

不完备信息系统中对象的相似性刻画

罗珺方 秦克云

(西南交通大学数学学院 成都 610031)

摘要 信息系统知识约简与知识发现是粗糙集理论的重要研究方向。通过对不完备信息系统中对象的相似性刻画方式以及相应的粗糙集模型进行对比分析,基于限制容差关系提出了一种新的限制相似关系,建立了基于限制相似关系的粗糙集模型,讨论了模型的基本性质及其与已有模型之间的关系。

关键词 粗糙集,容差关系,限制容差关系,限制相似关系

中图分类号 TP18 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.1.052

Similarity of Objects in Incomplete Information Systems

LUO Jun-fang QIN Ke-yun

(College of Mathematics, Southwest Jiaotong University, Chengdu 610031, China)

Abstract The knowledge reduction and knowledge discovery in the information systems are important topics of rough set theory. Based on the analysis of existing similarity relations and related rough set models in incomplete information systems, we proposed a new kind of similarity relation, called limited similarity relation, to characterize the objects similarity in incomplete information systems. The rough set model based on the limited similarity relation was established with its basic properties and the relationships between the existing rough set models were discussed.

Keywords Rough set, Tolerance relation, Limited tolerance relation, Limited similarity relation

1 引言

粗糙集理论^[1]是 Pawlak 于 1982 年提出的一种处理不确定性问题的数学工具。经过 30 多年的研究与发展,粗糙集理论已成为一种重要的智能信息处理技术,目前已在信息系统知识发现、决策分析、模式识别、医疗诊断、文本分类、海量数据处理等领域获得了成功的应用。

Pawlak 最初提出的经典粗糙集理论用于处理精确无缺损数据,不包含对原始不完备、模糊性和随机性数据的处理机制。然而随着科学技术的飞速发展,各个领域的数据随着时间的推移急剧增加,同时由于数据采集技术的限制和数据的丢失,使得我们所面临的大量信息系统都是不完备的。如何将粗糙集理论应用于具有不完整、不精确或缺损数据的信息系统,对于粗糙集理论的发展具有重要的理论意义。人们针对不完备信息系统提出了若干刻画对象相似性的方法,并研究了相应粗糙集模型中近似算子的性质^[2-10]。本文对已有的不完备信息系统中对象相似性刻画方式进行了对比分析,在此基础上提出了限制相似关系,讨论了相应的粗糙集模型中近似算子的性质,并研究了相应粗糙集模型之间的关系。

2 基于不完备信息系统的粗糙集模型

Pawlak 粗糙集模型^[1]可以处理完备信息,其中的知识形式化为论域上的等价关系。为拓广粗糙集理论的应用范围,

人们提出了多种形式的粗糙集推广模型,其中 Yao^[9]基于论域上的一般二元关系提出了广义粗糙集模型。

定义 1^[9] 设 U 是非空集合,称为论域, R 是 U 上的一个二元关系,称 (U, R) 为一个广义近似空间,对于任意 $X \subseteq U$, X 关于 (U, R) 的上、下近似分别定义为:

$$\bar{R}(X) = \{x \in U; R_s(x) \cap X \neq \emptyset\} \quad (1)$$

$$\underline{R}(X) = \{x \in U; R_s(x) \subseteq X\} \quad (2)$$

其中, $R_s(x) = \{y \in U; (x, y) \in R\}$ 为 x 关于 R 的右邻域。

令 $R_s^{-1}(x) = \{y \in U; (y, x) \in R\}$ 。容易证明,对于任意 $X \subseteq U$, 有 $\bar{R}(X) = \bigcup_{x \in X} R_s^{-1}(x)$ 。因此,如果 R 是对称二元关系,则 $\bar{R}(X) = \bigcup_{x \in X} R_s(x)$ 。

信息系统属性约简与知识获取是粗糙集理论的重要研究方向。一个信息系统是一个二元组 $T = (U, A)$, 其中 U 是非空有限集合,称为论域,其中的元素称为对象; A 是非空有限集合,其中的元素称为属性,且对于任意 $a \in A, a: U \rightarrow V_a$ 为一映射, V_a 是属性 a 的取值构成的集合。如果某些对象关于某些属性的属性值未知(用 * 表示),则称此信息系统为不完备信息系统。

