

DOI:10.11992/tis. 21704025

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20170703.1853.010.html>

广义分布保持属性约简研究

高学义^{1,2}, 张楠^{1,2}, 童向荣^{1,2}, 姜丽丽^{1,2}

(1. 烟台大学 数据科学与智能技术山东省高校重点实验室, 山东 烟台 264005; 2. 烟台大学 计算机与控制工程学院, 山东 烟台 264005)

摘要: 属性约简是粗糙集理论的重要研究内容之一。分布约简保证约简前后每个对象的概率分布保持不变, 即保证每条规则的置信度在约简前后不发生改变。实际应用中, 人们往往更加关注可信度较高或较低的规则。因此, 在本文中引入了广义分布保持属性约简, 该属性约简可以保证规则的置信度 $P(P \in [0, \alpha] \text{ 或 } [\beta, 1])$ 在约简前后不变。同时, 给出了广义分布保持属性约简的判定方法与基于差别矩阵的广义分布保持属性约简算法, 深入讨论了几种特殊情形下的广义分布保持约简。最后, 在 4 个 UCI 数据集上进行的实验分析表明, 几种特殊情形下的广义分布保持属性约简可退化为已有的一些属性约简, 且在不同置信区间下求得的广义分布保持属性约简存在包含关系, 验证了相关结论的正确性。

关键词: 分布保持; 属性约简; 粗糙集; 概率分布; 差别矩阵

中图分类号: TP181 **文献标志码:** A **文章编号:** 1673-4785(2017)03-0377-09

中文引用格式: 高学义, 张楠, 童向荣, 等. 广义分布保持属性约简研究[J]. 智能系统学报, 2017, 12(3): 377-385.

英文引用格式: GAO Xueyi, ZHANG Nan, TONG Xiangrong, et al. Research on attribute reduction using generalized distribution preservation[J]. CAAI transactions on intelligent systems, 2017, 12(3): 377-385.

Research on attribute reduction using generalized distribution preservation

GAO Xueyi^{1,2}, ZHANG Nan^{1,2}, TONG Xiangrong^{1,2}, JIANG Lili^{1,2}

(1. Key Lab for Data Science and Intelligent Technology of Shandong Higher Education Institutes, Yantai University, Yantai 264005, China; 2. School of Computer and Control Engineering, Yantai University, Yantai 264005, China)

Abstract: Attribute reduction is a pertinent issue in rough set theory. Distribution reduction ensures that the probability distribution of each target does not change before and after reduction; i.e., it ensures that the confidence of every rule remains unchanged before and after reduction. In actual applications, people are often interested in rules that have higher or lower confidences. Thus, attribute reduction based on generalized distribution preservation is proposed in this paper. Confidences in $[0, \alpha]$ or $[\beta, 1]$ were unchanged using the proposed technique. We also propose judgment methods for generalized-distribution-preservation attribute reduction and investigate the generalized attribute-reduction algorithm based on a discernibility matrix. Some special cases with respect to generalized-distribution-preservation attribute reduction are discussed in depth. Finally, experiments on four data sets downloaded from UCI show that some special cases with respect to generalized distribution preservation reduction could degenerate into some existing attribute reductions and inclusion relations exist in generalized distribution preservation attribute reduction under different confidence intervals, verifying the correctness of the relevant conclusions.

Keywords: distribution preservation; attribute reduction; rough sets; probability distribution; discernibility matrix

粗糙集理论是由波兰学者 Pawlak 教授于 1982

年提出的一种用于处理和分析不确定、不精确数据的数学方法与工具^[1-4]。目前, 粗糙集理论在机器学习、决策分析、模式识别、数据挖掘和智能信息处理等领域得到了广泛应用。

属性约简或知识约简是粗糙集理论的重要研

收稿日期: 2017-04-19. 网络出版日期: 2017-07-03.

基金项目: 国家自然科学基金项目(61403329, 61572418, 61502410, 61572419); 山东省自然科学基金项目(ZR2013FQ020, ZR2015PF 010); 山东省高等学校科技计划项目(J15LN09, 116LN17).

通信作者: 张楠. E-mail: zhangnan0851@163.com.

究内容之一,其本质是获取保持知识库某种分类能力在约简前后不发生变化的最小属性子集描述,国内外学者做了大量的相关研究工作。1992年,Skowron^[5]提出了差别矩阵的概念,为获取信息系统或决策表的所有约简或最小约简提供了理论基础;1998年,Kryszkiewicz^[6]讨论了基于差别矩阵的不完备信息系统广义决策保持属性约简问题;2003年,张文修等^[7]给出了分布约简和分配约简的差别矩阵约简方法,并提出了最大分布约简;2007年,徐伟华等^[8]给出了优势关系下基于差别矩阵的分布约简和最大分布约简;2009年,苗夺谦等^[9]提出了不可分辨关系保持属性约简和相应的差别矩阵构造方法;2010年,张楠等^[10]讨论了区间值信息系统下的属性约简问题。为了提高属性约简的算法效率,多种启发式属性约简算法相继被提出。1999年,苗夺谦等^[11]从信息论的角度给出了属性重要度的度量方法,在此基础上提出了基于互信息的启发式约简算法;2002年,王国胤等^[12]提出了基于条件信息熵的启发式属性约简算法;2010年,钱宇华等^[13]提出了正向近似的基本概念并将其应用于启发式属性约简的构造过程,提高了属性约简的计算效率;2011年,钱宇华等^[14-15]进一步将正向近似应用于不完备决策表的启发式属性约简,改善了不完备决策表下启发式属性约简的求取效率;陈红梅等^[16-17]在动态属性约简方面做了大量的研究工作;文献^[18-19]对现有的属性约简之间的关系进行了深入讨论与研究。

分布约简保证每个对象在约简前后的概率分布保持不变,即保证每条规则的置信度在约简前后不发生改变。在实际应用中,人们往往更关注可信度较高或较低的规则^[20],分布约简的标准过于严格,很多对实际决策无用的规则的置信度在约简前后也要保持不变,很可能使得最终约简过于冗长,对实际决策造成一定的干扰。本文在分布约简的基础上,通过弱化分布约简的约简标准,提出了一种新的属性约简,即广义分布保持属性约简,该属性约简可以保证规则的置信度($P \in [0, \alpha]$ 或 $[\beta, 1]$)在约简前后不变,并对广义分布保持属性约简的方法和相关性质进行了研究和讨论。

1 广义分布保持属性约简

定义 1^[7] 设决策表 $DT=(U, AT \cup D, V, f)$, 论域 $U=\{u_1, u_2, \dots, u_n\}$, 则 $\forall u_i \in U, A \subseteq AT$, 对象 u_i 在属性 A 下关于决策属性集 D 的概率分布定义为 $\mu_A(u_i) =$

