

## 基于抗噪加权模糊粒度量的样本和特征双选择

李嘉豪<sup>1</sup>, 折延宏<sup>2\*</sup>, 贺晓丽<sup>2</sup>, 钱婷<sup>2</sup>, 郑文利<sup>2</sup>

(1. 西安石油大学 计算机学院, 陕西 西安 710065;

2. 西安石油大学 理学院, 陕西 西安 710065)

**摘要:** 尽管现有基于相对模糊粗糙集的特征选择方法在相似关系计算中已尝试刻画样本的离群分布来增强鲁棒性, 但仍未能有效抑制潜在噪声的干扰, 且未能进一步压缩数据规模。鉴于此, 本文提出基于抗噪加权模糊粒的双选择方法: 首先设计基于抗噪离散因子的相对距离度量, 实现依据局部样本密度分布的自适应调整。其次构建抗噪加权模糊粒, 完成双选择框架下的模型粒化。最终设计基于该粒结构的双选择算法BS-RFRS, 在最大化数据约简的同时提升分类性能。通过在12个基准数据集上的实验验证, 该算法在分类准确率与有效性方面显著优于其他5种所比较的双选择算法, 其中在医疗诊断数据集和工业控制数据集上取得非常显著的准确率提升, 且有效性较传统双选择模型有所提高。在标签噪声影响下, BS-RFRS的分类准确率相比于BSNID (bi-selection approach based on neighborhood importance degree)模型和BSFRS (bi-selection method based on fuzzy rough sets)模型分别平均提升19.9%和42.7%。

**关键词:** 模糊粗糙集; 样本分布密度; 抗噪权重; 相对距离度量; 粒计算

中图分类号: TP181

文献标志码: A

文章编号: 0253-2395(2026)01-0029-13

## Bi-selection of Instances and Features Based on Denoising Weighted Fuzzy Granular Measure

LI Jiahao<sup>1</sup>, SHE Yanhong<sup>2\*</sup>, HE Xiaoli<sup>2</sup>, QIAN Ting<sup>2</sup>, ZHENG Wenli<sup>2</sup>

(1. School of Computer Science, Xi'an Shiyou University, Xi'an 710065, China;

2. School of Science, Xi'an Shiyou University, Xi'an 710065, China)

**Abstract:** Existing feature selection methods based on relative fuzzy rough sets have attempted to characterize instance outlier distribution in similarity calculation to enhance robustness, but still fail to effectively suppress potential noise interference and cannot further compress data scale. To overcome this issue, this paper proposed a bi-selection method using denoise-weighted fuzzy granules (BS-RFRS). A relative distance measure with a denoise discretization factor for adaptive adjustment based on local instance density was designed. Denoise-weighted fuzzy granules were then constructed for model granulation within the bi-selection framework. Based on this granular structure, the paper proposed the BS-RFRS algorithm to maximize data reduction while improving classification performance. Experiments on 12 benchmark datasets demonstrated that BS-RFRS significantly outperforms five other bi-selection algorithms in classification accuracy and effectiveness. It achieves particularly notable accuracy gains on medical diagnosis and industrial control datasets, and shows improved effectiveness over traditional models. Under label noise, the classification accuracy of BS-RFRS is on average improved by 19.9% and 42.7% compared with the BSNID model and the (bi-selection method based on

收稿日期: 2025-08-20; 修回日期: 2025-09-23

基金项目: 国家自然科学基金(12471442); 陕西省自然科学基金(2023-JC-YB-027; 2025JC-YBMS-034); 陕西省教育厅科学研究计划青年创新团队项目(23JP132)

作者简介: 李嘉豪(2000-), 男, 山东青岛人, 硕士研究生, 研究方向为粒计算。E-mail: JiahaoLi1783@outlook.com

\* 通信作者: 折延宏(SHE Yanhong), E-mail: yanhongshe@xsyu.edu.cn

引文格式: 李嘉豪, 折延宏, 贺晓丽, 等. 基于抗噪加权模糊粒度量的样本和特征双选择[J]. 山西大学学报(自然科学版), 2026, 49(1): 29-41. DOI: 10.13451/j.sxu.ns.2025110.

fuzzy rough sets) (BSFRS) model, respectively.

**Key words:** fuzzy rough sets; instance distribution density; denoising weight; relative distance metric; granular computing;

## 0 引言

随着数据规模的不断变大,在数据处理以及数据挖掘<sup>[1]</sup>等领域都面临着前所未有的挑战。受限于有限的计算机硬件性能,人们也越来越需要发展数据约简技术<sup>[2]</sup>来对海量数据进行约简处理。数据约简技术不仅可以消除冗余数据,减少计算量和运行时间,也可以剔除影响算法最终效果的噪声以及低质量的样本和特征,从而提高准确率等性能指标。

样本选择<sup>[3]</sup>是有效的数据约简技术之一,其通过减少冗余样本以及噪声样本的方式来减少数据规模并提炼出原始数据集的关键信息。García等<sup>[4]</sup>使用了一种模因算法来选择样本并将群体智能算法和局部搜索算法相结合。Tsai等<sup>[5]</sup>提出了一种有效的遗传算法来进行样本选择,通过适应生物的进化方式最终得到适应值最高的个体。Zhang等<sup>[6]</sup>提出了一种基于邻域分布理论的样本选择算法,首次同时结合了数据空间的密度特征和决策信息来评估样本的邻域环境。此外,基于神经网络、模糊粗糙集等的样本选择方法也已被提出<sup>[7-9]</sup>。

特征选择<sup>[10]</sup>作为数据约简中较为重要的技术之一,在解决维数灾难<sup>[11]</sup>、数据规模过大等问题上发挥了重要的作用。在Pawlak<sup>[12]</sup>首次提出知识粒的概念后,在粒计算<sup>[13-16]</sup>等领域对于特征选择技术的研究非常广泛,包括但不限于模糊粗糙集、邻域粗糙集、粒球计算等。为了解决经典粗糙集只能评估分类特征的不足,Hu等<sup>[17]</sup>提出了一种基于邻域粗糙集的前向特征搜索算法。此外,对于经典Shannon信息熵<sup>[18]</sup>的拓展,Hu等<sup>[19]</sup>还提出了一种新的邻域熵来衡量特征集在算法中的重要性以及相关性。与传统邻域粗糙集方法不同,An等<sup>[20]</sup>提出了基于相对模糊粗糙集模型的特征选择方法,在经典的模糊粗糙集的基础上考虑了样本分布的密集程度。Huang等<sup>[21]</sup>为解决分层语义的数据约简问题,提出了一种基于模糊依赖度的用于动态层次分类的特征选择方法。与此同

时,基于粒计算进行特征选择的方法<sup>[22-26]</sup>也在不断涌现。以上提到的特征选择模型中,大多集中于相似关系及隶属度的计算,而未能有效处理噪声因素<sup>[27]</sup>对于分类结果的影响。为了更好地减小噪声对分类的影响,本文提出了一种新型的抗噪加权模糊粒模型,该模型能够在相似度量时对噪声进行模型修正。

为了进一步进行数据约简,将样本选择和特征选择同时进行的双选择技术也受到不同领域学者的关注。在去除冗余样本的基础上进一步对特征集进行约简,可以更大程度上缩小数据规模。Ślęzak等<sup>[28]</sup>最先提出双向约简的概念,将样本选择和特征选择技术相结合。Zhang等<sup>[29]</sup>提出了一种基于模糊粗糙集模型的双选择方法BSFRS (bi-selection method based on fuzzy rough sets),定义了一种新的重要性程度函数作为评价指标来选择代表性样本和约简的属性集。同时,Zhang等<sup>[30]</sup>注意到在邻域粗糙集领域关于双选择方法的研究较少,于是提出了一种基于邻域粗糙集模型的双选择方法。Quan等<sup>[31]</sup>通过引入 $\beta$ 参数,提出了基于 $\beta$ 一致性粒化机制的模糊粗糙集双选择模型,来应对大规模数据的挑战。值得注意的是,Zhang等<sup>[29]</sup>提出了删除潜在噪声的处理方法,但此方法对于位于噪声边界的潜在噪声无法有效识别。针对这一局限,本文提出的抗噪加权模糊粒模型能够对潜在噪声进行有效修正,避免忽视靠近噪声边界的噪声影响。

