文章编号: 1000-6788(2010)09-1679-05 中图分类号: TP18 文献标志码: A

基于信息量的序信息系统的属性约简

马建敏 [1], 张文修 [2], 朱朝晖 [3]

(1. 长安大学 理学院 西安 710064; 2. 西安交通大学 理学院, 西安 710049;
3. 深圳卓成混凝土模块研究所, 深圳 518000)

摘　要 属性约简是粗糙集理论研究的核心内容之一. 在序信息系统中引入信息量和属性重要性, 给出它们与属性约简之间的关系. 针对序信息系统提出了一种基于信息量和属性重要性的属性约简算法, 讨论了算法的时间复杂度. 实例证明了该算法的有效性.

关键词 粗糙集; 序信息系统; 信息量; 属性重要性; 属性约简

Information quantity-based attribute reduction in ordered information systems

MA Jian-min[1], ZHANG Wen-xiu[2], ZHU Chao-hui[3]

(1. Faculty of Science, Chang'an University, Xi'an 710064, China; 2. Faculty of Science, Xi'an Jiaotong University, Xi'an 710049, China; 3. Research Institute of Shenzhen Zhuo Cheng Concrete Module, Shenzhen 518000, China)

Abstract One of the focuses of rough set theory is attribute reduction. In an ordered information system, the information quality and significance of attributes are defined. Relationships between them and attribute reductions are investigated, based on which, a heuristic algorithm for obtain attribute reductions is presented, and the time complexity of the algorithm is also shown. Finally, the validity of the algorithm have been depicted by an practical example.

Keywords rough set; ordered information system; information quality; significance of attribute; attribute reduction

1 引言

粗糙集理论是由波兰数学家 Pawlak 于 1982 年提出的一种数据分析理论[1−2]. 该理论由于能够分析处理不精确、不协调和不完备等信息引起人工智能工作者的广泛关注, 并被成功应用在机器学习与知识发现、数据挖掘、决策支持与分析、过程控制、模式识别等领域[3−4].

属性约简是粗糙集理论中的重要内容之一[1,5−6]. 所谓属性约简, 就是在保持信息系统的分类能力不变的前提下删除其中的冗余属性. 而当信息系统中的数据是随机采集时, 其冗余性更为普遍. 由于信息系统的属性约简并不唯一, 人们希望找出信息系统的所有约简或最小约简. 但寻找信息系统的最小约简是一个 NP-hard 问题[7]. 解决这类问题的一般方法是采用启发式搜索方法求出最优或次最优约简[8]. 苗夺谦等人[9−10]从信息角度, 给出了决策表中属性的重要性度量, 提出了基于互信息的知识相对约简的启发式算法. 王国胤等人研究了信息论观点和代数观点下的属性约简关系[11], 并提出了基于条件信息熵的决策表约简算法[12]. 梁吉业等人[13−14] 将信息论中信息量的概念[15] 引入到信息系统中, 提出了基于信息量的属性约简算法. 黄兵等人[16] 在不完备信息系统中引入了信息量和条件信息量, 给出了不完备信息系统的属性约简算法. 这些研究都是基于信息系统上的等价关系提出的. 然而, 现实中存在着大量的信息并不是基于等价关系的, 这极

收稿日期: 2009-06-23
资助项目: 国家自然科学基金 (10901025, 60703117); 中央高校基本科研业务费专项基金 (CHD2009JC028); 长安大学基础研究支持计划专项基金
作者简介: 马建敏 (1978–), 女, 山东日照人, 博士, 讲师, 研究方向为粗糙集, 概念格与粒计算; 张文修 (1940–), 男, 山西晋城人, 教授, 博士生导师, 研究方向为粗糙集、模糊集理论、信息科学的数学基础等.

大地限制了粗糙集的属性约简等方法的研究与应用. 1998 年 Greco 等人[17−18] 就信息系统中属性值排序问题提出了基于优势关系的粗糙集研究方法, 促进了粗集理论的应用和发展.

