

УДК 519.711.3

МОДЕЛЬ ПРОГНОЗИРОВАНИЯ С ИСПОЛЬЗОВАНИЕМ НЕЧЕТКОГО ОБУЧАЮЩЕГО МНОЖЕСТВА

Ю.О. GERMAN, О.В. GERMAN

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

Поступила в редакцию 7 апреля 2016

Рассматривается задача построения численного прогноза с использованием обучающего нечеткого множества. Поставленная в статье общая проблема связана с доопределением значений нечеткого вектора и его оценкой. При этом решаются две задачи: разработка метода для генерации нечеткого прогнозного значения с ожидаемым (в статистическом смысле) значением нечеткой меры и оценка качества прогноза. Представленный математический аппарат базируется на технике построения четкого многомерного классификатора и его использования для определения нечеткой меры принадлежности с последующей оценкой вероятностей нечетких векторов по Р. Ягеру.

Ключевые слова: нечеткое обучающее множество, мера принадлежности, классификация, прогноз.

Введение

Рассматривается задача построения численного прогноза с использованием обучающего нечеткого множества. Известны нечеткие классификаторы [1, 2]. Поставленная в статье задача расширяет рамки проблемы – нужно доопределить нечеткий вектор и дать его оценку (указать наиболее вероятное значение нечеткой меры принадлежности, т.е. математическое ожидание нечеткой меры). Известны подходы к прогнозированию нечетких временных рядов [3–5]. Они используют различный математический аппарат – генетические алгоритмы, функции распределения нечетких значений, нейронные сети и др. При этом качество прогнозирования в значительной степени увязывается с качеством используемой математической модели. Здесь имеются принципиальные моменты: оценка качества модели в общем случае, обеспечение статистически адекватного прогноза. Первый вопрос, вообще говоря, увязан со вторым. Если принципиально модель регулируема, то ее правильная коррекция позволит добиться требуемого качества прогноза. В противном случае модель следует заменить. Нечеткие данные описываются функцией меры принадлежности. Первая задача – эту функцию найти и статистически ее обосновать. По найденной функции легко найти меру прогнозного значения, а затем – само прогнозное значение. Можно обозначить две стороны этой проблемы. Во-первых, нужно определить метод для генерации нечеткого прогнозного значения с ожидаемым (в статистическом смысле) значением нечеткой меры. Эту задачу обозначим **31**. Во-вторых, нужно оценить качество прогноза (задача **32**). Цель последующего – изложить решение обеих задач.

Определение нечеткого прогнозного значения

Пусть дана следующая табл. 1 с нечеткими векторами, в которой представлены 8 векторов и значения меры их принадлежности к некоторому нечеткому множеству (скажем, A) и к его дополнению ($\sim A$). Рассматриваем только непрерывные случайные величины в качестве разрядов нечетких векторов. Для каждой случайной величины (вообще говоря) закон

распределения а priori не известен. Диапазон изменения случайной величины может быть известным или нет.

Таблица 1. Обучающая таблица T

x	x_1	x_2	x_3	μ_A	$\mu_{\sim A}$
1	1	4	8	0,6	0,4
2	2	2	5	0,8	0,2
3	3	3	2	1	0
4	4	3	1	0,9	0,1
5	4	2	1	0,8	0,2
6	5	1	2	0,6	0,4
7	5	2	2	0,55	0,45
8	5	3	1	0,55	0,45

Столбец $\mu_{\sim A}$ указывает меру принадлежности к дополнительному множеству $\sim A$. Пусть дан вектор $x_z = \langle 3, 5, ? \rangle$. Нужно доопределить (спрогнозировать) недостающее значение третьего разряда и указать меру принадлежности полученного вектора. При этом нужно указать такое значение для третьего разряда, при котором мера принадлежности к множеству A была бы максимальной из всех возможных вариантов.

Решение задачи 31

Для решения задачи 31 будем использовать модель нечеткого многомерного классификатора, предложенного в [6]. Эта модель, в свою очередь, базируется на формализме четкого многомерного классификатора [7]. Представляется необходимым для понимания смысла настоящей работы дать краткое введение по проблеме. Прежде всего от «нечеткой» таблицы T перейдем к четкой TT (табл. 2).

