

DOI:10.19651/j.cnki.emt.2108639

## 改进 DBSCAN 聚类的信道状态信息定位算法\*

刘宇<sup>1,2,3</sup> 余学祥<sup>1,2,3</sup> 谢世成<sup>1,2,3</sup> 刘双<sup>1,2,3</sup> 朱平<sup>1,2,3</sup>

(1.安徽理工大学空间信息与测绘工程学院 淮南 232001; 2.安徽理工大学矿山采动灾害空天地协同监测与预警安徽普通高校重点实验室 淮南 232001; 3.安徽理工大学矿区环境与灾害协同监测煤炭行业工程研究中心 淮南 232001)

**摘要:**近年来,利用 WiFi 信道状态信息的无线信号在室内定位、跌倒检测、身份识别等场景都发挥了重要应用价值。然而,复杂环境下多径效应的影响使得指纹定位的精度仍有待提高。针对这一问题,本文在降噪时提出了一种基于密度的自适应聚类算法,并在定位阶段联合动态加权 K 邻近算法进行匹配。首先,使用 Hampel 算法去除幅值信息的离群点;然后,将改进的 DBSCAN 算法自动调节参数对数据聚类;最后,用动态加权 K 邻近算法进行指纹库与实时定位点的匹配。仿真实验表明,在约  $5 \times 10 \text{ m}^2$  的定位区域内,DBSCAN 算法的平均定位精度达到 1.579 m,其中定位精度在 2 m 内的占比相对于传统指纹法提高了 42.9%。

**关键词:**室内定位;信道状态信息;无线感知;定位精度

**中图分类号:** TN92 **文献标识码:** A **国家标准学科分类代码:** 510.5015

## Channel state information localization based on improved DBSCAN clustering algorithm

Liu Yu<sup>1,2,3</sup> Yu Xuexiang<sup>1,2,3</sup> Xie Shicheng<sup>1,2,3</sup> Liu Shuang<sup>1,2,3</sup> Zhu Ping<sup>1,2,3</sup>

(1. School of Geomatics, Anhui University of Science and Technology, Huainan 232001, China;

2. Key Laboratory of Aviation-Aerospace-Ground Cooperative Monitoring and Early Warning of Coal Mining-Induced Disasters of Anhui Higher Education Institutes, Anhui University of Science and Technology, Huainan 232001, China;

3. Coal Industry Engineering Research Center of Mining Area Environmental and Disaster Cooperative Monitoring, Anhui University of Science and Technology, Huainan 232001, China)

**Abstract:** In recent years, wireless signals using WiFi channel state information have played important roles in scenarios such as indoor positioning, fall detection, and identification. However, the impact of multipath effects in complex environments makes the accuracy of fingerprint positioning to be improved. To solve this problem, this paper proposes an improved density-based spatial clustering of applications with noise during the process of noise reduction combined with enhanced weighted K-nearest neighbor algorithm in the online stage. First, the Hampel algorithm is used to remove outliers of the amplitude information; then, the improved DBSCAN algorithm automatically adjusts the parameter to cluster data. Finally, the enhanced weighted K-nearest neighbor algorithm is used to match the real-time positioning points from the fingerprint database. The experimental results show that the average positioning accuracy of the DBSCAN algorithm reaches 1.579 m in a positioning area of about  $5 \times 10 \text{ m}^2$ , and the percentage of error within 2 m is increased by 42.9% compared to the traditional fingerprint method.

**Keywords:** indoor positioning; channel state information; wireless position; positioning accuracy

## 0 引言

21 世纪是一个以大数据与物联网为载体的时代,信息化测绘及多传感器集成技术得到了突破性的发展,人们也

越来越关注自身的精确位置与相关导航定位的服务。目前,全球导航定位系统(global navigation satellite system, GNSS)已经成为人们外出的必备工具。然而,因自然或人工障碍物遮挡、导致卫星信号严重干扰的复杂室内环境(如

收稿日期:2021-12-24

\* 基金项目:国家自然科学基金(41474026)、安徽省重点研究与开发计划(202104a07020014)、安徽省科技重大科技专项(202103a05020026)资助

