

# 模糊粗糙集中基于条件熵的特定类属性约简

闵艳玲<sup>1</sup>, 杨霖琳<sup>1</sup>, 懂梦梦<sup>1</sup>, 张贤勇<sup>2</sup>

(1. 四川师范大学计算机科学学院, 成都 610101; 2. 四川师范大学数学科学学院, 成都 610066)

**摘要:** 传统的属性约简方法面向决策分类构建统一的属性约简集, 忽视了不同决策类间属性的差异性表征, 导致特定类的关键属性被冗余覆盖, 且特定类分类准确率不高等问题。针对以上问题, 考虑到模糊粗糙集在广泛存在的数值型和模糊性数据上的处理优势, 本文提出了一种条件熵驱动基于模糊粗糙集的特定类属性约简方法。首先, 通过融合模糊粗糙集的决策包含度与信息熵理论, 定义了面向特定决策类的条件熵, 量化了条件属性对目标类别的局部区分能力; 其次, 给出了基于条件熵的特定类属性约简条件, 并分别定义了基于特定类条件熵的内部和外部属性重要度, 进一步提出了基于属性重要度的特定类属性约简前向算法(Forward algorithm based on class-specific conditional entropy, FA-CE)和后向算法(Backward algorithm based on class-specific conditional entropy, BA-CE)。最后, 在UCI的7个数据集和特征选择数据的2个数据集上与邻域条件熵、互信息、邻域粗糙集和一种传统依赖度约简方法进行了特定类约简比较, 并在支持向量机(Support vector machine, SVM)、K最近邻(K-nearest neighbor, KNN)和分类与回归树(Classification and regression tree, CART)3种分类器上比较了特定类分类的准确率和 $F_1$ -score, 验证了本文提出的特定类属性约简方法的合理性与有效性。

**关键词:** 特定类属性约简; 模糊粗糙集; 条件熵; 属性约简; 决策包含度

**中图分类号:** TP18 **文献标志码:** A

**引用格式:** 闵艳玲, 杨霖琳, 懂梦梦, 等. 模糊粗糙集中基于条件熵的特定类属性约简[J]. 数据采集与处理, 2026, 41(3): 825-840. MIN Yanling, YANG Jilin, DONG Mengmeng, et al. Research on class-specific attribute reduction based on conditional entropy in fuzzy rough sets[J]. Journal of Data Acquisition and Processing, 2026, 41(3): 825-840.

## 引言

真实世界中的数据集通常包含大量无关紧要的信息。在处理和析数据时, 过多的属性或数据会增加处理时间并干扰数据分析结果。属性约简是知识发现与数据挖掘、模式识别等领域的重要概念<sup>[1-3]</sup>, 主要目的是在保持系统分类能力不变的前提下, 删除数据中的冗余属性, 从而实现数据降维, 并降低噪声对分类器的影响, 提升分类准确率<sup>[2-3]</sup>。

经典粗糙集主要针对符号型的数据, 在处理连续的数值型数据时需要将数据离散化, 这样可能造成信息丢失<sup>[4]</sup>。针对数值型数据, 经典粗糙集的许多扩展模型被提出<sup>[3-7]</sup>, 其中邻域粗糙集和模糊粗糙集是两个重要的模型。邻域粗糙集<sup>[5-6]</sup>通过对象之间的距离和一个邻域阈值来建立对象的邻域类, 但通常邻域阈值是根据一般经验或者多次实验来确定, 这将导致阈值具有主观性或增加计算成本和时间开销<sup>[6]</sup>。模糊粗糙集是通过对象的相似度直接建立模糊相似类而无需阈值, 避免了信息丢失, 具有更好的客观性且对不确定性和模糊性数据处理更友好<sup>[7]</sup>。

**基金项目:** 教育部人文社科规划基金(23YJA630114); 四川省自然科学基金(2024NSFSC0486, 2026NSFSC0444); 四川师范大学研究生创新能力培养项目(KY2025016)。

**收稿日期:** 2025-06-12; **修订日期:** 2025-08-15

信息熵是量化不确定性或随机性的一种数学工具<sup>[8]</sup>。在基于粗糙集理论的属性约简研究中,熵理论被用来刻画不确定性并进一步用来度量属性的重要度,从而建立一种启发式属性约简算法。Mi等<sup>[9]</sup>将信息熵引入模糊粗糙集,讨论了熵的基本性质。Dai等<sup>[10]</sup>针对区间值数据的属性约简问题,从信息论的角度提出了一种区间值数据属性约简的框架。Ji等<sup>[11]</sup>为解决现有属性约简算法未充分利用上近似信息可能导致的决策风险的问题,提出了一种基于融合信息熵的属性约简算法,有效利用了上近似中包含的决策信息和不确定信息。Liang等<sup>[12]</sup>提出相对信息熵和增强信息熵并基于这两种信息熵提出了两种新的约简方法,使得重要度的选择更加客观。Xie等<sup>[13]</sup>提出局部信息熵的概念,并利用距离计算模糊相似关系,解决了传统基于信息熵的方法存在计算效率低和过拟合等不足。目前大部分基于信息熵的属性约简研究主要针对决策分类来构建统一的属性约简。

然而,在实际应用中往往只需要考虑对某一个或某一些特定决策类的数据简化,即特定类的属性约简<sup>[14]</sup>,由于它关注不同决策类的差异性表征,因此可以更好地提取决策类的关键属性,从而使得约简率更高,并在多数情况下获得更高的分类准确率。因此,越来越多的研究者开始对特定类的属性约简展开研究<sup>[14-19]</sup>。Yao等<sup>[14]</sup>系统地比较了基于分类的约简、基于特定类的约简以及基于对象的约简,并考虑了它们与约简族之间的关系。Zhang等<sup>[15]</sup>基于粒计算、信息论和三支决策理论,提出了3种信息论类特定约简并建立了不同约简方法之间的横向和层次连接,为基于粒计算的属性约简提供了新的视角。Zhang等<sup>[16]</sup>提出基于邻域粗糙集的特定类属性约简并分析了它与邻域分类约简的层次关系。Luo等<sup>[17]</sup>针对邻域决策表,提出了一种基于邻域条件熵的特定类属性约简方法。现有基于特定类的属性约简方法能更准确地保留特定类的关键信息,从而提高特定决策类的分类性能。

模糊粗糙集和特定类约简在数值信息系统中有着广泛的应用。值得注意的是,模糊粗糙集理论中特定类属性依赖关系的量化仍存在不足,传统条件熵模型以全局优化为导向,难以捕捉局部类别的特征,制约多类不平衡数据的属性约简精度和分类准确性。因此,本文引入条件熵,基于模糊粗糙集特定类的决策包含度建立了面向特定决策类的条件熵。然后,进一步提出了属性约简条件并构造了两种约简算法:基于条件熵的特定类属性约简前向算法(Forward algorithm based on class-specific conditional entropy, FA-CE)和后向算法(Backward algorithm based on class-specific conditional entropy, BA-CE),并分析了面向特定决策类的约简属性和面向决策分类的约简属性之间的关系。最后通过对比实验比较了本文约简算法的约简精度、分类准确率及其 $F_1$ -score。实验验证了本文提出的约简方法对特定决策类属性约简的合理性和有效性。

## 1 基础理论

### 1.1 模糊粗糙集

设 $\langle U, A \cup D, V, f \rangle$ 是一个模糊决策信息系统(Fuzzy decision information system, FDIS),其中, $U = \{x_1, x_2, \dots, x_n\}$ 表示一个非空有限对象集, $A$ 和 $D$ 分别表示条件属性集和决策属性集,且 $A \cap D = \emptyset$ 。 $\pi_D = U/D = \{D_1, D_2, \dots, D_m\}$ 表示 $U$ 在 $D$ 下的决策划分。 $V = \{V_a | a \in A\}$ 是属性值集, $f: U \times A \rightarrow V$ 表示信息函数, $f(x, a) \in [0, 1]$ 是 $x$ 在属性 $a$ 上的属性值。

**定义 1** 在 $\langle U, A \cup D, V, f \rangle$ 中,对于任意对象 $x, y \in U$ , $x$ 和 $y$ 之间的相似度利用欧氏距离为

$$R_A(x, y) = 1 - \frac{1}{\sqrt{|A|}} \sqrt{\sum_{h=1}^{|A|} (f(x, a_h) - f(y, a_h))^2} \quad (1)$$

并且 $R_A(x, y) \in [0, 1]$ 。若 $R_A(x, y)$ 的值越接近1,则对象 $x$ 和 $y$ 越相似;若 $R_A(x, y)$ 的值越接近0,则对象 $x$ 和 $y$ 越不相似。