$(P(D_1|[u_i]_A), P(D_2|[u_i]_A), \dots, P(D_{|U/D|}|[u_i]_A))$ 。
其中, $P(D_j|[u_i]_A) = |D_j \cap [u_i]_A| / |[u_i]_A|, i \in \{1, 2, \dots, n\}, j \in \{1, 2, \dots, |U/D|\}$ 。

定义 2 设决策表 $DT=(U, AT \cup D, V, f), U=\{u_1, u_2, \dots, u_n\}, U/D=\{D_1, D_2, \dots, D_{|U/D|}\}$, 记 d_j 为 D_j 对应的决策值, 则 $\forall u_i \in U, A \subseteq AT, u_i$ 在 A 下关于决策属性 D 的 $[\alpha, \beta]$ 决策-置信度序偶集定义为

$$Y_A^{[\alpha, \beta]}(u_i) = \{ \langle d_j, P(D_j|[u_i]_A) \rangle \}$$

$\alpha \leq P(D_j|[u_i]_A) \leq \beta \wedge \alpha \leq P(D_j|[u_i]_{AT}) \leq \beta$
式中: $i \in \{1, 2, \dots, n\}, j \in \{1, 2, \dots, |U/D|\}, \alpha$ 和 β 满足 $(\alpha=0 \wedge \beta \in [0, 1])$ 或 $(\alpha \in [0, 1] \wedge \beta=1)$ 。

根据定义 2, 若对于 $\forall u \in U$, 均满足 $Y_A^{[\alpha, \beta]}(u) = Y_{AT}^{[\alpha, \beta]}(u)$, 则称 A 是广义分布协调集。若 A 是广义分布协调集, 且 A 的任意子集不是广义分布协调集, 则称 A 为广义分布保持约简。据此, 给出广义分布保持约简的形式化定义如下。

定义 3 给定决策表 $DT=(U, AT \cup D, V, f), U=\{u_1, u_2, \dots, u_n\}, \forall A \subseteq AT$, 若 A 是一个广义分布保持约简, 当且仅当以下两个条件成立:

$$1) \forall u_i \subseteq U, Y_A^{[\alpha, \beta]}(u_i) = Y_{AT}^{[\alpha, \beta]}(u_i);$$

$$2) \forall B \subset A, Y_B^{[\alpha, \beta]}(u_j) \neq Y_A^{[\alpha, \beta]}(u_j)。$$

式中: α 和 β 满足 $(\alpha=0 \wedge \beta \in [0, 1])$ 或 $(\alpha \in [0, 1] \wedge \beta=1), i, j \in \{1, 2, \dots, n\}$ 。

由定义 3 可知, 对于置信度在 $[\alpha, \beta]$ 内的规则, 它们的置信度在广义分布保持约简前后保持不变。

2 广义分布保持属性约简的判定与方法

首先, 给出广义分布协调集的等价证明。

定理 1 设决策表 $DT=(U, AT \cup D, V, f), U=\{u_1, u_2, \dots, u_n\}, \forall A \subseteq AT$, 则 A 是广义分布协调集当且仅当对于 $\forall u_i, u_j \in U$, 当 $Y_{AT}^{[\alpha, \beta]}(u_i) \neq Y_{AT}^{[\alpha, \beta]}(u_j)$ 成立时, 有 $[u_i]_A \cap [u_j]_A = \emptyset$ 。其中, α 和 β 满足 $(\alpha=0 \wedge \beta \in [0, 1])$ 或 $(\alpha \in [0, 1] \wedge \beta=1)$ 。

证明 不妨记 $\rho([u_i]_A) = \{[u_j]_{AT} : [u_j]_{AT} \subseteq [u_i]_A\}$, 其中 $i, j \in \{1, 2, \dots, n\}$ 。由于 $A \subseteq AT$, 故 $\rho([u_i]_A)$ 构成 $[u_i]_A$ 的一个划分。

“ \Rightarrow ”: 设 A 是决策表 DT 上的广义分布协调集。 $\forall u_i, u_j \in U$, 当 $[u_i]_A \cap [u_j]_A \neq \emptyset$ 时, 有 $[u_i]_A = [u_j]_A$ 。因此 $Y_A^{[\alpha, \beta]}(u_i) = Y_A^{[\alpha, \beta]}(u_j)$ 。但有 $Y_{AT}^{[\alpha, \beta]}(u_i) = Y_A^{[\alpha, \beta]}(u_i)$ 成立, 并且有 $Y_{AT}^{[\alpha, \beta]}(u_j) = Y_A^{[\alpha, \beta]}(u_j)$, 从而 $Y_{AT}^{[\alpha, \beta]}(u_i) = Y_{AT}^{[\alpha, \beta]}(u_j)$ 。因此, 若 $Y_{AT}^{[\alpha, \beta]}(u_i) \neq Y_{AT}^{[\alpha, \beta]}(u_j)$, 有 $[u_i]_A \cap [u_j]_A = \emptyset$ 。

“ \Leftarrow ”: 对于 $\forall u_i \in U$, 当 $[u_j]_{AT} \subseteq [u_i]_A$ 时, 有

$[u_i]_A \cap [u_j]_A \neq \emptyset$, 故 $Y_{AT}^{[\alpha,\beta]}(u_i) = Y_{AT}^{[\alpha,\beta]}(u_j)$ 。即对于 $\forall D_k \in U/D$, 记 d_k 为 D_k 对应的决策值, 若 $\langle d_k, P(D_k | [u_i]_{AT}) \rangle \in Y_{AT}^{[\alpha,\beta]}(u_i)$, 则必有 $\langle d_k, P(D_k | [u_j]_{AT}) \rangle \in Y_{AT}^{[\alpha,\beta]}(u_j)$ 成立, 并且 $P(D_k | [u_i]_{AT}) = P(D_k | [u_j]_{AT})$, 其中, $k \in \{1, 2, \dots, |U/D|\}$ 。为后续证明方便, 不妨记 $\rho([u_i]_A) = \{\rho_1, \rho_1, \dots, \rho_{|\rho([u_i]_A)|}\}$ 。由于 $\rho([u_i]_A) = \{[u_i]_{AT} : [u_i]_{AT} \subseteq [u_i]_A\}$, 故有

$$P(D_k | [u_i]_A) = \frac{\sum_{s=1}^{|\rho([u_i]_A)|} \{|\rho_s \cap D_k| : \rho_s \in \rho([u_i]_A)\}}{|[u_i]_A|} = \frac{\sum_{s=1}^{|\rho([u_i]_A)|} \left\{ \frac{|\rho_s \cap D_k|}{|\rho_s|} \cdot \frac{|\rho_s|}{|[u_i]_A|} : \rho_s \in \rho([u_i]_A) \right\}}{\sum_{s=1}^{|\rho([u_i]_A)|} \left\{ P(D_k | \rho_s) \cdot \frac{|\rho_s|}{|[u_i]_A|} : \rho_s \in \rho([u_i]_A) \right\}} = \frac{\sum_{s=1}^{|\rho([u_i]_A)|} \left\{ P(D_k | [u_i]_{AT}) \cdot \frac{|\rho_s|}{|[u_i]_A|} : \rho_s \in \rho([u_i]_A) \right\}}{P(D_k | [u_i]_{AT})}$$