本文在基于优势关系的序信息系统中建立了信息量的概念, 给出了信息量的性质, 以及信息量与属性约简之间的关系. 在此基础上给出了属性重要性的定义, 研究了属性重要性与约简之间的关系. 进一步, 基于信息量和属性重要性, 给出了获取序信息系统中属性约简的算法, 并通过实例验证了该算法的有效性.

2 序信息系统

定义 1 称三元数组 (U,A,F) 为信息系统 (IS), 其中, $U=\{x_1,x_2,\cdots,x_n\}$ 是非空有限对象集合, 称为论域; $A=\{a_1,a_2,\cdots,a_m\}$ 是非空有限属性集合; $F=\{f_a:\ \forall a\in A\}$ 是 U 到 A 上的函数集合, 其中, $f_a:\ U\to V_a\ (\forall a\in A)$ 称为信息函数; V_a 是属性 a 的值域.

设 $\leq$ 是论域 U 上的二元关系. 若 $\leq$ 满足自反、反对称及传递的, 则称 $(U,\leq)$ 为半序集. 对任意的 $x,y\in U$, 如果 $x\leq y$ 或 $y\leq x$ 之一成立时, 则称 x,y 是可比的, 否则称为不可比的. 如果半序集 $(U,\leq)$ 的一个子集中的任两个元素都是可比的, 则称这个子集为链[19].

定义 2 设 (U,A,F) 为信息系统. 对任意属性 $a\in A$, 若 a 的属性值 V_a 在偏序关系 $\leq$ 下构成链, 则称 a 是规范的. 若对任意的 $a\in A$, a 都是规范的, 则称 (U,A,F) 为序信息系统 (OIS).

定义 3 设 (U,A,F) 为序信息系统. 对任意的 $B\subseteq A$, 定义 U 上的二元关系

$$R_B^{\leqq}=\{(x,y)\in U\times U:\ f_a(x)\leq f_a(y),\quad \forall a\in B\},$$

称 $R_B^{\leqq}$ 为序信息系统 (U,A,F) 上的优势关系. 显然, $R_B^{\leqq}$ 是自反、传递的, 但不是对称的, 故不是等价关系, 记 $R_B^{\leqq}(x)=\{y\in U:\ (x,y)\in R_B^{\leqq}\}$, 称为 x 的优势类. 则 $U/R_B^{\leqq}=\{R_B^{\leqq}(x):\ x\in U\}$ 构成 U 的一个覆盖.

性质 1[6] 设 (U,A,F) 为序信息系统, $R_B^{\leqq}$ 为 (U,A,F) 上的优势关系. 则对任意的 $x,y\in U$,

1) 若 $B\subseteq A$, 则 $R_A^{\leqq}\subseteq R_B^{\leqq}$ 且 $R_B^{\leqq}=\bigcap\limits_{a\in B}R_a^{\leqq}$;

2) 若 $B\subseteq A$, 则 $R_A^{\leqq}(x)\subseteq R_B^{\leqq}(x)$ 且 $R_B^{\leqq}(x)=\bigcap\limits_{a\in B}R_a^{\leqq}(x)$;

3) 若 $y\in R_B^{\leqq}(x)$, 则 $R_B^{\leqq}(y)\subseteq R_B^{\leqq}(x)$ 且 $R_B^{\leqq}(x)=\cup\{R_B^{\leqq}(y):\ y\in R_B^{\leqq}(x)\}$;

4) $R_B^{\leqq}(y)=R_B^{\leqq}(x)\Leftrightarrow\forall a\in B,\ f_a(x)=f_a(y)$.

下面引入信息系统中的几个概念[1−2].

定义 4 设 (U,A,F) 是序信息系统, $a\in A$. 如果 $R_{A-\{a\}}^{\leqq}=R_A^{\leqq}$, 则称属性 a 在 A 中是不必要的 (或冗余的); 否则, 称 a 在 A 中是必要的.

不必要的属性在序信息系统中是冗余的, 如果将它从信息系统中去掉, 不会改变序信息系统的分类能力; 相反, 若从序信息系统中去掉一个必要的属性, 则一定改变该序信息系统的分类.

定义 5 设 (U,A,F) 是序信息系统. 如果每个属性 $a\in A$ 在 A 中都是必要的, 则称属性集 A 是独立的; 否则, 称 A 是相依的.

