КОМПЬЮТЕРНАЯ
МАТЕМАТИКА

УДК 519.9

И.И. РЯСНАЯ

О МОДЕЛИРОВАНИИ КЛАСТЕРОВ ПРЕДПРИЯТИЙ

На основе меры сходства на мультимножествах построена модель кластера предприятий. В качестве модели кластера предприятий используется понятие сегмента, как совокупности достаточно сходных нечетких мультимножеств. Предлагаемый подход позволяет формализовать и решать задачи кластеризации с учетом сходства и мощности классов эквивалентности.

Введение. Кластеры предприятий – стратегические группировки, деятельность которых направлена на достижение конкурентного преимущества [1].

Новые интеллектуальные информационные технологии, в настоящее время, оказывают существенное влияние на конкурентное преимущество отдельных предприятий, так как позволяют собирать и получать информацию, которая была недоступна ранее. Такие технологии основываются на разработке новых и развитии известных математических моделей, связанных с обработкой данных в условиях неопределенности, которая, чаще всего, порождена вербальным описанием характеристик предприятий.

Кластеры можно исследовать на разных уровнях взаимодействия внутри них, выявляя при этом различные особенности. Задача выявления и анализа кластеров может разделяться на два этапа [2]. На первом этапе формализуется понятие кластера, на втором этапе определяются дополнительные критерии и ограничения для решения задачи кластеризации как задачи оптимизации. В данной работе рассматривается модель кластера в виде совокупности достаточно сходных между собой нечетких мультимножеств, т. е. кластер определяется как специального вида нечеткое мультимножество (сегмент), сущность которого определяется его ядром.

Формализация модели кластеров. Пусть $X=\{x_i\}_{i=1}^{n}$ – заданное множество объектов, $W=\left\{w_j\right\}_{j=1}^{m}$ – множество совместимых качественных признаков объектов. Пусть $\psi : X \to W$ – соответствие, порождающее матрицу $\mathbf{\Psi}$ «объект-признак». Полагаем, что каждый объект $x_i \in X$ имеет непустой образ $\psi(x_i) \subseteq W$, а каждый признак $w_j \in W$ имеет непустой прообраз $\psi^{-1}(w_j) \subseteq X$. Очевидно, что $\psi(x_i)$ – четкое подмножество множества W. Если признаки бинарные и совместимые, то $\psi : X \to \{0,1\}^m$.

Соответствие ψ порождает отношение эквивалентности $\approx$ на X:

$$x_i \approx x_j \Leftrightarrow \psi(x_i) = \psi(x_j), \tag{1}$$

т. е. $x_i, x_j \in X$ эквивалентны, если они обладают одним и тем же набором признаков.

Введем на X бинарное отношение $T = \mathbf{\Psi} \times \mathbf{\Psi}'$, где $\mathbf{\Psi}'$ – транспонированная матрица, $\times$ – произведение матриц, $t(x_i, x_j) = \left|\psi(x_i) \cap \psi(x_j)\right|$ – мощность пересечения образов объектов $x_i, x_j \in X$.

Очевидно, что отношение T симметрично, а $t(x_i, x_i) \geq t(x_i, x_j)$. Кроме того, $t(x_i, x_i)$ равно количеству (ненулевых) признаков объекта x_i, причем $t(x_i, x_j) = 0 \Leftrightarrow \psi(x_i) \cap \psi(x_j) = \varnothing$.

Положим $\pi(x_i, x_j) = t(x_i, x_j) \Big/ \left|\psi(x_i) \cup \psi(x_j)\right|$, $i, j \in \{1,...,n\}$. Нетрудно показать, что, если $x_i \approx x_j$, то $\pi(x_i, x_j) = 1$. Если x_i и x_j принадлежат различным классам эквивалентности, то $\pi(x_i, x_j) < 1$ и $\pi(x_i, x_j) = \pi(x_j, x_i)$. Если $\psi(x_i) \cap \psi(x_j) = \varnothing$, т. е. объекты не имеют общих признаков, то $\pi(x_i, x_j) = 0$. Таким образом, функция $\pi : X \times X \to [0,1]$ порождает рефлексивное и симметричное нечеткое бинарное отношение R_π на X:

$$(x_i, x_j) \in R_\pi \Leftrightarrow \pi(x_i, x_j) > 0 .$$

Отношение R_π будем называть отношением сходства на X, а функцию $\mu_i(x_j) = \pi(x_i, x_j)$ – мерой сходства элементов $x_j \in X$, $j \in \{1,...,n\}$, с элементом $x_i \in X$.

