ДИНАМИЧЕСКИЕ ПРИЗНАКИ В ЗАДАЧАХ КЛАССИФИКАЦИИ МОДЕЛЕЙ

Д. А. Вятченин (Минск, Беларусь)

При выборе той или иной модели для решения некоторой прикладной задачи неизбежна классификация моделей [5]. Если χ – некоторый объект или процесс, который может быть описан n различными моделями $x_1,...,x_n$, а каждая модель $x_i, i=1,\dots,n$, независимо от того, является она предметной или абстрактной, может описываться m признаками, так что $x_i=(x_i^1,\dots,x_i^m)$, то для классификации моделей оказывается возможным применение традиционных статистических методов классификации объектов [4]. При решении задач классификации для представления исходных данных используются, как правило, матрицы вида «объект–свойство» и матрицы вида «объект–объект» [4]. Если $X=\{x_1,...,x_n\}$ – множество n объектов, каждый из которых характеризуется m признаками и может рассматриваться как точка в m-мерном признаковом пространстве $I^m(X)$, то исходные данные могут быть представлены в виде матрицы «объект–свойство» вида $\hat{X}_{n\times m}=[\hat{x}_i^t], i=1,\dots,n, t=1,\dots,m$, где x_i^t представляет собой значение t-го признака у i-го объекта, либо в виде матрицы «объект–объект», значениями ρ_{ij}, $i,j=1,\dots,n$ которой могут быть либо значения взаимных расстояний d_{ij}, либо значения коэффициентов близости r_{ij} между объектами $x_1,\dots,x_n$ множества X, так что матрица $\rho_{n\times n}=[\rho_{ij}]$ принимает вид матрицы попарных расстояний $d_{n\times n}=[d_{ij}]$, либо, соответственно, вид матрицы попарных коэффициентов близости $\rho_{n\times n}=[\rho_{ij}]$. Переход от матрицы «объект–свойство» к матрице «объект–объект» осуществляется с помощью различных расстояний между объектами или мер близости объектов друг к другу [4]. Поскольку признаки объектов измеряются в различных шкалах, как качественных, так и количественных [8], то возникает необходимость перехода к единообразному описанию для всех признаков, для чего применяются различные виды нормировки. В частности, для предварительной обработки данных, заданных в виде матрицы «объект–свойство», используется нормализация, определяемая формулой [8]

$$x_i^t=\frac{\hat{x}_i^t}{\max\limits_i \hat{x}_i^t}, \tag{1}$$

где $\hat{x}_i^t$ – значение t-го признака, измеренного в некоторой шкале, у i-го объекта.

При использовании нормализации все признаки $x^t, t=1,\dots,m$, характеризующие объекты исследуемой совокупности $X=\{x_1,...,x_n\}$, будут принимать значения в интервале $[0,1]$, так что объекты $x_i, i=1,\dots,n$ исследуемой совокупности могут быть интерпретированы как нечеткие множества на универсуме $x^t, t=1,\dots,m$ признаков, причем каждое значение $x_i^t, i=1,\dots,n, t=1,\dots,m$ может быть представлено в виде функции принадлежности $\mu_{x_i}(x^t)$, которая показывает степень выраженности t-го признака у i-го объекта. Поскольку для нечетких множеств определены различные расстояния [3], к примеру, обобщенное относительное расстояние Хемминга

$$d_H(x_i,x_j)=\frac{1}{m}\sum_{t=1}^{m}|\mu_{x_i}(x^t)-\mu_{x_j}(x^t)|,\quad i,j=\overline{1,n}, \tag{2}$$

где $n = card(X), \forall x_i, x_j \subseteq X, i = 1,\dots,n$, и имеет место соотношение $0 \le d_H(x_i, x_j) \le 1$, то, применяя к матрице «объект–свойство» $X_{n\times m} = [x_i^t], i = 1,\dots,n, t = 1,\dots,m$ нормированных данных расстояние (2), можно построить матрицу попарных расстояний $d_{n\times n}$ между объектами исследуемой совокупности. Матрица $d_{n\times n} = [\mu_I(x_i, x_j)], i, j = 1,\dots,n$ попарных расстояний может интерпретироваться как матрица нечеткого отношения несходства, или, иными словами, матрица нечеткой интолерантности I, дополнение T которой, вычисляемое в соответствии с формулой [7]

$$\mu_D(x_i, x_j) = 1 - \mu_I(x_i, x_j), \forall x_i, x_j \in X\ , \tag{3}$$

является нечеткой толерантностью, так что матрица T представляет собой матрицу коэффициентов попарного сходства объектов.

