一种基于 Vague 关系的模糊聚类方法*)

赵法信[1,2] 马宗民[1]

(东北大学信息科学与工程学院 沈阳 110004)[1] (通化师范学院 通化 134002)[2]

摘 要 Vague 关系作为模糊关系的一种推广，在某些情况下，比直觉模糊关系具有更强的模糊信息表达能力。本文基于 max-t & min-s 合成运算，将 Yang 和 Shih 的用于模糊关系的 n 步过程扩展到 Vague 关系上，扩展后的 n 步过程可以将相似 Vague 关系矩阵转换为等价 Vague 关系矩阵，进而提出了一种适用于所有 max-t & min-s 等价 Vague 关系矩阵的聚类算法。

关键词 模糊关系，vague 关系，max-t & min-s 合成，n 步过程，聚类算法

Fuzzy Cluster Analysis for Vague Relations

ZHAO Fa-Xin[1,2] MA Zong-Min[1]

(School of Information Science and Engineering, Northeastern University, Shenyang 110004)[1]

(Tonghua Normal University, Tonghua 134002)[2]

Abstract Vague relation, as a generalized fuzzy relation, has more powerful ability to process fuzzy information than intuitionistic fuzzy relation in some situations. On the basis of max-t & min-s compositions, Yang and Shih's fuzzy relation based n-step procedure is further extended. With the proposed n-step procedure, a similarity vague relation matrix is obtained by beginning with a proximity vague relation matrix. Then a clustering algorithm is proposed for all max-t & min-s similarity vague relation matrix.

Keywords Fuzzy relation, Vague relation, max-t & min-s composition, n-step procedure, Clustering algorithm

1 引言

现实世界中存在许多具有模糊性的不确定性信息和数据，而这些信息和数据是 Contor 创立的集合论所无法处理的。为此，Zadeh 于 1965 年提出了模糊集理论[1]。在模糊集理论中，每个属于论域 U 的对象 u 被赋予一个[0, 1]中的隶属度。但该单值隶属度既包含了支持 $u\in U$ 的证据，也包含了反对 $u\in U$ 的证据，它不可能同时表示支持和反对的证据。为了解决此类问题，Gau 等人提出了 Vague 集理论[2]。在一个 Vague 集 V 中，用一个真隶属函数 tr_V 和假隶属函数 f_V 来描述其隶属度的界，从而构成一个隶属度区间 $[tr_V(u),\ 1-f_V(u)]$，其中 $0\leqslant tr_V(u)\leqslant 1-f_V(u)\leqslant 1$。作为模糊集的一般化，Vague 集并不等同于直觉模糊集（IFS）。而且，在处理模糊数据时，Vague 集还具有一些独有的特点[3]。

聚类分析是数据挖掘中一个重要的分析方法。基于模糊集理论，模糊聚类分析已经在一些不同的领域得以广泛的研究[4,5]。模糊聚类分析大体上可以分为两大类，即基于模糊关系的聚类分析和基于目标函数的聚类分析。对于第一种聚类方法，Tamura 等在文[6]中基于 max-min 合成运算构造了一个 n 步过程，并利用该过程对相似模糊关系进行聚类分析，但该过程不能很好地挖掘数据之间的关系。为此，Yang 等在文[7]中将 Tamura 等提出的 n 步过程所基于的 max-min 合成运算扩展至所有的 max-t 合成运算，并基于所得到的 max-t 等价关系提出了一个聚类算法。

Vague 关系和直觉模糊关系都是模糊关系的一般化。De 等已经将直觉模糊关系应用到医疗诊断上[8]。他们断言，作为模糊集的补充，非隶属度（假隶属度）函数与隶属度函数相比具有更加重要的作用，这一点在决策支持和医疗诊断领域的应用上尤为显著。从文[3]中可以看出，相对于其它模糊关系，Vague 关系在诸如分析数据关系和相似度等方面的应用更具合理性。因此，本文主要研究了基于 Vague 关系的模糊聚类；基于 max-t & min-s 复合运算，对 Yang 等[7]的 n 步过程进行了进一步扩展，扩展后的 n 步过程可以将相似 Vague 关系改造为等价 Vague 关系；进而提出了一种适用于所有 max-t & min-s 等价 Vague 关系矩阵的聚类算法。

2 基本知识

定义 1(Vague 集) 设 U 为论域，$\forall u\in U$。U 中的一个 Vague 集 V 用一个真隶属函数 tr_V 和一个假隶属函数 f_V 表示，$tr_V(u)$是从支持 u 的证据所导出的 u 的隶属度下界，$f_V(u)$则是从反对 u 的证据所导出的 u 的否定隶属度下界，$tr_V(u)$和 $f_V(u)$都将区间[0,1]中的一个实数与 U 中的每一个元素联系起来，即

$$tr_V: U\rightarrow[0,\ 1]\quad f_V: U\rightarrow[0,\ 1]$$

其中，$tr_V(u)+f_V(u)\leqslant 1$。

假设 $U=\{u_1, u_2, \cdots, u_n\}$，那么论域 U 中的一个 Vague 集 V 可以表示如下：

$$V=\sum_{i=1}^{n}[tr_v(u_i), 1-f_v(u_i)]/u_i,\ \forall u_i\in U$$

其中 $tr_V(u)\leqslant \mu_V(u)\leqslant 1-f_V(u)$ 且 $1\leqslant i\leqslant n$

这里 u 的隶属度$[tr_V(u),\ 1-f_V(u)]$是区间[0, 1]的一个子区间。也就是说，尽管可能并不知道 u 的精确隶属度，但它落在某个子区间却是确定的。另外，我们还可以用差(1－

