пчелиной колонии. Предлагается применение метода пчелиной колонии для синтеза моделей сложных объектов и систем. Разработаны модификации метода пчелиной колонии для отбора информативных признаков и для многомерной оптимизации.

МОДЕЛЬ, ОПТИМИЗАЦИЯ, АГЕНТ, МЕТОД ПЧЕЛИНОЙ КОЛОНИИ, ВЫБОРКА, ЭКЗЕМПЛЯР, ПРИЗНАК

### Введение

Классические методы решения оптимизационных задач при синтезе моделей сложных объектов и систем являются высокоитеративными и накладывают определённые требования (например, унимодальность, непрерывность, монотонность, дифференцируемость и т.п.) на исследуемые зависимости, а также, как правило, являются методами локального поиска.

Поэтому весьма актуальной является задача разработки новых методов решения задач оптимизации, не требующих выполнения определённых требований от исследуемых зависимостей.

Для решения многих оптимизационных задач часто применяются методы, основанные на случайном поиске. К таким методам относятся интеллектуальные методы мультиагентной оптимизации, основанные на моделировании общественного интеллекта (Swarm Intelligence) [1, 2].

К интеллектуальным методам мультиагентной оптимизации, основанным на моделировании общественного интеллекта, относятся: метод муравьиных колоний (Ant Colony Optimization, ACO) [3,4], метод пчелиной колонии (Bee Colony Optimization, BCO) [5], оптимизация с помощью групп частиц (Particle Swarm Optimization, PSO) [6] и другие [7]. Данные методы являются новыми, и во многом ещё недостаточно освоены, однако, некоторые из них достаточно эффективно применяются для решения различных задач оптимизации. Например ACO применяется для решения задачи коммивояжёра [8], задачи календарного планирования [9], отбора информативных признаков [10, 11], кластеризации [12] и других [13–15]; BCO – для решения задачи календарного планирования [16], решения задачи коммивояжёра [17], решения транспортной задачи [18] и других [19, 20]. Одной из основных особенностей интеллектуальных методов мультиагентной оптимизации является то, что они основаны на агентно-ориентированном

подходе [21, 22], что способствует лучшему решению оптимизационных задач.

### 1. Постановка задачи

Метод пчелиной колонии, несмотря на ещё недостаточную известность, показал хорошие результаты при решении ряда оптимизационных задач [16–20]. Поэтому в данной работе ставится задача дальнейшего применения метода пчелиной колонии к решению оптимизационных задач.

Целью работы является исследование метода пчелиной колонии и его применение для синтеза моделей сложных объектов и систем.

Основными задачами работы являются:

- анализ основных принципов коллективного интеллекта, его основных характеристик и свойств, способствующих решению оптимизационных задач;
- формализация описания поведения пчёл, исследование метода пчелиной колонии, его преимуществ и недостатков;
- разработка новых методов оптимизации, основанных на методе пчелиной колонии, для синтеза моделей сложных объектов и систем, позволяющих решать задачу многомерной оптимизации и выполнять выделение наиболее информативных признаков из исходного набора.

### 2. Основные принципы коллективного интеллекта

Мультиагентные методы интеллектуальной оптимизации (методы коллективного интеллекта, Swarm Intelligence) имеют бионическую природу, то есть они основаны на моделировании поведения животных, носящего коллективный характер, за счёт чего достигается, так называемый коллективный интеллект.

Поэтому для определения основных принципов коллективного интеллекта необходимо рассмотреть принципы поведения различных коллективных животных. К коллективным животным относятся животные, которые живут колониями, например,

муравьи, пчёлы, термиты, некоторые виды ос, птиц, рыб и т.п. В основе поведения колоний таких животных лежит самоорганизация [23].

Самоорганизация – множество динамических механизмов, в соответствии с которыми система регулируется на глобальном уровне за счёт взаимодействия её компонентов на нижнем уровне без прямого взаимодействия между этими компонентами. Самоорганизация базируется на четырёх составляющих.

1. Положительная обратная связь: достигается за счёт выполнения простых поведенческих эмпирических приёмов, которые обеспечивают создание решений. Например при фуражировке муравьёв увеличение феромонов на пути к источнику пищи – это разновидность положительной обратной связи, поскольку таким образом создаётся сеть из возможных путей к источнику пищи.

2. Отрицательная обратная связь: уравнивает положительную обратную связь, что ведёт к стабилизации коллективного поведения. Например, при фуражировке муравьёв отрицательная обратная связь может быть вызвана следующими факторами: ограниченное количество фуражиров, истощение источника пищи, испарение феромона.

3. Неустойчивость положительной обратной связи: поведение общественных насекомых основано на стохастических правилах. Случайные отклонения в решениях – основа для новых решений. Кроме того, хаотичность может быть полезной, поскольку это позволяет колонии обнаружить новые решения.

4. Требование множественного взаимодействия между особями, за счёт чего достигается появление новых хороших и выносливых решений.

Для понимания коллективного интеллекта необходимо определить функции, которые выполняют общественные животные в процессе решения различных задач. Можно выделить четыре функции такого вида: координация, кооперация, коллективное принятие решения, специализация. Они не являются взаимно исключаемыми, а наоборот – совместно вносят вклад в достижение поставленных перед колонией целей.

Координация – соответствующая организация задач отдельных индивидов во времени и пространстве, позволяющая решить возникшую перед колонией проблему. Данная функция приводит к определённому пространственно-временному распределению особей, их действий и/или результатов их действий для достижения поставленной цели.

В качестве примера координации можно привести строительную деятельность в колониях насекомых. В процессе строительства гнезда у некоторых разновидностей социальных ос [24] или термитов за счёт процессов стигмержи (stigmergy)

особи улучшают текущее состояние гнезда (пространственная организация), предварительно достигнутое (временная организация) другими особями.

Кооперация достигается за счёт того, что особи выполняют вместе общую задачу, которая не могла бы быть решена отдельным индивидом. Индивиды должны объединять свои усилия, чтобы успешно решить общую задачу, которая находится вне пределов возможностей отдельных индивидов.

Кооперация наблюдается в процессе добычи пищи, когда отдельного индивида недостаточно, чтобы доставить пищу. Известно много примеров совместной транспортировки пищи различными видами муравьёв [25]. Такая транспортировка может быть очень эффективна для доставки пищи в гнездо. К примеру, муравьи *Pheidologeton diversus* сообщают могут транспортировать в десять раз больше веса, чем если бы делали это отдельно друг от друга.

Коллективное принятие решения относится к механизмам, которые срабатывают, когда колония сталкивается с проблемой выбора. Этот механизм завершается коллективным выбором одного из возможных решений.

Специализация заключается в том, что различные действия выполняются отдельными специализированными группами индивидов, например, фуражировка или присмотр за потомством [26]. Такая специализация может осуществляться как за счёт поведенческой дифференциации, так и в зависимости от возраста индивида.

Таким образом, организация коллективного поведения у общественных животных может быть рассмотрена как совокупность четырёх составляющих: координация, кооперация, коллективное принятие решений и специализация. Каждая из этих составляющих проявляется на коллективном уровне за счёт непрямого взаимодействия между особями. Такое взаимодействие обеспечивает возможность оценивания внешней информации в соответствии с двумя основными направлениями:

1. Координация и специализация обеспечивают формирование пространственных, временных и социальных связей, возникающих в процессе работы колонии. Координация регулирует пространственно-временное размещение индивидов, в то время как за счёт специализации достигается распределение их действий.

2. Кооперация и коллективное принятие решения являются инструментами, с помощью которых колония может реагировать на изменения во внешней среде. Коллективное принятие решения предоставляет механизмы, которые обеспечивают решения колонии, в то время как кооперация предоставляет механизмы, которые позволяют выйти за рамки возможности отдельных индивидов.

Вместе четыре составляющих в поведении общественных насекомых формируют представление, что колония в целом планирует свою работу для достижения поставленных целей.

### 3. Формализация поведения пчёл. Метод пчелиной колонии

Поскольку пчёлы являются общественными насекомыми, то их поведение соответствует всем принципам коллективного интеллекта, описанным выше.

– У пчёл положительная обратная связь заключается в том, что пчёлы, основываясь на полученной информации от других пчёл, начинают лететь к указанному источнику ресурсов.

– Отрицательная обратная связь заключается в том, что пчела, основываясь на информации, полученной от других пчёл, может решить, что найденный ею источник хуже.

– Неустойчивость положительной обратной связи: сбившийся с курса фуражир может найти новые, неразработанные источники питания, и затем привести за собой других рабочих особей для разработки этого источника пищи.

– Множественность взаимодействия: информация об источнике ресурсов, найденном одной пчелой, доступна для всех других в улье посредством выполнения так называемого виляющего танца.

– Координация выполняется при организации перемещения гнезда пчёл. В этом случае взаимодействия между отдельными особями приводят к синхронному (временная организация) и направленному (пространственная организация) движению особей к определённой цели.

– Коллективное принятие решения: медоносные пчёлы выбирают более производительные участки с цветами путём вербовки незанятых пчёл с помощью виляющего танца, выполняемого фуражиром, возвратившимся от источника нектара.

– Специализация: в сообществе медоносных пчел труд особей строго разделен. Одни из них добывают и приносят в улей пищу, другие – забирают ее и складывают в соты, третьи – чистят ячейки, четвертые – крыльями, как вентилятором, проветривают улей, пятые – несут караульную службу, дежуря у входа, и не пропускают врагов. В семье существуют разведчики для поиска пропитания, няни для ухода за малышами, пчелы-уборщики, могильщики, которые уносят из улья мертвых. Пчелы-солдаты организованы в специальные воинские подразделения, которые несут службу по защите улья от похитителей меда.

Для формализации, учитывающей соответствующие принципы и особенности поведения пчёл в природе при фуражировке, используются три основных понятия: источник нектара (цветок), занятые фуражиры, незанятые фуражиры.

Источник нектара характеризуется значимостью, определяемой различными факторами, такими как: удалённость от улья, концентрация нектара, удобство добычи нектара.

Занятые фуражиры закреплены за отдельным источником, на котором они добывают нектар, то есть они «заняты» им. Занятые фуражиры владеют такой информацией о данном источнике нектара, как: расстояние и направление от улья, полезность источника.

Незанятые фуражиры продолжают искать источники нектара для их использования. Существует два типа незанятых фуражиров: разведчики, которые ищут новые источники нектара, и наблюдатели, которые ждут в улье и могут выполнять другие действия в улье. Среднее количество разведчиков в рое составляет 5–10%.

Каждая незанятая пчела может полететь к источнику нектара, следуя за пчелой-разведчиком, которая нашла путь к цветку. Это достигается за счёт того, что каждый улей имеет так называемую закрытую площадку для танца, на которой пчёлы, обнаружившие источники нектара, выполняют виляющий танец, тем самым, пытаясь привлечь других незанятых пчёл последовать за ними. Если пчела решает оставить улей, чтобы получить нектар, она следует за одной из пчел-разведчиков к области с нектаром. Таким образом, незанятая пчела становится занятой.

По достижению области с нектаром занятый фуражир добывает нектар и возвращается в улей, оставляя нектар там. После того, как пчела оставляет нектар, она может выполнить одно из следующих трех действий:

- оставить источник нектара и снова стать незанятым фуражиром;
- продолжить фуражировку к тому же источнику нектара, не вербуя других особей своего улья;
- выполнить танец и таким образом выполнить вербовку.

Пчела выбирает одну из вышеупомянутых альтернатив с некоторой вероятностью. В пределах области танца, пчелы «рекламируют» различные области нектара. Механизмы, в соответствии с которыми пчела решает следовать за другой пчелой, исследованы недостаточно хорошо, но предполагается, что вербовка среди пчел с математической точки зрения всегда является функцией качества источника нектара [27]. Также отмечено, что не все пчелы начинают фуражировку одновременно.

Таким образом, выполняется разделение функций между занятыми фуражирами и разведчиками на улучшенное изучение найденных областей с нектаром и на нахождение новых областей с нектаром соответственно. За счёт такого разделения труда достигается эффективная работа всего роя пчёл.

В [27] предложена поведенческая модель самоорганизации колонии пчёл, в которой пчёлы-фуражиры, которые перемещаются к цветкам, возвращаются в улей, а также обладают информацией о полезности соответствующего источника. Собранный нектар обеспечивает обратную связь с текущим состоянием потока нектара в улье. Полезность — функция качества нектара, количества нектара и расстояния от улья. Обратная связь устанавливает ответный порог для сигнала вербовки, также известный как виляющий танец, длина которого зависит и от порога и от полезности. Виляющий танец выполняется на специальной площадке для танцев, где отдельные фуражиры могут наблюдать за их выполнением. Фуражиры могут случайным образом выбирать танец для наблюдения, и в соответствии с танцем они могут определить место расположения участка цветка и покинуть улей для фуражировки. Модель самоорганизации обеспечивает пропорциональную обратную связь с качеством источника нектара.

Таким образом, метод пчелиной колонии является эвристическим, итеративным, интеллектуальным методом мультиагентной оптимизации, основанным на случайном поиске.

