



基于广义自适应多粒度的多源信息融合研究

钱进, 童志钢, 余鹰, 洪承鑫, 苗夺谦

引用本文:

钱进,童志钢,余鹰,洪承鑫,苗夺谦. 基于广义自适应多粒度的多源信息融合研究[J]. *智能系统学报*, 2023, 18(1): 173–185.

QIAN Jin,TONG Zhigang,YU Ying,HONG Chengxin,MIAO Duoqian. Multi-source information fusion through generalized adaptive multi-granulation[J]. *CAAI Transactions on Intelligent Systems*, 2023, 18(1): 173–185.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202208030>

您可能感兴趣的其他文章

广义优势多粒度直觉模糊粗糙集及规则获取

Generalized dominance-based multi-granularity intuitionistic fuzzy rough set and acquisition of decision rules
智能系统学报. 2017, 12(6): 883–888 <https://dx.doi.org/10.11992/tis.201706034>

基于证据理论刻画多粒度覆盖粗糙集的数值属性

Evidence-theory-based numerical characterization of multi-granulation covering rough sets
智能系统学报. 2016, 11(4): 481–486 <https://dx.doi.org/10.11992/tis.201606011>

元素最小描述并集下的多粒度覆盖粗糙集模型

Multigranulation covering rough sets based on the union of minimal descriptions of elements
智能系统学报. 2016, 11(4): 534–538 <https://dx.doi.org/10.11992/tis.201605034>

从人类智能到机器实现模型——粒计算理论与方法

From human intelligence to machine implementation model: theories and applications based on granular computing
智能系统学报. 2016, 11(6): 743–757 <https://dx.doi.org/10.11992/tis.201612014>

采用粒计算的属性权重确定方法

A method for ascertaining the weight of attributes based on granular computing
智能系统学报. 2015(2): 273–280 <https://dx.doi.org/10.3969/j.issn.1673-4785.201312008>

DOI: 10.11992/tis.202208030

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20221104.1628.002.html>

基于广义自适应多粒度的多源信息融合研究

钱进¹, 童志钢¹, 余鹰¹, 洪承鑫¹, 苗夺谦^{1,2}

(1. 华东交通大学 软件学院, 江西 南昌 330013; 2. 同济大学 电子与信息工程学院, 上海 201804)

摘要: 多粒度粗糙集模型是一种有效的信息融合策略。利用该策略能从多个角度将多源信息进行融合, 并转化成一致的信息表示。现有的大多数多粒度信息融合方法对每个知识粒度都采用相同的阈值, 然而, 众所周知, 不同的信息源的来源和噪声都不尽相同, 其对应的知识粒度的阈值也应不同。为此, 首先在广义多粒度粗糙集中引入单参数决策理论粗糙集, 提出了广义自适应多粒度粗糙集模型。然后, 利用经典的融合策略设计了 4 种广义多粒度模型, 所有模型都可以通过一个参数补偿系数 ζ 来自适应地获得知识粒度对应的阈值对, 并讨论了这些模型的相关性质。最后, 通过实验结果证明, 所提模型在实际应用中灵活性更高, 决策更为合理。

关键词: 多源信息; 信息融合; 决策理论粗糙集; 粗糙集; 广义自适应多粒度; 多粒度; 自适应阈值; 知识粒度
中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-4785(2023)01-0173-13

中文引用格式: 钱进, 童志钢, 余鹰, 等. 基于广义自适应多粒度的多源信息融合研究 [J]. 智能系统学报, 2023, 18(1): 173-185.

英文引用格式: QIAN Jin, TONG Zhigang, YU Ying, et al. Multi-source information fusion through generalized adaptive multi-granulation[J]. CAAI transactions on intelligent systems, 2023, 18(1): 173-185.

Multi-source information fusion through generalized adaptive multi-granulation

QIAN Jin¹, TONG Zhigang¹, YU Ying¹, HONG Chengxin¹, MIAO Duoqian^{1,2}

(1. College of Software Engineering, East China Jiaotong University, Nanchang 330013, China; 2. College of Electronic and Information Engineering, Tongji University, Shanghai 201804, China)

Abstract: The multi-granulation rough set model is an effective information fusion strategy. In this paper, this strategy is used to fuse multi-source information from multiple perspectives, and then this information is translated into a consistent information representation. However, most existing multi-granulation information fusion methods use the same threshold value for each knowledge granularity. As we all know, the origin and noise differ among information sources, and the threshold values of the corresponding knowledge granularity should differ. To this end, in this paper, a generalized adaptive multi-granulation rough set model is proposed by combining a single-parameter decision-theoretic rough set with a generalized multi-granulation rough set. Then, four types of generalized multi-granulation models are designed based on typical fusion strategies so that all models can obtain threshold pairs corresponding to knowledge granularity by setting a compensation coefficient ζ . Furthermore, the relevant properties of these models are discussed. Finally, the experimental results demonstrate that the proposed model is more flexible and reasonable in practical applications.

Keywords: multi-source information; information fusion; decision-theoretic rough set; rough set; generalized adaptive multi-granulation; multi-granulation; adaptive thresholds; knowledge granularity

近些年, 迅猛发展的信息技术已经被广泛应用于企业和个人生活中, 特别是互联网、云计算等技术的发展, 促进了大数据产业的稳定增长。

由于传感器和数据收集技术的发展, 导致数据呈现多源化, 如何挖掘多源数据中有价值的信息变得尤为重要。

根据信息的特征, 可以把信息分为同构和异构两种数据类型。对于同构多源系统, 将多源信息系统集成到单个信息表中是最常见的融合策略。

收稿日期: 2022-08-22. 网络出版日期: 2022-11-07.

基金项目: 国家自然科学基金项目(62066014, 62163016); 江西省自然科学基金项目(20202BABL202018, 20212ACB202001).

通信作者: 钱进. E-mail: qjqilqyf@163.com.

Yu 等^[1]针对模糊不完全信息系统提出了基于信息熵的多信息源融合为单信息源的方法; Xu 等^[2]通过将每个对象的原始信息转化为三角模糊信息粒, 将多个信息源融合为单个源; Sang 等^[3]基于决策过程不确定性提出了包括 3 种多源决策方法, 其中包括基于信息熵和均值法融合为单源信息系统的两种方法; Zhang 等^[4]提出了多源同构数据的数据级融合模型, 并采用无监督的属性选择方法来消除单源信息系统的冗余属性; Zhang 等^[5]针对多源不完备区间值数据, 提出了基于信息熵的信息融合方法以及 4 种以信息源和属性变化为特征的增量融合机制。

多粒度粗糙集理论为信息融合提供了一条新的途径。众所周知, 人们解决问题时总是从不同的角度进行考虑, 借鉴此思想, Qian 等^[6]提出了多粒度粗糙集。此后引起大量学者对该模型进行了扩展研究。例如, Yang 等^[7]从代价敏感的角度研究了多粒度粗糙集; Liu 等^[8]提出了多粒度覆盖粗糙集模型; Qian 等^[9]将多粒度思想与贝叶斯决策理论相结合, 提出了多粒度决策理论粗糙集模型; Sun 等^[10]从单论域扩充到双论域, 提出双论域多粒度粗糙集; Liang 等^[11]提出了广义优势关系多粒度直觉模糊粗糙集模型及规则获取方法; Yu 等^[12]在多粒度近似空间中提出了一种双量化决策理论; Qian 等^[13]提出了一个基于多粒度的序贯三支决策模型; Li 等^[14]从粒计算的角度对多粒度数据分析方法进行综述; Zhang 等^[15]提出了自适应多粒度决策理论粗糙集模型。

在信息融合领域, 多粒度粗糙集模型利用一系列源产生的多粒度结构来实现多源信息融合。在这方面, 已有许多成果呈现。例如, 林国平等^[16]分析了多粒度粗糙集理论和证据理论之间的联系, 并提出了基于多粒度视角下的 D-S 证据理论融合策略; Che 等^[17]从多粒度的角度研究了多源信息系统的信息融合及相关数字特征; Sang 等^[18]提出了多源信息系统中的广义多粒度双量化决策理论粗糙集模型; Che 等^[19]基于多源粗糙近似和证据理论之间的关系, 提出了多源粗糙近似下的属性集约简; Wang 等^[20]提出了基于多粒度的多源数据知识获取方法; Yang 等^[21]通过构造支持聚合算子和可能聚合算子, 从多粒度的视角提出了多源决策信息融合方法; Luo 等^[22]提出了多源覆盖信息系统下的加权广义多粒度粗糙集模型及其应用; Chu 等^[23]针对混合信息系统, 提出了多粒度优势粗糙概念及属性约简并应用于实际的中西医临床诊疗决策过程。

根据前文的描述, 现有的多源信息融合方法

主要分为两种类型。这两种方法都存在一定的局限性, 第 1 种将多源信息表转化为到单信息表的方法, 在融合的过程中会导致大量信息丢失; 而第 2 种利用同一阈值的多粒度粗糙集模型进行信息融合的方法没有考虑多源信息的复杂性和多样性。在多源信息系统中, 每个子信息系统的采集标准通常不相同, 其含有的噪音数据也各不相同。因此, 采用相同阈值进行融合多源信息并不合理。

本文的研究动机是获取不同信息源导出的知识粒度对应的阈值对。为此, 本文将单参数决策理论粗糙集融入多粒度中, 提出了一种广义自适应多粒度粗糙集模型。该模型利用单信息源导出的不可分辨关系构造知识粒度, 并通过设置一个补偿系数 ζ 来自适应获得不同知识粒度的概率阈值对。之后设计了悲观-悲观自适应多粒度粗糙集、乐观-乐观自适应多粒度粗糙集、悲观-乐观自适应多粒度粗糙集和乐观-悲观自适应多粒粗糙集 4 种经典多粒度粗糙集模型, 并讨论了该设计模型的相关性质。最后, 在 6 个数据集上进行实验对比, 验证了本模型能够较好实现多源信息融合。

1 基础知识

本节介绍相关概念^[16,24-26]。

1.1 信息系统和 Pawlak 粗糙集

定义 1^[24] 信息系统(I)可用四元组 $(U, A, \{V_a | a \in A\}, \{M_a | a \in A\})$ 表示, 其中 U 是包含对象 x_1, x_2, \dots, x_n 的非空集合, 也称为论域; A 为描述对象的非空有限属性集; V_a 表示属性 a 的取值, 属性集 A 的取值集合可用 $V = \cup_{a \in A} V_a$ 表示; M 是映射函数, $M_a(x) = V_a$ 表示对象 x 在属性 a 的取值为 V_a 。信息系统可以简写为 (U, A, V, M) 。决策信息系统(H)是特殊的信息系统, 其属性集 A 由条件属性集 C 和决策属性集 D 组成, 因此也可以表达为 $(U, C \cup D, V, M)$ 。

