

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ УКРАИНЫ
ХАРЬКОВСКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ
РАДИОЭЛЕКТРОНИКИ

ISSN 0555-2656

БИОНИКА ИНТЕЛЛЕКТА

ИНФОРМАЦИЯ, ЯЗЫК, ИНТЕЛЛЕКТ

№ 1 (70)

2009

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> Идентификация объектов, описываемых векторами и линейными операциями над ними	3
<i>Бондаренко М. Ф., Шабанов-Кушнарченко С. Ю., Шабанов-Кушнарченко Ю. П.</i> Разработка систем условий, обеспечивающих существование линейного конечномерного оператора	11
<i>Бондаренко М. Ф., Шабанов-Кушнарченко С. Ю., Шабанов-Кушнарченко Ю. П.</i> Разработка теории структурной метризации конечномерного арифметического пространства	17
<i>Асеев Г. Г.</i> Методы интеллектуального анализа данных в электронных хранилищах	28
<i>Бодянский Е. В., Волкова В. В., Колчигин Б. В.</i> Самообучающаяся нейро-фаззи система для адаптивной кластеризации текстовых документов	34
<i>Ерохина А. Л., Кольченко А. В.</i> Интеллектуальный контроль пользователей в информационных системах	39
<i>Иванов В. Г., Шевченко И. В.</i> Представление имитационных моделей дискретно-событийных систем с помощью CASE средства E-NetSim	44
<i>Кораблев Н. М., Сорокина И. В.</i> О сходимости иммунных алгоритмов адаптации систем нечеткого вывода	50
<i>Четвериков Г. Г., Федорова Т. М., Вечірська І. Д., Богданов Д. Ф.</i> Моделювання фрагментів природної мови алгебро-логічними засобами (на прикладі k -значної реалізації морфологічного аналізу)	55
<i>Кригін М. Ю.</i> Дослідження відповідності між фонетичними системами та графічним представленням лексики східнослов'янських мов	60
<i>Ситник Л. Г.</i> Решение систем конъюнктивных и смешанных систем и их графовая интерпретация	64
<i>Хашан Т. С.</i> Метод автоматического определения границ шумового турбулентного источника в звуковом сигнале ...	68
<i>Удовенко С. Г., Дибее Г.</i> Робастное управление динамическим процессом на основе нечеткой модели	72
<i>Левыкин В. М., Чалый С. Ф., Селезнев С. Н.</i> Знание-ориентированная модель представления и обработки бизнес-объектов в процессах электронного документооборота	77
<i>Oleinick A. I.</i> Brownian motion simulation of vesicle actin tail formation	82
<i>Винарский В. Я., Маиталир С. В., Сакало Е. С., Шербинин К. С.</i> Робастные алгоритмы самообучения карты Кохонена в задаче обработки изображений	86
<i>Любченко В. А., Кобылин О. А.</i> Нормализация проективных преобразований методом одномерных корреляций	90
<i>Агарков А. В.</i> Структурное описание изображений в виде графа для решения задач распознавания	95
<i>Аксак Н. Г., Тыхун А. Ю., Барковская О. Ю., Солдатов А. С.</i> Распознавание изображений антропоморфного объекта	102
<i>Олейник А. А., Субботин С. А.</i> Упрощение структуры нейросетей на основе островной модели эволюционного поиска	107
<i>Одейчук А. Н.</i> Обобщенный критерий эффективности моделей прогнозирования временных рядов в информационных системах	113
<i>Фролова Г. А.</i> Проблемы организации интеллектуальной поддержки принятия решений в САПР ТП на мелкосерийных приборостроительных предприятиях	120
<i>Бісікало О. В.</i> Формалізація образної пам'яті людини у вигляді графу	127
<i>Фесенко У. А., Ренка В. Б.</i> Статистический анализ информативности теста Маллампати как предиктора сложной интубации трахеи у детей	132
<i>Ляховец А. В.</i> Преимущества модели определения кредитоспособности субъекта образования, основанной на анализе данных, разбитых на смысловые подгруппы	137
<i>Шубкина О. В.</i> Интеллектуальный анализ текстов в системах менеджмента знаний	142
Об авторах	147
Правила оформления рукописей для авторов научно-технического журнала «Бионика интеллекта»	149

УДК 519.7



ИДЕНТИФИКАЦИЯ ОБЪЕКТОВ, ОПИСЫВАЕМЫХ ВЕКТОРАМИ И ЛИНЕЙНЫМИ ОПЕРАЦИЯМИ НАД НИМИ

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

^{1, 2, 3}ХНУРЭ, г. Харьков, Украина

Развивается теория компараторной идентификации для случая, когда объект идентификации описывается векторами арифметического пространства. Приведены содержательные интерпретации теории. Предложена модель компараторной идентификации, позволяющая идентифицировать линейные конечномерные объекты.

КОМПАРАТОРНАЯ ИДЕНТИФИКАЦИЯ, ЦВЕТОВОЕ ЗРЕНИЕ, АРИФМЕТИЧЕСКОЕ ПРОСТРАНСТВО

Введение

Многие сигналы естественно представлять в виде векторов, то есть наборов отдельных компонентов (признаков) [1]. Число компонентов вектора называется его размерностью. В роли компонентов вектора обычно выступают вещественные числа [2]. Множество сигналов, каждый из которых представлен в виде вектора размерности n , называется пространством размерности n . Например световые излучения обычно представляют в виде спектров, то есть конечных или бесконечных наборов вещественных чисел. Говорят о двумерности поля зрения человека, имея в виду, что каждая его точка может быть охарактеризована парой вещественных чисел – ее координатами. Говорят также о трехмерном пространстве цветовых ощущений человека, о многомерном пространстве векторов, характеризующих место работы человека (зарплата, продолжительность отпуска, расстояние от места жительства до места работы и тому подобное).

Действия над векторами чаще всего описываются линейными операциями. Пространство, на котором определены линейные операции, называется линейным [3]. Примерами линейных операций могут служить сложение световых излучений, усиление или ослабление интенсивности светового излучения, преобразование светового излучения в цвет его опущения, оценка человеком места работы и тому подобное. В этой статье ставится задача отыскания таких полных перечней свойств, с помощью которых можно было бы осуществлять компараторную идентификацию объектов, подпадающих под понятие линейного пространства и под понятие линейной операции над векторами линейных пространств.

1. Идентификация объектов, описываемых векторами

Разработку сформулированной проблемы начнем с идентификации сигналов, описываемых в виде векторов некоторого пространства.

Пусть A – множество каких-нибудь сигналов. Предположим, что на A определены два предиката эквивалентности $E_1(x, y)$ и E_2 , удовлетворяющие условию:

$$\forall x, y \in A (E_1(x, y) \wedge E_2(x, y)) \supset x = y. \quad (1)$$

С помощью условия (1) в дальнейшем будет введено понятие двумерного пространства сигналов. Чтобы пояснить, как это делается, мы предварительно решим задачу о введении поля зрения человека. Рассмотрим испытуемого, голова которого неподвижна в пространстве и находится в вертикальном положении. Один глаз испытуемого смотрит прямо перед собой, его взгляд направлен на неподвижную точку фиксации, второй глаз закрыт. Испытуемому по очереди предъявляют различные точки ξ окружающего физического пространства и предлагают ответить на вопрос, видит ли он их или нет. В роли предъявляемой точки может использоваться, например, точечный источник света, перемещаемый в пространстве.

Пусть M – множество всех точек пространства. $N(\xi)$ – предикат на M , реализуемый испытуемым в данном опыте. Предикат N делит все пространство M на две части: если $N(\xi) = 1$, то точка ξ находится в зоне видимости $N \subseteq M$ испытуемого; если же $N(\xi) = 0$, то – за ее пределами, то есть в области $N \setminus M$. Область N имеет вид конуса неправильной формы (приблизительно кругового) с вершиной в оптическом центре глаза.

Из множества N произвольно выбираем две точки ξ и η и предъявляем их испытуемому. Последний должен установить, находятся ли они в точности одна за другой или нет. Своими ответами испытуемый реализует некоторый предикат $Q(\xi, \eta)$ на $N \times N$. Если $Q(\xi, \eta) = 1$, то субъективные образы x и y точек ξ и η совмещаются друг с другом в поле зрения испытуемого; если же $Q(\xi, \eta) = 0$, то не совмещаются. Опыт показывает, что предикат Q рефлексивен, симметричен и транзитивен (с той точностью, в пределах которой значения предиката Q можно считать однозначными). Действительно, пусть точкам ξ, η и ζ физического пространства N соответствуют их образы x, y, z – точки поля зрения испытуемого. Тогда при совпадении точек ξ и η совпадут и их субъективные образы x и y в поле зрения; если образы x и y двух точек в паре (ξ, η) совпадают, то они совпадут и для пары точек (η, ζ) : если образы x, y и y, z точек в парах (ξ, η) и (η, ζ) совпадают, то образы x и z совпадут и для пары точек (ξ, ζ) .

Следовательно, предикат $Q(\xi, \eta)$ есть эквивалентность. Он разбивает всю видимую часть N физического пространства M на слои, каждый из которых представляет собой бесконечный луч, исходящий из оптического центра глаза. Каждый такой луч определяет одну точку поля зрения. Множество всех точек поля зрения обозначаем символом A . Будем предъявлять испытуемому пары (x, y) точек x и y поля зрения, предлагая ему установить, видятся ли они на одной вертикали или нет. В другой серии экспериментов испытуемому предлагается установить, видятся ли точки x и y на одной горизонтали или нет.

Важно подчеркнуть, что от испытуемого требуется определить не взаимное положение двух точек физического пространства, а лишь субъективно воспринимаемое взаимное положение образов этих точек. Если испытуемому кажется, что предъявленные ему две точки поля зрения находятся на одной вертикали, то отсюда еще не следует, что соответствующие им точки физического пространства тоже лежат на одной вертикали. Своими ответами испытуемый реализует два предиката $E_1(x, y)$ и $E_2(x, y)$. Если $E_1(x, y) = 1$, то точки x и y поля зрения кажутся испытуемому находящимися на одной вертикали, если же $E_1(x, y) = 0$, то они кажутся не находящимися на ней. Если $E_2(x, y) = 1$, то точки x и y поля зрения воспринимаются лежащими на одной горизонтали, если же $E_2(x, y) = 0$, то они лежат на разных горизонталях.

Опыты показывают, что предикаты E_1 и E_2 суть эквивалентности. Они формируют два разбиения поля зрения: одно – в виде семейства горизонтальных линий и другое – в виде семейства вертикальных линий. Ясно, что любые вертикальная и горизонтальная линии могут пересекаться не более, чем в одной точке поля зрения. Случай, когда вертикаль и горизонталь не пересекаются, возможен: это происходит тогда, когда точка их пересечения попадает в область слепого пятна или же выходит за границы поля зрения. Таким образом, предикаты E_1 и E_2 удовлетворяют условию (1), которое в данной интерпретации гласит: если точки x и y поля зрения A лежат одновременно на одной вертикальной линии и на одной горизонтальной, то они совпадают друг с другом, то есть $x = y$. Образует множество B_1 всех вертикальных и множество B_2 всех горизонтальных линий в поле зрения A . Теперь каждую точку x поля зрения A можно представить в виде пары соответствующих ей координат (u_1, u_2) , где $u_1 \in B_1$ – вертикаль, проходящая через точку x , и $u_2 \in B_2$ – горизонталь, проходящая через ту же точку. Таким образом, множество точек поля зрения A мы превратили за счет введения двух эквивалентностей E_1 и E_2 в двумерное пространство T , являющееся подмножеством декартова произведения $B_1 \times B_2$ множеств B_1 и B_2 .

В общем случае двумерное пространство T для множества A вводится предикатами E_1 и E_2 , удовлетворя-

ющими условию (1) следующим образом. Формируем разбиения B_1 и B_2 и характеристические функции $f_1: A \rightarrow B_1$ и $f_2: A \rightarrow B_2$, соответствующие эквивалентностям E_1 и E_2 . Классы $u_1 = f_1(x)$ и $u_2 = f_2(x)$ разбиений B_1 и B_2 принимаем в качестве абсциссы и ординаты точки x . Пара (u_1, u_2) однозначно определяет точку x . Совокупность всех пар (u_1, u_2) , соответствующих всем точкам x множества A , принимаем в роли двумерного пространства T для множества A . Пространство T является подмножеством декартова произведения $B_1 \times B_2$ множеств B_1 и B_2 . Множества B_1 и B_2 принимаем в роли координатных осей пространства T . В том случае, когда пространство T совпадает с $B_1 \times B_2$, оно называется полным. Для полноты пространства T необходимо и достаточно, чтобы эквивалентности E_1 и E_2 , его вводящие, дополнительно удовлетворяли условию:

$$\forall x_1, x_2 \in A \exists y \in A (E_1(x_1, y) \wedge E_2(x_2, y)). \quad (2)$$

Если T – полное пространство, то существует сюръекция $y: B_1 \times B_2 \rightarrow A$, взаимно однозначно переводящая векторы (u_1, u_2) пространства T в соответствующие им точки x множества A . Обратный перевод точки x множества A в компоненты u_1, u_2 ее вектора (u_1, u_2) осуществляется функциями $f_1: A \rightarrow B_1$ и $f_2: A \rightarrow B_2$.

Изложенный выше способ введения двумерного пространства легко обобщается на случай пространства произвольной размерности n . Неполное пространство T для множества A вводится эквивалентностями E_1, E_2, \dots, E_n на A , удовлетворяющими условию:

$$\forall x, y \in A (E_1(x, y) \wedge E_2(x, y) \wedge \dots \wedge E_n(x, y) \supset x = y).$$

Для полноты пространства T необходимо и достаточно, чтобы эквивалентности E_1, E_2, \dots, E_n , его вводящие, дополнительно удовлетворяли условию:

$$\forall x_1, x_2, \dots, x_n \in A \exists y \in A (E_1(x_1, y) \wedge E_2(x_2, y) \wedge \dots \wedge E_n(x_n, y)). \quad (4)$$

Пространство T для множества A вводится предикатами E_1, E_2, \dots, E_n , удовлетворяющими условию (3), следующим образом. Формируем разбиения B_1, B_2, \dots, B_n и характеристические функции $f_1: A \rightarrow B_1, f_2: A \rightarrow B_2, \dots, f_n: A \rightarrow B_n$, соответствующие эквивалентностям E_1, E_2, \dots, E_n . Классы $u_1 = f_1(x), u_2 = f_2(x), \dots, u_n = f_n(x)$ разбиений E_1, E_2, \dots, E_n принимаем в качестве координат точки x . Набор (u_1, u_2, \dots, u_n) однозначно определяет точку x . Совокупность всех наборов (u_1, u_2, \dots, u_n) , соответствующих точкам x множества A , принимаем в роли n -мерного пространства для множества A . Пространство T является подмножеством декартова произведения $B_1 \times B_2 \times \dots \times B_n$ множеств B_1, B_2, \dots, B_n . Множества B_1, B_2, \dots, B_n принимаем в роли координатных осей пространства T . Только в том случае, когда пространство T полно, оно совпадает с де-

картовым произведением $B_1 \times B_2 \times \dots \times B_n$. Любое n -мерное пространство T для множества A полностью характеризуется множествами B_1, B_2, \dots, B_n сюръекциями $f_1 : A \rightarrow B_1, f_2 : A \rightarrow B_2, \dots, f_n : A \rightarrow B_n$ и однозначным отображением g , действующим из $B_1 \times B_2 \times \dots \times B_n$ на A . Для полного пространства T отображение g превращается в сюръекцию $g : B_1 \times B_2 \times \dots \times B_n \rightarrow A$. Сюръекции f_1, f_2, \dots, f_n однозначно определяют эквивалентности E_1, E_2, \dots, E_n , а наличие однозначного отображения g обеспечивает выполнение условия (3). Наличие же сюръекции g обеспечивает выполнение условия (4).

2. Идентификация объектов, описываемых векторами арифметического пространства

Пусть A – множество и $T = B_1 \times B_2 \times \dots \times B_n$ – его полное пространство. Предположим, что $B_1 = B_2 = \dots = B_n = R$ – множество вещественных чисел с заданными на нем операциями сложения и умножения. Множество R называется числовым полем, а его элементы – скалярами. Множество $T = R^n$ называется n -мерным арифметическим пространством [4], если на нем введены операции $x + y$ сложения векторов $x, y \in R^n$ и операция $\alpha * x$ умножения вещественного числа $\alpha \in R$ на вектор $x \in R^n$, определяемые следующим образом: если $x = (\alpha_1, \alpha_2, \dots, \alpha_n)$ и $y = (\beta_1, \beta_2, \dots, \beta_n)$, то

$$x + y = (\alpha_1 + \beta_1, \alpha_2 + \beta_2, \dots, \alpha_n + \beta_n), \tag{5}$$

если $x = (\alpha_1, \alpha_2, \dots, \alpha_n)$, то

$$\alpha x = (\alpha * \alpha_1, \alpha * \alpha_2, \dots, \alpha * \alpha_n). \tag{6}$$

Из этих определений следует, что для любых $\alpha, \beta, \gamma \in R$ и $x, y, z \in R^n$ $\alpha + \beta = \beta + \alpha$, $\alpha\beta = \beta\alpha$, $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$, $(\alpha\beta)\gamma = \alpha(\beta\gamma)$, $(\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma$, $0 + \alpha = \alpha$, $\alpha + (-\alpha) = 0$, $\alpha 1 = \alpha$, $\alpha \frac{1}{\alpha} = 1$, $x + y = y + x$, $(x + y) + z = x + (y + z)$, $x + 0 = x$, $x + (-x) = 0$, $1x = x$, $\alpha(\beta x) = (\alpha\beta)x$, $(\alpha + \beta)x = \alpha x + \beta x$, $\alpha(x + y) = \alpha x + \alpha y$. Здесь $0 = (0, 0, \dots, 0)$, $-x = (-1)x$.

Пользуясь приведенными определениями, покажем, что поле зрения человека можно с определенным приближением идентифицировать как двумерное арифметическое пространство. С этой целью сначала установим, что совокупности B_1 и B_2 всех вертикальных и горизонтальных линий поля зрения A можно отождествить с множествами вещественных чисел. Проведем горизонталь α и вертикаль β через точку фиксации 0 поля зрения (рис. 1), называя эти линии координатными осями поля зрения – осью абсцисс и осью ординат. Поставим во взаимно однозначное соответствие каждой вертикали u_1 точку ее пересечения с осью абсцисс α и каждой горизонтали u_2 – точку ее пересечения с осью ординат β .

Вместо вертикалей и горизонталей поля зрения теперь будем рассматривать соответствующие им точки на осях абсцисс и ординат. Выберем на оси абсцисс

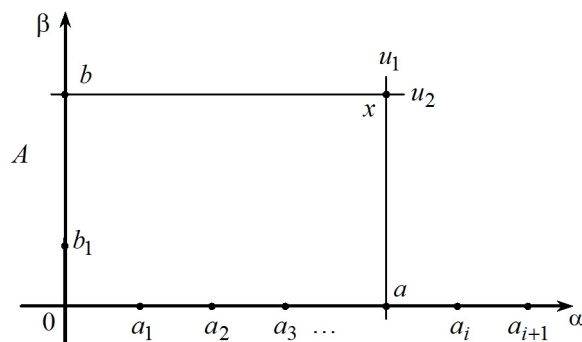


Рис. 1

какую-нибудь точку a_1 , близкую к точке фиксации 0, но не совпадающую с ней. Затем правее точки a_1 находим на оси α точку a_2 такую, чтобы расстояния между точками 0, a_1 и 0, a_2 совпали друг с другом. После этого справа от точки a_2 отыскиваем точку a_3 , исходя из условия равенства расстояний между точками 0, a_1 и a_2 , a_3 , и так далее.

Операцию отыскания точки a_{i+1} по точке a_i отождествляем с функцией счёта $a_{i+1} = a_i + 1$, точку a_1 отождествляем с единицей натурального ряда, а бесконечный ряд точек a_1, a_2, \dots на оси абсцисс отождествляем со всем натуральным рядом. Все четыре аксиомы натурального ряда, приведенные в предыдущей главе, выполняются. Приближенность такого способа идентификации поля зрения обнаруживается в том, что после выполнения некоторого конечного числа шагов мы доходим до границы поля зрения и ряд точек обрывается. Аналогичным образом на оси α выявляются точки, соответствующие рациональным и вещественным числам, а также выявляются операции над этими точками, которые может производить испытуемый, соответствующие сложению и умножению чисел. Кроме того, выявляется способность испытуемого упорядочивать точки поля зрения на оси абсцисс, которую идентифицируем как отношение порядка на множестве вещественных чисел. Неточность такого способа идентификации состоит в том, что точки, достаточно близкие друг к другу, глазом не различаются ввиду его ограниченной разрешающей способности, кроме того, операции сложения и умножения точек поля зрения на оси α оказываются не всюду определенными, а частичными всякий раз, когда сумма или произведение представляет собой точку, выходящую за пределы поля зрения или попадающую в область слепого пятна. В остальном все аксиомы теории вещественных чисел, перечисленные в предыдущей главе, выполняются. Аналогично, точки оси ординат поля зрения также идентифицируем как вещественные числа. При введении точки b_1 , соответствующей единице натурального ряда на оси β , следует учесть, что расстояния между точками 0, a_1 и 0, b_1 должны совпадать. Сложение произвольных точек поля зрения

и умножения их на число формализуется с помощью определений (5) и (6). Оказывается, что испытуемый обладает способностью производить такие операции. Возможность производить эти действия в конечном счете основывается на способности испытуемого устанавливать порядок на множестве расстояний между точками поля зрения.

Рассмотрим еще одну необходимую для дальнейшего изложения содержательную интерпретацию арифметического пространства. Речь идет о представлении световых излучений векторами n -мерного арифметического пространства. Как известно, любое световое излучение можно разложить призмой в спектр, то есть на простые составляющие. Субъективно спектр светового излучения воспринимается как линейно упорядоченное в поле зрения множество зрительных ощущений разной цветности. Полученный отрезок разбиваем на n участков a_1, a_2, \dots, a_n , выбирая число n и размеры участков с таким расчетом, чтобы цветность на каждом участке не менялась. Мощность светового излучения измеряем болометром, представляющим собой термометр специальной конструкции. Опыт показывает, что имеет место взаимно однозначное соответствие между показаниями болометра и яркостью зрительного ощущения соответствующей линии в спектре. Этот факт дает возможность идентифицировать совокупность B_i всех яркостей i -той линии спектра как множество вещественных чисел. Пусть A – множество всех световых излучений. Каждое из множеств B_i ($i = \overline{1, n}$) абстрактно вводится с помощью эквивалентности E_i , определяемой следующим образом: если яркости зрительных ощущений i -той линии спектров двух световых излучений совпадают, то $E_i(x, y) = 1$, если же не совпадают, то $E_i(x, y) = 0$.

Опыт показывает, что испытуемый способен практически воспроизводить своим поведением каждый из предикатов E_i с довольно высокой точностью. Кроме того, оказывается, что предикаты E_i подчиняются условиям (3) и (4). Следовательно, эквивалентностями E_i можно ввести полное пространство $T = B_1 \times B_2 \times \dots \times B_n$ размерности n , соответствующее множеству A . Сложение, умножение и порядок на множествах B_i определяем как соответствующие операции над мощностями соответствующих спектральных линий, рассматривая эти мощности просто как вещественные числа. Сложение световых излучений определяем равенством (5) как покоординатное сложение их спектральных линий. Умножение вещественного числа на световое излучение определяем аналогично равенством (6). Физически сложение излучений осуществляется простым совмещением их в пространстве. Умножение числа на световое излучение достигается диафрагмированием светового потока (при этом число представлено площадью отверстия диафрагмы) или же приближением или удалением источника све-

та от освещаемой поверхности (в этом случае множитель светового излучения обратно пропорционален квадрату расстояния).

В роли базисных элементов p_1, p_2, \dots, p_n пространства T можно взять n излучений $p_1(\lambda), p_2(\lambda), \dots, p_n(\lambda)$, заданных соответственно на интервалах длин волн $[\lambda_0, \lambda_1], [\lambda_1, \lambda_2], \dots, [\lambda_{n-1}, \lambda_n]$. Каждое из этих излучений имеет на интервале своего задания постоянную интенсивность и охватывает единичную площадь. Здесь $[\lambda_{n-1}, \lambda_n]$ – диапазон длин волн электромагнитных колебаний, видимых глазом ($\lambda_0 = 380$ нм, $\lambda_n = 780$ нм) [5]. Пусть $x(\lambda)$ – непрерывный спектр светового излучения x , понимаемый в физическом смысле этого слова. Определим числа $\alpha_1, \alpha_2, \dots, \alpha_n$ следующим образом:

$$\alpha_i = \int_{\lambda_{i-1}}^{\lambda_i} x(\lambda) d\lambda. \quad (7)$$

Выбирая в роли n достаточно большое натуральное число и разбивая достаточно равномерно интервал $[\lambda_0, \lambda_n]$ точками $\lambda_1, \lambda_2, \dots, \lambda_{n-1}$, всегда можно добиться того, чтобы спектр $x(\lambda)$ любого излучения x с требуемой точностью совпал со спектром

$$\sum_{i=1}^n \alpha_i p_i(\lambda)$$

аппроксимирующего излучения $\alpha_1 p_1, \alpha_2 p_2, \dots, \alpha_n p_n$. Это означает, что с достаточной точностью выполняется равенство

$$x = \sum_{i=1}^n \alpha_i p_i(\lambda), \quad (8)$$

следовательно числа $\alpha_1, \alpha_2, \dots, \alpha_n$ можно принять в роли координат вектора x .

Физические свойства световых излучений таковы, что все свойства конечномерного векторного пространства для них выполняется в пределах точности измерения спектров излучений. Правда, действуют эти свойства на несколько суженной основе. Дело в том, что спектр $(\alpha_1, \alpha_2, \dots, \alpha_n)$ любого реального излучения не может иметь отрицательных компонентов, поскольку числа $\alpha_1, \alpha_2, \dots, \alpha_n$ задают мощность излучения на отдельных участках его спектра, которая всегда неотрицательна. Поэтому выполнение свойств арифметического векторного пространства может быть экспериментально продемонстрирована не для всех векторов пространства T , а только для некоторой его части T_0 , называемой положительным конусом [6]. Таким образом, операции над скалярами и векторами, введенные в векторном пространстве, в физическом смысле оказываются не всюду определенными, а частичными. Тем не менее, ничто не мешает при математических действиях со скалярами и векторами использовать также и физически неинтерпретируемые спектры с отрицательными компонентами, когда это окажется целесообразным. Так, например, к спектру реального излучения можно прибавить спектр фиктивного излучения с отрицатель-

ными компонентами при условии, что в результате получится спектр излучения из множества T_0 , то есть такого излучения, которое можно физически предъявить испытуемому. Действие свойств арифметического пространства можно условно распространить на все векторы полного пространства T , но при этом надо иметь в виду, что эти свойства приобретают физический смысл только в том случае, когда в них будут фигурировать лишь векторы из множества T_0 .

3. Установление условий существования линейного предиката

Пусть на декартовом квадрате m -мерного векторного пространства M над полем G задан предикат E . Базис (p_1, p_2, \dots, p_n) пространства M произвольно фиксирован. Предикат E назовем линейным, если для любых $x, y \in M$ он может быть выражен в виде

$$E(x, y) = D(F(x), F(y)), \quad (9)$$

где

$$F(x) = (xk_1, xk_2, \dots, xk_n), \quad (10)$$

$$xk_i = \xi_1 \chi_{1i} + \xi_2 \chi_{2i} + \dots + \xi_m \chi_{mi}. \quad (11)$$

Здесь k_1, k_2, \dots, k_n – фиксированные линейно независимые векторы, $\xi_1, \xi_2, \dots, \xi_m$ – координаты вектора x ; $\chi_{1i}, \chi_{2i}, \dots, \chi_{mi}$ – координаты вектора k_i ; D – предикат равенства, заданный на $G^n \times G^n$. Имеется в виду, что $n \leq m$. Символом F обозначен линейный оператор, отображающий M в G^n .

Предикат E назовем аддитивным, если для любых $x_1, x_2, y_1, y_2 \in M$ из $x_1 E y_1$ и $x_2 E y_2$ следует $(x_1 + x_2) E (y_1 + y_2)$. Предикат E назовем однородным, если для любых $\alpha \in G$ и $x, y \in M$ из $x E y$ следует $\alpha x E \alpha y$. Предикат E назовем n -мерным, если существуют векторы $e_1, e_2, \dots, e_n \in M$ такие, что равенство

$$E(x, \sum_{i=1}^n F_i(x) e_i) = 1 \quad (12)$$

выполняется для каждого $x \in M$ при единственном наборе коэффициентов $F_1(x), F_2(x), \dots, F_n(x)$. Здесь F_1, F_2, \dots, F_n – фиксированные функции, определенные на множестве M со значениями в множестве G .

Ниже формулируется и доказывается теорема об условиях существования линейного предиката.

Теорема. Предикат E линейен в том и только том случае, когда он рефлексивен, симметричен, транзитивен, аддитивен, однороден и n -мерен.

Доказательство. Необходимость. Предположим, что предикат E линейен, и выведем отсюда его рефлексивность, симметричность, транзитивность, аддитивность, однородность и n -мерность.

Выводим рефлексивность. Для любого $x \in M$ $F(x)DF(x)$. В силу линейности предиката E имеем: xEx .

Выводим симметричность. Предположим, что $x, y \in M$ таковы, что xEy . Тогда $F(x)DF(y)$, $F(y)DF(x)$, yEx .

Выводим транзитивность. Пусть $x, y, z \in M$ таковы, что xEy и xEz . Тогда $F(x)DF(y)$, $F(y)DF(z)$, $F(x)DF(z)$, xEz .

Выводим аддитивность. Пусть $x_1, x_2, y_1, y_2 \in M$ выбраны так, что $x_1 E y_1$ и $x_2 E y_2$. Отсюда следует, что $F(x_1)DF(y_1)$ и $F(x_2)DF(y_2)$. Ввиду линейности предиката E для любого $i = \overline{1, n}$ имеем $x_1 k_i = y_1 k_i$ и $x_2 k_i = y_2 k_i$. Отсюда по свойству (1) выводим $x_1 k_i + x_2 k_i = y_1 k_i + y_2 k_i$. Пользуясь законами (9), (13) и (14), получаем $(x_1 + x_2) k_i = (y_1 + y_2) k_i$. Следовательно $F(x_1 + x_2)DF(y_1 + y_2)$, а значит $(x_1 + x_2)E(y_1 + y_2)$.

Выводим однородность. Пусть $x, y \in M$ таковы, что xEy . Это означает, что $xk_i = yk_i$. Для произвольного $\alpha \in G$ в силу (17) имеем $\alpha(xk_i) = \alpha(yk_i)$. Пользуясь законами (7)–(9), получаем $(\alpha x)k_i = (\alpha y)k_i$. Отсюда следует $F(\alpha x)DF(\alpha y)$ $\alpha x E \alpha y$.

Выводим n -мерность. Выберем векторы $e_1, e_2, \dots, e_n \in M$ так, чтобы набор (e_1, e_2, \dots, e_n) был дуален набору векторов (k_1, k_2, \dots, k_n) . Как известно [7], такой набор всегда существует. Требование n -мерности предиката E означает, что уравнение (12) при каждом $x \in M$ однозначно разрешимо относительно набора коэффициентов $(F_1(x), F_2(x), \dots, F_n(x))$. Докажем это. Уравнение (12) согласно (9)–(11) равносильно системе уравнений

$$xk_j = \left(\sum_{i=1}^n F_j(x) e_i \right) k_j,$$

где $j = \overline{1, n}$. Используя законы (18)–(20), последнюю систему равенств переписываем в виде:

$$xk_j = \left(\sum_{i=1}^n F_i(x) \right) (e_i k_j).$$

В силу дуальности наборов векторов (e_1, e_2, \dots, e_n) и (k_1, k_2, \dots, k_n) имеем $e_i k_j = 0$ при $i \neq j$ и $e_i k_j = 1$ при $i = j$. Поэтому последняя система $F_1(x), F_2(x), \dots, F_n(x)$ равенств запишется в виде $F_j(x) = xk_j$. Это означает, что коэффициенты $F_1(x), F_2(x), \dots, F_n(x)$ при каждом x однозначно определены. Необходимость доказана.

Достаточность. Предположим, что предикат E рефлексивен, симметричен, транзитивен, аддитивен, однороден и n -мерен, и выведем отсюда его линейность. Докажем сначала, что функции $F_1(x), F_2(x), \dots, F_n(x)$, фигурирующие в условии n -мерности, линейны. Для этого нужно убедиться в их аддитивности и однородности [8]. В силу n -мерности предиката E для любых $x, y \in M$ имеет место равенство (12), а также равенства

$$E(y, \sum_{i=1}^n F_i(y) e_i) = 1, \quad (13)$$

$$E(x + y, \sum_{i=1}^n F_i(x + y) e_i) = 1, \quad (14)$$

Пользуясь свойством аддитивности предиката E , из (12) и (13) выводим

$$E(x + y, \sum_{i=1}^n F_i(x) e_i + \sum_{i=1}^n F_i(y) e_i) = 1. \quad (15)$$

В силу n -мерности предиката E множители при векторах e_i в (14) и (15) совпадают, поэтому $F_i(x+x) = F_i(x) + F_i(y)$ при всех $i = \overline{1, n}$. Таким образом, функции $F_1(x), F_2(x), \dots, F_n(x)$ аддитивны.

В силу n -мерности предиката E для любых $\alpha \in G$ и $x \in M$ имеем:

$$E(\alpha x, \sum_{i=1}^n F_i(\alpha x)e_i) = 1. \quad (16)$$

Используя однородность предиката E из (11) выводим

$$E(\alpha x, \sum_{i=1}^n F_i(x)e_i) = 1. \quad (17)$$

Производим преобразования в последнем равенстве с помощью законов (13) и (14):

$$E(\alpha x, \sum_{i=1}^n (\alpha F_i(x))e_i) = 1. \quad (18)$$

В силу n -мерности предиката E множители при векторах e_i в (16) и (17) совпадают, поэтому $F_i(\alpha x) = \alpha F_i(x)$ при любых α и x для всех $i = \overline{1, n}$. Таким образом, функции $F_1(x), F_2(x), \dots, F_n(x)$ однородны, следовательно, они линейны.

Любая линейная функция $F(x)$ на M со значениями в множестве G может быть представлена в виде $F(x) = xk$ [8], где k – некоторый вектор из M . Значит, найдутся векторы $k_1, k_2, \dots, k_n \in G$ такие, что

$$F_1(x) = xk_1, \quad F_2(x) = xk_2, \dots, \quad F_n(x) = xk_n. \quad (19)$$

Докажем далее, что для любых $x, y \in M$ xEu в том и только том случае, когда $F_i(x) = F_i(y)$ для всех $i = \overline{1, n}$. Пусть xEu . Из этого соотношения и из (13) с помощью свойства транзитивности предиката E выводим:

$$E(x, \sum_{i=1}^n F_i(y)e_i) = 1.$$

Из (12) и только что записанного равенства, используя свойство n -мерности предиката E , находим $F_i(x) = F_i(y)$ для всех $i = \overline{1, n}$. Пусть теперь $F_i(x) = F_i(y)$ для всех $i = \overline{1, n}$. Отсюда следует:

$$\sum_{i=1}^n F_i(x)e_i = \sum_{i=1}^n F_i(y)e_i.$$

Из последнего равенства и (13) находим

$$E(y, \sum_{i=1}^n F_i(x)e_i) = 1.$$

Из (12) и только что записанного равенства с помощью свойств симметричности и транзитивности выводим xEu .

Доказанное означает, что условие xEu равносильно равенству

$$(F_1(x), F_2(x), \dots, F_n(x)) = (F_1(y), F_2(y), \dots, F_n(y)).$$

Для каждой линейной функции $F: M \rightarrow G$ найдется такое $k \in M$, что $F(x) = xk$ для всех $x \in M$. Поэтому $F_i(x) = xk_i$ и $F_i(y) = yk_i$ для любых $x, y \in M$ и всех $i = \overline{1, n}$. Это означает, что условие xEu равносильно равенству $(xk_1, xk_2, \dots, xk_n) = (yk_1, yk_2, \dots, yk_n)$, то есть ра-

венству $F(x) = F(y)$. Итак, мы доказали, что любой рефлексивный, симметричный, транзитивный, аддитивный, однородный и n -мерный предикат может быть выражен в виде (9)–(11) при подходящем выборе векторов k_1, k_2, \dots, k_n . Осталось доказать линейную независимость векторов k_1, k_2, \dots, k_n . С этой целью установим, что функции $F_1(x), F_2(x), \dots, F_n(x)$ линейно независимы. Для этого достаточно доказать, что равенство

$$\sum_{i=1}^n \gamma_i F_i(x) = 0$$

выполняется для всех $x \in M$ лишь в том случае, когда все коэффициенты γ_i равны нулю.

Докажем последнее утверждение. Доказательство ведем от противного. Предположим, что это утверждение неверно. Тогда найдется такой номер j из множества $\{1, 2, \dots, n\}$, для которого при любом x

$$F_j(x) = \sum_{i=1, i \neq j}^n \beta_i F_i(x), \quad (20)$$

где $\beta_1, \beta_2, \dots, \beta_{j-1}, \beta_{j+1}, \dots, \beta_n$ – подходящие коэффициенты. Предикат E рефлексивен, значит $E(e_i, e_i) = 1$ для любого $j = \overline{1, n}$. Иначе говоря,

$$E(e_j, 0 * e_1 + 0 * e_2 + \dots + 0 * e_{j-1} + 0 * e_j + 0 * e_{j+1} + \dots + 0 * e_n) = 1.$$

По свойству n -мерности предиката E из последнего равенства выводим $F_j(e_j) = 1$; для всех же $i \neq j$ имеем $F_j(e_j) = 0$. Вместе с тем, подставляя в (20) $x = e_j$, находим:

$$F_j(e_j) = \beta_1 F_1(e_j) + \beta_2 F_2(e_j) + \dots + \beta_{j-1} F_{j-1}(e_j) + \beta_{j+1} F_{j+1}(e_j) + \dots + \beta_n F_n(e_j) = \beta_1 0 + \beta_2 0 + \dots + \beta_{j-1} 0 + \beta_{j+1} 0 + \dots + \beta_n 0.$$

Получили равенство $1 = 0$. Однако, известно [9], что в любом векторном пространстве $1 \neq 0$. Мы пришли к противоречию. Отсюда вытекает линейная независимость функций $F_1(x), F_2(x), \dots, F_n(x)$, а следовательно и линейная независимость векторов k_1, k_2, \dots, k_n . Достаточность доказана.

Итак, мы доказали, что все рефлексивные, симметричные, транзитивные, аддитивные, однородные и n -мерные предикаты, и только такие предикаты, могут быть представлены в виде

$$E(x, y) = D((F_1(x), F_2(x), \dots, F_n(x)), (F_1(y), F_2(y), \dots, F_n(y))),$$

где $F_1(x), F_2(x), \dots, F_n(x)$ – подходящие линейно независимые линейные функции. Теорема доказана.

4. Содержательная интерпретация линейного предиката

Прикладное значение теоремы состоит в том, что она указывает полную систему признаков линейной конечномерной операции F , идентифицируемой компараторным методом. Если предикат E , реализуемый системой компараторной идентификации объекта F , обладает свойствами рефлексивности, симметрично-

сти, транзитивности, однородности, аддитивности и n -мерности, то объект идентификации F можно описать в виде линейного оператора. В противном случае это невозможно. Таким образом, теорема указывает практический способ распознавания любого объекта, который можно идентифицировать рассматриваемым в диссертации способом. Рассмотрим способ практической проверки характеристических свойств линейного объекта на примере зрительной системы человека. О свойствах рефлексивности, симметричности и транзитивности системы цветового зрения человека речь уже шла в [10].

Специфика проверки свойств аддитивности, однородности и n -мерности в случае цветового зрения человека состоит в том, что реально не существует световых излучений со спектрами, имеющими отрицательные компоненты. Кроме того, испытуемому нельзя предъявить излучения слишком большой интенсивности, иначе глаз ослепнет. Таким образом, в множестве M всевозможных теоретически возможных входных сигналов органа зрения имеется некоторая часть M_1 всех тех световых излучений, которые реально могут быть предъявлены испытуемому в процессе идентификации его зрительной системы. Поскольку законы рефлексивности, симметричности и транзитивности выполняются для всех входных сигналов органа зрения, содержащихся в множестве M_1 , то они могут быть чисто формально распространены на все множество M . Производя такое доопределение предиката E , мы не вводим в противоречие с фактами зрения.

Закон аддитивности следует проверять на всех тех реальных излучениях $x_1, x_2, y_1, y_2 \in M_1$, суммы которых $x_1 + y_1$ и $x_2 + y_2$ также не выходят за пределы множества M_1 . Кроме того, в роли равновыглядящих излучений можно брать любые равные элементы x и x множества M с физически нереализуемыми спектрами, поскольку закон рефлексивности только что был распространен на все множество M . Таким образом, если реальные излучения x_1, y_1 выглядят для данного испытуемого равноцветными, то должны выглядеть равноцветными и излучения $x_1 + x$ и $y_1 + x$, даже если сигнал x имеет физически нереализуемый спектр с отрицательными компонентами. Единственное ограничение при выборе сигнала x состоит в том, чтобы сигналы $x_1 + x$ и $y_1 + x$ были физически реализуемы, то есть входили в состав множества M_1 . Если это условие не будет выполняться, то опыт с предъявлением суммарных излучений просто нельзя будет выполнить на практике. Известно, что при всевозможных проверках такого рода закон аддитивности выполняется с той точностью, с которой реализуется предикат E [11]. Это обстоятельство дает нам право, не вступая в противоречие с фактами, распространить действие закона аддитивности на все элементы множества M .

Закон однородности следует проверять на всех тех реальных излучениях $x, y \in M_1$ и числах $\alpha \in G$, произведения которых αx и αy не выходят за пределы множества M_1 . Известно [12], что при всевозможных проверках такого рода закон однородности выполняется с той точностью, с которой фактически реализуется предикат E . Опираясь на этот факт, мы распространяем действие закона однородности на все элементы множества M и множества G . Пользуясь законом однородности, можно еще более разнообразить проверку закона аддитивности. Пусть имеются две пары x_1, x_2 и y_1, y_2 одноцветных излучений. Тогда при любых $\alpha, \beta \in G$ элементы $\alpha x_1, \alpha x_2$ и $\beta y_1, \beta y_2$ следует признать одноцветными, даже если они физически нереализуемы. Если суммарные излучения $\alpha x_1 + \beta y_1, \alpha x_2 + \beta y_2$ могут быть предъявлены испытуемому (то есть принадлежат множеству M_1), то они должны породить в сознании испытуемого одинаковые цвета. В практике колориметрических измерений проводились и такого рода опыты, они неизменно подтверждали закон аддитивности [13].

Закон n -мерности при данной интерпретации в эксперименте проверяется при $n = 3$. Предположим, что на левом поле сравнения сформировано световое излучение $x + \alpha'_1 e_1 + \alpha'_2 e_2 + \alpha'_3 e_3$, а на правом – излучение $\alpha''_1 e_1 + \alpha''_2 e_2 + \alpha''_3 e_3$. Здесь e_1, e_2, e_3 – специально подобранные элементы множества M . В роли этих элементов не обязательно использовать физически реализуемые световые излучения. Коэффициенты $\alpha'_1, \alpha'_2, \alpha'_3, \alpha''_1, \alpha''_2, \alpha''_3$ также можно брать произвольными из множества G . Символом x обозначен произвольный элемент множества M . В частности, им может быть элемент множества M_1 , то есть физически реализуемое световое излучение. Эксперимент состоит в том, что испытуемый сравнивает цвета полей сравнения и устанавливает их совпадение или несовпадение.

В качестве значений функций $F_1(x), F_2(x), F_3(x)$, фигурирующих в законе трехмерности, следует брать коэффициенты $\alpha'_1 - \alpha''_1, \alpha'_2 - \alpha''_2, \alpha'_3 - \alpha''_3$. Закон трехмерности будет выполняться, если для каждого элемента x , принадлежащего множеству M , в эксперименте наблюдается равенство цветов полей сравнения при единственном наборе значений коэффициентов $F_1(x), F_2(x), F_3(x)$. Обширная практика колориметрических измерений [14] показывает, что закон трехмерности выполняется во всех без исключения случаях (для лиц с нормальным зрением, в патологических случаях этот закон выполняется в двумерной или одномерной формулировке) с той точностью, с которой испытуемый реализует предикат E . Коэффициенты $F_1(x), F_2(x), F_3(x)$ называются координатами цвета, соответствующего световому излучению x .

Итак, можно с полным основанием утверждать, что все условия, при которых вступает в силу теорема, выполняются применительно к зрительной системе человека. Это означает, что преобразование светового излучения в цвет, осуществляемое органом зрения чело-

века, может быть математически описано в форме линейного оператора, отображающего t -мерное пространство световых излучений, где t – число линий в спектре светового излучения, в трехмерное пространство цветов. Подобно тому, как это только что сделано для цветового зрения человека, с помощью теоремы можно провести структурную идентификацию любого физического, технического или социально-экономического объекта и в результате выяснить, можно ли его отнести к классу конечномерных линейных объектов. Если да, то этот объект можно будет математически описать, пользуясь изложенными ниже методами параметрической компараторной идентификации.

Выводы

В статье разработан способ компараторной идентификации систем сигналов, формально описываемых как конечномерные линейные пространства. С помощью этого метода математически описано поле зрения человека как часть двумерного линейного пространства и множество всех световых излучений – как часть многомерного линейного пространства.

Введено понятие линейного предиката, с помощью которого можно выполнить компараторную идентификацию любого линейного преобразователя сигналов многомерных векторных пространств. Сформулирована и доказана теорема о необходимых и достаточных признаках линейного предиката, при наличии которых обеспечивается линейность охватываемого им преобразователя сигналов. С помощью этой теоремы произведена идентификация преобразования светового излучения в зрительное ощущение, осуществляемое органом зрения человека, в виде линейного оператора, отображающего многомерное пространство световых излучений в трехмерное пространство цветов.

Список литературы: 1. Хедли Дж. Линейная алгебра. – М.: Наука, 1984. – 415 с. 2. Шилов Г.Е. Математический анализ (конечномерные линейные пространства). – М.: Наука, 1969. – 432 с. 3. Шикин Е.В. Линейные пространства и отображения. – М.: Изд-во МГУ, 1987. – 312 с. 4. Данфорд Н., Шварц Дж.Т. Линейные операторы, общая теория. – М.: ИЛ, 1974. – 895 с. 5. Мешков В.В. Основы светотехники. Ч. 2. – М.-Л.: Госэнергоиздат, 1961. – 416 с. 6. Фаддеев Д.К. Лекции по алгебре. – М.: Наука, 1984. – 415 с. 7. Гельфанд И.М. Лекции по линейной алгебре. – М.: Наука, 1971. – 271 с. 8. Глазман И.М., Любич Ю.И. Конечномерный линейный анализ. – М.: Наука, 1969. – 475 с. 9. Варден В.Д. Алгебра. – М.: Наука, 1979. – 623 с. 10. Бондаренко М.Ф., Шабанов-Кушнарченко С.Ю., Шабанов-Кушнарченко Ю.П. Об общей теории компараторной идентификации // Бионика интеллекта: науч.-техн. журнал. – 2008. – № 2. – С. 13–22. 11. Нюберг Н.Д. Грассмана законы // Физический энциклопедический словарь. Т. 1. – М.: Сов. энциклопедия, 1960. – С. 136. 12. Нюберг Н.Д. Курс цветоведения. – М.: Гизлегпром, 1932. – 140 с. 13. Гуревич М.М. Цвет и его измерение. – М.-Л.: Изд-во АН СССР, 1950. – 268 с. 14. Федоров Н.Т. Общее цветоведение. – М.: Гостехтеориздат, 1939. – 183 с.

Поступила в редакцию 19.01.2009

УДК 519.7

Ідентифікація об'єктів, що описуються векторами й лінійними операціями над ними / М.Ф. Бондаренко, С.Ю. Шабанов-Кушнарченко, Ю.П. Шабанов-Кушнарченко // Біоніка інтелекту: наук.-техн. журнал. – 2009. – № 1(70). – С. 3–10.

У статті розвивається теорія компараторної ідентифікації для випадку, коли об'єкт ідентифікації описується векторами арифметичного простору. Запропоновано модель компараторної ідентифікації, що дозволяє ідентифікувати лінійні об'єкти. Наведено змістовні інтерпретації математичних положень.

Л.: 1. Бібліогр.: 14 найм.

UDC 519.7

Identification of the objects described by vectors and linear operations over them / M.F. Bondarenko, S.Yu. Shabanov-Kushnarenko, Yu.P. Shabanov-Kushnarenko // Bionics of Intelligence: Sci. Mag. – 2009. – №1(70). – P. 3–10.

In article the comparator identifications theory for a case when the object of identification is described by vectors of arithmetic space develops. The identification comparator model which allows to identify linear objects is offered. Substantial interpretations of mathematical positions are resulted.

Fig.: 1. Ref.: 14 items.

УДК 519.7



РАЗРАБОТКА СИСТЕМ УСЛОВИЙ, ОБЕСПЕЧИВАЮЩИХ СУЩЕСТВОВАНИЕ ЛИНЕЙНОГО КОНЕЧНОМЕРНОГО ОПЕРАТОРА

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

^{1, 2, 3}ХНУРЭ, г. Харьков, Украина

В статье предлагается экспериментально проверяемая система аксиом, достаточная для проведения компараторной идентификации линейных конечномерных объектов. Рассмотрены варианты сокращения системы аксиом, при которых сохраняется возможность идентификации заданного объекта компараторным методом. Доказана несократимость упрощенных систем условий, которые определяют конечномерный линейный предикат. Рассмотрены возможности упрощения отдельных условий, характеризующих конечномерный линейный предикат.

КОМПАРАТОРНАЯ ИДЕНТИФИКАЦИЯ, ЛИНЕЙНЫЙ КОНЕЧНОМЕРНЫЙ ОПЕРАТОР, ГИЛЬБЕРТОВО ПРОСТРАНСТВО

Введение

Настоящая статья является продолжением работы [1]. В ней рассматривались вопросы идентификации объектов, которые можно описать векторами и линейными операциями над ними. Здесь анализируется другая важная проблема компараторной идентификации – формулировка и обоснование минимальных с точки зрения экспериментальной проверки систем условий (или аксиом), характеризующих модель идентифицируемого объекта.

1. Упрощение систем условий, характеризующих конечномерный линейный предикат

Оказывается, что для обеспечения линейности предиката E выполнение всех условий, указанных в теореме [1], не обязательно. Об этом свидетельствуют два нижеследующих утверждения.

Утверждение 1. Для того чтобы предикат E был линейным, необходимо и достаточно, чтобы он обладал свойствами симметричности, транзитивности, аддитивности, однородности и n -мерности.

Доказательство. Выведем рефлексивность предиката E из его симметричности, транзитивности и n -мерности. В силу n -мерности предиката E для любого x имеем:

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

К равенству (1) применяем свойство симметричности:

$$E(\sum_{i=1}^n F_i(x)e_i, x) = 1. \quad (2)$$

По свойству транзитивности предиката E из равенств (1) и (2) выводим $E(x, x) = 1$. Утверждение 1 доказано.

Утверждение 2. Для того чтобы предикат E был линейным, необходимо и достаточно, чтобы он обладал свойствами рефлексивности, аддитивности, однородности и n -мерности.

Доказательство. Выведем симметричность предиката E из его рефлексивности, аддитивности, одно-

родности и n -мерности. При доказательстве утверждения 5 из [2] было установлено, что из аддитивности и n -мерности предиката E вытекает аддитивность функций $F_i(x)$, где $i = \overline{1, n}$. Пусть x и y таковы, что $E(x, y) = 1$. В силу рефлексивности имеем $E(-y, -y) = 1$. К двум последним равенствам применяем свойство аддитивности. В результате получаем $E(x-y, 0) = 1$. Из свойства n -мерности выводим:

$$E(x-y, \sum_{i=1}^n F_i(x-y)e_i) = 1. \quad (3)$$

Значения $F_i(x-y) = 0$, где $i = \overline{1, n}$, обращают равенство (3) в тождество. В силу единственности этих значений и аддитивности функций F_i имеем:

$$F_i(x) - F_i(y) = 0, \quad F_i(y) - F_i(x) = 0, \quad F_i(y-x) = 0.$$

Отсюда следует

$$\sum_{i=1}^n F_i(y-x)e_i = \sum_{i=1}^n 0e_i = 0.$$

Таким образом, согласно свойству n -мерности

$$E(y-x, 0) = E(y-x, \sum_{i=1}^n F_i(y-x)e_i) = 1,$$

иными словами, $E(y-x, 0) = 1$. Поскольку предикат E рефлексивен, то $E(x, x) = 1$. Применяя к последним двум равенствам свойство аддитивности, получаем $E(y, x) = 1$.

Выводим транзитивность предиката E из его рефлексивности и аддитивности. Предположим, что векторы x, y, z таковы, что xEy и yEz . В силу аддитивности предиката E имеем $E(x+y, y+z) = 1$. Применяя к двум последним равенствам свойство аддитивности, получаем xEz . Утверждение 2 доказано.

Прикладное значение утверждений 1 и 2 состоит в том, что достигается уменьшение числа свойств, которые надо проверить опытным путем для выяснения вопроса о возможности идентификации заданного объекта компараторным методом. Найдено два варианта упрощения системы свойств. В первом отпадает необходимость проверки одного из шести свойств – рефлексивности, во втором – двух свойств: симметричности и транзитивности. Это приводит к упроще-

нию и удешевлению процедуры структурной идентификации линейных конечномерных объектов.

2. Несократимость упрощенных систем условий, определяющих конечномерный линейный предикат

Следующее утверждение свидетельствует о том, что при $n < m$ дальнейшее уменьшение числа указанных выше характеристических свойств линейного предиката невозможно.

Утверждение 3. При $n < m$ системы характеристических свойств линейного предиката E , фигурирующие в утверждениях 1 и 2, несократимы.

Доказательство. Для доказательства сформулированного утверждения достаточно установить независимость каждого свойства от совокупности остальных свойств этих двух систем. Независимость доказываем путем введения на декартовом квадрате m -мерного векторного пространства M над некоторым полем G такого предиката E , для которого данное свойство не выполняется, а остальные свойства выполняются. В роли векторного пространства M используем арифметическое пространство R^m . Любой вектор x в R^m имеет вид набора $x = (\xi_1, \xi_2, \dots, \xi_m)$, составленного из m вещественных чисел $\xi_1, \xi_2, \dots, \xi_m$, принимаемых в качестве координат вектора x . Обратно, любой набор, составленный из m вещественных чисел, является вектором арифметического пространства R^m . Коэффициентами служат произвольные вещественные числа. В роли поля G используется множество R^1 . В арифметическом пространстве в роли операций сложения и умножения скаляров используются сложение и умножение вещественных чисел. Арифметическое пространство R^m является разновидностью m -мерного векторного пространства M над полем G , для него выполняются все аксиомы векторного пространства. Будем считать, что в R^m введен базис (p_1, p_2, \dots, p_m) , где $p_1 = (1, 0, 0, \dots, 0)$, $p_2 = (0, 1, 0, \dots, 0)$, ..., $p_m = (0, 0, \dots, 0, 1)$. Оставшаяся часть доказательства состоит из пяти частей.

1. Докажем, что из рефлексивности, симметричности, транзитивности, аддитивности и однородности не следует n -мерность предиката E . Выберем предикат E следующим образом: $E(x, y) = D(x, y)$. Здесь D – предикат равенства, заданный на $R^m \times R^m$. Рефлексивность, симметричность, транзитивность, аддитивность и однородность предиката E очевидны. Однако n -мерностью предикат E не обладает. Для того, чтобы убедиться в этом, выберем произвольным образом векторы $e_1, e_2, \dots, e_m \in R^m$ и образуем их линейную оболочку L . Возьмем какой-нибудь вектор $x \in R^m$ вне L . Поскольку $n < m$, такой вектор всегда существует. Для любого вектора y , принадлежащего L , имеем $x \neq y$, следовательно $E(x, y) = D(x, y) = 0$. Вместе с тем, вектор y , как принадлежащий линейной оболочке L , выражается через e_1, e_2, \dots, e_n при подходящем наборе коэффициентов $(\alpha_1, \alpha_2, \dots, \alpha_n)$. Иными словами:

$$y = \sum_{i=1}^n \alpha_i e_i.$$

Таким образом,

$$E(x, \sum_{i=1}^n \alpha_i e_i) = 0.$$

Следовательно, не существует такого набора векторов (e_1, e_2, \dots, e_n) , при котором равенство

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

выполнялось для каждого $x \in R^m$ хотя бы при каком-нибудь наборе коэффициентов $(\alpha_1, \alpha_2, \dots, \alpha_n)$. А это как раз и означает, что предикат E не n -мерен.

2. Докажем, что из рефлексивности, симметричности, транзитивности, однородности и n -мерности не следует аддитивность предиката E . Выберем предикат E следующим образом:

$$E(x, y) = D((F_1(x), F_2(x), \dots, F_n(x)), (F_1(y), F_2(y), \dots, F_n(y))). \quad (4)$$

Здесь $F_i : R^m \rightarrow R^1$ $i = \overline{1, n}$ – это функции, определенные следующим образом:

$$F_i(x) = \begin{cases} \xi_i, & \text{если } \xi_{n+1} = 0, \\ 2\xi_i, & \text{если } \xi_{n+1} \neq 0. \end{cases} \quad (5)$$

Символом D обозначен предикат равенства, $R^n \times R^n$. Очевидно, предикат E рефлексивен, симметричен и транзитивен. Он также однороден. Действительно, пусть α выбрано произвольно. Если вектор x выбран так, что его координата $\xi_{n+1} = 0$, то согласно (5) $F_i(\alpha x) = \alpha \xi_i = \alpha F_i(x)$. Если же x выбран так, что для него $\xi_{n+1} \neq 0$, то $F_i(\alpha x) = 2\alpha \xi_i = \alpha 2\xi_i = \alpha F_i(x)$. Следовательно, функции $F_i(x)$ однородны. Из только что полученного результата и определения (4) предиката E выводим, что для любых x, y и α из условия $E(x, y) = 1$ следует $E(\alpha x, \alpha y) = 1$.

Докажем n -мерность предиката E . Имеем:

$$\sum_{i=1}^n F_k(x) p_k = F_1(x) p_1 + F_2(x) p_2 + \dots + F_n(x) p_n + 0 p_{n-1} + 0 p_m.$$

Записанный вектор представим в виде набора его координат:

$$\sum_{i=1}^n F_k(x) p_k = (F_1(x), F_2(x), \dots, F_n(x), 0, \dots, 0).$$

У этого вектора $n+1$ -я координата равна нулю. Таким образом, согласно (5) находим:

$$F_i(\sum_{k=1}^n F_k(x) p_k) = F_i(x)$$

для любого $i = \overline{1, n}$.

С помощью определения (4) предиката E из только что полученной системы равенств выводим:

$$E(x, \sum_{i=1}^n F_k(x) p_k) = 1. \quad (6)$$

Коэффициенты $F_k(x)$ при векторах p_k в (6) однозначно определяются выбором вектора x . Следовательно, существует набор векторов (p_1, p_2, \dots, p_n) такой, что равенство (6) выполняется для каждого x при единственном наборе коэффициентов $(F_1(x), F_2(x), \dots, F_n(x))$. Вместе с тем, предикат E не аддитивен. Действительно, пусть $x = (\xi_1, \xi_2, \dots, \xi_m)$, $y = (\eta_1, \eta_2, \dots, \eta_m)$, причем $\xi_i \neq 0$ ($i \in \{1, 2, \dots, n\}$). Тогда согласно (5) $F_k(x+y) = 2(\xi_i + \eta_i)$, $F_i(x) + F_i(y) = \xi_i + 2\eta_i$. Следовательно, для так выбранных x и y $F_i(x+y) \neq F_i(x) + F_i(y)$.

3. Докажем, что из рефлексивности, симметричности, транзитивности, аддитивности и n -мерности не следует однородность предиката E . Выберем предикат E следующим образом:

$$E(x, y) = D((F_1(x), F_2(x), \dots, F_n(x)), (F_1(y), F_2(y), \dots, F_n(y))). \quad (7)$$

Здесь функции F_i ($i = \overline{1, n}$) для любого $x = (\xi_1, \xi_2, \dots, \xi_m)$ принимают значения

$$F_i(x) = x_i + f(x_{n+1}). \quad (8)$$

В роли функции f используем какую-нибудь аддитивную, но не однородную вещественную функцию вещественного аргумента. Как известно [3], такие функции существуют.

Для вектора

$$\sum_{i=1}^n F_k(x) p_k$$

коэффициент F_i играет роль i -той его координаты, а $n+1$ -я координата этого вектора равна нулю. Следовательно, согласно (8), имеем:

$$F_i(\sum_{k=1}^n F_k(x) p_k) = F_i(x) + f(0).$$

Для любой аддитивной функции $f(0) = 0$, поэтому

$$F_i(\sum_{k=1}^n F_k(x) p_k) = F_i(x) + f(0).$$

По определению (7) предиката E последняя система равенств означает, что

$$E(x, \sum_{i=1}^n F_k(x) p_k) = 1.$$

Отсюда непосредственно следует, что предикат E n -мерен. Он, кроме того, рефлексивен, симметричен и транзитивен. Координаты ξ_i ($i = \overline{1, n}$) вектора x являются аддитивными функциями аргумента x . Поэтому функции $F_k(x)$ ($k = \overline{1, n}$), согласно определению (8), аддитивны. Вместе с ними аддитивен и предикат E , определяемый равенством (7). Однако, в силу неоднородности функции f , неоднородны также и функции $F_k(x)$. Следовательно, неоднороден и предикат E .

4. Докажем, что из транзитивности, аддитивности, однородности и n -мерности предиката E не вытекает его симметричность.

Выберем предикат E следующим образом:

$$E(x, y) = D((\xi_1, \xi_2, \dots, \xi_m), (\eta_1, \eta_2, \dots, \eta_m)). \quad (9)$$

Возьмем x и y такие, что $\xi_i = \eta_i$ ($i = \overline{1, n}$), $\xi_i \neq \eta_{n+1} = 1$. Тогда $E(x, y) = 1$, тем не менее $E(y, x) = 0$. Отсюда следует, что предикат E не симметричен. Вместе с тем, предикат E транзитивен. Действительно, пусть $E(x, y) = E(y, z) = 1$. Тогда $\xi_i = \eta_i = \zeta_i$ ($i = \overline{1, n}$) $\xi_{n+1} \neq \zeta_{n+1} = 0$. Следовательно, $E(x, z) = 1$.

Предикат E аддитивен. Действительно, пусть x, x', y, y' таковы, что $E(x, y) = E(x', y') = 1$. Тогда $\xi_i = \eta_i$, $\xi_i' = \eta_i'$, $\xi_i + \xi_i' = \eta_i + \eta_i'$ ($i = \overline{1, n}$), $\eta_{n+1} = \eta_{n+1}' = 0$. Следовательно, $E(x+x', y+y') = 1$. Для доказательства однородности предиката E произвольно выберем x и y так, чтобы $E(x, y) = 1$. Тогда $\xi_i = \eta_i$ ($i = \overline{1, n}$), $\xi_{n+1} = 0$. При любом a имеем $a\xi_i = a\eta_i$, $\eta_{n+1} = 0$. Следовательно, $E(ax, ay) = 1$.

Докажем, что предикат E n -мерен. Для любого x принимаем $F_i(x) = \xi$ ($i = \overline{1, n}$). Тогда

$$\sum_{i=1}^n F_i(x) p_i = (\xi_1, \xi_2, \dots, \xi_n, 0, \dots, 0).$$

Следовательно, согласно (9), имеем:

$$E(x, \sum_{i=1}^n F_i(x) p_i) = 1,$$

причем при любом другом наборе коэффициентов, стоящих при p_i , последнее равенство не выполняется.

5. Докажем, что из симметричности, аддитивности, однородности и n -мерности не следует транзитивность предиката E . Выберем предикат E следующим образом:

$$E(x, y) = D((\xi_1, \xi_2, \dots, \xi_n), (-\eta_1, -\eta_2, \dots, -\eta_n)). \quad (10)$$

Очевидно, что этот предикат симметричен, аддитивен и однороден. Он также n -мерен. Действительно, примем для любого x $F_i(x) = -\xi_i$ ($i = \overline{1, n}$). Тогда

$$E(x, \sum_{i=1}^n F_i(x) p_i) = 1,$$

причем ни при каком другом наборе коэффициентов при векторах p_i это равенство не выполняется. Вместе с тем, предикат E не транзитивен. В самом деле, пусть x, y, z таковы, что $E(x, y) = E(y, z) = 1$, причем $x \neq 0$. Тогда $\xi_i = -\eta_i$, $\eta_i = -\zeta_i$, $\xi_i = \zeta_i$ ($i = \overline{1, m}$). Согласно определению (10) предиката E имеем $E(x, z) = 0$. Утверждение 3 доказано.

Полученный результат доказывает, что из свойств рефлексивности, симметричности, транзитивности, однородности и n -мерности другие несократимые системы характеристических свойств линейного предиката, кроме тех, которые указаны в утверждениях 1 и 2, образовать невозможно. Таким образом, в определенном смысле мы достигли предела в упрощении экспериментально проверяемых свойств линейного конечномерного объекта, которыми он должен обладать для возможности его компараторной идентификации.

3. Упрощение систем условий, определяющих конечномерный линейный предикат при $n = m$

Утверждение 4. Для того чтобы при $n = m$ предикат E был линейным, необходимо и достаточно, чтобы он обладал свойствами рефлексивности и n -мерности.

Доказательство. При $n = m$ любые векторы x и y выражаются через линейно независимые векторы e_1, e_2, \dots, e_n :

$$x = \sum_{i=1}^n \alpha_i e_i, \quad y = \sum_{i=1}^n \beta_i e_i.$$

Предположим, что $E(x, y) = 1$. Тогда

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

В силу рефлексивности предиката E для любого x имеем $E(x, x) = 1$, иначе говоря:

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

Согласно свойству n -мерности $\alpha_i = \beta_i$ при любом $i = \overline{1, n}$. Это означает, что $x = y$. Очевидно также, что из $x = y$ вытекает $E(x, y) = 1$. Следовательно, E – это предикат равенства: $E(x, y) = D(x, y)$. Он обладает свойствами симметричности, транзитивности и однородности.

Утверждение 5. Для того чтобы при $n = m$ предикат E был линейным, необходимо и достаточно, чтобы он обладал свойствами симметричности, транзитивности и n -мерности.

Доказательство. Согласно доказанному в утверждении 1, рефлексивность вытекает из симметричности, транзитивности и n -мерности предиката E . По утверждению 4 устанавливаем линейность предиката E .

Утверждение 6. При $n = m$ системы характеристических свойств линейного предиката E , фигурирующие в утверждениях 4 и 5, несократимы.

Доказательство. Будем, как и в утверждении 3, в роли векторного пространства M использовать арифметическое пространство R^m ($m \geq 1$). Покажем, что рефлексивность предиката E не следует из его n -мерности. Пусть $E(x, y) = D(x, -y)$. Очевидно, что этот предикат n -мерен, но не рефлексивен. Покажем, далее, что n -мерность предиката E не следует из его рефлексивности. В роли предиката E принимаем предикат $E(x, y)$, равный единице при любых x и y . Предикат E в этом случае, очевидно, рефлексивен. Однако он не n -мерен, поскольку при любом наборе векторов (e_1, e_2, \dots, e_n) существует бесконечно много наборов вещественных чисел $(\alpha_1, \alpha_2, \dots, \alpha_n)$ таких, что

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

Докажем, что симметричность предиката E не следует из его транзитивности и n -мерности. Положим

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

Предикат E транзитивен: если xEy и yEz , то $z = 0$, следовательно xEz . Он также n -мерен. Действительно, при любом x xEy только при $y = 0$. Вместе с тем, вектор 0 выражается в любом базисе (e_1, e_2, \dots, e_n) по формуле

$$0 = \sum_{i=1}^n \alpha_i e_i$$

с помощью единственно возможного набора коэффициентов $\alpha_i = 0$, $i = \overline{1, n}$. Следовательно, равенство

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

для любого x выполняется при единственном наборе коэффициентов $(\alpha_1(x), \alpha_2(x), \dots, \alpha_n(x))$. Однако предикат E несимметричен, поскольку при $a \neq 0$ $E(a, 0) = 1$, но $E(0, a) = 0$.

Докажем, что транзитивность предиката E не следует из его симметричности и n -мерности. Положим $E(x, y) = D(x, -y)$. Очевидно, что такой предикат симметричен, n -мерен, но не транзитивен. Наконец, установим, что n -мерность предиката E не следует из его симметричности и транзитивности. Положим xEy для всех $x, y \in R^m$. Предикат E симметричен и транзитивен, но не n -мерен, так как существует бесконечно много наборов $\alpha_1(x), \alpha_2(x), \dots, \alpha_n(x)$ таких, что

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

где (e_1, e_2, \dots, e_n) – какой-нибудь базис. Утверждение 6 доказано.

Как видим, в специальном случае, когда размерность пространства входных сигналов идентифицируемого объекта совпадает с размерностью его выходных сигналов, системы характеристических свойств объекта существенно упрощаются. В первом варианте число условий сокращается с шести до двух, во втором – до трех. Тем самым еще более упрощается и удешевляется процедура структурной идентификации линейных конечномерных объектов. Важно подчеркнуть, что линейные конечномерные объекты, у которых входные и выходные сигналы имеют одинаковую размерность, весьма распространены в природе, обществе и технике.

4. Упрощение отдельных условий, характеризующих конечномерный линейный предикат

Можно ли утверждать, что системы характеристических свойств линейного конечномерного предиката теперь упрощены до предела? Нет, поскольку не были использованы возможности каждого из свойств, входящих в эти системы. О том, что такие возможности имеются, свидетельствует нижеприведенное утверждение.

Пусть E – предикат, заданный на $M \times M$, где M – какое-нибудь векторное пространство с базисом (p_1, p_2, \dots, p_m) над некоторым полем G . Предикат E назовем примитивно рефлексивным, если для любого $k = \overline{1, m}$ справедливо равенство $E(p_k, p_k) = 1$. Предикат E назовем примитивно n -мерным, если существует набор векторов (e_1, e_2, \dots, e_n) такой, что равенства

$$E(p_k, \sum_{i=1}^n H_i(p_k)e_i) = 1 \quad (11)$$

выполняются для каждого $k = \overline{1, m}$ при единственном наборе коэффициентов $(H_1(p_k), H_2(p_k), \dots, H_n(p_k))$. Здесь H_1, H_2, \dots, H_n – это некоторые функции, заданные на множестве $\{p_1, p_2, \dots, p_m\}$ со значениями в множестве G .

Утверждение 7. Для того чтобы предикат E был линейным, необходимо и достаточно, чтобы он обладал свойствами примитивной рефлексивности, аддитивности, однородности и примитивной n -мерности.

Доказательство. Для доказательства достаточно из свойств, перечисленных в условии утверждения, вывести рефлексивность и n -мерность предиката E , а затем сослаться на утверждение 2. Выведем рефлексивность. Возьмем произвольный вектор x и выразим его в виде:

$$x = x_1 p_1, x_2 p_2, \dots, x_m p_m.$$

Здесь x_1, x_2, \dots, x_m – координаты вектора x . Из ограниченной рефлексивности и однородности предиката E выводим

$$\begin{aligned} E(x_1 p_1, x_1 p_1) &= E(x_2 p_2, x_2 p_2) = \\ &= \dots = E(x_m p_m, x_m p_m) = 1. \end{aligned}$$

Применяя $m - 1$ раз к полученным равенствам свойство аддитивности предиката E , получаем:

$$E(x_1 p_1 + x_2 p_2 + \dots + x_m p_m, x_1 p_1 + x_2 p_2 + x_m p_m) = 1.$$

То есть $E(x, x) = 1$.

Выводим n -мерность. Рассмотрим уравнение

$$E(0, \sum_{i=1}^n \gamma_i e_i) = 1 \quad (12)$$

относительно неизвестного набора коэффициентов $(\gamma_1, \gamma_2, \dots, \gamma_n)$. Здесь (e_1, e_2, \dots, e_n) – фиксированный набор векторов, фигурирующий в условии примитивной n -мерности предиката E . Докажем, что уравнению (12) удовлетворяет единственный набор коэффициентов $(0, 0, \dots, 0)$. Действительно, пользуясь свойством аддитивности предиката E , из (11) и (12) выводим:

$$E(p_k, \sum_{i=1}^n (H_i(p_k) + \gamma_i) e_i) = 1, k = \overline{1, m}.$$

Сравнивая полученные равенства с равенствами (11), с учетом единственности коэффициентов при векторах e_i получаем по условию примитивной n -мерности: $H_i(p_k) = H_i(p_k) + \gamma_i$ ($i = \overline{1, n}$). Отсюда следует $\gamma_1 = \gamma_2 = \dots = \gamma_n = 0$.

Пусть x – произвольный вектор с координатами $\xi_1, \xi_2, \dots, \xi_m$. Обозначим

$$F_i(x) = \sum_{k=1}^m \xi_k H_i(p_k), i = \overline{1, n} \quad (13)$$

и докажем, что при таком выборе функций F_1, F_2, \dots, F_n имеет место равенство (14). Действительно, из однородности и ограниченной n -мерности предиката E выводим:

$$E(\xi_k p_k, \sum_{i=1}^n \xi_k H_i(p_k) e_i) = 1, k = \overline{1, m}. \quad (14)$$

Применение свойства аддитивности предиката E и только что полученной системе равенств дает:

$$E(\xi_k p_k, \sum_{k=1}^m \sum_{i=1}^n \xi_k H_i(p_k) e_i) = 1.$$

Иначе говоря:

$$E(x, \sum_{k=1}^m e_i \sum_{i=1}^n \xi_k H_i(p_k)) = 1.$$

Подставляя $F_i(x)$ по (13) в последнее равенство, приходим к равенству (14). Это означает, что набор коэффициентов $(F_1(x), F_2(x), \dots, F_n(x))$, фигурирующий в законе n -мерности, существует.

Осталось доказать единственность этого набора. Предположим, что имеются два набора $(\mu_1, \mu_2, \dots, \mu_n)$ и $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$, для которых выполняются равенства

$$E(x, \sum_{k=1}^n \mu_i e_i) = 1, E(x, \sum_{k=1}^n \varepsilon_i e_i) = 1.$$

Из предпоследнего равенства с помощью свойства однородности предиката E выводим

$$E(-x, \sum_{k=1}^n (-\mu_i e_i)) = 1.$$

Применяем к двум последним равенствам свойство аддитивности предиката E :

$$E(0, \sum_{k=1}^n (\varepsilon_i - \mu_i) e_i) = 1.$$

По доказанному ранее имеем $\varepsilon_i - \mu_i = 0$ при всех $i = \overline{1, n}$. Отсюда следует равенство наборов $(\mu_1, \mu_2, \dots, \mu_n)$ и $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$. Итак, набор коэффициентов $(F_1(x), F_2(x), \dots, F_n(x))$, фигурирующий в равенстве (14), существует и единственен, то есть свойство n -мерности предиката E выполняется. Утверждение 7 доказано.

Практическая ценность утверждения 7 состоит в том, что оно обосновывает упрощенные формулировки свойств идентифицируемого объекта – примитивной рефлексивности и примитивной n -мерности, которые требуют выполнения гораздо меньшего объема экспериментальной работы, чем исходные свойства рефлексивности и n -мерности. Исходные формулировки свойств рефлексивности и n -мерности необходимо проверять для всех входных сигналов, новые же формулировки свойств примитивной рефлексивности и примитивной n -мерности достаточно проверить всего лишь на m входных сигналах. За счет этого процедура структурной компараторной идентификации объекта еще более упрощается и удешевляется.

Выводы

Авторами достигнуто упрощение систем необходимых и достаточных экспериментально проверяемых условий, с помощью которых распознается линейность и конечномерность объекта, идентифицируемого компараторным методом. Доказана несократимость систем условий, определяющих конечномерный линейный объект при его компараторной идентификации.

Выполнено упрощение систем условий, определяющих конечномерный линейный объект для случая совпадения размерности его входных и выходных сигналов. Достигнуто упрощение формулировок отдельных свойств объекта, обеспечивающее существенное удешевление процедуры его структурной идентификации компараторным методом как линейного и конечномерного объекта.

Список литературы: 1. Бондаренко М.Ф., Шабанов-Кушнарченко С.Ю., Шабанов-Кушнарченко Ю.П. Идентификация объектов, описываемых векторами и линейными операциями над ними // Бионика интеллекта: науч.-техн. журнал.—2009.—№1(70).—С. 3–10. 2. Бондаренко М.Ф., Шабанов-Кушнарченко С.Ю., Шабанов-Кушнарченко Ю.П. Об общей теории компараторной идентификации // Бионика интеллекта: науч.-техн. журнал.—2008.—№2(69).—С. 13–22. 3. Гелбаум Б., Олмстед Дж. Контрпримеры в анализе.—М.: Мир, 1967.—251 с.

Поступила в редколлегию 21.01.2009.

УДК 519.7

Розробка систем умов, що забезпечують існування лінійно-скінченномірному оператору / М.Ф. Бондаренко, С.Ю. Шабанов-Кушнарченко, Ю.П. Шабанов-Кушнарченко // Біоніка інтелекту: наук.-техн. журнал.—2009.—№ 1.—С. 11–16.

Пропонується система аксіом, що може бути експериментально перевірена, достатня для проведення компараторної ідентифікації лінійних кінченномірних об'єктів. Розглянуто варіанти скорочення системи аксіом, при яких зберігається можливість ідентифікації заданого об'єкта компараторним методом.

Бібліогр.: 3 найм.

UDC 519.7

Working up of the conditions system providing the linear finite-dimensional operator existence / M.F. Bondarenko, S.Yu. Shabanov-Kushnarenko, Yu.P. Shabanov-Kushnarenko // Bionics of Intelligence: Sci. Mag. —2009. —№1(70).—P. 11–16.

In article is offered the experimentally checked axioms system, sufficient for carrying out to linear finite-dimensional objects comparator identification. Axioms system reduction variants at which remains the possibility of the object comparator identification are considered.

Ref.: 3 items.

УДК 519.7



РАЗРАБОТКА ТЕОРИИ СТРУКТУРНОЙ МЕТРИЗАЦИИ КОНЕЧНОМЕРНОГО АРИФМЕТИЧЕСКОГО ПРОСТРАНСТВА

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

^{1, 2, 3}ХНУРЭ, г. Харьков, Украина

В статье сформулирована и решена задача о компараторной идентификации метрики пространства выходных сигналов линейного конечномерного объекта. Установлены необходимые и достаточные экспериментально проверяемые условия, при которых возможна компараторная идентификация этой метрики. Показано практическое значение этой задачи для математического описания технических и природных объектов.

КОМПАРАТОРНАЯ ИДЕНТИФИКАЦИЯ, ЦВЕТОВОЕ ЗРЕНИЕ, ГОМЕОМОРФИЗМ, МЕТРИКА

Введение

Настоящая статья является продолжением работ [1, 2]. В [1] проанализировано применение теории компараторной идентификации к цветовому зрению человека и социально-экономическому поведению людей. В [2] предложена модель компараторной идентификации, позволяющая идентифицировать линейные объекты. В данной статье обоснованы необходимые и достаточные экспериментально проверяемые условия, при которых возможна компараторная идентификация пространства выходных сигналов линейного конечномерного объекта. Разработан практический способ построения этой метрики.

1. Введение метризирующего предиката

Пусть N и N' – n -мерные арифметические пространства, связанные гомеоморфизмом $\varphi: N \rightarrow N'$, который взаимно однозначно и взаимно непрерывно отображает пространство N на пространство N' . Точка $x' \in N'$ называется образом точки $x \in N$, если $x' = \varphi(x)$, а пространство N' – образом пространства N . Пусть, кроме того, $\rho(x', y')$ – евклидово расстояние между точками x', y' пространства N' , определяемое равенством:

$$\rho(x', y') = \sqrt{(\xi'_1 - \eta'_1)^2 + (\xi'_2 - \eta'_2)^2 + \dots + (\xi'_n - \eta'_n)^2}, \quad (1)$$

где $\xi'_1, \xi'_2, \dots, \xi'_n$ и $\eta'_1, \eta'_2, \dots, \eta'_n$ – координаты точек x', y' в пространстве N' . Рассмотрим предикат Φ на N^4 , определяемый выражением:

$$\Phi(x_1, y_1, x_2, y_2) = D(\rho(\varphi(x_1), \varphi(y_1)), \rho(\varphi(x_2), \varphi(y_2))). \quad (2)$$

Здесь символ D обозначает предикат равенства, заданный на декартовом квадрате вещественной полуоси $[0, \infty)$, рассматриваемой как множество всех расстояний между точками пространства N' . Предикат вида (1) устанавливает, равны или нет расстояния между образами точек x_1, y_1 и x_2, y_2 . Будем говорить, что предикат Φ метризует пространство N' .

Проинтерпретируем сказанное на примере задачи о введении поля зрения человека, о которой говорилось в [2]. Выделим в физическом пространстве перед испытуемым какую-нибудь плоскость, перпендикулярную зрительной оси того глаза испытуемого, поле зрения которого исследуется, и проведем на ней через точку фиксации горизонтальную и вертикальную прямые, приняв их за координатные оси двумерного пространства N . Такая плоскость называется кампиметрической [3]. На кампиметрической плоскости исследователь произвольно выбирает две пары точек x_1, y_1 и x_2, y_2 (для этой цели следует брать точечные источники света в темноте) и предлагает испытуемому определить, одинаковы или нет расстояния $\rho(x_1, y_1)$ и $\rho(x_2, y_2)$ между субъективными образами $x'_1 = \varphi(x_1)$, $y'_1 = \varphi(y_1)$, $x'_2 = \varphi(x_2)$, $y'_2 = \varphi(y_2)$ точек x_1, y_1 и x_2, y_2 . Символом φ обозначено преобразование кампиметрической плоскости N в ее субъективный образ N' (называемый полем зрения глаза), осуществляемое зрительной системой человека. Если предъявить точки при наличии освещения, то испытуемый невольно будет оценивать расстояние между точками кампиметрической плоскости, а не между их субъективными образами.

Опыты показывают, что испытуемый способен решать предложенную задачу при предъявлении ему любых точек x_1, y_1 и x_2, y_2 кампиметрической плоскости с довольно высокой точностью. Это следует из того факта, что изменение расстояний между точками кампиметрической плоскости всего лишь на несколько процентов способно нарушить равенство расстояний между образами этих точек. Таким образом, в данном эксперименте испытуемый своим поведением реализует с весьма высокой точностью предикат Φ . Отображение φ не является тождественным, поэтому образы точек нельзя отождествлять с самими точками кампиметрической плоскости. Это доказывается тем, что образы точек x_1, y_1 и x_2, y_2 , лежащих на кампиметрической плоскости и расположенных на равных расстояниях друг от друга, часто субъективно воспринимаются как неравноудаленные. Вместе с тем наблюда-

ются случаи, когда образы точек в поле зрения находятся друг от друга на равных расстояниях, а соответствующие им точки на кампиметрической плоскости лежат на разных расстояниях. Субъективные наблюдения испытуемого свидетельствуют также и о том, что непрерывному перемещению точки на кампиметрической плоскости соответствует непрерывное перемещение ее образа в поле зрения и что каждой точке кампиметрической плоскости соответствует свой образ — точка в поле зрения. Следовательно, судя по ощущениям испытуемого, отображение φ в данной интерпретации можно считать гомеоморфизмом.

Сделанные выводы хотелось бы освободить от элемента субъективизма и обосновать чисто физическими экспериментами, используя метод компараторной идентификации. Представляется, что для решения этой задачи достаточно располагать лишь предикатом Φ и его свойствами, не опираясь ни на какие данные субъективного характера. Основываясь на предикате Φ и его свойствах, требуется доказать, что поле зрения представляет собой двумерное арифметическое пространство N' , получаемое деформацией φ (то есть взаимно однозначным и взаимно непрерывным преобразованием) кампиметрической плоскости N . Кроме того, требуется найти конкретный вид гомеоморфизма φ . Первая задача является задачей структурной метризации поля зрения, вторая — задачей параметрической метризации поля зрения.

В формальной постановке, независимой от какой бы то ни было интерпретации, задача структурной метризации пространства N' формулируется следующим образом. Дано n -мерное арифметическое пространство N с определенными на нем операциями сложения векторов и умножения вещественного числа на вектор. На N^4 определен предикат $\Phi(x_1, y_1, x_2, y_2)$. Требуется сформулировать такую систему A свойств предиката Φ , при выполнении которой предикат Φ можно было бы представить в виде (2). Вместе с тем, если хотя бы одно из свойств системы A для предиката Φ не выполняется, то его нельзя будет представить в виде (2). Задача параметрической метризации пространства N' сводится к отысканию конкретного вида функции φ , извлекаемого из значений предиката Φ , получаемых в опытах на испытуемом. Если бы задачу удалось решить, тогда можно было бы математически описать механизм формирования поля зрения человека. Имея такое описание, можно было бы применить его для целей совершенствования «зрения» роботов, узнающих автоматов и других технических систем.

Задачей идентификации поля зрения человека далеко не исчерпывается область возможного применения теории метризации конечномерного арифметического пространства. Другой важной содержательной интерпретацией этой теории является проблема введения естественной метрики в пространстве цветов.

Назовем ее цветовой интерпретацией задачи метризации. Содержательную интерпретацию задачи метризации, рассмотренную первой, назовем пространственной. В случае цветовой интерпретации множество n понимается как объективное цветовое пространство, в котором каждый цвет представлен набором колориметрических координат. Сумма цветов в цветовом пространстве N определяется суммой световых излучений, породивших данные цвета. Произведение числа на цвет в пространстве N определяется произведением этого числа на световое излучение, соответствующее данному цвету. Гомеоморфизм φ понимается как такое искривление цветового пространства N , в результате которого получается субъективное цветовое пространство N' , называемое еще иначе равноконтрастным цветовым пространством [4]. Равноконтрастность цветового пространства означает, что любые пары цветов, находящиеся в пространстве N' на одном и том же расстоянии друг от друга, выглядят равноудаленными также и в психофизическом смысле. Если б удалось построить равноконтрастное цветовое пространство, то это было бы равносильно отысканию наиболее экономного способа кодирования информации о цвете воспринимаемых объектов, что дало бы существенную экономию объема машинной памяти, требуемой для запоминания цветных изображений.

Еще одна интерпретация задачи метризации пространства естественно возникает при автоматическом управлении объектами. Назовем ее технической. Пусть некоторый объект преобразует входные сигналы из m -мерного векторного пространства M в выходные сигналы арифметического пространства N меньшей или той же размерности n . Выходной сигнал представлен набором числовых параметров, по которым осуществляется управление объектом. Управление ведется по расстоянию ρ между текущим выходным сигналом и некоторым эталонным набором чисел, которые могут меняться во времени. Цель управления объектом состоит в том, чтобы, меняя его параметры, постоянно держать выходной сигнал объекта достаточно близким к эталонному. При этом важно, чтобы фактическая точность такого приближения находилась в заранее заданных пределах. Как показывает практика управления объектами, заданная точность сравнения текущего сигнала с эталонным обычно не остается постоянной и меняется вместе с изменением эталонного сигнала. Например, текущий сигнал требуется сравнивать с малым эталонным сигналом обычно с меньшей ошибкой, чем при высоком уровне эталонного сигнала.

Трудно рассчитывать на то, чтобы естественным образом формируемое объектом управления пространство выходных сигналов всегда само собой удовлетворяло указанному выше требованию. Поэтому пространство выходных сигналов обычно нуждается в некотором «исправлении». Стандартный прием такого ис-

правления состоит в том, что пространство выходных сигналов деформируют, причем с таким расчетом, чтобы равным геометрическим расстояниям между точками пространства после его деформации всегда соответствовала одинаковая их удаленность друг от друга в некотором содержательном смысле, диктуемом соображениями, направленными на достижение максимальной эффективности процесса управления объектом. Для этого придется растянуть те области пространства, внутри которых точность сравнения фактического выходного сигнала объекта с эталонным недостаточна, и сжать те области, где эта точность избыточна.

