УДК 519.8

И.И. РЯСНАЯ

МЕТОД НЕЧЕТКОЙ КЛАСТЕРИЗАЦИИ

Рассматривается метод нечеткой кластеризации при условии, что дополнение нечеткого отношения сходства является метрикой. На базе введения понятий пороговой нормы и пороговой конормы решается задача разбиения исследуемого множества на непересекающиеся кластеры. Предлагаемый метод отличается от методов кластеризации, использующих нечеткое отношение эквивалентности, тем, что позволяет создавать более быстрые алгоритмы построения кластеров.

Введение. Методы нечеткой кластеризации с использованием аппроксимационного подхода при условии, что дополнение нечеткого отношения сходства является метрикой на исследуемом конечном множестве, рассматриваются в [1]. Для решения задачи кластеризации в метрическом пространстве используется транзитивное замыкание нечеткого отношения сходства, порождающее ультраметрику [2]. Однако в [3] отмечено, что классы эквивалентности, полученные на основе транзитивного замыкания, не всегда поддаются содержательной интерпретации из-за специфических свойств ультраметрики.

В данной статье показано, что в метрическом пространстве на базе введения понятий пороговой нормы и пороговой конормы можно, не используя операцию транзитивного замыкания, решить задачу разбиения исследуемого множества на непересекающиеся кластеры. Для упрощения записи формул будем, как правило, использовать одни и те же символы для обозначения нечетких множеств и функций принадлежности этих нечетких множеств (символ $\Diamond$ означает конец доказательства леммы или теоремы). Для нечетких множеств символы $\bigcup$ и $\bigcap$ соответствуют применению стандартных операций объединения ($\max$) и пересечения ($\min$) этих множеств.

Определения, теоремы, описание метода. Пусть τ – нечеткое отношение сходства на конечном множестве X , т. е. рефлексивное и симметричное нечеткое бинарное отношение. Функция принадлежности нечеткого отношения сходства удовлетворяет условиям $\tau(x,x)=1 \ \ \forall x\in X$, $\tau(x,y)=\tau(y,x)$

$\forall x, y \in X$. Элементы $x, y \in X$ называются сходными, если $\tau(x, y) > 0$, т. е. пара $(x, y) \in \tau$. Для несходных элементов $(x, y) \notin \tau$ или $\tau(x, y) = 0$. Для нечеткого отношения сходства на конечном множестве X существует пара элементов, для которой значение сходства минимально $\tau(x, y) = \alpha_{\min}$, $0 < \alpha_{\min} \le 1$.

Дополнение нечеткого отношения сходства $\rho = \overline{\tau}$ называется нечетким отношением несходства, функция принадлежности которого удовлетворяет условиям: $\rho(x, y) = 1 - \tau(x, y)$, $\rho(x, x) = 0 \quad \forall x \in X$, $\rho(x, y) = \rho(y, x) \quad \forall x, y \in X$. Для сходных элементов $x, y \in X$ – $\rho(x, y) < 1$, для несходных – $\rho(x, y) = 1$.

Определение 1. Нечеткое подмножество C_x множества X, $x \in X$, функция принадлежности которого $c_x(y) = \tau(x, y)$, $y \in X$, назовем C_x-кластером.

В матричном представлении отношения сходства $c_x(y)$ – строка матрицы, соответствующая элементу $x \in X$.

Определение 2. Нечетким отношением сходства уровня α, $\alpha \in [0,1]$, назовем нечеткое отношение сходства $\tau^{(\alpha)}$, функция принадлежности которого $\forall x, y \in X$ удовлетворяет следующим условиям:

$$\tau^{(\alpha)}(x, y) = \begin{cases} \tau(x, y), & \text{если } \tau(x, y) \ge \alpha, \\ 0, & \text{если } \tau(x, y) < \alpha. \end{cases} \tag{1}$$

Нетрудно видеть, что $\tau^{(\alpha)}(x, y) \le \tau(x, y)$, т.е. $\tau^{(\alpha)} \subseteq \tau$. Если выбрать $\alpha \le \alpha_{\min}$, то $\tau^{(\alpha)} = \tau$, если $\alpha = 1$, то $\tau^{(\alpha)}$ – диагональное отношение.

Определение 3. Нечеткое подмножество $C_x^{(\alpha)}$ множества X, $x \in X$, $\alpha \in [0,1]$, функция принадлежности которого $c_x^{(\alpha)}(y) = \tau^{(\alpha)}(x, y)$, $y \in X$, назовем $C_x^{(\alpha)}$-кластером.