定义 2^[24] 设 $I = (U, A, V, M)$, 对于 $\forall E \subseteq A$, R_E 被称为 U 上的等价关系, 定义为

$$R_E = \{(x, y) \in U^2 : \forall a \in E, M_a(x) = M_a(y)\}$$

进而得到目标概念 $X(X \subseteq U)$ 关于 R_E 的下近似和上近似分别为

$$\underline{R}(X) = \{x \in U : [x]_E \subseteq X\}$$

$$\overline{R}(X) = \{x \in U : [x]_E \cap X \neq \emptyset\}$$

通过 R_E 可将论域 U 划分为多个块, $[x]_E$ 则指含有对象 x 的划分块, 也称为 x 的等价类。

1.2 决策理论粗糙集

在概率粗糙集的基础上, Yao 等^[25]通过引入贝叶斯风险决策理论提出了决策理论粗糙集(DTRS)。

定义 3 设 $H = (U, C \cup D, V, M)$, 对于任何一个属性子集 $E \subseteq C$, 参数满足 $0 \leq \beta \leq \alpha \leq 1$, 关于目标概念 $X \subseteq U$ 的下近似与上近似分别定义为

$$\frac{a_{pr}(\alpha, \beta)}{(X)} = \{x \in U \mid P(X|[x]_E) \geq \alpha\}$$

$$\frac{\bar{a}_{pr}(\alpha, \beta)}{(X)} = \{x \in U \mid P(X|[x]_E) > \beta\}$$

式中: $[x]_E$ 为 E 下的等价类, $P(X|[x]_E)$ 代表目标概念 X 在 $[x]_E$ 的条件概率, 表示为 $P(X|[x]_E) = \frac{|[x]_E \cap X|}{|[x]_E|}$ 。

进而可以得到 $X \subseteq U$ 关于 E 的正域、边界域和负域, 定义为

$$P_{OS}(X) = \{x \in U \mid P(X|[x]_E) \geq \alpha\}$$

$$B_{ND}(X) = \{x \in U \mid \beta < P(X|[x]_E) < \alpha\}$$

$$N_{EG}(X) = \{x \in U \mid P(X|[x]_E) \leq \beta\}$$

在 DTRS 中, 阈值参数是根据最小风险决策准则来获得的。在二分类问题中, 假设对象集用 $U = \{x_1, x_2, \dots, x_n\}$ 来表示; 状态集为 $\Lambda = \{X, \sim X\}$, 分别表示对象 x 属于目标概念 X 和不属于目标概念 X ; 行为集表达为 $a = \{a_P, a_B, a_N\}$, 其中 a_P 、 a_B 和 a_N 分别代表将对象 x 划分到正域 (P_{OS})、边界域 (B_{ND}) 和负域 (N_{EG}) 的 3 种决策行为。 $\lambda_{\Delta P}(\Delta = P, B, N)$ 表示 x 属于 X 时采取行为 Δ 的损失, 同理, 当 x 不属于 X 时, 采取行为 Δ 的损失为 $\lambda_{\Delta N}$ 。则采取 3 种行为 a_P 、 a_B 和 a_N 的预期损失可分别表示为

$$\mathfrak{R}(a_P|[x]_E) = \lambda_{PP}P(X|[x]_E) + \lambda_{PN}P(X|\sim x|_E)$$

$$\mathfrak{R}(a_B|[x]_E) = \lambda_{BP}P(X|[x]_E) + \lambda_{BN}P(X|\sim x|_E)$$

$$\mathfrak{R}(a_N|[x]_E) = \lambda_{NP}P(X|[x]_E) + \lambda_{NN}P(X|\sim x|_E)$$

根据贝叶斯风险决策理论, 若损失函数满足 $\lambda_{PP} \leq \lambda_{BP} < \lambda_{NP}$, $\lambda_{NN} \leq \lambda_{BN} < \lambda_{PN}$, 则可获得以下决策规则:

- 1) 若 $P(X|[x]_E) \geq \alpha$, 设定 $x \in P_{OS}(X)$, 表示接受决策;
- 2) 若 $\beta < P(X|[x]_E) < \alpha$, 设定 $x \in B_{ND}$, 表示延迟决策;
- 3) 若 $P(X|[x]_E) \leq \beta$, 设定 $x \in N_{EG}(X)$, 表示拒绝决策。

其中阈值对 (α, β) 为

$$\alpha = \frac{\lambda_{PN} - \lambda_{BN}}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})}$$

$$\beta = \frac{\lambda_{BN} - \lambda_{NN}}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})} \quad (1)$$

1.3 广义多粒度粗糙集

定义 4^[26] 设信息系统 $I = (U, A, V, M)$, $E_i \subseteq A$, $i = 1, 2, \dots, l (l \leq 2^{|A|})$ 。对于 $X \subseteq U$, $S_X^{E_i}(x)$ 称为 X 的支持特征函数, 用于描述等价类 $[x]_{E_i}$ 和 X 的包含关系, 定义为

$$S_X^{E_i}(x) = \begin{cases} 1, & [x]_{E_i} \subseteq X \\ 0, & \text{其他} \end{cases}$$

定义 5^[26] 设信息系统 $I = (U, A, V, M)$, 参数满

足 $0.5 \leq \beta < 1$, $E_i \subseteq A$, $i = 1, 2, \dots, l (l \leq 2^{|A|})$, $S_X^{E_i}(x)$ 称为 X 的支持特征函数。对于 $X \subseteq U$, X 关于 $\sum_{i=1}^l E_i$ 的下近似和上近似分别定义为

$$\frac{G_M}{\sum_{i=1}^l E_i} (X)^\beta = \left\{ x \in U \mid \frac{\sum_{i=1}^l S_X^{E_i}(x)}{l} \geq \beta \right\}$$

$$\frac{\bar{G}_M}{\sum_{i=1}^l E_i} (X)^\beta = \left\{ x \in U \mid \frac{\sum_{i=1}^l (1 - S_X^{E_i}(x))}{l} > 1 - \beta \right\}$$

1.4 多源信息系统

定义 6^[16] 多源信息系统 (M_S) 由多个信息系统 $I_i = (U, A_i, V_i, M_i)$ 组成。对于任何 $i \in s$, I_i 表示 M_S 中第 i 个信息系统。其中论域 U 是有限非空对象集合; A_i 表示 I_i 下的属性集合; M_i 是 I_i 下的映射函数, 用来表示对象和属性之间的关系, V_i 是属性集的值域。则多源信息系统可定义为

$$M_S = \{I_1, I_2, \dots, I_s\}$$

特殊地, 当 $A_i = C_i \cup D$ 时, 多源信息系统也称为多源决策信息系统 (M_H)。其由多个决策信息系统 $H_i = (U, C_i \cup D, V_i, M_i)$ 组成, 对于任何 $i \in s$, H_i 表示 M_H 中第 i 个决策信息系统。 D 表示决策属性, $U/D = \{D_1, D_2, \dots, D_m\}$ 为对 U 的一个划分, 则多源决策信息系统可定义为

$$M_H = \{H_1, H_2, \dots, H_s\}$$

2 广义自适应多粒度粗糙集模型

从不同视角和层次采集的多源数据的标准不同, 数据的准确性和完备性也不同, 因此对多源数据生成的知识粒度采用相同的阈值来获取决策规则并不妥当。为解决这个问题, 本文提出了一个用于多源信息系统的广义自适应多粒度粗糙集模型, 该模型与传统的两个融合方法对比如图 1 所示。首先, 本文介绍如何在计算 3 个区域时自适应获得阈值对的方法。

2.1 自适应阈值对获取

经典决策理论模型需要预设多个损失参数才能获得阈值对 (α, β) , 在实际中过于复杂, 且很难确定参数值。为此, Suo 等^[27] 提出单参数决策理论粗糙集, 只需要预设单个参数, 即可获得数据驱动的损失函数矩阵, 并计算得到阈值对 (α, β) 。假设 $H = (U, C \cup D, V, M)$, 条件属性子集 $E \subseteq C$, 记 $\Gamma = \{X, X^c\}$ 表示对象状态集, 表 1 为数据驱动的损失函数。其中, $\zeta (0 < \zeta \leq 1)$ 称为补偿系数, 表示对中立

的容忍程度。当 $|X| \neq 0$ 和 $X \cup X^c = U$ 时, $S(X|[x]_E) = \frac{|X \cap [x]_E|}{|X|}$, $S^c(X|[x]_E) = \frac{|X^c \cap [x]_E|}{|X|}$, 分别代表 $[x]_E$ 中

