

УДК 519.711.3

## ПОСТРОЕНИЕ МНОГОМЕРНОГО НЕЧЕТКОГО РАСПОЗНАВАТЕЛЯ НА ОБУЧАЮЩЕМ МНОЖЕСТВЕ

А.Р. САМКО, Н.Л. БОБРОВА, О.В. ГЕРМАН

Белорусский государственный университет информатики и радиоэлектроники  
П. Бровки, 6, Минск, 220013, Беларусь

Поступила в редакцию 27 сентября 2011

Работа содержит описание модели многомерного нечеткого распознавателя на основе классифицирующего дерева, которое обеспечивает следующие преимущества: модель не использует функции нечеткой меры для оценки степени сходства нечетких векторов; модель ориентирована на вход векторов и не зависит от взаимного или группового соотношения элементов входных векторов. Это обстоятельство делает ненужной процедуру обучения, которая имеет решающее значение в моделях нейро-сетей с экспертом

*Ключевые слова:* распознавание, нечеткий распознаватель, классифицирующее дерево.

### Введение

Проблема реализации нечеткого распознавателя может решаться с разных позиций, которым присущи те или иные ограничения.

Часто применяют методы нечеткого логического типа Мамдани, Сугено, Ларсена и т.д. [1, 2]. Эти методы требуют определения меры близости входного экземпляра  $x' = \langle x'_1, \dots, x'_n \rangle$ , и посылочной части каждого логического правила, представленного в форме «если  $x_1 = a_1 \& \dots \& x_n = a_n$ , то  $y(\mu_y)$ ».

Данная задача не имеет простого (видимого) решения тогда, когда диапазоны изменения значений  $x_i (i = \overline{1, n})$  не известны заранее.

Можно рассмотреть подход на основе нечетких нейро-сетевых моделей [3]. Его ограничение состоит в том, что функция распознавания может быть весьма сложной, так что-либо процесс обучения окажется невозможным с практической точки зрения, либо качество распознавателя будет низким.

Большая часть алгоритмов нечеткого распознавания [4] использует функцию нечеткой меры для определения расстояния от входного образца до произвольного кластера. Выбор функций нечеткой меры в значительной степени субъективен и эвристичен.

В настоящей работе предлагается модель нечеткого многомерного распознавателя на основе классифицирующего дерева, исследование которого для четких объектов сообщалось в [5, 6]. С некоторыми модификациями это модель вполне может служить целям нечеткого распознавания. Достоинства описываемой модели таковы:

- она не использует функции нечеткой меры для оценки близости нечетких объектов;
- она ориентирована на большую размерность входных объектов и не критична к наличию индивидуальной или групповой корреляции разрядов входных объектов. Это обстоятельство отменяет необходимость в обучении распознавателя в смысле, принятом в нейро-сетевых моделях.

## Классифицирующее дерево

Ограничимся фрагментарным описанием концепции иерархического классифицирующего дерева, поскольку детали его построения можно найти в [5, 6].

В качестве примера сошлемся на рис. 1.

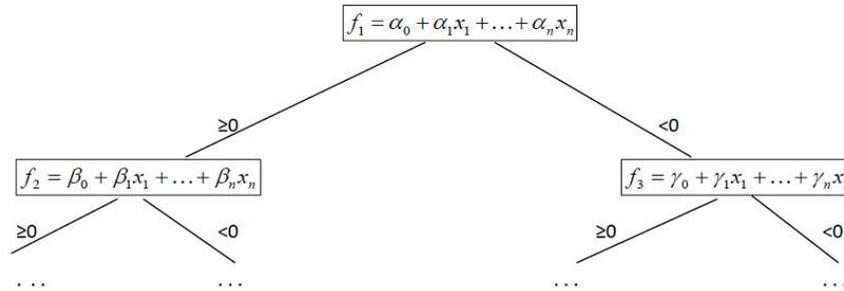


Рис. 1. Иерархическое классифицирующее дерево

В узлах дерева помещены линейные дискриминаторные функции вида  $f_v = \sum_{i=0}^n \mu_i x_i$  (где  $x_0=1$ ). Пусть на вход системы поступает объект  $x' = \langle x'_1, \dots, x'_n \rangle$ . Вычисляется значение  $f_1 = \alpha_0 + \alpha_1 x_1 + \dots + \alpha_n x_n$  в корневом узле дерева. Если это значение  $\geq 0$ , то переходим в вершину для вычисления  $f_2 = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$ , в противном случае – в вершину  $f_3 = \gamma_0 + \gamma_1 x_1 + \dots + \gamma_n x_n$ . Процесс последовательно распространяется на последующие вершины дерева, выбор которых зависит от значения функции, вычисленного в родительской вершине. Рано или поздно попадаем в вершину, являющуюся тупиковой (листовой), которая однозначно определяет, к какому классу принадлежит объект  $x$ .

