

Rough 集之间的相似度量^{*}

徐久成^{1,2} 沈钧毅¹ 王国胤³

(西安交通大学电信学院软件所 西安 710049)¹ (河南师范大学计算机科学与技术系 新乡 453002)²

(重庆邮电学院计算机科学与技术研究所 重庆 400065)³

Measure of Similarity between Rough Sets

XU Jiu-Cheng^{1,2} SHEN Jun-Yi¹ WANG Guo-Yin³

(College of Electron & Communication, Xi'an Jiaotong University, Xi'an 710049)¹

(Department of Computer Science & Technology, Henan Normal University, Xinxing 453002)²

(Institute of Computer Science & Technology, Chongqing University of Posts & Telecommunications, Chongqing 400065)³

Abstract Applications of rough set theory in incomplete information systems are a key of putting rough set into real applications. In this paper, after analyzing some basic concepts of classical rough set theory and extended rough set theory, the measure of similarity is developed between two rough sets in the classical rough set theory based on indiscernibility relation and between two rough sets in the extended rough set theory based on limited tolerance relation. Then, some properties of these two methods for measuring similarity are developed respectively. At last, these two measure methods of rough set theory are compared.

Keywords Rough set, Indiscernibility relation, Limited tolerance relation, Measure of similarity

1 引言

近年来由 Z. Pawlak 教授提出的经典 Rough 集理论在边界不确定信息处理和知识获取方面取得了很大成功。但遗憾的是,它是基于完备信息系统这样一个假设的,即所处理的信息表都是完备的,每个样本对象的所有属性值都是已知的。而在现实生活中,由于数据测量的误差、对数据理解或获取的条件限制等原因,使得在知识获取时往往面临的是不完备信息系统,即可能存在部分对象的一些属性值未知的情况。基于传统不分明关系的 Rough 集理论是不能处理不完备信息系统的,这就需要对传统不分明关系的 Rough 集理论进行扩充,把不分明关系扩充到广义 Rough 集理论下讨论,一般是通过弱化二元关系的一些属性得来的,如自反性、对称性、传递性^[1-3]。对传统不分明关系的 Rough 集理论进行扩充,一般有三种扩充:一是基于容差关系的扩充;二是基于非对称相似关系的扩充;三是基于量化容差关系的扩充^[6]。这三种扩充关系都有一定的局限性,王国胤教授又提出了一种限制容差关系^[4],限制容差关系具有自反性、对称性,但不具有传递性。限制容差关系吸取了容差关系和非对称相似关系的优点,丢弃了二者的缺陷(极端情况),更加符合客观实际。

在智能系统的推理过程中,经常需要将两个知识模式(谓词公式、框架片断、模糊断言等)进行比较和耦合,即检查这两个知识模式是否完全一致或近似一致。如果两者完全一致,或者虽不完全一致,但两者间的相似程度落在限定的阈值范围内,就称这两个知识模式是匹配的,否则称为不匹配。

在不确定性推理中,知识和证据都具有某种程度的不确

定性,这就为推理机的设计与实现增加了复杂性和难度。它除了必须解决推理方向、推理方法、控制策略等基本问题外,还需要解决不确定性的表示与度量、不确定性匹配、不确定性的传递算法等重要问题,这就需要涉及有关知识和证据之间的相似度量等问题。不确定性集合之间相似性度量是边界不确定信息处理的一个重要概念,是进行不确定性知识获取和不确定性信息处理的基础^[5,7,8]。在不确定性推理中,不确定性集合之间的相似性度量是推理的重要基础。本文所提出的 Rough 集之间的相似度量,可以进一步用于不确定性推理中。比如在基于模糊产生式规则下的不确定性推理中,需要考虑产生式中前提条件和证据间的匹配问题,该问题最终转化为对其前提条件和证据所对应的不确定集合之间的相似程度的计算。为了度量两个 Rough 集之间的相似程度,本文首先分析研究了经典 Rough 集与广义 Rough 集的一些基本理论及其联系和共性,然后分别提出了基于不分明关系的经典 Rough 集之间和基于限制容差关系的广义 Rough 集之间的相似性度量方法,并分别讨论了这两种度量方法所满足的性质。最后对这两种度量方法进行了对比分析。由于该方法具有较好的性质,更具合理性,因而是一种理想的度量 Rough 集之间相似程度的度量方法。

