

# 基于随机子集抽样的高效室内指纹定位算法

雷若兰, 乐燕芬

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

**摘要:** 针对复杂室内环境中无线信号受传播环境突发噪声影响而存在异常的情况, 提出了一种基于在线粗定位和随机子集抽样的 WiFi 指纹定位算法。首先, 该算法利用接入点(AP)信号的特定覆盖区域, 将目标快速、准确地定位到局部区域, 确定目标真实位置附近的候选参考点(RP); 随后, 采用随机子集抽样获取多个子集内的信号距离, 对其进行均值处理, 由此降低异常 AP 信号对信号距离的影响; 最后, 采用加权 K 近邻算法确定目标的估计位置。在实际室内环境采集的公共数据集中开展实验, 对在线粗定位和随机子集抽样方法的准确度进行分析, 验证了所提算法的高效性, 给出了参数选择对于定位性能的影响, 并将本算法与其他指纹定位算法进行了对比。结果表明, 在线粗定位方法能有效获取用户真实位置附近的候选 RP, 随机子集抽样能有效抑制大的定位误差, 即使在 AP 信号出现异常时, 也能获得优于已有算法的定位性能。

**关键词:** 指纹定位; AP 覆盖向量; 在线粗定位; 随机子集抽样; AP 异常

**中图分类号:** TP 391 **文献标志码:** A

## Efficient indoor fingerprinting localization algorithm based on random subset sampling

LEI Ruolan, LE Yanfen

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

**Abstract:** In view of the fact that the strength of wireless signals in complex indoor environments has anomalies due to the influence of the sudden noise in the propagation environment, an algorithm for WiFi fingerprinting localization based on online coarse positioning and random subset sampling was proposed. First, the algorithm utilized the specific coverage characteristics of access point (AP) signals to rapidly and accurately constrain the target's position to a local region, thereby identifying candidate

收稿日期: 2023-12-29

基金项目: 国家自然科学基金资助项目 (62172281)

第一作者: 雷若兰(1999—), 女, 硕士研究生。研究方向: 室内定位、信息隐藏。E-mail: 1665844775@qq.com

通信作者: 乐燕芬(1978—), 女, 副教授。研究方向: 物联网、室内定位、信息隐藏。E-mail: leyanfen@usst.edu.cn

引文格式: 雷若兰, 乐燕芬. 基于随机子集抽样的高效室内指纹定位算法[J]. 上海理工大学学报, 2025, 47(2): 230-238.

Citation: LEI Ruolan, LE Yanfen. Efficient indoor fingerprinting localization algorithm based on random subset sampling[J]. Journal of University of Shanghai for Science and Technology, 2025, 47(2): 230-238.

reference points (RP) near the true location. Subsequently, random subset sampling was employed to generate multiple subsets of signal distances, which were then averaged to mitigate the influence of abnormal AP signals on distance measurements. Finally, a weighted K-nearest neighbors algorithm was applied to estimate the target's position. Experiments were conducted using a publicly available dataset collected from real indoor environments to evaluate the accuracy of both online coarse positioning and random subset sampling. The results demonstrate the effectiveness of the proposed algorithm and reveal the impact of parameter selection on localization performance. Comparative studies with existing fingerprint-based localization algorithms show that the online coarse localization method effectively identifies candidate reference points near the user's true position, while the random subset sampling technique significantly reduces large localization errors. Even when AP signals exhibit abnormalities, the proposed algorithm maintains superior positioning accuracy compared to conventional algorithms.

**Keywords:** *fingerprinting localization; AP covering vector; online coarse positioning; random subset sampling; AP abnormality*

随着互联网和智能终端的快速发展,基于位置服务(location based services, LBS)受到各行业的广泛关注。在开放的室外环境中,全球定位系统(global positioning system, GPS)能够提供高精度的定位服务,但在复杂的室内环境中,GPS信号会快速衰减,无法完成有效定位<sup>[1]</sup>。近年来,研究人员提出了多种不同类型的无线信号,如蓝牙<sup>[2]</sup>、WiFi<sup>[3]</sup>、超宽带(ultra wide band, UWB)<sup>[4]</sup>、射频识别(radio frequency identification, RFID)<sup>[5]</sup>等,用于实现不同定位场景中的室内定位。

