

## О распознавании экстремальных объектов с помощью интервальных решающих правил

Н.В. Киншт, Н.Н. Петрунько\*, Г.Ш. Цициашвили†

**Аннотация.** Задано конечное множество объектов  $\mathbf{A} = \{a_i\}$ ,  $i = 1 \dots n$ , характеризуемых определяющими параметрами  $\mathbf{X} = \{x_i\}$  и экспериментально полученными интервальными свойствами  $\mathbf{Y} = \{y_{ij}\}$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, m$ . На множестве  $\mathbf{A}$  заданы подмножества характерных представителей  $\{\mathbf{A}_1, \dots, \mathbf{A}_r\}$ . Обсуждаются вопросы корректности, эффективности и оптимизации классификации, основанной на имеющихся экспериментальных данных.

Задачи классификации объектов на основе экспериментальных наблюдений давно и многократно привлекали к себе внимание исследователей. Особо оправдывает себя интервальный подход к проблеме при обработке ограниченного числа наблюдений и отсутствии надежд описать процессы, управляющие состоянием объектов, реальными статистическими закономерностями. Один из вариантов практически эффективного и простого интервального решения задачи обработки естественнонаучных наблюдений продемонстрировал с соавторами в своих работах Г.Ш. Цициашвили [1, 2] и др. Представляется, что подобный подход мог бы быть полезным в задачах диагностики (оценки технического состояния) высоковольтного электроэнергетического оборудования, тем более, что в настоящее время идет мощный процесс реструктуризации электроэнергетики, сопровождающийся массовым обследованием энергетических мощностей. Имея это в виду, в настоящей работе делается попытка более глубоко формализовать и (или) развить отмеченный выше подход.

Задано конечное множество объектов  $\mathbf{A} = \{a_i\}$ ,  $i = 1, \dots, n$ , характеризуемых определяющими параметрами  $\mathbf{X} = \{x_i\}$ , которые являются основой для классификации объектов. (Для простоты в дальнейшем объекты  $a_i$  будем отождествлять с их номерами, т. е. индексами  $i$ , что не может вызвать недоразумений). Рассматривается также множество интервальных свойств объектов, задаваемых матрицей  $\mathbf{Y} = \{y_{ij}\}$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, m$ .

Считается, что интервальные значения элементов матрицы  $\mathbf{Y}$  получены экспериментальным путем и отражают либо тренд соответствующего измеряемого значения во время его наблюдения, либо несколько повторных наблюдений в серии, либо априори интервально заданные ошибки измерений.

На множестве  $\mathbf{A}$  в соответствии с некоторой классификацией заданы подмножества характерных представителей  $\mathbf{A} = \cup \mathbf{A}_k$ ,  $k = 1, \dots, r$ . Им соответствуют непересекающиеся интервальные определяющие параметры:

$$\mathbf{X}_k = [\min(x_i), \max(x_i)], \quad i \in \mathbf{A}_k;$$

\*Институт автоматизации и процессов управления ДВО РАН.

†Институт математики ДВО РАН.

$$\mathbf{X}_{k_1} \cap \mathbf{X}_{k_2} = \emptyset, \quad k_1 \neq k_2, \quad k_1, k_2 = 1, \dots, r.$$

Ясно, что каждому множеству  $\mathbf{A}_k$  соответствует  $m$ -мерный интервал (гиперпараллелепипед) в пространстве свойств объектов  $\mathbf{Y}_k = \cup \mathbf{Y}_i, i \in \mathbf{A}_k$ . Если ассоциировать множество номеров свойств объекта с множеством точек плоскости на оси абсцисс, а их значения – с ординатами, то удобно говорить также о трубке свойств (значений  $\mathbf{y}_i$ ) для каждого объекта  $i$  и, соответственно, о трубках свойств  $\mathbf{Y}_k$  для характерных представителей.

В терминах используемой интервальной модели можно ставить и решать следующие задачи:

- оценивать корректность классификации;
- оценивать эффективность классификации, основанной на заданном конкретном экспериментальном материале;
- эксплуатировать классификацию, т. е. принимать практическое решение о принадлежности нового объекта ( $a_i \notin \mathbf{A}$ ) при получении соответствующих экспериментальных данных  $y_i$ ;
- оптимизировать классификацию, т. е. пересматривать и уточнять множества характерных представителей;
- оценивать значимость каждого из свойств при решении задачи классификации;
- исследовать задачу прогнозирования поведения объектов.

