

引用格式:蒋能,张红敏,李一鸣.基于分步变门限孤立森林的MFR波形单元无监督提取[J].信息对抗技术,2022,1(3):76-85. [JIANG Neng, ZHANG Hongmin, LI Yiming. Unsupervised extraction of MFR waveform units based on step-by-step variable threshold isolated forests[J]. Information Countermeasure Technology, 2022, 1(3):76-85. (in Chinese)]

基于分步变门限孤立森林的MFR波形单元无监督提取

蒋能,张红敏*,李一鸣

(战略支援部队信息工程大学数据与目标工程学院,河南郑州 450001)

摘要 波形单元是多功能雷达(multi-function radar, MFR)发射信号的基本构成,其提取的准确性直接影响后续MFR行为意图的分析。针对非协作MFR截获信号在无先验信息情况下波形单元提取时效性不佳的问题,提出了一种基于分步变门限孤立森林的波形单元提取方法。对接收的MFR脉冲信号进行漏脉冲和伪脉冲的检验,在获得脉冲参数一阶差分数据后,采用改进的分步变门限孤立森林算法进行波形单元起始脉冲搜索,实现无监督波形单元提取。理论分析和实验表明,该方法在保证波形单元提取鲁棒性的同时,显著降低了计算复杂度。

关键词 多功能雷达;波形单元;孤立森林;无监督学习

中图分类号 TN 95 **文献标志码** A **文章编号** 2097-163X(2022)03-0076-10

DOI 10.12399/j.issn.2097-163x.2022.03.007

Unsupervised extraction of MFR waveform units based on step-by-step variable threshold isolated forests

JIANG Neng, ZHANG Hongmin*, LI Yiming

(School of Data and Target Engineering, PLA Strategic Support Force Information Engineering University, Zhengzhou 450001, China)

Abstract The transmitted signal of Multi-Function Radar(MFR) consists of waveform unit, and the extraction accuracy of waveform unit directly affects the analysis of the subsequent behavior intent of the MFR. Aiming at the problem that the timing of waveform unit extraction is not efficient in the absence of prior information, this paper proposes a waveform unit extraction method based on step-by-step variable threshold isolated forest. Firstly, the received MFR pulse signal is connected to the leakage pulse and pseudo pulse test, and after obtaining the first-order differential data of the pulse parameters, the improved step-by-step variable threshold isolation forest algorithm is used to search for the initial pulse of the waveform unit to achieve unsupervised waveform unit extraction. Theoretical analysis and experiments show that the proposed method significantly reduces the computational complexity and ensures the robustness of waveform unit extraction.

收稿日期:2022-07-09

修回日期:2022-08-17

通信作者:张红敏, E-mail: zhmin1206@163.com

作者简介:蒋能(1998—),男,硕士研究生,研究方向为雷达数据处理;张红敏(1984—),女,博士,副教授,研究方向为雷达信号与信息处理;李一鸣(1995—),女,讲师,研究方向为雷达信号处理

基金项目:军队装备综合研究资助项目(20200113-4)

Keywords multi-function radar; waveform units; isolation forest; unsupervised learning

0 引言

多功能雷达(multi-function radar, MFR)是一种先进体制雷达,因其同时具备搜索、跟踪与制导等多种能力而被广泛部署和应用^[1]。随着MFR的大规模使用,其工作模式识别已经成为现代电子侦察系统的重要组成部分,其水平高低决定了电子情报侦察、电子对抗等重要军事活动的效果^[2]。

为了有效描述MFR信号的动态规律,不少研究人员力图简洁高效地建立MFR信号模型。文献[3-4]将MFR等效为有限状态机,使用隐马尔可夫模型(hidden Markov model, HMM)直接对脉冲列进行建模。文献[5-9]构建并完善了句法模型(syntactic model),将MFR信号等效为符合一定语法规则的语言,从雷达字(radar word)、雷达短语(radar phrase)和雷达句子(radar sentence)3个层次对MFR信号进行描述。雷达字概念对应MFR信号的波形单元,是句法模型的最底层,其提取的准确性和完整性直接影响后续对雷达短语和雷达句子的分析工作,对波形单元提取工作的研究具有十分重要的意义。

根据对先验知识需求情况的不同,当前MFR波形单元提取方法可分为有监督框架下^[10-13]和无监督框架下^[14-17]2类。有监督学习框架下的波形单元提取方法,需要利用先验知识构建“雷达字”模板库。考虑到MFR的非协作性,这一要求基本难以满足。相比较而言,无监督学习框架下的提取方法能在缺乏先验信息的条件下提取波形单元,更加贴合实际应用场景。文献[14]研究了雷达波束指向切换造成的脉冲幅度突变,提出了脉冲幅度变化点检测的波形单元提取方法;文献[15]研究了MFR信号脉间参数的变化规律,并使用参数自适应的密度聚类算法(density adaptive density-based spatial clustering of applications with noise, DA-DBSCAN)完成波形单元的提取工作,对各种波形单元都具有较强的适应性;文献[16]直接对多功能雷达脉冲序列进行复杂度最小化的语义编码,可提取和分类脉冲重复间隔(pulse repetition interval, PRI)跳变的雷达字;文献[17]提出一种利用聚类粗分类的样本进行循环神经网络(recurrent neural network,

RNN)模型训练,以提取不同雷达字的可区分度特征。文献[14-17]依据雷达信号自身特点完成波形单元提取任务,无需先验知识,但是普遍存在计算复杂度高的问题。

为解决非协作条件下波形单元提取计算复杂度高的问题,本文提出了一种基于改进的孤立森林的波形单元提取方法。通过分析现有文献中MFR信号的特点发现,当波形单元切换时,脉冲参数会出现明显变化。依据这一特点,选取测量误差相对较小的射频(radio frequency, RF)、脉宽(pulse wide, PW)和PRI这3种脉冲特征对MFR信号进行描述,并取得其一阶差分数据构建波形单元提取模型,将波形单元提取问题转化为波形单元起始脉冲提取的二分类问题;此外,该方法还专门针对MFR信号的特点对漏脉冲和伪脉冲问题进行了处理。

1 MFR波形单元提取基础

1.1 MFR信号模型

Visnevski等^[5]构建了句法模型应用于MFR信号识别,从雷达字、雷达短语、雷达句子3个层级描述MFR信号及其组合形式,他们的关系如图1所示。

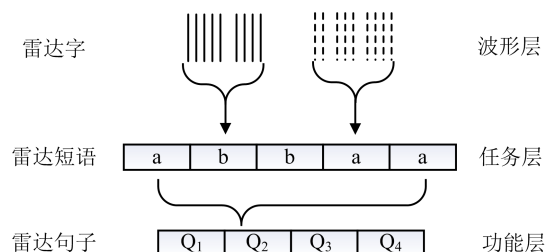


图1 MFR句法模型

Fig. 1 MFR syntax model

雷达字是具有一定规律的、按有限数目顺序排列的脉冲,是MFR完成单次检测的最小整体结构,与雷达发射的波形单元一一对应。图1中的a和b为2种不同的雷达字;雷达短语是一定数目的雷达字的顺序排列,图1中 Q_2 由“abbaa”5个雷达字组成,1个雷达短语一般对应1个雷达任务;雷达句子则是由一定数目的雷达短语串联组成。与波形单元相对应的雷达字作为MFR句法模型的最底层,其提取的准确性和完整性会直接

