

DOI: 10.13718/j.cnki.xdzk.2024.05.006

张晓燕, 李璐. 区间值决策表中基于相对优势邻域粒度的属性约简 [J]. 西南大学学报(自然科学版), 2024, 46(5): 67-76.

区间值决策表中基于相对优势邻域粒度的属性约简

张晓燕, 李璐

西南大学 人工智能学院, 重庆 400715

摘要: 现实生活中大量数据以区间值形式存在, 此时区间值决策表并不是基于等价关系, 传统的决策方法并不能解决这一问题. 为此, 本文在区间值决策表中引入相邻关系、相邻类的定义, 进而由相邻类建立了区间决策表的相对优势邻域粒度, 拓展了经典决策信息系统的相关方法, 并利用相对优势邻域粒度研究了区间决策表属性约简的启发式算法, 通过具体案例将得到的属性约简结果与代数约简进行了有效性验证, 进一步丰富和完善了信息系统属性约简理论.

关键词: 粗糙集; 区间值决策表; 相邻关系; 相对优势邻域粒度; 属性约简

中图分类号: TP18

文献标志码: A

开放科学(资源服务)标识码(OSID):



文章编号: 1673-9868(2024)05-0067-10

Attribute Reduction Based on Relative Dominance Neighborhood Granularity in Interval-valued Decision Tables

ZHANG Xiaoyan, LI Lu

College of Artificial Intelligent, Southwest University, Chongqing 400715, China

Abstract: In real life, a large number of data exists in the form of interval values. The interval value decision table is not based on the equivalence relation, and the traditional decision-making method cannot solve this problem. For this reason, this paper introduced the definitions of the adjacent relation and adjacent classes in interval valued decision tables. Moreover, the relative dominant neighborhood granularity was established in interval valued decision tables from adjacent classes, which expanded the relevant methods of classical decision information systems. Furthermore, the heuristic algorithm of attribute reduction was

收稿日期: 2023-05-30

基金项目: 国家自然科学基金项目(12371465); 重庆市自然科学基金项目(CSTB2023NSCQ-MSX1063).

作者简介: 张晓燕, 博士, 教授, 硕士研究生导师, 主要从事不确定性推理、人工智能的数学基础、粒计算与知识发现等研究.

established by using the relative advantage neighborhood granularity for the interval valued decision tables. The effectiveness of the obtained attribute reduction results and algebraic reduction was verified by specific cases. These results enriched and perfected the attribute reduction theory of information system.

Key words: rough set; interval-valued decision table; adjacent relation; relative dominant neighborhood granularity; attribute reduction

1965 年, 美国加利福尼亚州的控制论专家 L. A. Zadeh 引入了模糊集合的概念, 建立了模糊集合的基本理论框架, 并揭示了模糊集合与传统集合的关系^[1]. 这一概念的提出标志着模糊数学学科的诞生, 而模糊数学在解决许多经典数学难题上展现出了广泛的适用性. 模糊数学的研究成果已经在模糊控制、模糊聚类分析和模糊模式识别^[2-3]等领域取得了显著成效. 与此同时, 粗糙集理论可以很好地处理现实生活中的各种不确定性现象^[4-5]. 然而, 在经典粗糙集理论中, 对上下近似的定义要求过于严格, 从而限制了该理论的应用范围. 为了克服这些局限性, 研究人员在经典粗糙集理论的基础上进行了推广和延伸, 提出了决策粗糙集模型(DTRS)和程度粗糙集模型(GRS)^[6-7], 进一步完善了粗糙集理论.

在经典粗糙集和粗糙集的推广研究中, 属性约简起着至关重要的作用. 属性约简的目标是找到最小的属性子集, 尽可能保持决策精度的同时解决现实生活中数据信息之间的冗余问题, 进而找到对决策起决定性作用的条件属性数据. 然而, 在区间值决策表中, 如何有效地进行属性约简仍然是一个具有挑战性的问题. 为此, 本文将深入探索区间值决策表中基于相对优势邻域粒度的属性约简方法, 以提升决策精度. 已有研究在完备决策信息系统下提出了基于正域的属性约简方法^[8], 文献[9]等进一步推广了区间值决策信息系统的保正域属性约简方法. 此外, 还涌现了其他多种属性约简方法, 如基于相对知识粒度的属性约简^[10-11]和基于依赖度的区间集决策信息表属性约简^[12-15]等. 然而, 在区间值决策表的研究中, 基于相对优势邻域粒度的属性约简方法尚未充分探索. 本文在经典决策表相对优势邻域粒度的属性约简方法的基础上, 将其拓展应用于区间值决策表, 利用相对优势邻域粒度研究了区间决策表属性约简的启发式算法.