Как известно [4], носителем нечеткого подмножества $A \subseteq X$ называется множество $\operatorname{supp} A = \{x \mid \mu_A(x) > 0\}$, где $\mu_A(x)$ – функция принадлежности нечеткого подмножества A.

Лемма 1. Если $\alpha_1 > \alpha_2$, то $\tau^{(\alpha_1)} \subseteq \tau^{(\alpha_2)}$, $C_x^{(\alpha_1)} \subseteq C_x^{(\alpha_2)}$ и $\operatorname{supp} C_x^{(\alpha_1)} \subseteq \subseteq \operatorname{supp} C_x^{(\alpha_2)}$.

Доказательство следует из теоремы о разложении нечетких множеств [3].

Определение 4. Элементы $x, y \in X$ назовем α-сходными, если $(x, y) \in \tau^{(\alpha)}$.

Лемма 2. Если два элемента $x, y \in X$ α-сходны, то они входят в носитель общего для них кластера.

Доказательство леммы очевидно: существует, как минимум, два таких кластера $C_x^{(\alpha)}$ и $C_y^{(\alpha)}$. Обратная лемма неверна.

Множество нечетких подмножеств $\mathbf{B}=\{B_i\}$ называется нечетким покрытием множества A, если $\bigcup_i B_i = A$ [5].

Лемма 3. Множество нечетких кластеров $\mathbf{C}^{(\alpha)} = \left\{C_x^{(\alpha)} \mid x \in X\right\}$, $\alpha \in [0,1]$, является нечетким покрытием множества X :

$$\bigcup_{x\in X} C_x^{(\alpha)} = X . \tag{2}$$

Доказательство. Если $\alpha = 1$, то $x \in C_x^{(1)}$ $\forall x \in X$, и равенство (2) выполняется. В соответствии с леммой 1, если $\alpha < 1$, то $C_x^{(\alpha)} \supseteq C_x^{(1)}$, и равенство (2) тем более выполняется. ◊

Определение 5. Нечетким отношением несходства уровня δ, $\delta \in [0,1]$, назовем нечеткое отношение несходства $\rho^{(\delta)}$, функция принадлежности которого $\forall x, y \in X$ удовлетворяет условиям

$$\rho^{(\delta)}(x,y) = \begin{cases} \rho(x,y), & \text{если } \rho(x,y) \le \delta, \\ 1, & \text{если } \rho(x,y) > \delta. \end{cases} \tag{3}$$

Очевидно, что $\rho^{(\delta)}(x,y) \ge \rho(x,y)$, т.е. $\rho^{(\delta)} \supseteq \rho$. Если $\delta = 1-\alpha$, то согласно (1), (3) следует, что $\tau^{(\alpha)}(x,y) + \rho^{(\delta)}(x,y) = 1$.

Треугольной нормой называется функция $T:[0,1]\times[0,1] \to [0,1]$, удовлетворяющая условиям монотонности, ассоциативности и коммутативности, а также $T(0,0)=0$, $T(1,a)=a$ [4].

Треугольной конормой называется функция $S:[0,1]\times[0,1] \to [0,1]$, удовлетворяющая условиям монотонности, ассоциативности и коммутативности, а также $S(1,1)=1$, $S(0,a)=a$ [4].

Для двойственных треугольных норм и конорм имеет место равенство

$$T(a,b) + S(1-a,1-b) = 1 . \tag{4}$$

Определение 6. Пороговой треугольной нормой или треугольной нормой с порогом α, $\alpha \in [0,1]$, назовем функцию

$$T^{(\alpha)}(a,b) = \begin{cases} T(a,b), & \text{если } T(a,b) \ge \alpha, \\ 0, & \text{если } T(a,b) < \alpha, \end{cases} \tag{5}$$

где $T(a,b)$ – треугольная норма.

Функция $T^{(\alpha)}$ удовлетворяет условиям монотонности, ассоциативности и коммутативности, а также $T^{(\alpha)}(0,0)=0$ и, если $a \geq \alpha$, то $T^{(\alpha)}(1,a)=a$, иначе $T^{(\alpha)}(1,a)=0$. Кроме того, $T^{(\alpha)}(a,b) \leq T(a,b)$, если $\alpha_1 > \alpha_2$, то $T^{(\alpha_1)}(a,b) \leq T^{(\alpha_2)}(a,b)$.

