О. М. Васильев, Д. П. Ветров, Д. А. Кропотов, Представление и обнаружение знаний в экспертных системах для задач распознавания образов, *Ж. вычисл. матем. и матем. физ.*, 2007, том 47, номер 8, 1428–1454

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2007, том 47, № 8, с. 1428–1454

УДК 519.6:519.710.24

ПРЕДСТАВЛЕНИЕ И ОБНАРУЖЕНИЕ ЗНАНИЙ В ЭКСПЕРТНЫХ СИСТЕМАХ ДЛЯ ЗАДАЧ РАСПОЗНАВАНИЯ ОБРАЗОВ[1)]

© 2007 г. О. М. Васильев, Д. П. Ветров, Д. А. Кропотов

(119992 Москва, Ленинские Горы, МГУ, ВМиК)

e-mail: ovasiliev@inbox.ru; vetrovd@yandex.ru; dkropotov@yandex.ru

Поступила в редакцию 01.09.2006 г.
Переработанный вариант 21.02.2007 г.

Предложен новый подход к проектированию нечетких экспертных систем. Подробно рассмотрены вопросы представления знаний и формирования высказываний средствами нечеткой логики, а также описана модель нечетких рассуждений. Основное внимание уделено вопросам автоматического получения знаний (нечетких правил вывода) по множеству прецедентов. В частности, введены различные критерии качества правил и предложен алгоритм их генерации (метод эффективных сужений). Описаны возможности расширения вида допустимых правил путем введения операции нечеткой дизъюнкции. Также исследованы возможности последующей оптимизации найденных правил. Представлены результаты экспериментов, демонстрирующие ценность предлагаемых подходов. Библ. 47. Фиг. 10. Табл. 2.

Ключевые слова: распознавание образов, поиск закономерностей в данных, нечеткая логика, экспертные системы.

ВВЕДЕНИЕ

Теория распознавания образов оперирует с задачами, связанными с предсказанием набора зависимых переменных по множеству независимых переменных, доступных для измерений и оценок, в ситуациях, когда известна некоторая эмпирическая информация об исследуемой взаимосвязи (см. [1, 2]). Для решения подобных задач разработано большое количество методов (см. [3]). Большинство из них (например, нейтронные сети, см. [4]) используют концепцию "черного ящика", т.е. просто выдают результат предсказания для данного набора признаков. Таким образом, связь между переменными внутри процесса остается неясной. Хотя для ряда задач одного прогноза оказывается достаточно, но зачастую возникают ситуации, когда необходимы дополнительные знания об исследуемом процессе (см. [5]). Последнее становится особенно актуальным в тех случаях, когда значения независимых переменных могут быть модифицированы исследователем. Тогда приходят к так называемой задаче data mining. После извлечения необходимых знаний можно вернуться к прогнозу зависимых переменных, принимая во внимание полученную информацию. В широком поле прикладных задач подобную дополнительную информацию получают при помощи экспертов данной прикладной области. В связи с этим необходимо, чтобы знания, извлекаемые автоматически при помощи компьютера в процессе data mining, были сформулированы в тех же терминах, что и информация, получаемая от экспертов. Таким образом, эксперт будет в состоянии оценить, добавить или отвергнуть часть полученной дополнительной информации. Системы подобного рода, в которых важную роль играют экспертные дополнительные знания об исследуемой области, получили название экспертных систем.

Исследования в области экспертных систем берут свое начало с середины 50-х годов прошлого столетия и уже пережили несколько этапов развития. За это время был реализован ряд успешных проектов по созданию экспертных систем, работающих на практике (см. [6, 7]). Однако большинство из них следует отнести к области нечеткого логического управления. Дело в том, что в задачах управления база знаний системы (совокупность управляющих правил) зачастую состоит из небольшого количества правил, непосредственно подсказываемых предметной областью. Поэтому основной проблемой при разработке таких систем становится реализация нечетких рассуждений или выводов. В этой области существует большое количество наработок,

[1)] Работа выполнена при финансовой поддержке РФФИ (коды проектов 05-07-90333, 06-01-00492, 06-01-08045, 07-01-00211), ИНТАС (YS 04-83-2942, 04-77-7036) и программы ОМН РАН № 02.

которые приводят к успешным результатам (см. [8–10]). При построении экспертных систем, ориентированных на решение задач распознавания и прогнозирования, проблема построения базы данных (представления знаний в системе), организация базы знаний (получение знаний от экспертов предметной области) и получение модели вычисления прогноза играют ключевую роль. Необходимость решения комплекса сложных проблем приводит к ограниченному распространению экспертных систем распознавания и прогноза на сегодняшний день.

В области экспертных систем выделяют три основные проблемы:

1) представление знаний;

2) использование знаний;

3) приобретение знаний.

Наиболее естественной формой **представления знаний** является совокупность утверждений типа "ЕСЛИ..., ТО..." (см. [9]). При этом условия в посылке правила (антецеденте), а также в заключении правила (выводе) должны формулироваться достаточно просто, чтобы эксперт мог относительно легко проинтерпретировать то или иное правило. Для задач распознавания образов в качестве условий, входящих в правила, традиционно выбираются набор отрезков (значение признака лежит в определенном диапазоне $a \le x \le b$, см. [11, 12]) или совокупность нечетких подмножеств значений признаков (см. [13]). Нечеткие подмножества предоставляют возможность задавать правила на естественном языке, т.е. в форме, наиболее приближенной к представлению знаний эксперта. В настоящей статье исследуются различные способы представления знаний в виде совокупности нечетких подмножеств, рассматриваются вопросы генерации и оптимизации представления.

Использование знаний для прогнозирования, как правило, представляет собой различные схемы голосования по набору закономерностей (правил, см. [3, 11]). В случае представления знаний с помощью нечетких подмножеств здесь подключается развитый аппарат нечеткой логики. При этом в конечном итоге механизм нечеткого вывода также может быть проинтерпретирован с точки зрения схемы голосования (см. [6]). В статье рассматриваются различные способы голосования и настройки весов правил.

Наиболее трудоемкой является задача **приобретения знаний**. Для решения этой проблемы требуются как талант исследователя находить общий язык с экспертами, так и специальные автоматические алгоритмы извлечения знаний из эмпирической информации. Алгоритмы автоматической генерации знаний, как правило, связаны с решением многопараметрических оптимизационных задач, в которых целевая функция обладает большим количеством локальных оптимумов (см. [11]). Более того, в случае экспертных систем возникает ряд дополнительных требований к множеству искомых закономерностей. Среди них можно отметить следующие:

а) относительно небольшое количество информативных закономерностей, для того чтобы эксперт мог просмотреть, оценить и/или изменить набор правил;

б) небольшое количество условий в посылках правил;

в) правила должны быть сформулированы относительно *исходных* признаков, выражения типа $\sqrt[3]{\text{Вес}}$ или sin(Рост) являются недопустимыми.

Для поиска набора информативных закономерностей в экспертных системах обычно используют алгоритмы на базе нечетких нейтронных сетей (neuro-fuzzy approach, см. [14, 15]), генетических алгоритмов (см. [16], [17], 13]), а также гибридные подходы с использованием кластерного анализа (см. [18]). К сожалению, эти подходы обладают рядом недостатков:

– нечеткие нейронные сети (ННС) генерируют огромное количество правил относительно низкой информативности, что соответствует ситуации переобучения, а также не позволяет интерпретировать и редактировать набор правил;

– высокий порядок генерируемых правил для ННС;

– большая зависимость ННС от начального приближения;

– значительное время обучения ННС и генетических алгоритмов;

– необходимость определения числа кластеров в алгоритмах, использующих кластеризацию;

– большое количество параметров, настраиваемых пользователем, в генетических алгоритмах.

Можно отметить также такие алгоритмы генерации знаний, как решающие списки (см. [19]) и подходы с использованием процедуры адаптивной коррекции или бустинга (boosting, см. [20–23]). В этих алгоритмах применяется последовательная схема использования правил, что на практике значительно затрудняет или делает невозможной интерпретацию решений для эксперта.

Большое количество исследований посвящено алгоритмам с использованием решающих деревьев (см. [5, 24–26]). Эти алгоритмы обладают высокими точностными показателями и часто используются на практике. Однако интерпретация древовидной структуры в виде набора правил часто приводит к большому количеству "длинных" правил, посылки которых очень похожи друг на друга (см. [24]). Это также затрудняет интерпретацию решения.

В данной статье предлагается алгоритм генерации знаний, лишенный большинства упомянутых недостатков. Он строит небольшой набор компактных информативных правил, которые могут быть успешно использованы при решении задач.

Работа состоит из пяти разделов. В разд. 1 рассматриваются способы представления знаний в виде совокупности нечетких подмножеств. В разд. 2 рассматриваются вопросы использования знаний, описывается модель получения прогноза в случае нечетких закономерностей. Разд. 3 посвящен проблеме генерации знаний. В нем предлагаются автоматические алгоритмы поиска закономерностей, а также исследуются свойства алгоритмов. Разд. 4 посвящен вопросам оптимизации модели, поиска начального представления, настройки весов правил, поиска границ множеств в посылках правил и пр. В разд. 5 приводятся результаты экспериментов по применению разрабатываемых технологий для решения реальных прикладных задач, подводятся итоги и делаются выводы.

В дальнейшем будут использоваться следующие обозначения. Предположим, что имеется d действительных признаков (независимых переменных) и одна зависимая переменная, принимающая значения на конечном множестве $\{1, 2, \ldots, l, \Delta\}$ для задачи распознавания образов (классификации) с l классами или на множестве действительных чисел для задачи восстановления регрессии (прогнозирования). Прецедентная (эмпирическая) информация (обучающая выборка) представляет собой совокупность пар $\{\mathbf{x}^i, y^i\}_{i=1}^q$, где $\mathbf{x}^i \in \mathbb{R}^d$, $y^i \in \{1, 2, \ldots, l\}$, или $y^i \in \mathbb{R}$.

1. ПРЕДСТАВЛЕНИЕ ЗНАНИЙ

Особенностью экспертных систем является представление знаний в легко интерпретируемых терминах. Как уже отмечалось выше, наиболее доступной формой представления знаний являются выражения, сформулированные в лингвистическом виде. Такие знания формулируются на естественном и, следовательно, нечетком языке. В работе [9] предлагается формулировать знания в форме продукционных правил вида "ЕСЛИ..., ТО...", при этом посылка правила представляет собой логическое выражение относительно нечетких подмножеств для признаков.

Определение 1. Пусть $\mathcal{I}$ – произвольное множество, называемое *полным пространством*. *Нечетким подмножеством* M полного пространства $\mathcal{I}$ называется множество упорядоченных пар

$$\{(x, \mu_M(x)) \mid x \in \mathcal{I}, \mu_M : \mathcal{I} \longrightarrow [0, 1]\}.$$

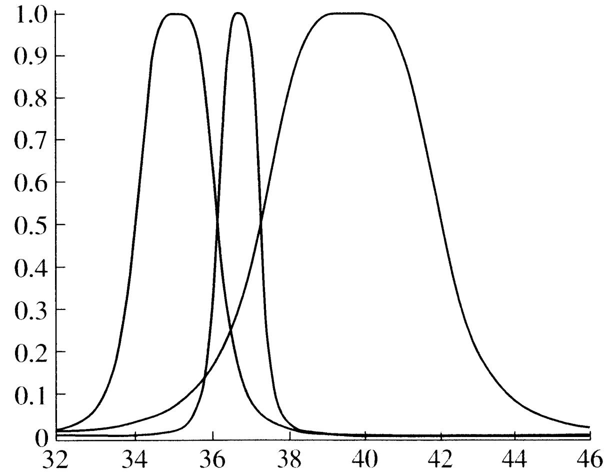
При этом, отображение $\mu_M(\cdot)$ называется *характеристической функцией* нечеткого подмножества M.

Рассмотрим произвольный числовой признак (один из тех, относительно которых сформулирована прецедентная информация). С точки зрения эксперта, область его изменения может быть описана вполне упорядоченным семейством подмножеств, каждому из которых соответствует некоторое лингвистическое значение, сформулированное на естественном языке. В качестве примера на фиг. 1 изображены характеристические функции трех нечетких подмножеств признака "температура тела пациента". Соответствующие им лингвистические значения (слева направо): "НИЗКАЯ", "НОРМАЛЬНАЯ", "ВЫСОКАЯ".