1 预备知识

本节先给出一些基本定义和相关性质, 详细可见文献[16].

区间值决策表指由论域 U 、条件属性集 C 、决策属性集 d , 以及属性值域 V 组成的四元组 $IVDT = \{U, AT = C \cup d, V, f\}$, 其中论域 U 是非空有限对象的集合, 条件属性集 C 是有限个条件属性的集合, 决策属性集 d 是有限个决策属性的集合, 且 $C \cap d = \emptyset$; 属性值域 $V = \bigcup_{a \in AT} V_a$, V_a 是任意属性 $a \in C \cup d$ 的值域; 信息函数 $f: U \times C \rightarrow V_a$ 满足对于 $\forall x_i \in U$, 属性 a 上的值 $f(x_i, a) \in V_a$ 是一个区间值, 决策属性值仍为单值.

设区间值决策表 $IVDT = \{U, AT = C \cup d, V, f\}$, 且属性子集 $B \subseteq C$, 阈值 $\alpha \in [0, 1]$, 则关于 B 的相邻关系定义为

$$NR_B^\alpha = \{(x_i, x_j) \subseteq U \times U \mid Dis \leq \alpha; \forall a_k \in B; x_i, x_j \in U\}$$

式中

$$Dis = \sqrt{\sum_{k=1}^{|B|} (P(x_i^{a_k} \geq x_j^{a_k}) - P(x_j^{a_k} \geq x_i^{a_k}))^2}$$

这里

$$P(x_i^{a_k} \geq x_j^{a_k}) = \min \left\{ 1, \max \left\{ \frac{a^+ - b^-}{(a^+ - a^-) + (b^+ - b^-)}, 0 \right\} \right\}$$

且

$$f(x_i, a_k) = [a^-, a^+], f(x_j, a_k) = [b^-, b^+]$$

对于区间值决策表 $IVDT = \{U, AT = C \cup d, V, f\}$, 基于相邻关系基础的相邻类为

$$NC_B^\alpha(x_i) = \{x_j \mid (x_i, x_j) \in NR_B^\alpha; x_j \in U\}$$

可知决策表 $IVDT$ 所有的相邻类构成一个对应于每一个元素的 $|U|$ 维粒化结构

$$N_B^\alpha = (NC_B^\alpha(x_1), NC_B^\alpha(x_2), \dots, NC_B^\alpha(x_{|U|}))$$

进一步, 区间值决策表 $IVDT = \{U, AT = C \cup d, V, f\}$ 对任意一个条件属性子集 $B \subseteq C$, $0 \leq \alpha_1 \leq \alpha_2 \leq 1$ 有

$$1) \forall x_i \in U, NC_C^{\alpha_1}(x_i) \subseteq NC_B^{\alpha_1}(x_i)$$

$$2) \forall x_i \in U, NC_B^{\alpha_1}(x_i) \subseteq NC_B^{\alpha_2}(x_i)$$

由于相对优势邻域粒度可以很直观地体现条件邻域粒化对决策属性划分的分辨力, 故可借助相邻关系导出粒化结构研究区间值决策表的属性约简方法.

2 区间值决策表中相对优势邻域粒度的属性约简

虽然经典相对优势邻域粒度能够体现优势邻域的粗细程度, 但却不适用于区间值决策表. 为此, 本节给出了基于区间相对优势粒度的启发式约简算法.

2.1 区间相对优势邻域粒度

首先, 区间值决策表的区间相对优势邻域粒度定义如下.

定义 1 设 $IVDT = \{U, AT = C \cup d, V, f\}$ 为一个区间值决策表, 对于任意一个条件属性子集 $B \subseteq C$ 及阈值 $\alpha \in [0, 1]$, 则 B 相对于 d 的区间相对优势邻域粒度为

$$RN(\alpha; B; d) = AN(\alpha; B) - AN(\alpha; B \cup d) = \sum_{i=1}^{|U|} \left(\frac{|NC_B^\alpha(x_i)|}{|U|} \right)^2 - \sum_{i,h=1}^{|U| \cdot m} \left(\frac{|NC_B^\alpha(x_i) \cap D_h|}{|U|} \right)^2$$

式中 D_h 为决策属性的决策类, 且 $\sum_{i=1}^{|U|} \left(\frac{|NC_B^\alpha(x_i)|}{|U|} \right)^2 = AN(\alpha; B)$, 及 $\sum_{i,h=1}^{|U| \cdot m} \left(\frac{|NC_B^\alpha(x_i) \cap D_h|}{|U|} \right)^2 = AN(\alpha; B \cup d)$.