基于WiFi指纹的室内定位技术由于低成本、低功耗、易部署的特点成为研究热点。WiFi指纹定位包括两个阶段:离线阶段和在线阶段。离线阶段:定位服务提供商(localization service provider, LSP)在室内环境中采集参考点(reference point, RP)中来自各接入点(access point, AP)的接收信号强度(received signal strength, RSS),并与其对应的位置信息一起构建离线指纹数据库;在线阶段:实时获取目标对象的RSS向量,通过匹配算法与离线数据库进行比较,求得目标估计位置。

最早基于WiFi的定位研究出现于RADAR系统<sup>[6]</sup>中,服务商通过计算不同的RSS向量之间的距离,采用K近邻<sup>[7]</sup>算法得到定位结果。后续研究为改善性能,引入加权K近邻<sup>[8]</sup>(weighted K-nearest neighbor, WKNN)算法来优化权重。随着机器学习算法的发展,神经网络<sup>[9]</sup>、回归模型<sup>[10]</sup>、贝叶斯网络模型<sup>[11]</sup>、支持向量机<sup>[12-13]</sup>等方法被用到定位中来完成目标位置估计,但这类算法通常

需要离线训练大量数据,复杂度较高。且定位环境的变化,如短时布局变动、人员短时活动等,会导致在线RSS出现瞬时大的波动,此时,在线指纹与离线指纹匹配度较低,离线训练模型的准确性会有一定的降低,引起定位误差的增大。针对动态环境中信号存在大波动的应用场景,He等<sup>[14]</sup>首次提出了一种基于指纹子空间的定位方案(LAAFU),通过构建多个指纹子集来生成多个估计位置。考虑到当指纹子集中包含异常AP信号时,生成的估计位置会远离目标的真实位置,该方案引入了AP(affinity propagation)聚类算法来确定可信度最高的若干参考点。该方法无需离线阶段训练定位模型,但要求获得足够数量的指纹子集才能形成有效聚类。针对LAAFU的完全随机子集生成方式,乐燕芬等<sup>[15]</sup>对其进行改进,综合考虑了目标可接收和不可接收的AP信号对定位的影响,提出了新的子集生成方式。但该方式同样需要生成足够数量的估计位置来形成有效的分簇,不适用于AP数量不多的定位环境。

为减小定位过程中的计算开销,也有文献提出对目标进行粗定位,用计算开销小的算法首先把目标定位到某一局部区域内,再进行二次精确定位。如田洪亮等<sup>[16]</sup>提出基于K-means算法对离线指纹RSS进行分簇的方法,计算目标RSS与各个簇中心向量的距离,选取相似度最高的目标簇,取簇中RP作为粗定位的候选RP。基于RSS信号特征分簇,同一簇内RP可能出现信号特征相似但实际物理位置较分散的情况。Feng等<sup>[17]</sup>采用

AP 聚类算法对参考位置点进行分簇,仅考虑各指纹中对应 RSS 值较高的 AP,以确定其对应的 AP 集合,通过 AP 聚类实现分簇,达到粗定位目的。该方案要求离线阶段 AP 信号分布较稳定,以确定较为准确的 AP 集合。Khalajmehrabadi 等<sup>[18]</sup>利用离线指纹的 AP 覆盖向量对其进行分簇,在线阶段对目标 AP 覆盖向量进行簇匹配,覆盖向量相似的两点通常物理位置相近。但是,当目标接收的 RSS 信号出现波动或者目标处于多个簇的临界位置时,以上方案可能出现簇的误匹配问题,由于后续是在粗定位簇内的候选 RP 中进行二次精确定位,上述情况可能导致最终定位结果出现较大误差。

针对上述问题,本文提出了一种基于 AP 覆盖向量的在线粗定位方法。考虑到具有相似覆盖向量的两点物理位置通常较接近,将目标的 AP 覆盖向量与离线指纹的 AP 覆盖向量进行距离比较,选定  $N$  个最近邻作为候选 RP。此时,各 RP 物理位置较接近,有效避免了上述相关方案中由信号波动而引起的定位误差,且无需进行离线分操作以及面对在线阶段簇的误匹配问题,大幅度减小了二次精确定位阶段指纹匹配的计算开销。同时,引入一种新的随机子集抽样方法来选取更优  $K$  近邻,降低外界突发噪声的干扰。该算法相较于 LAAFU,无需形成足够多的估计位置,定位过程只涉及一次位置估计,可以有效抑制大的定位误差,提高了定位性能。

