

面向距离判定的低功耗蓝牙信号强度特征研究

汪文博, 骆冰清

(南京邮电大学 计算机学院、软件学院、网络空间安全学院, 江苏 南京 210023)

摘要: 低功耗蓝牙因其能耗和成本低, 且设备兼容性好等特点, 而在智能设备中的应用愈加广泛, 特别是在距离判定方面具有重要作用。研究了低功耗蓝牙信号强度特征选择与特征分组在支持向量机、随机森林和卷积神经网络3种不同模型中对距离判定精度的影响, 通过使用皮尔逊相关系数对从移动设备信号强度数据中所选取的11个特征进行分组, 评估了在不同机器学习模型下引入单一特征时的性能增益, 由此定义了显著特征和非显著特征, 同时进一步研究了在组内多特征和组间多特征情况下对距离判定精度的影响。结果表明, 不同特征组合在各模型中的表现存在明显差异, 多个显著特征的引入能有效提高距离判定的准确性, 为低功耗蓝牙信号强度特征的选取和应用提供了一定指导和参考。

关键词: 低功耗蓝牙; 信号强度特征; 距离判定; 支持向量机; 随机森林; 卷积神经网络

DOI: 10.11907/rjdk.241591

开放科学(资源服务)标识码(OSID):

中图分类号: TP393; TP181

文献标识码: A

文章编号: 1672-7800(2025)008-0072-10



Research on Bluetooth Low Energy Signal Strength Feature for Distance Measurement

WANG Wenbo, LUO Bingqing

(School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

Abstract: Bluetooth Low Energy is increasingly used in smart devices due to its low energy consumption, low cost and wide range of device compatibility, especially in distance measurement. This paper studies the influence of Bluetooth Low Energy signal strength feature selection and feature grouping on the accuracy of distance measurement in three different models: support vector machine, random forest and convolutional neural network. By using the Pearson correlation coefficient to group 11 features selected from the signal strength data of mobile devices, the performance gain when a single feature is introduced under different machine learning models is evaluated, and key features and minor features are defined. At the same time, the influence on the accuracy of distance measurement under the conditions of multiple features within the group and multiple features between groups is further studied. The results show that there are obvious differences in the performance of different feature combinations in each model, and the introduction of multiple key features can effectively improve the accuracy of distance determination. This paper provides certain guidance and reference for the selection and application of low-power Bluetooth signal strength features.

Key Words: bluetooth low energy; signal strength feature; distance measurement; support vector machine; random forest; convolutional neural network

0 引言

随着第五代移动通信技术和物联网的迅速发展, 在当今智能设备普及的时代, 基于低功耗蓝牙(Low Energy Bluetooth, BLE)的技术正逐步成为关键驱动力^[1-2]。特别

是, 针对基于低功耗蓝牙的距离判定研究备受瞩目, 其对于实现智能环境感知、位置识别和距离测量等具有重要意义^[3]。

相比于到达时间定位法^[4](Time of Arrival, TOA)、时间差定位法^[5](Time Difference of Arrival, TDOA)、到达角定位法^[6](Angle of Arrival, AOA)、飞行时间法^[7](Time of

收稿日期: 2024-08-08

扫描二维码阅读全文:



作者简介: 汪文博(1998-), 男, 南京邮电大学计算机学院、软件学院、网络空间安全学院硕士研究生, 研究方向为室内定位技术、计算机应用技术; 骆冰清(1987-), 女, 博士, 南京邮电大学计算机学院、软件学院、网络空间安全学院副教授、硕士生导师, 研究方向为无线网络通信、区块链、物联网。本文通讯作者: 骆冰清。

Flight, TOF), 基于信号强度指示^[8] (Received Signal Strength Indicator, RSSI) 的距离估计方法通过信号强度推断设备之间的距离, 因其简单易实现且成本低的特点而常作为距离估计方法。RSSI 虽然可以直接用于距离判定, 但由于环境的复杂性和信号的传播特性, 其在复杂的室内环境(如机场、超市等)中容易发生衰减、反射、多径效应等现象^[9]。同时, 由于信号的多样性和复杂性, 信号强度不仅与距离有关, 还与发射功率、接收器的灵敏度和天线方向等因素有关。简单的 RSSI 值不能全面反映这些复杂因素的影响, 导致直接使用原始 RSSI 数据进行距离判定会影响精度^[10]。文献[11]的研究表明, 通过特征提取可以将这些因素加以综合考虑, 过滤原始 RSSI 数据中的噪声和冗余信息, 提取出更加稳定和鲁棒的信号特征, 这些特征也能更准确地反映信号传播特性, 从而减小环境干扰等因素对距离判定的影响。虽然特征提取增加了数据维度, 提升了计算复杂度和能耗, 但这是为了换取更高的判定精度和更好的稳定性。在现代硬件设备性能不断提升的情况下, 适度增加计算复杂度以换取更好的实际应用效果是可取的一种权衡。研究 RSSI 数据特征提取和特征分组对距离判定的影响, 能够有效减少信号强度波动带来的精度误差。其具体方法可分为利用机器学习和深度学习技术以学习 RSSI 数据样本进行特征提取和处理 RSSI 数据样本进行特征选取两方面。

一方面, 文献[12–15]利用机器学习和深度学习技术学习 RSSI 数据样本进行特征提取以应用于距离判定任务。文献[12]通过移动行走过程中采集的 RSSI 指纹建立对应的移动采集指纹库, 根据移动采集指纹特征构建特征向量, 建立基于自适应压缩感知算法的指纹匹配室内定位模型。文献[13]根据 RSSI 向量大小生成排序特征码以用于形成指纹库, 并提出一种基于排序特征匹配和距离加权的蓝牙定位算法。文献[14]提出一种基于卷积神经网络且具有 WiFi 指纹识别功能的室内定位系统, 通过使用堆叠的自编码器对 RSSI 数据样本进行特征提取, 提升室内定位精准度。文献[15]提出一种利用 BLE 信号的到达角和 RSSI 数据进行加权指纹特征匹配的室内定位算法。

另一方面, 文献[16–18]通过处理 RSSI 数据样本进行特征选取以应用于距离判定任务。文献[16]通过使用 RSSI 测量和 CSI 测量的 9 种不同类型的统计特征生成特征向量, 并形成指纹, 从而提出一种多模态学习方法以提升定位性能。研究表明, 低功耗蓝牙信号强度的统计特征对提高距离判定精度起到了一定作用。文献[17]使用 3 种不同的算法对 BLE 信号的 13 个空间、时间、频率和一般统计特征进行了分类, 以实现 COVID-19 接触追踪系统准确性和可靠性。该研究针对 BLE 信号的 13 个特征进行分类, 分析了基于信号特征的空间、时间和频率 3 个维度对距离判定的影响。文献[18]依据位置度量、离散度量和形状度量, 针对从 BLE 信号强度中选取的 15 个特征进行分组, 以

