

概率粗糙集模型

Probabilistic Rough Set Models

王基一 许黎明

(浙江师范大学计算机学院 浙江金华321004)

Abstract In this paper, rough set models based on a probabilistic measure are studied. The notions of lower and upper approximations based on a probabilistic approximation space are derived and properties of the approximation operators are discussed. The relationship between the probabilistic rough set model and Pawlak's rough set model is compared. At last, the notions of approximation accuracy, attribute dependence, and reduction of attribute related to this model are defined.

Keywords Approximation space, Approximation operators, Reduction, Rough set

1. 引言

原始的RS模型(常称为Pawlak RS模型)^[1,2]是建立在二元等价关系的基础上的,但由于实际问题的需要,Pawlak RS模型的应用受到了限制,因此人们将二元等价关系推广成一般的二元关系,得到了一般关系下的RS模型^[3],Yao还在文[4]中讨论了基于邻域算子的RS模型。另一方面Pawlak粗糙集模型是基于可利用信息的完全性的,因而忽视了可利用信息的不完全性和可能存在的统计和随机信息,这类模型对于不协调的决策表的规则提取往往显得无能为力。本文我们从概率论的观点出发来研究粗糙集理论,为研究不确定信息系统提供了新的粗糙集模型。

2. Pawlak 粗糙近似算子

定义1 设 U 是非空有限集,称为论域, $R \subseteq U \times U$ 是 U 上二元等价关系,称 $A = (U, R)$ 为Pawlak近似空间,常表示为可利用的已知的信息库,记 $[x]_R$ 为对象 x 所在的 R -等价类。对于任意的 $X \subseteq U$, X 不一定能用近似空间的知识来精确地描述,于是就用 X 关于近似空间 $A = (U, R)$ 的一对下近似 $\underline{R}X$ 和上近似 $\overline{R}X$ 来近似描述

$$\underline{R}X = \{x \in U | [x]_R \subseteq X\},$$

$$\overline{R}X = \{x \in U | [x]_R \cap X \neq \emptyset\}.$$

$\underline{R}X$ 解释为根据现有知识判断肯定属于 X 的那些对象的全体, $\overline{R}X$ 解释为根据现有知识判断可能属于 X 的那些对象的全体。当 $\underline{R}X = \overline{R}X$ 时,称 X 关于近似空间 A 是可定义的,否则称 X 关于 A 是不可定义的,这时称 X 为粗糙集(rough set)。

由于 R 是等价关系,因此 $\underline{R}X$ 和 $\overline{R}X$ 又等价于下列定义形式:

$$\underline{R}X = \bigcup \{[x]_R | [x]_R \subseteq X\},$$

$$\overline{R}X = \bigcup \{[x]_R | [x]_R \cap X \neq \emptyset\}.$$

X 关于近似空间 (U, R) 的正域、边界和负域分别定义为

$$POS(X) = \underline{R}(X) = \{x \in U | [x]_R \subseteq X\}$$

$$BN(X) = \overline{R}(X) - \underline{R}(X)$$

$$NEG(X) = \sim \overline{R}(X) = \{x \in U | [x]_R \cap X = \emptyset\}$$

3. 信息系统与决策规则

定义2 称 $S = (U, A, F)$ 为一个信息系统,其中 U 为有限对象集(论域),即 $U = \{x_1, x_2, \dots, x_n\}$; A 为有限属性集,即

$A = \{a_1, a_2, \dots, a_m\}$; F 为 U 和 A 的关系集,即 $F = \{f_j | j \leq m\}$ 。

其中 $f_j: U \rightarrow V_j, (j \leq m), V_j$ 为 a_j 的值域, V_j 中的值可以是定量值也可以是定性值。 $F(x)$ 反映了对象 x 在系统中的完全信息,比如 $f_j(x) = v$ 表示对象 x 的属性 a_j 具有属性值 v ,因此 F 也称为信息函数。

对于这样的一个信息系统,每一个属性子集 $B \subseteq A$ 就定义了论域 U 上的一个二元等价关系 R_B :

$$xR_By \Leftrightarrow f_j(x) = f_j(y), \forall a_j \in B$$

有时属性集 A 还分为条件属性集 C 和决策属性集 D ,其中 $C \cap D = \emptyset$,这时的信息系统称为决策表,记为 $(U, C \cup D, F)$ 。

