

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ УКРАИНЫ

ХАРЬКОВСКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ
РАДИОЭЛЕКТРОНИКИ

ISSN 0555-2656

БИОНИКА ИНТЕЛЛЕКТА

ИНФОРМАЦИЯ, ЯЗЫК, ИНТЕЛЛЕКТ

№ 2 (69)

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> Компараторная идентификация цветового зрения человека.....	3
<i>Бондаренко М.Ф., Шабанов-Кушнарченко С.Ю., Шабанов-Кушнарченко Ю.П.</i> Об общей теории компараторной идентификации	13
<i>Бондаренко М.Ф., Шабанов-Кушнарченко С.Ю., Шабанов-Кушнарченко Ю.П.</i> Идентификация объектов, описываемых числовыми системами	23
<i>Ерошенко А.Р.</i> Веб-сервисы в лингвистических информационных системах.....	32
<i>Єрохін А.Л., Турута О.П.</i> Параметричне управління пріоритетами обслуговування web-додатка сервером.....	36
<i>Машталир В.П., Путятин А.Е., Сакало Е.С.</i> Автоассоциативная многослойная нейронная сеть и алгоритм ее обучения при сжатии изображений	40
<i>Гороховатский В.А.</i> Метрики на множествах ключевых точек изображений	45
<i>Кобылин О.А.</i> Определение оптимального уровня вейвлет-разложения изображения для нормализации	51
<i>Гребенник И.В., Романова Т.Е., Шеховцов С.Б.</i> Интервальное оценивание альтернатив при принятии решений в геометрическом проектировании	56
<i>Збитнева М.В., Дударь В.В., Деревянко А.Г.</i> Сбор и обработка информации в интернет с использованием интеллектуальных агентов	61
<i>Алексієв В.О.</i> Концепція застосування GRID-технології на транспорті	67
<i>Авраменко В.П., Калистратова С.Ю.</i> Интеллектуальные процедуры управления бизнес-процессами предприятий.....	73
<i>Сироджа И.Б.</i> Оценивание качества идентификационных и прогнозных решений в инженерии квантов знаний.....	77
<i>Четвериков Г.Г., Федорова Т.Н.</i> Моделирование технологических процессов управления активными элементами.....	84
<i>Бісікало О.В.</i> Орієнтувальний рефлекс як задача моделювання образного механізму оперативної пам'яті людини....	89
<i>Марьин С.А., Мишин А.П., Репка В.Б.</i> Функционирование программного агента интерактивного взаимодействия пользователей системе дистанционного обучения	95
<i>Коляда М.Г.</i> О психологической классификации моделей обучаемых.....	101
<i>Бондаренко М.Ф., Работягов А.В., Щепковский С.В.</i> Современные методы кодирования речевого сигнала.....	106
<i>Левыкин В.М., Чалый С.Ф.</i> Оценивание полноты и непротиворечивости данных для знание-ориентированного представления бизнес-процессов	115
<i>Петров Э.Г., Чайникова Т.С.</i> Методы принятия решений в инвестиционном анализе	120
<i>Кораблев Н.М., Сорокина И.В.</i> Адаптивные нечеткие модели идентификации нелинейных объектов на основе искусственных иммунных систем	125
<i>Кочина М.Л., Калиманов В.Г.</i> Исследование и моделирование поляризационно-оптических свойств роговицы глаза при различных состояниях экстраокулярных мышц.....	132
<i>Олейник А.А., Субботин С.А.</i> Метод оптимизации на основе моделирования перемещения бактерий с применением эволюционных операторов.....	137
<i>Ольшевский А.И., Починский М.Ю.</i> Решение задачи Штейнера с помощью генетического алгоритма.....	145
<i>Кобозева А.А.</i> Анализ сингулярных чисел матрицы цифрового сигнала как основа подхода к обнаружению его фальсификации	152
<i>Герасин С.Н., Козлов М.А.</i> Оценка параметров численности популяции в дискретной модели Рикера	159
<i>Коноплянко З.Д.</i> Концепції організації інформаційно-інтелектуальних технологій	164
<i>Сайковская Л.Ф.</i> Результаты исследования частотных характеристик зрительной системы с использованием автоматизированного прибора	173
Об авторах	177
Правила оформления рукописей для авторов научно-технического журнала «Бионика интеллекта».....	179

УДК 519.7



КОМПАРАТОРНАЯ ИДЕНТИФИКАЦИЯ ЦВЕТОВОГО ЗРЕНИЯ ЧЕЛОВЕКА

М.Ф. Бондаренко¹, С.Ю. Шабанов-Кушнаренко², Ю.П. Шабанов-Кушнаренко³

^{1, 2, 3} ХНУРЭ, г. Харьков, Украина

Хотя теория компараторной идентификации представляет собой вполне самостоятельную область знания, однако, она имеет много точек соприкосновения с теорией прямой идентификации. В связи с этим здесь рассматриваются те достижения классической теории идентификации, на которые пришлось опираться при разработке вопросов компараторной идентификации.

КОМПАРАТОРНАЯ ИДЕНТИФИКАЦИЯ, ЦВЕТОВОЕ ЗРЕНИЕ, ГИЛЬБЕРТОВО ПРОСТРАНСТВО

Введение

В настоящее время теория компараторной идентификации еще накрепко привязана к конкретной задаче математического описания процесса цветового зрения человека, излагается в ее терминах. Исторически это вполне объяснимо, поскольку именно из недр теории цветового зрения она и выросла. Но теория компараторной идентификации во многом бы выиграла, если бы она была изложена всецело в нейтральных математических терминах. Это способствовало бы ее более успешному применению в других областях науки и техники. В работе ставится задача изложить теорию идентификации линейных конечномерных объектов на чисто математическом языке, отделив ее от содержательных интерпретаций и приложений. Вместе с тем, ставится задача дать психологическую, техническую и социально-экономическую интерпретации теории компараторной идентификации с целью облегчить и стимулировать ее распространение на практике.

1. Прямая идентификация линейных объектов

Наиболее полно теория прямой идентификации разработана для линейных объектов, описываемых отображением вида:

$$y(t) = \int_a^b x(\tau) K(r, t) dr. \quad (1)$$

Отображение (1) ставит в соответствие входному сигналу $x(\tau)$ объекта его выходной сигнал $y(t)$. Предполагается, что $x(\tau)$ и $y(t)$ — функции вещественных аргументов с вещественными значениями. Аргументы τ и t этих функций определены на интервале $[a, b]$, где a и b — фиксированные числа. Под знаком интеграла в выражении (1) стоит функция Грина, которая выступает в роли числовой характеристики линейного объекта [1].

В теории идентификации объектов, описываемых зависимостью (1), формулируют следующие две основные задачи: 1) задачу структурной идентификации, заключающуюся в отыскании системы свойств объекта, которая определяла бы вид отображения (1); 2) задачу параметрической иден-

тификации, состоящую в определении функции K , характеризующей конкретное преобразование вида (1) [2, 3].

Для решения задачи структурной идентификации обычно используют свойства аддитивности

$$F(x_1(\tau) + x_2(\tau)) = F(x_1(\tau)) + F(x_2(\tau)) \quad (2)$$

и однородности

$$F(\alpha x(\tau)) = \alpha F(x(\tau)) \quad (3)$$

объекта [4]. Символом F обозначено произвольное отображение сигнала $x(\tau)$ в сигнал $y(t)$. Отображение (1) удовлетворяет свойству аддитивности при любом выборе функций $x_1(\tau)$, $x_2(\tau)$ и свойству однородности при любом выборе функции $x(\tau)$ и вещественного числа α [5].

В трудах по идентификации иногда упоминается еще свойство непрерывности объекта [6]. Стандартный способ введения непрерывности отображения F сигнала $x(\tau)$ в сигнал $y(t)$ состоит в следующем [7]. Предполагается, что совокупность всевозможных входных сигналов $x(\tau)$ объекта образует вещественное гильбертово пространство $L_2[a, b]$, выходные сигналы $y(t)$ принадлежат тому же пространству. Для каждой функции $x(\tau)$ из $L_2[a, b]$, вводится ее норма по формуле:

$$\|x\| = \sqrt{\int_a^b x^2(\tau) d\tau}. \quad (4)$$

По этой же формуле вычисляется норма и для функций $y(t)$. Далее определяется сходимость по норме, а именно принимается, что x_n сходится к x , если $\|x_n - x\|$ сходится к нулю. Сходимость числовой последовательности c_n к нулю означает, что для любого вещественного числа $\varepsilon > 0$ найдется такой номер N , что при всех $n > N$ будет выполняться неравенство $|c_n| < \varepsilon$. Отображение F называют непрерывным относительно функции $x(\tau)$, если из сходимости x_n к x следует сходимость $y_n = F(x_n)$ к $y = F(x)$. Отображение называют непрерывным, если оно непрерывно относительно всех функций, принадлежащих пространству $L_2[a, b]$ [8].

Согласно теореме об общем виде линейного оператора [9] любое аддитивное и непрерыв-

ное отображение из пространства $L_2[a, b]$ в то же пространство имеет вид (1). Вместе с тем, известно [10], что аддитивность и однородность задают класс отображений более широкий, чем множество всех отображений вида (1).

Из учения о параметрической идентификации объектов, описываемых отображением (1), упомянем методы импульсного и ступенчатого возмущения. Метод импульсного возмущения [11] состоит в том, что на вход объекта при различных значениях параметра $r = \theta$ подается достаточно узкий импульс, диаграмма которого охватывает единичную площадь. Формально импульс описывается функцией Дирака $\delta(\tau - \theta)$. Реакция объекта $y_\theta(t) = F(\delta(\tau - \theta))$ совпадает с функцией Грина

$$K(\tau, t) = y_\tau(t). \quad (5)$$

Метод ступенчатого возмущения [12] заключается в том, что на вход объекта при различных значениях параметра $\tau = \theta$ подается единичный скачок:

$$J(\tau - \theta) = \begin{cases} 0, & \text{если } \tau < \theta, \\ 1, & \text{если } \tau \geq \theta. \end{cases} \quad (6)$$

Пусть $y_\theta(t) = F(J(\tau - \theta))$ — реакция объекта на единичный скачок. Тогда функция Грина для отображения (1) может быть вычислена по формуле:

$$K(\tau, t) = \frac{\partial y_\tau(t)}{\partial t}. \quad (7)$$

При численных расчетах прибегают к квантованию интервала $[a, b]$, на котором заданы сигналы $x(\tau)$ и $y(t)$ [13]. Обычно используют равномерный шаг квантования. Число квантов на интервале может быть различным для входного и выходного сигналов. Пусть m — число квантов для входного сигнала, n — для выходного. Схема квантования входного сигнала показана на рис. 1.

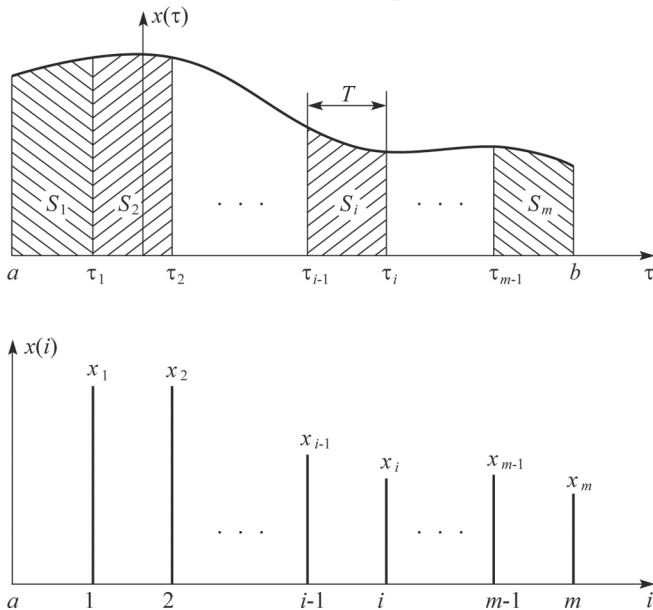


Рис. 1

Шаг квантования $T = (b - a)/m$. Символами $S_1, S_2, \dots, S_i, \dots, S_m$ обозначены площади под кривой $x(\tau)$ на интервалах $[a, \tau_1], [\tau_1, \tau_2], \dots, [\tau_{i-1}, \tau_i], \dots, [\tau_{m-1}, b]$, где $\tau_i = a + iT$. Функция $x(\tau)$ после квантования превращается в решетчатую функцию $x(i)$, значения которой $x_i = x(i)$ равны $x_i = S_i / T$. Они теперь зависят от номера i , который изменяется в пределах от 1 до m . Аналогично квантуется выходной сигнал $y(t)$. После квантования он превращается в решетчатую функцию $y(j)$, зависящую от номера j , который изменяется в пределах от 1 до n [14].

После квантования входной сигнал объекта можно рассматривать как m -мерный вектор $x = (x_1, x_2, \dots, x_m)$, а его выходной сигнал — как n -мерный вектор $y = (y_1, y_2, \dots, y_n)$. Функция Грина отображения (1), реализуемого объектом, превращается в матрицу $K = \|k_{ij}\|$ размера $m \times n$. Само же отображение (1) превращается в линейный оператор, отображающий m -мерное вещественное векторное пространство в n -мерное вещественное векторное пространство и описываемый равенством:

$$y = x * k. \quad (8)$$

Операция, обозначенная звездочкой, есть умножение вектора на матрицу [15, 16].

2. Математическое описание цветового зрения человека

Идея компараторной идентификации впервые возникла при разработке учения о цветовом зрении человека. Именно там она получила свое развитие и признание. Поэтому учение о цветовом зрении человека выполняет роль фундамента, на котором основывается разработка теории компараторной идентификации. Глаз человека воспринимает световые лучи, исходящие от предметов окружающего мира. Преобразуясь в зрительной системе, они превращаются в субъективный цветовой образ этих предметов. Образ воспринимаемого внешнего мира формируется в поле зрения, в каждой точке которого возникает вполне определенный цвет. Последний представляет собой ответную реакцию органа зрения на световое излучение, действующее на определенный участок сетчатки глаза. Зрительную систему человека можно рассматривать как преобразователь сигналов, а именно — как преобразователь светового излучения, исходящего из данной точки внешнего предмета, в цвет этой точки.

Входные сигналы зрительной системы человека, то есть световые лучи, вполне доступны для объективного измерения и анализа. Этого, однако, нельзя сказать о цветах — выходных сигналах органа зрения. Цвет зрительного ощущения — это субъективное состояние человека, поэтому он не-

доступен для прямого физического наблюдения. Правда, можно измерить физические параметры процессов, разыгрывающихся в тех участках коры мозга, которые формируют зрительные ощущения. Но как убедиться в том, что именно эти параметры несут информацию о цвете? Сделать это можно единственным способом — сравнить результаты физиологических измерений с информацией о самом зрительном ощущении. А для этого надо располагать данными о субъективном цвете. Таким образом, круг замыкается.

Выход из этой тупиковой ситуации впервые указал И. Ньютон [17]. Он воспользовался тем обстоятельством, что сам испытуемый (то есть человек, над которым производятся опыты) способен установить, равны или нет цвета любых предъявленных ему световых излучений. Делает он это субъективным способом, анализируя цвета собственным сознанием. В результате такого анализа человек может выработать физический сигнал, принимающий одно из двух значений: 1, если цвета совпадают, и 0, если не совпадают. Эта двоичная реакция испытуемого вполне объективна, ее можно зарегистрировать физическими приборами. Гениальность Ньютона проявилась в том, что он обнаружил в таком, казалось бы побочном и малоинформативном сигнале, принимающем всего два значения, богатый источник информации о цвете. В работе [18] показано, что этот сигнал содержит в себе вообще всю информацию о цвете.

Другое принципиально важное достижение Ньютона состояло в том, что он ввел понятие спектра светового излучения, разработал способ измерения спектра и показал, что цвет однозначно определяется спектром вызвавшего его светового излучения. Этим Ньютон фактически создал метод математического описания входных сигналов зрительной системы человека. Благодаря этому результату входные сигналы органа зрения оказалось возможным формально представлять в виде точек линейного пространства.

Выбрав три фиксированные световые излучения, Ньютон с помощью специально созданной им аппаратуры сумел осуществить регулирование интенсивности каждого из них и последующее совмещение полученных в результате этого регулирования трех излучений в пространстве. Этим он практически осуществил сложение излучений и умножение их на коэффициенты. Пользуясь введенными операциями, Ньютон установил, что любое световое излучение можно подравнять по цвету к смеси трех фиксированных излучений, причем пропорции интенсивности излучений в смеси однозначно определяются цветом сравниваемого излучения. Этот опыт в дальнейшем послужил основой для математического описания цвета в виде трехмерного вектора. Ньютон первым стал изобра-

жать цвет в виде точки физического пространства, что, по существу, означает введение векторного пространства выходных сигналов зрительной системы человека.

Изучая закономерности подравнивания излучений по цвету, Ньютон установил, что вектор цвета смеси излучений может быть найден по тому же правилу, по которому в механике отыскивается центр тяжести системы точечных масс, расположенных в пространстве. Массы точек принимаются численно равными интенсивностям смешиваемых излучений. Эти массы должны быть расположены в точках трехмерного пространства, которые соответствуют цветам смешиваемых излучений. Тогда центр тяжести системы всех масс всегда оказывается совпадающим с точкой, характеризующей цвет смеси излучений. Этими опытами Ньютон вплотную подошел к выводу о линейности преобразования светового излучения в цвет зрительного ощущения.

Основываясь на опытах Ньютона, Т. Юнг [19, 20] первым сформулировал тезис о линейности преобразования светового излучения в цвет. Это преобразование он представил в виде следующих трех интегралов:

$$a_1 = \int_{\lambda_1}^{\lambda_2} x(\lambda) K_1(\lambda) d\lambda,$$

$$a_2 = \int_{\lambda_1}^{\lambda_2} x(\lambda) K_2(\lambda) d\lambda,$$

$$a_3 = \int_{\lambda_1}^{\lambda_2} x(\lambda) K_3(\lambda) d\lambda.$$

Спектр светового излучения Юнг представил в виде вещественной функции $x(\lambda)$ вещественного аргумента λ , взаимно однозначно связанного с длиной волны света. Числа λ_1 и λ_2 характеризуют граничные длины волн света, видимого испытуемым. Символами a_1 , a_2 , a_3 обозначены вещественные числа, играющие роль координат цвета, порождаемого излучением $x(\lambda)$ в сознании испытуемого. Юнг впервые ввел три функции $K_1(\lambda)$, $K_2(\lambda)$, $K_3(\lambda)$ для числовой характеристики преобразования светового излучения в цвет, которые назвал функциями спектральной чувствительности зрения.

К. Максвелл [21] был первым, кто поставил задачу опытного определения функций спектральной чувствительности зрения. Для ее решения он создал специальную оптическую установку, с помощью которой подравнивал по цвету монохроматические световые излучения единичной интенсивности с различными длинами волн к смеси трех фиксированных излучений (тоже монохроматических). В результате этих опытов Максвелл построил специальный график, названный им цветовым треугольником. Этими опытами Максвелл

положил начало параметрической идентификации линейного объекта компараторным методом. По существу, здесь он впервые использовал прием, который в классической теории идентификации впоследствии был назван методом единичного импульсного возмущения.

Начало развитию структурной идентификации линейных объектов компараторным методом положил Г. Грассман [22]. Базируясь на опытах Ньютона подравнивания излучений по цвету, он впервые сформулировал некоторые из характеристических свойств поведения испытуемого, устанавливающего равенство или неравенство цветов световых излучений. Впоследствии эти свойства были названы законами Грассмана. Формулируются они следующим образом. Закон непрерывности: «Непрерывному изменению излучения соответствует непрерывное изменение цвета» [22, с. 72]. Закон трехмерности: «Для любого излучения можно подобрать одинаково выглядящую смесь белого излучения с некоторым чистым спектральным или же пурпурным излучением. Под пурпурным излучением понимается смесь крайних видимых излучений спектра» [22, с. 78]. Закон аддитивности: «Одинаково выглядящие излучения дают при сложении одинаково выглядящие излучения» [22, с. 82]. Впоследствии эти законы послужили отправным пунктом для поисков полной системы свойств, определяющей линейный объект при его компараторной идентификации.

Первым, кто поставил вопрос о разработке объективно проверяемой полной системы свойств поведения испытуемого, сравнивающего цвета световых излучений, был Э. Шредингер [23]. Он впервые дал строго физическое определение цвета как класса всевозможных световых излучений, порождающих один и тот же субъективный цвет (такие излучения называют метамерными). На этой основе он смог дать объективное определение операций сложения цветов и умножения цвета на вещественные коэффициенты. Это, в свою очередь, позволило ему по-новому сформулировать закон трехмерности: «Существует линейно независимая тройка цветов. Любые четыре цвета всегда линейно зависимы» [23, с. 431] и исключить элементы субъективизма при интерпретации закона непрерывности (теперь непрерывность изменения цвета понималась в строго физическом смысле). Шредингер был первым, кто привлек внимание к закону однородности как к одному из важных характеристических свойств линейного объекта при его компараторной идентификации. Он сообщает, что формулировка закона однородности принадлежит Э. Герингу: «Одинаково выглядящие излучения остаются одинаково выглядящими после повышения или понижения интенсивности каждого из них в одинаковых размерах» [23, с. 422].

Наконец, упомянем работу Г. Шайбнера [24], в которой впервые указывается на необходимость включения в число характеристических свойств поведения испытуемого, сравнивающего цвета излучений, законов рефлексивности (равные излучения порождают равные цвета), симметричности (равноцветные излучения после изменения порядка их предъявления испытуемому остаются равноцветными) и транзитивности (если цвет излучения x совпадает с цветом излучения y и цвет излучения y совпадает с цветом излучения z , то цвет излучения x совпадает с цветом излучения z).

3. Компараторная идентификация линейных объектов

Перечисленные в предыдущем разделе достижения теории цветового зрения стимулировали появление ряда чисто математических работ по компараторной идентификации линейных объектов. Впервые полная система свойств линейного объекта была описана и корректно математически обоснована в работе [25] для случая, когда этот объект отображает вещественное гильбертово пространство $L_2[0, 1]$ в n -мерное арифметическое пространство R^n . В этой работе введено понятие линейно-функционального предиката E , заданного на декартовом квадрате пространства $L_2[0, 1]$, который может быть представлен в виде:

$$E(x, y) = D\left(\int_0^1 x(t)a(t) dt, \int_0^1 y(t)a(t) dt\right). \quad (10)$$

Здесь D — предикат равенства; $a(t) = (a_1(t), a_2(t), \dots, a_n(t))$; $\{a_i\}_{i=1}^n$ — система линейно независимых функций из $L_2[0, 1]$.

Предикат E называется симметричным, если из $E(x, y) = 1$ следует $E(y, x) = 1$; транзитивным, если из $E(x, y) = E(y, z) = 1$ следует $E(x, z) = 1$; и аддитивным, если из $E(x, y) = E(u, v) = 1$ следует $E(x + u, y + v) = 1$ для любых x, y, z, u, v из $L_2[0, 1]$. Предикат E называется n -мерным, если существует система функций $\{e_k\}_{k=1}^n$ из $L_2[0, 1]$ такая, что для всякой функции x из $L_2[0, 1]$ найдется единственный набор вещественных чисел $(\alpha_1, \alpha_2, \dots, \alpha_n)$, удовлетворяющий условию:

$$E(x, \sum_{i=1}^n \alpha_i e_i) = 1. \quad (11)$$

Если функционалы $\alpha_1(x), \alpha_2(x), \dots, \alpha_n(x)$ непрерывны в $L_2[0, 1]$, то предикат E называется непрерывным.

Формулируется и доказывается следующая теорема об условиях существования линейно-функционального предиката. Для того чтобы предикат E был линейно-функциональным, необходимо и достаточно, чтобы он обладал свойствами симметричности, транзитивности, аддитивности, n -мерности и непрерывности. Кроме того, доказывает-

ся теорема о том, что свойства симметричности, транзитивности, аддитивности, n -мерности и непрерывности предиката E независимы.

Условия симметричности и транзитивности представляют собой формальную формулировку условий симметричности и транзитивности, указанных Шайбнером. Условие аддитивности математически формулирует закон аддитивности Грассмана. Условие n -мерности есть результат обобщения и формализации процедуры подравнивания излучений по цвету, предложенной Ньютоном. Условие непрерывности математически оформляет закон непрерывности, намеченный в работе Шредингера.

Смысл теоремы об условиях существования линейно-функционального предиката состоит в том, что он указывает полную систему свойств, определяющих линейный объект. Эта теорема решает задачу структурной идентификации линейного объекта в случае, когда он преобразует функции одного вещественного аргумента в n -мерные векторы с числовыми компонентами. Вторая теорема доказывает, что указанная система из пяти аксиом не может быть сокращена, то есть что все эти аксиомы должны быть проверены в эксперименте для решения вопроса о том, является ли данный объект линейным или нет. В работе [26] было установлено, что аксиомы симметричности и транзитивности в полной системе могут быть заменены одним условием рефлексивности (для любого x $E(x, x) = 1$). Этим достигнуто определенное упрощение системы характеристических свойств линейного объекта.

В терминах учения о цветовом зрении человека изложенные математические результаты интерпретируются следующим образом. Элементы $x(t)$ пространства $L_2[0, 1]$ интерпретируются как спектры излучений $x(\lambda)$, переход от переменной t к переменной λ осуществляется по формуле $\lambda = \lambda_1 + t(\lambda_2 - \lambda_1)$. Предикат $E(x, y)$ формально описывает поведение испытуемого, сравнивающего цвета излучений x и y . Значение 1 предиката E соответствует факту равенства цветов, значение 0 — факту их неравенства. Сложение элементов пространства $L_2[0, 1]$ и их умножение на вещественные числа понимаются как сложение излучений и их умножение на коэффициенты. Число n принимается равным трем. Числа $\alpha_1, \alpha_2, \alpha_3$, фигурирующие в законе n -мерности (при $n = 3$), понимаются как пропорции интенсивности излучений e_1, e_2, e_3 в смеси, которые обеспечивают цветовое равенство этой смеси с излучением x . Выполнение в опытах на испытуемом системы свойств симметричности, транзитивности, аддитивности, n -мерности и непрерывности (или рефлексивности, аддитивности, n -мерности и непрерывности) означает, что преобразование светового

излучения в цвет, осуществляемое органом зрения испытуемого, может быть формально описано интегралами Юнга (9).

Имеются работы [27–29], в которых исследуются условия существования линейного оператора, идентифицируемого компараторным методом, для положительной части линейного пространства $L_2[0, 1]$ и для произвольного выпуклого подмножества. Появление этих работ мотивировано тем, что спектры световых излучений всегда неотрицательны и, кроме того, глаз человека не способен воспринимать излучения очень малой и очень большой интенсивности. Во всех этих работах авторы приходят к весьма сложным формулировкам характеристических свойств линейного объекта.

Следует отметить работу [30], в которой для случая положительной части гильбертова пространства удалось использовать простые характеристические свойства поведения испытуемого, совпадающие по форме с законами аддитивности, трехмерности и непрерывности, полученные Шредингером для всего пространства. Упрощение достигнуто за счет введения функции $\beta(\lambda) = x(\lambda) - y(\lambda)$, представляющей собой разность спектров $x(\lambda), y(\lambda)$ световых излучений, предъявляемых испытуемому для сравнения их цветов. В законах цветового зрения теперь вместо излучений используются соответствующие им функции $\beta(\lambda)$.

В работе [28] описаны три метода экспериментального определения функций $K_1(\lambda), K_2(\lambda), K_3(\lambda)$ для интегралов Юнга (9). Первый метод родственен методу единичного импульсного возмущения, используемого при прямой идентификации линейного объекта. В роли входного сигнала $x(\lambda)$ используются всевозможные монохроматические излучения единичной интенсивности $\delta(\lambda - \lambda_0)$ с длиной волны λ_0 , изменяющейся в пределах от λ_1 до λ_2 . Отыскиваемые для них значения подравнивающих коэффициентов $\alpha_1(\lambda_0), \alpha_2(\lambda_0)$ и $\alpha_3(\lambda_0)$ используются в роли функций спектральной чувствительности зрения $K_1(\lambda_0), K_2(\lambda_0), K_3(\lambda_0)$. Этот метод по существу идентичен методу Максвелла и является его формальным оформлением в современных математических терминах.

Второй метод является аналогом метода единичного ступенчатого возмущения, используемого при прямой идентификации линейного объекта. Испытуемому предъявляется световое излучение, спектр которого имеет вид единичного скачка при произвольном значении длины волны λ_0 :

$$J(\lambda - \lambda_0) = \begin{cases} 0, & \text{если } \lambda < \lambda_0, \\ 1, & \text{если } \lambda \geq \lambda_0. \end{cases} \quad (12)$$

Для него отыскиваются в эксперименте значения подравнивающих коэффициентов $\alpha_1(\lambda_0), \alpha_2(\lambda_0)$ и $\alpha_3(\lambda_0)$. Функции спектральной чувствительности зрения отыскиваются по формуле:

$$K_i(\lambda) = \frac{d(\alpha_i(\lambda))}{d\lambda}, \quad (i=1,2,3). \quad (13)$$

Третий метод можно назвать методом обобщенного скачка. Суть его состоит в том, что берется произвольное световое излучение $x(\lambda)$, с тем, однако, ограничением, что оно ни в одной из точек на интервале $[\lambda_1, \lambda_2]$ не обращается в нуль. С помощью призмы и ширмы от него отрезается участок спектра до длины волны λ_0 , а для оставшейся части излучения $x_{\lambda_0}(\lambda)$ отыскиваются подравняющие коэффициенты $\alpha_1(\lambda_0), \alpha_2(\lambda_0)$ и $\alpha_3(\lambda_0)$. Функции спектральной чувствительности зрения отыскиваются по формуле:

$$K_i(\lambda) = \frac{1}{x(\lambda)} \cdot \frac{d(\alpha_i(\lambda))}{d\lambda}, \quad (i=1,2,3). \quad (14)$$

4. Компараторная идентификация метрики конечномерного линейного пространства

Возьмем какое-нибудь световое излучение A со спектром $x(\lambda)$, которое на заданном фоне воспринимается испытуемым как белый цвет. Если уменьшить интенсивность излучения A в 50 раз, то получим излучение со спектром $0,02x(\lambda)$. Обозначим его буквой B . Если испытуемому предъявить излучение B на том же фоне, то он воспримет его как черный цвет. Возьмем, наконец, излучение C , среднее между A и B . Оно имеет спектр $0,51x(\lambda)$. На том же фоне испытуемый воспримет его как серый цвет. Поскольку преобразование излучения в цвет, осуществляемое зрительной системой человека, описывается линейной зависимостью, то естественно ожидать, что цвет излучения C будет выглядеть как средний между цветами излучений A и B . Однако опыт этого не подтверждает. Фактически цвет излучения C оказывается гораздо ближе к цвету излучения A , чем к цвету излучения B . Чтобы получить цвет, равноотстоящий от цветов, вызываемых излучениями A и B , потребуется излучение $0,25x(\lambda)$ по интенсивности примерно в два раза меньшее, чем излучение C [31].

Описанный факт наводит на мысль, что утверждение о линейности преобразования светового излучения в цвет неверно. В действительности противоречия между опытом и линейной теорией цветового зрения в данном случае нет, однако, дать объяснение этому парадоксу непросто. Чтобы разобраться в нем, придется опереться на результаты, изложенные в первом разделе этой статьи. Если предикат $E(x, y)$, заданный на каком-либо множестве M , рефлексивен, симметричен и транзитивен (а именно таков предикат, реализуемый испытуемым, который сравнивает цвета двух световых излучений), то всегда найдется множество N и функция $F: M \rightarrow N$ такие что

$$E(x, y) = D(F(x), F(y)) \quad (15)$$

для всех $x, y \in M$. Имеется в виду, что D — это предикат равенства, заданный на множестве N . Элементы множества N можно интерпретировать как цвета, а функцию F — как преобразование светового излучения в цвет. Предикат D интерпретируется как выполняемая испытуемым операция сравнения цветов.

Если предикат E удовлетворяет также условиям аддитивности, трехмерности и непрерывности и задан на гильбертовом пространстве $L_2[\lambda_1, \lambda_2]$, то, как было сказано ранее, в роли множества N можно взять трехмерное векторное пространство R^3 , а в роли функции F — линейный оператор, отображающий пространство $L_2[\lambda_1, \lambda_2]$ в пространство R^3 . Обратим внимание на существенное различие между прямой идентификацией линейного объекта и его компараторной идентификацией. При прямой идентификации, если объект удовлетворяет условиям аддитивности и непрерывности, то отсюда с необходимостью следует, что он линейен и что его формальное описание должно иметь вид (1). При компараторной же идентификации, если предикат E удовлетворяет условиям аддитивности, n -мерности и непрерывности, то отсюда следует лишь то, что объект можно описать с помощью линейного оператора.

При прямой идентификации объекта описывающий его оператор оказывается единственным [32], при компараторной идентификации вариантов возможного формального описания объекта имеется много, и линейный оператор — это всего лишь один из них.

Если существуют два представления одного и того же предиката $E(x, y) = D(F(x), F(y))$ и $E(x, y) = D_1(F_1(y))$, где $F: M \rightarrow N$ и $M \rightarrow N$, то существует биекция $\varphi: N \rightarrow N$ такая, что $\varphi F(x) = F_1(x)$ при любом $x \in M$.

Кроме того, если предикат E может быть описан в виде $E(x, y) = D(F(x), F(y))$, и известно, что существует биекция $\varphi: N \rightarrow N'$ такая, что $\varphi(F(x)) = F_1(x)$ при любом $x \in M$, то предикат E может быть также формально представлен и в виде $E(x, y) = D_1(F_1(x), F_1(y))$. Перейдем от общего случая произвольной функции F к частному случаю, когда F — это линейный оператор из $L_2[\lambda_1, \lambda_2]$ на R^3 . Тогда множество N превращается в трехмерное векторное пространство R^3 . Будем считать, что множество N' — тоже трехмерное векторное пространство. Тогда φ — это произвольный взаимно однозначный оператор (в общем случае нелинейный), отображающий пространство N на пространство N' . Каждый объект идентификации, удовлетворяющий условиям рефлексивности, симметричности, транзитивности, аддитивности, трехмерности и непрерывности может быть формально описан любым из операторов вида $\varphi(F(x))$, причем других операторов, пригодных для такого описания, не существует.

Если брать в роли φ различные взаимно однозначные линейные операторы, отображающие пространство R^3 на R^3 , то окажется, что один и тот же объект, удовлетворяющий условиям рефлексивности, симметричности, транзитивности, аддитивности, трехмерности и непрерывности, может быть формально описан многими линейными операторами вида $\varphi F(x) = \varphi(F(x))$. Таким образом, даже в классе линейных операторов объект, удовлетворяющий только что упомянутым условиям, при компараторной идентификации допускает различные варианты своего математического описания. Важно отметить, что все эти варианты описания объекта совершенно равноправны. Невозможно отдать предпочтение ни одному из них, если основываться только на той информации, которая содержится в предикате E . Таким образом, вид оператора φ остается неопределяемым, он не может быть выведен из поведения испытуемого, устанавливающего равенство или неравенство цветов предъявляемых ему световых излучений.

После сделанных замечаний уже нетрудно объяснить встретившийся выше парадокс. Несовпадение фактически наблюдаемого в опыте среднего цвета с цветом, определяемым как средний в предположении линейности оператора F_φ , означает лишь то, что функция φ не является линейной. Таким образом, оператор $F_\varphi = \varphi(F(x))$ нелинеен, он не совпадает с линейным оператором F . При таком положении дела идентификацию преобразования светового излучения в цвет нельзя считать завершенной. Математическое описание этого преобразования будет окончено лишь после того, когда, кроме линейной части F преобразования φF , будет найдена также и его нелинейная часть φ . Ниже описываются литературные данные, которые вносят вклад в решение этой задачи.

Впервые на факт нелинейности преобразования физического стимула в ощущение обратил внимание Э. Вебер [33]. Он сформулировал закон, которому впоследствии было присвоено его имя. Согласно ему едва заметные различия интенсивности светового излучения dJ пропорциональны уровню самого излучения J . Иными словами:

$$\frac{dJ}{J} = \text{Const.} \quad (16)$$

Г. Фехнер [34] вывел отсюда, пользуясь некоторыми дополнительными предположениями, логарифмическую зависимость, согласно которой светлота B цвета равна

$$B = a \lg bJ, \quad (17)$$

где a и b — константы, определяемые опытным путем.

Д. Шкловер [35] распространил эту зависимость на все трехмерное пространство цветов, предложив

производить логарифмическое преобразование его по осям физиологической системы КЗС. Пользуясь этими предпосылками, он нашел конкретный вид преобразования φ , обеспечивающий приблизительное совпадение фактических средних цветов с цветами, найденными по зависимости F_φ как средние. Поискам положения осей физиологической системы посвящены исследования многих авторов [36, 37]. Информация о положении осей извлекается из сравнения реакций на свет лиц с патологией цветового восприятия, у которых зрение двухкомпонентно (дихроматов). Имеется три вида такой частичной цветовой слепоты — протанопия, дейтеранопия и тританопия. Дихроматы не различают цвета, которым в цветовом пространстве N соответствуют точки, лежащие на определенном образом ориентированных прямых линиях. Каждому виду цветовой слепоты соответствует свое семейство прямых линий, параллельных одной из осей цветового пространства в системе КЗС.

Другое направление исследований, также приводящее к установлению нелинейной связи светлоты цвета с интенсивностью соответствующего светового излучения, основывается на практике построения цветных атласов. Дело в том, что цветные атласы стремятся строить таким образом, чтобы соседние цветные накраски в любом месте атласа находились друг от друга по цвету на одном и том же расстоянии [38]. Наилучшим с этой точки зрения считается атлас А. Манселла [39], построенный на базе степенной зависимости светлоты цвета от интенсивности вызвавшего его излучения [40]:

$$B = cJ^d, \quad (18)$$

где c и d — константы, определяемые опытным путем. Вопрос о том, какая зависимость — логарифмическая или степенная (или какая-либо иная) лучше соответствует фактам зрения, пока еще окончательно не решен.

Большое число исследований было предпринято с целью построения равноконтрастной диаграммы цветности. Задача состоит в том, чтобы так деформировать цветовой треугольник Максвелла, чтобы геометрическое расстояние между любыми его точками стало равным психологическому расстоянию между любыми цветами, соответствующими этим точкам. Первая практически приемлемая диаграмма цветности была построена Д. Джаддом [41, 42], изучавшим величину цветового порога по разным направлениям для различных точек цветового треугольника. На основании этих опытов Джадд предложил проективное преобразование цветового треугольника в диаграмму цветности, обеспечивающее приближенно ее равноконтрастность.

Д. Мак-Адам [43] на основании более точных экспериментов скорректировал вид искомого проективного преобразования. Л. Зильберштейн [44]

исследовал погрешность приближения, получаемую при проективных преобразованиях цветового треугольника и сопоставил ее с точностью выполненных Мак-Адамом опытов. В результате он пришел к обоснованному выводу о невозможности получения проективными преобразованиями равноконтрастной диаграммы, согласующейся с экспериментальными данными. Н.Г. Болдырев и К.Б. Мартынов [45] применили к цветовому треугольнику преобразование более общего вида, а именно — гомеоморфизм, то есть произвольное взаимно однозначное непрерывное отображение плоскости на себя. Ими впервые эмпирически получена равноконтрастная диаграмма и тем самым доказана возможность введения в цветовом пространстве евклидовой метрики.

В свете описанных исследований возникает задача идентификации нелинейного отображения φ цветового пространства N на цветовое пространство N' . Цветовое пространство N называют объективным, пространство N' — субъективным. Из приведенных выше фактов и соображений явствует, что само существование нелинейного отображения φ выводится из факта изменения расстояния между цветами при переходе от пространства N к пространству N' . Значит, существует вполне определенный механизм, с помощью которого зрительная система человека определяет расстояние между точками пространства N' . С математической точки зрения этот факт означает, что на пространстве N' введена некоторая метрика, по всей вероятности евклидова. Именно она и определяет вид функции φ .

Выводы

Выполненный выше анализ источников свидетельствует о том, что учение о компараторной идентификации объектов имеет длительную и богатую историю и большие достижения. Имеется реальный объект — цветовое зрение человека, для которого компараторная идентификация является пока единственным методом математического описания. Весьма вероятно, что компараторная идентификация может быть применена также и для целей математического описания других функций человеческого интеллекта (слух, осязание, узнавание, понимание). Компараторная идентификация была успешно применена для математического описания технических [26, 28, 29] и социально-экономических [46] объектов. Вместе с тем, многие задачи, выдвинутые компараторной идентификацией, еще ждут своего научного и практического решения.

Прежде всего, обнаружилось, что компараторная идентификация имеет принципиальные отличия от прямой идентификации объектов. В отличие от прямой идентификации она расчленяется на две самостоятельные задачи. Первая задача состоит в математическом описании процесса факторизации (объединения в классы) входных сигналов. Если

объект линейный, то факторизация представляет собой процесс линейного проектирования некоторого пространства в пространство меньшей размерности. Вторая задача состоит в математическом описании второго этапа преобразования сигналов. Это преобразование нелинейно, оно осуществляет взаимно однозначное преобразование (деформацию) пространства.

Оба преобразования, встречающиеся как в природных, так и технических объектах, имеют ясно выраженный целесообразный характер. Первое (линейное) преобразование производит сжатие потока информации, сокращение его объема до приемлемых размеров, то есть своего рода обогащение информации. Второе (нелинейное) — преобразует информацию к виду, наиболее удобному для использования. Второе преобразование вводит на множестве выходных сигналов линейного объекта некоторую метрику, обеспечивающую равнозначность измерения сигналов. В то время как компараторная идентификация первого преобразования сигналов во многом родственна задаче прямой идентификации линейных объектов, компараторная идентификация второго преобразования сигналов представляет собой нечто новое, не имеющее аналога в учении о прямой идентификации объектов.

В статье поставлена задача исследования обоих видов компараторной идентификации. В области идентификации линейного преобразования сигналов в качестве центральной стоит задача избавиться от использования свойства непрерывности объекта. Непрерывность формулируется гораздо более сложно, чем все остальные свойства линейного объекта. Вместе с тем возникает подозрение, что она привлекается искусственно, и по существу не обязательна. Дело в том, что пространство выходных сигналов линейного объекта аффинно, оно не имеет своей метрики. Как мы видели, метрика появляется только в результате выполнения второго этапа преобразования сигналов. Однако без фактического использования метрических свойств пространства выходных сигналов линейного объекта непрерывность на нем ввести непросто.

Задача построения теории структурной компараторной идентификации линейного объекта без привлечения свойства непрерывности решается за счет перехода от бесконечномерного гильбертова пространства входных сигналов идентифицируемого объекта к конечномерному. Выше говорилось, что при практических расчетах переходят от описания входных сигналов, заданных первоначально в виде функций, к их описанию в виде конечномерных векторов. При этом никакой потери информации об объекте идентификации фактически не происходит. Ничто не мешает совершить такой переход не только в практике, но и в теории. Ограничение задачи исследования конечномерной постановкой дает еще и ту дополнительную выгоду, что позволяет перевести теорию компаратор-

ной идентификации полностью на абстрактную основу, как это принято в линейной алгебре, рассматривая коэффициенты как скаляры произвольной природы, а не обязательно как вещественные числа. Тем самым достигается более высокий уровень общности теории.

Переход на конечномерный вариант теории дает возможность, кроме того, избавиться от необходимости специально рассматривать различные части пространства (например, положительный конус, выпуклые части пространства). Благодаря существенному упрощению теории оказывается возможным переходить ко всему пространству от любой его части простым ее доопределением и тем самым подводить все встречающиеся задачи линейной идентификации под простейшую теорию, разработанную для всего пространства.

Методы компараторной идентификации метризирующего отображения (то есть деформирующего преобразования пространства), как это видно из обзора литературы, разработаны в гораздо меньшей степени, чем методы идентификации факторизующего (линейного) отображения. Здесь можно говорить пока лишь о разрозненных находках в области параметрической идентификации, вопрос же о структурной идентификации в данной области фактически еще и не ставился.

Дж. Нейман и О. Моргенштерн [47] впервые обратили внимание на глубокую аналогию между задачами идентификации цветового зрения и задачами математического описания социально-экономического поведения людей, подчеркнув психофизический характер последнего. Этим они, по существу, открыли дорогу для применения методов компараторной идентификации при изучении социально-экономических процессов. Значительные результаты в этом направлении получены проф. Э.Г. Петровым [46].

Продemonстрируем метод формального описания социально-экономических процессов на примере выбора человеком места работы. Место работы оценивается человеком по ряду параметров: размер зарплаты, удаленность от места жительства, срок предоставления жилплощади и ее размер, уровень обеспеченности дошкольными учреждениями и тому подобное. Оценка места работы характеризуется вещественным числом, содержательно интерпретируемым как степень привлекательности места работы для лица, занимающегося трудоустройством. Сравнивая два места работы, человек устанавливает, какое из них для него более предпочтительно. Он может также сказать, равноценны ли эти места или нет.

Таким образом, задача выбора места работы аналогична задаче компараторной идентификации процесса преобразования светового излучения в цвет, о которой шла речь выше. Набор чисел, характеризующих место работы, является аналогом спектра светового излучения. Отдельные числа в этом наборе соответствуют линиям спектра све-

тового излучения. Оценка места работы соответствует цвету светового излучения. Она, как и цвет, субъективна и поэтому недоступна для измерения физическими приборами. В отличие от цвета, который трехмерен, оценка места работы одномерна и в этом она больше похожа на яркость, которая тоже одномерна. Человек способен установить, какое место работы лучше, подобно этому наблюдатель определяет, какое световое излучение ярче. Усиливая или ослабляя световое излучение, можно добиться его равенства по яркости любому другому излучению. Подобно этому, изменяя параметры, характеризующие место работы, можно подравнять его по привлекательности к другому месту работы. Операция сравнения оценок двух мест работы соответствует операции сравнения цветов (или яркостей) двух световых излучений.

Отмеченная аналогия обуславливает возможность того, что процесс формирования человеком оценки степени привлекательности места работы можно будет идентифицировать тем же методом, что и процесс преобразования светового излучения в цвет или яркость. Однако при попытке двигаться в этом направлении обнаруживается серьезное препятствие, заключающееся в следующем. Механизм цветового зрения человека, как указывалось выше, был формально представлен в виде линейной модели. Для применения этой модели к только что рассмотренной задаче из области социально-экономического поведения людей необходимо, чтобы поведение человека, сравнивающего оценки мест работы, подчинялось условиям аддитивности и однородности.

Однако опыт свидетельствует, что фактически эти условия не выполняются. И тем не менее, из этого еще не следует, что линейная модель в данном случае неприменима. Не исключено, что изучаемый процесс выглядит нелинейным лишь потому, что были неудачно выбраны шкалы, в которых численно выражены параметры, характеризующие место работы. Поясним сказанное следующим примером. Предположим, что изучая преобразование светового излучения в цвет, мы измеряли бы мощность на отдельных участках спектра светового излучения не в линейной шкале мощности, а в шкале логарифма мощности. Ясно, что при этом мы не получили бы для данного преобразования ни аддитивности, ни однородности. Однако при использовании для измерения световых излучений другой шкалы, а именно — обычной шкалы мощности, рассматриваемое преобразование становится линейным.

Таким образом, столкнувшись с фактом неаддитивности и неоднородности какого-либо социально-экономического поведения людей, не следует отсюда сразу же выводиться его нелинейность. Надо прежде попытаться найти такие шкалы, при которых это поведение стало бы линейным. Итак, в рамках метода компараторной идентификации

мы приходим к задаче отыскания шкал, превращающих нелинейный процесс в линейный. Если ее удастся решить, то это может привести к такому расширению области применения линейных моделей, которое позволит включить в нее многие виды социально-экономического поведения людей [48].

Список литературы: 1. Юсупов Р.М. Получение информации об управляемом процессе в самонастраивающихся системах. — М.-Л.: Энергия, 1966. — 140 с. 2. Ильин В.А., Позняк Э.Г. Линейная алгебра. — М.: Наука, 1974. — 296 с. 3. Основы кибернетики. Теория кибернетических систем / Под ред. К.А. Пупкова. — М.: Высшая школа, 1976. — 405 с. 4. Ордынцев В.М. Математическое описание объектов автоматизации. — М.: Машиностроение, 1965. — 357 с. 5. Ахизер Н.И., Глазман И.М. Теория линейных операторов в гильбертовом пространстве. — Х.: Вища школа, 1977. — 315 с. 6. Райбман Н.С., Чадеев В.М. Построение моделей процессов производства. — М.: Энергия, 1975. — 372 с. 7. Люстерник Л.А., Соболев В.И. Элементы функционального анализа. — М.: Наука, 1965. — 520 с. 8. Попов Е.П. Теория линейных систем автоматического регулирования и управления. — М.: Наука, 1989. — 190 с. 9. Леви П. Конкретные проблемы функционального анализа. — М.: Наука, 1967. — 510 с. 10. Рисс Ф., Секефальви-Надь Б. Лекции по функциональному анализу. — М.: Мир, 1979. — 581 с. 11. Современные методы идентификации систем / Под редакцией П. Эйххофа. — М.: Мир, 1983. — 399 с. 12. Сейдж Э., Мелса Дж. Идентификация систем управления. — М.: Наука, 1974. — 246 с. 13. Галкин В.Я. Численные методы обработки сигналов в задачах идентификации. — М.: Энергия, 1987. — 90 с. 14. Спиди К., Браун Р., Гудвин Дж. Теория управления. Идентификация и оптимальное управление. — М.: Мир, 1973. — 247 с. 15. Гантмахер Ф.Р. Теория матриц. — М.: Наука, 1988. — 548 с. 16. Райбман Н.С. Что такое идентификация? — М.: Наука, 1970. — 117 с. 17. Ньютон И. Оптика или трактат об отражениях, преломлениях, изгибаниях и цветах света. — М.: Гостехтеориздат, 1954. — 365 с. 18. Шабанов-Кушнарченко С.Ю. Компараторная идентификация конечномерных процессов количественной оценки. — дисс. ... доктора техн. наук. Харьков. 1994. — 270 с. 19. Кравков С.В. Глаз и его работа. — М.: Изд-во АН СССР, 1950. — 487 с. 20. Young T. On the Theory of Light and Colours // Phil. Trans. Roy. Soc., V.21, 1801. — P. 12-49. 21. Maxwell J.C. On the theory of compound colours and the relations of the colours of the spectrum // Proc. Roy. Soc. V. 10, 1860. — P. 404-409. 22. Grassman H. Zur Theorie der Farbenmischung // Ann. d. Phys. u. Chemie, Bd. 89, №5, 1853. — P. 69-84. 23. Schrodinger E. Grundlinien einer Theorie der farbenmetrik im Tagessehen. (I Mitteilung) // Ann. d. Phys, Vierte folge, Bd. 63, 1920. — S. 397-426. 24. Scheibner H. On colours of the same appearance // Optica acta, 1966, Vol. 13, №3. — P. 73. 25. Ицков Ф.Э., Шабанов-Кушнарченко Ю. П. Об условиях существования линейно-функционального предиката // Проблемы бионики. — Харьков: Вища школа, 1983. — Вып. 30. — С. 3-10. 26. Шляхов В.В. Математические модели линейных процессов рецепции и их технические приложения: дисс. ... кандидата техн. наук: Харьков, 1984. — 137 с. 27. Сарнавский Н.Г., Шабанов-Кушнарченко Ю. П. Формальные аспекты теории цветового зрения // АСУ и приборы автоматики. — Харьков: Вища школа, 1981. — Вып. 58. — С. 107-114. 28. Пчелинов В.П. Математические модели спектральной чувствительности органа зрения и их технические приложения: дисс. ... кандидата техн. наук: Харьков, 1979. — 176 с. 29. Наталуха Ю.В. Идентификация линейных динамических систем методом нуль-органа:

дисс. ... кандидата техн. наук: Харьков, 1985. — 147 с. 30. Путятин Е.П. Математические модели статических процессов зрения и их использование в технике ввода оптической информации: дисс. ... кандидата техн. наук: Харьков, 1967. — 165 с. 31. Майзель С.О., Ратнер Е.С. Цветовые расчеты и измерения. — М.: Госэнергиздат, 1941. — 187 с. 32. Ту Дж., Гонсалес Р. Принципы распознавания образов: Пер. с англ. — М.: Мир, 1978. — 411 с. 33. Weber E. Tastsinn und Gemeingefuhl // Wagner's Handbuch der Physiologie, 1846, Bd. 3. — P. 96. 34. Fechner G. Elemente der Psychophysik. — Leipzig: 1860. — 237 p. 35. Schklover D. Methodes photoelectriques et appareils pour mesurer la couleur. — Zurich, Comm. int. de l'eclairage, 1955. — 30 p. 36. Нюберг Н. Д. Определение положения в цветовом треугольнике основного синего цвета // ДАН, АН СССР, том XV, 1949, № 2. 37. Мешков В.В. Основы светотехники. Ч. 2. — М.-Л.: Госэнергоиздат, 1961. — 416 с. 38. Луизов А.В. Глаз и свет. — Л.: Энергоатомиздат, 1983. — 137 с. 39. Munsell book of color. — Standard edition, Baltimore, 1929.- 192 p. 40. Мешков В.В., Матвеев А.Б. Основы светотехники. Ч.2. — М.: Энергоатомиздат, 1989. — 429 с. 41. Джадд Д., Вышецки Г. Цвет в науке и технике. — М.: Мир, 1978. — 592 с. 42. Judd D. A Maxwell triangle yielding uniform chromaticity diagram // JOSA, 1935, Vol. 25. — P. 24. 43. Mac-Adam D.L. Representations of colour tolerance on the chromaticity diagram // JOSA, 1939, Vol. 4. — P. 46. 44. Silberstein L., Mac-Adam D. The distribution of color matchings around a color center // JOSA, 1945, Vol. 1. — P.22. 45. Болдырев Н.Г., Мартынов К.Б. Равноконтрастный цветовой график // Проблемы физиологической оптики, т. VIII, 1953. — С.54-63. 46. Петров Э.Г. Организационное управление городом и его подсистемами (методы и алгоритмы). — Харьков: Вища школа. 1986. — 142 с. 47. Нейман Дж., Моргенштерн О. Теория игр и экономическое поведение. — М.: ИЛ, 1970. — 520 с. 48. Орлов А.И. Устойчивость в социально-экономических моделях. — М.: Наука, 1979. — 296 с.

Поступила в редколлегию 21.08.2008

УДК 519.7

Компараторна ідентифікація кольорового зору людини / М.Ф. Бондаренко, С.Ю. Шабанов-Кушнарченко, Ю.П. Шабанов-Кушнарченко // Біоніка інтелекту: наук.-техн. журнал — 2008. — № 2 (69). — С. 3-12.

У статті розвивається теорія компараторної ідентифікації. Розглядаються досягнення класичної теорії ідентифікації, на які довелося опиратися при розробці питань компараторної ідентифікації. Проаналізовано застосування теорії компараторної ідентифікації до кольорового зору людини та соціально-економічного поведіння людей.

Ил. 1. Бібліогр.: 48 найм.

UDC 519.7

The person colour sight comparator identification / M.F. Bondarenko, S.Yu. Shabanov-Kushnarenko, Yu.P. Shabanov-Kushnarenko // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 3-12.

In article the comparator identifications theory is develops. The classical identification theory achievements which it was necessary to lean by working out of questions comparator identifications are considered. The application of the comparator identifications theory to colour sight of the person and social and economic behaviour of people is analysed.

Fig. 1. Ref.: 48 items.

УДК 519.7



ОБ ОБЩЕЙ ТЕОРИИ КОМПАРАТОРНОЙ ИДЕНТИФИКАЦИИ

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

^{1, 2, 3} ХНУРЭ, г. Харьков, Украина

Использование метода компараторной идентификации обеспечивает объективность, высокую точность и надежность математического описания поведения и внутренних состояний человека. Решение этой проблемы позволяет лучше предвидеть последствия взаимодействия человека с машинами, благодаря чему достигается повышение эффективности и экономности АСУ. Метод компараторной идентификации обеспечивает полное (с точностью до обозначений) математическое описание идентифицируемого объекта и благодаря этому не уступает по своей описательной силе классическому методу прямой идентификации.

КОМПАРАТОРНАЯ ИДЕНТИФИКАЦИЯ, ЦВЕТОВОЕ ЗРЕНИЕ, ГИЛЬБЕРТОВО ПРОСТРАНСТВО

Введение

Идентификацией объекта называют определение характеристик этого объекта на базе его опытного исследования. Идентификация является наиболее трудоемкой и ответственной операцией при анализе объектов. Разработка методов идентификации — одна из важнейших проблем современной теории управления. Особо важное значение имеет задача идентификации объектов, которые можно свести к классу линейных объектов. Последние, как известно, наиболее распространены в природе и технике. К настоящему времени теория идентификации превратилась в обширное по содержанию и богатое по методам учение, тем не менее многие актуальные проблемы в ней еще ждут своего решения. В частности, актуальна задача расширения класса объектов, поддающихся эффективной идентификации.

Классическая задача идентификации состоит в том, чтобы по входным и выходным сигналам объекта определить закон преобразования сигналов этим объектом. Такую идентификацию называют прямой, поскольку она осуществляется при непосредственном доступе к выходным сигналам объекта. Вместе с тем, в ряде случаев возникает необходимость в косвенной идентификации объекта, когда у исследователя нет прямого доступа к его выходному сигналу. В данном случае эффективны для применения только методы косвенной идентификации. Наиболее удобен из них метод компараторной идентификации.

1. Об условиях применимости метода компараторной идентификации

Пусть $E(x, y)$ — предикат, заданный на декартовом квадрате непустого множества M . Будем писать xEy , если $E(x, y) = 1$, и $\bar{x}E\bar{y}$, если $E(x, y) = 0$. Предикат E называется рефлексивным, если xEx для всех $x \in M$; симметричным, если для всех $x, y \in M$ из xEy следует yEx ; и транзитивным, если для всех $x, y, z \in M$ из xEy и yEz следует xEz . Любой рефлексивный, симметричный и транзитивный предикат называется предикатом эквивалентности [1]. Пусть N — непустое множество, F — функция, отображающая множество M на мно-

жество N , D — предикат равенства, заданный на $N \times N$. Будем писать uDv , если $D(u, v) = 1$, и $\bar{u}\bar{D}\bar{v}$, если $D(u, v) = 0$.

Утверждение 1. Любой предикат E , заданный на $M \times M$ и выражающийся при любых $x, y \in M$ в виде

$$E(x, y) = D(F(x), F(y)), \quad (1)$$

есть предикат эквивалентности.

Функцию F , фигурирующую в (1), назовем характеристической функцией предиката эквивалентности.

Доказательство. Рефлексивность, симметричность и транзитивность предиката E непосредственно следуют из (1) и из рефлексивности, симметричности и транзитивности предиката равенства D .

Утверждение 2. Для любого предиката эквивалентности E , заданного на $M \times M$, найдутся непустое множество N и функция $F: M \rightarrow N$ такие, что при любых $x, y \in M$ будет выполняться (1).

Доказательство. Для каждого $x \in M$ существует единственное множество S_x всех y таких, что xEy . В роли множества N принимаем систему всех множеств S_x . Множество N не пусто. В роли F принимаем функцию, которая ставит в соответствие каждому элементу $x \in M$ множество S_x так, что $F(x) = S_x$. Докажем, что при таком выборе функции F равенство (1) выполняется при любых $x, y \in M$. Рассмотрим случай, когда x и y таковы, что xEy . Чтобы убедиться в том, что в этом случае $F(x)DF(y)$, достаточно доказать, что $S_x = S_y$. Докажем это. Пусть $z \in S_y$, тогда xEz . По свойству симметричности предиката E из xEy выводим yEx . По свойству транзитивности предиката E из yEx и xEz выводим yEz . Отсюда следует, что $z \in S_y$. Итак, мы получили, что $S_x \subseteq S_y$. Предположим теперь, что $z \in S_y$. Тогда yEz . По свойству транзитивности из xEy и yEz выводим xEz . Отсюда следует, что $z \in S_x$. Итак, мы получили, что $S_y \subseteq S_x$. Вместе взятые, эти два включения дают равенство $S_y = S_x$. Рассмотрим оставшийся случай, при котором x и y таковы, что $\bar{x}\bar{E}\bar{y}$. Чтобы убедиться в том, что теперь $F(x)\bar{D}F(y)$, достаточно

доказать, что $S_y \neq S_x$. Докажем это. Из $x\bar{E}y$ следует $y \notin S_x$. По свойству рефлексивности предиката E имеем yEy , откуда выводим $y \notin S_x$. Следовательно, $S_x \neq S_y$. Мы доказали, что значения предикатов $E(x, y)$ и $D(F(x), F(y))$ совпадают при любых $x, y \in M$. Утверждение 2 доказано.

Из утверждений 1 и 2 непосредственно следует, что любые предикаты эквивалентности и только они могут быть представлены в виде (1) при подходящем выборе множества N и функции F . Таким образом, правая часть равенства (1) представляет собой общий вид предиката эквивалентности.

С математической точки зрения полученный результат тривиален, однако, он весьма важен для теории компараторной идентификации, поскольку указывает систему необходимых и достаточных признаков, с помощью которых всегда можно установить, допускает ли объект, реализующий предикат E , идентификацию компараторным методом. Если система, имеющая два входа x, y и один выход t , реализует предикат $t = E(x, y)$, и этот предикат удовлетворяет условиям рефлексивности, симметричности и транзитивности, то ее можно идентифицировать компараторным методом. Если же хотя бы одно из этих трех условий не выполняется, то компараторный метод для такого объекта неприменим. Получаемые результаты по компараторной идентификации могут быть применены к любым физическим объектам, удовлетворяющим только что перечисленным условиям.

В применении к зрительной системе человека элементы $x, y \in M$ интерпретируются как световые излучения, предъявляемые испытуемому для восприятия, M — это множество всех таких излучений. Элементы множества N $u = F(x)$ и $v = F(y)$ интерпретируются как цвета, возбуждаемые в сознании испытуемого излучениями x и y . Устанавливая совпадение или различие цветов u и v , испытуемый реализует предикат $D(u, v)$. Реагируя на излучения x и y , испытуемый реализует предикат $E(x, y) = D(F(x), F(y))$. Значение предиката $D(u, v) = 1$ соответствует реакции испытуемого, выражающей равенство цветов u и v . Значение предиката $D(u, v) = 0$ соответствует реакции испытуемого, выражающей несовпадение цветов u и v . Множество N представляет собой совокупность всех цветов, которые могут быть возбуждены в сознании испытуемого излучениями из множества M . Функцию F содержательно интерпретируем как преобразование светового излучения в цвет, реализуемое зрительной системой испытуемого.

Требование, что E есть предикат, означает: двичный ответ испытуемого существует и единственен для каждой пары световых излучений множества M . Рефлексивность предиката E означает, что одинаковым световым излучениям соответствуют одинаковые цвета. Симметричность предиката E означает, что изменение порядка предъявления

излучений испытуемому не влияет на его реакцию. Транзитивность предиката E означает: если для некоторого испытуемого излучения x, y и y, z одноцветны, то для того же испытуемого будут одноцветными также и излучения x, z .

Постулат о том, что испытуемый в опытах со сравнением цветов световых излучений реализует вполне определенный предикат E , не будет выполняться, если во время проведения опытов менять фон, на котором предъявляется каждое из излучений. Это вызвано эффектом цветовой индукции [2], заключающейся в том, что цвет тест-поля зависит от цвета окружающего его фона. Таким образом, стабильность фона является необходимым условием корректности компараторной идентификации цветового зрения человека. Но и при достижении стабильности фона реакцию испытуемого на пару световых излучений можно считать однозначной только с определенной степенью приближения. Дело в том, что ответы испытуемого при определенных условиях наблюдения носят вероятностный характер. Когда цвета излучений оказываются на границе равенства и неравенства, испытуемый ощущает неуверенность, которая выражается в нестабильности его ответа.

Сказанное проиллюстрируем следующим экспериментом. При помощи диска Максвелла [3] испытуемому предъявлялась на одном поле сравнения цветная поверхность (темно-красная), имеющая следующие трехцветные координаты МКО [4] $x_1 = 0,483$, $y_1 = 0,308$, $\rho_1 = 0,13$. На втором поле сравнения предъявлялась цветная поверхность, излучающая свет того же спектрального состава, но несколько иной интенсивности, отличающаяся в K раз ($x_2 = x_1$, $y_2 = y_1$, $\rho_2 = K\rho_1$). Отношение интенсивности излучения второго поля к интенсивности первого равно $K = \rho_2/\rho_1$. Определение трехцветных координат поверхностей определялось с помощью атласа цветов Рабкина [4]. Значение коэффициента K регулировалось углом раствора полей на диске Максвелла. Во время опытов диск Максвелла освещался зеркальной лампой типа ЗК220-500-2 мощностью 500 Вт с расстояния 2 м под углом 45° . В опытах использовались следующие численные значения коэффициента $K = 0,96; 0,97; 0,98; 0,99; 1,00; 1,01; 1,02; 1,03; 1,04$. При каждом значении K опыт повторялся 100 раз и для него вычислялось среднее арифметическое σ из ответов испытуемого. При этом ответу «цвета равны» приписывалось численное значение 1, ответу «цвета не равны» — численное значение 0. Испытуемый заранее не знал, при каком значении K проводится каждый опыт.

На рис. 1 изображена полученная диаграмма. Такого рода диаграммы называют пороговыми кривыми [5]. По оси абсцисс отложено значение коэффициента K , характеризующего интенсивность излучения второго поля; по оси ординат — значение параметра σ , характеризующего веро-

ятность формирования испытуемым ответа «цвета равны». Как видно из диаграммы, однозначность ответа испытуемого нарушается в двух зонах при $K = 0,96 - 0,99$ и $K = 1,01 - 1,04$. За пределами этих зон ответ испытуемого детерминирован и однозначно определяется предъявленными ему излучениями. Таким образом, постулат о существовании предиката E нарушается в довольно узких зонах изменения интенсивности излучения в пределах 3% ее величины. Следует учесть, что диск Максвелла является довольно грубым измерительным прибором. Проводя эти же опыты на оптическом субъективном колориметре (например, на колориметре Демкиной [6]), можно было бы зону неустойчивости ответа испытуемого существенно сузить.

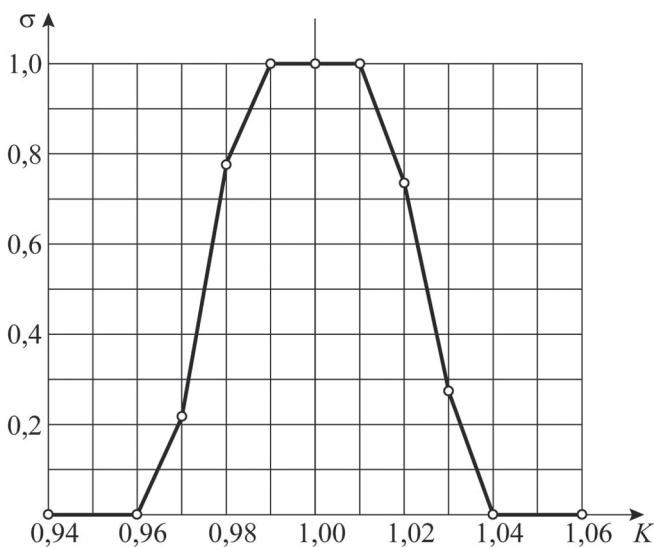


Рис. 1

С помощью опытов только что описанного типа можно оценить степень выполнения закона рефлексивности для цветового зрения человека. После того, как получена пороговая кривая, определяется ее ось симметрии (на рис. 1 — это вертикальная линия ($K = 1,00$)). Величина отклонения оси симметрии от положения $K = 1,00$ характеризует степень нарушения закона рефлексивности. Специально поставленные опыты показали, что это отклонение составляет величину порядка $\pm 0,1\%$ от величины K . Таким образом, можно утверждать, что для цветового зрения человека закон рефлексивности выполняется по крайней мере с той точностью, с которой осуществляется в эксперименте дозирование величины K . По аналогичной методике были выполнены опыты по проверке законов симметричности и транзитивности. Они показали, что и эти законы выполняются для цветового зрения человека практически точно. Таким образом, цветовое зрение человека можно идентифицировать компараторным методом, при этом точность такой идентификации лимитируется лишь величиной зоны неустойчивости ответа испытуемого.

2. Единственность представления объекта при его компараторной идентификации

В предыдущем параграфе было показано, что пара (N, F) , состоящая из множества N и функции $F: M \rightarrow N$, определяет единственный предикат эквивалентности $E(x, y) = D(F(x), F(y))$, заданный на множестве $M \times M$. Но справедливо ли обратное утверждение? Будет ли каждый предикат эквивалентности E единственным образом определять пару (N, F) ? Оказывается, нет. Существуют такие различные пары (N, F) и (N', F') , которые задают один и тот же предикат E . Формулируемое ниже утверждение указывает необходимое и достаточное условие, при котором две пары (N, F) и (N', F') определяют один и тот же предикат эквивалентности E .

Утверждение 3. Для того чтобы две пары (N, F) и (N', F') определяли один и тот же предикат эквивалентности E , заданный на декартовом квадрате множества M , необходимо и достаточно, чтобы существовала биекция T с областью определения N и областью значений N' такая, что для всех $x \in M$ $F'(x) = T(F(x))$.

Доказательство. Достаточность. Предположим, что существует биекция $T: N \rightarrow N'$ такая, что для всех $x \in M$ $F'(x) = T(F(x))$. Докажем, что в этом случае значения предикатов $E(x, y) = D(F(x), F(y))$ и $E'(x, y) = D'(F'(x), F'(y))$ совпадают при любых $x, y \in M$. Пусть x и y таковы, что $x E y$. Тогда $F(x) = D F(y)$, $T(F(x)) = D' T(F(y))$, $F'(x) = D' F'(y)$, $x E' y$. Если же $x \bar{E} y$, то $F(x) \bar{D} F(y)$, $T(F(x)) \bar{D}' T(F(y))$, $F'(x) \bar{D}' F'(y)$, $x \bar{E}' y$. Следовательно, для любых $x, y \in M$ $E(x, y) = E'(x, y)$.

Необходимость. Пусть $E(x, y) = E'(x, y)$ для любых $x, y \in M$. Докажем, что в этом случае найдется биекция $T: N \rightarrow N'$ такая, что $F'(x) = T(F(x))$ для любого $x \in M$. Рассмотрим отношение $T \subseteq N \times N'$, представляющее собой множество всех пар вида $(F(x), F'(x))$, где x — произвольный элемент множества M . Покажем, что T есть взаимно однозначная функция. Пусть x и y таковы, что $F(x) = D F(y)$, тогда $x E y$, $F'(x) = D' F'(y)$. Если же $F'(x) = D' F'(y)$, то $x E' y$, $x E y$, $F(x) = D F(y)$. Таким образом, отношение T есть взаимно однозначная функция, причем, $F'(x) = T(F(x))$ для любого $x \in M$. Областью определения функции T служит область значений функции F , то есть множество N . Областью значений функции T служит область значений функции F' , то есть множество N' . Отсюда, а также из определения понятия биекции [7] следует, что функция T — это биекция, отображающая все множество N на множество N' . Утверждение 3 доказано.

Из утверждения 3 непосредственно вытекает, что если при любых $x \in M$ предикат $E(x, y)$ можно представить в виде (1), то его также можно представить в виде:

$$E(x, y) = D(T(F(x)), T(F(y))), \quad (2)$$

где T — произвольно выбранная биекция. Из утверждения 3 также следует, что если предикат E представлен двумя различными способами $E(x, y) = D(F(x), F(y)) = D(F'(x), F'(y))$, то всегда найдется такая биекция T , которая связывает функции F и F' зависимостью $F'(x)T(F(x))$, справедливой при любом $x \in M$. Следовательно, получается, что нельзя указать единственно возможную характеристику F для предиката эквивалентности E .

Таким образом, если найдена некоторая функция F , математически описывающая объект идентификации, то на роль описания этого объекта с тем же правом может претендовать также целое семейство других функций. Поэтому при выборе функции F имеется большой произвол. В значительной мере произволен и выбор области значений функции F . Иными словами, выходные сигналы объекта компараторной идентификации также допускают различные варианты математического описания. Такая множественность представления объекта может навести на мысль о неполноте его описания методом компараторной идентификации и, следовательно, об ущербности этого метода по сравнению с классическими методами прямой идентификации. На самом деле степень полноты описания объекта при этих двух способах идентификации абсолютно одинакова. Дело в том, что при прямой идентификации описание объекта получается единственным лишь по той причине, что способ описания его выходных сигналов был выбран еще до начала идентификации. При компараторной же идентификации способ описания выходных сигналов выбирается в самом процессе идентификации, именно это и приводит к множественности описаний объекта.

Точно такая же множественность описаний объекта может возникнуть и при его прямой идентификации. Дело в том, что и при прямой идентификации выходные сигналы объекта могут описываться самыми различными способами. Так, компоненты вектора выходного сигнала можно нумеровать по-разному. Численные значения каждого из компонентов изменяются, если перейти к новой шкале при их измерении. При этом шкалу эту не обязательно брать линейной. Для каждого компонента можно брать свою собственную шкалу. Изменение же способа описания выходных сигналов объекта автоматически влечет за собой также и изменение описания самого объекта. После такого изменения описания линейный объект может даже превратиться в нелинейный. Таким образом, и при прямой идентификации математическое описание объекта может оказаться многовариантным. Это означает, что математическое описание объекта всегда получается лишь с точностью до произвольного биективного отображения. Требовать от

компараторной или некомпараторной идентификации, чтобы она давала единственно возможное описание объекта — это значит считать, что имеется единственная истинная система обозначений его выходных сигналов, что, конечно, неверно.

Нижеследующее утверждение доказывает, что методом компараторной идентификации объект описывается с точностью до изоморфизма. Содержательно это означает, что компараторная идентификация (так же, как и прямая) дает с точностью до обозначений единственное описание объекта. Рассмотрим два предиката эквивалентности $E(x, y)$ и $E'(x', y')$. Первый из них задан на декартовом квадрате множества M , второй — на декартовом квадрате множества M' . Выразим предикаты E и E' в виде:

$$E(x, y) = D(F(x), F(y)), \quad (3)$$

$$E'(x', y') = D'(F'(x), F'(y)). \quad (4)$$

Согласно утверждению 2 это всегда можно сделать. Здесь D — предикат равенства, заданный на декартовом квадрате множества N ; D' — предикат равенства, заданный на декартовом квадрате множества N' . Символами F и F' обозначены характеристические функции предикатов E и E' . Первая из них определена на множестве M и принимает значения на множестве N , вторая определена на M' и принимает значения на множестве N' .

Утверждение 4. Если модели $\langle M, E \rangle$ и $\langle M', E' \rangle$ изоморфны, то также изоморфны модели $\langle N, D \rangle$ и $\langle N', D' \rangle$.

Доказательство. Изоморфность моделей $\langle M, E \rangle$ и $\langle M', E' \rangle$ означает [8], что существует биекция $G: M \rightarrow M'$ такая, что

$$E(x, y) = E'(G(x), G(y)) \quad (5)$$

при любых $x, y \in M$. Существование биективной функции G означает, что множества M и M' равномощны. Рассмотрим отношение H , заданное на декартовом произведении $N \times N'$ и образованное всеми парами вида $(F(x), F'(G(x)))$. Покажем, что отношение H функционально. Возьмем какие-нибудь элементы $x, y \in M$ и предположим, что $F(x)DF(y)$. Тогда в силу (3) справедливо xEy . Из (5) следует, что $G(x)E'G(y)$, откуда согласно (4) получаем $F'G(x), (F(x), F'(G(x)))$. Итак, отношение H есть функция.

Покажем, что функция H взаимно однозначна. Пусть $x, y \in M$ таковы, что $F'(G(x))D'F'G(y)$. Тогда в силу (4) $G(x)E'G(y)$. Отсюда согласно (5) выводим xEy . Последнее соотношение в силу (3) влечет $F(x)DF(y)$. Итак, H — взаимно однозначная функция. Область значений функции F совпадает с множеством N , поэтому функция H определена на всем множестве N . Поскольку область значений функции F' совпадает с множеством N' ,

то областью значений функции H служит все множество N' . Итак, отношение H есть биекция, заданная на множестве N со значениями на множестве N' . Она может быть представлена равенством

$$F'(G(x)) = H(F(x)), \quad (6)$$

справедливым для всех $x \in M$. Из биективности отношения H следует, что множества N и N' равномощны.

Доказываем изоморфность предикатов D и D' . Для любых $u, v \in N$ из uDv следует $G(u)D'G(v)$. Точно так же uDv из $G(u)D'G(v)$. Итак,

$$D(u, v) = D'(G(u), G(v)) \quad (7)$$

при любых $u, v \in N$. Утверждение 4 доказано.

Пусть $u = F(x)$ – математическое описание объекта, заданного предикатом $E(x, y)$; $u' = F'(x')$ – описание того же объекта, заданного предикатом $E'(x', y')$. Функция F отображает множество M на множество N , а функция F' отображает множество M' на множество N' . Если множество M отличается от множества M' , а множество N — от множества N' , то это может означать лишь то, что входные сигналы x, x' и выходные сигналы u, u' объекта описаны в различных системах обозначений. Поэтому должны существовать биекции G и H такие, что $x' = G(x)$ и $u' = H(u)$. Они задают переход от одной системы обозначений сигналов объекта к другой. Пользуясь только что записанными соотношениями, можем равенство $u' = F'(x')$ переписать в виде $H(u) = F'(G(x))$. Учитывая (6), имеем $H(u) = H(F(x))$. Так как отображение H биективно, то $u = F(x)$. Мы превратили описание F объекта в описание F' . Аналогичным способом можно превратить описание F' объекта в описание F . Полученный результат означает, что объект компараторной идентификации может иметь различные описания F и F' только за счет того, что его входные и выходные сигналы представлены в различных системах обозначений. Если потребовать, чтобы входные и выходные сигналы объекта всегда описывались в одной и той же системе отсчета (то есть каким-то стандартным способом), то при компараторной идентификации, так же как и при прямой, описание объекта будет единственным.

3. Слабый и сильный изоморфизмы предикатов

Ввиду важности вопроса об изоморфности характеристических функций эквивалентностей в следующих двух параграфах дается его более глубокая проработка. Предикаты P и P' на $A \times B$ и $A' \times B'$ называются слабо изоморфными (или просто изоморфными), если существуют биекции $\varphi: A \rightarrow A'$ и $\psi: B \rightarrow B'$ такие, что для всех $x \in A$ и $y \in B$ выполняется равенство

$$P(x, y) = P'(\varphi(x), \psi(y)). \quad (8)$$

Будем говорить также, что предикат P (φ, ψ) изоморфен предикату P' . Биекции φ и ψ , удовлетворяющие условию (8), называются левым и правым изоморфизмами предикатов P и P' .

Предикаты $P(x, y)$ и $P'(x', y')$ на $A \times B$ и $A' \times B'$ называются сильно изоморфными, если существует биекция $\varphi: A \cup B \rightarrow A' \cup B'$ такая, что для всех $x \in A$ и $y \in B$ выполняется равенство

$$P(x, y) = P'(\varphi(x), \varphi(y)). \quad (9)$$

Будем говорить также, что предикат P φ -изоморфен предикату P' . Биекция φ , удовлетворяющая условию (9), называется изоморфизмом предикатов P и P' .

Понятия слабого и сильного изоморфизмов предикатов играют в теории компараторной идентификации важную роль. Дело в том, что выбор обозначений для сигналов идентифицируемой системы находится всецело во власти исследователя и определяется принятой им системой единиц. Если два исследователя, изучая поведение одной и той же системы, используют разные обозначения для ее входных сигналов, то они получают для нее различные предикаты. Если все входные сигналы одной и той же изучаемой системы каждым исследователем записываются в единой (но своей) системе единиц, то получаемые ими предикаты будут сильно изоморфными, если же в разных, то предикаты будут слабо изоморфными. В этом случае будем говорить, что изучаемые системы идентифицированы с точностью до обозначений (общих или отдельных). В случае сильного изоморфизма предикатов будем говорить, что индуцируемые системы совпадают с точностью до обозначений в единой системе единиц. В случае слабого изоморфизма предикатов будем говорить, что идентифицируемые системы совпадают с точностью до обозначений в разных системах единиц.

Утверждение 5. Изоморфизм φ предикатов P и P' , определенных соответственно на $A \times B$ и $A' \times B'$, биективно отображает $A \cap B$ на $A' \cap B'$, $A \setminus B$ на $A' \setminus B'$ и $B \setminus A$ на $B' \setminus A'$.

Доказательство. Доказываем, что φ отображает $A \cap B$ на $A' \cap B'$. Поскольку $A \cap B \subseteq A \cup B$, то, по определению сильно изоморфных предикатов P и P' для всех $x \in A \cap B$, $y \in B$ и $z \in A$ имеют место равенства $P(x, y) = P'(\varphi(x), \varphi(y))$ и $P(z, x) = P'(\varphi(z), \varphi(x))$. Следовательно, $\varphi(x) \in A'$ и $\varphi(x) \in B'$, то есть $\varphi(x) \in A' \cap B'$. Так как φ — биекция, то для всех $x \in A' \cap B'$, $y \in B'$, $z \in A'$ имеют место равенства $P'(x, y) = P(\varphi^{-1}(x'), \varphi^{-1}(y'))$, $P'(z, x) = P(\varphi^{-1}(z'), \varphi^{-1}(x'))$. Следовательно, $\varphi^{-1}(x') \in A \cap B$. Таким образом, ограничение функции φ на $A \cap B$ имеет значения в множестве $A' \cap B'$ и сюръективно. Кроме того, φ инъективна. Поэтому φ биективно отображает множество $A \cap B$ на множество $A' \cap B'$.

Доказываем, что φ отображает $A \setminus B$ на $A' \setminus B'$. Пусть $x \in A \setminus B$. Предположим, что $\varphi(x) \in B'$. Тогда для всех $z' \in A'$ имеет место равенство $P'(z', \varphi(x)) = P(\varphi^{-1}(z'), x)$. Следовательно, $x \in B$, что противоречит условию. Таким образом, имеем: $\varphi(x) \in A' \setminus B'$. Аналогично выводим, что для всех $x \in A' \setminus B'$ $\varphi^{-1}(x) \in A \setminus B$. Значит, φ сюръективно отображает $A \setminus B$ на $A' \setminus B'$. Будучи инъективной по определению, φ биективно отображает $A \setminus B$ на $A' \setminus B'$. Доказательство того, что φ биективно отображает $A' \setminus B'$ на $A \setminus B$ аналогично предыдущему. Утверждение 5 доказано.

Содержательно утверждение 5 означает, что если сигналы x и y системы $P(x, y)$ имеют имена в единой системе обозначений, то при переобозначении этих имен необходимо имена из области $A \cap B$ превратить в имена из области $A' \cap B'$, имена из $A \setminus B$ — в имена из $A' \setminus B'$ и имена из $B \setminus A$ — в имена из $B' \setminus A'$. Если сделать иначе, то предикат P' , получаемый из предиката P в результате переобозначения его входных сигналов, может оказаться неизоморфным предикату P , что недопустимо.

Утверждение 6. Если $A \cap B = \emptyset$ и $A' \cap B' = \emptyset$, то слабо изоморфные предикаты P и P' , определенные на $A \times B$ и $A' \times B'$, будут также и сильно изоморфными.

Доказательство. Если предикаты P и P' слабо изоморфны, то существуют биекции $\varphi_1: A \rightarrow A'$ и $\varphi_2: B \rightarrow B'$ такие, что для всех $x \in A$ и $y \in B$ выполняется равенство $P(x, y) = P'(\varphi_1(x), \varphi_2(y))$. Положим

$$\varphi(x) = \begin{cases} \varphi_1(x), & \text{если } x \in A; \\ \varphi_2(x), & \text{если } x \in B. \end{cases}$$

Очевидно, что φ биективно отображает $A \cup B$ на $A' \cup B'$, и для всех $x \in A$ и $y \in B$ выполняется равенство $P(x, y) = P'(\varphi(x), \varphi(y))$. Это означает, что предикаты P и P' сильно изоморфны. Утверждение 6 доказано.

Содержательно утверждение 6 означает: два различных описания одной и той же идентифицируемой системы $P(x, y)$, входные сигналы которой определены на непересекающихся областях, всегда совпадают с точностью до сильного изоморфизма, то есть совпадают с точностью до обозначений входных сигналов x и y системы P , описываемых в единой системе обозначений.

Утверждение 7. Пусть P и P' — предикаты, определенные на $A \times B$ и $A' \times B'$; $A \cap B$ не пусто; $\varphi_1: A \rightarrow A'$ и $\varphi_2: B \rightarrow B'$ — левый и правый изоморфизмы предикатов P и P' . Если образом множества $A \cap B$ при отображениях φ_1 и φ_2 является множество $A' \cap B'$ и на множестве $A \cap B$ значения биекций φ_1 и φ_2 совпадают, то предикаты P и P' сильно изоморфны.

Доказательство. Так как предикаты P и P' слабо изоморфны, то для всех $x \in A$ и $y \in B$ $P(x, y) = P'(\varphi_1(x), \varphi_2(y))$. Положим

$$\varphi(x) = \begin{cases} \varphi_1(x), & \text{если } x \in A, \\ \varphi_2(x), & \text{если } x \in B \setminus A. \end{cases}$$

Поскольку $\varphi_1(x)$ биективно отображает множество A на множество A' , а $\varphi_2(x)$ — множество $B \setminus A$ на $B' \setminus A'$, то $\varphi(x)$ биективно отображает $A \cup B$ на $A' \cup B'$. Докажем, что для всех $x \in A$ и $y \in B$ $P(x, y) = P'(\varphi_1(x), \varphi_2(y))$. Пусть $x \in A$ и $y \in B$. Возможны два случая: $y \in A \cap B$ и $y \in B \setminus A$. Пусть $y \in A \cap B$. Тогда, поскольку по условию утверждения 7 значения биекций φ_1 и φ_2 совпадают, то $\varphi_2(x) = \varphi_1(x) = \varphi(x)$. Если же $y \in B \setminus A$, то $\varphi_2(y) = \varphi(y)$. Следовательно, при любых $x \in A$ и $y \in B$ имеем: $P(x, y) = P'(\varphi_1(x), \varphi_2(y)) = P'(\varphi(x), \varphi(y))$. Это означает, что предикаты P и P' сильно изоморфны. Утверждение 7 доказано.

Содержательно утверждение 7 означает, что если для системы $P(x, y)$ найдется сигнал a , который можно подать как на вход x , так и на вход y (то есть $x = y = a$), и если все такие сигналы переобозначаются биекциями $\varphi_1(x) = x'$ и $\varphi_2(y) = y'$ по-одинаковому, то в результате получаем предикат $P'(x, y)$ вне зависимости от способа переобозначения остальных сигналов (при условии, что для них не используются имена сигналов a).

Утверждение 8. Если предикаты эквивалентности E и E' на $A \times A$ и $A' \times A'$ слабо изоморфны, то они также и сильно изоморфны.

Доказательство. Поскольку предикаты E и E' слабо изоморфны, то существуют биекции $\varphi_1: A \rightarrow A'$ и $\varphi_2: A \rightarrow A'$ такие, что для любых $x, y \in A$ имеет место равенство $E(x, y) = E'(\varphi_1(x), \varphi_2(y))$ (а). Подставляя в (а) $y = x$, по свойству рефлексивности предиката E получаем $E'(\varphi_1(x), \varphi_2(x)) = 1$ (б). Далее, используя свойство симметричности предиката E' , выводим $E'(\varphi_2(x), \varphi_1(x)) = 1$ (в). Пусть $x, y \in A$ таковы, что $E(x, y) = 1$. Тогда из (а) следует $E'(\varphi_1(x), \varphi_2(y)) = 1$ (г). По свойству транзитивности предиката E' из (в) и (г) выводим $E'(\varphi_2(y), \varphi_2(x)) = 1$ (д). Если же $x, y \in A$ таковы, что $E(x, y) = 0$, то из (а) следует $E'(\varphi_1(x), \varphi_2(y)) = 0$ (е). Если бы выполнялось равенство (д), то, в силу транзитивности предиката E' из (б) и (д) следовало бы $E'(\varphi_1(x), \varphi_2(y)) = 1$, что противоречит (е). Следовательно, (д) не выполняется, то есть $E'(\varphi_2(x), \varphi_2(y)) = 0$. Итак, для всех $x, y \in A$ имеет место равенство $E(x, y) = E'(\varphi_2(x), \varphi_2(y))$. Это означает, что предикаты E и E' сильно изоморфны. Утверждение 8 доказано.

Содержательно утверждение 8 означает, что сигналы x и y предиката эквивалентности $E(x, y)$ нельзя описывать в разных системах обозначений, а только в одной и той же. Желая описать систему $E(x, y)$ моделью эквивалентности, исследователь должен выражать ее выходные сигналы x и y в одной системе обозначений.

Утверждение 9. Пусть E — предикат эквивалентности на $A \times A$ и $f: A \rightarrow B$ — его характе-

ристическая функция. Тогда эквивалентность E изоморфна равенству D на $B \times B$ в том и только в том случае, когда f инъективна.

Доказательство. Необходимость. Пусть предикаты E и D изоморфны. Тогда существует биекция $\varphi: A \rightarrow B$ такая, что для всех $x, y \in A$ $E(x, y) = D(\varphi(x), \varphi(y))$. Проверяем инъективность функции f . Пусть $x, y \in A$ таковы, что $f(x) = f(y)$. Тогда $E(x, y) = 1$, следовательно, $\varphi(x) = \varphi(y)$. Поскольку φ биективна, то $x = y$.

Достаточность. Пусть функция f инъективна. Поскольку f сюръективна, то она биективна. По определению характеристической функции f предиката E для всех $x, y \in A$ $E(x, y) = D(f(x), f(y))$. Следовательно, предикаты E и D изоморфны. Утверждение 9 доказано.

Утверждение 9 дает ответ на вопрос, в каких случаях при восприятии предметов человек получает о них всю информацию, а в каких — не всю. Информация не теряется в тех случаях, когда органы чувств человека каждому предмету ставят в соответствие свой субъективный образ (неважно какой). Если же образов меньше, чем воспринимаемых предметов, то часть информации о предметах теряется. Глаз человека при восприятии световых излучений теряет часть информации о них. Это доказывается тем, что существуют такие различные световые излучения, которые воспринимаются глазом в виде одного и того же цвета. Например существует такая смесь красного и зеленого монохроматических излучений, которая неотличима по цвету от желтого монохроматического излучения.

4. Изоморфизм характеристических функций эквивалентностей

Любую функцию $y = f(x)$, отображающую множество A в множество B , можно задать, указывая соответствующий ей предикат $F(x, y)$, определенный на $A \times B$. Для этого при любых $x \in A$, $y \in B$ полагаем $F(x, y) = 1$, если $f(x) = y$, и $F(x, y) = 0$, если $f(x) \neq y$. Пусть $f: A \rightarrow B$ и $f': A' \rightarrow B'$ — функции, а F и F' на $A \times B$ и $A' \times B'$ — соответствующие им предикаты. Будем говорить, что функция f (φ, ψ) -изоморфна функции f' , если предикат F (φ, ψ) -изоморфен предикату F' . Если биекции φ и ψ существуют, то будем говорить, что функции f и f' слабо изоморфны. Если, кроме того, $\varphi = \psi$, то будем говорить, что они сильно изоморфны.

Пусть имеется преобразователь сигналов, реализующий функцию $y = f(x)$, которая отображает множество A на множество B . Переименовывая его входные и выходные сигналы x и y с помощью биекций $\varphi: A \rightarrow B$ и $\psi: A \rightarrow B$, получаем $x' = \varphi(x)$, $y' = \psi(y)$. В результате тот же преобразователь сигналов опишется уже другой функцией $y' = f'(x')$, отображающей множество A' на множество B' . Обозначая через φ^{-1} функцию, обратную биекции φ выражаем функцию f' через функцию f :

$$f(x') = \psi(f(\varphi^{-1}(x^{-1}))). \quad (10)$$

Аналогично выражается функция f через функцию f' :

$$f(x') = \psi^{-1}(f'(\varphi(x))), \quad (11)$$

где ψ^{-1} — функция, обратная биекции ψ .

Утверждение 10. Пусть E и E' — эквивалентности на $A \times A$ и $A' \times A'$; D и D' — предикаты равенства на $B \times B$ и $B' \times B'$. Если предикат E φ -изоморфен предикату E' , то существует такая биекция $\psi: A \rightarrow B$, что функция $f(\varphi, \psi)$ -изоморфна функции f' , а предикат D ψ -изоморфен предикату D' .

Доказательство. Поскольку предикаты E и E' φ -изоморфны, то существует биекция $\varphi: A \rightarrow A'$ такая, что для любых $x, y \in A$ $E(x, y) = E'(\varphi(x), \varphi(y))$. Рассмотрим отношение ψ , заданное на $B \times B'$ и образованное всеми парами вида $(f(x), f'(\varphi(x)))$, где $x \in A$. Пусть $x, y \in A$ таковы, что $f(x) = f(y)$. Тогда $E(x, y) = 1$, $E'(\varphi(x), \varphi(y)) = 1$, $f'(\varphi(x)) = f'(\varphi(y))$. Таким образом, отношение ψ удовлетворяет условию однозначности. Область значений функции $f(x)$ совпадает с множеством B , поэтому отношение ψ всюду определено слева. Следовательно, отношение ψ есть функция. Область значений функции $f'(\varphi(x))$ совпадает с множеством B' , поэтому функция ψ сюръективна. Пусть $x, y \in A$ таковы, что $f'(\varphi(x)) = f'(\varphi(y))$, тогда $E'(\varphi(x), \varphi(y)) = 1$, $E(x, y) = 1$, $f(x) = f(y)$. Следовательно, функция ψ инъективна. Итак, отношение ψ есть биекция, отображающая B на B' . Из определения отношения ψ непосредственно следует, что для всех $x \in A$ $f'(\varphi(x)) = \psi(f(x))$. Согласно (11) это означает, что функции f и f' (φ, ψ) -изоморфны. Доказываем изоморфность предикатов D и D' . В силу биективности ψ для любых $u, v \in B$ из $D(u, v) = 1$ следует $u = v$, $\psi(u) = \psi(v)$, $D'(\psi(u), \psi(v)) = 1$; из $D(u, v) = 0$ следует $u \neq v$, $\psi(u) \neq \psi(v)$, $D'(\psi(u), \psi(v)) = 0$. Итак, для любых $u, v \in B$ $D(u, v) = D'(\psi(u), \psi(v))$, а это означает, что предикаты D и D' ψ -изоморфны. Утверждение 10 доказано.

Утверждение 11. Если функция f (φ, ψ) -изоморфна функции f' , то предикат E (φ, ψ) -изоморфен предикату E' , а предикат D ψ -изоморфен предикату D' .

Доказательство. Предположим, что функция f (φ, ψ) -изоморфна функции f' . Тогда, согласно (9), $f'(x) = \psi^{-1}(f(\varphi(x)))$ для любого $x \in A$. Пусть $x, y \in A$ таковы, что $E(x, y) = 1$. Тогда $D(f(x), f(y)) = 1$, $f(x) = f(y)$, $\psi^{-1}(f(\varphi(x))) = \psi^{-1}(f(\varphi(y)))$, $D'(f(x), f'(\varphi(y))) = 1$. $E'(\varphi(x), \varphi(y)) = 1$. Если же $x, y \in A$ таковы, что $E(x, y) = 0$, то $D(f(x), f(y)) = 0$, $f(x) \neq f(y)$, $\psi^{-1}(f(\varphi(x))) \neq \psi^{-1}(f(\varphi(y)))$, $f'(\varphi(x)) \neq f'(\varphi(y))$, $D'(f'(\varphi(x)), f'(\varphi(y))) = 0$, $E'(\varphi(x), \varphi(y)) = 0$. Итак,

при любых $x, y \in A$ $E(x, y) = E'(\varphi(x), \varphi(y))$. Мы доказали, что предикат E φ -изоморфен предикату E' . Пусть $u, v \in B$ таковы, что $D(u, v) = 1$. Тогда $u = v$, $\psi(u) = \psi(v)$, $D'(\psi(u), \psi(v)) = 1$. Если же $u, v \in B$ таковы, что $D(u, v) = 0$, тогда $u \neq v$, $\psi(u) \neq \psi(v)$, $D'(\psi(u), \psi(v)) = 0$. Итак, при любых $u, v \in B$ $D(u, v) = D'(\psi(u), \psi(v))$. Мы доказали, что предикат D ψ -изоморфен предикату D' . Утверждение 11 доказано.

Из утверждений 10 и 11 непосредственно следует.

Утверждение 12. Для того чтобы эквивалентность E была φ -изоморфна эквивалентности E' , необходимо и достаточно, чтобы функция f была (φ, ψ) -изоморфна функции f' .

Содержательно утверждения 10 – 12 означают, что поведение $E(x, y) = D(f(x), f(y))$ идентифицируемой системы E полностью (то есть с точностью до обозначений) определяется действием идентифицируемого объекта f и наоборот. Кроме того, действие нулевого прибора $D(u, v)$ полностью определяется как поведением системы E , так и действием объекта f . Все сказанное свидетельствует о том, что компараторный метод является эффективным средством идентификации объекта f , внутреннего состояния $u = f(x)$ системы E и нулевого прибора $D(u, v)$. Например при изучении механизма цветового зрения человека исследователь объективно наблюдает двоичные ответы испытуемого, сравнивающего цвета световых излучений, и только из этих ответов получает исчерпывающие описания: 1) субъективного цвета; 2) преобразования физических световых излучений в психологические цвета; 3) способа субъективного анализа цветов. Верно и обратное утверждение: если исследователь уже нашел вид преобразования $u = f(x)$ светового излучения x в цвет u , то он может заранее вычислить ответ $t = E(x, y)$ испытуемого, сравнивающего цвета произвольных световых излучений x и y .

Имеется некоторое неравноправие между внешним поведением E испытуемого и соответствующим ему внутренним информационным процессом f , поскольку сильной изоморфности предикатов E и E' соответствует слабая изоморфность функций f и f' . Следующее утверждение устанавливает условие, при котором предикат E и функция f становятся в этом смысле равноправными.

Утверждение 13. Для того чтобы φ -изоморфность любых эквивалентностей E и E' на $A \times A$ и $A' \times A'$ была равносильна φ -изоморфности их характеристических функций $f: A \rightarrow B$ $f': A' \rightarrow B'$, необходимо и достаточно, чтобы множества A и B , A' и B' не пересекались.

Доказательство. Необходимость. Для доказательства достаточно привести пример множеств A , B , и B' таких, что $A \cap B = \emptyset$, а также φ -изоморфных, эквивалентностей E и E' , заданных на $A \times A$, характеристические функции которых не являются φ -изоморфными. Положим

$A = \{a_1, a_2, a_3\}$, $B = \{a_1, a_2, a_3\}$, $B' = \{a_1, a_3\}$. Эквивалентность $E = E'$ определим разбиением множества $M: \{\{a_1, a_2\}, \{a_3\}\}$. Функцию $f: A \rightarrow B$ определим следующим образом: $f(a_1) = f(a_2) = a_1$, $f(a_3) = a_2$. Функцию $f': A' \rightarrow B'$ определим так: $f'(a_1) = f'(a_2) = a_1$, $f'(a_3) = a_2$. Очевидно, E и E' , будучи равными, сильно изоморфны. Тем не менее, сильного изоморфизма для функций f и f' не существует, так как в противном случае такой изоморфизм должен отображать множество $\{a_1, a_2\}$ на себя и в то же время множество B на B' , что невозможно.

Достаточность непосредственно следует из утверждений 6 и 10. Утверждение 13 доказано.

Утверждение 13 содержит в себе рекомендацию инженеру, создающему искусственные анализаторы предметов для технических систем. Если требуется, чтобы предметы и их образы можно было измерять в одной и той же системе физических единиц и при этом всегда получать действие системы в виде предиката эквивалентности, то нужно обеспечить, чтобы множество всех анализируемых предметов и множество их образов не пересекались. Например при создании системы искусственного цветового зрения нужно сделать так, чтобы цвета как физические объекты были представлены не световыми излучениями, а какими-то иными физическими процессами, например, магнитными полями.

5. Обобщение метода компараторной идентификации

Метод компараторной идентификации, описанный во введении, допускает обобщение. Предикат $P(x, y)$, определенный на $A_1 \times A_2$, называется вполне определенным слева на $A_1 \times A_2$, если для каждого $x \in A_1$ существует $x \in A_2$ такой, что $P(x, y) = 1$. Предикат $P(x, y)$, определенный на $A_1 \times A_2$, называется вполне определенным справа на $A_1 \times A_2$, если для каждого $x \in A_2$ существует $x \in A_1$ такой, что $P(x, y) = 1$. Предикат $P(x, y)$ называется вполне определенным на $A_1 \times A_2$, если он вполне определен слева и справа на $A_1 \times A_2$. Пусть предикат $P(x, y)$ не вполне определен на $A_1 \times A_2$. Тогда, сужая области определения переменных x и y предиката A_1 и A_2 до множеств A'_1 и A'_2 соответствующих предикатам

$$A'_1(x) = \exists y \in A_2 P(x, y), \quad (12)$$

$$A'_2(y) = \exists x \in A'_1 P(x, y), \quad (13)$$

всегда можно получить предикат P , вполне определенный на $A'_1 \times A'_2$.

Приведенные определения можно обобщить на случай произвольного числа n переменных $x_1, x_2, \dots, x_n$ ($n \leq m$). Предикат $P(x_1, x_2, \dots, x_n)$, определенный на $A_1 \times A_2 \times \dots \times A_n$, называется вполне определенным на $A_1 \times A_2 \times \dots \times A_n$ по переменной x_i , если для каждого $x_i \in A_i$ существуют $x_1 \in A_1$, $x_2 \in A_2, \dots, x_{i-1} \in A_{i-1}, x_{i+1} \in A_{i+1}, \dots, x_n \in A_n$ такие, что $P(x_1, x_2, \dots, x_{i-1}, x_{i+1}, x_n) = 1$. Предикат $P(x_1, x_2, \dots, x_n)$ называется вполне определенным

на $A_1 \times A_2 \times \dots \times A_n$, если он вполне определен на $A_1 \times A_2 \times \dots \times A_n$ по всем переменным $x_1, x_2, \dots, x_n$. Когда нет особых причин поступить иначе, исследователь всегда стремится так ограничить области определения входных сигналов идентифицируемой системы, чтобы ее поведение описывалось вполне определенными предикатами. Нетрудно видеть, что любая эквивалентность на $A \times A$ вполне определена на $A \times A$.

Пусть E и E' — эквивалентности на $A \times A$. Будем говорить, что эквивалентность E вложена в эквивалентность E' , и писать $E \leq E'$, если для любых $x, y \in A$ из $E(x, y) = 1$ следует $E'(x, y) = 1$. Если $E \leq E'$, и $E \neq E'$, то будем писать $E < E'$ и говорить, что эквивалентность E строго вложена в эквивалентность E' . Если $E < E'$, то будем говорить, что разбиение R , соответствующее эквивалентности E , мельче разбиения R' , соответствующего эквивалентности E' . Будем говорить также, что разбиение R' крупнее разбиения R . Если $E \leq E'$, то будем говорить, что разбиение R мельче или равно разбиению R' . Если $E \leq E'$, то разбиение R , соответствующее эквивалентности E , называется подразбиением разбиения R' , соответствующего эквивалентности E' . Нетрудно видеть, что отношение вложения, определенное на множестве предикатов эквивалентности, рефлексивно, транзитивно и антисимметрично, то есть оно есть отношение порядка.

Теорема 1. Пусть P — предикат на $A_1 \times A_2$. Предикаты E_L и E_R на $A_1 \times A_2$ и $A_1 \times A_2$, определяемые тождествами

$$E_L(x_1, x_2) = \forall y \in A_2 (P(x_1, y) \sim P(x_2, y)), \quad (14)$$

$$E_R(y_1, y_2) = \forall x \in A_1 (P(x, y_1) \sim P(x, y_2)) \quad (15)$$

называются сопровождающими эквивалентностями (левой и правой) предиката P . Нетрудно видеть, что предикаты E_L и E_R на $A_1 \times A_1$ и E_R на $A_2 \times A_2$ — его сопровождающие эквивалентности; E_L на $A_1 \times A_1$ и E_2 на $A_2 \times A_2$ — эквивалентности; удовлетворяющие условиям $E_1 \leq E_L, E_2 \leq E_R$; $F_1: A_1 \rightarrow B_1$ и $F_2: A_2 \rightarrow B_2$ — характеристические функции эквивалентностей E_1 и E_2 . Тогда найдется единственный предикат L на $B_1 \times B_2$ такой, что для любых $x \in A_1$ и $y \in A_2$:

$$P(x, y) = L(f_1(x), f_2(y)). \quad (16)$$

Доказательство. Покажем, что предикат L с требуемыми свойствами существует. Для любых $u \in B_1, v \in B_2$ положим $L(u, v) = 1$, если и только если найдутся $x \in A_1, y \in A_2$ такие, что $F_1(x) = u, F_2(y) = v, P(x, y) = 1$. Докажем справедливость (16). Пусть $x \in A_1$ и $y \in A_2$ таковы, что $P(x, y) = 1$. Тогда из способа построения предиката L непосредственно следует $L(F_1(x), F_2(y)) = 1$. Пусть теперь $x \in A_1$ и $y \in A_2$ таковы, что $L(F_1(x), F_2(y)) = 1$. Тогда найдутся $x_1 \in A_1$ и $y_2 \in A_2$ такие, что $P(x_1, y_2) = 1, F_1(x_1) = F_1(x), F_2(y_2) = F_2(y)$, откуда $E_1(x, x_1) = E_2(y, y_2) = 1$. Так

как $E_1 \leq E_L, E_2 \leq E_R$, то $E_L(x, x_1)E_R(y, y_2) = 1$. Из определения сопровождающих эквивалентностей предиката P окончательно получаем $P(x, y) = 1$. Мы доказали справедливость (16) для построенного нами предиката L . Покажем теперь единственность L . Предположим, что, кроме (16), справедливо еще представление $P(x, y) = L'(F_1(x), F_2(y))$, где L' — некоторый предикат, заданный на $B_1 \times B_2$. Тогда для любых $x \in A_1$ и $y \in A_2$ $L(F_1(x), F_2(y)) = L'(F_1(x), F_2(y))$. Следовательно, для любых $u \in B_1, v \in B_2$ $L(u, v) = L'(u, v)$, что доказывает единственность представления P в виде (16). Теорема доказана.

Выражение (16) называется общим видом бинарного предиката P . Легко видеть, что любой бинарный предикат P можно представить в виде (16). Сюръекции E_1 и E_2 называются характеристическими функциями (левой и правой) предиката P . Предикат L называется образом предиката P при эквивалентностях E_1 и E_2 . Согласно теореме 1, любую систему преобразования сигналов (см. рис 2, а), реализующую бинарный предикат $P(x, y) = t$, можно представить в виде трехблочной схемы, изображенной на рис. 2, б. Сравнивая эту схему со схемой, представленной на рис. 1, а, видим, что первая является обобщением последней. Теперь вместо одной функции F используются две функции F_1 и F_2 , а вместо предиката равенства D — произвольный бинарный предикат L . Теорему 1 можно обобщить на случай произвольного n -арного предиката $P(x_1, x_2, \dots, x_n)$. Она гласит, что любой предикат можно представить в виде:

$$P(x_1, x_2, \dots, x_n) = L(F_1(x_1), F_2(x_2), \dots, F_n(x_n)). \quad (17)$$

Следовательно, любую систему сигналов, реализующую предикат $P(x_1, x_2, \dots, x_n) = t$ (см. рис. 3, а), можно представить в виде схемы, изображенной на рис. 3, б.