Исходя из различных применений данного метода [16–20] можно выделить следующие его достоинства:

- метод не склонен к заикливанию в локальных оптимумах, поскольку основан на случайном поиске;

- поиск лучшего решения основывается на решениях агентов всей колонии пчёл;

- метод может применяться в динамических приложениях, поскольку способен адаптироваться к изменениям окружающей среды;

- может использоваться для решения как дискретных, так и непрерывных задач оптимизации.

К недостаткам метода пчелиной колонии можно отнести:

- итеративность процесса поиска;

- трудности теоретического анализа процесса получения решений, обусловленные тем, что поиск решения имеет стохастическую природу;

- априорную неопределённость времени сходимости, хотя сходимость гарантируется;

- зависимость метода от настройки параметров, подбираемых экспериментально.

Метод пчелиной колонии обладает следующими особенностями:

1. Все агенты делятся на различные типы в соответствии с действиями, которые они выполняют в процессе решения задачи.

- 1.1 Занятые фуражиры обеспечивают использование уже найденных источников нектара, то есть незначительно изменяют уже найденные ранее решения задачи;

- 1.2 незанятые фуражиры обеспечивают продолжение поиска новых источников нектара, то есть агенты такого типа выполняют поиск новых допустимых решений задачи. Незанятые фуражиры в свою очередь бывают двух типов.

- 1.2.1 Наблюдатели — ожидают в улье других агентов. Они не выполняют никаких действий и фактически ожидают момента, когда им нужно будет также начать поиск решений;

- 1.2.2 Разведчики — обеспечивают поиск новых источников нектара. При этом поиск осуществляется случайным образом, то есть они случайно выбирают в пространстве поиска возможное решение.

2. Связь между решениями агентов осуществляется путём моделирования выполнения пчёлами виляющего танца. При этом выполнение виляющего танца обеспечивает образование двух типов обратной связи: положительная обратная связь и отрицательная обратная связь.

3. Процесс поиска решения обеспечивается двумя процедурами.

- 3.1. Поиск новых источников нектара во всём пространстве поиска, который достигается с помощью агентов-разведчиков. Таким образом, обеспечивается исследование всего пространства поиска.

- 3.2. Углублённое использование областей, в которых находятся уже найденные источники нектара (достигается с помощью занятых фуражиров). То есть находятся решения, находящиеся в пространстве поиска вблизи от рассматриваемого решения.

#### 4. Многомерная оптимизация на основе метода пчелиной колонии

Одной из часто возникающих задач в процессе моделирования сложных объектов и систем является нахождение глобального оптимума многомерной функции. Для решения этой задачи существует ряд традиционных методов (например, метод Коши, Ньютона, Левенберга-Марквардта и т.п.), которые, однако, требуют непрерывности, дифференцируемости и унимодальности целевых функций. Поэтому в рамках данной работы предлагается применять метод пчелиной колонии для решения задачи оптимизации многомерной функции.

Предложенный метод можно представить в виде следующих шагов.

*Шаг 1.* Задаются начальные параметры работы метода: начальное количество агентов-разведчиков  $B_s$ , максимальная энергия, коэффициент  $\alpha$ , начальная температура  $T_{init}$ , конечная температура  $T_{final}$ , максимальное количество итераций  $iter_{max}$ . Также задаются параметры, связанные с оптимизируемой функцией: количество переменных  $argCnt$ ; ссылка на саму функцию; минимальный  $Range^{min}$

и максимальный  $Range^{max}$  пределы для каждой из переменных, то есть область, в которой можно искать возможные решения; направление оптимума  $optOrient$  (определяет максимизацию или минимизацию функции). В случае использования локальной оптимизации задаётся один из традиционных методов локальной оптимизации, который будет использоваться, и задаётся класс агентов, для решения которых будет применяться локальная оптимизация – агент с лучшим решением или агенты, выполнившие моделирование виляющего танца.

**Шаг 2.** Создаются начальные агенты-разведчики.

**Шаг 2.1.** Для каждого начального агента-разведчика создаётся случайное решение:

$$bee_i.x_j = rand \cdot (Range_j^{max} - Range_j^{min}) + Range_j^{min},$$

$$\forall i = \overline{1, B_s}, j = \overline{1, OrgCnt},$$

где  $bee_i$  –  $i$ -ый агент-разведчик;  $rand$  – случайное число в интервале  $[0; 1]$ .

**Шаг 2.2.** Для полученных случайных решений рассчитывается полезность данного источника нектара как значение оптимизируемой функции:

$$bee_i.profitability = optOrient \cdot f(bee_i.x_1, \dots, bee_i.x_{OrgCnt}),$$

$$\forall i = \overline{1, B_s},$$

где  $OptOrient = 1$ , если выполняется максимизация, в противном случае –  $OptOrient = -1$ .

Текущее количество итераций устанавливается в  $1$ :  $iter = 1$ ; количество агентов-разведчиков  $B$  устанавливается в  $B_s$ :  $B = B_s$ ; текущая температура  $T$  устанавливается в  $T_{init}$ :  $T = T_{init}$ .

**Шаг 3.** Выбираются рабочие агенты, то есть такие агенты, на базе которых будут создаваться новые агенты с помощью процедуры скрещивания.

**Шаг 3.1.** Определяется агент  $best$  с наибольшей полезностью.

**Шаг 3.2.** Процедура имитации отжига [29]. Агент относится к рабочим агентам  $workBee$ , если выполняется условие:

$$\exp\left(-\frac{|bees_i.profitability - best.profitability|}{T}\right) > rand,$$

$$\forall i = \overline{1, B}.$$

**Шаг 4.** Скрещивание. Поскольку реальные пчёлы-разведчики при выборе источника нектара пользуются также генетическим материалом (в биологии ещё не изучено, каким именно образом разведчики выбирают одни цветки и пропускают другие, то есть предполагается, что разведчики основываются на генетическом опыте), то с помощью процедуры скрещивания моделируется именно этот момент поведения пчёл. Для скрещивания используются ранее отобранные с помощью процедуры имитации отжига рабочие агенты  $workBee$  и лучший агент за все итерации  $best$ . Новые аген-

ты создаются в два этапа: на базе решений рабочих агентов и на базе решения лучшего агента.

**Шаг 4.1.** Создание новых агентов на базе рабочих агентов  $workBee$ :

$$newWorkBee_i.x_j =$$

$$= workBee_i.x_j \pm rand \cdot (workBee_i.x_j - best.x_j),$$

$$\forall i = \overline{1, B_w},$$

где  $B_w$  – количество рабочих агентов  $workBee$ ; знак преобразования «+» или «-» выбирается случайным образом.

**Шаг 4.2.** Создание новых агентов на базе лучшего агента  $best$ :

$$newWorkBee_i.x_j =$$

$$= best.x_j \pm rand \cdot (workBee_i.x_j - best.x_j).$$

**Шаг 4.3.** Для всех новых агентов производится корректировка полученных решений, поскольку полученные таким образом значения переменных могут выходить за пределы  $[Range^{min}; Range^{max}]$ .

**Шаг 4.4.** Рассчитывается полезность полученных решений:

$$newWorkBee_i.profitability =$$

$$= f(newWorkBee_i.x_1, \dots, newWorkBee_i.x_{OrgCnt}),$$

$$\forall i = \overline{1, B_n},$$

где  $B_n$  – количество созданных при скрещивании агентов  $newWorkBee$ .

**Шаг 4.5.** Выбирается новый лучший агент  $best$ .

**Шаг 5.** Моделирование выполнения виляющего танца. К возможному выполнению танца допускаются рабочие агенты  $workBee$ , агенты, созданные путём скрещивания,  $newWorkBee$ , лучший агент за все итерации  $best$ . Моделирование выполнения виляющего танца происходит в несколько этапов. В результате данного моделирования выбираются те агенты, которые за счёт выполнения танца выполнят вербовку других агентов для исследования найденного ими решения.

**Шаг 5.1.** Выполняется нормирование полезностей агентов, допущенных к возможности выполнения танца. При этом нормирование учитывает направление оптимизации  $optOrient$ .

**Шаг 5.2.** Добавление шумов к полученным нормированным полезностям и их корректировка:

$$np_i = \begin{cases} 1, & \text{если } np_i + w_i > 1; \\ np_i + w_i, & \text{если } e_n < np_i + w_i < 1; \\ 0, & \text{если } 0 < np_i + w_i < e_n, \end{cases}$$

где  $np_i$  – нормированная полезность  $i$ -го агента;  $w_i$  – шум в полезности. Шум равномерно распределён в пределах  $(-w; +w)$ . Значение  $w$  выбирается экспериментально (предлагается  $w = 0, 1$ ),  $e_n$  – минимальный порог полезности. Минимальный порог выбирается экспериментально (предлагается  $e_n = 0, 1$ ).

**Шаг 5.3.** Определение преимущества танца каждого агента.

$$L_i = \max\{np_i - \eta \cdot \overline{np}, 0\},$$

где  $L_i$  – преимущество танца  $i$ -го агента;  $\eta$  – коэффициент, управляющий влиянием величины  $\overline{np}$  на  $L_i$ ;  $\overline{np}$  – среднее значение нормированной полезности всех агентов, которые допущены к моделированию выполнения танца:

$$\overline{np} = \frac{1}{B_c} \sum_{i=1}^{B_c} np_i,$$

где  $B_c$  – количество агентов, допущенных к выполнению танца.

**Шаг 5.4.** Выбор тех агентов, которые за счёт выполнения танца выполняют вербовку других агентов для исследования найденного ими решения. Агент считается выполнившим вербовку, если

$$\frac{L_i}{\beta} > \gamma \cdot \overline{np}, \forall i = \overline{1, B_c},$$

где  $\beta > 0$  – коэффициент, понижающий влияние достоинства танца на вероятность выполнения вербовки;  $\gamma \in [0; 1]$  – пороговой коэффициент, определяющий, насколько выше должно быть преимущество танца данного  $i$ -го агента относительно средней полезности решений всех  $B_c$  агентов.

**Шаг 6.** Если требуется проводить локальную оптимизацию для найденных лучших решений, тогда выполняется локальная оптимизация с помощью указанного метода локальной оптимизации (например, метод Нелдера-Мида, Хука-Дживса, Пауэлла и т.п. [28]). В зависимости от установленных параметров локальная оптимизация выполняется либо для решений тех агентов, которые произвели вербовку, либо только для решения лучшего агента.

**Шаг 7.** Выбирается агент с лучшим решением *best*.

**Шаг 8.** Перезапуск агентов. Создаются агенты, которые будут рассматриваться как агенты-разведчики для следующей итерации.

К новым агентам-разведчикам будут относиться:

– агенты, выполнившие посредством танца вербовку, лучший агент;

– агенты, которые стали занятыми фуражирами вследствие вербовки. Поскольку такие агенты должны выполнять улучшенное изучение уже существующего источника с нектаром, то при создании решений для данных агентов должны учитываться решения завербовавших агентов. В связи с этим для завербованных агентов решение создаётся следующим образом:

$$x_j = \text{dancedBee}.x_j + \text{range} \cdot \text{rand} - \frac{\text{range}}{2}, \forall j = \overline{1, \text{OrgCnt}},$$

где *range* – это предел, в котором величина переменной может отклоняться от значения данной  $j$ -ой

переменной в решении агента, моделировавшего танец, *dancedBee*.

– агенты, решение которых создаётся случайным образом:

$$x_j = \text{rand} \cdot (\text{Range}_j^{\max} - \text{Range}_j^{\min}) + \text{Range}_j^{\min}, \\ j = \overline{1, \text{OrgCnt}}.$$

Также для всех созданных агентов рассчитывается полезность выбранного решения.

**Шаг 9.** Обновление динамических параметров метода:

– увеличивается количество итераций:

$$\text{iter} = \text{iter} + 1;$$

– изменяется текущая температура:  $T = \alpha \cdot T$ ;

– изменяется предел *range*:

$$\text{range} = \text{range} \cdot \frac{\text{iter}_{\max} - \text{iter}}{\text{iter}_{\max}}.$$

**Шаг 10.** Проверка на останов. Проверка на останов считается успешной, если выполняется хотя бы одно из двух условий:

– достигнуто максимальное количество итераций:  $\text{iter} = \text{iter}_{\max}$ ;

– текущая температура равна конечной температуре:  $T = T_{\text{final}}$ .

Если проверка на останов дала успешный результат, то выполняется переход к шагу 11, в противном случае – к шагу 3.

**Шаг 11.** Останов.

## 5. Отбор информативных признаков на основе метода пчелиной колонии

В [20] описано применение метода пчелиной колонии для решения задач, основанных на распределении ресурсов (например, транспортной задачи). В соответствии с предложенными в [20] математическими моделями поведения пчёл авторами данной статьи разработана модификация метода пчелиной колонии для отбора информативных признаков при построении классифицирующих моделей в задачах диагностики и распознавания образов по признакам, состоящая из следующих шагов.

**Шаг 1.** Инициализация. Задаются основные параметры метода пчелиной колонии: количество агентов  $B$ , максимальное количество итераций  $T_{\max}$ , начальное количество агентов-разведчиков  $Ex_{\text{start}}$ , ограничение максимального количества агентов-разведчиков  $Ex_{\max}$ , пороговое значение полезности  $s_{\min}$ . Также задаётся общее количество признаков  $M$  и количество признаков  $N$ , которое следует оставить. После чего создаётся пространство поиска порядка  $N \times M$ .