地下停车场、超市、机场大厅等)使得室内定位技术的需求越来越强烈。回顾近年来普及的室内导航系统,人们已开发了不同的技术,如无线保真(wireless fidelity, WiFi)<sup>[1-2]</sup>、蓝牙(BLE)<sup>[3]</sup>、ZigBee<sup>[4]</sup>、超宽带(ultra-wideband, UWB)<sup>[5]</sup>、RFID<sup>[6]</sup>、地磁场<sup>[7]</sup>、可见光<sup>[8]</sup>、微机电系统(micro-electro-mechanical system, MEMS)传感器<sup>[9]</sup>等。不同的定位方法精度各异,从分米级、亚米级、米级到 10 m 不等,并且在抗干扰性、相对成本和典型覆盖范围等方面也各有不同<sup>[10]</sup>。

基于信号指纹定位的实现过程主要分成两个阶段:在离线阶段中,主要任务是采集信道状态信息(channel state information, CSI)信号,并在预处理后建立指纹库;再到在线阶段的工作是将待定位点与数据库的指纹匹配,从而解算估计出测量信息的位置<sup>[11]</sup>。文献[12]首次提出在 Intel5300 卡上实现的 PinLoc 的功能系统,该系统进行的 CSI 指纹定位平均准确率达到 89%,验证了物理层信息在 WiFi 定位的可行性。为了减少手动提取信号特征的工作量,在文献[13-14]中,作者分别提出基于误差逆传播与广义回归神经网络的方法对 CSI 信息进行训练分类,再通过模型进行目标对象预测。类似的也在文献[15]中,作者提出主成分分析和广义回归神经网络训练模型进行目标对象预测。文献[16]提出了基于 Keans 的加权 KNN 算法,提高匹配精度。为了获得嘈杂环境下节点的最优位置,文献[17]提出了最优 WKNN 算法,首先利用自适应卡尔曼滤波器降低无线传感器网接收的信号强度噪声,再用文化基因算法优化权值,来实现最优节点的目标定位。为了减少 CSI 数据的训练成本和复杂度,文献[18]提出基于张量分量和最小二乘迭代算法来提高定位性能。文献[19]提出基于矩阵填充理论构建数据库,并在定位阶段使用卡方距离代替欧氏距离,相比于传统 K 近邻法提高了精度。

为了提高聚类时手动输入参数的效率,本文提出一种改进的基于密度的自适应聚类算法(density-based spatial clustering of applications with noise, DBSCAN)结合动态加权 K 近邻方法(enhanced weighted K-nearest neighbor algorithm, EWKNN)。首先在离线阶段进行异常数据的剔除和填补,再通过优化的聚类算法对不同的链路进行自适应训练分类,进一步去除噪声。最后在实时定位阶段,通过动态修正 K 值的 WKNN 算法对待定位点进行匹配。通过 MATLAB 仿真结果表明,本文的定位方法具有可行性。

## 1 相关理论基础

### 1.1 信道状态信息

CSI 信号采取正交频分复用(orthogonal frequency division multiplexing, OFDM)的方式进行调制,它具有良好的抗干扰性。CSI 描述了无线信号从发射端传播到接收端的过程中,经由多条路径传播形成的散射、信号的衰弱、功率的衰减等组合信息,能反映环境特征的多径叠加效应。

以  $\mathbf{Y}$  表示接收端信号向量、 $\mathbf{X}$  为发射端信号向量、 $\mathbf{H}$  为状态矩阵、 $N$  表示高斯白噪声,则一个 CSI 的频域模型被建模表示为:

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + N \quad (1)$$

其中,信道矩阵是多维子载波信息的集合,表示为  $N_r \times N_t \times 30$  的复数矩阵,其中  $N_r$  为接收端的天线数,  $N_t$  为发送端的天线数,30 为 OFDM 子载波数。以  $\mathbf{H}(i)$  表示子载波  $i$  的信道状态信息矩阵,分别描述振幅和相位,CSI 信道矩阵被描述为:

$$\mathbf{H}(i) = |\mathbf{H}(i)| e^{j\sin\angle\mathbf{H}(i)} \quad (2)$$

由于普通的 WiFi 硬件很难输出描述物理层的 CSI 信息,目前采集 CSI 值有 3 种方式。第 1 种是利用构建软件定义无线电硬件平台(SDR platform),但所需成本较高。第 2 种是使用 Atheros9390 NIC 获取 CSI 原始数据包。第 3 种是通过 Intel5300 无线网卡并修改软件驱动的方式<sup>[20]</sup>,采集过程中每个信道包括 30 个子载波,实验采集并提取 CSI 振幅曲线,如图 1 所示。

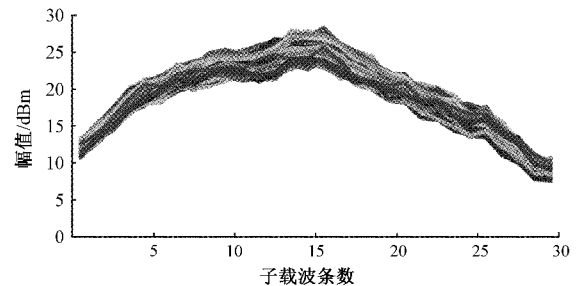


图 1 CSI 振幅曲线

### 1.2 精度评价指标

定位精度描述了某一定位算法或系统在比较估测结果与真实位置时的相近程度,主要的衡量指标有如下几个。

#### 1) 均方误差

均方误差(mean squared error, MSE)用于评价一组数据自身的离散变化程度,是衡量定位算法针对某个标签单次定位结果的最常用指标<sup>[21]</sup>。设待测目标真实位置的坐标个数为  $n$  个,表示为  $x_i = (x_i, y_i), i = 1, 2, 3, \dots, N$ , 对应的估计坐标为  $\hat{x}_i = (\hat{x}_i, \hat{y}_i), i = 1, 2, 3, \dots, N$ , 则第  $i$  个位置的均方误差可表示为:

$$MSE_i = \sqrt{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2} \quad (3)$$

#### 2) 均方根误差

均方根误差(root mean square error, RMSE)是均方误差的算术平方根,用来评价一组观测数据与真实值的偏差。在同上的估计二维位置的假设中,均方根误差可表示为:

$$RMSE_i = \sqrt{\frac{\sum_{i=1}^N \|x_i - \hat{x}_i\|^2}{N}} \quad (4)$$

#### 3) 累积分布函数

累积分布函数(cumulated density function, CDF)描述

了真实位置与估计位置之间的距离误差的累积分布概率,在概率统计里即表示统计所有小于或等于某一设定数值出现的概率和。

## 2 数据预处理及相关算法

### 2.1 Hampel 去除异常值处理

Hampel 滤波去除异常值的原理是寻找信号序列与中位数的差超出 3 个标准偏差的样本,即在任何位于区间  $[\mu - \gamma\delta, \mu + \gamma\delta]$  以外的数据点视为离群点,其中  $\mu$  与  $\delta$  分别表示样本序列的中值和绝对中误差,  $\gamma$  为相关参数,通常取 3。Hampel 针对某条子载波的 120 个样本去除离群点的过程,如图 2 所示。图中实线描述检测出的离散程度偏大的点,虚线为处理后的子载波。