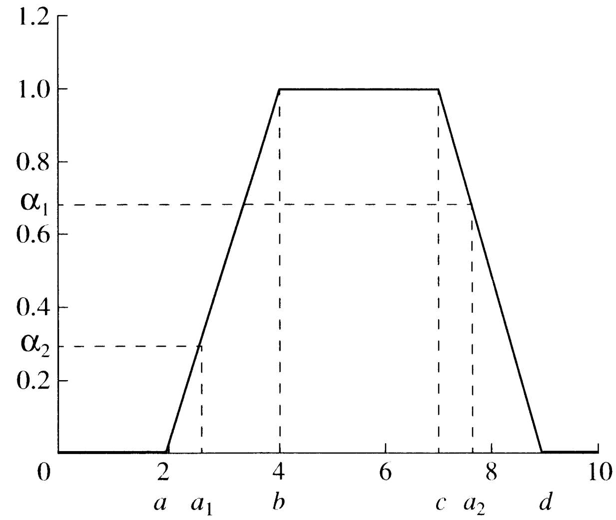
В общем случае можно предположить, что с точки зрения эксперта существует некоторое разбиение области значений признака, определяющее условные границы между различными состояниями. Это положение отражает следующее

Определение 2. *Экспертной интерпретацией признака* $i \in \{1, 2, \ldots, d\}$ для разбиения области его значений $a_i^1 < \ldots < a_i^{n_i}$ будем называть множество всех нечетких подмножеств числовой прямой с условными границами в соседствующих точках заданного разбиения:

$$\mathfrak{T}_i = \left\{M_i^j, j = \overline{1, n_i - 1} \mid \mu_{M_i^j}(\cdot) = \mu(x; a_i^j, a_i^{j+1})\ \forall x \in \mathbb{R}\right\}.$$



Фиг. 1.



Фиг. 2.

При этом

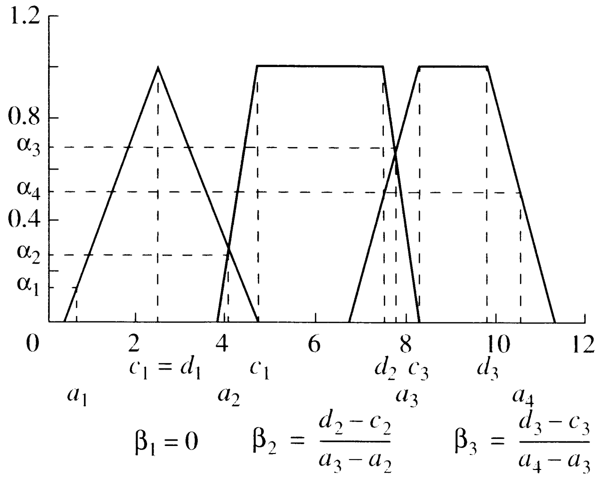
$$\forall j = \overline{1, n_i - 1} \quad \exists x_* \in [a_i^j, a_i^{j+1}] : \mu_{M_i^j}(x_*; a_i^j, a_i^{j+1}) = \max_x [\mu_{M_i^j}(x; a_i^j, a_i^{j+1})].$$

Здесь выражение $\mu(x; a, b)$ обозначает характеристическую функцию нечеткого подмножества с условными границами a и b. Вид характеристической функции $\mu(\cdot)$ можно выбирать поразному. В дальнейшем будут рассматриваться семейства трапециевидных и колоколообразных функций. Связь между функцией нечеткого подмножества и условными границами будет уточнена ниже.

1.1. Семейство трапециевидных функций

В качестве базовой формы нечеткого подмножества можно рассматривать равнобедренную трапецию (фиг. 2). В частном случае она может иметь форму прямоугольника (если $a = b$, $c = d$), т.е. задавать четкое множество, а также треугольника (если $b = c$).

Определение 3. $\{\alpha, \beta\}$-*покрытием* признака для разбиения области его значений $a_1 < \ldots < a_n$ будем называть совокупность нечетких подмножеств $\{M_i\}_{i=1}^{n-1}$ с характеристическими трапецие-



Фиг. 3.

видными функциями (фиг. 2) такую, что верно следующее:

$$\alpha_1, \ldots, \alpha_n \in (0, 1), \quad \beta_1, \ldots, \beta_{n-1} \in [0, 1), \tag{1.1}$$

$$\mu_{M_i}(a_{i+1}) = \mu_{M_{i+1}}(a_{i+1}) = \alpha_{i+1} \quad \forall i = \overline{1, n-2}, \tag{1.2}$$

$$\mu_{M_1}(a_1) = \alpha_1, \quad \mu_{M_{n-1}}(a_n) = \alpha_n, \tag{1.3}$$

$$\exists x_* \in [a_i, a_{i+1}] : \mu_{M_i}(x_*) = \max_x[\mu_{M_i}(x)] \quad \forall i \in \overline{1, n-1}, \tag{1.4}$$

$$\frac{\left|\{x \in \mathbb{R} \mid \mu_{M_i}(x) = 1\}\right|}{a_{i+1} - a_i} = \beta_i \quad \forall i \in \overline{1, n-1}. \tag{1.5}$$

Понятие $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$-покрытия проиллюстрировано на фиг. 3.

Справедлива следующая

Теорема 1. $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$*-покрытие для разбиения* $a_1 < \ldots < a_n$ *определено однозначно.*

Доказательство. Трапеция определяется однозначно посредством четырех параметров – точек излома трапеции: a, b, c, d (фиг. 2). Рассмотрим $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$-покрытие для заданного разбиения $\{a_i\}_{i=1}^n$. Требование (1.4) в определении покрытия означает, что для нечеткого подмножества M_i вершина трапеции располагается внутри интервала $[a_i, a_{i+1}]$. Отсюда, учитывая требования (1.1) и (1.2) в определении покрытия, можно получить, что $a_i \le b \le c \le a_{i+1}$, где b, c – соответствующие параметры характеристической функции для M_i.

Для простоты рассмотрим отдельно интервал разбиения (обозначим его через $[a_1, a_2]$) и соответствующее ему нечеткое подмножество – трапецию. Левая сторона трапеции представляет собой прямую линию, проходящую через три точки соответственно: $(0, a)$, (α_1, a_1), $(1, b)$ (фиг. 2). Это условие дает следующую систему уравнений:

$$0 = ka + b_1,$$

$$\alpha_1 = ka_1 + b_1,$$

$$1 = kb + b_1.$$

Здесь коэффициенты k, b_1 – угол наклона и параметр сдвига прямой.

Аналогично, правая сторона трапеции есть прямая линия, проходящая через точки $(0, d)$, (α_2, a_2), $(1, c)$ (фиг. 2). Ввиду равнобедренности трапеции угол наклона прямой $k_1 = -k$. Таким образом, получаем следующую систему:

$$0 = -kd + b_2,$$

$$\alpha_2 = -ka_2 + b_2,$$
$$1 = -kc + b_2.$$

Здесь b_2 – параметр сдвига прямой.

Условие (1.5) в определении покрытия приводит к равенству

$$c - b = \beta(a_2 - a_1).$$

Объединяя все полученные соотношения в единую систему, получаем

$$\begin{aligned} 0 &= ka + b_1,\\ \alpha_1 &= ka_1 + b_1,\\ 1 &= kb + b_1,\\ 0 &= -kd + b_2,\\ \alpha_2 &= -ka_2 + b_2,\\ 1 &= -kc + b_2,\\ c - b &= \beta(a_2 - a_1). \end{aligned}$$

Данная система представляет собой систему из семи линейных уравнений относительно семи неизвестных с невырожденным определителем и приводит к однозначному решению для a, b, c, d:

$$a = a_1 - \frac{\alpha_1(1-\beta)(a_2 - a_1)}{2 - \alpha_1 - \alpha_2},$$

$$b = a_1 + \frac{(1-\alpha_1)(1-\beta)(a_2 - a_1)}{2 - \alpha_1 - \alpha_2},$$

$$c = a_2 - \frac{(1-\alpha_2)(1-\beta)(a_2 - a_1)}{2 - \alpha_1 - \alpha_2},$$

$$d = a_2 + \frac{\alpha_2(1-\beta)(a_2 - a_1)}{2 - \alpha_1 - \alpha_2}.$$

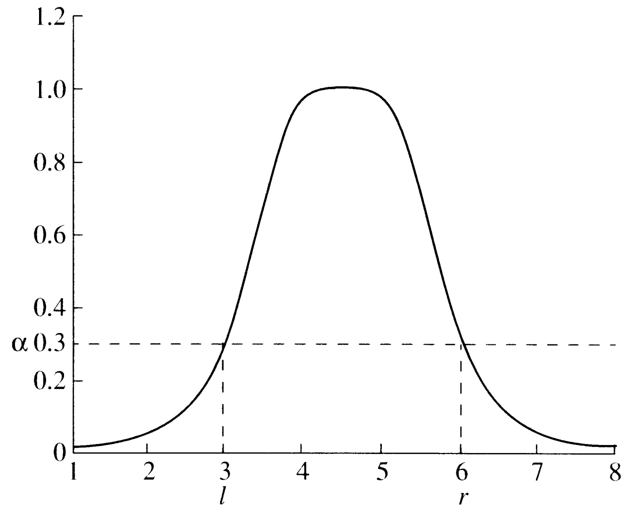
Теорема доказана.

Таким образом, для определения экспертной интерпретации для заданного разбиения можно воспользоваться понятием $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$-покрытия. Для этого необходимо определить $2n - 1$ параметров, где n – число элементов разбиения. Как было указано выше, на практике элементы разбиения представляют собой примерные границы между различными состояниями, задаваемыми экспертом, а введение нечетких подмножеств отражает степень уверенности эксперта относительно выбранных границ. Поэтому можно предположить, что коэффициенты $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ отражают особенности знания эксперта о значениях признака в целом, а не об отдельных нечетких подмножествах. Это означает, что можно положить $\alpha_1 = \ldots = \alpha_n = \alpha$, $\beta_1 = \ldots = \beta_{n-1} = \beta$. Таким образом, пространство оптимизируемых параметров существенно сокращается.

Характеристическая функция нечетких подмножеств $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$-покрытия имеет следующий вид:

$$\mu_l(x; l, r, \alpha, \beta) = \begin{cases} 0, & x < l - \dfrac{\alpha(r-l)(1-\beta)}{2(1-\alpha)},\\[2mm] 2\dfrac{(x-l)(1-\alpha)}{(r-l)(1-\beta)} + \alpha, & l - \dfrac{\alpha(r-l)(1-\beta)}{2(1-\alpha)} \le x < l + \dfrac{(1-\beta)(r-l)}{2},\\[2mm] 1, & l + \dfrac{(1-\beta)(r-l)}{2} \le x < r - \dfrac{(1-\beta)(r-l)}{2},\\[2mm] 2\dfrac{(r-x)(1-\alpha)}{(r-l)(1-\beta)} + \alpha, & r - \dfrac{(1-\beta)(r-l)}{2} \le x < r + \dfrac{\alpha(r-l)(1-\beta)}{2(1-\alpha)},\\[2mm] 0, & x \ge r + \dfrac{\alpha(r-l)(1-\beta)}{2(1-\alpha)}. \end{cases}$$

Здесь параметры l, r обозначают левую и правую условную границу соответственно.



Фиг. 4.

1.2. Семейство колоколообразных бесконечно дифференцируемых функций

Семейство трапециевидных функций может оказаться не очень удобным в тех случаях, когда необходимо проводить оптимизацию по условным границам разбиений. Дело в том, что трапеция не является дифференцируемой функцией в точках излома (которые соответствуют границам разбиений), что может осложнить применение градиентных методов оптимизации. Для того чтобы преодолеть эту проблему, введем семейство колоколообразных функций. Форму нечетких подмножеств данного семейства задают следующие характеристические функции (фиг. 4):

$$\mu_{\mathrm{II}}(x; l, r, \alpha, \beta) = \frac{1}{1 + \left(\frac{1}{\alpha} - 1\right)\left(\frac{x - (l + r)/2}{(r - l)/2}\right)^{2\beta}}, \quad \alpha \in (0, 1), \quad \beta \in \mathbb{N}, \quad l, r, x \in \mathbb{R}, \quad l < r. \tag{1.6}$$

Параметры l и r определяют *условные границы* нечеткого подмножества ($\mu_{\mathrm{II}}(l; l, r, \alpha, \cdot) = \mu_{\mathrm{II}}(r; l, r, \alpha, \cdot) = \alpha$); β показывает "степень нечеткости" подмножества. Предлагаемое семейство является достаточно гибким. В частности, при стремлении β к бесконечности получаем характеристическую функцию классического множества – отрезка $[l, r]$. Основным достоинством предлагаемого семейства является аналитичность μ_{II} по всем своим аргументам (x, l и r).

По аналогии с $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$-покрытием можно также ввести понятие покрытия для характеристических функций (1.6).

Определение 4. $[\boldsymbol{\alpha}, \boldsymbol{\beta}]$-*покрытием* признака для разбиения области его значений $a_1 < \ldots < a_n$ будем называть совокупность нечетких подмножеств $\{M_i\}_{i=1}^{n-1}$ с характеристическими колоколообразными функциями $\mu_{M_i}(x) = \mu_{\mathrm{II}}(x; a_i, a_{i+1}, \alpha_i, \beta_i)$. При этом $\alpha_1, \ldots, \alpha_{n-1} \in (0, 1)$, $\beta_1, \ldots, \beta_{n-1} \in \mathbb{N}$.

Для $[\boldsymbol{\alpha}, \boldsymbol{\beta}]$-покрытия значения параметров $\boldsymbol{\alpha}$ определяют значения характеристических функций принадлежности в точках разбиения. В том случае, когда отсутствуют веские причины задавать значения данного параметра, то разумным выбором является значение 0.5. По аналогии с $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$-покрытием значения параметров β в $[\boldsymbol{\alpha}, \boldsymbol{\beta}]$-покрытии можно трактовать как степень уверенности эксперта относительно выбранных границ разбиения. Потому можно положить одно и то же значение β для всех нечетких подмножеств покрытия, т.е. $\beta_1, \ldots, \beta_{n-1} = \beta$.

