

黄慎鹏,张烈平,刘津铭,等.基于改进KNN-ERT的Wi-Fi指纹补全方法[J].辽宁工程技术大学学报(自然科学版),2026,45(3):378-384.DOI:10.11956/j.issn.1008-0562.20250568

HUANG Shenpeng, ZHANG Lieping, LIU Jinming, et al. Wi-Fi fingerprint completion method based on improved KNN-ERT[J]. Journal of Liaoning Technical University (Natural Science), 2026, 45(3): 378-384. DOI: 10.11956/j.issn.1008-0562.20250568

基于改进KNN-ERT的Wi-Fi指纹补全方法

黄慎鹏¹, 张烈平^{2*}, 刘津铭³, 郑新鹏¹, 邵兰青⁴, 王守峰⁵

- (1. 桂林理工大学 机械与控制工程学院, 广西 桂林 541006; 2. 桂林航天工业学院 机电工程学院, 广西 桂林 541004;
3. 桂林航天工业学院 人工智能学院, 广西 桂林 541004; 4. 桂林航天工业学院 计算机科学与工程学院, 广西 桂林 541004;
5. 桂林理工大学南宁分校 电气与电子工程学院, 广西 南宁 532100)

摘要: 针对室内指纹定位中指纹库数据容易丢失的问题, 提出一种融合改进K近邻算法(KNN)和极端随机树算法的数据补全算法。在预处理阶段, 采用中值滤波与高斯滤波相结合的混合滤波策略提升指纹数据质量; 在模型训练阶段, 设计融合欧氏距离与曼哈顿距离的KNN算法确定邻近样本集合, 借助ERT模型对该集合进行训练, 并集成各决策树的输出以得到最终预测值。对补全后的指纹数据库进行数据增强操作, 进一步提升样本多样性。研究表明, 与KNN、ERT等算法相比, 所提算法在数据补全方面表现更优, 补全误差最高降低19%, 预测精度达到93%。研究结论为提高指纹数据库的完整性提供参考。

关键词: K近邻; 极端随机树; 混合滤波; 指纹数据库; 信息补全; 数据增强

中图分类号: TP181; TN92

文献标志码: A

文章编号: 1008-0562(2026)03-0378-07

Wi-Fi fingerprint completion method based on improved KNN-ERT

HUANG Shenpeng¹, ZHANG Lieping^{2*}, LIU Jinming³, ZHENG Xinpeng¹,
SHAO Lanqing⁴, WANG Shoufeng⁵

- (1. School of Mechanical and Control Engineering, Guilin University of Technology, Guilin 541006, China;
2. School of Mechanical and Electrical Engineering, Guilin University of Aerospace Technology,
Guilin 541004, China; 3. School of Artificial Intelligence, Guilin University of Aerospace Technology,
Guilin 541004, China; 4. School of Computer Science and Engineering, Guilin University of Aerospace
Technology, Guilin 541004, China; 5. School of Electrical and Electronic Engineering,
Nanning Branch of Guilin University of Technology, Nanning 532100, China)

Abstract: A completion algorithm based on an improved K-nearest neighbor (KNN) algorithm and extremely randomized trees (ERT) algorithm is proposed to address the problem of data loss in indoor fingerprint localization fingerprint databases. To improve the quality of fingerprint data, a hybrid filtering strategy combining median filtering and Gaussian filtering is employed in the preprocessing process. During the model training phase, the fingerprint data is divided into a training set and a testing set. The KNN algorithm, which combines Euclidean distance and Manhattan distance, is applied to determine the set of neighboring samples. Then, the ERT model is trained this set, and the outputs of all decision trees are integrated to generate the final prediction. Finally, perform data augmentation operations on the completed fingerprint database. The experimental results show that the proposed algorithm is superior in data completion compared with other algorithms, with a maximum completion error reduction of 19% and a prediction accuracy of up to 93%. The research conclusions provide a reference for improving the integrity of the fingerprint database.

Keywords: K-nearest neighbor; extremely randomized trees; hybrid filtering; fingerprint database; information completion; data enhancement

收稿日期: 2025-11-25 修回日期: 2026-03-05 接受日期: 2026-03-16 责任编辑: 朱含露

基金项目: 国家自然科学基金项目(61741303); 广西自然科学基金项目(2025GXNSFAA069942);

广西高校中青年教师科研基础能力提升项目(2023KY0263)

作者简介: 黄慎鹏(2001-), 男, 广西 贵港人, 硕士研究生, 主要从事传感器与智能信息处理技术方面的研究。E-mail: 3197236507@qq.com

通信作者: 张烈平(1971-), 男, 江西 赣州人, 博士, 教授, 主要从事传感器与智能信息处理技术方面的研究。E-mail: zlp@guat.edu.cn