Таблица 2. «Четкая» таблица TT

x	x_1	x_2	x_3	μ_A	Y
1	1	4	8	0,6	+
2	1	4	8	0,4	-
3	2	2	5	0,8	+
4	2	2	5	0,2	-
5	3	3	2	1	+
6	3	3	2	0	-
7	4	3	1	0,9	+
8	4	3	1	0,1	-
9	4	2	1	0,8	+
10	4	2	1	0,2	-
11	5	1	2	0,6	+
12	5	1	2	0,4	-
13	5	2	2	0,55	+
14	5	2	2	0,45	-
15	5	3	1	0,55	+
16	5	3	1	0,45	-

Новый столбец Y указывает класс, к которому принадлежит многомерный объект. Здесь два класса (+ и - (равносильно A и $\sim A$)). Значение нечеткой меры (μ_A) становится четвертым разрядом векторов x , причем если $\mu_A \geq 0,5$, то объект относится к классу A . Четкий классификатор для TT представляет в общем случае дерево. Узлы дерева (для нашего примера) представляют линейные алгебраические неравенства вида

$$z = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_3 + a_4 \cdot \mu. \quad (1)$$

Коэффициенты неравенств (a) (1) находятся с помощью процедуры устранения невязок [7] по обучающей таблице TT (и ее производным). Если коэффициенты a_k ($k=0,4$) известны, то, подставляя значения вместо переменных x_k, μ , получим значение z , причем, если $z \geq 0$, то «идем» по дереву по ветви влево, иначе – по ветви вправо. Если узел конечный, то при $z \geq 0$ объект относится к классу A , иначе к классу $\sim A$. Наш пример дает дерево с

единственной корневой вершиной и соответствующим алгебраическим неравенством (решение получено в EXCEL Solver).

$$z = 0,923 - 0,368 \cdot x_1 - 0,164 \cdot x_2 - 0,164 \cdot x_3 + 3,273 \cdot \mu \geq 0. \quad (2)$$

Подставляя, например, данные из первой и второй строк таблицы *TT*, получим

$$z = 0,923 - 0,368 \cdot 1 - 0,164 \cdot 4 - 0,164 \cdot 8 + 3,273 \cdot 0,6(0,4) = 0,54(-0,10).$$

Положительное число $z = 0,54$ соответствует классу A , отрицательное $(-0,10)$ классу $\sim A$. Итак, строим прогнозный ряд для вектора $\langle 3,5,?,? \rangle$. Для этого вектора не известна третья координата и мера принадлежности к множеству A . Из таблицы *TT* видим, что третья координата изменяется в диапазоне $[1;8]$. Мы далее предполагаем, что диапазон изменения переменной известен. В противном случае потребуется усложнить технику расчетов. Выберем в этом диапазоне $n > 2$ последовательно возрастающих равноудаленных значений (чем больше n , тем точнее ожидаемое прогнозное значение). Например, возьмем $n = 10$:

$$1; 1,77; 2,45; 3,24; 4,02; 4,8; 5,58; 6,36; 7,13; 8. \quad (3)$$

Возьмем теперь первый вектор $\langle 3, 5, 1 \rangle$ ($x_1 = 3, x_2 = 5, x_3 = 1$). Для этого вектора нужно найти меру принадлежности к множеству A . Опираясь на работы [6, 7], выполняем следующий эксперимент. Последовательно рассматриваем ряд значений меры принадлежности μ_A , начиная с $\mu_A = 0$ и каждый раз увеличивая μ_A на δ , где δ – достаточно малая величина, например, $\delta = 0,1$. Для всех векторов, получаемых в ходе эксперимента, определяем принадлежность к множеству A на базе оценки (2), полученной выше. Заносим данные в таблицу *TTT* (табл. 3).

Таблица 3. Экспериментальная таблица *TTT* для прогнозного вектора $\langle 3,5,1 \rangle$

x	x_1	x_2	x_3	μ_A	Y
1	3	5	1	0	-1,16363636
2	3	5	1	0,1	-0,83636364
3	3	5	1	0,2	-0,50909091
4	3	5	1	0,3	-0,18181818
5	3	5	1	0,4	0,145454545
6	3	5	1	0,5	0,472727273
7	3	5	1	0,6	0,8
8	3	5	1	0,7	1,127272727
9	3	5	1	0,8	1,454545455
10	3	5	1	0,9	1,781818182
11	3	5	1	1	2,109090909

Обнаруживаем точку «перехода» вектора $\langle 3,5,1 \rangle$ из класса $\sim A$ в класс A с пятой строки (значения в столбце Y становятся положительными). По этим данным нетрудно получить общий результат: $\mu_A \langle 3,5,1 \rangle \approx 0,65, \mu_{\sim A} \langle 3,5,1 \rangle \approx 0,35$ (среднее между двумя разделяющими значениями: 0,3 и 0,4). Теперь строим такую же экспериментальную таблицу для вектора $\langle 3, 5, 1,77 \rangle$ из ряда (3) и т.д. Итоговые оценки систематизированы авторами в таблице *TTTT* (табл. 4).