基于上述分析,经典的模糊粗糙集双选择模型忽视了样本分布密集程度的影响,同时现有策略未能有效处理位于噪声边界的潜在噪声。因此,为克服上述局限,本文提出了基于抗噪加权的相对模糊粗糙集模型来构建一种新型的抗噪加权模糊粒,以便在双选择算法中可以考虑到样本分布的密集程度以及噪声因素的影响。基于此抗噪加权模糊粒,进一步提出一种新的双选择算法BS-RFRS。经过12个不同规模的数据集以及3项实验指标的数据实验,我们验证了BS-RFRS算法在分类准确率、有效

性以及抗噪鲁棒性等方面具有显著的优势。

本文结构安排如下:第1节回顾模糊粗糙集的预备理论知识;第2节简要概述基于抗噪加权的相对模糊粗糙集模型以及对于抗噪加权模糊粒的构造;第3节详细阐述新提出的基于抗噪加权模糊粒的BS-RFRS算法;第4节在数值实验中展示了BS-RFRS算法的性能表现;第5节为本文的总结部分。

## 1 预备知识

### 1.1 模糊粗糙集

在模糊粗糙集<sup>[14]</sup>模型中,设 $U$ 为一个非空论域且 $F(U \times U)$ 是在 $U \times U$ 上的模糊幂集。如果 $R \in F(U \times U)$ ,则 $R$ 被称为模糊关系,其中 $R(x, y)$ 度量了 $x$ 与 $y$ 之间的模糊关系。关于模糊关系更加详细的定义见文献[14]。在本文中,任意两个样本 $x \in U, y \in U$ 的模糊关系 $R(x, y)$ 的计算定义为:

$$R(x, y) = \exp\left(-\frac{D(x, y)}{\delta}\right),$$

其中 $D(x, y)$ 刻画了样本 $x$ 与 $y$ 之间的“距离”。任意的两个样本 $x \in U, y \in U$ ,如果模糊关系满足 $R(x, y) = R(y, x)$ ,则我们称其满足对称性。若模糊关系满足 $R(x, x) = 1$ ,则称其满足自反性。任意的三个样本 $x, y, z \in U$ 如果满足 $R(x, y) \geq T(R(x, y), R(y, z))$ 则称其为 $T$ -传递<sup>[15]</sup>的,其中 $T$ 为三角范数。特别地,当 $T = \min$ 时, $R$ 为模糊等价关系。

在模糊粗糙集理论中,还用到了上下近似来衡量某一元素属于一个特定集合的可能性程度和确定性程度:

$$\underline{R}X(x_i) = \inf_{x_j \in U} \max\{1 - R(x_i, x_j), X(x_j)\}, \quad (1)$$

$$\bar{R}X(x_i) = \sup_{x_j \in U} \min\{R(x_i, x_j), X(x_j)\}, \quad (2)$$

其中 $\underline{R}X(x_i)$ 代表了 $x_i$ 在 $R$ 关系下,属于 $X$ 的下近似的隶属度,其中下近似表达了 $x_i$ 属于 $X$ 的确定性程度。 $\bar{R}X(x_i)$ 代表了 $x_i$ 在 $R$ 关系下,属于 $X$ 的上近似的隶属度,其中上近似表达了 $x_i$ 属于 $X$ 的可能性程度。

### 1.2 模糊信息系统以及模糊决策系统

设 $(U, A)$ 为一个模糊信息系统<sup>[16]</sup>,其中论域 $U = \{x_1, x_2, \dots, x_n\}$ ,特征集 $A = \{a_1, a_2, \dots, a_m\}$ 。对于 $a_i \in A$ 存在一个映射使得 $a_i: U \rightarrow V_{a_i}$ ,其中 $V_{a_i}$ 为 $a_i$ 的定义域,由此模糊关系 $R_{\{a_i\}}$ 可以被称为在模糊幂集 $F(U \times U)$ 上的一个模糊集合。 $R_{\{a_i\}}(x_i, x_j)$ 用来描述 $x_i, x_j$ 关于特征 $a_i$ 的相似程度。

在单决策特征的情况下,通过加入决策特征 $D$ 到 $(U, A)$ 中,其中 $A \cap D = \emptyset$ ,我们可以得到一个模糊决策系统 $(U, A \cup D)$ ,其中 $A$ 被称为条件特征集, $D$ 被称为决策特征。论域 $U$ 可以通过 $D$ 划分为相互非交的子集 $U/D = \{[x_i]_D; x_i \in U\}$ ,其中 $[x_i]_D = \{x_j \in U: D(x_i) = D(x_j)\}$ 被称作 $x_i$ 所属的决策类。此外,关于决策类 $[x_i]_D$ 的隶属函数为:

$$[x_i]_D(x_j) = \begin{cases} 1, & x_j \in [x_i]_D \\ 0, & x_j \notin [x_i]_D \end{cases}. \quad (3)$$

由此,设有一个模糊决策系统 $(U, A \cup D)$ ,其中 $U = \{x_1, x_2, \dots, x_n\}$ 且 $B \subseteq A$ ,则对每个 $x_i \in U$ ,若 $x_j \in [x_i]_D$ ,我们有:

$$\underline{R}_B[x_i]_D(x_j) = \inf_{x_k \in [x_i]_D} \{1 - R_B(x_j, x_k)\}. \quad (4)$$

若 $x_j \notin [x_i]_D$ :

$$\bar{R}_B[x_i]_D(x_j) = \sup_{x_k \in [x_i]_D} \{R_B(x_j, x_k)\}. \quad (5)$$

## 2 抗噪加权的相对模糊粗糙集

### 2.1 基于标准差的相对距离

在如欧氏距离和切比雪夫距离等经典的距离度量中,并没有考虑到样本之间的密集程度。例如在图1中,设样本 $x$ 与样本 $y, z$ 的距离相同。按照最近邻(Nearest Neighbor, NN)分类器的定义, $x$ 属于类别1的可能性和 $x$ 属于类别2的可能性相同。但图中类别2的样本分布更加稀疏,类别1的样本分布更加紧密,直观上来看, $x$ 属于类别2的可能性应该更高。因此在进行距离的计算时需要考虑样本的分布情况。为解决此类问题,引入相对距离<sup>[20]</sup>的概念。

**定义1<sup>[20]</sup>** 设 $(U, A \cup D)$ 为一个模糊决策系统,其中 $U = \{x_1, x_2, \dots, x_n\}$ , $A = \{a_1, a_2, \dots, a_m\}$

且特征集  $B \subseteq A$ 。将任意两个样本  $x, y \in U$  在特征集  $B$  度量下的相对距离记为  $D_{rB}^k(x, y)$ :

$$D_{rB}^k(x, y) = \frac{D_B(x, y)}{1 + S_k(y)}, \quad (6)$$

其中上标  $k$  为超参数,  $S_k(y)$  为样本  $y$  距离最近的  $k$  个同类样本(包含其自身)之间的标准差。

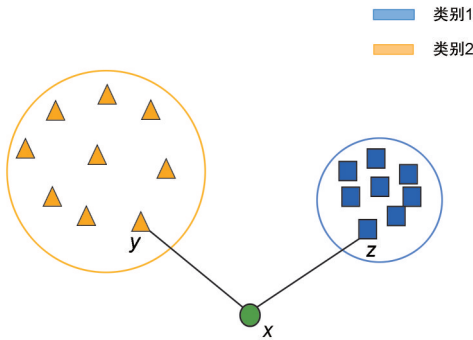


图1 经典距离度量

Fig. 1 Classical distance measure

根据图2所示,样本  $x$  与样本  $y, z$  在特征集  $B$  度量下的距离相同。设超参数  $k=4$ , 经过计算会得到  $S_4(y) > S_4(z)$ 。由公式(6)可知,使用相对距离来度量时,  $D_{rB}^4(x, y) < D_{rB}^4(x, z)$ 。故在NN分类器下可得出  $x$  应该属于类别2,符合直观结果。可见由于考虑了样本分布的离散程度,相对距离相较于普通的距离计算更具有适用性。

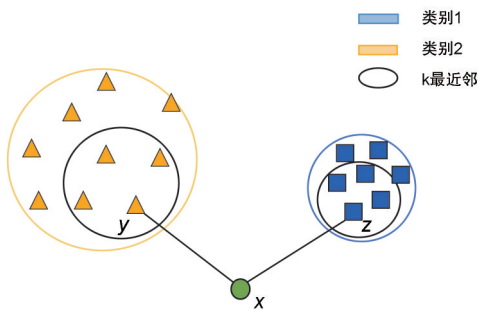


图2 相对距离度量

Fig. 2 Relative distance measure

### 2.2 基于抗噪加权的相对距离

在定义1中所定义的相对距离度量虽然考虑了样本的分布离散程度,但忽略了潜在噪声的影响,这将导致其在噪声较多的数据环境中表现不佳。如图3所示,样本  $x$  与样本  $y_1, z_1$  在特征集  $B = \{a_1, a_2\}$  度量下的距离相同,但样本

