

doi:10.13756/j.gtxyj.2024.230082.

无线通信技术

梁智儒,边东明,张更新. 麻雀搜索算法优化极端梯度提升的扩频码研究[J]. 光通信研究, 2024(6):230082.

Liang Z R, Bian D M, Zhang G X. Research on Spread Spectrum Codes Optimized by Sparrow Search Algorithm and Extreme Gradient Boosting [J]. Study on Optical Communications, 2024(6):230082.

麻雀搜索算法优化极端梯度提升的扩频码研究

梁智儒,边东明,张更新

(南京邮电大学 通信与信息工程学院,南京 210003)

摘要:【目的】直接序列扩频系统(DSSS)在军事和民用通信中都得到了广泛应用,因其具有对各种常见干扰较强的抵抗能力和安全性较高等优点,且易于实现,被广泛应用于码分多址(CDMA)之中。然而,在非协作通信的应用场景中,检测 DSSS 信号,对 DSSS 信号参数进行估计,甚至截获信息,都是文章需要考虑的问题。在 DSSS 中,正确识别所使用的扩频序列是正确解扩的重要前提条件。针对低信噪比 DSSS 信号扩频码识别成功率低的问题,文章通过结合 m 序列的三阶相关函数(TCF)峰值特性,在降噪预处理的前提下,通过功率谱二次处理识别 DSSS 信号的伪码周期作为先验信息,将扩频码的识别问题具体成为一个峰值检测分类的问题,进而对峰值识别分类进行了研究。【方法】文章提出了使用麻雀搜索算法(SSA)优化极端梯度提升(XGBOOST)的 DSSS 信号三阶相关峰值分类方法,提高了对 m 序列分类识别的准确率。【结果】通过在不同信噪比下对比常规峰值检测和决策树分类方法以及对比不同序列周期的识别分类准确率,仿真结果表明,经过预处理的 SSA XGBOOST 扩频码识别分类方法比起常规机器学习与峰值检测方法分类识别成功率更高,在高序列周期下性能逐步提升。【结论】文章所提方法能在较低的信噪比条件下更准确地识别分类扩频码 m 序列。

关键词: m 序列;三阶相关函数;麻雀搜索算法;优化极端梯度提升

中图分类号:TN914

文献标志码:A

Research on Spread Spectrum Codes Optimized by Sparrow Search Algorithm and Extreme Gradient Boosting

LIANG Zhiru, BIAN Dongming, ZHANG Gengxin

(College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract:【Objective】Direct Sequence Spread Spectrum (DSSS) has been widely used in military and civilian communications due to its strong resistance to various common interferences, high security, and ease of implementation. It has been widely used in Code Division Multiple Access (CDMA) system. However, in non-cooperative communication scenarios, detecting DSSS signals, estimating DSSS signal parameters, and even intercepting information are all issues that need to be considered. In DSSS, correctly identifying the spread spectrum sequence used is an important prerequisite for correcting despreading. To address the problem of low success rate of spread code identification for low signal-to-noise ratio DSSS signals, this paper combines the Third-order Correlation Function (TCF) of m -sequences and its peak characteristics to identify the pseudo-code period of DSSS signals as prior information through power spectrum secondary processing on the premise of denoising preprocessing. The problem of spread code identification is transferred into a peak detection classification problem. The peak identification and classification is then studied. 【Methods】This paper proposes a method of using Sparrow Search Algorithm (SSA) to optimize Extreme Gradient Boosting (XGBOOST) for third-order correlation peak classification of direct spread signals to improve the accuracy of m -sequence classification and identification. 【Results】By comparing conventional peak detection and decision tree classification methods at different signal-to-noise ratios and comparing the classification accuracy of different sequence periods, the simulation results show that the spread code identification and classification method optimized by SSA with XGBOOST after preprocessing has a higher classification and identification success rate than conventional machine learning and peak detection methods. Its performance gradually improves at high sequence periods. 【Conclusion】This method can more accurately identify and classify m -sequence spread codes under low signal-to-noise ratio conditions.