*)教育部博士点基金资助项目(2005145024)；教育部新世纪优秀人才支持计划(NCET-05-0288)。**赵法信** 博士研究生；**马宗民** 教授，博士生导师。

$tr_v(u)-fV(u)$)来描述对 u 了解的精度。差越小,对 u 的了解就越精确,反之,对 u 的了解就越少。如果 $tr_V(u)$ 等于 $1-f_V(u)$,说明我们对 u 的了解是较精确的,此时 Vague 集还原至模糊集;如果 $tr_V(u)$ 和 $1-f_V(u)$ 同时为 0(u 不属于集合 V)或 1(u 属于集合 V),就说明我们对 u 的了解最精确,此时 Vague 集还原至经典集合。

设 $VS_S(X)$ 表示论域 X 中所有 Vague 集的集合。那么,对于 $\forall A,B\in VS_S(X)$,有

$A\leqslant B$ 当且仅当 $tr_A(x)\leqslant tr_B(x)$ 和 $1-f_A(x)\leqslant 1-f_B(x)$ 同时成立,$\forall x\in X$。

定义 2(Vague 关系) 设 X 和 Y 是两个论域,称笛卡尔积 $X\times Y$ 上的 Vague 子集为从 X 到 Y 之间的二元 Vague 关系,表示为

$$R=\{\langle(x,y),tr_R(x,y),1-f_R(x,y)\rangle\mid x\in X,y\in Y\}$$

其中,$tr_R: X\times Y\to[0,1]$,$f_R: X\times Y\to[0,1]$ 且满足 $0\leqslant t_R(x,y)+f_R(x,y)\leqslant 1$,$\forall(x,y)\in X\times Y$。

在这里,我们将 $X\times Y$ 上所有 Vague 关系的全体表示为 $VSR(X\times Y)$。

定义 3(max-t & min-s 合成) 给定一个 t-范数、一个 s-范数和一个初始的 Vague 关系矩阵 $R^{(0)}=[(tr_{ij}^{(0)},1-f_{ij}^{(0)})]$,则 max-$t$ & min-s 合成定义为

$$R^{(n)}=[(tr_{ij}^{(n)},1-f_{ij}^{(n)})]$$

其中,$tr_{ij}^{(n)}=\max_k\{t(tr_{ik}^{(n-1)},tr_{kj}^{(n-1)})\}$,$f_{ij}^{(n)}=\min_k\{s(f_{ik}^{(n-1)},f_{kj}^{(n-1)})\}$,$n=1,2,3,\cdots$

定义 4(Vague 相似关系) 如果论域 X 中的 Vague 关系 R 满足

1)R 是自反的,即 $tr_R(x,x)=1$ 且 $f_R(x,x)=0\ \forall x\in X$

2)R 是对称的,即 $tr_R(x,y)=tr_R(y,x)$ 且 $f_R(x,y)=f_R(y,x)\ \forall x,y\in X$

则称 R 为 Vague 相似关系。

定义 5(Vague 等价关系) 如果论域 X 中的 Vague 相似关系 R 满足传递性(max-t & min-s 传递)$R\geqslant R\circ R$ 即 $t_R(x,z)\geqslant\bigvee_{y\in X}\{t[tr_R(x,y),tr_R(y,z)]\}$ 且 $1-f_R(x,z)\geqslant 1-\bigwedge_{y\in X}\{s[f_R(x,y),f_R(y,z)]\}\ \forall x,z\in X$,则称 R 为 vague 等价关系。

t-范数和 s-范数[9]分别是广义的"交"运算和广义的"并"运算。下面列出两个典型的 t-范数和 s-范数对偶算子,其中 $0\leqslant x,y\leqslant 1$:

(1)$t_V(x,y)=\max\{0,x+y-1\}$,$s_V(x,y)=\min\{1,x+y\}$;

(2)$t_p(x,y)=xy$,$s_p(x,y)=x+y-xy$。

3 *n* 步过程

扎德[10]将论域 X、Y 之间的模糊关系 R 定义为笛卡尔积 $X\times Y$ 的子集,并用隶属函数 $\mu_R(x,y)$ 表示元素 x 与 y 之间关系的程度,其取值范围为[0,1]。而且,他还给出了基于 max-min 的模糊等价关系 R 的定义,该关系必须满足三个条件,即:1)$\mu_R(x,x)=1$,$\forall x\in X$(自反性);2)$\mu_R(x,y)=\mu_R(y,x)$,$\exists x,y\in X$(对称性);3)$\mu_R(x,z)\geqslant\max_{y\in X}\{\min\{\mu_R(x,y),\mu_R(y,z)\}\}$,$\forall x,z\in X$(传递性)。很明显,基于 max-min 的模糊等价关系 R 是普通等价关系的一个模糊扩展。