Krystkiewicz^[2]提出了一种处理不完备信息系统的粗糙集模型。若对象关于某属性的属性值未知,则认为它可能取到任意已知属性值,即未知属性值被解释为任意已知属性值。基于此, Krystkiewicz 提出了描述对象相似性的容差关系。

到稿日期:2015-01-10 返修日期:2015-04-03 本文受国家自然科学基金(61473239, 61175044, 61372187), 中央高校基础研究基金(2682014ZT28), 西华大学省部级学科平台开放课题(szjj2014-052), 西南交通大学研究生创新实验实践项目(YC20154010)资助。

罗珺方(1990—),女,硕士生,主要研究方向为粗糙集理论与应用, E-mail:junfangluo@163.com; 秦克云(1962—),男,教授,博士生导师,CCF 高级会员,主要研究方向为粗糙集理论、粒计算、多值逻辑等, E-mail:keyunqin@263.net。

定义 2^[2] 给定不完备信息系统 $IT=(U,A), B\subseteq A$ 。由 B 确定的容差关系为:

$$T_B = \{(x, y) \in U \times U; \forall b \in B(b(x) = b(y) \vee b(x) = * \vee b(y) = *)\} \quad (3)$$

显然 T_B 是自反、对称关系,但一般不传递。对于任意 $x \in U, x$ 关于 T_B 的容差类记为 $T_B(x) = \{y \in U; (x, y) \in T_B\}$ 。基于广义粗糙集模型,对象集合 $X \subseteq U$ 基于容差关系 T_B 的上、下近似分别定义为:

$$\begin{aligned} \underline{T}_B(X) &= \{x \in U; T_B(x) \subseteq X\} \\ \overline{T}_B(X) &= \{x \in U; T_B(x) \cap X \neq \emptyset\} \end{aligned}$$

Stefanowski^[3]认为,不完备信息系统中对象被不完全描述的原因可能是由于知识不精确,还可能是由于本质上就不可能用所有的属性来描述它们。因此,未知属性值在当前不存在的,不允许比较未知值。基于此,Stefanowski 提出了关于对象的非对称相似关系。

定义 3^[3] 给定不完备信息系统 $IT=(U,A), B\subseteq A$ 。由 B 确定的非对称相似关系 S_B 为:

$$S_B = \{(x, y) \in U \times U; \forall b \in B(b(x) = b(y) \vee b(x) = *)\} \quad (4)$$

显然 S_B 是自反、传递关系,但不一定对称。

记 $S_B(x) = \{y \in U; (x, y) \in S_B\}$ 为 x 的非对称相似类, $S_B^{-1}(x) = \{y \in U; (y, x) \in S_B\}$ 。对象集合 $X \subseteq U$ 关于非对称相似关系 S_B 的上、下近似分别定义为:

$$\begin{aligned} \underline{S}_B(X) &= \{x \in U; S_B(x) \subseteq X\} \\ \overline{S}_B(X) &= \{x \in U; S_B(x) \cap X \neq \emptyset\} \end{aligned}$$

由于 S_B 的非对称性,一些明显具有大量相同的已知属性值的信息从直觉上就可以判定为相似的对象,可能会被分于不同的相似类。另外,对于某些不完备信息系统,容差关系的要求过于宽松,而非对称相似关系的要求过于严格。为此,王国胤提出了限制容差关系来描述对象的相似性。

定义 4^[4] 给定不完备信息系统 $IT=(U,A), B\subseteq A$ 。由 B 确定的限制容差关系记为 L_B ,且:

$$L_B = \{(x, y) \in U \times U; (\forall b \in B(b(x) = b(y) = *) \vee ((P_B(x) \cap P_B(y) \neq \emptyset) \wedge (\forall b \in P_B(x) \cap P_B(y)(b(x) = b(y))))))\} \quad (5)$$

