

DOI: 10.13382/j.jemi.B2003379

基于 kNN 优化算法的密封电子设备多余物定位技术*

孙志刚¹ 王国涛^{1,2} 高萌萌¹ 郜雷阵² 蒋爱平¹

(1. 黑龙江大学 电子工程学院 哈尔滨 150080; 2. 哈尔滨工业大学 军用电器研究所 哈尔滨 150001)

摘要:在密封电子设备的生产制造过程中,对多余物进行检测至关重要。微粒碰撞噪声检测法是我国军标规定的用于航天电子器件多余物检测的方法。针对密封电子设备体积大和检测出的多余物位置难以确定的问题,使用参数优化的 k 邻近(kNN)算法对多余物进行定位。通过搭建定位实验系统和设计试件模型,得到多通道的多余物信号,提取性能优良的时域和频域特征作为 kNN 算法学习的数据集。采用网格搜索法寻找 kNN 算法最优的 k 值选择、距离度量和权重设置,然后采用参数优化的 kNN 算法分别建立平面与空间定位模型。实验结果表明,采用参数优化的 kNN 算法进行多余物定位,平面与空间定位精度分别达到 81.18% 和 79.34%,有效提高了传统情况下的定位准确度。

关键词:密封电子设备;多余物;kNN 算法;特征提取;定位

中图分类号: TM58; TP181 **文献标识码:** A **国家标准学科分类代码:** 510

Sealed electronic equipment loose particle positioning technology based on kNN algorithm of parameter optimization

Sun Zhigang¹ Wang Guotao^{1,2} Gao Mengmeng¹ Gao Leizhen² Jiang Aiping¹

(1. Electronic Engineering College, Heilongjiang University, Harbin 150008, China;

2. Military Apparatus Research Institute, Harbin Institute of Technology, Harbin 150001, China)

Abstract: Detection of the loose particles is urgently required in the production of sealed electronic equipment. Particle impact noise detection is a national aerospace electronic component loose particles detection method. Aiming at the problem of the large volume of the sealed electronic equipment and the difficulty in determining the position of the loose particle, this paper uses the parameter optimized kNN algorithm to locate the loose particle. After building a positioning experiment system and designing a specimen model, a multi-channel loose particle signal is obtained, and the time domain and frequency domain features with excellent performance are extracted as the data set for kNN algorithm learning. The grid search method is used to find the optimal k value selection, distance measurement and weight setting of the kNN algorithm, then the kNN algorithm of parameter optimization is used to establish the plane and space positioning models respectively. The experimental results show that using the kNN algorithm of parameter optimization for loose particle positioning, the accuracy of plane and space positioning reaches 81.18% and 79.34% respectively, which effectively improves the positioning accuracy under traditional conditions.

Keywords: sealed electronic equipment; loose particle; kNN algorithm; feature extraction; positioning

0 引言

密封电子设备是指那些具有密封防水和电磁屏蔽等

特定功能的封闭式构造的组件,如航天继电器、运载火箭、人造卫星、火箭发动机和空间探测器等,主要用于航空航天等领域的定位、导航、通信和飞行控制等任务。密封电子设备的生产和装配过程工序繁多,各个环节都有

收稿日期: 2020-08-05 Received Date: 2020-08-05

* 基金项目:国家自然科学基金(51607059,51077022)、黑龙江省自然科学基金(QC2017059)、黑龙江省博士后基金(LBH-Z16169)、黑龙江省高校基本科研业务费(HDRCCX-201604)、黑龙江省教育厅科技成果培育(TSTAU-C2018016)、黑龙江大学校内项目(HDJMRH201912,2012TD007,QL2015)资助

可能将焊渣碎末、金属碎片、垫圈、涂层杂质和导线段等微粒物封装在设备内部^[1],这些留存在设备中的由外部引入或内部生成的破坏设备原有状态的一切物质,称为多余物。多余物在振动、冲击、超失重等外力的作用下,可能会碰撞损坏元器件或者附着在电路表面引起短路或线路异常,进而导致整个电子设备工作异常,甚至引发事故造成人员伤亡^[2]。因此,对密封电子设备进行多余物检测、定位及故障清除,是排除密封电子设备工作故障、保障整个航天系统可靠运行的重要手段。

目前,对存在于密封电子设备内多余物的研究重点大多聚集在判断是否存在多余物、设备内多余物信号与内部组件信号的判断与识别等方面^[3],关于多余物定位方面的研究不够深入。实质上,多余物碰撞产生的信号属于广义上的声发射信号,关于声发射源的定位问题一直是领域内关注的重点,其常用方法包括时差定位和区域定位,相关学者对此进行了深入的研究。Duff 等^[4]提出了一种利用两个信号包络的短时互相关函数来估计信号时延的方法。Lympertos 等^[5]提出了利用多个传感器和估计不同频率下的到达时间,作为一个平方误差最小化问题解决了声发射源到传感器的距离的问题,进而推断出声发射源的位置。行鸿彦等^[6]建立四元传声器阵列模型,将各阵元间的相对位置关系转化为三维坐标,利用声源传播到各阵元时延值对声源所在方位进行反演,得到声源的坐标和角度。随着声发射源定位研究的深入,相关学者将研究推广到各领域的应用中。Dris 等^[7]提出了一种板状结构声发射源定位的几何方法,通过实验比较了阈值交叉法和连续小波变换法,证明连续小波变换可以提高声发射源定位的精度。Nivesrangsan 等^[8]采用二维平面定位的方法,基于传统波速和声发射能量技术解决发动机故障源的定位问题,通过比较声信号到达各传感器的时间差和信号的信噪比,最终识别出故障源的位置。丁丹丹等^[9]基于传感器阵列采集信号的同步性、多余物活动范围和多余物信号传播的衰减规律等特征,采集导弹舱体内的多余物碰撞声信号,通过累加多次碰撞信号能量的方法,确定多余物的位置。虽然通过上述两种定位方法取得了一定的声发射定位研究成果,但传统的时差定位法在组成材质不均匀或内部结构复杂的物体中,难以测量声信号的时延和传播速度,进而影响了声发射源的定位精度;区域定位法同样不能获得声发射源的精确三维坐标位置,只能大概确定声发射源的区域性位置,且应用该方法得到的测试结果具有不确定性。

近年来,随着机器学习的兴起与快速发展,越来越多的学者将机器学习领域的方法应用到声发射源的定位上,这类方法主要是通过应用机器学习的相关算法来完成对目标源的定位。Yang 等^[10]基于 BP 神经网络设计了机器人声源定位系统,通过获取各通道信号的强度和