对于相依的属性集来说, 其中包含有多余属性, 可以对其约简.

定义 6 设 (U,A,F) 是序信息系统, A 中所有必要属性组成的集合称为属性集 A 的核, 记作 $\mathrm{Core}(A)$.

定义 7 设 (U,A,F) 是序信息系统, $B\subseteq A$. 如果 $R_B^{\leqq}=R_A^{\leqq}$, 则称 B 为序信息系统 (U,A,F) 的协调集; 若 B 为 (U,A,F) 的协调集, 且对任意的 $a\in B$, $R_{B-\{a\}}^{\leqq}\neq R_A^{\leqq}$, 称 B 为序信息系统 (U,A,F) 的约简.

易证, $\mathrm{Core}(A)=\cap\{D:\ D\subseteq A,\ D$ 是 (U,A,F) 的约简$\}$.

3 知识的信息量及属性重要性

文献 [13] 将信息论中信息量的概念[15] 引入到信息系统中, 给出了信息系统的属性约简方法. 下面将文献 [13] 中的信息量概念引入到序信息系统中, 讨论序信息系统的属性约简.

定义 8 设 (U,A,F) 是序信息系统, $B\subseteq A$, $U/R_B^{\leqq}=\{R_B^{\leqq}(x_i):\ x_i\in U\}$. 定义 B 的信息量为

$$I(B)=1-\frac{1}{|U|^2}\sum_{i=1}^{|U|}|R_B^{\leqq}(x_i)|,$$

其中, $|X|$ 表示集合 X 的基数.

性质 2 设 (U,A,F) 是序信息系统. 则对任意的 $B\subseteq A$, $I(B)\leq I(A)$.

证明　设 $B \subseteq A$, 由性质 1 知, 对任意的 $x_i \in U$, $R_A^{\leqslant}(x_i) \subseteq R_B^{\leqslant}(x_i)$. 于是 $|R_A^{\leqslant}(x_i)| \leq |R_B^{\leqslant}(x_i)|$. 从而由定义 8 即得 $I(B) \leq I(A)$.

由此性质即得协调集的判定定理.

定理 1　设 (U, A, F) 是序信息系统. 则对任意的 $B \subseteq A$,

$$B \text{ 是 } (U, A, F) \text{ 的协调集 } \Leftrightarrow I(A) = I(B).$$

定理 2　设 (U, A, F) 是序信息系统. 则对任意的 $a \in A$,

$$a \text{ 是必要的 } \Leftrightarrow I(A - \{a\}) < I(A).$$

证明　设 a 在A 中是必要的. 则 $R_{A-\{a\}}^{\leqslant} \neq R_A^{\leqslant}$. 从而存在 $x \in U$, 使得 $R_{A-\{a\}}^{\leqslant}(x) \neq R_A^{\leqslant}(x)$. 由 $R_A^{\leqslant} \subseteq R_{A-\{a\}}^{\leqslant}$ 知对任意 $x \in U$, $R_A^{\leqslant}(x) \subset R_{A-\{a\}}^{\leqslant}(x)$. 故由定义 8 可得 $I(A - \{a\}) < I(A)$. 反之, 设 $a \in A$, 且 $I(A - \{a\}) < I(A)$. 则由定义 8 知 $1 - \frac{1}{|U|^2} \sum_{i=1}^{|U|} |R_{A-\{a\}}^{\leqslant}(x_i)| < 1 - \frac{1}{|U|^2} \sum_{i=1}^{|U|} |R_A^{\leqslant}(x_i)|$. 由 $R_A^{\leqslant} \subseteq R_{A-\{a\}}^{\leqslant}$ 可知, 存在 $x_i \in U$ 使得 $R_{A-\{a\}}^{\leqslant}(x_i) \neq R_A^{\leqslant}(x_i)$.　于是 $R_{A-\{a\}}^{\leqslant} \neq R_A^{\leqslant}$. 即 a 为必要的.

性质 3　设 (U, A, F) 是序信息系统. 则对任意的 $B \subseteq A$, B 是独立的 $\Leftrightarrow \forall a \in B,\ I(B - \{a\}) < I(B)$.