Следующая теорема доказывает, что методом компараторной идентификации (роль компаратора на схеме рис. 3, б выполняет блок L) можно исчерпывающим образом (то есть с точностью до обозначений) идентифицировать одновременно n объектов $F_1, F_2, \dots, F_n$. Это значит, что в наиболее общем случае метод компараторной идентификации по степени глубины анализа объектов не уступает методу прямой идентификации.

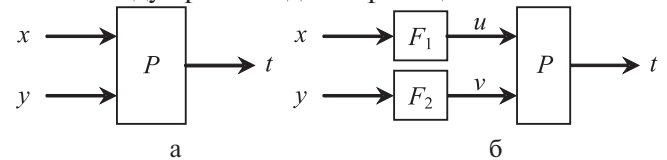


Рис. 2

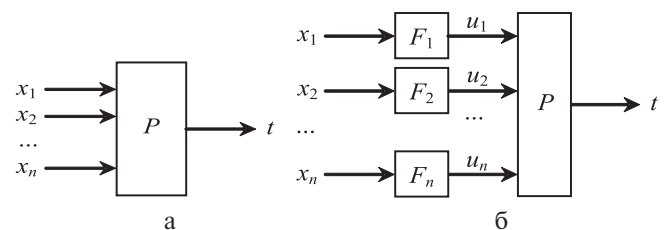


Рис. 3

Теорема 2. Пусть предикат $P(x_1, x_2, \dots, x_n)$ задан на множестве $A_1 \times A_2 \times \dots \times A_n$, предикат $P'(x'_1, x'_2, \dots, x'_n)$ задан на множестве $A'_1 \times A'_2 \times \dots \times A'_n$, предикаты $L(u_1, u_2, \dots, u_n)$ и $L'(u'_1, u'_2, \dots, u'_n)$ — на множествах $B_1 \times B_2 \times \dots \times B_n$, $B'_1 \times B'_2 \times \dots \times B'_n$ соответственно, причем L есть образ предиката при эквивалентностях $E_1 \times E_2 \times \dots \times E_n$, L' — образ P при эквивалентностях $E'_1 \times E'_2 \times \dots \times E'_n$. Пусть предикаты P и P' ($\varphi_1, \varphi_2, \dots, \varphi_n$)-изоморфны, а предикаты E_i и E'_i и φ -изоморфны, где $\varphi_i: A_i \rightarrow A'_i, i = \overline{1, n}$. Тогда найдутся такие биекции $\psi_i: A_i \rightarrow A'_i, i = \overline{1, n}$, что предикаты L и L' ($\psi_1, \psi_2, \dots, \psi_n$)-изоморфны.

Доказательство. Пусть $f_i: A_i \rightarrow B_i$ — характеристическая функция эквивалентности E_i , $F_i: A'_i \rightarrow B'_i$ — характеристическая функция эквивалентности E'_i . Из условий теоремы для любых $x_i \in A_i, x'_i \in A'_i, i = \overline{1, n}$, вытекает справедливость следующих равенств:

- 1) $P(x_1, x_2, \dots, x_n) = L(F_1(x_1), F_2(x_2), \dots, F_n(x_n))$,
- 2) $P'(x'_1, x'_2, \dots, x'_n) = L'(F'_1(x'_1), F'_2(x'_2), \dots, F'_n(x'_n))$,
- 3) $P(x_1, x_2, \dots, x_n) = P'(F_1(x_1), F_2(x_2), \dots, F_n(x_n))$,

откуда для любых $x_i \in A_i, i = \overline{1, n}$, верно

4) $L(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) = L'(F'_1(x'_1), F'_2(x'_2), \dots, F'_n(x'_n))$. Так как для любого $i = \overline{1, n}$ $F'_i(x'_i)$ эквивалентности E_i и E'_i φ_i -изоморфны, то, в силу утверждения 10, найдутся биекции $\psi_i: B_i \rightarrow B'_i, i = \overline{1, n}$ такие, что функции F и F' будут (φ_i, ψ_i) -изоморфны. Это означает выполнение для любого $x_i \in A_i$ равенств $F'(\varphi_i(x_i)) = \psi_i(F(x_i)), i = \overline{1, n}$. Подставляя правые части этих равенств в правую часть 4), получаем изоморфизм L и L' . Теорема доказана.

Чтобы продемонстрировать возможность использования на практике описанного здесь обобщенного метода компараторной идентификации, рассмотрим его применение для идентификации способности человека осмысливать воспринимаемые им ситуации. Испытуемому предъявляются для анализа ситуации и тексты. Ситуации — это все то, что может быть воспринято органами чувств человека; тексты — это все, что может быть им понято. В результате восприятия ситуации в сознании человека возникает образ этой ситуации; в результате понимания текста в сознании человека возникает смысл этого текста. Пусть A_1 — множество всевозможных ситуаций, A_2 — множество всех текстов. Испытуемый должен выполнить следующее задание: если ситуация $x \in A_1$ согласуется с текстом $y \in A_2$, то он реагирует сигналом $t = 1$, если не согласуется, то — сигналом $t = 0$. Например, взглянув из окна комнаты на улицу, испытуемый затем читает текст «Идет дождь». Если за окном на самом деле идет дождь, то он реагирует ответом 1, если же дождя нет, то — ответом 0. Своим поведением испытуемый реализует некоторый предикат $P(x, y) = t$, заданный на $A_1 \times A_2$, который называется ситуационно-текстовым.

Идентификация поведения испытуемого в данном случае имеет своей целью определение функций F_1 и F_2 и предиката L , фигурирующих на

схеме рис. 2, б. Функция F_1 формально описывает процесс преобразования ситуации в ее образ, функция F_2 описывает преобразование текста в его смысл. Предикат L описывает процесс осознания соответствия образа ситуации смыслу текста. Ситуации, образы которых совпадают, называются метамерными. Левая сопровождающая эквивалентность предиката P вводит разбиение множества A_1 на классы метамерных ситуаций. Тексты, смыслы которых совпадают, называются тождественными. Правая сопровождающая эквивалентность предиката P вводит разбиение множества A_2 на классы тождественных текстов. Множество всех ситуаций, согласующихся с данным текстом, может быть принято в качестве объективного эквивалента смысла данного текста.

Выводы

Идентификация интеллектуальной деятельности человека по данной схеме открывает путь к математическому описанию и искусственному воспроизведению таких важных для машинного интеллекта сторон разума как восприятие, понимание, узнавание и осознание. Ясно, что методы прямой идентификации в данном случае неприемлемы, поскольку образы ситуаций и смыслы текстов, являясь субъективными состояниями человека, в принципе недоступны прямому физическому измерению.

Список литературы: 1. Мальцев А.И. Алгебраические системы. — М.: Наука, 1970. — 392 с. 2. Кравков С.В. Цветовое зрение. — М.: Изд-во АН СССР, 1951. — 175 с. 3. Нюберг Н.Д. Грассмана законы // Физический энциклопедический словарь. Т. 1. — М.: Сов. энциклопедия, 1960. — С. 136. 4. Рабкин Е.Б. Атлас цветов. — М.: Медгиз, 1956. — 52 с. 5. Ивахненко А.Г. Долгосрочное прогнозирование и управление сложными системами. — К.: Техніка, 1975. — 311 с. 6. Раутиан Г.Н. Колориметрические приборы. Справочная книга оптика-механика. Ч. 1. — М.: ОНТИ, 1936. — 281 с. 7. Курош А.Г. Лекции по общей алгебре. — М.: Наука, 1975. — 400 с. 8. Мальцев А.И. Алгебраические системы. — М.: Наука, 1970. — 392 с.

Поступила в редколлегию 25.08.2008.

УДК 519.7

Про загальну теорію компараторної ідентифікації / М.Ф. Бондаренко, С.Ю. Шабанов-Кушнаренко, Ю.П. Шабанов-Кушнаренко // Біоніка інтелекту: наук.-техн. журнал — 2008. — № 2 (69). — С. 13-22.

У статті розвивається загальна теорія компараторної ідентифікації. Проаналізовано межі застосування методу компараторної ідентифікації. Розглянуто питання одиничності подання об'єкта при його компараторній ідентифікації та ізоморфізму моделей, що отримуються у результаті компараторної ідентифікації об'єктів. Узагальнено метод компараторної ідентифікації на випадок входів.

Л. 3. Бібліогр.: 8 найм.

UDC 519.7

About comparator identification general theory / M.F. Bondarenko, S.Yu. Shabanov-Kushnarenko, Yu.P. Shabanov-Kushnarenko // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 13-22.

In article the comparator identifications general theory is develops. The comparator identifications method application borders is analysed. The considered questions of uniqueness object representation at it comparator identification and isomorphism of models which turn out as a result comparator identifications of objects. The comparator identifications method on case of inputs is generalized.

Fig. 3. Ref.: 8 items.

УДК 519.7



ИДЕНТИФИКАЦИЯ ОБЪЕКТОВ, ОПИСЫВАЕМЫХ ЧИСЛОВЫМИ СИСТЕМАМИ

М.Ф. Бондаренко¹, С.Ю. Шабанов-Кушнаренко² Ю.П. Шабанов-Кушнаренко³^{1, 2, 3} ХНУРЭ, г. Харьков, Украина

В процессе компараторной идентификации объектов приходится записывать различные условия, которым удовлетворяют исследуемые предикаты. Чтобы иметь в дальнейшем возможность делать это беспрепятственно, надо располагать полным языком формального описания логических условий. В данной статье описывается разработанная нами для этой цели алгебра предикатных операций.

КОМПАРАТОРНАЯ ИДЕНТИФИКАЦИЯ, ЦВЕТОВОЕ ЗРЕНИЕ, ГИЛЬБЕРТОВО ПРОСТРАНСТВО

Введение

Представленные в настоящей статье результаты по идентификации понятия числа имеют много общего с известными положениями из учения об основаниях арифметики, поэтому необходимо проанализировать различие между ними. В нашей постановке речь идет только об идентификации (то есть математическом описании) понятия числа, вопрос об обосновании этого понятия не ставится. При решении задачи идентификации объектов все средства формального описания хороши, лишь бы они были надежны; нет необходимости их ограничивать, как это делается в математической логике при обосновании понятий арифметики. Снятие запрета на средства формального описания дает возможность идентифицировать именно ту арифметику, которая фактически используется в математической практике, а не тот ее вариант, который носит название формальной арифметики.

1. Разработка полного формального языка идентификации

Пусть U – какое-либо множество предметов, называемое универсумом. На множестве U вводим предметные переменные $x_1, x_2, \dots, x_m$. Если $a_1, a_2, \dots, a_m \in U$ и $x_1 = a_1, x_2 = a_2, \dots, x_m = a_m$, то будем говорить, что предметный вектор $(a_1, a_2, \dots, a_m)$ принадлежит пространству U^m размерности m . Любое подмножество P пространства U^m называется отношением на U^m . Любая функция $P(\xi) = (x_1, x_2, \dots, x_m)$, отображающая U^m в множество $\Sigma = \{0, 1\}$, называется предикатом на U^m . Пусть L – множество всех отношений на U^m , M – множество всех предикатов на U^m . Отношение $P \in L$ и предикат $P \in M$ называются соответствующими друг другу, если при любом $\xi = (x_1, x_2, \dots, x_m)$

$$P(x) = \begin{cases} 1, & \text{если } \xi \in P, \\ 0, & \text{если } \xi \notin P. \end{cases}$$

Предикатом узнавания предмета $a \in U$ по переменной x_i называется предикат x_i^a из M , определяемый условием

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

Отрицанием предиката P называется предикат $\bar{P}$, соответствующий отношению $\bar{P}$. Дизъюнкцией предикатов P и Q называется предикат $P \vee Q$, соответствующий отношению $P \cup Q$. Конъюнкцией предикатов P и Q называется предикат $P \wedge Q$, соответствующий отношению $P \cap Q$. Нулевым предикатом 0 называется предикат, соответствующий пустому отношению. Множество M всех предикатов на U^m называется универсумом предикатов.

На множестве M вводим предикатные переменные $X_1, X_2, \dots, X_n$. Множество M^n называется предикатным пространством размерности n . Элементы множества M^n называются предикатными векторами. Любая функция $F(X_1, X_2, \dots, X_n) = Y$, отображающая множество M^n в множество M , называется предикатной операцией. Образует множество R всех предикатных операций, отображающих M^n в M . Алгеброй предикатных операций над R называется любая алгебра, заданная на носителе R . Отрицанием предикатной операции F называется операция $\bar{F}$, отображающая R в R , значения которой определяются по правилу:

$$\bar{F}(X_1, X_2, \dots, X_n) = \overline{F(X_1, X_2, \dots, X_n)}.$$

Дизъюнкцией предикатных операций F и T называется операция $F \vee T$, отображающая $R \times R$ в R , значения которой определяются по правилу:

$$(F \vee T)(X_1, X_2, \dots, X_n) = F(X_1, X_2, \dots, X_n) \vee T(X_1, X_2, \dots, X_n).$$

Конъюнкцией предикатных операций F и T называется операция $F \wedge T$, отображающая $R \times R$ в R , значения которой определяются по правилу:

$$(F \wedge T)(X_1, X_2, \dots, X_n) = F(X_1, X_2, \dots, X_n) \wedge T(X_1, X_2, \dots, X_n).$$

Константной предикатной операцией называется любая предикатная операция вида $F(X_1, X_2, \dots, X_n) = P$, где P – предикат из M . Будем обозначать ее символом P и называть константным предикатом. Тожественной предикатной операцией по переменной $X_i (i = \overline{1, n})$ называется

операция вида $F(X_1, X_2, \dots, X_i) = X_i$. Будем обозначать ее символом X_i и называть предикатной переменной. Предикатной операцией узнавания предиката P по переменной $X_i (i = \overline{1, n})$ называется операция из R вида:

$$X_i^P = \begin{cases} 1, & \text{если } X_i = P, \\ 0, & \text{если } X_i \neq P. \end{cases}$$

Дизъюнктивно–конъюнктивной алгеброй предикатных операций называется такая алгебра над R , у которой базисными операциями служат дизъюнкция и конъюнкция предикатных операций, а базисными элементами являются любые константные предикаты и любые предикатные операции узнавания предиката.

Утверждение 1. (О полноте дизъюнктивноконъюнктивной алгебры предикатных операций). Любая предикатная операция выражается в дизъюнктивноконъюнктивной алгебре в виде:

$$F(X_1, X_2, \dots, X_n) = \vee F(P_1, P_2, \dots, P_n) X_1^{P_1} X_2^{P_2} \dots X_n^{P_n} \\ P_1, P_2, \dots, P_n \in M.$$

Доказательство очевидно.

Подстановкой $x_i / a(P) = Q$ значения $a \in U$ на место аргумента $x_i (i = \overline{1, m})$ называется предикатная операция, отображающая M в M и определяемая правилом:

$$Q(x_1, x_2, \dots, x_m) = P(x_1, x_2, \dots, x_{i-1}, a, x_{i+1}, \dots, x_m).$$

Подстановку можно понимать не только как предикатную операцию, но и как операцию над предикатными операциями. При таком понимании она определяется следующим образом:

$$x_i / a(F(X_1, X_2, \dots, X_n)) = \\ = F(x_i / a(X_1), x_i / a(X_2), \dots, x_i / a(X_n)).$$

Алгеброй подстановочных операций называется такая алгебра предикатных операций над R , у которой базисными операциями служат всевозможные подстановки вида:

$$x_i / a(i = \overline{1, m}, a \in U),$$

а также операции отрицания и дизъюнкции, а базисными элементами являются предикаты равенства $D(x_1, x_i), (i = \overline{2, m})$ и переменные $X_i (i = \overline{1, m})$.

Утверждение 2. (О полноте алгебры подстановочных операций). Любая предикатная операция выражается в алгебре подстановочных операций.

Доказательство. Теорема будет доказана, если все базисные элементы и базисные операции дизъюнктивноконъюнктивной алгебры будут выражены через базисные элементы и операции алгебры подстановочных операций. Операция дизъюнкции в алгебре подстановочных операций имеется. Выражаем конъюнкцию: $F \vee T = F \vee T$. Выражаем нулевой предикат: $0 = X_1 \wedge \overline{X_1}$. Выражаем уз-

навания предмета: $x_1^a = x_2 / a(D(x_1, x_2)), (a \in U),$ $x_i^a = x_i / a(D(x_1, x_i)), (i = \overline{1, m})$. Выражаем предикаты узнавания предикатов: $X_1^P = \vee \xi(X_1(\xi) \sim P(\xi))$. Любой постоянный предикат выражаем с помощью операций дизъюнкции и конъюнкции, примененных к нулевому предикату и предикатам узнавания предметов. Переменный предикат X_i в базисе алгебры подстановочных операций имеется. Выражаем операцию равнозначности: $F \sim FT = \overline{F} T \vee FT$. Выражаем квантор общности по предметному вектору $\vee \xi(F) = \vee x_1 \vee x_2 \dots \vee x_m (F)$. Выражаем квантор общности по переменной $x_i (i = \overline{1, m}) : \vee x_i (F) = \bigwedge x_i / a(F)$. Мы выразили все базисные элементы и все базисные операции дизъюнктивноконъюнктивной алгебры через базисные элементы и базисные операции алгебры подстановочных операций. Утверждение доказано.

Кванторы общности и существования по произвольному множеству R выражаются формулами:

$$\forall x \in A P(x) = \forall x (A(x) \supset P(x)), \\ \exists x \in A P(x) = \exists x (A(x) \wedge P(x)).$$

Квантор существования и единственности выражается в виде:

$$\exists! x P(x) = (\exists x P(x)) \wedge \forall x \forall y (P(x) \wedge P(y) \supset D(x, y)).$$

Кванторы общности и существования по предикату P имеют вид:

$$\forall P F(P) = \bigwedge_{P \in M} F(P); \\ \exists P F(P) = \bigvee_{P \in M} F(P).$$

Здесь $F(P)$ – произвольно выбранная формула алгебры предикатных операций, в которую входит переменный предикат P . Кванторы общности и существования по переменному отношению P , включенному в фиксированное отношение M , выражаются следующим образом:

$$\forall P \subseteq M F(P) = \bigwedge_{P \in M} \vee \xi (P(\xi) \supset M(\xi) \supset F(P)), \\ \exists P \subseteq M F(P) = \bigvee_{P \in M} \vee \xi (P(\xi) \supset M(\xi) \wedge F(P)).$$

2. Идентификация объектов, описываемых натуральными числами

Абстракция числовой системы широко используется на практике для целей математического описания совокупностей сигналов различного вида. Например световые излучения часто описывают в форме спектров с конечным числом линий, каждая из которых характеризуется неотрицательным числом. Ощущение цвета формально описывают в виде тройки чисел. Место работы человека характеризуется конечным набором числовых характеристик, таких как зарплата, продолжительность оплачиваемого отпуска, расстояние от места проживания до места работы и тому подобное. Как узнать, подходит та или иная совокупность

сигналов, встречающихся в реальном мире, под понятие числовой системы? Эта задача может быть решена методом логической идентификации. Под логической идентификацией понимается формальное описание предикатов, характеризующее изучаемые системы преобразования сигналов в виде некоторой совокупности логических условий, связывающих эти предикаты. Компараторная идентификация систем является частным случаем логической. В рассматриваемом случае задача логической идентификации сводится к формулировке достаточного числа свойств (признаков), по которым можно было бы распознать, является ли данная совокупность сигналов числовой системой или нет. Решение этой задачи начнем с установления характеристических признаков натурального ряда чисел. Затем это же будет сделано для системы рациональных чисел и, наконец, для системы вещественных чисел.

С содержательной точки зрения натуральный ряд чисел представляет собой такую совокупность N сигналов, каждый из которых можно занумеровать, то есть снабдить собственным именем, взятым из последовательности знаков $1, 2, \dots$. Такой процесс нумерации сигналов называется их счетом. Процесс счета можно формально охарактеризовать функцией счета $y = q(x)$, отображающей множество N в себя. Функция счета ставит в соответствие каждому сигналу x из множества N следующий за ним в натуральном ряду сигнал $y = q(x)$. Множество N и функция счета q обладают многими свойствами. Идентификация натурального ряда будет осуществлена, если удастся отобрать из числа таких свойств часть из них так, чтобы последние определяли множество N и функцию q в абстрактном смысле единственным образом (то есть с точностью до обозначений). Свойства, образующие такую исходную совокупность, называются аксиомами натурального ряда. Один из возможных вариантов аксиом натурального ряда впервые указал в 1889 году итальянский математик Пеано [1]. Эти аксиомы (записываемые нами на языке логики предикатов) будут далее использоваться при решении задачи идентификации понятия натурального ряда чисел.

Введем предикат $Q(x, y)$ на $N \times N$, соответствующий отношению, задаваемому равенством $y = q(x)$. Если $Q(x, y) = 1$, то будем говорить, что число y следует за числом x или что число x предшествует числу y . Значение предиката Q определяем для любых $x, y \in N$ следующим образом:

$$Q(x, y) = \begin{cases} 1, & \text{если } y = q(x), \\ 0, & \text{если } y \neq q(x). \end{cases} \quad (1)$$

Предикат Q называется предикатом счета. Пусть U – универсум, то есть. множество, интерпретируемое как множество всевозможных сигналов, а N

– какое-нибудь его подмножество. Множество N называется натуральным рядом чисел, если на его декартовом квадрате определен предикат Q , удовлетворяющий приведенным ниже четырем аксиомам. Выражение «предикат Q определен на $N \times N$ » содержательно интерпретируем в том смысле, что реально существует некий преобразователь сигналов $t = Q(x, y)$ с двумя входами x и y и одним выходом t , который на любую пару сигналов, взятых из множества N , однозначно реагирует сигналом 0 или сигналом 1. В частном случае в роли такого преобразователя сигналов может выступать человек или автомат.

Ниже формулируются четыре аксиомы натурального ряда чисел.

$$1) \forall x \in N \exists! y \in N Q(x, y).$$

Аксиома 1 означает, что за каждым числом натурального ряда следует единственное натуральное число. Она постулирует существование функции счета $y = q(x)$, отображающей множество N в N , которая соответствует предикату $Q(x, y)$.

$$2) \exists! y \in N \forall x \in N \neg Q(x, y).$$

Аксиома 2 означает, что в натуральном ряду имеется единственное число y , у которого нет предшествующего числа. Число y , обладающее свойством 2, то есть первое в натуральном ряду, называется единицей и обозначается символом 1.

$$3) \forall x_1, x_2, y \in N (Q(x_1, y) \wedge Q(x_2, y) \supset D(x_1, x_2)).$$

Аксиома 3 означает, что в натуральном ряду каждому числу предшествует не более, чем одно число. В аксиоме 3 запись $D(x_1, x_2)$ обозначает предикат равенства на U^2 .

$$4) \forall M \subseteq N (M(1) \wedge \forall x, y \in N (M(x) \wedge Q(x, y) \supset M(y)) \supset \forall x \in N M(x)).$$

Аксиома 4 представляет собой принцип математической индукции: если свойством M обладает число 1, и, кроме того, из предположения, что свойством M обладает число x , следует, что тем же свойством обладает и число $q(x)$, то свойством M обладают все натуральные числа. В аксиоме 4 символ M используется в двойной роли: в записи $M \subseteq N$ он обозначает множество, а в записях $M(1)$, $M(x)$ и $M(y)$ он обозначает предикат, соответствующий множеству M . Предикат, соответствующий множеству M , определяется для всех $x \in U$ следующим образом:

$$M(x) = \begin{cases} 1, & \text{если } x \in M, \\ 0, & \text{если } x \notin M. \end{cases} \quad (2)$$

В универсуме U может содержаться много разных подмножеств сигналов, которые идентифицируются как натуральные ряды чисел. Ранее мы содержательно определили предикат счета, выразив его через функцию счета. Однако предикат счета можно определить формально, не опираясь на содержательно введенное понятие функции счета, а именно: любой бинарный предикат Q на N

, удовлетворяющий аксиомам натурального ряда, называется предикатом счета. Теперь мы можем определить функцию счета формально, выразив ее через предикат счета: $y = q(x)$, если $Q(x, y) = 1$; $y \neq q(x)$, если $Q(x, y) = 0$.

Утверждение 3. Если Q на $N \times N$ и Q' на $N' \times N'$ — предикаты, удовлетворяющие аксиомам натурального ряда, то существует биекция $\varphi: N \rightarrow N'$ такая, что для всех $x, y \in N$ имеет место равенство $Q(x, y) = Q'(\varphi(x), \varphi(y))$.

Доказательство. Определим предикат Φ на $N \times N'$ следующим образом $\Phi(1, 1') = 1$; если $\Phi(x, x') = 1$ и $Q'(x', y') = 1$, то $\Phi(y, y') = 1$ (a). Покажем, что такое определение корректно, то есть данные условия определяют единственный предикат Φ , а также тот факт, что Φ — биекция. Предположим, некоторый предикат Φ удовлетворяет условиям (a). Рассмотрим элементы $a_1 \in N, a'_1 \in N'$ такие, что $Q(1, a_1) = Q'(1', a'_1) = 1$. Тогда $\Phi(a_1, a'_1) = 1$. Найдем элементы a_2, a'_2 такие, что $Q(a_1, a_2) = Q'(a'_1, a'_2) = 1$. Тогда $\Phi(a_2, a'_2) = 1$. Продолжая этот процесс, получим не более чем счетные множества $M = \{1, a_1, a_2, \dots, a_k, \dots\}$, $M' = \{1', a'_1, a'_2, \dots, a'_k, \dots\}$, причем, $\Phi(a_i, a'_i) = 1 \quad i = 1, 2, \dots, k, \dots$. Из аксиомы 4 следует $M = N, M' = N'$. Так как на каждом шаге в силу аксиомы 1 элементы a_i, a'_i определяются однозначно, и мы исчерпали множества N и N' , то предикат Φ условиями (a) задается единственным образом. Для доказательства биективности Φ достаточно показать, что на каждом шаге мы получали новые элементы a_i, a'_i , то есть не может существовать таких $p, s (p < i, s < i)$, что $a_i = a_p, a_i = a'_s$. Покажем это для элементов множества N . Предположим, $a_i = a_p, p < i$. Тогда $Q(a_{p-1}, a_p) = Q(a_{i-1}, a_p)$, откуда в силу 3 $a_{i-1} = a_{p-1}$. Точно так же показывается, что $a_{i-2} = a_{p-2}, \dots, a_{i-p+1} = a_1, a_{i-p} = 1$, но так как $p < i$, в силу 2 получаем противоречие. Аналогично показывается $a'_i \neq a'_s, s < i$. Мы показали биективность Φ . Положим $\varphi(x) = x' \Leftrightarrow \Phi(x, x') = 1$ для всех $x \in N, x' \in N'$. Из способа построения предиката Φ непосредственно вытекает $Q(x, y) = Q'(\varphi(x), \varphi(y))$ для всех $x, y \in N$. Утверждение доказано.

Таким образом, система 1 — 4 полна в том смысле, что она содержит достаточный набор признаков, с помощью которых можно распознать натуральный ряд чисел в любой совокупности сигналов. При этом идентификация производится с исчерпывающей полнотой, поскольку сигналы в любых совокупностях, распознанных как натуральные ряды чисел, отличаются только обозначениями, то есть несущественно. В абстрактном смысле натуральный ряд единственный. Любое множество, фактически имеющее структуру натурального ряда, всегда может быть идентифицировано как натуральный ряд чисел с помощью аксиом 1 — 4. Вместе с тем, любое реальное множество,

фактически не вписывающееся в содержательное представление о натуральном ряде, при идентификации с помощью свойств 1 — 4 не пройдет и будет отвергнуто.

3. Идентификация объектов, описываемых арифметикой натуральных чисел

Под арифметикой натуральных чисел здесь понимается натуральный ряд чисел вместе с операциями сложения и умножения и отношением линейного порядка на нем. Сначала формально введем понятие сложения натуральных чисел. Отправляясь от натурального ряда чисел, формально представленного предикатами N и Q , определяем на нем предикат $S(x, y, z)$ на N^3 следующими двумя аксиомами:

- 5) $\forall x \in N \exists u \in N (S(x, 1, u) \wedge Q(x, u))$,
- 6) $\forall x, y \in N \exists u, v, w \in N (S(x, u, w) Q(y, u) \wedge \wedge S(x, y, v) Q(u, w))$,

которые называются аксиомами сложения натуральных чисел. Предикат S , удовлетворяющий аксиомам 5 и 6, называется предикатом сложения. Можно доказать, что для каждого натурального ряда, представленного множеством N и предикатом Q , существует единственный предикат сложения, и что он функционален. Это означает, что предикату $S(x, y, z)$ соответствует функция $z = x + y$, которая называется сложением натуральных чисел. Сложение отображает $N \times N$ на N . С использованием функций счета и сложения свойства 5 и 6 записываются более просто в виде равенств $x + 1 = q(x)$ и $x + (y + 1) = (x + y) + 1$, справедливых для всех $x, y \in N$. Очевидно, что свойства так формально определенной операции $+$ совпадают со всеми свойствами рассматриваемой в арифметике операции сложения натуральных чисел. Выражение «множество сигналов N идентифицируется как натуральный ряд, на котором определена операция сложения чисел» содержательно означает, что на заданном множестве сигналов N действуют два преобразователя сигналов, один $N = Q(x, y)$ с двумя входами и второй $t' = S(x, y, z)$ — с тремя, выходные сигналы t и t' которых принимают значения 0 и 1 и подчиняются условиям 1 — 6.

После того, как на множестве натуральных чисел N введено сложение S , формально вводим на нем операцию умножения. Для этого определяем на множестве C предикат $P(x, y, z)$ следующими двумя аксиомами:

- 7) $\forall x \in P(x, 1, x)$,
- 8) $\forall x, y \in N \exists u, v, w \in N (P(x, u, w) Q(y, u) \wedge \wedge S(v, x, w) P(x, y, v))$,

называемыми аксиомами умножения натуральных чисел.

Предикат P на N^3 , удовлетворяющий аксиомам 7 и 8, называется предикатом умножения. Можно доказать, что для каждого натурального

ряда, характеризуемого предикатами N и Q , существует единственный предикат умножения, и что он функционален. Это означает, что предикату $P(x, y, z)$ может быть поставлена в соответствие функция $z = x * y$, которая называется умножением натуральных чисел. Умножение отображает $N \times N$ в N . С использованием функций сложения и умножения чисел свойства 7 и 8 запишутся более просто в виде равенств $x * 1 = x$ и $x * (y + 1) = x * y + x$, справедливых для всех $x, y \in N$. Очевидно, что свойства так определенной операции $*$ совпадают со свойствами рассматриваемой в арифметике операции умножения натуральных чисел.

Для того чтобы идентифицировать заданную совокупность сигналов N как натуральный ряд, на котором определены операции сложения и умножения чисел, достаточно убедиться в том, что на ней действуют три преобразователя сигналов Q , S и P , первый – с двумя входами, второй и третий – с тремя, реагирующие на входные сигналы двоичным ответом, поведение которых подчиняется требованиям 1 – 8. Именно благодаря действию этих трех преобразователей, совокупность сигналов N превращается в множество натуральных чисел с операциями сложения и умножения. Если действие преобразователей Q , S и P прекращается, то множество N тотчас же превращается в такую совокупность сигналов, о которой можно сказать лишь то, что ее мощность счетна. Утверждать, что сигналы совокупности N суть полноценные натуральные числа, которые можно реально сосчитать, складывать и умножать, теперь нет никаких оснований. В данном случае на множестве N можно ввести операции сложения и умножения его элементов, но фактически они не введены.

Вслед за сложением S на множестве N натуральных чисел, кроме умножения можно определить бинарный предикат порядка T . Он характеризуется единственным свойством, называемым аксиомой порядка:

$$9) \forall x, y \in N (T(x, y) \rightarrow \exists u \in N S(x, u, y)).$$

Предикат порядка условием (а) определяется единственным образом по предикату сложения S . Если $x, y \in N$ таковы, что $T(x, y) = 1$, то будем писать $x < y$. Аксиома 9 означает, что при любых $x, y \in N$ утверждение $x < y$ равносильно утверждению о существовании такого $u \in N$, при котором $x + u = y$. Предикат порядка введен прямым определением. В отличие от этого предикаты счета, сложения и умножения натуральных чисел были введены косвенными определениями. Косвенным определением было введено также и само множество натуральных чисел N .

Прямое определение характеризуется тем, что вновь вводимое понятие явно выражается через понятия, введенные ранее. Каждое понятие представлено своим предикатом. Косвенные опреде-

ления выражают вводимые ими понятия неявно, связывая их одним или несколькими логическими уравнениями (условиями) с уже введенными ранее понятиями. Множество N и предикат Q не выражаются через другие понятия, являясь первичными для арифметики, они неявно задаются системой связывающий их логических уравнений 1 – 4. Аналогично отношению $x < y$ прямым определением вводится отношение $x \leq y$. Утверждение $x \leq y$ равносильно утверждению « $x < y$ или $x = y$ ». С содержательной точки зрения отношения $<$ и $\leq$, равно как и отношение равенства $=$, можно считать заданными на множестве сигналов N . Кроме преобразователей Q , S и P , действует также и преобразователь $T(x, y)$, обладающий свойством 9, а также преобразователи, соответствующие отношениям $\leq$ и $=$.

Необходимо особо отметить, что на базе изложенной выше теории идентификации числовых объектов не представляется возможным распознавать реальные процессы, подпадающие под другие, еще не введенные формально, понятия арифметики натуральных чисел (например, вычитания и деления, четного и простого числа и тому подобного). Чтобы получить возможность распознавать и эти процессы, требуется для каждого из них предварительно ввести их характеристические свойства. Делается это просто: на основе прямого определения соответствующего понятия, уже введенного в применяемой на практике арифметике натуральных чисел. Каждое такое определение добавляется к числу аксиом (свойств), с помощью которых осуществляется идентификация объектов. При этом каждому вновь привлекаемому понятию ставится в соответствие некоторый предикат, явно выражаемый через ранее введенные предикаты. С содержательной точки зрения это означает, что на множестве сигналов, формально представляемых как натуральные числа, придется вводить все новые и новые преобразователи с двоичным ответом. Реализация этих преобразователей средствами вычислительной техники позволит искусственно воспроизводить арифметические способности человека. Возможности искусственного интеллекта будут постепенно расширяться по мере освоения машиной все новых и новых арифметических понятий. Подобным способом можно развивать не только арифметические, но и любые другие интеллектуальные способности машины.

4. Идентификация объектов, описываемых арифметикой рациональных чисел

Обозначим множество всех положительных рациональных чисел символом B . Ближайшая задача будет состоять в том, чтобы формально задать множество B такими логическими условиями, которые бы определили понятие положительно-

го рационального числа, в точности соответствующее тому понятию, которое под тем же именем фактически используется в классической арифметике рациональных чисел. Будем предполагать, что множество N натуральных чисел, предикат Q , а также операции $+$, $*$ и отношение $<$ уже выбраны в соответствии с их формальными определениями, указанными выше. Пусть R — некоторый предикат на $N \times N \times B$, называемый предикатом деления натуральных чисел. Содержательно предикат $R(x, y, z)$ интерпретируем как связь $x * z = y$, где x и y — натуральные числа, z — положительное рациональное число. Операция, обозначенная « $*$ », понимается как умножение натурального числа x на такое положительное рациональное число z , что в результате получается натуральное число y .

Ниже формулируются четыре аксиомы положительных рациональных чисел:

$$10) \forall x, y, z \in N \forall z_2 \in B P(y, z_1, x) \wedge \\ \wedge R(x, y, z_2) \supset z_1 = z_2.$$

Содержательно аксиома 10 означает следующее: если имеется положительное рациональное число z_1 , удовлетворяющее условию $yz_1 = x$, где x, y — натуральные числа, и операция $\frac{x}{y} = z_2$ дает в результате положительное рациональное число z_2 , то всегда $z_1 = z_2$. Иначе говоря, операция $\frac{x}{y} = z_2$ дает, когда x нацело делится на y , тот же результат, что и решение уравнения $yz_1 = x$. Это означает, что предикат R совпадает (после соответствующей перестановки аргументов) с предикатом P , если область значений его аргумента z_2 сузить с множества B до множества N . Следствием этого свойства является то, что все натуральные числа оказываются включенными в множество положительных рациональных чисел.

$$11) \forall x, y \in N \exists! z \in B R(x, y, z).$$

Согласно аксиоме 11, предикату R соответствует функция $r(x, y) = z$, отображающая $N \times N$ в B . Иначе говоря, решение уравнения $y * z = x$ относительно переменной z всегда существует и единственно в области положительных рациональных чисел при условии, что x и y — натуральные числа. Функцию $r(x, y)$ будем записывать в виде дроби: $r(x, y) = \frac{x}{y}$. Подчеркнем, что дробь — это не положительное рациональное число, а операция над двумя натуральными числами x и y . Горизонтальная черта, разделяющая натуральные числа x и y — это просто другой способ обозначения операции r .

$$12) \forall z \in B \exists x, y \in N R(x, y, z).$$

Аксиома 12 утверждает, что любое положительное рациональное число z можно представить в виде дроби $\frac{x}{y}$, иначе говоря, функция R сюръективна.

$$13) \forall x_1, x_2, y_1, y_2 \in N (\exists z \in B R(x_1, y_1, z) \wedge \\ \wedge R(x_2, y_2, z) \sim \exists t \in N P(x_1, y_1, t) \wedge P(x_2, y_1, t)).$$

Аксиома 13 формально выражает известное правило, определяющее равенство дробей: для любых натуральных чисел x_1, x_2, y_1, y_2

$$\frac{x_1}{y_1} = \frac{x_2}{y_2} \Leftrightarrow x_1 y_2 = x_2 y_1.$$

Любое множество B , для которого существует предикат R на $N \times N \times B$, удовлетворяющий аксиомам 10–13, называется множеством положительных рациональных чисел, согласованным с натуральным рядом, определяемым множеством N и предикатом Q . Характеризуя натуральный ряд чисел, приходится указывать не только множество N , но и предикат Q . Вместе с тем, предикат P указывать не обязательно, поскольку он однозначно определяется выбором N и Q .

Пусть N и Q фиксированы, $R(x, y, z)$ и $R'(x, y, z')$ — предикаты на $N \times N \times B$ и $N \times N \times B'$, удовлетворяющие условиям 10–13. Тогда можно доказать, что существует биекция $\phi: B \rightarrow B'$ такая, что для всех $x, y \in N$ и $z \in B$ $R(x, y, z) = R'(x, y, \phi(z))$. Иными словами, положительные рациональные числа определяются аксиомами 10–13 с точностью до изоморфизма, то есть в абстрактном смысле однозначно.

Сложение положительных рациональных чисел выражается через сложение и умножение натуральных чисел следующим прямым определением:

$$\forall x_1, x_2, y_1, y_2 \in N (\frac{x_1}{x_2} + \frac{y_1}{y_2} = \frac{x_1 * y_2 + x_2 * y_1}{x_2 * y_2}). \quad (3)$$

Аналогично определяется умножение положительных рациональных чисел:

$$\forall x_1, x_2, y_1, y_2 \in N (\frac{x_1}{x_2} * \frac{y_1}{y_2} = \frac{x_1 * y_1}{x_2 * y_2}). \quad (4)$$

Отношение $<$ порядка на множестве положительных рациональных чисел формально вводим следующим определением:

$$\forall x_1, x_2, y_1, y_2 \in N (\frac{x_1}{x_2} < \frac{y_1}{y_2} \sim x_1 * y_2 < x_2 * y_1), \quad (5)$$

которое выражает его через отношение порядка и операцию умножения натуральных чисел. Нетрудно доказать, что определения операций сложения и умножения и отношения порядка для положительных рациональных чисел корректны, то есть значения этих функций и отношения не зависят от того, какими дробями представлены числа.

Переходим к формальному определению понятия произвольного рационального числа. Множество всех рациональных чисел обозначаем символом C . Пусть T — некоторый предикат на $B \times B \times C$. Содержательно предикат $T(x, y, z)$ интерпретируем как связь $x + z = y$, где x, y — положительные рациональные числа, z — рациональное число.

Операция, обозначенная знаком «+», в данном случае понимается как сложение положительного рационального числа x с таким рациональным числом z , что в результате получается положительное рациональное число y .

Пусть $S_B(x, y, z)$ — предикат на B^3 , соответствующий отношению $x + y = z$, где x, y, z — положительные рациональные числа. Ниже формулируются четыре аксиомы рациональных чисел:

$$14) \forall x, y, z_1 \in B \quad \forall z_2 \in C \quad S_B(y, z_1, x) \wedge T(x, y, z) \supset$$

$$\supset z_1 = z_2,$$

$$15) \forall x, y \in B \quad \exists! z \in C \quad T(x, y, z),$$

$$16) \forall z \in C \quad \exists x, y \in B \quad T(x, y, z),$$

$$17) \forall x_1, x_2, y_1, y_2 \in B (\exists z \in C \quad T(x_1, y_1, z) \wedge \\ \wedge T(x_2, y_2, z) \supset \exists S \in B \quad S_B(x_1, y_2, S) \wedge S_B(x_2, y, S)).$$

Содержательно аксиома 14 означает следующее: если имеется рациональное число z_1 , удовлетворяющее условию $x + z_1 = y$, где x, y — положительные рациональные числа, и операция $x - y = z_2$ дает в результате рациональное число z_2 , то всегда $z_1 = z_2$. Иначе говоря, операция $x - y = z_2$, когда существует положительная разность $x - y$, дает тот же результат, что и решение уравнения $yz_1 = x$. Это означает, что предикат T совпадает (после соответствующей перестановки аргументов) с предикатом S_B , если область значения его аргумента z_2 сузить с множества C до множества B . Следствием этого свойства является то, что все положительные рациональные числа оказываются включенными в множество рациональных чисел.

Согласно аксиоме 15, предикату T соответствует функция $t(x, y) = z$, отображающая $B \times B \rightarrow C$. Иначе говоря, решение уравнения $y + z = x$ относительно переменной z всегда существует и единственно в области рациональных чисел при условии, что x и y — положительные рациональные числа. Функцию $t(x, y)$ будем записывать в виде разности: $t(x, y) = x - y$. Аксиома 16 утверждает, что любое рациональное число z можно представить в виде разности $x - y$, иначе говоря, функция t сюръективна. Аксиома 17 формально выражает известное правило, определяющее равенство разностей: для любых положительных рациональных чисел x_1, x_2, y_1, y_2

$$x_1 - y_1 = x_2 - y_2 \Leftrightarrow x_1 + y_2 = x_2 + y_1.$$

Любое множество C , для которого существует предикат T на $B \times B \rightarrow C$, удовлетворяющий аксиомам 10-17, называется множеством рациональных чисел, согласованным с натуральным рядом, определяемым множеством N и предикатом Q . Можно доказать, что множество C и предикат T определяются множеством N и предикатом Q с точностью до обозначений, то есть в абстрактном смысле однозначно.

Сложение рациональных чисел выражается через сложение положительных рациональных чисел следующим прямым определением:

$$\forall x_1, x_2, y_1, y_2 \in B \\ (x_1 - y_1) + (x_2 - y_2) = (x_1 + x_2) - (y_1 + y_2). \quad (6)$$

Аналогично определяется умножение рациональных чисел:

$$\forall x_1, x_2, y_1, y_2 \in B \quad ((x_1 - y_1) * (x_2 - y_2) = \\ = (x_1 * x_2) - (y_1 * y_2) - (x_1 * y_2 + y_1 * x_2)). \quad (7)$$

Отношение порядка на множестве рациональных чисел формально вводим следующим определением:

$$\forall x_1, x_2, y_1, y_2 \in B \\ (x_1 - y_1 < x_2 - y_2 \sim y_1 + x_2 < y_1 + x_2). \quad (8)$$

Число 0 вводим определением $0=11$. Между различными арифметиками рациональных чисел, удовлетворяющими аксиомам 1-17, существует параллелизм понятий. Параллельные понятия в таких арифметиках отличаются между собой только обозначениями и по существу идентичны.

5. Идентификация объектов, описываемых арифметикой вещественных чисел

При формальном введении вещественных чисел будем предполагать, что множество рациональных чисел уже формально определено вместе с операциями сложения и умножения и отношением порядка. Пусть C — множество всех рациональных чисел, а $+$ и $*$ — суть операции сложения и умножения рациональных чисел и $<$ есть отношение порядка на C . Множество всех вещественных чисел обозначаем символом V . Пусть $L(x, u)$ — предикат на $C \times V$. С его помощью мы свяжем натуральные числа x с вещественными числами u . В роли аксиом вещественного числа будут использованы некоторые свойства предиката L . Предварительно нам потребуется понятие сечения, введенное в 1872 году немецким математиком Дедекиндом [2].

Любое подмножество множества C называется сечением, если оно удовлетворяет следующим трем условиям: 1) подмножество содержит рациональное число, но не каждое рациональное число; 2) каждое содержащееся в подмножестве рациональное число меньше каждого не содержащегося; 3) в подмножестве нет наибольшего рационального числа. Наглядно сечение α можно представить как левую часть оси рациональных чисел (рис. 1). В множестве α нет крайней правой рациональной точки. Но зато может существовать крайняя левая точка a вне сечения. Между точкой a и любой точкой сечения α можно вставить еще одну рациональную точку. Наименьшее рациональное число, не принадлежащее сечению α , называется границей сечения α . Не исключаются такие сечения, у которых нет границы в виде рационального числа.

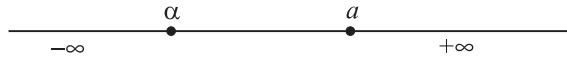


Рис. 1

Множество V и предикат L требуется связать такими условиями, чтобы для каждого вещественного числа u множеством всех корней уравнения $L(x, u) = 1$ относительно переменной x являлось некоторое сечение α , причем, когда переменная u пробегает все свои значения, то множество всех корней уравнения $L(x, u) = 1$ пробегает всевозможные сечения. Если существует рациональная граница a сечения α , то в роли вещественного числа u принимается рациональное число a , то есть $u = a$. Такое вещественное число называется рациональным. Все рациональные числа входят в состав множества вещественных чисел V . Если для сечения α граница не существует, то вещественное число u называется иррациональным.

После пополнения множества рациональных чисел всеми иррациональными ось рациональных чисел превращается в ось вещественных чисел. В результате у каждого сечения появляется единственная граница — рациональная или иррациональная. И каждой вещественной границе всегда соответствует единственное сечение. В силу наличия взаимно однозначного соответствия между вещественными числами и сечениями можно вещественные числа считать просто именами сечений. Оказывается, что вещественные числа, как и числа рациональные, линейно упорядочены. Это позволяет представить множество всех вещественных чисел как числовую ось.

Ниже приводится формулировка семи свойств множества V и предиката L , которые принимаются нами в роли аксиом, определяющих понятие вещественного числа.

- 18) $\forall u \in V \exists x \in C L(x, u)$,
- 19) $\forall u \in V \exists x \in C \neg L(x, u)$,
- 20) $\forall x, y \in C \forall u \in V (L(x, u) \wedge \neg L(x, u) \supset x < y)$,
- 21) $\forall u \in V \forall x \in C (L(x, u) \supset \exists y \in C (L(y, u) \wedge (x < y)))$,
- 22) $\forall u \in V \forall x \in C (\neg L(x, u) \wedge (\forall y \in C (\neg L(y, u) \supset x \leq y)) \supset x = u)$,
- 23) $\forall u, v \in V (\forall x \in C (L(x, u) \sim L(x, v)) \supset u = v)$,
- 24) $\forall M \subseteq C ((\exists x \in C M(x) \wedge \exists x \in C \neg M(x)) \wedge \forall x \in C \forall y \in C M(x) \wedge \neg M(y) \supset x < y) \wedge \forall x \in C (M(x) \supset \exists y \in C (M(y) \wedge (x < y))) \supset \exists u \in V \forall x \in C (M(x) \sim L(x, u)))$.

Аксиома 18 означает, что сечение u не пусто, то есть содержит по крайней мере одно рациональное число. Аксиома 19 означает, что в каждом сечении u содержатся не все рациональные числа. Вместе взятые, эти две аксиомы формально задают первое свойство сечения. Аксиома 20 формально характеризует второе свойство сечения: любое рациональное число, содержащееся в сечении, меньше любого

рационального числа, не содержащегося в нем. Аксиома 21 формально выражает третье свойство сечения: какое бы число сечения мы не взяли, в этом же сечении найдется большее число. Аксиома 22 гласит: если сечение u имеет рациональную границу x , то оно отождествляется с рациональным числом x . Аксиома 23 утверждает, что каждому сечению соответствует не более одного вещественного числа. Одновременно с этим очевидно, что каждому вещественному числу соответствует некоторое сечение. Поэтому имеет место взаимно однозначное соответствие между всеми сечениями и всеми вещественными числами. Аксиома 24 означает, что для каждого сечения найдется соответствующее ему вещественное число. Свойства 18-24 называются аксиомами вещественных чисел.

Утверждение 4. Если L на $C \times V$ и L' на $C \times V'$ — предикаты, удовлетворяющие аксиомам 18-24, то существует биекция $\varphi: V \rightarrow V'$ такая, что для всех $x \in C$ и $u \in V$ $L(x, u) = L(x, \varphi(u))$.

Доказательство. Пусть W — система всех подмножеств множества C . Построим на множестве $W \times V$ отношение ψ_1 следующим образом: $M \psi_1 u$, если и только если $x \in C$ ($M(x) \sim L(x, u)$). Покажем, что это отношение является биективным отображением множества сечений рациональных чисел на множество V . Действительно, аксиомы 18-21 показывают, что $M \psi_1 u$ влечет выполнение для M всех свойств сечения. Из самого определения отношения ψ_1 следует, что для каждого $u \in V$ найдется единственный предикат M такой, что $M \psi_1 u$. Остается показать, что для любого предиката M , соответствующего сечению рациональных чисел, найдется единственный элемент $u \in V$ такой, что $M \psi_1 u$. Существование такого элемента вытекает из аксиомы 24, а единственность следует из 23. Таким образом, ψ_1 — биекция. Зададим отношение ψ_2 на множестве $W \times V'$ следующим образом: $M \psi_2 u$, если и только если $x \in C$ ($M(x) \sim L'(x, u')$). Так же, как и раньше, показывается, что ψ_2 — биекция. Определим биекцию φ следующим образом: $\varphi = \psi_2 \psi_1^{-1}$. Тогда из определений отношений ψ_1 и ψ_2 , а также из 21 следует, что для всех $x \in C$ и $u \in V$ $L(x, u) = L'(x, \varphi(u))$. Утверждение доказано.

Утверждение 4 означает, что условия 18-24 определяют множество вещественных чисел в абстрактном смысле единственным образом. Иными словами, система условий 18-24, определяющая понятие вещественного числа, полна. Вместе с тем, используемое на практике понятие вещественного числа, очевидно, удовлетворяет всем условиям 18-24. Таким образом, условия 18-24 можно рассматривать в качестве полной системы аксиом, формально определяющей понятие вещественного числа.

Сложение вещественных чисел формально определяется следующим образом:

$$\forall x, y \in C \forall u, v \in V (x \in u \wedge y \in v \sim x + y \in u + v). \quad (9)$$

Здесь символы u и v удобно рассматривать двояко: и как сечения, которым принадлежат рациональные числа (в таком смысле понимаем записи $x \in u$ и $y \in v$), и как вещественные числа, которые можно складывать (такое понимание естественно для записи $u + v$). Запись $x \in u$ равносильна утверждению $L(x, u) = 1$. Согласно определению (9) сумма $u + v$ сечений u и v образуется из всевозможных сумм $x + y$ рациональных чисел, одно из которых x принадлежит сечению u , а другое y сечению v . Умножение вещественных чисел формально определяется так:

$$\forall x, y \in C \forall u, v \in V (x \in u \wedge y \in v \sim xy \in uv). \quad (10)$$

Согласно определению (10) произведение $u * v$ сечений u и v образуется из всевозможных произведений $x * y$ рациональных чисел, одно из которых x принадлежит сечению u , а другое y сечению v .

Порядок $<$ на множестве вещественных чисел определяется следующим образом:

$$\forall u, v \in V (u \subset v \sim u < v). \quad (11)$$

Согласно этому определению вещественное число u меньше вещественного числа v в том и только том случае, когда сечение u строго включено в сечение v . Из условия (11) следует, что для любых вещественных чисел либо $u < v$, либо $u = v$, либо $u > v$. Это означает, что все вещественные числа вытягиваются в линию, поэтому естественно говорить о множестве всех вещественных чисел как об оси. Если ограничить сферу действия операций $+$, $*$ и отношения $<$ рациональными вещественными числами, то они переходят в соответствующие операции и отношение над рациональными числами. Операции вычитания и деления вещественных чисел вводятся аналогично сложению и умножению с помощью прямых определений. Вычитание представляет собой операцию, отображающую $V \times V$ в V , а операция деления отображает $V \times (V \setminus 0)$ в V .

Выводы

Наиболее близкую к формулируемой здесь постановку задачи мы находим в классической работе Ландау [3], опубликованной впервые в 1930 году. Насколько нам известно, до настоящего времени результаты этой работы не пересматривались и не улучшались. Главный недостаток данной работы, оцениваемой нами с точки зрения задачи идентификации (а такая задача в ней не ставится, поскольку речь там идет только об обосновании ариф-

метики), заключается в том, что в ней все аксиомы арифметики записаны на неформализованном логическом языке, то есть на том языке, который был общепринятым среди математиков в то время, когда эта работа была написана. При решении задачи идентификации этого недостаточно: необходимо использовать язык логики предикатов. Именно это мы и осуществили в данной статье.

Аксиомы 1-9 по существу повторяют формулировки Ландау, различие заключается лишь в языке описания. Аксиомы 10-13 у Ландау вовсе отсутствуют. Это обусловлено тем, что он не различает дроби как пары натуральных чисел и как положительные рациональные числа и поэтому не вводит связывающий их предикат R . Однако при использовании языка логики предикатов сделать это необходимо. То же самое относится и к аксиомам 14-17, в которых фигурирует отсутствующий у Ландау предикат T , связывающий пары положительных рациональных чисел с соответствующими им рациональными числами. У Ландау отсутствует также предикат L , связывающий сечения с соответствующими им вещественными числами. Введение предиката L потребовало существенно иной системы аксиом для теории вещественного числа (аксиомы 18-24).

Список литературы: 1. Клини С. Математическая логика. — М.: Мир, 1973. — 480 с. 2. Дедекин Ю. Непрерывность и иррациональные числа. Одесса: 1923. — 40 с. 3. Ландау Э. Основы анализа. — М.: ИЛ, 1947. — 182 с.

Поступила в редколлегию 28.08.2008

УДК 519.7

Ідентифікація об'єктів, що описуються числовими системами / М.Ф. Бондаренко, С.Ю. ШабановКушнаренко, Ю.П. ШабановКушнаренко // Біоніка інтелекту: наук.-техн. журнал. — 2008. — № 2 (69). — С. 23-31.

У статті розвивається теорія компараторної ідентифікації. Розроблено повну формальну мову компараторної ідентифікації. Розроблено методи ідентифікації об'єктів, що описуються натуральними числами, арифметиками натуральних, раціональних та дійсних чисел.

Лл. 1. Бібліогр.: 3 найм.

UDC 519.7

Identification of the objects which are described by numbers systems / M.F. Bondarenko, S.Yu. ShabanovKushnarenko, Yu.P. ShabanovKushnarenko // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 23-31.

In article the comparator identifications theory is develops. The comparator identifications full formal language is developed. The objects identification methods which are described by natural numbers, natural, rational and real numbers arithmetics are developed.

Fig. 1. Ref.: 3 items.

УДК 004.75:004.451.82



ВЕБ-СЕРВИСЫ В ЛИНГВИСТИЧЕСКИХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

А. Р. Ерошенко

УМИФ, г. Киев, Украина, demode@ukrpost.ua

В статье рассмотрены принципы работы веб-сервиса. Автором выделены основные компоненты сервис-ориентированной архитектуры (SOA) и обоснована технология взаимодействия компонентов веб-сервисов. Проанализирована возможность применения технологии CORBA к задаче создания и использования лингвистических ресурсов.

ВЕБ-СЕРВИСЫ, ИНТЕРНЕТ, CORBA, SOA, JAVA, XML, SOAP, WSDL

Введение

С ростом популярности технологии WEB 2.0 в последнее время значительно увеличилось количество ресурсов Интернет, предоставляющих не только доступ к информации, но и позволяющие решать гораздо более широкий круг задач. Совсем недавно использование Интернет для поиска информации и поиска файлов было вполне достаточным. Сегодня же естественной является функция создания и редактирования посредством Интернет-браузера сложных документов, графических файлов, управления научными, производственными и бизнес-процессами. Кроме того, неуклонно возрастает роль лингвистических технологий в создании информационных систем, особенно для класса интеллектуальных. В связи с этим особо актуализируются задачи создания лингвистических ресурсов и возможности применения компьютерных технологий в лексикографии.

В отечественной научной литературе уделяется определенное внимание компьютерным технологиям лексикографической деятельности. Прежде всего, это научные разработки В. Широкова [1], А. Костышина [2], А. Рабульца [3], К. Якименко [4]. В этих работах ставится и решается целый ряд задач: перевод в цифровую форму традиционного лексикографического наследия, разработка инструментальных систем для создания электронных лингвистических продуктов, проведение разноплановых лингвистических исследований.

Существующие на сегодняшний день лексикографические исследования и проекты, целью которых является создание лингвистических ресурсов, выполняются значительными коллективами исследователей, зачастую распределенных территориально. Следовательно, стоит задача создания инструмента, позволяющего выполнять большие лингвистические проекты в режиме сетевого взаимодействия. Такая постановка задачи является совершенно новой для прикладной лингвистики.

Целью данного исследования является обоснование возможностей применения технологии CORBA для создания и использования лингвистических ресурсов. Реализация поставленной цели обусловила необходимость решения таких задач:

- описание принципа работы веб-сервиса;
- выделение основных компонентов сервис-ориентированной архитектуры;
- обоснование технологии взаимодействия компонентов веб-сервисов и операций, которые они выполняют;
- описание алгоритма взаимодействия информационных технологий и лингвистических ресурсов.

1. Теория веб-сервисов

Современный веб-сайт можно представить как совокупность взаимодействующих между собой сервисов. Обмен информацией может происходить также с сервисами, расположенными на других сайтах. Таким образом, мы получаем взаимодействующую гетерогенную среду, построенную по технологии, которая называется *CORBA*¹ [5].

CORBA является интегральной технологией взаимодействия веб-сервисов. Одной из ее главных составляющих является технология сервисно-ориентированной архитектуры *SOA* (*Service-Oriented Architecture*) [6]. На основе *SOA*-технологии нами создан веб-сервис, дающий возможность пользователям Интернет в режиме просмотра работать с реестром нескольких украинских словарей. Уникальность данной системы состоит в том, что пользователь может получить доступ к любому слову или словосочетанию словарной статьи, тогда как большинство мировых лингвистических систем позволяют выполнять данную процедуру только с реестром слов.

По своей сути *SOA* является новым направлением развития информационных технологий, представляющих собой создание глобальной сети распределенных вычислений. Наиболее подходящей средой функционирования данной системы являются интернет-сети. Инструментом для создания *SOA*-приложения могут быть многие популярные системы разработки: *.NET*, *PHP*, *Java*, и так далее. Этот список с каждым годом будет все более расширяться. Данный инструментальный средств

¹ *CORBA* (*Common Object Request Broker Architecture*) — общая архитектура брокера объектных запросов. Это технологический стандарт написания распределённых приложений, продвигаемый консорциумом *OMG*.

разработки может быть интегрирован в различные аппаратные платформы. Главное в том, что столь различные приложения могут обмениваться данными, то есть эффективно функционировать.

Прежде чем приступить к рассмотрению лингвистического веб-сервиса, опишем теорию функционирования распределенных вычислительных ресурсов, которыми в нашем случае являются веб-сервисы. Технологии Web-сервисов — это набор основанных на XML спецификаций, обеспечивающих универсальный метод технического описания сервисов и взаимодействия с ними. Сами сервисы, реализованные в соответствии с этими спецификациями, называют Web-сервисами. Возможно, именно технологии Web-сервисов сыграли свою роль в наименовании сервис-ориентированной архитектуры. На самом деле принципы SOA существуют уже как минимум третье десятилетие, в то время как о Web-сервисах мы узнали не так давно. На протяжении большей части этого времени доминировала объектно-ориентированная идеология построения систем, задачи технической реализации которой сводилась главным образом к возможности вызова удаленных объектов

Концепция SOA подразумевает существование информационных процессов в виде совокупности сервисов, взаимодействующих друг с другом.

Различают следующие три основных компонента сервис-ориентированной архитектуры:

- пользователь сервиса: приложение, программный модуль либо сервис, осуществляющий поиск и вызов необходимого сервиса из реестра сервисов по описанию сервиса, а также использующий сервис, предоставляемый провайдером сервиса, в соответствии с интерфейсом сервиса;
- провайдер сервиса: приложение, программный модуль либо сервис, осуществляющий реализацию сервиса в виде веб-сервиса, прием и исполнение запросов пользователей сервиса, а также публикацию сервиса в реестре сервисов;
- реестр сервисов: библиотека сервисов, предоставляющая пользователям сервиса средства поиска

и вызова необходимого сервиса и принимающая запросы провайдеров сервисов на публикацию сервисов.

Современный веб-сервис может быть как компонентом пользователя сервиса, так и сочетать в себе компонент провайдера, что само по себе увеличивает его работоспособность и стабильность работы. Введем обозначения, которые будут нами использоваться при описании принципа работы данной системы (рис. 1):

WSDL — язык описания веб-сервисов, основанный на языке XML; определяет веб-интерфейс для любых функций, которые используют удаленный вызов процедур (RPC);

UDDI — платформонезависимая структура описания, поиска и интегрирования веб-сервисов в Интернет. Данная структура представляет собой средство регистрации вычислительных ресурсов в формализованном XML-формате;

SOAP — протокол обмена структурированными сообщениями в распределенной вычислительной среде. Если раскрыть аббревиатуру, получим *Simple Object Access Protocol*, то есть простой протокол доступа к объектам. SOAP используется для удаленного вызова процедур (функций), а также для обмена произвольными сообщениями в формате XML;

XML — расширенный язык разметки, представляющий собой текстовый формат, предназначенный для хранения структурированных данных и обмена информацией между программами. XML может эффективно работать с различными типами данных, и реальной альтернативы ему на сегодняшний день не существует.

Каждый компонент может играть либо лишь одну роль (быть, например, только пользователем сервиса), либо одновременно сразу несколько ролей (например, быть провайдером одних сервисов и пользователем других).

В ходе взаимодействия друг с другом компоненты сервис-ориентированной архитектуры выполняют следующие основные операции (рис. 2.):

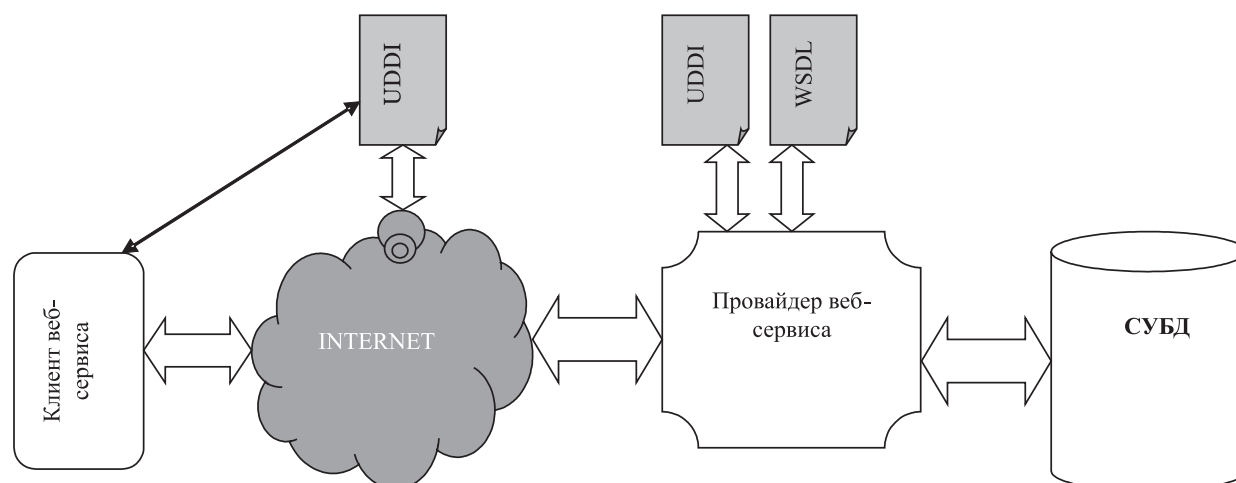


Рис. 1. Принцип работы веб-сервиса

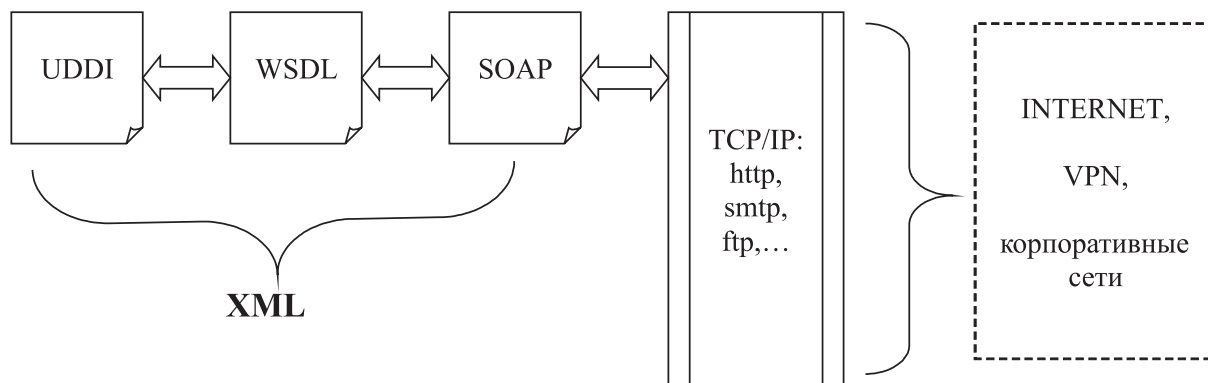


Рис. 2. Технология взаимодействия компонентов веб-сервисов

- публикация: для того, чтобы сервис был доступным (вызываемым) пользователям сервиса, необходимо сделать его интерфейс известным им;
- поиск: пользователь сервиса должен иметь возможность найти в реестре сервисов необходимый сервис, удовлетворяющий заданным критериям;
- связывание и вызов: после получения описания сервиса пользователь сервиса должен иметь возможность вызвать и использовать сервис в соответствии с его описанием.

Для удаленного взаимодействия с веб-сервисами используется *SOAP* [7]. Он обеспечивает взаимодействие распределенных систем независимо от объектной модели, операционной системы или языка программирования. Данные передаются в виде *XML* документов особого формата. Согласно определению *W3C*², веб-сервисы — это приложения, которые доступны по протоколам и являются стандартными для Интернет. Нет требования, чтобы веб-сервисы использовали какой-то определенный транспортный протокол. Спецификация *SOAP* определяет, каким образом связываются сообщения *SOAP* и транспортный протокол (рис. 3).

Рис. 3. Структура *SOAP* пакета

Как мы видим, веб-сервисы состоят из клиентской и серверной части, а часто и сами имеют клиентский и серверный модули. Дело в том, что современные приложения, ввиду их сложности, требуют наличия системы обработки ошибок, которые возникают при работе в интернет-сети. Данные системы обеспечивают стабильность работы приложения. Например, если клиент веб-сервиса перестает получать информационные пакеты от серверной части приложения, он может динамически изменять тайм-ауты ожидания сигнала, либо производить поиск другого источника информации. Таким образом, клиентская часть приложения уже перестает быть таковой фактически, а превращается в некое подобие серверной.

Этапы взаимодействия сервера и клиента будут выглядеть следующим образом:

- Клиентское приложение создает экземпляр объекта *SOAPClient*.
- *SOAPClient* читает файлы описания методов веб-сервиса (*WSDL*). Эти файлы могут храниться и на клиентской части приложения.
- Клиентское приложение, используя возможности позднего связывания методов объекта *SOAPClient*, вызывает метод сервиса. *SOAPClient* формирует пакет запроса (*SOAP Envelope*) и отправляет на сервер. Возможно использование любого транспортного протокола, но, как правило, используется *HTTP*.
- Пакет принимает серверное приложение *Listener* (может представлять собой *ISAPI* приложение или *ASP* страницу), создает объект *SOAPServer* и передает ему пакет запроса.
- *SOAP-Server* читает описание веб-сервиса, загружает описание и пакет запроса в *XML DOM* дерева.
- *SOAP-Server* вызывает метод объекта/приложения, реализующего сервис.
- Результаты выполнения метода или описание ошибки конвертируются объектом *SOAPServer* в пакет ответа и отправляются клиенту.
- Объект *SOAPClient* проводит разбор принятого пакета и возвращает клиентскому приложению результаты работы сервиса или описание возникшей ошибки.

² World Wide Web Consortium — организация, разрабатывающая и внедряющая технологические стандарты для всемирной сети Интернет

Помимо передачи данных, протокол *SOAP* предусматривает реакцию веб-сервиса на различные ошибки. Например на неверный вызов методов. Информация об ошибках в протоколе *SOAP* передается при помощи специальных дескрипторов.

В том случае, если необходима конфиденциальность обмена, возможно использование *Secure SOAP (SSOAP)*, в котором предусмотрено шифрование данных.

2. Особенности применения веб-сервисов в лингвистических системах

Проанализируем возможность применения технологии *CORBA* к задаче создания и использования лингвистических ресурсов.

В нашей модели лингвистическим ресурсом является интегрированный программный комплекс «Словники України», функционирующий на платформе операционной системы *Windows*, взаимодействовать с которым можно из локальной сети и через Интернет. Интерфейс программного взаимодействия через интернет-канал реализуется при помощи веб-сервиса, работающего на одной платформе с лингвистической системой.

Процедура доступа и взаимодействия с веб-сервисом согласно с технологией *SOA* одинакова для любых аппаратных платформ и операционных систем, что значительно упрощает процесс написания программ-клиентов.

Таким образом, веб-сервис опубликован в реестрах *UDDI*, что позволяет его найти и идентифицировать, имеет *WSDL*, в котором представлены функции программного интерфейса и теперь остается написать программу-клиент, используя подходящий инструментарий (*Perl, PHP, C#, Java, ...*).

Схема работы клиент-серверного модуля системы «Словники України» представлена на рис. 4.

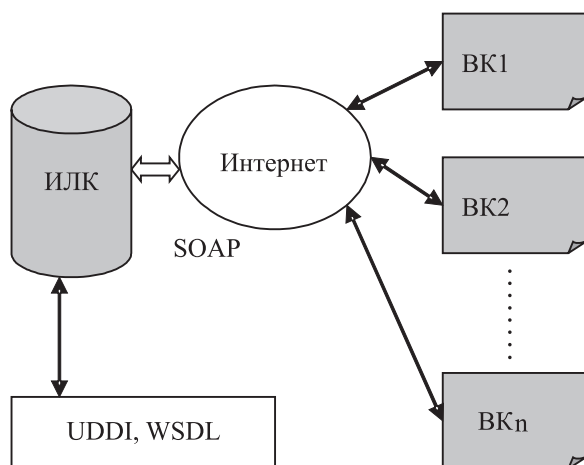


Рис. 4. Схема работы клиент-серверного модуля системы «Словники України»: ИЛК – интегрированный программный комплекс «Словники України»; BK1 – BK_n – множество веб-клиентов с различными операционными системами и аппаратной платформой

В результате мы получаем гибкий программный инструмент, позволяющий выполнять, например, задачи такого рода:

- выбор словаря;
- получение всего реестра словаря или его части;
- поиск по заданному слову;
- получение словарной статьи;
- доступ к любому слову словарной статьи;
- получение скалярных величин, касающихся того или иного словаря.

Выводы

Таким образом, проведенное исследование обосновывает необходимость использования технологии сервисно-ориентированной архитектуры *SOA* для создания лингвистических ресурсов и работы с ними. Преимущество веб-сервисов, разработанных с помощью *SOA*-технологии, дают возможность пользователям Интернет в режиме просмотра работать с реестром нескольких украинских словарей.

Список литературы. 1. Широков В.А. Элементы лексикографии. – К.: Довіра, 2005. – 304 с. 2. Костышин А. Использование сервис-ориентированной архитектуры и Web-сервисов для построения распределенных лексикографических сред: концепции и технологии // Прикладна лінгвістика та лінгвістичні технології: MegaLing-2006. – К.: Довіра, 2007. – С. 176-183. 3. Рабулець О.Г. Інтегровані лексикографічні системи: Автореф. дис. на здобуття наук. ступеня канд. техн. наук. – К., 2002. – 18 с. 4. Корпусна лінгвістика / В.А. Широков, О.В. Бугаков, Т.О. Грязнухіна та ін. – К.: Довіра, 2005. – 471 с. 5. *Common Object Request Broker Architecture*: <http://ru.wikipedia.org/wiki/CORBA>. 6. Фундамент для SOA: Инновации в технологиях и бизнесе. – Стратегия в области программного обеспечения IBM. – 2006. – № 1. 7. *World Wide Web Consortium, W3C*: <http://ru.wikipedia.org/wiki/W3C>.

Поступила в редколлегию 29.08.2008

УДК 004.75:004.451.82

Веб-сервіси в лінгвістичних інформаційних системах / О.Р. Єрошенко // Біоніка інтелекту: наук.-техн. журнал. – 2008. – № 2 (69). – С. 32-35.

У статті розглянуто принципи роботи веб-сервісу. Автором виділено основні компоненти сервіс-орієнтованої архітектури (SOA) і обґрунтована технологія взаємодії компонентів веб-сервісів. Проаналізована можливість застосування технології CORBA до задачі створення і використання лінгвістичних ресурсів.

Лл. 4. Бібліогр.: 7 найм.

УДК 004.75:004.451.82

Web-services in the Linguistic Informative Systems / A.R. Yeroshenko // Bionica intellecta: Sci. Mag. – 2008. – № 2 (69). – P. 32-35.

Principles of work of web-service are considered in the article. An author selects the basic components of Service-Oriented Architecture (SOA) and technology of co-operation of web-services components is grounded. Possibility of application of the CORBA-technology to the task of creations and uses of linguistic resources is analysed.

Fig. 4. Ref.: 7 items.

УДК 651.2:004.38



ПАРАМЕТРИЧНЕ УПРАВЛІННЯ ПРІОРИТЕТАМИ ОБСЛУГОВУВАННЯ WEB-ДОДАТКА СЕРВЕРОМ

А.Л. Єрохін, О.П. Турута

Харківський національний університет внутрішніх справ, м. Харків, Україна, ayerokhin@ukr.net
Харківський національний університет радіоелектроніки, м. Харків, Україна, science.ukraine@gmail.com

Розглядається вирішення задачі покращення обслуговування web-додатка в мережі Інтернет за допомогою параметричного управління пріоритетами обслуговування сервером. При цьому за рахунок використання механізмів пріоритетів сервера забезпечується можливість ізоляції web-додатка.

УПРАВЛІННЯ ПРІОРИТЕТАМИ, WEB-СЕРВЕР, ОБСЛУГОВУВАННЯ WEB-ДОДАТКА, ПАРАМЕТРИЧНЕ УПРАВЛІННЯ

1. Вступ

Сучасний Інтернет стає все більш складним та набуває яскраво виражених ознак інтелектуальної розподіленої системи. Користувачі звертаються до сучасних web-сервісів Інтернету як до системи, яка вирішує інтелектуальні задачі людини. Тому розроблення принципів, методів й архітектурних багатоагентних систем та засобів управління обслуговуванням є актуальною задачею.

Динамічний розвиток технологій розробки web-сайтів призводить до того, що прості web-сторінки замінюються web-додатками. Усе більше завдань виконується на стороні клієнта. Заміна архітектурних принципів проектування додатків викликає наступні проблеми:

- складність керування пріоритетом обслуговування процесів;
- визначення границь web-додатка;
- можливий вплив небажаного коду, іншого web-додатка;
- неможливість ізолювання програм у браузері одна від одної;
- виникнення передумов втручання в роботу web-додатка, підміна та аналіз даних, у тому числі в межах сесії.

Вказаними проблемами активно займається світова науково-технічна спільнота. Так, у роботі [1] представлені проблеми ізоляції web-додатків, у роботах [2,3,4] розглядаються моделі роботи web-сервісу і принципи проектування web-додатків, у роботі [5] описані методи формування web-додатка, критерії створення безпечних додатків охарактеризовані в [6, 7].

2. Постановка задачі

Виходячи з аналізу попередніх розробок, необхідно вирішити наступні задачі:

- 1) дослідити модель процесу авторизації користувача на основі існуючого підходу параметричного управління пріоритетами обслуговування сервером;
- 2) розробити метод покращення обслуговування web-додатка. Під час розробки методу необхідно

забезпечити можливість ізоляції web-додатка, для чого будемо використовувати механізми розподілу пріоритетів сервера;

3) розробити метод формування маркерів для захисту сесії web-додатка;

4) забезпечити захист даних автентифікації під час передачі неконтрольованими каналами зв'язку;

5) розробити спосіб розподілу загального пріоритету web-додатком на дочірні процеси.

3. Розподіл пріоритетів обслуговування агентів сервером

Процес авторизації агента відповідно до [6] виконується в три етапи (рис. 1):

1. Функції, дозволені агентом до ідентифікації (I) і вибір моменту ідентифікації (FIA_UID.1).

2. Функції, що реалізують ідентифікацію агента (FIA_UID.2), функції, дозволені до автентифікації (II), вибір моменту автентифікації (FIA_UAU.1).

3. Функції, що реалізують автентифікацію (FIA_UAU.2) і сам процес роботи користувача після авторизації (III).



Рис. 1. Схема авторизації агента

Клас (FIA) і підкласи (FIA_UAU, FIA_UID) відповідають стандарту [7].

Пропонується використовувати номер етапу авторизації користувача як ознаку для формування сервером пріоритету обслуговування агентів. Модифікуємо відому модель [8] параметричного керування пріоритетами обслуговування:

$$W_{t+1} = \begin{cases} W_t - 1, & A_t(I) = 1 \wedge A_t(I) = 2 \vee (\beta - D_t') * P_t < 0 \\ W_t, & A_t(I) = 2 \wedge (\beta \leq D_t' \vee A_t(I) = 3) \wedge \\ & \wedge (P_t > \alpha \vee A_t(I) = 1) \\ W_t + 1, & A_t(I) = 3 \vee (\beta - D_t') > 0, \end{cases} \quad (1)$$

де $A_i(t)$ — функція авторизації агента сесії i у момент часу t ; D_i^t — значення функції довіри агенту сесії i у момент часу t ; α — граничні значення продуктивності сервера; P_i — стан сервера в момент часу t ; β — нормоване значення рівня довіри, а різниця $\beta - D_i^t$ — визначає знак виразу.

Розробка функції D детально буде розглядатися в наступній роботі, а тут визначимо, що остаточна повна оцінка довіри агенту з боку системи визначається успішністю автентифікації. Однак існує багато видів атак на процес автентифікації з метою спрацювання в системі помилки I-го роду, коли доступ до системи отримає той, хто такого права не має, або спрацювання в системі помилки II-го роду, коли доступ до системи не зможе отримати той, хто має право на це та намагається його отримати. Необхідно зменшувати ймовірність виникнення помилок I-го та II-го роду відповідно P_I , P_{II} . Однак посилення правил політики безпеки призводить до збільшення ймовірності P_I , послаблення правил безпеки призводить до збільшення ймовірності P_{II} , а використання функції D оцінки довіри агенту дозволить покращити відношення $\frac{P_I}{P_{II}}$. D визначається від P^1 , P^2 , де P^1 — ознаки агента; P^2 — ознаки стану системи.

4. Метод формування маркера

Для забезпечення надійної роботи сесії на етапі III необхідно сформулювати множину маркерів — M^l для сесії l , де елемент M_k^l ($M_k^l \subset M$, $k=1, M(M^l)$) — маркер k -го запиту в сесії l .

Ідентифікатор агента є несекретним та може передаватися відкрито, проте дані автентифікації є секретними, тому передача їх неконтрольованими каналами зв'язку є небажаною. Наприклад компрометація біометричних даних є проблемою для подальшої роботи системи захисту [6].

В [9] відзначено високу ефективність використання одноразового блокноту, тому для реалізації автентифікації і захисту сесії пропонується сформувати множину маркерів M^l . Кожен маркер, як і шифрограми одноразового блокноту, буде використаний один раз.

На рис. 2 наведено схему проведення безпечної автентифікації та формування множини M^l .

Послідовність SA генерується сервером з урахуванням ознаки клієнта, сесії та випадкової величини. Номер SN вказує, який елемент множини M^l використовується для формування відповіді SB. Відмітимо, що з послідовностей SA, SB математично важко отримати образ даних автентифікації (паролі) — S. Клієнт знає (володіє) даними авторизації, web-додаток за відомим алгоритмом знаходить S-образ, а сервер зберігає замість даних автентифікації його S-образ.

Розглянемо метод формування множини маркерів M^l для сесії l .

Нехай SN — кількість маркерів, яку необхідно сформувати, значення SN вибирається випадково в інтервалі $L2 \geq L1 \geq SN$. У [10] вважається для Інтернет-сесій достатнім $L1=500$ та $L2=1000$.

Використаємо послідовність SA та образ даних автентифікації S для формування останнього елемента множини

$$M^l : M_{L2}^l = \text{shal}(SA + S).$$

Кожний попередній елемент знаходиться за формулою

$$M_j^l = \text{shal}(M_{j+1}^l),$$

де $j = L2-1, 1$.

На сервері випадковим чином вибирається SN ($L1 \geq SN$), формується послідовність SA, далі формується множина M^l . M_{SN}^l зберігається у сесії l , а клієнту відправляється SA, SN.

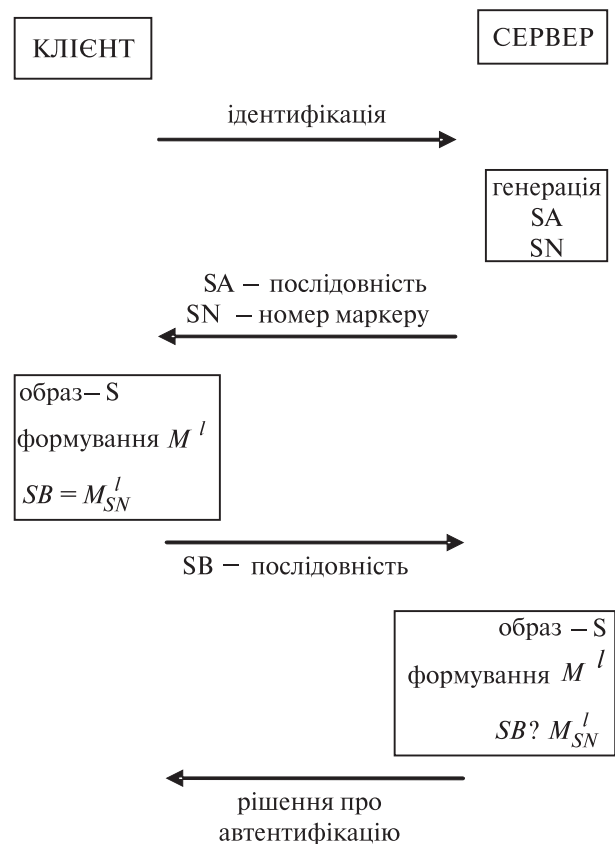


Рис. 2. Схема проведення безпечної автентифікації

Web-додаток аналогічно формує множину M^l , та зберігає перші SN елементів, серверу повертає SB ($SB = M_{SN}^l$). Якщо клієнт і сервер мають однакові образи S, то $SB = SB'$, тобто автентифікація пройшла успішно, в іншому випадку — помилка.

Відмітимо, що на сервері немає необхідності зберігати всю множину маркерів M^l , достатньо зберігати останній успішно переданий маркер. Автентичність наступних маркерів перевіряється відповідно до способу побудови множини M^l : $M_j^l = \text{shal}(M_{j+1}^l)$. За допомогою величини параметра SN-сервер може задавати довжину сесії, для якої формується одноразовий блокнот.

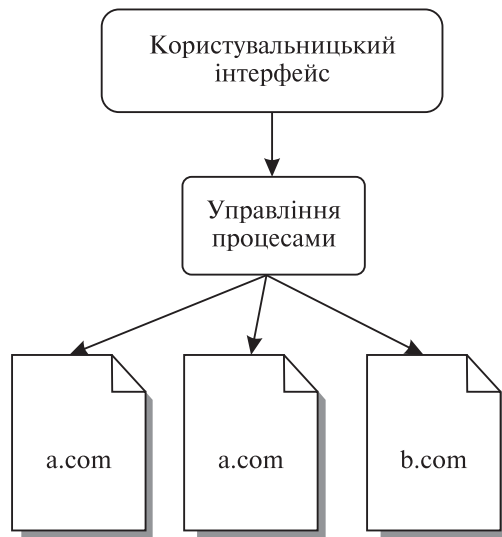
5. Особливості методу управління пріоритетами під час проектування web-додатків

При проектуванні клієнтської частини [2,3] необхідно відзначити тенденцію заміщення статичних web-сторінок динамічними web-додатками, зростанням пропорції активного коду [1]. Отже, при проектуванні web-додатків виникають наступні проблеми:

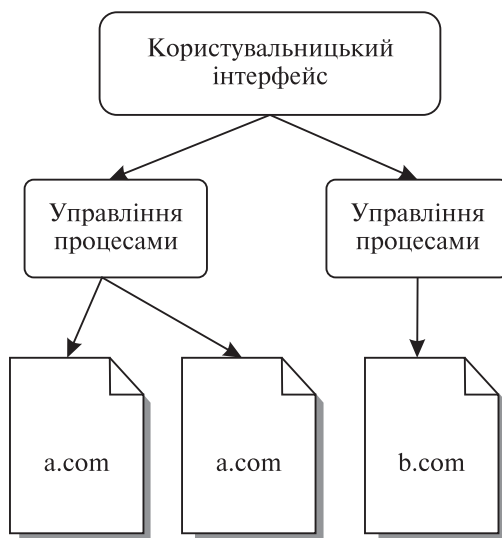
- проблема перехресного впливу одних web-додатків на інші;
- розподіл обчислювальних ресурсів.

Розглянемо схему обслуговування браузером завантажених документів, які в загальному випадку браузер обробляє за схемою, представленою на рис. 3а.

При ізоляції додатків [1] обробка відбувається за схемою, що представлена на рис. 3б.



а



б

Рис. 3. Схеми обслуговування процесів браузером: а – схема обслуговування процесів звичайним браузером; б – схема обслуговування процесів браузером з функцією ізоляції процесу

У роботах [4,5] розглядаються способи формування вмісту додатка, наповнення різним контентом. Web-додаток обробляє декілька процесів, браузер обслуговує кілька web-додатків.

На рис. 3б показано браузер з функцією ізоляції; у такому випадку web-додаток може управляти деякими процесами.

Пропонується використовувати маркер M_k^l , який одержується від сервера, де l – ідентифікатор сесії, у якій працює агент, а k – номер запиту в групі запитів сесії l , $k = \overline{SN, LI}$.

6. Обслуговування ідентифікованих агентів

Відзначимо, що для ряду сервісів тривалість роботи агентів на етапі III найбільша, отже, функція (1) не може управляти пріоритетом доступу, тому пропонується метод для параметричного управління обслуговуванням сервера з урахуванням маркера M_k^l .

Для кожного запитуваного ресурсу оцінимо пріоритет обслуговування за формулою

$$W_{t+1}^{M_j^l} = \begin{cases} W_t^l + 1, & \text{для } j = k \\ W_t^l * (1 - \frac{1}{k-1}), & \text{для } j < k, \end{cases} \quad (2)$$

де $W_t^l = \frac{\sum_{i=1}^k W_t^{M_i^l}}{k}$ – середній пріоритет для сесії;

$M_j^l \in M$; l – ідентифікатор сесії; $i = \overline{SN, LI}$.

Таким чином, забезпечується можливість ідентифікації потоків одного агента та підвищення пріоритету обслуговування активної сторінки й зниження пріоритетів сторінок, які обслуговуються у фоновому режимі.

Висновки

У роботі проаналізовано стандартний метод параметричного розподілу пріоритетів обслуговування агентів в мережі Інтернет.

Запропоновано спосіб інтелектуального управління завантаженням web-додатка шляхом виділення пріоритету між своїми процесами.

Запропоновано метод захисту даних автентифікації та формування множини маркерів для ідентифікації web-додатка протягом сесії.

Удосконалено метод параметричного управління обслуговуванням web-додатків, який дозволяє інтелектуально змінювати параметри обслуговування в залежності від завантаження сервера й пріоритету потоку даних, а також етапу автентифікації.

Список літератури: 1. Reis C., Bershad B., Gribble S., Levy H. Using Processes to Improve the Reliability of Browser-based Applications // University of Washington Technical Report UW-CSE-2007-12-01. 2. Быков М.Ю. Модель процесса обработки запросов системой управления web-сайтами

// Вычислительные методы и программирование. — М. — 2005. Т. 6. Раздел 2. С. 52-56. 3. *Алиев Р.Т.* Методы управления трафиком в мультисервисных сетях // Научно-технический вестник СПбГИТМО (ТУ). Выпуск 6. Информационные, вычислительные и управляющие системы. — СПб.: СПбГИТМО(ТУ), 2002. С. 10–13. 4. *Kappel G., Proll B., Retschitzegger W., Schwinger W., Hofer T.*, Modeling Ubiquitous Web Application — A Comparison of Approaches // Journal of Network and Computer Application — 2001. 5. *Christos B., Giorgos K., Ioannis M.* A web content manipulation technique based on page Fragmentation // Journal of Network and Computer Application, 2006. 6. *Галатенко В.А.* Стандарты информационной безопасности / Под ред. Бетенина В.Б. — М.: Интернет-университет Информационных технологий, 2006. — 208с. 7. ГОСТ Р ИСО/МЭК 15408-2-2002. Информационная технология. Методы и средства обеспечения безопасности. Критерии оценки безопасности информационных систем. Ч.2. Функциональные требования. Введен впервые. — М. Узд. стандартов, 2002. — 160 с. 8. *Коваленко А.А.* Методы та засоби підвищення оперативності передачі даних у мультисервісних мережах: Автореф. дис. на здобуття наук. ступеня канд. техн. наук // Харківський національний університет радіоелектроніки. —Харків. — 2008. — 19 с. 9. *Шеннон К.* Работы по теории информации и кибернетике. — М.: ИЛ, 1963. 10. *Полянская О.Ю.* Инфраструктуры

открытых ключей [электронный ресурс] — режим доступа <http://www.intuit.ru/departments/security/pki/>.

Поступила в редколлегию 2.09.2008

УДК 651.2:004.38

Параметрическое управление приоритетами обслуживания web-приложения сервером / А.Л.Ерохин, А.П.Турута // Бионика интеллекта: научн.-техн. журнал. — 2008. — № 2(69). — С. 36-39.

Рассматривается подход к управлению загрузкой web-приложения путем выделения приоритетов между собственными процессами. Усовершенствован метод параметрического управления обслуживанием web-приложений, который изменяет параметры обслуживания в зависимости от загрузки сервера, приоритетов потоков данных и особенностей аутентификации.

Рис.: 3. Библиогр.: 10 назв.

UDC 651.2:004.38

Light devices for visualization of traffic management / A.L.Yerokhin, O.P.Turuta // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 36-39.

The approach to management loading web-applications by separations priority between own process is considered. The method of parametric management of web-applications service is improved, it changes the parameters of the service depending on loading the server, priority dataflow and particularities to authentications.

Fig.:3. Ref.: 10 items.

УДК 004.032 26



АВТОАССОЦИАТИВНАЯ МНОГОСЛОЙНАЯ НЕЙРОННАЯ СЕТЬ И АЛГОРИТМ ЕЕ ОБУЧЕНИЯ ПРИ СЖАТИИ ИЗОБРАЖЕНИЙ

В.П. Машталир¹, А.Е. Путятин², Е.С. Сакало³

¹ХНУРЭ, г. Харьков, Украина, mashtalir@kture.kharkov.ua

^{2,3}ХНУРЭ, г. Харьков, Украина

Рассмотрена задача понижения размерности вектора признаков, часто встречающаяся при обработке изображений, с помощью гетерогенной автоассоциативной четырехслойной нейронной сети с прямой передачей информации. Предложен алгоритм ее обучения, обладающий улучшенными характеристиками по сравнению с традиционными градиентными процедурами обратного распространения ошибок.

НЕЙРОННАЯ СЕТЬ, СЖАТИЕ ИЗОБРАЖЕНИЙ, АНАЛИЗ ГЛАВНЫХ КОМПОНЕНТ

Введение

При решении широкого класса задач, связанных с обработкой сигналов высокой размерности и, прежде всего, обработкой изображений, достаточно часто возникает необходимость понижения размерности исходных векторов-образов с минимальной потерей информации. Наиболее часто для этого используется аппарат анализа главных компонент и главных подпространств, основанный на линейном преобразовании Карунена-Лоева. Вместе с тем, линейная техника не всегда способна выявить более сложные соотношения, присущие реальным изображениям. В ряде случаев более эффективным представляется нелинейный подход, основанный на нейросетевых технологиях [1].

Так, в [2] была рассмотрена задача сжатия информации с помощью двухслойной нейронной сети с прямой передачей информации и показано, что независимо от вида используемых активационных функций, такая архитектура реализует стандартный анализ главных компонент. В [1] предложена трехслойная автоассоциативная нейронная сеть обратного распространения и описан эксперимент, связанный с обработкой изображений, однако, сравнительный анализ с другими технологиями не проводился. В [3–5] показано, что оптимальное сжатие может быть обеспечено с помощью четырехслойной автоассоциативной нейронной сети с чередующимися линейными и нелинейными слоями и приведены примеры обработки изображений. Данная сеть реализует нелинейный анализ главных компонент (NLPCA) и обучается с помощью стандартного градиентного алгоритма распространения ошибок. Этот алгоритм, однако, обладает низкой скоростью сходимости, не учитывает неоднородности архитектуры сети и весьма чувствителен к воздействию различного рода возмущений.

В связи с этим представляется целесообразным оптимизировать процесс обучения четырехслойной гетерогенной автоассоциативной нейронной сети с прямой передачей информации, предназначенной для понижения размерности обрабатываемых векторов-образов, с учетом влияния помех и возможным нарушением выпуклости целевых функций скрытых слоев.

1. Архитектура нейронной сети

Архитектура рассматриваемой автоассоциативной нейронной сети, предназначенной для понижения размерности пространства исходных образов, приведена на рис. 1 и содержит четыре последовательно соединенных слоя нейронов.

На вход сети (рецепторный нулевой слой) подается последовательность векторов образов

$$x(1), x(2), \dots, x(k), \dots, x(N), \\ x(k) = (x_1(k), \dots, x_i(k), \dots, x_n(k))^T \in R^N,$$

предварительно пронормированных на гиперкубе так, что $x_i(k) \in [-1, 1], i = 1, 2, \dots, n$.

Первый скрытый слой содержит $h \geq n$ нейронов — элементарных персептронов Розенблатта с сигмоидальными активационными функциями. В случае отсутствия сигнала смещения первый скрытый слой содержит hn настраиваемых синаптических весов и описываемых соотношениями вида:

$$O_j^{[1]}(k) = \psi_j^{[1]}(u_j^{[1]}(k)) = \psi_j^{[1]}(\sum_{i=1}^n w_{ji}^{[1]} x_i(k)), j = 1, 2, \dots, h, \\ O^{[1]}(k) = \Psi^{[1]}(W^{[1]} x(k)),$$

где $O_j^{[1]}(k)$ — выходной сигнал j -го нейрона первого скрытого слоя $x(k)$; $\psi_j^{[1]}(\bullet)$ — активационная функция j -го нейрона первого скрытого слоя; $u_j^{[1]}(k)$ — сигнал внутренней активации j -го нейрона первого скрытого слоя; $w_{ji}^{[1]}$ — синаптический вес i -го входа j -го нейрона первого скрытого слоя; $O^{[1]}(k) = (O_1^{[1]}(k), \dots, O_j^{[1]}(k), \dots, O_h^{[1]}(k))^T$, $\Psi^{[1]} = \text{diag}(\Psi_1^{[1]}, \dots, \Psi_j^{[1]}, \dots, \Psi_h^{[1]})$ — $(h \times h)$ — диагональная матрица активационных функций, $W^{[1]} = \{w_{ji}^{[1]}\}$ — $(h \times n)$ — матрица синаптических весов.

Если нейроны первого скрытого слоя включают сигнал смещения, то приведенные выше соотношения принимают вид:

$$O_j^{[1]}(k) = \psi_j^{[1]}(u_j^{[1]}(k)) = \psi_j^{[1]}(\sum_{i=1}^n w_{ji}^{[1]} x_i(k) + \Theta_j^{[1]}) = \\ = \psi_j^{[1]}(\sum_{i=0}^n w_{ji}^{[1]} x_i(k)), \\ O^{[1]}(k) = \Psi^{[1]}(W^{[1]} x(k)),$$

где $\Theta_j^{[1]}$ – сигнал смещения j -го нейрона первого скрытого слоя; $w_{j0}^{[1]} = \Theta_j^{[1]}$; $x_0(k) = 1, W^{[1]} - (h \times n + 1)$ – матрица синаптических весов.

$x(k) = (1, x_1(k), \dots, x_n(k))^T$. В этом случае слой содержит $h(n+1)$ настраиваемых параметров.

В качестве активационных функций используются либо логистическая функция

$$\psi_j^{[1]}(u_j^{[1]}) = \frac{1}{1 + e^{-\gamma_j u_j^{[1]}},$$

либо гиперболический тангенс

$$\psi_j^{[1]}(u_j^{[1]}) = \tanh(\gamma_j u_j^{[1]}) = \frac{1 - e^{-2\gamma_j u_j^{[1]}}}{1 + e^{-2\gamma_j u_j^{[1]}}},$$

где γ_j – положительный параметр (возможно настраиваемый), задающий «крутизну» функции. Полезно также ввести в рассмотрение производные логистической функции

$$\frac{\partial \psi_j^{[1]}(u_j^{[1]})}{\partial u_j^{[1]}} = \gamma_j \psi_j^{[1]}(u_j^{[1]}) (1 - \psi_j^{[1]}(u_j^{[1]}))$$

и гиперболического тангенса

$$\frac{\partial \psi_j^{[1]}(u_j^{[1]})}{\partial u_j^{[1]}} = \gamma_j (1 - (\psi_j^{[1]}(u_j^{[1]}))^2)$$

соответственно.

Второй скрытый слой содержит $m < n$ нейронов – адаптивных линейных ассоциаторов, а его выходной сигнал $y = (y_1, \dots, y_m)^T$ является выходом

нейронной сети в целом и представляет собой «сжатый» входной образ x .

В отсутствие смещения второй скрытый слой описывается соотношениями:

$$y_j(k) = \sum_{i=1}^h w_{ji}^{[2]} o_i^{[1]}(k), j = 1, 2, \dots, m,$$

$$y(k) = W^{[2]} O^{[1]}(k),$$

где $W^{[2]} = \{w_{ji}^{[2]}\} - (m \times h)$ – матрица синаптических весов.

Если же нейроны второго скрытого слоя включают сигнал смещения, то приведенные выше соотношения принимают вид:

$$y_j(k) = \sum_{i=1}^h w_{ji}^{[2]} o_i^{[1]}(k) + \Theta_j^{[2]} = \sum_{i=0}^h w_{ji}^{[2]} o_i^{[1]}(k)$$

$$y(k) = W^{[2]} O^{[1]}(k),$$

где $W^{[2]} = \{w_{ji}^{[2]}\} - (m \times (h + 1))$ – матрица синаптических весов: $O^{[1]}(k) = (1, O_1^{[1]}(k), \dots, O_h^{[1]}(k))^T$.

Третий скрытый слой подобен первому и также содержит h элементарных перцептронов Розенблатта. В отсутствие смещения этот слой описывается соотношениями:

$$O_j^{[3]}(k) = \psi_j^{[3]}(u_j^{[3]}(k)) = \psi_j^{[3]}(\sum_{i=1}^m w_{ji}^{[3]} y_i(k)), j = 1, 2, \dots, h,$$

$$O^{[3]}(k) = \Psi^{[3]}(W^{[3]} y(k)),$$

где $W^{[3]} - (h \times m)$ – матрица синаптических весов.

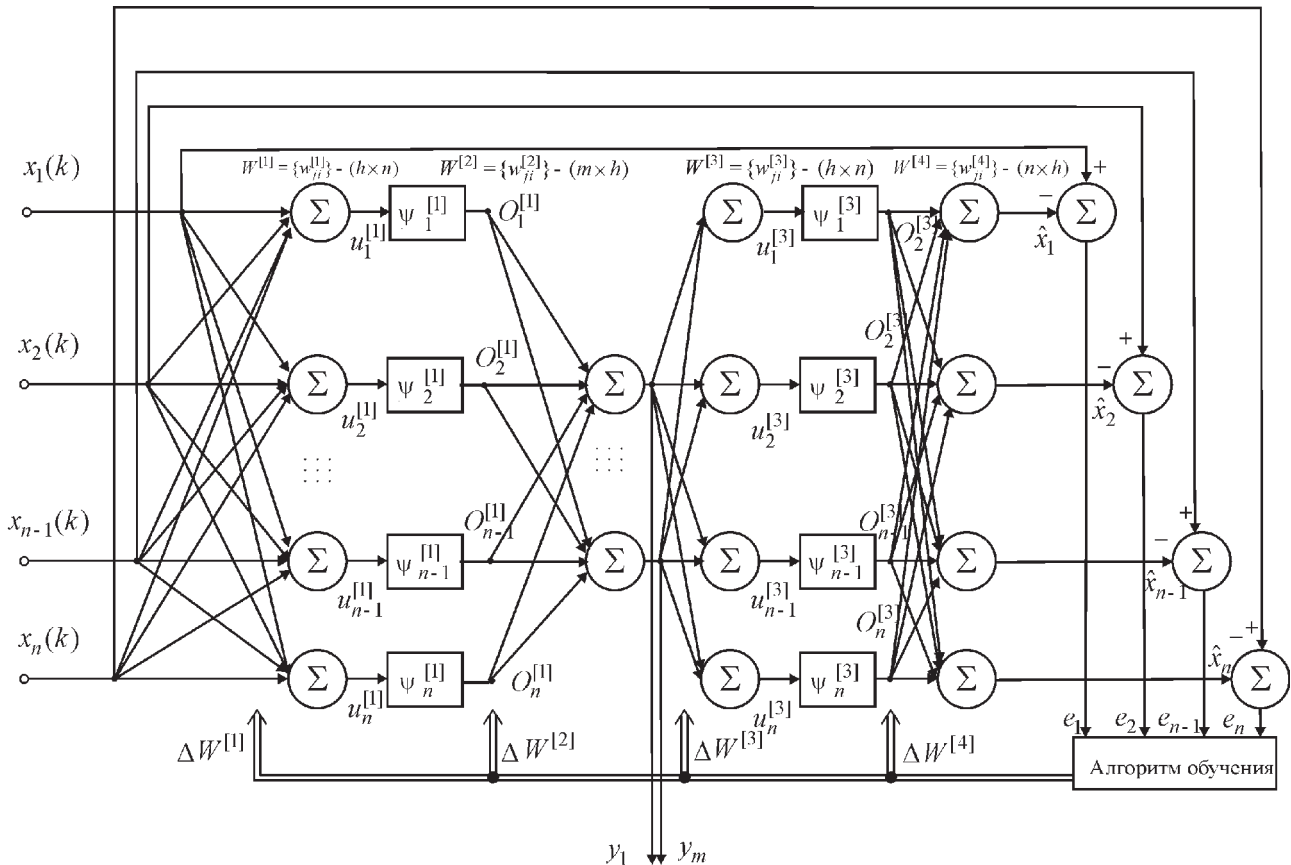


Рис. 1. Автоассоциативная многослойная нейронная сеть

Если же нейроны третьего скрытого слоя включают сигнал смещения, то приведенные выше соотношения имеют вид:

$$\begin{aligned} O_j^{[3]}(k) &= \Psi_j^{[3]}(u_j^{[3]}(k)) = \\ &= \Psi_j^{[3]}(\sum_{i=1}^m w_{ji}^{[3]} y_i(k) + \Theta_j^{[3]}), j=1,2,\dots,h, \\ O^{[3]}(k) &= \Psi^{[3]}(W^{[3]}y(k)), \end{aligned}$$

где $W^{[3]} - (h \times (m+1))$ – матрица синаптических весов.

Четвертый выходной слой подобен второму скрытому и содержит n нейронов – адаптивных линейных ассоциаторов. В отсутствие смещения четвертый слой описывается соотношениями:

$$\begin{aligned} \hat{x}_j(k) &= \sum_{i=1}^h w_{ji}^{[4]} o_i^{[3]}(k), j=1,2,\dots,n, \\ \hat{x}(k) &= W^{[4]} O^{[3]}(k), \end{aligned}$$

где $\hat{x}(k) - (n \times 1)$ – вектор, являющийся оценкой входного сигнала $x(k)$, восстановленного после сжатия; $W^{[4]} - (n \times h)$ – матрица синаптических весов.

Если адаптивные линейные ассоциаторы четвертого слоя включают сигнал смещения, то приведенные выше соотношения имеют вид:

$$\begin{aligned} \hat{x}_j(k) &= \sum_{i=1}^h w_{ji}^{[4]} o_i^{[3]}(k) + \Theta_j^{[4]} = \sum_{i=1}^h w_{ji}^{[4]} o_i^{[3]}(k), o_0^{[3]}(k) = 1, \\ \hat{x}(k) &= W^{[4]} O^{[3]}(k), \end{aligned}$$

где $W^{[4]} - (n \times (h+1))$ – матрица синаптических весов.

Таким образом, отображение, реализуемое четырехслойной автоассоциативной нейронной сетью, приведенной на рис. 1, имеет вид:

$$\hat{x}(k) = W^{[4]}(\Psi^{[3]}(W^{[3]}W^{[2]}\Psi^{[1]}(W^{[1]}x(k)))).$$

2. Обучение нейронной сети

В качестве критерия обучения будем использовать стандартную квадратичную функцию

$$\begin{aligned} E(k) &= \sum_{j=1}^n E_j(k) = \frac{1}{2} \sum_{j=1}^n (x_j(k) - \hat{x}_j(k))^2 = \\ &= \frac{1}{2} \sum_{j=1}^n e_j^2(k) = \frac{1}{2} \|e(k)\|^2 = \frac{1}{2} \|x(k) - \hat{x}(k)\|^2, \end{aligned}$$

где $e(k) = (e_1(k), \dots, e_n(k))^T$ – вектор ошибок восстановления (обучения) входного сигнала по всем координатам.

Сам же процесс обучения будем проводить на основе градиентной процедуры вида:

$$w_{ji}^{[s]}(k+1) = w_{ji}^{[s]}(k) - \eta^{[s]}(k) \frac{\partial E(k)}{\partial w_{ji}^{[s]}}, s=1,2,3,4,$$

или

$$\Delta w_{ji}^{[s]} = w_{ji}^{[s]}(k+1) - w_{ji}^{[s]}(k) = -\eta^{[s]}(k) \frac{\partial E(k)}{\partial w_{ji}^{[s]}},$$

где $\eta^{[s]}(k)$ – параметр шага обучения s -го слоя, определяющий скорость процесса настройки.