2. Формулировка и обоснование свойств метризирующего предиката

Ниже формулируются некоторые свойства метризирующего предиката Φ , которые в дальнейшем будут использованы при построении аксиоматической теории компараторной идентификации метризирующего отображения. Очевидно, что любой предикат Φ вида (2) рефлексивен, симметричен и транзитивен относительно пар точек, для которых определяется расстояние ρ между их образами по формуле (1). Таким образом, приходим к следующим трем свойствам, которым подчиняется предикат Φ :

– закону парной рефлексивности

$$\forall x, y \in N \quad \Phi(x, y, x, y); \quad (3)$$

– закону парной симметричности

$$\forall x_1, y_1, x_2, y_2 \in N \quad \Phi(x_1, y_1, x_2, y_2) \supset \Phi(x_2, y_2, x_1, y_1); \quad (4)$$

– закону парной транзитивности

$$\forall x_1, y_1, x_2, y_2, x_3, y_3 \in N \quad \Phi(x_1, y_1, x_2, y_2) \wedge \Phi(x_2, y_2, x_3, y_3) \supset \Phi(x_1, y_1, x_3, y_3). \quad (5)$$

Евклидово расстояние ρ обладает симметрией [5], это означает, что $\rho(x', y') = \rho(y', x')$ для любых $x', y' \in N'$. Поэтому для метризирующего предиката справедлив закон одиночной симметричности

$$\forall x_1, y_1, x_2, y_2 \in N \quad \Phi(x_1, y_1, x_2, y_2) \supset \Phi(y_2, x_2, x_1, y_1); \quad (4)$$

Евклидово расстояние ρ удовлетворяет также аксиоме тождества [6]. Это означает, что расстояние $\rho(x', y')$ между любыми совпадающими точками $x' = y'$ пространства N' должно равняться нулю. Кроме того, если x' и y' таковы, что $\rho(x', y') = 0$, то всегда $x' = y'$. Отсюда непосредственно вытекают следующие свойства метризирующего предиката: первый закон тождества

$$\forall x, y \in N \quad \Phi(x, x, y, y) \quad (7)$$

и второй закон тождества

$$\forall x, x_1, y_1 \in N \quad (\Phi(x, y, x_1, y_1) \supset x_1 = y_1). \quad (8)$$

Рассмотрим предикат R на N^3 , значения которого определяются через значения предиката Φ при любых $x, y, z \in N$ следующим образом:

$$R(x, y, z) = \Phi(x, z, z, y) \wedge \forall t \in N (\Phi(x, t, t, y) \wedge \Phi(z, x, x, t) \supset z = t). \quad (9)$$

Равенство $R(x, y, z) = 1$ означает, что точка $\varphi(z)$ лежит посередине отрезка прямой, соединяющего точки $\varphi(x)$ и $\varphi(y)$ в пространстве N' . В самом деле, из условия $R(x, y, z) = 1$ согласно (8) следует, что:

- а) $\Phi(x, z, z, y) = 1$,
- б) $\forall t \in N (\Phi(x, t, t, y) \wedge \Phi(z, x, x, t) \supset z = t) = 1$.

Утверждение а) означает, что точка $\varphi(z)$ равноудалена от точек $\varphi(x)$ и $\varphi(y)$. Утверждение б), взятое вместе с утверждением а), означает, что если какая-то точка $\varphi(t)$ равноудалена от точек $\varphi(x)$ и $\varphi(y)$ на такое расстояние, что и точка $\varphi(z)$ от точек $\varphi(x)$ и $\varphi(y)$, то точка $\varphi(t)$ всегда совпадает с точкой $\varphi(z)$. Таким образом, точка $\varphi(z)$ делит пополам отрезок прямой, соединяющий точки $\varphi(x)$ и $\varphi(y)$. Сказанное иллюстрируется для случая двумерного пространства диаграммой, изображенной на рис 1.

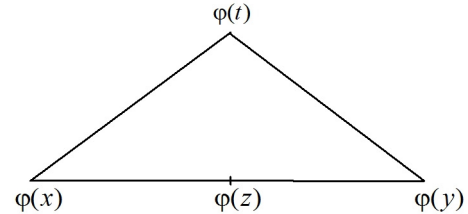


Рис. 1

Пусть $\varphi(t)$ – точка, лежащая на расстоянии ρ от точек $\varphi(x)$ и $\varphi(y)$. Такая точка единственна, если она лежит на середине отрезка прямой, соединяющего точки $\varphi(x)$ и $\varphi(y)$. На диаграмме эта единственная точка обозначена символом z . Если же точка $\varphi(t)$ не совпадает с точкой $\varphi(z)$, то всегда найдется еще одна точка $\varphi(t_1) \neq \varphi(t)$, которая так же, как и точка $\varphi(t)$, лежит на расстоянии ρ от точек $\varphi(x)$ и $\varphi(y)$.

Равенство же $R(x, y, z) = 0$ означает, что точка $\varphi(z)$ не лежит на середине отрезка прямой, соединяющего точки $\varphi(x)$ и $\varphi(y)$. В самом деле, если $R(x, y, z) = 0$, то условие а) или условие б) не выполняются. В первом случае точка $\varphi(z)$ не лежит посередине отрезка прямой, соединяющего точки $\varphi(x)$ и $\varphi(y)$, во втором случае точка $\varphi(z)$ не лежит на этом отрезке. Таким образом, отношение, соответствующее предикату $R(x, y, z)$, задает операцию отыскания средней точки $\varphi(z)$ между точками $\varphi(x)$ и $\varphi(y)$ пространства N' . Эту операцию назовем внутренним равноделением точек x и y . Для нее, очевидно, выполняются свойства всюду определенности

$$\forall x, y \in N \quad \exists z \in N \quad R(x, y, z) \quad (10)$$

и однозначности

$$\forall x, y, z, z_1 \in N \quad (R(x, y, z) \wedge R(x, y, z_1) \supset z = z_1). \quad (11)$$

Операцию внутреннего равноделения точек x и y , которая ставит им в соответствие точку z , обозначим

кружком: $z = x \circ y$. Она определена на N^2 со значениями в множестве N . Из определения предиката Φ с очевидностью следует, что в арифметическом пространстве N для любых двух точек x и y всегда найдется единственная точка z такая, что точка y будет результатом внутреннего равноделения интервала между точками x и z . Эту операцию назовем внешним равноделением точек x и y , обозначая ее звездочкой: $z = x^*y$. Последнее равенство равносильно равенству $x^*z = y$. Внешнее равноделение обладает свойством всюду определенности

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

и свойством однозначности

$$\forall x, y, z \in N (R(x, z, y) \wedge R(x, z_1, y) \supset z = z_1). \quad (13)$$

Очевидно, что операция внутреннего равноделения обладает свойствами коммутативности

$$\forall x, y \in N (x \circ y = y \circ x) \quad (14)$$

и идемпотентности

$$\forall x, y \in N (x \circ x = x). \quad (15)$$

Заметим, что свойство (14) логически следует из свойств (3), (6) и (8). Действительно, согласно определению (9) предиката R имеем:

$$R(x, x, x) = \Phi(x, x, x, x) \wedge \forall t \in N$$

$$\Phi(x, t, t, x) \wedge \Phi(x, x, x, t) \supset x = t.$$

По закону парной рефлексивности (3) находим: $\Phi(x, x, x, x) = 1$. По законам парной симметричности (6) имеем: $\Phi(x, t, t, x) = 1$. По второму закону тождества (8) из $\Phi(x, x, x, t)$ следует $x = t$. Поэтому $R(x, x, x) = 1 \forall t \in N \ 1 \wedge \Phi(x, x, x, t) \supset x = t = 1$, а значит, $x \circ x = x$.

Из зависимостей (1) и (2), определяющих метризирующий предикат Φ , следует, что непрерывное изменение положения точек x и y в пространстве N влечет непрерывное изменение положение точек $x \circ y$ и x^*y , являющихся результатом их внутреннего и внешнего равноделения. Соответственно этому имеет место свойство непрерывности операций $x \circ y$ и x^*y :

$$\text{Функции } x \circ y \text{ и } x^*y \text{ непрерывны по совокупности переменных } x \text{ и } y. \quad (16)$$

Имеется в виду непрерывность, индуцируемая евклидовой метрикой в пространстве N .

В любом четырехугольнике n -мерного арифметического пространства N' отрезки прямых, соединяющие середины противоположных сторон, пересекаются в точке 0 , которая делит их пополам [7]. Это свойство иллюстрируется в двумерном случае диаграммой, изображенной на рис. 2. Его истинность вытекает из тождества

$$\frac{\frac{\varphi(x)+\varphi(y)}{2} + \frac{\varphi(z)+\varphi(t)}{2}}{2} = \frac{\frac{\varphi(t)+\varphi(x)}{2} + \frac{\varphi(y)+\varphi(z)}{2}}{2}.$$

Отсюда следует свойство четырехугольника:

$$\forall x, y, z, t \in N ((x \circ y)(z \circ t) = (t \circ x)(y \circ z)). \quad (17)$$

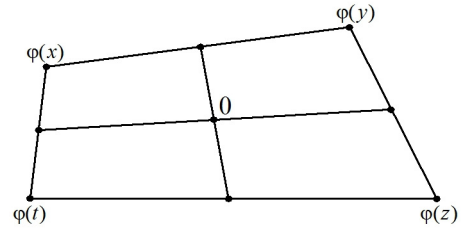


Рис. 2

Обратим внимание на то обстоятельство, что выражения (9)–(17) задают свойства предиката Φ , несмотря на то, что имя этого предиката в них не фигурирует. Это вытекает из того, что предикат R выражается зависимостью (1) через предикат Φ , а операция \circ определяется предикатом R . Таким образом, выражения (9)–(17) представляют собой сокращенную запись свойств предиката Φ . Все эти выражения можно при желании записать в виде высказываний, зависящих только от предиката Φ , если заменить в них предикат R и операцию \circ через предикат Φ .

3. Содержательные интерпретации свойств метризирующего предиката

Продолжим изучение пространственной, цветовой и технической интерпретаций метризирующего предиката. Важность рассмотрения пространственной интерпретации определяется, во-первых, тем, что она проще цветовой и технической и поэтому может служить как бы введением к описанию двух последних интерпретаций и, во-вторых, тем, что она сама по себе интересна как важная область практического применения разрабатываемых авторами методов компараторной идентификации объектов.

Когда мы говорим о пространственной интерпретации, то имеем в виду процесс зрительного восприятия испытуемым окружающего его трехмерного физического пространства. Глаз человека отображает трехмерное физическое пространство M на двумерное поле зрения N , причем только те точки исходного пространства, которые лежат на одном луче, исходящем из центра глаза, проектируются в одну и ту же точку поля зрения (рис. 3).

Процесс перехода от точки X трехмерного физического пространства, описываемой в координатной системе $\Xi_1\Xi_2\Xi_3$, к точке двумерного поля зрения x , описываемой в системе $\xi_1\xi_{21}$, может быть идентифицирован компараторным методом, который устанавливает факт совпадения или различия точек x и y поля зрения, являющихся образами $x = F(X)$, $y = F(Y)$ точек X и Y физического пространства.

Человек обладает способностью субъективно оценивать расстояние между точками поля зрения [8]. Если предъявить испытуемому две пары точек поля зрения (x_1, y_1) и (x_2, y_2) , то он с достаточно высокой точностью сможет определить, равны ли расстояния меж-

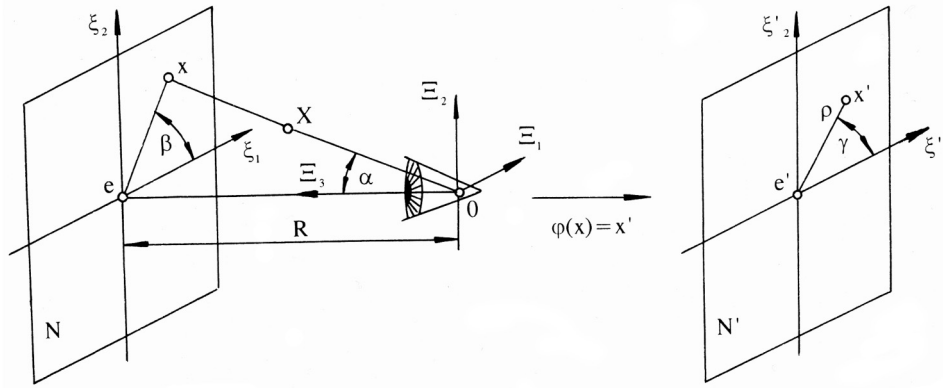


Рис. 3

ду точками этих пар или нет. Тем самым испытуемый фактически реализует метрический предикат $\Phi(x_1, y_1, x_2, y_2)$, о котором шла речь в п. 6. Факты свидетельствуют, что далеко не всегда расстояния между точками поля зрения, равные в физическом смысле, будут равными также и в психофизическом смысле. Это означает, что исходное поле зрения (назовем его объективным) в процессе восприятия искривляется, следовательно, может быть поставлен вопрос о субъективном поле зрения N' и о существовании гомеоморфизма $\varphi(x) = x'$, отображающего объективное поле зрения в субъективное.

Следующий опыт доказывает наличие искривления поля зрения. В ясную ночь выберем для наблюдения на небе какие-нибудь далеко отстоящие друг от друга три яркие звезды, лежащие на дуге окружности с центром в точке наблюдения так, чтобы средняя звезда делила пополам интервал между крайними (в определенные моменты времени в таком положении оказываются, например, планеты Венера, Марс и Юпитер). Если зафиксировать взгляд на средней звезде y , то остальные две звезды x и z будут восприниматься как равноудаленные от средней звезды и лежащие на одной прямой с ней (рис. 4а). Но если точку фиксации e сместить куда-нибудь в сторону (например, вправо и вниз), то крайние звезды теперь будут видаться на разном удалении от средней, а прямая линия превратится в ломаную, два отрезка которой образуют тупой угол (рис. 4б).

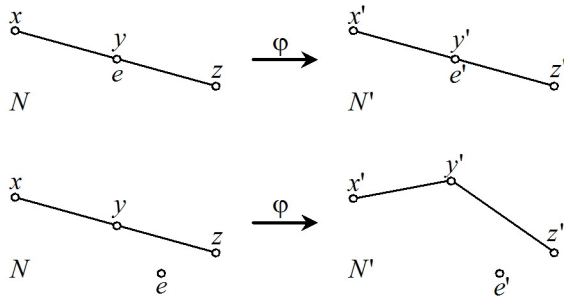


Рис. 4а

Рис. 4б

Специально проведенные нами опыты по сравнению расстояний между точками, зрительно воспринимаемыми испытуемым, свидетельствуют о том, что все перечисленные в предыдущем параграфе свойства метризирующего предиката в случае его пространственной интерпретации выполняются с той точностью, которую обеспечивает точность ответа испытуемого. Отклонение оси симметрии пороговой кривой при проверке, например, свойства парной рефлексивности не превышает $\pm 0,5\%$ [9]. Свойство парной рефлексивности в пространственной интерпретации означает, что при двукратном предъявлении одной и той же пары точек испытуемый будет устанавливать равенство расстояний. Свойство парной симметричности означает, что при перемене порядка предъявления пар точек испытуемому его реакция не изменится. Свойство парной транзитивности означает, что если при сравнении субъективных расстояний между точками первой и второй пары, а также второй и третьей пары испытуемый устанавливает их равенство, то он установит равенство также и при сравнении субъективных расстояний между точками первой и третьей пар.

Свойство одиночной симметричности означает, что после перестановки точек внутри отдельной пары результат сравнения расстояний не меняется. Первое свойство тождества означает, что при совпадении точек в обеих парах субъективные расстояния всегда будут одинаковыми. Второе свойство тождества означает, что если в одной паре точки совпадают, то при равенстве расстояний они будут совпадать и во второй паре. Существо операции внутреннего равноделения состоит в том, что к двум заданным точкам подыскивается такая третья, чтобы, во-первых, расстояния между первой и третьей, а также третьей и второй точками совпали, и, во-вторых, точка, удовлетворяющая первому условию при таком расстоянии, оказалась единственной. Третья точка субъективно воспринимается испытуемым как лежащая посередине отрезка прямой, соединяющего первую и вторую точки. Существо операции внешнего равноделения состоит в

том, что к первой и второй точкам подыскивается такая третья, чтобы отрезок прямой, соединяющий первую и третью точки, делился второй точкой пополам.

Закон внутреннего равноделения означает, что для любых двух точек поля зрения субъективно средняя точка между ними всегда существует и единственна. Исключением является случай, когда средняя точка попадает в зону слепого пятна [10], тогда она не существует. Закон внешнего равноделения означает, что для любых двух точек результат внешнего равноделения всегда существует и единственен. Этот закон также выполняется, но с одной оговоркой: так как поле зрения человека ограничено [11], то точка внешнего равноделения выходит за пределы поля зрения при некоторых исходных данных и поэтому не существует. Также остается в силе и замечание относительно зоны слепого пятна. Вследствие этого теорию компараторной идентификации метризирующего отображения приходится применять в случае ее пространственной интерпретации на суженной основе со всеми вытекающими отсюда предосторожностями [9]. Закон непрерывности означает, что при непрерывном изменении положения исходных точек будет также непрерывно меняться положение третьей точки, получающейся в результате выполнения операции внутреннего или внешнего равноделения. Смысл закона четырехугольника ясен из рассмотрения рис. 2. Этот закон применим во всех случаях, когда предъявляемые и получаемые в результате выполнения операций точки (всего 9 точек) лежат в пределах поля зрения испытуемого. В противном случае закон четырехугольника утрачивает смысл.

Перейдем теперь к рассмотрению цветовой интерпретации. По сравнению с пространственной цветовой интерпретацией имеет некоторые особенности. В то время, как расстояния между далекими и близкими точками в поле зрения определяются испытуемым с достаточно высокой и примерно одинаковой относительной точностью, с цветами дело обстоит иначе. Расстояния между близкими цветами сравниваются испытуемым на тождество и различие довольно точно. Когда же испытуемый устанавливает совпадение или различие расстояний между далекими цветами, то относительная точность сравнения резко падает, и он испытывает большую неуверенность в своих ответах. Поэтому в случае цветовой интерпретации можно говорить о достаточно точной экспериментальной проверке аксиом теории компараторной идентификации метризирующего отображения лишь в локальном смысле. В связи с этим в теории цветового зрения практикуется измерять расстояние числом пороговых значений (или кратной ему величиной), разделяющих данные цвета [12]. Следует обратить внимание на некоторую спорность такого измерения, поскольку понятия порога и расстояния теоретически между собой никак не связаны, их природа совершенно различна. Поэтому они не обязаны строго соответствовать друг другу.

Сказанное подтверждается исследованиями, выполненными для точек поля зрения человека. Пространственный порог различения, определяющий остроту зрения и измеряемый в угловых минутах, для различных участков поля зрения достаточно хорошо изучен [13]. При движении от центра (точки фиксации) к краям поля зрения острота зрения очень быстро снижается. Субъективные расстояния также сокращаются, но не так быстро, поэтому точного соответствия между порогами и расстояниями для пространственной интерпретации нет, но и попытка мерять субъективные расстояния между точками поля зрения числом порогов, разделяющих их, терпит фиаско. Это, однако, не означает, что таким же будет положение и для цветовой интерпретации. Быть может, измерение цветовых расстояний числом порогов правомерно, но это надо доказать специальным экспериментальным исследованием, демонстрирующим наличие корреляции между цветовыми расстояниями и порогами. Судя по литературным источникам, вопрос о проведении таких исследований до сих пор не ставился.

Что же касается технической интерпретации метризирующего предиката и его свойств, то здесь определяющим фактором является точность измерения сигналов пространства N . Именно она должна быть выравнена во всех областях пространства N' после перехода к нему с помощью гомеоморфизма φ от пространства N . Поэтому при экспериментальной проверке свойств метризирующего предиката в случае его технической интерпретации следует измерять расстояние между точками пространства N' не иначе, как числом порогов различения.

Возвращаясь от технической интерпретации метризирующего предиката к его цветовой и пространственной интерпретациям, можно прийти к выводу, что и там изучение порогового варианта метрики представляет определенный интерес. Дело в том, что реализуя модель цветового зрения или модель пространственного видения предметов человеком в виде технических устройств, последние можно будет использовать в системах автоматического управления (например, в системах слежения за движущимися объектами). В этом случае придется решать проблему выравнивания пространства выходных сигналов управляемого объекта с целью достижения рациональной точности их измерения [14–16], и будет небезынтересно узнать, как этот же вопрос решен в зрительной системе человека, чтобы дать возможность инженеру воспользоваться подсказкой природы.

При этом важно иметь в виду, что метрики одного и того же пространства выходных сигналов объекта, получаемые на базе изучения порогов и на базе изучения расстояний, вообще говоря, различны. Изучение этих метрик представляет собой две независимые друг от друга задачи. Сравнивая результаты раздельного

решения этих задач, можно будет узнать, насколько близко друг к другу получаемые гомеоморфизмы. В частном случае эти гомеоморфизмы могут даже совпадать, однако ожидать такого совпадения во всех случаях нет никаких оснований.

4. Введение структуры линейного пространства, согласованной с операциями внутреннего и внешнего равноделения точек

Выше было введено понятие метризирующего предиката и сформулированы некоторые из его характеристических свойств. Используя эти и некоторые другие свойства как исходные положения, поставим теперь задачу получить аксиоматическое, то есть абстрактное определение метризирующего предиката. Для решения этой задачи сначала введем на множестве N с помощью операций внутреннего и внешнего равноделения точек структуру n -мерного векторного пространства. Она может не совпадать с исходной структурой n -мерного арифметического пространства N . Множество N с введенной на нем новой структурой векторного пространства будем обозначать символом N^* .

Выберем в качестве нулевого вектора пространства N^* какой-нибудь элемент множества N . Обозначим его символом 0 . Операцию сложения векторов x, y пространства N^* определяем следующим образом:

$$x + y = 0 * (x \circ y). \quad (18)$$

С содержательной точки зрения (то есть исходя из первоначального конкретного определения метризирующего предиката) вектор $\varphi(x + y)$ в пространстве N' образуется из векторов $\varphi(x)$ и $\varphi(y)$ по правилу параллелограмма (рис. 5), в котором, как известно, диагонали делятся пополам точкой их пересечения.

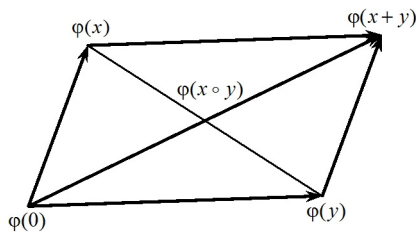


Рис. 5

Точка $\varphi(0)$ играет в пространстве N' роль нулевого вектора. Вектор $\varphi(x + y)$ представляет собой сумму векторов $\varphi(x)$ и $\varphi(y)$ пространства N' .

Множество N вместе с заданной на нем операцией сложения (18) образует абелеву группу [17]. Действительно, коммутативность сложения

$$x + y = y + x \quad (19)$$

вытекает из определения (18) и коммутативности операции внутреннего равноделения \circ . Проверяем свойство нуля

$$0 + x = x. \quad (20)$$

Операция $*$ является обратной по отношению к операции \circ , поэтому $0 + x = 0 * (0 \circ x) = x$.

Доказываем единственность нуля. Предположим, что существуют два нулевых вектора $0'$ и $0''$, так что для любого $x \in N$ имеем $0' + x = x$ и $0'' + x = x$. В частности, $0' + 0'' = 0''$ и $0' + 0'' = 0'$. В силу коммутативности сложения получаем $0' = 0''$, что и требовалось доказать. Однозначность решения уравнения $x + y = 0$ при любом $x \in N$ относительно y выводится следующим образом: из равенства $x + y = 0$ и (18) следует $0 * (x \circ y) = 0$, откуда с учетом свойства $x \circ x = x$ получаем $x \circ y = 0 \circ 0 = 0$, $y = x * 0$. Таким образом,

$$-x = x * 0. \quad (21)$$

Доказываем ассоциативность сложения

$$(x + y) + z = x + (y + z). \quad (22)$$

Полагая в законе четырехугольника $t = z$, имеем $(x \circ y) \circ (z \circ z) = (z \circ x) \circ (y \circ z)$. Пользуясь коммутативностью операции внутреннего равноделения и свойством $x \circ x = x$, приходим к закону дистрибутивности для операции \circ : при любых $x, y, z \in N$

$$(x \circ y) \circ z = (x \circ z) \circ (y \circ z). \quad (23)$$

В равенстве (23) заменяем x на $z * x$, y на $z * y$. Тогда:

$$((z * x) \circ (z * y)) \circ z = ((z * x) \circ z) \circ ((z * y) \circ z),$$

$$((z * x) \circ (z * y)) \circ z = x \circ y,$$

$$(z * x) \circ (z * y) = z * (x \circ y).$$

После замены x на y , y на z и z на x приходим еще к одному закону дистрибутивности: при любых $x, y, z \in N$

$$(x * y) \circ z = (x * z) \circ (x * y). \quad (24)$$

Полагая в аксиоме четырехугольника $t = 0$, имеем: $(x \circ y) \circ (0 \circ z) = (0 \circ x) \circ (y \circ z)$. Из последнего равенства с помощью (24) выводим:

$$0 * ((x \circ y) \circ (0 \circ z)) = 0 * ((0 \circ x) \circ (y \circ z)),$$

$$(0 * (x \circ y)) \circ (0 * (0 \circ z)) = (0 * (0 \circ x)) \circ (0 * (y \circ z)),$$

$$(x + y) \circ z = x \circ (y + z).$$

Наконец, приходим к закону ассоциативности:

$$0 * ((x + y) \circ z) = 0 * (x \circ (y + z)),$$

$$(x + y) + z = x + (y + z).$$

Итак, множество N вместе с заданной на нем операцией сложения (18) образует абелеву группу.

Приступаем к введению операции умножения вещественного числа на вектор x пространства N^* . Полагаем по определению операции умножения

$$0x = 0, \quad (25)$$

$$1x = x, \quad (26)$$

$$(-1)x = -x. \quad (27)$$

Определяем операцию удвоения $2x$ и операцию деления пополам $\frac{1}{2}x$ вектора x . Полагаем $2x = x + x$. Это означает, что

$$2x = 0x. \quad (28)$$

Принимаем, что $y = \frac{1}{2}x$ тогда и только тогда, когда $2y = x$. Отсюда выводим

$$\frac{1}{2}x = 0 \circ x. \quad (29)$$

Определяем умножение числа 2^k на вектор x k -кратным применением операции удвоения к вектору x :

$$2^k x = 2(2\dots(2x)\dots). \quad (30)$$

Умножение числа 2^{-k} на вектор x определяем k -кратным применением операции деления пополам к вектору x :

$$2^{-k} x = \frac{1}{2}\left(\frac{1}{2}\dots\left(\frac{1}{2}x\right)\dots\right). \quad (31)$$

Здесь k – произвольное натуральное число.

Далее определяем операцию умножения суммы и произведения чисел 2^k и 2^l на вектор x :

$$(2^k + 2^l)x = 2^k x + 2^l x, \quad (32)$$

$$(2^k 2^l)x = 2^k (2^l x). \quad (33)$$

Здесь k, l – целые числа. Для любого целого k по определению полагаем:

$$(-2^k)x = 2^k(-x). \quad (34)$$

Докажем, что

$$-2x = 2(-x). \quad (35)$$

Для этого вводим векторы $u = 0 * x$ и $v = u * 0$. Согласно тождеству $x \circ x = x$ имеем $(u \circ v) \circ (u \circ v) = (u \circ v)$. Применяя к полученному равенству свойство коммутативности операции \circ и ее свойство дистрибутивности (23), получаем $((u \circ v) \circ u) \circ ((u \circ v) \circ v) = u \circ v$. Из равенства $u * 0 = v$ выводим $u \circ v = 0$, поэтому $(0 \circ u) \circ (0 \circ v) = 0$. Из равенства $0 * x = u$ выводим $0 * u = x$, поэтому $x \circ (0 \circ v) = 0$. Отсюда $0 \circ v = x * 0$, $0 * (x * 0) = v = u * 0 = (0 * x) * 0$.

Итак, $(0 * x) * 0 = 0 * (x * 0)$. Используя (21) и (28), последнее равенство переписываем в виде $(2x)0 = 0 * (-x)$, а затем в виде $-(2x) = 2(-x)$. Тождество (35) доказано.

Докажем, что

$$-\left(\frac{1}{2}x\right) = \frac{1}{2}(-x). \quad (36)$$

Для этого вводим вектор $y = x * 0$. Согласно тождеству $x \circ x = x$ имеем: $(x \circ y) \circ (x \circ y) = (x \circ y)$. Применяя к полученному равенству свойство коммутативности операции \circ и ее свойство дистрибутивности (23), получаем: $((x \circ y) \circ x) \circ ((x \circ y) \circ y) = x \circ y$. Из равенства $y = x * 0$ выводим $x \circ y = 0$, поэтому $(0 \circ x) \circ (0 * y) = 0$. Из последнего равенства выводим $(0 \circ x) * 0 = 0 \circ y = 0 \circ (x * 0)$. Итак, $(0 \circ x) * 0 = 0 \circ (x * 0)$. Используя (22) и (29), последнее равенство переписываем в виде:

$\left(\frac{1}{2}x\right) * 0 = 0 \circ (-x)$. Тождество (36) доказано.

Из определений (30) и (31) и тождеств (35), (36) непосредственно вытекает тождество

$$-(2^k x) = 2^k(-x), \quad (37)$$

справедливое при любом целом k . Из определений (30), (31) также следует тождество

$$2^k(2^l x) = 2^{k+l} x, \quad (38)$$

справедливое при любых целых k и l .

Докажем, что при любом целом k имеет место тождество

$$2^k x + 2^k y = 2^k(x + y). \quad (39)$$

Действительно, при $k = 0$ тождество (39) выполняется: $2^0 x + 2^0 y = 1x + 1y = x + y = 1(x + y) = 2^0(x + y)$. Выполняется оно и при $k = 1$:

$$2^1 x + 2^1 y = 2x + 2y = (x + x) + (y + y) = 2(x + y) = 2^1(x + y).$$

Предположим, что при некотором $k > 0$ тождество (39) выполняется.

Тогда $2^{k+1}x + 2^{k+1}y = 2^k(2x) + 2^k(2y) = 2^k(2x + 2y) = 2^k(2(x + y)) = 2^{k+1}(x + y)$. Для случая $k < 0$ свойство (39) доказывается аналогично.

Из (37) и (39) вытекает, что

$$2^k x - 2^k y = 2^k(x - y). \quad (40)$$

Назовем двоично-рациональным любое число вида

$$\lambda = \sum_{k=-m}^n a_k 2^k, \quad (41)$$

где $a_k \in \{-1, 0, 1\}$, а m и n – некоторые натуральные числа. При любом двоично-рациональном λ определяем операцию λx равенством

$$\lambda x = \sum_{k=-m}^n a_k (2^k x). \quad (42)$$

Доказательство корректности этого определения, а также доказательство тождеств

$$\lambda(x + y) = \lambda x + \lambda y, \quad (43)$$

$$(\lambda + \mu)x = \lambda x + \mu x, \quad (44)$$

$$\lambda(\mu x) = (\lambda \mu)x \quad (45)$$

для любых двоично-рациональных чисел λ, μ достаточно громоздко и носит технический характер, поэтому здесь не приводится.

Чтобы перейти от двоично-рациональных чисел к вещественным, придется опереться на закон непрерывности и еще на одно дополнительное свойство метризирующего предиката: закон сходимости – для любой сходящейся последовательности двоично-рациональных чисел $\{\lambda_n\}_{n=1}^{\infty}$ и произвольного $x \in N$ существует $\lim \lambda_n x$. Законы непрерывности и сходимости позволяют определить операцию умножения произвольного вещественного числа λ на любой элемент x множества N :

$$\lambda x = \lim \lambda_n x, \quad (46)$$

где $\{\lambda_n\}$ – последовательность двоично-рациональных чисел, для которой $\lim \lambda_n x = \lambda$.

Корректность определения (46) обосновывается тем, что любое вещественное число λ единственным способом представляется бесконечным двоичным кодом, а следовательно, предел последовательности (46) для числа λ может быть только один. Из определения (46) и перечисленных выше свойств метризирующего предиката следует справедливость равенств (13)–(15) для любых вещественных чисел λ и μ . Итак, мы ввели на множестве N структуру линейного пространства над полем вещественных чисел, согласованную с операциями внутреннего и внешнего равноделения точек. Этим доказаны следующие утверждения:

Утверждение 1. Из законов парной рефлексивности, симметричности и транзитивности, одиночной симметричности, тождества, равноделения, непрерывности, четырехугольника и сходимости следует существование на области задания метризирующего предиката структуры линейного пространства над полем вещественных чисел с операцией сложения векторов, определяемой равенством (17).

Кроме того, имеет место следующее очевидное

Утверждение 2. Любой метризирующий предикат подчиняется закону сходимости.

С содержательной точки зрения закон сходимости означает следующее. Если задаться произвольным вещественным числом λ и попытаться экспериментально найти точку λx для произвольно выбранной точки x множества N , то это всегда удастся сделать, причем при повторном выполнении процедуры отыскания точки λx придется выполнить ту же самую последовательность внутренних и внешних равноделений. Поскольку любая физическая система, являющаяся объектом идентификации, имеет конечную чувствительность, то процесс отыскания точки λx не будет длиться бесконечно долго, а закончится за конечное число шагов.

5. Установление условий существования метризирующего предиката

В этом разделе формулируются два новых свойства метризирующего предиката и доказывается, что они, вместе со свойствами, введенными ранее в этой главе, образуют полную систему условий, обеспечивающих существование метризирующего предиката. В связи с недостатком места некоторые детали доказательства здесь опущены.

Сформулируем закон n -мерности – пространство N^* , введенное в предыдущем параграфе, n -мерно.

В развернутой форме закон n -мерности запишем в следующем виде: существуют векторы $e_1, e_2, \dots, e_n \in M^*$ такие, что равенство

$$\Phi(a, a, x, \sum_{i=1}^n \alpha_i(x) e_i^*) = 1 \quad (47)$$

выполняется для каждого $x \in N^*$ при единственном наборе коэффициентов $\alpha_1(x), \alpha_2(x), \dots, \alpha_n(x)$. Здесь a – какой-нибудь вектор из N^* ; $\alpha_1, \alpha_2, \dots, \alpha_n$ – фиксирован-

ные функции, определенные на N^* со значениями в поле вещественных чисел.

Согласно второму закону тождества условие (47) равносильно равенству

$$x = \sum_{i=1}^n \alpha_i(x) e_i^*, \quad (48)$$

которое означает, что любой вектор x выражается в виде линейной комбинации векторов $e_1^*, e_2^*, \dots, e_n^*$ при единственном наборе коэффициентов $\alpha_1, \alpha_2, \dots, \alpha_n$, а именно к этому сводится содержание закона n -мерности.

С содержательной точки зрения закон n -мерности означает следующее. Пусть в пространстве N^* произвольно выбрана точка x . Возьмем n фиксированных точек $e_1^*, e_2^*, \dots, e_n^* \in N^*$ и попытаемся из них и из нулевой точки пространства N^* , комбинируя их многократно и в различной последовательности с помощью операций внутреннего и внешнего равноделения, получить точку x . Закон n -мерности требует, чтобы существовала такая система точек $e_1^*, e_2^*, \dots, e_n^*$, при которой попытки такого рода всегда заканчиваются успешно. Он также требует, чтобы при использовании меньшего числа каких бы то ни было фиксированных точек не было возможности получить из них и из точки 0 точку x . В случае пространственной интерпретации закон n -мерности выполняется при $n = 2$, он означает, что поле зрения двумерно. В случае цветовой интерпретации закон n -мерности выполняется при $n = 3$, он означает, что пространство цветов трехмерно.

Возьмем в роли N' n -мерное арифметическое пространство R^n , а в роли φ' – функцию, отображающую пространство N^* на N' так, что вектору $x \in N^*$ с координатами $\xi_1, \xi_2, \dots, \xi_n$ соответствует в N' точка $x' = (\xi_1, \xi_2, \dots, \xi_n)$. Предполагается, что базис в N^* зафиксирован. Очевидно, что φ' есть гомеоморфизм. В арифметическом пространстве любой размерности может быть введена евклидова метрика. Вместе с тем, все евклидовы пространства одинаковой размерности гомеоморфны [18]. Пространства φ и N' – арифметические и имеют одну и ту же размерность n , следовательно, существует гомеоморфизм φ , отображающий пространство N на пространство N' . Отсюда непосредственно следует существование гомеоморфизма $\varphi'' = (\varphi')^{-1} \varphi$, отображающего пространство N на пространство N^* .

Для аксиоматического обоснования структуры метрического предиката Φ , задаваемой соотношениями (1) и (3), осталось доказать возможность введения такого базиса $\{e_1, e_2, \dots, e_n\}$ в N , при котором метрика, индуцируемая предикатом Φ , оказалась бы согласованной с метрикой пространства N' . Сказанное означает, что расстояние $\rho(x', y')$ между точками x', y' пространства N' , определяемое формулой (3), должно удовлетворить условию (1).

Сформулируем закон метричности – существует вектор $h \in N^*$ такой, что для любых $x, y \in N^*$ найдется единственное неотрицательное число $\alpha(x, y)$, для которого

$$\Phi(x, y, \alpha(x, y)h, 0) = 1. \quad (49)$$

Справедливы следующие утверждения:

Утверждение 3. Любой метризирующий предикат подчиняется законам n -мерности и метричности.

Утверждение 4. Функция

$$\rho(x', y') = \alpha(\varphi^{-1}(x'), \varphi^{-1}(y')) \quad (50)$$

обладает всеми свойствами евклидова расстояния [19] и выражается в виде (3).

Утверждение 5. Если принять $e_1 = \varphi^{-1}(1, 0, 0, \dots, 0, 0)$, $e_2 = \varphi^{-1}(0, 1, 0, \dots, 0, 0)$, ..., $e_n = \varphi^{-1}(0, 0, 0, \dots, 0, 1)$, то при любых $x_1, x_2, y_1, y_2 \in N$ равенство $\rho(x_1, y_1) = \rho(x_2, y_2)$ будет равносильно равенству $\Phi(x_1, y_1, x_2, y_2) = 1$.

Из утверждений 1–5 непосредственно следует теорема об условиях существования метризирующего предиката:

Теорема. Для того, чтобы предикат $\Phi(x_1, y_1, x_2, y_2)$ был метризирующим, необходимо и достаточно, чтобы он подчинялся законам парной рефлексивности, симметричности и транзитивности, одиночной симметричности, первому и второму законам тождества, законам внутреннего и внешнего равноделения, непрерывности, четырехугольника, сходимости, n -мерности и метричности.

С содержательной точки зрения закон метричности означает следующее. Пусть в пространстве N^* произвольно выбраны точки x , y и точка h , отличная от нулевой точки. Тогда можно будет найти, причем единственным образом, такую точку $t = \alpha(x, y)h$, лежащую на луче, исходящем из точки 0 и проходящем через точку h , которая находится на таком же расстоянии от точки 0 , что и точки x , y друг от друга. Это означает, что расстоянию между любыми двумя точками пространства N^* , индуцируемому предикатом Φ , можно поставить в соответствие числовую меру. Таким образом, можно говорить не только о равенстве или неравенстве расстояний между точками пространства N^* , но и о самом расстоянии, которое может быть измерено вещественным числом.

Смысл теоремы для цветовой интерпретации состоит в том, что существует субъективное цветовое пространство N' , а также взаимно однозначная и взаимно непрерывная функция φ , отображающая объективное цветовое пространство N в субъективное, такие что геометрическое расстояние между точками x', y' пространства N' в точности соответствует психологическому расстоянию между цветами, представленными этими точками. В пространственной интерпретации теорема означает, что существует гомеоморфизм φ , преобразующий объективное поле зрения в субъективное, причем в последнем расстояние между точками из-

меряется посредством евклидовой метрики. В технической интерпретации теорема указывает полный набор свойств системы управления, при выполнении которых возможно задать такое искривление пространства выходных сигналов объекта, которое обеспечит равную точность представления этих сигналов.

Выводы

Сформулирована задача о компараторной идентификации метрики пространства выходных сигналов линейного конечномерного объекта и показано ее практическое значение для математического описания технических и природных объектов.

Установлены необходимые и достаточные экспериментально проверяемые условия, при которых возможна компараторная идентификация метрики пространства выходных сигналов линейного конечномерного объекта.

Даны цветовая, пространственная и техническая интерпретации задачи о компараторной идентификации метрики пространства выходных сигналов линейного конечномерного объекта, определяющие возможные области ее практического применения.

Сформулирована и доказана теорема об условиях существования метризирующего предиката, решающая задачу структурной компараторной идентификации метрики пространства выходных сигналов линейного конечномерного объекта.

Список литературы: 1. Бондаренко М.Ф., Шабанов-Кушнарченко С.Ю., Шабанов-Кушнарченко Ю.П. Компараторная идентификация кольцевого зору людини // Біоніка інтелекту: наук.-техн. журнал. – 2008. – № 0. – С. 000–000. 2. Бондаренко М.Ф., Шабанов-Кушнарченко С.Ю., Шабанов-Кушнарченко Ю.П. Идентификация объектов, описываемых векторами и линейными операциями над ними // Біоніка інтелекту: наук.-техн. журнал. – 2009. – №0. – С. 000–000. 3. Гебер Р. Курс физиологии человека. – М.-Л.: Биомедгиз, 1936. – 678 с. 4. Матвеев А.Б. Проблемы построения равноконтрастного пространства // Светотехника. – 1964. – № 12. – С. 1–6. 5. Люстерник Л.А., Соболев В.И. Элементы функционального анализа. – М.: Наука, 1965. – 520 с. 6. Эдвардс Р. Функциональный анализ. Теория и приложения. – М.: ИЛ, 1969. – 1072 с. 7. Бондаренко М.Ф., Шабанов-Кушнарченко С.Ю. Об условиях существования линейного предиката. – Киев: Деп. в УкрНИИИТИ 11.05.88, №1130. – 9 с. 8. Гельмгольц Г. Новейшие успехи теории зрения. – Л.: Изд. «П.П. Сойкин», 1929. – 47 с. 9. Бондаренко М.Ф., Шабанов-Кушнарченко С.Ю., Шабанов-Кушнарченко Ю.П. Об общей теории компараторной идентификации // Бионика интеллекта: науч.-техн. журнал. – 2008. – №2(69). – С. 13–22. 10. Артамонов И.Д. Иллюзии зрения. – М.: Наука, 1964. – 102 с. 11. Авербах М.И. Офтальмологические очерки. – М.: Изд-во АН СССР, 1949. – 788 с. 12. Майстровская Л.М. Математические модели сенсорных систем и их технические приложения: дисс....канд. техн. наук. – Харьков, 1977. – 145 с. 13. Румянцев М.В. Видимость малых объектов // Проблемы физиологической оптики. – 1949. – т. VII. – С. 163. 14. Подиновский В.В., Гаврилов В.М. Оптимизация по последовательно применяемым критериям. – М.: Сов. радио, 1975. – 192 с.

15. Теория выбора и принятия решений / И.М. Макаров, Т.М. Виноградская, А.А. Рубчинский, В.Б. Соколов. – М.: Наука, 1982. – 328 с. 16. Шавров А.В., Солдатов В.В. Многокритериальное управление в условиях статистической неопределенности. – М.: Машиностроение, 1990. – 159 с. 17. Шилов Г.Е. Математический анализ (конечномерные линейные пространства). – М.: Наука, 1969. – 432 с. 18. Понтрягин Л.С. Непрерывные группы. – М.-Л.: Гостехиздат, 1938. – 136 с. 19. Рисс Ф., Секефальви-Надь Б. Лекции по функциональному анализу. – М.: Мир, 1979. – 581 с.

Поступила в редколлегию 26.01.2009.

УДК 519.7

Розробка теорії структурної метризації скінченномірному арифметичного простору / М.Ф. Бондаренко, С.Ю. Шабанов-Кушнарєнко, Ю.П. Шабанов-Кушнарєнко // Біоніка інтелекту: наук.-техн. журнал. – 2009. – №1(70). – С. 17–27.

У статті сформульовано й вирішено завдання про компараторну ідентифікацію метрики простору вихідних сигналів лінійного скінченномірному об'єкта. Установлені необхідні й достатні умови, що перевіряються експериментально, за яких можлива компараторна ідентифікація цієї метрики.

Л.: 5. Бібліогр.: 19 найм.

UDC 519.7

Working out of the finite-dimensional arithmetic space structural metrization theory / M.F. Bondarenko, S.Yu. Shabanov-Kushnarenko, Yu.P. Shabanov-Kushnarenko // Bionics of Intelligence: Sci. Mag. – 2009. – №1(70). – P. 17–27.

In article the problem about comparator identifications of the linear finite-dimensional object target signals space metrics is formulated and solved. Necessary and sufficient experimentally checked conditions at which comparator identification of this metrics is possible are established.

Fig.: 5. Ref.: 19 items.

УДК 519.237.8.:004.8



МЕТОДЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В ЭЛЕКТРОННЫХ ХРАНИЛИЩАХ

Г. Г. Асеев

Харьковская государственная академия культуры, г. Харьков, Украина
aseev@ic.ac.kharkov.ua

Сегодня наблюдается небывалый подъем комплексной компьютеризации корпоративных организаций, причем первостепенная роль отводится построению автоматизированных систем документооборота. В статье рассматриваются современные методы статистики для построения задач классификации и многомерного анализа данных в хранилищах электронного документооборота.

СЫРЫЕ ДАННЫЕ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, КЛАСТЕРНЫЙ АНАЛИЗ, МЕРА БЛИЗОСТИ, АЛГОРИТМ ТАКСОНОМИИ, АССОЦИАТИВНЫЕ ПРАВИЛА, ТРАНЗАКЦИЯ

Введение

Кластерный анализ применяется для исследования и обнаружения «машиной» (алгоритмами, средствами искусственного интеллекта) в сырых данных электронных хранилищ скрытых структур или зависимостей, которые: ранее не были известны, нетривиальны, практически полезны, доступны для интерпретации знаний человеком, необходимых для принятия решений в различных сферах человеческой деятельности.

Ранее была рассмотрена проблема обнаружения нового знания в хранилищах данных методами Knowledge Discovery in Databases (KDD) и Data Mining [1, 2].

1. Методы извлечения и анализа данных

Информация, найденная в процессе применения методов Data Mining, должна быть нетривиальной и ранее неизвестной, например, средние продажи не являются таковыми. Знания должны описывать новые связи между свойствами, предсказывать значения одних признаков на основе других и так далее. Найденные знания должны быть применимы и на новых данных с некоторой степенью достоверности. Полезность заключается в том, чтобы эти знания могли принести определенную выгоду при их применении. В случае, когда извлеченные знания непрозрачны для пользователя, должны существовать методы постобработки, позволяющие привести их к интерпретируемому виду.

Методы извлечения и анализа данных делятся на **описательные**: математическую статистику (оценивание параметров распределения; проверка статистических гипотез; дисперсионный и регрессионный анализ; анализ временных рядов, который, в свою очередь, состоит из спектрального и корреляционного анализа и фильтрации; многомерный анализ, который, в свою очередь, состоит из кластерного, дискриминантного и факторного анализа, метода главных компонент и шкалирование) [3-7] и **предсказательные**: эволюционное моделирование (генетические алгоритмы; искусственные нейронные сети, которые, в свою очередь, делятся на

RBF и ART сети; сети обратного и встречного распространения; сети Хемминга, Хопфилда, Кохонена и гибридно нечетко-нейронные и пр.); машинное обучение (деревья решений, которые, в свою очередь, делятся на энтропийную меру, ID3, C4.5, NewID и прочее) [8]. В связи с тем, что в короткой статье мы не в состоянии провести анализ, сравнение и рекомендации практического применения всех перечисленных методов, пока ограничимся основными моделями извлечения и анализа данных с помощью кластерного анализа и метода k_n ближайших соседей. Остальные модели будут рассмотрены в следующих публикациях.

2. Кластерный анализ

Задача кластерного анализа [9–11] – выделение групп в заданной (обучающей) совокупности элементов (путем итерационного слияния наиболее близких кластеров).

Кластерный анализ позволяет группировать данные вокруг нескольких центров в n -мерном пространстве. В литературе [9-11] описываются методы кластеризации полным перебором (теоретически), методами математического программирования, на основе матриц сходств, на основе оценивания функции плотности.

Первая группа методов, которые будут рассмотрены – алгоритмы таксономии. Кластер определяется как совокупность элементов, лежащих на расстоянии не больше r от центра (внутри гиперболы радиуса r или гиперкуба со сторонами $2r$). При этом в качестве центра выбирается один из элементов и формируется кластер из элементов, удаленных от него не далее, чем на r . Боннер предлагает выбирать очередной центр случайно, Хиверинен предлагает в качестве очередного центра брать «типическую» точку – лежащую на минимальном расстоянии от центра оставшегося множества объектов. Далее процедура повторяется для оставшихся элементов. Элементы, не попавшие ни в один кластер после определенного числа шагов или образования кластеров с требуемыми показателями, считаются нераспознанными, могут трактоваться как шум

в системе распознавания и исключаются из обучающей совокупности элементов.

Другая важная группа эвристических методов кластеризации – методы, основывающиеся на *последовательной агломеративной процедуре*.

Все эти методы дают оптимальное решение в классе меньшем, чем класс всех возможных разбиений (кластеров), однако достоинством этих методов является простота вычислительной процедуры и алгоритмов.

Пусть анализируемая совокупность состоит из p элементов, каждый из которых характеризуется значениями n дискриминантных переменных.

На первом шаге итеративной процедуры имеется p кластеров, каждый из которых включает по одному элементу.

Определяются два наиболее близких или сходных кластера, объединяются в один кластер, количество кластеров сокращается на 1: $p \rightarrow p - 1$.

Мера близости определяется расстоянием между элементами, заносимыми в симметричную матрицу расстояний D :

$$D = \begin{pmatrix} 0 & d_{12} & \dots & d_{1p} \\ d_{21} & 0 & \dots & d_{2p} \\ \cdot & \cdot & \ddots & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ d_{p1} & d_{p2} & \dots & 0 \end{pmatrix}.$$

Наиболее близкими считаются объекты с *наименьшим расстоянием*.

Расстояние между точками, между центрами кластеров определяются разными метриками. Так, расстояние между l -ой и r -ой точками в евклидовой метрике

$$\text{равно: } d_{rj}^{(2)} = \sqrt{\sum_{i=1}^n (x_{ri} - x_{ji})^2}.$$

Другие метрики.

Норма:

$$d_{rj}^{(1)} = \sum_{i=1}^n (x_{ri} - x_{ji}) - l_i.$$

Супремум-норма:

$$d_{rj}^{(\infty)} = \sup_{i=1,2,\dots,n} \{|x_{ri} - x_{ji}|\}.$$

Норма, которая охватывает и предыдущие (при $p = 2$, $p = 3$ и $p = \infty$):

$$d_{rj}^{(p)} = \left[\sum_{i=1}^n (|x_{ri} - x_{ji}|^p) \right]^{1/p} - l_p.$$

Махаланобиса:

$$D^2(X_i, X_j) = (X_i - X_j)^T W^{-1} (X_i - X_j),$$

где W - матрица рассеяния.

Возможны и другие способы определения *расстояния между классами*. Пусть объекты $\{X_i\}, i = 1, \dots, k_1$ принадлежат к одному классу, а $\{Y_j\}, j = 1, \dots, k_2$ – к другому.

Минимальное локальное расстояние:

$$D_1 = \min_{\substack{i=1,\dots,k_1 \\ j=1,\dots,k_2}} d(X_i, X_j).$$

Максимальное локальное расстояние:

$$D_2 = \max_{\substack{i=1,\dots,k_1 \\ j=1,\dots,k_2}} d(X_i, X_j).$$

Среднее расстояние:

$$D_3 = \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} \frac{d(X_i, X_j)}{k_1 k_2}.$$

Статистическое расстояние между кластерами:

$$D_4 = \frac{k_1 k_2}{k_1 + k_2} (\bar{X} - \bar{Y})^T (\bar{X} - \bar{Y}),$$

где

$$\bar{X} = \sum_{i=1}^{k_1} \frac{x_i}{k_1}, \quad \bar{Y} = \sum_{j=1}^{k_2} \frac{y_j}{k_2}.$$

Расстояние между центроидами:

$$D_5 = (\bar{X} - \bar{Y})^T (\bar{X} - \bar{Y}).$$

На каждом последующем шаге агломеративной процедуры потребуется пересчет лишь для одной строки (и одного столбца) D , то есть рассчитываются расстояния от образованного кластера до каждого из оставшихся кластеров. Существует несколько методов пересчета расстояний с использованием старых значений расстояний для объединяемых кластеров, отличающихся коэффициентами в формуле:

$$d_{rs} = \alpha_p d_{ps} + \alpha_q d_{qs} + \beta d_{pq} + \gamma |d_{ps} - d_{qs}|.$$

Если кластеры p и q объединяются в кластер r и требуется рассчитать расстояние от нового кластера до кластера s , применение того или другого метода зависит от способа определения расстояния между кластерами, различные методы различаются значениями коэффициентов $\alpha_p, \alpha_q, \beta$ и γ .

Основанием для слияния кластеров может быть не только мера их близости (расстояние между ними), но и *мера сходства* – неотрицательная функция $s(X_i, X_j) = s_{ij}$, причем $0 \leq s_{ij} \leq 1, i \neq j, s_{ii} = 1, s_{ij} = s_{ji}$. Меры сходства всех кластеров объединяются в симметричную матрицу S :

$$S = \begin{pmatrix} 1 & s_{12} & \dots & s_{1p} \\ s_{21} & 1 & \dots & s_{2p} \\ \cdot & \cdot & \ddots & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ s_{p1} & s_{p2} & \dots & 1 \end{pmatrix}.$$

Наиболее сходными считаются объекты со значением s_{ij} наиболее близким к 1, процедура образования кластеров при этом аналогична описанной выше с использованием меры близости.

В качестве меры сходства можно использовать коэффициент корреляции:

$$r_{ij} = \sum_{l=1}^n x_{li} x_{lj} / \sqrt{\sum_{l=1}^n x_{li}^2 \sum_{l=1}^n x_{lj}^2},$$

причем

$$\overset{\circ}{x}_{li} = x_{li} - \bar{x}_i, \quad \bar{x}_i = \sum_{l=1}^n x_{li}, \quad \overset{\circ}{x}_{ij} = x_{ij} - \bar{x}_j, \quad \bar{x}_j = \sum_{j=1}^n x_{ij}.$$

После выполнения очередного шага агломеративной процедуры выясняется, что желательное разбиение достигнуто. Существуют различные методы определения критерия остановки процедуры:

получено определенное заранее количество кластеров;

все кластеры содержат более определенного числа элементов;

кластеры обладают требуемым соотношением внутренней однородности и разнородности между собой.

После анализа обучающей выборки можно решать задачу классификации новых состояний $\{X_i\}$, относя их к той или другой группе, что определяется по минимальному расстоянию до центров групп.

3. Метод k_n ближайших соседей

Здесь идея состоит в том, что вокруг распознаваемого объекта \bar{x} строится ячейка объёма V . При этом неизвестный объект относится к тому образу, число обучающих представителей которого в построенной ячейке оказалось в большинстве. Если использовать статистическую терминологию, то число объектов образа s_i , попавших в данную ячейку, характеризует оценку усреднённой по объёму V плотности вероятности $p(\bar{x}/s_i)$ [12].

Для оценки усреднённых $p(\bar{x}/s_i)$ нужно решить вопрос о соотношении между объёмом V ячейки и количеством попавших в эту ячейку объектов того или иного класса (образа). Вполне разумно считать, что чем меньше V , тем более тонко будет охарактеризована $p(\bar{x}/s_i)$. Но при этом тем меньше объектов попадёт в интересующую нас ячейку, а следовательно, тем меньше достоверность оценки $p(\bar{x}/s_i)$. При чрезмерном увеличении V возрастает достоверность оценки $p(\bar{x}/s_i)$, но теряются тонкости её описания из-за усреднения по слишком большому объёму, что может привести к негативным последствиям (увеличению вероятности ошибок распознавания). При небольшом объёме обучающей выборки V целесообразно брать предельно большим, но обеспечить при этом, чтобы внутри ячейки плотности $p(\bar{x}/s_i)$ мало изменялись.

Тогда их усреднение по большому объёму не очень опасно. Таким образом, вполне может случиться, что объём ячейки, уместный для одного значения \bar{x} , может совершенно не годиться для других случаев.

Предлагается следующий порядок действий (пока что принадлежность объекта тому или иному образу учитывать не будем).

Для того чтобы оценить $p(\bar{x})$ на основании обучающей выборки, содержащей n объектов, центрируем ячейку вокруг \bar{x} и увеличиваем её объём до тех пор, пока она не вместит k_n объектов, где k_n есть некоторая функция от n . Эти k_n объектов будут ближайшими соседями \bar{x} . Вероятность P попадания вектора \bar{x} в

область R определяется выражением $P = \int_R p(\bar{x}') d\bar{x}'$.

Это сглаженный (усреднённый) вариант плотности распределения $p(\bar{x})$. Если взять выборку из n объектов (простым случайным выбором из генеральной совокупности), то k из них окажется внутри области R . Вероятность попадания k из n объектов в R описывается биномиальным законом, имеющим резко выраженный максимум около среднего значения nP . При этом k/n является неплохой оценкой для P .

Если теперь допустить, что R настолько мала, что $p(\bar{x})$ внутри неё меняется незначительно, то

$$\int_R p(\bar{x}') d\bar{x}' \approx p(\bar{x})V,$$

где V – объём области R ; \bar{x} – точка внутри R .

Тогда $P \approx p(\bar{x})V$. Но $P \approx \frac{k}{n}$, следовательно, $p(\bar{x}) \approx \frac{k/n}{V}$.

Итак, оценкой $p_n(\bar{x})$ плотности $p(\bar{x})$ является величина

$$p_n(\bar{x}) = \frac{k_n/n}{V_n}. \quad (*)$$

Без доказательства приведём утверждение, что условия

$$\lim_{n \rightarrow \infty} k_n = \infty \quad \text{и} \quad \lim_{n \rightarrow \infty} \frac{k_n}{n} = 0 \quad (**)$$

являются необходимыми и достаточными для сходимости $p_n(\bar{x})$ к $p(\bar{x})$ по вероятности во всех точках, где плотность $p(\bar{x})$ непрерывна.

Этому условию удовлетворяет, например, $k_n = \sqrt{n}$.

Теперь будем учитывать принадлежность объектов к тому или иному образу и попытаемся оценить апостериорные вероятности образов $p(s_i/\bar{x})$.

Предположим, что мы размещаем ячейку объёма V вокруг \bar{x} и захватываем выборку с количеством объектов k_n, k_{ni} из которых принадлежат образу s_i . Тогда в соответствии с формулой (*) оценкой совместной вероятности $p(\bar{x}, s_i)$ будет величина

$$p_n(\bar{x}, s_i) = \frac{k_{ni}/n}{V_n},$$

а

$$P_n(s_i/\bar{x}) = \frac{p_n(\bar{x}, s_i)}{\sum_{j=1}^M p_n(\bar{x}, s_j)} = \frac{k_{ni}}{k_n}.$$

Таким образом, апостериорная вероятность $P_n(s_i/\bar{x})$ оценивается как доля выборки в ячейке, относящаяся к s_i . Чтобы свести уровень ошибки к минимуму, нужно объект с координатами \bar{x} отнести к классу (образу), количество объектов обучающей выборки которого в ячейке максимально. При $n \rightarrow \infty$ такое правило является байесовским, то есть обеспечивает теоретический минимум вероятности ошибок распознавания (разумеется, при этом должны выполняться условия(**)).

Правило ближайшего соседа. Пусть $X_n = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}$ – множество объектов обучающей последовательности, то есть принадлежность каждого из них тому или иному образу достоверно известна. Пусть также $\bar{x}^* \in X_n$ является объектом, ближайшим к распознаваемому $\bar{x} \notin X_n$. Напомним, что при этом правило ближайшего соседа для классификации \bar{x} состоит в том, что \bar{x} относят к тому классу (образу), которому принадлежит \bar{x}^* . Естественно, такое отнесение носит случайный характер. Вероятность того, что \bar{x} будет отнесён к s_i , есть апостериорная вероятность $P(s_i/\bar{x}^*)$. Если n очень велико, то вполне можно допустить, что \bar{x} расположен достаточно близко к \bar{x}^* , настолько близко, что $P(s_i/\bar{x}^*) \approx P(s_i/\bar{x})$. А это есть не что иное, как рандомизированное решающее правило: \bar{x} относят к s_i с вероятностью $P(s_i/\bar{x})$. Байесовское решающее правило основано на выборе максимальной апостериорной вероятности, то есть \bar{x} относят к s_i в том случае, если

$$P(s_j/\bar{x}) = \max_i P(s_i/\bar{x}).$$

Отсюда видно, что если $P(s_j/\bar{x})$ близка к единице, то правило ближайшего соседа даёт решение, в большинстве случаев совпадающее с байесовским. Напомним, что эти рассуждения имеют достаточные основания лишь при очень больших n (объёмах обучающей

выборки). Такие условия на практике встречаются не так часто, но позволяют понять статистический смысл правила ближайшего соседа.

4. Анализ рыночной корзины

Анализом рыночной корзины называется задача поиска ассоциативных правил, которые описывают типичные шаблоны покупок, совершаемых в супермаркетах. Задача поиска ассоциативных правил впервые была представлена для анализа рыночной корзины, за что и получила свое название. Однако, сфера применения не ограничивается лишь одной торговлей. Ассоциативные правила также успешно применяют и в других областях: медицине, для анализа посещений веб-страниц (Web Mining), для анализа текста (Text Mining) для анализа данных по переписи населения [13], в анализе и прогнозировании сбоев телекоммуникационного оборудования и так далее.

Ассоциативные правила (Association Rules) [8, 14].

Впервые эта задача была предложена при поиске ассоциативных правил для нахождения типичных шаблонов покупок, совершаемых в супермаркетах, поэтому иногда ее еще называют анализом рыночной корзины (market basket analysis).

Пусть имеется база данных, состоящая из покупательских транзакций. Каждая транзакция – это набор товаров, купленных покупателем за один визит. Такую транзакцию еще называют рыночной корзиной.

Определение 1. Пусть $I = \{i_1, i_2, i_3, \dots, i_n\}$ – множество (набор) товаров, называемых элементами. Пусть D – множество транзакций, где каждая транзакция T – это набор элементов из $I, T \subseteq I$. Каждая транзакция представляет собой бинарный вектор, где $t[k]=1$, если i_k элемент присутствует в транзакции, иначе $t[k]=0$. Мы говорим, что транзакция T содержит X , некоторый набор элементов из I , если $X \subseteq T$. Ассоциативным правилом называется импликация $X \Rightarrow Y$, где $X \subset I, Y \subset I$ и $X \cap Y = \emptyset$. Правило $X \Rightarrow Y$ имеет поддержку s (support), если $s\%$ транзакций из D , содержат $X \cup Y$, $\text{supp}(X \Rightarrow Y) = \text{supp}(X \cup Y)$. Достоверность правила показывает какова вероятность того, что из X следует Y . Правило $X \Rightarrow Y$ справедливо с достоверностью (confidence) c , если $c\%$ транзакций из D , содержащих X , также содержат Y , $\text{conf}(X \Rightarrow Y) = \text{supp}(X \cup Y)/\text{supp}(X)$.

Покажем на конкретном примере: «75% транзакций, содержащих одну книгу при покупке, также содержат две одновременно. 3% от общего числа всех транзакций содержат обе книги». 75% – это достоверность (confidence) правила, 3% это поддержка (support), или «одна книга» \Rightarrow «две книги» с вероятностью 75%.

Другими словами, целью анализа является установление следующих зависимостей: если в транзакции

встретился некоторый набор элементов X , то на основании этого можно сделать вывод о том, что другой набор элементов Y также же должен появиться в этой транзакции. Установление таких зависимостей дает нам возможность находить очень простые и интуитивно понятные правила.

Алгоритмы поиска ассоциативных правил предназначены для нахождения всех правил $X \Rightarrow Y$, причем поддержка и достоверность этих правил должны быть выше некоторых наперед определенных порогов, называемых соответственно минимальной поддержкой (minsupport) и минимальной достоверностью (minconfidence).

Задача нахождения ассоциативных правил разбивается на две подзадачи:

- нахождение всех наборов элементов, которые удовлетворяют порогу minsupport. Такие наборы элементов называются часто встречающимися;

- генерация правил из наборов элементов, найденных согласно п. 1. с достоверностью, удовлетворяющей порогу minconfidence.

Один из первых алгоритмов, эффективно решающих подобный класс задач – это алгоритм APriori [15]. Кроме этого алгоритма в последнее время был разработан ряд других алгоритмов: DHP [16], Partition [19], DIC [13] и другие [20].

Значения для параметров минимальной поддержки и минимальной достоверности выбираются таким образом, чтобы ограничить количество найденных правил. Если поддержка имеет большое значение, то алгоритмы будут находить правила, хорошо известные аналитикам или настолько очевидные, что нет никакого смысла проводить такой анализ. С другой стороны, низкое значение поддержки ведет к генерации огромного количества правил, что, конечно, требует существенных вычислительных ресурсов. Тем не менее, большинство интересных правил находится именно при низком значении порога поддержки. Хотя слишком низкое значение поддержки ведет к генерации статистически необоснованных правил.

Поиск ассоциативных правил совсем не тривиальная задача, как может показаться на первый взгляд. Одна из проблем – алгоритмическая сложность при нахождении часто встречающихся наборов элементов, так как с ростом числа элементов в I ($|I|$) экспоненциально растет число потенциальных наборов элементов.

Обобщенные ассоциативные правила (Generalized Association Rules).

При поиске ассоциативных правил мы предполагали, что все анализируемые элементы однородны. Возвращаясь к анализу рыночной корзины, это товары, имеющие совершенно одинаковые атрибуты, за исключением названия. Однако не составит большого труда дополнить транзакцию информацией о том, в какую товарную группу входит товар и построить иерархию товаров.

Пусть нам дана база транзакций D и известно в какие группы (таксоны) входят элементы. Тогда можно извлекать из данных правила, связывающие группы с группами, отдельные элементы с группами и так далее.

Например, если «Покупатель купил товар из группы «Безалкогольные напитки», то он купит и товар из группы «Молочные продукты» или «Сок» \Rightarrow «Молочные продукты». Эти правила носят название обобщенных ассоциативных правил.

Определение 2. Обобщенным ассоциативным правилом называется импликация $X \Rightarrow Y$, где $X \subset I$, $Y \subset I$ и $X \cap Y = \emptyset$ и где ни один из элементов, входящих в набор Y , не является предком ни одного элемента, входящего в X . Поддержка и достоверность подсчитываются так же, как и в случае ассоциативных правил (см. *Определение 1*).

Введение дополнительной информации о группировке элементов в виде иерархии даст следующие преимущества:

- это помогает установить ассоциативные правила не только между отдельными элементами, но и между различными уровнями иерархии (группами);

- отдельные элементы могут иметь недостаточную поддержку, но в целом группа может удовлетворять порогу minsupport.

Для нахождения таких правил можно использовать любой из вышеназванных алгоритмов. Для этого каждую транзакцию нужно дополнить всеми предками каждого элемента, входящего в транзакцию. Однако применение «в лоб» этих алгоритмов неизбежно приведет к следующим проблемам:

- элементы на верхних уровнях иерархии стремятся к значительно большим значениям поддержки по сравнению с элементами на нижних уровнях;

- с добавлением в транзакции групп увеличилось количество атрибутов и соответственно размерность входного пространства. Это усложняет задачу, а также ведет к генерации большего количества правил.

Появление избыточных правил, противоречащих определению обобщенного ассоциативного правила, например, «Сок» \Rightarrow «Прохладительные напитки». Очевидно, что практическая ценность такого «открытия» нулевая при 100% достоверности. Следовательно, нужны специальные операторы, удаляющие подобные избыточные правила.

Для нахождения обобщенных ассоциативных правил желательное использование специализированного алгоритма [17], который устраняет вышеописанные проблемы и к тому же работает в 2–5 раз быстрее, чем стандартный APriori.

Группировать элементы можно не только по вхождению в определенную товарную группу, но и по другим характеристикам, например по цене (дешево, дорого), брэнд и так далее.

Численные ассоциативные правила (Quantitative Association Rules).

При поиске ассоциативных правил задача была существенно упрощена. По сути все сводилось к тому, присутствует в транзакции элемент или нет. То есть, если рассматривать случай рыночной корзины, то мы рассматривали два состояния: куплен товар или нет, проигнорировав, например, информацию о том, сколько было куплено, кто купил, характеристики покупателя и так далее. И можно сказать, что рассматривали «булевские» ассоциативные правила. Если взять любую базу данных, каждая транзакция состоит из различных типов данных: числовых, категориальных и так далее. Для обработки таких записей и извлечения численных ассоциативных правил был предложен алгоритм поиска [18].

Пример численного ассоциативного правила:

[Возраст: 30–35] и [Семейное положение: женат]
 ⇒ [Месячный доход: 2000–2500 грн].

Помимо описанных выше ассоциативных правил существуют косвенные ассоциативные правила, ассоциативные правила с отрицанием, временные ассоциативные правила для событий связанных во времени и другие.

Выводы

Как указывалось выше, существуют два вида моделей: предсказательные и описательные. Описательная модель не сможет претендовать на абсолютное знание, но даст аналитику некоторое преимущество уже самим фактом обнаружения альтернативного статистически значимого описания. Даже богатый арсенал классической статистики используется далеко не полностью, не говоря уже о более современных методах нелинейного анализа.

Если конкретные данные хранилища данных характеризуются конечной совокупностью значимых факторов, определяющих анализируемый процесс или развитие процесса, которые могут быть объективно представлены, то в этом случае в задачах многомерного анализа можно успешно применять современные методы статистики с классификатором, построенным по принципу максимального правдоподобия на основе кластерного анализа, метода k_n ближайших соседей и анализа рыночной корзины.

Список литературы: 1. *Арсеньев, С. Б.* Использование технологии анализа данных в интеллектуальных информационных системах. Управление информационными потоками / С. Б. Арсеньев, В. Б. Бритков, Н. А. Маленкова // Сб. тр. Ин-та систем. анализа РАН. – М.: Эдиториал УРСС, 2002. – С. 47–68. 2. *Асеев Г. Г.* Проблема обнаружения нового знания в хранилищах данных методами Knowledge Discovery in Databases / Г.Г. Асеев // Вестник НТУ «ХПИ». –Х.: НТУ «ХПИ», 2006. № 19. –С. 62–70. 3. *Соколов Г.А., Гладких И.М.* Математическая статистика: учеб. / Г.А. Соколов, И.М. Гладких - М: Экзамен, 2004. - 432 с. 4. *Елисеева И.И., Егорова И.И., Курьшьева С.В. и др.* Статистика: учеб. / Под ред. И.И. Ели-

сеевой. - М: ТК Велби / Проспект, 2004. - 448 с. 5. Статистика: учеб. / Под ред. В.С. Мхитаряна. - М: Экономист, 2005. - 671 с. 6. *Орлов А.И.* Прикладная статистика: учеб. / Орлов А.И. - М.: Экзамен. 2006. –452 с. 7. *Хастингс Н., Пикок Дж.* Справочник по статистическим распределениям / Н. Хастингс, Дж. Пикок. - М.: Статистика, 1990. –480 с. 8. *Асеев Г.Г.* Электронный документооборот: учеб. / Г.Г. Асеев. –К.: Кондор, 2007. –500с. 9. Факторный, дискриминантный и кластерный анализ / Дж.-О. Ким, Ч. У. Мьюллер, У. Р. Клекка и др. –М.: Финансы и статистика, 2003. 10. *Дюран Б.* Кластерный анализ / Б. Дюран, П. Оделл. –М.: Статистика, 2007. 11. *Елисеева И. И.* Группировка, корреляция, распознавание образов / И. И. Елисеева, В. О. Рукавишников. –М.: Статистика, 1999. 12. *Волошин Г. Я.* Методы распознавания образов (конспект лекций) [Электронный ресурс] / Г. Я. Волошин. –Режим доступа: <http://disser.h10.ru/raspobraz.html>. –Загл. с экрана. 13. Dynamic Itemset Counting and Implication Rules for Market Basket Data / S. Brin etc. // In Proc. ACM SIGMOD Int'l Conf. Management of Data. –NY: ACM Press, 1999. 14. *Шахиди А.* Введение в анализ ассоциативных правил [Электронный ресурс] / А. Шахиди. – Режим доступа: <http://www.basegroup.ru/rules/intro.html>. –Загл. с экрана. 15. *Agrawal R.* Fast Discovery of Association Rules / R. Agrawal, R. Srikant // In Proc. of the 20th International Conference on VLDB. – Santiago, Chile, September 2004. 16. *Savasere E.* An Efficient Algorithm for Mining Association Rules in Large Databases / E. Savasere, E. Omiecinski, S. Navathe // In Proc. 21st Int'l Conf. Very Large Data Bases. –San Francisco: Morgan Kaufmann, 2005. 17. *Srikant R.* Mining Generalized Association Rules / R. Srikant, R. Agrawal // In Proc. of the 21th International Conference on VLDB. –Zurich, Switzerland, 1995. 18. *Srikant R.* Mining quantitative association rules in large relational tables / R. Srikant, R. Agrawal // In Proceedings of the ACM SIGMOD Conf. on Management of Data. –Montreal, Canada, June, 2006. 19. *Park J. S.* An Effective HashBased Algorithm for Mining Association Rules / J. S. Park, M.-S Chen, S. Y. Philip // In Proc. ACM SIGMOD Int'l Conf. Management of Data. –NY: ACM Press, 2005. 20. *Bremermann H. J.* Global properties of evolution processes. Natural automata and useful simulations / H. J. Bremermann, J. Roghson, S. Salaff. –London: Macmillan, 2006. –P. 3–42.

Поступила в редколлегию 29.01.2009

УДК 519.237.8.:004.8

Методи інтелектуального аналізу даних в електронних сховищах / Г. Г. Асеев // Біоніка інтелекту: наук.-техн. журнал. – 2009. – №1(70). – С. 28–33.

Стаття присвячена проблемі дослідження і виявлення «машиною» у сирих даних схованих структур або залежностей, якщо конкретні дані сховища даних характеризуються кінцевою сукупністю значущих факторів, що обумовлюють аналізований процес чи розвиток процесу, методами кластерного аналізу, асоціативних правил та іншими.

Бібліогр.: 20 найм.

UDC 519.237.8.:004.8

Methods of intellectual analysis of data in electronic warehouses / G. G. Aseyev // Bionics of Intelligence: Sci. Mag. –2009. – №1(70). – P. 28–33.

An unprecedented increase of complex computerization of corporate organizations is being observed at present, and at that the development of automated systems of document flow is of great importance. The paper deals with up-to-date methods of statistics for development of classification tasks and multivariate data analysis in electronic document warehouses.

Ref.: 20 item.

УДК 004.91



САМООБУЧАЮЩАЯСЯ НЕЙРО-ФАЗЗИ СИСТЕМА ДЛЯ АДАПТИВНОЙ КЛАСТЕРИЗАЦИИ ТЕКСТОВЫХ ДОКУМЕНТОВ

Е. В. Бодянский¹, В. В. Волкова², Б. В. Колчигин³

¹ХНУРЭ, Харьков, Украина, bodya@kture.kharkov.ua

²ХНУРЭ, Харьков, Украина, volkova@kture.kharkov.ua

³ХНУРЭ, Харьков, Украина, quasimail@gmail.com

Предложена самообучающаяся нейро-фаззи система, предназначенная для кластеризации в реальном времени текстовых документов, последовательно поступающих на обработку. Система образована двумя самоорганизующимися картами Кохонена, обучаемыми с помощью рекуррентных алгоритмов, реализующих вероятностный и возможностный подходы к нечёткой кластеризации. Показано, что введённые алгоритмы являются обобщением правила самообучения Кохонена на основе принципа «победитель получает больше».

КЛАСТЕРИЗАЦИЯ, КАРТЫ КОХОНЕНА, НЕЙРО-ФАЗЗИ СИСТЕМЫ, НЕЧЁТКАЯ ЛОГИКА, ТЕОРИЯ ВОЗМОЖНОСТЕЙ

Введение

В общей проблеме интеллектуального анализа данных – Data Mining, Exploratory Data Analysis и, особенно, Web-Mining достаточно часто возникает задача поиска и классификации информации, содержащейся в текстовых документах, количество которых в Internet практически неограниченно и постоянно увеличивается. Фактически речь идет об очень больших и непрерывно растущих в реальном времени базах данных, образованных, как правило, не связанными между собой текстами самого различного содержания и происхождения, поиск в которых также должен производиться в on-line режиме.

В настоящее время существует достаточно много подходов к решению этой задачи, однако, большинство из них связано с использованием человеческого интеллекта, а следовательно, и с высоким уровнем субъективизма. В связи с этим представляется перспективным использование методов искусственного и вычислительного интеллекта для решения этой задачи в автоматическом режиме без участия человека. Среди таких методов высокую эффективность продемонстрировали искусственные нейронные сети и, прежде всего, самоорганизующиеся карты Т. Кохонена (SOM) [1], положенные в основу систем автоматической классификации больших массивов документов WEBSOM [2,3] и WEBSOM2 [4]. Процесс настройки этих нейросетей реализуется в режиме самообучения на основе принципов «победитель получает все» (WTA) или «победитель получает больше» (WTM), при этом априори предполагается, что структура обрабатываемых данных такова, что образуемые ими кластеры взаимно не пересекаются, то есть в процессе обучения сети теоретически может быть построена разделяющая гиперповерхность, четко разграничивающая разные классы.

Вместе с тем, при обработке реальных данных часто возникает ситуация, когда один образ-документ принадлежит разным классам, а сами эти классы взаимно пересекаются (перекрываются) [4]. В рамках тра-

диционных самоорганизующихся карт это обстоятельство никак не учитывается, однако может быть рассмотрено с позиций нечеткого кластерного анализа, который к настоящему времени также получил достаточное развитие и распространение [5,6].

Представляется естественным объединить простоту и быстрдействие самоорганизующихся карт Кохонена с возможностью работы в условиях взаимно перекрывающихся классов.

В связи с этим, в настоящей работе предлагается в качестве развития SOM адаптивная нечеткая самоорганизующаяся нейронная сеть, настраиваемая с помощью рекуррентного алгоритма самообучения, являющегося обобщением правила обучения Кохонена и позволяющая находить в реальном времени не только прототипы (центроиды) формируемых кластеров, но и оценивать уровни принадлежности каждого вновь поступившего образа к конкретному кластеру.

1. Адаптивная нечеткая самоорганизующаяся нейронная сеть

Архитектура рассматриваемой нечеткой нейронной сети приведена на рис. 1 и содержит единственный слой нейронов $N_i, i = 1, 2, \dots, p$, в общем случае отличающихся от традиционных адаптивных линейных ассоциаторов, образующих SOM Кохонена.

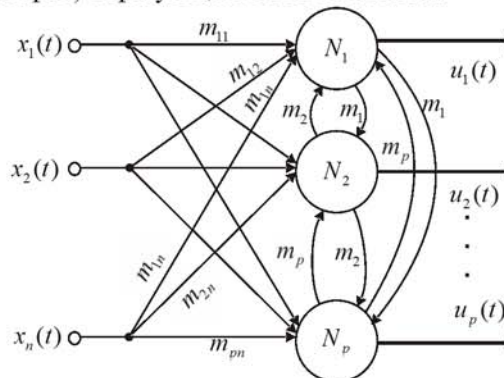


Рис. 1. Адаптивная нечеткая самоорганизующаяся нейронная сеть

На рецепторный слой сети последовательно подаются образы, подлежащие кластеризации, в виде $(n \times 1)$ -векторов признаков документа $x(t) = (x_1(t), x_2(t), \dots, x_n(t))^T$, где $t = 1, 2, \dots, V$ имеет смысл номера образа в обучающей выборке, или текущего дискретного времени. При этом сами векторы признаков $x(t)$ формируются на основе усеченных гистограмм частот появления отдельных слов в обрабатываемых текстах [2–4].

Настраиваемые синаптические веса $m_{ij}, i = 1, 2, \dots, p; j = 1, 2, \dots, n$ определяют координаты центроидов p взаимно перекрывающихся кластеров $m_i(t) = (m_{i1}(t), m_{i2}(t), \dots, m_{in}(t))^T$, а выходом сети, в отличие от SOM, выходной сигнал которой определяется только нейроном-победителем, является $(p \times 1)$ -вектор $u(t) = (u_1(t), u_2(t), \dots, u_p(t))^T$, определяющий уровень принадлежности образа $x(t)$ к каждому из p формируемых кластеров и вычисляемый нейронами N_i . По латеральным связям нейроны обмениваются координатами $m_i(t)$, необходимыми для вычисления принадлежностей $u_i(t)$.

2. Адаптивный вероятностный алгоритм самообучения

В основе самообучения лежит вероятностный алгоритм кластеризации, основанный на оптимизации целевой функции вида [7]:

$$E(u_i, m_i) = \sum_{t=1}^V \sum_{i=1}^p u_i^\beta(t) \|x(t) - m_i\|^2 \quad (1)$$

при ограничениях

$$\sum_{i=1}^p u_i(t) = 1, t = 1, 2, \dots, V, \quad (2)$$

$$0 \leq \sum_{i=1}^p u_i(t) \leq V, i = 1, 2, \dots, p, \quad (3)$$

где $u_i(t) \in [0, 1]$; β – неотрицательный параметр, именуемый «фаззификатором», определяющий нечеткую границу между классами и влияющий на уровень нечеткости в окончательном разбиении данных по кластерам.

Применение стандартного аппарата нелинейного программирования, основанного на неопределенных множителях Лагранжа и решении системы уравнений Куна-Таккера, ведет к известному результату:

$$\left\{ \begin{aligned} m_i^* &= \frac{\sum_{t=1}^V u_i^\beta(t) x(t)}{\sum_{t=1}^V u_i^\beta(t)}, \\ u_i(t) &= \frac{\left(\|x(t) - m_i^*\|^2 \right)^{\frac{1}{1-\beta}}}{\sum_{l=1}^p \left(\|x(t) - m_l^*\|^2 \right)^{\frac{1}{1-\beta}}}, \end{aligned} \right. \quad (4)$$

который при $\beta = 2$ совпадает с популярным FCM-алгоритмом Бездека:

$$\left\{ \begin{aligned} m_i^* &= \frac{\sum_{t=1}^V u_i^2(t) x(t)}{\sum_{t=1}^V u_i^2(t)}, \\ u_i(t) &= \frac{\|x(t) - m_i^*\|^{-2}}{\sum_{l=1}^p \|x(t) - m_l^*\|^{-2}}. \end{aligned} \right. \quad (5)$$

Применение процедуры (5) подразумевает работу данных в пакетном режиме, то есть весь набор данных, подлежащих кластеризации, должен быть задан заранее. Если же информация поступает на обработку последовательно, то есть количество обрабатываемых образов V не фиксировано, реализация данного подхода становится весьма затруднительной с вычислительной точки зрения.

С целью преодоления этого ограничения в [8,9] на основе процедуры нелинейного программирования Эрроу-Гурвица был введен вероятностный рекуррентный алгоритм нечеткой кластеризации вида:

$$\left\{ \begin{aligned} m_i(t+1) &= m_i(t) + \alpha(t) u_i^\beta(t) (x(t+1) - m_i(t)), \\ u_i(t+1) &= \frac{\left(\|x(t+1) - m_i(t+1)\|^2 \right)^{\frac{1}{1-\beta}}}{\sum_{l=1}^p \left(\|x(t+1) - m_l(t+1)\|^2 \right)^{\frac{1}{1-\beta}}}, \end{aligned} \right. \quad (6)$$

где $i = 1, 2, \dots, p$; $\alpha(t)$ – параметр шага поиска, влияющий на скорость сходимости и выбираемый обычно из эмпирических соображений в соответствии с условиями Дворецкого [10].

Анализируя (6), можно заметить, что рассматривая сомножитель $u_i^\beta(t)$ в качестве функции соседства $h_{c(x),i}$, приходим к правилу самообучения Кохонена на основе WTM-принципа:

$$m_i^{WTM}(t+1) = m_i^{WTM}(t) + h_{c(x),i}(t) (x(t+1) - m_i^{WTM}(t)), \quad (7)$$

где $c(x) = \arg \min_i \{ \|x - m_i\| \}$ определяет координаты нейрона-победителя; $h_{c(x),i}(t)$ – колоколообразная функция соседства, аргументом которой есть расстояние в принятой метрике между центроидом нейрона-победителя и нейрона N_i .

Заметим также, что в пакетной форме рекуррентной формуле (7) соответствует выражение [4]:

$$m_i^{*WTM} = \frac{\sum_{x(t) \in V_i} h_{c(x),i} x(t)}{\sum_{x(t) \in V_i} h_{c(x),i}}, \quad (8)$$

где V_i определяет множество всех образов, прототипом которых является m_i^{*WTM} ,

$$\sum_{i=1}^p V_i = V.$$

Как видно, формула (8) структурно совпадает с первым выражением в (4), что опять-таки подтверждает тождество понятий «принадлежности» и «соседства».

Полагая далее в (6) $\beta = 1$, приходим к алгоритму С-средних (hard C-means – HCM), а $\beta = 0$ соответствует стандартному WTA-правилу Кохонена для нейрона-победителя:

$$m_i^{WTA}(t+1) = m_i^{WTA}(t) + \alpha(t)(x(t+1) - m_i^{WTA}(t)). \quad (9)$$

Несложно заметить также, что рекуррентная процедура (9) минимизирует целевую функцию вида:

$$E(m_i) = \sum_{x(t) \in V_i} \|x(t) - m_i^{WTA}\|^2,$$

её прямая оптимизация ведет к обычной оценке среднего арифметического

$$m_i^{*WTA} = \frac{\sum_{x(t) \in V_i} x(t)}{V_i},$$

запись которой в рекуррентной форме – к соотношению

$$m_i^{WTA}(t+1) = m_i^{WTA}(t) + \frac{1}{t+1}(x(t+1) - m_i^{WTA}(t)).$$

Такой выбор параметра шага $\alpha(t)$ согласуется с требованиями стохастической аппроксимации и придает результатам ясный физический смысл.

Таким образом, в окончательном виде адаптивный алгоритм самообучения нечеткой самоорганизующейся сети может быть записан в простой форме:

$$\begin{cases} m_i(t+1) = m_i(t) + \frac{u_i^\beta(t)}{t+1}(x(t+1) - m_i(t)), \\ u_i(t+1) = \frac{(\|x(t+1) - m_i(t+1)\|^2)^{\frac{1}{1-\beta}}}{\sum_{l=1}^m (\|x(t+1) - m_l(t+1)\|^2)^{\frac{1}{1-\beta}}}, \end{cases} \quad (10)$$

объединяющей в себе вычислительную простоту и последовательную обработку кохоненовского самообучения с возможностями нечеткой кластеризации ($i = 1, 2, \dots, p$).

3. Адаптивный возможностный алгоритм самообучения

Известно, что основные недостатки вероятностного подхода связаны с ограничениями

$$\sum_{i=1}^p u_i^{PR}(t) = 1 \quad \forall t = 1, 2, \dots, V,$$

которые приводят к тому, что образ, не принадлежащий ни к одному из уже сформированных кластеров (то есть фактически образующий новый кластер) всё равно будет отнесён с равными уровнями принадлежности ко всем имеющимся «старым» кластерам. Это означает, что вероятностные процедуры нечеткой кластеризации эффективно работают при априори известном числе классов.

Данное обстоятельство привело к появлению возможностных (possibilistic) подходов к нечеткой кластеризации [11,12].