影响到后续对雷达短语和雷达句子的分析工作。

当需要在目标区域执行特定任务时, MFR 会在自身波形库中选择对应的波形朝目标区域发射。不同波形单元对应的波形在 R_{RF} (RF)、 W_{PW} (PW) 和 P_{PRI} (PRI) 等参数上会有明显的变化^[14-15]。图 2 为某 MFR 脉冲列时序图, 将波形

单元起始脉冲标记为红点, 能够看出不同波形单元在 PRI、RF 和 PW 参数范围上存在部分重叠, 波形单元内部参数也存在一定程度的随机变化, 但波形单元进行切换时参数变化更为明显^[15]。因此, 可根据 RF、PW、PRI 参数间的联合变化规律来实现 MFR 波形单元的提取工作。

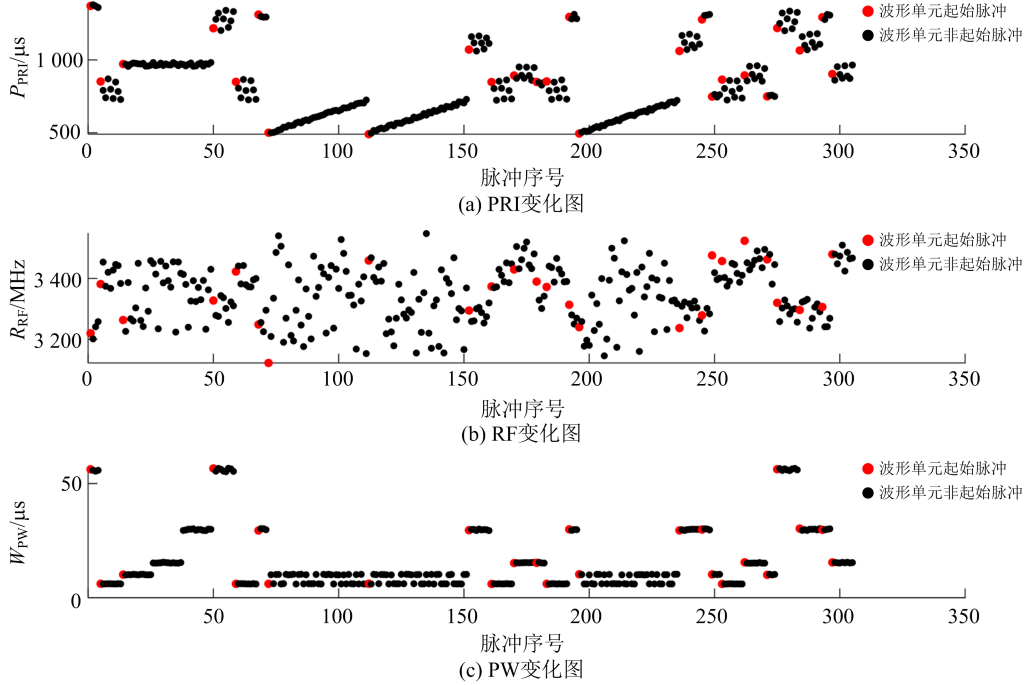


图 2 某 MFR 脉冲列时序图

Fig. 2 Pulse sequence diagram of a MFR

1.2 现有方法计算复杂度分析

现有无监督方法本质上都是通过聚类不同波形单元间的可区分特征完成提取任务的。用 m 表示测试序列的波形单元数量, w_i 表示第 i 个波形单元, 则测试序列的脉冲总数 N 为:

$$N = \sum_{i=1}^m l(w_i) \quad (1)$$

式中, $l(w_i)$ 表示 w_i 的脉冲个数。以经典密度聚类算法 (DBSCAN) 为例, 用于波形单元提取时的计算复杂度 T_{DB} 为:

$$T_{DB} = \underbrace{\sum_{i=1}^m 3l(w_i)}_{\langle \Delta P_{PRI}, \Delta R_{RF}, \Delta W_{PW} \rangle \text{ 处理}} + \underbrace{\frac{1}{2} \left[\sum_{i=1}^m 9l(w_i) \right]^2}_{\text{两两样本点距离计算}} + \underbrace{2 \left[\sum_{i=1}^m l(w_i) \right]^2}_{\text{核心点判断}} + \underbrace{\sum_{i=1}^m l(w_i)}_{\text{簇分类}} = \frac{85}{2} N^2 + 4N \quad (2)$$

DBSCAN 算法有 2 个超参数需要人为设置, 对聚类效果影响较大。为此有文献提出利用 DA-DBSCAN 自适应设置参数^[15], 以完成波形单

元的提取。然而, 该方法的计算复杂度显著增加, 难以满足雷达对抗的高实时性要求。其计算复杂度 (T_{DA}) 为:

$$T_{DA} = \underbrace{\sum_{i=1}^m 3l(w_i)}_{\langle \Delta P_{PRI}, \Delta R_{RF}, \Delta W_{PW} \rangle \text{ 处理}} + \underbrace{\frac{1}{2} \left[\sum_{i=1}^m 9l(w_i) \right]^2}_{\text{两两样本点距离计算}} + \underbrace{\left[\sum_{i=1}^m 3l(w_i) \right] \lg \left[\sum_{i=1}^m 3l(w_i) \right]}_{\text{Minpts-dist 排序}} + \underbrace{2 \left[\sum_{i=1}^m 3l(w_i) \right] \lg \left[\sum_{i=1}^m 3l(w_i) \right]}_{\Delta \text{ Minpts-dist 排序}} + \underbrace{\sum_{i=1}^N (3i)}_{\Delta \text{ Minpts-dist 方差计算}} + \underbrace{\sum_{i=1}^N \sum_{j=1}^i (2+N)}_{\text{熵极值搜索}} + \underbrace{2 \left[\sum_{i=1}^m l(w_i) \right]^2}_{\text{核心点判断}} + \underbrace{\sum_{i=1}^m l(w_i)}_{\text{簇分类}} = \frac{1}{2} N^3 + 45 N^2 + 6 N \cdot \lg^3 N + 7 N \quad (3)$$

