Лабораторная работа №3. Кластерный анализ систем.

Цель работы: Ознакомиться с применением теории нечетких отношений к кластерному анализу систем.

Задание к работе: ***Вариант 1.*** Построить матрицу парных сравнений (нечеткое отношение) по результатам операции нечёткого сравнения строк, задаваемых пользователем. Построить транзитивное замыкание нечеткого отношения, определить класс нечеткого отношения и сделать вывод о «схожести» входных строк.

Вариант 2. Построить матрицу парных сравнений (нечеткое отношение) по результатам операции нечёткого сравнения монохромных изображений, задаваемых пользователем. Построить транзитивное замыкание нечеткого отношения, определить класс нечеткого отношения и сделать вывод о «схожести» входных изображений.

Теоретическая часть

Приложения теории нечетких отношений к кластерному анализу систем

В кластерном анализе (автоматической классификации) предложена процедура кластеризации, основанная на транзитивном замыкании исходного нечеткого отношения, получаемого в результате опроса экспертов.

Например, эксперты в некоторой шкале сравнений указывают силу сходства между портретами людей, принадлежащих к нескольким семьям (универсуме *U,* мощность $|U| = N$), и на основе по-парного сравнения всех портретов строится отношение, в виде матрицы сходства $M = \|m_{i,j}\|, i, j = \overline{1, N}, N = |U|$, на множестве являющимся декартовым произведением *U*×*U*. Если наложить ограничение на элементы матрицы сходства, такое что $0 \le m_{i,j} \le 1, i, j = \overline{1..N}$, получим бинарное нечеткое отношение первого типа. Далее ищется транзитивное замыкание этого НО и выбирается порог (уровень) α – таким образом, чтобы число классов разбиения, получаемое из α-уровней, равнялось числу семей *N*.

Процедура классификации относила портреты, попавшие в один класс разбиения, к одной семье. В проведенных экспериментах результаты классификации дали хорошее согласование с истинным разбиением портретов по семьям.

Нечеткие отношения

Определение *Нечетким отношением* $\tilde{R}$ на множествах $X_1, X_2, ..., X_n$ называется нечеткое подмножество декартова произведения $X_1 \times X_2 \times ... \times X_n$. Степень принадлежности $\mu_{\tilde{R}}(X_1, X_2, ..., X_n)$ показывает степень выполнения отношения $\tilde{R}$ между элементами $(x_1, x_2, ..., x_n)$, $x_i \in X_i$, $i = \overline{1..n}$. Если значение $0 \le \mu_{\tilde{R}}(X_1, X_2, ..., X_n) \le 1$, то такое НО называется НО отношением типа 1.

Рассмотрим *бинарные нечеткие отношения*, которые задаются на декартовом произведении двух множеств. Обозначим эти множества через *X* и *Y*. Тогда задание бинарного нечеткого отношения $\tilde{R}$ на $X \times Y$ состоит в указании всех троек $\left(x, y, \mu_{\tilde{R}}(x, y)\right)$, где $x \in X$, $y \in Y$, или, что тоже самое, $(x, y) \in \left(X \times Y\right)$.

Пример 1. Задать нечеткое отношение *x≈y* ("*x* приблизительно равно *y*").

Пусть $(x,y)\in\{1,2,3\}$. Тогда нечеткое отношение удобно задавать матрицей вида:

$$\begin{array}{cc} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} & \leftarrow y \quad x\downarrow \\ \tilde{R}=\begin{bmatrix} 1 & 0.5 & 0.2 & 0.1 \\ 0.5 & 1 & 0.6 & 0.3 \\ 0.2 & 0.6 & 1 & 0.8 \\ 0.1 & 0.3 & 0.8 & 1 \end{bmatrix} & \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} \end{array}.$$

Для непрерывных множеств *X=[0,3]* и *Y=[0,3]* нечеткое отношение можно задать следующей функцией принадлежности: $\mu_{\tilde{R}}(x,y)=e^{-0.2(x-y)^2}$.

Пример 2. Задать нечеткое отношение "*x* намного меньше, чем *y*".

Пусть $(x,y)\in\{1,2,3\}$. Тогда нечеткое отношение можно задать матрицей вида:

$$\begin{array}{cc} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} & \leftarrow y \quad x\downarrow \\ \tilde{R}=\begin{bmatrix} 0 & 0.2 & 0.6 & 1 \\ 0 & 0 & 0.2 & 0.6 \\ 0 & 0 & 0 & 0.2 \\ 0 & 0 & 0 & 0 \end{bmatrix} & \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} \end{array}.$$

Пример 3. Задать отношение "схожий менталитет" для следующих национальностей {Украинцы(У), Чехи (Ч), Австрийцы (А), Немцы (Н)}.