X 的重要性和 X^c 的重要性, 记 $S = S(X|[x]_E)$, $S^c = S^c(X|[x]_E)$ 。

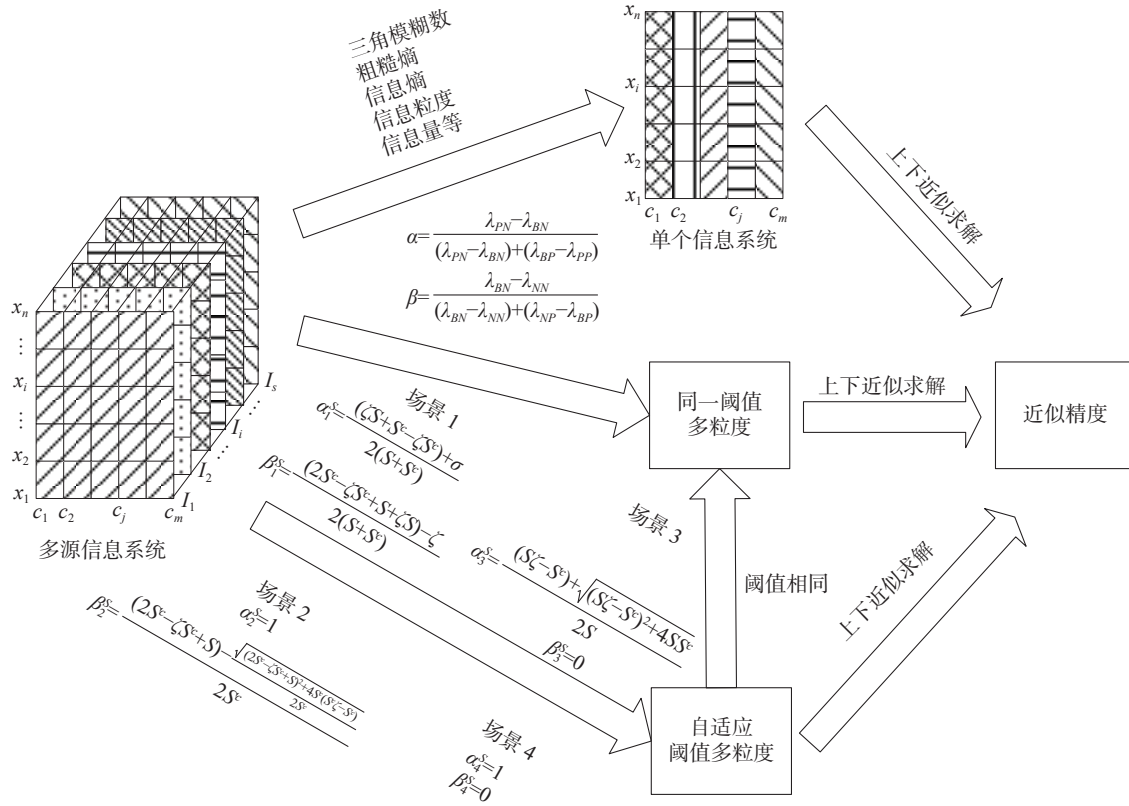


图 1 信息融合方法对比

Fig. 1 Comparison of information fusion methods

表 1 数据驱动的损失函数

Table 1 Data-driven loss function

决策动作	X	X^c
a_P	$\lambda_{PP} = 0$	$\lambda_{PN} = S^c(X [x]_E)$
a_B	$\lambda_{BP} = S(X [x]_E)(P(X [x]_E) - \zeta)$	$\lambda_{BN} = S^c(X [x]_E)(1 - P(X [x]_E) - \zeta)$
a_N	$\lambda_{NP} = S(X [x]_R)$	$\lambda_{NN} = 0$

根据表 1 和式 (1), 则新的阈值参数 α 和 β 可从以下 4 种场景计算得到:

1) 当 $P - \zeta > 0$ 且 $P + \zeta < 1$, 有 $\lambda_{BP} \neq 0$ 且 $\lambda_{BN} \neq 0$, 则阈值参数定义为

$$\begin{aligned} \alpha_1^S &= \frac{(\zeta S + S^c - \zeta S^c) + \sigma}{2(S + S^c)} \\ \beta_1^S &= \frac{(2S^c - \zeta S^c + S + \zeta S) - \zeta}{2(S + S^c)} \end{aligned} \quad (2)$$

式中:

$$\begin{aligned} \sigma &= \sqrt{(\zeta S + S^c - \zeta S^c)^2 + 4\zeta S^c(S^c + S)} \\ \varsigma &= \sqrt{(2S^c - \zeta S^c + S + \zeta S)^2 + 4(S + S^c)(\zeta S^c - S^c)} \end{aligned}$$

2) 当 $P \leq \zeta$ 且 $P + \zeta < 1$, 有 $\lambda_{BP} = 0$ 且 $\lambda_{BN} \neq 0$, 则阈值参数定义为

$$\begin{aligned} \alpha_2^S &= 1 \\ \beta_2^S &= \frac{(2S^c - \zeta S^c + S)}{2S^c} - \frac{\sqrt{(2S^c - \zeta S^c + S)^2 + 4S^c(S^c \zeta - S^c)}}{2S^c} \end{aligned} \quad (3)$$

3) 当 $P - \zeta > 0$ 且 $P + \zeta \geq 1$, 有 $\lambda_{BP} \neq 0$ 且 $\lambda_{BN} = 0$, 则阈值参数定义为

$$\begin{aligned} \alpha_3^S &= \frac{(S \zeta - S^c) + \sqrt{(S \zeta - S^c)^2 + 4S S^c}}{2S} \\ \beta_3^S &= 0 \end{aligned} \quad (4)$$

4) 当 $P \leq \zeta$ 且 $P + \zeta \geq 1$, 有 $\lambda_{BN} = 0$ 且 $\lambda_{BP} = 0$, 则阈值参数定义为

$$\alpha_4^S = 1, \quad \beta_4^S = 0 \quad (5)$$

通过对 4 种场景下阈值参数 α 和 β 求解的描

述, 我们知道以上两个参数由 S 、 S^c 和 ζ 共同决定, 因此可表示为 $\alpha^S = \phi(\zeta, S, S^c)$, $\beta^S = \nu(\zeta, S, S^c)$, 其中 $\alpha^S = \{\alpha_1^S, \alpha_2^S, \alpha_3^S, \alpha_4^S\}$, $\beta^S = \{\beta_1^S, \beta_2^S, \beta_3^S, \beta_4^S\}$ 。又因为 S 、 S^c 只取决于数据本身, 则参数又可记为 $\alpha^S = \phi(\zeta)$, $\beta^S = \nu(\zeta)$ 。进而单参数决策理论粗糙集模型相应的正域、边界域和负域可表示为

$$\begin{aligned} P_{OS}^S &= \{x \in U \mid P(X|[x]_E) \geq \alpha^S\} \\ B_{ND}^S &= \{x \in U \mid \beta^S < P(X|[x]_E) < \alpha^S\} \\ N_{EG}^S &= \{x \in U \mid P(X|[x]_E) \leq \beta^S\} \end{aligned}$$

式中: $0 < \zeta \leq 1$, 参数 α^S 和 β^S 由上述 4 种场景可计算得到。由文献 [27] 的理论 4 和理论 5 知道, $0 \leq \beta^S < \alpha^S \leq 1$ 恒成立。此外, 当 $0 < \zeta_2 \leq \zeta_1 \leq 1$ 时, 有 $0 \leq \beta^S(\zeta_1) \leq \beta^S(\zeta_2) < \alpha^S(\zeta_2) \leq \alpha^S(\zeta_1)$ 。

2.2 广义自适应多粒度粗糙集模型

多源信息系统本质上是由多个单源信息系统组成, 每个源诱导的不可分辨关系都会产生知识粒度。通过采用上节对知识粒度求解阈值的方法, 并结合多粒度的思想, 我们提出广义自适应多粒度粗糙集模型来实现多源信息系统的知识获取。

定义 7 称 $M_S = \{I_1, I_2, \dots, I_s\}$ 是多源信息系统, $I_i = (U, A_i, V_i, M_i)$, 参数满足 $0 < \zeta \leq 1$, $0 < \varphi \leq 1$, $0 < \gamma \leq 1$ 。在广义自适应多粒度粗糙集模型 MS-AMG 中, 对任意的 $X \subseteq U$, X 关于 M_S 的下近似和上近似分别定义为

$$\begin{aligned} \underline{M_S - A_{MG\varphi}^\zeta}(X) &= \left\{ x \in U \mid \frac{\sum_{i=1}^s F_X^{I_i, \zeta}(x)}{s} \geq \varphi \right\} \\ \overline{M_S - A_{MG\gamma}^\zeta}(X) &= \left\{ x \in U \mid \frac{\sum_{i=1}^s R_X^{I_i, \zeta}(x)}{s} \geq \gamma \right\} \end{aligned}$$

式中: $F_X^{I_i, \zeta}$ 和 $R_X^{I_i, \zeta}$ 分别为 X 的支持特征函数和相关特征函数, 定义为

$$\begin{aligned} F_X^{I_i, \zeta}(x) &= \begin{cases} 1, & P(X|[x]_{I_i}) \geq \alpha_i^\zeta \\ 0, & \text{其他} \end{cases} \\ R_X^{I_i, \zeta}(x) &= \begin{cases} 1, & P(X|[x]_{I_i}) > \beta_i^\zeta \\ 0, & \text{其他} \end{cases} \end{aligned}$$

其中: ζ 称为补偿系数; φ , γ 是近似参数。 $[x]_{I_i}$ 表示对象 x 在信息源 I_i 下所有属性形成的类。 $P(X|[x]_{I_i})$ 表示条件概率。 $\alpha_i^\zeta = \{\alpha_{i1}^\zeta, \alpha_{i2}^\zeta, \alpha_{i3}^\zeta, \alpha_{i4}^\zeta\}$, $\beta_i^\zeta = \{\beta_{i1}^\zeta, \beta_{i2}^\zeta, \beta_{i3}^\zeta, \beta_{i4}^\zeta\}$ 代表知识粒度上的阈值, 可由式 (2)~(5) 得到。

容易看出, 广义自适应多粒度粗糙集模型每个知识粒度的对应的阈值都是不尽相同, 而且是由数据驱动而自适应得到的。此外, 其还可以通过改变两个近似参数 φ 、 γ 调整模型的近似范围, 在实际融合过程中更具有灵活性。

定义 8 称 $M_S = \{I_1, I_2, \dots, I_s\}$ 是多源信息系统, $I_i = (U, A_i, V_i, M_i)$, 参数满足 $0 < \zeta \leq 1$, $0 < \varphi \leq 1$, $0 < \gamma \leq 1$ 。在广义自适应多粒度粗糙集模型 MS-AMG 中, 对任意的 $X \subseteq U$, X 关于 M_S 的正域、边界域和负域分别为

$$\begin{aligned} P_{OS}^{AMG}(X) &= \underline{M_S - A_{MG\varphi}^\zeta}(X) \\ B_{NG}^{AMG}(X) &= \overline{M_S - A_{MG\gamma}^\zeta}(X) - \underline{M_S - A_{MG\varphi}^\zeta}(X) \\ N_{EG}^{AMG}(X) &= U - \overline{M_S - A_{MG\gamma}^\zeta}(X) \end{aligned}$$

广义自适应多粒度粗糙集模型的有效性通常可以由多源决策信息系统中的融合近似精度来衡量。

定义 9 称 $M_H = \{H_1, H_2, \dots, H_s\}$ 是多源决策信息系统, $H_i = (U, C_i \cup D, V_i, M_i)$, 参数满足 $0 < \zeta \leq 1$, $0 < \varphi \leq 1$, $0 < \gamma \leq 1$, 决策等价类 $U/D = \{D_1, D_2, \dots, D_m\}$, 对于 $\forall D_j \in U/D (j = 1, 2, \dots, m)$, 融合近似精度定义为

$$\alpha_{AMG}^\zeta(U/D) = \frac{\sum_{j=1}^m |\underline{M_S - A_{MG\varphi}^\zeta}(D_j)|}{\sum_{j=1}^m |\overline{M_S - A_{MG\gamma}^\zeta}(D_j)|}$$