$y_1$  由于离群程度较高,有可能为数据集中的噪声。图3实例中的简要模糊信息系统如表1所示。

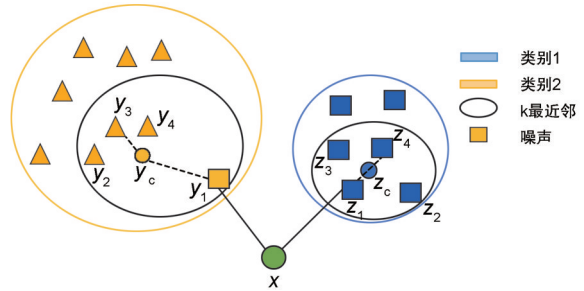


图3 有标签噪声下的相对距离度量

Fig. 3 Relative distance measure with label noise

表1 简要模糊信息系统

Table 1 Simple fuzzy information system

索引	类别1(y)		类别2(z)	
	$a_1$	$a_2$	$a_1$	$a_2$
1	0.500	0.4	0.70	0.3
2	0.100	0.6	0.80	0.2
3	0.200	0.7	0.70	0.5
4	0.300	0.7	0.80	0.5
c	0.275	0.6	0.75	0.4

假设待分类样本  $x$  与样本  $y_1, z_1$  的距离同为  $d$ , 经过标准差计算可得  $S_k(y_1) = 0.0612$ ,  $S_k(z_1) = 0.0408$ , 即  $S_k(y_1) > S_k(z_1)$ 。此时,根据定义1来计算相对距离会有  $D_{rB}^k(x, y_1) < D_{rB}^k(x, z_1)$ , 在NN分类器下  $x$  依然会被判为类别2。在样本  $y_1$  为噪声的情况下,这显然是不正确的分类结果。为了解决此类问题,我们提出了抗噪加权距离来进行复杂环境的距离度量。

定义2 设  $(U, A \cup D)$  为一个模糊决策系统,其中  $U = \{x_1, x_2, \dots, x_n\}$ ,  $A = \{a_1, a_2, \dots, a_m\}$  且特征集  $B \subseteq A$ 。将任意两个样本  $x, y \in U$  在特征集  $B$  度量下的抗噪加权距离记为  $D_{wB}^k(x, y)$ 。计算公式如下:

$$D_{wB}^k(x, y) = \frac{D_B(x, y)}{1 + \xi_k(y)}, \quad (7)$$

其中  $\xi_k(y)$  为抗噪离散因子。如图3所示,设距离样本  $y$  最近的  $k$  个同类样本为  $\{y_1, y_2, \dots, y_k\}$  (包含  $y$  自身,为方便后续公式表达记  $y$  为  $y_1$ ),

$y_c$  为这  $k$  个样本的中心点, 对于  $\forall a_i \in B$ ,  $a_i(y_c) = \sum_{p=1}^k a_i(y_p)/k$ 。基于上述条件, 将抗噪离散因子  $\xi_k(y)$  的计算公式定义为:

$$\xi_k(y) = \sqrt{\frac{\sum_{p=1}^k (D(y_p, y_c) - \bar{D})^2}{(k-1) \times \omega_k(y)^2}}, \quad (8)$$

其中  $\bar{D}$  为  $k$  个样本  $\{y_1, y_2, \dots, y_k\}$  与中心点  $y_c$  的距离平均值,  $\omega_k(y)$  为  $y$  的离群权重, 计算公式为:

$$\omega_k = \frac{D(y, y_c)}{\min_{p=1}^k D(y_p, y_c)}. \quad (9)$$

由上式可知随着  $y$  的离群程度变大, 权重  $\omega_k(y)$  随之增加, 导致  $\xi_k(y)$  减小, 最终引起抗噪加权距离  $D_{wB}^k(x, y)$  变远。如图 3 中所示, 若  $y_1$  的离群程度较大, 使得  $D_{wB}^k(x, y_1) > D_{wB}^k(x, z_1)$ , 为辅助理解, 本文在此对图 3 所示情况进行数值实例的推导: 根据公式 (9) 可以计算出  $\omega_k(y_1) = 2.25$ ,  $\omega_k(z_1) = 1$ , 进一步根据公式 (8) 可得  $\xi_k(y_1) = 0.0612/2.25 = 0.0272$ , 同理可得  $\xi_k(z_1) = 0.0408$ , 故  $\xi_k(y_1) < \xi_k(z_1)$ 。由于  $D_B(x, y_1) = D_B(x, z_1) = d$ , 故由公式 (7) 可推得  $D_{wB}^k(x, y_1) > D_{wB}^k(x, z_1)$ 。借此, NN 分类器会将其判为类别 1, 符合直观判断。基于上述分析, 抗噪加权距离在具有潜在噪声的环境下具有更强的适用性。

### 2.3 抗噪加权模糊粗糙集

**定义 3** 设  $(U, A \cup D)$  为一个模糊决策系统, 其中  $U = \{x_1, x_2, \dots, x_n\}$ ,  $A = \{a_1, a_2, \dots, a_m\}$ 。对于任意样本  $x_i \in U$  和样本集合  $X \subseteq U$ ,  $x_i$  对于  $X$  的抗噪加权模糊上下近似的隶属度计算如下:

$$\underline{R}^w X(x_i) = \inf_{x_j \in U} \max \{1 - R^w(x_i, x_j), X(x_j)\},$$

$$\overline{R}^w X(x_i) = \sup_{x_j \in U} \min \{R^w(x_i, x_j), X(x_j)\},$$

其中  $R^w(x_i, x_j)$  为样本  $x_i$  和  $x_j$  的抗噪加权模糊关系, 上标  $w$  为权重的语义。当  $B \subseteq A$ , 对于任意的两个样本  $x \in U, y \in U$ ,  $R^w_B(x, y)$  的计算定义为:

$$R^w_B(x, y) = \exp\left(-\frac{D_{wB}^k(x, y)}{\delta}\right). \quad (10)$$

由上式可知当  $x_j \in [x_i]_D$ :

$$\underline{R}^w_B[x_i]_D(x_j) = \inf_{x_k \in [x_i]_D} \{1 - R^w_B(x_j, x_k)\}, \quad (11)$$

以及当  $x_j \notin [x_i]_D$ :

$$\overline{R}^w_B[x_i]_D(x_j) = \sup_{x_k \in [x_i]_D} \{R^w_B(x_j, x_k)\}. \quad (12)$$

### 2.4 抗噪加权模糊粒

在接下来第 3 节介绍的双选择算法中, 判断一个样本能否“覆盖”另一个样本是非常重要的。在这里通过构建抗噪加权模糊粒的方法来进行判断。

**定义 4** 设  $(U, A \cup D)$  为一个模糊决策系统, 其中  $U = \{x_1, x_2, \dots, x_n\}$ ,  $A = \{a_1, a_2, \dots, a_m\}$ , 且特征集  $B \subseteq A$ 。对于任意样本  $x_i, x_j \in U$ , 抗噪加权模糊粒  $[x_i]_B^{\tau_i}$  定义为:

$$[x_i]_B^{\tau_i}(x_j) = \begin{cases} \tau_i, & D_{wB}^k(x_i, x_j) < \tau_i \\ 0, & D_{wB}^k(x_i, x_j) \geq \tau_i \end{cases}, \quad (13)$$

其中  $\tau_i = \underline{R}^w_A[x_i]_D(x_i)$ ,  $\tau_i$  为以下近似形式存在的覆盖边界。当  $[x_i]_B^{\tau_i}(x_j) > 0$  时, 则有  $D_{wB}^k(x_i, x_j) < \tau_i$  成立, 由下近似计算公式 (11) 可知, 这代表了样本  $x_i$  和  $x_j$  的抗噪加权距离小于  $x_i$  与其最近异类样本  $x_k \notin [x_i]_D$  的抗噪加权距离。因此可以认为这种情况下粒  $[x_i]_B^{\tau_i}$  可以覆盖样本  $x_j$ 。