Характеристические функции нечетких подмножеств в семействах трапециевидных и колоколообразных функций определяются относительно условных границ l и r. В дальнейшем обозначение $\mu(x; l, r)$ может восприниматься в смысле I или II без ограничения корректности построений.

2. МОДЕЛЬ НЕЧЕТКИХ РАССУЖДЕНИЙ

Предположим, что для каждого признака $i \in \{1, 2, \ldots, d\}$ задана экспертная интерпретация $\mathfrak{T}_i$ (база данных системы). Она может определяться с помощью $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$- или $[\boldsymbol{\alpha}, \boldsymbol{\beta}]$-покрытия. Пусть также известен набор правил, связывающих нечеткие подмножества входных и выходных переменных (база знаний системы). Каждое правило может быть представлено следующим образом:

$$\text{“ЕСЛИ } x_{i_1} \in M_{i_1}^{j_1} \;\&\ldots\&\; x_{i_r} \in M_{i_r}^{j_r}, \text{ ТО } y \in N^k\text{”.} \tag{2.1}$$

Здесь $M_{i_1}^{j_1}, \ldots, M_{i_r}^{j_r}$, N^k – нечеткие подмножества соответствующих переменных. Множество $\mathrm{Sump}(R) = \{M_{i_1}^{j_1}, \ldots, M_{i_r}^{j_r}\}$ будем называть *посылкой* правила R (2.1), метку класса в задаче классификации (индекс нечеткого подмножества вещественной оси в задаче восстановления регрессии) $\mathrm{Res}(R) = N^k$ – *заключением* правила R; число $\mathrm{Ord}(R) = r$ – *порядком* правила R. Такие правила могут быть получены от эксперта или сгенерированы по прецедентной информации (см. разд. 3).

Рассмотрим задачу проведения нечетких рассуждений (классификации новых объектов) с использованием базы данных и базы знаний. К сожалению, на настоящий момент не существует единых подходов по отношению к тому, какие вычисления необходимо проделать в программе, где встроены функции принадлежности, чтобы смоделировать нечеткий вывод. Существует более 100 методов преобразования нечетких выводов на лингвистическом уровне в вычислении (см. [6]). Какому именно методу отдать предпочтение, определяется прежде всего практическими результатами использования экспертных систем. В связи с этим представляется разумным принять стратегию выбора методов, лучше всего зарекомендовавших себя на практике.

Для определения нечеткого вывода необходимо задать способ вычисления функций принадлежности антецедента правила (нечеткой конъюнкции) и вывода (нечеткой импликации), а также определить композиционное правило вывода.

Нечеткая конъюнкция традиционно задается t-нормой [см. [7]). В настоящей статье будет использоваться операция min, которая является типичной t-нормой.

Знание эксперта (правило) $\mathrm{Sump}(R) \longrightarrow \mathrm{Res}(R)$ отражает некоторое нечеткое причинное отношение предпосылки и заключения, т.е.

$$R = \mathrm{Sump}(R) \longrightarrow \mathrm{Res}(R).$$

Посылку R можно рассматривать как нечеткое подмножество на прямом произведении $X \times Y$ полного пространства предпосылок X и полного пространства заключений Y. Таким образом, процесс получения результата вывода B' с использованием данных наблюдения A' и знания $\mathrm{Sump}(R) \longrightarrow \mathrm{Res}(R)$ можно представить в виде формулы

$$B' = A' \bullet R = A' \bullet (\mathrm{Sump}(R) \longrightarrow \mathrm{Res}(R)).$$

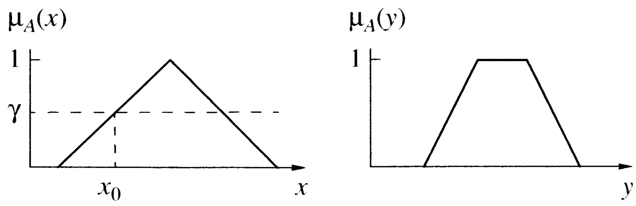
Здесь знак $\bullet$ является композиционным правилом нечеткого вывода, а знак $\longrightarrow$ в правиле $\mathrm{Sump}(R) \longrightarrow \mathrm{Res}(R)$ – нечеткой импликацией.

На практике нередки случаи, когда из-за особенностей системы информацию с достаточно хорошей точностью получить не удается либо нет возможности установить устройство измерения показателя, что вынуждает оценивать его по косвенным характеристикам. В подобных случаях удобно принимать за информационное наблюдение представление с помощью нечеткого подмножества. Поэтому в более общем случае наблюдение A' представляет собой нечеткое подмножество, а закон композиции $\bullet$ становится отдельным предметом исследования. Однако для систем распознавания можно положить, что значения входных переменных известны в виде конкретных значений, вследствие чего изучение этого вопроса опускается.

2.1. Выбор способа введения нечеткой импликации

Так же как и для операций НЕ, И, ИЛИ, расширение понятия импликации до нечеткого случая приводит к семействам обобщений. Однако в данном случае не удается построить аксиоматическую базу для расширения (по аналогии с t-нормой и t-конормой), вследствие чего обобщение данного понятия носит скорее эвристический характер.

Рассмотрим два нечетких подмножества A и B с их функциями принадлежности $\mu_A(x)$ и $\mu_B(y)$ (см. фиг. 5). С целью обоснованного выбора определения нечеткой импликации $\mu_{A \to B}(x, y)$ было



Фиг. 5.

проведено обширное исследование (см. [10]). В качестве критерия выбора была взята выполнимость некоторых схем нечетких рассуждений. Среди них можно выделить такие:

$$\frac{\text{если } X \text{ есть } A, \text{ то } Y \text{ есть } B;\ X \text{ есть } A}{Y \text{ есть } B};$$

$$\frac{\text{если } X \text{ есть } A, \text{ то } Y \text{ есть } B;\ X \text{ есть очень } A}{Y \text{ есть } B};$$

$$\frac{\text{если } X \text{ есть } A, \text{ то } Y \text{ есть } B;\ X \text{ есть более или менее } A}{Y \text{ есть более или менее } B};$$

$$\frac{\text{если } X \text{ есть } A, \text{ то } Y \text{ есть } B;\ X \text{ есть не-}A}{Y \text{ неизвестно}};$$

$$\frac{\text{если } X \text{ есть } A, \text{ то } Y \text{ есть } B;\ Y \text{ есть не-}B}{X \text{ есть не-}A};$$

$$\frac{\text{если } X \text{ есть } A, \text{ то } Y \text{ есть } B;\ Y \text{ есть } B}{X \text{ неизвестно}}.$$

Для исследования были взяты все известные в литературе по нечетким множествам определения нечеткой импликации с добавлением нескольких своих вариантов. В результате проверок был сделан вывод, что можно использовать в основном следующие правила (R_c, R_s, R_g – обозначения, введенные авторами, см. [10]):

(а) правило R_c (правило Мамдани):

$$\mu_{A\to B}(x, y) = \min(\mu_A(x), \mu_B(y));$$

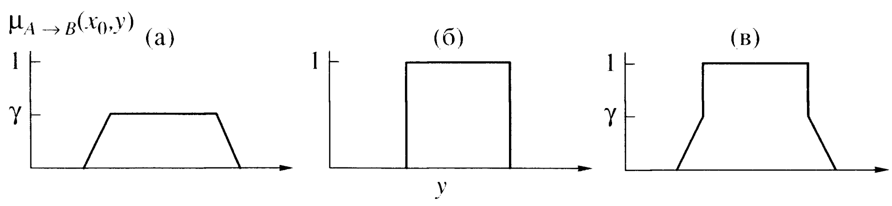
(б) правило R_s:

$$\mu_{A\to B}(x, y) = \begin{cases} 1, & \mu_A(x) \le \mu_B(y), \\ 0, & \mu_A(x) > \mu_B(y); \end{cases}$$

(в) правило R_g:

$$\mu_{A\to B}(x, y) = \begin{cases} 1, & \mu_A(x) \le \mu_B(y), \\ \mu_B(y), & \mu_A(x) > \mu_B(y). \end{cases}$$

Смысл каждого из этих выводов для примера, представленного фиг. 5, дается на фиг. 6. Как нетрудно видеть, правила вывода R_s и R_g не учитывают степень принадлежности $\gamma = \mu_A(x_0)$ в явном виде, а используют эти значения лишь для “усечения нечеткости”. Таким образом, в этом случае на выходе мы получаем практически четкие подмножества с уровнем принадлежности, равным 1, причем чем меньше γ, тем больше носитель подмножества вывода. Для правила вывода R_c значение γ играет большую роль. В связи с этим для задач распознавания представляется разумным использовать именно правило R_c для построения модели нечеткого вывода по правилам.



Фиг. 6.

2.2. Модель вычисления прогноза

Пусть имеется m лингвистических правил $\{R_i\}_{i=1}^m$ вывода вида (2.1). Согласно выводам предыдущего раздела, в качестве функции принадлежности нечеткой импликации выберем функцию, предложенную Мамдани (см. [8]):

$$\mu_{\mathrm{Sump}(R) \to \mathrm{Res}(R)}(x, y) = \min(\mu_{\mathrm{Sump}(R)}(x), \mu_{\mathrm{Res}(R)}(y)).$$

Тогда с учетом того, что в качестве нечеткой конъюнкции используется функция min, для правила вывода типа (2.1) соответствующая функция принадлежности будет иметь вид

$$\mu_{R_i}(\mathbf{x}, y) = \min\left(\mu_{M_{i_1}^{j_1}}(x_{i_1}), \ldots, \mu_{M_{i_r}^{j_r}}(x_{i_r}), \mu_{N^k}(y)\right).$$

Для группы правил вывода с одним и тем же результирующим нечетким подмножеством вычисление функции принадлежности происходит по формуле

$$\mu_{R_{N^k}}(\mathbf{x}, y) = \frac{\sum\limits_{\{i\,:\,\mathrm{Res}(R_i) = N^k\}} \mu_{R_i}(\mathbf{x}, y)}{\sum\limits_{\{i\,:\,\mathrm{Res}(R_i) = N^k\}} 1}, \tag{2.2}$$

которая несколько отличается от классического взятия максимума (см. [6]). Однако авторы считают (2.2) более гибким средством, в большей степени учитывающим индивидуальность каждого правила вывода. Предположим далее, что помимо самих правил вывода известна также информация о весах этих правил. Эти веса могут быть заданы экспертом или получены в результате оптимизации по обучающей выборке (см. разд. 4). В качестве весов также могут выступать значения эффективностей правил (см. разд. 3). Обозначим вес i-го правила через w_i. Тогда (2.2) преобразуется к виду

$$\mu_{R_{N^k}}(\mathbf{x}, y) = \frac{\sum\limits_{\{i\,:\,\mathrm{Res}(R_i) = N^k\}} w_i \mu_{R_i}(\mathbf{x}, y)}{\sum\limits_{\{i\,:\,\mathrm{Res}(R_i) = N^k\}} w_i}.$$

Пусть в пространстве прогнозной переменной определено K нечетких подмножеств $\{N^k\}_{k=1}^K$. Обозначим через $\mathrm{med}(N^k)$ медиану k-го нечеткого подмножества, т.е.

$$\mathrm{med}(N^k) = \frac{a_k + a_{k+1}}{2}.$$

Здесь a_k, a_{k+1} – соответствующие элементы разбиения в экспертной интерпретации пространства результата.

На последнем этапе остается определить по нечеткому подмножеству вывода точечное значение прогноза. Для этого в основном используются следующие методы (см. [6, 7, 8, 10]):

$$z = \arg\max_{k = \overline{1, K}} \mu_{R_{N^k}}(\mathbf{x}, \mathrm{med}(N^k)), \tag{2.3}$$

$$z = \frac{\sum_{k=1}^{K} \mathrm{med}(N^k)\mu_{R_{N^k}}(\mathbf{x}, \mathrm{med}(N^k))}{\sum_{k=1}^{K} \mu_{R_{N^k}}(\mathbf{x}, \mathrm{med}(N^k))}. \tag{2.4}$$

Дефаззификация по формуле (2.3) получила название дефаззификации по моде, в то время как формула (2.4) приводит к методу центра тяжести (ЦТ). Выражение (2.3) используется при построении экспертной системы для решения задачи классификации, т.е. когда необходимо разделить генеральную совокупность на непересекающиеся классы. Формула (2.4) используется в том случае, когда необходимо построить непрерывный прогноз.

Таким образом, описана следующая модель вычисления прогноза:

$$\mathfrak{M}[\{a_i^j\}_{i=0,\, j=1}^{d,\, n_i}, \{\alpha_i\}_{i=1}^{d}, \{\beta_i\}_{i-1}^{d}, \{R_i\}_{i=1}^{m}, \{w_i\}_{i=1}^{m}]. \tag{2.5}$$

Здесь $\{a_i^j\}_{i=0,\, j=1}^{d,\, n_i}$ – разбиения множеств значений каждого из d признаков и пространства результата, $\{\alpha_i\}_{i=1}^{d}$, $\{\beta_i\}_{i=1}^{d}$ – параметры $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$- или $[\boldsymbol{\alpha}, \boldsymbol{\beta}]$-покрытия экспертной интерпретации признаков, $\{R_i\}_{i=1}^{m}$ – набор правил вывода, $\{w_i\}_{i=1}^{m}$ – соответствующие веса правил.