**定义 2**<sup>[20]</sup> 在  $\langle U, A \cup D, V, f \rangle$  中, 给定  $\forall x \in U, B$  为  $A$  的任意子集, 即  $B \subseteq A$ , 对象  $x$  在属性集  $B$  上的模糊相似类定义为

$$[x]_B = \frac{R_B(x, x_1)}{x_1} + \frac{R_B(x, x_2)}{x_2} + \dots + \frac{R_B(x, x_n)}{x_n} \quad (2)$$

值得注意的是, 模糊相似类  $[x]_B$  由对象  $x$  与所有对象的相似度来表示, 而非利用阈值进行简单的“相似”和“不相似”表示。

**定义 3**<sup>[20]</sup> 在  $\langle U, A \cup D, V, f \rangle$  中, 给定  $\forall x \in U, B \subseteq A, U/D = \{D_1, D_2, \dots, D_m\}$ , 对于任意  $D_j \in U/D, 1 \leq j \leq m$ , 对象  $x$  的模糊决策利用包含度定义为

$$\tilde{D}_j(x) = \frac{|[x]_B \cap D_j|}{|[x]_B|} \quad (3)$$

式中  $\tilde{D}_j(x)$  表示对象  $x$  关于决策类  $D_j$  的隶属程度。这里,  $\tilde{D}_j$  是一个模糊集,  $\forall x \in U$  有  $\sum_{j=1}^m \tilde{D}_j(x) = 1$ 。

**定义 4**<sup>[21]</sup> 在  $\langle U, A \cup D, V, f \rangle$  中,  $\forall x \in U, B \subseteq A, U/D = \{D_1, D_2, \dots, D_m\}$ , 对象  $x$  在条件属性集  $B$  下, 关于决策类  $D_j$  的模糊上下近似定义为

$$\begin{cases} \underline{B}D_j(x) = \inf_{y \in U} \max \{1 - R_B(x, y), \tilde{D}_j(y)\} \\ \overline{B}D_j(x) = \max_{y \in U} \inf \{R_B(x, y), \tilde{D}_j(y)\} \end{cases} \quad (4)$$

**定义 5**<sup>[14]</sup> 在  $\langle U, A \cup D, V, f \rangle$  中,  $\forall x \in U, B \subseteq A, U/D = \{D_1, D_2, \dots, D_m\}$ , 在属性集  $B$  下对象  $x$  关于决策类  $D_j$  的正域为

$$\text{POS}_B(D_j)(x) = \underline{B}D_j(x) \quad (5)$$

所有对象  $x_i (1 \leq i \leq n)$  在决策类  $D_j$  下的正域为

$$\text{POS}_B(D_j) = \sum_{i=1}^n \text{POS}_B(D_j)(x_i) \quad (6)$$

在决策分类  $\pi_D$  的正域为

$$\text{POS}_B(D) = \sum_{j=1}^m \text{POS}_B(D_j) \quad (7)$$

**定义 6**<sup>[7]</sup> 在  $\langle U, A \cup D, V, f \rangle$  中,  $D$  在属性集  $B$  下的依赖度为

$$\gamma_B(D) = \frac{\text{POS}_B(D)}{|U|} \quad (8)$$

**定义 7** 在  $\langle U, A \cup D, V, f \rangle$  中, 如果  $R$  满足以下条件, 那么属性子集  $R$  是  $A$  的约简: (1)  $\gamma_R(D) = \gamma_A(D)$ ; (2)  $\forall r \in R, \gamma_{R-\{r\}}(D) < \gamma_R(D)$ 。这里, 令  $\text{Red}(D)$  为决策分类的所有属性约简集合。

### 1.2 信息论度量

在一个信息系统中,  $B \subseteq A$  为任意条件属性子集。基于经典粗糙集, 论域  $U$  上的等价关系  $E_B$  定义为

$$E_B = \{(x, y) \in U \times U \mid \forall b \in B, I(x, b) = I(y, b)\} \quad (9)$$

根据  $E_B$ , 可以得到 1 个等价划分  $U/E_B = \{X_B^i \mid i = 1, 2, \dots, n\}$ 。根据决策属性  $D$ , 有等价关系  $E_D = \{D_j \mid j = 1, 2, \dots, m\}$ , 它表示存在  $m$  个特定的决策类。

**定义 8**<sup>[17]</sup> 在一个信息系统中, 决策分类  $\pi_D$  在属性子集  $B \subseteq A$  下的条件熵定义为

$$H(D|B) = - \sum_{i=1}^n p(X_B^i) \sum_{j=1}^m p(D_j|X_B^i) \log_{10}(p(D_j|X_B^i)) \quad (10)$$

式中:  $p(X_B^i) = \frac{|X_B^i|}{|U|}$ ,  $p(D_j|X_B^i) = \frac{|X_B^i \cap D_j|}{|X_B^i|}$ 。

条件熵  $H(D|B)$  表示在给定的属性子集  $B$  下  $D$  的不确定性。条件熵越高, 表示  $D$  的不确定性越大, 反之表示  $D$  的不确定性越小。基于条件熵可以定义关于决策分类  $\pi_D$  的约简。

**定义 9** 在一个信息系统中, 当且仅当满足以下条件时,  $R$  是  $A$  的决策分类属性约简: (1)  $H(D|R) = H(D|A)$ ; (2)  $\forall r \in R, H(D|R - \{r\}) \neq H(D|R)$ 。

## 2 基于条件熵的特定类属性约简

基于模糊粗糙集, 本节引入熵理论融合模糊决策包含度定义了面向特定类的条件熵, 提出了条件熵驱动的特定类属性约简条件。进一步地, 讨论了特定决策类和决策分类的属性约简关系。

### 2.1 特定类的条件熵

**定义 10** 设  $\langle U, A \cup D, V, f \rangle$  是一个模糊决策信息系统,  $\forall B \subseteq A, x_i \in U$ , 属性子集  $B$  在论域  $U$  上的信息熵可定义为

$$H(B) = - \sum_{i=1}^n p([x_i]_B) \log(p([x_i]_B)) \quad (11)$$

式中  $[x_i]_B$  为对象  $x_i$  在属性集  $B$  下的模糊相似类。概率  $p([x_i]_B)$  的定义为

$$p([x_i]_B) = \frac{\sum_{l=1}^n R_B(x_i, x_l)}{|U|} \quad (12)$$

式中:  $R_B(x_i, x_l)$  为对象  $x_i$  和  $x_l$  的模糊相似度,  $|U|$  为对象集合的个数,  $H(B)$  度量了论域  $U$  在属性子集  $B$  下的不确定性程度。

**定义 11** 在  $\langle U, A \cup D, V, f \rangle$  中, 对于  $\forall D_j \in \pi_D$ , 特定决策类  $D_j$  在论域  $U$  上的信息熵为

$$H(\tilde{D}_j) = - \sum_{i=1}^n p([x_i]_{D_j}) \log(p([x_i]_{D_j})) \quad (13)$$

式中:  $p([x_i]_{D_j}) = \frac{\sum_{l=1}^n R_{D_j}(x_i, x_l)}{|U|}$ ,  $R_{D_j}(x_i, x_l)$  为对象  $x_i$  和  $x_l$  在决策类  $D_j$  中的相似度, 若  $x_i \in D_j$  且  $x_l \in D_j$  则  $R_{D_j}(x_i, x_l) = 1$ ; 若  $x_i \in D_j$  且  $x_l \notin D_j$ , 则  $R_{D_j}(x_i, x_l) = 0$ 。若  $x_i \notin D_j$ , 则  $\forall x_l \in U$ , 有  $R_{D_j}(x_i, x_l) = 0$ 。 $H(D_j)$  度量了论域  $U$  在决策类  $D_j$  下的不确定性程度。

**定义 12** 在  $\langle U, A \cup D, V, f \rangle$  中, 有属性子集  $B \subseteq A$ , 在  $D_j$  上的条件熵定义为

$$H(\tilde{D}_j|B) = - \sum_{i=1}^n p([x_i]_B) p(D_j|[x_i]_B) \log(p(D_j|[x_i]_B)) \quad (14)$$

式中条件概率  $p(D_j|[x_i]_B) = \frac{|[x_i]_B \cap D_j|}{|[x_i]_B|}$ 。

根据式(3)和式(13), 条件熵可以简化为

$$H(\tilde{D}_j|B) = - \sum_{i=1}^n \frac{|[x_i]_B \cap D_j|}{|U|} \log \tilde{D}_j(x_i) \quad (15)$$