其中, $P_B(x) = \{b; b \in B \text{ 且 } b(x) \neq *\}$ 。

显然 L_B 是自反、对称关系,但不一定传递。对于任意 $x \in U, x$ 关于 L_B 的限制容差类记为 $L_B(x) = \{y \in U; (x, y) \in L_B\}$,对象集合 $X \subseteq U$ 关于限制容差关系 L_B 的上、下近似分别为:

$$\begin{aligned} \underline{L}_B(X) &= \{x \in U; L_B(x) \subseteq X\} \\ \overline{L}_B(X) &= \{x \in U; L_B(x) \cap X \neq \emptyset\} \end{aligned}$$

尹旭日^[5]提出了介于容差关系 T_B 和限制容差关系 L_B 之间的约束非对称相似关系 C_B ,定义如下。

定义 5^[5] 给定不完备信息系统 $IT=(U,A), B\subseteq A$ 。由 B 确定的约束非对称相似关系记为 C_B ,且:

$$C_B = \{(x, y) \in U \times U; (\forall b \in B(b(x) = *) \vee ((P_B(x) \cap P_B(y) \neq \emptyset) \wedge (\forall b \in P_B(x) \cap P_B(y)(b(x) = b(y))))))\} \quad (6)$$

C_B 是自反、传递关系,但不一定对称。记 $C_B(x) = \{y \in U; (x, y) \in C_B\}$ 为对象 $x \in U$ 关于 C_B 的约束非对称相似类, $C_B^{-1}(x) = \{y \in U; (y, x) \in C_B\}$ 。对象集合 $X \subseteq U$ 关于约束非

对称相似关系 C_B 的上、下近似分别为:

$$\begin{aligned} \underline{C}_B(X) &= \{x \in U; C_B(x) \subseteq X\} \\ \overline{C}_B(X) &= \{x \in U; C_B(x) \cap X \neq \emptyset\} \end{aligned}$$

例 1 设对象集 $U = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}\}$, 条件属性集 $B = \{b_1, b_2, b_3, b_4\}$, 决策属性为 $d, \Phi = \{c_1, c_2, c_3, c_4, c_8, c_{10}\}, \Psi = \{c_5, c_6, c_7, c_9, c_{11}\}$ 为决策类。信息函数如表 1 所列。

表 1 不完备信息系统

A	b ₁	b ₂	b ₃	b ₄	d
c ₁	3	2	1	1	Φ
c ₂	3	2	1	0	Φ
c ₃	3	2	1	*	Φ
c ₄	3	*	1	*	Φ
c ₅	*	1	*	0	Ψ
c ₆	1	1	3	*	Ψ
c ₇	0	1	3	1	Ψ
c ₈	*	*	*	1	Φ
c ₉	*	*	*	1	Ψ
c ₁₀	*	*	*	*	Φ
c ₁₁	*	*	*	*	Ψ

依据容差关系得到容差类及决策类的上、下近似:

$$\begin{aligned} T_B(c_1) &= \{c_1, c_3, c_4, c_8, c_9, c_{10}, c_{11}\} \\ T_B(c_2) &= \{c_2, c_3, c_4, c_{10}, c_{11}\} \\ T_B(c_3) &= \{c_1, c_2, c_3, c_4, c_8, c_9, c_{10}, c_{11}\} \\ T_B(c_4) &= \{c_1, c_2, c_3, c_4, c_5, c_8, c_9, c_{10}, c_{11}\} \\ T_B(c_5) &= \{c_4, c_5, c_6, c_{10}, c_{11}\} \\ T_B(c_6) &= \{c_5, c_6, c_8, c_9, c_{10}, c_{11}\} \\ T_B(c_7) &= \{c_7, c_8, c_9, c_{10}, c_{11}\} \\ T_B(c_8) &= T_B(c_9) = \{c_1, c_3, c_4, c_6, c_7, c_8, c_9, c_{10}, c_{11}\} \\ T_B(c_{10}) &= T_B(c_{11}) = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}, c_{11}\} \end{aligned}$$

$$\overline{T}_B(\Phi) = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}, c_{11}\}$$

$$\underline{T}_B(\Phi) = \emptyset$$

$$\overline{T}_B(\Psi) = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}, c_{11}\}$$

$$\underline{T}_B(\Psi) = \emptyset$$

注意到,在容差关系下 c_3 和 c_4, c_9 和 c_{10} 没有一个已知的属性值相同却被分在同一类中。依据非对称相似关系得到对象的相似类及决策类的上、下近似:

$$S_B(c_1) = \{c_1\}, S_B(c_2) = \{c_2\}, S_B(c_3) = \{c_1, c_2, c_3\}$$

$$S_B(c_4) = \{c_1, c_2, c_3, c_4\}, S_B(c_5) = \{c_5\}, S_B(c_6) = \{c_6\}$$

$$S_B(c_7) = \{c_7\}, S_B(c_8) = S_B(c_9) = \{c_1, c_7, c_8, c_9\}$$

$$S_B(c_{10}) = S_B(c_{11}) = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}, c_{11}\}$$

$$\overline{S}_B(\Phi) = \{c_1, c_2, c_3, c_4, c_8, c_9, c_{10}, c_{11}\}$$

$$\underline{S}_B(\Phi) = \{c_1, c_2, c_3, c_4\}$$

$$\overline{S}_B(\Psi) = \{c_5, c_6, c_7, c_8, c_9, c_{10}, c_{11}\}$$

$$\underline{S}_B(\Psi) = \{c_5, c_6, c_7\}$$

基于非对称相似关系的分类方法有时比较严苛。比如 c_1 和 c_3 有 3 个已知的属性值相等,相似概率很大,但没有被分在同一个类中,即 $c_3 \notin S_B(c_1)$ 。依据限制容差关系可以得到:

$$L(c_1) = \{c_1, c_3, c_4, c_8, c_9\}, L(c_2) = \{c_2, c_3, c_4\}$$

$$L(c_3) = \{c_1, c_2, c_3, c_4\}, L(c_4) = \{c_1, c_2, c_3, c_4\}$$

$$L(c_5) = \{c_5, c_6\}, L(c_6) = \{c_5, c_6\}$$

$$L(c_7) = \{c_7, c_8, c_9\}, L(c_8) = L(c_9) = \{c_1, c_7, c_8, c_9\}$$

$$L(c_{10}) = \{c_{10}, c_{11}\}, L(c_{11}) = \{c_{10}, c_{11}\}$$

$$\overline{L}_B(\Phi) = \{c_1, c_2, c_3, c_4, c_7, c_8, c_9, c_{10}, c_{11}\}$$

$$\underline{L}_B(\Phi) = \{c_2, c_3, c_4\}$$

$$\overline{L}_B(\Psi) = \{c_1, c_5, c_6, c_7, c_8, c_9, c_{10}, c_{11}\}$$

$$\underline{L}_B(\Psi) = \{c_5, c_6\}$$

根据约束非对称相似关系,有:

$$C_B(c_1) = \{c_1, c_3, c_4, c_8, c_9\}, C_B(c_2) = \{c_2, c_3, c_4\}$$

$$C_B(c_3) = \{c_1, c_2, c_3, c_4\}, C_B(c_4) = \{c_1, c_2, c_3, c_4\}$$

$$C_B(c_5) = \{c_5, c_6\}, C_B(c_6) = \{c_5, c_6\}, C_B(c_7) = \{c_7, c_8, c_9\}$$

$$C_B(c_8) = C_B(c_9) = \{c_1, c_7, c_8, c_9\}$$

$$C_B(c_{10}) = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}, c_{11}\}$$

$$C_B(c_{11}) = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}, c_{11}\}$$

$$\overline{C}_B(\Phi) = \{c_1, c_2, c_3, c_4, c_7, c_8, c_9, c_{10}, c_{11}\}$$

$$\underline{C}_B(\Phi) = \{c_2, c_3, c_4\}$$

$$\overline{C}_B(\Psi) = \{c_1, c_5, c_6, c_7, c_8, c_9, c_{10}, c_{11}\}$$

$$\underline{C}_B(\Psi) = \{c_5, c_6\}$$

3 限制相似关系

设 $IT=(U, A)$ 为不完备信息系统, $B \subseteq A$ 。基于限制容差关系,两个对象 $x, y \in U$ 不可区分包括以下两种情况:(1) x, y 关于 B 中属性的取值全为未知值;(2) x, y 关于 B 中某些属性取值同为已知值,且关于这些属性取值相同。当两个对象关于 B 中属性全为未知值时, B 中属性不足以描述 x, y , 某种意义上可以认为它们不可分辨的可能性很小,但在容差关系和限制容差关系下均是不可分辨的。为解决上述问题,本文提出一种新的对象相似性关系,称为限制相似关系,定义如下。

