

信息熵约简与代数约简差异性的数学本质分析*

曾凡智^{1,2}, 卢炎生¹, 黄国顺², 陈章³

(1. 华中科技大学 计算机科学与技术学院, 湖北武汉 430074;
2. 佛山科学技术学院 计算机系, 广东佛山 528000; 3. 广东商学院 计算机学院, 广东广州 510320)

摘 要: 讨论了基于正区域与近似分类质量代数约简定义的等价性, 证明了基于决策强度的属性约简与信息熵约简是等价的, 从条件概率的角度将代数约简与信息熵约简的数学模型在形式上给出了统一表示, 分析了它们在一致决策表下是一致的以及在不一致决策表下是不一致的根本原因。

关键词: 近似分类质量; 决策强度; 代数约简; 信息熵约简

中图分类号: TP18 **文献标识码:** A **文章编号:** 0529-6579(2009)03-0028-05

The Mathematical Nature for the Differences between Algebraic and Information Entropy Reductions

ZENG Fanzhi^{1,2}, LU Yansheng¹, HUANG Guoshung², CHEN Zhang³

(1. School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China;

2. Department of Computer Science, Foshan University, Foshan 528000, China)

3. School of Computer, Guangdong University of Business Studies, Guangzhou 510320, China)

Abstract: The equivalence between algebraic reduction definitions based on approximate classification quality and positive region is discussed. It is proved that the attribute reduction concept of decision power is just the one based on conditional information entropy. From the view of conditional probability, the mathematic models of algebraic reduction and information entropy reduction are united formally. The consistence and inconsistency between them are analyzed.

Key words: approximation classification quality; decision power; algebraic reduction; information entropy reduction

属性约简是 Rough 集理论的核心问题之一^[1], 也是知识获取的关键步骤, 因此属性约简研究深受各研究者的关注。目前已有多种属性约简的定义被提出, 归纳起来主要有代数观点下的属性约简, 简称为代数约简, 有基于信息熵的属性约简, (简称为信息熵约简)^[2-3]、分布约简、最大分布约简、分配约简及近似约简等^[4]。讨论各种约简概念之间的关系, 有利于澄清各种约简的本质, 现有的研究结果表明, 分配约简与近似约简是等价的^[4],

信息熵约简与分布约简也是一致的^[5]。在不一致决策表下, 代数约简与信息熵约简并不完全相同, 王国胤、黄国顺等分别从定量和差别矩阵的角度讨论了代数核和信息核的差异性问题^[6-7]。由于现有代数约简与信息熵约简的定义形式相差太大, 很难将两者作进一步的比较和分析。

最近, 蒋思宇等提出一种基于决策强度的属性约简概念^[8], 徐章艳等^[9]学者对其作了进一步完善, 本文在他们工作的基础上作进一步的讨论, 指

* 收稿日期: 2008-11-16

基金项目: 广东省自然科学基金资助项目(06023960)

作者简介: 曾凡智(1964年生)男, 副教授; E-mail: coolhead@126.com

出决策强度属性约简本质上即是信息熵约简, 从条件概率的角度给出了代数约简与信息熵约简数学模型的统一表示形式, 分析了它们在一致决策表下的一致性、不一致决策表下的不一致性问题及其本质原因。

1 相关基本概念与理论

定义1 决策表 $S = \langle U, V, f, C \cup D \rangle$ 是一类特殊的信息系统, 其中 U 是一组对象的非空有限集合, 称为论域。 C 为有限的条件属性集, D 为有限的决策属性集, $C \cap D = \emptyset$, $V = \bigcup_{a \in C} V_a$, V_a 为属性 a 的值域; $f: U \times (C \cup D) \rightarrow V$ 是信息函数。对 U 上的任意属性集 $B \subseteq C \cup D$, 定义不可分辨关系 $ind(B) = \{(x, y) \in U^2 \mid \forall a \in B, f(x, a) = f(y, a)\}$, 关系 $ind(B)$ 构成 U 的一个划分, 用 U/B 表示, 其中的任何元素 $[x]_B = \{y \mid \forall a \in B, f(x, a) = f(y, a)\}$ 称为等价类。

