

基于邻域粗糙集和 K 近邻分类的核素识别方法研究

陈 宸 吴 桓 王国帆 姚正勇
(重庆建安仪器有限公司 重庆 400060)

摘要 针对 γ 能谱数据的高维特性和噪声干扰导致传统核素识别方法普遍存在识别精度不足与计算效率较低的问题,提出一种融合邻域粗糙集与 K 近邻分类的核素识别方法,以满足便携式核素识别设备在资源受限环境下的实际应用需求。该方法首先采用主成分分析法(Principal Component Analysis, PCA)对原始 1 024 维 γ 能谱数据进行降维,提取出累积贡献率 $\geq 99\%$ 的主成分,以消除冗余信息并抑制噪声干扰;随后引入邻域粗糙集的属性约简算法,通过定义样本邻域关系并计算属性依赖度,结合启发式前向贪心搜索策略对特征属性进行约简,最大限度保留关键判别信息并提升特征子集的分类鉴别能力;最后将约简后的低维特征输入 K 近邻分类器,依据距离加权投票机制计算各类核素的置信度,实现高效率、准确的核素识别。基于一套 $\text{LaBr}_3(\text{Ce})$ 探测系统与 1 024 道多道分析器对识别算法进行实验测试,采集包括 12 种单一核素及 2 种混合核素在内的 224 组 γ 能谱样本,在 STM32F407ZGT6 单片机硬件平台上进行测试,在邻域半径 $\delta=0.2$ 时对包含特征峰重叠的混合核素在内的测试集的平均识别精度达到 98.5%,单次核素识别运行时间控制在 140 ms 内,显著提升了核素识别的准确性与计算效率,为后续便携式核素识别仪的算法部署提供了可靠保障。

关键词 γ 谱仪, 邻域粗糙集, K 近邻分类, 核素识别

中图分类号 TL81

DOI: 10.3724/j.0253-3219.2026.hjs.49.250382

CSTR: 32193.14.hjs.CN31-1342/TL.2026.49.250382

Nuclide identification method based on neighborhood rough set and KNN classification

CHEN Chen WU Huan WANG Guofan YAO Zhengyong
(Chongqing Jian An Instrument Co., Ltd., Chongqing 400060, China)

Abstract [Background] The high-dimensionality and inherent noise in gamma spectrum data often lead to low identification accuracy and poor computational efficiency in conventional nuclide identification methods. Therefore, a new method that can provide high accuracy and computational efficiency in portable devices is needed. [Purpose] This study aims to develop an efficient method for nuclide identification by combining neighborhood rough set (NRS) theory and K-nearest neighbor (KNN) classifier. [Methods] This method involved three-stage process. Firstly, principal component analysis (PCA) was applied to gamma spectrum data for dimensionality reduction and reduction of linear redundancy. Subsequently, a heuristic search strategy based on NRS was employed to optimize the feature subset by eliminating redundant attributes. Lastly, the KNN classifier was used for efficient nuclide identification in the reduced low dimensional feature space. In the meantime, the identification method was implemented on an

第一作者: 陈宸, 男, 1987 年出生, 2013 年于重庆邮电大学获硕士学位, 正高级工程师, 从事人工智能及 γ 能谱分析

通信作者: 吴桓, E-mail: wuh_1982@163.com

收稿日期: 2025-10-10, 修回日期: 2025-11-10

First author: CHEN Chen, male, born in 1987, graduated from Chongqing University of Posts and Telecommunications with a master's degree in 2013, professor-level senior engineer, focusing on artificial intelligence and γ spectrum analysis

Corresponding author: WU Huan, E-mail: wuh_1982@163.com

Received date: 2025-10-10, revised date: 2025-11-10

STM32F407ZGT6 platform, and experimental verification was performed with a LaBr₃(Ce) detection system and a 1 024-channel multichannel analyzer. **[Results]** The experiment used a dataset of 224 gamma spectra, comprising 12 single-nuclide and 2 mixed-nuclide categories was collected in the experiment, and verification results show that an average identification accuracy of 98.5% was achieved by proposed method on the test set with a neighborhood radius of $\delta=0.2$. The processing time for a single identification is within 140 ms. **[Conclusions]** Experimental results confirm the superior generalization performance and high computational efficiency of the proposed method across various scenarios, including mixed-nuclide conditions, hence providing a reliable and efficient technical pathway for efficient nuclide identification in portable radiation detection devices.

Key words Gamma spectrometer, Neighborhood rough set, K-nearest neighbor, Nuclide identification

γ 核素识别技术在辐射监测、核应急响应及核扩散等领域发挥着至关重要的作用,其核心任务是从复杂的 γ 能谱中快速、准确地鉴别出放射性核素种类。传统分析方法以特征峰匹配法为代表^[1-3],该类方法依赖于特征峰能量与核素数据库进行匹配,虽原理直观、计算复杂度较低,但在低计数率、高本底或复杂的混合核素场景下,特征峰易被康普顿连续区域淹没或难以从统计涨落中提取,导致漏判与误判。为克服这些缺陷,机器学习(如人工神经网络、支持向量机等)与深度学习(如卷积神经网络等)智能算法逐渐成为主流^[4-9],通过处理全谱数据以自动学习深层特征,在应对重叠峰和复杂能谱时表现出较高的识别精度与鲁棒性。然而,其庞大的模型参数和复杂的计算结构导致较大的计算开销,使其难以在资源有限、算力较低的便携式核素识别设备中实现高效运行。另一方面,基于时间序列分析的核素识别方法近年来取得了显著进展^[10-14],该类方法通过利用脉冲信号到达时间、能量分布等先验信息,构建检测模型以实现快速核素识别。但这类方法在面对特征峰重叠的混合核素场景时难以有效解析能量相近的叠加信号,导致分类精度显著下降。因此,研究一种在有限算力下仍能保持高识别精度与计算效率的核素识别方法,已成为该领域的迫切需求。

为处理高维数据下的“维度灾难”问题,粗糙集理论应运而生。该理论自Pawlak提出以来^[15],已成为处理不确定性数据的强有力工具,尤其在属性约简领域始终发挥着核心作用^[16-17],其本质是通过粒化机制将论域划分为等价类,并基于上下近似逼近原理,实现对冗余属性的剔除。经典粗糙集基于等价关系处理离散数据,难以直接适用于连续数值型数据。邻域粗糙集^[18-22](Neighborhood Rough Set, NRS)创新性地引入样本邻域粒化机制,通过定义样本邻域构建邻域粒子系统,实现对连续空间的粒化与逼近。NRS的核心优势在于引入邻域半径 δ 作为容错机制,可有效刻画样本间的相似性,并通过属性