Следуя концепции обратного распространения ошибок, обучение сети начинается с четвертого (выходного) слоя. При этом:

$$\begin{aligned} \Delta w_{ji}^{[4]} &= -\eta^{[4]}(k) \frac{\partial E(k)}{\partial w_{ji}^{[4]}} = -\eta^{[4]}(k) \frac{\partial E_j(k)}{\partial \hat{x}_j} \cdot \frac{\partial \hat{x}_j}{\partial w_{ji}^{[4]}} = \\ &= \eta^{[4]}(k) e_j(k) O_i^{[3]}(k). \end{aligned}$$

Поскольку выходной сигнал слоя $\hat{x}(k)$ линейно зависит от настраиваемых параметров, для обучения можно использовать либо оптимальный по быстродействию алгоритм Качмажа-Уидроу-Коффа [6-8]:

$$\Delta w_{ji}^{[4]} = \frac{e_j(k) O_i^{[3]}(k)}{\|O^{[3]}(k)\|^2},$$

в векторной форме, имеющий вид:

$$\Delta w_j^{[4]} = \frac{e_j(k) O^{[3]}(k)}{\|O^{[3]}(k)\|^2}, j=1,2,\dots,n,$$

либо, если требуется дополнительная фильтрация возмущений и сглаживание, процедуру вида [9]:

$$\begin{cases} \Delta w_j^{[4]} = (p^{[4]}(k))^{-1} e_j(k) O^{[3]}(k), \\ p^{[4]}(k) = \alpha p^{[4]}(k-1) + \|O^{[3]}(k)\|^2, \end{cases} \quad (1)$$

где $0 \leq \alpha \leq 1$ – параметр сглаживания.

Обучение параметров третьего скрытого слоя производится с помощью алгоритма

$$\begin{aligned} \Delta w_{ji}^{[3]} &= -\eta^{[3]}(k) \frac{\partial E(k)}{\partial w_{ji}^{[3]}} = -\eta^{[3]}(k) \frac{\partial E(k)}{\partial u_j^{[3]}} \cdot \frac{\partial u_j^{[3]}}{\partial w_{ji}^{[3]}} = \\ &= \eta^{[3]}(k) \delta_j^{[3]}(k) y_i(k), \end{aligned} \quad (2)$$

где $\delta_j^{[3]}(k) = \frac{\partial E(k)}{\partial u_j^{[3]}}, j=1,2,\dots,h$ так называемая начальная ошибка (δ -ошибка) третьего скрытого слоя, которая также может быть представлена в форме

$$\delta_j^{[3]}(k) = -\frac{\partial E(k)}{\partial O_j^{[3]}(k)} \cdot \frac{\partial O_j^{[3]}(k)}{\partial u_j^{[3]}}$$

с учетом того, что

$$\partial O_j^{[3]}(k) = \Psi_j^{[3]}(u_j^{[3]}(k))$$

логично записать

$$\delta_j^{[3]}(k) = -\frac{\partial E(k)}{\partial O_j^{[3]}(k)} \cdot \frac{\Psi_j^{[3]}(u_j^{[3]}(k))}{\partial u_j^{[3]}},$$

а представив $-\frac{\partial E(k)}{\partial O_j^{[3]}(k)}$ в виде

$$\begin{aligned} -\frac{\partial E(k)}{\partial O_j^{[3]}(k)} &= -\sum_{i=1}^n \frac{\partial E(k)}{\partial \hat{x}_i(k)} \cdot \frac{\partial \hat{x}_i(k)}{\partial O_j^{[3]}} = \\ &= \sum_{i=1}^n \left(-\frac{\partial E(k)}{\partial \hat{x}_i(k)} \right) \cdot \frac{\partial}{\partial O_j^{[3]}} \left(\sum_{p=1}^h w_{ip}^{[4]}(k) O_p^{[3]}(k) \right) = \\ &= \sum_{i=1}^n e_i(k) \cdot \frac{\partial}{\partial O_j^{[3]}} \left(\sum_{p=1}^h w_{ip}^{[4]}(k) O_p^{[3]}(k) \right) = \sum_{i=1}^n e_i(k) w_{ij}^{[4]}(k), \end{aligned}$$

получить окончательное выражение:

$$\delta_j^{[3]}(k) = \frac{\partial \psi_j^{[3]}(u_j^{[3]}(k))}{\partial u_j^{[3]}} \sum_{i=1}^n e_i(k) w_{ij}^{[4]}(k).$$

Тогда процедура (2) приобретает форму:

$$\begin{aligned} \Delta w_{ji}^{[3]} &= \eta^{[3]}(k) \delta_j^{[3]}(k) y_i(k) = \\ &= \eta^{[3]}(k) y_i(k) \frac{\partial \psi_j^{[3]}(u_j^{[3]}(k))}{\partial u_j^{[3]}} \sum_{i=1}^n e_i(k) w_{ij}^{[4]}(k), \end{aligned}$$

или в векторном виде:

$$\Delta w_j^{[3]} = \eta^{[3]}(k) \delta_j^{[3]}(k) y(k). \quad (3)$$

Свойства общей процедуры (3) определяются выбором параметра $\eta^{[3]}(k)$, от обоснованного выбора которого зависит время обучения в целом.

Введем в рассмотрение модификацию регуляризованного алгоритма обучения [10, 11] вида:

$$\Delta w_j^{[3]} = \frac{\eta^{[3]}(k) \delta_j^{[3]}(k) y(k) + \rho^{[3]} \Delta w_j^{[3]}(k-1)}{\|y(k)\|^2}, \quad (4)$$

где $\eta^{[3]} = \text{const} > 0, 0 \leq \rho < 1$ — параметр регуляризации.

Несложно заметить, что при $\rho^{[3]} = 0$ процедура приобретает форму нелинейной модификации мультипликативного алгоритма Качмажа, характеризующегося высоким быстродействием. Если же требуется дополнительная фильтрация, можно от процедуры (4) перейти к процедуре общего вида:

$$\begin{cases} \Delta w_j^{[3]} = (p^{[3]}(k))^{-1} (\eta^{[3]}(k) \delta_j^{[3]}(k) y(k) + \Omega_3), \\ p^{[3]}(k) = \alpha p^{[3]}(k-1) + \|y(k)\|^2, \end{cases} \quad (5)$$

где $\Omega_3 = \rho^{[3]} \Delta w_j^{[3]}(k-1)$, объединяющей в себе быстродействие алгоритма Качмажа-Уидроу-Коффа, сглаживающие свойства алгоритма, введенного в [12], и возможность проходить «плоские» участки целевой функции присущие регуляризованным алгоритмам с $\rho^{[3]} > 0$.

Аналогичным образом можно записать алгоритм обучения второго скрытого слоя в виде:

$$\Delta w_{ji}^{[2]} = \eta^{[2]}(k) \delta_j^{[2]}(k) O_i^{[1]}(k),$$

где

$$\delta_j^{[2]}(k) = \sum_{i=1}^h \delta_i^{[3]}(k) w_{ij}^{[3]}(k).$$

Окончательно же:

$$\begin{cases} \Delta w_j^{[2]} = (p^{[2]}(k))^{-1} (\eta^{[2]}(k) \delta_j^{[2]}(k) O^{[1]}(k) + \Omega_2), \\ p^{[2]}(k) = \alpha p^{[2]}(k-1) + \|O^{[1]}(k)\|^2, \end{cases} \quad (6)$$

где $\Omega_2 = \rho^{[2]} \Delta w_j^{[2]}(k-1)$.

И, наконец, для первого скрытого слоя справедливы соотношения:

$$\Delta w_{ji}^{[1]} = \eta^{[1]}(k) \delta_j^{[1]}(k) x_i(k),$$

$$\delta_j^{[1]}(k) = \frac{\partial \psi_j^{[1]}(u_j^{[1]}(k))}{\partial u_j^{[1]}} \sum_{i=1}^h \delta_i^{[2]}(k) w_{ij}^{[2]}(k).$$

$$\begin{cases} \Delta w_j^{[1]} = (p^{[1]}(k))^{-1} (\eta^{[1]}(k) \delta_j^{[1]}(k) x(k) + \Omega_1), \\ p^{[1]}(k) = \alpha p^{[1]}(k-1) + \|x(k)\|^2, \end{cases} \quad (7)$$

где $\Omega_1 = \rho^{[1]} \Delta w_j^{[1]}(k-1)$.

Таким образом, процесс обучения нейронной сети, показанной на рис. 1, сводится к последовательному использованию процедур (1), (5), (6) и (7) для каждого предъявляемого вектора-образа $x(k)$. Для достаточно длинной обучающей выборки процесс настройки может быть реализован в реальном времени. Если же объем исходных данных ограничен, обучение может быть организовано по эпохам путем многократного «прогона» данных через нейронную сеть.

3. Результаты эксперимента

Для анализа качества компрессии изображений, получаемых при помощи построенной нейронной сети, был взят критерий, основанный на среднеквадратичной ошибке значений пикселей (mean square error— MSE):

$$d(x, y) = \sqrt{\sum_{i,j=1}^n ((x_{i,j} - y_{i,j})^2 / n^2)}. \quad (8)$$

Он подразумевает сравнение исходного изображения и восстановленного после сжатия. Так как предложенный метод сжатия следует отнести к сжатию с потерями, то естественно желание выяснить степень потерь. Однако, следует отметить, что данный критерий дает правильные результаты при отсутствии шума на изображении. При наличии же шума получим серьезные ошибки для данного критерия. В такой ситуации лучше использовать критерий отношения пикового значения сигнала к шуму (PSNR — peak signal-to-noise ratio), который является более устойчивым к подобным воздействиям.

В результате сравнения исходных изображений и после применения предлагаемого метода сжатия, использующего представленную четырехслойную нейронную сеть на основе критерия (8) получены результаты, представленные на рис. 2.

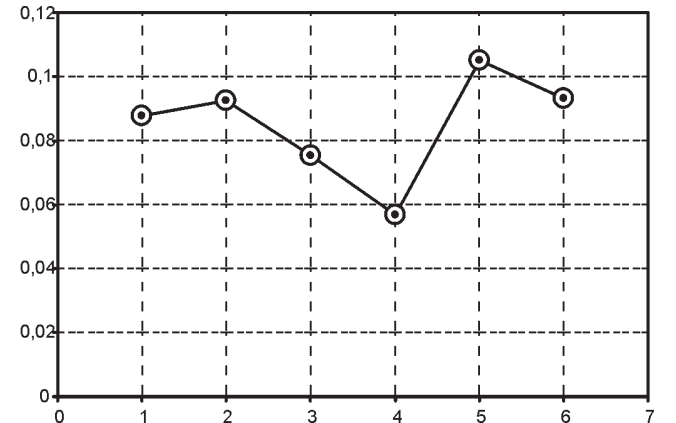


Рис. 2. Среднеквадратическое отклонение пикселей изображений, полученное при сжатии

Таким образом, полученные результаты свидетельствуют о том, что приведенный метод сжатия приводит к потерям информации в районе 8-9%, что для методов сжатия с потерями является приемлемым результатом.

Выводы

Рассмотрена задача сжатия данных высоких размерностей с помощью автоассоциативной гетерогенной четырехслойной нейронной сети типа многослойного персептрона. Предложена процедура ее обучения на основе обратного распространения ошибки, характеризующаяся высоким быстродействием, сглаживающими и следящими свойствами, введением дополнительной регуляризации и численного критерия. По сравнению с процедурами сжатия на основе преобразования Карунена-Лоэва предложенный нейросетевой подход позволяет выявить более «тонкие» эффекты, присутствующие в реальном изображении.

Список литературы: 1. *Dony R.D., Haykin S.* Neural network approaches to image compression // *Proc. IEEE.* — 1995. — 83. — P. 288-303. 2. *Bishop C.M.* Neural network for pattern recognition. — Oxford: Clarendon Press, 1995. — 482p. 3. *Kramer M.L.* Nonlinear principal component analysis using autoassociative neural networks / *AICHE J.* — 1991. — 32. — No 2. — P.233-243. 4. *Tan S., Mavrouniotis M.* Reducing data dimensionality through optimizing neural-network inputs // *AICHE J.* — 1995. — 41. — No 6. — P. 1471-1480. 5. *Malthouse E.C.* Limitations of nonlinear PCA as performed with generic neural networks // *IEEE Trans. on Neural Networks.* — 1998. — 9. — P. 165-173. 6. *Kaczmarz S.* Approximate solution of system of linear equations // *Int. J. Control.* — 1993. — 57. — No 6. — P. 1269-1271. 7. *Widrow B., Hoff Jr. M.E.* Adaptive switching circuits // 1960 IRE Western Electric Show and Con-

nection Record. — 1960. — Part 4. — P. 96-105. 8. *Widrow B., Lee M.* 30 years of adaptive neural networks perceptron, adaptive and backpropagation // *Proc. IEEE.* — 1990. — 78. — P. 1415-1442. 9. *Бодянский Е.В., Плисс И.П., Соловьева Т.В.* Многошаговые оптимальные упредители многомерных нестационарных стохастических процессов // *Докл. АН УССР.* — 1986. — Сер. А. — № 12. — С. 47-49. 10. *Cichocki A., Unbehauen R.* Neural networks for optimization and signal processing. — Stuttgart: Feubner, 1993. — 526p. 11. *Haykin S.* Neural networks. A comprehensive foundation. — Upper saddle river, N.J.: Prentice Hall, Inc., 1999. — 842p. 12. *Otto P., Bodyanskiy Ye., Kolodyazhnyi V.* A new learning algorithm for a forecasting neuro-fuzzy network // *Integrated Computer-Aided Engineering.* — 2003. — 10. — No 4. — P399-405.

Поступила в редколлегию 3.09.2008

УДК 004.032 26

Автоасоціативна багатосарова нейронна мережа та алгоритм її навчання при стисненні зображень / В.П. Машталір, О.Є. Путятін, Є.С. Сакало // Біоніка інтелекту: наук.-техн. журнал — 2008. — № 2 (69). — С. 40-44.

Запропоновано та розглянуто побудову багатосарової нейронної мережі при розв'язанні задачі стиснення зображень. Також запропоновано алгоритм її навчання та проведено аналіз отриманих результатів.

Л. 2. Бібліогр.: 12 найм.

UDC 004.032 26

Autoassociative multi-layer neural network and educational algorithm for image compression / V.P. Mashtalir, A.E. Putiatina, E.S. Sakalo // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 40-44.

Multi-layer neural network for image compression are proposed. Educational algorithm are offered and results analysis are conducted.

Fig. 2. Ref.: 12 items.

УДК 004.932.2



МЕТРИКИ НА МНОЖЕСТВАХ КЛЮЧЕВЫХ ТОЧЕК ИЗОБРАЖЕНИЙ

В.А. Гороховатский

ХНУРЭ, г. Харьков, Украина

Приведены новые результаты по синтезу и применению метрик в методах распознавания, устойчивых к частичным искажениям визуальных объектов. В качестве структурных признаков используется множество характерных точек. Описаны результаты имитационного моделирования в целях сравнительной оценки помехозащищенности обсуждаемых метрик.

КОМПЬЮТЕРНОЕ ЗРЕНИЕ, ИЗОБРАЖЕНИЕ, КЛЮЧЕВЫЕ ТОЧКИ, МНОЖЕСТВО, МЕТРИКА НА МНОЖЕСТВАХ, ЧАСТИЧНОЕ ИСКАЖЕНИЕ, РАСПОЗНАВАНИЕ, ГОЛОСОВАНИЕ, АДДИТИВНЫЙ ШУМ, ПОМЕХОЗАЩИЩЕННОСТЬ

Введение

Метрики – важный инструмент решения многих задач распознавания образов и интеллектуального анализа данных. Наличие метрики в пространстве позволяет принимать решение о принадлежности к множеству или о сходстве множеств на основе количественного показателя. Величина метрики (расстояния) часто непосредственно связана с вероятностными характеристиками отнесения элемента к классу [1]. В настоящее время известно огромное количество метрик, но задача поиска и синтеза новых метрик остается актуальной [2–7]. Основной целью создания метрик есть повышение эффективности при решении новых практических задач [8–11].

В задачах распознавания объектов, описываемых наборами ключевых точек (КТ) [4], исходная информация представляется в виде конечного множества векторов (дескрипторов) с действительными компонентами. По сути, такой набор является не вектором, а множеством, так как порядок следования элементов в нем зависит от процесса обработки. Поэтому мера сходства должна рассматриваться как метрика на множествах. Кроме того, дескрипторы дополнительно характеризуются координатами КТ, отражающими пространственную структуру объекта. Распознавание на множестве КТ представляют как оптимизацию некоторой меры, отражающей величину «пересечения» множеств.

Цель исследования – формализация и анализ свойств метрик на множествах КТ применительно к задаче распознавания изображений в условиях неполного представления.

Задачи исследования – построение метрик на множестве одиночных КТ, а также на множестве наборов КТ, анализ способов конкретного применения метрик, оценка эффективности разновидностей метрик.

1. Метрики на множестве КТ

Формально систему распознавания $\mathfrak{R}$ можно рассматривать в виде пары $\mathfrak{R} = \langle \mathfrak{w}, \mathfrak{v} \rangle$, где $\mathfrak{v}$ – фун-

кция формирования, а $\mathfrak{w}$ – функция обработки альтернатив классов. Значения альтернатив включают числовую оценку принадлежности к j -му классу. Эта оценка представляет собой расстояние или степень близости [1, 5]. К расстоянию предъявляется более строгая совокупность требований: симметричность, удовлетворение неравенству треугольника, равенство нулю только при совпадении элементов [3]. Степень близости должна быть симметричной, максимальной при сравнении объекта с самим собой, а также обладать свойством монотонного убывания при увеличении значения соответствующей метрики [5]. Метрики относятся к основным способам формирования альтернатив, где оценки принадлежности вычисляются однозначным образом.

При анализе дескрипторов возникает два основных вида помех [8]. Это, в первую очередь, помехи аддитивного типа, приводящие к небольшим отклонениям от истинных значений. С другой стороны, действуют локальные помехи и фон, часто приводящие к весомерному изменению дескриптора и вообще к возможному попаданию его в другой класс. Удачно построенная мера позволяет, в частности, учитывать как погрешность из-за незначительных отклонений, а также путем отбрасывания ложных соответствий дескрипторов обеспечить надежное распознавание объектов при существенном изменении анализируемого множества [2, 4].

В задачах распознавания по множеству КТ рассматривают два условных класса методов: основанные на низкоуровневом анализе отдельных КТ и на анализе глобальных характеристик, построенных путем интеграции имеющихся значений КТ изображения. В данной работе основное внимание уделяется первой группе методов, основой которых есть структурный анализ множеств.

При работе с множествами оперируют понятием «мера множества» [3, 5]. Суть меры для конечных множеств, как правило, состоит в оценке мощности N (количества элементов). Мера пересечения двух множеств A и C сводится к подсчету количества «общих» элементов A и C . С этой точки зрения

такие общепринятые метрики как пересечение и симметрическая разность имеют близкую трактовку. Для их мощностей выполняются соотношения, связывающие их с общей суммой числа элементов

$$\begin{aligned} N_U &= N_A + N_C - N_I, \\ N_\Delta &= N_A + N_C - 2N_I, \end{aligned} \quad (1)$$

где индексы обозначают операции объединения U , пересечения I и симметрической разности Δ множеств A, C .

Учитывая сложность задач распознавания изображений, связанную с близостью между собой значений отдельных дескрипторов КТ для разных объектов, основное внимание при построении метрик сосредоточивается на том, чтобы на первом этапе путем бинарных решений оценить «совпадение» одиночных КТ для сравниваемых объектов, а затем подсчетом прецедентов (голосов) принять окончательное решение.

Пусть $\mathcal{U} = \{u\}$ — множество КТ (универсум), а $\mathfrak{S} = (\mathcal{U}, \rho)$ — метрическое пространство с метрикой ρ на этом множестве. В большинстве задач анализа изображений $\mathcal{U}$ является множеством n -мерных векторов с действительными компонентами, то есть $u = (u_1, \dots, u_n)$, $u \in \mathcal{U}$, $u_i \in R$ [2].

Одним из наиболее простых вариантов построения ρ есть метрика «изолированных точек» [3]

$$\rho(a, c) = \begin{cases} 0, & a = c, \\ 1, & a \neq c, \end{cases} \quad (2)$$

основанная на «совпадении» элементов $a \in \mathcal{U}, c \in \mathcal{U}$. Учитывая естественные погрешности при вычислении дескрипторов, для оценки совпадений более практично использовать понятие δ -окрестности элемента

$$O_\delta(u) = \{x : \rho(x, u) \leq \delta\} \quad (3)$$

как множество элементов, расстояние до которых от элемента u не превышает величины некоторого порога δ (образует «шар» с радиусом δ и центром в точке u). В выражении (3) фигурирует отличная от (2) метрика. Рассматривая теперь каждый элемент как шар с радиусом δ , метрику (2) преобразуем к виду

$$\rho(a, c) = \begin{cases} 0, & O_\delta(a) \cap O_\delta(c) \neq \emptyset, \\ 1, & O_\delta(a) \cap O_\delta(c) = \emptyset, \end{cases} \quad (4)$$

то есть элементы a, c считаются одинаковыми в случае непустого пересечения их окрестностей $O_\delta(a), O_\delta(c)$. Соотношение (4) часто применяют как практическую аппроксимацию для реализации классической метрики (2). Выражения (2), (4) при использовании их в метриках для множеств дают возможность подсчитать количество несовпадающих элементов.

Другим вариантом применения метрики с учетом отклонений может быть проверка условия

вида $\rho(a, c) \leq \delta$, определяющего «эквивалентность» элементов a, c путем ограничения на величину расстояния между ними. Примером подходящей метрики может быть расстояние R_1^n

$$\rho(a, c) = \sum_{k=1}^n |a_k - c_k|. \quad (5)$$

Введенная на $\mathcal{U}$ метрика порождает отношение эквивалентности и, как следствие, разбиение пространства на непересекающиеся классы эквивалентных элементов. Элементы $a \in \mathcal{U}, c \in \mathcal{U}$ считаются эквивалентными, если выполняются условия вида $\rho(a, c) = 0$ либо $\rho(a, c) \leq \delta$. Однако в практике задач компьютерного зрения установление такой эквивалентности представляет собой довольно сложную задачу, так как при определенном значении δ из-за попарной близости значительное число дескрипторов может оказаться «эквивалентными». Поэтому порог δ часто выбирают экспериментально, анализируя фиксированную базу данных [4, 8].

Чаще всего используемые для сравнения конечных множеств метрики основаны либо на оценке степени совпадения (2) в виде числа голосов, либо на количественной мере соответствия аналогично (5). В частности, в известном методе SIFT для сопоставления дескрипторов используется классическое евклидово расстояние [9].

2. Расстояния для множеств

Вычисление метрики для множеств реализуется двухэтапной процедурой, где вначале определяются значения для отдельных элементов, а далее осуществляется некоторая «интеграция» этих значений [3, 5].

Содержание этапов задает качественные характеристики метрики. Второй этап можно рассматривать как в интегральном («средней связи» [5]), так и в дифференциальном («ближнего, дальнего соседа») плане. В структурных подходах более применимы методы второго типа, позволяющие формировать решения по подмножествам элементов, отбраковывая в некоторой мере «искаженные» дескрипторы. Интегральные же способы без анализа суммируют информацию от истинных и ложных дескрипторов и применимы в условиях высокого соотношения сигнал-шум.

Сопоставляемые множества A, C будем рассматривать как конечные наборы элементов из $\mathcal{U}$, то есть $A = \{a^1, a^2, \dots, a^{N_A}\}$, $C = \{c^1, c^2, \dots, c^{N_C}\}$, $a^k \in \mathcal{U}, c^k \in \mathcal{U}$, где N_A, N_C — мощности множеств. В задачах анализа КТ изображений множества A, C часто являются мультимножествами, то есть могут иметь повторяющиеся (эквивалентные) элементы.

Таким образом, при распознавании на множестве КТ имеем дело с двумя метрическими пространствами: $\mathfrak{S}$ с метрикой ρ для одиночных КТ и $\mathfrak{S}_A$ с метрикой ρ_A для множеств КТ, где $\mathfrak{S}_A = (\mathcal{U}_A, \rho_A)$,

$\mathcal{U}_A$ – множество элементов из $\mathcal{U}$. Пространство $\mathfrak{Z}$ есть подпространство $\mathfrak{Z}_A$.

Пример расстояния ρ_A между множествами – метрика Хаусдорфа

$$\rho_A(A, C) = \max[\max_{a \in A} \min_{c \in C} \rho(a, c), \max_{c \in C} \min_{a \in A} \rho(a, c)], \quad (6)$$

где $\rho(a, c)$ – метрика из пространства $\mathfrak{Z}$. Расстояние (6) обладает всеми свойствами метрики, то есть симметрично, удовлетворяет неравенству треугольника и обращается в нуль тогда и только тогда, когда $A = C$ [3]. Фактически метрика (6) устанавливает расстояние между множествами в виде выбранного специальным образом одного из расстояний между представителями этих множеств.

Метрика (6) может быть описана в терминах расстояния от точки до множества, которое представляет собой число

$$\rho(x, A) = \min_{a \in A} \rho(x, a). \quad (7)$$

Тогда (6) можно описать через (7) как

$$\rho_A(A, C) = \max\{\max_{a \in A} \rho(a, C), \max_{c \in C} \rho(c, A)\}.$$

Метрика вида (6) основана на одном из расстояний между элементами множеств и по этой причине не может считаться устойчивой к искажениям типа пропадания отдельных КТ и появления ложных КТ, вызванных локальными воздействиями на изображение. Аналогичным недостатком обладает и метрика «ближайшего соседа» [5]

$$\rho_A(A, C) = \min_{a \in A, c \in C} \rho(a, c). \quad (8)$$

Примером метрики для конечных множеств, интегрально учитывающей схожесть всех без исключения элементов, может быть величина «средней связи» между множествами A, C как арифметическое среднее всевозможных попарных расстояний между элементами [5]

$$\rho_A(A, C) = \frac{1}{N_A N_C} \sum_{a \in A} \sum_{c \in C} \rho(a, c). \quad (9)$$

Метрика (9) более устойчива к аддитивным искажениям КТ, чем метрики (6), (8) так как основана на усреднении, однако ее также нельзя считать устойчивой к локальным искажениям. Путь построения метрики с заданными свойствами состоит в синтезе модификаций классических метрик (6)–(9).

По аналогии с (3) для учета отклонений в пространстве $(\mathfrak{Z}_A, \rho_A)$ можно использовать понятие ε -окрестности множества [3], которая определяется в виде множества $O_\varepsilon(A) = \{x : \rho(x, A) \leq \varepsilon\}$ элементов, расстояние от которых до множества A не превышает величины ε . Понятие O_ε используется в теории приближения функций, где сходство между функциями устанавливается с учетом ε -окрестностей.

3. Метрики с учетом модели засорения

Процесс формирования элемента $z \in Z$ из множества $Z \subset \mathcal{U}$ в соответствии с моделью действия локальных помех [4] можно представить в виде «засорения» выборки [7]

$$z = \begin{cases} z^0, & \text{с вероятностью } (1 - \alpha), \\ t, & \text{с вероятностью } \alpha, \end{cases} \quad (10)$$

где z^0 – некоторое «истинное» значение дескриптора, t – случайное значение помехи, α – вероятность засорения, z, z^0, t принадлежат $\mathcal{U}$. Кроме того, дополнительно измерение z подвергается действию аддитивного шума [6].

Идеальным случаем можно считать вариант, когда дескрипторы разных классов имеют отличающиеся между собой распределения. В таком случае можно применить классификацию, в основе которой лежат интегральные характеристики классов, например, математическое ожидание, дисперсия и другое [1]. К сожалению, для КТ как структурных признаков объектов эта простая ситуация не характерна. Здесь, наоборот, имеем дело с дескрипторами, которые бывают достаточно близкими для разных классов.

Возможные подходы к построению метрик с необходимыми свойствами: а) усиление различий в пространстве дескрипторов разных классов [9]; б) использование вспомогательной информации на основе отношений дескрипторов в амплитудном и в пространственном смыслах [10]; в) отбраковка «ложных» дескрипторов в процессе вычисления метрики.

В условиях действия искажений целесообразно применить расстояния, согласованные с моделью действия помех (10). Основываясь на априорно заданной величине вероятности α , необходимо так синтезировать расстояние между объектами, чтобы оно было устойчиво к структурным искажениям заданного уровня. Если анализируемое множество КТ имеет мощность N , то решение должно основываться на количестве элементов равном или не меньшем значения $N_\alpha = (1 - \alpha)N$.

Для метода частных корреляций сравниваемые изображения представляются множествами из s фрагментов $B^j = \{b_i^j\}, i = \overline{1, s}$, между которыми существует строгое структурное соответствие [8]. Если $\rho(b_i^j, b_i^l)$ – метрика для фрагментов, то вариант метрики для изображений может быть представлен как

$$\rho_A(A, C) = \frac{1}{s} \sum_{i=1}^s \Pr[\rho(a_i, c_i) \leq \delta], \quad (11)$$

где предикат $\Pr$ задается в виде

$$\Pr[\rho(a_i, c_i) \leq \delta] = \begin{cases} 0, & \rho(a_i, c_i) > \delta, \\ 1, & \rho(a_i, c_i) \leq \delta. \end{cases} \quad (12)$$

Соотношение (11) есть линейная комбинация метрик, поэтому само является метрикой. Порог

δ задает радиус шара для одинаковых элементов, а нормировка проводится по общему количеству элементов s , посредством чего обеспечивается представление значений метрики в виде доли несовпадающих элементов. В результате метрика (11) вычисляет нормированное количество отличающихся элементов для множеств A, C равной размерности (пересечение), при этом структурное соответствие элементов a_i, c_i зафиксировано. По результатам вычисления (11) нужно проверить условие на количество s_α общих элементов с учетом параметра засорения: $s_\alpha \geq (1-\alpha)s$. Если это условие не выполнено, то величина (11) не определяется. Таким образом, метрика (11) отражает свойства подмножеств «одинаковых» фрагментов, мощность которых превышает величину $(1-\alpha)s$.

С применением базовой метрики «средней связи» (9) можно также рассмотреть вариант расстояния с учетом засорения в виде

$$\rho_A(A, C) = \frac{1}{s} \sum_{i=1}^s \rho(a_i, c_i) \Pr[\rho(a_i, c_i) \leq \delta], \quad (13)$$

где суммируются значения $\rho(a_i, c_i)$, соответствующие неэквивалентным элементам. Метрика (13) представляет собой линейную комбинацию значений метрики для несовпадающих элементов и отражает смысл «средней связи» [5]. Нормировка в (11), (13) также может быть осуществлена путем деления на количество неэквивалентных элементов [4].

Для произвольных множеств КТ, представленных в виде конечных наборов $A = \{a_i\}_{i=1}^{N_A}$, $C = \{c_j\}_{j=1}^{N_C}$, структурное соответствие между элементами в отличие от корреляционного метода устанавливается в процессе сопоставления [2] и фактически является его результатом. Поэтому основным способом реализации метрик в большинстве случаев остается полный перебор, для которого временные затраты количественно определяются величиной произведения мощностей $N_A N_C$. С учетом параметра α количество «одинаковых» элементов множеств A, C должно превышать величину $N_\alpha = (1-\alpha) \min(N_A, N_C)$, то есть число элементов в пересечении этих множеств должно превышать N_α . Значение метрики можно определить классически как пересечение $\rho_A(A, C) = \mu(A \cap C)$, где μ — некоторая мера для множеств. Конкретно такая метрика аналогично (11) выглядит как

$$\rho_A(A, C) = \frac{1}{N_A N_C} \sum_{i=1}^{N_A} \sum_{j=1}^{N_C} \Pr[\rho(a_i, c_j) \leq \delta]. \quad (14)$$

Дополнительно к (14) проверяется условие, что бы количество одинаковых элементов множеств A, C превышало величину N_α . Таким образом, за счет предиката, используемого в (11)–(14), осуществляется отбор неодинаковых элементов множеств.

Полезной может оказаться также идея ранжирования измерений, применяемая при обработке засоренных выборок в статистических задачах [7]. Применим ее для анализа множества значений метрики ρ , полученных на локальном этапе при сопоставлении дескрипторов КТ. Учитывая условие $\rho \geq 0$, проведем ранжирование величин $\{\rho_1, \dots, \rho_M\}$ по возрастанию $\{\rho_1 \leq \rho_2 \leq \dots \leq \rho_M\}$. Теперь будем использовать на втором этапе вычисления метрики выборку $\{\rho_1, \rho_2, \dots, \rho_{M_\alpha}\}$, усеченную справа до размера M_α , связанного с числом N_α допустимых общих элементов множеств. При этом считается, что «большие» значения метрики соответствуют сильно отличающимся элементам. На базе выборки $\{\rho_1, \rho_2, \dots, \rho_{M_\alpha}\}$ можно построить оценки, устойчивые к существенным отклонениям дескрипторов. В частности, эффективны оценки среднего, медианы и другое. Похожие процедуры применяют в статистике для анализа выборки с целью отбрасывания «хвостов» распределения. В результате будут отброшены значения, соответствующие незначительному сходству элементов, а итоговое решение будет устойчиво к искажениям с параметром α .

4. Голосование как способ реализации метрик

Пусть Z — распознаваемое множество КТ, а $Z(j)$, $j=1, \dots, J$ — эталонные множества, $Z \subset \mathcal{U}$, $Z(j) \subset \mathcal{U}$. Мера пересечения $I(j) = Z \cap Z(j)$ множеств $Z, Z(j)$ является классической метрикой [3, 11]. В качестве меры конечного множества $I(j)$ выберем количество элементов. Эту меру можно вычислить путем «голосования» элементов множества Z . Мера увеличивается на 1 (инкрементируется), если элемент $z \in Z$ имеет эквивалентный ему элемент из множества $Z(j)$. Эквивалентность будем понимать в смысле метрик (2), (4).

В теоретико-множественном представлении голосование сводится к построению на множестве Z представления $Z = \bigcup_j Z_j$ в виде разбиения (непересекающиеся между собой подмножества, $Z_i \cap Z_j = \emptyset$) или в виде покрытия ($Z_i \cap Z_j \neq \emptyset$), состоящего из элементов, каждый из которых получает метку класса j путем реализации отображения $Z \rightarrow J$. Решение о классе объекта, представленного множеством Z , принимается решающей коалицией, количество голосов которой соответствует наибольшей мощности среди пересечений Z с эталонами разных классов. Управляя порогом для размера решающей коалиции, можно согласованно принимать решение в зависимости от уровня искажений [4, 6].

Наряду с такими преимуществами, как устойчивость и гибкость принятия решения при влиянии частичных искажений КТ, методы голосования обладают таким важным достоинством, как простота построения, что при достаточно высокой надежности делает их конкурентоспособными среди известных подходов.

Использование различного типа отношений R_Z , построенных на множествах $Z, Z(j)$, дает возможность усовершенствования систем голосования путем дополнения их информационных ресурсов высокоуровневой системой признаков, включающей комбинации признаков [10]. Голос элемента из множества отношений R_Z считается значимым, если аналогичное отношение выполнено и для соответствующих элементов из множества $Z(j)$. Например голос пары в бинарном отношении значим, если обе КТ имеют достаточное сходство с одним и тем же эталоном. Голосование отношений как подмножеств Z в целом повышает надежность решений, однако, связано с некоторым ростом объема вычислений.

5. Эксперименты

В качестве экспериментальной модели для множества КТ выбран числовой вектор фиксированной размерности $s=10$, элементы которого имеют равномерное распределение на отрезке $[0, 9]$. Голосование за j -й класс проводится путем перебора возможных соответствий КТ объекта Z и эталонов $Z(j)$ как вычисление и анализ множества значений метрики в виде матрицы L_j размером $s \times s(j)$. Значения голосов $p[j]$ для класса j вычисляются на основе анализа L_j . Принятие решения сводится к анализу множества матриц $\{L_j\}, j=1, \bar{J}$.

Для сравнительной оценки помехозащищенности методов при действии аддитивного шума с нормальным распределением на вход подавался один из эталонов под влиянием шума на уровне $\eta = E/\sigma$, где E – среднее значение признака, σ – СКО шума [8]. Мерой ρ сравнения элементов выбран модуль разности. Использовалось правило: КТ принимает участие в голосовании (значимая КТ), если по величине ρ она отклоняется от 0 в пределах 10% своего максимального значения.

Целью компьютерного моделирования был качественный анализ методов различного типа, в частности, сравнение схем однозначных и множественных решений, а также одиночного и комбинированного (построенного на отношениях) типов голосования.

Исследовалась следующая группа одноуровневых методов, где каждая КТ из Z голосовала независимо от остальных:

1. Однозначное определение голосов одиночных КТ.
2. Множественный учет голосов значимых одиночных КТ.
3. Множественный учет голосов значимых пар КТ.
4. Множественный учет голосов значимых троек КТ.
5. Однозначное определение голосов пар КТ.

Количество сравниваемых признаков КТ, с чем непосредственно связан объем вычислительных

затрат, для разных методов существенно отличается. Например число сравнений для 1-го и 2-го методов будет равно $K = s \cdot \prod_{j=1}^J s(j)$, а для метода 3 оно равняется $K = A_s^2 \cdot \prod_{j=1}^J A_{s(j)}^2$, где A – число перестановок.

Рассмотрены также двухуровневые методы, где для формирования $p[j]$ применены следующие схемы нелинейной обработки голосов КТ по множеству Z :

6. $p[j]$ равно количеству значимых голосов элементов $z \in Z$, для которых имеется хотя бы одно соответствие в $Z(j)$. При условии $p[j] > 0,5s$ класс j участвует в решении, в противном случае считается $p[j] = 0$.

7. $p[j]$ равно значению расстояния «ближнего соседа» между множествами Z и $Z(j)$.

Результаты компьютерного моделирования приведены на диаграмме (рис. 1) в виде гистограммы значений η , характеризующих предельные возможности для помехозащищенности разных методов (1-7), когда вероятность распознавания остается достаточно высокой ($P > 0,95$). Чем ниже столбик, тем выше помехозащищенность метода.

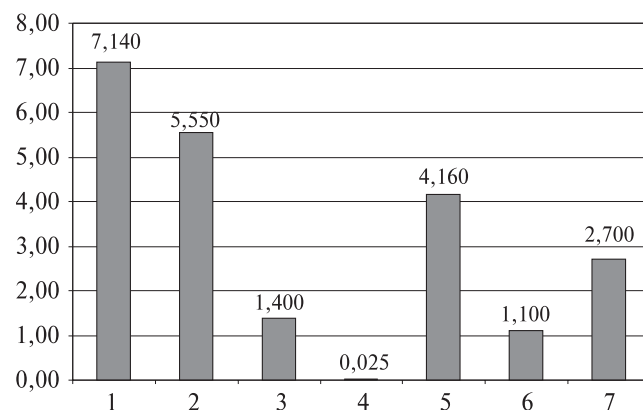


Рис. 1. Значения предельных соотношений сигнал-шум для разных методов

В целом в плане помехозащищенности голосование имеет достаточно высокие показатели. Как видим, наилучшую помехозащищенность имеют методы 3 и 4. Множественные решения более устойчивы к действию шума, чем однозначные. Это видно из сравнения столбиков 1 и 2. Еще более значительные отличия характерны для пар (столбики 3 и 5). Признаки на основе отношений обладают значительно лучшей помехозащищенностью (сравнение столбцов 3, 4 и 2). Преимущества отношений сохраняются с усилением шума. Результаты распознавания в целом зависят от меры ρ , а также от используемых признаков.

Понятно, что объективным в приведенных оценках есть только качественное сравнение, так как величина вероятности распознавания зависит также от количества и видов эталонов, порога для отклонений ρ и ряда других параметров.

Введение элементов иерархичности в процесс принятия решений упрощает вычисления, и по показателю помехозащищенности сравнимо с множественными решениями для пар КТ, причем метод 6 в этом плане несколько сильнее классического метода 7 «ближнего соседа». Учитывая, что вычислительные затраты для методов 3, 4 несколько выше, чем для подходов 6, 7 из-за существенно большего количества сравниваемых элементов, можно отдать предпочтение методу 6.

Проведенные исследования говорят о возможности согласованного выбора параметров и типа метода голосования в зависимости от условий распознавания, задаваемых уровнем помех и ограничениями на быстроедействие.

Выводы

Метрики, построенные на множествах КТ объектов с учетом частичного представления, расширяют круг решаемых задач и позволяют повысить надежность функционирования распознающих систем.

Научная новизна состоит в формализации применения метрик при распознавании по множеству КТ, изучении особенностей использования известных метрик и синтезе метрик с новыми свойствами.

Практическая значимость работы заключается в построении и сравнении между собой одноуровневых и иерархических способов построения метрик, получении конкретных характеристик помехозащищенности подходов.

Перспективы исследования состоят в построении методов распознавания по совокупности разнотипных признаков, включающих как значения, так и отношения разной местности. Вторым направлением есть оценка характеристик методов при действии локальных мешающих воздействий.

Список литературы: 1. Дуда Р., Харт П. Распознавание образов и анализ сцен. — М.: Мир, 1976. — 512 с. 2. Шапиро Л., Стокман Дж. Компьютерное зрение: Пер. с англ. — М.: Бином, 2006. — 752 с. 3. Колмогоров А.Н., Фомин С.В.

Элементы теории функций и функционального анализа. — Москва: Наука, 1976. — 544 с. 4. Гороховатский В.А. Распознавание изображений в условиях неполной информации. — Харьков: ХНУРЭ, 2003. — 112 с. 5. Прикладная статистика: Классификация и снижение размерности: Справ. изд./ Под ред. С.А. Айвазяна. — М.: Финансы и статистика, 1989. — 607 с. 6. Гороховатский В.А., Шляхов В.В. Анализ иерархических алгоритмов совмещения изображений на фоне пространственных локальных помех // Изв. высш. учебн. заведений. Радиоэлектроника. — Киев, 1991. Т. 34, № 1. С. 75–78. 7. Классификация и кластер/ Под ред. Дж. Вэн Райзина. — М.: Мир, 1980. — 390 с. 8. Путятин Е.П., Гороховатский В.О., Матат О.О. Методы та алгоритми комп'ютерного зору: Навч. посіб. — Харків: ТОВ «Компанія СМІТ», 2006. — 236 с. 9. Lowe D.G. Distinctive Image Features from Scale-Invariant Keypoints. — International Journal of Computer Vision, 60, 2, 2004, p. 91-110. 10. Гороховатский В.А. Совместное голосование признаков особенных точек изображений // Тези міжн. наук. конф. «Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту» (ISDMCI'2008). — Херсон: ХНТУ, 2008. — Т. 1, Ч. 1, С. 46-49. 11. Местецкий Л.М. Сравнение изображений гибких объектов на основе нормализации. — Труды междунар. конф. «GraphiCon'2007», Москва, 2007.

Поступила в редколлегию 5.09.2008

УДК 004.932.2

Метрики на множествах ключевых точек изображений / В.О. Гороховатский // Біоніка інтелекту: наук.-техн. журнал — 2008. — № 2 (69). — С. 45-50.

Досліджуються можливості застосування метрик на множествах ключевых точек при розпізнаванні зображень. Особливостями цих метрик є стійкість до часткових викривлень множини ознак. Наведено результати дослідження завадостійкості.

Л. 1. Бібліогр.: 11 найм.

UDK 004.932.2

The metrics in the sets of images dominant points / V.A. Gorohovatskij // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 45-50.

The opportunities of metrics application on the sets of key points are researched at images recognition. The features of these metrics are stable to partial distortions of the set of attributes. The results of noise immunity researches are shown.

Fig. 1. Ref.: 11 items.

УДК 681.3.07



ОПРЕДЕЛЕНИЕ ОПТИМАЛЬНОГО УРОВНЯ ВЕЙВЛЕТ-РАЗЛОЖЕНИЯ ИЗОБРАЖЕНИЯ ДЛЯ НОРМАЛИЗАЦИИ

О.А. Кобылин

ХНУРЭ, м. Харьков, Украина, kblin@kture.kharkov.ua

За счет определения оптимального уровня вейвлет-разложения возможно повысить скорость обработки изображения как традиционными методами, так и эффективно проводить нормализацию. Рассмотрены основные модели геометрических искажений, возникающие в системах технического зрения. Использование различных уровней разложения изображения позволяет существенно сократить количество операций, затраченных для 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 этих методов обработки изображений зависит от их способа представления. Именно от способа представления изображения зависит результативность анализа изображения. 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. Сокращение времени на геометрические преобразования остается актуальным на сегодняшний день.

Цель работы — определить оптимальный уровень вейвлет разложения изображения для нормализации.

Вейвлет-преобразование (wavelet transform) в настоящее время получило распространение во многих сферах применения обработки изображений: медицине, астрономии, в промышленности, системах компьютерной графики и других.

1. Вейвлет-разложения изображения

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 $\int_{-\infty}^{+\infty} \varphi(t) dt = 0$.

Произведем разложение изображения $B(x, y) \in L_2(R)$, имеющего размер 256×256 (рис. 1), как банк фильтров, применив следующие формулы:

$$vv_j(k, l) = \frac{1}{2} \sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} h'(2k-n) h'(2l-m) vv_{j-1}(n, m),$$

$$wv_j(k, l) = \frac{1}{2} \sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} g'(2k-n) h'(2l-m) vv_{j-1}(n, m),$$

$$vw_j(k, l) = \frac{1}{2} \sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} h'(2k-n) g'(2l-m) vv_{j-1}(n, m),$$

$$ww_j(k, l) = \frac{1}{2} \sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} g'(2k-n) g'(2l-m) vv_{j-1}(n, m),$$

где j — степень разрешения; $m, n = 0 \dots 2^j - 1$, $k, l = 0 \dots 2^j - 1$; $j = 0 \dots J$, $L = 2^j - 1$, $h'(n) = h(L - n - 1)$, $g'(n) = g(L - n - 1)$; n и m — координаты пикселей

изображения $vv_{j,k}^{(i)}$; k и l — координаты пикселей изображения $vv_{j,k}^{(i)}$, $vw_{j,k}^{(i)}$, $ww_{j,k}^{(i)}$; $h(t)$ и $g(t)$ — импульсные характеристики низкочастотного и высокочастотного фильтра.

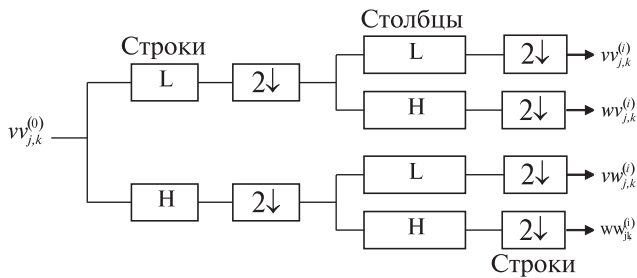


Рис. 1. Пример изображения

Рассмотрим уровни разложения, полученные в результате разложения. Схема разложения и пример разложения тестового изображения на 3 уровня представлен на рис. 2 и 3–5 соответственно. Ограничимся числом уровней разложения, равным 3.

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 — 128×128 , второго — 64×64 и третьего — 32×32 .

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

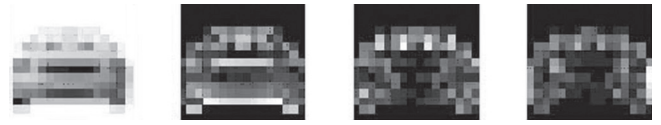


Рис. 5. Результат разложения по уровням декомпозиции 3-й уровень

В результате разложения получим на каждом уровне 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 .

Изображения, полученные на каждом уровне разложения $WB^{(1)}(x,y) \supset \dots \supset WB^{(k)}(x,y)$, оптимально подходят для выполнения процедур нормализации изображения.

При этом для каждого такого изображения можно выделить детализирующую компоненту, дополняющую его до изображения предыдущего уровня

$$d^{(i)} = WB^{(i+1)} * g^* = WB^{(i+1)} - WB^{(i)},$$

где g^* — высокочастотный декомпозирующий фильтр.

При этом очевидно, что

$$WB^{(0)} = d^{(1)} + d^{(2)} + \dots + d^{(k-1)} + WB^{(k)}.$$

Полученный набор изображений позволяет проводить более детальный анализ исходного изображения. Переход к другим уровням разложения позволяет отойти от мелких и случайных деталей, лучше выявить «внутреннюю» структуру при анализе геометрических искажений.

В связи с этим представляет интерес проводить нормализацию на разных уровнях разложения изображения. Изображения, полученные в результате вейвлет-разложения, лучше подходят для нормализации, чем исходное полутонное изображение.

Однако в результате вейвлет-преобразования на каждом шаге происходит изменение размера изображения в два раза (изображение пропускается через фильтры с полосой пропускания 0,5). Очевидно, что не каждый уровень разложения изображения подходит для операции нормализации. Поэтому необходимо определять оптимальный уровень вейвлет-разложения. Это особенно важно при проведении операций, связанных с геометрическими преобразованиями — нормализацией изображения.

Для определения оптимального уровня разложения проведем анализ полученных коэффициентов $vv_{j,k}^{(i)}$, $vw_{j,k}^{(i)}$, $ww_{j,k}^{(i)}$, $j,k \in Z$, которые характеризуют энергетическую структуру изображения на различных уровнях декомпозиции.

Сформируем из полученных коэффициентов $vv_{j,k}^{(i)}$, $wv_{j,k}^{(i)}$, $vw_{j,k}^{(i)}$, $ww_{j,k}^{(i)}$, $j, k \in Z$ признаки G_B^i , G_H^i , V_B^i , V_H^i , D_B^i , D_H^i , которые характеризуют горизонтальную, вертикальную и диагональную направленность эталонного В и Н анализируемого изображения на различных уровнях декомпозиции.

Горизонтальные коэффициенты ВП эталонного и анализируемого изображения:

$$G_B^{(i)} = \sum_{j=1}^n \max \{ vv_{j,k}^{(i)2}, j, k = 1..2^i \};$$

$$G_H^{(i)} = \sum_{j=1}^n \max \{ ww_{j,k}^{(i)2}, j, k = 1..2^i \}.$$

Вертикальные коэффициенты ВП эталонного и анализируемого изображения:

$$V_B^{(i)} = \sum_{j=1}^n \max \{ vv_{j,k}^{(i)2}, j, k = 1..2^i \};$$

$$V_H^{(i)} = \sum_{j=1}^n \max \{ vw_{j,k}^{(i)2}, j, k = 1..2^i \}.$$

Диагональные коэффициенты ВП эталонного и анализируемого изображения:

$$D_B^{(i)} = \sum_{j=1}^n \max \{ ww_{j,k}^{(i)2}, j, k = 1..2^i \};$$

$$D_H^{(i)} = \sum_{j=1}^n \max \{ ww_{j,k}^{(i)2}, j, k = 1..2^i \}.$$

Сформированные признаки можно использовать для определения оптимального уровня разложения.

Однако не каждый уровень разложения изображения подходит для операции нормализации.

Это объясняется тем, что сигнал претерпевает значительные изменения сразу на многих масштабах исследования и, соответственно, характеризуется возрастанием вейвлет-коэффициентов для многих уровней детализации в то время, как на стационарных участках вейвлет-коэффициенты оказываются сгруппированными вблизи определенных масштабов. Выбор уровня разложения может быть сведен к отысканию моментов увеличения вейвлет-коэффициентов на значительном количестве уровней масштабирования.

Таким образом, нормализацию необходимо проводить на том уровне разложения, который максимально сохраняет структуру изображения.

В этом случае возможно проводить анализ каждого уровня разложения и определять его оптимальный уровень разложения изображения для нормализации с помощью формулы:

$$\frac{G_B^i + V_B^i + D_B^i}{N^i} \leq \frac{G_H^1 + V_H^1 + D_H^1}{N^1}, \quad (1)$$

где N^i — количество коэффициентов разложения на каждом уровне декомпозиции; N^1 — количество

коэффициентов разложения на 1-ом уровне декомпозиции.

Первый уровень, для которого выполняется условие (1), является наиболее информативным.

2. Компенсация геометрических искажений для различных моделей смещений на различных уровнях разложения

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:

— Для модели $B(x, y) = B_1(x-n, y)$, группы $G_{c,x}$ нормализатор примет вид:

$$WF_C(B) = WB(x + \Phi w_1^i, y),$$

где

$$\Phi w_1^i = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} n v v_0^i(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v_0^i(n, m)} - \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} n v v^i(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v^i(n, m)}.$$

— Для модели $B(x, y) = B_1(x, y-m)$, группы $G_{c,y}$ нормализатор примет вид:

$$WF_C(B) = WB(x, y + \Phi w_2^i),$$

где

$$\Phi w_2^i = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} m v v_0^i(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v_0^i(n, m)} - \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} m v v^i(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v^i(n, m)}.$$

— Для модели $B(x, y) = B(x-n, y-m)$, группы $G_{c,x,y}$ нормализатор примет вид:

$$WF_C(B) = WB(x + \Phi w_1^i, y + \Phi w_2^i),$$

где

$$\Phi w_1^i = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} n v v_0^i(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v_0^i(n, m)} - \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} n v v^i(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v^i(n, m)},$$

$$\Phi w_2^i = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} m v v_0^i(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v_0^i(n, m)} - \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} m v v^i(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} v v^i(n, m)}.$$

— Для модели $B(x, y) = B(kx, ky)$, группы G_m нормализатор в случае равномерного и неравномерного изменения масштаба имеет вид:

$$WF_d(B) = WB(x \Phi w_1^i, y \Phi w_2^i),$$

$$WF_d(B) = WB(x \Phi_1 w_1^i, y \Phi_2 w_2^i),$$

где

$$\Phi w^i = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} vv^i(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} vv_0^i(n, m)},$$

$$\Phi_1 w^i = \sqrt{\frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} n^2 vv^i(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} vv^i(n, m)}},$$

$$\Phi_2 w^i = \sqrt{\frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} m^2 vv^i(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} vv^i(n, m)}}.$$

– Для модели $B_0(x, y) = B(x + hy, y)$, группы $G_{h,x}$ нормализатор примет вид:

$$WF_{Hx}(B) = WB(x + \Phi w^i y, y).$$

где

$$\Phi w^i = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} nm vv^i(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} m^2 vv^i(n, m)}.$$

– Для модели $B_0(x, y) = B(x, y + hx)$, группы $G_{h,y}$ нормализатор примет вид:

$$WF_{Hy}(B) = WB(x, y + \Phi w^i x),$$

где

$$\Phi w^i = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} nm vv^i(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} m^2 vv^i(n, m)}.$$

– Для модели:

$$B(x, y) = B_0(x \cos(\theta) + y \sin(\theta) - x \sin(\theta) + y \cos(\theta)),$$

группы G_u нормализатор примет 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

$$\Phi w^i = \frac{1}{2} \arctg \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} 2vv^i(n, m)(n^3 m + nm^3)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} vv^i(n, m)(n^4 + m^4)}.$$

– Для модели: $B_0(x, y) = B\left(\frac{x}{\lambda y + 1}, \frac{y}{\lambda y + 1}\right)$, груп-

пы G_p нормализатор перспективы примет вид:

$$F_p(B) = B\left(\frac{x}{\Phi w^i y + 1}, \frac{y}{\Phi w^i y + 1}\right),$$

где

$$\Phi w^i = \frac{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} vv^i(n, m)}{\sum_{n=0}^{2^j-1} \sum_{m=0}^{2^j-1} \frac{vv^i(n, m)}{n^4}}.$$

3. Эксперименты

Исследования проводились на базе компьютерной модели, позволяющей моделировать различные ситуации, возникающие в СТЗ. В компьютерной модели реализована возможность загрузки тестовых видеопоследовательностей синтезированных в 3DMAX. На синтезированные видеопоследовательности были наложены функция шума с нормальным и равномерным законами распределения.

Экспериментальные исследования устойчивости методов нормализации проводились с помощью компьютерного моделирования. Эталонное изображение объекта искажалось перспективными преобразованиями. Также с геометрическими искажениями накладывался нормальный шум с соотношением сигнал/шум от 10/2 до 10/5.

Испытания проводились для нескольких десятков изображений и видеопоследовательностей. В таблицах приводятся полученные усредненные значения.

Нормализация проводилась двумя способами:

– с использованием различных уровней вейвлет разложения;

– с помощью «классической» нормализации.

Проведем нормализацию изображения, искаженного перспективным преобразованием, с помощью нормализатора:

$$F_{p_y}(B) = B\left(\frac{x}{\Phi y + 1}, \frac{y}{\Phi y + 1}\right),$$

где функционал Φ определяется как:

$$\Phi = \frac{\iint_D \frac{B(x, y)}{x^4} dx dy}{\iint_D \frac{B(x, y)}{x^3} dx dy}.$$

Результат исследований нахождения параметров с помощью «классической» нормализации для перспективной группы приведен в табл. 1.

Таблица 1

Параметры преобразований	Погрешность Δ , %
3,27	1,5
8,33	1,7

Следующим этапом экспериментальных исследований была нормализация изображения, искаженного перспективным преобразованием, на различных уровнях разложения.

Таблица 2

Параметры преобразований	Погрешность Δ , %
1-й уровень разложения	
3,23	1,5
8,27	1,7
2-й уровень разложения	
3,13	2,7
8,28	1,8
3-й уровень разложения	
3,18	3,2
8,21	4,1

Результаты проведенных экспериментов показывают, что возможно эффективно применять различные уровни вейвлет разложения в СТЗ.

Выводы

В статье предложено использовать различные уровни разложения для нормализации изображения, полученные в результате вейвлет-преобразования. Рассмотрены различные типы вейвлетов. Использование различных типов вейвлетов влияют на определение оптимального уровня разложения. Определение оптимального уровня разложения изображения позволяет существенно сократить время 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. Вейвлеты 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 9.09.2008

УДК 681.3.07

Визначення оптимального рівня вейвлет-розкладання зображення для нормалізації / О.А. Кобилін // Біоніка інтелекту: наук.-техн. журнал — 2008. — № 2 (69). — С. 51-55.

Розглядається можливість компенсації геометричних перекошувань у системах технічного зору із застосуванням вейвлет-розкладання зображення. Використання різноманітних рівнів вейвлет-розкладання зображення дозволяє скоротити час нормалізації в системах технічного зору.

Табл. 2. Іл. 5. Бібліогр.: 4 найм.

UDK 681.3.07

Detection of image wavelet-decomposition optimal level for normalization / O.A. Kobylin // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 51-55.

A possibility of geometric distortions compensation in the technical vision field via wavelets decomposition approach is considered. Using various level the image wavelets decomposition allows to reduce the time for image normalization in the technical vision systems.

Tab. 2. Fig. 5. Ref.: 4 items.

УДК 519.859



ИНТЕРВАЛЬНОЕ ОЦЕНИВАНИЕ АЛЬТЕРНАТИВ ПРИ ПРИНЯТИИ РЕШЕНИЙ В ГЕОМЕТРИЧЕСКОМ ПРОЕКТИРОВАНИИ

И. В. Гребенник¹, Т.Е. Романова², С.Б. Шеховцов³

¹ ХНУРЭ, г. Харьков, Украина, grebennik@onet.com.ua

² ИПМаш НАНУ, г. Харьков, Украина, sherom@kharkov.ua

³ ХНУВД, г. Харьков, Украина, tarom7@yahoo.com

В статье рассматриваются задачи принятия решений в случае интервальной неопределенности, когда исходные данные заданы в интервальном виде. Строятся интервальные математические модели указанных задач. Преплагаются подходы, основанные на модификациях известных детерминированных методов принятия решений и интервальном анализе.

ИНТЕРВАЛЬНАЯ МАТЕМАТИЧЕСКАЯ МОДЕЛЬ, ПРИНЯТИЕ РЕШЕНИЙ, МНОГОФАКТОРНЫЕ ИНТЕРВАЛЬНЫЕ ОЦЕНКИ, ГЕОМЕТРИЧЕСКОЕ ПРОЕКТИРОВАНИЕ

Введение

Системы поддержки принятия решений используются при решении многих научных и прикладных задач. Основу математического обеспечения таких систем составляют математические модели задач принятия решений, относящихся к различным классам, и методы их решения [1]. На практике значительную часть решений приходится принимать в условиях неопределенности. В частности, эта проблема актуальна в классе оптимизационных задач упаковки, раскрой и покрытия [2].

Существуют различные методы учета неопределенности в математических моделях задач принятия решений [3–5].

В данном исследовании в качестве средства математического моделирования процесса выбора решений в условиях неопределенности используются методы интервального анализа [6].

Целью настоящей статьи является распространение известных методов решения детерминированных задач принятия решений на случай интервальной неопределенности. Здесь и далее под интервальной неопределенностью понимается случай, когда все исходные данные задачи принятия решений заданы в интервальном виде.

Постановка задачи. Пусть X — множество допустимых решений, каждое из которых оценивается множеством критериев. Полагаем, что значения критериев определены с точностью до некоторых интервалов $[a, b] \subset R^1$. Необходимо найти решение $x^0 \in X$, лучшее по всем заданным критериям в условиях интервальной неопределенности оценок альтернатив по заданным критериям.

Пусть $\{k_1(x), k_2(x), \dots, k_n(x)\}$ — множество интервальных отображений вида $k_i: X \rightarrow E_i \subset I_s R$, $i \in J_n = \{1, 2, \dots, n\}$. Здесь $I_s R = I_s R \cup \bar{I}_s R$ — пространство центрированных интервалов, где

$$I_s R = \{ \langle X \rangle = \langle x, v_x \rangle \mid x = \frac{a+b}{2} \in R^1, v_x = \frac{b-a}{2} \in R^1, [a, b] \subset R^1 \},$$

$$\bar{I}_s R = \{ \langle \bar{X} \rangle = \langle x, -v_x \rangle \mid \forall \langle X \rangle \in I_s R \},$$

$$E_i = \{ \langle X \rangle \in I_s R : x \in R^1, |v_x| \leq \delta_i \},$$

где δ_i — верхняя оценка «неопределенности» задания критерия $k_i(x)$.

1. Интервальная математическая модель задачи принятия решений

Определить

$$x^0 = \arg \operatorname{extr}_{x \in X} \{k_1(x), k_2(x), \dots, k_n(x)\}, \quad (1)$$

где экстремум интервального отображения понимается в смысле отношения порядка и способа определения максимума и минимума в пространстве $I_s R$.

Заметим, что в случае, когда для всех $k_i(x) = \langle k_i(x), v_{k_i(x)} \rangle$ выполняется соотношение $v_{k_i(x)} = 0$, задача (1) представляет собой детерминированную многокритериальную задачу принятия решений.

Отношение порядка в пространстве $I_s R$ вводится следующим образом [6]:

$$\forall \langle X \rangle = \langle x, v_x \rangle \in I_s R, \forall \langle Y \rangle = \langle y, v_y \rangle \in I_s R$$

$$(\langle X \rangle < \langle Y \rangle) \Leftrightarrow ((x < y) \vee (x = y) \wedge (v_x < v_y)), \quad (2)$$

$$(\langle X \rangle > \langle Y \rangle) \Leftrightarrow ((x > y) \vee (x = y) \wedge (v_x > v_y)),$$

$$(\langle X \rangle = \langle Y \rangle) \Leftrightarrow ((x = y) \wedge (v_x = v_y)).$$

На основе (2) минимум из n интервальных чисел $\langle Z_1 \rangle, \langle Z_2 \rangle, \dots, \langle Z_n \rangle$ определяется так:

$$\langle Z^* \rangle = \min \{ \langle Z_1 \rangle, \langle Z_2 \rangle, \dots, \langle Z_n \rangle \}, \quad (3)$$

где

$$\langle Z^* \rangle = \langle z^*, v_z^* = \min_{i \in J_n} v_{z_i} \rangle, \text{ если } z_1 = z_2 = \dots = z_n = z^*,$$

$$\langle Z^* \rangle = \langle z^* = z_{i_j} = \min_{i \in J_n} z_i, v_{z_{i_j}} \rangle, i_j \in J_n, j \in J_n,$$

$$\text{если } z_{i_j} \neq z_{i_k}, i \in J_k, j \neq k,$$

$$\langle Z^* \rangle = \left\langle z^* = \min_{i \in J_n} z_i, v^* = v_{z_{i_j}} = \min_{k \in J_r} v_{z_{i_k}} \right\rangle,$$

если $z_{i_1} = z_{i_2} = \dots = z_{i_r} = z^*$,

$$\{z_{i_1}, z_{i_2}, \dots, z_{i_r}\} \subset \{z_1, z_2, \dots, z_n\}, \quad 1 \leq r < n.$$

Аналогичным образом определяется максимум из n интервалов. Данный подход к определению минимума (максимума) распространяется и на случай бесконечного множества центрированных интервалов.

2. Анализ особенностей интервальной задачи

Задаче принятия решений в условиях интервальной неопределенности вида (1) присущи многие особенности детерминированных задач принятия решений.

Для решения задачи (1) предлагается построение и использование интервальных математических моделей, реализация которых основана на применении известных методов решения многокритериальных задач при $k_i: X \rightarrow R^1$, модифицированных для случая интервальных значений частных критериев $k_i(x)$ в $I_s R$.

При реализации неконструктивного подхода к выбору решения [7], обязательным является построение области компромиссов $X^c \subset X$. При конструктивном подходе (построение математической модели) эта процедура не обязательна, но часто желательна из вычислительных соображений.

Рассмотрим интервальные математические модели определения области компромиссов X^c на множестве допустимых решений X .

Если множество X дискретно и содержит небольшое число элементов, то для построения области компромиссов X^c используется попарное сравнение альтернатив $x \in X$ по интервальным оценкам.

Для любого решения $x \in X$ зададим отображение $f: X \rightarrow Y$ такое, что

$$f(x) = Y = (\langle Y_1 \rangle, \langle Y_2 \rangle, \dots, \langle Y_n \rangle), \quad (4)$$

где $\langle Y_i \rangle = k_i(x^j) \in I_s R$, $i \in J_n$. Тогда образом множества X будет множество $Y \subset I_s^n R$, где $I_s^n R$ — n -мерное интервальное пространство [8].

3. Область компромиссов для конечного множества X

Пусть $X = \{x^1, x^2, \dots, x^m\}$, а каждому $x^j \in X$ соответствует кортеж $Y^j = (\langle Y_1 \rangle, \langle Y_2 \rangle, \dots, \langle Y_n \rangle)$ оценок $\langle Y_i \rangle = k_i(x^j) \in I_s R$, $i \in J_n$, $j \in J_m$, по всем критериям из множества $\{k_1, k_2, \dots, k_n\}$. Построим область компромиссов X^c .

С этой целью осуществим попарное сравнение элементов множества X по каждому из критериев $k_1, k_2, \dots, k_n$ на основе соотношения (2) следующим образом.

Выберем альтернативы $x^p, x^q \in X$ и сравним $k_i(x^p)$ и $k_i(x^q)$ для всех $i \in J_n$. Возможны следующие ситуации: либо x^q доминирует x^p ; либо x^p доминирует x^q ; либо $x^p, x^q \in X$ несравнимы, то есть существуют такие $i_1, i_2 \in J_n$, что $k_{i_1}(x^p) < k_{i_1}(x^q)$ и $k_{i_2}(x^p) > k_{i_2}(x^q)$ или $k_{i_1}(x^p) > k_{i_1}(x^q)$ и $k_{i_2}(x^p) < k_{i_2}(x^q)$, и вывод о доминировании элементов $x^p, x^q \in X$ сделать нельзя.

Здесь символ $>$ означает предпочтение на множестве Y значений критериев. В частности, когда $Y \equiv I_s R$, то $k_i(x^p) > k_i(x^q)$ означает $k_i(x^p) > k_i(x^q)$, если критерий k_i максимизируется, и $k_i(x^p) < k_i(x^q)$ в противном случае. Соотношение $k_i(x^p) \approx k_i(x^q)$ означает $k_i(x^p) = k_i(x^q)$.

В результате сравнения элементов $x^p, x^q \in X$, $p = 1, 2, \dots, m-1$, $q = 2, \dots, m$ по всем критериям $k_i \in K$, $i \in J_n$, сформируем область компромиссов $X^c \subset X$ как множество всех недоминируемых альтернатив. Временная сложность приведенного алгоритма имеет оценку $\sigma = \frac{n \cdot m \cdot (m-1)}{2}$.

Когда определить область компромиссов на континуальном множестве X достаточно сложно или когда мощность конечного множества X и множества критериев достаточно велика, то целесообразно строить *приближенную* область компромиссов $\tilde{X}^c$ такую, что $X^c \subseteq \tilde{X}^c \subseteq X$.

Построение области $\tilde{X}^c$ основано на решении задач оптимизации $k_i(x)$ на множестве X . Суть решения задач оптимизации с интервальными функциями цели состоит в уменьшении диаметра (радиуса) интервала — значения функции цели. Указанное свойство может быть учтено в результате решения задач минимизации на основе соотношений (2)–(3). Для перехода к задачам минимизации интервальных функций введем интервальные отображения полезности $p_i(x)$ частных критериев $k_i \in K$, $i \in J_n$, задачи (1) с целью нормализации интервальных критериев, следуя [7].

Интервальное отображение полезности $p_i(x)$ частного критерия $k_i(x)$ должно удовлетворять следующим требованиям:

- 1) $Y_i = \{p_i(x) | \langle 0, 0 \rangle \leq p_i(x) \leq \langle 1, v_{p_i} \rangle, x \in X\}$;
- 2) $p_i(x)$ инвариантно к размерности частного критерия $k_i(x)$ и
- 3) $p_i(x)$ инвариантно к виду экстремума частного критерия $k_i(x)$.

Последнее требование означает, что независимо от вида экстремума (минимум или максимум) частного критерия $k_i(x)$, его наилучшему значению на множестве X должен соответствовать максимальный $(p_i(x) = \langle 1, v_{p_i} \rangle)$, а наихудшему — минимальный $(p_i(x) = \langle 0, 0 \rangle)$ результат отображения $p_i(x)$. Здесь v_{p_i} определяется как радиус интервала, задающего максимальную полезность решения $x \in X$, соответствующий δ_i .

Указанным требованиям отвечает интервальное отображение $\mathbf{p}_i(x)$ вида:

$$\mathbf{p}_i(x) = \left(\frac{\mathbf{k}_i(x) - \mathbf{k}_i^-}{\mathbf{k}_i^+ - \mathbf{k}_i^-} \right)^{\alpha_i}, \quad (5)$$

где $\mathbf{k}_i(x)$ — значение частного критерия, $\mathbf{k}_i^+, \mathbf{k}_i^-$ — соответственно наилучшее и наихудшее значение частного критерия $\mathbf{k}_i(x)$ на области допустимых решений $x \in X$, при этом

$$\mathbf{k}_i^+ = \begin{cases} \max_{x \in X} \mathbf{k}_i(x), & \text{если } \mathbf{k}_i(x) \rightarrow \max, \\ \min_{x \in X} \mathbf{k}_i(x), & \text{если } \mathbf{k}_i(x) \rightarrow \min; \end{cases}$$

$$\mathbf{k}_i^- = \begin{cases} \min_{x \in X} \mathbf{k}_i(x), & \text{если } \mathbf{k}_i(x) \rightarrow \max, \\ \max_{x \in X} \mathbf{k}_i(x), & \text{если } \mathbf{k}_i(x) \rightarrow \min; \end{cases}$$

$\alpha_i \in R^1$ определяет характер нелинейности функции полезности $p_i(x)$ в детерминированном случае [7].

Рассмотрим интервальное отображение $\hat{\mathbf{p}}_i(x)$ потери полезности частного критерия $\mathbf{k}_i \in \mathbf{K}$ как $\hat{\mathbf{p}}_i(x) = \langle 1, v_{p_i} \rangle - \bar{\mathbf{p}}_i(x)$, $\bar{\mathbf{p}}_i(x)$ — отображение, сопряженное отображению $\mathbf{p}_i(x)$ [6]. Очевидно, что независимо от вида экстремума частного критерия $\mathbf{k}_i(x)$, наилучшему результату отображения $\hat{\mathbf{p}}_i(x)$ соответствует минимальное значение $\langle 0, 0 \rangle$, а наихудшему — $\langle 1, v_{p_i} \rangle$. В дальнейшем, ориентируясь на решение задач минимизации интервальных отображений, будем осуществлять выбор наилучшего решения из множества X с помощью интервальных оценок потери полезности $\hat{\mathbf{p}}_i(x)$.

4. Приближенная область компромиссов

Рассмотрим один из возможных способов построения приближенной области компромиссов, который состоит в следующем.

Прежде всего построим приближенную область компромиссов для двух критериев $\mathbf{k}_i(x), i=1,2$. В этом случае согласно [7] на множестве допустимых решений X последовательно решаются две однокритериальные оптимизационные задачи:

$$x_i^0 = \arg \min_{x \in X} \hat{\mathbf{p}}_i(x). \quad (6)$$

Для каждого решения $x_i^0, i=1,2$, определяются $\hat{\mathbf{p}}_i(x)$. Обозначим $\hat{\mathbf{p}}_1(x_1^0) = \hat{\mathbf{p}}_1^+, \hat{\mathbf{p}}_2(x_2^0) = \hat{\mathbf{p}}_2^+, \hat{\mathbf{p}}_1(x_2^0) = \hat{\mathbf{p}}_1^-, \hat{\mathbf{p}}_2(x_1^0) = \hat{\mathbf{p}}_2^-$.

Образ $\tilde{Y}^c$ приближенной области компромиссов $\tilde{X}^c$ в пространстве Y задается следующим образом:

$$\begin{aligned} \tilde{Y}^c &= \{(\mathbf{k}_1(x), \mathbf{k}_2(x)) = \\ &= (\langle \mathbf{k}_1(x), v_{k_1}(x) \rangle, \langle \mathbf{k}_2(x), v_{k_2}(x) \rangle) \in Y \mid \\ &\hat{\mathbf{p}}_i^+ \leq \hat{\mathbf{p}}_i(x) \leq \hat{\mathbf{p}}_i^-, i=1,2\}. \end{aligned}$$

Пусть $i > 2$ и множество X конечно. Осуществим сравнение альтернатив в множестве X по каждой паре критериев из множества $\mathbf{K} = \{\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_n\}$.

Для этого сформируем множество пар критериев $\Pi = \{\Pi_{ij}\}$, $\Pi_{ij} = (\mathbf{k}_i, \mathbf{k}_j)$, $i \in J_n, j \in J_n$. Множество Π содержит $n(n-1)/2$ элементов.

Выберем элемент $\Pi_{ij} \in \Pi$ и реализуем приведенный выше способ определения приближенной области компромиссов для элементов множества X по критериям $\mathbf{k}_i, \mathbf{k}_j$. В результате получим множество $\tilde{X}_{ij}^c \subset X$. Выполним эту процедуру для всех $i, j=1, \dots, n, i < j$.

Приближенную область компромиссов $\tilde{X}^c$ можно представить в виде $\tilde{X}^c = \bigcup_{i,j=1, \dots, n, i < j} \tilde{X}_{ij}^c$.

Если X — непрерывное множество, то приближенную область компромиссов $\tilde{X}^c$ можно представить в виде $\tilde{X}^c = \prod_{i,j=1, \dots, n, i < j} \tilde{X}_{ij}^c$.

5. Формирование интервальных оценок альтернатив

Рассмотрим теперь способы формирования интервальных многофакторных оценок альтернатив на основе отображений обобщенной полезности альтернатив. Выделим следующие ситуации [7].

1. Пусть известны точные значения a_i коэффициентов относительной важности интервальных критериев $\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_n$, а, следовательно, их отображений локальной потери полезности $\hat{\mathbf{p}}_i(x)$. Здесь

$$a_i \in R^1, 0 \leq a_i \leq 1, i \in J_n, \sum_{i=1}^n a_i = 1. \quad (7)$$

Тогда в соответствии с детерминированным случаем сформируем аддитивную обобщенную интервальную оценку потери полезности вида:

$$\hat{\mathbf{P}}(x) = \sum_{i=1}^n a_i \cdot \hat{\mathbf{p}}_i(x), \quad (8)$$

а решение задачи примет вид:

$$x^0 = \arg \min_{x \in X} \hat{\mathbf{P}}(x). \quad (9)$$

В пространстве $\mathbf{I}_s \mathbf{R}$ определены отображения сопряжения, операции сложения и умножения на число $\lambda \in R^1$ [6]:

$$\begin{aligned} \langle \bar{X} \rangle &= \langle x, -v_x \rangle, \langle X \rangle + \langle Y \rangle = \langle x + y, v_x + v_y \rangle, \\ \lambda \langle X \rangle &= \langle \lambda x, |\lambda| v_x \rangle, \end{aligned}$$

где $\langle X \rangle = \langle x, v_x \rangle \in \mathbf{I}_s \mathbf{R}$, $\langle Y \rangle = \langle y, v_y \rangle \in \mathbf{I}_s \mathbf{R}$.

Заметим, что в модели (8) значения a_i являются действительными числами, а значения отображений локальной потери полезности $\hat{\mathbf{p}}_i(x)$ — элементами интервального пространства $\mathbf{I}_s \mathbf{R}$. В соответствии с аддитивной моделью (8) и операциями сложения и умножения на действительное число, введенными в пространстве $\mathbf{I}_s \mathbf{R}$, для каждой альтернативы $x \in X$ сформируем интервальные многофакторные оценки $\tilde{\mathbf{P}}(x, A)$, по которым найдем решение x^0 .

Полученное решение x^0 соответствует минимальной в смысле соотношения (3) интервальной

обобщенной оценке потери полезности альтернатив $x \in X$.

2. Пусть теперь значения коэффициентов относительной важности интервальных критериев $k_1, k_2, \dots, k_n$ известны с точностью до интервалов вида

$$\alpha_i = [\alpha_{i_{\min}}, \alpha_{i_{\max}}], \quad i \in J_n. \quad (10)$$

Каждому интервалу вида (10) поставим в соответствие центрированный интервал $\langle A_i \rangle = \langle a_i, v_{a_i} \rangle \in \mathbf{I}_s \mathbf{R}$, $i \in J_n$, [6] следующим образом:

$$\langle A_i \rangle \Leftrightarrow [a_i - v_{a_i}, a_i + v_{a_i}] = [\alpha_{i_{\min}}, \alpha_{i_{\max}}],$$

где $a_i = \frac{1}{2}(\alpha_{i_{\min}} + \alpha_{i_{\max}})$; $v_{a_i} = \frac{1}{2}(\alpha_{i_{\max}} - \alpha_{i_{\min}})$.

С целью выполнения для интервальных коэффициентов относительной важности критериев $k_1, k_2, \dots, k_n$ условий, аналогичных (7), осуществим нормализацию коэффициентов $\langle A_i \rangle$ по формуле $\langle A_i^H \rangle = \sigma^H \cdot \langle A_i \rangle$, $i \in J_n$, где $\sigma^H = \frac{1}{\sum_{i=1}^n \langle A_i \rangle}$.

Воспользуемся формулами интервального умножения, введенными в пространстве $\mathbf{I}_s \mathbf{R}$ [6].

Пусть $\langle X \rangle, \langle Y \rangle \in \mathbf{I}_s \mathbf{R}$. Интервальное произведение $\langle Y \rangle * \langle X \rangle$ задается следующим образом:

$$\begin{cases} \langle Y \rangle \cdot \langle X \rangle, & \text{если } \langle Y \rangle, \langle X \rangle \in \mathbf{I}_{s1} \\ \overline{\langle Y \rangle} \cdot \langle X \rangle, & \text{если } \langle X \rangle \in \mathbf{I}_{s2}, \langle Y \rangle \in \mathbf{I}_{s1} \\ \langle y + s | v_y, 0 \rangle \cdot \langle X \rangle, & \text{если } \langle X \rangle \in \mathbf{I}_{s3}, \langle Y \rangle \in \mathbf{I}_{s1} \end{cases},$$

$$s = \begin{cases} 1, & \text{если } v_x \cdot v_y \geq 0 \\ -1, & \text{если } v_x \cdot v_y < 0 \end{cases},$$

где $\mathbf{I}_{si}$, $i = 1, 2, 3$ — множество точек $\langle X \rangle \in \mathbf{I}_s \mathbf{R}$, которые удовлетворяют соответствующим условиям:

$$x - |v_x| > 0; \quad x + |v_x| < 0; \quad \begin{cases} x - |v_x| \leq 0, & \text{если } x \geq 0 \\ x + |v_x| \geq 0, & \text{если } x \leq 0 \end{cases},$$

где $\langle A \rangle \cdot \langle B \rangle = \langle ab + v_a v_b, av_b + bv_a \rangle$ — операция гиперболического умножения A и B .

Воспользовавшись формулой интервального деления [6], имеем

$$\langle A_i^H \rangle = \frac{\langle A_i \rangle}{\langle B \rangle} = \frac{1}{b^2 - v_b^2} \cdot \langle A_i \rangle * \langle B \rangle, \quad |b| \neq |v_b|,$$

$$\langle B \rangle = \langle b, v_b \rangle = \sum_{i=1}^n \langle A_i \rangle = \left\langle \sum_{i=1}^n a_i, \sum_{i=1}^n v_{a_i} \right\rangle. \quad (11)$$

Поскольку $a - |v_a| > 0$, то согласно таблице умножения интервальных чисел, формула (11) примет следующий вид:

$$\langle A_i^H \rangle = \frac{1}{b^2 - v_b^2} \cdot \langle a_i b + v_{a_i} v_b, b v_{a_i} + a_i v_b \rangle.$$

Следовательно,

$$\langle A_i^H \rangle = \langle a_i^H, v_i^H \rangle, \quad (12)$$

где

$$a_i^H = \frac{1}{\Delta_i} \left(a_i \sum_{i=1}^n a_i + v_{a_i} \sum_{i=1}^n v_{a_i} \right),$$

$$v_i^H = \frac{1}{\Delta_i} \left(v_{a_i} \sum_{i=1}^n a_i + a_i \sum_{i=1}^n v_{a_i} \right), \quad \Delta_i = \frac{1}{\left(\sum_{i=1}^n a_i \right)^2 - \left(\sum_{i=1}^n v_{a_i} \right)^2}.$$

Учитывая (12), определим

$$\sum_{i=1}^n \langle A_i^H \rangle = \langle a^H, v_a^H \rangle = \frac{1}{b^2 - v_b^2} \langle b^2 + v_b^2, 2b v_b \rangle =$$

$$= \frac{1}{\Delta_i} \left\langle \left(\sum_{i=1}^n a_i \right)^2 + \left(\sum_{i=1}^n v_{a_i} \right)^2, 2 \sum_{i=1}^n a_i \sum_{i=1}^n v_{a_i} \right\rangle,$$

$$\langle 0, 0 \rangle \leq \langle A_i^H \rangle \leq \langle a^H, v_a^H \rangle, \quad i \in J_n. \quad (13)$$

Замечание. Если положить все v_{a_i} равными нулю, то условие (13) будет эквивалентно выражению (7).

С учетом нормализованных интервальных коэффициентов относительной важности интервальных критериев соотношения, аналогичные (8), (9), примут вид:

$$\hat{P}(x, A^H) = \sum_{i=1}^n \langle A_i^H \rangle * \hat{p}_i(x),$$

$$x^0 = \arg \min_{x \in X} \hat{P}(x, A^H),$$

где $A^H = (\langle A_1^H \rangle, \langle A_2^H \rangle, \dots, \langle A_n^H \rangle) \in \mathbf{I}_s^n \mathbf{R}$.

3. Количественные значения весовых коэффициентов относительной важности критериев неизвестны, но интервальные критерии упорядочены по важности, например, следующим образом:

$$k_1 \succ k_2 \succ \dots \succ k_n.$$

Такое задание предпочтений частных критериев означает, что $\langle A_1 \rangle > \langle A_2 \rangle > \dots > \langle A_n \rangle$.

В этой ситуации используется принцип последовательной оптимизации, в соответствии с которым из двух решений $u \in X$, $v \in X$ первое предпочтительно, то есть $u \succ v$, если [9]

$$\hat{p}_1(u) < \hat{p}_1(v) \quad \text{или}$$

$$(\hat{p}_1(u) = \hat{p}_1(v)) \wedge (\hat{p}_2(u) < \hat{p}_2(v)) \quad \text{или}$$

$$(\hat{p}_1(u) = \hat{p}_1(v)) \wedge (\hat{p}_2(u) = \hat{p}_2(v)) \wedge (\hat{p}_3(u) < \hat{p}_3(v))$$

и так далее.

$\exists t \in J_{n-1}$ такое, что

$$(\hat{p}_j(u) = \hat{p}_j(v), j \in J_t) \wedge (\hat{p}_{t+1}(u) < \hat{p}_{t+1}(v)).$$

Выбор решения сводится к решению последовательности однокритериальных задач [7]:

$$X_i^0 = \arg \min_{x \in X_{i-1}^0} \hat{p}_i(x),$$

здесь $i \in J_n$, $X_0^0 \equiv X$.

4. Информация о предпочтениях относительно критериев $k_1, k_2, \dots, k_n$, а, следовательно, и о коэффициентах a_i , $i \in J_n$, отсутствует.

В этом случае следует использовать модель вида:

$$x^0 = \arg \min_{x \in X} \max_{i=1,2,\dots,n} \hat{p}_i(x).$$

Выводы

В настоящей работе предложен способ формирования интервальных многофакторных оценок альтернатив для принятия решений в геометрическом проектировании. Предложенный подход является комбинацией известных методов многокритериального принятия решений и методов интервального анализа.

Список литературы: 1. Кини Р.Л., Райфа Х. Принятие решений при многих критериях: предпочтения и замещения. - М.: Радио и связь, 1981. 2. Dyckhoff, G. Scheithauer, and J. Terno. Cutting and packing. In M. Dell'Amico, F. Maffioli, and S. Martello, editors, Annotated Bibliographies in Combinatorial Optimization, chapter 22, pages 393-412. John Wiley & Sons, Chichester, 1997. 3. Подиновский В.В. Задача оценивания коэффициентов важности как симметрически лексикографическая задача оптимизации // Кибернетика и системный анализ. — 2003. — № 3. — С. 150–162. 4. Овезгельдыев А.О., Петров К.Э. Оценка и ранжирование альтернатив в условиях интервальной неопределенности // Кибернетика и системный анализ. — 2005. — № 5. — С. 148–153. 5. Гребенник И.В., Романова Т.Е., Шеховцов С.Б. Оценка и ранжирование альтернатив при интервально заданных исходных данных в геометрическом проектировании // Системы обработки информации.—2007.— Вип. 2(60).—С. 109–112. 6. Стоян Ю.Г. Метрическое пространство центрированных интервалов// Докл. НАН Украины. Сер. А. — 1996. — № 7. — С. 23–25. 7. Овезгельдыев А.О.,

Петров Э.Г., Петров К.Э. Синтез и идентификация моделей многофакторного оценивания и оптимизации. — Киев: Наук. думка, 2002.—164 с. 8. Романова Т.Е. Интервальное пространство $I^n \mathbf{R}$. Интервальные уравнения // Докл. НАН Украины. Сер.А. — 2000.— № 9.— С. 36–41. 9. Подиновский В.В., Гаврилов В.М. Оптимизация по последовательно применяемым критериям. — М.: Сов. радио, 1975. — 192 с.

Поступила в редколлегию 11.09.2008

УДК 519.859

Інтервальне оцінювання альтернатив при прийнятті рішень у геометричному проектуванні / І.В. Гребенник, Т.Є. Романова, С.Б. Шеховцов // Біоніка інтелекту: наук.-техн. журнал — 2008. — № 2 (69). — С. 56-60.

У статті будуються інтервальні математичні моделі задач прийняття рішень в умовах інтервальної невизначеності. Для розв'язання задач, вихідні дані яких задані в інтервальному вигляді, пропонуються модифіковані методи на базі детермінованих методів прийняття рішень та інтервального аналізу.

Бібліогр. 9 найм.

UDC 519.859

Interval estimation of alternatives in decision making for geometric design / I.V. Grebennik, T.E. Romanova, S.B. Shekhovtsov // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 56-60.

In the article interval mathematical models of decision making problems under interval uncertainty are built. For solving problems with interval initial data modified methods based on determined decision making methods and interval analysis are proposed.

Ref.: 9.

УДК 519.6



СБОР И ОБРАБОТКА ИНФОРМАЦИИ В ИНТЕРНЕТ С ИСПОЛЬЗОВАНИЕМ ИНТЕЛЛЕКТУАЛЬНЫХ АГЕНТОВ

М.В. Збитнева, В.В. Дударь, А.Г. Деревянко

ХНУРЭ, г. Харьков, Украина

В статье предлагается усовершенствованный метод обработки информации о товарах по запросу пользователя с использованием интеллектуальных агентов в среде Интернет. В качестве формального аппарата использована алгебра предикатов первого порядка. Выполнен анализ существующих архитектур программных интеллектуальных агентов с точки зрения их применимости к задаче сбора и обработки web информации.

АГЕНТ, ОБРАБОТКА ИНФОРМАЦИИ, ИНТЕРНЕТ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Введение

Деятельность отдельных людей и организаций сейчас все больше зависит от имеющейся у них информации и способности ее эффективно использовать. Одним из основных средств получения информации сегодня является глобальная сеть Интернет — динамичная гетерогенная распределенная среда. Эффективный поиск информации в Интернете по мере увеличения объема и рассредоточения ее источников становится все более сложным и трудоемким. При этом критичным является не столько время поиска, сколько отбор информации, релевантной запросу пользователя. Информационный поиск представляет собой процесс сопоставления запроса пользователя со сведениями об информационных ресурсах, известных информационно-поисковой системе, к которой поступил этот запрос. В настоящее время доступ к информации, размещенной в глобальной сети Интернет, в подавляющем большинстве случаев обеспечивается при помощи поисковых машин. По оценкам, к поисковым службам обращается 71% всех пользователей Интернет [5]. Для формирования базы данных об информационных ресурсах поисковая система либо самостоятельно их индексирует, либо использует базы данных других поисковых систем. Второй способ значительно менее трудоемок, однако, то, что структура этих БД уже задана разработчиками, ограничивает параметры поиска.

В зависимости от потребности пользователей сети Интернет скорость поиска определенной информации в ней варьируется и зачастую требует автоматизации для облегчения работы. Это актуально в случае необходимости постоянного отслеживания какой-либо информации в Сети, её обновления, либо манипуляций с уже полученными данными (сортировка, вывод по определенным критериям и так далее). Для такого рода действий используются специальные прикладные программы — «агенты Интернет» [4].

Интеллектуальными агентами Интернет называют программы, автоматизирующие поиск, распознавание, извлечение и анализ информации

во всемирной Сети, ориентированные на нужды конкретного пользователя (или группы пользователей) [4]. Интеллектуальные агенты отличаются от широко используемых в настоящее время поисковых систем Интернет тем что:

1. Они способны работать самостоятельно в течение длительных промежутков времени (дни, недели и более), выполняя задание, порученное пользователем.

2. Как любая программа, однажды созданный агент может быть использован в будущем любое количество раз, в то время как запрос, посланный универсальной поисковой системе, вызывает однократную операцию сбора информации.

3. Поисковый агент передает пользователю не просто результаты работы машины (машин) поиска, а предварительно просматривает документы и выбирает из них наиболее релевантные с его точки зрения. Для этого агент использует определенную технологию оценки релевантности.

4. Агент может настраиваться на предпочтения пользователя, то есть учитывать ряд ограничений на поиск, формулируемых владельцем агента.

5. Некоторые агенты могут работать в off-line режиме, то есть когда пользователь дает задание агенту и отключается от Интернет, а агент отправляется на сервер, где выполняет работу. Когда пользователь снова подключается, агент посылает ему выполненное задание.

6. Агенты могут быть настроены на поиск по расписанию — искать информацию каждый час, день, неделю, месяц и так далее. Эта опция полезна при поиске, например, новостей, какой-то информации, обновление которой происходит довольно часто или информации, которая пользователю необходима постоянно в его работе.

7. Агенты могут обучаться. Пользователь оценивает работу агента, а агент, учитывая эти оценки, может корректировать свои критерии отбора информации.

Таким образом, поисковые агенты могут рассматриваться как интеллектуальная надстройка над машинами поиска.

Агент — это аппаратная или программная сущность, способная действовать в интересах достижения целей, поставленных пользователем. Они описываются также рядом свойств, которые характеризуют понятие агента. Обычно агент обладает набором из следующих свойств [2,3]:

1. Адаптивность: агент обладает способностью обучаться.

2. Автономность: агент работает как самостоятельная программа, ставя себе цели и выполняя действия для достижения этих целей.

3. Коллаборативность: агент может взаимодействовать с другими агентами несколькими способами, например, играя роль поставщика/потребителя информации или одновременно обе эти роли.

4. Способность к рассуждениям: агенты могут обладать частичными знаниями или механизмами вывода, например, знаниями, как приводить данные из различных источников к одному виду.

5. Коммуникативность: агенты могут общаться с другими агентами.

6. Мобильность: способность к передаче кода агента с одного сервера на другой.

В настоящей статье рассматриваются способы получения информации в системе World Wide Web — не просто в искусственной моделированной среде, а в той сложной, постоянно развивающейся среде, которая используется каждый день миллионами людей. Исследование будет проводиться на примере работы программных интеллектуальных агентов для поиска товаров в Интернет.

1. Постановка задачи

В данной статье проводится исследование способов получения информации о товарах в сети Интернет. Далее рассматриваются варианты создания приложений интеллектуальных агентов для поиска информации о предлагаемых интернет-магазинами товарах, определяется оптимальная архитектура и тип программы агента для приложений такого рода, и, в завершении, проводится формализация метода получения информации из Сети.

Задачи данного исследования:

1. Усовершенствовать существующие технологии по автоматическому извлечению информации из интернет-магазинов [1, 4], разработав приложение-агент со следующими характеристиками:

- работа в незнакомых средах (множество интернет-магазинов, разных по своей структуре);
- работа в долговременной перспективе, выполнение задач, растянутых во времени;
- самостоятельный анализ и сравнение предложений, выдача лишь релевантных результатов;
- невысокие требования к ресурсам клиентской машины и каналу в Интернет;
- обслуживание большого количества пользователей одновременно;

— возможность хранения индивидуальной информации;

— удобная обратная связь с пользователем.

2. Определение оптимальной архитектуры приложения для работы с интернет-магазинами, выбор типа интеллектуального агента.

3. Формализация предлагаемого метода получения информации о товарах в Интернет в виде формул алгебры предикатов первого порядка.

2. Выбор типа интеллектуального агента и определение архитектуры приложения

Разработку торгового агента можно проводить на основе одного из нескольких типов интеллектуальных агентов. Функцию агента реализует программа агента. Существует целый ряд основных проектов программ агента, соответствующих характеру явно воспринимаемой информации, которая используется в процессе принятия решения. Разные проекты характеризуются различной эффективностью, компактностью и гибкостью, а выбор наиболее подходящего проекта зависит от характера среды и предъявляемых агенту требований. Различают четыре основных типа программ агентов [1]:

- простые рефлексные агенты;
- рефлексные агенты, основанные на модели;
- агенты, действующие на основе цели;
- агенты, действующие на основе полезности.

Последние три типа интеллектуальных агентов могут также иметь способность к обучению, что намного усложняет их архитектуру, но позволяет выполнять задания в первоначально неизвестных вариантах среды и поэтому иметь более широкие сферы применения.

Простые рефлексные агенты отвечают непосредственно на акты восприятия, тогда как рефлексные агенты, основанные на модели, поддерживают внутреннее состояние, прослеживая те аспекты, которые не наблюдаются в текущем акте восприятия. Агенты, действующие на основе цели, организуют свои действия так, чтобы достигнуть своих целей, а агенты, действующие с учетом полезности, пытаются максимизировать свою собственную ожидаемую «удовлетворенность».

В зависимости от стоящих перед системой задач можно выделить два типа приложений: приложения, работающие с кратковременными (единичными) запросами и приложения для работы со сложными долговременными запросами. Работа с кратковременными запросами аналогична работе поисковых машин и заключается в предоставлении списка обработанных страниц интернет-магазинов с интересующим пользователя товаром и только. Работа со сложными запросами, растянутыми во времени, в полной мере использует все свойства интеллектуальных агентов. В данном случае мо-

жет потребоваться самостоятельная работа агента в течение дней, недель и даже более. Также здесь необходима возможность обучения и постоянного пополнения базы знаний для более эффективной работы и логика принятия некоторых решений для достижения необходимого результата.

Исходя из сложности архитектуры и масштабируемости проекта, можно выделить несколько типов приложений. Самым простым и доступным (с точки зрения затрат на разработку и вычислительной мощности компьютера пользователя) вариантом реализации программы является локальная архитектура приложения. Это Win32 (Java, WinForm для .Net и так далее) приложение, которое, используя пользовательский канал в Интернет, сканирует заданные Интернет-ресурсы и возвращает список страниц с результатами поиска. Этот вариант имеет недостаток в том, что накладывает определенные ограничения на скорость и гибкость поиска, так как в данном случае не учитывается история всей предыдущей работы агента, что может помешать при долговременном поиске, отслеживании изменений в интернет-магазинах и хранении информации о внутренней структуре этих веб-сайтов для облегчения последующего функционирования. Для устранения этих недостатков могут использоваться различные усовершенствования клиентских приложений, такие как: добавление базы данных для хранения и анализа результатов уже обработанных запросов; использование базы знаний для осуществления самостоятельного функционирования агента в сети и выполнения различного рода долговременных задач. В базе знаний уже могут храниться не только результаты выполненных запросов, а и собственная оценка агента уже обработанных страниц, а также вся важная информация, приведенная в удобную для хранения форму.

С точки зрения распределения модулей (поискового механизма приложения и клиентской части) существует также несколько способов реализации системы. Способом, лишенным недостатка в ограниченной пропускной способности пользовательского Интернет-канала, является использование технологии Клиент-Сервер, а особенно — использование Веб-сервисов XML. В этом случае интеллектуальная часть реализуется на сервере, доступном для клиентских подключений, однако, скорость поиска уже не будет зависеть от пропускной способности клиентского канала. Еще одним достоинством данного способа является то, что функциональную часть сервера могут одновременно использовать множество клиентов.

Для использования всех достоинств интеллектуальных агентов, приложение следует строить на программе рефлексного агента, основывающегося на модели с возможностью обучения. Такой интел-

лектуальный агент с успехом справится с долговременными задачами, функционируя даже в неизвестной среде. Его свойства будут способствовать работе с множеством интернет-магазинов, структура которых до сих пор была не известна. Для того чтобы повысить эффективность работы с множеством интернет-магазинов (а не с фиксированным их количеством), приложение необходимо связать с базой данных. Такой шаг позволит хранить в базе данных всю необходимую информацию о магазинах и их структуре, чтобы не тратить время на обучение каждый раз, когда возникнет необходимость в использовании агента.

База данных будет представлена в виде онтологии, которая представляет собой формализованное описание понятий и связей между ними. Применение онтологий позволяет представить предметную область в систематизированном виде, дает возможность оперировать абстрактными понятиями в проведении рассуждений.

Также имеет смысл вынести интеллектуальную часть на сервер, а клиентскую выполнить в виде WinForms-приложения. Работая в режиме веб-сервиса, интеллектуальный агент позволит разгрузить Интернет-канал пользователя и передавать на компьютер клиента лишь результаты выполнения запросов. Многопользовательская поддержка в таком случае также является плюсом, так как здесь практически не используются клиентские компьютеры в вычислениях, а вся нагрузка приходится на предназначенный под такие задачи сервер.

3. Обзор некоторых товарных агентов в Интернет

MySimon.com. В Web доступен интеллектуальный агент «MySimon.com», который производит быстрый поиск товаров в более чем 2 тыс. онлайн-магазинов, и сравнивает цены миллионов товаров. Это веб-сайт, поисковым движком которого и есть интеллектуальный агент.

В распоряжении пользователя имеется несколько базовых функциональных возможностей: поиск товара, ограничение по категориям в выпадающем списке рядом с формой поиска. Результатом выполнения поискового запроса будет список предложений из магазинов, доступных в Сети (в базе знаний агента «MySimon» их несколько тысяч). Также доступна сортировка, различные фильтры и возможность сравнения выбранных товаров вручную.

Наличие исключительно веб-интерфейса говорит о невозможности выполнения некоторых операций, таких как: многоразовый поиск, слежение за поступлениями, уведомление пользователя об определенных событиях. Также отсутствует определенная часть интеллектуальности, заключающаяся в самостоятельном анализе характеристик

товара агентом, которая перекладывает плечи пользователя.

Достоинством технологии, по которой реализован MySimon.com, можно считать реализацию интеллектуальной части на сервере, за счет чего обеспечивается высокая скорость поиска и поддержка работы в многопользовательском режиме.

PricePirates. Windows-приложение, агент для поиска товаров в Сети. PricePirates представляет собой полноценное приложение, реализованное в виде набора нескольких, связанных между собой сервисов. Под сервисами понимаются: отдельная страница гибкого поиска по аукционам, страница поиска по веб-магазинам со сравнением, поиск непосредственно по магазину Amazon.com, поиск в Веб, страница ведения статистики, популярные в eBay. Работа с Интернет-магазинами (раздел «Price comparison») осуществляется путем удаленного взаимодействия с некоторыми (участвующими в проекте «PricePirates») сайтами из каталога веб-агента «Shopping.com». Этот факт можно считать минусом данной программы. Действительно, пользователю хотелось бы иметь в распоряжении максимум выборки всех веб-магазинов Сети, а не только те, в которых заинтересованы разработчики «PricePirates». Здесь (также как и при работе с аукционами) существует возможность гибкого поиска товара, но отсутствует функция отслеживания или периодического сканирования магазинов на предмет поступления временно отсутствующей позиции.

Shopping Assistant. Это приложение-агент для работы исключительно с Интернет-магазином Amazon.com. Программа распространяется shareware (то есть бесплатно), и требует для работы Microsoft .Net Framework 1.1. Для осуществления поиска по магазину «Shopping Assistant» сканирует категории товаров, существующие в Amazon.com, после чего можно свободно по ним перемещаться и просматривать их содержимое. Также есть возможность поиска по ключевым словам, доступного из меню «Search». Полезной особенностью данной программы является возможность работы с учетными записями, совершение покупок под определенной записью и так далее. Здесь отсутствует какая-либо система для многократного поиска, автоматизации поиска, система оповещений, статистики. Другими словами: «Shopping Assistant» полностью оправдывает своё название как удобного «помощника» в совершении покупок в Интернет, не более.

4. Формализация метода обработки web информации торговым агентом

Набор отношений, характеризующих базу данных торгового агента, представим при помощи формул логики первого порядка. Пусть существует

база данных KnowledgeBase — множество знаний, тогда

$$cat_1, \dots, cat_n \in Categories,$$

$$Categories \subset KnowledgeBase,$$

где Categories — множество категорий товаров в Интернет, cat — элементы множества категорий, такие, что

$$\forall cat \ cat \in Categories \wedge Name(cat, name) \wedge \wedge Synonym(cat, syn_1) \wedge \dots \wedge Synonym(cat, syn_n),$$

где Name — отношение именования, Synonym — отношение синонима, а множество Synonyms — множество синонимов, которое содержит словарь синонимов для категорий. Здесь каждая категория имеет название Name, и у каждой категории имеется $syn_1, \dots, syn_n$ синоним.

Знания торгового агента об интернет-магазинах представляются таким образом:

$$store_1, \dots, store_n \in Stores,$$

$$Stores \subset KnowledgeBase,$$

где Stores — множество онлайн магазинов, store — элементы множества. При этом каждый элемент store имеет адрес главной страницы url, и имеет список категорий товаров cat, которые в нем можно приобрести:

$$\forall store, \exists cat \ store \in Stores \wedge cat \in Categories \wedge \wedge Url(store, page) \wedge Contain(store, cat),$$

где Url(store, page) — отношение, при котором каждому магазину store ставится в соответствие домашняя страница магазина page, а Contain(store, cat) — отношение содержания магазином store какой либо категории товаров cat.

В общем виде метод обработки информации торговым агентом описывается следующим образом:

$$GetOffers(query) \wedge CompareOffers(offers) \wedge \wedge ShowResult(result),$$

где query — запрос пользователя, offers — найденные предложения магазинов по запросу, result — релевантные предложения, выбранные торговым агентом. Здесь получение результата товарным агентом выполняется при совершении действий GetOffers (получить страницы различных магазинов с предложением товара), CompareOffers (сравнить найденный товар и выстроить градацию предложений в соответствии с определенными предпочтениями) и ShowResult (вывод результата в соответствии с градацией).

Ниже представлено описание получения предложений онлайн магазинов GetOffers(query) по запросу query, который состоит из действий GetStores(query) — получить список магазинов и

ряда действий $\text{GetRelevantPages}(\text{store}, \text{query})$ – получить список релевантных предложений (предложений, которые соответствуют пользовательскому запросу).

$$\begin{aligned} \text{GetOffers}(\text{query}) &= \text{GetStores}(\text{query}) \wedge \\ &\wedge \text{GetRelevantPages}(\text{store1}, \text{query}) \wedge \\ &\wedge \dots \wedge \text{GetRelevantPages}(\text{storeN}, \text{query}). \end{aligned}$$

Получение списка магазинов отличается от других аналогичных решений тем, что возможен поиск магазинов в Сети, которые не находятся в базе данных посредством поисковых систем:

$$\begin{aligned} \text{GetStores}(\text{query}) &= \\ &= \text{GetStoresFromBase}(\text{query}) \vee \\ &\vee \text{GetStoresFromWeb}(\text{query}), \end{aligned}$$

где действие $\text{GetStoresFromBase}(\text{query})$ – получение списка магазинов из базы данных, формально будет выглядеть как:

$$\begin{aligned} \text{GetStoresFromBase}(\text{query}) &= \\ &= \text{ParseQuery}(\text{query}) \wedge \\ &\wedge \text{LookForRelevantShops}(\text{pquery}), \end{aligned}$$

где действие $\text{ParseQuery}(\text{query})$ означает нормализацию пользовательского запроса в удобный для агента вид (pquery), а действие $\text{LookForRelevantShops}(\text{pquery})$ означает проведение поиска по базе данных нужных магазинов в соответствии с нормализованным запросом.

Действие по поиску магазинов в сети Интернет, если в базе данных не оказалось достаточного количества магазинов:

$$\begin{aligned} \text{GetStoresFromWeb}(\text{query}) &= \\ &= (\text{ParseQuery}(\text{query}) \wedge \\ &\wedge \text{SendQuery}(\text{pquery}) \wedge \\ &\wedge \text{ParseResultPage}(\text{page}) \wedge \\ &\wedge \text{LookForRelevantShops}(\text{pquery}) \wedge \\ &\wedge \text{UpdateKnowlageBase}(\text{shops})) \vee \\ &\vee \text{NotValidQuery}(\text{query}), \end{aligned}$$

где действие $\text{SendQuery}(\text{pquery})$ обозначает отправку запроса pquery поисковой системе, действие $\text{ParseResultPage}(\text{page})$ обозначает разбор страницы, которая была сформирована поисковой системой в ответ на запрос, $\text{UpdateKnowlageBase}(\text{shops})$ значит обновление базы данных новыми знаниями об онлайн магазинах, а действие $\text{NotValidQuery}(\text{query})$ будет совершаться в случае пустого результата.