Классификацию будем считать корректной, если выполняется условие

$$\forall k_1, k_2 \in \{1, \dots, r\}, \quad k_1 \neq k_2 \quad \exists i_1 \in \mathbf{A}_{k_1}, \quad i_2 \in \mathbf{A}_{k_2}, \quad \mathbf{y}_{i_1} \setminus (\mathbf{y}_{i_1} \cap \mathbf{Y}_{k_2}) \neq \emptyset. \quad (1)$$

Соотношение (1) постулирует принципиальную возможность классификации. При необходимости можно говорить о слабой либо сильной корректности, опираясь на свойства интервалов  $\mathbf{Y}_k$  в целом, а именно, если никакая пара интервалов не находится в отношении полного включения (слабая корректность):

$$\forall k_1, k_2 \in \{1, \dots, r\}, \quad k_1 \neq k_2, \quad \mathbf{Y}_{k_1} \setminus (\mathbf{Y}_{k_1} \cap \mathbf{Y}_{k_2}) \neq \emptyset, \quad (2)$$

либо не пересекаются вообще (сильная корректность):

$$\forall k_1, k_2 \in \{1, \dots, r\}, \quad k_1 \neq k_2, \quad (\mathbf{Y}_{k_1} \cap \mathbf{Y}_{k_2}) = \emptyset. \quad (3)$$

В дальнейшем, не оговаривая специально, рассматриваемые классификации будем считать корректными (например, в смысле (1)).

Оценим эффективность некоторой классификации. Для каждого  $i \in \mathbf{A}$  проверим соблюдение условия принадлежности  $\mathbf{y}_i$  трубке  $\mathbf{Y}_k$  и построим множества  $\mathbf{B}_k$ , соответствующие попаданию  $\mathbf{y}_i$  в трубку  $\mathbf{Y}_k$  и непопаданию параметра  $x_i$  в подмножество  $\mathbf{X}_k$ :

$$\forall i \in \mathbf{A} \quad (\mathbf{y}_i \subset \mathbf{Y}_k) \vee (x_i \notin \mathbf{X}_k) \Rightarrow i \in \mathbf{B}_k. \quad (4)$$

В соотношении (4) подразумевается полное попадание объекта  $\mathbf{y}_i$  в трубку  $\mathbf{Y}_k$ ; однако аналогично свойствам (2, 3) можно различать слабое либо сильное (как в (4)) соответствии объекта характерным подмножествам. Можно говорить также о несоответствии объекта  $\mathbf{y}_i$  трубке  $\mathbf{Y}_k$ , если хотя бы по одному параметру  $y_{ij}$  объект не попадает в трубку свойств  $\mathbf{Y}_k$ :

$$\forall i \in \mathbf{A} \quad \exists j (y_{ij} \notin \mathbf{Y}_k) \Rightarrow i \in \mathbf{B}_k. \tag{5}$$

Теперь эффективность классификации (вероятность распознавания объектов на основе имеющихся экспериментальных данных) отражается вектором неудачно классифицированных случаев  $\mathbf{b} = (b_1, \dots, b_r) = (|\mathbf{B}_1|, \dots, |\mathbf{B}_r|)$ , или же, например, суммарным интегральным показателем  $(1 - (b_1 + \dots + b_r)/n)$ . Естественно, что для оценки эффективности классификации можно было бы покоординатно сравнивать пересечения гиперпараллелепипедов  $\mathbf{Y}_k$  и оценивать их объемы; формально здесь нет проблем, но, имея в виду обработку естественнонаучных или технических экспериментальных данных с не вполне ясными внутренними связями и физически несоизмеримыми параметрами, идеи о гиперобъеме необходимо убедительно аргументировать в каждой конкретной задаче.