В возможностных алгоритмах классификации целевая функция имеет вид:

$$\begin{aligned} E(u_i, m_i) &= \sum_{t=1}^V \sum_{i=1}^p u_i^\beta(t) \|x(t) - m_i\|^2 + \sum_{i=1}^p \mu_i \sum_{t=1}^V (1 - u_i(t))^\beta = \\ &= \sum_{t=1}^V \sum_{i=1}^p u_i^\beta(t) d^2(x(t), m_i) + \sum_{i=1}^p \mu_i \sum_{t=1}^V (1 - u_i(t))^\beta, \end{aligned} \quad (11)$$

где скалярный параметр $\mu_i > 0$ определяет расстояние, на котором уровень принадлежности принимает значение 0,5, то есть если $d^2(x(t), m_i) = \mu_i$, то $u_i(t) = 0,5$.

Минимизация (11) по $u_i(t)$, m_i и μ_i производится путём решения системы уравнений:

$$\begin{cases} \partial E(u_i, m_i) / \partial u_i = 0, \\ \nabla_{m_i} E(u_i, m_i) = 0, \\ \partial E(u_i, m_i) / \partial \mu_i = 0, \end{cases}$$

которое имеет вид

$$u_i^{POS}(t) = \left(1 + \left(\frac{d^2(x(t), m_i^{*POS})}{\mu_i} \right)^{\frac{1}{\beta-1}} \right)^{-1}, \quad (12)$$

$$m_i^{*POS} = \frac{\sum_{t=1}^V (u_i^{POS}(t))^\beta x(t)}{\sum_{t=1}^V (u_i^{POS}(t))^\beta}, \quad (13)$$

$$\mu_i = \frac{\sum_{t=1}^V (u_i^{POS}(t))^\beta d^2(x(t), m_i^{*POS})}{\sum_{t=1}^V (u_i^{POS}(t))^\beta}. \quad (14)$$

Если сравнивать этот результат с результатами нечеткой вероятностной кластеризации в виде:

$$u_i^{PR}(t) = \frac{(\|x(t) - m_i^{*PR}\|^2)^{\frac{1}{1-\beta}}}{\sum_{l=1}^p (\|x(t) - m_l^{*PR}\|^2)^{\frac{1}{1-\beta}}} = \quad (15)$$

$$\begin{aligned} &= \frac{(d^2(x(t), m_i^{*PR}))^{\frac{1}{1-\beta}}}{\sum_{l=1}^p (d^2(x(t), m_l^{*PR}))^{\frac{1}{1-\beta}}}, \\ m_i^{*PR} &= \frac{\sum_{t=1}^V u_i^\beta(t) x(t)}{\sum_{t=1}^V u_i^\beta(t)}, \end{aligned} \quad (16)$$

можно заметить, что возможностные и вероятностные процедуры очень похожи и переходят друг в друга заменой выражения (12) на формулу (15) и наоборот.

Решение, полученное с помощью вероятностного алгоритма, рекомендуется использовать в качестве начальных условий для возможностного алгоритма [13].

Параметры расстояния μ_i рассчитываются в соответствии с (14) по результатам работы вероятностного алгоритма, то есть на основе $u_i^{PR}(t)$ и $d^2(x(t), m_i^{*PR})$.

Для режима последовательной обработки информации, когда количество обрабатываемых образов заранее неизвестно и они поступают один за другим, вводя локальную модификацию целевой функции (11) в виде

$$E_t(u_i(t), m_i) = \sum_{i=1}^p u_i^\beta(t) d^2(x(t), m_i) + \sum_{i=1}^p \mu_i (1 - u_i(t))^\beta,$$

используя технику рекуррентной нечёткой кластеризации [9], приходим к возможностному рекуррентному алгоритму нечёткой кластеризации:

$$\left\{ \begin{array}{l} m_i^{POS}(t+1) = m_i^{POS}(t) + \alpha(t) (u_i^{POS}(t))^\beta \times \\ \times (x(t+1) - m_i^{POS}(t)), \\ u_i^{POS}(t+1) = \left(1 + \left(\frac{d^2(x(t), m_i^{POS}(t+1))}{\mu_i(t)} \right)^{\frac{1}{\beta-1}} \right)^{-1}, \\ \mu_i(t+1) = \frac{\sum_{p=1}^{t+1} (u_i^{POS}(p))^\beta d^2(x(p), m_i^{POS}(t+1))}{\sum_{p=1}^{t+1} (u_i^{POS}(p))^\beta}. \end{array} \right.$$

Для фаззификатора $\beta = 2$ этот алгоритм приобретает простую форму [14]:

$$\left\{ \begin{array}{l} m_i^{POS}(t+1) = m_i^{POS}(t) + \alpha(t) (u_i^{POS}(t))^2 \times \\ \times (x(t+1) - m_i^{POS}(t)), \quad i = 1, 2, \dots, p, \\ u_i^{POS}(t+1) = \frac{\mu_i}{\mu_i(t) + \|x(t+1) - m_i^{POS}(t+1)\|^2}, \\ \mu_i(t+1) = \frac{\sum_{p=1}^{t+1} (u_i^{POS}(p))^2 \|x(p) - m_i^{POS}(t+1)\|^2}{\sum_{p=1}^{t+1} (u_i^{POS}(p))^2}. \end{array} \right.$$

4. Самообучающаяся нейро-фаззи система

Параллельное применение адаптивных вероятностного и возможностного алгоритмов приводит к объединённой процедуре (для $\beta = 2$):

$$\left\{ \begin{array}{l} m_i^{PR}(t+1) = m_i^{POS}(t) + \alpha(t) (u_i^{POS}(t))^2 \times \\ \times (x(t+1) - m_i^{POS}(t)), \\ u_i^{PR}(t+1) = \frac{\|x(t+1) - m_i^{PR}(t+1)\|^{-2}}{\sum_{l=1}^p \|x(t+1) - m_l^{PR}(t+1)\|^{-2}}, \\ m_i^{POS}(t+1) = m_i^{PR}(t+1) + \alpha(t) (u_i^{POS}(t))^2 \times \\ \times (x(t+1) - m_i^{PR}(t+1)), \\ u_i^{POS}(t+1) = \frac{\mu_i(t)}{\mu_i(t) + \|x(t+1) - m_i^{POS}(t+1)\|^2}, \\ \mu_i(t+1) = \frac{\sum_{p=1}^{t+1} (u_i^{POS}(p))^2 \|x(p) - m_i^{POS}(t+1)\|^2}{\sum_{p=1}^{t+1} (u_i^{POS}(p))^2}. \end{array} \right. \quad (17)$$

Процедура (17) по сути является алгоритмом самообучения нейро-фаззи системы, образованной двумя параллельно работающими самоорганизующимися картами Кохонена и обменивающимися между собой информацией так, как это показано на рисунке:

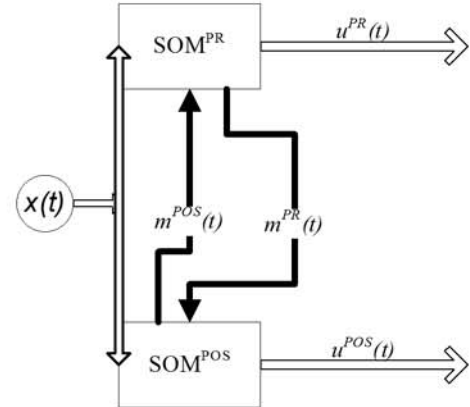


Рис. 2. Самообучающаяся нейро-фаззи система

Признаком правильного восстановления прототипов (а следовательно, и корректной кластеризации) с помощью алгоритма (17) является выполнение неравенства

$$\sum_{l=1}^p d^2(m_l^{PR}(t), m_l^{POS}(t)) \leq \varepsilon,$$

где малый параметр ε определяет приемлемую точность кластеризации.

Как мы уже отмечали, важное достоинство возможностного подхода связано с условием

$$\sum_{i=1}^p u_i^{POS}(t) \neq 1,$$

что позволяет обнаруживать появление новых классов в процессе кластеризации.

Вводя некоторое пороговое значение θ ($\theta \approx 0,1 + 0,2$) и контролируя условие

$$\sum_{i=1}^p u_i^{POS}(t) \leq \theta,$$

можно говорить о появлении нового класса при его выполнении.

Результаты численных экспериментов, проведенных на тестовой выборке, содержащей около ста текстовых документов по трём различным тематикам (Neural Network, Semantic Web, Text Mining), показали, что процедура нечёткой кластеризации обеспечивает более высокое качество классификации по сравнению с традиционными чёткими алгоритмами.

Выводы

Предложены адаптивные алгоритмы самообучения нечёткой самоорганизующейся нейронной сети, предназначенной для кластеризации больших массивов текстовых документов, позволяющие осуществлять в реальном времени нечёткую классификацию данных, последовательно поступающих на обработку. Алгоритмы не содержат свободных параметров, просты в реализации и объединяют в себе достоинства самоорганизующихся карт Кохонена и вероятностных процедур нечёткой кластеризации.

Список литературы: 1. Kohonen T. Self-Organizing Maps / T. Kohonen // Berlin: Springer-Verlag. – 1995. – 362 p. 2. Kaski S. WEBSOM – Self-organizing maps of document collections / S. Kaski, T. Honkela, K. Lagus, T. Kohonen // Neurocomputing. – 1998. – 21. – P. 101–117. 3. Lagus K. WEBSOM for textual data mining / K. Lagus, T. Honkela, S. Kaski, T. Kohonen // Artificial Intelligence Review. – 1999. – 13. – P. 345–364. 4. Kohonen T. Self organization of a massive document collection / T. Kohonen, S. Kaski, K. Lagus, J. Salojärvi, J. Honkela, V. Paatero, A. Saarela // IEEE Trans. on Neural Networks. – 2000. – 11. – P. 574–585. 5. Huppner F. Fuzzy-Klusteranalyse. Verfahren für die Bilderkennung, Klassifikation und Datenanalyse / F. Huppner, F. Klawonn, R. Kruse // Braunschweig: Vieweg. – 1996. – 280 p. 6. Huppner F. Fuzzy Clustering Analysis: Methods for Classification, Data Analysis, and Image Recognition / F. Huppner, F. Klawonn, R. Kruse, T. Runkler // Chichester: John Wiley & Sons Ltd. – 1999. – 289 p. 7. Bezdek J.C. Pattern Recognition with Fuzzy Objective Function Algorithms / J.C. Bezdek // N.Y.: Plenum Press. – 1981. – 272 p. 8. Bodyanskiy Ye. Recursive fuzzy clustering algorithms / Ye. Bodyanskiy, V. Kolodyazhnyi, A. Stephan // Proc. East West Fuzzy Coll. 2002. Zittau / Gurlitz: HS. – 2002. – P. 164–172. 9. Bodyanskiy Ye. Computational intelligence techniques for data analysis / Ye. Bodyanskiy // Lecture Notes in Informatics. – Bonn: GI, 2005. – P. 72. – P. 15–36. 10. Dvoretzky A. On stochastic approximation / A. Dvoretzky // Proc. 3-rd Berkley Symp. Math. Statistics and Probability. – 1956. – 1. – P. 39–55. 11. Krishnapuram R., Keller J. A possibilistic approach to clustering // IEEE Trans. on Fuzzy Systems. – 1993. – 1. – P. 98–110. 12. Krishnapuram R., Keller J. Fuzzy and possibilistic clustering methods for computer vision // Neural Fuzzy Systems. – 1994. – 12. P. 133–159. 13. Klawonn F., Kruse R. Constructing a fuzzy controller from data // Fuzzy Sets and Systems. – 1992. – 85. – P. 117–193. 14. Бодянский Е.В., Волкова В.В., Колчигин Б.В. Нечёткая возможностная кластеризация текстовых документов на основе самоор-

ганизирующей карты Кохонена // Тези доп. II міжнар. науково-практ. конф. «Математичне та програмне забезпечення інтелектуальних систем (MPZIS-2008)». – Днепропетровськ, 2008. – С.49–50.

Поступила в редколлегию 02.02.2009

УДК 004.91

Нейро-фаззі система, що самонавчається, для адаптивної кластеризації текстових документів / Є. В. Бодяньський, В. В. Волкова, Б. В. Колчигін // Біоніка інтелекту: наук.-техн. журнал. – 2009. – №1(70). – С. 34–38.

Запропоновано нейро-фаззі систему, що самонавчається, для кластеризації текстових документів в реальному часі. Для навчання системи введено рекурентні алгоритми, що базуються на ймовірнісному та можливісному підходах до нечіткої кластеризації.

Л.: 2. Бібліогр.: 14 найм.

UDK 004.91

Self-learning neuro-fuzzy system for text documents adaptive clustering / Ye.V. Bodyanskiy, V.V. Volkova, B.V. Kolchigin // Bionics of Intelligence: Sci. Mag. – 2009. – №1(70). – P. 34–38.

In the article self-learning neuro-fuzzy system for text documents adaptive clustering is proposed. For its learning recurrent algorithms based on possibilistic and probabilistic fuzzy clustering approaches are introduced.

Fig.: 2. Ref.: 14 items.

УДК 004.81, 004.93



ИНТЕЛЛЕКТУАЛЬНЫЙ КОНТРОЛЬ ПОЛЬЗОВАТЕЛЕЙ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

А. Л. Ерохин¹, А. В. Кольченко²¹ Харьковский национальный университет внутренних дел, г. Харьков, Украина, ayerokhin@ukr.net² ХНУРЭ, г. Харьков, Украина, akolchenko@mail.ru

Рассматривается решение задачи интеллектуального контроля над пользователем информационной системы на основе предложенного метода мониторинговой верификации.

ИНТЕЛЛЕКТУАЛЬНЫЙ КОНТРОЛЬ, ВЕРИФИКАЦИЯ ПОЛЬЗОВАТЕЛЯ, ВЗАИМОДЕЙСТВИЕ ОПЕРАТОР–КОМПЬЮТЕР

Введение

В настоящее время наиболее распространенными способами контроля пользователя в информационных системах (ИС) являются: аутентификация, активный аудит и верификация (рис. 1).



Рис. 1. Способы контроля пользователей ИС

Контроль пользователей призван обеспечить безопасность ИС. Каждый способ контроля имеет собственную модель взаимодействия пользователя с ИС и ориентирован на решение определенного круга задач. Традиционные способы контроля, к которым относятся аутентификация и активный аудит, не решают задачу проверки истинности пользователя в течение сеанса взаимодействия с ИС.

1. Обзор традиционных способов контроля пользователей ИС

Аутентификация является наиболее распространенным способом проверки подлинности пользователя. Подлинность при аутентификации может быть подтверждена путем предъявления пользователем одной из сущностей:

- нечто, что он знает: пароль, ПИН-код;
- нечто, чем он владеет: аппаратный ключ, карточка;
- нечто, что является частью его самого: отпечаток пальца, радужная оболочка глаза, голос.

В существующих системах контроля пользователя процедура аутентификация проводится, как правило, однократно, в начале сеанса работы пользователя. Образец сущности, предъявляемой пользователем для аутентификации, хранится в базе данных. Предъявляемая в момент аутентификации сущность сравнивается с образцом, в результате чего принимается решение о предоставлении или отклонении доступа пользо-

вателя в информационную систему. Наиболее распространенной формой аутентификации является парольная, когда пользователю в начале сеанса работы необходимо представиться (ввести логин) и подтвердить свою подлинность (ввести пароль).

Особым случаем является биометрическая аутентификация, когда пользователь для проверки подлинности предъявляет нечто, что является частью его самого. Биометрические идентификаторы, в зависимости от степени устойчивости в течение длительного времени, делятся на статические и динамические. Другими названиями указанных идентификаторов являются физиологические и поведенческие признаки соответственно. К физиологическим биометрическим идентификаторам относят радужную оболочку глаза, отпечаток пальца, лицо, кисть, сетчатку глаза. Поведенческие биометрические идентификаторы основаны на данных, получаемых путём измерения действий человека. Характерной чертой для поведенческих параметров является их протяжённость во времени. К поведенческим биометрическим признакам относят клавиатурный почерк, подпись, голос, жесты, походка. Результаты биометрической аутентификации подвержены ошибкам первого и второго рода. Ошибки первого рода происходят, когда ненадлежащий пользователь успешно проходит аутентификацию. Ошибки второго рода происходят в случае отказа в аутентификации надлежащего пользователя.

Системы аутентификации работают в «явном» режиме, то есть открыто для пользователя. Добавление нового пользователя в информационную систему требует внесения в базу данных системы аутентификации данных в виде эталонного образца, предъявляемого пользователем объекта для проверки подлинности. Такой эталонный образец и необходимые сведения о его владельце называются учетной записью пользователя.

Таким образом, системам аутентификации присущи следующие характеристики:

- однократная работа в начале сеанса взаимодействия;
- работа с «явным» режиме, с прерыванием пользователя;

- наличие образца объекта для сличения – учетная запись пользователя;
- механизм регистрации предъявляемого пользователем объекта;
- механизм сравнения эталонного образца и предъявленного пользователем объекта.

Системы **активного аудита** обеспечивают оперативное *обнаружение подозрительной активности* пользователя или компонента информационной системы – злоумышленное или нетипичное. Системы активного аудита также могут предоставлять средства для автоматического реагирования на нарушения. Системы активного аудита строятся в архитектуре менеджер/агент. Агент занимается извлечением регистрационной информации. К такой информации относятся выполняемые пользователем команды: запуск приложений, работа с объектами файловой системы, работа с объектами операционной системы. Анализ и принятие решений выполняет менеджер [1, 2]. Анализ поведения пользователей в системах активного аудита выполняется с помощью набора правил – профилей пользователей. Следует отметить, что профили пользователей описываются с точки зрения объектов взаимодействия, то есть, по сути, являются профилями объектов. Профиль объекта отражает принятую политику безопасности как допустимый набор действий с ним. Например, для данной группы пользователей некоторый объект файловой системы может быть доступен только для чтения, и факт попытки выполнения записи передается агентом менеджеру для анализа и принятия последующего решения. Результатом такого решения может быть обнаружение нарушения и отказ в обслуживании запроса пользователя, либо разрешение на его обработку.

Поскольку системы активного аудита ориентированы на объекты взаимодействия, наличие информации о каждом пользователе системы в базе данных не является необходимым. Следовательно, новый пользователь может приступать к работе с информационной системой без предварительной настройки системы активного аудита для этого пользователя.

Важным является тот факт, что системы активного аудита работают с «скрытым» режиме, то есть незаметно для пользователя. При этом работа пользователя в случае штатного поведения не прерывается.

Таким образом, для систем активного аудита при-
сущи:

- работа в течение всего сеанса взаимодействия;
- работа с «скрытым» режиме, без прерывания пользователя;
- наличие профиля объекта взаимодействия;
- механизм регистрации первичной информации о поведении пользователя;
- механизм детектирования нетипичного поведения пользователя;
- автоматическое реагирование на нарушения – опционально.

Описанные способы контроля пользователя имеют ряд недостатков. Так, в случае предоставления ненадлежащим пользователем данных чужой учетной записи, процедура аутентификации выполняется успешно. Более надежными в таком случае представляются способы биометрической аутентификации. Средства активного аудита ориентированы на объекты взаимодействия и не предоставляют возможностей различения надлежащих и ненадлежащих пользователей. Кроме того, ни один из указанных способов контроля не предполагает возможности подмены пользователя в течение сеанса взаимодействия с информационной системой.

Таким образом, актуальной задачей является проверка истинности пользователя в течение сеанса работы, когда аутентификация успешно выполнена. Несмотря на некоторое сходство с задачей проверки подлинности пользователя, решаемой с помощью аутентификации, решение задачи контроля пользователя в течение всего сеанса работы с информационной системой требует разработки новых методов и средств.

2. Разработка метода мониторинговой верификации

Под **верификацией пользователя** будем понимать *установление факта работы надлежащего пользователя* в течение сеанса работы (рис. 2). Для указания различия с процедурой аутентификации считаем наиболее приемлемым (обоснованным) использование термина «верификация» (позднелат. *verificatio* – доказательство, подтверждение, от лат. *verus* – истинный и *facio* – делаю) – проверка, контроль, удостоверение в подлинности, установление истинности.



Рис. 2. Аутентификация и верификация пользователя

Таким образом, верификация пользователя позволяет подтвердить истинность пользователя или детектировать его подмену. Проверку подлинности пользователя можно выполнять в режиме аутентификации, путем предъявления начального диалога аутентификации – повторная аутентификация. Однако такой подход вызывает прерывание выполнения текущей деятельности пользователя, что является негативным фактором [ссылка]. Другим подходом является выполнение процедуры верификации в скрытом режиме, незаметно для пользователя, без прерывания его действий.

Мониторинговой верификацией пользователя будем называть периодически выполняющиеся в течение сеанса взаимодействия процедуры подтверждения истинности пользователя. Очевидно, мониторинговый режим верификации позволит решить задачу проверки истинности пользователя в течение всего сеанса работы.

Мониторинговый режим верификации подразумевает выполнение процедуры верификации с некоторой периодичностью. В разных типах решаемых пользователем задач предлагаются следующие режимы работы системы мониторинговой верификации:

- синхронная верификация;
- верификация по требованию;
- комбинированный режим.

В синхронном режиме верификация выполняется без привязки к действиям пользователя, через заданные равные временные промежутки. Указанный режим работы является удовлетворительным для таких типов задач, где нет разделения потока действий по приоритетам (безопасности). Обнаружение подмены пользователя блокирует дальнейшее взаимодействие. В таком случае временной промежуток выполнения верификации является одним из параметров системы мониторинговой верификации пользователя.

Верификация по требованию применима к задачам, где в потоке пользовательских действий возможно выделение критических действий. Для выполнения критического действия необходимо удостовериться в истинности пользователя, что и выполняет система мониторинговой верификации. В случае обнаружения подмены пользователя обслуживание пользователя прекращается. Множество критических действий пользователя является параметром системы мониторинговой верификации, работающей в режиме верификации по требованию.

Комбинированный режим верификации использует указанные выше режимы в комплексе.

Представляется, что использование мониторинговой верификации совместно с аутентификацией и активным аудитом обеспечит более полное и лучшее решение задачи контроля пользователя при работе в информационных системах. Сравнение способов контроля пользователей приведено в таблице.

Таблица

Метод	Задача	Процедура
Аутентификация	Установление подлинности пользователя	Однократно, открыто для пользователя, в начале сеанса. Подтверждает или отвергает подлинность пользователя
Мониторинговая верификация		Периодически, в мониторинговом режиме, скрытно от пользователя. Подтверждает или отвергает подлинность пользователя
Активный аудит	Обнаружение подозрительной активности	Периодически, скрытно от пользователя. Не устанавливает подлинность пользователя. Разрешает или отклоняет операцию

Для разработки системы мониторинговой верификации пользователя информационной системы необходимо решить задачу выбора объекта, предъявляемого пользователем системе для установления своей подлинности. При этом необходимо, чтобы работа системы происходила в скрытном режиме, без прерывания деятельности пользователя.

Для решения поставленной задачи выполним анализ человеко-машинного взаимодействия в информационной системе.

Большинство задач в информационной системе решается пользователем с помощью прикладных программ, имеющих графический пользовательский интерфейс. Взаимодействие пользователя-оператора с компьютером происходит в виде диалога (рис. 3). Диалог состоит из набора реплик с обеих сторон и продолжается до решения задачи пользователем, либо завершается отказом от выполнения задачи. В любом случае, взаимодействие имеет темпоральный характер (то есть продолжается в течение некоторого отрезка времени).



Рис. 3. Схема взаимодействия оператор – компьютер

Взаимодействие пользователя с компьютером может происходить в разных модальностях: кинестетической, аудиальной, тактильной. Одной их характеристик пользователя во время взаимодействия является его двигательная активность. Двигательная активность включает воздействия пользователя на наиболее распространенные в настоящее время устройства ввода: клавиатуру и мышь. В связи с этим возникает идея использования двигательных характеристик движений мыши в качестве объекта, предъявляемого для установления истинности пользователя в системе мониторинговой верификации. Предлагаемый подход имеет следующие достоинства:

- отсутствие необходимости в специальной аппаратуре;
- сбор данных о взаимодействии производится в скрытном порядке, без нарушения деятельности пользователя;
- возможность работы в течение всего сеанса взаимодействия.

Данные исследований в области психофизиологии и инженерной психологии подтверждают возможность использования предложенного подхода:

1. Психомоторика – одно из сложных индивидуальных свойств человека, связанных не только с его половозрастными и нейродинамическими характеристиками, особенностями билатерального регулирования и темперамента, но и с рядом личностных и субъектно-деятельностных качеств [3, 4].

2. Многие виды операторской деятельности складываются из совокупности различных сенсомоторных реакций.

3. Время реакции может использоваться как один из показателей психофизиологического состояния оператора или готовности его к выполнению определенного вида деятельности.

4. Время реакции часто используется как индикатор при инженерно-психологических измерениях и исследованиях [5].

Некоторые аналоги предлагаемого подхода можно увидеть в биометрических методах аутентификации. Так, данные двигательной активности используются при аутентификации по клавиатурному подерку, динамике росписи, жестам, походке. В то же время существенным отличием от биометрической аутентификации является то, что нами решается иная задача: задача верификации пользователя.

Представляется возможным также использование сенсомоторных характеристик пользователя для решения ряда задач, среди которых:

- определение психофизиологического состояния оператора;
- оценка компетентности пользователя при работе с компьютером;
- определение степени трудности задачи, решаемой пользователем с помощью компьютера;
- анализ удобства использования интерфейсов приложений.

3. Разработка модели взаимодействия пользователя с информационной системой

Для разработки метода мониторинговой верификации пользователя требуется построить модель взаимодействия пользователя с информационной системой.

Анализ процесса взаимодействия, реализующегося в форме диалога между пользователем и интерфейсом информационной системы, позволяет выделить сущности, которые требуют формального описания. Такими сущностями являются *Пользователь* и *Интерфейс* информационной системы.

Активность *Пользователя*, наблюдаемая вовне, выражается в совершении двигательных действий с помощью манипулятора. Эти действия являются средством коммуникации с *Интерфейсом*. Представим модель *Пользователя* в виде черного ящика. Входами являются сигналы, посылаемые *Интерфейсом*, выходами – двигательные акты, выполняющиеся с помощью манипулятора. Возможными способами взаимодействия пользователя с помощью манипулятора явля-

ются движения в двумерном пространстве, шелчки кнопками: левой, средней, правой, прокрутка колеса мыши, перетаскивание.

Другим элементом модели взаимодействия является *Интерфейс* информационной системы. *Интерфейс* принимает управляющие воздействия *Пользователя* и генерирует ответные сигналы. Следует заметить, что *Интерфейс* является системой, состоящей из множества управляющих элементов: кнопки, поля, списки и другие элементы управления. Именно на них и совершает воздействия *Пользователь*.

Рассмотрим интеллектуальный интерфейс ИС в виде системы отношений

$$S \subset G \times Z, \quad (1)$$

где G – множество элементов интерфейса (ЭИ), выход системы;

Z – множество координат ЭИ на плоскости, вход системы.

В классах эквивалентности S выделим подмножества выходов и входов в виде объединений их элементов:

$$\begin{cases} (Z)_s = \bigcup_{s \in S} (z_i)_s, \\ (G)_s = \bigcup_{s \in S} (g_i)_s, \end{cases} \quad (2)$$

где $G = \times \{V_i : i \in I_G\}$ – выходной объект;

$Z = \times \{V_i : i \in I_Z\}$ – входной объект системы S , заданные декартовым произведением.

Каждой паре чисел $(q, r)_G$ выхода соответствует пара на входе $(i, j)_Z$

$$G \times Z = (g_{qr} \times z_{ij}; \quad \forall i, j \in I_Z \quad \forall q, r \in I_G). \quad (3)$$

Зададим формальную модель интерфейсной части ИС в виде объединения единичных образов ЭИ, каждый из которых представляется декартовым произведением множества элементов интерфейса G и множества координат на плоскости интерфейсной части Z :

$$I = \bigcup_{k=1}^p (G \times Z)_k = \bigcup_{k=1}^p \{ \langle g, x \rangle \mid g \in 1..p, z \in 1..p \}, \quad (4)$$

где k – индекс единичного образа ЭИ, $k = 1, \dots, p$;

p – количество ЭИ в ИС.

Интеллектуальная активность человека при взаимодействии с ИС может быть представлена множеством (набором) единичных образов с записями посещений элементов интерфейсной части.

На основе событийно-ориентированной парадигмы рассмотрим модели активности человека-пользователя ИС. Под событием E будем понимать элементарное действие человека-пользователя, связанное с его сенсомоторной активностью при взаимодействии с ИС. Взаимодействие с ИС – это посещение множества элементов интерфейса G .

Пусть E_i – асинхронное событие, а $i = 1, \dots, n$ – количество событий в одном сеансе взаимодействия с ИС. Введем:

$ETRi$ – источник сенсомоторного события:

$$ETRi = \{ETRxi, ETRyi\}, \quad (5)$$

где $ETRxi, ETRyi$ – пиксельные координаты события в интерфейсной зоне ИС;

2) $ETPi$ – множество типов событий в интерфейсной зоне:

$$ETPi = \{MM, MC, MD, KP, SS\}, \quad (6)$$

где MM – перемещение мыши,

MC – нажатие кнопки мыши,

MD – двойное нажатие кнопки мыши,

KP – нажатие клавиши на клавиатуре,

SS – изменение положения полосы прокрутки;

3) ELi – множество объектов-получателей (подписчиков событий в ИС):

$$ELi = \{g_i | g_i \in G\}, \quad (7)$$

где g_i – элемент интерфейса, с которым связана функция–обработчик соответствующего события;

4) EHi – множество обработчиков событий – функций, код которых выполняется при срабатывании события.

Введем понятие единичного образа (паттерна) сенсомоторной активности человека-пользователя в виде:

$$P_i = \bigcup_{j=1}^m \langle L_j, R_j, M_j, C_j, D_j, K_j, S_j \rangle, \quad (8)$$

где j – индекс итерации дообучения SMP, $j = 1, \dots, m$;

m – количество событий посещения ЭИ во время сеанса взаимодействия с ИС;

L_j – множество координат элемента интерфейса;

R_j – множество размеров (площадей) элемента интерфейса;

M_j, C_j, D_j – множества координат манипулятора «мышь»;

K_j – множество кодов нажатых клавиш управления;

S_j – множество координат положений полос прокрутки окна интерфейса.

Тогда под сенсомоторным портретом человека-пользователя ИС (SMP) будем понимать

$$SMP = \bigcap_{i=1}^n \bigcup_{j=1}^m \langle L, R, M, C, D, K, S \rangle_i, \quad (9)$$

где i – индекс итерации уточнения (дообучения) SMP, $i = 1, \dots, n$;

n – минимально допустимое доверительное количество итераций формирования SMP.

При выполнении операции уточнения (дообучения) SMP на первых $n-k$ шагах процедуры начальной верификации вычисляется симметрическая разность множеств.

Таким образом, построена модель интеллектуальной деятельности пользователя, которая применяется для выполнения процедуры мониторинговой верификации.

Выводы

Таким образом, в статье предложен оригинальный метод мониторинговой верификации пользователя ИС, который основывается на модели интеллектуальной деятельности пользователя в процессе взаимодействия с информационной системой. В отличие от известных методов аутентификации и активного аудита разработанный метод позволяет проводить неинвазивный скрытый контроль активности пользователя. Дальнейшими направлениями исследований являются исследование свойств модели, что позволит уточнить метод мониторинговой верификации пользователя ИС.

Список литературы:

1. Стандарты 2006: Стандарты информационной безопасности. Курс лекций: Учеб. пособие.– 2-е изд. / В. А. Галатенко / Под ред. акад. РАН В.Б. Бетелина.– М.: Интернет-университет Информационных Технологий, 2006.– 264 с.
2. Бетелин В.Б., Галатенко А.В., Галатенко В.А., Кобзарь М.Т., Сидак А.А. Классификация средств активного аудита в терминах «Общих критериев» // Сб. «Информационная безопасность. Инструментальные средства программирования. Базы данных» / Под ред. В.Б. Бетелина.– М.: НИИСИ РАН, 2001.– С.4–26.
3. Бернштейн Н.А. Очерки по физиологии движений и физиологии активности.– М.: Медицина, 1966.– 349 с.
4. Коренкова Н. Е., Олейник Ю. Н. Психомоторика в структуре интегральной индивидуальности человека // Психологический журнал.– 2006.– № 1.– С. 54–66.
5. Психофизиология оператора в системах человек-машина / К.А. Иванов-Муромский, О.Н. Лукьянова, В.А. Черноморец и др. – Киев: Наукова думка, 1980.– 344 с.

Поступила в редколлегию 06.02.2006

УДК 004.81, 004.93

Интеллектуальний контроль користувачів в інформаційних системах / А. Л.Єрохін, О. В. Кольченко // Біоніка інтелекту: наук.-техн. журнал.– 2009.– №1(70).– С. 39–43.

В роботі запропоновано метод моніторингової верифікації для встановлення справжності користувача інформаційних систем. Метод базується на використанні даних сенсомоторної активності користувача під час сеансу роботи. У роботі побудовано формальну модель інтелектуального інтерфейсу та модель взаємодії користувача з інформаційною системою, що використовується для моніторингової верифікації користувача.

Табл.:1. Іл.:3. Бібліогр.: 2 найм.

UDC 004.81, 004.93

Intelligent user control in information systems / A. L.Yerokhin, A. V. Kolchenko // Bionics of Intelligence: Sci. Mag. – 2009. – №1(70).– P. 39–43.

The monitoring verification method for user genuine in information systems is offered. The method based on the user sensomotoric activity during interaction. The formal models of intelligent interface and user interaction which is used for monitoring verification are built.

Tab.:1. Fig.:3. Ref.: 2 items.

УДК 681.32



Представление имитационных моделей дискретно-событийных систем с помощью CASE средства E-NetSim

В.Г. Иванов¹, И.В. Шевченко²^{1,2} ХНУРЭ, г. Харьков, Украина

Рассмотрены существующие программно-технические комплексы, реализующие анализ исследуемой системы при помощи теории оценочных сетей. Описана система имитационного моделирования E-NetSim. Рассматриваемое CASE средство предназначено для создания и исследования имитационных моделей дискретно-событийных систем, представленных в виде оценочных сетей Петри. Показан механизм взаимодействия пользователя с системой при создании моделей и задании ее основных параметров. Показана возможность простого, интуитивно понятного описания сложных функциональных и логических связей в создаваемой модели.

ПРОГРАММНО-ТЕХНИЧЕСКИЙ КОМПЛЕКС, CASE-СРЕДСТВО, СЕТИ ПЕТРИ, ОЦЕНОЧНЫЕ СЕТИ, МОДЕЛИРУЮЩАЯ МОЩНОСТЬ, ДИАЛоговый ИНТЕРФЕЙС, СИСТЕМА ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ

Введение

Создание сложных распределенных информационно-технических систем невозможно без глубокого изучения процессов взаимодействия между отдельными составляющими в таких системах. Характерными особенностями таких систем являются глубокий параллелизм, взаимная неопределенность состояний, отсутствие однозначной зависимости между событиями и условиями их порождающими, многообразие вариантов возможного развития каждого процесса и системы в целом.

Использование сетей Петри, особенно их расширений E-сетей, по мнению группы разработчиков, позволяет приблизиться к созданию инструментального средства, позволяющего с единых позиций создавать комплексное описание системы, проводить ее исследование и выработать рекомендации по дальнейшему развитию.

В настоящее время разработано множество программных средств, позволяющих производить анализ сложных систем, базируясь на теории сетей Петри и их расширениях, однако для пользования ними разработчик помимо знания самой методологии СП должен разбираться в области программирования, что в значительной мере ограничивает область их использования. Таким образом, задача разработки функционально мощного и в тоже время предельно простого и понятного программно-технического комплекса, позволяющего проводить анализ системы при помощи E-сетей, является актуальной.

1. Анализ последних исследований

Анализ существующего программного обеспечения, позволяющего проводить исследования системы с использованием теории оценочных сетей, показал, что на данный момент не существует относительно про-

стого для пользователя программного средства [1-3]. Существует множество интересных средств, как правило, разработанных в 80х-90х годах, функционально мощных, однако не обладающих гибким графическим интерфейсом и уже давно морально устаревших. Современное программное обеспечение носит скорее демонстративный характер, единственным применением которого может быть визуализация принципов работы оценочных сетей в образовательных целях для большей наглядности. Другими распространенными средствами имитационного моделирования в наше время являются языки имитационного моделирования (ЯИМ)[4], такие как: GPSS, Simula, SimScript проч. Также был рассмотрен программный комплекс MATLAB [5,6], в основу которого положен язык высокого уровня для выполнения технических и научных вычислений. Обладая высокой моделирующей мощностью, функциональным, математическим аппаратом, комплекс является одним из лучших средств исследования различных систем. Однако наряду с достоинствами комплекса одним из очевидных недостатков является сложность и длительность изучения.

Целью данной работы является описание моделирующей мощности разработанного ранее программно-технического комплекса E-NetSim, раскрывающего имитационные возможности математического аппарата E-сетей.

2. Особенности описания моделей с помощью оценочных сетей

Мобильная инструментальная система имитационного моделирования предназначена для создания и исследования имитационных моделей дискретно-событийных систем. В настоящее время разработано множество программно-технических комплексов, позволяющих производить анализ сложных систем, ба-

зируясь на теории сетей Петри и их расширений, однако для пользования ними разработчик помимо знания самой методологии оценочных сетей (ОС) должен разбираться в области программирования, что в значительной мере ограничивает область их использования. Основным условием использования рассматриваемой системы имитационного моделирования является то, что пользователь должен быть квалифицирован в области исследуемой системы и быть знакомым с принципами построения моделей при помощи оценочных сетей.

Модель создается на базе основных элементов оценочных сетей: *позиции(условия), переходы (действия), дуги, маркеры.*

Состояние системы отображается в модели совокупностью выполнившихся условий. Выполнение условия моделируется маркировкой соответствующей позиции. Происходящие в системе события в модели представляются с помощью переходов, которые срабатывают под воздействием одного или нескольких условий. Основными параметрами перехода являются: *время работы, условия для срабатывания, функция преобразования, тип перехода.*

Время работы перехода – количество тактов моделирования, которые необходимы для переноса маркера из входных позиций перехода в выходные.

Функция преобразования является математической зависимостью, выполняющей преобразования над заданными атрибутами маркеров, при срабатывании перехода.

Все переходы, используемые в оценочных сетях, делятся на два вида: *безусловные (переходы типа T, F, J) и условные (переходы типа X и Y).*

Дуги служат для образования логических связей типа «предусловие» и «постусловие» для переходов.

Маркеры, которые перемещаются по сети, кроме отображения выполнения условия могут отождествляться с объектами, существующими в исследуемой системе, и нести в себе один или более параметров, каждый из которых может принимать дискретный набор значений (атрибутов). В предлагаемой системе моделирования предусмотрена возможность гибкого задания количества атрибутов маркеров и их значений.

Существует несколько подходов формального представления E-сети. Будем использовать следующее представление:

$$E = \langle (P, V, R), T, I(T), O(T), Z, V, Q, \Psi, M \rangle,$$

где:

$P = \{P_i\}$ – конечное непустое множество позиций;

$V = \{b_k\} \in P$ – множество периферийных, не внутренних: $I(b_k) = 0$ либо $O(b_k) = 0$ и $I(b_k) \cup (b_k) \neq 0$ позиций, которые используются в сети для определения связей с внешней средой;

$R = \{r_m\} \subset P$ – множество решающих позиций, на которых определяется множество Q;

$T = \{t_n\}$ – непустое конечное множество переходов;
 $I(T), O(T)$ – функция связи переходов и позиций соответственно по входам и выходам;

$z: T \rightarrow R^+$ – функция, с помощью которой задается время выполнения переходов;

$v = \{V_{is}\}$ – конечное непустое множество переменных (атрибутов маркера), изменяющих свои значения при выполнении сети;

$Q = \{q_n\}$ – конечное множество решающих процедур, с помощью которых задаются правила выполнения условных переходов;

$\Psi = \{\Psi_n\}$ – множество процедур перехода, выполняющих преобразования над атрибутами маркеров;

$M = (M_0(p_i)), p_i \in P, M_0(p_i) \in \{0, 1\}$ – начальная маркировка сети.

Переход срабатывает при наличии маркера во входной позиции и отсутствии его в выходной позиции. На примере безусловного T перехода это продемонстрировано на рис. 1. Формально это можно записать так: $M_1(p_1)\{0\}, M_1(p_2)\{1\}$.

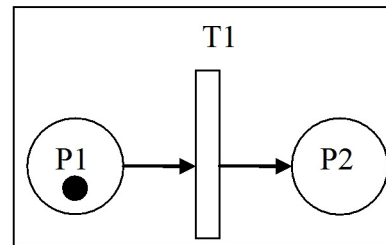


Рис. 1. T-переход

Формальное описание процесса преобразования значений атрибутов инициализирующего маркера представлено ниже:

$$V_{is} = \begin{cases} \Psi_k(R), k = s \\ V_{is}, k \neq s \end{cases}, \quad s = \overline{1, n}, \quad (1)$$

где:

R – параметры, от которых зависит функция преобразования значения атрибута Ψ_k . В качестве параметра может выступать: значение изменяемого атрибута, значения других атрибутов маркера, параметром может быть случайная величина с известным законом распределения;

V_{is} – значение s -го атрибута маркера, находящегося в позиции i ;

n – количество атрибутов маркера, находящегося в позиции i .

Время работы перехода зависит от функции $z: T \rightarrow R^+$. Предлагается в CASE-средстве реализовать возможность гибкого управления временем работы перехода. Для этого можно учитывать значения атрибутов маркеров, которые активизируют переход. При этом появляется возможность изменять время работы переходов по заданному закону, что в значительной мере повышает моделирующие возможности системы.

Условные переходы типа X и Y помимо перечисленных особенностей в своей структуре имеют условный переключатель, задающий направление ветвления или выборки в зависимости от условия. Рассмотрим процесс срабатывания X-перехода, представленного на рис. 2. В общем случае разрешающая процедура может быть сколь угодно сложной, но сущность ее работы заключается в проверке условий разветвления потока маркеров.

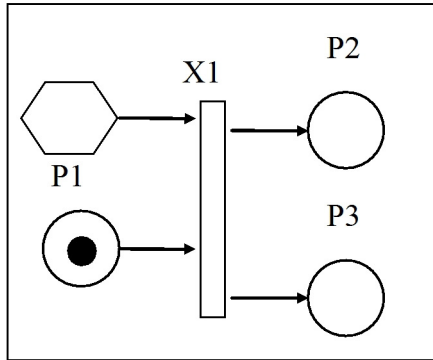


Рис. 2. Переход типа X

Ниже приведено формальное описание работы X-перехода в случае, когда выходные позиции не содержат маркеров:

$$\begin{cases} M_1(p_1) = M_0(p_1) - 1 \\ M_1(p_2) = M_0(p_2) + 1 \\ M_1(p_3) = M_0(p_3) \end{cases} \quad \forall V_{1s} \in V : q_1(q_1^i(p_2), \dots, q_n^i(p_2)) \in Q \quad (2)$$

$$\begin{cases} M_1(p_1) = M_0(p_1) - 1 \\ M_1(p_3) = M_0(p_3) + 1 \\ M_1(p_2) = M_0(p_2) \end{cases} \quad \forall V_{1s} \in V : q_1(q_1^i(p_3), \dots, q_n^i(p_3)) \in Q, \quad s = \overline{0, n}$$

где:

n – количество проверяемых атрибутов (условием для маркирования определенной позиции могут служить значения нескольких атрибутов иницирующего маркера);

q_1 – конечное множество решающих процедур $q_i^i(p_i)$, одним из параметров которых является позиция, которая будет маркирована в результате работы перехода и удовлетворения условий.

При функционировании перехода типа Y (рис. 3) наблюдается противоположная ситуация. Когда маркирована только одна из входных позиций $M_x(p_i) = 1, p_i \in I(t_m)$, где t_m – описываемый переход типа Y, работа перехода схожа с работой T-перехода. Сложность возникает в случае, когда маркированы две и более позиций, связанных по входам, и состоит в определении очередности извлечения маркеров.

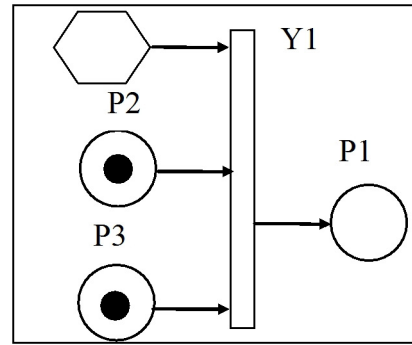


Рис. 3. Переход типа Y

Формальное определение решения такой ситуации представлено ниже:

$$\begin{cases} M_1(p_1) = M_0(p_1) + 1 \\ M_1(p_2) = M_0(p_2) - 1 \\ M_1(p_3) = M_0(p_3) \end{cases} \quad \forall V_{2s} \in V : q_1(q_1^i(p_1), \dots, q_n^i(p_1)) \in Q$$

$$\begin{cases} M_1(p_1) = M_0(p_1) + 1 \\ M_1(p_3) = M_0(p_3) - 1 \\ M_1(p_2) = M_0(p_2) \end{cases} \quad (3)$$

$$\forall V_{3s} \in V : q_1(q_1^i(p_1), \dots, q_n^i(p_1)) \in Q, \quad s = \overline{0, n'}$$

3. Система имитационного моделирования E-NetSim

Приведенный подход довольно трудоемок, из-за чрезвычайно большого объема информации, которую необходимо ввести пользователю при задании параметров модели. В большинстве известных систем имитационного моделирования на базе сетей Петри задание этих параметров осуществляется с помощью специализированного языка, что требует специальных навыков. В предлагаемой системе имитационного моделирования используется интерактивная схема описания модели. Суть ее заключается в объединении графического и математического методов. При помощи графического редактора из готовых простейших элементов E-сети создается модель исследуемой системы, а при помощи окон свойств элементов вносится информация о функциональных и параметрических характеристиках каждого объекта в модели.

Средство обладает гибким графическим интерфейсом, позволяющим продемонстрировать динамическое изменение положения того или иного маркера относительно позиций и обеспечить хорошую визуализацию процессов, происходящих в модели. Пользователю достаточно правильно выделить «причинно-следственные» связи в исследуемом объекте и создать имитационную модель.

Задание математических соотношений, определяющих поведение отдельных элементов моделируемой системы пользователем, также максимально упрощено и сведено к некоторому набору простых, логически

понятных, действий. Существует также возможность динамического изменения времени срабатывания перехода. Для этого в свойствах перехода в поле «время срабатывания» необходимо ввести переменную, которая является атрибутом маркера, инициирующего работу перехода. После этого при попадании маркера во входную позицию такого перехода происходит динамическое изменение времени работы перехода (рис. 4).

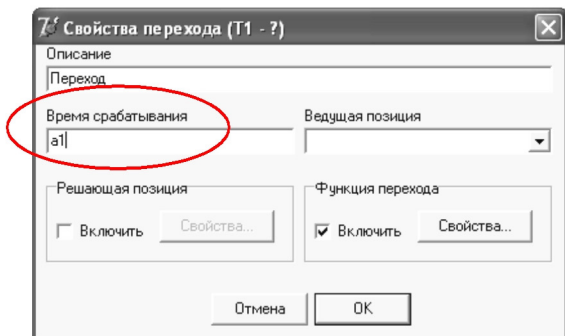


Рис. 4. Динамическое определение времени перехода

Основными источниками маркеров в системе являются генераторы. Генератор представляет собой базовый элемент теории E-сетей и является совокупностью простых элементов. Схематическое изображение генератора представлено на рис.5.

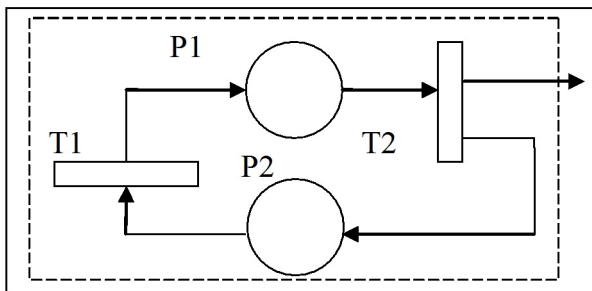


Рис. 5. Схематическое изображение объекта «Генератор»

В предлагаемом CASE-средстве для инициализации генератора, необходимо задать характеристики временного распределения и характеристики распределений атрибутов маркеров, создаваемых генератором. Количество атрибутов при этом неограниченно. В системе реализовано задание распределений, таких как экспоненциальное, равномерное, нормальное, также значение определяемого параметра может быть константой.

После того как модель построена и заданы математические соотношения, определяющие все её параметры, проводится проверка на корректность соединения элементов модели. В случае если будут найдены ошибки в определении элементов модели, в отдельном окне выводится соответствующая информация. Процесс управления моделированием также довольно гибок. В системе имеется возможность выбора нескольких режимов моделирования: пошаговый, по-

зволяющий эксперту произвести «доводку модели» и проверку на правильность функционирования; моделирование выбранного количества шагов; прогон модели до заданного события, условия или момента времени. В последнем случае возможно установить требуемое время выполнения одного такта, тем самым ускоряя или замедляя процесс моделирования.

В конце эксперимента с моделью происходит генерация календаря событий, который хранит в себе полную информацию о всех имевших место событиях во время моделирования. Обработывая отчетную информацию, эксперт может получить интересные его сведения, начиная от вероятности пребывания системы в том или ином состоянии до временных характеристик элементов исследуемой системы.

Как уже было сказано ранее, для описания модели используются математический и графический методы. В соответствии с ними в средстве реализован графический редактор и набор диалоговых форм, позволяющих задать характеристики элемента в понятной для пользователя форме. Рассмотрим более подробно процесс описания модели при помощи диалоговых форм.

Основную часть главного окна средства занимает рабочая область редактора сети. На ней пользователь при помощи кнопок управления и базовых элементов оценочных сетей методом Drag&Drop формирует структуру имитационной модели исследуемой системы или процесса. Чтобы добавить элемент на рабочую область редактора сети, пользователю необходимо воспользоваться следующим меню, представленным на рис. 6.

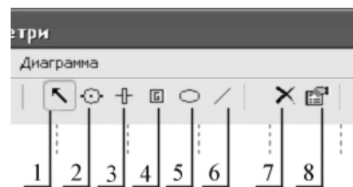


Рис. 6. Панель инструментов

Здесь:

- 1 – выбор объекта и перемещение по области редактора;
- 2 – добавление позиции;
- 3 – добавление перехода;
- 4 – добавление генератора;
- 5 – добавление очереди;
- 6 – добавление дуги (соединение разнотипных элементов);
- 7 – удаление выбранного объекта либо группы объектов;
- 8 – свойства выбранного объекта.

После того, как требуемый объект расположен в рабочей области, его можно перемещать, масштабировать, вращать, а также выполнять редактирование свойств элемента при помощи диалогового интерфейса, вид которого будет различен в зависимости от выбранного элемента.

На рис. 7 представлена главная форма, используемая при задании свойств позиции:

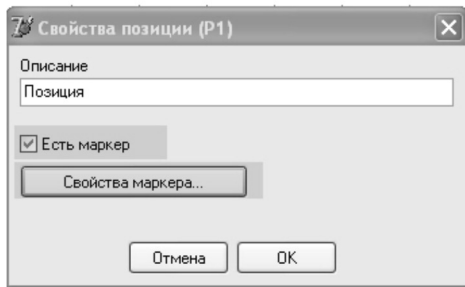


Рис. 7. Задание свойств элемента «Позиция»

При инициализации редактирования свойств позиции на экране возникает диалоговая форма «Свойства позиции», позволяющая задавать смысловое описание позиции, а также добавить маркер стартовой разметки сети. Если в позицию принудительно помещается маркер, то возможна модификация его параметров. Для этого достаточно при помощи кнопки «Свойства маркера» вызвать одноименную форму (рис. 8).

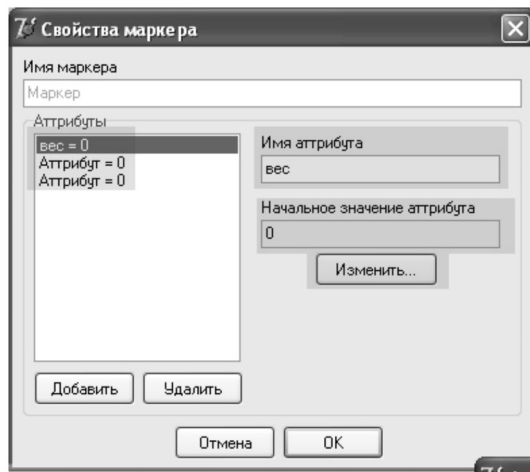


Рис. 8 Задание свойств элемента «Маркер»

Поле «Имя маркера» является неизменным, так как в процессе моделирования имена маркерам присваиваются автоматически и отличаются номером-идентификатором. В левой области окна отображается список атрибутов маркера. Редактирование свойств атрибута выполняется при помощи двух полей – «Имя атрибута» и «Начальное значение». Поле «Имя атрибута» позволяет дать смысловое описание характеристик маркера, зачастую заимствованное из терминологии предметной области исследуемой системы. Редактирование поля «Начальное значение атрибута» осуществляется при помощи формы «Описание функции», которая позволяет установить тип значения и его параметр. В средстве реализовано три наиболее часто используемых закона распределения случайных величин (экспоненциальный, нормальный, равномерный) и функция постоянного значения.

Задание свойств генератора маркеров выполняется с помощью формы, представленной на рис. 9.

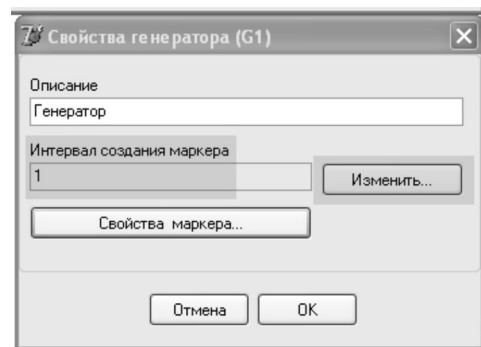


Рис. 9. Задание свойств объекта «Генератор»

Для задания свойств генератора маркеров, необходимо указать интервал времени между появлением маркеров на выходе генератора, а также определить набор атрибутов, характеризующих каждый создаваемый маркер. Значением поля «Интервал создания маркера» может также быть либо константа, либо случайная величина, соответствующая одному из трех законов распределения

Объект «Очередь» в модели описывается с помощью формы, представленной на рис. 10.

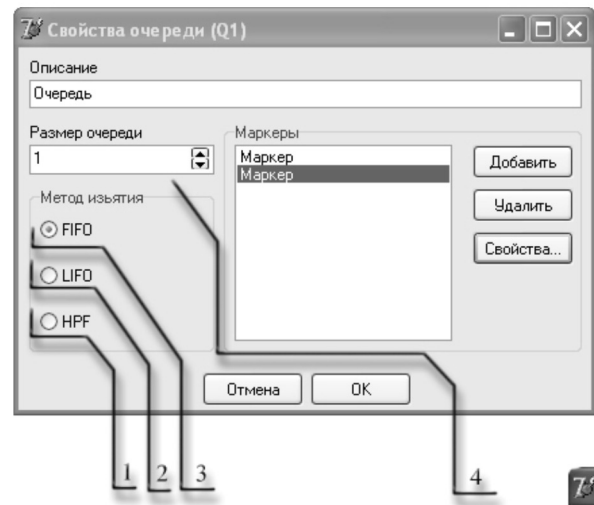


Рис. 10. Задание свойств объекта «Очередь»

Очередь характеризуется размером (это максимальное число помещаемых в неё маркеров) и типом диспетчеризации (это метод изъятия маркеров из очереди). Имеется возможность задавать три наиболее часто встречаемых метода освобождения очереди – FIFO, LIFO, HPF. В отдельной области выводится информация о маркерах, присутствующих в очереди, а при помощи трех управляющих кнопок возможно добавление маркеров в очередь и редактирование значений их атрибутов при создании стартовой разметки сети.

На рис. 10 представлены по пунктам управляющие кнопки и поля, среди которых:

- 1 – тип диспетчеризации NPF;
- 2 – тип диспетчеризации LIFO;
- 3 – тип диспетчеризации FIFO;
- 4 – размер очереди, устанавливаемый перед началом процесса проигрывания модели.

Описание объекта «Переход» может быть осуществлено при помощи следующей формы, представленной на рис. 11.

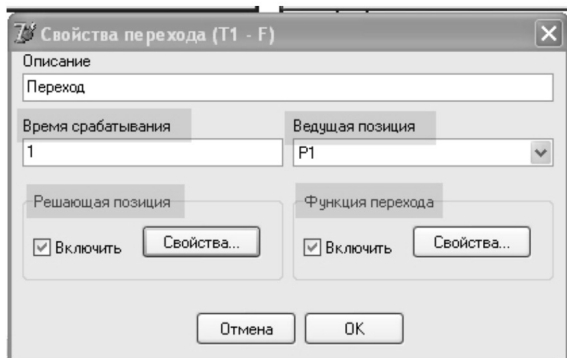


Рис. 11. Задание свойств объекта «Переход»

Характеристиками объекта «Переход» являются: время срабатывания перехода – это время от момента инициализации работы до срабатывания перехода; функция перехода – это правило преобразования атрибутов маркера, выполняемое в результате запуска перехода; решающая позиция – это определение функции ветвления или выборки для разрешения конфликтных ситуаций в переходах условного типа. Ведущая позиция задает номер позиции, из которой маркер переносится в выходную позицию перехода типа F. Время работы каждого перехода может быть величиной постоянной или изменяться динамически в процессе моделирования.

Формы задания правил преобразования атрибутов маркера при срабатывании перехода и определения функции, задающей правила работы условных переходов, схожи по своему интерфейсу. Для формирования функции необходимо воспользоваться специальными полями.

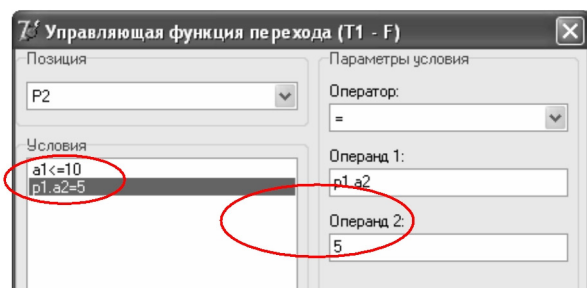


Рис. 12. Определение функций перехода

Функционально форма «Управляющая функция перехода» (рис. 12) разбита на три области: «Позиция» – привязка условий осуществляется с учетом позиций, которые будут либо маркированы в результате конф-

ликта, либо из которых будет изъят маркер; «Условия» – собственно конечное множество условий, разрешающих конфликтную ситуацию; область «Оператор» – выполняет функцию составления условий.

Выводы

Проведенный анализ позволяет сделать вывод о том, что использование оценочных сетей при исследовании сложных систем открывает возможности комплексного описания изучаемого объекта. Существующие средства программного представления моделей в виде E-сети сложны в использовании и требуют специальной программной подготовки от пользователя. В работе описано CASE-средство, предназначенное для создания и исследования имитационных моделей дискретно-событийных систем в виде оценочных сетей Петри. Приведено описание разработанного программно-технического комплекса, описаны его функциональные и моделирующие возможности, показана диалоговая система взаимодействия с пользователем. В настоящее время происходит модернизация комплекса путем добавления новых функциональных возможностей.

Список литературы: 1. Питерсон Дж. Теория сетей Петри и моделирование систем: Пер. с англ. – М.: Мир, 1984 г. – 264 с. 2. Котов В.Е. Сети Петри. – М.: Наука, 1984. – 160 с. 3. Костин А. Е., Шаньгин В. Ф. Организация и обработка структур данных в вычислительных системах: Учеб. пособ. для вузов. – М.: Высш. шк., 1987. – 248 с. 4. Варжапетян А. Г. Имитационное моделирование на GPSS/H: Учеб. пособ.; ГУАП. – СПб., 2007. – 384 с.: ил. 5. <http://www.mathworks.com/products/stateflow>. 6. Гуляев А. Визуальное моделирование в среде MATLAB: Уч. пособ. – СПб.: Питер, 2000. – 432 с.

Поступила в редколлегию 09.02.2009

УДК 681.32

Подання імітаційних моделей дискретно-подійних систем за допомогою CASE засобу E-NetSim/ В. Г. Иванов, І. В. Шевченко // Біоніка інтелекту: наук.-техн. журнал. – 2009. – №1(70). – С. 44-49.

У роботі описано моделюючі можливості та гнучкість системи графічного моделювання оцінних мереж – E-NetSim. Розглянуто особливості створення імітаційної моделі, введення параметрів окремих елементів, «програвання» моделі й обробки отриманих результатів.

Лі.: 12. Бібліогр.: 6

UDK 681.32

Presentation of the simulation models discrete - event systems by means of CASE facility E-NetSim/ V.G. Ivanov, I.V. Shevchenko // Bionics of Intelligence: Sci. Mag. – 2009. – №1(70). – P. 44-49.

Given work considers the possibility and show flexible of the graphic system of modeling of the merit networks – E-NetSim. This work is described particularities of the creation a simulation model, tasks parameter of its component parts, «playing» models and processing got data

Fig.: 12. Ref.: 6 items.

УДК 004.89



О СХОДИМОСТИ ИММУННЫХ АЛГОРИТМОВ АДАПТАЦИИ СИСТЕМ НЕЧЕТКОГО ВЫВОДА

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 сходимости иммунных алгоритмов адаптации систем нечеткого вывода. Процесс адаптации состоит в настройке формы и параметров функций принадлежности (параметрическая адаптация) и параметров и структуры базы нечетких правил (структурная адаптация). В работе приведены псевдокоды рассматриваемых алгоритмов, рассмотрен используемый оператор мутации. Показана сходимость иммунных алгоритмов адаптации моделей нечеткого вывода при условии $0 < p_m < 1$, где p_m – вероятность мутации.

ИСКУССТВЕННЫЕ ИММУННЫЕ СИСТЕМЫ, АДАПТИВНАЯ МОДЕЛЬ НЕЧЕТКОГО ВЫВОДА, ФУНКЦИЯ ПРИНАДЛЕЖНОСТИ, БАЗА НЕЧЕТКИХ ПРАВИЛ, СХОДИМОСТЬ, ЦЕПИ МАРКОВА

Введение

Искусственные иммунные системы (ИИС) начали развиваться с 1990 года как новое направление в области вычислительного интеллекта. Число моделей ИИС, которые применяются в различных областях научных и инженерных проектных решений, постоянно увеличивается. В ИИС используется способность естественной иммунной системы вырабатывать новые типы антител и отбирать наиболее подходящие из них для взаимодействия с попавшими в организм антигенами [1]. Для объяснения иммунологических механизмов существуют различные теории – теория иммунной сети, принципы клонального и негативного отбора. Существует несколько вычислительных моделей ИИС: CLONALG [2], VCA [3], MISA [4], основанных на принципе клонального отбора; модель иммунной сети для задач распознавания и анализа данных aiNET [5] и другие.

ИИС используют вычислительные модели обработки информации в иммунологических взаимодействиях с практическими применениями для решения многих проблем, таких как распознавание образов, поиск данных, компьютерная безопасность, обнаружение ошибок, классификация, оптимизация и другие. Свойства ИИС, такие как распознавание, разнообразие, обучение, память, распределенное обнаружение и другие, позволяют использовать иммунные принципы для решения задач адаптации систем нечеткого вывода. Иммунные алгоритмы адаптации систем нечеткого вывода, предложенные в [6], основанные на принципе клонального отбора и теории иммунной сети, позволяют сформировать базу нечетких правил по данным обучающей выборки, а также выполнить оптимизацию параметров нечетких правил.

Несмотря на разнообразие вычислительных моделей ИИС, вопросам сходимости иммунных алгоритмов в литературе посвящено очень мало внимания. Одной из первых работ в данной области является [4],

в которой представлено полное доказательство сходимости иммунного алгоритма MISA, основанного на принципе клонального отбора с использованием двоичного кодирования и дополнительной популяции лучших решений. В [7] доказана сходимость иммунного алгоритма VCA, основанного на принципе клонального отбора с использованием двоичного кодирования и специфичного оператора мутации. В [4,7] показано, что алгоритмы являются полностью стохастическими и могут быть описаны с использованием теории Марковских цепей. В [8] рассмотрены обобщенные подходы к доказательству сходимости иммунных алгоритмов в зависимости от используемых иммунных операторов.

В данной работе рассматривается сходимость иммунных алгоритмов структурной и параметрической адаптации системы нечеткого вывода [6] с использованием теории Марковских цепей.

1. Постановка задачи

Адаптивную систему нечеткого вывода можно представить в виде кортежа:

$$AFIS = \langle \{X_i, y_i\}, RB, DB, I, G(RB), L(DB), F \rangle, \quad (1)$$

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; $L(DB)$ – оптимизация базы данных; F – функция, оценивающая эффективность нечеткой модели.

Основная трудность использования данной системы заключается в необходимости структурной и параметрической адаптации (оптимизации) – генерации и оптимизации базы правил и оптимизации параметров базы данных [6].

Для выполнения структурной и параметрической адаптации модели нечеткого вывода применяются иммунные алгоритмы, которые можно формально представить следующим образом:

$$ImmAlg = \left(Ag, Ab, L, Alph, Aff, Cl, \right. \\ \left. Mut, Edit, Supp, N, p_m, d, \delta, \vartheta \right), \quad (2)$$

где $Ag = \{Ag_1, Ag_2, \dots, Ag_{N_{Ag}}\}$ – популяция антигенов; N_{Ag} – размер популяции антигенов; $Ab = \{Ab_1, Ab_2, \dots, Ab_{N_{Ab}}\}$ – популяция антител, где каждое антитело Ab_i представляет собой вектор $Ab_i = \langle ab_1, ab_2, \dots, ab_L \rangle, i = 1, N_{Ab}$; N_{Ab} – размер популяции антител; L – количество параметров вектора антитела; $Alph$ – используемый алфавит мощностью A ; Aff – функция аффинности; Cl – оператор клонирования; Mut – оператор мутации; $Edit$ – оператор редактирования популяции; $Supp$ – оператор супрессии; N – количество антител для клонирования; p_m – вероятность мутации; d – количество худших антител для замены случайно сгенерированными новыми антителами; δ – порог сжатия иммунной сети; ϑ – критерий останова работы алгоритма.

Алгоритмы структурной и параметрической адаптации являются полностью стохастическими, и, следовательно, могут быть описаны в терминах Марковских цепей. Таким образом, в работе ставится задача на основе теории Марковских цепей доказать сходимость иммунных алгоритмов адаптации системы нечеткого вывода.

2. Иммунные алгоритмы адаптации систем нечеткого вывода

Рассматриваются иммунные алгоритмы адаптации моделей нечеткого вывода по алгоритму Такаги-Сугено с гауссовыми функциями принадлежности, подробно описанные в [6]. Адаптация данной модели нечеткого вывода включает в себя как структурную адаптацию – формирование базы нечетких правил, так и параметрическую адаптацию конечного набора нечетких правил. По результатам выполнения структурной и параметрической адаптации формируется база знаний системы нечеткого вывода.

2.1. Иммунный алгоритм структурной адаптации

В задаче структурной адаптации в роли антигенов выступают примеры обучающей выборки, в роли антител – правила нечеткого вывода. Для кодирования антител используется подход, при котором каждое антитело соответствует одному правилу из базы правил.

Иммунный алгоритм структурной адаптации основан на принципе клонального отбора, в соответствии с которым использование механизмов мутации клонов антител обеспечивает поддержание разнообразия антител в популяции (правил нечеткого вывода). Для сокращения числа правил применяется теория иммунной сети. Использование сетевого взаимодействия антител внутри популяции клеток памяти позволяет удалять избыточные и противоречивые правила путем вычисления аффинности антитело-антитело и сжатия сети.

Иммунный алгоритм структурной адаптации можно представить с помощью следующего псевдокода:

1. Генерация начальной популяции антител Ab .
2. Цикл для каждого антигена Ag_i :
 - 2.1. В систему нечеткого вывода подставляются параметры функций принадлежности 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. Вычисление аффинности популяции C^* к антигену Ag_i .
 - 2.6. Редактирование популяции антител Ab и заполнение популяции клеток памяти M .
 - 2.7. Супрессия клеток памяти внутри популяции Ab .
 - 2.8. Замена d худших антител новыми антителами в популяции антител Ab .
3. Проверка критерия останова. Если достигнут – переход к шагу 4, в противном случае – к шагу 2.
4. Конец. Результатом является популяция клеток памяти M , которая представляет собой набор правил вывода с вычисленными коэффициентами нечетких правил.

2.2. Иммунный алгоритм параметрической адаптации

Этап параметрической адаптации конечного набора правил сводится к нахождению, в соответствии с имеющимися экспериментальными данными и принятым критерием, оптимальных значений параметров для оставленных в базе правил.

В роли антигенов в алгоритме параметрической адаптации выступают примеры обучающей выборки. В роли антител – параметры функций принадлежности для входных переменных. Иммунный алгоритм параметрической адаптации основан на принципе клонального отбора.

Алгоритм параметрической адаптации можно представить при помощи следующего псевдокода:

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 функций принадлежности 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 достигнут — переход к шагу 4, в противном случае — к шагу 2.

4. Конец. База данных адаптивной нечеткой системы формируется по лучшему антителу популяции.

Основными отличиями представленных алгоритмов структурной и параметрической адаптации является использование дополнительной популяции клеток памяти для хранения найденных коэффициентов нечетких правил и ее супрессия для удаления избыточных и противоречивых правил в соответствии с положениями теории иммунной сети на этапе выполнения структурной адаптации.

2.3. Мутация антител при вещественном кодировании

Характерной особенностью описанных алгоритмов является использование специфичного оператора мутации, необходимого при использовании вещественного кодирования антител. Мутация вектора вещественных параметров выполняется путем добавления гауссовского шума. Мутация заключается в добавлении к текущему значению параметров в антителе случайной переменной с нормальным законом распределения и нулевым математическим ожиданием: $x_{i+1} = x_i + N(0, \sigma_i)$, где σ_i — дисперсия случайной величины N , которая ассоциируется с каждым параметром x_i антитела. Большие значения σ_i приводят к большей мутации. Ожидаемое значение амплитуды мутации каждого параметра можно контролировать путем настройки σ_i — шага мутации.

Настройка шага мутации выполняется с использованием самоадаптирующегося алгоритма, используемого в эволюционных системах и эволюционном программировании [9], при котором параметр σ_i сначала подвергается мутации, а соответствующие параметры антитела изменяются уже с использованием новых значений σ_i . Для изменения значений σ_i используется логнормальная мутация:

$$\sigma_{i+1} = \sigma_i \exp(\tau' N(0,1) + \tau N_i(0,1)), \quad (3)$$

в которой в качестве τ и τ' используются значения, предложенные в [9]: $\tau = (\sqrt{2\sqrt{n}})^{-1}$, $\tau' = (\sqrt{2n})^{-1}$, где n — количество параметров.

Предлагаемый оператор самоадаптирующейся гауссовской мутации предполагает одновременную мутацию всех параметров в антителе, что не всегда необходимо. Использование вероятности мутации позволяет вносить гауссовский шум в антитело в случайные параметры. Вероятность мутации определяет, какой процент параметров, выбранных из антитела случайным образом, будет подвержен мутации.

Вероятность мутации вычисляется при каждой операции мутации обратно-пропорционально аффинности мутируемого антитела.

Вероятность мутации и шаг мутации влияют на скорость сходимости алгоритма. Большие шаг и вероятность мутации поддерживают разнообразие антител в иммунной системе на достаточно высоком уровне, делая случайные скачки по всей области поиска, и позволяют находить глобальный экстремум, не застревая в точках локального оптимума, однако большинство изменений могут привести к нефункциональным антителам с плохой аффинностью. Небольшие шаг и вероятность мутации позволяют иммунной системе исследовать локальную область вокруг антитела, делая небольшие шаги в поиске антител с лучшей аффинностью, ведущих к оптимальному значению, но скорость сходимости алгоритма уменьшится.

За счет использования обратно-пропорциональной самоадаптирующейся гауссовской мутации механизм мутации регулируется в процессе работы иммунного алгоритма и для антител с улучшенной в результате мутации аффинностью уменьшает вероятность мутации в последующих поколениях работы иммунного алгоритма, чтобы не потерять полученные преимущества.

3. Сходимость иммунных алгоритмов параметрической и структурной адаптации

За основу доказательства сходимости взято доказательство сходимости алгоритма ВСА с двоичным представлением антител и специфичным для данного алгоритма оператором мутации [4]. Так как в процессе клонального отбора антитела в популяции не взаимодействуют друг с другом, может быть использована модель популяции с одним антителом, для увеличения которой до любого размера используются правила независимых вероятностей [4].

В задачах структурной и параметрической адаптации состояние системы в момент времени t описывается случайной величиной X_t , значение которой моделируется вектором распределения вероятностей $V_t = (v_{t,1}, v_{t,2}, \dots, v_{t,N})$, где $v_{t,j} = P(X_t = j)$ — вероятность того, что в момент времени t система находится в состоянии j ; N — количество возможных состояний системы $N = A^L$; A — мощность используемого алфавита; L — количество параметров вектора антитела.

Вероятность перехода из состояния j в состояние k не зависит от времени t и может быть представлена матрицей переходов P размерности $N \times N$ с элементами $P_{jk} = P(X_{t+1} = k | X_t = j)$.

Распределение вероятностей для момента времени $t+1$ записывается как $v_{t+1,k} = \sum_{j=1}^N v_{t,j} P_{jk}$, или в матричной форме: $V_{t+1} = V_t P$. Так как матрица P не зависит от времени, вектор распределения вероятностей V_t в момент времени t можно записать как $V_t = V_0 P^t$,

где V_0 – вектор начального распределения вероятностей.

Для доказательства сходимости алгоритма необходимо рассмотреть поведения матрицы P^t при $t \rightarrow \infty$ и показать, что если существует хотя бы одно оптимальное состояние, оно достижимо с вероятностью 1 при $t \rightarrow \infty$ и является поглощающим.

Матрица переходов P может быть определена при помощи двух матриц: матрицы мутаций и матрицы редактирования.

Матрица мутаций определяет влияние оператора мутации на возможность перехода из состояния в состояние и зависит от пространства поиска и представления антител. Матрица мутаций содержит вероятности того, что алгоритм найдет оптимальное решение в пространстве поиска, которые формируются по маске мутаций:

$$M_m = p_m^k (1 - p_m)^{L-k} \left(\frac{1}{A-1} \right)^k, \quad (4)$$

где p_m – вероятность мутации параметров антитела; k – количество изменяемых параметров в антителе; L – длина строки антитела; A – мощность используемого алфавита.

Матрица редактирования определяет влияние оператора редактирования на состояние популяции и отвечает за то, что если оптимальное состояние будет найдено, оно останется в популяции и не будет потеряно. Матрица редактирования зависит от пространства поиска и не зависит от способа кодирования антител. Если аффинность конечного состояния лучше аффинности начального состояния, элементу матрицы, соответствующему данному переходу, присваивается 1. Если аффинность хуже или равна – присваивается 0.

Для получения матрицы переходов выполняется поэлементное умножение матрицы мутаций на матрицу редактирования.

Для случая параметрической адаптации модель иммунного алгоритма с одним антителом в популяции является полностью общей, так как результатом работы алгоритма будет одно лучшее по популяции антитело, которое как только будет найдено, будет сохранено в популяции, что гарантирует оператор редактирования популяции.

Для структурной адаптации результатом является набор правил нечеткого вывода – то есть набор антител. Поэтому в иммунном алгоритме структурной адаптации используется дополнительная популяция клеток памяти, в которую помещаются лучшие антитела из основной популяции антител. Эти клетки не участвуют в клональном отборе и используются только для сохранения лучших результатов, найденных в ходе работы алгоритма. В каждом поколении иммунного алгоритма в клетки памяти заносятся антитела, найденные в предыдущем поколении, которые улучшают соответствующие антитела (правила), записанные в популяции клеток памяти. Для устранения избыточно-

сти правил в популяции клеток памяти применяется оператор супрессии, не нарушающий при этом полноты базы нечетких правил.

Доказательство сходимости:

Утверждение 1. Все неоптимальные состояния являются переходными при $0 < p_m < 1$.

Доказательство. При условии $0 < p_m < 1$ матрица мутаций содержит только ненулевые элементы. Таким образом, возможно достичь оптимальное состояние за один шаг из любого начального состояния в соответствии с матрицей редактирования, позволяющей выполнять переходы из неоптимальных состояний в оптимальные с вероятностью 1. Следовательно, для неоптимального состояния j вероятность остаться в том же состоянии на следующей итерации $p = P_{jj} < 1$. Как только состояние будет изменено на состояние с лучшей аффинностью, в него уже нельзя будет вернуться в соответствии с матрицей редактирования. Следовательно $P_{jj}(n) = p^n$, где n – количество поколений, то есть

$$\sum_{n=0}^{\infty} P_{jj}(n) = \sum_{n=0}^{\infty} p^n = \frac{1}{1-p} < \infty, \quad (5)$$

следовательно, состояние j является переходным.

Теория цепей Маркова является эффективной в случае неприводимых цепей. Цепь, соответствующая иммунному алгоритму, является приводимой. Однако, применив теорему декомпозиции [10], цепь можно разбить на множество переходных неоптимальных состояний и множество возвратных состояний, соответствующих оптимуму. Поэтому, если задача оптимизации имеет M оптимумов, то матрицу переходов можно представить в следующем виде:

$$P = \begin{bmatrix} 1 & 0 \\ R & Q \end{bmatrix}, \quad (6)$$

где 1 – матрица переходов из оптимального состояния в оптимальное размерностью $M \times M$; 0 – матрица переходов из оптимального состояния в неоптимальное размерностью $M \times (N - M)$; R – матрица переходов из неоптимального состояния в оптимальное размерностью $(N - M) \times M$; Q – матрица переходов из неоптимального состояния в неоптимальное размерностью $(N - M) \times (N - M)$.

В соответствии со стандартными свойствами стохастических матриц [10] $Q^t \rightarrow 0$ при $t \rightarrow \infty$.

Утверждение 2. Все оптимальные состояния являются поглощающими.

Доказательство. Матрица редактирования запрещает переходы из глобального оптимума даже в другой глобальный оптимум. Как только алгоритм попадает в состояние глобального оптимума, оно удовлетворяет условию возвратности $\sum_{n=0}^{\infty} P_{jj}(n) = \infty$. Для оптимального состояния j $p = P_{jj} = 1$, то есть как только

оптимальное состояние найдено, вероятность остаться в нем равна 1. Следовательно, сумма (6) сходится и состояние оптимума является возвратным, а так как состояния не взаимодействуют, то оно является и поглощающим.

Из этого следует, что алгоритм сходится при условии $0 < p_m < 1$.

Выводы

В работе приведено доказательство сходимости иммунных алгоритмов структурной и параметрической адаптации систем нечеткого вывода с использованием вещественного кодирования и со специфичным оператором мутации на основе теории Марковских цепей. Так как алгоритмы основаны на принципе клонального отбора, при котором антитела не взаимодействуют друг с другом, модель иммунного алгоритма рассмотрена для случая одного антитела в популяции, увеличить которую до любого размера можно на основе правил независимых вероятностей.

Доказательство сходимости алгоритмов основано на двух основных утверждениях: во-первых, в результате мутации, в соответствии с матрицей мутации, можно достичь оптимальное состояние из неоптимального за один шаг (в матрице мутации не учитывается шаг мутации, то есть предполагается, что за один шаг каждый параметр в антителе может изменяться на любую случайную величину); и, во-вторых, как только оптимальное состояние будет найдено, оно сохранится в популяции и не будет утеряно, что гарантирует матрица редактирования, определяющая влияние оператора редактирования на состояние популяции. Показано, что использование супрессии клеток памяти в случае структурной адаптации не противоречит данным утверждениям. Доказано, что иммунные алгоритмы структурной и параметрической адаптации сходятся при условии $0 < p_m < 1$, где p_m – вероятность мутации.

Список литературы. 1. *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. **2.** *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. – 2002. – P. 239-251. **3.** *Kelsey, J., Timmis, J.* Immune Inspired Somatic Contiguous Hypermutation for Function Optimisation. *Proc. of Genetic and Evolutionary Computation Conference-2003.* – Springer Lecture Notes in Computer Science 2723. – P. 207-218. **4.** *Villalobos-Arias, M., Coello Coello, C.A., Hernández-Lerma, O.* Convergence Analysis of a Multiobjective Artificial Immune System Algorithm // *Proc. of ICARIS.* – 2004. – Springer Lecture Notes in Computer Science 3239. – P. 226-235. **5.** *L.N. de Castro and F.J. Von Zuben.* aiNet: an artificial immune network for data analysis // *Data Mining: A Heuristic Approach*, Chapter XII, H.A. Abbass, R.A. Sarker, and C.S. Newton, Eds. USA: Idea Group Publishing, 2001. – P. 231-259. **6.** *Korablyov M.M., Ovcharenko I.V.*

Adaptation of fuzzy inference models using artificial immune systems // Proc. of Advanced Computer Systems and Networks: Design and Application Conference-2007. – Lviv. – P. 89-91. **7.** *Edward Clark, Andrew Hone, Jon Timmis.* A Markov Chain Model of the B-Cell Algorithm // *Proc. of ICARIS-2005.* – P.318-330. **8.** *Cutello, V.; Nicosia, G.; Romeo, M.; Oliveto, P.S.* On the Convergence of Immune Algorithms. *Foundations of Computational Intelligence*, 2007. IEEE Symposium on Volume, 2007. – P. 409 – 415. **9.** *H.-P. Schwefel.* Numerical Optimization of Computer models. – John Wiley & Sons, New-York, 1981, 1995 (2nd edition). **10.** *Grimmett, G.R. & Stirzaker, D.R.* Probability and Random Processes. – Oxford: Oxford University Press, 1982.

Поступила в редколлегию 11.02.2009

УДК 004.89

Про збіжність імунних алгоритмів адаптації систем нечіткого виводу / М.М. Кораблев, І.В. Сорокіна // Біоніка інтелекту: наук.-техн. журнал. – 2009. – № 1 (70). – С. 50-54.

В роботі розглянуто збіжність імунних алгоритмів адаптації моделей нечіткого виводу. Процес адаптації полягає в оптимізації форми та параметрів функцій належності (параметрична адаптація) та параметрів і структури бази нечітких правил (структурна адаптація). Приведено псевдокоди імунних алгоритмів адаптації, проаналізовано оператор обернено-пропорційної гауссовської мутації, що самоадаптується. Показано збіжність імунних алгоритмів адаптації моделей нечіткого виводу за умови $0 < p_m < 1$, де p_m – ймовірність мутації.

Бібліогр.: 10 найм.

UDK 004.89

On the Convergence of immune algorithms of fuzzy inference system adaptation. / N.M.Korablev, I.V.Sorokina // Bionics of Intelligence: Sci.Mag. – 2009. – № 1 (70). – P. 50-54.

The given work considers the convergence of immune algorithms of fuzzy inference system adaptation. The process of adaptation consists in adjustment of form and parameters of membership functions (parametric adaptation) and parameters and structure of fuzzy rule base (structural adaptation). The convergence of the immune algorithms of the fuzzy inference system adaptation under condition $0 < p_m < 1$, where p_m is mutation probability.

Ref.: 10 items.

УДК 519.62



Моделювання фрагментів природної мови алгебро-логічними засобами (на прикладі *k*-значної реалізації морфологічного аналізу)

Г.Г. Четвериков¹, Т.М. Федорова², І.Д. Вечірська³, Д.Ф. Богданов⁴

¹ХНУРЕ, м. Харків, Україна, chetvergg@kture.kharkov.ua.

²ХНУРЕ, м. Харків, Україна, tatanic_fedorova@mail.ru.

³ХНУРЕ, м. Харків, Україна, ira_se@list.ru.

⁴Харківський національний політехнічний університет «ХПІ», м. Харків, Україна

Здійснено перехід до системного та структурного синтезу функціонально-орієнтованих процесорів на основі структур алгебри скінченних предикатів для аналізу та синтезу мови. На прикладі задачі морфологічного аналізу запропоновано методику проектування лінгвістичних об'єктів *k*-значними засобами структур алгебри скінченних предикатів просторового типу та наведено інтерпретацію їх функціонування.

АЛГЕБРА СКІНЧЕННИХ ПРЕДИКАТИВ, ВІДНОШЕННЯ, ЛІНГВІСТИЧНИЙ ОБ'ЄКТ, МОРФОЛОГІЧНИЙ АНАЛІЗ, СТРУКТУРА ПРОСТОРОВОГО ТИПУ, ФУНКЦІОНАЛЬНО-ОРІЄНТОВАНИЙ ПРОЦЕСОР

Вступ

Останнім часом в науці багато уваги приділяється питанням розробки систем автоматичної та автоматизованої обробки мовної субстанції, її аналізу, синтезу, нормалізації, тощо. Досягнення найбільшого економічного та соціального успіху стає можливим за умови активного використання сучасних алгебро-логічних методів опису структури інформаційних процесів, зокрема, «інтелектуального опрацювання текстів на предмет екстракції з них певних знань» [1]. Наразі у світі вже існують програмні засоби, які певною мірою можуть виконувати ці операції [2]. Але такі засоби і відповідні їм лінгвістичні технології повинні мати комплексний характер. Актуальною є задача забезпечення їх достатньо «високоорганізованим інтелектом», що базується як на програмних, так і на апаратних засобах формального моделювання природної мови, наближених за своїми можливостями та принципами організації до роботи мозку людини [3].

Наявність алгебри скінченних предикатів (АСП) [3] дає можливість переходу від алгоритмічного опису інформаційних процесів до опису їх у вигляді рівнянь. Рівняння ж у свою чергу задають відношення між змінними. Усі змінні у рівнянні рівноправні, будь-які з них можуть виступати як у ролі незалежних, так і в ролі залежних. Рівняння дають перевагу перед алгоритмами, яка полягає у можливості передбачення реакції системи навіть при неповній визначеності вхідних сигналів, у той час як неповністю розроблений алгоритм є непрацездатним. Крім того, зміна знань про об'єкт у вигляді системи рівнянь АСП, що описують її структуру, на протигагу алгоритму, не вимагає докорінної зміни структури самої системи. За допомогою формул АСП можна побудувати АСП-структури, які реалізують відповідні скінченні предикати. Це є аналогом побудови комбінаційних схем за формулами

алгебри логіки. В залежності від рівня функціонально-структурної реалізації отримуємо АСП-структури різної складності та призначення [4].

Стаття є логічним продовженням досліджень, які наведені у роботах [5-7].

1. Стан проблеми та постановка задачі

Насамперед необхідно визначитись, що таке «природна мова» або «мова людська»? Однозначної відповіді на це запитання немає. Нашим дослідженням відповідає наступне означення: «Язык — это находящаяся в распоряжении того или иного коллектива система элементов — единиц разных ярусов (морфем, слов, словосочетаний, предложения и тому подобного) плюс система правил функционирования этих единиц. Систему единиц называют инвентарем языка, а систему правил функционирования, то есть правил генерации и понимания осмысленных высказываний — грамматикой (в широком смысле) этого языка» [8], тобто мова — це система знаків, де мовний знак часто відображають у вигляді трикутника Фреге [9].

Таким чином, об'єктами мови виступають певні психофізичні стани та процеси, які відбуваються в мовнорозумовому апараті людини, а усна та писемна її репрезентації слугують елементами інфраструктури мовного процесу [1].

Відомо, що семантичному аналізу речення або зв'язного тексту передують морфологічний та синтаксичний аналіз. В прикладній лінгвістиці під морфологічним аналізом (МА) розуміють аналіз окремих словоформ поза контекстом. В автоматичній обробці природномовного тексту МА означає процедуру, в результаті виконання якої з форми, зовнішнього оформлення слова в тексті можна одержати відомості про будь-які рівні мовної структури.

