

基于SLIB与时频网格—能量特征融合的未知辐射源个体识别

姜 豪¹, 张佳飞¹, 王首斌¹, 沈 雷²

(1. 中国电子科技集团第36研究所, 浙江 嘉兴 314000; 2. 杭州电子科技大学通信工程学院, 浙江 杭州 310018)

摘 要: 针对射频指纹识别中未知辐射源识别的自适应聚类需求、单一特征在信号表征中的局限性, 提出一种基于电磁信号SLIB与时频网格—能量特征融合和熵惩罚K-means聚类的未知辐射源分类识别算法。首先, 对时频谱网格划分的时频网格—能量特征与围线积分双谱(SLIB)特征, 进行PCA+t-SNE双重降维并融合, 形成既包含信号的非线性高阶统计特征, 又保留了时频域的局部细粒度能量变化信息的综合特征向量, 以更全面地表征信号的本质特征; 其次, 构建一种基于熵惩罚的K-means聚类方法, 采用无动态簇中心调整与信息熵权重更新机制的监督自适应策略, 自适应调整聚类簇数; 最后, 引入Gamma动态调整因子、Beta权重机制优化样本归属计算, 提升聚类稳定性和鲁棒性。实验表明, 相较于单独使用SLIB、时频网格—能量特征方法, 所提方法分类识别准确率更高; 基于熵惩罚K-means的聚类方法相较于多种聚类模型的分类效果更好、鲁棒性更强。

关键词: SLIB; 时频网格; 双重降维; 特征融合; 熵惩罚

DOI: 10.11907/rjcdk.251327

中图分类号: TP391.14

文献标识码: A

文章编号: 1672-7800(2026)009-0098-08

扫描二维码阅读全文:



Identification of Unknown Emitter Based on SLIB and Time-Frequency Grid-Energy Feature Fusion

JIANG Hao¹, ZHANG Jiafei¹, WANG Shoubin¹, SHEN Lei²

(1. 36th Research Institute of China Electronics Technology Group Corporation, Jiaxing 314000, China;

2. School of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310018, China)

Abstract: Aiming at the adaptive clustering requirements for unknown radiation source identification in radio frequency fingerprint recognition and the limitations of single feature in signal representation, a classification and recognition algorithm for unknown radiation sources based on electromagnetic signal SLIB and time-frequency grid energy feature fusion and entropy penalty K-means clustering is proposed. Firstly, the time-frequency grid energy features and the line integral bispectrum (SLIB) features of the time-frequency grid division are subjected to PCA+t-SNE dual dimensionality reduction and fusion to form a comprehensive feature vector that contains both the nonlinear high-order statistical features of the signal and the local fine-grained energy change information in the time-frequency domain, in order to more comprehensively characterize the essential characteristics of the signal; Secondly, a K-means clustering method based on entropy penalty is constructed, which adopts a supervised adaptive strategy without dynamic cluster center adjustment and information entropy weight update mechanism to adaptively adjust the number of clusters; Finally, Gamma dynamic adjustment factor and Beta weighting mechanism are introduced to optimize sample assignment calculation, improving clustering stability and robustness. Experiments have shown that compared to using SLIB and time-frequency grid energy feature methods alone, the proposed method has higher classification and recognition accuracy; The clustering method based on entropy penalty K-means has better classification performance and stronger robustness compared to various clustering models.

Key Words: SLIB; time-frequency grid; double dimensionality reduction; feature fusion; entropy penalty

收稿日期: 2025-05-04

基金项目: 国防重点实验室基金项目(62471163)

作者简介: 姜豪(1983-), 男, 中国电子科技集团公司第36研究所高级工程师, 研究方向为信号处理与人工智能; 张佳飞(1998-), 男, 杭州电子科技大学通信工程学院硕士研究生, 研究方向为信号处理; 王首斌(1984-), 男, 博士, 中国电子科技集团公司第36研究所高级工程师, 研究方向为电子战系统总体技术; 沈雷(1979-), 男, 博士, 杭州电子科技大学通信工程学院教授、博士生导师, 研究方向为图像处理和信号处理。本文通讯作者: 张佳飞。

0 引言

在非协作场景中,传统监督学习方法受限于带类别标注数据的不足和未知类型辐射源信号的存在,无法直接训练已知辐射源的样本^[1-3]。在此情况下,无监督学习对辐射源的分类识别尤为关键。作为无监督学习中的一种经典方法,聚类算法可将辐射源信号按照其内在特征进行分组,即使在没有标注信息的情况下,也能高效地将无标签数据划分为不同类别^[4-8]。

李昕等^[9]提出一种基于密度峰值算法的通信电台个体识别方法,利用矩形双谱提取信号特征,并采用欧几里德距离聚类分析特征,在2 000个样本信号的实验中实现了51.9%的识别率。然而,该方法的聚类结果对超参数选择敏感,无法有效处理高维特征数据中的冗余信息,从而影响了识别性能。彭进先等^[10]提出基于多特征参数聚类分析的特定辐射源识别方法,通过分析辐射源相关的10种特征参数构建特征数组,并利用聚类分析进行辐射源识别。然而,该方法依赖选定特征的权重分配和专家评分系统,未充分考虑复杂电磁环境中信号的多变换域特性,在噪声和干扰较强环境下的识别准确性和鲁棒性受限。朱玲^[11]提出基于变分模态分解(Variational Mode Decomposition, VMD)与t-分布随机领域嵌入(t-distributed stochastic neighbor embedding, t-SNE)降维的未知辐射源识别方法,以提取信号特征并降低维度。然而,VMD在信号分解时存

在过度分解信号或模式混叠问题,将丧失一些有用的信息且运算量大,t-SNE倾向于优化局部结构而忽视了全局结构,单独使用t-SNE会导致类别间边界模糊,存在如何选择最优聚类数问题。

为此,本文提出一种基于SLIB与时频网格—能量特征融合和熵惩罚K-means聚类的未知辐射源分类识别算法。为了更好地表征信号的指纹特征,通过主成分分析(Principal Component Analysis, PCA)与t-SNE双重降维时频网格—能量特征、围线积分双谱(Surrounding-line Integrated Bispectrum, SLIB)特征,以形成既包含信号的非线性高阶统计特征,又保留了时频域的局部细粒度能量变化信息的综合特征向量,避免了冗余信息和高维特征的影响,为构建的熵惩罚的K-means聚类提供了高效且具辨识力的数据输入。K-means聚类根据数据的内在结构,通过熵惩罚机制动态调整簇数、优化样本归属计算,减少错误分类概率。此外,引入Gamma与Beta动态调整因子进一步提升聚类算法的稳定性与适应性,使系统在面对噪声和复杂数据时仍能够保持较高的识别准确率。