用于机器学习模型的接近估计, 研究了在 3 种不同机器学习模型中引入不同组之间更多特征时的性能增益。此外, 针对 BLE 信号强度特征进行了分组, 并通过引入不同特征在不同机器学习模型中的性能进行了比较分析, 但分析结果缺少对具体选取特征种类的指导。

对 RSSI 数据进行处理分析以预测判定任务的机器学习模型和深度学习模型多种多样。文献[19]提出一种基于 Wi-Fi 信号的室内定位系统, 使用支持向量机(Support Vector Machine, SVM)在处理 RSSI 数据方面表现出色。文献[20]提出一种改进的基于 RSSI 的室内定位方法, 使用随机森林(Random Forest, RF)在特征提取和处理非线性 RSSI 信号方面效果显著, 具有较强鲁棒性和抗噪能力。文献[21]的研究表明, k-近邻算法(k-Nearest Neighbor, k-NN)在基于 RSSI 蓝牙指纹的定位中表现良好, 但 k-NN 算法存在计算复杂度高、内存占用大、对噪声敏感等缺点。文献[22]提出一种基于卷积神经网络(Convolutional Neural Network, CNN)的双分支多尺度特征提取网络, 通过其非线性激活函数和多层结构, 可以有效捕捉和建模 RSSI 数据与距离之间的非线性关系。文献[23]提出一种基于混合 CNN 模型的三相逆变器的自适应特征提取方法, CNN 通过卷积操作能够有效捕捉复杂环境中变化的 RSSI 信号局部模式和特征, 以提升特征提取准确性和鲁棒性, 并提高模型适应能力和预测精度。文献[24]提出一种基于 RSSI-LSTM 的 CR-IOT 自私攻击检测方法, 长短期记忆网络(Long Short-Term Memory, LSTM)可以通过学习 RSSI 数据的时间依赖性进行预测, 但 LSTM 也存在计算复杂度高、训练时间长、参数调整复杂等缺点。

综上所述, 为研究信号强度特征选择和特征分组对距离判定精度产生的影响, 并比较不同特征分别在机器学习模型和深度学习模型中对距离判定精度的影响, 本文选取 SVM、RF 和 CNN 进行实验研究。使用皮尔逊相关系数对从移动设备信号强度数据中所选取的 11 个特征进行分组, 评估了在不同模型下引入单一特征时的性能增益, 由此定义了显著特征和非显著特征, 同时进一步研究了其在组内多特征和组间多特征情况下对距离判定精度的影响。

1 研究方法

1.1 技术路径

本文以提高距离判定精度为主要目标, 研究了不同低功耗蓝牙信号强度特征在不同模型中对距离判定精度的影响。如图 1 所示, 首先从移动手机设备获取 RSSI 作为数据集, 接着从低功耗蓝牙信号强度中选取了 11 个特征, 再利用皮尔逊相关系数对所选取的 11 个特征按相关性进行分组, 最后使用多特征分组化方法, 分别探究“组内单一特征”“组内多特征”“组间多特征”在不同模型中对距离判定精度的影响。

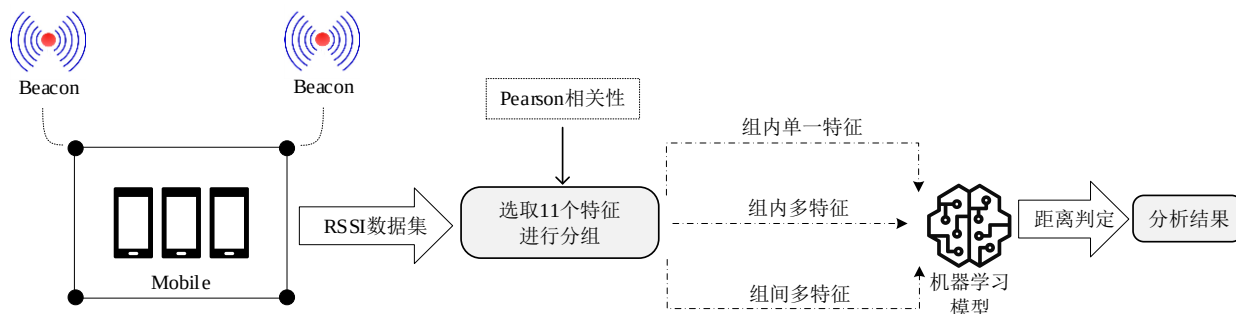


Fig. 1 Technology path in this paper

图1 本文研究技术路径

1.2 特征定义

本文基于文献[16-18]的实验验证,从统计全面性和冗余复杂性两方面加以综合考虑,从低功耗蓝牙信号强度中选取了11个特征,在不同机器学习模型中展开针对距离判定精度的研究。这些特征包括:平均值($mean_{id}$)、中位数($median_{id}$)、最小值(min_{id})、最大值(max_{id})、方差($variance_{id}$)、第一四分位数($one\ quarter_{id}$)、第三四分位数($three\ quarter_{id}$)、L1到平均值的距离($L1\ to\ mean_{id}$)、L1到中位数的距离($L1\ to\ median_{id}$)、峰度($kurt_{id}$)和偏度($skew_{id}$)。所选取的11个特征能够全面反映RSSI信号的统计特性,包括位置参数(如平均值、中位数)、离散程度(如方差、四分位数)和形态参数(如峰度、偏度)。将这些多样化的特征应用于距离判定任务,旨在更全面地考虑信号传播的复杂性,从而提高距离判定准确性和可靠性。

1.3 特征分组

Pearson相关性,又称为积差相关,用于衡量两个变量之间的线性关系强度和方向^[25]。它计算简单且高效,能够提供一个清晰直观的数值,有助于特征之间相关性的快速判断。在RSSI信号强度数据中,许多特征之间存在线性关系,考虑到RSSI数据特性的适用性,故在特征分组过程中采用Pearson相关性分析方法,以探究各低功耗蓝牙信号强度特征之间的关联性,从而将高相关性特征进行分组,以减少冗余信息、简化模型并减少计算复杂度。如图2所示,通过皮尔逊相关系数可以清晰了解某两个特征之间的相关程度。基于各特征之间的相关程度,将相关性较高的特征分到同一组,如表1所示。“平均值、中位数、第一四分位数、第三四分位数、最小值、最大值”为组一,“方差、L1到平均值的距离、L1到中位数的距离”为组二,“偏度、峰度”为组三。通过特征分组,旨在进一步研究组内相关性较高的多特征和组间相关性较低的多特征在不同模型中对距离判定精度的影响。

1.4 距离判定方法

在对RSSI数据进行特征选取后,获得了3组共11个特征。本文通过不同的特征组合和支持向量机、随机森林、卷积神经网络3种不同的模型运算,研究了不同特征对距离判定预测值精度的影响。下文将介绍3种不同模型在距离判定中的应用方法。