定义2 给定决策表 $S = \langle U, V, f, C \cup D \rangle$, 对 $\forall B \subseteq C \cup D, X \subseteq U$, 记 $U/ind(B) = \{B_1, B_2, \dots, B_t\}$, 称 $\underline{B}X = \{x \in U: [x]_B \subseteq X\}$ 为 X 关于 B 的下近似集, 称 $\overline{B}X = \{x \in U: [x]_B \cap X \neq \emptyset\}$ 为 X 关于 B 的上近似集。

定义3 给定决策表 $S = \langle U, V, f, C \cup D \rangle$, 设 $U/D = \{D_1, D_2, \dots, D_r\}$ 表示决策属性集 D 对 U 的划分。 $U/B = \{B_1, B_2, \dots, B_t\}$ 表示条件属性集 B ($B \subseteq C$) 对 U 的划分, 称 $POS_B(D) = \bigcup_{x \in U/D} \underline{B}x$ 为 B 关于 D 的正区域。若 $POS_C(D) = U$ 称其为一致决策表, 如果 $POS_C(D) = \emptyset$ 称其为完全不一致决策表, 其余情形称为部分不一致决策表。记 $U_1 = POS_C(D)$, $U_2 = U - U_1$, 称 U_1 为一致对象集, U_2 为矛盾对象集或不一致对象集。

定义4 给定决策表 $S = \langle U, V, f, C \cup D \rangle$, 设 $U/D = \{D_1, D_2, \dots, D_r\}$, $B \subseteq C$, 则 B 对 D 近似分类质量定义为:

$$\gamma_B(D) = \sum_{j=1}^r |\underline{B}D_j| / |U|$$

其中 $|\cdot|$ 表示集合的基数。

特别地, 若 $\gamma_C(D) = 1$, 那么决策表是一致的, 否则是不一致的。

定义5 给定决策表 $S = \langle U, V, f, C \cup D \rangle$, $B \subseteq C$, 若 $POS_B(D) = POS_C(D)$ 且对于任意的 $B' \subset B$ 有 $POS_{B'}(D) \neq POS_C(D)$, 则称 B 是 C 相对于 D 的属性约简, 称为代数约简。

定义6 给定决策表 $S = \langle U, V, f, C \cup D \rangle$, 由 $B \subseteq C$ 及 D 在 U 上导出的划分分别为 $U/B =$

$\{B_1, B_2, \dots, B_t\}$, $U/D = \{D_1, D_2, \dots, D_r\}$ 那么 D 相对于 B 的条件熵定义为:

$$H(D|B) = - \sum_{i=1}^t (p(B_i) \sum_{j=1}^r p(D_j|B_i) \cdot \log(p(D_j|B_i)))$$

其中 $p(B_i) = \frac{|B_i|}{|U|}$, $p(D_j|B_i) = \frac{|D_j \cap B_i|}{|B_i|}$, $i = 1, 2, \dots, t, j = 1, 2, \dots, r$. 若对于 $B \subseteq C$ 相对于决策属性 D 有 $H(D|B) = H(D|C)$ 且对于 B 中任意属性 b 都有 $H(D|B) \neq H(D|B - \{b\})$, 则称 B 是 C 信息熵观点下的属性约简, 简称为信息熵约简。

定义7 在决策表 $S = \langle U, V, f, C \cup D \rangle$ 中, 设 $U/D = \{D_1, D_2, \dots, D_r\}$, $P(D_j/[x]_B) = \frac{|D_j \cap [x]_B|}{|[x]_B|}$, 记 $\mu_B(x) = (P(D_1/[x]_B), P(D_2/[x]_B), \dots, P(D_r/[x]_B))$, 若对 $\forall x \in U, B \subseteq C$ 有 $\mu_B(x) = \mu_C(x)$, 则称 B 是分布协调集, 若 B 是分布协调集且 B 的任何真子集不是分布协调集, 则称 B 为 C 的分布约简。

2 近似分类质量与决策强度

根据定义5, 判断 $B \subseteq C$ 是否为 C 的代数约简的关键在于判断是否有 $POS_B(D) = POS_C(D)$ 成立, 而判断两个集合是否相等是比较费时的。文献[10]给出一种简便方法, 它将判断两正区域是否相等简化为判断它们的基数是否相等, 得到如下定理1。