1 本文方法

本文为了实现非合作场景下未知辐射源的分类识别,提出基于SLIB与时频网格—能量特征融合的未知辐射源个体识别算法^[12]。该算法涉及变换域特征提取、数据降维与融合和分类识别3个环节,如图1所示。

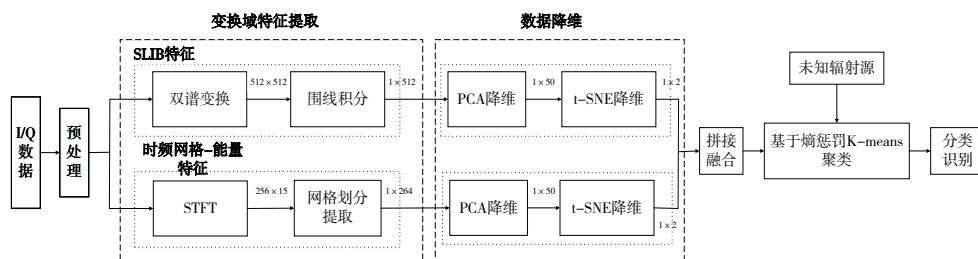


Fig. 1 Individual emitter identification flow based on SLIB and time-frequency grid-energy feature fusion

图1 基于SLIB与时频网格—能量特征融合辐射源个体识别流程

在预处理阶段,裁剪掉信号中的无信号部分后进行幅值归一化与分段处理。在变换域特征提取阶段,经过二阶差分增强特征可分性后,利用双谱变换获取信号维度为 512×512 的双谱矩阵,沿积分路径对双谱矩阵进行围线积分,以获得 1×512 维的SLIB特征。

在时频网格—能量特征提取阶段,根据短时傅里叶变换计算信号维度为 256×15 的时频矩阵,先在频域段将时频矩阵划分为8段,计算其整体频段能量特征,再将时间和频率轴均划分为8段,每个网格区域提取4个网格特征,最终提取到 1×264 维的时频网格—能量特征。

在数据降维阶段,首先,利用PCA将获取到的两个特征提取主成分降维到50;其次,通过t-SNE非线性降维算法将其降到2维,在包含全局信息的同时,又不丢失局部特征;再次,将融合的 1×4 维的特征向量输入基于熵惩罚

的K-means聚类算法进行迭代,以输出分类结果;最后,将未知设备的数据集进行特征提取、数据降维和融合,与已训练的数据集组合起来进行聚类识别。

2 变换域特征提取

2.1 积分双谱

双谱是一种高阶统计量,既可保留信号的幅度和相位信息,又具备抑制高斯噪声的能力,对辐射源设备的分类识别十分有益^[13,14]。然而,由于双谱体现的是三维空间中的谱信息,计算结果为二维矩阵形式的特征向量,若直接将其用于分类识别,存在维数较高、数据量较大的问题,会导致算法复杂度较高。为了充分发挥双谱分析的优势,需将二维双谱数据降维转换为一维表示,以提升后续处理

效率。

积分双谱是一种常用的降维方法,其原理是沿双谱平面的特定路径对双谱值进行积分,实现二维特征向一维特征的映射,旨在降低数据维度的同时,完整保留双谱特征的信息,使其能更适用于辐射源目标的识别和分类任务。目前,基于不同的积分路径可将积分双谱分类为:径向积分双谱、轴向积分双谱、圆周积分双谱及围线积分双谱^[15-18]。

SLIB 按照特定路径提取双谱值,确保了每个双谱数据既不遗漏也不重复,保证了双谱特征的完整性。因此,SLIB 不仅能准确提取信号中所有有效信息,保留信号的尺度特征及部分相位细节,还具备平移不变性。图 2 为积分路径。

$$\text{SLIB}(\omega) = \oint_{S_l} B(\omega_1, \omega_2) d\omega_1 d\omega_2 \quad (1)$$

式中: S_l 为围线积分双谱的积分路径; $B(\omega_1, \omega_2)$ 为信号的双谱估计值。

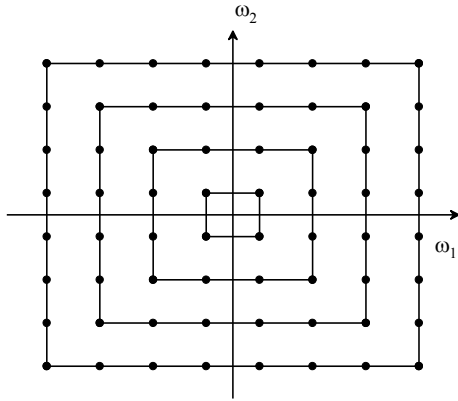


Fig. 2 Integral path of bispectrum of confinement integral

图2 围线积分双谱的积分路径

2.2 时频网格—能量特征

在信号分类与识别领域,时频分析方法能揭示信号在时间和频率域的多维信息,直接分类可能导致计算复杂度过高。为了有效地利用时频图中的信息,避免维度灾难的发生,需对时频图进行降维处理,以提取更紧凑且能充分反映信号特性的特征。

时频网格—能量特征提取方法通过网格划分时频图,在每个网格内提取多维统计特征,包括频段能量特征、网格统计特征、网格能量特征等,将高维时频信息转化为低维特征向量,在实现了降维的同时保留了信号的时频特性信息。同时,提取每个网格区域的统计特征(能量、均值、标准差、熵等),既能显著降低数据维度、计算复杂度,还能提升信号处理和分类效率。此外,该网格划分方式灵活,能根据不同信号的特性进行调整,在不同类型的信号处理中都具有良好的适应性。

