

偏联系数聚类和随机森林算法在雷达信号分选中的应用

张萌萌^{**}, 刘以安^{*}, 宋萍

江南大学物联网工程学院, 江苏 无锡 214122

摘要 为了提高雷达调制信号在电子对抗环境中的分选准确度, 建立了基于偏联系数模糊聚类(PCFCM)算法和教与学随机森林(TLRF)算法的雷达调制信号分选(PCFCM-TLRF)模型。该模型引入偏联系数(PCN)改进 K 均值聚类(K-means)算法, 优化模糊 C 均值聚类(FCM)算法, 用优化后的 FCM 算法对信号样本集进行预处理; 使用“教与学”优化(TLBO)算法优化随机森林(RF)算法, 使优化后的 RF 算法能够以更低的复杂度构成更优的分类器; 将预处理后的样本作为 TLRF 中的训练样本实现信号分选。研究表明, 与其他分选模型相比, PCFCM-TLRF 模型具有更高的分选准确度, 能够有效地实现雷达调制信号的分选。

关键词 遥感; 信号分选; 偏联系聚类算法; 教与学随机森林算法; 集对分析

中图分类号 TN958

文献标识码 A

doi: 10.3788/LOP56.062804

Applications of Partial Connection Clustering Algorithm and Random Forest Algorithm in Radar Signal Sorting

Zhang Mengmeng^{**}, Liu Yi'an^{*}, Song Ping

School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China

Abstract In order to improve the sorting accuracy of radar modulated signals in the electronic countermeasure environment, based on the partial connection fuzzy C-means (PCFCM) algorithm and the teaching-learning random forest (TLRF) algorithm, a radar modulated signal sorting model PCFCM-TLRF is proposed. In this model, we introduce the partial connection number (PCN) to improve the K-means clustering algorithm and optimize the fuzzy C-means (FCM) algorithm. Then the signal sample is pre-processed with the improved FCM algorithm. The teaching-learning-based optimization (TLBO) algorithm is used to optimize the random forest (RF) algorithm, so that the optimized RF algorithm can form a better classifier with much lower complexity. The pre-processed sample is used as the training sample in the TLRF algorithm to realize the sorting of radar signals. The research results show that the sorting accuracy of the PCFCM-TLRF model is higher than those of other sorting models. This model can realize the effective sorting of radar modulated signals.

Key words remote sensing; signal sorting; partial connection clustering algorithm; teaching and learning random forest algorithm; set pair analysis

OCIS codes 280.5600; 070.1060; 070.5010

1 引言

电子信息技术在军事领域的广泛应用, 使得电子对抗方式逐渐由传统机械化向信息化转变, 信息优势成为主要的争夺优势。雷达信号分选作为现代电子战中的重要组成环节, 成为电子对抗的关键技术^[1-2]。然而随着雷达技术的不断发展, 越来越多样

化的雷达体制使得战场环境日趋复杂, 依靠雷达外部参数进行信号匹配的分选技术已经很难满足战场要求^[3-4]。同时, 信号在被动接收的过程中, 电子对抗的非合作特性使先验信息不能被完整地获取, 先验信息的缺失也会使雷达信号无法进行准确识别^[5]。因此, 寻找新的有效的分选方法具有重要的研究意义。

收稿日期: 2018-09-19; 修回日期: 2018-10-14; 录用日期: 2018-10-17

基金项目: 国家自然科学基金(21706096)

* E-mail: lya_wx@jiangnan.edu.cn; ** E-mail: mengzhangm@qq.com

目前雷达调制信号的分选方法主要可以划分为监督学习、无监督学习和半监督学习。监督学习在训练过程中需要大量的标签样本,利用标签样本中的有用信息进行训练学习,学习过程中需要大量的先验信息。但在战场环境中,新型雷达的出现和雷达体制的多样化,使实际应用中有标签的样本数据较少,应用具有一定的局限性。无监督学习不需要使用标签样本,在雷达信号标签样本不能事先获取的分选领域得到广泛应用。聚类方法作为一类操作简单、复杂度低的无监督学习方法,在信号分选中的应用层出不穷,较为经典的有支持向量聚类(SVC)算法、K均值聚类(K-means)算法和模糊C均值聚类(FCM)算法。SVC算法需要计算大量的核函数,时间耗费较长。K-means算法属于硬划分算法,忽略了样本之间的联系,聚类结果往往不太理想。FCM算法考虑了样本之间的联系,不对样本进行强制分类,但FCM算法对初始化敏感。对此,文献[6]利用K-means算法对FCM算法的聚类中心进行初始化。K-means算法的计算复杂度远低于FCM算法,且在绝大部分情况下,其最终聚类中心与FCM聚类中心较为接近,以K-means算法的聚类中心作为FCM算法的初始聚类,可以提高FCM算法的效率。然而K-means算法在寻优过程中会出现极小解的情况,且FCM算法不能直接对不完整数据进行聚类。半监督学习^[7]是在少量的有标签样本数据下,利用大量无标签样本数据进行训练和分类。目前被广泛地运用于数据挖掘与分类问题中的随机森林(RF)算法^[8]是一种监督学习方法。RF算法作为一种集成式机器学习算法,可以改善单模型分类方法(如支持向量机、决策树、K近邻算法等)精度不高、容易过拟合的问题,且RF算法对噪声容忍度较高。但RF算法中并非所有的基分类器都对集成分类器的准确率产生积极作用,而且对于目前的雷达信号分选,如何有效减少训练所需的样本标签仍是一个具有挑战性的问题。