## 1 算法框架和设计

本文提出了一种基于在线粗定位和随机子集抽样的指纹定位算法。离线阶段, LSP 建立各 RP 对应的指纹数据库;在线阶段,目标对象采集自身 RSS 向量,服务器端计算得到相应的 AP 覆盖向量,同时计算其与离线 RP 的 AP 覆盖向量的 Jaccard 相似度,选取相似度较高的  $N$  个 RP 作为候选 RP 集,实现粗定位。然后,服务器端随机生成一个 AP 子集索引向量,对目标对象的 RSS 向量进行子集抽样操作。基于该 AP 索引,计算第  $s$  个候选 RP 的离线指纹与目标指纹子集向量间的信号距离  $d_s$ ,  $s \in [1, N]$ 。重复  $M$  次子集抽样过程,各候选 RP 与目标间得到  $M$  个距离,对  $M$  个距离进行均值处理,将此均值作为两者间的最终信号

距离  $D_s$ 。方案中采用多次子集抽样形成的平均信号距离,可降低目标 RSS 信号中可能包含的异常值对真实最近邻选择的影响。本文所提出的方案依据  $D_s$  选取  $K$  近邻,采用 WKNN 算法求得目标估计位置。算法流程图如图 1 所示。

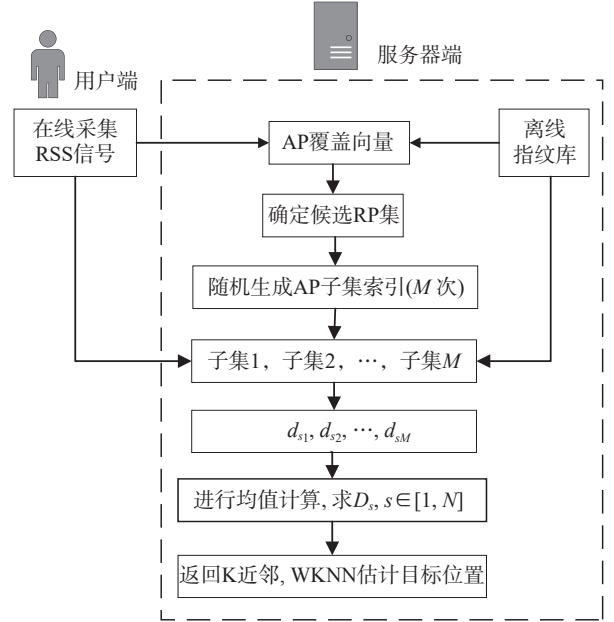


图 1 定位算法流程图

Fig.1 Flow chart of positioning algorithm

### 1.1 离线阶段

LSP 采集各 RP 的位置信息以及来自所有 AP 的 RSS 信号,构成原始离线指纹数据库。假设目标监测区域部署有  $m$  个 RP、 $n$  个 AP,构建的指纹库中第  $i$  条指纹可表示为

$$F_i = [L_i, R_i] = [x_i, y_i, r_{i1}, r_{i2}, \dots, r_{in}] \quad (1)$$

式中:  $L_i = [x_i, y_i]$ , 表示第  $i$  条指纹的位置;  $R_i$  表示第  $i$  个 RP 接收的来自所有 AP 的 RSS 向量;  $r_{in}$  表示来自第  $n$  个 AP 的 RSS。对于未检测到信号强度的 AP,将其信号值定义为 0。同时,由 RSS 指纹信息可确定每个 AP 覆盖向量。设第  $i$  个离线指纹的覆盖向量为

$$C_i = (c_{i1}, c_{i2}, \dots, c_{in}) \quad (2)$$

式中,  $c_{in}$  为二进制向量值,如果该 AP 处 RSS 值不为 0,  $c_{in}$  就为 1,否则取 0。

### 1.2 在线阶段

在线阶段, LSP 首先要利用覆盖向量的相似性确定候选 RP 集,完成目标粗定位;然后再通过基于子集抽样的二次精确定位,完成对目标的最终位置估计。

### 1.2.1 确定候选 RP

当原始离线指纹库的数据量较大时, 指纹匹配过程需要在整个监控区域内进行, 这会造成较大的计算开销。本文通过下述方法确定一个候选 RP 集, 将目标定位在局部区域内, 以减小后续定位过程中目标指纹与离线指纹匹配时的计算开销。在线阶段, 假设目标接收到的 RSS 向量为