2 Rough 集与限制容差关系

为了便于本文后续部分的讨论,在此先对经典 Rough 集和广义 Rough 集理论的一些有关基本概念进行简单介绍。

设 U 是一个非空有限论域, R 是 U 上的二元不分明关系, $K = (U, R)$ 称为近似空间, $\forall (x, y) \in U \times U$, 若 $(x, y) \in$

^{*} 本文得到国家自然科学基金(69803014)、国家科技部攀登——特别支持费、教育部骨干教师资助计划(GG-520-10617-1001)、教育部留学回国人员科研启动基金、重庆市应用基础研究基金、信息产业部计算机网络与通信技术重点实验室研究基金资助。徐久成 博士研究生,副教授,主要研究方向: Rough 集与不确定信息的处理。沈钧毅 教授,博导,主要研究方向: 数据库与软件集成, 数据仓库, 数据挖掘。王国胤 博士,教授,主要研究方向: 神经网络, Rough 集, 知识获取, 智能信息系统等。

R , 则称元素 x 与 y 在 K 中是不可分辨的。 $U|R$ 是 U 上由 R 生成的全体不分明类, 它构成 U 上的一个划分。

定义 1^[6] 给定完备知识表达系统 $S = \langle U, R, V, f \rangle$, 对于每个子集 $X \subseteq U$ 和不分明关系 $B (B \subseteq R)$, X 的上近似集和下近似集分别可以由 B 的基本集定义如下:

$$B^+(X) = \bigcup \{Y_i | (Y_i \in U | IND(B) \wedge Y_i \cap X \neq \emptyset)\},$$

$$B_-(X) = \bigcup \{Y_i | (Y_i \in U | IND(B) \wedge Y_i \subseteq X)\}.$$

其中, $U | IND(B) = \{X | (X \subseteq U \wedge \forall x \forall y (x \in X \wedge y \in X \wedge b \in B \rightarrow b(x) = b(y)))\}$ 是不分明关系 B 对 U 的划分。

Rough 集的上、下近似集也可通过集合元素来定义:

$$B^+(X) = \{x | (x \in U \wedge [x]_B \cap X \neq \emptyset)\},$$

$$B_-(X) = \{x | x \in U \wedge [x]_B \subseteq X\}.$$

在经典 Rough 集中, 定义经典 Rough 集意义下的集合 X 的 Rough 隶属函数为:

$$\mu_X^R(x) = Card(X \cap I(x)) / Card(I(x)), \quad (1)$$

其中, $x \in U$, $I(x)$ (即 $[x]_B$) 是包含 x 的不分明类, $Card(M)$ 表示集合 M 的基数, $0 \leq \mu_X^R(x) \leq 1$ 。

给定不完备信息系统 $NS = \langle U, A, V, f \rangle$, 其中, $A = C \cup \{d\}$, C 是条件属性集合, d 是决策属性, 对于具有遗漏属性值的属性子集 $B \subseteq A$, 记遗漏值为“*”, 我们引入不完备信息系统 NS 中一些关系的定义。

定义 2^[6] 容差关系 T 的定义为:

$$\forall x, y \in U (T_B(x, y) \Leftrightarrow \forall c_j \in B (c_j(x) = c_j(y) \vee c_j(x) = * \vee c_j(y) = *)).$$

显然, 容差关系 T 具有自反性、对称性, 但不一定具有传递性。

定义 3^[6] 非对称相似关系 S 的定义为:

$$\forall x, y \in U (S_B(x, y) \Leftrightarrow \forall c_j \in B (c_j(x) = c_j(y) \vee c_j(x) = *)).$$