Построение дерева и получение дискриминаторных функций в его узлах основаны на использовании метода устранения невязок [7] как частного варианта симплексного алгоритма для задачи линейного программирования.

Иерархическое классифицирующее дерево строится на основе обучающего множества, представленного экспертом (-ами). Типичный вид обучающего множества приведен в табл. 1.

Таблица 1. Обучающее множество

NN	$x_1$	$x_2$	$\dots$	$x_n$	Кластер
1	$x_1^1$	$x_2^1$		$x_n^1$	$K_1$
2	$x_1^2$	$x_2^2$		$x_n^2$	$K_2$
3	$x_1^3$	$x_2^3$		$x_n^3$	$K_2$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$N$	$x_1^N$	$x_2^N$		$x_n^N$	$K_1$

Количество кластеров  $P$  в общем случае более двух; в «базовом» случае  $P=2$  (случай  $P>2$  просто сводится к построению нескольких деревьев, для каждого из которых  $P=2$ ).

Исследование зависимости размера дерева от характера вероятностного распределения значений входных векторов, вида пересечений различных кластеров представляют самостоятельную задачу. Отметим, что эти размеры во многих практических ситуациях определяют дерево с числом ярусов, не превосходящих несколько единиц даже для обучающих таблиц с сотней строк.

### Задача нечеткой многомерной классификации

Обучающее множество зададим в качестве иллюстраций в виде табл. 2 (случай трехмерных объектов).

Таблица 2. Обучающее множество трехмерных объектов

NN	$x_1$	$x_2$	$x_3$	Кластер	Мера принадлежности ( $\mu$ )
1	0	1	-2	A	1,0
2	0	0	-1	A	0,8
3	-1	-5	-4	B	1,0
4	2	5	3	B	0,9
5	-1	-1	0	A	0,6
6	3	3	7	B	1,0
7	-1	1	-1	A	1,0

Здесь семь объектов, для которых указаны значения нечеткой меры ( $\mu_i$ ), принадлежности к соответствующему кластеру.

Будем интерпретировать нечеткую меру принадлежности кластеру, как, например, субъективную вероятность, с которой эксперт относит данный объект в соответствующий кластер.

Мы хотим построить на основании табл. 2 классифицирующее дерево. Качество распознавания можно связать со статистическим критерием  $\chi^2$  таким образом. Пусть  $q_k=0$ , если объект с номером  $k$  из табл. 2 отнесен в кластер  $B$ , и  $q_k=1$ , если объект отнесен в кластер  $A$ . Вычисляем критерий  $\chi^2$ :

$$\chi_p^2 = \sum_{\substack{r=1 \\ \mu_k \neq 0}}^N \frac{(\mu_k - q_k)^2}{\mu_k}. \quad (1)$$

Чтобы считать распознавание удовлетворительным с точки зрения (1), необходимо

$$\chi_p^2 \leq \chi_{\text{табл}}^2(\alpha, n^{\text{св}}), \quad (2)$$

где  $\alpha$  – выбранная вероятность ошибки, например,  $\alpha=0,05$ ;  $n^{\text{св}}$  – число степеней свободы.

Число степеней свободы  $n^{\text{св}}$  зависит от количества независимых коэффициентов в линейной дискриминаторной функции  $f_v = \sum_{i=0}^r \alpha_i x_i$ . Проблема состоит в том, что в классифицирующем дереве несколько узлов, в общем случае, и для каждого узла используется собственная дискриминаторная функция, определенная на собственном обучающем множестве. Пусть мощность этого обучающего множества есть  $M_z$ ,  $f_z = \sum_{i=0}^r \alpha_i x_i$ . Тогда число степеней свободы

$$n_z^{\text{св}} = M_z - r. \quad (3)$$

Из (2) следует, что наилучшее качество обеспечивается тогда, когда

$$\begin{cases} \mu_k \geq 0,5 \rightarrow q_k = 1 \\ \mu_k < 0,5 \rightarrow q_k = 0 \end{cases} \quad (4)$$

Это условие положено в основу нечеткого распознавателя, описанного в [8], однако метод построения такого распознавателя при большом числе объектов в обучающем множестве может оказаться крайне трудоемким, т.к. связан с решением  $NP$ -полной задачи.