针对以上问题,本文尝试将无监督学习方法和监督学习方法融合,提出一种基于FCM算法和RF算法的雷达调制信号分选方法。在传统K-means算法的基础上引入偏系数(PCN)^[9],用PCN距离矩阵对样本进行迭代聚类,返回K-means算法的最终聚类中心,以此优化FCM算法的初始聚类,建立基于PCN改进的FCM算法(PCFCM),使其能够对具有不确定性的雷达数据进行有效的聚类分析。为进一步提高雷达信号的分选准确率,再结合“教与

学”优化(TLBO)算法^[10]优化的RF算法(TLRF),建立基于PCFCM-TLRF的雷达调制信号分选模型。

2 改进模糊C均值聚类算法

2.1 模糊C均值聚类算法

FCM算法利用隶属度建立样本与各类别之间的联系,是一种软划分算法。其基本思想是通过反复迭代更新聚类中心 C 和隶属度矩阵 U 进行聚类,使被划分到同一类样本之间的相似度最大,不同类样本之间的相似度最小。设样本集 $X = \{x_1, x_2, \dots, x_n\}$, n 表示样本个数。FCM算法将样本集划分为 K 个子集,聚类中心 $C = \{c_1, c_2, \dots, c_K\}$,则FCM算法的目标函数为

$$J(U, C) = \sum_{i=1}^n \sum_{j=1}^K u_{ij}^{\alpha} d_{ij}^2, \quad (1)$$

式中: $d_{ij} = |x_i - c_j|$ 表示样本 x_i 到类中心 c_j 的距离; i 表示 n 个样本中的第 i 个样本; j 表示 K 个子集中的第 j 个子集; α 表示模糊加权指数,最佳取值区间为 $[1.5, 2.5]$,没有特殊要求的条件下取 $\alpha = 2$;隶属度矩阵 U 的矩阵元为 u_{ij} , u_{ij} 表示样本 x_i 对类 j 的隶属度,满足约束条件

$$\sum_{j=1}^K u_{ij} = 1, 1 \leq i \leq n, u_{ij} \geq 0, 1 \leq j \leq K, \quad (2)$$

根据Lagrange乘数法,将约束条件代入(1)式,可得

$$c_j = \frac{\sum_{i=1}^n x_i u_{ij}^{\alpha}}{\sum_{i=1}^n u_{ij}^{\alpha}}, \quad (3)$$

$$u_{ij} = \left[\sum_{l=1}^K (d_{ij}/d_{il})^{2/(\alpha-1)} \right]^{-1}, \quad (4)$$

式中: l 表示第 K 个子集中的第 l 个样本。

相比于其他聚类算法,FCM算法应用最为广泛。然而,FCM算法虽然给出了样本属于各类别的程度,但是没有给出这种程度的不确定性;而且对于实际应用中的不完整数据,FCM算法无法直接应用。

2.2 基于PCN改进的模糊C均值聚类算法

PCN能够反映出研究对象在微观层面上各种不确定信息综合后的宏观发展趋势,揭示了系统的确定性和不确定性。PCN最早由赵克勤先生提出,是集对分析^[11-12]中的一种伴随函数。

把PCN理论引入K-means算法,用PCN替代K-means算法中的欧式距离来度量样本之间的联系,统一分析因处理不精确、不一致、不完整所致的各种不确定性。利用样本之间的PCN距离矩阵,对

K-means 算法进行迭代聚类, 获得最终的聚类中心, 以此优化 FCM 的初始聚类, 并用 PCN 更新聚类结束后的隶属度矩阵, 建立基于 PCN 改进的 FCM 算法。改进点如下:

1) 优化 FCM 算法的初始聚类中心

根据文献[13]可知, 一组优秀的初始聚类中心可以缩短 FCM 算法的迭代寻优过程, 提高聚类速度。用优化后的 K-means 算法对 FCM 算法的聚类中心初始化, 能够有效地减少 FCM 算法的迭代次数, 具体过程如下。

a) 初始化类别数 K , 采用最大最小距离法初始化 K-means 算法的簇中心 $C_m = \{m_k\}, k = 1, 2, 3, \dots, K$ 。

b) 计算待分类样本 x 与初始聚类中心 m_k 的 PCN。

首先将待分类样本 x 和聚类中心 m_k 看作具有 Q 个属性的集合 S 和集合 T , 组成集对 $H = (S, T)$ 。然后计算样本 x 属性集中单个属性的偏联系数 ∂u , 即

$$\partial u_{(q)} = \frac{a}{a+b} + \frac{b}{b+c}\lambda + \frac{b}{b+a}\lambda + \frac{c}{b+c}\eta, \quad (5)$$

式中: $a = S_q/T_q$, 表示两个集合的同一程度; $b = S_q/[T_q(T_q - 1)]$, 表示两个集合的差异不确定程度; $c = (T_q^2 - S_q T_q - 1)/(T_q^2 - 1)$, c 表示两个集合的对立程度; a, b, c 满足条件 $a + b + c = 1$; $q = 1, 2, \dots, Q$ 表示属性 Q 中的第 q 个属性; λ 表示差异表示符, 取值范围为 $[-1, 1]$; η 表示对立标记符, 取值 $\eta = -1$ 。

最后计算样本 x 和聚类中心 m_k 的 PCN, 即

$$D_{xk} = \sum_{q=1}^Q \partial u_{(q)}, \quad (6)$$

式中: D_{xk} 表示样本 x 和聚类中心 m_k 之间联系的确定性程度。

c) 确定样本标签所属类别。根据(6)式计算样本 x 与 K 个簇中心 C_m 的 PCN 距离, 若 $D_{xk} = \max\{D_{xk}\}, k = 1, 2, 3, \dots, K$, 则认为样本 x 与聚类中心 m_k 的联系程度最大, 即 $x \in k$ 。

d) 修正簇中心 C_m 。根据(7)式重新计算各个簇的聚类中心, 即

$$m_k = \frac{1}{|k|} \sum_{x \in k} x, \quad (7)$$

式中: $|k|$ 表示属于第 k 类的样本总数。

e) 重复步骤 c)~d), 直至各个类中心前后两次的误差小于 10^{-4} , 返回最终聚类中心。

2) 更新 FCM 算法聚类后的距离矩阵

样本标签的确定不仅依赖于簇中心 C_m , 还和样本与 C_m 的联系程度有关, 聚类结束后, 计算样本与 C_m 的 PCN 更新距离矩阵, 过程如下。

a) 利用返回的聚类结果作为初始簇中心 C_m , 根据(4)式确定隶属度 U 。

u_{xk} 表示样本 x 对 m_k 的隶属度, 满足约束条件

$$\sum_{k=1}^K u_{xk} = 1, 1 \leq x \leq n, u_{xk} \geq 0, 1 \leq k \leq K. \quad (8)$$

b) 利用隶属度矩阵更新算法的簇中心 C_m , 直至前后两次得到的各个簇的聚类中心 m_k 的误差小于 10^{-4} 。

c) 根据(6)式计算待分类样本与簇中心 C_m 的 PCN 距离矩阵, 以此确定待分类样本的所属类别, 若 $D_{xk} = \max\{D_{xk}\}, k = 1, 2, 3, \dots, K$, 则 $x \in k$ 。

3 优化随机森林算法

3.1 随机森林算法

RF 算法是一种由若干基础分类器“决策树”组成的集成学习方法。其基本思想是将 bagging 算法和 random subspace 算法结合, 生成多棵互不相关的决策树, 每棵决策树都能够独立地利用训练样本进行训练, 训练完成后, 对这些决策树采取投票选取机制, 确定最终分类结果。本文 RF 算法中的基础分类器采用 C4.5 分类算法生成。C4.5 算法选择信息增益率最大的属性作为当前节点的分裂属性, 生成决策树。

信息增益率的计算过程如下:

1) 信息增益

$$G(A) = I(D) - I_A(D), \quad (9)$$

式中: $I(D) = -\sum_{i=1}^m p_y \lg p_y$ 表示样本集 D 的熵;