对于决策表 $(U, C \cup D, F)$,记由 R_C 生成的 U 上的划分为 $C' = \{C_1, C_2, \dots, C_l\}$;记由 R_D 生成的 U 上的划分为 $D' = \{D_1, D_2, \dots, D_j\}$,则当某一个 C_i 包含于某一个 D_j ,或等价地 $C_i \subseteq \overline{R_D}(D_j)$ 时,可以导出确定性的决策规则:

$$r: Desc(C_i) \Rightarrow Des_D(D_j)$$

其中 $Desc(C_i)$ 表示 C_i 中的对象的条件描述,即

$$Desc(C_i) = \bigwedge_{a \in C} (a, a(x)), x \in C_i$$

其中 $Des_D(D_j)$ 表示 D_j 中的对象的决策描述,即

$$Des_D(D_j) = \bigwedge_{d \in D} (d, d(x)), x \in D_j$$

当某一个 C_i 与某一个 D_j 的交非空,或等价地 $C_i \subseteq \overline{R_D}(D_j)$ 时,可以导出可能性的决策规则:

$$r: Desc(C_i) \xrightarrow{c} Des_D(D_j)$$

其中 c 为规则的可信度。

4. 概率粗糙近似算子的定义及其性质

定义3 设 U 是有限对象构成的论域, R 是 U 上的等价关系,其构成的等价类为 $U/R = \{X_1, X_2, \dots, X_n\}$,仍记 x 所在的等价类为 $[x]$,令 P 为定义在 U 的子集类构成的 σ 代数上的概率测度,三元组 $A_p = (U, R, P)$ 称为概率近似空间。 U 中的每个子集称为概念,它代表了一个随机事件, $P(X|Y)$ 表示事件 Y 发生下 X 出现的条件概率,也可解释为随机选择的对象在概念 Y 的描述下属于 X 的概率。设 $0 \leq \beta < \alpha \leq 1$,对于任意 $X \subseteq U$,我们定义 X 关于概率近似空间 $A_p = (U, R, P)$ 依参数 α, β 的下近似 $\underline{PA}_\alpha(X)$ 和上近似 $\overline{PA}_\beta(X)$ 如下:

$$\underline{PA}_\alpha(X) = \{x \in U | P(X|[x]_R) \geq \alpha\}$$

$$\overline{PA}_\beta(X) = \{x \in U | P(X|[x]_R) > \beta\}$$

$\underline{PA}_\alpha$ 与 $\overline{PA}_\beta$ 分别称为概率下近似算子和概率上近似算子。 X 关于 A_p 依参数 α, β 的概率型正域、边界和负域分别为

$$POS(X, \alpha) = \underline{PA}_\alpha(X) = \{x \in U | P(X|[x]_R) \geq \alpha\}$$

$$BN(X, \alpha, \beta) = \{x \in U | \beta < P(X|[x]_R) < \alpha\}$$

$$NEG(X, \beta) = U - \overline{PA}_\beta = \{x \in U | P(X|[x]_R) \leq \beta\}$$

显然,

$$\underline{PA}_\alpha(X) = POS(X, \alpha) = \bigcup \{[x]_R | P(X|[x]_R) \geq \alpha\}$$