相位特征生成训练数据集,系统取得了对声源定位的良好效果,但其提取的特征数量有限,且未对提取特征的可靠性进行分析说明。李卓^[11]针对航天密封舱中撞击源定位的问题,通过研究机器学习中的 k 邻近(kNN)分类、k 均值聚类等算法的可行性并最终设计定位方法,与传统定位方法相比,在定位精度上有了明显改善,但该方法仍然是在区域定位法得到局部范围的基础上展开研究的,本质上来说没有摆脱传统定位方法的限制。郝雷阵^[12]将机器学习分类算法与密封电子设备的多余物定位问题相结合,综合比较多种分类算法在多余物定位上的表现效果,但其缺少对其中的具体某种算法展开深入的研究。本文在前人的研究基础上,分析机器学习分类算法的基本原理,提出了应用 kNN 算法的多余物定位方法。通过网格搜索法对 kNN 算法的 k 值选择、距离度量和权重设置等参数进行了寻优,最终使用参数优化的 kNN 算法建立分类器模型实现对多余物的定位。

1 相关工作

1.1 kNN 算法

kNN 算法是机器学习中常用的分类算法,它的工作原理如下:对于待预测的样本,基于某种距离度量找出训练集中与其距离最近的 k 个样本,然后基于这 k 个样本的类别信息,根据某种决策规则来进行预测^[13],通常使用多数表决的规则,也称“投票法”,即选出这 k 个样本中多数样本所属的类别作为待预测样本的预测结果。另外,还可以基于距离的远近对 k 个训练样本进行加权投票,距离越近的样本“投票”权重越大。所以本文在使用 kNN 算法进行模型分类时,主要对上述三个参数进行研究 with 优化。

1) k 值选择

k 值的选择对 kNN 算法的预测结果有重大的影响,如果选择较小的 k 值,只有训练集中距离待预测样本最近的 k 个样本才会对预测结果起作用。虽然一定程度上会减少学习的近似误差,但是会增大学习的估计误差,且预测结果会对距离较近的 k 个训练样本非常敏感。除此之外,如果距离较近的 k 个训练样本中存在噪声样本,预测结果就可能出错。

如果选择距离较大的 k 值,相当于选择较大邻域范围内的 k 个训练样本用于预测,虽然一定程度上会减少学习的估计误差,但是对应增大学习的近似误差。这时训练集中距离待预测样本较远的样本也会对预测结果起作用,使预测结果发生错误。

2) 距离度量

两个具有相同维度特征空间的样本之间的距离,反映了两者的相似性^[13],kNN 算法的特征空间一般是

n 维实数向量空间 \mathbf{R}^n 。为了测量两个样本之间的相似度,需要引入距离函数进行度量,最常见的选择是闵可夫斯基距离。

设特征空间 X 是 n 维实数向量空间 $\mathbf{R}^n, \mathbf{x}_i, \mathbf{x}_j \in X, \mathbf{x}_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(n)})^T, \mathbf{x}_j = (x_j^{(1)}, x_j^{(2)}, \dots, x_j^{(n)})^T, \mathbf{x}_i, \mathbf{x}_j$ 的闵可夫斯基距离 L_p 定义为:

$$L_p(\mathbf{x}_i, \mathbf{x}_j) = \left(\sum_{l=1}^n |x_i^{(l)} - x_j^{(l)}|^p \right)^{\frac{1}{p}} \quad (1)$$

式中: $p \geq 1$ 。

当 $p = 2$ 时,称为欧氏距离,即:

$$L_2(\mathbf{x}_i, \mathbf{x}_j) = \left(\sum_{l=1}^n |x_i^{(l)} - x_j^{(l)}|^2 \right)^{\frac{1}{2}} \quad (2)$$

当 $p = 1$ 时,称为曼哈顿距离,即:

$$L_1(\mathbf{x}_i, \mathbf{x}_j) = \sum_{l=1}^n |x_i^{(l)} - x_j^{(l)}| \quad (3)$$

当 $p = \infty$ 时,称为切比雪夫距离,即:

$$L_\infty(\mathbf{x}_i, \mathbf{x}_j) = \max_l |x_i^{(l)} - x_j^{(l)}| \quad (4)$$

另外,马氏距离能够有效计算两个未知样本集之间相似度的方法,是表征样本间的协方差距离^[14]。在同种分布的样本集中,可以通过协方差矩阵计算两个样本之间的马氏距离。 $\mathbf{x}_i, \mathbf{x}_j$ 的马氏距离 L_M 定义为:

$$L_M(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \mathbf{x}_j)} \quad (5)$$

式中: $\boldsymbol{\Sigma}$ 是 n 维实数变量的协方差矩阵。

如果协方差矩阵是单位向量,即各维度独立同分布,马氏距离就变成了欧氏距离。

3) 权重设置

kNN 算法是采用多数表决的分类决策规则,也就是“投票法”。即由待预测样本的 k 个邻近训练样本中的多数类别决定待预测样本的类别。在这种情况下, k 个邻近训练样本投票的权重相等。当距离待预测样本较远且与之类别不一致的训练样本被包含在邻域范围内时,其具有的等同投票权重会影响正确类别的预测结果。

因此,可以在邻域范围内基于距离的远近,对 k 个训练样本进行加权投票,使得距离越近的训练样本“投票”权重越大,代表着投票可靠性的增强,降低了部分距离较远的训练样本对预测结果的投票影响。

1.2 实验系统

为解决设备内存在的多余物问题,美国宇航局最早提出微粒碰撞噪声检测法来判断多余物的存在,并将其应用到航天发动机、密封继电器、火箭密封舱等电子元件的多余物检测上^[15]。本文在此基础上,设计了多余物定位实验系统,系统由被测试件和摆台、数据采集系统和摆台驱动系统、上位机 3 部分组成,多余物定位实验系统结构框图如图 1 所示。