Данная статья излагает метод, который позволяет сохранить стратегию устранения невязок в качестве основы построения нечеткого классифицирующего дерева, отказавшись от требования (4), но сохранив условие (2). Таким образом, даже если обучающаяся таблица содержит сотни объектов, сложность построения нечеткого классифицирующего дерева не выходит за пределы сложности симплексного алгоритма.

## Построение нечеткого классифицирующего дерева

Ограничим рассмотрение построением одного узла. По данным табл.2 составляем неравенства с неизвестными коэффициентами  $\alpha_0, \alpha_1, \alpha_2, \alpha_3$  :

$$\begin{aligned}
 \alpha_0 + \alpha_1 \cdot 0 + \alpha_1 \cdot 1 + \alpha_3 \cdot (-2) &\geq 0 & (1,0) \\
 \alpha_0 + \alpha_1 \cdot 0 + \alpha_2 \cdot 0 + \alpha_3 \cdot (-1) &\geq 0 & (0,8) \\
 \alpha_0 - \alpha_1 \cdot 1 - \alpha_2 \cdot 5 - \alpha_3 \cdot 4 &< 0 & (1,0) \\
 \alpha_0 - \alpha_1 \cdot 2 + \alpha_2 \cdot 5 + \alpha_3 \cdot 3 &< 0 & (0,9) \\
 \alpha_0 + \alpha_1 \cdot (-1) + \alpha_2 \cdot (-1) + \alpha_3 \cdot 0 &\geq 0 & (0,6) \\
 \alpha_0 + \alpha_1 \cdot 3 + \alpha_2 \cdot 3 + \alpha_3 \cdot 7 &< 0 & (1,0) \\
 \alpha_0 + \alpha_1 \cdot (-1) + \alpha_2 \cdot 1 + \alpha_3 \cdot (-1) &\geq 0 & (1,0)
 \end{aligned} \tag{5}$$

Каждое неравенство получаем в виде  $\sum_{i=0}^3 \alpha_i \cdot x_i \geq 0$

где  $x_0' = 1$ ,  $x_i'$  – координата вектора из табл. 2. Каждому неравенству приписывается соответствующая мера нечеткости. Далее заменяем каждое жесткое неравенство ( $<$ ) нежестким ( $\leq$ ) с помощью дополнительного малого слагаемого и приводим все неравенства к виду  $\geq$ . Заметим, что в (5) неравенства вида  $\geq 0$  относятся к кластеру  $A$ , а неравенства вида  $< 0$  – к кластеру  $B$ . Стартовая система имеет такой вид:

$$\begin{aligned}
 \alpha_0 + \alpha_1 \cdot 0 + \alpha_1 - \alpha_3 \cdot 2 &\geq 0 & (1,0) \\
 \alpha_0 + \alpha_1 \cdot 0 + \alpha_2 \cdot 0 - \alpha_3 \cdot 1 &\geq 0 & (0,8) \\
 -\alpha_0 + \alpha_1 + 5 \cdot \alpha_2 + 4 \cdot \alpha_3 &\geq 1 & (1,0) \\
 -\alpha_0 + 2 \cdot \alpha_1 - 5 \cdot \alpha_2 - 3 \cdot \alpha_3 &\geq 1 & (0,9) \\
 \alpha_0 - \alpha_1 - \alpha_2 + \alpha_3 \cdot 0 &\geq 0 & (0,6) \\
 -\alpha_0 - 3 \cdot \alpha_1 - 3 \cdot \alpha_2 - 7 \cdot \alpha_3 &\geq 1 & (1,0) \\
 \alpha_0 - \alpha_1 + \alpha_2 - \alpha_3 &\geq 0 & (1,0)
 \end{aligned} \tag{6}$$

Стратегия устранения невязок (СУН) реализуется в две стадии. На первой стадии система неравенств приводится к виду с четырьмя (по числу переменных) базовыми неравенствами вида  $\alpha_k' \geq 0$ , устанавливающими неотрицательность переменных. С этой целью строим замены для переменных из неравенств с положительной правой частью. Например, из

$$-\alpha_0 + \alpha_1 + 5 \cdot \alpha_2 + 4 \cdot \alpha_3 \geq 1$$

получим

$$\begin{aligned}
 -\alpha_0 &\geq 1 - \alpha_1 - 5 \cdot \alpha_2 - 4 \cdot \alpha_3 \\
 -\alpha_0 &= 1 - \alpha_1 - 5 \cdot \alpha_2 - 4 \cdot \alpha_3 + \omega_0
 \end{aligned}$$