0 引言

在室内定位方法中,基于接收信号强度指示(received signal strength indication, RSSI)的Wi-Fi指纹定位技术,凭借无需额外硬件改造、成本低廉且易于部署等优势,成为室内定位领域的研究和应用热点^[1]。该技术的核心在于构建一个能够准确反映室内空间信号分布特征的离线指纹数据库^[2]。然而,在实际应用过程中,指纹数据库中的RSSI数据极易出现缺失现象^[3],这主要是复杂室内环境的固有特性所致,如环境噪声、多径效应、设备性能差异以及信号遮挡等^[4-5]。这些缺失值不仅直接破坏了数据的完整性和可靠性,更严重削弱了指纹库的内在关联性和结构一致性,导致后续定位匹配算法的精度显著下降。

近年来,针对指纹数据补全算法的相关研究取得一定进展。吕博等^[6]提出基于降噪自编码器超分辨率生成对抗网络的降噪指纹数据库增强方法,提升了指纹数据库的完整性。张雯涛等^[7]将稀疏采样多时间段指纹数据与Kriging插值方法相结合,补全未采样指纹点,提升了建库效率。FOUAD等^[8]设计了一种结合模糊C均值聚类、K近邻(K-nearest neighbor, KNN)算法,以及迭代填充技术的复合数据填补策略,用以修复数据集中的不完整信息。SANJAR等^[9]运用K近邻算法融合特征选取方法应对数据缺失情况。NIKFA LAZAR等^[10]通过决策树与迭代式模糊聚类相结合的方式,完成数据缺失值填补。综上分析,现有补全方法多依赖传统插值技术或者单一机器学习模型进行缺失数据填补,难以充分表征实际环境的复杂空间特性,导致Wi-Fi数据库的填补准确率不高、补全效率低和预测效果不佳,制约室内定位的可靠性和准确性。

针对上述问题,作者提出一种基于K近邻算法与极端随机树(ERT)算法的集成机器学习模型,通过对不完整Wi-Fi指纹数据的精准填补,提升指纹数据库的完整性和准确性。

1 算法设计

1.1 中值滤波和高斯滤波融合的混合滤波

融合中值滤波与高斯滤波的互补优势,有效滤除指纹数据库中的噪声,从而提升指纹数据的准确性。先利用中值滤波识别并剔除异常值,再通过高斯滤波对指纹数据进行平滑处理。

中值滤波通过特定形态的二维滑动窗口遍历图像,将窗口内的像素值按大小排序,生成递增或递减的有序序列^[11]。滤波处理后的RSSI结果为

$$\tilde{r}_t = \text{median}(r_{t-k}, r_{t-k+1}, \dots, r_t, r_{t+k}), \quad (1)$$

式中: r_t 为初始RSSI向量的第 t 个观测值; k 为窗

口半长参数。

在Wi-Fi指纹数据库构建过程中,每个定位节点将采集多个无线接入点(access point, AP)发出的RSSI。通过对每个节点进行连续 N 次测量,可获得对应的RSSI向量,即

$$\mathbf{R}_{\text{raw}} = (r_1, r_2, \dots, r_N). \quad (2)$$

对序列 $\mathbf{R}_{\text{raw}}$ 进行中值滤波处理,获得平滑的RSSI数值序列 $\mathbf{R}_{\text{fil}} = \text{median}(\mathbf{R}_{\text{raw}})$ 。将此滤波结果作为参考点的RSSI数据,保存到指纹数据库中。

高斯滤波通过对邻域内的像素值进行基于高斯分布的加权平均,实现数据的平滑与降噪^[12]。高斯核权重计算式为

$$w_g(i, j) = \exp\left[-\frac{(i-j)^2}{2\sigma_g^2}\right], \quad (3)$$

式中: σ_g 为高斯滤波的标准差; i, j 分别为当前数据点与邻域点的数据序号。

对中值滤波结果 $\mathbf{R}_{\text{fil}}$ 序列执行高斯滤波操作,最终获得平滑后的信号 $\mathbf{R}'_{\text{fil}}$, 即

$$\mathbf{R}'_{\text{fil}} = \frac{\sum_{j \in \Omega} w_g(i, j) \mathbf{R}_{\text{fil}}}{\sum_{j \in \Omega} w_g(i, j)}, \quad (4)$$

式中, Ω 为高斯滤波所选取的邻域集合。

1.2 KNN和ERT算法结合的信息补全方法

(1) KNN数据补全算法

KNN数据补全算法是一种基于邻近样本的缺失值填补策略,该方法通过度量目标样本与邻近数据点之间的相似度(通常采用距离函数),分析样本间的相关性,从而对缺失信息进行填补^[13]。在Wi-Fi指纹补全任务中,该方法的主要执行步骤如下。

步骤1 计算待填补数据与其他全部指纹数据 X_i 间的距离,计算公式为

$$D(\mathbf{X}_{\text{mis}}, \mathbf{X}_i) = \sqrt{\sum_{j=1}^n (\mathbf{X}_{\text{mis},j} - \mathbf{X}_{i,j})^2}, \quad (5)$$