接下来我们设计了算法 1 来求解自适应多粒度粗糙集模型融合近似精度。

算法 1 MS-AMG 融合近似精度求解算法

输入 补偿系数 ζ , 参数 φ 、 γ 和多源决策信息系统 $M_H = \{H_1, H_2, \dots, H_s\}$;

输出 融合的近似精度 $\alpha_{AMG}^\zeta(U/D)$ 。

- 1) for 每一个 $D_j \subseteq U/D$ do
- 2) {for 每一个 $x_i \in U$ do
- 3) {初始化 $F(i) = 0$, $R(i) = 0$;
- 4) {for 每一个 $H_k \in M_H$ do
- 5) 计算对象 x_i 关于决策类 D_j 的条件概率 $P(D_j|[x]_{H_k})$ 及阈值对 $\alpha_k^\zeta, \beta_k^\zeta$;
- 6) 若 $P(D_j|[x]_{H_k}) \geq \alpha_k^\zeta$, 则
 $F(i) = F(i) + 1$;
- 7) 若 $P(D_j|[x]_{H_k}) > \beta_k^\zeta$, 则
 $R(i) = R(i) + 1$;
- 8) 若 $\frac{F(i)}{s} \geq \varphi$, 则将对象 x_i 划入决策类 D_j 的下近似 $\underline{M_S - A_{MG\varphi}^\zeta}(D_j)$;
- 9) 若 $\frac{R(i)}{s} \geq \gamma$, 则将对象 x_i 划入决策类 D_j 的上近似 $\overline{M_S - A_{MG\gamma}^\zeta}(D_j)$;
- 10) 返回融合近似精度

$$\alpha_{AMG}^\zeta(U/D) = \frac{\sum_{j=1}^m |\underline{M_S - A_{MG\varphi}^\zeta}(D_j)|}{\sum_{j=1}^m |\overline{M_S - A_{MG\gamma}^\zeta}(D_j)|}$$

算法 1 的时间复杂度为 $O(|U| \times m \times s)$, 其中 $|U|$ 为对象个数, m 是决策数量, s 表示信息源的数量。

2.3 经典自适应多粒度粗糙集模型

通过调整参数 φ 、 γ 的取值, 广义自适应多粒度粗糙集可以转换为 4 种经典的多粒度粗糙集模型。根据上下近似采取的策略, 可分为悲观-悲观自适应多粒度粗糙集模型 MS-PPAMG、乐观-乐观自适应多粒度粗糙集模型 MS-OOAMG、悲观-乐观自适应多粒度粗糙集模型 MS-POAMG 和乐观-悲观自适应多粒度粗糙集模型 MS-OPAMG。

假设 $M_S = \{I_1, I_2, \dots, I_s\}$ 是多源信息系统, 其中 $I_i = (U, A_i, V_i, M_i)$, $0 < \zeta \leq 1$, 则上述 4 种多粒度粗糙集模型的定义如下。

定义 10 在悲观-悲观自适应多粒度粗糙集模型 MS-PPAMG 中, 对任意的 $X \subseteq U$, X 关于 M_S 的下近似和上近似分别定义为

$$\begin{aligned} \underline{M_S - P_{PAMG}}^\zeta(X) &= \left\{ x \in X \mid \frac{\sum_{i=1}^s F_X^{I_i, \zeta}(x)}{s} = 1 \right\} \\ \overline{M_S - P_{PAMG}}^\zeta(X) &= \left\{ x \in X \mid \frac{\sum_{i=1}^s R_X^{I_i, \zeta}(x)}{s} = 1 \right\} \end{aligned}$$

MS-PPAMG 模型的正域、边界域和负域的表达与 MS-AMG 模型相同。

定义 11 在乐观-乐观自适应多粒度粗糙集模型 MS-OOAMG 中, 对任意的 $X \subseteq U$, X 关于 M_S 的下近似和上近似分别定义为

$$\begin{aligned} \underline{M_S - O_{OAMG}}^\zeta(X) &= \left\{ x \in X \mid \frac{\sum_{i=1}^s F_X^{I_i, \zeta}(x)}{s} > 0 \right\} \\ \overline{M_S - O_{OAMG}}^\zeta(X) &= \left\{ x \in X \mid \frac{\sum_{i=1}^s R_X^{I_i, \zeta}(x)}{s} > 0 \right\} \end{aligned}$$

MS-OOAMG 模型的正域、边界域和负域的表达与 MS-AMG 模型相同。

定义 12 在悲观-乐观自适应多粒度粗糙集模型 MS-POAMG 中, 对任意的 $X \subseteq U$, X 关于 M_S 的下近似和上近似分别定义为

$$\underline{M_S - P_{OAMG}}^\zeta(X) = \left\{ x \in X \mid \frac{\sum_{i=1}^s F_X^{I_i, \zeta}(x)}{s} = 1 \right\}$$

$$\overline{M_S - P_{OAMG}}^\zeta(X) = \left\{ x \in X \mid \frac{\sum_{i=1}^s R_X^{I_i, \zeta}(x)}{s} > 0 \right\}$$

MS-POAMG 模型的正域、边界域和负域的表达与 MS-AMG 模型相同。

定义 13 在乐观-悲观自适应多粒度粗糙集模型 MS-OPAMG 中, 对任意的 $X \subseteq U$, X 关于 M_S 的下近似和上近似分别定义为

$$\begin{aligned} \underline{M_S - O_{PAMG}}^\zeta(X) &= \left\{ x \in X \mid \frac{\sum_{i=1}^s F_X^{I_i, \zeta}(x)}{s} > 0 \right\} \\ \overline{M_S - O_{PAMG}}^\zeta(X) &= \left\{ x \in X \mid \frac{\sum_{i=1}^s R_X^{I_i, \zeta}(x)}{s} = 1 \right\} \end{aligned}$$

MS-OPAMG 模型的正域、边界域和负域的表达与 MS-AMG 模型相同。然而, MS-OPAMG 在某些情况下是不符合粗糙集理论的。例如, 假如存在 $I_i, I_j \in M_S$, 如果对象 x 有 $P(X|[x]_{I_i}) \geq \alpha_i^\zeta$, 且 $P(X|[x]_{I_j}) \leq \beta_j^\zeta$, 则 $F_X^{I_i, \zeta}(x) = 1, R_X^{I_j, \zeta}(x) = 0$ 。根据定义 13 知, $x \in \underline{M_S - O_{PAMG}}^\zeta(X)$, 并且 $x \notin \overline{M_S - O_{PAMG}}^\zeta(X)$ 。故 $\underline{M_S - O_{PAMG}}^\zeta(X) \supseteq \overline{M_S - O_{PAMG}}^\zeta(X)$ 不成立。

根据定义 10~13 可知, 给定不同的近似参数, 所提出的广义自适应多粒度粗糙集模型能够退化为经典策略的自适应多粒度粗糙集模型。这也体现所提模型具有灵活性, 能适应不同的场景。

2.4 广义自适应多粒度粗糙集模型的相关性质

这部分将介绍广义自适应多粒度粗糙集模型常见性质以及不同自适应多粒度粗糙集模型之间的关系。

定理 1 称 $M_S = \{I_1, I_2, \dots, I_s\}$ 是多源信息系统, 其中 $I_i = (U, A_i, V_i, M_i)$, $0 < \zeta \leq 1$, 如果 $0 < \varphi \leq 1$, $0 < \gamma \leq 1$, 对于任何 $X \subseteq U$, 则广义自适应多粒度粗糙集满足以下常见结论:

- 1) $\underline{M_S - A_{MG_\varphi}}^\zeta(\emptyset) = \overline{M_S - A_{MG_\gamma}}^\zeta(X) = \emptyset$;
- 2) $\underline{M_S - A_{MG_\varphi}}^\zeta(X) \subseteq X$;
- 3) $\overline{M_S - A_{MG_\gamma}}^\zeta(X) \subseteq X$;
- 4) $\underline{M_S - A_{MG_\varphi}}^\zeta(U) = \overline{M_S - A_{MG_\gamma}}^\zeta(U) = U$ 。

证明 根据定义 7, 显然结论 1)~3) 成立。对于结论 4, $\forall x \in U$, 有 $P(U|[x]_{I_i}) = 1$ 。则对于 $\forall I_i \in M_S$, 有 $F_X^{I_i, \zeta_1}(x) = 1 \geq \alpha_i^\zeta$, 进而可以得出

$$\frac{\sum_{i=1}^s F_X^{I_i, \zeta}(x)}{s} = 1 \geq \varphi$$

即 $x \in \underline{M_S - A_{MG}}_{\varphi}^{\zeta}(U)$, 可证的 $\underline{M_S - A_{MG}}_{\varphi}^{\zeta}(U) = U$ 。类似地, 我们也能证明 $\overline{M_S - A_{MG}}_{\gamma}^{\zeta}(U) = U$, 则定理 1 可证。

定理 2 称 $M_S = \{I_1, I_2, \dots, I_s\}$ 是多源信息系统, 其中 $I_i = (U, A_i, V_i, M_i)$, $0 < \zeta \leq 1$, 如果 $0 < \varphi_1 \leq \varphi_2 \leq 1$, $0 < \gamma_1 \leq \gamma_2 \leq 1$, 对于任何 $X \subseteq U$, 则以下结论成立:

- 1) $\underline{M_S - A_{MG}}_{\varphi_1}^{\zeta}(X) \supseteq \underline{M_S - A_{MG}}_{\varphi_2}^{\zeta}(X)$;
- 2) $\overline{M_S - A_{MG}}_{\gamma_1}^{\zeta}(X) \supseteq \overline{M_S - A_{MG}}_{\gamma_2}^{\zeta}(X)$ 。

证明 1) 对于 $\forall x \in \underline{M_S - A_{MG}}_{\varphi_2}^{\zeta}(X)$, 由定义 7 可知

$$\frac{\sum_{i=1}^s F_X^{I_i, \zeta}(x)}{s} \geq \varphi_2$$

若存在 $0 < \varphi_1 \leq \varphi_2 \leq 1$, 则有

$$\frac{\sum_{i=1}^s F_X^{I_i, \zeta}(x)}{s} \geq \varphi_1$$

即 $x \in \underline{M_S - A_{MG}}_{\varphi_1}^{\zeta}(X)$, 从而可证 $\underline{M_S - A_{MG}}_{\varphi_1}^{\zeta}(X) \supseteq \underline{M_S - A_{MG}}_{\varphi_2}^{\zeta}(X)$ 。