显然, 非对称相似关系 S 不具有对称性, 但具有自反性、传递性。

定义 4^[4] 限制容差关系 L 的定义为: 设 $P_B(x) = \{b | b \in B \wedge b(x) \neq *\}$, 则

$$\forall x, y \in U (L_B(x, y) \Leftrightarrow \forall b \in B (b(x) = b(y) = *) \vee ((P_B(x) \cap P_B(y) \neq \emptyset) \wedge \forall b \in B ((b(x) \neq *) \wedge (b(y) \neq *) \rightarrow (b(x) = b(y)))))).$$

显然, 限制容差关系 L 具有自反性、对称性, 但不具有传递性。

定义 5^[4] 限制容差类 $J_B^L(x) = \{y | y \in U \wedge L_B(x, y)\}$, $D \subseteq U$, 集合 D 相应的上、下近似集定义为:

$$D_L^L = \{x | x \in U \wedge J_B^L(x) \cap D \neq \emptyset\},$$

$$D_B^L = \{x | x \in U \wedge J_B^L(x) \subseteq D\}$$

以后为书写方便起见, 把限制容差 $J_B^L(x)$ 类简写为 $J(x)$ 。

定义 6 在广义 Rough 集中, 基于限制容差关系 L 意义下的集合 X 的 Rough 隶属函数为:

$$\mu_X^L(x) = Card(X \cap J(x)) / Card(J(x)) \quad (2)$$

其中, $x \in U$, $J(x)$ 是包含 x 的限制容差类, $Card(M)$ 表示集合 M 的基数, $0 \leq \mu_X^L(x) \leq 1$ 。

经典 Rough 集是广义 Rough 集在完备信息系统情况下的特例。基于不分明关系的经典 Rough 集的上、下近似集定义与基于限制容差关系的广义 Rough 集的上、下近似集定义类同, 前者用不分明关系的不分明类来定义, 后者用限制容差关系的限制容差类来定义。经典 Rough 集按某种不分明关系

把 U 划分成若干个不分明类, 不分明类间互不相交; 与经典 Rough 集的不分明关系相比, 广义 Rough 集中按某种扩充关系 (如限制容差关系) 不能对 U 进行划分, 而只能分成若干个具有相应关系的相似类 (如限制容差类), 相似类间互相交叠。

3 基于不分明关系的经典 Rough 集之间的相似度量及性质

在经典 Rough 集中, 基于不分明关系, 我们提出了集合间的相似性度量方法, 并给出了其相应性质:

设论域为 $U = \{u_1, u_2, \dots, u_n\}$, A 和 B 是论域 U 上的两个 Rough 集, $A, B \subseteq U$, $u_i \in U$, u_i 在某指定的不分明关系 I 下分别关于 A, B 的 Rough 隶属度分别为 $a_i = \mu_A^I(u_i)$ 和 $b_i = \mu_B^I(u_i)$ ($i = 1, 2, \dots, n$), 则 A 和 B 关于不分明关系 I 下的 Rough 隶属函数分别表示为 A', B' :

$$A' = \mu_A^I(u_1)/u_1 + \mu_A^I(u_2)/u_2 + \dots + \mu_A^I(u_n)/u_n, \quad (3)$$

$$B' = \mu_B^I(u_1)/u_1 + \mu_B^I(u_2)/u_2 + \dots + \mu_B^I(u_n)/u_n,$$

则集合 A 和 B 之间的相似程度可由下面的公式计算得到:

$$Sim_{D_I}(A, B) = \begin{cases} 1, & \text{if } A = B = \emptyset, \\ \frac{\sum_{i=1}^n \min\{a_i, b_i\}}{\sum_{i=1}^n \max\{a_i, b_i\}}, & \text{else} \end{cases} \quad (4)$$

若集合 A 和 B 之间的相似程度越大, 则 $Sim_{D_I}(A, B)$ 的值越大; 若 A 和 B 之间的相似程度越小, 则 $Sim_{D_I}(A, B)$ 的值越小。