依赖度量化特征区分能力,为决策表提供可解释的属性约简框架。利用邻域粗糙集对使用主成分分析法(Principal Component Analysis, PCA)降维后的 γ 能谱数据进行属性约简,可在显著降低数据维度的同时提升后续分类模型的泛化能力。

本文将邻域粗糙集与K近邻分类方法(K-Nearest Neighbor, KNN)^[23-24]相结合,首先利用PCA对原始高维 γ 能谱数据进行降维以消除线性冗余;进而引入邻域粗糙集的属性约简方法,依据属性依赖度进行启发式搜索,实现对特征子集的约简,保留最具判别性的特征组合;在此基础上,采用KNN分类器在约简后的低维特征空间中进行快速识别,该分类器无需复杂训练过程、仅依赖距离度量与近邻投票机制,可显著降低计算复杂度。理论分析与实验表明,本方法具有更高的识别精度与计算效率,十分适合在资源受限的便携式设备上实现实时、准确的核素识别。

1 理论基础

1.1 邻域粗糙集

定义1:给定实数空间上的一个非空有限集合 $U = \{u_1, u_2, \dots, u_n\}$,对于 U 上的任意对象 u_i ,定义 $\delta(u_i) = \{u | u \in U \wedge \Delta(u, u_i) \leq \delta\}$ 。其中, δ 称为邻域半径^[18-22]; Δ 为 N 维实数空间 R^N 上的度量函数。本文选择欧氏距离 $\Delta(x, y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2}$ 作为度量函数。

定义2:给定一个三元组结构 $\langle U, A, D \rangle$,其中 $U = \{u_1, u_2, \dots, u_n\}$ 表示对象全集(称为论域,本文指 γ 能谱样本集); A 是连续数值型的条件属性集合(本文指对 γ 能谱进行PCA降维后的特征向量); D 为决策属性(本文指核素类别标签),称 $\langle U, A, D \rangle$ 为一个邻域决策系统(Neighborhood Decision System, NDS)^[18-22]。

定义3:在邻域决策系统 $NDS = \langle U, A, D \rangle$ 中,决

策属性 D 将论域 U 划分为 N 个互斥的等价类 $\{U_1, U_2, \dots, U_N\}$ 。对于任意条件属性子集 $B \subseteq A$, 定义决策属性 D 关于 B 的上下近似集及边界域^[18-22]。

上近似集: $\bar{N}_B D = \bigcup_{k=1}^N \bar{N}_B U_i$, 其中 $\bar{N}_B U_i = \{u_i | \delta(u_i) \cap U_i \neq \emptyset \wedge u_i \in U\}$;

下近似集: $\underline{N}_B D = \bigcup_{k=1}^N \underline{N}_B U_i$, 其中 $\underline{N}_B U_i = \{u_i | \delta(u_i) \subseteq U_i \wedge u_i \in U\}$;

边界域: $BN(D) = \bar{N}_B D - \underline{N}_B D$, 为上下近似集之差。

粗糙集的概念示意图如图1所示, 黑色折线所围成的区域表示论域的子集 U_i ; 红色区域表示由集合 U_i 所确定的下近似集 $\underline{N}_B U_i$; 红色加灰色区域表示由集合 U_i 所确定的上近似集 $\bar{N}_B U_i$; 灰色区域为集合 U_i 所确定的边界域 $\bar{N}_B U_i - \underline{N}_B U_i$ 。

定义4: 决策属性 D 对条件属性子集 B 的依赖度^[18-22]为: $\gamma_B(D) = |\underline{N}_B D|/|U|$ 。 $\gamma_B(D)$ 量化了条件属

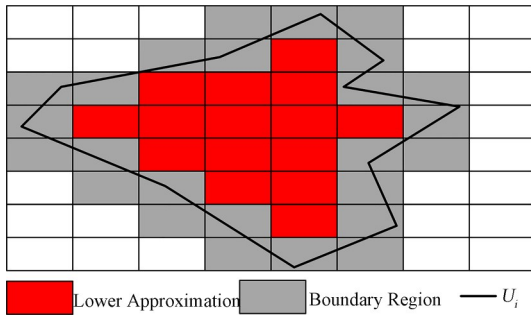


图1 粗糙集概念示意图(彩图见网络版)
Fig.1 Schematic diagram of the rough set concept (color online)

性子集 B 对决策类的分辨能力。

定义5: 条件属性 a 相对于条件属性子集 $B(B \subseteq A)$ 的属性重要度^[18-22]定义为:

$$\text{SIG}(a, B, D) = \gamma_{B \cup a}(D) - \gamma_B(D)$$

1.2 K近邻分类

KNN分类方法的核心思想是一种基于实例的惰性学习算法。它假定相似样本在特征空间中彼此邻近, 其分类决策无需预先训练, 而是通过计算待识别样本与特征空间中所有已标记样本之间的距离, 选取距离最近的 K 个样本(即近邻), 并根据距离加权投票的机制来判定待识别样本的所属类别^[23-24]。该方法结构简单, 且适用于混合核素场景。本文使用的KNN分类算法中度量样本间的距离公式为欧式距离, 与邻域粗糙集中度量样本间差异性的度量函数保持一致。

2 核素识别方法

本文提出的核素识别方法主要涵盖特征提取与分类识别两个关键阶段。在特征提取阶段, 首先对原始 γ 能谱数据构成的论域 U 进行标准化处理, 随后采用PCA方法实现数据降维, 并进一步借助邻域粗糙集的属性约简算法筛选降维后的特征, 最终获得约简后的特征集合 $reduct$ 及对应的论域 U' 。在分类识别阶段, 将 $reduct$ 与 U' 输入KNN分类器, 通过计算待识别样本与训练样本之间的欧氏距离来衡量相似性, 结合基于距离加权的近邻投票机制判定并输出核素类别 ω 及置信度 $conf$ 。整体工作流程如图2所示。