**定义 5** 设  $(U, A \cup D)$  为一个模糊决策系统, 其中  $U = \{x_1, x_2, \dots, x_n\}$ ,  $A = \{a_1, a_2, \dots, a_m\}$ , 且特征集  $B \subseteq A$ 。对于任意样本  $x_i, x_j \in U$ , 粒  $[x_i]_B^{\tau_i}$  的覆盖集合  $C_{ov}([x_i]_B^{\tau_i})$  定义为:

$$C_{ov}([x_i]_B^{\tau_i}) = \{x_j: [x_i]_B^{\tau_i}(x_j) > 0\}. \quad (14)$$

## 3 样本与特征双选择算法

### 3.1 抗噪加权样本选择

样本选择算法的核心目标是通过筛选具有代表性的样本子集实现对原始大规模样本集的有效表征。在模型训练过程中, 噪声干扰的普遍存在性与离群样本的模糊性导致难以准确判定特定样本是否为噪声。鉴于此, 相较于模糊

粗糙集双选择模型<sup>[29]</sup>采用的噪声删除策略,抗噪加权样本选择通过引入抗噪加权模糊粒,对潜在离群点的距离计算进行离散校准,而非直接删除。为此,我们引入了模糊粗糙集双选择模型中的重要性程度<sup>[29]</sup>的定义并对其进行抗噪加权的自适应改造,改造后的定义6和定义7如下。

**定义6** 设 $(U, A \cup D)$ 为一个模糊决策系统,其中 $U = \{x_1, x_2, \dots, x_n\}$ ,  $A = \{a_1, a_2, \dots, a_m\}$ 。对于任意样本 $x_i \in U$ 构造的抗噪加权模糊粒 $[x_i]_A^{\tau_i}$ ,其重要性程度 $F([x_i]_A^{\tau_i})$ 定义为:

$$F([x_i]_A^{\tau_i}) = \tau_i \cdot |C_{ov}([x_i]_A^{\tau_i})|, \quad (15)$$

其中 $|\cdot|$ 为基数,抗噪加权模糊粒 $[x_i]_A^{\tau_i}$ 覆盖的样本个数 $|C_{ov}([x_i]_A^{\tau_i})|$ 越多,则该粒的覆盖能力越强。借此可引申到样本 $x_i$ 所具有的覆盖能力大小并以此判断样本的重要性。由公式(15)可知,对于样本重要性程度的考虑因素不仅有样本的覆盖能力 $|C_{ov}([x_i]_A^{\tau_i})|$ ,同时也考虑样本属于其自身决策类的确定性程度。这代表了如果出现两个样本的覆盖能力相同的情况,则 $F$ 较大的样本其重要性程度 $F$ 较大。完成 $F$ 的定义后,抗噪加权样本选择算法的具体过程可以分为以下几步:

首先以数据集中所有样本构建抗噪加权模糊粒 $[x]_A^{\tau}$ 并计算其覆盖能力 $|C_{ov}([x]_A^{\tau})|$ 。然后根据公式(15)计算出所有粒的重要性程度 $F([x]_A^{\tau})$ ,挑选出重要性程度值最大的样本 $x_0$ ,即 $x_0 = \operatorname{argmax}_{x \in U} F([x]_A^{\tau})$ ,将样本 $x_0$ 加入到前向搜索集合 $U_*$ 中并将集合 $C_{ov}([x_0]_A^{\tau_0})$ 中所包含的样本(包括 $x_0$ 自身)从原数据集 $U$ 中删除。最后进行迭代判断,如果所有的样本都被挑选出或者被删除,则循环结束,输出前向搜索集合 $U_*$ ,否则继续挑选 $F([x]_A^{\tau})$ 最大的样本并进行循环。算法伪代码如算法1所示。

为了在加速样本选择的同时避免因实时更新下近似而造成大量有效样本被剔除。在迭代步骤的7-12剔除样本时,沿用步骤1-5中计

算的 $F([x_i]_A^{\tau_i})$ 。

#### 算法1 抗噪加权样本选择算法

输入:模糊决策系统 $(U, A \cup D)$ ,超参数 $k$	
输出:前向搜索样本集 $U_*$	
初始化样本集 $U_* \leftarrow \emptyset$ ;	
1: for $i = 1:  U $	//遍历所有样本
2:     计算覆盖边界 $\tau_i$ ;	
3:     根据 $[x_i]_B^{\tau_i}$ 定义及公式(14)得到覆盖集合 $C_{ov}([x_i]_B^{\tau_i})$ ;	
4:     根据公式(15)和参数 $k$ 计算 $F([x_i]_A^{\tau_i})$ ;	
5: end for	
6: $\max F \leftarrow \max_{x_i \in U} F([x_i]_A^{\tau_i})$ ;	
7: while $\max F > 0$	// $\max F \leq 0$ 时已无未选样本
8: $x_0 = \operatorname{argmax}_{x_i \in U} F([x_i]_A^{\tau_i})$ ;	//选择 $F$ 最大的粒
9: $U_* \leftarrow U_* \cup \{x_0\}$ ;	//前向搜索
10: $U \leftarrow U - C_{ov}([x_0]_A^{\tau_0})$ ;	//剔除 $x_0$ 覆盖样本
11: $\max F \leftarrow \max_{x_i \in U} F([x_i]_A^{\tau_i})$ ;	//更新当前最大 $F$
12: end while	
13: 输出 $U_*$ 。	

现在对算法1进行最大时间复杂度分析。算法1中步骤2-4的时间复杂度为 $O(|U|^2 \times |A| \times k)$ ,步骤5-6的时间复杂度为 $O(|U| \times |A|)$ ,步骤7-12的时间复杂度为 $O(|U|^2 \times |A|)$ 。综合来看,算法1总共的最大时间复杂度为 $O(|U|^2 \times |A| \times k)$ 。

### 3.2 抗噪加权特征选择

经过算法1后,因为原来的样本集 $U$ 经过选择后为样本集 $U_*$ ,模糊决策系统变为 $(U_*, A \cup D)$ 。特征选择算法则是在此基础上对特征集 $A$ 进行选择得到特征集 $A_*$ ,借此将整个数据集进一步压缩,减少模型训练时的计算量。在特征选择算法中我们依然需要用到先前提出的抗噪加权模糊粒 $[x_i]_B^{\tau_i}$ ,同时对重要性程度的计算进行改造,使其适用于抗噪加权特征选择任务。

**定义7** 设 $(U_*, A \cup D)$ 为一个模糊决策系统,其中 $U_* = \{x_1, x_2, \dots, x_n\}$ ,  $A = \{a_1, a_2, \dots, a_m\}$ ,且有任意特征集 $B \subseteq A$ 。 $\forall x_i \in U_*$ 构造的粒

$[x_i]_B^{r_i}$  的重要性程度  $L([x_i]_B^{r_i})$  定义为:

$$L([x_i]_B^{r_i}) = \mathbf{r}_i^* \cdot \frac{|C_{ov}([x_i]_A^{r_i})|^2}{|C_{ov}([x_i]_B^{r_i})|}. \quad (16)$$

值得注意的是公式(16)定义的  $L([x_i]_B^{r_i})$  与公式(15)定义的  $F([x_i]_A^{r_i})$  不同,区别在于前者量化的是特征集  $B$  的选取影响。由于样本集由  $U$  变为了  $U_*$ ,因此对于任意样本  $x_i \in U_*$ ,有  $\mathbf{r}_i^* = \underline{R}_A^w[x_i]_D^*(x_i)$ 。在特征选择算法中,我们不对单个粒的重要性程度进行比较,而是对  $U_*$  整体进行考虑。因此我们将所有抗噪加权模糊粒在任意特征集  $B$  下的重要性程度之和记为  $\phi_{U_*}(B)$ ,其计算公式如下:

$$\phi_{U_*}(B) = L([x_i]_B^{r_i}), \quad (17)$$

其中根据文献[29]中的定理1,  $\phi_{U_*}$  满足单调性,即  $\phi_{U_*}(B - \{a_i\}) < \phi_{U_*}(B)$ 。根据以上定义,在相同样本集的情况下,  $\phi_{U_*}(B)$  越大,代表所有抗噪加权模糊粒在特征集  $B$  下的重要性程度之和越大。因此可以用重要性程度之和  $\phi_{U_*}$  来衡量特征集重要与否。完成  $\phi_{U_*}$  的定义后,本文中抗噪加权的前向搜索特征选择算法具体过程可以分为以下几步:

初始化前向搜索特征集  $A_*$ 。首先计算出所有抗噪加权模糊粒在样本集  $U_*$  下的重要性程度  $L([x_i]_A^{r_i})$ ,然后根据公式(16)和(17)计算出  $\phi_{U_*}(A)$ 。从未选择的特征中选择单个特征  $a_0 \in A - A_*$ ,其中  $a_0 = \operatorname{argmax}_{a_i \in A - A_*} \phi_{U_*}(A_* \cup \{a_i\})$ ,并更新  $A_*$  为  $A_* \cup \{a_0\}$ 。最后判断条件  $\phi_{U_*}(A_*) < \phi_{U_*}(A)$ ,如果满足则继续循环,不满足则输出  $A_*$ 。

现在对算法2进行最大时间复杂度分析。算法2中步骤2—4的最大时间复杂度为  $O(|U_*|^2 \times A \times k)$ ,步骤6—9的最大时间复杂度为  $O(|U_*|^2 \times A^2)$ 。综上所述,算法2的最大时间复杂度为  $O(|U_*|^2 \times A \times (A + k))$ 。

经过算法2选择的特征集  $A_*$  并不一定是最佳约简的特征集,最后我们通过包装器算法对特征集  $A_*$  进行进一步约简为  $A_w$ ,在此不对此

包装器算法进行赘述。最后,本文所提出的双选择算法(BS-RFRS)过程如算法3所示。

#### 算法2 抗噪加权特征选择算法

输入:模糊决策系统  $(U_*, A \cup D)$ ,超参数  $k$   
 输出:前向搜索样本集  $A_*$

- 1: 初始化特征集  $A_* \leftarrow \emptyset$ ;
- 2: for  $i = 1: |U_*|$  //遍历所有样本
- 3: 计算覆盖边界  $\mathbf{r}_i^*$ ;
- 4: 根据公式(14)获取覆盖集合  $C_{ov}([x_i]_B^{r_i})$  和  $C_{ov}([x_i]_A^{r_i})$ ;
- 5: 计算  $L([x_i]_A^{r_i})$ ;
- 6: end for
- 7: 计算  $\phi_{U_*}(A)$ ; //计算终止条件
- 8: while  $\phi_{U_*}(A_*) < \phi_{U_*}(A)$
- 9: 挑选  $a_0 = \operatorname{argmax}_{a_i \in A - A_*} \phi_{U_*}(A_* \cup \{a_i\})$ ;
- 10:  $A_* \leftarrow A_* \cup \{a_0\}$ ; //前向搜索
- 11: end while
- 12: 输出  $A_*$ 。

#### 算法3 BS-RFRS

输入:原始模糊决策系统  $(U, A \cup D)$ ,超参数  $k$   
 输出:最终的模糊决策系统  $(U_*, A_w \cup D)$

- 1: 对数据预处理;
- 2: 算法1得到  $U_*$ ;
- 3: 算法2得到  $A_*$ ;
- 4: 包装器算法得到  $A_w$ ;
- 5: 输出  $(U_*, A_w \cup D)$ 。

## 4 数值实验

在此部分,我们使用了从UCI等来源的12个数据集进行数值实验来评价BS-RFRS算法的性能表现。这些数据集在表2简述。

### 4.1 评价指标

于此,我们采用了3个指标来对BS-RFRS算法进行性能评价,分别是分类准确率、约简率、有效性。

(1)分类准确率:在双选择中,样本选择和特征选择都会对模型的预测性能产生影响。样本选择通过消除冗余数据和噪声来更好地服务于特征选择,特征选择则是在约简样本集的基础上消除冗余特征。最终的准确率用来评估所选择约简特征集的质量。分类准确率计算公式如下:

$$r_{acc} = \frac{|I_R|}{|I|}, \quad (18)$$

其中集合  $I_R$  为分类正确的样本,  $I$  为全部样本。

表2 数据集简述

Table 2 Description of the datasets

数据集缩写	样本数量	特征数量	类别数量
Wine	178	13	3
Iono	351	34	2
WDBC	569	30	2
Sonar	208	61	2
Libras	360	91	15
Darwin	174	451	2
CTG	2 126	21	3
WDG1	5 000	21	3
WDG2	5 000	41	3
Robot	5 456	25	4
Thyroid	9 127	29	2
Gamma	19 020	11	2

(2) 约简率: 约简率反映了算法在数据压缩上的效率。高约简率可以减小数据规模和计算负担, 但需要保证准确率不显著下降。约简率计算公式如下:

$$r_{\text{red}} = 1 - \frac{|U_*|}{|U|} \times \frac{|A_w|}{|A|}. \quad (19)$$

(3) 有效性: 有效性反映了权衡考虑约简率和分类准确率两者, 由约简率和准确率的乘积计算。计算公式如下:

$$r_{\text{eff}} = r_{\text{red}} \times r_{\text{acc}}. \quad (20)$$

## 4.2 实验设计

(1) 数据集预处理: 首先, 数据集中所有数据需要进行归一化处理, 使得  $a'(x_i) \in [0, 1]$ , 即对每个  $x_i \in U$  通过下式归一化:

$$a'(x_i) = \frac{a(x_i) - \min_j a(x_j)}{\max_j a(x_j) - \min_j a(x_j)}. \quad (21)$$

在数据归一化处理之后, 为了满足算法所需要的数据集模式, 对每个数据集进行划分。实验中使用了十折交叉验证方法, 每一折实验中, 取出随机不同的 10% 的样本作为测试集, 剩下 90% 的样本作为训练集。同时, 在更进一步的抗噪对比实验中, 我们分别随机将每个数据集的 10%、30%、50% 的数据标签异化, 使其成为标签噪声数据。通过不同的噪声比例来对比方法之间的抗噪性能。

对于任意的两个样本  $x_i, x_j \in U$  的距离  $D(x, y)$  采用切比雪夫距离进行计算:

$$D(x, y) = \max_{a_i \in A} |a_i(x) - a_i(y)|. \quad (22)$$

(2) 对比算法: 为了检验 BS-RFRS 算法的性能表现, 我们挑选了 5 个双选择算法进行对比。其中第一个算法为基于模糊粗糙集的双选择算法 BSFRS<sup>[29]</sup>, 第二个算法为基于邻域粗糙集的双选择算法 BSNID (bi-selection approach based on neighborhood importance degree)<sup>[30]</sup>。为了有更多对比算法, 我们用较为具有代表性的样本选择算法 RIS<sup>[9]</sup> 和 NID<sup>[30]</sup> 与三个特征选择算法 RFRS<sup>[20]</sup>、FSI<sup>[25]</sup> 以及 NRBO<sup>[26]</sup> 进行组合, 形成三个双选择算法 RIS-FSI、RIS-RFRS 和 NID-NRBO 来进行对比。

(3) 实验参数设置及环境: 对于 BS-RFRS 算法, 在抗噪加权距离  $D_{wB}^k(x_i, x_j)$  的计算中, 需要找到  $x_j$  的  $k$  个同类近邻来进行抗噪离散因子  $\xi_k(x_j)$  的计算。因此超参数  $k$  不同的设置可能对算法结果产生影响, 在本次实验中, 我们使参数  $k$  分别从 1 到 10 依次进行选取。实验中统一采用 KNN 分类器, 其中  $K=3$ , 即 3NN 分类方法。

我们使用 Matlab R2021b 进行数值实验。实验工作站硬件配置为 12th Gen Intel (R) Core (TM) i5-12500H 3.10 GHz, 以及 16.0 GB 内存和 windows11 操作系统。所有实验重复 20 次。

## 4.3 实验结果

数值实验中的六个算法在全部的 12 个数据集上的实验结果将在表格中展示, 其中包括分类准确率、约简率、有效性三个实验指标。实验结果将在如表 3—表 5 中展示。对于超参数  $k$  对算法性能表现所带来的影响将在图 4 中进行展示。混合不同比例标签噪声的抗噪性能对比实验结果在图 5 展示。

(1) 分类准确率: 从表 3 中可以看到 BS-RFRS 算法在分类准确率方面具有明显的优势。其中 BS-RFRS 算法在 12 个数据集中有 10 个数据集达到了最优, 2 个数据集中达到了次优。在小规模数据集中例如 Wine、WDBC (Wisconsin Diagnostic Breast Cancer)、Sonar 等为最优算法, 同时在高维数据集 Darwin 中 BS-RFRS 算法也为最优算法。在大规模数据集中如 CTG (Cardiotocography)、WDG2 (Waveform