因此, $Y_{AT}^{[\alpha,\beta]}(u_i) = Y_A^{[\alpha,\beta]}(u_i)$, 从而 A 是广义分布协调集, 证毕。

定理1给出了判断属性子集是广义分布协调集的方法, 由此可进一步得到广义分布保持约简的方法, 在此可给出广义分布差别矩阵的概念。

定义4 给定决策表 $DT = (U, AT \cup D, V, f)$, $U = \{u_1, u_2, \dots, u_n\}$, $AT = \{a_1, a_2, \dots, a_{|AT|}\}$ 为条件属性集, $D = \{d\}$ 为决策属性, 记

$$D^* = \{(u_i, u_j) | Y_{AT}^{[\alpha,\beta]}(u_i) \neq Y_{AT}^{[\alpha,\beta]}(u_j)\}$$

定义

$$M_{ij}^{[\alpha,\beta]} = \begin{cases} \{a \in AT | f(a, u_i) \neq f(a, u_j)\}, & (u_i, u_j) \in D^* \\ AT, & (u_i, u_j) \notin D^* \end{cases}$$

为 (u_i, u_j) 的广义分布可辨识属性集, $M^{[\alpha,\beta]} = \{M_{ij}^{[\alpha,\beta]} | i, j \in \{1, 2, \dots, n\}\}$ 为决策表的一个 $n \times n$ 的广义分布差别矩阵, 其为对称矩阵, α 和 β 满足 $(\alpha = 0 \wedge \beta \in [0, 1])$ 或 $(\alpha \in [0, 1] \wedge \beta = 1)$ 。

定理2 设 $DT = (U, AT \cup D, V, f)$, $\forall A \subseteq AT$, 则 A 是广义分布协调集 $\Leftrightarrow \forall u_i, u_j \in U$, 若 $M_{ij}^{[\alpha,\beta]} \neq \emptyset$, 有 $A \cap M_{ij}^{[\alpha,\beta]} \neq \emptyset$ 。其中, α 和 β 满足 $(\alpha = 0 \wedge \beta \in [0, 1])$ 或 $(\alpha \in [0, 1] \wedge \beta = 1)$ 。

证明 若 $(u_i, u_j) \notin D^*$, 显然有 $A \cap M_{ij}^{[\alpha,\beta]} \neq \emptyset$ 。反之, 则有

“ \Rightarrow ”: 由于 A 是广义分布协调集, 故对于 $\forall (u_i, u_j) \in D^*$, $Y_{AT}^{[\alpha,\beta]}(u_i) \neq Y_{AT}^{[\alpha,\beta]}(u_j)$, 由定理1可得 $[u_i]_A \cap [u_j]_A = \emptyset$ 。因此, $\exists a \in A$, 不等式 $f(a, u_i) \neq f(a, u_j)$ 成立, 故 $a \in M_{ij}^{[\alpha,\beta]}$, 即 $A \cap M_{ij}^{[\alpha,\beta]} \neq \emptyset$ 。

“ \Leftarrow ”: 不妨假设 $\exists (u_i, u_j) \in D^*$, 使得 $A \cap M_{ij}^{[\alpha,\beta]} = \emptyset$, 显然 $Y_{AT}^{[\alpha,\beta]}(u_i) \neq Y_{AT}^{[\alpha,\beta]}(u_j)$ 。 $\forall a \in A$, 必然有 $a \notin M_{ij}^{[\alpha,\beta]}$, 也即 $f(a, u_i) = f(a, u_j)$ 。故有 $[u_i]_A = [u_j]_A$, $[u_i]_A \cap [u_j]_A \neq \emptyset$, 由定理1可得 A 不是广义分布协调集。定理得证。

定义5 设 $DT = (U, AT \cup D, V, f)$, $M^{[\alpha,\beta]}$ 为广义分布保持约简的差别矩阵, 其对应的差别函数为

$$DF(M^{[\alpha,\beta]}) = \bigwedge \{ \bigvee M_{ij}^{[\alpha,\beta]} | 1 \leq i < j \leq n \}$$

式中: $\bigvee M_{ij}^{[\alpha,\beta]} = \bigvee a (a \in M_{ij}^{[\alpha,\beta]})$ 表示 $M_{ij}^{[\alpha,\beta]}$ 中所有属性的析取, 且 α 和 β 满足 $(\alpha = 0 \wedge \beta \in [0, 1])$ 或 $(\alpha \in [0, 1] \wedge \beta = 1)$ 。

通过化 $DF(M^{[\alpha,\beta]})$ 的主合取范式转化为主析取范式即可得到所有广义分布保持属性约简。

定理3 设 $DT = (U, AT \cup D, V, f)$, $M^{[\alpha,\beta]}$ 为 DT 的广义分布保持约简的差别矩阵, 且 α 和 β 满足 $(\alpha = 0 \wedge \beta \in [0, 1])$ 或 $(\alpha \in [0, 1] \wedge \beta = 1)$ 。 $DF(M^{[\alpha,\beta]})$ 是由 $M^{[\alpha,\beta]}$ 导出的差别函数, $DF(M^{[\alpha,\beta]})$ 的极小析取范式为

$$DF(M^{[\alpha,\beta]}) = \bigvee_{k=1}^t \left(\bigwedge_{s=1}^{q_k} a_{i_s} \right)$$

记 $A_k = \{a_{i_s} | s = 1, 2, \dots, q_k\}$, 则 $\{A_k | k = 1, 2, \dots, t\}$ 是决策表的所有广义分布保持约简, 其中 t 表示 $DF(M^{[\alpha,\beta]})$ 的极小析取范式中的合取项数目。

证明 $\forall k \leq t, \forall M_{ij}^{[\alpha,\beta]} \in M^{[\alpha,\beta]}$, 由极小析取范式的定义知 $A_k \cap M_{ij}^{[\alpha,\beta]} \neq \emptyset$, 再由定理2可得 A_k 是广义分布协调集。同时, $DF(M^{[\alpha,\beta]}) = \bigvee_{k=1}^t (A_k)$, 若在 A_k 中去掉任意一个属性形成 A'_k , 则必然 $\exists M_{ij}^{[\alpha,\beta]} \in M^{[\alpha,\beta]}$, 使得 $A'_k \cap M_{ij}^{[\alpha,\beta]} = \emptyset$, 故 A'_k 不是广义分布协调集, 从而 A_k 是广义分布保持约简。

由于 $DF(M^{[\alpha,\beta]})$ 中包含了所有的 $M_{ij}^{[\alpha,\beta]}$, 因此不存在其他的广义分布保持约简, 定理得证。

3 广义分布保持属性约简算法

本节给出广义分布保持约简算法 (generalized distribution preservation reduction algorithm, GDPRA), 算法描述如下。