Key words: m -sequence; TCF; SSA; XGBOOST

0 引 言

利用三阶相关函数(Third-order Correlation Function, TCF)实现对 m 序列的估计,这种方法提

出的理由是因为三阶相关法比常用的二阶统计量含有更多的信息,该方法的出现受到了国内外学者的广泛关注。 m 序列 TCF 的应用,还停留在峰值检测的角度,在寻找合适阈值的前提下可用于进行是否

收稿日期:2023-08-01; 修回日期:2023-09-27; 纸质出版日期:2024-12-10

基金项目:国家自然科学基金资助项目(61801445)

作者简介:梁智儒(1998—),男,安徽亳州人。硕士,主要研究方向为卫星物联网。

通信作者:梁智儒,硕士。E-mail:574799923@qq.com

© Editorial Office of Study on Optical Communications. This is an open access article under the CC BY-NC-ND license.

是峰值点的判决^[1-4]。将伪随机(Pseudo Noise, PN)码序列估计问题抽象为一个信号提取问题,并通过神经网络进行盲估计^[5],上述方法的应用说明了 TCF 峰值特性作为分类的可操作性,但是该常规方法进行分类时受环境参数的影响较大。

而近年来,机器学习方法在通信领域的应用逐渐增多,机器学习可以自动从大量数据中学习特征从而提高分类的准确性和效率^[6]。同周期不同反馈系数下,直接序列扩频(Direct Sequence Spread Spectrum, DSSS)信号三阶峰值识别本质上来说就是识别特征问题,而极端梯度提升(Extreme Gradient Boosting, XGBOOST)算法^[7-8]是一种基于提升树的学习系统,其在分类问题上通过多轮的树迭代来不断减小误差,使得该算法能最终构建成一个强大的分类器。常规 XGBOOST 模型训练找到最优参数组合需要对参数进行遍历,会耗费大量的计算资源和时间,而麻雀搜索算法(Sparrow Search Algorithm, SSA)作为群智能优化算法的一种,其优势在于优秀的收敛性能,可以减少调优过程中的迭代次数,从而达到节省资源与时间的目的^[9]。

综上,本文提出了基于 SSA-XGBOOST 的扩频码研究。本文对 XGBOOST 模型中重要参数采用 SSA 进行最优化处理,将 SSA 应用到 XGBOOST 算法的参数选择中,以 XGBOOST 算法训练集分类准确率最高为优化目标,选择最优参数,建立 SSA-XGBOOST 模型,解决了低信噪比下 DSSS 信号采用 m 序列扩频码识别成功率低的问题。

1 信号模型

图 1 所示为信号处理流程图,本课题所采用的接收信号时域表达式为

$$x(t) = s(t) + n(t) = A \sum_{i=0}^{\infty} d_i c(t) g_d(t - iT_d) + n(t), \quad (1)$$

式中: $x(t)$ 为中频 DSSS 信号,由 $s(t)$ 和 $n(t)$ 组成; $s(t)$ 为 DSSS 信号; $n(t)$ 为零均值高斯白噪声; A 为幅度; d_i 为信息码序列,取值范围为 $\{-1, +1\}$; $g_d(t)$ 为门函数; R_d 为其传输码元速率, T_d 为其码元宽度, $T_d = 1/R_d$; $c(t)$ 为 PN 序列。设 T_p 为 PN 序列码片宽度,速率为 R_p 。由于研究信号为 DSSS 信号,即一位信息码采用完整周期的 PN 序列进行扩频,即 $T_d = NT_p$, N 为 PN 序列周期, $N \gg 1$, 所以扩频后信息流序列的速率还是 PN 码速率 R_p , 信息码与 PN 码的乘积才如式(1)所示。

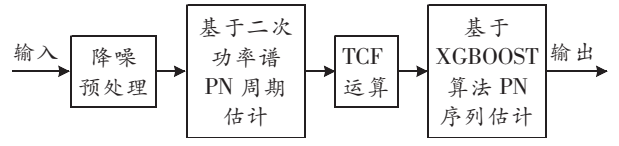


图 1 信号处理流程图

Figure 1 The flowchart of signal processing

2 信号预处理过程

在图 1 所示的流程图中,降噪预处理是为了降低噪声 $n(t)$ 带来的影响,采用常见的小波变换进行降噪。对于一个有限能量信号,基于小波理论^[10],选取合适的小波基进行分解,通过不同的基分解,将信号转换为

$$s(t) = \sum_k c_{(j_0)}(k) 2^{(j_0/2)} \varphi(2^{(j_0/2)} t - k) + \sum_k \sum_j d_j(k) 2^{(j/2)} \psi(2^{j/2} t - k), \quad (2)$$