由式 (3) 能够看出, DA-DBSCAN 算法运算量会随着脉冲数量 N 急剧上升。实际积累大约 4 s 的脉冲信号, 利用 DA-DBSCAN 进行波形单

元的提取,在 i9-11900K + 32 G DDR Memory + Windows 10 平台上的计算耗时将超过 183 s。因此,有必要提出一种低计算复杂度的波形单元提取方法。

2 分步变门限孤立森林 MFR 波形单元提取

在句法模型的基础上,本文提出一种基于分步变门限的孤立森林波形单元提取方法。该方法的基本思想是根据波形单元参数切换规律,将波形单元提取问题转化为波形单元起始脉冲和非起始脉冲的二分类问题,并利用孤立森林构建波形单元提取模型,分步变门限完成波形单元的提取工作。下面按照该方法流程进行详细说明。

2.1 漏脉冲伪脉冲检测

设侦察接收机依次接收到去交错后的某 MFR 的 n 个脉冲: $p = \{p_1, p_2, \dots, p_n\}$, 对于每个脉冲 $p_i (i=1, 2, \dots, n)$, 由 3 个特征参数对其进行描述, 即: $\{P_{\text{PRI}}, R_{\text{RF}}, W_{\text{PW}}\}$ 。其中,

$$\begin{cases} P_{\text{PRI}} = \{x_{11}, x_{12}, x_{13}, \dots, x_{1n}\} \\ R_{\text{RF}} = \{x_{21}, x_{22}, x_{23}, \dots, x_{2n}\} \\ W_{\text{PW}} = \{x_{31}, x_{32}, x_{33}, \dots, x_{3n}\} \end{cases} \quad (4)$$

侦收雷达信号时,由于脉冲去交错算法的局限、复杂电磁环境干扰等原因^[14],容易出现遗漏脉冲和伪脉冲。在本文选择的 3 种脉冲描述参数中, PRI 是根据相邻脉冲时间间隔得到的,当接收脉冲列中出现漏脉冲和伪脉冲时,会对 PRI 产生较大影响。因此,需要根据 MFR 的实际情况对这 2 种错误脉冲情况进行参数修正。

(1) 漏脉冲。对于 MFR 脉冲信号,其雷达字内部相邻脉冲参数的变化程度较小^[15]。因此,若某脉冲的 PRI 突然剧烈变化为相邻脉冲 PRI 的大约 2 倍时,可以考虑出现了漏脉冲;但为了避免将相邻雷达字起始脉冲的 PRI 变化误识别为漏脉冲,还需比较当前脉冲与上一时刻脉冲 RF、PW 参数的变化程度来综合判断。若 p_i 与 p_{i-1} 间可能存在漏脉冲,则满足:

$$\begin{aligned} x_{1i} &\in [2(1-\tau)x_{1(i-1)}, 2(1+\tau)x_{1(i-1)}], \\ x_{2i} &\in [(1-\tau)x_{2(i-1)}, (1+\tau)x_{2(i-1)}] \vee \\ x_{3i} &\in [(1-\tau)x_{3(i-1)}, (1+\tau)x_{3(i-1)}] \end{aligned} \quad (5)$$

式中, τ 为匹配门限调节参数,“ $\vee$ ”表示“或”,“ $\in$ ”表示“属于”。若满足式(5),则认为存在漏脉冲,直接在 p_i 与 p_{i-1} 中间时刻添加一个与 p_{i-1} 脉冲 RF、PW 相同的脉冲并更新脉冲位置信息。

(2) 伪脉冲。同样如果 PRI 出现剧烈变化,可以初步将相邻两段明显减小的 PRI 相加并与正常 PRI 对比。若 p_i 为潜在的伪脉冲,则满足:

$$\begin{aligned} x_{1(i-1)} + x_{1i} &\in [(1-\sigma)x_{1(i-2)}, (1+\sigma)x_{1(i-2)}], \\ x_{2i} &\notin [(1-\tau)x_{2(i-1)}, (1+\tau)x_{2(i-1)}] \wedge \\ x_{3i} &\notin [(1-\tau)x_{3(i-1)}, (1+\tau)x_{3(i-1)}] \end{aligned} \quad (6)$$

式中,“ $\wedge$ ”表示“且”,“ $\notin$ ”表示“不属于”。若满足式(6),则认为 p_i 为伪脉冲,直接将 p_i 从脉冲集中删除并更新脉冲位置信息。

2.2 特征参数一阶差分处理

由于相邻雷达字转换时脉冲参数会产生剧烈变化,故利用相邻脉冲参数的差值 $\{\Delta P_{\text{PRI}}, \Delta R_{\text{RF}}, \Delta W_{\text{PW}}\}$ 来衡量脉冲参数时序上的震荡程度。其具体形式为:

$$\begin{cases} \Delta P_{\text{PRI}} = \{x'_{11}, x'_{12}, x'_{13}, \dots, x'_{1n}\}, \\ x'_{11} = 0, x'_{1i} = x_{1i} - x_{1(i-1)}, i = \{2, 3, \dots, n\} \\ \Delta R_{\text{RF}} = \{x'_{21}, x'_{22}, x'_{23}, \dots, x'_{2n}\}, \\ x'_{21} = 0, x'_{2i} = x_{2i} - x_{2(i-1)}, i = \{2, 3, \dots, n\} \\ \Delta W_{\text{PW}} = \{x'_{31}, x'_{32}, x'_{33}, \dots, x'_{3n}\}, \\ x'_{31} = 0, x'_{3i} = x_{3i} - x_{3(i-1)}, i = \{2, 3, \dots, n\} \end{cases} \quad (7)$$

用 $L_{\text{label}} = \{l_1, l_2, l_3, \dots, l_n\}$ 对每个脉冲进行标记,若脉冲 p_i 是某个雷达字的起始脉冲,则 $l_i = 1$, 否则 $l_i = 0$ 。将某 MFR 脉冲信号按本文所述方法处理后,得到 $\{\Delta P_{\text{PRI}}, \Delta R_{\text{RF}}, \Delta W_{\text{PW}}\}$ 在三维特征空间中的联合分布,如图 3 所示。

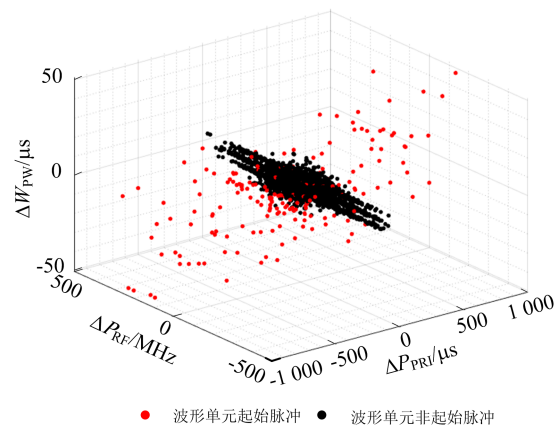


图3 MFR 脉冲信号联合分布图

Fig. 3 Distribution map of MFR pulse signal

图 3 中红点为波形单元起始脉冲,黑点为波形单元非起始脉冲。由于单个波形单元内部脉冲参数变化较为稳定,黑点集中且高密度分布在靠近原点的一小片特征空间中;由于相邻波形单

