

基于粒矩阵的相容关系的知识表示

陈泽华 谢刚 阎高伟

(太原理工大学信息工程学院 太原 030024)

摘要 Rough集以等价关系为基础建立了知识表达系统和知识发现算法,相容关系是等价关系的推广。从粒矩阵角度定义了相容关系的知识表达系统、知识的依赖性与关联规则发现方法。在此基础上,把粒矩阵运算由等价关系推广到相容关系。

关键词 粒计算, Rough集理论, 粒矩阵, 相容关系, 关联规则

中图分类号 TP182 **文献标识码** A

Granular Matrix-based Knowledge Representation for Tolerance Relation

CHEN Ze-hua XIE Gang YAN Gao-wei

(College of Information Engineering, Taiyuan University of Technology, Taiyuan 030024, China)

Abstract Based on equivalence relation, knowledge representation system (KRS) and knowledge reduction algorithms were established by rough set theory (RST). Tolerance relation is extension of equivalence relation. KRS, upper and lower approximation, knowledge dependency and discovery of association rules of tolerance system were defined and computed by granular matrix.

Keywords Granular computing (GrC), Rough set theory (RST), Granular matrix (GrM), Tolerance relation, Association rule

1 引言

粒计算是一种解决复杂问题的方法和手段^[1-6],不同粒化方法产生不同粒度的知识。Pawlak教授以等价关系为基础,建立了 Rough集理论,完成了知识的逼近、表达、约简和推理^[7]。笔者曾经通过定义粒矩阵,将 Rough集的代数表达系统转化为粒矩阵的数值定义和运算,简化了问题的描述和求解^[8]。相容关系是等价关系的推广,本文从粒矩阵角度讨论了相容关系的知识表达、近似、依赖和关联规则发现。

2 相容关系的粒计算

2.1 基本概念

定义1(二元邻域与二元邻域系统^[2,10]) 令 $B \subseteq V \times U$ 为一个二元关系。给定一个 B , 对任一对象 $p \in V$, 都存在映射 $B: p \rightarrow N(p)$, 子集 $N(p) \subseteq U$, $N(p) = \{u | (p, u) \in B\}$, 它由 U 中所有和 p 点具有二元关系 B 的元素 u 组成。称 $N_B(p)$ 为二元关系 B 导出的 p 点的右邻域(简称邻域)。如果 $V = U$, 称 B 是 U 上的二元关系。

B 实现了粒化, 产生了二元关系的粒结构。粒结构是粒子(子集)的集合, 粒的内容可见, 粒子间的关系(重叠关系)由粒的内容确定。

二元邻域系统: 称 $\{N(p) | N(p) \subseteq U\}$ 为 U 上的右二元邻

域系统(简称二元邻域系统)。

定义2(诱导划分与中心知识^[9]) 给定一个 B , 令 $p_1, p_2 \in U$, 用下列方法在 U 上定义一个等价关系 $E_B: (p_1, p_2) \in E_B \Leftrightarrow N_B(p_1) = N_B(p_2)$, 容易证明 E_B 是一个由 B 诱导出的等价关系。称 E_B 在 U 上形成的划分 π_{E_B} 为诱导划分, $\pi_{E_B} = \{[p]_{E_B} | p \in U\}$ 。其中, $[p]_{E_B} = \{u \in U | N_B(p) = N_B(u), p \in U\}$ 。

定义 $[p]_{E_B}$ 为 $N(p)$ 的中心集(也称中心知识), 记作 $C(p) = [p]_{E_B}$ 。

定义3(等价关系与相容关系) 令 $R \subseteq U \times U$, $x, y \in U$, 如果 R 具有自反、对称和传递特性, 则称 R 是 U 上的等价关系。如果 $(x, y) \in R$, 则称 x 与 y 具有等价关系, 记作 xRy 。

令 $T \subseteq U \times U$, $x, y \in U$, 如果 T 具有自反、对称特性, 则称 T 是 U 上的相容关系。如果 $(x, y) \in T$, 则称 x 与 y 具有相容关系, 记作 xTy 。