例 1: 给定一个知识库 $K = \langle U, I \rangle$, 其中 $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$, I 是一个不分明关系, $U | I = \{E_1, E_2, E_3, E_4\}$, 其中:

$$E_1 = \{x_1, x_4, x_8\}, E_2 = \{x_2, x_5, x_7\},$$

$$E_3 = \{x_3\}, E_4 = \{x_6\}.$$

对于集合 $A = \{x_1, x_4, x_7\}$, $B = \{x_1, x_4\}$, $C = \{x_2, x_3\}$, $D = \{x_3, x_6\}$, A, B, C, D 在不分明关系 I 下的 Rough 隶属函数表示分别为 A', B', C', D' , 由 Rough 隶属函数 $\mu_x^I(x)$ (公式(1)) 经计算有:

$$A' = \frac{2}{3}/x_1 + \frac{1}{3}/x_2 + 0/x_3 + \frac{2}{3}/x_4 + \frac{1}{3}/x_5 + 0/x_6 + \frac{1}{3}/x_7 + \frac{2}{3}/x_8,$$

$$B' = \frac{2}{3}/x_1 + 0/x_2 + 0/x_3 + \frac{2}{3}/x_4 + 0/x_5 + 0/x_6 + 0/x_7 + \frac{2}{3}/x_8,$$

$$C' = \frac{1}{3}/x_1 + \frac{1}{3}/x_2 + 0/x_3 + \frac{1}{3}/x_4 + \frac{1}{3}/x_5 + 0/x_6 + \frac{1}{3}/x_7 + \frac{1}{3}/x_8,$$

$$D' = 0/x_1 + 0/x_2 + 1/x_3 + 0/x_4 + 0/x_5 + 1/x_6 + 0/x_7 + 0/x_8.$$

则集合 A 和 B 之间的相似度为:

$$Sim_{D_I}(A, B) = (\sum_{i=1}^8 \min\{a_i, b_i\}) / (\sum_{i=1}^8 \max\{a_i, b_i\}) = (2/3 + 2/3 + 2/3) / (2/3 + 1/3 + 2/3 + 1/3 + 1/3 + 2/3) \approx 0.67,$$

同理, $Sim_{D_I}(A, A) = 1$, $Sim_{D_I}(B, C) \approx 0.33$, $Sim_{D_I}(A, D) = 0$ 。

设论域为 $U = \{u_1, u_2, \dots, u_n\}$, A 和 B 是论域 U 上的两个 Rough 集, $A, B \subseteq U$, $u_i \in U$, u_i 在某指定的不分明关系 I 下分

别关于 A, B 的 Rough 隶属度分别为 $a_i = \mu_A^L(u_i)$ 和 $b_i = \mu_B^L(u_i)$ ($i=1, 2, \dots, n$), 上述经典 Rough 集理论下不确定集合相似性度量满足如下性质:

定理 1 $SimD_I(A, B) \in [0, 1]$.

证明: 由公式(4)显然.

定理 2 对经典 Rough 集来说, $SimD_I(A, B)$ 满足:

- (1) 自反性: $SimD_I(A, A) = 1$;
- (2) 对称性: $SimD_I(A, B) = SimD_I(B, A)$;

证明: 自反性和对称性由公式(4)直接可证.

公式(4)不满足传递性.

定理 3 $SimD_I(A, B) = 0$, 当且仅当对 $\forall u_i (i=1, 2, \dots, n)$, $\mu_A^L(u_i)$ 和 $\mu_B^L(u_i)$ 至少有一个为 0, 且集合 A 与 B 不能同时为空集.

证明: (必要性) 若 $SimD_I(A, B) = 0$, 则根据公式(4)知集合 A 与 B 不能同时为空集, 且 $(\sum_{i=1}^n \min\{a_i, b_i\}) / (\sum_{i=1}^n \max\{a_i, b_i\}) = 0$. 因为集合 A 与 B 不能同时为空集, 从而有 $(\sum_{i=1}^n \max\{a_i, b_i\}) \neq 0$, 因此 $(\sum_{i=1}^n \min\{a_i, b_i\}) = 0$, 因而对 $\forall i (i=1, 2, \dots, n)$ 有 $\min\{a_i, b_i\} = 0$, 即对 $\forall u_i (i=1, 2, \dots, n)$, $\mu_A^L(u_i)$ 和 $\mu_B^L(u_i)$ 至少有一个为 0.