式中: t 为时间序列; j 和 k 分别为尺度和平移的离散化索引; j_0 为当前尺度选择; φ 函数为尺度函数; ψ 函数为小波函数; $c_{(j_0)}$ 和 d_j 函数均为对应尺度和平移索引的内积表示。式中,第 1 个求和为原信号的主干(近似值),第 2 个求和为细节值。通过逐级分解,依次降低细节值的比重,保留近似值的特征,从而实现对一个能量有限信号的降噪。在实际过程中,逐级分解用低通滤波器实现。对降噪后信号进行估计。

通过对 DSSS 信号 $s(t)$ 的傅氏变换去模平方作为功率谱 $\hat{S}(e)$:

$$\hat{S}(e) = |FT\{|FT[s(t)]|^2\}|^2 \cong |T_c \sum_k (1 - |e - kT_d|/T_c)|^2, \quad (3)$$

式中: FT 为傅氏变换; T_c 为伪码序列周期; e 为常数。

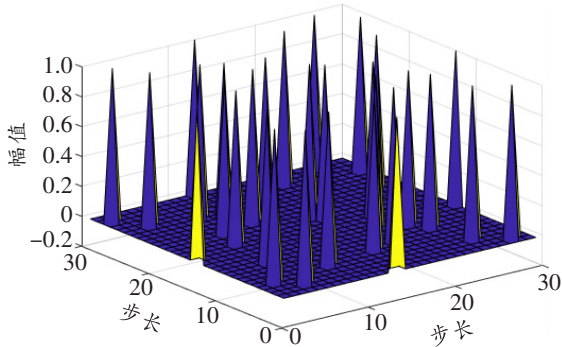
通过进行二次谱的操作,波形会呈现出规律性的脉冲,在这些脉冲处汇聚着信号能量,这些脉冲的间距对应的正是所用扩频 PN 序列的周期长度或者是周期长度的整数倍^[11]。

m 序列的 TCF 是通过 m 序列的平移相加性以及 m 序列自相关特性好的特点推导出 m 序列的 TCF 有峰值特性^[5]。 n 阶 m 序列 TCF 峰值点数为 $2^n - 2$ 个, m 序列的 TCF $R_m(\tau_1, \tau_2)$ 计算公式为

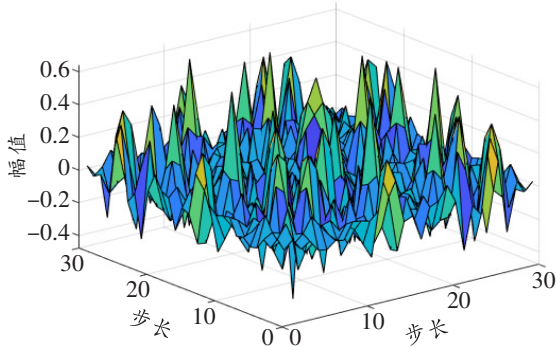
$$R_m(\tau_1, \tau_2) = R(p, q) = \frac{1}{M} \sum_{i=1}^M v(i) v_p(i) v_q(i), \quad (4)$$

式中: $v(i)$ 为 $s(t)$ 采样; $v_p(i)$ 和 $v_q(i)$ 分别为 $v(i)$ 平移 p 和 q 单位后的序列; τ_1 与 τ_2 为平移步长; M 为峰值点数和。

m 序列 TCF 峰值图与经过噪声峰值图如图 2 所示,平面均表示平移步长,纵轴为赋值。



(a) 5 阶 m 序列无噪声 TCF 峰值
(a) Peak of 5th order m -sequence TCF without noise



(b) 5 阶 m 序列有噪声 TCF 峰值
(b) Peak of 5th order m -sequence TCF with noise

图 2 m 序列 TCF 峰值图

Figure 2 Peak graph m -sequence TCF

3 SSA 优化 XGBOOST 的流程

SSA 是从麻雀觅食行为获取思路,将麻雀分为两种类型:寻找实物的麻雀 A 类和跟随 A 类麻雀进行觅食的麻雀 B 类。本身特性是高适应性的麻雀被划分为 A 类,被赋予了寻找丰富食物所在范围的职责,并肩负着确定搜索方向的任务。而 B 类麻雀是 A 类麻雀的跟随者,跟随发现者获取食物。为了提高自身捕食率,给 B 类麻雀赋予一部分侦查机制,即当危险发生时立刻执行反捕食的行为逻辑^[9]。而 XGBOOST 是一种基于提升树的学习系统,其对分类识别特征具有良好的作用。