Задачею морфологічного аналізу є ідентифікація словоформ та присвоєння кожній словоформі комп-

лексу морфологічної інформації (КМІ) [8]. Такий комплекс складається із морфологічно-інформаційних рядків (МІ-рядків), структура яких така: номер, <(основа чи ознаки основи), МІ>, де номер – порядковий номер даної словоформи у висловлюванні; основа (ознака основи) – код семантичної ознаки, номер синтаксичної чи семантичної моделі керування, що присвоєні даній основі у словнику основ; МІ – частина мови та її граматичні категорії: рід, число, відмінок, час, особа, тощо.

У зв'язку з цим вважається доцільним вирішити задачу побудови і дослідження АСП – структур шляхом декомпозиції багатозначних структур на дискретно-аналогові та цифрові субблоки (у вигляді універсальних функціональних перетворювачів просторового типу [10]) для схемної реалізації математичних моделей мовної субстанції на прикладі k -значної реалізації морфологічного аналізу.

2. Джерела розробки та алгоритмічна реалізація

Існує два методи реалізації [6, 7] морфологічного аналізу: словниковий (декларативний), що використовується для аналізу мов із нерозвинутим відмінюванням слів (англійська, французька, тощо) та алгоритмічний (процедурний) – для аналізу мов із розвинутим відмінюванням слів (наприклад, українська). В МА розчленовують словоформи на основу та закінчення, після чого в словниках зберігаються як основи, так і їх закінчення. Для МА виконують пошук у складі словоформи, яку аналізують, деякої словникової основи та певного словникового закінчення. Потім порівнюють інформацію про основу та закінчення і отримують КМІ для всієї словоформи.

Для МА змінюваної словоформи її кінцеву частину по черзі порівнюють із закінченнями словника. Після порівняння ту частину словоформи, що збіглася, відокремлюють і отримують припустиму основу (ПОС), припустиме закінчення (ПЗК) та припустиму морфологічну інформацію (ПМІ).

Дані про ПЗК (ПМІ) зчитують із словника закінчень (морфологічної інформації). Потім переходять до пошуку інших ПЗК, ПОС та ПМІ.

На другому кроці аналізу словоформи виконують ідентифікацію її можливих основ шляхом перевірки збігів отриманих ПОС із основами машинного словника основ.

На третьому кроці МА словоформи порівнюють інформацію з тими ПОС та ПЗК, що отримали підтвердження за допомогою словника основ.

Ефективність МА суттєво залежить від виду подання машинних словників у пам'яті комп'ютера та способу їх кодування. Тому доцільно мати окремий допоміжний словник перенумерованих основ, що наявні в одному примірнику та розташовані в алфавітному порядку.

Розглянемо приклад можливої системи кодування. Для подання значень граматичних категорій будь-якої словоформи використаємо 9-розрядний 10-значний код. Позначимо його розряди: $P(1)$, $P(2)$, ..., $P(9)$. Призначення розрядів:

$P(1)$, $P(2)$ – для кодування частин мови словоформи: 01 – іменник; 02 – повна форма прикметника; 03 – коротка форма прикметника; 04 – дієслово і так далі.

$P(3)$ – для позначення типу (прислівника, числівника, сполучника, частки), класу займенника або параметра словозміни (іменника, повного прикметника) зі значенням «живий», «неживий». $P(3) = 0$ позначає відсутність значень будь-якої з перелічених категорій або їх неозначеність. Якщо $P(3) \neq 0$, то його значення розглядається разом з $P(1)$ і $P(2)$.

$P(4)$ – для передачі значення граматичного роду словоформи: 0 – його відсутність або неозначеність; 1, 2, 3, – чоловічий, жіночий, середній рід відповідно.

$P(5)$ – для кодування значення числа: 0, 1, 2, – відсутність, одинна, множина.

$P(6)$ – для позначення відмінку: 0, 1, 2, 3, 4, 5, 6, 7 – відсутність, називний, родовий, давальний, знахідний, орудний, місцевий, кличний.

$P(7)$ – для передачі значення (коду) категорії пасивності-активності: 0, 1, 2 – відсутність / неозначеність, активність, пасивність.

$P(8)$ – для вираження значення (коду) часу; 0, 1, 2, 3 – відсутність, минулий, теперішній, майбутній.

$P(9)$ – для передачі значення (коду) категорії виду: 0, 1, 2 – відсутність / неозначеність (закінчення), доконаний, недоконаний.

Кодування даних тезауруса (машинного словника) закінчень можна представити як файл запису таким чином:

< XXX , $P(1)$, $P(2)$, ..., $P(9)$ > ,

де XXX – номер морфологічного типу; $P(1)$, $P(2)$, ..., $P(9)$ – код морфологічних ознак закінчень.

Усі записи, що відносяться до одного й того ж закінчення, треба розміщувати один за одним; а для спрощення процедури обробки файла наприкінці кожної групи записів треба додати один чи декілька записів, заповнених нулями. Це спрощує процедуру поповнення файла новою інформацією [11].

Для формування одного МІ-рядка для всієї словоформи порівнюють код основи та код закінчення на відповідність їх перших п'яти розрядів; якщо вони не співпадають, то дані несумісні. Для порівняння вибирають черговий код закінчення. Якщо відповідність визначена, то решта розрядів результуючого коду формуються згідно з правилами 10-значної диз'юнкції / кон'юнкції (див. табл. 1, 2) значень відповідних розрядів кодів основи та закінчення. Крім того, попередньо перевіряють умову збігів операндів чи рівність одного з них нулеві.

Приклад

Словоформа «аналізу» має ПЗК = «у», ПОС = «аналізу»:
 «аналізу» – <001 01 21 00 000>
 «у» – <001 01 01 12 000>
 «аналізу» – <001 01 21 12 000>

Отримана МІ відповідає таким ознакам: 001 – морфологічний тип, 01 – іменник, неживий, чоловічий рід, однина, родовий відмінок.

Таблиця 1

Таблиця істинності 10-значної функції $\max(X_1, X_2)$

$f(X_1, X_2) = \max(X_1, X_2)$		X_2									
		0	1	2	3	4	5	6	7	8	9
X_1	0	0	1	2	3	4	5	6	7	8	9
	1	1	1	2	3	4	5	6	7	8	9
	2	2	2	2	3	4	5	6	7	8	9
	3	3	3	3	3	4	5	6	7	8	9
	4	4	4	4	4	4	5	6	7	8	9
	5	5	5	5	5	5	5	6	7	8	9
	6	6	6	6	6	6	6	6	7	8	9
	7	7	7	7	7	7	7	7	7	8	9
	8	8	8	8	8	8	8	8	8	8	9
	9	9	9	9	9	9	9	9	9	9	9

Таблиця 2

Таблиця істинності 10-значної функції $\min(X_1, X_2)$

$f(X_1, X_2) = \min(X_1, X_2)$		X_2									
		0	1	2	3	4	5	6	7	8	9
X_1	0	0	0	0	0	0	0	0	0	0	0
	1	0	1	1	1	1	1	1	1	1	1
	2	0	1	2	2	2	2	2	2	2	2
	3	0	1	2	3	3	3	3	3	3	3
	4	0	1	2	3	4	4	4	4	4	4
	5	0	1	2	3	4	5	5	5	5	5
	6	0	1	2	3	4	5	6	6	6	6
	7	0	1	2	3	4	5	6	7	7	7
	8	0	1	2	3	4	5	6	7	8	8
	9	0	1	2	3	4	5	6	7	8	9

3. Апаратні платформи та побудови

Стосовно апаратної реалізації алгоритму в розглянутому прикладі МА: найближчим підходом є застосування неоднорідних систем типу «базова машина + функціонально-орієнтований процесор» (БМ+ФОП) (рис.1) [12]. ФОП – це процесори ефективної реалізації достатньо складних функцій, які мають значну питому вагу в поставлених задачах.

ФОП поділяються на програмовані та непрограмовані. Непрограмовані ФОП називають спеціалізованими процесорами. Вони обчислюють елементарні функції, здійснюють перетворення координат, тощо. Програмовані – мають свою систему команд, яка відрізняється від системи команд БМ.

Очевидно, що в основі апаратної побудови проблемно-орієнтованої системи для нашого випадку буде лежати структура «базова машина + програмований

ФОП на базі універсального двохходового десятизначного елемента» (рис. 2 а, б) [13, 14].

Оскільки алгоритм МА передбачає ряд кроків автоматичного перебору, порівняння та покрокового аналізу результатів обчислень, до апаратних засобів БМ відносяться: пам'ять словників основ та закінчень, а також оперативна пам'ять припустимих основ, припустимих закінчень та припустимої морфологічної інформації, оперативна пам'ять налагоджень універсального елемента на виконання певної логічної функції (у нашому прикладі це функція вибору максимального з двох значень). Стикування десятизначного елемента з БМ, що є на сьогоднішній день двозначною, вимагає наявності перетворювачів кодів (ПК) з двозначного коду в десятизначний та навпаки (ПК 2→10 та ПК 10→2). Процедура морфо-логічного аналізу вимагає також окремого лічильника на п'ять кроків для перевірки даних на сумісність (див.рис. 1) [15].



Рис. 1. Неоднорідна система типу «БМ+ФОП»

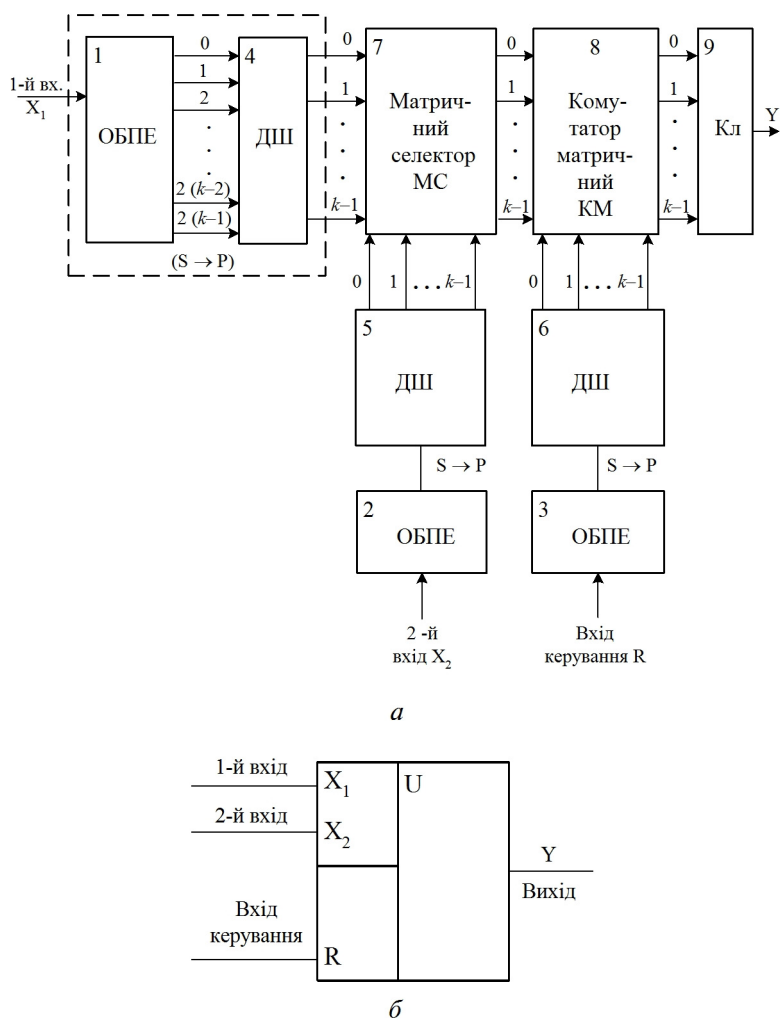


Рис. 2. Двовходовий універсальний багатозначний логічний елемент:
 а – структурна схема;
 б – умовне позначення

Таким чином, основні переваги застосування такого апаратного вирішення задачі МА полягають у наступному: підвищенні міри регулярності структури ФОРМ; скорочення термінів розроблення системи; можливості використання працівників нижчої кваліфікації під час проектування; зниження трудомісткості під час конструювання; гнучкість архітектури; спрощення ремонту та заміни модулів, що вийшли з ладу; зручність від використання скороченої експлуатаційної документації; зменшення числа типів кристалів та підвищення регулярності внутрішньої структури великих та надвеликих інтегральних схем; підвищення швидкодії за рахунок граничного паралелізму в роботі АСП-структур просторового типу.

Висновки

На підставі вивчення даних морфології української мови та за допомогою її опису у вигляді рівнянь АСП здійснено перехід до системного та структурного синтезу функціонально-орієнтованих процесорів на базі АСП-структур для аналізу та синтезу мови.

Запропоновано методику проектування лінгвістичних об'єктів (на прикладі задачі морфологічного аналізу) k -значними засобами на базі АСП-структур просторового типу та наведено інтерпретацію їх функціонування.

Наведені вище результати дозволяють зробити наступний важливий висновок: використання нових алгебро-логічних засобів моделювання природномовних конструкцій у вигляді системи рівнянь мовою АСП та явного способу завдання скінченного алфавітного оператора, який лежить в основі апаратного методу розв'язання цих рівнянь, забезпечують реалізацію властивості оборотності АСП-структур і широке розпаралелення обробки символічної інформації. Фундаментальні дослідження алгебро-логічної структури природної мови, а також алгебро-логічних засобів її моделювання у вигляді відповідних АСП-структур дають можливість впритул підійти до розв'язання важливої наукової проблеми: досягнення високоякісних технологій обробки символічної інформації на основі концепції уніфікації і методів синтезу оборотних просторових багатозначних структур мовних систем.

Список літератури: 1. Широков В. А. Елементи лексикографії. – К.: Довіра, 2005. – 304 с. 2. Широков В. А. Феноменологія лексикографічних систем. – К.: Наукова думка, 2004. – 326 с. 3. Бондаренко М. Ф., Шабанов-Кушнарченко Ю. П. Теория интеллекта. – Х.: СМІТ, 2006. – 576 с. 4. Четвериков Г. Г. Формалізація принципів побудови універсальних k -значних структур мовних систем штучного інтелекту // Доповіді НАН України. – 2001. – №1 (41). – С. 76 – 79. 5. Четвериков Г. Г. Алгебро-логічні засоби моделювання природної мови. Част. 1 // Бионика интеллекта. – 2004. – №1(61). – С. 42 – 49. 6. Четвериков Г. Г. Алгебро-логічні засоби моделювання природної мови. Част. 2 // Бионика интеллекта: науч.-техн. журнал. – 2005. – №2(62). – С. 25 – 31. 7. Четвериков Г. Г. Алгебро-логічні

засоби моделювання природної мови. Част. 3 // Бионика интеллекта. – 2007. – №1(66). – С. 48 – 52. 8. Апресян Ю. Д. Лексическая семантика. – М.: Наука, 1974. – 324 с. 9. Попов Э. В. Общение с ЭВМ на естественном языке. – М.: Наука, 1982. – 360 с. 10. Четвериков Г. Г., Бондаренко М. Ф., Шабанов-Кушнарченко Ю. П. Модуль для преобразования кодов, заданных логическими уравнениями // А. с. 1241483 СССР, МКИ Н03М 7/00 / (СССР). – №3844735/24; Заявлено 11.12.84; Опубл. 30.06.86. Бюл. №4. 11. Бондаренко М. Ф., Коноплянко З. Д., Четвериков Г. Г. Основы теории синтеза надшвидкокодирующих структур мовних систем штучного інтелекту. – К.: ІЗМН, 1997. – 264 с. 12. Функционально ориентированные процессоры / А. И. Воляхо, В. Б. Смолов, В. У. Плюснин и др. / Под ред. В. Б. Смолова. – Л.: Машиностроение, 1988. – 224 с. 13. Бондаренко М. Ф., Коноплянко З. Д., Четвериков Г. Г. Основы теории багатозначних структур і кодування в системах штучного інтелекту. – Х.: Фактор-друку, 2003. – 336 с. 14. Бондаренко М. Ф., Коноплянко З. Д., Четвериков Г. Г. Функциональный преобразователь с многозначным кодированием // Пат. 2147789 РФ, МКВ Н 03 К 19/02, Н 03 М 1/00. (Україна). – №97101717/09; Заявл. 04.02.97; Опубл. 24.04.2000. Бюл. №11. 15. Бондаренко М. Ф., Осыка А. Ф. Автоматическая обработка информации на естественном языке: Учеб. пособ. – К.: УМК ВО, 1991. – 142 с.

Надійшла до редколегії 13.02.2009

УДК 519.62

Моделирование фрагментов естественного языка алгебро-логическими средствами (на примере k -значной реализации морфологического анализа) / Г. Г. Четвериков, Т. Н. Федорова, И. Д. Вечирская, Д. Ф. Богданов // Бионика интеллекта: науч.-техн. журнал. – 2009. – № 1 (70). – С. 55–59.

В статье рассмотрены основные принципы проектирования лингвистических объектов k -значными средствами структур пространственного типа. Исследования проводились на данных морфологического анализа украинского языка средствами алгебры конечных предикатов.

Табл.: 2. Ил.: 2. Библиогр.: 15 назв.

UDK 519.62

Modeling fragments of natural language by algebra-logic means (on an example of k -unit realization of the morphological analysis) / G. G. Chetverikov, T. N. Fedorova, I. D. Vechirsky, D. F. Bogdanov // Bionics of Intelligence: Sci. Mag. – 2009. – № 1 (70). – P. 55–59.

In article main principles designing of linguistic objects are considered by k -unit means structures of spatial type. Researches were spent on the data of the morphological analysis of the Ukrainian language by means of algebra final predicates.

Tab.: 2. Fig.: 2. Ref.: 15 items.

УДК 658.012.011.56



ДОСЛІДЖЕННЯ ВІДПОВІДНОСТІ МІЖ ФОНЕТИЧНИМИ СИСТЕМАМИ ТА ГРАФІЧНИМ ПРЕДСТАВЛЕННЯМ ЛЕКСИКИ СХІДНОСЛОВ'ЯНСЬКИХ МОВ

М. Ю. Кригін

Український мовно-інформаційний фонд НАН України, м. Київ, Україна
maxus@zeos.net

Здійснено порівняння лексичного складу природної мови з його фонетичним відповідником, на основі чого зроблено висновок про ступінь відповідності між орфографічною та фонетичною системами мови. Наведено порівняльні характеристики ступеня фонетичності на рівні окремих літер, слів, а також за частинами мови. Дослідження проведено на прикладі української мови. Сформульовано певні припущення та пропозиції щодо проведення відповідних досліджень для інших східнослов'янських мов (білоруської та російської).

ЛЕКСИЧНИЙ СКЛАД, ФОНЕТИЧНА ТРАНСКРИПЦІЯ, СТУПІНЬ ФОНЕТИЧНОСТІ, ФОНЕТИЧНИЙ ВІДПОВІДНИК

Вступ

Останніми роками в Україні досить жваво обговорюються правописні питання та принципи, на яких мусять ґрунтуватися український правопис. Зокрема, постає питання, наскільки чинний правопис відповідає фонетичному принципу, тобто наскільки фонетика української мови близька до її орфографії. Актуальним також є завдання побудови систем автоматичного озвучування українських текстів, представлених в електронній формі. Цьому присвячено чимало наукових праць [1–4]; створено й реально діючі системи озвучування [5]. Тим часом проблема якості систем озвучування з огляду на адекватне передання ними української літературної вимови та наголосу досі залишається нерозв'язаною. На нашу думку, для вирішення цієї проблеми необхідно проведення більш глибоких досліджень відповідності між знаковою системою українського письма та фонетичними процесами, притаманними українській мові. Дослідження такого роду можуть і мають проводитися на різних рівнях мови, насамперед, на лексичному, який репрезентує основну, центральну підсистему мовної системи. З огляду на значний обсяг української лексики (внаслідок розвиненості словозмінної системи української мови на реєстрі приблизно в 200 тис. одиниць він сягає 4 млн одиниць) такі дослідження неминуче набудуть статистичного характеру. Оскільки всі східнослов'янські мови засновані на кирилиці, мають схожі мовні системи і правила транскрибування, ми поширили дослідження на три мови: білоруську, російську та українську. Саме статистичному дослідженню відповідності між орфографією та фонетичним принципом письма для східнослов'янських мов і присвячена ця робота, проте найбільш детально ми зупинимося на аналізі української мови.

Метою роботи є дослідити на максимально формалізованому рівні, наскільки орфографічне представлення за правилами українського правопису слова x

відповідає його фонетичній транскрипції x' за правилами транскрибування [6] для української мови, коли x пробігає деяку досить велику репрезентативну множину українських слів.

1. Принципи транскрибування текстів

Українська писемність, взагалі, заснована на засаді відповідності «звук – літера», але цей принцип витримується не абсолютно. Якщо знак фонетичної транскрипції відповідає написанню слова за правописними правилами, можемо робити висновок, що написання слова є фонетично адекватним. Якщо має місце лише часткова відповідність, то можна оцінити *ступінь фонетичності* написання, алгоритм розрахунку якого буде викладено далі.

Вважатимемо, що знаки письма відповідають знакам транскрипції, якщо вони позначаються на письмі однаково. З 33 букв українського алфавіту 27 мають відповідники у фонетичній транскрипції, і лише 6 не мають (є, ї, ю, я, ь та щ). До цих останніх слід додати і апостроф, який також не має свого відповідника у фонетичній транскрипції.

У багатьох випадках написання алфавітних літер збігається з транскрипційними знаками: а, б, в, г, ґ, д, ж, з, к, л, м, н, п, р, с, т, у, ф, х, ц, ч, ш. Якщо приголосні є пом'якшеними, за транскрипційним знаком, яким позначається літера, ставиться знак, яким позначається відповідний ступінь пом'якшення, тобто пом'якшений приголосний в транскрипції позначається принаймні двома знаками – знак, що відповідає звукові та символ пом'якшення, що йде за ним. Подовжені звуки позначаються двокрапками. Літера в у кінці складів при транскрибуванні передається як v' ; літери е, и, о мають по два транскрипційні відповідники: е та e'' , и та i'' , о та o'' відповідно; знаком й у транскрипції позначається як літера й, так і перша частина йотованих голосних; літера щ транскрипційно позначається двома літерами шч; буквосполучення дз та дж мають власні транскрипційні знаки; м'який знак позначається за до-

помогою вертикальної рисочки: '. У транскрипції також позначаються основні та другорядні наголоси. З правилами транскрибування слів української мови, які було взято за основу роботи програми транскрибування, можна ознайомитися в Орфоепічному словнику української мови [6].

2. Апарат порівняння орфографії та транскрипції

Введемо позначення. Символом $F(x)$ позначатимемо ступінь фонетичності слова x ; через $L^C(x)$ позначимо кількість символів у слові x , написання яких в орфографічному представленні збігається з транскрипційним представленням; через $L^W(x)$ позначимо кількість символів у слові x в орфографічному представленні; через $L^T(x)$ позначимо кількість символів у слові x в транскрипційному представленні. Тоді ступінь фонетичності $F(x)$ слова x природно визначити такою формулою:

$$F(x) = \frac{2L^C(x)}{L^W(x) + L^T(x)} \cdot 100\% . \quad (1)$$

Ступінь фонетичності F для всього масиву тексту вираховується як середнє $F(x)$:

$$F = \frac{\sum_{x \in G} F(x)}{N[G]} , \quad (2)$$

де $N[G]$ – кількість слів у досліджуваному масиві G .

Дослідження може бути проведене на масиві текстів (наприклад, на певному лінгвістичному корпусі) або на достатньо широкому списку словоформ. У першому випадку є можливість врахування частотності вживання тих чи інших слів і відповідно комбінацій тих чи інших літер, але при цьому деякі графемні комбінації можуть взагалі не потрапити до вибірки. У другому випадку частоти вживання слів не враховуються, але подається повна картина словоформ та морфемних комбінацій у реальних та граматично коректних словоформах. З огляду на викладене саме другий шлях було обрано для нашого дослідження. Предметом вивчення став масив українських словоформ, який для змінюваних лексем виводиться за допомогою програми парадигматизації (побудови повної словозмінної парадигми) на основі словозмінної класифікації, розробленої в Українському мовно-інформаційному фонді НАН України [7].

Обраний у якості дослідного масиву перелік словоформ літературної української мови було сформовано автоматично програмою граматичного словника, створеного в Українському мовно-інформаційному фонді НАН України, лексеми в якому проіндексовано за словозмінними класами української мови згідно із словозмінною класифікацією [7]. При цьому враховувалися зміни наголосів у похідних словоформах відповідно до акцентуаційної класифікації української мови [8]. Наголоси в початкових словоформах було розстав-

лено відповідно до Орфографічного словника української мови [9]. Загалом досліджуваний масив налічує понад 186 тис. лексем та близько 3,7 млн словоформ. Кожну з словоформ було автоматично протранскрибовано за допомогою спеціальної програми, розробленої відповідно до правил української фонетичної транскрипції, поданої в Орфоепічному словнику; зазначена програма використовується також в українському електронному словникові [10]. Таким чином, було отримано пари «словоформа – її фонетична транскрипція» у такому вигляді (див. табл. 1).

Таблиця 1

Транскрибування словоформ

мова	[мува]
мови	[муви ^е]
мові	[мув'і]
мову	[муву]
мовою	[мувойу]

Тотожність знаків письма і транскрипції визначалася за повною відповідністю або за відповідністю знаків з наголосами звичайним ненаголошеним буквам українського письма. Так, у першому рядку знак **ó** з правої колонки вважається тотожним з буквою **о** лівої колонки. Таке спрощення прийнято через відсутність знаків наголосу в українських текстах (крім деяких спеціальних: книжки для молодших школярів, навчальна література для іноземців тощо).

3. Результати дослідження

Для кожної пари «слово – транскрипція» отримано числову характеристику, яка дорівнює кількості співпадаючих символів. Це проілюстровано в табл. 2.

Таблиця 2

Приклад аналізу збігу символів у словоформі

Слово x	Транскрипція	Кількість символів у слові, $L^W(x)$	Кількість символів у транскрипції, $L^T(x)$	Кількість співпадаючих символів, $L^C(x)$
тигр	[тигр]	4	4	4
хліб	[хл'іб]	4	5	4
будь-хто	[буд'хто]	8	7	6
починається	[почи ^е на-йе ^і ц ^а :а]	11	12	5
щит	[щит]	3	4	2

Таким чином, ступінь фонетичності словоформи становить для першого рядка табл. 2

$$F(x) = \frac{2 \cdot 4}{4 + 4} \cdot 100 = 100\% ,$$

для другого – $F(x) = \frac{2 \cdot 4}{4 + 5} \cdot 100 = 88,89\% ,$

для третього – $F(x) = \frac{2 \cdot 6}{8 + 7} \cdot 100 = 80\% .$

Наголошене **é** в транскрипції ототожнюється зі звичайним **е**.

Результати порівняння для табл. 1 подано в табл. 3.

Таблиця 3

Приклади обчислення ступенів фонетичності

№	Словоформа x	Транскрипція	$L^C(x)$	$L^W(x)$	$L^T(x)$	$F(x)$, (%)
1	мова	мóва	4	4	4	100
2	мови	мóви ^е	3	4	4	75
3	мові	мóв'ї	4	4	5	88,88889
4	мову	мóву	5	5	5	100
5	мовою	мóвоюу	4	5	6	72,72727

Ступінь фонетичності окремого символу алфавіту підраховано як відношення кількості вживань певної літери, які у транскрипції передаються її орфографічним відповідником до кількості всіх вживань цієї літери. Ступені фонетичності українських літер представлено у табл. 4.

Таблиця 4

Ступені фонетичності символів у масиві українських словоформ

Сим-вол	Кількість у масиві	Кількість відповістей у масиві транскрипцій	Ступінь фонетичності
'	48015	0	0
-	68634	0	0
a	2760049	2757090	99,89
б	514549	508305	98,79
в	1913692	1460382	76,31
г	567803	563357	99,22
г	7469	7412	99,24
д	856574	734626	85,76
е	1744011	356509	20,44
є	199814	0	0
ж	217677	164120	75,4
з	712748	663714	93,12
и	2245339	557877	24,85
і	1737610	1735282	99,87
ї	105560	0	0
й	458373	458196	99,96
к	1239151	1230113	99,27
л	1382647	1374111	99,38
м	1584210	1577046	99,55
н	2401237	2254700	93,9
о	3500959	3257258	93,04
п	1181773	1174116	99,35
р	1875913	1871017	99,74
с	1507714	1428001	94,71
т	1710612	1629247	95,24
у	1441999	1426272	98,91
ф	151271	148976	98,48
х	388302	387357	99,76
ц	227980	226601	99,4
ч	459448	452804	98,55
ш	385905	369918	95,86
щ	81401	0	0
ь	686208	0	0
ю	426560	0	0
я	628415	0	0

В результаті проведення підрахунків ступеня фонетичності за формулами (1) та (2) на всьому сформованому масиві словоформ було встановлено, що українські тексти – на рівні слів – на 76,7% збігаються з власною транскрипцією.

Нерівномірність розподілу літер за їх відповідністю у транскрипції пояснюється особливістю української правописної системи. Шість вищевказаних літер (**є, ї, щ, ь, ю, я**), а також апостроф і дефіс не мають графічних відповідників у фонетичній транскрипції, хоча знак м'якшення фактично однозначно відповідає знаку вертикальної риски у транскрипції. Голосні **є, ю, я** транскрибуються у залежності від місця в слові та передаються, наприклад, через **й** та відповідний нейотований голосний або через відповідний нейотований приголосний з пом'якшенням попереднього голосного. Зрозуміло, що всі ці символи мають нулі у відповідних колонках табл. 4.

Цікавішим є факт досить низьких рівнів фонетичності для літер **в** (76,31%), **д** (85,74%), **ж** (75,4%) і особливо **е** (20,44%) та **и** (24,85%). Це пояснюється регулярним вживанням цих літер на позначення різних звуків. Зокрема, буква **в** у нашій мові вживається для вираження щільного приголосного **в** та напівголосного **ў**. Велика частотність цього останнього звуку, зокрема в дієслівних формах чоловічого роду минулого часу, зумовлює високий відсоток невідповідності букви **в** її фонетичному репрезентантові у масиві українських словоформ. Низький ступінь фонетичності літер **д** та **ж** зумовлений вживанням сполучення цих букв для позначення африкати **дж**. Нарешті, напрочуд низький ступінь фонетичності для високочастотних голосних **е** та **и** пояснюється регулярним вживанням транскрипційних символів **е^u** та **и^e**, відповідно, на позначення специфічних голосних звуків у ненаголошеній позиції.

Було також досліджено залежність фонетичності графічного репрезентанта словоформи від її частини-мовної приналежності. Результати подано в табл. 5.

Таблиця 5

Ступінь фонетичності словоформ різних частин мови

Частина мови	Загальна кількість словоформ	Збіг з транскрипцією F_x
Іменники	922707	79,323350060
Прикметники	959898	80,516120718
Дієслова	1015115	73,614291915
Прислівники	8712	81,321647214
Дієприслівники	186	72,911394735
Дієприкметники	330338	81,576514615
Числівники	1738	72,09770427
Прийменники	149	81,988834356
Частки	124	84,685310025
Сполучники	118	82,307622942
Займенники	2945	74,090968710
Вигуки	492	78,744486483
Присудкове слово	173	82,837217673
Вставне слово	31	81,956417207

Високий відсоток невідповідності з транскрипцією для дієслів та числівників неважко пояснити регулярною появою у словозмінній парадигмі форм на -ться (для дієслів) та -дять (для числівників), для яких написання лише віддалено відповідає вимові і відповідно графічній транскрипції: [ц:а] та [ц':ат'] відповідно.

Аналогічно можна провести аналіз відповідності транскрипції початкових форм слів та найбільш частотних форм, які використовуються в мові, а також відібраних за іншими критеріями груп словоформ, наприклад, термінів — в залежності від кінцевої мети дослідження.

Наявність переліку найуживаніших слів (зокрема, зафіксованих у найбільш авторитетних словниках), а також наявність системи парадигматизації української лексики, яка дозволяє отримати в явному вигляді всі непочаткові форми слів, робить цілком можливим кількісний аналіз відповідності написання українських слів їх звуковому вираженню.

Висновки

З огляду на викладене можна сказати, що ступінь фонетичності української мови досить високий, і отже, цілком реальна можливість побудови високоякісних систем для озвучування українських текстів, заснованих на простих правилах. Для цього необхідно провести статистичні дослідження розподілу транскрипційних символів (та їхніх комбінацій) в українських текстах і за результатами проведеного дослідження сформувати експериментальні масиви акустичних відповідників із урахуванням інтонаційних та синтагматичних характеристик.

Серед східнослов'янських мов найбільш фонетичною, на нашу думку, є білоруська, оскільки в ній є окремих символ в алфавіті для позначення нескладового у — ў, немає ефекту переходу о — а, е — и при вимові та передачі дзвінких приголосних глухими в усному мовленні. В російській мові, як ми прогнозуємо, спостерігається найменший серед східнослов'янських мов ступінь фонетичності за рахунок того, що в словах типу *солнце, радостный* відбувається випадіння приголосного, в словах типу *дуб, пруд* кінцевий дзвінкий вимовляється глухо, голосні о переходять на вимові в а у словах типу *пошел, родник*; спостерігаються також інші системні фонетично-правописно розходження. Більш докладному аналізу цих явищ буде присвячено окрему роботу.

Автор висловлює подяку В. А. Широкову та І. В. Шевченку, у тісній співпраці з якими була написана ця стаття.

Список літератури: 1. Сажок М. М. Автоматизовані засоби дослідження синтезу українського мовлення на основі фонемно-трифонної моделі // Автоматизовані системи управління та прогресивні інформаційні технології. Вип. І.— Київ, 2003.— С. 101–113. 2. Сажок М. М. Усномовний паспорт дик-

тора для мовленевих діалогових систем // Автоматизовані системи управління та прогресивні інформаційні технології. Вип. ІІ.— Київ, 2004.— С. 101–111. 3. Сажок М. Генерування правил розставляння наголосів у багатомовному аспекті // Праці 5-ї Всеукр. міжнар. конф. «Оброблення сигналів і зображень та розпізнавання образів» УкрОбраз'2000.— Київ, 2000.— С. 111–112. 4. Т. В. Людовик, Н. Н. Сажок. Использование речевых баз данных большого объема при синтезе речи в системах искусственного интеллекта // Проблемы управления и информатики.— Київ.— 2003.— №6.— С. 82–87. 5. Винцук Т., Людовик Т., Сажок М., Селюх Р. Автоматичний озвучувач українських текстів на основі фонемно-трифонної моделі з використанням природного мовного сигналу // Праці 6-ї Всеукр. міжнар. конф. «Оброблення сигналів і зображень та розпізнавання образів» УкрОбраз'2002.— Київ, 2002.— С. 79–84. 6. Орфоепічний словник української мови: в 2 т. / За ред. М. М. Пешак, В. М. Русанівського.— К.: Довіра, 2001. 7. Шевченко І. В. Алгоритмічна словозмінна класифікація української лексики // Мовознавство.— 1996.— № 4—5.— С. 40–44. 8. Шевченко І. В. Автоматизована дистрибуція наголосів у словозмінній парадигмі українського дієслова. // Мовознавство.— 2001.— №5.— С. 26–30. 9. Український орфографічний словник. Вид. 4-те / За ред. В. М. Русанівського.— К.: Довіра, 2005. 10. Широков В. А., Рабулець О. Г., Шевченко І. В., Костишин О. М., Якименко К. М. Інтегрована лексикографічна система «Словники України».— К., 2004.

Поступила до редколегії 16.02.2009

УДК 658.012.011.56

Исследование соответствия между фонетическими системами и графическим представлением лексики восточнославянских языков / М. Ю. Крыгин // Бионика интеллекта: науч.-техн. журнал. — 2009.— №1 (70).— С. 60–63.

Статья посвящена исследованию того, насколько близки фонетика и письменность восточнославянских языков. Построена модель степени соответствия между орфографическим и фонетическим представлением слов. Вычислительный эксперимент проводился на выборке, состоящей из словоформ, зафиксированных в грамматическом словаре украинского языка (около 3,7 млн. слов). Сделан вывод, что степень фонетичности украинской орфографии составляет более 76%. Обсуждаются аналогичные вопросы для белорусского и русского языков.

Табл.: 5. Библиогр.: 7 назв.

UDK 658.012.011.56

Research of the correspondence between phonetic systems and graphic representation of vocabulary for East Slavonic languages / M. Yu. Krygin // Bionics of Intelligence: Sci. Mag. — 2009. — №1(70).— P. 60–63.

The article is devoted research of the similarity of phonetics and written language for the East Slavonic languages. The model of correspondence degree between orthographic and phonetic representations of the words is built.

The computing experiment has been conducted on a sample that contains word forms fixed in the Ukrainian Grammatical Dictionary (about 3.7 million word forms). A conclusion is made that the degree of the correspondence of Ukrainian orthography and phonetics is more than 76%. Analogical problems for Byelorussian and Russian are discussed.

Tabl.: 5. Ref.: 7 items.

УДК 519.62



РЕШЕНИЕ СИСТЕМ КОНЪЮНКТИВНЫХ И СМЕШАНЫХ СИСТЕМ И ИХ ГРАФОВАЯ ИНТЕРПРЕТАЦИЯ

Л. Г. Ситник

Сумы, Сумской областной институт последипломного педагогического образования

В статье рассматриваются алгебраические методы решения систем логических уравнений конъюнктивного и смешанного типа, основанные на теории логических определителей. Получен общий вид решений таких систем и предложена интерпретация решений с помощью языка теории графов.

КОНЪЮНКТИВНАЯ СИСТЕМА, ДИЗЪЮНКТИВНО-КОНЪЮНКТИВНАЯ СИСТЕМА, ДЕДИЗЪЮНКЦИЯ, ИНТЕРПРЕТИРУЮЩИЙ ГРАФ

Введение

Данная статья является продолжением работы [1] на случай конъюнктивных и смешанных систем. Для них с использованием аппарата логических определителей построены частные и общие решения, показана принципиальная разрешимость таких систем. Используются интерпретации на языке теории графов.

1. Система конъюнктивных уравнений

В графовой интерпретации дизъюнктивной системы (см. формулу (3) из [1]) каждой переменной сопоставляется вершина, наделенная функцией дизъюнкции от значений вершин-соседей, связанных с ней единичными заходящими дугами. Пусть теперь каждая вершина интерпретирующего графа G будет наделена функцией конъюнкции, то есть будет принимать единичное значение, если все ее соседи уже наделены единичными значениями. Систему уравнений, сопоставляемую такому графу, будем называть конъюнктивной.

Будем пользоваться записью булевой степени x^a , означающей

$$x^a = \begin{cases} 1, a = 0 \\ x, a = 1 \end{cases} \quad (1)$$

Тогда каждое уравнение конъюнктивной системы можно представить в виде:

$$x_i = x_1^{a_{i1}} x_2^{a_{i2}} \dots x_n^{a_{in}} b_i, i = \overline{1, n}, \quad (2)$$

где знаки конъюнкции опущены [2].

Систему (2) будем условно изображать в матричной форме как

$$X = X^A B, \quad (3)$$

где A – матрица показателей степеней переменных из X , объединяемых в (2) знаками конъюнкции.

Решение системы (2), как и дизъюнктивной системы (см. формулу (3) из [1]), неоднозначно. Поэтому среди прочих будем выделять 0-решения и 1-решения, имеющие тот же смысл, что и при решении дизъюнктивной системы. Найдем их.

Записи (1) соответствует эквивалентная форма $(x \vee \bar{a})$, а поэтому, перейдя от (2) к двойственной системе относительно \bar{x}_i , с помощью теоремы 1 из [1] получим, что справедлива следующая теорема.

Теорема 1. Система конъюнктивных уравнений (2) имеет единственное 1-решение X^1 :

$$x_i^1 = \bigcap_{j=1}^n b_j^{A_{ji}^{(1)}}, i = \overline{1, n}. \quad (4)$$

На графе G определитель $|A_{ji}^{(1)}|$ интерпретируется объединением всех путей, исходящих из вершин x_j и заканчивающихся вершиной x_i . Следовательно, значением переменной x_i^1 в 1-решении системы (2) будет конъюнкция всех b_j , из которых существует хотя бы один путь в вершину x_i , то есть справедливо следующее соотношение:

$$x_i^1 = \bigcap_{b_j \in B_i} b_j, B_i \subset B, i = \overline{1, n}, \quad (5)$$

где B_i – множество всех истоков, из которых существуют пути в x_i на графе G ; B – множество всех истоков, а $x_i^1 = 1$, если $B_i = \emptyset$.

Для интерпретации 1-решения на графе G в терминах его окраски удобно использовать не конъюнктивный принцип, определяемый (5), а «дизъюнктивный», приняв истоки с нулевой краской за определяющие. Тогда в нулевой цвет должны перекраситься все те вершины, у которых хотя бы одна соседняя уже окрашена в нулевой цвет.

Под однородной конъюнктивной системой (2), в которой в отличие от дизъюнктивной $b_j = 1, j = \overline{1, n}$.

Справедлива также следующая теорема.

Теорема 2. Любое частное решение системы (2) может быть представлено покомпонентной конъюнкцией ее 1-решения и одного из решений, соответствующей ей однородной системы.

Любое решение однородной системы (2) может быть найдено с помощью G_0 тем же путем, что и решение однородной дизъюнктивной системы (см. теорему 2 из [1]), с той лишь разницей, что вершинам с петлями в графе G_c произвольным образом могут присваиваться не единичные, а нулевые значения. Множество всех частных решений системы (2), найденных указанным путем, и составляет ее общее решение.

Любое частное решение однородной системы (2) может быть найдено через его базисные решения, порожденные вектором «автонулей» с компонентами:

$$x_l = \bigcap_{k=1}^n a_k l^{|A_{kl}^{(1)}|} = |\widehat{A}_l^0|, l = \overline{1, n}. \quad (6)$$

Тогда любое частное решение однородной системы (2) суть покомпонентная конъюнкция любых наборов следующих базисных решений x_{0i}^0 с компонентами:

$$x_{0i}^0 = |\widehat{A}_i^0|^{A_{ii}^{(1)}}, i = \overline{1, n},$$

Покомпонентная конъюнкция всех базисных решений является 0-решением однородной системы (2).

Для однородной системы конъюнктивных уравнений справедлива следующая теорема.

Теорема 3. Однородная система конъюнктивных уравнений, соответствующая системе (2), имеет единственное 0-решение X_0^0 :

$$x_{0i}^0 = \bigcap_{l=1}^n \left(\bigcap_{k=1}^n a_{kl}^{|A_{ki}^{(1)}|} \right)^{|A_{ii}^{(1)}|}, i = \overline{1, n}.$$

Если граф G ацикличен, то в нем не могут возникнуть «автонули», и, следовательно, имеют место:

Следствие 1. Интерпретирующий систему (2) граф G ацикличен, если и только если 0-решение, соответствующее однородной системе, представлено только единичными компонентами.

Следствие 2. Если интерпретирующий систему (2) граф G ацикличен, то она имеет единственное решение, совпадающее с 1-решением (0-решением).

Следствие 3. Интерпретирующий систему (2) граф G ацикличен, если и только если все компоненты вектора автонулей единичны:

$$|\widehat{A}_l^0| = 1, l = \overline{1, n}.$$

2. Дизъюнктивно-конъюнктивная система уравнений

Рассмотрим теперь случай системы, которая может включать в себя как дизъюнктивные уравнения, так и конъюнктивные:

$$\begin{aligned} x_i &= a_{i1}x_1 \vee a_{i2}x_2 \vee \dots \vee a_{in}x_n \vee b_i, i = \overline{1, l}, \\ x_j &= x_1^{a_{j1}} x_2^{a_{j2}} \dots x_n^{a_{jn}} b_j, j = \overline{l+1, n}. \end{aligned} \quad (7)$$

Назовем такую систему уравнений дизъюнктивно-конъюнктивной

Соответствующий интерпретирующий граф G будет содержать дизъюнктивные вершины $x_i, i = \overline{1, l}$ и конъюнктивные вершины $x_j, j = \overline{l+1, n}$. Такой граф можно представить как результат объединения нескольких графов $G^{(k)}$, в которых каждая дизъюнктивная вершина имеет только одну заходящую дугу. Каждый из таких графов можно трактовать как сугубо конъюнктивный, полагая, что дизъюнктивная вершина с одной заходящей дугой эквивалентна такой же конъюнктивной вершине. Будем называть графы $G_v^{(k)}, v = \overline{1, r}$ конъюнк-

тивными составляющими графа G , а процедуру разложения графа G на конъюнктивные составляющие будем называть его дедизъюнкцией. Соответствующее представление исходной системы (7) в виде r систем конъюнктивных уравнений также будем называть ее дедизъюнкцией [2]. Эти системы будем называть конъюнктивными составляющими дедизъюнкции системы (7). Имеет место следующая лемма.

Лемма. Для всякого решения системы (7) найдется конъюнктивная составляющая ее дедизъюнкция, допускающая тоже решение.

Доказательство этой леммы строится на той же основе, что в графе G можно у дизъюнктивных вершин, принявших нулевое значение, оставить любую одну заходящую дугу, а у принявших единичное значение оставить только одну заходящую дугу, связанную с любой из вершин, также принявших единичные значения.

Используя дедизъюнкцию системы (7) можно показать, что справедлива следующая теорема.

Теорема 4. 0-решение (1-решение) системы (7) есть покомпонентная дизъюнкция 0-решений (1-решений) конъюнктивных составляющих ее дедизъюнкций.

Доказательство этой теоремы основывается на том, что ни 0-решения, ни 1-решения конъюнктивных составляющих не могут иметь больше единичных компонент, чем исходная система, и эти компоненты не могут быть новыми. А то, что 0-решение и 1-решение исходной системы совпадает с каким-либо из решений конъюнктивной составляющей, определяется сформулированной выше леммой.

Следствие 4. Если интерпретирующий систему (7) граф G ацикличен, то он имеет единственное решение, совпадающее с 0-решением (1-решением).

При проверке графа G на ацикличность нет необходимости учитывать свойства его вершин (дизъюнктивность, конъюнктивность). Достаточно лишь исследовать матрицу A , используя следствие 3 или следствие 4, эквивалентные в силу их двойственности.

Пример 1. Для построения графов $G_v^{(k)}$ можно воспользоваться следующей процедурой. Пусть вместо матрицы A система представлена D -матрицей, отображающей дизъюнктивные связи, и K -матрицей, отображающей конъюнктивные связи в графе G . Например:

$$D = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, K = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

Выполним лексикографическую дизъюнкцию D -матрицы:

$$\begin{aligned} \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} &= \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \vee \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \vee \\ &\vee \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \vee \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

Тогда исходящий граф представим следующим набором конъюнктивных графов (см. рис.1) с матрицами смежностей:

$$G_1^{(k)} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, G_2^{(k)} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix},$$

$$G_3^{(k)} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, G_4^{(k)} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

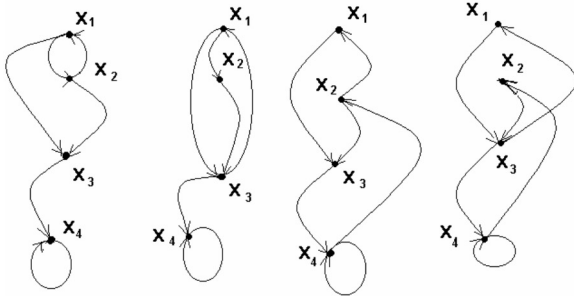


Рис.1. Набор конъюнктивных графов

Теперь можно убедиться, что соответствующие этим графам системы уравнений имеют только нулевые решения при любых B , и, следовательно, 0-решение в этом примере при любых B также нулевое.

Пример 2. (использование решения как формулы).

В системах с искусственным интеллектом предметная область зачастую представляется семантической сетью в виде И-ИЛИ-графа, в котором вершинам типа ИЛИ сопоставляются понятия, а вершинам типа И — отношения (модули), отображающие одни понятия и другие. Если одно и то же понятие может быть сформулировано различными модулями, то в соответствующую ему вершину типа ИЛИ будут заходить несколько дуг.

В СС решаются различные задачи, в том числе :

- выводимо (вычислимо) ли требуемое понятие по заданному множеству исходных понятий;
- какие из модулей принимают участие в конкретном выводе.

Для этой цели СС сопоставим граф, в котором дизъюнктивные вершины отмечаются именами понятий, а конъюнктивные — именами отношений.

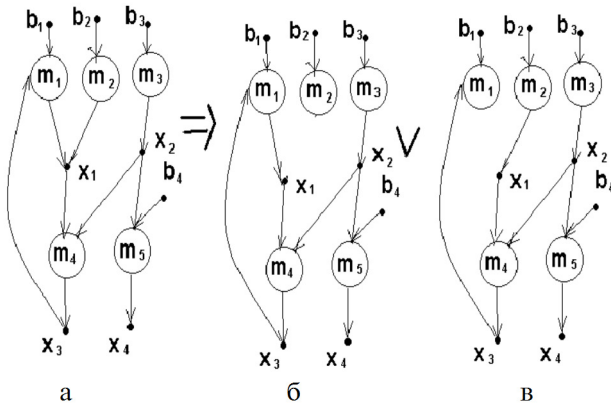


Рис.2. Структура семантической сети:
а — граф семантической сети;
б, в — конъюнктивные компоненты графа

На рис. 2а представлен такой граф, причем точками изображены дизъюнктивные вершины, а кружочками — конъюнктивные. После дедизъюнкции этого графа получим две его конъюнктивные компоненты $G_1^{(k)}$ и $G_2^{(k)}$ (рис. 2б и 2в), потенциально содержащие различные варианты решений задач на СС.

Конъюнктивные компоненты преобразуем в интерпретирующие графы G_1 и G_2 следующим образом.

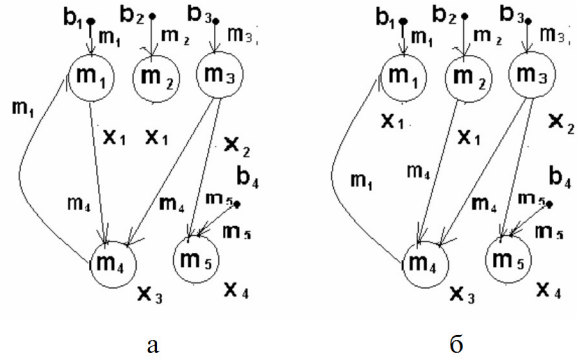


Рис.3. Интерпретирующие графы:
а — связь $m_1 \rightarrow m_4$; б — связь $m_2 \rightarrow m_4$

Имена отношений приписываются дугам, заходящим в конъюнктивные вершины, дизъюнктивные вершины исключаются, их именами помечаются предшествующие дизъюнктивным вершинам (рис.3а и 3б). При таком построении некоторые из конъюнктивных вершин будут помечены одними и теми же именами понятий. Поэтому матрицы смежности для полученных графов будем записывать относительно уникальных индексов, сопоставленных вершинам модулей:

$$A_1 = \begin{pmatrix} 0 & 0 & 0 & m_1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ m_4 & 0 & m_4 & 0 & 0 \\ 0 & 0 & m_5 & 0 & 0 \end{pmatrix}, A_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & m_4 & m_4 & 0 & 0 \\ 0 & 0 & m_5 & 0 & 0 \end{pmatrix},$$

где дуги представляются приписанными им именам модулей, причем значение $m_i = 1, i = \overline{1,5}$. Пусть названные выше задачи решаются относительно понятий x_3 и x_4 , или, что то же, относительно m_4 и m_5 . Ответы на поставленные вопросы можно получить, найдя 0-решения для графов G_1 и G_2 . Но так как G_1 содержит цикл, включающий m_4 , то он не может быть использован для вывода m_4 . Остается использовать G_2 , для которого получим:

$$x_3 = (m_1 b_1)^0 (m_2 b_2)^{m_4} (m_3 b_3)^{m_4} (m_5 b_4) = (m_2 b_2 m_3 b_3)^{m_4},$$

$$x_4 = (m_1 b_1)^0 (m_2 b_2)^0 (m_3 b_3)^{m_5} (m_5 b_4)^{m_5} = (m_3 b_3 m_5 b_4)^{m_5}.$$

Таким образом, x_3 вычислимо, если определены b_2 и b_3 , а участвуют в вычислениях модули m_2, m_3 и m_4 . Для вычисления x_4 достаточно определить b_3 и b_4 , а в вычислениях участвуют модули m_3 и m_5 .

Решения систем логических уравнений, рассмотренные в данной работе и статье [1], были реализова-

ны программно. Если проанализировать соответствующие формулы решений, то не сложно заметить, что для их программной реализации необходимо вычислить булев определитель. Решение данной задачи может быть реализовано по следующему алгоритму.

Шаг 1. Ищется элемент определителя $A(i, j)$ равный 1 и запоминается номер его строки и столбца.

Шаг 2. Осуществляется поиск 1 в столбце $j+1$ и во всех строках, кроме i -той. Если единичный элемент отсутствует, то определитель равен нулю, в противном случае шаг 2 повторяется до тех пор, пока не будет найдена последовательность из n единиц, каждая из которых стоит в отдельном столбце и отдельной строке. Это означает, что булев определитель равен 1.

Выводы

В статье предложен метод решения конъюнктивных и смешанных систем логических уравнений матричными методами, позволивший применять известные аналитические методы решения линейной алгебры (аналог метода Крамера). Разработан общий алгоритм нахождения решения систем логических уравнений и его графоаналитическая интерпретация.

Для частных случаев решения систем логических уравнений предложены понятия: «0-решение» и «1-решение» этих систем, позволяющие представить в компактном виде структуру общего решения.

Список литературы: 1. Ситник Л. Г., Шабанов-Кушнаренок С. Ю. Решение систем дизъюнктивных уравнений методом логических определителей // Бионика интеллекта: науч.-техн. журнал.—2008.—№1(68). 2. Чистов В. П. Аналитические решения логических уравнений // Техническая кибернетика (Изв. АН СССР).—1994.—№ 2.—С. 219–224.

Поступила в редколлегию 18.02.2009

УДК 519.62

Розв'язання систем кон'юнктивних і змішаних систем та їх графова інтерпретація / Сітнік Л. Г. // Біоніка інтелекту: наук.-техн. журнал.—2009.—№1(70).—С. 64–67.

У статті розглядається питання про розв'язання систем логічних рівнянь за допомогою логічних визначників. Знайдено вигляд розв'язків даних систем і запропонована інтерпретація з використанням теорії графів.

Лл.: 3. Бібліогр.: 2 найм.

UDK 519.62

Decision of the systems of the conjunct and mixed systems and their graphs interpretation / Sitnik L. G. // Bionics of Intelligence: Sci. Mag. —2009. —№1(70).—P. 64–67.

In the article a question is about the decision of the systems of logical equations through logical determinants. The general view of decisions of these systems is found and interpretation with the use of theory of graphs offered.

Fig.3.: Ref.: 2 items.

УДК 004.896, 004.934



МЕТОД АВТОМАТИЧЕСКОГО ОПРЕДЕЛЕНИЯ ГРАНИЦ ШУМОВОГО ТУРБУЛЕНТНОГО ИСТОЧНИКА В ЗВУКОВОМ СИГНАЛЕ

Хашан Т.С.

Институт прикладной математики и механики НАН Украины, г. Донецк, khashan@mail.ru

В статье изложен метод автоматического поиска границ шумового турбулентного источника в звуковом сигнале. Предложенный метод основан на многомерном анализе энергетической завершенности исследуемых отрезков сигнала, сравнение последовательностей его характеристик и их сопоставлении. На основе рассматриваемого метода разработан алгоритм и проведены численные исследования.

ШУМОВОЙ ТУРБУЛЕНТНЫЙ ИСТОЧНИК, СЕГМЕНТАЦИЯ, РАСПОЗНАВАНИЕ РЕЧИ, АМПЛИТУДНО-ВОЛНОВОЕ ПРЕДСТАВЛЕНИЕ ВОЛНОВЫХ ПРОЦЕССОВ

Введение

Одним из важных направлений исследований в области искусственного интеллекта является разработка интеллектуальных систем автоматического распознавания звуковых сигналов (ЗС). Качество работы этих систем зависит от первоначальной обработки сигнала и подготовки его к дальнейшим исследованиям. Данная статья является продолжением исследований, связанных с разработкой методов и алгоритмов для интеллектуальной системы технического слуха малых мобильных роботов [1-3], одной из подсистем которой является система автоматической сегментации ЗС.

Следует отметить, что под сегментацией ЗС подразумевают определение границ фрагментов сигнала, на котором могут присутствовать голосовой, шумовой турбулентный, шумовой нетурбулентный источники, а также голосовой и шумовой источники, функционирующие одновременно [4].

Ранее в [2-6] было отмечено, что на сегодняшний день все разрабатываемые системы автоматического сегментации ЗС являются дикторозависимыми, так как идеология их построения основывается на анализе спектральных характеристик, определенных на окнах фиксированной длины (фреймах). При этом размер окна, а также его границы не зависят от природы сигнала. Такой подход приводит к назначению порогов, на основании которых принимается решение о наличии границы между источниками различной природы [7-10]. Очевидно, что заданные значения порогов одного и того же участка для одного и того же диктора изменяются в широких пределах.

Поэтому актуально создание дикторонезависимых методов и алгоритмов автоматического определения границ фрагментов ЗС, соответствующих одному из вышеприведенных источников.

В работах [4-6] предложена идея, позволяющая отказаться от окон фиксированной длины, основанная исключительно на анализе амплитудно-временного представления волнового процесса (АВП ВП) и энер-

гетической завершенности исследуемых фрагментов (ЭЗФ), выдвигении гипотез и проверке их истинности с учетом факторов закономерности и случайности поведения характеристик сигнала на исследуемых ЭЗФ.

Эта идея развита в [2,3], где решена задача автоматического выбора границ фрагментов сигнала, на котором функционирует голосовой источник. В [5] предложен алгоритм, позволяющий выделять сегменты сигнала, соответствующие функционированию шумового турбулентного источника. Однако анализ данного метода показал, что он эффективен лишь в том случае, когда в исследуемом сигнале присутствует только один шумовой источник, либо характеристики других источников близки к первому найденному. Кроме того, границы искомым фрагментов определяются с ошибками. На основании сказанного сформулируем цель и задачи данной работы.

Целью данной работы является усовершенствование дикторонезависимого метода и алгоритма автоматического определения границ фрагментов речевого сигнала, соответствующих функционированию шумовых турбулентных источников.

Для достижения поставленной цели необходимо решить следующие задачи:

- проанализировать метод, представленный в [5];
- на основании анализа усовершенствовать метод автоматического выделения фрагментов АВП ВП, соответствующих функционированию шумовых турбулентных источников (ШТИ);
- программно реализовать алгоритм и провести численные исследования.

1. Метод автоматического определения границ фрагментов, соответствующие функционированию ШТИ. Анализ его эффективности

В основе метода решения поставленной задачи лежит анализ функции $U = \{u_n\}_{n=1}^N$ – АВП ВП (рис.1) и ее характеристик: $D = \{d_i\}_{i=1}^M$ – длин полных колебаний (ПК), $L = \{[l_1^i, l_2^i, l_3^i, l_4^i, l_5^i, l_6^i]\}_{i=1}^M$ – границ ПК,

$R = \{r_i\}_{i=1}^M$ – размахов амплитуд на ПК (рис.2), где M – количество ПК функции U , а также энергетически завершенных фрагментов R_j^1 (ЭЗФ) с границами $\{\{lr_1^j, lr_2^j, lr_3^j, lr_4^j, lr_5^j\}\}$, где $j = \overline{1, J}$ – количество ЭЗФ функции U (рис.3). Отметим, что характеристики функции U получены согласно правилам, представленным в [2].

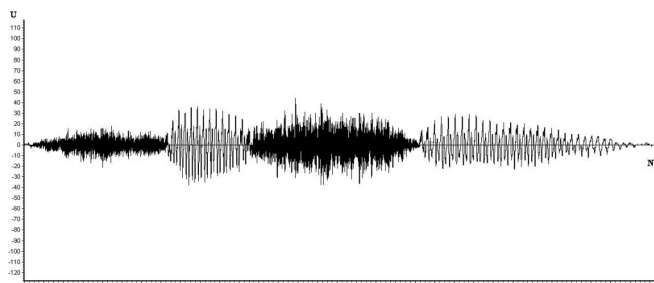


Рис.1. АВП ВП слова «СУШУ»

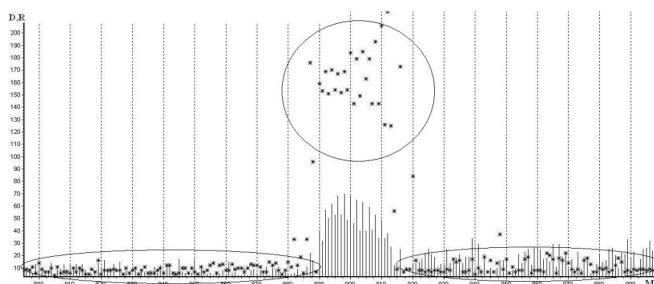


Рис. 2. Фрагмент «СУШ» слова «сушу» в виде $\{d_i, r_i\}_{i=1}^N$; d_i представлено знаком *, r_i – знаком |

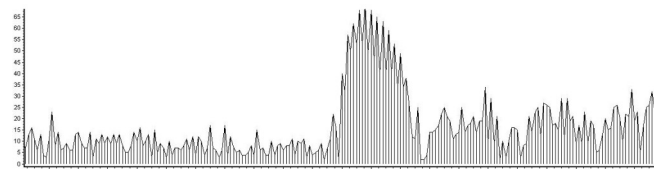


Рис. 3. Фрагмент «СУШ» слова «сушу», представленного R_j^1

Анализируя характеристики функции U на фрагментах, соответствующих функционированию ШТИ и его окружения, в [5] были сделаны следующие выводы: длины ПК, принадлежащих фрагментам ШТИ, меньше длин ПК, не принадлежащих объекту (см. рис. 2); длины ПК, принадлежащих фрагментам ШТИ, близки.

Однако, как было показано авторами [5], невозможно определить близость ПК на исследуемых фрагментах, так как они ведут себя случайным образом. Анализ ЭЗФ также подтверждает случайность поведения характеристик. Поэтому для описания поведения длин ПК вводятся:

– функция F , которая отражает количество ПК в единицу времени:

$$F = \{F(R_j^1) : F(R_j^1) = \frac{N(R_j^1) \cdot f}{D(R_j^1)}, j = \overline{1, J}\}, \quad (1)$$

где $N(R_j^1)$ – количество ПК в ЭЗФ R_j^1 ; $D(R_j^1)$ – дли-

на ЭЗФ R_j^1 ; f – частота дискретизации сигнала по времени; J – количество ЭЗФ.

– множество

$$G^1 = \{R_k^1 : F(R_k^1) > \frac{(\overline{M} + \overline{m})}{2},$$

$$k = \overline{k', k''}, F(R_{k-1}^1) \leq \frac{(\overline{M} + \overline{m})}{2}, \quad (2)$$

$$F(R_{k+1}^1) \leq \frac{(\overline{M} + \overline{m})}{2}\},$$

где $\overline{M} = \max_{1 \leq j \leq J} (F(R_j^1))$ и $\overline{m} = \min_{1 \leq j \leq J} (F(R_j^1))$.

Значения функции $F(R_k^1)$, принадлежащие фрагментам ШТИ, будут больше, чем значения функции $F(R_{k+1}^1)$ и $F(R_{k-1}^1)$, принадлежащие их окружению, $R_{k+1}^1 \notin G^1$ $R_{k-1}^1 \notin G^1$ соответственно время существования ЭЗФ

$$R_k^1 < R_{k+1}^1 \text{ и } R_k^1 < R_{k-1}^1.$$

$$R_k^1 \in G^1 \quad R_{k+1}^1 \notin G^1 \quad R_k^1 \in G^1 \quad R_{k-1}^1 \notin G^1$$

Из определения множества G^1 следует, что

$$F(R_s^1) \in \left[\min_{R_k^1 \in G^1} F(R_k^1); \overline{M} \right], F(R_s^1) \in G^1.$$

Алгоритм, разработанный на основании данного метода позволяет «близко» определить границы объектов, соответствующие ШТИ. На рис. 4 показан пример функционирования алгоритма на примере слова «силища». Типизация выделенных фрагментов (их объединения с учетом случайности поведения функции на соседних фрагментах) (рис. 5а, б) устраняет «провалы и всплески», однако не уточняет границы искомого фрагмента.

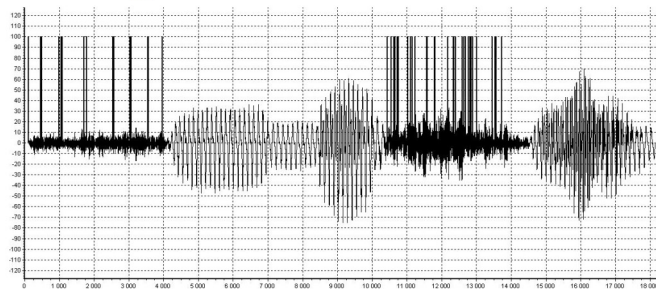


Рис. 4. Фрагменты ШТИ, выделенные в результате работы алгоритма (слово «силища»)

2. Усовершенствование метода автоматического определения границ ШТИ в ЗС

Для уточнения границ фрагментов, соответствующих функционированию ШТИ, проведем анализ последовательностей ЭЗФ первого и второго уровней, а также ПК для каждой из последовательностей.

Отметим, что последовательности ЭЗФ второго уровня строятся из последовательностей ЭЗФ первого уровня следующим образом: выбираются максимальные значения размахов амплитуд на R_j^1 и анало-

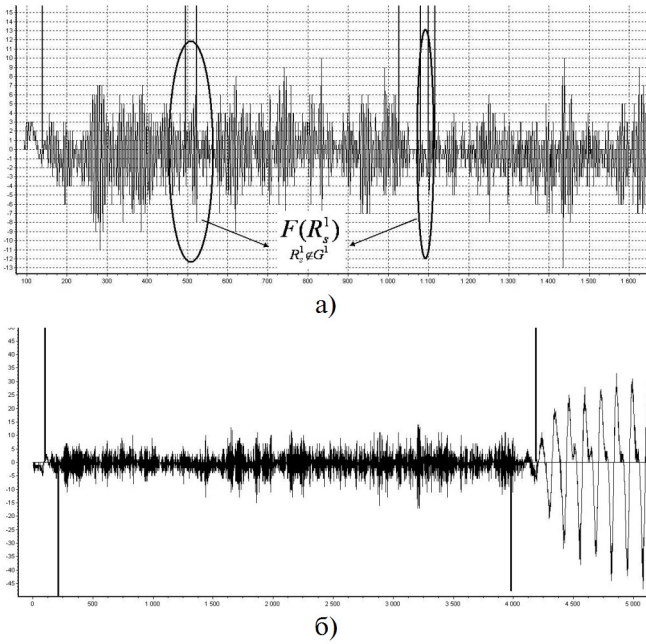


Рис. 5. Фрагмент АВП ВП, соответствующий функционированию ШТИ: а) до типизации объектов (звук «с»); б) после типизации объектов (границы после типизации отмечены знаком |, нижняя часть графика, «си»)

гично, только уже для второго уровня, из этих значений строится ЭЗФ уровня 2 – R_j^2 .

$$R^2 = (r_{i_2}^2 : r_{i_2}^2 = \max_{r_{\mu}^1 \in R_j^1} (r_{\mu}^1), \mu \in [lr_1^{j,1}; lr_5^{j,1}], i_2 = \overline{1; M_2}) \quad (3)$$

$$\begin{aligned} R_{j_2}^2 : r_{i-1}^2 > r_i^2 = r_{i+p_1}^2, \\ r_{i+p_1}^2 < r_{i_2}^2 \leq r_{i_2+p_2}^2 < r_{i_3}^2 = r_{i_3+p_3}^2, \\ r_{i_3+p_3}^2 > r_{i_4}^2 \geq r_{i_4+p_4}^2 > r_{i_4+p_4}^2, \\ r_{i_4+p_4}^2 = r_{i_5}^2 = r_{i_5+p_5}^2 < r_{i_5+p_5+1}^2, \end{aligned} \quad (4)$$

где $\exists s1 \in N : \forall p1 = \overline{0, s1}$, $s2 \in N : \forall p2 = \overline{0, s2}$,

$$s3 \in N : \forall p3 = \overline{0, s3}, \quad s4 \in N : \forall p4 = \overline{0, s4},$$

$$s5 \in N : \forall p5 = \overline{0, s5}, \quad j_2 = \overline{1, J_2},$$

J_2 – количество ЭЗФ второго уровня;

M_2 – количество ПК и, соответственно, размахов амплитуд, составляющих уровень 2.

Рассматривая длины ПК на R_j^1 , а также определяя их близость согласно условиям (5) и (6), принимается решение о выборе границ искомым фрагментов. В случае, когда эти условия не выполняются (то есть на таких ЭЗФ фиксируются «всплески» или «провалы») определяется количество «близких» между собой ПК внутри ЭЗФ. Если их большинство, то принимается решение о случайности поведения функции U на этих

участках (рис. 5б – границы отмечены знаком |, верхняя часть рисунка).

$$\begin{aligned} l_1^{i+k1} &\leq l_{\max}^i + d_i + S < l_4^{i+k1}, \\ l_3^{i+k1} &\leq l_{\min}^i + d_i + S < l_6^{i+k1}, \\ l_1^i &\leq l_{\max}^{i+k1} - d_{i+k1} - S < l_4^i \leq l_{\min}^{i+k1} - d_{i+k1} - S < l_6^i. \end{aligned} \quad (5)$$

$$\begin{aligned} 0 < \frac{r_i^+}{2} \leq r_{i+k1}^+, \quad 0 < \frac{r_{i+k1}^+}{2} \leq r_i^+, \\ r_{i+k1}^- \leq \frac{r_i^-}{2} < 0, \quad r_i^- \leq \frac{r_{i+k1}^-}{2} < 0. \end{aligned} \quad (6)$$

$S = l_1^{i+k1} - l_6^i$ – расстояние между ПК.

Дополнительным условием принадлежности к искомым фрагментам является условие (7) (это условие выполнимо для большинства сигналов, однако оно не является обязательным):

$$G^m = (r_m^1 : r_m^1 < \frac{\sum_{i=1}^M r_i^1}{M}, m = \overline{m_1, m_2}). \quad (7)$$

На отрезках АВП ВП, определенных вышеуказанным способом, строится ЭЗФ 2 уровня – условия (3), (4). Далее на нем уточняется поведение функции $F(R_{j_2}^2)$.

$$F^2 = \{F(R_{j_2}^2) : F(R_{j_2}^2) = \frac{N(R_{j_2}^2) \cdot f}{D(R_{j_2}^2)}, j_2 = \overline{1, J_2}\}, \quad (8)$$

где $N(R_{j_2}^2)$ – количество ПК в ЭЗФ $R_{j_2}^2$; $D(R_{j_2}^2)$ – длина ЭЗФ $R_{j_2}^2$; f – частота дискретизации сигнала по времени; J_2 – количество ЭЗФ второго уровня.

Очевидно, что значения функции $F(R_{k_2}^2)$, принадлежащие области $R_{k_2}^2 \in G^2$

лежащие фрагментам ШТИ, будут больше, чем значения функций $F(R_{k_2+1}^2)$ и $F(R_{k_2-1}^2)$, принадлежащих его окружению. Область G^2 строится следующим образом:

$$\begin{aligned} G^2 &= \{R_{k_2}^2 : F(R_{k_2}^2) > \frac{(\overline{M_2} + \overline{m_2})}{2}, \\ k_2 &= \overline{k'_2, k''_2}, F(R_{k_2-1}^2) \leq \frac{(\overline{M_2} + \overline{m_2})}{2}, \\ F(R_{k_2+1}^2) &\leq \frac{(\overline{M_2} + \overline{m_2})}{2}\}, \end{aligned} \quad (9)$$

где $\overline{M_2} = \max_{1 \leq j_2 \leq J_2} (F(R_{j_2}^2))$ и $\overline{m_2} = \min_{1 \leq j_2 \leq J_2} (F(R_{j_2}^2))$.

На основании предложенного метода был разработан алгоритм и проведены следующие численные исследования.

3. Численные исследования

Основные параметры: частота дискретизации $f = 22050$ Гц, 8-битный сигнал, запись сигнала производилась при помощи программы SoundForge.

Аналізу подвергались около 300 слов, произнесенных 10 дикторами (3 женщины и 7 мужчин, возраст от 22 до 60 лет). Во всех случаях границы фрагментов ШТИ были выявлены точно и совпадают с результатами сегментации ЗС оператором вручную на основании прослушивания соответствующих фрагментов. Это свидетельствует о высоком показателе по сравнению с базовым методом. На рис.6 показан результат функционирования модифицированного алгоритма на примере слова «лошак» (диктор – мужчина, 23 года). Как видно из рисунка, алгоритм точно определил все фрагменты, соответствующие ШТИ.

Следует отметить, что в данном методе не устранен пока недостаток неразделимости двух подряд идущих фрагментов, соответствующих ШТИ, например слово «шить» (рис.7). Этот недостаток поясняется тем, что характер поведения функции F на данных фрагментах, а также все характеристики U идентичны.

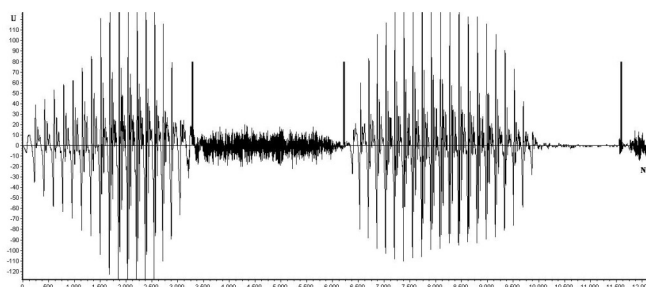


Рис. 6. Выявленные на АВП ВП фрагменты ШТИ слова «лошак»: границы фонем «ш» и «к»

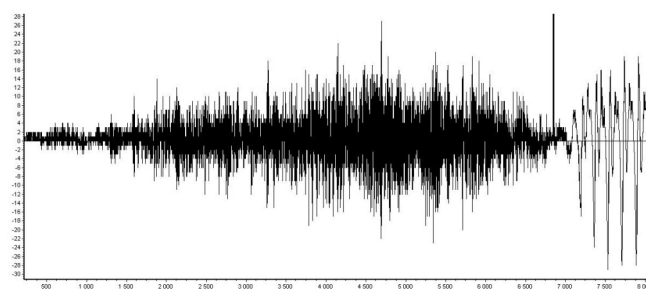


Рис. 7. Фрагмент АВП ВП «СШИ» слова «шить»

Выводы

В статье предлагается модифицированный метод выделения границ фрагментов звуковых сигналов, соответствующих функционированию шумового турбулентного источника. Предлагаемый метод не зависит от диктора, а также от фонетического и фонологического состава. В отличие от базового метода [5], подход, описанный в данной статье, позволяет выделить все искомые объекты, а также точно определить их границы.

На основании базового и предложенного методов, представляется возможным перейти к решению задач по автоматическому определению границ фрагментов

АВП ВП, соответствующих функционированию шумового нетурбулентного источника, а также голосового и шумового источников, функционирующих одновременно. Решение указанных задач позволит в дальнейшем перейти к решению задачи распознавания звуковых сигналов.

Список литературы: 1. *V.E.Pavlovsky, A.S.Myagkov, T.S.Khashan, V.V.Pavlovsky.* Concept, Simulation and Elaboration of Audition Sensors for Robots // Proc. of The IARP Workshop «Adaptive and Intelligent Robots: Present and Future». – Moscow, Nov 24-26 2005. – IPMech RAS, RAS. – vol.1. – pp. 90–100. 2. *Хашан Т.С.* Сегментация речевого сигнала // Искусственный интеллект. – 2007. – №.4. – С. 386–397. 3. *Хашан Т.С., Писаренко Ю.В.* Алгоритм пеленгаии речевых и специальных сигналов системой технического слуха в интеллектуальной робототехнике // КЗИМ. – 2008. – №.7. – с. 66–75. 4. *Мышко С.В., Блохин С.В., Доценко А.В.* Определение на амплитудно-временном представлении звуковых волн элементарных и однородных квазипериодических структур // Биофизика сложных систем. – 1993. – С. 92–100. 5. *Доценко А.В., Мышко С.В., Шевченко А.И., Блохин С.В.* Выделение фрагментов амплитудно-временного представлении звуковых волн, являющихся результатом функционирования шумового турбулентного источника // Биофизика сложных систем. – 1993. – С. 85–92. 6. *Мышко Н.В., Доценко А.В., Блохин С.В., Мышко С.В.* Определение энергетически завершенных фрагментов амплитудно-временного представления звуковых волн // Биофизика сложных систем. – 1993. – С. 103–117. 7. *Старушко Д.Г., Шелепов В.Ю.* Новая система признаков для распознавания речевых единиц // Искусственный интеллект. – 2002. – № 4. – С. 286–288. 8. *Yermolenko T.V., Fedorov E.E.* The procedures of the noise clipping in the signal, based on Fourier- and wavelet-transform and on classification of sounds of speech // Труды межд. конф. SPECOM'2006. – 2006. – P.245–250. 9. *Карпов О.Н.* Технология построения устройств распознавания речи: Моногр. – Днепропетровск: Из-во Днепропетровского ун-та, 2001. – 184 с. 10. *Шелепов В.Ю., Ниценко А.В.* Амплитудная сегментация речевого сигнала, использующая фильтрацию и известный фонетический состав // Искусственный интеллект. – 2003. – № 3. – С. 421–426.

Поступила в редколлегию 19.02.2009

УДК 004.896, 004.934

Метод автоматичного визначення границ шумового турбулентного джерела у звуковому сигналі / Т. С. Хашан // Біоніка інтелекту: наук.-техн. журнал. – 2009. – №1(70). – С. 68–71.

Запропоновано вдосконалений метод виділення границь фрагментів звукового сигналу, що відповідають функціонуванню шумового турбулентного джерела. Запропонований метод не залежить від диктора, а також від фонетичного й фонологічного складу.

Л.: 7. Бібліогр.: 10 найм.

UDC 004.896, 004.934

Method for automatically determining the boundaries of the turbulent noise source in audio signal / Khashan Tatiana S. // Bionics of Intelligence: Sci. Mag. – 2009. – №1(70). – P. 68–71.

Proposed an improved method for the speech signal fragment's boundaries allocation, which corresponds to the turbulent noise source activity. The proposed method is a speaker-independent and does not depends from phonetic and phonological structure of the speech.

Fig.: 7. Ref.: 10 items.

УДК 681.513



РОБАСТНОЕ УПРАВЛЕНИЕ ДИНАМИЧЕСКИМ ПРОЦЕССОМ НА ОСНОВЕ НЕЧЕТКОЙ МОДЕЛИ

С. Г. Удовенко, Г. Дибе

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

Приведен метод нечеткого управления динамическим процессом, основанный на использовании дополнительных аддитивных составляющих. Предложенный подход позволяет повысить робастность нечеткого цифрового управления по отношению к неопределенностям, характерным для нечеткой системы. Такая робастность обеспечивает компенсацию или снижение влияния неопределенностей и возмущений на качество управления. Результаты моделирования подтверждают работоспособность рассмотренного в настоящей работе нечеткого робастного регулятора.

НЕЧЕТКАЯ ЛОГИКА, РОБАСТНОСТЬ, АППРОКСИМАЦИЯ НЕЛИНЕЙНОСТИ, ОШИБКА НЕЧЕТКОГО УПРАВЛЕНИЯ

Введение

В последнее время получили широкое распространение интеллектуальные системы управления детерминированными и стохастическими объектами, реализуемые с помощью универсальных или специализированных микропроцессорных средств. Однако успешные попытки практического применения таких систем связаны в основном с задачами управления линейными объектами, решение которых основано на компьютерной реализации линейных цифровых регуляторов. В то же время линейная аппроксимация динамики целого ряда технических объектов является недостаточно корректной (например, для объектов робототехники, мехатроники и так далее).

С математической точки зрения можно рассмотреть широкий спектр нелинейных процессов (от детерминированных до хаотических) и использовать для дискретного управления такими процессами различные методы. Однако на практике к реальным системам цифрового управления предъявляются требования, ограничивающие возможность получения точных решений в реальном масштабе времени. В связи с этим для синтеза нелинейных цифровых регуляторов получили распространение методы вычислительного интеллекта, в том числе применение нечеткой логики. Эти методы позволяют рассмотреть модель нелинейного объекта как некоторую номинальную модель, в которой учтены неопределенности. В этом случае задача цифрового управления сводится к достижению робастности синтезируемых регуляторов к учитываемым типам неопределенностей. Решение проблемы создания цифровых регуляторов для сложных многопараметрических объектов (в энергетике, химической технологии и так далее) может быть основано на описании динамики этих объектов нечеткими моделями, в которых зависимости между входными, выходными переменными и переменными состояниями представляются с помощью нечетких правил [1]. Задачи нечеткой идентификации и нечеткого управления связаны

с необходимостью применения различных методов аппроксимации нечетких систем [2]. Несмотря на определенные теоретические достижения в области синтеза нечетких систем управления, остаются неисследованными важные аспекты их практического применения, в том числе:

- обоснование гарантированной устойчивости нечетких систем;
- обеспечение робастности нечетких систем к различным типам неопределенностей [3].

В настоящей работе исследуется возможность создания нечетких цифровых регуляторов, реализуемых на базе микропроцессорной техники, обладающих свойствами гарантированной устойчивости и робастности по отношению к некоторым видам неопределенностей.

1. Постановка задачи

Любая модель физического процесса основана на использовании аппроксимации, в той или иной мере учитывающей неопределенности. Теорема нечеткой аппроксимации подтверждает возможность представления динамики нелинейных процессов с помощью нечеткой системы. Однако такое представление может потребовать использования очень большого числа нечетких правил. Для преодоления этих трудностей необходимо провести анализ допустимых ошибок моделирования. В этом случае нелинейный процесс можно будет представить номинальной нечеткой моделью, количество правил в которой является приемлемым и фиксированным априори, а ошибки моделирования рассматриваются как некоторая неопределенность. Представление неопределенности может иметь различные формы в зависимости от характера параметра информации, которая должна быть определена. В общем случае, такое представление должно быть основано на знании физических причин, обуславливающих неопределенности, и объема информации, необходимой для описания неопределенностей в удобном для практического использования виде. Неопределенно-

сти, присутствующие в системах управления, можно разделить на два больших класса: неструктурные неопределенности и структурные неопределенности. Неструктурные неопределенности наиболее характерны для моделирования систем. Они позволяют установить условия робастности и определить необходимые процедуры для синтеза регуляторов. Как правило, неструктурные неопределенности могут быть представлены в моделях аддитивными, мультипликативными на входе или мультипликативными на выходе составляющими. Такая форма представления неопределенностей названа неструктурной, потому что единственной используемой здесь информацией является верхнее ограничение неопределенности. В данной работе будем рассматривать все неопределенности моделирования в общей структуре, полагая, что нечеткая модель процесса определяется суммой номинальной нечеткой модели и некоторой дополнительной составляющей Δ , отражающей неопределенности и удовлетворяющей заданному неравенству:

$$|\Delta| < \Delta_0.$$

В общем случае такой тип представления неопределенности используется для того, чтобы учесть внутренние явления (такие, как параметрические ошибки и ошибки аппроксимации, связанные, например, с линеаризацией). Аддитивное представление используется для отображения внутренних неопределенностей (параметрических ошибок и ошибок аппроксимации). Учитывая, что при нечетком моделировании основным источником ошибок является неточность аппроксимации, аддитивное представление неопределенностей является наиболее приемлемым для построения нечетких моделей нелинейных объектов. Решим задачу определения аддитивной составляющей алгоритма управления нечеткой линеаризованной динамической системой, гарантирующей устойчивость структуры робастного аддитивного регулятора, то есть обеспечивающей сходимость к нулю ошибки управления вдоль всей траектории системы.

2. Решение задачи

Нечеткая дискретная линеаризация объектов цифрового управления. Линеаризация дискретных зависимостей «вход-выход» (ЛДВ) позволяет упростить процедуру синтеза нечетких цифровых регуляторов для нелинейных объектов управления.

Принцип ЛДВ основан на вводе в модель нового входа:

$$v(k) = \Psi_\alpha [z_1, u(k)], \quad (1)$$

где Ψ_α – нелинейная функция, отражающая зависимость «вход-выход» и позволяющая получить равенство:

$$y(k+r) = v(k), \quad (2)$$

которое линеаризует исходную систему.

В общем случае можно определить закон цифрового управления в виде:

$$U(k) = \Psi_\alpha^{-1} [z_1, v(k)]. \quad (3)$$

Рассмотрим принцип ЛДВ для нечеткой линеаризации дискретных зависимостей «вход-выход» (НЛДВ). В этом случае функция FRML6 аппроксимируется нечеткой системой (например, системой Такаги-Сугено).

Выход, генерируемый нечеткой системой Такаги-Сугено, можно представить в виде:

$$y_f(k+r) = \Psi_{fd} [z_1(k), u(k)]. \quad (4)$$

Представим зависимость (3) следующей суммой:

$$y_f(k+r) = \Psi_{a1} [Z_1(k)] + \Psi_{a2} [Z_1(k)]u(k), \quad (5)$$

где Ψ_{a1} и Ψ_{a2} – функции, позволяющие выразить выход нечеткой системы с помощью активизированной ячейки.

Очевидно, что выход глобальной системы можно представить в виде:

$$y(k+r) = y_f(k+r) + \Delta, \quad (6)$$

где Δ – ошибка аппроксимации, не превышающая допустимого значения Δ_0 .

В [2] показано, что закон управления нечеткой системой можно представить зависимостью:

$$u(k) = -\frac{\Psi_{a1}}{\Psi_{a2}} + \frac{v(k)}{\Psi_{a2}}, \quad (7)$$

которая требует инверсии функции Ψ_{fd} .

Следует отметить, что применение нового входа $v(k)$ требует знания выхода системы для $(r-1)$ будущих моментов времени.

На практике можно для определения управлений использовать задержку входа на r тактов.

Кроме того, значения выходов системы в будущих тактах можно заменить их прогнозом.

В [2] показано, что дополнительная переменная $v(k)$ линеаризованной системы может быть выбрана следующим образом:

$$\begin{aligned} v(k) &= y_d(k+r) + \sum_{j=0}^{(r-1)} \eta_j [y_d(k+j) - y(k+j)] = \\ &= y_d(k+r) + \eta^T e, \end{aligned} \quad (8)$$

где

$$\begin{aligned} \eta^T &= [\eta_0, \eta_1, \dots, \eta_{r-1}]; \\ e(k) &= [e_0(k), e_0(k+1), \dots, e_0(k+r-1)]^T; \\ e_0(k) &= y_\alpha(k) - y(k). \end{aligned}$$

Значения $e_0(k)$ соответствуют ошибке системы управления, а коэффициенты η_j выбираются такими, чтобы в z -полиноме

$$z^r + \eta_{r-1}z^{r-1} + \dots + \eta_0 = 0$$

все корни лежали внутри единичной окружности комплексной z -плоскости.

Тогда уравнение ошибки можно записать в виде:

$$e_0(k+r) = -\eta^T e(k) - \Delta$$

или (в матричной форме)

$$e(k+1) = Ae(k) - B\Delta, \quad (9)$$

где

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ -\eta_0 & -\eta_1 & -\eta_2 & \dots & -\eta_{r-1} \end{bmatrix}; \quad (10)$$

$$B = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 1 \end{bmatrix}.$$

Формирование аддитивного робастного управления.

Целью синтеза робастного нечеткого регулятора является компенсация (или уменьшение) влияния ошибки нечеткой аппроксимации на качество управления, обеспечивающая робастность структуры системы по отношению к этой ошибке.

Добавим в закон управления (7) составляющую аддитивного управления u_A :

$$u = -\frac{\Psi_{a1}}{\Psi_{a2}} + \frac{1}{\Psi_{a2}}v - \frac{1}{\Psi_{a2}}u_A. \quad (11)$$

В этом случае уравнение ошибки управления примет вид:

$$e_0(k+r) = -\eta^T e(k) + u_A - \Delta$$

или (в матричной форме):

$$e(k+1) = Ae(k) + Bu_A - B\Delta. \quad (12)$$

Схема, приведенная на рисунке, иллюстрирует принцип робастного аддитивного управления.