(充分性) 若对 $\forall u_i (i=1, 2, \dots, n)$, $\mu_A^L(u_i)$ 和 $\mu_B^L(u_i)$ 至少有一个为 0. 则 $(\sum_{i=1}^n \min\{a_i, b_i\}) = 0$, 又集合 A 与 B 不能同时为空集, 则 $(\sum_{i=1}^n \max\{a_i, b_i\}) \neq 0$, 从而有 $SimD_I(A, B) = 0$.

4 基于限制容差关系的广义 Rough 集之间的相似度量及性质

在不完备信息系统中, 基于限制容差关系, 我们提出了广义 Rough 集间的相似性度量方法, 并给出了其相应性质:

设论域为 $U = \{u_1, u_2, \dots, u_n\}$, A 和 B 是论域 U 上的两个广义 Rough 集, $A, B \subseteq U$, $u_i \in U$, 在某指定的限制容差关系 L 下分别关于 A, B 的广义 Rough 隶属度分别为 $a_i = \mu_A^L(u_i)$ 和 $b_i = \mu_B^L(u_i)$ ($i=1, 2, \dots, n$), 则 A 和 B 关于某指定的限制容差关系 L 下 Rough 隶属函数表示分别为 A', B' :

$$\begin{aligned} A' &= \mu_A^L(u_1)/u_1 + \mu_A^L(u_2)/u_2 + \dots + \mu_A^L(u_n)/u_n, \\ B' &= \mu_B^L(u_1)/u_1 + \mu_B^L(u_2)/u_2 + \dots + \mu_B^L(u_n)/u_n, \end{aligned} \quad (5)$$

则集合 A 和 B 之间的相似程度可由下面的公式计算得到:

$$GSimD_L(A, B) = \begin{cases} 1, & \text{if } A=B=\Phi, \\ \frac{\sum_{i=1}^n \min\{a_i, b_i\}}{\sum_{i=1}^n \max\{a_i, b_i\}}, & \text{else} \end{cases} \quad (6)$$

若集合 A 和 B 之间的相似程度越大, 则 $GSimD_L(A, B)$ 的值越大; 若 A 和 B 之间的相似程度越小, 则 $GSimD_L(A, B)$ 的值越小.

表 1 不完备信息表

A	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
c_1	3	2	2	*	*	2	3	*	3	1	*	3
c_2	2	3	3	2	2	3	*	0	2	*	2	2
c_3	1	2	2	*	*	2	*	0	1	*	*	1
c_4	0	0	0	1	1	1	3	*	3	*	*	*
d	Ω	Ω	Ψ	Ω	Ψ	Ψ	Ω	Ψ	Ψ	Ω	Ψ	Ω

例 2: 表 1 是在讨论不完备信息系统中经常用到的一个典型不完备信息表^[4,6], 其中 $U = \{x_1, x_2, \dots, x_{12}\}$ 是对象集合, c_1, c_2, c_3, c_4 是条件属性, 值域均为 $[0, 1, 2, 3]$, d 是决策属性, 将对象分为 Ψ 和 Ω 两个决策集合(决策表). L 是一个限制容差关系(即在定义 4 中, 令 $B = \{c_1, c_2, c_3, c_4\}$ 即可), $J(x)$ 是包含对象 x 的限制容差类. 其中,

$$\begin{aligned} J(x_1) &= \{x_1, x_{11}, x_{12}\}, J(x_2) = \{x_2, x_3\}, J(x_3) = \{x_2, x_3\}, \\ J(x_4) &= \{x_4, x_5, x_{11}, x_{12}\}, J(x_5) = \{x_4, x_5, x_{11}, x_{12}\}, \\ J(x_6) &= \{x_6\}, \\ J(x_7) &= \{x_7, x_9, x_{12}\}, J(x_8) = \{x_8\}, \\ J(x_9) &= \{x_7, x_9, x_{11}, x_{12}\}, \\ J(x_{10}) &= \{x_{10}\}, J(x_{11}) = \{x_1, x_4, x_5, x_9, x_{11}, x_{12}\}, \\ J(x_{12}) &= \{x_1, x_4, x_5, x_7, x_9, x_{11}, x_{12}\} \end{aligned}$$