Использование обычного, не нечеткого отношения позволяет выделить только одну пару наций со схожими менталитетами – немцев и австрийцев. Этим отношением не отражается тот факт, что по менталитету чехи более близки к немцам, чем к украинцам. Нечеткое отношение позволяет легко представить такую информацию:

$$\begin{array}{cc} \begin{array}{cccc} \text{У} & \text{Ч} & \text{А} & \text{Н} \end{array} & \\ \tilde{R}=\begin{bmatrix} 1 & 0.4 & 0.2 & 0.1 \\ 0.4 & 1 & 0.4 & 0.3 \\ 0.2 & 0.4 & 1 & 0.8 \\ 0.1 & 0.3 & 0.8 & 1 \end{bmatrix} & \begin{array}{c} \text{У} \\ \text{Ч} \\ \text{А} \\ \text{Н} \end{array} \end{array}.$$

Операции над нечеткими отношениями

Перейдем теперь к рассмотрению операций над нечеткими отношениями. Некоторые из этих операций являются аналогами соответствующих операций для обычных отношений, однако, как и в случае нечетких множеств, существуют операции, характерные лишь для нечетких отношений. Заметим, что так же, как и в случае нечетких множеств, операции объединения и пересечения нечетких отношений (и операцию произведения) можно определить различными способами.

Пусть на множестве X заданы два нечетких отношения A и B, т, е. в декартовом произведении X×X заданы два нечетких множества A и B, Нечеткие множества C=A $\bigcup$ B и D=A $\bigcap$ B называются соответственно объединением и пересечением нечетких отношений A и B на множестве X. Если воспользоваться определением объединения и пересечения нечетких множеств, то для функций принадлежности отношений C и D получаем

$$\mu_C(x,y)=\max\{\mu_A(x,y),\mu_B(x,y)\},$$

$$\mu_D(x,y)=\min\{\mu_A(x,y),\mu_B(x,y)\}.$$

Говорят, что нечеткое отношение B включает в себя нечеткое отношение A, если для нечетких множеств A и B выполнено A $\subseteq$ B. Для функций принадлежности этих множеств неравенство

$\mu_A(x,y) \le \mu_B(x,y)$ выполняется при любых x, y ∈ X. В рассмотренном выше примере отношений (≥) и (>>) нечеткое отношение R содержится в отношении R, т. е. должно быть $\mu_R(x,y) \le \mu_R(x,y)$ для любых чисел x, y из интервала [0, 1].

Композиция нечетких отношений

Операция композиции нечетких отношений R_1 в X×Y и R_2 в Y×Z позволяет определить нечеткое отношение в X×Z.

Пусть R_1 есть нечеткое отношение в X×Y, R_2 - нечеткое отношение в Y×Z. (Max-min) – композиция $R_1 \circ R_2$ определяется выражением: $\mu_{R_1 \circ R_2}(x,z) = \max\limits_y \left[\min\left\{ \mu_{R_1}(x,y), \mu_{R_2}(y,z) \right\} \right]$, где: $x \in X, y \in Y, z \in Z$.

Вычисление композиции нечетких отношений аналогично вычислению произведения матриц, ("столбец на строку"), только вместо произведения и суммы выполняются операции взятия минимума и максимума соответственно.

Пример 4: Пусть X = {x_1, x_2, x_3}, Y = {y_1,y_2,y_3,y_4,y_5}, Z = {z_1,z_2,z_3,z_4}. Отношение R_1 на X×Y задано следующей матрицей:

R_1	y_1	y_2	y_3	y_4	y_5
x_1	0.1	0.2	0.5	0.7	0.3
x_2	0.3	0	1	0.3	0.7
x_3	0.1	0.8	0	0	1

Нечеткое отношение R_2 на Y×Z задано следующей матрицей:

R_1	z_1	yz_2	z_3	z_4
y_1	0.9	0	0.4	0.7
y_2	0.3	0.4	0.9	1
y_3	1	0.4	0.2	0
y_4	0.3	0.4	0.7	0.2
y_5	0.5	0.5	0.9	0.1

Тогда нечеткое отношение R_1° R_2 определено на X ×Z и выражается следующей матрицей:

R_1	z_1	z_2	z_3	z_4
x_1	0.5	0.4	0.7	0.2
x_2	1	0.5	0.7	0.8
x_3	0.5	0.5	0.9	0.8

Свойства нечетких отношений

- *Рефлексивность.* Нечеткое отношение R на множестве X называется рефлексивным, если для любого x∈X выполнено равенство $\mu_R(x,x) = 1$ ($R(x,x) = 1, \forall x \in X$). В случае конечного множества X главная диагональ матрицы рефлексивного нечеткого отношения R состоит целиком из единиц. Примером рефлексивного нечеткого отношения может служить отношение «примерно равны» в множестве чисел.
- *Антирефлексивность*: $R(x,x) = 0,\ \forall x \in X$. Функция принадлежности антирефлексивного нечеткого отношения обладает свойством $\mu_R(x,x) = 0$ при любом x ∈ X. Антирефлексивно, например, отношение «много больше» в множестве чисел. Ясно, что дополнение рефлексивного отношения антирефлексивно.

- *Симметричность.* Нечеткое отношение R на множестве X называется симметричным, если для любых x, y$\in$X выполнено равенство: $\mu_R(x,y)=\mu_R(y,x)$. Матрица симметричного нечеткого отношения, заданного в конечном множестве, симметричная. Пример симметричного нечеткого отношения – отношение «сильно различаться по величине».
- *Антисимметричность.* Нечеткое бинарное отношение называется антисимметричным, если $\forall(x,y)\in U\times U: x\neq y \Rightarrow \mu_R(x,y)\neq\mu_R(y,x)$ или $\mu_R(x,y)=\mu_R(y,x)=0$.
- *Транзитивность.* Нечеткое отношение R на множестве X называется транзитивным, если $R\circ R\subseteq R$, $R(x,z)\geq R(x,y)\wedge R(y,z),\ \forall x,y,z\in X$.

Из этого определения видно, что свойство транзитивности нечеткого отношения зависит от способа определения композиции нечетких отношений, под которой будем понимать ***max-min*** композицию.

Транзитивное замыкание нечеткого бинарного отношения

Пусть R - нечеткое отношение в U×U. Определим $R^2=R\circ R$ с функцией принадлежности

$$\mu_{R^2}(x,z)=\max_y\left[\min\left\{\mu_R(x,y),\mu_R(y,z)\right\}\right],$$

где $x, y, z \in U$.

Сравнивая последнее выражение с определением транзитивности для нечетких бинарных отношений, не трудно увидеть, что свойство транзитивности можно записать в следующем виде:

$$R^2\subset R.$$

Аналогично можно определить по индукции R^n:

$$R^n=R^{n-1}\circ R,\ n=2,3,\ldots$$

Пусть $R^2\subset R$ и $R^{n+1}\subset R^n$, тогда $R^n\subset R$.

Транзитивное замыкание $\tilde{R}$ нечеткого бинарного отношения R определяется как:

$$\tilde{R}=R^1\cup R^2\cup\ldots\cup R^n\ldots$$

где R^n определяется рекурсивно: $R^1=R,\ R^n=R^{n-1}\circ R,\ n=2,3,\ldots$

Как проверить, построили мы или нет транзитивное замыкание конкретного нечеткого бинарного отношения? Ответ на этот вопрос дает следующее утверждение: Пусть R - некоторое нечеткое бинарное отношение. Если существует k такое, что $R^{k+1}=R^k$, то $\tilde{R}=R\cup R^2\cup\ldots\cup R^k$. Для конечных НО, заданных на счетных множествах число k равняется рангу матрицы НО (k = rang R).

Вычисление транзитивного замыкания нечеткого отношения – довольно утомительное дело, даже если множество состоит из небольшого числа элементов. $\tilde{R}=R^1\cup R^2\cup\ldots\cup R^n\ldots$

Декомпозиция нечетких отношений

Одно из важнейших свойств НО заключается в том, что они могут быть представлены в виде совокупности обычных отношений, причем эти отношения могут быть упорядочены по включению, представляя собой иерархическую совокупность отношений. Разложение НО на совокупность обыкновенных отношений основано на понятии α-уровня нечеткого отношения.

Множества уровня определяются так:

$$R_\alpha=\left(\left(x,y\right)\mid(x,y)\in X\times X,\mu_R(x,y)\geq\alpha\right)$$

Если обычное отношение R_α подобно НО отождествлять с его характеристической функцией R_α: $X\times Y\rightarrow[0,1]$, то соотношение α-уровня можно переписать в виде:

$$R_\alpha=\begin{cases}1, \text{ если } R(x,y)\geq\alpha,\\ 0, \text{ в противном случае.}\end{cases}$$

Нетрудно заметить, что α-уровни нечеткого отношения удовлетворяют соотношению:

$$\alpha \le \beta \Rightarrow R_\alpha \supseteq R_\beta\,,$$

представляя собой совокупность вложенных друг в друга отношений.