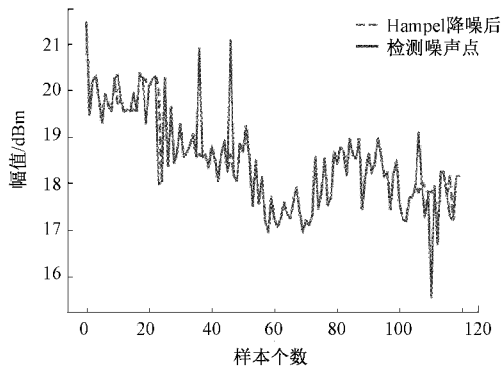


图 2 Hampel 处理异常值

### 2.2 改进 DBSCAN 聚类算法原理

基于机器学习的 DBSCAN 算法的定义可概括为由密度可达关系导出的最大密度相连的样本集合(或簇)<sup>[22-23]</sup>。DBSCAN 需指定两个参数 ( $\epsilon, MinPts$ ), 其算法过程可解释为:导入一个包括  $m$  个样本的数据集  $D = (x_1, x_2, x_3, \dots, x_m)$ ; 首先分别初始化输入参数  $D, \epsilon, MinPts$ ; 从选取某样本开始出发,以距离参数  $\epsilon$  为半径的圆的邻域范围内,当数据点的数量  $\geq$  最小点参数  $MinPts$  时,则所有满足条件的点都成一个聚类集合;此后,集合内部的点将继续以  $\epsilon$  为半径向外辐射,形成新的聚类集合,不断遍历循环,直到无法找到新的数据点结束。此时,无法形成聚类的点作为噪声,进行剔除或填补,最终输出基于密度的簇划分集合  $C$ 。

DBSCAN 算法原理简单,优势体现在无需指定种类数,通过数据集中对象堆积的密度自动聚类成簇,从而识别出异常点。但缺点在于参数的敏感性,人工输入的半径参数和密度阈值往往易造成参数取值不合理,导致准确率降低。例如当半径参数过大时,集合个数太少会使得聚类毫无意义,为了使簇的个数更加准确,本文因此将改进的自适应参数密度聚类算法引入进来。

自适应的核心在于对样本集中所有点先进行两两欧氏距离计算,对各样本的对象密度进行升序排列,并在约束上

下限的范围内将最大的差值作为半径,根据距离空间中异常数据的对象具有分布离散且无规律等特点,由统计原则找到合理的半径参数,从而把有意义的对象和异常数据区分开,来降低标记噪声对象的错误率。DBSCAN 对数据处理的对比图如图 3 与 4 所示。从图 4 可以看出,经算法处理噪声后的子载波数据更加集中。

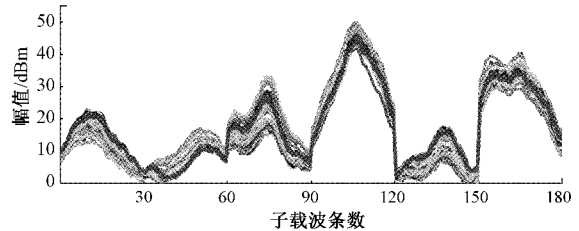


图 3 DBSCAN 处理前幅值图

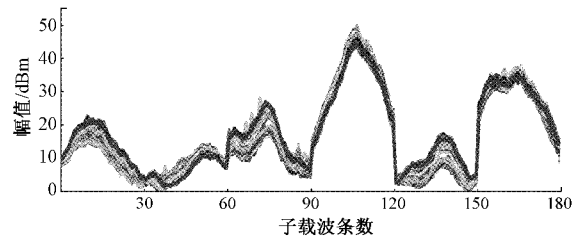


图 4 DBSCAN 处理后幅值图

### 2.3 改进 WKNN 定位算法原理

加权最邻近算法(WKNN)是基于 KNN 算法的一种改进,当选取了最邻近  $K (K \geq 2)$  个参考点后,KNN 采用均值结果作为待定位位置的输出,而 WKNN 在本实验中增加了距离倒数为加权系数,即指纹点到待测点的欧氏距离越大,所对应的权值越小。WKNN 的定位公式可描述为:

$$(\hat{x}_i, \hat{y}_i) = \sum_{j=1}^k \frac{\frac{1}{dis_j + \epsilon}}{\sum_{j=1}^k \frac{1}{dis_j + \epsilon}} (x_j, y_j) \quad (5)$$

其中,  $(\hat{x}_i, \hat{y}_i), (x_i, y_i)$  分别表示第  $i$  个 WKNN 定位最终坐标、参考点坐标、待定位点到指纹数据库点的欧氏距离;  $\epsilon$  为一个趋于 0 的极小数字,是为了防止分母距离出现 0 的情况。

