

引用格式:高晓峰,王青海. $U \times W$ 型双论域变精度粗糙集模型[J]. 微电子学与计算机, 2020, 37(3): 33-36. [GAO X F, WANG Q H. $U \times W$ type dual domain variable precision rough set model [J]. Microelectronics & Computer, 2020, 37(3): 33-36.]

$U \times W$ 型双论域变精度粗糙集模型

高晓峰, 王青海

(青海师范大学 计算机学院, 青海 西宁 810008)

摘 要: 针对现实世界中多个论域之间信息分类和属性约简问题, 提出了 $U \times W$ 型双论域变精度粗糙集模型, 通过对 $U \times W$ 型双论域中等价关系的研究, 得到了二维等价关系的等价类, 在此基础上对该模型结构和数学特性进行了深入研究, 得到了一些有用的结论, 最后利用粗糙熵和信息熵对其不确定性进行了度量, 并通过实例对模型及度量结果进行了验证.

关键词: $U \times W$ 型双论域; 变精度粗糙集; 粗糙熵; 信息熵

中图分类号: TP18

文献标识码: A

文章编号: 1000-7180(2020)03-0033-04

$U \times W$ type dual domain variable precision rough set model

GAO Xiao-feng, WANG Qing-hai

(School of Computer, Qinghai Normal University, Xining 810008, China)

Abstract: Aiming at the information classification and attribute reduction between multiple universes in the real world, a $U \times W$ double-disciplinary variable precision rough set model is proposed. By studying the equivalence relation in the $U \times W$ double domain, Based on the equivalence class of two-dimensional equivalence relation, the structure and mathematical characteristics of the model are deeply studied, and some useful conclusions are obtained. Finally, the uncertainty is measured by rough entropy and information entropy. The model and measurement results are verified by examples.

Key words: $U \times W$ type double domain; variable precision rough set; rough entropy; information entropy

1 引言

粗糙集理论是波兰数学家 Pawlak 于 1982 年提出用来处理实际问题中规则挖掘和属性约简的数学工具^[1-2], 已在诸多领域有成功的应用^[3-4]. 由于实际问题的复杂性和不确定性, 经典粗糙集就表现出许多不足. 于是模型在数据格式上不断地被推广, 双论域及多论域粗糙集模型在新的领域不断被扩展^[5-13].

本文提出 $U \times W$ 型双论域变精度粗糙集模型,

旨在解决现实世界中不同论域之间信息分类和属性约简问题.

2 $U \times W$ 型双论域变精度粗糙集

定义 1 设 $X \times Y$ 和 $N \times M$ 是有限论域 $U \times W$ 的非空子集, 如果对于每一对 $(e, s) \in X \times Y$ 有 $(e, s) \in N \times M$, 则 $N \times M$ 包含 $X \times Y$, 令

$$c(X \times Y, N \times M) = \begin{cases} 1 & -| (X \times Y) \cap (N \times M) | / | X \times Y | \\ 0 & | X \times Y | > 0 \end{cases}$$

收稿日期: 2019-06-24; 修回日期: 2019-07-05

基金项目: 国家社会科学基金(17XTQ013); 青海省应用基础研究项目(2017-ZJ-768); 青海省物联网重点实验室建设专项(2017-Z-Y21); 赛尔网络下一代互联网技术创新项目(NGII20160504); 教育部春晖计划项目(Z2015051); 青海师范大学教学研究重点项目(qhnujy2015102)

式中, $|X \times Y|$ 表示集合 $X \times Y$ 的基数. 称 $c(X \times Y, N \times M)$ 为集合 $X \times Y$ 关于集合 $N \times M$ 的二维相对错误分辨率. 令 $0 \leq \beta < 0.5$, 多数包含关系定义为 $N \times M^\beta \supseteq X \times Y \Leftrightarrow c(X \times Y, N \times M) \leq \beta$.