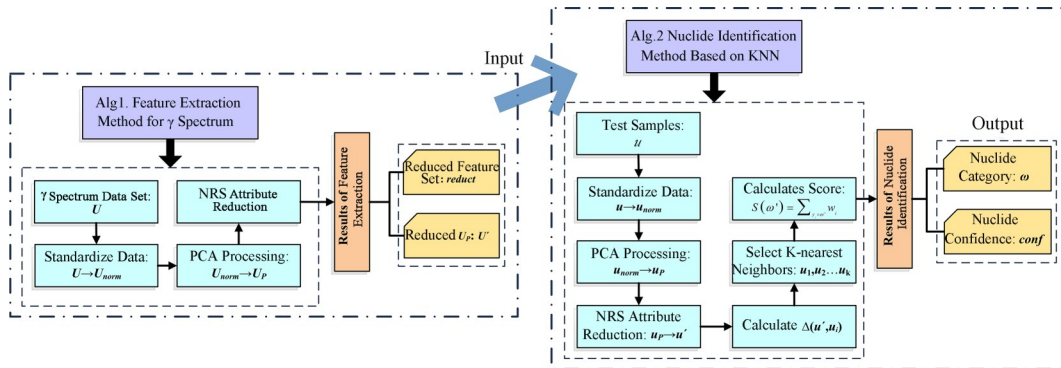


图2 核素识别方法整体工作流程图(彩图见网络版)
Fig.2 Overall workflow diagram of the nuclide identification method (color online)

2.1 算法1 基于邻域粗糙集的 γ 能谱特征提取方法

本算法旨在通过邻域粗糙集对PCA降维后的 γ 能谱特征进行属性约简。算法流程如下:

输入: 论域 $U = \{(x_i, y_i)\}_{i=1}^n$ 。其中, x_i 为原始 γ 能谱数据, $y_i \in \{\omega_1, \omega_2, \dots, \omega_c\}$ 为核素类别标签, c 为核素种类个数, n 为样本总数。

输出: 约简后的特征集合 $reduct$ 、约简后的论域

U' 。

Step1: 计算训练样本集的均值 μ 与标准差 σ , 对样本进行标准化处理: $U_{\text{norm}} = (U - \mu)/\sigma$;

Step2: 在 U_{norm} 上训练 PCA 模型, 按特征贡献率排序, 选取累积贡献率 $\geq 99\%$ 的前 P 个主成分, 得到投影矩阵 $W \in \mathbb{R}^{P \times 1024}$ 。将标准化数据投影至 PCA 空间: $U_p = W \times U_{\text{norm}}^T$, 同时得到降维后的条件属性集合 $A = \{1, 2, \dots, P\}$;

Step3: 构建邻域决策系统 $\langle U_p, A, D \rangle$, D 为决策属性(核素类别), 初始化约简集 $\text{reduct} = \emptyset$;

Step4: 对 $\forall a_i \in A \setminus \text{reduct}$, 计算 $\text{SIG}(a_i, \text{reduct}, D) = \gamma_{\text{reduct} \cup a_i}(D) - \gamma_{\text{reduct}}(D)$;

Step5: 选择 a_k , 使其满足 $\text{SIG}(a_k, \text{reduct}, D) = \max \{ \text{SIG}(a_i, \text{reduct}, D) \}$;

Step6: 若 $\text{SIG}(a_k, \text{reduct}, D) > 0$, 则令 $\text{reduct} = \text{reduct} \cup a_k$, 返回 Step4; 否则跳转至 Step7。

Step7: 根据 reduct 对 U_p 中的条件属性进行筛选, 得到约简后的论域 U' 。输出 U' 和 reduct 。

算法1中邻域半径 δ 的大小对属性约简的性能

至关重要: 当 δ 取值过小时, 其细粒度的划分方式导致每个样本的邻域范围过窄, 使大量样本被孤立地划入下近似集(图3(a)中红色方块), 仅少数处于真实边界附近的样本被标记为边界域(图3(a)中浅灰色方块)。这种情形虽然表面上扩大了确定性区域, 但会使模型过度关注训练样本中局部的、偶然的甚至是噪声引起的特征差异, 从而学习到缺乏泛化能力的特定规则, 导致最终分类模型过拟合、在未知数据上表现不佳。

当 δ 取值过大时, 其粗粒度的划分方式导致每个样本的邻域范围过宽, 使大量原本属于不同类别的样本被包含进同一邻域内, 从而扩大了边界域(图3(b)中浅灰色方块), 而能够明确分类的样本则被压缩到较小的核心区域(图3(b)中红色方块)。这种情形意味着分类不确定性显著增加, 算法无法找到足够多的确定性样本来构建清晰的决策规则, 导致属性约简时难以剔除冗余特征, 最终使分类模型欠拟合、精度下降。

因此, 需找到一个合适的 δ 值, 使得下近似集最大化覆盖确定性样本的同时控制边界域范围, 从而实现有效的属性约简。

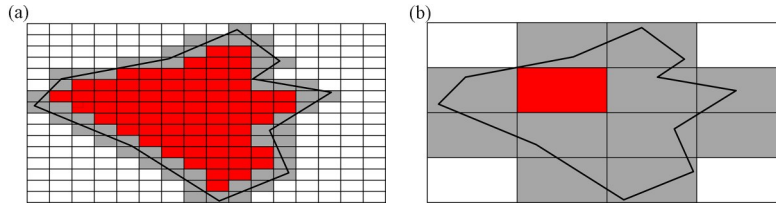


图3 不同 δ 下的样本划分示意图(彩图见网络版) (a) 细粒度划分, (b) 粗粒度划分

Fig.3 Schematic diagram of sample partitioning under different δ (color online)

(a) Fine-grained partitioning, (b) Coarse-grained partitioning

2.2 算法2 基于KNN的核素识别方法

本算法利用算法1约简后输出的结果, 结合K近邻分类器对待测样本进行核素识别。算法流程如下。

输入:

1) 约简后的论域 U' 、预计算的均值 μ 和标准差 σ ;

2) 预训练的PCA投影矩阵 $W \in \mathbb{R}^{P \times 1024}$;

3) 特征集合 reduct ;

4) 待识别的 γ 能谱样本 $u \in \mathbb{R}^{1024}$ 。

输出: 核素类别标签 ω 、置信度 conf 。

Step1: 对待测样本进行标准化处理: $u_{\text{norm}} = (u - \mu)/\sigma$;

Step2: 将标准化数据投影至 PCA 空间: $u_p = W \times u_{\text{norm}}^T$;

Step3: 基于 reduct 对 u_p 进行特征筛选, 得到低维特征向量 u' ;

Step4: 使用K近邻分类方法, 计算 u' 与 U' 中所有样本的欧氏距离, 选择距离最小的 K 个样本作为近邻; 随后为每个近邻 u_i 分配一个权重 $w_i = 1/(d_i + \varepsilon)$, 其中, d_i 为欧式距离, ε 为极小正数, 以防止除零错误; 接着对每个核素类别 ω' , 计算加权投票得分: $S(\omega') = \sum_{y_i = \omega'} w_i$, 最后计算置信度:

$$\text{conf}(\omega') = S(\omega') / \sum_{i=1}^K w_i;$$

Step5: 输出置信度超过阈值的核素类别 ω 。

2.3 时间复杂度分析

算法的预测流程可分解为4个核心计算阶段, 其时间复杂度分别如下:

1) 对输入样本进行标准化处理, 涉及每个道址的计算。该操作的时间复杂度为 $O(chl)$, 其中, chl

为能谱道址总数(本文 $chl=1\ 024$)。

2)将标准化后的数据与投影矩阵 $W \in \mathbb{R}^{P \times 1024}$ 相乘,将其从原始高维空间转换到 P 维主成分空间。该矩阵乘法的时间复杂度为 $O(chl \times P)$,其中, P 为PCA降维后保留的主成分数量。

3)根据约简后的特征集合 $reduct$,从 P 维特征中筛选出 d 个子特征。此操作为索引映射,其时间复杂度为 $O(d)$ 。

4)进行K近邻分类。首先计算待识别样本与 n 个训练样本在 d 维空间中的欧氏距离,时间复杂度为 $O(n \times d)$;随后,从 n 个距离值中找出 k 个最小值,可使用部分快速排序(如基于堆的排序算法),时间复杂度为 $O(n \times \log k)$;最后为 k 个近邻分配权重,并进行距离加权投票,时间复杂度为 $O(k \times c)$,其中 c 为核素类别数。因此,KNN分类阶段的总时间复杂度为 $O(n \times d + n \times \log k + k \times c)$ 。

综上所述,算法的总时间复杂度 $O(T)=O(chl)$

$+O(chl \times P)+O(d)+O(n \times d + n \times \log k + k \times c)$ 。通过对各单项复杂度的量级比较(通常有 $chl > n > P > d$ 且 $n \gg k$ 、 $n \gg c$),可以发现PCA投影操作 $O(chl \times P)$ 是整个时间复杂度的主导项,后续的分类算法对计算效率的影响较小。

3 实例分析

为清晰阐述本方法的计算流程,通过一个简化的数值实例,逐步演示从属性约简到核素识别的完整过程。给定一个PCA降维后的邻域决策系统 $NDS = \langle U, A, D \rangle$ 。其中论域 $U = \{e_1, e_2, \dots, e_9\}$ 包含9个样本;条件属性集合 $A = \{a_1, a_2, a_3, a_4, a_5\}$ 为5维特征向量;决策属性 D 将样本划分为三类核素: ^{241}Am 、 ^{133}Ba 、 ^{137}Cs 。样本的具体属性值与类别标签如表1所示。待识别的样本向量为 $u = \{0.0, 0.001\ 123\ 97, -0.005, 0.255, 1\}$ 。计算参数设定为邻域半径 $\delta = 0.03$,近邻数 $k = 3$ 。

表1 邻域决策系统示例
Table 1 Example of neighborhood decision system

编号 No.	a_1	a_2	a_3	a_4	a_5	决策属性 Decision attribute
e_1	0.002 336	0.009 507	0.016 100	-0.208 161	-1	^{241}Am
e_2	0	0.007 295	0.009 722	-0.231 081	-1	^{241}Am
e_3	0	0.056 084	0.016 744	-0.173 528	-1	^{241}Am
e_4	0	0.003 199	-0.003 494	0.244 238	1	^{133}Ba
e_5	0	0.001 099	-0.003 913	0.242 789	1	^{133}Ba
e_6	0	0.003 558	-0.003 709	0.245 653	1	^{133}Ba
e_7	0	0.041 249	0.003 037	0.265 501	1	^{137}Cs
e_8	0	0.061 119	0.002 337	0.260 522	1	^{137}Cs
e_9	0.007 880	0.030 943	-0.002 066	0.257 323	1	^{137}Cs

首先,决策属性 D 将论域 U 划分为三个等价类:
 $U/D = \{U_1, U_2, U_3\} =$

$$\{\{e_1, e_2, e_3\}, \{e_4, e_5, e_6\}, \{e_7, e_8, e_9\}\}$$

随后,执行算法1,初始化约简集 $reduct = \emptyset$ 。以计算属性 a_4 的重要度为例。

根据定义1和定义3,由 $\{a_4\}$ 诱导出的邻域关系为:

$$\delta(e_1): \{e_1, e_2\};$$

$$\delta(e_2): \{e_1, e_2\};$$

$$\delta(e_3): \{e_3\};$$

$$\delta(e_4) \sim \delta(e_9): \{e_4, e_5, e_6, e_7, e_8, e_9\}。$$

计算下近似集: $\underline{N}_{\{a_4\}}U_1 = \{e_1, e_2, e_3\}$ (因为其邻域均包含于 U_1), $\underline{N}_{\{a_4\}}U_2 = \emptyset$, $\underline{N}_{\{a_4\}}U_3 = \emptyset$ 。故
 $\underline{N}_{\{a_4\}}D = \bigcup_{i=1}^3 \underline{N}_{\{a_4\}}U_i = \{e_1, e_2, e_3\}$ 。

$$\text{依赖度 } \gamma_{\{a_4\}}(D) = \left| \underline{N}_{\{a_4\}}D \right| / |U| = \frac{3}{9} = \frac{1}{3}。$$

$$\text{属性重要度 } \text{SIG}(a_4, reduct, D) = \gamma_{\{a_4\}}(D) - \gamma_{\emptyset}(D) = \frac{1}{3}。$$

同理,计算其他属性重要度: $\text{SIG}(a_1, reduct, D) = 0$, $\text{SIG}(a_2, reduct, D) = 0$, $\text{SIG}(a_3, reduct, D) = 0$, $\text{SIG}(a_5, reduct, D) = 1/3$ 。选择重要度最大的属性 a_4 加入约简集,即 $reduct = \{a_4\}$ 。重复上述过程,最终得到约简集 $reduct = \{a_4, a_2, a_1\}$ 。根据此结果对特征进行筛选,得到约简后的邻域决策系统,如表2所示。

接下来,采用算法2对样本 u 进行类别预测。

首先,按 $reduct$ 对 u 进行筛选,得到 $u' = \{0.0, 0.001\ 123\ 97, 0.255\}$ 。

然后,计算 u' 与表2中各样本的欧式距离,并按距离升序排序,结果如表3所示。

选取前三个最近邻样本(e_6, e_4, e_5),其类别均为 ^{133}Ba 。采用距离加权投票,计算得 ^{133}Ba 的总权重为 $1/(d_6+\varepsilon)+1/(d_4+\varepsilon)+1/(d_5+\varepsilon)$,置信度 $conf=1.0$ 。