定理1^[10] 给定决策表 $S = \langle U, V, f, C \cup D \rangle$, $B \subseteq C$, $POS_C(D) = POS_B(D)$ 的充分必要条件是 $|POS_C(D)| = |POS_B(D)|$ 。

定理2 给定决策表 $S = \langle U, V, f, C \cup D \rangle$, $B \subseteq C$, $\gamma_B = \frac{|POS_B(D)|}{|U|}$ 。

证明 设 $U/B = \{B_1, B_2, \dots, B_t\}$, $U/D = \{D_1, D_2, \dots, D_r\}$, 则 $POS_B(D) = \bigcup_{D_j \in U/D} \underline{B}D_j$, 由于对任意的 $1 \leq i, j \leq r$, 只要 $i \neq j$ 有 $D_i \cap D_j = \emptyset$, 从而 $\underline{B}D_i \cap \underline{B}D_j = \emptyset$, 所以 $|POS_B(D)| = \sum_{j=1}^r |\underline{B}D_j|$, 从而 $\gamma_B = \frac{|POS_B(D)|}{|U|}$ 。

根据定理1、2可得到代数约简的另一等价定义形式。

定义8^[11] 给定决策表 $S = \langle U, V, f, C \cup D \rangle$, 设 $U/D = \{D_1, D_2, \dots, D_r\}$, $B \subseteq C$, 若 $\gamma_B = \gamma_C$, 且不存在 $B' \subset B$, 使得 $\gamma_{B'} = \gamma_C$, 则称 B 为 C 的代数约简。

与定义5不同,定义8的意义在于可以从概率的角度去刻画代数约简。由于下近似集是以等价类为基本单位进行运算的,因此很容易产生“非此即彼”的问题。为了克服近似分类质量这种过于“粗糙”的缺点,多位学者提出了多种改进方法,其基本思想是同时考虑一致对象与不一致对象的影响,如在文献[8],蒋思宇等提出的一种平均决策强度的概念,徐章艳等针对它不具有单调性的缺点,提出一种修正的平均决策强度公式,以保证其单调性;Hu等提出的一种刻画决策表的决策性能公式^[12]。

定义9^[9,12] 给定决策表 $S = \langle U, V, f, C \cup D \rangle$, $B \subseteq C$, 记 $U/B = \{B_1, B_2, \dots, B_t\}$, $U/D = \{D_1, D_2, \dots, D_r\}$, 其决策强度定义为 σ_B :

$$\sigma_B(D) = \sum_{i=1}^t \sum_{k=1}^r \frac{|B_i \cap D_k|}{|B_i|} \times \frac{|B_i \cap D_k|}{|U|}$$

如果将 $\frac{|B_i \cap D_k|}{|B_i|}$ 理解为 $B_i \rightarrow D_k$ 的决策可信度, $\frac{|B_i \cap D_k|}{|U|}$ 理解为决策的样本覆盖度, 那么

$\sigma_B(D)$ 表示的就是综合决策强度。特别地, 如果决策表是一致的, 那么对任意的 $C_i \in U/C$, 一定存在相应的某个 $D_{k_0} \in U/D$, 使得 $C_i \subseteq D_{k_0}$, 否则就与 $POS_C(D) = U$ 矛盾, 因此有 $\frac{|C_i \cap D_{k_0}|}{|C_i|} = 1$ 且对任意的 $k \neq k_0, 1 \leq k \leq r$ 有 $\frac{|C_i \cap D_k|}{|C_i|} = 0$,

从而 $\sigma_C(D) = \sum_{k=1}^r \frac{|CD_k|}{|U|} = \gamma_C(D) = 1$ 。这意味着决策强度在一致决策表下与近似分类质量的是完全相同的。对于不一致决策表, 两者之间并不完全一致, 具体考察算例1中的情形。

算例1 在表1的决策表 S_1 中, $U = \{x_1, x_2, \dots, x_9\}$, $C = \{a, b, c\}$, $D = \{d\}$ 。