式中  $\tilde{D}_j(x_i)$  为对象  $x_i$  在  $D_j$  下的模糊决策包含度。这里  $H(\tilde{D}_j|B) \geq 0$ , 它衡量了论域  $U$  在属性子集  $B$  下关于决策类  $D_j$  的不确定性。 $H(\tilde{D}_j|B)$  越小, 表明所有对象关于特定决策类  $D_j$ , 在属性子集  $B$  下的不确定性越小, 即属性子集  $B$  针对决策类  $D_j$  提供了较多的相关信息。反之,  $H(\tilde{D}_j|B)$  越大, 表明所有对象关于特定决策类  $D_j$ , 在属性子集  $B$  下的不确定性越大, 即属性子集  $B$  针对决策类  $D_j$  提供了较少的相关信息。

根据式 (15), 基于决策包含度的决策类条件熵有如下性质: (1) 非负性:  $\forall \tilde{D}_j(x_i) \in (0, 1]$ ,  $\log \tilde{D}_j(x_i) \leq 0$ , 则  $H(\tilde{D}_j|B) \geq 0$  恒成立; (2) 局部单减: 设  $h_i = -\frac{|[x_i]_B \cap D_j|}{|U|} \log \tilde{D}_j(x_i)$ , 在  $\frac{|[x_i]_B \cap D_j|}{|U|}$  相同的情况下,  $h_i$  值随  $\tilde{D}_j(x_i)$  的增大而减小。证明如下:

(1) 非负性: 因为  $\forall \tilde{D}_j(x_i) \in (0, 1]$ ,  $\log \tilde{D}_j(x_i) \leq 0$ , 所以  $-\log \tilde{D}_j(x_i) \geq 0$ ; 又因为  $\frac{|[x_i]_B \cap D_j|}{|U|} \geq 0$ , 所以每一项都非负, 则求和也为非负, 所以  $H(\tilde{D}_j|B) \geq 0$  恒成立。

(2) 局部单减: 在条件属性子集  $B$  相同的情况下,  $\forall x_m, x_i \in U$ , 可能有  $\frac{|[x_m]_B \cap D_j|}{|U|} = \frac{|[x_i]_B \cap D_j|}{|U|}$ 。所以设  $\omega_i = \frac{|[x_i]_B \cap D_j|}{|U|}$ ,  $d_i = \tilde{D}_j(x_i)$ , 于是  $h_i$  可简写成  $h_i = -\omega_i \log d_i$ 。设  $f(d_i) = -\log d_i$ , 通过求导可得  $f'(d_i) = -\frac{1}{d_i}$ , 因为  $d_i \in (0, 1]$ , 所以  $f'(d_i) < 0$ , 因此  $f(d_i)$  单调递减, 所以随着  $d_i$  增大  $f(d_i)$  减小, 所以当  $\omega_i = \frac{|[x_i]_B \cap D_j|}{|U|}$  相同时,  $h_i$  值随  $\tilde{D}_j(x_i)$  的增大而减小。证毕。

进一步地, 在不同条件属性子集下, 决策类条件熵有如下性质。

**性质 1** 在  $\langle U, A \cup D, V, f \rangle$  中,  $\forall D_j \in \pi_D$ , 对于  $\forall B_1, B_2 \subseteq A$ , 且  $B_1 \subseteq B_2$ , 有

$$H(\tilde{D}_j|B_2) \leq H(\tilde{D}_j|B_1) \tag{16}$$

根据性质 1, 将属性添加到当前初始属性子集中至少不会增加熵值。因此, 对于  $B \subseteq A$ , 条件属性集  $A$  有最小条件熵, 即

$$H(\tilde{D}_j|A) \leq H(\tilde{D}_j|B) \tag{17}$$

一般而言, 属性约简的目标是从条件属性集  $A$  中剔除冗余属性, 且找到与属性集  $A$  在分类时具有相同表征能力的属性子集。在一个模糊决策信息系统中, 特定决策类的属性约简目标是找到最小的属性子集并且该子集在特定的决策类  $D_j$  中具有与  $A$  相同的信息表征能力。因此, 根据上述特性, 可以利用条件熵来构建特定类的属性约简条件, 具体如下。

**定义 13** 在  $\langle U, A \cup D, V, f \rangle$  中, 假设  $R \subseteq A$  是决策类  $D_j$  的一个约简, 当且仅当满足以下两个条件: (1)  $H(\tilde{D}_j|R) = H(\tilde{D}_j|A)$ , (2)  $\forall r \in R, H(\tilde{D}_j|R - \{r\}) > H(\tilde{D}_j|R)$ 。

令  $\text{Red}_h(\tilde{D}_j)$  为基于条件熵的特定类  $D_j$  的所有约简的集合。条件 (1) 保证了所选的属性子集与初始属性集  $A$  具有相同的信息量, 而条件 (2) 排除了约简集  $R$  中的冗余属性。换句话说, 可以通过上述条件判定  $A$  的子集是否为特定类的约简, 该子集具有与初始属性集相同的并且是最小的条件熵。

## 2.2 特定类和分类之间的属性约简分析

本节将通过一个例子来计算基于条件熵的特定类属性约简, 进一步讨论面向决策分类的属性约简与特定决策类的属性约简之间的关系。

**例 1** 给定 1 个模糊决策表  $\langle U, A \cup D, V, f \rangle$  如表 1 所示, 其中,  $U = \{x_1, x_2, \dots, x_8\}$ ,  $A = \{a_1, a_2, a_3, a_4\}$ , 并且  $\forall a \in A$ , 有  $f(x, a) \in [0, 1]$ ,  $D = \{d\}$ ,  $\pi_D = U/D = \{D_1, D_2, D_3\}$ 。根据式 (1) 计算所

有对象在属性A下对象的相似度如表2所示。

表1 模糊决策信息系统  $\Omega = \langle U, A \cup D, V, f \rangle$

Table 1 FDIS  $\Omega = \langle U, A \cup D, V, f \rangle$

$U$	$a_1$	$a_2$	$a_3$	$a_4$	$d$
$x_1$	0.9	0.9	0.2	0.1	1
$x_2$	0.8	0.7	0.3	0.4	1
$x_3$	0.2	0.9	0.2	0.2	1
$x_4$	0.3	0.9	0.1	0.4	2
$x_5$	0.4	0.1	0.9	0.5	2
$x_6$	0.4	0.3	0.8	0.9	2
$x_7$	0.9	0.5	0.1	0.8	3
$x_8$	0.8	0.4	0.1	0.8	3

表2 相似度表

Table 2 Similarity table

$U$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$
$x_1$	1							
$x_2$	0.81	1						
$x_3$	0.65	0.66	1					
$x_4$	0.66	0.71	0.88	1				
$x_5$	0.38	0.53	0.44	0.43	1			
$x_6$	0.37	0.55	0.44	0.47	0.77	1		
$x_7$	0.59	0.75	0.50	0.59	0.47	0.56	1	
$x_8$	0.56	0.73	0.51	0.59	0.51	0.59	0.93	1

根据式(2)以  $x_1$  为例,其模糊相似类可表示为:  $[x_1] = \frac{1}{x_1} + \frac{0.81}{x_2} + \frac{0.65}{x_3} + \frac{0.66}{x_4} + \frac{0.38}{x_5} + \frac{0.37}{x_6} + \frac{0.59}{x_7} + \frac{0.56}{x_8}$ 。

然后利用式(3)可计算所有对象对于每个决策类的模糊决策包含度矩阵,有

$$\tilde{D} = \begin{bmatrix} \tilde{D}_1 \\ \tilde{D}_2 \\ \tilde{D}_3 \end{bmatrix} = \begin{bmatrix} 0.49 & 0.43 & 0.45 & 0.42 & 0.30 & 0.29 & 0.34 & 0.33 \\ 0.28 & 0.31 & 0.35 & 0.36 & 0.49 & 0.47 & 0.30 & 0.31 \\ 0.23 & 0.26 & 0.20 & 0.22 & 0.22 & 0.24 & 0.36 & 0.36 \end{bmatrix}_{3 \times 8}$$

根据式(8)和定义7计算决策分类D的属性约简如表3所示,即为  $\{a_1, a_3, a_4\}$ 。利用特定决策类的包含度  $\tilde{D}_j(x_i)$  和式(15),可计算得到  $H(\tilde{D}_1|A) = 0.81, H(\tilde{D}_1|\{a_2, a_3, a_4\}) = H(\tilde{D}_1|\{a_1, a_3, a_4\}) = H(\tilde{D}_1|\{a_1, a_2, a_4\}) = 0.81, H(\tilde{D}_1|\{a_3, a_4\}) = H(\tilde{D}_1|\{a_2, a_4\}) = 0.81, H(\tilde{D}_1|\{a_1, a_2, a_3\}) = H(\tilde{D}_1|\{a_2, a_3\}) = 0.82, H(\tilde{D}_1|\{a_1, a_4\}) = H(\tilde{D}_1|\{a_1, a_2\}) = H(\tilde{D}_1|\{a_2\}) = 0.83, H(\tilde{D}_1|\{a_1, a_3\}) = H(\tilde{D}_1|\{a_4\}) = 0.84, H(\tilde{D}_1|\{a_3\}) = 0.87, H(\tilde{D}_1|\{a_1\}) = 0.89$ 。