Теорема о декомпозиции: Любое отношение R можно представить в форме: $R = \max\limits_{\alpha}[\alpha \times R_\alpha]$ $0 < \alpha \le 1$.

Некоторые специальные типы нечетких отношений, отношение подобия

Рассмотрим отношение подобия и связанные с ним отношения различия, сходства и их свойства. Эти отношения интересны для нас тем, что они имеют интересные приложения в задачах обработки информации, демонстрирующие новые возможности такой обработки, предоставляемые введением и учетом нечеткости.

- Нечетким отношением подобия называется транзитивное рефлексивное симметричное нечеткое бинарное отношение. Очевидно, что отношение подобия является предпорядком.

Пример 5: Для любого а: 0<a<1 нечеткое бинарное отношение R является отношением подобия.

1	a	a	a	a
a	1	a	a	a
a	a	1	a	a
a	a	a	1	a
a	a	a	a	1

- Нечеткое бинарное отношение, обладающее свойствами антирефлексивности, симметричности и (min-max) - транзитивности называется отношением ***различия***.

Алгоритм построения матрицы парных сравнений строк

Данная операция используется для поиска «похожих до определённой степени» строк. Скажем, если в большом списке поставщиков есть организация «Копыта и Рога, ООО», а вы пытаетесь найти в нём по памяти «ЗАО “Рога и Копыта”», то при использовании традиционных методов поиска вас, скорее всего, ждёт неудача. Нечёткое сравнение позволяет без труда находить такие совпадения.

Алгоритм построения матрицы парных сравнений строк использует в качестве аргументов две строки и параметр сравнения – максимальную длину сравниваемых подстрок (кластеров). Результатом работы для каждой подстроки является число, лежащее в пределах от 0 до 1. 0 соответствует полному несовпадению двух строк, а 1 – полной (в определённом ниже смысле) их идентичности.

Алгоритм сравнения составляет всевозможные комбинации подстрок с длинной, вплоть до максимально указанной, и подсчитывает их совпадения в двух сравниваемых строках. Количество совпадений, разделённое на число вариантов, объявляется коэффициентом схожести строк и заносится в матрицу парных сравнений.

Поясним работу алгоритма на примере, где в качестве аргументов заданы две строки “test” и “text” и максимальная длина сравниваемых подстрок равной 4 (т.е. максимально сравниваются слова целиком). Рассмотрим, каким образом происходит сравнение строк и заполнение матрицы парных сравнений.

Таблица № 1.
Пояснения к алгоритму построения матрицы парных сравнений строк.

Сравниваемая подстрока	Подстроки второй строки	Есть совпадение?	Количество совпадений	Количество вариантов	Результат
Сравниваем строку *test* со строкой *text* по подстрокам длины 1.					
T	t, e, x, t	да	3	4	3 / 4
E	t, e, x, t	да			
S	t, e, x, t	нет			
T	t, e, x, t	да			
Сравниваем строку *text* со строкой *test* по подстрокам длины 1.					
T	t, e, s, t	да	3	4	3 / 4
E	t, e, s, t	да			
X	t, e, s, t	нет			
T	t, e, s, t	да			
Сравниваем строку *test* со строкой *text* по подстрокам длины 2.					
Te	te, ex, xt	да	1	3	1 / 3
Es	te, ex, xt	нет			
St	te, ex, xt	нет			
Сравниваем строку *text* со строкой *test* по подстрокам длины 2.					
Te	te, es, st	да	1	3	1 / 3
Ex	te, es, st	нет			
Xt	te, es, st	нет			
Сравниваем строку *test* со строкой *text* по подстрокам длины 3.					
Tes	tex, ext	нет	0	2	0 / 2
Est	tex, ext	нет			
Сравниваем строку *text* со строкой *test* по подстрокам длины 3.					
Tex	tes, est	нет	0	2	0 / 2
Ext	tes, est	нет			
Сравниваем строку *test* со строкой *text* по подстрокам длины 4.					
Test	Text	нет	0	1	0 / 1
Сравниваем строку *text* со строкой *test* по подстрокам длины 4.					
Text	Test	нет	0	1	0 / 1

Приведем пример полностью заполненной матрицы парных сравнений для строк «test» и «text»:

Таблица № 2.

Пояснения к алгоритму построения матрицы парных сравнений строк.