表1 决策表 S_1
Table 1 Decision table S_1

U	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
a	1	1	1	1	1	1	1	1	2
b	1	1	1	1	1	2	2	2	2
c	1	1	2	2	2	2	2	2	1
d	1	2	1	1	2	1	3	3	2

记 $\frac{U}{D} = \{\{x_1, x_3, x_4, x_6\}, \{x_2, x_5, x_9\}, \{x_7, x_8\}\}$,

$\frac{U}{C} = \{\{x_1, x_2\}, \{x_3, x_4, x_5\}, \{x_6, x_7, x_8\}, \{x_9\}\}$, 易得其正区域 $POS_C(D) = \{x_9\}$, $\gamma_C = (1/9)$, $\sigma_C(D) = (2/3)$ 。显然, $\sigma_C(D) \geq \gamma_C(D)$, 这是因为平均决策强度不仅考虑了一致决策对象的影响还同时考虑到了不一致决策对象的影响。

3 代数约简与决策强度约简的异同性分析

定理3^[9] 给定决策表 $S = \langle U, V, f, C \cup D \rangle$, 对任意 $B \subseteq A \subseteq C$, 则有 $\sigma_B(D) \leq \sigma_A(D)$ 。

定理3保证了决策强度的单调性, 因此可以给出基于决策强度的属性约简定义。

定义10^[9] 给定决策表 $S = \langle U, V, f, C \cup D \rangle$, $B \subseteq C$, 若 $\sigma_B(D) = \sigma_C(D)$, 则称 B 是 C 关于 D 的一个决策强度协调集, 如果 $\sigma_B(D) = \sigma_C(D)$ 且对任意的 $b \in B$, 均有 $\sigma_{B-b}(D) < \sigma_C(D)$, 则称 B 是 C 的一个基于决策强度的属性约简, 简称为决策强度约简。

在一致决策表下, 决策强度协调集与代数协调集是等价的。

定理4 给定一致决策表 $S = \langle U, V, f, C \cup D \rangle$, $B \subseteq C$, B 是 C 的决策强度协调集当且仅当 B 是 C 的代数协调集。

证明 若 B 是 C 的代数协调集, 又因决策表是一致的, 从而 $\sigma_C(D) = \gamma_C(D) = \gamma_B(D) = \sigma_B(D)$, 即 B 是 C 的决策强度协调集。

反之, 若 B 是 C 的决策强度协调集, 那么有 $\sigma_B(D) = \sigma_C(D)$, 设 $U/B = \{B_1, B_2, \dots, B_t\}$, $U/C = \{C_1, C_2, \dots, C_s\}$, $U/D = \{D_1, D_2, \dots, D_r\}$, 由 $B \subseteq C$ 知 C 是比 B 更细的划分, 记 $T(B_j) = \{C_i \mid C_i \subseteq B_j, i = 1, 2, \dots, s\}$, 则 $T(B_j)$ 构成 B_j 的一个划分, 假设 $T(B_j) = \{C_{j1}, C_{j2}, \dots, C_{jl_j}\}$, 那么

$$\sum_{i=1}^{l_j} \frac{p(C_{ji})}{p(B_j)} p(D_k/C_{ji}) = \sum_{i=1}^{l_j} \frac{|C_{ji}|}{|B_j|} \frac{|D_k \cap C_{ji}|}{|C_{ji}|} = \sum_{i=1}^{l_j} \frac{|D_k \cap C_{ji}|}{|B_j|} = \sum_{i=1}^{l_j} \frac{|D_k \cap B_j|}{|B_j|} = p(D_k/B_j)$$

所以

$$\sigma_C(D) = \sum_{i=1}^s \frac{|C_i|}{|U|} \sum_{k=1}^r \left(\frac{|C_i \cap D_k|}{|C_i|} \right)^2 = \sum_{i=1}^t \frac{|B_j|}{|U|} \left(\sum_{j=1}^{l_j} \frac{|C_{ji}|}{|B_j|} \sum_{k=1}^r \left(\frac{|C_{ji} \cap D_k|}{|C_{ji}|} \right)^2 \right)$$