**Шаг 2.** Запуск разведчиков. Разведчики случайным образом размещаются в пространстве поиска. При этом выбранное расположение (точка в про-

странстве поиска) должно быть таким, чтобы не было одинаковых значений координат.

**Шаг 3.** Отправка занятых фуражиров. Занятые фуражиры прикреплены к определённым источникам ресурса. Начальное значение занятых фуражиров  $B_e = 0$ , поскольку в начале работы метода ещё нет источников ресурсов, за которыми могут быть закреплены занятые фуражиры.

Полезность пребывания агента в источнике  $h$  на итерации  $t$ , при условии, что в этом источнике находится  $x_h$  агентов, рассчитывается по формуле:

$$s_h(t) = \frac{a_h}{x_h(t)}, \quad h = \overline{1, N \times M},$$

где  $a_h$  – количество полезного вещества, вырабатываемое источником, в единицу времени. Количество полезного вещества  $a_h$  определяется после составления модели на основе положения соответствующего источника. В аспекте задачи отбора признаков количество полезного вещества  $a_h$  предлагается рассчитывать как обратное значение ошибки модели  $\epsilon_h$ :

$$a_h = \frac{E}{\epsilon_h},$$

где  $E$  – коэффициент, понижающий степень влияния ошибки  $\epsilon_h$ .

Если полезность пребывания  $s_h(t)$  достигает порогового значения ( $s_h < s_{min}$ ), то агент помещается в близлежащую точку от точки  $h$  пространства поиска. Новое положение определяется путём изменения значения одной из координат текущего положения агента:

$$z_r = z_r + k \cdot \Delta z,$$

где  $z_r$  – координата, которая меняется;  $r$  – случайным образом выбранный номер координаты для изменения;  $k$  – коэффициент, определяющий направление изменения значения координаты, может быть равен  $+1$  или  $-1$ ;  $\Delta z$  – предел, в котором может изменяться переменная.

За счёт этого выполняется исследование области, в которой находится ранее выявленный источник ресурсов.

**Шаг 4.** Расчёт полезности полученного ресурса. Суммарная полезность фуражировки занятого фуражира или разведчика  $i$  рассчитывается по формуле:

$$F^i(t) = \begin{cases} 1, & \text{если } J_f(h^i(t)) + w_f^i(t) > 1; \\ J_f(h^i(t)) + w_f^i(t), & \text{если } e_n < J_f(h^i(t)) + w_f^i(t) < 1; \\ 0, & \text{если } 0 < J_f(h^i(t)) + w_f^i(t) < e_n, \end{cases}$$

где  $F^i(t)$  – полезность фуражировки  $i$ -го агента;  $w_f^i(t)$  – шум в суммарной полезности. Шум равномерно распределён в пределах  $(-w_f; +w_f)$ . Значение

$w_f$  выбирается экспериментально (предлагается  $w_f = 0,1$ ),  $e_n$  – минимальный порог полезности. Минимальный порог выбирается экспериментально (предлагается  $e_n = 0,1$ );  $J_f(h^i(t))$  – полезность источника  $h^i$ , в котором побывал  $i$ -ый агент на итерации  $t$ . Полезность источника  $h$  предлагается рассчитывать по формуле:

$$J_f(h) = \frac{\epsilon^*}{\epsilon_h},$$

где  $\epsilon^*$  – заданная (требуемая) точность решения.

Полезность незанятых фуражиров и отдыхающих полагается 0:  $F^i(t) = 0$ .

**Шаг 5.** Выбор лучшего результата и проверка, достигается ли заданная точность  $\epsilon^*$ . Если точность достигается, то выполняется переход к шагу 9, в противном случае – переход к шагу 6.

**Шаг 6.** Моделирование выполнения танца, за счёт чего достигается обмен информацией. Каждый агент принимает решение выполнять или не выполнять танец. При этом вероятность выполнения виляющего танца  $i$ -ым агентом на итерации  $t$  рассчитывается по формуле:

$$p(i,t) = \frac{1}{\beta} L_f^i(t),$$

где  $\beta > 0$  – коэффициент, понижающий влияние преимущества пути на вероятность выполнения танца;  $L_f^i(t)$  – достоинство танца  $i$ -го агента на итерации  $t$ .  $L_f^i(t)$  рассчитывается по формуле:

$$L_f^i(t) = \max\{F^i(t) - \alpha \bar{F}(t), 0\},$$

где  $\bar{F}(t)$  – среднее значение полезности всех источников;  $\alpha$  – коэффициент, управляющий влиянием величины  $\bar{F}(t)$  на  $L_f^i(t)$ .

**Шаг 7.** Выделение новых разведчиков и вербовка. Каждый незанятый фуражир может стать разведчиком или последовать за другим агентом.

Вероятность того, что незанятый фуражир станет разведчиком, рассчитывается по формуле:

$$p_e(t) = \exp\left(-\frac{1}{2} \frac{L_i^2(t)}{\sigma^2}\right),$$

где  $\sigma$  – коэффициент, который необходим для моделирования поведения фуражировки;  $L_i(t)$  – сумма преимущества танцев разных агентов:

$$L_i(t) = \sum_{j=1}^B L_f^j(t).$$

Кроме того, незанятый фуражир может быть подвергнут вербовке, то есть последовать за  $i$ -ым агентом. Вероятность того, что незанятый фуражир последует за  $i$ -ым агентом, предлагается рассчитывать по формуле:

$$p_i(t) = \frac{L_f^i(t)}{\sum_{j=1, j \neq i}^B L_f^j(t)}.$$

**Шаг 8.** Увеличивается счётчик итераций:  $t = t + 1$ . Если  $t < T_{max}$ , то выполнить переход к шагу 2, в противном случае – переход к шагу 9.

**Шаг 9.** Останов.

**6. Эксперименты и результаты**

На основе предложенных модификаций метода пчелиной колонии для оптимизации многомерной функции и решения задачи отбора информативных признаков было разработано программное обеспечение на языке пакета Matlab 7.0.

С помощью созданного программного обеспечения проводились эксперименты по нахождению оптимального значения тестовых многомерных функций с целью проверки эффективности применения метода пчелиной колонии к оптимизации многомерной функции. В экспериментах применялись тестовые функции, представленные в табл. 1. Также в табл. 1 представлены результаты, полученные с помощью разработанного программного обеспечения.

Основным свойством предложенных тестовых функций является полимодальность, то есть большое количество локальных оптимумов, что позволяло проверить, склонен ли предложенный метод

к закливанию в локальных оптимумах. Также важной особенностью некоторых тестовых функций является сложность дифференцируемости.

Графики тестовых функций представлены на рис. 1–4.

Результаты работы разработанного программного обеспечения для соответствующих функций представлены на рис. 5, 6, 7 и 8.

Для проведения экспериментов были выбраны следующие значения настраиваемых параметров:  $B_s = 20$ ,  $\alpha = 0,99$ ,  $T_{init} = 30$ ,  $T_{final} = 10$ ,  $iter_{max} = 100$ .

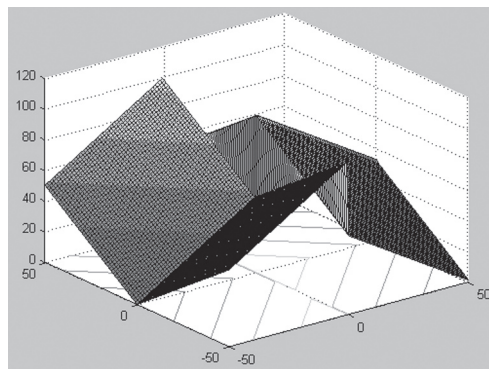


Рис.1. График тестовой функции  $f_1$

Таблица 1

Тестовые функции и найденные для них оптимумы

Функция	Фактический оптимум	Полученный оптимум
$f_1 = p(x_2) \cdot (1 + p(x_1)) +  x_1 + 50p(x_2)(1 - 2p(x_1))  +  x_2 + 50p(x_2)(1 - 2p(x_2)) ,$ <p>где <math>p(x) = \begin{cases} 1, &amp; x \geq 0 \\ 0, &amp; x &lt; 0 \end{cases}</math></p> $-50 \leq x_1 \leq 50;$ $-50 \leq x_2 \leq 50.$	<p>min:</p> $f_1^*(0; -50) = 0$	$x_1 = 0,000513$ $x_2 = -50$ $f_1 = 0,000958$
$f_2 = \sum_{i=1}^n  x_i \cdot \sin(x_i) + 0,1 \cdot x_i ,$ $n = 2$ $-5 \leq x_1 \leq 5;$ $-5 \leq x_2 \leq 5.$	<p>min:</p> $f_2^*(0; 0) = 0$	$x_1 = 0,000952$ $x_2 = 0,000863$ $f_2 = 0,0001831$
$f_3 = 20 + e - 20 \exp\left(-0,2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}\right) - \exp\left(\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i)\right),$ $n = 2$ $-5 \leq x_1 \leq 5;$ $-5 \leq x_2 \leq 5.$	<p>min:</p> $f_3^*(0; 0) = 0$	$x_1 = 0,0001065$ $x_2 = 1,33 \cdot 10^{-5}$ $f_3 = 0,0003039$
$f_4 = 21,5 + x_1 \cdot \sin(4\pi x_1) + x_2 \cdot \sin(20\pi x_2)$ $-3 \leq x_1 \leq 12,1;$ $-4 \leq x_2 \leq 5,8.$	<p>max:</p> $f_4^*(11,63; 5,725) = 38,8503$	$x_1 = 11,63$ $x_2 = 5,725$ $f_4 = 38,8503$