2) 结论 2) 证明相似。

定理 3 称 $M_S = \{I_1, I_2, \dots, I_s\}$ 是多源信息系统, 其中 $I_i = (U, A_i, V_i, M_i)$, 对于任何 $X \subseteq U$, 如果 $0 < \zeta_2 \leq \zeta_1 \leq 1$, $0 < \varphi \leq 1$, $0 < \gamma \leq 1$, 则以下结论成立:

- 1) $\underline{M_S - A_{MG}}_{\varphi}^{\zeta_1}(X) \subseteq \underline{M_S - A_{MG}}_{\varphi}^{\zeta_2}(X)$;
- 2) $\overline{M_S - A_{MG}}_{\gamma}^{\zeta_1}(X) \supseteq \overline{M_S - A_{MG}}_{\gamma}^{\zeta_2}(X)$;
- 3) $\alpha_{AMG}^{\zeta_1}(X) \leq \alpha_{AMG}^{\zeta_2}(X)$ 。

证明 1) 对于 $\forall x \in \underline{M_S - A_{MG}}_{\varphi}^{\zeta_1}(X)$, 由定义 7 可知

$$\frac{\sum_{i=1}^s F_X^{I_i, \zeta_1}(x)}{s} \geq \varphi$$

若存在 $\zeta_2 \leq \zeta_1$, 则有 $\alpha_i^{\zeta_1} \geq \alpha_i^{\zeta_2}$ 。因此对于 $\forall I_i \in M_S$, 有 $F_X^{I_i, \zeta_1}(x) \leq F_X^{I_i, \zeta_2}(x)$, 进而可以得出

$$\frac{\sum_{i=1}^s F_X^{I_i, \zeta_2}(x)}{s} \geq \varphi$$

即 $x \in \underline{M_S - A_{MG}}_{\varphi}^{\zeta_2}(X)$, 则结论 1 可证明。

2) 结论 2、3 证明相似。

定理 4 称 $M_S = \{I_1, I_2, \dots, I_s\}$ 是多源信息系统, 其中 $I_i = (U, A_i, V_i, M_i)$, $0 < \zeta \leq 1$, $0 < \varphi \leq 1$, $0 < \gamma \leq 1$, 对于任何 $X \subseteq U$, 以下结论成立。

- 1) $\underline{M_S - P_{PAMG}}^{\zeta}(X) = \underline{M_S - P_{OAMG}}^{\zeta}(X) \subseteq \underline{M_S - A_{MG}}_{\varphi}^{\zeta}(X) \subseteq \underline{M_S - O_{AMG}}^{\zeta}(X)$;
- 2) $\overline{M_S - P_{PAMG}}^{\zeta}(X) \subseteq \overline{M_S - A_{MG}}_{\gamma}^{\zeta}(X) \subseteq \overline{M_S - O_{AMG}}^{\zeta}(X) = \overline{M_S - P_{OAMG}}^{\zeta}(X)$ 。

证明 1) 对于 $x \in \underline{M_S - P_{PAMG}}^{\zeta}(X)$, 由定义 10 知

$$\frac{\sum_{i=1}^s F_X^{I_i, \zeta}(x)}{s} = 1 \geq \varphi$$

则 $x \in \underline{M_S - A_{MG}}_{\varphi}^{\zeta}(X)$, 即

$$\underline{M_S - P_{PAMG}}^{\zeta}(X) \subseteq \underline{M_S - A_{MG}}_{\varphi}^{\zeta}(X);$$

同理, 当 $x \in \overline{M_S - A_{MG}}_{\gamma}^{\zeta}(X)$ 时, 有

$$\frac{\sum_{i=1}^s F_X^{I_i, \zeta}(x)}{s} \geq \varphi > 0$$

即 $x \in \underline{M_S - O_{AMG}}^{\zeta}(X)$, 则结论 1 可证明。

2) 结论 2 证明类似。

2.5 案例分析

本节将用一个癌症诊断案例来解释广义自适应多粒度粗糙集的应用。假设表 2 是某医院癌症诊断的多源决策信息表 M_H 。 $M_H = \{H_1, H_2, H_3\}$, 其中 $H_i (i=1, 2, 3)$ 分别表示检测团块厚度、细胞大小的均匀性和细胞形状的均匀性得到的单源决策信息系统, $U = \{x_1, x_2, \dots, x_{10}\}$ 是一个包含了 10 个病人的集合, d 是决策属性, 表示病人是否诊断为癌症。

表 2 多源决策表

Table 2 Multi-source decision table

U	H_1		H_2			H_3			d
	a_1	a_2	b_1	b_2	b_3	c_1	c_2	c_3	
x_1	1	0	1	1	0	0	1	1	0
x_2	1	0	1	1	0	0	1	1	1
x_3	1	0	1	0	1	0	0	0	1
x_4	0	1	0	1	0	1	1	0	0
x_5	0	1	1	0	1	0	0	0	1
x_6	1	0	0	1	0	0	1	1	0
x_7	1	0	1	0	1	0	0	0	0
x_8	0	1	1	0	1	1	1	0	1
x_9	1	0	1	0	1	1	1	0	1
x_{10}	0	1	0	1	0	1	1	0	1

首先, 分别通过 3 个单源决策信息系统的条件属性集 C_i 和决策属性 d 对 U 进行划分, 结果如下:

$$U/R_{C_1} = \{\{x_1, x_2, x_3, x_6, x_7, x_9\}, \{x_4, x_5, x_8, x_{10}\}\},$$

$$U/R_{C_2} = \{\{x_1, x_2\}, \{x_3, x_5, x_7, x_8, x_9\}, \{x_4, x_6, x_{10}\}\},$$

$$U/R_{C_3} = \{\{x_1, x_2, x_6\}, \{x_3, x_5, x_7\}, \{x_4, x_8, x_9, x_{10}\}\},$$

$$U/R_d = \{\{x_1, x_4, x_6, x_7\}, \{x_2, x_3, x_5, x_8, x_9, x_{10}\}\}。$$

然后计算 $[x]_{H_i}$ 相对于决策属性 D_0 、 D_1 的条件概率 P_0 和 P_1 , 如表 3 所示。

令 $\zeta = 0.3$, 然后根据式 (2) 到 (5) 计算 $[x]_{DIS}$ 相对于决策 D_0 、 D_1 的阈值对, 分别如表 4 和表 5 所示。

表 3 条件概率
Table 3 Conditional probability

U	H ₁		H ₂		H ₃	
	P ₀	P ₁	P ₀	P ₁	P ₀	P ₁
x ₁	0.50	0.50	0.50	0.50	0.67	0.33
x ₂	0.50	0.50	0.50	0.50	0.67	0.33
x ₂	0.50	0.50	0.20	0.80	0.33	0.67
x ₄	0.25	0.75	0.67	0.33	0.25	0.75
x _i	0.25	0.75	0.20	0.80	0.33	0.67
x _b	0.50	0.50	0.67	0.33	0.67	0.33
x ₁	0.50	0.50	0.20	0.80	0.33	0.67
x ₂	0.25	0.75	0.20	0.80	0.25	0.75
x ₉	0.50	0.50	0.20	0.80	0.25	0.75
x ₁₀	0.25	0.75	0.67	0.33	0.25	0.75

表 4 决策 D₀ 上的阈值对
Table 4 A pair of thresholds of D₀

U	H ₁	H ₂	H ₃
	(a ₁ , b ₁)	(a ₂ , b ₂)	(a ₃ , b ₃)
x ₁	(0.71, 0.29)	(0.71, 0.29)	(0.60, 0.19)
x ₂	(0.71, 0.29)	(0.71, 0.29)	(0.60, 0.19)
x ₃	(0.71, 0.29)	(1.00, 0.47)	(0.81, 0.40)
x ₄	(1.00, 0.44)	(0.60, 0.19)	(1.00, 0.44)
x ₅	(1.00, 0.44)	(1.00, 0.47)	(0.81, 0.40)
x ₆	(0.71, 0.29)	(0.60, 0.19)	(0.60, 0.19)
x ₇	(0.71, 0.29)	(1.00, 0.47)	(0.81, 0.40)
x ₈	(1.00, 0.44)	(1.00, 0.47)	(1.00, 0.44)
x ₉	(0.71, 0.29)	(1.00, 0.47)	(1.00, 0.44)
x ₁₀	(1.00, 0.44)	(0.60, 0.19)	(1.00, 0.44)

表 5 决策 D₁ 上的阈值对
Table 5 A pair of thresholds of D₁

U	H ₁	H ₂	H ₃
	(a ₁ , b ₁)	(a ₂ , b ₂)	(a ₃ , b ₃)
x ₁	(0.71, 0.29)	(0.71, 0.29)	(0.81, 0.40)
x ₂	(0.71, 0.29)	(0.71, 0.29)	(0.81, 0.40)
x ₃	(0.71, 0.29)	(0.53, 0.00)	(0.60, 0.19)
x ₄	(0.56, 0.00)	(0.81, 0.40)	(0.56, 0.00)
x ₅	(0.56, 0.00)	(0.53, 0.00)	(0.60, 0.19)
x ₆	(0.71, 0.29)	(0.81, 0.40)	(0.81, 0.40)
x ₇	(0.71, 0.29)	(0.53, 0.00)	(0.60, 0.19)
x ₈	(0.56, 0.00)	(0.53, 0.00)	(0.56, 0.00)
x ₉	(0.71, 0.29)	(0.53, 0.00)	(0.56, 0.00)
x ₁₀	(0.56, 0.00)	(0.81, 0.40)	(0.56, 0.00)

1) 在 M_H 中, 令 $\varphi = 0.6, \gamma = 0.2$, 则 MS-AMG 模型在每个决策类的下近似和上近似为

$$\begin{aligned} \underline{M}_S - A_{MG0.6}^{0.3}(D_0) &= \{x_6\}, \\ \overline{M}_S - A_{MG0.2}^{0.3}(D_0) &= \{x_1, x_2, x_3, x_4, x_6, x_7, x_9, x_{10}\}, \\ \underline{M}_S - A_{MG0.6}^{0.3}(D_1) &= \{x_3, x_4, x_5, x_7, x_8, x_9, x_{10}\}, \\ \overline{M}_S - A_{MG0.2}^{0.3}(D_1) &= \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}. \end{aligned}$$