Chen 等^[7]在 XGBOOST 的提出中,将通用参数、提升参数和学习任务参数称为 XGBOOST 算法的 3 大类可调参数。上述 3 类参数分别对应着设置整体、选择每一步树与指导优化任务执行的功能。在文献[12]中详细阐述了在基于 XGBOOST 的模型搭建过程中对算法性能影响最大的是提升参数,将学习率和树的最大深度作为主要优化目标,SSA 本身参数最大迭代次数作为次要优化目标。因此在

采用 SSA 对 XGBOOST 算法参数进行优化时需将搜索的空间维度设置为 $l=2$ 。

图 3 所示为 SSA 优化 XGBOOST 算法的流程图,优化步骤如下:

步骤 1:数据集的产生与预处理。首先确定识别分类模型的输入和输出。将预处理信号作为输入,建立样本集。

步骤 2:对数据集进行划分。训练集用于训练基于 TCF 数据的 SSA-XGBOOST 学习模型,构建内部的联系。验证集是为了验证模型的准确性。

步骤 3:使用 SSA 进行参数优化。依次更新麻雀发现者、加入者与察觉到危险的侦查者的位置。用适应度作为评判依据,通过适应度比较,确定各最佳位置,并将全局最优的数据输出构建 SSA-XGBOOST 模型。

步骤 4:采用 SSA-XGBOOST 模型进行验证集验证,获取同周期不同反馈系数下 m 序列的识别分类准确率。

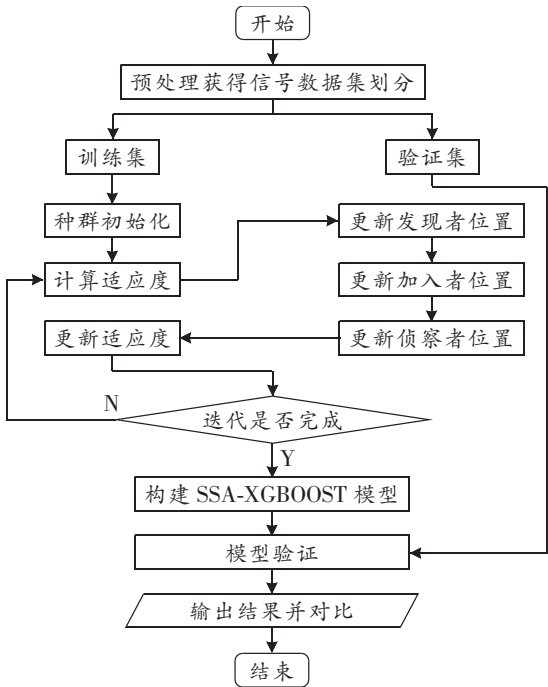


图 3 SSA 优化 XGBOOST 流程图

Figure 3 Flowchart of SSA optimization of XGBOOST

4 仿真结果与分析

4.1 实验样本选取

为了较为直观地展现周期谱线之间的间距,以及为了在数据集较少的情况下更能凸显出性能的差距,仿真采用的是采用 5 阶 m 序列扩频的短码 DSSS 信号,信号采用二进制相移键控 (Binary

Phase Shift Keying, BPSK) 调制, 信息码长度取 100, 反馈系数依次取 45、67 和 75 生成输入, 信噪比采用 -15 dB。

根据式(2)的小波降噪理论, 采用 Mallat 算法, 基于高低频滤波器对输入信号进行五阶小波分解重构, 小波采用 db2 小波。在经过降噪预处理后, 基于式(3)进行基于二次功率谱的 PN 周期识别。采用延迟法以消除信息码影响, 将信号 $s(t)$ 逐步延迟码片长度后与自身相乘, 直至得到较为规律的峰值间隔, 选取最小的码片长度作为序列周期。在得到 PN 序列周期后, 忽略失步点, 将截取 PN 序列周期看作无失步的完整周期, 并通过式(4)求得 TCF 数据, 为了让使用 XGBOOST 训练模型时得到的模型识别性能更加准确, 在仿真中采用对每个反馈系数对应扩频序列分别生成 100 组数据的方法生成数据集。将整个数据集合分为训练集和验证集, 进行分割的比例为 8 : 2。