元之间的脉冲参数变化更加剧烈,因此波形单元起始脉冲对应的红点广泛分布在远离原点的大片特征空间中。因此可根据脉冲参数在 $\{\Delta P_{\text{PRI}}, \Delta R_{\text{RF}}, \Delta W_{\text{PW}}\}$ 中的分布情况将波形单元起始脉冲筛选出来。

2.3 分步变门限孤立森林模型构建

2.3.1 孤立森林提取模型

孤立森林算法(isolation forest, iForest)是一种高效的异常数据检测算法^[18],也称离群点数据挖掘,广泛应用于网络安全^[19]、金融安全^[20]等领域。在孤立森林中,将异常点数据定义为“分布稀疏且离高密度区较远的点”,这与 MFR 雷达字起始脉冲在 $\{\Delta P_{\text{PRI}}, \Delta R_{\text{RF}}, \Delta W_{\text{PW}}\}$ 特征空间中的分布相吻合。因此,本文将构建孤立森林作为 MFR 雷达字的提取模型。

孤立森林由多个孤立树(isolation tree, iTree)组成。设由脉冲参数差值 $\{\Delta P_{\text{PRI}}, \Delta R_{\text{RF}}, \Delta W_{\text{PW}}\}$ 组成数据集 $D = \{P_1, P_2, \dots, P_n\}$, $P_i = \{x'_{1i}, x'_{2i}, x'_{3i}\}$, $i \in [1, n]$ 。以数据集 D 为对象构建孤立森林波形单元提取模型的过程如下:

步骤 1 预设孤立树数量 t 和单棵孤立树叶子样本数 ϕ 。

步骤 2 从给定的数据集 $D = \{P_1, P_2, \dots, P_n\}$ 中随机抽取 ϕ 个样本组成新的数据集 D' , 将 D' 放入要训练的孤立树的根节点。

步骤 3 从样本集的 3 个维度中随机指定一个维度 q , 在当前节点数据集 D'^k 中随机产生一个分隔点 e , 满足:

$$\begin{aligned} \min(x_{ji} | j=q, x_{ji} \in D'^k) &< e < \\ \max(x_{ji} | j=q, x_{ji} \in D'^k) \end{aligned} \quad (8)$$

式中, D'^k 表示在此孤立树上按先后顺序产生的第 k 个节点数据集, 按照其产生的顺序依次进行处理。

步骤 4 通过切割点 e 在所选维度 q 的特征空间中生成一个超平面, 将当前节点中的样本点划分到 2 个子结点中: $x_{ji} < p$ 的样本点放入左子节点, $x_{ji} > p$ 的样本点放入右子节点。

步骤 5 重复步骤 3 和步骤 4, 直到所有的叶子节点都只有一个样本点, 或者当前孤立树达到了预设的高度时, 退出。

步骤 6 重复步骤 2~步骤 5, 直到生成全部 t 棵孤立树。

至此, 基于孤立森林的波形单元提取模型构

建完成。

在得到孤立森林模型后, 把 D 中所有数据点输入到全部孤立树组成的孤立森林中, 遍历所有的孤立树, 即可得到每个脉冲数据点对应的异常得分(abnormal score)。异常得分 $s(P_i, \phi)$ 的计算公式如下:

$$s(P_i, \phi) = 2^{-\frac{E(h(P_i))}{c(\phi)}} \quad (9)$$

式中, $h(P_i)$ 是 P_i 在每棵孤立树中从根节点到叶子节点的路径长度, $E(\cdot)$ 表示取均值, $c(\phi)$ 起到了归一化的作用, 为给定抽取样本数为 ϕ 时孤立树路径长度的平均值, 即:

$$c(\phi) = \begin{cases} 2H(\phi-1) - \frac{2(\phi-1)}{n}, & \phi > 2 \\ 1, & \phi = 2 \\ 0, & \phi < 2 \end{cases} \quad (10)$$

式中, $H(\phi) = \ln \phi + \xi$, ξ 为欧拉常数。

$s(P_i, \phi)$ 和 $E(h(P_i))$ 满足如下关系:

$$\begin{cases} E(h(P_i)) \rightarrow 0, & s(P_i, \phi) \rightarrow 1 \\ E(h(P_i)) \rightarrow \max(h(\cdot)), & s(P_i, \phi) \rightarrow 0 \\ E(h(P_i)) \rightarrow c(\phi), & s(P_i, \phi) \rightarrow 0.5 \end{cases} \quad (11)$$

由于波形单元起始脉冲对应的数据点远离高密度区域, 因此其在孤立森林中的平均高度 $E(h(P_i))$ 较非波形单元起始脉冲对应的数据点小, 故其异常得分 $s(P_i, \phi)$ 相对较高。这里只关心路径较短的数据点, 因此将单颗孤立树的路径长度进行如下限制:

$$\max(h(\cdot)) \leq \lceil (\ln \phi) \rceil \quad (12)$$

式中, $\lceil \cdot \rceil$ 表示向上取整。

将单颗孤立树的最大路径长度限制在 $\lceil (\ln \phi) \rceil$ 内, 可以有效降低训练孤立树时的运算量。

2.3.2 分步变门限策略

从 MFR 脉冲信号在 $\{\Delta P_{\text{PRI}}, \Delta R_{\text{RF}}, \Delta W_{\text{PW}}\}$ 特征空间的分布来看, 孤立森林算法可以完成波形单元起始脉冲的提取。但孤立森林算法应用于 MFR 的波形单元提取, 还存在以下问题^[21]:

(1) 孤立森林算法要求异常数据占总样本量的比例很小。若同时将大量的脉冲一阶差分特征值输入孤立森林中进行脉冲起始点的搜索, 波形单元起始脉冲对应的数据点大致会占到总数据量的 5%~10% (单个波形单元脉冲数量从几个到几十个不等)。这与孤立森林要求的异常数

据的小比例相违背。

(2) 对于非协作MFR信号,由于接收机只有本次侦收脉冲的参数数据,波形单元起始脉冲的准确率未知,区分波形单元起始脉冲和非起始脉冲的异常得分阈值无法确定。

针对以上问题,采用多次构建孤立森林,分多步、变门限完成波形单元的提取工作。预设一个门限 ϵ ,当完成完整的孤立树训练和样本点异常得分计算流程后,将所有异常得分 $s(P_i, \psi) > \epsilon$ 的样本点从数据集 D 中剔除,并放入波形单元起始脉冲数据集 $W_f = \{P_i, i | s(P_i, \psi) > \epsilon\}$ 中,而后再将剔除后的数据集作为新的数据集 D 。重复上述流程,直到新计算的异常得分都没有超过门限 ϵ 或到达最大搜索次数 r 时,结束整个波形单元起始脉冲的搜索。