1.4.1 支持向量机模型

支持向量机是一种广泛应用于分类和回归的监督学习模型^[26]。其应用于距离判定任务中的核心思想是使用带标签的训练数据集训练SVM模型,从而找到一个最佳的超平面,将不同类别的数据点分隔开来,使得不同类别的数据点在该超平面上具有最大间隔。如式(1)所示,对于每个数据点,计算其到超平面的距离Distance,从而判断数据点的相似性或离群程度。距离越小,表明数据点越接近决策边界;距离越大,表明数据点越远离决策边界。

$$Distance = \frac{|w \cdot x + b|}{\|w\|} \quad (1)$$

其中, $w \cdot x + b$ 是数据点到超平面的垂直距离, $\|w\|$ 是超平面法向量的范数。

首先,读取数据集并将其划分成特征变量和预测值,并实例化一个K折交叉验证器。将数据集分成训练集和测试集后,定义SVM模型,并对测试集和训练集进行预测,最终计算并输出模型的评估指标(均方误差和绝对误差)。鉴于实验的预测值仅为单一变量——实际距离 d ,故本文采用线性核函数('linear')进行训练,并使用10折交叉验证策略。同时,通过调整正则化参数C的值进行模型优化,最终模型性能通过训练集和测试集的评估指标进行度量。为了减少分类错误并提高训练和测试时的准确率,本文将惩罚参数设定为10。

1.4.2 随机森林模型

随机森林模型是一种集成学习方法,通过构建多个决策树并将其结果进行平均或投票以提高预测性能^[27]。其应用于距离判定任务中的核心思想是以原始数据为基础进行多次随机抽样,训练多棵决策树,分析输入数据点在各树中的路径及其叶节点的分布情况以进行相似性或异常性判定。若两个数据点在大多数树中经过相似的路径或落在相同或相近的叶节点中,则它们的距离较近,表明它们更相似;相反,若距离越大,表明数据点可能是异常点或离群点。

与SVM模型相似,首先读取数据集,将其划分成特征变量和预测值,并实例化一个K折交叉验证器。将数据集分成训练集和测试集后,定义一个随机森林回归模型,并通过交叉验证和网格搜索技术进行参数调优。最终,对训

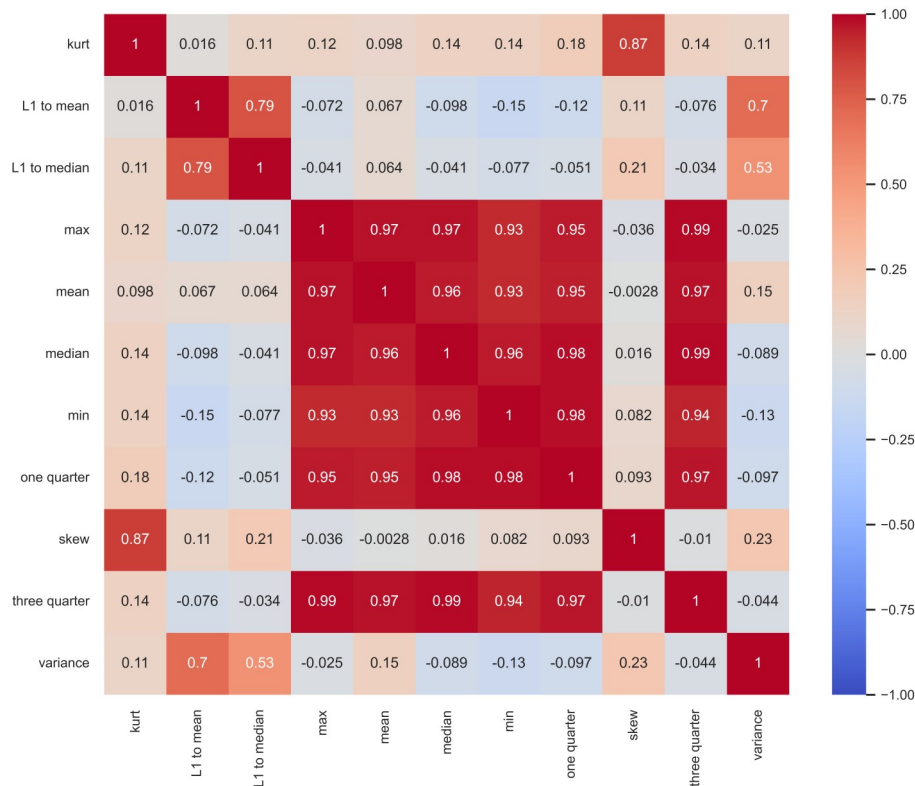


Fig. 2 Correlation of the 11 selected features

图2 选取的11个特征相关性

Table 1 Feature grouping results

表1 特征分组结果

组别	特征
组一	平均值(mean)
	中位数(median)
	第一四分位数(one quarter)
	第三四分位数(three quarter)
	最小值(min)
	最大值(max)
组二	方差(variance)
	L1到平均值的距离(L1 to mean)
	L1到中位数的距离(L1 to median)
组三	偏度(skew)
	峰度(kurt)

练集和测试集进行预测,计算并输出模型的评估指标(均方误差和绝对误差)。本文采用 cross_val_score 函数进行交叉验证,并对模型进行了特征重要性分析。在构建 RF 模型时,本文只考虑室内定位距离判定,即 2 m 以内的范围。因此,将模型的' max_feature' 设置为 2,并使用 20 个决策树,采用有放回采样模式以提高距离判定精度。

1.4.3 卷积神经网络模型

卷积神经网络是一种深度学习模型,其应用于距离判定任务的核心思想是在卷积层和池化层提取高层次特征^[28]。模型包括 3 个卷积层、3 个池化层、3 个全连接层和 2 个 dropout 层,输入至 CNN 模型的数据经过卷积层和池化

层,形成特征提取器。卷积核作为特征探测器,对特征进行提取和计算,生成新的特征层。最终,CNN 模型输出的 2 个元素即为对二维坐标的预测。在具体训练过程中,每个客户端根据自身的数据集训练本地模型,将经过预处理的 RSSI 数据和坐标值纳入训练过程。在每次本地客户端训练结束后,计算误差梯度、更新权值并将其反馈至本地模型。在测试集上对训练好的模型进行测试和评价,将每个样本的特征值作为输入,输出对应预测值,并使用评价指标(均方误差和绝对误差等)对模型表现进行评价。

为了控制模型复杂度,并防止因过拟合而导致训练结果不佳,在损失函数中引入 L2 正则化项,如式(2)所示。它通过向模型的损失函数添加一个权重参数 L2 范数的惩罚项加以实现。实验中,设置 L2=0.01 的正则化参数,以提升模型泛化能力,减少过拟合现象。

$$L_2 = L_{data} + \lambda ||W||_2^2 \tag{2}$$

其中: L_{data} 是模型的数据损失,即本地模型训练预测的坐标和真实坐标之间的误差; λ 是正则化参数,用于控制正则化惩罚的强度; $||W||_2^2$ 是权重向量 W 的 L2 范数的平方,表示权重向量中各参数的平方和。