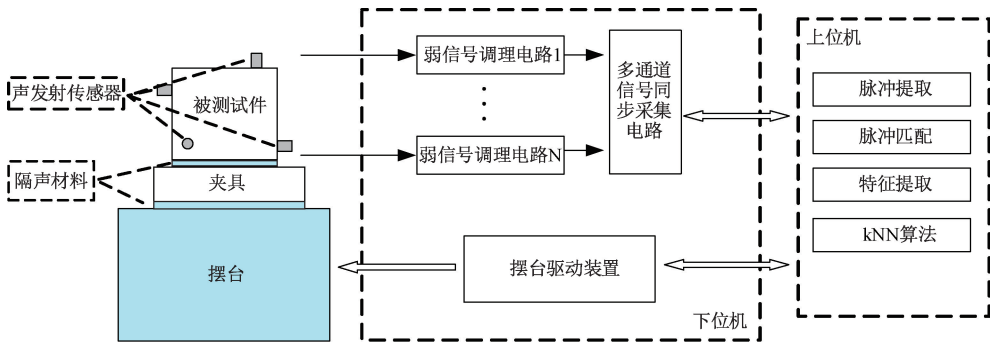


图 1 多余物定位实验系统结构框图

Fig. 1 Block diagram of the loose particle positioning experiment system

与微粒碰撞噪声检测法类似,实验系统中的上位机通过驱动摆台装置来激励被测试件,使存在于被测试件内部的多余物产生碰撞或滑动状态,均匀放置在被测试件表面不同位置的声发射传感器,用来采集多余物运动产生的声信号,转化成电信号后送入弱信号调理电路进行放大、滤波等处理。多通道信号同步采集电路将各通道调理后的多余物信号由模拟信号转换成数字式的信号数据,同步采集后通过 USB 控制器发送至上位机保存。上位机对保存的信号数据进行预处理,包括脉冲提取和脉冲匹配等,之后提取各通道信号中能反映多余物位置的特征信息,包括时域特征和频域特征,大量特征的集合

组成了用于 kNN 算法的数据集,将其按照固定比例划分为训练集和测试集,训练集中的数据用于 kNN 算法的训练,测试集中的数据用于对 kNN 算法构建的定位模型进行性能评估。

1.3 实验模型

由于密封电子设备的特殊性,被测试件利用实物进行测试的机会较少。因此,本文制作了被测试件的实验模型,图 2 所示是用钢板制作的实验模型示意图,其规格为 400 mm×400 mm×400 mm,模型内部用 3 块互相正交的钢板分隔为 8 个相等的互不相通的空间区域。本文中的多余物定位实验分为平面定位和空间定位两部分,其

中,平面定位实验利用被测试件模型的一个表面,模拟那些厚度可忽略不计的板状密封电子设备,如铝板结构的设备等,并按照如图 2(a) 所示划分成 16 个相等的互不相通的平面区域,因此每个平面区域的定位精度达到 100 mm,并将 4 个声发射传感器放置在平面区域的 4 个直角。空间定位实验是利用整个被测试件模型,并按照如图 2(b) 所示在模型表面的不同位置放置 4 个声发射传感器,每个空间区域的定位精度达到 200 mm。

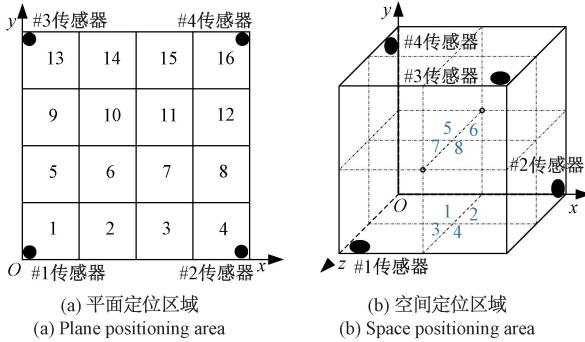


图 2 定位实验方案

Fig. 2 Positioning experiment plan

2 特征提取

传统声发射源定位方法通过确定声发射信号到达传感器的先后顺序或时间差,或者声发射信号在介质中的传播速度。针对组成结构简单的被测试件,这种方法的实现效果还算理想,只需对测试环境和装置严格要求。一旦被测试件的内部结构趋于复杂,声发射信号的传播速度就难以获取,传统的时差定位法也将难以胜任,区域定位法也达不到要求的精度,因此需要探寻新的定位方法。由于声发射信号的衰减特性和传播介质的各向异性,多余物信号存在的某些特征会随传播方向和传播距离的不同而有所差异,本文主要从时域和频域两个方面来分析多余物信号的特征,利用这些特性组成数据集并用于 kNN 算法的训练学习,得到性能优良的定位模型去预测未知多余物的位置。

2.1 传播特性

多余物信号属于广义上的声发射信号,所以它与传统声发射信号具有类似的传播机理。受密封电子设备组成材质、内部结构和热传导性能等因素的影响,在进行多余物检测的过程中信号的传播形式和波形会变得复杂。同时,信号在介质的传播过程中会发生散射、衰减甚至波形畸变的现象^[16],主要包括几何扩展衰减、材料吸收衰减和散射衰减 3 种形式。

假设多余物产生的声发射信号如下:

$$u(x, t) = Ae^{-i\omega(t-x/C_0)} \quad (6)$$

式中: A 为信号的振幅; ω 为信号的角频率; C_0 为信号的相速度。

则在 $x = b$ 处声发射信号的形式如下:

$$u(b, t) = \frac{1}{2} \pi \int_{-\infty}^{\infty} A(\omega) e^{-i\omega(t-b/C_0)} d\omega \quad (7)$$

若不考虑信号的衰减时,分析其频谱特性可得:

$$U(b, \omega) = A(\omega) e^{i\omega b/C_0} = |U(b, \omega)| e^{i\varphi(b, \omega)} \quad (8)$$

若考虑传播过程中的衰减因素,则有:

$$U'(b, \omega) = A(\omega) e^{-\alpha(\omega)b} e^{i\omega b C(\omega)} = |U'(b, \omega)| e^{i\varphi'(b, \omega)} \quad (9)$$

式中: $\alpha(\omega)$ 是频率的衰减系数; $C(\omega)$ 是相速度的衰减系数; $\varphi'(b, \omega)$ 是衰减之后的相位谱。

通过对比衰减前后的频谱特性,可知:

$$|U'(b, \omega)| = |U(b, \omega)| e^{-\alpha(\omega)b} \quad (10)$$

$$\varphi'(b, \omega) = \varphi(b, \omega) + \omega b g(\omega) \quad (11)$$

$$g(\omega) = \frac{1}{C(\omega)} - \frac{1}{C_0} \quad (12)$$

由上述分析可知,多余物信号在介质的传播过程中,在时域和频域上产生了衰减和畸变现象,正是信号的传播特性为多余物定位研究提供了理论基础。考虑到声发射传感器的空间布局,因为多余物信号到各传感器的传播路径和距离不同,所以各传感器接收到的多余物信号存在着较大差异,图 3 所示为多通道信号同步采集电路获取的多余物信号。