对于集合 $A = \{x_1, x_2, x_7, x_9, x_{12}\}$, $C = \{x_1, x_2, x_7, x_9, x_{11}\}$, $D = \{x_1, x_2, x_7, x_8, x_{11}\}$, $G = \{x_1, x_2, x_7, x_8\}$, A, C, D, G 在限制容差关系 L 下的广义 Rough 隶属函数表示分别为 A', C', D', G' , 由广义 Rough 隶属函数公式(2)经计算有:

$$\begin{aligned} A' &= \frac{2}{3}/x_1 + \frac{1}{2}/x_2 + \frac{1}{2}/x_3 + \frac{1}{4}/x_4 + \frac{1}{4}/x_5 + 0/x_6 + 1/x_7 + 0/x_8 + \frac{3}{4}/x_9 + 0/x_{10} + \frac{1}{2}/x_{11} + \frac{4}{7}/x_{12} \\ C' &= \frac{2}{3}/x_1 + \frac{1}{2}/x_2 + \frac{1}{2}/x_3 + \frac{1}{4}/x_4 + \frac{1}{4}/x_5 + 0/x_6 + \frac{2}{3}/x_7 + 0/x_8 + \frac{3}{4}/x_9 + 0/x_{10} + \frac{1}{2}/x_{11} + \frac{4}{7}/x_{12} \\ D' &= \frac{2}{3}/x_1 + \frac{1}{2}/x_2 + \frac{1}{2}/x_3 + \frac{1}{4}/x_4 + \frac{1}{4}/x_5 + 0/x_6 + \frac{1}{3}/x_7 + 1/x_8 + \frac{1}{2}/x_9 + 0/x_{10} + \frac{1}{3}/x_{11} + \frac{3}{7}/x_{12} \\ G' &= \frac{1}{3}/x_1 + \frac{1}{2}/x_2 + \frac{1}{2}/x_3 + 0/x_4 + 0/x_5 + 0/x_6 + \frac{1}{3}/x_7 + 1/x_8 + \frac{1}{4}/x_9 + 0/x_{10} + \frac{1}{6}/x_{11} + \frac{2}{7}/x_{12} \end{aligned}$$

则集合 A 和 C 之间的广义相似度为:

$$GSimD_L(A, C) = (2/3 + 1/2 + 1/2 + 1/4 + 1/4 + 2/3 + 3/4 + 1/2 + 4/7) / (2/3 + 1/2 + 1/2 + 1/4 + 1/4 + 1 + 3/4 + 1/2 + 4/7) \approx 0.93,$$

同理, $GSimD_L(A, A) = 1$, $GSimD_L(A, D) \approx 0.63$,

$$GSimD_L(A, G) \approx 0.40.$$

设 A 和 B 是有限非空论域 $U = \{u_1, u_2, \dots, u_n\}$ 上的两个广义 Rough 集合, 则在广义 Rough 集理论中基于限制容差关系 L 下的不确定集合相似性度量满足如下的性质:

定理 4 $GSimD_L(A, B) \in [0, 1]$.

定理 5 对基于限制容差关系 L 下的广义 Rough 集来说, $GSimD_L(A, B)$ 满足:

- (1) 自反性: $GSimD_L(A, A) = 1$;
- (2) 对称性: $GSimD_L(A, B) = GSimD_L(B, A)$;

公式(6)不满足传递性.

