УДК 004(075.32)

О СВОЙСТВАХ НЕЧЕТКОГО ОТНОШЕНИЯ СХОДСТВА

Т. М. Леденева, Р. К. Стрюков

Воронежский государственный университет

Поступила в редакцию 13.11.2014 г.

Аннотация. В статье рассматриваются свойства нечеткого отношения сходства и его взаимосвязь с покрытиями заданного множества объектов..
Ключевые слова: сходство, покрытие, степень уверенности, подобие.
Annotation. The article deals with the properties of fuzzy relations of similarity and its relationship with coatings of objects.
Keywords: similarity, cover, degree of certainty, the similarity.

ВВЕДЕНИЕ

Сходство является ключевым понятием интеллектуальной обработки информации. Оно используется в алгоритмах классификации/кластеризации, в процедурах информационного поиска, при построении баз знаний на основе наблюдаемых данных. Во всех перечисленных случаях понятие сходства является основой структуризации заданного множества объектов.

Преимуществом понятия сходства является то, что оно может применяться к данным произвольной природы.

Цель статьи заключается в представлении методов построения отношения сходства как основы классификационных процедур.

1. ПОСТАНОВКА ЗАДАЧИ

Пусть $X = \{x_i\}_{i=\overline{1,N}}$ – заданное конечное множество объектов, $\sigma = \{\sigma_j\}_{j=\overline{1,M}}$ – набор свойств объектов из X, причем выполняется следующее предположение: каждый из объектов x_i не обязательно обладает всеми свойствами из σ, но во множестве X существует хотя бы один объект, обладающий данным свойством σ_j.

Требуется сгруппировать похожие в некотором смысле объекты.

Заметим, что приведенная постановка задачи предполагает построение разбиения множества X на классы. Особенностью задачи является то, что множество свойств не является фиксированным в том смысле, что группируемые объекты могут обладать различными наборами свойств, в общем случае пересекающихся.

2. ОБЩИЙ ПОДХОД К ФОРМИРОВАНИЮ МЕХАНИЗМОВ «ПОХОЖЕСТИ»

В рамках классического подхода всякое разбиение заданного множества X основывается на предположении, что на X определено некоторое отношение эквивалентности E, и задача заключается в том, чтобы выявить это отношение, тогда классы эквивалентности, индуцированные E, описывают группировки схожих объектов. Пусть $\aleph = \{X_1, ..., X_U\}$ – система подмножеств множества X, такая что $\bigcup_{i=1}^{U} X_i = X$ (1) и для любых индексов $i \neq j$ выполняется условие $X_i \cap X_j = \varnothing$ (2), тогда $\aleph$ называется *разбиением* множества X.

Если для системы $\aleph$ выполняется только условие (1), то она называется *покрытием* множества X.

Если $\aleph$ является разбиением множества X, то на нем можно определить отношение эквивалентности $E \subseteq X^2$ по правилу

$$E = \left\{ \left(x_i, x_j \right) / \; x_i \in X_i \; и \; x_j \in X_j \right\}, \qquad (3)$$

тогда фактор-множество множества X по отношению E совпадает с $\aleph$, т. е. $X/_E = \aleph$.

С другой стороны, если E – отношение эквивалентности на множестве X, то множество классов эквивалентности $\left\{ E\left(x_i\right) = \left\{ x_j : \left(x_i, x_j\right) \in E \right\} \right\}_{i=\overline{1,N}}$ образует разбиение множества X.

Заметим, что классификация является результатом установления отношения похожести, как меры близости, между элементами множества, при этом в рамках классического подхода не существует а) элементов, не охваченных классификацией и б) элементов, попавших одновременно в два класса (принцип непротиворечивости). Однако в реальных задачах могут встретиться элементы, для которых можно говорить лишь о степени принадлежности к некоторому классу.

Рассмотрим общий подход к формированию механизмов «похожести».

Формально отношение похожести двух объектов представляется бинарным отношением *сходства*, которое по определению является рефлексивным и симметричным. Это отношение может обладать или не обладать свойством транзитивности. Сходство является отношением эквивалентности, если оно транзитивно.