<table>
<tr><td colspan="2" rowspan="2">text
test</td><td>Подстроки длины 1</td><td>Подстроки длины 2</td><td>Подстроки длины 3</td><td>Подстроки длины 4</td></tr>
<tr><td>t
e
x
t</td><td>te
ex
xt</td><td>tex
ext</td><td>Text</td></tr>
<tr><td>Подстроки длины 1</td><td>t
e
s
t</td><td>5 / 16</td><td>6 / 12</td><td>6 / 8</td><td>3 / 4</td></tr>
<tr><td>Подстроки длины 2</td><td>te
es
st</td><td>6 / 12</td><td>1 / 9</td><td>1 / 6</td><td>1 / 3</td></tr>
<tr><td>Подстроки длины 3</td><td>tes
est</td><td>6 / 8</td><td>1 / 6</td><td>0 / 4</td><td>0 / 2</td></tr>
<tr><td>Подстроки длины 4</td><td>test</td><td>3 / 4</td><td>1 / 3</td><td>0 / 2</td><td>0 / 1</td></tr>
</table>

Покажем, каким образом вычисляется элемент таблицы с номером (3,2) т.е. когда подстроки длины три строки «test» сравниваются с подстроками длины два строки «text»:

Таблица № 3.

Пояснения к алгоритму построения матрицы парных сравнений строк.

<table>
<tr><td>Подстроки длины три строки «test»</td><td>Подстроки длины два строки «text»</td><td colspan="2">Совпадения</td><td>Результат</td></tr>
<tr><td rowspan="3">tes</td><td>te</td><td>Да</td><td rowspan="3">1 / 3</td><td rowspan="6">1 / 6</td></tr>
<tr><td>ex</td><td>Нет</td></tr>
<tr><td>xt</td><td>Нет</td></tr>
<tr><td rowspan="3">est</td><td>te</td><td>Нет</td><td rowspan="3">0 / 3</td></tr>
<tr><td>ex</td><td>Нет</td></tr>
<tr><td>xt</td><td>Нет</td></tr>
</table>

Алгоритм построения матрицы парных сравнений изображений

Алгоритм построения матрицы парных сравнений изображений использует в качестве аргументов два монохромных изображения и параметр сравнения – количество элементов (кластеров) разбиения изображения в строках и столбцах (обозначим его через *M*). Результатом работы, для каждых двух элементов различных изображений находящихся в одной строке и столбце разбиения, является число, лежащее в пределах от 0 до 1. 0 соответствует полному несовпадению элементов изображения, а 1 – полной (в определённом ниже смысле) их идентичности. После выполнения попарных сравнений всех элементов изображений получаем матрицу парных сравнений.

Значения заносимые в матрицу парных сравнений, вычисляются по правилу алгебраической суммы, отношений количества значащих пикселей к общему количеству пикселей в кластере:

$$m_{i,j} = \frac{N_{1\ \ i,j}^{Black}}{N_{1\ \ i,j}^{All}} + \frac{N_{2\ \ i,j}^{Black}}{N_{2\ \ i,j}^{All}} - \frac{N_{1\ \ i,j}^{Black}}{N_{1\ \ i,j}^{All}} \times \frac{N_{2\ \ i,j}^{Black}}{N_{2\ \ i,j}^{All}}, \qquad i,\, j = \overline{1..M}\,.$$

Где: $N_{1\ \ i,j}^{Black}$, $N_{2\ \ i,j}^{Black}$ - число значащих (для черно-белого изображения это черный цвет) пикселей в *i,j-ом* кластере для 1-го и 2-го изображений соответственно;
$N_{1\ \ i,j}^{All}$, $N_{2\ \ i,j}^{All}$ - общее число пикселей в *i, j-ом* кластере для 1-го и 2-го изображений соответственно.

Получение вывода об идентичности строк и изображений

После заполнения матрицы парных сравнений, в соответствии с вышеуказанными алгоритмами, которые в свою очередь являются нечеткими отношениями на множестве являющимся декартовым произведением разбитых на кластеры множеств входных данных (строк или изображений), требуется найти транзитивное замыкание этого нечеткого отношения, по ***max-min*** композиции. На основании полученного транзитивного замыкания сделать вывод об идентичности входных данных (строк или изображений).

Литература

1. Заде Л. Понятие лингвистической переменной и его применение для принятия приближенных решений. М.: Мир, 1976. -165с.
2. Алтунин А.Е., Семухин М.В. Модели и алгоритмы принятия решений в нечетких условиях: Монография. Тюмень: Издательство Тюменского государственного университета, 2000. 352 с.