Определение 7. Пороговой треугольной конормой или треугольной конормой с порогом δ, $\delta \in [0,1]$, назовем функцию

$$S^{(\delta)}(a,b) = \begin{cases} S(a,b), & \text{если } S(a,b) \leq \delta, \\ 1, & \text{если } S(a,b) > \delta, \end{cases} \tag{6}$$

где $S(a,b)$ – треугольная конорма.

Функция $S^{(\delta)}$ удовлетворяет условиям монотонности, ассоциативности и коммутативности, а также $S^{(\delta)}(1,1)=1$ и, если $a \leq \delta$, то $S^{(\delta)}(0,a)=a$, иначе $S^{(\delta)}(0,a)=1$. Очевидно, что $S(a,b) \leq S^{(\delta)}(a,b)$. Если $\delta_1 > \delta_2$, то $S^{(\delta_1)}(a,b) \leq S^{(\delta_2)}(a,b)$.

Лемма 4. Если T и S – двойственные треугольные норма и конорма, соответственно, и $\delta = 1-\alpha$, то пороговые треугольные норма $T^{(\alpha)}$ и конорма $S^{(\delta)}$ также двойственные:

$$T^{(\alpha)}(a,b) + S^{(\delta)}(1-a,1-b) = 1. \tag{7}$$

Доказательство. Согласно (4)–(6), для $T(a,b) \geq \alpha$, $S(1-a,1-b) \leq 1-\alpha = \delta$, следовательно, $T^{(\alpha)}(a,b) = T(a,b) = 1 - S(1-a,1-b) = 1 - S^{(\delta)}(1-a,1-b)$, т.е. $T^{(\alpha)}(a,b) + S^{(\delta)}(1-a,1-b) = 1$. Если $T(a,b) < \alpha$, то $T^{(\alpha)}(a,b)$=0, а $1-T(a,b) =$ $= S(1-a,1-b) > \delta$, т.е. $S^{(\delta)}(1-a,1-b) = 1$ и равенство (7) также выполняется.◊

В дальнейшем полагаем, что $\delta = 1-\alpha$.

Для конормы Лукасевича $S_L(a,b) = \min(1, a+b)$ пороговая треугольная конорма вычисляется следующим образом:

$$S_L^{(\delta)}(a,b) = \begin{cases} a+b, & \text{если } a+b \leq \delta, \\ 1, & \text{если } a+b > \delta. \end{cases} \tag{8}$$

Определение 8. Функцию принадлежности $\rho(x,y)$ отношения несходства будем называть S_L-метрикой на X, если

$$\rho(x,y) = \min_z S_L(\rho(x,z), \rho(z,y)).$$

Иначе говоря,

$$S_L\left(\rho(x,z),\rho(z,y)\right)\geq S_L\left(\rho(x,y),\rho(y,y)\right)=\rho(x,y). \qquad (9)$$

Обозначим $S_L(a,b)$ как $\oplus$, тогда из (9) получим неравенство треугольника

$$\rho(x,z)\oplus\rho(z,y)\geq\rho(x,y). \qquad (10)$$

Теорема 1. Если $\rho(x,y)$ – S_L-метрика, то $\rho^{(\delta)}(x,y)$ – метрика по пороговой треугольной конорме $S_L^{(\delta)}$.

Доказательство. Следует доказать, что, если выполняется условие (10), то

$$S_L^{(\delta)}\left(\rho^{(\delta)}(x,z),\rho^{(\delta)}(z,y)\right)\geq S_L^{(\delta)}\left(\rho^{(\delta)}(x,y),\rho^{(\delta)}(y,y)\right)=\rho^{(\delta)}(x,y). \qquad (11)$$

Рассмотрим возможные случаи.

а) Если $\rho(x,z)+\rho(z,y)>\delta$, то, поскольку $\rho(x,y)\leq\rho^{(\delta)}(x,y)$, согласно (8) $S_L^{(\delta)}\left(\rho^{(\delta)}(x,z),\rho^{(\delta)}(z,y)\right)=1$. Если $\rho(x,y)\leq\delta$, то $\rho^{(\delta)}(x,y)=\rho(x,y)$, и условие (11) выполняется. Если $\rho(x,y)>\delta$, то $\rho^{(\delta)}(x,y)=1$, и условие (11) также выполняется.