由函数 $f(z_1, z_2, \dots, z_r) = \sum_{j=1}^r z_j^2$ 为下凸函数知

$$\sum_{i=1}^{l_j} \frac{|C_{ji}|}{|B_j|} \sum_{k=1}^r \left(\frac{|C_{ji} \cap D_k|}{|C_{ji}|} \right)^2 \geq$$

$$\sum_{k=1}^r \left(\sum_{j=1}^{l_j} \frac{|C_{jk}|}{|B_j|} \left(\frac{|C_{jk} \cap D_k|}{|C_{jk}|} \right) \right)^2 = \sum_{k=1}^r p^2(D_k/B_j)$$

从而有 $\sigma_c(D) \geq \sigma_B(D)$, 且等号成立的充分必要条件为:

$$\begin{aligned} p(D_k/C_{j1}) &= p(D_k/C_{j2}) = \cdots = \\ p(D_k/C_{jl_j}) &= p(D_k/B_j) \end{aligned} \quad (1)$$

由于 S 是一致的, $POS_c(D) = U$, 从而对任意 $C_{jk} \in T(B_j)$, 存在某个 $D_{k_0} \in U/D$, 使得 $C_{jk} \subseteq D_{k_0}$, 根据式 (1) 知 $B_j \subseteq D_{k_0}$, 由 B_j 的任意性知 $POS_B(D) = U$, 即 B 是 C 的代数协调集。

根据定理 4、定义 5 和 8 知在一致决策表下, 决策强度属性约简和代数约简是等价的。

然而在不一致决策表下, 决策强度协调集和代数协调集并不完全等价, 但有如下结论。

定理 5 给定不一致决策表 $S = \langle U, V, f, C \cup D \rangle$, 若 $B \subseteq C$ 是 C 的决策强度协调集, 则其必为 C 的代数协调集。

证明 若 $B \subseteq C$, 所以有 $POS_B(D) \subseteq POS_c(D)$, 下面只须证明 $POS_c(D) \subseteq POS_B(D)$ 。

对于任意 $x \in POS_c(D)$, 存在某等价类 $C_{jk} \in U/C$ 及 $D_{k_0} \in U/D$, 使得 $x \in C_{jk} \subseteq D_{k_0}$, 从而 $p(D_{k_0}/C_{jk}) = 1$ 。如果 B 是 C 的决策强度协调集, 那么 $\sigma_c(D) = \sigma_B(D)$, 根据定理 4 后半部分的证明过程知, 等号成立的充分必要条件为: $p(D_k/C_{j1}) = p(D_k/C_{j2}) = \cdots = p(D_k/C_{jl_j}) = p(D_k/B_j)$, 因此 $p(D_{k_0}/B_j) = 1$, 这意味着 $B_j \subseteq D_{k_0}$, 即 $x \in POS_B(D)$ 。所以 $POS_c(D) \subseteq POS_B(D)$ 。

在不一致决策表下, 代数协调集不一定是决策强度协调集, 具体可考察例 1 中的算例。

算例 2 (续例 1)

由 $POS_{|a|}(D) = \{x_9\}$ 及 $POS_c(D) = \{x_9\}$ 知 $\{a\}$ 为其代数协调集, 但 $\sigma_c(D) \neq \sigma_{|a|}(D)$, 其中 $\sigma_c(D) = \frac{2}{3}$, $\sigma_{|a|}(D) = \frac{4}{9}$, 所以 $\{a\}$ 不是 C 的决策强度协调集。进一步可知 $\{a\}$ 为 C 的代数约简但不是决策强度约简。

4 代数约简与信息熵约简差异性的数学本质分析

定理 6 给定决策表 $S = \langle U, V, f, C \cup D \rangle$, $B \subseteq C$, B 是 C 的决策强度约简当且仅当 B 是 C 的信息熵约简。

证明 令 $p(B_i) = \frac{|B_i|}{|U|}$, $p(D_j | B_i) =$

$$\frac{|D_j \cap B_i|}{|B_i|}, f(z_1, z_2, \dots, z_r) = \sum_{k=1}^r z_k^2 \text{ 则 } \sigma_B(D) =$$

$$\sum_{i=1}^l p(B_i) f(p(D_1/B_i), p(D_2/B_i), \dots, p(D_r/B_i)),$$

因 $f(z_1, z_2, \dots, z_r)$ 为下凸函数, 根据文献 [5] 的结论知, B 是 C 的决策强度约简当且仅当 B 是 C 的信息熵约简。