定义 2 设 $(U \times W, \mathfrak{R})$ 为近似空间, 其中论域 $U \times W$ 为非空有限集合, $\mathfrak{R}$ 为 $U \times W$ 上的等价关系, $U \times W / \mathfrak{R} = \{E_1 \times T_1, E_1 \times T_2, \dots, E_n \times T_m\}$ 为 $\mathfrak{R}$ 的等价类构成的集合. 对于 $X \times Y \subseteq U \times W$ 定义 $X \times Y$ 的 β 下近似 (也称为 β 正区域, 记为 $\text{pos}_\beta(X \times Y)$), 上近似、边界域和负区域为: $\mathfrak{R}_\beta(X \times Y) = \bigcup \{E \times T \in U \times W / \mathfrak{R} \mid c(E \times T, X \times Y) \leq \beta\}$; $\bar{\mathfrak{R}}_\beta(X \times Y) = \bigcup \{E \times T \in U \times W / \mathfrak{R} \mid c(E \times T, X \times Y) < 1 - \beta\}$; $\text{bnr}_\beta(X \times Y) = \bigcup \{E \times T \in U \times W / \mathfrak{R} \mid \beta < c(E \times T, X \times Y) < 1 - \beta\}$; $\text{negr}_\beta(X \times Y) = \bigcup \{E \times T \in U \times W / \mathfrak{R} \mid c(E \times T, X \times Y) \geq 1 - \beta\}$.

定理 1 对于 $\forall X \times Y \subseteq U \times W$, 下列关系式成立 ($\sim(X \times Y) = U \times W - X \times Y$)

$$\text{posr}_\beta(\sim(X \times Y)) = \text{negr}_\beta(X \times Y)$$

证明 $\text{posr}_\beta(\sim(X \times Y)) = \text{posr}_\beta(U \times W - X \times Y) = \bigcup \{E \times T \in U \times W / \mathfrak{R} \mid c(E \times T, U \times W - X \times Y) \leq \beta\}$, $\text{negr}_\beta(X \times Y) = \bigcup \{E \times T \in U \times W / \mathfrak{R} \mid c(E \times T, X \times Y) \geq 1 - \beta\}$ 又 $c(E \times T, U \times W - X \times Y) = 1 - |(E \times T) \cap (U \times W - X \times Y)| / |E \times T| \leq \beta \Leftrightarrow |(E \times T) \cap (X \times Y)| / |E \times T| \leq \beta$, $c(E \times T, X \times Y) = 1 - |(E \times T) \cap (X \times Y)| / |E \times T| \geq 1 - \beta \Leftrightarrow |(E \times T) \cap (X \times Y)| / |E \times T| \leq \beta$. 故 $\text{posr}_\beta(\sim(X \times Y)) = \text{negr}_\beta(X \times Y)$ 成立.

对于 $0 \leq \beta < 0.5$, 下列定理成立.

定理 2 $\mathfrak{R}_\beta(X \times Y) \supseteq \mathfrak{R}(X \times Y)$,

$$\bar{\mathfrak{R}}(X \times Y) \supseteq \bar{\mathfrak{R}}_\beta(X \times Y),$$

$$\text{bnr}_\mathfrak{R}(X \times Y) \supseteq \text{bnr}_\beta(X \times Y),$$

$$\text{negr}_\beta(X \times Y) \supseteq \text{negr}_\mathfrak{R}(X \times Y).$$

定理 2 结论由定义 2 显然可得.

3 基于 $U \times W$ 型双论域变精度粗糙集的不确定性度量

3.1 粗糙熵度量

定义 3 设 $(U \times W, \mathfrak{R})$ 为近似空间, $\mathfrak{R}$ 为 $U \times W$ 上的等价关系, 对于 $X \times Y \subseteq U \times W$, $(\mathfrak{R}_\beta, \bar{\mathfrak{R}}_\beta)$ 为 $X \times Y$ 的 β 下近似和上近似, 则 $X \times Y$ 关于 β ($0 \leq \beta < 0.5$) 的近似精度 $\alpha_{\mathfrak{R}_\beta}$ 和粗糙度 $\rho_{\mathfrak{R}_\beta}$ 分别为

$$\alpha_{\mathfrak{R}_\beta}(X \times Y) = \frac{|\mathfrak{R}_\beta(X \times Y)|}{|\bar{\mathfrak{R}}_\beta(X \times Y)|}$$

$$\rho_{\mathfrak{R}_\beta}(X \times Y) = 1 - \alpha_{\mathfrak{R}_\beta}(X \times Y)$$

粗糙熵是以信息熵和粗糙度的视角对信息的不确定性进行度量, 相比较于传统粗糙度, 其对粗糙集模型的不确定性度量更为客观. 本文基于 Liang [14] 提出的粗糙熵对 $U \times W$ 型双论域变精度粗糙集模型进行不确定性度量.