Метод получения релевантных страниц $\text{GetRelevantPages}(\text{store}, \text{query})$:

$$\begin{aligned} \text{GetRelevantPages}(\text{store}, \text{pquery}) &= \\ &= \exists \text{page } \text{page} \in \text{Pages} \wedge \\ &\wedge \text{Relevant}(\text{page}, \text{pquery}) \wedge \\ &\wedge \text{Offer}(\text{page}), \end{aligned}$$

где существует хотя бы одна страница page , которая принадлежит множеству страниц Pages , является релевантной нормализованному запросу пользователя $\text{Relevant}(\text{page}, \text{pquery})$ pquery и которая предоставляет возможность совершить покупку $\text{Offer}(\text{page})$.

Алгоритм работы торгового агента представляется следующим образом:

- 1) Получение запроса пользователя.
- 2) Нормализация запроса – синтаксический разбор запроса и его представление с помощью определений, описанных в базе данных, как базовые.
- 3) Интеллектуальный поиск магазинов:
 - а. Обращение к базе данных, поиск магазинов соответствующей направленности в случае, если категория товара была определена агентом автоматически из запроса.
 - б. Расширение базы данных путем поиска магазинов соответствующей направленности в Сети с помощью поисковых систем, в случае, если знаний недостаточно.
 - в. Если агент не обладает дополнительными знаниями о товаре, осуществляется сразу поиск по страницам в магазинах, которые находятся в базе данных (функционал пункта 4).
 - г. Если страниц найдено не было – переход к поиску магазинов в Интернет с помощью поисковых систем.
 - е. В случае успеха – разбор найденных страниц и сохранение знаний о товаре и линейке товаров для ускорения последующего поиска.
- 4) Выбор страницы с предложением в найденных Интернет-магазинах.
- 5) Разбор страницы, выделение характеристик товара.
- 6) Сохранение полученных описаний товара в нормализованной форме.
- 7) Сравнение товара по критериям, очевидным для агента, либо по критериям, которые выбрал пользователь, сортировка.
- 8) Выдача отсортированного результата.

Преимуществами данного алгоритма перед алгоритмами других торговых агентов (*MySimon.com*, *PricePirates*, *ShoppingAssistant* и т.д.) и алгоритмами, которые предлагаются [1,4], являются: гибкий подход к выбору магазинов – самостоятельный поиск таковых в сети; интеллектуальный поиск, благодаря которому возможен не простой перебор страниц существующих онлайн магазинов, а выбор таких магазинов, в которых наличие данного товара наиболее вероятно.

Выводы

В результате исследований была разработана архитектура программы интеллектуального агента для осуществления покупок в Интернет. Также был предложен усовершенствованный алгоритм работы такого интеллектуального агента. Предло-

женное формальное описание с помощью алгебры предикатов первого порядка данного агента в виде метода обработки информации торговым агентом. Разработанный метод и алгоритм торгового агента отличается от аналогов [1,4] тем, что позволяет агенту более эффективно использовать вычислительные ресурсы, затрачивая на поиск меньше времени, благодаря усовершенствованному подходу в поиске магазинов. В данной программе агент вначале анализирует запрос и на основании своих знаний может спрогнозировать, в каких магазинах наиболее вероятно наличие такого товара. Если же такой прогноз невозможен, то агент возвращается к традиционному способу поиска товаров — к перебору страниц с предложениями по цепочке категорий. Еще одним нововведением можно считать возможность самостоятельного поиска агентом новых магазинов в сети Интернет посредством поисковых систем. Таким образом, расширяется база данных агента, что повышает эффективность его дальнейшей работы при повторным обращении к этим магазинам.

Список литературы: 1. Рассел С., Норвиг П. Искусственный интеллект: современный подход, 2-е изд.. Москва: Издательский дом «Вильямс», 2006. 1408 с. 2. Люгер Д.Ф. Искусственный интеллект: стратегии и методы решения сложных проблем, 4-е изд.. Москва: Издательский дом «Вильямс», 2003. 864 с. 3. Тарасов В.Б. От многоагентных систем к интеллектуальным организациям: философия, психология, информатика. — М.: УРСС, 2002.

4. А.А. Морозов, Ю.В. Обухов. Разработка математического аппарата логического программирования интеллектуальных агентов интернет // Журнал радиоэлектроники № 10, 2003. 5. Ю.В. Рогушина Использование онтологического описания предметной области для повышения релевантности информационного поиска // Проблемы программирования (4). 2003.

Поступила до редколегії 15.09.2008

УДК 519.6

Збір і обробка інформації в Інтернет з використанням інтелектуальних агентів / В.В. Дудар, М.В. Збітнева, А.Г. Дерев'янка // Біоніка інтелекту: наук.-техн. журнал — 2008. — № 2 (69). — С. 61-66.

Дана стаття присвячена розгляду питань розробки елементів агентної технології, таких як методи та алгоритми. Запропоновано удосконалений метод обробки інформації в Інтернет торговим агентом, реалізацію даного методу у вигляді алгоритму, архітектуру програмного застосування.

Бібліогр.: 5 найм.

УДК 519.6

Collection and treatment of information in internet with the use of the intellectual agents / V. Dudar, M. Zbitneva, A. Derevyanko // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 61-66.

This article is devoted to consideration of questions of development of elements of agent technologies, such as methods and algorithms. An improved method of information treatment by the commercial agent in Internet, realization of this method as an algorithm, architecture of program appendix are offered.

Ref.: 5 items.

УДК 681.324

**КОНЦЕПЦІЯ ЗАСТОСУВАННЯ GRID-ТЕХНОЛОГІЇ НА ТРАНСПОРТІ****В.О. Алексієв**

Харківський національний автомобільно-дорожній університет,
м. Харків, Україна, mechatronics@list.ru

Проведено аналіз створення комп'ютеризованої системи забезпечення розвитку транспортної інфраструктури. Розглянуто концептуальне обґрунтування застосування GRID-технології для інформаційної підтримки інтелектуальних транспортних технологій. Запропоновано шляхи сумісного використання гетерогенних комп'ютерних ресурсів різних транспортних організацій.

КОМП'ЮТЕРНІ РЕСУРСИ, GRID, ТРАНСПОРТНА ІНФРАСТРУКТУРА, РОЗВИТОК**Вступ**

Сучасна транспортна інфраструктура міст та регіонів є сукупністю транспортних засобів, шляхів сполучення, дорожнього середовища та інтелектуальних систем керування дорожнім рухом. До неї входять телематичні системи, які надають оперативну інформацію про стан дорожнього середовища та дозволяють взаємодіяти із всіма учасниками дорожнього руху. Для розвитку та експлуатації таких транспортних систем потрібні додаткові потужні комп'ютерні ресурси. Джерелом таких ресурсів може бути застосування існуючих регіональних комп'ютерних корпоративних мереж, що підтримують новітні GRID-технології. Розглянемо концептуальне обґрунтування отримання таких додаткових комп'ютерних ресурсів для розвитку транспортної інфраструктури великого міста за рахунок доступу до регіональних комп'ютерних систем.

1. Стан проблеми

Обчислювальні комп'ютерні ресурси на транспорті є порівняно спеціалізованим програмно – апаратним комплексом. Він виконує завдання інформаційної підтримки управління роботою різномірних локальних обчислювальних мереж інформаційно-обчислювальних центрів, які керують роботою терміналів автоматизованих робочих місць системи диспетчерського управління рухом транспортних одиниць. Потрібно надання якості гнучкості таким мережам з метою їх розвитку та адаптації до вимог безперервного розвитку транспортної інфраструктури міст та регіонів. Ці якості необхідні для впровадження новітніх інформаційних інтелектуальних технологій управління обчислювальними процесами з метою придбання додаткових комп'ютерних ресурсів за рахунок оптимального планування розподілу завдань в системі. Більшість існуючих систем та мереж на транспорті мають такі можливості, що особливо властиво для ВНЗ транспортного профілю з порівняно нерівномірним завантаженням обчислювальних потужностей. Це має відношення не тільки до транспортних організацій та підприємств.

Об'єднання усіх комп'ютерних ресурсів великих міст та регіонів у загальній обчислювальній мережі дозволить підвищити ефективність використання комп'ютерних ресурсів для організації швидкого та своєчасного електронного обороту даних. Для більшості організацій або їх підрозділів таке одночасне об'єднання гетерогенних комп'ютерних ресурсів на фізичному, логічному та організаційному рівні, як правило, є проблемою. Її вирішення – методологія GRID [1].

2. Джерела розробки

Ключова проблема GRID – розподіл завдань між обчислювальними вузлами мережі та динамічне планування, управління розподіленим обчислювальним процесом визначені недостатньо. Відомі підходи до цього обмежені застосуванням кластерних рішень, що потребують значної доробки програмного забезпечення окремих завдань. Впровадження GRID потребує розвитку інформаційного аналізу та синтезу складних обчислювальних систем, узагальнення, уніфікації, структурування та віртуалізації обчислювальних процесів, удосконалювання топології складових інформаційного простору. Такий комплексний підхід має сполучатися з застосуванням класичних рішень теорії розподілених систем [2,3,4], віртуалізації у комп'ютерних системах, планування обчислювальних процесів [5,6,7].

Змістовно концептуальне становлення GRID базується на самоорганізації, отриманні синергетичного ефекту від сумісної роботи різних комп'ютерів однієї мережі. Сполучення принципу віртуалізації ЕОМ з самоорганізацією і синхронізацією обчислювальних процесів дозволяє організувати сумісну роботу різних комп'ютерів такої гетерогенної системи, що базується на узагальненні цієї проблеми у фундаментальних дослідженнях [8,9,10]. Однак у прикладних додатках впровадження GRID – систем на транспорті слід урахувати відповідне концептуальне обґрунтування властивостей сучасної транспортної інфраструктури [11,12].

Слід відзначити, що транспортна інфраструктура великих міст та регіонів розвивається у напрямку

автоматизації, впровадження засобів мехатроніки, телематики, застосування штучних нейронних мереж, інтелектуалізації систем управління та зв'язку [13, 14, 15]. Розглянемо постановку задачі на впровадження GRID – систем на транспорті як концептуальне визначення основних напрямів цієї роботи.

3. Постановка задачі

В основі формального визначення принципів застосування GRID полягає така аксіоматика. Припустимо, що комп'ютерна система складається з елементів керування – P , виконання – U , інформаційних елементів – I , загальної структури – S . Якщо структури не існує, то $S=0$. Введемо поняття операції додавання (+), зменшення (–), операцію зв'язку ($\rightarrow$). Операція $+(–)$ означає введення в систему додаткових елементів, ($\rightarrow$) оцінює спрямованість інформації або підпорядкованість елементів. Послідовність операцій визначають дужки як засіб формування груп елементів.

Будь – яке додавання елементів одного типу не змінює функції групи цих елементів:

$$P + P = P; U + U = U; I + I = I. \quad (1)$$

Якщо прийняти, що S_L – лінійна структура; S_M – матрична; S_i – інформаційна; (=) – знак еквівалентності, то

$$\left. \begin{aligned} S_L &= P + U, S_i = (P \leftrightarrow +Y,) \\ S_M &= (P + U) + \dots + (P + U) \dots \end{aligned} \right\}. \quad (2)$$

Тоді структура S існує тільки у тому випадку, якщо є хоча б один елемент P та елемент U , а додавання елемента I не змінює якості об'єкту, вузла або модуля комп'ютерної системи. Структура – не тільки кількісна, але і якісна побудова. Елемент, що введено у конкретну структуру, має характеристики, відмінні від тих, котрі він міг би мати в іншій структурі $S=0$, якщо $P=0$ або $U=0$. Впливає: усяка структура є похідною від лінійної. Потенційно будь-яка структура такої організаційно – технічної системи має два функціональних рівня: елемент I теж знаходиться на рівні керування:

$$(P \leftrightarrow I) = P, \text{ а } S_i = S_L. \quad (3)$$

Аналогічно визначимо таку властивість як структурна стійкість комп'ютерної системи. Це значить, що тип і характеристики системи принципово не змінюються при доповненні, введенні в існуючу структуру ідентичних елементів:

$$\left\{ \begin{aligned} S\{[P + P = P], [U + U = U], \\ [I + I = I]\} = S(P, U, I), \end{aligned} \right\}. \quad (4)$$

де P, U, I – елементи керування, виконання й інформаційні відповідно.

З цього впливає властивість «структурного збігу»: структура системи, кількість рівнів, підсистем і ланок оптимальні тоді, коли його функ-

ціональна структура збігається з організаційною. Таким чином, інформаційні елементи визначають інформаційний характер будь-якої комп'ютерної інтелектуальної системи. Наслідком цього є властивість «ефективної реалізації». Система працює ефективно при однаковому фізичному і споживчому розрізненні її підсистем і ланок. Позначивши через $\xi_{\text{фіз}}$, $\xi_{\text{спож}}$ фізичний і споживчий порогові розрізнення результатів оцінки вузла, окремого комп'ютеру або іншої одиниці мережевого обладнання комп'ютерної системи, одержимо:

$$\xi_i = \xi_{i+1}, \text{ де } \xi_i \rightarrow \xi_{\text{фіз}}, \text{ а } \xi_{i+1} \rightarrow \xi_{\text{спож}} \quad (5)$$

Постановка задачі полягає у концептуальному обґрунтуванні отримання обчислювальних потужностей для підтримки транспортної інфраструктури за рахунок раціонального розподілу завдань комп'ютерам існуючого інформаційного простору корпоративних мереж. Цей розподіл має відповідати твердженням (1-5) та властивостям прозорості, віртуалізації, масштабування, з'єднання користувачів з ресурсами [2, 3, 4].

Розглянемо послідовно потрібні удосконалювання топології, особливості застосування операційних систем, моніторинг та впровадження GRID – систем на транспорті

4. Удосконалювання топології

Будь-які комп'ютерні ресурси організацій та підприємств, що мають стабільні тенденції до розвитку, мають і тенденції розвитку комп'ютерних ресурсів. У ролі їх обчислювального середовища можливості координації використання таких розподілених ресурсів покладають на Grid-технології. Вони надають можливості застосування різноманітних ресурсів: обчислювальних, накопичення даних та комунікаційних. Розвиток топології обчислювальних мереж практично проходить три рівня: Intragrid (внутрішній Grid); Extragrid (зовнішній Grid, що об'єднують вже декілька організацій); найвищий за рівнем масштабу розміру Grid-систем є Intergrid (глобальні системи, які об'єднують вже багато організацій, партнерів, кластерних рішень, рис. 1).

Саме цей рівень Grid має відповідати за розвиток транспортної інфраструктури великого міста або регіону. Звичайно таке об'єднання координується Grid-системою, а відповідна віртуальна мережа дозволяє технічно об'єднати розрізнені внутрішні Grid та кластери у єдиний інформаційний простір, що координується вже єдиною Grid – системою, що надається користувачеві як єдина віртуальна платформа. Відповідна топологія на нижньому рівні – окрема обчислювальна лабораторія, у якій користувач через механізм віртуальної ЕОМ має доступ до Intergrid – ресурсів (рис. 2).

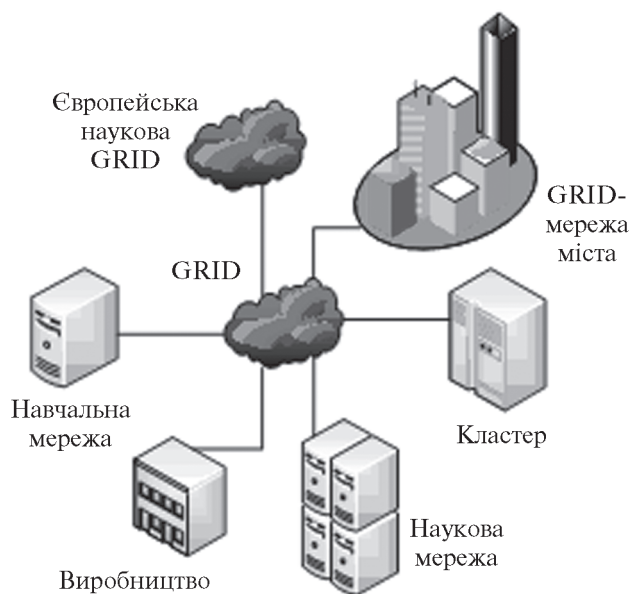


Рис. 1. Intergrid-система



Рис. 2. Окрема лабораторія

Своєрідне занурення автоматизованих робочих місць окремої транспортної лабораторії до комп'ютеризованого простору існуючих у будь-якому великому місті окремих систем, що мають Intergrid-ресурси, дозволяє отримати значний зиск від використання отриманих таким чином додаткових комп'ютерних ресурсів. У цьому випадку немає необхідності створення великої корпоративної мережі для підтримки транспортної інфраструктури. Причому надійність та продуктивність окремих систем може бути порівняно невеликою, але користувач такої розподіленої системи отримує єдину надійну та продуктивну платформу для обчислень, отримання доступу до баз даних та знань, також може зберігати свої дані та користуватися різноманітними комунікаційними технологіями. Такого зиску потребує залежність від цілеспрямованого розвитку архітектури мережі та від визначення програмної платформи, типу та властивос-

тей відповідної окремої або мережевої операційної системи.

5. Програмна платформа

Якщо проаналізувати тенденцію використання в існуючих комп'ютерних системах різних операційних систем, то можна стверджувати про перевагу Microsoft Windows (Windows XP, рис.3). Інші версії цієї операційної системи використовуються рідше. Це обумовлено тим, що старі версії працюють на комп'ютерах, які мають невелику продуктивність обчислень. Новітні версії операційних систем впроваджуються порівняно повільно, завдяки необхідності підвищення кваліфікації користувачів. Також можна затверджувати, що новітні операційні системи, як правило, встановлюються на нове обладнання, що теж опосередковано впливає на популярність їх застосування.

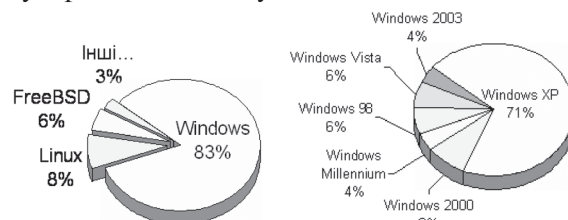


Рис. 3. Застосування операційних систем

Однак більшість реалізацій розподілених додатків та GRID-систем зараз реалізовано для платформи Linux. Тому доцільним є розглянути ці системи у якості альтернативи чи додаткової операційної системи.

На початковому етапі впровадження GRID-технологій з практичної точки зору найбільш зручним та багатофункціональним може бути застосування дистрибутиву Instant-Grid версії 1.0 (<http://www.instant-grid.org>). Він базується на системі Knoppix 5.0.0 (Live-CD) та системі Globus Toolkit 4.0.3 (програмне забезпечення проміжного рівня – middleware, що забезпечує можливість застосування GRID-технологій). Система Instant-Grid має зручний інтерфейс користувача, який оснований на використанні технологій подання даних web-браузером. Ця система поєднує як програми, що виконуються у консольному режимі, так і графічні додатки. У дистрибутиві реалізовано багато рішень Globus Toolkit, наприклад: компонент WSGRAM для управління задачами, що вирішуються у системі, система GridFTP для обміну файлами та система віддаленого входу до машин-клієнтів по захищеному каналу GSISSH. Таким чином, можна визначити концепцію застосування програмного забезпечення для окремої комп'ютерної лабораторії, що входить до складу розвиненої мережі універсального призначення. Впровадження GRID-технологій потребує використання мережевої операційної системи. На робочих станціях можна

виконати подвійну установку операційних систем. Одна буде із родини Windows, а інша — на базі вільної операційної системи типу Linux.

6. Моніторинг та впровадження

Розвиток гетерогенних комп'ютерних ресурсів передбачає принцип конкуренції, що забезпечує «виживання» найбільш ефективних зв'язків згідно з властивістю самоорганізації. У такому випадку можливо, що у комп'ютерних мережах — подібно тому, як це відбувається в інших синергетичних системах, — раптово виникнуть нові більш ефективні об'єднання комп'ютерів та розподіл завдань між окремими комп'ютерами. Особливо це важливо для вирішення проблем комп'ютерного забезпечення розвитку транспортної інфраструктури міст та регіонів. Саме комп'ютерні ресурси у транспортних системах мають гетерогенний характер та потребують застосування GRID-технологій. Застосування GRID-технологій в існуючих комп'ютерних системах будь-якої обчислювальної мережі це чітке визначення її топології та засобів моніторингу серверних додатків для раціональної маршрутизації завдань. Проведена автором оцінка навантаження окремих вузлів обчислювальної мережі транспортного ВНЗ, що виконує завдання різних користувачів у розрізі доби, тижня, місяцю, року, свідчить про досить нерівномірне навантаження комп'ютерного обладнання. Навіть візуальний аналіз результатів моніторингу окремих вузлів будь-якої мережі доводить, що їх навантаження досить нерівномірне, як, наприклад у випадку оцінки фізичної моделі корпоративної мережі (рис. 4) [16].

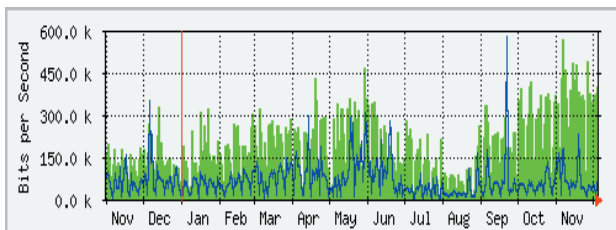


Рис. 4. Завантаження окремого вузла мережі: чорна лінія — вихідна інформація, сіра — вхідна

Така нерівномірність завантаження комп'ютерів усередині багатокомп'ютерної системи є джерелом отримання додаткових комп'ютерних потужностей. Для їх використання потрібна своєрідна рекуперація таких резервів комп'ютерних ресурсів. Вона полягає у одночасному використанні комп'ютерів для вирішення завдань різних користувачів, реалізації паралельної обробки інформації. Практично замість багатокомп'ютерної системи треба отримати синергетичний комп'ютер, принцип роботи якого засновано на розподілених обчисленнях, віртуалізації комп'ютерного середовища користувачів. Задача полягає в організації комп'ютерного сере-

довища існування GRID-технології, яке відповідає характеристикам типової (концептуальної) організації окремої комп'ютерної лабораторії, програмно-апаратним засобам відповідної обчислювальної мережі, яка складається з ланцюга Intragrid — Extragrid — Intergrid.

Тому застосування у будь-якій організації GRID-технологій, перш за все, вимагає виконання систематизації та уніфікації наявних комп'ютерних ресурсів. Для транспортної інфраструктури потрібна інвентаризація існуючих у регіоні комп'ютерних систем та одночасно повне уявлення про їх працездатність, що є основним завданням моніторингу комп'ютерних систем. Виконання спостереження за станом локальної обчислювальної мережі треба повторювати постійно. Дані, що отримані у цьому моніторингу, бажано зберігати для отримання статистичних даних про роботу мережі. Це надає можливості приймати рішення про масштабування локальної мережі, доповнення її новими користувачами.

Система, що базується на безперервному моніторингу усіх складових мережі, передбачає не тільки нормалізацію вимірювальної інформації, але і її структурний, змістовий (семантичний) та прагматичний аналіз. Наслідок — оцінка ефективності рішення відповідної задачі у залежності від попереднього стану комп'ютерного комплексу та прогнозування його зміни, що оцінюється.

На відміну від відомої послідовності перетворень сигналів теорії інформаційної техніки, у такій гнучкій комп'ютерній системі звичайні вимірювальні операції змінюються на процедури структурного, семантичного й прагматичного аналізу. Звичайна лінійна послідовність дій, прийнята в інформаційно-вимірювальних системах, доповнюється рекурсивними процедурами імітації й прийняття рішень і, як наслідок, — доповненням зворотного зв'язку у контурі прийняття оптимальних рішень щодо вибору раціонального шляху передачі даних та відповідного розподілу завдань машинам мережі (рис. 5). Така узагальнена схема проходження моніторингових сигналів у комп'ютерних системах передбачає не тільки нормалізацію вимірювальної інформації, але і її структурний, змістовий (семантичний) та прагматичний аналіз з точки зору оцінки ефективності рішення відповідної задачі у залежності від попереднього стану машини, що оцінюється.

Таким чином, ми одержуємо не просто інформаційну систему, а своєрідний інформаційний регулятор, що сполучає переваги систем програмного керування із системами, що працюють на основі синтезу керуючого впливу.

В таких мережах при виконанні перетворення великих обсягів інформації кожна машина працює автономно. Однак для вирішення проблем їх

сумісної роботи основним стає можливість обміну не тільки результатами цієї роботи, але і вирішенням завдань формування керуючої інформації. З одного боку, якщо ці комп'ютери жорстко з'єднані один з одним, вони знов-таки являють собою подобу великого комп'ютера. При цьому розподіл завдань між окремими комп'ютерами визначається засобом з'єднання. З іншого боку, ці комп'ютери автоматично встановлюють між собою нові зв'язки. У цьому випадку окремі комп'ютери мають послати іншим інформацію про те, звідки й куди спрямоване повідомлення і запит про готовність машини-адресата взяти на себе виконання певного завдання. У відповідь на цей сигнал комп'ютер – приймач, що одержав повідомлення, має дати відправникові відповідь про свою готовність (або неготовність) узятися за виконання поставленого завдання. Після обміну повідомленнями у випадку готовності відбувається направлення завдання з однієї машини на іншу. Такі дії вимагають деяких витрат комп'ютерних ресурсів мережі. Тому система повинна знайти оптимальне рішення, що має місце при самоорганізації шляхів спрямування інформації. Знайти рішення можна у тому випадку, коли у комп'ютерній системі постійно виконується моніторинг стану кожної машини, вузлів обробки та передачі даних.

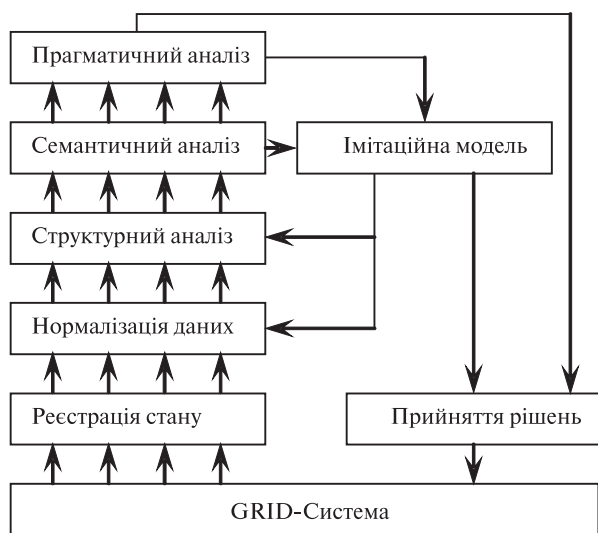


Рис. 5. Узагальнена схема моніторингу GRID

Комп'ютери, об'єднані в подібну мережу, розподіляють завдання між собою й вирішують їх паралельно, потім переходять до нового розподілу й тому подібне. У цьому випадку при розподілі завдань між окремими комп'ютерами виникають періодичні коливання, у результаті чого кожний окремий комп'ютер пропускає через себе колосальний потік інформації. Можливі ситуації її недостатньої стійкості (можливість виникнення коливань). Багаторазовий розподіл завдань у комп'ютерній мережі можна використовувати для

зміцнення спочатку слабкого зв'язку між окремими комп'ютерами. При цьому в дію вступає принцип конкуренції, що забезпечує «виживання» найбільш ефективних зв'язків за аналогією з процесами навчання згідно з теорією штучних нейронних мереж [13, 14].

Висновки та пропозиції

GRID-технології є складовою частиною інтелектуальної технології управління рухом транспортної інфраструктури міст та регіонів. Впровадження GRID-технологій у практику ведення комп'ютерних ресурсів базується на досвіді створення розподілених систем. Концептуальне обґрунтування застосування GRID-систем базується на прозорості, віртуалізації, масштабуванні, з'єднанні користувачів з ресурсами; захищеності та надійності передачі та обробки інформації та сумісному використанні комп'ютерних ресурсів різних організацій, які є гетерогенними системами. Ці принципи співпадають з властивостями комп'ютерних систем структурної стійкості, структурного збігу та ефективної реалізації, як управління нової комп'ютерної системи, що визначені у постановці задачі.

Першим етапом до застосування GRID-технологій у існуючій структурі обчислювальної мережі є чітке визначення її структурної схеми та засобів моніторингу серверних додатків. Ця визначеність підвищить рівень надійності, швидкодії та працездатності усіх інформаційних технологій, що підтримуються в існуючій системі. У цьому разі можна одержати додаткові комп'ютерні ресурси для вирішення задач комп'ютерного забезпечення транспортної інфраструктури паралельно з забезпеченням окремих обчислювальних процесів «донорської» організації, що є власником Intergrid-системи. GRID-технології дозволяють отримати додаткові комп'ютерні ресурси за рахунок відповідної структуризації локальної мережі, упорядкування комунікаційних каналів та підвищення рівня управління процесом моніторингу серверних додатків.

Це дослідження виконувалось за замовленням МОН України науково-технічної роботи «Методологія застосування GRID-технологій для проведення наукових досліджень у транспортному ВНЗ» (фінансування за рахунок бюджетних коштів).

Список літератури: 1. IBM Corp. (2005) Introduction to Grid Computing. SG24-6778-00. <http://ibm.com/redbooks>. 2. Э. Таненбаум, М. Ван Стеен. Распределенные системы. Принципы и парадигмы. СПб.: Питер, 2003. — 877 с. 3. Буров Є. Комп'ютерні мережі. 2-ге основне і доповнен. вид. Львів: Бак, 2003. — 548 с. 4. Олифер В.Г., Олифер Н.А. Компьютерные сети. Принципы, технологии, протоколы: Учебник для вузов. 3-е изд. — СПб.: Питер, 2006. — 958 с.

5. Спилман Д., Хадсон К., Крафт М. Планирование, внедрение и поддержка инфраструктуры Active Directory Microsoft Windows Server 2003. Учебный курс Microsoft: Пер. с англ. — М.: Русская Редакция, СПб.: Питер, 2006. — 656 с. 6. Хетч Б., Колесников О. Linux: создание виртуальных частных сетей (VPN): Пер. с англ. — М.: КУДИЦ — ОБРАЗ, 2004. — 464 с. 7. Алексієв В.О. Новітня GRID-технологія для вирішення задач дослідження мехатронних систем у автомобільно-дорожньому ВНЗ // Вестник ХНАДУ / Сб. науч. тр. — Харьков. — Вып. 38. — 2007. — С. 111-113. 8. Хакен Г. Тайны природы. Синергетика: учение о взаимодействии. — Москва-Ижевск: Институт компьютерных исследований, 2003. — 320 с. 9. Хакен Г., Хакен-Крелль М. Тайны восприятия. — М.: Институт компьютерных исследований, 2002. — 272 с. 10. Капица С.П., Курдюмов С.П., Малиновский Г.Г. Прогнозы будущего. — Москва: Едиториал УРСС, 2003. — 288 с. 11. Алексієв В.О. Інтелектуальна технологія моніторингу рухомих одиниць транспортної системи міста // Вестник ХНАДУ / Сб. науч. тр. — Харьков. — Вып. 25. — 2004. — С. 11-113. 12. Алексієв В.О. Концептуальний аналіз автомобільних мехатронних систем // Автомобільний транспорт. — Харків, ХНАДУ, вип. 16, 2005. — С. 321-323. 13. Пржибил П., Свитек М. Телематика на транспорте. — М.: МАДИ (ГТУ), 2003 — 540 с. 14. Руденко О.Г., Бодянский Е.В. Основы теории искусственных нейронных сетей. — Харьков: ТЕЛЕТЕХ, 2002. — 317 с. 15. Комашинский В.И., Смирнов Д.А. Нейронные сети и их применение в системах управления и связи. — М.: Горячая линия — Телеком, 2003. — 94 с. 16. Методологія застосування GRID-технологій для проведення наукових досліджень у транспортному ВНЗ // ХНАДУ, звіт з наукової роботи № держреєстрації 0107U008063. — Харків, 2007. — 78 с.

Надійшла до редколегії 17.09.2008

УДК 681.324

Концепция использования GRID-технологий на транспорте / В.О. Алексеев // Бионика интеллекта: науч.-техн. журнал. — 2008. — № 2 (69). — Р. 67-72.

Проведен анализ создания компьютерной системы, предназначенной для развития транспортной инфраструктуры. Рассмотрено концептуальное обоснование использования для этого новейшей GRID технологии. Определено соответствие принципов внедрения GRID-систем на транспорте и пути совершенствования архитектуры, выбора операционной системы, мониторинга вычислительных процессов распределенной компьютерной систем. Предложено совместное использование компьютерных ресурсов региональной транспортной инфраструктуры разными транспортными организациями.

Ил. 5. Библ.: 15 наим.

UDC 681.324

The concept of use GRID of technologies on transport / V.O. Alekseyev // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 67-72.

The analysis of creation of the computer system intended for development of a transport infrastructure is lead. The conceptual substantiation of use for this purpose by the newest GRID technologies is considered. Conformity of principles of introduction GRID-systems on transport and a way of perfection of architecture, choice of operational system, monitoring of computing processes of the distributed system is certain. Sharing computer resources of a regional transport infrastructure by the different transport organizations is offered.

Fig. 5. Ref.: 15 items.

УДК 681.5.015



ИНТЕЛЛЕКТУАЛЬНЫЕ ПРОЦЕДУРЫ УПРАВЛЕНИЯ БИЗНЕС-ПРОЦЕССАМИ ПРЕДПРИЯТИЙ

В.П. Авраменко¹, С.Ю. Калистратова²

^{1,2} Харьковский национальный университет радиоэлектроники, Харьков, Украина

Исследованы проблемы повышения эффективности функционирования предприятий за счет интеллектуализации процедур управления бизнес-процессами. Разработаны модели и интеллектуальные процедуры организационного управления бизнес-процессами предприятий.

ПРИНЯТИЕ УПРАВЛЕНЧЕСКИХ РЕШЕНИЙ, ИНТЕЛЛЕКТУАЛИЗАЦИЯ УПРАВЛЕНИЯ,
МОДЕЛИРОВАНИЕ БИЗНЕС-ПРОЦЕССОВ

Введение

Важная роль в повышении эффективности управления бизнес-процессами предприятий принадлежит средствам искусственного интеллекта, которые позволяют согласовать локальные качества составляющих принимаемого решения с глобальной целью управления с учетом условий функционирования и проявления внешней среды.

Запланированная эффективность функционирования предприятия может быть достигнута только в случае неизменности характеристик бизнес-процессов и отсутствия внешних воздействий. Наличие неопределенности исходных данных и действие внешней среды приводят к худшей эффективности бизнес-процессов, чем при полной определенности.

При наличии дрейфа бизнес-процессов и проявлении внешней среды управленческие решения принимаются не безусловно лучшие, а лучшие в определенном смысле, например, в среднем, или близкие к некоторому решению, предпочтительному с точки зрения лица, принимающего решения.

Для повышения эффективности управления бизнес-процессами в условиях неопределенности требуется накопление статистических данных, которые позволили бы построить модели дрейфа характеристик бизнес-процессов, а также выбрать наиболее предпочтительные методы и инструментальные средств, ориентированных на определенный тип возмущений бизнес-процессов.

Целью данной работы является создание и апробация интеллектуальных процедур управления бизнес-процессами предприятий, а также разработка и применение методологии и инструментальных средств интеллектуального управления бизнес-процессами различного назначения.

Анализ состояния проблемы интеллектуализации управления бизнес-процессами [1–3] свидетельствует об актуальности этого научного направления.

1. Проблема интеллектуализации управления

Основной процесс функционирования предприятия заменяется совокупностью составляющих

бизнес-процесса, каждая из которых вносит свою лепту в общую стоимость создаваемой продукции. Глобальный критерий эффективности функционирования предприятия при этом аппроксимируется совокупностью локальных критериев, каждый из которых относится к конкретной составляющей бизнес-процесса.

В качестве локальных критериев эффективности принимаются конкретные технико-экономические показатели качественного и количественного типа. Значения локальных критериев зависят от технологических, финансовых и информационных *ресурсов, времени* выполнения составляющих бизнес-процесса и его обслуживания, *координации* и *согласованности* потребления имеющихся бизнес-ресурсов.

Проблема интеллектуализации управления бизнес-процессами состоит в том, чтобы сформировать последовательность управляющих воздействий, которая обеспечит целенаправленное поведение соответствующей компоненты бизнес-процесса предприятия в соответствии с заданными локальными критериями качества. Для этого необходимо построить модели компонентов бизнес-процесса, вносящих определенное долевое участие в конечную стоимость произведенного продукта или услуги.

Перспективными направлениями управления производством на современном этапе научно-технического прогресса являются следующие основные подходы [3]:

- TQM (Total Quality Management) – система всеобщего управления качеством;
- PIQS (Process Integrated Quality System) – набор требований к системам менеджмента качества;
- WFMS (Work Flow Management System) – система управления потоками работ;
- ERP (Enterprise Resource Management System) – система планирования и управления ресурсами.

2. Результаты выполненных исследований

Инструменты уровня управления бизнес-процессом должны поддерживать абстрактное описание логики выполнения операций бизнес-про-

цесса, средства представления этих операций в информационных системах, интеграцию отдельных операций в единый процесс и мониторинг его выполнения. Если все эти требования выполнены, появляется возможность вносить изменения в бизнес-логику, меняя определение процесса. Другими словами, можно менять компоновку операций, не внося изменений в программный код соответствующих приложений. Таким образом, становится реальностью быстрая трансляция динамики бизнеса в поддерживающую его прикладную инфраструктуру.

Предлагается апробированная методика создания автоматизированной системы управления бизнес-процессами предприятия, которая основана на использовании компьютерных технологий и процессного подхода к управлению. В основе процессного подхода к управлению предприятием лежит выделение в предприятии бизнес-процессов, пересекающих совокупность функциональных подразделений предприятия, и управление этими бизнес-процессами посредством информационных технологий.

Концептуальная модель предприятия на основе бизнес-процессного описания определяет логику последовательной детализации бизнес-процессов предприятия. На первом этапе строится самая

грубая модель предприятия — функциональная модель, дающая представление о функциях предприятия и о распределении ответственности за их выполнение. Функциональная модель предприятия не содержит информации о взаимосвязях функций. Для этого строится процессная модель, которая дает описание потоков ресурсов (материальные, информационные и так далее), используемых при выполнении функций, но она не дает их стоимостную оценку, что необходимо иметь для расчета финансового плана предприятия. Поэтому необходимо уточнить описание бизнес-процессов с помощью построения количественной модели бизнес-процессов, которая определяет стоимость выполнения бизнес-процесса. Построив количественную модель бизнес-процессов, можно построить и модель финансового плана предприятия (рис. 1).

Чтобы построить эффективную систему процессно-технологического управления, необходимо однозначно определить все объекты, основным из которых является объект управления. Объектом процессного управления может выступать один или несколько бизнес-процессов.

Интеллектуальное управление бизнес-процессами возможно только при участии человека, который способен четко сформулировать цель уп-



Рис. 1. Основные бизнес-процессы производственного предприятия

правления, имеет доступ к информации о реальных значениях параметров составляющих бизнес-процесса и наделен полномочиями принятия управленческих решений.

Компьютерный мониторинг и аудит составляющих бизнес-процесса улучшает наблюдаемость и управляемость бизнес-процесса, что в конечном итоге дает возможность принятия альтернативных управленческих решений. Выделенный уровень бизнес-логики позволяет компоновать бизнес-процессы из подпроцессов, тем самым делегируя ответственность за выполнение заданных бизнес-правил различным подразделениям предприятия.

Применение программных платформ с комплексной интеллектуализацией управления бизнес-процессами дает возможность не только повысить эффективность бизнес-процессов, но и позволяет формировать среды постоянного совершенствования этих процессов.

Блок-схема интеллектуализированного управления бизнес-процессами предприятия приведена на рис. 2. Этой графической модели однозначно соответствует модель логического регулярного выражения алгебры событий. Алгебра событий определяет для множества всех событий операции дизъюнкции, конъюнкции и итерации.

Дизъюнкция двух событий $S_1 \vee S_2$ представляет собой теоретико-множественное объединение событий S_1 и S_2 , означающее наступление одного из событий S_1 или S_2 . Произведение двух событий $S_1 \circ S_2$ представляет собой событие, состоящее из всех слов вида pq , где p — любое слово события S_1 , а q — любое слово события S_2 . Итерацией события S является событие, состоящее из пустого слова ϵ и всевозможных слов вида $p_1 p_2 \dots p_n$, где $p_1, p_2, \dots, p_n$ — произвольные слова события S , а n — любое натуральное число.

Аппарат регулярных выражений составляют алгебра операторов R_{on} и алгебра условий R_{yc} . Элементами алгебры операторов R_{on} являются все операторы Y_j , определяющие деятельность бизнес-процесса и обозначающие конкретные акты деятельности по преобразованию входных ресурсов в выходной продукт. Кроме основных операторов Y_j , в алгебру R_{on} входят два вспомогательных оператора: тождественный оператор E и пустой оператор $\emptyset$.

Элементами алгебры условий R_{yc} являются все логические условия X_i , которые принимают значения только $X_i = 1$ или $X_i = 0$. Кроме основных условий X_i , в алгебру R_{yc} входят два вспомогательных условия: «1» — тождественно-истинное значение и «0» — тождественно-ложное значение.

Представление любого оператора из алгебры R_{on} через алфавит этой алгебры, содержащий все операторы и условия на множестве n , называется регулярным выражением алгебры событий (РВАС) и обозначается

$$R = (Y_j, X_i, E, \text{III}, 1, 0), \quad (1)$$

причем условным и линейным операторам РВАС соответствуют переменные, последовательности операторов — операции их произведения (конкатенации), параллельно соединенным ветвям — операция дизъюнкции переменных, циклам — операция итерации переменных; условные операторы указываются в скобках для того, чтобы отличить их от линейных операторов.

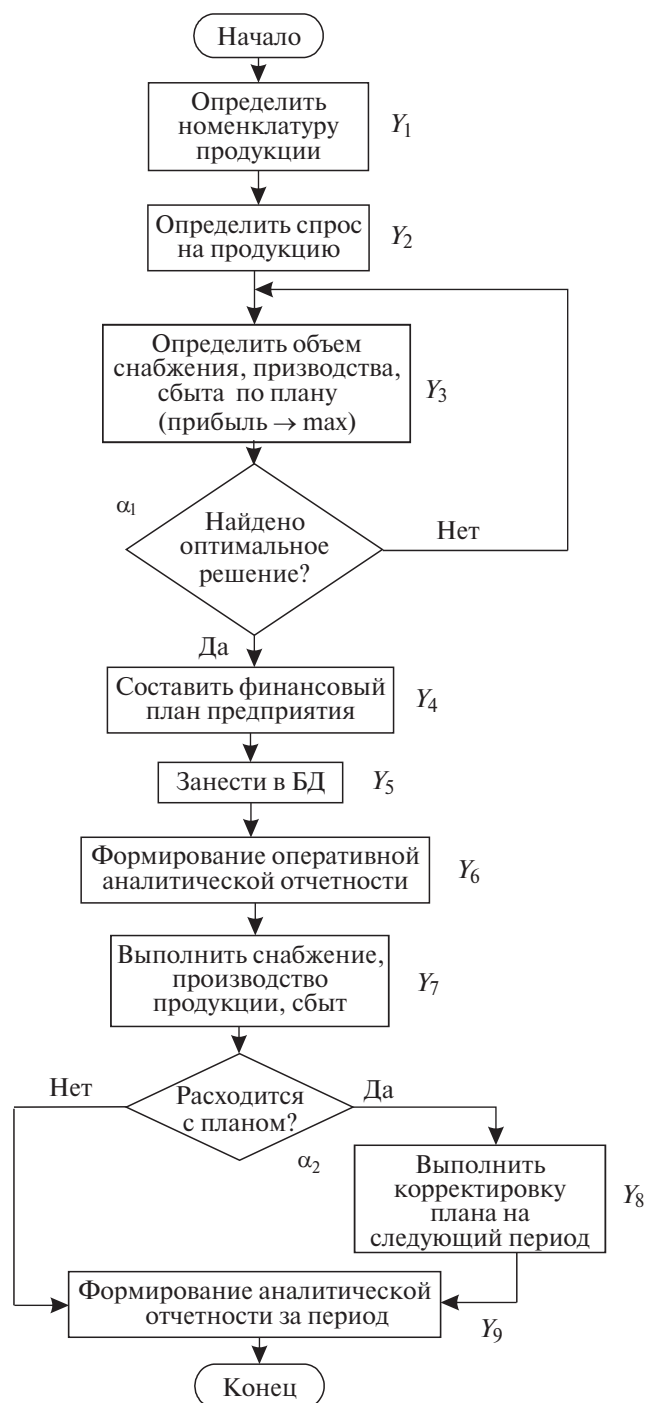


Рис. 2. Математическая модель управления бизнес-процессами предприятия

Математическая модель бизнес-процессов оперативного управления выполнением план-графи-

ков производства (рис. 1), построенная в классе регулярных выражений алгебры событий, имеет следующий вид:

$$R = Y_1 \cdot Y_2 \cdot Y_3 \cdot {}_{\alpha 1}(Y_3 \vee Y_4)^{\alpha 1} \cdot Y_5 \cdot Y_6 \cdot Y_7 \cdot {}_{\alpha 2} \\ \cdot (Y_9 \vee Y_8 \cdot Y_9)^{\alpha 2}, \quad (2)$$

где Y_1 — оператор определения номенклатуры продукции; Y_2 — оператор определения спроса на продукцию; Y_3 — оператор определения объемов снабжения, производства, сбыта по плану; Y_4 — оператор составления финансового плана предприятия; Y_5 — оператор занесения данных в БД; Y_6 — оператор формирования оперативной аналитической отчетности; Y_7 — оператор выполнения снабжения, производства продукции, сбыта; Y_8 — оператор выполнения корректировок плана на следующий период; Y_9 — оператор формирования аналитической отчетности за период; α_1 — условный оператор проверки нахождения оптимального решения; α_2 — условный оператор проверки расхождения с планом.

Заключение

Впервые разработаны методология, алгоритмы и эффективные вычислительные процедуры управления бизнес-процессами производства за счет использования функционально-логических алгоритмов управления, построенных на основе регулярных выражений алгебры событий.

Список литературы: 1. Авраменко В.П., Калачева В.В. Интеллектуальные технологии повышения эффективности бизнес-процессов // Бионика интеллекта: науч.-техн. журнал. — № 1 (66), 2007. С. 95–97. 2. Андерсен Б. Бизнес-процессы. Инструменты совершенствования. — М.: Стандарты и качество, 2004. — 272 с. 3. Железо Б.А., Ермакова Т.А. Реинжиниринг бизнес-процессов. — Минск: Книжный дом Мисанта, 2006. — 213 с. 4. Репин В.В., Елиферов В.Г. Процессный подход к управлению. Моделирование бизнес-процессов. — М.: Стандарты и качество, 2004. — 398 с.

Поступила в редколлегию 18.09.2008

УДК 681.5.015

Інтелектуальні процедури управління бізнес-процесами підприємств / В.П. Авраменко, С.Ю. Калістратова // Біоніка інтелекту: наук.-техн. журнал — 2008. — № 2 (69). — С. 73–76.

Досліджено проблеми підвищення ефективності функціонування підприємств за рахунок інтелектуалізації процедур управління бізнес-процесами. Розроблено моделі та інтелектуальні процедури організаційного управління бізнес-процесами підприємств.

Іл.: 2. Бібліогр.: 4 назв.

UDK 681.5.015

The intelligence procedures of control business processes of enterprises / V.P. Avramenko, S.U. Kalistratova // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 73–76.

The paper is devoted to the research of increase of the functioning of enterprises through intellectualization management business processes. There are mathematical models and computational procedures intellectualization management of business processes of enterprises.

Fig.: 2. Ref.: 4 items.

УДК 004.8:004.89



ОЦЕНИВАНИЕ КАЧЕСТВА ИДЕНТИФИКАЦИОННЫХ И ПРОГНОЗНЫХ РЕШЕНИЙ В ИНЖЕНЕРИИ КВАНТОВ ЗНАНИЙ

И.Б. Сироджа

Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», г. Харьков, Украина

Рассматриваются плохо формализованные задачи знаниеориентированного принятия идентификационных и прогнозных решений средствами инженерии квантов знаний в условиях неопределенности. Благодаря введенной логической модели экстраполяции наблюдений в виде системы импликативных закономерностей предметной области как базы знаний, задача прогнозирования в частном случае сводится к задаче идентификации ситуаций. На этой основе предложена единая количественная мера качества идентификационных и прогнозных решений, дедуктивно выводимых из предварительно построенной импликативной базы квантов знаний (БкЗ).

ИНЖЕНЕРИЯ КВАНТОВ ЗНАНИЙ, ИМПЛИКАТИВНЫЕ ЗАКОНОМЕРНОСТИ, БАЗА ЗНАНИЙ, ИДЕНТИФИКАЦИОННЫЕ И ПРОГНОЗНЫЕ РЕШЕНИЯ, ДЕДУКТИВНЫЙ ВЫВОД

Введение

Квантовый подход к инженерии знаний в искусственном интеллекте привел к образованию нового научного направления — *инженерии квантов знаний* (ИКЗ) для решения плохо формализованных задач принятия *идентификационных* и *прогнозных* решений в условиях неопределенности [1, 2]. В отличие от существующих подходов, основанных на фреймах, продукциях, семантических сетях и предикатных исчислениях [4, 5], ИКЗ базируется на представлении знаний посредством алгоритмических структур 0-го, 1-го и 2-го уровней сложности, имеющих три составляющих — *семантическую*, *информационную* и *операторную*. Эти структуры названы *разноуровневыми алгоритмическими квантами знаний* (РАКЗ). Условно кванты знаний (*к-знания*) относятся к 0-му уровню, если они представимы *числом* или *символом*; к 1-му уровню, если — *числовым* или *символьным вектором* и ко 2-му уровню, если — *числовой* или *символьной матрицей*. Важнейшим преимуществом РАКЗ-моделей представления и преобразования *к-знаний* является возможность их описания в множественной, векторно-матричной и аналитической (конечно-предикатной) формах, что позволяет манипулировать знаниями посредством машинных алгебр и процедур логического вывода. Основы теории, практические приложения ИКЗ, а также оценивание качества знаниеориентированных решений посредством вычисления риска ошибочного решения по контрольной выборке подробно изложены в монографии [2].

В данной работе предложена единая количественная *мера качества идентификационных и прогнозных* решений, определяемая эмпирической степенью корреляции между наблюдаемыми и модельными прогнозными процессами. Такая мера позволяет обоснованно судить о потенциальной *предсказуемости* динамических процессов и об основных причинах неудачных прогнозов.

1. Особенности средств инженерии квантов знаний

Главная особенность математических РАКЗ-моделей и методов ИКЗ состоит в представлении базы *к-знаний* (БкЗ) системой *импликативных* и/или *функциональных* закономерностей, которые находятся индуктивно по выборочным обучающим *к-знаниям* из предметной области. *Импликативной* закономерностью или *запретом* *r*-го ранга согласно [2, 5], называется такая устойчивая логическая связь между *r* двоичными признаками из общего их числа *n* ($2 \leq r \leq n$) $x_1, x_2, \dots, x_n$, характеризующих объект принятия решений (ОПР), когда для реальных объектов недопустима (запрещена) хотя бы одна комбинация $\binom{n}{r}$ значений признаков. *Функциональная* связь как частный случай вытекает из *импликативной*. Их отличие состоит в том, что при *функциональной* связи значения признаков-аргументов всегда однозначно определяют значение признака-функции, а при *запретной* связи — не всегда, только при определенных комбинациях значений аргументов. Все эти связи и характеристики ОПР представляются квантами знаний соответствующих уровней. Если указанные признаки $x_1, x_2, \dots, x_n$ принимают значения из конечных двоичных множеств $X_1, X_2, \dots, X_n$, то декартово произведение $X_1 \times X_2 \times \dots \times X_n$ образует бинарное пространство *B*, элементами которого служат всевозможные наборы элементов, взятых по одному из множеств $X_i, i = \overline{1, n}$. Наборы значений признаков, отвечающие *реальным* объектам, образуют подмножество $B_r \subset B$, а отвечающие *выборочным обучающим* объектам — подмножество $B_0 \subset B_r$. Очевидно, мощности множеств B_0, B_r и *B* находятся в отношении «<<» — «намного меньше»:

$$|B_0| \ll |B_r| \ll |B|. \quad (1)$$

Кванты знаний об ОПР и связях между их признаками удобно интерпретировать как *интервалы*

пространства B или соответствующие элементарные конъюнкции. Из соотношения (1) следует, что *импликативные* закономерности для искомой БкЗ можно обнаруживать путем нахождения *запретных интервалов* во множестве B_0 как *обучающей* выборке. Действительно, если существует связь между признаками, которой отвечает *запретный* интервал J в пространстве B , то должно выполняться отношение $J \cap B_r = \emptyset$, а значит, и отношение $J \cap B_0 = \emptyset$, то есть интервал J – пустой для B_0 , так как в обучающей выборке отсутствует соответствующая запретная комбинация значений признаков. Таким образом, если $J \cap B_0 = \emptyset$, то можно выдвинуть гипотезу о существовании запретной связи определенного ранга r , то есть числа связанных переменных в соответствующей элементарной конъюнкции. Достоверность гипотезы оценивается величиной математического ожидания $M_z\{m, n, r\}$ [2] как средним числом пустых интервалов ранга r , возникающих в пространстве B при отсутствии каких-либо закономерностей, когда B_0 объемом $m \times n$ (m – количество квантов-наблюдений, n – число признаков ОПР) выбирается из B случайно с равновероятным выбором элементов. Очевидно, что гипотеза принимается лишь при достаточно малом значении величины $M_z\{m, n, r\}$ из интервала $[0, 1]$, не большего от заданного допустимого порога M_z^* при ранге r запрета, не превышающего $r_{\max}$. Минимизированную и безизбыточную БкЗ представляют в виде матричного запретного кванта знаний 2-го уровня $k_2 \bar{\Sigma}_{BM}$, строки которого описывают определенные запреты и могут интерпретироваться как интервалы пространства B либо как отвечающие им элементарные конъюнкции [1, 2]. Следовательно, *импликативные* k -знания $k_2 \bar{\Sigma}_{BM}$ в целом определяют запретную область $\bar{\Sigma}_B$ в пространстве B , которая описывается характеристической функцией запрета (ХФЗ) Φ в виде конечного бинарного предиката, принимающего значение «1» только на элементах области $\bar{\Sigma}_B$. Предикат Φ удобно выразить в дизъюнктивной нормальной форме (ДНФ) как дизъюнкции элементарных конъюнкций, отвечающих найденным *запретам*. Именно представление БкЗ в виде ХФЗ Φ позволяет решать задачу прогнозирования динамических процессов как задачу экстраполяции интересующих свойств ОПР по имеющимся наблюдениям и в частном случае получить решение задачи идентификации состояний процессов, то есть распознавания объектов. Специфику предлагаемой экстраполяции поясним геометрическими соображениями с неявным учетом времени прогноза с $\tau_{\text{пр}}$ в отличие от общепринятого использования явной зависимости наблюдаемого процесса $Y(t)$ от времени t .

Пусть известна некоторая БкЗ $= k_2 \bar{\Sigma}_{BM}$ в форме ХФЗ Φ , устойчивость закономерностей которой гарантируется на заданное время прогноза τ . Ре-

зультаты частичных наблюдений заданы векторным квантом знаний 1-го уровня $k_1 y_\omega$, представляющим интервал y_ω пространства $B_r \subset B$, в котором локализован ОПР ω . Очевидно, что если этот интервал пересекается с запретной областью $\bar{\Sigma}_B$ данной БкЗ, то путем удаления из y_ω области пересечения получим остаток $\hat{y}_\omega = y_\omega \setminus (\bar{\Sigma}_B \cap y_\omega)$, объем и форма которого определяют возможный результат экстраполяции. Если интервал y_ω не пересекается с $\bar{\Sigma}_B$, то имеем результат «не знаю» в силу недостатка знаний в БкЗ. В случае полного пересечения y_ω с $\bar{\Sigma}_B$ приходим к результату, свидетельствующему о противоречивости БкЗ. Формально такое использование системы запретов БкЗ $= \bar{\Phi}$ при экстраполяции данных, представленных квантом наблюдений $k_1 y_\omega$, полностью реализуется оператором редукции БкЗ $k_1 y_\omega$ вида $RED(k_2 \bar{\Sigma}_{BM}, k_1 y_\omega)$. Получаемые при этом редуцированные k -знания $k_1 \hat{y}_\omega = RED(k_2 \bar{\Sigma}_{BM}, k_1 y_\omega)$ содержат необходимую информацию для принятия *идентификационных* (C -задача) и *прогнозных* (B -задача) решений при анализе текущих ситуаций. B -задача состоит в определении (предсказании) неизменных значений s признаков ОПР по известным значениям $(n-s)$ оставшихся признаков из общего числа n путем редукции БкЗ $= \bar{\Phi}$ по кванту наблюдений $k_1 y_\omega$ в данной предметной области. В C -задаче как в частном случае B -задачи требуется определить (предсказать) неизвестное значение i -го идентификационного признака ($i = \overline{1, n}$) по измеренным значениям $(n-1)$ -го признака распознаваемого класса объектов. Отсюда следует возможность построения единой количественной меры качества принимаемых прогнозных и идентификационных решений как некоторой степени сходства между наблюдаемым процессом $Y(t)$ и модельным прогнозным процессом $Z(t)$. Процессы Y и Z связаны с не регистрируемым напрямую реальным процессом $X(t)$ при знаниеориентированном решении проблемы предсказуемости. Количественно степень сходства между $Y(t)$ и $Z(t)$ будем представлять показателем корреляции между наблюдением и прогнозом, спустя время τ после начала наблюдения. Далее сформулируем рассматриваемую задачу оценивания качества знаниеориентированных решений и приведем методику ее решения средствами ИКЗ с обсуждением вопросов, связанных с предсказуемостью динамических процессов.

2. Постановка задачи

Согласно изложенным в пункте 1 посылкам, содержательная постановка задачи оценивания качества знаниеориентированных идентификационных и прогнозных решений состоит в определении единой количественной меры W их качества, опираясь на возможность решения B -, C -задач пу-

тем редуцирования характеристической функции $\bar{\Phi}$ имплицитивной БкЗ с неявным учетом времени. При этом мера W должна характеризовать потенциальную предсказуемость наблюдаемого (регистрируемого) процесса Y с использованием модельного процесса Z и знаний о реальном процессе X , который не подвергается прямой регистрации, а также позволять судить об основных причинах некачественного прогноза. Заметим, что динамический процесс $Z(t)$ в нашем случае представляется РАКЗ-моделью в виде $X\Phi Z \bar{\Phi}$, где время t не является явным аргументом, а лишь определяет временной период гарантированной устойчивости системы закономерностей БкЗ. При этом $X\Phi Z \bar{\Phi}$ одновременно выполняет роль БкЗ и служит аксиоматической основой дедуктивного механизма вывода искомых решений.

Формальная постановка задачи заключается в следующем. Пусть наблюдаемый процесс $Y(t)$ связан с реальным процессом $X(t)$ соотношением

$$Y(t) = X(t) + \sigma_v + \sigma_f + \sigma_{\Delta Q}, \quad (2)$$

где величины σ_v , σ_f и $\sigma_{\Delta Q}$ отражают соответственно искажения от инструментальных шумов при измерениях, от флуктуационных физических шумов и от влияния неадекватности модели прогнозирования $\Delta Q = Q_x - Q_z$ как от «шумов незнания». Реальный динамический процесс $X(t)$ может быть описан реальным оператором Q_x в символическом виде:

$$Q_x(S_x, x; \alpha_k, f_i) = 0, \quad (3)$$

где S_x — математический язык представления процесса X ; x — набор существенных переменных на принятом уровне описания; α_k — параметры процесса X ; f_i — несущественные и флуктуационные переменные. Например оператор (3) для многих динамических систем часто имеет вид дифференциального уравнения

$$Q_x(d/dt, x; \alpha_k, f_i). \quad (4)$$

В зависимости от того, как решаются вопросы выбора принципа прогнозирования, существенных и несущественных переменных, возникает потребность в упрощенном, идеализированном представлении реального процесса модельным процессом $Z(t)$ с оператором Q_z , аналогичным оператору (3):

$$Q_z(S_z, z, \hat{\alpha}_k, \emptyset) = 0. \quad (5)$$

Здесь $\hat{\alpha}_k$ — оценки параметров процесса Z , а символ « $\emptyset$ » означает пренебрежение несущественными факторами f_i , что и составляет идеализацию. Поскольку в нашем случае Z представляется РАКЗ-моделью в виде $X\Phi Z \bar{\Phi}(z)$, модельный оператор (5) с использованием квантов знаний задан выражением:

$$Q_z\left(\text{RED}\left(k_2 \bar{\Sigma}_{BM}, k_1 y_{\omega}, k_1 \hat{y}_{\omega}\right), \bar{\Phi}(z), \hat{\alpha}_k\right) = 0 \quad (6)$$

В развернутом виде выражение (6) является конечно-предикатным уравнением относительно набора существенных переменных z и оценок $\hat{\alpha}_k$ параметров процесса Z , а в квантовом представлении — дедуктивным преобразователем матричного кванта $k_2 \bar{\Sigma}_{BM} = \text{БкЗ}$ по векторному кванту наблюдаемой ситуации $k_1 y_{\omega}$ (то есть ОПР ω) в результирующий прогнозный квант знаний $k_1 \hat{y}_{\omega}$ посредством оператора редукции (RED). При этом значения наборов z разнотипных признаков ОПР измеряются как в количественных, так и в качественных шкалах, но для машинного манипулирования k -знаниями преобразуются в булеву форму (значение признака $z_i=1$, если ОПР обладает этим признаком и значение $z_i=0$, если не обладает). Таким образом, согласно квантовой прогнозной модели (6) принятие знаниеориентированных решений при условиях неопределенности в сформулированных B -и C -задачах реализуются с помощью редуцирования базы квантов знаний по текущей ситуации наблюдаемого процесса Y для получения прогнозируемой ситуации модельного процесса Z . Поэтому мы будем использовать общепринятую характеристику качества прогноза [3] в виде эмпирического усреднения квадрата ошибки $\eta = Y - Z$:

$$\langle \eta^2 \rangle = \langle (Y - Z)^2 \rangle, \quad (7)$$

где угловые скобки $\langle \rangle$ упрощенно означают выбранный исследователем способ эмпирического усреднения. Например, если обозначить через $y_i = Y(t_j^0 + \tau)$ значение процесса $Y(t)$, спустя время τ после момента t_j^0 начала j -го наблюдения, а через $z_j = Z(t_j^0 + \tau_{\text{пр}})$ — соответствующий прогноз на интервале $(t_j^0, t_j^0 + \tau_{\text{пр}})$, удовлетворяющий начальному условию $z_j^0 = y_j^0$, то средний квадрат ошибки (7) можно определить по формуле

$$\langle \eta^2 \rangle = \frac{1}{N} \sum_{j=1}^N [Y(t_j^0 + \tau_{\text{пр}}) - Z(t_j^0 + \tau_{\text{пр}})]^2, \quad (8)$$

где N — число наблюдений.

Очевидно, чем больше N , тем надежнее оценка погрешности (8). Для того, чтобы характеризовать при прогнозировании потенциальную предсказуемость, в роли единой количественной меры качества знаниеориентированных решений будем использовать безразмерную величину

$$W(\tau) = \frac{\langle Y \cdot Z \rangle}{\sqrt{\langle Y^2 \rangle \cdot \langle Z^2 \rangle}}, \quad (9)$$

которая является эмпирическим коэффициентом корреляции между наблюдением и прогнозом через время τ после наблюдения.

Руководствуясь приведенными выше соображениями, принимаем в качестве исходных следующие данные:

– заглаговременно синтезираная

$$\text{БкЗ} = k_2 \bar{\Sigma}_{\text{ВМ}} = \bar{\Phi}_\tau$$

с гарантированной устойчивостью ее закономерностей на заданном интервале времени прогнозирования τ для рассматриваемой предметной области;

– выборочные N пар значений (реализаций) наблюдаемого процесса $Y(t)$ (2) и модельного процесса $Z(t)$, вычисляемых по формуле (6) для данной предметной области.

Требуется на основании использования исходных данных и формул (7)–(9) разработать методику определения единой меры качества $W(\tau)$ принимаемых знаниеориентированных и прогнозных решений независимо от природы предметной области. Учитывая корреляционный характер количественной меры $W(\tau)$, установить ее связь с абсолютной погрешностью прогноза $\langle \eta^2 \rangle$ (7), с причинами неуспешного прогнозирования и с некоторыми характеристиками предсказуемости динамических процессов.

3. Решение поставленной задачи: методика вычисления меры качества $W(\tau)$

Изложим методику определения единой меры качества идентификационных и прогнозных решений $W(\tau)$ по заданным $\text{БкЗ} = k_2 \bar{\Sigma}_{\text{ВМ}} = \bar{\Phi}_\tau$ и N выборочным парам реализаций двух процессов $Y(t)$ и $Z(t)$ как эмпирического распределения двух случайных величин Y и Z . Статистические оценки характеристик такой двумерной генеральной совокупности при неизвестном теоретическом распределении будем вычислять по рекомендациям в [6]. Заметим, что трактовка случайности как непредсказуемости является лишь одним из многих соглашений о случайности, принятых в современной науке. Мы будем следовать соглашению, принятому в экспериментальной физике, где случайность (непредсказуемость) связывают с потерей корреляции. Например для процесса $Y(t)$, характеризуемого коэффициентом корреляции

$$K_y(\tau) = \frac{\langle Y(t) \cdot Y(t-\tau) \rangle}{\langle Y^2 \rangle}, \quad (10)$$

значения $Y(t)$, которые разделены временем, превышающим время корреляции τ_c , считаются непредсказуемыми [3].

Итак, методику определения меры качества $W(\tau)$ по формуле (9) изложим в виде последовательности действий и формул статистического оценивания необходимых эмпирических характеристик распределения значений случайных величин Y и Z , обоснованных в справочнике [6]. Сначала для N выборочных пар реализаций Y и Z выбирают размеры и положение разрядов (y'_{i-1}, y'_i) , (z'_{j-1}, z'_j) , на которые делят область эмпирического распределения Y и Z в отдельности. Число разрядов для

Y и Z рекомендуется брать не более 10, так как при увеличении их числа точность оценок изменяется незначительно, а объем вычислений быстро возрастает. Полученные данные сводят в корреляционную таблицу (табл. 1), в каждую клетку которой заносится частота N_{ij} – число пар измерений, удовлетворяющих условиям

$$y'_{i-1} < y < y'_i, \quad z'_{j-1} < z < z'_j. \quad (11)$$

Этим реализациям приписываются значения:

$$y_i = \frac{1}{2} \cdot (y'_{i-1} + y'_i), \quad z_j = \frac{1}{2} \cdot (z'_{j-1} + z'_j). \quad (12)$$

В нижнюю строки табл. 1 заносят частоты N_{yi} , а в крайний правый столбец – частоты N_{zj} , определяемые по формулам:

$$N_{yi} = \sum_{j=1}^{m_z} N_{ji}, \quad N_{zj} = \sum_{i=1}^{m_y} N_{ij}, \quad (13)$$

$$(i = 1, 2, \dots, m_y), \quad (j = 1, 2, \dots, m_z),$$

где m_y – число разрядов для величины Y ; m_z – число разрядов для величины Z .

По данным табл. 1, используя формулы (11)–(13), определяют условные выборочные средние $\bar{Y}(z_j)$ и $\bar{Z}(y_i)$:

$$\bar{Z}(y_i) = \frac{1}{N_{yi}} \cdot \sum_{j=1}^{m_z} N_{ji} \cdot z_j, \quad (i = 1, 2, \dots, m_y); \quad (14)$$

$$\bar{Y}(z_j) = \frac{1}{N_{zj}} \cdot \sum_{i=1}^{m_y} N_{ij} \cdot y_i, \quad (j = 1, 2, \dots, m_z). \quad (15)$$

Таблица 1

	$i \backslash j$	y_i				N_{zj}	$\bar{Y}(z_j)$
		1	2	...	m_y		
z_j	1						
	2						
	...						
	m_z						
N_{yi}							
$\bar{Z}(y_i)$							

Далее, на основании использования формул (12)–(15) и данных табл. 1 вычисляют выборочные средние $\bar{Y}$ и $\bar{Z}$ по формулам:

$$\bar{Y} = \frac{1}{N} \cdot \sum_{i=1}^{m_y} N_{yi} \cdot y_i = \frac{1}{N} \cdot \sum_{j=1}^{m_z} N_{zj} \cdot \bar{Y}(z_j); \quad (16)$$

$$\bar{Z} = \frac{1}{N} \cdot \sum_{j=1}^{m_z} N_{zj} \cdot z_j = \frac{1}{N} \cdot \sum_{i=1}^{m_y} N_{yi} \cdot \bar{Z}(y_i). \quad (17)$$

Используя формулы (12), (13), (16) и (17), а также соответствующие данные табл. 1, определяют эмпирические дисперсии S_{1y}^2 и S_{1z}^2 по формулам:

$$S_{1y}^2 = \frac{1}{N-1} \cdot \sum_{i=1}^{m_y} N_{yi} (y_i - \bar{Y})^2; \quad (18)$$

$$S_{1z}^2 = \frac{1}{N-1} \cdot \sum_{j=1}^{m_z} N_{zj} (z_j - \bar{Z})^2. \quad (19)$$

Наконец, на основании использования выражений (12), (13), (16)-(19) приходим к соотношению для вычисления эмпирического коэффициента корреляции $\bar{r}$:

$$\bar{r} = \frac{\sum_{i=1}^{m_y} \sum_{j=1}^{m_z} N_{ij} \cdot y_i \cdot z_j - N \cdot \bar{Y} \cdot \bar{Z}}{(N-1) \cdot \sqrt{S_{1y}^2 \cdot S_{1z}^2}} = W(\tau), \quad (20)$$

которое и является рабочей формулой для вычисления единой количественной меры качества знаниеориентированных решений $W(\tau)$ (9). Численное значение меры $W(\tau)$ получаем путем подстановки в формулу (20) значений вычисленных величин N_{ij} , y_i , z_j , $\bar{Y}$, $\bar{Z}$, S_{1y}^2 и S_{1z}^2 для заданной эмпирической N – выборки по указанным формулам. При $N > 70 \div 80$ величина $\bar{r} = W(\tau)$ (20) приближенно подчиняется нормальному распределению, что позволяет оценивать меру качества $W(\tau)$ с помощью доверительной вероятности

$$\gamma = P \left\{ \bar{r} - \beta \frac{1 - \bar{r}^2}{\sqrt{N}} < W(\tau) < \bar{r} + \beta \frac{1 - \bar{r}^2}{\sqrt{N}} \right\} \approx 2\Phi_0(\beta),$$

используя методику расчетов [6] (см. § 14.3, пример 14.3.1, табл. III).

4. Связь меры качества $W(\tau)$ с потенциальной предсказуемостью

Условимся называть сроком (временем) прогнозирования интервал времени $\tau = t - t^0$ после начала наблюдения t^0 . При исследовании проблемы предсказуемости мы всегда вынуждены рассматривать три процесса: реальный процесс $X(t)$, наблюдаемый процесс $Y(t)$ и модельный прогнозный процесс $Z(t)$, которые схематично изображены на рис. 1.

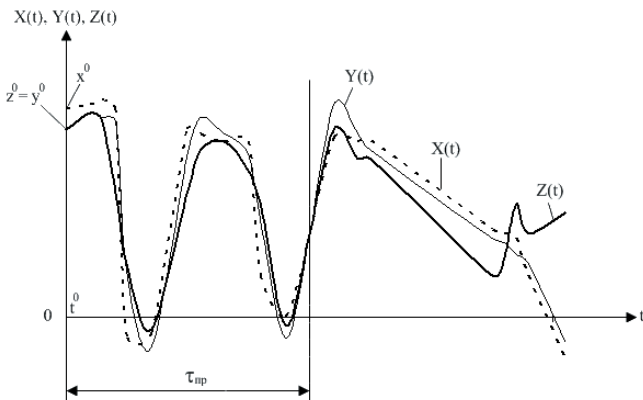


Рис. 1. Процессы $X(t)$, $Y(t)$, $Z(t)$

В силу корреляционного характера величина меры качества прогнозирования $W(\tau)$ равна 1 при $\tau = t - t^0 = 0$ поскольку начальное значение прогно-

за $z_0 = y_0$, а с течением времени уменьшается до 0. Общий вид зависимости $W(\tau)$ показан на рис. 2. Близкие к 1 значения $W(\tau)$ отвечают удовлетворительному прогнозу, а малые значения соответствуют неудовлетворительной предсказуемости из-за слабой согласованности наблюдения и прогноза. Поэтому величину меры $W(\tau)$ будем называть *степенью предсказуемости*. Время, за которое величина $W(\tau)$ снижается до уровня 0,5, назовем *временем предсказуемости* (см. рис. 2) с обозначением $\tau_{пр}$. Степень предсказуемости $W(\tau)$ однозначно связана с абсолютной погрешностью прогноза $\langle \eta^2 \rangle$ (7):

$$W(\tau) = \frac{\langle Y^2 \rangle + \langle Z^2 \rangle}{2 \cdot \sqrt{\langle Y^2 \rangle \cdot \langle Z^2 \rangle}} \cdot \left[1 - \frac{\langle \eta^2 \rangle}{\langle Y^2 \rangle + \langle Z^2 \rangle} \right]. \quad (21)$$

Времени предсказуемости $\tau_{пр}$ отвечает абсолютная погрешность $\langle \eta^2 \rangle$ порядка

$$0,5 \cdot (\langle Y^2 \rangle + \langle Z^2 \rangle) \approx \langle Y^2 \rangle,$$

то есть порядка дисперсии наблюдаемого процесса $Y(t)$. Если высокую степень предсказуемости ($W(\tau) \approx 1$) рассматривать как свидетельство предсказуемой (детерминированной) динамики процесса, а малые значения $W(\tau)$ приписать непредсказуемому, случайному поведению, то $W(\tau)$ можно трактовать как степень детерминированности процесса $Y(t)$ по отношению к модельному процессу $Z(t)$ [3].

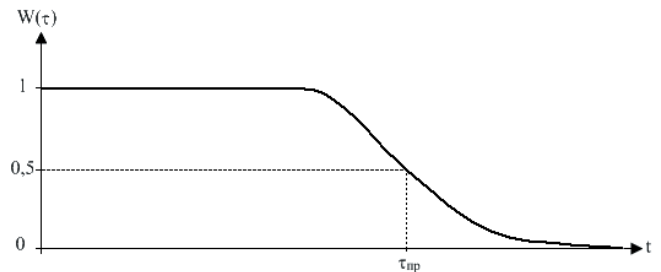


Рис. 2. Зависимость степени предсказуемости W от времени $\tau = t - t^0$, прошедшего после начала наблюдения

В экспериментальной физике установлено, что коэффициент корреляции $K_Y(\tau)$ (10) наблюдаемого процесса $Y(t)$ является наименьшей (наихудшей) степенью предсказуемости, а время корреляции τ_c – наименьшим (наихудшим временем прогнозируемого поведения $Y(t)$). В интервале $\tau_c < \tau < \tau_{пр}$ один и тот же процесс $Y(t)$ следует классифицировать как случайный, если судить в рамках соглашения, отождествляющего случайность с потерей корреляции и как детерминированный с точки зрения отождествления случайности с непредсказуемостью [3]. Нижний предел $\tau_{пр} = \tau_c$ проявляется при условиях, когда исследователь не располагает динамическим уравнением (4) и для прогноза использует принцип «завтра будет как сегодня». В этом случае в качестве прогноза $z(t) = z(t^0 + \tau)$ при-

нимает значение $y(t^0)$, наблюдаемое в данный момент t :

$$z(t+\tau) = y(t^0) \text{ или } z(t) = y(t-\tau). \quad (22)$$

При этом степень предсказуемости $W(\tau)$ (9) превращается в коэффициент корреляции $K_Y(\tau)$ (10).

Результат сопоставимости $\tau_{\text{пр}}$ со временем корреляции τ_c проиллюстрируем на примере зашумленного динамического хаоса из [3]. Рассматривается наблюдаемый хаос $Y(t)$ с наибольшим положительным показателем Ляпунова λ_+ со средним квадратом ошибки прогноза, нарастающем по экспоненциальному закону

$$\langle \eta^2 \rangle = (\sigma_v^2 + \sigma_f^2 + \sigma_{\Delta Q}^2) \cdot \exp^{(2\lambda_+ \cdot \tau)}.$$

Здесь величины σ_v^2 , σ_f^2 и $\sigma_{\Delta Q}^2$ отражают соответственно вклад инструментальных шумов, физических шумов и «шумов незнания» подобно (2). Полученная оценка времени предсказуемого поведения $Y(t)$ имеет вид:

$$\tau_{\text{пр}} \approx \frac{1}{2 \cdot \lambda_+} \cdot \ln \frac{\langle Y^2 \rangle}{\sigma_v^2 + \sigma_f^2 + \sigma_{\Delta Q}^2}. \quad (23)$$

Поскольку порядок величины $1/(2\lambda_+)$ отвечает времени корреляции процесса $\tau_c \approx 1/2\lambda_+$, отношение $\tau_{\text{пр}}/\tau_c$ характеризуется значением величины, стоящей под знаком логарифма в (23). Если, например, шумовые слагаемые σ_v^2 , σ_f^2 и $\sigma_{\Delta Q}^2$ составляют одну тысячную (10^{-3}) от среднеквадратичной амплитуды $A_Y \approx \langle Y^2 \rangle^{1/2}$, то $\tau_{\text{пр}} \approx \tau_c \cdot \ln 10^3 \approx 7\tau_c$, то есть $\tau_{\text{пр}}$ в 7 раз превышает τ_c . Здесь вскрывается важность понятия *горизонта предсказуемости* τ_{hor} , под которым будем понимать предельное время предсказуемого поведения процесса $Y(t)$, которое нельзя превзойти ни усовершенствованием измерений, ни улучшением предсказательной модели $Z(t)$. Иными словами, τ_{hor} отвечает пределу пренебрежимо малых измерительных шумов и шумов незнания, что равносильно в нашем примере хаоса пренебрежению вкладами σ_v^2 и $\sigma_{\Delta Q}^2$ по сравнению с вкладом физических шумов σ_f^2 , влияние которых невозможно устранить ни при каких условиях:

$$\tau_{\text{hor}} = \frac{1}{2\lambda_+} \ln \frac{\langle Y^2 \rangle}{\sigma_f^2}.$$

Таким образом, изложенные выше соображения позволяют представить на рис. 3 общий характер изменения степени предсказуемости $W(\tau)$ по мере совершенствования прогностической модели $Z(t)$.

Кривая ① показывает, что на малых временах, не превышающих время корреляции τ_c , делается удовлетворительный прогноз по принципу «завтра как сегодня», когда $W(\tau)$ (9) совпадает с коэффициентом корреляции $K_Y(\tau)$ (10). Кривые ②

и ③ отвечают более или менее удовлетворительным прогнозным моделям $Z_1(t)$ и $Z_2(t)$, когда при $\tau_{\text{пр}} > \tau_c$ по мере совершенствования моделей степень предсказуемости $W(\tau)$ вместе со временем $\tau_{\text{пр}}$ увеличиваются. На этом этапе посредством использования РАКЗ-модели (6) процесс $Y(t)$ можно превратить из непредсказуемого (точка w_1 на кривой ②, отвечающей модели $Z_1(t)$) в предсказуемый (детерминированный) (точка w_2 на кривой ③, отвечающей модели $Z_2(t)$). Переход $w_1 \rightarrow w_2$ от модели Z_1 к более совершенной РАКЗ-модели Z_2 переводит процесс из разряда случайных (непредсказуемых) к детерминированным (предсказуемым) и возможен только в рамках концепции, отождествляющей случайное с непредсказуемым. В рамках аксиоматической теории вероятностей такой переход невозможен в принципе, так как процесс остается случайным независимо от наличия либо отсутствия модели для его предсказания. Наконец, кривая ④ на рис. 3 отвечает предельно достижимой степени предсказуемости $W(\tau)$ за предельное время предсказания τ_{hor} («горизонт предсказуемости»), начиная с которого никакая модель не может обеспечить удовлетворительное предсказание.

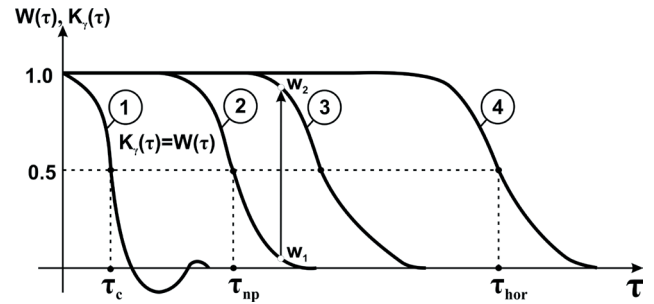


Рис. 3. Характер изменения степени предсказуемости $W(\tau)$ по мере совершенствования прогностической модели $Z(t)$

Следовательно, предложенные оценки степени предсказуемости как количественной меры $W(\tau)$ (20) вместе с горизонтом предсказуемости τ_{hor} имеют принципиальное значение — нет смысла строить прогнозы на сроки, превышающие τ_{hor} , в силу неоправданных потерь времени и средств.

Выводы

1. На основе использования прогнозных РАКЗ-моделей вида (6) и меры качества решений $W(\tau)$ (20) в инженерии квантов знаний выделены и исследованы основные причины, препятствующие успешному прогнозированию. Это — локальная неустойчивость исследуемых процессов, действие флуктуационных сил и шумов измерения, эволюция динамики процессов, неадекватность (неточность) модельного оператора Q_z (5) относительно реального оператора Q_x (3) и финансовые ограничения.

2. Любой прогнозный процесс имеет ограниченное время предсказуемого поведения, которое зависит от указанных в п. 1 основных причин. Существует максимальное значение этого времени — так называемый «горизонт предсказуемости», обусловленный действием только неустраняемых флуктуаций самого процесса, за пределы которого выйти невозможно.

3. Успешный прогноз достигается главным образом за счет удачного выбора модельного прогнозного оператора Q_z (6), что послужило мотивацией для разработки знаниеориентированной РАКЗ-модели принятия идентификационных и прогнозных решений в виде характеристической функции запретов Φ с неявным учетом времени.

Эффективность предложенной меры качества $W(\tau)$ знаниеориентированных решений и приведенные в заключении выводы подтверждены результатами решения многочисленных практических задач [2].

Список литературы. 1. Сироджа И.Б. Математическое и программное обеспечение интеллектуальных компьютерных систем: Учеб. пособ. — Харьков: ХАИ, 1992. — 100 с. 2. Сироджа И.Б. Квантовые модели и методы искусственного интеллекта для принятия решений и управления. — К.: Наукова думка, 2002. — 490 с. 3. Кравцов Ю.А. Фундаментальные и практические пределы предсказуемости // Пределы предсказуемости: Науч.-техн. сб. — М.: Наука, 1996. С. 170-199. 4. Люгер Дж. Ф. Искусственный интеллект: стратегии и методы решения сложных проблем, 4-е издание: Пер. с англ. — М.: Издательский дом «Вильямс», 2003. — 863 с. 5. Закревский А.Д. Выявление имплицитных закономерностей в булевом пространстве признаков и распознавание образов // Кибернетика. — 1982. — №1. — С. 1-6. 6. Абезгауз Г.Г., Тронь А.П., Копенкин Ю.Н., Корovina И.А. Справочник по вероятностным расчетам. — М.: Воен. издат. МО СССР, 1970. — 532 с.

Поступила в редколлегию 22.09.2008

УДК 004.89

Оцінювання якості ідентифікаційних та прогнозних рішень в інженерії квантів знань / І.Б. Сіроджа // Біоніка інтелекту: наук.-техн. журнал — 2008. — № 2 (69). — С. 77-83.

Розроблено логічну модель екстраполяції спостережень у вигляді предикатної системи імплікативних закономірностей предметної області як бази квантів знань. Вдалося звести задачу прийняття ідентифікаційних рішень до частого випадку прийняття прогнозних рішень. При цьому загальна задача прийняття прогнозних рішень запропонована по новому, з неявним обліком змінної часу (на відміну від традиційної постановки з явним обліком часу), завдяки використанню предикатної характеристичної функції для екстраполяції стійких закономірностей динаміки процесу, що прогнозується у фазовому просторі.

Л. 3. Бібліогр.: 6 найм.

UDK 004.89

Estimating qualities identification and prognostic decisions in engineering of quantes of knowledge / I. Sirodza // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 77-83.

The logic model are develops of extrapolation observation in the form of predicate system implicational laws of a object domain, as bases of quantes of knowledge (БкЗ). It was possible to reduce a problem of acceptance the identification decision to an acceptance special case prognostic decisions. Thus the general problem of acceptance prognostic decisions is posed with the implicit account of a time variable (unlike traditional posed problem with the obvious account of time), thanks to using of predicate characteristic function for extrapolation of steady laws of dynamics predictable process in phase space.

Fig. 3. Ref.: 6 items

УДК 519.62



МОДЕЛИРОВАНИЕ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ УПРАВЛЕНИЯ АКТИВНЫМИ ЭЛЕМЕНТАМИ

Г.Г. Четвериков¹, Т.Н. Федорова²

¹ ХНУРЭ г. Харьков, Украина, chtvergg@kture.kharkov.ua;

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

Рассмотрены методы моделирования режимов функционирования инженерных сетей, оптимальных по критериям устойчивости и энергетических потерь.

СЕТИ ИНЖЕНЕРНЫЕ, МОДЕЛИРОВАНИЕ, ОПТИМИЗАЦИЯ, УСТОЙЧИВЫЙ РЕЖИМ

Введение

Инженерные сети, к которым относятся системы водо-, тепло- и газоснабжения можно рассматривать как совокупность элементов двух типов: активных и пассивных. Структура такой сети определяется графом, в котором каждая дуга соответствует определенному элементу сети. Режим функционирования сети характеризуется вектором расходов целевого продукта по дугам и вектором давлений в узлах сети. Функционирование каждого элемента описывается уравнением, связывающим значения давлений в узлах, являющихся началом и концом дуги, соответствующей элементу, и расходом целевого продукта по этой дуге. Вид уравнения зависит от вида рассматриваемой инженерной сети и от типа элемента. Управление технологическими процессами, протекающими в инженерной сети, происходит путем включения/выключения активных элементов, а также задания параметров некоторых пассивных элементов.

Задача моделирования технологических процессов управления активными элементами заключается в выборе определенных значений параметров пассивных элементов, определяющих режимы работы активных элементов, при изменении состава работающих активных элементов.

1. Поиск устойчивого режима при добавлении в сеть активного элемента

Рассмотрим инженерную сеть, граф которой представлен на рис. 1.

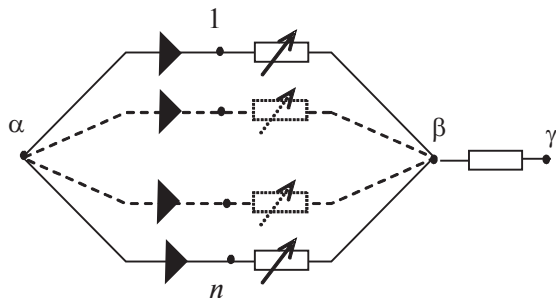


Рис. 1. Пример графа инженерной сети с одним входом

Дуги $(\alpha, 1), \dots, (\alpha, n)$ — соответствуют активным элементам.

Дуги $(\beta, 1), \dots, (\beta, n)$ — соответствуют пассивным элементам с изменяемой характеристикой (изменяемым сопротивлением)

Дуга (β, γ) — соответствует распределительной части инженерной сети.

Уравнения, описывающие процесс протекания целевого продукта по элементам сети, выглядят следующим образом:

$$P_i = P_\alpha + a_i - b_i q_i^2 \text{ — для активных дуг;}$$

$$P_\beta = P_i - r_i q_i^2 \text{ — для пассивных дуг с изменяемым сопротивлением;}$$

$$P_\beta - P_\gamma = R_s Q_s^2 \text{ — для дуги соответствующей распределительной части инженерной сети.}$$

В приведенных формулах $P_\alpha, P_\beta, P_\gamma, P_i$ — давления в соответствующих узлах сети; q_i — расход целевого продукта по дугам (α, i) и (β, i) ; Q_s — расход по дуге (β, γ) ; a_i, b_i — параметры дуг (α, i) ; r_i — сопротивление дуги (i, β) ; R_s — обобщенное сопротивление распределительной части сети.

Система уравнений, описывающая процесс функционирования приведенной на рисунке сети, имеет следующий вид:

$$\begin{cases} P_i = P_\alpha + a_i - b_i q_i^2, & (1) \\ P_\beta = P_i - r_i q_i^2, & (2) \\ P_\beta - P_\gamma = R_s Q_s^2, & (3) \\ Q_s = \sum_{i=1}^n q_i. & (4) \end{cases}$$

$$i = 1 \dots n$$

Значения P_α, P_β, P_i могут быть измерены, P_γ — полагаем равным 0. Измеряется также значение Q_s . Значения a_i, b_i — это паспортные характеристики активных элементов. Значения q_i определяются из уравнений (1), затем подставляются в уравнения (2), откуда определяются значения r_i . Из (3) определяем значение R_s .

В дальнейшем будем считать что величина R_s не меняется в момент изменения состава работающих активных элементов.

Для каждого активного элемента есть диапазон допустимых значений расходов целевого продукта $[q_i^+, q_i^{++}]$.

При добавлении активного элемента в группу уже работающих (что физически означает, например, включение насоса на насосной станции водопроводной сети), значения r_i в пассивных дугах, стоящих в графе после ранее действовавших активных элементов, как правило не меняются (что означает, что оператор не меняет положение задвижек, стоящих после работавших ранее насосов), если значение расхода целевого продукта по активным дугам не выходит за пределы заданных ограничений. Для пассивной дуги, стоящей после добавленного активного элемента, необходимо выбрать такое значение сопротивления, чтобы, во-первых, расход по добавленной активной дуге был в допустимых пределах, а во-вторых, чтобы этот расход был близок к точке наиболее устойчивой работы.

Пусть номер добавленного активного элемента n . Тогда математически задача выбора допустимого решения состоит в определении такого значения $r_n \geq 0$, при котором после решения системы (1) – (4) относительно переменных $q_i, P_i (i=1, \dots, n), P_\beta, Q_s$ при заданных $a_i, b_i, r_i (i=1, \dots, n), R_s$ полученные значения $q_i (i=1, \dots, n)$ удовлетворяли бы неравенствам:

$$q_i^+ \leq q_i \leq q_i^{++}. \quad (5)$$

Подставим значения P_i из (1) в (2) и таким образом упростим нашу систему, приведя ее к виду:

$$\begin{cases} P_\beta = P_\alpha + a_i - b_i - r_i q_i^2 & (i=1, \dots, n), \end{cases} \quad (6)$$

$$\begin{cases} P_\beta = R_s Q_s^2, \end{cases} \quad (7)$$

$$\begin{cases} Q_s = \sum_{i=1}^n q_i. \end{cases} \quad (8)$$

Примем $r_n=0$ и решим систему (6) – (8). Если окажется, что полученное значение $q_n < q_n^+$, то это означает, что для данной конфигурации активных элементов при зафиксированных значениях $r_i (i=1, \dots, n-1)$ допустимого режима не существует, так как для любых значений $r_n > 0$ величина q_n будет только уменьшаться.

Если при значении $q_n > q_n^+$ ограничения (5) выполняются не для всех активных элементов, то мы можем начать поиск допустимого решения. При этом надо учитывать, что $P_\beta (r_n=0)$ – есть максимально возможное давление на входе распределительной части сети при рассматриваемой конфигурации активных элементов.

Если задаться некоторым значением P_β^0 , то из (7) можно определить Q_s , а из (6) – значения $q_i (i=1, \dots, n-1)$. Значение q_n определяем из (8), записав его в виде:

$$q_n = Q_s - \sum_{i=1}^{n-1} q_i,$$

после чего находим

$$r_n = \frac{P_\alpha + a_n - b_n q_n^2 - r_n q_n^2 - P_\beta}{q_n^2}.$$

Таким образом, достаточно простой и эффективный алгоритм моделирования допустимого решения представляется следующим: на отрезке $[P_\beta^+, P_\beta^{++}]$ с шагом для значений

$$P_\beta^{(k)} = P_\beta^{++} - k \Delta P_\beta \quad (k=0, \dots, t)$$

решаем систему (6) – (8). Процесс прекращается, как только будут выполнены ограничения (5). Для реальных инженерных сетей диапазон значений $(P_\beta^+ - P_\beta^{++})$ можно выбирать равным от 20 до 30, а величину $\Delta P_\beta = 1$.

Введем функцию

$$F(q_n) = \min_{q_n^+ \leq q_n \leq q_n^{++}} (q_n - q_n^+, q_n^{++} - q_n). \quad (9)$$

Значения этой функции характеризуют степень удаления текущего значения q_n от границ допустимой области.

После определения допустимого решения, используя описанную выше процедуру, можно найти такое значение r_n^* , при котором функция (9) будет достигать максимального значения.

2. Выбор наиболее устойчивого режима группы активных элементов

Теперь рассмотрим более сложную задачу. Будем считать, что при включении /отключении активных элементов мы можем изменять все сопротивления $r_i (i=1, \dots, n)$. Тогда задача заключается в подборе таких значений $r_i (i=1, \dots, n)$, чтобы значения $q_i (i=1, \dots, n)$, полученные в результате решения системы (6)–(8), удовлетворяли ограничениям (5).

Зададим $q_i = q_i^+ (i=1, \dots, n)$. Из (8) определим Q_s^+ , а из (7) – P_β^+ . Затем из уравнений (6) определим $r_i (i=1, \dots, n)$.

Если для какого-либо k -го активного элемента мы получим, что $r_k < 0$, то это означает, что допустимых решений нет. Расход Q_s^+ – это наименьший из возможных расходов по сети, а соответственно P_β^+ – наименьшее из возможных давлений, поэтому при всех остальных значениях Q_s и P_β полученное отрицательное значение r_k будет только еще меньше.

Если среди значений $r_i (i=1, \dots, n)$ есть равные 0, то это означает, что найденное допустимое значение является единственно возможным.

Если все значения $r_i (i=1, \dots, n)$ больше 0, то можно начать поиск таких значений $r_i (i=1, \dots, n)$, которые обеспечили бы наиболее устойчивый режим работы инженерной сети. Для этого введем в рассмотрение функцию

$$\varphi(r_1, \dots, r_n) = \max_{i=1, \dots, n} \left| \frac{q_i - q_i^*}{\Delta_i} \right|, \quad (10)$$

где q_i^* – расход целевого продукта, соответствующий

щий наиболее устойчивому режиму работы активного элемента;

$$\Delta_i = \begin{cases} q_i^+ - q_i^*, & \text{если } q_i < q_i^*, \\ q_i^{++} - q_i^*, & \text{если } q_i \geq q_i^*. \end{cases}$$

Считаем, что $q_i = q_i(r_1, \dots, r_n)$, то есть $q_i (i=1, \dots, n)$ являются неявными функциями, которые определяются системой уравнений (6) – (8).

Для задания наиболее устойчивого режима работы сети необходимо найти такие значения из множества допустимых, которые обеспечивали бы минимум функции (10). Областью значений функции $\varphi(r_1, \dots, r_n)$ является отрезок $[0, 1]$. Значение 0 функции $\varphi(r_1, \dots, r_n)$ принимается при $q_i = q_i^* (i=1, \dots, n)$, а значение 1 достигается, если хотя бы одна из величин q_i равна q_i^+ или q_i^{++} .

Поиск оптимальных значений $r_i (i=1, \dots, n)$ начинаем с проверки на допустимость режима, при котором $q_i = q_i^* (i=1, \dots, n)$. По заданным значениям $q_i^* (i=1, \dots, n)$ определяем $Q_s^* = \sum_{i=1}^n q_i^*$, а из уравнения (8) – значение P_β^* . Затем из уравнения (6) определяем $r_i^* (i=1, \dots, n)$. Если какие-то из значений r_i^* окажутся меньше 0, то запускается итерационная процедура, в которой значения расходов в активных элементах на очередном шаге определяются по формулам:

$$q_i^{(k+1)} = q_i^k + \delta_i, \quad k = 0, \dots, m.$$

$$q_i^{(0)} = q_i^+, \quad \delta_i = \frac{(q_i^* - q_i^+)}{m}.$$

Величины $r_i (i=1, \dots, n)$ находятся из системы (6) – (8).

Значение функции $\varphi(r_1, \dots, r_n)$ при каждом шаге уменьшается на $(1/m)$. Процесс прекращается, если на очередном шаге одно или несколько сопротивлений r_i окажутся отрицательными. Тогда значения сопротивлений $r_i (i=1, \dots, n)$, полученные на предыдущем шаге, принимаются в качестве оптимальных. Величина m определяет точность полученного решения. Процедуру можно усовершенствовать, сделав ее двухуровневой. Сначала, задавшись небольшим значением m_1 и, следовательно, достаточно большим шагом $\delta_i^{(1)}$, найдем грубое приближение к оптимальному решению: $r_i^{*(1)}, q_i^{*(1)}, Q_s^{*(1)}, P_\beta^{*(1)} (i=1, \dots, n)$. Затем, используя эти значения как начальные, будем искать оптимальные расходы на отрезке $[q_i^{*(1)}, q_i^*]$ с уже меньшим шагом $\delta_i^{(2)}$, соответствующим некоторому значению m_2 . Такой подход позволяет получить более точное решение с меньшими вычислительными затратами.

3. Минимизация энергетических потерь при заданном уровне устойчивости

Решение, найденное при минимизации функции (10), позволяет найти наиболее устойчивые,

по отношению к изменениям внешних условий, режимы работы активных элементов, но никак не учитывает энергетические потери, которые происходят на пассивных элементах, используемых для регулирования работы активных элементов – это потери на элементах с изменяемым сопротивлением. Для учета таких потерь можно рассмотреть функцию

$$\psi(\Delta P_i, q_i) = \sum_{i=1}^n \Delta P_i q_i, \quad (11)$$

где q_i – расход по i -ому активному элементу; ΔP_i – потери напора на этом элементе; $\Delta P_i = P_i - P_\beta = r_i q_i^2$.

Наиболее выгодным будет, безусловно, режим, при котором достигался бы минимум функции (11), но при этом обеспечивалась бы достаточно устойчивая работа активных элементов. Рассчитать такой режим можно, если решить следующую задачу:

$$\sum_{i=1}^n r_i q_i^3 \rightarrow \min_{q_i (i=1, \dots, n)}, \quad (12)$$

$$P_\beta = P_\alpha + a_i - b_i q_i^2 - r_i q_i^2, \quad (13)$$

$$P_\beta = R_s Q_s^2, \quad (14)$$

$$Q_s = \sum_{i=1}^n q_i, \quad (15)$$

$$r_i \geq 0, \quad (16)$$

$$q_i^{*+} \leq q_i \leq q_i^{*++} \quad (i=1, \dots, n). \quad (17)$$

Значения q_i^{*+} и q_i^{*++} выбираются так, чтобы $[q_i^{*+}, q_i^{*++}] \subset [q_i^+, q_i^{++}]$. Область допустимых значений специально сужается, чтобы в случае, когда оптимальное решение окажется на границе, и по истечении некоторого времени после изменения внешних условий функционирования сети расходы по активным элементам выйдут за пределы $[q_i^{*+}, q_i^{*++}]$, это не привело бы к выходу их за пределы диапазона $[q_i^+, q_i^{++}]$. Задача (12) – (17) – это задача минимизации нелинейной функции при ограничениях. Выразив r_i через q_i и исключив переменные Q_s и P_β , задачу (12) – (17) можно представить в виде:

$$\sum_{i=1}^n (P_\alpha + a_i - b_i q_i^2 - R_s \sum_{j=1}^n q_j) q_i \rightarrow \min_{q_i (i=1, \dots, n)}, \quad (18)$$

$$P_\alpha + a_i - b_i q_i^2 - R_s \sum_{j=1}^n q_j \geq 0, \quad (19)$$

$$q_i - q_i^{*+} \geq 0, \quad (20)$$

$$q_i^{*++} - q_i \geq 0. \quad (21)$$

Для решения задачи (18) – (21) можно воспользоваться методом штрафных функций [1]. Будем минимизировать функцию

$$\Phi_k(q_1, \dots, q_n) = F(q_1, \dots, q_n) + B_k(q_1, \dots, q_n).$$

Для последовательности $\{v_k\}_1^\infty$ такой, что $v_{k+1} > v_k$ $\lim_{k \rightarrow \infty} v_k = \infty$

$$F(q_1, \dots, q_n) = \sum_{i=1}^n (P_\alpha + a_i - b_i q_i^2 - R_s \sum_{j=1}^n q_j),$$

$$B_k(q_1, \dots, q_n) = -v_k \sum_{i=1}^n \left(\frac{q_i^2}{P_\alpha + a_i - b_i q_i^2 - R_s \sum_{j=1}^n q_j} + \frac{1}{q_i - q_i^{*+}} + \frac{1}{q_i^{*++} - q_i} \right).$$

При использовании этого метода начальное приближение должно принадлежать допустимой области. В качестве начального приближения можно выбрать оптимальный с точки зрения максимальной устойчивости режим.

4. Работа активных элементов в системе с несколькими входами

На рис. 2 схематически показан граф инженерной сети с двумя входами. Реальные инженерные сети (например, системы водоснабжения), как правило имеют несколько входов и достаточно большое количество выходов. Поэтому при моделировании технологических процессов управления активными элементами в таких сетях вместо обобщенного уравнения (3), описывающего работу распределительной части сети, необходимо использовать систему, уравнения которой задают функциональную связь между значениями давлений в начале и конце элемента сети и расходом целевого продукта, а также учитывают соблюдение первого закона Кирхгофа для каждого узла.

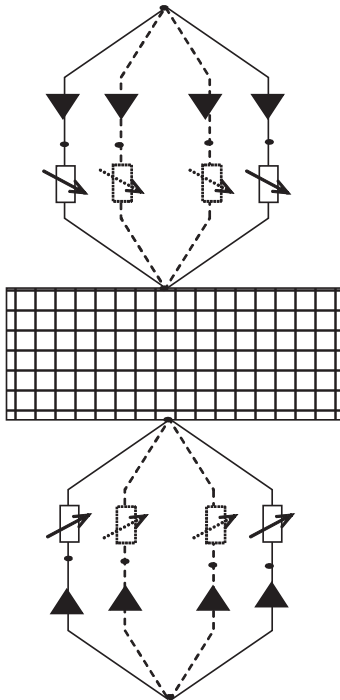


Рис. 2. Пример графа инженерной сети с двумя входами

Математическая модель инженерной сети с несколькими входами и выходами может быть представлена следующей системой уравнений:

$$P_{ki} - P_{ni} = a_i - b_i q_i^2, \quad i \in M_a, \quad (22)$$

$$P_{ni} - P_{ki} = r_i q_i^2, \quad i \in M_r, \quad (23)$$

$$P_{ni} - P_{ki} = c_i q_i^2, \quad i \in M_c, \quad (24)$$

$$P_{ni} - P_{ki} = R_i q_i^2, \quad i \in M_p, \quad (25)$$

$$\sum_{i \in M_j^+} q_i - \sum_{i \in M_j^-} q_i, \quad j \in V. \quad (26)$$

В этой системе: P_{ni} и P_{ki} — давление в начальном и конечном соответственно узлах i -ой дуги; q_i — расход целевого продукта по i -ой дуге; a_i, b_i — параметры, характеризующие работу используемых активных элементов; r_i — сопротивления пассивных элементов, которые могут задаваться оператором, для выбора режима работы активных элементов; c_i — сопротивления пассивных дуг распределительной части сети, которые не меняются; R_i — сопротивления пассивных дуг, соответствующих потребителям сети; эти значения меняются в процессе функционирования сети, но считается, что в момент включения или выключения активных элементов они не изменяются; M_a — множество индексов, соответствующих активным дугам; M_r — множество индексов, соответствующих дугам с изменяемыми сопротивлениями; M_c — множество индексов, соответствующих дугам с постоянными сопротивлениями; M_p — множество индексов, соответствующих дугам, описывающим состояние потребителей сети; V — множество индексов узлов сети; M_j^+ — множество индексов дуг, по которым целевой продукт поступает в i -ый узел; M_j^- — множество индексов дуг, по которым целевой продукт отбирается из i -го узла.

Считаем, что для решения задачи моделирования технологических процессов управления активными элементами значения P_{ni} ($i \in M_a$), P_{ki} ($i \in M_p$), a_i, b_i ($i \in M_a$), r_i ($i \in M_r$), c_i ($i \in M_c$), R_i ($i \in M_p$) нам известны. В этом случае значения P_{ki} ($i \in M_a$), P_{ni}, P_{ki} ($i \in M_r \cup M_c$), P_{ni} ($i \in M_p$), q_i ($i \in M_a \cup M_r \cup M_c \cup M_p$) могут быть определены в результате решения системы (22)–(26). Методы решения таких систем хорошо разработаны и описаны в [2–4].

В отличие от инженерных систем с одним входом для систем с несколькими входами при задании значений q_i ($i \in M_a$) нет возможности получить аналитическое выражение для P_{ki} ($i \in M_r$). Значения P_{ki} ($i \in M_r$) могут быть получены только в результате решения системы (24) – (26) при заданных значениях расходов целевого продукта на входах сети и давлений на ее выходах. В данном случае величины P_{ki} ($i \in M_r$) мы можем считать

неявными функциями от вектора расходов целевого продукта по активным элементам.

При решении поставленных в этой статье задач поиска допустимых и оптимальных в смысле критериев (10) и (18) режимов работы активных элементов сети необходимо для каждого набора значений q_i ($i \in M_a$) решать систему (24) – (26).

Выводы

Задача моделирования технологических процессов управления активными элементами является важной составляющей частью задачи построения системы управления режимами функционирования инженерных сетей. Основными управляющими воздействиями в инженерной сети являются включение или отключение активных элементов. Решение об этом принимается диспетчером, контролирующим режимы работы сети и управляющим ее функционированием. Перед принятием решения о включении или выводе из работы активных элементов диспетчер должен иметь возможность оценить последствия принятого им решения, а это возможно сделать только путем расчета предложенной в работе математической модели.

Кроме того, диспетчер может выбрать также оптимальный в смысле критерия устойчивости режим, или режим, обеспечивающий минимальные потери на регулирующей аппаратуре (задвижках) при соблюдении заданного уровня устойчивости. Параметры такого режима он может передать операторам непосредственно обслуживающим активные элементы.

Процедуры, позволяющие рассчитать допустимые и наиболее устойчивые режимы, предложенные в статье, достаточно просты и эффективны. Для поиска режима с наименьшими энергетическими потерями предложен стандартный подход,

применяемый при минимизации функций с ограничениями. В плане развития этих работ было бы целесообразно провести исследования и разработать метод определения оптимального по энергетическим потерям режима с учетом специфики оптимизируемой функции.

Список литературы: 1. Раскин Л. Г. Анализ сложных систем и элементы теории оптимального управления. — М.: Сов. Радио, 1976. — 344 с. 2. Евдокимов А. Г., Дубровский В. В., Тевяшев А. Д. Потокораспределение в инженерных сетях. — М.: Стройиздат, 1979. — 199 с. 3. Евдокимов А. Г., Тевяшев А. Д. Оперативное управление потокораспределением в инженерных сетях. — Х.: ХГУ, 1980. — 144 с. 4. Евдокимов А. Г. Минимизация функций и ее приложения к задачам автоматизированного управления инженерными сетями. — Х.: Вища шк., 1985. — 288 с.