Значения параметров модели $\mathfrak{M}$ можно находить путем оптимизации по множеству прецедентов $\{\mathbf{x}^i, y^i\}_{k=1}^{q}$ с помощью выбранного функционала качества. Конкретные функционалы качества и методы их оптимизации будут рассмотрены в разд. 4.

3. АВТОМАТИЧЕСКОЕ ПРИОБРЕТЕНИЕ ЗНАНИЙ

Рассмотрим продукционное правило вывода (2.1):

$$\text{“ЕСЛИ } x_{i_1} \in M_{i_1}^{j_1} \,\&\, \ldots \,\&\, x_{i_r} \in M_{i_r}^{j_r}, \text{ ТО } y \in N^k\text{”}.$$

Множество всех возможных правил данного вида $\mathfrak{R}^* = \{R\}$ будем называть *универсальной базой знаний*. Посылку правила $\mathrm{Sump}(R) = \{M_{i_1}^{j_1}, \ldots, M_{i_r}^{j_r}\}$ можно рассматривать как нечеткое подмножество полного пространства $\mathbb{R}^r$. Характеристическая функция этого нечеткого подмножества, в дальнейшем называемая *степенью принадлежности объекта посылке*, в соответствии с выбранной характеристической функцией нечеткой конъюнкции имеет вид

$$\nu_R(x) = \min\left(\mu_{M_{i_1}^{j_1}}(x_{i_1}), \ldots, \mu_{M_{i_r}^{j_r}}(x_{i_r})\right).$$

Характеристическая функция заключения правила имеет вид

$$\mu_{N^k}(y) = \begin{cases} 1 & \text{при } y = c_k, \\ 0 & \text{иначе,} \end{cases} \qquad c_k \in \{1, 2, \ldots, l\},$$

в случае задачи классификации и

$$\mu_{N^k}(y) = \mu(y; l_k, r_k), \quad l_k < r_k,$$

в случае задачи восстановления регрессии.

Проблема приобретения знаний заключается в том, чтобы выделить подмножество $\mathfrak{R}$ универсальной базы знаний. При этом правила, составляющие $\mathfrak{R}$ должны удовлетворять некоторому критерию значимости, основанному на прецедентной информации. В следующем подразделе формулируется ряд критерием значимости, заданные предикатами C на единичном квадрате числовой плоскости. Важными характеристиками, при помощи которых можно сформулировать многие известные критерии значимости, являются репрезентативность и эффективность.

Определение 5. *Репрезентативностью* правила R называется величина

$$\operatorname{rep}(R) = \frac{\sum_{k=1}^{q} \nu_R(x^k)}{q}. \tag{3.1}$$

Определение 6. *Эффективностью* правила R называется величина

$$\operatorname{eff}(R) = \frac{\sum_{k=1}^{q} \min(\nu_R(x^k), \mu_{\operatorname{Res}(R)}(y^k))}{\operatorname{rep}(R)q}.$$

Если все множества в посылке правила четкие, то тогда репрезентативность правила является долей объектов обучающей выборки, удовлетворяющих посылке правила, а эффективность – доля объектов, удовлетворяющих правилу целиком, среди объектов, удовлетворяющих посылке правила. Задача автоматического приобретения знаний по прецедентной информации заключается в нахождении всех правил $R \in \mathfrak{R}^*$ таких, что некоторый предикат значимости правила равен единице, т.е. $C(\operatorname{rep}(R), \operatorname{eff}(R)) = 1$.

3.1. Критерии значимости правил

Задачей процесса генерации правил вывода является поиск правил, которые по возможности обладают высокими значениями репрезентативности и эффективности. Более формально: правило $R \in \tilde{\mathfrak{R}} \subset \mathfrak{R}$ является значимым, если $C(\operatorname{rep}(R), \operatorname{eff}(R)) = 1$, где $C : [0, 1]^2 \longrightarrow \{0, 1\}$.

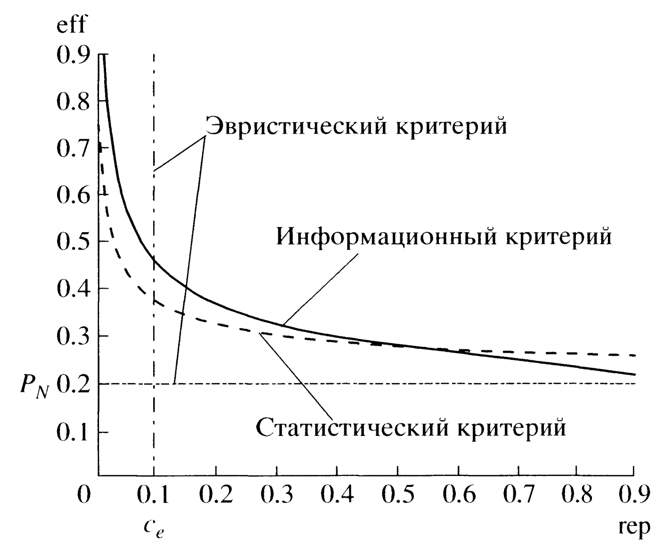
Эвристический критерий значимости. В наиболее простой форме предикат значимости правила задается фиксированными порогами для значений эффективности и репрезентативности.

Определение 7. Пусть даны числа $0 < c_r \le 1$ и $0 < c_e \le 1$. *Эвристическим критерием* будем называть следующий предикат на единичном квадрате числовой плоскости (фиг. 7):

$$C^h(v, w) = \begin{cases} 1, & \text{если } v \ge c_r \text{ и } w \ge c_e, \\ 0 & \text{иначе,} \end{cases} \qquad (v, w) \in [0, 1]^2.$$

Правило R будем называть *значимым по эвристическому критерию*, если $C^h(\operatorname{rep}(R), \operatorname{eff}(R)) = 1$.

Однако константные пороги для значений эффективности и репрезентативности не совсем точно отражают суть значимости правила. Понятно, что для низких значений репрезентативно-



Фиг. 7.

сти значимое правило должно обладать высоким значением эффективности. С другой стороны, для высоких значений репрезентативности значение эффективности немного выше априорной вероятности появления класса уже будет свидетельствовать о значимости правила. Следующий критерий значимости принимает во внимание указанные замечания.

Статистический критерий значимости основывается на технике проверки статистических гипотез. Предположим, что все множества в посылке правила четкие. Тогда правило R является незначимым, если информация о том, что объект удовлетворяет посылке правила Sump(R), не дает информации о его принадлежности заключению этого правила Res(R) = N:

$$\mathbb{P}(y \in N \mid \mathbf{x} \in \mathrm{Sump}(R)) = \mathbb{P}(y \in N).$$

Обозначим оценку априорной вероятности заключения N по обучающей выборке через P_N:

$$P_N = \frac{\sum_{k=1}^{q} \mu_N(y^k)}{q}. \tag{3.2}$$

Если гипотеза верна, тогда случайная величина rep(R)eff(R)q представляет собой n = rep(R)q испытаний Бернулли с вероятностью успеха P_N и может быть приближена нормальным распределением с центром rep(R)qP_N и дисперсией rep(R)$qP_N(1 - P_N)$, т.е.

$$\frac{\mathrm{eff}(R) - P_N}{\sqrt{\frac{P_N(1-P_N)}{\mathrm{rep}(R)q}}} \sim \mathcal{N}(0, 1).$$

Выбирая z, соответствующее достаточно высокому уровню значимости, имеем доверительный интервал, в котором принимается гипотеза

$$\frac{\mathrm{eff}(R) - P_N}{\sqrt{\frac{P_N(1-P_N)}{\mathrm{rep}(R)q}}} \in (-z, z).$$

В области $(-\infty, -z]$ расположены антиправила, из принадлежности объекта посылке которых следует сделать вывод о непринадлежности объекта их заключению. Таким образом, правило признается значимым, если

$$\frac{\mathrm{eff}(R) - P_N}{\sqrt{\frac{P_N(1-P_N)}{\mathrm{rep}(R)q}}} > z.$$

Определение 8. Пусть дан уровень значимости $\alpha > 0$. *Статистическим критерием* будем называть следующий предикат на единичном квадрате числовой плоскости (фиг. 7):

$$C^s(v, w) = \begin{cases} 1, \text{ если } \dfrac{w - P_N}{\sqrt{\dfrac{P_N(1-P_N)}{vq}}} \ge z_\alpha \\ 0 \text{ иначе,} \end{cases} \qquad (v, w) \in [0, 1]^2.$$

Здесь z_α – квантиль стандартного нормального распределения, отвечающая уровню значимости α. Правило R будем называть *значимым по статистическому критерию*, если C^s(rep(R), eff(R)) = 1.

Статистический критерий сформулирован для случая, когда все множества в посылке правила четкие. В этом случае рассмотрение испытаний Бернулли и аппроксимация нормальным распределением являются вполне законными. Данный критерий можно использовать и в нечетком случае, однако тогда его следует считать эвристикой. Тем не менее в практических приложениях его использование часто приводит к хорошим результатам.

Информационный критерий значимости основывается на понятиях энтропии и информационного выигрыша. Напомним, что энтропией дискретного распределения называется математическое ожидание количества информации; при этом если вероятность исхода равна $P(w)$, то количество информации, связанное с ним, равно $-\log(P(w))$. Таким образом, энтропию обучающей выборки до получения информации о существовании правила R с заключением N, учитывая соотношение (3.2), определяем по формуле

$$H_1 = -P_N\log(P_N) - (1-P_N)\log(1-P_N).$$

После получения информации о существовании правила R энтропия принимает более сложный вид:

$$H_2(\mathrm{rep}(R), \mathrm{eff}(R)) = -\mathrm{rep}(R)\{\mathrm{eff}(R)\log[\mathrm{eff}(R)] + [1-\mathrm{eff}(R)]\log[1-\mathrm{eff}(R)]\} -$$
$$-(1-\mathrm{rep}(R))\left[\frac{P_N - \mathrm{rep}(R)\mathrm{eff}(R)}{1-\mathrm{rep}(R)}\log\left(\frac{P_N - \mathrm{rep}(R)\mathrm{eff}(R)}{1-\mathrm{rep}(R)}\right) +\right.$$
$$\left.+\frac{1-\mathrm{rep}(R)[1-\mathrm{eff}(R)] - P_N}{1-\mathrm{rep}(R)}\log\left(\frac{1-\mathrm{rep}(R)[1-\mathrm{eff}(R)] - P_N}{1-\mathrm{rep}(R)}\right)\right].$$

Так же как и в рассуждениях для статистического критерия, при $\mathrm{eff}(R) < P_N$, значимым объявляется антиправило. Для устранения этой проблемы достаточно домножить информационный выигрыш на $\mathrm{sign}(\mathrm{eff}(R) - P_N)$.

Определение 9. Пусть дано число $c_i > 0$. *Информационным критерием* будем называть следующий предикат на единичном квадрате числовой плоскости (фиг. 7):

$$C^i(v, w) = \begin{cases} 1, & \text{если } \mathrm{sign}(w - P_N)(H_1 - H_2(v, w)) \ge c_i, \\ 0 & \text{иначе,} \end{cases} \qquad (v, w) \in [0, 1]^2.$$

Правило R будем называть *значимым по информационному критерию*, если $C^i(\mathrm{rep}(R), \mathrm{eff}(R)) = 1$.

На фиг. 7 представлены различные критерии значимости. Правила, лежащие выше обозначенных кривых, признаются значимыми по соответствующему критерию. Возможные значения эффективности и репрезентативности удовлетворяют соотношению $vw \le P_N$.

Критерий значимости Колмогорова–Смирнова. Рассмотрим распределение значений характеристической функции правила R:

$$G(t) = \mathbb{P}\{\mu_R(x) < t\}.$$

Можно построить эмпирические функции распределения $G'(t)$ и $G''(t)$, используя объекты $y_i \in \mathrm{Res}(R)$ и $y_i \notin \mathrm{Res}(R)$ соответственно. Тогда для определения значимых различий между распределениями может быть использован непараметрический статистический тест Колмогорова–Смирнова.

Заметим, что, в отличие от предыдущих критериев значимости, данный критерий сформулирован и обоснован для применения в нечетком случае.

3.2. Общие требования к предикатам значимости правил

За исключением эвристического критерия значимости, предикаты для остальных критериев задаются в форме

$$C(v, w) = \begin{cases} 1, & \text{если } F(v, w) \ge F_0, \\ 0 & \text{иначе,} \end{cases}$$

где $F : [0, 1]^2 \longrightarrow \mathbb{R}$ – функция критерия и F_0 – некоторый порог. Чем больше значение функции критерия F, тем более значимым является правило. Можно сформулировать следующее требование к функциям критерия для "разумных" критериев значимости правил.