定义 4 设 $IS = (U \times W, \mathfrak{S}, V, f)$ 为信息系统, 其中: $U \times W$ 为论域; $\mathfrak{S}$ 为有限属性集; $V = \bigcup_{a \in A} V_a$, V_a 为属性 a 的值域; $f: (U \times W) \times \mathfrak{S} \rightarrow V$ 为信息函数. 属性子集 $\mathfrak{S}_1$ ($\mathfrak{S}_1 \subseteq \mathfrak{S}$) 对论域的划分 $U \times W / \mathfrak{S}_1 = \{X_1 \times Y_1, X_1 \times Y_2, \dots, X_n \times Y_m\}$, $X \times Y \subseteq U \times W$, 则属性集合 $\mathfrak{S}_1$ 的熵定义为

$$E(\mathfrak{S}_1) = - \sum_{i=1}^m \sum_{j=1}^n \frac{|X_i \times Y_j|}{|U \times W|} \log_2 \frac{1}{|X_i \times Y_j|}$$

式中, $i = 1, 2, \dots, m, j = 1, 2, \dots, n$

$X \times Y$ 在 β ($0 \leq \beta < 0.5$) 精度和等价关系 $\mathfrak{R}$ 上的粗糙熵定义为

$$E_{\mathfrak{R}_\beta}(X \times Y) = \rho_{\mathfrak{R}_\beta}(X \times Y) E(\mathfrak{S}_1)$$

定理 3 在等价关系 $\mathfrak{R}$ 下, 若 $\beta_1 < \beta_2$, 则 $E_{\mathfrak{R}_\beta}(X \times Y) > E_{\mathfrak{R}_\beta}(X \times Y)$.

证明 由定理 4 可知, 若 $\beta_1 < \beta_2$ 时,

$$\frac{|\mathfrak{R}_{\beta_1}(X \times Y)|}{|\bar{\mathfrak{R}}_{\beta_1}(X \times Y)|} < \frac{|\mathfrak{R}_{\beta_2}(X \times Y)|}{|\bar{\mathfrak{R}}_{\beta_2}(X \times Y)|}, \text{ 即}$$

$$1 - \frac{|\mathfrak{R}_{\beta_1}(X \times Y)|}{|\bar{\mathfrak{R}}_{\beta_1}(X \times Y)|} > 1 - \frac{|\mathfrak{R}_{\beta_2}(X \times Y)|}{|\bar{\mathfrak{R}}_{\beta_2}(X \times Y)|}, \text{ 即}$$

$$\rho_{\mathfrak{R}_{\beta_1}}(X \times Y) > \rho_{\mathfrak{R}_{\beta_2}}(X \times Y), \text{ 由 } E_{\mathfrak{R}_\beta}(X \times Y) = \rho_{\mathfrak{R}_\beta}(X \times Y) E(\mathfrak{S}_1) \text{ 可知, } E_{\mathfrak{R}_{\beta_1}}(X \times Y) > E_{\mathfrak{R}_{\beta_2}}(X \times Y).$$

3.2 信息熵度量

定义 5 设 $IS = (U \times W, \mathfrak{S}, V, f)$ 是一个信息系统, 其中: $U \times W$ 为论域; $\mathfrak{S}$ 为有限属性集; $V = \bigcup_{a \in A} V_a$, V_a 为属性 a 的值域; $f: (U \times W) \times \mathfrak{S} \rightarrow V$ 为信息函数. $\mathfrak{S}_1, \mathfrak{S}_2 \subseteq \mathfrak{S}$, $U \times W / \mathfrak{S}_1 = \{X_1 \times Y_1, X_1 \times Y_2, \dots, X_i \times Y_j\}$, $U \times W / \mathfrak{S}_2 = \{N_1 \times M_1, N_1 \times M_2, \dots, N_d \times M_k\}$, 若 $\forall X_e \times Y_g \in U \times W / \mathfrak{S}_1, \exists N_p \times M_q \in U \times W / \mathfrak{S}_2 (X_e \times Y_g \subseteq N_p \times M_q)$, 则称 $\mathfrak{S}_1$ 比 $\mathfrak{S}_2$ 更细, 记为 $\mathfrak{S}_1 \leq \mathfrak{S}_2$.