$I_A(D) = \sum_{z=1}^v \frac{|D_z|}{|D|} I(D_z)$ 表示属性 A 对样本集 D

划分的期望信息; $1 \leq y \leq m$ 表示样本集 D 按类别划分为 m 个子集; v 表示样本集 D 按属性 A 的属性值划分为 v 个子数据集; $|D_z|$ 表示第 z 个子数据集

2) 信息增益率

$$R_G(A) = \frac{G(A)}{S_A(D)}, \quad (10)$$

式中: $S_A(D) = -\sum_{z=1}^v \frac{|D_z|}{|D|} \times \lg \frac{|D_z|}{|D|}$ 表示属性 A

的“分裂信息”。

RF算法的性能依赖于各个基础分类器的性能,然而在分类过程中并非所有的基础分类器都具有优秀的分类性能^[14-16]。如何保留优质的分类器,剔除不好的分类器,找出最优的决策树组合成为问题的关键。

3.2 TLBO 算法优化随机森林

TLBO算法是一种通过模拟班级中教师“教学”过程与学生“学习”过程,找出最优解的群智能优化算法。TLBO算法因其全局搜索能力强、收敛速度快、参数设置少、结构简单等特点,已成功应用于很多工程问题^[17]。在TLBO算法中,班级是搜索空间中的种群,教师和学生是种群中的个体,成绩是算法中的适应度。

利用TLBO算法对随机森林中的决策树进行优化搜索,寻找最优的决策树组合,构成新的集成,以此提高RF算法的分选性能。改进过程如下:

a) 利用随机采样技术对训练集进行随机采样,生成 T_{TN} 棵决策树。

b) 初始化不同的决策树组合作为初始种群,计算训练样本在不同决策树组合下的分类结果,即计算不同决策树组合的适应度。

c) 选择种群中适应度最大的个体作为教师 $X_{teacher}$,通过学生和教师之间、学生和学生之间的差异度进行个体迭代更新,寻找最优个体。

1) 学生和教师之间的差异度表示为

$$\phi = r(X_{teacher} - T_{TF}X_{mean}), \quad (11)$$

式中: $r = \text{rand}(0, 1)$ 表示学习步长; $T_{TF} = \text{round}[1 + \text{rand}(0, 1)]$ 表示教学因子; X_{mean} 表示班级的平均成绩。

2) 学生和学生之间的差异度表示为

$$\phi = \begin{cases} r_h(X_h - X_g), f(X_h) > f(X_g) \\ r_h(X_g - X_h), f(X_g) > f(X_h) \end{cases}, \quad (12)$$

式中: $r_g = \text{rand}(1, N_{NP})$ 表示第 g 个学生的学习因子, N_{NP} 表示种群个数; $f(X_g)$ 表示第 g 个学生的成绩; $f(X_h)$ 表示第 h 个学生的成绩。

d) 判断是否达到最大迭代次数,若是,则返回最优个体,即最优决策树组合;若不是,则继续步骤c)~步骤d)。

4 PCFCM-TLRF 信号分选模型

信号分选主要包括两个阶段,一是特征提取;二是分类器设计。特征提取是一种非线性变换,提取能够反映信号特征的信息,完成信号到特征的转换。

分类器设计是根据第一阶段提取的信号特征,选择合适的分类器对特征进行分选识别。根据信号分选的阶段划分,建立基于PCFCM-TLRF的雷达调制信号分选模型。首先提取雷达调制信号的熵特征作为信号样本集,利用PCFCM软划分的特点将具有高度相似性的样本进行归类,提取性能较好的训练样本,使样本规律更加容易识别;结合具有较强学习能力和泛化能力的TLRF算法,进一步提高分选准确率。模型优化过程如图1所示,具体步骤包括:数据准备、聚类分析、优化决策树组合、信号识别。