Определение 10. Функция критерия F *удовлетворяет условию монотонности*, если $\forall(v_0, w_0) : F(v_0, w_0) \ge F_0$ выполняется условие

$$F(v, w) \ge F(v_0, w_0) \quad \forall(v, w) > (v_0, w_0). \tag{3.3}$$

Для выполнения условия (3.3) необходимо и достаточно, чтобы $\forall(v_0, w_0) : F(v_0, w_0) \geq F_0$ выполнялись условия

$$F(v, w_0) \geq F(v_0, w_0) \quad \forall v > v_0, \tag{3.4}$$

$$F(v_0, w) \geq F(v_0, w_0) \quad \forall w > w_0, \tag{3.5}$$

где (3.4) означает, что с ростом репрезентативности с сохранением значения эффективности правило не становится менее значимым, а (3.5) – что для фиксированного уровня репрезентативности увеличение значения эффективности только увеличивает значимость правила.

Теорема 2. *Функции критерия для критерия значимости Колмогорова–Смирнова в четком случае и для статистического, информационного критериев значимости в общем случае удовлетворяют условию* (3.3).

Доказательство. Для доказательства (3.3) необходимо и достаточно проверить выполнение условий (3.4) и (3.5).

1. Для статистического критерия функция критерия определяется выражением

$$\frac{w - P_N}{\sqrt{\dfrac{P_N(1 - P_N)}{vq}}}.$$

При фиксированном значении репрезентативности v знаменатель является постоянной величиной и функция критерия возрастает с ростом эффективности w. С другой стороны, при фиксированном значении w числитель является константой, а знаменатель убывает с ростом v. Это означает, что функция критерия возрастает с ростом v.

2. Для информационного критерия функция критерия определяется выражением

$$vH(w) + (1 - v)H\left(\frac{P_N - vw}{1 - v}\right), \tag{3.6}$$

где $H(x) = x\log(x) + (1 - x)\log(1 - x)$. Вычисляя производную выражения (3.6) по w, получаем

$$v\log\left(\frac{w}{1 - w}\right) - v\log\left(\frac{P_N - vw}{1 - v - P_N + vw}\right) = v\log\left(\frac{w - vw - wP_N + vw^2}{P_N - vw - wP_N + vw^2}\right).$$

Исключая из рассмотрения антиправила, т.е. принимая $w \geq P_N$, получаем, что производная всюду неотрицательна. Следовательно, функция монотонно неубывает по w. Аналогичные рассуждения для v приводят к тому, что функция (3.6) монотонно неубывает по v.

3. В четком случае функция критерия для критерия Колмогорова–Смирнова имеет вид

$$p/P - n/N, \tag{3.7}$$

где P и N – количество своих и чужих объектов в выборке, а p и n – количество своих и чужих объектов, выделяемых правилом. При фиксированном значении репрезентативности $\text{rep}(R) = (p + n)/(P + N)$ увеличение эффективности $\text{eff}(R) = p/(p + n)$ эквивалентно увеличению p. При этом значение n уменьшается так, что $p + n = \text{const}$. Тогда функция критерия (3.7) возрастает.

При фиксированном значении эффективности имеем $p = Cn$, где $C \in (0, 1)$ – некоторая константа. В этом случае увеличение значения репрезентативности эквивалентно увеличению p. При этом функция критерия (3.7) также возрастает.

3.3. Характеристика предиката, определяющего критерий значимости

Для дальнейших построений необходимо ввести некоторые дополнительные обозначения.

Определение 11. Число $\alpha \geq 0$ будем называть *характеристикой* предиката $C : [0, 1]^2 \longrightarrow \{0, 1\}$, если выполнено следующее:

$\forall(v, w) \in [0, 1]^2 : C(v, w) = 1 \Rightarrow vw \geq \alpha$;

$\forall\varepsilon > 0\ \exists(v, w) \in [0, 1]^2 : C(v, w) = 1$ и $vw < \alpha + \varepsilon$.

Определение 12. Предикат $C^\alpha : [0, 1]^2 \longrightarrow \{0, 1\}$ с характеристикой α будем называть *максимальным*, если для любого другого предиката $C' : [0, 1]^2 \longrightarrow \{0, 1\}$ с характеристикой α выполнено условие $\forall(v, w) \in [0, 1] : C'(v, w) = 1 \Rightarrow C^\alpha(v, w) = 1$.

Теорема 3. *Для любого* $\alpha \in [0, 1]$ *существует единственный максимальный предикат с характеристикой* α:

$$C^{\alpha}(v, w) = \begin{cases} 1, & \textit{если } vw \geq \alpha, \\ 0 & \textit{иначе}, \end{cases} \quad (v, w) \in [0, 1]^2. \tag{3.8}$$

Доказательство. *Существование* следует из того, что C^{α}, определяемый формулой (3.8), удовлетворяет определению максимального предиката.

Единственность. Пусть C' – произвольный максимальный предикат с характеристикой α. Тогда, согласно определению максимального предиката,

$$\forall (v, w) \in [0, 1]^2 : \begin{cases} C'(v, w) \Rightarrow C^{\alpha}(v, w), \\ C^{\alpha}(v, w) \Rightarrow C'(v, w), \end{cases}$$

т.е. $C'(v, w) \Leftrightarrow C^{\alpha}(v, w)\ \forall (v, w) \in [0, 1]^2$. Утверждение доказано.

3.4. Метод эффективных сужений

Определение 13. Правило R_2 будем называть *сужением* правила R_1 и обозначать $R_2 \subset R_1$, если выполнены следующие условия:

1) $\mathrm{Res}(R_1) = \mathrm{Res}(R_2)$;

2) $\mathrm{Sump}(R_1) \subset \mathrm{Sump}(R_2)$.

Теорема 4. *Если* $R' \supset R$, *то выполнены неравенства*

$$\mathrm{rep}(R') \geq \mathrm{rep}(R),$$

$$\mathrm{rep}(R')\mathrm{eff}(R') \geq \mathrm{rep}(R)\mathrm{eff}(R).$$

Доказательство. Допустим, что $\mathrm{Sump}(R') = \{ M_{i_1}^{j_1}, \ldots, M_{i_r}^{j_r} \}$, а посылка правила R отличается от посылки R' на одно множество $\mathrm{Sump}(R) = \mathrm{Sump}(R') \cup \{ M_{i_{r+1}}^{j_{r+1}} \}$. Тогда

$$\mathrm{rep}(R') = \{\text{опр. } (3.1)\} = \frac{1}{q}\sum_{k=1}^{q} \min\left(\mu_{M_{i_1}^{j_1}}(x_{i_1}^k), \ldots, \mu_{M_{i_r}^{j_r}}(x_{i_r}^k)\right) \geq$$

$$\geq \frac{1}{q}\sum_{k=1}^{q} \min\left(\mu_{M_{i_1}^{j_1}}(x_{i_1}^k), \ldots, \mu_{M_{i_r}^{j_r}}(x_{i_r}^k), \mu_{M_{i_{r+1}}^{j_{r+1}}}(x_{i_{r+1}}^k)\right) = \mathrm{rep}(R).$$

Соответственно,

$$\mathrm{rep}(R')\mathrm{eff}(R') = \frac{1}{q} \sum_{k = \overline{1, q};\ y^k = \mathrm{Res}(R')} \min\left(\mu_{M_{i_1}^{j_1}}(x_{i_1}^k), \ldots, \mu_{M_{i_r}^{j_r}}(x_{i_r}^k)\right) \geq$$

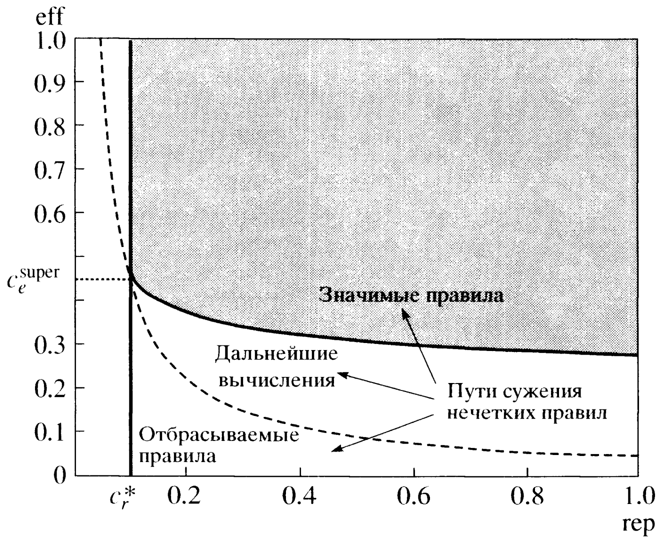
$$\geq \frac{1}{q} \sum_{k = \overline{1, q};\ y^k = \mathrm{Res}(R')} \min\left(\mu_{M_{i_1}^{j_1}}(x_{i_1}^k), \ldots, \mu_{M_{i_r}^{j_r}}(x_{i_r}^k), \mu_{M_{i_{r+1}}^{j_{r+1}}}(x_{i_{r+1}}^k)\right) = \{\mathrm{Res}(R) = \mathrm{Res}(R')\} = \mathrm{rep}(R)\mathrm{eff}(R).$$

Доказательство для случая произвольного порядка правила R проводится аналогично. Утверждение доказано.

Пусть заданы экспертные интерпретации всех признаков $\mathfrak{T} = \{\mathfrak{T}_1, \ldots, \mathfrak{T}_d\}$, заключение правил N и предикат C с характеристикой $\alpha \geq \alpha_0 > 0$. Пусть

$$c_r^* = \inf\{ v \in [0, 1] \mid \exists w \in [0, 1] : C(v, w) = 1 \}.$$

Следующий алгоритм позволяет находить все значимые правила минимального порядка.



Фиг. 8.

Шаг 1. Построим все возможные правила первого порядка для заданной экспертной интерпретации признаков $\mathfrak{I}$:

$$\mathfrak{R}' := \{R \in \mathfrak{R}^* \mid \operatorname{Res}(R) = N, \operatorname{Sump}(R) = M_i^{j_i}, j_i = \overline{1, n_i}, i = \overline{1, d}\}.$$

Шаг 2. Вычислим репрезентативность и эффективность каждого правила и исключим из рассмотрения нерепрезентативные:

$$\mathfrak{R}' := \{R \in \mathfrak{R}' \mid \operatorname{rep}(R) \ge c_r^*\}.$$

Шаг 3. Исключим из полученного множества правила, не удовлетворяющие максимальному критерию с характеристикой α_0:

$$\mathfrak{R}' := \{R \in \mathfrak{R}' \mid C^{\alpha_0}(\operatorname{rep}(R), \operatorname{eff}(R)) = 1\}.$$

Шаг 4. Если все правила исчерпаны, то алгоритм заканчивает работу. В противном случае, если правило удовлетворяет критерию C, то оно является значимым и заносится в результирующее множество $\tilde{\mathfrak{R}}$, иначе – участвует в процессе сужений:

$$\tilde{\mathfrak{R}} := \tilde{\mathfrak{R}} \cup \{R \in \mathfrak{R}' \mid C(\operatorname{rep}(R), \operatorname{eff}(R)) = 1\}$$

$$\mathfrak{R}' := \{R \in \mathfrak{R}' \mid C(\operatorname{rep}(R), \operatorname{eff}(R)) = 0\}.$$

Шаг 5. Остальные правила (если они есть) используются в процессе сужения следующим образом. Объединение посылок любых двух правил, которые сужаются к некоторому правилу большего порядка, и есть в точности посылка нового правила, если порядок последнего на единицу больше порядка исходных:

$$\mathfrak{R}' := \{R \in \mathfrak{R}^* \mid \operatorname{Res}(R) = N, \operatorname{Sump}(R) = \operatorname{Sump}(R_1) \cup \operatorname{Sump}(R_2), R_1, R_2 \in \mathfrak{R}',$$
$$\operatorname{Ord}(R) = \operatorname{Ord}(R_1) + 1 \ \forall R_+ : R \subset R_+, \operatorname{Ord}(R_+) = \operatorname{Ord}(R) - 1 \Rightarrow R_+ \in \mathfrak{R}'\}.$$

Шаг 6. Если все правила обработаны, т.е. $\mathfrak{R}' = \emptyset$, то конец работы алгоритма, иначе переходим к шагу 2.

На фиг. 8 представлена схема работы метода эффективных сужений. Заштрихованная область – область координат значимых правил. Вертикальная черная линия пересекает ось абсцисс в точке c_r^*. Все точки, в которых значение максимального предиката для заданного критерия равно единице, расположены выше штриховой линии.

Теорема 5 (основная). *Метод эффективных сужений строит все значимые правила минимального порядка, т.е.*

$$\mathfrak{R} = \{R \in \mathfrak{R}^* \mid C(\mathrm{rep}(R), \mathrm{eff}(R)) = 1\ \forall R' \supset R : C(\mathrm{rep}(R'), \mathrm{eff}(R')) = 0\}.$$

Доказательство. Рассмотрим произвольное минимальное по порядку значимое правило R, пусть $\mathrm{Ord}(R) = r$. Докажем, что оно войдет в результирующее множество $\mathfrak{R}$ на r-й итерации шага 4. Если это не так, то найдется правило меньшего (или такого же) порядка $\exists R' \supset R$, $\mathrm{Ord}(R') = r' \le r$, для которого R является сужением (или $R' = R$), исключаемое из $\mathfrak{R}''$ на r'-й итерации шага 2 или 3. Рассмотрим оба случая.