定义 6 给定不完备信息系统 $IT=(U, A), B \subseteq A$ 。由 B 确定的限制相似关系记为 M_B ,且:

$$M_B = \{(x, y) \in U \times U; ((P_B(x) \cap P_B(y) \neq \emptyset) \wedge (\forall b \in P_B(x) \cap P_B(y) (b(x) = b(y))))\} \cup I_U \quad (7)$$

其中, $I_U = \{(x, x); x \in U\}$ 是论域 U 上的恒等关系。

显然 M_B 是自反、对称的关系,但一般不传递。对于任意 $x \in U, x$ 关于 M_B 的容差类记为 $\underline{M}_B(x) = \{y \in U; (x, y) \in M_B\}$ 。对象集合 $X \subseteq U$ 关于限制容差关系 M_B 的上、下近似分别定义为:

$$\underline{M}_B(X) = \{x \in U; M_B(x) \subseteq X\}$$

$$\overline{M}_B(X) = \{x \in U; M_B(x) \cap X \neq \emptyset\}$$

定理 1 设 $IT=(U, A)$ 为不完备信息系统, $B \subseteq A, X, Y \subseteq U$ 。

$$(1) \underline{M}_B(\sim X) = \sim \overline{M}_B(X), \overline{M}_B(\sim X) = \sim \underline{M}_B(X)$$

$$(2) \underline{M}_B(U) = U, \overline{M}_B(\emptyset) = \emptyset$$

$$(3) \underline{M}_B(X \cap Y) = \underline{M}_B(X) \cap \underline{M}_B(Y), \overline{M}_B(X \cup Y) = \overline{M}_B(X) \cup \overline{M}_B(Y)$$

$$(4) \text{若 } X \subseteq Y, \text{则 } \underline{M}_B(X) \subseteq \underline{M}_B(Y), \overline{M}_B(X) \subseteq \overline{M}_B(Y)$$

$$(5) \underline{M}_B(X) \subseteq X \subseteq \overline{M}_B(X)$$

$$(6) X \subseteq \underline{M}_B(\overline{M}_B(X)), \overline{M}_B(\underline{M}_B(X)) \subseteq X$$

对于不完备信息系统 $IT=(U, A), B \subseteq A, x \in U$, 下列性质一般不成立:

$$(1) M_B = \bigcap_{b \in B} M_b;$$

$$(2) M_B(x) = \bigcap_{b \in B} (M_b(x));$$

$$(3) B_1 \subseteq B_2 \Rightarrow M_{B_2} \subseteq M_{B_1};$$

$$(4) B_1 \subseteq B_2 \Rightarrow M_{B_2}(x) \subseteq M_{B_1}(x)$$

例 2 考虑不完备信息系统 $IT=(U, A)^{[8]}$, 其中论域 $U = \{x_1, x_2, x_3, x_4\}$, 属性集 $A = \{c_1, c_2, c_3\}$, 如表 2 所列。

表 2 一个不完备信息系统 $IT=(U, A)$

	c_1	c_2	c_3
x_1	*	1	2
x_2	2	*	2
x_3	3	*	*
x_4	2	2	2

令 $B = \{c_1, c_2\}, I_U = \{(x, x); x \in U\}$ 是论域 U 上的恒等关系, 则计算得到如下结果:

$$M_A = \{(x_1, x_2), (x_2, x_1), (x_2, x_4), (x_4, x_2)\} \cup I_U$$

$$M_B = \{(x_2, x_4), (x_4, x_2)\} \cup I_U$$

$$M_{\{c_1\}} = \{(x_2, x_4), (x_4, x_2)\} \cup I_U$$

$$M_{\{c_2\}} = \{(x_2, x_3), (x_3, x_2)\} \cup I_U$$

$$\text{于是 } \bigcap_{b \in B} M_{\{b\}} = M_{\{c_1\}} \cap M_{\{c_2\}} = I_U \neq M_B$$

