

一种基于集成学习的干扰频点检测方法

朱吉利¹, 范振军^{2*}

¹中国电子科技集团公司第十研究所, 四川 成都

²成都天奥集团有限公司, 四川 成都

收稿日期: 2024年12月16日; 录用日期: 2025年1月8日; 发布日期: 2025年1月20日

摘要

针对传统干扰频点检测方法所存在的精度不够高、易受噪声干扰、实时性较差问题, 提出了一种基于集成学习的干扰频点检测方法。首先, 对仿真数据进行归一化、类别标注等处理并生成3种不同规模的数据集; 然后, 构建多个基于XGBoost的频点干扰二分类模型; 最后, 通过并联所构建的二分类模型对通信帧数据进行频点干扰识别以得到所有被干扰的频点。仿真实验表明, 基于并联XGBoost模型的干扰频点检测在仿真数据测试集上的精确率达96.8%, 平均推理时间小于5 ms, 验证了所提方法的高效性。

关键词

集成学习, 干扰频点检测, XGBoost

A Detection Method of Jamming Frequency Based on Ensemble Learning

Jili Zhu¹, Zhenjun Fan^{2*}

¹The 10th Research Institute of China Electronics Technology Group Corporation, Chengdu Sichuan

²Chengdu Spaceon Group Co., Ltd., Chengdu Sichuan

Received: Dec. 16th, 2024; accepted: Jan. 8th, 2025; published: Jan. 20th, 2025

Abstract

Aiming at the problems of low precision, easily disturbed by noise, and poor real-time performance of traditional jamming frequency detection methods, a new method of jamming frequency detection based on ensemble learning was proposed. Firstly, the simulation data were processed by normalization, category labeling, etc., and three datasets of different sizes were generated. Then, multiple binary classification models based on XGBoost are constructed for a single frequency. Finally, the

*通讯作者。

binary classification model constructed in parallel is used to identify the frequency jamming of communication frame data to obtain all the jamming frequencies. Simulation experimental results show that the proposed method has an average accuracy of more than 96% and an average inference time of less than 5 ms on the simulation data set, which verifies the high efficiency of the proposed method.

Keywords

Ensemble Learning, Jamming Frequency Detection, XGBoost

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着无线通信系统面临的电磁环境越来越复杂,使得通信过程中受到的干扰日趋严重。快速准确的干扰频点检测有助于提升通信系统的抗干扰能力。

传统的干扰检测方法主要有基于预设门限值的能量检测法、连续均值消除(Consecutive Mean Excision, CME)与双阈值法[1]-[4],以及匹配滤波器检测法[5]和循环平稳特征检测法[6]等。其中,能量检测法通过计算接收信号在一定时间内的能量,并与预设的门限值进行比较,从而判断是否存在干扰信号,该方法实现简单、适用范围较广,但预设的门限值对检测结果影响较大且容易受到噪声干扰,导致精度不高。CME方法是一种有效的频谱干扰检测方法,尤其适用于窄带干扰,但依然受预设门限值的影响。为克服CME法的局限性,双阈值法通过引入两个不同的阈值(即高阈值和低阈值)以便更好区分正常信号和干扰信号,在认知无线电领域应用广泛。匹配滤波器法能够在输出端使信号的信噪比(SNR)达到最大,当接收到的信号中包含与已知信号相似的成分时,匹配滤波器能够最大化地放大这些成分,从而有效地检测出信号中的干扰,特别适用于加性噪声的场景,但依赖于已知信号的准确性。循环平稳特征检测法是一种基于循环统计量的时频分析方法,通过利用信号的循环平稳特性来完成对干扰信号的检测和参数估计,具有抗噪声能力强、性能稳定等优点,特别适用于处理具有周期性和非平稳性的信号,但计算较慢且依赖于先验信息,限制了其应用范围。王正欢[7]等人通过改进周期图法估计信号频谱并使用前向连续均值消除(FCME)实现鲁棒性的干扰频点检测。李孟超[8]等人将FCME法用于跳频通信系统的干扰检测,提升了门限设计的有效性。上述这些方法主要依赖于门限值设定、人工特征设计等传统方式实现,存在检测精度不够高、应用范围受限、易受噪声干扰等问题。

机器学习(Machine Learning, ML),它能够从大量数据中学习到所需的潜在规律,并生成可以应用于新数据的预测模型,已在干扰识别[9][10]、智能抗干扰[11][12]、智能通信[13]等相关领域进行实践应用,为干扰检测[14]提供了一个新的思路。本文针对传统干扰检测方法存在的不足,利用机器学习技术对跳频通信系统的干扰频点检测进行了实验研究,结果表明,基于集成学习的方法能显著提升干扰频点检测的精度,且实时性高。