1. R' исключено из $\mathfrak{R}''$ в результате шага 2, т.е. $\mathrm{rep}(R') < c_r^*$. Тогда, на основании утверждения 4, имеем $\mathrm{rep}(R) < c_r^*$. Согласно определению точной нижней грани числового множества,

$$\forall \varepsilon \in (0, c_r^*]\ \forall w \in [0, 1] : C(c_r^* - \varepsilon, w) = 0,$$

а значит, в частности, $C(\mathrm{rep}(R), \mathrm{eff}(R)) = 0$, что противоречит значимости правила R.

2. R' исключено из $\mathfrak{R}''$ в результате шага 3, т.е. $\mathrm{rep}(R')\mathrm{eff}(R') < \alpha_0 \le \alpha$. Аналогично предыдущему рассуждению, на основании теоремы 4 имеем $\mathrm{rep}(R)\mathrm{eff}(R) < \alpha$. Следовательно, ввиду того что C имеет характеристику α имеем $C(\mathrm{rep}(R), \mathrm{eff}(R)) = 0$, что также противоречит значимости правила R. Утверждение доказано.

Замечание 1. Доказанное утверждение обосновывает применимость метода эффективных сужений для произвольного предиката с положительной характеристикой. В случае если характеристика предиката равна нулю и $c_r^* = 0$, шаги 2 и 3 не имеют смысла и алгоритм вырождается в тривиальный перебор.

Теорема 6. *Предикаты C^h, C^s и C^i имеют положительные характеристики.*

Доказательство. 1. Докажем, что характеристика C^h равна $c_r c_e$:

$$\forall (v, w) \in [0, 1]^2 : C^h(v, w) = 1 \Rightarrow \mathrm{rep} \ge c_r \text{ и } w \ge c_e \Rightarrow vw \ge c_r c_e = \alpha,$$

$$\forall \varepsilon > 0\ \exists v = c_r,\quad w = c_e : C^h(v, w) = 1 \text{ и } vw = c_r c_e = \alpha < \alpha + \varepsilon.$$

Значит, $\alpha = c_r c_e$ является характеристикой предиката C^h согласно определению (11).

2. Докажем, что характеристика C^s положительна. Предположим противное: $\alpha = 0$. Заметим, что $z \ge z_0 > 0$. Положим

$$\varepsilon = \varepsilon_0 < \min\left\{P_c, \left[z_0\sqrt{\frac{P_c(1-P_c)}{q}}/(1-P_c)\right]^2 = r_0\right\}^2;$$

тогда, по определению (11),

$$\exists (v, w) \in [0, 1]^2 : vw < \varepsilon \text{ и } \frac{w - P_N}{\sqrt{\frac{P_N(1-P_N)}{vq}}} \ge z \Rightarrow (w < P_N \vee v < r_0) \text{ и } \frac{w - P_N}{\sqrt{\frac{P_N(1-P_N)}{vq}}} \ge z \Rightarrow z < z_0.$$

Таким образом, получаем противоречие. Следовательно, характеристика C^s является положительной.

3. Докажем, что характеристика C^i положительна. Предположим противное: $\alpha = 0$. Заметим, что $c_i \ge c_0 > 0$ и $\lim\limits_{v \to +0} (H_1 - H_2(v, w)) = 0$, т.е. $\exists \delta : 0 < v < \delta \Rightarrow H_1 - H_2 < c_0$. Положим $\varepsilon = \delta P_N$; тогда, по определению (11), $\exists (v, w) \in [0, 1]^2 : vw < \varepsilon$ и $w > P_N$ и $H_1 - H_2(v, w) > c_i \Rightarrow c_i < c_0$. Противоречие. Утверждение доказано.

Замечание 2. Чем выше нижняя оценка характеристики предиката, тем больше незначимых в перспективе правил исключаются из рассмотрения на ранних этапах метода эффективных сужений. Получение точных значений характеристик является важной теоретической задачей, имеющей прямое отношение к практике. В настоящей работе сформулированы и доказаны теоремы существования положительных характеристик для известных критериев: эвристического, статистического и информационного.

3.5. Дизъюнктивная модификация метода эффективных сужений

В дальнейшем правила, посылка которых является конъюнкцией нечетких подмножеств, будем называть элементарными. В некоторых задачах закономерности, удовлетворяющие критерию значимости, имеют существенно "непрямоугольную" форму. При этом такие закономерности могут быть приближены множеством элементарных правил, каждое из которых, вообще говоря, не удовлетворяет предикату C. Такие правила назовем *сложными*, или *дизъюнктивными*.

Определение 14. Рассмотрим правила R_1, R_2 и их посылки $\mathrm{Sump}_1 = \{ M_{i_1(1)}^{j_1(1)}, \ldots, M_{i_{r(1)}(1)}^{j_{r(1)}(1)} \}$ и $\mathrm{Sump}_2 = \{ M_{i_1(2)}^{j_1(2)}, \ldots, M_{i_{r(2)}(2)}^{j_{r(2)}(2)} \}$. Соответственно, ν_1 и ν_2 – степени принадлежности этим посылкам. Данные правила будем называть *расположенными по соседству*, если найдется объект $\mathbf{x} \in \mathbb{R}^d$ такой, что

$$\mu_{M_{i_k(1)}^{j_k(1)}}(x_{i_k(1)}) \geq \alpha_{i_k(1)}, \quad k = \overline{1, r(1)},$$

$$\mu_{M_{i_k(2)}^{j_k(2)}}(x_{i_k(2)}) \geq \alpha_{i_k(2)}, \quad k = \overline{1, r(2)}.$$

Здесь $\boldsymbol{\alpha}$ задает соответствующее покрытие в экспертной интерпретации признаков.

Фактически это определение означает, что у соседних правил "ядра", определяемые условными границами соответствующих разбиений, пересекаются. В практических приложениях разумным является использование критерия значимости правил, определяемого пересечением предикатов C^h и C^s (фиг. 8). При этом ордината пересечения левой и нижней границ области значимых правил (обозначим ее c_e^{super}, и назовем порогом суперэффективности) обладает интересным свойством: результат объединения любого правила с эффективностью, превосходящей порог суперэффективности, и произвольного значимого правила является значимым дизъюнктивным правилом.

В настоящей работе предлагается алгоритм генерации подобных дизъюнктивных правил, являющийся расширением схемы МЭС. Он позволяет отыскивать значимые правила, состоящие из расположенных по соседству элементарных (некоторые, но не все).

Данная модификация касается только шага 3, вместо которого выполняется шаг 3'.

Шаг 3'. Вычислим репрезентативность и эффективность каждого правила. Если $C^{\alpha_0}(\cdot, \cdot) = 0$ и $\mathrm{eff}(R) < c_e^{\mathrm{super}}$, то правило исключается из дальнейших рассмотрений. Если $C^{\alpha_0}(\cdot, \cdot) = 0$, а его эффективность $\mathrm{eff}(R) \geq c_e^{\mathrm{super}}$ и при этом в базе знаний уже содержится правило, являющееся соседним по отношению к рассматриваемому, то тогда правило заносится в базу результирующих правил.

Модифицированная схема МЭС позволяет получить, вообще говоря, более точные решения по сравнению с МЭС за счет большего покрытия множества прецедентов правилами. Результаты испытаний на некоторых практических задачах приведены в разд. 5. Отметим, что данная модификация метода эффективных сужений решает задачу поиска всех дизъюнктивных правил приближенно. Любой метод получения точного решения имеет более высокую алгоритмическую сложность.

4. ОПТИМИЗАЦИЯ В МОДЕЛИ АЛГОРИТМОВ

Рассматриваемая задача синтеза экспертной системы распознавания образов с использованием прецедентной информации, как отмечалось выше, допускает построение решения в три этапа: задание экспертной интерпретации каждого признака, получение набора правил (знаний) и выбор наилучшего алгоритма в модели (2.5) (уточнение весов правил, параметров покрытий, границ разбиений и пр.). Решение каждого из этих этапов может проводиться как экспертом, так и автоматически (а также совместно). Рассмотрим последовательно вопросы поиска и оптимизации значений для каждой из групп параметров модели (2.5).

4.1. Первичное формирование границ нечетких подмножеств

В [16], [17], [13] задача поиска оптимального разбиения, а также параметров покрытия нечеткими подмножествами решается в одной связке с генерацией множества правил. В данной работе понятие экспертной интерпретации предлагается разделить на понятие разбиения области значений признаков и понятие набора покрывающих нечетких подмножеств. Первое из этих понятий является независимым и определяет трактовку экспертом самого понятия признака. Второе понятие отвечает за меру неточности в определении экспертной интерпретации и вполне может оптимизироваться с учетом прецедентной информации и обнаруженного набора правил вывода. В качестве покрытия можно выбрать $\{\alpha, \beta\}$- или $[\alpha, \beta]$-покрытия и оптимизировать функционал качества по параметрам α и β.

Пусть требуется построить некоторое прямоугольное разбиение пространства признаков $\{\{a_i^j\}_{j=1}^{n_i}\}_{i=1}^{d}$. Обозначим проекцию множества объектов обучающей выборки на i-й признак через множество $\{x_i^k\}_{k=1}^{q}$. Пусть также дано требуемое количество интервалов для каждого признака: n_i, $i = \overline{1, d}$. Тогда для нахождения первичного разбиения $a_i^1 < \ldots < a_i^{n_i}$ i-го признака разумно использовать результат одномерной кластеризации. Для решения этой задачи может быть задействован, например, метод равномерной массы (проводить разбиение таким образом, чтобы в каждой группе было примерно одинаковое число объектов, или в случае, если объекты имеют индивидуальные веса, группы объектов должны иметь примерно одинаковые веса). В простейшем виде этот метод таков.

Шаг 1. Упорядочим множество $\{x_i^k\}_{k=1}^{q}$: $x_i^{k_1} \le \ldots \le x_i^{k_q}$.

Шаг 2. Положим

$$a_i^1 = x_i^{k_1} - \frac{x_i^{k_q} - x_i^{k_1}}{2(n_i - 2)}, \quad a_i^{n_i} = x_i^{k_q} + \frac{x_i^{k_q} - x_i^{k_1}}{2(n_i - 2)}.$$

Шаг 3. Для каждого j от 2 до $n_i - 1$ вычислим $t = \left\lfloor \dfrac{(j-1)q}{n_i - 1} \right\rfloor$ и положим $a_i^j = \dfrac{x_i^{k_t} + x_i^{k_{t+1}}}{2}$. Конец работы алгоритма.

Также для решения задачи первичных границ могут быть использованы другие методы кластеризации, например K-средних (см. [3], [27]), иерархическая группировка (см. [3], [27]), восстановление смеси нормальных распределений (см. [28], [29]) и др.

Значения параметров $\{n_i\}_{i=1}^{d}$ (число кластеров для каждого признака) могут быть определены экспертом вручную путем визуального оценивания качества кластеризации в одномерном пространстве. Если количество признаков велико и визуальное решение требует значительных временных затрат, могут быть использованы методы автоматического определения числа кластеров. Среди них можно выделить байесовский подход для смеси нормальных распределений (см. [28], [29]), байесовский подход для смеси распределений Стьюдента (см. [30]), а также подход, основанный на многоразовой кластеризации пар подвыборок из исходной выборки с последующим построением гистограммы значений критерия сравнения двух кластеризаций (см. [31], [32]).

Разбиения множеств значений признаков могут последовательно уточняться в процессе исследования данной прикладной области. Например, объекты, неправильно распознанные на очередной итерации исследования, могут менять свое влияние на разбиение значений в будущем. В [23] предлагается использование процедуры адаптивной коррекции или бустинга (см. [21]) для формирования границ разбиений.

4.2. Оптимизация весов правил

Пусть известны экспертные интерпретации всех признаков $\mathfrak{T} = \{\mathfrak{T}_1, \ldots, \mathfrak{T}_d\}$ и построена база знаний $\mathfrak{R} = \{R \in \mathfrak{R}^* \mid C(\mathrm{rep}(R), \mathrm{eff}(R)) = 1\} = \{R_i\}_{i=1}^{m}$. Требуется провести оптимизацию значений весов правил $\{w_i\}_{i=1}^{m}$ с целью минимизации вероятности ошибки алгоритма на произвольном объекте признакового пространства. Однако эта величина недоступна для измерения, а ее оценивание представляет собой отдельную математическую проблему (см. [33]). В связи с этим рассматривают альтернативные функционалы качества в терминах обучающей выборки и модели алгоритмов.

Случай задачи классификации. Предположим, что решается задача распознавания образов (классификации). Тогда решающее правило (2.3) может рассматриваться как выбор класса, которому соответствует максимальная оценка. В качестве такой оценки выступает величина $\mu_{R_{V^k}}(\mathbf{x}, \mathrm{med}(N^k))$ для класса k. Введем следующую функцию правдоподобия данных и модели (заданного набора весов правил):

$$p(\mathbf{y} \mid \mathbf{w}) = \prod_{i=1}^{q}\left[\frac{\lambda + \mu_{R_{N^{y^i}}}(\mathbf{x}^i, \mathrm{med}(N^{y^i}))}{\lambda l + \sum_{c=1}^{l}\mu_{R_{N^c}}(\mathbf{x}^i, \mathrm{med}(N^c))}\right].$$

Здесь $\lambda > 0$ – произвольный положительный, заранее заданный параметр.