Таблица 4. Экспериментальная таблица *TTTT* для векторов $\langle 3,5, _ \rangle$

x	x_1	x_2	x_3	μ_A
1	3	5	1	0,65
2	3	5	1,77	0,65
3	3	5	2,45	0,55
4	3	5	3,24	0,55
5	3	5	4,02	0,45
6	3	5	4,8	0,45
7	3	5	5,58	0,45
8	3	5	6,36	0,35
9	3	5	7,13	0,35
10	3	5	8	0,25

По сути, таблица $T T T T$ задает определение нечеткой меры на заданном диапазоне. Однако эта таблица неполна, поскольку не представлены нечеткие значения, позволяющие «охватить» весь диапазон от 0 до 1. Здесь возможны варианты, о чем говорилось ранее. Если диапазон значений для x_3 представляется исследователю полным (других значений нет), то первая задача **31** решена. В противном случае к экспериментальной таблице $T T T T$ следует применить технику экстраполяции (например, [8]). Этим мы получаем решение первой задачи **31**.

Решение задачи 32

Для решения задачи **32** мы прибегнем к методу Р. Ягера [9, 10], который устанавливает связь между вероятностью нечеткого объекта и его нечеткой мерой принадлежности и который мы несколько модифицируем. Получив вероятности, можно далее вычислить ожидаемое значение нечеткого вектора, что и будет являться прогнозным значением.

В алгоритме Ягера нужно разбить нечеткие объекты сначала на интервальные множества W_α :

$$W_\alpha = \{x \mid \mu(x) > \alpha\}. \quad (4)$$

Имеем

$$W_{0,65} = \{ \},$$

$$W_{0,55} = \{1, 2\},$$

$$W_{0,45} = \{1, 2, 3, 4\},$$

$$W_{0,35} = \{1, 2, 3, 4, 5, 6, 7\},$$

$$W_{0,25} = \{1, 2, 3, 4, 5, 6, 7, 8, 9\},$$

$$W_0 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}.$$

Здесь в записи множеств уровня указаны в фигурных скобках номера векторов x из таблицы $T T T T$. С каждым множеством уровня, по Ягеру, связывается (полу)интервал, на котором это множество остается «неизменным». Так, например, множество $W_{0,55} = \{1, 2\}$ сохраняет свой вид на (полу)интервале $[0,55; 0,65)$. Далее, следуя Ягеру, определим вероятность выбора «наугад» произвольного множества уровня (4) как величину, пропорциональную длине (полу)интервала $\beta - \alpha$, на котором это множество остается неизменным и при условии, что сумма вероятностей выборов интервалов равна 1.

$$P(W_{\alpha_i}) = \frac{\beta_i - \alpha_i}{\sum_k (\beta_k - \alpha_k)}, \quad (5)$$

где β_i – верхняя граница полуинтервала. Получаем

$$P(W_{0,55}) = \frac{0,65 - 0,55}{0,65 - 0} = 0,154;$$

$$P(W_{0,45}) = \frac{0,55 - 0,45}{0,65 - 0} = 0,154;$$

$$P(W_{0,35}) = 0,154;$$

$$P(W_{0,25}) = 0,154;$$

$$P(W_0) = 0,385.$$

Определим теперь вероятность выбора произвольного объекта x как

$$P(x) = \sum_j P(W_j) \cdot P(x \mid W_j). \quad (6)$$

Используя формулу (6) и полагая, что векторы внутри каждого множества уровня выбираются с равной вероятностью, сведем данные расчетов в следующую табл. 5. Из этой таблицы без труда находим математическое ожидание координаты x_3 :

$$M[x_3] = \sum x_{3i} \cdot P(x_{3i}), \quad (7)$$

получая в ответе $M[x_{3i}] = 3,37$.

Таблица 5. Вероятности векторов <3,5,>

x	x1	x2	x3	P(x)
1	3	5	1	0,193
2	3	5	1,77	0,193
3	3	5	2,45	0,116
4	3	5	3,24	0,116
5	3	5	4,02	0,077
6	3	5	4,8	0,077
7	3	5	5,58	0,077
8	3	5	6,36	0,056
9	3	5	7,13	0,056
10	3	5	8	0,039

Итак, прогнозным значением нечеткого вектора является <3, 5, 3,37>. Этому прогнозному значению можно указать меру нечеткости (вероятность), если построить зависимость между x_3 и $\mu_A(x_3)$ ($P(x_3)$). Воспользуемся Excel. Представлена экспоненциальная аппроксимация функции меры принадлежности $\mu_A(x_3)$. Аппроксимирующая функция имеет следующий вид

$$y = 0,7439 \cdot e^{-0,087x} \quad (8)$$

Для найденного прогнозного значения $M[x_3] = 3,37$ найдем $\mu_A(3,37) = 0,55$. Хороший прогноз может быть получен и на основании «плохой» модели и наоборот. Таким образом, требуется оценка самой модели прогнозирования. В нашем случае качество модели изначально определяется тем, насколько модель четкого классификатора адекватно описывает (разбивает) множество нечетких векторов. Именно с этой позиции обосновывается качество введенного здесь нечеткого предсказателя. Этот вопрос нашел решение в [6]. Рассмотрим два последних столбца таблицы TT (в столбце Y знак «+» заменен на 1, а «-» на 0) (табл. 6).