根据定义13,可以得到  $\text{Red}_h(\tilde{D}_1) = \{\{a_3, a_4\}, \{a_2, a_4\}\}$ 。同样地可以得到  $\text{Red}_h(\tilde{D}_2)$  和  $\text{Red}_h(\tilde{D}_3)$  的属性约简集如表3所示。通过观察表3,有特定决策类  $D_2$  和  $D_3$  的属性约简与决策分类  $\pi_D$  的属性约简存在完全包含的关系即  $\text{Red}_h(\tilde{D}_j) \subseteq \text{Red}(D)$ , 而特定决策类  $D_1$  与决策分类  $D$  却不存在该关系。实际上进一步分析两种属性约简,两者的关系应存在以下可能:(1)独立关系:  $\text{Red}_h(\tilde{D}_j) \not\subseteq \text{Red}(D)$ ; (2)部分包含:  $\text{Red}_h(\tilde{D}_j) \cap \text{Red}(D) \neq \emptyset$ ; (3)完全包含:  $\text{Red}_h(\tilde{D}_j) \subseteq \text{Red}(D)$ 。在例1中,  $\text{Red}_h(\tilde{D}_1)$  与  $\text{Red}(D)$  满足部分包含,说明条件属性  $a_4$  是局部  $D_1$  和全局  $D$  的公共核心属性,而  $a_2$  只对  $D_1$  具有局部重要性。在全局约简中的  $a_1$  和  $a_4$  则分别对决策类  $D_3$  和  $D_2$  是冗余属性。

表3 分类和特定类的属性约简

Table 3 Attribute reducts of classification and class-specific

$U$	属性约简
$D$	$\{a_1, a_3, a_4\}$
$D_1$	$\{a_3, a_4\}, \{a_2, a_4\}$
$D_2$	$\{a_1, a_3\}$
$D_3$	$\{a_3, a_4\}$

### 3 基于条件熵的特定类前后向约简算法

在基于条件熵的特定类属性约简中,条件熵可以客观度量任意一个属性子集对于某一特定类的不确定性。类似于决策分类的属性约简,针对某一个特定决策类,本文利用条件熵分别定义了内部和外部属性重要度的概念,进一步提出了基于模糊粗糙集的特定类的前向和后向属性约简算法。

**定义 14** 在 $\langle U, A \cup D, V, f \rangle$ 中,对于 $B \subseteq A$ ,属性 $b \in B$ ,属性 $b$ 的内部重要度定义为

$$\text{Sig}^{\text{inter}}(b, B, \tilde{D}_j) = H(\tilde{D}_j|B - \{b\}) - H(\tilde{D}_j|B) \quad (18)$$

定义 14 中,属性 $b$ 的重要度通过从属性集 $B$ 中移除 $b$ 时条件熵的变化来评估,当条件熵变化越大,则表明该属性的重要性越大;反之,则该属性重要度越小,若内部重要度为零,则说明 $b$ 是冗余的。基于属性的内部重要度可设计特定类的后向属性约简启发式算法,如算法 1 所示。

**算法 1 基于条件熵的特定类后向属性约简算法**

输入:模糊决策信息系统  $\Omega = \langle U, A \cup D, V, f \rangle$

输出:约简集  $R$

- (1)  $A \rightarrow R$
- (2) 通过式(15)计算初始条件熵  $H_0(D_j|A)$
- (3) repeat
- (4) for each  $a \in A$  do
- (5) 计算  $H(D_j|R - \{a\})$
- (6) 通过式(18)计算  $\text{Sig}^{\text{inter}}(a, R, D_j) = H(D_j|R - \{a\}) - H(D_j|R)$
- (7) end for
- (8) 选择  $a_{\min} = \text{argminSig}^{\text{inter}}(a, R, D_j)$
- (9)  $R = R - a_{\min}$
- (10) 计算  $H(D_j|R)$
- (11) 直到  $H(D_j|R) = H_0(D_j|A)$
- (12) 返回  $R$

**定义 15** 在 $\langle U, A \cup D, V, f \rangle$ 中,对于 $B \subseteq A$ ,属性 $b \in B$ ,属性 $b$ 的外部重要度定义为

$$\text{Sig}^{\text{exter}}(b, B, \tilde{D}_j) = H(\tilde{D}_j|B) - H(\tilde{D}_j|B \cup \{b\}) \quad (19)$$

定义 15 中,当属性 $b$ 添加到属性集 $B$ 时,条件熵可能会发生变化,因此可以用来评估属性 $b$ 的重要度。从空集开始,迭代的添加属性,通过嵌套循环计算属性重要度来获得属性约简。具体过程如算法 2 所示。

**算法 2 基于条件熵的特定类前向属性约简算法**

输入:模糊决策信息系统  $\Omega = \langle U, A \cup D, V, f \rangle$

输出:约简集  $R$

- (1)  $R = \emptyset$
- (2) 通过式(15)计算初始条件熵  $H_0(D_j|A)$
- (3) repeat

- (4) for each  $a \in A - R$  do
- (5)     计算  $H(D_j | R \cup \{a\})$
- (6)     通过式(19)计算  $\text{Sig}^{\text{exter}}(a, R, D_j) = H(D_j | R) - H(D_j | R \cup \{a\})$
- (7)     end for
- (8) 选择  $a_{\max} = \text{argmax} \text{Sig}^{\text{exter}}(a, R, D_j)$
- (9) if  $\text{Sig}^{\text{exter}}(a, R, D_j) > 0$
- (10)     $R = R \cup \{a_{\max}\}$
- (11) else
- (12) break
- (13) 返回  $R$

算法1中内部重要度是衡量在删除某一属性时这一属性在当前属性集内部的重要程度,在后向属性约简中每次优先选择删除内部重要度最小的属性;算法2中外部重要度是从当前属性子集外部选择某一属性时,需要衡量外部重要度,通常选择外部重要度最大的属性添加到当前属性集。

算法1的条件熵计算时间复杂度为  $O(|U|^2 \cdot |A|)$ ,重要度计算时间复杂度为  $O(1)$ ,两个循环次数为  $O(|A|^2)$ ,算法1的总时间复杂度为  $O(|U|^2 \cdot |A|^3)$ ;存储数据  $O(|A| \cdot |U|)$ ,相似度矩阵  $O(|U|^2)$ ,所以总的空间复杂度为  $O(|A| \cdot |U| + |U|^2)$ 。算法2与算法1在时间与空间复杂度上相同,在实际运用中算法1比算法2更慢,因为算法1需要遍历所有属性。在面对属性数较大且对象数适中的数据集时,两种算法都是可行的。

## 4 实验与分析

为了验证提出的基于条件熵的特定类FA-CE和BA-CE算法的有效性,本节从UCI数据集官网中下载了7个数据集(序号1~7),从特征选择(<https://jundongli.github.io/scikit-feature/datasets.html>)中下载了2个数据集(序号8~9),其中序号7~9为高维数据集,其余为低维数据集。数据集描述如表4所示(Yale和WarpAR10P数据集选择了其中3个决策类)。进行比较的模型如下:(1)CFARNRS<sup>[18]</sup>,3层属性约简中基于邻域粗糙集的特定类前向属性约简方法;(2)NCECAR<sup>[17]</sup>,基于邻域条件熵的特定类属性约简方法;(3)FRS,基于模糊粗糙集传统依赖度的特定类约简方法;(4)CSRRWFS<sup>[22]</sup>,基于相关性-冗余权重的邻域互信息的特定类的特征选择方法,分别在KNN(最近邻),SVM(支持向量机),CART(分类与回归树)上比较了这几个约简算法的分类准确率及 $F_1$ -score。KNN分类器邻居数 $K$ 值为3。在CFARNRS算法中,9个数据集的邻域半径都设置为0.3;在NCECAR算法中,Climate数据集的邻域半径为0.4,其余数据集的邻域半径设为0.2;在CSRRWFS算法中,所有数据集邻域半径均取0.2。

表4 数据集描述

Table 4 Description of datasets

序号	数据集	条件属性	样本	决策类
1	Breast cancer	9	116	2
2	Climate	19	540	3
3	Heart failure records	12	299	2
4	Ionosphere	34	351	2
5	Parkinsons	23	195	3
6	Wine	13	178	3
7	Toxicity	1 203	171	2
8	Yale	1 024	166	3
9	WarpAR10P	2 400	130	3

模拟实验在 Pycharm 环境下进行,该代码采用 Python 编写,并在配备 Intel(R) Core(TM) i5-8265U CPU @ 1.80 GHz、内存 8 GB 的 PC 机上执行。

