

加权自适应多粒度决策理论粗糙集模型

张宇, 汪小燕

(安徽工业大学 计算机科学与技术学院, 安徽 马鞍山 243032)

摘要: 多粒度粗糙集涉及属性权重时, 通常利用指定的方式设定, 而对于属性颗粒结构对应的上、下近似集计算所需的概率阈值也常依赖于专家建议设定, 这使得现有粗糙集模型在实际应用中缺乏适应性。为此, 提出1种加权自适应多粒度决策理论粗糙集模型 (weighted adaptive multi-granulation decision-theoretic rough sets, WAMG-DTRS)。通过信息增益计算属性粒度的权重, 并通过设置权重系数控制颗粒结构数目。同时, 利用单参数决策理论粗糙集中的阈值公式确定属性颗粒结构中不同对象对应的上、下近似集计算所需的概率阈值, 从而使模型更好地适应实际应用需求。在 WAMG-DTRS 模型的基础上, 进一步构建5种平均加权自适应多粒度决策理论粗糙集模型, 以进一步提高模型应用的适应性。通过实例和实验验证, 结果表明: 调整权重系数可灵活调整 WAMG-DTRS 模型的上、下近似集规模; 不同平均条件下的 WAMG-DTRS 模型展现出不同的下近似集特性, 并具备 WAMG-DTRS 模型灵活调控权重系数的能力。通过综合考虑不同平均条件, 可以进一步提升模型的适应性。

关键词: 自适应阈值; 权重; 多粒度; 决策理论; 粗糙集; 信息增益; 粒计算

中图分类号: TP 18 **文献标志码:** A **doi:** 10.12415/j.issn.1671-7872.23182



Weighted Adaptive Multi-granularity Decision-theoretic Rough Set Model

ZHANG Yu, WANG Xiaoyan

(School of Computer Science & Technology, Anhui University of Technology, Maanshan 243032, China)

Abstract: In multi-granulation rough sets, attribute weights are typically assigned through predetermined methods, while the probability thresholds required for computing upper and lower approximations of attribute granular structures often rely on expert recommendations. This makes existing rough set models lack adaptability in practical applications. To address this, a weighted adaptive multi-granularity decision-theoretic rough set (WAMG-DTRS) model was proposed. The weights of the attribute granularity were calculated according to information gain, and the number of granular structures was controlled by the setting weight coefficients. The probability thresholds required for calculating the upper and lower approximate sets corresponding to different objects under the attribute granular structure were determined by the threshold formula in the single-parameter decision theory rough sets to better adapt to practical applications. Based on this model, five types of average weighted adaptive multi-granularity decision-theoretic rough sets models were constructed, further improving the adaptability of the model. The feasibility of these models was proved through practical examples and experiments. The results show that adjusting the weight coefficients can flexibly control the scale of the upper and lower approximation sets in the WAMG-DTRS model. The average weighted adaptive multi-granulation decision-theoretic rough set models under different average conditions exhibit varying characteristics in the lower approximation sets and retain the ability of the WAMG-DTRS

收稿日期: 2023-12-23

基金项目: 国家自然科学基金青年基金项目(61806004); 安徽省自然科学基金项目(1808085MF196)

作者简介: 张宇(1997—), 女, 安徽阜阳人, 硕士生, 主要研究方向为数据挖掘、粗糙集理论、粒计算。

通信作者: 汪小燕(1974—), 女, 安徽桐城人, 副教授, 主要研究方向为数据挖掘、粗糙集理论、粒计算。

引文格式: 张宇, 汪小燕. 加权自适应多粒度决策理论粗糙集模型[J]. 安徽工业大学学报(自然科学版), 2025, 42(2):178-188.

model to flexibly adjust weight coefficients. By considering different average conditions, the adaptability of the model can be further enhanced.

Keywords: adaptive threshold; weight coefficient; multi-granularity; decision theory; rough sets; information gain; granular computing

随着计算机和互联网技术的迅猛发展,数据的不完整性、不精确性和不确定性问题日益凸显。这些问题不仅对企业的战略制定、项目规划和资源分配产生深远的影响,还使得机器学习任务变得更加复杂,增加数据提取有用知识和规律的难度。粗糙集理论^[1]是 1 种处理不确定性和不完整性的数学工具,在数据挖掘^[2-4]、模式识别^[5]、智能信息处理^[6]等领域展现了广泛的应用潜力。然而,由于实际应用中数据的多样性和差异性,粗糙集理论的适用性受到限制。因此,研究如何提升粗糙集理论模型的适应性,增强模型对信息不确定性的处理能力,对于深入理解复杂系统的内在规律,并为不同领域系统决策提供有力支持具有重要意义。

自粗糙集理论提出以来,学者们不断推动该理论的发展。针对概率粗糙集中给定风险阈值参数缺乏合理语义解释的问题,Yao^[7]在传统粗糙集理论的基础上,结合贝叶斯风险最小值和三支决策理论^[8]的思想,提出决策理论粗糙集 (decision-theoretic rough sets, DTRS),通过考虑风险决策问题,提供 1 种更准确和有效的决策支持方法。然而,DTRS 在实际应用中面临概率阈值参数难以确定的问题。为此,Suo 等^[9]引入数据驱动的损失函数矩阵和基于样本的显著性和概率,提出单参数获取相对颗粒结构概率阈值的方法,在传统 DTRS 模型中仅需预先设置 1 个参数,但模型的灵活性受到限制。为了提高 DTRS 模型的实用性和灵活性,学者们进一步提出了多粒度决策理论粗糙集 (multi-granulation decision-theoretic rough set, MG-DTRS)^[10],通过引入多个粒度的概念,能够更全面地描述数据的特征和规律。随着信息量的增长,MG-DTRS 模型的适用性受到挑战。为此,进一步提出了多种改进措施。Zhang 等^[11]将 MG-DTRS 的概念应用于犹豫模糊语言背景下的双空间框架中,解决了犹豫模糊群体决策中的信息分析和融合问题;Li 等^[12]针对 MG-DTRS 仅考虑颗粒结构中的绝对定量(相对定量)信息而忽略相对定量(绝对定量)信息问题,提出多粒度双定量决策理论粗糙集模型,实现了信息反映的全面性;Qian 等^[13]针对多层次粒度的成本敏感决策问题,提出多粒度顺序三支决策广义模型,帮助决策者在面对 3 个成对不相交区域时明确行动策略;此

外,Zhang 等^[14]针对 MG-DTRS 模型的固有弱点(主观给定概率参数),提出自适应多粒度决策理论粗糙集 (adaptive multi-granulation decision-theoretic rough sets, AMG-DTRS) 模型,通过设置补偿系数 ζ 自适应地获取概率阈值,显著增强模型的灵活性和实用性。在持续改进 MG-DTRS 模型以适应复杂应用需求的同时,学者们也认识到合理设定属性粒度权重对提升决策可靠性和模型效率的重要性。为此,提出了多种属性粒度权重设定方法:张明等^[15]通过考虑粒度质量赋予权重;Guo 等^[16]基于决策树分类准确度对粒度进行加权;Pang 等^[17]提出了专家和属性的多粒度权重挖掘与融合方法;Liu 等^[18]则通过最小化样本点到类别中心点的距离确定属性权重,克服了传统层次分析法 (analytic hierarchy process, AHP) 等方法的局限性。

尽管现有研究在 MG-DTRS 模型的扩展和属性粒度权重设定方面已取得了一定成果,但在属性粒度权重设定和平均多粒度决策理论粗糙集模型的细化上仍有不足,限制了模型的适用性。现实应用中,合理设置概率阈值和属性粒度权重是提高模型适用性的关键。为此,提出 1 种加权自适应多粒度决策理论粗糙集 (weighted adaptive multi-granulation decision-theoretic rough sets, WAMG-DTRS) 模型,在此基础上,根据不同的平均条件构建 5 种平均加权自适应多粒度决策理论粗糙集 (average weighted adaptive multi-granulation decision-theoretic rough sets, AWAMG-DTRS) 模型,以期通过灵活整合不同粒度下的信息特征,进一步提升模型决策的准确性和适应性,为复杂决策问题提供更有效的解决方案。

1 预备知识

1.1 多粒度决策理论粗糙集

对于传统的 DTRS 模型,其并未考虑从多个粒度空间分析问题。为此,Qian 等^[10]结合 DTRS 与多粒度粗糙集的思想,提出了多粒度决策理论粗糙集 (MG-DTRS),其中包括乐观 MG-DTRS、悲观 MG-DTRS、平均 MG-DTRS 模型。这些模型通过引入多粒度视角,能够更全面地描述和分析复杂数据,从而提升了模型的适用性和决策能力。