在将模式识别应用于诸如视觉图像、气味、图形等实际方面时,人的主观判断提供了重要的信息。这些主观信息可以用相似模糊关系来表达。Tamura 等在文[6]中首先对基于模糊关系的模糊聚类进行了研究。由于相似模糊关系没有传递性,不能将其直接应用于聚类,因而,Tamura 等人利用模糊关系的合成运算构造了一个 n 步过程,用于将模糊相似关系改造为模糊等价关系。近来,Yang 等在文[7]中对 Tamura 等的 n 步过程进行了进一步扩展,并提出了一个基于模糊关系的聚类算法。本文将对 Yang 等的聚类算法进行扩展,并将之应用到 Vague 关系上。

定理 1(n 步过程) 设 $R^{(0)}=[(tr_{ij}^{(0)},1-f_{ij}^{(0)})]$ 是一个相似 vague 关系矩阵,那么利用 Vague 关系的 max-t & min-s 合成运算,有

$$I<R^{(0)}<R^{(1)}<\cdots<R^{(n)}=R^{(n+1)}=\cdots$$

这里 $R^{(n)}$ 是一个等价关系且 $I=[(0,0)]$。当 n 无限大时,有 $\lim_{n\to\infty}R^{(n)}=R^{(\infty)}=$,$R^{(\infty)}$ 为等价关系,即

$$I<R^{(0)}<R^{(1)}<\cdots<R^{(n)}<R^{(n+1)}<\cdots<R^{(\infty)}$$

证明:因为 $R^{(0)}$ 是一个近似 Vague 关系矩阵,故有 $I<R^{(0)}$,设 $R^{(1)}=[(tr_{ij}^{(1)},1-f_{ij}^{(1)})]$,其中 $tr_{ij}^{(1)}=\bigvee_k\{t(tr_{ik}^{(0)},tr_{kj}^{(0)})\}$,$f_{ij}^{(1)}=\bigwedge_k\{s(f_{ik}^{(0)},f_{kj}^{(0)})\}$。那么有

$$\left.\begin{aligned}&tr_{ij}^{(0)}=t(tr_{ij}^{(0)},1)=t(tr_{ij}^{(0)},tr_{jj}^{(0)})\leqslant\bigvee_k t(tr_{ik}^{(0)},f_{kj}^{(0)}=tr_{ij}^{(1)}\\&f_{ij}^{(0)}=s(f_{ij}^{(0)},0)=s(f_{ij}^{(0)},f_{jj}^{(0)})\geqslant\bigwedge_k s(f_{ik}^{(0)},f_{kj}^{(0)})=f_{ij}^{(1)}\end{aligned}\right\}$$

$\Leftrightarrow R^{(0)}\leqslant R^{(1)}$

假设 $R^{(0)}$ 没有传递性,那么对于 m 的某个值,必然存在 i 和 j,使得

$$\left.\begin{aligned}&tr_{ij}^{(0)}<t(tr_{im}^{(0)},tr_{mj}^{(0)})\Rightarrow tr_{ij}^{(0)}<\bigvee_k t(tr_{ik}^{(0)},\\&\quad rt_{kj}^{(0)})=tr_{ij}^{(1)}\\&f_{ij}^{(0)}>s(f_{im}^{(0)},f_{mj}^{(0)})\Rightarrow f_{ij}^{(0)}>\bigwedge_k s(f_{ik}^{(0)},\\&\quad f_{kj}^{(0)})=f_{ij}^{(1)},i.e.\ 1-f_{ij}^{(0)}<1-f_{ij}^{(1)}\end{aligned}\right\}\Rightarrow R^{(0)}<R^{(1)}$$

如果 $R^{(1)}$ 没有传递性,同样还可以得出 $R^{(1)}<R^{(2)}$。依次类推,如果经过 $(n-1)$ 次合成后所得到的 Vague 关系仍没有传递性,那么有

$$I<R^{(0)}<R^{(1)}<\cdots<R^{(n)}$$

假设 $R^{(n)}$ 具有传递性,那么对于所有的 i 和 j,有

$tr_{ij}^{(n)}\geqslant\bigvee_k t(tr_{ik}^{(n)},tr_{kj}^{(n)})=tr_{ij}^{(n+1)},tr_{ii}^{(n)},tr_{ij}^{(n)})=t(1,tr_{ij}^{(n)})=tr_{ij}^{(n)}\Rightarrow tr_{ij}^{(n+1)}=\bigvee_k t(tr_{ik}^{(n)},tr_{kj}^{(n)})\geqslant t(tr_{ii}^{(n)},tr_{ij}^{(n)})=tr_{ij}^{(n)}$