При получении дополнительных экспериментальных данных  $\mathbf{y}_i, i = n, \dots, p$ , соответствующих новым объектам, множество  $\mathbf{A}$  расширяется, практическое решение о принадлежности объекта соответствующему подмножеству характерных представителей принимается по принципу, аналогичному (4) или (5):

$$\forall i (i = n, \dots, p) \quad (\mathbf{y}_i \subset \mathbf{Y}_k) \Rightarrow i \in \mathbf{A}_k, \tag{6}$$

причем уверенность в правильности решения обеспечивается с вероятностью  $(1 - b_k/n)$ .

Обратимся к иллюстративному примеру. Пусть исходные данные заданы табл. 1 (для простоты введена нормировка:  $x_i \in [0, 1], y_{ij} \in [0, 1]$ ).

Таблица 1

A	X	Y <sub>1</sub>	Y <sub>2</sub>	Y <sub>3</sub>	Y <sub>4</sub>	Y <sub>5</sub>	Y <sub>6</sub>	Y <sub>7</sub>	Y <sub>8</sub>
1	0.0000	0.5133	0.5494	0.7391	0.6429	0.8030	0.7500	0.8750	0.7273
2	0.0000	0.6467	0.9699	0.9348	0.9286	0.7121	0.5000	0.9375	0.7273
3	0.0000	0.7067	0.9699	0.7826	0.5714	1.0000	0.7500	0.6250	0.9091
4	0.0760	0.5933	0.7651	0.9130	1.0000	0.6061	0.7500	0.5015	0.5000
5	0.0770	0.5933	0.8735	0.8478	0.5714	0.7727	0.7500	0.6625	0.6818
6	0.0800	0.4667	0.6265	0.8478	0.9286	0.8636	0.7500	0.7500	1.0000
7	0.1538	0.5733	0.7711	0.4783	0.2857	0.3030	0.5000	0.8125	0.3636
8	0.1698	0.5733	0.8434	0.3043	0.0000	0.3030	0.2500	0.5625	1.0000
9	0.2308	0.0000	0.2771	0.5217	0.7143	0.3030	0.5000	0.7500	0.1818
10	0.2564	0.5733	0.3855	1.0000	0.2857	0.2424	0.5000	0.7500	0.1818
11	0.2809	0.2800	0.1566	0.5217	0.2857	0.2727	0.2500	1.0000	0.9091
12	0.3077	0.0400	0.4217	0.7826	0.1429	0.3030	0.2500	0.4375	0.4545
13	0.3468	0.4267	0.2169	1.0000	0.7143	0.0909	0.2500	0.6875	0.7273
14	0.3846	0.1333	0.6265	0.7391	0.7143	0.0000	0.2500	0.3750	0.7727
15	0.4156	0.3067	0.2169	0.3913	0.7143	0.1515	0.5000	0.3125	0.1364
16	0.5611	0.3067	0.0000	0.5217	1.0000	0.7576	0.2500	0.0000	0.6364
17	0.8851	0.8400	0.8072	0.9130	0.5714	0.4545	0.5000	0.3125	0.3182
18	0.9231	0.7067	0.3976	0.6087	1.0000	0.4545	0.7500	0.8125	0.0455
19	1.0000	1.0000	0.9880	0.3043	0.1429	0.2121	1.0000	0.4375	0.0909
20	1.0000	0.2800	0.8795	0.7391	0.2857	0.4848	0.5000	0.7500	0.2727

Пусть также дано, что  $A = A_1 \cup A_2 \cup A_3 = \{1, \dots, 6\} \cup \{7, \dots, 15\} \cup \{16, \dots, 20\}$ , и, следовательно,  $\mathbf{X}_1 = [0, 0.08]$ ;  $\mathbf{X}_2 = [0.1538, 0.4156]$ ;  $\mathbf{X}_3 = [0.5611, 1.0]$ . Если предполагается, что определяющий параметр  $\mathbf{X}$  может принимать любые значения на интервале  $[0, 1]$ , учитывая, что  $[0, 1] \setminus (\mathbf{X}_1 \cup \mathbf{X}_2 \cup \mathbf{X}_3) \neq \emptyset$ , необходимо