定义 1^[10] 假设 $(U, C \cup \{d\})$ 为决策信息系统 (decision information system, DIS), $A = \{A_1, A_2, A_3, \dots, A_m\}$

$(A_i \subseteq C, i = 1, 2, \dots, m)$, 定义 $X (X \subseteq U)$ 的乐观 MG-DTRS 下近似、上近似为:

$$\sum_{i=1}^m A_i^{O,\alpha}(X) = \{x : P(X|[x]_{A_1}) \geq \alpha \vee P(X|[x]_{A_2}) \geq \alpha \vee \dots \vee P(X|[x]_{A_m}) \geq \alpha, x \in U\} \quad (1)$$

$$\sum_{i=1}^m A_i^{O,\beta}(X) = U - \{x : P(X|[x]_{A_1}) \leq \beta \wedge P(X|[x]_{A_2}) \leq \beta \wedge \dots \wedge P(X|[x]_{A_m}) \leq \beta, x \in U\} \quad (2)$$

定义 2^[10] 假设 $(U, C \cup \{d\})$ 为 DIS, $A = \{A_1, A_2, A_3, \dots, A_m\}$ ($A_i \subseteq C, i = 1, 2, \dots, m$), 定义 $X (X \subseteq U)$ 的悲观 MG-

DTRS 下近似、上近似为:

$$\sum_{i=1}^m A_i^{P,\alpha}(X) = \{x : P(X|[x]_{A_1}) \geq \alpha \wedge P(X|[x]_{A_2}) \geq \alpha \wedge \dots \wedge P(X|[x]_{A_m}) \geq \alpha, x \in U\} \quad (3)$$

$$\sum_{i=1}^m A_i^{P,\beta}(X) = U - \{x : P(X|[x]_{A_1}) \leq \beta \vee P(X|[x]_{A_2}) \leq \beta \vee \dots \vee P(X|[x]_{A_m}) \leq \beta, x \in U\} \quad (4)$$

定义 3^[10] 假设 $(U, C \cup \{d\})$ 为 DIS, $A = \{A_1, A_2, A_3, \dots, A_m\}$ ($A_i \subseteq C, i = 1, 2, \dots, m$), 定义 $X (X \subseteq U)$ 的平均

MG-DTRS 上、下近似为:

$$\sum_{i=1}^m A_i^{M,\alpha}(X) = \left\{ x \in U : \frac{(\rho(X|[x]_{A_1}) + \rho(X|[x]_{A_2}) + \dots + \rho(X|[x]_{A_m}))}{m} \geq \alpha \right\} \quad (5)$$

$$\sum_{i=1}^m A_i^{M,\beta}(X) = U - \left\{ x \in U : \frac{(\rho(X|[x]_{A_1}) + \rho(X|[x]_{A_2}) + \dots + \rho(X|[x]_{A_m}))}{m} \leq \beta \right\} \quad (6)$$

其中: U 为非空有限对象的集合; $A_1, A_2, A_3, \dots, A_m$ 为 m 个条件的属性粒度; C 为条件属性粒度集合; d 为决策属性; X 为决策类别的对象集; $[x]_{A_i}$ 为 x 关于 A_i 的等价类; $\rho(X|[x]_{A_i})$ 表示条件概率; α, β 为概率阈值。

1.2 单参数概率阈值和信息增益权重

在 DTRS 模型中, 参数确定是 1 个难点, 限制了模型在数据处理中的应用。为此, 单参数决策理论粗糙集 (single-parameter decision-theoretic rough sets, SP-DTRS)^[9] 基于样本的显著性和概率, 引入数据驱动的损失函数矩阵。该模型只需预先设置 1 个参数, 即可计算 4 种场景不同对象对应颗粒结构的概率阈值。这不仅简化了 DTRS 模型的参数设定, 还使得 DTRS 模型更易在实际问题中得到应用和推广。SP-DTRS 模型中, 单参数计算属性颗粒结构下不同对象对应的概率阈值, 具体包括如下 4 种情况:

当 $|X| \neq 0, X \cup X^c = U, X^c$ 为非决策类别的对象集, ζ 为补偿系数 ($0 < \zeta \leq 1$), $S(X|[x]_{A_i}) = \frac{|X \cap [x]_{A_i}|}{|X|} = S, S^c(X|[x]_{A_i}) = \frac{|X^c \cap [x]_{A_i}|}{|X|} = S^c$ 。 α_i, β_i 表示属性粒度 A_i 下不同对象的概率阈值, $\alpha_i = (\alpha_1^s, \alpha_2^s, \alpha_3^s, \alpha_4^s)$, $\beta_i = (\beta_1^s, \beta_2^s, \beta_3^s, \beta_4^s)$ 。

情况 1^[9]: 当 $\rho(X|[x]_{A_i}) - \zeta > 0$ 且 $\rho(X|[x]_{A_i}) + \zeta < 1$ 时

$$\alpha_1^s = \frac{(\zeta S + S^c - \zeta S^c) + \theta}{2(S + S^c)}; \beta_1^s = \frac{(\zeta S + S + 2S^c - \zeta S^c) - \varphi}{2(S + S^c)} \quad (7)$$

其中: $\theta = \sqrt{(\zeta S + S^c - \zeta S^c)^2 + 4\zeta S^c(S + S^c)}$;
 $\varphi = \sqrt{(\zeta S + S + 2S^c - \zeta S^c)^2 + 4(S + S^c)(\zeta S^c - S^c)}$ 。

情况 2^[9]: 当 $\rho(X|[x]_{A_i}) \leq \zeta$ 且 $\rho(X|[x]_{A_i}) + \zeta < 1$ 时

$$\alpha_2^s = 1; \beta_2^s = \frac{(S + 2S^c - \zeta S^c) - f}{2S^c} \quad (8)$$

其中 $f = \sqrt{(S + 2S^c - \zeta S^c)^2 + 4S^c(\zeta S^c - S^c)}$ 。

情况 3^[9]: 当 $\rho(X|[x]_{A_i}) > \zeta$ 且 $\rho(X|[x]_{A_i}) + \zeta \geq 1$ 时

$$\alpha_3^s = \frac{(\zeta S - S^c) + \sqrt{(\zeta S - S^c)^2 + 4S^c S}}{2S}; \beta_3^s = 0 \quad (9)$$

情况 4^[9]: 当 $\rho(X|[x]_{A_i}) \leq \zeta$ 且 $\rho(X|[x]_{A_i}) + \zeta \geq 1$ 时

$$\alpha_4^s = 1; \beta_4^s = 0 \quad (10)$$

粗糙集中的属性权重^[19] 用于衡量不同属性对决策属性的重要程度, 在数据特征重要性评估、决策分类、属性约简等方面具有重要意义。文中通过计算属性的信息增益^[20] 来确定属性粒度权重。若 $H(Y)$ 为信息熵, $H(Y) = -\sum_{j=1}^n p_j \log_2 p_j$, 其中 n 为事物的分类, p_j 为事物类型 j 的概率; $H(Y|A_i)$ 为条件信息熵, $H(Y|A_i) = \sum_{i=1}^m p_i H(Y|A_i)$, $p_i = P(A_i)$; $g(Y|A_i)$ 为信息增益, $g(Y|A_i) = H(Y) - H(Y|A_i)$, 则 (属性粒度) 权重可表示为:

$$w_{A_i} = \frac{g(Y|A_i)}{\sum_{A_i \in A} g(Y|A_i)} = \frac{H(Y) - H(Y|A_i)}{\sum_{A_i \in A} (H(Y) - H(Y|A_i))} \quad (11)$$

2 加权自适应多粒度决策理论粗糙集

信息增益^[21]是衡量 1 个属性对决策属性分类能力的重要指标。为避免在乐观情况下下近似集过于宽松以及在悲观情况下下近似集过于严谨的问题, 文中根据信息增益计算属性粒度权重, 并通过设定合适的权重参数来灵活调整颗粒结构的数量, 进而提出 WAMG-DTRS 模型。

定义 4 假设 $(U, C \cup \{d\})$ 为 DIS,

$A = \{A_1, A_2, A_3, \dots, A_m\} (A_i \subseteq C, i = 1, 2, \dots, m)$, 对于任意 $\zeta \in (0, 1), 0 < \gamma \leq 1, w = \{w_{A_1}, w_{A_2}, \dots, w_{A_m}\}$ 为各属性粒度对应的权重, $T_1 \subseteq A,$

$$T_1 = \left\{ A_i | A_i \in A \wedge \rho(X|[x]_{A_i}) \geq \alpha_i \wedge \sum_{i=1}^{|T_1|} w_{A_i} \geq \gamma \right\}, \quad T_2 \subseteq A,$$

$$T_2 = \left\{ A_i | A_i \in A \wedge \rho(X|[x]_{A_i}) \geq \beta_i \wedge \sum_{i=1}^{|T_2|} w_{A_i} \geq \gamma \right\}, \quad \text{定义}$$

