

Раздел III. Искусственный интеллект и нечеткие системы

УДК 681.3: 519.8: 517.11

В.М. Глушань, В.П. Карелин, О.Л. Кузьменко

ФОРМИРОВАНИЕ КЛАСТЕРОВ НА МНОЖЕСТВЕ НЕЧЁТКО ОПИСАННЫХ СИТУАЦИЙ ПРИ ПОСТРОЕНИИ МОДЕЛИ ПРИНЯТИЯ РЕШЕНИЙ

Для управления сложными системами или процессами важно наличие адаптивных быстро реагирующих управленческих механизмов, позволяющих принимать решение в условиях неопределенности, неполноты и нечеткости исходной информации. Эффективное управление требует не только большого опыта, знаний, квалификации, но и использования современных информационных технологий и соответствующих интеллектуальных систем поддержки принятия решений (ИСППР). Такие системы, обрабатывая и анализируя имеющуюся информацию о состояниях объекта управления (ОУ) и внешней среде, предлагают лицу, принимающему решения (ЛПР), несколько альтернативных вариантов решений (действий). Право выбора конкретного решения остается за ЛПР.

Чтобы управлять сложным объектом, необходимо иметь либо достаточно точную, но весьма сложную математическую модель ОУ, либо модель управления объектом, что более целесообразно. Иначе говоря, имеет смысл моделировать не сам ОУ, а действия человека-оператора или ЛПР в процессе управления объектом. Подобные модели, использующие не только традиционные, четкие алгоритмы и количественные данные, но и нечеткие алгоритмы и качественные оценки, реализуют в виде нечетких систем управления, включающих ИСППР. Основными компонентами этих систем являются база знаний, база моделей, база данных, блок логического вывода – интеллектуальный решатель, блок лингвистического представления ситуаций, блок числового представления ситуаций [1-3].

Ситуацией называется набор значений признаков, описывающих состояние ОУ в некоторый момент времени. Под нечетким описанием ситуации понимается такое, где отображены не только количественные, но и ряд качественных характеристик системы. Вид нечеткого описания объектов существенно зависит от области приложения. В ряде случаев нечеткое описание представляет собой совокупность нечетких множеств первого уровня, например: $S = \{ \text{качество среднее, производительность низкая, себестоимость высокая и т.д.} \}$. В еще более неопределенных обстоятельствах принятия решений (ПР) описание ситуации представляется совокупностью нечетких множеств второго уровня [1, 4]. Примером такого описания ситуации, характеризующей некоторое состояние, возникшее при управлении технологическим процессом, является следующий: $S = \{ \langle 0.4/\text{«низкое»}, \langle 0.8/\text{«среднее»}, \langle 0.2/\text{«высокое»} \text{ «качество»}, \langle 0.1/\text{«очень низкая»}, \langle 0.4/\text{«низкая»}, \langle 0.8/\text{«средняя»}, \langle 0.6/\text{«высокая»}, \text{ «производительность»}, \langle 0.1/\text{«низкая»}, \langle 0.6/\text{«средняя»}, \langle 0.9/\text{«высокая»}, \text{ «себестоимость»} \rangle \}$. Здесь ка-

ждому признаку (качество, производительность, себестоимость) соответствует лингвистическая переменная, значения которой (низкая, средняя, высокая и т.п.) в описании ситуации также заданы нечетко.

При разработке моделей, имитирующих процессы ПР человеком-оператором, исходят из того, что ЛПР представляет собой нечёткую систему распознавания образов, соответствующих возможным управлениям в системе, а процесс принятия решений ЛПР есть процесс распознавания ситуаций в многомерном пространстве признаков. Это предположение при условии существования эффективных процедур формирования эталонных классов ситуаций, даёт возможность строить эффективные схемы ПР. Они основаны либо на вычислении для данной конкретной ситуации её степени принадлежности каждому из эталонных классов, либо на вычислении сходства описания текущей ситуации с каждой из эталонных ситуаций, с последующим выбором решения, соответствующего тому классу или той эталонной ситуации, которой данная ситуация принадлежит с наибольшей степенью. Такие модели принятия решений (МПР), следуя [5], будем называть нечеткими классификационными моделями.