本文采用短时傅里叶变换对信号进行时频分析,得到信号的时频矩阵 $S(f, t)$ 。

$$S(f, t) = \sum_n x(n) \omega(n - t) e^{-j2\pi f n} \quad (2)$$

式中: f 为频率; t 为时间; $x(n)$ 为原始信号; $\omega(n)$ 为窗函数; $S(f, t)$ 为信号在特定时间和频率点的幅度。

为了将整个时频图的高维信息转化为较低维度的特征,并保留信号的时频信息。首先,沿频率轴将频率范围等分为 N 个频带,沿时间轴将时频图等分为 M 个等间隔的时间窗口,每个网格包含不同时间和频率范围内的时频信息;其次,为了全面捕捉信号在时频网格内的多样性,从多个维度描述信号的时频特性,提取每个网格内的能量、均值、标准差、熵等统计特征;最后,对所有网格提取的特征进行组合,形成一个一维特征向量,每个网格包含多个特征,最终的特征向量维度为网格数与每个网格特征数的乘积,例如每个网格提取 k 个特征,则最终的时频网格—能量特征维度为 $M \times N \times k$ 。

2.3 数据降维与融合

在信号处理和特征提取过程中,通过 SLIB 与时频网格—能量特征提取可获取丰富的信息,但这些特征仍为高维向量,存在冗余特征和维度灾难的问题。具体地,高维特征包含大量信息,存在计算复杂度过高、过拟合风险大和难以可视化的问题。为了提升模型的分类与识别性能,提出 PCA+t-SNE 降维方法来解决这些问题。其中,PCA 主要用于处理线性特征,对于数据中潜在的非线性结构,无法完整地保留数据的局部细节和结构信息^[19];t-SNE 作为一种非线性降维方法,虽能有效保留高维数据的局部结构,但提取后的特征向量维度较高,数据特征值过多,计算数据点之间成对距离的时间复杂度很高,随着维度增加,计算量呈指数级增长,算法效率低下^[20]。此外,t-SNE 对噪声点很敏感,降维过程中会计算高维数据点之间的相似度,噪声点会导致其在低维空间中被错误地分配,从而影响分类效果。

由于双谱的对称性,积分双谱值存在冗余特征,在时频图进行网格划分时,存在信道噪声等冗余信息。为了解决上述问题,采用 PCA 对数据进行线性降维,去除冗余特征和噪声点,保留数据的全局结构,以减少计算量、提升数据质量;应用 t-SNE 进行非线性降维,充分保留局部特征,使降维后的数据既包含全局信息,又不丢失局部特征。综上,双重降维后的特征在低维空间中有效保留了数据的聚类结构,使相似的数据点尽可能接近,不相似的数据点则会分开。

3 基于熵惩罚 K-means 的聚类算法

针对传统的 K-means 聚类算法在处理降维后的数据时,存在的初始化敏感性、对噪声的高敏感度及对不规则簇形的适应性差等问题,构建了一种基于熵惩罚的 K-means 聚类算法。具体地,引入动态簇调整机制,采用全局熵与样本局部熵加权策略,结合 Gamma 与 Beta 动态调整

因子,使算法能根据数据的分布特征自适应确定最优簇数,优化样本隶属关系。这一机制有效减少了噪声点对聚类过程的干扰,使算法能适应数据的复杂性和不规则性,在无标签数据的辐射源识别任务中展现出更强的适应性和准确性。

3.1 理论分析

本文基于EM算法的思想,对K-means聚类算法的目标函数进行改进,引入熵惩罚项,使用簇的比例熵来引导算法自动找到最佳的簇数量^[21]。目标函数为:

$$J_1(z, A, \alpha) = \sum_{i=1}^n \sum_{k=1}^c z_{ik} KL(\tilde{x}_i \| \tilde{a}_{ik}) - \beta n \sum_{k=1}^c \alpha_k \ln \alpha_k - \gamma \sum_{i=1}^n \sum_{k=1}^c z_{ik} \ln \alpha_k, \beta \geq 0 \quad (3)$$

式中: $\sum_{k=1}^c \alpha_k \ln \alpha_k$ 为第 k 个簇权重(出现概率) α_k 的全局熵; β 为平衡全局熵的权重,控制各簇的竞争力度; $\sum_{i=1}^n \sum_{k=1}^c z_{ik} \ln \alpha_k$ 为基于样本分配的熵项,将簇比例 α_k 与样本隶属 z_{ik} 耦合; γ 为平衡样本分配熵的权重,从局部角度衡量 α_k 与真实数据分布的一致性。

同时,利用基于概率分布的KL散度量度样本的概率隶属关系^[22]。

$$KL(\tilde{x}_i \| \tilde{a}_{ik}) = \sum_{j=1}^n \tilde{x}_{ij} \ln \frac{\tilde{x}_{ij}}{\tilde{a}_{ik}} \quad (4)$$

式中: $\tilde{x}_i = x_i / \sum_{j=1}^n x_j$, $\tilde{a}_{ik} = a_{ik} / \sum_{j=1}^n a_{ij}$ 为样本和簇中心的概率分布。

为了方便求解,以 λ 为拉格朗日乘子,约束条件为 $\sum_{k=1}^c \alpha_k = 1$,则式(3)的拉格朗日函数为:

$$\tilde{J}_1(z, A, \alpha, \lambda) = \sum_{i=1}^n \sum_{k=1}^c z_{ik} KL(\tilde{x}_i \| \tilde{a}_{ik}) - \beta n \sum_{k=1}^c \alpha_k \ln \alpha_k - \gamma \sum_{i=1}^n \sum_{k=1}^c z_{ik} \ln \alpha_k - \lambda (\sum_{k=1}^c \alpha_k - 1) \quad (5)$$

(1) 样本隶属。对 $\tilde{J}_1$ 关于 z_{ik} 求偏导数,并令其为0,则样本隶属 z_{ik} 的更新方程为:

$$z_{ik} = \begin{cases} 1, & KL(\tilde{x}_i \| \tilde{a}_{ik}) - \gamma \ln \alpha_k = \min_{1 \leq k \leq c} KL(\tilde{x}_i \| \tilde{a}_{ik}) - \gamma \ln \alpha_k \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

由式(6)可知,采用KL散度量度样本概率分布之间的信息差异或信息损失程度,能灵敏地捕捉分布形状、峰值和概率质量分配的差异。因此,数据点 x_i 隶属 k 类,除了要考虑KL散度外,还要考虑 $\ln \alpha_k$; 若 α_k 很小则该簇样本量很少, $-\gamma \ln \alpha_k$ 会很大,该簇被排斥的概率会增大。