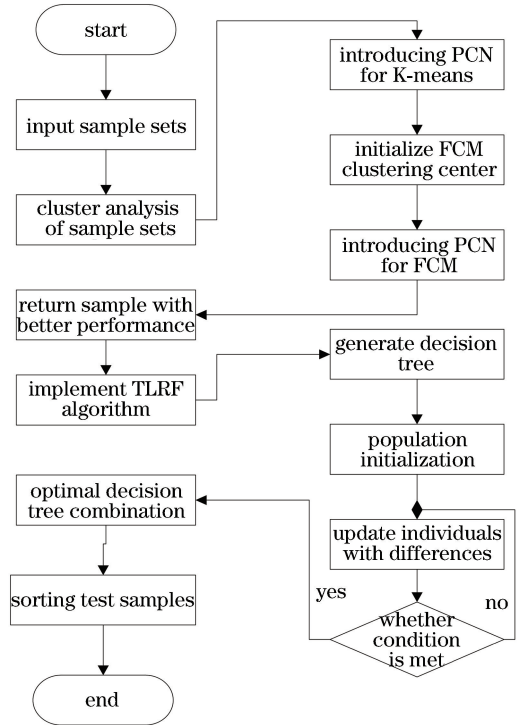


图1 基于PCFCM-TLRF的雷达调制信号分选模型

Fig. 1 Radar modulated signal sorting model based on PCFCM-TLRF

1) 数据准备

由文献[18-19]可知,利用信号的熵特征进行分选可以在一定程度上弥补使用脉内特征和外部参数特征分选时存在的识别种类少、噪声考虑少的不足。本研究提取7种雷达调制信号(参数设置见表1)的香农熵、范数熵、奇异值熵作为雷达信号的特征数据样本集。信噪比(SNR)范围为 $-10 \text{ dB} \sim 10 \text{ dB}$,每种信号每隔 2 dB 产生100个样本,共7700个样本。其中,训练样本是每隔 2 dB 的信噪比下的700个特征样本,测试样本是从对应的训练样本中随机抽取的130个特征样本。

表1 雷达调制信号参数设置

Table 1 Parameter setting of radar modulated signals

Signal	Sampling	Carrier	Encoding	Pulse	Bandwidth /MHz
	frequency /MHz	frequency /MHz		width / μ s	
LFM	120	30		10	20
NLFM	120	30		10	20
CW	120	30		10	20
BPSK	120	30	13-bit Barker code	13	20
FSK	120	30	13-bit Barker code	13	20
MPSK	120	30	Frank code	10	20
QPSK	120	30	Frank code	10	20

Note: LFM is linear frequency modulation; NLFM is nonlinear frequency modulation; CW is continuous wave; BPSK is binary phase shift keying; MPSK is M -ary phase shift keying; QPSK is quaternary phase shift keying.

2) 聚类分析

以训练样本集作为聚类对象,采用 PCFCM 算法对其进行聚类,返回 PCN 高的样本组成新的样本集 $D = \{F_{\text{date}}, F_{\text{label}}\}$, F_{date} 表示样本集中的样本特征, F_{label} 表示样本集中的样本标签。聚类过程中, PCFCM 算法参数设置如下:聚类簇个数 $K = 7$,模糊加权指数 $\alpha = 2$,聚类过程中用 PCN 衡量迭代过程中样本与簇中心 C_m 的关系。

3) 优化决策树组合

将聚类结果 $D = \{F_{\text{date}}, F_{\text{label}}\}$ 作为随机森林的数据集,采用 bootstrap 方法随机从数据集 D 中抽取 300 个训练样本作为训练集 S ,抽取 15 次 ($T_{\text{TN}} = 15$),生成 T_{TN} 棵决策树;然后采用 TLBO 算法优化随机森林,根据适应度函数寻找最优决策树组合。寻优过程中,TLRF 算法参数设置如下:种群个数 $N_{\text{NP}} = 15$,最大迭代次数 $N_{\text{NI}} = 150$,适应度函数为

$$f(S) = \operatorname{argmax}_{c=1,2,3,\dots,K} \{I[f_i(S) = c]\}, \quad (13)$$

式中: K 表示样本中的簇个数; $f_i(S) = i$ 表示样本 S 在第 t 棵决策树上的输出; c 表示 K 簇中的某一簇; $I(\cdot)$ 表示满足括号中条件的样本个数。

4) 信号识别

利用步骤 3) 中返回的最优决策树组合对测试样本进行分选,将样本分选结果与样本实际标签作比较,检验分选效果。

5 仿真实验与分析

5.1 改进算法性能分析

选取来自加利福尼亚大学 UCI 数据库里的 Iris Flowers 数据集、Wine Quality 数据集作为实验数据集,测试 PCFCM 算法和 TLRF 算法的有效性。Iris Flowers 数据集以鸢尾花的特征作为数据来源,

数据集包含 3 类鸢尾花,每类 50 个数据,共 150 个数据,每个数据包含 4 个属性(花萼长度、花萼宽度、花瓣长度、花瓣宽度);Wine Quality 数据集以葡萄酒的化学成分作为数据来源,数据集包含 3 种不同起源的葡萄酒,共 178 个数据,每个数据包含 13 个属性。Iris Flowers、Wine Quality 是多分类问题中常用的数据集。