Известный метод максимума правдоподобия приводит к следующей оптимизационной задаче на значения весов правил:

$$\begin{gathered}\log p(\mathbf{y} \mid \mathbf{w}) \longrightarrow \max_{\mathbf{w}},\\ w_i \geq 0, \quad i = \overline{1, m}.\end{gathered} \tag{4.1}$$

Однако простая максимизация правдоподобия (4.1) может приводить к существенному переобучению алгоритма распознавания. Для того чтобы избежать данной ситуации, используются различные методы регуляризации обучения (см. [4], [29], [34], [35]). В данной статье предлагается использовать так называемый байесовский подход (см. [4], [29], [34], [36]–[38]). В этом подходе регуляризация задается при помощи априорного распределения в пространстве оптимизируемых параметров. Функция принятия решений в решающем правиле (2.3) может трактоваться, как взвешенная линейная комбинация функций принадлежности правилам. Таким образом, данная модель близка по своему виду к обобщенным линейным моделям, которые используются в таких популярных алгоритмах распознавания, как метод опорных векторов (см. [39], [40]), метод релевантных векторов (см. [38]), предсказание с помощью гауссовских процессов (см. [34], [41]) и др. Обычно при применении байесовского подхода в обобщенных линейных моделях в качестве априорного распределения используется нормальное распределение (см. [37], [38]) или двухстороннее распределение Лапласа (см. [42]). В задаче (4.1) веса являются неотрицательными величинами. Поэтому в качестве априорного распределения для весов правил предлагается использовать одностороннее нормальное распределение (4.2) или экспоненциальное распределение (4.3):

$$p(w_i \mid \alpha_i) = \sqrt{\frac{2\alpha_i}{\pi}}\exp\left(-\frac{1}{2}\alpha_i w_i^2\right), \quad i = \overline{1, m}, \tag{4.2}$$

$$p(w_i \mid \alpha_i) = \alpha_i \exp(-\alpha_i w_i), \quad i = \overline{1, m}. \tag{4.3}$$

Заметим, что здесь веса правил предполагаются независимыми случайными величинами, т.е. априорное распределение в пространстве весов есть произведение априорных распределений для каждого весового коэффициента. При этом набор параметров $\{\alpha_i\}_{i=1}^m$ определяет конкретный вид априорного распределения для каждого веса.

Для подбора гиперпараметров (параметров модели) $\{\alpha_i\}_{i=1}^m$ в байесовском подходе используется процедура максимизации обоснованности (правдоподобия модели, см. [36], evidence см. [38]):

$$p(\mathbf{y} \mid \boldsymbol{\alpha}) = \int p(\mathbf{y} \mid \mathbf{w})p(\mathbf{w} \mid \boldsymbol{\alpha})d\mathbf{w} \longrightarrow \max_{\boldsymbol{\alpha}}. \tag{4.4}$$

Формально в байесовском подходе для гиперпараметров также необходимо определить априорное распределение и использовать процедуру максимизации правдоподобия гипермодели. Такая процедура введения гиперпараметров очередного уровня может быть продолжена до тех пор, пока на каком-то этапе в качестве априорного распределения не выбирается равномерное распределение в области изменения параметров. В большинстве практических задач обычно ограничиваются одним уровнем.

Интеграл (4.4) не может быть вычислен аналитически. Поэтому при практическом применении процедуры максимизации обоснованности часто используется так называемое приближение Лапласа – замена логарифма подынтегральной функции параболой, что влечет за собой приближение подынтегральной функции гауссианой, интеграл от которой является легко вычислимым выражением (см. [37], [38]). При этом значения, соответствующие максимуму подынтегральной функции (регуляризованного функционала), являются искомыми весами правил.

Для случая априорного распределения типа (4.2) или (4.3) в [37], [38] получена эффективная итерационная процедура максимизации лапласовского приближения обоснованности для произвольного вида функции правдоподобия данных и модели $p(\mathbf{y} \mid \mathbf{w})$. Для ее получения необходимо приравнять к нулю значения производных логарифма обоснованности по параметрам и использовать прием из (см. [37]).

Случай задачи восстановления регрессии. Для задачи восстановления регрессии и решающего правила (2.4) введем следующую функцию правдоподобия данных и модели:

$$p(\mathbf{y} \mid \mathbf{w}) = (\sqrt{2\pi}\sigma)^{-m}\exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y}-\hat{\mathbf{y}}\|^2\right). \tag{4.5}$$

Здесь $\hat{y}^i = M[\ldots](\mathbf{x}^i)$ – результат предсказания выходной переменной с помощью экспертной системы M для объекта обучения $\mathbf{x}^i$, а σ – некоторый положительный параметр, определяющий уровень шума в данных.

Так же как и в случае задачи классификации, для подбора значений весов правил здесь может быть использована простая процедура максимума правдоподобия или процедура регуляризации в рамках байесовского подхода. Заметим, что в случае регрессии использование байесовского подхода позволяет также оценить значение параметра σ при помощи максимизации обоснованности.

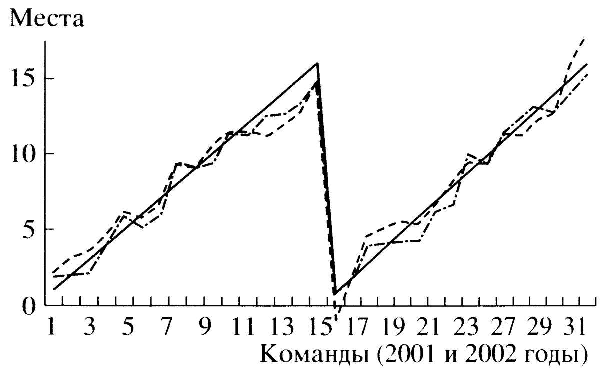
В более общем случае функция правдоподобия $p(\mathbf{y} \mid \mathbf{w})$ выбирается в виде $\exp[-\Sigma_{i=1}^{q} l(y^i, \hat{y}^i)]$, где $l : \mathbb{R}^2 \longrightarrow \mathbb{R}$ – некоторая функция потерь. Выбор квадратичной функции потерь (4.5) не всегда является адекватным, так как в этом случае большие отклонения штрафуются слишком сильно. В [43] предложена универсальная функция потерь, которая не штрафует близкие значения, средние отклонения штрафует квадратично, а большие отклонения – линейно. Кроме того, эта функция является аналитической. Значения параметров, регулирующих границы между близкими, средними и большими отклонениями, могут быть найдены при помощи байесовского подхода. При этом процедура совершенно аналогична случаю с квадратичной функцией потерь.

4.3. Оптимизация границ разбиений

Пусть заданы экспертные интерпретации признаков $\mathfrak{T}$ и множество правил $\mathfrak{R}$. Тогда имеется модель алгоритмов $\mathfrak{M} = \mathfrak{M}_{\mathfrak{R}}[\{\{a_i^j\}_{j=1}^{n_i}\}_{i=1}^{d};\ \{w_m\}_{m=1}^{N}]$. На практике условные границы разбиений могут быть выбраны неадекватно, например если они выбираются как результат одномерной кластеризации данных. Введение покрытий признаков нечеткими подмножествами позволяет оптимизировать некоторый функционал качества Q по границам разбиений. Получается задача оптимизации с линейными ограничениями:

$$\begin{gathered} Q \underset{\mathbf{a}}{\longrightarrow} \max, \\ a_i^1 < \ldots < a_i^{n_i}, \quad i = \overline{1, d}. \end{gathered} \tag{4.6}$$

В качестве функционала качества выступает правдоподобие данных и модели. Как и во многих других нетривиальных моделях алгоритмов распознавания, рассматриваемый функционал качества является многоэкстремальным. Для бесконечно дифференцируемого семейства характеристических функций (1.6) обобщенный градиент вычисляется аналитически. Таким образом, для поиска приближенного решения поставленной задачи можно воспользоваться любым известным методом оптимизации, учитывающим обобщенный градиент.



Фиг. 9.

Итак, полностью изложена схема построения решения задачи распознавания образов с заданным функционалом качества. Результатом является алгоритм распознавания, принимающий решение о классификации объектов на основании знаний, сформулированных в терминах эксперта. Экспертная интерпретация признаков и правила, сформулированные относительно нее, могут быть синтезированы автоматически по заданной прецедентной информации. Впоследствии эксперт имеет возможность верифицировать и модифицировать полученные правила.

5. ЭКСПЕРИМЕНТЫ

Предложенная модель нечеткой экспертной системы (ЭС) была протестирована на нескольких задачах.

Сначала рассмотрим результаты тестирования в качестве системы для получения непрерывного прогноза. Первая задача заключается в определении мест команд в российском футбольном чемпионате за два года по турнирной таблице. Данные представляют собой таблицу с 32 объектами и 6 признаками. В качестве признаков выступают, соответственно, итоговое место в чемпионате, число побед, число ничьих, число поражений, количество забитых мячей и количество пропущенных мячей.

Необходимо определить итоговое положение команды в таблице (предсказать значение первого признака). Результаты испытаний сравнивались с множественной линейной регрессией. На фиг. 9 представлены результаты прогнозирования. Сплошная линия – точные значения, штрихпунктирная – предсказание ЭС, штриховая – предсказание линейной регрессии. Команды были отсортированы в соответствии с их итоговым положением. Несмотря на высокую корреляцию между признаками и прогнозируемой переменной (линейная зависимость между числом побед, ничьих, поражений и количеством набранных очков), для линейной регрессии среднеквадратичное отклонение составило 38.056, в то время как для ЭС оно равнялось 21.936. Кроме того, система ЭС генерировала небольшое количество легко интерпретируемых правил. Вот, например, некоторые правила, которые нашла система для футбольного чемпионата:

ЕСЛИ Пропущенных мячей = Немного И Поражений = Мало, ТО Место = Высокое

ЕСЛИ Побед = Немало И Забитых мечей = Много И Ничьих =

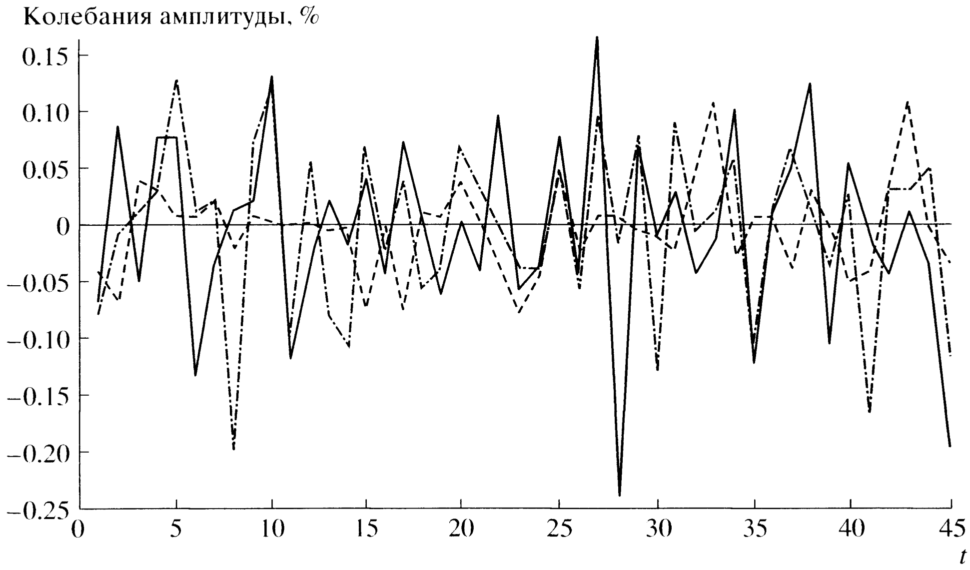
= Несколько, ТО Место = Призовое

ЕСЛИ Забитых мячей = Мало, ТО Место = Низкое

ЕСЛИ Пропущенных мячей = Немного И Ничьих = Много, ТО Место = Невысокое

Знания подобного рода особенно полезны в ситуации, когда значения некоторых признаков можно контролировать. Учитывая степень влияния этих признаков на скрытое значение прогноза, мы можем эффективно управлять процессом путем модификации значений соответствующих признаков.

Также был проведен ряд экспериментов по сравнению описанной нечеткой системы с альтернативным подходом системы MatLab fuzzy logic toolbox. Последняя представляет собой развитие так называемого нейро-нечеткого подхода (см. [14], [15], [44]). Она также позволяет генерировать автоматически нечеткие правила с использованием прецедентной информации и опреде-



Фиг. 10.

лять формы нечетких множеств. Для исследования была выбрана следующая задача: предсказание колебаний амплитуды магнитного поля в ускоряющем элементе "cavity" клистрона. Необходимые данные были получены для гамбургского линейного ускорителя в институте DESY (город Гамбург, Германия). Данная задача имеет важное практическое значение, так как точность работы ускорителя в значительной мере подвержена многочисленным внешним воздействиям (например, колебаниям земной поверхности из-за движения автомобилей, нестабильности геомагнитной обстановки, наличием различного рода шумов, состоянием атмосферы и пр.). Предсказание амплитуды колебаний магнитного поля позволяет применять нейтрализующие воздействия для предотвращения колебаний с целью добиться более надежной работы системы. В качестве исходного признакового пространства в данном случае выступали значения колебаний в других ускоряющих элементах того же клистрона. Результаты работы ЭС и MatLab на контрольной таблице представлены на фиг. 10 (колебания магнитного поля в % по времени). Сплошная линия – точные значения, штриховая – предсказание ЭС, пунктирная – MatLab. Система MatLab склонна к переобучению, несмотря на использование дополнительной выборки для его предотвращения, а также в значительной мере хуже улавливает основные пики колебаний и общую тенденцию.