### 4.1 属性约简率分析

为了验证特定类属性约简的约简效率,首先,在 9 个数据集中对决策分类  $\pi_D$  用基于信息熵的约简方法进行约简,以获取  $Red(D)$ ;然后,对数据集中每个决策类分别用 BA-CE 和 FA-CE 算法进行特定类属性约简。对于任意的  $D_j \in \pi_D$ ,约简率公式为

$$Reduction\ Rate(D_j) = \frac{|A| - |Red(D_j)|}{|A|} \times 100\% \quad (20)$$

约简率比较如图 1 所示,观察 9 个数据集不难发现有 8 个数据集特定决策类的属性约简率都高于或等于决策分类的属性约简,这也验证了全局约简对局部约简有冗余覆盖。只有 Ionosphere 数据集的  $D_1$

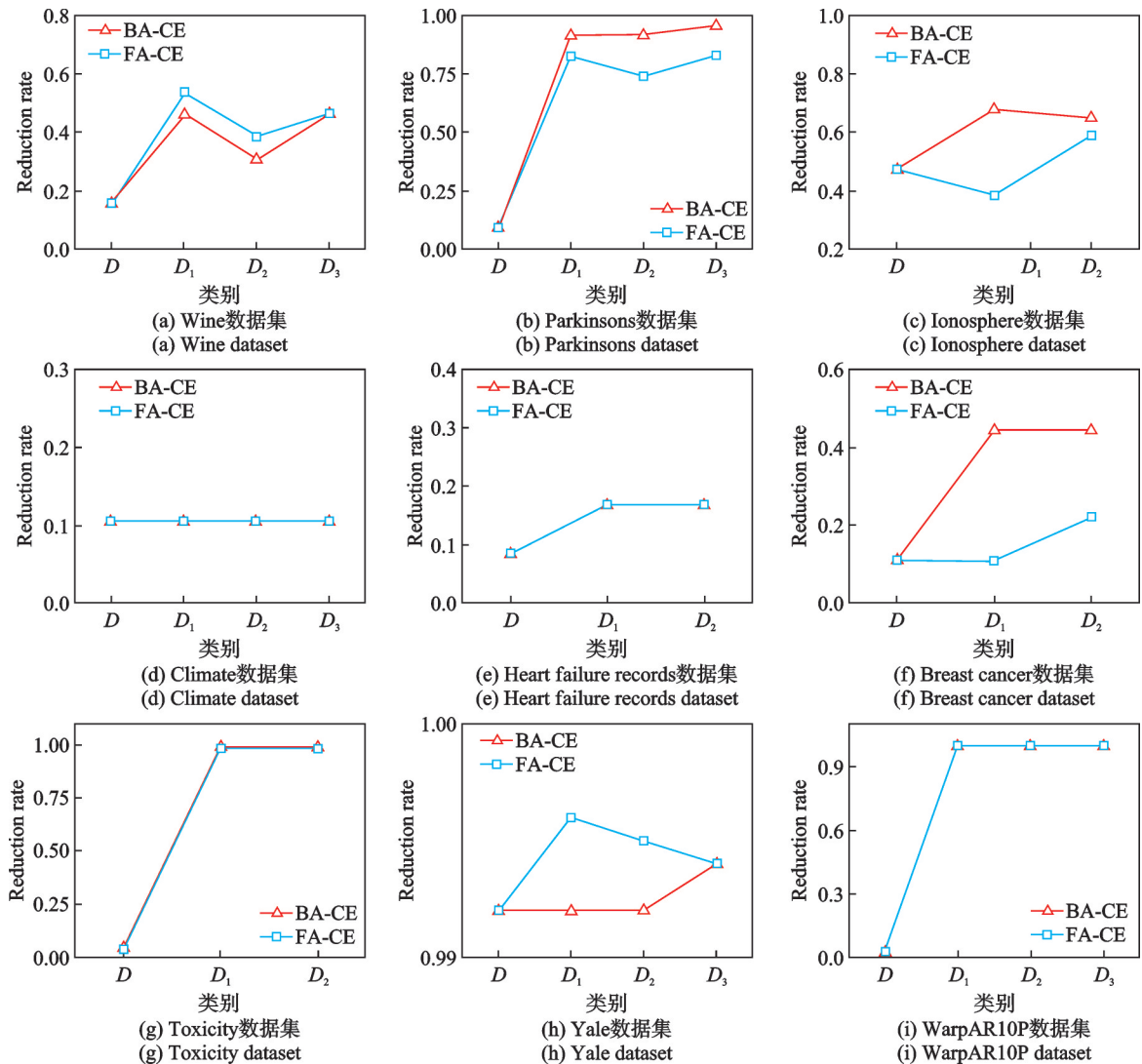


图 1 约简率比较

Fig.1 Reduction rate comparison

在FA-CE上略低于 $D$ 的约简率,其原因是为了局部表达更准确,特定决策类 $D_1$ 约简保留了更多属性。

以Breast cancer数据集为例,后向约简有 $R_h(D_1)=\{a_1, a_2, a_6, a_7, a_9\}$ , $R_h(D_2)=\{a_1, a_2, a_7, a_8, a_9\}$ ;前向约简有 $R_h(D_1)=\{a_1, a_2, a_3, a_4, a_6, a_7, a_8, a_9\}$ , $R_h(D_2)=\{a_1, a_2, a_3, a_6, a_7, a_8, a_9\}$ ;而 $\text{Red}(D)=\{a_1, a_2, a_4, a_5, a_6, a_7, a_8, a_9\}$ 。特定决策类 $D_1$ 的约简率分别为44.4%和11.1%,决策类 $D_2$ 的约简率分别为44.4%和22.2%,对比决策分类 $\pi_D$ 分别提升了33.3%和11.1%。这一实验结果验证了从局部特征信息出发进行特定类约简的BA-CE和FA-CE的约简有效性。

## 4.2 准确率数值分析

为了验证BA-CE算法和FA-CE算法的约简正确性,本节选用3种常见的分类器,即KNN、SVM和CART。在9个数据集上进行分类准确率的对比,并采用十折交叉验证,将数据集划成10份大小相等不重叠的数据集进行10轮训练和测试。表5~7展示了未约简(Raw)情况下和5种不同的约简算法分别在KNN、SVM和CART上的性能对比结果,其中黑色加粗加下划线字体表示最高准确率,黑色加粗字体表示次高准确率。在9个数据集中共有23个特定决策类。在表5中,FA-CE算法在6个数据集一共

表5 KNN分类器上的准确率  
Table 5 Accuracy of KNN classifier

数据集	$D_j$	Raw	CFARNRS	NCECAR	FRS	CSRRWFS	BA-CE	FA-CE
Wine	$D_1$	0.905	0.905	0.905	0.905	0.893	<u>0.911</u>	0.910
	$D_2$	0.778	<b>0.944</b>	0.725	0.777	0.762	0.782	<u>0.983</u>
	$D_3$	0.760	<b>0.966</b>	0.742	0.760	0.797	<b>0.978</b>	<u>0.978</u>
Parkinsons	$D_1$	0.964	0.970	<b>0.980</b>	0.950	0.959	<u>0.985</u>	0.970
	$D_2$	0.953	0.969	<u>1.000</u>	<b>0.974</b>	0.969	<u>1.000</u>	0.964
	$D_3$	0.969	<u>0.985</u>	0.959	<u>0.985</u>	0.964	<b>0.975</b>	<b>0.975</b>
Ionosphere	$D_1$	0.849	<u>0.863</u>	0.846	<b>0.852</b>	0.846	<u>0.863</u>	<b>0.852</b>
	$D_2$	0.849	<b>0.863</b>	0.846	0.846	0.829	<u>0.906</u>	0.849
Heart failure records	$D_1$	0.592	0.626	0.626	0.626	0.627	<b>0.632</b>	<u>0.652</u>
	$D_2$	0.592	0.592	0.626	0.626	0.611	<b>0.629</b>	<u>0.646</u>
Climate	$D_1$	0.572	0.554	0.528	0.559	<b>0.571</b>	<u>0.604</u>	<u>0.604</u>
	$D_2$	0.556	0.544	<b>0.563</b>	0.554	0.562	<u>0.580</u>	<u>0.580</u>
	$D_3$	0.952	0.950	0.950	<b>0.956</b>	<b>0.954</b>	0.952	0.950
Breast cancer	$D_1$	0.527	0.492	0.477	<b>0.527</b>	0.522	<u>0.537</u>	<b>0.527</b>
	$D_2$	0.527	<b>0.579</b>	0.477	0.527	0.530	<u>0.590</u>	0.519
Toxicity	$D_1$	0.538	0.603	0.614	<b>0.627</b>	<u>0.641</u>	0.557	0.620
	$D_2$	0.538	0.591	<b>0.614</b>	0.601	0.612	0.578	<u>0.637</u>
Yale	$D_1$	0.875	0.808	0.725	0.708	0.717	<b>0.817</b>	<u>0.858</u>
	$D_2$	0.800	0.908	0.683	0.800	0.867	<b>0.933</b>	<u>0.967</u>
	$D_3$	0.833	0.800	<b>0.875</b>	0.775	0.783	0.833	<u>0.933</u>
WarpAR10P	$D_1$	0.675	0.875	0.742	0.617	0.600	<b>0.892</b>	<u>0.925</u>
	$D_2$	0.850	0.925	0.867	0.875	0.733	<u>1.000</u>	<b>0.950</b>
	$D_3$	0.850	0.900	0.892	<b>0.917</b>	0.683	0.892	<u>1.000</u>