性质 1 设 $IVDT = \{U, AT = C \cup d, V, f\}$ 为一个区间值决策表, 对于任意一个条件属性子集 $B \subseteq C$ 及阈值 $\alpha \in [0, 1]$, 则有

$$RN(\alpha; B; d) = \sum_{i=1}^{|U|} \left[\left(\sum_{h=1}^m \frac{|NC_B^\alpha(x_i) \cap D_h|}{|U|} \right)^2 - \sum_{h=1}^m \left(\frac{|NC_B^\alpha(x_i) \cap D_h|}{|U|} \right)^2 \right]$$

证 由于

$$\frac{U}{d} = \{D_1, D_2, \dots, D_m\}, D_h \cap D_j = \emptyset \quad 1 \leq h \leq j \leq m$$

因此对于 $\forall i \in [1, |U|]$, 可以得出

$$\frac{|NC_B^\alpha(x_i)|}{|U|} = \sum_{h=1}^m \frac{|NC_B^\alpha(x_i) \cap D_h|}{|U|}$$

于是, 性质 1 成立.

引理 1 设 $f(g_1, g_2, \dots, g_{m-1}, g_m) = \left(\sum_{s=1}^m g_s \right)^2 - \sum_{s=1}^m g_s^2$, 其中 $g_s (1 \leq s \leq m)$ 为函数自变量, 则有

$$1) f(g_1, g_2, \dots, g_{m-1}, g_m) = 2(g_1 g_2 + g_1 g_3 + \dots + g_{m-1} g_m);$$

2) 若 $0 \leq g_1 \leq t_1, \dots, 0 \leq g_m \leq t_m$, 则 $f(g_1, g_2, \dots, g_m) \leq f(t_1, t_2, \dots, t_m)$.

证 1) 很显然成立. 接下来证明 2). 由于

$$f(t_1, t_2, \dots, t_m) - f(g_1, g_2, \dots, g_m) = 2(t_1 t_2 + \dots + t_{m-1} t_m) - 2(g_1 g_2 + \dots + g_{m-1} g_m) = \\ 2(t_1 t_2 - g_1 g_2) + \dots + 2(t_{m-1} t_m - g_{m-1} g_m)$$

当且仅当 $t_1 = g_1, t_2 = g_2, \dots, t_m = g_m$ 时等号成立. 证毕.

性质 2 设 $IVDT = \{U, C \cup d, V, f\}$ 是一个区间值决策表, 对于任意两个条件属性子集 $A, B \subseteq C$ 以及阈值 $\alpha \in [0, 1]$, 则以下结论成立:

1) 若 $A \subseteq B$, 则 $RN(\alpha; B; d) \leq RN(\alpha; A; d)$;

2) 若 $0 \leq \alpha_1 \leq \alpha_2 \leq 1$, 则 $RN(\alpha_1; B; d) \leq RN(\alpha_2; B; d)$.

证 1) 令

$$\frac{|NC_B^\alpha(x_i) \cap D_h|}{|U|} = B_h^i, \quad \frac{|NC_A^\alpha(x_i) \cap D_h|}{|U|} = A_h^i$$

可知关于 B 和 A 的区间相对优势邻域粒度可以简化为

$$RN(\alpha; B; d) = \sum_{i=1}^{|U|} \left[\left(\sum_{h=1}^m B_h^i \right)^2 - \sum_{h=1}^m (B_h^i)^2 \right] \\ RN(\alpha; A; d) = \sum_{i=1}^{|U|} \left[\left(\sum_{h=1}^m A_h^i \right)^2 - \sum_{h=1}^m (A_h^i)^2 \right]$$

又因 $A \subseteq B$, 可知 $\forall x_i \in U$ 满足 $NC_B^\alpha(x_i) \subseteq NC_A^\alpha(x_i)$, 于是

$$0 \leq \frac{|NC_B^\alpha(x_i) \cap D_h|}{|U|} \leq \frac{|NC_A^\alpha(x_i) \cap D_h|}{|U|}$$

即 $0 \leq B_h^i \leq A_h^i$.