分析多通道多余物信号的时域图可知,幅值是各通道多余物信号之间最显著的差别,借助幅值进一步分析可得,通道 2 的声发射传感器距离多余物信号源最近,通道 4 的传感器距离最远,其他 2 个通道则难以辨别,故包括幅值在内的时域特征在一定程度上能表征多余物的位置信息。分析多通道多余物信号的频域图可知,各通道采集多余物信号的频率分布在 20~120 kHz 的范围内,且距离信号源越近,信号在高频点处生成的能量就越大,功率谱密度也越大,说明信号的频率特性会因为传播距离的不同而产生差异。在时域上难以对通道 1 和 3 进行明显的区分,但在频域上却能将两者进行有效区分,所以同样可以利用多通道信号的频域特征预测多余物的位置。

在本文研究的多余物定位实验中,被测试件内部放置的多余物包括铁屑、金属丝、橡皮粒,对它们产生的多余物信号进行分析,最终选取了能较好反映它们位置信息的 7 个时域特征和 4 个频域特征。需要说明的是,本文选取的特征是在现有语音信号处理领域研究的基础上进行分析筛选的,可能存在本文未纳入考虑的更能反映位置信息的特征;对于其他类型多余物产生的多余物信号进行特征提取时,可以同样使用本文选取的时域和频域特征,也可以借鉴本文提出的声信号特征提取方法,提取适用的时域和频域特征。

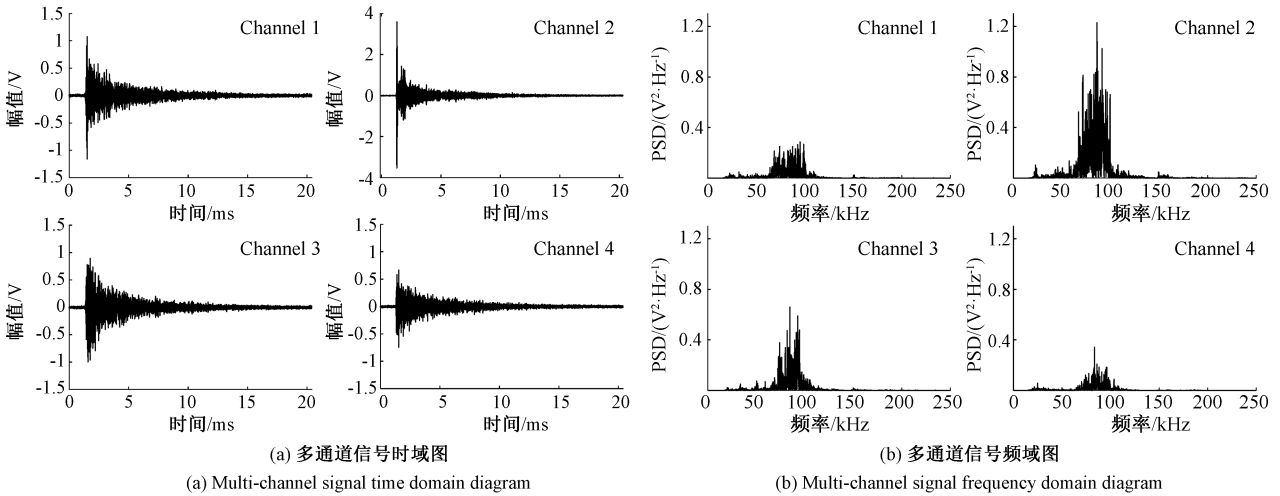


图 3 多通道同步采集的多余物信号

Fig. 3 Multi-channel simultaneous acquisition of loose particle signal

2.2 时域特征

在对多余物信号进行时域特征提取时,本文主要从时间特征、能量特征和脉冲过零率 3 方面进行考虑。

1) 时间特征

因为声信号在介质中传播距离的长短表征时间消耗的多少,所以时间特征是表征声发射源位置最重要的特征之一,具体包括上升时间、时间延迟和脉冲对称度等。目前互相关法是时延估计最常用的方法之一,即通过计算时域上两个信号的互相关函数来对比两者的相关程度^[17]。由于各通道采集的多余物信号来自同一声发射源,信号之间存在一定的关联性,故本文根据 Knapp 等^[18]提出的广义互相关法进行算法设计。

考虑任意 2 个通道的多余物信号 $x_1(n)$ 和 $x_2(n)$, 将其分别表示为:

$$x_1(n) = s(n - \tau_1) + n_1(n) \quad (13)$$

$$x_2(n) = s(n - \tau_2) + n_2(n) \quad (14)$$

式中: $s(n)$ 为多余物产生的位移信号; $n_1(n)$ 、 $n_2(n)$ 是声发射传感器收到的噪声信号; τ_1 、 τ_2 是多余物信号传播到声发射传感器的时差。

$x_1(n)$ 和 $x_2(n)$ 的互相关函数 $R_{12}(\tau)$ 为:

$$R_{12}(\tau) = E[x_1(n)x_2(n - \tau)] \quad (15)$$

将式(13)和(14)代入式(15)中,由于 $s(n)$ 、 $n_1(n)$ 和 $n_2(n)$ 互不相关,则:

$$R_{12}(\tau) = E[s(n - \tau_1)s(n - \tau_2 - \tau)] = R_s[\tau - (\tau_1 - \tau_2)] \quad (16)$$

广义互相关法在求两信号的互相关函数前,先对信号进行滤波,则互相关函数可由互功率谱表示为:

$$R_{12}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G_{12}(\omega) e^{j\omega\tau} d\omega \quad (17)$$

式中: $G_{12}(\omega)$ 是两信号的互功率谱。

通过对互相关函数进行加权处理,能够提高信号抑制噪声的能力,进而提升估计信号时延的精度。因此,广义互相关函数可以进一步表示为:

$$R_{12}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi_{12}(\omega) G_{12}(\omega) e^{j\omega\tau} d\omega \quad (18)$$

式中: $\psi_{12}(\omega)$ 是加权函数,在不同的应用场合,可选择不同的加权函数。在本文多余物定位的时延估计中,一般取 $\psi_{12}(\omega) = \frac{1}{|G_{12}(\omega)|}$, 即 PHAT 加权函数。