12个决策类取得最高准确率,在5个数据集一共5个决策类上取得次高准确率,在Breast cancer的 $D_2$ 上比Raw略低0.8%,其原因为Breast cancer数据集的特征数过少,强行约简后导致部分有效信息丢失。BA-CE算法在6个数据集11个决策类达到最高准确率,在4个数据集6个决策类上达到次高准确率,在Toxicity高维数据集上的准确率比其他约简算法低,因为后向约简在高维数据集中可能导致误删。在表6中FA-CE算法在8个数据集13个决策类上达到最优,在6个数据集9个决策类上达到次优,在Parkinsons数据集集中的 $D_1$ 略低于CSRRWFS算法0.4%的分类准确率。BA-CE算法在8个数据集12个决策类上达到最高准确率,在7个数据集9个决策类上取得次高准确率。从表7可以发现,FA-CE算法在7个数据集14个决策类上取得最高准确率,在3个数据集3个决策类上取得次高准确率,BA-CE算法在7个数据集8个决策类上达到最高准确率,在6个数据集8个决策类中准确率次高。

综合表5~7的分析结果可以得出,本文提出的FA-CE和BA-CE算法取得最高准确率的特定决策类个数明显大于其他算法,也就是说,本文提出的两种算法对于大部分决策类数据在高约简率的情况下,分类仍有一定程度的提升。

表6 SVM分类器上的准确率  
Table 6 Accuracy of SVM classifier

数据集	$D_j$	Raw	CFARNRS	NCECAR	FRS	CSRRWFS	BA-CE	FA-CE
Wine	$D_1$	0.927	0.915	0.915	0.921	0.922	<b>0.927</b>	<b>0.928</b>
	$D_2$	0.759	0.771	<b>0.776</b>	0.771	0.762	<b>0.774</b>	0.770
	$D_3$	0.731	<b>0.978</b>	0.730	0.730	0.728	<b>0.994</b>	<b>0.994</b>
Parkinsons	$D_1$	0.970	<b>0.970</b>	<b>0.970</b>	<b>0.970</b>	<b>0.974</b>	0.970	0.970
	$D_2$	0.969	0.969	0.969	0.969	0.969	<b>0.970</b>	<b>0.970</b>
	$D_3$	0.969	0.969	0.969	0.969	0.969	<b>0.980</b>	0.970
Ionosphere	$D_1$	0.929	0.909	0.835	<b>0.943</b>	0.937	0.932	<b>0.952</b>
	$D_2$	0.929	0.929	0.835	0.932	0.931	<b>0.949</b>	<b>0.940</b>
Heart failure records	$D_1$	0.679	<b>0.679</b>	<b>0.679</b>	<b>0.679</b>	<b>0.681</b>	0.679	0.679
	$D_2$	0.679	<b>0.679</b>	<b>0.679</b>	<b>0.679</b>	<b>0.681</b>	0.679	0.679
Climate	$D_1$	0.643	<b>0.667</b>	0.656	0.654	<b>0.662</b>	<b>0.667</b>	<b>0.667</b>
	$D_2$	0.637	<b>0.667</b>	<b>0.665</b>	0.639	0.657	<b>0.667</b>	<b>0.667</b>
	$D_3$	0.959	<b>0.959</b>	0.667	<b>0.959</b>	<b>0.959</b>	<b>0.959</b>	0.959
Breast cancer	$D_1$	0.529	0.529	0.529	0.529	<b>0.533</b>	<b>0.542</b>	<b>0.542</b>
	$D_2$	0.529	0.529	0.529	0.529	0.533	0.541	<b>0.554</b>
Toxicity	$D_1$	0.649	<b>0.673</b>	<b>0.673</b>	<b>0.673</b>	<b>0.665</b>	<b>0.673</b>	<b>0.673</b>
	$D_2$	0.649	<b>0.667</b>	<b>0.673</b>	<b>0.673</b>	0.665	<b>0.673</b>	<b>0.673</b>
Yale	$D_1$	0.900	0.775	0.608	0.667	<b>0.867</b>	<b>0.867</b>	0.825
	$D_2$	0.805	0.875	0.583	0.708	0.808	0.908	<b>0.917</b>
	$D_3$	0.683	0.767	0.708	0.717	0.717	<b>0.917</b>	0.900
WarpAR10P	$D_1$	0.646	<b>0.900</b>	0.825	0.642	0.692	0.808	<b>0.950</b>
	$D_2$	0.868	0.850	0.867	0.850	0.867	<b>0.967</b>	0.925
	$D_3$	0.875	<b>0.900</b>	0.892	0.892	0.817	<b>0.900</b>	<b>1.000</b>

表7 CART分类器上的准确率  
Table 7 Accuracy of CART classifier

数据集	$D_j$	Raw	CFARNRS	NCECAR	FRS	CSRRWFS	BA-CE	FA-CE
Wine	$D_1$	0.933	0.937	0.938	<b>0.950</b>	0.938	0.938	<u>0.955</u>
	$D_2$	0.916	0.909	0.899	0.911	0.927	<u>0.932</u>	0.928
	$D_3$	0.922	<b>0.961</b>	0.916	0.933	0.921	<b>0.961</b>	<u>0.972</u>
Parkinsons	$D_1$	0.979	<b>0.980</b>	0.960	0.959	<b>0.980</b>	0.979	<u>0.985</u>
	$D_2$	0.995	<b>0.974</b>	<u>1.00</u>	0.959	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>
	$D_3$	0.969	<u>0.970</u>	<b>0.969</b>	<u>0.970</u>	0.964	<b>0.969</b>	<u>0.970</u>
Ionosphere	$D_1$	0.886	<b>0.886</b>	0.838	0.878	0.883	0.849	<u>0.892</u>
	$D_2$	0.892	0.863	0.829	<b>0.895</b>	0.846	<u>0.900</u>	0.889
Heart failure records	$D_1$	0.766	<u>0.796</u>	0.783	0.780	0.641	0.786	<b>0.792</b>
	$D_2$	0.780	0.760	<b>0.790</b>	0.780	0.786	<u>0.810</u>	0.786
Climate	$D_1$	0.563	0.457	0.469	<b>0.574</b>	0.568	<u>0.611</u>	0.572
	$D_2$	0.561	0.507	0.511	0.569	0.568	<u>0.604</u>	<b>0.583</b>
	$D_3$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Breast cancer	$D_1$	0.698	0.681	0.583	0.680	0.696	<b>0.699</b>	<u>0.769</u>
	$D_2$	0.681	0.683	0.574	0.698	0.678	<b>0.733</b>	<u>0.789</u>
Toxicity	$D_1$	0.510	0.522	0.520	0.469	0.529	<b>0.544</b>	<u>0.597</u>
	$D_2$	0.509	0.527	0.527	<b>0.579</b>	0.535	<u>0.585</u>	0.520
Yale	$D_1$	0.783	<u>0.842</u>	0.808	0.775	0.786	<b>0.817</b>	0.742
	$D_2$	0.843	<b>0.900</b>	0.850	0.725	0.781	0.742	<u>0.917</u>
	$D_3$	0.775	0.817	0.783	0.767	0.657	<b>0.900</b>	<u>0.933</u>
WarpAR10P	$D_1$	0.750	0.750	0.750	0.692	0.761	<b>0.800</b>	<u>0.850</u>
	$D_2$	0.875	<u>0.900</u>	<u>0.900</u>	0.850	<b>0.893</b>	<u>0.900</u>	<u>0.900</u>
	$D_3$	0.950	0.792	0.867	<b>0.892</b>	0.836	0.817	<u>0.950</u>