对应融合近似精度为

$$\alpha_{AMG}^{0.3}(U/D) = \frac{\sum_{j=1}^2 \left| \underline{M}_S - A_{MG0.6}^{0.3}(D_j) \right|}{\sum_{j=1}^2 \left| \overline{M}_S - A_{MG0.2}^{0.3}(D_j) \right|} = \frac{4}{9}$$

2) 在 M_H 中, 则 MS-PPAMG 模型在每个决策类的下近似和上近似分别为

$$\begin{aligned} \underline{M}_S - P_{PAMG}^{0.3}(D_0) &= \emptyset \\ \overline{M}_S - P_{PAMG}^{0.3}(D_0) &= \{x_1, x_2, x_6\} \\ \underline{M}_S - P_{PAMG}^{0.3}(D_1) &= \{x_5, x_8\} \\ \overline{M}_S - P_{PAMG}^{0.3}(D_1) &= \{x_3, x_5, x_7, x_8, x_9\} \end{aligned}$$

对应融合近似精度为

$$\alpha_{PPAMG}^{0.3}(U/D) = \frac{\sum_{j=1}^2 \left| \underline{M}_S - P_{PAMG}^{0.3}(D_j) \right|}{\sum_{j=1}^2 \left| \overline{M}_S - P_{PAMG}^{0.3}(D_j) \right|} = \frac{1}{4}$$

3) 在 M_H 中, 则 MS-OOAMG 模型在每个决策类的下近似和上近似分别为

$$\begin{aligned} \underline{M}_S - O_{OAMG}^{0.3}(D_0) &= \{x_1, x_2, x_4, x_6, x_{10}\} \\ \overline{M}_S - O_{OAMG}^{0.3}(D_0) &= \{x_1, x_2, x_3, x_4, x_6, x_7, x_9, x_{10}\} \\ \underline{M}_S - O_{OAMG}^{0.3}(D_1) &= \{x_3, x_4, x_5, x_7, x_8, x_9, x_{10}\} \\ \overline{M}_S - O_{OAMG}^{0.3}(D_1) &= \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\} \end{aligned}$$

对应融合近似精度为

$$\alpha_{OOAMG}^{0.3}(U/D) = \frac{\sum_{j=1}^2 \left| \underline{M}_S - O_{OAMG}^{0.3}(D_j) \right|}{\sum_{j=1}^2 \left| \overline{M}_S - O_{OAMG}^{0.3}(D_j) \right|} = \frac{2}{3}$$

4) 在 M_H 中, 则 MS-POAMG 模型在每个决策类的下近似和上近似分别为

$$\begin{aligned} \underline{M}_S - P_{OAMG}^{0.3}(D_0) &= \emptyset \\ \overline{M}_S - P_{OAMG}^{0.3}(D_0) &= \{x_1, x_2, x_3, x_4, x_6, x_7, x_9, x_{10}\} \\ \underline{M}_S - P_{OAMG}^{0.3}(D_1) &= \{x_5, x_8\} \\ \overline{M}_S - P_{OAMG}^{0.3}(D_1) &= \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\} \end{aligned}$$

对应融合近似精度为

$$\alpha_{\text{POAMG}}^{0.3}(U/D) = \frac{\sum_{j=1}^2 \left| \overline{M_S - P_{\text{OAMG}}^{0.3}}(D_j) \right|}{\sum_{j=1}^2 \left| \overline{M_S - P_{\text{OAMG}}^{0.3}}(D_j) \right|} = \frac{1}{9}$$

5) 在 M_H 中, 则根据 MS-OPAMG 模型在每个决策类的下近似和上近似分别为

$$\begin{aligned} \overline{M_S - O_{\text{PAMG}}^{0.3}}(D_0) &= \{x_1, x_2, x_4, x_6, x_{10}\} \\ \overline{M_S - O_{\text{PAMG}}^{0.3}}(D_0) &= \{x_1, x_2, x_6\} \\ \overline{M_S - O_{\text{PAMG}}^{0.3}}(D_1) &= \{x_3, x_4, x_5, x_7, x_8, x_9, x_{10}\} \\ \overline{M_S - O_{\text{PAMG}}^{0.3}}(D_1) &= \{x_3, x_5, x_7, x_8, x_9\} \end{aligned}$$

对应融合近似精度为

$$\alpha_{\text{OPAMG}}^{0.3}(U/D) = \frac{\sum_{j=1}^2 \left| \overline{M_S - O_{\text{PAMG}}^{0.3}}(D_j) \right|}{\sum_{j=1}^2 \left| \overline{M_S - O_{\text{PAMG}}^{0.3}}(D_j) \right|} = \frac{3}{2}$$

通过分析以上结果, 所提出的广义自适应多粒度粗糙集 MS-AMG 可以自适应求得知识粒度 $[x]_H$ 的阈值, 相对于取相同阈值的多粒度决策理论粗糙集模型来说更具有合理性。并且与采取 4 种经典策略的多粒度粗糙集模型相比, 模型 MS-AMG 在决策 $D_i (i=1, 2)$ 的上下近似显然有

$$\begin{aligned} \overline{M_S - P_{\text{PAMG}}^{0.3}}(D_i) &= \overline{M_S - P_{\text{OAMG}}^{0.3}}(D_i) \subseteq \\ \overline{M_S - A_{\text{MG}, 0.6}^{0.3}}(D_i) &\subseteq \overline{M_S - O_{\text{OAMG}}^{0.3}}(D_i) = \\ &= \overline{M_S - O_{\text{PAMG}}^{0.3}}(D_i) \\ \overline{M_S - P_{\text{PAMG}}^{0.3}}(D_i) &= \overline{M_S - O_{\text{PAMG}}^{0.3}}(D_i) \subseteq \\ \overline{M_S - A_{\text{MG}, 0.2}^{0.3}}(D_i) &\subseteq \overline{M_S - O_{\text{OAMG}}^{0.3}}(D_i) = \\ &= \overline{M_S - P_{\text{OAMG}}^{0.3}}(D_i) \end{aligned}$$

此外, 对于 MS-OPAMG 模型, 其关于决策 D_i 的下近似并不包含在上近似内, 导致其融合近似精度也大于 1, 这种情况在粗糙集理论中并不合理, 而 MS-AMG 模型可以通过控制参数 φ 、 γ 范围来避免这种情况发生。

3 实验

为了进一步验证广义自适应多粒度粗糙集模

型的有效性, 本节选择 6 个 UCI 数据集进行实验分析, 数据集详细信息如表 6 所示。在本文中, 算法是使用软件 IntelliJ IDEA 来实现的, 实验环境为 PC 机, 其配置为 Intel(R) Core(TM) i5-6300HQ @ 2.30 GHz, 20 GB 内存, Windows 10 操作系统。

表 6 数据集的详细信息

Table 6 Specific information about the data sets

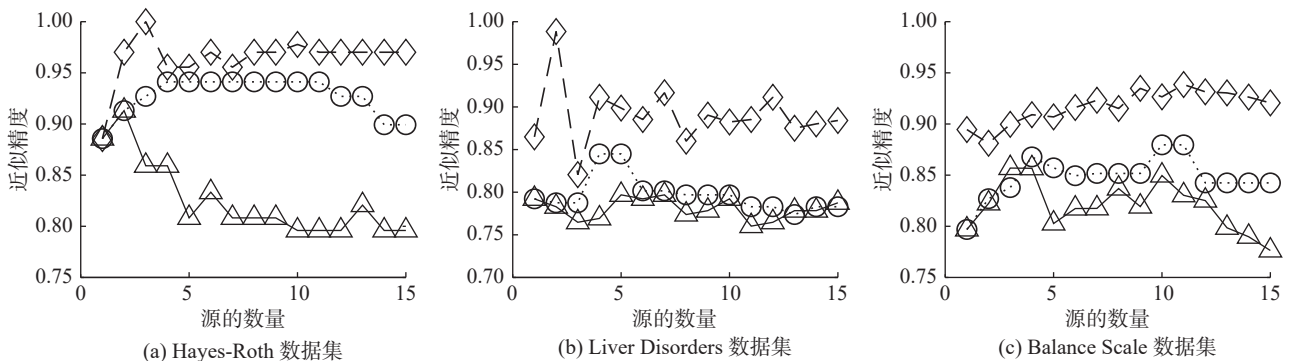
序号	数据集	数目	属性数	决策类
1	Hayes-Roth	160	4	3
2	Liver Disorders	345	7	2
3	Balance Scale	625	4	3
4	Banknote authentication	1372	5	2
5	Wireless Indoor Localization	2000	7	4
6	Abalone	4177	8	3

多源数据集在生活中并不常见, 为了建立多源数据集, 我们将原始数据添加噪声的方式生成包含 15 个信息源的多源数据集, 添加噪声来扩充为多源数据集的方法可参考文献 [21]。

3.1 本文模型与不同集成单源的模型对比分析

本节主要对比 MS-AMG 模型与两种集成为单源信息系统的模型之间的融合近似精度, 图 2 显示了 3 种模型在不同数据集下近似精度的实验结果。

其中, CE-MSD(conditional entropy multi-source decision method) [3] 是基于条件熵将多源信息系统集成为单源信息系统的模型, M-MSD(mean multi-source decision method) [3] 是基于均值法合成单源信息系统的模型。MS-AMG 的补偿系数设置 $\zeta = 0.1$ 。通过实验结果发现, MS-AMG 可以在不同的数据集上设置合适的近似参数, 使得其融合效果都好于模型 CE-MSD 和 M-MSD。参数具体设置如表 7 所示。例如, 对于“Hayes-Roth”数据集近似参数设置为 $\varphi = 0.3$ 和 $\gamma = 0.7$, 其融合效果会高于其他两种模型, 这也反应了该数据集生成的多个信息系统差异性很大。