当 $\tau = \tau_1 - \tau_2$ 时,互相关函数 $R_{12}(\tau)$ 可取最大值,故有时间延迟:

$$\tau = \arg[\max(R_{12}(\tau))] \quad (19)$$

假设多余物信号的起始时刻为 T_b , 峰值时刻为 T_m , 结束时刻为 T_e , 则多余物信号的脉冲上升时间为:

$$t_r = T_m - T_b \quad (20)$$

多余物信号的脉冲对称度为:

$$\theta = \frac{t_r}{T_e - T_m} \times 100\% \quad (21)$$

2) 能量特征

信号的总能量是指在一段时间范围内对信号的平方求积分,对应离散信号的总能量可以表示为:

$$E = \sum_{i=1}^n x^2(i) \quad (22)$$

式中: $x(i)$ 表示信号的幅值,信号的总能量表征信号的强弱。所以信号的衰减幅值越大,能量就越小。

均方根值也被称为有效值,本文用来表示信号发送功率的能力。对于多余物信号,均方根电压同样是表征其信号状态的特征值,均方根速度(RMSV)定义为:

$$RMSV = \sqrt{\frac{\sum_{i=1}^n x^2(i)}{n}} = \sqrt{\frac{E}{n}} \quad (23)$$

3) 过零率

过零率是指一段信号符号变化的比率,本文通过多余物信号通过零点的次数进行量化。信号过零率的大小与其频率密切相关^[12],不同频率的多余物信号在传播过程中的衰减速率不同,导致各通道声发射传感器接收到的过零率也会有所差异,故把过零率作为多余物定位的特征。

$$zcr = \frac{1}{2} \sum_{i=2}^n | \text{sgn}[x(i)] - \text{sgn}[x(i-1)] | \quad (24)$$

式中:

$$\text{sgn}[x(i)] = \begin{cases} 1, & x(i) \geq 0 \\ -1, & x(i) < 0 \end{cases} \quad (25)$$

2.3 频域特征

设 $X(f)$ 为声发射信号经过傅里叶变换后的幅频谱,则频谱质心可以表示为:

$$S_f = \frac{\int_{f_1}^{f_2} f |X(f)| df}{\int_{f_1}^{f_2} |X(f)| df} \quad (26)$$

式中: f_1 和 f_2 为频谱的上下限频率。

对应离散信号的频谱质心,可使用式(27)进行计算。

$$S_f = \frac{\sum_{i=N_1}^{N_2} f_i |X(i)|}{\sum_{i=N_1}^{N_2} |X(i)|} \quad (27)$$

式中: N_1 和 N_2 分别表示离散信号频谱的频率范围。由此可见,通过频谱质心可以看出多余物信号的频率分布情况。

频谱均方差表示对幅频谱 $X(f)$ 求取标准差,其公式定义为:

$$\sigma_f = \sqrt{\frac{\sum_{i=N_1}^{N_2} (X(i) - \bar{X})^2}{N_2 - N_1}} \quad (28)$$

式中: $\bar{X} = \frac{1}{N_2 - N_1} \sum_{i=N_1}^{N_2} X(i)$ 是频域上各分量的均值。频谱均方差 σ_f 表示信号在频域上能量变化的大小,其值越大,信号能量越大,频率变化越剧烈。

通过在能量谱上统计信号的特征值同样可以对其进行频域分析,包括重心频率(FC)、均方频率(MSF)、均方根频率(RMSF)、频率方差(VF)和频率标准差(RVF),上述特征的定义分别如式(29)~(31)所示。

$$FC = \frac{\sum_{i=N_1}^{N_2} f_i S(i)}{\sum_{i=N_1}^{N_2} S(i)} \quad (29)$$

$$MSF = \frac{\sum_{i=N_1}^{N_2} f_i^2 S(i)}{\sum_{i=N_1}^{N_2} S(i)} \quad (30)$$

$$VF = MSF - FC \quad (31)$$

其中, $S(i) = |X(i)|^2$, 表示信号的能量谱,分别对 MSF 和 VF 求均方根得到 $RMSF$ 和 RVF , 本文同样选择这两个频域特征用于多余物定位研究。

2.4 特征分析

在机器学习中,组成结构良好的数据集与性能优良的算法同样重要。数据特性是影响分类器模型预测性能的关键因素,往往选择的特征越好,分类器模型性能就越好,最终对应模型的预测准确度就越高。本文从信息增益和定位准确度两个方面对提取的时域和频域特征进行分析。

1) 信息增益

信息增益是指在得到特征 X 的信息后,能减少类 Y 的信息不确定性程度的度量^[12]。特征 X 对训练集 D 的信息增益 $g(D, X)$ 可表示为:数据集 D 的经验熵 $H(D)$ 减去特征 X 给定条件下 D 的经验条件熵 $H(D|X)$, 即:

$$g(D, X) = H(D) - H(D|X) \quad (32)$$

通过信息增益来分析多余物信号的时域和频域特征,计算得到每个特征的重要度,如图4(a)所示,是在实验中计算得到的各个特征的重要度,图中横坐标上每4个特征值为一组,对应各特征在4个通道上的取值。同样可以从图4(a)看出,上升时间 $risTime$ 的重要度最大,过零率 $zerorate$ 的重要度最小,均方根频率 $speStdDev$ 和均方根电压 $RMSV$ 在其四通道的某个通道上的重要度较低,相比之下其他3个通道的特征值较高。总而言之,所提取的7个时域和4个频域特征有一定区分多余物位置的能力。

2) 定位准确度

信息增益更多的是考虑特征与特征、特征与类别之间的关系,忽略了分类器模型的性能。利用不同特征数据集的子集来训练分类器模型,可以得到对应各特征的分类器模型的定位准确度,进而可以直观的得到各特征分类性能的好坏。图4(b)所示为不同特征数据集训练的 kNN 分类器模型用于多余物定位的准确度对比结果。其中,“All”指综合所有特征训练的分类器模型的定位准确度。从图4(b)可以看出,在使用单一特征进行分类模

型训练时,上升时间 risTime 的定位准确度最高,过零率 zerorate 的定位准确度最差,分析结果与信息增益一致。还可以得出时域特征比频域特征更能表征多余物的位置

信息,因为信号的衰减更多体现在能量减少上。总体来说,使用所有的 11 个特征训练的分类器模型的定位准确度最高。

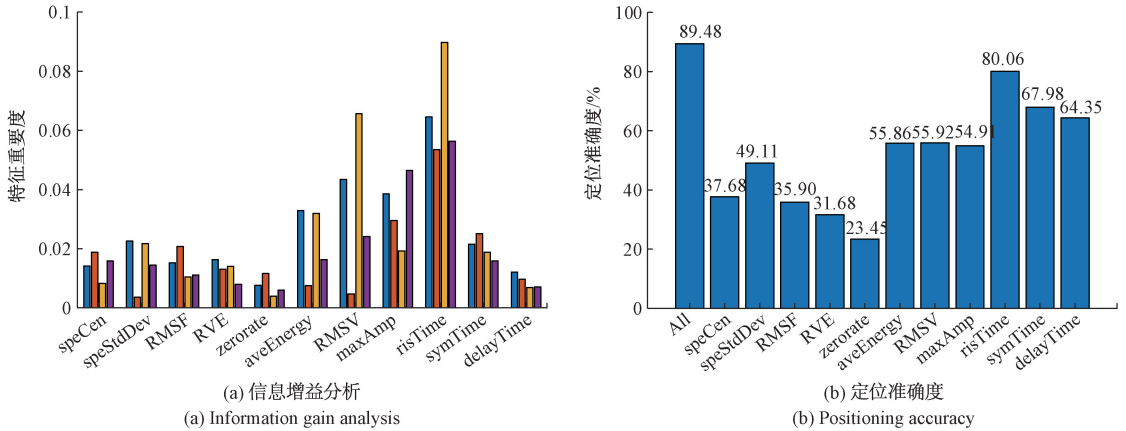


图 4 特征的信息增益与定位准确度分析效果
Fig. 4 Analysis effect diagram of feature information gain and classification accuracy