为了解决实际测量中使用传统 WKNN 算法仍面临的精度有限问题,引入了动态加权 K 邻近算法(enhanced weighted K-nearest neighbor algorithm, EWKNN)<sup>[24]</sup>,旨在指纹定位时能够动态的剔除距离较大的参考点。该算法的基本思想流程如下。

- 1) 计算待测点与指纹数据库所有点的欧氏距离。
- 2) 由  $K$  值分别求出  $dis(1)$  到  $dis(k)$  的差值,并计算平均距离,将其作为阈值,剔除差值里大于该阈值的数,设保留剩余的近邻点个数  $L$ ,代替原  $K$  值。
- 3) 对新的距离值进行权值分配,分别与适量坐标相乘求和。

4)重复步骤 1)和 2),直到遍历结束。

EWKNN 的公式表示为:

$$(\hat{x}_i, \hat{y}_i) = \sum_{i=1}^L \frac{1}{\sum_{i=1}^L \frac{1}{dis(i)}} (x_i, y_i) \quad (6)$$

### 3 实验结果分析

#### 3.1 实验环境

本实验的定位场景如图 5 所示。实验测试场景位于安徽理工大学空间信息与测绘工程学院的一楼实验室。整个试验区域中,在离线阶段均匀地设置了 50 个参考点,如图中圆圈表示,区域面积约为  $5 \times 10 \text{ m}^2$ 。而在线定位阶段,随机选取 15 个位置作为测试点,图中以三角形表示,来验证定位算法的性能。以左上角的参考点坐标为例设为(0,0),则该区域的 4 个顶点位置分别为(0,0)、(0,4)、(9,0)、(9,4)。

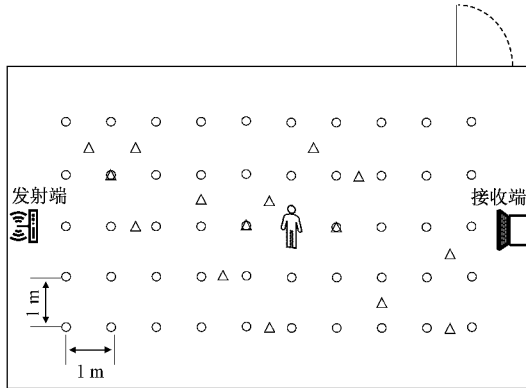


图 5 实验环境

在测试中使用 R4CM 型号的路由器作为信号接入点 (AP) 发送数据包,接收装置为装有 Intel5300 网卡的联想笔记本电脑,且操作系统为 Linux 内核的 Ubuntu14.04,并将其设置为 Ap 模式,最终通过开源软件 CSITool 驱动程序可采集接收的 CSI 信息。实验人员在定位区域内保持站立不动,随后开启发包,用来接收发射端的数据包速度为 10 Packets/s,每次采集时长约 30 s,即每个待测点采集 300 个样本数据。2 根发射端和 3 根接收端天线距离地面的距离约为 1 m,设备的工作频段为 2.4GHz,通信信道为 6 个链路。各链路又包含 30 个子载波,所以共 180 个特征维度 ( $2 \times 3 \times 30$ )。

#### 3.2 实验验证与分析

在试验区域随机选择 15 个待定位坐标,分别以 DBSCAN-EWKNN、KEANS-EWKNN、DBSCAN-WKNN、KMEANS-WKNN 算法完成定位,对应的定位结果分布如图 6 所示。从图 6 可以看出,DBSCAN-EWKNN 算法相较于其他 3 种具有更高的精确度,且不同算法的具体误差如表 1 所示,并将不同算法误差指标形成箱型图如图 7 表示。