Поступила в редколлегию 24.09.2008

УДК 519.62

Моделювання технологічних процесів керування активними елементами / Г.Г. Четвериков, Т.М. Федорова // Біоніка інтелекту: наук.-техн. журнал — 2008. — № 2 (69). — С. 84-88.

У статті дано математичну постановку і приведено методи рішення задач пошуку стійкого режиму інженерних мереж, що мають один або кілька входів. Розглянуто також задачу визначення оптимального, з погляду енергетичних утрат, режиму.

Л. 2. Бібліогр.: 4 найм.

UDK 519.62

Simulation technological processes to control active elements / G.G. Chetverikov, T.M. Fedorova // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 84-88.

In article the mathematical statement is given and the methods of problem solving of looking up of a stable conditions of engineering networks having one or several entrances are adduced. The problem of definition optimum regime is reviewed also, from the view of power losses.

Fig. 2. Ref.: 4 items.

УДК 519.876.5:004.81



ОРІЄНТУВАЛЬНИЙ РЕФЛЕКС ЯК ЗАДАЧА МОДЕЛЮВАННЯ ОБРАЗНОГО МЕХАНІЗМУ ОПЕРАТИВНОЇ ПАМ'ЯТІ ЛЮДИНИ

О.В. Бісікало

Вінницький державний аграрний університет, м. Вінниця, Україна, bisikalo@vsau.org

Запропоновано формальний підхід до взаємодії феноменологічних складових образного механізму оперативної пам'яті людини. Розглянуто алгебраїчну модель орієнтувального рефлексу у вигляді алгоритму навігації в образному просторі.

ОРІЄНТУВАЛЬНИЙ РЕФЛЕКС, АНСАМБЛЬ ОБРАЗІВ, ВЕКТОР ЕМОЦІЙ, ФОКУС УВАГИ, АЛГЕБРАІЧНА МОДЕЛЬ

Введення

Парадокси асоціативного мислення тривалий час привертають до себе прискіпливу увагу дослідників з різних наукових галузей. Проте навіть в психології, яка починалася з асоціативної психології та в межах якої З. Фрейдом було розроблено метод вільних асоціацій як основу психоаналізу, визнається феноменологічний характер складових образного асоціативного мислення [1]. Відповідно до цього моделювання штучного інтелекту і дотепер не може ґрунтуватися на сталих та загальновизначених наукових засадах [2]. Актуальність досліджень в області моделювання механізму образного мислення визначається широким спектром впливу окресленої проблеми на сучасну програмну індустрію.

Аналіз останніх досліджень і публікацій. З функціональної точки зору маніпулювання образами в оперативній пам'яті людини схоже на базові операції сортування і пошуку даних, які належать до класичних основ комп'ютерних технологій [3]. Проте необхідно враховувати, що з механізмом оперативної пам'яті тісно пов'язані такі відомі психологічні феномени, як ансамбль образів, вектор емоцій, орієнтувальний рефлекс, надоперативна, асоціативна і довготривала пам'ять [2].

Слід також розрізняти два режими інтелектуальної діяльності — в природних умовах фізіологічний орієнтувальний рефлекс призводить до збудження аферентних виконавчих механізмів і, внаслідок цього, появи нових образів через надоперативну пам'ять [4]. У складнішому випадку виключно вербального сприйняття інформації поява нових образів в ансамблі оперативної пам'яті можлива тільки через асоціативну і довготривалу пам'ять [5]. Більшість авторів відзначають, що існуючі підходи до моделювання операцій образного мислення мають на даний момент скоріше концептуальний, ніж практичний характер [6, 7].

Виділення невирішених раніше частин проблеми. Значна частина досліджень щодо вибору і пошуку потрібних образів зосереджена в області фізіології і психології. Проте практично відсутні публікації щодо механізму взаємодії згаданих психологічних феноменів в контексті оперативної або коротко-

часної пам'яті людини. Не перевірена у повній мірі гіпотеза про те, що саме образ слід вважати тією одиницею спілкування, в процесі дослідження якої можна буде описати взаємодію всіх чинників, що визначають успішну комунікацію [8]. У зв'язку з цим необхідно відповісти на запитання — для чого і як змінюється ансамбль образів оперативної пам'яті, яка при цьому виконує роль фокусу уваги.

Постановка завдання. Виходитимемо з того, що в кожен дискретний проміжок часу оперативна пам'ять об'єднує 7 ± 2 активних образів, на один з яких спрямовано фокус уваги. Принципи побудови механізму оперативної пам'яті мають показати, як в процесі зміни початкового складу ансамблю образів можна знайти і обрати за певним числовим критерієм такий образ, який був би найбільш або прийнятно наближеним до ідеального образу—мети. Вихідними даними для пошуку та вибору в загальному випадку можна вважати множину існуючих образів—ознак невідомого образу—мети. Зрештою формально приведемо завдання до алгоритму навігації між вузлами булеану [9].

1. Концептуальна модель оперативної пам'яті

Визначення основних понять концептуальної моделі оперативної пам'яті обмежує формальні рамки психологічних феноменів, що є значимими з точки зору запропонованого підходу.

Вважатимемо *ансамблем образів* (АО) сукупність від 5 до 9 образів, що одночасно знаходяться в оперативній пам'яті та доступні для усіх задекларованих операцій з образами.

Вектор емоцій (ВЕ) є числовою характеристикою образу з погляду унікальної системи переваг, бажань та цінностей людини. Оскільки природна фізіологія базується на складних гормональних механізмах емоцій, недоступних в принципі для штучного інтелекту, обмежимо значення ВЕ сумою асоціативних зв'язків даного образу з певною підмножиною інших образів, заданих як «центри емоцій». В задачах пошуку та вибору образів значенням ВЕ можна вважати сумарну силу асоціативного зв'язку даного образу з образами—ознаками мети поточного пошуку.

Ваги образів визначаються в результаті розрахунку ВЕ для усіх складових АО, що знаходяться в оперативній пам'яті. Вага образу, в загальному випадку, є змінною в часі величиною, яка залежить, наприклад, від ступеня «новизни» образу в АО.

Фокус уваги може бути представлений у вигляді вказівника, який в кожен момент часу спрямований на образ з найбільшою вагою в АО.

Для продуктивної роботи образного механізму оперативна пам'ять може взаємодіяти з іншими типами пам'яті. Це, в першу чергу, багатофункціональний *додатковий стек* для планування та регулювання процесу вибору, що має свій фізіологічний прототип в лобових долях головного мозку [3]. Джерелом нових образів для моделі будемо вважати *асоціативну та довготривалу пам'ять*, які формалізуються у вигляді простору асоціативних пар [9]. Підхід, що пропонується, побудований на тому, що всі розглянуті види пам'яті обмінюються образами як неподільними операційними одиницями, а змістовне навантаження образних конструкцій визначається через асоціативні зв'язки між парами образів [5].

Тоді, значно обмежуючи в рамках запропонованого підходу відомий психологічний феномен, вважатимемо *орієнтувальним рефлексом* алгоритм, що дозволяє уникати зациклення в роботі механізму оперативної пам'яті.

Визначення загальних принципів моделювання образного механізму оперативної пам'яті дозволяє враховувати особливості природного вирішення задач, пов'язаних з пошуком та вибором кінцевого образу—мети. Запропоновані в [10] принципи декларують можливість розв'язання поставленої задачі на основі визначених раніше понять.

Принцип 1. АО оперативної пам'яті може змінюватися в кожному такті часу тільки на один образ, а саме:

а — образ додається до ансамблю оперативної пам'яті, при цьому загальна кількість образів не може перевищувати 9;

б — образ вилучається з ансамблю оперативної пам'яті, при цьому загальна кількість образів не може бути меншою за 5;

с — новий образ додається до АО, а замість нього вилучається інший образ.

Фізіологічною підставою цього принципу є особливості *надоперативної пам'яті* людини, в яку одночасно і паралельно потрапляють різноманітні сигнали зі всіх органів чуття, проте моменти часу «розпізнавання» образів з безлічі сигналів можна вважати послідовними та кратними дискретному проміжку часу моделювання [3].

Принцип 2. При появі кожного нового образу з інших видів пам'яті для нього в той же такт часу визначається вага. Тому можна вважати, що значення ВЕ в кожному дискретному такті часу відоме

для всіх образів оперативної пам'яті, причому окремим фіксуються складові АО з мінімальною і максимальною (у якості фокусу уваги) вагами.

Принцип 3. Власне механізм оновлення АО здійснюється таким чином:

а — якщо вага нового образу більша за максимальну вагу, то склад АО змінюється за принципом 1а або 1с (новий образ замінює образ з найменшою вагою);

б — якщо вага нового образу менша за мінімальну вагу, то склад АО не змінюється;

с — проміжне значення ваги нового образу приводить до застосування принципу 1а або 1с (так само новий образ замінює в АО образ з найменшою вагою).

Принцип 4. Під час кожного модельного такту образу у фокусі уваги ставиться у відповідність відсортований у порядку зменшення ряд найбільш близьких за силою асоціативного зв'язку образів, для яких послідовно визначається вага. За необхідності значення ВЕ також можуть бути розрахованими для образів певної події, пов'язаної з образом у фокусі уваги.

Принцип 5. Суть обміну образів в оперативній пам'яті полягає у тому, що при кожній наступній ітерації склад АО глобально поліпшується з погляду сумарного значення ВЕ. Образний механізм оперативної пам'яті закінчує свою роботу при досягненні достатнього значення ваги для образу у фокусі уваги.

Принцип 6. Головна проблема образного механізму полягає у зацикленні, яке принципово можливе в наступних випадках:

а — жоден з ряду образів, що асоціюються з образом у фокусі уваги, не набирає потрібної ваги для входження в АО;

б — образ у фокусі уваги не змінюється, оскільки нові образи—претенденти до АО набирають меншу вагу, ніж у нього;

с — класичне зациклення — фокус уваги переходить по замкнутому колу всередині групи з двох, трьох чи більше образів.

Принцип 7. З метою запобігання зациклення необхідно відстежувати розглянуті вище випадки і адекватно на них реагувати, а саме:

а — вноситься на розгляд нова підмножина образів за допомогою непрямої асоціації через події, пов'язані з образом у фокусі;

б — образ у фокусі уваги переноситься з АО в стек контролю за принципом 1б або 1с (найвагоміший з образів-претендентів замінює в АО образ з найбільшою вагою), а фокус уваги спрямовується на другий за значенням ваги образ АО;

с — *n*-вимірне зациклення визначається тоді, коли в стеку контролю з'являється дубль вже наявного у ньому образу — в цей момент необхідно закінчити роботу образного механізму з видачею

повідомлення про неможливість вирішення задачі пошуку прийнятного з точки зору БЕ образу за існуючих даних.

Моделі задач оперативної пам'яті. До складу основних задач, які природно розв'язує оперативна пам'ять людини, можна віднести [2]:

- орієнтувальний рефлекс (пошук типу АО → образ);
- вибір образу–рішення (вибір типу стек → АО → образ);
- генерацію повідомлення (побудова типу АО → синтагма);
- відповідь на запитання (побудова типу синтагма → АО → синтагма).

Моделювання цих задач може бути досягнуте за рахунок введення певних обмежень в розглянутий концептуальний механізм оперативної пам'яті.

2. Формалізація механізму оперативної пам'яті

На основі введених вище принципів можна запропонувати формальну постановку задачі алгебраїчного моделювання механізму оперативної пам'яті.

Будемо вважати булеаном множину усіх підмножин множини образів *Image* [9]. Якщо поставити у відповідність кожній з цих підмножин бінарний код, то власне образи знаходяться на першому шарі булеану та мають бінарні коди *Bi-I* тільки з однією одиницею та іншими нулями. Тоді бінарний код *Bi-Sy* (синтагми) чи *Bi-OM* (оперативної пам'яті) довільного *i*-го шару складається з *i* одиниць та *n-i* нулів, де *n* – потужність множини *Image*. В межах запропонованого підходу [11] досягається накопичення сили прямого *Force+* та зворотного *Force-* асоціативного зв'язку у парах образів з бінарними кодами *Bi-I-1.Assoc-Twice* та *Bi-I-2.Assoc-Twice*.

Для реалізації моделі будемо використовувати такі змінні:

- *Weigh-Current* – вага поточного образу;
- *Bi-Current* – бінарний код поточного образу;
- *Focus-Weight* – вага образу у фокусі уваги (найбільша з образів АО);
- *Focus-Bi* – бінарний код образу у фокусі уваги;
- *Weight-Min* – найменша вага образу в АО;
- *Bi-Min* – бінарний код образу АО з найменшою вагою;
- *Weigh-Top* – достатнє (прийнятне) значення БЕ;

та наступні додаткові конструкції:

- *Vector-Set* – стек з множиною образів–складових вектора емоцій;
- *Choice-Set* – стек з множиною нових образів–претендентів для вибору;
- *Check-Set* – стек для контролю за процесом вибору образу.

Розглянемо формальні операції і предикати на булеані, які відповідають вищевикладеним принципам.

1. Операція переходу на вузол вищого шару *Level-Up* (до АО з кодом *Bi-OM* додається образ з кодом *Bi-I*) і предикат *Top-Border*, який істинний тоді, коли не досягнута верхня межа обсягу оперативної пам'яті – принцип 1а:

$$Bi-OM \cup Bi-I \xrightarrow{Level-Up} Bi-OM ; \quad (1a)$$

$$Level < 9 \rightarrow Top-Border . \quad (1b)$$

2. Операція переходу на вузол нижчого шару *Level-Down* (від АО з кодом *Bi-OM* віднімається образ з кодом *Bi-I*) і предикат *Bottom-Border*, який істинний тоді, коли не досягнута нижня межа обсягу оперативної пам'яті – принцип 1б:

$$Bi-OM \cap (\neg Bi-I) \xrightarrow{Level-Down} Bi-OM ; \quad (2a)$$

$$Level > 5 \rightarrow Bottom-Border . \quad (2b)$$

3. Операція переходу на інший вузол поточного шару *Change-Image* (в АО з кодом *Bi-OM* образ *Bi-Current* замінює образ *Bi-I*) – принцип 1с.

$$(Bi-OM \cap (\neg Bi-I)) \cup Bi-Current \xrightarrow{Change-Image} Bi-OM . \quad (3)$$

4. Операція *Evaluate-Vector*, що визначає БЕ для образу *Bi-Current* – принцип 2:

$$Evaluate-Vector(Bi-Current) ::= \{[Vector-Set] \sum_{i=1}^k (Force- + Force+ | Bi-I-1.Assoc-Twice = Bi-Current \cup Bi-I-2.Assoc-Twice = Bi-I_i)\}, \quad (4)$$

де *k* – загальна кількість образів – складових вектора емоцій *Vector-Set*.

5. Предикат *Weight+*: показує, що вага поточного образу більша за максимальну вагу в АО – принцип 3а:

$$Weight-Current > Focus-Weight \rightarrow Weight+ . \quad (5)$$

6. Предикат *Weight-*: показує, що вага поточного образу менша за мінімальну вагу АО – принцип 3б:

$$Weight-Current < Weight-Min \rightarrow Weight- . \quad (6)$$

7. Операція побудови хеш-таблиці образів *Hash-Table*, що асоціюються з образом *Focus-Bi* – принцип 4:

$$Focus-Bi \xrightarrow{Hash-Table} Choice-Set ; \quad (7)$$

8. Предикат *Stop-Find*, що показує закінчення успішного пошуку образу, – принцип 5:

$$Focus-Weight \geq Weight-Top \rightarrow Stop-Find . \quad (8)$$

9. Предикат *Assoc-False*: показує, що ні один образ-претендент з *Choice-Set* не набрав потрібної ваги для входження в АО – принцип 6а:

$$\{[Choice - Set : n] \text{ Weight} - = True\} \rightarrow \text{Assoc} - False, \quad (9)$$

де n – кількість образів в стеку для вибору *Choice-Set*.

10. Предикат *New-False* : показує, що образ у фокусі уваги не змінився, оскільки нові образи-претенденти до АО мають меншу вагу, ніж у нього – принцип 6б:

$$\{[Choice - Set : n] \text{ Weight} + = True\} \rightarrow \text{New} - False. \quad (10)$$

11. Операція *Add-Check*, яка додає образ *Focus-Bi* до стека контролю *Check-Set*:

$$\text{Focus} - Bi \xrightarrow{\text{Add-Check}} \text{Check} - Set. \quad (11)$$

12. Операція *Image-to-Event* для пошуку події, найтісніше пов'язаної з образом у фокусі уваги *Focus-Bi* – принцип 7а:

$$\text{Focus} - Bi \xrightarrow{\text{Image-to-Event}} Bi - Sy. \quad (12)$$

13. Операція *Update-Choice*, яка заносить образи події в стек *Choice-Set*:

$$Bi - Sy \xrightarrow{\text{Update-Choice}} \text{Choice} - Set. \quad (13)$$

14. Предикат *Stop-False*, що показує неуспішне закінчення пошуку, оскільки в стеку *Check-Set* з'явився дубль вже існуючого там образу – принципи 6с і 7с:

$$\{[Check - Set : m] \text{ Focus} - Bi = Bi - I.Check\} \rightarrow \text{Stop} - False, \quad (14)$$

де m – кількість образів у стеку контролю *Check-Set*; *Bi-I.Check* – бінарний код чергового образу в стеку контролю.

Тоді формальною постановкою задачі моделювання орієнтувального рефлексу можна вважати:

$$Bi - OM, \text{ Vector} - Set \rightarrow \text{Focus} - Bi | \text{Focus} - Weight \geq \text{Weight} - Top. \quad (15)$$

3. Алгебраїчна модель орієнтувального рефлексу

Для розв'язання поставленої задачі (15) з урахуванням формальних операцій та предикатів (1)-(14) представимо двохосновну алгебраїчну систему у вигляді:

$$\text{Algebra} = \langle B; \Omega \rangle, \quad (16)$$

де

$$B = \{Image, Bi\} \quad (17)$$

основи, а

$$\Omega = \{OP, IF\} \quad (18)$$

– сигнатура системи, що складається з операцій *OP* та предикатів *IF*. В якості основ використовуються *Image* – множина образів та *Bi* – бінарні коди множини всіх підмножин (булеану) *Image*. До складу операцій та предикатів системи входять розглянуті раніше

$$OP = \left\{ \begin{array}{l} \text{Level} - Up, \text{Level} - Down, \text{Change} - Image, \\ \text{Evaluate} - Vector, \text{Hash} - Table, \text{Add} - \\ \text{Check}, \text{Image} - to - Event, \text{Update} - Choice \end{array} \right\}, \quad (19)$$

$$IF = \left\{ \begin{array}{l} \text{Top} - Border, \text{Bottom} - Border, \\ \text{Weight} +, \text{Weight} -, \text{Stop} - Find, \\ \text{Assoc} - False, \text{New} - False, \text{Stop} - False \end{array} \right\}. \quad (20)$$

Покажемо, що, задача моделювання механізму оперативної пам'яті в межах алгебраїчної системи *Algebra* фактично зводиться до алгоритму навігації між вузлами (підмножинами) булеану від 5-го шару до 9-го на основі операцій *OP* та предикатів *IF*. В граф-схемах для алгебраїчних конструктів, що будуть розглядатися з цією метою, використані такі позначення для операторів структурного програмування [12]:

Do – цикл за параметром або за умовою;
If (+ -) – альтернатива;
* – композиція.

Достатньою для розв'язання задачі (15) є послідовна побудова наступних операторів з сигнатури (18).

1) Оператор *Op-Plus*, що додає до АО новий образ (*Bi-Current*) або замінює образ *Bi-Min* на образ *Bi-Current* – якщо не досягнуто верхнього кордону оперативної пам'яті (предикат *Top-Border* істинний), то виконується оператор *Level-Up*, інакше виконується оператор *Change-Image*:

$$\text{Op} - Plus ::= ([\text{Top} - Border] \text{Level} - Up, \text{Change} - Image). \quad (21)$$

Граф-схему оператора *p-Plus* представлено на рис. 1.

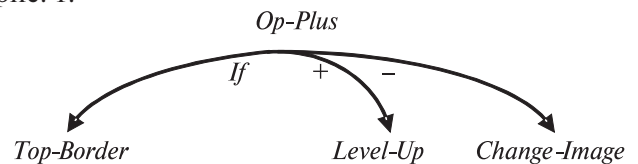


Рис. 1. Граф-схема оператора *Op-Plus*

2) Оператор *Op-Minus* (рис. 2), що вилучає один образ (*Focus-Bi*) з АО або замінює образ *Focus-Bi* на образ *Bi-Current* – якщо не досягнуто нижнього кордону оперативної пам'яті (предикат *Bottom-Border* істинний), то виконується оператор *Level-Down*, інакше виконується оператор *Change-Image*:

$$\text{Op} - Minus ::= ([\text{Bottom} - Border] \text{Level} - Down, \text{Change} - Image). \quad (22)$$

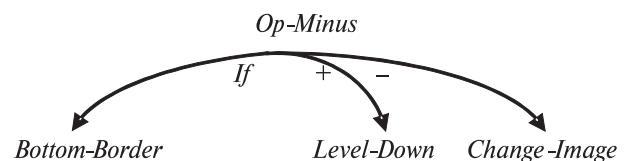


Рис. 2. Граф-схема оператора *Op-Minus*

3) Оператор *Mechanism* (рис. 3) для моделювання механізму обміну образами в АО – якщо вага поточного образу більша за вагу фокуса уваги

(предикат *Weight+* істинний), то виконується оператор *Op-Plus*, а предикат *New-False* отримує значення *False*, інакше, якщо вага поточного образу більша та дорівнює мінімальній вазі АО, то виконується оператор *Change-Image*, що заміняє в АО образ *Bi-Min* на образ *Bi-Current*, а предикат *Assoc-False* отримує значення *False*:

$$\begin{aligned} Mechanism ::= & ([Weight+] (Op - Plus^* \\ & New - False : False), ([Weight-] E, \\ & (Change - Image^* Assoc - False : False))), \end{aligned} \quad (23)$$

де E – оператор, що не виконує жодних дій.

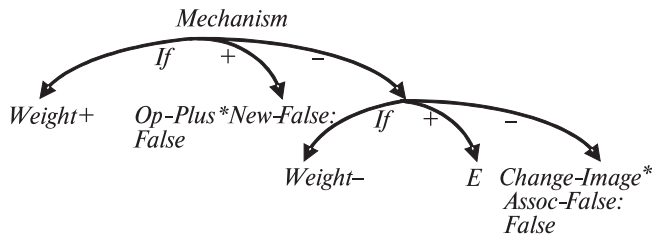


Рис. 3. Граф-схема оператора Mechanism

4) Оператор *Mech-Begin* (рис. 4), що моделює запуск механізму оперативної пам'яті – предикатам *New-False* та *Assoc-False* надається початкове значення *True* та вперше запускається операція *Hash-Table*:

$$\begin{aligned} Mech-Begin &::= Assoc-False:True^* \\ New-False:True^* Hash-Table. \end{aligned} \quad (24)$$

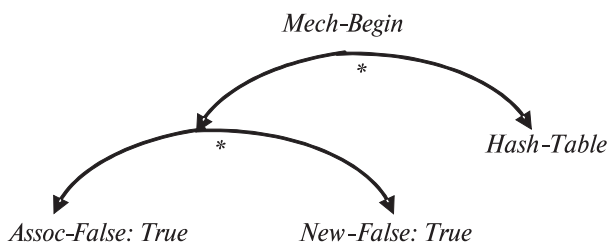


Рис. 4. Граф-схема оператора Mech-Begin

5) Оператор *Focus-Plus* (рис. 5) для зміни фокусу на крайній з точки зору значення ВЕ – виконується операція *Add-Check*, яка додає образ *Bi-Current* до стеку контролю *Check-Set* та перевіряється предикат *Stop-False*, якщо такий самий образ в стеку контролю уже присутній, то алгоритм закінчується операцією *Abort*, інакше значення *Bi-Current* присвоюється змінній *Focus-Bi*, а потім виконується операція *Hash-Table*:

$$\begin{aligned} Focus - Plus ::= & Add - Check * ([Stop - False] \\ Abort, Focus - Bi = & Bi - Current) * Hash - Table. \end{aligned} \quad (25)$$

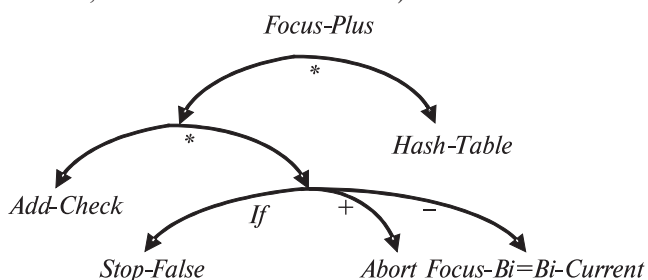


Рис. 5. Граф-схема оператора Focus-Plus

6) Оператор *Focus-Minus* (рис. 6) для зміни фокусу на гірший з точки зору значення ВЕ – виконується оператор *Op-Minus*, потім змінна *Bi-Current* отримує найбільше значення ваги образу з АО та виконується оператор *Focus-Plus*:

$$\begin{aligned} Focus - Minus &::= Op - Minus^* \\ Bi - Current &= Max(Bi_i)^* Focus - Plus. \end{aligned} \quad (26)$$

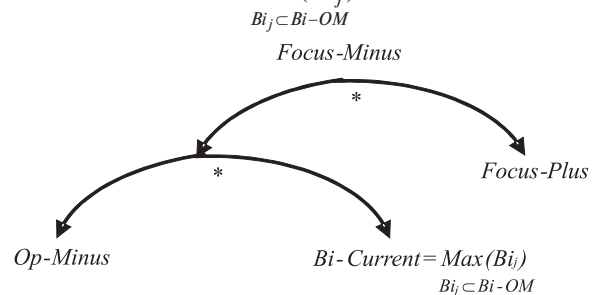


Рис. 6. Граф-схема оператора Focus-Minus

7) Оператор *New-Event* (рис. 7), що моделює згадування події за асоціацією з образом — послідовно виконуються операції *Image-to-Event* та *Update-Choice*:

$$\begin{aligned} \text{New-Event} &::= \text{Image-to-Event}^* \\ \text{Update-Choice.} \end{aligned} \quad (27)$$

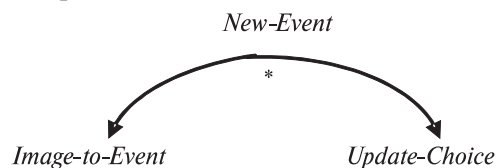


Рис. 7. Граф-схема оператора New-Event

8) Оператор *Operative-Memory* (рис. 8) для моделювання мнемонічних процесів в оперативній пам'яті — перед початком головного циклу виконується оператор *Mech-Begin* та, за кількістю образів n в стеку вибору *Choice-Set*, перевіряється предикат *New-False*, якщо фокус уваги не змінився, то виконується оператор *Mechanism*, в іншому випадку — оператор *Focus-Plus*, після головного циклу перевіряється предикат *Assoc-False*, якщо склад АО не обновлювався, то виконується оператор *New-Event*, інакше — оператор *Focus-Minus*:

$$\begin{aligned} \text{Operative} - \text{Memory} &::= \text{Mech} - \text{Begin}^* \\ \{[\text{Choice} - \text{Set} : n] \ (\&[\text{New} - \text{False}] \ \text{Mechanism}, \\ \text{Focus} - \text{Plus})\}^* (\&[\text{Assoc} - \text{False}] \\ \text{New} - \text{Event}, \text{Focus} - \text{Minus}). \end{aligned} \quad (28)$$

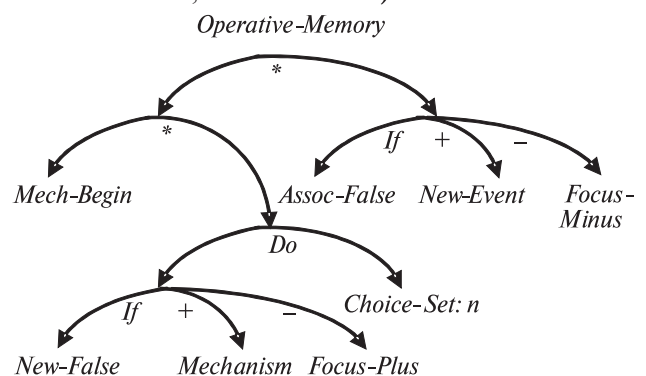


Рис. 8. Граф-схема оператора Operative-Memory

9) Оператор *Orient-Reflect* (рис. 9) для моделювання орієнтовного рефлексу — оператор *Operative-Memory* виконується в циклі доки предикат *Stop-Find* не стане істинним:

$$\text{Orient-Reflect} ::= \{[\text{Stop-Find}] \text{ Operative-Memory}\}. \quad (29)$$

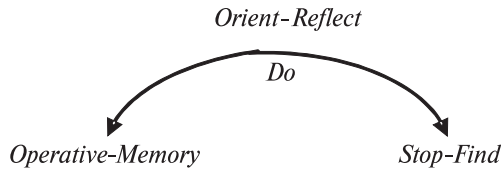


Рис. 9. Граф-схема оператора Orient-Reflect

Алгоритм моделювання орієнтувального рефлексу за допомогою алгебраїчних операторів (21)÷(29) програмно реалізовано на основі технології DrScheme + SQLite, яка поєднує можливості мов Lisp та SQL.

Висновки

В роботі визначені поняття та сформульовано принципи, що дозволяють послідовно побудувати концептуальну модель механізму оперативної пам'яті людини за допомогою формального погляду на такі відомі феномени як ансамбль образів, вектор емоцій, фокус уваги, орієнтувальний рефлекс. Подальша формалізація запропонованої концептуальної моделі призвела до алгебраїчної постановки задачі алгоритму навігації між вузлами булеану від 5-го шару до 9-го, які відповідають ансамблю образів або подіям.

На основі введених обмежень відомого психологічного феномену орієнтувальний рефлекс розглядається як одна з задач оперативної пам'яті. Для вирішення поставленої задачі використано двохосновну алгебраїчну систему (16)÷(20), яка дозволяє шляхом введення послідовності з операторів (21)÷(29) представити орієнтувальний рефлекс у вигляді алгоритму навігації в образному просторі.

Перспективним напрямком розвитку запропонованого алгебраїчного підходу можна вважати побудову алгоритмів для розв'язання інших задач оперативної пам'яті, що традиційно належать до предметної області лексичного процесора — вибір образу—рішення, генерацію повідомлення та отримання відповіді на запитання.

Список літератури: 1. Основные направления психологии в классических трудах. Ассоциативная психология. Г. Эббингауз. Очерк психологии. А. Бэн. Психология. — М.: ООО «Издательство АСТ-ЛТД», 1998. — 544 с. 2. Лурия А.Р. Язык и сознание / Под ред. Е.Д.Хомской. — М., Изд-во Московского университета, 1979. — 320 с. 3. Соколов Е.Н., Вайткявичус Г.Г. Искусственный интеллект: от нейрона к нейрокомпьютеру. — М.: Наука, 1990. — 237 с.

4. Данилова Н.Н., Крылова А.Л. Физиология высшей нервной деятельности: Учебник. — М.: Изд-во МГУ, 1989. — 399 с. 5. Бисикало О.В. Принципы построения концептуальной модели образного мышления: В расширенных материалах первой международной конференции «Новые информационные технологии в образовании для всех», (29-31 мая 2006 г.). — Киев, 2006. — С. 25-34. 6. Демьянков В.З. «Теория речевых актов» в контексте современной лингвистической литературы: (Обзор направлений) // Новое в зарубежной лингвистике: Вып. 17. Теория речевых актов. М.: Прогресс, 1986. С. 223-235. 7. Валькман Ю.Р., Исмаилова Л.Р. О языке образного мышления: Докл. междунар. конф. «Диалог 2004». — С. 90-97. 8. Демьянков В.З. Предикаты и концепция семантической интерпретации // Известия АН СССР. Серия литературы и языка. — 1980. — Т. 39, № 4. — С. 336-346. 9. Бісикало О.В. Дослідження простору асоціативних пар в контексті бази знань електронного підручника // Вимірювальна та обчислювальна техніка в технологічних процесах. — 2006. — № 2 (28). — С. 109-113. 10. Бисикало О.В. Подход к моделированию образного механизма оперативной памяти: Сб. трудов Второй междунар. конф. «Новые информационные технологии в образовании для всех: состояние и перспективы развития» 21-23 ноября 2007 г». — Киев, 2007. — С. 336-344. 11. Бісикало О.В. Структура блоку пам'яті на основі моделі образного мислення людини // Искусственный интеллект — 2007. — № 3. — С. 461-468. 12. Цейтлин Г.Е. Введение в алгоритмику. — Киев: Сфера, 1998. — 310 с.

Поступила до редколлегии 25.09.2008

УДК 519.876.5:004.81

Ориентировочный рефлекс как задача моделирования образного механизма оперативной памяти человека / О.В. Бисикало // Бионика интеллекта: науч.-техн. журнал. — 2008. — № 2 (69). — С. 89-94.

Работа посвящена формализации понятий ансамбля образов, вектора эмоций, фокуса внимания и других составляющих образного механизма оперативной памяти человека. На основе введенных ограничений известного психологического феномена ориентировочный рефлекс рассматривается как одна из задач оперативной памяти. Предлагается алгебраическая модель ориентировочного рефлекса в виде алгоритма навигации в образном пространстве.

Ил. 9. Библиогр.: 12 назв.

UDC 519.876.5:004.81

Reference reflex as task of modeling of imaging mechanism of the human operative memory / Oleg V. Bisikalo // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 89-94.

The article is devoted to the formalization of concepts of images ensemble, vector of emotions, focus of attention and other constituents of imaging mechanism of the human operative memory. The reference reflex is examined as one of tasks of the operative memory on the basis of imposed restrictions of the known psychological phenomenon. The algebraic model of the reference reflex is offered as an algorithm of navigation in imaging space.

Fig. 9. Ref.: 12 items.

УДК 519.62



ФУНКЦИОНИРОВАНИЕ ПРОГРАММНОГО АГЕНТА ИНТЕРАКТИВНОГО ВЗАИМОДЕЙСТВИЯ ПОЛЬЗОВАТЕЛЕЙ СИСТЕМЕ ДИСТАНЦИОННОГО ОБУЧЕНИЯ

С. А. Марьин¹, А. П. Мишин², В. Б. Репка³¹Харьковская государственная академия культуры, г. Харьков, Украина, maryin_sergey@mail.ru²ХНУРЭ, г. Харьков, Украина, askell@rambler.ru³ХНУРЭ, г. Харьков, Украина, victoria_repka@kture.kharkov.ua

Рассмотрен подход использования агентных технологий в организации систем дистанционного обучения. На примере агента интерактивного взаимодействия участников процесса обучения изложены основные принципы создания и функционирования программных агентов в мультиагентной системе дистанционного образования. Приведены модели и алгоритмы функционирования и взаимодействия агента.

ДИСТАНЦИОННОЕ ОБУЧЕНИЕ, АГЕНТ, ЗНАНИЯ, ЦЕЛИ, ФАКТЫ, ОНТОЛОГИЧЕСКАЯ МОДЕЛЬ, BDI, JADEX, XML

Введение

Основываясь на возможных видах взаимодействия в системах дистанционного образования (СДО): координатор-группа; координатор-слушатель; слушатель-координатор; слушатель-слушатель; группа-слушатель; слушатель-группа, можно достичь желаемого результата — повышения эффективности обучения и вовлечения слушателей в активную самостоятельную деятельность в системе ДО. В принципе для большинства людей это наиболее оптимальный выход, но тут же мы сталкиваемся с проблемой, что не всегда у человека существует возможность спокойно обучаться из-за нехватки времени, места расположения учебного заведения и проблем, связанных с теми или иными факторами, которые ограничивают доступ человека к получению образования.

СДО использует всемирную паутину как средство интерактивного взаимодействия пользователей и реальных преподавателей, организующих процесс обучения, используя для этого программные средства, базу знаний и управляющие утилиты, организующие процесс обучения. Популярными в данное время стали агентные технологии.

В данной работе предлагается создать агента интерактивного взаимодействия пользователей системы дистанционного обучения как средство, выполняющее коммутацию пользователей (координатор-слушатель; слушатель-координатор) на основании набора логически заданных функций и рекомендуемых действий и в соответствии с поставленным запросом.

1. Концепция и общее понятие мультиагентной системы дистанционного обучения

Классический подход к процессу ДО — наличие структурированных, объединённых в единый информационно-образовательный «пакет данных», предоставляемый пользователям посредством каких-либо информационных носителей или средс-

твами коммутации. Пакет содержит необходимые данные, непосредственного требующиеся для организации процесса обучения студента. На данном этапе развития информационных и телекоммуникационных технологий данный подход изжил себя. Основным недостатком такого подхода есть то, что весь процесс обучения организовывается самим обучаемым. Таким образом, целесообразно представить процесс образования как систему вида: «база данных», содержащая материалы обучения + набор утилит, непосредственно организующих процесс обучения. Для этого требуется первоначально определить функции и задачи СДО, наиболее точно имитирующие реальный процесс обучения, используя имеющиеся цели, задачи и базу знаний системы и максимально ограничивающие вмешательство в процесс обучения реального тьютора.

Решая вопросы «самостоятельности системы» с ограничением вмешательства тьютора, необходимо рассматривать наличие какого-то механизма, который будет проводить самоадаптацию в соответствии с поставленными целями и действиями пользователей. В качестве такого механизма может выступать совокупность взаимосвязанных агентов: пользователи взаимодействуют с системой через агентов, система взаимодействует с пользователями через агентов, и агенты могут взаимодействовать с агентами «советуясь» и «общаясь» друг с другом.

В данной работе для решения вышеуказанной проблемы предлагается рассмотреть агента как интеллектуальное средство, выполняющее интерактивное взаимодействие объектов на основании набора логически заданных функций и рекомендуемых действий в соответствии с поставленным запросом для повышения эффективности работы системы. Предлагается создать блок интерактивного взаимодействия системы ДО, обучаемого и тьютора посредством общения в виде передачи

текстовой информации, что позволит сократить или полностью исключить интервал адаптации студентов и преподавателей к возможностям новых средств обучения.

2. Идея разработки агента взаимодействия пользователей

Вопрос о том, как обеспечить пользователей связью при дистанционном обучении лежит на поверхности и ответ типа: используя различные средства коммутации (электронная почта, различные пейджинговые системы (ICQ, MSN и др.), текстовые и голосовые чаты, конференц-связь [1,2]) – слишком общий, поскольку не учитывает того множества факторов, которые в совокупности влияют на обучение: затрачиваемые ресурсы, интерактивность общения и синхронность функционирования в СДО.

Выявив наиболее приемлемые функции, такие как: текстовая передача данных; доставка сообщений в online/offline режимах пользователей; отсутствие установки специализированного программного обеспечения; запуск с серверной стороны; поддержка многопоточного использования системы пользователями, создадим средство интерактивного взаимодействия пользователей СДО в виде агента (рис. 1). Программный агент интерактивного взаимодействия пользователей СДО состоит из двух модулей: модуль коммутации и модуль поддержки пользователя в отсутствии тьютора (так называемый модуль чата). Модуль коммутации использует службу рассылки текстовых сообщений типа чат, устанавливая двухстороннюю связь пользователя системы и тьютора.

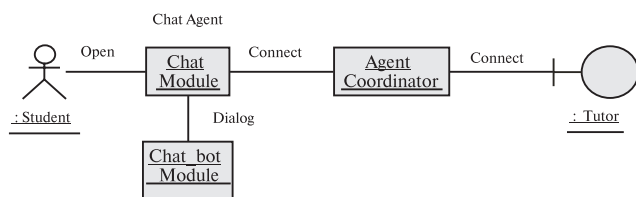


Рис. 1. Схема взаимодействия агентов с пользователями

Реагируя на заданный запрос, агент обрабатывает событие и принимает решение о связи с преподавателем; при отсутствии тьютора передаёт управление агенту, отвечающему за координацию взаимодействия в СДО. Модуль поддержки ожидает получения вопроса и выбора ответа на него. Модуль поддержки пользователя вырабатывает решение, базируясь на онтологической модели диалога.

Пользователи непосредственно взаимодействуют через агентов системы, не предпринимая никаких действий, кроме подачи запроса на установление возможности связи с участниками системы. Затем агент интерактивного взаимодействия

выполняет действия, связанные с построением диалогового окна общения пользователей и обработкой полученной информации об имеющихся пользователях. Эти действия выполняются по требованию каждого пользователя одними и теми же агентами [3].

Рассмотрим процесс совместного функционирования агентов и пользователя для более четкого определения рентабельности агента: при возникновении необходимости общения с тьютором в СДО для пользователей существует возможность запустить чат. Чат, связанный с модулем коммутации Start_Chat.agent, приводит в состояние активации класс-конструктор формирования: 1) интерфейса диалогового окна чата и 2) функцию связи с агентом координатором. Агент определяет состояние тьюторов, отправляя PING-запрос и формируя контакт-лист, после чего переходит в состояние ожидания дальнейших действий пользователя. «Пробуждение» модуля коммутации происходит в случае, если пользователь устанавливает связь с требуемым ему тьютором. Агент коммутации от лица обучаемого отправляет алерт-приглашение на коммутацию пользователей и устанавливает связь, запуская окно чата на стороне тьютора. Когда невозможно установить связь с тьютором, агент активизирует модуль Chat_bot, предоставляющий варианты выхода из ситуации, оставляя тьютору сообщение в offline-режиме для последующего ответа или связи со студентом.

3. Критерии функционирования агента Chat_Bot в системе ДО

При общении тьютора и студента диалог осуществляется по схеме вопрос-ответ. Рассмотрим все возможные варианты разговоров студента и тьютора для определения ключевых моментов диалога, которые не привязывают к конкретным вопросам и ответам, и для выявления закономерностей, дающих возможность применить их к определению задач агента Chat_Bot. Он является средством поддержки пользователей при отсутствии тьютора и выполняет роль преподавателя. Для формализации отношений понятий «Вопрос» – «Ответ» воспользуемся онтологиями [4], преимуществами которых для разрабатываемой агентной модели взаимодействия есть: однозначное определение терминов с использованием логических аксиом; логическая согласованность определений; монотонное расширение и/или специализации без ревизии существующих понятий; влияние кодирования. Концептуализация при создании онтологии специфицирована на уровне представления, а не символического кодирования. Для разработки онтологической модели O , воспользуемся следующей нотацией, – упорядоченная тройка:

$$O = (\mathcal{K}, \mathcal{R}, \Phi), \quad (1)$$

где $\aleph$ – конечное множество концептов (понятий, терминов) предметной области, которую представляет онтология O ; $\aleph$ – конечное множество отношений между концептами (понятий, терминов) заданной предметной области; Φ – конечное множество функций интерпретации (аксиоматизации), заданных на концептах и/или отношениях онтологии O .

Рассмотрев отношение «вопрос пользователя – ответ бота» в виде онтологической модели, построенной на таксономической модели вариантов, классифицируем таксоны диалога, адаптируя их к среде ДО (рис. 2).

Первоначальное распознавание вопроса происходит в соответствии с представленной веткой таксономии Y_1 (вопрос пользователя) и продвигается по цепочке, присваивая идентификаторы вероятности совпадения вопроса с элементом дерева, такие как: $Y_{1,1}$ (тип) – первоначальное со-

стояние системного запроса, который определяет вопрос и относит его к одному из вариантов: требующему или не требующему ответа для возможного нормального функционирования системы и предотвращая ситуацию, при которой невозможно предоставить ответ пользователю. Если пользователь запрашивает у системы ответ на вопрос по несуществующей тематике, агент или игнорирует вопрос, или предлагает сформулировать его более точно, осуществляя запись во временный файл, содержащий неопределенные вопросы для дальнейшей обработки преподавателем или администратором системы. Когда установлен тип вопроса, система принимает соответствующее действие по обработке запроса и переходит на следующий шаг классификации вопроса к метке $Y_{1,2}$ (суть), выполняя алгоритм определения основных сущностей категории вопроса (схема алгоритма приведена на рис. 3). Раздел категорий подразумевает набор общепринятых ключевых слов, использующихся в учебных курсах, таких как: знания, дополнения, уточнения.

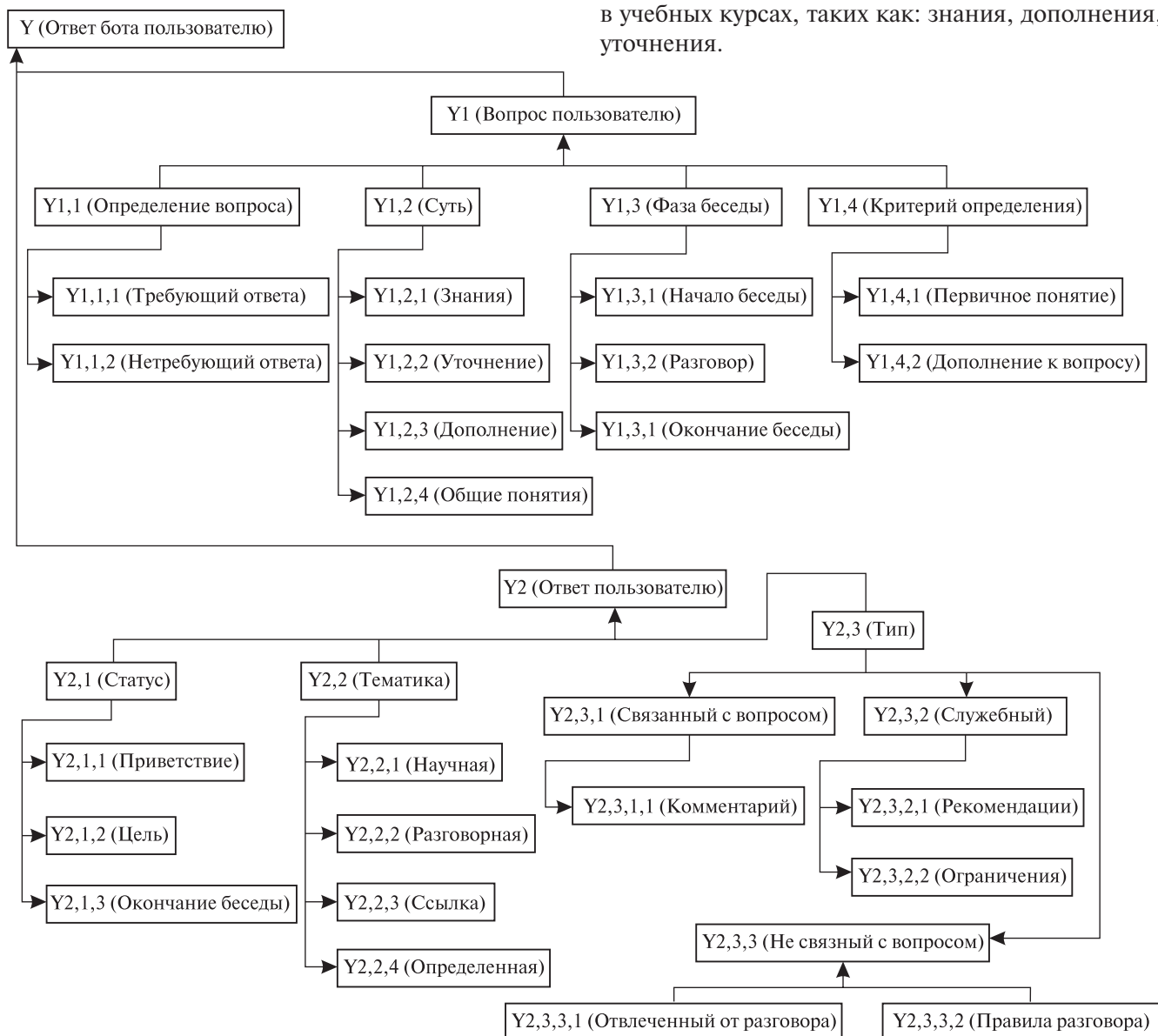


Рис. 2. Онтологическая модель разговора

Предполагается возможность уточнения у бота данного ответа и отнесения вопроса к классу критерия, расклассифицировав его на два подтипа: первичное восприятие вопроса системой или дополнение к ответу $Y_{1.4}$ (критерий определения). Тем самым происходит запись вопроса в класс ответа по уже заранее известной тематике. Предлагается также внедрить в систему плагин-ответ и плагин-вопрос $Y_{1.3}$ (статус) — используемые для определения вопроса, не внесенного в стандартную базу ответов. Расклассифицировав вопрос, чат-бот автоматически по такому же алгоритму начинает генерировать ответ пользователю.

Ответ — наиболее значимая часть работы бота, отвечающая на поставленный вопрос пользователем с минимальным расхождением от реальной сути вопроса.

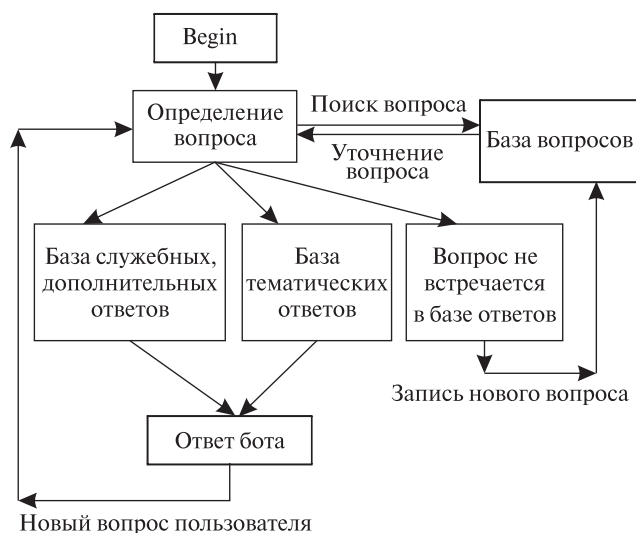


Рис. 3. Схема алгоритма прохода по онтологической модели

Агент, определив множество понятий предметной области X (вопрос) и конечное множество отношений между концептами заданной предметной области R (идентификаторы отношения к ответу) переходит к шагу определения конечного множества функций аксиоматизации, заданных на концептах Φ (ответ) для окончательного представления ответа пользователю $O(1)$. По принципу обработки вопроса агент производит такие же манипуляции с базой имеющихся у него ответов.

С известными параметрами вопроса X , R система относит вопрос к имеющимся ответам. Таксономическая модель ответа предполагает продвижение по ветке Y_2 (ответа) для выявления из базы ответов представления и отношения вопроса к ответу с целью получения Y (ответ бота пользователю). Используются следующие этапы определения наиболее адекватного ответа $Y_{2.1}$ (статус разговора) — как указано выше в разговоре имеются стадии: начало разговора $Y_{2.1.1}$ (определенное действиями агента в системе, для приветствия пользователя);

цель разговора $Y_{2.1.2}$ (реагируя на вопрос или утверждение пользователя, агент придерживается определенных тематик в ответах при разговоре); логическое завершение $Y_{2.1.3}$.

Такая онтологическая модель записывается в XML-файл, который обрабатывает агент и тем самым выполняет разговор с пользователем средствами чата от имени преподавателя.

4. BDI модель программного агента интерактивного взаимодействия пользователей в Jadex

В данной работе используется технология Jadex, которая позволяет моделировать BDI агентов с наборами фактов, целей, планов и другое [5], которые используются для обеспечения поддержки персонализированного обучения.

Функционирование агента интерактивного взаимодействия пользователей осуществляется средствами двух вложенных агентов: модуль коммутации «start_chat» и модуль поддержки пользователей в отсутствии тьютора «chat_bot».

При получении команд вызова чата из приложения СДО вызывается агент «start_chat», который начинает выполнять свои действия согласно установленному приоритету в первоначальных настройках агента в соответствии с планом (рис. 4).

Агент выполняет план «starting», который создает окно чата в соответствии с установленными параметрами и формирует список пользователей, данные о которых получены от агента координатора.

```

<initialstates>
<!-- Используется для установки выполнения первоначальных
      планов агентов -->
<initialstate name="default">
  <plans> <initialplan ref="starting"/> </plans>
  <plans> <initialplan ref="review"/> </plans>
</initialstate>
</initialstates>
  
```

Рис. 4. Начальное состояние агента

Список пользователей, определенный в базе знаний системы, представляет собой системный xml-файл, обрабатываемый java-классом «TutorList» (рис. 5). Агент интерактивного взаимодействия в соответствии с полученными адресами передаёт необходимые данные (IP адрес) агенту координатору для установления тестовой связи с тьютором.

```

<beliefs>
  <beliefset name="sdo_list_tutor" class="TutorList">
    <fact> new TutorList("user_tutor.xml",null) </fact>
  </beliefset>
</beliefs>
  
```

Рис. 5. Знания агента

Файл user_tutor.xml содержит данные, используемые агентами «coordinator» и модулем агента

коммутации «start_chat» для обеспечения интерактивной связи пользователя с тьюторами, проверки статуса и предоставления информации о тьюторе студенту [6].

Определяя статус тьюторов, агент устанавливает графические индикаторы в состояние online/offline, используя план «start_chat», путем отправки сервисного запроса каждому из тьюторов по плану «review» (рис. 8).

```

<goals>
  <!-- Запуск coordinator агента для установления связи -->
  <achievegoalref name=»coordinator»>
    <concrete ref=»startercap.start_coordinator»/>
  </achievegoalref>
  <!-- Запуск Chat_bot агента для выполнения ответа
  пользователю-->
  <achievegoalref name=»start_chat_bot»>
    <concrete ref=»startercap.start_chat_bot»/>
  </achievegoalref>
</goals>

```

Рис. 6. Цели агента

Установление связи пользователей модуля агента «start_chat» с терминалами тьюторов осуществляется агентом координатором «coordinator» по запросу пользователя к агенту (рис. 6). Выполнив первоначальные планы, агент переходит в состояние ожидания дальнейших действий пользователя.

По команде пользователя «передать сообщение тьютору» агент пытается передать его, проверяя состояние тьютора. Если тьютор находится в режиме online, агент реагирует на сообщение (рис. 7), последовательно направляя алерт-уведомление на установление связи, используя план «connection» (рис. 8).

```

<events>
  <!-- Обработка запроса на общение с тьютором -->
  <messageevent name=»connect_tutor» direction=»receive»
    type=»fipa»>
    <parameter name=»performative» class=»String»
      direction=»fixed»>
      <value>SFipa.REQUEST</value>
    </parameter>
  </messageevent>
</events>

```

Рис. 7. Сообщения, на которые реагирует агент

При подтверждении тьютора на алерт-сообщение, агент устанавливает соединение для общения, используя окно чата.

Если тьютор отказывается принять сообщение или не реагирует, на него находясь в статусе online/offline, агент, используя план «connection», ожидая некоторый промежуток времени, затем выполняет вызов модуля агента «start_chat_bot» для возможного начала диалога с машиной. Модуль «start_chat_bot» работает по заданной онтологической

модели, используя базу знаний, содержащую ответы, вопросы и ключи, определяющие принадлежность ответа к вопросу.

```

<plans>
  <plan name=»starting»>
    <!-- Открытие диалогового окна чата и заполнение контакт
    листа тьюторов -->
    <body>new StartingChat()</body>
  </plan>
  <plan name=»review»>
    <!-- Проверка состояния тьюторов -->
    <body> new CheckSource()</body>
  </plan>
  <plan name=»connection» class=»ConnectionTutor»>
    <!-- Установить связь с выбранным тьютором -->
    <body>new ConnectionTutor()</body>
    <waitqueue> messageevent ref=»connect_tutor»/>
  </waitqueue>
  </plan>
</plans>

```

Рис. 8. Планы агента

Выводы

Разработка и внедрение агента интерактивного взаимодействия участников мультиагентной системы дистанционного обучения обеспечивает динамику общения и предоставляет требуемую информацию студенту при отсутствии тьютора. Предложенный в работе программный агент выражен как BDI– модель, для него созданы факты, цели, планы, сообщения, также разработана схема взаимодействия модулей агента и его связь с другими агентами. Реализовано программное приложение с возможностью обработки запросов пользователя: запуск агента, построение GUI, получение информации от агента–координатора для генерирования информации об участниках системы, интерактивная коммутация пользователей, запуск агента чат-бота для оказания поддержки обучаемому, формирования логической связи между полученным запросом и действиями агента.

Модуль агента чат-бот представляет пользователю ответ, заменяя тьютора в его отсутствие, на основании полученного вопроса к системе. Обработка вопроса происходит по разработанной онтологической модели, ограниченной пределами тематики системы ДО.

Таким образом, использование агентных технологий для решения поставленной задачи решает проблему коммутации, используя модуль чата в системе дистанционного образования и взаимодействует с остальными агентами системы, дополняя их функциональные возможности.

Конечным результатом реализации [7,8] агента является подключение к мультиагентной системе дистанционного образования для совместного взаимодействия с другими агентами в соответствии с заданной BDI–моделью.

Список литературы: 1. Copernic Products <http://www.copernic.com>. 2. Mulberry Technologies <http://www.mulberrytech.com>. 3. The World Wide Web Consortium (W3C) <http://www.w3.org>. 4. John Davies, Rudi Studer Semantic Web technologies: trends and research in ontology-based systems, 2006. 5. Alexander Pokahr, Lars Braubach, Andrzej Walczak, Jadex User Guide. 6. Hamburg: Systems Group University of Hamburg, 2005. 7. Программная среда IntelliJ IDEA 6.0 для разработки Java приложений. 8. <http://jadex.sourceforge.net/jadex>.

Поступила в редколлегию 29.09.2008

УДК 519.62

Функціонування програмного агента інтерактивної взаємодії учасників мультиагентної системи дистанційного навчання / С.А. Марьин, О.П. Мишин, В.Б. Репка // Біоніка інтелекту: наук.-техн. журнал — 2008. — № 2 (69). — С. 95-100.

У статті викладено ідею розробки програмного агента інтерактивної взаємодії користувачів в мультиагентній системі дистанційного навчання. Також представлено схему взаємодії агентів з користувачами, критерій функціонування агента, онтологічну модель розмови агента з користувачами за відсутності тьютора.

Лл. 8. Бібліогр.: 8 найм.

UDK 519.62

Functioning program agent interactive communication of users in distant learning system / S. Marin, A. Mishin, V. Repka // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 95-100.

In this article the idea of program agent development for interactive communication of users in multi-agents distant learning system is given. Also the scheme of agents' interaction with users, and criteria of agent functioning, and ontological model of conversation between the agent and the tutor (when tutor is absent) are presented.

Fig.8. Ref.: 8 names.

УДК 378.147



О ПСИХОЛОГИЧЕСКОЙ КЛАССИФИКАЦИИ МОДЕЛЕЙ ОБУЧАЕМЫХ

М.Г. Коляда

Донецкий национальный технический университет, г. Донецк, Украина, kolyada_mihail@mail.ru

Проведена классификация моделей обучаемых, которые могут быть использованы в автоматизированных обучающих системах с точки зрения психологических составляющих. Рассмотрены формализованные психологические модели, элементы которых могут быть реализованы в интеллектуальных компьютерных системах.

МОДЕЛЬ ОБУЧАЕМОГО, ФАКТОРНАЯ МОДЕЛЬ ИНТЕЛЛЕКТА, МОДЕЛЬ ПАМЯТИ, СВЯЗАННЫЙ ИНТЕЛЛЕКТ, СКОРОСТЬ ИНТЕЛЛЕКТА, УРОВЕНЬ ИНТЕЛЛЕКТА

Введение

Современный этап использования моделей обучения в автоматизированных обучающих системах характеризуется тем, что какие бы мы не использовали подходы в «технизации» этого процесса, несомненным остается тот факт, что без понимания психологической сущности любого процесса обучения и психологических механизмов управления познавательной деятельностью обучаемого мы не сможем далеко продвинуться в решении этих задач.

В связи с усиливающимся интересом к проектированию компьютерных обучающих систем возникает много неясностей, неувязок и нестыковок в подходах трактовки видов моделей и лежащих в их основе идей. Попытка классифицировать виды моделей обучаемых в интеллектуальных обучающих системах нами уже сделана [1; 2], но мы коснулись лишь общих аспектов моделирования обучаемого и дидактической составляющей. Сейчас появились новые направления в моделировании обучаемых: были созданы прогрессивные педагогические теории, разработаны инновационные дидактические технологии процесса обучения, которые опираются на психологические теории научения. Поэтому назрела большая необходимость в осмыслении современных видов и классов моделей обучаемых с психологической точки зрения, используемых в автоматизированных обучающих системах.

Данная статья в этом плане очень своевременна и актуальна, особенно, если рассматривать психологический аспект воплощения теоретических идей в практические наработки, выражающиеся в проектировании и использовании интеллектуальных обучающих программных системных комплексов.

Целью статьи является выделение классификационных признаков моделей обучаемых с точки зрения психологических составляющих.

Опираясь на цели, мы определяем узкий круг задач, которые в основном сводятся к пониманию психологической сущности идей успешного обучения и составляющих ее элементов.

В настоящее время в педагогической психологии выделяют четыре основные *психологические теории научения* и соответственно им четыре *дидактические модели процесса обучения*:

1. Ассоциативная теория научения. *Модель обучения как управления процессом накопления и переработки чувственного опыта.*

2. Деятельностная (условно-рефлекторная) теория научения. *Модель процесса обучения как стимуляции познавательной и исследовательской активности обучаемых через направление и организацию их практической деятельности.*

3. Знаковая теория научения. *Модель процесса обучения как формирования у обучаемых обобщенных понятийных систем и приемов умственной деятельности.*

4. Операциональная теория научения. *Модель процесса обучения как управления психической деятельностью через организацию предметно-речевой деятельности* (Л.Б. Ительсон).

Кроме названных теорий различные научные школы еще в советское время выделили свои психологические теории научения, в том числе: *теория поэтапного формирования новых знаний и действий* (П.Я. Гальперин), *теория установки* (Д.Н. Узнадзе), *теория развивающего обучения* (Д.Б. Эльконина, Б.Г. Ананьев, Г.С. Костюк, Л.В. Занков, М.А. Менчинская, А.А. Смирнов), *теория формирования теоретического мышления* (В.В. Давыдов), *теория проблемного обучения* (М.И. Махмутов, Т.В. Кудрявцев, Ц.Я. Лернер, А.М. Матюшкин).

Ни одну из этих теорий не следует абсолютизировать. Каждая из них имеет свою преимущественную область применимости, а в определенных условиях теоретические основы разных теорий дополняют друг друга.

Рассматривая модели обучаемых с психолого-педагогической точки зрения необходимо отметить, что построение их, как правило, отождествляется с построением дидактических технологий обучения.

Под технологией надо понимать совокупность и последовательность методов и процессов пре-

образования исходных материалов, позволяющих получить продукцию с заданными параметрами [3]. Под педагогическими технологиями понимают комплексный, интегративный процесс, который охватывает обучаемых, идеи, средства и способы организации деятельности для анализа проблем и планирования, обеспечения, оценивания и управления решением проблем, которые касаются всех аспектов усвоения знаний [4].

Для технологий обучения характерны следующие особенности: неопределенность результата, отсутствие методов и средств, дающих сразу после одного цикла взаимодействия (обучения) необходимый 100% результат; периодическое проведение контроля по совершенствуемому параметру; дополнительная работа с обучаемыми, которые не усвоили материал (повторение цикла взаимодействия); вторичное проведение контроля после дополнительной работы; в случае устойчивого непонимания нового материала проводится диагностика причин непонимания или отставания.

В самом широком смысле под *моделью обучаемого* понимают знания об обучаемом, используемые для организации процесса обучения. Это множество точно представленных фактов об обучаемом, которые описывают различные стороны его состояния: знания, личностные характеристики, профессиональные и психологические качества и другое [5].

По сути дела, модель обучаемого — это знания преподавателя (обучающей системы) об обучаемом, используемые для организации процесса обучения. Это общее определение, по мнению В.А. Петрушина, допускает две интерпретации: 1) модель обучаемого является моделью текущего состояния знаний и умений индивидуального обучаемого; 2) она представляет собой «идеальную» модель знаний об обучаемом, включающую знания о предметной области, типичных ошибках и когнитивных механизмах [6, с 85].

Дж. Селф под моделью обучаемого понимает множество точно представленных фактов об обучаемом, которые могут, например, описывать предпочтения, представления, навыки или действия [7]. Л.А. Растрин под моделью обучаемого понимает представление того процесса, который происходит в обучаемом в результате психологического восприятия им той или иной обучающей информации [8]. Эта модель учитывает не только текущее состояние знаний и умений обучаемого, но и такие важнейшие при обучении психические процессы как запоминание и забывание.

Обобщая представленные определения в контексте нашей темы исследования, приведем рабочее определение: *модель обучаемого* — это психолого-дидактические механизмы и знания об обучаемом как объекте педагогического воздей-

ствия, необходимые для организации процесса обучения с целью получения наиболее оптимального и эффективного результата обучения.

Представим известные и малоизвестные психологические составляющие моделей обучаемых.

Психологические составляющие моделей обучаемых

Исследованиями проблем обучения первыми занялись психологи через изучение психофизиологических особенностей обучаемых. В психологии обучение понимается так же как в педагогике — усвоение обучаемым определенной системы знаний и выработка им необходимых умений и навыков. При этом, с точки зрения психологии, важную роль в обучении играет память, то есть такие важнейшие психические процессы как запоминание и забывание, характеризующие усвоение знаний.

Одной из первых формализованных моделей с точки зрения психологии была *математическая модель памяти*, Г. Эббингауза [9]. Он стал первым, кто показал, что процессы запоминания и забывания имеют нелинейный характер. Осмысленный материал запоминается несколько быстрее, чем бессмысленный; при запоминании информации, несущей конкретную смысловую нагрузку, действуют определенные эффекты и закономерности. Так, например, при запоминании задачи действует особенный эффект. Если решение задачи не доведено до конца, она запоминается лучше и остается в памяти дольше, решенная же задача запоминается значительно хуже. Кроме этого, действует эффект края. Та информация, которая находится ближе к краю списка, то есть либо в конце, либо в начале, запоминается лучше, а то, что расположено в середине, быстрее исчезает из памяти. Герман Эббингауз установил, что срок, на который запоминается информация, зависит от установки, действующей во время запоминания, то есть при запоминании нужно рассчитывать на то, что данная информация обязательно потребуется в дальнейшем. Так же при запоминании большого объема различной информации существует эффект «затапывания следов». Чем больше сходных по содержанию и форме данных человек пытается запомнить, тем хуже ему это удается.

От исследований памяти Эббингауз вполне естественно перешел к изучению различных проблем педагогики. Память молодого человека очень активная и емкая, он при желании может запомнить гораздо больше информации, чем взрослый, скорость запоминания у него также гораздо выше. Однако он при запоминании совершает гораздо больше ошибок, что становится причиной быстрой потери информации.

Эббингауз первый указал на то, что информация, которую обучаемый должен получить от

преподавателя или обучающей системы, должна восприниматься им активно. Если в процессе получения знаний он старается обдумывать их, формулировать вопросы и комментарии, то процесс запоминания будет более эффективным. Информация, которую получают обучаемые, должна быть эмоционально насыщенной, тогда она будет легче запоминаться, если же эмоциональная окраска отсутствует, то ее нужно «придумать» и это непростая задача создателей автоматизированных обучающих систем. При запоминании большого объема материала нужно особенно тщательно заучивать информацию, содержащуюся в середине учебного материала, поскольку именно она обычно быстрее всего выпадает из памяти. Для того чтобы избежать действия эффекта «затапывания следов», нужно постоянно менять специфику запоминаемого материала, например, чередовать материал естественных предметов с гуманитарными, а также менять форму подачи материала.

В психологии существуют также *факторные модели интеллекта обучаемого*. К ним относят модель Ч. Спирмена, Л. Терстоуна, Дж. Гилфорда, Р. Б. Кеттелла, Ф. Вернона и Д. Векслера, Г. Ю. Айзенка и Л. Т. Ямпольского. Как отмечает В.Н. Дружинин, «критерием интеллектуального поведения обучаемого является не преобразование среды, а открытие возможностей этой среды для адаптивных действий его в ней» [10, с. 45].

Определение интеллекта как некоторой способности, обуславливающей успешность адаптации обучаемого к новым условиям, является наиболее общим и должно обязательно учитываться при построении компьютерной обучающей системы. В первую очередь интеллект рассматривается, как способность обучаемого решать определенным образом сконструированные тестовые задания. Суть любого измерительного подхода заключается в процедуре и содержании тестовых заданий [10].

В модели Ч. Спирмена [11] успех любой интеллектуальной деятельности в процессе обучения определяют через некий общий фактор (общая способность к обучению) и фактор, специфический для данной деятельности (специальная способность обучаемого).

В отличие от Спирмена, Л. Терстоун [12] отрицал наличие единственного фактора, обеспечивающего продуктивность обучающих действий. По его предположениям, каждый интеллектуальный акт является результатом взаимодействия множества отдельных факторов. Наиболее часто в исследованиях Терстоуна воспроизводились семь следующих факторов: *словесное понимание* — тестируется заданиями на понимание текста; *словесные аналогии* — понятийное мышление, интерпретация пословиц и так далее; *речевая беглость* — измеряется тестами на нахождение рифмы, называние слов,

принадлежащих к определенной категории; *числовой фактор* — тестируется заданиями на скорость и точность арифметических вычислений; *пространственный фактор* — делится на два подфактора: первый определяет успешность и скорость восприятия пространственных отношений (узнавание плоских геометрических фигур), а второй — связан с мысленным манипулированием зрительными представлениями в трехмерном пространстве; *ассоциативная память* — измеряется тестами на механическое запоминание словесных ассоциативных пар; *скорость восприятия* — определяется по быстрому и точному восприятию деталей, сходств и различий в изображениях. Разделяют вербальный и образный подфакторы. *Индуктивный фактор* тестируется заданиями на нахождение правила и на завершение последовательности.

Однако, как показали дальнейшие исследования, факторы Терстоуна оказались зависимыми, то есть коррелировали между собой, что подтверждает предположение Ч. Спирмена о существовании единого фактора.

Дж. Гилфорд в результате систематизации своих исследований предложил модель «структуры интеллекта» [13]. Эта модель является трехмерной по схеме: содержание (задания) — умственный процесс (операции) — результат.

Содержанием задачи являются особенности материала или информации (изображение, символы (числа и буквы), семантика (слова), поведение). *Операция*, по Гилфорду, это психический процесс. Им может быть познание, память, дивергентное и конвергентное мышление, оценивание. *Результатами* является та форма, в которой испытуемый дает ответ. Ими может быть: элемент, классы, отношения, системы, типы преобразований и выводы.

Факторы в этой модели независимы, и каждый из них образуется сочетанием категорий трех измерений интеллекта. Названия факторов условны. По мнению Гилфорда, в настоящее время идентифицировано более 100 факторов, а всего их — 120.

Р. Б. Кеттеллом в результате анализа результатов большого количества тестов были предложены два фактора: фактор «*связанного интеллекта*» и фактор «*текущего интеллекта*» [14]. «Связанный интеллект» определяет меру овладения культурой общества, к которому принадлежит обучаемый. «Текущий интеллект» определяет возможности нервной системы быстро и точно перерабатывать информацию.

Фактор «связанного интеллекта» диагностируется тестами на запас слов, чтение, учет социальных нормативов, а фактор «текущего интеллекта» — тестами на выявление закономерности в ряду фигур и цифр, объем оперативной памяти, пространственные операции. Эти факторы, по Кет-

теллу, являются базовыми. Кроме них, он выделил три дополнительные парциальные фактора: «визуализация» — как способность манипулировать образами, «память» — как способность сохранять и воспроизводить информацию и «скорость» — как способность поддерживать высокий темп реагирования. Уровень развития парциальных факторов определяется опытом взаимодействия обучаемого с окружающей его средой.

В дальнейшем было показано, что фактор «связанного» и фактор «текущего» интеллекта коррелируют и в ходе исследования невозможно отделить «текущий интеллект» от «связанного интеллекта», так как они сливаются в единый общий фактор (как и общий фактор по Спирмену).

Факторы в модели *Ф. Вернона* располагаются на четырех уровнях [15]. Первый уровень занимает *общий фактор* (как и по Спирмену), на втором — расположены два основных фактора: *вербально-образовательный и практико-технический*. На третьем находятся *специальные способности* (техническое мышление, арифметическая способность и другие). На последнем расположены более частные субфакторы.

Модель *Д. Векслера* [16] отличается наличием только трех уровней. На первом находится *общий интеллект* (по Спирмену), на втором — *групповые* факторы (невербальный и вербальный интеллекты) и на третьем — *специфические факторы*, определяющиеся успешностью выполнения отдельных субтестов.

Г. Ю. Айзенк является представителем одномерного подхода к интеллекту [17]. По Айзенку, можно выделить три типа концепций интеллекта: *биологическую, психометрическую и социальную*. Эти концепции соответствуют трем структурным уровням интеллекта.

«Биологический интеллект» связан со структурами головного мозга, обеспечивающими осмысленное поведение. Методами его измерения могут являться: *электроэнцефалография (ЭЭГ), измерение усредненных вызванных потенциалов (УВП), кожно-гальваническая реакция (КГР), измерение времени реакции (ВР)*. «Психометрический интеллект» определяется *успешностью выполнения тестов IQ*. Эта успешность зависит как от биологического интеллекта, так и от культурных факторов. «Социальный интеллект» определяется успешностью адаптации в социуме. По мнению Айзенка, биологический уровень является фундаментальным для остальных.

Решая проблему соотношения скорости переработки информации и когнитивной дифференцированности, Айзенк объединяет *фактор сложности* (зависящий от успешности выполнения сложных задач за ограниченное время) и *фактор скорости* (зависящий от скорости выполнения простых за-

даний), так как существует корреляция результатов простых тестов, выполненных с ограничением времени и таких же тестов без ограничения времени, близкая к единице. Делая выводы из результатов своих исследований, Айзенк предполагает наличие трех основных параметров характеризующих IQ. Это — *скорость, настойчивость* (зависит от числа попыток решения задачи) и *число ошибок*.

Главным параметром, характеризующим уровень интеллекта по Айзенку, является скорость переработки информации. Айзенк предлагает использовать в качестве ее показателя *время реакции выбора из множества альтернатив*. Как видно, Айзенку не удастся выйти из измерений скорости-трудности. Таким образом, *уровень интеллекта* характеризуется не только скоростью работы мыслительных процессов, но и способностью человека работать с множеством альтернатив. Фактор, обеспечивающий переработку сложной информации и детерминирующий индивидуальную продуктивность, В.Н. Дружинин называет «индивидуальным когнитивным ресурсом» [10].

Л. Т. Ямпольским была сделана попытка разрешить дилемму «сложности» и «скорости» [18]. Так в результате обработки результатов созданного им теста на логико-комбинаторное мышление было выявлено три фактора. Эти факторы характеризуют продуктивность выполнения испытуемым данного теста. Первый фактор — *фактор времени решения*; второй — *правильности решения простых задач*; третий — *фактор правильности решения сложных задач*. Факторы не являются ортогональными, а связаны друг с другом. Корреляция первого фактора со вторым составляет 0,202, второго с третьим — 0,832, третьего с первым — 0,389.

Подводя итог по факторным моделям интеллекта, можно сказать, что выявлено два главных фактора, от которых зависит успешность выполнения задач обучаемыми. Это: фактор *«скоростного интеллекта»* и фактор *«когнитивной сложности»* (предельных когнитивных возможностей).

Именно используя эти психологические модели, были получены различные коэффициенты и зависимости, на основе которых были созданы первые модели обучения. Позднее данные модели были переведены в вероятностную форму [19]. Данные модели используются разработчиками обучающих систем на последующих этапах развития моделей обучения.

Выводы

Рассмотренные виды и классификация моделей обучаемых в автоматизированных обучающих системах представлены разноплановым подходом, имеющим три основных вектора объединения:

— *учет психологических особенностей и резервов личности,*

— учет дидактических основ эффективного обучения и

— технологическое воплощение через интеллектуализацию и автоматизацию всех составляющих компонентов в единое целое.

Задача интеграции и взаимодействия рассмотренных моделей обучаемых в консолидирующем единстве достаточно сложна, требует новых научных исследований как в области концептуальных и методологических подходов, так и в сфере конкретных практических воплощений.

Список литературы: 1. Коляда М.Г. Виды моделей обучаемых в автоматизированных обучающих системах // Искусственный интеллект. — 2008. — № 3. — С. 142–147. 2. Коляда М.Г. Дидактическая составляющая моделей обучаемых в автоматизированных обучающих системах // Искусственный интеллект. — 2008. — № 4. — С. 267–274. 3. Советский энциклопедический словарь. М., 1981. С. 1338. 4. Глоссарий терминов по технологиям обучения. Париж, ЮНЕСКО, 1986. — С. 43. 5. Атанов Г.А. Моделирование учебной предметной области, или предметная модель обучаемого // Educational Technology & Society. — 2000. — № 3 (3) — С. 111–124. — ISSN 1436-4522. 6. Петрушин В.А. Экспертно-обучающие системы. — К.: Наукова Думка, 1992. — 196 с. 7. Self J., Paiva A. Learner Model Resonance Maintenance System // Lancaster University Press, 3, 1993. — Р. 23–31. 8. Расстригин Л. А., Эренштейн М.Х. Адаптивное обучение с моделью обучаемого. — Рига: Зинатне, 1988. — 160 с. 9. Эббингауз Герман. Режим доступа: http://mirslivare.com/content_psy/JEBBINGAUZ-GERMAN-3097.html. — Название с экрана. 10. Дружинин В.Н. Психология общих способностей. — М.: Латерна, Вита, 1995. 11. Spearman Ch. General Intelligence, Objectively Determined and Measured // American Journal of Psychology. № 15. 1904. Р. 201–293. 12. Терстоун Л. (Thurstone) Аттитууд может быть измерен. — 1929. — Р. 45. 13. Гилфорд Дж. Три стороны интеллекта // Психология мышления. — М.: Прогресс, 1965. 14. Cattell R.B., Eber H.W., Tatsuoka M.M. Handbook for the sixteen personality factors questionnaire. Champaign, IL: Inst. Personality and ability testing, 1970. (Кемтелл Р.Б. Умственные тесты и измерения // Mind. — 1890.) 15. Allport G. & Vernon P. Изучение ценностей // A Study of Values, 1931. 16. Wechsler D. Manual for the Wechsler adult Intelligence Scale. N. Y. — 1955. 17. Айзенк Г. Интеллект: новый взгляд // Вопросы психологии. — 1995. — № 1. — С. 111–129. 18. Ямпольский Л.Т. Измерение продуктивности интеллектуальной деятельности // Вопросы психологии. — 1984. — № 5. — С. 142–147. 19. Суходольский Т.В. Введение в математико-психологическую теорию деятельности. Спб.: изд. СПб ун-та, 1998. — 220 с.

Поступила в редколлегию 30.09.2008

УДК 378.147

Про психологічну класифікацію моделей тих, яких навчають / М.Г. Коляда // Біоніка інтелекту: наук.-техн. журнал — 2008. — № 2 (69). — С. 101–105.

В статті розглядається класифікація моделей тих, яких навчають в автоматизованих навчаючих системах, з точки зору психологічних складових. Представлено психологічні формалізовані моделі тих, яких навчають, які можна використовувати в інтелектуальних навчаючих системах.

Бібліогр.: 19 найм.

UDK 378.147

Kinds and classification of student models in the automated learning systems / M.G. Kolyada // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 101–105.

In the articles the different kinds and classifications of models student in the automated learning systems are esteemed. The general approaches to classification of models, psychological, pedagogical aspects, and also problems of automation and technical intrusion of construction of student models are submitted.

Ref.: 19 items.

УДК 004.934



СОВРЕМЕННЫЕ МЕТОДЫ КОДИРОВАНИЯ РЕЧЕВОГО СИГНАЛА

М.Ф. Бондаренко¹, А.В. Работягов², С.В. Щепковский³

¹ХНУРЭ, г. Харьков, Украина,

²ХНУРЭ, г. Харьков, Украина, beloswet@kture.kharkov.ua

³ХНУРЭ, г. Харьков, Украина, svserg@kture.kharkov.ua

В статье проведен обзорный анализ современного состояния исследований в области кодирования речевого сигнала. Рассмотрены основные применяемые и перспективные методы кодирования.

КОДИРОВАНИЕ ФОРМЫ ВОЛНЫ, ПАРАМЕТРИЧЕСКОЕ КОДИРОВАНИЕ, ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЕ

Введение

В настоящее время в связи с активным развитием техники связи, особенно мобильной, решение задачи кодирования речевой информации имеет большое значение.

Приоритетными областями применения кодирования речевой информации являются:

1) уплотнение канала связи для обеспечения большей пропускной способности;

2) обеспечение конфиденциальности (защиты) передаваемой по техническому каналу связи речевой информации.

Первая область применения обусловлена увеличением объемов передаваемой речевой информации по каналу связи: при возрастании количества передаваемой информации (например, увеличение объемов телефонных переговоров внутри фирмы, увеличение количества входящих звонков) возникает необходимость обеспечения большой про-

пускной способности используемых каналов связи для поддержания информационного обмена.

Вторая область применения обусловлена тем, что одновременно с развитием радиоэлектронных средств связи развиваются и средства электронного шпионажа. Это обстоятельство ставит задачу защиты речевой информации в ряд наиболее актуальных.

Каналы утечки информации достаточно многочисленны. Они могут быть как естественными, обусловленными, например, техническим несовершенством канала связи, так и искусственными, то есть созданными с помощью технических средств. В табл. 1 рассматриваются некоторые характеристики и особенности применения основных средств радиоэлектронного шпионажа.

Отметим, что речевая информация принципиально отличается от текстовой информации. При

Таблица 1

Контролируемое устройство	Приемник информации	Место установки	Стоимость аппаратуры	Вероятность применения	Качество перехвата
Телефон	Индуктивный или контактный датчик	Телефонная линия от аппарата до АТС	Низкая	Высокая	Хорошее
Телефон	Контактный датчик	Телефонная линия от аппарата до АТС	Низкая	Низкая	Плохое
Телефон, любое устройство с питанием от сети	Радиомикрофон с передачей по телефонной сети или по сети 220В	Тел. аппарат, тел. розетка, любое устройство с питанием	Низкая	Высокая	Хорошее
Любое место в помещении	Автономные радиомикрофоны, направленные микрофоны, в т.ч. лазерные	Любое место в помещении	Высокая	Средняя	Хорошее
Радиотелефон, радиостанция	Панорамный радиоприемник	Прием с эфира	Средняя	Высокая	Хорошее
Сотовый телефон	Устройство прослушивания сотовой сети	Прием с эфира	Высокая	Высокая	Хорошее
Монитор ПК	Широкополосная антенна	Прием с эфира	Очень высокая	Низкая	Посредственное
Монитор ПК	Широкополосный контактный датчик	Питающая электросеть	Очень высокая	Низкая	Посредственное
Магистраль компьютерной сети	Индуктивный или контактный датчик	Кабель магистральной	Высокая	Высокая	Хорошее

шифровании текста используют определенный ограниченный набор символов, поэтому при работе с текстом можно использовать такие шифры, как шифры перестановки, шифры замены, шифры взбивания и так далее. Речь же нельзя (во всяком случае, на современном уровне развития технологии распознавания речи) представить набором каких-либо символов. Поэтому применяются иные принципы кодирования.

При кодировании речевой информации ставят задачу максимального сжатия речевого сигнала [1]. Существует предел сжатия речевой информации. Теорема Грея-Бергера определяет предел сжатия последовательностей по минимально достижимой среднеквадратической ошибке. В соответствии с данной теоремой этот предел составляет 100 бит/с. Эффективность (качество) кодирования оценивается по субъективному восприятию речи.

Качество сигнала измеряется часто по пятибалльной шкале MOS (*mean opinion score* — средняя субъективная оценка) (см. табл. 2). Оценка по шкале MOS определяется путем обработки усредненных оценок, даваемых группами слушателей нескольким речевым сигналам, воспроизводимым различными громкоговорителями.

Каждый слушатель дает оценку для каждого сигнала.

Таблица 2

1	2	3	4	5
Плохо	Слабо	Разборчиво	Хорошо	Отлично

Основные усилия в разработке методов и систем цифровой связи осуществляют высокоразвитые страны — США, Япония, Франция, Германия и Англия. Конечные результаты исследований формулируются в виде национальных стандартов и рекомендаций Международного института телекоммуникаций ITU (International Telecommunication Union).

1. Методы кодирования речевого сигнала

Современные методы кодирования речевого сигнала можно классифицировать по принципам кодирования, которые положены в основу метода. Различают следующие принципы кодирования речевого сигнала:

- принцип кодирования формы волны речевого сигнала;
- принцип кодирования параметров речевого тракта человека и источника возбуждения;
- принцип кодирования символьной информации (фонем);
- принцип кодирования лингвистической информации (слов, фраз и тому подобное).

Данные принципы различаются в зависимости от используемых основных свойств речевого сигнала, образования и восприятия речи:

1) основные свойства речевого сигнала: амплитуда, частота (время), площадь (как производная амплитуды и времени);

- 2) свойства образования речи:
 - изменения амплитуды,
 - деление речи на звуки, паузы и шумы,
 - особенности языковой и фонетической структуры,
 - кратковременная корреляция (особенности формантной структуры),
 - особенности структуры тона звука (звонкие звуки),
 - особенности структуры шума (глухие звуки).
- 3) свойства восприятия речи:
- 4) локальный спектральный динамический диапазон речевого сигнала,
- 5) слуховое маскирование.

На современном этапе используются, в основном, методы, которые базируются на первых двух принципах кодирования.

Принцип кодирования формы волны речевого сигнала. Первым шагом кодирования является измерение значения амплитуды сигнала. Для этого 12–14-ти разрядный динамический диапазон амплитуды разбивают на 8 логарифмических поддиапазонов, в каждом из которых значение амплитуды кодируют 5 разрядами и таким образом достигают сокращения информации до 64000 бит/с (кодирование по μ - и A- законам в соответствии со стандартом ITU G.711) [2]. Следующим шагом является адаптивная дифференциальная импульсно-кодовая модуляция (*adaptive differential pulse code modulation* — ADPCM), (например, в соответствии со стандартами G.721 или G.726), с помощью которой осуществляют кодирование (аппроксимацию) степени приращения амплитуды сигнала во времени. Таким путем удастся достичь степени сжатия речевого сигнала порядка 32000–16000 бит/с, причем приемлемое (коммерческое) качество речи (по критерию отношения: полезный сигнал/шум) обеспечивается до 24000 бит/с. При более низких скоростях кодирования сохраняется разборчивость речи, но характерны сильные нелинейные и частотные искажения сигнала и ухудшение отношения сигнал/шум. Дальнейшее уменьшение информационной емкости сигнала с помощью данного подхода считается неэффективным.

Принципы параметрического кодирования. Низкоскоростное кодирование складывается из двух основных процессов:

- параметрическое представление речевого сигнала минимальным набором параметров, характеризующих источник возбуждения и акустический артикуляторный фильтр;
- дискретизация речевых параметров для их передачи по каналу связи при использовании минимальной емкости канала.

Для параметрического описания речи обычно используется подход, основанный на вычислении параметров, описывающих передаточную функцию речевого тракта человека и функцию возбуждения. Такими параметрами, например, являются: коэффициенты линейного предсказания (модель

авторегрессии) и связанные с ними коэффициенты отражения или отношение площадей поперечного сечения смежных акустических резонаторов в соответствии с моделью речевого тракта человека, представленной системой акустических резонаторов. В последнее время наибольшее распространение получил метод, позволяющий вычислять непосредственно полюса передаточной функции речевого тракта в частотной области, упорядоченные по возрастанию частоты (*liner spectral frequency* – LSF). Обычно для кодирования речи используются 8-10 параметров (один из вышеперечисленных наборов), вычисляемых на интервалах порядка 5-40 мс. Кроме того, вычисляется параметр, характеризующий изменение амплитуды либо мощности сигнала, период основного тона речи, а также признак типа тон/шум/пауза, характеризующий способ возбуждения речевого сигнала.

В качестве функции возбуждения речевого сигнала используется дельта функция.

Полученный набор параметров, оптимизированный по критерию точности и минимальной разрядности представления, передается в цифровом виде по каналу связи в реальном времени; на приемном конце осуществляется синтез речевого сигнала по перечисленным параметрам. Таким путем удастся снизить информационную емкость речевого сигнала до уровня 16000-1200 бит/с, причем с сохранением разборчивости и индивидуальных особенностей речи говорящего.

На рис. 1 представлены результаты тестирования фирмой AT&T различных систем цифровой связи на основе субъективного восприятия качества речи, передаваемой по цифровым каналам связи [3].

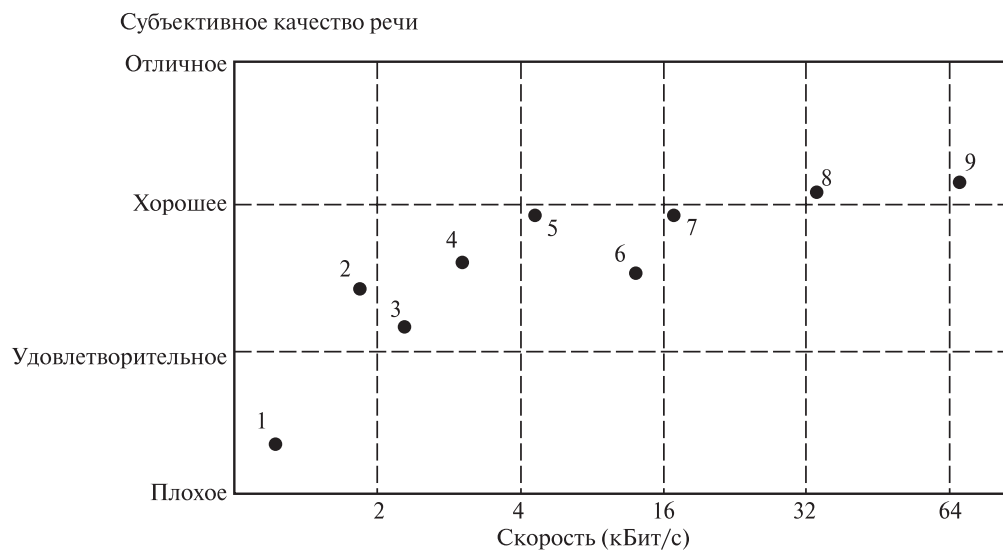


Рис. 1. Субъективное восприятие качества речи

В табл. 3 указаны основные стандарты кодирования.

Как следует из графика, качество некоторых систем параметрического кодирования приближается к качеству ADPCM. Подчеркнем, что такой подход позволил снизить информационную емкость речевого сигнала с исходных 96000 бит/с до 2400 бит/с, то есть в 40 раз. При этом сохраняется не только разборчивость речи, но и индивидуальные особенности речи человека.

Следующим шагом в направлении дальнейшего увеличения сжатия речевого сигнала является создание фонемного вокодера. Как известно, минимальной слогоразличительной (и словоразличительной) единицей речи является фонема. Ожидается, что создание устойчивого метода распознавания фонем позволит снизить скорость передачи данных до 100 бит/с, что соответствует информационной скорости передачи текста. Заметим, что на приемном конце речь будет восстановлена синтезатором речи по фонемному тексту, при этом информация об индивидуальности диктора будет утрачена.

И наконец, последним этапом совершенствования систем кодирования может явиться создание системы на основе автоматического распознавания слов и, возможно, целых фраз. В этом случае по

Таблица 3

№ пп	Стандарт кодирования	Организация	Тип кодирования	Скорость Кбит/с
1	FS1015	Secure Voice Communication USA	LSF	2,4
2	JDC-HR	Japanese Cecular	PSI-CELP	3,67
3	FS1016	Secure Voice Communication USA	CELP	4,8
4	GSM-HR	Eropean Cecular	RPE-LTP	7,8
5	ITU	ITU-T	VCELP	8
6	GSM	Eropean Cecular	RPE-LTP	13
7	G.728	ITU-T	LD-CELP	16
8	G.726	ITU-T	ADPCM	32
9	G.711 m, А закон	ITU-T	PCM	64

каналу связи может быть передан только его код, а на приемном конце будет получен номер слова из некоторого ограниченного словаря и с помощью синтезатора преобразован в речевой сигнал.

Рассмотрим основные методы кодирования речевого сигнала.

1.1. Базовое преобразование речевого сигнала — метод импульсно-кодовой модуляции

При разработке практически любой технической системы кодирования необходимо представить речевой сигнал в цифровом виде. Этот процесс осуществляется на основе метода импульсно-кодовой модуляции (pulse code modulation — PCM).

Процесс базируется на теореме Котельникова (Найквиста), в соответствии с которой цифровой сигнал, полученный выборкой с частотой в два раза выше максимальной частоты аналогового сигнала, с помощью интерполяции обратно преобразуется в аналоговую форму. Человеческая речь воспроизводится с приемлемым качеством в полосе частот 100–4000 Гц, чему соответствует частота выборки 8 кГц (8000 отсчетов в секунду), каждый отсчет преобразуется в 8-битовый цифровой код. Общая скорость цифрового потока РСМ-сигнала равна 8×8000 отсчетов в секунду, то есть 64 кбит/с.

Метод импульсно-кодовой модуляции принят в 1960 г. в качестве международного стандарта кодирования речи для телефонного канала (стандарт ITU G.711), работающего на скорости 64 кбит/с. Этот алгоритм используется при передаче речевой информации в коммерческих телефонных сетях. Оцифровка голосового сигнала включает измерение уровня аналогового сигнала через равные промежутки времени. В соответствии со стандартом G.711 необходимо обеспечить передачу частотных составляющих речевого сигнала в диапазоне от 200 до 3400 Гц. Стандарт обеспечивает неискаженную передачу сигнала в полосе человеческого голоса (до 4 кГц) с отношением сигнал/шум 40 дБ.

Хотя РСМ-сигнал со скоростью 64 кбит/с и гарантирует качество речи аналогового телефонного сигнала, ограниченная общая ширина канала, особенно в спутниковых и радиочастотных системах, вынуждает снижать скорость битовых потоков, отводимых для каждого речевого сигнала. С этой точки зрения весьма эффективны алгоритмы сжатия речи, дополняющие РСМ-кодирование математическими функциями, такими как фильтры,

квантизаторы и предсказатели. Они преобразуют РСМ-сигнал так, чтобы передавать его более эффективным способом, обеспечивая тем самым не менее точное воспроизведение сигнала на приемном конце. Соответствующие устройства называют кодерами (при прямом преобразовании), декодерами (при обратном преобразовании) или кодеками (другое название — вокодер VOice CODer). Основной технической характеристикой кодеков является скорость передачи кодированного речевого сигнала по каналу связи, которая измеряется в бодах (бит/с). Скорость обработки измеряется в миллионах инструкций в секунду (*millions of instructions per second — mips*).

1.2. Метод нелинейного квантования

Метод нелинейного квантования основан на уменьшении числа уровней квантования по амплитуде. На рис. 2 представлен сигнал, который квантуется по времени в 11 точках с использованием 8 уровней квантования по амплитуде сигнала.

Уровни квантования (по амплитуде) кодируются по правилу, представленному в табл. 4:

Номер уровня квантования преобразован в код с 3 битами.

Таблица 4

Кодирование уровней квантования для несжатого сигнала

Уровень квантования	0	1	2	3	4	5	6	7
Код	000	001	010	011	100	101	110	111

Сигнал кодируется следующим образом: 101 111 110 001 010 100 111 100 011 010 101. Общее количество — 33 бита.

Уменьшение числа уровней квантования по амплитуде состоит в том, чтобы соединить каждые два соседних уровня в один. Результатом такого преобразования является только 4 уровня квантования вместо 8. Они могут быть закодированы с использованием только кодов с 2 битами (см. табл. 5).

Таблица 5

Кодирование уровней квантования для сжатого сигнала

Уровень	0	1	2	3
Код	00	01	10	11

Восстановленный сигнал после сжатия имеет те же самые базисные контуры, но искажения в таком

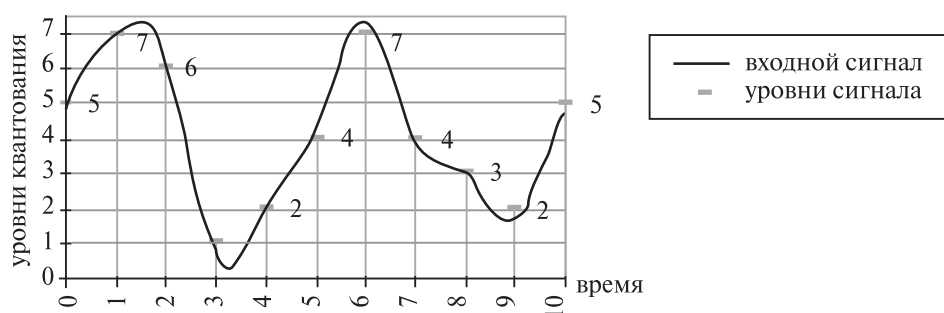


Рис. 2. Исходный несжатый сигнал

сигнале больше из-за грубого приближения, так как шум квантования увеличился. Это — следствие того, что шаг квантования стал двойным по сравнению с первым случаем (несжатым сигналом).

Сигнал закодирован следующим образом (см. рис. 3): 10 11 11 00 01 10 11 10 01 01 10, общее количество 22 бита (11 отсчетов, 2 бита на отсчет), коэффициент сжатия — 1,5:1 (уменьшение 33% в размере). Уровни квантования неравномерно распределены внутри диапазона квантования — они плотнее вблизи уровня нуля и реже вблизи максимального уровня.

Стандарт кодирования, применяемый на практике с коэффициентом сжатия 2:1, получил название A-law.

1.3. Метод кодирования на основе уменьшения числа выборок

Метод основан на уменьшении числа выборок квантования по времени.

Например можно заменить каждые две соседних выборки только на одну, равную их среднему значению. Плотность выборок становится меньше, чем прежде (число выборок уменьшается в два раза). В результате восстановленный по выборкам сигнал может быть существенно искажен за счет потери части гармоник.

При использовании этого метода необходимо отобрать те частоты (гармоники), которые не имеют существенного значения и могут не учитываться, не внося существенной погрешности. Сигнал закодирован следующим образом (см. рис. 4): общее количество битов 18 (6 отсчетов, 3 бита на отсчет), коэффициент сжатия — 1,8:1 (уменьшение 45,5% в размере).

1.4. Метод адаптивно-дифференциальной импульсно-кодовой модуляции

Прямое аналого-цифровое преобразование (PCM-преобразование) является низкоэффектив-

ным, но высококачественным методом кодирования. Кодеки, построенные на базе данного метода, работают на скоростях не ниже 32 кбит/с. При этом полоса входного аналогового сигнала ограничена диапазоном 0,3–3,4 кГц.

Значительные результаты в области эффективного кодирования речи достигнуты на базе общего подхода «кодирования с предсказанием». Большая часть стандартизированных ITU алгоритмов кодирования относится именно к этому направлению.

Метод адаптивной дифференциальной импульсно-кодовой модуляции (ADPCM) принят в качестве стандарта в 1984 г. под названием G.726. Он воспроизводит речь почти с такой же субъективной оценкой качества, как и PCM, используя только 32 кбит/с, и обеспечивает на порядок более высокую помехоустойчивость. Однако он теряет работоспособность при вероятности одиночной ошибки, составляющей около 5×10^{-3} , и передаче пакетов ошибок малой длительности.

Метод основан на том обстоятельстве, что в аналоговом речевом сигнале невозможны резкие скачки интенсивности. Поэтому, если кодировать не саму амплитуду сигнала, а ее изменение по сравнению с предыдущим значением, то можно обойтись меньшим числом разрядов. В ADPCM изменение уровня сигнала кодируется четырехразрядным числом, при этом частота измерения амплитуды сигнала сохраняется неизменной. Таким образом, ADPCM снижает скорость битового потока вдвое путем обработки разности между двумя соседними отсчетами, а не самих отсчетов. Позволяет снизить скорость с 64 кбит/с до 24–48 кбит/с (в зависимости от того, насколько точно передается приращение). Приводит к снижению отношения сигнал/шум и менее точному воспроизведению исходного сигнала.

Среди кодеров формы сигнала первыми появились методы адаптивной дельта-модуляции. Анали-

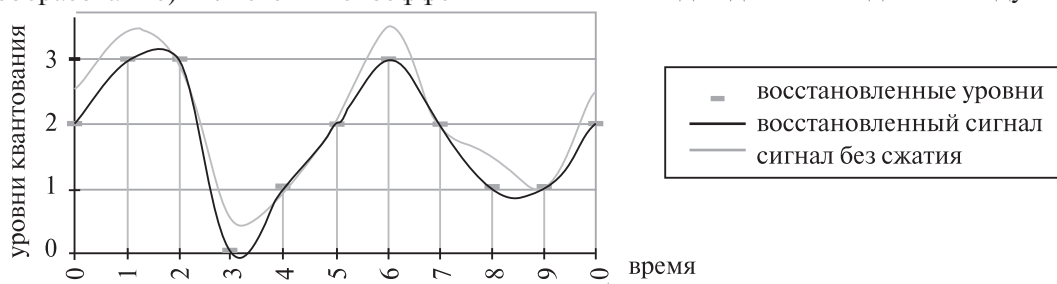


Рис. 3. Сжатый (закодированный) сигнал — уменьшение числа уровней квантования

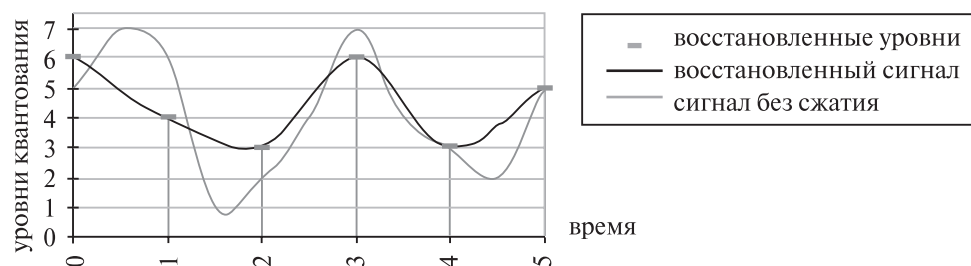


Рис. 4. Сжатый сигнал (уменьшение числа выборок)

тически они являются предельными случаями дифференциальной РСМ, но по ряду причин могут быть выделены в отдельный класс: скорость передачи при дельта-модуляции соответствует частоте дискретизации (одноразрядное квантование); при скоростях 40–30 кбит/с этот метод обеспечивает более высокое качество восстановления, чем РСМ. Дельта-модуляция обладает наилучшими параметрами помехоустойчивости среди всех методов кодирования. Соответствующие системы не теряют работоспособности при возникновении одиночных ошибок и их пакетов малой длительности.

Принципиальное отличие ADPCM от РСМ заключается в использовании адаптивного аналого-цифрового преобразователя (АЦП) и дифференциального кодирования. Адаптивный АЦП отличается от стандартного РСМ-преобразователя тем, что в любой момент времени уровни квантования расположены однородно (а не логарифмически), причем шаг квантования меняется в зависимости от уровня сигнала. Применение адаптивного метода базируется на том, что в речи последовательные уровни сигнала не являются независимыми. Поэтому, преобразуя и передавая лишь разницу между предсказанным и реальным значением, можно заметно снизить нагрузку линии, а также требования к широкополосности канала. Следует иметь в виду, что метод не лишен серьезных недостатков: уровень шумов, связанный с квантованием сигнала, выше, а при резких изменениях уровня сигнала, превышающих диапазон АЦП, возможны серьезные искажения.

РСМ и ADPCM — методы кодирования волновой функции речевого сигнала. Это означает, что они рассматривают входной речевой сигнал как чисто аналоговый. Однако для получения высокого качества сигнала при скоростях ниже 32 кбит/с такое кодирование неэффективно.

1.5. Метод кодирования с линейным предсказанием

При кодировании с линейным предсказанием (*linear predictive coding* — LPC) моделируются различные параметры речи, которые передаются вместо отсчетов или их разности, требующих значительно большей пропускной способности канала. LPC работает с блоками отсчетов, а не с отдельными отсчетами, как РСМ или ADPCM. Для каждого блока алгоритм LPC вычисляет и передает частоту основного тона, его амплитуду, «флаг» речевого или неречевого происхождения сигнала и другие параметры.

При таком подходе к кодированию речи, во-первых, возрастают требования к вычислительным мощностям микропроцессоров, используемых для обработки сигнала, а во-вторых, увеличивается задержка при передаче, поскольку кодирование применяется не к отдельным значениям, а к некоторому их набору, который перед началом преобразования следует накопить в определенном буфере. Подчеркнем, что задержка в передаче речи

при использовании этого метода связана не только с необходимостью обработки цифрового сигнала (эту задержку можно уменьшать, увеличивая мощность процессора), но и непосредственно следует из специфики метода.

Этот метод позволяет достигать очень больших степеней сжатия, которым соответствует полоса пропускания 2,4 или 4,8 кбит/с, однако, при этом качество звука получается низким. Поэтому в коммерческих приложениях он не используется, а применяется, в основном, для ведения служебных переговоров.

Более сложные алгоритмы на базе LPC комбинируют LPC с элементами кодирования звуковой волны. Эти алгоритмы используют замкнутый LPC-кодер (называемый также «анализ через синтез» — *analysis-by-synthesis* — AbS), в котором при передаче сигнала осуществляется оптимизация кода. Ее выполняет алгоритм, который находит наилучшую аппроксимацию каждого речевого сегмента. Закодирав сигнал, процессор пытается восстановить его форму и сличает результат с исходным сигналом, после чего начинает варьировать параметры кодирования, добиваясь наилучшего совпадения. Для использования такого метода требуются большие вычислительные мощности.

Вариант замкнутых LPC-алгоритмов — это метод линейного предсказания с кодовым возбуждением (*code-excited linear prediction* — CELP), метод регулярного импульсного возбуждения (*Regular Pulse Excitation* — RPE), используемый в европейских сотовых системах при скорости 13,2 кбит/с и метод LD-CELP с низкой задержкой (*low delay CELP*). Для кодирования речевых сигналов, например, в стандарте TETRA используется кодер с линейным предсказанием и многоимпульсным возбуждением от кода — CELP со скоростью преобразования 4,8 кбит/с. Данный метод кодирования основан на линейной авторегрессионной модели процесса формирования и восприятия речи и входит в группу так называемых методов анализа через синтез, реализующих современные и эффективные алгоритмы кодирования речевых сигналов. Алгоритмы данного класса занимают промежуточное положение между кодерами формы сигнала, в которых сохраняется форма колебания речевого сигнала в процессе его дискретизации и квантования, и параметрическими вокодерами, основанными на процедурах оценки и кодирования небольшого числа параметров речи, объединяя преимущества каждого из них.

LD-CELP принят ITU в 1992 г. как стандарт кодирования речи G.728 на скорости 16 кбит/с. Алгоритм LD-CELP применяется к последовательности цифр, получаемых в результате аналого-цифрового преобразования речевого сигнала с 16-разрядным разрешением. Пять последовательных цифровых значений кодируются одним 10-битовым блоком, что позволяет достигнуть скорости 16 кбит/с. Для применения этого метода требуются большие вычислительные мощности, в частности, для прямо-

линейной реализации G.728 необходим процессор с быстродействием 44 mips [4].

1.6. Методы кодирования с предсказанием

Широкое распространение для различных приложений получило и множество нестандартных методов кодирования. В частности: варианты адаптивного кодирования с предсказанием (*adaptive predictive coding* – APC), разработанные в лабораториях компании Bell; метод линейного предсказания с векторным возбуждением (*vector-sum-excited linear prediction* – VSELP), предложенный фирмой Motorola в качестве стандарта для цифровых сотовых систем США, работающих на скорости 8 кбит/с; метод линейного предсказания с предикативным кодовым возбуждением (*predictive code-excited linear prediction* – PCELP), созданный DSP Group в 1992 г.