表2 约简后的邻域决策系统示例
Table 2 Example of reduced neighborhood decision system

编号 No.	a_1	a_2	a_4	决策属性 Decision attribute
e_1	0.002 336	0.009 507	-0.208 161	^{241}Am
e_2	0	0.007 295	-0.231 081	^{241}Am
e_3	0	0.056 084	-0.173 528	^{241}Am
e_4	0	0.003 199	0.244 238	^{133}Ba
e_5	0	0.001 099	0.242 789	^{133}Ba
e_6	0	0.003 558	0.245 653	^{133}Ba
e_7	0	0.041 249	0.265 501	^{137}Cs
e_8	0	0.061 119	0.260 522	^{137}Cs
e_9	0.007 880	0.030 943	0.257 323	^{137}Cs

因此,待测样本被判别为核素 ^{133}Ba ,且置信度为100%。

表3 u' 与各样本的欧式距离
Table 3 Euclidean distances between u' and each sample

$\Delta(u', e)$	决策属性 Decision attribute	编号 No.
0.009 658	^{133}Ba	e_6
0.010 960	^{133}Ba	e_4
0.012 211	^{133}Ba	e_5
0.030 930	^{137}Cs	e_9
0.041 476	^{137}Cs	e_7
0.060 248	^{137}Cs	e_8
0.432 038	^{241}Am	e_3
0.463 242	^{241}Am	e_1
0.486 120	^{241}Am	e_2

4 实验分析

4.1 实验设置与数据预处理

为验证本方法的有效性,实验采用配备北京玻璃研究院有限公司生产的3.81 cm×3.81 cm LaBr₃(Ce)探测器及自研的1 024道多道分析器对 γ 能谱进行采集。核素识别算法运行于STM32F407 ZGT6单片机硬件平台,该处理器基于ARM Cortex-M4内核,主频168 MHz,配备192 KB RAM与1 MB Flash存储空间,为实时谱处理与核素识别提供硬件支持。

测试源包括12种单一放射性核素(^{241}Am 、 ^{133}Ba 、 ^{57}Co 、 ^{60}Co 、 ^{137}Cs 、 ^{152}Eu 、 ^{18}F 、 ^{131}I 、 ^{226}Ra 、 $^{99\text{m}}\text{Tc}$ 、 ^{232}Th 、 ^{238}U)及两组混合核素($^{57}\text{Co}+^{152}\text{Eu}$ 、 $^{133}\text{Ba}+^{226}\text{Ra}$)。混合核素的选取基于其特征峰能量的邻近性: ^{152}Eu 的121.78 keV特征峰与 ^{57}Co 的122.05 keV特征峰几乎完全重叠, ^{226}Ra 的295.22 keV、351.93 keV特征峰则

分别与 ^{133}Ba 的302.85 keV、356.01 keV特征峰非常接近,它们的叠加能谱如图4所示(为提升谱图可比性并突出关键区域,图中仅展示前500道区间的数据,且所有能谱的计数均进行归一化处理)。

在数据采集前,首先使用天然本底 ^{40}K 的1 460 keV特征峰作为参考峰,通过调整多道分析器的高压和增益参数,使能谱峰位漂移被控制在±2道以内。然后对上述14种测试用例(12种单一核素与两组混合核素)分别采集能谱。每种测试用例测量了三组共16个1 min的能谱数据,按剂量率大小分为三个梯度:第一组包含6个统计涨落较高、剂量率仅高于本底水平 $0.5 \mu\text{Sv} \cdot \text{h}^{-1}$ (误差≤±30%)的能谱;第二组包含5个剂量率高于本底 $2 \mu\text{Sv} \cdot \text{h}^{-1}$ (误差≤±30%)的能谱;第三组包含5个剂量率高于本底 $10 \mu\text{Sv} \cdot \text{h}^{-1}$ (误差≤±30%)的能谱。混合核素样本摆放时,使各核素至探测器中心的剂量率保持一致(误差≤±30%)。共计获取224个能谱,每组选取2个样本共构成84个训练样本,其余140个样本作为测试集,用于后续算法验证。

4.2 核素识别试验

首先基于算法1对训练集的84个样本进行特征提取与属性约简:对原始1 024道 γ 能谱数据进行PCA降维。图5为主成分贡献率的帕累托图。实验结果表明,当选取前25个主成分时,累积贡献率可达99%以上,能够有效保留原始数据的判别信息。

随后采用邻域粗糙集方法对生成的25维邻域决策系统进行属性约简。通过调整邻域半径 δ 从0~0.8(步长0.05),分别计算不同 δ 取值下对应的约简后的特征集合。

最后将其作为输入,采用算法2(近邻数 $k=5$ 、置信度阈值 $\text{conf}=0.8$)对140个测试样本进行核素识别,记录其平均分类精度。图6(a)展示了不同邻域半径 δ 取值下约简得到的特征数量 n ,图6(b)展示了相应条件下的平均分类精度。

以上实验结果表明,邻域半径 δ 的取值显著影响属性约简效果与分类性能。当 δ 取值较小($0 \leq \delta \leq 0.15$)时,样本邻域范围过窄,多数邻域仅包含自身或极少数相似样本,导致下近似集规模偏大,属性依赖度取值过高,使算法1过早终止。尽管约简后特征数较少,但其未能充分捕获类别间判别信息,分类精度较低。当 δ 取值较大($\delta \geq 0.3$)时,样本邻域重叠严重,多数邻域包含多类样本,下近似集规模偏小,属性依赖度取值过小,使算法1需添加较多属性以提升依赖度,导致约简后含有较多冗余特征,模型泛化能力下降。

当 $\delta=0.2$ 时,属性约简按重要度排序,共获得9

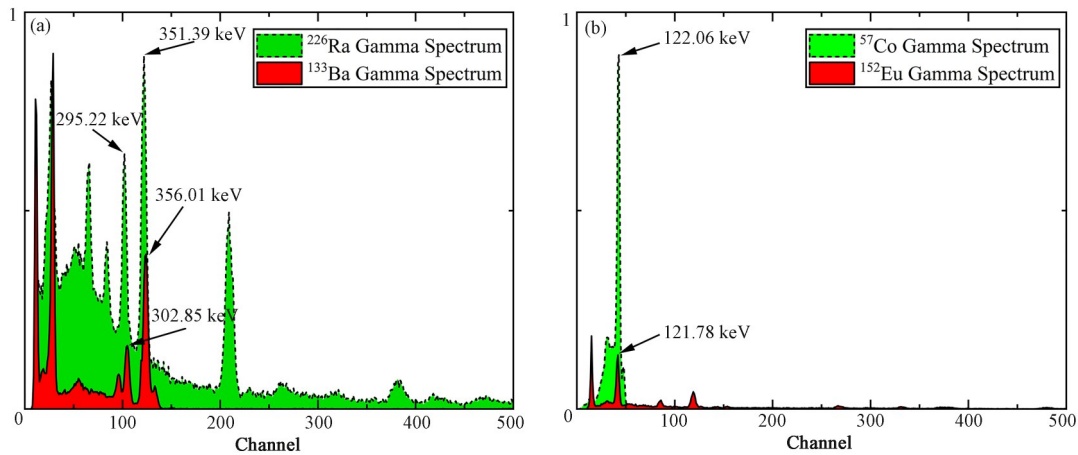


图4 不同核素的叠加能谱(彩图见网络版) (a) $^{133}\text{Ba}+^{226}\text{Ra}$, (b) $^{57}\text{Co}+^{152}\text{Eu}$
Fig.4 Overlapped spectrum of different nuclides (color online) (a) $^{133}\text{Ba}+^{226}\text{Ra}$, (b) $^{57}\text{Co}+^{152}\text{Eu}$