### 4.3 $F_1$ -score数值分析

为了综合评估这5种约简算法和更全面反映各个算法的实际表现,本文分别验证并比较了5种算法在这3个分类器下的 $F_1$ -score,如图2~4所示。通过观察图2,在KNN分类器下,FA-CE算法在5个数据集中的决策类达到最高的 $F_1$ -score值,在剩下4个数据集中有2个都处于次优水平。BA-CE算法在7个数据集上达到最高的 $F_1$ -score值,在Yale数据集上效果不太理想。其原因是在高维少样本数据集上,基于BA-CE的约简在面对KNN分类器时,容易误删对局部结构有重要贡献的特征,导致了对邻域结构的破坏,从而使得少数类样本的识别能力下降,因此 $F_1$ -score明显降低。同样,观察图3和图4可以发现BA-CE和FA-CE算法在另外两种分类器上的 $F_1$ -score普遍高于其他对比算法。综上,其实验结果可以验证本文提出的约简算法的合理性和有效性。并且通过实验分析可得出,BA-CE算法更适合约简属性冗余度低以及中等及以下维度的数据集,FA-CE算法则在大部分数据场景下都适用。

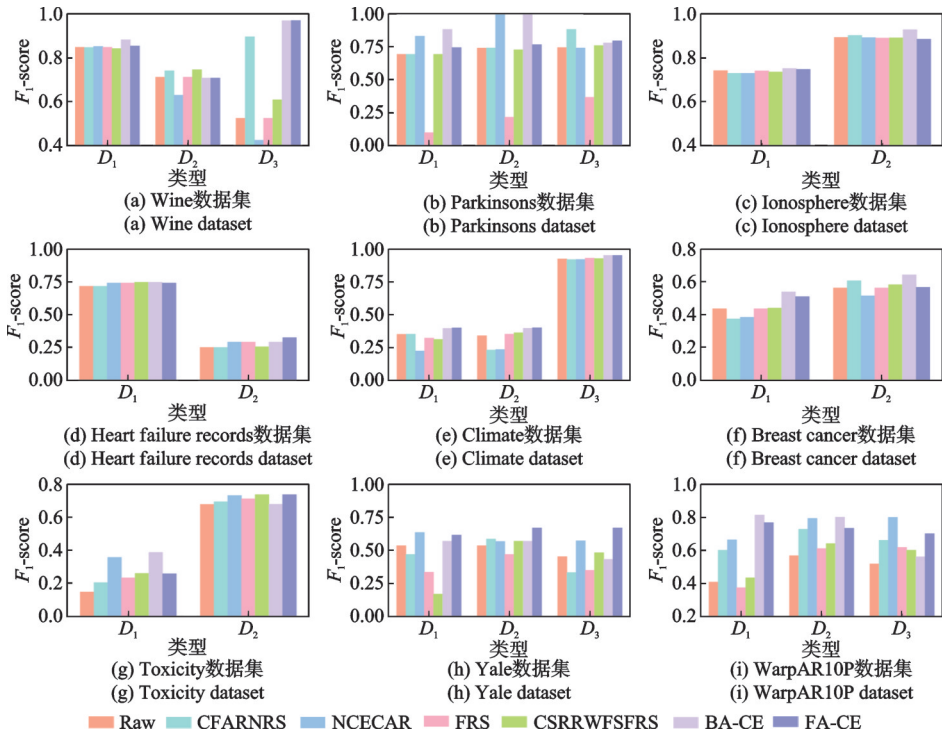


图2 在KNN分类器下9个数据集的 $F_1$ -score

Fig.2  $F_1$ -score of nine datasets under the KNN classifier

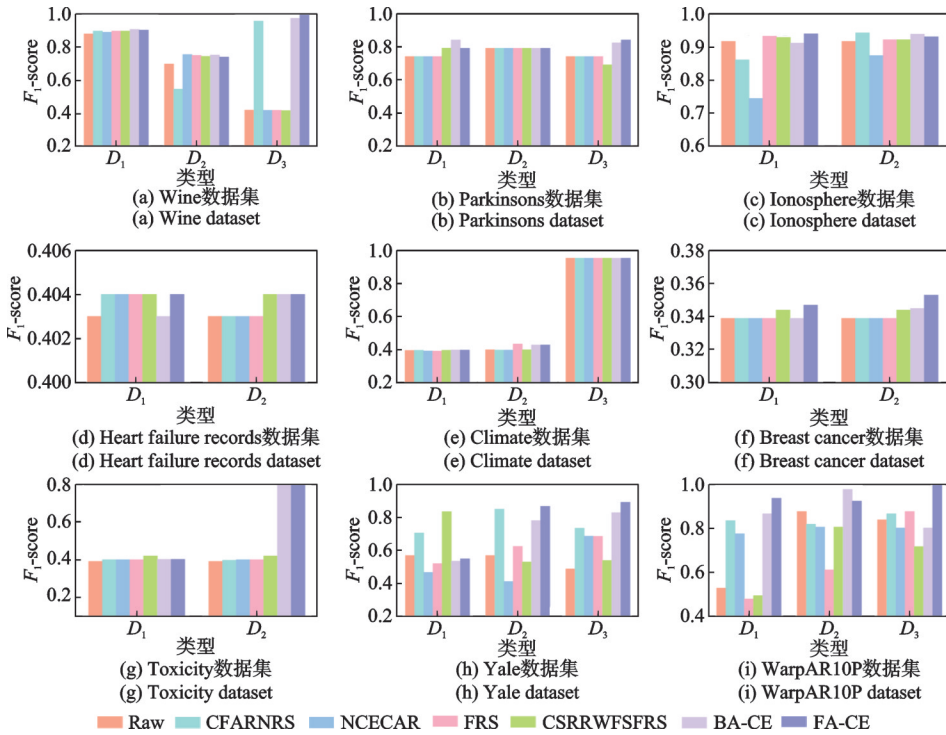
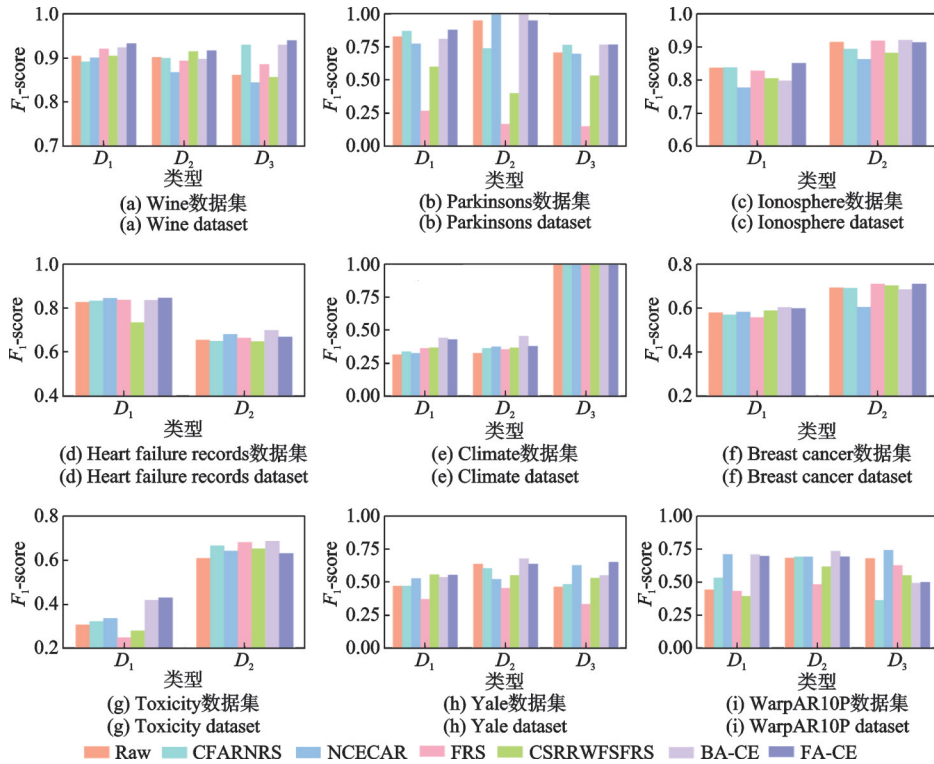


图3 在SVM分类器下9个数据集的 $F_1$ -score

Fig.3  $F_1$ -score of nine datasets under the SVM classifier

图4 在CART分类器下9个数据集的 $F_1$ -scoreFig.4  $F_1$ -score of nine datasets under the CART classifier

## 5 结束语

本文针对特定决策类的约简,提出了一种基于模糊粗糙集的特定类属性约简方法。通过融合模糊粗糙集的决策包含度与信息熵理论,定义了面向特定决策类的条件熵,并设计了FA-CE与BA-CE约简算法。最后实验结果证明了这两种约简算法相对于决策分类的属性约简在约简率上有大幅度的提升,在3个分类器中,分类准确率和 $F_1$ -score在大部分数据集上都有更好的表现。在未来的研究工作中,可以考虑针对符号类型数据和数值型的混合数据进行特定类属性约简,或者在不完备信息系统中研究特定类的属性约简,从而进一步拓展特定类属性约简的应用场景。