4.2 SSA 参数寻优比较与结果

在 -15 dB 信噪比下, 仿真使用 SSA 对参数进行全局最优搜索, 设置最大迭代次数上下限区间为 [0, 30], 树最大深度设置区间为 [1, 12], 学习率设置区间为 [0.01, 1.00]。对比实验分别采用常见遗传算法与粒子群算法优化 XGBOOST 训练模型, 设置最大迭代次数均为 500 次。适应度迭代曲线如图 4 所示。

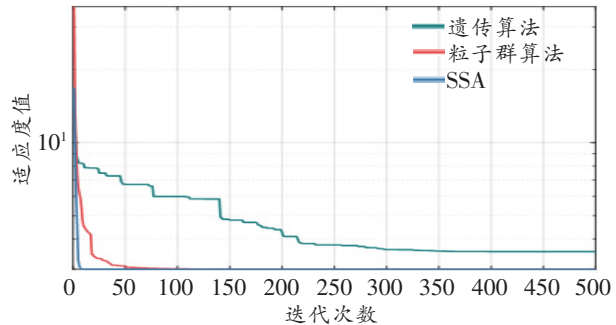


图 4 适应度变化曲线图

Figure 4 The value of fitness versus iteration number

由图可知, 用 SSA 优化 XGBOOST, 只需要两次迭代就达到最佳位置, 同时遗传算法的准确率对比 SSA 有明显的差距。SSA 优化 XGBOOST 可以比目前常见的遗传算法和粒子群算法更准确快速地找到最优参数。

通过 SSA 对学习率和树的最大深度参数进行全局最优搜索得到构建 XGBOOST 的最优参数, 具体如表 1 所示。用表 1 中最优参数对 XGBOOST

进行训练, 并采用训练好的 XGBOOST 对测试样本进行测试, 在 -15 dB 信噪比下, 测试结果如图 5 所示, 测试集的预测结果对比准确率达到 66.777 8%。

表 1 模型最优参数

Table 1 Optimal parameter of the model	
参数名	最佳参数值
学习率	0.943 4
树的最大深度	5
提升器	gbtree
列采样率	1
随机采样比例	0.62
Gamma 函数	0

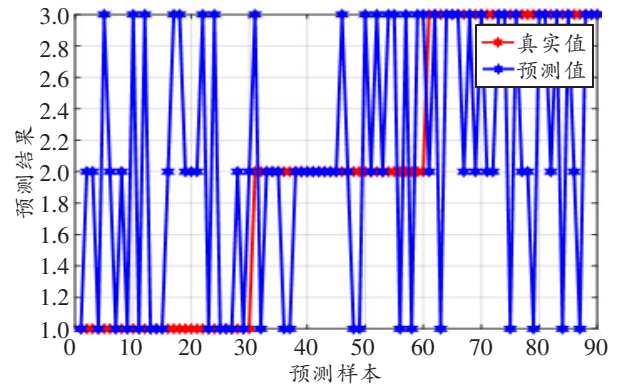


图 5 预测值与真实值对比图

Figure 5 Comparison graph of predicted values and actual values

数据说明经过 SSA 优化的 XGBOOST 模型在较低信噪比下依然有较高识别分类 m 序列的准确率, 因此本文所提采用 SSA 优化的 XGBOOST 模型在低信噪比下对 m 序列分类识别的方法是可行的。

4.3 不同模型识别效果对比

为了显示 SSA-XGBOOST 模型性能的优劣, 采用对比实验的方法在相同的步骤下分别用训练集通过常规的峰值检测法、决策树的机器学习分类算法和 XGBOOST 算法训练模型, 同时为了展现降噪预处理的操作是否有必要, 再分别通过有无降噪预处理的训练集导入上述 3 种算法, 进行对比验证。信噪比取 -20 ~ 0 dB, 对每个信噪比做 100 次蒙特卡洛仿真。而且为了验证预处理是否产生作用, 在对比实验中加入了有无预处理的对比, 实验结果如图 6 所示。