2. 基于集成学习的干扰频点检测方法

2.1. 设计思路

在机器学习任务中,分类是一种监督学习方法,用于根据输入数据预测标签。通过使用可用的训练

数据训练机器学习模型, 以对输入数据执行分类任务。机器学习模型构建流程如图 1 所示, 涵盖了数据、特征设计、算法模型训练与参数调优、评测和应用等环节。

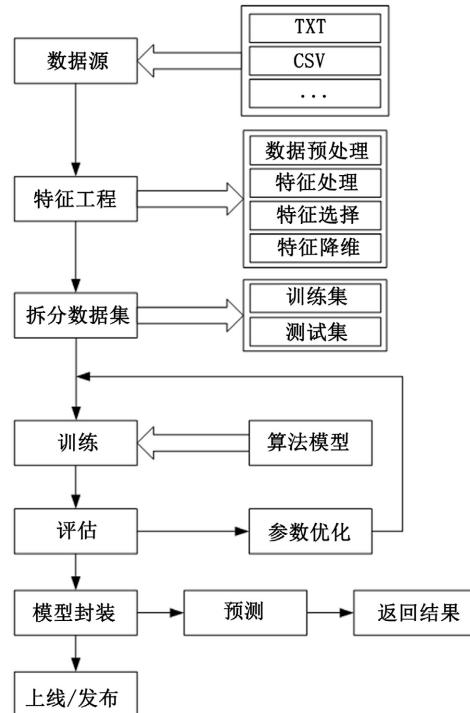


Figure 1. Flowchart of machine learning model construction
图 1. 机器学习模型构建流程图

对于干扰频点检测, 这个问题在机器学习领域属于多标签分类的研究领域。想要检测出跳频序列中具体被干扰的频点, 可以将该问题转换为一个多个二分类模型训练预测的任务, 即针对单个频点的干扰判别, 可以看作是基于信号的特征参数来训练分类器从而实现跳频频点是否被干扰的分类识别, 因此, 可以针对跳频频率集中的每个频率, 通过设计合适的二分类器以及相应的识别准则, 分别训练 N 个分类器, 对每个样本预测 N 次得到最后的组合输出, 以此实现对整体输入信号的干扰频点检测。本文采用典型的机器学习分类器中的极致梯度提升(eXtreme Gradient Boosting, XGBoost) [15]模型来构建干扰频点检测算法, 其流程如图 2 所示。

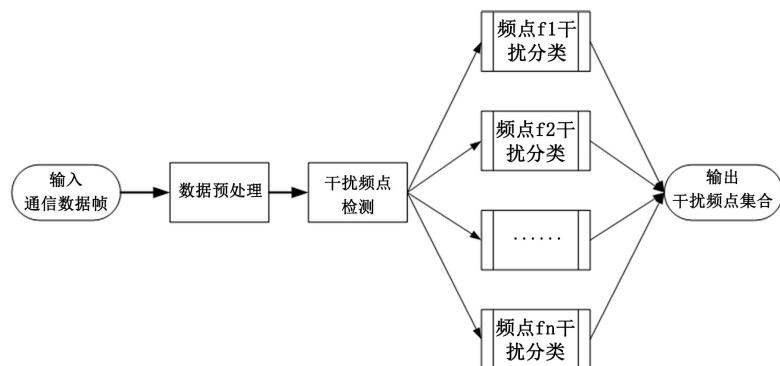


Figure 2. Flowchart of jamming frequency detection
图 2. 干扰频点检测算法流程图

2.2. 集成学习模型

集成学习即分类器集成, 是一种机器学习范式, 通过构建并结合多个学习器来完成学习任务。一般流程是: 先通过学习生成一组“个体学习器”, 再用某种策略将它们结合起来[16]。XGBoost 算法是在梯度提升决策树(Gradient Boosting Decision Tree, GBDT) [17]的基础上提出的一种基于分类回归树的集成学习算法, 通过采用泰勒二次展开公式求解目标函数, 多用于分类、预测、回归等问题[15] [18], 其具有高准确度、不易过拟合、可扩展性强、性能高等优点, 并且能分布式处理高维特征值, 效率高于同类算法 10 倍以上[19]。