Введем множества

$$P_k = \left\{ x_i \in X : x_i \; обладает \; свойством \; \sigma_k \right\}$$

$k = \overline{1,M}$. Для любого k $P_k \neq \varnothing$, так как во множестве X существует хотя бы один элемент, обладающий свойством σ_k. Система множеств $\left\{P_k\right\}_{k=\overline{1,M}}$ образует покрытие множества X, т. е. $\bigcup_{k=1}^{M} P_k = X$, на основе которого определим отношение

$$R = \left\{ \left(x_i, x_j\right) : \exists k = \overline{1,M} \left(x_i \in P_k \; и \; x_j \in P_k \right) \right\} \subseteq X^2, \; (4)$$

содержащее такие пары объектов x_i и x_j, которые обладают хотя бы одним из свойств $\sigma_1, \ldots, \sigma_M$.

Заметим, что данное отношение рефлексивно и симметрично, но не обязательно транзитивно, что обусловлено тем, что объекты могут сравниваться по разным свойствам. Таким образом, отношение R является сходством.

Будем говорить, что объекты x_i и x_j сходны, если $\left(x_i, x_j\right) \in R$.

<u>Утверждение 1.</u> Если каждый объект $x_i \in X$ обладает только одним свойством, то отношение сходства является эквивалентностью.

В самом деле, в этом случае каждое из множеств P_i содержит единственный элемент, причем $P_i \cap P_j = \varnothing$ для любых индексов $i \neq j$. Следовательно, система множеств $\left\{P_k\right\}_{k=\overline{1,M}}$ образует разбиение множества X, причем каждый класс является тривиальным [], т.е. содержит единственный элемент из множества X.

Пусть x_i и x_j – объекты из множества X, $\nu^i = \left(\nu_1^i, \ldots, \nu_M^i\right)$, $\nu^j = \left(\nu_1^j, \ldots, \nu_M^j\right)$ – векторы, каждая компонента которого $\nu_k^i \left(\nu_k^j\right)$ есть *степень уверенности* в том, что объект $x_i \left(x_j\right)$ обладает свойством σ_k. Будем считать, что степень уверенности оценивается числом из [0,1], и ее значение, равное 0, соответствует тому, что объект не обладает данным свойством. Если значение степени уверенности равно 1, то объект в полной мере обладает данным свойством. Таким образом, все объекты имеют векторные оценки уверенности одинаковой размерности.

Поскольку во множестве X по предположению обязательно существует элемент, который наверняка обладает свойством σ_k, то $\max\limits_i \left\{\nu_k^i\right\} = 1$, а так как каждый объект обладает хотя бы одним из свойств, то $\max\limits_k \left\{\nu_k^i\right\} = 1$.

Величины $\nu_k^i \left(i = \overline{1,N}, k = \overline{1,M}\right)$ образуют матрицу степеней уверенности, ее строки соответствуют элементам x_i множества X, а столбцы – свойствам σ_k. В каждой строке и в каждом столбце матрицы имеется, по крайней мере, одна 1.

Нечеткое множество объектов, обладающих свойством σ_k, определим следующим образом:

$$\tilde{P}_k = \left\{ \left(x_i / \nu_k^i \right) \right\}_{i=\overline{1,N}}. \qquad (5)$$

Утверждение 2. Семейство нечетких множеств $\tilde{\Omega} = \left\{ \tilde{P}_1, ..., \tilde{P}_M \right\}$ образует покрытие множества X.

Доказательство тривиально: $\bigcup_{k=1}^{M} \tilde{P}_k = X$, так как для любого x_i имеет место $\max_k \left\{ \nu_k^i \right\} = 1$.

Обобщая отношение (4) на случай нечетких множеств $\tilde{P}_k$, получим нечеткое отношение с функцией принадлежности

$$\mu_{\tilde{R}}\left(x_i, x_j \right) = \max_k \min \left\{ \nu_k^i, \nu_k^j \right\}. \qquad (6)$$

Чтобы подчеркнуть связь с покрытием $\tilde{\Omega}$, используем обозначение $\tilde{R}_{\tilde{\Omega}}$.

Утверждение 3. Отношение $\tilde{R} = \tilde{R}_{\tilde{\Omega}}$ обладает следующими свойствами:

1) $\tilde{R}$ – симметричное и рефлексивное отношение;

2) $\mu_{\tilde{R}}\left(x_i, x_j \right) \le \max_k \left\{ \mu_{\tilde{R}}\left(x_i, x_k \right), \mu_{\tilde{R}}\left(x_k, x_j \right) \right\}$ для всех x_i, x_j.