输入 决策表 $DT = (U, AT \cup D, V, f)$, α 和 β 。

输出 DT 的所有广义分布保持属性约简。

1) 计算每个对象在条件属性集下关于决策属性的置信度分布 μ_{AT} 。

2) 根据每个对象的置信度分布 μ_{AT} 获取每个对象的 $[\alpha, \beta]$ 决策-置信度序偶集。

3) 根据对象之间的决策-置信度序偶集构造相应的广义分布差别矩阵。

4) 根据广义分布差别矩阵构造广义分布差别函数,并通过吸收率进行简化。

5) 在 $DF(M^{[\alpha,\beta]})$ 基础上通过结合律获取所有的广义分布保持约简。

其中, α 和 β 满足 $(\alpha=0 \wedge \beta \in [0,1])$ 或 $(\alpha \in [0,1] \wedge \beta=1)$ 。

由于上述算法是通过差别矩阵获取决策表的所有广义分布保持约简,故算法在最坏情况下的时间复杂度为 $O(|AT|^{U|2})$,最坏情况下的空间复杂度为 $O(|AT||U|^2)$,其中 $|U|$ 为样本空间中的对象数目, $|AT|$ 为条件属性数,下面通过例1简要说明 GDPRA 的执行过程。

例1 如表1所示,论域为 $U = \{u_1, u_2, u_3, u_4\}$, $AT = \{a_1, a_2, a_3, a_4\}$ 为条件属性集, $D = \{d\}$ 为决策属性,分别求 $[\alpha, \beta] = [0, 0.3]$ 以及 $[\alpha, \beta] = [0.8, 1]$ 时的所有广义分布保持约简。

表1 决策表

Table 1 Decision table

U	a_1	a_2	a_3	a_4	d
u_1	1	1	0	1	0
u_2	1	0	0	1	1
u_3	1	0	1	1	2
u_4	1	0	1	1	1

1) 获取每个对象的置信度分布

$$U/AT = \{E_1, E_2, E_3\}$$

$$U/D = \{D_1, D_2, D_3\}$$

$$E_1 = \{u_1\}$$

$$E_2 = \{u_2\}$$

$$E_3 = \{u_3, u_4\}$$

$$D_1 = \{u_1\}$$

$$D_2 = \{u_2, u_4\}$$

$$D_3 = \{u_3\}$$

$$\mu_{AT}(u_1) = (1, 0, 0)$$

$$\mu_{AT}(u_2) = (0, 1, 0)$$

$$\mu_{AT}(u_3) = (0, 0.5, 0.5)$$

$$\mu_{AT}(u_4) = (0, 0.5, 0.5)$$

2) 获取每个对象的 $[\alpha, \beta]$ 决策-置信度序偶集

当 $\alpha=0, \beta=0.3$ 时

$$Y_{AT}^{[0,0.3]}(u_1) = \{ \langle 1, 0 \rangle, \langle 2, 0 \rangle \}$$

$$Y_{AT}^{[0,0.3]}(u_2) = \{ \langle 0, 0 \rangle, \langle 2, 0 \rangle \}$$

$$Y_{AT}^{[0,0.3]}(u_3) = \{ \langle 0, 0 \rangle \}$$

$$Y_{AT}^{[0,0.3]}(u_4) = \{ \langle 0, 0 \rangle \}$$

当 $\alpha=0.8, \beta=1$ 时

$$Y_{AT}^{[0.8,1]}(u_1) = \{ \langle 0, 1 \rangle \}$$

$$Y_{AT}^{[0.8,1]}(u_2) = \{ \langle 1, 1 \rangle \}$$

$$Y_{AT}^{[0.8,1]}(u_3) = \emptyset$$

$$Y_{AT}^{[0.8,1]}(u_4) = \emptyset$$

3) 构造广义分布差别矩阵

$[\alpha, \beta] = [0, 0.3]$ 时对应的广义分布差别矩阵如表2所示, $[\alpha, \beta] = [0.8, 1]$ 时对应的广义分布差别矩阵如表3所示。

表2 广义分布差别矩阵1

Table 2 Generalized distribution discernibility matrix 1

$M^{[0,0.3]}$	u_1	u_2	u_3	u_4
u_1	AT	$\{a_2\}$	$\{a_2, a_3\}$	$\{a_2, a_3\}$
u_2	$\{a_2\}$	AT	$\{a_3\}$	$\{a_3\}$
u_3	$\{a_2, a_3\}$	$\{a_3\}$	AT	AT
u_4	$\{a_2, a_3\}$	$\{a_3\}$	AT	AT

表3 广义分布差别矩阵2

Table 3 Generalized distribution discernibility matrix 2

$M^{[0.8,1]}$	u_1	u_2	u_3	u_4
u_1	AT	$\{a_2\}$	$\{a_2, a_3\}$	$\{a_2, a_3\}$
u_2	$\{a_2\}$	AT	$\{a_3\}$	$\{a_3\}$
u_3	$\{a_2, a_3\}$	$\{a_3\}$	AT	AT
u_4	$\{a_2, a_3\}$	$\{a_3\}$	AT	AT

4) 获取差别函数并进行简化

$$DF(M^{[0,0.3]}) = (a_2) \wedge (a_3)$$

$$DF(M^{[0.8,1]}) = (a_2) \wedge (a_3)$$

5) 通过结合律获取所有的广义分布保持约简

由计算得, $[\alpha, \beta] = [0, 0.3]$ 时和 $[\alpha, \beta] = [0.8, 1]$ 时的所有广义分布保持约简均为 $\{a_2, a_3\}$ 。

4 一些特殊情形下的讨论

值得注意的是,给定决策表 $DT = (U, AT \cup D, V, f)$,当 α 和 β 取某些特殊值时,广义分布保持约简可以退化为目前已存在的一些约简,本节将根据 α 和 β 不同的特殊取值情况展开讨论,并给出相应的结论。其中,将 M_{GEN} 、 M_{POS} 以及 M_{DIS} 分别记为广义决策可辨识矩阵、正域可辨识矩阵以及分布可辨识矩阵,同时,将 M_{ij}^{GEN} 、 M_{ij}^{POS} 以及 M_{ij}^{DIS} 分别记为对象 u_i 和 u_j 对应广义决策可辨识矩阵、正域可辨识矩阵以及分布可辨识矩阵的可辨识属性集,其中论域为 $U = \{u_1, u_2, \dots, u_n\}, i, j \in \{1, 2, \dots, n\}$ 。

1) $\alpha = \beta = 0$ 时

当 α 和 β 取值均为 0 时,广义分布保持约简实质是保证对于置信度为 0 的规则在约简前后的置信

度均为 0,而对于置信度不为 0 的规则在约简前后的置信度均不为 0,由此可得如下结论。

定理 4 设决策表 $DT=(U, AT \cup D, V, f)$, 对于 $\forall R \subseteq AT$ 且 $R \neq \emptyset, \alpha = \beta = 0$, 若 R 是决策表 DT 的一个广义分布保持约简, 则 R 必定同时是决策表 DT 的一个广义决策保持约简。