阈值 ϵ 的设置根据当前轮次异常得分的分布情况确定。对于异常得分序列 $S = \{s(P_1), s(P_2), \dots, s(P_n)\}$,由于孤立树的最大高度限制在 $\lceil \lg \phi \rceil$ 内,故 $s(P_i), i \in [1, n]$ 的取值集中分布在 $[0.5, 1]$ 区间内。构建序列 S' :

$$S' = \{S, 1 - S\} \quad (13)$$

对于 S' ,其均值 μ 和方差 σ^2 为:

$$\begin{cases} \mu = 0.5 \\ \sigma^2 = \frac{1}{2n-1} \sum_{i=1}^{2n} (S'_i - \mu)^2 \end{cases} \quad (14)$$

S' 的得分分布大致拟合均值为 μ ,方差为 σ^2 的正态分布,如图4所示。故将 ϵ 确定为:

$$\epsilon = u + 3\sigma \quad (15)$$

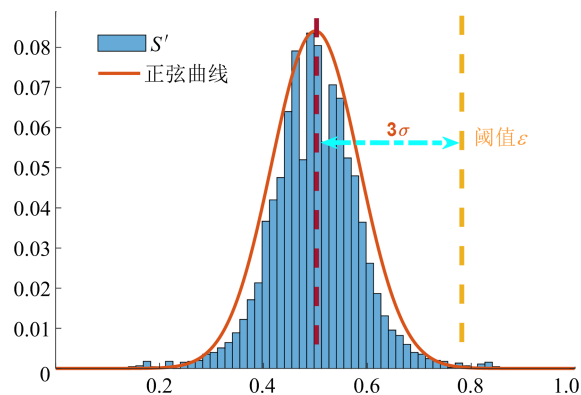


图4 阈值 ϵ 示意图

Fig. 4 Diagram of threshold ϵ

2.4 算法流程

本文方法的基本流程如图5所示,主要分为漏脉冲伪脉冲检测、特征参数一阶差分处理、分步变门限孤立森林提取模型训练构建3个步骤。

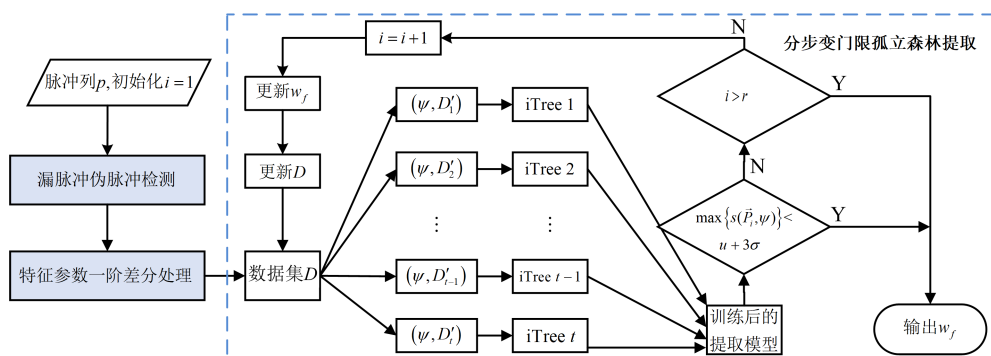


图5 分步变门限孤立森林波形单元提取流程

Fig. 5 Extraction process of isolated forest waveform unit with Multi-step and variable threshold

2.5 波形单元提取评价指标

(1) 本文方法对波形单元的提取,相当于将波形单元起始脉冲和非起始脉冲进行二分类。因此,定义如下评价指标:真正例(true positive, T_{TP})表示波形单元起始脉冲且被标记为起始脉冲的样本;假正例(false positive, F_{FP})表示波形单元非起始脉冲且被标记为起始脉冲的样本;真反例(true negative, T_{TN})表示波形单元非起始脉冲且被标记为非起始脉冲的样本;假反例(false

negative, F_{FN})表示波形单元起始脉冲且被标记为非起始脉冲的样本。

使用真正率(true positive rate, T_{TPR})、假正率(false positive rate, F_{FPR})和 F -score^[22]这3个指标来衡量波形单元提取算法的性能,其定义如下:

$$T_{TPR} = \frac{T_{TP}}{T_{TP} + F_{FN}} \quad (16)$$

$$F_{FPR} = \frac{F_{FP}}{F_{FP} + T_{TN}} \quad (17)$$

$$F\text{-score} = \frac{2PR}{P+R} \quad (18)$$

$$\text{式中, } P = \frac{T_{TP}}{T_{TP} + F_{FP}}, R = \frac{T_{TP}}{T_{TP} + F_{FN}}.$$

(2) 孤立森林算法是具有线性时间复杂度的异常检测算法。为评价本文方法在时间复杂度上的性能优势,分析和测试在不同样本数量情况下的计算复杂度和时间消耗。

3 实验结果与分析

本节将通过实验,验证基于孤立森林提取波形单元起始脉冲实验的可行性和有效性。实验基于某 MFR 模拟器生成 40 种波形单元数据进行分析。

3.1 算法有效性分析

首先将本文所提基于孤立森林的波形单元提取算法与同为无监督算法的 DBSCAN、DA-DBSCAN^[15] 进行对比,并考虑漏脉冲、伪脉冲和参数误差情形。脉冲丢失率 (ratio of dropped pulses, R_{RDP})、虚假脉冲率 (ratio of false pulses, R_{RFP}) 和误差偏离水平 (error deviation level, E_{EDL}) 的定义如下:

$$R_{RDP} = \frac{n_{\text{drop}}}{n} \times 100\% \quad (19)$$

$$R_{RFP} = \frac{n_{\text{false}}}{n} \times 100\% \quad (20)$$

$$E_{EDL} = \frac{\xi_{pi}}{x_{pi}} \times 100\% \quad (21)$$

式中, n 为脉冲列 $p = \{p_1, p_2, p_3, \dots, p_n\}$ 包含的脉冲个数, n_{drop} 为随机丢弃的脉冲个数, n_{false} 为混入的虚假脉冲个数, $x_{pi} = \{x_{1i}, x_{2i}, x_{3i}\}$ 为无误差脉冲参数数据, $\xi_{pi} = \{\xi_{1i}, \xi_{2i}, \xi_{3i}\}$ 为随机测量误差。