Пусть P_1 – симметричная, положительно определенная матрица, удовлетворяющая уравнению Ляпунова

$$A^T P_1 A - P_1 = -Q, \quad (13)$$

где Q – произвольная симметрическая, положительно определенная матрица.

Введем квадратичную функцию $V(k)$ вида:

$$V(k) = 0,5e^T P_1 e(k). \quad (14)$$

Закон аддитивного робастного управления u_A , формируемого по методу сдвига, непосредственно вытекает из следующего утверждения.

Утверждение. Если аддитивное управление u_A формируется по закону

$$u_A = -\Delta_0 \text{sign} [e^T(k+1)P_1 B], \quad (15)$$

то квадратичная функция (14) является функцией Ляпунова, гарантирующей устойчивость структуры робастного аддитивного управления, то есть обеспечивающей сходимость к нулю ошибки управления вдоль всей траектории системы.

Доказательство. Определим разность вида:

$$\Delta V(k) = V(k+1) - V(k) = 0,5e^T(k+1)P_1 e(k+1) - 0,5e^T(k)P_1 e(k). \quad (16)$$

Подставляя значение ошибки управления из (12) в (16), получаем:

$$\Delta V(k) = 0,5e^T(k) [A^T P_1 A - P_1] e + e^T(k) A^T P_1 B [u_A - \Delta] + 0,5B^T P_1 B [u_A - \Delta]^2$$

или

$$\Delta V(k) = -0,5e^T(k) Q e(k) + e^T(k) A^T P_1 B [u_A - \Delta] + 0,5B^T P_1 B [u_A - \Delta]^2. \quad (17)$$

С учетом (8) можно показать, что

$$e^T(k) A^T P_1 B = e^T(k+1) P_1 B - B^T P_1 B e^T(k) \eta - B^T P_1 B e^T(k+r). \quad (18)$$

Подставляя (18) в (17), получаем:

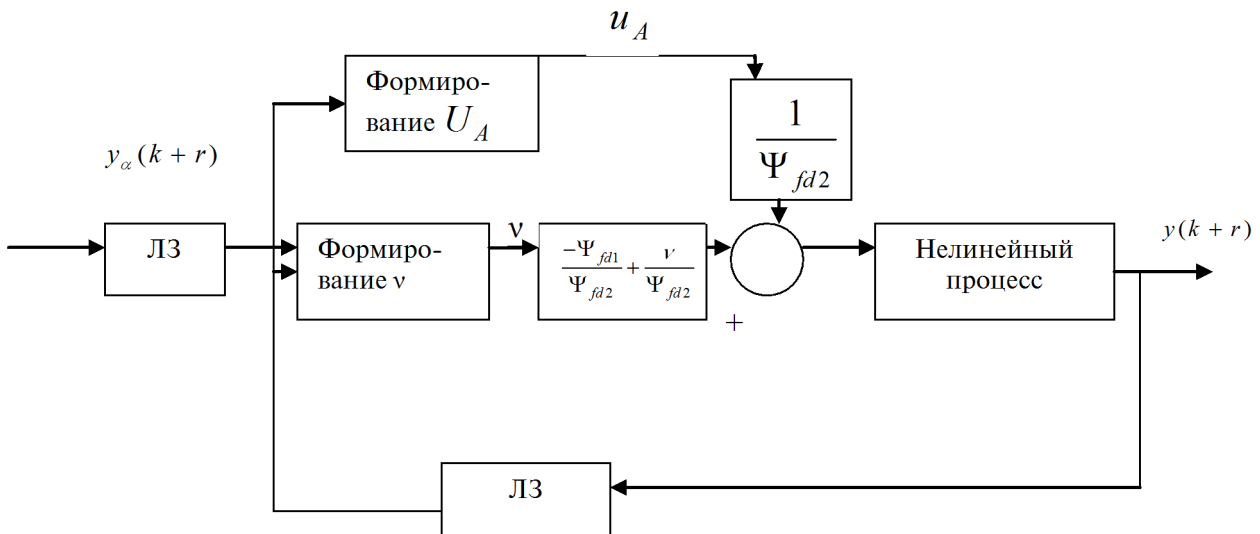


Рис. Схема робастного аддитивного управления

$$\Delta V(k) = -0,5e^T(k)Qe(k) + e^T(k+1)P_1B[u_A - \Delta] - B^T P_1 B e^T(k)\eta[u_A - \Delta] + B^T P_1 B \eta^T e(k)[u_A - \Delta] - B^T P_1 B [u_A - \Delta]^2 + 0,5B^T P_1 B [u_A - \Delta]^2. \quad (19)$$

Для упрощения уравнение (19) представим в виде:

$$\Delta V(k) = -0,5e^T(k)Qe(k) + e^T(k+1)P_1B[u_A - \Delta] - B^T P_1 B e^T(k)\eta[u_A - \Delta] + B^T P_1 B \eta^T e(k)[u_A - \Delta] - B^T P_1 B [u_A - \Delta]^2 + 0,5B^T P_1 B [u_A - \Delta]^2. \quad (20)$$

После подстановки управления u_A из (15) в (20) получаем:

$$\Delta V(k) \leq -0,5e^T(k)Qe(k) \leq 0. \quad (21)$$

Неравенство (21) подтверждает устойчивость рассматриваемой системы управления.

3. Результаты моделирования

Рассмотрим нелинейную систему 2-го порядка, описываемую следующим динамическим уравнением:

$$y(k+2) = \frac{y(k)}{1+y^2(k)} - 0,15y^2(k-1) + u(k). \quad (22)$$

Эту нелинейную систему, предполагая неизвестной ее модель, представим набором следующих нечетких правил:

Если $y(k)$ соответствует $A_1^{i_1}$ и $y(k-1)$ соответствует $A_2^{i_2}$ и $u(k)$ соответствует $B_1^{j_1}$, то $y_f(k+2) = \phi_{fd}^{(i_1, i_2, j_1)}$; $i_1, i_2 = \overline{1,5}; j_1 = \overline{1,2}$.

Пусть желаемая траектория системы имеет следующий вид:

$$y_d(k) = \sin\left(\frac{2\pi k}{100}\right) + 0,75\cos\left(\frac{2\pi k}{50}\right). \quad (23)$$

В соответствии с этим составим набор нечетких правил. Правила, формируемые в соответствии с приведенной ниже таблицей, имеют вид:

Если $y(k)$ соответствует A_1^3 и $y(k-1)$ соответствует A_2^3 и $u(k)$ соответствует B_1^2 , то $y_f(k+2) = 1,3714$.

Значения переменных $y(k)$, $y(k-1)$ и $u(k)$ изменяются в следующих диапазонах:

$$\begin{aligned} [y(k)^{\min}, y(k)^{\max}] &= \\ [y(k-1)^{\min}, y(k-1)^{\max}] &= [-2, 0; 2, 0]; \\ [u(k)^{\min}, u(k)^{\max}] &= [-1, 50; 1, 50]. \end{aligned}$$

Зададим желаемые значения полюсов замкнутой системы цифрового управления: $z_1 = -0,8; z_2 = -0,9$, что соответствует характеристическому уравнению вида:

$$z^2 + 1,7z + 0,72 = 0. \quad (24)$$

В качестве начальных условий примем:

$$y(0) = y_\alpha(0); y(1) = y_\alpha(1); V(0) = 0.$$

Вначале определим норму H_∞ дискретной передаточной функции $H_{e_0\Delta}(z)$ без учета управления U_A :

$$\|H_{e_0\Delta}(z)\|_\infty = 50.$$

Норма ошибки слежения для горизонта времени $[k_0, k_f] = [0, 500]$ равна:

$$\|e_0\| = 20,242.$$

Определим закон аддитивного управления U_A , обеспечивающий выполнение условия (21) при $\gamma = 2,50$. Матрицу Q выберем равной $0,001 I$.

С учетом приведенного выше утверждения, получаем симметричную положительно определенную матрицу вида:

$$P = \begin{bmatrix} 2,6151 & 2,8015 \\ 2,8015 & 6,2113 \end{bmatrix}.$$

Таблица

Набор нечетких правил для системы (22)

$\begin{matrix} y(k) \\ y(k-1) \end{matrix}$	A_1^1	A_1^2	A_1^3	A_1^4	A_1^5
A_2^1	$B_1^1 : -2,2508$ $B_1^2 : 0,0653$	$B_1^1 : -2,0643$ $B_1^2 : 0,3964$	$B_1^1 : -0,8539$ $B_1^2 : 1,3641$	$B_1^1 : -0,5648$ $B_1^2 : 1,1989$	$B_1^1 : -0,3783$ $B_1^2 : 1,4563$
A_2^2	$B_1^1 : -1,5306$ $B_1^2 : 0,5450$	$B_1^1 : -1,5828$ $B_1^2 : 0,1981$	$B_1^1 : -1,6516$ $B_1^2 : 0,4903$	$B_1^1 : -0,4430$ $B_1^2 : 1,2779$	$B_1^1 : -0,3594$ $B_1^2 : 1,2661$
A_2^3	$B_1^1 : -0,1895$ $B_1^2 : 1,1960$	$B_1^1 : -1,5819$ $B_1^2 : 0,6573$	$B_1^1 : -1,3405$ $B_1^2 : 1,3714$	$B_1^1 : -0,7626$ $B_1^2 : 1,7056$	$B_1^1 : -0,8446$ $B_1^2 : 1,3796$
A_2^4	$B_1^1 : -0,5648$ $B_1^2 : 1,1989$	$B_1^1 : -0,3829$ $B_1^2 : 1,4533$	$B_1^1 : -0,4623$ $B_1^2 : 1,4935$	$B_1^1 : -0,6365$ $B_1^2 : 1,3894$	$B_1^1 : -1,3816$ $B_1^2 : 0,9469$
A_2^5	$B_1^1 : -0,4469$ $B_1^2 : 1,2767$	$B_1^1 : -0,3594$ $B_1^2 : 1,2661$	$B_1^1 : -0,2255$ $B_1^2 : 1,1542$	$B_1^1 : -0,5028$ $B_1^2 : 1,4336$	$B_1^1 : 0,2350$ $B_1^2 : 1,3600$

В этом случае, норма H_∞ дискретной передаточной функции между неопределенностью Δ и ошибкой $e_0(k)$ с учетом управления U_A имеет следующее значение:

$$\|H_{e_0\Delta}(z)\|_\infty = 1,8233,$$

что подтверждает выполнение условия (21).

Норма ошибки слежения для горизонта времени $[k_0, k_f] = [0, 500]$ равна:

$$\|e_0\| = 0,5028.$$

Для начальных условий $y(0) = 1.0$; $y(1) = 1.25$ ($V(0) \neq 0$) норма ошибки слежения составила: 0,8883 (с учетом аддитивной составляющей управления) и 20,50 (без учета аддитивной составляющей управления).

Результаты моделирования подтверждают высокое качество робастного нечеткого управления.

Полученные теоретические результаты были частично использованы применительно к задачам нечеткого управления распределением нагрузки между фильтрами блока подпитки теплосети.

Выводы

Учитывая, что при нечетком моделировании основным источником ошибок является неточность аппроксимации, аддитивное представление неопределенностей является наиболее приемлемым для построения нечетких моделей нелинейных объектов. Введение в закон нечеткого управления аддитивных составляющих, полученных с помощью методов H_∞ -оптимизации, позволяет повысить робастность нечеткого цифрового управления по отношению к неопределенностям и возмущениям, характерным для нечеткой системы. Такая робастность может обеспечить компенсацию или снижение влияния неопределенностей и возмущений на качество управления.

Результаты моделирования подтверждают работоспособность рассмотренных в настоящей работе нечетких робастных регуляторов.

Список литературы. 1. В.В. Круглов. Интеллектуальные информационные системы: компьютерная поддержка систем нечеткой логики и нечеткого вывода. – М.: Физматлит, 2002. – 315 с. 2. P. Congdon. Fuzzy Modeling for Control. – Boston: Kluwer Academic Publishers, 1998. – 275 p. 3. Каюша В.А., Кирияк Р.Д., Удовенко С.Г. Алгоритм робастного управления стохастическими объектами с дрейфом параметров // Вестник Харьковского государственного политехнического университета. – 2000. – Вып. 99. – С. 70–73.

Поступила в редакцию 24.02.2009

УДК 681.513

Робастне керування динамічним процесом на основі нечіткої моделі / С.Г. Удовенко, Г. Дібе // Біоніка інтелекту: наук.-техн. журнал. – 2009. – №1(70). – С. 72–76.

У статті розглянуто задачу представлення байєсовської моделі керованого стохастичного процесу у просторі станів. Запропоновано алгоритм рекурентного оцінювання станів згідно з цією моделлю. Наведено результати експериментального моделювання, що підтверджують працездатність запропонованого підходу.

Табл.: 1. Лл.: 1. Бібліогр.: 3 найм.

UDK 681.513

Robust control for dynamic process based of fuzzy model / S.G. Udovenko, G. Dibe // Bionics of Intelligence: Sci. Mag. – 2009. – №1(70). – P. 72–76.

In the article the tasks of presentation of bayesian model of the controlled stochastic process are examined in states space. The algorithm of recursive estimation of the states is offered in accordance with this model. The results of experimental modeling confirmative the capacity of offered approach are offered.

Tabl.: 1. Fig.: 1. Ref.: 3 items.

УДК 007.5



ЗНАНИЕ-ОРИЕНТИРОВАННАЯ МОДЕЛЬ ПРЕДСТАВЛЕНИЯ И ОБРАБОТКИ БИЗНЕС-ОБЪЕКТОВ В ПРОЦЕССАХ ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА

В.М. Левыкин¹, С.Ф. Чалый², С.Н. Селезнев³

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

²ХНУРЭ, г. Харьков, Украина, chaliy@datasvit.net

³ХНУРЭ, г. Харьков, Украина, rp01@mail.ru

Предложена модель представления и обработки бизнес-объектов в процессах электронного документооборота, которая включает в себя представление знаний в форме бизнес-правил, отражающих требования к порядку обработки таких объектов. Использование предложенной модели позволяет адаптировать последовательность обработки документов с учетом наличия исполнителей с соответствующими полномочиями.

БИЗНЕС-ПРАВИЛА, БИЗНЕС-ОБЪЕКТЫ, ДОКУМЕНТ, ИСПОЛНИТЕЛЬ, ПРОЦЕДУРА

Введение

Процессный подход к управлению организацией основывается на выделении бизнес-процессов (БП) и управлении этими процессами с целью повышения эффективности ее деятельности. Бизнес-процессы являются ключевыми элементами процессно-ориентированной системы управления организацией либо предприятием и представляют собой «последовательности работ, которые по определенной технологии преобразует входы процесса в его выходы, представляющие ценность для потребителя» [1].

Традиционный подход к описанию бизнес-процессов, ориентированный на разработку неизменяемой workflow-модели БП, входит в противоречие с практикой функционирования отечественных предприятий. Последняя характеризуется эволюцией целей предприятия, а также влиянием непредвиденных внешних воздействий, вызывающих изменение БП, что требует преобразования структуры бизнес-процесса.

В связи с изложенным, возникает проблема представления и моделирования бизнес-процессов предприятия, которые изменяются в процессе своего функционирования под влиянием внешних возмущений, а также эволюции целей предприятия.

В рамках решения данной проблемы, БП представляется в виде многокомпонентной структуры и включает в себя компоненту функциональных требований, выраженных в форме бизнес-правил, компоненту бизнес-процедур, которые оперируют бизнес-объектами (БО), а также последовательность бизнес-процедур (workflow) [2].

Отметим, что в качестве бизнес-объектов в документообороте (ДО) выступают документы. Обработка таких бизнес-объектов выполняется на основе знаний о процессах документооборота, выраженных в форме бизнес-правил.

Вопросы управления бизнес-процессами достаточно детально рассмотрены в работах [1,3], где основное

внимание уделено разработке последовательности процедур БП с учетом организационной структуры современных предприятий. Вопросы представления, моделирования и оценки БО с учетом бизнес-правил, в указанных работах рассмотрены недостаточно.

Таким образом, задача формирования знание-ориентированного представления бизнес-объектов, является актуальной.

1. Формулировка задачи

Задача разработки знание-ориентированной модели представления бизнес-объектов при процессном подходе к управлению документооборотом предполагает учет особенностей таких БО, в частности структуры их атрибутов, операций их обработки, правил применения таких операций.

Соответственно, данная задача формулируется следующим образом.

Требуется разработать модель представления БО, отражающую взаимосвязь элементов бизнес-объекта, а также правила обработки таких элементов, базирующиеся на знаниях о предметной области и учитывающие последовательность процедур при реализации процессов документооборота.

Исходными данными являются атрибуты БО и операции по их обработке в рамках процессов документооборота.

2. Обобщенная последовательность и правила обработки документа в процессах электронного документооборота

Проведенный анализ процессов документооборота позволил выделить обобщенную последовательность обработки документов, ориентированную на распределение исполнителей (рис. 1). Такое распределение выполняется в соответствии с правилами, учитывающими функциональные обязанности сотрудников структурных подразделений. Полученная обобщенная последовательность обработки позволяет обосновать ди-

аграмму четырех основных состояний документа: поступление документа, документ на обработке, документ на регистрации, документ на исполнении (рис. 2). Перевод документа из одного состояния в другое выполняется на основе соответствующих бизнес-правил, отражающих знания о процессах документооборота.

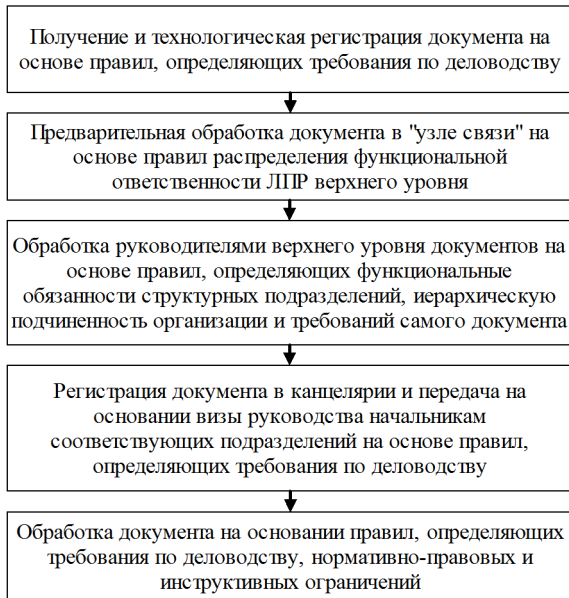


Рис. 1. Последовательность обработки электронного входящего документа

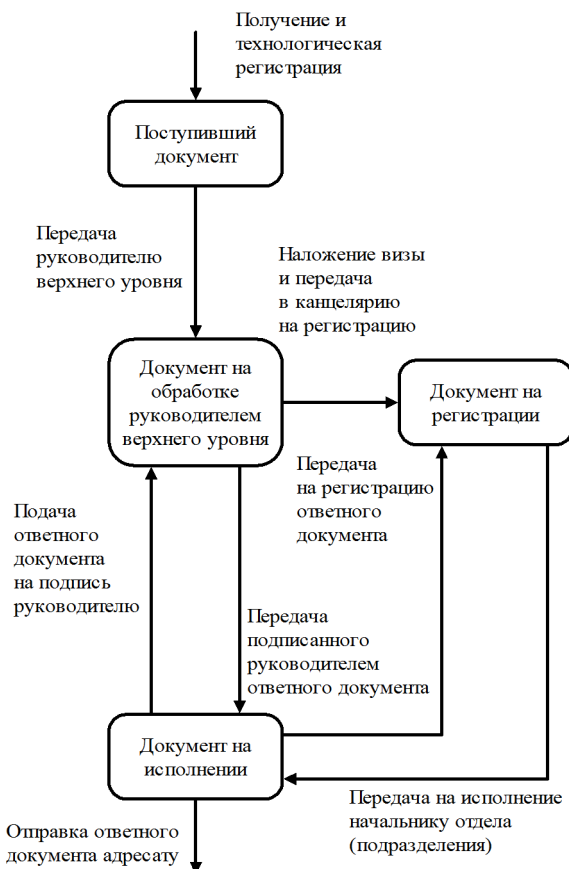


Рис. 2. Диаграмма состояний документа в процессе реализации электронного документооборота

Такие правила подразделяются на следующие группы [4]:

- определение объектов и исполнителей бизнес-процесса;
- классификация объектов и исполнителей бизнес-процесса;
- управление бизнес-процедурами и распределение исполнителей;
- определение структурных взаимосвязей между элементами БП.

Под правилом определения объектов и исполнителей $Bpr_{ОО}$ будем понимать бизнес-правило, которое характеризуется следующими особенностями:

- аргументом правила является объект (объекты) и/или исполнители бизнес-процесса;
- результатом выполнения правила является определение нового объекта и/или исполнителя БП.

Правилем классификации объектов и исполнителей бизнес-процесса $Bpr_{КО}$ будем называть бизнес-правило, которое характеризуется следующими особенностями:

- аргументом правила является объект и/или исполнитель бизнес-процесса;
- результатом выполнения правила является множество объектов и исполнителей БП.

Правилем управления бизнес-процедурами и распределения исполнителей $Bpr_{УП}$ будем называть бизнес-правило, которое характеризуется следующими особенностями:

- аргументом правила является бизнес-процедура;
- результатом применения правила является запуск или останов соответствующей процедуры.

Правилем $Bpr_{СВС}$ определения структурных взаимосвязей между элементами БП будем называть бизнес-правило, которое характеризуется следующими особенностями:

- аргументом правила является объект, бизнес-процедура либо последовательность процедур;
- результатом применения правила является элемент бизнес-процесса, включающий аргумент правила в качестве своей составной части.

3. Выделение структурных элементов бизнес-объекта для процесса документооборота

Необходимым условием построения модели представления бизнес-объекта является выделение его структурных элементов, отражающих специфику процессов документооборота. Иерархия таких элементов охватывает:

1. Заглавную часть документа, которая включает в себя следующий набор элементов:

- информация об отправителе (авторе);
- информация о получателе (исполнителе) документа;
- дата подписания;

- номенклатурный регистрационный номер;
 - категория документа: приказ, распоряжение, письмо, договор, заявление, служебная записка;
 - информацию о грифе, если таковой есть;
2. Основную часть документа, которая имеет следующие составляющие:
- уточнение категории документа: служебная записка; представление; прошение; докладная; письмо; обращение; запрос; разъяснение; уведомление; ответ на обращение;
 - сроки выполнения и/или дату вступления в силу (могут быть выражены нечетко);
3. Реквизиты руководителя, подписавшего документ: должность; звание; фамилия; подпись.
4. Реквизиты исполнителя, подготовившего документ с контактной информацией.
- Обобщенная структура документа, которая отражает рассмотренные его особенности, представлена на рис. 3.

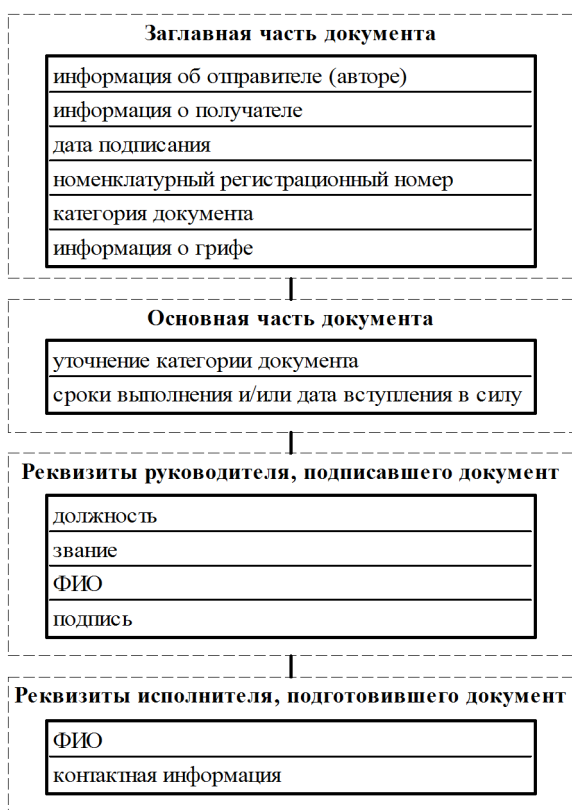


Рис. 3. Обобщенная структура БО-документа

На основании выделенных структурных составляющих типового документа приведем формальное представление бизнес-объекта для процессов документооборота.

Формально модель документа M^D имеет вид:

$$M^D = \langle Z, T, B, I, Vpr^D \rangle, \quad (1)$$

где Z – заглавная часть документа; T – основная часть документа; B – подпись руководителя под документом; I – исполнитель, подготовивший документ; Vpr^D – множество правил обработки (исполнения) документа.

Приведенное представление отражает общую типовую структуру документа, но не раскрывает связей между элементами и их внутреннюю иерархию. С целью детализации структуры БО рассмотрим особенности составных частей документа с учетом специфики их обработки.

Заглавная часть документа Z обычно содержит информацию об отправителе и получателе, и используется в операциях предварительного анализа и логической сортировки документов процедуры предварительной обработки в узле связи на основе **правил распределения функциональной ответственности ЛПР верхнего уровня**. Исполнителями являются сотрудники узла связи.

Заглавная часть документа может содержать гриф секретности («секретно») или ограниченности использования («ДСП» – для служебного пользования и тому подобное). Наличие грифа накладывает определенные ограничения на процесс обработки документа, выраженные в форме **правил ограничения доступа к информации** и, обычно, не дает прямой информации для содержательного исполнения документа.

Информация об отправителе включает в себя следующий набор атрибутов: наименование организации, в которой подготовлен (создан) данный документ; информация о юридическом адресе отправителя; контактная информация – если документ создан на фирменном бланке.

Информация об отправителе используется в: **правилах выбора ЛПР верхнего уровня для визирования документа; правилах формирования обратной связи и выбора ЛПР, который подпишет ответный документ; правилах оценки правомочности исполнения документа и юридической значимости документа.**

Информация о получателе дает, в первую очередь, возможность убедиться, что документ поступил по соответствующему адресу. В некоторых случаях (обычно при поступлении документа от подчиненных территориальных подразделений) информация о получателе может содержать в себе дополнительные данные для передачи документа на исполнение в соответствующие структурные подразделения.

Информация о получателе используется в **правилах наложения визы на исполнение.**

Информация о дате документа может быть использована для определения даты вхождения документа в силу (закон) или максимально допустимых сроков рассмотрения (исполнения) документа, если об этом нет явного упоминания в теле документа.

Информация о дате документа используется в **правилах проверки сроков рассмотрения документов и правилах вступления законодательных документов в силу.**

Номер документа может дать информацию о структурном подразделении, подготовившем документ (обычно, если документ поступил от вышестоящей организации или подчиненных территориальных подразделений).

Информация о номере документа используется в **правилах выделения подразделения организации, подготовившего документ.**

Заголовок документа может явно определять тип документа.

Таким образом, заглавную часть документа можно представить в виде кортежа:

$$Z = \langle G^D, O^D, D^D, N^D, P^D, Z^D, Vpr^{DZ} \rangle, \quad (2)$$

где G^D – гриф документа; O^D – отправитель документа; D^D – дата документа; N^D – номер документа; P^D – получатель документа; Z^D – заголовок документа; Vpr^{DZ} – правила обработки заголовка документа.

Подпись под документом определяет уровень руководителя, который подписал документ.

Информация о руководителе, подписавшем документ, используется в **правилах оценки правомочности исполнения документа и юридической значимости документа.**

Данные об исполнителе обычно содержат фамилию, имя и отчество ответственного за подготовку документа и его контактные данные. Данная информация используется для уточнения сути документа и для реализации обратной связи с исполнителями.

Данные об исполнителе документа используются в **правилах диалогового уточнения сути документа и/или поставленной в нем задачи/запроса.**

Тело документа является мало структурированной и, в тоже время, наиболее информативной частью документа. Объем тела документа может быть от нескольких предложений до десятков и сотен страниц. Анализ информационного содержания тела документа является наиболее трудоемкой задачей особенно на этапе обработки руководителем верхнего уровня и исполнении документа.

Основное внимание телу документа уделяется при реализации следующих процедур:

– обработка руководителями верхнего уровня документов на основе **правил, определяющих функциональные обязанности структурных подразделений, иерархическую подчиненность организации и требований самого документа;**

– обработка документа на основании правил, определяющих требования по деловодству, нормативно-правовых и инструктивных ограничений.

Кроме того, информация, содержащаяся в теле документа, может иметь отражение на все группы правил процесса документооборота организации.

4. Выделение правил обработки бизнес-объектов

Проведенная структуризация бизнес-объекта позволила выделить ряд групп правил обработки БО. С учетом приведенных выше четырех типов бизнес-правил выполним классификацию и формализацию таких правил.

Правила определения объектов и исполнителей Vpr_{OO} включают в себя:

– правила, определяющие функциональные обязанности структурных подразделений;

– правила, определяющие требования по деловодству;

– правила нормативно-правовых и инструктивных ограничений.

Правила классификации объектов и исполнителей бизнес-процесса Vpr_{KO} включают в себя:

– правила распределения функциональной ответственности ЛПР верхнего уровня;

– правила ограничения доступа к информации;

– правила выбора ЛПР верхнего уровня для визирования документа;

– правила выделения подразделения организации, подготовившего документ;

– правила, определяющие иерархическую подчиненность организации.

Правила управления бизнес-процедурами и распределения исполнителей Vpr_{VI} включают в себя:

– правила оценки правомочности исполнения документа и юридической значимости документа;

– правила проверки сроков рассмотрения документов;

– правила вступления законодательных документов в силу.

Правила Vpr_{CBC} определения структурных взаимосвязей между элементами БП включают в себя:

– правила формирования обратной связи и выбора ЛПР, который подпишет ответный документ;

– правила наложения визы на исполнение;

– правила подчиненности взаимозаменяемости ЛПР верхнего уровня;

– правила распределения ответственности исполнителей по документу.

Исходя из изложенной классификации, можно сделать вывод, что при формализации обработки бизнес-объектов могут быть использованы логические модели 4-х типов правил, приведенные в работе [4].

Выводы

В рамках статьи впервые предложена знание-ориентированная модель представления и обработки бизнес-объектов для процессов электронного документооборота, которая охватывает иерархию атрибутов документа и связанных с каждым атрибутом правил определения и классификации объектов, управления такими объектами и назначения исполнителей, а также определения структурных взаимосвязей между элементами. Указанные правила отражают знания о порядке функционирования электронного документооборота.

Полученная модель обеспечивает возможность повышения эффективности обработки документов путем автоматизации назначения процедур обработ-

ки и распределения исполнителей на основе соответствующих правил.

С практической точки зрения предложенная модель предоставляет возможность реализации базы знаний бизнес-правил по данному процессу в виде фреймовой структуры.

Представление документа, основанное на правилах, обеспечивает гибкую адаптацию последовательности его обработки в зависимости от наличия или отсутствия исполнителей, сроков исполнения, ограничений уровня доступа.

Список литературы: 1. Репин В.В., Елиферов В.Г. Процессный подход к управлению. Моделирование бизнес-процессов.—М.: РИА «Стандарты и качество», 2004.—408 с. 2. Левикин В.М., Чалий С.Ф. Формализованное описание бизнес-процессов в АИС // Штучний інтелект.—2003.—№ 2.—С. 28–34. 3. Кальянов Г.Н. Теория и практика реорганизации бизнес-процессов.—М.: СИНТЕГ, 2000. 4. Модели представления знаний для бизнес-процессов с изменяемой структурой / В. М. Левикин, С. Ф.Чалий // Бионика интеллекта: науч.-техн. журнал.— 2008.—№ 1(68). — С. 40–44.

Поступила в редколлегию 25.02.2009

УДК 007.5

Знання-орієнтована модель представлення та обробки бізнес-об'єктів в процесах електронного документообігу / В.М. Левикин, С.Ф. Чалий, С.М. Селезньов // Біоніка інтелекту: наук.-техн.-журнал. — 2009. — №1(70). —С. 77–81.

Запропоновано модель представлення та обробки бізнес-об'єктів в процесах електронного документообігу, яка включає в себе представлення знань у формі бізнес-правил, що відображають вимоги до порядку обробки таких об'єктів. Використання запропонованої моделі дозволяє адаптувати послідовність обробки документів з урахуванням наявності виконавців з відповідними повноваженнями.

Л.: 3. Бібліогр.: 4 найм.

UDC 007.5

Knowledge-oriented model of process and presentation of business-objects for electronic documents / V.M. Levikin, S.F. Chaliy, S.N. Seleznev // Bionics of Intelligence: Sci.Mag.— 2009. — №1(70).—P. 77–81.

Knowledge-oriented model of process and presentation of business-objects for electronic documents is offered. The model includes representation of knowledges in a form of business-rules of four types. The use of the offered model allows to adapt the sequence of treatment of documents taking into account the presence of performers with the proper plenary powers.

Fig.: 3. Ref.: 4 items.

UDC 57:51-76; 57.02.001.57; 517.958:57



BROWNIAN MOTION SIMULATION OF VESICLE ACTIN TAIL FORMATION

A. I. Oleinick

KNURE, Kharkov, Ukraine, oleinick@kture.kharkov.ua

The mathematical model of actin tail formation at the surface of a spherical vesicle is proposed in this article. The simulation of the protein diffusion process, which is complicated by the formation of a protein cluster at the vesicle surface leading to the development of an actin tail, was performed using the Brownian motion approach. Two relevant mechanisms of tail growth are considered and discussed.

VESICLE, BROWNIAN MOTION, PROTEIN DIFFUSION, CLUSTERISATION, ACTIN TAIL

Introduction

Movement of some vesicles inside the cell occurs because of the catalysed formation of an actin tail which forces the vesicle to move in a certain direction. For example, the investigations of the *Xenopus* eggs show that within several minutes from their activation by PMA (Phorbol Myristate Acetate) actin tails appeared at some vesicles (endosome or lysosome) in these cells thus propelling the vesicles in the cytoplasm of the cell [1,2]. This process is related solely to the vesicles since it could be reproduced *in vitro*, i.e. out of the cell environment. In previous works [1,2] it was assumed that the origin of this process is the gathering (by diffusion) and growth of the protein (N-WASP/Cdc42 complexes) population in specific regions of the vesicle surface. As a result of this sequence actin tails begin to grow from these regions. However the relationship between the rate of tail growth and rate of protein accumulation at the vesicle surface is still unknown since experimentally one can observe only the rate of actin tail growth.

The aim of this work is to delineate the relation between the rate of tail growth and the rate of protein agglomeration (or its amount) in order to quantify these processes.

1. Mathematical model

Let us consider a vesicle or any other appropriate object enclosed in a membrane (e. g., endosome or lysosome). In the following we will assume it to be spherical with radius R . This does not affect the generality of the results presented below unless some shape features significantly alter the probability of the protein displacement or reaction ability. We also assume that N_0 proteins are present in the vesicle membrane, which initially are uniformly and randomly distributed over the membrane spherical surface.

The cytoskeleton structures create nanometric compartments at the vesicle surface [3]. The proteins can move freely and randomly within these compartments, but the walls, imposed by the cytoskeleton, restrict protein movement out of these compartments. However a protein molecule can «jump» over these barriers by accumulating some threshold energy ΔE^\ddagger . Therefore the displacement of a protein is equivalent to its «jump» diffusion between

nanometric compartments, because the mean protein displacement within the compartment is equal to zero. If the temperature (T) is fixed, the frequency of these jumps is controlled by the Gaussian probability $\exp(-\Delta E^\ddagger / k_B T)$ and represents the local diffusion coefficient (D).

Hence in the following we will consider the Brownian motion of proteins over a spherical surface. The Brownian motion of free proteins continues till they become activated by specific cell pathways. The latter provoke a cascade of reactions leading to the formation of a cluster of the N-WASP/Cdc42 protein complexes on the vesicle surface coupled with the activation of Arp2/3 complex and beginning of actin nucleation [1,2]. The main features of the utilised algorithm of Brownian motion are the following. During the free stage of their movement the proteins are considered as rigid cylinders with radius r_{pr} on the vesicle surface and they are not allowed to intersect. At the aggregation phase (i.e. during the formation of a cluster) each protein which approaches the cluster closer than to within r_{react} (the distance between cylindrical surfaces) joins the cluster, i.e. the aggregation rate constant k_{agg} is assumed to be sufficiently high. When the aggregation phase begins the first two proteins which appear at the required distance and are able to react form the cluster.

It should be noted that such diffusion limited aggregation process will lead to the formation of branched fractal clusters [4]. However in this work we will assume that the cluster always represents a spherical segment with the area equal to the area of the aggregated proteins. The latter assumption has a significant biological basis since it is known for this kind of structures that they possess the ability of regrouping to avoid the forms with long branches [5]. Moreover this statement is also confirmed by electron microscopy images [1].

The elementary step of each protein on a sphere was implemented in a computer program on the basis of the algorithm described in [6] which consists in the following. Let x be a vector which represents the position of the protein on a sphere and contains its Cartesian coordinates. First, a random 3D vector ζ is generated. Each coordinate of this vector is generated using the Gaussian normal distribution with zero mean and a standard deviation $\sigma = 1$ (in fact the outcome of this process does not depend on the precise value

of σ). Then the vector ζ undergoes the orthonormalisation procedure with respect to vector x (i. e. made orthogonal to vector x and is normalized to a unit length):

$$\bar{\zeta}_i = \zeta_i - (\zeta \circ x) x_i, \quad (1a)$$

$$\zeta_{ort} = \frac{\bar{\zeta}}{|\bar{\zeta}|}, \quad (1b)$$

where (\circ) denotes the inner product of two vectors. Vector ζ_{ort} defines the direction of the protein displacement. The new position of the protein (x') is calculated according to the following expression:

$$x' = x + \delta \zeta_{ort}, \quad (2)$$

where δ is a normally distributed random quantity with zero mean and a standard deviation $\sigma = 2\sqrt{D\Delta t}$.

When the new coordinates of the protein have been evaluated they are tested for acceptability (i. e. the cylinders representing the proteins should not intersect). If the step was not accepted the protein remains at the location where it arrived during the previous time step. After that the protein is checked for its ability to aggregate with the cluster. Since the cluster is represented by a spherical segment and the number of clustered proteins and their area are known, one can easily evaluate the angle corresponding to the cluster edge (θ_{cl}) assuming that its centre is located at the north pole of the sphere. Indeed, the area of the cluster (S^{aggr}) is given by

$$S^{aggr} = 2\pi r_{ves}^2 (1 - \cos\theta_{cl}), \quad (3a)$$

$$\theta_{cl} = \arccos \left[1 - \frac{S^{aggr}}{2\pi r_{ves}^2} \right]. \quad (3b)$$

The use of dimensionless coordinates (i. e. when all the coordinates and characteristic dimensions are scaled with respect to the vesicle radius, which results into consideration of a unit sphere) coupled with the latter relation allows one to check only the z-coordinate of the particle when deciding whether the particle can join the cluster or not. This fact allows a significant acceleration of the simulation process.

Iterative repeating of the described procedure for all proteins until the required duration of the experiment (t_{exp}) is achieved or all the proteins are aggregated, yields the number of aggregated proteins as a function of time, i. e. the dependence $N(t_k)$, where t_k are the acquisition times. However the available experimental data are given in the form of actin tail length vs. time [2]. Hence it is necessary to establish the correspondence between these two transients. The correspondence law contains fundamental information about how the rate of «comet tail» growth and size of the cluster (or its variation) are linked. Therefore formulating an appropriate correspondence law between the mentioned above transients and its verification will allow one to disclose the exact mechanism of tail formation.

The vesicle comet tail growth transients demonstrate a wide range of behaviours depending on the exact conditions of the experiment. Up to now there are two main hypothetic mechanisms which are considered by the experts in the field [2]. These hypotheses are the following.

1) The growth rate is proportional to the number of clustered proteins, i. e.

$$\frac{dL}{dt} = k_1 N, \quad (4a)$$

where k_1 is the rate constant (in cm/s) whose value corresponds to the growth rate stimulated by one protein. Hence the tail length can be computed on the basis of the $N(t)$ curve as follows:

$$L(t) = k_1 \int_0^t N(\tau) d\tau, \quad (4b)$$

where τ is a dummy variable.

2) The growth rate is proportional to the number of proteins which join the cluster through its perimeter per unit of time:

$$\frac{dL}{dt} = k_2 \frac{dN}{dt}, \quad (5a)$$

where k_2 is the rate constant (in cm) whose value corresponds to the tail length supported by one protein. The integration of the latter equation shows that the tail length in this case is proportional to the number of clustered proteins (i. e. $L \sim N$):

$$L(t) = k_2 N(t). \quad (5b)$$

2. Results and discussion

We refer to the experimental data given in [2]. However before the discussion of experimental data and simulation results let us consider some features of the system at hand. The paper [2] contains plots of the actin tail growth dynamics for glass beads covered with the lipid bilayer containing all the components required for the tail growth. The estimated number of N-WASP proteins is around 50000 of molecules per square micrometer with a maximum distance of 5 nm between the molecules. Using these data we can evaluate the average area available for each protein:

$$\frac{1\mu\text{m}^2}{5 \times 10^4} = \frac{10^6 \text{ nm}^2}{5 \times 10^4} = 20 \text{ nm}^2. \quad (6)$$

Assuming square packing of proteins in the cluster (for other packings an additional factor commensurable to unity will appear, which, however, does not change the present analysis [7]) and ignoring for the time being the sphericity of the system, we deduce from the latter equation that the size of the packed array cell, and hence the distance between the protein centres, is equal to $a = \sqrt{20 \text{ nm}^2} \approx 4.5 \text{ nm}$. This value shows that the protein radius (r_{pr}) is at most a few nanometers.

The radius of the glass beads employed in the experiments was approximately 1 μm [2]. This implies that the overall number of proteins located at the bead surface, N_0 , is:

$$N_0 = 4\pi \mu\text{m}^2 \times 5 \times 10^4 \frac{\text{molecules}}{\mu\text{m}^2} = 6.3 \times 10^5 \text{ molecules.} \quad (7)$$

Using such a large number of particles in simulations is extremely time consuming. In order to overcome this difficulty we rely on the approach utilizing quasi-particles, i.e. whenever the number of particles (N_{part}) used in the simulations is less than the number of proteins in the real system (N_0), each particle is considered as a collection of $w = N_0 / N_{part}$ proteins (we also will say that a particle has weight w). The radius of such a weighted particle is evaluated as $\sqrt{w} r_{pr}$, i. e. it occupies the same area on the vesicle surface as a collection of proteins which it represents. Simulations results given below were obtained for $w = 600$, since this value was sufficient to achieve convergence in w as was checked by numerical experiments.

It should be noted also that the use of weighted particles also requires scaling of other relevant parameters, in our case of r_{react} . However the clusterisation algorithm described above allows neglecting this scaling since r_{react} and r_{pr} are decoupled because they are not encountered simultaneously in the aggregation condition and only r_{react} and the coordinates of the quasi-particle matter here. Another important parameter is the sampling time (dt). However the used algorithm does not impose a severe condition on its value as was established by inspecting the time convergence. Therefore the sampling time was chosen in order to produce an average particle displacement $\Delta l = 2\sqrt{D\Delta t}$ to be a few (2-3) times larger than r_{react} , but not exceeding the average distance between the particles (l) which is given by

$$l = \sqrt{\frac{4\pi r_{ves}^2}{N_{part}}}. \quad (8)$$

For the results given below the following parameters were used: $t_{exp} = 3600$ s, $r_{ves} = 1$ μm , $D = 5 \times 10^{-13}$ cm^2/s , $r_{pr} = 2$ nm, $r_{react} = 4$ nm, $N_0 = 6 \times 10^5$, $N_{part} = 10^3$. Note that the value of the diffusion coefficient was taken to be smaller than the lateral diffusivity of lipids in the membrane which is generally of the order of $10^{-8} - 10^{-10}$ cm^2/s .

Figures 1 and 2 illustrate the actin tail growth dynamics computed on the basis of $N(t)$ curves according to equations (4) and (5), respectively. The $N(t)$ curves were obtained by averaging ten individual runs of Brownian motion simulations with the same set of parameters. It should be noted that the amplitude of these transients is controlled by the rate constants defined in Eqs. (4)–(5). For the results given in Figures 1 and 2 their values ($k_1 = 3 \times 10^{-8}$ $\mu\text{m}/\text{s}$, $k_2 = 6.14 \times 10^{-5}$ μm) were adjusted in order to produce the tail length of several tens of microns per hour since this is the inherent growth rate for the system at hand [2]. However despite the same amplitude at the end of the experiment both dependences behave in two qualitatively different ways.

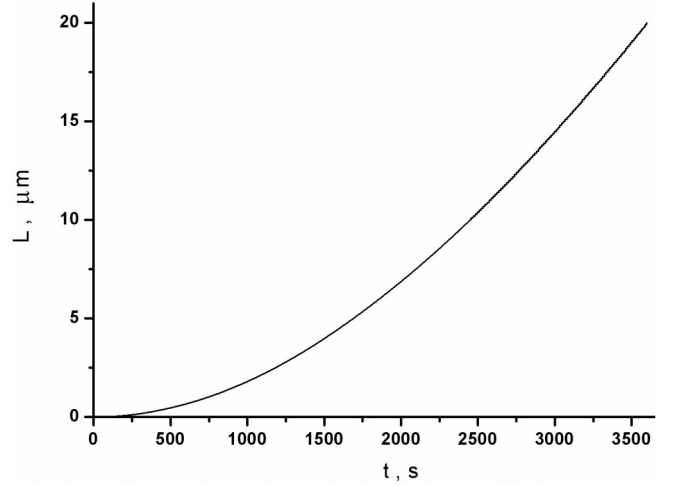


Fig. 1. Tail growth transient obtained via Eq. (3), i. e. growth rate is proportional to the number of aggregated proteins, $k_1 = 3 \times 10^{-8}$ $\mu\text{m}/\text{s}$

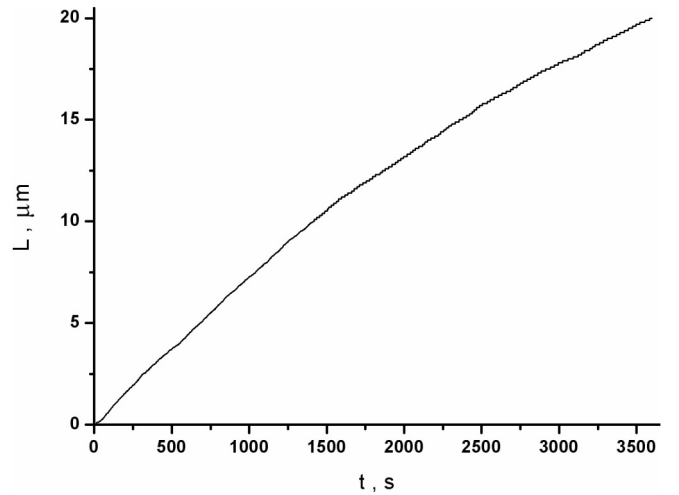


Fig. 2. Tail growth transient obtained via Eq. (4), i. e. tail length is proportional to the number of proteins in the cluster, $k_2 = 6.14 \times 10^{-5}$ μm

Indeed, since each aggregated protein steadily increases the tail length, the first mechanism (Figure 1) results in an ever faster growing curve with an asymptote $L \sim k_1 N_0 t$. On the other hand, the curve obtained via the second mechanism (Figure 2) clearly shows the exhaustion of the protein population with time leading to a stationary value of the actin tail length at large times. The latter behaviour is reminiscent to the one observed experimentally (see Figure 5D in [2]). At the same time the behaviour obtained with the first mechanism clearly does not have place at least under the conditions described in [2].

For more precise data fitting more accurate experimental data are required since each reported curve in [2] was obtained from 10-15 experimental curves and, as a consequence of such poor statistics, has sufficiently large error bars. However these data are sufficient for excluding one of the possible mechanisms on qualitative grounds.

Conclusions

The mathematical model introduced in this work describes complex processes involved in the actin tail formation at the vesicle living membrane. The proposed approach to the simulation of the considered problem can be easily modified to account for other mechanisms or interactions if anything of the like is discovered experimentally. Hence it can serve as a basis for the investigation of more general and complex biological phenomena.

Owing to this mathematical model two existing hypotheses of actin tail growth kinetics could be tested: i) the growth rate is proportional to the number of clustered proteins; ii) the growth rate is proportional to the number of proteins which join the cluster per unit of time. The results of this test lead to the conclusion that the first hypothetical mechanism cannot occur in practice (at least it is not observed in experimental data under the conditions given in [2]).

References: 1. *J. Taunton, B.A. Rowning, M.L. Coughlin, M. Wu, R.T. Moon, T.J. Mitchison, C.A. Larabell.* Actin-dependent propulsion of endosomes and lysosomes by recruitment of N-WASP // *Journal of Cell Biology.*—2000.—148.—pp. 519-530. 2. *C. Co, D.T. Wong, S. Gierke, V. Chang, J. Taunton.* Mechanism of actin network attachment to moving membranes: barbed end capture by N-WASP WH2 domains // *Cell.*—2007.—128.—pp. 901-913. 3. *A. Kusumi, C. Nakada, K. Ritchie, K. Murase, K. Suzuki, H. Murakoshi, R. S. Kasai, J. Kondo, T. Fujiwara.* Paradigm shift of the plasma membrane concept from the two-dimensional continuum fluid to the partitioned fluid: high-speed single-molecule tracking of membrane molecules // *Annu. Rev. Biophys. Biomol. Struct.* 34, 2005, pp. 351-378. 4. *P. Pelce.* New visions on form and growth.—Oxford University Press, 2004.—388 p. 5. *A. Lehninger.* Principles of Biochemistry.—W. H. Freeman & Co, 2004.—1110 p. 6. *W. Krauth.* Statistical Mechanics: Algorithms and Computations.—Oxford University Press, 2007.—320 p. 7. *C. Amatore.* Is there an Intrinsic Limit to the Size of 2D Supracrystals Built from Weakly Interacting Nanoparticles? // *Chem. Eur. J.*—2008.—14(28).—pp. 8615-8623.

Поступила в редколлегию 27.02.2009

УДК 57:51-76; 57.02.001.57; 517.958:57

Моделирование роста актинового волокна у везикул при помощи броуновского движения / А.И. Олейник // Бионика интеллекта: науч.-техн. журнал.—2009.—№ 1 (70).—С. 82-85.

В статье предложена математическая модель роста актинового волокна на поверхности везикулы и дано количественное описание данного процесса с использованием броуновского движения. Рассмотрены два гипотетических механизма образования актинового волокна, известные из литературы. Сравнение результатов моделирования, соответствующих этим механизмам с экспериментальными данными, позволили сделать заключение, что один из механизмов не приводит к поведению, наблюдаемому в эксперименте.

Ил.: 2. Библиогр.: 7 назв.

УДК 57:51-76; 57.02.001.57; 517.958:57

Моделирование роста актинового волокна у везикул за допомогою броунівського руху / О.І. Олійник // Біоніка інтелекту: наук.-техн. журнал. —2009. —№ 1(70). —С. 82-85.

У статті запропоновано математичну модель росту актинового волокна на поверхні везикули та наведено кількісний опис цього процесу з використанням броунівського руху. Розглянуто два гіпотетичних механізми створення актинового волокна, що відомі з літератури. Порівняння результатів моделювання, що відповідають цим механізмам, з експериментальними даними дозволило зробити висновок, що один з механізмів не призводить до поведінки, яка спостерігається в експерименті.

Іл.: 2. Бібліогр.: 7 найм.

УДК 004.93*1



РОБАСТНЫЕ АЛГОРИТМЫ САМООБУЧЕНИЯ КАРТЫ КОХОНЕНА В ЗАДАЧЕ ОБРАБОТКИ ИЗОБРАЖЕНИЙ

В. Я. Винарский¹, С. В. Машталир², Е. С. Сакало³, К. С. Щербинин⁴

¹ Университет прикладных наук и искусств, г. Ганновер, Германия, vladimir.vinarski@fh-hannover.de,
^{2, 3, 4} ХНУРЭ, г. Харьков, Украина, mashtalir_s@kture.kharkov.ua

В работе предложены алгоритмы робастного конкурентного самообучения самоорганизующейся карты Т. Кохонена (SOM), предназначенной для анализа изображений, когда обрабатываемые образы «загрязнены» возмущениями с аномальными выбросами (негауссовские помехи). Введенные алгоритмы вычислительно просты, обладают фильтрующими свойствами и предназначены для обработки информации в последовательном режиме по мере ее поступления.

НЕЙРОННАЯ СЕТЬ, КОХОНЕН, ОБРАБОТКА ИЗОБРАЖЕНИЙ

Введение

В задачах обработки изображений, связанных с сегментацией, кластеризацией, сжатием, выделением признаков и тому подобным, широкое распространение получили искусственные нейронные сети с латеральными связями, реализующие принцип конкурентного самообучения, введенные и теоретически обоснованные Т. Кохоненом [1–8]. Эти сети имеют простейшую однослойную архитектуру и образованы нейронами — адаптивными линейными ассоциаторами, при этом информация передается как в прямом направлении от входа к выходу сети, так и по поперечным латеральным каналам обмена данными, связывающим все нейроны между собой. Таким образом, целью работы является создание нейронной сети для обработки изображений в условиях воздействия на них помех разного рода.

1. Анализ существующих подходов

Обработка изображений с помощью самоорганизующейся карты обычно производится следующим образом: входное изображение размером $M \times N$ пикселей разбивается на блоки размером $m \times n$, каждый из которых преобразуется в векторную форму размерности $mn \times 1$. Затем с помощью обучающей выборки, содержащей $MN(mn)^{-1}$ векторов образов $x(k)$, $k = 1, 2, \dots, MN(mn)^{-1}$, пронормированных в принятой метрике, нейронная сеть настраивается с помощью принятого правила самообучения. При этом предполагается, что архитектура в слое Кохонена содержит p нейронов с синаптическими весами, заданными в форме $mn \times 1$ — векторов $w_q = (w_{q,1}, w_{q,2}, \dots, w_{q,mn})^T$, $q = 1, 2, \dots, p$. После замены каждого из q -го векторов $x(k)$ на номер нейрона-победителя j , соответствующего этому образу, запоминаем номер j , соответствующий этому образу, и соответствующий ему вектор синаптических весов w_j^* . Для восстановления исходного изображения каждый номер нейрона-победителя j заменяется на w_j^* , девекторизованный к матричной форме. Понятно, что чем меньше отношение $p \cdot mn / MN$, тем выше эффективность сети.

В основе процедуры самоорганизации карты Кохонена лежат принципы конкурентного самообучения, а ее работа, как и любого другого алгоритма обучения, начинается с инициализации синаптических весов сети, которые обычно выбираются случайным образом, при этом весьма желательно, чтобы для каждого из нейронов выполнялось условие $\|w_q(0)\| = 1$. Сама же процедура самоорганизации выполняется в три основных этапа [9]: конкуренции, кооперации, синаптической адаптации.

Процесс конкуренции начинается с анализа образа $x(k)$, поступающего с рецепторного слоя на все нейроны слоя Кохонена. Для каждого из нейронов вычисляется его выходной сигнал

$$y_q(k) = w_q^T(k-1)x(k), q = 1, 2, \dots, p$$

и расстояние (как правило, в евклидовой метрике)

$$D(x(k), w_q(k-1)) = \|x(k) - w_q(k-1)\|,$$

причем если входы и синаптические веса предварительно пронормированы так, что $\|x(k)\| = \|w_q(k-1)\| = 1$, а в качестве расстояния используется традиционная евклидова метрика, то мерой близости между векторами $x(k)$ и $w_q(k-1)$ может служить также скалярное произведение

$$D(x(k), w_q(k-1)) = w_q^T(k-1)x(k) = \cos(w_q(k-1), x(k)) = y_q(k).$$

Далее определяется нейрон-победитель, ближайший ко входному образу такой, что

$$D(x(k), w_j^*(k-1)) = \min D(x(k), w_q(k-1)),$$

то есть

$$\|x(k) - w_j^*(k-1)\| < \|x(k) - w_q(k-1)\|, \forall q \neq j,$$

после чего в простейшем случае, «перепрыгивая» через этап кооперации, можно подстроить синаптические веса сети с помощью правила конкурентного самообучения [1]

$$w_j(k) = \begin{cases} w_j(k-1) + \eta(k)(x(k) - w_j(k-1)), & \text{если } j\text{-нейрон победил} \\ w_j(k-1) & \text{в противном случае.} \end{cases} \quad (1)$$

Процедура (1) реализует принцип «победитель получает все» (WTA), при этом вектор синаптических весов нейрона-победителя подтягивается ко входному образу на расстояние, определяемое шагом поиска $\eta(k)$.

Регулирование величины шага поиска обычно производится, исходя из эмпирических соображений [9], а общая рекомендация сводится к тому, что он должен монотонно уменьшаться в процессе самообучения. С этой рекомендацией хорошо согласуется введенная в [10] процедура

$$\eta(k) = p^{-1}(k),$$

$$p(k) = \alpha p(k-1) + \|x(k)\|^2, \quad 0 \leq \alpha \leq 1,$$

которая для нормированных входов изображений простой вид

$$p(k) = \alpha p(k-1) + 1,$$

а при $\alpha = 1$ автоматически дает оценку

$$\eta(k) = \frac{1}{k},$$

удовлетворяющую условиям стохастической аппроксимации А. Дворецкого.

Можно заметить, что правило настройки (1) является обобщением на многомерный случай простейшего линейного абсолютно-оптимального алгоритма обучения [11], минимизирующего критерий

$$E^k = \sum_{s=1}^k (x(s) - w)^2,$$

где $x(s)$ и w – скаляры.

С позиций же теории оптимизации (1) есть не что иное, как градиентная процедура стохастической аппроксимации, минимизирующая критерий

$$E_j^k = \sum_{s=1}^k \|x(s) - w\|^2,$$

и ведущая, при $\eta(k) = \frac{1}{k}$, к векторной оценке среднего арифметического

$$w_j(k) = \frac{1}{k} \sum_{s=1}^k x(s),$$

при этом естественно предполагается, что j -й нейрон k раз оказывается победителем.

Этап же кооперации предполагает, что нейрон-победитель определяет так называемую локальную область топологического соседства, в которой обучается не только он сам, но и его ближайшее окружение, при этом более близкие к победителю нейроны активируются сильнее, чем удаленные. Эта топологическая область определяется функцией соседства $h_{jq}(k)$, зависящей от расстояния $D(w_j^*(k), w_q(k))$ между победителем $w_j^*(k)$ и любым из нейронов слоя Кохонена $w_q(k)$ и некоторого скалярного параметра σ , задающего ее «ширину». Как правило, $h_{jq}(k)$ – это потенциальная функция, симметричная относительно максимума в точке $w_j^*(k)$ ($D(w_j^*(k), w_j^*(k)) = 0$), принима-

ющая в ней единичное значение и монотонно стремящаяся к нулю с ростом расстояния. Часто в качестве функции соседства используются гауссиан [12], треугольная функция [13], перевернутая парабола [14], «мексиканская шляпа» [15] и целый ряд других.

Использование функции соседства приводит к правилу самообучения вида

$$w_q(k) = w_q(k-1) + \eta(k) h_{jq}(k-1) (x(k) - w_q(k-1)), \quad (2)$$

реализующему принцип «победитель получает больше» (WTA) вместо традиционного WTA. Естественно, что при $h_{jq}(k-1) = \delta_{jq} = 1$ при $j = q$ и 0 в остальных случаях, вновь приходим к стандартному алгоритму (1), обеспечивающему на каждом шаге настройку единственного нейрона победителя w_j^* . Использование же в качестве $h_{jq}(k)$ потенциальных функций ведет к тому, что все нейроны сети в большей или меньшей мере подтягивают векторы своих синаптических весов $w_q(k-1)$ к предъявляемому образу $x(k)$.

Можно показать [1], что алгоритм (2) минимизирует критерий

$$E_j^k = \sum_{s=1}^k h_{jq}(s) \|x(s) - w_j\|^2$$

и сходится к оценке взвешенного среднего арифметического

$$w_j(k) = \frac{\sum_{s=1}^k h_{jq}(s) x(s)}{\sum_{s=1}^k h_{jq}(s)}.$$

Известно, что статистические оценки, связанные с квадратичными критериями и средними арифметическими, являются эффективными лишь в случае, когда наблюдения, принадлежащие конкретной группе (кластеру), подчинены нормальному закону распределения. Если же распределение данных в кластере отличается от гауссовского, а сами данные загрязнены аномальными выбросами (что достаточно часто бывает в задачах обработки изображений), эффективность оценок среднего арифметического быстро падает с «утяжелением» хвостов распределений [16]. Ясно, что в этом случае алгоритмы (1) и (2) также становятся неэффективными.

2. Медианный алгоритм Кохонена

В случае, когда в данных присутствуют наблюдения, резко выделяющиеся на фоне остальных, целесообразно использовать критерии, менее реагирующие на выбросы. Таковым, например, является достаточно популярный критерий наименьших модулей [17]

$$E_j^k = \sum_{s=1}^k \sum_{i=1}^{mn} |x_i(s) - w_{ji}|, \quad (3)$$

порождаемый манхэттенской метрикой

$$D(x(k), w_q(k-1)) = \sum_{i=1}^{mn} |x_i(k) - w_{qi}(k-1)|.$$

Алгоритм обучения самоорганизующейся карты в этом случае приобретает вид [1]

$$w_j(k) = \begin{cases} w_j(k-1) + \eta(k) \text{sign}(x(k) - w_j(k-1)), \\ \text{если } j\text{-нейрон победил} \\ w_j(k-1) \text{ в противном случае,} \end{cases} \quad (4)$$

где $\text{sign } x(k) = (\text{sign } x_1(k), \text{sign } x_2(k), \dots, \text{sign } x_{mn}(k))^T$.

Минимизация критерия (3) приводит к оценке, являющейся выборочной медианой, которая по сравнению с оценками, основанными на вычислении среднего, менее чувствительна к различного рода выбросам. Фактически с помощью алгоритма (4) производится не только кластеризация данных, но и их медианная фильтрация.

Известно [17], что медиана является эффективной оценкой математического ожидания случайной величины, когда данные распределены по закону Лапласа с функцией плотности распределения

$$f(x_i, w_i) = \frac{1}{2\sigma_i} \exp\left(-\frac{|x_i - w_i|}{\sigma_i}\right),$$

где σ_i – параметр ширины (масштаба) распределения по i -й координате, в общем случае неизвестный; w_i – математическое ожидание, подлежащее оцениванию.

Оптимальная оценка математического ожидания в этом случае связана с минимизацией критерия

$$E_j^k = \sum_{s=1}^k \sum_{i=1}^{mn} \frac{1}{\sigma_i} |x_i(s) - w_{ji}|$$

и может быть получена путем обобщения на многомерный случай абсолютно оптимального релейного алгоритма оценивания [18] и его адаптации к форме (4), то есть

$$w_j(k) = \begin{cases} w_j(k-1) + \frac{\Sigma}{k} \odot \text{sign}(x(k) - w_j(k-1)), \\ \text{если } j\text{-нейрон победил} \\ w_j(k-1) \text{ в противном случае,} \end{cases} \quad (5)$$

где $\Sigma = (\sigma_1, \sigma_2, \dots, \sigma_{mn})^T$; \odot – символ прямого (скоттова) произведения.

Поскольку вектор Σ в (5) неизвестен, в [20] был введен одномерный алгоритм настройки параметра масштаба, который применительно к рассматриваемому случаю приводит к обобщению процедуры (5) в форме

$$w_j(k) = \begin{cases} w_j(k-1) + \frac{\Sigma(k)}{k} \odot \text{sign}(x(k) - w_j(k-1)), \\ \Sigma(k) = \Sigma(k-1) - \frac{1}{k}(\Sigma(k-1) - |x(k) - w_j(k-1)|), \\ \text{если } j\text{-нейрон победил,} \\ w_j(k-1) \text{ в противном случае.} \end{cases} \quad (6)$$

Переходя к этапу кооперации следует заметить, что поскольку медиана является ранговой оценкой, для построения функции соседства целесообразно использовать также ранговые конструкции, к каковым, например, относится так называемый «нейронный газ» [19].

В рамках этой концепции все нейроны карты Кохонена ранжируются в порядке возрастания расстояний так, что

$$D(x(k), w_j^*(k-1)) < \\ < D(x(k), w^1(k-1)) < \dots < D(x(k), w^{mn-1}(k-1)),$$

где верхний индекс обозначает не номер нейрона в сети, а его ранг в смысле расстояния от предъявленного образа $x(k)$, то есть

$$R(w^*(k-1)) = 0 < R(w^1(k-1)) = 1 < \\ < R(w^2(k-1)) = 2 < \dots < R(w^{mn-1}(k-1)) = mn-1.$$

Далее для каждого из нейронов определяются значения функции соседства

$$h_q^*(k-1) = \exp\left(-\frac{R(w_q(k-1))}{\sigma(k)}\right) \quad (7)$$

(здесь $\sigma(k)$ – параметр ширины) и производится уточнение синаптических весов сети согласно правилу

$$w_q(k) = w_q(k-1) + \eta(k) h_q^*(k-1) \text{sign}(x(k) - w_q(k-1)), \quad (8)$$

$q = 1, 2, \dots, p$,

являющемуся робастной модификацией алгоритма (2).

Несложно видеть, что параметры нейрона с нулевым рангом $w^*(k)$ (фактически нейрона-победителя) при этом уточняются согласно первому соотношению алгоритма (4).

Аналогично параметру шага $\eta(k)$ параметр ширины функции соседства (7) также должен монотонно уменьшаться в процессе самообучения, для чего в [19] была предложена простая процедура, которая для рассматриваемого нами случая может быть записана в виде

$$\begin{cases} \eta(k) = \eta(0) \left(\frac{\eta(\min)}{\eta(\max)}\right)^{\frac{kmn}{MN}}, \\ \sigma(k) = \sigma(0) \left(\frac{\sigma(\min)}{\sigma(\max)}\right)^{\frac{kmn}{MN}}, \end{cases} \quad (9)$$

где $\eta(\min)$, $\sigma(\min)$ и $\eta(\max)$, $\sigma(\max)$ – минимально и максимально возможные значения соответствующих параметров.

Заметим, что соотношения (9) также удовлетворяют условиям А. Дворецкого.

3. Комбинированный робастный алгоритм самообучения карты Кохонена

Следует заметить, что в «чистом» виде ни гауссовское, ни лапласово распределение на практике не встречается, в связи с чем Я. З. Цыткиным [18] было рассмотрено приблизительно нормальное распределение, отличающиеся от классического гауссовского наличием «тяжелых» хвостов. Функция плотности этого распределения имеет вид

$$f(x_i, w_i) = \frac{1}{2\sigma_i} \sec^2 h^2 \left(\frac{x_i - w_i}{\sigma_i}\right), \quad (10)$$

связанная с ним метрика –

$$D(x(k), w_q(k-1)) = \sum_{i=1}^{mn} \sigma_i \ln \cosh \frac{x_i(k) - w_{qi}(k-1)}{\sigma_i}, \quad (11)$$

а критерий обучения –

$$E_j^k = \sum_{s=1}^k \sum_{i=1}^{mn} \sigma_i \ln \cosh \frac{x_i(k) - w_{ji}(k-1)}{\sigma_i}.$$

Далее несложно записать алгоритм самообучения карты Кохонена на основе WTA-принципа в виде

$$w_j(k) = \begin{cases} w_j(k-1) + \eta(k) \tanh \frac{x(k) - w_j(k-1)}{\sigma}, \\ \text{если } j\text{-нейрон победил,} \\ w_j(k-1) \text{ в противном случае,} \end{cases}$$

где $\tanh \frac{x(k) - w_j(k-1)}{\sigma} = (\tanh \frac{x_1(k) - w_{j1}(k-1)}{\sigma_1}, \dots, \tanh \frac{x_{mn}(k) - w_{jmn}(k-1)}{\sigma_{mn}})^T$, а параметры в данном случае определяют крутизну функции гиперболического тангенса, при этом чем меньше значение σ_i , тем ближе алгоритм (12) по свойствам к медианной процедуре (4). С ростом же σ_i (12) приближается по своим свойствам к стандартному правилу (1).

Что же касается выбора функции соседства, резонно предположить, что если для стандартной процедуры (2) в качестве таковой наиболее часто используется гауссиан, то в данном случае с таким же успехом может быть применена функция (10) в виде

$$h_q(k-1) = \frac{1}{2\sigma_i} \sec h^2 D(w_j^*(k-1), w_q(k-1)),$$

где $D(\cdot, \cdot)$ определяется выражением (11).

Тогда комбинированный робастный алгоритм самообучения карты Кохонена на основе WTM-принципа окончательно может быть записан в виде

$$w_q(k) = w_q(k-1) + \eta(k) h_q(k-1) \tanh \frac{x(k) - w_q(k-1)}{\sigma}, \quad (13)$$

$q = 1, 2, \dots, p$.

Алгоритмы (12) и (13) в зависимости от принятого вектора параметров могут обладать свойствами как традиционных правил Кохонена (1), (2), так и их медианных модификаций (4), (5), обеспечивая устойчивую обработку информации при наличии различного рода возмущений и помех.

Выводы

В работе предложена группа робастных алгоритмов самообучения самоорганизующейся карты Кохонена в задачах обработки изображений, искаженных различного рода шумами и возмущениями, содержащими аномальные выбросы. Алгоритмы достаточно просты в вычислительном отношении и предназначены для обработки данных по мере их поступления в последовательном режиме.

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: Clarandon 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., Mavrovouniotis 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 Connection Record. –1960. –Part 4. –P. 96-105. 8. Бодянский Е.В., Плисс И.П., Соловьева Т.В. Многошаговые оптимальные упреждители многомерных нестационарных стохастических процессов // Докл. АН УССР. –1986. –Сер. А. – № 12. – С. 47-49. 9. Haykin S. Neural networks. A comprehensive foundation. – Upper saddle river, N.J.: Prentice Hall, Inc., 1999. –842 p. 10. Бодянский Е.В., Королькова Е.Е., Ламонова Н.С. Модифицированные алгоритмы самообучения самоорганизующихся карт Т. Кохонена // Проблемы бионики. –2003. –Вып. 58. –С. 3-9. 11. Пыткин Я.З. Основы теории обучающихся систем. – М.: Наука, 1970. –252с. 12. Lo Z.-P., Yu Y., Bavarian B. Analysis of the convergence properties of topology preserving neural networks // IEEE Trans/ on Neural Networks. –1993. –4. –P. 207-220. 13. Rojas R. Neural Networks. A Systematic Introduction. –Berlin: Springer-Verlag, 1996. –502p. 14. Bodianskiy Ye., Chaplanov O., Kolodyazhnyi V., Otto P. Adaptive quadratic radial basis function network for time series forecasting / Proc. East West Fuzzy Coll. –Zittau-Goerlitz & HS, 2002. –P. 164-172. 15. Advances in Intelligent Control / Ed. By C.J. Harris. –London: Taylor and Francis, 1994. –373p. 16. Смоляк С.А., Титаренко Б.П. Устойчивые методы оценивания. –М.: Статистика, 1980. –208 с. 17. Мудров В.И., Кушко В.А. Методы обработки измерений. – М.: Сов. Радио, 1976. –192 с. 18. Пыткин Я.З. Основы информационной теории идентификации. – М.: Наука, 1984. –320с. 19. Martinez M., Berkovich S., Schulten K. «Neural gas» network for vector quantization and its application to the time series prediction // IEEE Trans. on Neural Networks. –1993. –4. –P. 558-569.

Поступила в редколлегию 02.03.2009

УДК 004.93'1

Робастні алгоритми самовивчення карти Кохонена до задачі обробки зображень / В.Я. Вінарський, С.В. Машталір, Є. С. Сакало, К. С. Шербінін // Біоніка інтелекту: наук.-техн. журнал. –2009. – №1(70). – С. 86–89.

В роботі запропоновані алгоритми робастного конкурентного самовивчення карти Т. Кохонена (SOM), яка призначена для аналізу зображень у випадку наявності збурень різного роду.

Бібліогр.: 20 найм.

UDC 004.93'1

Robust Algorithms of self-learning Kohonan maps in image processing / V. Vinarski, S. Mashtalir, S. Sakalo, K. Scherbinin // Bionics of intelligence: Sci. Mag. –2009. – № 1 (70). – P. 86-89.

We proposed algorithms for robust self-learnig competition T. Kohonen maps (SOM), which is designed to image analysis in case of disturbances.

Ref.: 20 items.

УДК 681.3.07



НОРМАЛИЗАЦИЯ ПРОЕКТИВНЫХ ПРЕОБРАЗОВАНИЙ МЕТОДОМ ОДНОМЕРНЫХ КОРРЕЛЯЦИЙ

Любченко В. А.¹, Кобылин О. А.²^{1, 2}ХНУРЭ, Харьков, Украина

kblin@kture.kharkov.ua

За счет перехода от двумерного изображения к его ограничениям на прямые переходим от нормализации девятипараметричной проективной группы на плоскости к нормализации проективной группы на прямой с двумя параметрами. Использование математических моделей разложения полной проективной группы на базовые подгруппы и результатов одномерной нормализации позволяет построить алгоритм нормализации полной проективной группы.

НОРМАЛИЗАЦИЯ ИЗОБРАЖЕНИЙ, ГРУППА ПРОЕКТИВНЫХ ПРЕОБРАЗОВАНИЙ, ОГРАНИЧЕНИЯ НА ПРЯМЫЕ, СИСТЕМА КОМПЬЮТЕРНОГО ЗРЕНИЯ, ЭТАЛОН

Введение

В настоящее время интенсивное развитие технических средств позволяет решать все более широкий класс задач, связанных с обработкой изображений, возникающих, в частности, при распознавании образов. Однако уровень развития соответствующего математического обеспечения значительно отстает от возможностей технических систем. Поэтому созданию новых моделей и методов, позволяющих повысить эффективность распознавания соответствующих объектов технической системой или человеком-оператором, уделяется большое внимание. Среди подобных задач важное место занимают задачи, связанные с обработкой изображений, подверженных геометрическим искажениям [1].

Чаще всего объектом исследования в системах компьютерного зрения (СКЗ) является трехмерный объект O , не изменяющий своих абсолютных размеров в евклидовом пространстве. Поведение такого объекта описывается аффинной моделью в пространстве. При получении двумерных снимков такого объекта образуется множество изображений, связанных между собой проективными преобразованиями [2]. Если рассматривать геометрические искажения как результат действия некоторой группы G , то множество входных изображений B , отвечающих некоторому эталону B_0 , рассматривают как класс эквивалентности W заданный эталоном $W(B_0) = \{B_1, B_2, \dots, B_s\}$. Таким образом, современная СКЗ должна учитывать преобразования, возникающие под воздействием геометрических искажений.

Одним из известных методов, работающих в рамках геометрических преобразований, является нормализация [1]. Процедура нормализации заключается в автоматическом определении неизвестных параметров преобразования, которому подвергнуто входное изображение, и последующем приведении его к эталонному виду. В процессе нормализации изображение замещается эквивалентным ему. Преобразова-

ния осуществляются с помощью операторов нормализации F , называемых нормализаторами, а вычисление параметров выполняется функционалами Φ , действующими на множестве изображений. Существуют параллельные, последовательные, параметрические и следящие нормализаторы. Однако известные нормализаторы эффективно работают только для базовых преобразований, таких как смещения, повороты, растяжения, косые сдвиги.

Сложность проективной группы обуславливается тем, что изображение под действием проективного преобразования изменяется непропорционально (в отличие от аффинных преобразований) [4], поэтому известные методы нормализации, хорошо работающие в рамках аффинной модели восприятия, непригодны.

Под изображением будем понимать отсегментированный снимок B исследуемого объекта O в некотором поле зрения D . Для математического описания изображения используют функцию от двух переменных $B(x, y)$, которая обозначает точку (пиксел) изображения B с координатами (x, y) . Значение этой функции определяет интенсивность (яркость) в точке с различными уровнями градации в зависимости от формата изображения. Под интерпретацией (распознаванием) изображения понимают сравнение входного изображения B (признаков входного изображения) с заранее заданным эталонным изображением B_0 (признаками эталонного изображения). Зависимость входного и эталонных изображений в рамках проективной модели зрения описывается математической моделью:

$$B(x', y') = B_0 \left(\frac{b_{11}x + b_{12}y + b_{13}}{b_{31}x + b_{32}y + b_{33}}, \frac{b_{21}x + b_{22}y + b_{23}}{b_{31}x + b_{32}y + b_{33}} \right), \quad (1)$$

где детерминант матрицы Π преобразования не равен нулю: $\det(\Pi) \neq 0$.

Существует только один классический метод в теории обработки изображений, работающий с моделью зрения (1) – корреляционный [1]. Он заключается в

нахождении меры близости входного изображения и эталонного. Пусть определена некоторая мера близости $\xi(B_0, B_k)$. Определим через $\bar{\xi}$ значение меры в идеальном случае, то есть $\bar{\xi} = \xi(B_0, B_k)$. Тогда корреляционный метод нормализации будет заключаться в нахождении такого преобразования $g \in G$, при котором значение $\xi_g = \xi(B_0, gB_k)$ будет наиболее близко к истинному значению $\bar{\xi}$. На практике задают порог близости ε , при котором некоторое входное изображение B_k (k – задает номер изображения из множества B_0 , если существует $g \in G$, при котором выполняется условие $|\bar{\xi} - \xi_g| < \varepsilon$. Коэффициенты корреляции могут вычисляться различными способами, после чего выбирают минимальный или максимальный из них. Классической мерой сходства является определение максимума функции:

$$\xi_g = \frac{\sum_{(x,y) \in D} B(x,y)B_0(x,y)}{\sqrt{\sum_{(x,y) \in D} B^2(x,y)} \sqrt{\sum_{(x,y) \in D} B_0^2(x,y)}} \rightarrow \max_{g \in G}, \quad (2)$$

где D – поле зрения; $B(x, y)$, $B_0(x, y)$ – входное и эталонное изображения соответственно.

Из (2) следует, что ξ_g стремится к единице и в идеале $\bar{\xi} = 1$. На основании найденных параметров $g \in G$ можно построить некоторый оператор, для которого будет выполнимо условие:

$$F(B(x, y)) = gB(x, y) = B_0(x, y), \quad (3)$$

который и является нормализатором относительно действия группы G . Из определения нормализатора (3) следует, что для любых изображений $B_1, B_2 \in W$ справедливо $F(B_1) = F(B_2)$, что дает возможность построения нормализатора без определения некоторого эталонного изображения. Построение нормализатора определяет задачу построения отображения $\Phi: W \rightarrow G$, то есть для каждого $B_i \in W$ находим элемент $g_i \in G$ такой, что $F(B) = gB$; иными словами, отображение Φ должно удовлетворять условию: $\Phi(gB) = \Phi(B)g^{-1}$.

Можно считать, что для группы G и множества входных изображений W выполнено условие единственности: если $g_e B = B$, то $g_e = e$ является единицей группы. Тогда выполняется равенство $(\Phi(B))^{-1} \Phi(gB)g = e$, из которого следует, что единицей группы G является преобразование $\Phi(B_0)$, поэтому $\Phi(B) = g^{-1}$ обеспечивает определение неизвестных параметров преобразования группы G путем построения отображения $\Phi: W \rightarrow G$ [1].

Очевидно, что корреляционный метод нормализации наиболее надежный, но зато и наиболее трудоемкий, поскольку предполагается полный перебор в пространстве сигналов. Трудоемкость метода корреляции напрямую зависит от количества параметров группы G . Увеличение этого количества приводит к столь значительным вычислениям, что непосредственное приме-

нение корреляционных алгоритмов порой невозможно в реальном режиме времени. Это обусловило создание модификаций корреляционных методов, которые обеспечивали бы ту же надежность при сокращении трудоемкости. К таким модификациям относятся методы частных корреляций. Основная идея их состоит в том, что входное и эталонное изображения разбиваются некоторым образом на участки, для каждого участка вычисляются свои коэффициенты корреляции, которые сворачиваются в один общий коэффициент корреляции по определенному правилу. В рамках многопараметрической группы преобразований способ разбиения изображения на участки остается открытым.

В работе [3] показано, что полная проективная группа преобразований G_{II} может быть представлена в виде композиции базовых. Поскольку композиция преобразований равна произведению матриц соответствующих преобразований, в дальнейшем при описании разложений многопараметрических группы будем использовать обозначение матриц:

$$II = DP_y P_x H_x UC_{xy}, \quad (4)$$

где C_{xy} – смещение; D – сжатие вдоль осей; P_y, P_x – преобразования перспектив вдоль осей ординат и абсцисс соответственно; H_x – косой сдвиг вдоль оси абсцисс; U – поворот [1, 3]. На основе разложения (4) построим последовательный нормализатор проективной группы в следующем виде:

$$F_{II} = F_D F_{P_y} F_{P_x} F_{H_x} F_U F_{C_{xy}}, \quad (5)$$

где F_{II} – нормализатор центропроективной группы; $F_{C_{xy}}$ – нормализатор смещений; F_U – нормализатор группы поворотов; F_{P_y}, F_{P_x} – нормализаторы группы перспективы вдоль осей ординат, абсцисс соответственно; F_{H_x} – нормализатор группы косоугольного сдвига вдоль оси абсцисс; F_D – нормализатор группы сжатий.

Для использования нормализатора (5) и метода частичных корреляций введем следующее определение и выполним ряд построений.

Определение. Ограничением $B(\ell_k)$ изображения B на прямую ℓ_k называется их общая часть: $B(\ell_k) = B \cap \ell_k$, то есть та часть изображения, которая совпадает с расположением прямой ℓ_k .

Пусть семейство $\{\ell_k\}$ состоит из s ($k = \overline{1, s}$) прямых, равноотстоящих друг от друга, параллельных оси ординат и находящихся в поле зрения D . Тогда ограничения на прямые $B(\ell_k)$ формируются из тех участков изображения B , которые пересекаются с прямыми рассматриваемого семейства (рис. 1).

Аналогично можно получить ограничения на горизонтальные прямые. По существу, дискретизация – это и есть ограничения на вертикальные или горизонтальные прямые с достаточно малым шагом. Однако использование горизонтальных и вертикальных прямых при сложных преобразованиях часто ограничено по возможностям. Более универсальным является анализ

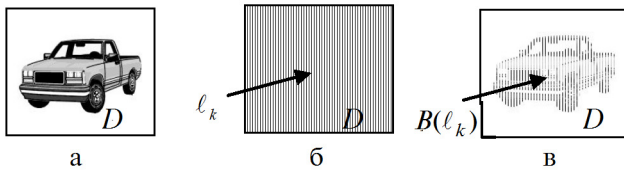


Рис. 1. Переход от изображения к его ограничениям на семейство прямых $\{l_k\}$, параллельных оси ординат: а – исходное изображение B ; б – семейство прямых $\{l_k\}$; в – ограничения $B(l_k)$ изображения B на семейство прямых $\{l_k\}$.

прямых, проходящих через заданную точку. Заддим некоторую опорную точку и рассмотрим набор прямых $\{l_k\}$, проходящих через нее, с одинаковым углом между l_k и l_{k+1} . Пусть задано эталонное и искаженное изображения (рис. 2).



Рис. 2. Тестовые изображения:

а – эталонное изображение; б – входное изображение.

Случайным образом выберем на эталоне точку M и перейдем от изображения к его ограничениям на прямые, проходящие через точку M (рис. 3а), и представим полученные семейства прямых в виде матрицы (рис. 3б).

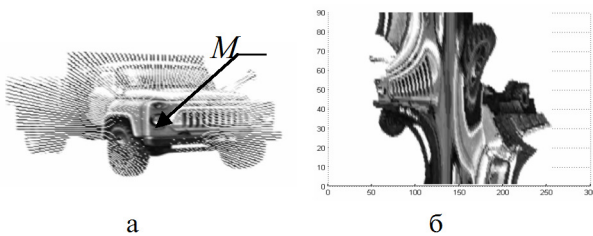


Рис. 3. Представление эталона в виде ограничений на 180 прямых: а – ограничения на прямые; б – визуализация матрицы ограничений

Входное изображение было получено путем воздействия на эталонное проективным преобразованием (рис. 4а). Построим теперь множество ограничений на прямые $\{N_j\}$, проходящие через точку N , на основе искаженного изображения (рис. 4б).

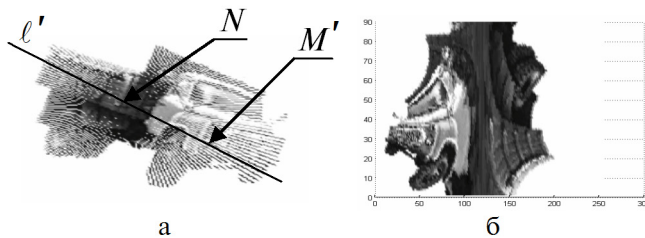


Рис. 4. Представление входного изображения в виде ограничений на 180 прямых: а – ограничения на прямые; б – матрица ограничений

Из классической проективной геометрии известно, что прямая переходит в прямую, следовательно, прямая l перейдет в прямую l' . Поскольку прямая l содержит точку M , а прямая l' – точку N , то можно сделать вывод, что существует некоторая прямая M_i из множества $\{M_i\}$, которая соответствует прямой M_i , и существует некоторая прямая N_j из множества $\{N_j\}$, которая соответствует прямой l' . А поскольку прямая l' является образом прямой l , то и прямая N_j является образом прямой M_i . Эти прямые связаны одномерной проективной группой преобразований P_o

$x' = \frac{mx}{tx+1}$, матрица которого может быть представлена в виде разложения

$$P_o = \begin{pmatrix} m & 0 \\ t & 1 \end{pmatrix} = \begin{pmatrix} m & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix} = D_o P_o,$$

где D_o – одномерное сжатие, P_o – преобразование перспективы [3]. Таким образом, одним из способов нормализации проективной прямой является корреляционная процедура по 2-м параметрам: сжатию и перспективе, с помощью которой можно в множествах $\{M_i\}$ и $\{N_j\}$ найти соответствующие l и l' с некоторыми параметрами одномерного сжатия и перспективы, связывающие эти прямые. На рис. 5 приведены примеры найденных прямых l и l' .

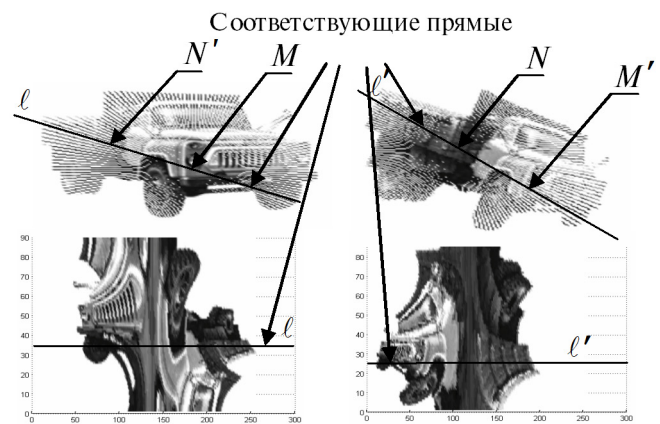


Рис. 5. Образ прямой эталонного изображения в проективно искаженном изображении

На следующем этапе произвольным образом выберем на эталоне и входном изображениях другие, отличные от M и N , точки D и K соответственно, относительно которых и получим ограничения на прямые рис. 6.

Таким образом, найдем прямую η' входного изображения, которая является образом прямой η эталона. В результате получаем две прямые l и η , для которых известны их образы l' и η' . Пусть прямые l и η пересекаются в точке O , тогда точка O' , расположенная на пересечении прямых l' и η' , является ее образом. Используя координаты O и O' можно установить параметры нормализатора F_{Cy} , который вы-

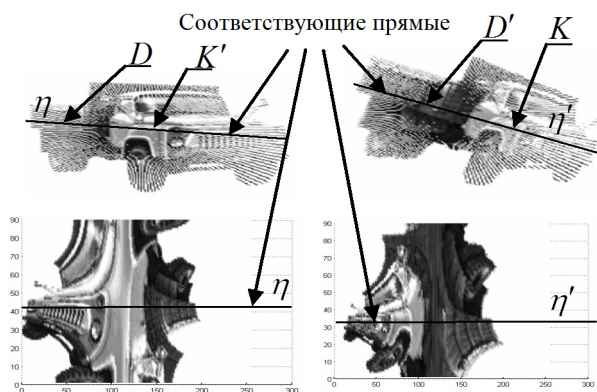


Рис. 6. Образ прямой эталонного изображения в проективно искаженном изображении через точки D и K

полнит центрирование двух изображений. Расположим начало координат эталонного и входного изображений соответственно в точки O и O' (рис. 7).