Database Generator Version 2)、Thyroid 以及 Gamma 中,达到了最佳的性能表现且在 Robot 和 WDG1 (Waveform Database Generator Version 1)数据集中表现为次优。由表 3 观察得,基于抗噪加权模糊粒构造的 BS-RFRS 算法在不同数据规模的数据集中皆表现优越。

(2)约简率:从表 4 中可以看到 BS-RFRS 算法在约简率方面表现较为平庸。在同样样本的情况下,由于相对距离的计算小于经典的切比雪夫距离,因此 BS-RFRS 算法所选择的样本数量偏多。不过,BS-RFRS 算法在所有对比算法当中,约简率依然处于一个中等的性能表现。

表 3 BS-RFRS 和其他算法的分类准确率对比(数值范围 0~1)

Table 3 Average classification accuracies of BS-RFRS and the compared algorithms (values range: 0 to 1)

数据集	BS-RFRS	BSNID	BSFRS	NID-NRBO	RIS-FSI	RIS-RFRS	(w/t/l)
Wine	<u>0.983 0</u>	0.960 6	0.937 6	0.943 1	0.803 3	0.832 4	(5/0/0)
Iono	<u>0.931 6</u>	0.914 5	0.908 8	0.926 1	0.743 4	0.720 6	(4/1/0)
WDBC	<u>0.978 9</u>	0.949 1	0.961 3	0.971 9	0.870 0	0.871 8	(4/1/0)
Sonar	<u>0.942 8</u>	0.898 9	0.909 5	0.906 1	0.658 8	0.672 1	(5/0/0)
Libras	<u>0.788 8</u>	0.780 5	0.644 4	0.738 4	0.452 8	0.475 0	(4/1/0)
Darwin	<u>0.931 7</u>	0.902 2	0.904 7	0.923 8	0.613 4	0.722 2	(4/1/0)
CTG	<u>0.911 6</u>	0.883 3	0.900 3	0.888 5	0.808 6	0.690 5	(5/0/0)
WDG1	0.809 4	0.794 1	<u>0.815 0</u>	0.679 8	0.575 0	0.693 4	(4/1/0)
WDG2	<u>0.822 2</u>	0.795 3	0.814 6	0.755 2	0.568 6	0.756 0	(4/1/0)
Robot	0.908 3	0.880 0	0.890 7	0.880 3	0.810 4	<u>0.910 0</u>	(4/1/0)
Thyroid	<u>0.825 5</u>	0.331 5	0.805 7	0.299 7	0.724 1	0.812 5	(5/0/0)
Gamma	<u>0.809 1</u>	0.777 6	0.798 5	0.712 3	0.710 7	0.762 7	(5/0/0)

注:下划线表示最优数据。

表 4 BS-RFRS 和其他算法的约简率对比(数值范围 0~1)

Table 4 Average reduction rates of BS-RFRS and the compared algorithms (values range: 0 to 1)

数据集	BS-RFRS	BSNID	BSFRS	NID-NRBO	RIS-FSI	RIS-RFRS	(w/t/l)
Wine	0.935 1	0.953 9	<u>0.959 6</u>	0.909 2	0.950 2	0.932 4	(1/1/3)
Iono	0.945 5	0.929 3	<u>0.969 3</u>	0.901 1	0.953 6	0.806 6	(3/1/1)
WDBC	0.972 5	<u>0.994 1</u>	0.983 3	0.969 4	0.982 6	0.982 9	(0/1/4)
Sonar	0.948 7	0.909 3	0.944 4	0.902 4	<u>0.975 2</u>	0.947 4	(2/2/1)
Libras	0.937 2	0.922 6	0.967 4	<u>0.995 7</u>	0.978 3	0.963 4	(1/0/4)
Darwin	0.991 3	0.980 8	0.993 9	0.989 0	<u>0.997 0</u>	0.977 0	(2/3/0)
CTG	0.952 0	<u>0.982 0</u>	0.956 0	0.958 0	0.937 6	0.928 1	(2/2/1)
WDG1	0.794 8	0.987 6	0.838 9	<u>0.998 8</u>	0.893 7	0.898 5	(0/0/5)
WDG2	0.987 3	0.973 5	0.743 0	<u>0.995 8</u>	0.942 9	0.932 7	(4/1/0)
Robot	0.957 1	0.970 7	<u>0.973 4</u>	0.909 9	0.879 3	0.950 8	(2/1/2)
Thyroid	0.947 1	0.982 5	0.954 7	<u>0.996 3</u>	0.897 4	0.889 4	(2/1/2)
Gamma	0.915 1	0.987 0	0.895 6	<u>0.999 1</u>	0.800 8	0.788 1	(3/0/2)

注:下划线表示最优数据。

(3)有效性:从表 5 中可以看到 BS-RFRS 算法在有效性方面具有明显的优势。其中 BS-RFRS 算法在 12 个数据集中有 9 个数据集达到了最优,2 个数据集中达到了次优。根据有效性的定义,BS-RFRS 算法在权衡分类准确率和约简率方面性能表现出色,可以很好地体现出 BS-RFRS 算法在双选择算法中的优势。

此外对于超参数  $k$  来说,其很大程度上影响了算法在不同数据集上的性能表现。因此在图 4 中展示以上实验在不同数据集上的最佳  $k$  值选取。观察图 4 可知在 12 个数据集中有 10 个在  $k > 1$  时的性能表现最佳,这代表 BS-RFRS 中使用参数  $k$  值来判别噪声在大多数情况下可以提升算法表现。

表5 BS-RFRS和其他算法的有效性对比(数值范围0~1)

Table 5 Average effectiveness of BS-RFRS and the compared algorithms (values range: 0 to 1)

数据集	BS-RFRS	BSNID	BSFRS	NID-NRBO	RIS-FSI	RIS-RFRS	(w/t/1)
Wine	<u>0.919 2</u>	0.916 2	0.899 4	0.857 3	0.763 6	0.776 5	(4/1/0)
Iono	0.880 1	0.849 9	<u>0.880 5</u>	0.833 5	0.708 8	0.565 7	(4/1/0)
WDBC	<u>0.952 0</u>	0.943 4	0.945 3	0.941 9	0.855 0	0.856 4	(3/2/0)
Sonar	<u>0.894 5</u>	0.817 3	0.858 1	0.816 5	0.642 5	0.635 9	(5/0/0)
Libras	<u>0.739 0</u>	0.719 4	0.623 2	0.735 2	0.442 9	0.457 0	(4/1/0)
Darwin	<u>0.923 4</u>	0.885 1	0.899 1	0.913 1	0.611 6	0.705 5	(5/0/0)
CTG	<u>0.873 4</u>	0.867 4	0.860 7	0.850 0	0.758 3	0.640 8	(4/1/0)
WDG1	0.643 3	<u>0.784 2</u>	0.683 5	0.679 0	0.514 0	0.622 8	(2/0/3)
WDG2	<u>0.811 8</u>	0.774 3	0.605 4	0.752 1	0.536 2	0.705 1	(5/0/0)
Robot	<u>0.869 3</u>	0.854 3	0.867 0	0.802 3	0.712 7	0.865 3	(3/2/0)
Thyroid	<u>0.781 8</u>	0.325 7	0.769 2	0.298 6	0.649 8	0.722 6	(5/0/0)
Gamma	0.740 4	<u>0.767 5</u>	0.715 1	0.710 1	0.626 0	0.601 1	(4/0/1)