(a) Hayes-Roth 数据集

(b) Liver Disorders 数据集

(c) Balance Scale 数据集

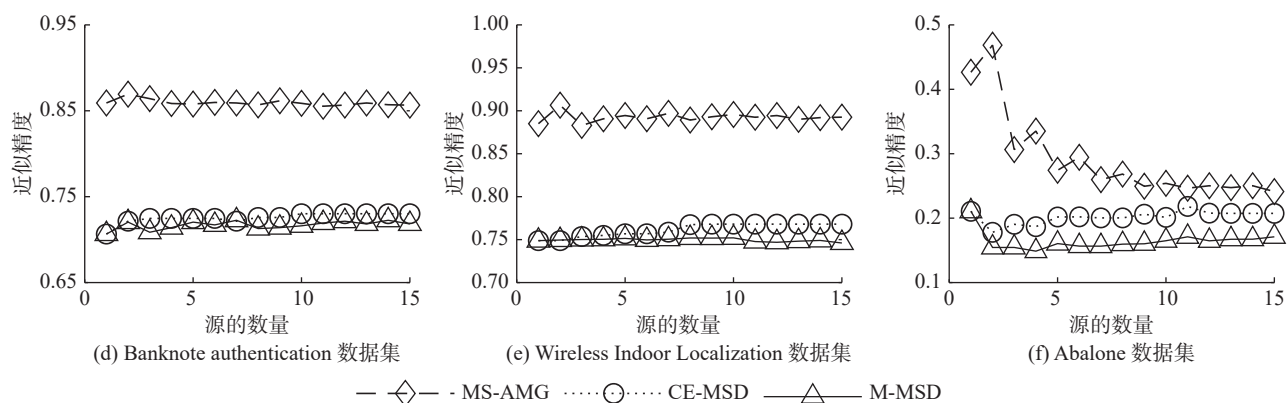


图 2 模型 MS-AMG、CE-MSD 和 M-MSD 的近似精度对比

Fig. 2 Comparison of approximation accuracy between MS-AMG, CE-MSD and CE-MSD models

表 7 数据集的参数信息

Table 7 Parameters information about the data sets

序号	数据集	φ	γ
1	Hayes-Roth	0.3	0.7
2	Liver Disorders	0.4	0.6
3	Balance Scale	0.5	0.5
4	Banknote authentication	0.5	0.5
5	Wireless Indoor Localization	0.4	0.6
6	Abalone	0.5	0.5

由此可知, MS-AMG 模型可以针对不同的多

源信息系统采取不同的近似参数,从而达到良好的融合效果。

3.2 本文模型与不同多粒度融合模型的对比分析

为了进一步验证 MS-AMG 模型在融合近似精度中的有效性,本节将对不同的多粒度融合模型进行近似精度对比。其中,包括两组补偿系数 ζ 分别取为 0.1、0.3 的 MS-AMG 模型,两组固定阈值对设置为 (0.9, 0.1) 和 (0.7, 0.3) 多粒度模型以及文献 [21] 提出的多粒度融合模型 (MG)。所有多粒度模型的上下近似参数都设置为 0.4 和 0.6。各个模型的近似精度如图 3 所示。

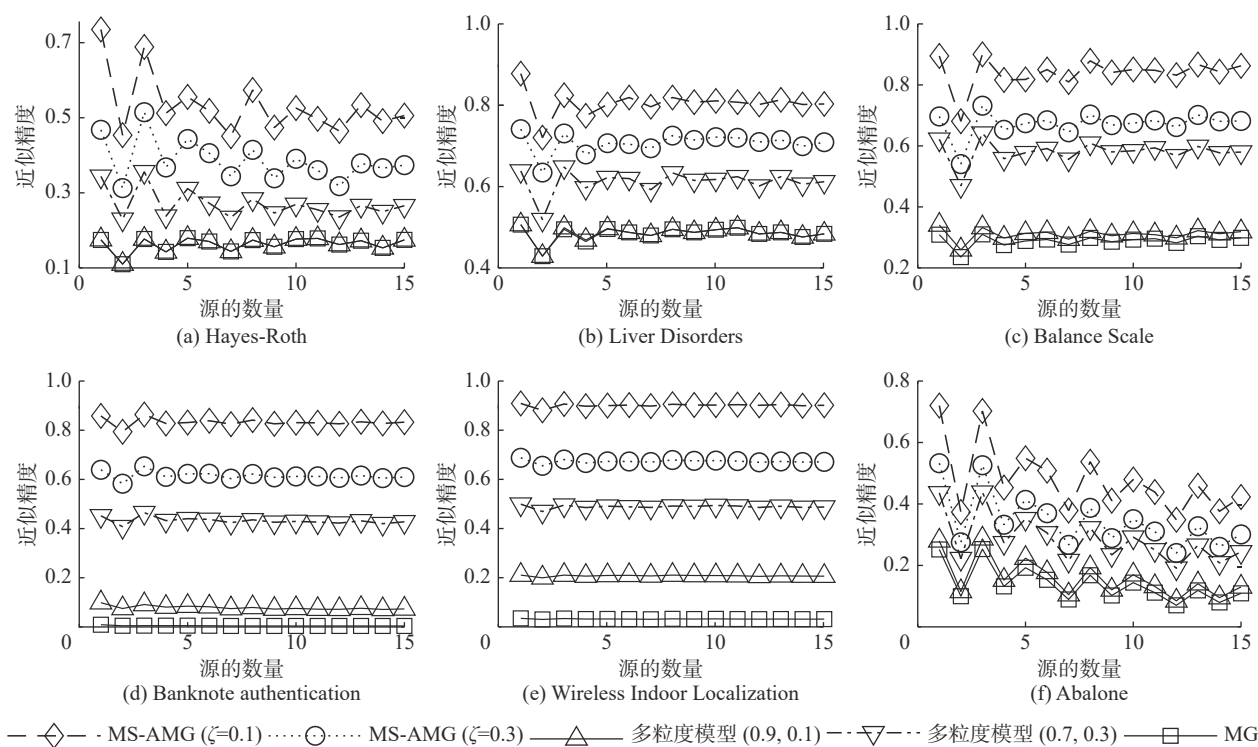


图 3 不同多粒度模型的融合近似精度比较

Fig. 3 Comparison of fusion approximation accuracy of different multi-granulation models

其中,模型 MG 的近似精度最低,而采用固定阈值的多粒度模型以及自适应阈值的模型 MS-

AMG,由于对决策规则具有容忍性,其融合效果相对较好。但从参数数量和实际应用上看,采用

自适应阈值的方式的 MS-AMG 更加具有优越性。此外, 实验结果显示 MS-AMG 模型的近似精度随着 ζ 的减少而增大, 符合定理 3。

3.3 4 种广义自适应多粒度粗糙集模型对比分析

在本小节中, 我们通过近似精度来比较广义自适应多粒度粗糙集 MS-AMG、悲观-悲观自适应多粒度粗糙集 MS-PPAMG、乐观-乐观自适应多粒度粗糙集 MS-OOAMG 和悲观-乐观自适应

多粒度粗糙集 MS-POAMG。4 种模型的补偿系数 $\zeta=0.1$, MS-AMG 的近似参数为 $\varphi=0.5$, $\gamma=0.5$ 。上述 4 种模型的知识粒度对应的阈值都是自适应获得的, 并且所有模型可以看作是 MS-AMG 近似参数取不同值而得到的。从图 4 展示的结果可以看出, 不同近似参数的 MS-AMG 融合效果也不相同, 因而可以根据实际情况选择合适的策略和模型, 灵活解决复杂的信息融合问题。

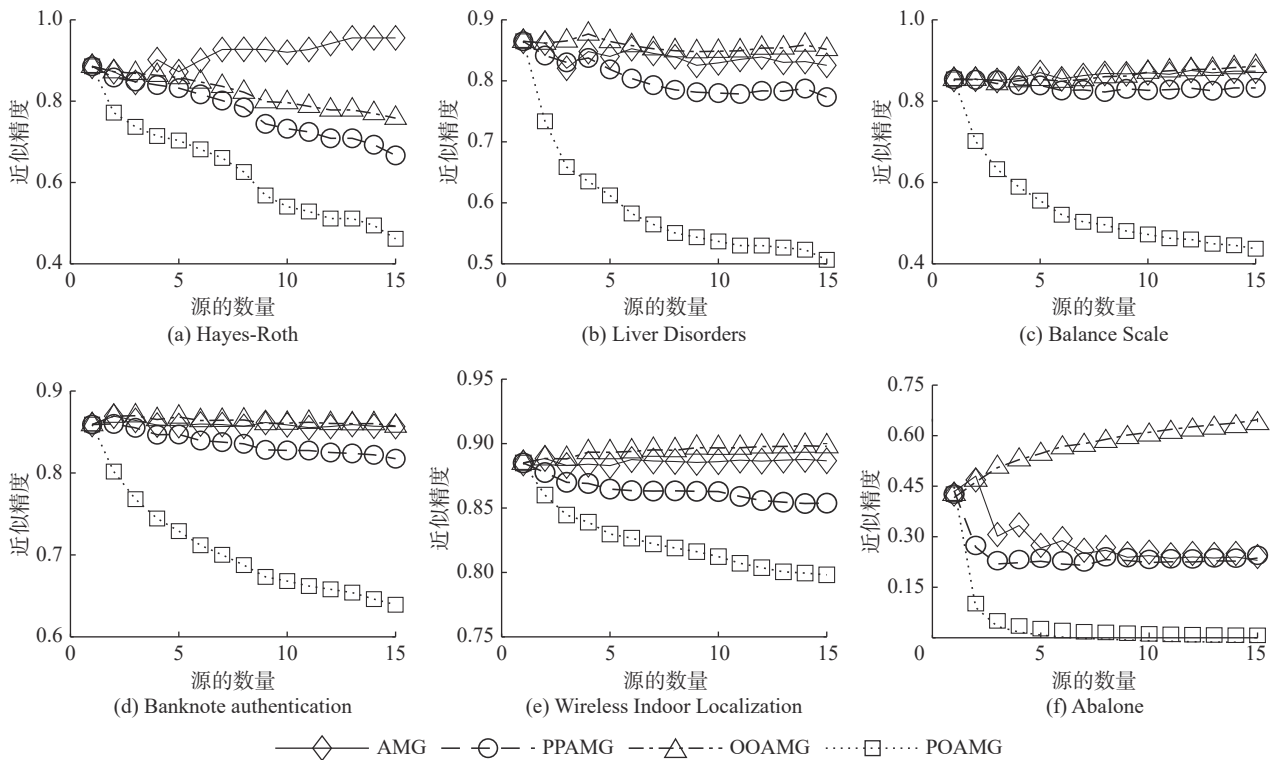


图 4 4 种自适应多粒度模型的融合近似精度比较

Fig. 4 Comparison of fusion approximation accuracy of four adaptive multi-granulation models