## 参考文献:

- [1] 郑颖春, 郭玲. 一种改进的基于知识粒度的增量属性约简算法[J]. 小型微型计算机系统, 2025, 46(1): 128-134.  
ZHENG Yingchun, GUO Ling. Improved knowledge granularity-based incremental attribute reduction algorithm[J]. Journal of Chinese Computer Systems, 2025, 46(1): 128-134.
- [2] 骆公志, 张尚蕾. 混合层次依赖度下的邻域粗糙集多目标特征选择算法[J]. 数据采集与处理, 2025, 40(1): 117-133.  
LUO Gongzhi, ZHANG Shanglei. A multi-objective feature selection algorithm based on neighborhood rough set with hybrid hierarchical dependency [J]. Journal of Data Acquisition and Processing, 2025, 40(1): 117-133.
- [3] 庞文莉, 于潇, 郑宇, 等. 基于后悔理论的多粒度直觉模糊三支决策模型[J]. 数据采集与处理, 2025, 40(2): 501-516.  
PANG Wenli, YU Xiao, ZHENG Yu, et al. A multi-granularity intuitionistic fuzzy three-way decision model based on regret theory[J]. Journal of Data Acquisition and Processing, 2025, 40(2): 501-516.
- [4] ZHANG Xiaoyan, ZHAO Weicheng. Uncertainty measures and feature selection based on composite entropy for generalized multigranulation fuzzy neighborhood rough set[J]. Fuzzy Sets and Systems, 2024, 486: 108971.

- [5] HU Qinghua, YU Daren, LIU Jinfu, et al. Neighborhood rough set based heterogeneous feature subset selection[J]. Information Sciences, 2008, 178(18): 3577-3594.
- [6] CHEN Hongmei, LI Tianrui, CAI Yong, et al. Parallel attribute reduction in dominance based neighborhood rough set[J]. Information Sciences, 2016, 373: 351-368.
- [7] JI Xia, DUAN Wanyu, PENG Jianhua, et al. Fuzzy rough set attribute reduction based on decision ball model[J]. International Journal of Approximate Reasoning, 2025, 179: 109364.
- [8] SHANNON C E. A mathematical theory of communication[J]. The Bell System Technical Journal, 1948, 27(3): 379-423.
- [9] MI Jusheng, YEE Leung, WU Weizhi. An uncertainty measure in partition-based fuzzy rough sets[J]. International Journal of General Systems, 2005, 34(1): 77-90.
- [10] DAI Jianhua, HU Hu, ZHENG Guojie, et al. Attribute reduction in interval-valued information systems based on information entropies[J]. Frontiers of Information Technology & Electronic Engineering, 2016, 17(9): 919-928.
- [11] JI Xia, LI Jie, YAO Sheng, et al. Attribute reduction based on fusion information entropy[J]. International Journal of Approximate Reasoning, 2023, 160: 108949.
- [12] LIANG Baohua, WANG Lin, LIU Yong. Attribute reduction based on improved information entropy[J]. Journal of Intelligent & Fuzzy Systems, 2019, 36(1): 709-718.
- [13] XIE Linlin, LIN Guoping, LI Jinjin, et al. A novel fuzzy-rough attribute reduction approach via local information entropy[J]. Fuzzy Sets and Systems, 2023, 473: 1.1-1.22.
- [14] YAO Yiyu, ZHANG Xianyong. Class-specific attribute reducts in rough set theory[J]. Information Sciences, 2017, 418: 601-618.
- [15] ZHANG Xianyong, YANG Jilin, TANG Lingyu. Three-way class-specific attribute reducts from the information viewpoint[J]. Information Sciences, 2020, 507: 840-872.
- [16] ZHANG Xianyong, FAN Yunrui, YAO Yuesong, et al. Class-specific attribute reducts based on neighborhood rough sets[J]. Journal of Intelligent & Fuzzy Systems, 2022, 43: 7891-7910.
- [17] LUO Lianhui, YANG Jilin, XIE Yu. Class-specific attribute reduction based on neighborhood conditional entropy[C]// Proceedings of 2023 International Conference on Cyber-Physical Social Intelligence (ICCSI). [S.l.]: IEEE, 2023: 479-484.
- [18] LUO Lianhui, YANG Jilin, ZHANG Xianyong, et al. Tri-level attribute reduction based on neighborhood rough sets[J]. Applied Intelligence, 2024, 54(5): 3786-3807.
- [19] MA Xiao, XU Hao, LIU Yi, et al. Class-specific feature selection using fuzzy information-theoretic metrics[J]. Engineering Applications of Artificial Intelligence, 2024, 136: 109035.
- [20] WANG Changzhong, QI Yali, SHAO Mingwen, et al. A fitting model for feature selection with fuzzy rough sets[J]. IEEE Transactions on Fuzzy Systems, 2017, 25(4): 741-753.
- [21] YANG Jilin, ZHANG Xianyong, QIN Keyun. Constructing robust fuzzy rough set models based on three-way decisions[J]. Cognitive Computation, 2021(14): 1955-1977.
- [22] MA Xiao, LU Kecheng. Class-specific feature selection using neighborhood mutual information with relevance-redundancy weight[J]. Knowledge-Based Systems, 2024, 300: 112212.

作者简介:



闵艳玲(2001-),女,硕士研究生,研究方向:粒计算与属性约简。



杨霁琳(1981-),通信作者,女,教授,硕士生导师,研究方向:粒计算与智能信息处理, E-mail: jilinyang@sicnu.edu.cn。



懂梦梦(2000-),男,硕士研究生,研究方向:不确定性数据分析。



张贤勇(1978-),男,教授、博士后,博士生导师,研究方向:不确定性分析与机器学习。

## Class-Specific Attribute Reduction Based on Conditional Entropy in Fuzzy Rough Sets

MIN Yanling<sup>1</sup>, YANG Jilin<sup>1\*</sup>, DONG Mengmeng<sup>1</sup>, ZHANG Xianyong<sup>2</sup>

(1. School of Computer Science, Sichuan Normal University, Chengdu 610101, China; 2. School of Mathematical Science, Sichuan Normal University, Chengdu 610066, China)

**Abstract:** Traditional attribute reduction methods construct a unified attribute reduction set for decision classification, ignoring the differentiated representation of attributes among various decision classes, which often results in the pivotal attributes of specific classes being redundantly covered and leads to suboptimal classification accuracy for those specific classes. To address these issues, this paper proposes a class-specific attribute reduction method driven by conditional entropy based on fuzzy rough sets, considering the advantages of fuzzy rough sets in handling widely existing numerical and fuzzy data. Firstly, by integrating the decision inclusion degree of fuzzy rough sets with information entropy theory, a class-specific conditional entropy is defined to quantify the local discriminative power of conditional attributes with respect to the target class. Secondly, the paper presents a class-specific attribute reduction condition based on conditional entropy and defines both internal and external attribute significance measures based on this class-specific conditional entropy. Furthermore, forward (FA-CE) and backward (BA-CE) attribute reduction algorithms are proposed based on attribute significance. Finally, the class-specific attribute reduction is conducted on seven UCI datasets and two feature selection datasets, and comparative analyses are performed against methods based on neighborhood conditional entropy, mutual information, neighborhood rough sets, and a conventional dependency-based reduction approach. The classification accuracy and  $F_1$ -score of the proposed method are evaluated using support vector machine (SVM), K-nearest neighbor (KNN) and classification and regression tree (CART) classifiers, demonstrating the rationality and effectiveness of the proposed class-specific attribute reduction approach.

### Highlights

1. Propose a class-specific attribute reduction method based on conditional entropy-driven fuzzy rough sets, which addresses the problem that attribute differences between different decision classes are ignored by traditional classification-based reduction algorithms.
2. Define conditional entropy for specific decision classes by integrating fuzzy rough set-based decision inclusion degree and information entropy theory, to quantify the local discriminative ability of conditional attributes for target decision classes.
3. Define the attribute significance measures based on the proposed conditional entropy for specific decision classes, and constructs two attribute reduction algorithms including the forward algorithm (FA-CE) and backward algorithm (BA-CE).

**Key words:** class-specific attribute reduction; fuzzy rough sets; conditional entropy; attribute reduction; decision inclusion degree

---

**Foundation items:** Humanities and Social Sciences Planning Fund of Ministry of Education(No.23YJA630114);Sichuan Provincial Natural Science Foundation(Nos.2024NSFSC0486, 2026NSFSC0444);Sichuan Normal University Postgraduate Innovation Ability Training Program(No.KY2025016).

**Received:** 2025-06-12; **Revised:** 2025-08-15

\***Corresponding author, E-mail:** jilinyang@sicnu.edu.cn.