$$\mathbf{R}_t = [r_{t1}, r_{t2}, \dots, r_{tn}] \quad (3)$$

对目标指纹也采用与离线指纹类似的方法, 确定当前 AP 的覆盖向量  $\mathbf{C}_t$ , 表示为

$$\mathbf{C}_t = (c_{t1}, c_{t2}, \dots, c_{tn}) \quad (4)$$

如果  $r_{tn}$  不为 0,  $c_{tn}$  就为 1, 否则取 0。

在复杂的室内环境中, AP 无线信号的衰减特性与环境布局密切相关, 且都有其特定的信号覆盖区域。因此, 覆盖向量可以表明该位置处 AP 的分布情况。为了考察每个 AP 的覆盖区域, 现选取某实际室内场景的公开数据集<sup>[19]</sup>中的 2 个 AP (AP23 和 AP29) 进行实验, 绘制出其信号覆盖情况, 如图 2 所示。可以发现, 被同一 AP 所覆盖的 RP 位置相近, 说明具有较相似覆盖向量的两点其物理位置通常较接近。

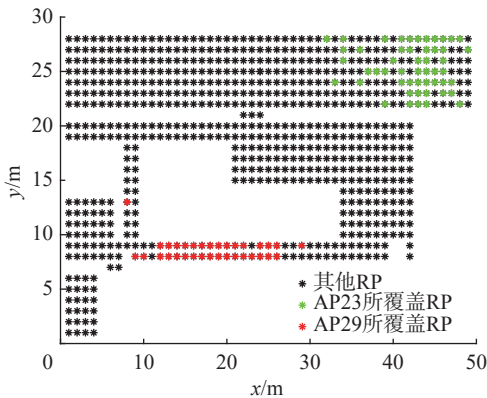


图 2 不同 AP 的信号覆盖情况

Fig.2 Signal coverage of different access points

本文采用 Jaccard 系数来衡量  $\mathbf{C}_i$  和  $\mathbf{C}_t$  的相似度, Jaccard 系数通常用于比较有限样本集之间的相似性与差异性。Jaccard 系数越大, 说明  $\mathbf{C}_i$  和  $\mathbf{C}_t$  在相同位置有更多的相同元素, 即可接收或不可接收的 AP 集越相似, 说明目标对象与该参考点位置越接近。本方案提出计算  $\mathbf{C}_i$  与  $\mathbf{C}_t$  之间的 Jaccard 相似度, 表示为

$$J(\mathbf{C}_t, \mathbf{C}_i) = \frac{|\mathbf{C}_t \cap \mathbf{C}_i|}{|\mathbf{C}_t \cup \mathbf{C}_i|} \quad (5)$$

对于两个  $n$  维二进制向量, 它们之间的

Jaccard 相似度计算可转化为

$$J(\mathbf{C}_t, \mathbf{C}_i) = \frac{\mathbf{C}_t \cdot \mathbf{C}_i}{|\mathbf{C}_t| + |\mathbf{C}_i| - \mathbf{C}_t \cdot \mathbf{C}_i} \quad (6)$$

式中:  $|\mathbf{C}_t| = \mathbf{C}_t \cdot \mathbf{C}_t$ ;  $|\mathbf{C}_i| = \mathbf{C}_i \cdot \mathbf{C}_i$ 。

利用式 (6) 计算用户与指纹库内所有 RP 覆盖向量的相似度, 选取相似度较高的前  $N$  个覆盖向量对应的 RP 作为候选 RP, 其所在区域构成了目标的粗定位区域。候选 RP 的 RSS 向量可以表示为

$$\mathbf{R}_s = [r_{s1}, r_{s2}, \dots, r_{sn}] \quad (7)$$

### 1.2.2 二次精确定位

本文所提出的二次精确定位算法具体步骤如下:

#### a. AP 随机子集抽样

对目标接收到的信号  $\mathbf{R}_t$  进行 AP 子集抽样, 根据随机生成的长度为  $l$  ( $l < n$ ) 的 AP 索引向量来选取子集, 则目标指纹的一个随机子集抽样结果可表示为

$$\mathbf{R}'_t = [r_1, r_2, \dots, r_l] \quad (8)$$

根据所抽取的 AP 子集, 任一候选 RP 相应的指纹子集表示为

$$\mathbf{R}'_s = [r'_{s1}, r'_{s2}, \dots, r'_{sl}] \quad (9)$$

式中,  $s \in [1, N]$ 。

计算  $\mathbf{R}'_t$  与  $\mathbf{R}'_s$  信号空间的余弦距离, 得到第  $s$  个候选 RP 的离线指纹与目标指纹子集向量间的信号距离  $d_s$ :

$$d_s = 1 - \cos(\mathbf{R}'_t, \mathbf{R}'_s) \quad (10)$$

重复该子集随机抽样过程  $M$  次, 则各离线指纹对于目标指纹均求得  $M$  个距离, 对这  $M$  个距离求取平均值, 得到两者间的最终距离  $D_s$ :

$$D_s = \frac{1}{M} \sum_{j=1}^M d_{sj} \quad (11)$$

式中,  $d_{sj}$  表示第  $s$  个候选 RP 的离线指纹在第  $j$  次子集抽样过程中与目标指纹间的信号距离。

由于抽样的随机性, 子集可能抽到大量信号值为 0 的 AP 或者异常 AP。AP 有效信息量的不足会对定位结果产生较大影响, 通过重复多次随机抽样操作取平均值的方式, 不仅能充分利用 AP 信息, 同时, 最小化异常子集对信号空间距离计算的影响, 得出目标指纹与候选 RP 间更精准的距离值, 以确保后续选出目标真实位置附近的  $K$  近

邻，达到提高定位性能的目的。

b. 求取目标估计位置

选取距离最小的前  $K$  个指纹对应的 RP 位置作为最近邻，取距离倒数作为各 RP 位置  $P_i(x,y)$  的初始权重，而后归一化该权重，对各 RP 位置依照归一化后权重求和，得到目标估计位置  $P(x,y)$ ：

$$P(x,y)=\sum_{i=1}^K\frac{\frac{1}{D_i}}{\sum_{s=1}^K\frac{1}{D_s}}P_i(x,y)\tag{12}$$

该算法具体过程如表 1 所示。

表 1 随机子集抽样定位算法伪代码

Tab.1 Pseudocode of random subset sampling localization algorithm

|                                                      |
|------------------------------------------------------|
| 输入：目标指纹 $R_t$ 、候选RP指纹 $R_s$ 、随机子集抽样次数 $M$ 、候选RP数 $N$ |
| 输出：目标估计位置 $P(x,y)$                                   |
| 1 for $j=1:M$ do                                     |
| 2 随机生成长度为 $l$ 的 AP 索引向量                              |
| 3 根据 AP 索引向量选取目标的指纹子集 $R'_t$                         |
| 4 for $s=1:N$ do                                     |
| 5 根据 AP 索引向量提取候选 RP 的指纹子集 $R'_s$                     |
| 6 计算 $R'_t$ and $R'_s$ 的距离                           |
| 7 end                                                |
| 8 end                                                |
| 9 依据式(11)计算各 RP 与目标指纹间最终距离 $D_s$                     |
| 10 依据 $D_s$ 返回 $K$ 近邻对应坐标及其权重                        |
| 11 依据式(12)输出估计目标位置                                   |

2 实验结果及分析

通过在实际室内场景的公开数据集<sup>[19]</sup>中进行实验来验证本文算法的有效性。该数据集共采集 3 层楼的 RSS 信息，每层楼布局一致，面积为 1425 m<sup>2</sup>，共部署有 32 个 AP。数据集包含各楼层不同 RP 处所测得的 RSS 信息及其对应位置信息。其中，一楼采集到 397 条指纹信息，二楼采集到 711 条指纹信息，三楼采集到 431 条指纹信息。以二楼为例，实验布局如图 3 所示，“\*”代表参考位置点，“#”代表从指纹库随机抽取的在线测试点。

在每层数据库中随机抽取 100 个 RSS 指纹信息作为在线测试数据，其余作为离线指纹信息，后续实验此参数保持不变。测试点的定位误差定义为其真实坐标与估计位置坐标之间的欧氏距离，计算所有测试点的平均定位误差及定位误差的累积分布函数 (cumulative distribution function, CDF)