б) Если $\rho(x,z)+\rho(z,y)\leq\delta$, то $\rho(x,z)\leq\delta$ и $\rho(z,y)\leq\delta$, а, согласно (10), $\rho(x,y)\leq\delta$. Из (3) следует, что $\rho^{(\delta)}(x,z)=\rho(x,z)$, $\rho^{(\delta)}(z,y)=\rho(z,y)$, $\rho^{(\delta)}(x,y)=\rho(x,y)$, тогда $S_L^{(\delta)}\left(\rho^{(\delta)}(x,z),\rho^{(\delta)}(z,y)\right)=S_L\left(\rho(x,z),\rho(z,y)\right)\geq$ $\geq\rho(x,y)=\rho^{(\delta)}(x,y)$. ◊

Метрику по пороговой треугольной конорме $S_L^{(\delta)}$ будем называть $S_L^{(\delta)}$-метрикой.

Далее полагаем, что $\rho(x,y)$ – S_L-метрика, тогда $\rho^{(\delta)}(x,y)$ – $S_L^{(\delta)}$-метрика.

Определение 9. Расстоянием на множестве нечетких кластеров $\mathbf{C}=\{C_x\mid x\in X\}$ назовем такую функцию $d:\mathbf{C}\times\mathbf{C}\to[0,1]$, что

$$d\left(C_x,C_y\right)=1-c_x(y)=\min_z S_L\left(1-c_x(z),1-c_y(z)\right), \qquad (12)$$

где $S_L(a,b)$ – конорма Лукасевича.

Функция d – симметрична и антирефлексивна, а условие (11) порождает ограничение, представляющее собой аналог неравенства треугольника

$$S_L\left(1-c_x(z),1-c_y(z)\right)\geq S_L\left(1-c_x(y),1-c_y(y)\right)=1-c_x(y). \qquad (13)$$

Пусть $\xi(x)=C_x$, $\xi^{-1}(C_x)=x$, $x\in X$, $C_x\in\mathbf{C}$.

Лемма 5. Биекция ξ между метрическим пространством (X,ρ) и метрическим пространством $(\mathbf{C},d)$ является изометрией.

Доказательство. Так как $1-c_x(z)=\rho(x,z)$, $1-c_y(z)=\rho(y,z)$, $1-c_x(y)=$ $=\rho(x,y)$, то (13) можно, с учетом симметричности отношения несходства, записать как $S_L(\rho(x,z),\rho(z,y))\ge S_L(\rho(x,y),\rho(y,y))=\rho(x,y)$, что в точности совпадает с (9). Следовательно, $d(C_x,C_y)=\rho(x,y)$, где $\rho(x,y)$ – S_L-метрика.◊

Определим биективное отображение $\xi_\alpha : X\to\mathbf{C}^{(\alpha)}$ следующим образом: $\forall x\in X$ $\xi_\alpha(x)=C_x^{(\alpha)}$ и $\forall C_x^{(\alpha)}\in\mathbf{C}^{(\alpha)}$ обратное отображение $\xi_\alpha^{-1}\left(C_x^{(\alpha)}\right)=x$.

Лемма 6. Биекция ξ_α между метрическим пространством $\left(X,\rho^{(\delta)}\right)$ и метрическим пространством $\left(\mathbf{C}^{(\alpha)},d\right)$, где расстояние $d\left(C_x^{(\alpha)},C_y^{(\alpha)}\right)=1-c_x^{(\alpha)}(y)$ $\forall C_x^{(\alpha)},C_y^{(\alpha)}\in\mathbf{C}^{(\alpha)}$, является изометрией.

Иначе говоря, $d\left(C_x^{(\alpha)},C_y^{(\alpha)}\right)=\rho^{(\delta)}(x,y)$, где $\rho^{(\delta)}(x,y)$ – $S_L^{(\delta)}$-метрика.

Доказательство аналогично доказательству леммы 5.

Определение 10. Кластеры $C_x^{(\alpha)},C_y^{(\alpha)}\in\mathbf{C}^{(\alpha)}$, $x,y\in X$, назовем соседними, если расстояние между ними $d\left(C_x^{(\alpha)},C_y^{(\alpha)}\right)\le\delta$, где $\delta=1-\alpha$.