定义4(商集与商结构^[2]) U 上的等价关系 E_B 产生一个 U 上的商集 Q 。 Q 中的点 q 是把等价类 $[p]_{E_B}$ (中心集) 当作一个内容不可见的点(黑箱)抽象得到的。存在自然映射 $P: U \rightarrow Q$ 或者 $p \mapsto q$, 其中: $p \in U, q \in Q$ 。令 $q_1, q_2 \in Q, q_1 = [p_1]_{E_B}, q_2 = [p_2]_{E_B}$ 。 U 上的二元关系 B 引起 Q 上的二元关系 $B_Q: (x_1, x_2) \in B \Rightarrow (q_1, q_2) \in B_Q$, 其中: $x_1 \in [p_1]_{E_B}, x_2 \in [p_2]_{E_B}$ (请注意, B 与 B_Q 不能相互确定)。

到稿日期: 2012-02-21 返修日期: 2012-07-11 本文受国家自然科学基金(60975032), 山西省青年自然科学基金(2010021016-1, 2011 011012-2), 2011 年山西省留学回国人员科技活动择优项目资助。

陈泽华(1974—), 女, 副教授, 主要研究领域为粗糙集、粒计算、智能信息处理, E-mail: chenzehua@tyut.edu.cn; 谢刚(1972—), 男, 教授, 主要研究领域为智能信息处理、图像处理、故障诊断; 阎高伟(1970—), 男, 副教授, 主要研究领域为智能信息处理、进化计算。

等价关系、相容关系都是特殊的二元关系。如果二元关系 B 是等价关系,诱导划分就是划分,诱导划分产生的中心集 $C(p)$ 就是划分产生的等价类。如果 B 是相容关系,则中心集 $C(p)=[p]_{E_T}$,其中 E_T 是相容关系 T 在 U 上产生的等价关系, $(p_1, p_2) \in E_T \Leftrightarrow N_T(p_1) = N_T(p_2)$,容易证明: $(x_1, x_2) \in T \Leftrightarrow (q_1, q_2) \in T_Q$ (T 与 T_Q 彼此确定, T_Q 是 U 上的相容关系 T 引起的 Q 上的相容关系)。

在二元关系与相容关系下,商集中的点(中心集)之间不会有交集,但是点的邻域有交集。在 Rough 集中,商集中点的邻域是空集。本文在相容关系下讨论相关概念。

定义 5(知识集与知识结构) 用对人类有意义的词汇描述商集中的每一个点,产生知识集 $KS = f(U/E_T) = \{words\}$ 。知识集上的相容关系具有知识结构,它反映了词与词之间的关系。

定义 6(相容知识库与相容知识表达系统) 令 $K=(U, T)$ 为一个相容知识库, U 是论域, T 是相容关系族,定义相容知识表达系统为: $KRS=(U, T, f, w)$,其中 U 是论域; T 是相容关系, $T=C \cup D$, C 为条件属性集, D 为决策属性集; f 是命名函数; w 是词。

在二元关系下,知识库与 KRS 不是一一对应的,易证明在相容关系与等价关系下,二者是一一对应的。把知识库表达成 KRS (知识表的形式),以便进行数据挖掘。

2.2 相容关系的知识近似

在 Rough 集理论中, Pawlak 教授分别从点集拓扑和集合的角度定义上、下近似,前者没有人工智能解释,而后者用集合表示知识。在等价关系下,这两个定义得出的解一样, Rough 集以集合定义为主。

本文采用拓扑定义做相容关系的上、下近似,易证明,拓扑定义可以用中心知识表示,具有知识的含义^[14]。

定义 7(相容关系的集合近似) 给定知识库 $K=(U, T)$,其中 $U \neq \emptyset$ 是一个有限集, T 是 U 上的相容关系族,对任意子集 $X \subseteq U$,定义:

$$T_* X = \{x | x \in U; N_T(x) \subseteq X\} \quad (1)$$

$$T^* X = \{x | x \in U; N_T(x) \cap X \neq \emptyset\} \quad (2)$$

$T_* X$ 与 $T^* X$ 分别为 X 的 T 下近似和 T 上近似。

3 基于粒矩阵的相容系统

3.1 基于粒矩阵的基本定义

定义 8(粒结构矩阵) 给定知识库 $K=(U, T)$,假设论域中有 n 个元素: $u_1, \dots, u_i, \dots, u_p, \dots, u_j, \dots, u_n$ ($1 \leq i, j < n$),对每一个 $u_p \in U$, T 产生一个二元邻域系统 $\{N_T(u_p) | N_T(u_p) \subseteq U\}$ 。