式中: $\mathbf{X}_{\text{mis}}$ 为包含缺失值的指纹数据; n 为Wi-Fi指纹的坐标特征维度; $\mathbf{X}_{\text{mis},j}$ 为 $\mathbf{X}_{\text{mis}}$ 中的第 j 个指纹数据; $\mathbf{X}_{i,j}$ 为 $\mathbf{X}_i$ 中的第 j 个指纹数据, $i = 1, 2, 3, \dots, n$ 。

步骤2 将计算得到的距离值按大小排序,选取与待填补数据距离最近的 C 个样本,构成最近样本集 M 。

步骤3 设 $d_1, d_2, \dots, d_K$ 为 M 中各数据与 $\mathbf{X}_{\text{mis}}$ 之间的距离, $E_1, E_2, \dots, E_K$ 为 $d_1, d_2, \dots, d_K$ 对应数据的RSSI, 计算 $\mathbf{X}_{\text{mis}}$ 的填补值 $\hat{E}_{\text{mis}}$ 为

$$\hat{E}_{\text{mis}} = \left(\sum_{i=1}^K \frac{1}{d_i} E_i \right) / \left(\sum_{i=1}^K \frac{1}{d_i} \right). \quad (6)$$

步骤4 重复步骤3,对Wi-Fi指纹数据库中的所有缺失指纹进行补全。

(2) ERT数据补全算法

ERT算法是一种非线性数据补全方法,通过高度随机性来估计缺失数据值^[14]。在构建决策树时,随机选择特征与特征的分割阈值,从而增强模型的随机性与泛化能力。设完整Wi-Fi指纹库为 D ,其中 D_{com} 为仅包含完整指纹数据的集合, D_{mis} 为仅包含缺失指纹数据的集合。该算法的主要步骤如下。

步骤1 从 D_{com} 中进行有放回抽样(每次记录抽取的样本后将其放回 D_{com}),重复该过程生成多个训练子集,每个子集包含少量指纹数据。每棵决策树的随机训练集为

$$D_{sub} = \frac{1}{T} \sum_{t=1}^T D_t, \quad (7)$$

式中: D_t 为第 t 棵决策树的随机抽样; T 为决策树的数量。

步骤2 使用训练集 D_{sub} 对模型进行训练。

步骤3 重复步骤1和步骤2,根据缺失指纹的数量创建多棵决策树,形成极端随机森林。

步骤4 将 D_{mis} 输入步骤3训练完成的ERT补全模型,得到多棵决策树的预测值,以其算术平均值作为最终的填补值(缺失数据的预测值),即

$$\hat{V}_{mis} = \frac{1}{T} \sum_{t=1}^T T_t(D_{mis}), \quad (8)$$

式中, $T_t(D_{mis})$ 为第 t 棵决策树的估算值。

(3) 算法融合与改进设计

在KNN算法中,传统欧式距离难以准确反映实际环境特性,预测值与真实值的误差较大^[15]。相比之下,曼哈顿距离对异常值不敏感,具有更强的鲁棒性,可有效降低因数据波动带来的偏差。然而,曼哈顿距离仅在特征尺度相近的场景下表现良好,当距离值过大时,其主导作用会掩盖其他特征的相似程度^[16]。因此,提出一种融合曼哈顿距离与欧式距离的联合距离度量方法:首先分别计算两种距离,再取平均值作为最终的距离度量值。

此外,KNN算法和ERT算法虽然能够对不完整数据进行修复,但均存在局限性。KNN方法虽易于实现,但实际场景中信号特征具有多样性和非线性关联,仅通过基础距离计算难以精确表征。ERT算法虽能捕捉数据的整体分布模式,却在处理大规模数据时运算负担显著增加。为此,提出基于联合距离度量的改进KNN-ERT补全算法。首先计算待填补数据与数据库中其他样本的联合距离度量值,通过KNN算法确定邻近样本集合,再利用ERT模型的全局特征学习能力对邻近集合进行训练,强化指纹间的相关性,实现对指纹库的优化补全,计算步骤如下。

步骤1 在离线阶段,采集实验环境中每个

AP的RSSI值。实验场地共部署 m 个AP,第 j 次采集到第 i 个AP的RSSI $_{ij}$ 构成初始指纹数据集,其表达式为

$$r^{ij} = (RSSI_{i1}, RSSI_{i2}, \dots, RSSI_{ij})。 \quad (9)$$

将同一采样点采集到的 m 个AP的RSSI值进行整合,得到指纹数据库为

$$C = (r^1, r^2, \dots, r^m)。 \quad (10)$$

步骤2 为减少数量级差异对补全结果的影响,对指纹数据进行归一化处理^[17],计算式为

$$RSSI_i^* = \frac{RSSI_i - RSSI_{min}}{RSSI_{max} - RSSI_{min}}, \quad (11)$$

式中: $RSSI_{max}$ 、 $RSSI_{min}$ 分别为采集到的RSSI最大值、最小值; $RSSI_i$ 为采集到的第 i 个RSSI值。

步骤3 为减轻异常数据对模型补全的干扰,采用混合滤波算法对指纹数据进行平滑处理。此时,经过步骤2和步骤3处理的指纹数据库因环境因素仍存在一定缺失值。

步骤4 设 $X_{mis} = \{X_1, X_2, \dots, X_n\}$,其中每个样本均由 m 个特征维度组成。计算缺失样本与其他样本 X_{cls} 之间的联合距离度量值为