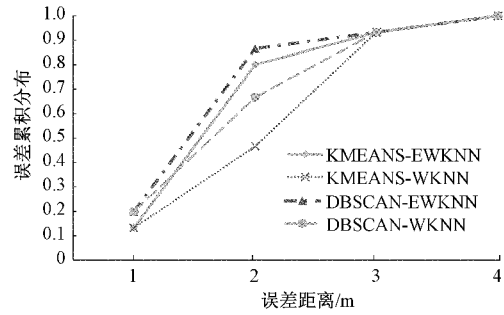


图 6 定位算法比较图

表 1 不同算法的误差对比

定位算法	KMEANS-WKNN	DBSCAN-WKNN	KMEANS-EWKNN	DBSCAN-EWKNN
Error <sub>min</sub>	0.470	0.445	0.300	0.409
Error <sub>max</sub>	3.077	3.160	3.637	3.028
AVERAGE	1.842	1.630	1.640	1.579

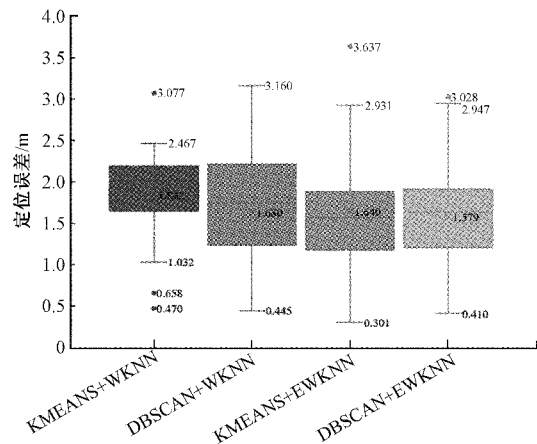


图 7 不同算法的相关指标比较图

从表 1 和图 7 中可看到分别以最小定位误差 Error<sub>min</sub>、最大定位误差 Error<sub>max</sub> 以及误差均值 AVERAGE 作为指标衡量各算法定位的结果。采用 KMEANS-EWKNN 算法得到的误差均值为 1.640 m,而以 DBSCAN-EWKNN 算法得到误差均值为 1.579 m,最大定位误差由 3.637 m 减小至 3.028 m。其中,通过本文设计的算法,比例为 86.7% 的定位误差可以在 2 m 以内,相比于 KMEANS-WKNN 算法提高了 42.9%。

### 4 结 论

Wi-Fi 指纹定位法为目前最常见的定位方法,本文设计了一种自适应 DBSCAN 联合 EWKNN 算法,该方法首先通过 Hampel 算法去除异常值噪声,再使用改进的聚类算法自动调节参数构成数据指纹库,最后用改进的 WKNN 算法进行实时定位点的一一匹配。实验结果表明,在约为  $5 \times 10 \text{ m}^2$  的定位区域内,该方法的平均定位精度达到

1.579 m, 优于传统的指纹定位, 验证了其可行性。在本文后续的工作中, 将继续引入深度学习网络模型, 同时使用目标跟踪滤波算法融合其他传感器信号, 达到定位准确性、复杂性与成本之间整体平衡。