为比较代数约简与信息熵约简在数学模型语义的差异性, 下面将近似分类质量改写。

由于近似分类质量 γ_B 的计算只与下近似集 $\underline{BD}_k$ 有关, 而 $\underline{BD}_k \subseteq D_k$, 从而只要 $\underline{BD}_k \neq \emptyset$, 就有 $\frac{|\underline{BD}_k \cap D_k|}{|\underline{BD}_k|} = 1$, 因此可给出近似分类质量的另一等价定义形式。

定义 4' 给定决策表 $S = \langle U, V, f, C \cup D \rangle$, 设 $U/D = \{D_1, D_2, \dots, D_r\}$, $B \subseteq C$, 则近似分类质量定义为:

$$\gamma_B(D) = \sum_{k=1}^r \frac{|\underline{BD}_k \cap D_k|}{|\underline{BD}_k|} \times \frac{|\underline{BD}_k \cap D_k|}{|U|}$$

定理 7 给定决策表 $S = \langle U, V, f, C \cup D \rangle$, $B \subseteq C$, 则 $\gamma_c(D) \geq \gamma_B(D)$, 等号成立的充分必要条件是任意的 $D_k \in \frac{U}{D}$, $|\underline{BD}_k| = |\underline{CD}_k|$ 。

证明 由于 $B \subseteq C$, 所以对任意的 $D_k \in \frac{U}{D}$, $\underline{BD}_k \subseteq \underline{CD}_k$, 从而 $|\underline{BD}_k| \leq |\underline{CD}_k|$, 由近似分类质量定义知 $\gamma_c(D) \geq \gamma_B(D)$ 。若 $\gamma_B(D) = \gamma_c(D)$, 那么由 $|\underline{BD}_k| \leq |\underline{CD}_k|$ 必有 $|\underline{BD}_k| = |\underline{CD}_k|$, 否则有 $\gamma_c(D) > \gamma_B(D)$, 与题设矛盾。反之结论显然成立。

定理 7 表明, 由于 $\underline{BD}_k$ 中的元素在 D_k 的条件概率都为 1, 因此判断 $B \subseteq C$ 是否为 C 的代数协调集简化为判断 D_k 关于 B 的下近似集元素的个数与关于 C 的下近似集元素的个数是否相等, 即 $|\underline{BD}_k| = |\underline{CD}_k|$ 是否成立, 这意味着判断 $B \subseteq C$ 是否为 C 的代数协调集简化为判断其样本覆盖率 $\frac{|\underline{BD}_k|}{|U|} = \frac{|\underline{CD}_k|}{|U|}$ 是否成立即可。

若 $B \subseteq C$ 是 C 的信息熵协调集, 那么对任意的 $B_j \in \frac{U}{B}$ 是由若干具有相同条件概率的 C 等价类的

并集构成, 不妨设 $B_j = \bigcup_{l=1}^{l_j} C_{jl}$, 那么有 $p(D_k/C_{jl}) = p(D_k/C_{j2}) = \cdots = p(D_k/C_{jl_j}) = p(D_k/B_j)$, 特别地, 如果此时条件概率的取值为 1, 则 $B_j \subseteq \underline{BD}_k$,

$\bigcup_{j=1}^{l_j} C_{j1} \subseteq \underline{CD}_k$, 且 $|\bigcup_{j=1}^{l_j} C_{j1}| = |B_j|$, 易得 $|\underline{BD}_k| = |\underline{CD}_k|$, 即 B 是 C 的代数协调集。这意味着信息熵协调集 (保持各对象的条件概率不变) 必为代数协调集 (保持条件概率为 1 的对象不变)。反之则不一定成立, 因为条件概率不一定都能取到 1, 除非决策表是一致的。这就是代数约简与信息熵约简在一致决策表下等价, 在不一致决策表下不相一致的本质原因。

5 结 语

本文主要在文献 [9] 的基础上讨论了代数约简与信息熵约简差异性。将代数约简与信息熵约简的数学模型统一写成基于条件概率的形式, 从语义上分析了它们在不一致决策表下不同的本质原因。对于信息熵协调集, 它保持条件概率不变, 从而对那些条件概率取值为 1 的对象也保持不变, 即必为代数协调集, 反之则不一定成立, 除非决策表是一致的。