Для случая задач классификации система ЭС сравнивалась с некоторыми распространенными методами распознавания – линейным дискриминантом Фишера (LDF), q-ближайших соседей (QNN), тестовым алгоритмом (TA), комитетом гиперплоскостей (LM), методом опорных векто-

Таблица 1

Метод распознавания	Меланома (малая выборка)	Фонемы (большая выборка)	Почка (средняя выборка, много классов)
MLP	65.6	78.2	77.5
LDF	56.3	77.4	77.5
TA	62.5	65.5	65.7
LM	50.0	77.2	79.3
SVM	59.4	77.4	83.1
QNN	62.5	84.7	80.3
ЭС	66.6	77.5	76.5

Таблица 2

Задача	МЭС	ДМЭС	ДМЭС$_Q$
AUSTRALIAN	19 ± 2.40	18.9 ± 1.41	18.42 ± 1.90
BUPA	43.61 ± 4.71	44.56 ± 4.64	38.16 ± 4.59
CLEVELAND	19.45 ± 2.10	19.08 ± 1.90	18.73 ± 1.82
CREDIT	18.94 ± 1.93	19.51 ± 1.77	19.26 ± 2.37
HEPATITIS	40.23 ± 5.22	39.91 ± 5.35	40.34 ± 5.26
HUNGARY	18.8 ± 1.74	18.93 ± 1.79	17.9 ± 1.83
Оценка	13	14	9

ров (SVM) многослойным перцептроном (MLP). Результаты работы (точность на контрольной выборки в %) представлены в табл. 1.

В качестве задач выступали:

МЕЛАНОМА. Определение степени заболевания по данным амбулаторного наблюдения; 33 признака, 3 класса. Всего 48 объектов для обучения, 32 – для контроля.

ФОНЕМЫ. Определение типа фонемы по ряду фильтрационных признаков сигнала; 6 признаков, 2 класса. Всего 2200 объектов для обучения, 1404 – для контроля.

ПОЧКА. Определение степени заболевания по данным амбулаторного наблюдения; 8 признаков, 7 классов. Всего 170 объектов для обучения, 150 – для контроля.

Также были проведены эксперименты по сравнению предлагаемых методов автоматической генерации знаний по прецедентной информации. Для сравнения их качества было произведено 120 экспериментов на основе 6 задач из UCI-репозитория (см. [45]). Данные для каждой задачи случайным образом разбивались на обучающую (33% всех объектов) и контрольную выборку (67% всех объектов) двадцатью различными способами. Для каждого из сравниваемых методов по обучающей выборке автоматически строился алгоритм классификации. Количество ошибок на контрольной выборке, усредненное по различным разбиениям, заносилось в соответствующую методу и задаче ячейку табл. 2. Столбец МЭС содержит результаты испытаний экспертной системы распознавания образов без оптимизации функционала качества, для которой правила генерировались методом эффективных сужений. В следующем столбце – аналогичные результаты экспертной системы распознавания образов без оптимизации функционала качества, основанной на дизъюнктивной модификации метода эффективных сужений. Столбец ДМЭС$_Q$ содержит результаты экспертной системы распознавания образов, также основанной на дизъюнктивной модификации метода эффективных сужений; веса правил и экспертные интерпретации признаков оптимизировались с точки зрения рассматриваемого функционала качества.

Для сравнения результатов в каждой строчке методам присваивалась оценка: 1 – лучший метод, 2 – следующий, 3 – худший. Суммы оценок в столбцах записаны в последней строке табл. 2.

В результате можно сделать несколько выводов. Метод эффективных сужений в основном работает не хуже, чем его дизъюнктивная модификация. Ввиду того что множество правил последнего метода всегда больше, ДМЭС склонен к переобучению. Однако оптимизация экспертных интерпретаций признаков и весов правил приводит к нужному улучшению результатов. Отсутствие переобучения на этапе оптимизации может быть объяснено наличием ограничений в формуле (4.6), связанных с наличием порядка между нечеткими подмножествами, составляющими экспертную интерпретацию каждого признака.

6. ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

Данные испытаний показали, что предложенный метод построения нечеткой экспертной системы может быть использован для поддержки принятия решений в тех случаях, когда не существует строгой математической модели изучаемой области знаний, но имеется некоторая априорная информация и количество прецедентов не позволяет использовать нелинейные статистические методы. По-видимому, данный метод не стоит применять в тех случаях, когда необходимо получить абсолютно точное значение прогноза и/или существует строгая математическая модель явления.

Следует заметить, что оптимизация набора нечетких правил, по сути, представляет собой подбор колоколообразной ядровой функции. Как уже отмечалось выше при оптимизации весов правил, предлагаемая модель близка к обобщенным линейным моделям, где в качестве элементарных признаков фигурируют нечеткие правила. Использование $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$- и $[\boldsymbol{\alpha}, \boldsymbol{\beta}]$-покрытий фактически означает подбор колоколообразной ядровой функции. Решение подобной проблемы в ядровых методах распознавания, таких как методы опорных и релевантных векторов, является очень трудоемкой задачей (см. [46], [47]). В данном случае ее удается избежать за счет конструирования набора правил из элементарных блоков, задаваемых экспертом либо определяемых в результате кластеризации.

СПИСОК ЛИТЕРАТУРЫ

1. *Заде Л.* Размытые множества и их применение в распознавании образов и кластер-анализе // Классификация и кластер. М.: Мир, 1980. С. 208–247.
2. *Журавлев Ю.И.* Избранные научные труды. М.: Магистр, 1998.
3. *Журавлев Ю.И., Рязанов В.В., Сенько О.В.* РАСПОЗНАВАНИЕ. Математические методы. Программная система. Практические применения. М.: Фазис, 2006.
4. *Bishop C.M.* Neural networks for pattern recognition. Oxford: Univ. Press, 1995.
5. *Breiman L., Friedman J.H., Olshen R.A., Stone C.J.* Classification and regression trees. Belmont, CA: Wadsworth Internal. Group, 1984.
6. *Перфильева И.* Приложения теории нечетких множеств // Итоги науки и техн. Сер. Теория вероятностей. Матем. статистика. Теоретич. кибернетика. М.: ВИНИТИ, 1990. Т. 29. С. 83–151.
7. *Тэрано Т., Асаи К., Сугено М.* Прикладные нечеткие системы. М.: Мир, 1993.
8. *Mamdani E.* Advances in the linguistic synthesis of fuzzy controllers // Proc. 6th Internat. Symp. Multiple-Values Logic, 1976. P. 196–202.
9. *Заде Л.* Понятие лингвистической переменной и его применение к принятию приближенных решений. М.: Мир, 1980.
10. *Fukami S., Mizumoto M., Tanaka K.* Some considerations of fuzzy conditional inference // Fuzzy Sets and Systems. 1980. V. 3. P. 243–273.
11. *Разанов В.В., Сенько О.В.* О некоторых моделях голосования и методах их оптимизации // Распознавание, классификация, прогноз. 1990. Т. 3. С. 106–145.
12. *Воронцов К.В.* Лекции по логическим алгоритмам классификации; http://www.ccas.ru/voron/download/LogicAlgs.pdf.2006.
13. *Паклин Н.Б.* Адаптивные модели нечеткого вывода для идентификации нелинейных зависимостей в сложных системах: Дис. ... канд. техн. наук. Ижевск: ИжГГУ, 2004.
14. *Ojala T.* Neuro-fuzzy systems in control. Master of science thesis. Tampere. Finland, 1994.
15. *Jang J.-S.R., Sun C.-T., Mizutani E.* Neuro-fuzzy and soft computing. Upper Saddle River, NJ: Prentice-Hall, 1997.
16. *Ishibuchi H., Nozaki K., Yamamoto N., Tanaka C.* Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms // Fuzzy Sets and System. 1994. V. 65. № 2/3. P. 237–253.
17. *Inoue H., Kamei K., Inoue K.* Rule pairing methods for crossover in GA for automatic generation of fuzzy control rules; citeseer.ist.psu.edu/200265.html.
18. *Gomez-Scarmeta A.F., Jimenez F.* Generating and tuning fuzzy rules using hybrid systems // Proc. 6th IEEE Internat. Conf. Fuzzy System. 1997. V. 1. P. 247–252.
19. *Rivest R.L.* Learning decision lists // Mach. Learn. 1987. V. 2. № 3. P. 229–246.
20. *Cohen W.W.* Fast effective rule induction // Proc. 12th Internat. Conf. Mach. Learn. CA: Morgan Kaufmann, 1995. P. 151–163.
21. *Freund Y., Schapire R.E.* Experiments with a new boosting algorithm // Proc. 13th Internat. Conf. Mach. Learn., 1996. P. 148–156.
22. *Cohen W.W., Singer Y.* A simple, fast, and effective rule learner // Proc. 16th Nat. Conf. Artificial Intelligence, 1999.
23. *Ветров Д.П., Кропотов Д.А.* Использование методов Boosting для генерации знаний // Тр. XII Всерос. конф. "Матем. методы распознавания образов". М.: Макс-Пресс, 2005. С. 48–51.
24. *Quinlan J.R.* C4.5: Programs for machine learning. San Mateo, CA: Morgan Kaufmann, 1993.
25. *Breslow L.A., Aha D.W.* Simplifying decision trees: a survey // Knowledge Eng. Rev. 1997. V. 12. № 1. P. 1–40. siteseer.ist.psu.edu/breslow96simplifying.html.
26. *Дюличева Ю.Ю.* Модели коррекции редуцированных бинарных решающих деревьев: Дис. канд. физ.-матем. наук. Симферополь: ТавРГУ, 2004.
27. *Дуда Р., Харт П.* Распознавание образов и анализ сцен. М.: Мир, 1976.

28. *Corduneanu A., Bishop C.* Variational model selection for mixture distributions // Artificial Intelligence and Statistics. Morgan Kaufmann, 2001. P. 27–34.
29. *Шумский С.А.* Байесова регуляризация обучения // Лекции по нейроинформатике. Ч. 2. М.: МИФИ, 2002. С. 30–93.
30. *Bishop C.M., Svensen M.* Robust Bayesian mixture modeling // Proc. ESANN, 2004.
31. *Ben-Hur A., Elisseeff A., Guyon I.* A stability based method for discovering structure in clustered data // Proc. Symposium on Biocomputing. Lihue, Hawaii, 2002. P. 6–17.
32. *Kuncheva L.* Combining pattern classifiers: methods and algorithms. Wiley, 2004.
33. *Гуров С.И.* Оценка надежности классифицирующих алгоритмов. М.: Издат. отд. ВМиК МГУ, 2002.
34. *MacKay D.J.C.* Information theory, inference, and learning algorithms. Cambridge Univ. Press, 2003.
35. *Kropotov D.A., Tolstov I.V., Vetrov D.P.* Decision trees regularization based on stability principle // Pattern Recognition and Image Analysis. 2005. V. 15. № 1. P. 107–109.
36. *Berger J.O.* Statistical decision theory and bayesian analysis. Berlin etc: Springer, 1985.
37. *MacKay D.J.C.* Bayesian interpolation // Neural Computation. 1992. V. 4. № 3. P. 415–447.
38. *Tipping M.E.* Sparse bayesian learning and the relevance vector machine // J. Mach. Learn. Res. 2001. V. 1. P. 211–244.
39. *Burges C.* A tutorial on support vector machines for pattern recognition // Data Mining and Knowledge Discovery. 1998. V. 2. P. 121–167.
40. *Vapnik V.N.* Statistical learning theory. Wiley, 1998.
41. *Williams C.K.I.* Prediction with gaussian processes: from linear regression to linear prediction and beyond // Learning in Graphical Models. MIT. 1999. P. 599–621.
42. *Williams P.M.* Bayesian regularization and pruning using a laplace prior // Neural Computation. 1995. V. 7. № 1. P. 117–143.
43. *Chu W.* Bayesian approach to support vector machines: PhD thesis. Nat. Univ. of Singapore, 2003.
44. *Дьяконов В., Круглов В.* Математические пакеты расширения MATLAB. Специальный справочник. М.: Питер, 2001.
45. *Murphy P., Aha D.* UCI repository of machine learning databases // Univ. California, Dept. Informat. and Comput. Sci. California: Irvine, 1996; http://www.ics.uci.edu/~mlearn/MLRepository.html.
46. *Kropotov D.A., Ptashko N.O., Vetrov D.P.* The use of bayesian framework for kernel selection in vector machines classifiers // Progress in Pattern Recognition, Image Analysis and Applic. LNCS 3773. Berlin etc: Springer, 2005. P. 252–261.
47. *Kropotov D.A., Vetrov D.P., Ptashko N.O., Vasiliev O.M.* The use of stability principle for kernel determination in relevance vector machines // ICONIP2006, Part I, LNCS 423. Berlin etc: Springer, 2006. P. 727–736.