(2) 簇中心 a_k 与簇权重 α_k 。对于簇中心 a_k ,将其更新为该簇内的样本均值位置。

$$a_k = \sum_{i=1}^n z_{ik} x_{ij} / \sum_{i=1}^n z_{ik} \quad (7)$$

对 $\tilde{J}_1$ 关于 α_k 求偏导数, $\frac{\partial \tilde{J}_1}{\partial \alpha_k} = -\beta n (\ln \alpha_k + 1) -$

$\gamma \sum_{i=1}^n \frac{z_{ik}}{\alpha_k} - \lambda = 0$, 将等式两边乘以 α_k , 并对每一个簇累计求和。

$$-\sum_{k=1}^c \eta \beta \alpha_k \ln \alpha_k - \sum_{k=1}^c \eta \beta \alpha_k - \gamma \sum_{k=1}^c \sum_{i=1}^n z_{ik} - \sum_{k=1}^c \lambda \alpha_k = 0 \quad (8)$$

由于其约束条件为 $\sum_{k=1}^c \alpha_k = 1$, 则:

$$\lambda = -\eta \beta \sum_{k=1}^c \alpha_k \ln \alpha_k - \eta \beta - n \gamma \quad (9)$$

因此, α_k 的更新公式为:

$$\alpha_k^{(t+1)} = \sum_{i=1}^n z_{ik} / n + (\beta / \gamma) \alpha_k^{(t)} \left(\ln \alpha_k^{(t)} - \sum_{s=1}^c \alpha_s^{(t)} \ln \alpha_s^{(t)} \right) \quad (10)$$

式中: t 为迭代次数; $\sum_{s=1}^c \alpha_s \ln \alpha_s$ 为加权的 $\ln \alpha_k$ 的均值,权重为 $\alpha_1, \alpha_2, \dots, \alpha_c$ 。

如果第 k 个权重 $\alpha_k^{(t)}$ 小于加权均值,则新的权重 $\alpha_k^{(t+1)}$ 将减小,即较小的比例减小,较大的比例在下次迭代中增大,进而出现竞争。此时,如果 $\alpha_k < 1/n$, $1 \leq k \leq c^{(t)}$, 则 α_k 为不合法的比列,需要丢弃并更新簇的数量 c , 剩余权重 α_k' 及相应的 z_{ik}' 也需更新并进行归一化处理。

$$\alpha_k' = \alpha_k' / \sum_{s=1}^{c^{(t+1)}} \alpha_s' \quad (11)$$

$$z_{ik}' = z_{ik}' / \sum_{s=1}^{c^{(t+1)}} z_{is}' \quad (12)$$

(3) γ, β 的更新策略。学习率 γ 影响着局部熵在目标隶属的相对权重,随着迭代次数增加,为了逐步减弱熵惩罚的影响并避免过早收敛, γ 不应下降太慢或太快。本文算法中 $\gamma = e^{-d^{(t)}/250}$ 。在竞争学习模式下,算法自适应地减少聚类数量,为下一轮迭代提供参数的估计值。此外,参数 β 影响着全局熵,控制竞争过程,本文算法利用规则设置 $-e^{-1} \leq \alpha_k \ln \alpha_k < 0$, 当 $0 < \alpha_k \leq 1$ 且 $E = \sum_{s=1}^c \alpha_s \ln \alpha_s < 0$ 时,有 $\alpha_k E = \alpha_k \sum_{s=1}^c \alpha_s \ln \alpha_s < 0$, 则存在:

$$-e^{-1} \beta < \beta \alpha_k \left(\ln \alpha_k - \sum_{s=1}^c \alpha_s \ln \alpha_s \right) < \beta (-\alpha_k E) \quad (13)$$

在考虑约束条件 $\sum_{k=1}^c \alpha_k = 1$ 且当 $\alpha_k < 1/2$ 时,存在 $\ln \alpha_k - \sum_{s=1}^c \alpha_s \ln \alpha_s < 0$, 避免所有 $\alpha_k \leq 0$, 即上述不等式左侧必须大于所有 α_k 中小于 $1/2$ 的最大值。

$$-e^{-1} \beta > -\max \{ \alpha_k | \alpha_k < 1/2, k = 1, 2, \dots, c \} \quad (14)$$

由式(14)可得 $\beta < \max \{ \alpha_k | \alpha_k < 1/2, k = 1, 2, \dots, c \} < e/2$, 为了避免 β 过大,令 $\beta \in [0, 1]$ 。此外,若 $\alpha_k^{(t+1)}$ 和 $\alpha_k^{(t)}$ 之间差异较小,则增大 β 以增强其竞争力;若差异较大,则减小 β 以保持稳定性。因此, β 的更新公式为:

$$\beta = \sum_{k=1}^c \exp \left\{ -\eta n \left| \alpha_k^{(t+1)} - \alpha_k^{(t)} \right| \right\} / c \quad (15)$$

式中: $\eta = \min \{ 1, 1/l^{(d/2-1)} \}$; $\lfloor \cdot \rfloor$ 表示不大于 $\cdot$ 的最大整数。

根据 α_k 的更新公式得到的不等式为:

$$\begin{aligned} \max_{1 \leq k \leq c} \alpha_k^{(t+1)} &\leq \max_{1 \leq k \leq c} \left(\frac{1}{n} \sum_{i=1}^n z_{ik} \right) + \frac{\beta}{\gamma} \max_{1 \leq k \leq c} \alpha_k^{(t)} \left(\ln \max_{1 \leq k \leq c} \alpha_k^{(t)} - \sum_{s=1}^c \alpha_s^{(t)} \ln \alpha_s^{(t)} \right) \\ &< \max_{1 \leq k \leq c} \left(\frac{1}{n} \sum_{i=1}^n z_{ik} \right) + \beta \left(- \max_{1 \leq k \leq c} \alpha_k^{(t)} \sum_{s=1}^c \alpha_s^{(t)} \ln \alpha_s^{(t)} \right) \end{aligned} \quad (16)$$

当 $\max_{1 \leq k < c} \left(\frac{1}{n} \sum_{i=1}^n z_{ik} \right) - \beta \max_{1 \leq k < c} \alpha_k^{(t)} \sum_{s=1}^c \alpha_s^{(t)} \ln \alpha_s^{(t)} \leq 1$