因而可得到 $tr_{ij}^{(n+1)}=tr_{ij}^{(n)}$。类似地,还可以得到 $f_{ij}^{(n+1)}=f_{ij}^{(n)}$,从而有 $R^{(n+1)}=R^{(n)}$。类似地,有 $R^{(n+2)}=R^{(n+1)}$。由此可得

$$I<R^{(0)}<R^{(1)}<\cdots<R^{(n)}=R^{(n+1)}=\cdots$$

如果不存在一个有限值 n 使得 $R^{(n)}=R^{(n+1)}=\cdots$,那么有

$$I<R^{(0)}<R^{(1)}<\cdots<R^{(n)}<R^{(n+1)}<\cdots<R^{(*)}$$

其中,$R^{(*)}=\begin{bmatrix}[1,1]&\cdots&[1,1]\\\vdots&&\vdots\\{}[1,1]&\cdots&[1,1]\end{bmatrix}$

因为 $\{R^{(k)}\mid k=0,1,2,\cdots\}$ 是单调有界的,因而必有 $R^{(\infty)}=\lim_{n\to\infty}R^{(n)}$ 存在。下面,我们来证明 $R^{(\infty)}$ 是一个等价的 Vague 关系。

已知 $tr_{ij}^{(n)}=\bigvee_k t(tr_{ik}^{(n-1)},tr_{kj}^{(n-1)})$,$f_{ij}^{(n)}=\bigwedge_k s(f_{ik}^{(n-1)},f_{kj}^{(n-1)})$

设 $tr'^{(n)}_{ij}=\bigvee_k t(tr'^{(n-1)}_{ik}),tr'^{(n-1)}_{kj}$,$f'^{(n)}_{ij}=\bigwedge_k s(f'^{(n-1)}_{ik},f'^{(n-1)}_{kj})$

尽管 $R^{(n)}$ 不同于 $R'^{(n)}$,但有 $\lim_{n\to\infty}R'^{(n)}=\lim_{n\to\infty}R'^{(n)}=R^{(\infty)}$ 成立。

$tr'^{(2)}_{ij}=\bigvee_{k_1}t(tr'^{(1)}_{ik_1},tr'^{(1)}_{k_1j})$,

$tr'^{(3)}_{ij}=\bigvee_{k_2}t(tr'^{(2)}_{ik_2},tr'^{(2)}_{k_2j})=\bigvee_{k_1,k_2}t(tr'^{(1)}_{ik_1},tr'(1)_{k_1k_2}$,

$tr'^{(1)}_{k_2,j})$

…

$$tr'^{(n)}_{ij}=\vee_{k_1,\cdots,k_{n-1}}t(\cdots(t(t(tr'^{(1)}_{ik_1},tr'^{(1)}_{k_1k_2}),tr'^{(1)}_{k_2k_3}),\cdots,tr'^{(1)}_{k_{n-1}j}))$$

$$tr'^{(m+n)}_{ij}=\vee_{k_1,\cdots,k_{m+n-1}}t(\cdots(t(t(tr'^{(1)}_{ik_1},tr'^{(1)}_{k_1k_2}),tr'^{(1)}_{k_2k_3}),\cdots,tr'^{(1)}_{k_{m+n-1}j}))$$

$$\geqslant\vee_{k_1,\cdots,k_{m-1}}\vee_{k_{m+1},\cdots,k_{m+n-1}}t(\cdots(t(t(tr'^{(1)}_{ik_1},tr'^{(1)}_{k_1,k_2}),tr'^{(1)}_{k_2k_3}),\cdots,tr'^{(1)}_{k_{m+n-1}j}))$$

$$=t(\vee_{k_1,\cdots,k_{m-1}}t(\cdots t(tr'^{(1)}_{ik_1}\ tr'^{(1)}_{k_1k_2}),\cdots,tr'^{(1)}_{k_{m-1}l}),\vee_{k_{m+1},\cdots,k_{m+n-1}}t(\cdots t(tr'^{(1)}_{lk_{m+1}},tr'^{(1)}_{k_{m+1}k_{m+2}}),\cdots,tr'^{(1)}_{k_{m+n-1}j}))$$

$$=t(tr'^{(m)}_{il},tr'^{(n)}_{lj})$$

类似地,可以得到 $1-f'^{m+n}_{ij}\geqslant1-s(f'^{(m)}_{il},f^{(n)}_{lj})$。

因此,对于所有的 l,有 $tr'^{(m+n)}_{ij}\geqslant t(tr'^{(m)}_{il},tr'^{(n)}_{lj})$ 和 $1-f'^{(m+n)}_{ij}\geqslant1-s(f'^{(m)}_{(il)},f'^{(n)}_{lj})$ 成立。当 $m\to\infty$ 和 $n\to\infty$ 时,对于所有的 l,可以得到

$$tr'^{(\infty)}_{ij}\geqslant t(tr'^{(\infty)}_{il},tr'^{(\infty)}_{il}),1-f'^{(\infty)}_{il}\geqslant1-s(f'^{(\infty)}_{il},f'^{(\infty)}_{lj})$$