Упорядоченное множество значений меры сходства, порожденной объектом x_i, представляет собой вектор $\vec{\mu}_i = \left(\mu_i(x_j)\right)_{j=1}^{n}$, а нечеткое множество $\tilde{X} = \{(x_i, \vec{\mu}_i)\}$ – векторнозначное нечеткое множество.

Определение 1. Нечетким C_{x_i}-кластером назовем нечеткое множество $C_{x_i} = =\left\{\left(x_j,\mu_{C_{x_i}}\left(x_j\right)\right)\right\}_{j=1}^{n} =\left\{\left(x_j,\mu_i\left(x_j\right)\right)\right\}_{j=1}^{n}$, где $\mu_i\left(x_j\right)=\pi\left(x_i,x_j\right)$, $i\in\{1,...,n\}$.

Нетрудно показать, если ядра C_{x_i}-кластера и C_{x_j}-кластера, $x_i,x_j\in X$, пересекаются, то $C_{x_i}=C_{x_j}$. Кроме того, элементы $x_i,x_j\in X$ эквивалентны тогда и только тогда, когда соответствующие им кластеры C_{x_i} и C_{x_j} совпадают:

$$x_i \approx x_j \Leftrightarrow C_{x_i} = C_{x_j}. \qquad (2)$$

Пусть $\mathbf{C}=\left\{C_{x_i}\right\}_{i=1}^{n}$. Определим биективное отображение $\xi\left(x_i\right)=C_{x_i}$, $\xi^{-1}\left(C_{x_i}\right)=x_i$, $x_i\in X$, $C_{x_i}\in\mathbf{C}$.

Согласно (2) системы $\left(X,\approx\right)$ и $\left(\mathbf{C},=\right)$ изоморфны.

Пусть $E=\{e\}$ – конечное множество элементов произвольной природы, но не являющееся мультимножеством, $N_h=\{0,1,2,...,h\}$. Определим на множестве E функцию $f:E\to N_h$, которую назовем конечной (ограниченной сверху) функцией кратности: $f\left(e\right)=k_e e=\left\{\underbrace{e,e,\,...,e}_{k_e}\right\}$, $k_e\in N_h$, где k_e – количество повторений элемента e. Очевидно, что $\forall A\subseteq E$ $f\left(A\right)$ представляет собой мультимножество. Нетрудно показать, что $\forall A\subseteq E$ мультимножество $f:2^A\to N_h$ представляет собой решетку по включению, где 2^A – множество всех подмножеств множества A.

Представим множество кластеров $\mathbf{C}=\left\{C_{x_i}\right\}_{i=1}^{n}$ в виде мультимножеств. Пусть $\mathbf{K}=\{K_l\}_{l=1}^{p}$ – множество классов эквивалентности множества X, определяемое выражением (1), $\xi\left(K_l\right)=\left\{\,C_{x_i}\,\middle|\,x_i\in K_l\right\}=Q_l$ – образ класса K_l, представляющий собой множество равных C_{x_i}-кластеров. Обозначим $\left|K_l\right|=k_l$. В соответствии с (2) $\left|K_l\right|=\left|Q_l\right|$. Пусть u_l – эталон класса эквивалентности K_l, представляющий собой произвольный элемент этого класса, $U=\{u_l\}_{l=1}^{p}$ – множество элементов, принадлежащих различным классам эквивалентности: $u_l\neq u_q$, если $l\neq q$. Множество U будем называть множеством базовых элементов или универсумом мультимножеств на множестве X.

Рефлексивную и симметричную матрицу $\left(a_{lq}\right)$, $l,q=\overline{1,p}$, будем называть нечеткой матрицей сходства элементов универсума U либо нечеткой матрицей сходства множества классов эквивалентности $\mathbf{K}=\{K_l\}_{l=1}^{p}$ множества X, где $a_{lq}=\pi\left(x_i\in K_l, x_j\in K_q\right)$, $a_{lq}\in[0,1]$, $l,q\in\{1,...,p\}$. Если $l\neq q$, то $a_{lq}=a_{ql}$. Кроме того, $a_{ll}>a_{lq}$.

Пусть $x_i\in K_l$. Очевидно, что $C_{x_i}=\left\{\left(x_j,\mu_i\left(x_j\right)\right)\right\}_{j=1}^{n}=\left\{\left(k_q\cdot\left(u_q,a_{lq}\right)\right)\right\}_{q=1}^{p}=$
$=\left\{k_1\cdot\left(u_1,a_{l1}\right),...,k_q\cdot\left(u_q,a_{lq}\right),...,k_p\cdot\left(u_p,a_{lp}\right)\right\}$, $k_1+...+k_q+...+k_p=n$.

Таким образом, C_{x_i}-кластер можно представить как нечеткое мультимножество либо как элемент *векторнозначного* мультинечеткого множества (Multi-Fuzzy Set [3]) на универсуме U. Введем следующие операции.