用一个长度为 l 的向量表示邻域 $N_T(u_p)$,如果 $u_i Tu_p$,则向量对应的第 i 位为 1,否则为 0。子集 $N_T(u_p)$ 被映射到二进制位串空间: $X \rightarrow \{0, 1\}^l$ 。

$\{N_T(u_p)\}$ 可以用粒结构矩阵 GrS_T 定义:

$$GrS_T = [r_{ij}]_{n \times n} = \begin{bmatrix} r_{11} & \dots & r_{1n} \\ \dots & r_{ij} & \dots \\ r_{n1} & \dots & r_{nn} \end{bmatrix} r_{ij} = \begin{cases} 1, & u_i Tu_j \\ 0, & \text{else} \end{cases} \quad (3)$$

定义 9(商集矩阵 QS_T) 用商集矩阵 QS_T 表达商集 Q 。 E_T 产生等价类的集合 $\{\{*\}\}$, U/E_T 表示把等价类当作点的集合 $\{[\{*\}]\}$ 。其中, $[\{*\}]$ 表示点。保留 GrS_T 中相同行

的第一行,删掉其余行,获得 $QS_T = [r_{ij}]_{m \times n}$ 。

定义 10(知识结构矩阵 KS_T) 保留 $QS_T = [r_{ij}]_{m \times n}$ 中相同列的第一列,删掉其余列,得到商结构矩阵,它具有自反、对称性。为商结构矩阵的每一行命名,就得到知识结构 $KS = f(U/E_T) = \{a, b, \dots, q\}$,称具有名字的矩阵为知识结构矩阵 $KS_T = [r_{ij}]_{m \times m}$ 。从形式上看,商结构矩阵和知识结构矩阵是一样的。

在本文,粒结构矩阵、商集矩阵、商结构矩阵和知识结构矩阵以及后面的粒关系矩阵统称为粒矩阵。

3.2 基于粒矩阵的相容关系的上下近似

给定相容知识库 $K=(U, T)$,用粒矩阵定义相容知识库中任一子集 X 的上、下近似。

定义 11 用二进制串表示子集 X ,定义粒关系矩阵 $GrRM$:

$$GrRM = GrS_T \times X' \quad (4)$$

$GrRM$ 反映了粒结构中所有子集和 X 的关系。令 $sum(i)$ 为 GrS_T 中第 i 行元素的和, $r(j)$ 为 $GrRM$ 中第 j 个元素。若 $sum(i) = r_j$,且 $i = j$,则 $N_T(i) \subseteq X$;若 $r_j \neq 0$ 则 $N_T(i) \cap X \neq \emptyset$;若 $r_j = 0$,则 $N_T(i) \cap X = \emptyset$ 。

上下近似可以定义如下:

$$T_* X = \bigcup \{C(p) | p \in U; sum(i) = r(j) \text{ for } i = j\} \quad (5)$$

$$T^* X = \bigcup \{C(p) | p \in U; r_j \neq 0\} \quad (6)$$

在这里及下文中,用 i 表示论域的第 i 个元素 u_i 。

3.3 基于粒矩阵的相容关系的连续性与依赖性

定义 12(连续性) 一个映射 $f: (U_1, B_1) \rightarrow (U_2, B_2)$ 在点 $x \in U_1$ 是连续的,当且仅当对 $y = f(x)$ ($y \in U_2$) 的每一个二元邻域 $N_{B_2}(y)$, $N_{B_1}(x)$,使 $f(N_{B_1}(x)) \subseteq N_{B_2}(y)$ 。 f 是连续映射当且仅当 f 在 U_1 上的每一点都连续。

Lin 教授于 1998 年^[2,10,11]开始讨论二元知识的强弱依赖关系。

定义 13 假设 B_1, B_2 是 U 上的二元关系,定义知识的 3 种依赖关系:

知识 B_2 弱依赖于知识 B_1 当且仅当恒等映射 $I_d: (U, B_1) \rightarrow (U, B_2)$ 是连续的,或者 $\forall x \in U$,总有 $N_{B_1}(x) \subseteq N_{B_2}(x)$,即 $B_1 \subseteq B_2$,表示为 $B_1 \rightarrow_w B_2$ 。

知识 B_2 强依赖于知识 B_1 当且仅当 $B_1 \rightarrow_w B_2$,并且每一个 B_2 -邻域都是一些 B_1 -邻域的并集,表示为 $B_1 \Rightarrow B_2$ 。

知识 B_2 中心依赖于知识 B_1 ,表示为 $B_1 \Rightarrow_c B_2$ 当且仅当:

$$(1) B_1 \rightarrow_w B_2 \text{ 即 } N_{B_1}(x) \subseteq N_{B_2}(x), B_1 \subseteq B_2.$$

$$(2) E_{B_1} \rightarrow_w E_{B_2} \text{ 即 } [p]_{E_{B_1}} \subseteq [p]_{E_{B_2}}, E_{B_1} \subseteq E_{B_2}.$$

推论 中心依赖会引起 Q 层空间的连续映射。

$$(1) E_{B_1} \rightarrow_w E_{B_2} \text{ 即 } E_{B_1} \subseteq E_{B_2}, \text{使 } Q_1 \text{ 到 } Q_2 \text{ 存在映射。}$$

$$(2) B_1 \rightarrow_w B_2 \text{ 即 } B_1 \subseteq B_2, Q_{B_1} \subseteq Q_{B_2} \text{ 使 } Q_1 \text{ 到 } Q_2 \text{ 的映射连续。}$$

强依赖一定是弱依赖,强依赖保证了某些概念可用已知知识进行表示。在等价关系下,与易证明知识的强、弱、中心依赖一样,统称知识的依赖性。在二元关系下,三者不同。文献^[12]讨论了相容关系下 3 种依赖的关系。本文用粒矩阵定义做相容关系下的相关定义和运算。

定义 14(相容关系中心依赖的粒矩阵定义) 令 $GrRM = (c_{ij})_{m \times n} = U/E_{T_1} \times (U/E_{T_2})'$

其中, $NE(i)$ 为 $GrRM$ 中第 i 行非零元素的个数。

$$(1) T_1 \subseteq T_2 \Leftrightarrow GrS_{T_1} \cap GrS_{T_2} = GrS_{T_1}.$$

$$(2) E_{T_1} \subseteq E_{T_2} \Leftrightarrow k = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^n \{c_{ij} | NE(i)=1\} = 1.$$

4 计算实例

4.1 求相容系统的粒矩阵和 KRS

令 $K=(U, T), U=\{1, 2, 3, 4, 5\}, T_1, T_2, T_3 \subseteq T$:

$T_1: N_1(1)=\{1, 2\}, N_1(2)=\{1, 2\}, N_1(3)=\{3, 5\}, N_1$

$(4)=\{4, 5\}, N_1(5)=\{3, 4, 5\}$

$T_2: N_2(1)=\{1, 3\}, N_2(2)=\{2, 3\}, N_2(3)=\{1, 2, 3, 4, 5\}, N_2(4)=\{3, 4, 5\}, N_2(5)=\{3, 4, 5\}$

$T_3: N_3(1)=\{1, 2, 4, 5\}, N_3(2)=\{1, 2, 4, 5\}, N_3(3)=\{3, 4, 5\}, N_3(4)=\{1, 2, 3, 4, 5\}, N_3(5)=\{1, 2, 3, 4, 5\}$