Эти высокоэффективные кодеры обеспечивают отличное качество звука при низких скоростях (2,4–8 кбит/с). Для кодирования погрешности предсказания в них используются так называемые «кодовые книги», состоящие из блоков с конечным числом символов. Эти методы описаны стандартом G.729, в котором вместо оцифрованного сигнала передаются номера выборок из хранящейся в памяти «кодовой книги», где описаны типичные элементы человеческого голоса.

Перечисленные разновидности кодеров различаются способами формирования и хранения этих последовательностей. Чаще всего последовательность хранится в сжатом виде. Дополнительные буквы в названии кодера (LD, V и другие) указывают на способ реализации предсказателя, синтеза квантователя или кодовой книги.

Стандарт G.729A (G.729 Annex A) – это алгоритм сжатия звука преимущественно для передачи речи, генерирующий кадры (фреймы) длительностью 10 мс. Принцип G.729A похож на G.729, но требует меньше вычислений. Это упрощение приводит к незначительному ухудшению качества речи.

Особенности этого стандарта кодирования:

- частота сэмпирования 8 кГц/16 бит (80 сэмплов в 10-ти миллисекундном фрейме);
- фиксированный битрейт (8 кбит/с 10-тимиллисекундные фреймы);
- фиксированный размер фрейма (10 байт для 10-тимиллисекундного фрейма);
- алгоритмическая задержка 15 мс на фрейм, с 5-тимиллисекундной предварительной задержкой;
- сложность алгоритма оценивается в 15 баллов, используя относительную оценочную шкалу, где G.711 равен 1 и G.723.1 равен 25 баллам;
- условно-объективная оценка MOS по методу экспертных оценок, протестированная в идеальных условиях, дает результат 4,04 для G.729A, по сравнению с 4,45 для G.711 u-law;
- условно-объективная оценка MOS по методу экспертных оценок, протестированная в перегруженной сетевой среде, дает результат 3,51 для G.729A, по сравнению с 4,13 для G.711 u-law.

1.7. Метод множественной импульсной многоуровневой квантизации

В марте 1995 г. ITU выбрал метод сжатия речи для своих будущих стандартов в области мультимедиа и видеотелефонов, подключаемых к коммутируемым телефонным сетям. Стандарт сжатия G.723 частично базируется на новом методе сжатия речи – множественной импульсной многоуровневой квантизации (*Multipulse Maximum Likelihood Quantization* – MP-MLQ), разработанном израильской фирмой AudioCodes, создателем передовых речевых и факсимильных технологий и ее корпоративным партнером – американской фирмой DSP Group.

Метод MP-MLQ относится к семейству алгоритмов AbS. Речевой кодер MP-MLQ использует LPC-анализатор 10-го порядка и работает на скоростях 6,4; 7,2 и 8,0 кбит/с. Его структура поддерживает перепрограммирование «на лету» для одной или нескольких скоростей. Масштабируемость алгоритма MP-MLQ позволяет разрабатывать производные реализации для скоростей вплоть до 4,0 кбит/с и более низких коммуникационных задержек (до 20 мс), осуществлять кодирование на нескольких скоростях и с переменной скоростью, выполнять многоканальную обработку (благодаря низкой вычислительной нагрузке – менее 10 mips) и достигать высокого качества.

В отличие от других кодеров с низкими скоростями, метод обеспечивает минимальный уровень искажений при парном кодировании, когда речевой сигнал проходит через два или более последовательных цикла компрессии/декомпрессии. Эта особенность имеет практическое значение в приложениях, в которых сеанс речевой связи в цифровом канале коммутируется через центральную АТС. Тесты, проведенные в фирмах AT&T Labs и France Telecom (CNET), показали, что оценка качества сигнала по шкале MOS после двух кодирований в тандеме методом MP-MLQ составила 3,409, что лучше оценки G.726 ADPCM на 32 кбит/с после четырех кодирований в тандеме (3,102) и почти эквивалентно после двух кодирований в тандеме (3,491).

2. Достигнутые результаты

В настоящее время считается достигнутой скорость передачи кодированного речевого сигнала 2400 бит/с (американский федеральный стандарт FS1015). При этом кроме разборчивости от вокодера требуется обеспечение узнаваемости говорящего, возможность определения его эмоционального состояния и тому подобное. Такие требования заставляют разрабатывать методики испытаний, отличающиеся от обычного тестирования на разборчивость. Разработка таких методик является нетривиальной задачей.

Следует заметить, что для реализации алгоритмов кодирования речевых сигналов в реальном масштабе времени требуются процессоры с производительностью в 15–20 mips.

Исследовательские работы, проводимые в этих направлениях Санкт-Петербургским университетом, позволили в настоящее время создать экспериментальные алгоритмы, программы и устройства, обеспечивающие:

- кодирование/декодирование речевого сигнала на скоростях 3600, 2400 и 1200 бит/с;
- синтеза речи по произвольному тексту.

Системы кодирования и передачи речи могут быть выполнены с использованием как стандартных технических средств, ставших неотъемлемой частью современных персональных компьютеров: средства мультимедиа (звуковые карты) и модемы, так и на базе специальных сигнальных процессоров серии ADSP-21XX.

Экспериментальный макет кодирования/декодирования речи обеспечивает компрессию речевых сигналов в реальном масштабе времени, а затем после передачи по каналу связи восстанавливает речь с сохранением разборчивости и индивидуальных особенностей диктора на компьютерах даже такого уровня, как IBM PC 486 DX4-100 без использования спецвычислителей.

Результаты испытаний экспериментального компьютерного вокодера для русскоязычных дикторов приведены в табл. 6.

Таблица 6

Результаты испытаний разборчивости
по ГОСТ-В 20775-75

Скорость (бит/с)	1200	2400	3600
Разборчивость слогов	89%	91	94%

В настоящее время завершается отладка ПО, обеспечивающего разборчивость слов на скорости 1200 бит/с не хуже 93-94 % при передаче речи в дуплексном режиме по линии связи.

Ведутся исследования по созданию алгоритмов, которые должны обеспечить кодирование речевого сигнала на скорости 600 бит/с с разборчивостью не менее 0,90.

Существующий экспериментальный макет синтезатора речи обеспечивает разборчивость синтезируемого сигнала не хуже 95% при неограниченном объеме словаря.

3. Перспективы кодирования речевой информации

Одним из перспективных направлений уменьшения информационной емкости систем кодирования речи является использование моделей речеобразования, которые не предполагают независимое функционирование источника возбуждения звука при речеобразовании и акустического фильтра, формирующего лингвистическое качество речевого сигнала. Это направление в настоящий момент развивается по линии решения обратной задачи, то есть определения формы артикуляторного тракта (так называемые функции площади) и характеристик источника возбуждения звука по

речевому сигналу. В отличие от классического параметрического описания речевого сигнала при таком подходе этот сигнал будет характеризоваться (опосредованно) с помощью медленно изменяющихся параметров артикуляторного тракта. Кроме того, характеристики взаимодействия источника звука и акустического фильтра в этом случае автоматически учитываются (в противоположность к предположению о независимости акустического фильтра и источника звука, которое привело к основным трудностям в совершенствовании малобитных и высококачественных вокодеров в последние 15 лет).

Есть еще одно преимущество при таком подходе. В случае использования в качестве параметрического описания коэффициентов линейного прогноза при интерполяции могут возникать неустойчивые состояния. В случае физического моделирования медленно меняющейся формы артикуляторного тракта такое невозможно.

В Московском техническом университете связи и информатики разработаны новые методы эффективного кодирования речи в классе линейного предсказания с анализом через синтез (ЛПАС), основанные на синтезе сигнала погрешности предсказания ортогональными полиномами Чебышева как в спектральной, так и во временной областях [5]. На скоростях 8-16 кбит/с данные методы кодирования обеспечивают первый класс качества по разборчивости в соответствии с ГОСТ Р 51061-97. Повышенное качество синтеза речи здесь обеспечивается за счет оптимизации (по критерию максимума отношения сигнал/суммарная погрешность синтеза) метода рекуррентной (со взвешиванием) оценки коэффициентов линейного предсказания для нестационарной речи. Получены оптимальные оценки множителя забывания.

Разработан новый метод низкоскоростного кодирования речи в классе ЛПАС на основе модели голосового возбуждения с динамической частотно-импульсной модуляцией (ДЧИМ). На скоростях 3,5-4,0 кбит/с метод обеспечивает первый класс качества по разборчивости в соответствии с ГОСТ Р 51061-97. По сравнению с методом ADPCM, здесь достигается сжатие цифрового представления речи в 8 и более раз.

Перспективным направлением развития систем кодирования речи представляется направление, связанное с моделью речи, предложенной К. Стивенсом. Центральной идеей этой модели является представление о том, что информационное ядро речевого сообщения привязано к ограниченными речевым участкам этого сообщения. Остальные временные отрезки являются только неким наполнением, связанным с физическими ограничениями артикуляторного тракта. Практически на этой идее основан подход, который получил название *waveform interpolation*. В этом случае в речевой волне вы-

деляются и передаются в канал только отдельные периоды основного тона, а остальные при декодировании вычисляются методом декодирования.

На аналогичной идее основан еще один интересный подход (*temporal decomposition*) к проблеме кодирования. При этом предполагается, что набор спектральных параметров может быть представлен линейной комбинацией, набора перекрывающихся компактных функций.

В последнее десятилетие в мире возникло и оформилось новое научное направление, связанное с так называемым *вейвлет-преобразованием*. Слово «wavelet», являющееся переводом с французского «ondelette», означает «небольшие волны, следующие друг за другом».

Вейвлет-преобразование сигналов обеспечивает возможность весьма эффективного сжатия сигналов (не только речевых) и их восстановления с малыми потерями информации, а также решение задач фильтрации сигналов. Основная идея вейвлет-преобразования состоит в представлении некоторой случайной функции (исследуемого сигнала) как суперпозиции определенных базисных негармонических функций — вейвлетов (см. рис. 5).

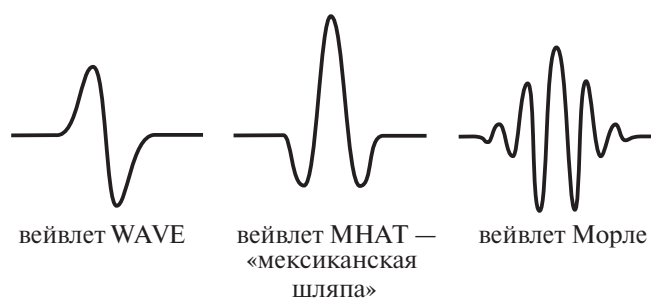


Рис. 5. Базовые вейвлет-функции

Для того чтобы вейвлеты хорошо аппроксимировали исходный сигнал, они подвергаются масштабированию (сжатию или растяжению) и сдвигу (смещению). Результат вейвлет-преобразования — обычный массив числовых коэффициентов. Такая форма представления информации о сигнале очень удобна, поскольку числовые данные легко обрабатывать.

После масштабирования выполняется этап порогового преобразования. На этом этапе отбрасываются коэффициенты, значение которых близко к нулю. Следует помнить, что при этом происходит необратимая потеря информации, так как «отброшенные» коэффициенты участвуют в формировании (восстановлении) сигнала. Поэтому выбранное пороговое значение коэффициентов сильно влияет на качество сигнала — задание слишком высокого порога повлечет за собой снижение качества.

Уникальные свойства вейвлетов позволяют построить базис, в котором данные будут представлены всего лишь несколькими ненулевыми коэффициентами. Это означает, что массив коэффициентов можно сильно сжать обычными методами без потери информации.

Заключение

В ближайшее время можно ожидать появление систем компрессии до 600–900 бит/с. Скорее всего, они будут опираться на интерполяцию речевого сигнала во временной области. Возможность сохранения индивидуальных особенностей в этом случае весьма сомнительна. Во всяком случае, эта сторона проблемы требует дополнительного изучения, что потребует разработки специальных методик.

В настоящее время сотрудниками лаборатории исследования речи Харьковского национального университета радиоэлектроники разрабатывается новый метод кодирования речевого сигнала, который имеет принципиальные отличия от методов, описанных в статье. В основе метода лежит бионический принцип распознавания речи, в частности, продолжает свое развитие структурное (геометрическое) направление исследования речевого сигнала [6].

Список литературы: 1. Александр Крейнс. Как налить море в наперсток? Технологии компрессии голоса // Сети и системы связи. — 1996. — № 9-10. — С. 119-121. 2. <http://www.itu.ch/itudoc/itu-t/rec/g/g700-799.html>, ITU Coding Standards. 3. О. Варламова. Помехоустойчивые коды — будущее цифровой телефонии // Сети и системы связи. — 1997. — № 10. — С. 26-32. 4. ITU Recommendation G.728. Coding of Speech at 16 kbit/s Using Low-Delay Code Excited Linear Prediction, November 1994. 5. Медведев О. Н. Автореф. дисс. на соискание ученой степени канд. техн. наук. — Москва: Технический университет связи и информатики, 2007. — 19 с. 6. Бондаренко М.Ф., Дрюченко А.Я., Шабанов-Кушнаренко Ю.П. Гласные звуки в теории и эксперименте. — Харьков: ХНУРЭ, 2002. — 348 с.

Поступила в редколлегию 2.10.2008

УДК 004.934

Сучасні методи кодування мовного сигналу / М.Ф. Бондаренко, А.В. Работягов, С.В. Щепковський // Біоніка інтелекту: наук.-техн. журнал — 2008. — № 2 (69). — С. 106-114.

Аналізуються сучасні методи кодування мовного сигналу. Розглянуті методи, що вже мають статус міжнародних стандартів, а також перспективні методи, які знаходяться у стадії впровадження і розвитку. В результаті проведеного дослідження зроблений висновок щодо можливості створення найближчим часом методів з більш високим ступенем компресії мови.

Табл.: 6. Іл.: 5. Бібліогр.: 6 найм.

UDC 004.934

Modern methods of encoding of speech signal / M. Bondarenko, A. Robotyagov, S. Schepkovsky // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 106-114.

The modern methods of encoding of speech signal are analysed. Methods, already having status of international standards, and also perspective methods which are in the stage of introduction development, are considered. As a result of the conducted research a conclusion is done about possibility of creation in the near time methods with more high degree of compression of speech.

Tab.: 6. Fig.: 5. Ref.: 6 items.

УДК 007.5



ОЦЕНИВАНИЕ ПОЛНОТЫ И НЕПРОТИВОРЕЧИВОСТИ ДАННЫХ ДЛЯ ЗНАНИЕ-ОРИЕНТИРОВАННОГО ПРЕДСТАВЛЕНИЯ БИЗНЕС-ПРОЦЕССОВ

В.М.Левыкин¹, С.Ф.Чалый²¹ХНУРЭ, г.Харьков, Украина, levykin@kture.kharkov.ua²ХНУРЭ, г.Харьков, Украина, chaliy@datasvit.net

Предложены критерии и метод оценивания полноты и непротиворечивости данных знание-ориентированной модели представления БПИС, что позволяет выявить несоответствия данных модели БПИС и реального процесса и выполнить соответствующую реорганизацию процесса на основе знаний, выраженных в форме бизнес-правил.

БИЗНЕС-ПРОЦЕССЫ С ИЗМЕНЯЕМОЙ СТРУКТУРОЙ, БИЗНЕС-ПРАВИЛА, ПОЛНОТА, НЕПРОТИВОРЕЧИВОСТЬ

Введение

Современный этап процессного управления предприятием характеризуется необходимостью управления гибкими бизнес-процессами, ориентированными на пользователя и изменяющимися на основе знаний о функционировании предприятия.

Такие бизнес-процессы с изменяемой структурой (БПИС) характеризуются изменением временных параметров и последовательности входящих в состав процесса бизнес-процедур во время его функционирования, распределенностью БПИС между подразделениями предприятия, значительной степенью параллелизма, наличием временных, финансовых, материальных ограничений, взаимодействием нескольких предприятий при функционировании бизнес-процессов.

БПИС представляется множеством ситуаций St_l , $l = \overline{1, L}$ [1]. Для обработки каждой ситуации выполняется последовательность бизнес-процедур в соответствии со сценарием Sc_l , входящим в состав ситуации. Бизнес-процедуры сценария связаны между собой через общий набор данных D_l . Важной особенностью данного представления БПИС является использование знаний, которые определяют запуск процедур сценария. Знания могут быть представлены двояко — либо в модели в форме бизнес-правил, либо как знания исполнителей. Следует отметить, что знания в форме бизнес-правил формализуют функциональные требования к бизнес-процессам [2], что дает возможность реорганизовать БПИС во время его функционирования.

Важным аспектом процессного управления на основе БПИС является управление жизненным циклом бизнес-процессов.

Управление жизненным циклом бизнес-процессов направлено на поддержку достижения поставленных целей БПИС путем документирования, тестирования, анализа и оптимизации критически важных процессов.

Документирование БПИС позволяет получить формальное описание текущего состояния про-

цессов и является необходимым условием для их тестирования, анализа и оптимизации. Тестирование БПИС позволяет проверить корректность функционирования процесса в заданных условиях. Анализ БПИС позволяет найти узкие места и неэффективные процедуры в процессах, а также возможности по оптимизации бизнес-процессов, которые прозрачны (невидимы) со стороны соответствующего предприятия.

Таким образом, тестирование БПИС ориентировано на оценку полноты и непротиворечивости имеющегося формального знание-ориентированного представления бизнес-процесса путем выявления и локализации несоответствий с реальным процессом. Выявление таких несоответствий во время функционирования бизнес-процессов на реальном предприятии приводит к значительным материальным и временным затратам. Поэтому проблема оценки полноты и непротиворечивости на этапе получения формального представления бизнес-процесса является актуальной.

Отметим, что решение указанной проблемы является особенно важным для БПИС в силу характерных для них динамических изменений во время выполнения.

1. Формулировка задачи

В ряде исследований [3, 4] показано, что наиболее типичными ошибками для бизнес-процессов являются ошибки доступа к данным. Соответственно, рассматриваемая задача оценки полноты и непротиворечивости сводится к задаче оценивания полноты и непротиворечивости данных для формального представления БПИС.

Исходными данными задачи являются множества данных D_l ситуаций St_l , каждое из которых включает наборы входных и выходных данных бизнес-процедур, а также наборы свободных данных, используемых для принятия решений о реорганизации бизнес-процесса.

Требуется проверить использование данных для каждой ситуации БП по критериям полноты и непротиворечивости.

Формально задача оценки полноты и непротиворечивости данных для знание-ориентированного представления БПИС состоит в следующем.

Пусть $BPS_i \in BPS$, $St_l \in BPS_i$, Sc_l ,

$$D_l \in St_l, i = \overline{1, I}, l = \overline{1, L},$$

$$\begin{aligned} \text{тогда } (A_{K_{r_{p1}}}(D_l) \wedge A_{K_{r_{p2}}}(Sc_l) \wedge A_{K_{r_N}}(Sc_l)) &= true, \\ \Rightarrow A(BPS_i) &= true, \\ (A_{K_{r_{p1}}}(D_l) \wedge A_{K_{r_{p2}}}(Sc_l) \wedge A_{K_{r_N}}(Sc_l)) &= false \quad (1) \\ \Rightarrow A(BPS_i) &= false, \end{aligned}$$

где $A_{K_{r_{p1}}}(D_l)$ – результаты проверки по критерию полноты данных ситуации в целом; $A_{K_{r_{p2}}}(Sc_l)$ – результаты проверки по критерию полноты входных и выходных данных процедур; $A_{K_{r_N}}(Sc_l)$ – результаты проверки непротиворечивости свободных данных.

$$\begin{aligned} A_{K_{r_{p1}}}(D_l), A_{K_{r_{p2}}}(Sc_l), A_{K_{r_N}}(Sc_l) &= \\ = \begin{cases} true, \\ \text{при удовлетворении критерия для } BPS_i, \\ false, \text{ в противном случае.} \end{cases} \quad (2) \end{aligned}$$

Тогда задача оценки полноты и непротиворечивости данных имеет вид:

$$A(BPS_i) \rightarrow true. \quad (3).$$

Таким образом, в соответствии с (3) необходимо добиться, чтобы набор данных формальной модели БПИС удовлетворял всем трем рассмотренным критериям.

2. Оценивание полноты и непротиворечивости данных модели БПИС

При решении поставленной задачи функционирование бизнес-процесса целесообразно представлять в виде ориентированного графа, вершины которого представляют последовательность состояний процесса, а дуги – переходы между этими состояниями. Тогда состояние процесса в модели в каждый момент времени определяется наборами состояний составляющих его ситуаций. Состояние ситуации характеризуется соответствующей структурой данных D_l и значением этих данных.

Изменение состояния ситуации определяется результатом выполнения процедур и внешних по отношению к процессу воздействий. Изменение порядка выполнения процедур, заданного в сценарии ситуации, определяется на основе знаний путем реализации соответствующих бизнес-правил. Следовательно, переходы между состояниями процесса в общем случае являются результатом выполнения процедур и бизнес-правил, изменяющих наборы данных соответствующих ситуаций.

Очевидно, что полная проверка всех возможных вариантов реализации бизнес-процесса требует значительных временных и материальных затрат.

На практике, например при тестировании программного обеспечения, используются критерии

выбора тестов. Такие критерии обеспечивают лишь определенную степень полноты проверки тестируемого объекта и его элементов. Примерами таких критериев являются:

- однократная проверка всех функций, реализуемых программой;
- однократная проверка всех ветвей.

Считается, что использование последнего критерия позволяет выявить 67-90% ошибок в программном обеспечении. В то же время бизнес-процесс представляет собой более сложный объект, чем программа в силу своей непредсказуемости, например вследствие непредвиденных внешних воздействий, возникновения нештатных состояний.

Наиболее характерными ошибками и несоответствиями для бизнес-процессов современного предприятия являются ошибки данных [5]. Примерами таких ошибок являются: создание объектов данных (либо атрибутов объектов данных), которые не используются при реализации БП; отсутствие необходимых объектов данных или их атрибутов; дублирование объектов данных или их атрибутов. Указанные ошибки обычно проявляются при доступе к данным во время запуска и выполнения бизнес-процедур.

Основная проблема, которая возникает при оценке полноты и непротиворечивости модели – это проблема выбора критерия оценки. Указанный критерий определяет те элементы БП, на основании проверки состояния которых выполняется анализ адекватности модели.

В силу специфики модели БПИС, характеризующейся управлением на основе данных, а также рассмотренных ошибок данных для бизнес-процессов предприятия, при создании критерия проверки адекватности модели БПИС целесообразно в первую очередь учитывать структуру данных процесса.

При обработке структуры данных БПИС, характеризующей выполняющиеся ситуации, могут возникать следующие события:

- доступ к объекту данных;
- определение (изменение) значения объекта данных;
- определение атрибутов объекта данных.

Соответственно, определение элемента данных модели БПИС будем рассматривать как определение его атрибутов.

Определение объекта данных для рассмотренного представления БПИС происходит во время определения ситуации.

Определение значений объекта данных происходит либо при возникновении ситуации, либо при выполнении процедур сценария, связанных с ситуацией. Данные используются при запуске ответствующих процедур.

С целью проверки полноты и непротиворечивости данных рассмотренного представления БПИС определим функцию оценки доступа к данным $O(d)$, которая учитывает изменения состояний данных модели.

Пусть $\forall d \in D, D \in St, s_d \in S_D$,

$$S_D = \{S_{\text{Начальное}}, S_{\text{Определенное}}, S_{\text{Неопределенное}}\},$$

$$S_{\text{Определенное}} = \{true, false\},$$

тогда

$$\begin{aligned} (s_d = S_{\text{Начальное}}) &\rightarrow \\ (s_d \in S_{\text{Определенное}}) &\Rightarrow O(d) = O(d) + 1, \\ (s_d \in S_{\text{Определенное}}) &\rightarrow \\ (s_d = S_{\text{Неопределенное}}) &\Rightarrow O(d) = O(d) + 1, \\ s_d \rightarrow \neg s_d &\Rightarrow O(d) = O(d) + 1, (\forall s_d \in S_{\text{Определенное}}). \end{aligned} \quad (4)$$

Как видно из (4), функция оценки доступа к данным приобретает ненулевое значение в трех случаях: при изменении состояния данных из начального в определенное; при изменении s_d из определенного в неопределенное; при изменении значения s_d в определенном состоянии.

Введение функции оценки доступа к данным позволяет путем подсчета количества изменений s_d выполнить оценку полноты данных модели.

В соответствии с особенностями рассматриваемого представления БПИС выполнение каждой процедуры сценария связано с упорядоченным набором свободных данных.

Определение. Упорядоченный набор свободных данных Kt_{Br_j} бизнес-процедуры Br_j представляет собой упорядоченные пары подмножеств свободных данных $D_{ADD_j}^1, D_{ADD_j}^2$, причем данные первого подмножества изменяют свое значение до начала реализации процедуры Br_j , а второго – в процессе ее выполнения:

$$\begin{aligned} \forall D_{ADD} \exists Kt_{Br_j}, Kt_{Br_j} &= (D_{ADD_j}^1, D_{ADD_j}^2), \\ t_{D_{ADD_j}^1}^{end} < t_{D_{ADD_j}^2}^{end}, t_{D_{ADD_j}^1}^{end} < t_{Br_j}^{beg}, t_{D_{ADD_j}^2}^{end} &\geq t_{Br_j}^{beg}, \quad j = \overline{1, J}, \end{aligned} \quad (5)$$

где D_{ADD} – множество свободных данных ситуации; $t_{D_{ADD_j}^1}^{end}, t_{D_{ADD_j}^2}^{end}$ – моменты времени последнего изменения состояния подмножеств $D_{ADD_j}^1, D_{ADD_j}^2$ соответственно; $t_{Br_j}^{beg}$ – момент начала выполнения процедуры Br_j .

Таким образом, упорядоченный набор Br_j свободных данных определяет последовательность появления свободных данных, необходимых для принятия решений по управлению процедурой Br_j на основе знаний в форме бизнес-правил.

Введение упорядоченного набора данных позволяет проверить непротиворечивость свободных данных, используемых соответствующей бизнес-процедурой. Действительно, упорядоченный набор данных с одной стороны задает возможный

порядок принятия решений по управлению процедурой, а с другой – фиксирует изменение состояния элементов данных.

Для выполнения оценки данных на основе (1) и (2) предлагаются критерии оценки полноты и непротиворечивости данных ситуационно-сценарной модели, отражающие особенности описания бизнес-процессов с изменяемой структурой.

Критерии оценки полноты и непротиворечивости имеют следующий вид.

Критерий полноты данных для ситуации Kr_{p1} определяется следующим образом: каждый элемент структуры данных анализируемой ситуации должен быть проверен функцией оценки доступа к данным $O(d)$ по крайней мере один раз.

Если $\forall d \in D_l, \exists O(d) \geq 1$, то $Kr_{p1} = true$,

$$D_l \in St_l, l = \overline{1, L}. \quad (6)$$

Критерий полноты данных для процедуры Kr_{p2} определяется следующим образом: каждый элемент d_{IN} входных и d_{OUT} выходных данных анализируемой бизнес-процедуры Br_j , входящей в состав сценария ситуации, должен быть проверен функцией оценки доступа к данным $O(d)$ по крайней мере один раз.

Пусть $d_{IN} \in D_{IN_j}, Br_j \in Sc_i, d_{OUT} \in D_{OUT_j}$,

$$D_{IN_j}, D_{OUT_j} \subset D_l, j = \overline{1, J}$$

$$\text{Если } \forall Br_j (\exists O(d_{IN}) \geq 1) \wedge (\exists O(d_{OUT}) \geq 1), \quad (7)$$

то $Kr_{p2} = true$.

Критерий непротиворечивости данных Kr_N определяется следующим образом: каждый элемент упорядоченного набора свободных данных $Kt_{D_{ADD_j}}$ анализируемой бизнес-процедуры Br_j должен быть проверен по крайней мере один раз.

Пусть $d_{ADD}^1 \in D_{ADD_j}^1, d_{ADD}^2 \in D_{ADD_j}^2$,

$$D_{ADD_j}^1, D_{ADD_j}^2 \subset Kt_{D_{ADD_j}}, j = \overline{1, J}.$$

Если $(\forall Kt_{D_{ADD_j}}) \wedge$

$$\wedge (\exists O(d_{ADD}^1) \geq 1) \wedge (\exists O(d_{ADD}^2) \geq 1), \quad (8)$$

то $Kr_N = True$.

Предложенные критерии не зависят от типа оцениваемой бизнес-процедуры. Следовательно, в общем случае указанные критерии могут применяться к подпроцессам либо бизнес-процессам в целом.

Для решения сформулированной задачи предлагается метод оценивания полноты и непротиворечивости данных знание-ориентированной модели бизнес-процесса.

В соответствии с данным методом оценивание полноты и непротиворечивости данных включает в себя следующие основные этапы:

1. Присвоение исходных значений объектов данных в соответствии с тестовым набором данных.

2. Анализ ситуации в целом соответствии с критерием Kr_{p1} :

$$\forall St_l \in BPS_i, D_l \in St_l, i = \overline{1, I}, l = \overline{1, L}.$$

Если $\forall d \in D_l \quad Kr_{p1} = true$,

то $A_{Kr_{p1}}(D_l) = true$,

иначе $A_{Kr_{p1}}(D_l) = false$. (9)

3. Принятие решений о переопределении объектов данных на основании результатов этапа 1.

Если $A_{Kr_{p1}}(Br_j) = false$,

то корректировать D_l ,

иначе перейти к этапу 3. (10)

4. Выбор сценария, содержащего последовательность процедур БП в соответствии с тестовым набором данных.

5. Выбор процедуры в соответствии со сценарием.

6. Анализ процедуры в соответствии с критерием Kr_{p2} :

$$\forall St_l \in BPS_i, Sc_l \in St_l, i = \overline{1, I}, l = \overline{1, L}.$$

Если $\forall Br_j \in Sc_l \quad Kr_{p2} = True \Rightarrow$

$$\Rightarrow A_{Kr_{p2}}(Sc_l) = true. (11)$$

7. Анализ процедуры в соответствии с критерием непротиворечивости Kr_N :

$$\forall St_l \in BPS_i, Sc_l \in St_l, i = \overline{1, I}, l = \overline{1, L}.$$

Если $\forall Br_j \in Sc_l \quad Kr_N = True \Rightarrow$ (12)

$$A_{Kr_N}(Sc_l) = true.$$

8. Принятие решений о корректировке модели БП на основании результатов этапов 6 и 7.

Если $A_{Kr_{p2}}(Br_j) = false$,

то корректировать $Br_j, R_{IN_j}, R_{OUT_j}$. (13)

Если $A_{Kr_N}(Br_j) = false$,

то корректировать $Sc, Br_j, R_{IN_j}, R_{OUT_j}$. (14)

Этапы 5-8 повторяются для всех процедур сценария.

Реализация предлагаемого метода связана с выполнением набора бизнес-правил, отражающих знания о взаимосвязи элементов в модели БПИС.

На основе указанных правил в результате анализа адекватности выполняется корректировка модели БПИС путем изменения взаимосвязей между ее элементами.

Разработанный метод оценки полноты и непротиворечивости данных ориентирован на выявление и локализацию ошибок и несоответствий при

реинжиниринге БПИС во время его функционирования с учетом изменившихся из-за различных воздействий элементов данных, а также до выполнения измененных фрагментов БП, что дает возможность достигнуть запланированного уровня результативности процесса.

Оценим количество проверок данных при тестировании полноты и непротиворечивости данных ситуации предложенной модели БП.

Число возможных проверок $A_{Kr_{p1}}^*$ при оценке полноты данных ситуационно-сценарной модели по критерию Kr_{p1} определяется следующим образом:

$$A_{Kr_{p1}}^* \leq \delta * |D|, (15)$$

где $|D|$ – мощность множества данных рассматриваемой ситуации; δ – коэффициент, соответствующий количеству переходов между состояниями данных в модели.

Действительно, согласно критерию Kr_{p1} проверяется каждый элемент структуры данных анализируемой ситуации. Изменение состояния каждого элемента данных проверяется согласно выражению (4) и, следовательно, $\delta = 3$. Таким образом, выражение (15) принимает вид: $A_{Kr_{p1}}^* \leq 3 * |D|$.

Число возможных проверок $A_{Kr_{p2}}^*$ при оценке полноты данных процедуры по критерию Kr_{p1} определяется следующим образом:

$$A_{Kr_{p2}}^* \leq \delta * \prod_j (|D_{IN_j}| * |D_{OUT_j}|), (16)$$

где D_{IN_j} – подмножество входных данных бизнес-процедуры Br_j ; D_{OUT_j} – подмножество выходных данных бизнес-процедуры Br_j .

В соответствии с (16) максимальное количество проверок данных по каждой Br_j определяется как $|D_{IN_j}| * |D_{OUT_j}|$ в силу зависимости между входными и выходными данными бизнес-процедуры. Зависимость между процедурами сценария определяет мультипликативный характер оценки (16).

Число возможных проверок данных $A_{Kr_N}^*$ при оценке непротиворечивости данных процедуры по критерию Kr_N определяется следующим образом:

$$A_{Kr_N}^* \leq \delta * \sum_j (|D_{ADD_j}^1| + |D_{ADD_j}^2|), (17)$$

где $D_{ADD_j}^1$ – подмножество свободных данных, элементы которого изменяют свое значение до начала реализации процедуры Br_j ; $D_{ADD_j}^2$ – подмножество свободных данных, элементы которого изменяют свое значение в процессе выполнения процедуры Br_j .

Элементы подмножеств свободных данных $D_{ADD_j}^1$ и $D_{ADD_j}^2$ в общем случае являются взаимно независимыми, поэтому оценка (17) является аддитивной.

Выводы

Полученные критерии и метод оценивания полноты и непротиворечивости данных модели БПИС характеризуется проверкой набора данных, определяющих ситуации процесса, а также входных и выходных наборов данных для процедур сценария (включая, при необходимости, последовательность данных). Использование полученных результатов позволяет выявить несоответствия данных модели БПИС и реального процесса и выполнить соответствующую реорганизацию процесса на основе знаний, выраженных в форме бизнес-правил.

Список литературы: 1. Чалый С.Ф. Разработка технологии управления слабоструктурированными бизнес-процессами // Автоматизированные системы управления и приборы автоматики. — 2006. — № 135. — С. 63-71. 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. Калянов Г.Н. Теория и практика реорганизации бизнес-процессов. — М.: СИНТЕГ, 2000. 4. W.M.P. van der Aalst. Business Process Management Demystified: A Tutorial on Models, Systems and Standards for Workflow Management. In J. Desel, W. Reisig, and G. Rozenberg, editors, Lectures on Concurrency and Petri Nets, volume 3098 of Lecture Notes in Computer Science, pages 1-65. Springer-Verlag, Berlin, 2004. 5. Автоматизация управления предприятием / В.В. Баронов, Г.Н. Калянов, Ю.И. Попов и др. — М.: Инфра-М, 2000.

Поступила в редколлегию 3.10.2008

УДК 007.5

Оцінювання повноти і несуперечності даних для знання-орієнтованого представлення бізнес-процесів/ В.М. Леви́кін, С.Ф. Чалий // Біоніка інтелекту: наук.-техн. журнал — 2008. — № 2 (69). — С. 115-119.

Сформовано критерії та розроблено метод оцінювання повноти і несуперечності даних знання орієнтованої моделі представлення бізнес-процесів із змінною структурою. Запропонований метод характеризується перевіркою набору даних, що визначають ситуації процесу, а також вхідних і вихідних наборів даних для процедур сценарію в моделі. Метод дозволяє виявити невідповідності даних моделі бізнес-процесу із змінною структурою та реального процесу. Реорганізація процесу виконується на основі визначених невідповідностей за допомогою знань про його функціонування, виражених у формі бізнес-правил.

Бібліогр.: 5 найм.

UDC 007.5

Evaluation of plenitude and uncontradiction of information for knowledge-oriented presentation of business-process/ V.M.Levikin, S.F.Chaliy // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 115-119.

The criteria and the method of evaluation of plenitude and uncontradiction of information for knowledge-oriented presentation of business-process with a variable structure is developed. The offered method is characterized by verification of datasets. That datasets are connected to scenario of model and determine the situations of process, input and output procedure's data. The method allows to find out data disparities of model of business process with a variable structure and real process. Reorganization of process is executed on the basis of certain disparities by knowledges about it functioning, expressed as business rules.

Ref.:5 items.

УДК 519.08



МЕТОДЫ ПРИНЯТИЯ РЕШЕНИЙ В ИНВЕСТИЦИОННОМ АНАЛИЗЕ

Э.Г. Петров¹, Т.С. Чайникова²¹ХНУРЭ, г. Харьков, Украина²ХНУРЭ, г. Харьков, Украина, e-mail: chaynikova@gmail.com

В данной статье рассматривается решение задачи выбора альтернатив в инвестиционном анализе. Представлены модели выбора решений для инвестирования в условиях риска и неопределенности.

ИНВЕСТИРОВАНИЕ, ПРАВИЛО ВАЛЬДА, ПРАВИЛО ГУРВИЦА, ЭВОЛЮЦИОННОЕ ПРОГРАММИРОВАНИЕ, МЕТОД ГРУППОВОГО УЧЕТА АРГУМЕНТОВ

Введение

Инвестирование — это процесс прямого или косвенного приумножения капитала: возврат денежных средств либо получение новых возможностей, которые влияют на снижение затрат или оптимизацию процессов предприятия. Из-за различных факторов (нестабильность социально-экономической среды, изменение спроса, «человеческого фактора») цели инвестирования часто не достигаются.

На протяжении многих лет инвесторы безуспешно пытаются найти универсальный метод, который позволил бы подвергать анализу инвестиционные решения, учитывать все возможные риски и вероятные результаты рыночного поведения ценных бумаг, анализировать текущее состояние портфелей ценных бумаг инвесторов, желаемые финансовые результаты и находить инвестиционные решения, вложения в которые приносило бы наилучший результат.

1. Постановка задачи

Финансовые результаты инвестирования, которые интересуют инвестора, зависят от множества разнородных факторов $K_i, i = \overline{1, n}$, которые характеризуют как состояния внешней социально-экономической среды, так и собственно объекта инвестирования, то есть $R = F(k_i), i = \overline{1, n}$.

При этом следует учесть, что многие из характеристик $K_i, i = \overline{1, n}$ являются не детерминированными (точечными) значениями, а носят интервальный характер с различной степенью информированности о распределении возможных значений на интервале, а значит и различными формами представления этой информации (вероятностной, нечеткой, интервальной).

В связи с этим возникает три задачи.

1. Синтез модели (1) формирования ожидаемого финансового эффекта от инвестирования.

2. Учет неопределенности факторов $K_i, i = \overline{1, n}$, характеризующих ситуацию инвестирования и определения интервальной оценки R .

3. Принятие решения о возможности и размерах инвестирования.

Целью настоящей статьи является рассмотрение подходов и методов решения указанных задач.

2. Проблема синтеза модели оценивания результатов инвестирования

Это проблема структурно-параметрической идентификации многофакторного оценивая модели (1), что связано с определением:

- количества и состава кортежа значимых факторов $\langle K_i \rangle, i = \overline{1, n}$;
- структуры и параметров модели (1).

В настоящее время перечисленные задачи решаются на эвристическом уровне путем методологии экспертного оценивания. Недостатки этой методологии широко известны, поэтому существует острая необходимость разработки и использования альтернативных формальных методов. Это направление в настоящее время широко известно как анализ данных (data mining).

Цель анализа данных состоит в выявлении скрытых правил и закономерностей в наборах данных. Человеческий разум не способен улавливать более двух-трех взаимосвязей даже в небольших выборках. Но и традиционная математическая статистика, долгое время претендовавшая на роль основного инструмента анализа данных, также нередко пасует при решении задач из реальной сложной жизни.

Анализ данных — это процесс обнаружения в сырых данных:

- 1) ранее неизвестных;
- 2) нетривиальных;
- 3) практически полезных;
- 4) доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

Существуют различные подходы к классификации самой области анализа данных. В частности, выделяют «*дескриптивный*» и «*предиктивный*» подходы. В первом случае обнаруживаются знания описательного характера, а во втором — знания, которые можно использовать для прогноза. Предметом анализа могут быть как хорошо структурированные данные, например, таблицы в реляционной базе, так и неструктурированные данные, например, текст или изображения. Можно выделить, по крайней мере, семь методов выявления и анализа знаний: классификация, регрессия, клас-

теризация, анализ ассоциаций, прогнозирование временных последовательностей (рядов), агрегирование (обобщение), обнаружение отклонений. Вероятно, наиболее распространенной сегодня операцией интеллектуального анализа данных является классификация. С ее помощью выявляются признаки, характеризующие группу, к которой принадлежит тот или иной объект.

Однажды определенный эффективный классификатор используется для классификации новых записей в базе данных в уже существующие классы, и в этом случае он приобретает характер прогноза. Например, классификатор, идентифицирующий риск инвестирования, может быть использован для целей принятия решения, велик ли риск приобретения акций определенной компании. То есть классификатор используется для прогнозирования возможности возврата займа.

На практике классификация является решением проблемы о возможной целесообразности инвестирования определенного предприятия. Этот вопрос, тревожащий любого инвестора можно, конечно, разрешить интуитивно. Если образ предприятия в сознании инвестора соответствует его представлению о выгодной инвестиции, то инвестировать можно, иначе — нет. По схожей схеме работают установленные в инвестиционных компаниях системы анализа данных. Лишенные субъективной предвзятости, они опираются в своей работе только на ретроспективную базу данных компании, где записывается детальная информация о каждой инвестиции и, в конечном итоге, факт ее прибыльности. Инвестиции в этих системах интерпретируются как векторы в пространстве, атрибутам которых соответствуют различные данные о них. Факт прибыльности (неприбыльности) инвестиции играет роль функции $y_i = \{+1; -1\}$. Часть упомянутой ретроспективной базы можно рассматривать как проверочных набор данных, а другую часть — как обучающий набор. При таком подходе задача определения риска, связанного с инвестициями, сводится к задаче построения классификатора. Решить ее можно с помощью нескольких методик. Также очевидно, что она имеет решение, поскольку интуитивно понятно, какого рода закономерности в данных обуславливают риск, связанный с инвестицией.

Анализ данных является мультидисциплинарной областью. Отсюда обилие методов и алгоритмов, реализованных в различных действующих системах анализа данных. Многие из таких систем интегрируют в себе сразу несколько подходов. Тем не менее, как правило, в каждой системе имеется какой-то ключевой компонент, на который делается главная ставка. Ниже приводится классификация указанных ключевых компонентов: статистические методы, нейронные сети, деревья решений, системы рассуждения на основе аналогичных случаев (теория прецедентов), нечеткая

логика, генетические алгоритмы, эволюционное программирование, алгоритмы ограниченного перебора, комбинированные методы.

Эволюционное программирование сегодня — самая молодая и наиболее перспективная ветвь анализа данных. Суть метода заключается в том, что гипотезы о виде зависимости целевой переменной от других переменных формулируются системой в виде программ на некотором внутреннем языке программирования. Если это универсальный язык, то теоретически на нем можно выразить зависимость любого вида. Процесс построения этих программ строится подобно эволюции в мире программ. Когда система находит программу, достаточно точно выражающую искомую зависимость, она начинает вносить в нее небольшие модификации и отбирает среди построенных таким образом дочерних программ те, которые повышают точность. Таким образом, система «выращивает» несколько генетических линий программ, которые конкурируют между собой в точности выражения искомой зависимости. Специальный транслирующий модуль переводит найденные зависимости с внутреннего языка системы на понятный пользователю язык (математические формулы, таблицы и прочее), делая их легкодоступными. Для того чтобы сделать полученные результаты еще понятнее для пользователя-нематематика, имеется богатый арсенал разнообразных средств визуализации обнаруживаемых зависимостей.

Поиск зависимости целевых переменных от остальных ведется в форме функций какого-то определенного вида.

Для финансовых рынков характерна нестабильность и неустойчивость. При этом известные модели на практике часто оказываются непригодными для прогнозирования. В такой ситуации для анализа, моделирования и прогнозирования этих процессов целесообразно применять методы прямого построения моделей по данным наблюдений (статистики). Цель таких методов — выявление неявных причинно-следственных связей и закономерностей, скрытых в ретроспективных данных, и построение математических моделей в явной форме. При этом необходимо искать как структуру, так и параметры моделей, то есть решать задачу структурно-параметрической, или просто структурной, идентификации.

Одним из наиболее удачных подходов к построению алгоритмов поиска зависимости переменных является метод группового учета аргументов (МГУА) [3]. С помощью этого метода зависимость ищется в форме полиномов. Причем сложные полиномы заменяются несколькими более простыми, учитывающими только некоторые признаки (группы аргументов). Обычно для этого используются попарные объединения признаков.

В качестве класса возможных структур предлагается полином Колмогорова-Габора.

$$P(x) = a_0 + \sum_{i=1}^n a_i k_i^-(x) + \sum_{r=n+1}^m \sum_{j=1}^n \sum_{g=1}^n a_r k_i^H(x) * k_g^H(x) + \dots$$

$$i = \overline{1, n}; g = \overline{1, n}, r = \overline{n+1, m}.$$

Он позволяет описать любую нелинейную зависимость и не накладывает никаких априорных ограничений, так как содержит в своем составе как первые, так и более высокие степени характеристик $k_i(x)$ и все возможные их комбинации [4].

МГУА основан на разбиении множества переменных на комбинаторные группы (чаще всего на пары). Для каждой группы формируется простейший полином (структура модели), и для него на основе экспериментальной выборки решается задача параметрической идентификации. Эта задача решается или путем составления и решения системы нормальных алгебраических уравнений, базирующихся на экспериментальных данных о численных значениях входа и выхода исследуемого объекта, или путем аппроксимации тех же данных по критерию наименьших квадратов (МНК). Задача идентификации параметров решается для каждой локальной групповой модели на обучающей экспериментальной последовательности и затем определяется ее качество по заданному критерию на проверочной последовательности. На основе этого производится селекция заданного числа лучших по качеству моделей первого уровня, которые опять группируются, и для них строятся аналогичные первому уровню полиномы. Процедура циклически повторяется до тех пор, пока будет получен полином оптимальной сложности [5].

3. Формирование оценки эффективности инвестиций

После синтеза модели (1) следует этап анализа, он состоит в вычислении ожидаемого инвестиционного эффекта R . Решение этой задачи существенно осложняется тем, что характеристики $K_i, i = \overline{1, n}$ содержат неопределенности различных видов. Это обусловливается неполнотой знаний, погрешностями измерений, малым интервалом ретроспективного наблюдения процесса, сознательным искажением данных и так далее. В результате будем предполагать, что любой из элементов K_i кортежа $\langle K_i \rangle, i = \overline{1, n}$ представляет собой интервал возможных значений K_i , которые характеризуют значения левой и правой границы $K_i^{\min}$ и $K_i^{\max}$:

$$K_i^{\max} \leq K_i \leq K_i^{\min} \quad (2)$$

Если $K_i^{\max} = K_i^{\min}$, величина K_i является детерминированным точечным значением, в противном случае это интервальная величина. Распределение значений внутри интервала может характеризоваться:

- статистической информацией (математическое ожидание, дисперсия, функция плотности распределения вероятности);
- нечеткой информацией в виде функции принадлежности некоторому нечеткому множеству;

- отсутствием информации, то есть представлять собой интервальное значение.

В этих условиях необходимо вычислить оценки финансовой эффективности инвестиции. Это можно сделать двумя путями:

- методом Монте-Карло [6];
- методом варьирования значений кортежей $\langle K_i \rangle, i = \overline{1, n}$.

Достоинство метода Монте-Карло заключается в том, что на основе статистического моделирования можно получить функцию плотности распределения интегральной оценки, ее математическое ожидание и дисперсию. Недостаток — высокая трудоемкость метода. Альтернативой является формирование матрицы платежей вида представленного в табл. 1.

4. Правила и критерии принятия решений в условиях риска и неопределенности

Результаты инвестирования зависят от многих условий, являются ненадежными и могут иметь следующий характер:

1) Оценка ожидаемой прибыли от инвестиций производится при известной вероятности получения того или иного результата.

2) Оценка производится в условиях, когда известна некоторая интервальная вероятность получения результата.

3) Оценка производится в условиях, когда нет четкой информации о результате, либо неполная информация для определения вероятности получения результата.

На основе субъективных эвристических соображений задачи второго и третьего типов можно свести к задачам первого типа. Рассмотрим различные варианты решения подобных задач.

При наличии конечного числа рассматриваемых альтернатив $A_j (j = \overline{1, m})$ инвестирования и известных значений кортежа различных факторов $\langle K_i \rangle, i = \overline{1, n}$, определяющих результат, а также при наличии функции, с помощью которой можно вычислить скалярную оценку результата инвестирования CK_{ji} , исходную информацию для принятия решения можно представить в виде следующей матрицы (табл. 1).

Таблица 1

Альтернатива	Действие различных факторов				
	$\langle K_1 \rangle$	...	$\langle K_i \rangle$	...	$\langle K_n \rangle$
A_1	CK_{11}		CK_{1i}		CK_{1n}
...					
A_j	CK_{j1}		CK_{ji}		CK_{jn}
...					
A_l	CK_{l1}		CK_{li}		CK_{ln}

Элементы матрицы (CK_{ji}) соответствуют значениям ожидаемого эффекта, которые принимает альтернатива A_j при состоянии среды Z_i , и исходя

из которых инвестор или ЛПР с помощью некоторых правил принятия решений может обоснованно выбрать ту или иную альтернативу.

Под неопределенностью следует понимать отсутствие, неполноту или недостаточность информации об объекте, процессе, явлении или неуверенность в достоверности информации. Неопределенность обуславливает появление ситуаций, не имеющих однозначного исхода для различных экономических, инвестиционных и других объектов. К основным правилам принятия решений при неопределенности относятся [1]:

- 1) правило Вальда, или правило максимина;
- 2) правило максимакса;
- 3) правило Гурвица.

В соответствии с правилом Вальда из альтернативных инвестиций (см. табл. 1) выбирают ту альтернативу A_j , которая при самом неблагоприятном состоянии внешней деловой среды $\langle K_i \rangle, i = \overline{1, m}$ имеет наибольшее значение эффекта CK_{ji} . Альтернатива A^* , которой принадлежит это максимальное значение, предлагается к реализации:

$$A^* = \left\{ A_j \max_j \min_i CK_{ji} \right\}.$$

Инвестор, принимающий решение по правилу Вальда, имеет малую склонность к риску и высокую — к пессимизму: предполагая экстремально негативное развитие состояния внешней деловой среды, он учитывает наименее благоприятное развитие каждой инвестиционной альтернативы.

Оптимистический инвестор, напротив, для выбора альтернатив использует правило максимакса, выбирая инвестицию с наивысшим достигаемым значением стоимости капитала и не учитывая при принятии решения риска, связанного с неблагоприятным развитием состояния внешней деловой среды. Следовательно, оптимальная альтернатива A^* , определяется по формуле:

$$A^* = \left\{ A_j \max_j \max_i CK_{ji} \right\}.$$

Основной недостаток правил максимина и минимакса при принятии решения заключается в использовании только одного варианта развития для каждой инвестиции. Таким образом, часть информации при этом не учитывается.

Правило Гурвица (правило оптимизма-пессимизма), сочетая правила максимина и минимакса, позволяет ЛПР выбирать оптимальную альтернативу A^* , по следующей формуле:

$$A^* = \left\{ A_j \max_j \left[(1 - \alpha) \cdot \min_i CK_{ji} + \alpha \max_i CK_{ji} \right] \right\},$$

где α — коэффициент оптимизма, $\alpha = [0, 1]$.

Если $\alpha = 1$, то выбор альтернативы проводится по правилу максимакса, если $\alpha = 0$ — по правилу максимина.

При применении правила Гурвица информации используется больше, чем при правилах максимина и минимакса в отдельности, но не вся имеющаяся в распоряжении ЛПР информация.

Под ситуацией риска следует понимать сочетание, совокупность различных обстоятельств и условий, создающих обстановку вокруг того или иного вида деятельности. Ей соответствуют три условия:

- 1) наличие неопределенности;
 - 2) необходимость выбора альтернатив;
 - 3) возможность оценить вероятность реализации выбираемых альтернатив.
- К основным правилам и критериям принятия решений в ситуации риска относятся правило Байеса; $\mu\sigma$ — критерий, или критерий среднего значения и стандартного отклонения; критерий Бернулли.

Правило Байеса применяется, если известна вероятность наступления p_i возможных состояний внешней среды, тогда в качестве критерия выбора можно использовать значение математического ожидания j -й альтернативы A . При наличии множества инвестиционных альтернатив оптимальная A^* альтернатива находится по формуле:

$$A^* = \left\{ A_j \max_j M[A_j] \right\}; \quad M[A_j] = \sum_{i=1}^n CK_{ji} \cdot p_i,$$

где $M[A_j]$ — математическое ожидание, CK_{ji} — эффективность j -й альтернативы при i -м состоянии внешней среды с вероятностью p_i его наступления.

Заметим, что элементы матрицы CK_{ji} являются оценками инвестиционных эффектов. Таким образом, изменение эффекта по отношению к изменению значения стоимости капитала принимается пропорциональным, а отношение к риску — нейтральным.

Критерий среднего значения и стандартного отклонения, или $\mu\sigma$ -критерий, позволяет оценивать отношение инвестора к риску.

Рассматривая, кроме математического ожидания $M[x]$, дисперсию Dx или стандартное среднеквадратическое отклонение σ — результатов как степень риска в критерии принятия решения, можно отметить, что риск тем выше, чем выше значение σ . Полезность альтернативных решений, так называемая полезность риска, зависит от $M[x]$ и σ и отражается функцией приоритетности риска, или коэффициентом вариации V .

Есть возможность замены $M[x]$ и моментов риска целевых функций на ожидаемую полезность; при этом вместо монетарных целевых функций используется полезность, которую ЛПР связывает с целями и ожидаемой степенью их достижения с учетом личного отношения к риску.

По критерию Бернулли, ЛПР способно оценить выгоду различных возможных инвестиций, отыскивая максимум «морального ожидания» (МО) для каждой альтернативы по следующей формуле:

$$MO = \sum_{i=1}^n f(CK_{ji}) \cdot p_i; \quad \sum_{i=1}^n p_i = 1,$$

где $f(CK_{ji})$ — дегрессивно растущая функция полезности, (CK_{ji}) — стоимость капитала j -й альтернативы при i -м состоянии внешней среды; p_i — вероятность его наступления.

С помощью теории полезности Бернулли можно определить функции полезности ненадежных результатов (ожиданий), например, стоимости капитала. Для этого находят надежный результат, так называемый надежный эквивалент, имеющий сходную выгоду с двумя ненадежными результатами, характеризующимися определенными значениями вероятности получения.

С помощью функции полезности можно определить ожидаемые значения полезности альтернатив; при этом в величины полезности трансформируются все возможные результаты. Альтернатива, имеющая максимальное значение математического ожидания полезности, считается оптимальной. При нейтральном отношении к риску этот критерий соответствует правилу Байеса.

Прогнозирование риска связано с разработкой сценариев улучшения финансового состояния субъекта рынка, с предвидением наступления определенных и неожиданных случайных событий, с проработкой альтернатив развития, рыночных механизмов и принятием гибких экстренных решений. Выбор варианта решения может осуществляться в разнообразных условиях возможных хозяйственных ситуаций, когда их вероятности могут быть как известными заранее, так и неизвестными, но при существовании оценок их относительных значений или знания основных направлений оценки результата вложения капитала.

Заключение

Принятие решений в таких сферах как анализ процессов в макроэкономике, прогнозировании поведения финансовых рынков, проверке надежности фирм, требуют инструментальных средств, которые способны находить точные модели процессов. При этом возникают проблемы, связанные с большим числом переменных, небольшим количеством наблюдений и неизвестными динамическими связями между переменными. Такие экономические объекты являются сложными плохо-обусловленными системами, которые характеризуются:

- 1) недостаточной априорной информацией;
- 2) большим количеством параметров, которые не измеряются;
- 3) зашумленными или короткими выборками данных;
- 4) плохо-обусловленными объектами с характеристиками.

Проблемы моделирования сложных экономических систем могут быть решены с помощью дедуктивных логико-математических или с помощью индуктивных переборных методов. Дедуктивные и имитационные методы имеют преимущества в случае простых задач моделирования, если известна теория объекта, который моделируется, и потому возможно построение модели, исходя из физически обоснованных принципов, применяя знания человека относительно процессов в объекте. Но эти методы не в состоянии дать удовлетворитель-

ный результат для сложных систем. В этом случае получение знаний по данным, то есть нахождение модели на основе экспериментальных измерений, имеет преимущества. Такие объекты содержат минимальное априорное знание или не имеют теоретических основ вообще. Эти проблемы могут быть решены с помощью Метода Группового Учета Аргументов (МГУА), который находит знание об объекте непосредственно по выборке данных. Этот индуктивный переборный метод имеет преимущества при исследовании сложных объектов, которые не имеют определенной теории, в частности объектов с размытыми характеристиками.

Список литературы: 1. Овезгельдыев А.О., Петров Э.Г., Петров К.Э. Синтез и идентификация моделей многофакторного оценивания и оптимизации. — К.: Наукова думка, 2002. 161с. 2. http://www.citforum.ru/consulting/BI/data_mining/3.shtml. 3. Ивахненко А.Г., Лана В.Г. Предсказание случайных процессов, АН УССР, 1971. 416с. 4. Петров Э.Г., Булавин Д.А., Петров К.Э. Использование генетических алгоритмов для решения задачи структурно-параметрической идентификации модели индивидуального многофакторного оценивания // Проблемы бионики: Всеукр. межвед. науч.-техн. сб. — 2003. — № 60. 5. Петров Э.Г., Булавин Д.А., Петров К.Э. Решение задачи структурно-параметрической идентификации модели индивидуального многофакторного оценивания методом группового учета аргументов // АСУ та прилади автоматики: Всеукр. міжвід. наук.-техн. зб. — 2004. — Вип. 4. — С. 65-76. 6. Словарь по кибернетике / Под ред. В.М. Глушкова. Киев, 1979. — С.335.

Поступила в редколлегию 7.10.2008

УДК 519.08

Методи прийняття рішень в інвестиційному аналізі / Е. Г. Петров, Т.С. Чайнікова // Біоніка інтелекту: наук.-техн. журнал — 2008. — № 2 (69). — С. 120-124.

Прийняття рішень у таких сферах, як аналіз процесів у макроекономіці або прогнозуванні фінансових ринків потребують засобів, які можуть знаходити точні моделі на основі прогнозів процесів. Проблеми моделювання складних економічних систем можуть бути вирішені за допомогою дедуктивних або індуктивних переборних методів. Вони мають переваги у випадку роботи з простими моделями. Та у разі моделювання складних систем вони не здатні надати задовільний результат. Ця проблема може бути вирішена за допомогою Метода Групового Урахування Аргументів (МГУА), який знаходить знання про об'єкт безпосередньо за вибіркою даних, зокрема, для об'єктів з розмитими характеристиками.

Табл.: 1. Бібліогр.: 6 найм.

UDC 519.08

The decision-making techniques in investment analysis / E.G. Petrov, T.S. Chaynikova // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 120-124.

Decision making in the spheres of process analysis in macroeconomics or financial markets forecasting require tools that are able to search accurate models on the base of process estimate. The problem of the complex economic systems modeling can be solved with the help of deductive or inductive selective methods. They have an advantage when there are plain models but in case of complex systems modeling these methods are not able to give reliable results. This problem can be solved by the Group Method of Data Handling (GMDH). It finds knowledge about the object directly by the data access. This inductive selective method has advantages while investigating complex objects, particularly for the objects with degraded characteristics.

Tab.: 1. Ref.: 6 items.

УДК 004.89



АДАПТИВНЫЕ НЕЧЕТКИЕ МОДЕЛИ ИДЕНТИФИКАЦИИ НЕЛИНЕЙНЫХ ОБЪЕКТОВ НА ОСНОВЕ ИСКУССТВЕННЫХ ИММУННЫХ СИСТЕМ

Н.М. Кораблев¹, И.В. Сорокина²

¹ХНУРЭ, г.Харьков, Украина, korablev@kture.kharkov.ua

²ХНУРЭ, г.Харьков, Украина, i.sorokina@kture.kharkov.ua

В данной работе рассмотрены вопросы адаптации нечетких моделей идентификации нелинейных объектов, представленных в виде системы нечеткого вывода и нечеткой нейронной сети, с использованием искусственных иммунных систем. Процесс адаптации состоит в настройке формы и параметров функций принадлежности, а также параметров и структуры (в случае системы нечеткого вывода) базы нечетких правил. Показана эффективность предложенных иммунных алгоритмов адаптации моделей нечеткого вывода на тестовых функциях.

ИСКУССТВЕННЫЕ ИММУННЫЕ СИСТЕМЫ, АДАПТИВНАЯ МОДЕЛЬ НЕЧЕТКОГО ВЫВОДА, АДАПТИВНАЯ НЕЧЕТКАЯ НЕЙРОННАЯ СЕТЬ, ФУНКЦИИ ПРИНАДЛЕЖНОСТИ, БАЗА НЕЧЕТКИХ ПРАВИЛ

Введение

Активное использование современных информационных технологий для создания математических основ интеллектуальных систем в последние годы позволило по-новому взглянуть на проблему обработки многомерных данных в условиях неполной и неточной информации. Одной из наиболее актуальных в этой сфере является задача идентификации нелинейных объектов. Классические методы многомерной статистики, используемые для решения такой задачи, основаны на определенных предположениях о структуре исходных данных и, как следствие, попытки их применения приводят к большим ошибкам при условии несоответствия обучающих наборов этим предположениям. Основные трудности при использовании классических систем нечеткого вывода для решения задачи идентификации нелинейных объектов связаны с априорным определением компонентов этих моделей (нечетких высказываний, функций принадлежности (ФП) для каждого значения лингвистических переменных, структуры базы нечетких правил и другое). При наличии набора неструктурированных данных определение этих компонентов представляет собой сложную проблему, требующую привлечения знаний экспертов в предметной области. Поскольку эти компоненты зачастую выбираются субъективно, они могут быть не вполне адекватны моделируемой системе или процессу. Специфика задачи идентификации нелинейных объектов приводит к необходимости устранения вышеуказанных недостатков.

Наиболее перспективным направлением при решении данной задачи является использование универсальных аппроксиматоров широкого класса многомерных нелинейных функций — адаптивных моделей нечеткого логического вывода и адаптивных нечетких нейронных сетей [1]. Параметры адаптивных моделей обоих типов настраиваются

путем их оптимизации в смысле некоторого критерия, формируемого по данным из обучающей выборки.

Решение подобной задачи оптимизации представляет собой трудную задачу по ряду причин. Во-первых, это большая размерность вектора параметров нечеткой модели. Во-вторых, большая размерность пространства поиска тесно связана с проблемой многоэкстремальности целевой функции. Кроме того, характер функции полностью зависит от данных обучающей выборки, которые меняются с каждым новым моделируемым процессом или объектом. В настоящее время накоплен большой арсенал средств для решения задач оптимизации, но поиск новых, более эффективных методов продолжается. Стандартными и хорошо изученными являются градиентные методы оптимизации [2]. Данная группа методов использует для поиска экстремума значение градиента функции. Недостатком этих методов является необходимость вычисления градиента функции, зависимость от начального приближения.

Широко распространена оптимизация на основе случайного поиска, когда градиент функции заменяется случайным вектором. Генетические алгоритмы как часть парадигмы эволюционных вычислений позволяют решать многие проблемы классических методов оптимизации. В [3] рассмотрено использование генетических алгоритмов в задачах оптимизации систем нечеткого вывода. Однако классические ГА и их разновидности не всегда эффективны при решении задачи мультимодальной оптимизации.

Развитие эволюционных вычислений с целью повышения эффективности алгоритмов обучения нечетких моделей ведется по пути создания новых методов, использующих возможности динамического изменения параметров алгоритмов оптимизации и различные варианты распараллеливания

процессов обработки нечеткой информации. Таким требованиям удовлетворяют искусственные иммунные системы (ИИС), которым присущи такие свойства как распознавание, разнообразие, обучение, память, распределенное обнаружение, саморегулирование, метадинамика, сетевая организация и другое [4]. В работе [5] рассмотрено применение ИИС в нечетких системах классификации, в частности для генерации базы нечетких правил.

Рассматривается также настройка адаптивных моделей нечеткого вывода, реализованных в виде системы нечеткого вывода и нечеткой нейронной сети на основе ИИС. Для оценки эффективности предложенных иммунных алгоритмов их работа проверяется на тестовых примерах. Проведенные эксперименты показывают, что предлагаемые иммунные алгоритмы способны выполнять адаптацию нечетких моделей идентификации нелинейных объектов.

1. Постановка задачи

Для адаптации нечеткой модели строится целевая функция:

$$\Phi(P) = \frac{1}{n} \sum_{j=1}^n [F(X_j, P) - y_j]^2 \Rightarrow \min_P, \quad (1)$$

минимизирующая среднеквадратическое отклонение между фактическими значениями выходной переменной y_j и полученной на основе нечеткого прогноза точечной оценкой $F(X_j, P)$. Вектор P представляет собой вектор параметров нечеткой модели и его состав определяется исходя из выбранного алгоритма обучения.

В качестве базовой модели нечеткого вывода используется модель Такаги-Сугено первого порядка с правилами вида:

$$P_i: \text{ЕСЛИ } x_i \text{ есть } A_{i1} \text{ И } \dots \text{ И } x_j \text{ есть } A_{ij} \text{ И } \dots \text{ И } x_m \text{ есть } A_{im} \\ \text{ТО } y = k_{i1}x_1 + \dots + k_{ij}x_j + \dots + k_{im}x_m + k_{i0}, \quad i = 1, \dots, n,$$

где A_i — это нечеткие множества и соответствующие им ФП, построенные в пространстве входных $x_i, 1, m$ значений переменных; $k_{ij} (j = 0, \dots, m)$ — коэффициенты аргументов функции.

Используемые гауссовы ФП описываются выражением:

$$\mu(x) = \exp \left[- \left(\frac{x - c}{\sigma} \right)^2 \right], \quad (2)$$

где c — центр нечеткого множества; σ отвечает за крутизну функции.

Для адаптации нечетких моделей (системы нечеткого вывода, нечеткой нейронной сети) ставится задача разработки алгоритмов на основе искусственных иммунных систем с использованием принципа клонального отбора [6-7] и теории им-

мунной сети [8]. Для настройки нечетких моделей необходимо оптимизировать параметры ФП и нечетких правил по данным из обучающей выборки.

2. Адаптивные системы нечеткого вывода

2.1. Модель системы нечеткого вывода

Основой адаптивной системы нечеткого вывода является база знаний, представляющая собой совокупность информации о множестве лингвистических переменных в виде соответствующих ФП и о множестве правил нечеткого вывода [1]. База знаний формируется на основе информации из обучающей выборки.

Представим адаптивную систему нечеткого вывода в виде кортежа:

$$AFIS = \langle \{X_i, y_i\}, RB, DB, I, G(RB), L(DB), F \rangle,$$

где $\{X_i, y_i\}, i = \overline{1, n}$ обучающая выборка; RB — база правил; DB — база данных; I — механизм нечеткого логического вывода; $G(RB)$ — генерация и оптимизация базы правил; $L(DB)$ — генерация и оптимизация базы данных (параметров ФП); F — функция, оценивающая эффективность нечеткой модели.

Основная трудность использования данной системы заключается в необходимости структурной и параметрической адаптации (оптимизации) — поиске оптимального набора параметров термов лингвистических переменных и коэффициентов нечетких правил, и структуры множества правил логического вывода [9].

2.2. Иммунный алгоритм адаптации правил нечеткого вывода (структурная адаптация)

Алгоритм формирования нечеткой базы знаний представляет собой итерационную процедуру последовательной идентификации наблюдений из обучающей выборки с помощью искусственных иммунных систем.

Антигены представляют собой примеры обучающей выборки. Каждое антитело кодирует одно правило базы знаний. Количество антител в популяции (правил в базе знаний) соответствует количеству примеров в обучающей выборке. Правила кодируются вещественными числами, представляющими собой индекс нечетких множеств входных переменных и коэффициенты нечетких правил:

$$Ab = \langle IFS_1, IFS_2, \dots, IFS_n, k_0, k_1, \dots, k_n \rangle,$$

где $IFS_1, IFS_2, \dots, IFS_n$ — индексы нечетких множеств для n входных переменных; $k_0, k_1, \dots, k_n$ — коэффициенты нечетких правил.

Аффинность антител популяции вычисляется по результатам нечеткого вывода в соответствии с (1). В результате работы иммунного алгоритма клонального отбора лучшие по популяции антитела формируют популяцию клеток памяти, которая представляет множество нечетких правил вида:

R_1 : IF x_1 is IFS_{11} ...AND... x_n is IFS_{1n} THEN
 $y = k_{10} + k_{11}x_1 + \dots + k_{1n}x_n$;

R_2 : IF x_1 is IFS_{21} ...AND... x_n is IFS_{2n} THEN
 $y = k_{20} + k_{21}x_1 + \dots + k_{2n}x_n$;

...

Rq : IF x_1 is IFS_{q1} ...AND... x_n is IFS_{qn} THEN
 $y = k_{q0} + k_{q1}x_1 + \dots + k_{qn}x_n$.

Полное множество правил в АСНВ для реальных задач может быть достаточно велико. Если для описания наблюдения используются n переменных, каждая из которых имеет m термов, то верхняя оценка числа правил составляет m^n .

Для сокращения мощности множества правил необходимо в этом множестве выделить некоторое минимальное подмножество наблюдений, исключив избыточные и противоречивые правила. Данную задачу в предлагаемом иммунном алгоритме выполняет оператор супрессии, который обеспечивает сжатие популяции путем связывания антител в зависимости от значения их аффинности относительно других антител популяции. Антитела с минимальной аффинностью удаляются из популяции в соответствии с установленным порогом сжатия сети, что приводит к уменьшению числа правил в базе знаний адаптивной системы нечеткого вывода.

Алгоритм структурной адаптации можно представить в виде следующей последовательности шагов:

1. Генерация начальной популяции антител Ab .
2. Цикл для каждого антигена Ag_i :
 - 2.1. В систему нечеткого вывода подставляются параметры ФП и правило вывода с коэффициентами, взятыми из антитела, вычисляется аффинность антител Ab к антигену Ag_i .
 - 2.2. Выбор n лучших антител для клонирования N раз.
 - 2.3. Клонирование, формирование популяции клонов C .
 - 2.4. Мутация популяции C , формирование популяции C^* .
 - 2.5. Вычисления аффинности популяции C^* к антигену Ag_i .
 - 2.6. Редактирование популяции антител Ab и заполнение популяции клеток памяти M .
 - 2.7. Супрессия антител внутри популяции Ab .
 - 2.8. Замена d худших антител новыми антителами в популяции антител Ab .
3. Проверка критерия останова. Если не достигнут — переход к шагу 2, в противном случае — переход к шагу 4.
4. Конец.

Результат — популяция клеток памяти M , которая представляет собой набор правил вывода с вычисленными коэффициентами нечетких правил.

2.3. Иммунный алгоритм адаптации ФП (параметрическая адаптация)

Параметрическая адаптация включает в себя процедуру оптимизации первичных значений параметров ФП. Принцип действия процедуры адаптации заключается в изменении параметров термов всех лингвистических переменных. Целью параметрической адаптации является поиск такой модели, с таким набором параметров, для которой качество идентификации будет наилучшим.

В задаче параметрической адаптации нечетких моделей в роли антигена выступает обучающая выборка; в роли антител — параметры ФП вида (2) для входных переменных. Антитело представляет собой строку фиксированной длины:

$$Ab = \langle c_{11}, \dots, c_{1m}, \dots, c_{n1}, \dots, c_{nm}, \sigma_{11}, \dots, \sigma_{1m}, \dots, \sigma_{n1}, \dots, \sigma_{nm} \rangle,$$

где c_{ij}, σ_{ij} , $i = \overline{1, n}$; $j = \overline{1, m}$ — параметры гауссовых ФП вида (2) для n входных переменных, каждая из которых имеет m термов.

Каждая из ФП задана на универсуме $X = [x_{\min}, x_{\max}]$, в качестве которого выбран замкнутый интервал действительных чисел. Исходный вектор параметров ФП генерируется случайным образом. Предлагается вещественное кодирование антител, что позволяет повысить точность найденных решений.

На вход адаптивной системы нечеткого вывода подается множество значений входных переменных из обучающей выборки. На основании параметров ФП, взятых из антител, и сгенерированной на этапе структурной адаптации базы правил выполняется нечеткий вывод. По результатам нечеткого вывода вычисляется аффинность каждого антитела со всеми антигенами популяции в соответствии с (1). Поиск минимума функции (1) ведется в допустимом множестве изменения входных переменных x_i : $x_i^{\min} \leq x_i \leq x_i^{\max}$, $i = \overline{1, n}$. Антитела с лучшей аффинностью подвергаются клонированию и мутации. Антитела с худшим значением аффинности удаляются из популяции и заменяются новыми. Работа иммунного алгоритма прекращается при достижении критерия останова. База данных адаптивной нечеткой системы формируется по лучшему антителу популяции.

Алгоритм параметрической адаптации можно представить в виде следующей последовательности шагов:

1. Генерация начальной популяции антител Ab .
2. Цикл для каждого антитела Ab_i :
 - 2.1 В систему нечеткого вывода с базой нечетких правил, полученных на этапе структурной адаптации, подставляются параметры ФП из антитела Ab_i , вычисляется аффинность антитела к популяции антигенов Ag .

2.2. Выбор n лучших антител для клонирования N раз.

2.3. Клонирование, формирование популяции клонов C .

2.4. Мутация популяции C , формирование популяции C^* .

2.5. Вычисление аффинности популяции C^* к популяции антигенов Ag .

2.6. Редактирование популяции антител.

2.7. Замена d худших антител новыми антителами в популяции антител Ab .

3. Проверка критерия останова. Если не достигнут — переход к шагу 2, в противном случае — переход к шагу 4.

4. Конец.

Результат — антитело с лучшей аффинностью, содержащее параметры ФП.

2.4. Взаимодействие алгоритмов структурной и параметрической адаптации

Обобщенный алгоритм обучения адаптивной модели нечеткого вывода для n входных переменных $x_1, x_2, \dots, x_n$ и одной выходной переменной y представляет собой итерационную процедуру, в которой организовано взаимодействие процедур параметрической и структурной адаптации.

На подготовительном этапе для каждого идентифицируемого объекта формируется обучающая выборка вида: $(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}, y^{(k)})$, $k = \overline{1, K}$, где $x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}, y^{(k)}$ — значения входных переменных $x_1, x_2, \dots, x_n$ и выходной переменной y в k -ом примере; K — общее количество примеров в обучающей выборке.

На основании известных минимальных и максимальных значений каждой входной переменной область определения этих переменных разбивается на отрезки, на каждом из которых задается гауссов ФП.

Далее на этапе структурной адаптации формируется N наборов нечетких правил, в каждом поколении иммунного алгоритма происходит модификация популяции в соответствии с выбранной схемой применения иммунных операторов, вычисление приспособленности каждого антитела, ранжирование антител по степени аффинности в популяции, а также формирование базы правил вывода, на основе которой выполняется параметрическая адаптация. Иммунный алгоритм параметрической адаптации итерационно повторяется заданное число раз. В конце из полученной популяции выбирается наиболее приспособленное антитело, которое является результатом адаптации системы нечеткого вывода.

3. Адаптивные нечеткие нейронные сети

3.1. Архитектура нечетких нейронных сетей

Модели нечеткого вывода могут быть реализованы в виде нечетких сетей, по структуре идентичных многослойным нейронным сетям с прямым

распространением сигнала, элементы каждого слоя которой реализуют отдельный этап нечеткого вывода в нечеткой модели [1]. В настоящее время наиболее распространенной архитектурой является архитектура ANFIS [10].

Используемый класс нечетких нейронных сетей является функционально эквивалентным системе нечеткого вывода. Для устранения недостатков классических систем нечеткого вывода нечеткие нейронные сети создаются адаптивными, с коррекцией параметров нечетких правил и ФП.

Адаптивная нечеткая нейронная сеть с двумя входными переменными, каждая из которых представляется двумя ФП, и четырьмя нечеткими правилами вывода представлена на рис. 1.

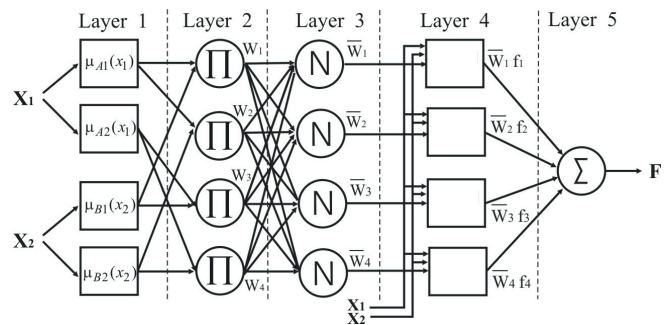


Рис. 1. Архитектура нечеткой нейронной сети

Выходы нейронов первого слоя ($\mu_{A_i}(x_1)$ и $\mu_{B_i}(x_2)$, $i = \overline{1, 2}$) представляют собой значения гауссовых ФП, вычисленных в соответствии с (2). Нейроны второго слоя выполняют умножение входных сигналов и представляют вес правил: $w_i = \mu_{A_i}(x_1) \times \mu_{B_i}(x_2)$, $i = \overline{1, 4}$. В третьем слое i -й нейрон вычисляет отношение веса i -го правила к сумме весов всех правил: $\bar{w}_i = \frac{w_i}{\sum w_i}$, $i = \overline{1, 4}$. Функ-

ция активации нейронов четвертого слоя описывается как $\bar{w}_i f_i = p_i x_1 + q_i x_2 + r_i$, $i = \overline{1, 4}$, где $\{p_i, q_i, r_i\}$ — параметры i -го правила вывода Такаги-Сугено первого порядка. Единственный нейрон пятого слоя вычисляет общий выход как сумму входных сигналов: $F = \sum \bar{w}_i f_i$, $i = \overline{1, 4}$ [10].

Нейроны, обозначенные квадратами, являются адаптивными с настройкой параметров ФП в слое 1 (c и σ для гауссовой ФП (2)) и параметров нечетких правил в слое 4 ($\{p_i, q_i, r_i\}$).

3.2. Иммунный алгоритм адаптации нечетких нейронных сетей

В задаче обучения АННС в качестве антигена выступают примеры обучающей выборки. Каждое антитело кодирует параметры правил нечеткого вывода и ФП входных переменных:

$$Ab = \langle c_{11}, \dots, c_{1m}, \dots, c_{n1}, \dots, c_{nm}; \sigma_{11}, \dots, \sigma_{1m}, \dots, \sigma_{n1}, \dots, \sigma_{nm}; k_{10}, \dots, k_{1n}, \dots, k_{q0}, \dots, k_{qn} \rangle,$$

где c_{ij}, σ_{ij} , $i = \overline{1, n}$; $j = \overline{1, m}$ — параметры гауссовых ФП вида (2) для n входных переменных, каждая из

которых имеет m термов; $k_{i0}, \dots, k_{in}$, $i = \overline{1, q}$; – коэффициенты q нечетких правил.

Такое представление параметров нечеткой нейронной сети в антители позволяет одновременно выполнять настройку как параметров нечетких правил, так и параметров ФП. Популяция антителей генерируется случайным образом, используется вещественное кодирование.

В АННС подставляются параметры, закодированные в антители, и для значений входных переменных, закодированных в антигене, рассчитываются значения выходной переменной. Реализуется режим пакетного обучения, при котором значения параметров изменяются только после предъявления сети всех обучающих образов. Аффинность антителя определяется как среднеквадратическое отклонение реального выхода АННС от желаемого для всех антигенов в соответствии с (1). Результатом работы алгоритма будет антитело с минимальной по популяции аффинностью.

Алгоритм обучения нечеткой нейронной сети можно представить в виде следующей последовательности шагов:

1. Генерация начальной популяции антителей Ab .
 2. Цикл для каждого антителя Ab_i :
 - 2.1 Параметры из антителя Ab_i подставляются в нечеткую нейронную сеть, вычисляется аффинность антителя к популяции антигенов Ag .
 - 2.2. Выбор n лучших антителей для клонирования N раз.
 - 2.3. Клонирование, формирование популяции клонов C .
 - 2.4. Мутация популяции C , формирование популяции C^* .
 - 2.5. Вычисление аффинности популяции C^* к популяции антигенов Ag .
 - 2.6. Редактирование популяции антителей.
 - 2.7. Замена d худших антителей новыми антителями в популяции антителей Ab .
 3. Проверка критерия останова. Если не достигнут – переход к шагу 2, в противном случае – переход к шагу 4.
 4. Конец.
- Результат – антитело с лучшей аффинностью, содержащее параметры правил нечеткого вывода и ФП входных переменных.

4. Экспериментальные исследования

4.1. Параметры иммунных алгоритмов

Параметры иммунных алгоритмов структурной и параметрической адаптации правил нечеткого вывода и обучения нечеткой нейронной сети определены предварительными экспериментами и приведены в табл. 1.

Размер популяции антителей для структурной адаптации системы нечеткого вывода соответствует количеству примеров в обучающей выборке.

Количество антителей для клонирования n влияет на скорость сходимости алгоритма. С увеличением данного параметра скорость сходимости алгоритма увеличивается, однако увеличивается и вычислительная сложность алгоритма.

Используется оператор пропорционального клонирования [5], который клонирует n антителей пропорционально их аффинности:

$$N_c(Ab_i) = \begin{cases} \text{Min_}N_c, & \text{if } \text{Aff}(Ab_i) < \text{Max_Aff} * 0.3 \\ \text{MaxO } N_c, & \text{if } \text{Aff}(Ab_i) < \text{Max_Aff} * 0.7 \\ \text{Aff}(Ab_i) / \beta, & \text{в остальных случаях,} \end{cases} \quad (3)$$

где $\text{Min_}N_c$ и $\text{Max_}N_c$ – минимальное и максимальное количество клонов соответственно; $\text{Aff}(Ab_i)$ – значение аффинности антителя Ab_i ; Max_Aff – лучшее значение аффинности, полученное в текущем поколении; β – параметр, который может изменяться в процессе выполнения поколений.

Вероятность мутации влияет на скорость сходимости алгоритма. Большая вероятность мутации поддерживает разнообразие антителей в иммунной системе на достаточно высоком уровне, делая случайные скачки по всей области поиска, и позволяет находить глобальный экстремум, не застревая в точках локального оптимума, однако большинство изменений могут привести к нефункциональным антителям с плохой аффинностью. Небольшая вероятность мутации позволяет иммунной системе исследовать локальную область вокруг антителя, делая небольшие шаги в поиске антителя с лучшей аффинностью, ведущих к оптимальному значе-

Таблица 1

Параметры иммунного алгоритма

Параметр	Адаптивная система нечеткого вывода		Адаптивная нечеткая нейронная сеть
	Структурная адаптация	Параметрическая адаптация	
Размер популяции	121	15	15
Количество антителей для клонирования	30	7	7
Минимальное количество клонов	5	5	5
Максимальное количество клонов	7	7	7
Максимальная вероятность мутации	0,3	0,3	0,3
Количество антителей для редактирования	10	1	1
Коэффициент суппрессии (порог сжатия сети)	0,001	-	-

нию, но скорость сходимости алгоритма уменьшится. Следовательно, механизм мутации необходимо регулировать в процессе работы иммунного алгоритма и для антител с улучшенной в результате мутации аффинностью необходимо уменьшать вероятность мутации в последующих поколениях работы иммунного алгоритма, чтобы не потерять полученные преимущества. Используется оператор обратно-пропорциональной мутации [5], согласно которому вероятность мутации обратно пропорциональна аффинности антитела:

$$p_m(Ab_i) = c * \frac{Aff(Ab_i) - Max_Aff}{Min_Aff - Max_Aff}, \quad (4)$$

где c — максимальная вероятность мутации i -го клона; $Aff(Ab_i)$ — значение аффинности антитела Ab_i ; Max_Aff — лучшее значение аффинности, полученное в текущем поколении; Min_Aff — худшее значение аффинности, полученное в текущем поколении;

При редактировании популяции антитела с худшей по популяции аффинностью заменяются случайно сгенерированными новыми антителами, что обеспечивает поддержку разнообразия в популяции.

Оператор супрессии удаляет похожие антитела, для того чтобы избежать избыточности и поддерживать разнообразие в популяции. Согласно теории иммунной сети оператор супрессии обеспечивает сжатие популяции путем связывания антител в зависимости от значения их аффинности относительно других антител популяции. Антитела с минимальной аффинностью могут быть удалены из популяции в соответствии с установленным порогом сжатия сети.

Критерием останова является достижение заданного количества поколений работы алгоритма.

4.2. Результаты экспериментов

Для оценки эффективности предложенных алгоритмов были проведены эксперименты с использованием тестовых функций, приведенных в табл. 2.

Отметим, что тестовая функция 1 является унимодальной, а тестовая функция 2 — мультимодальной. Обучающая выборка для тестовых функций сформирована в соответствии с диапазоном входных переменных и составляет 121 пример. По данным из обучающей выборки каждая тестовая функция восстанавливается с помощью адаптивной системы нечеткого вывода и адаптивной

нечеткой нейронной сети. Начальные значения параметров ФП устанавливаются таким образом, чтобы ФП были равномерно распределены по области определения функций. Эксперименты проводились с использованием разного количества лингвистических термов для каждой из двух входных переменных.

В табл. 3 приведены данные по среднеквадратическому отклонению по результатам 10 запусков алгоритма для тестовых функций 1 и 2.

Таблица 3

Результаты экспериментов

Количество лингвистических термов	Тестовая функция 1	Тестовая функция 2
Система нечеткого вывода		
3	0,2	0,69
5	0,051	0,082
7	0,012	0,05
Нечеткая нейронная сеть		
3	0,27	0,501
5	0,035	0,057
7	0,024	0,047

Из табл. 3 видно, что использование большего количества ФП для каждой переменной позволяет повысить точность восстановления функции.

Восстановленные поверхности и полученные ФП представлены на рис. 2 и 3. Алгоритм адаптации нечеткой нейронной сети позволяет достичь лучших результатов за счет одновременной настройки как параметров нечетких правил, так и параметров ФП. Однако при адаптации нечеткой нейронной сети структура базы правил задается строго фиксированной, так как при изменении количества правил необходимо изменять структуру нейронной сети.

При адаптации системы нечеткого вывода изменяются не только коэффициенты нечетких правил, но и количество правил нечеткого вывода за счет выполнения супрессии. Так, в адаптивной нечеткой нейронной сети количество правил для двух входных переменных с пятью ФП составляет 25 для обеих тестовых функций 1–2, в то время как для системы нечеткого вывода количество правил сокращается до 19 для тестовой функции 1 и — до 23 для тестовой функции 2.

Следует отметить, что процесс адаптации не-

Таблица 2 четкой нейронной сети требует значительных временных затрат на вычисление, что связано с длиной строки оптимизируемых параметров, представленных в антителе. Так, в случае, когда для лингвистической оценки входных переменных используется 3 терма, длина строки антитела составляет 39 параметров, в случае 5 термов — 95 параметров, в случае 7 термов — 175 параметров.

Тестовые функции

	Тестовая функция	Область определения функции
1	$F_1(x_1, x_2) = x_1^2 + x_2^2$	$x_1, x_2 \in [-3, 3]$
2	$F_2(x_1, x_2) = 3(1 - x_1)^2 \exp(-x_1^2 - (x_2 + 1)^2) - 10\left(\frac{x_1}{5} - x_1^3 - x_2^5\right) \exp(-x_1^2 - x_2^2) - \frac{1}{3} \exp(-(x_1 + 1)^2 - x_2^2)$	$x_1, x_2 \in [-3, 3]$

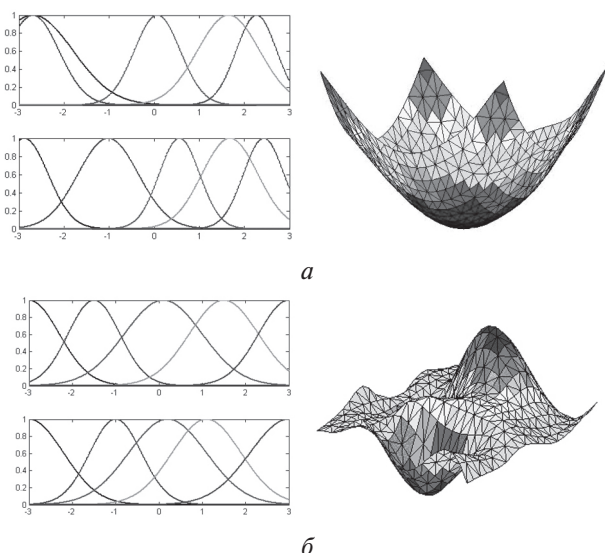


Рис. 2. Адаптивная система нечеткого вывода: настроенные ФП и восстановленная поверхность для тестовой функции 1 (а) и тестовой функции 2 (б)

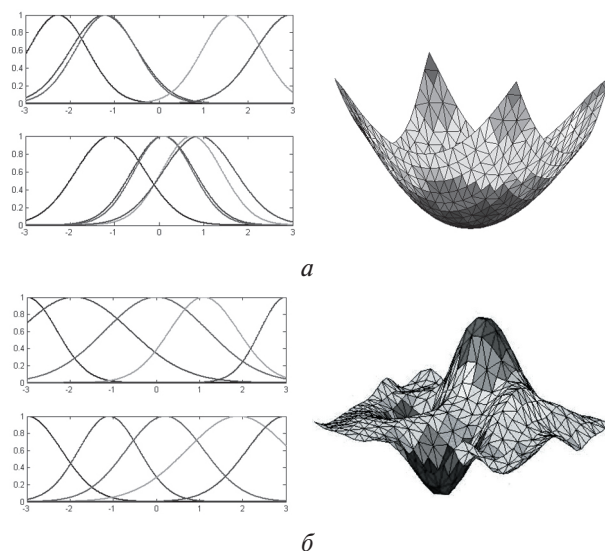


Рис. 3. Адаптивная нечеткая нейронная сеть: настроенные ФП и восстановленная поверхность для тестовой функции 1 (а) и тестовой функции 2 (б)

Выводы

В работе рассмотрено решение задачи идентификации нелинейных объектов на основе использования искусственных иммунных систем. Предложены адаптивные нечеткие модели идентификации объектов, реализованные в виде системы нечеткого вывода и нечеткой нейронной сети с иммунной настройкой. В адаптивной системе нечеткого вывода реализованы процедуры структурной и параметрической оптимизации. В адаптивной нечеткой нейронной сети реализована одновременная параметрическая адаптация ФП и правил нечеткого вывода. Проведенные экспериментальные исследования на двух тестовых функциях показали эффективность применения искусственных иммунных систем для обучения адаптивных систем нечеткого вывода и нечетких нейронных сетей, описывающих нелинейные функции.

Список литературы: 1. Борисов В.В., Круглов В.В., Федюлов Ф.С. Нечеткие модели и сети. — М: Горячая линия. — Телеком, 2007. — 284 с. 2. Петров Е.Г., Новожилова М.В., Гребенник И.В. Методи і засоби прийняття рішень у соціально-економічних системах: Навч. посібн. / За ред. Е.Г. Петрова. — К., Техніка, 2004. — 256с. 3. Castro, P.A.D.; Camargo, H.A. «Learning and Optimization of Fuzzy Rule Base by means of Self Adaptive Genetic Algorithms», Proceedings of IEEE International Conference on Fuzzy Systems, Vol. 2, pp. 1037-1042. 4. Timmis J.I., Knight T., De Castro L.N., Hart E. An Overview of Artificial Immune Systems // Computation in Cells and Tissues: Perspectives and Tools for Thought, Natural Computation Series, Springer, 2004 — p. 51-86. 5. Castro, P.A.D.; Coelho, G. P.; Caetano, M. F.; Von Zuben, F. J. «Designing Ensembles of Fuzzy Classification Systems: An Immune-Inspired Approach», Proceedings of the 4th International Conference on Artificial Immune Systems (ICARIS), Lecture Notes in Computer Science, Springer-Verlag, Berlin, vol. 3627, pp. 469-482, 2005. 6. De Castro, L., Von Zuben, F.J.: Learning and optimization using the clonal selection principle. IEEE Transactions on Evolutionary Computation, Special Issue on Artificial Immune Systems 6. 239-251, 2002. 7. De Castro L. N., Von Zuben F. J. The Clonal Selection Algorithm with Engineering Applications // In Proceedings of GECCO'00, Workshop on Artificial Immune Systems and Their Applications, 2000. — p. 36-37. 8. L.N. de Castro and F.J. Von Zuben, «aiNet: an artificial immune network for data analysis,» in Data Mining: A Heuristic Approach, Chapter XII, H.A. Abbass, R.A. Sarker, and C.S. Newton, Eds. USA: Idea Group Publishing, 2001, pp.231-259. 9. Korablyov M.M., Ovcharenko I.V. Adaptation of fuzzy inference models using artificial immune systems // Advanced Computer Systems and Networks: Design and Application: Proceedings of the 3-rd International Conference ACSN-2007, Lviv. P. 89-91. 10. Jang J.-S. R. ANFIS: Adaptive-Neuro-Based Fuzzy Inference System // IEEE Trans. Systems & Cybernetics. - 1993. - Vol. 23. - P. 665 - 685. - 1991.

Поступила в редакцию 8.10.2008

УДК 004.89

Адаптивні нечіткі моделі ідентифікації нелінійних об'єктів на основі штучних імунних систем / М.М. Корабльов, І.В.Сорокіна // Біоніка інтелекту: наук.-техн. журнал. — 2008. — № 2 (69). — С. 125-131.

В роботі розглянуто задачу адаптації нечітких моделей ідентифікації нелінійних об'єктів за допомогою штучних імунних систем. Нечіткі моделі представлені у вигляді системи нечіткого виводу та нечіткої нейронної мережі. Процес адаптації полягає в оптимізації форми та параметрів функцій належності та параметрів і структури бази нечітких правил. Показано ефективність запропонованих імунних алгоритмів адаптації моделей нечіткого виводу.

Таб. 3. Лл.: 3. Бібліогр.: 10 найм.

UDK 004.89

Adaptive fuzzy models of nonlinear objects identification based on artificial immune systems / N.M. Korablyov, I.V. Sorokina // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 125-131.

The given work considers the problem of adaptation of fuzzy models of nonlinear objects identification using artificial immune systems. Fuzzy models are represented in the form of fuzzy inference systems and fuzzy neural networks. The process of adaptation consists in adjustment of form and parameters of membership functions and parameters and structure of fuzzy rule base. The efficiency of proposed immune algorithms of fuzzy models adaptation is shown.

Tab. 3. Fig.: 3. Ref.: 10 items.

УДК 1612.841/.844:611.84:620.171.51.001.5:616-07



ИССЛЕДОВАНИЕ И МОДЕЛИРОВАНИЕ ПОЛЯРИЗАЦИОННО-ОПТИЧЕСКИХ СВОЙСТВ РОГОВИЦЫ ГЛАЗА ПРИ РАЗЛИЧНЫХ СОСТОЯНИЯХ ЭКСТРАОКУЛЯРНЫХ МЫШЦ

М.Л. Кочина¹, В.Г. Калиманов²^{1,2} Харьковский национальный медицинский университет, г.Харьков, Украина, m_kochina@yahoo.com

Представлена модель напряженно-деформированного состояния глаза и результаты расчета внутренних напряжений, возникающих в роговице под действием экстраокулярных мышц и внутриглазного давления. Рассмотрены различные варианты изменения интерференционных картин глаз при патологии экстраокулярных мышц.

ПОЛЯРИЗОВАННЫЙ СВЕТ, ВНУТРЕННИЕ НАПРЯЖЕНИЯ, КОСОГЛАЗИЕ, ЭКСТРАОКУЛЯРНЫЕ МЫШЦЫ, ИНТЕРФЕРЕНЦИОННАЯ КАРТИНА

Введение

Роговица живого глаза обладает оптической анизотропией, обусловленной целым рядом факторов. Это состав и строение вещества роговицы, ее форма, механические воздействия со стороны экстраокулярных мышц (ЭОМ) и внутриглазного давления (ВГД) [1,2,3]. При освещении роговицы живого глаза поляризованным светом на ней наблюдается интерференционная картина, состоящая из темного креста по центру (изоклина) и фигуры в форме ромба по периферии. Интерференционный ромб образован набором цветных полос, называемых изохромами. На основании большого количества исследований установлено, что цветовой состав картины изохром обусловлен пигментацией радужной оболочки глаза. На рис. 1 представлены фотография и схема интерференционной картины здорового глаза.

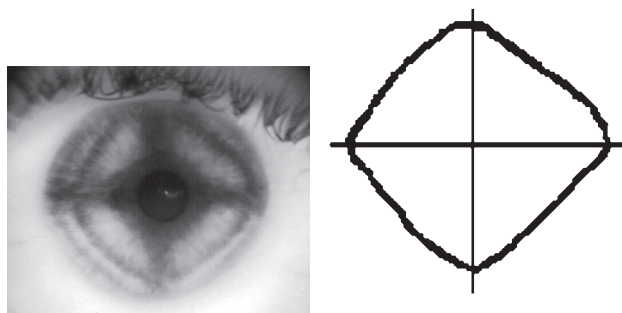


Рис.1. Фотография и схема интерференционной картины нормального глаза

Использование поляризованного света позволяет исследовать физическую природу оптической анизотропии роговицы живого глаза и установить вклад каждого из действующих факторов в формирование интерференционной картины, наблюдаемой на ней. Внутриглазное давление растягивает роговицу, действуя на каждую ее точку с одинаковой силой. Однако внутренние напряжения, возникающие в ней в разных точках, будут различны, что обусловлено градиентом толщины роговицы (в центре она тоньше, чем по периферии). Усилия прямых ЭОМ, приложенные к склере в области

главных меридианов, растягивают роговицу, также создавая в ней внутренние напряжения. В каждой точке роговицы возникают суммарные напряжения, обусловленные суперпозицией напряжений от действия ЭОМ и ВГД. Усилия косых ЭОМ приложены за экватором глаза и вклад в напряженное состояние роговицы не вносят.

Достаточно распространенной патологией детского возраста является косоглазие [4]. Термином «косоглазие» объединяют различные по происхождению и топике поражения зрительной и глазодвигательной систем, вызывающие периодическое или постоянное отклонение зрительной оси одного из глаз от точки фиксации. По данным разных авторов явным косоглазием страдают 1,5–2,0 % детей. В структуре детской глазной патологии косоглазие занимает второе место после аномалий рефракции, удельный вес его составляет 25,62% случаев, частота – 3,92 на 1000 [4,5].

Косоглазие может быть явным и скрытым (гетерофория). Явное косоглазие проявляется отклонением одного или обоих глаз от точки фиксации. Скрытое косоглазие в нормальных условиях не проявляется, глаза стоят симметрично, однако, оно может проявиться, например, при резком снижении зрения одного из глаз (заболевания или травма) и при значительной зрительной нагрузке. На рис. 2 представлена упрощенная классификация видов косоглазия в зависимости от причин его возникновения.

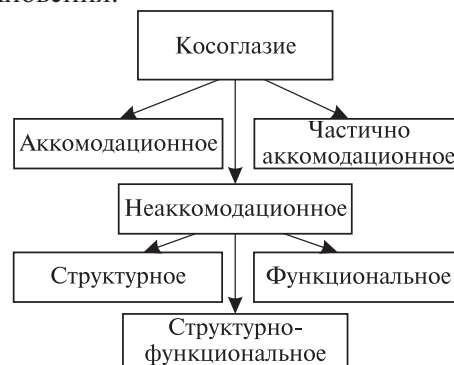


Рис. 2. Упрощенная классификация видов косоглазия в зависимости от причин его возникновения

Аккомодационное косоглазие обусловлено аномалиями рефракции и может быть устранено с помощью оптимальной очковой коррекции. Частично аккомодационное косоглазие, кроме рефракционного, может иметь и мышечный компонент, поэтому, во многих случаях, коррекцией полностью не устраняется. Аккомодационное косоглазие появляется обычно у детей не ранее, чем в 2-3 летнем возрасте. При одностороннем аккомодационном косоглазии имеется большая дальнорезкость или анизометропия, при альтернирующем — обычно высокая гиперметропия, почти одинаковая на обоих глазах, достаточно часто встречается гиперметропический астигматизм. Хирургическое лечение не показано [4, 5].

Наибольший интерес для нас представляет неаккомодационное косоглазие, поскольку причины его возникновения связаны с состоянием ЭОМ. Причиной неаккомодационного косоглазия и гетерофории является патология ЭОМ, которая может быть обусловлена структурными, функциональными или структурно-функциональными причинами. При неаккомодационном косоглазии отклонение глаза не исчезает под влиянием атропинизации и постоянного ношения очков. Косоглазие появляется рано: или сразу после рождения ребенка, или в первый год жизни. Аномалия рефракции мало выражена, обычно это слабая гиперметропия или эметропия. Косоглазие этого типа плохо поддается консервативному лечению. Исследования ряда авторов показали, что аккомодационное косоглазие на первом году жизни возникло у 7% детей, тогда как в более старшем возрасте оно составило 26,1% — 29,3% от общего количества косоглазия, пик возникновения аккомодационного косоглазия приходится на 3-й — 4-й год жизни, тогда как неаккомодационное косоглазие, возникшее сразу после рождения и на первом году жизни, составляет по данным разных авторов от 42,8% до 65% от общего количества больных [4, 5].

Структурное косоглазие обусловлено аномалиями прикрепления ЭОМ. Эти аномалии могут проявляться смещением места прикрепления мышцы или группы мышц как вдоль линии их действия (что эквивалентно усилению или ослаблению силы действия мышц), так и в сторону от линии действия. Во втором случае мышцы вертикального действия могут вызывать отклонения глазного яблока по горизонтали, а горизонтального — по вертикали. Диагностика поражения мышц в случае структурного косоглазия представляет значительную сложность, поскольку в соответствии с существующими методами врач исследует направления отклонения глаз от ортофорического положения и определяет тактику хирургического лечения в соответствии с существующими стандартными представлениями. Если глаз отклоняется по гори-

зонтали, то проводят хирургическую коррекцию мышц горизонтального действия, не учитывая возможность более сложного биомеханического соотношения усилий. Обычно такое хирургическое лечение малоэффективно, поскольку не устраняется реальная причина отклонения глаза.

Функциональное косоглазие обусловлено усилением или ослаблением силы действия мышцы или группы мышц, обычно оно проявляется отклонением глаза в сторону более сильной мышцы.

Комбинированным является структурно-функциональное косоглазие, обусловленное как смещением места приложения усилий мышц, так и неравновесным их действием.

Поляризационно-оптический метод, основанный на исследовании формы интерференционных картин на роговице глаза, позволяет дифференцировать причины возникновения отклонений глаз в каждом конкретном случае. Действие ЭОМ вносит наиболее существенный вклад в формирование интерференционной картины именно в виде ромба. Эти мышцы, удерживая глазное яблоко в определенном положении, всегда находятся в тонусе. Многочисленные исследования формы интерференционных картин в норме и при различной патологии экстраокулярных мышц позволили выявить закономерности их изменения и разработать объективные бесконтактные способы диагностики [6-9].

Так, установлено, что ослабление усилия со стороны одной из мышц приводит к скруглению соответствующего ей угла интерференционного ромба и уменьшению длины соответствующего ей отрезка диагонали. При увеличении мышечного усилия происходит вытяжение соответствующего угла и увеличение длины отрезка диагонали [6]. Смещение места прикрепления мышцы вдоль линии ее действия эквивалентно изменению усилия, а смещение в сторону от линии действия приводит к искажению фигуры ромба и смещению углов ромба от соответствующих меридианов.

При изменении тонуса глазодвигательных мышц, обусловленного конвергенцией (сведением зрительных осей, обычно при рассматривании близко расположенных предметов), в процессе напряженного зрительного труда, выполняемого длительное время, форма интерференционного ромба также изменяется, причем эти изменения сходны с теми, которые наблюдаются при изменении тонуса глазодвигательных мышц при косоглазии [9].

Таким образом, поляризационно-оптический метод может быть использован для исследования структурно-функционального состояния глазодвигательных мышц как в задачах диагностики, так и при оценке участия ЭОМ в решении зрительных задач. Однако получить на практике и соответствующим образом классифицировать интерференци-

онные картины глаз при всех вариантах патологии ЭОМ достаточно сложно, поэтому нами проведено моделирование распределения напряжений в роговице при разных сочетаниях их силы действия и места прикрепления.

Целью работы является исследование с помощью модели напряженно-деформированного состояния глаза влияния структурно-функционального состояния глазодвигательных мышц на распределение внутренних напряжений в роговице.

1. Модель напряженно-деформированного состояния глаза

Для моделирования различных вариантов состояния ЭОМ была использована модель напряженно-деформированного состояния глаза [2, 9]. На рис. 3 представлена схема сферической оболочки, к которой приложены 4 сосредоточенных усилия Q_i со стороны прямых ЭОМ.

Для моделирования напряженно-деформированного состояния глазного яблока были сделаны следующие упрощения, облегчающие расчеты:

1. Считаем, что глазное яблоко представляет собой тонкую сферическую оболочку, заполненную жидкостью. Так как мы исследуем распределение внутренних напряжений только в роговице, то считаем, что радиус кривизны глазного яблока равен среднему радиусу кривизны роговицы. Оболочку глаза можно считать тонкой, так как отношение толщины ее стенки к радиусу кривизны величина малая. Градиентом напряжений по толщине оболочки, исходя из условия ее тонкости, пренебрегаем. Для упрощения расчетов считаем оболочку сферической.

2. Считаем, что упругие свойства оболочки одинаковые.

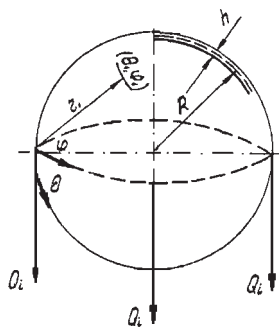


Рис. 3 Схема сферической оболочки с приложенными усилиями ЭОМ

3. Считаем, что оболочка заполнена однородной жидкостью (не учитываем сложное внутреннее строение глазного яблока).

4. При расчетах учитываем только градиент толщины роговой оболочки, остальную же часть глазного яблока считаем однородной по толщине.

5. Считаем, что на роговую оболочку действуют только наружные прямые мышцы, действие косых мышц не учитываем.

При проведении моделирования было учтено реальное место прикрепления каждой из наружных прямых мышц и проведены исследования всех возможных вариантов их анатомических изменений.

Оболочка глаза в отсутствии действия ЭОМ растянута давлением внутриглазной жидкости, которая, по закону Паскаля, оказывает равное давление на все точки оболочки, поэтому на каждый ее элемент будут действовать растягивающие усилия, вызывающие появление напряжений, которые определяются по формуле:

$$\sigma_p = \frac{PR}{2h},$$

где σ_p — напряжения в оболочке; P — давление внутриглазной жидкости; R — радиус кривизны оболочки; h — ее толщина.

Каждая сила Q_i оказывает растягивающее действие на роговую оболочку и сжимает склеру, причем при определенных условиях (например, при конвергенции) еще и деформирует склеру.