$$D_{ij}(X_{mis}, X_{cls}) = \frac{1}{2} \left(\sum_{b=1}^n |X_{mis,a} - X_{cls,b}| + \sqrt{\sum_{a=1}^n (X_{mis,a} - X_{cls,b})^2} \right), \quad (12)$$

式中: $\sum_{b=1}^n |X_{mis,a} - X_{cls,b}|$ 为第 a 个待填补指纹与第 b 个

其他数据间的曼哈顿距离; $\sqrt{\sum_{a=1}^n (X_{mis,a} - X_{cls,b})^2}$ 为第 a 个待填补指纹与第 b 个其他数据间的欧氏距离。

步骤5 遍历缺失指纹数据库 X_{mis} 中的每个待填补点,计算该点与所有其他数据点 X_{cls} 之间的距离,筛选出与待填补点最接近的 C 个样本。

步骤6 利用完整的指纹数据库,采用联合距离度量与最优近邻参数 K 值,对KNN模型进行训练。通过训练完成的KNN模型,计算获得与缺失数据距离最近的 C 个邻近样本,构成邻近样本集为

$$N(C) = \{p_1, p_2, \dots, p_C\}。 \quad (13)$$

步骤7 将步骤6得到的邻近样本集与含缺失指纹的数据库进行合并,将合并后的指纹库用于训练ERT模型。采用有放回随机抽样的方式,从合并指纹库中生成多个训练子集,选取最具代表性的特征进行训练,重复此过程,直至所有特征均参与训练或训练结果收敛,完成多决策树ERT模型的训练与构建。

步骤8 利用训练完成的ERT模型对缺失数据进行估计,以各决策树估计值的平均值作为缺失数据点的最终填补值,填补点的预测值为

$$\hat{X}_{\text{mis}} = \frac{1}{n} \sum_{t=1}^n T_t(X_{\text{input}}), \tag{14}$$

式中： $T_t(X_{\text{input}})$ 为第 t 棵树对填补点的估算值； n 为决策树的数量。

1.3 数据增强的信息补全

由于室内环境干扰因素众多，不同时段与位置的 RSSI 往往存在差异。本文引用数据增强技术，通过构造多元化样本分布，提升模型对各类信号波动场景的学习能力与抗干扰能力^[18]。对改进 KNN-ERT 模型补全后的数据集实施数据增强，具体步骤如下。

- 步骤 1 对补全后的数据指纹集进行复制，得到两个完全相同的指纹数据集。
- 步骤 2 对其中一个指纹数据集进行数据增强，为每一个 RSSI 值叠加标准差为 0.1、平均值为 0 的高斯噪声。
- 步骤 3 将原始数据集与增强后的数据集整合为一个新数据集。
- 改进算法流程见图 1。

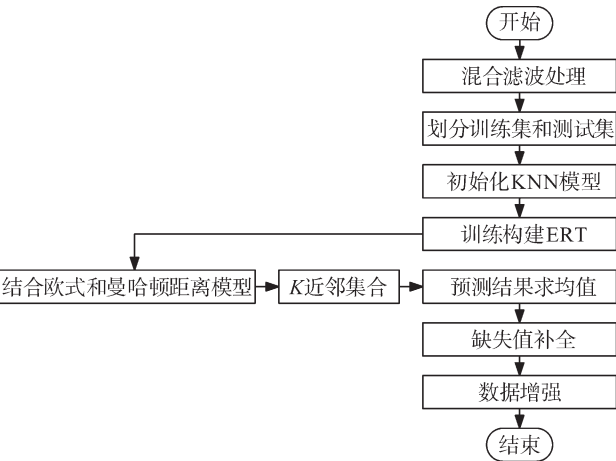


图 1 改进算法流程

Fig.1 flow chart of the improved algorithm

2 实验环境与数据采集

2.1 实验环境搭建

实验场地设于校内某实验楼内，面积约为 100 m²。选取 6 个不同位置，分别部署型号为 MW305R 的无线路由器及其信号接收装置，各设备持续向接收装置传输信号。完成设备布置后，按网格划分方式，以 1 m 为间隔将实验区域均匀划分为 100 个采样点位，各采样点均位于网格中心点，用于采集 6 个 Wi-Fi 热点的 RSSI 值，其具体位置如图 2 中绿色圆点所示。