根据引理 1 可得

$$f(B_1^i, B_2^i, \dots, B_m^i) \leq f(A_1^i, A_2^i, \dots, A_m^i)$$

所以

$$RN(\alpha; A; d) \leq RN(\alpha; B; d)$$

2) 令

$$\frac{|NC_B^{\alpha_1}(x_i) \cap D_h|}{|U|} = B_h^{\alpha_1}, \quad \frac{|NC_B^{\alpha_2}(x_i) \cap D_h|}{|U|} = B_h^{\alpha_2}$$

则关于 α_1, α_2 的区间相对优势邻域粒度可以简化为

$$RN(\alpha_1; B; d) = \sum_{i=1}^{|U|} \left[\left(\sum_{h=1}^m B_h^{\alpha_1} \right)^2 - \sum_{h=1}^m (B_h^{\alpha_1})^2 \right] \\ RN(\alpha_2; B; d) = \sum_{i=1}^{|U|} \left[\left(\sum_{h=1}^m B_h^{\alpha_2} \right)^2 - \sum_{h=1}^m (B_h^{\alpha_2})^2 \right]$$

又因为 $0 \leq \alpha_1 \leq \alpha_2 \leq 1$, 则

$$\forall x_i \in U, NC_B^{\alpha_1}(x_i) \subseteq NC_B^{\alpha_2}(x_i)$$

所以

$$0 \leq \frac{|NC_B^{\alpha_1}(x_i) \cap D_h|}{|U|} \leq \frac{|NC_B^{\alpha_2}(x_i) \cap D_h|}{|U|}$$

即 $0 \leq B_h^{\alpha_1} \leq B_h^{\alpha_2}$.

再由引理 1 可得

$$f(B_1^{\alpha_1}, B_2^{\alpha_1}, \dots, B_m^{\alpha_1}) \leq f(B_1^{\alpha_2}, B_2^{\alpha_2}, \dots, B_m^{\alpha_2})$$

所以有

$$RN(\alpha; A; d) \leq RN(\alpha; B; d)$$

故

$$RN(\alpha_1; B; d) \leq RN(\alpha_2; B; d)$$

性质 3 设 $IVDT = \{U, C \cup d, V, f\}$ 为一个区间值决策表, 对于任意一个条件属性子集 $B \subseteq C$ 及阈值 $\alpha \in [0, 1]$, 有 $RN(\alpha; B; d) \in [0, |U| - 1]$.

特别地, 当粒化结构 N_B^α 最细时, 即 $\forall x_i \in U, NC_B^\alpha(x_i) = \{x_i\}$, 此时 $RN(\alpha; B; d) = 0$; 当粒化结构 N_B^α 最粗时, 即 $\forall x_i \in U, NC_B^\alpha(x_i) = \{U\}$, 此时 $RN(\alpha; B; d) = |U| - 1$.

由性质 2 和性质 3 可以进一步有属性的必要性和独立性的定义. 下文中均设阈值 $\alpha \in [0, 1]$.

定义 2 设 $IVDT = \{U, C \cup d, V, f\}$ 是一个区间值决策表, 对于 $\forall a \in C$, 若 $RN(\alpha; C - a; d) = RN(\alpha; C; d)$, 则称 a 是 C 中的不必要属性, 否则称 a 是必要属性.

定义 3 设 $IVDT = \{U, C \cup d, V, f\}$ 是一个区间值决策表, 由 C 中所有关于决策属性 d 的必要属性组成的集合, 称为属性集合 C 相对于决策属性 d 的核, 记为 $Core_C^\alpha(d)$.

任意一个条件属性子集 $B \subseteq C$, 若 B 中任意一个条件属性在 B 中相对于决策属性 d 都必要, 则称 B 是 C 的一个相对于决策属性 d 的属性约简, 且易证 $Core_C^\alpha(d)$ 恰好是全部属性约简的交集.

定义 4 设 $IVDT = \{U, C \cup d, V, f\}$ 是一个区间值决策表, $B \subseteq C$, 则 B 是 C 的一个相对于决策属性 d 的属性约简, 当且仅当

- 1) $RN(\alpha; B; d) = RN(\alpha; C; d)$
- 2) $\forall a \in B, RN(\alpha; B; d) \neq RN(\alpha; B - \{a\}; d)$

属性约简的核心主要来源于相对优势邻域粒度.

定义 5 对于区间值决策表 $IVDT = \{U, C \cup d, V, f\}$, 若 $\forall a \in C$, 则条件属性 a 关于属性集 C 相对于决策属性 d 的属性内重要度为

$$Sig^{inner}(\alpha; a; C; d) = RN(\alpha; C - \{a\}; d) - RN(\alpha; C; d)$$

另外, 对于 $\forall B \in C$ 以及 $\forall a \in C - B$, a 关于 B 相对于决策属性 d 的区间属性外重要度为

$$Sig^{outer}(\alpha; a; B; d) = RN(\alpha; B; d) - RN(\alpha; B \cup \{a\}; d)$$

由定义 5 可知, 区间属性内外重要度变化越大说明该属性对于决策的影响越大. 因此区间属性内重要度与区间属性外重要度两者提供了快速实现约简的属性选择原理.