Таблица 6. Измененная таблица TT

μ_A	0,6	0,4	0,8	0,2	1	0	0,9	0,1	0,8	0,2	0,6	0,4	0,55	0,45	0,55	0,45
Y	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0

Тогда все сводится к вопросу о статистической «близости» двух этих столбцов (например, в смысле критерия Фишера или χ^2 -квadrat). Следовательно, если нет оснований считать эти ряды статистически не адекватными, то четкий распознаватель правильно кластеризует обучающее множество и модель дает обоснованные выводы. Предположим, что нет статистической адекватности двух рассматриваемых рядов чисел. Например, возьмем первую строку приведенной выше таблицы с наибольшим отклонением между μ_A и Y . В таблице TT ей соответствует следующий вектор (табл. 7).

Таблица 7. Вектор расширенной таблицы наблюдений

x	x1	x2	x3	μ_A	Y
1	1	4	8	0,6	+

Используя подход [6, 7], считаем, что в 6 из 10 случаев ($\mu_A = 0,6$) данный вектор ассоциировался с классом A . Заменим данный вектор, например, 10 «похожими» на него случайными векторами, из которых 6 отнесем к классу A , а 4 – к классу $\sim A$ (табл. 8). Значения координат «группируются» вокруг «старых» разрядов как $x_i \pm \Delta$ со случайной «ошибкой» Δ (Δ имеет нулевое математическое ожидание и среднеквадратичное отклонение, равное доверительному интервалу для среднего значения соответствующего разряда). Этим можно добиться уменьшения расчетного значения критерия χ^2 -квadrat, используемого для проверки адекватности рядов μ_A и Y . Описанный прием позволяет «подогнать» четкий классификатор под нечеткие объекты с требуемой доверительной точностью.

Таблица 8. Замена нечеткого вектора четким

x	x_1	x_2	x_3	μ_A	Y
1	1	4	8	1	+
1,04	1	4	7,97	1	+
0,95	0,95	4,02	8	1	+
1,05	1,03	4	8,05	1	+
0,98	0,96	3,88	8,02	1	+
1	1	4,05	8	1	+
1	1,05	3,96	7,92	0	-
0,96	0,96	3,98	8,02	0	-
0,94	1,02	4,03	7,94	0	-
1,06	0,92	4	7,96	0	-

Заключение

Описанный подход к прогнозированию нечетких многомерных объектов использует технику кластеризации, позволяющую достаточно просто обрабатывать весьма значительные по объему обучающие выборки (сотни и тысячи объектов). Не предполагается знания законов распределения разрядов случайных многомерных объектов, взаимосвязи (наличия выраженной парной и групповой корреляции) между разрядами. Прогнозирование можно выполнить при достаточно общих допущениях.

A FORECASTING MODEL ON THE BASIS OF A FUZZY LEARNING SET

Yu.O. GERMAN , O.V. GERMAN.

Abstract

A problem of constructing a numeric forecasting evaluator on the basis of a fuzzy learning set is considered. The stated general problem is connected to the definition of the missing fuzzy vector co-ordinates and their evaluation. The general formulation is divided into two tasks: to build a method producing missing fuzzy forecasting values with expected value of a fuzzy measure and forecasting quality estimation. The given mathematical backgrounds are based on the model of a multidimensional crisp classifier and its usage for the fuzzy measure definition with the following evaluation on the basis of the fuzzy vectors probabilities by R. Yager.

Keywords: fuzzy learning set, membership measure function, classification, forecasting.

Список литературы

1. Вятчин Д.А. Нечеткие методы автоматической классификации. Минск, 2004.
2. Ishibuchi H., Nakashima T., Murata T. // IEEE Transactions on Systems, Man, and Cybernetics. 1999. Vol. 29. P. 601–618.
3. Тукаева Э.М., Мухаметзянов И.З. // УЭКС. 2013. № 8. С. 65–69.
4. Chen S.M. // Fuzzy sets Systems. 1996. Vol. 81, № 3. P. 311–319.
5. Демидов Л.А., Скворцова Г.С. // Вестн. РГРТУ. 2010. № 1 (31). С. 28–35.
6. Боброва Н.Л., Герман О.В. // Матер. Междунар. НК «Информационные технологии и системы». Минск, октябрь 2013. С. 242–244.
7. Герман О.В., Боброва Н.Л. // Докл. БГУИР. 2013. № 6 (76). С. 67–71.
8. Фильчаков П.Ф. Численные и графические методы прикладной математики. Киев, 1970.
9. Нечеткие множества и теория возможностей / Под ред. Р. Ягера. М., 1986.
10. Герман О.В. Введение в теорию экспертных систем и обработку знаний. Минск, 1995.