$X (X \subseteq U)$ 的 WAMG-DTRS 下近似和上近似为:

$$\begin{aligned} \underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^\gamma} (X)} &= \\ \bigcup_{\substack{T_1 \subseteq A \wedge \\ \sum_{i=1}^{|T_1|} w_{A_i} \geq \gamma}} \left\{ x \in U : \bigwedge_{A_i \in T_1} \rho(X|[x]_{A_i}) \geq \alpha_i \right\} \end{aligned} \quad (12)$$

$$\begin{aligned} \overline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)^\gamma} (X)} &= \\ \bigcup_{\substack{T_2 \subseteq A \wedge \\ \sum_{i=1}^{|T_2|} w_{A_i} \geq \gamma}} \left\{ x \in U : \bigwedge_{A_i \in T_2} \rho(X|[x]_{A_i}) > \beta_i \right\} \end{aligned} \quad (13)$$

其中 γ 为权重的控制系数, γ 越大, 下近似越小; γ 越小, 下近似越大。 w_{A_i} 为属性粒度 A_i 的权重; $\alpha_i > \beta_i$ 。

称 $\left(\underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^\gamma} (X)}, \overline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)^\gamma} (X)} \right)$ 为集合 X 关于 A 的

WAMG-DTRS 模型的粗糙集, 可得到 WAMG-DTRS 模型的正域、负域和边界域为:

$$P_w(X) = \underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^\gamma} (X)}$$

$$N_w(X) = U - \overline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)^\gamma} (X)}$$

$$B_w(X) = \overline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)^\gamma} (X)} - \underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^\gamma} (X)}$$

定理 1 假设 $(U, C \cup \{d\})$ 为 DIS,

$A = \{A_1, A_2, A_3, \dots, A_m\} (A_i \subseteq C, i = 1, 2, \dots, m)$, 对于任意 $\zeta \in (0, 1), 0 < \gamma \leq 1, X \subseteq U$, 可得

$$1) \quad 0 < \gamma_1 < \gamma_2 \leq 1, \quad \underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^{\gamma_2}} (X)} \subseteq \underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^{\gamma_1}} (X)};$$

$$\underline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)^{\gamma_2}} (X)} \subseteq \underline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)^{\gamma_1}} (X)}。$$

$$2) \quad X \subseteq Y \subseteq U, \quad \underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^\gamma} (X)} \subseteq \underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^\gamma} (Y)};$$

$$\underline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)^\gamma} (X)} \subseteq \underline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)^\gamma} (Y)}。$$

$$3) \quad X \subseteq U, Y \subseteq U, \quad \underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^\gamma} (X) \cup \sum_{i=1}^m A_i^{(\zeta, \alpha_i)^\gamma} (Y)} \subseteq$$

$$\underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^\gamma} (X \cup Y)}; \quad \underline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)^\gamma} (X) \cup \sum_{i=1}^m A_i^{(\zeta, \beta_i)^\gamma} (Y)} \subseteq$$

$$\underline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)^\gamma} (X \cup Y)}。$$

$$4) \quad X \subseteq U, Y \subseteq U, \quad \underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^\gamma} (X \cap Y)} \subseteq \underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^\gamma} (X) \cap}$$

$$\underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^\gamma} (Y)}; \quad \underline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)^\gamma} (X \cap Y)} \subseteq \underline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)^\gamma} (X) \cap}$$

$$\underline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)^\gamma} (Y)}。$$

证明 1) 设 $x \in \underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^{\gamma_2}} (X)}$, 由定义 4 可知

$\forall A_i \in T_1, \rho(X|[x]_{A_i}) \geq \alpha_i$, 其中 $T_1 \subseteq A \wedge \sum_{i=1}^{|T_1|} w_{A_i} \geq \gamma_2$, 又因

为 $\gamma_1 < \gamma_2$, 所以 $x \in \underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^{\gamma_1}} (X)}$, 所以

$$\underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^{\gamma_2}} (X)} \subseteq \underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^{\gamma_1}} (X)}, \quad \text{同理可证,}$$

$$\underline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)^{\gamma_2}} (X)} \subseteq \underline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)^{\gamma_1}} (X)}。$$

2) 设 $x \in \underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^\gamma} (X)}$, 由定义 4 可知 $\forall A_i \in T_1$, 由

于 $\rho(X|[x]_{A_i}) \geq \alpha_i$, 其中 $T_1 \subseteq A \wedge \sum_{i=1}^{|T_1|} w_{A_i} \geq \gamma$, 又因为

$X \subseteq Y, \rho(Y|[x]_{A_i}) \geq \alpha_i$, 所以 $x \in \underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^\gamma} (Y)}$, 所以

$$\underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^\gamma} (X)} \subseteq \underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^\gamma} (Y)}, \quad \text{同理可证}$$

$$\underline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)^\gamma} (X)} \subseteq \underline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)^\gamma} (Y)}。$$

3) 设 $\forall x \in \underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^\gamma} (X) \cup \sum_{i=1}^m A_i^{(\zeta, \alpha_i)^\gamma} (Y)}$, 若

$x \in \underline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)^\gamma} (X)}$, 根据定义 4 可知

$\forall A_i \in T_1, \rho(X|[x]_{A_i}) \geq \alpha_i$, 其中 $T_1 \subseteq A \wedge \sum_{i=1}^{|T_1|} w_{A_i} \geq \gamma$, 对于

$\forall Y \subseteq U, \rho(X \cup Y | [x]_{A_i}) \geq \rho(X | [x]_{A_i}) \geq \alpha_i$, 所以
 $x \in \sum_{i=1}^m \overline{A_i^{(\zeta, \alpha_i)^\gamma}(X \cup Y)}$ 。同理, 若 $x \in \sum_{i=1}^m \overline{A_i^{(\zeta, \alpha_i)^\gamma}(Y)}$, 则
 $x \in \sum_{i=1}^m \overline{A_i^{(\zeta, \alpha_i)^\gamma}(X \cup Y)}$, 所以 $\sum_{i=1}^m \overline{A_i^{(\zeta, \alpha_i)^\gamma}(X \cup Y)} \supseteq$
 $\sum_{i=1}^m \overline{A_i^{(\zeta, \alpha_i)^\gamma}(X)} \cup \sum_{i=1}^m \overline{A_i^{(\zeta, \alpha_i)^\gamma}(Y)}$; 同理可证,
 $\sum_{i=1}^m \overline{A_i^{(\zeta, \beta_i)^\gamma}(X \cup Y)} \supseteq \sum_{i=1}^m \overline{A_i^{(\zeta, \beta_i)^\gamma}(X)} \cup \sum_{i=1}^m \overline{A_i^{(\zeta, \beta_i)^\gamma}(Y)}$ 。
 4) 设 $\forall x \in \sum_{i=1}^m \overline{A_i^{(\zeta, \alpha_i)^\gamma}(X \cap Y)}$, 根据定义 4 可知
 $\forall A_i \in T_1, \rho(X \cap Y | [x]_{A_i}) \geq \alpha_i$, 对于 $\forall X \subseteq U$,
 $\rho(X | [x]_{A_i}) \geq \rho(X \cap Y | [x]_{A_i}) \geq \alpha_i$, 对于 $\forall Y \subseteq U$,
 $\rho(Y | [x]_{A_i}) \geq \rho(X \cap Y | [x]_{A_i}) \geq \alpha_i$, 所以 $\forall x \in \sum_{i=1}^m \overline{A_i^{(\zeta, \alpha_i)^\gamma}(X)}$,
 $\forall x \in \sum_{i=1}^m \overline{A_i^{(\zeta, \alpha_i)^\gamma}(Y)}$, 所以 $\sum_{i=1}^m \overline{A_i^{(\zeta, \alpha_i)^\gamma}(X \cap Y)} \subseteq \sum_{i=1}^m \overline{A_i^{(\zeta, \alpha_i)^\gamma}(X)} \cap$
 $\sum_{i=1}^m \overline{A_i^{(\zeta, \alpha_i)^\gamma}(Y)}$; 同理可证 $\sum_{i=1}^m \overline{A_i^{(\zeta, \beta_i)^\gamma}(X \cap Y)} \subseteq \sum_{i=1}^m \overline{A_i^{(\zeta, \beta_i)^\gamma}(X)} \cap$
 $\sum_{i=1}^m \overline{A_i^{(\zeta, \beta_i)^\gamma}(Y)}$ 。