设置 K-means、FCM、PCFCM 三种算法的聚类簇个数 $K = 3$,迭代终止条件为前后两次相同类中心的误差 $e < 10^{-4}$ 。表 2 给出 K-means、FCM、PCFCM 三种算法对 Iris Flowers 数据集、Wine Quality 数据集的聚类结果。从表 2 可知,K-means 算法结构简单、复杂度低,用时相对较短,但准确率不高;FCM 算法准确率有所提升,但对初始聚类敏感,导致用时较长;PCFCM 算法在准确率、时间上均优于 FCM 算法和 K-means 算法,说明引入 PCN 改进 FCM 算法可以提高 FCM 算法的运行速度、获得较高的分类准确率。

表 2 K-means 算法、FCM 算法、PCFCM 算法性能比较

Table 2 Performance comparison among K-means algorithm, FCM algorithm and PCFCM algorithm

Data set	Algorithm	Mean	Time
		accuracy /%	average /s
Iris Flowers	K-means	85.53	0.1713
Iris Flowers	FCM	93.00	0.2326
Iris Flowers	PCFCM	99.75	0.1045
Wine Quality	K-means	68.54	0.4387
Wine Quality	FCM	70.22	0.6890
Wine Quality	PCFCM	75.39	0.4264

设置 TLRF 算法、RF 算法的决策树个数为 $T_{\text{TN}} = 15$ 。随机抽取 Iris Flowers、Wine Quality 各数据集的 90 组数据作为训练样本,30 组数据作为测试样本,将测试样本分别输入已经训练好的 SVM

算法、RF 算法、TLRF 算法。三种算法对 Iris Flowers 数据集、Wine Quality 数据集的分类结果如表 3 所示。从表 3 可知,与 RF 算法、SVM 算法相比,TLRF 算法在对测试样本分选时所用时间更少、准确率更高,说明用 TLBO 优化 RF 算法后构成的新集成算法 TLRF 能够用更少的基础分类器获得更高的准确率,提高了 RF 算法的集成性能。PCFCM 算法、TLRF 算法的良好表现,为基于 PCFCM-TLRF 的雷达调制信号分选模型奠定了基础。

表 3 SVM 算法、RF 算法、TLRF 算法性能比较

Table 3 Performance comparison among SVM algorithm, RF algorithm and TLRF algorithm

Data set	Algorithm	Mean accuracy / %	Test time average / s
Iris Flowers	SVM	92.67	0.0122
Iris Flowers	RF	95.00	0.0350
Iris Flowers	TLRF	99.17	0.0014
Wine Quality	SVM	70.78	0.0437
Wine Quality	RF	95.33	0.0272
Wine Quality	TLRF	96.61	0.0017

5.2 PCFCM-TLRF 模型性能分析

为验证 PCFCM-TLRF 模型的有效性,将训练样本作为 TLRF 模型、传统 K-means 优化 FCM (KFCM)算法与 SVM 算法结合的 KFCM-SVM 模型^[6]、FCM-TLRF 模型、PCFCM-TLRF 模型的输入,比较算法的输出结果。

图 2 给出同等条件下 TLRF 模型、KFCM-SVM 模型、FCM-TLRF 模型、PCFCM-TLRF 模型在信噪比大于 -10 dB 时雷达调制信号分选准确率与信噪比的拟合图。从图 2 可知,4 种模型对雷达调制信号的分选准确率基本都随着信噪比的增加而提高。当信噪比小于 0 dB 时,信号受噪声影响较大,特征空间容易出现重叠,使得分选结果较差。PCFCM-TLRF 模型通过对样本进行聚类分析,返回数据集中 PCN 较高的样本,使得 PCFCM-TLRF 模型的分选准确率与其他 3 种模型相比具有明显优势。当信噪比大于 0 dB 时,PCFCM-TLRF 模型的优势减小,但 PCFCM-TLRF 模型的分选准确率先与其他模型达到 100%,说明 PCFCM-TLRF 模型的性能优于其他 3 种模型,能够较好地应用于雷达信号分选。