定理 6 $GSimD_L(A, B) = 0$, 当且仅当对 $\forall u_i (i=1, 2, \dots, n)$, $\mu_A^L(u_i)$ 和 $\mu_B^L(u_i)$ 至少有一个为 0 且集合 A 与 B 不能同时为空集.

定理 4~定理 6 的证明类同定理 1~定理 3, 这里不再赘述.

5 两种集合相似度量方法的联系

上面我们分别讨论了基于不分明关系的经典 Rough 集

(下转第 60 页)

$= \{ \frac{0.64}{\text{排球}}, \frac{0.28}{\text{游泳}}, \frac{0.24}{\text{举重}} \}, (r=0.71)$ 。

(6) 若晴天天凉则适宜于三种运动的可能性分布 $d_6 = \{ \frac{0.00}{\text{排球}}, \frac{0.30}{\text{游泳}}, \frac{0.70}{\text{举重}} \}, (r=1.00)$ 。

(7) 若天凉阴天则适宜于三种运动的可能性分布 $d_7 = \{ \frac{0.38}{\text{排球}}, \frac{0.01}{\text{游泳}}, \frac{0.61}{\text{举重}} \}, (r=0.76)$ 。

(8) 若天凉下雨则适宜于三种运动的可能性分布 $d_8 = \{ \frac{0.02}{\text{排球}}, \frac{0.00}{\text{游泳}}, \frac{0.98}{\text{举重}} \}, (r=0.97)$ 。

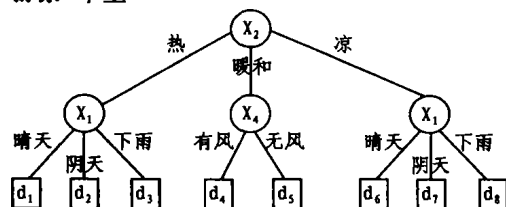


图1 $d_i, i=1, \dots, 8$ 为对应终端结点的可能性分布 (见规则的结论)

在可信度大于0.7的条件下,上述各条模糊规则同湿度没有关系。假设已知一个对象属于属性值的隶属度为(0.0, 1.0, 0.0; 0.5, 0.5, 0.0; 0.7, 0.3; 1.0, 0.0)。则该对象对8条规则的条件的满足程度分别为0.0, 0.5, 0.0, 0.5, 0.0, 0.0, 0.0, 0.0。由于满足第二条规则和第四条规则的程度都达到最大,大多

数模糊决策树都是在适宜于游泳和适宜于举重中任选一个,作为该对象所属的类,这带有很大的不确定性。若假定输出对应的是可能性分布,利用简单模糊推理可以得到该对象的可能性分布为 $\{ \frac{0.30}{\text{排球}}, \frac{0.30}{\text{游泳}}, \frac{0.44}{\text{举重}} \}$ 。此时适宜于三种运动的程度相差不远,而适宜于举重的程度略大。

结论 本文在假设一个对象可以属于多个类的情况下,给出了模糊决策树的算法,所给算法得出的模糊决策树的终端结点赋给的是一个可能性分布。简单模糊推理被用于确定一个属性值已知对象所属类的可能性分布。本文所得的模糊树能够避免在确定属性值已知的对象时所属类出现不一致的情况,对所建的模糊决策树提取的规则解释更加自然。

参考文献

- Janikon C Z. Fuzzy Decision Trees: Issues and Methods. IEEE Trans. Syst., Man, Cybern., B, 1988, 28: 1~14
- Yuan Y, Shaw M J. Induction of Fuzzy decision Trees. Fuzzy sets and Systems, 1995, 69: 125~139
- Quinlan J R. Induction of Decision Trees. Machine learning, 1986, 1: 71~99
- Quinlan J R. C4. 5, Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann, 1993
- Chang I, Hsu J. Fuzzy Classification Trees for Data Analysis. Fuzzy Sets and Systems, 2002, 130: 87~99
- Myung L K, Mi K L, Hyong L J, Kwang H L. A Fuzzy Decision Tree Induction Method for Fuzzy Data. In: Proc. IEEE Int. Fuzzy Systems Conf. 1999. 1~16
- 张文修, 梁广锡. 模糊控制与系统. 西安交通大学出版社, 1998