Вместе с тем, часто оказывается возможной ситуация, когда некоторому признаку x^t объекта x_i соответствует не одно, а несколько значений, либо признак x^t принимает значения в некотором интервале. Характерным примером объектов с варьирующимися в интервале значениями признаков могут послужить самолеты с изменяемой геометрией крыла. Подобные ситуации, возникающие в квалиметрии моделей, рассматриваются в работе [5]. Для классификации объектов с варьирующимися значениями признаков в работе [2] было предложено обобщение нормализации

$$x_i^{t(u)} = \hat{x}_i^{t(u)} \Big/ \max_{i,u} \hat{x}_i^{t(u)}\ , \tag{4}$$

где символом $\hat{x}^{t(u)}, u \in \{1,\dots,s\}$ обозначается наличие у признака $\hat{x}^t, t \in \{1,\dots,m\}$ для некоторого объекта $\hat{x}_i, i \in \{1,\dots,n\}$ s градаций, так что если признак $\hat{x}^t, t \in \{1,\dots,m\}$ принимает для объекта $\hat{x}_j, j \in \{1,\dots,n\}, i \ne j$ только одно значение, то $\hat{x}_j^{t(u)} = const, \forall u \in \{1,\dots,s\}$. Кроме того, в работе [2] было предложено следующее обобщение расстояния (2):

$$d_{gH}(x_i^{(s)}, x_j^{(s)}) = \frac{1}{m}\sum_{t=1}^{m}\left(\frac{1}{s^2}\left(\sum_{u=1}^{s}\sum_{v=1}^{s} |\,\mu_{x_i}(x^{t(u)}) - \mu_{x_j}(x^{t(v)})\,|\right)\right). \tag{5}$$

Для демонстрации информативности динамических признаков при определении сходства между объектами с помощью расстояния (5), в качестве иллюстративного примера реальных объектов были отобраны десять самолетов различного назначения американской разработки [6], а в качестве описывающих их признаков – пять достаточно информативных характеристик, для двух из которых значения могут приниматься в интервале. Данные об указанной исследуемой совокупности представлены табл. 1.

Таблица 1

Некоторые характеристики самолетов ВВС стран НАТО

Номер объекта	Тип самолета	Размах крыла, м.		Длина, м.	Высота, м.	Взлетная масса, кг.		Практ. потолок, м.
		макс.	мин.			номин.	макс.	
1	F-104A	6,68		16,69	4,11	11650	12400	17400
2	F-104G	6,68		16,69	4,11	9428	13054	17680
3	F-5A	7,70		14,68	4,06	6080	9379	15390
4	F-5E	8,13		14,68	4,06	7030	10922	16460
5	F-16A	9,45		14,52	5,01	10205	15500	15240
6	F-4B	11,70		17,76	4,96	20865	24765	19000
7	F-4E	11,77		19,19	5,02	18818	28030	19685
8	F-14A	19,45	11,64	18,89	4,88	24300	33724	21000
9	F-111A	19,20	9,74	22,40	5,22	32000	41500	15500
10	FB-111A	21,34	10,34	22,40	5,19	45360	54000	18300

При обозначении размаха крыла символом $\hat{x}^1$, длины – $\hat{x}^2$, высоты – $\hat{x}^3$, взлетной массы – $\hat{x}^4$, практического потолка – $\hat{x}^5$, а самих самолетов – символами $\hat{x}_i, i=1,\ldots,10$, где i соответствует номеру самолета в табл. 1, и применении к данным, представленным в табл. 1, формулы (4), получается таблица нормированных данных, каждая строка которой может интерпретироваться как нечеткое множество, определенное на интервале [2]. В качестве абстрактной модели объекта $x_i, i=1,\ldots,10$ исследуемой совокупности может рассматриваться соответствующая строка $i=(x_i^1,\ldots,x_i^m)$ таблицы нормированных данных. Поскольку каждому объекту исследуемой совокупности может быть поставлена в соответствие модель, представляющая собой значения любого непустого подмножества признаков множества $\{x^1,x^2,x^3,x^4,x^5\}$, то различные наборы признаков будут определять различные структуры сходства на исследуемой совокупности. Таким образом, представляется целесообразным рассмотреть структуру сходства между объектами исследуемой совокупности для различных наборов признаков. Матрица сходства, полученная путем применения операции дополнения (3) к матрице попарных коэффициентов различия, полученной, в свою очередь, путем применения расстояния (5) к таблице нормированных данных, для множества признаков $\{x^2,x^3\}$, представлена табл. 2, и является матрицей обычной нечеткой толерантности T_2.