3 实验介绍

3.1 实验数据

本实验采用上文搭建的多余物定位实验系统,结合设计的被测件模型,通过对多通道多余物信号的特征提取,形成用于多余物定位的 kNN 算法训练学习的数据集。在平面定位实验中,分别在 16 个平面区域放置一些多余物颗粒,所以经过 16 次多余物检测后,得到 16 个对应平面区域内多余物信号特征的数据集,具体描述如表 1 所示。在空间定位实验中,分别在 8 个空间区域放置一些多余物颗粒,所以经过 8 次多余物检测后,得到 8 个对应空间区域内多余物信号特征的数据集,具体描述如表 2 所示。

表 1 平面定位数据具体描述

Table 1 Detailed description of plane positioning data

区域号	数据总数	特征数	区域号	数据总数	特征数
1	4 041	11	9	4 583	11
2	3 333	11	10	5 443	11
3	5 167	11	11	5 919	11
4	4 416	11	12	5 159	11
5	4 855	11	13	3 821	11
6	5 822	11	14	5 528	11
7	4 904	11	15	5 793	11
8	4 475	11	16	3 525	11

在 Python 程序工程中依次读取平面定位实验中的 16 个平面区域对应的特征数据集,最终形成用于多余物平面定位的 kNN 算法训练学习的数据集。同样依次读

取空间定位实验的 8 个空间区域对应的特征数据集,形成用于多余物空间定位的 kNN 算法训练学习的数据集。对两个数据集中缺失部分特征的数据进行舍弃处理,最终得到有效的平面定位数据 76 765 个,空间定位数据 81 768 个。把处理后的两个数据集分别按 3:1 的比例分成训练集和测试集,训练集用于建立 kNN 分类器模型和对模型进行调参优化,测试集用于检测最优分类器模型的定位性能,最终得到泛化能力良好且定位精度较高的 kNN 分类器模型。

表 2 空间定位数据具体描述

Table 2 Detailed description of spatial positioning data

区域号	数据总数	特征数	区域号	数据总数	特征数
1	10 828	11	5	10 151	11
2	11 123	11	6	8 772	11
3	10 410	11	7	10 242	11
4	9 776	11	8	10 483	11

因此,在平面定位实验中,57 573 个数据用作训练,19 192 个数据用作测试;同样在空间定位实验中,61 326 个数据用作训练,20 442 个数据用作测试。

3.2 评估指标

为了直观评估训练好的 kNN 分类器模型的定位性能,需要建立量化模型泛化能力的评估指标,本文主要选取了定位精度作为性能度量,评价分类模型在测试集上的定位效果。

$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ 为假定数据测试集,其中 y_i 是特征数据 x_i 对应的多余物的真实位置, $f(x_i)$ 是应用 kNN 分类器 f 预测的多余物位置。多余物

的定位精度可表示为预测位置与真实位置一致的样本数占整体样本数的比例,即:

$$acc(f;D) = \frac{1}{m} \sum_{i=1}^m I(f(x_i) = y_i) \quad (33)$$

式中: I 是指示函数,当 $f(x_i) = y_i$ 时, $I(f(x_i) = y_i) = 1$ 。

3.3 实验结果

本文基于 Scikit-Learn 类库完成对 kNN 算法的参数优化,包括 k 值选择、距离度量与权重设置。通过调用 Scikit-Learn 的 KNeighborsClassifier 类函数,优化设计其内部参数值,利用定位性能评估指标来完成一个 kNN 分类模型的构建。其中,影响多余物定位精度的 KNeighborsClassifier 类参数包括 $n_neighbors$ 、 $metric$ 和 $weights$ 。默认情况下, $n_neighbors$ 、 $metric$ 和 $weights$ 的值分别设定为 5、“euclidean”和“uniform”,分别表示 k 值取 5、欧氏距离度量和相同权重。此时,分别加载平面和空间定位数据集进行训练,得到 kNN 分类器模型进行测试,定位效果如表 3 所示。

本文采用控制变量法对 kNN 算法寻优过程进行设计,即在保持其他两个参数不变的情况下,控制其中一个

表 3 默认情况下 kNN 分类模型的定位效果

Table 3 The positioning effect of the kNN classification model by default

定位类别	测试样本数	预测正确 样本数	预测错误 样本数	准确率/%
平面定位	19 192	14 916	4 276	77.72
空间定位	20 442	14 937	5 505	73.07

参数选取不同数值进行分析比较,得出当前参数的最优取值,因此下文分别进行 3 次寻优的实验。

1) 实验 1:对 k 值进行不同的选择

本文基于 Scikit-Learn 类库的 KNeighborsClassifier 类函数, $n_neighbors$ 是表征 k 值选择的自变量,分别代入平面和空间定位数据集用于当前参数的训练和测试。通过给定 $n_neighbors$ 的取值范围或取值集合,在软件程序中对选取不同 k 值的 kNN 分类器在训练集和测试集的分类性能进行评估。考虑到 k 值选取过大训练效率低下且容易造成过拟合,选取过小会产生误差,所以本文最终选取 k 值的范围为 3~10,对应输出结果如图 5 所示。