И далее  $\alpha_0 = -1 + \alpha_1 + 5 \cdot \alpha_2 + 4 \cdot \alpha_3 - \omega_0$ . Проводим эту замену в систему (6). Затем аналогичным образом заменяем  $\alpha_1, \alpha_2$  и  $\alpha_3$ . В результате стадия 1 заканчивается следующей системой:

$$\begin{aligned}
 \omega_1 &\geq 0 & (1,0) \\
 \omega_3 &\geq 0 & (0,8) \\
 \omega_0 &\geq 1 & (1,0) \\
 \frac{2}{3} \omega_1 - \frac{9}{8} \omega_0 - 9 \cdot \omega_3 + \frac{7}{2} \omega_2 &\geq \frac{17}{8} & (0,9) \\
 \omega_2 &\geq 0 & (0,6) \\
 -\frac{11}{8} \omega_0 + \frac{11}{2} \omega_1 + \frac{13}{2} \omega_2 - 19 \cdot \omega_3 &\geq \frac{19}{8} & (1,0) \\
 -\frac{7}{8} \omega_0 + \frac{3}{2} \omega_1 + \frac{9}{2} \omega_2 - 3 \omega_3 &\geq \frac{3}{8} & (1,0)
 \end{aligned} \tag{7}$$

На второй стадии СУН последовательно устраняются невязки.

1. Невязкой называется неравенство ( $\geq$ ) с положительной правой частью.
2. Стоп-невязкой называется невязка, в левой части которой нет положительных коэффициентов при переменных.

СУН требует выражать из невязки переменную с положительным коэффициентом. Для случая нечеткой системы неравенств в первую очередь используем невязки с мерой (1,0). Так, из предпоследнего неравенства в (7) получим

$$\omega_1 = \frac{19}{44} + \frac{1}{4}\omega_0 - \frac{13}{11}\omega_2 + \frac{38}{11}\omega_3 + z_1 \quad (8)$$

Подставляя (8) в (7), получим новую систему

$$\frac{1}{4}\omega_0 - \frac{13}{11}\omega_2 + \frac{38}{11}\omega_3 + z_1 \geq -\frac{19}{44} \quad (1,0)$$

$$\omega_3 \geq 0 \quad (0,8)$$

$$\omega_0 \geq 1 \quad (1,0)$$

$$-\frac{3}{4}\omega_0 + \frac{19}{11}\omega_2 - \frac{42}{11}\omega_3 + \frac{3}{2}z_2 \geq \frac{65}{44} \quad (0,9) \quad (9)$$

$$\omega_2 \geq 0 \quad (0,6)$$

$$z_3 \geq 0 \quad (1,0)$$

$$-\frac{1}{2}\omega_0 + \frac{30}{11}\omega_2 - \frac{24}{11}\omega_3 + \frac{3}{2}z_1 \geq -\frac{3}{11} \quad (1,0)$$

Хотя СУН заканчивает итерации, когда нет невязок, в случае нечеткой системы неравенств мы ориентируемся на следующее условие. Если невязок нет среди неравенств с мерой (1,0), то СУН останавливает итерации.

Мы как раз имеем это условие. Подставим в систему (9) нулевые значения для переменных:  $\omega_0 = 0, \omega_2 = 0, \omega_3 = 0, z_1 = 0$ . Все неравенства будут выполнены, кроме четвертого по порядку с мерой (0,9). Проверим значение критерия  $\chi_p^2$ . Вычисляем  $\chi_p^2$  только (!) для нечетких неравенств, т.к. все четкие неравенства выполнены.

$$\chi_p^2 = \frac{(0,8-1)^2}{0,8} + \frac{(0-0,9)^2}{0,9} + \frac{(1-0,6)^2}{0,6} = 1,21$$

Табличное значение  $\chi_m^2$  находим для  $\alpha=0,05$  и числа степеней свободы  $n^{cb} = 7 - 4 = 3$  (если неравенство использовано для описания четырех коэффициентов). Имеем  $\chi_m^2 = 7,81 > \chi_p^2$ . Следовательно, найденное нами решение статистически значимо (адекватно) и должно быть принято. Отправляясь от «решения»  $\omega_0 = 0, \omega_2 = 0, \omega_3 = 0, z_1 = 0$ , получим (на основании произведенных подстановок)

$$\alpha_0 = -\frac{1}{11} \quad \alpha_1 = -\frac{19}{44} \quad \alpha_2 = \frac{15}{44} \quad \alpha_3 = -\frac{1}{11}$$

Таким образом, нами получено линейное дискриминаторное неравенство в корневом узле дерева вида

$$-\frac{1}{11} - \frac{19}{44}x_1 + \frac{15}{44}x_2 - \frac{1}{11}x_3 \quad (\geq 0)$$

Полученное дерево состоит из одного корневого узла.