参考文献:

- [1] PAWLAK Z. Rough sets. *International Journal of Information and Computer Science*, 1982, 11(5): 341 ~ 356.
- [2] 苗夺谦, 王 珏. 粗糙集理论中的概念与运算的信息表示[J]. *软件学报*, 1999, 10(2): 113 ~ 116.
MIAO Duoqian, WANG Jue. An information representation of the concepts and operations in rough set theory [J]. *Journal of Software*, 1999, 10(2): 113 ~ 116.
- [3] 王国胤, 于 洪, 杨大春. 基于条件信息熵的决策表约简[J]. *计算机学报*, 2002, 25(7): 759 ~ 766.
WANG Guoyin, YU Hong, YANG Dachun. Decision table reduction based on conditional information entropy [J]. *Chinese Journal of Computers*, 2002, 25(7): 759 ~ 766.
- [4] 张文修, 米据生, 吴伟志. 不协调目标信息系统的知识约简[J]. *计算机学报*, 2003, 26(1): 12 ~ 18.
ZHANG Wenxiu, MI Jusheng, WU Weizhi. Knowledge reductions in inconsistent information systems [J]. *Chinese Journal of Computers*, 2003, 26(1): 12 ~ 18.
- [5] 袁修久, 张文修. 决策表的分布约简和严凸函数下约简的等价性[J]. *系统工程*, 2003, 21(5): 5 ~ 7.
YUAN Xiujia, ZHANG Wenxiu. Studies on equivalence

of the distribution reduction and strictly convex function based reduction in decision tables [J]. *System Engineering*, 2003, 21(5): 5 ~ 7.

- [6] 王国胤, 安久江, 吴渝. Rough 集理论代数观与信息观的差异量化分析[J]. *小型微型计算机系统*, 2005, 26(7): 1187 ~ 1190.
WANG Guoyin, AN Jiujiang, WU Yu. Quantitative analysis of the difference between the algebra view and information view of rough set theory [J]. *Journal of Chinese Computer Systems*, 2005, 26(7): 1187 ~ 1190.
- [7] 黄国顺, 刘云生. 不一致决策表信息熵约简与代数约简的核计算与转化[J]. *小型微型计算机系统*, 2008, 29(2): 308 ~ 312.
HUANG Guoshun, LIU Yunsheng. Calculation and translation for core of information entropy and algebra reductions in inconsistent decision table [J]. *Journal of Chinese Computer Systems*, 2008, 29(2): 308 ~ 312.
- [8] 蒋思宇, 卢炎生. 两种新的决策表属性约简概念[J]. *小型微型计算机系统*, 2006, 27(3): 512 ~ 515.
JIANG Siyu, LU Yansheng. Two new definitions of decision table [J]. *Journal of Chinese Computer Systems*, 2006, 27(3): 512 ~ 515.
- [9] 徐章艳, 宋 威, 杨炳儒, 等. 关于“两种新的决策表属性约简概念”的注记[J]. *小型微型计算机系统*, 2007, 28(9): 1686 ~ 1689.
XU Zhangyan, SONG Wei, YANG Bingru, et al. Note on “two new definitions of decision table” [J]. *Journal of Chinese Computer Systems*, 2007, 28(9): 1686 ~ 1689.
- [10] 黄国顺. 基于数据库系统的决策表核与属性约简算法[J]. *计算机应用*, 2008, 28(5): 1180 ~ 1182.
HUANG Guoshun. Core and attribute reduction algorithms based on database system for decision table [J]. *Journal of Computer Applications*, 2008, 28(5): 1180 ~ 1182.
- [11] PAWLAK Z. GRZYMALA B J, SLOWINSKI R, et al. Rough sets [J]. *Communication of the ACM*, 1995, 38(11): 89 ~ 95.
- [12] QIAN Yuhua, LIANG Jiye, LI Deyu, et al. Measures for evaluating the decision performance of a decision table in rough set theory [J]. *Information Sciences*, 2008, 178: 181 ~ 202.