定理 3　设 (U, A, F) 是序信息系统. 则 $\mathrm{Core}(A) = \{a \in A:\ I_{A-\{a\}}(a) < I(A)\}$.

证明　$a \in \mathrm{Core}(A) \Leftrightarrow A - \{a\}$ 是不协调的 $\Leftrightarrow R_{A-\{a\}}^{\leqslant} \neq R_A^{\leqslant} \Leftrightarrow R_A^{\leqslant} \subset R_{A-\{a\}}^{\leqslant} \Leftrightarrow I(A) > I(A - \{a\})$.

定理 2, 3 给出了利用信息量判定序信息系统的属性特征的方法.

定理 4　设 (U, A, F) 是序信息系统. 则对任意的 $B \subseteq A$,

$$B \text{ 是约简 } \Leftrightarrow I(B) = I(A),\ \text{且 } \forall a \in B,\ I(B - \{a\}) < I(A).$$

通过信息量, 不仅可以得到序信息系统的协调集及属性约简的判定定理, 还可获得必要属性的属性特征. 但对于其他属性是否重要, 重要性程度如何却难以度量. 为此, 在信息量基础上引入属性重要性的概念. 以信息量的变化大小来量度属性的重要性程度.

定义 9　设 (U, A, F) 是序信息系统. 属性 $a \in A$ 在 A 中的重要性定义为:

$$Sig_{A-\{a\}}(a) = I(A) - I(A - \{a\}).$$

特别地, 当 $A = \{a\}$, 用 $Sig(a)$ 表示 $Sig_{\emptyset}(a)$, 则 $Sig(a) = I(\{a\})$, 其中, $U/R_{\emptyset}^{\leqslant} = U$, $I(\emptyset) = 0$.

上述定义表明属性 $a \in A$ 在 A 中的重要性由 A 中去掉 a 后引起的信息量变化的大小来度量.

性质 4　设 (U, A, F) 是序信息系统. 则对任意的 $a \in A$,

$$0 \leq Sig_{A-\{a\}}(a) \leq 1 - \frac{1}{|U|}.$$

由定理 2 及定义 9 即得属性特征与属性重要性的关系.

定理 5　设 (U, A, F) 是序信息系统. 则对任意的 $a \in A$,

$$a \text{ 在} A \text{ 中是必要的 } \Leftrightarrow Sig_{A-\{a\}}(a) > 0.$$

定理 6　设 (U, A, F) 是序信息系统. 则对任意的 $B \subseteq A$,

$$B \text{ 是约简 } \Leftrightarrow I(B) = I(A),\ \text{且对任意的 } a \in B,\ Sig_{B-\{a\}}(a) > 0.$$

定理 7　设 (U, A, F) 是序信息系统. 则 $\mathrm{Core}(A) = \{a \in A:\ Sig_{A-\{a\}}(a) > 0\}$.

4　基于信息量的序信息系统属性约简算法

由定理 5 及定理 7 可知, 序信息系统 (U, A, F) 的核 $\mathrm{Core}(A)$ 就是属性重要性大于零的那些属性, 即必要属性构成的集合. 由于核是唯一的且核为任何约简的子集, 因此, 核可作为求最小约简的基础. 由定理 6 知, 在核 $\mathrm{Core}(A)$ 的基础上, 只需找到集合 $B \subseteq A$, 满足 $\mathrm{Core}(A) \subseteq B$, 且对任意的 $a \in B, Sig_{B-\{a\}}(a) > 0$, 即 $\forall a \in B - \mathrm{Core}(A),\ Sig_{B-\{a\}}(a) > 0$ 的属性即可. 由定义 9, 给出下面定义:

定义 10 设 (U, A, F) 是序信息系统, $B \subseteq A$. 对任意属性 $a \in A - C$, a 关于属性集 C 的重要性定义为:

$$Sig_C(a) = Sig_{C \cup \{a\} - \{a\}}(a) = I(C \cup \{a\}) - I(C).$$

上述定义表明不在属性集 C 中的属性 $a \in A - C$ 关于属性集 C 的重要性由 C 中添加 a 后所引起的信息量的变化大小来度量. $Sig_C(a)$ 的值越大, 说明属性 $a \in A - C$ 关于属性集 C 就越重要. 从而可把 $Sig_C(a)$ 作为寻找最小属性约简的启发式信息, 以减少搜索空间.