图 3 给出各雷达调制信号的分选准确率与信噪比的曲线拟合图,从图 3 可知,各信号的分选准确率都随着信噪比的增加而上升。所有信号中,CW 信

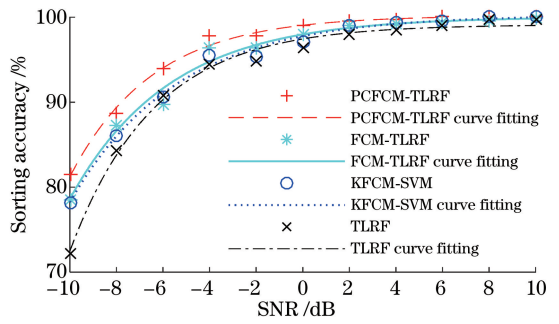


图 2 4 种模型的分选准确率与信噪比的拟合曲线
Fig. 2 Fitting curves of sorting accuracy and SNR for four models

号和 BPSK 信号分选效果最好,信噪比为 -6 dB 时准确率可以达到 100%;NLFM 信号、FSK 信号的分选效果在小于 -6 dB 的条件下不太理想。但 NLFM 信号、FSK 信号总体上升速度很快,且在大于 -4 dB 条件下,分选准确率有明显提升,说明 NLFM 信号、FSK 信号受信噪比的影响较大。其他信号的分选效果虽然有所波动,但在信噪比大于 2 dB 时分选准确率都逐渐趋于 100%,说明 PCFCM-TLRF 算法可以有效实现多种雷达调制信号的分选。

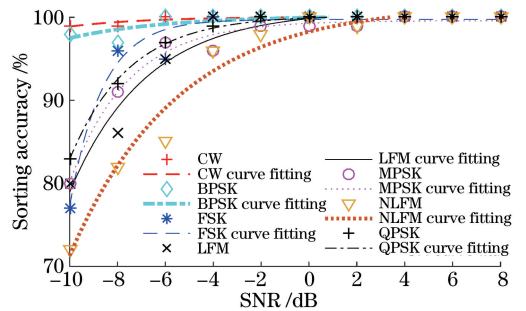


图 3 各种调制信号的分选准确率与信噪比的拟合曲线
Fig. 3 Fitting curves of sorting accuracy and SNR for various kinds of modulated signals

6 结 论

采用改进的 FCM 算法对雷达信号的特征数据集进行聚类分析;采用优化的 RF 算法对聚类分析后的样本集采样训练,获得最优决策树组合。采用无监督学习和监督学习相结合的方法,应对雷达信号分选过程中存在的标签样本数据少、低信噪比条件下分选准确率不高等问题。仿真实验结果表明,PCFCM-TLRF 模型相比于 TLRF 模型、KFCM-SVM 模型和 FCM-TLRF 模型具有更有效、更准确的雷达信号识别效果,为雷达信号分选准确度的提高提供了一种新的方法。但在信噪比小于 -6 dB