Таблица 2

Матрица обычной нечеткой толерантности T_2

i	1	2	3	4	5	6	7	8	9	10
1	1,0000									
2	1,0000	1,0000								
3	0,9503	0,9503	1,0000							
4	0,9503	0,9503	1,0000	1,0000						
5	0,8654	0,8654	0,9054	0,9054	1,0000					
6	0,8947	0,8947	0,8450	0,8450	0,9229	1,0000				
7	0,8570	0,8570	0,8074	0,8074	0,8948	0,9623	1,0000			
8	0,8771	0,8771	0,8275	0,8275	0,8900	0,9671	0,9799	1,0000		
9	0,7662	0,7662	0,7166	0,7166	0,8040	0,8715	0,9092	0,8891	1,0000	
10	0,7691	0,7691	0,7194	0,7194	0,8069	0,8744	0,9121	0,8920	0,9971	1,0000

В свою очередь, сравнение объектов исследуемой совокупности по трем признакам $\{x^2,x^3,x^5\}$, порождает структуру сходства, описываемую сильной нечеткой толерантностью T_3, матрица которой представлена табл. 3.

Таблица 3

Матрица сильной нечеткой толерантности T_3

i	1	2	3	4	5	6	7	8	9	10
1	1,0000									
2	0,9956	1,0000								
3	0,9350	0,9305	1,0000							
4	0,9520	0,9475	0,9830	1,0000						
5	0,8760	0,8715	0,9346	0,9176	1,0000					
6	0,9044	0,9088	0,8394	0,8564	0,8889	1,0000				
7	0,8684	0,8729	0,8034	0,8204	0,8593	0,9640	1,0000			
8	0,8609	0,8654	0,7959	0,8129	0,8352	0,9463	0,9657	1,0000		
9	0,8140	0,8095	0,8093	0,7958	0,8652	0,8588	0,8730	0,8388	1,0000	
10	0,8318	0,8362	0,7668	0,7838	0,8227	0,9052	0,9194	0,8851	0,9536	1,0000

При сравнении объектов по признакам $\{x^1, x^2, x^3\}$ получается представленная табл. 4 матрица нормальной слабой нечеткой толерантности T_1.

Таблица 4

Матрица нормальной слабой нечеткой толерантности T_{1n}

i	1	2	3	4	5	6	7	8	9	10
1	1,0000									
2	1,0000	1,0000								
3	0,9510	0,9510	1,0000							
4	0,9442	0,9442	0,9933	1,0000						
5	0,8670	0,8670	0,9096	0,9163	1,0000					
6	0,8514	0,8514	0,8342	0,8409	0,9134	1,0000				
7	0,8252	0,8252	0,8080	0,8147	0,8936	0,9738	1,0000			
8	0,7796	0,7796	0,7624	0,7692	0,8315	0,9171	0,9256	0,9390		
9	0,7225	0,7225	0,7053	0,7120	0,7909	0,8405	0,8656	0,8502	0,9261	
10	0,7030	0,7030	0,6858	0,6925	0,7714	0,8304	0,8555	0,8421	0,9075	0,9141

Сравнение объектов по признакам $\{x^1, x^2, x^3, x^5\}$ дает нормальную строгую слабую нечеткую толерантность T_0, матрица которой представлена табл. 5.

Таблица 5

Матрица нормальной строгой слабой нечеткой толерантности T_{0n}

i	1	2	3	4	5	6	7	8	9	10
1	1,0000									
2	0,9967	1,0000								
3	0,9393	0,9360	1,0000							
4	0,9470	0,9437	0,9822	1,0000						
5	0,8745	0,8712	0,9304	0,9227	1,0000					
6	0,8695	0,8728	0,8327	0,8505	0,8903	1,0000				
7	0,8417	0,8450	0,8049	0,8227	0,8673	0,9722	1,0000			
8	0,7919	0,7952	0,7551	0,7728	0,8050	0,9140	0,9285	0,9543		
9	0,7692	0,7659	0,7777	0,7726	0,8401	0,8387	0,8494	0,8222	0,9446	
10	0,7665	0,7699	0,7297	0,7475	0,7921	0,8644	0,8751	0,8494	0,8973	0,9356

Сходство объектов по всем признакам $\{x^1, x^2, x^3, x^4, x^5\}$ определяется представленной табл. 6 матрицей субнормальной строгой слабой нечеткой толерантности T_0.