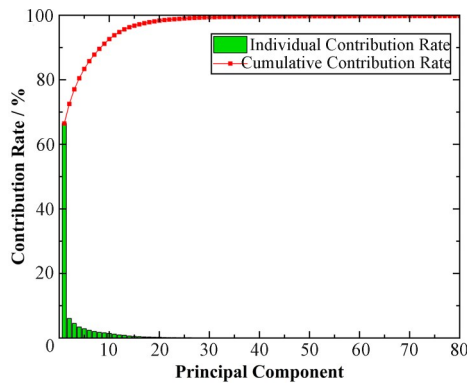


图5 主成分贡献率的帕累托图(彩图见网络版)
Fig.5 Pareto diagram of principal component contribution rates (color online)

个特征: $\{a_5, a_{13}, a_6, a_{20}, a_{11}, a_1, a_{18}, a_{23}, a_9\}$, 平均分类精度达到98.5%, 且混合核素全部正确识别, 实验结果见表4。说明此邻域范围既能有效捕捉局部样本相似性, 又能保持良好泛化能力, 属性依赖度可准确反映特征判别能力, 可使算法1能逐步添加显著重要特征, 最终得到高判别性与低冗余度的特征子集。

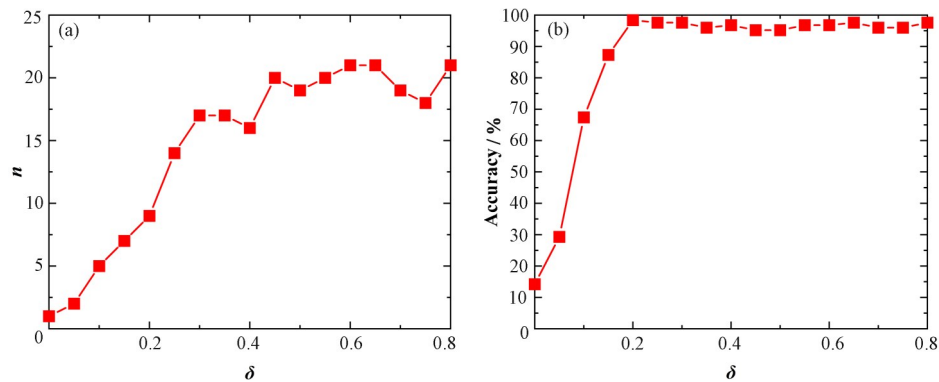


图6 不同 δ 取值对应的实验结果 (a) 约简后特征个数, (b) 平均分类精度
Fig.6 Experimental results under different δ (a) Number of features after reduction, (b) Average classification accuracy

4.3 计算效率消融实验

在STM32F407ZGT6单片机硬件平台对计算效率进行评估, 通过消融实验定量分析各算法模块的时间开销, 实验模型如下:

实验组1仅执行1024~25维的PCA投影操作, 记为 t_1 ;

实验组2在组1基础上增加特征选择模块, 依次生成1~25维的特征集合并进行K近邻分类($k=5$), 记为 t_2 。

图7表示两个实验组的算法运行时间随特征个数的变化。

实验结果表明, PCA投影操作为主要计算瓶颈, 其单次执行耗时 $t_1=112\text{ ms}$, 占整体计算开销的主要部分。当在实验组2中逐步增加特征个数时, 分类算法运行时间呈现近似线性增长趋势, 但整体波动范围控制在110~140 ms的区间内。这一现象与算法时间复杂度分析结果一致, 验证了理论模型的准确性, 同时也表明本方法特别适合部署在资源受限的嵌入式设备上。

表4 核素识别实验结果($\delta=0.2$)
Table 4 Experimental results of nuclide identification ($\delta=0.2$)

核素 Nuclide	识别成功次数 Number of successful identifications	总样本数 Total number of samples	核素 Nuclide	识别成功次数 Number of successful identifications	总样本数 Total number of samples
²⁴¹ Am	10	10	¹³³ Ba	10	10
⁵⁷ Co	10	10	⁶⁰ Co	10	10
¹³⁷ Cs	10	10	¹⁵² Eu	10	10
¹⁸ F	10	10	¹³¹ I	10	10
²²⁶ Ra	10	10	^{99m} Tc	10	10
²³² Th	9	10	²³⁸ U	9	10
⁵⁷ Co+ ¹⁵² Eu	10	10	¹³³ Ba+ ²²⁶ Ra	10	10

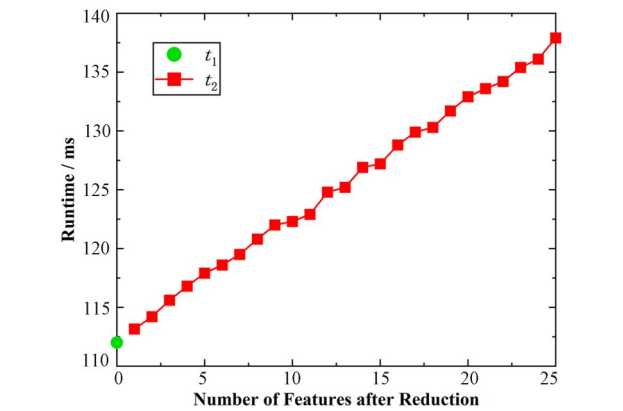


图7 算法运行时间随特征个数的变化(彩图见网络版)
Fig.7 Variation of runtime of the algorithm with the number of features (color online)

4.4 属性约简性能消融实验

为系统评估属性约简对核素识别分类性能的影响,通过消融实验对比不同特征选择策略的有效性,具体包括以下两组实验设置:

实验组1在取消算法1使用的邻域粗糙集属性约简模块的基础上,采用无监督特征选择方法直接基于PCA降维后的25个主成分,按主成分贡献率从高到低依次选取1~25个特征子集作为输入,采用K近邻分类器($k=5$)进行核素识别,并记录各特征组合对应的平均分类精度。

实验组2将算法1中使用的启发式前向搜索方法改为后向搜索方法:算法从完整的条件属性集合(即PCA降维后的25维特征)开始迭代,在每一步中,移除某个条件属性后计算属性重要度 $SIG(a, B - a, D) = \gamma_B(D) - \gamma_{B-a}(D)$,从当前属性集合中去掉使属性重要度取值最小的属性,迭代上述过程,直至移除任何属性均导致属性重要度降低时终止。

图8展示了平均分类精度随特征个数的变化,其中实验组1的结果记为 n_1 ,实验组2在不同 δ 对应的(特征数,分类精度)数据点的结果记为 n_2 。作为

对比,将图6中的数据点也绘制于图8中(记为 n_3)。

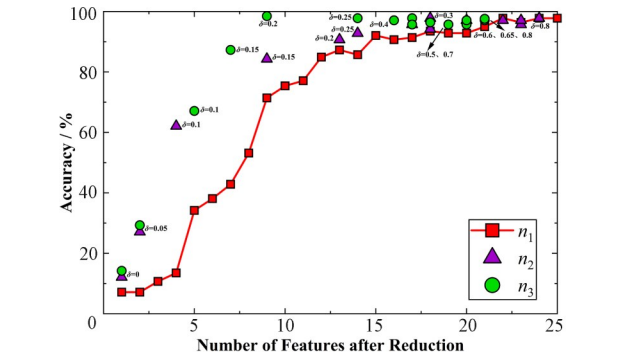


图8 平均分类精度随特征个数的变化(彩图见网络版)
Fig.8 Variation of average classification accuracy with the number of features (color online)

实验结果表明,不同的特征选择策略对分类性能产生显著影响。其中,算法1使用的前向搜索方法整体性能最优,后向搜索方法在分类精度上虽优于无监督特征选择方法,但整体性能仍低于前向搜索方法。

实验组1采用的无监督特征选择方法在特征个数为22、24和25时达到最大分类精度97.8%,低于前向搜索方法在 $\delta=0.2$ 时获得的98.5%的最高精度。其根本原因在于PCA降维是一种无监督的学习方式,其优化目标在于最大限度地保留原始数据的方差,然而方差最大的方向未必与类别判别边界最为相关,其选出的特征子集可能包含对分类任务而言冗余甚至构成干扰的高方差噪声(例如某些 γ 能谱的康普顿平台或噪声也可能具有较高的方差)。

实验组2采用的启发式后向搜索方法在 $\delta=0.3$ 和0.8时,约简后特征个数分别为18和24,获得的最大分类精度为97.8%。其整体性能优于无监督特征选择方法,低于前向搜索方法。这是因为后向搜索方法从属性全集开始进行迭代,由于冗余特征的存在,移除单个特征对整体依赖度影响较小,导致算法保留了更多的判别性较弱的冗余特征,最终影响了分类器的性能。

算法1使用的启发式前向搜索方法,其核心是一种有监督的贪心搜索策略的学习方式,该方式以最大化决策的下近似集为直接目标,能够主动甄别并筛选出那些对区分不同核素类别判别性最强的特征(即使这些特征的能量方差较低),从而在整体上提升了模型的分类准确率和泛化能力,尤其体现在对特征峰重叠的混合核素样本的精确识别上。

5 结语

本文提出的基于邻域粗糙集与K近邻分类的核素识别方法,通过融合PCA降维、邻域粗糙集属性约简和KNN分类器,有效解决了 γ 能谱数据高维特性与嵌入式设备资源约束之间的矛盾。实验结果表明,在邻域半径 $\delta=0.2$ 的最优参数下,该方法对包含特征峰重叠的混合核素在内的140组样本测试集的平均识别精度达到了98.5%,且单次识别的算法运行时间控制在140 ms以内。消融实验进一步验证,相较于无监督特征选择和后向搜索方法,本文采用的邻域粗糙集前向搜索方法在属性约简后仅保留9个特征,显著提升了特征子集的判别能力与分类模型泛化性能。在STM32F407ZGT6硬件平台上的计算效率分析显示,PCA投影操作作为系统主要计算瓶颈,而属性约简与KNN分类过程耗时较低,验证了本方法在有限算力下仍能兼顾高准确率与实时处理能力,为便携式设备部署核素识别算法提供了一条可靠的技术路径。

作者贡献声明 陈宸负责研究思路、理论推导、数据分析、论文撰写;吴桓负责选题指导、技术支持、论文审核;王国帆、姚正勇负责论文修改、实验操作、数据采集。

参考文献

- 1 陈亮, 魏义祥, 屈建石. 便携式 γ 谱仪中的核素识别算法[J]. 清华大学学报(自然科学版), 2009, 49(5): 635 - 638. CHEN Liang, WEI Yixiang, QU Jianshi. Nuclide identification algorithm for a portable gamma spectrometer[J]. Journal of Tsinghua University (Science and Technology), 2009, 49(5): 635 - 638.
- 2 Huang Y, Liu M Z, Liu X H, *et al.* Nuclide spectrum peak searching algorithm based on multiple morphological structuring elements[J]. Journal of Physics: Conference Series, 2020, 1634(1): 012078. DOI: 10.1088/1742-6596/1634/1/012078.
- 3 Luo R, Huang Y, Liu X H. Comparative study on peak-seeking methods of the mixed radioactive energy spectrum [J]. OALib, 2020, 7(8): 1 - 10. DOI: 10.4236/oalib.1106676.
- 4 Jhung S, Hur S, Cho G, *et al.* A neural network approach for identification of gamma-ray spectrum obtained from silicon photomultipliers[J]. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2020, 954: 161704. DOI: 10.1016/j.nima.2018.12.019.
- 5 Koo B T, Lee H C, Bae K, *et al.* Development of a radionuclide identification algorithm based on a convolutional neural network for radiation portal monitoring system[J]. Radiation Physics and Chemistry, 2021, 180: 109300. DOI: 10.1016/j.radphyschem.2020.109300.
- 6 Zhao S Q, Zhou S M, Chen R. Research on radionuclide identification method based on GASF and deep residual network[J]. Applied Sciences, 2025, 15(3): 1218. DOI: 10.3390/app15031218.
- 7 杨馨, 张坤明, 陈涛. 基于机器学习预测膨润土中阳离子的分配系数[J]. 核技术, 2025, 48(6): 060302. DOI: 10.11889/j.0253-3219.2025.hjs.48.240431. YANG Xin, ZHANG Kunming, CHEN Tao. Machine learning-based prediction of cation distribution coefficients in bentonite[J]. Nuclear Techniques, 2025, 48(6): 060302. DOI: 10.11889/j.0253-3219.2025.hjs.48.240431.
- 8 Li F, Luo C Y, Wen Y Z, *et al.* A nuclide identification method of γ spectrum and model building based on the transformer[J]. Nuclear Science and Techniques, 2025, 36(1): 7. DOI: 10.1007/s41365-024-01564-5.
- 9 Huang X, Yuan Y G, Zhu Y X, *et al.* A low-consumption multiple nuclides identification algorithm for portable gamma spectromete[J]. Nuclear Science and Techniques, 2025, 36(7): 121. DOI: 10.1007/s41365-025-01701-8.
- 10 Candy J V, Breitfeller E, Guidry B L, *et al.* Physics-based detection of radioactive contraband: a sequential Bayesian approach[J]. IEEE Transactions on Nuclear Science, 2009, 56(6): 3694 - 3711. DOI: 10.1109/TNS.2009.2034374.
- 11 万琳健, 张璇, 张春雷, 等. 基于序贯贝叶斯方法的核素快速检测方法研究[J]. 核技术, 2025, 48(1): 010404. DOI: 10.11889/j.0253-3219.2025.hjs.48.240190. CSTR: 32193.14.hjs.CN31-1342/TL.2025.48.240190. WAN Linjian, ZHANG Xuan, ZHANG Chunlei, *et al.* Fast detection method for radionuclides based on Sequential Bayesian approach[J]. Nuclear Techniques,