3 平均加权自适应多粒度决策理论粗糙集

在风险评估和决策分析中, 通常通过联合概率来评估复杂事件或系统的风险, 而联合概率由 m 个条件概率的平均值估计。基于这一思想, Li 等^[5]提出了平均 MG-DTRS 模型, 通过给定单参数 ζ 计算属性颗粒结构中不同 (某) 对象对应的概率阈值 (α_i, β_i), 并通过设置属性粒度权重的控制参数对颗粒结构的数量进行调控。这一改进使得平均 MG-DTRS 模型能够更灵活地调整颗粒结构的大小, 从而更符合实际应用需求。在此基础上, 结合式 (14)~(16), 构建 5 种 AWAMG-DTRS 模型。为区分这 5 种模型的上下近似集, 将 First, Second, Third, Fourth 和 Fifth 的缩写 F, Se, Th, Fo 和 Fi 作为上标进行标识。

$$E(\rho(X | [x]_{A_i})) = \frac{\rho(X | [x]_{A_1}) + \rho(X | [x]_{A_2}) + \dots + \rho(X | [x]_{A_m})}{\alpha_1 + \alpha_2 + \dots + \alpha_m} = E_\rho \quad (14)$$

$$\frac{\alpha_1 + \alpha_2 + \dots + \alpha_m}{m} = \bar{\alpha}; \quad \frac{\beta_1 + \beta_2 + \dots + \beta_m}{m} = \bar{\beta} \quad (15)$$

设 $(U, C \cup \{d\})$ 为 DIS, $A = \{A_1, A_2, A_3, \dots, A_m\}$ ($A_i \subseteq C, i = 1, 2, \dots, m$), $w = \{w_{A_1}, w_{A_2}, \dots, w_{A_m}\}$ 为各属性粒度对应的权重, $M = \{A_i | A_i \in A \wedge \rho(X | [x]_{A_i}) \geq \alpha_i\}$,

$T = \left\{ T_l | T_l \subseteq M \wedge w_{T_l} \geq \gamma, w_{T_l} = \sum_{j=1}^{|T_l|} w_{A_j} \right\}$, 则平均属性粒度权重为

$$\bar{w} = \frac{\sum_{l=1}^{|T|} w_{T_l}}{|T|} \quad (16)$$

3.1 第一种 AWAMG-DTRS 模型

定义 5 假设 $(U, C \cup \{d\})$ 为 DIS, $A = \{A_1, A_2, A_3, \dots, A_m\}$ ($A_i \subseteq C, i = 1, 2, \dots, m$), 对于任意 $\zeta \in (0, 1), 0 < \gamma \leq 1, w = \{w_{A_1}, w_{A_2}, \dots, w_{A_m}\}$ 为各属性粒度对应的权重, $M_1 = \{A_i | A_i \in A \wedge \rho(X | [x]_{A_i}) \geq \alpha_i\}$,

$T_1 = \{T_l | T_l \subseteq M_1 \wedge w_{T_l} \geq \gamma, w_{T_l} = \sum_{j=1}^{|T_l|} w_{A_j}\}, \bar{w}_1 = \frac{\sum_{l=1}^{|T_1|} w_{T_l}}{|T_1|};$

$M_2 = \{A_i | A_i \in A \wedge \rho(X | [x]_{A_i}) \geq \beta_i\}$,

$T_2 = \left\{ T_k | T_k \subseteq M_2 \wedge w_{T_k} \geq \gamma, w_{T_k} = \sum_{j=1}^{|T_k|} w_{A_j} \right\}, \bar{w}_2 = \frac{\sum_{k=1}^{|T_2|} w_{T_k}}{|T_2|},$

定义 $X (X \subseteq U)$ 关于 A 的第一种 AWAMG-DTRS 模型下近似和上近似为:

$$\sum_{i=1}^m \overline{A_i^{(\zeta, \alpha_i)F, \gamma}(X)} = \bigcup_{T_l \in T_1 \wedge \sum_{i=1}^{|T_l|} w_{A_i} \geq |T_l| \bar{w}_1} \left\{ x \in U : \bigwedge_{A_i \in T_l} \rho(X | [x]_{A_i}) \geq \alpha_i \right\} \quad (17)$$

$$\sum_{i=1}^m \overline{A_i^{(\zeta, \beta_i)F, \gamma}(X)} = \bigcup_{T_k \in T_2 \wedge \sum_{i=1}^{|T_k|} w_{A_i} \geq |T_k| \bar{w}_2} \left\{ x \in U : \bigwedge_{A_i \in T_k} \rho(X | [x]_{A_i}) > \beta_i \right\} \quad (18)$$

称 $\left\langle \sum_{i=1}^m \overline{A_i^{(\zeta, \alpha_i)F, \gamma}(X)}, \sum_{i=1}^m \overline{A_i^{(\zeta, \beta_i)F, \gamma}(X)} \right\rangle$ 为集合 X 关于 A

的第一种 AWAMG-DTRS 模型的粗糙集。

3.2 第二种 AWAMG-DTRS 模型

定义 6 假设 $(U, C \cup \{d\})$ 为 DIS, $A = \{A_1, A_2, A_3, \dots, A_m\}$ ($A_i \subseteq C, i = 1, 2, \dots, m$), 对于任意 $\zeta \in (0, 1), 0 < \gamma \leq 1, w = \{w_{A_1}, w_{A_2}, \dots, w_{A_m}\}$ 为各属性粒度对应的权重,

$T_1 \subseteq A, T_1 = \left\{ A_i | A_i \in A \wedge \rho(X | [x]_{A_i}) \geq \alpha_i \wedge \sum_{i=1}^{|T_1|} w_{A_i} \geq \gamma \right\},$

$T_2 \subseteq A, T_2 = \left\{ A_i | A_i \in A \wedge \rho(X | [x]_{A_i}) \geq \beta_i \wedge \sum_{i=1}^{|T_2|} w_{A_i} \geq \gamma \right\},$ 定义 $X (X \subseteq U)$ 关于 A 的第二种 AWAMG-DTRS 模型下

近似和上近似为:

$$\frac{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)Se, \gamma}(X)}{\bigcup_{T_1 \subseteq A \wedge \sum_{i=1}^{|T_1|} w_{A_i} \geq \gamma} \left\{ x \in U : \bigwedge_{A_i \in T_1} \rho(X|[x]_{A_i}) \geq \bar{\alpha} \right\}} \quad (19)$$

$$\frac{\sum_{i=1}^m A_i^{(\zeta, \beta_i)Se, \gamma}(X)}{\bigcup_{T_2 \subseteq A \wedge \sum_{i=1}^{|T_2|} w_{A_i} \geq \gamma} \left\{ x \in U : \bigwedge_{A_i \in T_2} \rho(X|[x]_{A_i}) > \bar{\beta} \right\}} \quad (20)$$

其中: $(\alpha_1 + \alpha_2 + \dots + \alpha_m) / m = \bar{\alpha}$, $(\beta_1 + \beta_2 + \dots + \beta_m) / m = \bar{\beta}$, $\bar{\alpha}$ 和 $\bar{\beta}$ 为 2 个概率约束。

称 $\langle \sum_{i=1}^m A_i^{(\zeta, \alpha_i)Se, \gamma}(X), \sum_{i=1}^m A_i^{(\zeta, \beta_i)Se, \gamma}(X) \rangle$ 为集合 X 关于 A

的第二种 AWAMG-DTRS 模型的粗糙集。

3.3 第三种 AWAMG-DTRS 模型

定义 7 假设 $(U, C \cup \{d\})$ 为 DIS, $A = \{A_1, A_2, A_3, \dots, A_m\} (A_i \subseteq C, i = 1, 2, \dots, m)$ 。对于任意 $\zeta \in (0, 1), 0 < \gamma \leq 1, w = \{w_{A_1}, w_{A_2}, \dots, w_{A_m}\}$ 为各属性粒度对应的权重, $T_1 \subseteq A, T_1 = \left\{ A_i | A_i \in A \wedge \rho(X|[x]_{A_i}) \geq \alpha_i \wedge \sum_{i=1}^{|T_1|} w_{A_i} \geq \gamma \right\}$, $T_2 \subseteq A, T_2 = \left\{ A_i | A_i \in A \wedge \rho(X|[x]_{A_i}) \geq \beta_i \wedge \sum_{i=1}^{|T_2|} w_{A_i} \geq \gamma \right\}$, 定义 $X (X \subseteq U)$ 关于 A 的第三种 AWAMG-DTRS 模型下近似和上近似为:

$$\frac{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)Th, \gamma}(X)}{\bigcup_{T_1 \subseteq A \wedge \sum_{i=1}^{|T_1|} w_{A_i} \geq \gamma} \left\{ x \in U : \bigwedge_{A_i \in T_1} E_\rho \geq \alpha_i \right\}} \quad (21)$$

$$\frac{\sum_{i=1}^m A_i^{(\zeta, \beta_i)Th, \gamma}(X)}{\bigcup_{T_2 \subseteq A \wedge \sum_{i=1}^{|T_2|} w_{A_i} \geq \gamma} \left\{ x \in U : \bigwedge_{A_i \in T_2} E_\rho > \beta_i \right\}} \quad (22)$$