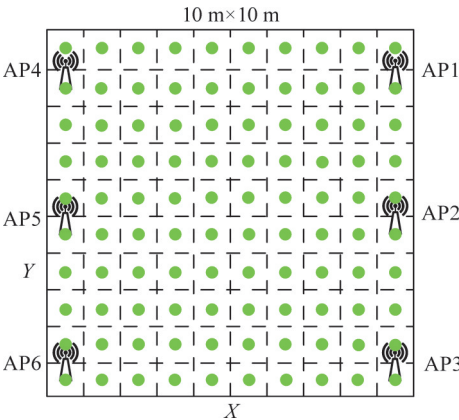


图 2 实验环境布置示意

Fig.2 schematic of the experimental environment layout

2.2 指纹信息采集

为完成指纹数据的采集，在手机上安装自主开发的 Wi-Fi 指纹采集软件，实时采集 6 个 AP 的 RSSI 值，部分数据见表 1。

表 1 Wi-Fi 数据库结构

坐标	RSSI 值/dBm					
	AP1	AP2	AP3	AP4	AP5	AP6
(30,1)	-43.88	-45.54	-66.79	-50.32	-55.29	-49.11
(32,1)	-45.88	-44.29	-61.35	-52.94	-55.04	-48.82
(34,1)	-44.79	-43.19	-66.56	-49.89	-51.23	-51.28
...	...	...	...	...	...	...
(58,1)	-42.25	-42.47	-68.29	-53.00	-52.95	-50.76
(60,1)	-40.82	-43.50	-67.73	-51.89	-53.36	-50.24

3 实验与结果分析

3.1 实验参数设置

不同接入点的参考位置均存在 RSSI 值缺失现象，且缺失程度存在差异，致使指纹库数据不完整。存在缺失值的部分指纹数据见表 2。

表 2 各采样点缺失指纹数据

坐标	RSSI 值/dBm					
	AP1	AP2	AP3	AP4	AP5	AP6
(30,1)	-43.88	-44.66	-66.70	-51.42	-	-49.11
(32,1)	-	-43.84	-	-52.94	-54.14	-48.82
(34,1)	-44.79	-42.99	-66.64	-54.31	-51.23	-
...	...	...	...	...	...	...
(58,1)	-42.25	-42.47	-68.29	-	-	-50.76
(60,1)	-40.82	-	-67.73	-51.89	-53.36	-50.24

注：“-”代表指纹数据缺失。

对采集到的指纹数据进行混合滤波处理，完成 KNN 模型初始化。将指纹数据集按 8 : 2 比例划分为训练集和测试集，输入改进 KNN-ERT 融合模

型完成缺失数据填补。对填补后的数据进行数据增强。实验时重点调节每棵决策树的最小叶子节点数、随机特征比例、最小样本数和最大树深。对于关键参数 K 的优化,则通过信息熵进行多维度性能评估,结果见表3。

表3 参数 K 设定和实验结果
Tab.3 parameter K settings and experimental results

K	随机特征比例	最小叶子节点数	最大树深	最小样本数	信息熵
10	0.8	4	-1	6	0.840
20	0.8	4	-1	6	0.815
30	0.8	4	-1	6	0.752
50	0.8	4	-1	6	0.781
80	0.8	4	-1	6	0.801

信息熵数值降低时,预测性能显著提升。据此,根据表3确定改进补全模型的最优参数为:随机特征比例为0.8,最小叶子节点数为4,最大树深为-1,最小样本数为6, K 为30。

3.2 混合滤波性能分析

为验证所提混合滤波策略对指纹数据稳定性的影响,采集AP1和AP4在同一位置的30组数据,分别采用原始数据、单一中值滤波、中值-高斯混合滤波3种预处理方式开展对比实验,实验结果见图3。

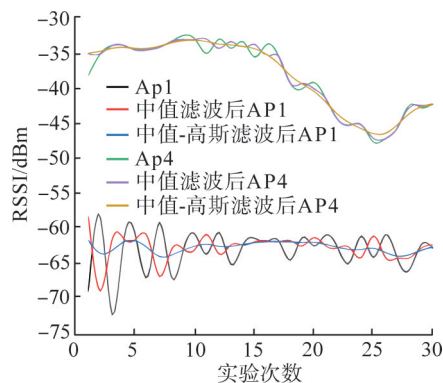


图3 不同滤波方式效果对比

Fig.3 comparison of the effects of different filtering methods

由图3可见,经混合滤波处理后,AP1与AP4的RSSI曲线均呈现更平滑的趋势,波动幅度得到有效抑制。AP1的RSSI值由处理前的-71.98~-62.53 dBm缩小至-62.59~-61.17 dBm;AP4的RSSI则由-46.65~-31.57 dBm缩减为-45.13~-34.78 dBm。由此可见,混合滤波能够在一定程度上滤除指纹集中的干扰,提高指纹数据的准确性。