Если $C_x^{(\alpha)}$, $C_y^{(\alpha)}$ – соседние кластеры, то $y\in\operatorname{supp}C_x^{(\alpha)}$, $x\in\operatorname{supp}C_y^{(\alpha)}$. Если кластеры $C_x^{(\alpha)},C_y^{(\alpha)}$ не являются соседними, то $d\left(C_x^{(\alpha)},C_y^{(\alpha)}\right)=1$.

Определение 11. Множество кластеров $\mathbf{O}_x=\left\{C_y^{(\alpha)}\mid y\in\operatorname{supp}C_x^{(\alpha)},\ y\ne x\right\}$ назовем δ-окрестностью кластера $C_x^{(\alpha)}$.

Очевидно, что $\forall C_y^{(\alpha)}\in\mathbf{O}_x$ $d\left(C_x^{(\alpha)},C_y^{(\alpha)}\right)\le\delta$, т. е. любой из этих кластеров является соседним с кластером $C_x^{(\alpha)}$.

Определение 12. Кластеры $C_x^{(\alpha)},C_y^{(\alpha)}\in\mathbf{C}^{(\alpha)}$, $x,y\in X$, назовем смежными, если $\bigcap\left(C_x^{(\alpha)},C_y^{(\alpha)}\right)\ne\varnothing$, а расстояние между ними $d\left(C_x^{(\alpha)},C_y^{(\alpha)}\right)=1$.

Для смежных кластеров существует хотя бы один общий соседний кластер $C_z^{(\alpha)}$, такой, что $z \in \bigcap\left(\operatorname{supp} C_x^{(\alpha)}, \operatorname{supp} C_y^{(\alpha)}\right)$.

Если $\bigcap\left(C_x^{(\alpha)}, C_y^{(\alpha)}\right) \neq \varnothing$, то эти кластеры – либо соседние, либо смежные.

Определение 13. Кластеры $C_x^{(\alpha)}, C_y^{(\alpha)} \in \mathbf{C}^{(\alpha)}$ назовем связанными, если существует такая совокупность кластеров $C_x^{(\alpha)} = C_{z_0}^{(\alpha)}$, $C_{z_1}^{(\alpha)}$, $C_{z_2}^{(\alpha)}$, …, $C_{z_{n-1}}^{(\alpha)}$, $C_{z_n}^{(\alpha)} = C_y^{(\alpha)}$, что $\bigcap\left(C_{z_0}^{(\alpha)}, C_{z_1}^{(\alpha)}\right) \neq \varnothing, \bigcap\left(C_{z_1}^{(\alpha)}, C_{z_2}^{(\alpha)}\right) \neq \varnothing, \ldots, \bigcap\left(C_{z_{n-1}}^{(\alpha)}, C_{z_n}^{(\alpha)}\right) \neq \varnothing$; $C_{z_i}^{(\alpha)} \in \mathbf{C}^{(\alpha)}$, $i \in \{0, \ldots, n\}$.

Очевидно, что соседние и смежные кластеры являются связанными. Однако связанные кластеры не являются, в общем случае, соседними или смежными.

Лемма 7. Между связанными кластерами $C_x^{(\alpha)}, C_y^{(\alpha)} \in \mathbf{C}^{(\alpha)}$ существует путь, в котором расстояние между кластерами не превышает δ в метрике $S_L^{(\delta)}$.

Другими словами, такой путь проходит через соседние кластеры. Назовем такой путь δ-связным.

Доказательство. Пусть $n = 1$ (определение (13)), тогда $\bigcap\left(C_x^{(\alpha)}, C_y^{(\alpha)}\right) \neq \varnothing$. Кластеры $C_x^{(\alpha)}$, $C_y^{(\alpha)}$ – соседние или смежные. В первом случае $d\left(C_x^{(\alpha)}, C_y^{(\alpha)}\right) \leq \delta$. Во втором случае существует общий соседний кластер $C_z^{(\alpha)}$, такой, что $C_z^{(\alpha)} \in \mathbf{O}_x$ и $C_z^{(\alpha)} \in \mathbf{O}_y$, $z \in \operatorname{supp} C_x^{(\alpha)}$, $z \in \operatorname{supp} C_y^{(\alpha)}$. Иначе говоря, $d\left(C_x^{(\alpha)}, C_z^{(\alpha)}\right) \leq \delta$, $d\left(C_z^{(\alpha)}, C_y^{(\alpha)}\right) \leq \delta$, и путь между кластерами $C_x^{(\alpha)}, C_y^{(\alpha)}$ – δ-связный. Если кластеры $C_{z_0}^{(\alpha)} = C_x^{(\alpha)}$ и $C_{z_{k-1}}^{(\alpha)}$, $k < n$, связаны, то аналогично доказывается существование δ-связного пути между кластерами $C_{z_{k-1}}^{(\alpha)}, C_{z_k}^{(\alpha)}$, т.е. кластеры $C_{z_0}^{(\alpha)}, C_{z_k}^{(\alpha)}$ связаны. По индукции следует наличие δ-связного пути для $k = n$. ◊