Доказательство.

Из определения (6) следует, что $\tilde{R}$ симметричное отношение. Кроме того, $\mu_{\tilde{R}}\left(x_i, x_i \right) = \max_k \left\{ \nu_k^i \right\} = 1$, и, следовательно, $\tilde{R}$ рефлексивно. Таким образом, $\tilde{R}_i$ удовлетворяет определению нечеткого сходства. Покажем, что выполняется свойство 2).

$$\max_k \left\{ \mu_{\tilde{R}}\left(x_i, x_k \right), \mu_{\tilde{R}}\left(x_k, x_j \right) \right\} =$$

$$= \max_k \left\{ \max_r \left\{ \min\left(\nu_r^i, \nu_r^k \right), \min\left(\nu_r^j, \nu_r^k \right) \right\} \right\} =$$

$$= \max_k \left\{ \min \left\{ \max \left\{ \nu_k^i, \nu_k^j \right\}, \underbrace{\max_r \left\{ \nu_k^r \right\}}_{=1} \right\} \right\} =$$

$$= \max_k \max \left\{ \nu_k^i, \nu_k^j \right\} \ge$$

$$\ge \max_k \min \left\{ \nu_k^i, \nu_k^j \right\} = \mu_{\tilde{R}}\left(x_i, x_j \right).$$

Из утверждения 2 следует, что если задано семейство нечетких подмножеств X, образующих покрытие, то на X можно определить отношение, являющееся нечетким сходством.

Верно обратное

Утверждение 4 [1]. Пусть $\tilde{R}$ – нечеткое отношение сходства на множестве X с функцией принадлежности $\mu_{\tilde{R}}(x)$, тогда существует покрытие $\tilde{\Omega}$, такое что $\tilde{R} = \tilde{R}_{\tilde{\Omega}}$.

Вполне возможно, что для различных покрытий $\tilde{\Omega}_1$ и $\tilde{\Omega}_2$ окажется, что $\tilde{R}_{\tilde{\Omega}_1} = \tilde{R}_{\tilde{\Omega}_2}$.

Каждому элементу $x_i \in X$ соответствует нечеткое множество свойств, которыми элемент обладает в различной степени, т. е.

$$\tilde{S}_i = \left\{ \left(\sigma_k / \nu_k^i \right) \right\}_{k=\overline{1,M}}.$$

Множества $\left\{ \tilde{S}_1, ..., \tilde{S}_N \right\}$ образуют покрытие множества свойств σ.

Любое покрытие можно интерпретировать как отображение. В самом деле, положим

$$F(x,i) = \begin{cases} 1, & \text{если объект } x \\ & \text{обладает свойством } i, \\ 0, & \text{иначе.} \end{cases} \qquad (7)$$

которое можно рассматривать как соответствие между множеством объектов X и множеством свойств σ. Заметим, что в четком случае отношение сходства R_F на X, опирающееся на соответствие (7), представляет собой ядро F, т. е. $R_F = F^{-1} \circ F$. Последнюю формулу можно обобщить на нечеткий случай с помощью композиционного правила

$$R_F(x,y) = \left(F^{-1} \circ F \right)(x,y) =$$

$$\max_i \min \left\{ F(x,i), F(i,y) \right\} =$$

$$= \max_i \min \left\{ \tilde{S}_i(x), \tilde{S}_i(y) \right\} = \tilde{R}_{\Omega}(x,y). \qquad (8)$$

С другой стороны, для каждого нечеткого отображения $F: X \to \sigma$ можно рассмотреть покрытие $\Omega_F = \left\{ F(x,i) \right\}_{i \in \sigma}$ нечеткого множества $U(x) = \max_i F(x,i)$.

Таким образом, покрытия и нечеткие отображения дают эквивалентное описание структур схожести.

3. ОТНОШЕНИЕ ПОДОБИЯ – ОСНОВА ПРОЦЕДУР КЛАСТЕРИЗАЦИИ

Отношение подобия – важный частный случай нечеткого отношения сходства. Отношение $S \subseteq X^2$ называется отношением подобия, если оно рефлексивно $\left(\forall x \left(S(x,x) = 1 \right) \right)$, симметрично $\left(\forall x, y \left(S(x,y) = S(y,x) \right) \right)$ и транзитивно, причем транзитивность можно определить различными способами [2, 3].