时,关于 $\alpha_k^{(t)}$ 的约束条件 $\max_{1 \leq k \leq c} \alpha_k^{(t+1)} \leq 1$ 成立,则存在:

$$\beta \leq \left(1 - \max_{1 \leq k \leq c} \left(\frac{1}{n} \sum_{i=1}^n z_{ik} \right) \right) / \left(- \max_{1 \leq k \leq c} \alpha_k^{(t)} \sum_{s=1}^c \alpha_s^{(t)} \ln \alpha_s^{(t)} \right) \quad (17)$$

根据式(15)、式(17)可得算法参数 β 的迭代过程为:

$$\begin{aligned} \beta^{(t+1)} &= \\ \min &\left[\frac{\sum_{k=1}^c \exp(-\eta n |\alpha_k^{(t+1)} - \alpha_k^{(t)}|)}{c}, \frac{1 - \max_{1 \leq k \leq c} \left(\frac{1}{n} \sum_{i=1}^n z_{ik} \right)}{\left(- \max_{1 \leq k \leq c} \alpha_k^{(t)} \sum_{k'=1}^c \ln \alpha_{k'}^{(t)} \right)} \right] \end{aligned} \quad (18)$$

式中: β 可在任何时刻跳跃,当聚类数 c 稳定时 $\beta = 0$,即聚类数不再减少,达到最佳聚类数目。本文将所有数据点作为初始均值, $a_k = x_k$ 且 $c_{\text{initial}} = K$,并将 $\alpha_k = 1/c_{\text{initial}}$ 作为初始权重参数; K 为根据样本总量预设的一个粗略聚类数目。

3.2 实现流程

为了在数据稀缺、缺乏带标签数据的情况下,仍能对未知类别的样本进行有效地聚类 and 分类识别,根据样本内在的类别分布自适应地确定最佳聚类数目。本文提出基于熵惩罚的 K-means 聚类算法,通过动态更新和竞争机制自动识别样本类别数。

步骤 1: 令 $\varepsilon > 0$, 给定初始值 $c^{(0)} = K, \alpha_k^{(0)} = 1/n, a_k^{(0)} = x_i$, 初始化学学习率 $\gamma^{(0)} = \beta^{(0)} = 1, t = 0$ 。

步骤 2: 根据 $a_k^{(t)}, \alpha_k^{(t)}, c^{(t)}, \gamma^{(t)}, \beta^{(t)}$ 和式(16), 计算数据点在第 t 次迭代中被分配到第 k 类的隶属度 $z_{ik}^{(t)}$ 。

步骤 3: 根据 $\gamma^{(t+1)}$ 的计算公式得到新一轮迭代的学学习率。

步骤 4: 根据步骤 2 中获得的第 $t+1$ 轮各数据点属于每类的隶属度 $z_{ik}^{(t+1)}$ 、第 t 轮迭代中的 $\alpha_k^{(t)}$, 并根据式(10)更新第 $t+1$ 轮的每个类的混合比例参数 $\alpha_k^{(t+1)}$ 。

步骤 5: 根据更新后的 $\alpha_k^{(t+1)}, \alpha_k^{(t)}$ 计算新一轮的竞争参数 $\beta^{(t+1)}$ 。

步骤 6: 根据 $\alpha_k^{(t+1)} \leq 1/n$ 丢弃不合适的聚类簇, 并调整 $\alpha_k^{(t+1)}, z_{ik}^{(t+1)}$ 、聚类数目 $c^{(t+1)}$ 。如果 $t \geq 60$ 且 $c^{(t-60)} - c^{(t)} = 0$ 时, 令 $\beta^{(t+1)} = 0$ 。

步骤 7: 根据更新后的聚类数目 $c^{(t+1)}$ 、隶属度 $z_{ik}^{(t+1)}$ 更新聚类中心 $a_k^{(t+1)}$ 。

步骤 8: 比较 $a_k^{(t+1)}, \alpha_k^{(t)}$, 如果满足收敛条件 $\max_{1 \leq k < c^{(0)}} \|a_k^{(t+1)} - a_k^{(t)}\| < \varepsilon$, 则停止迭代过程; 否则, 令 $t =$

$t+1$, 且返回步骤 2 进行新一轮迭代。

4 仿真实验与分析

本文在不同信噪比下, 将所提方法与基于积分双谱的未知辐射源非监督算法、基于双谱直方图的通信辐射源个体识别算法进行比较, 以分析所提方法的识别性能^[23, 24]。

4.1 实验环境与数据集

本文实验的服务器为 NVIDIA GEFORCE RTX 3090 GPU, 在 Matlab 2019b 中使用改进的熵惩罚 K-means 聚类、PCA 与 t-SNE 降维算法进行特征降维与聚类。

在真实环境下, 7 台同型号超短波发射机发射同种通信信号样本, 并由同一台超短波接收机采集以文本形式存储 IQ 数据。通信信号的调制方式为 BPSK, 信号采样率为 96 kHz, 载频为 24 kHz, 每台设备提取 600 个样本, 每个信噪比采集 2 400 个样本, 按照 8:2 的比例划分训练集、测试集。

为了比较算法在不同信噪比下的识别性能, 在仿真软件 Matlab 2019a 中添加高斯白噪声以模拟不同信噪比下的噪声环境, 建立了信噪比范围为 $\{-10 \text{ dB}, -7 \text{ dB}, -5 \text{ dB}, -2 \text{ dB}, 0 \text{ dB}, 5 \text{ dB}, 8 \text{ dB}, 10 \text{ dB}\}$ 的数据集。同时, 为了进一步验证所提算法的泛化能力, 还引入公开的手机蓝牙 (Bluetooth, BT) 信号开源数据集 (<https://zenodo.org/records/3876140>), 在 5 GHz 采样率下的 BT 信号进行测试验证。此外, 对 5 台不同型号的数据样本进行测试, 采用跨域数据验证以确保所提算法的鲁棒性和适应性。

4.2 实验参数

在围线积分双谱特征的提取过程中, 首先在预处理阶段引入二阶差分算子, 通过相邻信号点间的相位差分运算增强特征可分性; 其次, 利用 bispecd() 双谱函数计算双谱矩阵, 设置滞后数为 49, 采用分段长度为 512 点的数据切片, 片段重叠率为 50%, 每段数据进行 512 点的 FFT 运算并禁用偏移量补偿; 最后, 对双谱矩阵进行围线积分双谱, 以获取 1×256 维的特征向量。