定义 6 设 $IS = (U \times W, \mathfrak{S}, V, f)$ 是一个信息系统, $\mathfrak{S}_1, \mathfrak{S}_2 \subseteq \mathfrak{S}$, $U \times W / \mathfrak{S}_1 = \{X_1 \times Y_1, X_1 \times Y_2, \dots, X_i \times Y_j\}$, $U \times W / \mathfrak{S}_2 = \{N_1 \times M_1, N_1 \times M_2, \dots, N_d \times M_k\}$, 若 $\forall X_e \times Y_g \in U \times W / \mathfrak{S}_1, \exists N_p \times M_q \in U \times W / \mathfrak{S}_2 (X_e \times Y_g \subset N_p \times M_q)$, 则称 $\mathfrak{S}_1$ 比 $\mathfrak{S}_2$ 严格细, 记为 $\mathfrak{S}_1 < \mathfrak{S}_2$.

定义 7 设 $IS = (U \times W, \mathfrak{S}, V, f)$ 为信息系

统, $U \times W/\mathfrak{S} = \{X_1 \times Y_1, X_1 \times Y_2, \dots, X_m \times Y_n\}$, 则 $\mathfrak{S}$ 的信息熵定义为

$$H(\mathfrak{S}) = - \sum_{i=1}^m \sum_{j=1}^n P(X_i \times Y_j) \log P(X_i \times Y_j)$$

式中, $i = 1, 2, \dots, m, j = 1, 2, \dots, n, |E \times T|$ 为集合 $E \times T$ 的基数.

定理4 设 $IS = (U \times W, \mathfrak{S}, V, f)$ 是一个信息系统, $\mathfrak{S}_1, \mathfrak{S}_2 \subseteq \mathfrak{S}, U \times W/\mathfrak{S}_1 = \{X_1 \times Y_1, X_1 \times Y_2, \dots, X_i \times Y_j\}, U \times W/\mathfrak{S}_2 = \{N_1 \times M_1, N_1 \times M_2, \dots, N_d \times M_k\}$, 若 $\mathfrak{S}_1 \leq \mathfrak{S}_2$, 则 $H(\mathfrak{S}_1) \geq H(\mathfrak{S}_2)$.

证明 (1) 由 $\mathfrak{S}_1 \leq \mathfrak{S}_2$ 知 $\forall X_e \times Y_g \in U \times W/\mathfrak{S}_1, \exists N_p \times M_q \in \frac{U \times W}{\mathfrak{S}_2} (X_e \times Y_g \subset N_p \times M_q)$,

反之对于任意 $N_p \times M_q$ 存在 S 个 $X_e \times Y_g$ 满足 $\frac{|N_p \times M_q|}{|U \times W|} = \frac{|X_1 \times Y_1|}{|U \times W|} + \frac{|X_1 \times Y_2|}{|U \times W|} + \dots + \frac{|X_s \times Y_s|}{|U \times W|}$. 有

$$\begin{aligned} & \frac{|X_1 \times Y_1|}{|U \times W|} \log \frac{|X_1 \times Y_1|}{|U \times W|} + \frac{|X_1 \times Y_2|}{|U \times W|} \log \frac{|X_1 \times Y_2|}{|U \times W|} + \dots + \frac{|X_s \times Y_s|}{|U \times W|} \log \frac{|X_s \times Y_s|}{|U \times W|} < \\ & \frac{|N_p \times M_q|}{|U \times W|} \log \frac{|N_p \times M_q|}{|U \times W|}, \text{ 则 } - \left(\frac{|X_1 \times Y_1|}{|U \times W|} \log \frac{|X_1 \times Y_1|}{|U \times W|} + \frac{|X_1 \times Y_2|}{|U \times W|} \log \frac{|X_1 \times Y_2|}{|U \times W|} + \dots + \right. \\ & \left. \frac{|X_s \times Y_s|}{|U \times W|} \log \frac{|X_s \times Y_s|}{|U \times W|} \right) > - \left(\frac{|N_p \times M_q|}{|U \times W|} \log \frac{|N_p \times M_q|}{|U \times W|} \right), \text{ 则 } - \left(\sum_{e=1}^i \sum_{g=1}^j \frac{|X_e \times Y_g|}{|U \times W|} \log \frac{|X_e \times Y_g|}{|U \times W|} \right) > - \left(\sum_{p=1}^d \sum_{q=1}^k \frac{|N_e \times M_g|}{|U \times W|} \log \frac{|N_e \times M_g|}{|U \times W|} \right), \text{ 因此有 } H(\mathfrak{S}_1) > H(\mathfrak{S}_2). \end{aligned}$$

$$(2) \text{ 由 } \mathfrak{S}_1 = \mathfrak{S}_2 \text{ 知 } \frac{|X_e \times Y_g|}{|U \times W|} = \frac{|N_p \times M_q|}{|U \times W|},$$

因此有 $H(\mathfrak{S}_1) = H(\mathfrak{S}_2)$.