В работе рассмотрена процедура кластеризации нечетких ситуаций.

Задача кластеризации состоит в разбиении множества объектов на несколько подмножеств (классов, кластеров), в которых объекты более схожи между собой, чем с объектами из других кластеров. Чтобы сформировать на множестве объектов кластер, необходимо, в первую очередь, ввести меру сходства (подобия), которая может быть положена в основу правила отнесения объектов к кластеру, характеризующему некоторым центром (эталонном, представителем). Для кластеризации наиболее часто используют метод K -средних, который основан на минимизации суммы квадратов расстояний между каждым из исходных объектов и центром его кластера [6].

Одной из важных задач, которую необходимо решать при построении классификационных МПР, является нахождение для каждого класса представителя (эталона), с каждым из которых будут сравниваться исследуемые объекты. Способы нахождения для каждого класса своего представителя – эталона класса рассмотрены авторами в [7]. При проведении процедуры кластеризации методом K -средних представители (эталонные) классов находятся в процессе разбиения объектов (данных) на кластеры.

Процедура кластеризации методом K -средних состоит из следующих шагов:

- ◆ выбираются или назначаются k объектов, которые будут первичными центрами кластеров;
- ◆ остальные объекты приписываются к ближайшим заданным кластерным центрам;
- ◆ текущие координаты первичных кластерных центров заменяются на кластерные средние;
- ◆ предыдущие два шага повторяются до тех пор, пока изменения координат кластерных центров не станут минимальными.

Нами предлагается адаптация процедуры кластеризации к условиям нечетко описанных ситуаций, позволяющая снизить трудоемкость этой процедуры.

Для разбиения множества из N объектов (ситуаций) на классы (кластеризация по сходству описаний) необходимо определить степень сходства (или расстояние) для каждой пары ситуаций из заданного множества. Поскольку каждая ситуация описывается совокупностью нечетких множеств, то при определении близости (степени сходства или несходства) ситуаций можно использовать те же формулы, по которым сравниваются нечеткие множества.

Пусть на базовом множестве $X = \{x_1, x_2, \dots, x_n\}$ заданы нечеткие множества: $A = \{\mu_A(x)/x\}$ и $B = \{\mu_B(x)/x\}$, где $x \in X$, $\mu_A(x), \mu_B(x) \in [0, 1]$. Для определения степени сходства (несходства) нечетких множеств A и B применяют различные меры близости и расстояний, в частности, меры сходства Дейка, Танимото, расстояние Хэмминга и др. [5].

Предлагается на всех этапах построения и применения классификационных МПР оперировать не значениями дискретных функций принадлежности сравнимых нечетких множеств, а их репрезентативными числами [6]. В качестве репрезентативного числа r_A , характеризующего данное нечеткое множество A , будем использовать значение центра масс (тяжести) его функции принадлежности, которое определяется по формуле:

$$r_A = \frac{\sum_{i=1}^n \mu_A(x_i) \cdot x_i}{\sum_{i=1}^n \mu_A(x_i)}. \quad (1)$$

В нечетких множествах второго уровня элементами базового множества являются словесные (лингвистические) строго упорядоченные (по индексам i) значения t_i , поэтому в формуле для получения репрезентативного числа r_A такого нечеткого множества в качестве значения базовой переменной t_i правомерно использовать ее порядковый номер – индекс i [6]. Тем самым, формула (1) примет вид:

$$r_A = \frac{\sum_{i=1}^n \mu_A(t_i) \cdot i}{\sum_{i=1}^n \mu_A(i)}. \quad (2)$$

Поскольку любая ситуация задана нечеткими значениями каждой из k лингвистических переменных (нечеткими множествами второго уровня), то для построения усредненной ситуации необходимо для каждого нечеткого множества рассчитать репрезентативное число r (выполнить переход от нечеткого представления к четкому). В результате каждая из N ситуаций будет представлена репрезентативным вектором (РВ) $R = (r_1, r_2, \dots, r_k)$, состоящим из k репрезентативных чисел (k – количество признаков, характеризующих объект). Усредненная ситуация также будет представлена аналогичным репрезентативным вектором $R_{ср}$, каждый элемент r_i^* ($i = 1, 2, \dots, k$) которого получен как среднее арифметическое соответствующих i -х элементов всех N векторов R .