3.3 补全性能分析

基于3.1节确定的参数配置,对表2中的缺失指纹数据进行补全,补全后的数据见表4。

表4 补全后的Wi-Fi指纹库
Tab.4 completed Wi-Fi fingerprint library

坐标	RSSI值/dBm					
	AP1	AP2	AP3	AP4	AP5	AP6
(30,1)	-43.88	-44.66	-66.70	-51.42	-52.84	-49.11
(32,1)	-44.11	-43.84	-67.84	-52.94	-54.14	-48.82
(34,1)	-44.79	-42.99	-66.64	-54.31	-51.23	-49.84
...	...	...	...	...	...	...
(58,1)	-42.25	-42.47	-68.29	-53.00	-52.95	-50.76
(60,1)	-40.82	-43.50	-67.73	-51.89	-53.36	-50.24

为评估所提联合距离度量方式的性能,分别对单一欧式距离、单一曼哈顿距离以及联合距离度量方式开展对比实验。实验过程中,为模拟实际复杂环境下的数据丢失情形,对指纹数据库中的部分数据进行随机删除,该组样本RSSI真实值为-56.66 dBm,三类方式均采用KNN算法对缺失数据进行补全,通过比对预测值与实际删除值之间的偏差,实现对其预测性能的量化评估。不同距离度量方式的预测效果对比见图4。

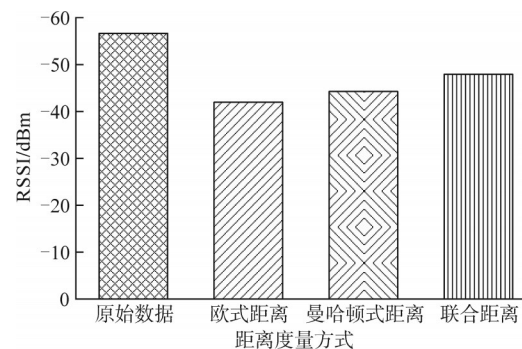


图4 不同距离度量方式效果对比

Fig.4 comparison of different distance measurement effects

由图4可知,联合距离度量方式对缺失数据的预测值为-47.92 dBm,预测精度达84.57%,最接近真实值。曼哈顿距离度量方式的预测精度次之,准确率为78.1%,与真实值相差12.4 dBm。而欧式距离度量方式的预测效果最差,误差为14.69 dBm,准确率仅为74%。对比实验结果表明,相较于单一距离度量方式,联合距离度量方式可有效提升缺失数据的估计精度。

为进一步评估所提KNN-ERT填补算法的性能,选取KNN模型、ERT模型、改进KNN-RF模型^[19]、GAIN模型^[20]、KNN-ERT模型和改进KNN-ERT模型,在相同数据集划分、缺失率设置与评

估指标下开展数据补全对比实验。各模型的预测效果见图 5。

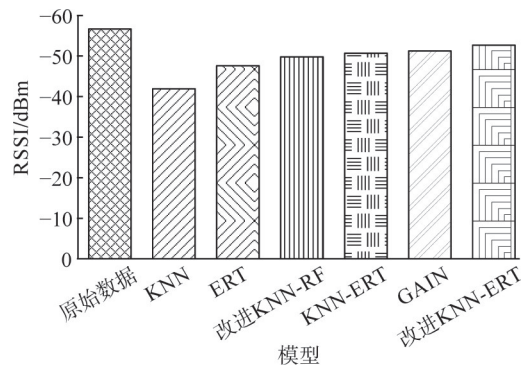


图 5 不同模型预测效果

Fig.5 prediction performance of different models

由图 5 可知, 改进 KNN-ERT 模型的填补预测

值为-52.69 dBm, 误差仅为 3.97 dBm, 预测准确率高达 92.9%, 预测结果最接近真实值。GAIN 模型的填补效果次之, 误差为 5.42 dBm, 预测准确率为 90.4%。KNN-ERT 模型的预测精度较差, 准确率为 89.4%, 与真实值相差 5.96 dBm。改进 KNN-RF 模型的预测准确率为 87.8%, ERT 预测准确率为 83.8%。而 KNN 模型的填补效果最差, 其估算值与真实值的误差高达 14.73 dBm, 准确率仅为 73.9%。实验结果表明, 改进 KNN-ERT 模型相较于其他模型, 预测准确率最高。

3.4 数据增强对定位性能的影响分析

完成混合滤波与改进 KNN-ERT 模型填补后, 对补全后的 Wi-Fi 数据库进行数据增强处理。增强前后的部分 Wi-Fi 指纹数据见表 5。

表 5 数据增强前后部分指纹数据对比

Tab.5 comparison of partial fingerprint data before and after data enhancement

坐标	RSSI 值/dBm					
	AP1	AP2	AP3	AP4	AP5	AP6
(30,1)	-43.88/-41.20	-44.66/-51.56	-66.70/-64.40	-51.42/-52.99	-52.84/-52.11	-49.11/-52.67
(32,1)	-44.11/-39.52	-43.84/-52.78	-67.84/-63.61	-52.94/-50.63	-54.14/-51.27	-48.82/-50.97
(34,1)	-44.79/-37.99	-42.99/-49.80	-66.64/-63.42	-54.31/-52.49	-51.23/-53.20	-49.84/-52.83
...	...	...	...	...	...	...
(58,1)	-42.25/-46.78	-42.47/-42.85	-68.29/-64.77	-53.00/-54.01	-52.95/-53.48	-50.76/-52.96
(60,1)	-40.82/-47.88	-43.50/-42.39	-67.73/-67.94	-51.89/-51.87	-53.36/-53.46	-50.24/-50.69