本文将针对以上 3 种模型在基于低功耗蓝牙的距离判定任务中的表现进行实验研究,以探究多特征分组化方法应用于这些模型对距离判定精度的影响。

2 实验方法与分析

2.1 实验环境

如图3所示,实验场地为长宽高为480 cm×400 cm×270 cm的室内测试环境。实验使用CSR1010芯片和BLE4.0协议的iBeacon,抓包工具为TI Sniffer + CC2540 dongle,手机工具为nRF Connect + oeBecaon。参考节点布置了20个,测试节点布置了15个。在每个坐标点针对每个Beacon节点采集25个RSSI数据,共100组数据,参考节点数据总量为2 000组,测试点数据总量为1 500组。

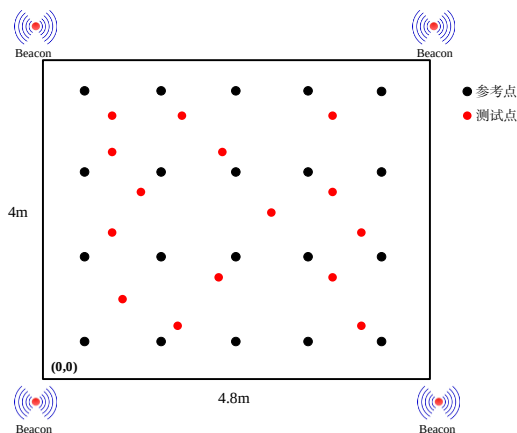


Fig. 3 Experimental site

图3 实验场地

2.2 单特征在不同模型中对距离判定精度的影响

本文主要研究在不同模型下,单一特征对距离判定精度的影响程度,并根据平均绝对误差这一判定条件确定每组组内特征对距离判定精度的影响程度。

2.2.1 SVM 模型

在SVM模型中,仅代入单一特征作为模型训练的特征变量,通过测试集预测值与真实值的差值,即平均绝对误差,评判该特征对距离判定精度的影响程度。若某个特征的平均绝对误差最小,表明该特征会提高距离判定精度,即该特征为显著特征;相反,若某个特征的平均绝对误差最大,表明该特征会降低距离判定精度,即该特征为非显著特征。因特征共分为3组,故实验分3次进行。

组一、组二、组三的实验结果如表2所示。在组一内,仅使用单一特征时,特征mean的均方误差和平均绝对误差都为最低,特征max的均方误差和平均绝对误差都为最高,即mean为组内显著特征,max为组内非显著特征;在组二内,仅使用单一特征时,特征L1 to mean的均方误差和平均绝对误差都为最低,特征L1 to median的均方误差和平均绝对误差都为最高,即L1 to mean为组内显著特征,L1 to median为组内非显著特征;在组三内,仅使用单一特征时,特征skew的均方误差和平均绝对误差都为最低,特征kurt的均方误差和平均绝对误差都为最高,即skew为组内显著特征,kurt为组内非显著特征。

Table 2 Feature error data for group 1, group 2 and group 3 under

SVM model

表2 SVM模型下组一、组二、组三特征误差数据

组别	特征	均方误差	绝对误差
组一	max	2.949 5	1.443 4
	mean	1.060 2	0.828 2
	median	1.566 0	1.029 2
	min	1.897 6	1.127 1
	one quarter	1.728 0	1.082 1
组二	three quarter	1.588 2	1.044 2
	L1 to mean	1.060 2	0.828 2
	L1 to median	2.168 6	1.203 4
组三	variance	1.230 1	0.874 2
	kurt	1.542 2	1.023 9
	skew	1.542 1	0.981 2

综上所述可得SVM模型下3个组内特征对距离判定的影响程度,如表3所示。在组一中,显著特征是平均值(mean),非显著特征是最大值(max),影响程度一般的特征是中位数(median)、第三四分位数(three quarter)、第一四分位数(one quarter)、最小值(min)。在组二中,显著特征是L1到平均值的距离(L1 to mean),非显著特征是L1到中位数的距离(L1 to median),影响程度一般的特征是方差(variance)。在组三中,显著特征是偏度(skew),非显著特征是峰度(kurt)。

Table 3 Analysis of the influence degree of various features under

SVM model

表3 SVM模型下各组特征影响程度分析

SVM模型	组一	组二	组三
显著特征	mean	L1 to mean	skew
非显著特征	max	L1 to median	kurt

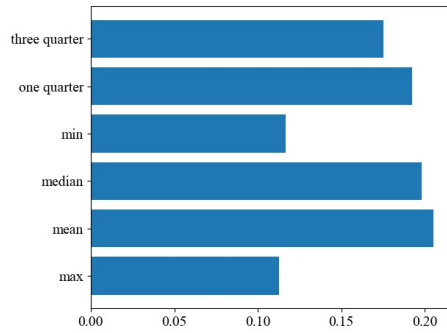
2.2.2 RF 模型

在RF模型中,通过引入RF模型库中的Feature_importances函数,以便直观研究不同特征对距离判定的影响。Feature_importances是机器学习模型中评估特征影响大小的属性,其原理是通过模型中每个特征的影响进行分析,确定每个特征对模型预测的影响程度。

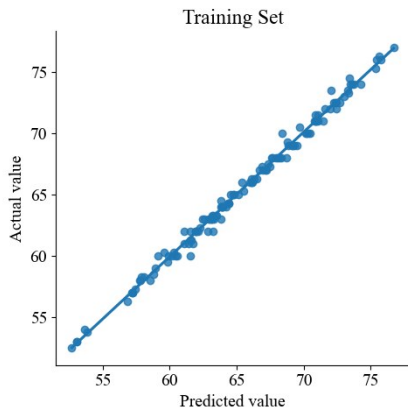
实验以各组单一特征为特征变量,实际距离 d 作为预测值,将其代入RF模型进行距离判定,通过 C 值($0 < C < 1$)输出, C 值代表了某一特征对距离判定精度的影响程度。若某个特征的输出 C 最大,表明该特征会提高距离判定精度,即该特征为显著特征;相反,若某个特征的输出 C 最小,表明该特征会降低距离判定精度,即该特征为非显著特征。各组特征对距离判定影响大小和相关训练集预测情况如图4—图6所示。

在组一内,特征mean的输出 C 值最大,特征max的输出 C 值最小,即mean为组内显著特征,max为组内非显著特征;在组二内,特征L1 to median的输出 C 值最大,特征L1 to mean的输出 C 值与L1 to median输出的 C 值几乎相同,而variance的输出 C 值最小,即L1 to median和L1 to mean为组内显著特征,variance为组内非显著特征;在组三