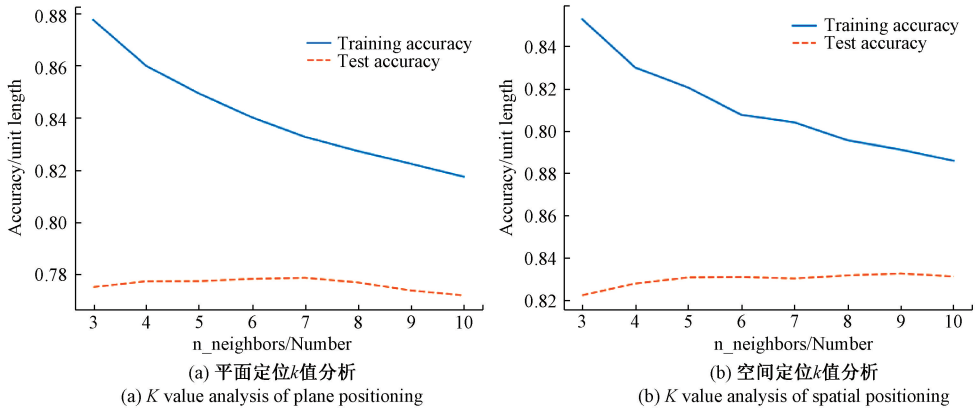


图 5 不同 k 值的性能评估

Fig. 5 Performance evaluation of different k values

从图 5 可以看出,兼顾训练与测试准确率,平面定位实验中 k 值选取 3、5、7 较为合适,空间定位实验中 k 值取 4、5、9 较为合适,将上述选取的 k 值分别代入 KNeighborsClassifier 类中进行训练,得到表 4 所示的不同 k 值的 kNN 分类模型定位效果。可以发现,在其他两个参数保持默认不变的情况下,平面定位中 k 值取 7、空间定位中 k 值取 9 的定位效果最好。

2) 实验 2: 选择不同的距离函数

本文 $metric$ 是表征距离度量的自变量,代入平面和空间定位数据集用于训练和测试,用不同的距离函数对训练集和测试集的性能进行评估,在 KNeighborsClassifier 类函数中,可选择的距离函数包括欧氏距离、曼哈顿距离和切比雪夫距离。

在平面与空间定位实验中, $metric$ 取值的范围为“euclidean”、“manhattan”和“chebyshev”,分别代入 KNeighborsClassifier 类中进行训练,得到如表 5 所示的不同距离函数的 kNN 分类器模型定位效果。可以发现,在其他两个参数保持默认不变的情况下,平面与空间定位中选择曼哈顿距离函数的定位效果最好。

3) 实验 3: 设置不同的权重

本文 $weights$ 是表征权重设置的自变量,代入平面和空间定位数据集用于训练和测试,设置不同的“投票”权重对训练集和测试集的性能进行评估,在 KNeighborsClassifier 类函数中,对权重的设置包括均等权重和距离反比权重两种。在平面与空间定位实验中 $weights$ 依次取“uniform”和“distance”,分别代入

表 4 不同 k 值的定位效果Table 4 Positioning effect of different k values

定位类别	$n_neighbors$	测试样本数	预测正确 样本数	准确率/%
平面定位	3	19 192	14 872	77.49
	5	19 192	14 916	77.72
	7	19 192	14 941	77.85
空间定位	4	20 442	14 878	72.78
	5	20 442	14 937	73.07
	9	20 442	14 975	73.26

表 5 不同距离函数的定位效果

Table 5 Positioning effect of different distance functions

定位类别	$metric$	测试样本数	预测正确 样本数	准确率/%
平面定位	euclidean	19 192	14 916	77.72
	manhattan	19 192	15 369	80.08
	chebyshev	19 192	14 043	73.17
空间定位	euclidean	20 442	14 937	73.07
	manhattan	20 442	15 567	76.15
	chebyshev	20 442	13 907	68.03

KNeighborsClassifier 类函数中进行训练,得到如表 6 所示的不同权重设置的 kNN 分类模型定位效果。可以发现,在其他两个参数保持默认不变的情况下,平面与空间定位中设置距离反比权重的定位效果最好。

表 6 不同权重设置的定位效果

Table 6 Positioning effect of different weight settings

定位类别	$weights$	测试样本数	预测正确 样本数	准确率/%
平面定位	uniform	19 192	14 916	77.72
	distance	19 192	15 177	79.08
空间定位	uniform	20 442	14 937	73.07
	distance	20 442	15 195	74.33

上述过程是采用控制变量法对单独某个参数进行寻优设计和分析比较,得出了各自的最优取值。但本文基于 KNeighborsClassifier 类函数训练的 kNN 分类器模型,是对 3 个参数的综合考虑,得出三者之间的最优组合,以取得最佳的多余物定位效果。因此,本文采用网格搜索法对基于 k 值选择、距离度量和权重设置的 kNN 分类器模型进行综合的参数寻优,即在软件程序中对 3 个参数各自的取值范围,采用排列组合的方式对范围内的取值进行组合,得到多个组合的定位性能,在此基础上得到定位性能最优的参数取值组合。最终,经过网格搜索法得到的平面定位参数最优组合为 $n_neighbors$ 取值 5、 $metric$ 选取曼哈顿距离、 $weights$ 设置为距离反比权重;平面定位参数最优组合为 $n_neighbors$ 取值 9、 $metric$ 选取曼哈顿

距离、 $weights$ 设置为距离反比权重。对应的定位效果如表 7 所示。

表 7 参数寻优 kNN 分类模型的定位效果

Table 7 Positioning effect of kNN classification model based on parameter optimization

定位类别	测试样本数	预测正确 样本数	预测错误 样本数	准确率/%
平面定位	19 192	15 581	3 611	81.18
空间定位	20 442	15 829	4 613	77.43

在 Scikit-Learn 类库之外,本文还通过独立 Python 编程的方式引入马氏距离进行 kNN 算法的优化,同样通过网格搜索法,进行本文的第 4 次实验。

4) 实验 4: 引入马氏距离

在本实验中,将 $metric$ 恒定设置为马氏距离函数,通过网格搜索法对平面与空间定位中的其他两个参数,即 $n_neighbors$ 和 $weights$ 进行寻优,所使用的参数取值范围与上文网格搜索法使用的相一致。

最终,在 $metric$ 保持恒定不变情况下,得到的平面定位参数最优组合为 $n_neighbors$ 取值 7、 $weights$ 设置为距离反比权重;平面定位参数最优组合为 $n_neighbors$ 取值 9、 $weights$ 设置为距离反比权重。对应的定位效果如表 8 所示。

表 8 马氏距离 kNN 分类模型的定位效果

Table 8 Positioning effect of kNN classification model based on Mahalanobis distance

定位类别	测试样本数	预测正确 样本数	预测错误 样本数	准确率/%
平面定位	19 192	15 581	3 611	79.79
空间定位	20 442	15 829	4 613	79.34

4 结 论

本文提出了使用参数优化的 kNN 算法对密封电子设备内部的多余物进行定位研究,搭建了多余物定位实验系统,设计了密封电子设备的试件模型,通过提取多通道多余物信号的时域和频域特征,建立用于 kNN 算法学习的数据集。最终通过实验验证,得出使用参数优化的 kNN 算法建立的分类器模型,能够有效地提升多余物定位的性能。当 kNN 选取 k 值为 5、曼哈顿距离度量、设置距离反比权重时,用于平面定位的 kNN 分类模型的定位精度最高,由默认的 77.72% 提升到 81.18%;当 kNN 选取 k 值为 9、马氏距离度量、设置距离反比权重时,用于空间定位的 kNN 分类模型的定位精度最高,由默认的