注:“/”前为处理前 RSSI 值,“/”后为处理后 RSSI 值。

为全面评估混合滤波、改进 KNN-ERT 数据填补方法及数据增强对定位精度的影响, 设置 6 种数据预处理方式开展对比实验。方式 1: 不进行处理; 方式 2: 高斯滤波+改进 KNN-RF; 方式 3: 混合滤波+改进 KNN-ERT; 方式 4: 混合滤波处理; 方式 5: 混合滤波+GAIN; 方式 6: 混合滤波+改进 KNN-ERT+数据增强。每种方式处理后的指纹数据均采用 KNN 模型进行位置估计。不同预处理方式的定位误差结果见图 6。

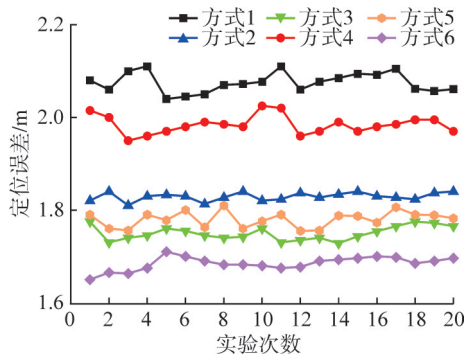


图 6 不同预处理方式的定位误差

Fig.6 positioning errors of different preprocessing methods

由图 6 可见, 方式 6 较其他预处理方法, 平均定位误差降低 0.39 m, 定位准确率提升 23.9%, 在稳定性和准确度方面均得到明显改善。

不同组合方式的误差箱线图见图 7。

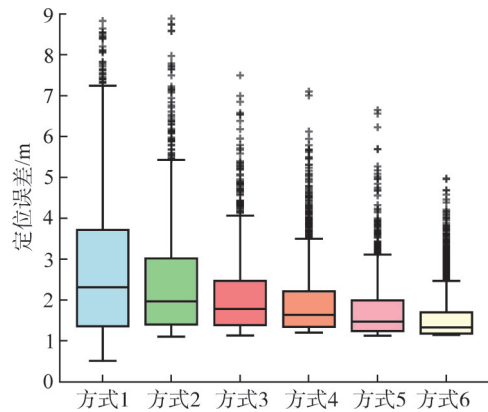


图 7 error box plot of different combination methods

由图 7 可见, 方式 6 的四分位数间距最小, 定位误差分布最为集中, 其中位误差约为 1.5 m, 异常值最少, 误差范围最小。方式 3 的误差范围略大于前者, 异常值更多, 但其中位误差仍处于较低

水平,表明该算法保持了相对较高的定位精度与较好的性能。方式5的误差分布相对紧凑,表现出良好的稳定性,其中位误差约为1.8 m,且异常值较多,说明其定位精度较低且稳定性不足。方式2的误差分布较宽,箱体较大,表明其定位精度有限。方式4的误差分布范围更为宽泛,箱体偏大,定位精度不佳。方式1表现最差,箱形图最宽,中位误差约为2.2 m,波动显著,信号波动明显,定位精度最低。上述实验结果表明,方式6的预处理方法,在定位误差幅值与误差分布方面均优于其他对比方法。

4 结论

针对Wi-Fi信号RSSI值的波动特性,提出一种结合中值滤波与高斯滤波的混合滤波算法,用于剔除指纹数据库中的异常值。通过融合欧氏距离与曼哈顿距离的改进KNN算法,筛选出近邻数据集,并将其输入ERT模型进行训练。利用训练完成的ERT模型(包含多棵决策树)进行预测,以各决策树估计值的算术平均值作为缺失数据点的填补值。对填补后的Wi-Fi指纹库进行数据增强处理。实验结果表明,相较于KNN、ERT、改进KNN-RF、KNN-ERT及GAIN算法,本文算法的预测准确率最高,整体性能更优。后续研究将进一步优化算法,探索自适应优化与多模态数据融合方法,以进一步提升补全算法的综合性能。

参考文献(References):