注:下划线表示最优数据。

各数据集最优k值分布

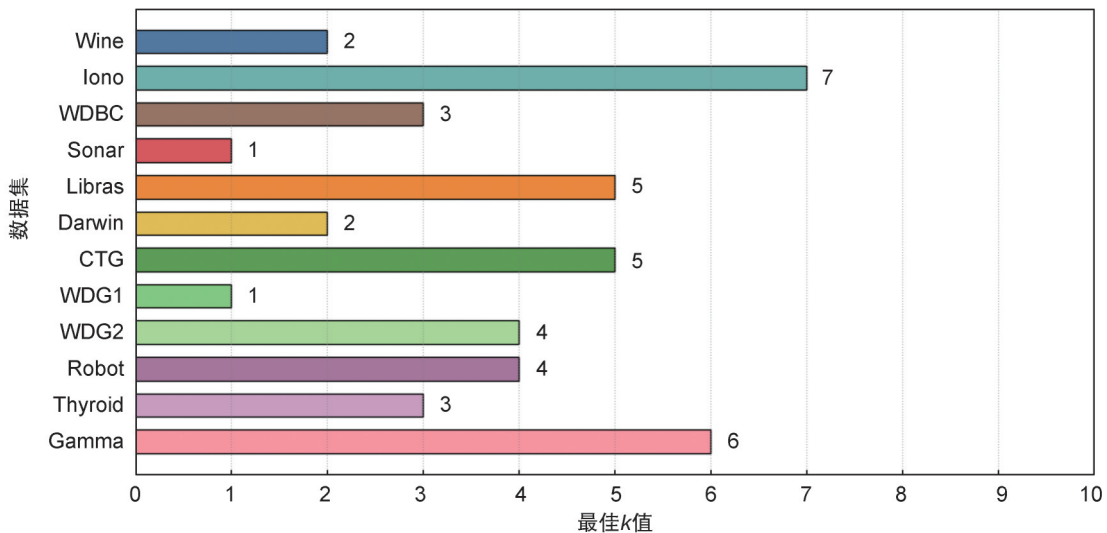


图4 不同数据集下的最佳k值

Fig 4 Optimal k values across different datasets

更进一步地,我们对原始数据集实验中表现最优的三个算法BS-RFRS、BSNID、BSFRS进行抗噪对比实验。通过随机制造10%、30%、50%的标签噪声进行对比,对比结果如图5所示。在30%以及50%的噪声环境下,BS-RFRS的抗噪性能优势显著,BS-RFRS在三种噪声比例下的平均分类准确率为0.8137,而BSNID和BSFRS分别为0.7037和0.6498。特别的,在50%的高标签噪声环境下BS-RFRS相对于BSNID和BSFRS分别提升了27.24%和40.17%。由此可见,在抗噪性能方面BS-RFRS有着明显优势。

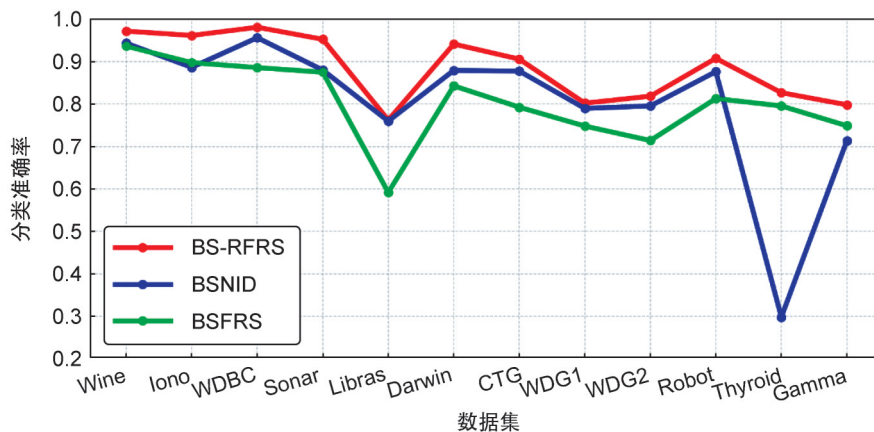
图6和图7分别展现了分类准确率和有效性的临界差异图,据图中显示,BS-RFRS在分类准确率和有效性的表现中处于领先地位。

### 5 结论

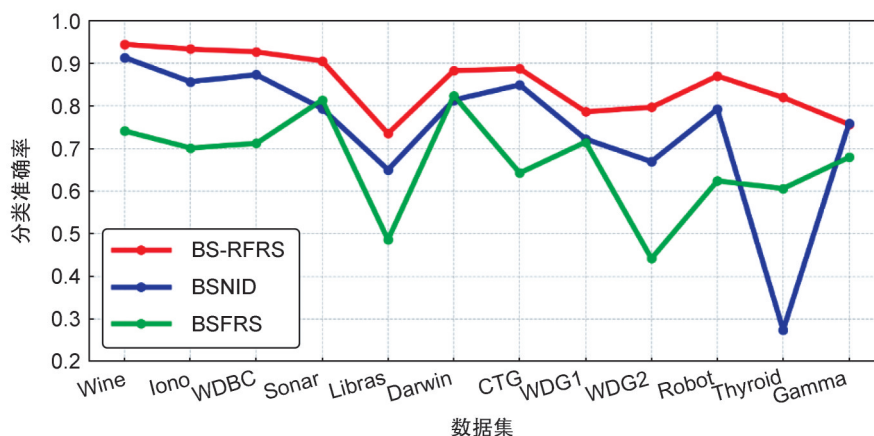
在本文中,我们通过对基于标准差的相对距离进行改进,并提出抗噪加权模糊粒的构造方法,最终提出了BS-RFRS算法。该算法能够在考虑样本分布离散程度及噪声影响的情况下完成双选择的任务。BS-RFRS算法在分类准确率以及有效性两个实验指标上有显著

的优势,这代表着BS-RFRS算法不光有着优异的分类准确率表现,而且在约简率和分类准确率的权衡上也有着优异的效果。在抗噪性

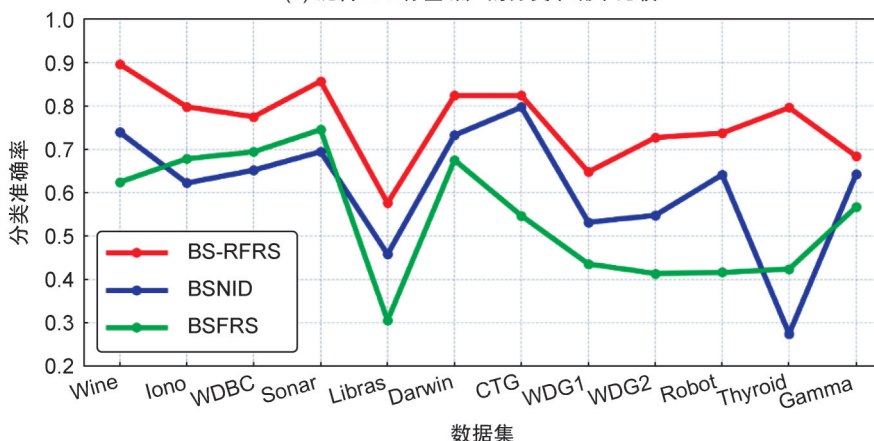
能对比实验中,BS-RFRS相较于传统双选择算法的分类准确性平均提升了42.7%,性能显著提升。



(a) 混合10%标签噪声的分类准确率比较



(b) 混合30%标签噪声的分类准确率比较



(c) 混合50%标签噪声的分类准确率比较

图5 不同标签噪声比例环境下的分类准确率(数值范围0~1)对比

Fig 5 Comparison of classification accuracy (range: 0 to 1) under different label noise ratio environments