性质 4 对于区间值决策表 $IVDT = \{U, C \cup d, V, f\}$, 若 $\forall a \in C$, 则条件属性 a 为 C 中对 d 必要的属性当且仅当 $Sig^{inner}(\alpha; a; C; d) > 0$, 即

$$Core_C^\alpha(d) = \{a \in C \mid Sig^{inner}(\alpha; a; C; d) > 0\}$$

证 若 a 为 C 中对 d 必要的属性, 则 $RN(\alpha; C; d) \neq RN(\alpha; C - \{a\}; d)$, 再根据区间相对优势邻域粒度的粒化单调性可知, $RN(\alpha; C; d) \leq RN(\alpha; C - \{a\}; d)$, 所以 $Sig^{inner}(\alpha; a; C; d) > 0$. 反之, 若 $Sig^{inner}(\alpha; a; C; d) > 0$, 则 $RN(\alpha; C; d) \leq RN(\alpha; C - \{a\}; d)$, 那么 $RN(\alpha; C; d) \neq RN(\alpha; C - \{a\}; d)$. 再根据定义 2 可知, a 为 C 中对 d 必要的属性, 于是性质成立.

2.2 相对优势邻域粒度属性约简的启发式算法

下面给出相对优势邻域粒度属性约简的启发式算法.

算法 1 基于区间值相对优势邻域粒度的属性约简算法

输入 区间值决策表 $IVDT$ 及阈值 α

输出 决策表 $IVDT$ 的一个属性约简 S

step1 计算 C 相对于 d 的区间相对优势邻域粒度 $RN(\alpha; C; d)$;

step2 设 $Core_C^a(d) = \emptyset, \forall a \in C$, 计算 $Sig^{inner}(\alpha; a; C; d)$, 若 $Sig^{inner}(\alpha; a; C; d) > 0$,

$Core_C^a(d) = Core_C^a(d) \cup \{a\}$, 并令 $S = Core_C^a(d)$;

step3 判断 $RN(\alpha; S; d) = RN(\alpha; C; d)$ 是否成立, 若成立则 goto step5, 否则 goto step4;

step4 进入 step4, 则说明 $\forall a \in \{C - S\}, RN(\alpha; C; d) \leq RN(\alpha; S; d)$, 计算 $Sig^{outer}(\alpha; a; S; d)$, 将 $\max(Sig^{outer}(\alpha; a^*; S; d))$ 对应属性 a^* 并到 S 中, 重复该步骤过程直到满足 $RN(\alpha; S; d) = RN(\alpha; C; d)$, 转到 step5;

step5 输出 S

算法 1 的时间复杂度分析如下:

step1 计算 $RN(\alpha; C; d)$ 的时间复杂度为

$$O(|U|^2 \cdot |C| \cdot |\frac{U}{d}|)$$

step2 对于 $\forall a \in C$, 计算 $Sig^{inner}(\alpha; a; C; d)$ 的时间复杂度为

$$O(|U|^2 \cdot |C|^2 \cdot |\frac{U}{d}|)$$

step3 计算 $RN(\alpha; S; d)$ 的时间复杂度为

$$O(|U|^2 \cdot |S| \cdot |\frac{U}{d}|)$$

step4 将属性 a^* 添加到属性约简集中, 对 $\forall a \in \{C - S\}$, 计算 $Sig^{outer}(\alpha; a; S; d)$, 时间复杂度为

$$O(|U|^2 \cdot |S| \cdot |C - S| \cdot |\frac{U}{d}|)$$

最复杂情况下, step4 的复杂度为

$$O(|U|^2 \cdot [|C| \cdot 1 + (|C| - 1) \cdot 2 + \dots + 2 \cdot (|C| - 1) + 1 \cdot |C|] \cdot |\frac{U}{d}|)$$

从而此式的复杂度不超过

$$O\left(\frac{|U|^2 \cdot (|C| + 1)^2 \cdot |C| \cdot |\frac{U}{d}|}{4}\right)$$

所以算法 1 的时间复杂度为

$$O\left(\frac{|U|^2 \cdot (|C| + 1)^2 \cdot |C| \cdot |\frac{U}{d}|}{4}\right)$$

由算法 1 可知, 条件属性子集基数越大, 相对优势邻域粒度越小.

2.3 属性约简与代数约简等价性证明

基于正域的属性约简是粗糙集理论中常用的方法之一, 用于在决策信息系统中降低数据维度并提高决策效率. 该方法通过寻求能够唯一确定决策属性值的最小条件属性子集(即正域), 并删除冗余属性, 构建简化后的数据集. 基于正域的属性约简方法在保持决策精度的同时减少了数据冗余, 提高了决策过程的效率, 为处理大规模、高维度的数据提供了有效的工具和技术.