В процедурах нечеткой классификации [3] исходная информации о близости объектов рассматриваемого множества задается в виде отношения сходства S, которое является рефлексивным и симметричным, но не обязательно транзитивным. Для перехода к отношению подобия P применяется процедура транзитивного замыкания с использованием подходящей композиции, которая сохраняет симметричность и рефлексивность, но добавляет транзитивность.

В соответствии с теоремой о декомпозиции [1] любое нечеткое отношение можно представить через систему обычных отношение $S_\alpha = \{(x,y): S(x,y) \geq \alpha\}$ для $\alpha \in (0,1]$. Заметим, что если $\alpha_1 < \alpha_2$, то $S_{\alpha_2} \subset S_{\alpha_1}$.

Из рефлексивности и симметричности S следует рефлексивность и симметричность S_α для любого для $\alpha \in (0,1]$. Кроме того, если S транзитивно, то S_α также транзитивно, а, следовательно, для любого для $\alpha \in (0,1]$ отношение S_α является обычным отношением эквивалентности, которое индуцирует на X некоторое разбиение на классы эквивалентности.

Подмножество $A \subset X$ будем называть *классом подобия* отношения подобия S, если $S(x,y) > S(x,z)$ для всех $x, y \in A$ и $z \notin A$. Множество классов подобия отношения S совпадает с множеством классов эквивалентности отношений S_α [4].

Два объекта x и y будем считать *равносильными* в смысле отношения S, если $S(x,y) = 1$ и $S(x,z) = S(y,z)$ для всех объектов $z \in X \setminus \{x,y\}$.

Два объекта x и y будем называть *неразличимыми на уровне* α, если $S(x,y) \geq \alpha$ и для всех $z \in X$ выполняется $S(x,z) \geq \alpha$ тогда и только тогда, когда $S(y,z) \geq \alpha$.

Ясно, что два объекта, неразличимые на некотором уровне α, будут равносильными в смысле отношения S_α.

Отношение сходства S, определенное на X, будет отношением подобия тогда и только тогда, когда все объекты X неразличимы в S.

ЗАКЛЮЧЕНИЕ

Нечеткое отношение сходства лежит в основе представления естественных кластеров в методах классификации/кластеризации. Переход к нечетким множествам, функции принадлежности которых отражают степени уверенности в наличии у объектов свойств, позволяют решить проблему несравнимости объектов по различным наборам свойств. Предложенный подход может быть использован для решения задач медицинской диагностики, когда различные симптомы проявляются по-разному для тех или иных заболеваний.

СПИСОК ЛИТЕРАТУРЫ

1. *Леденева Т. М.* Обработка нечеткой информации / Т.М. Леденева. – Воронеж : ИПЦ ВГУ, 2006. – 232 с.

2. *Леденева Т. М.* Транзитивность как особое свойство нечетких отношений / Т. М. Леденева, Н. А. Каплиева. – Воронеж : ВГУ, 2006. – 51 с. – Деп. в ВИНИТИ 07.12.2006, № 1523-В2006.

3. *Каплиева Н. А.* Исследование различных типов транзитивности в приложении к нечеткой классификации / Н. А. Каплиева, Т. М. Леденева // Вестник Воронежского государственного университета. Сер. Физика. Математика. – 2006. – № 2. – С. 206–216.

4. *Леденева Т. М.* О влиянии функции подобия на результаты нечеткой классификации / Т. М. Леденева, Н. Х. Нгуен // Информационные технологии. – М. : Новые технологии, 2011. – № 11. С. 15–23.

Леденева Татьяна Михайловна – доктор технических наук, профессор, заведующий кафедрой вычислительной математики и прикладных информационных технологий, Воронежский государственный университет, e-mail ledeneva-tm@yandex.ru

Стрюков Руслан Константинович – аспирант кафедры вычислительной математики и прикладных информационных технологий, Воронежский государственный университет, e-mail 79204605031@ya.ru

Ledeneva Tatiana M. – Doctor of Technic Science, Professor, Head of the Department of Computational Mathematics and Applied Information Technologies, Voronezh State University, e-mail ledeneva-tm@yandex.ru

Stryukov Ruslan K. – Post-graduate student of the Department of Computational Mathematics and Applied Information Technology, Voronezh State University, e-mail 79204605031@ya.ru