- 2025, **48**(1): 010404. DOI: 10.11889/j.0253-3219.2025.hjs.48.240190. CSTR: 32193.14.hjs.CN31-1342/TL.2025.48.240190.
- 12 曹仁龙, 李孝哲, 黄亭璋, 等. 一种用于核素识别的序贯贝叶斯方法[J]. 核电子学与探测技术, 2025, **45**(3): 343 – 352. DOI: 10.20173/j.cnki.ned.20250110.001.
CAO Renlong, LI Xiaozhe, HUANG Tingzhang, *et al.* A sequential Bayesian method for nuclide recognition[J]. Nuclear Electronics & Detection Technology, 2025, **45**(3): 343 – 352. DOI: 10.20173/j.cnki.ned.20250110.001.
- 13 罗庚, 石睿, 曾树鑫, 等. 基于核脉冲峰值序列二维卷积神经网络的核素识别方法[J]. 核技术, 2025, **48**(3): 030502. DOI: 10.11889/j.0253-3219.2025.hjs.48.240301. CSTR: 32193.14.hjs.CN31-1342/TL.2025.48.240301.
LUO Geng, SHI Rui, ZENG Shuxin, *et al.* Radionuclide identification method of two-dimensional convolutional neural network based on nuclear pulse peak sequence[J]. Nuclear Techniques, 2025, **48**(3): 030502. DOI: 10.11889/j.0253-3219.2025.hjs.48.240301. CSTR: 32193.14.hjs.CN31-1342/TL.2025.48.240301.
- 14 Li X Z, Zhang Q X, Tan H Y, *et al.* Fast nuclide identification based on a sequential Bayesian method[J]. Nuclear Science and Techniques, 2021, **32**(12): 143. DOI: 10.1007/s41365-021-00982-z.
- 15 Pawlak Z. Rough sets[J]. International Journal of Computer & Information Sciences, 1982, **11**(5): 341 – 356. DOI: 10.1007/BF01001956.
- 16 陈宸, 赵军. 一种新的基于二进制分辨矩阵的属性约简方法[J]. 计算机应用与软件, 2013, **30**(9): 123 – 127. DOI: 10.3969/j.issn.1000-386x.2013.09.035.
CHEN Chen, ZHAO Jun. A new attribute reduction method based on binary discernibility matrix[J]. Computer Applications and Software, 2013, **30**(9): 123 – 127. DOI: 10.3969/j.issn.1000-386x.2013.09.035.
- 17 Yu B, Hu Y, Kang Y, *et al.* A novel variable precision rough set attribute reduction algorithm based on local attribute significance[J]. International Journal of Approximate Reasoning, 2023, **157**: 88 – 104. DOI: 10.1016/j.ijar.2023.03.002.
- 18 胡清华, 于达仁, 谢宗霞. 基于邻域粒化和粗糙逼近的数值属性约简[J]. 软件学报, 2008, **19**(3): 640 – 649. DOI: 10.3724/SP.J.1001.2008.00640.
HU Qinghua, YU Daren, XIE Zongxia. Numerical attribute reduction based on neighborhood granulation and rough approximation[J]. Journal of Software, 2008, **19**(3): 640 – 649. DOI: 10.3724/SP.J.1001.2008.00640.
- 19 郑文彬, 李进金, 何秋红. 基于属性重要度的变精度邻域粗糙集属性约简算法[J]. 计算机科学, 2019, **46**(12): 261 – 265. DOI: 10.11896/jsjx.181102184.
ZHENG Wenbin, LI Jinjin, HE Qiuhong. Attribute reduction algorithm for neighborhood rough sets with variable precision based on attribute importance[J]. Computer Science, 2019, **46**(12): 261 – 265. DOI: 10.11896/jsjx.181102184.
- 20 Wu S Z, Wang L T, Ge S Y, *et al.* Neighborhood rough set with neighborhood equivalence relation for feature selection[J]. Knowledge and Information Systems, 2024, **66**(3): 1833 – 1859. DOI: 10.1007/s10115-023-01999-z.
- 21 Qian D M, Liu K Y, Wang J, *et al.* Attribute reduction based on directional semi-neighborhood rough set[J]. International Journal of Machine Learning and Cybernetics, 2025, **16**(4): 2523 – 2535. DOI: 10.1007/s13042-024-02406-x.
- 22 Sun L, Zhang X Y, Xu J C, *et al.* An attribute reduction method using neighborhood entropy measures in neighborhood rough sets[J]. Entropy, 2019, **21**(2): 155. DOI: 10.3390/e21020155.
- 23 Biau G, Devroye L. Lectures on the Nearest Neighbor Method[M]. Cham: Springer International Publishing, 2015. DOI: 10.1007/978-3-319-25388-6.
- 24 Li W, Zhang Y, Sun Y F, *et al.* Approximate nearest neighbor search on high dimensional data: experiments, analyses, and improvement[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, **32**(8): 1475 – 1488. DOI: 10.1109/TKDE.2019.2909204.