Если между двумя кластерами не существует δ-связного пути, то эти кластеры назовем несвязанными.

Определение 14. Связным предклассом назовем множество кластеров $\left\{C_x^{(\alpha)}\right\}$, в котором любые два кластеры связаны.

Другими словами, между любыми двумя элементами связного предкласса существует δ-связный путь. Очевидно, что множество кластеров $\mathbf{O}_x$ представляет собой связный предкласс.

Определение 15. Связным классом или Δ-кластером назовем максимальный предкласс.

Пусть $N = |X|$.

Теорема 2. Множество Δ-кластеров $\mathbf{\Sigma} = \left\{\Delta_j\right\}_{j=1}^{m}$, $2 \le m \le N$, представляет собой разбиение множества $\mathbf{C}^{(\alpha)}$.

Доказательство. Предположим, что Δ-кластеры пересекаются, т. е. $\bigcap\left(\Delta_i, \Delta_j\right) \ne \varnothing$, $i \ne j$, $i, j \in \{1,...,m\}$, тогда существует кластер $C_z^{(\alpha)} \in \bigcap\left(\Delta_i, \Delta_j\right)$ и кластеры Δ_i, Δ_j оказываются связанными, т.е. не являются максимальными предклассами, что противоречит определению 15. Обозначим $|\Delta_j| = k_j$, $j = \overline{1,m}$. Если $k_j = 1$, то кластер Δ_j содержит только один элемент. Так как $\forall C_x^{(\alpha)} \in \mathbf{C}^{(\alpha)}$ $\exists \Delta_j \in \mathbf{\Sigma}$ и $\forall \Delta_j \in \mathbf{\Sigma}$ $\exists C_x^{(\alpha)} \in \mathbf{C}^{(\alpha)}$, то $\bigcup\limits_{j=1}^{m} \Delta_j = \mathbf{C}^{(\alpha)}$ и $\sum\limits_{j=1}^{m} k_j = N$. ◊

Назовем D-кластерами множества вида $D_j = \left\{x \mid x = \xi_\alpha^{-1}\left(C_x^{(\alpha)}\right), C_x^{(\alpha)} \in \Delta_j\right\}$.

Теорема 3. Объединение носителей кластеров $C_x^{(\alpha)} \in \mathbf{C}^{(\alpha)}$, входящих в один и тот же кластер Δ_j, $j \in \{1,...,m\}$, порождает классы эквивалентности на множестве X, которые представляют собой D-кластеры, а множество этих кластеров $C_x^{(\alpha)}$ является нечетким покрытием соответствующего кластера D_j.

Доказательство. Поскольку Δ_j – максимальный предкласс, $j \in \{1,...,m\}$, то $\forall z \in \bigcup\limits_{C_x^{(\alpha)} \in \Delta_j} \operatorname{supp} C_x^{(\alpha)}$ $\exists C_z^{(\alpha)} \in \Delta_j$, т.е. $z = \xi_\alpha^{-1}\left(C_z^{(\alpha)} \mid C_z^{(\alpha)} \in \Delta_j\right)$. Следовательно, $\bigcup\limits_{C_x^{(\alpha)} \in \Delta_j} \operatorname{supp} C_x^{(\alpha)} = D_j$, $|\Delta_j| = |D_j|$, $\bigcup\limits_{C_x^{(\alpha)} \in \Delta_j} C_x^{(\alpha)} = D_j$. Согласно теореме 2 множество $\left\{\Delta_j\right\}_{j=1}^{m}$ – разбиение множества $\mathbf{C}^{(\alpha)}$, а между элементами $x \in D_j$ и кластерами $C_x^{(\alpha)} \in \Delta_j$ существует взаимно-однозначное соответствие. Поэтому $\bigcap\left(D_i, D_j\right) = \varnothing$, $i \ne j$, $i, j \in \{1,...,m\}$, $\bigcup_{j=1}^{m} D_j = X$. Таким образом, множество