$$\overline{PA}_\beta(X) = \bigcup \{[x]_R | P(X|[x]_R) > \beta\}$$

$$BN(X, \alpha, \beta) = \bigcup \{[x]_R | \beta < P(X|[x]_R) < \alpha\}$$

$$NEG(X, \beta) = \bigcup \{[x]_R | P(X|[x]_R) \leq \beta\}$$

X 关于 A_r 依参数 α, β 的正域, 边界和负域构成了论域 U 的划分. 而且显然有

$$\overline{PA}_\beta(X) = POS(X, \alpha) \cup BN(X, \alpha, \beta)$$

或者

$$BN(X, \alpha, \beta) = \overline{PA}_\beta(X) - POS(X, \alpha)$$

命题1 $\forall X \subseteq U, \underline{PA}_\alpha(\sim X) = NEG_{1-\alpha}(X)$

证明: $\underline{PA}_\alpha(\sim X) = \{x \in U | P(\sim X|[x]_R) \geq \alpha\}$

$$= \{x \in U | 1 - P(X|[x]_R) \geq \alpha\}$$

$$= \{x \in U | P(X|[x]_R) \leq 1 - \alpha\} = NEG_{1-\alpha}(X)$$

命题2 $\forall X \subseteq U, BN(X, \alpha, \beta) = BN(\sim X, 1 - \beta, 1 - \alpha)$

证明: $BN(X, \alpha, \beta) = \{x \in U | \alpha < P(X|[x]_R) < \beta\}$

$$= \{x \in U | \alpha < 1 - P(\sim X|[x]_R) < \beta\}$$

$$= \{x \in U | 1 - \beta < P(\sim X|[x]_R) < 1 - \alpha\}$$

$$= BN(\sim X, 1 - \beta, 1 - \alpha)$$

命题3 设 $0 \leq \beta < \alpha \leq 1, X, Y \subseteq U$, 则概率近似算子满足下列对偶性质:

$$(1) \underline{PA}_\alpha(\phi) = \overline{PA}_\alpha(\phi) = \phi, \underline{PA}_\alpha(U) = \overline{PA}_\alpha(U) = U;$$

$$(2) \underline{PA}_\alpha(X) \subseteq \overline{PA}_\beta(X);$$

$$(3) \underline{PA}_\alpha(X) = \sim \overline{PA}_{1-\alpha}(\sim X),$$

$$\overline{PA}_\beta(X) = \sim \underline{PA}_{1-\beta}(\sim X);$$

$$(4) \overline{PA}_\beta(X \cup Y) \supseteq \overline{PA}_\beta(X) \cup \overline{PA}_\beta(Y),$$

$$\underline{PA}_\alpha(X \cap Y) \subseteq \underline{PA}_\alpha(X) \cap \underline{PA}_\alpha(Y);$$

$$(5) \overline{PA}_\beta(X \cap Y) \subseteq \overline{PA}_\beta(X) \cap \overline{PA}_\beta(Y),$$

$$\underline{PA}_\alpha(X \cup Y) \supseteq \underline{PA}_\alpha(X) \cup \underline{PA}_\alpha(Y);$$

(6) 若 $X \subseteq Y$, 则

$$\underline{PA}_\alpha(X) \subseteq \underline{PA}_\alpha(Y), \overline{PA}_\beta(X) \subseteq \overline{PA}_\beta(Y);$$

(7) 若 $\alpha_1 \leq \alpha_2, \beta_1 \leq \beta_2$, 则

$$\underline{PA}_{\alpha_2}(X) \subseteq \underline{PA}_{\alpha_1}(X), \overline{PA}_{\beta_2}(X) \subseteq \overline{PA}_{\beta_1}(X)$$

由命题3的(7)可以看出概率粗糙集的正域随着 α 的减少而增大, 负域随着 β 的增大而增大, 同时边界缩小. 但是一般情况下

$$\lim_{\alpha \uparrow r} \underline{PA}_\alpha(X) = \underline{PA}_r(X)$$

$$\lim_{\beta \uparrow r} \overline{PA}_\beta(X) = \underline{PA}_r(X)$$

却不成立.

例1 设 $U = \{x_i | 1 \leq i \leq 20\}, U/R = \{X_1, X_2, X_3, X_4, X_5, X_6\}$, 其中 $X_1 = \{x_1, x_2, x_3, x_4, x_5\}; X_2 = \{x_6, x_7, x_8\}; X_3 = \{x_9, x_{10}, x_{11}, x_{12}\}; X_4 = \{x_{13}, x_{14}\}; X_5 = \{x_{15}, x_{16}, x_{17}, x_{18}\}; X_6 = \{x_{19}, x_{20}\}$. 令 $X = \{x_6, x_7, x_8, x_{13}, x_{17}\}$, 则 $\underline{PA}_{0.5}(X) = X_2 \cup X_4$, 但 $\lim_{\alpha \uparrow 0.5} \underline{PA}_\alpha(X) = X_2$, 而 $\lim_{\beta \uparrow 0.5} \overline{PA}_\beta(X) = X_2 \cup X_4 \neq \overline{PA}_{0.5}(X) = X_2$.