由定义 10, 在由属性重要性获得核 $\mathrm{Core}(A)$ 的基础上, 逐次选择 A 中核中元素之外的最重要的属性添加到核中去, 直到其信息量等于整个属性集 A 的信息量, 即可获得序信息系统的属性约简.

序信息系统的核与属性约简算法:

输入 序信息系统 (U, A, F).

输出 序信息系统的核与约简.

Step 1 计算序信息系统中知识 A 的信息量 $I(A)$.

Step 2 $\mathrm{Core}(A) := \emptyset$. 计算每个属性 $a \in A$ 在 A 中的重要性 $Sig_{A-\{a\}}(a)$. 若 $Sig_{A-\{a\}}(a) > 0$, 则 $\mathrm{Core}(A) := \mathrm{Core}(A) \cup \{a\}$. 最后得到的 $\mathrm{Core}(A)$ 为属性集 A 的核.

Step 3 计算核 $\mathrm{Core}(A)$ 的信息量. 若 $I(\mathrm{Core}(A)) = I(A)$, 则输出核 $\mathrm{Core}(A)$ 即为序信息系统的属性约简 (此时 $\mathrm{Core}(A)$ 为 (U, A, F) 最小约简); 否则, 若 $I(\mathrm{Core}(A)) < I(A)$, 执行 Step 4.

Step 4 令 $C = \mathrm{Core}(A)$, 对属性集 $A - C$ 重复执行:

1) 对每个属性 $a \in A - C$, 计算属性重要性 $Sig_C(a)$;

2) 选择属性 a 使其满足 $Sig_C(a) = \max_{a' \in A-C} Sig_C(a')$, $C := C \cup \{a\}$.

3) 若 $I(C) = I(A)$, 则输出 C(此时 C 为 (U, A, F) 的一个属性约简); 否则, 转 1).

下面分析上述算法的时间复杂性:

计算核 $\mathrm{Core}(A)$ 共需计算 $|A|$ 次 $Sig_{A-\{a\}}(a)$.

计算属性约简需要计算 $Sig_C(a)$ 的次数最多为: $|A| + (|A| - 1) + \cdots + 1 = |A|(|A| + 1)/2 = O(|A|^2)$.

为了计算 $Sig_{A-\{a\}}(a)$(计算 $Sig_C(a)$ 与计算 $Sig_{A-\{a\}}(a)$ 的时间复杂性相同), 要求下列计算:

1) 计算 $|A|$ 个覆盖.

类似文献 [20] 可知, 计算每一个覆盖的时间复杂性为 $O(|U|^2)$, 因此计算 $|A|$ 个覆盖的时间复杂性为 $O(|A| \times |U|^2)$.

2) 为了计算 $U/R_A^{\leq}$ 和 $U/R_{A-\{a\}}^{\leq}$, 需要计算 $|A|-1$ 和 $|A|-2$ 次交. 计算一次交的时间复杂性为 $O(|U|^2)$. 因此计算这些交的时间复杂性为 $(|A| - 1 + |A| - 2) \times O(|U|^2) = O(|A| \times |U|^2)$. 因此, 计算一次 $Sig_{A-\{a\}}(a)$ 的时间复杂性为 $O(|A| \times |U|^2)$.

计算 $I(A)$ 和 $I(A - a)$ 的时间复杂性为 $O(|A| \times |U|^2)$.

整个算法的时间复杂性为: $(|A| + |A|(|A| + 1)/2) \times O(|A| \times |U|^2) = O(|A|^3 \times |U|^2)$.

例 1 表 1 给出了信息系统 (U, A, F), 其中 $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, $A = \{a_1, a_2, a_3, a_4\}$.