又因为 $tr'^{(\infty)}_{ij}=1,1-f'^{(\infty)}_{il}=1$ 以及 $tr'^{(\infty)}_{ij}=tr'^{(\infty)}_{ji},1-f'^{(\infty)}_{ij}=1-f'^{(\infty)}_{ji}$,所以 $R^{(\infty)}$ 是一个等价的 Vague 关系。

定理 1 将 Yang 等的 n 步过程扩展到 Vague 关系上。也就是说,我们能够利用 max-t & min-s 合成将相似 Vague 关系矩阵改造为等价 Vague 关系矩阵。

例 1 给定如下相似 Vague 关系矩阵 $R^{(0)}$

$$R^{(0)}=\begin{bmatrix}[1,1] & & \\ [0.8,0.8] & [1\ 1] & \\ [0.65,0.7] & [0.15,0.2] & [1,1]\end{bmatrix}$$

(1)利用 max-t_p & min-s_p 合成,有

$$R^{(1)}=\begin{bmatrix}[1,1] & & \\ [0.8,0.8] & [1\ 1] & \\ [0.65,0.7] & [0.52,0.56] & [1,1]\end{bmatrix}=R^{(2)}$$

(2)利用 max-t_v & min-s_v 合成,有

$$R^{(1)}=\begin{bmatrix}[1,1] & & \\ [0.8,0.8] & [1\ 1] & \\ [0.65,0.7] & [0.45,0.5] & [1,1]\end{bmatrix}=R^{(2)}$$

在实际应用中,通常只能得到 Vague 相似关系,该关系不具备传递性。利用定理 1,我们可以使该关系具有传递性。从例 1 可以看出,利用 max-$t_\triangledown$ & min-$s_\triangledown$ 合成得到的结果[0.45, 0.5]更加接近原始关系值[$tr^{(0)}_{23}$, $1-f^{(0)}_{23}$]=[0.15, 0.2],在某种程度上,利用 max-$t_\triangledown$ & min-$s_\triangledown$ 合成能够将 Vague 相似关系改造为 vague 等价关系并能够保持与原始关系最大的相似性。也就是说,max-$t_\triangledown$ & min-$s_\triangledown$ 合成的结果看起来比 max-t_p & min-s_p 合成的结果更有意义。

定理 2 设 t_1,t_2 是任意两个 t-范数,s_1,s_2 是任意两个 s-范数,$R^{(0)}$ 是一个相似 Vague 关系矩阵。假设 $I<R^{(0)}<R^{(1)}<\cdots<R^{(m)}=R^{(m+1)}$ 是利用基于 max-t_1 & min-s_1 合成的 n 步过程得到的结果,而 $I<R^{(0)}<R^{(1)}<\cdots<R^{(n)}=R^{(n+1)}$ 是利用基于 max-t_2 & min-s_2 合成的 n 步过程得到的结果。如果 $t_1\leqslant t_2$ 且 $s_1\geqslant s_2$,那么有 $m\leqslant n$ 成立。

证明:因为 $R^{(n)}$ 是一个 max-t_2 & min-s_2 等价关系,因而有

$$tr^{(n)}_{ij}\geqslant\vee_k t_2(tr^{(n)}_{ik},tr^{(n)}_{kj}),f^{(n)}_{ij}\leqslant\wedge_k s_2(f^{(n)}_{ik},f^{(n)}_{kj})$$

又因为对任意 k 都有

$$t_2(tr^{(n)}_{ik},tr^{(n)}_{kj})\geqslant t_1(tr^{(n)}_{ik},s_2(tr^{(n)}_{kj}),f^{(n)}_{ik},f^{(n)}_{kj})\leqslant s_1(f^{(n)}_{ik},f^{(n)}_{kj})$$

由此可得 $tr^{(n)}_{ij}\geqslant V_k t_1(tr^{(n)}_{ik},tr^{(n)}_{kj}),f^{(n)}_{ij}\leqslant\wedge_k s_1(f^{(n)}_{ik},f^{(n)}_{kj})$

所以有 $m\leqslant n$ 成立。

根据定理 2,可以发现,利用 max-$t_\triangledown$ & min-$s_\triangledown$ 合成求得 $R^{(0)}$ 的传递闭包 $R^{(m)}$ 需要的步数不多于利用 max-t_p & min-s_p 合成求得 $R^{(0)}$ 的传递闭包 $R^{(n)}$ 所需要的步数,即 $m\leqslant n$。

定义 6 对于 $0\leqslant\alpha,\beta\leqslant11$ 且 $\alpha+\beta\leqslant1$,vague 关系 R 的 α-β 截关系定义如下:

$$R_{(\alpha,\beta)}=\{(x,y)\mid tr_R(x,y)\geqslant\alpha,\ f_R(x,y)\leqslant\beta\}$$

因此,下列命题成立:

若 $0\leqslant\alpha_1\leqslant\alpha_2\leqslant1,0\leqslant\beta_2\leqslant\beta_1$,其中 $0\leqslant\alpha_1+\beta_1\leqslant1,0\leqslant\alpha_2+\beta_2\leqslant1$,则有 $R_{(\alpha_2,\beta_2)}\subset R_{(\alpha_1,\beta_1)}$。