证明 不妨设 $\forall u_i, u_j \in U$, 其中 $[u_i]_{AT} \cap [u_j]_{AT} = \emptyset$, 同时设 $U/D = \{D_1, D_2, \dots, D_{|U/D|}\}$ 。若有 $Y_{AT}^{[0,0]}(u_i) \neq Y_{AT}^{[0,0]}(u_j)$ 成立, 则 $\exists D_k \in U/D$, 其中 $k \in \{1, 2, \dots, |U/D|\}$, 使得 $[u_i]_{AT} \cap D_k = \emptyset \wedge [u_j]_{AT} \cap D_k \neq \emptyset$ 或 $[u_i]_{AT} \cap D_k \neq \emptyset \wedge [u_j]_{AT} \cap D_k = \emptyset$ 成立, 也即 $\delta_{AT}(u_i) \neq \delta_{AT}(u_j)$ 成立, 故由定义 4 可知有 $M_{ij}^{[0,0]} = M_{ij}^{GEN} = \{a \in AT | f(u_i, a) \neq f(u_j, a)\}$ 成立; 反之, 若 $Y_{AT}^{[0,0]}(u_i) = Y_{AT}^{[0,0]}(u_j)$, 则 $\forall D_k \in U/D$, 有 $[u_i]_{AT} \cap D_k \neq \emptyset \wedge [u_j]_{AT} \cap D_k \neq \emptyset$ 或 $[u_i]_{AT} \cap D_k = \emptyset \wedge [u_j]_{AT} \cap D_k = \emptyset$ 成立, 故 $\delta_{AT}(u_i) = \delta_{AT}(u_j)$, 由定义 4 可知 $M_{ij}^{[0,0]} = M_{ij}^{GEN} = AT$ 成立。由于 $\forall u_i, u_j \in U$, 故在 $\alpha = \beta = 0$ 条件下, 有 $M_{GEN} = M^{[0,0]}$ 成立, 故 R 是决策表 DT 的一个广义分布保持约简, 则 R 必定同时是决策表 DT 的一个广义决策保持约简, 证毕。

2) $\alpha = \beta = 1$ 时

显然, 当 $\alpha = \beta = 1$ 时, 广义分布保持约简实质是保证了置信度为 1 的规则在约简前后的置信度保持不变, 由此可得如下结论。

定理 5 决策表 $DT=(U, AT \cup D, V, f)$, 对于 $\forall R \subseteq AT$ 且 $R \neq \emptyset$, 令 $\alpha = \beta = 1$, 若 R 是决策表 DT 的一个广义分布保持约简, 则 R 必定同时是决策表 DT 的一个正域保持约简。

证明 不妨设 $\forall u_i, u_j \in U$, 其中, $[u_i]_{AT} \cap [u_j]_{AT} = \emptyset, U/D = \{D_1, D_2, \dots, D_{|U/D|}\}$, 分情况进行如下讨论。

当 $Y_{AT}^{[1,1]}(u_i) \neq Y_{AT}^{[1,1]}(u_j)$ 成立时, 有 $M_{ij}^{[1,1]} = \{a \in AT | f(u_i, a) \neq f(u_j, a)\}$ 成立。假设 $Y_{AT}^{[1,1]}(u_i) \neq \emptyset$ 且 $Y_{AT}^{[1,1]}(u_j) \neq \emptyset$, 则 $\exists D_k \in U/D$, 使得 $P(D_k | [u_i]_{AT}) = P(D_k | [u_j]_{AT}) = 1$ 成立, 故有 $u_i, u_j \in POS_{AT}(D)$ 且 $f(u_i, d) \neq f(u_j, d)$ 成立, 即 $M_{ij}^{POS} = \{a \in AT | f(u_i, a) \neq f(u_j, a)\} = M_{ij}^{[1,1]}$ 成立。若假设不成立, 则有 $(u_i \in POS_{AT}(D) \wedge u_j \in BND_{AT}(D))$ 或 $(u_j \in POS_{AT}(D) \wedge u_i \in BND_{AT}(D))$ 成立; 若 $f(u_i, d) \neq f(u_j, d)$ 成立, 则有 $M_{ij}^{POS} = \{a \in AT | f(u_i, a) \neq f(u_j, a)\} = M_{ij}^{[1,1]}$, 若 $f(u_i, d) = f(u_j, d)$, 则有 $M_{ij}^{POS} = AT$, 同时必然 $\exists u_{i'} \in [u_i]_{AT}, u_{j'} \in [u_j]_{AT}$, 使得 $f(u_{i'}, d) \neq f(u_{j'}, d)$, 故 $M_{i'j'}^{POS} = \{a \in AT | f(u_{i'}, a) \neq f(u_{j'}, a)\} = M_{i'j'}^{[1,1]}$, 故 $M_{ij}^{POS} = M_{ij}^{[1,1]}$ 。

当 $Y_{AT}^{[1,1]}(u_i) = Y_{AT}^{[1,1]}(u_j)$ 时, 有 $Y_{AT}^{[1,1]}(u_i) = Y_{AT}^{[1,1]}(u_j) = \emptyset$ 或 $Y_{AT}^{[1,1]}(u_i) = Y_{AT}^{[1,1]}(u_j) \neq \emptyset$ 成立。对于前者, $\forall D_k \in U/D$, 使得 $[u_i]_{AT} \cap D_k \neq [u_j]_{AT} \cap D_k$ 成立并且 $[u_j]_{AT} \cap D_k \neq [u_i]_{AT} \cap D_k$ 成立, 故有 $u_i, u_j \in BND_{AT}(D)$; 对于后者, $\exists D_k \in U/D$, 使 $([u_i]_{AT} \cap D_k = [u_j]_{AT} \cap D_k) \wedge ([u_j]_{AT} \cap D_k \neq [u_i]_{AT} \cap D_k)$ 成立, 即 $f(u_i, d) = f(u_j, d)$, 故由定义 4 可知 $M_{ij}^{POS} = M_{ij}^{[1,1]} = AT$ 。

综上, 由于 $\forall u_i, u_j \in U$, 故在 $\alpha = \beta = 1$ 的条件下, $M_{POS} = M^{[1,1]}$ 成立, 故 R 是决策表 DT 的广义分布保持约简, 则 R 必定同时是决策表 DT 的一个正域保持约简, 证毕。

3) $\alpha = 0, \beta = 1$ 时

当 $\alpha = 0, \beta = 1$ 时, 广义分布保持约简实质是保证了置信度在 $[0, 1]$ 内的所有规则在约简前后的置信度不变, 同时易得, 此时对象的 $[\alpha, \beta]$ 决策-置信度序偶集等价于在决策等价类划分上的置信度分布, 由此可得如下结论。