在提取时频网格—能量特征时, 配置汉明窗口函数, 分别设置 512 点、256 点的窗口长度与重叠量, 配合 512 点的 FFT 运算生成时频矩阵。同时, 通过 8×8 网格分区提取局部统计量及全局 8 频段能量分布。

在聚类模型训练时, 目标聚类簇数依据样本总量与单个设备样本量预设大于实际簇数的值, 本文算法 $n = 8$, 最大迭代次数设置为 20, 训练轮次设置为 50, Beta、Gamma 初始值为 1, 每个簇的初始概率均等为 $1/n$ 。

4.3 评价指标

为了衡量算法聚类结果的优劣, 使用调整兰德指数 (Adjusted Rand Index, ARI) 作为聚类性能评估指数。

$$ARI = \frac{RI - E[RI]}{\max(RI) - E[RI]} \quad (19)$$

式中: RI 为兰德因子, 且 $RI = (a + b) / \binom{n}{2}$; n 为样本的总数; a 指同属真实类且同属聚类的对数; b 指分属不同类且不同聚类的对数; $\binom{n}{2}$ 为所有可能的元素对的数量; ARI 的值在 $[-1, 1]$, 值越大表示真实类别划分和聚类结果之间的重叠程度越高。

4.4 实验结果

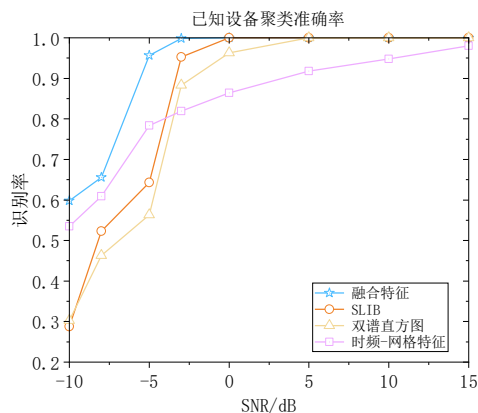
4.4.1 不同算法分类识别性能

将单独使用 SLIB、时频网格—能量特征、双谱直方图特征的方法与所提方法进行比较, 每个特征采用的聚类算法均为熵惩罚 K-means 聚类模型算法, 数据集进行了相同处理, 以测试所提模型的分类识别性能。图 3(a) 为超短波接收机信号在不同信噪比下 4 个设备的聚类性能。图 3(b) 为将另外 3 台同型号接收机作为未知辐射源后, 在不同信噪比下与原训练结果的比较结果。

由图 3 可见, 所提算法的识别率高于单独使用 SLIB、时频网格—能量特征、双谱直方图特征的方法, 原因为融合后的特征结合了信号的非线性高阶统计特征和时频域的细粒度能量变化信息, 充分表征了不同设备间的细微差异, 而单一特征无法充分体现信号在不同域间的复杂变化情况, 会遗漏一些关键细节。实验表明, 通过多特征融合能更准确地捕捉设备之间细微的差异, 可提升模型对未知辐射源设备个体的分类能力, 证明了算法的有效性。

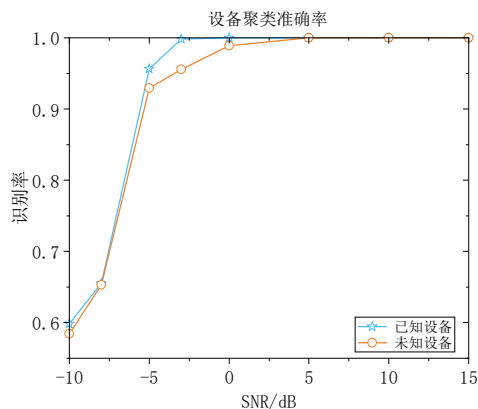
图 4 为 4 台超短波发射机在 -5 dB 下的聚类可视化结果和混淆矩阵。由此可见, 经过 t-SNE 降维后, 不同类别的样本分布在不同区域, 表明数据中的类别是可分的。图 4(b) 为聚类结果与真实标签之间的混淆矩阵。由此可见, 在第一个类别 1 的真实标签与预测标签之间, 93% 的样本被正确地分类为类别 1, 剩下的 7% 被错误分类为类别 2, 其他类别的预测正确率较高, 表明聚类算法的效果非常好, 与真实标签之间的匹配度很高。图 5 为在 -5 dB 下, 加入 3 台超短波发射机后的聚类可视化结果和混淆矩阵。由此可见, 添加新数据集后, 不同设备间的聚类可分性仍然较好。

为了进一步验证该算法泛化能力, 选择 5 台在 5 GHz 采样率下采集的 BT 信号, 采用相同的信号特征处理方式, 比较两种信号特征在单独使用 SLIB、时频网格—能量特征与使用本文方法的性能, 每个特征采用的聚类算法均为基于熵惩罚 K-means 聚类算法, 实验结果如图 6 所示。由此可见, 所提方法在手机 BT 数据集上的聚类准确率相较于单独使用 SLIB、时频网格—能量特征的性能更优, 能充分利用信号的非线性高阶统计特征和时频域的细粒度能量



(a) Clustering performance of 4 known devices

(a) 已知 4 台设备的聚类性能



(b) Classification performance after adding unknown devices

(b) 加入未知设备的分类性能

Fig. 3 Clustering results of ultrashort wave equipment at different SNR

图 3 超短波设备在不同信噪比下聚类性能

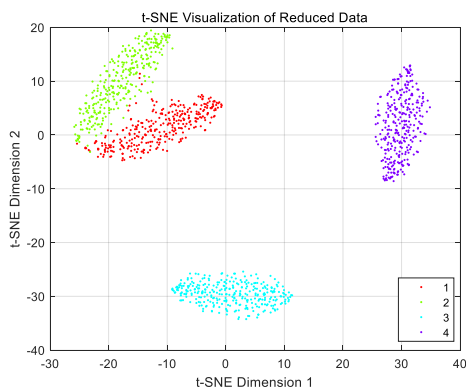
变化信息来分类手机 BT 设备。

由图 3、图 6 可见, 手机设备在不同信噪比下的识别性能明显低于超短波发射机设备, 可能与手机 BT 信号所采用的频率跳变扩频技术和复杂的数字调制方式等相关。尽管, 这些技术提升了设备的抗干扰能力和通信安全性, 但会使信号结构更复杂。具体为, 复杂的多信道环境、跳频机制和协议层叠加往往会掩盖信号中的非线性特征, 使特征提取困难, 识别性能变差。在信噪比低于 0 dB 时, 手机设备单独使用时频—网格特征受噪声影响严重, 识别准确率下降; 在信噪比高于 -8 dB 时, SLIB 特征能保持相对稳定的变化, 原因为预处理阶段引入了二阶差分因子, 增加了信号中与非线性特征相关的特性, 在低信噪比下具有一定的抗噪能力。