例 2 给定 $X=\{x_1,x_2,x_3\}$ 上的相似 Vague 关系矩阵 $R^{(0)}$ 如下:

$$R^{(0)}=\begin{bmatrix}[1,1] & & \\ [0.8,0.8] & [1\ 1] & \\ [0.65,0.7] & [0.15,0.2] & [1,1]\end{bmatrix}$$

与例 1 相同,有

$$R^{(1)}=\begin{bmatrix}[1,1] & & \\ [0.8,0.8] & [1\ 1] & \\ [0.65,0.7] & [0.45,0.2] & [1,1]\end{bmatrix}=R^{(2)}$$

是一个基于 max-$t_\triangledown$ & min-$s_\triangledown$ 合成的 max-t & min-s 等价关系,有

(1)当 $0<\alpha\leqslant0.45$, $0.5\leqslant\beta<1$(即 $0<1-f_R(x,y)\leqslant0.5$),$\alpha+\beta\leqslant1$ 时,可得$\{(x_1,x_1),(x_1,x_2),(x_1,x_3),(x_2,x_2),(x_2,x_3),(x_3,x_1),(x_3,x_2),(x_3,x_3)\}$,此时,$X$ 中的元素聚为一类:$\{x_1,x_2,x_3\}$;

(2)当 $0.65<\alpha\leqslant0.8$, $0.2\leqslant\beta<0.3$(即 $0.7<1-f_R(x,y)(0.8)$,$\alpha+\beta\leqslant1$ 时,可得 $R^{(1)}_{(\alpha,\beta)}=\{(x_1,x_1),(x_1,x_2),(x_2,x_1),(x_2,x_2),(x_3,x_3)$,此时,$X$ 中的元素聚为两类:$\{x_1,x_2\}$ 和$\{x_3\}$;

(3)当 $0.8<\alpha\leqslant1,0\leqslant\beta<0.2$(即 $0.8<1-f_R(x,y)\leqslant1$),$\alpha+\beta\leqslant1$ 时,可得 $R^{(1)}_{(\alpha,\beta)}=\{(x_1,x_1),(x_2,x_2),(x_3,x_3)\}$,此时,$X$ 中的元素聚为三类:$\{x_1\}$,$\{x_2\}$ 和$\{x_3\}$;

但是,

(4)当 $0.45<\alpha\leqslant0.65$, $0.3\leqslant\beta<0.5$(即 $0.5<1-f_R(x,y)\leqslant0.7$),$\alpha+\beta\leqslant1$ 时,可得 $R^{(1)}_{(\alpha,\beta)}=\{(x_1,x_1),(x_1,x_2),(x_1,x_3),(x_2,x_1),(x_2,x_2),(x_2,x_3),(x_3,x_1),(x_3,x_2),(x_3,x_3)\}$,此时,$x_1$ 和 x_2 以及 x_1 和 x_3 分别被聚为一类,但 x_2 和 x_3 却不在同一个类中。也就是说,我们不能从 $R^{(1)}_{(\alpha,\beta)}$ 中获得一个聚类结果。但是从 $R^{(1)}$ 中,可以看出,$tr_{R^{(1)}}(x_1,x_2)=0.8>0.65=tr_{R^{(1)}}(x_1,x_3)$,$f_{R^{(1)}}(x_1,x_2)=0.2<0.3=f_{R^{(1)}}(x_1,x_3)$。根据扩展的最大相似原则(称为 max-min 相似原则),应该首选将 x_1 和 x_2 归为一类,因此,有如下聚类结果:

当 $0.45<\alpha\leqslant0.65$, $0.3\leqslant\beta<0.5$,$\alpha+\beta\leqslant1$ 时,X 中的元素聚为两类:$\{x_1,x_2\}$,$\{x_3\}$。

4 聚类算法

在实际应用中,有许多关系是用相似关系矩阵表达的。由于相似关系不具备传递性,因而不能直接用于聚类。利用第 3 节提出来的基于 max-t & min-s 合成的 n 步过程可以将相似 Vague 关系矩阵改造为等价 Vague 关系矩阵,并可将该等价 Vague 关系矩阵用于聚类。从第 3 节可以看出,基于 max-$t_\triangledown$ & min-$s_\triangledown$ 合成的 n 步过程较基于 max-t_p & min-s_p 合成的 n 步过程更有意义,也更加有效。但是,从例 2 中可以看出,在某些情况下,并不能直接从 max-$t_\triangledown$ & min-$s_\triangledown$ 等价矩阵得到分类树,因而无法直接进行聚类。为解决此问题,本文基于扩展的最大相似原则,提出了一个适用于 max-t & min-s