其中: $E(\rho(X|[x]_{A_i})) = (\rho(X|[x]_{A_1}) + \rho(X|[x]_{A_2}) + \dots + \rho(X|[x]_{A_i})) / m = E_\rho$ 称 $\langle \sum_{i=1}^m A_i^{(\zeta, \alpha_i)Th, \gamma}(X), \sum_{i=1}^m A_i^{(\zeta, \beta_i)Th, \gamma}(X) \rangle$ 为集合 X 关于 A 第三种 AWAMG-DTRS 模型的粗糙集。

3.4 第四种 AWAMG-DTRS 模型

定义 8 假设 $(U, C \cup \{d\})$ 为 DIS, $A = \{A_1, A_2, A_3, \dots, A_m\} (A_i \subseteq C, i = 1, 2, \dots, m)$ 。对于任意

$\zeta \in (0, 1), 0 < \gamma \leq 1, w = \{w_{A_1}, w_{A_2}, \dots, w_{A_m}\}$ 为各属性粒度对应的权重, $M_1 = \{A_i | A_i \in A \wedge \rho(X|[x]_{A_i}) \geq \alpha_i\}$,

$$T_1 = \{T_l | T_l \subseteq M_1 \wedge w_{T_l} \geq \gamma, w_{T_l} = \sum_{j=1}^{|T_l|} w_{A_j}\}, \bar{w}_1 = \frac{\sum_{l=1}^{|T_1|} w_{T_l}}{|T_1|};$$

$$M_2 = \{A_i | A_i \in A \wedge \rho(X|[x]_{A_i}) \geq \beta_i\},$$

$$T_2 = \left\{ T_k | T_k \subseteq M_2 \wedge w_{T_k} \geq \gamma, w_{T_k} = \sum_{j=1}^{|T_k|} w_{A_j} \right\}, \bar{w}_2 = \frac{\sum_{k=1}^{|T_2|} w_{T_k}}{|T_2|},$$

定义 $X (X \subseteq U)$ 关于 A 的第四种 AWAMG-DTRS 模型下近似和上近似为:

$$\frac{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)Fo, \gamma}(X)}{\bigcup_{T_1 \in T_1 \wedge \sum_{i=1}^{|T_1|} w_{A_i} \geq |T_1| \bar{w}_1} \left\{ x \in U : \bigwedge_{A_i \in T_1} E_\rho \geq \alpha_i \right\}} \quad (23)$$

$$\frac{\sum_{i=1}^m A_i^{(\zeta, \beta_i)Fo, \gamma}(X)}{\bigcup_{T_k \in T_2 \wedge \sum_{i=1}^{|T_k|} w_{A_i} \geq |T_k| \bar{w}_2} \left\{ x \in U : \bigwedge_{A_i \in T_2} E_\rho > \beta_i \right\}} \quad (24)$$

称 $\langle \sum_{i=1}^m A_i^{(\zeta, \alpha_i)Fo, \gamma}(X), \sum_{i=1}^m A_i^{(\zeta, \beta_i)Fo, \gamma}(X) \rangle$ 为集合 X 关于 A

第四种 AWAMG-DTRS 模型的粗糙集。

3.5 第五种 AWAMG-DTRS 模型

定义 9 假设 $(U, C \cup \{d\})$ 为 DIS, $A = \{A_1, A_2, A_3, \dots, A_m\} (A_i \subseteq C, i = 1, 2, \dots, m)$ 。对于任意 $\zeta \in (0, 1), 0 < \gamma \leq 1, w = \{w_{A_1}, w_{A_2}, \dots, w_{A_m}\}$ 为各属性粒度对应的权重, $M_1 = \{A_i | A_i \in A \wedge \rho(X|[x]_{A_i}) \geq \alpha_i\}$,

$$T_1 = \{T_l | T_l \subseteq M_1 \wedge w_{T_l} \geq \gamma, w_{T_l} = \sum_{j=1}^{|T_l|} w_{A_j}\}, \bar{w}_1 = \frac{\sum_{l=1}^{|T_1|} w_{T_l}}{|T_1|};$$

$$M_2 = \{A_i | A_i \in A \wedge \rho(X|[x]_{A_i}) \geq \beta_i\},$$

$$T_2 = \left\{ T_k | T_k \subseteq M_2 \wedge w_{T_k} \geq \gamma, w_{T_k} = \sum_{j=1}^{|T_k|} w_{A_j} \right\}, \bar{w}_2 = \frac{\sum_{k=1}^{|T_2|} w_{T_k}}{|T_2|},$$

定义 $X (X \subseteq U)$ 关于 A 的第五种 AWAMG-DTRS 模型下近似和上近似为:

$$\frac{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)Fi, \gamma}(X)}{\bigcup_{T_l \in T_1 \wedge \sum_{i=1}^{|T_l|} w_{A_i} \geq |T_l| \bar{w}_1} \left\{ x \in U : \bigwedge_{A_i \in T_1} \rho(X|[x]_{A_i}) \geq \bar{\alpha} \right\}} \quad (25)$$

$$\sum_{i=1}^m \overline{A_i^{(\zeta, \beta_i)Fi, \gamma}}(X) = \bigcup_{T_k \in T_2 \wedge \sum_{i=1}^{|T_k|} w_{A_i} \geq |T_k| \bar{w}_2} \left\{ x \in U : \bigwedge_{A_i \in T_2} \rho(X|[x]_{A_i}) > \bar{\beta} \right\} \quad (26)$$

称 $\overline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)Fi, \gamma}}(X), \overline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)Fi, \gamma}}(X)$ 为集合 X 关于 A

第五种 AWAMG-DTRS 模型的粗糙集。

定理 2 假设 $(U, C \cup \{d\})$ 为 DIS, $A = \{A_1, A_2, A_3, \dots, A_m\} (A_i \subseteq C, i = 1, 2, \dots, m)$, 对于任意 $\zeta \in (0, 1), 0 < \gamma \leq 1, w = \{w_{A_1}, w_{A_2}, \dots, w_{A_m}\}$ 为各属性粒度对应的权重, $X \subseteq U$, 可得:

- 1) $\overline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)Fo, \gamma}}(X) \subseteq \overline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)Fi, \gamma}}(X)$;
- 2) $\overline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)Fo, \gamma}}(X) \subseteq \overline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)Fi, \gamma}}(X)$;
- 3) $\overline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)Fi, \gamma}}(X) \subseteq \overline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)F, \gamma}}(X)$;
- 4) $\overline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)Fi, \gamma}}(X) \subseteq \overline{\sum_{i=1}^m A_i^{(\zeta, \beta_i)F, \gamma}}(X)$ 。

证明 1) 设 $\forall x \in \overline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)Fo, \gamma}}(X)$, 由定义 8 知,

$$\forall A_i \in T_1, T_1 = \{T_l | T_l \subseteq M_1 \wedge w_{T_l} \geq \gamma, w_{T_l} = \sum_{j=1}^{|T_l|} w_{A_j}\},$$

$$M_1 = \{A_i | A_i \in A \wedge \rho(X|[x]_{A_i}) \geq \alpha_i\}, E_\rho \geq \alpha_i, \text{ 即}$$

$$\rho(X|[x]_{A_i}) \geq \alpha_i \text{ 且 } E_\rho \geq \alpha_i. \text{ 由定义 5 知,}$$

$$\forall A_i \in T_1, T_1 = \{T_l | T_l \subseteq M_1 \wedge w_{T_l} \geq \gamma, w_{T_l} = \sum_{j=1}^{|T_l|} w_{A_j}\},$$

$$M_1 = \{A_i | A_i \in A \wedge \rho(X|[x]_{A_i}) \geq \alpha_i\}, \rho(X|[x]_{A_i}) \geq \alpha_i, \text{ 即}$$

$$\rho(X|[x]_{A_i}) \geq \alpha_i, \text{ 所以 } x \in \overline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)Fi, \gamma}}(X), \text{ 所以}$$

$$\overline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)Fo, \gamma}}(X) \subseteq \overline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)Fi, \gamma}}(X).$$