(上接第57页)

和基于限制容差关系的广义 Rough 集下集合之间相似性度量的方法和性质,实际上这两种度量方法是有联系的,而且是统一的。下面将对这两种集合相似度量方法进行一些比较。

(1) 经典 Rough 集和广义 Rough 集中集合之间相似性度量的方法具有一致性,可以为它们建立一个统一模型。这两种度量方法实际上就是一种度量方法,即 Rough 集之间的相似性度量方法,它可用于不确定推理和不确定信息的处理。只是当信息系统完备时,用经典 Rough 集之间的相似性度量方法;当信息系统不完备时,用广义 Rough 集之间的相似性度量方法。

(2) 经典 Rough 集和广义 Rough 集中集合之间相似性度量方法满足一些共同的性质。如经典 Rough 集中定理1~3分别与广义 Rough 集中定理4~6相对应。

(3) 经典 Rough 集下集合之间相似性度量是基于完备信息系统下的不分明关系的;而广义 Rough 集下集合之间相似性度量是基于不完备信息系统下的限制容差关系的。

(4) 不管是经典 Rough 集还是广义 Rough 集之间的相似程度,可根据实际要求确定一个阈值,当集合之间的相似程度落在限定的阈值范围内时,就认为两个集合之间具有相似性。

(5) 经典 Rough 集中集合之间相似性度量方法是广义 Rough 集中集合之间相似性度量方法的特殊情况。当基于限制容差关系的广义 Rough 集中的信息从不完备状态变为完备状态时,则基于限制容差关系的广义 Rough 集下集合之间相似性度量的方法即为基于不分明关系的经典 Rough 集下集合之间相似性度量的方法。

(6) 除基于限制容差关系的广义 Rough 集之外,对于基于容差关系、非对称相似关系的广义 Rough 集,也可以定义类似的广义相似性度量,限于篇幅,不再详细讨论,有兴趣的读者可以自行设计。

结论 Rough 集理论在不完备信息系统中的应用,是将 Rough 集理论进一步推向实用的关键之一,因为现实需要处理的数据基本都是在一定程度上是不完备的,所以就有必要研究能够直接处理不完备数据处理的方法。本文在分析研究经典 Rough 集与广义 Rough 集的一些基本概念理论的基础上,分别提出并讨论了基于不分明关系的经典 Rough 集之间的相似度量及性质和基于限制容差关系的广义 Rough 集之间的相似度量及性质。最后对它们的度量方法进行了一些比较。对于本文提出的度量方法,经过仿真实验,效果良好,可用于完备信息系统或不完备信息系统中 Rough 集之间的相似度量,从而更好地应用于不确定信息的数据处理。本文所提出的 Rough 集之间的相似度量,也可用于不确定性推理中。基于不确定集合相似度量的匹配与推理系统,是我们下一步的研究工作。

参考文献

- Kryszkiewicz M. Rough set approach to incomplete information systems. Information Sciences, 1998, 112: 39~49
- Liu Qing. λ -Level Rough equality relation and the inference of rough paramodulation. In: Proc. of Intl. Workshop on Rough Set Theory and Granular Computing (RSTGC-2001), 2001, 5: 424~431
- 江娟, 刘清. 基于程度不分明关系的广义 Rough 集定义. 计算机科学, 2001, 28(5. 专刊): 8~9
- 王国胤. Rough 集理论在不完备信息系统中的扩充. 计算机研究与发展, 2002, 39(10)
- 李凡, 徐章艳. Vague 集之间的相似度量. 软件学报, 2001, 12(6): 922~927
- 王国胤. Rough 集理论与知识获取. 西安: 西安交通大学出版社, 2001
- Chen S M. Measures of similarity between vague sets. Fuzzy Sets and Systems, 1995, 74(2): 217~223
- Hong D H, Kim C. A note on similarity measures between vague sets and between elements. Information Sciences, 1999, 115(1): 83~96