生成由 200 个波形单元构成的脉冲列。 R_{RDP} 、 R_{RFP} 和 E_{EDL} 取值为 $0 \sim 20\%$ 。加入 2 部固定体制雷达信号作为伪脉冲,参数分别设置为 $P_{PRI_1} = 1\ 000\ \mu\text{s}$ 、 $R_{RF_1} = 3\ \text{GHz}$ 、 $W_{PW_1} = 20\ \mu\text{s}$ 和 $P_{PRI_2} = 500\ \mu\text{s}$ 、 $R_{RF_2} = 4\ \text{GHz}$ 、 $W_{PW_2} = 10\ \mu\text{s}$ 。由于 DBSCAN 与 DA-DBSCAN 算法基于密度进行聚类,将 P_{PRI} 、 R_{RF} 、 W_{PW} 按照 μs 、 MHz 、 μs 的数值进行聚类,聚类参数 Minpts 为 7, Eps 为 3。本文算法的参数设置如下:漏脉冲伪脉冲检验参数 $\tau = 0.15$,单棵树训练样本数 ψ 为数据集 D 样本数量的 $1/10$,孤立树的数量 $t = 100$,最大搜索次数 10 次。进行 500 次 Monte Carlo 实验后的平均结果分别如图 6~图 8 所示。

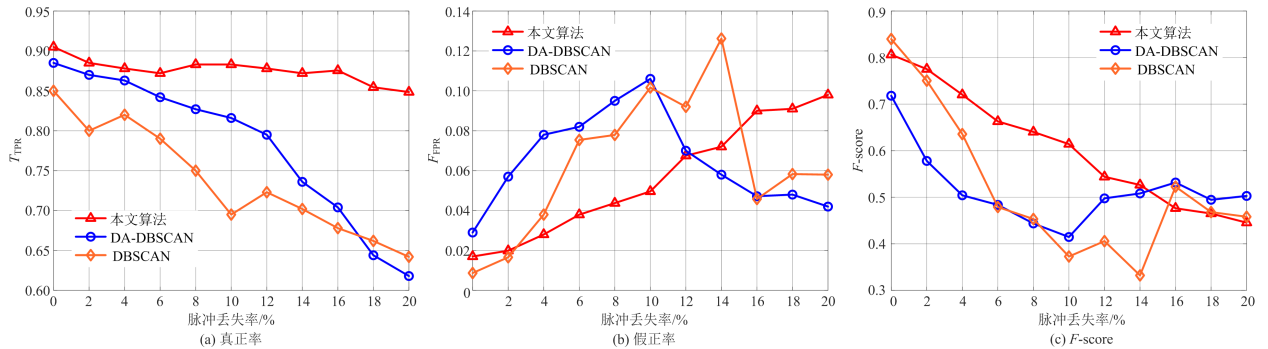


图 6 不同 RDP 下各算法提取性能

Fig. 6 Extraction effect of each algorithm under different RDP

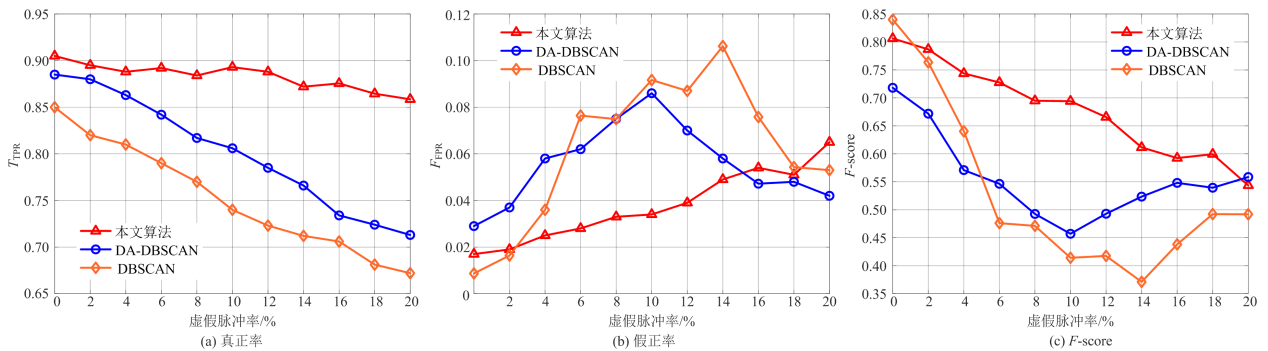


图 7 不同 RFP 下各算法提取性能

Fig. 7 Extraction effect of each algorithm under different RFP

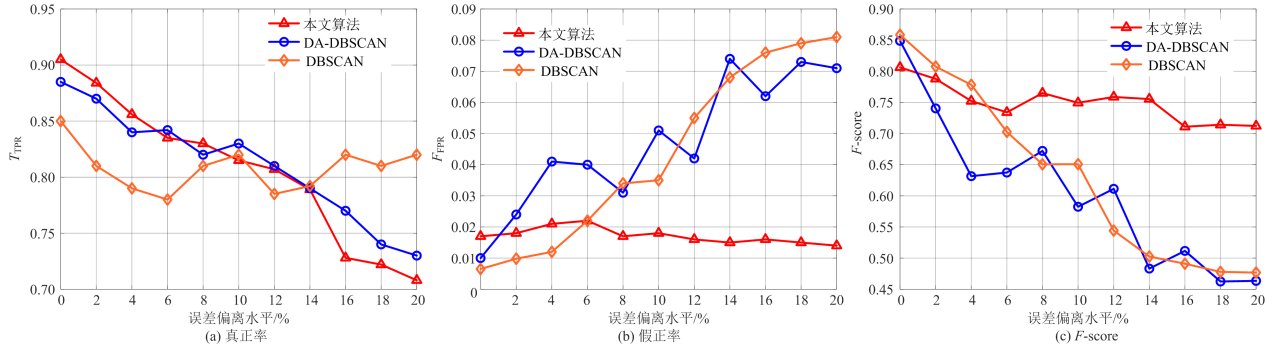


图 8 不同 EDL 下各算法提取性能

Fig. 8 Extraction effect of each algorithm under different EDL

从图 6 和图 7 可以看出,在面临不同程度的漏脉冲和伪脉冲情况时,本文的波形单元提取算法可以始终保持 T_{TPR} 在 85% 以上。同时,由于本文算法基于隔离来实现波形单元起始脉冲的搜索,随着 R_{RDP} 、 R_{RFP} 不断增加,异常 P_{PRI} 导致更多 ΔP_{PRI} 剧烈变化, F_{FPR} 逐渐增加到 10% 和 6%;而对于 DBSCAN 和 DA-DBSCAN 来说,随着 R_{RDP} 、 R_{RFP} 不断增加,其 F_{FPR} 先显著增加到 10% 左右,而后逐渐降低。这是由于其基于特征空间密度的大小来判断波形单元起始,随着脉冲错误水平导致 P_{PRI} 异常概率的不断上升, P_{PRI} 异常脉冲样本点在特征空间中的分布逐渐由稀疏变为密集,从而不再被误判。