4 结束语

本文通过分析认为, 现有多粒度粗糙集模型通常采用同一阈值融合多源信息, 并没有考虑信息源之间的差异性。本文为解决这个问题, 将单参数决策粗糙集理论运用到广义多粒度粗糙集中, 提出了广义自适应多粒度粗糙集模型 MS-AMG, 并且定义了 4 种经典多粒度环境下 MS-AMG 模型的表现形式。所提模型都能够基于数据驱动得到每个知识粒度的阈值, 解决了同一阈值多粒度粗糙集实现信息融合的限制性。然后给出了一个例子来说明 MS-AMG 模型的应用, 初步显示了所提出的模型的优越性。最后通过实验进一步表明, 我们所提出模型的融合方法更加具有灵活性, 能处理好多源信息的融合, 为信息融合提供了新的思路。在未来的研究中, 将开展基于自适应多粒度序贯三支决策模型的研究。

参考文献:

- [1] YU Jianhang, XU Weihua. Information fusion for multi-source fuzzy information system with the same structure[C]//2015 International Conference on Machine Learning and Cybernetics. Guangzhou: IEEE, 2015: 170-175.
- [2] XU Weihua, YU Jianhang. A novel approach to information fusion in multi-source datasets: a granular computing viewpoint[J]. *Information sciences*, 2017, 378: 410-423.
- [3] SANG Binbin, GUO Yanting, SHI Derong, et al. Decision-theoretic rough set model of multi-source decision systems[J]. *International journal of machine learning and cybernetics*, 2018, 9(11): 1941-1954.
- [4] ZHANG Pengfei, LI Tianrui, YUAN Zhong, et al. A data-level fusion model for unsupervised attribute selection in

- multi-source homogeneous data[J]. *Information fusion*, 2022, 80: 87–103.
- [5] ZHANG Xiaoyan, CHEN Xiuwei, XU Weihua, et al. Dynamic information fusion in multi-source incomplete interval-valued information system with variation of information sources and attributes[J]. *Information sciences*, 2022, 608: 1–27.
- [6] QIAN Yuhua, LIANG Jiye, YAO Yiyu, et al. MGRS: a multi-granulation rough set[J]. *Information sciences*, 2010, 180(6): 949–970.
- [7] YANG Xibei, QI Yunsong, SONG Xiaoning, et al. Test cost sensitive multigranulation rough set: model and minimal cost selection[J]. *Information sciences*, 2013, 250: 184–199.
- [8] LIU Caihui, MIAO Duoqian, QIAN Jin. On multi-granulation covering rough sets[J]. *International journal of approximate reasoning*, 2014, 55(6): 1404–1418.
- [9] QIAN Yuhua, ZHANG Hu, SANG Yanli, et al. Multi-granulation decision-theoretic rough sets[J]. *International journal of approximate reasoning*, 2014, 55(1): 225–237.
- [10] SUN Bingzhen, MA Weimin. Multigranulation rough set theory over two universes[J]. *Journal of intelligent & fuzzy systems*, 2015, 28(3): 1251–1269.
- [11] 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.
- LIANG Meishe, MI Jusheng, ZHAO Tianna. Generalized dominance-based multi-granularity intuitionistic fuzzy rough set and acquisition of decision rules[J]. *CAAI transactions on intelligent systems*, 2017, 12(6): 883–888.
- [12] YU Jianhang, ZHANG Biao, CHEN Minghao, et al. Double-quantitative decision-theoretic approach to multi-granulation approximate space[J]. *International journal of approximate reasoning*, 2018, 98: 236–258.
- [13] QIAN Jin, LIU Caihui, MIAO Duoqian, et al. Sequential three-way decisions via multi-granularity[J]. *Information sciences*, 2020, 507: 606–629.
- [14] 李金海, 王飞, 吴伟志, 等. 基于粒计算的多粒度数据分析方法综述 [J]. *数据采集与处理*, 2021, 36(3): 418–435.
- LI Jinhai, WANG Fei, WU Weizhi, et al. Review of multi-granularity data analysis methods based on granular computing[J]. *Journal of data acquisition and processing*, 2021, 36(3): 418–435.
- [15] ZHANG Pengfei, LI Tianrui, LUO Chuan, et al. AMG-DTRS: adaptive multi-granulation decision-theoretic rough sets[J]. *International journal of approximate reasoning*, 2022, 140: 7–30.
- [16] 林国平, 梁吉业, 钱宇华. 基于多粒度视角下的 D-S 证据理论融合策略 [J]. *计算机科学*, 2014, 41(2): 45–48.
- LIN Guoping, LIANG Jiye, QIAN Yuhua. Multigranulation view based fusing strategy of D-S evidence[J]. *Computer science*, 2014, 41(2): 45–48.
- [17] CHE Xiaoya, MI Jusheng, CHEN Degang. Information fusion and numerical characterization of a multi-source information system[J]. *Knowledge-based systems*, 2018, 145: 121–133.
- [18] SANG Binbin, YANG Lei, CHEN Hongmei, et al. Generalized multi-granulation double-quantitative decision-theoretic rough set of multi-source information system[J]. *International journal of approximate reasoning*, 2019, 115: 157–179.
- [19] CHE Xiaoya, MI Jusheng. Attributes set reduction in multigranulation approximation space of a multi-source decision information system[J]. *International journal of machine learning and cybernetics*, 2019, 10(9): 2297–2311.
- [20] 万青, 魏玲, 任睿思. 协调多源决策表的规则提取 [J]. *南京大学学报(自然科学)*, 2020, 56(4): 494–504.
- WAN Qing, WEI Ling, REN Ruisi. Rule acquisition of consistent multi-source decision tables[J]. *Journal of Nanjing University (natural science edition)*, 2020, 56(4): 494–504.
- [21] YANG Lei, XU Weihua, ZHANG Xiaoyan, et al. Multi-granulation method for information fusion in multi-source decision information system[J]. *International journal of approximate reasoning*, 2020, 122: 47–65.
- [22] 骆公志, 陈佳馨. 多源覆盖信息系统下的加权广义多粒度粗糙集模型及其应用 [J]. *计算机工程与科学*, 2021, 43(12): 2231–2237.
- LUO Gongzhi, CHEN Jiabin. A weighted generalized multi-granulation rough set model of multi-source covering information system and its applications[J]. *Computer engineering & science*, 2021, 43(12): 2231–2237.
- [23] CHU Xiaoli, SUN Bingzhen, CHU Xiaodong, et al. Multi-granularity dominance rough concept attribute reduction over hybrid information systems and its application in clinical decision-making[J]. *Information sciences*, 2022, 597: 274–299.
- [24] PAWLAK Z. Rough sets[J]. *International journal of computer & information sciences*, 1982, 11(5): 341–356.
- [25] YAO Y Y, WONG S K M. A decision theoretic framework for approximating concepts[J]. *International journal*

of man-machine studies, 1992, 37(6): 793–809.

- [26] XU Weihua, LI Wentao, ZHANG Xiantao. Generalized multigranulation rough sets and optimal granularity selection[J]. *Granular computing*, 2017, 2(4): 271–288.
- [27] SUO Mingliang, TAO Laifa, ZHU Baolong, et al. Single-parameter decision-theoretic rough set[J]. *Information sciences*, 2020, 539: 49–80.

作者简介:



钱进, 教授, 博士, 主要研究方向为粒计算、大数据挖掘和机器学习。主持国家自然科学基金项目 2 项、省部级自然科学基金项目 3 项。获江西省自然科学奖 1 项。发表学术论文 50 余篇。



童志钢, 硕士研究生, 主要研究方向为粗糙集、粒计算和大数据挖掘。



苗夺谦, 教授, 博士, 国际粗糙集学会理事长、中国人工智能学会会士、中国计算机学会杰出会员, 主要研究方向为粒计算、不确定性、大数据分析。荣获中国人工智能学会吴文俊人工智能自然科学二等奖 1 项; 主持国家自然科学基金面上项目 7 项, 发表

学术论文 180 余篇, ESI 高被引 8 篇; 出版教材和学术著作 10 余部。

2023 年中国粒子计算与知识发现学术会议 (CGCKD 2023) Conference on China Granular Computing and Knowledge Discovery Society 2023 (CGCKD 2023)

由中国人工智能学会主办, 中国人工智能学会粒计算与知识发现专委会协办, 国际粗糙集学会支持, 厦门理工学院承办的 2023 年中国粒计算与知识发现学术会议(第二十三届中国 Rough 集与软计算学术会议、第十七届中国粒计算学术会议、第十一届三支决策学术会议)将于 2023 年 7 月 7~9 日在福建省厦门市召开。

联合会议由三部分组成: CRSSC 始于 2001 年, 主要研讨 Z. Pawlak 教授所提出 Rough 集理论; CGrC 于 2007 年加入, 主要研讨 L.A. Zadeh 和 T.Y. Lin 提出的粒计算理论以及张钹院士和张铃教授提出的商空间理论; 三支决策学术会议于 2012 年加入, 主要研讨姚一豫教授提出的三支决策理论。

征文范围 (包括但不限于):

1. 粗糙集与软计算: 粗糙集理论及应用、模糊集理论及应用、近似推理与不确定性分析、多准则决策分析、进化计算、神经计算、类脑计算、认知计算、云模型理论及其应用等。
2. 粒计算理论及其应用: 商空间理论及应用、粒逻辑与推理、形式概念分析、信息粒的表示、信息粒化、粒度选择、粒计算的模态分析与处理、多粒度推理模型、多粒度认知计算、多粒度智能决策、多粒度时空数据挖掘等。
3. 三支决策模型与分析: 三支决策与粗糙集、三支决策与区间集、三支决策与粒计算、三支决策与模糊集、三支决策与概念格、三支决策的数学模型、三支决策空间、三支聚类、动态三支决策、序贯三支决策等。
4. 知识发现与数据挖掘: 分类学习、数据聚类、关联和相关性、时间序列挖掘、知识演化、衰退分析、决策树、神经网络、遗传算法、模糊分类和聚类、粗糙分类、规则归纳、多粒度深度学习、多粒度知识发现和建模、可解释机器学习等。
5. 人工智能创新应用: 智慧医疗、智能制造、智慧城市、智慧交通、智慧海洋、多尺度媒体数据分析、元宇宙、多粒度深度学习等。会议采用网上征稿方式, 详情请浏览网址: <https://cs.xmut.edu.cn/cgckd2023/pager-contributions.html>。

重要日期:

论文投稿截止日期: 2023 年 3 月 1 日

论文录用通知日期: 2023 年 4 月 30 日

论文修改与终稿提交日期: 2023 年 5 月 20 日