另外, 由于 $M_{\{c_1\}}(x_2) = \{x_2, x_4\}, M_{\{c_2\}}(x_2) = \{x_2, x_3\}, \bigcap_{b \in A} (M_{\{b\}}(x_2)) = \{x_2\}$, 因此 $M_B(x_2) = \{x_2, x_4\} \neq \bigcap_{b \in A} (M_{\{b\}}(x_2))$ 。类似地, (3)、(4)一般不成立。

限制相似关系和限制相似类随着属性集的增大不一定单调递减,即在限制相似关系下,原来不可分辨的对象随着属性的减少会变为可分辨。

下面讨论不完备信息系统的不可区分关系以及相应的近似算子之间的关系。

定理 2 设 $S=(U, A, V, f)$ 为不完备信息系统, $B \subseteq A$ 。

$$(1) S_B \subseteq C_B \subseteq T_B$$

$$(2) M_B \subseteq L_B \subseteq C_B \subseteq T_B$$

证明:(1)对于任意 $x, y \in U$, 设 $(x, y) \in S_B$, 于是对于任意 $b \in B$, 有 $b(x) = *$ 或 $b(x) = b(y)$ 。如果 $P_B(x) = \emptyset$, 则对于任意 $b \in B$, 有 $b(x) = *$, 从而 $(x, y) \in C_B$ 。如果 $P_B(x) \neq \emptyset$, 则对于任意 $b \in P_B(x)$, 有 $b(x) = b(y)$, 从而 $P_B(x) \subseteq P_B(y)$ 。于是 $P_B(x) \cap P_B(y) = P_B(x) \neq \emptyset$, 且对于任意 $b \in P_B(x) \cap P_B(y)$, 有 $b(x) = b(y)$, 从而有 $(x, y) \in C_B$ 。故 $S_B \subseteq C_B$ 。

对于任意 $x, y \in U$, 设 $(x, y) \in C_B$ 。如果 $\forall b \in B (b(x) = *)$, 显然有 $(x, y) \in T_B$ 。如果 $P_B(x) \cap P_B(y) \neq \emptyset$, 且对于任意 $b \in P_B(x) \cap P_B(y)$, 有 $b(x) = b(y)$, 则当 $b(x) \neq *$ 且 $b(y) \neq *$ 时, 有 $b(x) = b(y)$, 从而 $(x, y) \in T_B$ 。于是有 $C_B \subseteq T_B$ 。

(2)对于任意 $x, y \in U$, 设 $(x, y) \in M_B$ 。如果 $(x, y) \in I_U$, 则由 L_B 的自反性可得 $(x, y) \in L_B$ 。如果 $P_B(x) \cap P_B(y) \neq \emptyset$ 并且对任意的 $b \in P_B(x) \cap P_B(y)$, 有 $b(x) = b(y)$, 则显然有 $(x, y) \in L_B$ 。于是有 $M_B \subseteq L_B$ 。

对于任意 $x, y \in U$, 设 $(x, y) \in L_B$ 。如果对于任意 $b \in B$, 有 $b(x) = b(y) = *$, 则有 $\forall b \in B (b(x) = *)$, 从而 $(x, y) \in C_B$ 。如果 $P_B(x) \cap P_B(y) \neq \emptyset$ 并且对任意的 $b \in P_B(x) \cap P_B(y)$, 有 $b(x) = b(y)$, 则显然有 $(x, y) \in C_B$ 。故 $L_B \subseteq C_B$ 。

于是 $M_B \subseteq L_B \subseteq C_B \subseteq T_B$ 。

定理 3 设 $S=(U, A, V, f)$ 为信息系统, $B \subseteq A, X \subseteq U$ 。则有

$$(1) \underline{T}_B(X) \subseteq \underline{C}_B(X) \subseteq \underline{S}_B(X)$$

$$\overline{S}_B(X) \subseteq \overline{C}_B(X) \subseteq \overline{T}_B(X)$$

$$(2) \underline{T}_B(X) \subseteq \underline{C}_B(X) \subseteq \underline{L}_B(X) \subseteq \underline{M}_B(X)$$

$$\overline{M}_B(X) \subseteq \overline{L}_B(X) \subseteq \overline{C}_B(X) \subseteq \overline{T}_B(X)$$