Таблица 2

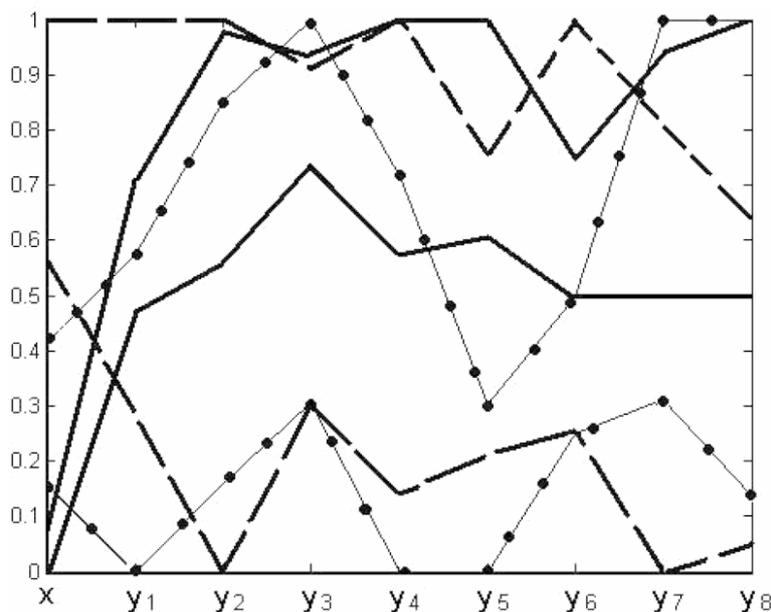
	$X_k$	$Y_{1k}$	$Y_{2k}$	$Y_{3k}$	$Y_{4k}$	$Y_{5k}$	$Y_{6k}$	$Y_{7k}$	$Y_{8k}$
$A_1$	[0, 0.1)	[0.4667 0.7067]	[0.5494 0.9699]	[0.7391 0.9130]	[0.5714 1.0000]	[0.6061 1.0000]	[0.5000 0.7500]	[0.5015 0.9375]	[0.5000 1.0000]
$A_2$	[0.1538, 0.4156]	[0.0000 0.5733]	[0.2771 0.8434]	[0.3043 1.0000]	[0.0000 0.7143]	[0.0000 0.3030]	[0.2500 0.5000]	[0.3125 1.0000]	[0.1364 1.0000]
$A_3$	[0.5611, 1.0]	[0.2800 1.0000]	[0.9880 0.0000]	[0.3043 0.9130]	[0.1429 1.0000]	[0.2121 0.7576]	[0.2500 1.0000]	[0.0000 0.8125]	[0.0909 0.6364]

соответствующим образом дополнить интервалы характерных представителей  $X_k$ , например:  $X_1 = [0, 0.1)$ ;  $X_2 = [0.1, 0.45)$ ;  $X_3 = [0.45, 1.0]$ . Впрочем, последнее представляет интерес лишь при физической интерпретации результатов.

Построив гиперпараллелепипеды  $Y_1, Y_2, Y_3$ , (табл. 2), изобразим их также, для наглядности, в виде трубок характерных представителей (рисунок).

Проверим корректность классификации. Построив гиперпараллелепипеды  $Y_1, Y_2, Y_3$ , выясним, что они находятся в соотношении (2) слабой корректности. Для оценки эффективности классификации построим множество неудачно классифицированных объектов:  $B_1 = \emptyset, B_2 = \emptyset, B_3 = \{4, 7\}$ . Таким образом, из 20 объектов неудачно классифицируются два, и эффективности классификации можно приписать значение 0.9.

Рассмотрим далее практическое использование построенной классификации. Пусть получены экспериментальные данные  $A' = \{21, \dots, 26\}$  (табл. 3).



Представление гиперпараллелепипедов  $Y_1$  (сплошные линии),  $Y_2$  (штриховые линии),  $Y_3$  (линии точек)

Таблица 3

A	Y <sub>1</sub>	Y <sub>2</sub>	Y <sub>3</sub>	Y <sub>4</sub>	Y <sub>5</sub>	Y <sub>6</sub>	Y <sub>7</sub>	Y <sub>8</sub>
21	0.2467	0.6783	0.3913	0.2857	0.2152	0.5000	0.5625	0.4091
22	0.7933	0.1387	0.3043	0.2857	0.2121	0.5000	0.6250	0.1364
23	0.5067	0.9398	0.3448	0.2857	0.5152	0.5000	0.6250	0.3636
24	0.2400	0.6386	0.3592	0.5714	0.3030	0.2500	0.4375	0.8636
25	0.5733	0.7843	0.8783	0.5714	0.8030	0.7500	0.6875	0.5909
26	0.3067	0.3566	0.5217	0.2857	0.3030	0.5000	0.6250	0.1364