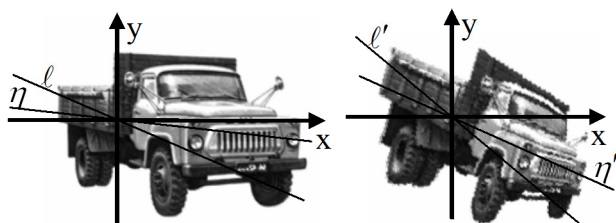


Рис. 7. Изображения с соответствующими прямыми: а – изображение эталона; б – входное изображение

Далее прямые, полученные на входном и эталонном изображениях, для удобства переносим на одну ось (рис. 8).

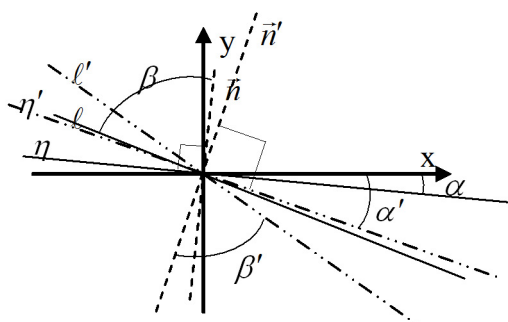


Рис. 8. Геометрическое представление полученных прямых

На рис. 9 α – угол между η и осью Ox ; α' – угол между η' и осью Ox ; β – угол между ℓ и \vec{n} ; \vec{n} – вектор нормали к прямой η ; β' – угол между ℓ' и \vec{n}' , где \vec{n}' – вектор нормали к прямой η' ; m, t – параметры сжатия и перспективы вдоль вектора ℓ' : m – сжатие, t – перспектива; n, k – параметры сжатия и перспективы вдоль вектора η' .

Нормализацию изображения (рис. 3. б) по полученным параметрам преобразования рассмотрим на при-

мере (рис. 10), где на найденные прямые эталонного и входного изображения наложены сетки.

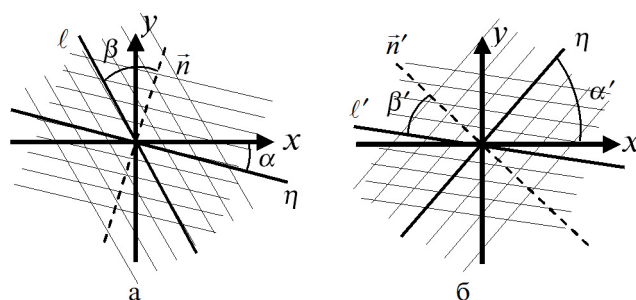


Рис. 9. Графическое представление изображения: а – изображение эталона; б – входное изображение

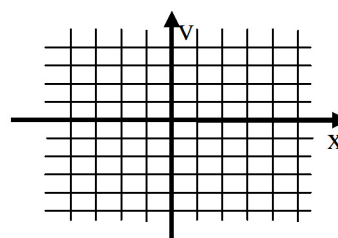


Рис. 10. Нормализованная сетка изображения

Необходимо осуществить переход от эталонного и входного изображений (рис. 10) к некоторым новым изображениям, чтобы нанесенная сетка приняла вид, представленный на рис. 10. Для этого воспользуемся последовательным нормализатором (5): воздействуем на эталонное и входное изображения нормализаторами соответственно поворота $F_{U(\alpha)}$, $F_{U(\alpha')}$ и косоуго сдвига $F_{H_x(\beta)}$, $F_{H_x(\beta')}$. То есть повернем эталонное и входное изображения таким образом, чтобы прямые изображения η и η' совпали с осью абсцисс, и воздействуем преобразованием косоуго сдвига соответственно на угол β и β' , чтобы между прямыми η и ℓ , η' и ℓ' был угол 90 градусов. Затем приведем входное изображение, полученное на предыдущем шаге, к видоизмененному эталонному, воспользовавшись нормализаторами перспектив вдоль оси абсцисс $F_{P_x(\text{tg}(\beta)^*t)}$ и оси ординат $F_{P_y(\text{tg}(\beta')^*k)}$. Полученная пара отличается только преобразованием сжатия, которое компенсируем нормализатором $F_{D(m,n)}$. В результате получим нормализованные изображения.

Выводы

В статье предложен новый метод нормализации проективных искажений, использующий один из инвариантов проективной геометрии: прямая при преобразовании переходит в прямую. Метод перехода от двумерных изображений к их ограничениям на прямые позволяет перейти от сложной задачи определения девяти параметров проективной группы к одномерной проективной группе с двумя параметрами. Разработан метод построения последовательного нормализатора исходного изображения, используя параметры, полученные при нормализации ограничений на прямые.

Список литературы: 1. Путьтин Е. П., Аверин С. И. Обработка изображений в робототехнике. – М.: Машиностроение, 1990. – 320 с. 2. Моденов П. С. Аналитическая геометрия. – М.: МГУ, 1969. – 598 с. 3. Любченко В. А., Путьтин Е. П. Математические модели разложения проективных преобразований в задачах нормализации // Радиоэлектроника и информатика. – 2002. – №2(19). – С. 57–59. 4. Любченко В. А., Яковлева Е. В., Передрий Е. О. Нормализация перспективных преобразований проективно искаженных изображений // Вестник НТУ «ХПИ». – 2007. – №26. – С. 174–184.

Поступила в редколлегию **04.03.2009**

УДК 681.3.07

Нормалізація проєктивних перетворень методом одновимірних кореляцій / В. А. Любченко, О. А. Кобилін // Біоніка інтелекту: наук.-техн. журнал. – 2009. – №1(70). – С. 90–94.

Розглядається нормалізація проєктивних перекручень на зображенні із застосуванням одновимірних кореляцій. Перехід від двовимірного зображення до його обмежень на прямі забезпечує можливість визначити невідомі параметри повної дев'ятипараметричної проєктивної групи перетворень, що дає змогу отримати нормалізатор зображення для систем комп'ютерного зору.

Лл.: 10. Бібліогр.: 4 найм.

UDK 681.3.07

Project transformation normalization via one-dimensional correlations method / V. A. Lyubchenko, O. A. Kobylin // Bionics of Intelligence: Sci. Mag. – 2009. – №1(70). – P. 90–94.

Normalization of the image project distortions by using one-dimensional correlations is considered. The transition from 2D image to its restriction on lines allows to detect unknown parameters of the full nine-parameters projective transformation group, what gives us image normalization tool (normalizator) for computer vision system.

Fig.: 5. Ref.: 4 items.

УДК 004.932



СТРУКТУРНОЕ ОПИСАНИЕ ИЗОБРАЖЕНИЙ В ВИДЕ ГРАФА ДЛЯ РЕШЕНИЯ ЗАДАЧ РАСПОЗНАВАНИЯ

А. В. Агарков

Институт проблем искусственного интеллекта, г. Донецк, Украина
aav@iai.donetsk.ua

Предложен способ построения структурного описания изображения для распознавания и выделения объектов. Данное описание строится на основе анализа мультимасштабного представления, построенного на основе свертки с функцией Гаусса. Предложен новый критерий для выделения структурных элементов, на основе которых производится построение описания изображения. Данное описание отражает структуру значимых деталей изображения, соответствующих областям с однородной яркостью, и позволяет восстанавливать исходное изображение и отдельные его области.

МУЛЬТИМАСШТАБНОЕ ПРЕДСТАВЛЕНИЕ, ОПИСАНИЕ ИЗОБРАЖЕНИЙ, РАСПОЗНАВАНИЕ ИЗОБРАЖЕНИЙ, ОБНАРУЖЕНИЕ ОБЪЕКТОВ, ГРАФЫ

Введение

В последние годы появилось значительное количество работ, посвящённых распознаванию изображений и обнаружению объектов, проводимых инвариантно относительно их масштаба и поворотов. Данные методы основаны на использовании структурного описания изображений и объектов их составляющих. Можно выделить три направления в этой области – (1) использование контурных представлений, (2) использование скелетных представлений для описания бинарных и силуэтных изображений, (3) использования локальных черт, инвариантных к масштабу и поворотом.

Методы, использующие контуры объектов [1] для описания их структуры, посвящены способам описания и сравнения кривых, образующих контуры. Сложность применения данного метода связана с проблемами, возникающими при выделении контуров объектов.

Подход, использующий скелетное представление [2,3], основан на выделении скелета бинарного объекта или силуэта и последующим описанием его в виде графа, каждой вершине которого соответствует участок скелета, имеющий определённые свойства. Данные участки выступают в качестве структурных элементов описания объекта. Выделение структурных элементов и описание их свойств производится таким образом, что описание объектов, получаемое в результате, оказывается инвариантным относительно масштаба и поворотов. Распознавание изображений сводится к сравнению графа, который описывает исследуемый объект, с графом-шаблоном, который представляет объекты одного класса.

Данный подход используется также при распознавании полутоновых и цветных изображений. Для этого вначале выделяются границы, затем на основе использования шоковых линий (shock graphs) проводится устранение дефектов и выделение скелета. После чего применяется методика, используемая для бинар-

ных и силуэтных изображений. Для бинарных изображений этот подход показывает хорошие результаты. Однако для полутоновых и цветных изображений его использование затруднено сложностями, связанными с выделением границ объектов, в результате которых объект может быть разбит на несколько сегментов, что затрудняет процедуру сравнения с шаблоном.

Для распознавания цветных и полутоновых изображений в настоящее время все шире используется подход, основанный на применении локальных черт, инвариантных относительно масштабу и поворотам [4–8]. Наиболее простым и эффективным методом для выделения таких черт в настоящий момент является использование мультимасштабного представления, построенного на основе свёртки с функцией Гаусса и её производными.

Использование мультимасштабного представления для анализа иерархической структуры изображений было начато в 90-х годах прошлого столетия в работах [9–13]. В этих же работах указывается, что применение функции Гаусса для построения мультимасштабного представления позволяет достичь инвариантности относительно масштаба и поворотов при анализе структуры изображений.

В настоящий момент достаточно сильно развит метод выделения локальных черт изображения, использующий мультимасштабное представление. Выделенные локальные черты используются для описания объектов. Обнаружение и распознавание объектов основано на сравнении описаний изображения и объекта, построенных на основе использования локальных черт. Основными направлениями развития данного подхода являются увеличение устойчивости выделения локальных черт и усовершенствование дескрипторов, которые описывают их яркостные свойства. При этом практически не затрагиваются вопросы, связанные с анализом и использованием иерархической структуры изображений, хотя именно с этого и начиналось. Оста-

лись в тени также вопросы, связанные с использованием других типов структурных элементов, которые можно выделить при помощи мультимасштабного представления. В частности, в ранних работах были предложены способы выделения хребтов изображений, которые соответствуют протяженным областям с однородной яркостью и которые по сути являются скелетами данных областей. Хотя следует заметить, что для выделения структурных элементов разной природы использовались мультимасштабные представления, построенные на основе различных ядер свёртки.

В настоящей работе вводятся критерии для выделения дополнительных структурных элементов, которые позволяют описывать структуру значимых деталей изображения, соответствующих однородным по яркости областям. На основе использования данных структурных элементов производится выделение и описание областей изображения с однородной яркостью. Показано, что на основе использования данных структурных элементов возможно построение скелетного представления однородных по яркости областей. Данные структурные элементы используются для построения описания изображений и объектов, их составляющих, которое предназначено для проведения распознавания. Данное описание позволяет также восстанавливать как исходное изображение, так и отдельные его части, в том числе соответствующие отдельным объектам.

1. Выделение локальных черт изображений с помощью мультимасштабного представления

Мультимасштабное представление изображения на основе применения функции Гаусса представляет собой упорядоченное множество изображений, полученных из исходного путём свёртки с ней:

$$L(x, y, \sigma_i) = G(x, y, \sigma_i) * I(x, y),$$

где $G(x, y, \sigma_i)$ – функция Гаусса с масштабом σ_i ; $\sigma_i = k\sigma_{i-1}$, $(*)$ – операция свертки; $I(x, y)$ – исходное изображение. Данное множество изображений образует в объединённом трехмерном пространстве плоскость изображения – масштаб сглаживания дискретную функцию $L(x, y, \sigma)$. Для увеличения скорости построения мультимасштабного представления используются октавы исходного изображения – при увеличении масштаба сглаживания в два раза во столько же раз уменьшается исходное. Соответственно, множитель k выбирается так, чтобы $k^{N_o} = 2$, где N_o – количество масштабов в одной октаве.

Однако для выделения локальных черт изображения используется мультимасштабное представление, построенное на основе свёртки с ядром, которое представляет собой разность функций Гаусса с различным масштабом сглаживания, известное как оператор DoG:

$$\begin{aligned} D(x, y, \sigma_i) &= (G(x, y, \sigma_i) - G(x, y, \sigma_{i-1})) * I(x, y) \\ &= L(x, y, \sigma_i) - L(x, y, \sigma_{i-1}). \end{aligned} \quad (1)$$

Данное представление также образует дискретную функцию $D(x, y, \sigma)$.

Как видно из формулы (1), функция $D(x, y, \sigma)$ получается из $L(x, y, \sigma)$ путём вычитания соседних слов мультимасштабного представления. Это обстоятельство будет использовано в дальнейшем при построении описания изображения.

Локальные черты изображения выделяются из условия достижения экстремума функции $D(x, y, \sigma)$. При этом координаты (x, y) определяют центр области, соответствующей локальной черте, а значение σ – её характерный масштаб. Локальные черты, выделяемые с помощью данного оператора DoG, соответствуют пятнообразным областям (blobs) однородной яркости.

Множество выделенных локальных черт $E = \{E_j(x_j, y_j, \sigma_j)\}$ составляет основу для описания изображения и объекта, который необходимо распознать.

2. Выделение структурных элементов изображения

Описание изображения, построенное на основе множества $E = \{E_j(x_j, y_j, \sigma_j)\}$, позволяет распознавать объекты инвариантно их масштаба и поворота. Однако данное описание не в полной мере отражает структуру изображения – не используется иерархия локальных черт, определяющая уровни детализации изображения. Также не в полной мере отражена сегментная структура изображения – разбиение изображения на области однородной яркости. Отдельные локальные черты могут соответствовать частям одной области. Для отражения структуры изображения и объектов их составляющих, которые в основном состоят из соседних областей однородной яркости, это представляется достаточно важным. К тому же, использование более подробной информации о структуре изображения и его отдельных областей позволит отказаться от использования достаточно громоздких дескрипторов для описания отдельных локальных черт.

Для того чтобы более подробно отразить структуру изображения, необходимо в описание добавить новые структурные элементы, которые бы с одной стороны, были связаны с уже имеющимся множеством $\{E_j(x_j, y_j, \sigma_j)\}$, а с другой – имели бы самостоятельную ценность для отражения структурных свойств.

В развитие подхода, предложенного автором в работах [14, 15], в качестве дополнительных структурных элементов предлагается использовать области изображения, которые соответствуют экстремумам по направлению в плоскостях мультимасштабного представления $D(x, y, \sigma_i)$. То есть использовать экстремум по направлению функций $D(x, y, \sigma = const)$. Данные экстремумы образуют хребты и тальвеги данных функций (далее характерные линии), которые, в свою очередь, соответствуют протяженным деталям изображения. Как следует из свойств оператора DoG такими

детальями могут быть либо границы, либо протяженные области однородной яркости. Причём характер детали можно определить из поведения характерной линии с изменением масштаба σ . Если положение линии изменяется, то она соответствует границе, если положение линии остаётся стабильным – протяжённой области с однородной яркостью.

На рис. 2–4 представлены результаты выделения характерных линий при использовании масштабов сглаживания $\sigma=3,8$ (рис. 2), $\sigma=11,2$ (рис. 3) и $\sigma=33,98$ (рис. 4). Характерные линии выделены белым цветом, их толщина соответствует половине масштаба сглаживания.

Предлагается следующая процедура для выделения дополнительных структурных элементов. В каждой плоскости мультимасштабного представления $D(x, y, \sigma_i)$ производится поиск множества экстремумов по направлению $T = \{p_j = p(x_j, y_j, \sigma_{i,j})\}$ функции $D(x, y, \sigma = const)$. Данное множество разбивается на непересекающиеся подмножества:

$$\bar{T} = \{t_j : \bigcup_j t_j = T; \forall j, k, j \neq k, t_j \cap t_k = \emptyset\},$$

каждое из которых состоит из экстремумов, имеющих одинаковые координаты в плоскости изображения и расположенные в последовательности из соседних плоскостей:

$$t = \{p(x_j, y_j, \sigma_{i,j}) : \forall j, k, j \neq k, (x_j, y_j) = (x_k, y_k); \{\sigma_{i,j} : i = i_1, \dots, i_{N_t}, i_{m+1} = i_m + 1\}\},$$

где N_t – количество экстремумов p в множестве t .

Множество \bar{T} разделяется на два непересекающихся подмножества \bar{T}_{\max} и \bar{T}_{\min} , которые образованы локальными максимумами и минимумами соответственно.

Каждое множество $t_i \subset T$ образует дополнительный структурный элемент. Данный структурный элемент характеризуется координатами в плоскости изображения (x_p, y_p) масштабом $\sigma_{extr} = \arg \min_{p(x, y, \sigma)} \max_{p(x, y, \sigma) \in t_i} (D(x, y, \sigma))$



Рис.1. Исходное изображение

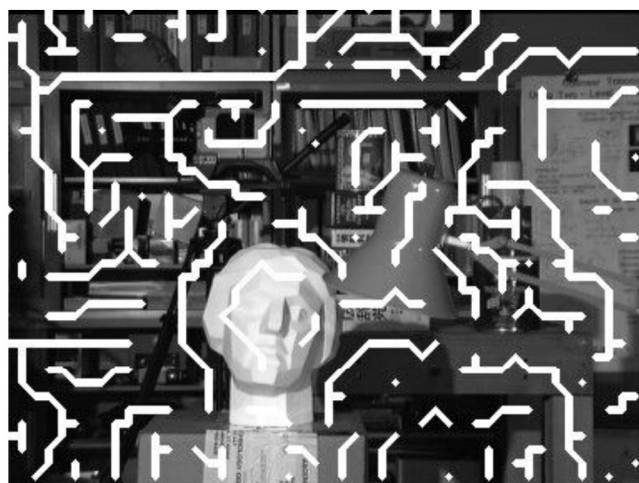


Рис. 3. Характерные линии, выделенные при использовании масштаба сглаживания $\sigma=11,2$



Рис.2. Характерные линии, выделенные при использовании масштаба сглаживания $\sigma=3,8$

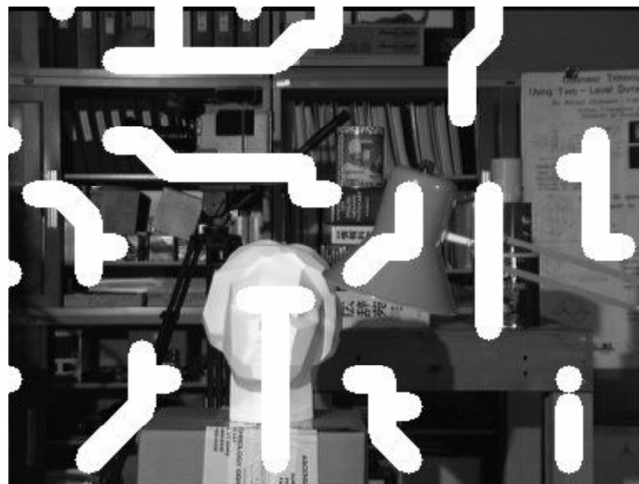


Рис. 4. Характерные линии, выделенные при использовании масштаба сглаживания $\sigma=33,98$

(далее – характерный масштаб). Причём следует учитывать, какого типа экстремумы составляют множество t_i – максимумы или минимумы. Если минимумы, то значение масштаба должно соответствовать минимальному значению $D(x, y, \sigma)$, если максимумы – максимальному. Также данный структурный элемент характеризуется значением $L(x, y, \sigma_{extr})$ – характерной яркостью.

Следует заметить, что $E \subset T$, поскольку экстремум функции $D(x, y, \sigma)$ также является экстремумом по направлению для $D(x, y, \sigma = const)$. Это позволяет описывать локальные черты при помощи дополнительных структурных элементов, расположенных в окрестности их центров.

Поскольку в каждой плоскости мультимасштабного представления выделенные экстремумы по направлению образуют характерные линии, то логично структурные элементы объединять так, чтобы они образовывали линии в плоскости изображения.

В соответствии с тем, к какому из подмножеств \bar{T}_{max} и \bar{T}_{min} относятся структурные элементы, разделяются на два типа детали изображения, которым соответствуют структурные элементы.

3. Построение описания изображения

Описание изображения строится на основе применения графа. Пусть граф $G_I^T(V^T, E^T)$ является описанием изображения $I(x, y)$. Каждой вершине данного графа ставится в соответствие структурный элемент, соответствующий множеству t_i . Причём каждой вершине соответствует только один структурный элемент. Ребрами соединяем вершины, если соответствующие им структурные элементы являются соседями.

Соседство разделяется на четыре типа: (1) соседние структурные элементы, относящиеся к одной детали изображения, у которых центры являются соседними пикселями; (2) ближайшие структурные элементы, относящиеся к соседним деталям изображения одного типа; (3) ближайшие структурные элементы, относящиеся к соседним деталям изображения разного типа; (4) структурные элементы, центры которых расположены рядом, и которые преобразуются друг в друга при изменении масштаба сглаживания.

Ребра, отражающие соседство разных типов, маркируются по-разному. Атрибуты вершин отражают свойства структурных элементов – координаты в плоскости изображения, характерные масштаб и яркость. Однако атрибут вершины также может содержать более подробную информацию о множестве t_i , которое соответствует данной вершине. Здесь отражены только те свойства, которые используются в данной работе.

Для того чтобы установить отношения соседства между структурными элементами, используется модифицированное дистанционное преобразование. Данная модификация позволяет определять для каждого пикселя изображения не только расстояние до ближайше-

го объекта, но и индексы этого объекта и его ближайшего структурного элемента (в частности – пикселя).

Как известно [16], дистанционное преобразование позволяет приближённо определять расстояние до объектов на изображении для любых пикселей. Расстояние определяется как длина минимального пути, построенного из элементарных перемещений (по горизонтали, вертикали и диагонали) до объекта.

Строится вспомогательное $F(x, y)$ изображение с такими же размерами, что и у исходного. Каждому пикселю, принадлежащему объекту, присваиваем значение ноль, остальным – бесконечность (на практике – просто очень большое значение).

$$F(x, y) = \begin{cases} 0, & (x, y) \in O \\ \infty, & (x, y) \notin O \end{cases}$$

Для проведения дистанционного преобразования используются различные маски. Рассмотрим наиболее простой случай маски 3×3 , которая соответствует элементарным перемещениям на один пиксель по горизонтали, вертикали и диагонали. Длина перемещения по диагонали приближённо принимается 1,4. Поскольку операции с целыми числами совершаются быстрее, чем с дробными, то длина вертикального и горизонтального перемещений принимается 5, а диагонального – соответственно 7. Таким образом, для дистанционного преобразования используется маска

$$\begin{bmatrix} 7 & 5 & 7 \\ 5 & 0 & 5 \\ 7 & 5 & 7 \end{bmatrix}.$$

Затем, начиная от левого края изображения, последовательно для всех колонок выполняется следующая процедура. К значению изображения $D(x, y)$ в пикселе, принадлежащем текущей колонке, прибавляется длина элементарного перемещения, и полученный результат сравнивается со значением в соседнем пикселе, соответствующем перемещению. Если результат суммы меньше, чем значение в соседнем пикселе, то данное значение меняется на полученную сумму. Данная процедура выполняется для всех пикселей колонки.

Поскольку данная процедура выполняется для всех колонок последовательно от левого края к правому, то достаточно рассматривать только перемещения в направлении движения, то есть используется только третий столбец маски.

Затем аналогичные процедуры выполняются от правой стороны изображения к левой, сверху вниз и снизу вверх. В результате каждое значение $F(x, y)$ соответствует примерному расстоянию до ближайшей точки объекта.

Если на изображении несколько объектов, и необходимо помимо расстояния знать и какой из объектов находится ближе всех к данной точке и (или) какая точка объекта, то в вышеописанный алгоритм вносится небольшое дополнение.

Вводится ещё одно дополнительное поле $M(x, y)$. Каждое значение $M(x, y)$ – это индекс (или набор индексов), соответствующий определённому объекту и (или) пикселю объекта.

При начальном заполнении полей каждому значению M в точках (x, y) , принадлежащих объекту (объектам), присваивается значение, соответствующее индексам, идентифицирующим объект и (или) пиксель.

Если при проведении дистанционного преобразования значение $F(x, y)$ меняется, то меняется и $M(x, y)$ на значение, соответствующее текущему пикселю.

Для каждого слоя мультимасштабного представления производится дистанционное преобразование. В качестве выделенных объектов выступают отдельные пиксели, соответствующие экстремумам. Данные пиксели индексируются в соответствии с тем, какому структурному элементу они соответствуют. Таким образом, после проведения дистанционного преобразования каждому пикселю изображения соответствует индекс структурного элемента, к центру которого он находится ближе всего, и значение, соответствующее расстоянию до данного центра.

Чтобы определить, изменилось ли положение экстремума при изменении масштаба, достаточно проверить значение индекса детали пикселя с теми же координатами, что и у рассматриваемого экстремума, но в соседней плоскости мультимасштабного представления. Если индекс изменился – значит экстремум сместился и данный индекс соответствует ближайшей детали. Между данными структурными элементами устанавливается отношение соседства типа (4). Данное отношение отражает иерархию между структурными элементами – та, которая имеет больший характерный масштаб, имеет больший уровень иерархии.

Взаимное расположение структурных элементов определяется также на основе анализа результатов дистанционного преобразования. Для этого необходимо найти пиксели, в которых значения, соответствующие расстоянию до ближайшего экстремума, достигают максимума по направлению. То есть необходимо выделить «хребты» функции дистанционного преобразования $F(x, y)$. Индексы пикселей по разные стороны от данных «хребтов» соответствуют паре структурных элементов, расположенных наиболее близко друг от друга и имеющих тип соседства (2) или (3).

4. Описание отдельных областей

Как показано выше, областям изображения, обладающим однородной яркостью, соответствуют характерные линии, не меняющие своего расположения в плоскости изображения в достаточно большом диапазоне масштабов сглаживания. Таким образом, такой области соответствует связный подграф графа G_I^T , состоящий из вершин, соединённых с друг другом рёбрами типа (1). Поскольку данному подграфу фактиче-

ски соответствует характерная линия (вообще говоря, разветвляющаяся), то для её описания уместно применить аппарат скелетного представления, достаточно хорошо разработанный к настоящему времени.

Отношения соседства между областями изображения определяются из того, какие вершины соединены рёбрами типов (2) и (3). Иерархические отношения, отражающие детализацию областей, определяются рёбрами типа (4).

Таким образом, граф G_I^T позволяет для описания отдельных областей изображения использовать два подхода: первый – на основе скелетного представления, второй – на основе структуры, образуемой структурными элементами более низких уровней иерархии.

5. Восстановление изображения и его отдельных областей

Граф G_I^T позволяет довольно точно восстанавливать исходное изображение. Действительно, поскольку каждый структурный элемент, соответствующий вершинам данного графа, представляет собой область с однородной яркостью и центром в определённой точке плоскости изображения, то представляется возможным восстановить исходное изображение. Однако такое изображение будет лишено мелких деталей, размер которых меньше, чем начальный масштаб сглаживания.

Процедура восстановления проводится по следующему принципу – чем меньше характерный масштаб структурного элемента, тем больший приоритет он имеет при определении яркости конкретного пикселя.

Поскольку целью является исследование возможности восстановления исходного изображения по его описанию в виде графа G_I^T , то используется наиболее простой способ – в центре каждого структурного элемента рисуется окружность с радиусом, равным характерному масштабу и характерной яркостью. Если окружности перекрываются, то яркость общих пикселей определяется характерной яркостью структурного элемента меньшего масштаба. Если масштабы равны – значение яркости в пикселе усредняется.

На рис. 5–7 показаны результаты восстановления изображения (рис. 1), описанного графом G_I^T , построенном на основе мультимасштабного представления, содержащего семь октав по семь масштабов в каждой октаве. Размер исходного изображения 384×288 . Начальное значение масштаба сглаживания – $\sigma_0 = 1,4$. Отражены этапы восстановления, соответствующие минимальному значению характерных масштабов структурных элементов $\sigma = 22,4$ (рис. 5), $\sigma = 6,18$ (рис. 6) и конечный результат при $\sigma = 1,4$ (рис. 7).

Как видно из рис. 5–7, чем меньше минимальный характерный масштаб структурных элементов, используемых для восстановления изображения, тем выше уровень детализации. Как и ожидалось, восстановленное изображение лишено мелких деталей, размер ко-



Рис. 5. Восстановленное изображение, соответствующее минимальному характерному масштабу структурных элементов $\sigma = 22,4$

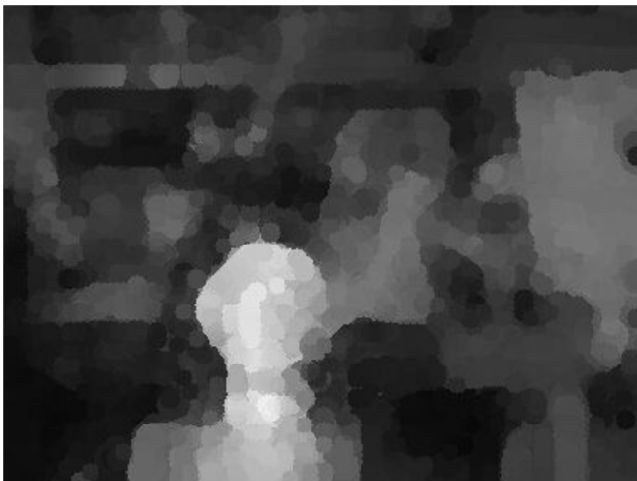


Рис. 6. Восстановленное изображение, соответствующее минимальному характерному масштабу структурных элементов $\sigma = 6,18$

торых меньше минимального масштаба сглаживания. Из-за этого оно выглядит несколько сглаженным по сравнению с исходным (рис. 1). Однако визуальная оценка показывает, что результат восстановления не сильно отличается от оригинала.

Таким же образом возможно и восстановление отдельных областей изображений, которым соответствуют подграфы графа G_I^T . При этом, как и в случае со всем изображением, возможно восстановление с различной степенью детализации, которая отражена в иерархических связях между отдельными структурными элементами и областями.

Это свойство позволяет выделять распознанный объект от фона без дополнительной обработки исходного выражения. Качество восстановленного изображения также отражает адекватность предложенного описания.



Рис. 7. Восстановленное изображение

Выводы

В работе предложен новый критерий для выделения структурных элементов на основе использования мультимасштабного представления. Данные структурные элементы соответствуют областям изображения, имеющим однородную яркость. Также предложен способ построения описания изображения, предназначенного для распознавания и выделения объектов. Данное описание основано на использовании структурных элементов, выделенных с помощью предложенного критерия. Данное описание представляет собой граф, вершины которого соответствуют отдельным структурным элементам, а рёбра отражают иерархические, пространственные и структурные отношения между ними.

Рассмотрены основные свойства данного графа, которые позволяют использовать для описания характерных областей изображения аппарат скелетного представления и структуру, определяемую структурными элементами нижних уровней иерархии и их отношениями. Также данный граф позволяет восстанавливать как всё изображение, так и его отдельные области. Данное свойство даёт возможность выделять объекты на фоне без использования дополнительной обработки исходного изображения.

Направление дальнейших исследований связано с более глубоким изучением свойств предложенного описания, которые позволяют эффективно проводить распознавание изображений.

Список литературы: 1. *Ozcanli, O.C. and Kimia, B.B.* Generic Object Recognition via Shock Patch Fragments // *BMVC.* – Warwick Print, 2007. – P.1030–1039. 2. *Kaleem Siddiqi.* Shock Graphs and Shape Matching // *International Journal of Computer Vision.* – 1999. – Vol. 35. – № 1. – P. 13–32. 3. *Tagare, H., Vos, F., Jaffe, C., Duncan, J.* Arrangement – a spatial relation between parts for evaluating similarity of tomographic section // *IEEE Trans. PAMI.* – 1995. – Vol. 17. – P.880–893. 4. *David G. Lowe.* Distinctive Image Features from Scale-Invariant Keypoints // *International Journal of Computer Vision.* – 2004. – Vol. 2, №60. –

P.91-110. **5.** *David G. Lowe.* Local Feature View Clustering for 3D Object Recognition // In Proc. CVPR. –Springer. –2001. –P.682–688. **6.** *David G. Lowe.* Local Feature View Clustering for 3D Object Recognition // In Proc. CVPR. –Springer. –2001. –P.682–688. **7.** *Hongli Deng, Wei Zhang, Eric Mortensen, Thomas Dietterich, Linda Shapiro.* Principal Curvature-Based Region Detector for Object Recognition // IEEE Conference on Computer Vision and Pattern Recognition – CVPR '07. –2007 –P.1–8. **8.** *Wei Zhang, Hongli Deng, Thomas G. Dietterich and Eric N. Mortensen.* A Hierarchical Object Recognition System Based on Multi-scale Principal Curvature Regions // 18th International Conference on Pattern Recognition. –2006. –Vol. 1 –P.778–782. **9.** *Arjan Kuijper, Luc Florack.* Logical Filtering in Scale Space // 12th Danish conference on pattern recognition and image analysis. –Kopenhagen. –2003. –P.4–11. **10.** *Joes Staal, Stiliyan Kalitzin, Bart ter Haar Romeny, and Max Viereger.* Detection of Critical Structures in Scale Space // Proceedings of the Second International Conference on Scale-Space Theories in Computer Vision. –1999. –P.105–116. **11.** *Kuijper A., Florack L.M.J., Viereger M.A.* Scale Space Hierarchy // Journal of Mathematical Imaging and Vision. –2003. –Vol. 18, № 2. –P.169–189. **12.** *Salden A., Florack L. and Haar B., Koenderink R., Viereger M.* Multi-Scale Analysis and Description of Image Structure // Nieuw Archief voor Wiskunde. –1992. –Vol. 10, № 3. –P.309–326. **13.** *Tony Lindeberg,* Scale-space: A framework for handling image structures at multiple scales // Proc. CERN School of Computing, Egmond aan Zee. –+The Netherlands. –1996. **14.** *Агарков А. В.* Ієрархічне представлення зображень для застосування при вирішенні широкого кола проблем // Оброблення сигналів і зображень та розпізнавання образів: Восьма всеукр. міжнар. конф. –Київ, 2006. –С. 119–122. **15.** *Агарков А. В.* Иєрархическое представление изображений с помощью графа // Искусственный интеллект. –2007. –№.1. –С.101–109. **16.** *Старовойтов В. В.* Локальные геометрические методы цифровой обработки и анализа изображений. –Минск: Институт технической кибернетики НАН Беларуси, 1997. –284 с. **17.** *Р. Дуда, П. Харт.* Распознавание образов и анализ сцен. –М.: Мир, 1976. –511 с.

Поступила в редколлегию 06.03.2009

УДК 004.932

Структурний опис зображень у вигляді графа для рішення задач розпізнавання / А. В. Агарков // Біоніка інтелекту: наук.-техн. журнал. –2009. –№1(70) . –С. 95–101.

Запропоновано спосіб побудови структурного опису зображення для розпізнавання і виділення об'єктів. Даний опис будується на основі аналізу мультимасштабного представлення, побудованого на основі згортки з функцією Гауса. Запропоновано новий критерій для виділення структурних елементів, на основі яких здійснюється побудова опису зображення. Даний опис відбиває структуру значущих деталей зображення, що відповідають областям з однорідною яскравістю, і дозволяє відновлювати вхідне зображення й окремі його області.

Л.: 7. Бібліогр.: 17 найм.

UDK 004.932

Structural Image Description For Solving Of Recognition Tasks / A.V. Agarkov // Bionics of Intelligence: Sci. Mag. –2009. –№1(70). –P. 95–101.

It is proposed new way to construct structural image description, which is based on Gauss multi-scale representation. It is proposed new criterion for extraction of unit extraction that use for image description construction. This image description reflects characteristic image features, which is brightness homogeneous regions.

Fig.: 7. Ref.: 17 items.

УДК 681.5; 004.272.26; 004.42



РАСПОЗНАВАНИЕ ИЗОБРАЖЕНИЙ АНТРОПОМОРФНОГО ОБЪЕКТА

Н. Г. Аксак¹, А. Ю. Тыхун¹, О. Ю. Барковская¹, А. С. Солдатов¹¹ ХНУРЭ, г. Харьков, Украина, axak@kture.kharkov.ua

Разработан параллельный алгоритм распознавания изображений антропоморфного объекта с помощью технологии MS-MPI. Предложенный параллельный алгоритм позволяет автоматизировать процесс определения правильности выполнения элементов спортивной гимнастики, а также повысить производительность программ и сократить время работы.

ПАРАЛЛЕЛЬНОЕ И РАСПРЕДЕЛЕННОЕ ПРОГРАММИРОВАНИЕ, АЛГОРИТМ, СКЕЛЕТ, ИЗОБРАЖЕНИЕ, СКЕЛЕТОНИЗАЦИЯ, БИНАРИЗАЦИЯ, РАСПОЗНАВАНИЕ, АНТРОПОМОРФНЫЙ ОБЪЕКТ

Введение

Одной из наиболее сложных задач, решаемых в различных областях (системы диагностики, тестирования, управления, аэрокосмические съемки, цифровая картография, медицина и так далее), является обработка изображений. Предметной областью предлагаемой работы являются элементы, выполняемые спортсменками-гимнастками во время упражнения. В спорте актуальной является разработка систем распознавания изображения спортсмена, например, в художественной гимнастике с целью выставления оценки за выполненный элемент. Такой подход к судейству соревнований предполагает облегчить работу судей, сделать процесс автоматизированным, избавиться от доли субъективности, которая присутствует при судействе.

При решении таких задач основную сложность представляет значительный объем входных данных. Как правило, при обработке изображения представляются в виде матриц пикселей, для более качественного изображения отводится матрица большего размера и, как следствие, растет время обработки.

Для удовлетворения минимальных требований пользователя современные методы обработки должны быть более производительными и интеллектуальными, добиться чего можно посредством использования высокопроизводительных вычислительных систем.

В связи с этим авторы поставили перед собой задачу разработать параллельный алгоритм распознавания изображений антропоморфного объекта с применением технологии MS-MPI.

1. Алгоритм распознавания изображений антропоморфного объекта

Алгоритм распознавания изображений антропоморфного объекта включает скелетонизацию бинарного изображения антропоморфного объекта и нейросетевое распознавание полученного скелета.

Скелетонизация бинарного изображения состоит из выделения угловых пикселей и их удаления. Скелетом изображения будем называть выделенный центральный остов изображения единичной толщины с сохранением

топологических и размерных (масштабных) характеристик исходного изображения. Программная реализация алгоритма скелетонизации состоит из четырех пунктов:

- 1) считывание изображения и распределение данных по процессорам;
- 2) параллельное выделение части угловых пикселей на каждом процессоре;
- 3) сбор данных на головном процессоре;
- 4) удаление выделенных угловых пикселей.

Программная реализация *нейросетевого распознавания полученного скелета* предполагает следующие этапы:

- 1) подача на вход нейронной сети предварительно обработанного изображения (скелета);
- 2) обучение нейронной сети;
 - 2.1) параллельное вычисление выходов нейронов первого, второго и третьего слоев;
 - 2.2) сбор данных о выходах нейронов на головном процессоре;
 - 2.3) параллельное вычисление и коррекция весовых коэффициентов первого и второго слоев;
 - 2.4) сбор откорректированных весовых коэффициентов на головном процессоре;
 - 2.5) для достижения необходимой величины ошибки повторить этапы 2.1–2.2;

Реализация алгоритма осуществлена на языке программирования C++ в среде разработки Microsoft Visual Studio 2005 с применением библиотеки MS-MPI под управлением операционной системы Windows Compute Cluster Server 2003. Распараллеливание алгоритма распознавания изображений антропоморфного объекта реализовано на кластерной системе с топологией «звезда», состоящей из 10 компьютеров (Intel Pentium D, 3000 МГц).

2. Алгоритм скелетонизации изображений антропоморфного объекта

Алгоритм распознавания изображений антропоморфного объекта предполагает обработку изображений людей (антропоморфные объекты), находящихся в статическом положении, то есть механические движения частей тела человека, а также влияние на него других объектов или предметов не учитываются.

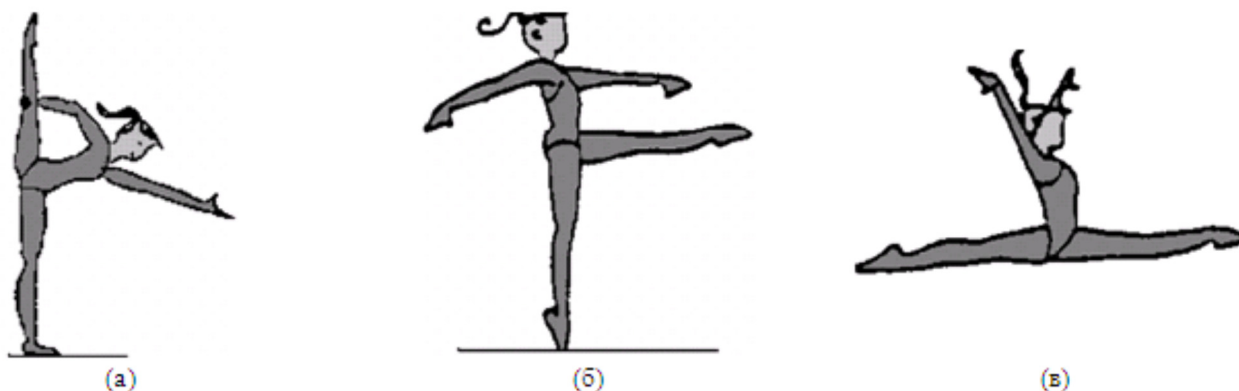


Рис. 1. Пример обрабатываемых элементов

Предметной областью являются элементы, выполняемые спортсменками-гимнастками во время упражнения (рис. 1).

При оценивании выступлений спортсменов в художественной гимнастике судьи принимают решение о получаемой оценке на основе правильности выполненного элемента. Примером может быть элемент, который называется «прыжок шагом» (рис. 1в). Элемент считается правильно выполненным, если гимнастка раскрывает ноги в воздухе в полный шпагат (угол между ногами составляет 180° или больше). В этом случае гимнастка получает 0,2 балла. Иначе, элемент считается не выполненным, и оценка - 0 баллов.

Антропоморфные объекты отличаются от неантропоморфных тем, что они испускают инфракрасное излучение, которое невидимо для человеческого глаза, но может быть обнаружено различными приёмниками теплового излучения и преобразовано в видимое изображение. Для перевода в видимую область спектра собственного теплового излучения людей используются тепловизоры. Видеоизображение, получаемое тепловизором, имеет бинарный вид – область интереса черная, фон – белый. Недостатком является то, что область интереса имеет размытый вид.

При использовании искусственных нейронных сетей исходные данные необходимо предварительно подготовить для вычисления более точного результата. Для этого выполняется скелетонизация – получение связного скелета бинарного изображения (выделение центрального остова изображения).

Выделение скелета бинарного изображения осуществляется посредством, во-первых, определения угловых пикселей и, во-вторых, – их удаления.

Обоснование необходимости выполнять скелетонизацию. Выделение центрального остова является необходимым шагом работы предлагаемого алгоритма, так как каждый антропоморфный объект (изображения гимнасток) имеет индивидуальные параметры тела – рост, вес, длина и толщина рук, длина и толщина ног, толщина корпуса, что вызывает дополнительные трудности при нейросетевом распознавании.

Исходное изображение преобразуется к бинарному виду или к изображению, содержащему два контрастных цвета одной цветовой шкалы. В случае черно-белого изображения фон является белым, а основное изображение (изображение переднего плана) черным (рис. 2а), что соответствует матричному виду, представленному на рис. 2б.

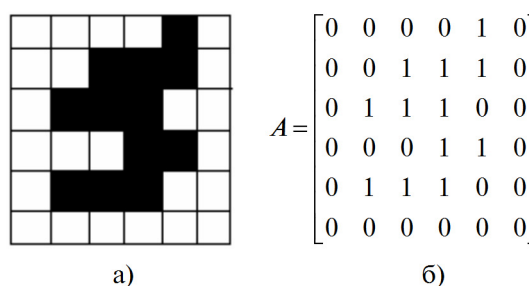


Рис. 2. Пример исходного изображения

Выделение угловых пикселей. Угловые пиксели выделяются согласно заданному правилу. Суть данного этапа заключается в нахождении «прямоугольных выпуклостей» на изображении и их выделении. Таким образом, формируется несвязный контур изображения.

В основе правила, согласно которому выделяются угловые пиксели, лежат структурные элементы (СЭ) для нахождения левого нижнего S_1 (рис. 3а), правого нижнего S_2 (рис. 3б), левого верхнего S_3 (рис. 3в) и правого верхнего S_4 угловых (рис.3 г) пикселей.

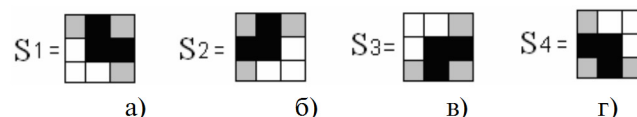


Рис. 3. Варианты расположения пикселей структурного элемента

СЭ представляют собой связную прямоугольную область пикселей. Основным результирующим пикселем является центральный пиксель прямоугольного окна. Поскольку обрабатывается изображение, состоящее из пикселей двух типов – фоновых белых «0» и основных черных «1», то подэлементами СЭ также являются 0 и 1.

СЭ представляется в виде матрицы, состоящей из девяти элементов:

$$S = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}.$$

Три угловых элемента матрицы S принимают значение равное 1, другие три обязательно равны нулю, а оставшиеся элементы, не влияющие на результат скелетонизации, могут принимать значение как «1», так и «0».

В матричном виде структурные элементы можно представить следующим образом:

$$S_1 = \begin{bmatrix} x_{11} & 0 & x_{13} \\ 1 & 0 & 0 \\ 1 & 1 & x_{33} \end{bmatrix}, S_2 = \begin{bmatrix} x_{11} & 0 & x_{13} \\ 0 & 0 & 1 \\ x_{31} & 1 & 1 \end{bmatrix},$$

$$S_3 = \begin{bmatrix} 1 & 1 & x_{13} \\ 1 & 0 & 0 \\ x_{13} & 0 & x_{33} \end{bmatrix}, S_4 = \begin{bmatrix} x_{11} & 1 & 1 \\ 0 & 0 & 1 \\ x_{13} & 0 & x_{33} \end{bmatrix}.$$

Угловым считается пиксель с координатами x_{22} .

Предположим, что исходное изображение имеет размер $n \times m$. Тогда задача выделения угловых пикселей требует для своего решения $4 \times (n - 2) \times (m - 2)$ операций сравнения матриц. С другой стороны, алгоритм выполнения сравнения матриц может быть рассмотрен как процесс $(n - 2) \times (m - 2)$ независимых подзадач сравнения элементов. Таким образом, первый этап выполнения скелетонизации ведет к использованию четырех процессоров. Процессор $p_i, i = \overline{1,4}$ отвечает за выделение угловых пикселей согласно p_i -му структурному элементу S_i (рис.3).

На этапе инициализации каждому процессору передаются исходное изображение и соответствующий СЭ.

Для каждой строки и каждого столбца выполняется сравнение расположения подэлементов СЭ S_i и обрабатываемого изображения:

– если подэлементы СЭ полностью покрывают пиксели фона и изображения, то граничные элементы СЭ становятся элементами фона – «0», а результирующий центральный пиксель становится элементом изображения – «1».

– иначе, если подэлементы СЭ не полностью покрывают пиксели соответствующего окна обрабатываемого изображения, то все пиксели данного окна изображения устанавливаются как фоновые – «0».

После выполнения описанных действий, получается несвязный контур из угловых пикселей.

Таким образом, на каждом процессоре образуется матрица, содержащая несвязный контур исходного изображения (элементы контура – единичные элемен-

ты). Результирующий несвязный контур исходного изображения получается путем сложения матриц, полученных на каждом процессоре (рис. 4а, 4б):

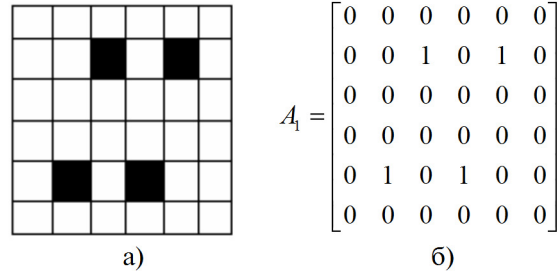


Рис. 4. Выделены угловые пиксели

Удаление выделенных угловых пикселей.

Полученный несвязный контур угловых пикселей исходного изображения вычитается из основного изображения (рис.5а,5б)

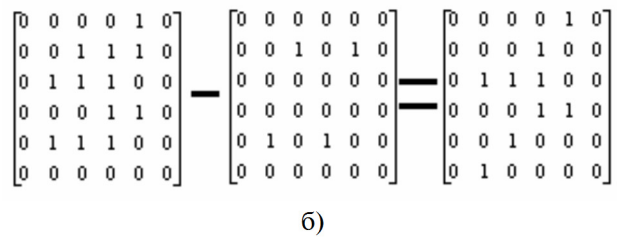
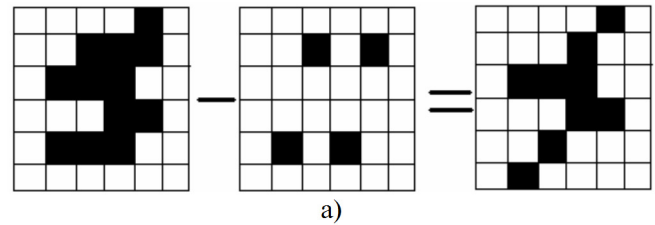


Рис.5. Вычитание несвязного контура:

- а) графическое представление;
- б) матричное представление

Все вышеописанные операции выполняются до тех пор, пока не будет достигнуто изображение толщиной в один пиксель.

Результатом работы предложенного алгоритма является неразрывное бинарное изображение единичной толщины – скелет: фон – белый, основное изображение (полученный скелет) – черное. Под неразрывностью понимается то, что каждый пиксель полученного скелета обязательно имеет один соседний пиксель [1].

3. Нейроалгоритм распознавания скелетонизированного изображения

Последним этапом является распознавание скелетонизированного бинарного изображения. Для реализации «расознавания с учителем» применяются искусственные нейронные сети, которые хорошо себя зарекомендовали при решении такого рода задач. Проблема распознавания в данной работе сведена к классификации образов к одному из двух классов («верно выполненное упражнение», «упражнение выполнено неверно»).

После выполнения скелетонизации на вход нейронной сети подается выделенный срединный остов изображения.

Решение задачи распознавания скелетонизированного бинарного изображения реализовано трехслойным персептроном, обучение которого осуществляется алгоритмом обратного распространения ошибки [2].

Функционирование трехслойной нейронной сети (НС) персептронного типа определяется следующим соотношением:

$$y^3 = f^3 \left(W^3 f^2 \left(W^2 f^1 \left(W^1 X + B^1 \right) + B^2 \right) + B^3 \right),$$

где для выходного, скрытого и входного слоев соответственно W^3, W^2, W^1 – матрицы весовых коэффициентов; B^3, B^2, B^1 – векторы пороговых значений; $f^3(*), f^2(*), f^1(*)$ – функции активации; X – вектор входных сигналов.

При анализе алгоритма обучения НС выявлен потенциальный параллелизм. Параллельное выполнение нейроалгоритма достигается декомпозицией на уровне реализации функций нейронов, при которой доли параллельного кода равномерно разделяются между выбранными процессорами, что повышает эффективность алгоритма.

В программной реализации используется топология процессоров типа master/slave (головной/подчиненный). Задачей головного процессора является послочное распределение нейронов по процессорам и окончательная сборка посредством обмена сообщениями библиотеки MPI. Подчиненные процессоры осуществляют параллельную настройку параметров нейронной сети и возвращают головному процессору полученные данные. В результате формируется выход каждого слоя.

Количество необходимых для распознавания итераций задается пользователем в зависимости от величин заданного параметра ошибки реакции сети и времени выполнения алгоритма.

В результате проведения экспериментов выявлено, что при ошибке $\delta = 0,02$ и количестве нейронов в скрытом слое, равном $N_j = 50$, точность классификации изображений составляет 92%.

Время работы последовательного нейроалгоритма прямо пропорционально разрешению обрабатываемого изображения.

На рис. 6 видно, что с увеличением разрешения изображения повышается эффективность параллельного алгоритма.

Выводы

Научная новизна работы заключается в разработке параллельного алгоритма распознавания изображений антропоморфного объекта с целью повышения производительности и сокращения времени обработки изображений.

Практическая ценность. Распараллеливание предложенного алгоритма на высокопроизводительные системы позволяет автоматизировать процесс распознавания статичных антропоморфных объектов. Отличительной чертой предложенного алгоритма является возможность распознавания зашумленных изображений антропоморфных объектов.

Список литературы: 1. Барковская О. Ю. Антропоморфные объекты для распознавания образов // XII Междунар. молодежный форум «Радиоэлектроника и молодежь в XXI в.». – Харьков, 1 – 3 апреля 2008. – С. 187. 2. Аксак Н. Г., Тыхун А. Ю. Вычислительная модель нейроалгоритма многослойного персептрона // Материалы Седьмой Междунар. конф.-семинара «Высокопроизводительные параллельные вычисления на кластерных системах». – Нижний Новгород, 2007. – С. 11–18.

Поступила в редколлегию 10.03.2009

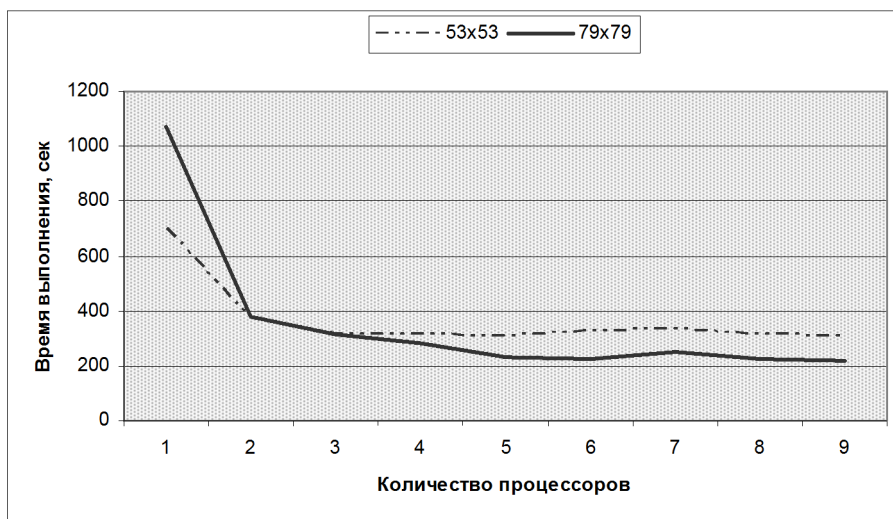


Рис. 6. Зависимость времени выполнения алгоритма от количества процессоров при обработке изображений разрешения 51×51 и 79×79

УДК 681.5; 004.272.26; 004.42

Розпізнавання зображень антропоморфного об'єкта / Н. Г. Аксак, А. Ю. Тихун, О. Ю. Барковська, А. С. Солдатов // Біоніка інтелекту: наук.-техн. журнал.—2009.—№1(70).—С. 102–106.

У статті описано паралельний алгоритм розпізнавання зображень антропоморфного об'єкта, що включає скелетонізацію бінарного зображення антропоморфного об'єкта та нейромережне розпізнавання отриманого скелета. Скелетонізація бінарного зображення складається з виділення кутових пікселів та їх вилучення. Програмна реалізація нейромережного розпізнавання припускає паралельну настройку параметрів нейронної мережі.

Л.: 6. Бібліогр.: 2 найм.

UDK 681.5; 004.272.26; 004.42

Recognition of anthropomorphic object images / N. G. Axak, A. U. Tihun, A. U. Barkovskaya, A. S. Soldatov // Bionics of Intelligence: Sci. Mag.—2009.—№1(70).—P. 102–106.

In the article is described parallel algorithm of anthropomorphic images recognition, including binary anthropomorphic image skeletonization and neuronet recognition of the received skeleton. The binary image skeletonization consists of angular pixels allocation and their removal. Program realization of neuronet recognition assumes parallel adjustment of neural network parameters.

Fig.: 6. Ref.: 2 items.

УДК 004.032.26:004.89



УПРОЩЕНИЕ СТРУКТУРЫ НЕЙРОСЕТЕЙ НА ОСНОВЕ ОСТРОВНОЙ МОДЕЛИ ЭВОЛЮЦИОННОГО ПОИСКА

А. А. Олейник¹, С. А. Субботин²¹ Запорожский национальный технический университет, г. Запорожье, Украина, olejnik@zntu.edu.ua;² Запорожский национальный технический университет, г. Запорожье, Украина, subbotin@zntu.edu.ua

Исследована проблема упрощения нейросетевых моделей. Разработан многокритериальный островной эволюционный метод, позволяющий организовать параллельный поиск по нескольким различным критериям. Проведены эксперименты по оптимизации нейромоделей режимов фрезерования нежестких деталей авиадвигателей.

НЕЙРОМОДЕЛЬ, ОПТИМИЗАЦИЯ, ЭВОЛЮЦИОННЫЙ ПОИСК

Введение

В настоящее время разработка автоматических систем распознавания образов, технической и биомедицинской диагностики, поддержки принятия решений остаётся достаточно сложной теоретической и технической проблемой, эффективность решения которой существенно зависит от качества классификации и идентификации исследуемых объектов, явлений, процессов, сигналов, что обуславливает необходимость разработки новых методов построения моделей зависимостей по конечным наборам данных, описывающих свойства и характеристики исследуемых объектов [1–2].

В качестве распознающих моделей эффективно могут использоваться искусственные нейронные сети прямого распространения [3–5], для построения которых необходимо предварительно определить структуру нейросети, учитывающую специфику решаемой задачи. Как правило, при синтезе нейромоделей структура сети задается пользователем, а сеть обучается путем оптимизационной настройки матрицы связей [6], вследствие чего эффективность данной нейромодели в значительной степени зависит от выбранной структуры.

Поскольку нейросети, имеющие небольшое количество нейронов из-за своих ограниченных аппроксимационных способностей не позволяют решать реальные практические задачи, число нейронов в сети выбирается с запасом, что приводит к увеличению времени вычисления значения целевого параметра по синтезированной нейромодели, увеличению объема памяти, необходимого для хранения значений весовых коэффициентов сети, снижению ее интерпретабельности, что вызвано избыточностью и неоптимальностью синтезированной нейромодели.

Поэтому целью настоящей работы является разработка метода оптимизации построенных нейромоделей, позволяющего сокращать количество нейроэлементов и связей между ними, синтезируя нейронные сети, обладающие приемлемыми аппроксимационными качествами.

1. Постановка задачи

Пусть задана обучающая выборка исходных данных в виде $\langle X, Y \rangle$, где $X = \{X_i\}$ – исходный набор значений признаков, характеризующих рассматриваемый объект или процесс; $Y = \{y_p\}$ – массив значений выходного параметра в выборке; $X_i = \{x_{ip}\}$ – i -ый признак в исходной выборке, $i = 1, 2, \dots, L$; L – общее количество признаков в исходном наборе; x_{ip} – значение i -го признака для p -го экземпляра выборки, $p = 1, 2, \dots, m$; y_p – значения прогнозируемого параметра для p -го экземпляра; m – количество экземпляров выборки.

Тогда задача оптимизации построенной нейросетевой модели вида $HC = HC(C, W, B, DF, TF)$ заключается в поиске таких новых значений $C' \subseteq C, W', B', DF', TF'$, при которых достигаются оптимальные значения заданных критериев оптимальности $\xi_1, \xi_2, \dots, \xi_K$, учитывающих основные характеристики нейромодели, где K – количество целевых критериев; C – матрица, определяющая наличие синаптических связей между элементами сети (рецепторами, нейронами); $W = W(C)$ – матрица весовых коэффициентов, соответствующих связям, присутствующим в сети HC ; $B = B(C)$ – вектор смещений нейронов сети; $DF = DF(C)$ – вектор дискриминантных функций нейроэлементов; $TF = TF(C)$ – вектор функций активации нейронов сети; $\xi(HC, X, Y)$ – критерий, определяющий эффективность использования нейросетевой модели HC для аппроксимации зависимости между набором входных параметров X и соответствующим ему вектором значений выходного параметра Y . Как правило, в качестве критерия оптимальности нейромодели используется среднеквадратическая ошибка [7].

В настоящее время существует два основных подхода к оптимизации структуры синтезированной нейросетевой модели [6]:

– методы удаления связей уменьшают количество элементов матрицы весов нейросети, упрощая таким образом ее структуру. При использовании таких методов удаляются связи, имеющие наименьшие значения синаптических весов или оказывающие наименьшее влияние на эффективность функционирования нейросетевой модели;

– методы удаления нейронов упрощают структуру нейромодели путем исключения из нее нейронов, удаление которых не приводит к значительному увеличению ошибки сети. Таким образом, при использовании методов удаления нейронов дополнительно необходимо оценивать их значимость путем вычисления ошибки упрощенной модели.

Однако существующие подходы упрощения нейромоделей, как правило, предполагают использование штрафных функций [6, 8], что во многих случаях не позволяет получать оптимальную структуру нейромодели, а иногда приводит к невозможности сходимости процесса упрощения нейромодели или к удалению значимых связей или нейронов.

Поэтому необходимо разработать новые методы упрощения нейросетевых моделей, способные выполнять поиск во всем пространстве возможных структур нейронных сетей прямого распространения.

Для оптимизации построенных нейромоделей эффективно могут использоваться методы эволюционного поиска [9–11], которые работают на каждой итерации с набором решений, что позволяет более детально исследовать пространство поиска, а также не накладывают ограничения на целевую функцию, в частности, не требуется дифференцируемость целевой функции, что позволяет производить поиск в пространстве нейронных сетей, содержащих нейроэлементы с недифференцируемыми функциями активации.

2. Оптимизация нейромоделей на основе эволюционного подхода

Известно [6, 12], что нейросетевые модели обладают различными характеристиками: качеством аппроксимации, точностью, интерпретабельностью, простотой аппаратной и программной реализации, вычислительной и пространственной сложностью и другими. Поэтому при оптимизации нейромодели целесообразным является использование набора критериев, наиболее полным образом её характеризующих.

При оценивании структурной сложности нейромодели целесообразно использовать количество синаптических соединений (весов) в сети N_w , количество нейронов N_n и количество слоев N_c в нейромодели. Предлагается использовать следующий критерий оценивания сложности структуры нейросети:

$$K_c = \left(1 + \frac{N_w}{N_n}\right) \left(1 + \frac{N_n}{N_c}\right) (1 + N_c), \quad N_n \geq 1, \quad N_c \geq 1.$$

Такой критерий учитывает количество слоев в нейросети, а также среднее количество нейронов в слое и среднее количество связей, приходящихся на один нейрон.

На вычислительную сложность сети существенное влияние оказывают функции активации нейроэлементов. Поэтому в качестве критерия оценивания вычислительной сложности предлагается использовать выра-

жение, учитывающее количество вычислительно сложных и вычислительно простых функций активации:

$$K_e = \left(\frac{1 + N_{сл.}}{1 + N_{пр.}}\right),$$

где $N_{сл.}$ – количество вычислительно сложных функций активации нейронов (логистическая сигмоидная, тангенциальная сигмоидная, радиальнобазисная и другие); $N_{пр.}$ – количество вычислительно простых функций активации (пороговая, линейная).

Для оценивания аппроксимационных качеств сети и сложности ее применения целесообразно использовать среднеквадратическую ошибку модели E и время расчета значения выходного параметра по синтезированной нейромодели t соответственно.

Предложенный набор критериев охватывает большинство характеристик нейросетей, что позволяет использовать его для многокритериальной оптимизации с целью получения достаточно простых и интерпретабельных нейромоделей, обеспечивающих высокую точность аппроксимации.

Поскольку оптимизация по нескольким критериям является эффективной при небольшом количестве целевых функций, предлагается для упрощения структуры нейромодели использовать два критерия, интегрирующие основные характеристики нейросетей:

$$f_1 = E \cdot K_c = E \left(1 + \frac{N_w}{N_n}\right) \left(1 + \frac{N_n}{N_c}\right) (1 + N_c),$$

$$f_2 = t \cdot K_e = t \left(\frac{1 + N_{сл.}}{1 + N_{пр.}}\right).$$

Первый критерий f_1 характеризует качество аппроксимации нейромодели и ее структурную сложность. Критерий f_2 объединяет время вычисления выходного параметра по синтезированной сети и ее вычислительную сложность.

С целью решения задачи оптимизации нейромоделей по нескольким заданным целевым критериям предлагается использовать многокритериальный островной эволюционный метод, выполнение которого состоит из двух этапов. На первом этапе выполняется разбиение популяции на подпопуляции (острова) и однокритериальный поиск оптимумов каждой из целевых функций, на втором – многокритериальный поиск в объединенной популяции хромосом.

Разработанный многокритериальный островной эволюционный метод предлагается выполнять в следующей последовательности шагов.

Шаг 1. Установить счетчик итераций: $t = 0$. Выполнить инициализацию начальной популяции P_0 хромосомами, содержащими информацию о значениях весовых коэффициентов и смещений нейроэлементов, а также о виде функции активации каждого нейрона сети, подлежащей упрощению.

Решения (хромосомы) при оптимизации нейромодели предлагается кодировать следующим образом

(рис. 1): в первой части хромосом хранить информацию о значениях весов нейронов входного слоя, во второй – значения весовых коэффициентов нейронов скрытых слоев, в третьей – значения смещений нейронов, в четвертой – функции активации каждого нейрона сети.

На рис. 1 используются следующие обозначения: $iw_{\nu\rho}$ – значение весового коэффициента связи от ν -го признака к ρ -му нейрону; $lw_{\nu\rho}$ – значение весового коэффициента связи от ν -го нейрона к ρ -му нейрону; b_ρ – значение смещения ρ -го нейрона; tf_ρ – функция активации ρ -го нейрона.

Шаг 2. Разбить популяцию P_t на K подпопуляций (островов) размером N/K каждая ($N/K \geq 2$), где K – количество целевых функций решаемой задачи.

Шаг 2.1. Для каждой хромосомы H_j рассчитать евклидово расстояние от нее до всех остальных хромосом в популяции. Евклидово расстояние d между хромосомами H_j и H_l вычисляется по формуле:

$$d(H_j; H_l) = \sqrt{\sum_{i=1}^L (h_{ij} - h_{il})^2},$$

где L – размер хромосом; h_{ij} и h_{il} – значения i -ых генов хромосом H_j и H_l , соответственно.

Шаг 2.2. Установить счетчик сформированных подпопуляций: $c = 1$. Сформировать множество не вошедших в подпопуляцию хромосом $A = P_r$.

Шаг 2.3. Инициализировать c -ую подпопуляцию: $V_c = \emptyset$.

Шаг 2.4. Выбрать из множества A две хромосомы H_j и H_l с максимальным расстоянием между ними.

Шаг 2.5. Включить выбранные хромосомы H_j и H_l в c -ую подпопуляцию: $V_c = V_c \cup \{H_j, H_l\}$. Исключить хромосомы H_j и H_l из A : $A = A \setminus \{H_j, H_l\}$.

Шаг 2.6. Если c -ая подпопуляция полностью сформирована ($|V_c| = N/K$), тогда перейти к шагу 2.9.

Шаг 2.7. Выбрать из множества A хромосому H_k , сумма расстояний от которой до хромосом из V_c является максимальной.

Шаг 2.8. Включить хромосому H_k в c -ую подпопуляцию: $V_c = V_c \cup \{H_k\}$. Исключить хромосому H_k из A : $A = A \setminus \{H_k\}$. Выполнить переход к шагу 2.6.

Шаг 2.9. Если сформированы все подпопуляции ($c = K$), тогда перейти к выполнению шага 3.

Шаг 2.10. Увеличить счетчик сформированных подпопуляций: $c = c + 1$. Выполнить переход к шагу 2.3.

Шаг 3. Увеличить счетчик итераций: $t = t + 1$.

Шаг 4. Выполнить основной цикл эволюционного поиска в каждой из K подпопуляций.

Шаг 4.1. Оценить приспособленность особей c -ой подпопуляции путем вычисления значений c -ой целевой функции $f_c(H_j)$, $j = 1, 2, \dots, N$.

Шаг 4.2. Выбрать часть хромосом для скрещивания и мутации.

Шаг 4.3. Выполнить оператор скрещивания, уменьшающий количество синаптических соединений в сети.

Для частей хромосом, содержащих информацию о значениях весовых коэффициентов нейронов входного и скрытых слоев, значение i -го гена потомков предлагается определять по формулам:

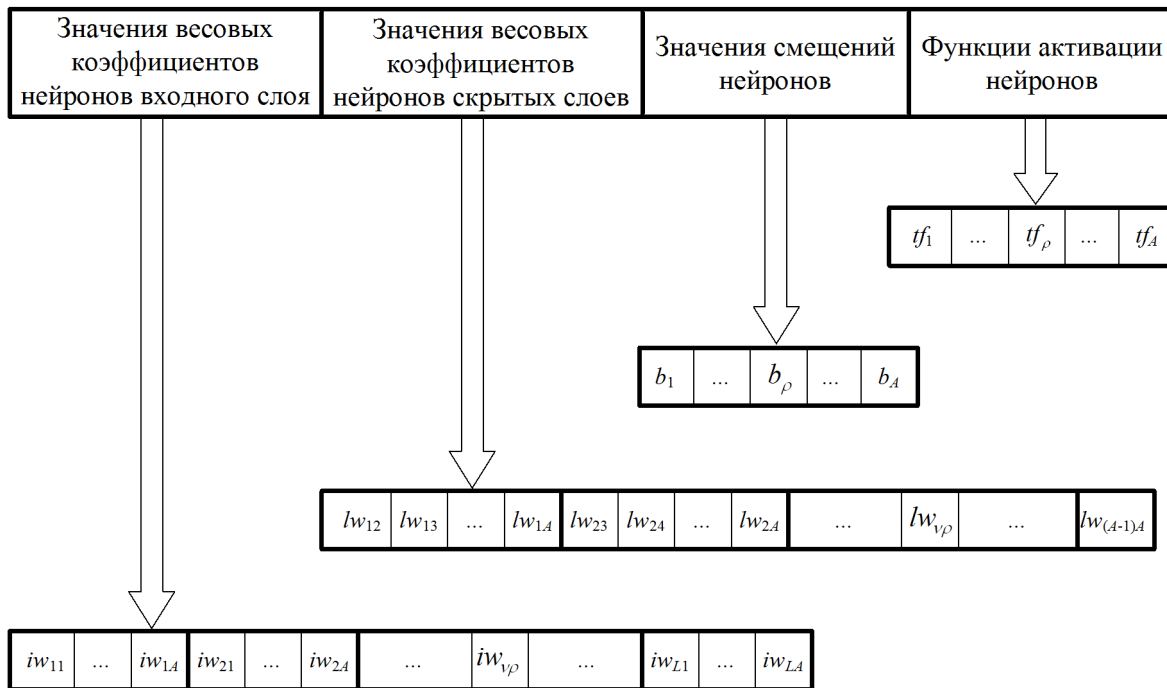


Рис. 1. Представление хромосомы при упрощении нейромоделей

$$h_{in1} = \begin{cases} 0, & \text{если } h_{i1} \cdot h_{i2} < 0; \\ kh_{i1} + (1-k)h_{i2}, & \text{в противном случае,} \end{cases}$$

$$h_{in2} = \begin{cases} 0, & \text{если } h_{i1} \cdot h_{i2} > 0; \\ (1-k)h_{i1} + kh_{i2}, & \text{в противном случае,} \end{cases}$$

где h_{in1} и h_{in2} – значения i -ых генов первого и второго потомков, соответственно; h_{i1} и h_{i2} – значения i -ых генов первого и второго родителей, соответственно; k – коэффициент, задаваемый пользователем, $k \in (0; 1)$.

Значения генов, соответствующих смещениям нейронов, определить по формулам:

$$h_{in1} = kh_{i1} + (1-k)h_{i2} \quad \text{и} \quad h_{in2} = kh_{i2} + (1-k)h_{i1}.$$

Значения генов, определяющих функции активации нейрона, предлагается определять по правилам:

$$h_{in1} = \begin{cases} h_{i1}, & \text{если } h_{i1} = h_{i2} \quad \text{или} \quad r > 0,5; \\ \text{rand[TF]}, & \text{в противном случае,} \end{cases}$$

$$h_{in2} = \begin{cases} h_{i2}, & \text{если } h_{i1} = h_{i2} \quad \text{или} \quad r \leq 0,5; \\ \text{rand[TF]}, & \text{в противном случае,} \end{cases}$$

где r – случайно сгенерированное число в интервале $(0; 1)$; rand[TF] – случайно выбранный элемент множества TF, содержащего информацию о функции активации, используемой для построения нейросети.

Шаг 4.4. Выполнить оператор точечной мутации.

В случае, если для мутации выбран ген хромосомы, содержащий информацию о значениях весовых коэффициентов нейронов сети, тогда новое значение i -го гена предлагается вычислять по формуле:

$$h_{in} = \begin{cases} 0, & \text{если } |r| < 0,5h_i; \\ r, & \text{в противном случае,} \end{cases}$$

где h_i и h_{in} – значения i -го гена до и после мутации соответственно; $r = \text{rand}[-h_i; h_i]$ – случайно сгенерированное число в интервале $[-h_i; h_i]$.

Если для мутации выбран ген хромосомы, соответствующий смещению нейрона, тогда значение i -го гена после мутации h_{in} предлагается определять по формуле:

$$h_{in} = \text{rand}[h_{i,\min}; h_{i,\max}],$$

где $h_{i,\min}$ и $h_{i,\max}$ – минимальное и максимальное значения i -го гена в текущей популяции хромосом.

В случае выбора для мутации гена хромосомы, содержащего информацию о функции активации нейрона, новое значение гена предлагается определять по формуле:

$$h_{in} = \text{rand[TF]}.$$

Шаг 4.5. Создать новое поколение c -ой подпопуляции из полученных на предыдущем шаге хромосом-потомков и наиболее приспособленных хромосом текущего поколения.

Шаг 5. Если $(\text{rem}(t, t_m) = 0)$, где $\text{rem}(a, b)$ – остаток от целочисленного деления a на b ; t_m – заданный пользователем интервал смены этапов эволюционной оптимизации, тогда перейти к шагу 6. В противном случае выполнить переход к шагу 3.

Шаг 6. Увеличить счетчик итераций: $t = t + 1$. Объединить хромосомы подпопуляций в единую популяцию P_t :

$$P_t = \bigcup_{c=1}^K V_c.$$

Шаг 7. Выполнить основной цикл эволюционного поиска над хромосомами из объединенной популяции. При этом оценивание хромосомы H_j проводить с помощью вычисления обобщенного значения целевой функции $F(H_j) = F(f_1(H_j), f_2(H_j), \dots, f_K(H_j))$, определяемого по правилу:

$$F(H_j) = \sum_{k=1}^K \frac{B_k}{\Delta_k},$$

где $B_k = \begin{cases} f_k(H_j) - \min(f_k), & \text{если } f_k \rightarrow \min; \\ \max(f_k) - f_k(H_j), & \text{если } f_k \rightarrow \max; \end{cases}$

$\Delta_k = \max(f_k) - \min(f_k)$ – диапазон изменения значений k -ой целевой функции; $\max(f_k)$ и $\min(f_k)$ – соответственно максимальное и минимальное значения k -ой целевой функции на текущей итерации. Таким образом, $F(H_j) \in [0; K]$.

При необходимости оценивание хромосом обобщенной популяции можно проводить с помощью подхода Парето, используя сортировку или ранжирование по принципу недоминируемости.

С целью уменьшения количества синаптических соединений в сети в качестве эволюционных операторов предлагается использовать операторы скрещивания и мутации, используемые на шагах 4.3 и 4.4, соответственно.

Шаг 8. Если $(\text{rem}(t, t_m) = 0)$, тогда перейти к шагу 9. В противном случае выполнить переход к шагу 7.

Шаг 9. Выполнить проверку критериев окончания поиска (достижение максимально допустимого количества итераций, приемлемого значения целевой функции). Если критерии останова удовлетворены, тогда выполнить переход к шагу 10. В противном случае перейти к выполнению шага 2.

Шаг 10. Останов.

Разработанный многокритериальный островной эволюционный метод, в отличие от классической островной модели эволюционного поиска [10], позволяет организовать параллельный поиск по нескольким различным критериям, вследствие чего он может применяться для решения многокритериальных задач. Кроме того, предложенный эволюционный метод упрощения нейромоделей позволяет ускорить поиск за счет уменьшения количества вычислений значений всех целевых функций на первом этапе эволюционной оптимизации, а также за счет использования специальных эволюционных операторов, уменьшающих количество синаптических соединений в сети, позволяет синтезировать нейронные сети, имеющие наиболее значимые нейроны и связи между нейроэлементами.

3. Эксперименты и результаты по синтезу моделей режимов фрезерования деталей авиадвигателей

С целью проверки эффективности применения разработанного многокритериального островного эволюционного метода проводились эксперименты по решению задачи синтеза моделей режимов фрезерования деталей авиадвигателей [13, 14].

При обработке нежестких деталей авиадвигателей применяют методы высокоскоростного фрезерования, которые, однако, сопровождаются вибрациями обрабатываемых деталей. Такие вибрации значительно повышают вероятность разрушения детали в процессе обработки. С целью уменьшения амплитуды вибраций предлагается синтезировать модели зависимости режимов высокоскоростного фрезерования от параметров вибраций. Затем на основе построенных моделей необходимо выбрать оптимальные параметры, обеспечивающие минимальный уровень вибраций.

Исходная выборка [13] характеризовалась значениями шести режимов фрезерования (y_1 – скорость резания, м/мин; y_2 – частота оборотов фрезы, об./мин; y_3 – подача на зуб, мм/зуб; y_4 – S – минутная подача, мм/мин; y_5 – глубина резания, мм; y_6 – вид фрезерования (попутное или встречное), а также получаемыми в результате обработки спектрами, которые численно представлены в виде отсчетов сигналов.

С целью сокращения описания исходных сигналов и представления их в форме, удобной для анализа зависимости режимов фрезерования от полученного сигнала, выполнялось преобразование Фурье [14] для каждого из полученных сигналов. В результате такого преобразования были получены 100 комплексных коэффициентов для каждого сигнала. Выделение наиболее информативной комбинации признаков выполнялось с помощью эволюционного метода с фиксацией части пространства поиска [15], в результате чего получили, что для синтеза математической модели y_1 было отобрано 11 признаков, y_2 – 14 признаков, y_3 – 13 признаков, y_4 – 12 признаков, y_5 – 12 признаков,

y_6 – 14 признаков.

На основе полученных входных данных выполнялся структурно-параметрический синтез нейромоделей [15, 16] зависимостей y_1 – y_6 .

Упрощение построенных нейромоделей выполнялось с помощью предложенного многокритериального островного эволюционного метода. Начальные значения параметров эволюционных методов устанавливались следующими: оператор отбора – отбор с использованием рулетки, оператор скрещивания – одноточечный, оператор мутации – гауссова мутация, количество особей в популяции $N = 100$ (для многокритериальных методов используются две подпопуляции по 50 особей каждая), вероятность скрещивания $p_{скр} = 0,8$, вероятность мутации $p_m = 0,05$, количество элитных особей $N_e = 2$. Критерии останова: максимально допустимое количество итераций $T = 100$, достижение приемлемого значения $\Pi\Phi f_n = 0,01$.

Сравнивались такие методы эволюционной оптимизации, как однокритериальный эволюционный поиск, использующий в качестве фитнес-функции f_1 , однокритериальный эволюционный поиск, использующий в качестве фитнес-функции f_2 , векторной генетический метод (Vector Evaluated Genetic Algorithm, VEGA), генетический метод с сортировкой по принципу недоминируемости (Nondominated Sorting Genetic Algorithm, NSGA), эволюционный метод с архивированием недоминируемых хромосом (Strength Pareto Evolutionary Algorithm, SPEA) [17] и многокритериальный островной эволюционный метод.

Результаты экспериментов по упрощению нейросетевой модели y_1 приведены в табл. 1, где t – время в секундах, затраченное на эволюционную оптимизацию; $E_{обуч}$ – ошибка модели по данным обучающей выборки; $E_{тест}$ – ошибка модели по тестовой выборке.

При упрощении моделей зависимостей других режимов фрезерования (y_2 – y_6) получены аналогичные результаты. Сравнение результатов, полученных с помощью применения однокритериального эволюцион-

Таблица 1

Результаты экспериментов по отбору признаков с помощью различных методов эволюционного поиска

Метод	τ	N_n	N_w	f_1	f_2	$E_{обуч}$	$E_{тест}$
Однокритериальный эволюционный поиск, использующий в качестве фитнес-функции f_1	218,2	8	39	0,78	0,097	0,0099	0,381
Однокритериальный эволюционный поиск, использующий в качестве фитнес-функции f_2	162,9	9	48	14,02	0,079	0,1108	0,1630
VEGA (лучшая модель, полученная для критерия f_1)	210,3	8	40	0,83	0,099	0,0098	0,2089
VEGA (лучшая модель, полученная для критерия f_2)	210,3	7	49	13,77	0,0087	0,0683	0,0936
Генетический метод с сортировкой по принципу недоминируемости	247,9	7	42	0,82	0,092	0,0099	0,2519
Эволюционный метод с архивированием недоминируемых хромосом	231,6	8	38	0,79	0,094	0,0098	0,2632
Многокритериальный островной эволюционный метод	182,7	7	37	0,78	0,095	0,0098	0,1931

ного поиска, VEGA, NSGA, SPEA и предложенного метода, показывает, что разработанный многокритериальный островной эволюционный метод позволяет более эффективно упрощать нейромодели по нескольким заданным критериям, а также сокращает время эволюционной оптимизации за счет меньшего количества вычислений значений всех целевых функций.

Выводы

С целью повышения интерпретируемости построенных нейросетевых моделей в работе решена задача упрощения их структуры на основе эволюционного подхода.

Научная новизна работы заключается в том, что разработан многокритериальный островной эволюционный метод, в отличие от классической островной модели эволюционного поиска, позволяет организовать параллельный поиск по нескольким различным критериям, вследствие чего он может применяться для решения многокритериальных задач. Кроме того, предложенный эволюционный метод упрощения нейромоделей позволяет ускорить поиск за счет уменьшения количества вычислений значений всех целевых функций на первом этапе эволюционной оптимизации, а также за счет использования специальных эволюционных операторов, уменьшающих количество синаптических соединений в сети, позволяет синтезировать нейронные сети, имеющие наиболее значимые нейроны и связи между нейроэлементами.

Практическая ценность результатов работы состоит в том, что решена задача упрощения нейромоделей режимов фрезерования нежестких деталей авиадвигателей.

Список литературы: 1. Интеллектуальные средства диагностики и прогнозирования надежности авиадвигателей: Монография / В. И. Дубровин, С. А. Субботин, А. В. Богуслаев, В. К. Яценко. – Запорожье: ОАО «Мотор-Сич», 2003. – 279 с. 2. Патрик Э. Основы теории распознавания образов. – М.: Сов. радио, 1980. – 408 с. 3. Минаев Ю. Н., Филимонова О. Ю., Лиес Б. Методы и алгоритмы идентификации и прогнозирования в условиях неопределенности в нейросетевом логическом базисе. – М.: Горячая линия-Телеком, 2003. – 205 с. 4. Хайкин С. Нейронные сети: полный курс. – СПб: Издательский дом «Вильямс», 2005. – 1104 с. 5. Круглов В. В., Борисов В. В. Искусственные нейронные сети: Теория и практика. – М.: Горячая линия-Телеком, 2001. – 382 с. 6. Руденко О. Г., Бодянский Е. В. Штучні нейронні мережі. – Харків: Компанія СМІТ, 2006. – 404 с. 7. Дли М. И. Нечеткая логика и искусственные нейронные сети. – М.: Физматлит, 2001. – 225 с. 8. Нейросетевые системы управления / В. А. Терехов, Д. В. Ефимов, И. Ю. Тюкин, В. Н. Антонов. – СПб.: Изд-во С.-Петербургского ун-та, 1999. – 265 с. 9. Эволюционные методы компьютерного моделирования: Монография / А. Ф. Верлань, В. Д. Дмитриенко, Н. И. Корсунов, В. А. Шорох. – К.: Наука думка, 1992. – 256 с. 10. Cantu-Paz E. Efficient and Accurate Parallel Genetic Algorithms. – Massachusetts: Kluwer Academic Publishers, 2001. – 162 p. 11. Haupt R., Haupt S. Practical Genetic Algorithms. – New Jersey: John Wiley & Sons, 2004. – 261 p. 12. Yao X. Evolving Artificial Neural Network // Proceedings of

the IEEE. – 1999. – № 9(87). – P. 1423–1447. 13. Качан А. Я., Внуков Ю. Н., Павленко Д. В. и др. Особенности колебаний деталей при высокоскоростном строчном фрезеровании // Вісник двигунобудування. – 2007. – № 1. – С. 69–76. 14. Олейник А. А., Павленко Д. В., Субботин С. А. Определение влияния режимов высокоскоростного фрезерования на параметры нежестких деталей на основе эволюционного подхода // Вісник двигунобудування. – 2008. – № 1. – С. 84–90. 15. Субботин С. О., Олейник А. О. Структурный синтез нейромоделей на основе полимодального эволюционного поиска // Радіоелектроніка. Інформатика. Управління. – 2008. – № 1. – С. 111–117. 16. Олейник А. А. Эволюционный метод обучения нейромоделей // Интеллектуальный анализ информации: Сборник трудов восьмой международной конференции ИАИ-2008 (14–17 мая 2008 г.). – К.: Просвіта, 2008. – С. 351–361. 17. Coello C. A Short Tutorial on Evolutionary Multiobjective Optimization // Evolutionary Multi-Criterion Optimization: Proceeding of the International Conference EMO2001 (7–9 March 2001). – Zurich: Springer-Verlag, 2001. – P. 21–40.

Поступила в редколлегию 12.03.2009

УДК 004.032.26:004.89

Спрощення структури нейромереж на основі островної моделі еволюційного пошуку / А. О. Олейник, С. О. Субботін // Біоніка інтелекту: наук.-техн. журнал. – 2009. – №1(70). – С. 107–112.

У статті досліджено проблему спрощення нейромережних моделей. Розроблено багатокритеріальний островний еволюційний метод, що дозволяє організувати паралельний пошук за декількома різними критеріями. Проведено експерименти з оптимізації нейромоделей режимів фрезерування нежорстких деталей авіадвигунів.

Табл.: 1. Іл.: 1. Бібліогр.: 17 найм.

UDC 004.032.26:004.89

Neural Network Structure Simplification Based on Island Model of Evolutionary Search / A.A. Oleynik, S.A. Subbotin // Bionics of Intelligence: Sci. Mag. – 2009. – №1(70). – P. 107–112.

The problem of neural network models simplification is investigated in the paper. The multiobjective island evolutionary method, allowing to organize a parallel search by several criteria is developed. The experiments on optimization of neural network models of airengine nonrigid details milling are conducted.