Объединение: $C_{x_i}\cup C_{x_l}=\left\{\left(x_j,\max\left(\mu_i\left(x_j\right),\mu_l\left(x_j\right)\right)\right)\right\}_{j=1}^{n}$, $\forall x_i,x_l\in X$.

Сложение: $C_{x_i}+C_{x_l}=\left\{\left(x_j,\mu_i\left(x_j\right)\right),\left(x_j,\mu_l\left(x_j\right)\right)\right\}_{j=1}^{n}$, $\forall x_i,x_l\in X$.

Умножения на число: $k\cdot C_{x_i}=\left\{k\cdot\left(x_j,\mu_i\left(x_j\right)\right)\right\}_{j=1}^{n}$, $\forall x_i\in X$.

Например,

$C_{x_i}+C_{x_i}=2\cdot C_{x_i}=\left\{2\cdot\left(x_j,\mu_i\left(x_j\right)\right)\right\}_{j=1}^{n}=\left\{\left(x_j,\mu_i\left(x_j\right)\right),\left(x_j,\mu_i\left(x_j\right)\right)\right\}_{j=1}^{n}$, $\forall x_i\in X$.

Для упрощения записи формул будем использовать следующий символизм: $k_q\cdot\left(u_q,a_{lq}\right)\equiv\left(u_q,\left(k_q\cdot a_{lq}\right)\right)$. Пусть $x_i\in K_l$. Тогда $\forall C_{\left(x_i\in K_l\right)}\in Q_l$ $C_{\left(x_i\in K_l\right)}=$
$=\left\{k_1\cdot\left(u_1,a_{l1}\right),...,k_q\cdot\left(u_q,a_{lq}\right),...,k_p\cdot\left(u_p,a_{lp}\right)\right\}=\left\{\left(u_1,\left(k_1\cdot a_{l1}\right)\right),...,\left(u_p,\left(k_p\cdot a_{lp}\right)\right)\right\}$.
Кроме того, $C_{\left(x_i\in K_l\right)}=\left(u_l,\vec{k}\otimes\vec{a}_l\right)$, $\vec{k}=\left(k_1,...,k_q,...,k_p\right)$, $\vec{a}_l=\left(a_{l1},...,a_{lq},...,a_{lp}\right)$, $\otimes$ – *операция*, представляющая собой “покомпонентное умножение” векторов:

$$\vec{k}\otimes\vec{a}_l=\left(k_1,...,k_q,...,k_p\right)\otimes\left(a_{l1},...,a_{lq},...,a_{lp}\right)=\left(k_1\cdot a_{l1},...,k_q\cdot a_{lq},...,k_p\cdot a_{lp}\right).$$

Если сложить все кластеры $C_{\left(x_i\in K_l\right)}\in Q_l$, то получим следующее мультинечеткое множество

$$Z_{Q_l}=\sum_{x_i\in K_l}C_{x_i}=k_l\cdot\left(u_l,\vec{k}\otimes\vec{a}_l\right).$$

В обычной записи $Z_{Q_l}=\left\{k_lk_1\cdot\left(u_1,a_{l1}\right),...,k_lk_q\cdot\left(u_q,a_{lq}\right),...,k_lk_p\cdot\left(u_p,a_{lp}\right)\right\}$.

Мощность множества Z_{Q_l} равна: $\left|Z_{Q_l}\right| = \left|\sum_{x_i \in K_l} C_{x_i}\right| = k_l \sum_{q=1}^{p} k_q a_{lq}$. Множество мультинечетких множеств $Z = \left\{Z_{Q_l}\right\}_{l=1}^{p} = \left\{k_l \Box\left(u_l, \vec{k} \otimes \vec{a}_l\right)\right\}_{l=1}^{p}$ можно также записать в виде матрицы

$$\begin{pmatrix} k_1 k_1 \Box(u_1, a_{11}) & \dots & k_1 k_p \Box(u_1, a_{1p}) \\ \vdots & \ddots & \vdots \\ k_p k_1 \Box(u_p, a_{p1}) & \cdots & k_p k_p \Box(u_p, a_{pp}) \end{pmatrix},$$

мощность множества Z равна: $|Z| = \left|\sum_{l=1}^{p} Z_{Q_l}\right| = \left|\sum_{l=1}^{p} \sum_{x_i \in K_l} C_{x_i}\right| = \sum_{l=1}^{p} \sum_{q=1}^{p} k_l k_q a_{lq} = \left|R_{\pi}\right|$.