接下来将证明在区间值决策表中, 区间相对优势邻域粒度的属性约简与代数约简是等价的.

定义 6 设有区间值决策表 $IVDT = \{U, C \cup d, V, f\}$, 若对于该决策表的 $\forall NC_C^a(x_i) \in N_C^a, \exists D_h \in$

$\left(\frac{U}{d}\right)$, 有 $NC_C^a(x_i) \subseteq D_h$, 则称该决策表为协调区间值决策表, 否则为不协调的.

性质 5 区间值决策表 $IVDT = \{U, C \cup d, V, f\}$ 是协调的, 当且仅当 $RN(\alpha; C; d) = 0$.

证 “ \Rightarrow ” 由于 $IVDT$ 是协调区间值决策表, 所以

$$\forall NC_C^a(x_i) \in N_C^a, \exists D_h \in \left(\frac{U}{d}\right)$$

有

$$NC_C^a(x_i) \subseteq D_h$$

又因为

$$\frac{U}{d} = \{D_1, D_2, \dots, D_m\}, D_h \cap D_j = \emptyset \quad 1 \leq h \leq j \leq m$$

故 $\forall i \in [1, |U|]$.

另外,

$$RN(\alpha; B; d) = \sum_{i=1}^{|U|} \left[\left(\sum_{h=1}^m \frac{|NC_B^a(x_i) \cap D_h|}{|U|} \right)^2 - \sum_{h=1}^m \left(\frac{|NC_B^a(x_i) \cap D_h|}{|U|} \right)^2 \right]$$

故可得

$$RN(\alpha; C; d) = 0$$

“ \Leftarrow ” 若 $RN(\alpha; C; d) = 0$, 则由

$$RN(\alpha; B; d) = AN(\alpha; B) - AN(\alpha; B \cup d)$$

可知

$$AN(\alpha; B) = AN(\alpha; B \cup d)$$

又由

$$RN(\alpha; B; d) = AN(\alpha; B) - AN(\alpha; B \cup d) = \sum_{i=1}^{|U|} \left(\frac{|NC_B^a(x_i)|}{|U|} \right)^2 - \sum_{i,h=1}^{|U|,m} \left(\frac{|NC_B^a(x_i) \cap D_h|}{|U|} \right)^2$$

可得

$$\forall i \in [1, |U|] \quad \frac{|NC_B^a(x_i)|}{|U|} = \sum_{h=1}^m \left(\frac{|NC_B^a(x_i) \cap D_h|}{|U|} \right)$$

也即

$$\forall N_B^a(x_i) \in N_B^a$$

于是有

$$N_B^a(x_i) = N_B^a(x_i) \cap \left(\frac{U}{d}\right)$$

因此有

$$\forall NC_C^a(x_i) \in N_C^a, \exists D_h \in \left(\frac{U}{d}\right)$$

于是 $NC_C^a(x_i) \subseteq D_h$.

故 $IVDT$ 是协调区间值决策表.

性质 6 设 $IVDT = \{U, C \cup d, V, f\}$ 是一个协调区间值决策表, 若 $B \subseteq C$, 则

$$RN(\alpha; B; d) = RN(\alpha; C; d) \Leftrightarrow pos_B^a(D) = pos_C^a(D)$$

式中 $pos_B^a(D)$, $pos_C^a(D)$ 分别指决策类 D 关于条件属性集 B, C 的正域.

证 先证充分性. 因为 $IVDT$ 是协调区间值决策表, 可知 $pos_C^{\alpha}(D) = U$; 由性质 6 可知 $RN(\alpha; C; d) = 0$, 因为 $RN(\alpha; B; d) = RN(\alpha; C; d)$, 所以 $RN(\alpha; B; d) = 0$.