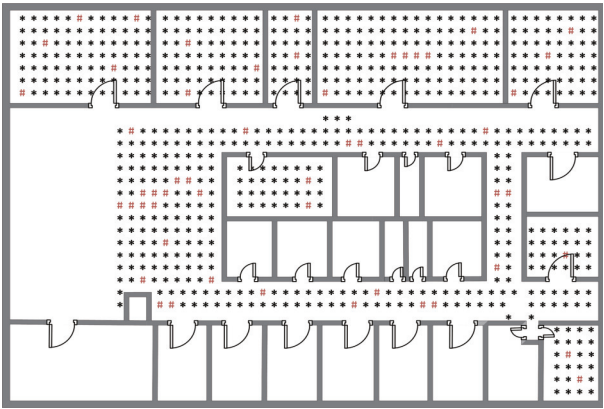


图 3 实验布局图

Fig.3 Experimental layout

作为衡量算法定位性能的指标。

2.1 在线粗定位的准确度分析

为验证本文所提出的在线粗定位方法的实用性，将其与文献 [16] 所提到的基于 K-means 的离线粗定位方法进行对比实验，两种实验仅存在前期粗定位方法的差异，后续实验均按照本文所提出的随机子集抽样方法进行二次精确定位。

图 4 给出了两种粗定位方法下各测试点的定位误差情况。实验参数如下：选取 50 个测试点，在粗定位阶段，本文所提出的在线粗定位方法根据覆盖向量选取 40 个候选 RP，离线粗定位方法将全指纹库数据划分为 5 簇；然后，在二次精确定位阶段，两种方法均对粗定位得到的 RP 集进行随机抽取 AP 子集，随机 10 次，AP 子集长度为 10；最终 WKNN 匹配时最近邻点  $K=5$ 。由结果可知：两种方法 80% 的测试点定位误差在 4 m 以下；本文所提出的在线粗定位方法平均定位误差为 2.62 m；而离线粗定位方法平均定位误差为 3.20 m，且其中部分测试点的定位误差过大，最大达到 20.03 m。造成这一差距的原因可能与离线粗定位时簇的划分有关。

为进一步分析定位误差，选取误差最大的测试点，图 5、图 6 分别给出了该测试点在离线、在线粗定位两种情况下的定位结果。

离线 K-means 算法将误差最大的测试点划分到簇 4 中，但由图 5 可知，该测试点实际距离簇 4 较远，距离类 1 更近。由于该测试点出现在不同簇类的边缘，在计算测试点与不同簇中心的距离时出现偏差，导致其所属簇的错误划分。后续是在粗定位簇内 (即簇 4) 的 RP 中进行二次精确定位，错误的簇匹配导致了最终定位结果出现较大误