2) 同理 1) 可证。

3) 设 $\forall x \in \overline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)Fi, \gamma}}(X)$, 由定义 9 知,

$$\forall A_i \in T_1, T_1 = \{T_l | T_l \subseteq M_1 \wedge w_{T_l} \geq \gamma, w_{T_l} = \sum_{j=1}^{|T_l|} w_{A_j}\},$$

$$M_1 = \{A_i | A_i \in A \wedge \rho(X|[x]_{A_i}) \geq \alpha_i\}, \rho(X|[x]_{A_i}) \geq \bar{\alpha}, \text{ 即}$$

$$\rho(X|[x]_{A_i}) \geq \alpha_i \text{ 且 } \rho(X|[x]_{A_i}) \geq \bar{\alpha}. \text{ 由定义 5 知,}$$

$$\forall A_i \in T_1, T_1 = \{T_l | T_l \subseteq M_1 \wedge w_{T_l} \geq \gamma, w_{T_l} = \sum_{j=1}^{|T_l|} w_{A_j}\},$$

$$M_1 = \{A_i | A_i \in A \wedge \rho(X|[x]_{A_i}) \geq \alpha_i\}, \rho(X|[x]_{A_i}) \geq \alpha_i, \text{ 即}$$

$\rho(X|[x]_{A_i}) \geq \alpha_i$, 所以 $x \in \overline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)Fi, \gamma}}(X)$, 所以

$$\overline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)Fi, \gamma}}(X) \subseteq \overline{\sum_{i=1}^m A_i^{(\zeta, \alpha_i)F, \gamma}}(X).$$

4) 同理 3) 可证。

4 实例分析与实验验证

为验证所提出模型在实际应用中的可行性和适应性,通过 1 个实例进行分析,并选取 UCI 机器学习数据集进行仿真实验,以评估模型在实际场景中的表现及应用效果。

4.1 实例分析

给定 1 个汽车信息系统 $(U, C \cup \{d\})$, 其中对象 $U = \{x_0, x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}$, 条件属性 $C = \{A_1, A_2, A_3, A_4\}$, A_1, A_2, A_3, A_4 分别代表汽车的价格、尺寸、能源属性和安全性。价格包括 Low 和 High, 尺寸包括 Compact 和 Full, 能源属性包括 Gasoline 和 Diesel, 安全性分为 Low, Medium 和 High。 d 为决策属性, 用于判断汽车性能, 分为 Poor, Good, Excellent 3 种等级。信息系统的信息如表 1。

表 1 信息系统

Tab. 1 Information systems

U	A_1	A_2	A_3	A_4	d
x_0	Low	Compact	Gasoline	Low	Poor
x_1	High	Full	Gasoline	High	Excellent
x_2	High	Full	Diesel	Medium	Good
x_3	High	Full	Diesel	Medium	Poor
x_4	Low	Full	Gasoline	High	Good
x_5	High	Full	Diesel	High	Excellent
x_6	High	Compact	Diesel	Low	Poor
x_7	Low	Full	Gasoline	Medium	Poor
x_8	High	Compact	Gasoline	High	Good
x_9	Low	Compact	Gasoline	High	Poor

根据式 (11) 计算表 1 所示信息系统各属性粒度的权重, 分别为 $w_{A_1} = 0.239, w_{A_2} = 0.239, w_{A_3} = 0.011, w_{A_4} = 0.511$ 。取 $\zeta = 0.3$, 根据式 (7)~(10), 利用单参数获取表 1 所示信息系统不同属性颗粒结构中不同对象对应的概率阈值。设定 $\gamma = 0, 0.3, 0.5$, 采用 WAMG, FAWAMG, SAWAMG, TAWAMG, FoAWAMG, FiAWAMG 模型, 通过对 γ 参数的设定控制颗粒结构数目。WAMG, FAWAMG, SAWAMG, TAWAMG, FoAWAMG, FiAWAMG 分别对应 WAMG-DTRS,

第一种 AWAMG-DTRS, 第二种 AWAMG-DTRS, 第三种 AWAMG-DTRS, 第四种 AWAMG-DTRS, 第五种 AWAMG-DTRS 模型。表 2~4 分别为权重系数 $\gamma=0, 0.3, 0.5$ 时 6 种模型的决策区域, 其中 POS 代表正域, BND 代表边界域, NEG 代表负域。

表 2 $\gamma=0$ 时 6 种模型的决策区域

Tab. 2 Decision regions of the six models when $\gamma=0$

模型	POS	BND	NEG
WAMG	$x_0, x_2, x_3, x_4, x_6, x_7, x_8, x_9$	x_1, x_5	\emptyset
FAWAMG	x_0, x_2, x_3, x_6, x_7	x_1, x_4, x_5, x_8, x_9	\emptyset
SAWAMG	$x_0, x_4, x_6, x_7, x_8, x_9$	x_1, x_2, x_3, x_5	\emptyset
TAWAMG	x_0, x_6	$x_1, x_2, x_3, x_4, x_5, x_7, x_8, x_9$	\emptyset
FoAWAMG	x_0, x_6	$x_1, x_2, x_3, x_4, x_5, x_7, x_8, x_9$	\emptyset
FiAWAMG	x_0, x_6	$x_1, x_2, x_3, x_4, x_5, x_7, x_8, x_9$	\emptyset

表 3 $\gamma=0.3$ 时 6 种模型的决策区域

Tab. 3 Decision regions of the six models when $\gamma=0.3$

模型	POS	BND	NEG
WAMG	$x_0, x_2, x_3, x_6, x_7, x_9$	x_1, x_4, x_5, x_8	\emptyset
FAWAMG	x_0, x_2, x_3, x_6, x_7	x_1, x_4, x_5, x_8, x_9	\emptyset
SAWAMG	x_0, x_6, x_9	$x_1, x_2, x_3, x_4, x_5, x_7, x_8$	\emptyset
TAWAMG	x_0, x_6	$x_1, x_2, x_3, x_4, x_5, x_7, x_8, x_9$	\emptyset
FoAWAMG	x_0, x_6	$x_1, x_2, x_3, x_4, x_5, x_7, x_8, x_9$	\emptyset
FiAWAMG	x_0, x_6	$x_1, x_2, x_3, x_4, x_5, x_7, x_8, x_9$	\emptyset

表 4 $\gamma=0.5$ 时 6 种模型的决策区域

Tab. 4 Decision regions of the six models when $\gamma=0.5$

模型	POS	BND	NEG
WAMG	x_0, x_2, x_3, x_6, x_7	x_1, x_4, x_5, x_8, x_9	\emptyset
FAWAMG	x_0, x_2, x_3, x_6, x_7	x_1, x_4, x_5, x_8, x_9	\emptyset
SAWAMG	x_0, x_6	$x_1, x_2, x_3, x_4, x_5, x_7, x_8, x_9$	\emptyset
TAWAMG	x_0, x_6	$x_1, x_2, x_3, x_4, x_5, x_7, x_8, x_9$	\emptyset
FoAWAMG	x_0, x_6	$x_1, x_2, x_3, x_4, x_5, x_7, x_8, x_9$	\emptyset
FiAWAMG	x_0, x_6	$x_1, x_2, x_3, x_4, x_5, x_7, x_8, x_9$	\emptyset

由表 2~4 可知: 6 种模型中, 权重的控制系数与下近似集之间的关系紧密。对于 WAMG-DTRS 模型, $\gamma=0$ 时, 下近似集包含的对象数为 8 个; $\gamma=0.3$ 时, 下近似集包含的对象减至 6 个; 而 $\gamma=0.5$ 时, 下近似集包含的对象进一步减至 5 个, 表明随 γ 值的增大, WAMG-DTRS 模型下近似集呈现减小的趋势。对于 FAWAMG, SAWAMG, TAWAMG, FoAWAMG, FiAWAMG 模型, $\gamma=0$ 时, 下近似集包含的对象分别为 5, 6, 2, 2, 2 个; $\gamma=0.3$ 时, 下近似集包含的对象分

别为 5, 3, 2, 2, 2 个; 而 $\gamma=0.5$ 时, 下近似集包含的对象分别为 5, 2, 2, 2, 2 个。由此表明: 相同 γ 值、不同平均条件下 AWAMG-DTRS 模型的下近似集大小会不同; 且随 γ 值的增大, AWAMG-DTRS 模型的下近似集逐渐减小。

综上所述可知, 通过调整 γ 值的大小能够有效控制颗粒结构的数量, 从而实现下近似集的调控, 验证了本文所提出模型的可行性。下近似集大小在粗糙集理论中直接关系到模型的性能、计算效率和可解释性, 因此实际应用中可通过调整 γ 的大小使下近似集能够在理论和实践中达到平衡, 从而提高模型的实用性和可操作性。