Tab.: 1. Fig.: 1. Ref.: 17 items.

УДК 004.415.2 – 519.246.8



ОБОБЩЕННЫЙ КРИТЕРИЙ ЭФФЕКТИВНОСТИ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

А. Н. Одейчук

Национальный научный центр «Харьковский физико-технический институт», г. Харьков, Украина
anodeychuk@kipt.kharkov.ua

Сформулирована постановка задачи для поиска критерия эффективности моделей временных рядов. Проведен сравнительный анализ существующих критериев эффективности, представленных в литературе. Построен обобщенный критерий эффективности, который был апробирован в условиях, когда традиционные критерии эффективности моделей прогнозирования не позволяют принять решения, и показал высокую результативность его использования.

КРИТЕРИЙ ЭФФЕКТИВНОСТИ, ПРИНЯТИЕ РЕШЕНИЙ, МОДЕЛИ ПРОГНОЗИРОВАНИЯ,
ИНФОРМАЦИОННАЯ СИСТЕМА

Введение

В условиях финансовой нестабильности и неопределенности на валютном и фондовом рынках возрастает необходимость в максимально точных прогнозах на краткосрочную и среднесрочную перспективу не только курсов валют или акций, но также и фундаментальных факторов развития экономик отдельных стран.

Стремительный рост и внедрение новых современных технологий в различных отраслях производства формируют острую потребность в своевременных прогнозах ключевых технических показателей производства (расчет некоторых параметров в ядерно-техническом комплексе и других).

Построение моделей прогнозирования технико-экономических показателей, представляемых в виде временных рядов, является сложной и наукоемкой задачей прикладного характера. С одной стороны, реальные процессы характеризуются нестационарностью, в результате чего требуется постоянно производить адаптацию моделей к новым тенденциям. Это служит причиной необходимости в многообразии перепределении структуры и параметров модели прогнозирования. В этой связи, разработка информационной системы, которая позволит автоматизировать процесс построения моделей прогнозирования, является необходимой и своевременной, о чем свидетельствуют многочисленные публикации в литературе [1–4, 23]. С другой стороны, задача построения моделей прогнозирования усложняется тем фактом, что для моделирования одного и того же временного ряда при помощи различных методологий и подходов может быть построено множество моделей, которые будут различаться по своей структуре и точности прогноза. Например на основе только одного класса моделей ARIMA или GARCH для временного ряда в общем случае может быть построено бесконечное множество различного вида моделей путем варьирования количества и значений параметров. Данный факт обуславливает необходимость в использовании для выбора

наиболее адекватной и точной модели прогнозирования одного из критериев, предложенных различными авторами [5–18]. Однако вопрос, связанный с тем, какой именно из них необходимо применять, освещен в литературе, на наш взгляд, недостаточно полно.

1. Состояние проблемы

В работах [1–4] рассматривается вопрос, связанный с проектированием информационной системы, реализующей разработку моделей прогнозирования временных рядов. Несмотря на масштаб проведенных исследований, задача определения критерия эффективности, который позволит принять решение о наиболее адекватной модели прогнозирования из множества конкурирующих между собой моделей, не нашла четкого отражения в общей постановке задачи проектирования системы и, вследствие чего, не была решена.

В этой связи является актуальным проведение исследования, направленного на решение задачи определения обобщенного критерия эффективности, позволяющего выделить из множества конкурирующих между собой моделей прогнозирования временных рядов модель, имеющую минимальную среднеквадратическую ошибку прогноза и оптимальную структуру. Это позволит восполнить пробел в проведенных исследованиях по проектированию информационной системы построения моделей прогнозирования, что и является целью данной статьи.

2. Постановка задачи исследования

Под эффективностью модели прогнозирования будем понимать набор количественных показателей, описывающих наиболее адекватную и оптимальную по своей структуре модель.

Постановка задачи исследования формулируется следующим образом.

Дано множество моделей прогнозирования Z . Каждая из моделей z :

– построена на основе методологии C^i , предпосылкам и предположениям которой удовлетворяет мно-

жество статистических характеристик H_D исследуемого временного ряда D ;

- имеет структуру S^{ij} ;
- оценки параметров определяются методом M^{ijl} .

Также дано множество критериев эффективности L , которые учитывают различные качественные показатели моделей прогнозирования и имеют некоторые ограничения.

Требуется определить критерий эффективности L^* , позволяющий из множества конкурирующих между собой моделей прогнозирования временных рядов выбрать модель, имеющую максимальную эффективность. Представим решаемую задачу следующим образом:

$$Z \xrightarrow{L^* \in L} Z^* (C^i \in C, S^{ij} \in S, M^{ijl} \in M, H_D), \quad (1)$$

где Z – множество моделей прогнозирования;

Z^* – модель прогнозирования, имеющая максимальную эффективность;

C – множество методологий построения моделей прогнозирования (множество классов моделей прогнозирования);

S – множество видов структур моделей прогнозирования;

M – множество методов определения оценок параметров моделей.

Основными требованиями к искомому критерию эффективности является учет следующих показателей качества:

- E_1 – среднее квадратическое отклонение;
- E_2 – количество параметров модели (сложность структуры модели);
- E_3 – ошибка выбора структуры модели;
- E_4 – смещенность прогноза;
- E_5 – автокорреляция остатков.

3. Анализ критериев эффективности модели прогнозирования

Существует достаточно большое количество критериев, применяемых для определения качества построенных моделей прогнозирования, их адекватности и точности.

Проведем краткий анализ возможностей использования этих показателей в качестве оценки эффективности модели.

При анализе критериев использованы следующие обозначения: Y_t – значение временного ряда в момент

t , \hat{Y}_t – прогноз значения Y_t ; $\bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t$ – среднее Y ; $e_t = Y_t - \hat{Y}_t$ – погрешность или ошибка прогноза; n – длина временного ряда.

Наиболее часто встречаемые в литературе критерии эффективности приведены ниже.

Среднее абсолютное отклонение (Mean Absolute Deviation, MAD) измеряет точность прогноза, усред-

няя величину ошибки прогноза (абсолютные значения каждой ошибки) [5]. Использование MAD наиболее полезно в тех случаях, когда аналитику необходимо измерить ошибку прогноза в тех же единицах, что и исходный ряд. Ошибка MAD вычисляется с помощью следующего уравнения:

$$MAD = \frac{1}{n} \sum_{t=1}^n |e_t|. \quad (2)$$

Однако недостатком данного критерия является нечувствительность к ошибкам, малым по абсолютной величине.

Среднеквадратическая ошибка (Mean Squared Error, MSE) и сумма квадратов ошибки (Sum Squared Error, SSE) рассчитываются по уравнениям, приведенным ниже:

$$MSE = \frac{1}{n} \sum_{t=1}^n e_t^2. \quad (3)$$

$$SSE = \sum_{t=1}^n e_t^2. \quad (4)$$

Данные критерии наиболее часто используются при выборе оптимальной модели прогнозирования в пакетах прикладных программ по прогнозированию [6]. Каждое значение ошибки прогноза в критериях (3) и (4) возводится в квадрат, что подчеркивает большие ошибки прогноза, в отличие от критерия (2). Данная особенность очень важна, так как модель прогнозирования, постоянно дающая средние по величине ошибки, в некоторых случаях может быть предпочтительнее другой модели, имеющей малые ошибки, но периодически дающей неприемлемые по абсолютной величине выбросы.

В тех случаях, когда проводится сравнение качества моделей прогнозирования разнородных объектов, оценка ошибок в абсолютных величинах не приемлема или вызывает трудности с интерпретацией получаемых результатов. В этой связи предпочтительным является переход к оценке ошибок в процентах [6,7].

С этой целью используют среднюю абсолютную ошибку в процентах (Mean Absolute Percentage Error, MAPE), вычисляемую по уравнению:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|e_t|}{Y_t} \cdot 100. \quad (5)$$

Этот показатель характеризует насколько велики ошибки прогноза в сравнении с действительными значениями ряда.

Более информативным критерием адекватности модели прогнозирования служит средняя процентная ошибка (Mean Percentage Error, MPE), определяющая смещенность (постоянное занижение или завышение) построенного прогноза [5,8]. Данный критерий рассчитывается по уравнению:

$$MPE = \frac{1}{n} \sum_{t=1}^n \frac{e_t}{Y_t} \cdot 100. \quad (6)$$

Значения, вычисляемые по уравнению (6), интерпретируются следующим образом. Если модель прогнозирования является несмещенной, критерий (6) будет иметь процентное значение, близкое к нулю и не превышающее 5 % [6]. Если в результате расчетов получается большое отрицательное процентное значение, то построенная модель является последовательно переоценивающей. Если получено большое положительное процентное значение, то – последовательно недооценивающей.

Заметим, что критерии MAPE и MPE не подходят для временных рядов с нулевыми членами. Однако в работе [6] рекомендуют в таких случаях пропускать нулевые элементы ряда и уменьшать значение n на единицу.

Для целей определения адекватной модели прогнозирования применяют коэффициент несовпадения Тейла [8,9], вычисляемый по уравнению:

$$U = \frac{\sqrt{\sum_{t=1}^n (\hat{Y}_t - Y_t)^2}}{\sqrt{\sum_{t=1}^n \hat{Y}_t^2 + \sum_{t=1}^n Y_t^2}}. \quad (7)$$

Если коэффициент принимает значение равно, нулю, то прогноз точный, а если значение, равно единице, то прогноз неадекватный.

Другим показателем адекватности модели может служить коэффициент детерминации (КД) [10, 11], который рассчитывается по уравнению:

$$R^2 = 1 - \frac{SSE}{\sum_{t=1}^n (Y_t - \bar{Y})^2}. \quad (8)$$

Чем ближе в выражении (8) значение КД к единице, тем лучше модель описывает реальные данные. На практике принято считать модель адекватной, если $KD > 0,75$ [1,10].

К недостаткам данного критерия, а также коэффициентов (2 – 8), можно отнести то, что их значения тем больше свидетельствуют об улучшении точности модели, чем больше новых параметров включается в модель, хотя это и не обязательно означает улучшение качества модели. Учесть данную особенность позволяет скорректированный коэффициент детерминации (СКД) [12,13], где r – количество параметров модели:

$$СКД = 1 - \frac{n-1}{n-r-1} (1 - R^2). \quad (9)$$

СКД построен таким образом, что добавление нового параметра в модель приводит к снижению его значения – штрафуются [18]. Эта особенность выгодно отличает СКД от критериев (2 – 8), которые фокусируют внимание исследователя на величине отклонения прогноза от реальных данных и не учитывают степень оптимальности структуры построенной модели.

Тем не менее, близкие к единице значения КД и СКД еще не свидетельствуют о хороших прогнозных качествах модели [1]: необходимо отсутствие автокорреляции ошибок, что может быть проверено с помощью статистики Дарбина-Уотсона. Однако она не применима, если в модели среди объясняющих переменных имеются лаговые значения результативного признака. В этой связи предлагается использовать Q-статистику Льюнга-Бокса (Ljung-Box Q-statistic, LBQ), имеющую распределение χ_m^2 и рассчитываемую по формуле [18]:

$$LBQ = n(n+2) \sum_{k=1}^m \frac{\rho_k^2}{n-k}, \quad (10)$$

где ρ_k – значение автокорреляционной функции. Если Q-статистика больше заданного квантиля распределения χ_m^2 , то признается наличие автокорреляции до m -го порядка в исследуемом ряду.

Также следует отметить, что критерий СКД может принимать значения больше единицы и отрицательные значения (пример приведен в (11)), что не дает возможности корректно интерпретировать данный критерий.

$$СКД = 1 - \frac{5-1}{5-6-1} (1-0,2) = 2,6, \quad (11)$$

$$СКД = 1 - \frac{10-1}{10-2-1} (1-0,2) = -0,0286.$$

На принципе штрафов за дополнительные параметры построены критерии Маллоуза, Акаике и Шварца [5,12,14,21].

Критерий Маллоуза:

$$C_p = \frac{SSE}{\sum_{i=1}^n (Y_i - \bar{Y})^2} + \frac{2(r+1)}{n-r-1}. \quad (12)$$

Информационный критерий Акаике:

$$AIC = \ln(MSE) + \frac{2(r+1)}{n}. \quad (13)$$

Байесовский информационный критерий (критерий Шварца):

$$BIC = \ln(MSE) + \frac{(r+1) \ln n}{n}. \quad (14)$$

Модель прогнозирования тем лучше, чем ниже показатель C_p , AIC , BIC . На данные критерии положительно влияет уменьшение остаточной дисперсии и отрицательно – количество включенных параметров. Основным различием между ними является степень жесткости, то есть, насколько велик штраф за большое количество параметров в модели.

Критерии C_p и AIC при больших n сильно коррелированы между собой. Однако C_p несколько жестче AIC , особенно при малых n . BIC является наиболее жестким критерием, причем, как можно увидеть из

приведенной формулы (14), в отличие от остальных критериев, его жесткость возрастает с ростом n .

Различие в жесткости проистекает из различия в предъявляемых требованиях к моделям прогнозирования. Критерии C_p и AIC направлены на достижение высокой точности прогноза: C_p направлен на минимизацию дисперсии ошибки прогноза, а AIC — на минимизацию расхождения между плотностью распределения по истинной модели и по выбранной модели. В основе критерия BIC лежит требование максимизации вероятности выбора истинной модели.

Кроме того, следует отметить существенный недостаток критерия Акаике, состоящий в переоценке истинного количества параметров в модели [16]. В свою очередь критерий Маллоуза применим только при нормальном распределении остатков [15].

При сравнении моделей прогнозирования важно учитывать, что если разные модели используют различные функциональные формы, то нельзя непосредственно сравнивать КД, СКД или значения критериев (2-8). Например, нельзя сравнить эти статистики для линейного и логарифмического варианта модели.

Значения, получаемые по логарифмической модели, значительно меньше соответствующих значений линейной модели, поэтому является естественным, что остатки также значительно меньше. Величина КД безразмерна, однако в двух уравнениях она относится к разным статистическим понятиям. В одном уравнении она измеряет объясненную моделью долю дисперсии зависимой переменной, а в другом — дисперсии логарифма зависимой переменной. Для сопоставимости значений критериев в работе [11] рекомендуется применение процедуры Бокса-Кокса. Данная процедура предполагает преобразования масштаба наблюдений зависимой переменной, при котором обеспечивалась бы возможность непосредственного сравнения суммы квадратов отклонений.

Проведенный сравнительный анализ критериев эффективности моделей прогнозирования показал, что для непосредственной оценки адекватности и точности модели прогнозирования могут быть использованы критерии (3–5, 7, 8).

Критериями, которые позволяют учитывать помимо адекватности модели степень оптимальности ее структуры, выражаемую через количество включенных в модель прогнозирования параметров, являются критерии, представленные в уравнениях (9, 12–14). На основе анализа их недостатков и ограничений наиболее предпочтительным при сравнении моделей является критерий (14). Это связано с тем, что данный критерий помимо того, что учитывает адекватность и степень сложности структуры модели, также ориентирован на минимизацию ошибки выбора структуры модели прогнозирования [1], что наиболее полно отвечает требованиям поставленной задачи. Результаты

исследований, приведенные в работе [17], свидетельствуют о том, что критерии MAD , MSE и MPE дают согласованные результаты, а критерии КД, СКД, C_p , AIC и BIC не противоречат друг другу [18].

Таким образом, для сокращения размерности решаемой задачи и сокращения количества альтернатив критериев предложено использовать в качестве критерия, удовлетворяющего требованиям $E_1 - E_3$ постановки задачи, Байесов информационный критерий BIC .

Смешенность прогноза относительно реальных данных позволяет оценить средняя процентная ошибка (6), а автокорреляцию ошибок — Q-статистика Льюнга-Бокса (10). Также было отмечено, что перед расчетом критериев эффективности при сравнении моделей, имеющих различные функциональные формы, необходимо производить нормировку исследуемого временного ряда процедурой Бокса-Кокса.

4. Определение обобщенного критерия эффективности моделей прогнозирования

Для формализации процедуры принятия решения о наиболее точной математической модели прогнозирования исследуемого процесса необходимо разработать обобщенный критерий оценки эффективности. В связи с тем, что наиболее информативным критерием с точки зрения постановки задачи, как показали результаты проведенного выше анализа, является Байесовский информационный критерий, а критерии MSE (6) и LBQ (10) являются второстепенными, то имеет место многокритериальная задача с равноважными критериями.

Существует два основных подхода сведения многокритериальной задачи такого вида к однокритериальной [19]:

- свертка локальных критериев с некоторыми весовыми коэффициентами;
- приоритет важнейшего критерия.

Учитывая то, что значения рассчитываемых критериев имеют неодинаковую размерность, а нормировка их значений не может дать возможность выполнять какие-либо действия с ними ввиду того, что критерии имеют различные ограничения и экстремумы, нахождение свертки критериев является затруднительным. Поэтому целесообразно ориентироваться на второй способ — приоритет важнейшего критерия. Для этого может быть использован один из методов, основанных на данном подходе [20]:

- уступок;
- условной оптимизации.

Метод уступок предполагает при выборе оптимального решения оперировать некоторыми наперед заданными уступками по критериям, которые неизвестны.

Метод условной оптимизации основан на нахождении условного экстремума главного критерия f_j , при условии, что дополнительные критерии (второстепенные) должны быть ограничены некоторыми установленными величинами C_i :

$$f_i(z) \rightarrow \text{extr},$$

$$z \in Z; f_i(z) \{ \leq \geq \} C_i; i = 1, 2, \dots, m; i \neq 1. \quad (15)$$

Таким образом, основным критерием эффективности представим (14), а его ограничения – уравнениями (6) и (10). Тогда искомым критерий, который назовем *обобщенный критерий эффективности моделей прогнозирования GCEFM* (Generalized Criterion for Efficiency of Forecasting Models, GCEFM), примет вид:

$$GCEFM = \min_{z \in Z} \{ BIC \mid LBQ < \chi_m^2, |MPE| \leq 5 \}. \quad (16)$$

На основе полученного обобщенного критерия эффективности *GCEFM* из всех конкурирующих моделей прогнозирования выбирается та модель, которая обеспечивает минимальную среднеквадратическую ошибку, минимальную ошибку выбора структуры модели, оптимальную структуру модели, отсутствие автокорреляции ошибок, а также характеризуется несмещенным получаемым на основе нее прогнозом.

Блок-схема принятия решения на основе обобщенного критерия эффективности *GCEFM* приведена на рис. 1.

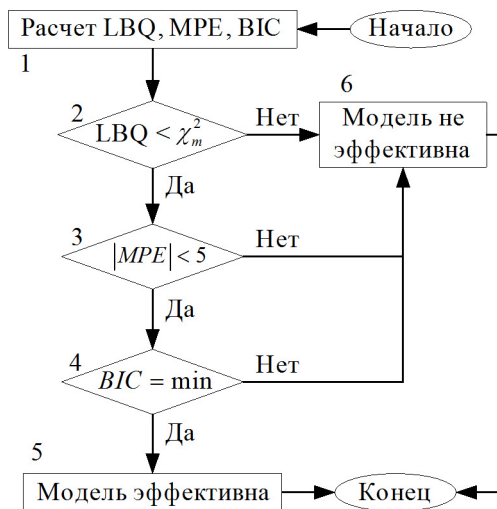


Рис. 1. Блок-схема *GCEFM*

Первоначально для каждой из анализируемых моделей производится расчет Q-статистики Льюнга-Бокса для ошибки модели, средней процентной ошибки и Байесовского информационного критерия. Затем на втором шаге осуществляется проверка $LBQ < \chi_m^2$, и в случае отсутствия автокорреляции ошибки – переход к шагу 3, иначе модель признается неэффективной. На шаге 3 выполняется проверка $|MPE| < 5$, и если условие выполняется, то переходим к шагу 4, иначе к шагу 6. На 4 шаге определяется, является ли минимальным *BIC* среди моделей удовлетворяющих условиям представленных на шаге 2 и шаге 3. Если условие выполняется ($BIC = \min$), то наиболее эффективная модель среди рассматриваемых моделей считается найденной (шаг 5), а при невыполнении этого условия – модель не эффективна (шаг 6).

5. Апробация обобщенного критерия эффективности *GCEFM*

Проведем моделирование исследуемого временного ряда y длиной 500 измерений с помощью уравнения регрессии:

$$y = 1,5x_1 + 0,7x_2 + 0,6x_3 + 0,35x_4 + 0,2x_5 + \varepsilon,$$

в котором регрессоры x_i и ошибка ε генерируются с помощью генератора псевдослучайных чисел с нормальным законом распределения, математическим ожиданием 0 и дисперсией 1. Полученный в результате моделирования временной ряд y представлен на рис. 2.

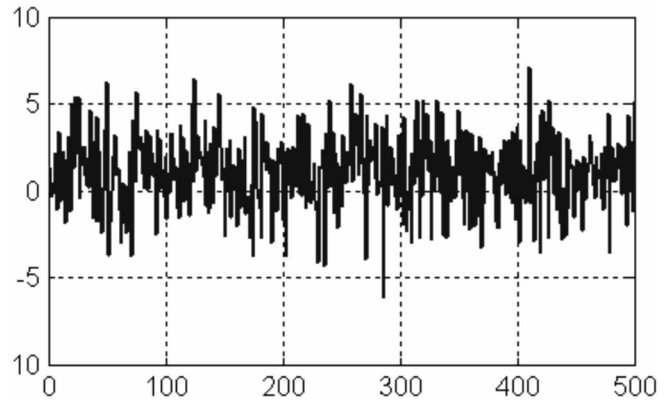


Рис. 2. Исследуемый временной ряд

Построим четыре регрессионные модели временного ряда y по неполным наборам регрессоров, что на практике является достаточно обыденным случаем при построении модели регрессии.

Первая модель z_1 включает регрессор x_1 , вторая модель z_2 – x_1 и x_2 , третья модель z_3 – x_1 и x_3 и четвертая модель z_4 – x_1 и x_4 . Определим оценки коэффициентов моделей z_1 и z_4 с помощью МНК и выполним расчет значений критериев эффективности (3–8, 10, 13, 14) для каждой из моделей (табл. 1).

Как видно из таблицы, критерии *MSE*, *SSE*, Тейла, *AIC* и *BIC* дают согласованный результат и выделяют модель z_4 как наиболее точную в сравнении с другими моделями. Однако ни один из перечисленных критериев не учитывает автокорреляцию ошибки модели. Q-статистика Льюнга-Бокса отвергает гипотезу об отсутствии автокорреляции ошибки для четвертой модели, вследствие чего она не может считаться адекватной и эффективной.

Критерии *MPE*, *MAPE* и КД хотя и дают возможность численно оценить степень точности и адекватности построенных моделей относительно исследуемого временного ряда, однако, как отмечалось выше, они не позволяют учитывать оптимальность структуры построенной модели. Следует заметить, что в соответствии с критерием *MPE* ($|MPE| < 5$) сразу две модели z_1 и z_3 являются адекватными и дают несмещенный прогноз, что требует от исследователя провести дополнительный анализ для определения эффективной модели. Критерии КД и *MAPE* также выделили четвертую модель как

Значение критериев эффективности

Модель	MSE	SSE	MPE	MAPE	Тейла	КД	AIC	BIC	LBQ
z_1	1,8631	931,5417	0,5477	27,8897	0,3242	0,5601	0,6342	0,6595	96,8884 < 124,3421
z_2	1,4408	720,4041	-40,455	47,5404	0,2779	0,6598	0,3812	0,4149	107,5718 < 124,3421
z_3	1,0381	519,1425	-4,7871	17,0696	0,2307	0,7549	0,0576	0,0997	117,2641 < 124,3421
z_4	0,9167	458,3526	7,3448	3,6251	0,2153	0,7836	-0,063	-0,0124	131,4965 > 124,3421

наиболее эффективную, что является неправомерным из-за автокорреляции остатков модели.

Полученные результаты проведенного анализа позволяют сделать вывод, что использование только одного критерия является недостаточным для принятия всестороннее обоснованного и адекватного решения при выборе эффективной модели прогнозирования.

Для принятия решения о том, какая из рассматриваемых моделей $z_1 \div z_4$ является эффективной и адекватной, воспользуемся обобщенным критерием эффективности *GCEFM*.

Расчет критериев *LBQ*, *MPE* и *BIC* выполнен и представлен в таблице. На основании значения Q-статистики Льюиса-Бокса модели $z_1 \div z_3$ подлежат дальнейшему анализу, а модель z_4 является неэффективной. Значение критерия *MPE* для второй модели не удовлетворяет неравенству $|MPE| \leq 5$, в результате чего она не может считаться эффективной. Минимальным значением критерия *BIC* среди оставшихся конкурирующих между собой моделей z_1 и z_3 обладает модель z_3 .

Таким образом, в соответствии с обобщенным критерием эффективности *GCEFM* адекватной и эффективной моделью прогнозирования для исследуемого временного ряда y является z_3 .

Следует отметить, что применение обобщенного критерия эффективности *GCEFM* является наиболее перспективным для отбора эффективных моделей прогнозирования в тех случаях, когда традиционные критерии не позволяют дать однозначный ответ.

Моделирование и расчеты были проведены с использованием математического пакета MatLab [22].

Выводы

Сформулирована и поставлена задача нахождения обобщенного критерия эффективности *GCEFM*, который дает возможность осуществить выбор из множества конкурирующих между собой моделей прогнозирования временных рядов модель, имеющую максимальную эффективность.

Проведен сравнительный анализ критериев эффективности, который позволил сделать вывод о жесткой специализации существующих критериев эффективности на определенных качественных показателях модели прогнозирования и об отсутствии единого критерия, который бы удовлетворял всем требованиям, представленным в постановке задачи. Кроме того, перед расчетом критерия эффективности была отмечена необхо-

димость нормирования временного ряда, используя процедуру Бокса-Кокса, что позволяет обеспечить сопоставимости отклонений получаемых моделями, имеющими различные функциональные формы.

Определен обобщенный критерий эффективности *GCEFM* по методу оптимизации с нежесткими критериями, основанному на приоритете важнейшего критерия. Найденный обобщенный критерий эффективности *GCEFM* позволяет производить выбор модели прогнозирования с учетом среднеквадратического отклонения, ошибки выбора структуры модели, количества параметров модели, степени смещенности прогноза, а также автокорреляции остатков.

Обобщенный критерий эффективности *GCEFM* прошел апробацию, в ходе которой показано, что в случае, когда традиционные критерии эффективности моделей прогнозирования не позволяют получить однозначное решение и выделить эффективную модель, использование критерия *GCEFM* дает возможность преодолеть ситуацию неопределенности и принять решение о выделении эффективной модели, либо отклонить рассматриваемые модели как неэффективные. Все процедуры и вычисления были рассчитаны с помощью математического пакета MatLab.

Результаты работы могут быть использованы при разработке информационных систем построения моделей прогнозирования, а также при решении задач составления прогнозов технико-экономических показателей и принятия решений в различных отраслях хозяйственной деятельности и науки.

Список литературы: 1. Шамша Б.В., Гуржий А.М., Дудар З.В., Левикін В.М. Математичне забезпечення інформаційно-управляючих систем: Підручник для студентів вищих навчальних закладів. – Харків: ТОВ «Компанія СМІТ», 2005. – 448 с. 2. Одейчук А.Н., Шамша Б.В., Федоров Е.Г. Интеллектуальная система выбора метода прогнозирования стохастических рядов в условиях гетероскедастичности // АСУ и приборы автоматизации. – 2007. – Вып. 138. – С. 9–14. 3. Andrey Odeychuk, Olesya Morozova, Anastasiya Gud. The expert system of search the forecasting method with using of neural network in volatility conditions of initial data // Modern problems of radio engineering, telecommunications and computer science: Proceedings of the international conference TCSET'2008. Lviv: Publishing house of Lviv polytechnic. – 2008. – P. 55–58. 4. Одейчук А.Н., Куклин Н.Н. Использование языка моделирования UML при проектировании информационной технологии разработки модели прогнозирования // АСУ и приборы автоматизации. – 2008. – Вып. 144. – С. 157–164. 5. Ханк Д.Э., Уичерн Д.У., Райс А.Дж. Биз-

нес-прогнозируемые, 7-е изд.: Пер. с англ. – М.: Изд. дом «Вильямс», 2003. – 656 с. **6.** *Льюис К.Д.* Методы прогнозирования экономических показателей / Пер. с англ. и предисл. Е.З. Демиленко. – М.: Финансы и статистика, 1986. – 133 с. **7.** *Ярова В.В. та ін.* Эконометрика: Лекції / Ярова В.В., Александрова А.І. / Харк. нац. аграр. ун-т ім. В.В. Докучаєва. – Х., 2003. – 124 с. **8.** *Чернышев С.Л.* Моделирование экономических систем и прогнозирование их развития: Учебник. – М.: Изд.-во МГТУ им. Н. Э. Баумана, 2003. – 232 с. **9.** *Уразаев Р.П.* Методы генерации алгоритмов прогнозирования при помощи операций над базовыми алгоритмами. – М.: Вычислительный центр АН СССР, 1988. – 25 с. **10.** *Демиленко Е.З.* Линейная и нелинейная регрессия. – М.: Наука, 1987. – 248с. **11.** *Ежеманская С.Н.* Эконометрика / Серия «Учебники, учебные пособия». – Ростов н/Д: Феникс, 2003. – 160 с. **12.** *Арженский С.В., Федосова О.Н.* Эконометрика: Учебное пособие. – Ростов н/Д.: Рост. гос. экон. ун-т., 2002. – 102 с. **13.** *Княженко І.І., Павленко Л.В.* Эконометрия: Навч.-метод. посіб. – Донецьк: ТОВ «Юго-восток, Лтд», 2006. – 187 с. **14.** *Магнус Я.Р., Катышев П.К., Пересецкий А.А.* Эконометрика. Начальный курс: Учеб. – 6-е изд., перераб. и доп. – М.: Дело, 2004. – 576 с. **15.** *Стадник М.П.* Модификация критерия Мэллоуза-Акаике для подбора порядка регрессионной модели // Автоматика и техника. – 1988. – №4. – С. 98–108. **16.** *Носко В.П.* Эконометрика: Введение в регрессионный анализ временных рядов. – М.: МФТИ, 2002. – 273 с. **17.** *Makridakis S., Hibon M.* The M-3 Competition: Results, Conclusions and Implications // International of Forecasting. – 16 (2000). – P. 451–476. **18.** *Суслов В.И., Ибрагимов Н.М., Талышева Л.П., Цыплаков А.А.* Эконометрия. – Новосибирск: Изд. СО РАН, 2005. – 744 с. **19.** *Березовский Б.А. и др.* Многокритериальная оптимизация: Математические аспекты. – М.: Наука, 1989. – 128 с. **20.** *Подиновский В.В., Ногин В.Д.* Парето-оптимальные решения многокритериальных задач. – М.: Наука, Главная редакция физ.-мат. литературы, 1982. – 256 с. **21.** *Грін, Вільям Г.* Эконометричний аналіз / Пер. з англ. А.Олійник, Р.Ткачук.; наук. ред. пер. О.Комашко; передм. О.І.Черняка, О.В.Комашка. – К.: Вид. Соломії Павличко «Основи», 2005. – 1197 с. **22.** *Ануфриев И.Е., Смирнов А.Б., Смирнова Е.Н.* MATLAB 7. – СПб.: БХВ-Петербург, 2005. – 1104 с. **23.** *Абдикеев Н.М.* Проектирование интеллектуальных систем в экономике: Учебник / Под. ред. Н.П. Тихомирова. – М.: Изд. «Экзамен», 2004. – 528 с.

УДК 004.415.2 – 519.246.8

Узагальнений критерій ефективності моделей прогнозування часових рядів в інформаційних системах / А.М. Олейчук // Біоніка інтелекту: наук.-техн. журнал. – 2009. – № 1(70). – С. 113–119.

У статті обговорюється проблема вибору моделі прогнозування в інформаційних системах. Проведено порівняльний аналіз критеріїв ефективності. Побудовано узагальнений критерій ефективності моделей прогнозування методом умовної оптимізації.

Табл.: 1. Іл.: 2. Бібліогр.: 23 найм.

UDC 004.415.2 – 519.246.8

Generalized criterion for efficiency of time series forecasting models in information systems / A.N. Odeychuk // Bionics of Intelligence: Sci. Mag. – 2009. – №1(70). – P. 113–119.

The article is devoted to the forecasting model selection problem in information systems. A comparative analysis of efficiency criteria has been made. The generalized criterion for efficiency of forecasting models has been developed using the conditional optimization method.

Tabl.: 1. Fig.: 2. Ref.: 23 items.

Поступила в редакцію 17.03.2009

УДК 004.8:004.89



ПРОБЛЕМЫ ОРГАНИЗАЦИИ ИНТЕЛЛЕКТУАЛЬНОЙ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В САПР ТП НА МЕЛКОСЕРИЙНЫХ ПРИБОРОСТРОИТЕЛЬНЫХ ПРЕДПРИЯТИЯХ

Г. А. Фролова

Харьковский Национальный аэрокосмический университет им. Н. Е. Жуковского «ХАИ», Украина
Frolova.Galka@gmail.com

Выполнен обзор и анализ интеллектуальных возможностей современных автоматизированных систем технологической подготовки производства (САПР ТП) на современных приборостроительных предприятиях. Рассмотрены проблемы внедрения информационных технологий, выделены задачи технологической подготовки производства, которые требуют автоматизации. Предлагается использовать методы инженерии квантов знаний для решения слабоструктурированных задач принятия решений в системах автоматизации технологической подготовки производства.

ТЕХНОЛОГИЧЕСКАЯ ПОДГОТОВКА ПРОИЗВОДСТВА, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, ИНЖЕНЕРИЯ ЗНАНИЙ, КВАНТОВЫЙ МЕТОД

Введение

Технологическая подготовка производства (ТПП) является сложным процессом, от эффективного выполнения которого во многом зависит не только трудоемкость и стоимость технологической подготовки запускаемого изделия, но и стоимость изготовления самого изделия. По-новому организовать ТПП, сократить время и стоимость ТПП, повысить качество принимаемых решений позволит использование современных информационных технологий. В связи с большим количеством систем автоматизированного проектирования, предлагаемых рынком в настоящее время, все большую актуальность приобретают вопросы, связанные с оптимальным выбором конкретных программ. Большинство систем оказывают значительную помощь в конструкторской подготовке производства, автоматизируют рутинные процессы технологической подготовки производства, но творческие задачи остаются за бортом. В настоящее время мало работ, посвященных вопросам автоматизации ТПП на базе информационных технологий, ориентированных на знания [1], что и определяет актуальность данной статьи.

Цель данной статьи состоит в обзоре и анализе интеллектуальных возможностей современных автоматизированных систем технологической подготовки производства (САПР ТП).

1. Состояние проблемы автоматизации технологической подготовки мелкосерийного приборостроительного производства в современных условиях

В настоящее время заметно возрос интерес предприятий к системам управления технологической подготовкой производства. Украинские предприятия столкнулись с проблемой выбора подходящего средства для своей инфраструктуры. Очевидно, что из имеющихся на рынке систем целесообразно выбрать ту,

которая наиболее эффективно сможет решать задачи предприятия как в настоящий момент, так и в перспективе. Мощные системы автоматизированного проектирования, такие как CATIA, UNIGRAPHICS создаются за рубежом, но не могут помочь из-за несоответствия принципам работы с традиционными бумажными элементами, регламентируемыми ЕСКД/ЕСТД, англоязычностью, неприемлемо высокой стоимостью решений. Таким образом, в решении вопросов автоматизации или поддержки решений технолога при проектировании технологических процессов (ТП) ни одна зарубежная система помощи оказать не может. Они обеспечивают описание геометрических форм, образмеривание, генерацию спецификаций. Но проектирование технологии обработки детали перекладывается на технологов.

Современные российские программные средства автоматизации подготовки производства, использующиеся на приборостроительных предприятиях, представлены следующими системами:

- ВЕРТИКАЛЬ — система автоматизированного проектирования, разработанная российской компанией «АСКОН», для конструкторско-технологической подготовки производства с возможностями оформления технологической документации.
- ADEM (англ. Automated Design Engineering Manufacturing) — российская интегрированная CAD/CAM/CAPP система предназначена для автоматизации конструкторско-технологической подготовки производства (КТПП).
- SprutTP - система предназначена для информационной поддержки процесса технического проектирования и автоматизации разработки технологических процессов от компании ЗАО «СПРУТ-Технология».
- T-FLEXТЕХНОЛОГИЯ (T-FLEX/ТехноПро) — система автоматизированного проектирования, разработанная компанией «Топ Системы», с возмож-

ностями решения задач технической подготовки производства в различных отраслях промышленности.

- **TECHCARD** — универсальная система автоматизации технологической подготовки производства, разработанная Intermech (авторизованный разработчик и системный центр компании Autodesk). TECHCARD может работать с эскизами, созданными в AutoCAD, AutoCAD LT, SolidWorks и в любой другой системе, являющейся OLE-сервером

- **APM Technology** — система автоматизированного расчета и проектирования в машиностроении и строительстве комплекса APM WinMachine.

- **Pro/ENGINEER** — система автоматизации деятельности конструкторов и технологов при разработке и проектировании технологических процессов новых изделий компании Parametric Technology Corporation (PTC).

Таким образом, необходимо, учтя специфику мелкосерийного приборостроительного производства, проанализировать рынок современных автоматизированных систем.

Технологическая подготовка производства изделий на каждом предприятии имеет свои особенности, поэтому подсистемы ТПП каждой компании являются уникальными, хотя и реализованы на базе общих принципов, зафиксированных в соответствующих стандартах ЕСТПП, ЕСКД, ЕСТД и так далее.

Для ТПП на мелкосерийном предприятии принципиальное значение имеют следующие взаимосвязанные факторы: тип производства, преобладание того или иного вида технологического процесса, номенклатура технологических операций и другое.

Это связано, прежде всего, со спецификой мелкосерийного производства (новизна и сложность изделий, многоменклатурность, сжатые сроки изготовления, а также высокие требования к качеству изделий), которая накладывает дополнительные условия на процесс производства.

Технолог при разработке технологического процесса использует ГОСТы, большое количество нормативных документов и справочной информации (альбомы технологической оснастки, классификаторы, метрологические данные и другое)

Разработка технологических процессов — это, с одной стороны, интеллектуальное творчество [2], а с другой — гигантская рутинная работа по составлению комплектов документов.

Во время развития информационных технологий знания становятся ценнейшим достоянием [1-2]. Профессиональные знания технологов являются их интеллектуальным базисом для принятия технологических решений на основе их профессионального опыта и знаний о текущем состоянии предприятия. Совершенно естественно, что эффективная разработка технологических процессов возможна только на основе сбо-

ра, обработки и анализа соответствующей достоверной информации технологами предприятия. В результате этих интеллектуальных операций с информацией создаются ТП, которые в виде определенной структурированной информации могут храниться в соответствующих базах знаний и базах данных. Это необходимо для того, чтобы накопленный положительный опыт находил отражение в базе знаний [3] системы и был доступен для всех, в том числе и для новых сотрудников. Для достижения этой цели нужно предоставить возможность сохранять технологический опыт в системе. Такую возможность и обеспечивают методы искусственного интеллекта. Поэтому сокращение времени подготовки производства, снижение себестоимости продукции и увеличение объемов/номенклатуры выпуска изделий происходит за счет поддержки принятия решений при разработке ТПП.

В силу названных причин возникает необходимость использования при ТПП, наряду с традиционными средствами автоматизации, методов и средств искусственного интеллекта, в частности, в форме знаниеориентированных систем принятия решений [4, 6]. Поэтому при анализе современных систем автоматизации ТПП особое внимание будет уделено их интеллектуальным возможностям.

2. Анализ системы ВЕРТИКАЛЬ

САПР ТП ВЕРТИКАЛЬ [7] является частью комплекса систем компании АСКОН, таких как КОМПАС-3D, ЛОЦМАН-Технолог и других. САПР ТП ВЕРТИКАЛЬ может работать как в комплексе с этими системами, так и автономно.

Проектирование ТП в системе осуществляется в различных режимах. Технологию предоставляется возможность выбора оптимального сочетания режимов проектирования, взаимодополняющих друг друга. Одним из основных режимов проектирования технологического процесса является режим проектирования на основе техпроцесса-аналога. А также в системе ТП ВЕРТИКАЛЬ пользователь может оперировать конструкторско-технологическими элементами. Они объединяют конструкторскую и технологическую информацию об элементах, из которых состоит деталь. Для отображения конструкторско-технологической информации служат дерево конструкторско-технологических элементов [7, 9] и панель для отображения планов обработки.

Формирование дерева конструкторско-технологических элементов [9] осуществляется с помощью специальной библиотеки, в которой конструктивные элементы связаны с типовыми технологическими планами их обработки. Библиотека содержит инструменты визуализации и быстрого поиска необходимых элементов. Эта библиотека представляет собой базу знаний, а создание новой версии объекта на основе имеющихся данных об изделии — процесс принятия решений.

3. Анализ системы ADEM

ADEM [8] был создан как единый продукт, включающий в себя инструментарий для проектантов и конструкторов (CAD), технологов (CAPP) и программистов ЧПУ (CAM). Поэтому он содержит нескольких различных предметно-ориентированных САПР под единой логикой управления и на единой информационной базе.

Наличие русскоязычной версии, поддержка стандартов ЕСКД, ANSI, многофункциональность системы, совместно с интуитивно понятным интерфейсом делают возможным ее применение как в конструкторских бюро, так и непосредственно на производстве.

В системе ADEM можно создавать маршрутные, операционные, маршрутно-операционные ТП [8, 10]. Причем большое количество вспомогательной информации технолог может почерпнуть из конструкторских и технологических библиотек. Например параметрическая библиотека содержит таблицу параметров. Каждая строка таблицы параметров — это вариант исполнения элемента, а каждому столбцу выбранного исполнения соответствует размер чертежа. Так же в системе представлены непараметрические библиотеки, которые содержат элементы стандартного размера. Как правило, это библиотеки для создания схем. В модуле ADEM CAM используются библиотеки режущего инструмента, базы данных материала с его свойствами. По свойствам материала система рассчитывает либо выбирает из базы данных режимы резания.

Вся имеющаяся информация хранится в базе данных и представляет техпроцесс в виде структуры этих данных. Для пользователя работа с системой проектирования техпроцессов состоит в создании сценариев работы технолога (интеллектуальная часть) и в последующем выполнении действий по выбранному сценарию (рутинная часть).

Система ADEM позволяет создать библиотеки, в которых пользователь может задействовать сложные расчеты, выполнять запросы к базам данных, заносить результаты в поля диалогов и многое другое [8]. Но для работы с системой ADEM технолог должен знать встроенный язык программирования системы, иметь навыки инженера по знаниям, так как систему необходимо предварительно обучить. Это потребует дополнительных временных затрат на обучение специалиста, получение знаний и затем на обучение системы.

4. Анализ системы СПРУТ

Инструментальная среда СПРУТ [10, 17] содержит полный набор инструментальных средств, необходимых для разработки конструкторских и технологических САПР.

Рассмотрим укрупненную структуру СПРУТ. Для описания структуры и характеристик объекта проектирования разработано соответствующее визуальное

интерактивное средство описания объекта в виде графа И/ИЛИ с возможными заменами и описанием свойств каждой структурной единицы. Описание структуры объекта проектирования — задача предметного специалиста, далеко не всегда владеющего программированием. Поэтому пользователю предлагается удобное интерактивное средство, требующее (разумеется, помимо экспертных знаний), только умения владеть мышью и клавиатурой. Объект при этом может постоянно изменяться и дополняться пользователем по мере необходимости в процессе развития системы. Для создания баз знаний или методик расчетов предметному специалисту предлагается инструмент, позволяющий в удобном для него виде непосредственно описывать методики расчета. При описании методик могут использоваться, в частности, таблицы, графики, монограммы и тому подобное. При этом в системе СПРУТ [17] поддерживается естественная форма описания методик и алгоритмов выполнения всех этапов проектирования изделия. Однако на практике часто возникают ситуации, когда справочные материалы, используемые для описания методик расчета, содержат ошибки в виде опечаток, своевременно выявить которые могут только предметники. Поэтому необходимо предусмотреть контроль и анализа ошибок, допущенных на стадии описания, и утверждения окончательного варианта руководством.

А также СПРУТ [10, 17] предлагает проводить мероприятия по реструктуризации предприятия, что требует кардинального изменения в сложившихся за десятилетия правилах организации работ предприятия

5. Анализ системы T-FLEX ТЕХНОЛОГИЯ

T-FLEX [11] Технология обеспечивает автоматизированную разработку ТП, а также маршрутной, маршрутно-операционной и операционной технологии.

В системе заложена возможность обучения пользователем и самообучение на примерах конкретного производства. Процесс проектирования технологом в системе T-FLEX Технология заключается в заполнении или редактировании содержимого полей экранных форм, выбору элементов из таблиц, справочников системы и работе с параметрическими чертежами из информационной базы системы.

Информационная база [11] T-FLEX Технология содержит большое количество справочников по составляющим технологических процессов. Это — наименования операций, оборудование, приспособления, вспомогательные материалы, тексты переходов, режущие, измерительные, вспомогательные инструменты, заготовки, комплектующие для сборочных технологических процессов и так далее. К каждому типу технологического оснащения в информационной базе можно добавлять параметры, признаки классификации и иллюстрации. Имеются средства поиска и ускоренно-

го подбора оснащения по различным условиям, а также механизмы фильтрации данных.

Рабочее окно системы T-FLEX Технология обеспечивает отображение маршрута обработки, справочников и расчетов в виде дерева [10, 11]. Данные по технологическому процессу, операциям и переходам структурированы на соответствующих закладках. Переход содержит все параметры поверхностей, которые можно ассоциировать с размером на чертеже или эскизе.

Деятельность технолога строится исходя из принципа «смотри и выбирай» [11]. При этом от технолога требуется: понимание сути процессов ТП, умение управлять системой. Процесс проектирования ТП сводится к заполнению или редактированию содержимого полей экранных форм, выбору элементов из таблицы. Вся остальная работа выполняется в автоматическом режиме.

6. Анализ системы TECHCARD

Одной из составляющих информационного обеспечения системы TECHCARD [12] является база данных технологического назначения, в которой предусмотрено наличие единых каталогов по материалам, операциям, оборудованию и его размещению, а также различных каталогов по переходам, оснастке, типовым технологическим процессам и справочникам.

В состав системы TECHCARD входит полномасштабная экспертная система TECHEXP, которая позволяет организовать произвольные расчеты с использованием математических формул и технологических таблиц.

Например необходимо произвести расчет некоторых параметров по формулам. В большинстве случаев не удастся ограничиться одной, которая была бы всегда верна. Для этого формулы сопровождаются условиями, чтобы экспертная система смогла правильно выбрать нужную формулу в конкретном случае. Каждое условие представляет собой особую формулу, которая возвращает логический результат: «ДА», если условие выполняется, в противном случае — «НЕТ». При необходимости система выберет ту формулу, условие которой выполняется [13].

Если формула, которая могла бы определить зависимость результата от исходных параметров, не найдена, то необходимо воспользоваться технологическими таблицами (аналогичные таблицам, приведенным в технологических справочниках).

По известной информации экспертная система обеспечивает поиск необходимых параметров [12, 13]. Но формула для расчета необходимого параметра может включать в себя несколько других параметров, которые также неизвестны и которые могут быть рассчитаны по другим формулам или таблицам, тогда этот поиск может быть очень сложным и запутанным. Такой процесс вычисления некоторого параметра и называется в TECHEXP выводом решения, а совокуп-

ность расчетных формул, таблиц и условий представляет собой базу знаний.

Рядовой пользователям [13] TECHEXP может настроить систему в режиме трассировки, воспользовавшись подробным списком необходимых действий предложенных системой. Однако компьютер будет выдавать правильные результаты только в том случае, если в него введут правильные знания (формулы, таблицы, условия и так далее.). Неверные знания, естественно, приведут к получению неверных результатов. Хотя система TECHCARD и позволяет создать единую интегрированную программную и информационную среду применительно к различным видам производства, но на практике это требует значительной перенастройки под конкретное предприятие, а также обучения пользователей для работы с экспертной системой TECHEXP [12, 13].

Систему TECHCARD можно использовать в технических отделах как крупных предприятий, так и небольших производственных организаций.

7. Анализ системы APM TECHNOLOGY

APM Technology [14] содержит в себе базу данных, в которой можно найти всю необходимую информацию: о типовых технологических процессах, операциях, материалах, приспособлениях, режущих инструментах и другом.

APM Technology предоставляет возможность настройки баз данных в соответствии с номенклатурой оборудования, технологической оснасткой и другой справочной информацией, используемой на данном предприятии [14, 16].

Технологическая информация в проекте представлена в виде дерева объектов, названия которых совпадают с названиями основных компонентов технологического процесса (операция, переход, приспособление, режущий инструмент, средства измерения и так далее.).

Как правило, разработка технологического процесса осуществляется не с нуля, а на основе технологического процесса-аналога (технологического процесса детали похожей конструкции) или типового технологического процесса, типовых операций и переходов [16]. Текст перехода при разработке операций система формирует автоматически, в зависимости от выбранного типа обрабатываемой поверхности и метода обработки. Поддерживается возможность копирования не только ТП целиком, но и различных его частей, например объектов «Операция», «Переход», «Вспомогательный материал», «Приспособление», «Режущий инструмент» и так далее. Это позволит избежать повторения уже выполненных ранее действий (при разработке других ТП), а следовательно, ускорит процесс разработки за счет используемых на предприятии наработок [14, 18].

8. Анализ системы Pro/ENGINEER

САПР Pro/ENGINEER [15] изначально ориентирована на автоматизацию деятельности конструкторов и технологов, сокращение времени разработки новых изделий и сроков подготовки производства.

В базе данных системы Pro/ENGINEER создаются уникальные записи относительно оборудования и его типа, инструмента, операции, материалов, цеха, используемых режущих, мерительных и вспомогательных инструментов.

Причем в базе данных указывается не только название инструмента, но также тип оборудования, на котором этот инструмент может применяться. Аналогичным образом представлена информация об оборудовании – указывается цех, в котором находится оборудование, а также его тип. В информации о станке указывается тип выполняемой на нем операции.

Благодаря такой структуре базы данных, проектирование ТП осуществляется при помощи многочисленных справочников. В Pro/ENGINEER существует возможность автоматического переноса обработки с одной детали на другую [15]. Для этого необходимо экспортировать созданную обработку в свою библиотеку, а затем, при необходимости, импортировать ее на другую деталь. При этом на новую деталь автоматически переносятся все переходы с определенными режимами резания и весь инструмент, то есть весь отработанный ТП целиком. Существует возможность корректировать любой объект. В некоторых случаях возможно автоматизировать даже передачу ссылок на новую обработку, например путем использования одного и того же имени контура в шаблоне и в новой детали. Указанная возможность может значительно сократить время разработки и обеспечить более качественную обработку за счет применения уже отработанных на конкретном предприятии технологий [15].

9. Возможности применения интеллектуальных методов в ТПП

В результате проведенного анализа было выявлено, что большинство САПР технологической подготовки производства, в силу своей специфики, не являются коробочным программным продуктом. Для нормальной работы с подобными системами необходимо провести целый комплекс настроечных мероприятий в соответствии с требованиями определенного производства. В основном это связано с использованием различного оборудования, технологической оснастки, методов обработки, разнообразных методик к назначению режимов обработки, а также с иной спецификой производств.

Известные САПР объединяет одно свойство: крайне низкий уровень их интеллектуального развития. Они не способны самостоятельно принять ни одного технического решения и в руках инженера, принима-

ющего все решения, являются не более чем усовершенствованным справочником. Таким образом, все богатство инженерных знаний остается в книгах и, по мере способностей и опыта, в человеческих головах. Но люди приходят и уходят, унося из фирм их бесценные сокровища – инженерный опыт. Квалификация современных конструкторско-технологических САПР – это уровень техника-чертежника, а должна она соответствовать уровню ведущего технолога.

При имеющихся достоинствах рассмотренных САПР остается открытым вопрос о том, когда технолог нуждается в интеллектуальной поддержке, какие знания, сколько и из каких источников нужны для решения поставленной задачи, как оценить качество базы знаний. Система не ведет технолога при проектировании ТП, отсутствуют обученные БЗ для поддержки принятия технологических решений.

Путь решения вышеперечисленных проблем [5, 6] в интеллектуальных САПР ТП, которые призваны помогать технологу принимать решения в условиях многокритериального выбора и неопределенностей, обусловленных неполнотой и нечеткостью данных, преобладанием качественных характеристик производственных ситуаций, недостаточной логической и вероятностной определенностью последствий принимаемых решений. Такие проблемы называются слабоструктурированными или плохоформализованными. Для решения подобных проблем необходимо использовать средства и методы инженерии знаний.

Существующие искусственные нейронные сети и методы инженерии знаний, основанные на логических, продукционных, фреймовых и других моделях знаний, недостаточно эффективны из-за несовершенства способов представления и машинного способа манипулирования ими [1-3]. Тем не менее, знаниеориентированное направление остается актуальным в моделировании интеллектуальных умений человека успешно принимать решения в различных условиях неопределенности, благодаря человеческой интуиции и знаниям.

Особенности рассматриваемого процесса ТПП [4] требуют реализации в модели свойства активности знаний, возможности отражения в базе знаний структурных отношений объектов, возможности оперирования нечеткими знаниями, учета различных неопределенностей, использования многоуровневых представлений (данные, модели, метамодели и так далее).

В этом направлении профессор И. Б. Сироджа предложил квантовый подход к инженерии знаний, реализованный посредством разработанного метода разноразрядных алгоритмических квантов знаний (РАКЗ-метод) для принятия идентификационных и прогнозных решений в условиях неопределенности [5, 6].

Идея квантового подхода кроется в новой формализованной структуризации (автоматическом квантовании) информации для компьютерного воспроизведе-

дения умозаключений и рассуждений средствами математической логики и теории алгоритмов [5]. Понятие кванта знаний, то есть k -знания определяется аксиоматически как алгоритмическая структура 0-го, 1-го и 2-го уровней сложности, которая описывает конкретное событие порцией (квантом) информации в виде высказывания и содержит три составляющие: содержательную (семантика), информационную (символы) и процедурную (операторы, алгоритмы). Если квантовому событию можно поставить в соответствие число, то имеем k – знания 0-го уровня, если кортеж чисел (вектор) или матрицу, то k -знания имеют 1-й или 2-й уровень соответственно независимо от типа условий неопределенности. Введенные k -знания представляются множественными, аналитическими, векторно-матричными квантовыми моделями. При этом средством РАКЗ- метода реализуется алгоритмизация РАКЗ- моделей представления знаний и дедуктивного вывода следствий из посылок, опираясь на базу k -знаний (БкЗ), которая строится индуктивно при обучении на примерах [5]. Такое структурирование k -знаний позволяет привлекать для анализа, синтеза и машинного манипулирования k -знаниями богатый математический арсенал алгебраических операторов. Поэтому предлагается использовать методы инженерии квантов знаний для решения слабоструктурированных технологических задач.

Таким образом, можно обеспечить представление процесса ТПП, извлечение, формализацию технологических знаний с помощью построения квантовых моделей для эффективного манипулирования имеющимися знаниями и логического вывода решений.

Выводы

Анализируя интеллектуальные возможности, реализованные в системах ВЕРТИКАЛЬ, ADEM, SprutTP, T-FLEX ТЕХНОЛОГИЯ, TECHCARD, APM Technology Data, Pro/ENGINEER, можно сделать вывод, что под накоплением технологического опыта подразумевается использование типовых и групповых технологических процессов, представление в электронном виде информации, содержащейся в ГОСТах, технологических справочниках, классификаторах, нормативной документации и другое. Под обучением в этих системах понимается возможность внесения информации о новом технологическом процессе для последующего использования и редактирования, а также редактирования расчетных формул и записей данных. А процесс принятия технологических решений заключается в выборе по технологии «смотри и выбери» из имеющихся данных операций технологического процесса.

Лишь в некоторых системах, таких как СПРУТ ТП и T-FLEX ТЕХНОЛОГИЯ, используется аппарат инженерии знаний. Моделирование и манипулирование знаниями основано в этих системах на классических

моделях представления знаний – продукционных, семантических сетях.

В качестве нового решения проблемы интеллектуализации САПР ТП автор предлагает для представления знаний технологов по разработке ТП и манипулирования ими использовать квантовые средства инженерии знаний. Использование квантового подхода для создания автоматизированной знаниеориентированной системы поддержки принятия решений при разработке технологических процессов позволит эффективно организовать ТПП.

Список литературы. 1. *Расел С., Норвег П.* Искусственный интеллект: современный подход (AI: A Modern Approach). – М.: Изд. дом «Вильямс», 2006. – 1408 с. 2. *Джарратано Дж., Райли Г.* Экспертные системы: принципы разработки и программирование : Пер. с англ. – М.: Изд. дом «Вильямс», 2006. – 1152 с. 3. *Гаврилова Т.А., Хорошевский В.Ф.* Базы знаний интеллектуальных систем. – Санкт-Петербург: Питер, 2001. – 384 с. 4. *Алиев Р.А., Абдикеев Н.М., Шахназаров М.М.* Производственные системы с искусственным интеллектом. – М.: Радио и связь, 1990. – 264 с. 5. *Сироджа И.Б.* Квантовые модели и методы искусственного интеллекта для принятия решений и управления. – К.: Наукова думка, 2002. – 490 с. 6. *Сироджа И.Б., Петренко Т.Ю.* Метод разноразмерных алгоритмических квантов знаний для принятия производственных решений при недостатке или нечетких данных. – К.: Наукова думка, 2000. – 247 с. 7. Новая версия САПР ТП ВЕРТИКАЛЬ V3: автоматизация технологического проектирования в едином информационном пространстве предприятия. [Электронный ресурс]. – Режим доступа: <http://vertical.ascon.ru>. 8. ADEM. Материал из Википедии – свободной энциклопедии. [Электронный ресурс]. – Режим доступа: <http://ru.wikipedia.org/wiki/ADEM>. 9. *Андриченко А.* «Вертикаль» – новое поколение технологических САПР: объектный подход. [Электронный ресурс]// САПР и графика. – № 6. – 2005. – Режим доступа: <http://www.sapr.ru/article.aspx?id=7515&iid=306> 10. *Евгеньев Г., Кузьмин Б., Лебедев С., Тагиев Д.* САПР XXI века: интеллектуальная автоматизация проектирования технологических процессов. [Электронный ресурс] // САПР и графика. – №4. – 2000. – Режим доступа: <http://www.sapr.ru/Article.aspx?id=7110>. 11. *Талдыкин В.* T-FLEX Технология – современная система автоматизации технологической подготовки производства. [Электронный ресурс]// САПР и графика. – № 3. – 2006. – Режим доступа: <http://www.sapr.ru/Article.aspx?id=16095>. 12. *Гинзбург И., Куприяничик А.* TECHCARD – самый мощный набор инструментов технолога. [Электронный ресурс]// САПР и графика. – № 6. – 2001. – Режим доступа: <http://www.sapr.ru/Article.aspx?id=7407> 13. *Макеенко С., Саган А.* echCard 7.0 – седьмой виток развития технологической САПР от «ИНТЕРМЕХ» [Электронный ресурс]// САПР и графика – № 6. – 2001. – Режим доступа: <http://www.sapr.ru/Article.aspx?id=16085> 14. *Макеенко С., Саган А.* APM Technology – модуль для проектирования технологических процессов. [Электронный ресурс]// САПР и графика. – №3. – 2006. – Режим доступа: <http://www.sapr.ru/Article.aspx?id=16071> 15. *Волков А., Пасынков И., Саранчин А., Чечиков С.* Pro/TechDoc – средство разработки технологических процессов и подготовки документации по ГОСТ в системе Pro/ENGINEER. [Электронный ресурс]// САПР и графика. –

№2.–2006.–Режим доступа: <http://www.sapr.ru/Article.aspx?id=15611>. 16. APM WinMachine 9.5. [Электронный ресурс].–Режим доступа: <http://www.csoft.vtm.ru/>.–Заголовок с экрана. 17. ЗАО «СПРУТ-Технология». –Режим доступа: <http://www.sprut.ru/index.asp>.–Заголовок с экрана. 18. APM WinMachine.–Режим доступа: <http://www.ascon.kiev.ua/cgi-bin/>. –Заголовок с экрана.

Поступила в редколлегию 19.03.2009

УДК 004.89

Проблеми організації інтелектуальної підтримки прийняття рішень у САПР ТП на дрібносерійних приладобудівельних підприємствах / Г. О. Фролова // Біоніка інтелекту: наук.-техн. журнал. – 2009. – №1(70). – С.120–126.

Зроблено огляд та аналіз сучасних систем, таких як ВЕРТИКАЛЬ, АДЕМ, SprutTP, T-FLEX ТЕХНОЛОГІЯ, TECHCARD, APM Technology Data, Pro/ENGINEER, а також розглянуто проблеми впровадження інформаційних технологій на сучасних дрібносерійних підприємствах. При аналізі САПР ТП розглянуто їх інтелектуальні можливості, котрі дозволяють зберігати та використовувати технологічний досвід на підприємстві. Задля збереження технологічного досвіду доцільно використовувати методи інженерії знань. Розглянуто один з можливих підходів представлення, маніпулювання знаннями та прийняття рішень –квантовий підхід.

Бібліогр.: 18 найм.

UDK 004.89

Problems of the organization of intellectual support of decision-making in CAD of technological process on the small-scale plant / G. Frolova // Bionics of Intelligence: Sci. Mag. – 2009. – №1(70). – P. 120–126.

The review and the analysis of modern systems, such as the VERTICAL, ADEM, SprutTP, T-FLEX TEKHNOLOGIYA, TECHCARD, APM Technology Data, Pro/ENGINEER, and also a problem of introduction of information technologies at the modern small-scale plant are considered. At analysis CAD of technological process, to give attention to their intellectual possibilities which allow to keep and make use of technological experience at the plant. For preservation of technological experience it is expedient to use methods of knowledge engineering. One of possible approaches of representation, manipulation by knowledge and decision-making - the quantes approach is considered.

Ref.: 18 items.

УДК 004.9:159.955



ФОРМАЛІЗАЦІЯ ОБРАЗНОЇ ПАМ'ЯТІ ЛЮДИНИ У ВИГЛЯДІ ГРАФУ

О. В. Бісикало

Вінницький державний аграрний університет, м. Вінниця, Україна
obisikalo@gmail.com

Розглядається загальний підхід до представлення образних потоків та інших складових моделі асоціативної пам'яті людини у вигляді графу. На основі декомпозиції інформаційних потоків визначено структуру та формальні ознаки орієнтованого графу. В цілях моделювання задач мовленнєвої діяльності запропоновано формальні критерії якості для оцінки сенсу образу та події.

ОБРАЗ, ПОДІЯ, АСОЦІАТИВНА ПАМ'ЯТЬ, ІНФОРМАЦІЙНІ ПОТОКИ, ОРІЄНТОВАНИЙ ГРАФ, КРИТЕРІЇ СЕНСУ

Вступ

Моделювання образного асоціативного мислення людини ще не сприймається науковою спільнотою як окремий напрямок досліджень, незважаючи на помітні успіхи у розпізнаванні образів. Такий стан речей пояснюється складним феноменологічним характером об'єкта дослідження та відсутністю загально визнаної теорії інтелекту. Внаслідок цього проблемним лишається низький рівень формального розуміння сенсу природномовних конструкцій в сучасних інформаційних технологіях комп'ютерної лінгвістики та електронного навчання. Актуальність результатів моделювання образного мислення людини визначається широким колом можливих програмних додатків, впровадження яких поки що обмежується недостатнім рівнем розуміння сенсу природномовних конструкцій.

Опубліковані підходи до теоретико-модельної формалізації процесів мислення та рефлексії, де враховуються поняття свідомості, представлення, образу, онтології, але лишається поза розглядом онтогенез психічних функцій людини [1]. Проте багато авторів відзначають, що існуючі підходи до моделювання операцій образного мислення мають на даний момент скоріше концептуальний, аніж практичний характер [2].

В роботі [3] запропонована функціональна модель психічної діяльності людини (рис. 1), яка враховує онтогенез та взаємодію таких блоків головного мозку:

- 1 – ефективна і афективна складові нервової системи, блок емоцій (стволові відділи мозку);
- 2 – виконавчі механізми (моторні відділи мозку);
- 3 – надоперативна пам'ять або буфер для створення складних образів (скронева область);
- 4 – блок створення складних рухів або кінестетичних мелодій (передмоторна область);
- 5 – оперативна пам'ять (зона ТРО);
- 6 – довготривала пам'ять (потилична область);
- 7 – асоціативна пам'ять (тім'яна область);
- 8 – блок моделювання або формування і вибору стратегій (постцентральні області);
- 9 – блок програмування, регулювання і контролю (лобові доли).

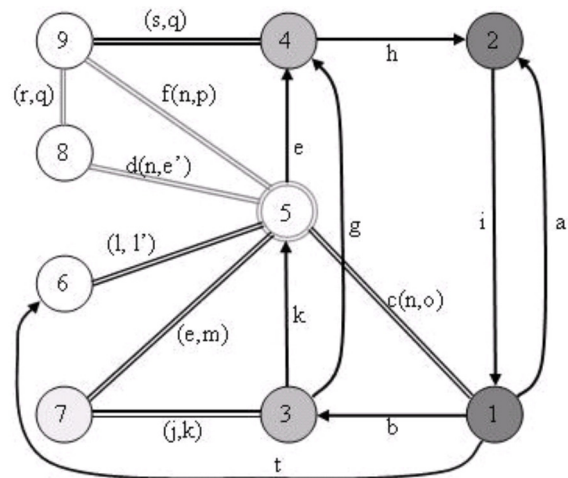


Рис. 1. Функціональна модель психічної діяльності

На рис.1 використані наступні позначення для інформаційних потоків, що відповідають нервовим та психічним процесам людини:

- a) – пряма рефлекторна дуга;
- b) – відчуття;
- c) – отримання вектора емоцій для образів активної ситуації;
- d) – отримання образу рішення для образів активної ситуації в результаті впливу (вливання) сенсів;
- e) – образ у фокусі уваги або інтегрований образ рішення (дії);
- f) – отримання образу контролю стратегії для образів активної ситуації;
- g) – сприйняття і уявлення, що беруть участь в умовних рефлексах;
- h) – складні команди (кінестетичні мелодії);
- i) – команди управління з моторних відділів;
- j) – сприйняття образів;
- k) – представлення образів;
- l) – записані та відтворені (l') події у вигляді ансамблю образів;
- m) – асоційовані образи;
- n) – образи активної ситуації;
- o) – вектор емоцій для образів активної ситуації;

p) – складовий образ неспівпадіння контрольної та активної ситуацій;

q) – складовий образ мети (стратегія);

r) – план (програма) дій у вигляді послідовності образів;

s) – набір стійких рефлексорних автоматизмів для виконання інтелектуальних дій.

Розглянутий підхід до функціонального моделювання психічної діяльності поки що не підкріплено формальною універсальною моделлю образної пам'яті, що підтримує функції всіх типів пам'яті людини – надоперативної, оперативної, довготривалої та асоціативної.

Постановка завдання. Виходячи з особливостей функціональної моделі психічної діяльності формалізувати модель образної пам'яті людини у вигляді графу.

1. Декомпозиція образних потоків

З метою визначення суттєвих ознак образної пам'яті людини розглянемо природні типи розпізнавання інформації, закодованої у вигляді сигналів нервової системи. Виходячи з фізіологічних та психологічних основ моделювання образного мислення [3] приймемо робочу гіпотезу про те, що траєкторія нервового збудження може передавати такі типи сигналів:

1) кількісний параметр – базовий тип інформаційного сигналу;

2) якісний параметр як обмежена множина кількісних параметрів;

3) образ (гештальт) як обмежена множина якісних параметрів;

4) подія як обмежена множина образів.

Така послідовність типів сигналів відповідає логічному принципу від простого до складного та, скоріш за все, має філогенетичну хронологію. Інформаційні потоки кількісної природи безперервно надходять через всі 5 органів чуття людини і так само безперервно розпізнаються в автоматичному (несвідомому) режимі. Базовий тип нервового збудження задається виключно кількісними параметрами – силою та тривалістю сигналу, а також специфічними для кожного органу чуття властивостями, наприклад, світових або звукових хвиль і таке інше [4]. Результатом первинного розпізнавання можна вважати ідентифікацію певної множини або комбінації кількісних параметрів у вигляді якісного параметру, що має безпосереднє відношення до одного з раніше розпізнаних образів. Звідси якісні параметри (інакше – відчуття образів) представляють собою ознаки або властивості, яким, як правило, ставиться у відповідність вербальний (символьний) еквівалент, наприклад, «зелений», «великий», «голосний», «приємний», «повільний», «небезпечний», «солодкий», тощо.

Згідно з нейропсихологічними основами моделювання [5] первинне розпізнавання відбувається на несвідомому рівні, проте в результаті інтроспективного

аналізу може згодом «спливати» у свідомість саме завдяки вербальним ознакам відповідних образів. Наступним етапом розпізнавання, що спочатку також відбувається поза свідомістю, є феномен сприйняття людиною цілісного образу або гештальту. Будь-яке представлення образу об'єднує в єдине ціле досить значну, але обмежену кількість якісних ознак та властивостей, що пов'язують його з іншими образами. Отже, представлення образів у пам'яті людини визначаються певною множиною найбільш значущих (в першу чергу – статистично) зв'язків, що можна продемонструвати відомими загадками на зразок «Сам – алий, сахарний, кафтан – зелений, бархатний» або «Два кольця, два кінця, посередине гвоздик».

Третім природним типом розпізнавання для людини можна вважати одночасне сприйняття ансамблю з 7 ± 2 взаємодіючих образів у вигляді події [3]. Під час події кожний образ може грати одну з ролей, найголовнішими з яких є об'єкт дії, власне дія (метод) та суб'єкт дії. Інші складові активної образної ситуації при необхідності характеризують ознаки (властивості) об'єкта та/або суб'єкта та/або обставини дії. Даний тип розпізнавання належить до природних, тому що, так само як і попередні два, спочатку здійснюється несвідомо. У певних випадках, пов'язаних з небезпечними для людини обставинами, ця первинна тваринна здатність до рефлексорного сприйняття та реагування на події є найбільш ефективним або, навіть, єдиним шляхом до самозбереження. Мовленнєва діяльність людини, хоча і вважається вищим типом психічної діяльності, так само побудована на комунікації подій та, відповідно, синтагматичному устрої мови [5].

Виходячи з запропонованої класифікації, проведемо аналіз введених в структурній моделі образного мислення інформаційних потоків. Загальна кількість можливих об'єктів аналізу згідно з [3] дорівнює 17-ти, оскільки функціональна схема вміщує позначення для 3-х петель, що мають змішану природу сигналу – с), d) та f). Сигнал з кількісними параметрами передається у 4-х потоках а), i), o) та t), з якісними параметрами – у 3-х потоках b), h) та s). Сигнал з цілісними образами передається у 5-х потоках e), g), j), k) та m), а сигнал з подіями – у 5-х потоках l) та l'), n), p), q) та r). Результати проведеного аналізу узагальнено в табл. 1.

Таблиця 1

Відповідність інформаційних потоків моделі типам сигналу нервового збудження

№ п/п	Тип сигналу	Позначення інформаційних потоків	Кількість потоків
1.	Кількісний параметр	a), i), o), t)	4
2.	Якісний параметр	b), h), s)	3
3.	Образ	e), g), j), k), m)	5
4.	Подія	l), n), p), q), r)	5

2. Формальне представлення образної пам'яті

Представимо образну пам'ять людини у вигляді асоціативної мережі, що відповідає ланцюгу для 3-х розглянутих типів розпізнавання: кількісний параметр – якісний параметр – образ – подія. В загальному випадку така мережа може складатися з n образів, потенційно пов'язаних між собою зв'язками типу «кожний з кожним». Поява нового i -го образу в асоціативній мережі пов'язується з новим результатом 1-го типу природного розпізнавання (якісним параметром). **Ідея формалізації** розглянутої асоціативної мережі образної пам'яті полягає у використанні орієнтованого графу, який характеризується певними особливостями. Вершиною графу можна вважати образ, що має унікальні вербальні ознаки сутності якості–об'єкт–метод [6]. Ребрам графу є спрямований асоціативний зв'язок від головного i -го до підлеглого j -го образу, причому вага ребра Ve_{ij} залежить від статистично зафіксованої кількості k_{ij} фактів появи такого зв'язку у попередніх подіях згідно з виразом

$$Ve_{ij} = 1 - e^{-k_{ij}} .$$

Відповідно вага Ve_{ji} зворотного ребра аналогічно залежить виключно від статистичної кількості k_{ji}

$$Ve_{ji} = 1 - e^{-k_{ji}} ,$$

причому, в загальному випадку

$$Ve_{ij} \neq Ve_{ji} .$$

Фрагмент орієнтованого графу представлено на рис. 2, де кольором виділено образи, що асоціативно пов'язані з центральним образом.

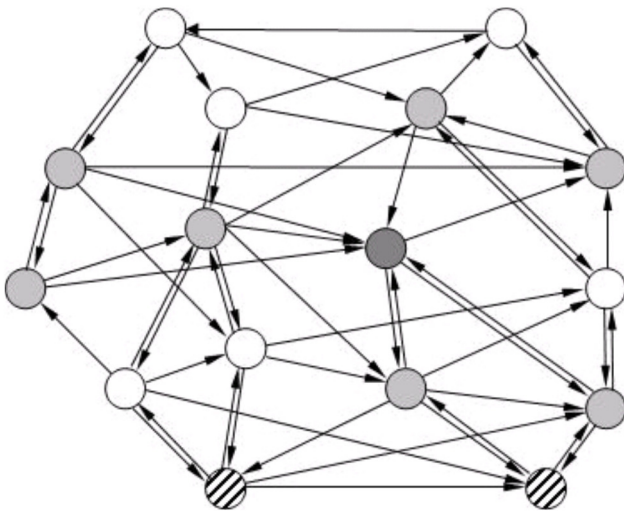


Рис.2. Фрагмент асоціативної мережі образів: визначення образу та пари образів

Розглянутий на рис.2 фрагмент можна вважати ілюстрацією формального опису окремого образу шляхом визначення його зв'язків з іншими образами мережі в результаті 2-го типу природного розпізнавання. Вага ребра графу при такому підході характеризує силу асо-

ціативного зв'язку між двома образами асоціативної мережі образів (АМО) людини – якщо зв'язок між i -м та j -м образами відсутній, то вага Ve_{ij} дорівнює 0; якщо цей асоціативний зв'язок максимально сильний, то вага відповідного ребра наближується до 1:

$$0 \leq Ve_{ij} < 1 .$$

Можна стверджувати, що значення введеної ваги ребра графу оцінює статистичний сенс асоціативного зв'язку між парою образів з урахуванням напрямку асоціації. Очевидно, що образна пам'ять кожної людини має унікальний характер, оскільки залежить від місця та обставин народження, оточення, виховання та освіти, тощо – всі ці чинники будують індивідуальний ланцюг подій довжиною в життя. Тому треба враховувати те, що сенс образів з тими ж самими вербальними ознаками для різних людей неоднаковий, це також поширюється на характеристики АМО у цих людей. Покажемо, що запропонована вага Ve_{ij} дозволяє оцінити сенс окремого образу АМО чи декількох пов'язаних образів, наприклад, дати певні кількісні характеристики для таких відомих вербальних понять як «багатий образ» або «значима подія». Так, будь-який i -й образ АМО, що добре видно на рис. 2, характеризується двома головними параметрами:

- n'_i – кількістю вхідних асоціативних зв'язків;
- n''_i – кількістю вихідних асоціативних зв'язків.

Тоді, знаючи вагу Ve_{ij} відповідного ребра графу для кожного з таких асоціативних зв'язків, можна визначити:

$$Se'_i = \frac{1}{n'_i} \sum_{j=1}^{n'_i} Ve_{ij} , \tag{1}$$

$$Se''_i = \frac{1}{n''_i} \sum_{j=1}^{n''_i} Ve_{ij} , \tag{2}$$

де Se'_i – середнє арифметичне значення «сенсу визначення» образу;

Se''_i – середнє арифметичне значення «вихідного сенсу» образу.

Зрозуміло, що внаслідок виразу (4) для ваги Ve_{ij} , вірними є співвідношення:

$$0 \leq Se'_i < 1 , \\ 0 \leq Se''_i < 1 .$$

Окрім того, для моделювання ряду задач образного мислення, в тому числі мовленнєвої діяльності, потрібно мати оцінку середнього арифметичного загального сенсу i -го образу у вигляді

$$Se_i = (Se'_i + Se''_i) / 2 \tag{3}$$

та значення загальної кількості асоціативних зв'язків i -го образу

$$n_i = n'_i + n''_i . \tag{4}$$

Розглянута сукупність характеристик $n'_i, n''_i, (1), (2), (3), (4)$ в залежності від мети дослідження може засто-

совуватися у якості формальної бази для побудови лінійки складних якісних критеріїв-ознак. Наприклад, при моделюванні мовленнєвої діяльності пропонується вважати вербальними ознаками образу, значення сенсу (9) якого визначено, такі:

- маловідомий образ ($0 \leq Se_i < 0,2$);
- простий образ ($0,2 \leq Se_i < 0,4$);
- популярний образ ($0,4 \leq Se_i < 0,6$);
- складний образ ($0,6 \leq Se_i < 0,8$);
- багатий образ ($0,8 \leq Se_i < 1$).

Парою образів будемо називати дві суміжні вершини графу, пов'язані між собою ребром з певною вагою – на рис. 2 пара образів відокремлена штрихуванням.

За визначенням [6], поняття події може бути представлено як сукупність декількох пар образів. Тому, якщо у фрагменті загального графа виділяти кольором зв'язки та образи окремої події, то отримуємо її формальний опис. На рис.3 таким чином зображено подію, що виникає внаслідок 3-го типу природного розпізнавання, – центральна темна вершина уособлює образ-дію, а інші сірі вузли – об'єкт, суб'єкт і якості (дії та об'єкта).

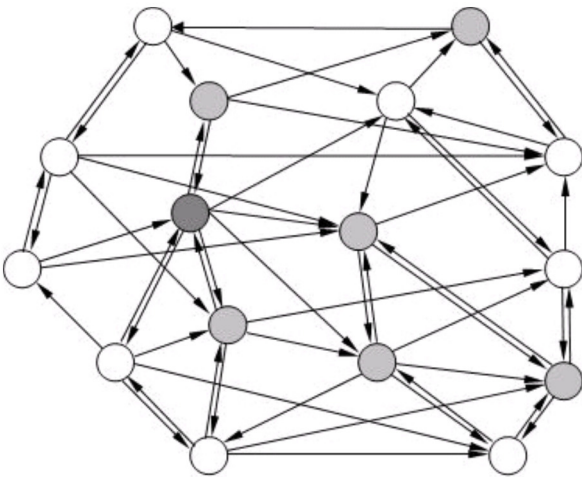


Рис.3. Фрагмент асоціативної мережі образів визначення події

Отже, представивши подію у вигляді послідовного або розгалуженого ланцюга (орієнтованого графа) образів, за основу оцінки сенсу події візьмемо сенс асоціативної пари образів. Загальний сенс пари з i -го головного та j -го підлеглого образів представимо у вигляді:

$$Se_{ij}^p = (Se_i + 2 \cdot Se_j + Se_j) / 4.$$

Тим самим у сенс пари закладається така сама питома вага асоціативного зв'язку, як і загальна середня вага образів, що складають цю пару. Твердження про те, що будь-яке складне дерево події розкладається на асоціативні пари, легко перевіряється за допомогою синтаксичного розбору речення. Відповідно сенс події є середнім значенням сенсу всіх пар, що складають подію

$$Se_i^e = \frac{1}{n_i^e} \sum_{j=1}^{n_i^e} Se_{(ij)}^p,$$

де Se_i^e – сенс i -ї події;

n_i^e – кількість асоціативних пар i -ї події;

$Se_{(ij)}^p$ – сенс j -ї пари i -ї події.

Зрозуміло, що так само, як і для попередньо розглянутих критеріїв сенсу

$$0 \leq Se_{ij}^p < 1, \\ 0 \leq Se_i^e < 1.$$

При моделюванні психічної, у тому числі мовленнєвої, діяльності оцінка сенсу i -ї події Se_i^e також дозволяє поставити у відповідність нормованому числовому параметру множину якісних критеріїв-ознак, наприклад:

- малоймовірна подія ($0 \leq Se_i^e < 0,2$);
- незначна подія ($0,2 \leq Se_i^e < 0,4$);
- типова подія ($0,4 \leq Se_i^e < 0,6$);
- гучна подія ($0,6 \leq Se_i^e < 0,8$);
- значуща подія ($0,8 \leq Se_i^e < 1$).

Формальні поняття та ознаки графа асоціативної мережі образів. На відміну від довільного орієнтованого графа запропонована модель образної пам'яті дозволяє врахувати суттєві особливості його побудови з метою уникнення NP-повної задачі в алгоритмах пошуку на графі. Для розгляду таких особливостей введемо наступні формальні визначення:

1. Образ – вершина графа, що з'єднуються з іншими вершинами не менше, ніж двома ребрами. Кожна вершина має унікальну множину зв'язків з іншими суміжними вершинами, тому у графі не можуть існувати випадки так званих симетричних зв'язків між двома вершинами (рис.4.), що необхідно відслідковувати при додаванні нових вершин та ребер в модель образної пам'яті.

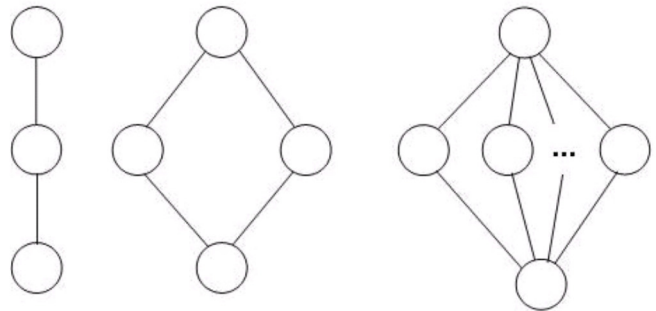


Рис.4. Неприпустимі випадки симетричних зв'язків між двома вершинами графа

2. Кількість ненульових (вага більша за 0) зв'язків від кожної вершини графа до інших вершин набагато менша за $n-1$. Згідно з практикою асоціативного експерименту, при великих значеннях n людина може підібрати не більше 7 ± 2 значущих слів-асоціатів на слово-стимул.

3. Між будь-якими двома вершинами графа розмірністю n обов'язково існує кінцевий шлях довжиною не більше за $(n+1)/2$ у зв'язку з визначенням образу. Найдовший кінцевий шлях за відсутності симетричних зв'язків показано на рис.5.

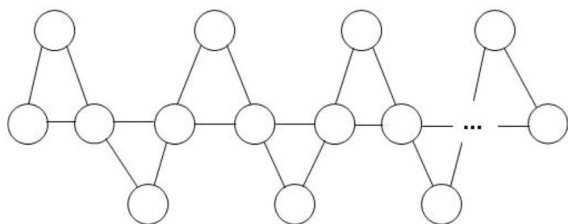


Рис.5. Найдовший шлях між двома вершинами графа

4. Пара образів – ребро між двома суміжними вершинами графа з вагою, що визначається на основі (1). З урахуванням напрямку зв'язку вага ребра $V_{e_{ij}}$ в загальному випадку не дорівнює вазі ребра $V_{e_{ji}}$.

5. Пара як компонент події фіксує одиничний асоціативний зв'язок від головного образу до підлеглого [6]. При внесенні нових подій в модель образної пам'яті необхідно збільшувати на 1 лічильники k_{ij} для тих пар, з яких складається подія.

6. Подія – підграф розмірністю m загального орієнтованого графа, такий що:

- а) може бути представлений у вигляді дерева, коренем якого є вершина-метод;
- б) загальна кількість спрямованих ребер (та відповідних пар) з урахуванням подвійного зв'язку між вершиною-об'єктом та вершиною-методом дорівнює m ;
- с) у кожній вершині дерева кількість підлеглих вершин не може бути більшою за $m-1$.

Висновки

В результаті декомпозиції функціональної моделі образного мислення визначено 4 інформаційних потоки кількісної природи, 3 – якісної, 5 потоків з образами та 5 – з подіями. Запропонована формалізація асоціативної мережі образів у вигляді орієнтованого графа дозволяє врахувати особливості головних концептів моделі образного мислення – образу, пари образів і події та поставити їм у відповідність кількісні характеристики сенсу. При такому підході основні функції образного мислення [3, 6] можна формально звести до двох типів операцій:

- модифікації графа;
- пошуку вершин у графі, в тому числі з метою визначення маршруту.

Перспективним напрямком розвитку запропонованого підходу необхідно вважати застосування та модифікацію відомих алгоритмів пошуку у графі для розв'язання кола специфічних задач образного мислення – моделювання інсайту, побудови асоціативного ланцюга, генерації відповіді на питання, тощо.

Список літератури: 1. Пальчинов Д.Е. Моделирование мышления и формализация рефлексии: I. Теоретико-модельная формализация онтологии и рефлексии // *Философия науки*. – 2006. – № 4 (31). – С. 86–113. 2. Валькман Ю.Р., Исмаилова Л.Р. О языке образного мышления: Доклады междунар. конф. «Диалог 2004». – С. 90–97. 3. Bisikalo O. Principles

of concept model developing of image thinking. First International Conference “New Information Technologies in Education for All”, Extended Conference Proceedings (29–31 May 2006). – Kiev: Akadempriodika, 2006. – P. 25–34 (in Russian). 4. Соколов Е.Н., Вайткявичус Г. Г. Искусственный интеллект: от нейрона к нейромикрокомпьютеру. – М.: Наука, 1990. – 237 с. 5. Лурия А.Р. Язык и сознание / Под ред. Е. Д. Хомской. – М.: Издательство Московского университета, 1979. – 320 с. 6. Бісікало О. В. Структура блоку пам'яті на основі моделі образного мислення людини // *Искусственный интеллект*. – 2007. – № 3. – С. 461–468.

Надійшла до редколегії 23.03.2009

УДК 004.9:159.955

Формализация образной памяти человека в виде графа / О. В. Бисикало // *Бионика интеллекта: науч.-техн. журнал*. – 2009. – № 1(70). – С. 127–131.

В статье рассматривается общий подход к представлению образных потоков и других составляющих модели ассоциативной памяти человека в виде графа. На основе декомпозиции информационных потоков определены структура и формальные признаки ориентированного графа. В целях моделирования задач речевой деятельности предложены формальные критерии качества для оценки смысла образа и события. Ил.: 5. Библиогр.: 6 назв.

UDK 004.9:159.955

Formalization of the human image thinking in a graph form / Oleg V. Bisikalo // *Bionics of Intelligence: Sci. Mag.* – 2009. – № 1(70). – P. 127–131.

The main approach to the introduction of image flows and other components of the human associative memory in the form of a graph form is considered in the given article. The structure and formal indications of a rough graph are defined on the decomposition of data flows. The formal quality criteria for evaluation of image and event sense are suggested aiming at modeling tasks of the speech activity.

Fig.: 5. Ref.: 6 items.

УДК 616.231 – 089.819. 3:614.1



СТАТИСТИЧЕСКИЙ АНАЛИЗ ИНФОРМАТИВНОСТИ ТЕСТА МАЛЛАМПАТИ КАК ПРЕДИКТОРА СЛОЖНОЙ ИНТУБАЦИИ ТРАХЕИ У ДЕТЕЙ

У. А. Фесенко¹, В. Б. Репка²¹ Харьковский национальный медицинский университет, г. Харьков, Украина² Харьковский национальный университет радиоэлектроники, г. Харьков, Украина, vic-repka@yandex.ru

Проведен статистический анализ информативности теста Маллампаати как предиктора сложной интубации трахеи у детей с челюстно-лицевой патологией. Для установления соответствия шкал оценки верхних дыхательных путей по Маллампаати и оценки ларингоскопического вида по Кормаку-Лихейну в работе использован статистический метод «анализ соответствий». Анализ проведен на основании значений процентов схожести категорий и графической интерпретации схожести и представлен для различных возрастных диапазонов детей с челюстно-лицевой патологией.