定理5 设 $IS = (U \times W, \mathfrak{S}, V, f)$ 是一个信息系统, $\mathfrak{S}_1, \mathfrak{S}_2 \subseteq \mathfrak{S}$, 有:

(1) 若 $\mathfrak{S}_1 \supset \mathfrak{S}_2$, 则 $H(\mathfrak{S}_1) \geq H(\mathfrak{S}_2)$.

(2) 若 $\mathfrak{S}_1 \subset \mathfrak{S}_2$, 则 $H(\mathfrak{S}_1) \leq H(\mathfrak{S}_2)$.

定理5结论由定理4显然可得.

定理4说明了信息熵和知识粗细变化的关系: 知识变粗, 信息熵变小, 不确定性变小; 定理5说明了属性变化对信息熵的影响: 属性变多, 划分变细, 信息熵变大, 不确定性变大.

4 应用实例

$$\text{设 } U = \left\{ \begin{matrix} x_1, x_2, x_3, x_4, \\ x_5, x_6, x_7, x_8 \end{matrix} \right\}, W = \left\{ \begin{matrix} y_1, y_2, y_3, \\ y_4, y_5, y_6 \end{matrix} \right\}.$$

U 关于等价关系 R 的划分 $U/R = \{\{x_1, x_2\}, \{x_3, x_4, x_5, x_6, x_7\}, \{x_8\}\}$, 记为 $U/R = \{E_1, E_2, E_3\}$; W 关于等价关系 S 的划分为 $W/S = \{\{y_1\}, \{y_2, y_3\}, \{y_4, y_5, y_6\}\}$, 记为 $W/S = \{T_1, T_2, T_3\}$. 则 $U \times W$ 关于等价关系 $\mathfrak{R}$ 的划分为:

$$E_1 \times T_1 = \{(x_1, y_1), (x_2, y_1)\},$$

$$E_1 \times T_2 = \left\{ \begin{matrix} (x_1, y_2), (x_1, y_3), \\ (x_2, y_2), (x_2, y_3) \end{matrix} \right\},$$

$$E_1 \times T_3 = \left\{ \begin{matrix} (x_1, y_4), (x_1, y_5), (x_1, y_6), \\ (x_2, y_4), (x_2, y_5), (x_2, y_6) \end{matrix} \right\},$$

$$E_2 \times T_1 = \left\{ \begin{matrix} (x_3, y_1), (x_4, y_1), (x_5, y_1), \\ (x_6, y_1), (x_7, y_1) \end{matrix} \right\},$$

$$E_2 \times T_2 = \left\{ \begin{matrix} (x_3, y_2), (x_3, y_3), (x_4, y_2), (x_4, y_3), \\ (x_5, y_2), (x_5, y_3), (x_6, y_2), (x_6, y_3), \\ (x_7, y_2), (x_7, y_3) \end{matrix} \right\},$$

$$E_2 \times T_3 = \left\{ \begin{matrix} (x_3, y_4), (x_3, y_5), (x_3, y_6), (x_4, y_4), \\ (x_4, y_5), (x_4, y_6), (x_5, y_4), (x_5, y_5), \\ (x_5, y_6), (x_6, y_4), (x_6, y_5), (x_6, y_6), \\ (x_7, y_4), (x_7, y_5), (x_7, y_6) \end{matrix} \right\},$$

$$E_3 \times T_1 = \{(x_8, y_1)\}, E_3 \times T_2 = \{(x_8, y_2), (x_8, y_3)\},$$

$$E_3 \times T_3 = \{(x_8, y_4), (x_8, y_5), (x_8, y_6)\}.$$

对于精度水平: $\beta_1 = 0, \beta_2 = 0.2, \beta_3 = 0.4$, 计算集合 $X \times Y$ 的上下近似.