Рассмотрим предложенный способ нахождения представителя класса нечетких ситуаций на конкретном примере.

Пусть $S = \{c_1, c_2, c_3, c_4\}$ – множество признаков, значениями которых описываются ситуации, возникающие на объекте управления. Каждый признак x_i соответствует лингвистической переменной (ЛП), множество значений t_i (термов) которой обозначим через T . Значение каждой ЛП в описании ситуации задается нечетким множеством второго уровня $G = \{\mu_G(t_i)/t_i\}$, где базовым множеством является терм-множество T . В общем случае терм-множества для различных ЛП также различаются.

В нашем примере терм-множества значений каждого из признаков (ЛП), в силу их однотипности, примем одинаковыми: $T = \{t_0 - \text{несущественный уровень}; t_1 - \text{очень низкий уровень}; t_2 - \text{низкий уровень}; t_3 - \text{средний уровень}; t_4 - \text{высокий уровень}; t_5 - \text{очень высокий уровень}; t_6 - \text{значительный уровень}\}$. Приведенные лингвистические значения t_i ЛП, в свою очередь, задаются нечеткими множества-

ми первого уровня, функции принадлежности которых будем представлять нечеткими треугольными функциями или, иначе, нечеткими треугольными числами, определенными на базовом множестве $U = \{0; 0,17; 0,33; 0,5; 0,67; 0,83; 1\}$. Нечеткие треугольные или трапециевидные числа являются частным случаем функций принадлежности типа (L-R), которые широко используются для описания экспертных суждений [3, 5, 7].

Зададим некоторое значение ЛП (признака) c_1 следующим нечетким множеством: {<0/несущественный уровень>, <0/очень низкий ур.>, <0/низкий ур.>, <0/средний ур.>, <0,7/высокий ур.>, <0,7/очень высокий ур.>, <0/значительный ур.>}. Для компактности описания заменим в нечетких множествах все названия термов (низкий ур., средний ур. и т.д.) их обозначениями t_i ($i = 0, 1, \dots, 6$) в термно-множестве T . Получим для c_1 запись: {<0/ t_0 >, <0/ t_1 >, <0/ t_2 >, <0/ t_3 >, <0,7/ t_4 >, <0,7/ t_5 >, <0/ t_6 >}

Пусть задано множество $S = \{S_1, S_2, S_3, S_4\}$, состоящее из следующих 4 нечетко описанных ситуаций:

Таблица 1

Набор нечетко описанных ситуаций

T	0	1	2	3	4	5	6
	Несуществств. t_0	Очень низкий t_1	Низкий t_2	Средний t_3	Высокий t_4	Очень высокий t_5	Значительный t_6
U	0	0,17	0,33	0,5	0,67	0,83	1
S_1							
c_1	0	0	0	0	0,3	1,0	0,3
c_2	0	0	0	0,7	0,7	0	0
c_3	0	0	0,3	1,0	0,3	0	0
c_4	0	0	0	1,0	0	0	0
S_2							
c_1	0	0	0,3	1,0	0,3	0	0
c_2	0	0	0	0	0,3	1,0	0,3
c_3	0	0	0	0	0	0,7	0,7
c_4	0	0	0	0	0,2	0,9	1,0
S_3							
c_1	0	0	0	0	0,7	0,7	0
c_2	0	0	0	0	0,7	0,7	0
c_3	0	0	0,3	1,0	0,3	0	0
c_4	0	0	0	0	1,0	0	0
S_4							
c_1	0	0	0	0,2	0,5	0,7	1,0
c_2	0	0	0	0,2	0,5	0,7	1,0
c_3	0	0	0	0	0,2	0,9	1,0
c_4	0	0	0	0,2	0,5	0,7	1,0

Проиллюстрируем процедуру кластеризации методом K -средних.

Построим матрицу сходства объектов по критерию:

$$C(A,B) = 1 - \frac{\sum_{i=1}^n |\mu_A(x_i) - \mu_B(x_i)|}{\sum_{i=1}^n \mu_A(x_i) + \sum_{i=1}^n \mu_B(x_i)},$$

где n – мощность базового множества X .