СТАТИСТИЧЕСКИЙ АНАЛИЗ СООТВЕТСТВИЙ, СЛОЖНАЯ ИНТУБАЦИЯ ТРАХЕИ, ЧЕЛЮСТНО-ЛИЦЕВАЯ ПАТОЛОГИЯ, ТЕСТЫ-ПРЕДИКТОРЫ

Введение

Статистические методы обработки данных давно завоевали популярность как методы обработки информации, получаемой практикующими врачами. Существует даже своеобразный регламент использования в научных публикациях того или иного способа обработки данных, различных критериев оценки полученных результатов, их оформления в текстовом и графическом виде. Наиболее часто применяются такие статистические методы обработки данных, как расчет критериев Стьюдента и Пирсона, коэффициентов корреляции и линейной регрессии и тому подобных. Однако успех применения любого метода зависит от того, насколько хорошо анализируемые данные соответствуют основным предположениям, принятым при разработке статистического метода. Методы анализа, разработанные для определенного типа признаков, могут привести к совершенно неверным выводам при их применении к признакам другого типа, поэтому нужно быть особенно внимательным при выборе метода, адекватного анализируемым данным. Тип исходных данных определяет, какими методами эти данные могут быть обработаны.

Целью нашего исследования был статистический анализ информативности теста Маллампаати у детей с челюстно-лицевой патологией. В предлагаемом исследовании рассматриваются качественные признаки, что и обусловило выбор соответствующего метода статистического анализа – анализа соответствий.

Дыхательные осложнения во время наркоза остаются основной угрозой для жизни пациентов. Частота их за последние десятилетия значительно снизилась с появлением мониторов, современных фиброоптических приборов для трудных интубаций трахеи. Трудные интубации трахеи не всегда бывают предвиденными. Прогноз в таких ситуациях во многом зависит от опыта анестезиолога и готовности необходимого оснащения.

По данным американского исследования документов судебных дел [1], причиной осложнений анестезии были в 34% случаев – дыхательные нарушения, в 6% – проблемы с аппаратурой, в 4% – осложнения со стороны кровообращения, в 4% – медикаменты и их дозы, в 2% – судороги. Причиной респираторных осложнений в 38% случаев была неадекватная вентиляция, в 18% – интубация пищевода и в 17% – сложная интубация трахеи.

В одном французском исследовании [2], в котором анализировали течение 198103 наркозов, зарегистрировано 268 (0,135%) тяжелых инцидентов, из которых 163 (0,082%) целиком отнесены к осложнениям анестезии. Среди этих 163 случаев у 90 (55%) пациентов имели место респираторные осложнения: у 28 – посленаркозная депрессия дыхания, у 27 – аспирация желудочного содержимого, у 16 – осложнения интубации, у 9 – бронхоспазм, у 5 – отказ аппаратуры, у 3 – ларингоспазм, у 2 – пневмоторакс.

Для предвидения трудной интубации трахеи используют разные тесты-предикторы: тест Маллампаати [3], тест «прикусывания верхней губы» [4], измерения тиро-ментального, стерно-ментального расстояний, высоты стояния гортани и другие. Сложность ларингоскопии оценивают по классам Кормака-Лихейна [5]. Классы по Маллампаати теоретически должны совпадать с классами Кормака-Лихейна и предсказывать степень сложности интубации трахеи. Однако информативность теста Маллампаати по данным литературы не очень высокая.

По нашим данным, истинно трудные интубации трахеи у детей с челюстно-лицевой патологией встречаются в 5,5% случаев и обусловлены в основном ограниченным открыванием рта, недоразвитием нижней челюсти, опухолями, которые обтурируют верхние дыхательные пути [6].

1. Постановка задачи

Как уже упоминалось, цель исследования – статистический анализ информативности теста Маллампа-ти у детей с челюстно-лицевой патологией.

Обследовано 249 детей в возрасте от 6 месяцев до 17 лет, которым проведены челюстно-лицевые операции в условиях общей анестезии с интубацией трахеи. Всем детям перед интубацией проведена оценка верхних дыхательных путей по Маллампа-ти в модификации Семсуна-Янга [7].

Ларингоскопический вид оценен по Кормаку-Лихейну. Дети разделены на возрастные группы (табл.1).

Таблица 1

Распределение детей по возрасту и их антропометрические показатели (средняя \pm стандартное отклонение)

Возраст, лет	Вес, кг	Рост, см	Количество детей	% от общего количества детей
От 6 месяцев до 1 года	8,4 \pm 1,1	68 \pm 9,4	28	11,6%
1-3 года	13,3 \pm 2,7	88,9 \pm 9,9	61	24,4%
4-6 лет	19,8 \pm 2,9	112,8 \pm 8,4	24	9,5%
7-10 лет	28,6 \pm 8,7	131,6 \pm 9,9	20	8%
11-13 лет	40,1 \pm 9,9	148,4 \pm 12,6	32	12,8%
14-17 лет	59,1 \pm 11,1	167 \pm 10,7	84	33,7%

Для установления соответствия шкал оценки верхних дыхательных путей по Маллампа-ти и оценки ларингоскопического вида по Кормаку-Лихейну в работе использован статистический метод «анализ соответствий» из-за категориальной природы исследуемых шкал. Существуют два подхода к анализу соответствий. При первом устанавливается взаимное соответствие градаций пары признаков на основании расчета расстояний между градациями (расстояние хи-квадрат); при втором объекты и категории неколичественных признаков представляются в виде точек на плоскости, что позволяет выделить аномальные наблюдения и возможные группировки, строить гипотезы о взаимосвязях. Для установления соответствия двух исследуемых шкал воспользуемся обоими подходами:

1) чем меньше значение статистики хи-квадрат между категориями двух переменных, тем больше эти категории сходны между собой. На основе анализа результатов таблицы процентов схожести категорий видно: чем выше процент, тем больше схожесть между категориями шкал;

2) для анализа графической интерпретации схожести категорий выбраны двух- и трехмерные графики координат. Кружочками обозначаются категории шкалы Маллампа-ти; квадратиками – шкалы Кормака-Лихейна.

Статистический анализ проведен с помощью пакета прикладных программ Statistica 8.0.

2. Результаты исследования у детей

Рассмотрим для каждого возрастного промежутка схожесть по двум переменным (шкалам) Кормаку-Лихейн и Маллампа-ти.

Результаты исследования у детей от 6 месяцев до 1 года. У детей до 1 года оценку по Маллампа-ти проводили при громком плаче, учитывая невозможность продуктивного контакта для максимального открывания рта.

При анализе результатов сравнения для возраста до 1 года по таблице процентов схожести категорий (табл. 2) можно сделать вывод, что по шкале Маллампа-ти однозначно можно установить соответствие классу «3» значения «1» шкалы Кормака-Лихейна (29,6%). «1» класс по Маллампа-ти имел одинаковую схожесть с классами «1» и «2» по Кормаку-Лихейну (по 3,7%). Класс «2» по Маллампа-ти имел одинаковую схожесть с классами «1» и «3» по Кормаку-Лихейну (по 22,2%). Класс «4» по Маллампа-ти во всех случаях совпал с классом «3» по Кормаку-Лихейну, схожесть – 7,4%.

Таблица 2

Проценты схожести категорий по Маллампа-ти и Кормаку-Лихейну у детей до 1 года

Percentages of Total (маллампа-ти)					
Row variables: malamp(4)					
Column variables: kormak(4)					
	1	2	3	4	Total
1	3,70370	3,703704	0,000000	0,000000	7,4074
2	22,22222	0,000000	22,22222	3,703704	48,1481
3	29,62963	0,000000	7,40741	0,000000	37,0370
4	0,000000	0,000000	7,40741	0,000000	7,4074
Total	55,55556	3,703704	37,03704	3,703704	100,0000

На трехмерном изображении соответствия классов (рис. 1) видно, что класс «1» по Маллампа-ти соответствует классу «2» по шкале Кормака-Лихейна (по таблице всего 3.7%, но на графике очевидно это соответствие). Исходя из графика, видно, что значение «2» по Маллампа-ти соответствует в равной степени значению «3» и «4» по Кормаку-Лихейну.

Результаты исследования у детей 1 – 3 лет. Как видно из табл.3, у малышей 1–3 лет наибольшую схожесть имел «1» класс по Маллампа-ти с классом «1» по Кормаку-Лихейну – 20,3%. Класс «2» по Маллампа-ти имел большую схожесть с классом «1» (37,2%) и меньшую с классом «2» (13,5%) по Кормаку-Лихейну. Класс «3» по Маллампа-ти одинаково распределялся по первым трем классам по Кормаку-Лихейну. Также и класс «4» по Маллампа-ти был неинформативен.

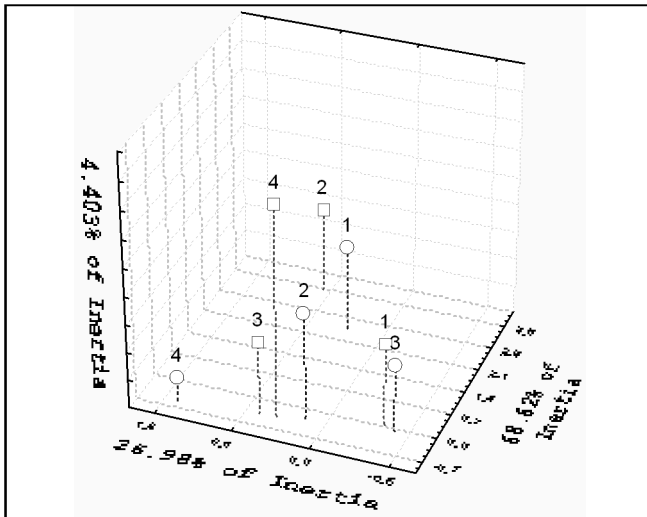


Рис.1. Трехмерное изображение схожести категорий по Маллампати и Кормаку-Лихейну у детей до 1 года

Таблица 3

Проценты схожести категорий по Маллампати и Кормаку-Лихейну у детей 1 – 3 лет

Percentages of Total (малампати)					
Row variables: malamp(4)					
Column variables: kormak(4)					
	1	2	3	4	Total
1	20,33898	1,69492	5,08400	0,000000	27,1186
2	37,28814	13,55932	5,08400	1,694915	57,6271
3	3,38983	3,38983	3,38900	0,000000	10,1695
4	1,69492	0,00000	1,69400	1,694915	5,0847
Total	62,71186	18,64407	15,25400	3,389831	100,0000

Однако в трехмерном изображении (рис.2) можно наблюдать большую схожесть соответствующих групп по Маллампати и по Кормаку-Лихейну.

Результаты исследования у детей 4 – 6 лет. Как видно в табл.4, класс «1» по Маллампати имел преимущественную схожесть с классом «1» по Кормаку-Лихейну (41,6%). Классы «2» и «3» по Маллампати тяготели к классам «1» и «3» ларингоскопического вида. Класс «4» по Маллампати во всех случаях совпал с соответствующим классом по Кормаку-Лихейну (4,16%).

Таблица 4

Проценты схожести категорий по Маллампати и Кормаку-Лихейну у детей 4 – 6 лет

Percentages of Total (малампати)					
Row variables: malamp(4)					
Column variables: kormak(4)					
	1	2	3	4	Total
1	41,66667	4,166667	8,333333	0,000000	54,1667
2	16,66667	0,000000	8,333333	0,000000	25,0000
3	8,333333	0,000000	8,333333	0,000000	16,6667
4	0,00000	0,000000	0,00000	4,166667	4,1667
Total	66,66667	4,166667	25,00000	4,166667	100,0000

На трехмерном изображении (рис.3) схожести категорий у детей 4 – 6 лет четко прослеживается соответствие классов «1», «3» и «4» по Маллампати аналогичным классам ларингоскопического вида. Кругочек и квадратик класса «4» наслоились друг на друга. А вот класс «2» по Маллампати довольно далеко расположен от соответствующего класса по Кормаку-Лихейну и приближается скорее к классу «1» ларингоскопического вида.

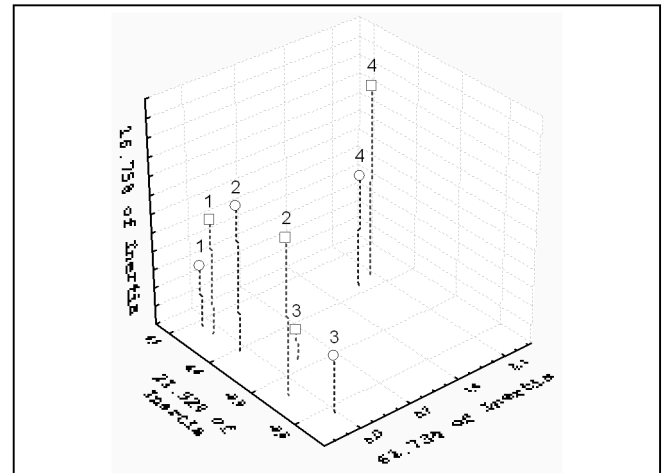


Рис.2. Трехмерное изображение схожести категорий по Маллампати и Кормаку-Лихейну у детей 1 – 3 лет

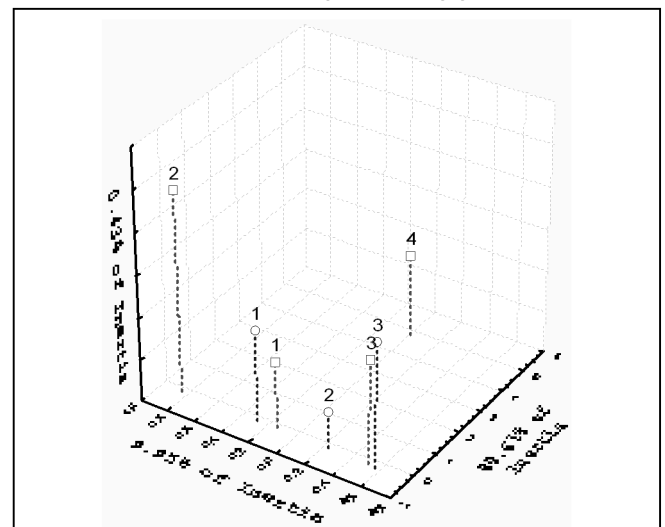


Рис.3. Трехмерное изображение схожести категорий по Маллампати и Кормаку-Лихейну у детей 4 – 6 лет

Результаты исследования у детей 7-10 лет. Как видно из табл.5, у детей 7 – 10 лет первые три класса по Маллампати имели большую схожесть с первым классом по Кормаку-Лихейну. Класс «4» по Маллампати совпадал с классами «3» и «4» при ларингоскопии.

На двухмерном графике (рис.4) эти тенденции видны более четко: классы «2» и «3» по Маллампати наслоились друг на друга и оба абсолютно тяготели к первому классу по Кормаку-Лихейну. А классы «4» по Маллампати и «3», «4» по Кормаку-Лихейну также наслоились друг на друга.

Таблица 5

Проценты схожести категорий по Маллампаати и Кормаку-Лихейну у детей 7 – 10 лет

Percentages of Total (маллампаати) Row variables: malamp(4) Column variables: kormak(4)					
	1	2	3	4	Total
1	36,84211	10,52632	0,00000	0,00000	47,3684
2	15,78947	0,00000	0,00000	0,00000	15,7895
3	10,52632	0,00000	0,00000	0,00000	10,5263
4	0,00000	0,00000	10,52632	15,78947	26,3158
Total	63,15789	10,52632	10,52632	15,78947	100,0000

Результаты исследования у детей 11-13 лет. Из табл.6 можно проследить закономерное смещение классов по Маллампаати к соответствующим классам по Кормаку-Лихейну. Исключением стал класс «2» по Маллампаати, который ни в одном случае не совпал с аналогичным классом при ларингоскопии и больше тяготеет к первому классу по Кормаку-Лихейну.

Таблица 6

Проценты схожести категорий по Маллампаати и Кормаку-Лихейну у детей 11 – 13 лет

Percentages of Total (маллампаати) Row variables: malamp (4) Column variables: kormak(4)					
	1	2	3	4	Total
1	40,62500	0,000000	9,37500	0,000000	50,000
2	25,00000	0,000000	3,12500	0,000000	28,125
3	3,12500	3,125000	0,00000	0,000000	6,250
4	0,00000	0,000000	6,25000	9,375000	15,625
Total	68,75000	3,125000	18,75000	9,375000	100,000

Результаты исследования у детей 14-17 лет. У подростков 14 – 17 лет информативность теста Маллампаати еще больше возрастает. Из табл.7 можно увидеть наибольшую информативность первого класса по Маллампаати (41,6%). Остальные классы также больше тяготеют к соответствующим классам по Кормаку-Лихейну.

Таблица 7

Проценты схожести категорий по Маллампаати и Кормаку-Лихейну у детей 14 – 17 лет

Percentages of Total (маллампаати) Row variables: malamp (4) Column variables: kormak(4)					
	1	2	3	4	Total
1	41,66667	2,380952	3,57143	0,000000	47,6190
2	25,00000	1,190476	2,38095	0,000000	28,5714
3	4,76190	0,000000	2,38095	3,571429	10,7143
4	2,38095	0,000000	4,76190	5,952381	13,0952
Total	73,80952	3,571429	13,09524	9,523810	100,0000

Двухмерное изображение (рис. 6) четко демонстрирует информативность теста Маллампаати по признаку совпадения с соответствующими классами по Кормаку-Лихейну при ларингоскопии. Трехмерный график (рис.5) позволяет увидеть большую информатив-

ность теста Маллампаати у детей 11 – 13 лет по сравнению с более младшими детьми.

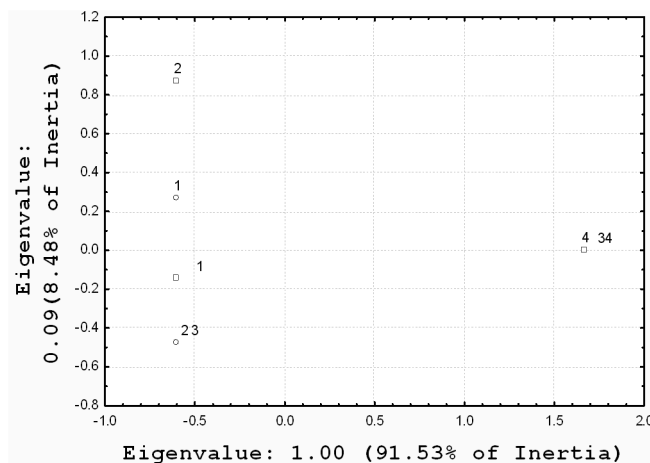


Рис. 4. Двухмерное изображение схожести категорий по Маллампаати и Кормаку-Лихейну у детей 7 – 10 лет

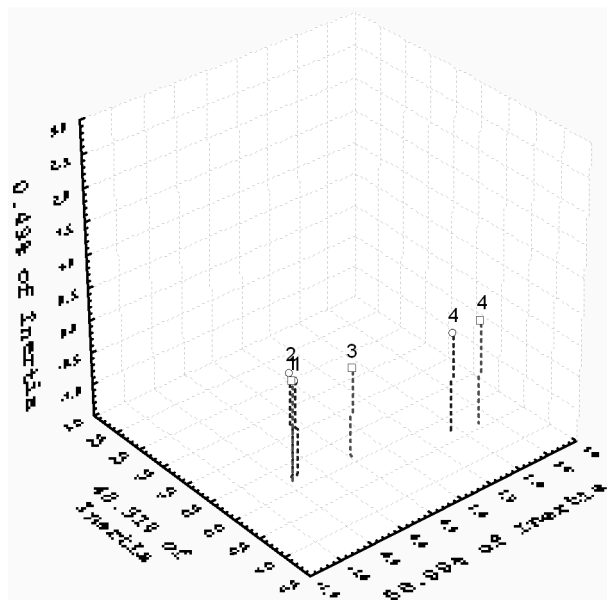


Рис.5. Трехмерное изображение схожести категорий по Маллампаати и Кормаку-Лихейну у детей 11 – 13 лет

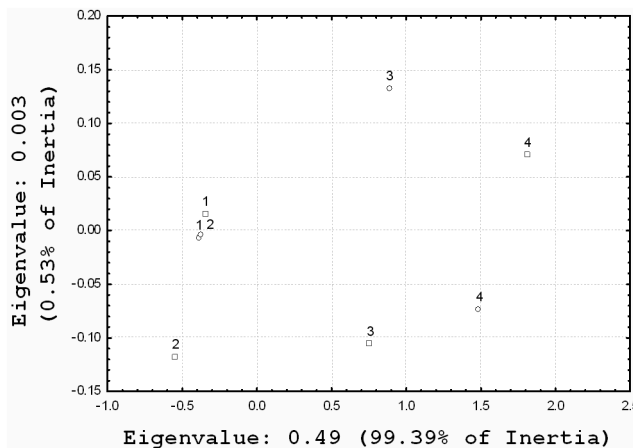


Рис. 6. Двухмерное изображение схожести категорий по Маллампаати и Кормаку-Лихейну у детей 14 – 17 лет

В табл.8 обобщены результаты статистической оценки информативности теста Маллампасти по признаку совпадения с классами по Кормаку-Лихейну. У детей до 4-х лет во всех случаях однозначного трактования результата эти классы не совпадают. А у детей старше 4-х лет все однозначные выводы подтверждают случаи совпадения классов по Маллампасти с классами по Кормаку-Лихейну при ларингоскопии.

Таблица 8

Обобщенные результаты исследования соответствия классов по Маллампасти и Кормаку-Лихейну

Возраст	Классы по Маллампасти	Классы по Кормаку-Лихейну	Вывод
От 6 месяцев до 1 года	1	2	однозначно
	2	3, 4	
	3	1	однозначно
	4	ближе к 4	
1-3 года	1	2	однозначно
	2	3, 4	
	3	1	однозначно
	4	ближе к 3,4	
4-6 лет	1	1	однозначно
	2	1	
	3	3	однозначно
	4	4	однозначно
7-10 лет	1	1	однозначно
	2	1	
	3	1	
	4	4, 3	
11-13 лет	1	1	однозначно
	2	1	
	3	2, 1	
	4	4	однозначно
14-17 лет	1	1	однозначно
	2	1	
	3	1, 4	
	4	4, 3	

Выводы

Таким образом, при применении статистического метода анализа соответствий для установления соответствия шкал оценки верхних дыхательных путей по Маллампасти и оценки ларингоскопического вида по Кормаку-Лихейну было установлено, что:

1. У детей до 4-х лет тест Маллампасти не соответствует ларингоскопической картине по Кормаку-Лихейну и мало информативен в качестве предиктора сложной интубации трахеи.

2. У детей старше 4-х лет тест Маллампасти чаще коррелирует с ларингоскопическим видом по Кормаку-Лихейну и может предсказать сложность интубации трахеи.

В перспективе представляет интерес подробный анализ других тестов-предикторов сложности интубации трахеи у детей.

Список литературы: 1. Caplan RA, Posner KL (1996) Medical-legal considerations: The ASA Closed Claim Project: In: Benumof JL (ed) Airway management, principles and practice. Mosby, St. Louis – Baltimore. 2. Tiret L, Desmonts JM, Hatton F, Vourch G (1986). Complications associated with anaesthesia – a prospective survey in France. Can Anaesth Soc J, 26: 472. 3. Mallampati S.R., Gatt S.P., Gugino L.D., Desai S.P., Waraksa B., Freiburger D., Liu P.L. (1985) A clinical sign to predict difficult tracheal intubation: A prospective study. Can Anaesth Soc J, 32: 429-434. 4. Khan Z.H., Kashfi A., Ebrahimkhani E. (2003) A comparison of the upper lip bite test (a simple new technique) with modified Mallampati classification in predicting difficulty in endotracheal intubation: a prospective blinded study. Anesth Analg, 96 (2): 595-599. 5. Cormack R.S., Lehane J. (1984) Difficult tracheal intubation in obstetrics. Anaesthesia, 39: 1105-1111. 6. Фесенко У.А. (2008) Складна інтубація трахеї у дітей // Медицина сьогодні і завтра. – №2, с.24-29.7. Samsoon G.L.T., Young J.R.B. (1987) Difficult tracheal intubation: A retrospective study // Anaesthesia. – 42. – P. 487-490.

Поступила в редколлегию 25.03.2009

УДК 616.231 – 089.819. 3:614.1

Статистичний аналіз інформативності тесту Маллампасти як предиктора складної інтубації трахеї у дітей / У.А. Фесенко, В.Б. Репка // Біоніка інтелекту: наук.-техн. журнал. – 2009. – №1(70). – С. 132–136.

Застосовано статистичний метод «аналіз відповідностей» для статистичного аналізу інформативності тесту Маллампасти як предиктора складної інтубації трахеї у дітей з щелепно-лицьовою патологією. Даний метод використано для встановлення відповідності шкал оцінки верхніх дихальних шляхів за Маллампасти і оцінки ларингоскопічного виду за Кормаком-Лихейном. На підставі отриманих значень процентів подібності категорій, 2х- і 3х-мірної графічної інтерпретації подібності виявлено збіг і розбіжності значень шкал для різних вікових груп дітей з щелепно-лицьовою патологією.

Табл.: 8. Лл.: 6. Бібліогр.: 7 найм.

UDC 616.231 – 089.819. 3:614.1

Statistical analysis of predictive value of Mallampati test in difficult tracheal intubation in children / U.A. Fesenko, V.B. Repka // Bionics of Intelligence: Sci. Mag. – 2009. – №1(70). – P. 132–136.

The predictive value of the Mallampati test in difficult tracheal intubation in children with maxillofacial diseases is analyzed statistically. Statistical method «correspondence analysis» is applied to the statistical analysis of the Mallampati test informativity as predictor of difficult tracheal intubation in children. Based on the values of similarity categories percent, 2D and 3D graphic interpretation of similarity coincidence and discrepancies of values of scales for various age groups of children with a maxillofacial pathology are exposed.

Tab.: 8. Fig.: 6. Ref.: 7 items.

УДК 007.52



ПРЕИМУЩЕСТВА МОДЕЛИ ОПРЕДЕЛЕНИЯ КРЕДИТОСПОСОБНОСТИ СУБЪЕКТА ОБРАЗОВАНИЯ, ОСНОВАННОЙ НА АНАЛИЗЕ ДАННЫХ, РАЗБИТЫХ НА СМЫСЛОВЫЕ ПОДГРУППЫ

А. В. Ляховец

ХНУРЭ, г. Харьков, Украина
alena-vl@yandex.ru

Проведен анализ ситуации на рынке кредитования образования. Рассмотрены методы и модели, применяемые при принятии решения о целесообразности кредитования объекта образования. Предложена модель системы для определения целесообразности кредитования. Система основана на анализе не только данных финансового состояния заемщика, но и успеваемости, данных о выбранном учебном заведении, специальности, а также других параметрах, отображающих учебный потенциал заемщика. Данные о заемщике и данные об учебном заведении обрабатываются отдельно, и для каждой из групп параметров методы и алгоритмы подбираются индивидуально.

МОДЕЛЬ, СИСТЕМА, ГРУППА ДАННЫХ, ПРОЦЕСС КРЕДИТОВАНИЯ, ЗАЕМЩИК, УЧЕБНОЕ ЗАВЕДЕНИЕ, ЦЕЛЕСООБРАЗНОСТЬ КРЕДИТОВАНИЯ

Введение

Образовательные кредиты до недавнего времени находились на периферии отечественного рынка различных банковских услуг, однако интерес к финансовым инструментам подобного рода в последние годы значительно растет. Программы кредитования учащихся являются одним из важнейших элементов системы финансирования украинского высшего образования. Они способствуют решению проблем доступности качественного образования и направляют в образовательные учреждения дополнительные ресурсы.

Создание в Украине эффективно действующего института образовательного кредитования является важной и актуальной задачей, особенно в условиях жесткого контроля выдачи кредитных средств после экономического кризиса. Эффективность становления этого института и закрепление паттернов поведения клиентов в отношении образовательных займов во многом зависит от политики банков, занимающихся разработкой и предоставлением соответствующих кредитных продуктов. В условиях недостатка информации для выработки политики оценки платежеспособности и вероятности дефолта потенциальных заемщиков одной из возможностей для банка (вместо установления высоких процентных ставок и осуществления дискриминации по доходу) является учет результатов анализа факторов потенциального спроса на образовательные кредиты.

Интегрировавшись в современную экономику, образование стало одной из отраслей, работа которой подлежит экономической оценке по параметрам «цена – качество». Переход к массовому производству работников высшей квалификации выдвинул требования, во-первых, гарантированного стандартного качества выпускников, способных эффективно работать и конкурировать на рынке труда; во-вторых, появилось тре-

бование оптимизации затрат в условиях усложняющейся и растущей подготовки специалистов [1,2].

Все это обуславливает изменение положения вузов, их превращение в рыночные субъекты, проводящие самостоятельную политику по привлечению ресурсов государства, собственного капитала и бизнеса.

Целью данной работы является разработка модели системы, предназначенной для определения целесообразности кредитования субъекта образования, и определение преимуществ данной модели.

1. Анализ ситуации на рынке кредитования образования

Студенческие займы стимулируют спрос на высшее образование, делая его более доступным и помогая преодолеть финансовые трудности, которые могут возникнуть во время обучения. У студентов появляется, и стимул хорошо учиться, чтобы по окончании вуза получить достойную работу и выплатить долг. Таким образом, высшее образование становится более «продуктивным». Можно также отметить, что у студента появляется широкий выбор места учебы, не зависящий от стоимости обучения. Это может привести к повышению конкуренции среди высших учебных заведений.

Законодательными актами определяются принципы выдачи государственных кредитов. Для частного кредитования также необходима четкая модель и принципы выдачи кредитных средств [3].

Учебное заведение со своей стороны получает преимущество перед другими учебными заведениями. Поток абитуриентов увеличивается, повышаются требования к знаниям и способностям учащегося, и образовательный уровень заведения растет. Следовательно, заключая договоры с кредитными учреждениями и продвигая образовательные кредиты, высшие учебные заведения привлекают абитуриентов, повышают конкурентоспособность среди подобных учебных заведений.

Анализируя зарубежный опыт, можно сделать вывод, что украинские схемы кредитования являются несовершенными для области образования. Одной из основных проблем, препятствующих развитию данной отрасли кредитования, является проблема выработки правил классификации новых клиентов. Необходимо достоверная оценка потенциального заемщика, отсеечение «плохих» заемщиков. Неверная классификация порождает проблему обеспечения возврата средств заемщиком в принудительном порядке. На данный момент из-за экономического кризиса предельно упало кредитное предложение. В связи с этим достаточно продолжительное время кредиты смогут быть выданы очень ограниченному количеству клиентов. В этой ситуации еще больше возрастает заинтересованность банка в выдаче кредитных средств более перспективному клиенту.

Кредитоспособность клиента в мировой банковской практике фигурирует как один из основных объектов оценки при определении целесообразности выдачи кредитных средств и выбора формы кредитных отношений. Как в отечественных, так и зарубежных банках (за исключением единичных программ отдельных вузов) основой кредита на образование является потребительский кредит, то есть при принятии решения о выдаче кредитных средств не учитывается важная информация об образовательных успехах абитуриента/студента, выбранном учебном заведении, специальности, а также перспективах абитуриента/студента как самостоятельной личности. Кредит выдается родителям или опекунам абитуриента или студента и перспективы возврата

кредита учащимся не рассматриваются. При этом теряется большая область данных для анализа, влияющих на возвратность кредитных средств.

Исходя из вышеперечисленных проблем в сфере кредитования, а также перспектив в образовательной системе можно сделать вывод, что для более успешного развития кредитования образования необходима система, которая при поддержке принятия решения о возможности кредитования и условиях выдачи кредита анализировало бы не только данные о родителях/опекунах учащегося, но и его данные, данные об учебном заведении, характеристики выбранной клиентом специальности и влияние фактора экономической ситуации [1-6].

2. Разработка модели системы для определения кредитоспособности субъекта образования

Основной целью работы является разработка модели системы для применения в процессе принятия решений о целесообразности выдачи кредитных средств в кредитовании образования. В нынешних рыночных условиях необходим стабильный четкий эффективный механизм для расширения клиентской базы, но в то же время позволяющий не увеличить риски невозврата кредитных средств, а уменьшить.

Для построения модели системы были проанализированы основные потребности пользователей и сформированы в виде запросов к системе, определены участники процесса кредитования образовательных нужд и ожидаемый результат от работы системы. Для достижения данной цели была построена Use case диаграмма (рис. 1).

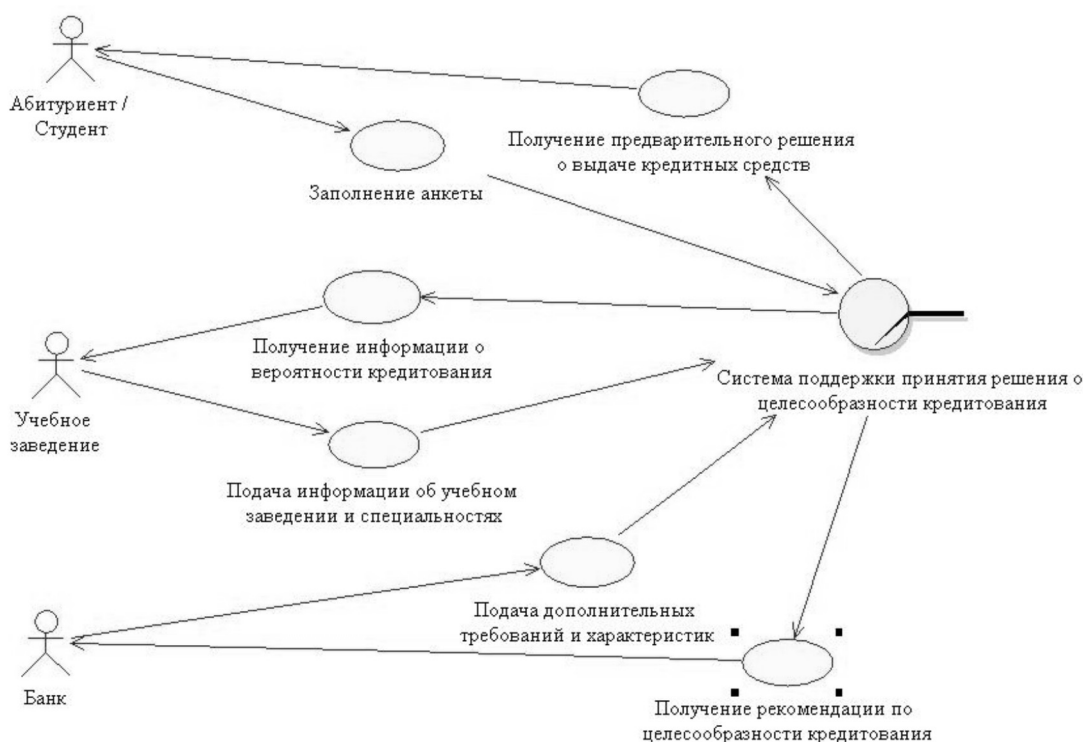


Рис. 1. UML диаграмма работы участников процесса кредитования с системой

На данной диаграмме представлены основные участники процесса получения кредита на образовательные нужды, а также процессы, с помощью которых объекты взаимодействуют. Следовательно, диаграмма отображает основные требования к системе.

Для определения отдельных этапов работы системы, детализации основных процессов и отображения всех необходимых потоков данных была разработана функциональная модель. Функциональная модель 1-го уровня предложенной системы представлена на рис. 2.

Входными данными для модели системы являются 3 группы характеристик:

1. Характеристики абитуриента/студента:
 - пол;
 - возраст;
 - семейное положение;
 - льготы;
 - средний бал аттестата;
 - средний бал по профилирующим предметам;
 - престиж школы;
 - наличие опыта работы после школы (связан с выбранной специальностью или нет);
 - наличие дополнительного образования (связанно с выбранной специальностью или нет);
 - наличие высшего образования у родителей;
 - вероятность материальной поддержки со стороны родителей;
 - награды;
 - участие в олимпиадах;
 - текущая успеваемость (для студентов).
2. Характеристики учебного заведения:
 - рейтинг учебного заведения;
 - уровень аккредитации учебного заведения;
 - рейтинг специальности;
 - процент трудоустройства выпускников специальности;

- востребованность специалистов на рынке;
- уровень зарплат молодых специалистов данной специальности;
- уровень зарплат специалистов данной специальности.

3. Банковские характеристики для определения кредитоспособности заемщика: характеристики, влияющие на выдачу кредитных средств на данный момент времени.

При необходимости в каждую из групп характеристик могут быть добавлены дополнительные поля. Также некоторые поля могут быть удалены из-за их неактуальности [6–11]. Для частного и государственного кредитования характеристики могут отличаться, а также может отличаться принцип анализа результата работы системы.

Наполнение системы данными и обучение моделей включает (рис. 3):

1. Выбор наиболее важных характеристик для дальнейшего анализа. Добавление новых или удаление имеющихся, утративших актуальность.

2. Подготовка обучающего множества. Этот этап является достаточно трудоемким и проблематичным, так как отсутствуют достаточные статистические данные, и отсутствует информация о выплате кредитов студентами, а финансовая информация является конфиденциальной.

3. Препроцессирование данных. На данном этапе проводится подготовка данных для дальнейшей работы и анализа. Данные обрабатываются к необходимому для модели виду.

4. Кластеризация объектов. Обучение моделей для кластеризации и определение кластеров, которые будут служить основой для классификации новых объектов и анализа результатов.

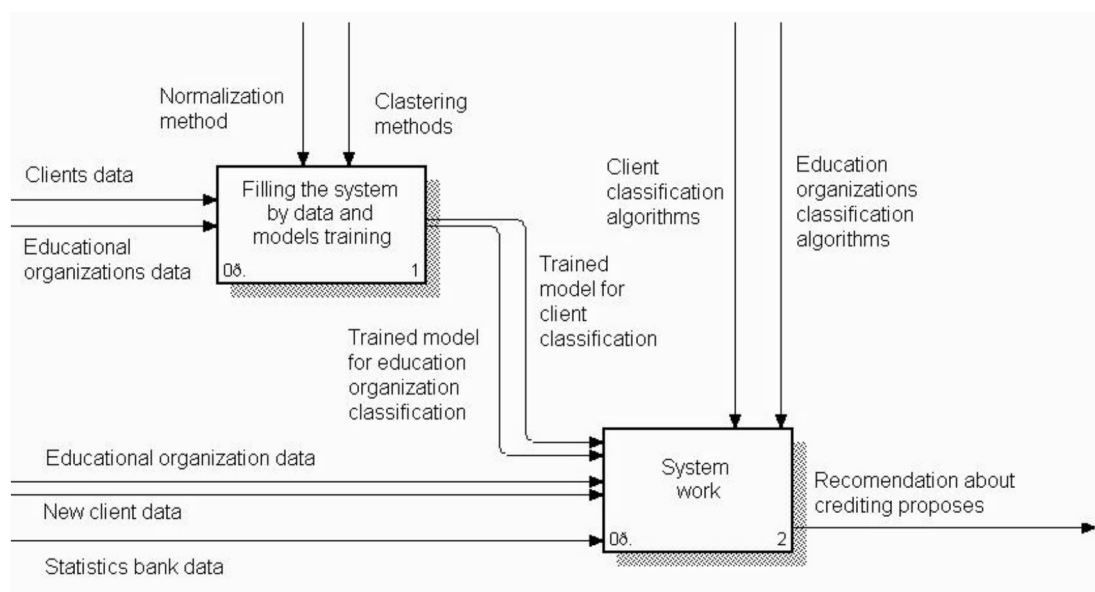


Рис. 2. Функциональная модель 1-го уровня системы кредитования в области образования

Функциональная модель второго уровня, представленная на рис. 3, описывает процесс наполнения системы данными и обучения моделей.

Работа системы подразумевает следующие этапы (рис 4.):

- на вход системы подается информация о клиенте (абитуриенте/студенте), информация об учебном заведении и специальности, возможно, некоторые дополнительные характеристики;
- характеристики клиента и учебного заведения анализируются в отдельных модулях. Для каждой из групп выбирается наиболее подходящий алгоритм. Производится классификация;
- результаты по группам характеристик суммируются, и на выходе суммирующего блока пользователь получает оценку о целесообразности кредитования клиента, рекомендуемые условия кредитования (сроки кредитования, сумма кредита, сроки начала выплат и так далее);
- затем полученный результат обрабатывается с помощью коэффициента значимости

3. Преимущества представленной модели

Преимуществами представленной модели для определения целесообразности кредитования субъекта образования являются:

- при анализе данных по представленной схеме учитываются все необходимые характеристики. На данный момент принципы выдачи образовательного кредита остаются такими же, как и для потребительского, то есть ни успеваемость студента, ни характеристики учебного заведения не учитываются. Эти дополнительные характеристики могут сыграть важную роль в принятии решения о выдаче кредитных средств, сделав его более точным и снизив кредитные риски;
- данные клиента и данные об учебном заведении анализируются отдельно по алгоритмам, подобранным для каждой группы данных. Эти группы отличаются не только смыслом, но и размерами выборок, следовательно, различные методы дадут наилучшие результаты. В условиях отсутствия необходимой статистической базы более аккуратная работа с имеющимися данными приобретает еще большее значение;

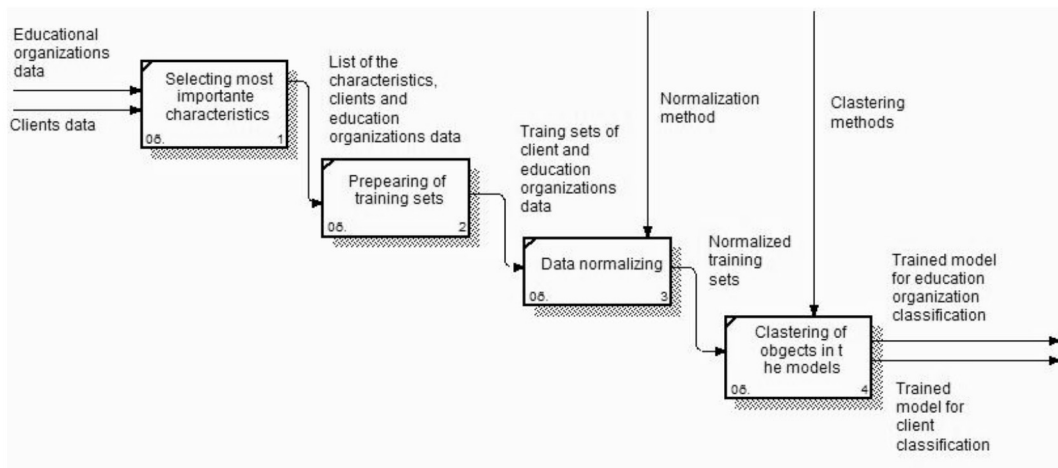


Рис. 3. Функциональная модель 2-го уровня: наполнение системы данными и обучение моделей

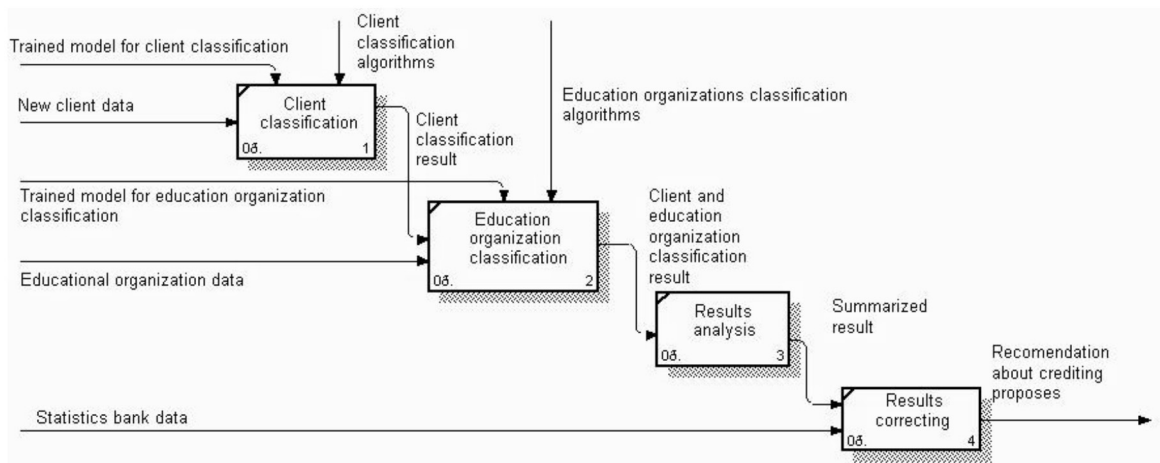


Рис. 4. Функциональная модель 2-го уровня: этапы работы системы

– использование интеллектуальных методов определения кредитоспособности. В системе кроме стандартных статистических, экономических методов, применяемых в банках, предлагаются интеллектуальные, которые являются средствами поддержки принятия решений.

Выводы

Украинский рынок образовательных кредитов еще мало развит, хотя уже наблюдается постоянное расширение предложений банков в области кредитования образования. К причинам непопулярности образовательного кредитования относятся, в частности, слабая информированность населения о программах кредитования, неудовлетворенность существующими программами (например, высокие процентные ставки), недоверие к организациям, выдающим кредит (например, недостаточная прозрачность схемы расчета суммы платежей), а также боязнь перспективы жизни в долг, обусловленная спецификой менталитета. Сегодня родители предпочитают делать сбережения в течение многих лет, чтобы затем инвестировать их в образование своих детей.

Внедрение предложенной системы поможет банку в определении целесообразности кредитования заемщика и соответственно уменьшит риск невозврата кредита, представит более точную картину о заемщике, следовательно, условия и процедура кредитования может быть более индивидуальна.

Упростив процедуру получения кредита на образование, повысив доступность образовательных кредитов для различных слоев населения, банки приобретут новых клиентов, университеты смогут привлечь большее количество абитуриентов, повысит свой уровень конкурентоспособности, а абитуриенты в свою очередь получают образование, независимо от нынешнего материального положения при выборе учебного заведения.

Список литературы: 1. Зидерман А. Роль образовательных займов в расширении доступности высшего образования // Социальная дифференциация высшего образования.— М.: НИСП, 2005.— С. 280–322. 2. Абанкина И. В., Домненко Б.И., Левшина Т. Л., Осовецкая Н. Я. Перспективы образовательного кредитования в России // Вопросы образования.—2004.— №4.— С. 64–88. 3. Постановление Правительства Украины №844 от 27.05.2000 «Про затвердження Порядку надання цільових пільгових державних кредитів для здобуття вищої освіти». 4. Абанкина И. В. и др. Перспективы образовательного кредитования в России // Вопросы образования.—2004.— №4.— С. 64–88. 5. Barr H. Высшее образование: способы и источники финансирования: Пер. с англ. М. Семеновой // Вопросы образования.—2005.— №2.— С. 129–142. 6. Кузьминов Я. И., Бендукидзе К. А., Юдкевич М. М. Курс институциональной экономики. Институты, сети, транзакционные издержки, контракты.— М.: ГУ-ВШЭ, 2006. 7. Паймеева Л. А. Образовательный кредит // Современное образование.— СГУ.— 2005.— №5 (98). [Точный адрес статьи: http://www.edit.muh.ru/newspaper/so_2005/05/st5.htm]. 8. J. Lmi.

Student Loans in an International Perspective: the World Bank Experience // World Bank.—2003.— http://www-wds.worldbank.org/external/default/WDSContentServer/WDS/IB/2003/12/08/000090341_20031208144148/Rendered/PDF/1272950studentloans.pdf. 9. Студенты в болонском процессе и болонский процесс для студентов// Материалы Всероссийского молодежного форума «Я участвую в реформе образования (г. Москва, 31 марта 2006 г.) / Под науч. ред. д-ра пед. наук, проф. В. И. Байденко. —М.: Исследовательский центр проблем качества подготовки специалистов, 2006.— 56 с. 10. Усошкин В. М. Современный коммерческий банк. Управление и операции.—М., 2003. 11. Altbach P.G. Patterns in Higher Education Development // American Higher Education in the Twenty-First Century, eds. Altbach P.G., Berdahl R.O., Gumpert P.J.—Baltimore: John Hopkins University Press, 1999.

Поступила в редколлегию 27.03.2009

УДК 007.52

Переваги моделі визначення кредитоспроможності суб'єкту освіти, побудованої на аналізі даних, розбитих на смислові підгрупи / А. В. Ляховець // Біоніка інтелекту: наук.-техн. журнал.— 2009.— № 1(70).— С. 137–141.

Ця робота містить опис ринку кредитування послуг в освітній сфері для подальшого розв'язання проблеми визначення доцільності видачі кредитних коштів абітурієнту/студенту. Для вирішення цієї проблеми розроблена модель системи. В статті подана Use Case діаграма для визначення основних учасників процесу кредитування та основних потоків даних та очікувані результати від роботи системи. На функціональних діаграмах описані основні процеси та потоки даних.

Л.: 4. Бібліогр.: 11 найм.

UDC 007.52

Advantages of the model for educational subject solvency determination based on data analysis which divided into sense groups / A. V. Lyakhovets // Bionics of Intelligence: Sci. Mag. —2009. — №1(70).— P. 137–141.

The report contains description of crediting market research in educational sphere for following solving task of determining of reasonability giving the credit money to school leaver or student. For this task solving model of the system has been developed. In the report Use Case diagram is presented. This diagram is for determining of main participants of crediting process, main data flows and expected results of the system work. At the functional diagrams main processes and data flows in the system described.

Fig.: 4. Ref.: 11 items.

УДК 519.52



ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ТЕКСТОВ В СИСТЕМАХ МЕНЕДЖМЕНТА ЗНАНИЙ

О. В. Шубкина

ХНУРЭ, г. Харьков, Украина

olga.shubkina@gmail.com

Проведен анализ задач, стоящих перед разработчиками систем менеджмента знаний, а также анализ существующих методов интеллектуального анализа текстов в системах управления знаниями. Описан метод концептуального аннотирования для формирования семантических профилей документов, на основе которого разработана архитектура системы.

УПРАВЛЕНИЕ ЗНАНИЯМИ, ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ТЕКСТОВ, КОНЦЕПТУАЛЬНОЕ АННОТИРОВАНИЕ, СЕМАНТИЧЕСКИЙ ПРОФИЛЬ ДОКУМЕНТА

Введение

На современном этапе социально-экономического развития человечества — этапе построения информационного общества — знание, его создание и управление им становится ключевым ресурсом развития мировой экономики. Деятельность как отдельных людей, так и организаций все в большей степени зависит от имеющихся у них знаний и способности эффективно их использовать.

Менеджмент или управление знаниями (Knowledge Management — КМ) сегодня рассматривается как мощное конкурентное преимущество. Однако ни информационные технологии (ИТ), ни данные сами по себе не могут обеспечить конкурентного преимущества на долгосрочный период. Они могут быть достигнуты только «переводом» информации в ценные, смысловые руководства к действию. Таким образом, знание состоит в действии: в эффективном представлении данных и информационных ресурсов для принятия решений, а также в самом выполнении принятого решения [1].

Ресурсы знаний различаются в зависимости от отраслей индустрии и приложений, но, как правило, представлены в текстовом виде. Документы представляют собой ресурс знаний организации, в котором находится около 85% всей информации [2, 3]. Таким образом, возрастает роль извлечения знаний из различного вида текстовых источников, накопленных на разных этапах развития организации. Этот процесс является основной задачей Text Mining (ТМ) — интеллектуального анализа текстовой информации [4]. В управлении знаниями ТМ играет важную роль, потому что является механизмом выявления закономерностей, характерных элементов или свойств, которые могут использоваться в качестве метаданных документа, ключевых слов, аннотаций. Другая важная задача ТМ состоит в отнесении документа к некоторым категориям из заданной схемы их систематизации. Таким образом, с помощью ТМ можно обратить в свою пользу всю имеющуюся у организации информацию,

опыт и квалификацию сотрудников с тем, чтобы повысить качество обслуживания клиентов и сократить время реакции на меняющиеся рыночные условия.

1. Постановка задач исследования

В рамках данной работы необходимо провести анализ задач, стоящих перед разработчиками систем КМ и возможности их решения с помощью интеллектуального анализа текстовой информации, предложить подход к алгоритмизации аннотирования текстовых документов на основе имеющейся концептуальной модели предметной области (ПрО). На основе анализа известных методов ТМ разработать метод концептуального аннотирования (КА) для создания профилей документов, а также сформировать обобщенную архитектуру системы КА на основе полученного алгоритма.

2. Основные задачи, решаемые с помощью систем КМ

Менеджмент знаний, несмотря на междисциплинарный характер данного направления исследований, в настоящее время наиболее активно развивается в искусственном интеллекте, предлагая прогрессивные решения, основанные на парадигме Semantic Web, по созданию систем корпоративной памяти предприятий, порталов знаний и другого [5]. К основным задачам КМ относят: достижение поставленных целей организации за счет роста интеллектуального капитала и эффективного его использования; повышение эффективности принимаемых решений; создание предпосылок для появления инноваций; повышение эффективности процессов проектирования. Основные задачи КМ решаются с помощью разработки корпоративных систем менеджмента или управления знаниями (СУЗ), реализуемых в настоящее время на основе парадигмы корпоративной памяти (КП). Основное предназначение КП состоит в накоплении, систематизации, управлении и совместном использовании профессиональными группами сотрудников всей необходимой и полезной информации в целях достижения конкурентного преимущества организации, эффективного и сво-

временного решения текущих задач, исключения дублирования, противоречивости и потери знаний, накопленных в процессе жизнедеятельности организации. Таким образом, наиболее важной задачей, стоящей перед разработчиками СУЗ, является разработка подхода к проектированию КП, от которой зависит эффективность работы системы в целом.

Среди интеллектуальных информационных технологий, используемых для разработки СУЗ, значительную роль играет ТМ, целью которого является извлечение полезной информации из текстовых источников путем идентификации и исследования полезных интересных шаблонов. Поскольку большая часть информации как в корпоративных Intranet-сетях, так и в WEB-пространстве содержится в текстовом виде, технологии интеллектуальной обработки текстов помогают решать многие задачи КМ, и прежде всего, построение пространства знаний (по сути, КП) путем формирования адекватной онтологии ПрО на основе извлечения знаний из предметно-ориентированных текстовых коллекций, их структурирования и анализа. Несмотря на то, что существуют классические известные подходы, модели и языки описания знаний ПрО, в настоящее время *de facto* наиболее эффективным стал онтологический подход к представлению знаний, развиваемый в рамках нового направления искусственного интеллекта – онтологического инжиниринга (Ontology Engineering - OE). Именно онтология позволяет получить эффективное представление эксплицитной концептуальной модели для сложноструктурированных и слабоформализованных ПрО в гетерогенных распределенных пространствах, к которым относится и КП предприятия.

3. Современные средства ТМ

В настоящее время наиболее востребованы системы ТМ с максимально автоматизированными ETL-процессами структурирования контента (Extract, Transfer, Load — извлечение, преобразование, загрузка). Важной характеристикой такого рода систем является функция оперативного анализа информации, полученной по запросу для выбора дальнейшего направления исследования документов (автопилотирование направления исследования), выполняемая с помощью методов ТМ.

В таких системах используется двухфазная технология аналитической обработки. В первой фазе (ETL) производится автоматизированный анализ отдельных документов, структуризация их контента и формирование хранилищ исходной и аналитической информации. Во второй фазе (OLAP, Text Mining, Data Mining) — извлечение в оперативном режиме знаний из хранилища или из полученной по запросу подборки документов. К наиболее интересным системам аналитической обработки относятся ClearForest, Convera

RetrievalWare, Hummingbird КМ, IBM Text Miner, инструменты компании IQMen, Inxight Smart Discovery Extraction Server, Ontos Miner, Oracle Text, ODB-Text, TextAnalyst, инструменты компании Smartware, XANALYS Link Explorer, X-Files [6].

К наиболее актуальным направлениям извлечения знаний из текста на сегодняшний день относятся: аналитическая обработка фактов; ведение досье; извлечение и структурирование фактографической информации; поиск информации по запросам на естественном языке с использованием тезаурусов; направление поиска информации, объектов в хранилище документов, в подборке документов; аннотирование документов, построение дайджестов по объектам; проведение тематического анализа документов (кластеризация и рубрицирование); построение и динамический анализ семантической структуры текстов; выделение ключевых тем и информационных объектов; определение общей и объектной тональности сообщений; исследование частотных характеристик текстов.

4. Аннотации как способ извлечения знаний из текстов

Обработка текстов подразумевает несколько последовательных этапов: на первом происходит нормализация слов с учетом морфологии языка; на втором — семантический анализ текста, когда уточняется конкретный смысл слова в зависимости от контекста. Затем строится семантический образ исходного документа, на основе которого делаются интеллектуальные запросы.

Классический подход к аннотированию документов предусматривает определение в тексте основных тем, позволяет выделить предложения, в которых тематика документа представлена наиболее ярко (представлены все основные тематические узлы). Данные предложения и образуют аннотацию текста.

Потребность в автоматическом составлении аннотаций для поисковых машин и каталогов довольно высока и со временем увеличивается. Держатели каналов вещания, списков рассылки новостей и корпоративных систем документооборота также начинают испытывать интерес к автоматизации этого процесса. Зачастую не только требование дать возможность быстрого просмотра содержания, но и желание не нарушать авторские права заставляет заменять полные тексты документов аннотациями. Пользователю также может пригодиться возможность быстро просматривать большие объемы документов, используя их краткие аннотации.

Следует отметить, что фактически во всех известных системах машинное аннотирование является экстрагированием — программа не «пересказывает» смысл текста, а просто извлекает из него те фрагменты, которые считает важными, и объединяет их в аннотацию. Важность конкретного предложения определяется по различным параметрам, в частности по так называемым маркерам важности (например, «в заклю-

чение нужно сказать, что...»), количеству содержательных слов в нем и так далее.

В наиболее развитых средствах аннотирования учитывается также зависимость предложений друг от друга с тем, чтобы не вносить в аннотацию обрывки, начинающиеся, например, со слов «К тому же...», «В-третьих...» и тому подобное. Чтобы аннотация получилась связной, программа подбирает группы взаимосвязанных (взаимозависимых) предложений, а затем «склеивает» их, для большей связности немного изменяя на стыках [7].

5. Концептуальное аннотирование

В данной работе предлагается подход ТМ к обработке текстовой информации в качестве построения так называемых семантических профилей документов для задач КА.

Онтологический подход к разработке такой системы позволит решать и комплекс прикладных задач, например, поиск экспертов в интересующей области знаний, выявление основных научных приоритетов отдельных ВУЗов, кафедр, преподавателей (создание семантических профилей). Под семантическим профилем документа будем понимать набор отобранных терминов с отсылкой документа к соответствующим понятиям, представленным в онтологической модели ПрО. Механизм КА предполагает связывание полученных терминов с концептами онтологии [8].

Примерами использования семантических профилей документов являются: построение профилей преподавателей с возможностью просмотра, например, основных публикаций, направлений научных исследований; поиск экспертов в заданной ПрО; выявление основных научных приоритетов отдельных ВУЗов, кафедр, преподавателей. Общий метод построения представлен на рис. 1.

Далее рассмотрим все шаги предложенного алгоритма подробнее.

Основной задачей лексического анализа является разбиение входного текста, состоящего из последовательности одиночных символов, на последовательность слов или лексем, то есть выделить эти слова из непрерывной последовательности. Все символы входной последовательности с этой точки зрения разделяются на символы, принадлежащие каким-либо лексемам, и символы, разделяющие лексемы (разделители) [9].

С точки зрения дальнейших фаз анализа лексический анализатор выдает информацию двух сортов: для синтаксического анализатора, работающего вслед за лексическим, существенна информация о последовательности классов лексем, ограничителей и ключевых слов, а для контекстного анализа, работающего вслед за синтаксическим, важна информация о конкретных значениях отдельных лексем (идентификаторов, чисел и так далее). В данной системе на этапе лексического анализа предложено использовать методологию

извлечения основных терминов, характеризующих специфическую область исследования.

После того как произведен анализ каждого слова, начинается анализ отдельных предложений (синтаксический анализ), позволяющий определить взаимосвязи между отдельными словами и частями предложения. Результатом такого анализа является граф, узлами которого выступают слова предложения; при этом, если два слова связаны каким-либо образом, то соответствующие им вершины графа связаны дугой с определенной окраской. Возможные окраски дуг зависят от языка, на котором написано предложение, а также от выбранного способа представления синтаксической структуры предложения.

Контекстный, или семантический, анализ текста базируется на результатах синтаксического анализа, получая на входе уже не набор слов, разбитых на предложения, а набор деревьев, отражающих синтаксическую структуру каждого предложения. Поскольку методы синтаксического анализа пока мало изучены, решения целого ряда задач семантической обработки текста базируются на результатах анализа отдельных слов, и вместо синтаксической структуры предложения, анализируются наборы стоящих рядом слов.

Семантический анализ, так или иначе, работает со смыслом слов. Следовательно, должна быть какая-то общая для всех методов анализа база, позволяющая выявлять семантические отношения между словами. Такой основой является тезаурус языка. На математическом уровне он представляет собой ориентированный граф, узлами которого являются слова в их основной словоформе. Дуги задают отношения между словами и могут иметь ряд окрасок.

В качестве одного из методов для кластеризации в данном случае целесообразно использовать *k*-кластеризацию. Это неиерархический алгоритм, кластеры представлены в виде центроидов, являющихся «центром массы» всех объектов, входящих в кластер, в которой каждый объект связан только с одной группой.

После указанных шагов система получает набор терминов, описывающих основные понятия документа. Под термином в данном случае понимаем слово или словосочетание, точно обозначающее понятие и его соотношение с другими понятиями в пределах исследуемой ПрО. В общем случае термины служат специализирующими, ограничительными обозначениями характерных для ПрО, явлений, их свойств и отношений. В идеале термин должен быть однозначным, систематичным, стилистически нейтральным.

На последнем шаге работы происходит связь концептов онтологии с полученным набором терминов. При этом предполагается наличие первоначальной версии обобщенной онтологии – концептуальной модели исследуемой ПрО, которая в процессе обработки коллекции текстов может редактироваться.

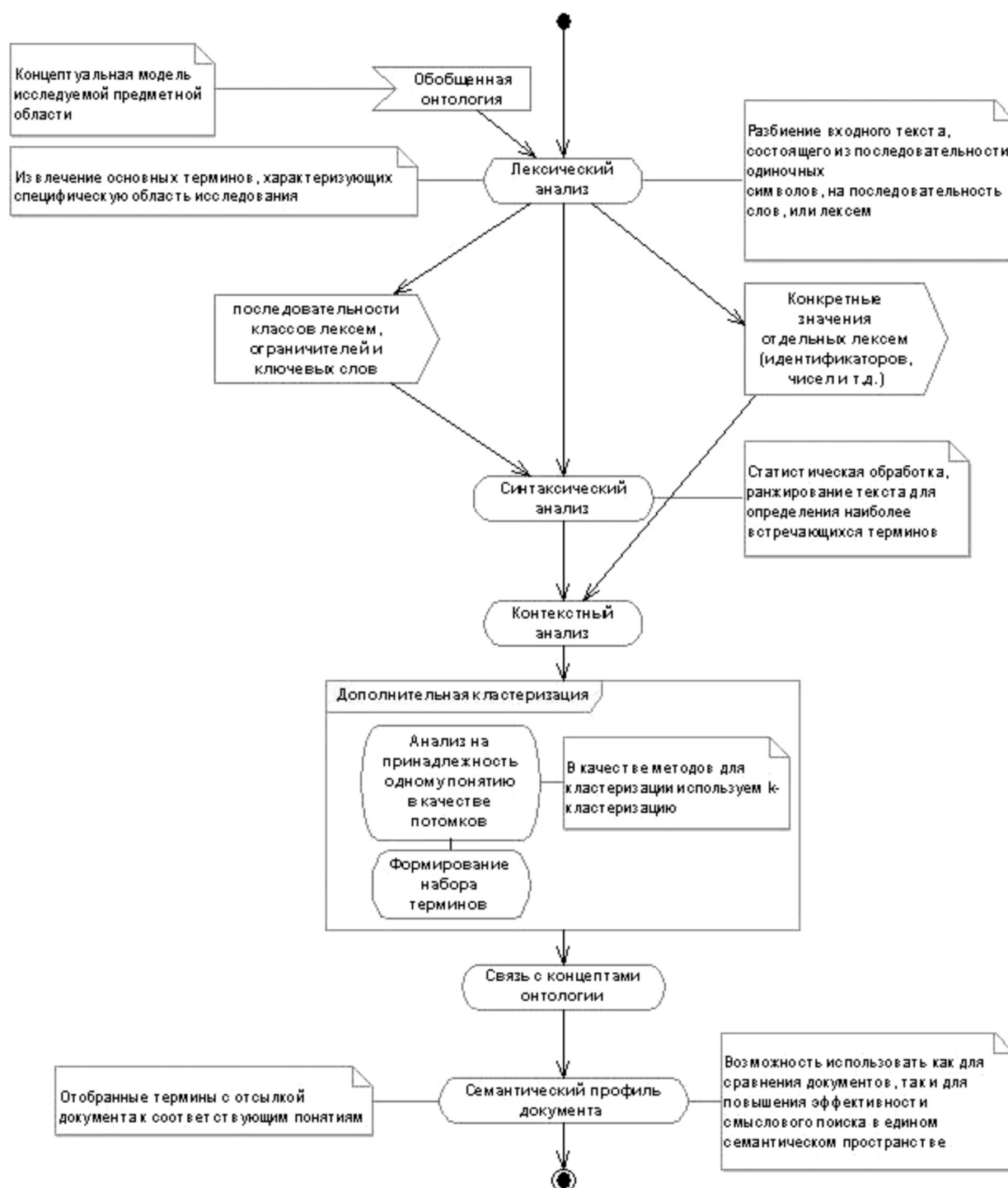


Рис. 1. Общая схема КА

В качестве инструментария создания и редактирования таких онтологий предлагается использование визуальной среды Protégé 3.1 со встроенным плагином PROMPT 2.4.8, который позволяет управлять различными версиями онтологии, осуществляя их слияние и выравнивание [10].

Согласно предложенному алгоритму разработана обобщенная архитектура системы, которая сможет

реализовать все описанные функции (рис. 2). Знания системы включают предметные, априорные, фактические (из текстов) и лингвистические. Система включает следующие взаимосвязанные модули: базу фактов, основанную на текстах ПрО, словарь терминов и словарь конечного набора терминов, формируемый в процессе работы системы, а также онтологию, описывающую взаимосвязь понятий в данной сфере.



Рис. 2. Обобщенная архитектура системы КА

Выводы

Проанализированы основные задачи, решаемые сегодня системами анализа текстов: формирование информационного портрета текста в терминах ключевых понятий; выявление смысловых связей между понятиями; автоматическое реферирование. Большинство существующих методов ТМ, основу которых составляют словари терминов, используют «плоские» тематические словари – набор ключевых слов, выделенных с помощью статистических методов обработки информации и сгруппированных в иерархическую структуру. Такие словари формируют статистический образ текста.

В работе предложен метод КА как перспективный, мало изученный подход в задачах анализа текстовой информации. Этот метод является основой для создания семантического профиля документа, который предполагает не только статистический анализ текста с определением наиболее часто встречающихся терминов, но и их дополнительную кластеризацию, и анализ на принадлежность одному понятию в качестве потомков.

В качестве понятий выступают концепты онтологии ПрО. Семантические профили в дальнейшем используются для сравнения документов, а также повышения эффективности смыслового поиска в едином семантическом пространстве.

На основе предложенного метода разработан алгоритм КА, подробно рассмотрены основные шаги для построения семантического профиля документа, а также его дальнейшего связывания с концептами онтологии ПрО. Кроме того, в данной работе была разработана архитектура системы КА.

Список литературы: 1. *Гладун А.Я., Рогушина Ю.В.* Онтологии в корпоративных системах // Корпоративные системы: журнал. – Киев: Изд. дом «Комиздат», 2006. – №1. 2. *Uren V., Cimiano Ph., Iria J., Handschuh S., Vargas-Vera M., Motta E., Ciravegna F.* Semantic annotation for knowledge management: Requirements and a survey of the state of the art // Web Semantics: Science, Services and Agents on the World Wide Web. – 2006. – Vol. 4, No. 1. – P. 14-28. 3. *Mertensson M.* A critical review of knowledge management as a management tool // Journal of Knowledge Management. – 2000. – V. 4, № 3. 4. *Feldman R.,*

Sanger J. The text mining handbook: advanced approaches in analyzing unstructured data. – Cambridge University Press, 2007. – 410 p. 5. *Рябова Н.В., Волошина Н.А., Шубкина О.В.* Применение методов text mining при решении задач менеджмента знаний // Научная сессия МИФИ-2008. Сб. науч. тр. В 15 томах. Т. 10. Интеллектуальные системы и технологии. – М.: МИФИ, 2008. – С.31-32. 6. *Ильин Н., Киселев С., Рябышкин В., Танков С.* Технологии извлечения знаний из текста // Открытые системы. СУБД. – 2006. – № 6. – С. 48-53. 7. *Ашманов И.* Информация и знания: невидимая грань. – <http://newasp.omskreg.ru/intellect/f5.htm>. 8. *Шубкина О.В.* Методы и модели построения семантических профилей документов // Материалы 12-ого Междунар. молод. форума «РАДИО-ЭЛЕКТРОНИКА И МОЛОДЕЖЬ В XXI веке». В 3 частях. Ч.2. – Харьков. – С. 173. 9. *Компаниец Р.И. и др.* Системное программирование. Основы построения трансляторов. – СПб.: КОРОНА принт, 2000. – 256 с. 10. www.protege.stanford.edu. 11. *Гаврилова Т.А., Хорошевский В.Ф.* Базы знаний интеллектуальных систем. – СПб: Питер, 2000. – 384 с. 12. *Luger G.* Artificial Intelligence: Structures and Strategies for Complex Problem Solving, 5th edition. – UK: Pearson Education, 2006. – 889 p. 13. *Tiwana A.* The knowledge management toolkit: orchestrating IT, strategy, and knowledge platforms. – 2nd ed. – Prentice Hall PTR, 2002. – 388 p. 14. *Feldman R., Sanger J.* The text mining handbook: advanced approaches in analyzing unstructured data. – Cambridge University Press, 2007. – 410 p.

Поступила в редколлегию 31.03.2009

УДК 519.52

Інтелектуальний аналіз текстів в системах менеджменту знань / О. В. Шубкіна // Біоніка інтелекту: наук.-техн. журнал. – 2009. – №1(70). – С. 142–146.

Розглядаються основні задачі систем аналізу текстової інформації. Запропонований метод концептуального анотування, визначені його головні етапи. Розглянута узагальнена архітектура системи концептуального анотування.

Лл.: 2. Бібліогр.: 14 найм.

UDK 519.52

Text mining in the knowledge management systems / O. V. Shubkina // Bionics of Intelligence: Sci. Mag. – 2009. – №1(70). – P. 142–146.

The main aims of the text mining systems are considered. Method of the concept annotation is offered, its basic stages are defined. Generic architecture of the concept annotation system is considered. Fig.: 2. Ref.: 14 items.

ОБ АВТОРАХ

Агарков Андрей Викторович	95	научный сотрудник отдела распознавания зрительных образов Донецкого института проблем искусственного интеллекта
Аксак Наталия Георгиевна	102	канд. техн. наук, старший научный сотрудник, доцент кафедры ЭВМ Харьковского национального университета радиоэлектроники
Асеев Георгий Георгиевич	28	д-р техн. наук, профессор, заведующий кафедрой информационных технологий Харьковской государственной академии культуры
Барковская Олеся Юрьевна	102	аспирант кафедры ЭВМ Харьковского национального университета радиоэлектроники
Бисикало Олег Владимирович	127	канд. техн. наук, доцент кафедры экономической кибернетики и информатики Винницкого государственного аграрного университета
Богданов Дмитрий Федорович	55	канд. техн. наук, доцент кафедры автоматического приборостроения национального технического университета «Харьковский политехнический институт»
Бодянский Евгений Владимирович	34	д-р техн. наук, профессор кафедры искусственного интеллекта Харьковского национального университета радиоэлектроники
Бондаренко Михаил Федорович	3, 11, 17	д-р техн. наук, профессор, ректор Харьковского национального университета радиоэлектроники
Вечирская Ирина Дмитриевна	55	канд. техн. наук, старший научный сотрудник кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники
Винарский Владимир Яковлевич	86	канд. техн. наук Университета прикладных наук и искусств, г. Ганновер, Германия
Волкова Валентина Владимировна	34	аспирант кафедры искусственного интеллекта Харьковского национального университета радиоэлектроники
Гассан Дибе	72	аспирант кафедры ЭВМ Харьковского национального университета радиоэлектроники
Ерохин Андрей Леонидович	39	д-р техн. наук, профессор, начальник кафедры информатики Харьковского национального университета внутренних дел
Иванов Валерий Геннадиевич	44	канд. техн. наук, доцент кафедры системотехники Харьковского национального университета радиоэлектроники
Кобылин Олег Анатольевич	90	канд. техн. наук, ассистент кафедры информатики Харьковского национального университета радиоэлектроники
Колчигин Богдан Владленович	34	студент Харьковского национального университета радиоэлектроники
Кольченко Алексей Вячеславович	39	аспирант кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники
Кораблев Николай Михайлович	50	канд. техн. наук, доцент кафедры ЭВМ Харьковского национального университета радиоэлектроники
Крыгин Максим Юрьевич	60	младший научный сотрудник Украинского языково-информационного фонда НАН Украины, г. Киев
Левыкин Виктор Макарович	77	д-р техн. наук, профессор, заведующий кафедрой информационных управляющих систем Харьковского национального университета радиоэлектроники
Любченко Валентин Анатольевич	90	доцент кафедры информатики Харьковского национального университета радиоэлектроники
Ляховец Алена Витальевна	137	аспирант кафедры программного обеспечения ЭВМ Харьковского национального университета радиоэлектроники
Машталер Сергей Владимирович	86	канд. техн. наук, доцент кафедры информатики Харьковского национального университета радиоэлектроники

Одейчук Андрей Николаевич	113	младший научный сотрудник научно-производственного комплекса «Возобновляемые источники энергии и ресурсосберегающие технологии» национального научного центра «Харьковский физико-технический институт»
Олейник Александр Игоревич	82	канд. техн. наук, старший научный сотрудник, докторант научно-исследовательской лаборатории математического и компьютерного моделирования Харьковского национального университета радиотехники
Олейник Андрей Александрович	107	аспирант кафедры программных средств Запорожского национального технического университета
Репка Виктория Борисовна	132	канд. техн. наук, доцент кафедры программного обеспечения ЭВМ Харьковского национального университета радиотехники
Сакало Евгений Сергеевич	86	аспирант кафедры информатики Харьковского национального университета радиотехники
Селезнев Сергей Николаевич	77	аспирант кафедры информационных управляющих систем Харьковского национального университета радиотехники
Ситник Лариса Григорьевна	64	преподаватель кафедры педагогики и инновационных технологий Сумского областного института последипломного педагогического образования
Субботин Сергей Александрович	107	канд. техн. наук, доцент кафедры программных средств Запорожского национального технического университета
Сорокина Ирина Витальевна	50	аспирант кафедры ЭВМ Харьковского национального университета радиотехники
Солдатов Антон Сергеевич	102	магистрант кафедры ЭВМ Харьковского национального университета радиотехники
Тыхун Алла Юрьевна	102	аспирант кафедры ЭВМ Харьковского национального университета радиотехники
Удовенко Сергей Григорьевич	72	д-р техн. наук, профессор кафедры ЭВМ Харьковского национального университета радиотехники
Федорова Татьяна Николаевна	55	стажер-исследователь кафедры программного обеспечения ЭВМ Харьковского национального университета радиотехники
Фесенко Улболган Абдулхамитовна	132	канд. мед. наук, анестезиолог, доцент кафедры детской челюстно-лицевой хирургии Харьковского государственного медицинского университета
Фролова Галина Александровна	120	аспирант Национального аэрокосмического университета им. Н.Е. Жуковского «Харьковский авиационный институт»
Хашан Татьяна Сергеевна	68	инженер 1 категории отдела теории управляющих систем Донецкого института прикладной математики и механики НАН Украины
Чалый Сергей Федорович	77	д-р техн. наук, профессор кафедры информационных управляющих систем Харьковского национального университета радиотехники
Четвериков Григорий Григорьевич	55	д-р техн. наук, профессор кафедры программного обеспечения ЭВМ Харьковского национального университета радиотехники
Шабанов-Кушнарченко Сергей Юрьевич	3, 11, 17	д-р техн. наук, профессор кафедры программного обеспечения ЭВМ Харьковского национального университета радиотехники
Шабанов-Кушнарченко Юрий Петрович	3, 11, 17	д-р техн. наук, профессор кафедры программного обеспечения ЭВМ Харьковского национального университета радиотехники
Шевченко Игорь Владимирович	44	студент кафедры системотехники Харьковского национального университета радиотехники
Щербинин Константин Сергеевич	86	аспирант кафедры информатики Харьковского национального университета радиотехники
Шубкина Ольга Васильевна	142	аспирант кафедры искусственного интеллекта Харьковского национального университета радиотехники

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 (210x297 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. 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).

Формулы, символы, переменные, которые нельзя набрать в Microsoft Word, 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.

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, напечатанный на лазерном принтере, в 2-х экземплярах) необходимо подать следующие документы: заявку утвержденного образца, подписанную всеми авторами; акт экспертизы о возможности опубликования материалов в открытой печати; рецензию; сведения об авторах; электронный вариант рукописи, реферата и сведений об авторах; оплату за публикацию.

Редакция журнала «Бионика интеллекта» обращает внимание авторов на необходимость составления списка литературы в соответствии с новым ДСТУ ГОСТ 7.1:2006.

Примеры библиографического описания

Книга одного автора

Андреев, В. В. Как организовать делопроизводство на предприятии [Текст] / В. В. Андреев. – М. : ИНФРА-М, 1997. – 94 с.

Книга двух авторов

Белов, А. В. Финансы и кредит [Текст] : учеб. / А. В. Белов, В. Н. Николаев ; КНУ им. Т. Г. Шевченко. – К. : Университет, 2004. – 215 с. – Библиогр.: с. 213–215. – ISBN 5-7042-1441-X.

Гроссе, Э. Химия для любознательных [Текст] : пер. с нем. / Э. Гроссе, Х. Вайсменталь. – М.: Химия, 1980. – 392 с.

Книга трех авторов

Агафонова, Н. Н. Гражданское право [Текст] : учеб. пособие / Н. Н. Агафонова, Т. В. Богачева, Л. И. Глушкова; под общ. ред. А. Г. Калпина; МОН Украины. – 2-е изд., перераб. и доп. – Х. : Фактор, 2000. – 542 с. – (Университетская книга).

Бахвалов, Н. С. Численные методы [Текст] : учеб. пособие для физ.-мат. специальностей вузов / Н. С. Бахвалов, Н. П. Жидков, Г. М. Кобельков ; под общ. ред. Н. И. Тихонова. — 2-е изд. — М.: Физматлит : Лаб. базовых знаний; СПб. : Нев. диалект, 2002. — 630 с. : ил. ; 25 см. — (Технический университет. Математика). — Библиогр.: с. 622–626. — Предм. указ.: с. 627–630. — 30000 экз. — ISBN 5-93208-043-4 (в пер.).

Книга четырех авторов

Элементи інформатики [Текст] : довідник / В. С. Височанський, А. І. Кардаш, В. С. Костев, В. В. Черняхівський. — К. : Наук. думка, 2003. — 192 с.

Книга пяти авторов и более

Коротковолновые антенны [Текст] : учеб. пособие / Г. З. Айзенберг, С. П. Белоусов, Я. М. Журбин и др. ; под общ. ред. А. А. Стогния. — 2-е изд. — М. : Радио и связь, 2003. — 192 с.

Переводное издание

Нойман, Э. Происхождение и развитие сознания [Текст] : пер. с англ. — К. : Ваклер ; М. : Реал-бук, 1998. — 462 с.

Книги под заглавием

Информационные технологии в маркетинге [Текст] : учеб. / под ред. Г. А. Титаренко. — М. : ЮНИТИ, 2000. — 335 с. — (Textbook). — ISBN 5-238-00154-1.

Многотомное издание, отдельный том

Савельев, И. В. Курс общей физики [Текст]. Т. 1. Механика. Молекулярная физика : учеб. Пособие / И. В. Савельев. — 2-е изд., перераб. — М.: Наука, 1982. — 432 с.

Статьи из журналов

Гончаров, В. А. Численная схема моделирования дозвуковых течений вязкого сжимаемого газа [Текст] / В. А. Гончаров, В. М. Кравцов // Журн. вычисл. математики и мат. физики. — 1988. — Т. 28, №12. — С. 1858–1866.

Сборники научных трудов

Отчет о выполнении плана научно-исследовательских работ за 2003 год [Текст] : сб. науч. тр. / Рос. Акад. мед. наук, Сиб. отд. — Новосибирск : СО РАМН, 2004. — 83 с.

Тезисы конференций

Образование, наука, производство: пути углубления интеграции и повышения качества инженерного образования [Текст] : тез. докл. науч.-практ. конф. (окт. 2000) / отв. ред. В. Г. Вдовенко. — Красноярск : САА, 2000. — 53 с.

Материалы конференций

Проблемы экономики, организации и управления реструктуризацией и развитием предприятий промышленности, сферы услуг и коммунального хозяйства [Текст] : материалы IV междунар. науч.-практ. конф., 30 марта 2005 г. Новочеркасск / редкол. : Б. Ю. Серебряков (отв. ред.). — Новочеркасск : Темп, 2005. — 58 с.

Стандарты, технико-экономические и технические документы

ГОСТ Р 517721-2001. Аппаратура радиоэлектронная бытовая. Входные и выходные параметры и типы соединений. Технические требования [Текст]. — Введ. 2002-01-01. — М. : Изд-во стандартов, 2001. — 27 с.

Инструкция по проектированию, строительству и эксплуатации гидротехнических сооружений на подра-

батываемых горными работами территориях [Текст] : СН 522-85. — Утв. Госстроем СССР 03.05.86. — Изд. офиц. — М.: Стройиздат, 1986. — 32 с.

Авторские свидетельства, патенты

А. с. 1007970 СССР, МПК В 25 J 15/00. Устройство для захвата деталей [Текст] / Ваулин В. С., Калов В. К. (СССР). — 3360585/25-08 ; заявл. 23.11.81 ; опубл. 30.03.83, Бюл. № 12. — 2 с.

Приемопередающее устройство [Текст] : пат. 2187888 Рос. Федерация: МПК7 Н 04 В 1/38, Н 04 J 13/00 / Чугаева В. И. ; заявитель и патентообладатель Воронеж. науч.-исслед. ин-т связи. — № 2000131736/09 ; заявл. 18.12.00; опубл. 20.08.02, Бюл. № 32. — 3 с.

Отчеты о НИР

Оценка эффективности автоматизированных ИПС [Текст]: отчет о НИР (промежуточ.) / ВНИИЦентр ; рук. М. И. Иванов. — ОЦ 02604-И5В ; ГР 01821100006 ; Инв. Б452743. — М., 1982. — 90 с.

Диссертации

Антопольский, А. Б. Лингвистическое описание и оценка информационных языков [Текст] : дис. ... канд. филол. наук / А. Б. Антопольский. — М., 1969. — 404 с.

Белозеров, И. В. Религиозная политика Золотой Орды на Руси в XIII-XIV вв. [Текст] : дис. ... канд. ист. наук : 07.00.02 : защищена 22.01.02 : утв. 15.07.02 / Белозеров Иван Валентинович. — М., 2002. — 215 с. — Библиогр.: с. 202-213. — 04200201565.

Авторефераты диссертаций

Бутковский, О. Я. Обратные задачи хаотичной динамики и проблемы предсказуемости хаотичных процессов [Текст] : автореф. дис. ... д-ра физ.-мат. наук : 01.04.03 / О. Я. Бутковский ; [Ин-т радиотехники и электроники РАН]. — М., 2004. — 39 с.

Электронные ресурсы

Удаленного доступа

Основные направления исследований, основанные на семантическом анализе текстов [Электронный ресурс] / С.-Петерб. гос. ун-т, фак. прикладной математики — процессов управления. — Режим доступа : <http://apcp.apmath.spbu.ru/ru/staff/tuzov/onapr.html/> — 10.12.2004 г. — Загл. с экрана.

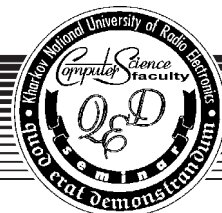
Локального доступа

Internet шаг за шагом [Электронный ресурс] : интеракт. учеб. — Электрон. дан. и прогр. — СПб. : Питер Ком, 1997. — 1 электрон. опт. диск (CD-ROM). — Систем. требования: ПК от 486 DX 66 МГц; RAM 1616 Мб; Windows 95; зв. плата. — Загл. с этикетки диска.

Художественная энциклопедия зарубежного классического искусства [Электронный ресурс]. — Электрон. текстовые, граф., зв. дан. и прикладная прогр. (546 Мб). — М. : Большая Рос. энцикл. [и др.], 1996. — 1 электрон. опт. диск (CD-ROM) : зв., цв. ; 12 см + рук. пользователя (1 л.) + открытка (1 л.). — (Интерактивный мир). — Систем. требования: ПК 486 или выше; 8 Мб ОЗУ; Windows 3.1 или Windows 95; SVGA 32768 и более цв.; 640x480; 4x CD-ROM дисковод; 16-бит. зв. карта; мыш. — Загл. с экрана. — Диск и сопровод. материал помещены в контейнер 20x14 см.

ИНФОРМАЦИОННОЕ СООБЩЕНИЕ

Факультет компьютерных наук ХНУРЭ
Ежемесячный научный семинар Q.E.D.



Если в рамках научных направлений факультета КН Вы

хотите обсудить в «полновесном» формате (до полутора часов):

- «Новый результат в ...»
- «Состояние и перспективы развития ...»
- «Сравнительный анализ подходов (методов...) ...»

сообщить в формате «лягушатника» (до 20 минут):

- обоснование темы диссертации
- цели и задачи диссертационных исследований
- теоретическую значимость и практическую ценность результатов
- полученные теоретические результаты и их новизну
- представить монографию, статью, доклад ...

приглашаем принять участие в работе семинара Q.E.D.

Постоянные члены семинара:

проф. Авраменко В.П.
проф. Бескорвайный В.В.
проф. Бодянский Е.В.
проф. Бондаренко М.Ф.
проф. Борячок М.Д.
проф. Гребенник И.В.
проф. Дударь З.В.
проф. Кучеренко Е.И.
проф. Левыкин В.М.
проф. Лесная Н.С.
проф. Машгалир В.П. (зам. председателя)
проф. Петров Э.Г. (председатель)
доц. Рябова Н.В.
проф. Филатов В.А.
доц. Чалая Л.Э. (ученый секретарь)
проф. Четвериков Г.Г.
проф. Шабанов-Кушнаренко Ю.П.

*Семинар проводится
каждый второй четверг
в 15⁰⁵, ауд. 259*

СЛЕДИТЕ ЗА ОБЪЯВЛЕНИЯМИ!

Запись для выступлений на семинаре
у секретаря (комн. 370)

ПРИГЛАШАЮТСЯ ВСЕ ЖЕЛАЮЩИЕ

Харьковский национальный университет
радиоэлектроники, г. Харьков, просп. Ленина, 14

Наукове видання

БІОНІКА ІНТЕЛЛЕКТУ
інформація, мова, інтелект

Науково-технічний журнал

№ 1 (70)
2009

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. Г. Ристенко*

Рекомендовано Вченою Радою
Харківського національного університету радіоелектроніки
(протокол № 52 від 24.04.2009)

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 24.04.2009. Формат 60 × 84 ¹/₈. Друк ризографічний.
Папір офсетний. Гарнітура Newton. Умов. друк. арк. 16,9. Обл.-вид. арк. 19,8.
Тираж 100 прим. Зам. № .

Надруковано в навчально-науковому видавничо-поліграфічному центрі ХНУРЕ
61166, Харків-166, просп. Леніна, 14