表 1 序信息系统 (U, A, F)

U	a_1	a_2	a_3	a_4
x_1	1	2	1	2
x_2	3	2	2	2
x_3	1	1	2	1
x_4	2	1	3	1
x_5	3	3	2	3
x_6	3	2	3	2

由表 1 知, 对任意的属性 $a \in A$, 属性值域 $V_a = \{1, 2, 3\}$ 构成链, 即 $a \in A$ 都是规范的. 从而信息系统 (U, A, F) 是序信息系统.

Step 1 由定义 3, $R_B^{\leq} = \{(x, y) \in U \times U :\ f_a(x) \leq f_a(y)\}$, 即得由属性集 A 构成的优势类:

$$R_A^{\leq}(x_1) = \{x_1, x_2, x_5, x_6\},\ R_A^{\leq}(x_2) = \{x_2, x_5, x_6\},\ R_A^{\leq}(x_3) = \{x_2, x_3, x_4, x_5, x_6\},$$
$$R_A^{\leq}(x_4) = \{x_4, x_6\},\ R_A^{\leq}(x_5) = \{x_5\},\ R_A^{\leq}(x_6) = \{x_6\}.$$

从而可得知识 A 的信息量 $I(A) = 1 - \frac{1}{|U|^2} \sum_{i=1}^{6} |R_A^{\leq}(x_i)| = \frac{5}{9}$.

Step 2 由定义 9 求得属性 a_i 在 A 中的重要性为:

$$Sig_{A-\{a_1\}}(a_1) = I(A) - I(A - \{a_1\}) = \frac{5}{9} - \frac{5}{9} = 0,\ Sig_{A-\{a_2\}}(a_2) = I(A) - I(A - \{a_2\}) = \frac{5}{9} - \frac{5}{9} = 0,$$
$$Sig_{A-\{a_3\}}(a_3) = I(A) - I(A - \{a_3\}) = \frac{5}{9} - \frac{5}{12} = \frac{5}{36},\ Sig_{A-\{a_4\}}(a_4) = I(A) - I(A - \{a_4\}) = \frac{5}{9} - \frac{5}{9} = 0.$$

故由定理 7 知, $\mathrm{Core}(A) = \{a_3\}$. 且 $I(\mathrm{Core}(A)) = I(\{a_3\}) = \frac{11}{36}$.

由于 $I(\mathrm{Core}(A)) \neq I(A)$, 执行 Step 3.

Step 3 令 $C = \mathrm{Core}(A) = \{a_3\}$. 对 $A - C = \{a_1, a_2, a_4\}$ 计算各个属性关于属性 C 的重要性:

1) $Sig_C(a_1) = I(\{a_1, a_3\}) - I(a_3) = \frac{4}{9} - \frac{11}{36} = \frac{5}{36}$, $Sig_C(a_2) = I(\{a_2, a_3\}) - I(a_3) = \frac{5}{9} - \frac{11}{36} = \frac{1}{4}$, $Sig_C(a_4) = I(\{a_3, a_4\}) - I(a_3) = \frac{5}{9} - \frac{11}{36} = \frac{1}{4}$.

2) 由于 $Sig_C(a_2) = Sig_C(a_4) = \max\limits_{a \in A-C} Sig_C(a) = \frac{1}{4}$, 取 $C_1 := C \cup \{a_2\}$, $C_2 := C \cup \{a_4\}$.

3) 对任意的 $i=1,2$, $I(C_i)=\frac{5}{9}$, 且 $I(C_i)=I(A)$.
故核 $\mathrm{Core}(A)=\{a_3\}$, $C_1=\{a_2,a_3\}$, $C_2=\{a_3,a_4\}$ 均为序信息系统的约简.

5 结论

属性约简是寻找保持分类能力不变的最小属性子集. 为了获取属性约简, 人们提出了基于信息熵、信息量等的属性约简算法. 本文针对优势关系构成的序信息系统提出了知识的信息量概念, 研究了基于信息量的协调集和属性约简的判定方法. 进一步, 定义了序信息系统的属性重要性, 给出了基于信息量和属性重要性的属性约简算法. 通过实例说明, 该算法能得到序信息系统的属性约简.