Определение 2. Нечетким ядром уровня α C_{x_i}-кластера назовем нечеткое множество $\Omega_{C_{x_i}}^{\alpha} = \left\{\left(x_j, \mu_i\left(x_j\right)\right) \middle| \mu_i\left(x_j\right) \geq 1-\alpha, \alpha \in [0,1]\right\}$, где α достаточно малое число.

Определение 3. Границами уровня β C_{x_i}-кластера с нечетким ядром уровня α назовем обычное множество $Bound_{\beta}\ C_{x_i} = \left\{x_j \in X \middle| 0 < \beta \leq \mu_{C_{x_i}}(x_j) < 1-\alpha\right\}$.

Пусть $x_i \in K_l$, $Q_l = \xi(K_l)$, $C_{x_i} \in Q_l$, $Ker_{\alpha} C_{x_i} = \left\{x_t \middle| \mu_{C_{x_i}}(x_t) \geq 1-\alpha\right\}$ – множество элементов в носителе нечеткого ядра уровня α C_{x_i}-кластера, $t \in J_i = \left\{j_{i_1}, \dots, j_{i_q}, \dots, j_{i_{r(i)}}\right\}$, где J_i, $r(i)$ – соответственно, перечень номеров элементов и количество элементов в носителе нечеткого ядра C_{x_i}-кластера.

Определение 4. Сегментом назовем нечеткое множество $S_i(\alpha,\beta) = \bigcup_{t \in J_i} C_{x_t}$, где $x_t \in Ker_{\alpha} C_{x_i}$, $0 < \beta < 1-\alpha$.

Функция принадлежности элемента $x_j \in X$ сегменту $S_i(\alpha,\beta)$, порожденному нечетким ядром C_{x_i}-кластера, $C_{x_i} \in Q_l$, вычисляется следующим образом:

$$\mu_{S_i(\alpha,\beta)}(x_j) = \begin{cases} 1, & 1-\alpha \leq \mu_{C_{x_i}}(x_j) \leq 1, \\ \max\limits_{j}\left\{\mu_{C_{x_t}}(x_j), t \in J_i\right\}, & \beta \leq \mu_{C_{x_t}}(x_j) < 1-\alpha, \\ 0, & 0 \leq \mu_{C_{x_t}}(x_j) < \beta, t \in J_i. \end{cases} \quad (3)$$

Сегмент $S_i(\alpha,\beta)$ – нечеткое мультимножество на универсуме U.

Мощность ядра сегмента имеет две оценки. Первая равна мощности нечеткого ядра уровня α C_{x_i} -кластера, $C_{x_i} \in Q_l$, которая вычисляется по формуле $\left|\Omega^{\alpha}_{\left(C_{x_i} \in Q_l\right)}\right| = \sum_{q=j_{i_1}}^{j_{i_{r(i)}}} k_q a_{lq}$. Вторая – мощности носителя нечеткого ядра уровня α этого кластера, равна мощности ядра нечеткого множества с функцией принадлежности, вычисляемой по формуле (3), $\left|Ker_{\alpha} C_{x_i}\right| = \sum_{q=j_{i_1}}^{j_{i_{r(i)}}} k_q$.

Заключение. Представление кластеров (сегментов), как совокупности достаточно сходных нечетких мультимножеств, позволяет формализовать и решать задачи кластеризации с учетом сходства и мощности классов эквивалентности. На первом этапе решения задачи кластеризации строится модель сегментов, на втором – используя дополнительные критерии и ограничения, решается оптимизационная задача.

І.І. Рясна

ПРО МОДЕЛЮВАННЯ КЛАСТЕРІВ ПІДПРИЄМСТВ

На основі міри подібності на мультимножинах побудована модель кластера підприємств. Як модель кластера використовується сукупність досить подібних нечітких мультимножин. Запропонований підхід дозволяє формалізувати і розв'язувати задачі кластеризації з урахуванням подібності та потужності класів еквівалентності.

I.I. Rjasnaja

ON MODELING OF CLUSTERS OF ENTERPRISES

A model of similarity relation of the enterprises on the basis of a measure of similarity on multisets is constructed. A set of sufficiently similar fuzzy multisets is used as a cluster model. The approach proposed allows to formalize and to solve problems of clustering taking into account similarity and cardinality of equivalence classes.

1. *Портер М.* Конкуренция. – М.: Издательский дом «Вильямс», 2003. – 496 с.
2. *Руспини Э.Г.* Последние достижения в нечетком кластер-анализе // Нечеткие множества и теория возможностей. Последние достижения. – М.: Радио и связь, 1986. – С. 114–132.
3. //nsmv2007.ulsu.ru/index.php?url=19&type=file.

Получено 15.03.2011

Об авторе:

Рясная Ирина Ивановна,
научный сотрудник Института кибернетики имени В.М. Глушкова НАН Украины.