Применяя принцип (5) к объектам {21, ..., 26}, получим, что {a<sub>21</sub>, a<sub>24</sub> ∈ A<sub>2</sub>, a<sub>22</sub>, a<sub>23</sub> ∈ A<sub>3</sub>, a<sub>25</sub> ∈ A<sub>1</sub>, а объект a<sub>26</sub> может быть отнесен как к подмножеству A<sub>2</sub>, так и подмножеству A<sub>3</sub>. Таким образом, объекты a<sub>21</sub>, ..., a<sub>25</sub> классифицируются с уверенностью 0.9, а объект a<sub>25</sub> — с уверенностью 0.45.

Об оптимизации классификации. Можно варьировать разбиение области определения определяющего параметра X на (заданное) число r интервалов определения характерных представителей. Каждому из этих разбиений будет в соответствии с (4) сопутствовать вектор b = (b<sub>1</sub>, ..., b<sub>r</sub>), оптимизация классификации сводится к минимизации его нормы.

Оценивая значимость каждого из свойств j, j = 1, ..., m, при решении задачи классификации, аналогично (4) либо (5) построим вектор

$$b_j = (b_{1j}, \dots, b_{rj}) = (|B_{1j}|, \dots, |B_{rj}|);$$

$$\forall j = 1 \dots m, \quad i = 1 \dots n, \quad (y_{ij} \in Y_{kj}) \vee (x_i \notin X_k) \Rightarrow i \in B_j. \quad (7)$$

Норма вектора b<sub>j</sub> в обратном соотношении ассоциируется со значимостью параметра j.

Наконец, остановимся вкратце на связи данной модели с задачей прогнозирования поведения объектов. Интерпретируя исходную таблицу как классификатор состояний объекта, а дополнительные экспериментальные данные — как изменение параметров некоторого объекта во времени, произведя экстраполяцию каждого из параметров, возможно на этой основе с помощью данной классификации осуществлять прогноз состояния объекта по его определяющему параметру. Так, к примеру, обратившись к табл. 4, представим себе, что данные строк 27 и 28 соответствуют наблюдению поведения некоторого объекта в паре временных точек с интервалом Δt. Нетрудно видеть, что состояние объекта характеризовалось интервалом характерного свойства X<sub>1</sub>. Произведя элементарную, единственно возможную (при отсутствии иных дополнительных данных) интерполяцию поведения объекта, получим строку 29 в таблице свойств объекта. Очевидно, что теперь прогнозируется переход его в состояние, соответствующее X<sub>3</sub>.

Таблица 4

A''	Y <sub>1</sub>	Y <sub>2</sub>	Y <sub>3</sub>	Y <sub>4</sub>	Y <sub>5</sub>	Y <sub>6</sub>	Y <sub>7</sub>	Y <sub>8</sub>
27	0.65	0.9	0.9	0.9	0.8	0.6	0.6	0.9
28	0.5	0.8	0.8	0.7	0.6	0.65	0.62	0.7
29	0.35	0.7	0.7	0.5	0.4	0.7	0.64	0.5

**Выводы.** Рассматриваемая формальная модель классификации достаточно прозрачна по постановке и проста в использовании. Представляется полезным более глубоко ее формализовать.

### Список литературы

- [1] Цициашвили Г.Ш. и др. Модификация метода интервальной математики применительно к прогнозу экстремальной ледовитости в Охотском море // Вестник ДВО РАН. – 2002. – № 4. – С. 111–114.
- [2] Болотин Е.И., Цициашвили Г.Ш., Голычева И.В. Некоторые аспекты и перспективы факторного прогнозирования эпидемического проявления очагов клещевого энцефалита на основе многомерного анализа временных рядов // Паразитология. – 2002. – Т. 36, № 2. – С.89–95.