命题4 设 $0 < r < 1$, 则对于任意 $X \subseteq U$ 有

$$(1) \lim_{\alpha \uparrow r} \underline{PA}_\alpha(X) = \bigcup_{\alpha > r} \underline{PA}_\alpha(X) = \overline{PA}_r(X);$$

$$(2) \lim_{\beta \uparrow r} \overline{PA}_\beta(X) = \bigcap_{\beta < r} \overline{PA}_\beta(X) = \underline{PA}_r(X)$$

证明: (1) 当 $\alpha > r$ 时, 由定义知

$$\underline{PA}_\alpha(X) = \{x \in U | P(X|[x]_R) \geq \alpha\} \subseteq \{x \in U | P(X|[x]_R) \geq r\}$$

$$> r\} = \overline{PA}_r(X)$$

又由命题3知 $\underline{PA}_\alpha(X)$ 随着 α 的减小而增大, 因此

$$\lim_{\alpha \uparrow r} \underline{PA}_\alpha(X) = \bigcup_{\alpha > r} \underline{PA}_\alpha(X) \subseteq \overline{PA}_r(X)$$

若存在 $x_0 \in \overline{PA}_r(X) \setminus \bigcup_{\alpha > r} \underline{PA}_\alpha(X)$, 则有 $P(X|[x_0]) > r$, 但是

对任意 $\alpha > r$ 有 $x_0 \notin \underline{PA}_\alpha(X)$, 即

$$\forall \alpha > r, \text{有 } P(X|[x_0]) < \alpha$$

这说明 $P(X|[x_0]) \leq r$ 与 $P(X|[x_0]) > r$ 矛盾! 这意味着

$$\bigcup_{\alpha > r} \underline{PA}_\alpha(X) = \overline{PA}_r(X)$$

因此, 结论(1)成立.

(2) 由 $\beta < r$ 可得

$$\overline{PA}_\beta(X) = \{x \in U | P(X|[x]_R) > \beta\} \supseteq \{x \in U | P(X|[x]_R) \geq r\} = \underline{PA}_r(X)$$

从而由 $\overline{PA}_\beta(X)$ 关于 β 的是单调递减性得

$$\lim_{\beta \uparrow r} \overline{PA}_\beta(X) = \bigcap_{\beta < r} \overline{PA}_\beta(X) \supseteq \underline{PA}_r(X)$$

若存在 $y \in \bigcap_{\beta < r} \overline{PA}_\beta(X) \setminus \underline{PA}_r(X)$, 即对于 $\forall \beta < r, y \in \overline{PA}_\beta(X)$, 但 $y \notin \underline{PA}_r(X)$, 于是

$$P(X|[y]) > \beta, \forall \beta < r \quad (1)$$

但是, $P(X|[y]) < r$. 由(1)式知 $P(X|[y]) \geq r$, 这与 $P(X|[y]) < r$ 矛盾! 故结论(2)成立.

命题5 设 $0 < r < 1$, 则

$$\lim_{\beta \uparrow r, \alpha \uparrow r} BN(X, \alpha, \beta) = \bigcap_{\beta \uparrow r, \alpha \uparrow r} (\overline{PA}_\beta(X) \setminus \underline{PA}_\alpha(X)) \\ = \underline{PA}_r(X) \setminus \overline{PA}_r(X) = \{x \in U | P(X|[x]_R) = r\}$$

证明: 显然, $BN(X, \alpha, \beta) \supseteq \underline{PA}_r(X) \setminus \overline{PA}_r(X) = \{x \in U | P(X|[x]_R) = r\}$, 而当 α 单调下降趋于 r, β 单调上升趋于 r 时, 命题3知 $BN(X, \alpha, \beta)$ 是单调下降的, 从而

$$\lim_{\beta \uparrow r, \alpha \uparrow r} BN(X, \alpha, \beta) = \bigcap_{\beta \uparrow r, \alpha \uparrow r} (\overline{PA}_\beta(X) \setminus \underline{PA}_\alpha(X)) \\ \supseteq \underline{PA}_r(X) \setminus \overline{PA}_r(X) = \{x \in U | P(X|[x]_R) = r\}$$