1) 粒结构矩阵 (GrS_T)

$$GrS_{T_1} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}, GrS_{T_2} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix},$$

$$GrS_{T_3} = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

2) 商集矩阵和知识结构矩阵

$E_{T_1} = \{\{1, 2\}, \{3\}, \{4\}, \{5\}\}$

$U/E_{T_1} = \{[\{1, 2\}], [\{3\}], [\{4\}], [\{5\}]\}$

$E_{T_2} = \{\{1\}, \{2\}, \{3\}, \{4, 5\}\}$

$U/E_{T_2} = \{[\{1\}], [\{2\}], [\{3\}], [\{4, 5\}]\}$

$E_{T_3} = \{\{1, 2\}, \{3\}, \{4, 5\}\}$

$U/E_{T_3} = \{[\{1, 2\}], [\{3\}], [\{4, 5\}]\}$

$$QS_{T_1} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}, QS_{T_2} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix},$$

$$QS_{T_3} = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$KS_1 = f(U/E_{T_1}) = \{a, b, c, d\}$

$KS_2 = f(U/E_{T_2}) = \{e, f, g, h\}$

$KS_3 = f(U/E_{T_3}) = \{x, y, z\}$

保留 QS_T 中相同列的第一列, 删掉其余列, 并给矩阵每行/列一个有意义的名字得到知识结构矩阵。同行为 1 的元素具有相容关系 T_Q 。

$$KS_{T_1} = \begin{matrix} a \\ b \\ c \\ d \end{matrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}, KS_{T_2} = \begin{matrix} e \\ f \\ g \\ h \end{matrix} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix},$$

$$KS_{T_3} = \begin{matrix} x \\ y \\ z \end{matrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$N_Q(a)=\{a\}, N_Q(b)=\{b, d\}, N_Q(c)=\{c, d\}, N_Q(d)=\{b, c, d\}, N_Q(e)=\{e, g\}, N_Q(f)=\{f, g\}, N_Q(g)=\{e, f, g, h\}, N_Q(h)=\{g, h\}, N_Q(x)=\{x, z\}, N_Q(y)=\{y, z\}, N_Q(z)=\{x, y, z\}$

3) 基于相容关系的知识表达系统

表 1 是例 1 中 $K=(U, T)$ 对应的 KRS。

由定义 4、定义 5, 以 T_1 列为例:

$P(1)=P(2)=q_1, P(3)=q_2, P(4)=q_3, P(5)=q_4$

$f: f(q_1)=a, f(q_2)=b; f(q_3)=c; f(q_4)=d$

同理可以得到表 1, 与 Rough 集的 KRS 不同之处在于, 在同一个 T 下, (粒) 知识之间的交集不为空。

表 1 基于相容关系的 KRS

U	T ₁	T ₂	T ₃
1	a	e	x
2	a	f	x
3	b	g	y
4	c	h	z
5	d	h	z

4.2 利用粒矩阵求相容关系下知识的上下近似

令 $K=(U, T): X, Y, Z \subseteq U$

$T: N_T(1)=\{1, 2\}, N_T(2)=\{1, 2\}, N_T(3)=\{3, 5\},$

$N_T(4)=\{4, 5\}, N_T(5)=\{3, 4, 5\}$

$E_T = \{E_1, E_2, E_3, E_4\} = \{\{1, 2\}, \{3\}, \{4\}, \{5\}\}$

$X=\{2, 3, 4\}, Y=\{1, 2, 3\}, Z=\{1, 2, 3, 4\}$

根据定义 7 求 $T_*X, T^*X; T_*Y, T^*Y; T_*Z, T^*Z$:

$T_*X = \phi, T^*X = U$

$T_*Y = \{1, 2\}, T^*Y = \{1, 2, 3, 5\}$

$T_*Z = \{1, 2\}, T^*Z = \{1, 2, 3, 4, 5\}$

利用粒矩阵:

$$GrS_T = [r_{ij}]_{5 \times 5} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

$X=[0 \ 1 \ 1 \ 1 \ 0]; Y=[1 \ 1 \ 1 \ 0 \ 0]; Z=[1 \ 1 \ 1 \ 1 \ 0]$

$GrRM_1 = GrS_T \times X' = [1 \ 1 \ 1 \ 1 \ 2]$

对于 GrS_T :

$sum(1)=sum(2)=sum(3)=sum(4)=2, sum(5)=3$

对于向量 X 和矩阵 $GrRM_1$:

$r(1)=r(2)=r(3)=r(4)=1, r(5)=2$

根据定义 11: $T_*X = \phi, T^*X = U$ 。

对于 $GrRM_2 = GrS_T \times Y' = [2 \ 2 \ 1 \ 0 \ 1]$ 有:

$r(1)=r(2)=2, r(3)=1, r(4)=0, r(5)=1$

$sum(1)=r(1), sum(2)=r(2)$

所以 $T_*Y = \{1, 2\}, T^*Y = \{1, 2, 3, 5\}$ 。

同理: $GrRM_3 = GrS_T \times Z' = [2 \ 2 \ 1 \ 1 \ 2]$

$T_*Z = \{1, 2\}, T^*Z = \{1, 2, 3, 4, 5\}$

矩阵简化了相容关系上下近似的计算。点集定义的中心知识表达方式使拓扑定义具有了知识的含义。

4.3 利用粒矩阵求相容关系下的关联规则

本节说明相容关系下关联规则的定义及粒矩阵描述。

(下转第 240 页)

测技术对话题进行聚类 and 预测评估。实验证明,该系统具有较高的响应速度和用户满意度,因而具有一定的应用价值和推广前景。未来的研究包括对等舆情监测系统体系结构、广域舆情监测与控制等。

参 考 文 献

[1] 郑魁,疏学明,袁宏永. 网络舆情热点信息自动发现方法[J]. 计算机工程,2010,36(3):4-6
 [2] 陈莉萍,杜军平. 突发事件热点话题识别系统及关键问题研究[J]. 计算机工程与应用,2011,47(2):19-22
 [3] 宋光慧,聂琰,郭建康. 高校网络舆情监测系统构建及应对机制研究[J]. 计算机与现代化,2011(11):120-122
 [4] 丁菊玲,勒中坚. 基于观点树的网络舆情危机预警方法[J]. 计算机应用研究,2011,28(9):3501-3505
 [5] 丁杰,徐俊刚. IPSMS:一个网络舆情监控系统的设计与实现

[J]. 计算机应用与软件,2010,27(4):188-190
 [6] 方薇,何留进,孙凯. 采用元胞自动机的网络舆情传播模型研究[J]. 计算机应用,2010,30(3):751-755
 [7] 闵可锐,赵迎宾,刘昕,等. 互联网话题识别与跟踪系统设计及实现[J]. 计算机工程,2008,34(19):212-214
 [8] 席耀一,林琛,李弼程,等. 基于语义相似度的论坛话题追踪方法[J]. 计算机应用,2011,31(1):93-95
 [9] 龙志伟,程葳. 基于词聚类的热点话题检测算法[J]. 计算机工程与设计,2011,32(6):2214-2217
 [10] 殷风景,肖卫东,葛斌,等. 一种面向网络话题发现的增量文本聚类算法[J]. 计算机应用研究,2011,28(1):54-57
 [11] 洪宇,张宇,范基礼,等. 基于语义域语言模型的中文话题关联检测[J]. 软件学报,2008,19(9):2265-2275
 [12] 吕楠,罗军勇,刘尧,等. 基于话题三层结构模型的话题演化分析算法[J]. 计算机工程,2009,35(23):71-73

(上接第 213 页)

定义 15(关联规则) 在相容关系下 $a \wedge e \rightarrow x$ 是关联规则,当且仅当 $(N_{T_1}(a) \subseteq N_{T_3}(x)) \wedge (N_{T_2}(e) \subseteq N_{T_3}(x))$ 。

相应地,定义 15 也可用粒矩阵进行表示。请注意,点的连续性是相容关系下关联规则成立的必要条件。

在表 1 中,条件属性 $C = \{T_1, T_2\}$,决策属性 $D = \{T_3\}$ 。由表 1 可看到决策规则:

$$a \wedge e \rightarrow x; a \wedge f \rightarrow x; b \wedge g \rightarrow y;$$

$$c \wedge h \rightarrow z; d \wedge h \rightarrow z$$

对于 T_1, T_3 ,有

$$GrRM = (c_{ij})_{m \times n} = U/E_{T_1} \times (U/E_{T_3})'$$

$$= \begin{bmatrix} 11000 \\ 00100 \\ 00010 \\ 00001 \end{bmatrix} \times \begin{bmatrix} 100 \\ 100 \\ 010 \\ 001 \\ 001 \end{bmatrix} = \begin{bmatrix} 200 \\ 010 \\ 001 \\ 001 \end{bmatrix}$$