由图可知, 常规的 TCF 峰值识别方法因其需确定峰值检测阈值等因素, 识别成功率一直维持在较低水平。对比峰值检测法, 虽然决策树算法得出训练模型在不同的信噪比下性能更高, 但基于 XG-

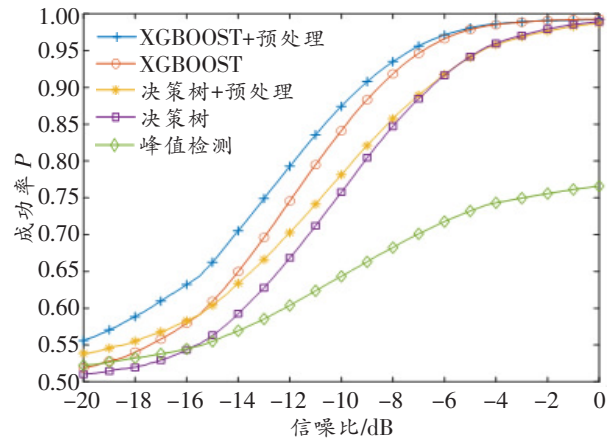


图 6 峰值检测、决策树和 XGBOOST 的识别结果对比图

Figure 6 Comparison graph of recognition results of peak detection, decision tree and XGBOOST

BOOST 算法的 PN 序列识别的识别准确率在不同信噪比下均高于其他方法。而且通过降噪预处理后的识别准确率均高于未经过降噪预处理的识别成功率,说明了降噪预处理对 PN 序列的识别成功率有一定的作用。

4.4 不同周期识别效果对比

为了进一步说明基于 XGBOOST 的 PN 序列识别算法的性能,选择不同周期、不同反馈系数的 PN 序列扩频序列进行对比实验,仿真采用的是不同级数 m 序列下常用的反馈系数,具体参数如表 2 所示,仿真步骤重复上述针对不同方法识别性能对比实验中 XGBOOST 算法的部分,依次更换 m 序列级数周期与反馈系数,仿真在信噪比为 -15 dB 的干扰条件下进行。性能仿真结果如图 7 所示。

表 2 m 序列级数的常用反馈级数

Table 2 Common feedback series table of m -sequence series

级数 n	周期 p	反馈系数
5	31	45,67,75
6	63	103,147,155
7	127	203,211,217,235
8	255	435,453,537,543,545
9	511	1 021,1 055,1 131,1 157,1 167,1 175

由图 7 可知,在级数 6~9 的范围内,随着级数 n 以及反馈级数种类的增加,基于 XGBOOST 算法的 PN 序列识别的准确率逐渐增长。级数 5~6 识别准确率的下降是源于两组对比实验中 TCF 峰值点数的增长、数据量的增加,但是反馈系数的种类数量一致,这导致了识别性能的下降。

图 7 说明本文所提 m 序列分类识别方法在低

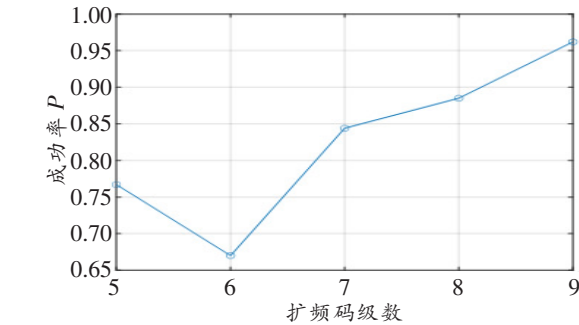


图 7 不同 m 序列阶数下识别结果对比图

Figure 7 Comparison graph of recognition results under different m -sequence orders

周期 PN 序列下识别分类效果不佳,但是随着周期的增长,识别性能呈现明显上涨趋势,更贴合现实情况下码分多址、宽带码分多址情况下常见的 8 阶扩频码周期,因此本文所提采用 SSA 优化的 XGBOOST 模型在低信噪比下对 m 序列分类识别的方法具有现实意义。