对所给定的样本和样本标签 X, Y , 该算法的目标函数为:

$$obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (1)$$

上式中, l 为损失函数, 表示模型的拟合程度, 其中 y_i 为样本 x_i 的真实标签, \hat{y}_i 为模型的预测标签; Ω 为正则化项, 有效降低了模型过拟合的可能性, 其中 f_k 为第 k 棵树的模型。XGBoost 算法的建树过程、Boosting 过程均以目标函数为基础进行, 中间迭代过程的最终目的就是最小化目标函数。

2.3. 基于并联 XGBoost 的干扰频点检测模型

基于并联 XGBoost 的干扰频点检测, 分为两大阶段:

阶段 1: 针对 fn 个频点, 设计合理的参数策略(主要是学习率、树深度、损失函数等), 让 XGBoost 直接通过数据中的特征学习这种变化的规律, 以获得样本数据集上的分类能力, 从而得到干扰分类器, 即分别训练 fn 个频点干扰分类器 c_{f_i} , 并最终获得跳频频点集的干扰分类器集合 $C = \{c_{f_0}, \dots, c_{f_{n-1}}\}$ 。

阶段 2: 基于分类器集合 C , 将通信帧按频点 f_i 进行拆分得到频点 f_i 的数据样本 d_{f_i} , 再用频点 f_i 对应的分类器 c_{f_i} 进行预测得到频点 f_i 的干扰识别结果, 以此类推, 得到该通信帧中所有频点的干扰识别结果, 从而获得被干扰的频点序列, 记为 $GR_{hp} \subseteq F_{hp}$ 。

3. 实验与分析

3.1. 实验仿真数据

本文按表 1 所示的主要参数构建仿真通信系统并生成仿真数据。本文构建的通信系统包括 3 个通信节点, 即干扰节点、通信发送节点和接收节点。假定跳频频率集 $F = \{f_0, \dots, f_{fn-1}\}$, 表示跳频系统的所有跳变频点, 其频点数记为 fn ; 非空子集 $F_{hp} \subseteq F$, 表示实际跳频通信帧中的频点集合, 其频率数 $fk \leq fn$ 。

Table 1. Main parameters of communication system

表 1. 通信系统主要参数

系统参数	取值
信道总数 L	11 条
干扰信号带宽	20 MHz
跳频带宽 bw	10 MHz
跳变频点记录数	6 个
跳频中心频率 cf	2500 MHz
工作频段	2450 MHz~2550 MHz
通信节点距离	100 m

生成仿真数据时假定干扰信号的功率为 6 W, 信道噪声为 5 dB。通过 MATLAB 实现仿真系统搭建并生成仿真数据共计 13,4000 条, 其中干扰与非干扰状态分别标识为 1 与 0, 其数据量分别为 1,2000 条与 12,2000 条, 记为跳频干扰数据集(FHJD)。

首先, 基于 FHJD, 将原始通信帧数据按频点 f_i 抽取成 fn 个子集 D_{f_i} , 其单个样本包含的信息如表 2 所示。

Table 2. The primary fields of a single frequency sample data
表 2. 单个频点样本数据的主要字段

序号	字段	数据类型
1	采样时间戳	字符串
2	速率	字符串
3	信噪比	浮点型
4	信号功率	浮点型
5	频率信号功率	浮点型

然后, 按随机切分的方式并以干扰与非干扰的比例分别为 1:1、1:5、1:10 三种情形生成不同规模(标识为 10 k、50 k、100 k)的数据集。最后, 将各规模的数据集以 6:2:2 的比例按分层抽样的方式划分为训练集、验证集、测试集, 用于模型训练与验证。

3.2. 实验环境

本文实验所用环境的基本信息为: 操作系统平台为 Windows 10, 处理器为 Intel(R)Core(TM)i7-10700CPU@2.90GHz, 内存不少于 64 GB, 算法代码的编程语言为 Python 3.7, 依赖包包括 Pandas, Numpy, Scikit-Learn, XGBoost 等。

3.3. 评价指标

从机器学习角度看, 干扰频点检测实则属于一个二分类问题。为进一步对比模型的训练结果及实际测试效果, 本文采用准确率、召回率、精确率及 F1 值等评价指标。

准确率是指预测正确的样本与样本集的比例, 其值越高越好, 计算公式如下:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

召回率是指预测为正样本与实际为正样本的比例, 其值越高越好, 计算公式如下:

$$R = \frac{TP}{TP + FN} \quad (3)$$

精确率是指实际为正样本与预测为正样本的比例, 其值越高越好, 计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (4)$$