73.07%提升到79.34%。为了进一步提高多余物定位的精度,下一步将考虑使用多个kNN分类器集成的算法,或采用现有的由多个决策树集成的随机森林算法,训练泛化性能良好的集成分类器模型,继续尝试对集成分类器模型进行参数寻优研究,与本文的单个分类器模型综合比较分析。

参考文献

- [1] 熊涛. 航天器总装多余物控制方法探讨[J]. 航天器环境工程, 2006, 23(5): 277-281.
XIONG T. Discussion on the control method of spacecraft's surplus [J]. Spacecraft Environment Engineering, 2006, 23(5): 277-281.
- [2] 许龙飞, 麻高占, 马利, 等. 运载火箭总装气密试验多余物控制方法分析[J]. 质量与可靠性, 2017(4): 6-9.
XU L F, MA G Z, MA L, et al. Analysis of surplus control method for carrier rocket assembly air tight test[J]. Quality and Reliability, 2017(4): 6-9.
- [3] 梁晓雯, 蒋爱平, 王国涛, 等. 参数优化决策树算法的密封继电器多余物信号识别技术[J]. 电子测量与仪器学报, 2020, 34(1): 178-185.
LIANG X W, JIANG A I P, WANG G T, et al. Sealed relay loose particle signal recognition technology based on decision tree algorithm of parameter optimization [J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(1): 178-185.
- [4] LE DUFF A, HAMDI S E, PLANTIER G, et al. Time delay estimation for acoustic source location by means of short-time cross-correlation [C]. Sensors, 2011 IEEE, 2011: 1885-1888.
- [5] LYMPERTOS E, DERMATAS E. Acoustic emission source location using multi-frequency arrival times [C]. 2006 IEEE Mediterranean Electrotechnical Conference, 2006: 389-392.
- [6] 行鸿彦, 杨旭, 张金玉. 基于四元传声器阵列的声源全方位定位算法[J]. 仪器仪表学报, 2018, 39(11): 43-50.
XING H Y, YANG X, ZHANG J Y. Omnidirectional sound source localization algorithm based on four-element microphone array [J]. Chinese Journal of Scientific Instrument, 2018, 39(11): 43-50.
- [7] DRIS E Y, DRAI R, BENAMMAR A, et al. Acoustic emission source localization in plate-like structure [C]. 2017 European Conference on Electrical Engineering and Computer Science (EECS), 2017: 193-197.
- [8] NIVESRANGSAN P, STEEL J A, REUBEN R L. Source location of acoustic emission in diesel engines [J]. Mechanical Systems and Signal Processing, 2007, 21(2): 1103-1114.
- [9] 丁丹丹, 朱永晓, 陈伟, 等. 导弹舱体多余物检测系统设计[J]. 宇航计测技术, 2019, 39(S1): 81-85.
DING D D, ZHU Y X, CHEN W, et al. Design of missile cabin surplus detection system [J]. Journal of Astronautic Metrology and Measurement, 2019, 39(S1): 81-85.
- [10] YANG G, JUNG J, SEOL D. Sound-source localization system based on neural network for mobile robots [C]. IEEE International Joint Conference on Neural Networks, 2008: 3126-3130.
- [11] 李卓. 基于声发射的载人密封舱撞击源定位技术研究[D]. 哈尔滨: 哈尔滨工业大学, 2016.
LI ZH. Research on location technology of impact source of manned airtight module based on acoustic emission [D]. Harbin: Harbin Institute of Technology, 2016.
- [12] 郜雷阵. 密封电子设备设备活动多余物定位方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2019.
GAO L ZH. Research on loose particle location method for sealed electronic devices [D]. Harbin: Harbin Institute of Technology, 2019: 225-229.
- [13] 张鸿祥. 基于人工智能的认知无线电频谱感知算法研究[D]. 南京: 东南大学, 2019.
ZHANG H X. Research on cognitive radio spectrum sensing algorithm based on artificial intelligence [D]. Nanjing: Southeast University, 2019.
- [14] 王印松, 马婧妍. 一种基于改进马氏距离的MIMO系统性能评价方法[J]. 中国测试, 2020, 46(6): 101-107.
WANG Y S, MA P Y. A Performance evaluation method of MIMO system based on improved mahalanobis distance [J]. China Measurement & Testing Technology, 2020, 46(6): 101-107.
- [15] 王世成, 王国涛, 翟国富, 等. 密封电子元器件微粒碰撞噪声自动检测系统的研究[J]. 计算机与数字工程, 2010, 38(9): 1-4, 12.
WANG SH CH, WANG G T, ZHAI G F, et al. Research on automatic detection system for particle impact noise of sealed electronic components [J]. Computer and Digital Engineering, 2010, 38(9): 1-4, 12.
- [16] 刘松平, CONNAN M, 陈积愈. 模态声发射检测技术[J]. 无损检测, 2000, 22(1): 38-41.
LIU S P, CONNAN M, CHEN J Y. Modal acoustic emission detection technology [J]. Nondestructive Testing, 2000, 22(1): 38-41.

- [17] YEGNANARAYANA B, RAMANI D. Processing of reverberant speech for time-delay estimation [J]. IEEE Trans. on Speech and Audio Processing, 2005, 13(6): 1110-1118.
- [18] KNAPP C H, CARTER G C. The generalized correlation method for estimation of time delay [J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1976, 24(4): 320-327.

作者简介



孙志刚, 2019 年于江苏理工学院获得学士学位, 现为黑龙江大学信息与通信工程专业硕士研究生, 主要研究方向为密封电子元器件和电子设备多余物定位技术。
E-mail: 2191313@s.hlj.u.edu.cn

Sun Zhigang received his B. Sc. degree

from Jiangsu University of Technology in 2019. Now he is a M. Sc. candidate in Information and Communication Engineering at Heilongjiang University. His main research interest includes sealed electronic components and electronic equipment loose particle positioning technology.



蒋爱平(通信作者), 2007 年于哈尔滨工业大学获得博士学位, 现为黑龙江大学教授, 主要研究方向为数字图像处理与计算机视觉。

E-mail: 1988014@hlju.edu.cn

Jiang Aiping (Corresponding author) received her Ph. D. degree from Harbin Institute of Technology in 2007. Now she is a professor at Heilongjiang University. Her main research interests include digital image processing and computer vision.