### 参考文献

- [1] 赖朝安, 龙漂. 基于高斯过程回归和 WiFi 指纹的室内定位方法[J]. 电子测量与仪器学报, 2021, 35(2): 186-193.
- [2] LIU F, LIU J, YIN Y Q, et al. Survey on WiFi-based indoor positioning techniques [J]. IET Communications, 2020, 14(9): 1372-1383.
- [3] BAI L, CIRAVEGNA F, BOND R, et al. A low cost indoor positioning system using bluetooth low energy[J]. IEEE Access, 2020, 8: 136858-136871.
- [4] DING X, DONG S. Improving positioning algorithm based on RSSI [J]. Wireless Personal Communications, 2020, 110(4): 1947-1961.
- [5] WANG J, WANG M M, YANG D, et al. UWB positioning algorithm and accuracy evaluation for different indoor scenes [J]. International Journal of Image and Data Fusion, 2021, 12(3): 203-225.
- [6] 李丽娜, 马俊, 徐攀峰, 等. RFID 室内定位技术研究综述[J]. 计算机应用与软件, 2015, 32(9): 1-3.
- [7] 金展, 康瑞清, 王威. 基于地磁序列匹配的大型室内动态定位方法[J]. 仪器仪表学报, 2021, 42(7): 199-207.
- [8] BAI L, YANG Y, GUO C, et al. Camera assisted received signal strength ratio algorithm for indoor visible light positioning [J]. IEEE Communications Letters, 2019, 23(11): 2022-2025.
- [9] 赵天贺, 汪伟. 基于 MEMS 技术的捷联惯导系统现状[J]. 飞航导弹, 2017(10): 17-21.
- [10] 闫大禹, 宋伟, 王旭丹, 等. 国内室内定位技术发展现状综述[J]. 导航定位学报, 2019, 7(4): 5-12.
- [11] FANG S, LIN T. A dynamic system approach for radio location fingerprinting in wireless local area networks[J]. IEEE Transactions on Communications, 2010, 58(4): 1020-1025.
- [12] SEN S, RADUNOVIC B, CHOUDHURY R R, et al. You are facing the Mona Lisa: Spot localization using PHY layer information[C]. Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services, New York: ACM, 2012: 183-196.
- [13] 刘帅, 王旭东, 吴楠. 一种基于卷积神经网络的 CSI 指纹室内定位方法[J]. 工程科学学报, 2021, 43(11): 1512-1521.
- [14] 蒋芳, 张南飞, 胡艳军, 等. 基于 BP 神经网络的 CSI 无源目标分类方法[J]. 北京邮电大学学报, 2020, 43(1): 40-45.
- [15] 李新春, 黄朝晖. 基于聚类的 PCA 和 GRNN 的 CSI 指纹定位算法[J]. 重庆邮电大学学报(自然科学版), 2021, 33(3): 449-457.
- [16] 吉彩云, 袁明辉, 李瑞祥, 等. 基于 K-means 的室内定位加权优化 k-NN 算法[J]. 电子测量技术, 2018, 41(10): 66-69.
- [17] FANG X, JIANG Z, LEI N, et al. Optimal weighted K-nearest neighbour algorithm for wireless sensor network fingerprint localisation in noisy environment [J]. IET Communications, 2018, 12(10): 1171-1177.
- [18] 周牧, 龙玥辛, 蒲巧林, 等. 基于 CSI 张量分解的室内 Wi-Fi 指纹定位方法[J]. 通信学报, 2021, 42(11): 159-171.
- [19] 贾鹏飞, 张月霞. 5G 超密集网络下的室内压缩重构指纹定位算法[J]. 电讯技术, 2021, 61(11): 1419-1423.
- [20] 王正杰, 杨伟丽, 王喆, 等. 基于 CSI 的行为识别研究综述[J]. 计算机工程与应用, 2018, 54(5): 14-23.
- [21] 徐小龙. 物联网室内定位技术[M]. 北京: 电子工业出版社, 2017: 37-39.
- [22] ESTER M, KRIEGEL H P, SANDER J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]. Kdd, 1996, 96(34): 226-231.
- [23] 李文杰, 闫世强, 蒋莹, 等. 自适应确定 DBSCAN 算法参数的算法研究[J]. 计算机工程与应用, 2019, 55(5): 1-7.
- [24] SHIN B, LEE J H, LEE T, et al. Enhanced weighted K-nearest neighbor algorithm for indoor Wi-Fi positioning systems [C]. Proceedings of IEEE International Conference on Computing Technology & Information Management, 2012, 2: 574-577.

### 作者简介

刘宇, 硕士研究生, 主要研究方向为大地测量与卫星导航定位。

E-mail: 2361203186@qq.com

余学祥, 教授, 博士生导师, 主要研究方向为 GPS 变形监测自动化、GPS/GIS 集成、灾害监测与预报。

E-mail: 1064365177@qq.com

谢世成, 硕士研究生, 主要研究方向为空间定位与导航技术。

E-mail: 1179718419@qq.com

刘双, 硕士研究生, 主要研究方向为大地测量与卫星导航定位。

E-mail: 2323866441@qq.com

朱平, 硕士研究生, 主要研究方向为大地测量与卫星导航定位。

E-mail: 599813267@qq.com