内, 特征 kurt 的输出 C 值最大, 特征 skew 的输出 C 值最小, 即 kurt 为组内显著特征, skew 为组内非显著特征。



(a) Feature importance
(a) 特征重要性



(b) Prediction of relevant test sets
(b) 相关测试集预测

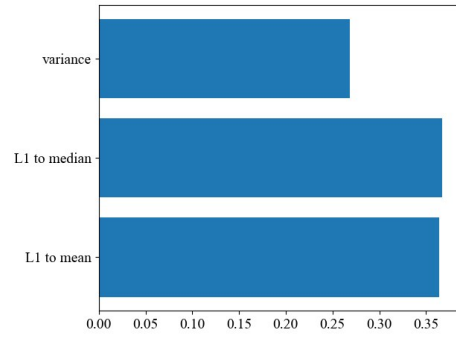
Fig. 4 Feature importance and associated test set predictions of group 1

图 4 组一特征重要性和相关测试集预测

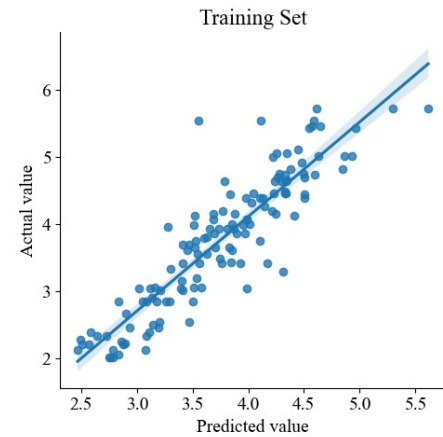
综上分析可得 RF 模型下 3 个组内特征对距离判定的影响程度, 如表 4 所示。在组一中, 显著特征是平均值 (mean), 非显著特征是最大值 (max), 影响程度一般的特征是中位数 (median)、第一四分位数 (one quarter)、第三四分位数 (three quarter)、最小值 (min)。在组二中, 显著特征是 L1 到平均值的距离 (L1 to mean)、L1 到中位数的距离 (L1 to median), 非显著特征是方差 (variance)。在组三中, 显著特征是峰度 (kurt), 非显著特征是偏度 (skew)。

2.2.3 CNN 模型

在 CNN 模型中, 仅代入单一特征作为模型训练的特征变量, 通过测试集预测值与真实值的差值, 即平均绝对误差, 评判该特征对距离判定精度的影响程度。若某个特征的平均绝对误差最小, 表明该特征会提高距离判定精度, 即该特征为显著特征; 相反, 若某个特征的平均绝对误差最大, 表明该特征会降低距离判定精度, 即该特征为非显著特征。因特征共分为 3 组, 故实验分 3 次进行。



(a) Feature importance
(a) 特征重要性



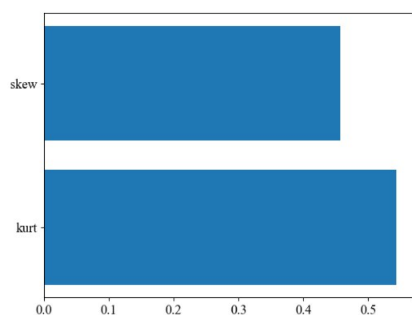
(b) Prediction of relevant test sets
(b) 相关测试集预测

Fig. 5 Feature importance and associated test set predictions of group 2

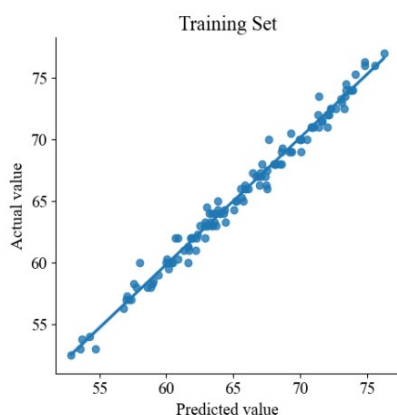
图 5 组二特征重要性和相关测试集预测

组一、组二、组三的实验结果如表 5 所示。在组一内, 仅使用单一特征时, 特征 three quarter 的均方误差和平均绝对误差都为最低, 特征 min 的均方误差和平均绝对误差与 three quarter 的均方误差和平均绝对误差几乎相同, 而特征 mean 的均方误差和平均绝对误差都为最高, 即 three quarter 和 min 为组内显著特征, mean 为组内非显著特征; 在组二内, 仅使用单一特征时, 特征 L1 to median 的均方误差和平均绝对误差都为最低, 特征 variance 的均方误差和平均绝对误差与 L1 to median 的均方误差和平均绝对误差几乎相同, 而特征 L1 to mean 的均方误差和平均绝对误差都为最高, 即 L1 to median 和 variance 为组内显著特征, L1 to mean 为组内非显著特征; 在组三内, 仅使用单一特征时, 特征 kurt 的均方误差和平均绝对误差都为最低, 特征 skew 的均方误差和平均绝对误差都为最高, 即 kurt 为组内显著特征, skew 为组内非显著特征。

综上分析可得 CNN 模型下 3 个组内特征对距离判定的影响程度, 如表 6 所示。在组一中, 显著特征是第三四



(a) Feature importance
(a) 特征重要性



(b) Prediction of relevant test sets
(b) 相关测试集预测

Fig. 6 Feature importance and associated test set predictions of group 3

图6 组三特征重要性和相关测试集预测

Table 4 Analysis of the influence degree of various features under RF model

表4 RF模型下各组特征影响程度分析

RF 模型	组一	组二	组三
显著特征	mean	L1 to mean, L1 to median	kurt
非显著特征	max	variance	skew

分位数(three quarter)、最小值(min),非显著特征是平均值(mean),影响程度一般的特征是中位数(median)、第一四分位数(one quarter)最大值(max)。在组二中,显著特征是L1到中位数的距离(L1 to median)、方差(variance),非显著特征是L1到平均值的距离(L1 to mean)。在组三中,显著特征是峰度(kurt),非显著特征是偏度(skew)。

2.3 组内多特征在不同模型中对距离判定精度的影响

分3组开展实验,以每组所有特征为特征变量,将其作为训练集进行模型训练,以研究各组多特征对距离判定精度的影响程度。

2.3.1 SVM 模型

在SVM模型中,组一至组三实验数据如表7所示,其

Table 5 Feature error data for group 1, group 2 and group 3 under CNN model

表5 CNN模型下组一、组二、组三特征误差数据

组别	特征	均方误差	绝对误差
组一	max	2.381 2	1.324 5
	mean	2.717 3	1.429 8
	median	1.141 4	0.888 8
	min	1.089 3	0.885 3
	one quarter	1.118 5	0.892 5
	three quarter	1.037 3	0.852 2
组二	L1 to mean	2.932 0	4.486 0
	L1 to median	0.852 2	0.784 4
	variance	0.869 3	0.774 2
组三	kurt	0.858 9	0.775 3
	skew	0.899 9	0.786 6