F1-score 是精确率和召回率的加权调和平均, 多用于评价模型分类效果的均衡性, 计算公式如下:

$$F1 = \frac{2 * P * R}{P + R} \quad (5)$$

上述各式中, P 为正例, N 为负例, T (True)表示真, F (False)表示否, TP 代表真正样本, FP 代表假正样本, TN 代表真负样本, FN 代表假负样本。

3.4. 模型训练参数

基于 XGBoost 模型, 在不同参数条件下训练频点干扰分类模型, 以获得性能最好的模型参数组合。主要对模型所依赖的学习率、gamma、决策树最大深度、类别数目进行调参, 具体实验参数如表 3 所示。

Table 3. Major parameter settings for the XGBoost model

表 3. XGBoost 模型的主要参数设置

参数名称	参数描述	参数取值
eta	指定学习率	0.02
gamma	叶节点进行分支所需损失减小的最小值	0.01
max_depth	树的最大深度	3
min_child_weight	孩子节点中的最小样本权重和	1
subsample	对训练集二次抽样比例, 防止过拟合	0.8
colsample_bytree	构建每棵树的子抽样比例	0.8
obj	指定学习任务及其目标函数	binary: logistic
num_class	数据集类别数	2

在训练中, 采用对数损失函数作为分类模型的目标函数, 且使用 Scikit-Learn 提供的 GridSearchCV 方法对模型的主要参数进行调优。

3.5. 实验结果及分析

模型的准确度主要体现在对测试集的预测准确性上。通过上述模型训练得到在不同数据规模下基于 XGBoost 的频点干扰分类模型, 具体实验结果如下。

实验 1: 基于 XGBoost 的频点干扰分类

按表 3 所述参数, 训练基于 XGBoost 的频点干扰分类模型, 其在训练集和验证集上的学习曲线如图 3 所示。

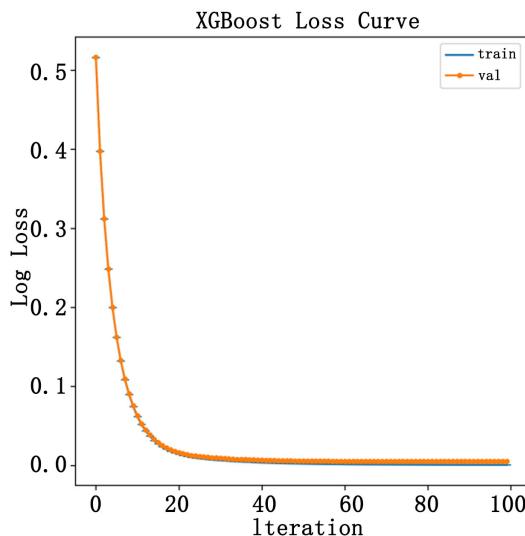


Figure 3. Loss curve of XGBoost model

图 3. XGBoost 模型的损失曲线图

从图3可知, XGBoost 模型的损失随着迭代次数的增加而逐渐减小, 在迭代 20 次后趋于平稳, 表明模型的训练效果较好。验证集和训练集的损失在趋于平稳后很接近, 表明模型的泛化性能较好。

取频点 f_0 在不同规模数据集上, 频点干扰分类模型在训练集与测试集上的实验结果如表4 所示。

Table 4. Experimental results of frequency jamming classification based on XGBoost

表4. 基于 XGBoost 的频点干扰分类实验结果

数据规模	准确率(%)		时间性能(ms)
	训练集	测试集	
10 k	96.35	96.13	4.18
50 k	96.63	96.59	4.21
100 k	96.88	96.75	4.33

从表4 可知, 在不同数据规模下, 基于 XGBoost 的频点干扰分类模型在所设定的参数条件下, 平均准确率最优为 96.88%, 平均推理时间在 4.4 ms 内。

实验 2: 干扰频点检测模型测试

根据实验 1 的结果, 构建基于并联 XGBoost 模型的干扰频点检测方法, 其在不同规模数据集下的验证结果如表5 所示。

Table 5. Experimental results of frequency jamming classification based on multi-XGBoost on test set

表5. 基于并联 XGBoost 模型的干扰频点检测方法在测试集上的实验结果

方法	数据规模	精确率(%)	召回率(%)	F1-score	时间性能(ms)
Multi-XGBoost	10 k	96.51	95.62	0.9606	4.38
	50 k	96.81	95.92	0.9636	4.41
	100 k	96.76	95.85	0.9630	4.58

从表5 可知, 所设方法在不同规模测试集下的干扰频点检测的精确率最优达到 96.81%, 且平均推理时间在 4.6 ms 以内。当干扰与非干扰的比例增大时, 精确率、召回率无明显变化, 这表明模型没有过拟合, 具有较强的鲁棒性。