- [1] ZHANG L P, CHEN Y, GAO X L, et al. An algorithm for indoor positioning based on LightGBM+ExtraTrees chaotic weighted ensemble: evaluation and comparison[J]. The Journal of Supercomputing, 2025, 81(7): 812.
- [2] LU Y W, MA Y T, HAN Y X, et al. DMLF: device-free multitarget adaptive backscatter link fingerprint localization[J]. IEEE Transactions on Instrumentation and Measurement, 2025, 74: 5501111.
- [3] 古玉锋, 杜雨洁, 王育阳, 等. 基于神经网络与自调节卡尔曼滤波的超宽带定位算法研究[J]. 中国机械工程, 2023, 34(12): 1504-1511. GU Yufeng, DU Yujie, WANG Yuyang, et al. Study on UWB positioning algorithm based on neural networks and self-adjusting Kalman filters[J]. China Mechanical Engineering, 2023, 34(12): 1504-1511.
- [4] ZHANG T W, ZHANG P, KALATHAS P, et al. A machine learning approach to improve ranging accuracy with AoA and RSSI[J]. Sensors, 2022, 22(17): 6404.
- [5] ZHANG L C, ZHANG W C, BAO X. Visible light fingerprint database recovery algorithm based on CP decomposition[J]. Optics express, 2023, 31(14): 22885-22902.
- [6] 吕博, 周蓉, 张甜楠, 等. 基于SRGAN-DAE的室内定位指纹生成[J]. 电子测量技术, 2025, 48(3): 154-160.
- LYU Bo, ZHOU Rong, ZHANG Tianyu, et al. Indoor localization fingerprint generation based on SRGAN-DAE[J]. Electronic Measurement Technology, 2025, 48(3): 154-160.
- [7] 张雯涛, 吴飞, 朱海, 等. 基于FDE-IRF的室内指纹定位方法[J]. 全球定位系统, 2021, 46(4): 117-126. ZHANG Wentao, WU Fei, ZHU Hai, et al. Indoor fingerprint localization method based on FDE-IRF[J]. GNSS World of China, 2021, 46(4): 117-126.
- [8] FOUAD K M, ISMAIL M M, AZAR A T, et al. Advanced methods for missing values imputation based on similarity learning[J]. PeerJ Computer Science, 2021, 7: e619.
- [9] SANJAR K, BEKHZOD O, KIM J, et al. Missing data imputation for geolocation-based price prediction using KNN - MCF method[J]. ISPRS International Journal of Geo-Information, 2020, 9(4): 227.
- [10] NIKFALAZAR S, YEH C H, BEDINGFIELD S, et al. Missing data imputation using decision trees and fuzzy clustering with iterative learning[J]. Knowledge and Information Systems, 2020, 62(6): 2419-2437.
- [11] SONIYA S L, RAJ T A B. An intelligent system for lung CT image denoising using a hybrid WT-NLM filter[J]. Automatika, 2025, 66(2): 188-200.
- [12] 杨旭升, 李唯诣, 张文安. 面向RTK定位的整数约束型渐进高斯滤波方法[J]. 自动化学报, 2025, 51(2): 366-375. YANG Xusheng, LI Weiyi, ZHANG Wenan. Integer-constrained progressive Gaussian filtering method for RTK positioning[J]. Acta Automatica Sinica, 2025, 51(2): 366-375.
- [13] CHANDRAN S, GOKARAJU R, NARENDRA K. An extended impedance-based fault location algorithm in power distribution system with distributed generation using synchrophasors[J]. IET Generation, Transmission & Distribution, 2024, 18(3): 479-490.
- [14] AMINIFAR A, SHOKRI M, RABBI F, et al. Extremely randomized trees with privacy preservation for distributed structured health data[J]. IEEE Access, 2022, 10: 6010-6027.
- [15] ÇETIN A İ, BÜYÜKLÜ A H. A new approach to K-nearest neighbors distance metrics on sovereign country credit rating[J]. Kuwait Journal of Science, 2025, 52(1): 100324.
- [16] WANG X W, YUE Y, ZHANG F, et al. Active detection of interphase faults in distribution networks based on energy relative entropy and Manhattan distance[J]. Electric Power Systems Research, 2025, 241: 111397.
- [17] LI H L, QIAN Z H, LIU G Q, et al. NQRELoc: AP selection via nonuniform quantization RSSI entropy for indoor localization[J]. IEEE Sensors Journal, 2022, 22(10): 9724-9732.
- [18] 石军, 王天同, 朱子琦, 等. 基于深度学习的医学图像分割方法综述[J]. 中国图象图形学报, 2025, 30(6): 2161-2186. SHI Jun, WANG Tiantong, ZHU Ziqi, et al. Deep learning-based medical image segmentation methods[J]. Journal of Image and Graphics, 2025, 30(6): 2161-2186.
- [19] 张烈平, 陈耀, 郑新鹏, 等. 基于改进KNN-RF的信息补全算法[J]. 电子测量技术, 2024, 47(13): 74-80. ZHANG Lieping, CHEN Yao, ZHENG Xinpeng, et al. Information completion algorithm based on improved KNN-RF[J]. Electronic Measurement Technology, 2024, 47(13): 74-80.
- [20] DONG W N, FONG D Y T, YOON J S, et al. Generative adversarial networks for imputing missing data for big data clinical research[J]. BMC Medical Research Methodology, 2021, 21: 78.