Все формулы представлены в сферических координатах θ и ϕ , радиус-вектор r отсчитывается от точки приложения силы. Так как воздействия глазодвигательных мышц симметричны, угол θ изменяется от 0 до $\pi/2$, угол ϕ — от 0 до π . Квадрат радиус-вектора r определяется по формуле:

$$r^2 = R^2(\theta^2 + \sin^2 \theta \phi^2).$$

Каждая сила Q_i создает в сферической оболочке нормальные напряжения $\sigma_i = \frac{N_i}{h}$ и напряжения смещения $\tau_i = \frac{T_i}{h}$, где N_i — нормальные усилия, T_i — усилия смещения, h — толщина оболочки.

Усилия в оболочке определяются по следующим формулам:

$$N_\theta = Q_i \frac{R\theta}{4\pi r^2} \left[2(1+\nu) \frac{R^2 \gamma^2 \sin^2 \theta}{r^2} - 3 - \nu \right],$$

$$N_\gamma = -Q_i \frac{R\theta}{4\pi r^2} \left[2(1+\nu) \frac{R^2 \gamma^2 \sin^2 \theta}{r^2} + 1 - \nu \right],$$

$$T_\theta = T_\gamma = -Q_i \frac{R\gamma}{4\pi r^2} \left[2(1+\nu) \frac{R^2 \theta^2}{r^2} + 1 - \nu \right].$$

В приведенных далее формулах ν — коэффициент Пуассона, характеризующий упругие свойства оболочки.

В каждой точке оболочки можно определить эквивалентные напряжения с учетом только нормальных напряжений $\sigma'_3 = \sqrt{\sigma_\theta^2 + \sigma_\gamma^2 - \sigma_\theta \sigma_\gamma}$ и эквивалентные напряжения с учетом напряжений сдвига $\sigma_3 = \sqrt{(\sigma'_3)^2 + 3\tau^2}$. Поскольку внутриглазная жидкость создает в оболочке напряжения σ_p , то в каждой точке оболочки существуют суммарные на-

пряжения, определяемые по формуле $\sigma = \sigma_{\theta} + \sigma_p$, и главные напряжения, определяемые по формулам:

$$\sigma_1 = \frac{1}{2} \left[\sigma_{\theta} + \sigma_{\gamma} + \sqrt{(\sigma_{\theta} - \sigma_{\gamma})^2 + 4\tau^2} \right] + \sigma_p,$$

$$\sigma_2 = \frac{1}{2} \left[\sigma_{\theta} + \sigma_{\gamma} - \sqrt{(\sigma_{\theta} - \sigma_{\gamma})^2 + 4\tau^2} \right] + \sigma_p,$$

где σ_1 — максимальные напряжения; σ_2 — минимальные.

Приведенные формулы являются асимптотическими, они описывают напряженное состояние оболочки в окрестности точки приложения силы Q_i . К главному яблоку приложены четыре такие силы, потому результирующие напряжения в каждой точке оболочки определяют как суперпозицию напряжений, создаваемых каждой силой.

2. Результаты и их обсуждение

При проведении расчета распределения напряжений в оболочке мы исходили из того, что вблизи точки приложения силы напряжения большие, но они существенно уменьшаются с расстоянием. Результирующие напряжения в оболочке рассчитывались путем сложения значений напряжений от каждой силы. Поскольку линии равных напряжений (изохром) наблюдаются на фоне радужной оболочки глаза, то есть как бы спроектированы на плоскость, то расчетные кривые строились в проекции на плоскость, что осуществлялось на основе геометрических представлений. При проведении расчетов были использованы значения параметров, соответствующие их значениям для реальных глаз.

В качестве иллюстрации осуществления метода расчета с использованием модели на рис. 4 представлены фотография и схема интерференционной картины глаза с косоглазием, обусловленным смещением места прикрепления внутренней прямой мышцы в сторону лимба и ослаблением силы действия наружной прямой мышцы.

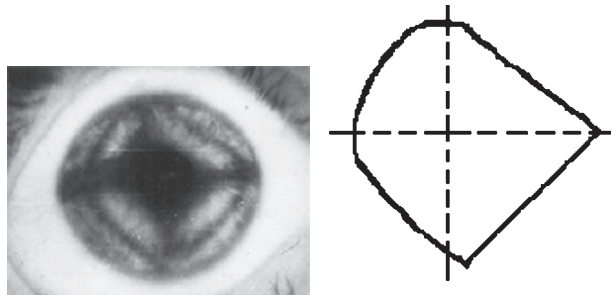


Рис. 4. Фотография и схема интерференционной картины глаза с косоглазием

На рис. 5 представлены расчетные изохромы для разных случаев смещения места прикрепления внутренней прямой мышцы.

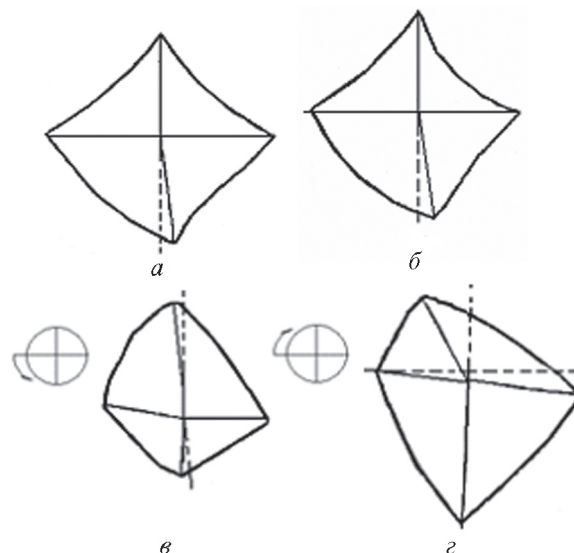


Рис. 5. Расчетные изохромы для случаев смещения места прикрепления внутренней прямой мышцы: вдоль линии действия к лимбу (а), к экватору (б), вниз от меридиана (в) и вверх от меридиана (г)

Расчетные изохромы по форме сходны с экспериментальными и могут быть использованы для диагностики структурно-функциональной патологии ЭОМ.

Выводы

1. Модель напряженно-деформированного состояния глаза может быть использована для расчета линий равных напряжений, поскольку форма расчетных изохром хорошо согласуется с экспериментальными данными.

2. Расчет линий равных напряжений при всех возможных вариантах дислокации ЭОМ позволит создать набор эталонов интерференционных картин, пригодных для диагностических целей.

Список литературы: 1. Пеньков М.А., Кочина М.Л. Интерференционный метод в диагностике косоглазия // Офтальмо-логический журнал. — 1979. — №8. — С. 497-498. 2. Пеньков М.А., Алтухер Г.М., Кочина М.Л. Расчет изохром и изоклин роговой оболочки глаза // Биофизика. — Вып.2. — 1982. — Т. XXVIII. — С. 313-316. 3. Пеньков М.А., Кочина М.Л., Уткин Ю.А. Анализ интерференционных изохром роговой оболочки глаза и выделение диагностических признаков // Проблемы бионики. — 1983. — Вып.30. — С. 22-28. 4. Аветисов Э.С. Содружественное косоглазие. — М.: Медицина, 1977. — 312 с. 5. Научно-практический анализ результатов хирургического этапа лечения детей с врожденным косоглазием / Е.И. Ковалевский, М.Р. Гусева, М.И. Грязнов, А.А. Холмский // Офтальмологический журнал. — 1979. — №5. — С. 272-275. 6. Кочина М.Л. Некоторые результаты моделирования напряженного состояния роговой оболочки глаза // Кибернетика и вычислительная техника. — 1991. — Вып.90. — С. 97-99. 7. Результаты моделирования врожденной патологии экстраокулярных мышц / М.А. Пеньков, С.Ф. Зубарев, М.Л. Кочина, Т.А. Босенко // Офтальмологический журнал. — 1990.

— N4. — С. 197-200. 8. *Босенко Т.О.* Диагностика асиметрії зовнішніх м'язів ока у поляризованому світлі при різних видах співдружної косоокості: Автореф. дис. на здобуття наук. ступ. канд. мед. наук. (14.00.08). — Одеса, 1994. — 16 с. 9. *Кочина М.Л., Калиманов В.Г.* Модель напружено-деформованого состояния роговицы глаза // Матер. І укр. наукової конф. «Проблеми біологічної і медичної фізики». 20-22 вересня 2004 р., Харків. — С. 146.

Поступила в редколлегию 10.10.2008

УДК: 1612.841/.844:611.84:620.171.51.001.5:616-07

Дослідження та моделювання поляризаційно-оптичних властивостей рогівки ока при різних станах екстраокулярних м'язів / М.Л. Кочина, В.Г.Каліманов // // Біоніка інтелекту: наук.-техн. журнал — 2008. — № 2 (69). — С. 132-136.

Наведена модель напружено-деформованого стану ока та результати розрахунків внутрішніх напружень, які виникають в рогівці під впливом екстраокулярних м'язів і внутрішньоочного тиску. Розглянути різні варіанти

змін інтерференційних картин при патології екстраокулярних м'язів.

Лл. 5. Бібліогр.: 9 найм.

UDK: 1612.841/.844:611.84:620.171.51.001.5:616-07

Research and modeling of polarization optical properties eye's cornea under various states of oculomotor muscles / M.L. Kochina, V.G.Kalimanov // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 132-136.

The optical anisotropy of alive eye's cornea is caused by properties of corneal collagen, interposition of its fibres in a tissue of a cornea, level of intraocular pressure and a structurally functional condition of oculomotor muscles. The interferential picture in a form of rhombus which parameters depend on force of action and a place of attachment of oculomotor muscles is observed in polarized light on a cornea in a usual self. A method of calculation tensions in the eye was described. These tensions were provoked by mechanical influence of oculomotor muscles and intraocular pressure. The method is suitable for calculating tensions for eyes with different pathology oculomotor muscles.

Fig. 5. Ref.: 9 items.

УДК 519.6:004.93



МЕТОД ОПТИМИЗАЦИИ НА ОСНОВЕ МОДЕЛИРОВАНИЯ ПЕРЕМЕЩЕНИЯ БАКТЕРИЙ С ПРИМЕНЕНИЕМ ЭВОЛЮЦИОННЫХ ОПЕРАТОРОВ

А.А. Олейник¹, С.А. Субботин²¹ Запорожский национальный технический университет, г. Запорожье, Украина, olejnik@zntu.edu.ua² Запорожский национальный технический университет, г. Запорожье, Украина, subbotin@zntu.edu.ua

Исследуется метод оптимизации на основе моделирования перемещения бактерий. Разработано расширение метода оптимизации на основе моделирования бактерий, учитывающее эволюционные процессы, протекающие в популяции бактерий. Предложено применение разработанного метода оптимизации для отбора информативных признаков.

ОПТИМИЗАЦИЯ, АГЕНТ, КОЛЕКТИВНЫЙ ИНТЕЛЛЕКТ, МЕТОД ОПТИМИЗАЦИИ
НА ОСНОВЕ МОДЕЛИРОВАНИЯ ПЕРЕМЕЩЕНИЯ БАКТЕРИЙ, ВЫБОРКА, ПРИЗНАК

Введение

В настоящее время активно развиваются направления искусственного интеллекта, основные идеи которых взяты из природы. К таким направлениям можно отнести: искусственные нейронные сети [1, 2], методы эволюционного поиска [3, 4], методы оптимизации, основанные на моделировании коллективного интеллекта общественных насекомых [5, 6], искусственные иммунные системы [7].

Одними из наиболее новых направлений являются мультиагентные методы интеллектуальной оптимизации. Они основаны на моделировании поведения общественных живых существ, таких, как птицы, рыбы, муравьи, пчёлы, бактерии. Таким образом, разработаны соответствующие методы оптимизации: метод оптимизации с использованием роя частиц (Particle Swarm Optimization, PSO) [8], метод муравьиных колоний (Ant Colony Optimization, ACO) [9], метод пчелиной колонии (Bee Colony Optimization, BCO) [10], метод оптимизации на основе моделирования перемещения бактерий (Bacteria Foraging Optimization, BFO) [11]. Данные методы показали хорошие результаты при решении различных практических задач оптимизации [8–11], засвидетельствовав тем самым эффективность своего использования.

Однако, поскольку методы основаны на поведении живых существ, то модели поведения живых существ могут обладать неполнотой или могут не полностью соответствовать своим природным аналогам.

В связи с этим актуальной является разработка дополнений существующих моделей, использующихся для моделирования поведения живых существ, которые обеспечивают более полное соответствие природным аналогам и лучшую эффективность работы методов оптимизации.

1. Постановка задачи

Метод BFO эффективно применяется для решения задач оптимизации [11]. Однако математические модели, разработанные для данного метода,

даже по мнению их разработчиков [11] являются не полными. Поэтому необходимо доработать известные математические модели BFO для более адекватного моделирования эволюционных процессов жизнедеятельности бактерий.

Целью данной работы является расширение математической модели метода BFO, учитывающей эволюционные процессы в популяции бактерий.

Основными задачами работы являются:

- исследование основных принципов поведения бактерий, которые должны учитываться при моделировании их поведения;
- анализ метода оптимизации на основе моделирования перемещения бактерий;
- включение операторов эволюционного поиска в уже существующую математическую модель, применяемую в метод оптимизации на основе моделирования перемещения бактерий.
- применение разработанного метода оптимизации для отбора информативных признаков;
- разработка программного обеспечения, реализующего предложенную математическую модель, и его тестирование.

2. Биологические основы поведения бактерий

Одной из наиболее изученных бактерий является бактерия *E. Coli* [12], которая живёт в кишечнике большинства млекопитающих (в том числе и в кишечнике человека). Бактерия *E. Coli* при соответствующих условиях может самовоспроизводиться (расщепляться) за 20 мин. Способность двигаться бактерии *E. Coli* обеспечивается за счёт шести или более жгутиков, которые вращаются с частотой 100–200 оборотов в секунду, каждый из которых управляется с помощью своего собственного биологического «мотора». Бактерия *E. Coli* чередует свои действия: то перемещается прямолинейно (со скоростью 10–20 мкм/сек.), то кувыркается (изменение направления). Когда жгутики вращаются по часовой стрелке (или против часовой стрелки), они работают как пропеллеры и, следовательно, *E. Coli* может плыть или кувыркаться.

Примерная схема движения бактерии выглядит следующим образом: когда «мотор» работает в одну сторону, все жгутики у бактерии складываются и вращаются вместе, и бактерия движется прямолинейно. В конце пробега бактерия останавливается, «мотор» переключается и начинает работать в другую сторону. Жгутики растопыряются и «бултыхаются» независимо друг от друга. Бактерия при этом переориентируется в пространстве случайным образом. После этого «мотор» опять переключается и начинает работать в ту сторону, в которую жгутики работают вместе, и возникает следующий отрезок прямолинейного движения.

Хемотаксис — это двигательная реакция бактерии в ответ на появление в среде аттрактанта (аттрактант — вещество, привлекающее бактерии) или репеллента (репеллент — вещество, отпугивающее бактерий) [13]. В естественных условиях аттрактантами являются вещества, полезные для бактерий, а репеллентами — те, которые бактериям вредны.

При наличии пространственных изменений концентрации аттрактантов или репеллентов частота кувырканий, а, следовательно, и длина свободного пробега бактерии изменяются. Длина свободного перемещения бактерии, плывущей в сторону возрастающей концентрации аттрактанта, увеличивается, а при движении в сторону возрастающей концентрации репеллента уменьшается.

Таким образом, можно выделить следующие хемотаксические действия бактерии *E. Coli* [13]:

- если бактерия находится в нейтральной среде, то чередуются кувырки и передвижения, за счёт чего выполняется поиск;

- если бактерия перемещается по градиенту аттрактанта, то перемещение продолжается в этом же направлении. Таким образом, обеспечивается поиск более благоприятной окружающей среды;

- если происходит перемещение в направлении, противоположном градиенту репеллента, то обеспечивается избегание неблагоприятной окружающей среды.

Таким образом, бактерия может перемещаться в областях с полезными веществами и в то же время избегать опасных веществ. Сенсоры, используемые *E. Coli* — белковые рецепторы, которые являются очень чувствительными (незначительное изменение в концентрации полезных веществ может вызвать существенное изменение в поведении бактерии). От рецепторов сигналы поступают на метилакцептирующие белки, которые собирают все сигналы от рецепторов, и результирующий сигнал выходит на «мотор» жгутика, который управляет движением бактерии в зависимости от соотношения полезных и опасных веществ в окружающей среде.

Бактерии часто умирают или растворяются, и это должно учитываться при моделировании их

деятельности. Мутации у *E. coli* происходят в нормальных условиях приблизительно 10^{-7} в ген и в поколение.

E. Coli могут формировать сложные устойчивые пространственно-временные структуры в некоторых полутвердых полезных средах. Бактерии могут поглощать питательные вещества по их радиусу, начиная от внешней границы, заканчивая серединой, даже в случае, если бактерии были изначально размещены в центре питательных веществ. Кроме того, при определённых условиях, они могут скрывать притягивающие сигналы от клетки к клетке, за счёт чего они группируются и защищают друг друга. Таким образом, эти бактерии могут группироваться в колонии.

3. Метод оптимизации на основе моделирования перемещения бактерий

Данный метод предназначен для нахождения минимума функции $J(X)$, $X \in R^p$ при неизвестном градиенте $\nabla J(X)$, где X — позиция бактерии в пространстве поиска R^p , а с помощью $J(\theta)$ моделируются полезные и вредные свойства среды, то есть $J(X)$ характеризует, где находятся аттрактанты и репелленты. Таким образом, $J < 0$, $J = 0$, $J > 0$ означает, что бактерия находится в полезной, нейтральной и вредной среде соответственно.

Пусть $P(j, k, l) = \{X^i(j, k, l), i = 1, 2, \dots, S\}$ описывает позицию каждого члена популяции S бактерий на j -ом хемотаксическом шаге, k -ом шаге воспроизведения и на l -ом событии исключения — рассеивания.

Тогда метод ВФО можно представить в виде последовательности шагов [11].

Шаг 1. Инициализация. Задаются параметры, влияющие на работу метода: S — количество бактерий, N_{re} — количество шагов воспроизведения, N_s — количество шагов-повторений на одном хемотаксическом шаге, N_c — количество хемотаксических шагов, N_{ed} — количество событий исключения-рассеивания; P_{ed} — вероятность рассеивания. Случайным образом распределяются начальные значения X^i , $i = 1, 2, \dots, S$, по пространству поиска. Рассчитываются начальные значения целевой функции для каждой бактерии J^i .

Шаг 2. Установить: $l = l + 1$.

Шаг 3. Установить: $k = k + 1$.

Шаг 4. Установить: $j = j + 1$.

Шаг 5. Для каждой бактерии моделируется хемотаксис: кувыркание, перемещение и скольжение.

Шаг 5.1. Установить: $i = i + 1$.

Шаг 5.2. Кувыркание. Моделирование кувыркания достигается за счёт генерации вектора случайных чисел $\varphi(j) \in R^p$:

$$\varphi = \frac{\Delta}{\sqrt{\Delta^T \Delta}},$$

где Δ – вектор случайных чисел в интервале $[-1; 1]$.

Вектор ϕ представляет собой множество длин для соответствующих измерений.

Шаг 5.3. Перемещение. Рассчитывается новое положение i -ой бактерии по формуле:

$$X^i(j+1, k, l) = X^i(j, k, l) + C(i)\phi(j),$$

где $C(i) > 0$ – размер шага в определённом направлении, позволяющий моделировать процесс кувыркания.

Для новой позиции $X^i(j+1, k, l)$ рассчитывается соответствующее значение целевой функции $J(i, j+1, k, l)$.

Шаг 5.4. Скольжение. Если в позиции $X^i(j+1, k, l)$ значение $J(i, j+1, k, l)$ лучше, чем в позиции $X^i(j, k, l)$, то есть выполняется условие: $J(i, j+1, k, l) < J(i, j, k, l)$, тогда производится следующий хемотаксический шаг с тем же вектором ϕ и в том же направлении (переход к шагу 5.3), и такое повторение может происходить N_s раз. Если условие не выполняется, то переход к шагу 5.5.

Шаг 5.5. Если $i < S$, то переход к шагу 5.1, в противном случае – переход к шагу 6.

Шаг 6. Если $j < N_c$, то переход к шагу 4, в противном случае – переход к шагу 7.

Шаг 7. Воспроизведение. Менее здоровые бактерии умирают, а остальные, более здоровые, разделяются на две бактерии, при этом новые бактерии размещаются в ту же самую точку пространства поиска. За счёт такого подхода обеспечивается неизменность общего количества бактерий. Для этого все бактерии сортируются в соответствии с полученными значениями целевой функции, после чего худшая половина (менее здоровые бактерии) отбрасывается, а лучшая (более здоровые бактерии) – дублируется.

Шаг 8. Если $k < N_{re}$, тогда выполняется переход к шагу 3, в противном случае – переход к шагу 9.

Шаг 9. Исключение и рассеивание. Жизнь популяции бактерий в окружающей среде может меняться либо постепенно (например, путём потребления полезных веществ), либо внезапно в связи с некоторым другим воздействием. Может произойти так, что все бактерии в области погибнут, или колония бактерий будет рассеяна в другую часть окружающей среды. Данный эффект может мешать возможному хемотаксическому прогрессу, но в то же время этот эффект и помогает, поскольку, в случае рассеивания, бактерии могут разместиться около хороших источников с полезными веществами. Исключение и рассеивание помогают понизить вероятность стагнации, то есть заикливания в локальном оптимуме, что часто наблюдается в традиционных градиентных методах оптимизации.

В соответствии с данным подходом каждая бактерия с вероятностью P_{ed} размещается в случайно выбранной точке пространства поиска.

Таким образом, проверяется условие: $U^i < P_{ed}$, где U^i – случайное число в интервале $[0; 1]$ для i -ой бактерии.

Если данное условие выполняется, то бактерия помещается в позицию $X^i(j, k, l)$, полученную случайным образом.

Шаг 10. Если $l < N_{ed}$, тогда выполняется переход к шагу 2, в противном случае – к шагу 11.

Шаг 11. Выбирается и сохраняется лучшее решение J_{best} и соответствующая позиция X_{best} , в которой достигается лучшее решение J_{best} .

Шаг 12. Проверка на окончание поиска. Если были выполнены все циклы для всех бактерий, то выполняется переход к шагу 14, в противном случае выполняется перезапуск – переход к шагу 13.

Шаг 13. Перезапуск агентов: выбираются новые случайные позиции для каждого агента X^i , $i = 1, 2, \dots, S$ и рассчитываются соответствующие значения целевой функции J^i , $i = 1, 2, \dots, S$. Устанавливаются счётчики циклов: $j = 0$; $k = 0$; $l = 0$.

Шаг 14. Останов.

Метод оптимизации на основе моделирования перемещения бактерий применялся для решения различных задач оптимизации: нахождения оптимального потока энергии [14], динамического распределения ресурсов [15] и других [16, 17]. Результаты проведенных экспериментов по решению тестовых задач [14–17] показали, что применение метода BFO является эффективным и перспективным, поскольку результаты, полученные с помощью данного метода, характеризовались более высокой или достаточной точностью по сравнению с традиционными методами.

Можно выделить следующие преимущества метода BFO по сравнению с традиционными методами многомерной оптимизации:

- метод не выдвигает абсолютно никаких требований к целевой функции, то есть функция может быть мультимодальной, негладкой, обладать разрывами;

- для нахождения оптимума целевой функции не надо знать градиент целевой функции;

- метод является методом глобального поиска, в то время как традиционные методы часто являются методами локальной оптимизации;

- метод обладает достаточным количеством настраиваемых параметров, что позволяет его лучшей адаптации для решения той или иной практической задачи оптимизации;

- метод является мультиагентным методом, за счёт чего достигается снижение времени выполнения поиска, поскольку поиск выполняется некоторым количеством агентов параллельно.

- метод является хорошо расширяемым, то есть в его работу можно легко добавлять новые операторы и тем самым создавать гибридные системы. Одним из перспективных является добавление

генетических алгоритмов, которые позволяли бы моделировать более правильные эволюционные процессы, которые наблюдаются в процессе деятельности бактерий.

К недостаткам метода можно отнести:

- случайный характер скорости сходимости, хотя нахождение оптимума гарантируется;
- сильная зависимость работы метода от настраиваемых параметров;
- неполнота модели поведения бактерий, которая может быть доработана, поскольку процедуры воспроизведения и рассеивания ещё не доработаны.

4. Метод оптимизации на основе моделирования перемещения бактерий с применением генетических операторов

Как указывалось выше, одним из недостатков разработанного метода является неполнота модели поведения бактерий. Поэтому в данной работе предлагается расширение существующей модели путём внедрения генетических операторов: оператора отбора, скрещивания и мутации.

Предлагается использовать генетические операторы:

- для отбора – пропорциональный отбор;
- для скрещивания – случайный выбор родительской пары с арифметическим скрещиванием;
- для мутации – простая мутация.

Разработанный метод, дополняющий метод оптимизации на основе моделирования перемещения бактерий с группировкой части [18] и использующий генетические операторы эволюционного поиска, можно представить в виде последовательности следующих шагов.

Шаг 1. Инициализация. Задаются параметры, влияющие на работу метода: S , N_{re} , N_s , N_c , N_{ed} ; P_{ed} . Случайным образом распределяются начальные значения X^i , $i = 1, 2, \dots, S$, по пространству поиска. Рассчитываются начальные значения целевой функции для каждой бактерии J^i .

Для каждой бактерии выполняется пересчёт: $J^i = J^i + J_{ar}(X^i)$. $J_{ar}(X^i)$ рассчитывается по формуле:

$$J_{ar}(X) = \exp(M - J(X))J_{cc}(X),$$

где M – настраиваемый параметр; $J_{cc}(X)$ рассчитывается по формуле:

$$J_{cc}(X) = \sum_{i=1}^S J_{cc}^i = \sum_{i=1}^S \left[-d_{attract} \exp \left(-w_{attract} \sum_{j=1}^p (X_j - X_j^i)^2 \right) \right] + \sum_{i=1}^S \left[h_{repellent} \exp \left(-w_{repellent} \sum_{j=1}^p (X_j - X_j^i)^2 \right) \right],$$

где $d_{attract}$ – глубина аттрактанта (полезных веществ) клетки; $w_{attract}$ – мера ширины аттрактанта;

$h_{repellent} = d_{attract}$ – высота вредного вещества (репеллент); $w_{repellent}$ – мера ширины репеллента.

Таким образом, клетки исключают друг друга посредством локального потребления, а также за счёт того, что клетки не являются пищей друг для друга.

Шаг 2. Установить: $l = l + 1$.

Шаг 3. Установить: $k = k + 1$.

Шаг 4. Установить: $j = j + 1$.

Шаг 5. Моделирование хемотаксиса.

Шаг 5.1. Установить: $i = i + 1$.

Шаг 5.2. Кувыркание и перемещение.

Рассчитывается новое положение i -ой бактерии по формуле:

$$X^i(j+1, k, l) = X^i(j, k, l) + C(i)\phi(j).$$

Рассчитывается значение целевой функции $J(i, j+1, k, l)$.

Устанавливается $J(i, j+1, k, l) = J(i, j+1, k, l) + J_{ar}(X^i(j+1, k, l))$.

Если $J(i, j+1, k, l) < J_{best}$, то установить:

$$J_{best} = J(i, j+1, k, l), X_{best} = X^i(j+1, k, l).$$

Шаг 5.3. Скольжение. Пока выполняется условие: $J(i, j+1, k, l) < J(i, j, k, l)$, повторяется шаг 5.2. Такое повторение может происходить N_s раз. Если условие не выполняется, то переход к шагу 5.4.

Шаг 5.4. Если $i < S$, то переход к шагу 5.1, в противном случае – переход к шагу 6.

Шаг 6. Если $j < N_c$, то переход к шагу 4, в противном случае – переход к шагу 7.

Шаг 7. Воспроизведение.

Шаг 7.1. Отбор. Выполнить пропорциональный отбор.

Шаг 7.1.1. Найти среднее значение целевой функции J_{cp} популяции как среднее арифметическое значений целевых функций всех особей:

$$J_{cp} = \frac{1}{S} \sum_{i=1}^S J^i.$$

Шаг 7.1.2. Для каждой бактерии вычислить отношение $P_s(i) = \frac{J^i}{J_{cp}}$.

Шаг 7.1.3. В зависимости от величины $P_s(i)$ сформировать массив бактерий, допущенных к скрещиванию. Формирование массива допущенных к скрещиванию бактерий осуществляется путём стохастического остаточного отбора: если $P_s(i) < 1$, тогда бактерия считается хорошо приспособленной и допускается к скрещиванию.

Шаг 7.2. Скрещивание. Вначале выполнить случайный выбор родительских пар.

Шаг 7.2.1. Для выбора пары родителей задать вероятность скрещивания P_c . P_c зависит от решаемой задачи и в общем случае лежит в интервале $[0,6; 0,99]$.

Шаг 7.2.2. Произвольным образом пронумеровать все бактерии, отобранные для скрещивания.

Шаг 7.2.3. Выбрать первого родителя. Для этого, начиная с первой, перебрать все бактерии популяции до тех пор, пока случайно выбираемое число из интервала $[0; 1]$ не будет меньше P_c . Когда это случится для одной из бактерий, эта элемент станет первым родителем.

Шаг 7.2.4. Продолжить просмотр популяции, начиная со следующего после первого родителя решения. Просмотр продолжать до тех пор, пока опять случайно выбранное число не будет меньше P_c . Бактерия, для которой выполнится такое условие, будет вторым родителем.

Описанным способом составляются пары до тех пор, пока не выберется нужное количество пар родителей.

Далее следует выполнить арифметическое скрещивание.

При использовании арифметического скрещивания на основе решений родителей X^1 и X^2 создаются два потомка X^{1n} и X^{2n} , значения i -ых координат x^{1n}_i и x^{2n}_i которых рассчитываются по формулам:

$$\begin{aligned}x^{1n}_i &= k \cdot x^1_i + (1 - k)x^2_i, \\x^{2n}_i &= k \cdot x^2_i + (1 - k)x^1_i,\end{aligned}$$

где $k \in [0; 1]$ — некоторый вещественный коэффициент, который выбирается случайным образом на каждой итерации.

Шаг 7.3. Мутация. Для всех бактерий выполнить простую мутацию.

Шаг 7.3.1. Выбрать случайным образом мутирующую координату.

Шаг 7.3.2. В заданном интервале допустимых значений координаты выбрать новое значение координаты, не равное текущему.

Шаг 8. Если $k < N_{re}$, то выполняется переход к шагу 3, в противном случае — переход к шагу 9.

Шаг 9. Исключение и рассеивание.

Проверяется условие: $U^i < P_{ed}$. Если данное условие выполняется, то бактерия помещается в позицию $X^i(j, k, l)$, полученную случайным образом.

Шаг 10. Если $l < N_{ed}$, то выполняется переход к шагу 2, в противном случае — к шагу 11.

Шаг 11. Выбирается и сохраняется лучшее решение J_{best} и соответствующая позиция X_{best} .

Шаг 12. Проверка на окончание поиска. Если были выполнены все циклы для всех бактерий, то выполняется переход к шагу 14, в противном случае выполняется перезапуск — переход к шагу 13.

Шаг 13. Перезапуск агентов: выбираются новые случайные позиции для каждого агента X^i , $i = 1, 2, \dots, S$, рассчитываются соответствующие значения целевой функции J^i , $i = 1, 2, \dots, S$. Счётчики циклов сбрасываются в 0: $j = 0$; $k = 0$; $l = 0$.

Шаг 14. Останов.

5. Отбор информативных признаков на основе моделирования перемещения бактерий

Задача отбора информативных признаков заключается в выборе набора признаков, которые наиболее полно характеризуют изучаемый сложный объект или процесс.

Для решения задачи отбора информативных признаков на основе метода ВФО бактерии должны осуществлять оптимизацию в пространстве поиска $X = (x_1, x_2, \dots, x_N)$, $x_i = \{0, 1\}$, где N — общее количество признаков, характеризующих выборку. Таким образом, каждая точка пространства поиска представляет собой битовую строку, координаты которой позволяют определить, считается ли признак информативным: информативным считается признак, если соответствующая координата равна 1, в противном случае — признак считается неинформативным. В качестве целевой функции $J(X)$ предлагается использовать ошибку модели, построенной на основе признаков, которые являются информативными для данной точки пространства поиска.

При выполнении перемещения по такому пространству поиска необходимо учитывать особенности этого пространства поиска, то есть при получении новой позиции она не должна выходить за пределы пространства поиска. Поэтому новое положение i -ой бактерии предлагается рассчитывать по следующей формуле (3.1):

$$X^i(j+1, k, l) = \begin{cases} 1, & \text{если } X^i(j, k, l) + C(i) \cdot \varphi(j) \geq 1; \\ \text{round}(X^i(j, k, l) + C(i) \cdot \varphi(j)), & \text{если } 0 < X^i(j, k, l) + C(i) \cdot \varphi(j) < 1; \\ 0, & \text{если } X^i(j, k, l) + C(i) \cdot \varphi(j) \leq 0, \end{cases}$$

где $X^i(j, k, l)$ — позиция i -ой бактерии на j -ом хемотаксическом шаге, k -ом шаге воспроизведения и на l -ом событии исключения-рассеивания; φ — представляет собой множество длин для соответствующих измерений; $C(i) > 0$ — размер шага в определённом направлении, позволяющий моделировать процесс кувиркания.

Тогда величины, определяющие процесс движения в пространстве поиска, должны принимать следующие значения:

— переменные, определяющие размер шага в определённом направлении, для i -ой бактерии: $C(i) \in (0; 1)$;

— вектор случайных чисел на j -ом шаге хемотаксиса: $\varphi(j) = \{-1, 0, +1\}$.

Пример перемещения по пространству поиска из трёх признаков, если бактерия в начале поиска находилась в точке $(0; 0; 0)$, вектор размера шага для данной бактерии $C = (1; 1; 1)$, вектор случайных чисел на 1-ом шаге хемотаксиса $\varphi(1) = (1; 0; +1)$, а на 2-ом шаге $\varphi(2) = (-1; +1; 0)$, приведен на рис. 1.

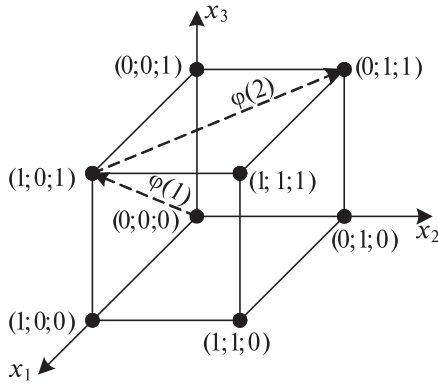


Рис. 1. Пример перемещения в пространстве поиска из трёх признаков

Разработанный метод оптимизации на основе моделирования перемещения бактерий для сокращения размерности набора исследуемых признаков позволяет получить распознающие модели, использующие набор наиболее информативных признаков.

6. Эксперименты и результаты

Для проверки эффективности предложенных изменений в методе оптимизации на основе моделирования перемещения бактерий было разработано программное обеспечение, реализующее предложенный и традиционный методы. Программное обеспечение разрабатывалось при помощи среды разработки пакета Matlab 7.0.

С помощью разработанного программного обеспечения решалась задача нахождения оптимума следующих тестовых функций:

$$f_1 = 5 \cdot e^{-0,1 \cdot ((x_1-15)^2 + (x_2-20)^2)} - 2 \cdot e^{-0,08 \cdot ((x_1-20)^2 + (x_2-15)^2)} + \\ + 3 \cdot e^{-0,08 \cdot ((x_1-25)^2 + (x_2-10)^2)} + 2 \cdot e^{-0,1 \cdot ((x_1-15)^2 + (x_2-20)^2)} - \\ - 2 \cdot e^{-0,5 \cdot ((x_1-5)^2 + (x_2-10)^2)} - 4 \cdot e^{-0,1 \cdot ((x_1-15)^2 + (x_2-5)^2)} - \\ - 2 \cdot e^{-0,5 \cdot ((x_1-21)^2 + (x_2-25)^2)} + 2 \cdot e^{-0,5 \cdot ((x_1-25)^2 + (x_2-16)^2)} + \\ + 2 \cdot e^{-0,5 \cdot ((x_1-5)^2 + (x_2-14)^2)}.$$

$$0 \leq x_1 \leq 30;$$

$$0 \leq x_2 \leq 30.$$

$$f_2 = 20 + e - 20 \exp \left(-0,2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} \right) -$$

$$- \exp \left(\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i) \right),$$

$$n = 2$$

$$-5 \leq x_1 \leq 5;$$

$$-5 \leq x_2 \leq 5.$$

$$f_3 = \sum_{i=1}^n |x_i \cdot \sin(x_i) + 0,1 \cdot x_i|,$$

$$n = 2$$

$$-5 \leq x_1 \leq 5;$$

$$-5 \leq x_2 \leq 5.$$

Графики тестовых функций приведены на рис. 2–4.

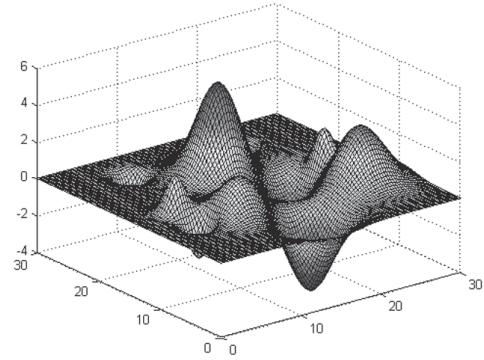


Рис. 2. График тестовой функции f_1

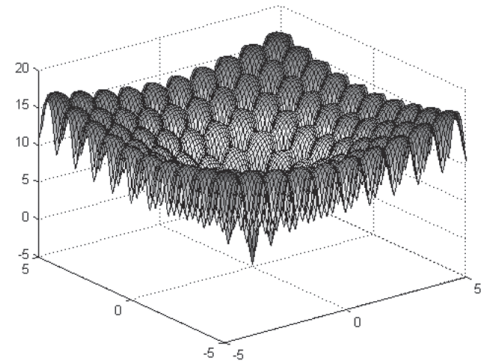


Рис. 3. График тестовой функции f_2

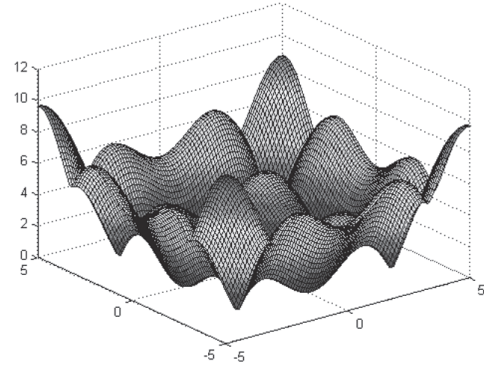


Рис. 4. График тестовой функции f_3

Каждый метод запускался по 10 раз для нахождения оптимума каждой функции. Использовались следующие значения настраиваемых параметров: $S = 20$; $N_{re} = 6$; $N_s = 4$; $N_c = 20$; $N_{ed} = 2$; $P_{ed} = 0,25$; $d_{attract} = 0,1$; $w_{attract} = 0,2$; $h_{repellant} = d_{attract}$; $w_{repellant} = 10$.

В процессе проведения экспериментов рассчитывались следующие показатели: f – лучший полученный оптимум; Δf – отклонение полученного оптимума от реального; c_f – количество вычислений целевой функции; i^* – номер хемотаксического шага, на котором было впервые найдено результирующее лучшее решение; $\bar{f}$ – среднее значение полученных оптимумов функции для всех запусков метода; $D(f)$ – дисперсия полученных значений оптимумов.

Полученные в ходе выполнения эксперименты значения показателей для традиционного метода оптимизации на основе моделирования перемещения бактерий приведены в табл. 1.

Таблица 1

Результаты экспериментов для метода BFO

	f^*	f	Δf	c_f	i^*	$\bar{f}$	$D(f)$
f_1	-3,9865	-3,9863	0,0002	9252	48	-3,7868	0,6317
f_2	0	0,0065	0,0065	8626	207	0,0657	0,0434
f_3	0	0,0004	0,0004	9721	76	0,0037	0,0045

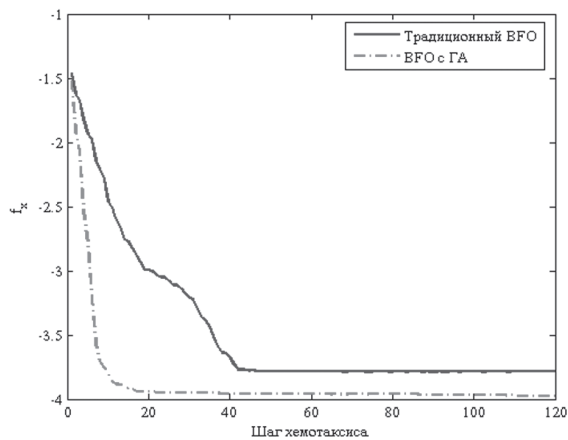
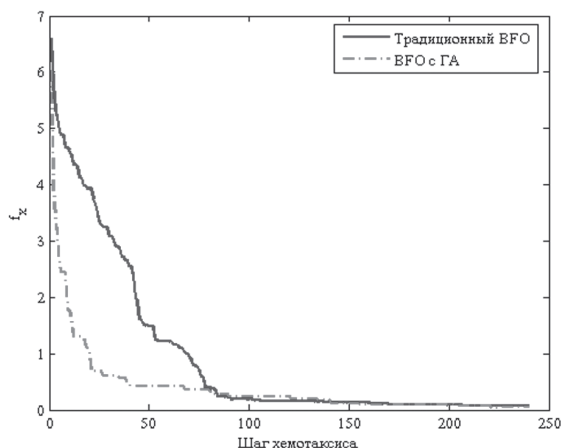
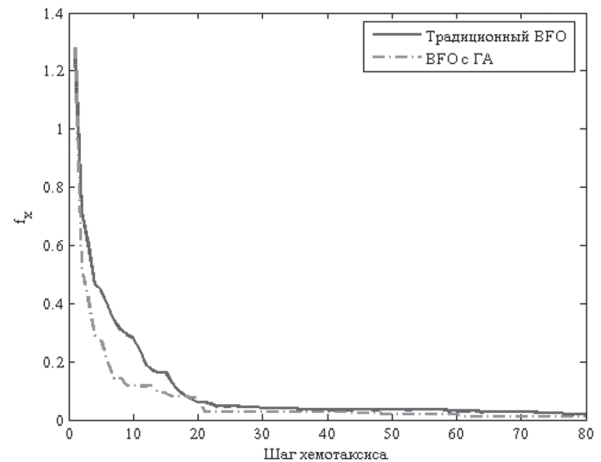
Значения показателей, полученных в ходе экспериментов, для метода оптимизации на основе моделирования перемещения бактерий с использованием генетических операторов представлены в табл. 2.

Таблица 2

Результаты экспериментов для метода BFO-GA

	f^*	f	Δf	c_f	i^*	$\bar{f}$	$D(f)$
f_1	-3,9865	-3,9864	0,0001	5982	88	-3,9821	0,0052
f_2	0	0,0018	0,0018	6527	228	0,0278	0,0198
f_3	0	0,0002	0,0002	5790	62	0,0025	0,0040

Графики работы методов в процессе нахождения оптимальных значений функций представлены на рис. 5–7.

Рис. 5. Процесс оптимизации тестовой функции f_1 Рис. 6. Процесс оптимизации тестовой функции f_2 Рис. 7. Процесс оптимизации тестовой функции f_3

На основании полученных результатов экспериментов можно сделать вывод, что предложенный метод характеризуется большей точностью полученных результатов, лучшей стабильностью получаемых решений, о чём свидетельствуют значения средних значений и дисперсий, полученных вследствие 10 запусков каждого из методов. При этом можно отметить, что оптимальные результаты предложенный метод находит на более поздних этапах своей работы, что в свою очередь способствует получению более оптимальных решений. Также преимуществом предложенного метода является то, что выполняется меньшее количество расчётов оптимизируемой функции.

Таким образом, предложенные дополнения к методу оптимизации на основе моделирования перемещения бактерий способствуют повышению эффективности оптимизационного процесса в целом.

Выводы

В работе получил дальнейшее развитие метод оптимизации на основе моделирования перемещения бактерий.

Вследствие изучения биологических основ метода и существующей математической модели, используемой в методе BFO, был сделан вывод, что данный метод может расширяться, и что используемая математическая модель не является полной. В связи с этим в данной работе было предложено применение генетических операторов отбора, скрещивания и мутации в процессе моделирования перемещения бактерий. Также разработаны подходы, позволяющие применять метод оптимизации на основе моделирования перемещения бактерий для решения задачи отбора информативных признаков. Таким образом, был расширен сам метод оптимизации на основе моделирования перемещения бактерий, и была расширена область его применения.

На основе результатов проведенных экспериментов можно сделать вывод, что предложенные расширения повышают эффективность оптимизационного процесса, что проявляется в: получении более оптимальных значений целевой функции, большей стабильности результатов, меньшем количестве вычислений целевой функции. Таким образом, предложенный метод может быть рекомендован к применению в процессе решения практических задач оптимизации.

Список литературы: 1. Хайкин С. Нейронные сети: полный курс. — СПб: Изд. дом «Вильямс», 2005. — 1104 с. 2. Руденко О.Г., Бодянский Е.В. Штучні нейронні мережі. — Харків: Компанія СМІТ, 2006. — 404 с. 3. Gen M., Cheng R. Genetic algorithms and engineering design. — New Jersey: John Wiley & Sons, 1997. — 352 p. 4. The Practical Handbook of Genetic Algorithms. Applications / Ed. L.D. Chambers. — Florida: CRC Press, 2000. — Vol. I. — 520 p. 5. Beni G., Wang J. Swarm Intelligence // Annual Meeting of the Robotics Society: Proceedings of Seventh International Conference. — Tokyo: RSJ Press, 1989. — P. 425-428. 6. Bonabeau E., Dorigo M., Theraulaz G. Swarm Intelligence: From Natural to Artificial Systems. — New York: Oxford University Press, 1999. — 320 p. 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. 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. 9. Dorigo M., Maniezzo V., Colnani 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. 10. 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. 11. Passino K.M. Biomimicry of bacterial foraging for distributed optimization and control // IEEE Control System Magazine, 2002. — №3 (22). — P. 52-67. 12. Madigan M.T., Martinko J.M., Parker J. Biology of Microorganisms. — New Jersey: Prentice Hall, 1997. — 1036 p. 13. Berg H.C. Random

Walks in Biology. — Princeton: Princeton University Press, 1993. — 164 p. 14. Tripathy M., Mishra S., Lai L.L., Zhang Q.P. Transmission Loss Reduction Based on FACTS and Bacteria Foraging Algorithm // Proceedings of 9th International Conference of Parallel Problem Solving from Nature. — Heidelberg: Springer, 2006. — P. 222 — 231. 15. Analysis And Design Of Intelligent Systems Using Soft Computing Techniques / Ed. Melin P., Castillo O.R., Ramirez E.G., Kacprzyk J. — Heidelberg: Springer, 2007. — 855 p. 16. Mishra S. A hybrid least square-fuzzy bacterial foraging strategy for harmonic estimation // IEEE Transactions on Evolutionary Computation, 2005. — №1(9). — P. 61-73. 17. Kim D.H., Cho C.H. Bacterial Foraging Based Neural Network Fuzzy Learning // Proceedings of the 2nd Indian International Conference on Artificial Intelligence (IICAI-2005). — Pune: IICAI, 2005. — P. 2030-2036. 18. Biswas A., Dasgupta S., Das S., Abraham A. Synergy of PSO and Bacterial Foraging Optimization: A Comparative Study on Numerical Benchmarks // Proceedings of the Second International Workshop on Hybrid Artificial Intelligent Systems (HAIS 2007). — Salamanca: Springer, 2007. — P. 255-263.

Поступила в редколлегию 14.10.2008

УДК 519.6:004.93

Метод оптимізації на основі моделювання переміщення бактерій з використанням еволюційних операторів / О.О. Олійник, С.О. Субботін // Біоніка інтелекту: наук.-техн. журнал — 2008. — № 2 (69). — С. 137-144.

У статті досліджується метод оптимізації на основі мультиагентної оптимізації. Запропоновано удосконалення методу шляхом впровадження в його роботу еволюційних операторів відбору, схрещування та мутації. Також розглянуто можливість його використання до вирішення задачі відбору інформативних ознак.

Табл. 2. Іл. 7. Бібліогр.: 18 найм.

UDC 519.6:004.93

Bacteria foraging optimization with evolutionary operators / A.A. Oleynik, S.A. Subbotin // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 137-144.

The bacteria foraging optimization is analyzed in the paper. The using of evolutionary operators for bacteria foraging optimization is proposed. Application of proposed method for feature selection is offered.

Tab. 2. Fig. 7. Ref.: 18 items.

УДК 519.6



РЕШЕНИЕ ЗАДАЧИ ШТЕЙНЕРА С ПОМОЩЬЮ ГЕНЕТИЧЕСКОГО АЛГОРИТМА

А.И. Ольшевский, М.Ю. Починский

Государственный университет информатики и искусственного интеллекта,
г. Донецк, Украина info@iai.donetsk.ua

Проведен анализ алгоритмов решения задачи Штейнера. Рассмотрено решение задачи с помощью модифицируемого генетического алгоритма. Показано, что для повышения эффективности работы алгоритма целесообразно выделять области исходных данных (создания начальных популяций) с возможностью распараллеливания процессов вычислений.

ЗАДАЧИ ШТЕЙНЕРА, СИНТЕЗ, АЛГОРИТМЫ РЕШЕНИЯ, КРАТЧАЙШАЯ СЕТЬ, ПОПУЛЯЦИЯ, ОБЛАСТЬ, БЛОК ВЫЧИСЛЕНИЙ

Введение

Важными оптимизационными задачами на графах являются задачи построения кратчайших связывающих деревьев. Для минимизации суммарной длины ребер графа в кратчайших связывающих деревьях предложено при соединении множества деревьев использовать дополнительные точки (вершины). Задачу построения минимального дерева при помощи введения дополнительных точек называют задачей Штейнера[1]. Задача Штейнера является NP-полной.

Задача Штейнера имеет практические приложения – конструирование интегральных электронных схем. Более короткая сеть проводящих линий на интегральной схеме требует меньшего времени зарядки-разрядки по сравнению с более длинной сетью и повышает, таким образом, быстродействие схемы. Однако задача отыскания кратчайшей сети на интегральной схеме имеет другую геометрию, так как проводники на ней обычно проходят лишь в двух направлениях – горизонтальном и вертикальном.

Разновидности задачи о кратчайшей сети из вертикальных и горизонтальных соединений применимы и при проектировании локальных компьютерных сетей.

Прямоугольная версия задачи поиска минимального остова дерева может быть эффективно решена алгоритмом, выбирающим на каждом шаге кратчайшее соединение, если это соединение не образует замкнутого пути.

В связи с возможностью быстрого получения набора различных деревьев Штейнера наиболее эффективным представляется применение методов генетического поиска для решения поставленной задачи.

В данной статье рассматриваются существующие алгоритмы решения задачи Штейнера и на основе их анализа разработан генетический алгоритм синтеза кратчайшей сети.

1. Генетический алгоритм

Генетический алгоритм (ГА) – это эвристический алгоритм поиска, используемый для решения

задач оптимизации и моделирования путем последовательного подбора, комбинирования и вариации искомых параметров с использованием механизмов, напоминающих биологическую эволюцию. Является разновидностью эволюционных вычислений. Отличительной особенностью генетического алгоритма является акцент на использование оператора «скрещивания», который производит операцию рекомбинации решений-кандидатов, роль которой аналогична роли скрещивания в живой природе.

Задача кодируется таким образом, чтобы её решение могло быть представлено в виде вектора («хромосома»). Случайным образом создаётся некоторое количество начальных векторов («начальная популяция»). Они оцениваются с использованием «функции приспособленности», в результате чего каждому вектору присваивается определённое значение («приспособленность»), которое определяет вероятность выживания организма, представленного данным вектором. После этого с использованием полученных значений приспособленности выбираются векторы (селекция), допущенные к «скрещиванию». К этим векторам применяются «генетические операторы» (в большинстве случаев «скрещивание» – crossover и «мутация» – mutation), создавая таким образом следующее «поколение». Особи следующего поколения также оцениваются, затем производится селекция, применяются генетические операторы и так далее. Так моделируется «эволюционный процесс», продолжающийся несколько жизненных циклов (поколений), пока не будет выполнен критерий останова алгоритма. Таким критерием может быть:

- нахождение глобального, либо субоптимального решения;
- исчерпание числа поколений, отпущенных на эволюцию;
- исчерпание времени, отпущенного на эволюцию.

Генетические алгоритмы служат, главным образом, для поиска решений в очень больших, сложных пространствах поиска.

Таким образом, можно выделить следующие этапы генетического алгоритма:

1. Создание начальной популяции.
2. Вычисление функций приспособленности для особей популяции (оценивание).
3. Начало цикла.
4. Выбор индивидов из текущей популяции (селекция).
5. Скрещивание и\или мутация.
6. Вычисление функций приспособленности для всех особей.
7. Формирование нового поколения.
8. Если выполняются условия останова, то конец цикла, иначе — в начало цикла на шаг 3.

Достоинством эвристических алгоритмов является их способность находить оптимальные или же близкие к оптимальным решения задачи большого числа точек за разумное время вычисления. Недостатком является то, что относительно найденного решения нельзя утверждать, что оно оптимально [2]. Двумя основными критериями, оценивающими эвристический алгоритм, будут скорость сходимости и близость получаемого решения к оптимальному или же к лучшему решению. ГА обладает механизмом обмена информацией между особями, который порождает из родительских решений решения потомков [3].

У обычного алгоритма генетического поиска есть существенный недостаток — очень важна последовательность, в которой будут заданы исходные точки. Исходя из этого, предлагается алгоритм с изменяющейся последовательностью точек.

На рис. 1 изображена двухуровневая общая схема функционирования алгоритма.

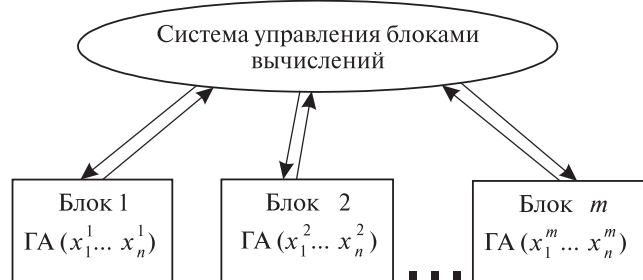


Рис. 1. Структурная схема алгоритма

Блок вычислений — это генетический алгоритм с заданным набором входных параметров и результатом каждого этапа эволюции. Фактически блоки представляют собой отдельные программы, работающие по модифицированному алгоритму генетического поиска с постоянной последовательностью соединения исходных точек. Сами соединения могут быть разными.

Система управления блоками вычислений (СУБВ) — это программа, которая подает на блоки вычислений последовательность точек и производит анализ результатов работы блоков вычислений.

2. Алгоритм работы СУБВ

Алгоритм состоит из 6 основных шагов.

1. Введем кодировку хромосом. Ген будет иметь вид числа от 1 до n , где n — количество исходных точек.

2. Случайным образом (путем генерации случайных последовательностей чисел от 1 до n) создаем начальную популяцию для системы управления блоками вычислений. Обнуляем счетчик популяций: $t = 0$.

3. Рассчитываем значение фитнес-функции. Для этого передаем каждую хромосому в блок вычислений, и он возвращает значение фитнес-функции.

4. Создаем новую популяцию.

4.1. С помощью оператора селекции выбирается 2 родителя.

4.2. Оператор кроссинговера скрещивает родителей и получается потомок (новая хромосома).

4.3. Для того чтобы алгоритм не вырождался, после кроссинговера запускается оператор мутации.

4.4. Если сгенерирована не вся популяция, то переходим на шаг 4.1.

5. Рассчитываем фитнес-функцию для хромосом новой популяции. Для этого передаем каждую хромосому в блок вычислений, и он возвращает значение фитнес-функции.

6. $t = t + 1$. Если $t = t_{\max}$, где $t_{\max}$ — ограничение на количество популяций, алгоритм прекращает работу и решением будет лучшая хромосома последней популяции, иначе на шаг 4.

После того, как ввели кодировку хромосом, создается начальная популяция и рассчитывается фитнес-функции. Для этого генерируется m последовательностей неповторяющихся случайных чисел от 1 до n , где m — количество блоков вычислений, а n — количество исходных точек.

Таким образом, имеем N хромосом, где N может быть задано произвольно.

Пусть $m = 4$ и заданы 4 исходные точки:

(1, 2), (2, 2), (3, 1), (4, 1).

Предположим, что в результате генерации последовательностей были получены следующие хромосомы:

3, 4, 1, 2;
1, 2, 3, 4;
1, 2, 4, 3;
3, 4, 2, 1.

После этого для каждой хромосомы рассчитываем фитнес-функцию. Передаем хромосому в блок вычислений, и он возвращает значение фитнес-функции.

Обнуляем счетчик популяций: $t = 0$.

Создание новой популяции. Для генерации одной хромосомы новой популяции необходимо последовательно запустить 3 оператора:

- 1) оператор селекции;
- 2) оператор кроссинговера;
- 3) оператор мутации.

Первым запускается оператор селекции. Оператор селекции осуществляет отбор хромосом в соответствии со значениями их функции приспособленности. Существуют как минимум три популярных типа оператора селекции: рулетка, турнирный и равномерный отбор.

Турнирный отбор реализует m турниров, чтобы выбрать m особей. Каждый турнир построен на выборке k элементов из популяции, и выбора лучшей особи среди них. Наиболее распространен турнирный отбор с $k=2$.

При равномерном отборе все хромосомы имеют равную вероятность быть выбранными. Выбор родительских хромосом производится случайно в независимости от приспособленности хромосом.

Метод рулетки — отбирает особей с помощью n «запусков» рулетки. Колесо рулетки содержит по одному сектору для каждого члена популяции. Размер i -го сектора пропорционален соответствующей величине $P_{sel}(i)$ вычисляемой по формуле:

$$P_{sel}(i) = \frac{f(i)}{\sum_{i=1}^n f(i)}.$$

При таком отборе члены популяции с более высокой приспособленностью с большей вероятностью будут чаще выбираться, чем особи с низкой приспособленностью. На рис. 2 показан пример распределения вероятности выбора особи.

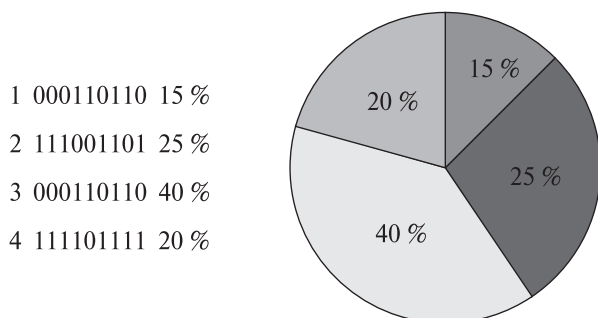


Рис. 2. Оператор селекции типа колеса рулетки с пропорциональными секторами функции приспособленности

Для того чтобы увеличить скорость работы алгоритма будет использоваться равномерный отбор, так как для его осуществления потребуется наименьшее количество математических операций по сравнению с другими видами отбора.

Таким образом, после оператора селекции имеется два числа, представляющие собой порядковые номера хромосом, которые станут родителями.

Оператор кроссинговера скрещивает родителей и получает потомка.

Далее начинает свою работу оператор кроссинговера или скрещивания, как его еще называют.

Оператор скрещивания (crossover) осуществляет обмен частями хромосом между двумя (может быть и больше) хромосомами в популяции. Может быть односточечным или многоточечным. Односточечный кроссовер работает следующим образом. Сначала случайным образом выбирается одна из 1-1 точек разрыва. Точка разрыва — участок между соседними битами в строке. Обе родительские структуры разрываются на два сегмента по этой точке. Затем соответствующие сегменты различных родителей склеиваются и получают два генотипа потомков.

Многоточечный кроссовер отличается только тем, что он имеет несколько точек разрыва вместо одной.

Односточечный оператор кроссинговера требует меньшего количества операций, поэтому использоваться в алгоритме будет он.

Для того чтобы алгоритм не вырождался, после кроссинговера запускается оператор мутации.

Мутация — стохастическое изменение части хромосом. Каждый ген строки, которая подвергается мутации, с вероятностью P_{mut} (обычно очень маленькой) меняется на другой ген.

На рис. 3 изображен пример мутации.

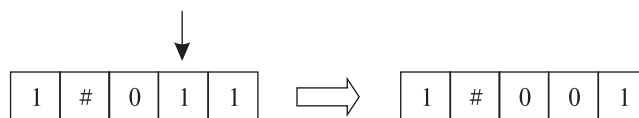


Рис. 3. Оператор мутации

Оператор мутации необходим для «выбивания» популяции из локального экстремума и способствует защите от преждевременной сходимости.

Так же как и кроссинговер, мутация проводится не только по одной случайной точке. Можно выбирать некоторое количество точек в хромосоме для изменения, причем их число также может быть случайным. Также можно мутировать сразу некоторую группу подряд идущих точек. Вероятность мутации значительно меньше вероятности кроссинговера и редко превышает 1%. Среди рекомендаций по выбору вероятности мутации нередко можно встретить варианты $1/L$ или $1/N$, где L — длина хромосомы, N — размер популяции.

Таким образом, в результате работы операторов селекции, кроссинговера и мутации получается новая хромосома.

Новая хромосома подается на вход блока вычислений, и тот возвращает значение фитнес-функции.

После запуска этих 3 операторов m раз, где m — число блоков вычислений, генерация новой популяции закончена.

Счетчик популяций t увеличиваем на 1. $t = t + 1$.

В случае, если выполняется условие окончания алгоритма, программа выдает лучшую хромосому из последней популяции.

Условие окончания может быть как ограничением на количество популяций, так и любым другим.

3. Алгоритм работы блока вычислений

Алгоритм блока вычислений можно разбить на 6 основных шагов.

1. Введем кодировку хромосом.

Используется следующая кодировка хромосом: П – вправо; Н – вниз; В – вверх.

Таким образом, указывается направление, в котором стоит следовать от начала до конца дерева.

То есть для заданных точек в пункте 4.3 может получиться следующая хромосома:

1П, 1П, 1Н, 1П.

Здесь гены разделены запятыми. Графически хромосому можно представить, как показано на рис. 4.

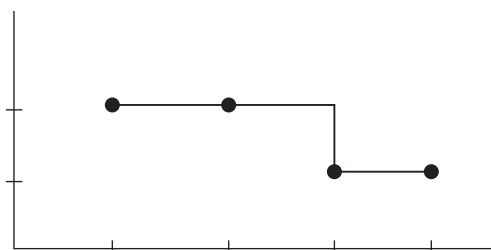


Рис. 4. Пример хромосомы

2. Создаем начальную популяцию для блока вычислений. Хромосомы строятся через всевозможные столбы Штейнера. Обнуляем счетчик популяций: $u = 0$.

Столб Штейнера – это прямая, проведенная через любую точку из исходных, такая, что из каждой другой из исходных точек опускается перпендикуляр до пересечения с этой прямой.

3. Рассчитываем фитнес-функцию для всех хромосом блока вычислений. Фитнес-функция – это длина дерева.

4. Создаем новую популяцию.

4.1. С помощью оператора селекции выбирается 2 родителя.

4.2. Оператор кроссинговера скрещивает родителей и получается потомок (новая хромосома).

4.3. Для того чтобы алгоритм не вырождался, после кроссинговера запускается оператор мутации.

4.4. Если сгенерирована не вся популяция, то – на шаг 4.1.

5. Рассчитываем фитнес-функцию для хромосом новой популяции.

6. $u = u + 1$. Если $u = u_{\max}$, где $u_{\max}$ – ограничение на количество популяций, алгоритм прекращает свою работу, и решением будет лучшая хромосома последней популяции, иначе – на шаг 4.

Создание начальной популяции и расчет фитнес-функции. Хромосомы строятся через всевозможные вертикальные или горизонтальные столбы Штейнера.

На рис. 5 приведен пример двух деревьев, построенных на начальной популяции с помощью горизонтальных столбов Штейнера.

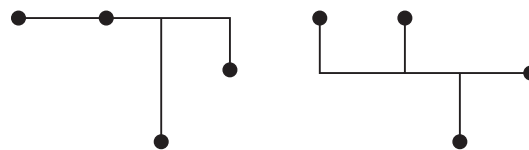


Рис. 5. Деревья с горизонтальными столбами Штейнера

На рис. 6 приведен пример деревьев, построенных с помощью вертикальных столбов Штейнера.

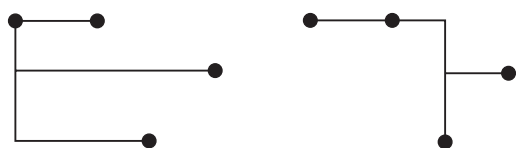


Рис. 6. Деревья с вертикальными столбами Штейнера

После создания нового поколения для каждой хромосомы рассчитывается фитнес-функция или длина дерева.

Создание новой популяции. Для генерации одной хромосомы новой популяции необходимо последовательно запустить 3 оператора:

- 1) оператор селекции;
- 2) оператор кроссинговера;
- 3) оператор мутации.

С помощью равномерного оператора селекции выбирается два родителя.

Одноточечный оператор кроссинговера скрещивает родителей и получает потомка.

Скрещивание производится одноточечным кроссинговером. На рис. 7 приведен пример скрещивания двух деревьев.

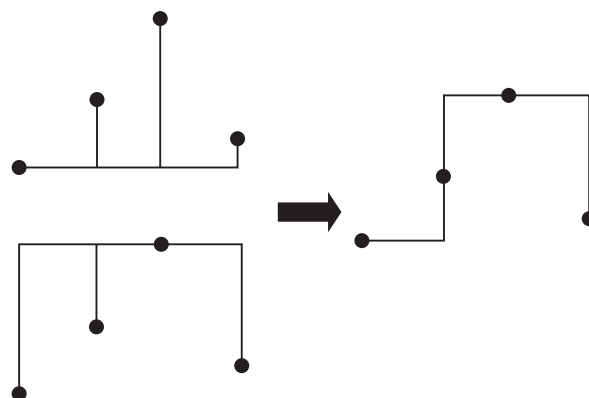


Рис. 7. Скрещивание двух деревьев

Запускается оператор мутации, который изменяет потомка с заданной вероятностью.

Для мутации был выбран отрезок между 2 и 3 точками. На рис. 8 приведена хромосома до и после работы оператора мутации.



Рис. 8. Хромосома до мутации и после

В данном случае мутация проводится между двумя исходными точками с одной дополнительной точкой и просто меняет координаты по осям X и Y местами. Но может возникнуть и другая ситуация. Например, когда дополнительных точек две и они заменяются одной (рис. 9).

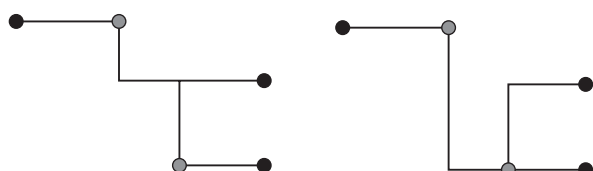


Рис. 9. Хромосома до мутации и после

Конечно, в приведенном примере эта мутация не вызывает никакого сокращения дерева Штейнера, но когда количество точек большее, это может приводить к заметному уменьшению длины дерева.

Но бывают ситуации, когда имеется две дополнительные точки. В этом случае имеет смысл уменьшить общую длину дерева, убрав обе дополнительные точки. В результате мутации будут удалены обе дополнительные точки и исходные точки будут соединены прямой линией (рис. 10).

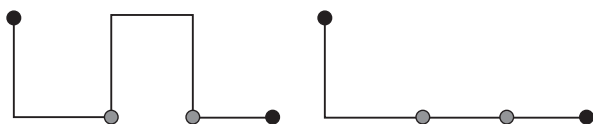


Рис. 10. Хромосома до мутации и после

Существует также другая ситуация, когда подвергаются мутации точки, соединенные прямой линией. Здесь имеет смысл, наоборот, добавить две дополнительные точки, так как есть вероятность, что линии между этими точками совпадут с уже существующими, и в итоге сократится дерево. Только добавление точек происходит в согласовании с ближайшими точками (рис. 11).



Рис. 11. Хромосома до мутации и после

Рассчитываем фитнес-функцию для хромосом новой популяции.

Счетчик популяций u увеличиваем на 1. $u = u + 1$.

В случае, если выполняется условие окончания алгоритма, программа выдает лучшую хромосому из последней популяции.

Условие окончания может быть как ограничением на количество популяций, так и любым другим.

Опционально, в зависимости от исходных данных, несколько хромосом остаются без изменений. Эти хромосомы называются лидерами популяции.

4. Анализ особенностей формирования исходных данных

Анализ проводился на основании результатов работы программы модифицированного генетического алгоритма.

Последовательно были заданы 6, 20 и 60 точек. Для каждой точки программа запускалась по 20 раз. После чего можно попытаться выявить закономерность, в которой должны соединяться исходные точки.

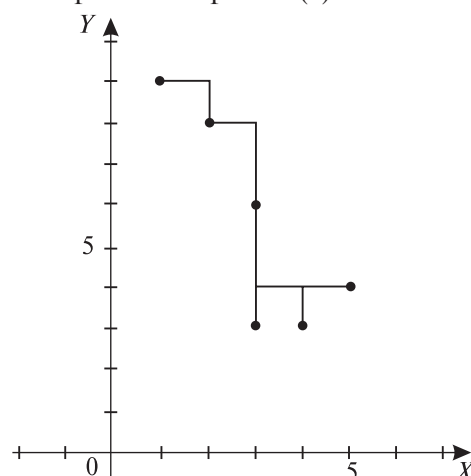
В табл. 1 приведены координаты 6-ти исходных точек.

Таблица 1

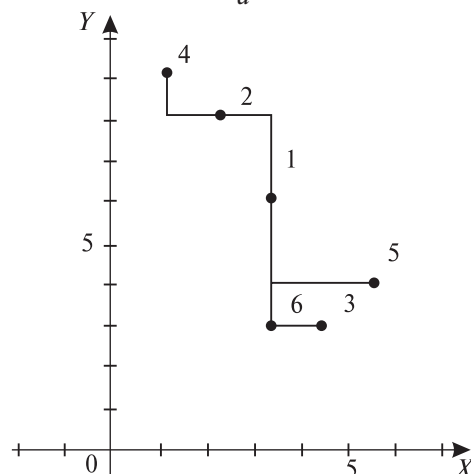
Исходные точки						
№	1	2	3	4	5	6
X	3	2	4	1	5	3
Y	6	8	3	9	4	3

В результате запуска алгоритма 16 раз из 20 получилось дерево, изображенное на рис. 12(а), длиной 11 единиц при порядке исходных точек – 3, 6, 5, 1, 2, 4 (8 раз из 16) и 5, 6, 3, 1, 2, 4 (7 раз из 16), и 1 раз, при последовательности точек 3, 6, 5, 1, 2, 4.

Остальные 2 раза было получено дерево такой же длины, но с другим порядком – 3, 6, 4, 2, 1, 5. Дерево изображено на рис. 12(б).



а



б

Рис. 12 Деревья для 6 точек, полученные при разном порядке данных

Порядок точек можно представить в виде табл. 2.

Таблица 2

Результат работы программы

№	Порядок точек	Доля от общего числа запусков	Длина дерева
1	3, 6, 5, 1, 2, 4	45%	11
2	5, 6, 3, 1, 2, 4	35%	11
3	3, 6, 4, 2, 1, 5	20%	11

Во всех трех последовательностях точек можно видеть, что точки 1, 2 и 4 все время стоят рядом и образуют область, которая обязательно присутствует во всех лучших решениях.

Также в двух первых случаях область образуют точки 3, 6 и 5, но в третьем случае эта область разбивается на две.

В результате запуска алгоритма 15 раз из 20 получались деревья, изображенные на рисунках 13(а), 13(б) и 13(в), длиной 80 единиц при разных последовательностях исходных точек. Остальные 5 раз были получены различные последовательности и деревья большей длины.

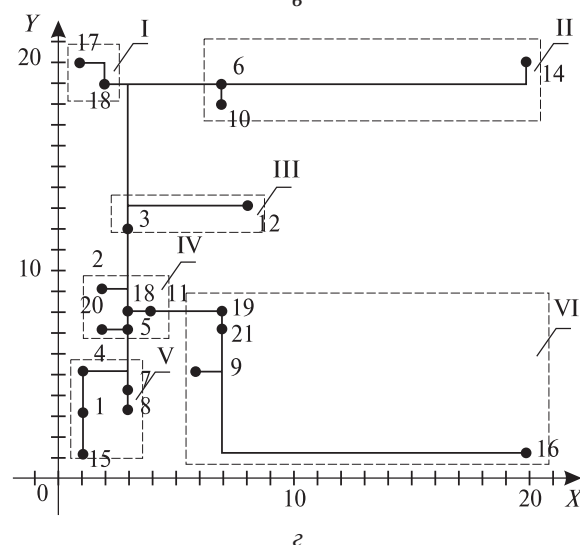
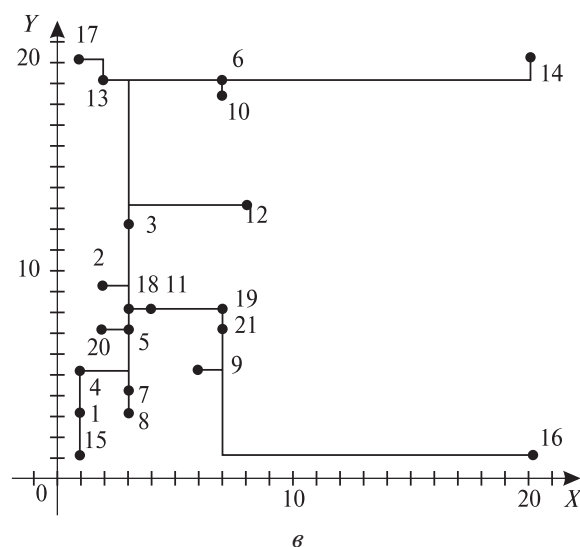
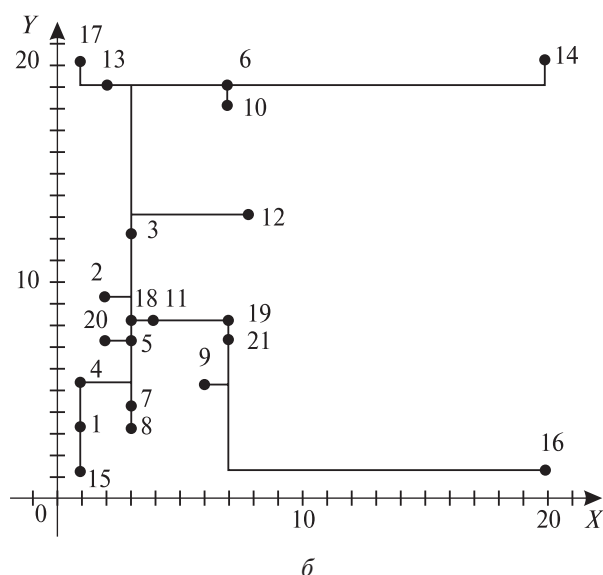
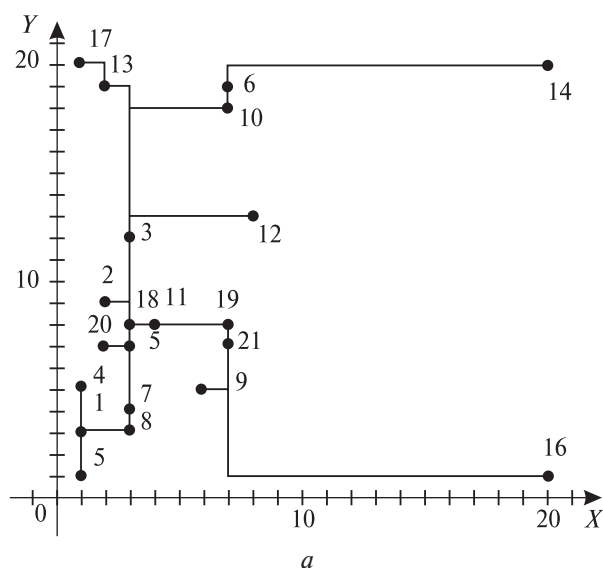


Рис. 13. Деревья для 21 точки, полученные при разном порядке данных

Результаты работы алгоритмов для 21 исходной точки приведены в табл. 3.

Таблица 3

Исходные точки

№	1	2	3	4	5	6	7	8	9	10	11	12	13
X	1	2	3	1	3	7	3	3	6	7	4	8	2
Y	3	9	12	5	7	19	4	3	5	18	8	13	19

№	14	15	16	17	18	19	20	21
X	20	1	20	1	3	7	2	7
Y	20	1	1	20	8	8	7	7

Порядок точек и результаты можно представить в виде табл. 4.

Если рассматривать только последовательности, в которых были получены самые короткие деревья, то во всех последовательностях можно выделить повторяющиеся области. Так, любая последовательность состоит из этих областей, соединенных между собой в разном порядке. В табл. 4 для наглядности каждая вторая область выделена жирным шрифтом.

В данном случае области выделялись аналитическим путем. Выделялись области, которые присутствовали во всех последовательностях и были неразрывны. Точки в областях менялись местами,

но не пересекались с другими областями. С большим количеством точек для выделения областей можно использовать любой алгоритм кластеризации по одному признаку.

Таблица 4

Результат работы программы

№	Порядок точек	Доля от общего числа запусков	Длина дерева
1	14, 6, 10, 2, 18, 20, 5, 11, 21, 16, 9, 19, 8, 4, 15, 1, 7, 12, 3, 13, 17	35%	80 (рис. 13(б))
2	17, 13, 12, 3, 7, 15, 1, 4, 8, 6, 14, 10, 19, 9, 16, 21, 18, 2, 20, 11, 5	25%	80 (рис. 13(б), рис. 13(в))
3	7, 8, 4, 15, 1, 5, 20, 11, 2, 18, 9, 16, 21, 19, 3, 12, 13, 17, 6, 10, 14	15%	80 (рис. 13(б), рис. 13(г))
4	Другие порядки	25%	82-86

На основании анализа табл. 4 можно выделить области, как показано на рис. 13(г).

Области образуются скоплениями точек, что хорошо видно из I, IV и V областей на рис. 13(г). Точка 14 была включена в одну область с точками 10 и 6, так как любые другие соединения образуют более длинное дерево. Так же произошло с точками 16 (область VI), 12 и 3 (область III).

На основании полученных результатов можно сделать вывод, что признаком, по которому выделяются области, является расстояние между точками.

Таким образом, имеет смысл предложить предобработку исходных данных для существующих алгоритмов решения задачи Штейнера на основании выделения областей.

5. Результаты численных экспериментов

Для программной реализации модифицированного генетического алгоритма потребовалось создать 3 класса: класс СУБВ, класс истории эволюции и класс самой эволюции.

Класс эволюции — это основная часть блока вычислений, в которой реализован генетический алгоритм. В нем реализованы операторы кроссинговера, селекции, мутации и вычисления фитнес-функции.

Класс истории эволюции используется для того, чтобы изучить возможности реализованного алгоритма. Он хранит все популяции, что требует немалого количества оперативной памяти.

Двухуровневая структура организации программного продукта позволяет выполнять параллельные действия, поэтому при увеличении количества ядер наблюдается заметное повышение производительности[4]. С целью определения прироста производительности были проведены эксперименты, при которых определялось время создания следующей популяции с заданными параметрами алгоритма: 8 блоков вычислений, 20 исходных точек, 500 популяций для каждого блока, 20% лидеров, 6% вероятность мутации. В системе с установленным процессором Core 2 Quad 2,4 GHz 8MB L2 cache путем отключения ядер были получены результаты, представленные в табл. 5.

Таблица 5

Зависимость производительности от количества ядер

	Среднее время выполнения, сек	Прирост по сравнению с 1-ядром, %
1-ядро	5,12	0
2-ядра	3,43	49
4-ядра	2,98	72

Таким образом, при переходе с одноядерного процессора на двухядерный наблюдается 49-процентное увеличение производительности, а при переходе с 2-ядерного на 4-ядерный еще 15-процентное.

Заключение

Приведенные исследования показывают, что для уменьшения времени вычислений целесообразно выделять области исходных данных для создания начальных популяций при решении задачи Штейнера с помощью генетического алгоритма. С целью повышения эффективности особое внимание при программной реализации необходимо уделить двухуровневой структуре организации параллельных вычислений.

На основании исследований разработан модифицированный генетический алгоритм решения задачи Штейнера как часть программного комплекса маршрутизации пакетов данных в сети дистанционного обучения. Программная реализация выполнена на языке программирования C++ с использованием библиотеки QT 4.2.

Список литературы: 1. Ольшевский А.И. Алгоритмы построения маршрута при групповой рассылке сетевых пакетов данных дистанционного обучения на базе ДонГИИИ // Искусственный интеллект. — 2007. — № 4. — С. 483-490. 2. Курейчик В.М. Генетические алгоритмы и их применение. Таганрог: ТРТУ, 2002. — 242 с. 3. Маршалл Берн У., Грэм Рональд Л. Поиск кратчайших сетей: Пер. с англ. — М.: Спайтер, 1993. — 104 с. 4. Наумов Л.А. Метод разбиения задач на подзадачи, рекурсия, «разделяй и властвуй» // Динамическое программирование. — 2002, № 3,4. — С. 94-106.

Поступила в редколлегию 20.10.2008

УДК 519.6

Вирішення задачі Штейнера за допомогою генетичного алгоритму / А.І. Ольшевський, М.Я. Починський // Біоніка інтелекту: наук.-техн. журнал — 2008. — № 2 (69). — С. 145-151.

У роботі розглядаються питання вирішення задачі Штейнера за допомогою генетичного алгоритму. Для зменшення часу обчислень пропонується дворівнева структура з передобробкою початкових точок. Описуються запропоновані підходи. Приводяться аналіз особливостей формування початкових даних і деякі результати чисельних експериментів.

Табл. 5. Іл. 13. Бібліогр.: 4 найм.

UDC 519.6

One Solution of Steiner's Problem Using the Genetic Algorithm / A.I. Olshevskiy, M.Ya. Pochinskiy // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 145-151.

In the article some questions of Steiner's problem using the genetic algorithm are considered. The two-level structure with preprocessing of initial points is proposed to reduce time of calculations. The suggested approaches are described. The analysis of initial data forming and some results of numerical experiments are presented.

Tab. 5. Fig. 13. Ref.: 4 items.

УДК 004.056.5:518:512.624.3



АНАЛИЗ СИНГУЛЯРНЫХ ЧИСЕЛ МАТРИЦЫ ЦИФРОВОГО СИГНАЛА КАК ОСНОВА ПОДХОДА К ОБНАРУЖЕНИЮ ЕГО ФАЛЬСИФИКАЦИИ

А.А. Кобозева

Одесский национальный политехнический университет, г. Одесса, Украина, Alla_kobozeva@ukr.net

Разработаны математические основы общего подхода к проблеме доказательства подлинности цифрового сигнала, детектирования его фальсификации независимо от вида сигнала и фальсифицирующего воздействия, обоснован выбор используемых математических инструментов, а также проведена адаптация предложенного общего подхода к решению конкретной задачи.

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ, ПОДЛИННОСТЬ ЦИФРОВЫХ СИГНАЛОВ (ИЗОБРАЖЕНИЙ), ЦИФРОВОЙ СИГНАЛ, СИНГУЛЯРНЫЕ ЧИСЛА, МАТРИЦЫ ИЗОБРАЖЕНИЯ, ФАЛЬСИФИЦИРОВАННЫЙ, ОСНОВНОЕ ИЗОБРАЖЕНИЕ, ЗАМЕЩАЮЩАЯ ОБЛАСТЬ, ЧАСТИЧНОЕ ВОССТАНОВЛЕНИЕ, ПОЛНОЕ ВОССТАНОВЛЕНИЕ

Введение

Для современного общества проблема информационного обеспечения всех сфер деятельности по своей значимости и актуальности превосходит проблему дальнейшей индустриализации производства, которая до недавнего времени считалась одной из центральных. Общество вступает в постиндустриальный период развития, который, по всеобщему мнению, можно назвать информационным [1].

Процесс внедрения новых информационных технологий невозможен без решения вопросов информационной безопасности [2], составной частью которой является обеспечение подлинности цифровых сигналов (изображений (ЦИ), видео, аудио (ЦА)), создание методов обнаружения их фальсификации. Этим вопросам в современной печати уделяется много внимания [3–13], однако, до настоящего момента не существовало общего подхода к проблеме идентификации несанкционированного вмешательства в цифровой сигнал (ЦС), основывающегося на единой математической базе, который мог бы результативно использоваться независимо от того, какой именно сигнал подвергся фальсификации: ЦИ, видео или ЦА. Кроме того, методы, предлагаемые в открытой печати, не лишены значительных недостатков: многие не имеют под собой строгой математической базы [6–9], некоторые опираются лишь на специфические возможности используемых для получения ЦС технических средств [3–5] и так далее. Все это заставляет искать новые математические инструменты и подходы к решению рассматриваемой задачи в целом, начало чему было положено автором настоящей работы в [14], где был предложен метод обнаружения фальсификации ЦИ, базирующийся на анализе возмущений сингулярных чисел (СНЧ) блоков матрицы изображения. Математическая основа метода делает его принципиально адаптируемым для работы не только с ЦИ, но и цифровыми видео, аудио,

перспективным с точки зрения использования его основы для создания общего подхода к проблеме выявления фальсификации произвольного ЦС.

1. Постановка задачи и цель исследования

Целью статьи является формирование математических основ единого подхода к решению проблемы доказательства подлинности ЦС, детектирования его фальсификации независимо от вида сигнала и фальсифицирующего воздействия, обоснование выбора математических инструментов, используемых для этого, а также попытка непосредственной адаптации формируемого подхода к решению конкретной задачи.

Для достижения поставленной цели необходимо решить следующие **задачи**:

- определить математическую формализацию для произвольного ЦС, процесса его фальсификации;
- на основе принятой формализации определить основные математические инструменты для анализа ЦС;
- выделить математические параметры, полностью определяющие ЦС при принятой формализации;
- установить характерные особенности выделенных параметров, определяющие свойства ЦС, позволяющие отделить фальсифицированный ЦС от оригинального при конкретном способе фальсификации;
- рассмотреть пример адаптации нового общего подхода для выявления фальсификации ЦС, проведенной определенным способом, путем разработки практического метода.

2. Математическое представление цифрового сигнала и его фальсификации

Произвольное ЦИ и видеосигнал, рассматриваемый как последовательность видеок кадров, естественным образом представимы в виде матрицы (совокупности матриц) [15]. Хотя общеиспользуемой

математической формализацией ЦА является вектор, возможны различные способы перехода к его двумерному матричному представлению. Таким образом, имеет место следующее утверждение.

Утверждение 1. В качестве математической модели любого из рассматриваемых ЦС (ЦИ, видео, ЦА) можно использовать матрицу (конечное множество матриц). Свойства ЦС, независимо от его конкретного вида, будут определяться математическими свойствами соответствующих матриц.

В силу утверждения 1 в качестве одного из математических инструментов при формировании единого подхода к проблеме обнаружения фальсификации ЦС используется матричный анализ [16].

Поскольку любая матрица однозначно определяется своим сингулярным спектром — множеством сингулярных чисел (СНЧ) — и набором сингулярных векторов (СНВ) специального вида, которые получаются при помощи нормального сингулярного разложения матрицы (SVD) [17], то имеет место следующее утверждение.

Утверждение 2. Любой ЦС при выбранном матричном способе формализации определяется сингулярным спектром (спектрами) и набором (наборами) СНВ соответствующей ему матрицы (матриц): СНЧ и СНВ несут в себе всю информацию о состоянии ЦС.

Произвольное преобразование ЦС, в том числе и фальсификация, представляется в виде возмущения соответствующей матрицы (матриц) [17], откуда, с учетом утверждения 2, следует следующее утверждение.

Утверждение 3. Любое преобразование ЦС формально представимо в виде совокупности возмущений СНЧ и СНВ соответствующей ему матрицы (матриц).

Таким образом, в качестве математического инструмента при формировании подхода к проблеме обнаружения фальсификации ЦС наряду с матричным анализом естественно использовать теорию возмущений [18].

СНЧ произвольной матрицы, в отличие от ее СНВ, являются хорошо обусловленными [18], их возмущения сравнимы с величиной возмущающего воздействия [17]. В силу этого для оценки состояния ЦС до и после фальсифицирующего преобразования и силы самого преобразующего воздействия имеет смысл анализировать совокупность возмущений только СНЧ соответствующей матрицы (матриц).

Утверждения 1–3 составляют математический базис для создаваемого общего подхода к рассматриваемой проблеме, основная идея которого заключается в сведении анализа произвольного ЦС, производимого с целью обнаружения его фальсификации, к анализу СНЧ соответствующей матрицы (матриц), выявлению их характерных

особенностей, сигнализирующих о несанкционированном изменении ЦС. Очевидно, что после представления сигнала в матричном виде для предлагаемого подхода конкретика ЦС (ЦИ, ЦА или видео) не играет роли.

Для различных возмущений (в частности, различных способов фальсификации) характерные особенности СНЧ, сигнализирующие о возмущающем воздействии (несанкционированном изменении сигнала), очевидно будут различны. Необходимо выделить и обосновать признаки СНЧ, наличие или отсутствие которых говорит об определенных характеристиках исследуемого сигнала (дает возможность отделить фальсифицированный сигнал от оригинального при конкретном способе фальсификации), начало чему было положено в [14].

3. Анализ сингулярных чисел

Для определенности и наглядности проводимых ниже рассуждений, не ограничивая в силу вышесказанного их общности, будем рассматривать в качестве ЦС изображение в градациях серого, конкретизируя, где необходимо, полученные результаты для ЦА, видео.

Рассмотрим адаптацию предложенного выше подхода, основанного на анализе СНЧ соответствующих матриц, для конкретной задачи фотомонтажа ЦИ, проделанного наиболее часто встречаемым на практике способом, заключающимся в следующем. Пусть часть ЦИ, которое будем называть *основным изображением* (ОИ), заменяется частью ЦИ, далее называемой *вклейкой*, или замещающей областью (ЗО). ОИ и изображение, из которого взята ЗО, получены при помощи современных видеокамер, большинство из которых использует для хранения изображений схему JPEG (с потерями), основанную на дискретном косинусном преобразовании (DCT) [15], включающую в качестве основного шага *квантование*: DCT-коэффициенты блоков 8×8 , полученных после предварительного разбиения матрицы ЦИ, квантуются при помощи массива нормализации и округляются до ближайшего целого. Восстановление изображения после JPEG-сжатия далее будем рассматривать в двух видах: частичное восстановление (ЧВ), включающее в себя обратное DCT без округления полученных коэффициентов, и полное восстановление (ПВ), при котором происходит округление коэффициентов матрицы, являющейся результатом ЧВ (введение их значений во множество целых чисел $\{0, 1, \dots, 255\}$) [14]. Обсуждаемые фотоснимки — ПВ после JPEG-сжатия изображения.

Пусть полученное фальсифицированное ЦИ снова сохраняется в формате JPEG. Это приводит к повторному квантованию коэффициентов DCT, причем матрица нормализации может быть отлич-

на от используемой при первом квантовании. Такой способ фальсификации для удобства изложения будем называть ниже F-2.

Задача заключается в выделении таких характерных особенностей СНЧ матриц блоков ЦИ, которые явно укажут на проведенную фальсификацию способом F-2.

Для наглядности результатов проводимых исследований поставим в соответствие ЦИ, матрица которого предварительно разбита на блоки, матрицу, каждый элемент которой равен количеству нулевых СНЧ в соответствующем блоке (МНСЧБ). В [14] были проанализированы и обоснованы характерные особенности этих матриц для ЦИ, хранимых без потерь, после квантования DCT-коэффициентов и ЧВ, после квантования и ПВ. Типичные примеры представлены на рис. 1 для изображения CELL.TIF, где выделены элементы, имеющие наименьшие значения, что позволяет наглядно увидеть соответствие между контурами ЦИ и блоками, содержащими наименьшее количество нулевых СНЧ [14]. МНСЧБ исходного изображения CELL.TIF – нулевая. В качестве матрицы нормализации, которая далее для удобства изложения обозначается KVANT, использовалась матрица, предложенная в [15].

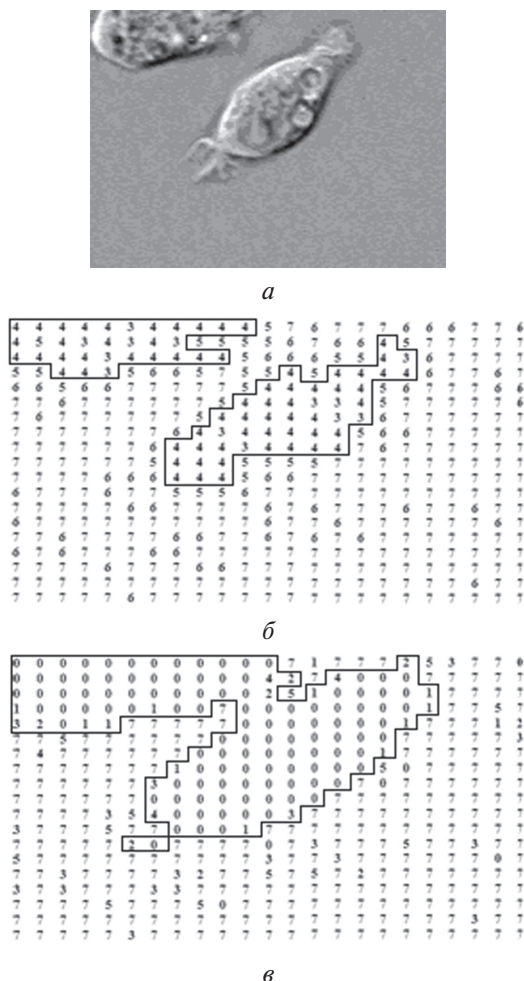


Рис. 1. Исходное ЦИ CELL.TIF (а); МНСЧБ после ЧВ (б); после ПВ (в)

Поскольку фальсифицированное изображение при фотомонтаже F-2 сохраняется в формате JPEG, определим характерные особенности СНЧ блоков при повторном квантовании коэффициентов DCT, которые впоследствии будут использованы для идентификации несанкционированного изменения ЦИ.

Элементы матрицы нормализации второго квантования, далее обозначаемой KVANT2, могут быть отличны от элементов KVANT, но это отличие не может быть значительным, поскольку матрицы нормализации строятся таким образом, чтобы учесть не только избыточность ЦИ с целью его сжатия, но и особенности зрительной системы человека, обеспечить отсутствие артефактов на изображении после его восстановления [15]. Перед квантованием (первым, вторым) используется стандартное разбиение матрицы ЦИ на блоки. Результаты для МНСЧБ изображения CELL приведены на рис. 2. Как видно при сопоставлении рис. 2а и 1б (рис. 2б и 1в), дающих типичную картину происходящего, повторное квантование практически не меняет МНСЧБ по сравнению с первым квантованием при соответствующих восстановлении. Такая картина, как показывает вычислительный эксперимент, проведенный в среде MATLAB, будет наблюдаться независимо от того, больше или меньше будут коэффициенты KVANT2 относительно соответствующих коэффициентов KVANT. Результат является естественным в силу следующих рассуждений.

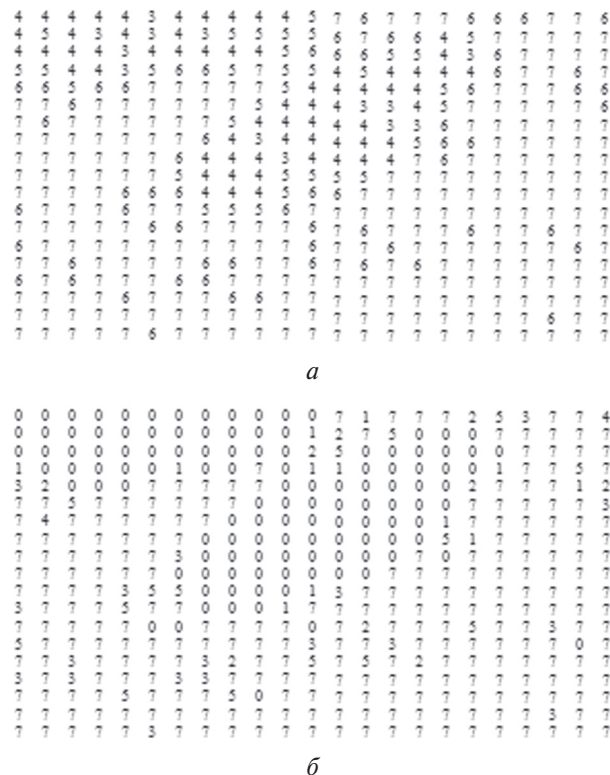


Рис. 2. МНСЧБ после второго квантования ЧВ (а); ПВ (б)

При проведении первого квантования обнулятся коэффициенты DCT, отвечающие высоким (и возможно, средним) частотам, что приведет к обнулению наименьших СНЧ матриц блоков [17], некоторые из которых после ПВ изображения снова станут ненулями, но будут лишь незначительно отличаться от нуля [14]. Это повлечет за собой в ПВ изображении малое отличие от нуля обнуленных при первом квантовании коэффициентов DCT, соответствующих высоким частотам. В силу этого повторное квантование в подавляющем большинстве случаев вновь обнулит обсуждаемые коэффициенты DCT, а значит, как следствие, соответствующие СНЧ, оставляя ненулевыми те, которые были ненулевыми после первого квантования. Таким образом, МНСЧБ, получаемые после одного и двух квантований, будут мало отличаться друг от друга как при частичном, так и при ПВ изображения, что и наблюдается на практике. Однако нужно отметить, что, очевидно, важным фактором здесь является совпадение сеток разбиения на блоки (СРБ) матрицы изображения при первом и втором квантовании.

Рассмотрим вариант, когда СРБ различны. Наличие нулевых СНЧ в блоке после первого квантования говорит о линейной зависимости столбцов (строк) его матрицы. Количество отличных от нуля СНЧ определяет ранг матрицы блока (максимальное количество линейно независимых столбцов (строк)). Изменим расположение СРБ в матрице ЦИ. Эта операция приводит к изменению МНСЧБ. Рассмотрим ее подробно. Пусть произошел сдвиг сетки на одну позицию вправо. Этот сдвиг исключит из блока один (первый) столбец и добавит новый, который окажется последним. Вновь введенный столбец, первоначально принадлежавший другому блоку, сформированный при его квантовании, как правило, оказывается линейно независимым со столбцами блока, в котором он очутился. Если при этом исключенный первый столбец был одним из тех, которые определяли ранг блока, то описанная сдвижка СРБ не изменит ранга матрицы. В противном случае, ранг матрицы блока увеличится на 1 (а количество нулевых СНЧ уменьшится на 1). Все сказанное выше будет иметь место и в том случае, если сдвиг сетки осуществляется на одну единицу вниз. Если сдвиг сетки происходит одновременно по обеим осям, на количество единиц, большее 1, то ничего определенного по поводу изменения значений МНСЧБ в общем случае сказать нельзя: как подключаемые, так и удаляемые столбцы (строки) и их части могут быть как линейно зависимыми, так и определять ранг. Однако, принимая во внимание, что для случайно выбранных m векторов из пространства размерности n , где $m < n$ (в рассматриваемом случае $m \leq 4$, $n = 8$), вероятность того, что они окажутся линейно зависимыми меньше, чем вероятность

того, что они окажутся линейно независимыми, логично предположить, что в большинстве случаев при описанной выше операции сдвига сетки это должно привести к неумению ранга матрицы (неувеличению количества нулевых СНЧ). И хотя векторы, фигурирующие в виде столбцов (строк) матриц блоков, подключаемые при сдвиге СРБ, нельзя в полной мере считать случайными, вычислительный эксперимент, проведенный в среде MATLAB, подтверждает неувеличение элементов МНСЧБ при операции сдвига. Типичная картина представлена на рис. 3 для изображения CELL (срав. с рис. 1б). В большинстве случаев количество нулевых СНЧ не увеличилось, однако, их уменьшение в общем случае, как и было отмечено выше, непредсказуемо (рис. 3б), приводит к значительному разбросу значений элементов МНСЧБ – от 1 до 7, что является характерной особенностью несовпадения СРБ матрицы. Заметим, однако, что упомянутое уменьшение и разброс значений элементов МНСЧБ характерен для областей, отвечающих фоновым частям исходного изображения. Что касается подобластей, содержащих многочисленные контуры, то они практически остаются неизменными при сдвиге СРБ: количество нулевых СНЧ таких блоков невелико первоначально (для рассматриваемого примера CELL нулевые СНЧ в них отсутствуют вообще (рис. 1б)), сдвиг сетки не меняет здесь картину линейной независимости столбцов (строк) матриц блоков, поскольку исключаемые из блока (подключаемые к блоку) столбцы (строки) или их части линейно независимы [14].



а



б

Рис. 3. МНСЧБ при сдвиге сетки разбиения на 1 по каждой оси (а); на 4 и 3 по осям ОХ и ОУ соответственно (б)

Интерес к поведению элементов МНСЧБ при изменении положения сетки разбиения матрицы ЦИ не случаен. Если фотомонтаж осуществляется так, как предполагалось выше, то есть ОИ и ЗО являются результатами одинарного квантования, то вероятность совпадения СРБ ОИ и ЗО после осуществления фотомонтажа мала, а значит, особенности МНСЧБ, возникающие за счет такого несовпадения, укажут на наличие фальсификации.

Учитывая конкретику рассматриваемого способа фальсификации F-2, продолжим исследования. В единожды отквантованном и ПВ изображении проведем сдвиг СРБ. Новое разбиение матрицы ЦИ используем для повторного квантования с ПВ. Пример результирующей МНСЧБ для изображения CELL представлен на рис. 4 (сдвиг сетки на 2 по каждой оси).

Если сравнить МНСЧБ с представленной на рис. 2б, то заметны значительные отличия. Полученный результат легко объясним, он является совокупностью двух причин: во-первых, сдвиг сетки по показанному выше приводит к значительному разбросу значений элементов МНСЧБ в фоновых подобластях изображения, во-вторых, из-за сдвига сетки структура матриц блоков, линейная независимость (зависимость) столбцов (строк) в пределах блока, являющаяся результатом первого квантования и ПВ изображения, разрушена. Проводимое повторное квантование по своей специфике является аналогом первого квантования с ПВ для исходного изображения: здесь появляется много блоков, имеющих малое число нулевых СНЧ и значительную часть ненулевых СНЧ, мало отличающихся от нуля [14]. Типичная картина наблюдается для блока (7,2), имеющего только одно нулевое СНЧ (рис. 4). Его сингулярный спектр определяется следующим образом: 960.3934, 3.6338, 1.0000, 0.7398, 0.6979, 0.4984, 0.2672, 0.0000. Значимым здесь очевидно является лишь максимальное СНЧ, все остальные лишь незначительно превышают ноль, являясь результатом округлений, выполняемых при ПВ изображения [14].

0	0	0	0	0	0	0	0	0	0	0	0	4	2	0	1	2	3	5	5	6
0	0	0	0	0	0	0	0	0	0	0	1	4	0	0	0	1	3	4	7	7
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	7	7
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	3	2	0
5	0	0	0	0	1	1	2	7	1	0	0	0	0	0	0	2	7	7	4	0
4	3	1	7	5	7	2	0	3	0	0	0	0	0	0	0	7	7	7	6	3
1	3	7	2	4	7	0	1	0	0	0	0	0	0	0	1	7	7	3	3	3
1	3	3	7	2	7	3	0	0	0	0	0	0	0	0	3	0	7	7	2	2
3	3	2	2	0	0	0	0	0	0	0	0	0	0	0	1	7	1	4	3	7
3	3	1	0	7	1	2	0	0	0	0	0	0	0	0	7	7	1	0	2	7
0	2	1	0	1	2	0	0	0	0	1	0	7	3	7	1	0	0	4	7	7
7	7	7	3	1	1	0	0	1	0	0	7	4	4	7	5	7	1	0	2	2
5	1	3	5	1	1	1	7	1	7	0	0	2	2	2	1	7	1	4	2	1
1	7	0	0	0	2	1	1	0	1	0	0	1	0	4	7	7	7	1	1	1
0	0	2	2	2	0	2	5	1	1	4	2	1	1	1	0	1	7	2	2	3
0	7	0	5	2	1	0	4	5	1	5	2	3	7	7	2	0	2	2	5	7
7	7	7	2	1	7	7	4	4	1	1	7	1	7	7	7	2	3	7	4	2
2	3	1	1	1	0	2	7	4	3	2	0	4	7	0	7	7	7	7	4	1

Рис. 4. МНСЧБ дважды квантованного ПВ
изображения при имеющемся сдвиге сетки разбиения

Воспользуемся выявленной особенностью СНЧ при повторном квантовании с несовпадающими СРБ для определения фотомонтажа F-2.

4. Метод обнаружения фотомонтажа F-2

Для наглядности изложения рассмотрим пример. Пусть изображение CELL подверглось JPEG-сжатию: один раз отквантовано и ПВ. На его основе формируется фотомонтаж CELL1, представленный на рис. 5а: ОИ и ЗО — части CELL, при этом замещающая область — это подобласть CELL, расположенная в нижнем правом углу, которая перемещена в верхний левый (такой прием, называемый копированием областей (region duplication), часто используется при фальсификации видео [19]). Хотя и ОИ, и ЗИ — части одной цифровой фотографии, СРБ их матриц, проделанных перед первым квантованием, после совершения фотомонтажа не совпадают. Если полученный фотомонтаж вновь сохранить в формате JPEG, что приведет к повторному квантованию и ПВ, что предполагает F-2, беря для этого СРБ стандартным образом (она совпадет с сеткой первого квантования только для ОИ и не совпадает для ЗО), то МНСЧБ будет выглядеть так, как представлено на рис. 5б. Внимание привлекает подобласть МНСЧБ, расположенная в ее левом верхнем углу, где без всякого «порядка» локализованы значения от 0 до 7. Такая ситуация для дважды квантованного и ПВ изображения, как следует из всего вышесказанного, принципиально возможна, если верхний левый угол изображения содержит контуры [14], что не соответствует действительности или, если имеет место несовпадение СРБ левой верхней подобласти изображения при первом и втором квантовании, которое возникло за счет фальсификации исходного ЦИ.

 a [illegible]

6

Рис. 5. Изображение CELL1 (а); МНСЧБ CELL1 (б)

Приведенный пример является наглядной иллюстрацией для обоснования основных шагов приведенного ниже метода обнаружения фотомонтажа F-2 ЦИ:

1. Разбить матрицу исследуемого ЦИ на блоки 8×8 .
2. Для каждого из полученных блоков определить множества СНЧ.
3. По результатам шага 2 построить МНСЧБ исследуемого ЦИ.
4. В МНСЧБ выделить связанные подобласти, для которых наблюдается значительный «беспорядочный» разброс значений элементов. Они могут соответствовать вклейкам либо областям исходного изображения, содержащим контуры.
5. Для выделенных на шаге 4 подобластей проанализировать сингулярные спектры всех блоков, соответствующих наименьшим значениям элементов МНСЧБ. Если среди этих блоков будет значительная часть таких, большая часть СНЧ которых сравнима с нулем, то данная подобласть соответствует вклейке.

Замечание 1. Схема JPEG-сжатия широко используется при хранении цифровых видеосигналов (стандарт MPEG), а часто используемые способы фальсификации видео приводят к двойному квантованию коэффициентов DCT матриц, соответствующих блокам кадров [20]. Таким образом, установленные особенности СНЧ и предложенный выше метод определения фальсификации, основанный на этих особенностях, может использоваться не только для обнаружения фальсификации цифровой фотографии, но и видео.