Преобразуем транзитивную матрицу сходства в матрицу расстояний по формуле: $D_{ij} = 1 - C_{ij}$. Получим табл. 2.

Таблица 2

Матрица расстояний

D_{ij}	S_1	S_2	S_3	S_4
S_1	0	0,60	0,46	0,60
S_2	0,60	0	0,60	0,35
S_3	0,46	0,60	0	0,60
S_4	0,60	0,35	0,60	0

Производим разбиение объектов на классы.

1. Задаем количество классов $K=2$. В качестве центров кластеров зададим объекты: $z_1(1) = S_1$; $z_2(1) = S_2$.

2. По матрице расстояний (табл. 2) проверяем расстояние объектов до центров кластеров: $|S_i - z_i(1)| < |S_i - z_j(1)|$, где $i = 3, 4$; $j = 1, 2$; если это неравенство выполняется, то S_i принадлежит первому классу, если не выполняется, то S_i принадлежит второму классу.

Таким образом, разбиение имеет следующий вид:

$$P_1(1) = \{S_1, S_3\}, P_2(1) = \{S_2, S_4\}.$$

3. Далее определяем в каждом классе новые центры.

На этом этапе предлагается каждый объект исходного множества объектов представить репрезентативным вектором $R = (r_1, r_2, \dots, r_k)$, состоящим из k репрезентативных чисел, рассчитанных по формуле (2).

Применение репрезентативного представления объектов позволяет работать со средними значениями РВ без восстановления нечеткого описания объектов на каждом этапе алгоритма.

Для отыскания РВ каждого объекта вычислим значения репрезентативных чисел r_i ($i=1, \dots, 4$) для каждого признака. Полученные значения представим в табл. 3.

Таблица 3

Представление исходных ситуаций посредством РВ

	S_1	S_2	S_3	S_4
c_1	5,0	3,0	4,5	5,1
c_2	3,5	5,0	4,5	5,1
c_3	3,0	5,5	3,0	5,4
c_4	3,0	5,4	4,0	5,1

Затем найдем среднее арифметическое значений r_i объектов каждого класса по каждому из признаков. Полученные значения представлены в табл. 4, 5.

Таблица 4

Значения элементов центра первого кластера

r_i	$R_1(S_1)$	$R_3(S_3)$	$z_1(2)=R_{cp}^1=\{r_i\}$
c_1	5,0	4,5	4,8
c_2	3,5	4,5	4,0
c_3	3,0	3,0	3,0
c_4	3,0	4,0	3,5

Таблица 5

Значения элементов центра второго кластера

r_i	$R_2(S_2)$	$R_4(S_4)$	$z_2(2)=R_{cp}^2=\{r_i\}$
c_1	3,0	5,1	4,0
c_2	5,0	5,1	5,0
c_3	5,5	5,4	5,5
c_4	5,4	5,1	5,2

Набор средних арифметических значений r_i по каждому признаку в каждом классе дает РВ, соответствующие новым усредненным объектам (центрам) полученных классов.

Так как $z_1(1) \neq z_1(2)$, $z_2(1) \neq z_2(2)$, необходимо пересчитать расстояния от объектов множества до новых центров кластеров.

4. Для расчета расстояния между центрами кластеров и остальными альтернативами воспользуемся формулой Хэмминга:

$$D(S_i, z_j) = \sum_{l=1}^k |r_l^i - r_l^j|. \quad (3)$$

Полученные результаты представлены в табл.б.

Таблица 6

Расчет расстояний объектов до центров кластеров

	$z_1(2)$	S_1	S_2	S_3	S_4
c_1	4,8	0,3	1,8	0,3	0,3
c_2	4,0	0,5	1,0	0,5	1,1
c_3	3,0	0,0	2,5	0,0	2,4
c_4	3,5	0,5	1,9	0,5	1,6
Ср. арифметическое		0,3	1,8	0,3	1,4

	$z_2(2)$	S_1	S_2	S_3	S_4
c_1	4,0	1,0	1,0	0,5	1,0
c_2	5,0	1,5	0,0	0,5	0,0
c_3	5,5	2,5	0,0	2,5	0,0
c_4	5,2	2,2	0,2	1,2	0,2
Ср. арифметическое		1,8	0,3	1,2	0,3
Номер кластера		1	2	1	2