$\Delta_j = \left\{C_x^{(\alpha)}\right\}$ – нечеткое покрытие кластера D_j, а множество $\left\{D_j\right\}_{j=1}^{m}$ – разбиение множества X. Обозначим R_D отношение эквивалентности, порождаемое множеством непересекающихся кластеров $\left\{D_j\right\}_{j=1}^{m}$, а отношение эквивалентности, порождаемое множеством непересекающихся кластеров $\left\{\Delta_j\right\}_{j=1}^{m}$, – R_Δ. Тогда $R_D(x, y) \Leftrightarrow R_\Delta\left(C_x^{(\alpha)}, C_y^{(\alpha)}\right)$.◊

Определим на множестве $\mathbf{C}^{(\alpha)}$ отношение θ, такое, что $\forall C_x^{(\alpha)}, C_y^{(\alpha)} \in \mathbf{C}^{(\alpha)}$ функция принадлежности $\theta\left(C_x^{(\alpha)}, C_y^{(\alpha)}\right) = 1 - d\left(C_x^{(\alpha)}, C_y^{(\alpha)}\right)$. Отношение θ – симметрично и рефлексивно, т. е. представляет собой нечеткое отношение сходства.

Вышеизложенное позволяет сформулировать следующую теорему, выражающую сущность предлагаемого метода кластеризации.

Теорема 4. Биективное отображение ξ_α представляет собой изоморфизм систем $\left\langle X, \tau^{(\alpha)}, \rho^{(\delta)}, R_D \right\rangle$ и $\left\langle \mathbf{C}^{(\alpha)}, \theta, d, R_\Delta \right\rangle$.

Доказательство. Согласно лемме 6 имеет место равенство $\rho^{(\delta)}(x, y) = d\left(C_x^{(\alpha)}, C_y^{(\alpha)}\right)$, тогда из определения отношения θ следует $\tau^{(\alpha)}(x, y) = \theta\left(C_x^{(\alpha)}, C_y^{(\alpha)}\right)$. Из теоремы 3 следует, что $R_D(x, y) = R_\Delta\left(C_x^{(\alpha)}, C_y^{(\alpha)}\right)$.◊

Опишем кратко предлагаемый метод кластеризации. Задачу кластеризации считаем решенной, когда число кластеров $m \geq 2$. Практический интерес представляет исследование связи между структурой и количеством Δ-кластеров, которая изменяется в зависимости от значения порога α. Для проведения такого исследования задается набор пороговых значений $\alpha_1 < \alpha_2 < \ldots < \alpha_n$, где $\alpha_1 \geq \alpha_{\min}$, $\alpha_n < 1$. Для каждого значения порога α строится множество кластеров $\mathbf{C}^{(\alpha)} = \left\{C_x^{(\alpha)} \mid x \in X\right\}$ и на основе объединения носителей связанных кластеров $C_x^{(\alpha)}$ строится множество кластеров $\left\{D_j\right\}$. Поскольку между кластерами D_j и Δ_j существует взаимно-однозначное соответствие, то кластер D_j содержит перечень (имена) кластеров $C_x^{(\alpha)}$, входящих в кластер Δ_j. Характеристическое свойство D-кластеров и Δ-кластеров состоит в существовании пути между любыми элементами одного и того же кластера, в котором расстояние между соседними элементами не превышает заданного значения δ, $\delta = 1 - \alpha$, и отсутствии такого пути между элементами различных кластеров.

Представляет интерес сравнить вышерассмотренный метод кластеризации и метод кластеризации на базе нечеткого отношения эквивалентности.

Пусть $\hat{\tau}^{(\alpha)}$ – нечеткое отношение эквивалентности, полученное путем $\max-\min$ транзитивного замыкания нечеткого отношения сходства $\tau^{(\alpha)}$, а $\hat{\rho}^{(\delta)}$ – дополнение нечеткого отношения эквивалентности, $\alpha = 1 - \delta$, $\hat{\rho}^{(\delta)}(x, y) = 1 - \hat{\tau}^{(\alpha)}(x, y)$.

Нетрудно показать, что $\min-\max$ транзитивное замыкание отношения несходства $\rho^{(\delta)}$ равно $\hat{\rho}^{(\delta)}$.