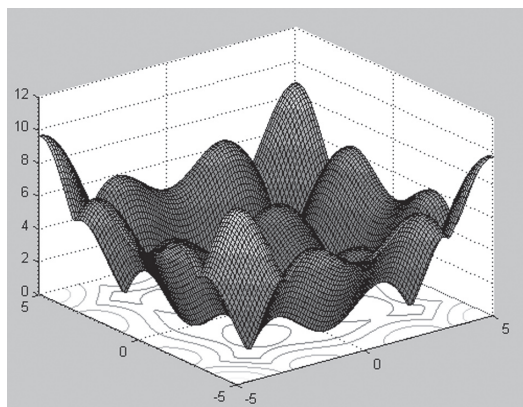


Рис. 2. График тестовой функции  $f_2$

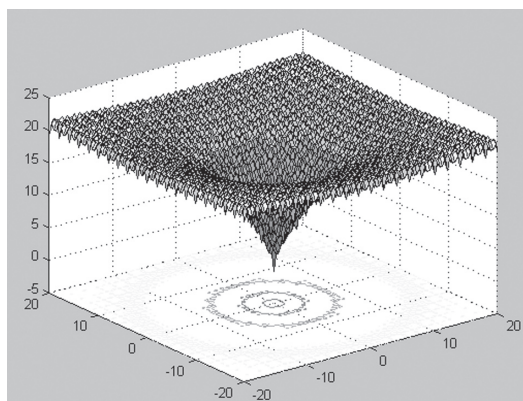


Рис. 3. График тестовой функции  $f_3$

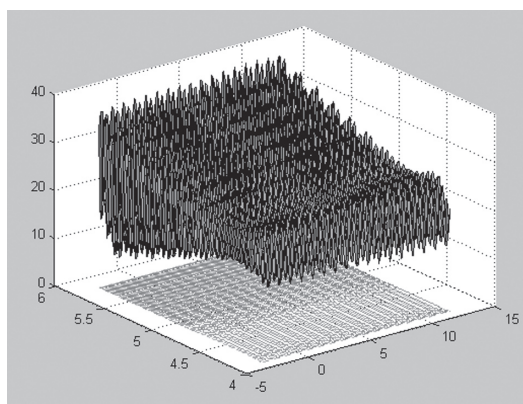


Рис. 4. График тестовой функции  $f_4$

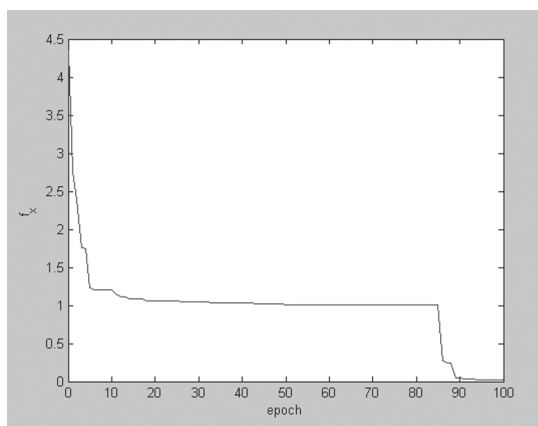


Рис. 5. Результаты работы программы при оптимизации тестовой функции  $f_1$

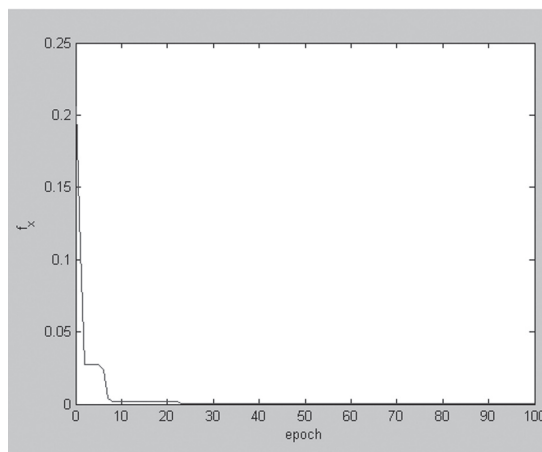


Рис. 6. Результаты работы программы при оптимизации тестовой функции  $f_2$

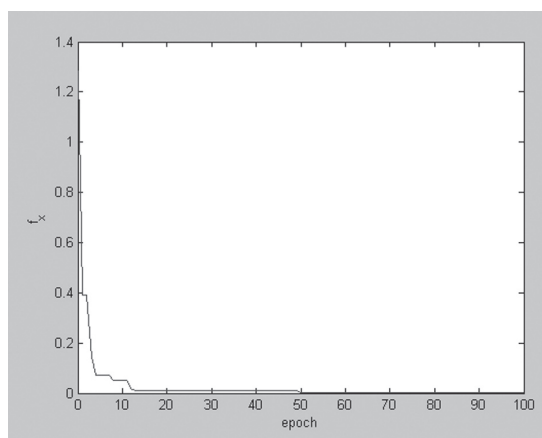


Рис. 7. Результаты работы программы при оптимизации тестовой функции  $f_3$

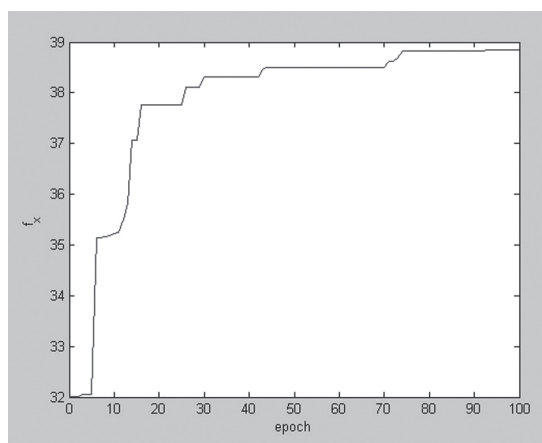


Рис. 8. Результаты работы программы при оптимизации тестовой функции  $f_4$

Как видно из рисунков, оптимальное значение функции предложенный метод пчелиной колонии находит на 20–40 итерации, что свидетельствует о достаточно быстро сходимости метода. При этом – отклонение полученных решений относительно действительных оптимумов составило: 0,00096; 0,00018; 0,0003 и 0 для тестовых функций  $f_1$ ,  $f_2$ ,  $f_3$  и  $f_4$  соответственно.

## Выводы

В данной работе исследован один из мультиагентных методов интеллектуальной оптимизации — метод пчелиной колонии.

Исходя из того, что метод пчелиной колонии может применяться для решения как дискретных, так и непрерывных задач оптимизации, было принято решение метод пчелиной колонии применить к решению актуальных задач: многомерной оптимизации и отбора информативных признаков. Были предложены соответствующие модификации метода пчелиной колонии, использующие основные преимущества метода пчелиной колонии, и имеющие некоторые особенности, например, применение процедуры имитации отжига и выполнение скрещивания, что позволяет улучшить процесс поиска оптимальных значений. Таким образом, была расширена область применения метода пчелиной колонии.

На основании полученных экспериментальных результатов можно сделать вывод, что применение метода пчелиной колонии является эффективным, поскольку найденные оптимальные значения тестовых функций адекватны реальным оптимальным значениям соответствующих тестовых функций.

**Список литературы:** 1. Beni G., Wang J. Swarm Intelligence // Annual Meeting of the Robotics Society: Proceedings of Seventh International Conference. — Tokyo: RSJ Press, 1989. — P. 425–428. 2. Bonabeau E., Dorigo M., Theraulaz G. Swarm Intelligence: From Natural to Artificial Systems. — New York: Oxford University Press, 1999. — 320 p. 3. Dorigo M., Maniezzo V., Colomi A. The Ant System: Optimization by a colony of cooperating agents // IEEE Transactions on Systems, Man, and Cybernetics. — 1996. — Part B, №26(1). — P. 29–41. 4. Олейник А.А. Сравнительный анализ методов оптимизации на основе метода муравьиных колоний // Компьютерне моделювання та інтелектуальні системи: Зб. наук. праць / За ред. Д.М. Пізи, С.О. Субботіна. — Запоріжжя: ЗНТУ, 2007. — С. 147–159. 5. Sumpter D.J.T., Broomhead D.S. Formalising the Link between Worker and Society in Honey Bee Colonies // Lecture Notes In Computer Science: Proceedings of the First International Workshop on Multi-Agent Systems and Agent-Based Simulation. — MABS '98 LNAI, 1998. — P. 95–110. 6. Kennedy J., Eberhart R.C. Particle swarm optimization // International Conference on Neural Networks: Proceedings. — NJ: IEEE, 1995. — Vol 4. — P. 1942–1948. 7. De Castro L.N., Von Zuben F.J. Artificial Immune Systems. Part I. Basic Theory And Applications. — Technical Report No. Rt Dca 01/99. — Brazil: Feec/Unicamp, 1999. — 95 p. 8. Gambardella L. M., Dorigo M. Solving symmetric and asymmetric TSPs by ant colonies // Proceedings of the 1996 IEEE International Conference on Evolutionary Computation (ICEC'96). — NJ: IEEE Press, 1996. — P. 622–627. 9. Colomi A., Dorigo M., Maniezzo V., Trubian M. Ant system for job-shop scheduling // Belgian Journal of Operations Research, Statistics and Computer Science (JORBEL). — 1994. — №34. — P. 39–53. 10. Субботин С.А., Олейник А.А., Яценко В.К. Отбор информативных признаков на основе модифицированного метода муравьиных колоний // Радіоелектроніка та інфор-

матика. — 2006. — № 1. — С. 65–69. 11. Subbotin S., Oleynik A. Modifications of Ant Colony Optimization Method for Feature Selection // The experience of designing and application of CAD systems in Microelectronics: Proceedings of the IX International Conference CADSM—2007 (20–24 February 2007). — Lviv: Publishing house of Lviv Polytechnic, 2007. — P. 493–494. 12. Субботин С.А., Олейник А.А. Кластер-анализ на основе мультиагентного подхода // Автоматика-2007: Тези доповідей чотирнадцятої міжнародної наук.-техн. конф. (10–14 вересня 2007 р.). — Севастополь: СТУЯЕП, 2007. — С. 50–53. 13. Maniezzo V., Colomi A., Dorigo M. The ant system applied to the quadratic assignment problem. — Bruxelles: Universite Libre de Bruxelles, 1994. — 24 p. 14. Bullnheimer B., Hartl R.F., Strauss C. An improved ant system algorithm for the vehicle routing problem. — Vienna: Institute of Management Science, University of Vienna, 1997. — 28 p. 15. Costa D., Hertz A. Ants can colour graphs // Journal of the Operational Research Society. — 1997. — №48. — P. 295–305. 16. Chong S.C., Low M.Y.H. A Bee Colony Optimization Algorithm to Job Shop Scheduling // Winter Simulation Conference: Proceedings of the 38th conference on Winter simulation. — Monterey: Monterey Press, 2006. — P. 1954–1961. 17. Lucic P., Teodorovic D. Bee System: Modeling Combinatorial Optimization Transportation Engineering Problems by Swarm Intelligence // Transportation Analysis: Proceedings of the Triennial Symposium TRISTAN IV. — Sao Miguel: Azores Press, 2001. — P. 441–445. 18. Teodorovic D., Dell'Orco M. Bee Colony Optimization — a Cooperative Learning Approach to Complex Transportation Problems // Advanced OR and AI Methods in Transportation: Proceedings of 16th Mini—EURO Conference and 10th Meeting of EWGT (13–16 September 2005). — Poznan: Publishing House of the Polish Operational and System Research, 2005. — P. 51–60. 19. Nakrani S., Tovey C. On Honey Bees and dynamic allocation in an internet server colony // Adaptive Behavior. — 2004. — №12. — P. 223–240. 20. Quijano N., Passino K.M. Honey Bee Social Foraging Algorithms for Resource Allocation: Theory and Application. — Columbus: Publishing house of the Ohio State University, 2007. — 39 p. 21. Parunak H.V.D. “Go to the ant”: Engineering principles from natural multi-agent systems // Annals of Operation Research — 1997. — №75. — P. 69–101. 22. Gasser L., Hill R.W. Coordinated problem solvers // Annual Review of Computer Science, — 1990. — №4. — P. 203–253. 23. Bonabeau E., Theraulaz G., Deneubourg J.L., Aron S., Camazine S. Self-organization in social insects // Trends in Ecology and Evolution. — №12(50). — 1997. — P. 188–193. 24. Downing H.A., Jeanne R.L. Nest construction by the paperwasp Polistes: a test of stigmergy theory // Animal Behaviour. — №36. — 1988. — P. 1729–1739. 25. Wojtusiak J., Godzinska E.J., Dejean A. Capture and retrieval of very large prey by workers of the African weaver ant *Oecophylla longinoda* (Latreille) // 12th congress of the international union for the study of social insects, Paris, Sorbonne, 21–27 August 1994. — Paris: Universite Paris Nord, 1994. — P. 548. 26. Gordon D.M. Dynamics of task switching in harvester ants // Animal Behaviour. — №38. — 1989. — P. 194–204. 27. Seeley T.D. The Wisdom of the Hive. — Cambridge: Harvard University Press, 1995. — 265 p. 28. Дубровін В.І., Субботін С.О. Методи оптимізації та їх застосування в задачах навчання нейронних мереж: Навч. посіб. — Запоріжжя: ЗНТУ, 2003. — 136 с.

Поступила в редколлегию 12.05.2008

УДК 111



## ФИЛОСОФСКИЕ ОСНОВАНИЯ ФОРМАЛЬНОЙ ОНТОЛОГИИ (ФЕНОМЕНОЛОГИЧЕСКИЙ И АНАЛИТИЧЕСКИЙ АСПЕКТЫ)

В.И. Штанько<sup>1</sup>, Е.В. Добровольская<sup>2</sup>

<sup>1</sup>ХНУРЭ, г.Харьков, Украина, valent@kture.kharkov.ua

<sup>2</sup>ХНУРЭ, г.Харьков, Украина, dobrovolska@gmail.com

В статье рассмотрены философские основания одного из направлений компьютерных наук – онтологического инжиниринга, в рамках которого разрабатываются формальные онтологии. Выявлены глубинные взаимосвязи между аналитической философией первой половины – середины XX в. и формальной онтологией, а также установлена связь между интерпретациями термина «формальная онтология» феноменологией и компьютерными науками.

ОНТОЛОГИЯ, ФОРМАЛЬНАЯ ОНТОЛОГИЯ, ОНТОЛОГИЧЕСКИЙ ИНЖИНИРИНГ, ФЕНОМЕНОЛОГИЯ, АНАЛИТИЧЕСКАЯ ФИЛОСОФИЯ, МАТЕРИАЛЬНАЯ ОНТОЛОГИЯ, КОНЦЕПТУАЛЬНАЯ СХЕМА

### Введение

В настоящее время одним из популярных направлений компьютерных наук является онтологический инжиниринг. В рамках этого направления разрабатываются и проектируются компьютерные (формальные) онтологии, соединившие в себе различные области знания: искусственный интеллект, логику, философию.

#### 1. Цель и постановка задачи

Термин «формальная онтология» имеет феноменологическое происхождение, тем не менее в его интерпретации компьютерными науками прослеживаются также некоторые идеи аналитической философии. В работах Р. Поли [19], Б. Смита [21], Е. Драгалиной-Черной [7], В. Васюкова [2], В. Рубашкина и Д. Лахути [10] и других рассмотрены некоторые аспекты философских оснований формальной онтологии, однако, ее взаимосвязи с феноменологией и аналитической философией исследованы недостаточно и фрагментарно.

Цель авторов — в данной статье выявить глубинные взаимосвязи между данными философскими направлениями (аналитической философией первой половины – середины XX в. и феноменологией Э. Гуссерля) и онтологией компьютерных наук.

#### 2. Формальная онтология Э. Гуссерля

Как отмечалось выше, термин «формальная онтология» имеет феноменологическое происхождение: авторство принадлежит основателю феноменологии Э. Гуссерлю. По мнению этого философа, формальная онтология — это общая форма для всех материальных онтологий, которая «...скрывает в себе формы всех возможных онтологий вообще (всех «настоящих» «материальных» онтологий), ...предписывает всем материальным онтологиям общую для всех них формальную устроенность...» [6, с.20]. Для Гуссерля характерно противопоставление формального материальному. Формальное — это некая высшая форма, подчиняющая и обоб-

щающая в себе все материальное многообразие.