等价关系矩阵的聚类算法。具体如下。

算法:基于 Vague 关系的模糊聚类

输入:相似 Vague 关系矩阵:$R^{(0)}=[tr_{ij}^{(0)}, 1-f_{ij}^{(0)}]_{n\times n}$

输出:类 C

Begin

1. 初始化 $I=\{1, 2, \cdots, n\}$并给定阈值(α,β);
2. 用基于 max-t & min-s 合成的 n 步过程计算 $R^{(0)}$ 相应的等价关系矩阵 $R=[tr_{ij}, 1-f_{ij}]_{n\times n}$;
3. 将所有满足$(i=j$ or $tr_{ij}\langle \alpha$ or $f_{ij}\rangle \beta)$的$[tr_{ij}, 1-f_{ij}]$置为$[0, 0]$;
4. $[tr_{pq}, 1-f_{pq}]=[\max\{tr_{ij} \mid i<j, i, j\in I\}, 1-\min\{f_{ij} \mid i<j, i, j\in I\}], \forall p,q\in I$;
   if 存在$[tr_{pq}, 1-f_{pq}]\neq[0,0]$ then
   $C=\{p, q\}$; GOTO 5;
   else
   GOTO 6;
5. for 每一个 $u\in I-C$
   if $\sum_{i\in C} tr_{iu}=\max\{\sum_{k\in C} tr_{ij} \mid j(\in I-C, tr_{ij}\neq 0$ for all $i\in C\}$ and $\sum_{i\in C} f_{iu}=\min\{\sum_{i\in C} f_{ij} \mid j\in I-C, f_{ij}\neq 0$ for all $i\in C\}$ then
   $\{C=C+\{u\}$; GOTO 5;$\}$
   else
   {输出类 C; $I=I-C$; GOTO 4; }
6. 将 I 中的所有值作为不同的类输出。

End

例 3 给定一个 8×8 相似 Vague 关系矩阵 $R^{(0)}$,并给定$(\alpha,\beta)=(0.5, 0.3)$。

$$R^{(0)}=\begin{bmatrix}
[1\ 1] & & & & & & & \\
[0.6\ 0.7] & [1\ 1] & & & & & & \\
[0.7\ 0.8] & [0.7\ 0.8] & [1\ 1] & & & & & \\
[0.9\ 0.9] & [0.3\ 0.4] & [0.2\ 0.25] & [1\ 1] & & & & \\
[0.2\ 0.2] & [0.5\ 0.5] & [0.4\ 0.4] & [0.1\ 0.2] & [1\ 1] & & & \\
[0.8\ 0.8] & [0.4\ 0.5] & [0.5\ 0.6] & [0.3\ 0.4] & [0.6\ 0.6] & [1\ 1] & & \\
[0.3\ 0.4] & [0.1\ 0.1] & [0.8\ 0.9] & [0.7\ 0.8] & [0.1\ 0.2] & [0.0\ 0.2] & [1\ 1] & \\
[0.2\ 0.2] & [0.3\ 0.4] & [0.91] & [0.2\ 0.2] & [0.0\ 0.2] & [0.2\ 0.3] & [0.1\ 0.2] & [1\ 1]
\end{bmatrix}$$

算法执行过程如下:

步骤 1 初始化:$I=\{1, 2, \cdots, 8\}$;$(\alpha,\beta)=(0.5, 0.3)$;

步骤 2 用基于 max-$t_\triangledown$ & min-$s_\triangledown$ 合成的 n 步过程计算相似 vague 关系矩阵 $R=[tr_{ij}, 1-f_{ij}]_{8\times 8}$;

$$R=\begin{bmatrix}
[1\ 1] & & & & & & & \\
[0.6\ 0.7] & [1\ 1] & & & & & & \\
[0.7\ 0.8] & [0.7\ 0.8] & [1\ 1] & & & & & \\
[0.9\ 0.9] & [0.5\ 0.6] & [0.6\ 0.7] & [1\ 1] & & & & \\
[0.4\ 0.4] & [0.5\ 0.5] & [0.4\ 0.4] & [0.3\ 0.3] & [1\ 1] & & & \\
[0.8\ 0.8] & [0.4\ 0.5] & [0.5\ 0.6] & [0.7\ 0.7] & [0.6\ 0.6] & [1\ 1] & & \\
[0.6\ 0.7] & [0.5\ 0.7] & [0.8\ 0.9] & [0.7\ 0.8] & [0.2\ 0.3] & [0.4\ 0.5] & [1\ 1] & \\
[0.6\ 0.8] & [0.6\ 0.8] & [0.91] & [0.5\ 0.7] & [0.3\ 0.4] & [0.4\ 0.6] & [0.7\ 0.9] & [1\ 1]
\end{bmatrix}$$