在未来,双选择算法具有应用于不同的特定工程领域的潜力。后续工作中,在如分层分类等领域继续探索双选择算法。更进一步,由于增量学习方法在整体时间复杂度上具有相对

优势,且更符合工程中数据不断扩增的应用场景。未来可以在增量式角度改进基于相对模糊粗糙集模型的双选择算法,以此来满足实际工程中数据不断增长变化的要求。

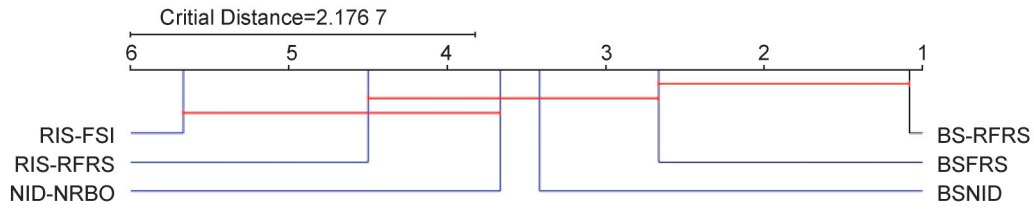


图 6 分类准确率的临界差异

Fig 6 Critical difference of classification accuracies

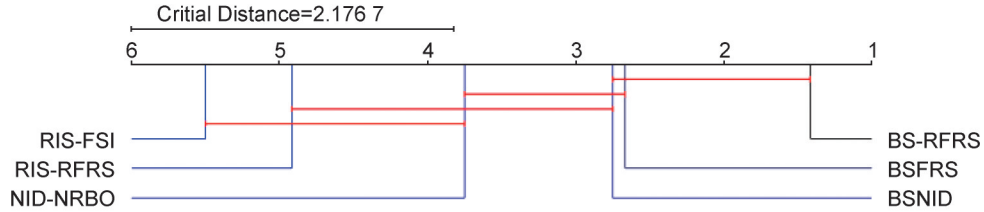


图 7 有效性的临界差异

Fig 7 Critical difference of effectiveness

## 参考文献:

- [1] CHEN M S, HAN J W, YU P S. Data Mining: an Overview from a Database Perspective[J]. *IEEE Trans Knowl Data Eng*, 1996, **8**(6): 866–883. DOI: 10.1109/69.553155.
- [2] LELEWER D A, HIRSCHBERG D S. Data Compression [J]. *ACM Comput Surv*, 1987, **19**(3): 261–296. DOI: 10.1145/45072.45074.
- [3] OLVERA-LÓPEZ J A, CARRASCO-OCHOA J A, MARTÍNEZ-TRINIDAD J F, et al. A Review of Instance Selection Methods[J]. *Artif Intell Rev*, 2010, **34**(2): 133–143. DOI: 10.1007/s10462-010-9165-y.
- [4] GARCÍA S, CANO J R, HERRERA F. A Memetic Algorithm for Evolutionary Prototype Selection: A Scaling up Approach[J]. *Pattern Recognit*, 2008, **41**(8): 2693–2709. DOI: 10.1016/j.patcog.2008.02.006.
- [5] TSAI C F, CHEN Z Y. Towards High Dimensional Instance Selection: an Evolutionary Approach[J]. *Decis Support Syst*, 2014, **61**: 79–92. DOI: 10.1016/j.dss.2014.01.012.
- [6] ZHANG X, YANG Q, QIAN T. Learning to Select Representative Instances Based on Neighborhood Distribution[J]. *Neurocomputing*, 2025, **654**: 131320. DOI: 10.1016/j.neucom.2025.131320.
- [7] BRIGHTON H, MELLISH C. Advances in Instance Selection for Instance-based Learning Algorithms[J]. *Data Min Knowl Discov*, 2002, **6**(2): 153–172. DOI: 10.1023/A:1014043630878.
- [8] ZHANG Q, ZHU Y, CORDEIRO F R, et al. PSSCL: a Progressive Sample Selection Framework with Contrastive Loss Designed for Noisy Labels[J]. *Pattern Recognit*, 2025, **161**: 111284. DOI: 10.1016/j.patcog.2024.111284.
- [9] ZHANG X, MEI C L, CHEN D G, et al. A Fuzzy Rough Set-based Feature Selection Method Using Representative Instances[J]. *Knowl Based Syst*, 2018, **151**: 216–229. DOI: 10.1016/j.knsys.2018.03.031.
- [10] VENKATESH B, ANURADHA J. A Review of Feature Selection and Its Methods[J]. *Cybern Inf Technol*, 2019, **19**(1): 3–26. DOI: 10.2478/cait-2019-0001.
- [11] ALTMAN N, KRZYWINSKI M. The Curse(s) of Dimensionality[J]. *Nat Methods*, 2018, **15**(6): 399–400. DOI: 10.1038/s41592-018-0019-x.
- [12] PAWLAK Z. Rough Sets[J]. *Int J Comput Inf Sci*, 1982, **11**(5): 341–356. DOI: 10.1007/BF01001956.
- [13] ZADEH L A. Granular Computing and Rough Set Theory[M]//Rough Sets and Intelligent Systems Paradigms. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007: 1–4. DOI: 10.1007/978-3-540-73451-2\_1.
- [14] DUBOIS D, PRADE H. Rough Fuzzy Sets and Fuzzy Rough Sets[J]. *Int J Gen Syst*, 1990, **17**(2/3): 191–209. DOI: 10.1080/03081079008935107.
- [15] 陈德刚, 徐伟华, 李金海, 等. 粒计算基础教程[M]. 北京: 科学出版社, 2019.
- [16] 陈德刚, 刘民, 吴澄, 等. 模糊信息系统的代数结构及其约简[J]. 清华大学学报(自然科学版), 2003, **43**(9): 1233–1235. DOI: 10.16511/j.cnki.qhdxxb.2003.09.022.
- [17] CHEN D G, LIU M, WU C, et al. Algebraic Structure and Reduction of Fuzzy Information Systems[J]. *J Tsinghua Univ Sci Technol*, 2003, **43**(9): 1233–1235. DOI: 10.16511/j.cnki.qhdxxb.2003.09.022.
- [17] HU Q H, YU D R, LIU J F, et al. Neighborhood Rough Set Based Heterogeneous Feature Subset Selection[J].

- Inf Sci*, 2008, **178**(18): 3577–3594. DOI: 10.1016/j.ins.2008.05.024.
- [18] SHANNON C E. A Mathematical Theory of Communication[J]. *Bell Syst Tech J*, 1948, **27**(3): 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.
- [19] HU Q H, YU D R. Neighborhood Entropy[C]//2009 International Conference on Machine Learning and Cybernetics. New York: IEEE, 2009: 1776–1782. DOI: 10.1109/ICMLC.2009.5212245.
- [20] AN S, ZHAO E H, WANG C Z, *et al.* Relative Fuzzy Rough Approximations for Feature Selection and Classification[J]. *IEEE Trans Cybern*, 2021, **53**(4): 2200–2210. DOI: 10.1109/TCYB.2021.3112674.
- [21] HUANG W L, SHE Y H, HE X L, *et al.* Fuzzy Rough Sets-based Incremental Feature Selection for Hierarchical Classification[J]. *IEEE Trans Fuzzy Syst*, 2023, **31**(10): 3721–3733. DOI: 10.1109/TFUZZ.2023.3300913.
- [22] ZHANG X Y, YAO Y Y. Tri-level Attribute Reduction in Rough Set Theory[J]. *Expert Syst Appl*, 2022, **190**: 116187. DOI: 10.1016/j.eswa.2021.116187.
- [23] HU M, TSANG E C C, GUO Y T, *et al.* Attribute Reduction Based on Overlap Degree and K-nearest-neighbor Rough Sets in Decision Information Systems[J]. *Inf Sci*, 2022, **584**: 301–324. DOI: 10.1016/j.ins.2021.10.063.
- [24] DAI J H, ZHU Z L, ZOU X T. Fuzzy Rough Attribute Reduction Based on Fuzzy Implication Granularity Information[J]. *IEEE Trans Fuzzy Syst*, 2024, **32**(6): 3741–3752. DOI: 10.1109/TFUZZ.2024.3381993.
- [25] WANG C Z, HUANG Y, DING W P, *et al.* Attribute Reduction with Fuzzy Rough Self-information Measures[J]. *Inf Sci*, 2021, **549**: 68–86. DOI: 10.1016/j.ins.2020.11.021.
- [26] SOWMYA R, PREMKUMAR M, JANGIR P. Newton-raphson-based Optimizer: A New Population-based Metaheuristic Algorithm for Continuous Optimization Problems[J]. *Eng Appl Artif Intell*, 2024, **128**: 107532. DOI: 10.1016/j.engappai.2023.107532.
- [27] FRÉNEY B, VERLEYSEN M. Classification in the Presence of Label Noise: A Survey[J]. *IEEE Trans Neural Netw Learn Syst*, 2014, **25**(5): 845–869. DOI: 10.1109/TNNLS.2013.2292894.
- [28] ŚLEZAK D, JANUSZ A. Ensembles of Bireducts: Towards Robust Classification and Simple Representation [M]//Future Generation Information Technology. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011: 64–77. DOI: 10.1007/978-3-642-27142-7\_9.
- [29] ZHANG X, MEI C L, LI J H, *et al.* Instance and Feature Selection Using Fuzzy Rough Sets: A Bi-selection Approach for Data Reduction[J]. *IEEE Trans Fuzzy Syst*, 2023, **31**(6): 1981–1994. DOI: 10.1109/TFUZZ.2022.3216990.
- [30] ZHANG X, HE Z Q, LI J H, *et al.* Bi-selection of Instances and Features Based on Neighborhood Importance Degree[J]. *IEEE Trans Big Data*, 2024, **10**(4): 415–428. DOI: 10.1109/TBDDATA.2023.3342643.
- [31] QUAN J S, QIAO F C, YANG T, *et al.* A Biselection Method Based on Consistent Matrix for Large-scale Datasets[J]. *IEEE Trans Fuzzy Syst*, 2025, **33**(6): 1992–2005. DOI: 10.1109/TFUZZ.2025.3543893.