若 $\exists x_0 \in \bigcap_{\beta \uparrow r, \alpha \uparrow r} (\overline{PA}_\beta(X) \setminus \underline{PA}_\alpha(X)) \setminus \{x \in U | P(X|[x]_R) = r\}$, 则对于任意 $\beta < r < \alpha$ 有 $x_0 \in BN(X, \alpha, \beta) = \overline{PA}_\beta(X) \setminus \underline{PA}_\alpha(X)$, 但 $P(X|[x_0]) \neq r$. 从而

$$P(X|[x_0]) > \beta, \forall \beta < r, \quad (2)$$

而且

$$P(X|[x_0]) < \alpha, \forall \alpha > r \quad (3)$$

由(2)式得 $P(X|[x_0]) \geq r$, 由(3)得 $P(X|[x_0]) \leq r$, 这样我们就得到了 $P(X|[x_0]) = r$ 这与假设矛盾! 命题得证.

注: 从以上命题可以看出, 随着 β 逐渐增大(到 r), 逐渐下降(到 r), 边界 $BN(X, \alpha, \beta)$ 逐渐收缩(成为 $\{x \in U | P(X|[x]_R) = r\}$), $\{x \in U | P(X|[x]_R) = r\}$ 称为 X 关于概率近似空间依参数 r 的绝对边界, 之所以给予这个名称是因为对于任意满足 $\beta < r < \alpha$ 的 α 与 β , 边界 $BN(X, \alpha, \beta)$ 都包含 $\{x \in U | P(X|[x]_R) = r\}$.

尽管例1表明 $\underline{PA}_\alpha(X)$ 关于 α 不是右连续的, $\overline{PA}_\beta(X)$ 关于 β 不是左连续的, 但是下面的命题表明它们分别关于 α 和 β 是左连续和右连续的.

命题6 设 $0 < r < 1$, 则对于任意 $X \subseteq U$ 有

$$(1) \lim_{\alpha \uparrow r} \underline{PA}_\alpha(X) = \bigcap_{\alpha < r} \underline{PA}_\alpha(X) = \underline{PA}_r(X);$$

$$(2) \lim_{\beta \uparrow r} \overline{PA}_\beta(X) = \bigcup_{\beta > r} \overline{PA}_\beta(X) = \overline{PA}_r(X)$$

证明: (1) 当 $\alpha < r$ 时, 由定理1知 $\underline{PA}_\alpha(X) \supseteq \underline{PA}_r(X)$, 又由命题3知 $\underline{PA}_\alpha(X)$ 随着 α 的增大而减少, 因此

$$\lim_{\alpha \uparrow r} \underline{PA}_\alpha(X) = \bigcap_{\alpha < r} \underline{PA}_\alpha(X) \supseteq \underline{PA}_r(X)$$

若存在 $x_0 \in \bigcap_{\alpha < r} \underline{PA}_\alpha(X) \setminus \underline{PA}_r(X)$, 则对任意 $\alpha < r$ 有 $x_0 \in \underline{PA}_\alpha(X)$, 但是 $x_0 \notin \underline{PA}_r(X)$, 即

$\forall \alpha < r$, 有 $P(X|[x_0]) \geq \alpha$, 而 $P(X|[x_0]) < r$
这是矛盾的, 因为由 $\forall \alpha < r$, 有 $P(X|[x_0]) \geq \alpha$ 可得 $P(X|[x_0]) \geq r$.

这意味着

$$\bigcap_{\alpha < r} \underline{PA}_\alpha(X) = \underline{PA}_r(X)$$

因此, 结论(1)成立.