$$X \times Y = \left\{ \begin{matrix} (x_1, y_1), (x_1, y_2), (x_1, y_4), (x_1, y_6), (x_2, y_1), \\ (x_2, y_4), (x_2, y_6), (x_2, y_7), (x_3, y_2), (x_3, y_3), \\ (x_3, y_4), (x_3, y_5), (x_3, y_6), (x_3, y_7), (x_4, y_2), \\ (x_4, y_3), (x_4, y_5), (x_5, y_3), (x_5, y_4), (x_5, y_7), \\ (x_6, y_2), (x_6, y_3), (x_7, y_2), (x_7, y_3), (x_8, y_4), \\ (x_8, y_6) \end{matrix} \right\}$$

$$c(E_1 \times T_1, X \times Y) = 0; c(E_1 \times T_2, X \times Y) = 0.75;$$

$$c(E_1 \times T_3, X \times Y) = 0.33; c(E_2 \times T_1, X \times Y) = 1;$$

$$c(E_2 \times T_2, X \times Y) = 0.1; c(E_2 \times T_3, X \times Y) = 0.67;$$

$$c(E_3 \times T_1, X \times Y) = 1; c(E_3 \times T_2, X \times Y) = 1;$$

$$c(E_3 \times T_3, X \times Y) = 0.33.$$

假设分类误差 $\beta_1 = 0$:

$$\mathfrak{R}_0(X \times Y) = E_1 \times T_1$$

$$\bar{\mathfrak{R}}_0(X \times Y) = (E_1 \times T_1) \cup (E_1 \times T_2) \cup (E_1 \times T_3) \cup (E_2 \times T_2) \cup (E_2 \times T_3) \cup (E_3 \times T_3)$$

$$\text{bnr}_0(X \times Y) = (E_1 \times T_2) \cup (E_1 \times T_3) \cup (E_2 \times T_2) \cup (E_2 \times T_3) \cup (E_3 \times T_3)$$

$$\text{negr}_0(X \times Y) = (E_2 \times T_1) \cup (E_3 \times T_1) \cup (E_3 \times T_2)$$

$$\alpha_{\mathfrak{R}_0} = 0.05 \quad \rho_{\mathfrak{R}_0} = 0.95$$

假设分类误差 $\beta_2 = 0.2$:

$$\mathfrak{R}_{0,2}(X \times Y) = (E_1 \times T_2) \cup (E_2 \times T_2)$$

$$\bar{\mathfrak{R}}_{0,2}(X \times Y) = (E_1 \times T_1) \cup (E_1 \times T_2) \cup (E_1 \times T_3) \cup (E_2 \times T_2) \cup (E_2 \times T_3) \cup (E_3 \times T_3)$$

$$\text{bnr}_{0,2}(X \times Y) = (E_1 \times T_2) \cup (E_1 \times T_3) \cup (E_2 \times T_3) \cup (E_3 \times T_3)$$

$$\text{negr}_{0,2}(X \times Y) = (E_2 \times T_1) \cup (E_3 \times T_1) \cup (E_3 \times T_2)$$

$$\alpha_{\mathfrak{R}_{0,2}} = 0.3 \quad \rho_{\mathfrak{R}_{0,2}} = 0.7$$

假设分类误差 $\beta_3 = 0.4$:

$$\mathfrak{R}_{0,4}(X \times Y) = (E_1 \times T_1) \cup (E_1 \times T_3) \cup (E_2 \times T_2) \cup (E_2 \times T_3)$$

$$\bar{\mathfrak{R}}_{0,4}(X \times Y) = (E_1 \times T_1) \cup (E_1 \times T_3) \cup (E_2 \times T_2) \cup (E_2 \times T_3)$$

$$\text{bnr}_{0,4}(X \times Y) = \varphi$$

$$\text{negr}_{0,4}(X \times Y) = (E_1 \times T_2) \cup (E_2 \times T_1) \cup (E_2 \times T_3) \cup (E_3 \times T_1) \cup (E_3 \times T_2)$$

$$\alpha_{\mathfrak{R}_{0,4}} = 0.53 \quad \rho_{\mathfrak{R}_{0,4}} = 0.47$$

若属性子集 $\mathfrak{S}_1$ 对论域划分和 $\mathfrak{R}$ 相同,我们可以计算出当分类误差分别 $\beta_2 = 0.2$ 和 $\beta_3 = 0.4$ 时,其对应粗糙熵分别为:

$$E_{\mathfrak{S}_1\beta_2}(X \times Y) = 2.26, E_{\mathfrak{S}_1\beta_3}(X \times Y) = 1.52.$$

则可得若 $\beta_2 < \beta_3$, 则 $E_{\mathfrak{S}_1\beta_2}(X \times Y) > E_{\mathfrak{S}_1\beta_3}(X \times Y)$. 这个结果验证了定理 3 的正确性.