由定义 14, $k=1$ 且 $GrS_{T_1} \cap GrS_{T_3} = GrS_{T_1}$,故 T_3 中心依赖于 T_1 。同理,对于 $T_2 T_3$,虽然 $E_{T_2} \subseteq E_{T_3}$,但是 $GrS_{T_2} \cap GrS_{T_3} \neq GrS_{T_2}$, T_2 到 T_3 只在 h 点连续, $N_Q(h) \subseteq N_Q(z)$ 。因此表 1 最终可以得到 2 条关联规则: $c \wedge h \rightarrow z; d \wedge h \rightarrow z$ 。

结束语 等价关系、相容关系都是特殊的二元关系,都可当作粒化方法,并因此衍生相应的知识处理方法和手段。

2005 年起,笔者从二进制粒矩阵理论的角度讨论了 Rough 集理论的上下近似^[13],并用矩阵系统定义了 Rough 集代数定义系统。本文把基于等价关系的矩阵运算拓展到相容关系,定义了基于粒矩阵的相容关系的上、下近似和知识表达系统,给出了相容关系的关联规则发现算法。算例表明了各种概念和定义的可计算性,这些定义可以进一步实现相容关系的知识约简,还可以拓展到普通二元关系中,相关研究还在进行中。

参 考 文 献

[1] Zadeh L A. Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems[J]. Soft Computing, 1998(2):

23-25
 [2] Lin T Y. Granular Computing: Practices, Theories, and Future Directions[C]// Encyclopedia of Complexity and Systems Science. 2009:4339-4355
 [3] 张铃,张钊. 问题求解理论及应用-商空间粒度计算理论及应用(第 2 版)[M]. 北京:清华大学出版社,2007
 [4] 刘清,孙辉,王洪发. 粒计算研究现状及基于 Rough 逻辑语义的粒计算研究[J]. 计算机学报,2008,31(4):543-555
 [5] 苗夺谦,王国胤,刘清,等. 粒计算:过去、现在与展望[M]. 北京:科学出版社,2007
 [6] 王国胤,张清华,胡军. 粒计算研究综述[J]. 智能系统学报,2007,2(6):8-26
 [7] Pawlak Z. Rough sets: Theoretical Aspects of Reasoning About Data[M]. Dordrecht: Kluwer Academic Publishers, 1991
 [8] 陈泽华,谢刚,谢璐,等. 基于粒计算的 Rough 集模型[J]. 计算机科学,2009,36(5):200-203
 [9] Chen Z H, Lin T Y. Matrix Theory for Tolerance Granular Computing[C]// Proceedings of IEEE International Conference on Granular Computing. San Jose, USA, Aug 2010: 673-676
 [10] Lin T Y. Granular Computing on Binary Relations I: Data Mining and Neighborhood Systems[M]// Skowron A, Polkowski L, eds. Rough Sets In Knowledge Discovery. Physica-Verlag, 1998:107-121
 [11] Lin T Y. Chinese Wall Security Policy Models: Information Flows and Confining Trojan Horses[M]// Vimercati S, Ray I, Ray I, eds. Data and Applications Security XVII: Status and Prospects. Kluwer Academic Publishers, 2003:275-297
 [12] Meng J, Wang P, Wang X K, et al. Rule Induction for Tolerance Relation-based Rough Sets[C]// Proceedings of IEEE Granular Computing. Kaohsiung, China, Nov 2011:445-450
 [13] Chen Z H, Xie G, Yan G W, et al. Application of a Matrix-Based Binary Granular Computing Algorithm in RST[C]// Proceedings of IEEE Granular Computing. Beijing, China, July 2005: 409-412
 [14] Chen Z H, Lin T Y, Xie G. Knowledge Approximations and Representations in Binary Granular Computing[C]// Proceedings of IEEE Granular Computing. Hangzhou, China, Aug 2012:75-79