(2) 由 $\beta > r$ 和定理1可得 $\overline{PA}_\beta(X) \subseteq \overline{PA}_r(X)$, 从而由 $\overline{PA}_\beta(X)$ 关于 β 的是单调递减性得

$$\lim_{\beta \uparrow r} \overline{PA}_\beta(X) = \bigcup_{\beta > r} \overline{PA}_\beta(X) \subseteq \overline{PA}_r(X)$$

若存在 $y \in \overline{PA}_r(X) \setminus \bigcup_{\beta > r} \overline{PA}_\beta(X)$, 即 $y \in \overline{PA}_r(X)$, 但对于 $\forall \beta > r$, $y \notin \overline{PA}_\beta(X)$, 于是 $P(X|[y]) > r$, 而

$$P(X|[y]) \leq \beta, \forall \beta > r \quad (4)$$

由(4)式知 $P(X|[y]) \leq r$, 这与 $P(X|[y]) > r$ 矛盾! 故结论(2)成立.

5. 概率粗糙近似算子与 Pawlak 近似算子的比较

命题7 若取 $P(X|[x]) = \frac{|X \cap [x]|}{|[x]|}$, 则

$$(1) \underline{PA}_1(X) = R(X)$$

$$(2) \overline{PA}_0(X) = \overline{R}(X)$$

$$(3) BN(X, 1, 0) = BN(X)$$

$$(4) NEG(X, 0) = NEG(X)$$

命题8 若取 $P(X|[x]) = \frac{|X \cap [x]|}{|[x]|}$, 则当 $0 \leq \beta < \alpha \leq 1$

时, 有

$$(1) \underline{R}(X) \subseteq \underline{PA}_\alpha(X)$$

$$(2) \overline{PA}_\beta(X) \subseteq \overline{R}(X)$$

$$(3) BN(X, \alpha, \beta) \subseteq BN(X)$$

$$(4) NEG(X, \beta) \supseteq NEG(X)$$

由命题8可以看出, 概率粗糙集中的边界一般要比相应的 Pawlak 粗糙集的边界小, 而正域和负域都比 Pawlak 粗糙集的正域和负域大, 其主要原因是概率粗糙集允许将 Pawlak 粗糙集边界中条件概率大于零小于1的某些等价类也计入到它的正域里面, 还将 Pawlak 粗糙集边界中一部分并入到它的负域中去. 从这里我们也可以看出概率粗糙集模型比 Pawlak 粗糙集模型的应用更加广泛.

6. 近似精度、可定义集与属性约简

定义4 若 $0 \leq \beta < \alpha \leq 1$, 则 X 关于概率近似空间 A_β 依参数 α, β 的近似精度 $\eta(X, \alpha, \beta)$ 定义为:

$$\eta(X, \alpha, \beta) = \begin{cases} \frac{|\underline{PA}_\alpha(X)|}{|\overline{PA}_\beta(X)|}, & \text{若 } \overline{AP}_\beta(X) \neq \emptyset \\ 0, & \text{若 } \overline{AP}_\beta(X) = \emptyset \end{cases}$$

当 $\underline{PA}_\alpha(X) = \overline{PA}_\beta(X)$ 时, 称 X 关于 A_β 是 (α, β) -可定义, 也称为 X 关于 A_β 是 (α, β) -可辨识的, 否则称 X 关于 A_β 是 (α, β) -不可辨识的或称为 (α, β) -粗糙集.

由定义4可见, X 关于概率近似空间 A_β 是 (α, β) -可定义的当且仅当其近似精度为1, 这时等价地有 $BN(X, \alpha, \beta) = \emptyset$. 若 X 关于概率近似空间 A_β 是 (α, β) -可定义的, 则相对于 X 而言我们可以对 U 中的任何一个对象 x , 都可以进行正确分类或辨识, 即要么 $x \in POS(X, \alpha)$, 要么 $x \in NEG(X, \beta)$; 而若 X 关于 A_β 是 (α, β) -粗糙集, 则相对于 X 而言, 我们不能对 U 中的所有的对象进行辨识.

命题9 若 X 关于概率近似空间 A_β 是 (α, β) -可辨识的, 则当 $0 \leq \alpha' \leq \alpha < \beta \leq \beta' \leq 1$ 时, X 关于概率近似空间 $A_{\beta'}$ 是 (α') -

β' -可辨识的.