证明:(1)对于任意 $x \in U$, 由 $C_B \subseteq T_B, C_B(x) = \{y \in U;$

$(x, y) \in C_B$, $T_B(x) = \{y \in U; (x, y) \in T_B\}$, 可得 $C_B(x) \subseteq T_B(x)$ 。若 $x \in \underline{T}_B(X)$, 则 $T_B(x) \subseteq X$, 从而 $C_B(x) \subseteq X$, 故有 $x \in \underline{C}_B(X)$ 。于是 $\underline{T}_B(X) \subseteq \underline{C}_B(X)$ 。

又因 $S_B \subseteq C_B$, $S_B(x) = \{y \in U; (x, y) \in S_B\}$, 故 $S_B(x) \subseteq C_B(x)$ 。若 $x \in \underline{C}_B(X)$, 则 $C_B(x) \subseteq X$, 从而 $S_B(x) \subseteq X$, 故有 $x \in \underline{S}_B(X)$ 。于是 $\underline{C}_B(X) \subseteq \underline{S}_B(X)$ 。

综上, $\underline{T}_B(X) \subseteq \underline{C}_B(X) \subseteq \underline{S}_B(X)$ 。

又因为 $\overline{T}_B(X) = \bigcup_{x \in X} T_B(x)$, $\overline{C}_B(X) = \bigcup_{x \in X} C_B^{-1}(x)$, $\overline{S}_B(X) = \bigcup_{x \in X} S_B^{-1}(x)$, $\overline{L}_B(X) = \bigcup_{x \in X} L_B(x)$, $S_B^{-1}(x) \subseteq C_B^{-1}(x) \subseteq T_B(x)$, 所以 $\overline{S}_B(X) \subseteq \overline{C}_B(X) \subseteq \overline{T}_B(X)$ 成立。

(2) 同理可证。

推论 1 若 $\forall x \in U, P_B(x) \neq \emptyset$, 则 $L_B = C_B = M_B$ 。

例 3 考虑例 1 中的不完备决策表, 根据本文提出的限制相似关系, 相应的限制相似类为:

$$\begin{aligned} M_B(c_1) &= \{c_1, c_3, c_4, c_8, c_9\}, M_B(c_2) = \{c_2, c_3, c_4\} \\ M_B(c_3) &= \{c_1, c_2, c_3, c_4\}, M_B(c_4) = \{c_1, c_2, c_3, c_4\} \\ M_B(c_5) &= \{c_5, c_6\}, M_B(c_6) = \{c_5, c_6\}, M_B(c_7) = \{c_7, c_8, c_9\} \\ M_B(c_8) &= M_B(c_9) = \{c_1, c_7, c_8, c_9\}, M_B(c_{10}) = \{c_{10}\} \\ M_B(c_{11}) &= \{c_{11}\} \\ \overline{M}_B(\Phi) &= \{c_1, c_2, c_3, c_4, c_7, c_8, c_9, c_{10}\} \\ \underline{M}_B(\Phi) &= \{c_2, c_3, c_4, c_{10}\} \\ \overline{M}_B(\Psi) &= \{c_1, c_5, c_6, c_7, c_8, c_9, c_{11}\} \\ \underline{M}_B(\Psi) &= \{c_5, c_6, c_{11}\} \end{aligned}$$

结束语 信息系统知识约简与规则获取是粗糙集理论的重要研究方向。针对不完备信息系统, 人们提出了多种刻画对象相似性的不可区分关系, 如容差关系、非对称相似关系、限制容差关系等。本文对以上不可区分关系进行了对比分析, 在此基础上提出了限制相似关系, 讨论了限制相似关系以及相应的粗糙集模型中近似算子的性质。借助本文的结果可以进一步研究基于限制相似关系的信息系统知识约简方法。

参 考 文 献

[1] Pawlak Z. Rough set [J]. International Journal of Computer and Information Science, 1982, 11, 341-356

(上接第 231 页)