При выполнении СУН для системы с нечеткими неравенствами используем следующие (кроме приведенных) правила:

- если стоп-невязка возникла в неравенстве с мерой  $\mu=1$  (неравенство четкое), то это неравенство исключается из системы, как и в случае четких неравенств [5, 6];

- если стоп-невязка возникла в неравенстве с нечеткой мерой  $\mu < 1$ , а среди четких неравенств ( $\mu = 1$ ) нет невязок, то проверяем критерий  $\chi_p^2 \leq \chi_m^2$ . При выполнении критерия СУН завершает операции, и получаем линейную дискриминаторную функцию для рассматриваемого узла. Если  $\chi_p^2 > \chi_m^2$ , то стоп-невязка удаляется из числа неравенств текущей системы;

- если невязок вообще нет, но  $\chi_p^2 > \chi_m^2$ , то последовательно заменяем каждое нечеткое неравенство с мерой  $\mu < 1$  набором четких неравенств (с  $\mu = 1$ ), как показано на следующем примере.

Пусть, например, нечеткое неравенство таково:

$$\varepsilon_1 \cdot x_1 + \varepsilon_2 \cdot x_2 + \dots + \varepsilon_n \cdot x_n \geq \varepsilon_0 \quad (0,6) \quad (10)$$

Мера  $\mu = 0,6$  интерпретируется как «факт», что в среднем в трех случаях из пяти неравенство выполняется, а в двух – нет. Сгенерируем случайным образом пять векторов  $\varepsilon_i$  в  $\Delta$ -окрестности точки  $\langle \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n \rangle$  для достаточно малого  $\Delta$  ( $\Delta \approx 0,05 \cdot |\varepsilon_i|$  для коэффициента  $\varepsilon_i$ ). Из этих векторов получим коэффициенты для трех неравенств смысла  $\geq \varepsilon_0$  и двух – смысла  $< \varepsilon_0$ . Нечеткое неравенство (10) заменяем на пять четких неравенств.

Из исключенных неравенств (стоп-невязок) формируем новую систему и выполняем СУН на этой системе для формирования нового узла классифицирующего дерева, в который будет выполнен переход при выполнении дискриминаторного неравенства в родительском узле [5, 6].

### Заключение

Изложенная техника естественным образом обобщает метод построения классифицирующего дерева для четких объектов [5, 6]. Анализ качества распознавания увязан с критерием  $\chi_1^2$ , что ориентирует систему на большое число объектов в обучающем множестве. Достоинства модели распознавателя связаны с отсутствием привязки к закону распределения и наличию корреляционной связи между группами разрядов (признаков).

## BUILDING MULTIDIMENSIONAL FUZZY RECOGNIZER ON A LEARNING SET

A.R. SAMKO, N.L. BOBROVA, O.V. GERMAN

### Abstract

The paper contains a model of a multidimensional fuzzy recognizer on the basis of a classifying tree which provides the following advantages: the model does not use a fuzzy measure function to evaluate a degree of similarity of fuzzy vectors; the model is oriented at highly dimension input vectors and does not depend on the mutual or group correlation of the elements of input vectors. This circumstance makes not necessary a learning procedure which is crucial in neuro-nets with a teacher model.

### Список литературы

1. Леоненков А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH. СПб., 2005.
2. Ярушклина Н.Г. Основы теории нечетких и гибридных систем. М., 2004.
3. Яхьяева Г.Э. Нечеткие множества и нейронные сети. М., 2006.
4. Вятчинин Д.А. Нечеткие методы автоматической классификации. Мн., 2004.
5. Герман О.В., Дорожкина Н.Н. // Вестник Ставропольского университета. 1999. Вып. 20. С. 85–99.
6. Герман О.В., Боброва Н.Л., Самко А.Р. // Докл. БГУИР. 2011. №4(58). С. 86–93.
7. Герман О.В., Дорожкина Н.Н., Самко А.Р. // Труды БГТУ. 2007. Сер. VI. Вып. XV. С. 160–164.
8. Герман О.В., Дорожкина Н.Н., Самко А.Р. // Труды БГТУ. 2007. Сер. IX. Вып. XVI. С. 116–118.