从图 8 可以看出,在面临不同程度的脉冲参数测量误差情况时,本文算法与对比算法的策略有所不同。随着脉冲参数测量误差的不断增大,本文算法的提取 T_{TPR} 由 90% 逐渐降低到 70%,但是 T_{FPR} 会始终保持在 2% 附近;而参与对比的 DBSCAN 和 DA-DBSCAN 的 TFR 都随着误差水平的提高而显著增加到 7% 以上,但 T_{TPR} 并没有明显优于本文算法。这导致随着脉冲参数测量误差的不断增大,其 F -score 值相比本文算法的差距逐渐增加。

总的来说,本文算法与对比算法相比在波形单元提取性能上各有优势。考虑到真实侦收脉冲分选错误水平一般较低,本文算法在面临漏脉冲和伪脉冲时的波形单元起始脉冲提取性能表现优于对比算法。而在脉冲参数测量误差较大时,本文算法能够更好地兼顾查全率和查准率这两者的平衡,具备更加优异的表现。

3.2 算法实时性分析

雷达辐射源可以在短时间内发射大量脉冲信号,因此波形单元提取算法的计算复杂度是衡

量其性能的一项重要指标。本节首先从理论上对上述 3 种算法的计算复杂度进行分析,并以仿真实验统计 3 种算法在处理不同数量波形单元时的平均所用时长。

为了对算法计算复杂度进行理论分析,首先明确如下变量: m 表示测试序列的波形单元数量; t 表示孤立树的数量; ψ 表示训练单棵树用到的样本数量; r 表示最大搜索轮次。测试序列的脉冲 N 总数由式(1)已给出。

对于本文算法,其计算复杂度 T_{IF} 为:

$$T_{\text{IF}} = \underbrace{\sum_{i=1}^m 3l(\omega_i)}_{\Delta P_{\text{PRI}}, \Delta R_{\text{RF}}, \Delta W_{\text{PW}} \text{ 处理}} + \underbrace{\sum_{i=1}^m 6[l(\omega_i) - 1]}_{\text{漏脉冲检测}} + \underbrace{\sum_{i=1}^m 6[l(\omega_i) - 1]}_{\text{伪脉冲检测}} + \underbrace{\left[\sum_{i=1}^t \sum_{j=1}^{h(P_{\text{tree}_i})} (\psi - j) \right]}_{\text{孤立树生成}} + \underbrace{\sum_{i=1}^t \sum_{j=1}^{h(P_i)} \sum_{k=1}^m l(\omega_k)}_{\text{孤立森林遍历所有样本}} r \quad (22)$$

由于 $h(P_{\text{tree}_i}) \geq h(p_i)$, 且对于所有孤立树 $h(P_{\text{tree}_i}) \leq \lceil \lg \psi \rceil$, 故:

$$T_{\text{IF}} \leq 15N + \left[\sum_{i=1}^t \sum_{j=1}^{\lceil \lg \psi \rceil} \psi + \sum_{i=1}^t \sum_{j=1}^{\lceil \lg \psi \rceil} \sum_{k=1}^m l(\omega_k) \right] r = 15 \cdot N + [t\psi \lceil \lg \psi \rceil + t \cdot N \cdot \lceil \lg \psi \rceil] r = 15 \cdot N + (\psi + N)t \lceil \lg \psi \rceil r \quad (23)$$

式(2)和式(3)已分别给出了 DBSCAN 和 DA-DBSCAN 的计算复杂度 T_{DB} 和 T_{DA} 。设波形单元的平均脉冲个数为 20, 孤立树数量 $t = 100$, 训练单棵树的样本数量 $\psi = 0.1N$, 最大搜索次数 $r = 10$, 将上述条件带入复杂度计算结果, 可得到不同样本数量下 3 种算法的计算复杂度比较

结果,如图 9 所示。

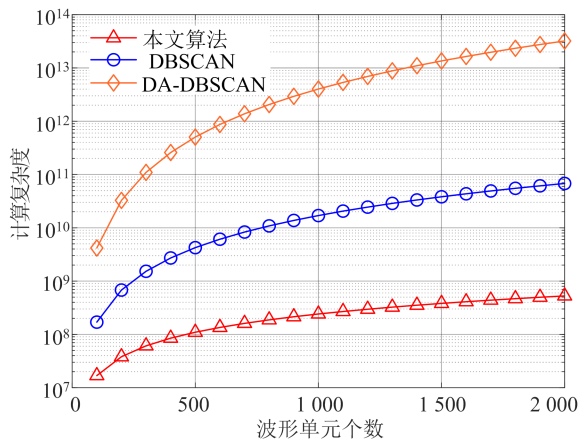


图 9 不同波形单元数量下各算法计算复杂度

Fig. 9 The computational complexity of each algorithm under different number of waveform units

图 9 表明,在各波形单元数量下,本文方法相比对比算法具有更低的计算复杂度。

条件设置与图 9 对比计算复杂度时保持不变,本文进行实验统计了 3 种算法在不同波形单元数量下的平均耗时。实验数据在 i9-11900K + 32G DDR memory + Windows 10 平台上,使用 Matlab 2020a 计算并测得。结果见表 1 所列。

表 1 不同波形单元数量下各算法运行时间

Table 1 The running time of each algorithm under different number of waveform units

单位:s			
波形单元数量	本文算法	DBSCAN	DA-DBSCAN
100	2.032	3.669	20.341
200	4.762	17.242	183.426
500	11.242	193.244	>600

从表 1 可以看出,本文方法的运行时间最短,并且运行时间基本随波形单元数量线性增加。当波形单元数量达到 500 时,本文算法运行时间分别只有 DBSCAN 和 DA-DBSCAN 的不到 6% 和 2%,优势明显。

从理论分析和仿真实验可以看出,本文算法相比 DBSCAN 和 DA-DBSCAN 算法的计算复杂度明显降低,特别在波形单元数量较多时,计算实时性的优势更加明显。此外,本文算法没有超参数需要人为设置,避免了因经验不足导致算法效果不佳的问题。综合本文算法与对比算法识别性能并没有太大差距以及本文算法在计算复

杂度上的优势,本文算法具有更好的工程应用前景与研究价值。

4 结束语

本文提出了一种基于孤立森林提取 MFR 波形单元的方法,能够做到在准确提取波形单元起始脉冲的同时,大大降低算法耗时。该方法根据 MFR 波形单元内部和波形单元间脉冲串参数变化的差异性,利用脉冲参数的一阶差分来描述波形单元切换时的特点,将波形单元提取问题转化为波形单元起始脉冲提取的二分类问题,并引入了改进的孤立森林算法进行波形单元起始脉冲的搜索,有效解决了传统基于聚类搜索起始脉冲算法的高运算量问题。实验表明,本文方法在面临漏脉冲、伪脉冲、参数误差时具有较好的鲁棒性,同时运算耗时短,性能优。但是,本文算法的提取准确率还有待提高,这还依赖于对波形单元脉冲参数的变化规律作更加准确的描述。