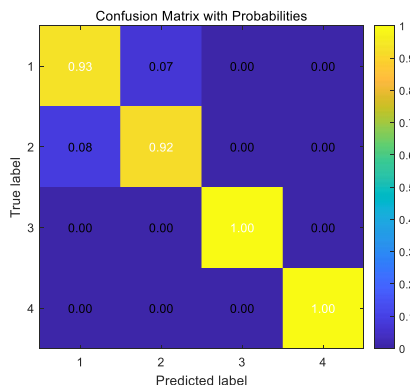
综上, 所提算法在手机设备中的识别准确率相较于超短波发射机设备较低, 但相较于与单一特征方法而言仍具有显著优势, 具有一定的实际应用价值。

4.4.2 不同聚类模型分类性

为了验证改进的熵惩罚 K-means 聚类算法的性能, 在多域特征降维融合后的数据集中, 将其与基于粒子群优化



(a) Visualization results
(a) 可视化结果



(b) Confusion matrix
(b) 混淆矩阵

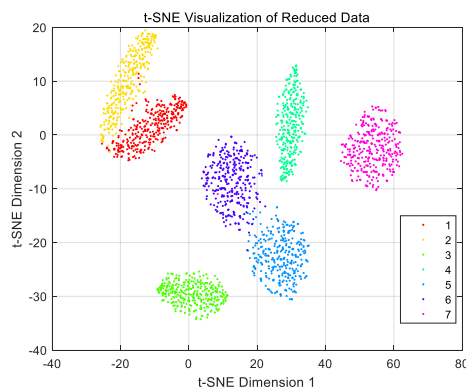
Fig. 4 Clustering visualization results and confusion matrix of four ultrashort wave transmitters

图4 4台超短波发射机聚类可视化结果和混淆矩阵

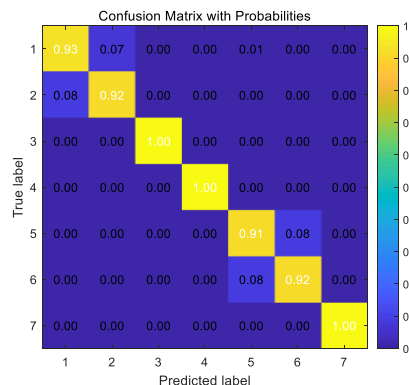
的自适应K-means聚类算法、高斯混合模型GMM聚类方法进行比较,实验结果如图7所示^[25]。由此可见,改进后的熵惩罚K-means聚类算法相较于基于粒子群优化的自适应K-means聚类方法和高斯混合模型GMM方法,在低信噪比下的识别准确率更高。在信噪比为-8 dB时,改进后的熵惩罚K-means聚类算法相较于其他两个模型,分类准确率分别提升7.79%、11.1%,证明了其在低信噪比的环境下能有效提升信号分类准确性,在复杂噪声环境中的性能更优、鲁棒性更强。

5 结语

本文针对射频指纹识别中未知辐射源识别的自适应聚类需求、单一特征在信号表征中的局限性,提出一种基于电磁信号SLIB与时频网格—能量特征融合和熵惩罚K-means聚类的未知辐射源分类识别算法。首先,将时频谱网格划分的时频网格—能量特征与围线积分双谱特征,进行PCA+t-SNE双重降维并融合,形成既包含信号的非线性高阶统计特征,又保留时频域的局部细粒度能量变化信息



(a) Visualization results
(a) 可视化结果



(b) Confusion matrix
(b) 混淆矩阵

Fig. 5 Clustering visualization results and confusion matrix after adding 3 ultrashort wave transmitters

图5 加入3台超短波发射机后聚类可视化结果和混淆矩阵

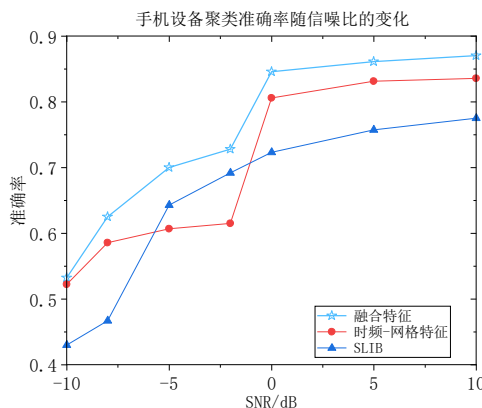


Fig. 6 Recognition results of mobile phone devices at different SNR

图6 手机设备在不同信噪比下的识别性能

的综合特征向量,以更全面表征信号的本质特征;其次,构建一种基于熵惩罚K-means聚类方法,采用结合了动态簇中心调整与信息熵权重更新机制的无监督自适应策略调整聚类簇数;最后,引入Gamma动态调整因子和Beta权重机制,优化样本归属计算,提升聚类稳定性和鲁棒性。

实验表明,相较于单独使用SLIB、时频网格—能量特征方法,所提方法的分类识别准确率更高;基于熵惩罚K-

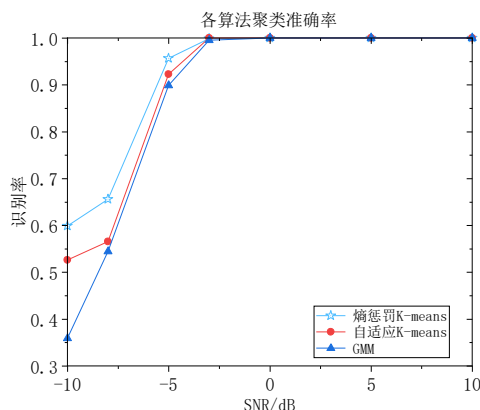


Fig. 7 Recognition results of different models at different SNR

图7 不同模型在不同信噪比下的识别性能

means的聚类方法相较于其他聚类模型的分类效果和鲁棒性更强。

参考文献:

- [1] YANG H F, ZHANG H, WANG H J, et al. A novel approach for unlabeled samples in radiation source identification[J]. Journal of Systems Engineering and Electronics, 2022, 33(2): 354–359.
- [2] ZHANG B H, LIU S T, et al. Research progress of unknown emitter identification technology[J]. Journal of China Academy of Electronics and Information Technology, 2023, 18(9): 864–872.
张本辉, 刘松涛. 未知辐射源识别技术研究进展[J]. 中国电子科学研究院学报, 2023, 18(9): 864–872.
- [3] TANG P, XU Y T, LI J H, et al. Development of non-cooperative radiation source identification technology[J]. Journal of Data Acquisition & Processing, 2020, 35(6): 1033–1048.
汤鹏, 徐以涛, 李京华, 等. 非合作辐射源识别技术发展动态分析[J]. 数据采集与处理, 2020, 35(6): 1033–1048.
- [4] CIOS K J, SWINIARSKI R W, PEDRYCZ W, et al. Unsupervised learning: clustering[J]. Data Mining: A Knowledge Discovery Approach, 2007, 12(3): 257–288.
- [5] COHN R, HOLM E. Unsupervised machine learning via transfer learning and K-means clustering to classify materials image data[J]. Integrating Materials and Manufacturing Innovation, 2021, 10(2): 231–244.
- [6] LI X, LEI Y K. Radiation source individual identification using machine learning method[C]// IEEE 8th Joint International Information Technology and Artificial Intelligence Conference, 2019: 1001–1005.
- [7] LU Y, LU S Y, FOTOUHI F, et al. FGKA: a fast genetic k-means clustering algorithm[C]// Proceedings of the 2004 ACM Symposium on Applied Computing, 2004: 622–623.
- [8] YANG Z T, WU Z L, YIN Z D, et al. Hybrid radar emitter recognition based on rough K-means classifier and relevance vector machine[J]. Sensors, 2013, 13(1): 848–864.
- [9] LI X, LEI Y K. Individual identification of communication station based on density peak algorithm[J]. Signal Processing, 2019, 35(7): 1242–1249.
李昕, 雷迎科. 基于密度峰值算法的通信电台个体识别[J]. 信号处理, 2019, 35(7): 1242–1249.
- [10] PENG J X, SANG M M, GENG D, et al. Research on identification of specific emitter based on multi-feature parameter clustering[J]. Metrology & Measurement Technology, 2023, 43(6): 45–53.
彭进先, 桑苗苗, 耿丹, 等. 多特征参数聚类分析的特定辐射源识别方法研究[J]. 计测技术, 2023, 43(6): 45–53.
- [11] ZHU L. Research on unsupervised identification algorithm of communication emitter individual[D]. Shanghai: University of Electronic Science and Technology of China, 2024.
朱玲. 通信辐射源个体非监督识别算法研究[D]. 上海: 电子科技大学, 2024.
- [12] LI Q X. Research on individual identification technology of communication emitter[D]. Chengdu: University of Electronic Science and Technology of China, 2021.
李秋雪. 通信辐射源个体识别技术研究[D]. 成都: 电子科技大学, 2021.
- [13] ZHANG X D. Modern signal processing[M]. Beijing: Tsinghua University Press, 2002.
张贤达. 现代信号处理[M]. 北京: 清华大学出版社, 2002.
- [14] DING L D, WANG S L, WANG F G, et al. Specific emitter identification via convolutional neural networks[J]. IEEE Communications Letters, 2018, 22(12): 2591–2594.
- [15] CHANDRAN V, ELGAR S L. Pattern recognition using invariants defined from higher order spectra—one-dimensional inputs[J]. IEEE Transactions on Signal Processing, 1993, 41(1): 205–212.
- [16] TUGNAT J K. Detection of non-Gaussian signals using integrated polyspectrum[J]. IEEE Transactions on Signal Processing, 1994, 42(11): 3137–3149.
- [17] LIAO X, BAO Z. Circularly integrated bispectra: novel shift invariant features for high-resolution radar target recognition[J]. Electronics letters, 1998, 34(19): 1879–1880.
- [18] YU F N. Application of bispectrum in underwater target feature extraction[D]. Harbin: Harbin Engineering University, 2010.
于凤楠. 双谱分析方法在水下目标特征提取中的应用[D]. 哈尔滨: 哈尔滨工程大学, 2010.
- [19] ZHANG Z, HU A Q, XU W, et al. An artificial radio frequency fingerprint embedding scheme for device identification[J]. IEEE Communications Letters, 2022, 26(5): 974–978.
- [20] XU Z Y. Application of t-SNE algorithm in dimensionality reduction of high dimensional data[D]. Xiamen: Xiamen University, 2021.
徐振宇. t-SNE算法在高维数据降维中的应用研究[D]. 厦门: 厦门大学, 2021.
- [21] DEMPSTER A P. Maximum likelihood from incomplete data via the EM algorithm[J]. Journal of Royal Statistical Society, 1977, 39(1): 1–22.
- [22] JI S, ZHANG Z Z, YING S H, et al. Kullback-leibler divergence metric learning[J]. IEEE Transactions on Cybernetics, 2020, 52(4): 2047–2058.
- [23] ZHANG K X. Research on individual identification of radiant source based on RF fingerprint[D]. Xi'an: Xi'an University of Technology, 2023.
张可欣. 基于射频指纹的辐射源个体识别方法研究[D]. 西安: 西安理工大学, 2023.
- [24] LI X. Research on individual identification of communication emitter based on unsupervised learning[D]. Changsha: National University of Defense Technology, 2019.
李昕. 基于无监督学习的通信辐射源个体识别技术研究[D]. 长沙: 国防科技大学, 2019.
- [25] WANG Y C, HE Z W, MA M J, et al. Unsupervised specific emitter identification based on feature parameter fusion and adaptive clustering[C]// IEEE/CIC International Conference on Communications in China, 2022: 269–274.