Table 6 Analysis of the influence degree of each feature group under CNN model

表6 CNN模型下各组特征影响程度分析

CNN 模型	组一	组二	组三
显著特征	Three quarter, min	L1 to median, variance	kurt
非显著特征	mean	L1 to mean	skew

均方误差、平均绝对误差分析如图7所示。组一的平均绝对误差最小,组二、组三的平均绝对误差几乎相同,但组三的均方误差最低。综上分析可知,在SVM模型下,组三多特征会提高距离判定精度,组二多特征会降低距离判定精度,组一多特征对距离判定精度影响一般。SVM模型在本次实验中表现较为稳定。通过适当选择核函数,SVM能够较好地处理非线性RSSI数据,并在不同的组内多特征实验设置下保持较高的准确率。

Table 7 Three sets of feature error data under SVM model

表7 SVM模型下3组特征误差数据

SVM 模型	组一	组二	组三
均方误差	0.998 7	1.542 2	0.910 5
绝对误差	0.809 6	1.022 5	1.023 9

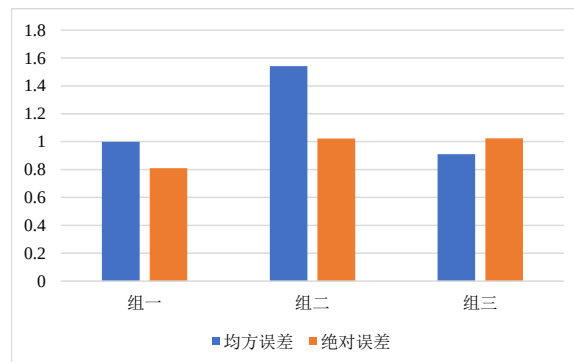


Fig. 7 Comparison of errors among groups under SVM model

图7 SVM模型各组误差比较

2.3.2 RF 模型

在RF模型中,组一至组三实验数据如表8所示,其均方误差、平均绝对误差分析如图8所示。组一的平均绝对误差、均方误差都最小,组三的平均绝对误差、均方误差都最大。综上分析可知,在RF模型下,组一多特征会提高距

离判定精度,组三多特征会降低距离判定精度,组二多特征对距离判定精度影响一般。RF 模型在本次实验中展示了强大的鲁棒性和抗噪能力。RF 通过多个决策树的集成进行预测,能够有效减少单一模型的偏差,从而提升整体预测性能。

Table 8 Three sets feature of error data under RF model

RF 模型	组一	组二	组三
均方误差	1.103 7	1.143 2	1.217 7
绝对误差	0.843 1	0.891 4	0.933 3

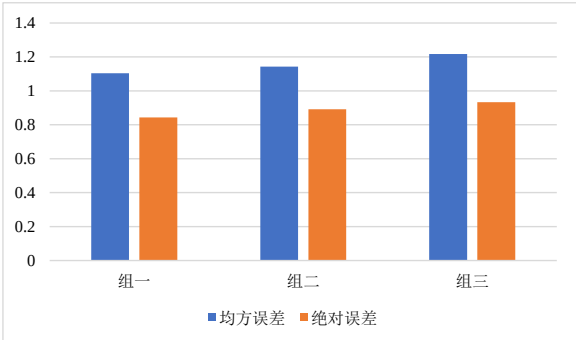


Fig. 8 Comparison of errors among groups under RF model

图 8 RF 模型各组误差比较

2.3.3 CNN 模型

在 CNN 模型中,组一至组三实验数据如表 9 所示,其均方误差、平均绝对误差分析如图 9 所示。组三的平均绝对误差、均方误差都最小,组二的平均绝对误差、均方误差都最大。综上分析可知,在 CNN 模型下,组三多特征会提高距离判定精度,组二多特征会降低距离判定精度,组一多特征对距离判定精度影响一般。尽管 CNN 模型在特征提取能力上优于 SVM 和 RF 模型,但 CNN 在本次实验中均方误差和绝对误差较大,其原因可能是本实验未对 RSSI 数据进行高维变换,导致模型未能充分学习数据特征。为保持 3 个模型的可对比性,其他两个模型(SVM 和 RF)并未对 RSSI 数据进行高维变换,这样的实验设计可确保 3 个模型在相同的数据预处理条件下进行性能评估,从而避免结果偏差。

Table 9 Three sets of feature error data under CNN model

CNN 模型	组一	组二	组三
均方误差	6.744 4	22.050 1	0.910 5
绝对误差	2.432 6	26.747 6	0.795 7

2.4 组间多特征在不同模型中对距离判定精度的影响

从不同组中选择 3 种特征在不同模型中训练,以研究不同特征组合对距离判定精度的影响,同时分析哪几种特征组合可以提高距离判定精度。

2.4.1 SVM 模型

在 SVM 模型中,选取 3 个不同组内特征作为特征变量训练,训练后通过测试集评估各组对距离判定精度的影响,并研究不同特征组合对距离判定精度的影响程度。实

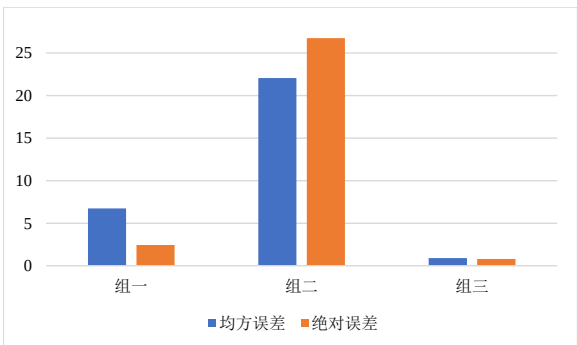


Fig. 9 Comparison of errors among CNN model groups

图 9 CNN 模型各组误差比较

验数据及其均方误差、平均绝对误差如表 10 所示(其中,√代表当前小组所选特征,下同)。

Table 10 Experimental data (1)

表 10 实验部分数据(1)		1	2	3	4	5	6	7	8
组一	max			√				√	√
	mean		√	√			√	√	√
	median								
	min					√			
组二	one quarter								
	three quarter	√			√	√			
	L1 to mean		√	√		√	√		
	L1 to median								
组三	variance	√							√
	kurt	√			√		√	√	
	skew		√		√				
	均方误差	1.6185	0.9336	1.1697	1.4532	1.4514	0.9198	1.2326	1.5768
测试集	绝对误差	1.0266	0.7691	0.8635	0.9583	0.9513	0.7921	0.8903	1.0187

由表 10 中各实验组的误差数据可知,2 组、3 组、6 组所选取的特征组合能明显提高距离判定精度,相较于 1 组、4 组、8 组所选取的特征组合,表现最好的 6 组可有效提升精度,达 20% 以上。分析 2 组、3 组、6 组的特征组合可知,2 组包含“mean、L1 to mean、skew”3 个显著特征,而 3 组和 6 组都只包含“mean、L1 to mean”2 个显著特征。综上所述,在 SVM 模型中,通过选取上述含有“mean、L1 to mean”2 个显著特征的特征组合,可有效提升距离判定精度。