4.2 实验验证

UCI 数据集能够提供标准化的测试环境, 有效评估模型的性能和适应性。为进一步验证 WAMG-DTRS 模型的可行性和可靠性, 选择 UCI 数据集为实验对象进行仿真实验, UCI 数据集的具体信息如表 5。

表 5 UCI 数据集

Tab. 5 UCI datasets

数据集	样本/个	条件属性/个	决策类别/个
Zoo	101	17	7
Wine-quality-red	1 599	11	6
Wine-quality-white	4 898	11	7
Diabetes	520	16	2

为提升模型的效率和精度, 实验分析前需要对 UCI 数据集进行预处理。删除权重小于 0.005 及重要性较小的属性粒度, 从而聚焦对模型性能有实质性影响的属性粒度。预处理的数据集信息如表 6。

表 6 UCI 数据集的预处理结果

Tab. 6 Preprocessing results of UCI dataset

数据集	样本/个	属性/个		决策类别/个
		删除	条件	
Zoo	101	1	16	7
Wine-quality-red	1 599	0	11	6
Wine-quality-white	4 898	0	11	7
Diabetes	520	3	13	2

基于表 6 中的数据集, 取 $\zeta=0.3$, 根据式 (7)~(10), 利用单参数获取不同样本对象在不同属性颗粒结构下对应的概率阈值 (α, β) 。以 Diabetes 数据集为例, 该数据集包含 2 个决策类别 Negative 和 Positive; 条件属性包括 Age, Gender, Polyuria 等 13 个。表 7 为 Diabetes 数据集在不同条件属性中各对象对应的概率阈值。

表 7 Diabetes 数据集的概率阈值

Tab. 7 Probability threshold of the Diabetes dataset

U	Negative			Positive		
	Age	Gender	...	Age	Gender	...
x_0	(0.69,0.27)	(0.60,0.19)	...	(0.73,0.31)	(0.81,0.40)	...
x_1	(0.71,0.29)	(0.60,0.19)	...	(0.71,0.29)	(0.81,0.40)	...
x_2	(1,0.70)	(0.60,0.19)	...	(0.3,0)	(0.81,0.40)	...
x_3	(0.67,0.25)	(0.60,0.19)	...	(0.75,0.33)	(0.81,0.40)	...
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots

基于表 6 中的数据集,以 Diabetes 数据集为例,取 $\gamma = 0.5, \zeta = 0.3$, 分别计算乐观 MG-DTRS、悲观 MG-DTRS、WAMG-DTRS 以及 5 种 AWAMG-DTRS 模型决策属性中各决策类别下的下近似集数量,结果见表 8。

表 8 Diabetes 数据集下各模型的下近似集

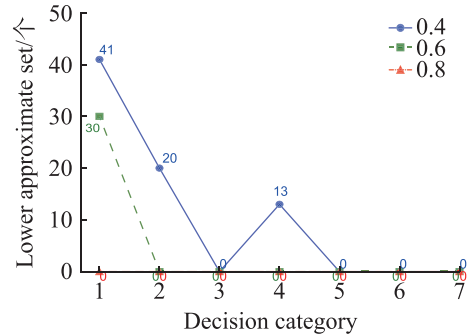
Tab. 8 The lower approximation set of each model under the Diabetes dataset

模型	Positive/个	Negative/个
乐观 MG-DTRS ($\alpha = 0.7, \beta = 0.3$)	507	284
悲观 MG-DTRS ($\alpha = 0.7, \beta = 0.3$)	0	0
WAMG-DTRS	226	83
第一种 AWAMG-DTRS	507	284
第二种 AWAMG-DTRS	223	0
第三种 AWAMG-DTRS	201	0
第四种 AWAMG-DTRS	226	83
第五种 AWAMG-DTRS	223	0

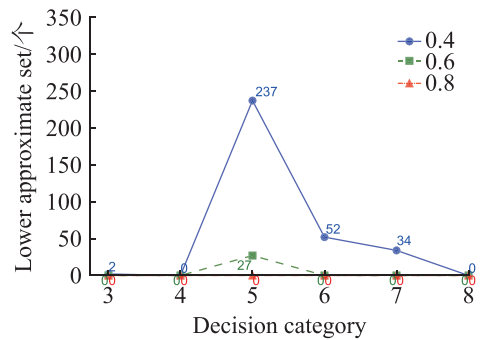
根据表 8 可知:当 $\gamma = 0.5$ 时, WAMG-DTRS 模型、乐观 MG-DTRS 和悲观 MG-DTRS 模型在决策类别 Positive 的下近似集包含的对象分别为 226, 507, 0 个, 在决策类别 Negative 的下近似集包含的对象分别为 83, 284, 0 个。可以看出, WAMG-DTRS 模型的下近似集规模介于乐观和悲观模型之间,可有效克服乐观 MG-DTRS 模型过于宽松和悲观 MG-DTRS 模型过于严谨的问题。 $\gamma = 0.5$ 时, 5 种 AWAMG-DTRS 模型在决策类别 Positive 的下近似集包含的对象分别为 507, 223, 201, 226, 223 个, 而在决策类别 Negative 的下近似集包含的对象分别为 284, 0, 0, 83, 0 个。这表明,不同平均条件下, AWAMG-DTRS 模型的下近似集规模和构成不同,下近似集的特性受平均条件变化的影响。

WAMG-DTRS 模型是 5 种 AWAMG-DTRS 模型的基础,具有更强的适用性和灵活性。为进一步验证所提模型的相关性质,特别是权重系数 γ 对模型上、下近似集的影响,选取 WAMG-DTRS 模型进一步进行仿真实验,实验数据来源于表 6 中的 4 个数据集。

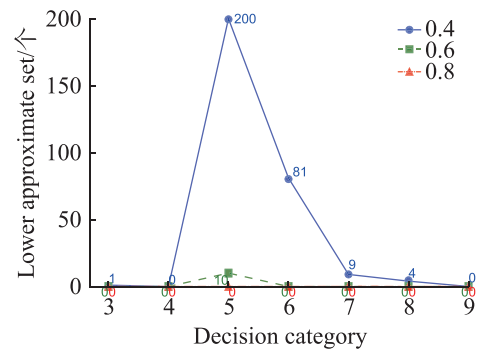
针对每个数据集,分别设置 3 组不同的权重系数 $\gamma(0.4, 0.6, 0.8)$, 仿真分析权重系数对模型输出结果的影响,结果如图 1, 2。



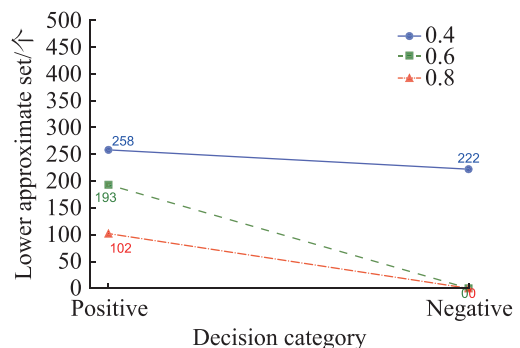
(a) Zoo



(b) Wine-quality-red



(c) Wine-quality-white



(d) Diabetes

图 1 不同数据集与权重控制系数下 WAMG-DTRS 模型的下近似集规模

Fig. 1 The scale of lower approximation sets of the WAMG-DTRS model under different datasets and weight control coefficients