Таблица 6

Матрица субнормальной строгой слабой нечеткой толерантности T_{0s}

i	1	2	3	4	5	6	7	8	9	10
1	0,9986									
2	0,9906	0,9933								
3	0,9355	0,9358	0,9939							
4	0,9463	0,9438	0,9768	0,9928						
5	0,8898	0,8857	0,9254	0,9225	0,9902					
6	0,8556	0,8554	0,8103	0,8291	0,8754	0,9928				
7	0,8311	0,8309	0,7858	0,8046	0,8547	0,9607	0,9829			
8	0,7706	0,7703	0,7252	0,7441	0,7842	0,9074	0,9152	0,9460		
9	0,7238	0,7182	0,7146	0,7152	0,7836	0,8193	0,8301	0,8259	0,9381	
10	0,6738	0,6735	0,6284	0,6472	0,6973	0,7920	0,8028	0,8030	0,8699	0,9325

Анализ нечетких толерантностей, описывающих структуру сходства на исследуемой совокупности десяти объектов в признаковом пространстве $I^m(X)$ при различных значениях m, позволяет сделать качественный вывод о корректности обобщения (4) нормализации (1) и обобщения (5) расстояния (2) между нечеткими множествами. Методологическое преимущество нормировки (4) и расстояния (5) заключается в адекватном отображении матрицей нечеткой толерантности $d_{n\times n} = [\mu_T(x_i, x_j)]$, полученной применением к матрице $d_{n\times n} = [\mu_I(x_i, x_j)]$ попарных расстояний между объектами исследуемой совокупности операции дополнения (3), структуры сходства между объектами исследуемой совокупности, не искажающим геометрии признакового пространства $I^m(X)$.

Таким образом, признаки, принимающие значения в интервале, оказываются достаточно информативными, а различные наборы признаков порождают различную структуру сходства при сравнении моделей, что приводит к различным результатам при их классификации. Кроме того, анализ результатов вычислительных экспериментов показывает, что для нечетких толерантностей, полученных при различных наборах признаков, описывающих объекты исследуемой совокупности, соотношение $T_0 \subseteq T_1 \subseteq T_3 \subseteq T_2$, приведенное в работе [1], в общем, не имеет места, так же, как и не выполняются соотношения $T_{0s} \subseteq T_{0n}$ и $T_{1s} \subseteq T_{1n}$.

Литература

1. **Вятченин Д. А.** Содержательная интерпретация нечетких отношений сходства // Полигнозис. – 2001. – № 1. – С. 20–25.
2. **Вятченин Д. А.** Обобщение понятия расстояния между нечеткими множествами для классификации объектов с динамическими признаками // Вестник Военной академии Республики Беларусь. – 2005. – № 3. – С. 32–37.
3. **Кофман А.** Введение в теорию нечетких множеств / Пер. с фр. В. Б. Кузьмина; под ред. С. И. Травкина. – М.: Радио и связь, 1982. – 432 с.
4. Прикладная статистика: Классификация и снижение размерности: Справ. изд./ С. А. Айвазян, В. М. Бухштабер, И. С. Енюков, Л. Д. Мешалкин; Под ред. С. А. Айвазяна. – М.: Финансы и статистика, 1989. – 607 с.
5. **Соколов Б. В., Юсупов Р. М.** Концептуальные основы оценивания и анализа качества моделей и полимодельных комплексов//Известия РАН. Теория и системы управления. – 2004. – № 6. – С. 5–16.
6. **Цихош Э.** Сверхзвуковые самолеты: Справочное руководство/Пер. с польск.; Под ред. В.Г. Микеладзе и Е.В. Зябрева. – М.: Мир, 1983. – 432 с.
7. **Viattchenin D. A.** Theoretical Notes on Fuzzy Intolerances. // Pattern Recognition and Information Processing: Proceedings of Sixth International Conference (15-17 May 2001, Minsk, Republic of Belarus). Vol. 2/Ed. by S Ablameyko et al. -Minsk, Institute of Engineering Cybernetics of the National Academy of Sciences of Belarus, 2001. – P. 142–146.
8. **Walesiak M.** Ugólniona miara odległości w statystycznej analizie wielowymiarowej. – Wrocław: Wydawnictwo Akademii Ekonomicznej im. Oskara Langego, 2002. – 107 s.