Формальная онтология часто противопоставляется онтологии региона (региональной онтологии), где регион — это наивысший материальный род, которому подчинена конкретная эмпирическая предметность со всеми составляющими ее материальными сущностями. Региональная онтология рассматривает предметы региона, которому она посвящена (например, онтология природы, онтология материальной вещи).

Итак, мы выяснили, что формальная онтология описывает структуру и подчиняет себе все материальные или региональные онтологии, которые являются описанием определенного среза реальности (предметной области) или региона. Теперь рассмотрим более подробно, какое определение дается формальной онтологии в компьютерных науках.

#### 3. Формальная онтология в компьютерных науках

За основу возьмем несколько авторитетных работ, на которые ссылаются исследователи как в области компьютерных наук (информационных технологий), в частности, онтологического инжиниринга [4], [18], так и в области философии [21], [10], в которых даются определения онтологии и проясняется контекст использования этого термина. К этим работам относятся, например, работы Н. Гуарино [16], [17] и Т. Грубера [14], [15].

Т. Грубер определяет онтологию как «эксплицитную спецификацию концептуализации». Под термином «концептуализация» подразумевается «абстрактное упрощенное видение мира, которое мы представляем с определенной целью» [14, р.908]. Далее Грубер поясняет, что в онтологии имена сущностей из универсума рассуждений (такие как классы, отношения, функции или другие объекты) связываются с текстом, понятным человеку, в котором описывается значение этих имен и формальные аксиомы, которые ограничивают интерпретацию и согласовывают использование

имен. В заключение Грубер определяет онтологию как изложение логической теории, говорит о важности такого качества онтологии как согласованность, для чего вводится термин «онтологические обязательства» — договоренность о согласованном использовании общей терминологии.

Н. Гуарино уточняет определения, данные Т. Грубером, и выделяет несколько интерпретаций понятия «онтология» [16]: 1 — онтологическая теория — логическая теория, предназначенная для выражения онтологического знания (все аксиомы онтологической теории должны быть истинны для любой возможной концептуализации, лежащей в основе этой теории); 2 — спецификация онтологических обязательств, где онтологические обязательства понимаются как установление соответствия между логическим языком и множеством семантических структур [17]; 3 — концептуализация, где концептуализация представляется в следующем виде:  $\langle W, D, R \rangle$ , где  $W$  — набор возможных миров,  $D$  — набор объектов из предметной области,  $R$  — набор интенциональных отношений в этой области. Иными словами, концептуализация — набор правил, ограничивающий структуру фрагмента реальности.

Заметим, что и в первой, и во второй работах говорится о различии онтологии на уровне знаний (концептуальной структуры) и онтологии на символическом уровне (прикладной онтологии, артефакта). Прикладная онтология описывает некоторую предметную область — понятие, родственное «региону» Гуссерля. Поэтому прикладную онтологию мы можем отождествить с региональной онтологией.

По мнению Б. Смита, который является специалистом в области компьютерных наук и философии, онтология является словарем терминов с общепринятыми определениями, созданным для получения лексической или таксономической системы для представления знаний, которая может быть принята различными сообществами. В идеале, онтология является формальной теорией, в которую включены не только определения, но и системы аксиом.

В статье Е.Г. Драгиной-Черной утверждается, что «абстрактные логики в целом могут рассматриваться как формальные онтологии абстрактных объектов, приближаясь тем самым к компьютерным онтологиям предметных областей» [7, с.169]. Под абстрактной логикой можно понимать любую совокупность, состоящую из: (1) класса изоморфных структур, (2) класса формальных выражений некоторого языка и (3) отношения выполнимости между ними.

Основываясь на вышеназванных и других работах, можно выделить для рассмотрения такие интерпретации онтологии: концептуальная структура, логическая теория, программно реализован-

ный артефакт (прагматически-ориентированная онтология). Кроме того, в основе создания онтологии мы можем отметить такие идеи: построение искусственного языка; создание концептуальной схемы; стремление к формализации знания, к четким, однозначным формулировкам; установление связи между языком и окружающим миром. Именно эти идеи находились в поле рассмотрения аналитической философии, которая характеризуется строгостью аргументации, обоснованностью, терминологической ясностью и однозначностью.

#### 4. Аналитический аспект философских оснований формальной онтологии

На начальном этапе аналитическая философия, исследуя язык науки, была уверена в возможности элиминации философских проблем с помощью анализа искусственного или естественного языка. Философия должна была помогать науке в ее развитии, выявлять ее логическую структуру, избавлять от неточности, многозначности путем анализа языка, совершенствовать его и создавать идеальный язык науки. Целью философии было создание единой формализованной системы знаний о мире — точной, строгой, лаконичной.

После лингвистического поворота в поле изучения аналитической философии попал обыденный язык и обыденный мир. Считалось, что реальность является человеку с помощью языка, речевых актов, язык и есть реальность. Признавались историчность и многозначность языка. Отказ от единой структуры языка, возможность различных интерпретаций обосновывались новым представлением языка — как множества языковых игр, теорий, основанных на различных предпосылках. Куайн, например, говорил о существовании логической формы теории, которую он назвал «теоретической формой», эта форма может быть интерпретирована по-разному в зависимости от выбранного универсума рассуждений. Признавалось, что мир можно описывать в различных системах, то есть каждая система специфицировалась своей онтологией — совокупностью общих категорий.

Начиная с 60-х годов 20 в. в рамках аналитической философии рассматривались теории референции, соотношения слов и объектов реального мира. Разрабатывались два подхода — экстерналистский и интерналистский — значение рассматривалось либо как внешний, социальный, либо как внутренний феномен. Задача философии виделась в выявлении глубинной грамматики, общей формальной структуры обыденного языка. Аналитическая философия на основе выявления формальной структуры языка, глубинной грамматики пыталась перейти к структурным основам бытия. По мнению Стросона, Хомского, Даммита, в основания грамматики входит онтология, потому необходимо

было создание особого онтологического словаря.

Возможно, именно идеи аналитической философии о постижении и описании реальности с помощью анализа языка, выявления и воспроизведения его глубинной структуры, построении искусственного языка и явились предпосылками возникновения такого направления в компьютерных науках как инженерия онтологий (онтологический инжиниринг) и самой онтологии компьютерных наук (компьютерной онтологии, онтологии информационных наук).

Рассмотрим более подробно, какие понятия аналитической философии могли явиться идейными предпосылками возникновения онтологии компьютерных наук.

На начальном этапе развития аналитической философии логический позитивист, один из разработчиков программных положений Венского кружка Р. Карнап ввел в рассмотрение термин «языковой каркас» — систему способов речи, подчиненную некоторым правилам: «Если кто-либо хочет говорить на своем языке о новом виде объектов, он должен ввести систему новых способов речи, подчиненную новым правилам; мы назовем эту процедуру построением языкового каркаса для рассматриваемых новых объектов» [12, с.20].

Идею «понятийного аппарата», подобную идее «языкового каркаса», мы находим у К. Айдукевича — представителя Львовско-Варшавской школы. Как «понятийный аппарат», так и «языковой каркас» можно поставить в соответствие онтологии как концептуальной структуре.

Понятийный аппарат К. Айдукевича являлся основой для картины мира: «...все суждения, которые мы принимаем и которые образуют картину мира, не определяются однозначно данными опыта, но зависят от выбора понятийного аппарата, с помощью которого мы интерпретируем эти данные» [1, с.231]. При этом Айдукевич говорил о том, что принятием того или иного понятийного аппарата мы можем изменять всю картину мира, настаивал на пластичности и изменчивости понятийного аппарата каждого человека.

У Айдукевича мы находим также основу для интерпретации онтологии как логической системы — понятие языка, которое отличается от обыденного. Так, например, немецкий язык (в обыденном понимании) является совокупностью языков, отличающихся друг от друга способом приписывания значений. При изменении значения, связываемого со словом, мы переходим от одного языка к другому, оставаясь при этом в рамках одного языка в обычном понимании этого слова.

Язык может быть согласованным и несогласованным, открытым и замкнутым. Айдукевич вывел признаки переводимости языков одного в другой и взаимосвязь переводимости с понятийным аппа-

ратом. Понятийным аппаратом, лежащим в основе замкнутого и согласованного языка, он назвал класс всех значений выражений, фигурирующих в этом языке.

Язык у Айдукевича являлся идеализированным, ему не соответствовали ни обыденные языки, ни язык науки, поэтому правила функционирования, переводимости языков один в другой необходимо было приспособлять к действительности, подобно тому, как в физике приспособляют к действительности законы, выведенные для идеального газа.

Еще у одного представителя Львовско-Варшавской школы — А. Тарского мы находим понятие «мета-язык», которое можно сравнить с моделью, структурой самой онтологии, так как это язык, в котором говорится о другом (объектном) языке: объектный язык — это «...язык, который «о чем-то говорит» и который является предметом всего нашего обсуждения...», мета-язык — язык, «...в котором мы «говорим о» первом языке и в терминах которого мы хотим, в частности, построить определение истины для первого языка» [11, с.101].

Тарский отмечал, что оба эти термина относительно, так как если нас заинтересует понятие истины, применимое к предложениям мета-языка, то он станет объектным языком, и чтобы определить истину для этого языка, мы должны перейти к новому мета-языку (более высокого уровня). Мета-язык содержит объектный язык как часть и термины общелогического характера.

Понятие «мета-язык» можно сравнить также и с предпосылочным языком, а следовательно, и с онтологией верхнего уровня (содержащей базовые понятия), так как значение определяемого термина объектного языка объясняется в терминах мета-языка, причем, эти термины должны быть «совершенно ясными и недвусмысленными». Это достигается тем, что в состав мета-языка входят термины объектного языка; термины, относящиеся к форме выражений объектного языка и используемые для образования их имен; и термины логики. Таким образом, как считает Тарский, мета-язык не будет включать никаких «неопределяемых терминов», семантические термины (говорящие об объектном языке) могут вводиться в мета-язык только посредством определений. К мета-языку Тарский выдвинул требование «быть существенно богаче», выполнение которого должно было сделать невозможной интерпретацию мета-языка в объектном языке и гарантировало удовлетворительное определение истины.

Еще один философ, придерживающийся взглядов логического позитивизма — В. Куайн, рассматривал в своей работе «Онтологическая относительность» понятие «предпосылочный язык», которое можно сравнить скорее с онтологией верхнего

уровня, так как оно содержит базисные термины, с помощью которых можно объяснить все остальные - для остановки «регресса» в определении. Правда, Куайн признавал, что с помощью предпосылочного языка можно получить только относительный смысл: «Предпосылочный язык дает нам искомый смысл (query sense), хотя бы относительный смысл, относительный в обращении к этому предпосылочному языку.» [20, р.205].

Куайн говорил также об онтологической относительности теории: универсум теории может быть осмысленным лишь относительно некоторой предпосылочной теории и лишь относительно некоторого выбора способа перевода одной теории в другую.

Идейные предпосылки возникновения компьютерной онтологии мы находим не только у логических позитивистов, но и у представителей лингвистической философии. Так, например, у самого инициатора «лингвистического поворота» в аналитической философии Л. Витгенштейна одними из ключевых понятий являлись «примитивный язык» и «языковая игра», которые сходны с прикладной онтологией, так как слова здесь употребляются в строго определенном смысле и относятся к определенной области значений. Вне игры вещь не имеет имени, слово имеет значение только в составе предложения.

Языковая игра для Витгенштейна – форма жизни, которых существует множество: «Легко представить себе язык, состоящий только из приказов и донесений в сражении. Или язык, состоящий только из вопросов и выражений подтверждения и отрицания. И бесчисленное множество других языков. Представить же себе какой-нибудь язык – значит представить некоторую форму жизни» [3, §19].

Еще один представитель лингвистической философии – Д. Дэвидсон – в работе «Об идее концептуальной схемы» [8] проанализировал взгляды своих предшественников и современников на вопросы о смене, переводимости и интерпретации концептуальных схем. Саму концептуальную схему Дэвидсон определил как способ организации опыта; систему категорий, придающую форму чувственным данным; точку зрения индивидов, культур и эпох на происходящие события. Если вспомнить определение онтологии как концептуальной структуры или схемы, то легко проследить аналогию между понятием «концептуальной схемы» Д. Дэвидсона и онтологией компьютерных наук.

Дэвидсон пытался установить отношения между языком, реальностью и концептуальной схемой. По его мнению, схема лежит в основе языка, причем, одной схеме может соответствовать множество языков, а разным концептуальным схемам

соответствуют разные языки: «когда концептуальные схемы различаются, то различаются и языки. Говорящие на разных языках могут разделять одну концептуальную схему при условии наличия способа перевода с одного языка на другой, поэтому изучение критериев перевода представляется способом рассмотрения критериев идентичности концептуальных схем» [8, с.145].

Концептуальная схема и язык взаимозависимы, так как с помощью языка мы организуем данные опыта, чувственные данные для построения схемы: «...нечто является языком и объединяется с концептуальной схемой независимо от того, можем ли мы это нечто перевести, если оно стоит в определенном отношении (предсказания, организации, согласования) к опыту (природе, реальности, чувственным данным)» [8, с.151].