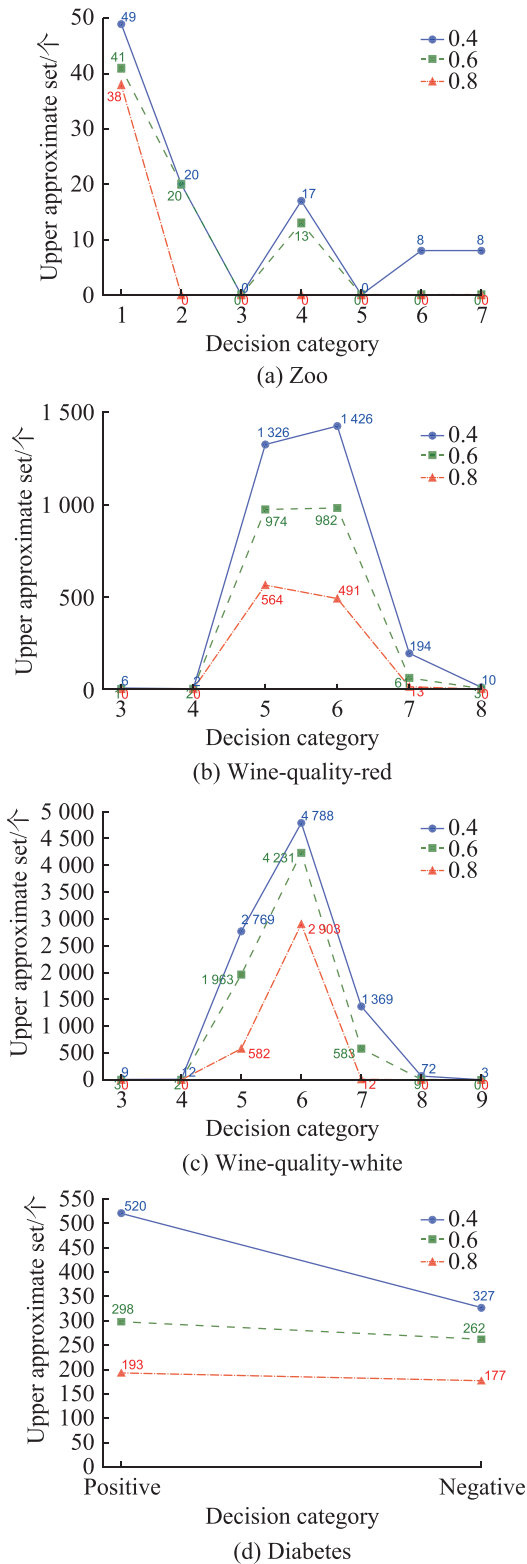


图2 不同数据集与权重控制系数下 WAMG-DTRS 模型的上近似集规模

Fig. 2 The scale of upper approximation set of the WAMG-DTRS model under different datasets and weight control coefficients

图中纵轴表示下近似集、上近似集规模, 横轴表示决策类别。由图 1, 2 可知: 权重系数 γ 从 0.4 增

至 0.8, WAMG-DTRS 模型在不同数据集及决策类别下的下近似集、上近似集呈现逐渐减小的趋势。这表明, 通过精细调控权重系数 γ , 可有效实现对粗糙集颗粒结构的动态调整。这一结果进一步验证了 WAMG-DTRS 模型在颗粒结构调控方面的可行性和灵活性; 同时也表明通过精细调控权重系数, 能使模型更好地适应不同数据集特性, 从而在模型复杂度与性能之间找到 1 个理想的平衡点。这种灵活性显著提升了 WAMG-DTRS 模型在实际应用场景中的适用性和有效性, 为复杂决策问题提供了更为精准和高效的解决方案。

5 结论

针对 MG-DTRS 模型在实际应用中适应性不足的问题, 提出 WAMG-DTRS 模型。该模型利用单参数自适应方法获取属性颗粒结构下不同对象对应的上、下近似集计算所需的概率阈值, 并通过信息增益方法精确计算各属性粒度的权重, 从而确保模型能够有效反映数据的内在特征, 提高模型的准确性。为了进一步增强 WAMG-DTRS 模型对不同数据集特点和复杂性的适应性, 引入权重系数调节属性粒度权重, 控制粒度结构的数量。此外, 为保证 WAMG-DTRS 模型的多样化和适应性, 进一步构建了 5 种 AWAMG-DTRS 模型。通过实例并结合 UCI 中数据集进行仿真验证, 结果表明: 通过调控 WAMG-DTRS 模型的权重系数, 能够有效解决乐观 MG-DTRS 模型过于宽松和悲观 MG-DTRS 模型过于严谨的问题; 随 WAMG-DTRS 模型权重系数的增大, 模型的上、下近似集逐渐减小; 在相同权重系数下, 构建的 5 种 AWAMG-DTRS 模型的下近似集呈现差异性, 且随权重系数的增大, AWAMG-DTRS 模型的下近似集规模均逐渐减小。这些结果表明, 通过精细调控权重系数, 能够显著提升模型对不同数据集特点的适应能力, 增强模型在实际场景中的适用性和可行性。

参考文献:

[1] PAWLAK Z. Rough sets[J]. International Journal of Computer & Information Sciences, 1982, 11(5):341-356.
 [2] MEHER TAJ S, SUDHA M, KUMARAVEL A. Predicting heart failure using data mining with Rough set theory and Fuzzy Petri Net[J]. Journal of Physics: Conference Series, 2021, 1724(1):012033.
 [3] ZHOU J Y, GUO Y, SUN Y, et al. Data mining method based on rough set and fuzzy neural network[J]. Journal of

- Intelligent & Fuzzy Systems, 2020, 38(4):3717–3725.
- [4] SASIKUMAR C S, KUMARAVEL A. E-learning attributes selection through rough set theory and data mining[J]. International Journal of Innovative Technology and Exploring Engineering, 2019, 8(10):3920–3924.
- [5] LI Q B, WEI Y, LI W J. Method for fine pattern recognition of space targets using the entropy weight fuzzy-rough nearest neighbor algorithm[J]. Journal of Applied Spectroscopy, 2021, 87(6):1018–1022.
- [6] ZHANG P F, LI T R, WANG G Q, et al. Multi-source information fusion based on rough set theory: a review[J]. Information Fusion, 2021, 68:85–117.
- [7] YAO Y Y. Decision-theoretic rough set models[C]// Rough Sets and Knowledge Technology. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007:1–12.
- [8] ZHAO X R, HU B Q. Three-way decisions with decision-theoretic rough sets in multiset-valued information tables[J]. Information Sciences, 2020, 507:684–699.
- [9] SUO M L, TAO L F, ZHU B L, et al. Single-parameter decision-theoretic rough set[J]. Information Sciences, 2020, 539:49–80.
- [10] QIAN Y H, ZHANG H, SANG Y L, et al. Multigranulation decision-theoretic rough sets[J]. International Journal of Approximate Reasoning, 2014, 55(1):225–237.
- [11] ZHANG C, LI D Y, LIANG J Y. Multi-granularity three-way decisions with adjustable hesitant fuzzy linguistic multigranulation decision-theoretic rough sets over two universes[J]. Information Sciences, 2020, 507:665–683.
- [12] LI M M, ZHANG C P, CHEN M H, et al. Multigranulation double-quantitative decision-theoretic rough sets based on logical operations[J]. International Journal of Machine Learning and Cybernetics, 2022, 13(6):1661–1684.
- [13] QIAN J, LIU C H, MIAO D Q, et al. Sequential three-way decisions via multi-granularity[J]. Information Sciences, 2020, 507:606–629.
- [14] ZHANG P F, LI T R, LUO C, et al. AMG-DTRS: adaptive multi-granulation decision-theoretic rough sets[J]. International Journal of Approximate Reasoning, 2022, 140:7–30.
- [15] 张明, 程科, 杨习贝, 等. 基于加权粒度的多粒度粗糙集[J]. 控制与决策, 2015, 30(2):222–228.
- ZHANG M, CHENG K, YANG X B, et al. Multigranulation rough set based on weighted granulations[J]. Control and Decision, 2015, 30(2):222–228.
- [16] GUO Y T, TSANG E C C, XU W H, et al. Adaptive weighted generalized multi-granulation interval-valued decision-theoretic rough sets[J]. Knowledge-Based Systems, 2020, 187:104804.
- [17] PANG J F, GUAN X Q, LIANG J Y, et al. Multi-attribute group decision-making method based on multi-granulation weights and three-way decisions[J]. International Journal of Approximate Reasoning, 2020, 117:122–147.
- [18] LIU F T, WANG J L. An accurate method of determining attribute weights in distance-based classification algorithms[J]. Mathematical Problems in Engineering, 2022, 2022:6936335.
- [19] 罗康, 彭长根, 谭伟杰, 等. 基于属性权重的隐私泄露风险评估方法[J]. 计算机技术与发展, 2022, 32(10):82–87.
- LUO K, PENG C G, TAN W J, et al. Privacy disclosure risk assessment method based on attribute weight[J]. Computer Technology and Development, 2022, 32(10):82–87.
- [20] QU K L, XU J C, HOU Q C, et al. Feature selection using Information Gain and decision information in neighborhood decision system[J]. Applied Soft Computing, 2023, 136:110100.
- [21] 李佳旭, 蔡梦思, 谭索怡, 等. 基于引文大数据的高阶网络建模及信息增益比较研究[J]. 系统科学与数学, 2021, 41(10):763–2775.
- LI J X, CAI M S, TAN S Y, et al. A comparison study of higher-order network modeling and information gain based on big citation data[J]. Journal of Systems Science and Mathematical Sciences, 2021, 41(10):763–2775.

责任编辑:丁吉海