2.4.2 RF 模型

在 RF 模型中,选取 3 个不同组内特征作为特征变量训练,训练后通过测试集评估各组对距离判定精度的影响,并研究不同特征组合对距离判定精度的影响程度。实验数据及其均方误差、平均绝对误差如表 11 所示。

由表 11 中各实验组的误差数据可得,4 组、6 组、7 组所选取的特征组合能提高距离判定精度,相较于 1 组、3 组、5 组所选取的特征组合,可提升精度约 10%。分析 4 组、6 组、7 组的特征组合可知,4 组包含“max、variance、kurt”3 个非显著特征,6 组和 7 组都包含“max、kurt”2 个非显著特征。综上所述,在 RF 模型中,通过选取上述含有“max、kurt”2 个非显著特征的组合,可在一定程度上提升距离判定精度。

Table 11 Experimental data(2)

表 11 实验部分数据(2)

组别	特征	1	2	3	4	5	6	7	8
组一	max				√		√	√	
	mean	√	√	√					√
	median								
	min						√		
	one quarter								
组二	three quarter								
	L1 to mean					√			
	L1 to median	√	√			√			
组三	variance			√	√				
	kurt	√		√		√	√	√	√
	skew		√		√			√	√
测试集	均方误差	1.1393	1.1305	1.1408	1.0370	1.3496	0.9887	0.9578	1.0872
	绝对误差	0.9013	0.8939	0.9123	0.8314	0.9549	0.8515	0.8651	0.8429

2.4.3 CNN 模型

在 CNN 模型中,选取 3 个不同组内特征作为特征变量训练,训练后通过测试集评估各组对距离判定精度的影响,并研究不同特征组合对距离判定精度的影响程度。实验数据及其均方误差、平均绝对误差如表 12 所示。

Table 12 Experimental data(3)

表 12 实验部分数据(3)

组别	特征	1	2	3	4	5	6	7	8
组一	max			√					
	mean			√					
	median								
	min		√		√	√	√		
	one quarter								
组二	three quarter	√	√		√	√	√		√
	L1 to mean		√						
	L1 to median							√	
组三	variance	√		√	√			√	
	kurt	√				√		√	√
	skew						√		√
测试集	均方误差	0.8263	3.5594	9.9954	1.1143	1.0253	1.0444	0.9236	1.0277
	绝对误差	0.7605	1.6764	3.0159	0.8871	0.8559	0.8644	0.8070	0.8475

由表 12 中各实验组的误差数据可得,1 组、7 组、8 组所选取的特征组合能明显提高距离判定精度,相较于 2 组、3 组、4 组所选取的特征组合,表现最好的 1 组可有效提升精度,达 59% 以上。分析 1 组、7 组、8 组的特征组合可知,1 组包含“Three quarter、variance、kurt”3 个显著特征,7 组也包含“L1 to median、variance、kurt”3 个显著特征。综上所述,在 CNN 模型中,通过选取上述“Three quarter、variance、kurt”或“L1 to median、variance、kurt”的特征组合,可有效提升距离判定精度。

3 讨论

本文使用支持向量机、随机森林和卷积神经网络 3 种模型对低功耗蓝牙信号强度数据进行了评估,并通过实验对这 3 种模型的性能进行了详细分析。支持向量机模型在低功耗蓝牙信号强度数据上的表现较为稳定,表明其在

距离判定任务中具有较高的预测精度。其在处理数据量较小且特征较简单的情况下表现优异,含有“mean、L1 to mean”2 个显著特征的特征组合可有效提升距离判定精度。随机森林模型展示了强大的鲁棒性和抗噪能力,通过集成多个决策树进行预测,有效减少了单一模型的偏差,从而提升了整体预测性能,其在避免过拟合方面表现出色,含有“max、kurt”2 个非显著特征的特征组合可在一定程度上提升距离判定精度。卷积神经网络模型在实验中的结果与支持向量机和随机森林模型存在一定差距,只有当选取含有“Three quarter、variance、kurt”或“L1 to median、variance、kurt”3 个显性特征的特征组合时,才能有效提升距离判定精度。但也有研究表明,CNN 在 RSSI 数据经高维变换后,模型性能表现整体优异。文献[29]的实验结果显示,经过转换后的 RSSI 数据在 CNN 模型中能更有效地进行位置识别;文献[30]的研究表明,CNN 在处理 RSSI 数据的高维特征提取方面表现出色,有助于提高定位准确性和鲁棒性。后续研究中,将尝试通过增加训练数据维度,以进一步研究 CNN 模型的性能。

建议在实际应用中,根据具体场景选择合适的模型以应用于距离判定任务。支持向量机模型在处理少量特征时表现较好,适用于实时性要求高、易于在本地部署的场景(例如,可在商场、图书馆、博物馆为用户提供实时导航,或在智能家居系统中优化设备管理和资源分配等)。随机森林模型在处理多特征组合时具有较好的鲁棒性和稳定性,适用于复杂环境下的距离判定(例如,在大型工厂、车间、仓库中优化物流管理和生产线控制,或在智能交通系统中实现车队管理、交通流量监控和事故预防)。卷积神经网络模型在特征提取和非线性关系捕捉上具有优势,尽管在实验中表现不如预期,但其强大的特征提取能力在未来研究中仍具有重要意义,对于数据量大且复杂的情况(例如,在智能农业系统中优化作业效率,或在医院中应用于医疗监护系统的病人追踪,确保病人安全并优化医疗资源配置),可进一步探索卷积神经网络的应用潜力。

4 结语

本文系统地探讨了低功耗蓝牙信号强度特征在距离判定任务中的应用,重点分析了特征分组对不同机器学习模型性能的影响。通过实验验证发现,特征的合理选取和组合能显著提升距离判定精度,尤其是在使用多个显著特征时效果更为显著。支持向量机、随机森林和卷积神经网络模型在不同特征组合下表现出不同的优势和局限,为未来 BLE 信号强度特征在实际应用中的优化提供了理论依据和实践参考。

同时,本文研究也存在一定局限性和不足,体现如下:①实验数据量有限;②基于深度学习模型的实验有待进一步增加。未来研究将着重解决这些问题,以进一步提高距

离判定准确性和实用性。

参考文献:

- [1] ERUNKULU O O, ZUNGERU A M, LEBEKWE C K, et al. 5G mobile communication applications: a survey and comparison of use cases [J]. IEEE Access, 2021, 9: 97251–97295.
- [2] LIANG M, HU X M, LI P, et al. Bluetooth low energy security experiment technology of “smart phone + wearable device” [J]. Computer Technology and Development, 2020, 30(11): 111–116.
梁敏, 胡曦明, 李鹏, 等. “手机+可穿戴设备”的低功耗蓝牙安全实验技术[J]. 计算机技术与发展, 2020, 30(11): 111–116.
- [3] ZHANG Y L, WANG J, HAN H Z, et al. Research and precision analysis of indoor ranging model based on Bluetooth [J]. Science of Surveying and Mapping, 2021, 46(3): 1–7, 66.
张亚磊, 王坚, 韩厚增, 等. 蓝牙室内测距模型研究与精度分析[J]. 测绘科学, 2021, 46(3): 1–7, 66.
- [4] ALMAMORI A, MOHAN S. Estimation of channel state information (CSI) in cell-free massive MIMO based on time of arrival (ToA) [J]. Wireless Personal Communications, 2020, 114(2): 1825–1831.
- [5] ZOU Y, LIU H. TDOA localization with unknown signal propagation speed and sensor position errors [J]. IEEE Communications Letters, 2020, 24(5): 1024–1027.
- [6] LE T K, HO K C. Joint source and sensor localization by angles of arrival [J]. IEEE Transactions on Signal Processing, 2020, 68(20): 6521–6534.
- [7] TORRES S E, LEAL R C, GARCIA R C. Application and perspectives of MALDI – TOF mass spectrometry in clinical microbiology laboratories [J]. Microorganisms, 2021, 9(7): 1539.
- [8] BOUSSAD Y, MAHFOUDI M N, LEGOUT A, et al. Evaluating smartphone accuracy for RSSI measurements [J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70(14): 1–12.
- [9] SUBHAN F, KHAN A, SALEEM S, et al. Experimental analysis of received signals strength in Bluetooth Low Energy (BLE) and its effect on distance and position estimation [J]. Transactions on Emerging Telecommunications Technologies, 2022, 33(2): e3793.
- [10] WANG Z K, LUO B Q. Hybrid fingerprint indoor localization method based on ECA-CNN [J]. Journal of Computer Research and Development, 2024, 61(2): 428–440.
王正康, 骆冰清. 基于ECA-CNN的混合指纹室内定位方法[J]. 计算机研究与发展, 2024, 61(2): 428–440.
- [11] ZHU X Q, QU W Y, QIU T, et al. Indoor intelligent fingerprint-based localization: principles, approaches and challenges [J]. IEEE Communications Surveys & Tutorials, 2020, 22(4): 2634–2657.
- [12] YANG F, LIU J B, GONG X D, et al. Mobile phone-based indoor positioning method using adaptive compressed sensing for walking-surveyed fingerprinting [J/OL]. Geomatics and Information Science of Wuhan University, 2024-02-24. <https://doi.org/10.13203/j.whugis.20230241>.
杨帆, 柳景斌, 龚晓东, 等. 基于移动指纹采集的手机自适应压缩感知室内定位方法[J/OL]. 武汉大学学报(信息科学版), 2024-02-24. <https://doi.org/10.13203/j.whugis.20230241>.
- [13] LU M C, WANG S H, LI Y K, et al. Bluetooth location algorithm based on feature matching and distance weighting [J]. Journal of Computer Applications, 2018, 38(8): 2359–2364.
陆明炽, 王守华, 李云柯, 等. 基于特征匹配和距离加权的蓝牙定位算法[J]. 计算机应用, 2018, 38(8): 2359–2364.
- [14] SONG X, FAN X, HE X, et al. CNNLoc: deep-learning based indoor localization with Wi-Fi fingerprinting [C]//2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CB-DCOM/IOP/SCI), 2019: 589–595.
- [15] SUN D, ZHANG Y, XIA W, et al. A BLE indoor positioning algorithm based on weighted fingerprint feature matching using AoA and RSSI [C]//Changsha: 2021 13th International Conference on Wireless Communications and Signal Processing (WCSP), 2021.
- [16] YAN J, MA C, KANG B, et al. Extreme learning machine and AdaBoost-based localization using CSI and RSSI [J]. IEEE Communications Letters, 2021, 25(6): 1906–1910.
- [17] SU Z, PAHLAVAN K, AGU E. Performance evaluation of COVID-19 proximity detection using bluetooth LE signal [J]. IEEE Access, 2021, 9: 38891–38906.
- [18] PABLO G M, RAMIRO D, LUCAS B, et al. Feature selection for proximity estimation in COVID-19 contact tracing apps based on Bluetooth Low Energy (BLE) [J]. Pervasive and Mobile Computing, 2021, 77: 101474–101487.
- [19] ABBAS H A, BOSKANY N W, GHAFOR K Z, et al. Wi-Fi based accurate indoor localization system using SVM and LSTM algorithms [C]//2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI), 2021: 416–422.
- [20] MADURANGA M W P, TILWARI V, ABEYSEKERA R. Improved-RSSI-based indoor localization by using pseudo-linear solution with machine learning algorithms [J]. Journal of Electrical Systems and Information Technology, 2024, 11(1): 10–30.
- [21] ADI S, WILLY P P, MUHAMMAD A A H, et al. Vehicles position tracking in parking lots using K-nearest neighbor and fingerprinting based on RSSI bluetooth [C]//Jakarta: 2021 Sixth International Conference on Informatics and Computing (ICIC), 2021.
- [22] YANG F, LIU J, ZHANG Q, et al. CNN-based two-branch multi-scale feature extraction network for retrosynthesis prediction [J]. BMC Bioinformatics, 2022, 23(1): 362–378.
- [23] SUN Q, YU X, LI H, et al. Adaptive feature extraction and fault diagnosis for three-phase inverter based on hybrid-CNN models under variable operating conditions [J]. Complex & Intelligent Systems, 2022, 8(3): 29–42.
- [24] SINDHUJA S, CHAKKARAVARTHY D M, SELVAM J. CR-IOT based selfish attack detection via RSSI-LSTM [J]. Measurement: Sensors, 2023, 27: 100729–100738.
- [25] ZHANG M, LI W, ZHANG L, et al. A Pearson correlation-based adaptive variable grouping method for large-scale multi-objective optimization [J]. Information Sciences, 2023, 639: 118737.
- [26] ROY A, CHAKRABORTY S. Support vector machine in structural reliability analysis: a review [J]. Reliability Engineering & System Safety, 2023, 233: 109126.
- [27] LU J, HUANG C, TAO X, et al. Annual forest disturbance intensity mapped using Landsat time series and field inventory data for the conterminous United States (1986–2015) [J]. Remote Sensing of Environment, 2022, 275: 113003.
- [28] AMANAT A, RIZWAN M, JAVED A R, et al. Deep learning for depression detection from textual data [J]. Electronics, 2022, 11(5): 676–689.
- [29] LEE H, LEE J. Convolutional model with a time series feature based on RSSI analysis with the Markov transition field for enhancement of location recognition [J]. Sensors, 2023, 23(7): 3453–3469.
- [30] HOU Y, YANG X, ABBASI Q H. Efficient AoA-based wireless indoor localization for hospital outpatients using mobile devices [J]. Sensors, 2018, 18(11): 3698–3715.

(责任编辑: 孙娟)