对比分析:

与传统干扰频点检测方法相比, 基于监督学习的干扰频点检测, 不需要人为设定阈值及构造特征, 仅需要构建合适的分类模型即可自动学习数据中隐藏的规律, 快速高效检测出干扰频点。

4. 结束语

本文提出了基于集成学习的干扰频点检测方法, 通过引入机器学习分类算法, 采用基于并联 XGBoost 模型的策略实现跳频通信干扰频点检测。实验结果表明, 基于 XGBoost 模型的单个频点干扰分类的平均准确率达 96.88%, 且实时性强; 此外, 基于并联 XGBoost 模型的干扰频点检测方法的准确率最优达到了 96.81%, 满足工程应用的条件。

本文所提方法仅应用了时域特征, 对数据的研究有待深入, 在今后的工作中需要对数据的多域特征及模型优化进一步研究。

参考文献

- [1] Yucek, T. and Arslan, H. (2009) A Survey of Spectrum Sensing Algorithms for Cognitive Radio Applications. *IEEE*

- Communications Surveys & Tutorials*, **11**, 116-130. <https://doi.org/10.1109/surv.2009.090109>
- [2] 范广伟, 邓江娜, 王振华, 等. 基于能量检测的 GNSS 干扰检测技术研究[J]. 无线电工程, 2013, 43(3): 33-35.
- [3] 吕再兴. 通信对抗中的干扰检测算法研究[D]: [硕士学位论文]. 成都: 电子科技大学, 2011.
- [4] Henttu, P. and Aromaa, S. (2002) Consecutive Mean Excision Algorithm. *IEEE 7th International Symposium on Spread Spectrum Techniques and Applications*, **2**, 450-454. <https://doi.org/10.1109/isssta.2002.1048582>
- [5] Wang, H., Yang, E., Zhao, Z. and Zhang, W. (2009) Spectrum Sensing in Cognitive Radio Using Goodness of Fit Testing. *IEEE Transactions on Wireless Communications*, **8**, 5427-5430. <https://doi.org/10.1109/twc.2009.081586>
- [6] Chen, J., Gibson, A. and Zafar, J. (2008) Cyclo-Stationary Spectrum Detection in Cognitive Radios. In: *IET Seminar Digest. IET Seminar on Cognitive Radio and Software Defined Radios: Technologies and Techniques*, IET Publication, 1-5.
- [7] 王正欢, 杨亚宁, 帅云开. 面向认知无线电的干扰频点鲁棒检测方法[J]. 遥测遥控, 2018, 39(6): 40-45.
- [8] 李孟超, 潘申富. 一种基于 FCME 的跳频通信干扰检测门限设计方法[J]. 河北工业科技, 2022, 39(2): 94-100.
- [9] 申鹏. 跳频通信系统中的干扰识别技术研究[D]: [硕士学位论文]. 西安: 西安电子科技大学, 2018.
- [10] 陶柱. 跳频通信干扰识别技术的研究[D]: [硕士学位论文]. 成都: 电子科技大学, 2021.
- [11] 王安强. 跳频通信智能抗干扰技术研究[D]: [硕士学位论文]. 杭州: 电子科技大学, 2022.
- [12] 张子恒. 智能抗干扰通信关键技术研究[D]: [硕士学位论文]. 成都: 电子科技大学, 2022.
- [13] 梁应敞, 谭俊杰, Dusit, 等. 智能无线通信技术研究概况[J]. 通信学报, 2020, 41(7): 1-17.
- [14] 李奇, 徐慨, 杨海亮. 干扰信号检测技术研究[J]. 信息通信, 2018(6): 28-31.
- [15] Chen, T. and Guestrin, C. (2016) XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 13-17 August 2016, 785-794. <https://doi.org/10.1145/2939672.2939785>
- [16] 黑新宏, 高苗, 张宽, 等. 基于 Nadam-TimeGAN 和 XGBoost 的时序信号故障诊断方法[J]. 通信学报, 2024, 45(4): 185-200.
- [17] Friedman, J.H. (2001) Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, **29**, 1189-1232. <https://doi.org/10.1214/aos/1013203451>
- [18] 赵然磊, 杨留栓, 徐晓, 等. 基于 XGBoost 算法的火山岩岩性识别方法与研究[J]. 地球物理学进展, 2025: 1-12.
- [19] 索基源, 李元奎, 崔金龙, 等. 基于 XGBoost 算法的船舶油耗预测模型[J]. 中国航海, 2024, 47(2): 153-159.