Об отношении между реальностью и концептуальной схемой Дэвидсон говорил так: «Даже сама реальность относительна к схеме: то, что считается реальным в одной системе понимания, может не считаться таковым в другой» [8, с.144]. Философ выделил понятие дуализма концептуальной схемы и эмпирического содержания (организующей схемы и того, что ожидает организации), которое назвал третьей догмой эмпиризма, свободного от «...непоследовательных догм аналитико-синтетической дистинкции и редукционизма, т.е. от бесполезной идеи, будто мы можем единственным путем – предложение за предложением – локализовать эмпирическое содержание» [8, с.150] Этот дуализм не может быть представлен рационально.

Дэвидсон отметил, что он не первый рассматривал понятие концептуальной схемы, ссылаясь на «концептуальную схему науки» Куайна, различные миры Стросона (отличающиеся друг от друга благодаря нашим фиксированным описательным ресурсам, предполагаемой системой понятий), различные миры Куна (мир в глазах различных наблюдателей, подходящих к нему с несоизмеримыми системами понятий), схемы Патнэма и Фейерабенда, порождаемые новой наукой.

Рассматривая историю развития аналитической философии и компьютерной онтологии, можно проследить общие тенденции.

Первоначально задачей аналитической философии было создание однозначной формализованной системы знаний о мире, что является идеей построения онтологии компьютерных наук. В ходе своего развития философия показала несостоятельность своей программы и отказалась от этой задачи. Была доказана невозможность формализации всех областей знания, невозможность создания универсальной структуры.

К. Айдукевич, например, показал, что развитие науки не может иметь тенденцию к универсальному языку или универсальной области значений, так

как в таком случае язык будет несогласованным, а область значений этого языка была бы грубым подбием понятийных аппаратов.

Гудмен говорил о невозможности выбора единственного основания классификации для онтологии и об ограниченности наших способов описания. Дэвидсон говорил о невозможности существования единой концептуальной схемы.

Витгенштейн показал, что реальный язык нельзя формализовать, привести к логически чистой системе: «Ведь кристальная чистота логики оказывается для нас недостижимой, она остается всего лишь требованием...требованию чистоты грозит превращение в нечто пустое. Оно заводит нас на гладкий лед, где отсутствует трение, стало быть, условия в каком-то смысле становятся идеальными, но именно поэтому мы не в состоянии двигаться. Мы хотим идти: тогда нам нужно трение. Назад, на грубую почву!» [3, §107]. Кроме того, Витгенштейн говорил о невозможности формализации языка, описывающего внутренние переживания и понятного «лишь мне одному».

Райл говорил о невозможности формализации: «...логика повседневных утверждений, и даже логика утверждений ученых, юристов, историков и игроков в бридж, в принципе не может быть адекватно представлена посредством формул формальной логики. Так называемые логические постоянные, отчасти благодаря продуманному ограничению, действительно имеют рассчитанную логическую силу. Однако неформальные выражения и повседневного, и технического дискурса имеют собственные нерегламентированные логические возможности, которые нельзя без остатка свести к логическим возможностям марионеток формальной логики» [9, с.172].

По мнению Тарского, и этого мнения придерживаются некоторые разработчики онтологий, возможна формализация (теоретически) некоторых областей науки (Тарский приводил в пример математику и теоретическую физику). «Было бы интересно и важно действительно построить язык такого типа, который оказался бы достаточно богатым для изложения обширной области эмпирической науки. Это послужило бы оправданием надежды на то, что языки с точной структурой в конце концов заменят повседневный язык в научных рассуждениях» [11, с.98].

После неоднократных попыток создания единой онтологии идея универсальности была оставлена большинством исследователей, и инженерия онтологий сосредоточилась на создании прикладных онтологий. Кроме того, появилась идея создания онтологии верхнего уровня, описывающей базовые термины, на основе которой можно было бы строить онтологии предметных областей. Райл говорил о существовании «меновых», «дотеорети-

ческих» понятий, обладающих нейтральностью, с помощью которых можно осуществлять переход от одной области науки к другой, но все же было доказано, что единой системы, даже базовой, быть не может. Так, например, Дэвидсон говорил о бессмысленности поисков общих оснований для сравнения концептуальных схем, если под этим подразумевать нечто общее для несоизмеримых схем. Этим основанием не может быть ни нейтральная по отношению к теории реальность, ни некий словарь (фиксированный запас значений). О существовании такого базисного словаря, полностью состоящего из слов, которые связаны с природой «непроблематическими» способами, на основании которого можно сравнивать теории, говорил Кун. Фейерабенд и Дэвидсон показали, что такой словарь не может существовать, так как «При переходе от одной теории к другой слова неуловимым образом изменяют свои значения или условия применимости. Хотя большая часть тех же самых знаков используется как до, так и после научной революции (например: сила, масса, элемент, состав, клетка), способ, которым они связываются с природой, изменился. Таким образом, следующие друг за другом теории, как мы утверждали, несоизмеримы».

Развитие аналитической философии показало, что содержание фактов, теорий, зависит и от других, вненаучных форм деятельности людей. Наука всегда существует в контексте культуры, общества, истории и зависит от экономики, политики, искусства, религии и тому подобного. Это затрагивает проблему интерпретации и интеграции онтологий — ведь не всегда даже исследователи в одной и той же области могут найти общий язык.

Построение онтологии предполагает определение входящих в нее понятий, для определения которых используются другие понятия. Витгенштейн обратил внимание на бесконечность определения. Карнап говорит о необходимости ввода протокольных предложений, из которых можно вывести все понятия. Куайн же указывает на относительность определения: «То, что делает онтологические вопросы бессмысленными, если они рассматриваются абсолютно (а не относительно), — это не их универсальность, а их свойство быть логическим кругом. Вопрос в форме «Что есть  $F$ ?» может получить ответ только обращением к следующему термину « $F$  есть  $G$ ». Ответ имеет только относительный смысл: смысл, относительный к некритическому принятию  $G$ » [20, p.208].

### Выводы

В заключение отметим, что после рассмотрения интерпретации понятия «формальная онтология» компьютерными науками и феноменологией Э.Гуссерля, можно утверждать, что они имеют и

общие черты, и различия. Общей является идея о подчиненности материального формальному, тем не менее, если в компьютерных науках установлена прямая связь материальный регион - формальная онтология, то в феноменологии эта связь осуществляется через сущностную форму. Прямая же связь установлена между регионом и региональной или материальной онтологией.

Выявив взаимосвязи между аналитической философией первой половины – середины XX в. и формальной онтологией, можно утверждать, что невнимание к факту существования данной связи привело разработчиков онтологий к повторению ошибок философов-аналитиков. Несмотря на неудачи, постигшие аналитическую философию, идея существования общей, универсальной структуры мира, вера в возможность отображения этой структуры с помощью изучения языка, продолжают существовать. Онтологический инжиниринг, у истоков которого стояли философы-аналитики, повторил попытки воплощения этой идеи и снова потерпел неудачу.

**Список литературы:** 1. *Айдукевич К.* Картина мира и понятийный аппарат // *Философия науки*. Вып. 2: Гносеологические и методологические проблемы. – М., 1996. – С. 231-254. 2. *Васюков В.Л.* Формальная онтология и искусственный интеллект // *Искусственный интеллект*. – М.: ИИнтелЛЛ, 2006. – С. 305-324. 3. *Витгенштейн Л.* Философские исследования // *Философские работы*. М., 1994. – Ч. 1. С. 80-130. (пар. 1-120). 4. *Гаврилова Т.А., Геллеверя Т.Е., Горовой В.А.* Онтологии как средство концептуализации web-порталов // *Искусственный интеллект*. – 2002. – №3. – С. 80-86. 5. *Гудмен Н.* Способы создания миров: Пер. с англ. – М.: Идея-Пресс, Логос, Праксис, 2001, – 376 с. 6. *Гуссерль Э.* Идеи к чистой феноменологии и феноменологической философии / Эдмунд Гуссерль / Пер. с нем. А. В. Михайлова. – М.: Дом интеллектуал. кн.: CEU Press, 1999. (<http://museum.philosophy.ru/books/>) 7. *Драгалина-Черная Е.Г.* Формальные онтологии как абстрактные логики // *Логические исследования*, 2005.

– №12. – С. 162-169. 8. *Дэвидсон Д.* Об идее концептуальной схемы // *Аналитическая философия*. Избранные тексты. (Сост. А.Ф.Грязнов). – М., Изд-во МГУ. – С. 144-159. 9. *Пайл Г.* Обыденный язык // *Аналитическая философия: Становление и развитие (антология)* – М.: «Дом интеллектуальной книги», «Прогресс-Традиция», 1998. – С.150-173. 10. *Рубашкин В.Ш., Лахути Д.Г.* Онтология: от натурфилософии к научному мировоззрению и инженерии знаний // *Вопросы философии*. 2005. – № 1. – С. 64-81. 11. *Тарский А.* Семантическая концепция истины и основания семантики // *Аналитическая философия: Становление и развитие (антология)* – М.: «Дом интеллектуальной книги», «Прогресс-Традиция», 1998. – С. 90-129. 12. *Carnap R.* Empiricism, Semantics and Ontology // *Revue Internationale de Philosophie* 1950. – Vol. 4. – p. 20-40. Reprinted // *The linguistic turn: Recent essays in philosophical method*. – Chicago; L., 1967. P.72-84. 13. *Davidson D.* The Method of Truth in Methaphysics // *Inquiries into Truth and Interpretation*. Oxford, 1985, p. 199–214. 14. *Gruber T.R.* Toward Principles for the Design of Ontologies Used for Knowledge Sharing//*International Journal of Human-Computer Studies*. Vol. 43, p.907-928. 15. *Gruber T.R.* The Role of Common Ontology in Achieving Sharable, Reusable Knowledge Bases // *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*. Cambridge, MA, Morgan Kaufmann, 1991. – p. 601-602. 16. *Guarino N., Giaretta P.* Ontologies and Knowledge Bases. Towards a Terminological Clarification // *Towards Very Large Knowledge Bases*. –IOS Press, Amsterdam, 1995. 17. *Guarino N., Carrara M., Giaretta P.* Formalizing Ontological Commitment. // *Proceedings of AAAI-94*, Vol. 1, 1994. – p. 560-567. 18. *Mizoguchi R., Ikeda M.* Towards Ontology Engineering, Proceedings of the Joint Pacific Asian Conference on Expert Systems / Singapore International Conference on Intelligent Systems, Singapore, 1997. – p. 259-266. 19. *Poli R.* Descriptive, Formal and Formalized Ontologies // *Husserl's Logical Investigations Reconsidered*. Berlin: Springer, 2003. p. 183-210. 20. *Quine W.V.O.* Ontological Relativity // *The Journal of Philosophy*. 1968. Vol. LXV, № 7. P. 185-212. 21. *Smith B.* Ontology // *The Blackwell Guide to Philosophy of Computing and Information*. – Oxford: Blackwell, 2003. – p. 155-166.

*Поступила в редколлегию 12.05.2008*

УДК 519.62



## О СПОСОБЕ КОДИРОВАНИЯ СИГНАЛОВ РАЗЛИЧНОЙ ПРИРОДЫ

С.Н. Герасин<sup>1</sup>, О.В. Калиниченко<sup>2</sup>, В.П. Лоцман<sup>3</sup>, А.С. Трошило<sup>4</sup>

<sup>1</sup> ХНУРЭ, г. Харьков, Украина, sgerasin@mail.ru

<sup>2, 3, 4</sup> ХНУРЭ, г. Харьков, Украина

На основе представления нового подхода кодирования звуковой информации проведен сравнительный анализ его с широко известными методами кодирования звуковых данных. Приведены описания применяемых при этом методик сравнения, а также итоговые результаты испытаний и перспективы дальнейших исследований.

КОДИРОВАНИЕ, СЖАТИЕ АУДИОДАНЫХ, МЕТРИКИ, ТЕСТИРОВАНИЕ, СОНОГРАММА

### Введение

Не так давно в университете проводились работы по разработке и тестированию эффективного способа кодирования информации. Основной задачей была разработка таких алгоритмов кодирования/декодирования сигналов различной природы, применение которых, например, в сфере звукозаписи позволило бы добиться большего на порядок коэффициента сжатия аудио данных по сравнению с ныне используемыми звуковыми форматами [1-4].

В представленном на тестирование способе многокомпонентного кодирования, как следует из документации, для получения высоких характеристик использовался ряд методологических предпосылок. Основная предпосылка заключалась в том, что сигнал любой природы представляет собой информационный объект, который нельзя разложить на составляющие элементы. Следующая методологическая предпосылка состояла в том, что не все цифровые значения сигнала имеют одинаковую информационную значимость. Поэтому часть цифровых значений сигнала должна фиксироваться в виде значимых компонентов. Другая часть может быть восстановлена при воспроизведении сигнала по информационно значимым компонентам. Алгоритм нахождения информационно значимых компонентов сигнала, а также восстановления всего множества значений составляет основную суть способа.