定理 6 决策表 $DT=(U, AT \cup D, V, f)$, 对于 $\forall R \subseteq AT$ 且 $R \neq \emptyset, \alpha = 0$ 且 $\beta = 1$, 若 R 是决策表 DT 的一个广义分布保持约简, 则 R 必定同时是决策表 DT 的一个分布保持约简。

证明 不妨设 $\forall u_i, u_j \in U$, 其中 $[u_i]_{AT} \cap [u_j]_{AT} = \emptyset, U/D = \{D_1, D_2, \dots, D_{|U/D|}\}$; 若 $Y_{AT}^{[0,1]}(u_i) \neq Y_{AT}^{[0,1]}(u_j)$, 则 $\exists D_k \in U/D$, 使得 $P(D_k | [u_i]_{AT}) \neq P(D_k | [u_j]_{AT})$, 故有 $\mu_{AT}(u_i) \neq \mu_{AT}(u_j)$ 成立, 所以由定义 4 可知 $M_{ij}^{DIS} = M_{ij}^{[0,1]} = \{a \in AT | f(u_i, a) \neq f(u_j, a)\}$; 反之, 若 $Y_{AT}^{[0,1]}(u_i) = Y_{AT}^{[0,1]}(u_j)$ 成立, 则 $\forall D_k \in U/D$, 有 $P(D_k | [u_i]_{AT}) = P(D_k | [u_j]_{AT})$ 成立, 则由定义 4 可知 $M_{ij}^{DIS} = M_{ij}^{[0,1]} = AT$; 由于 $\forall u_i, u_j \in U$, 故在 $\alpha = 0, \beta = 1$ 的条件下, 有 $M_{DIS} = M^{[0,1]}$ 成立, 故 R 是决策表 DT 的一个广义分布保持约简, 必定同时是决策表 DT 的一个分布保持约简, 证毕。

综上, 图 1 给出了广义分布保持约简与上述几种约简之间的关系。

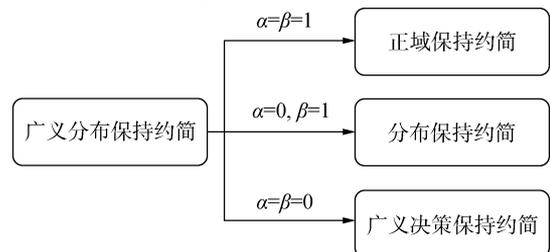


图 1 几种不同约简之间的关系

Fig.1 Relationships among different reductions

例 2 表 1 所示决策表,论域 $U = \{u_1, u_2, u_3, u_4\}$, $AT = \{a_1, a_2, a_3, a_4\}$ 为条件属性集, $D = \{d\}$ 为决策属性。由

$$U/AT = \{E_1, E_2, E_3\}$$

$$E_1 = \{u_1\}$$

$$E_2 = \{u_2\}$$

$$E_3 = \{u_3, u_4\}$$

$$U/D = \{D_1, D_2, D_3\}$$

$$D_1 = \{u_1\}$$

$$D_2 = \{u_2, u_4\}$$

$$D_3 = \{u_3\}$$

$$POS_{AT}(D) = \{u_1, u_2\}$$

$$\delta_{AT}(u_1) = \{0\}$$

$$\delta_{AT}(u_2) = \{1\}$$

$$\delta_{AT}(u_3) = \{1, 2\}$$

$$\delta_{AT}(u_4) = \{1, 2\}$$

$$\mu_{AT}(u_1) = (1, 0, 0)$$

$$\mu_{AT}(u_2) = (0, 1, 0)$$

$$\mu_{AT}(u_3) = (0, 0.5, 0.5)$$

$$\mu_{AT}(u_4) = (0, 0.5, 0.5)$$

求得正域保持约简为 $\{a_2, a_3\}$, 广义决策保持约简为 $\{a_2, a_3\}$, 分布保持约简为 $\{a_2, a_3\}$ 。

因此, 当 $\alpha = \beta = 1$ 时, 可得: $Y_{AT}^{[1,1]}(u_1) = \{\langle 0, 1 \rangle\}$, $Y_{AT}^{[1,1]}(u_2) = \{\langle 1, 1 \rangle\}$, $Y_{AT}^{[1,1]}(u_3) = \emptyset$, $Y_{AT}^{[1,1]}(u_4) = \emptyset$ 。

据此构造广义分布差别矩阵, 如表 4 所示。

表 4 广义分布差别矩阵

$M^{[1,1]}$	u_1	u_2	u_3	u_4
u_1	AT	$\{a_2\}$	$\{a_2, a_3\}$	$\{a_2, a_3\}$
u_2	$\{a_2\}$	AT	$\{a_3\}$	$\{a_3\}$
u_3	$\{a_2, a_3\}$	$\{a_3\}$	AT	AT
u_4	$\{a_2, a_3\}$	$\{a_3\}$	AT	AT

由广义分布差别矩阵可得所有的广义分布保持约简为 $\{a_2, a_3\}$, 与正域约简一致。同理, $\alpha = \beta = 0$ 时的广义分布保持约简为 $\{a_2, a_3\}$, 与广义决策约简一致; $\alpha = 0, \beta = 1$ 时的广义分布保持约简为 $\{a_2, a_3\}$, 与分布约简一致。

由定理 4~6 可得如下结论。

推论 1 设 $DT = (U, AT \cup D, V, f)$, 置信度区间为 $[\alpha, \beta]$, $\forall A \subseteq AT$, 且 A 是置信度区间 $[\alpha, \beta]$ 下的

一个广义分布保持约简, 进一步, 若给定置信度区间 $[\alpha', \beta']$, 且满足 $[\alpha', \beta'] \subseteq [\alpha, \beta]$, 则 $\exists A' \subseteq A$, 使得 A' 是置信度区间 $[\alpha', \beta']$ 下的一个广义分布保持约简, 且满足 $A' \subseteq A$ 。其中, α 和 β 满足 $(\alpha = 0 \wedge \beta \in [0, 1])$ 或 $(\alpha \in [0, 1] \wedge \beta = 1)$ 。

证明 由已知条件得, $[\alpha', \beta'] \subseteq [\alpha, \beta]$, $\forall u \in U$, 有 $Y_A^{[\alpha, \beta]}(u) = Y_{AT}^{[\alpha, \beta]}(u)$ 。则 $\forall u \in U$, 必然有 $Y_{A'}^{[\alpha', \beta']}(u) = Y_{AT}^{[\alpha', \beta']}(u)$, 故 A 是决策表在置信度区间 $[\alpha', \beta']$ 下的一个广义分布协调集。假设 A 是决策表在置信度区间 $[\alpha', \beta']$ 下的一个广义分布约简, 则有 $A' \subseteq A$; 反之, 必然 $\exists A' \subset A$, 使得 $Y_{A'}^{[\alpha', \beta']}(u) = Y_A^{[\alpha', \beta']}(u)$, 故 $\exists A' \subseteq A$, 使得 $A' \subseteq A$ 成立, 证毕。