证明: 由命题3知,

$$\underline{PA}_{\alpha'}(X) \subseteq \underline{PA}_\alpha(X) \subseteq \overline{PA}_{\beta'}(X) \subseteq \overline{PA}_\beta(X)$$

因此, 由 X 的 (α, β) -可辨识性可推得 X 的 (α', β') -可辨识性.

命题10 若 X 关于概率近似空间 A_β 是 (α, β) -不可辨识的, 则对于任意 $\alpha' \geq \alpha$ 和 $\beta' \leq \beta$, X 关于概率近似空间 $A_{\beta'}$ 是 (α', β') -不可辨识的.

证明: 类似于命题9可证.

定义5 设 $(U, C \cup D, F)$ 是决策表, P 是 U 上的概率测度, 记由 R_D 生成的 U 上的划分为

$$D^* = \{D_1, D_2, \dots, D_J\}$$

对于属性子集 $B \subseteq C$, 记由 B 生成的 U 上的二元等价关系为 R_B , 由 R_B 生成的概率近似空间记为 $B = (U, P, R_B)$, 则决策类 D_j 关于条件属性子集 B 的依赖性程度定义为

$$r(B, D_j, \alpha, \beta) = \frac{|\underline{PB}_\alpha(D_j)| + |\overline{PB}_\beta(D_j)|}{|U|}$$

决策类 D_j 关于条件属性子集 B 的依赖性程度反映了对记录在决策表中的信息进行随机试验时决策结果 D_j 或 $\sim D_j$ 进行成功预测的可能性.

定义6 条件属性子集 $B \subseteq C$ 称为关于决策类 D_j 是 (α, β) -依赖的, 若存在 B 的真子集 M 使得 $POS_M(D_j, \alpha) = POS_C(D_j, \alpha)$ 且 $NEG_M(D_j, \beta) = NEG_C(D_j, \beta)$; 否则称 B 关于决策类 D_j 是 (α, β) -独立的, 这时也称 B 是在决策表中关于决策类 D_j 的一个 (α, β) -约简.

由定义6可见, 若条件属性子集 $B \subseteq C$ 关于决策类 D_j 是 (α, β) -依赖的, 则在决策表中去掉 B 中的某些属性, 在 α 水平下不会影响对决策 D_j 的预测结果, 且在 β 水平下也不会影响对决策 $\sim D_j$ 的预测结果. 决策表中关于决策类 D_j 的一个 (α, β) -约简是决策表中关于决策类 D_j 的一个最大 (α, β) -独立属性子集.

结束语 作为处理不确定性和不精确性问题的一种数学工具, 目前粗糙集理论已在医学、生物学、化学、材料学、地理学和金融等其他学科得到了成功的应用. 粗糙集理论在人工智能中的应用主要有两大类, 一类是无决策的数据分析, 内容主要包括数据压缩、约简、聚类或分类; 另一类是有决策的数据分析, 主要包括决策分析、规则提取等, 当然也涉及对原始数据的预处理等. 经典的粗糙集模型是建立在绝对精确的包含关系上的, 这限制了粗糙集理论的应用. 本文提出基于概率测度的粗糙集模型, 给出概率粗糙近似算子, 并且给出了此类模型的近似精度、属性依赖性和约简等粗糙集理论中一些重要的概念. 本文提出的模型是经典粗糙集模型的推广, 其应用相当广泛.

参考文献

- 1 Pawlak Z. Rough sets. International Journal of Computer and Information Sciences, 1982, 11: 341~356
- 2 Pawlak Z. Rough sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Boston, 1991
- 3 Yao Y Y. Generalized rough set models. In: Rough sets in Knowledge Discovery: 1. Methodology and Applications, L. Polkowski, and A. Skowron (Eds), Physica-Verlag, Heidelberg, 286~318
- 4 Yao Y Y. Relational interpretations of neighborhood operators and rough set approximation. Information Sciences: An International Journal, 1998, 111: 239~259
- 5 张文修, 吴伟志, 梁吉业, 李德玉. 粗糙集理论与方法. 北京: 科学出版社, 2001