的条件下,NLFM信号和FSK信号的分选效果不理想,需要进一步提高。

参 考 文 献

- [1] Zhao G Q. Principle radar countermeasure[M]. 2nd ed. Xi'an: Xi'an University of Electronic Science and Technology Press, 2012: 87-105.
赵国庆. 雷达对抗原理[M]. 2版. 西安: 西安电子科技大学出版社, 2012: 87-105.
- [2] Yang F Q, Li Z, Luo Z L. Method of modulation recognition of mixed modulation signal [J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2014, 53(1): 42-46.
杨发权, 李赞, 罗中良. 混合调制信号调制识别方法[J]. 中山大学学报(自然科学版), 2014, 53(1): 42-46.
- [3] Zhu B. Study on feature extraction and evaluation method of radar emitter signals [D]. Chengdu: Southwest Jiaotong University, 2015: 17-24.
朱斌. 雷达辐射源信号特征提取与评价方法研究[D]. 成都: 西南交通大学, 2015: 17-24.
- [4] Zhang W X, Sun F L, Wang B. Radar signal intrapulse feature extraction based on improved wavelet transform algorithm [J]. International Journal of Communications, Network and System Sciences, 2017, 10(8): 118-127.
- [5] Fu Y, Wang X, Zhou D Q, *et al.* Recognition method of radar signal based on SVD of ambiguity function and improved S3VM algorithm [J]. Computer Engineering and Applications, 2017, 53(6): 264-270.
符颖, 王星, 周东青, 等. 基于模糊函数SVD和改进S3VM的雷达信号识别[J]. 计算机工程与应用, 2017, 53(6): 264-270.
- [6] Gao W W, Shen J X, Wang Y L, *et al.* Automated detection of bright lesions of diabetic retinopathy based on improved and fast FCM and SVM [J]. Chinese Journal of Biomedical Engineering, 2013, 32(3): 305-312.
高玮玮, 沈建新, 王玉亮, 等. 改进的快速FCM及SVM实现糖网白色病灶的自动检测[J]. 中国生物医学工程学报, 2013, 32(3): 305-312.
- [7] Wu J Q, Tian X L. Anarrow-band radar target recognition system based on semi-supervised learning [J]. Journal of China Academy of Electronics and Information Technology, 2015, 10(1): 49-53.
吴剑旗, 田西兰. 一种基于半监督学习的窄带雷达目标识别系统[J]. 中国电子科学研究院学报, 2015, 10(1): 49-53.
- [8] Breiman L. Random forest [J]. Machine Learning, 2001, 45(1): 5-32.
- [9] Zhao K Q. Partial connection number [C] // Proceedings of Chinese Artificial Intelligence 2005. Beijing: Beijing University of Posts and Telecommunications Press, 2005, 6: 884-885.
赵克勤. 偏联系数[C] // 中国人工智能进展2005论文集. 北京: 北京邮电大学出版社, 2005, 6: 884-885.
- [10] Rao R V, Savsani V J, Vakharia D P. Teaching-learning-based optimization: A novel method for constrained mechanical design optimization problems [J]. Computer-Aided Design, 2011, 43(3): 303-315.
- [11] Li R, Liu Y A, Wang G. Radar signal sorting method based on modified set pair analysis clustering [J]. Modern Electronics Technique, 2014, 37(9): 8-11.
黎蓉, 刘以安, 王刚. 基于改进集对分析聚类的雷达信号分选方法[J]. 现代电子技术, 2014, 37(9): 8-11.
- [12] Zhang X H, Liu Y A, Cao N S, *et al.* Radar signal sorting method based on set pair analysis [J]. Modern Radar, 2010, 32(2): 35-37.
张秀辉, 刘以安, 曹宁生, 等. 基于集对分析的雷达信号分选算法[J]. 现代雷达, 2010, 32(2): 35-37.
- [13] Zhang Y K, Yin L X, Sun J G. Fuzzy clustering algorithm based on the improved genetic algorithm [J]. CAAI Transactions on Intelligent Systems, 2015, 10(4): 627-635.
张永库, 尹灵雪, 孙劲光. 基于改进的遗传算法的模糊聚类算法[J]. 智能系统学报, 2015, 10(4): 627-635.
- [14] Li D W, Yang F B, Wang X X. Multisource remote sensing classification based on random forest and adaptive weighted D-S evidence synthesis [J]. Laser & Optoelectronics Progress, 2016, 53(3): 031001.
李大威, 杨风暴, 王肖霞. 基于随机森林与D-S证据合成的多源遥感分类研究[J]. 激光与光电子学进展, 2016, 53(3): 031001.
- [15] Kong Q Q, Ding X Q, Gong H L. Application of improved random forest pruning algorithm in tobacco origin identification of near infrared spectrum [J]. Laser & Optoelectronics Progress, 2018, 55(1): 013006.
孔清清, 丁香乾, 宫会丽. 改进的修剪随机森林算法在烟叶近红外光谱产地识别中的应用研究[J]. 激光

- 与光电子学进展, 2018, 55(1): 013006.
- [16] Cai J X, Feng G C, Tang X, *et al.* Human action recognition based on local image contour and random forest [J]. *Acta Optica Sinica*, 2014, 34 (10): 1015006.
蔡加欣, 冯国灿, 汤鑫, 等. 基于局部轮廓和随机森林的人体行为识别 [J]. *光学学报*, 2014, 34(10): 1015006.
- [17] Xu J H, Wang Y. Energy efficiency optimization for discrete manufacturing workshop based on discrete teaching-learning-based optimization algorithm [J]. *Journal of System Simulation*, 2016, 28(12): 3019-3026.
徐军辉, 王艳. 基于改进教与学算法的离散制造车间能效优化 [J]. *系统仿真学报*, 2016, 28(12): 3019-3026.
- [18] Li J C. Radar signal recognition based on three dimensional entropy features [J]. *Journal of Shanghai Dianji University*, 2015, 18(3): 136-140.
李靖超. 基于三维熵特征的雷达信号识别 [J]. *上海电机学院学报*, 2015, 18(3): 136-140.
- [19] Liu K, Han J, Huang Q H. Sorting radar signal based on the improved resemblance coefficient and singular spectrum entropy [J]. *Modern Radar*, 2015, 37(9): 80-85.
刘凯, 韩嘉宾, 黄青华. 基于改进相像系数和奇异谱熵的雷达信号分选 [J]. *现代雷达*, 2015, 37(9): 80-85.