5 实验分析

本节采用 4 个 UCI 数据集进行实验, 数据集信息如表 5 所示, 其中, $|U|$ 表示数据集的样本数, $|AT|$ 表示数据集的特征数, $|D|$ 表示分类数。对于数据集的预处理, 处理策略如下: 缺失特征值通过用该缺失特征值所对应特征下的多数特征值进行填充, 连续型特征进行等频离散化, 名词性特征值用整数进行替换, 所有数据集的预处理均在 Weka 3.6 下进行。实验环境如下: Windows 7 旗舰版 32 位操作系统, Intel Pentium G640C 处理器, 主频 2.8 GHz, 内存 6.0 GB, 所有算法均采用 MATLAB R2010b 编写实现。

表 5 UCI 数据集信息

数据集	$ U $	$ AT $	$ D $
Haberman's Survival	306	3	2
Blood Transfusion Service Center	748	4	2
Stone Flakes	79	7	3
Airfoil Self-Noise	1503	5	16

注: BTSC 为数据集 Blood Transfusion Service Center 的缩写

实验分为两部分。第 1 部分验证置信度区间分别为 $[1.0, 1.0]$ 、 $[0.0, 0.0]$ 以及 $[0.0, 1.0]$ 时, 广义分布保持约简可分别退化为正域保持约简、广义决策保持约简以及分布约简, 同时, 也可验证广义分布保持约简算法的正确性; 第 2 部分验证在较小的置信度区间下求得的广义分布保持约简是在较大的置信度区间下求得的广义分布保持约简的子集。

5.1 广义分布保持属性约简的退化情形

本节中,分别令 $[\alpha, \beta]$ 取值为 $[1.0, 1.0]$ 、 $[0.0, 0.0]$ 和 $[0.0, 1.0]$,并求4个UCI数据集的广义分布保持约简,然后,分别求它们在正域保持约简算法(positive region preservation reduction algorithm, PRPRA),广义决策保持约简算法(algorithm of generalized decision preservation reduction, AGDPR)以及分布保持约简算法(distribution preservation reduction algorithm, DPRA)下的约简,通过前后对比,验证广义分布保持约简在3个特殊置信度区间下的退化情况,实验结果如表6~8所示。

表6 GDPRA和PRPRA的约简结果($[\alpha, \beta]=[1, 1]$)

Table 6 Reduction results for GDPRA and PRPRA ($[\alpha, \beta]=[1, 1]$)

数据集	GDPRA	PRPRA
Haberman's Survival	{2,3}	{2,3}
BTSC	{1,4}	{1,4}
Stone Flakes	{2,6,7}	{2,6,7}
Airfoil Self-Noise	{1,2,3,4}	{1,2,3,4}

表7 GDPRA和GDECPRAs的约简结果($[\alpha, \beta]=[0, 0]$)

Table 7 Reduction results for GDPRA and GDECPRAs ($[\alpha, \beta]=[0, 0]$)

数据集	GDPRA	GDECPRAs
Haberman's Survival	{2,3}	{2,3}
BTSC	{1,4}	{1,4}
Stone Flakes	{2,3,4,5,6,7}	{2,3,4,5,6,7}
Airfoil Self-Noise	{1,2,3,4,5}	{1,2,3,4,5}

表8 GDPRA和DPRA的约简结果($[\alpha, \beta]=[0, 1]$)

Table 8 Reduction results for GDPRA and DPRA ($[\alpha, \beta]=[0, 1]$)

数据集	GDPRA	DPRA
Haberman's Survival	{1,2,3}	{1,2,3}
BTSC	{1,2,4}, {1,3,4}	{1,2,4}, {1,3,4}
Stone Flakes	{2,3,4,5,6,7}	{2,3,4,5,6,7}
Airfoil Self-Noise	{1,2,3,4,5}	{1,2,3,4,5}

当 $[\alpha, \beta]$ 分别为 $[1.0, 1.0]$ 、 $[0.0, 0.0]$ 以及 $[0.0, 1.0]$ 时,GDPRA的约简结果分别同PRPRA、AGDPR以及DPRA的约简结果一致,验证了相关结论的正确性。

5.2 不同置信度区间下约简的包含关系

本部分实验设置如下:首先,固定 α 的值为0.0,令 β 取值范围为0.0~1.0,取值间隔为0.2,记录随 β 取值的变化在不同置信度区间下求得的广义分布保持约简。同样的,固定 β 的值为1.0,令 α 取值范围为0.0~1.0,取值间隔为0.2,记录随 α 取值的变化在不同置信度区间下求得的广义分布保持约简,实验结果如表9~12所示。

表9 数据集1: Haberman's survival

Table 9 Data set 1: Haberman's survival

$[\alpha, \beta]$	约简	$[\alpha, \beta]$	约简
$[0.0, 0.0]$	{2,3}	$[0.0, 1.0]$	{1,2,3}
$[0.0, 0.2]$	{1,2,3}	$[0.2, 1.0]$	{1,2,3}
$[0.0, 0.4]$	{1,2,3}	$[0.4, 1.0]$	{1,2,3}
$[0.0, 0.6]$	{1,2,3}	$[0.6, 1.0]$	{1,2,3}
$[0.0, 0.8]$	{1,2,3}	$[0.8, 1.0]$	{1,2,3}
$[0.0, 1.0]$	{1,2,3}	$[1.0, 1.0]$	{2,3}

表10 数据集2: blood transfusion service center

Table 10 Data set 2: blood transfusion service center

$[\alpha, \beta]$	约简	$[\alpha, \beta]$	约简
$[0.0, 0.0]$	{1,4}	$[0.0, 1.0]$	{1,2,4}, {1,3,4}
$[0.0, 0.2]$	{1,2,4}, {1,3,4}	$[0.2, 1.0]$	{1,2,4}, {1,3,4}
$[0.0, 0.4]$	{1,2,4}, {1,3,4}	$[0.4, 1.0]$	{1,2,4}, {1,3,4}
$[0.0, 0.6]$	{1,2,4}, {1,3,4}	$[0.6, 1.0]$	{1,2,4}, {1,3,4}
$[0.0, 0.8]$	{1,2,4}, {1,3,4}	$[0.8, 1.0]$	{1,2,4}, {1,3,4}
$[0.0, 1.0]$	{1,2,4}, {1,3,4}	$[1.0, 1.0]$	{1,4}

表11 数据集3: stone flakes

Table 11 Data set 3: stone flakes

$[\alpha, \beta]$	约简	$[\alpha, \beta]$	约简
$[0.0, 0.0]$	{2,3,4,5,6,7}	$[0.0, 1.0]$	{2,3,4,5,6,7}
$[0.0, 0.2]$	{2,3,4,5,6,7}	$[0.2, 1.0]$	{2,3,4,5,6,7}
$[0.0, 0.4]$	{2,3,4,5,6,7}	$[0.4, 1.0]$	{2,3,4,5,6,7}
$[0.0, 0.6]$	{2,3,4,5,6,7}	$[0.6, 1.0]$	{2,3,4,5,6,7}
$[0.0, 0.8]$	{2,3,4,5,6,7}	$[0.8, 1.0]$	{2,4,5,6,7}
$[0.0, 1.0]$	{2,3,4,5,6,7}	$[1.0, 1.0]$	{2,6,7}