Таким образом, разбиение имеет следующий вид:

$$P_1(2) = \{S_1, S_3\}, P_2(2) = \{S_2, S_4\}.$$

5. На следующем шаге получаем следующие центры кластеров:

Таблица 7

Центры кластеров на третьей итерации

	$z_1(3)$	$z_2(3)$
c_1	4,8	4,0
c_2	4,0	5,0
c_3	3,0	5,5
c_4	3,5	5,2

Так как $z_1(3) = z_1(2)$, $z_2(3) = z_2(2)$, выполнение алгоритма закончено, получено искомое разбиение на 2 класса: $P_1(2) = \{S_1, S_3\}$, $P_2(2) = \{S_2, S_4\}$. Эталонные элементы (представители, центры) каждого класса сформированы в процессе алгоритма.

Так как алгоритм кластеризации методом K -средних предполагает заранее заданное количество кластеров, оценка результатов процесса кластеризации может быть проведена посредством анализа расстояний между центрами классов и расстояний между объектами одного класса. В нашем случае подобный анализ показал, что кластеризация при количестве классов равном двум является удовлетворительной.

Предложенный способ представления объекта посредством РВ выгодно использовать (с точки зрения трудоемкости вычислений) и при сравнении некоего нового объекта (ситуации) с эталонами классов. Для этого необходимо получить РВ для исследуемого объекта (для каждого представителя классов соответствующие РВ получены заранее). Расстояние D между объектами S_0 и S_j определяется по формуле (3). Ближайшим к анализируемому объекту S_0 считаем тот из S_j , до которого расстояние $D(S_0, S_j)$ меньше, чем для остальных эталонов.

В реальных условиях значимость выбранных параметров для оценки ситуаций не всегда одинакова. С точки зрения экспертов, одни из них более весомы, другие менее. Эта информация может изменить состав выделенных классов и, соответственно, повлиять на сходство анализируемого объекта с эталонными представителями кластеров. Значимость параметров может быть учтена путем нахождения среднего взвешенного значений сходства объектов по каждому из выбранных параметров.

Рассмотренные метод кластеризации нечетких ситуаций и метод распознавания текущей ситуации позволяют строить классификационные модели принятия управленческих решений, способные оперировать качественной информацией полученной от экспертов и находить лучшее решение в конкретной сложившейся ситуации.

Переход от нечеткой функции принадлежности к ее репрезентативному значению не приводит к потере информации об объекте и является допустимым в рамках процедуры кластеризации. Данные выводы подтвердились в процессе проведенных исследований, когда множество ситуаций было разделено на классы также и методом иерархической агломеративной процедуры кластеризации, и в результате были получены абсолютно идентичные представители классов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Мелихов А.Н., Берштейн Л.С., Коровин С.Я. Ситуационные советующие системы с нечеткой логикой. – М.: Наука, 1990. – 272 с.
2. Карелин В.П., Кузьменко О.Л. Средства и методы поддержки принятия управленческих решений в условиях нечеткости, неопределенности и многокритериальности // Вестник Таганрогского института управления и экономики. – 2007, № 1(5). – С.73-77.
3. Леденева Т.М., Татаркин Д.С. Особенности проектирования систем нечеткого логического вывода // Информационные технологии. – 2007, №7. – С.12-18.
4. Заде Л.А. Понятие лингвистической переменной и его применение к принятию приближенных решений. – М.: Мир, 1976. – 168 с.
5. Берштейн Л.С., Карелин В.П., Целых А.Н. Модели и методы принятия решений в интегрированных интеллектуальных системах. – Ростов-на-Дону: РГУ, 1999. – 276 с.
6. Ту Дж, Гонсалес Р. Принципы распознавания образов // Пер. с англ. И.Б. Гуревича под ред. Ю.И. Журавлева. – М.: Мир, 1978. – 414 с.
7. Карелин В.П., Кузьменко О.Л. Нахождение представителя класса нечетких ситуаций при построении модели принятия решений // Известия вузов Северо-Кавказский регион. Технические науки. – 2008, №4. – С. 50-54.