5 结束语

本文的理论分析和仿真实验表明,基于 SSA 优化的 XGBOOST 模型的 m 序列识别算法可以有效地提高其在低信噪比下的识别精度,且识别精度与 m 序列的周期成正比,对比常规的峰值检测法有着较高的识别效率,且对比常见的优化算法,SSA 寻优能力有着不错的提升。降噪预处理也对各种算法的精度起到了显著的作用。

但是预处理中通过延迟二次谱进行周期识别只在短码扩频中有着较好的应用。而且随着 m 序列周期的增加, m 序列的 TCF 峰值数据量呈指数型上升,特别是在阶数较高的情况下,因为常用的反馈系数较多,训练集数据较多,导致模型训练时间增长,计算量大为增加,该算法为追求识别准确率或针对已知扩频序列周期的场合进行扩频码周期识别提供了一种新的研究方向。

参考文献:

[1] Warner E S, Mulgrew B, Grant P M. Triple Correlation Analysis of m -sequences[J]. Electronics Letters, 1993,29(20):1755.

[2] Batty K E, Adams E R. Detection and Blind Identification of m -sequence Codes Using Higher Order Statistics[C]//Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics, SPW-HOS'99, Caesarea, Israel;IEEE,1999;778683.

[3] 白彥, 杨晓静, 张玉. 基于高阶统计处理技术的 m -序列帧同步码识别[J]. 电子与信息学报, 2012, 34(1): 33–37.
Bai Y, Yang X J, Zhang Y. A Recognition Method of m -sequence Synchronization Codes Using Higher-order Statistical Processing[J]. Journal of Electronics & Information Technology, 2012, 34(1): 33–37.

[4] 俎云霄. 基于高阶统计处理技术的 m -序列检测及识别[J]. 电子与信息学报, 2007, 29(7): 1576–1579.
Zu Y X. The Detection and Recognition of m -sequence Using Higher-order Statistical Processing[J]. Journal of Electronics & Information Technology, 2007, 29(7): 1576–1579.

[5] 张天骐, 林孝康, 周正中. 基于神经网络的低信噪比直扩信号扩频码的盲估计方法[J]. 电路与系统学报, 2007, 12(2): 118–123.
Zhang T Q, Lin X K, Zhou Z Z. A Neural Network Approach to Blind-estimation of PN Spreading Sequence in Lower SNR DS/SS Signals[J]. Journal of Circuits and Systems, 2007, 12(2): 118–123.

[6] 郑远攀, 李广阳, 李晔. 深度学习在图像识别中的应用研究综述[J]. 计算机工程与应用, 2019, 55(12): 20–36.
Zheng Y P, Li G Y, Li Y. Survey of Application of Deep Learning in Image Recognition[J]. Computer Engineering and Applications, 2019, 55(12): 20–36.

[7] Chen T, Guestrin C. XGBoost: a Scalable Tree Boosting System[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, CA, USA: ACM, 2016: 785–794.

[8] Friedman J, Hastie T, Tibshirani R. Special Invited Paper. Additive Logistic Regression: a Statistical View of Boosting[J]. The Annals of Statistics, 2000, 28(2): 337–374.

[9] Xue J, Shen B. A Novel Swarm Intelligence Optimization Approach: Sparrow Search Algorithm[J]. Systems Science & Control Engineering, 2020, 8(1): 22–34.

[10] 韩震宇, 申旭娟, 石章林. 信号的多分辨率分析及其在消噪中的应用[J]. 四川联合大学学报(工程科学版), 1999(1): 52–58.
Han Z Y, Shen X J, Shi Z L. The Multiresolution Analysis of Signal and Its Application in Noise Canceling[J]. Journal of Sichuan United University (Advanced Engineering Sciences), 1999(1): 52–58.

[11] 张天骐, 周正中. 直扩信号伪码周期的谱检测[J]. 电波科学学报, 2001, 16(4): 518–521.
Zhang T Q, Zhou Z Z. A New Spectral Method of Periodic Detection of PC Sequence in Lower SNR DS/SS Signals[J]. Chinese Journal of Radio Science, 2001, 16(4): 518–521.

[12] Aarshay J. Complete Guide to Parameter Tuning in XGBoost[DB/OL]. (2019-05-16) [2023-08-01]. <https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-XGBoost-with-codes-python>.