参考文献

[1] Pawlak Z. Rough sets[J]. International Journal of Computer and Information Science, 1982(11): 341–356.
[2] Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning About Data[M]. Boston: Kluwer Academic Publishers, 1991.
[3] Pawlak Z, Grzymala-Busse J W, Slowinski R, et al. Rough sets[J]. Communication of the ACM, 1995, 38(11): 89–95.
[4] Pawlak Z. Rough set theory and its application to data analysis[J]. Cybernetics and Systems, 1998(9): 661–668.
[5] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法 [M]. 北京: 科学出版社, 2001.
Zhang W X, Wu W Z, Liang J Y, et al. Rough Set Theory and Approaches[M]. Beijing: Science Press, 2001.
[6] 张文修, 梁怡, 吴伟志. 信息系统与知识发现 [M]. 北京: 科学出版社, 2003.
Zhang W X, Liang Y, Wu W Z. Information System and Knowledge Discovery[M]. Beijing: Science Press, 2003.
[7] Wang S K M, Ziarko W. On optimal decision rules in decision tables[J]. Bulletin of Polish Academy of Sciences, 1985, 33: 693–676.
[8] Miao D Q, Wang J. Information-based algorithm for reduction of knowledge[C]//IEEE International Conference on Intelligent Processing Systems, 1997: 1155–1158.
[9] 苗夺谦, 王珏. 粗糙集理论中概念与运算的信息表示 [J]. 软件学报, 1999, 10(2): 113–116.
Miao D Q, Wang J. An information representation of the concepts and operations in rough set theory[J]. Journal of Software, 1999, 10(2): 113–116.
[10] 苗夺谦, 胡桂荣. 知识约简的一种启发式算法 [J]. 计算机研究与发展, 1999, 36(6): 681–684.
Miao D Q, Hu G R. A heuristic algorithm for reduction of knowledge[J]. Journal of Computer Research & Development, 1999, 36(6): 681–684.
[11] Wang G Y. Algebra view and information view of rough sets theory[C]//Data Mining and Knowledge Discovery: Theory, Tools, and Technology, Proceedings of SPIE, 2001, 4384: 200–207.
[12] 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 G Y, Yu H, Yang D C. Decision table reduction based on conditional information entropy[J]. Chinese Journal of Computer, 2002, 25(7): 759–766.
[13] 梁吉业, 曲开社, 徐宗本. 信息系统的属性约简 [J]. 系统工程理论与实践, 2001, 21(12): 76–80.
Liang J Y, Qu K S, Xu Z B. Reduction of attribute in information systems[J]. Systems Engineering — Theory & Practice, 2001, 21(12): 76–80.
[14] 梁吉业, 李德玉. 信息系统中的不确定性与知识获取 [M]. 北京: 科学出版社, 2005.
Liang J Y, Li D Y. The Uncertainty and Knowledge Acquiring in Information Systems[M]. Beijing: Science Press, 2005.
[15] 张文修, 梁怡, 徐萍. 基于包含度的不确定性推理原 [M]. 北京: 清华大学出版社, 2007.
Zhang W X, Liang Y, Xu P. Uncertainty Reasoning Based on Inclusion Degree[M]. Beijing: Tsinghua University Press, 2007.
[16] 黄兵, 周献中, 张蓉蓉. 基于信息量的不完备信息系统属性约简 [J]. 系统工程理论与实践, 2005, 25(4): 55–60.
Huang B, Zhou X Z, Zhang R R. Attribute reduction based on information quantity under incomplete information systems[J]. Systems Engineering — Theory & Practice, 2005, 25(4): 55–60.
[17] Greco S, Matarazzo B, Slowingski R. Rough approximation of a preference relation by dominance by relation[J]. European Journal of Operation Research, 1999, 117: 63–83.
[18] Greco S, Matarazzo B, Slowingski R. A new rough set approach set approach to multicriteria and multiattribute classification[C]//Rough Sets and Current Trends in Computing (RSCTC'98), Lecture Notes in Artificial Intelligence, Berlin: Springer-Verlag, 1424, 1998: 60–67.
[19] Ganter B, Wille R. Formal Concept Analysis: Mathematical Foundations[M]. Berlin: Springer, 1999.
[20] Guan J W, Bell D A, Guan Z. Matrix computation for information systems[J]. Information Sciences, 2001, 131: 129–156.