Способ многокомпонентного кодирования (CD-способ) применительно к различным информационным сигналам показал широкие возможности его применения в различных областях науки и техники (системах связи). Его можно использовать в системах речевой обработки сигналов, управления и хранения данных, а также в сжатии при передаче по каналам связи.

Теоретические результаты получили подтверждение в проведенной серии испытаний (тестов), которые опирались на методы компьютерного моделирования [5, 6]. Кроме этого были получены данные успешной работы способа с речевыми сигналами, звуковыми данными CD качества, не уступающие при воспроизведению оригиналам [7].

Для проведения тестовых испытаний работа на подготовленных заранее тестовых сигналах различного происхождения, специалисты университета проанализировали и отработали на практике общепринятые и хорошо известные тестовые методики проверки. Для сравнения на испытание были отобраны широко применяемые алгоритмы сжатия данных [8,9,10].

Основной целью было сравнение качества работы кодеков при кодировании музыки и речи, а также выявление особенностей кодирования, отражающихся на качестве звучания [11]. С таких же позиций рассматривался универсальный программный кодек, отражающий (моделирующий) новый способ многокомпонентного кодирования. Задачи проведения более критических анализов, с детальными оценками реализации способа, перед исследователями не ставились по той причине, что способ проходит этап патентования.

В ходе исследования (испытаний) применялись различные методики сравнения: от оценки качества при прослушивании до использования автоматических методов сравнения [1]. Конечно, в большинстве случаев чаще доверяют метрикам, которые основаны на оценках прослушивания именно музыкальными экспертами. Однако применение автоматических методов сравнения позволяет существенно ускорить темпы и качество разработки новых кодеков, в особенности для предотвращения появления в коде ошибок, влияющих на воспроизведение.

В рамках тестирования были использованы следующие кодеки:

#### Речевые кодеки:

- 1) GSM 6.10 (встроенный кодек windows 98 SE);
- 2) CELP (встроенный кодек windows 98 SE);
- 3) TrueSpeech (встроенный кодек windows 98 SE).

#### Универсальные кодеки:

- 1) MP3 кодек Lame 3.93 MMX (RazorLame V 1.1.5.1342);
- 2) MP3 кодек из пакета Blaze Media Convert 1.4 (BMC);
- 3) MP3 Pro Fraunhofer (CoolEdit 2.0 mp3pro);
- 4) Advanced Audio Coding (AAC) (производитель неизвестен);

5) Windows Media Audio (WMA) кодек из пакета Blaze Media Convert 1.4;

6) Yamaha Sound VQ Format (VQF);

7) Coder/decoder (CD) – программная модель способа многокомпонентного кодирования/декодирования.

Основная задача тестирования кодеков – сравнение качества сжатия нескольких наиболее известных и распространенных аудио-кодексов с размером получаемого потока данных 32kbps (сжатие порядка двадцати раз) из формата PCM 44100 Hz, 16 bit, mono, 705600 kbps. Такой поток теоретически позволяет передавать звук через модем в реальном времени.

### 1. Используемые метрики тестирования

Автоматическое тестирование проводилось с использованием двух основных метрик: временной и частотной Peak-Signal-to-Noise-Ratio (PSNR). Общая методика тестирования включала цикл компрессии-декомпрессии исходного файла исследуемыми кодексами с последующим сравнением сжатого и исходного файлов программами реализующими метрики.

#### Временная PSNR-метрика.

Значение временного PSNR для двух файлов, представленных массивами сэмплов ( $a[i], i = 1..n; b[j], j = 1..m$ ), вычислялось по следующей формуле:

$$PSNR = 20 \times \text{Log} \left( \frac{\text{Max\_sample\_value}}{\sqrt{\frac{\sum_{i=1}^{\min(n,m)} (a[i] - b[i])^2}{\min(n,m)}}}} \right)$$

где  $\text{Max\_sample\_value}$  – максимальное значение амплитуды каждого сэмпла, в нашем случае равно  $2^{15} = 32768$ .

#### Частотно-временная PSNR-метрика.

Основная часть кодексов перед сжатием переводит сигнал в частотно-временное пространство, поэтому для них более корректно использовать метрику, работающую как с временными составляющими сигнала, так и с частотными.

Основная методика – сравнение отклонения амплитуд сигналов в частотно-временном пространстве. От исследуемых сигналов берется дискретное оконное преобразование Фурье с некоторым шагом по времени STFT (short-time Fourier transform), после чего вычисляются амплитуды по формуле:

$$A_n[j] = \log \left( \sqrt{\text{FFT}[j].\text{Im}^2 + \text{FFT}[j].\text{Re}^2} \right).$$

Каждый такой вектор является спектром сигнала небольшой области вокруг данной временной точки. Объединив все векторы  $A_n$  моментальных спектров в столбцы матрицы, можно получить сонограмму данного сигнала

$$\text{Sono}[i][j] = A_i[j].$$

Частотно-временное PSNR для двух сигналов рассчитывалось следующим образом:

$$PSNR = 20 \times \text{Log} \left( \frac{\text{Max\_value}}{\sqrt{\frac{\sum_{i=0}^{\min(T_1, T_2)} \sum_{j=f_1}^{f_2} (\text{Sono}_1[i][j] - \text{Sono}_2[i][j])^2}{\min(T_1, T_2)}}}} \right)$$

где  $\text{Max\_value}$  – амплитуда сигнала максимальной возможной мощности, допустимая в данном представлении звукового сигнала;  $T_1$  и  $T_2$  – количество векторов моментальных пиков для первого и второго сигнала соответственно;  $f_1$  и  $f_2$  – параметры, отфильтровывающие из общей спектрограммы частотную полосу для исследования. Преимущество данного метода заключается в том, что он не чувствителен к фазе сигнала, в результате чего подходит для кодексов, сохраняющих общее звучание, но не сохраняющих фазы (форму волны), а также в том, что появляется возможность рассмотрения искажений в отдельных частотных диапазонах.

### 2. Методы тестирования по сонограмме

Методика тестирования кодексов с потоком данных 32 кбит/с. по сонограмме. Изображения сонограмм исходного файла и конечного (прошедшего цикл компрессии/декомпрессии) сравнивались между собой. Большие отличия в сонограммах свидетельствуют о большем искажении звука в процессе кодирования. Для большей наглядности строилось, так называемое, изображение «разности сонограмм» массива разности значений сонограмм исходного и полученного сигналов. Соответственно, чем больше оказывались получаемые значения по модулю, тем больше отличался исходный файл при сжатии на данных частотах и моментах времени. Если при кодировании файла какие-то частоты исчезают, то на изображении разности сонограмм эта область помечается сиреневыми или красными цветами в зависимости от величины изменения. А если, наоборот, добавляются какие-то новые звуковые составляющие, которых не было в исходном сигнале, либо они имели меньшую амплитуду, то такие области помечаются синими и зелеными цветами.

Для упрощения изложения амплитуду сигнала будем рассматривать в логарифмической шкале.

Если имеются значения амплитуд сонограммы первого сигнала  $A_1[t][f]$  и второго  $A_2[t][f]$ , где  $t$  – координата по времени, а  $f$  – по частоте, то разница изображений сонограммы получается так:

$$A_r[i][j] = A_1[i][j] - A_2[i][j] \quad i = t_1..t_2 \quad j = [0, 22050] \text{ Hz}.$$

Если  $A_r[i][j] < 0$ , то данная точка помечается фиолетовым цветом при малом абсолютном значении и красным, если  $abs(A_r[i][j])$  велико. Также при  $A_r[i][j] > 0$  цвет переходит от белого к синему, а потом к зеленому с увеличением значения  $A_r[i][j]$ .

На обычных сонограммах цвета переходят от белого к синему, потом к зеленому, красному и желтому по мере увеличения значений  $A_n[i][j]$ ,  $n \in 1, 2$ .

Значения амплитуд  $A_n[i][j]$  получаются из дискретного преобразования Фурье от исходного сигнала следующим образом:

$$A_n[i][j] = \log\left(\sqrt{FFT[j].Im^2 + FFT[j].Re^2}\right), i \in [0, T],$$

где  $i$  – время.  $FFT$ -вектор, полученный в результате обработки сигнала быстрым преобразованием Фурье.

### Выводы и перспективы исследований

В итоге проделанной специалистами работы, можно сказать, что исследованные метрики не дают результата, полностью соответствующего человеческой оценке качества звука. Но результаты всех метрик в совокупности немного приближаются к корректному ответу (с точки зрения человека). Методы сравнения, сравнивающие только форму волны, оказываются наименее пригодными для сравнения звука среди всех метрик, описанных в этой статье. Однако по ним можно определить особенности сжатия того или иного кодека, например, по ним легко узнать, сохраняет ли кодек форму волны или нет. Следует отметить, что если кодек не сохраняет форму волны, то это совершенно не значит, что кодек некачественно сжимает файл, так как, например, шум можно сохранить как волну сложной формы, а можно просто записать параметры ее распределения и при декомпрессии восстановить шум с другой формой волны, но аналогичный по звучанию. Получается, что качественное сравнение кодеков с сохранением формы волны и без ее сохранения, по форме сигнала оказывается попыткой сопоставить практически несравнимые предметы, имеющие разные характеристики, как, например, сравнение кошки и собаки на предмет того, кто лучше. В этом случае мы лишь можем сказать о том, что исследуемые объекты имеют разные параметры. Использование кодеком психоакустической модели также вносит серьезные искажения в результаты тестирования, если в метрике не использовать информацию о таких моделях. В то же время, построение спектральных моделей и метрик на основе спектра может сильно помочь при анализе и классификации звука, а также при определении причин появления артефактов.

С точки зрения близости к оригиналу по описанным метрикам предпочтения следует отдать новому CD-способу. Он может определиться как

лучший по общим критериям сравнений.

В ходе испытаний программной модели CD-способа было установлено следующее:

1. *Работа со специальными сигналами.* В представленном спектре стандартных тестовых сигналов от 10 Гц до 22,5 кГц дополнительно вводились определенные функциональные «нагрузки» и тестовые характеристики. Получен максимальный коэффициент сжатия  $K=3000$ . Средний коэффициент на множестве специальных сигналов составил  $K=2000$ .

2. *Сжатие речевых сигналов.* На тестовых примерах сжатия речи был получен сравнительно высокий коэффициент сжатия, максимальный  $K=246$ . Тестировались примеры английской, немецкой, французской и русской связной речи. Средний коэффициент на множестве речевых сигналов составил  $K=124$ .

3. *Сжатие аудио-сигналов формата wav.* Тестовые примеры аудио-сигналов имели широкий жанрово-музыкальный спектр, при этом получен средний коэффициент  $K=40$ . Для музыкальных файлов данный способ кодирования может составить серьезную конкуренцию на рынке аудиозаписи. Он также вполне отвечает требованиям по применению в телефонии и системах радиосвязи.

4. Испытания показали, что представленный на исследование способ многокомпонентного кодирования и декодирования составляет серьезную конкуренцию формату mp3 и обеспечивает конкурентный уровень не только на рынке аудио-записей, но и рынке Embedded технологий.

Важным направлением совершенствования способов общения человека и ЭВМ является создание специальной аппаратуры и программ, обеспечивающих эффективный обмен информацией между ЭВМ и ее пользователем с помощью звучащей речи. Речь – это самый легкий и удобный способ общения, наиболее часто используемый человеком. Письмо, набор текста на компьютере – значительно менее мобильные и оперативные способы общения. Речь является самым быстрым способом связи, она эффективно действует на расстоянии, в темноте, она легко совмещается с другими действиями человека. Но эффективное решение проблемы автоматического речевого общения человека и ЭВМ невозможно без решения проблем, связанных с дискретизацией акустических колебаний, их отделения от различных шумов и помех и их кодирования таким образом, чтобы иметь возможность распознавать отдельные фонетические единицы.

Исходя из всего сказанного выше, можно сделать однозначный вывод: изучение свойств слухового восприятия человека является одной из важнейших задач, решение которой приведет к значительному облегчению общения человека и

ЭВМ и расширению сферы применения средств вычислительной техники.

В области изучения свойств слухового восприятия человека имеется множество задач, ждущих своего решения. Одной из них является выяснение принципов и математическое описание закономерностей, на основе которых человек воспринимает непрерывные колебания воздуха и превращает их в дискретную форму.

Данная работа посвящена изучению свойств и механизмов восприятия слуховым анализатором человека дискретизированных акустических колебаний и разработке системы кодирования звука, которая, опираясь на соответствующие научные исследования, позволяла бы принимать во внимание и кодировать тот минимальный объем информации, который необходим слуховому анализатору человека для восприятия кода звука как полноценной картины исходного звукового сигнала.