若属性子集 $\mathfrak{S}_1$ 对论域划分和 $\mathfrak{R}$ 相同,属性子集 $\mathfrak{S}_2$ 对论域划分为

$$\left\{ \begin{array}{l} \{E_1 \times T_1 \cup E_1 \times T_2 \cup E_1 \times T_3\} \\ \{E_2 \times T_1 \cup E_2 \times T_2 \cup E_2 \times T_3\} \\ \{E_3 \times T_1 \cup E_3 \times T_2 \cup E_3 \times T_3\} \end{array} \right\}, \text{ 则对应的信}$$

息熵分别为: $H(\mathfrak{S}_1) = 0.83, H(\mathfrak{S}_2) = 0.39$. 则可得若 $\mathfrak{S}_1 \leq \mathfrak{S}_2$, 则 $H(\mathfrak{S}_1) \geq H(\mathfrak{S}_2)$. 这个结果验证了定理 4 和定理 5 的正确性.

5 结束语

本文在二维等价关系下定义了 $U \times W$ 型双论域变精度粗糙集模型,该模型是 $U \times W$ 型双论域粗

糙集和变精度结合成的扩展形式. 该模型在现实世界中多论域之间的信息分类和属性约简中具有广阔的应用前景. 如何在 $U \times W$ 型双论域变精度粗糙集模型中获取决策规则及与机器学习模型相融合是下一步要研究的内容.

参考文献:

- [1] PAWLAK Z. Rough set[J]. International Journal of computer and Information Sciences. 1982 (11): 341-356.
- [2] PAWLAK Z. Rough sets; theoretical aspects of reasoning about data [M]. Boston: Kluwer Academic Publisher, 1991.
- [3] 黄锦静,陈岱,李梦天. 基于粗糙集的决策树在医疗诊断中的应用[J]. 计算机技术与发展, 2017, 27(12): 148-152.
- [4] 王虎,王锐,曾珠. 基于粗糙集与 D_S 证据理论的汽车售后服务知识推理[J]. 工业工程, 2014, 17(5): 35-40.
- [5] 张文修,吴伟志,梁吉业,等. 粗糙集理论与方法[J]. 北京:科学出版社,2003.
- [6] 余扬. 双论域的粗糙集模型[J]. 科学技术与工程, 2005, 5(10): 661-662.
- [7] 朱英丽,杨勇,朱晓钟. 双论域上的粗糙集模型[J]. 微电子学与计算机, 2011, 28(4): 57-60.
- [8] 孙文鑫. 双论域的一般多粒度粗糙集[J]. 计算机工程与应用, 2017, 53(4): 79-83.
- [9] 庾慧英,刘尚. 双论域上的变精度粗糙集模型[J]. 科学技术与工程, 2007, 7(1): 17.
- [10] 林群,阎瑞霞. 基于变精度双论域粗糙集的个性化推荐方法[J]. 计算机应用与软件, 2017, 34(6): 250-256.
- [11] 王国胤,张清华. 不同知识粒度下粗糙集的不确定性研究[J]. 计算机学报, 2008, 31(9): 1588-1598.
- [12] 王国胤,张清华,马希骛,等. 知识不确定性问题的粒计算模型[J]. 软件学报, 2011, 22(4): 676-694.
- [13] 阎瑞霞,郑建国,翟育明. 双论域粗糙集的不确定性度量[J]. 上海交通大学学报, 2011, 45(12): 1841-1845.
- [14] LIANG J Y, SHI Z Z. The information entropy, rough entropy and knowledge granulation in rough set theory [J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2004, 12(1): 37-46.

作者简介:

高晓峰 男, (1994-), 硕士研究生. 研究方向为粒计算与粗糙集、机器学习和人工智能. E-mail: 3355706780@qq.com.

王青海 男, (1968-), 硕士, 教授. 研究方向为神经网络、粗糙集理论、人工智能.