由

$$RN(\alpha; B; d) = AN(\alpha; B) - AN(\alpha; B \cup d)$$

可得

$$AN(\alpha; B) = AN(\alpha; B \cup d)$$

又由

$$N_B^{\alpha} = N_B^{\alpha} \cap \left(\frac{U}{d}\right)$$

得到

$$pos_B^{\alpha}(D) = U$$

因此

$$pos_B^{\alpha}(D) = pos_C^{\alpha}(D)$$

再证必要性. 因为 $IVDT$ 是协调区间值决策表, 所以 $pos_C^{\alpha}(D) = U$, 由性质 6 知

$$RN(\alpha; C; d) = 0$$

由

$$pos_B^{\alpha}(D) = pos_C^{\alpha}(D)$$

可得

$$pos_B^{\alpha}(D) = U$$

因此

$$N_B^{\alpha} = N_B^{\alpha} \cap \left(\frac{U}{d}\right)$$

即

$$AN(\alpha; B) = AN(\alpha; B \cup d)$$

所以

$$RN(\alpha; B; d) = 0$$

因此

$$RN(\alpha; B; d) = RN(\alpha; C; d)$$

3 案例分析

区间决策表 $IVDT$ 如表 1 所示^[8], 其中 $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$ 代表 10 个对象, $C = \{a_1, a_2, a_3, a_4\}$ 为条件属性集, d 为决策属性集. 文献[8] 计算表明当阈值 $\alpha = 0.65$ 时该区间值决策表的属性约简为 $S_1 = \{a_1, a_2\}$ 和 $S_2 = \{a_2, a_3\}$.

设阈值 $\alpha = 0.65$, 计算可知条件属性子集增链为

$$A_1 = \{a_1\} \subset A_2 = \{a_1, a_2\} \subset A_3 = \{a_1, a_2, a_3\} \subset A_4 = \{a_1, a_2, a_3, a_4\}$$

阈值增链为

$$\alpha_1 = 0.55 < \alpha_2 = 0.65 < \alpha_3 = 0.75 < \alpha_4 = 0.85 < \alpha_5 = 0.95$$

对条件属性子集增链和阈值增链分别计算区间相对优势邻域粒度.

对条件属性子集增链有

$$RN(0.65; A_1; d) = 0.64 > RN(0.65; A_2; d) =$$

$$0.34 > RN(0.65; A_3; d) = 0.34 > RN(0.65; A_4; d) = 0.34$$

对于阈值增链有

$$RN(0.55; B; d) = RN(0.65; B; d) = 0.34 > RN(0.75; B; d) = RN(0.85; B; d) = RN(0.95; B; d) = 0$$

故区间相对优势粒度满足属性和阈值的双重粒化单调性, 即性质 3 成立. 同时, 计算结果都满足双界范围: $[0, |U| - 1]$, 即性质 4 成立.

当 $\alpha = 0.65$ 时, 由算法 1 可计算出该决策表的一个属性约简.

表 1 给定一区间决策表

U	a_1	a_2	a_3	a_4	d
x_1	[0.5, 0.8]	[0.2, 0.4]	[0.6, 0.9]	[0.1, 0.5]	1
x_2	[0.4, 0.7]	[0.2, 0.3]	[0.5, 0.8]	[0.1, 0.4]	0
x_3	[0.2, 0.4]	[0.4, 0.5]	[0.3, 0.5]	[0.3, 0.6]	0
x_4	[0.5, 0.7]	[0.3, 0.4]	[0.6, 0.8]	[0.2, 0.5]	1
x_5	[0.5, 0.7]	[0.2, 0.4]	[0.6, 0.8]	[0.1, 0.5]	1
x_6	[0.5, 0.8]	[0.2, 0.3]	[0.6, 0.9]	[0.1, 0.4]	1
x_7	[0.2, 0.5]	[0.3, 0.5]	[0.3, 0.6]	[0.2, 0.6]	0
x_8	[0.5, 0.7]	[0.3, 0.4]	[0.6, 0.8]	[0.2, 0.5]	1
x_9	[0.2, 0.4]	[0.4, 0.5]	[0.3, 0.5]	[0.3, 0.6]	0
x_{10}	[0.6, 0.9]	[0.3, 0.4]	[0.7, 1.0]	[0.2, 0.5]	1

1) 计算相对优势邻域粒度 $RN(0.65; C; d) = 0.34$.

2) 设 $Core_C^\alpha(d) = \emptyset$, 计算区间属性内重要度

$$Sig^{inner}(0.65; a_1; C; d) = 0.00$$

$$Sig^{inner}(0.65; a_2; C; d) = 0.30$$

$$Sig^{inner}(0.65; a_3; C; d) = 0.00$$

$$Sig^{inner}(0.65; a_4; C; d) = 0.00$$

选择区间属性内重要度大于零的属性并入到核心 $Core_C^{0.65}(d)$ 中, 即 $Core_C^{0.65}(d) = \{a_2\}$, 实现对 S 的更新.

3) 计算 S 关于 d 的区间相对优势邻域粒度 $RN(0.65; S; d) = 0.84 > RN(0.65; C; d) = 0.34$.