При вводе в ЭВМ осциллограмма звука кодируется с помощью АЦП-преобразователей. При этом диаграмма звукового давления заменяется решетчатой функцией. Численные значения отдельных линий решетчатой функции запоминаются вычислительной машиной. Такой способ кодирования звука получил повсеместное применение. Между тем, исследование механизма слухового анализатора человека показывает, что природа использует совершенно иной способ кодирования звука, а именно - частотно-импульсный. Если с помощью микроэлектронной техники определить, какая информация передается по отдельным волокнам слухового нерва, то обнаружится, что по ним следуют одиночные электрохимические импульсы, причем тем реже, чем ниже громкость сигналов, и наоборот. Каждый импульс возникает в момент накопления некоторой фиксированной порции площади под кривой звукового давления, отсчитываемой от момента возникновения предыдущего импульса. Исследования показывают, что частотно-импульсный метод кодирования звука выгодно отличается от применяемого повсеместно в настоящее время АЦП как в смысле повышенной помехоустойчивости, так и в смысле экономности при хранении в памяти ЭВМ. Установлено, что частотно-импульсное кодирование звука обеспечивает, как минимум, семикратную экономию памяти при

хранении кода звука по сравнению с АЦП-кодированием. Значительная экономия достигается также за счет уменьшения трудоемкости обработки звуковых кодов.

Широкое практическое использование дискретизации речевых сигналов выдвигает для научной проработки ряд актуальных вопросов:

— можно ли строго обосновать правомерность дискретизации звука;

— является ли использование аналого-цифрового преобразователя стандартного типа наилучшим способом дискретизации звукового сигнала;

— с какой точностью и частотой следует проводить дискретизацию и на чем основывать выбор численных значений параметров дискретизации?

Анализ литературных данных показывает, что ни на один из этих вопросов пока нет исчерпывающих ответов.

Очевидно, что ответ на все эти вопросы может быть получен лишь на пути психофизического изучения свойств слухового восприятия.

**Список литературы:** 1. *Ватолин Д., Ратушняк А., Смирнов М.* Методы сжатия данных. — М.: Диалог-МИФИ, 2003. — 382 с. 2. *Ковалгин Ю.А.* Алгоритмы компрессии цифровых аудиоданных // Системы и средства связи, телевидения и радиовещания. — 2000. — № 3. — С. 17-29. 3. *Ковалгин Ю.А.* Компрессия цифрового звука: психоакустические основы и алгоритмы // Звукорежиссер. — 2000. — № 6. — С. 56-63. 4. *Ковалгин Ю.А., Вологдин Э.И.* Цифровое кодирование звуковых сигналов. — Санкт-Петербург, КОРОНА, принт. -2004. 5. *Jurgen Herre, Michael Schug.* Analysis of Decompressed Audio - The "Inverse Decoder" Presented at the 109th Convention 2000 September 22-25 Los Angeles, California, USA. 6. *M. Kowalski and B. Torr.esani* A STUDY OF BERNOULLI AND STRUCTURED RANDOM WAVEFORM MODELS FOR AUDIO SIGNALS // LAMP, CMI, 39 rue Joliot-Curie, 13453 Marseille cedex 13, France 2005. 7. Audio Test and Demonstration CD SOUND CHECK 2 // <http://www.interstudio.co.uk/schck2a.htm> 8. *Church S.* ISDN and ISO/MPEG Layer III Audio Coding: Powerful New Tools for Broadcast Audio Production / AES an audio engineering society preprint/ The 95<sup>th</sup> Convention/ — New-York, 1993. — October. — P. 7-10. 9. Digital Audio Compression Standard (AC-3) // Doc.A/52.-1995-12-20. 10. *Davis Pan,* "A Tutorial on MPEG/Audio Compression", IEEE Multimedia, Vol. 2, No. 2, pp. 60-74, 1995. 11. Structured Audio: Creation, Transmission, and Rendering of Parametric Sound Representations (1998) *Barry Vercoe, William Gardner, Eric Scheirer* // <http://citeseer.ist.psu.edu/context/342384/442890>.

*Поступила в редколлегию 16.05.2008*

## ОБ АВТОРАХ

<b>Авраменко Валерий Павлович</b>	45	д-р техн. наук, профессор кафедры информационных управляющих систем Харьковского национального университета радиоэлектроники
<b>Бондаренко Михаил Федорович</b>	3, 13, 23	д-р техн. наук, профессор, ректор Харьковского национального университета радиоэлектроники
<b>Бритик Владимир Иванович</b>	94	канд. техн. наук, доцент кафедры информатики Харьковского национального университета радиоэлектроники
<b>Бурцев Валерий Николаевич</b>	30	канд. техн. наук, заместитель директора научно-производственной фирмы «Цветодинамика» по научной работе
<b>Бурцев Владимир Николаевич</b>	30	директор научно-производственной фирмы «Цветодинамика»
<b>Бых Анатолий Иванович</b>	111	д-р физ.-мат. наук, профессор, заведующий кафедрой биомедицинских электронных устройств и систем Харьковского национального университета радиоэлектроники
<b>Герасин Сергей Николаевич</b>	177	д-р техн. наук, профессор кафедры высшей математики Харьковского национального университета радиоэлектроники
<b>Гороховатский Владимир Алексеевич</b>	87	канд. техн. наук, доцент кафедры информатики Харьковского национального университета радиоэлектроники
<b>Гребенник Игорь Валериевич</b>	79	д-р техн. наук, профессор кафедры системотехники Харьковского национального университета радиоэлектроники
<b>Добровольская Елена Витальевна</b>	171	аспирантка Харьковского национального университета радиоэлектроники
<b>Егорова Евгения Александровна</b>	94	аспирантка Харьковского национального университета радиоэлектроники
<b>Ерохин Андрей Леонидович</b>	30	д-р техн. наук, профессор, начальник кафедры информатики Харьковского национального университета внутренних дел
<b>Жемчужкина Татьяна Владимировна</b>	111	канд. техн. наук, старший преподаватель кафедры биомедицинских электронных устройств и систем Харьковского национального университета радиоэлектроники
<b>Ивченко Ольга Владимировна</b>	143	студентка Харьковского национального университета радиоэлектроники
<b>Калиниченко Ольга Викторовна</b>	177	доцент кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники
<b>Каменева Ирина Витальевна</b>	149	студентка Харьковского национального университета радиоэлектроники
<b>Касала Ольга Васильевна</b>	153	студентка Харьковского национального университета радиоэлектроники
<b>Кириченко Людмила Олеговна</b>	158	канд. техн. наук, доцент кафедры прикладной математики Харьковского национального университета радиоэлектроники
<b>Кобылин Олег Анатольевич</b>	74	канд. техн. наук, ассистент кафедры информатики Харьковского национального университета радиоэлектроники
<b>Кораблев Николай Михайлович</b>	106	канд. техн. наук, доцент кафедры ЭВМ Харьковского национального университета радиоэлектроники
<b>Левыкин Виктор Макарович</b>	40	д-р техн. наук, профессор, заведующий кафедрой информационных управляющих систем Харьковского национального университета радиоэлектроники
<b>Лесная Наталия Советовна</b>	143	профессор кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники
<b>Лесовец Татьяна Владимировна</b>	153	студентка Харьковского национального университета радиоэлектроники
<b>Лоцман Владимир Павлович</b>	177	студент Харьковского национального университета радиоэлектроники
<b>Любченко Татьяна Петровна</b>	52	младший научный сотрудник Украинского языково-информационного фонда НАН Украины
<b>Машталир Сергей Владимирович</b>	101	канд. техн. наук, доцент кафедры информатики Харьковского национального университета радиоэлектроники
<b>Набатова Светлана Николаевна</b>	34	аспирантка Харьковского национального университета радиоэлектроники
<b>Непечатова Виктория Дмитриевна</b>	139	аспирантка Харьковского национального университета радиоэлектроники
<b>Нечипоренко Алина Сергеевна</b>	65	аспирантка Харьковского национального университета радиоэлектроники

<b>Носова Татьяна Витальевна</b>	111	канд. техн. наук, младший научный сотрудник кафедры биомедицинских электронных устройств и систем Харьковского национального университета радиоэлектроники
<b>Олейник Алексей Александрович</b>	161	студент Запорожского национального технического университета
<b>Перепелица Валентин Иванович</b>	134	аспирант Харьковского национального университета радиоэлектроники
<b>Прасол Игорь Викторович</b>	65	канд. техн. наук, профессор кафедры биомедицинских приборов и устройств Харьковского национального университета радиоэлектроники
<b>Процай Наталья Тимофеевна</b>	69	аспирантка Харьковского национального университета радиоэлектроники
<b>Радивилова Тамара Анатольевна</b>	158	методист отдела организации учебного процесса Харьковского национального университета радиоэлектроники
<b>Репка Виктория Борисовна</b>	143	канд. техн. наук, доцент кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники
<b>Романова Татьяна Евгеньевна</b>	79	доктор технических наук, профессор, старший научный сотрудник Института проблем машиностроения им. А.Н.Подгорного НАН Украины
<b>Самер Лага</b>	139	аспирант Харьковского национального университета радиоэлектроники
<b>Самынина Марина Геннадиевна</b>	123	аспирантка Института животноводства Украинской академии аграрных наук
<b>Свередюк Надежда Леонидовна</b>	101	студентка Харьковского национального университета радиоэлектроники
<b>Ситник Лариса Григорьевна</b>	128	преподаватель кафедры педагогики и инновационных технологий Сумского областного института последипломного педагогического образования
<b>Субботин Сергей Александрович</b>	161	канд. техн. наук, лауреат премии Президента Украины, доцент кафедры программных средств Запорожского национального технического университета
<b>Тарасов Юрий Александрович</b>	149	студент Харьковского национального университета радиоэлектроники
<b>Тевяшев Андрей Дмитриевич</b>		академик Украинской нефте-газовой академии, д-р техн. наук, профессор, заведующий кафедрой прикладной математики Харьковского национального университета радиоэлектроники
<b>Тимофеев Владимир Александрович</b>	139	профессор, заведующий кафедрой экономической кибернетики Харьковского национального университета радиоэлектроники
<b>Трошило Александр Степанович</b>	177	ведущий научный сотрудник Института радиофизики и электроники им. А.Я. Усикова НАН Украины
<b>Тулупов Владимир Владимирович</b>	45	канд. техн. наук, доцент кафедры информатики Харьковского национального университета внутренних дел
<b>Удовенко Сергей Григорьевич</b>	134	д-р техн. наук, профессор кафедры ЭВМ Харьковского национального университета радиоэлектроники
<b>Чалый Сергей Федорович</b>	40	д-р техн. наук, профессор кафедры информационных управляющих систем Харьковского национального университета радиоэлектроники
<b>Четвериков Григорий Григорьевич</b>	153	д-р техн. наук, профессор кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники
<b>Шабанов-Кушнарченко Сергей Юрьевич</b>	3, 13, 23, 128	д-р техн. наук, профессор кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники
<b>Шабанов-Кушнарченко Юрий Петрович</b>	3, 13, 23	д-р техн. наук, профессор кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники
<b>Шатовская Татьяна Борисовна</b>	149	канд. техн. наук, доцент кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники
<b>Шерстнюк Андрей Витальевич</b>	143	студент Харьковского национального университета радиоэлектроники
<b>Шеховцов Сергей Борисович</b>	79	канд. техн. наук, доцент Харьковского национального университета внутренних дел
<b>Шигимага Виктор Александрович</b>	117	канд. с.-х. наук, старший научный сотрудник, биофизик, заведующий лабораторией клеточной и молекулярной биологии Института животноводства Украинской академии аграрных наук
<b>Штанько Валентина Игоревна</b>	171	д-р филос. наук, профессор кафедры философии Харьковского национального университета радиоэлектроники
<b>Яськов Георгий Николаевич</b>	84	канд. техн. наук, научный сотрудник отдела математического моделирования и оптимального проектирования Института проблем машиностроения НАН Украины им. А.Н. Подгорного

## 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 x 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 страниц (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, ris\_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 подрисовочную 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 редколлегию. 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. Реферат не должен дублировать текст аннотации.

Вместе с рукописью (на листах белой бумаги формата А4 плотностью 80–90 г/м<sup>2</sup>, 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. Оплату за публикацию.

*Наукове видання*

**БІОНІКА ІНТЕЛЕКТУ**  
**інформація, мова, інтелект**

Науково-технічний журнал

№ 1 (68)

2008

Головний редактор — *М. Ф. БОНДАРЕНКО*

Відповідальний редактор — *Ю. П. Шабанов-Кушнарєнко*

Заступник відповідального редактора — *Г. Г. Четвериков*

Відповідальний секретар — *І. Д. Вечірська*

Коректор — *Л. М. Денісова*

Комп'ютерна верстка — *О. Б. Ісаєва*

Рекомендовано Вченою Радою  
Харківського національного університету радіоелектроніки  
(протокол № 43 від 30.05.2008)

**Адреса редакції:**

Україна, 61166, Харків-166, просп. Леніна, 14,  
Харківський національний університет радіоелектроніки, к. 127, 285  
тел. 702-14-77, факс 702-10-13,  
e-mail: [bionics@kture.kharkov.ua](mailto:bionics@kture.kharkov.ua)

---

Підписано до друку 30.11.2007. Формат 60 x 84 <sup>1</sup>/<sub>8</sub>. Друк ризографічний.  
Папір офсетний. Гарнітура Newton. Умов. друк. арк. 21,4. Обл.-вид. арк. 22,0.  
Тираж 100 прим. Зам. № .

Надруковано в навчально-науковому видавничо-поліграфічному центрі ХНУРЕ  
61166, Харків-166, просп. Леніна, 14