步骤 3 将所有满足条件$(i=j$ or $tr_{ij}<0.5$ or $1-f_{ij}<0.7)$的$[tr_{ij}, 1-f_{ij}]$置为$[0, 0]$,如下所示

$$R=\begin{bmatrix}
[0\ 0] & & & & & & & \\
[0.6\ 0.7] & [0\ 0] & & & & & & \\
[0.7\ 0.8] & [0.7\ 0.8] & [1\ 1] & & & & & \\
[0.9\ 0.9] & [0\ 0] & [0.6\ 0.7] & [1\ 1] & & & & \\
[0\ 0] & [0\ 0] & [0\ 0] & [0\ 0] & [1\ 1] & & & \\
[0.8\ 0.8] & [0\ 0] & [0\ 0] & [0.7\ 0.7] & [0\ 0] & [1\ 1] & & \\
[0.6\ 0.7] & [0.5\ 0.7] & [0.8\ 0.9] & [0.7\ 0.8] & [0\ 0] & [0\ 0] & [1\ 1] & \\
[0.6\ 0.8] & [0.6\ 0.8] & [0.91] & [0\ 0] & [0\ 0] & [0\ 0] & [0.7\ 0.9] & [1\ 1]
\end{bmatrix}$$

步骤 4 因为$[\max tr_{ij}, 1-\min f_{ij}]=[tr_{38}, 1-f_{38}]=[0.9, 1]$,所以有 $p=3, q=8$, $C=\{3,8\}$;

步骤 5 由$(\max_{i\in I-C}(tr_{3i}+tr_{8i}), \min_{i\in I-C}(f_{3i}+f_{8i}))=[tr_{37}, f_{37}]+[tr_{87}, f_{87}]=(1.5, 0.2)$可得$C=\{3, 8, 7\}$;

步骤 5 由$(\max_{i\in I-C}(tr_{3i}+tr_{8i}+tr_{7i}), \min_{i\in I-C}(f_{3i}+f_{8i}+f_{7i}))=[tr_{31}, f_{31}]+[tr_{81}, f_{81}]+[tr_{71}, f_{71}]=(1.9, 0.7)$可得$C=\{3, 8, 7, 1\}$;

步骤 5 因为不存在 $u\in I-C$ 使 $tr_{3u}+tr_{8u}+tr_{7u}+tr_{1u}$ 取得最大值,使 $f_{3u}+f_{8u}+f_{7u}+f_{1u}$ 取得最小值,因而 $C=\{3, 8, 7, 1\}$, $I=I-C=\{2, 4, 5, 6\}$;

步骤 4 由$[\max tr_{ij}, 1-\min f_{ij}]=[tr_{46}, 1-f_{46}]=[0.7, 0.7]$可得 $C=\{4, 6\}$;

步骤 5 因为不存在 $u\in I-C$ 使 $tr_{4u}+tr_{6u}$ 取得最大值,使 $f_{4u}+f_{6u}$ 取得最小值,因而 $C=\{4, 6\}$, $I=I\ (\ C=\{2, 5\}$;

步骤 4 $[tr_{25}, 1-f_{25}]=[0, 0]$,执行转向步骤 6;

步骤 6 将 I 中的所有值作为两个不同的类$\{2\}$,$\{5\}$输出。

因此,$(\alpha,\beta)=(0.5, 0.3)$时,可以得到 4 个聚类:$\{3, 8, 7, 1\}$, $\{4, 6\}$, $\{2\}$和 $\{5\}$。

结束语 本文将 Yang 等人的 n 步过程扩展到 Vague 关系上。基于扩展的 max-t & min-s n 步过程,可以将相似 Vague 关系矩阵改造为基于 max-t & min-s 合成的等价 Vague 关系矩阵。虽然基于 max-$t_\triangledown$ & min-$s_\triangledown$ 合成的 n 步过程较基于 max-t_p & min-s_p 合成的 n 步过程具有更高的效率,但在某些情况下也无法直接对基于该合成的等价矩阵进行聚类。为此,本文提出了一种适用于所有 max-t & min-s 等价 vague 关系矩阵的聚类算法。

参考文献

1 Zadeh L A. Fuzzy Sets. Information and Control, 1965, 8(3): 338~353

2 Gau W L, Buehrer D J. Vague Sets. IEEE Transactions on Systems, Man, and Cybernetics, 1993, 23(2): 610~614

3 Lu A, Ng W. Vague Sets or Intuitionistic Fuzzy Sets for Handling Vague Data: Which One is Better? . Lecture Notes in Computer Science, 2005, 3716: 401~416

4 Bezdek J C. Pattern Recognition with Fuzzy Objective Function Algorithms. New York: Plenum Press, 1981

5 Trauwaert E, Kaufman L, Rousseeuw P. Fuzzy Clustering Algorithms Based on the Maximum Likelihood Principle. Fuzzy Sets and Systems, 1991, 42: 213~227

6 Tamura S, Higuchi S, Tanaka K. Pattern Classification Based on Fuzzy Relations. IEEE Trans Systems Man Cyberne, 1978(1): 61~66

7 Yang M S, Shih HM. Cluster Analysis Based on Fuzzy Relations. Fuzzy Sets and Systems, 2001, 120(2): 197~212

8 De S K, Biswas R, Roy A R. An Application of Intuitionistic Fuzzy Sets in Medical Diagnosis. Fuzzy Sets and Systems, 2001, 117: 209~213

9 Zimmermann H J. Fuzzy Set Theory and Its Applications. Dordrecht, Kluwer, 1991

10. Zadeh L A. Similarity Relations and Fuzzy Ordering. Inform Sci, 1971, 3: 177~200