Теорема 5. Классы эквивалентности множества уровня α нечеткого отношения эквивалентности $\hat{\tau}^{(\alpha)}$ представляют собой D-кластеры на множестве X.

Доказательство. Пусть $x, y \in K$, где K – класс эквивалентности, тогда $\hat{\tau}^{(\alpha)}(x, y) \geq \alpha$ и $\hat{\rho}^{(\delta)}(x, y) \leq \delta$. Аналогично [3] нетрудно показать, что построение $\min-\max$ транзитивного замыкания отношения несходства $\rho^{(\delta)}$, представленного в виде конечного нечеткого графа, эквивалентно вычислению длины путей с помощью операции $\max$ и выбору пути с помощью операции $\min$. Неравенство $\hat{\rho}^{(\delta)}(x, y) \leq \delta$ означает существование δ-связного пути между этими элементами в метрическом пространстве $(X, \rho^{(\delta)})$. Для любых трех элементов $x, y, z \in X$ из $\hat{\rho}^{(\delta)}(x, y) \leq \delta$ и $\hat{\rho}^{(\delta)}(y, z) \leq \delta$ следует, что $\hat{\rho}^{(\delta)}(x, z) \leq \delta$. Следовательно, между этими элементами существует δ-связный путь в пространстве $(X, \rho^{(\delta)})$, т.е. $x, y, z \in K$. По индукции следует, что δ-связный путь существует между всеми элементами, принадлежащими K. Следовательно, класс эквивалентности K представляет собой D-кластер на множестве X. $\Diamond$

Сравним вычислительную сложность алгоритмов построения классов эквивалентности отношения $\hat{\tau}^{(\alpha)}$ и D-кластеров на множестве X. Так как вычислительная сложность построения D-кластеров по матрице $\tau^{(\alpha)}(x, y)$ и классов эквивалентности по матрице $\hat{\tau}^{(\alpha)}(x, y)$ одинакова, а для построения матрицы $\hat{\tau}^{(\alpha)}(x, y)$ требуется $O\left(N^4\right)$ операций, то алгоритм построения D-кластеров оказывается существенно более быстродействующим.

Заключение. Предлагаемый метод кластеризации базируется на том, что дополнение нечеткого отношения сходства (отношение несходства) является метрикой. Использование набора пороговых значений сходства позволяет построить систему вложенных кластеров, на основе которой можно проводить дальнейший анализ исследуемого множества. Данный метод отличается от методов кластеризации на основе нечеткого отношения эквивалентности тем, что позволяет разрабатывать более быстрые алгоритмы построения кластеров.

І.І. Рясна

МЕТОД НЕЧІТКОЇ КЛАСТЕРИЗАЦІЇ

Розглянуто метод нечіткої кластеризації за умови, що доповнення нечіткого відношення схожості є метрикою. На основі введення понять норм та конорм з порогом розв'язується задача розбиття множини, що досліджується, на кластери, які не перетинаються. Запропонований метод відрізняється від методів, що використовують нечітке відношення еквівалентності тим, що дозволяє створювати більш швидкі алгоритми для побудови кластерів.

I.I. Rjasnaja

FUZZY CLUSTERING METHOD

The method of fuzzy clustering is examined under the condition that a complement of fuzzy similarity relation is a metric. On the basis of definition of concepts of threshold norm and threshold conorm, the problem of set partitioning on nonintersecting clusters is solved. The method proposed differs from the methods using the fuzzy equivalence relation by the fact that it allows to create faster algorithms of construction of clusters.

1. *Руспини Э.Г.* Последние достижения в нечетком кластер-анализе // Нечеткие множества и теория возможностей. Последние достижения. – М.: Радио и связь, 1986. – С. 114–132.
2. *Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И.* Методы и модели анализа данных: OLAP и Data Mining. – СПб.: БХВ-Петербург, 2004. – 336 с.
3. *Кофман А.* Введение в теорию нечетких множеств. – М.: Радио и связь, 1982. – 432 с.
4. *Нечеткие* множества в моделях управления и искусственного интеллекта / Под ред. Д.А. Поспелова. – М.: Наука, 1986. – 312 с.
5. *Леоненков А.* Нечеткое моделирование в среде MATLAB и fuzzyTECH. – Санкт-Петербург: БХВ–Петербург, 2003. – 724 с.

Получено 15.04.2010

Об авторе:

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