表 12 数据集 4: airfoil self-noise
Table 12 Data set 4: airfoil self-noise

$[\alpha, \beta]$	约简	$[\alpha, \beta]$	约简
[0.0,0.0]	{1,2,3,4,5}	[0.0,1.0]	{1,2,3,4,5}
[0.0,0.2]	{1,2,3,4,5}	[0.2,1.0]	{1,2,3,4,5}
[0.0,0.4]	{1,2,3,4,5}	[0.4,1.0]	{1,2,3,4,5}
[0.0,0.6]	{1,2,3,4,5}	[0.6,1.0]	{1,2,3,4}
[0.0,0.8]	{1,2,3,4,5}	[0.8,1.0]	{1,2,3,4}
[0.0,1.0]	{1,2,3,4,5}	[1.0,1.0]	{1,2,3,4}

6 结束语

实际中,具有较高或较低置信度的规则往往更易受到人们的关注,若通过分布约简进行规则提取,提取的规则可能过于冗长,不便于实际决策。因此,本文对分布约简的约简标准进行弱化,提出了广义分布保持约简的概念。理论与实验分析表明,当置信度区间取某些特殊值时,广义分布保持属性约简可退化为现有的一些属性约简,表明了广义分布保持属性约简具有一定的泛化性能,同时为深入研究不同属性约简之间的相互关系开阔了研究思路。实验数据表明,广义分布保持属性约简较分布约简可以获取更加简短的规则,且根据实际需要可以调整置信度区间以获取所需规则,使得广义分布保持属性约简可以适应不同的实际需求。但考虑到本文提出的算法主要是通过差别矩阵获取所有的广义分布保持属性约简,其时间和空间复杂度较高,不便于在实际应用中推广,具有一定的局限性,故开发更为高效的广义分布保持属性约简算法是未来主要的研究工作之一。

参考文献:

[1] PAWLAK Z. Rough sets[J]. International journal of computer and information sciences, 1982, 11(5): 341-356.
 [2] PAWLAK Z. Rough sets: theoretical aspects of reasoning about data[M]. Boston: Kluwer Academic Publishers, 1992.
 [3] 张文修. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001.
 [4] 王国胤, 姚一豫, 于洪. 粗糙集理论与应用研究综述[J]. 计算机学报, 2009, 32(7): 1229-1246.
 WANG Guoyin, YAO Yiyu, YU Hong. A survey on rough set theory and applications [J]. Chinese journal of

computers, 2009, 32(7): 1229-1246.
 [5] SKOWRON A, RAUSZER C. The discernibility matrices and functions in information systems[J]. Theory and decision library, 1992, 11: 331-362.
 [6] KRYSZKIEWICZ M. Rough set approach to incomplete information systems[J]. Information sciences, 1998, 112(1/2/3/4): 39-49.
 [7] 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.
 [8] 徐伟华, 张文修. 基于优势关系下不协调目标信息系统的分布约简[J]. 模糊系统与数学, 2007, 21(4): 124-131.
 XU Weihua, ZHANG Wenxiu. Distribution reduction in inconsistent information systems based on dominance relations[J]. Fuzzy systems and mathematics, 2007, 21(4): 124-131.
 [9] MIAO Duoqian, ZHAO Yan, YAO Yiyu, et al. Relative reducts in consistent and inconsistent decision tables of the Pawlak rough set model[J]. Information sciences, 2009, 179(24): 4140-4150.
 [10] 张楠, 苗夺谦, 岳晓冬. 区间值信息系统的知识约简[J]. 计算机研究与发展, 2010, 47(8): 1362-1371.
 ZHANG Nan, MIAO Duoqian, YUE Xiaodong. Approaches to knowledge reduction in interval-valued information systems[J]. Journal of computer research and development, 2010, 47(8): 1362-1371.
 [11] 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 Duoqian, HU Guirong. A heuristic algorithm for reduction of knowledge[J]. Journal of computer research and development, 1999, 36(6): 681-684.
 [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 Guoyin, YU Hong, YANG Dachun. Decision table reduction based on conditional information entropy [J]. Chinese journal of computers, 2002, 25(7): 759-766.
 [13] QIAN Yuhua, LIANG Jiye, PEDRYCZ W, et al. Positive approximation: an accelerator for attribute reduction in rough set theory [J]. Artificial intelligence, 2010, 174(9): 597-618.
 [14] QIAN Yuhua, LIANG Jiye, PEDRYCZ W, et al. An

efficient accelerator for attribute reduction from incomplete data in rough set framework [J]. Pattern recognition, 2011, 44(8): 1658-1670.

- [15] 钱宇华, 梁吉业, 王锋. 面向非完备决策表的正向近似特征选择加速算法[J]. 计算机学报, 2011, 34(3): 435-442.

QIAN Yuhua, LIANG Jiye, WANG Feng. A positive approximation based accelerated algorithm to feature selection from incomplete decision tables [J]. Chinese journal of computers, 2011, 34(3): 435-442.

- [16] CHEN Hongmei, LI Tianrui, RUAN Da, et al. A rough-set based incremental approach for updating approximations under dynamic maintenance environments[J]. IEEE transactions on knowledge and data engineering, 2013, 25(2): 274-284.

- [17] CHEN Hongmei, LI Tianrui, LUO Chuan, et al. A rough set-based method for updating decision rules on attribute values' coarsening and refining[J]. IEEE transactions on knowledge and data engineering, 2014, 26(12): 2866-2899.

- [18] JIA Xiuyi, SHANG Lin, ZHOU Bing, et al. Generalized attribute reduct in rough set theory[J]. Knowledge-based systems, 2015, 91: 204-218.

- [19] ZHOU Jie, MIAO Duoqian, PEDRYCZ W, et al. Analysis of alternative objective functions for attribute reduction in complete decision tables [J]. Soft computing, 2011, 15(8): 1601-1616.

- [20] ZHANG Xiao, MEI Changlin, CHEN Degang, et al. Multi-confidence rule acquisition and confidence-preserved attribute reduction in interval-valued decision systems[J]. International journal of approximate reasoning, 2014, 55(8): 1787-1804.

作者简介:



高学义,男,1992年生,硕士研究生,主要研究方向为粗糙集、数据挖掘与机器学习。



张楠,男,1979年生,博士,主要研究方向为粗糙集、认知信息学与人工智能。



童向荣,男,1975年生,教授,博士,主要研究方向为多 Agent 系统、分布式人工智能与数据挖掘技术。