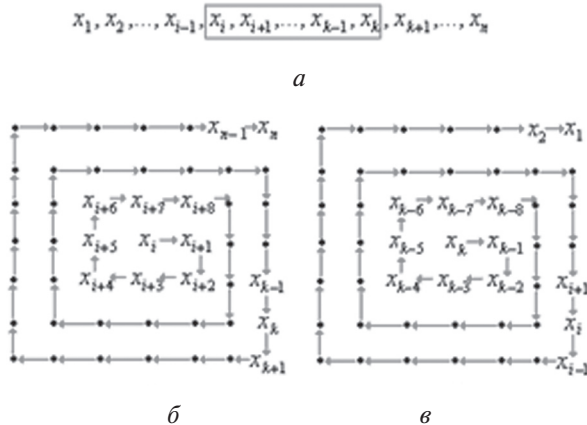


Рис. 6. Вектор, отвечающий ЦА (а); соответствующие матрицы, полученные спиралевидным способом (б, в)

Замечание 2. Предложенный метод обнаружения фальсификации ЦИ может быть использован и для ЦА, если фальсификация сигнала проводилась аналогичным описанному выше способом. Однако здесь важную роль будет играть способ перехода от векторного представления ЦА к матричному: необходимо, чтобы подозрительная с точки зрения фальсификации часть сигнала в соответс-

твующей матрице локализовалась компактно в связной области. Для обеспечения выдвинутого пожелания переход к матричному представлению имеет смысл производить спиралевидным способом, центр спирали будет соответствовать либо первому, либо последнему отсчету части звукового сигнала, проверяемой на подлинность (рис. 6). На рис. 6а «подозрительная» часть ЦА выделена.

Замечание 3. Основным недостатком предложенного подхода к решению проблемы обнаружения фальсификации ЦС, основанного на анализе СНЧ блоков соответствующих матриц, остается его несостоятельность для «улавливания» вклейки малого размера (копирования областей малых площадей). Если размеры (размер) вклейки сравним с размером блока, уловить описанные выше характерные особенности элементов МНСЧБ будет невозможно. Для увеличения эффективности предлагаемых методов обнаружения фальсификаций ЦС можно использовать прием, предложенный автором в [21]: матрице блока A , имеющей в общем случае произвольную структуру, ставятся в соответствие симметричные матрицы B, C той же размерности по правилу:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{18} \\ a_{21} & a_{22} & a_{23} & \dots & a_{28} \\ a_{31} & a_{32} & a_{33} & \dots & a_{38} \\ \dots & \dots & \dots & \dots & \dots \\ a_{81} & a_{82} & a_{83} & \dots & a_{88} \end{pmatrix} \rightarrow B = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{18} \\ a_{12} & a_{22} & a_{23} & \dots & a_{28} \\ a_{13} & a_{23} & a_{33} & \dots & a_{38} \\ \dots & \dots & \dots & \dots & \dots \\ a_{18} & a_{28} & a_{38} & \dots & a_{88} \end{pmatrix}, \quad (1)$$

$$C = \begin{pmatrix} a_{11} & a_{21} & a_{31} & \dots & a_{81} \\ a_{21} & a_{22} & a_{32} & \dots & a_{82} \\ a_{31} & a_{32} & a_{33} & \dots & a_{83} \\ \dots & \dots & \dots & \dots & \dots \\ a_{81} & a_{82} & a_{83} & \dots & a_{88} \end{pmatrix},$$

после чего анализируются собственные значения [18] B, C вместо СНЧ A . Соотношение (3) дает возможность увеличить вклад 3О в отдельный блок, если эта 3О занимала лишь часть A .

Выводы

В работе сформированы математические основы единого подхода к решению проблемы доказательства подлинности ЦС (ЦИ, ЦА, видео), детектирования его фальсификации независимо от вида сигнала и фальсифицирующего воздействия, основная идея которого заключается в сведении анализа ЦС, проводимого с целью выявления его несанкционированного изменения, к анализу сингулярных чисел соответствующей матрицы (матриц) для определения их характерных особенностей, сигнализирующих о фальсификации ЦС. В качестве математических инструментов использованы матричный анализ и теория возмущений.

Предложенный в работе подход адаптирован к решению конкретной задачи выявления фотомонтажа, результатом чего является разработка практического метода.

Таким образом, все поставленные задачи решены, цель работы достигнута.

Список литературы: 1. Куприянов А.И., Сахаров А.В., Шевцов В.А. Основы защиты информации. — М.: Изд. центр «Академия». — 2006. — 256 с. 2. Хорошко В.А., Чекатков А.А. Методы и средства защиты информации. — К.: Юниор, 2003. — 501 с. 3. Kundur D., Hatzinakos D. Digital watermarking for tell-tale tamper proofing and authentication // Proceedings of the IEEE. — 1999. — Vol. 87(7). — P. 1167-1180. 4. Fridrich J., Goljan M., Du M. Invertible authentication // In SPIE, Security and Watermarking of Multimedia Contents. — 2001. 5. Blythe P., Fridrich J. Secure digital camera // In Proc. Digital Forensic Research Workshop. — 2004. 6. Popescu A.C., Farid H. Exposing digital forgeries by detecting traces of re-sampling // IEEE Trans. Signal Process. — 2005. — Vol. 53(2). — P. 758-767. 7. Johnson M. K., Farid H. Exposing digital forgeries by detecting inconsistencies in lighting // In Proc. ACM Multimedia and Security Workshop. — New York. — 2005. — P. 1-10. 8. Bayram S., Sankur B., Memon N. Image manipulation detection // Journal of Electronic Imaging. — 2006. — Vol. 15(4). — P. 1-17. 9. He J., Lin Z., Wang L., Tang X. Detecting Doctored JPEG Images Via DCT Coefficient Analysis // ECCV(3). — 2006. — P. 423-435. 10. Wang W., Farid H. Exposing Digital Forgeries in Video by Detecting Double MPEG Compression // MM&Sec. — 2006. — P. 37-47. 11. Wang W., Farid H. Exposing Digital Forgeries in Video by Detecting Duplication // MM&Sec. — 2007. — P. 35-42. 12. Журавель В.В., Рыбальский О.В. К развитию теории выявления следов цифровой обработки сигналов // Захист інформації. — 2007. — № 1(32). — С. 83-85. 13. Рыбальский О.В. К экспериментальной проверке достоверности положений теории выявления следов цифровой обработки фонограмм // Реєстрація, зберігання та обробка даних. — 2004. — Т. 6, № 3. — С. 85-98. 14. Кобозева А.А.

Использование особенностей возмущений сингулярных чисел матрицы цифрового изображения для обнаружения его фальсификации // Искусственный интеллект. — 2008. — № 1. — С. 145-153. 15. Гонсалес Р., Вудс Р. Цифровая обработка изображений. — М.: Техносфера, 2005. — 1072 с. 16. Хорн Р., Джонсон Ч. Матричный анализ. — М.: Мир, 1989. — 656 с. 17. Кобозева А.А. Связь свойств стеганографического алгоритма и используемой им области контейнера для погружения секретной информации // Искусственный интеллект. — 2007. — № 4. — С. 531-538. 18. Деммель Дж. Вычислительная линейная алгебра. — М.: Мир, 2001. — 430 с. 19. Кобозева А.А. Применение сингулярного и спектрального разложения матриц в стеганографических алгоритмах // Вісник Східноукр. нац. ун-ту ім. В. Даля. — 2006. — № 9(103), Ч. 1. — С. 74-83.

Поступила в редколлегию 22.10.2008

УДК 004.056.5:518:512.624.3

Аналіз сингулярних чисел матриці цифрового сигналу як основа підходу до виявлення його фальсифікації / А.А. Кобозева // Біоніка інтелекту: наук.-техн. журнал — 2008. — № 2 (69). — С. 152-158.

Розроблено математичні основи загального підходу до проблеми доведення автентичності цифрового сигналу, виявлення його фальсифікації незалежно від виду сигналу і фальсифікуючої дії, обґрунтовано вибір використовуваних математичних інструментів, а також проведено адаптацію запропонованого загального підходу до рішення конкретної задачі.

Л. 5. Бібліогр.: 19 найм.

UDK 004.056.5:518:512.624.3

Analysis of Digital Signal Matrix Singular Values as Foundation for its Forgery Detection / A.A. Kobozeva / Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 152-158.

New approach to a problem of digital signal forgery detection based on analysis of singular values disturbances of digital signal matrix is proposed in this article. This approach may be used for any digital signal.

Fig. 5. Ref.: 19 items.

УДК 519.62



ОЦЕНКА ПАРАМЕТРОВ ЧИСЛЕННОСТИ ПОПУЛЯЦИИ В ДИСКРЕТНОЙ МОДЕЛИ РИКЕРА

С.Н. Герасин¹, М.А. Козлов²

^{1,2} ХНУРЭ, г. Харьков, Украина

В статье рассматриваются различные дискретные популяционные модели. Исследуется поведение модели Рикера в области изменения параметров. Получены оценки параметров модели Рикера по зашумленным наблюдениям для случая, когда ошибка внесена в модель мультипликативно.

ПОПУЛЯЦИЯ, МОДЕЛЬ РИКЕРА, ДИСКРЕТНАЯ МОДЕЛЬ ФЕРХЮЛЬСТА, ОЦЕНКА ПАРАМЕТРОВ МОДЕЛИ

Введение

Жизненные циклы многих биологических видов имеют четко выраженную временную периодичность, как правило, приуроченную к сезонной периодичности климата земли. Подавляющее большинство таких видов имеют точно очерченный (часто весьма небольшой) период размножения, во время которого каждая локальная популяция представляет собой совокупность дискретных, не пересекающихся возрастных классов [1].

Для многих видов насекомых, некоторых видов членистоногих, моллюсков, рыб, земноводных и пресмыкающихся каждая отдельная популяция представляет собой один возрастной класс (или стадию развития) и смежные поколения не перекрываются. Это обуславливается тем, что за время развития очередного поколения (от яйца до имаго) все предыдущее поколение, представленное в период размножения взрослыми особями, успевает вымереть. Динамика численности популяций таких видов с неперекрывающимися поколениями описывается особенно просто. Если условия среды от поколения к поколению меняются не очень сильно и можно сделать предположение об их постоянстве, то численность некоторого поколения будет определяться только численностью предыдущего поколения. Выбрав определенные моменты времени для регистрации этих численностей (например, начало какой-либо стадии развития) и обозначив через N_n численность n -го поколения, можно записать следующее детерминистическое уравнение, описывающее динамику численности такой одно возрастной популяции с неперекрывающимися поколениями:

$$N_{n+1} = F(N_n). \quad (1)$$

По сравнению с дифференциальными уравнениями модели с дискретным временем («дискретные» модели) типа (1) должны давать более реальное описание динамики численности популяции по крайней мере, для тех биологических видов, поколения которых не перекрываются. В начале 70-х годов появились достаточно глубокие исследования дискретных моделей динамики численности,

которые сильно изменили наши представления о роли плотностно зависимых лимитируемых факторов в поддержании устойчивости одновидовых сообществ [2].

1. Дискретные популяционные модели Ферхюльста и Рикера

Среди различных дискретных моделей популяционной динамики одной из наиболее простых является модель Ферхюльста с дискретным временем

$$N_{n+1} = aN_n(1 - N_n) = aN_n(1 - N_n/M), \quad (2)$$

где $\beta = b/a$ или $M = a/b$. Эта модель в такой или какой-либо иной форме рассматривается многими авторами [3]. Существенный ее недостаток заключается в том, что она предсказывает отрицательное значение численности в некотором поколении, если численность в предыдущем поколении была слишком велика (больше некоторой фиксированной величины $M = 1/b$) или если репродуктивный потенциал слишком высок ($a > 4$), а начальное значение численности было около $M/2$. Однако при $a < 4$ и при $N_0 < M$ величина численности популяции N_n не может превзойти M (а соответственно и оказаться отрицательной) ни при каких n и модель (2) оказывается корректной. Эта модель относится к типу логистического уравнения [4].

Подобные некорректности исключены в другом дискретном варианте логистического уравнения, предложенном в 1954 г. известным ихтиологом У.Е. Рикером [5]. Он изучал связь между запасом (численностью родительского стада) и пополнением (численностью потомков, доживших до репродуктивного возраста) в рыбной популяции и пришел к следующей зависимости:

$$N_{n+1} = aN_n e^{-bN_n}, \quad (3)$$

где N_n — численность запаса, а N_{n+1} — численность пополнения. Для одновозрастной популяции (например, популяции горбуши) это соответствует численностям популяций в n -ом и $(n+1)$ -м поколении соответственно. Параметр a в модели (3) характеризует скорость роста популяции в отсутствии

лимитирования, а параметр b связан с емкостью экологической ниши данной популяции.

Зависимость (3) можно обосновать, например, из следующих соображений. Рассмотрим популяцию со строго периодическим характером процесса размножения, предполагая при этом, что периоды размножения сравнительно кратковременны и строго регулярны (дискретны) во времени, гибель особей происходит непрерывно во времени между периодами размножения, а после размножения все родительское стадо погибает, как это наблюдается у лососей. Будем считать, что коэффициент рождаемости p (среднее число рожденных потомков, приходящихся на одну особь родительского стада) постоянен, то есть не меняется от поколения к поколению, а коэффициент мгновенной смертности β линейно зависит от числа рожденных особей R_0 : $\beta = \beta_0 + kR_0$. В этой формуле первое слагаемое отражает постоянную естественную смертность молодежи, а второе — смертность, обусловленную ограниченностью жизненных ресурсов, требуемых на начальных этапах развития молодежи. Возможна и другая интерпретация линейной связи между коэффициентом смертности и числом рожденных особей. Рикер, например, объяснял ее давлением хищников.

Пусть N_n — численность репродуктивной части популяции в n -ом поколении, тогда, согласно нашему предположению, численность родившихся потомков R_0 равна pN_n . Обозначим через $R(\tau)$ — численность потомков, доживших до возраста τ . Ясно, что $R(0) = R_0$. Поскольку за время τ , прошедшее после размножения, численность популяции непрерывно уменьшалась за счет смертности, то для $R(\tau)$ справедливо уравнение Мальтуса (без рождаемости) $R'(\tau) = -\beta R(\tau)$ с начальным условием $R(0) = R_0$ [4]. Известно, что решение этого уравнения имеет вид $R(\tau) = R_0 e^{-\beta\tau}$. Обозначим через T шаг дискретной модели — время между двумя периодами размножения. За это время все оставшиеся потомки достигают репродуктивного возраста и сами приступают к размножению, следовательно $R(T) = N_{n+1}$ или

$$N_{n+1} = R_0 e^{-\beta T}.$$

Подставляя сюда $\beta = \beta_0 + kR_0$ и $R_0 = pN_n$, получаем:

$$N_{n+1} = pN_n e^{-(\beta_0 + kpN_n)T} = pN_n e^{-\beta_0 T} e^{kpN_n T}.$$

Вводя обозначения $a = pe^{-\beta_0 T}$ и $b = kpT$, приходим к формуле (3):

$$N_{n+1} = aN_n e^{-bN_n}.$$

Проведем исследование модели Рикера. Легко видеть, что параметр b в уравнении (3) не влияет на характер динамического поведения. Действительно, если ввести преобразование переменной $x = bN$, то вместо (3) получаем

$$x_{n+1} = ax_n e^{-x_n}. \quad (4)$$

Таким образом, можно без ограничения общности заменить b на 1 и предполагать, что численность измеряется не в абсолютных, а в относительных единицах, которые меньше реальной численности в $1/b$ раз (c как правило весьма мало, существенно меньше 1). Рассмотрим характер поведения x_n при разных значениях a .

2. Стационарные точки модели Рикера

Хорошо известно, что достаточно долгий экспоненциальный рост численности в природе никогда не наблюдается. Рано или поздно сказывается действие лимитирующих факторов. В случае, когда подобное лимитирование целиком определяется действием плотностно зависимых факторов, имеет место следующая общая модель:

$$N_{n+1} = aN_n f(N_n), \quad (5)$$

где $f(N)$ — некоторая, как правило, убывающая функция численности, такая, что $f(0) = 1$, и a — репродуктивный потенциал популяции — максимальная скорость роста популяции (максимальное число выживших потомков, приходящееся на одну особь родительского поколения) или скорость роста популяции в пустоту, так как $q(0) = a$.

Начнем с нахождения равновесных (стационарных) значений численности (неподвижных точек), то есть таких значений N , при которых $N_{n+1} = N_n = N$. Эти значения численности удовлетворяют следующему алгебраическому уравнению: $\bar{N} = F(\bar{N})$, то есть соответствуют точкам пересечения графика функции $F(N)$ и биссектрисы первого координатного угла (рис. 1). Для (5) имеем:

$$\bar{N} = a\bar{N}f(\bar{N}). \quad (6)$$

У этого уравнения есть одно очевидное «тривиальное» решение $\bar{N} = 0$. Для оставшихся решений ($\bar{N} \neq 0$) имеем $1 = af(\bar{N})$ или

$$f(\bar{N}) = 1/a.$$

Если функция $f(N)$ убывающая и $f(0) = 1$, то график функции $f(N)$ целиком лежит под прямой $f \equiv 1/a$ при $a < 1$ и пересекает эту прямую в единственной точке $N = \bar{N}$ при $a > 1$ (рис. 1).

Таким образом, при $a < 1$ уравнение (5) имеет единственное (нулевое) положение равновесия, а при $a > 1$ существуют два положения равновесия: 1) $\bar{N} = 0$ и 2) $\bar{N}$ удовлетворяет уравнению (6).

Теперь перейдем к нахождению равновесных значений численности и анализа их устойчивости. Воспользуемся результатами, полученными при исследовании общего случая, учитывая что

$$F(x) = axe^{-x}, \quad f(x) = e^{-x} \quad \text{и} \quad F'(x) = ae^{-x}(1-x).$$

При $a < 1$ уравнение (4) имеет одно глобально устойчивое положение равновесия $\bar{x} = 0$, то есть для всех x_0 ($0 < x_0 < 8$) $x_n \rightarrow 0$, причем $x_{n+1} < x_n$.

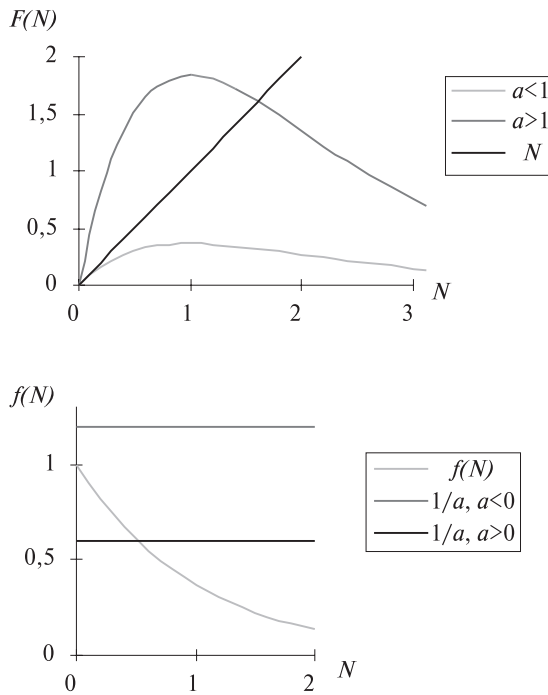


Рис. 1. Графики функций $F(N)$, $f(N)$ и прямых $1/a$ при $a < 1$ и $a > 1$

При $a > 1$ уравнение (4) имеет два равновесных решения. Тривиальное $\bar{x}_1 = 0$, неустойчивое при всех значениях $a > 1$, и нетривиальное, определяемое из уравнения $e^{-\bar{x}_2} = 1/a$, которое может быть как устойчивым, так и неустойчивым. Исследуем это решение.

Имеем: $\bar{x}_2 = \ln a$, $F'(\bar{x}_2) = 1 - \ln a$. Возможны следующие три ситуации:

1) $1 < a < e$. В этом случае $0 < F'(\bar{x}_2) < 1$, следовательно, неподвижная точка $\bar{x}_2$ устойчива и переход к равновесию вблизи $\bar{x}_2$ осуществляется монотонно.

2) $e < a < e^2$. В этом случае $-1 < F'(\bar{x}_2) < 0$. Неподвижная точка $\bar{x}_2$ устойчива и переход к равновесию осуществляется путем затухающих колебаний около $\bar{x}_2$.

3) $a > e^2$. В этом случае $F'(\bar{x}_2) < -1$. Неподвижная точка $\bar{x}_2$ неустойчива и вблизи этой точки поведение численности имеет вид расходящихся колебаний.

Для случаев 1 и 2 нетрудно показать, что точка $\bar{x}_2$ является устойчивой в целом для луча $(0, \infty)$, то есть, если $1 < a < e^2$, то $x_n \rightarrow \bar{x}_2$, при $0 < x_0 < \infty$ [6].

3. Циклы и хаотический режим

При $a > e^2$ нет устойчивых равновесных точек. При переходе параметра a через точку бифуркации $a = a_1 = e^2 \approx 7,39$ точка $\bar{x}_2$ из устойчивой превращается в неустойчивую и из нее рождается устойчивый 2-цикл. Аналитическое исследование границ устойчивости этого и последующих циклов оказывается достаточно сложно. Точки бифуркаций циклов можно найти только с помощью компьютера. При $a = a_2 \approx 12,49$ 2-цикл из устойчивого превращается

в неустойчивый и от него рождается устойчивый 4-цикл. При $a = a_3 \approx 14,24$ 4-цикл теряет устойчивость и от него рождается устойчивый 8-цикл, при $a = a_4 \approx 14,68$ 8-цикл теряет устойчивость и от него рождается устойчивый 16-цикл, при $a = a_5 \approx 14,75$ 16-цикл теряет устойчивость и от него рождается устойчивый 32-цикл и так далее.

Первая серия бифуркаций заканчивается в модели Рикера в точке накопления $a^* \approx 14,77$, а последовательность значений $\{a_k\}$, при которых происходит удвоение периода, удовлетворяет закону Фейгенбаума [3]. Так же, как и в предыдущей модели, когда параметр a превосходит точку накопления a^* , появляются области его значений, в которых поведение численности популяции теряет сколько-нибудь регулярный характер и становится хаотическим. Однако при возрастании a за точкой накопления зоны хаотического поведения численности перемежаются с «окнами» периодического, то есть регулярного поведения.

Так при $\sim 22,25 < a < \sim 23,49$ существует устойчивый цикл длины 3, который переходит в устойчивый 6-цикл. При $a \approx 24,26$ 6-цикл теряет устойчивость и от него рождается устойчивый 12-цикл, при $a \approx 24,40$ 12-цикл теряет устойчивость и от него рождается устойчивый 24-цикл так далее [7].

4. Оценка параметров модели Рикера по неточным наблюдениям. Постановка задачи

Среди рекуррентных моделей вида (1) для математической экологии большой интерес представляет модель численности популяций — модель Рикера:

$$x_{n+1} = ax_n e^{-bx_n}, \quad n = 0, 1, \dots, \quad (7)$$

где $x_0 > 0$, а параметры a, b удовлетворяют условиям:

$$a > 1, \quad b > 0. \quad (8)$$

Эта модель подробно рассмотрена в [3], [4].

В этих работах указаны области значений параметра (a, b) , при которых последовательность $x_n, n = 0, 1, \dots$ имеет либо притягивающий цикл, либо предельное распределение. Особое внимание здесь уделяется зависимости длины притягивающего цикла от параметра (a, b) .

Для модели (7)–(8) и практически, и теоретически важно оценивать параметры a и b по неточным наблюдениям, $y_n, n = 0, 1, \dots$ за состоянием последовательности $x_n, n = 0, 1, \dots$. Эта задача является непростой статистической задачей. В силу нелинейности соотношения (5) применение к оценке параметра (a, b) метода наименьших квадратов (широко используемого для оценки параметров случайных процессов в линейных системах) не вполне удачно. Это объясняется тем, что (как обнаружили экологи, обсчитывая модель Рикера на машине) при некоторых значениях параметров a и b последовательность $x_n, n = 0, 1, \dots$ обнаруживает псевдослучайный характер.

То есть сама детерминированная траектория обладает свойством псевдослучайности и неустойчивости. И эта неустойчивость привела к тому, что при включении еще и ошибок в наблюдения, качество оценок параметров модели Рикера, полученных по методу наименьших квадратов, стало невысоким. Обращение же к эргодическим свойствам модели Рикера дало существенно более точные оценки параметров a и b .

В настоящей работе строится статистическая оценка параметра (a, b) рекуррентной модели Рикера (7) (8) по наблюдениям y_n , $n = 0, 1, \dots$ за значениями x_n , $n = 0, 1, \dots$. При этом выбираем мультипликативную модель внесения ошибок ε_n , $n = 0, 1, \dots$ в наблюдения y_n :

$$y_n = x_n e^{\varepsilon_n}, \quad n = 0, 1, \dots, \quad (9)$$

более свойственную задачам оценки численности популяций.

Здесь ε_n , $n = 0, 1, \dots$ — последовательность независимых, нормально распределенных случайных величин со средним 0 и известной дисперсией σ^2 .

При этом используются такие качественные свойства последовательности (9), как существование у нее предельного цикла длины $q \geq 1$ или предельного распределения [4]. Использование этих свойств последовательности x_n для оценки параметра (a, b) по неточным наблюдениям y_n , $n = 0, 1, \dots$ является целью настоящей статьи.

5. Построение уравнений для вычисления параметров a и b через $\bar{f}(x)$ при разных $f(x)$

Получим уравнения для вычисления параметров модели Рикера для детерминированного (без случайных возмущений) случая методом средних по траектории.

Утверждение 1. Если последовательность (7) имеет предельные циклы, то справедливы соотношения:

$$\ln a = b\bar{x}, \quad (10)$$

$$b = 2 \frac{\overline{x \ln x} - \bar{x} \cdot \overline{\ln x}}{\overline{x^2} - \bar{x}^2}. \quad (11)$$

Доказательство. Логарифмируем соотношение (7).

$$\ln x_{n+1} = \ln a + \ln x_n - b x_n.$$

Тогда

$$\ln^2 x_{n+1} = \ln^2 a + \ln^2 x_n + b^2 x_n^2 + 2 \ln a \ln x_n - 2 \ln a \cdot b x_n - 2 b x_n \ln x_n.$$

Суммируем два последних равенства по n .

$$\frac{1}{N} \sum_{n=0}^{N-1} \ln x_{n+1} = \ln a + \frac{1}{N} \sum_{n=0}^{N-1} \ln x_n - b \cdot \frac{1}{N} \sum_{n=0}^{N-1} x_n,$$

$$\frac{1}{N} \sum_{n=0}^{N-1} \ln^2 x_{n+1} = \ln^2 a + \frac{1}{N} \sum_{n=0}^{N-1} \ln^2 x_n +$$

$$+ b^2 \cdot \frac{1}{N} \sum_{n=0}^{N-1} x_n^2 + 2 \ln a \cdot \frac{1}{N} \sum_{n=0}^{N-1} \ln x_n -$$

$$- 2 \ln a \cdot b \cdot \frac{1}{N} \sum_{n=0}^{N-1} x_n - 2 b \cdot \frac{1}{N} \sum_{n=0}^{N-1} x_n \cdot \ln x_n.$$

Переходя к пределу при $N \rightarrow \infty$ и используя обозначения в виде верхних черточек для средних значений, получаем следующие два уравнения:

$$\overline{\ln x} = \ln a + \overline{\ln x} - b \cdot \bar{x},$$

$$\overline{\ln^2 x} = \ln^2 a + \overline{\ln^2 x} + b^2 \overline{x^2} + 2 \ln a \overline{\ln x} - 2 \ln a \cdot b \cdot \bar{x} - 2 b \cdot \bar{x} \overline{\ln x}.$$

Из первого уравнения сразу получаем уравнение (7):

$$\ln a = b\bar{x}.$$

После упрощения второе уравнение принимает вид:

$$0 = \ln^2 a + b^2 \overline{x^2} + 2 \ln a \cdot \overline{\ln x} - 2 \ln a \cdot b \cdot \bar{x} - 2 b \cdot \bar{x} \overline{\ln x}.$$

Если подставить в последнее равенство выражение для $\ln a$ из (7), после преобразований получаем:

$$0 = b^2 (\overline{x^2} - \bar{x}^2) + 2b (\bar{x} \cdot \overline{\ln x} - \overline{x \ln x}).$$

Из данного равенства получаем уравнение (8) для вычисления параметра b :

$$b = 2 \frac{\overline{x \ln x} - \bar{x} \cdot \overline{\ln x}}{\overline{x^2} - \bar{x}^2}.$$

Утверждение доказано.

Замечание. Соотношения (7), (8) являются основой для оценки параметров a, b по наблюдениям y_n , $n \leq 0$ за последовательностью x_n , $n \leq 0$.

6. Оценка параметров a и b по наблюдениям y_n .

Включение в задачу случайных ошибок

Рассмотрим мультипликативную модель внесения ошибок в наблюдения y_n , $n = 0, 1, \dots$

$$y_n = x_n e^{\varepsilon_n}. \quad (12)$$

В рамках условий утверждения 1 и равенств (9) справедливы следующие оценки в пределе при $n \rightarrow \infty$:

$$MY_n = M \frac{1}{n} \sum_{i=0}^{n-1} y_i \rightarrow \bar{x} \cdot e^{\sigma^2/2}, \quad (13)$$

$$M(\ln Y)_n = M \frac{1}{n} \sum_{i=0}^{n-1} \ln y_i \rightarrow \overline{\ln x}, \quad (14)$$

$$M(Y \ln Y)_n = M \frac{1}{n} \sum_{i=0}^{n-1} y_i \ln y_i \rightarrow$$

$$\rightarrow (x \cdot \ln x + \sigma^2 \cdot \bar{x}) \cdot e^{\sigma^2/2},$$

$$\bar{x} = \lim_{n \rightarrow \infty} MY_n \cdot e^{-\sigma^2/2},$$

$$\overline{x^2} = \lim_{n \rightarrow \infty} M(Y^2)_n \cdot e^{-2\sigma^2},$$

$$\overline{\ln x} = \lim_{n \rightarrow \infty} M(\ln Y)_n,$$

$$\overline{x \ln x} = \lim_{n \rightarrow \infty} M((Y \cdot \ln Y)_n - \sigma^2 \cdot Y_n) \cdot e^{-\sigma^2/2}.$$

Последние четыре оценки есть прямые следствия оценок (13)-(15).

Из приведенных равенств очевидно: выполняются следующие соотношения в пределе при $n \rightarrow \infty$:

$$MY_n \cdot e^{-\sigma^2/2} \rightarrow \bar{x},$$

$$M(Y^2)_n \cdot e^{-2\sigma^2} \rightarrow \overline{x^2},$$

$$M(\ln Y)_n \rightarrow \overline{\ln x},$$

$$Me^{-\sigma^2/2} ((Y \cdot \ln Y)_n - \sigma^2 \cdot Y_n) \rightarrow \overline{x \ln x}.$$

Учитывая данные соотношения в уравнениях (10) и (11), получаем в качестве оценки параметров a и b по наблюдениям y_n , $n = 0, 1, \dots$ величины $\tilde{a}_n$ и $\tilde{b}_n$ соответственно, которые определяются следующими двумя равенствами:

$$\tilde{a}_n = \exp\left(\tilde{b}_n \cdot Y_n \cdot e^{-\sigma^2/2}\right), \quad (9.7)$$

$$\tilde{b}_n = 2e^{\sigma^2/2} \cdot \frac{(Y \ln Y)_n - Y_n \cdot \sigma^2 - Y_n \cdot (\ln Y)_n}{(Y^2)_n \cdot e^{-\sigma^2} - (Y_n)^2}. \quad (9.8)$$

Выводы

Из результатов данной работы видно, что даже в самых простых дискретных моделях динамики численности (точнее, моделях с дискретным временем) появляются нерегулярные, хаотические режимы динамического поведения популяции. Подобные динамические режимы ранее были обнаружены только в стохастических моделях, описывающих изменения численности популяции под действием случайных факторов. То есть, периодическое или хаотическое изменение численности может быть получено в детерминистических моделях и определяется чисто детерминистическими факторами, если среди этих факторов присутствует сезонность процессов размножения и плотностно зависимое экологическое лимитирование.

Для модели Рикера в случае, когда наблюдаемые величины зашумлены, получены статистические оценки параметров.

Список литературы: 1. Ризниченко Г.Ю., Рубин А.Б. Математические модели биологических продукционных процессов. — М.: МГУ, 1993. — 302 с. 2. Скалецкая Е.И., Фрисман Е.Я., Шапиро А.П. Дискретные модели численности популяций и оптимизация промысла. — М.: Наука, 1979. — 166 с. 3. Шапиро А.П., Луппов С.П. Реккурентные уравнения в теории популяционной биологии. М.: Наука, 1983. — 134 с. 4. Ляшенко И. М., Мукоед А. П. Моделі біологічних та екологічних процесів. — К.: Вид-во КНУ, 2002. — 450 с. 5. Рикер У.Е. Методы оценки и интерпретация биологических показателей популяций рыб. — М.: Пищевая промышленность, 1979. — 408 с. 6. Свирежнев Ю.М., Логофет Д.О. Устойчивость биологических сообществ. — М.: Наука, 1978. — 352 с. 7. Пых Ю.А. Равновесие и устойчивость в моделях популяционной динамики. — М.: Наука, 1983. — 184 с.

Поступила в редколлегию 23.10.2008

УДК 512.643.8

Оцінка параметрів чисельності популяції в дискретній моделі Рікера // С.М. Герасін, М.О. Козлов / Біоніка інтелекту: наук.-техн. журнал — 2008. — № 2 (69). — С. 159-163.

У статті розглядаються різні дискретні моделі популяцій. Досліджується поведінка моделі Рікера в області зміни параметрів. Отримано оцінки параметрів моделі Рікера за збуреними спостереженнями для випадку, коли помилка внесена до моделі мультиплікативно.

Лл.1. Бібліогр.: 7 найм.

UDK 512.643.8

Estimation of quantity parameters of population is in discrete Riker model / S.N. Gerasin, M.A. Koslov // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 159-163.

Different discrete population models are examined in the article. The conduct of model of Rikera is probed in area of change of parameters. The estimations of parameters of model of Rikera are got on supervisions for a case, when an error is borne in a model multiplicative.

Fig.1. Ref.: 7 items.

УДК 007.681.5:519.714:519.766



КОНЦЕПЦІЇ ОРГАНІЗАЦІЇ ІНФОРМАЦІЙНО-ІНТЕЛЕКТУАЛЬНИХ ТЕХНОЛОГІЙ

З.Д. Коноплянко

Львівський інститут банківської справи Університету банківської справи
Національного банку України, м. Львів, zenovij@ukr.net

Стаття присвячена аналізу проблеми створення систем штучного інтелекту, які дозволяють моделювати рішення на логічному та апаратному рівнях і які є елементами k -значної структурної організації інформаційно-інтелектуальних технологій.

БАЗА ЗНАНЬ, ЛОГІКА, ПРИРОДНА МОВА, ШТУЧНИЙ ІНТЕЛЕКТ, k -ЗНАЧНА СТРУКТУРА

Вступ

У ХХІ столітті виникла матеріально-енергетична та соціально-економічна криза суспільного виробництва, виходом із якої став перехід економічно розвинутих країн від матеріально-енергетичних технологій виробництва та споживання до інформаційно-інтелектуальних технологій, де економія матеріально-енергетичних ресурсів досягається шляхом переходу від опрацювання інформації до обробки знань. Інтелектуальна підтримка суспільно-економічних процесів дозволяє раціонально розв'язати триєдине завдання прогресу сучасної цивілізації — підвищення ефективності, гуманістичності та екологічності виробництва [1-3].

Найшвидші суперкомп'ютери сьогодні працюють набагато повільніше, ніж це потрібно для роботи з даними і знаннями, поданими мовними моделями, що описують описові науки. Розв'язання цієї проблеми залежить від того, чи будуть створені відповідні комп'ютерні системи, здатні працювати зі швидкістю у 1000 разів більшою, аніж існуючі найбільші суперкомп'ютери [1-5]. Поняття «нові інформаційні технології» передбачає вміння користуватися сучасними засобами та методами добування, опрацювання і систематизації знань.

1. Мета роботи

Основним завданням цієї роботи є виклад новостворюваної концепції організації інформаційно-інтелектуальних технологій та інтелектуальної підтримки суспільно-економічних процесів, зокрема діяльності фінансово-кредитних установ. Предметом досліджень є моделювання інтелектуальної діяльності людей як у зовнішньому її прояві (вирішення складних завдань, розуміння природної мови, інтерпретація візуальної інформації та мови), так і у внутрішньому (накопичення, надання і використання знань).

2. Основні поняття та визначення

У теорії пізнання постулюються для новостворюваних теорій два основоположних принципи: «якщо існує об'єкт, то його можна досліджувати» та «щоб уникнути труднощів взаєморозуміння,

необхідно перш за все порозумітись із основними визначеннями», які стосуються даної теорії. Відповідно, *наочною чи проблемною галуззю* ми називатимемо сукупність даних і знань, що є достатнім для інформаційного (заданого на контекстно-залежній мові) забезпечення рішення деякої задачі чи сукупності завдань інтелектуальної системи.

Складність поняття *знань*, завжди виражається на деякій мові відношень, полягає у множинності можливостей її реалізації і нерозривності з поняттям *дані*, в безперервному процесі їх зміни, врахування якого забезпечує реалізацію контекстного зв'язку даних.

Відзначимо сферу предметних галузей, де найдоцільніше працювати з даними і знаннями, поданими мовними моделями. Це галузі з переважанням емпіричного знань, де складність фактів і їх описів виключає використання мови математики — так звані описові науки, які, загалом, і стали такими, бо вивчають те, що не може бути адекватно описане мовою математики — відкриті системи в їх природному оточенні, системи, спостережувані на рівні апарату лінгвістики. Як буде видно з подальшого, саме такий підхід дозволяє добиватися найбільшого ступеня адекватності опису тієї чи іншої системи, особливо при введенні поняття зміни формально-логічних властивостей лінгвістичного апарату під дією вхідних інформаційних потоків.

Сказане вище означає, що сукупність знань—дані в завданнях управління подається деякою семіотичною системою. Взагалі кажучи, під семіотикою розуміється наука про знаки і знакові системи, а також про мови як знакові системи, причому загальний підхід до мовних знакових систем, що включає не тільки традиційне поняття мови, але і мови запахів, свисту, абстрактного живопису і тому подібне, вимагає ще свого теоретичного обґрунтування. Мовою може бути все, що забезпечує виникнення і існування комунікативної функції для своїх користувачів.

У семіотичній системі зазвичай виділяють три аспекти: *синтаксичний, семантичний і прагматич-*

ний. Такого роду розділення цікавить нас із наступної причини. Відповідно до перерахованих трьох аспектів семіотичних систем необхідно виділити *три типи знань, як три типи відношень між даними*: синтаксичні, семантичні і прагматичні.

Знання синтаксичного типу характеризує синтаксичну структуру потоку інформації, яка не залежить від сенсу і змісту використовуваних при цьому понять, тобто інтелектуальну систему не утворює.

Семантичне знання розглядається як структура, утворююча поточний контекст. Воно містить інформацію, безпосередньо пов'язану з поточними значеннями і сенсом описуваних понять і зумовлює стан зв'язків даних в інформаційній базі.

Прагматичне знання зумовлює найбільш вірогідні зв'язки, що описують дані з погляду вирішуваної задачі (узагальнений чи «об'єктивний» контекст), наприклад, із урахуванням специфічних критеріїв і угод, що діють у даному завданні.

Поняття об'єктивності відповідає трактуванню прагматики, прагматичного аспекту створення інтелектуальної системи як спрямованого обмеження її «свободи волі», обмеження непотрібних зв'язків і «фантазії зв'язків за контекстом» для системи, від якої ми не чекаємо надлишку інтелектуальності. Для порівняння пригадаємо, що ми маємо на увазі, коли даємо людині характеристику «прагматик».

З інженерної точки зору, синтаксична, семантична і прагматична сторони знань є різними зв'язками одного чи групи термінів (даних) з іншими записами в інформаційній базі.

Зі сказаного зрозуміло, що найбільш загальна проблема побудови системи управління семантичного чи семантико-прагматичного рівня взаємодії пов'язана з вибором технології контекстно-залежного подання знань, побудовою інформаційних баз (даних і знань) про предметну галузь і механізми висновку для отримання необхідних рішень.

Логіка цього механізму може мінятися залежно від контекстів взаємозв'язків знаків (їх прагматики), що є однією з найважливіших причин початкової організації всіх практичних робіт тільки на проблемно-орієнтованій підмножині контекстно-залежної мови, тобто там, де ми сподіваємося на деяку постійність контексту. Ця умова є корисною тільки в сенсі забезпечення деякої початкової простоти побудов. Але, на відміну від «чистих формально-логічних систем», умова постійності контексту не є для нас обов'язковою. Всі наші побудови орієнтовані, кінець кінцем, саме на мінливості поточного контексту, без чого ми не маємо права говорити про інтелектуальні системи.

Структурування мови йде через формування базового набору відношень (їх число не більше 200): часових, просторових, казуальних, квантифікуючих (бінарні, тернарні), тощо. Метапроцедури — це цілеспрямований пошук у лабіринті можливостей,

побудова чи пошук нових лабіринтів, структурування, декомпозиція задач на підзадачі, формування закономірностей на підставі спостережень. У основі наших знань про світ лежать системи класифікації, які дозволяють нам сприймати світ як структурований за мірою загальності об'єкт.

3. Оброблення мов декларативного типу

Мови декларативного типу зручні для опису (подання) знань у системах ШІ [6-10]. Оброблення мови декларативного типу можна розбити на дві частини:

- те, що декларує людина;
- те, що опрацьовує ЕОМ.

Попереднє оброблення людиною полягає в абстрагуванні реальної проблеми у фізичну чи математичну модель. У подальшому математична модель перетворюється у машинно-орієнтоване відображення, яке сприймає ЕОМ. Складання відповідної програми для перетворення форм подання інформації можливе лише з допомогою семантичного опрацювання. ЕОМ при цьому здійснює лише підтримку оброблення, що його здійснює людина.

Основу такої машинної мови складають:

- основні поняття;
- семантика (на основі відповідного синтаксису);
- процедурні зображення, що описують семантичні одиниці;
- відповідні правила структурування понять (синтаксичні правила).

Механізми оброблення (їх робота) визначаються семантикою та синтаксисом мови [12-19]. Мови процедурного типу є мовами опису, що базуються на семантиці процедур. Операційні механізми цих мов відповідають функціям оброблення та перетворення інформації, а структурні елементи мови не мають нічого спільного з об'єктами реального світу. Процедурна семантика розглянута й описана в теорії обчислень, що дозволяє отримати повноцінний теоретичний опис системи оброблення знань у ШІ.

Мова декларативного типу опирається на відповідну їй структуру даних. Причому структура даних відповідає концептуальній структурі світу. Структура даних мови декларативного типу включає в себе компоненти, які визначаються структурою пам'яті ЕОМ. У мовах декларативного типу співвідношення синтаксис-семантика аналогічні співвідношенню природна мова - рисунки, графі, просторові зображення, бази даних, тощо.

Бази даних — це абстрактні моделі обмеженої прикладної галузі (галузі знань), які поділяються на:

- а) інженерні БД;
- б) ієрархічні;
- с) мережеві;

- d) реляційні;
- e) постреляційні.

Подання моделей декларативних мов здійснюється предикатами, малюнками, кресленнями, графами, тощо. Математична структура даних у декларативних мовах базується на системах предикатних рівнянь в алгебрі скінченних предикатів (АСП) або на аксіоматичній теорії множин, у якій теорія множин інтерпретується як структура даних.

Таким чином, приходимо до можливо єдиного шляху, яким найповніше, найоб'єктивніше і найефективніше треба і можна йти шляхом створення систем ШІ — це шлях аналізу, моделювання та синтезу мовного інтелектуального інтерфейсу із допомогою АСП, засобів k -значної логічної системи та відповідних структур і кодування.

АСП [1] є узагальненням алгебри логіки і визначена як:

$$x^\delta = \begin{cases} x, & \delta = 1, \\ \bar{x}, & \delta = 0, \end{cases} \quad (1)$$

де δ — степінь аргументу.

Генеральною ідеєю АСП є розширення області пробігання δ до $\{a_0, a_1, \dots, a_{k-1}\}$, де $a_0, a_1, \dots, a_{k-1}$ — k -значні змінні. Виходячи з цього, задаються функції розпізнавання символів a_i змінної x :

$$x^{a_i} = \begin{cases} 1, & x = a_i, \\ 0, & x \neq a_i. \end{cases} \quad (2)$$

При цьому фіксується множина змінних $x_1 \dots x_n$ та скінченних областей пробігання $A_j = \{a_{i_1}, \dots, a_{i_{k_j}}\}$ кожної змінної x_j , які можуть бути довільної значності.

Функції розпізнавання з'єднуються операціями $\&$, $\vee$ булівської алгебри, у результаті чого отримуються вирази $x^a \& (y^b \vee x^c)$, названі в [1] скінченими предикатами. Вони трактуються як функції виду:

$$f: A_1 \times \dots \times A_n \rightarrow \{0, 1\}. \quad (3)$$

Це, по суті, характеристичні функції k -значної логіки, але, на відміну від неї, в АСП усі функції можуть приймати значення лише з множини $E_2 = \{0, 1\}$. Операції (1) розпізнавання, $\&$ та $\vee$ утворюють функціонально повний базис у множині функцій виду (3).

Теорія інтелекту — це наука про детерміновані, дискретні і скінченні функції індивідуального людського інтелекту [11-13]. Таке визначення істотно обмежує поле діяльності для дослідника інтелекту.

По-перше, це визначення вимагає, щоби теорія інтелекту обмежилася вивченням лише функціональної сторони інтелекту, залишаючи осторонь структуру його матеріального носія. При цьому за межами теорії інтелекту залишаються такі цікаві і важливі питання, як будова мозку і нервової системи людини, функціонування їх окремих частин і структур.

По-друге, теорія інтелекту визначається як теорія індивідуального людського інтелекту. Отже, з теорії інтелекту виключаються процеси взаємодії людей один із одним, а це веде до винесення соціальних явищ за її межі. Крім того, таким визначенням виключаються з розгляду інші, ще не відкриті наукою, можливі види інтелекту, а також взаємодія між інтелектом різних видів.

По-третє, прийняте тут визначення теорії інтелекту вимагає виключення з розгляду безперервних і нескінченних інформаційних процесів, веде до зменшення ролі чинника випадковості у діяльності інтелекту. Недолік такого підходу приводить, зокрема, до неможливості усестороннього вивчення на його основі системи «інтелект — зовнішній світ». Ігнорування статистичної природи сигналів, що надходять до інтелекту із зовнішнього світу, також істотно збіднює теорію інтелекту. Бажання обійтися без безперервних моделей сильно затрудняє вивчення роботи органів чуття, органів руху і мови людини.

Таким чином, прийняте нами визначення теорії інтелекту страждає багатьма недоліками, воно обмежує предмет цієї науки порівняно вузькими рамками, при цьому багато процесів і механізми, які люди схильні інтуїтивно відносити до розряду інтелектуальних, залишаються за межами цього визначення. Нижче приводяться міркування, що спонукали нас, не дивлячись на наголошені моменти, зупинитися на так недосконалomu визначенні теорії інтелекту.

Перш за все, слід взяти до уваги ту обставину, що, навіть якщо б ми врахували всі ті доповнення, які тільки що були згадані, все одно повнота визначення теорії інтелекту не була б досягнута. Інтелект — це така система, яка знаходиться в постійному розвитку; ніхто не в змозі передбачати всі можливі напрями і наслідки цього розвитку. Не виключено також, що з часом будуть відкриті нові види інтелекту, принципово відмінні від людського. Тому визначення теорії інтелекту, як би ми не старалися, у будь-якому випадку буде неповним. З іншого боку, в майбутньому у зв'язку з успіхами пізнання інтелекту ніщо не перешкодить його дослідникам розширити рамки теорії інтелекту, коли в цьому з'явиться необхідність. В даний момент, проте, така необхідність гостро не відчувається. Навіть те, порівняно вузьке, визначення теорії інтелекту, яке прийняте нами, ставить масу цікавих задач. Рішення цих задач і практичне використання досягнень теорії інтелекту обіцяє великі блага людству. В даний час, поки теорія інтелекту знаходиться ще у стадії зародження і робить перші кроки у своєму розвитку, вельми розширювальне тлумачення її предмету надало б їй «ведмедючу послугу», бо потурало б розпилюванню наукових сил і засобів, розтраті зусиль у безплідних спробах вирі-

шувати задачі, складність яких перевершує можливість сучасної науки. І, навпаки, вдале тимчасове обмеження предмету теорії інтелекту дозволило б сконцентрувати зусилля на першочергових задачах теорії інтелекту, послужило б стимулом до якнайшвидшого її розвитку.

Визначаючи теорію інтелекту таким чином, як це зроблено в роботах [11-13], ми виходили з того погляду, що детермінованим, дискретним і скінченним процесам в інтелектуальній діяльності належить провідна роль. Вельми важливою обставиною, що вплинула на вибір такого визначення, послужила також наявність на даний час універсальних цифрових обчислювальних машин. Детерміновані, дискретні і скінченні процеси — це в точності ті процеси, які в принципі можуть бути реалізовані на цифрових обчислювальних машинах. Це надає у розпорядження теорії інтелекту універсальний засіб для практичної перевірки і апробації всіх її висновків і рекомендацій. При прийнятому визначенні теорія інтелекту стає могутньою рушійною силою у справі подальшого вдосконалення засобів обчислювальної техніки. Це вдосконалення, у свою чергу, спричинить серйозну стимулюючу дію на розвиток самої теорії інтелекту.

Обмеження теорії інтелекту детермінованими, дискретними і скінченними процесами дає можливість забезпечити її універсальним математичним апаратом для формального опису довільних скінченних відношень. У роботі як такий апарат запропоновано алгебру скінченних предикатів. На наш погляд, АСП може служити міцним фундаментом теорії інтелекту. Саме завдяки наявності цієї алгебри ми зважилися виділити теорію інтелекту в розряд самостійної наукової дисципліни.

Створення АСП доводить до логічного кінця пошуки універсальної математичної мови для опису детермінованих, дискретних і скінченних інформаційних процесів. АСП у відомому значенні є узагальненням апарату булевих функцій (перехід від двійкових знаків до букв) і апарату багатозначної логіки (перехід від функцій до відношень). Використання АСП у теорії інтелекту приводить до виключення потенційної нескінченності з формальних описів інтелектуальних процесів, що забезпечує краще наближення одержуваних моделей до дійсності, ніж те, яке може дати апарат теорії алгоритмів. АСП вводить у вжиток скінченні алфавітні оператори, адекватніші для формального опису детермінованих, дискретних і скінченних інформаційних процесів, ніж ті, що широко використовуються в даний час для тієї ж мети потенційно нескінченні алфавітні оператори.

Важливо підкреслити, що АСП не є чимось зовнішнім стосовно теорії інтелекту. Вона є не тільки засобом формального опису функцій інтелекту, але в її формулах укладена сама суть інтелектуальних

процесів. Формули АСП природним чином інтерпретуються як поняття інтелекту, її рівняння — як думки, думки інтелекту, а процес рішення рівнянь — АСП — як процес мислення.

Наявність АСП відкриває можливість переходу від алгоритмічного опису інформаційних процесів до опису їх у вигляді рівнянь. Ми вважаємо, що мова рівнянь добре відповідає специфіці інтелектуальних процесів. На відміну від алгоритмів і програм для ЕОМ, рівняння АСП, як, втім, і будь-яким іншим рівнянням, не властива спрямованість дії. Алгоритмічний процес завжди реалізує деяку функцію, з його допомогою проводиться визначення значень цієї функції залежно від значень її аргументів. Рівняннями ж можна задати відносини між змінними. Всі змінні в рівнянні рівноправні, будь-які з них, залежно від нашого бажання, можуть виступати як у ролі незалежних, так і у ролі залежних змінних. Алгоритми і програми описують функціонування систем з входами і виходами, а рівняння АСП описують функціонування багатополіусних систем. У першому випадку вхід і вихід системи жорстко зафіксовані, у другому випадку вхідні сигнали можуть подаватися на будь-які полюси, а знімання вихідних сигналів також може проводитися з будь-яких полюсів.

Особлива перевага способу опису об'єкту за допомогою рівнянь у порівнянні з алгоритмічним описом полягає у тому, що за рівняннями можна визначити реакції об'єкту при будь-якому способі розподілу вхідних і вихідних сигналів між його полюсами. Якщо число змінних у рівняннях велике (а в рівняннях, що описують інтелектуальні процеси, це число, поза сумнівом, буде дуже великим), то таким же великим буде і число полюсів у об'єкта, описуваного цими рівняннями. Число ж усіх різних способів розподілу вхідних і вихідних сигналів між полюсами об'єкту буде астрономічно великим. У цих умовах практично неможливо створити повний набір алгоритмів, кожний з яких обчислював би реакції об'єкту при якомусь одному способі розподілу вхідних і вихідних сигналів між полюсами об'єкту.

Рівняння вигідно відрізняються від алгоритмів ще й тим, що дозволяють розраховувати (хоча і неточно) реакції описуваного ними об'єкту навіть при неповному його завданні. Не повністю ж розроблений алгоритм буде просто непрацездатним. Опис об'єкту у вигляді системи рівнянь можна легко удосконалювати, проводячи в ньому лише локальні зміни шляхом додавання або виключення окремих рівнянь. Навіть у процесі своєї реконструкції система рівнянь завжди залишається готовою до використання. Описуючи ж реакції об'єкту за допомогою алгоритму, ми ризикуємо зіткнутися з необхідністю корінної переробки алгоритму при щонайменших змінах наших знань про об'єкт.

Опис об'єктів за допомогою рівнянь - традиційний спосіб наукового опису світу. Цей спосіб широко поширений у фізиці, хімії, техніці і багатьох інших науках. Нам представляється, що прийшов час використовувати його також і при описі дискретних інформаційних процесів.

Алгоритми і програми для ЕОМ, з одного боку, і рівняння АСП — з іншого, що розглядаються як засоби математичного опису скінченних об'єктів, відрізняються ще в одному вельми важливому відношенні. Алгоритми і програми, описуючи деякий об'єкт, разом із тим містять у собі ще й рецепт обчислення реакцій цього об'єкту на зовнішні дії. Укладач алгоритму чи програми вимушений вирішувати відразу дві задачі: 1) природничонаукову задачу математичного опису об'єкту; 2) формально-логічну задачу визначення реакцій об'єкту на зовнішні дії. Рівняння ж, описуючи об'єкт, не містять у собі рецепту обчислення реакцій цього об'єкту. Щоби ці рівняння привести в дію, їх необхідно розв'язати за допомогою спеціально розробленого методу. При математичному описі об'єкту за допомогою рівнянь їх укладач вирішує не обидві задачі, а тільки одну з них - першу. Він не робить нічого зайвого, не захаращує опис інформацією, що не відноситься до справи, тому природно чекати, що математичний опис об'єктів за допомогою рівнянь буде простішим, ніж опис тих же об'єктів за допомогою алгоритмів і програм. Ми сподіваємося, що рівняння АСП стануть основним інструментом у справі моделювання інтелектуальних процесів.

Вважаємо, що мова рівнянь добре відповідає специфіці інтелектуальних процесів. На відміну від алгоритмів і програм для цифрових ЕОМ рівнянням АСП, як, втім, і будь-яким іншим рівнянням, не властива спрямованість дії. Алгоритмічний процес завжди реалізує деяку функцію, з його допомогою проводиться визначення значень цієї функції в залежності відношень тощо); поняття алгебри логіки і багатозначної логіки; деякі поняття, пов'язані з процесами обробки виразів (слово, довжина слова, з'єднання слів, упаковка виразу, об'єднання виразів, тощо); прості лінгвістичні поняття (голосні, шиплячі, елементи парадигматичної таблиці тощо). У цій галузі ще належить виконати величезний обсяг робіт. Їх завершення означатиме формалізацію всієї системи понять, якою володіє людський інтелект. У міру розширення круга понять, математично описаних засобами АСП, її мова ставатиме все більш ємною і абстрактною, поступово наближаючись по своїй силі і виразності до природної мови.

Важливою проблемою теорії інтелекту є розробка методів і засобів для вирішення рівнянь АСП. Вони необхідні для фізичного відтворення інформаційних процесів, формально описаних теорією інтелекту. Без вказаних методів і засобів математичні моделі функцій інтелекту залишаться тільки

мертвими записами на папері. На даний час зроблено перші кроки у напрямі рішення цієї проблеми: побудовано алгоритми рішення рівнянь АСП шляхом приведення їх до досконалої диз'юнктивної нормальної форми і шляхом послідовної підстановки в рівняння знайдених значень змінних; розроблено методи мінімізації формул і методи декомпозиції рівнянь АСП; знайдено способи явного виразу невідомих для рівнянь із постійними і змінними параметрами; описано ряд тотожностей, на основі яких можуть виконуватися спрощення формул за наявності в них символів вищого рівня абстракції, що скорочують (заперечення, імплікація, об'єднання множин, з'єднання слів, упаковка виразів, тощо).

Рішення рівнянь, що описують складні інтелектуальні процеси, можна подати як багаторушний процес, здійснюваний на різних рівнях абстракції. Щоб повністю автоматизувати процес, необхідно заздалегідь виконати дуже велику роботу щодо формального опису властивостей понять різного рівня абстракції, створити теорію тотожних перетворень і спрощення формул різних консервативних розширень АСП. Як перший крок у цій галузі розроблено початки теорії виразів, в якій формально описані прості перетворення послідовностей символів (селекція, комутація, зсув, перестановка, вибірка, розмноження, знищення, заміна букв виразу, підрахунок букв виразу, віднімання виразів, тощо).

Нагальною задачею теорії інтелекту є створення універсального вирішувача задач, який на першому етапі робіт можна подати у вигляді спеціальної програми для цифрових ЕОМ, розв'язку певного класу рівнянь АСП з символікою достатньо високого рівня абстракції. Рішення цієї задачі дозволило б взагалі відмовитися від складання програм для багатьох задач, складність яких не перевершує деякий заданий рівень. Достатньо б було записати умови задачі у вигляді системи рівнянь АСП і вирішити систему за допомогою універсального вирішувача.

Другий етап робіт у цьому напрямі представляється як створення спеціалізованої машини для вирішення рівнянь АСП заданого рівня складності. Така машина, по суті, виявилася б відправним пунктом у справі створення системи штучного інтелекту. У міру того, як у процесі вдосконалення машини, збільшувалася б гранична складність вирішуваних нею рівнянь, а також росли б рівень абстрактності і число освоєних нею понять, спостерігався б поступовий розвиток її інтелекту. Зараз зроблені певні зусилля у напрямі створення передумов до побудови такої машини. Розроблено початки теорії ланцюгів перемикачів і теорії цілком скінченних автоматів, розглянуто способи схемної реалізації деяких первинних математичних структур і операцій (множин і відношень, операцій над ними, булевих і перемикачів функцій, арифметичних дій, операцій над виразами і словами, тощо).

4. Фонетичний аналіз. Моделі букво-фонемних відношень для голосних та приголосних звуків української мови

Задача фонетичного аналізу полягає у тому, щоби засобами АСП описати букво-фонемне відношення, що зв'язує звук мови у визначеному фонетичному контексті з відповідною йому буквою українського алфавіту.

Перш ніж перейти до математичного опису фонетичного відношення, уведемо формальне визначення фонем як знаку. Це визначення фонем для української мови можна записати у вигляді наступного індивідуального предиката [1]:

$$\begin{aligned} \Phi_{OH}(X) = & x^a \vee x^b \vee x^B \vee x^r \vee x^I \vee x^e \vee x^c \vee x^J \vee \\ & x^3 \vee x^H \vee x^i \vee x^{\dot{i}} \vee x^K \vee x^l \vee x^M \vee x^H \vee x^O \vee \\ & x^l \vee x^p \vee x^c \vee x^T \vee x^y \vee x^\phi \vee x^x \vee x^u \vee x^H \vee \\ & \vee x^{III} \vee x^{III} \vee x^{JO} \vee x^R. \end{aligned} \quad (5)$$

Твердження, що X є фонема, еквівалентне рівності $\Phi_{on}(X) = 1$.

У [1] побудовано математичні моделі букво-фонемних відношень для голосних та приголосних звуків української мови на підставі методики формального опису букво-фонемних правил рівняннями алгебри скінчених предикатів (АСП).

Для створення математичних моделей букво-фонемних відношень уведено 12 граматичних ознак фонем: y_1 — ознака голосності звуку із значеннями г — голосна, п — приголосна; y_2 — ознака шумності приголосних із значеннями ш — шумова, н — не шумова; y_3 — ознака вокалізації звуку із значеннями дз — дзвінка, г — глуха; y_4 — ознака проривності звуку із значеннями пр — проривна, щ — щілинна; y_5, y_6 — ознаки веляризації та лабіалізації звуку із значеннями м, с, в — мала, середня, велика відповідно; y_7 — ознака палаталізації із значеннями т, м — тверда, м'яка відповідно; y_8 — ознака назалізації із значеннями р — ротова, н — носова; y_9 — вібрантності із значеннями с — спокійна, т — тремтяча; y_{10} — місця артикуляції звуку г — губна, п — передньоязикова, з — задньоязикова; y_{11} — ознака передньоязикової шумової фонемі із значеннями з — зубна, п — піднебінна; y_{12} — ознака аффрикативності передньоязикових шумових фонем із значеннями а — аффриката, н — не аффриката.

Під час переходу від фонетичних ознак до фонетичного зображення звука формується фонема – знак, що позначає клас звуків, які тяжіють до певної букви тексту. Зворотнє перетворення забезпечує заміну фонетичного знака набором відповідних йому значень фонетичних ознак звуку. Процедури таких перетворень детально розроблено і описано в [1] і тут не наводяться, щоби за дрібними деталями не втратити картину загального підходу.

Моделювання букво-фонемних відношень голосних і приголосних української мови здійснюється за допомогою рівнянь АСП, що зв'язують між собою звук Y мови, який перебуває у певному фонетичному контексті Z , та його графічне зображення X :

$$F(X,Y,Z)=1, \quad (6)$$

де X – алфавіт української мови; Y – звуки української мови; Z – фонетичний контекст.

У рівняннях типу (6) звук мови подається у вигляді вектора $Y = (p, q, t)$ із компонентами p, q, t , що виконують роль фонетичних ознак, суть яких описана нижче. Як формальний еквівалент *впливу фонетичного контексту* в [1] розглядається сукупність ознак (p_w, q_w, p_v, q_v) . Відповідно еквівалентом впливу наголосу є змінна t . Набір змінних $(X_1, \dots, X_{32})$ характеризує графічне зображення звуку.

Традиційний опис фонетики української мови використовує поняття *фонетичного пом'якшення* [1]. Будь-якому звуку мови властиве одне з двох можливих значень цієї ознаки: т — тверда, м — м'яка. У букво-фонемній моделі даний елемент характеристики подається змінною *q*. Наступна логічна рівність формально визначає область зміни для уведеної змінної:

$$q^T \vee q^M = 1. \quad (7)$$

Під час опису букво-фонемного відношення наступним є *поняття наголошення*. Воно визначає інформацію про наявність чи відсутність наголосу голосного звуку, що розглядається. Для формалізації граматичної *категорії наголошування* скористаємось змінною t із області визначення, що складається з трьох можливих значень даної категорії:

$$t_3 \vee t_2 \vee t_1 = 1, \quad (8)$$

де 3 – біжучий голосний звук, який перебуває у наголошеному складі, 2 – у переднаголошеному складі, 1 – у решті складів слова.

Уведемо сукупність змінних Q_z , необхідних для формального опису понять, зв'язаних із визначенням *фонетичного контексту*. До таких понять відноситься опис фонетичних позицій абсолютного початку та абсолютного кінця слова, середини слова, сусідства з м'якими чи твердими голосними й приголосними; опис передвокальної, поствокальної та інтервокальної позицій, положення після чи до групи приголосних, різних за місцем та способом утворення, за участю голосу тощо. Множину змінних Q_z назовемо *характеристикою фонетичного контексту*.

Для того, щоби виразити залежність між звуком Y та характеристикою фонетичного контексту Z у вигляді логічних рівнянь, знадобляться змінні, що описують оточення звуку в слові, що досліджується. Позначимо змінною u звук слова, що безпосередньо йде за досліджуваним звуком y . Якщо звук Y перебуває в положенні абсолютного кінця слова, приймаємо $u = -$. Областю визначення змінної u будемо вважати множину всіх звуків української мови. Позначимо предикат, що стоїть у лівій частині рівняння (3) через $J_1(p)$, тоді область зміни змінної u може бути задана рівнянням:

$$J_1(p_u) = 1. \quad (9)$$

Предикат $J_1(p_u)$ відповідає поняттю «звук української мови». Звук, що передує біжучому звуковій слова, експлікуємо за допомогою літерної змінної v . Якщо U звучить на початку слова, то приймаємо $v = L$.

На підставі вивчення даних фонетики і орфографії української мови в [1] розроблено систему фонетичних ознак, а також систему предикатів для формального опису системи фонетичних ознак, що дозволяє здійснювати їх автоматичне оброблення. Розроблено методику формального опису букво-фонемних відношень, побудовано математичні відношення голосних та приголосних звуків української мови, на основі яких формалізуються закономірності поєднання голосних і приголосних букв та звуків, що утворюють пари за дзвінкістю-глухістю й твердістю-м'якістю, а також поза парних приголосних. Розроблено математичну модель акцентних характеристик української мови. Усі ці результати дозволяють отримати компактну модель для економного запису словників природної мови в пам'яті ЕОМ.

Із допомогою опису букво-фонемних зв'язків у вигляді системи рівнянь АСП можна в інтелектуальному інтерфейсі здійснити формальний перехід від фонетичного запису слова до його графічного зображення (задача фонетичного аналізу), за даним орфографічним текстом отримати його фонетичну транскрипцію (задача синтезу мови), а також перейти до системного та структурного синтезу процесорів функціонально-зорієнтованих на аналіз та синтез мови, які є необхідною складовою частиною систем штучного інтелекту.

5. Морфологічний аналіз

Морфологічний аналіз — це оброблення словоформ без зв'язку з контекстом. Словоформа — це відтинки тексту між двома просвітками. Розділові знаки вважаються окремими словоформами.

У центрі уваги морфології стоїть слово з його граматичними змінами. У лінгвістичній літературі термін «морфологія» вживається для позначення відповідної системи всередині мови, у яку входить словозміна (словотворення розглядається як самостійна система) [1].

Під словозміною розуміється утворення форм одного й того ж слова, що існує в мові, як система форм (словоформ). Словоформи «можна визначити, як регулярні видозміни слова, об'єднані тотожністю його лексичного значення й неоднаковими морфологічними значеннями» [1]. Запропоноване визначення охоплює такі зміни, за яких зберігається цілим лексичне значення слова.

Нижче об'єктом формального опису є письмо, тобто орфографічна, а не фонематична реалізація словоформ української мови. Така форма запису усуває необхідність додаткового оброблення під час введення/виведення текстів, а також дозволяє

усунути із опису акцентний компонент. При цьому, окремою словоформою є відтинки тексту між двома проміжками.

Утворення слова звичайно відбувається за допомогою спеціальних словозмінних морфем. Під морфемою будемо розуміти мінімальну значущу одиницю мови: «Морфема — це найменша неподільна значуща частина слова, тобто одиниця мови, а в мові виступають окремі представники морфем, які зручно називати морфами» [1].

За значеннями і функцією в структурі слова морфеми поділяються на кореневі і службові. Морф — конкретний вияв морфеми, найменша значуща частина, що виділяється у складі певної конкретної словоформи слова. Особливу роль під час словозміни відіграють закінчення словоформ (флексії), що передають, як правило, граматичні значення в українських словах. Закінчення завжди виступає в певній системі словозміни змінюваних частин мови.

Типи відмінювання іменників, прикметників, числівників, займенників розрізняються системою закінчень, які виражають граматичні значення відмінка, роду і числа або лише відмінка (в кількісних числівниках). Наприклад у іменників флексії завершують матеріальну побудову, закріплюють лексико-граматичне значення слів, сприяють відображенню таких категорійних ознак як рід, одухотвореність, число, відмінок. Дієслова мають складну систему дієвідмінювання за особами, родами і числами, основними показниками яких є особово-, родово-числові флексії.

Будемо класифікувати флексії на регулярні та нерегулярні. Регулярні флексії — найрозповсюдженіші й приймаються за норму. Описана нижче математична модель базується на формальній моделі морфології української мови.

У морфології української мови існує певна залежність (відношення) між флексією та оточуючим її текстом. Завдання полягає у математичному описі існуючої залежності, тобто у формалізації поняття флексії. Текст, що охоплює закінчення, є неоднорідним щодо нього. Будемо розрізняти найближчий текст (що межує безпосередньо із закінченням у словоформі) і подальший текст (що межує з словоформою, у якій перебуває закінчення, що нас цікавить).

У роботі [1] введено поняття текстового відношення $T(X, Y)$, де X — набір граматичних ознак, що виділяє підмножину усіх текстів із деякої множини різноманітних рядків обмеженої довжини, упорядкованих із букв скінченного алфавіту; Y — текст. Текст розглядається як багаторівнева структура, у якій із букв утворюються морфи, а з морфів — словоформи, із словоформ — речення, та з речень — абзаци. Окремі частини цієї конструкції названі фрагментами тексту (букви, морфи, словоформи, речення, тощо). У математичній моделі системи української словозміни, як фрагмент тексту

виступають словозмінні морфеми, що можуть бути суфіксами й закінченнями. Враховуючи висловлене раніше зауваження щодо ролі флексій у процесах словозміни, зосередимо подальший розгляд на методиці формального опису процесів зміни слів (за допомогою флексій), що можна застосувати й до інших лінгвістичних одиниць тексту.

Дію тексту, що оточує закінчення (як ближчого, так і подальшого), замінимо введенням деякого набору ознак, який проінтерпретуємо як *смісл закінчення* [1]. Під смислом будемо розуміти набір проміжних змінних, що зв'язують рівняння, які описують закінчення, з рівняннями, що описують решту частини тексту. Набір значень цих змінних будемо інтерпретувати як значення смислу. Смісл розглядається як змінний вектор $X = (X_1, X_2, \dots, X_m)$, де m — число компонентів вектора смислу, а змінні $X_1, X_2, \dots, X_m$ — компоненти смислу.

Нехай $Y_1, Y_2, \dots, Y_n$ — закінчення, де значенням Y_i служить i -а за рахунком буква закінчення. Набір $Y = (Y_1, Y_2, \dots, Y_n)$ буквених змінних $Y_1, Y_2, \dots, Y_n$ назовемо *змінним закінченням*, набір значень цих змінних — *індивідуальним закінченням* чи просто закінченням. Змінні змінного закінчення, їх значення назовемо значеннями компонентів закінчення.

Під час побудови математичної моделі флексії описується *фрагментне морфологічне* відношення $L(X, Y)$, що відтворює зв'язок між смислом X і змінним фрагментом (флексійним морфом) тексту Y . Параметрами введеного відношення можуть бути будь-які із компонентів вектора смислу. При цьому $X \in H, Y \in M$, де H — множина наборів значень компонентів смислу, M — множина різних послідовностей скінченної довжини, упорядкованих із букв українського алфавіту та символу проміжка (L —), що служить для заповнення вільних від букв позицій у закінченнях.

Вважаємо, що морфологічне відношення $L(X, Y)$ виконується тоді і тільки тоді, коли Y передає значення компонентів смислу X . Мовна практика говорить про те, що результат перевірки умови виконання морфологічного відношення має бути однозначним (у протилежному випадку людина не змогла б ефективно формувати тексти).

Згідно з викладеним вище принципом однозначності, закінчення завжди однозначно залежить від смислу. Цей принцип можна інтерпретувати як вимогу повноти набору ознак (компонентів смислу), із допомогою якого здійснюється вибір індивідуального закінчення. Набір ознак будемо називати повним, якщо він при будь-яких значеннях ознак забезпечує однозначність вибору відповідної йому флексії. У якості смислу приймемо будь-який набір ознак, що задовольняє вимозі повноти. Під повним смислом флексії будемо розуміти набір, складений з деяких компонентів вектора смислу X . Смісл флексії, у прийнятому раніше значенні, будемо називати *повним смислом*. Поняття повного й неповного смислу флексії введені за аналогією

з відповідними поняттями для фрагменту тексту. Моделювання фрагментного відношення необхідно почати з вибору структури багатокомпонентних векторів X та Y . Як компоненти вектора Y необхідно вибрати букви закінчення. При цьому нумерація букв ведеться в межах флексії зліва направо. Що ж стосується вибору структури вектора смислу X , то тут значно все складніше.

Для того, щоб отримати компактний та повний опис фрагментного морфологічного відношення $L(X, Y)$, необхідно вибрати компоненти вектора X із урахуванням внутрішньої структури мови. Для раціонального вирішення цього завдання будемо спиратися на поняття, сформульовані морфологією української мови. Критерієм під час оцінки припустимості вжитку того чи іншого поняття, виробленого мовознавством, має бути, очевидно, успіх формалізації природної мови. Через те, що об'єкт, який моделюється, є достатньо складною для математичного опису системою, зручно розбити його на ряд підсистем (які будемо надалі називати «мовними блоками»), що пізніше об'єднуються теоремою про розкладання в алгебрі скінченних предикатів [1].

Для кожного мовного блоку фіксують значення деяких із компонентів смислу, які ми назовемо обмеженими. Наприклад одним із таких компонентів смислу може бути ознака частини мови, що дозволяє формалізувати флексійне оброблення кожної частини мови окремо. Другу групу ознак, що служать безпосередньо для вибору закінчення усередині деякого мовного блоку, назовемо змінними. Необхідно відзначити, що назви двох груп ознак умовні, оскільки насправді всі компоненти вектора смислу є змінними. Обмежувальним ознакам поставимо у відповідність змінну ξ , а змінним — X . Таким чином, маємо можливість уточнити структуру вектора смислу $X = (\xi, X) = (\xi_1, \xi_2, \dots, \xi_k, X_1, X_2, \dots, X_l)$, де k, l — число обмежувальних та змінних компонентів вектора смислу відповідно.

Змінні ознаки в свою чергу неоднорідні щодо закінчення. Раніше вже підкреслювалась неоднорідність тексту, що оточує флексію. Впливові ближнього тексту на закінчення будуть відповідати лексичні ознаки, а подальшого — синтаксичні.

Щоби визначити склад змінних ознак математичної моделі, необхідно сформулювати таку вимогу до неї: закінчення Y , яке перебуває в морфологічному відношенні зі смислом X , має однозначно визначатися смислом (у відповідності з принципом однозначності). Унаслідок цього, до складу змінних компонентів вектора смислу включенні як словозмінні категорії (для прикметників — рід, число, відмінок), так і несловозмінні, але такі, що керують утворенням різних варіантів форм. Наприклад наявність у деяких слів сучасних та архаїчних варіантів форм: уста, перст — *архаїчне*, губи, палець — *сучасне*. Тому до складу змінних ознак включено ознаку зі значеннями: сучасне (с), архаїчне (а). Не завжди закономірності в утворенні варіантів форм

настільки очевидні, що їм можна поставити у відповідність будь-яку морфологічну категорію. У подібних випадках будемо уводити так звані умовні морфологічні категорії, що відповідають простому перерахуванню варіантів форм.

Отже, завдання полягає у формальному описі засобами АСП фрагментного морфологічного відношення $L = (\xi, X) = (\xi_1, \xi_2, \dots, \xi_k, X_1, X_2, \dots, X_l, Y_1, Y_2, \dots, Y_n)$, що буде моделлю флексійного оброблення українських словоформ. Морфологічне відношення визначається своєю характеристичною функцією $L_\varphi(X, Y)$, що задається наступним чином:

$$L_\varphi(X, Y) = \begin{cases} 1, & \text{якщо } X \text{ і } Y \text{ зв'язані} \\ & \text{морфологічним відношенням;} \\ 0, & \text{у протилежному випадку.} \end{cases} \quad (10)$$

Відношення (6) є предикатом, визначеним на скінченній множині предметної галузі. Він дозволяє інтерпретувати різноманіття граматичного оброблення українських флексій (аналіз, синтез, нормалізація, корегування помилок, тощо) із допомогою розв'язків канонічних рівнянь виду $L_\varphi(X, Y) = 1$.

Подамо морфологічну функцію у явному вигляді:

$$y_j^\sigma = F_\sigma(\xi, X), \quad 1 \leq j \leq n. \quad (11)$$

Вона буде описана, якщо нам вдасться подати у вигляді формул предикати F_σ , де σ належить множині значень, які приймають букви закінчень. Предикати $F_\sigma(\xi, X)$ задаються різноманітними ідентифікаторами (розпізнавачами) y_j^σ для змінної y_j .

Тут прийнято паралельний метод опису множини флексійних морфів [1], згідно із яким утворюється множина місць для усіх флексій, що підлягають опису. Кожному місцю j ставиться у відповідність своя змінна y_j . Як формальний еквівалент флексії приймаємо набір значень усіх змінних y_j . Значенням змінної y_j є стан місця j для індивідуального морфа.

Таким чином, задача опису множини флексійних морфів зводиться до формального опису функції виду:

$$y_j^\sigma = F_\sigma(\xi_1, \xi_2, \dots, \xi_k, X_1, X_2, \dots, X_l), \quad 1 \leq j \leq n.$$

Висновки

На підставі вивчення даних фонетики і орфографії української мови та за допомогою опису букво-фонемних зв'язків у вигляді системи рівнянь АСП можна:

— в інтелектуальному інтерфейсі здійснити формальний перехід від фонетичного запису слова до його графічного зображення (задача фонетичного аналізу);

— за даним орфографічним текстом отримати його фонетичну транскрипцію (задача синтезу мови);

— перейти до системного та структурного синтезу процесорів, функціонально-зорієнтованих на аналіз та синтез мови, які є необхідною складовою частиною систем штучного інтелекту.

Список літератури: 1. Бондаренко М.Ф., Коноплянко З.Д., Четвериков Г.Г. Основи теорії синтезу надшвидкодійних структур мовних систем штучного інтелекту. — К.: ІЗМН, 1997. — 264 с. 2. Бондаренко М.Ф., Коноплянко З.Д., Четвериков Г.Г. Основи теорії багатозначних структур і кодування в системах штучного інтелекту. — Х.: Фактор-Друк, 2003. — 336 с. 3. Коноплянко З.Д., Чаплига В.М., Чаплига М.В. Багатозначні структури та кодування систем економічної кібернетики. — Львів: ЛБІ НБУ, 2004. — 314 с. 4. Коноплянко З.Д. Принципы построения многозначных систем искусственного интеллекта // Проблемы бионики. — 1990. — Вып. 45. — С. 27–35. 5. Коноплянко З.Д. Принцип симбиоза багатозначних структур штучного інтелекту та його формалізація // Искусственный интеллект. — 2001. — № 2. — С. 140–151. 6. Коноплянко З.Д., Четвериков Г.Г. Проблеми інтелектуалізації та українізації цифрових систем та мереж телекомунікацій // Труды УНИИРТ. — 1995. — № 4. — С. 72–81. 7. Коноплянко З.Д., Четвериков Г.Г. Аналіз лінгвістичних зв'язків елементів фонетичного рівня української мови // Праці УНДІРТ. — 1996. — № 1(5). — С. 70–77. 8. Четвериков Г.Г., Ляховець С.В., Коноплянко З.Д., Колесник А.С. Аналіз та дослідження букво-фонемних відношень української мови // Праці УНДІРТ. — 2001. — № 1(25). — С. 79–83. 9. Абашина Н.М., Коноплянко З.Д., Чаплига В.М. Структурно-функціональний аналіз системи штучного інтелекту та його підсистеми — національної (української) мови // Мандрівець. — 2002. — № 5(40). — С. 64–71. 10. Лачинов В.М., Поляков А.О. Информодинамика или Путь к Миру открытых систем. — Санкт-Петербург: Изд. СПбГТУ. — 1999. — 432 с. 11. Шабанов-Кушнарченко Ю.П. Теория интеллекта. Математические средства. — Харьков: Вища шк., 1984. — 144 с. 12. Шабанов-Кушнарченко Ю.П. Теория интеллекта. Технические средства. — Харьков: Вища шк., 1986. — 136 с. 13. Шабанов-Кушнарченко Ю.П. Теория интеллекта. Проблемы и перспективы. — Харьков: Вища шк., 1987. — 160 с.

Надійшла до редколегії 28.10.2008

УДК 007.681.5:519.714:519.766

Концепции организации информационно-интеллектуальных технологий / З.Д. Коноплянко // Бионика интеллекта: науч.-техн. журнал. 2008. — №2 (69). — С. 164–172.

В статье проанализированы проблемы создания систем искусственного интеллекта, которые позволяют моделировать решения, которые являются элементами k -значной организации информационно-интеллектуальных технологий.

Библиогр.: 13 наим.

UDK 007.681.5:519.714:519.766

Conceptions of organization of informatively-intellectual technologies / Konoplyanko Z. D. // Bionics of Intelligence: Sci. Mag. — 2008. — № 2 (69). — P. 164–172.

In the article problems are analysed of creation of the intelligence of systems, which allow to design the decisions which are the elements of multiple valued organization of informatively-intelligence technologies.

Ref.: 12 items.

УДК: 612,84-003,96:681,3



РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ ЧАСТОТНЫХ ХАРАКТЕРИСТИК ЗРИТЕЛЬНОЙ СИСТЕМЫ С ИСПОЛЬЗОВАНИЕМ АВТОМАТИЗИРОВАННОГО ПРИБОРА

Л.Ф. Сайковская

ХНУРЭ, г. Харьков, Украина, lily_s_f@mail.ru

Предложен способ и устройство для оценки функционального состояния организма в условиях напряженного зрительного труда, приведена схема устройства и результаты исследования частотных характеристик зрительной системы, полученные с его помощью.

**ФУНКЦИОНАЛЬНОЕ СОСТОЯНИЕ ЧЕЛОВЕКА, АВТОМАТИЗИРОВАННЫЙ КОМПЛЕКС,
КРИТИЧЕСКАЯ ЧАСТОТА СЛИЯНИЯ МЕЛЬКАНИЙ**

Введение

Компьютеры стремительно внедряются в производство, обучение и быт. Значительно облегчая получение и обработку информации, они оказывают негативное влияние на организм человека [1]. Работа профессиональных пользователей персональных компьютеров (ПК) требует повышенных умственных усилий и большого нервно-эмоционального напряжения, решения в ограниченное время сложных задач, высокой концентрации внимания и особой ответственности выполнения производственного задания. Высокие требования к зрительной системе, нервное напряжение, а также монотонный характер труда и вынужденная рабочая поза вызывают большое количество жалоб работающих на повышенное общее и зрительное утомление. Кроме того, у людей, постоянно пользующихся компьютером, отмечается целый ряд функциональных перестроек зрительной системы (например, астиопия, уменьшение объема абсолютной аккомодации, нарушения частотно-контрастной чувствительности, ложная близорукость с тенденцией развития истинной), которые, с одной стороны, обеспечивают максимально возможную адаптацию к специфическому зрительному труду, с другой – приводят к формированию донозологических и патологических состояний [2].

В связи с вышеизложенным, актуальной задачей является разработка способов и устройств для оценки функционального состояния организма человека в условиях напряженного труда, например, при работе за компьютером.

Исходя из доминирующей роли центральной нервной системы (ЦНС) во всех процессах в организме, при исследовании функционального состояния человека в процессе работы с ПК, в первую очередь необходимо контролировать состояние ЦНС. Динамика параметров, характеризующих ее состояние, отображает индивидуальные особенности процессов регулирования в организме в ходе адаптации к нагрузке. При работе с ПК на первое место большинство исследователей ставят его негативное влияние на зрительную систему,

что обусловлено особенностями изображения на экране монитора и характером зрительного труда [1, 2]. Причем помимо необходимости разглядывать мелкие объекты общение с монитором доставляет зрительной системе и еще одно неудобство: внимание оператора почти постоянно направлено в центр экрана, и основная нагрузка приходится на центральную зону сетчатки. Утомлению способствует в значительной мере и то, что снижается работа мышц глазного яблока, а следовательно, и циркуляция крови в сетчатке [1].

Существуют различные методы оценки зрительного утомления. Основываясь на применении традиционного набора методик исследования зрительных функций, зрительное утомление наиболее полно изучено в части мышечного компонента его развития.

В настоящее время наиболее информативными для исследования сенсорного утомления зрительной системы считаются следующие показатели: критическая частота слияния мельканий (КЧСМ), пороги контрастной, электрической и цветовой чувствительности, время моторного ответа на сенсорный зрительный раздражитель, динамика образов и устойчивость ясного видения [3–5].

Одной из широко используемых временных характеристик зрительной системы является критическая частота слияния световых мельканий – частота перехода от видимости мельканий к ощущению их субъективного слияния. КЧСМ является многофакторным индикатором психофизиологического состояния, отражающим текущий уровень активации ЦНС, зависит от лабильности (функциональной подвижности) нервных процессов и чувствительна к изменению психического состояния человека.

Уменьшение значения КЧСМ свидетельствует о развитии утомления ЦНС и организма в целом, увеличение – о наличии возбуждения или стресса [6, 7].

Метод КЧСМ применяется в офтальмологии, офтальмоэргономике, экспериментальной психологии, физиологии и гигиене труда и спорта и был

предложен также для исследования зрительного утомления человека.

Однако в настоящее время в Украине не выпускаются современные устройства, позволяющие проводить исследование и регистрацию КЧСМ как в ручном, так и в автоматическом режимах. Таким образом, целью данной работы является разработка автоматизированного прибора, пригодного для проведения исследований КЧСМ в динамике зрительного труда, а также показать возможность оценки функционального состояния человека при помощи полученного устройства.

1. Материалы и методы

Для исследования динамики КЧСМ у пользователей ПК нами был разработан автоматизированный прибор, включенный в комплекс для диагностики функционального состояния человека [8]. Его структурная схема представлена на рис. 1. Электрическая принципиальная схема данного устройства приведена на рис. 2.

Прибор для исследования КЧСМ построен на основе импульсного генератора с возможностью регулирования частоты импульсов, нагрузкой которого являются четыре тест-объекта (светодиоды разного цвета: красного, синего, зеленого и желтого). Диаметр светового источника составляет 10 мм, дискретность изменения частоты — 0,5 Гц. В существующих в настоящее время устройствах для исследования КЧСМ обычно используются

три цветовых стимула (красный, синий, зеленый). Нами введен дополнительный стимул желтого цвета, что позволяет исследовать особенности функционального состояния оппонентных полей сетчатки в динамике зрительного труда [3].

Суть метода состоит в том, что сетчатка по-разному воспринимает пульсирующий световой сигнал: если частота пульсаций невелика — глаз видит серию вспышек — мельканий, а при большой частоте сигнал воспринимается как ровное свечение. Таким образом, увеличивая частоту пульсаций, можно установить ее граничное значение, за которым глаз перестает различать мелькания. Это значение и называется КЧСМ [4,5]. Исследование КЧСМ является психофизическим методом оценки активности ряда процессов (электрофизиологических, энергетических, биохимических и других), обеспечивающих зрительное восприятие по высокочастотным каналам зрительной системы. КЧСМ характеризует функциональное состояние зрительного анализатора в целом (особенно чувствительна она при патологии сетчатки и зрительного нерва), не зависит от остроты зрения и рефракции, снижается с возрастом [5,9,10]. Кроме того, у разных людей величина КЧСМ при одних и тех же условиях различна, аналогично отличается и динамика показателя [11].

В ходе исследования последовательно, сначала на один, затем на другой глаз испытуемого воздействовали импульсами синего (460 нм), зелено-

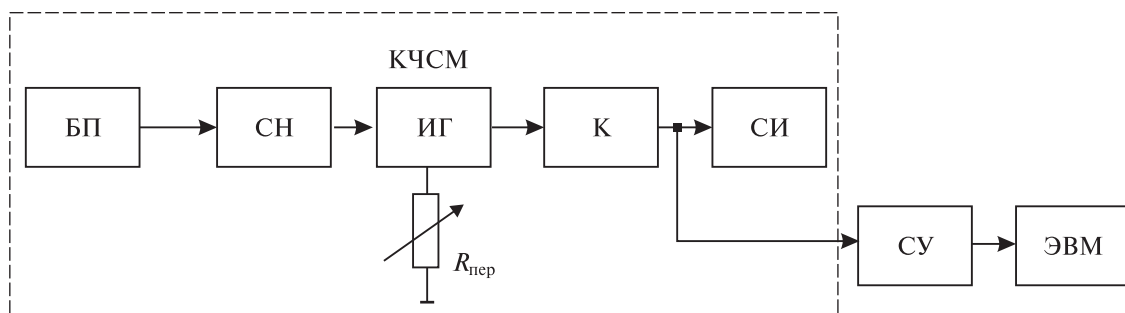


Рис. 1. Структурная схема автоматизированного прибора для исследования КЧСМ: БП — блок питания, СН — стабилизатор напряжения, ИГ — импульсный генератор, К — ключ, СИ — светоизлучатели, СУ — согласующее устройство

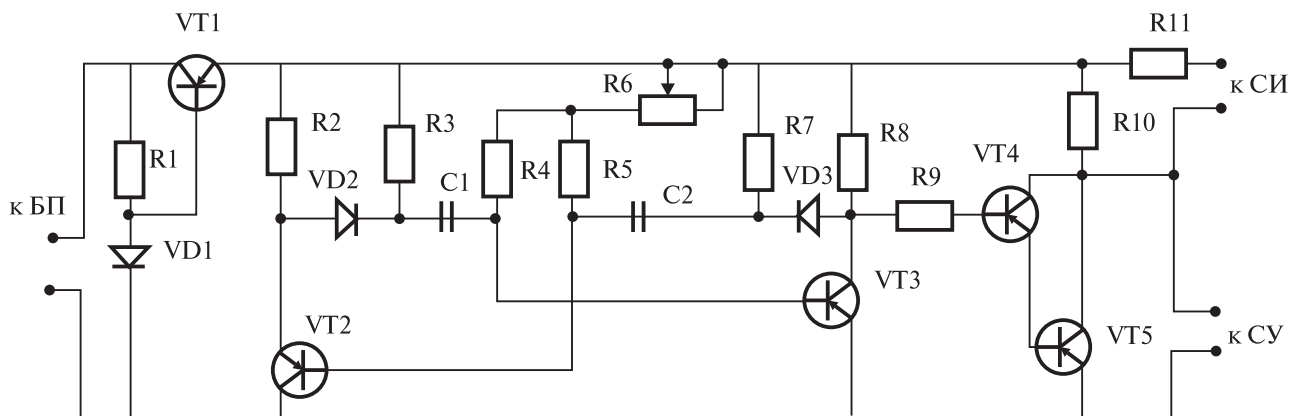


Рис. 2. Схема электрическая принципиальная автоматизированного прибора для исследования КЧСМ

го (525 нм), желтого (575 нм) и красного (622 нм) цветов, частоту которых плавно увеличивали от 1 до 60 Гц. Во время проведения эксперимента, вращая регулятор частоты мельканий от меньшей частоты к большей, подбирается такая частота цветового стимула, при которой испытуемый начинает воспринимать мелькающий тест-объект как непрерывный. Исследование повторяют трижды для каждого глаза и вычисляют среднюю величину критической частоты слияния мельканий.

Поскольку прибор с помощью согласующего устройства подключен к ПК, полученные в результате исследования КЧСМ данные заносятся в его память.

Под нашим наблюдением находилась группа пользователей из 54 человек в возрасте $21,6 \pm 0,4$ года. Определение КЧСМ было проведено в стандартных условиях до и после выполнения различных тестовых заданий на компьютере. В качестве тестовых заданий были использованы: работа с текстом, реализованным на компьютере, игры на компьютере «DX Ball» и «Пасьянс».

Каждый вид тестовых заданий на компьютере выполнялся испытуемыми в разные дни. До и после выполнения тестового задания испытуемого просили отметить момент полного слияния мель-

каний. КЧСМ—тест повторяли 2–3 раза и учитывали только те данные, которые совпадали при повторном исследовании.

2. Результаты и их обсуждение

Результаты исследования КЧСМ при различных видах нагрузки представлены в таблицах 1, 2 и 3.

В табл. 1 представлены результаты исследования КЧСМ до и после работы с текстовым тестом, реализованным на компьютере.

По данным, приведенным в табл. 1 очевидно, что достоверные изменения в показателях КЧСМ до и после работы имеются для желтого, зеленого и синего цветов. Значения КЧСМ для синего и красного цветов достоверно различаются как до, так и после работы. Очевидно, что наблюдается достоверный рост показателей для зеленого, синего и желтого цветов.

Можно отметить, что между показателями правого и левого глаз существует небольшая разница, которая увеличивается после работы. Такие различия характерны для многих функциональных показателей организма, поскольку рост асимметрии является признаком развития утомления.

В табл. 2 представлены результаты исследования КЧСМ до и после игры «DX Ball». По данным,

Таблица 1

Результаты исследования КЧСМ до и после работы с текстом

Цвет	Красный		Желтый		Зеленый		Синий	
Глаз	Правый	Левый	Правый	Левый	Правый	Левый	Правый	Левый
До	35,8 ± 0,6	36,3 ± 0,4	38,1 ± 0,6	39,6 ± 0,5	40,9 ± 0,8	42,1 ± 0,9	37,8 ± 0,8**	39,0 ± 0,5**
После	35,3 ± 0,7	37,1 ± 0,8	39,5 ± 0,6*	38,1 ± 0,9*	41,3 ± 0,7*	41,2 ± 0,8*	41,1 ± 0,9**	39,3 ± 0,4**

*различия между показателями до и после работы на компьютере достоверны ($P < 0,05$);

** различия между показателями КЧСМ для красного и синего цветов, как до, так и после работы достоверны ($P < 0,05$)

Таблица 2

Результаты исследования КЧСМ до и после игры «DX Ball»

Цвет	Красный		Желтый		Зеленый		Синий	
Глаз	Правый	Левый	Правый	Левый	Правый	Левый	Правый	Левый
До	37,3 ± 0,4	36,2 ± 0,5	39,7 ± 0,6	39,0 ± 0,2	41,4 ± 0,2	41,8 ± 0,3	38,2 ± 0,3	39,4 ± 0,5**
После	38,4 ± 0,3	37,1 ± 0,2	43,3 ± 0,2*	41,6 ± 0,2*	43,2 ± 0,3*	44,1 ± 0,2*	39,3 ± 0,2**	41,2 ± 0,2**

* различия между показателями до и после игры достоверны ($P < 0,05$);

** различия между показателями КЧСМ для красного и синего цветов после работы достоверны ($P < 0,05$)

Таблица 3

Результаты исследования КЧСМ до и после игры «Пасьянс»

Цвет	Красный		Желтый		Зеленый		Синий	
Глаз	Правый	Левый	Правый	Левый	Правый	Левый	Правый	Левый
До	37,3 ± 0,4	36,0 ± 0,4	39,7 ± 0,6	39,0 ± 0,2	41,6 ± 0,2	42,0 ± 0,3	38,7 ± 0,2	39,0 ± 0,5**
После	38,3 ± 0,2	37,3 ± 0,4	42,3 ± 0,2*	42,1 ± 0,2*	43,3 ± 0,4*	44,2 ± 0,4*	39,1 ± 0,6	41,3 ± 0,6**

* различия между показателями до и после игры достоверны ($P < 0,05$);

** различия между показателями КЧСМ для красного и синего цветов после работы достоверны ($P < 0,05$)

представленным в табл. 2, очевидно, что после игры отмечается достоверное повышение значения КЧСМ на желтый, зеленый и синий цвета. Причем после игры, также как и после работы с текстом, значения КЧСМ для красного и синего цветов достоверно различаются, что не согласуется с литературными данными [12].

В табл. 3 представлены результаты исследования динамики КЧСМ в процессе компьютерной игры «Пасьянс». В этом случае также отмечается достоверное увеличение показателей на все четыре цвета.

Сравнивая результаты, представленные в таблицах 2 и 3, можно заметить, что как исходные, так и конечные показатели для двух видов игр сходны, наблюдается достоверный рост показателей для зеленого, синего и желтого цветов. Можно отметить, что между показателями правого и левого глаз существует небольшая разница, которая увеличивается после работы, что также свидетельствует о развитии зрительного утомления.

Выводы

Таким образом, можно сделать следующие выводы:

1. Разработанный автоматизированный прибор пригоден для проведения исследований КЧСМ в динамике зрительного труда.
2. Введение дополнительного светового стимула желтого цвета позволяет получить новую информацию о состоянии зрительной и центральной нервной системы.
3. Использование прибора в составе комплекса для диагностики функционального состояния пользователей ПК позволит получить более полную информацию о механизмах развития зрительного утомления.

Список литературы: 1. Казарян Э.Э., Мамиконян В.Р. Влияние компьютеров на соматическое здоровье и орган зрения пользователей // Рефракционная хирургия и офтальмология. - 2003. - том 3 (№1). - С. 77-81. 2. Кочина М.Л., Будянская Э.Н., Яворский А.В. и др. Возможность ультрафиолетового повреждения глаз пользователей видеодисплейных терминалов. // Клиническая и экспериментальная медицина. - 2003. - № 3-4. - С.167-171. 3. Шамшинова А.М., Волков В.В. Функциональные методы исследования в офтальмологии. - М., 1998. - С. 33-35; 187-190. 4. Алферова Ю.А. Нарушения зрительных функций в ранней диагностике глаукомы. - Автореф. дис. канд. мед. наук. - М., 1998. 5. Голубцов К.В., Курман

И.Г., Хейло Т.С. и др. Мелькающий свет в диагностике и лечении патологических процессов зрительной системы человека // Информационные процессы. - 2003. - Т. 3. - №2. - С. 114-122. 6. Владимирский Б.М., Власкина Л.А. Метод обнаружения изменений функционального состояния человека-оператора // Физиология человека. - 1987. - Т.14, № 5. - С.863-865. 7. Нетудыхатка О.Ю. Роль критической частоты слияния мельканий в оценке напряженности труда моряков // Офтальмологический журнал. - 1987. - № 5. - С. 300 - 303. 8. Кочина М.Л., Сайковская Л.Ф. Автоматизированный комплекс для диагностики функционального состояния пользователей ПК // «Радиотехника»: Всеукр. Межвед. науч.-техн. сб. - Х.: 2006. - №146. - С. 49-54. 9. Рогатина Е.В., Голубцов К.В. Критическая частота слияния мельканий в дифференциальной диагностике патологии зрительного анализатора у детей // Вестник офтальмологии. - 1997. - №6. - С.20-22. 10. Хайненкен Э. Эффекты последствий и критическая частота мельканий в зависимости от возраста // Когнитивная геронтология. - 1994. - Т. 2, №1. - С.79-83. 11. Кочина М.Л., Чумаков В.И., Сайковская Л.Ф. и др. Результаты исследования динамики функциональных показателей зрительной системы и психофизиологических показателей студентов в процессе работы на компьютере // Матер.16 Междунар. крымской конф. «СВЧ-техника и телекоммуникационные технологии»: Севастополь. - 2006. - С.54-55. 12. Волков В.В., Луизов А.В., Овчинников Б.В. Эргономика зрительной деятельности человека. - Л.: Машиностроение, 1989. - 111 с.

Поступила в редколлегию 30.10.2008

УДК: 612,84-003,96:681,3

Результати дослідження частотних характеристик зорової системи з використанням автоматизованого приладу. / Л.Ф. Сайківська // Біоніка інтелекту: наук.-техн. журнал - 2008. - № 2 (69). - С. 173-176.

В статті наведено метод та прилад для дослідження частотних характеристик зорової системи людини. Стверджується, що використання наведеного приладу в складі комплексу для діагностики функціонального стану користувачів ПК дозволить отримати більш повну інформацію про механізм розвитку зорового стомлення.

Табл. 3. Іл. 2. Бібліогр.: 12 найм.

UDK 612,84-003,96:681,3

The results of research of frequency characteristics of visual system with use of the automated complex / L.F. Saikovskaya // Bionics of Intelligence: Sci. Mag. - 2008. - № 2 (69). - P. 173-176.

In work it is offered to use a method and equipment of research frequency characteristics of visual system in diagnostics of functional stays of hymen.

Tab. 3. Fig. 2. Ref.: 12 items.

ОБ АВТОРАХ

Авраменко Валерий Павлович	73	д-р техн. наук, профессор кафедры информационных управляющих систем Харьковского национального университета радиоэлектроники
Алексеев Владимир Олегович	67	канд. техн. наук, доцент Харьковского национального автомобильно-дорожного университета
Бисикало Олег Владимирович	89	канд. техн. наук, доцент кафедры экономической кибернетики и информатики Винницкого государственного аграрного университета
Бондаренко Михаил Федорович	3, 13, 23, 106	д-р техн. наук, профессор, ректор Харьковского национального университета радиоэлектроники
Герасин Сергей Николаевич	159	д-р техн. наук, профессор кафедры высшей математики Харьковского национального университета радиоэлектроники
Гороховатский Владимир Алексеевич	45	канд. техн. наук, доцент кафедры информатики Харьковского национального университета радиоэлектроники
Гребенник Игорь Валериевич	56	д-р техн. наук, профессор кафедры системотехники Харьковского национального университета радиоэлектроники
Деревянко Алексей Геннадиевич	61	магистр кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники
Дударь Виктория Валерьевна	61	ассистент кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники
Ерохин Андрей Леонидович	36	д-р техн. наук, профессор, начальник кафедры информатики Харьковского национального университета внутренних дел
Ерошенко Александр Русланович	32	младший научный сотрудник Украинского языково-информационного фонда НАН Украины, г. Киев
Збитнева Майя Вячеславовна	61	канд. техн. наук, доцент кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники
Калиманов Владимир Григорьевич	132	соискатель кафедры радиоэлектронных устройств Харьковского национального университета радиоэлектроники, зам. начальника учебного отдела ЦНИЛ Харьковского национального медицинского университета
Калистратова Светлана Юрьевна	73	студентка Харьковского национального университета радиоэлектроники
Кобозева Алла Анатольевна	152	канд. физ.-мат. наук, доцент Одесского национального политехнического университета
Кобылин Олег Анатольевич	51	канд. техн. наук, ассистент кафедры информатики Харьковского национального университета радиоэлектроники
Коляда Михаил Георгиевич	101	кандидат педагогических наук, доцент кафедры радиотехники и технической защиты информации Донецкого национального технического университета
Козлов Максим Александрович	159	аспирант кафедры высшей математики Харьковского национального университета радиоэлектроники
Коноплянко Зеновий Дмитриевич	164	д-р техн. наук, профессор кафедры компьютерных технологий Львовского института банковского дела Университета банковского дела НБУ
Кораблев Николай Михайлович	125	канд. техн. наук, доцент кафедры ЭВМ Харьковского национального университета радиоэлектроники
Кочина Марина Леонидовна	132	д-р биол. наук, старший научный сотрудник, профессор, зав. биофизическим отделом ЦНИЛ Харьковского национального медицинского университета
Левыкин Виктор Макарович	115	д-р техн. наук, профессор, заведующий кафедрой информационных управляющих систем Харьковского национального университета радиоэлектроники

Марьин Сергей Александрович	95	канд. техн. наук, доцент кафедры информационно-документных систем Харьковской государственной академии культуры
Машталир Владимир Петрович	40	д-р техн. наук, профессор, декан факультета компьютерных наук Харьковского национального университета радиоэлектроники
Мишин Алексей Петрович	95	студент Харьковского национального университета радиоэлектроники
Олейник Алексей Александрович	137	студент Запорожского национального технического университета
Ольшевский Андрей Иванович	145	доцент кафедры программного обеспечения интеллектуальных систем Донецкого государственного университета информатики и искусственного интеллекта
Петров Эдуард Георгиевич	120	д-р техн. наук, профессор, заведующий кафедрой системотехники Харьковского национального университета радиоэлектроники
Починский Максим Юрьевич	145	аспирант Донецкого государственного университета информатики и искусственного интеллекта
Путятин Александра Евгеньевна	40	аспирант кафедры информатики Харьковского национального университета радиоэлектроники
Работягов Андрей Валентинович	106	канд. техн. наук, научный сотрудник кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники
Репка Виктория Борисовна	95	канд. техн. наук, доцент кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники
Романова Татьяна Евгеньевна	56	д-р техн. наук, профессор, старший научный сотрудник Харьковского института проблем машиностроения им. А.Н.Подгорного НАН Украины
Сайковская Лилия Федоровна	173	аспирант кафедры радиоэлектронных устройств Харьковского национального университета радиоэлектроники
Сакало Евгений Сергеевич	40	аспирант кафедры информатики Харьковского национального университета радиоэлектроники
Сироджа Игорь Борисович	77	д-р техн. наук, профессор прикладной и вычислительной математики Национального аэрокосмического университета им. Н.Е. Жуковского «ХАИ»
Сорокина Ирина Витальевна	125	аспирант кафедры ЭВМ Харьковского национального университета радиоэлектроники
Субботин Сергей Александрович	137	канд. техн. наук, лауреат премии Президента Украины, доцент кафедры программных средств Запорожского национального технического университета
Турута Алексей Петрович	36	аспирант кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники
Федорова Татьяна Николаевна	84	стажер-исследователь кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники
Чайникова Татьяна Сергеевна	120	аспирант кафедры системотехники Харьковского национального университета радиоэлектроники
Чалый Сергей Федорович	115	д-р техн. наук, профессор кафедры информационных управляющих систем Харьковского национального университета радиоэлектроники
Четвериков Григорий Григорьевич	84	д-р техн. наук, профессор кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники
Шабанов-Кушнаренко Сергей Юрьевич	3, 13, 23,	д-р техн. наук, профессор кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники
Шабанов-Кушнаренко Юрий Петрович	3, 13, 23,	д-р техн. наук, профессор кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники
Шеховцов Сергей Борисович	56	канд. техн. наук, доцент Харьковского национального университета внутренних дел
Щепковский Сергей Вадимович	106	ведущий инженер кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники

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 (кегль 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/м², напечатанной на лазерном принтере, в 2-х экземплярах) необходимо подать следующие документы:

1. Заявку утвержденного образца, подписанную всеми авторами:

«Прошу принять статью ... (указать фамилии авторов, название статьи) объемом ... страниц (указать количество страниц) для опубликования в научно-техническом журнале «Бионика интеллекта». Оплату гарантируем.

Сведения об авторах: (указать фамилию, имя, отчество каждого автора, место работы, ученую степень, научное звание, контактный телефон, почтовый и электронный адреса).

Подписи авторов».

2. Акт экспертизы о возможности опубликования материалов в открытой печати.

3. Рецензию, подписанную доктором наук.

4. Сведения об авторах.

5. Электронный вариант рукописи, реферата и сведений об авторах.

6. Оплату за публикацию.

Наукове видання

БІОНІКА ІНТЕЛЕКТУ
інформація, мова, інтелект

Науково-технічний журнал

№ 2 (69)

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 № 48 від 28.11.2008)

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

Підписано до друку 17.12.2008. Формат 60 x 84 ¹/₈. Друк ризографічний.
Папір офсетний. Гарнітура Newton. Умов. друк. арк. 21,0. Обл.-вид. арк. 20,8.
Тираж 100 прим. Зам. № .

Надруковано в навчально-науковому видавничо-поліграфічному центрі ХНУРЕ
61166, Харків-166, просп. Леніна, 14