[12] Bacon D F, Graham S L, Shap O J. Compiler Transformations for High-Performance Computing [J]. ACM Computing Surveys, 1994, 26(4): 345-420

[13] Alexander M J, Bailey M W, Childers B R, et al. Memory bandwidth optimizations for wide-bus machines[C]//Proceedings of the 26th Hawaii International Conference on System Sciences, Wailea, Hawaii, 1993: 466-475

[14] Mowry T C. Tolerating Latency Through Software-Controlled Data Prefetching[D]. Stanford University, March 1994

[15] Fog A. Optimizing subroutines in assembly language [D]. Copenhagen University College of Engineering, September 2012

[16] Ma Y, Carr S. Register Pressure Guided Unroll-and-Jam[M]//The 2008 Open64 Workshop. Boston, MA, USA, April 6, 2008

[17] Carr S, Guan Y. Unroll and Jam using Uniformly Generated Sets [C]//Proceedings of the 30th Annual International Symposium on Microarchitecture (MICRO-30), 1997: 349-357

[2] Kryszkiewicz M. Rough set approach to incomplete information systems [J]. Information Sciences, 1998, 112: 39-49

[3] Slowinski R, Stefanowski J. Rough classification in incomplete information systems [J]. Computing Modeling, 1989, 12 (10/11): 1347-1357

[4] Wang G Y. Extension of rough set under incomplete information systems [J]. Journal of Computer Research and Development, 2002, 39(10): 1238-1243 (in Chinese)

王国胤. 粗糙集理论在不完备信息系统中的扩充 [J]. 计算机研究与发展, 2002, 39(10): 1238-1243

[5] Yin X R, Shang L. Expansion of Rough set model under incomplete information system [J]. Journal of Nanjing University, 2006, 42(4): 337-341 (in Chinese)

尹旭日, 商琳. 不完备信息系统中 Rough 集的扩充模型 [J]. 南京大学学报, 2006, 42(4): 337-341

[6] Wang G Y. Rough Set Theory and Knowledge Discovery [M]. Xi'an: Xi'an Jiaotong University Press, 2001 (in Chinese)

王国胤. 粗糙集理论与知识获取 [M]. 西安: 西安交通大学出版社, 2001

[7] Stefanowski J, Tsoukias A. On the extension of rough sets under incomplete information [C]//Zhong N, Skowron A, Ohsuga S, eds. Proc of the 7th Int'l Workshop on New Directions in Rough Sets, Data Mining, and Granular-Soft Computing. Berlin: Springer-Verlag, 1999: 73-81

[8] Guan L H. Incomplete information processing method based on Rough sets [D]. Chengdu: Southwest Jiaotong University, 2012 (in Chinese)

官礼和. 基于 Rough 集的不完备信息处理方法研究 [D]. 成都: 西南交通大学, 2012

[9] Yao Y Y. Relational interpretations of neighborhood operators and rough set approximation operators [J]. Information Sciences, 1998, 101: 239-259

[10] Qin K Y, Zhao H, Pei Z. The reduction of decision table based on generalized indiscernibility relation [J]. Journal of Xihua University, 2013, 32(4): 1-4 (in Chinese)

秦克云, 赵华, 裴峥. 基于广义不可区分关系的决策表约简 [J]. 西华大学学报, 2013, 32(4): 1-4

[18] Carr S, Kennedy K. Improving the Ratio of Memory Operations to Floating-Point Operations in Loops [J]. ACM Transactions on Programming Languages and Systems (ToPLAS), 1994, 16(6): 1768-1810

[19] Mark S, Saman A. Predicting Unroll Factors Using Supervised Classification [C]//Proceedings of the International Symposium on Code Generation and Optimization (CGO), 2003: 204-215

[20] Monsifrot A, Bodin F, Quiniou R. A Machine Learning Approach to Automatic Production of Compiler Heuristics [M]//Artificial Intelligence: Methodology, Systems, Applications, 2002: 41-50

[21] Mowry T C, Lam M S, Gupta A. Design and evaluation of a compiler algorithm for prefetching [C]//Proceeding of the Fifth International Conference on Architectural Support for Programming Languages and Operating Systems, Massachusetts: ACM Press, 1992: 62-73