4) $\forall a \in \{C - S\}$ 计算区间属性外重要度

$$Sig^{outer}(0.65; a_1; S; d) = 0.50$$

$$Sig^{outer}(0.65; a_3; S; d) = 0.50$$

$$Sig^{outer}(0.65; a_4; S; d) = 0.50$$

选择区间属性外重要度最大的条件属性 a^* 并入到 S 中, 实现对核心 S 的更新. 又因为属性 a_1 与属性 a_3 的外重要度都为 0.50. 为保证一般性, 对属性 a_1 与属性 a_3 都进行验证, 可得

$$RN(0.65; \{a_1, a_2\}; d) = 0.34 = RN(0.65; C; d) = 0.34$$

$$RN(0.65; \{a_2, a_3\}; d) = 0.34 = RN(0.65; C; d) = 0.34$$

5) 输出 $S_1 = \{a_1, a_2\}$ 或 $S_2 = \{a_2, a_3\}$.

即 $S_1 = \{a_1, a_2\}$ 和 $S_2 = \{a_2, a_3\}$ 为该区间值决策表的两个属性约简, 这与文献[8]基于正域属性约简的结果相同.

4 结论

本文给出了区间相对优势邻域粒度等概念,在此基础上构建出相应的启发式约简算法,并证明了在协调区间值决策表上,区间相对优势邻域粒度表示与代数表示是等价的.同时,结合具体实例分析验证了区间相对优势邻域粒度及属性约简算法的合理性.本文研究结果可应用于环境监测数据的属性约简,从而进一步选出对环境质量评估和污染防治具有关键影响的属性.然而,由于现实生活中的决策信息系统可能存在不协调或不完备的情况,将来可进一步研究不协调或不完备区间决策信息系统下的属性约简.同时,目前区间值决策表中基于相对优势邻域粒度的属性约简中阈值并未解释选取标准,这也是该领域未来需要研究的工作.

参考文献:

- [1] ZADEH L A. Fuzzy Sets [J]. Information and Control, 1965, 8(3): 338-353.
- [2] 徐淑静. 基于 FPGA 的并行模糊控制算法的研究与实现 [D]. 成都: 中国科学院大学(中国科学院光电技术研究所), 2022.
- [3] 陈娇花. 基于改进 K-means 模糊聚类的区域健康大数据智能分析方法研究 [J]. 电子设计工程, 2022, 30(19): 30-34.
- [4] PAWLAK Z. Rough Sets: Theoretical Aspects of Reasoning about Data [M]. Dordrecht: Kluwer Academic Publishers, 1991.
- [5] 余建航, 徐伟华. 序信息系统下基于精度与程度“逻辑与”和“逻辑或”的粗糙集 [J]. 计算机科学, 2016, 43(2): 269-272.
- [6] 姚晟, 吴照玉, 陈菊, 等. 基于决策理论粗糙集的一种新属性约简方法 [J]. 微电子学与计算机, 2019, 36(5): 76-81.
- [7] 陈德刚, 徐伟华, 李金海, 等. 粒计算基础教程 [M]. 北京: 科学出版社, 2019.
- [8] 陈华峰, 龙建武, 瞿先平. 区间值决策信息系统中基于正域的属性约简 [J]. 重庆理工大学学报(自然科学), 2019, 33(11): 130-136.
- [9] 苗夺谦, 范世栋. 知识的粒度计算及其应用 [J]. 系统工程理论与实践, 2002, 22(1): 48-56.
- [10] 陈玉明, 吴克寿, 谢荣生. 基于相对知识粒度的决策表约简 [J]. 山东大学学报(工学版), 2012, 42(6): 8-12.
- [11] 唐鹏飞, 莫智文, 谢鑫. 区间值决策表中基于相对知识粒度的属性约简 [J]. 重庆理工大学学报(自然科学), 2021, 35(11): 286-292.
- [12] 唐鹏飞, 张贤勇, 莫智文. 基于依赖度的区间集决策信息表属性约简 [J]. 计算机应用研究, 2021, 38(11): 3300-3303, 3309.
- [13] 唐玉凯. 不完备决策系统下的高效属性约简算法研究 [D]. 烟台: 烟台大学, 2021.
- [14] 梁美社, 米据生, 赵天娜. 广义优势多粒度直觉模糊粗糙集及规则获取 [J]. 智能系统学报, 2017, 12(6): 883-888.
- [15] 王翠翠, 李宝萍, 毛军军. 基于区间二型模糊熵的多属性决策方法 [J]. 计算机工程与应用, 2017, 53(18): 132-136, 162.
- [16] 徐伟华. 序信息系统与粗糙集 [M]. 北京: 科学出版社, 2013.

责任编辑 张枸