参考文献

- [1] 马爽,王莹桂,柳征,等. 基于序列比对的多功能雷达搜索规律识别方法[J]. 电子学报, 2012, 40(7): 1434-1439.
MA Shuang, WANG Yinggui, LIU Zheng, et al. A Method for search schema recognition of multifunction radars based on sequence alignment[J]. Acta Electronica Sinica, 2012, 40(7): 1434-1439. (in Chinese)
- [2] CHARLISH A B. Autonomous agents for multi-function radar resource management[D]. London: University College London, 2011.
- [3] LAVOIE P. Hidden Markov modeling for radar electronic warfare: U. S. Patent 6, 788, 243[P]. 2004-09-07.
- [4] 陈维高,贾鑫,朱卫纲,等. 基于 HMM 的雷达状态转移估计方法[J]. 北京航空航天大学学报, 2017, 43(10): 2171-2180.
CHEN Weigao, JIA Xin, ZHU Weigang, et al. Radar state transfer estimation method based on HMM[J]. Journal of Beijing University of Aeronautics and Astronautics, 2017, 43(10): 2171-2180. (in Chinese)
- [5] VISNEVSKI N, KRISHNAMURTHY V, HAYKIN S, et al. Multi-function radar emitter modelling: a stochastic discrete event system approach[C]//Proceedings of IEEE Conference on Decision & Control. [S.l.]: IEEE, 2004: 6295-6300.
- [6] WANG A, KRISHNAMURTHY V. Modeling and interpretation of multifunction radars with stochastic

- grammar[C]// Proceedings of IEEE Aerospace Conference. [S. l.]: IEEE, 2008: 1-13.
- [7] WANG A, KRISHNAMURTHY V. Threat estimation of multifunction radars: modeling and statistical signal processing of stochastic context free grammars [C]// Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. [S. l.]: s. n., 2007: 793-796.
- [8] HAYKIN S, MOULINES E. Modeling, identification, and control of largescale dynamical systems [C]// Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. [S. l.]: IEEE, 2005: 945-948.
- [9] VISNEVSKI N, KRISHNAMURTHY V, WANG A, et al. Syntactic modeling and signal processing of multifunction radars: a stochastic context free grammar approach[J]. Proceedings of the IEEE, 2007, 95(5): 1000-1025.
- [10] ARASARATNAM I, HAYKIN S, KIRUBARAJAN T, et al. Tracking the mode of operation of multifunction radars[C]// Proceedings of IEEE Conference on Radar. [S. l.]: IEEE, 2006: 1-6.
- [11] 刘海军,樊昀,李悦,等. 多功能雷达建模中的雷达字提取技术研究[J]. 国防科技大学学报, 2010, 32(2): 91-96.
- LIU Haijun, FAN Yun, LI Yue, et al. Research on extracting of radar words in modeling of multi-function radar[J]. Journal of National University of Defense Technology, 2010, 32(2): 91-96. (in Chinese)
- [12] 王勇军. 一种改进的事件驱动的 MFR 雷达字提取方法[J]. 现代雷达, 2019, 41(3): 17-20.
- WANG Yongjun. Novel approach of radar word extraction for MFRs based on event driven method[J]. Modern Radar, 2019, 41(3): 17-20. (in Chinese)
- [13] ZHU M, LI Y, WANG S. Model-based time series clustering and inter pulse modulation parameter estimation of multifunction radar pulse sequences [J]. IEEE Transactions on Aerospace and Electronic Systems, 2021, 57(6): 3673-3690.
- [14] 马爽,柳征,姜文利. 基于幅度变化点检测的多功能雷达脉冲列解析方法[J]. 电子学报, 2013, 41(7): 1436-1441.
- MA Shuang, LIU Zheng, JIANG Wenli. A method for multifunction radar pulse train analysis based on amplitude change point detection[J]. Acta Electronica Sinica, 2013, 41(7): 1436-1441. (in Chinese)
- [15] 阳榴,朱卫纲,吕守业,等. 面向非协作多功能雷达的波形单元提取方法[J]. 系统工程与电子技术, 2021, 43(10): 2843-2850.
- YANG Liu, ZHU Weigang, LV Shouye, et al. Waveform unit extraction method for noncooperative multifunction radar[J]. Systems Engineering and Electronics, 2021, 43(10): 2843-2850. (in Chinese)
- [16] 刘章孟,袁硕,康仕乾. 多功能雷达脉冲列的语义编码与模型重建[J]. 雷达学报, 2021, 10(4): 559-570.
- LIU Zhangmeng, YUAN Shuo, KANG Shiqian. Semantic coding and model reconstruction of multifunctional radar pulse train[J]. Journal of Radars, 2021, 10(4): 559-570. (in Chinese)
- [17] LIU Z M. Recognition of multi-function radars via hierarchically mining and exploiting pulse group patterns [J]. IEEE Transactions on Aerospace and Electronic Systems, 2020, 56(6): 4659-4672.
- [18] YANG L, ZHU W G, LV S Y, et al. Waveform unit extraction method for noncooperative multifunction radar[J]. Systems Engineering and Electronics, 2021, 43(10): 133-139.
- [19] LIU F T, TING K M, ZHOU Z H. Isolation forest [C]// Proceedings of the 8th IEEE International Conference on Data Mining. [S. l.]: IEEE, 2008: 413-422.
- [20] ALQURASHI S, SHIRAZI H, RAY I. On the performance of isolation forest and multilayer perceptron for anomaly detection in industrial control systems networks[C]// Proceedings of the 8th International Conference on Internet of Things: Systems, Management and Security (IOTSMS). [S. l.]: IEEE, 2021: 1-6.
- [21] KIRAN M, WANG C, PAPADIMITRIOU G, et al. Detecting anomalous packets in network transfers: investigations using PCA, autoencoder and isolation forest in TCP[J]. Machine Learning, 2020, 109(5): 1127-1143.
- [22] 杨晓晖,张圣昌. 基于多粒度级联孤立森林算法的异常检测模型[J]. 通信学报, 2019, 40(8): 133-142.
- YANG Xiaohui, ZHANG Shengchang. Anomaly detection model based on multi grained cascade isolation forest algorithm [J]. Journal on Communications, 2019, 40(8): 133-142. (in Chinese)
- [23] RUDD E M, JAIN L P, SCHEIRER W J, et al. The extreme value machine[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(3): 762-768.