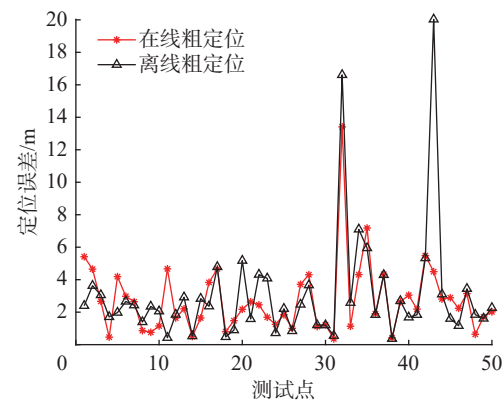


图 4 各测试点的定位误差

Fig.4 Positioning error of the test points

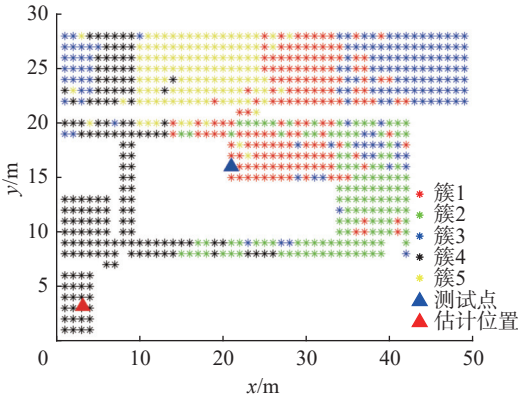


图 5 离线粗定位位置点分布情况

Fig.5 Distribution of offline coarse positioning locations

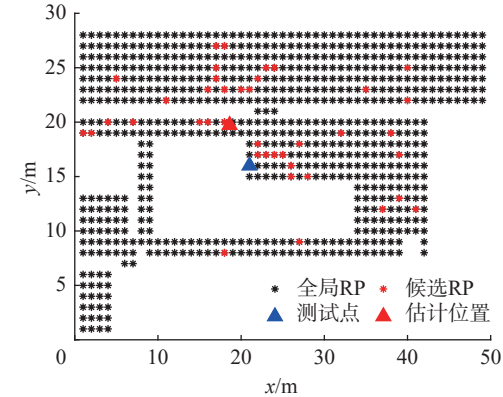


图 6 在线粗定位位置点分布情况

Fig.6 Distribution of online coarse positioning locations

差。采用本文所提出的基于覆盖向量的候选 RP 在线选择方式, 最终选出的  $N$  个 RP 如图 6 所示。尽管有部分候选 RP 分布在测试点较远的位置, 但其真实位置附近的大多数 RP 均被选为候选 RP, 有效避免了离线分簇可能导致的较大定位误差。

2.2 随机子集抽样的准确度分析

为验证本文所提出的随机子集抽样方法的有效性, 将其与固定子集抽样方法进行对比。在确

定测试点候选 RP 集后, 分别使用本文所提出的随机子集抽样与固定子集抽样进行测试点的二次精确定位。在 4 次实验中, 固定子集抽样方法分别选取指纹库中前 15 个 AP、中间 15 个 AP、后 15 个 AP, 随机子集抽样方法则随机选取 15 个 AP, 实验结果如图 7 所示。可以看出, 随机子集抽样方法表现出更加稳定的定位性能。在固定子集抽样实验中, 后 3 次实验相较于第 1 次实验的定位误差均明显增大, 达到 3 m 以上, 这可能由于其所抽取的 AP 子集中包含异常 AP, 影响了最终的定位性能。随机子集抽样方法降低了定位中采用异常 AP 的概率, 有利于定位性能的提高。

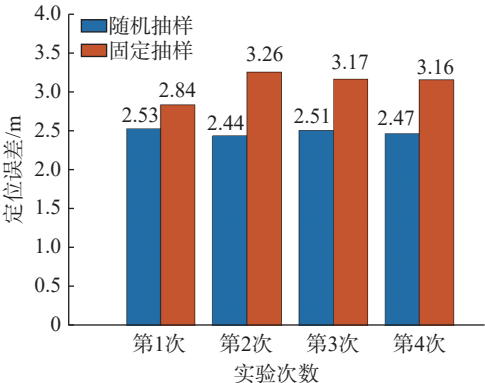


图 7 不同子集抽样方法定位误差比较

Fig.7 Comparison of positioning errors by different subset sampling methods

2.3 算法高效性分析

本文算法通过在线粗定位确定的候选 RP 分布如图 8 所示。相较于传统定位方案的全局指纹匹配, 在线粗定位可将目标快速、准确地定位到局部区域内, 可有效降低后续二次精确定位阶段指纹匹配的计算开销。

本文算法提出了在线粗定位阶段采用 Jaccard 相似度作为指纹匹配的依据, 为验证其高效性,

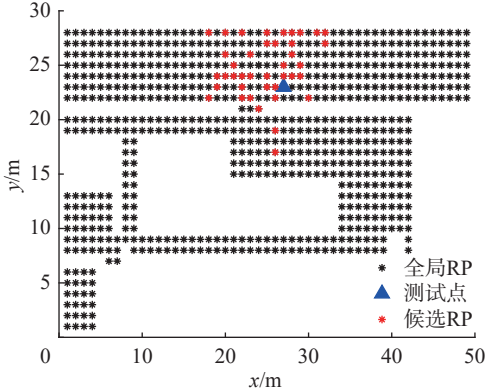


图 8 在线粗定位候选 RP 分布

Fig.8 Candidate RP distribution in online coarse localization

将其与传统的基于汉明距离、欧氏距离的指纹匹配方法进行实验对比。随机选取一条指纹库指纹作为在线指纹，与剩下的 710 条离线指纹进行匹配，比较不同情况下的时间开销，实验结果如表 2 所示。本文所提出的基于 Jaccard 相似度的指纹匹配方法具有较低时间开销，实用性优于其他两种方法。

表 2 不同指纹匹配方法下的时间开销

Tab.2 Time overhead of different fingerprint matching methods

| 指纹匹配方法     | 时间开销/ms |
|------------|---------|
| Jaccard相似度 | 16      |
| 汉明距离       | 27      |
| 欧氏距离       | 32      |

2.4 参数选择对性能的影响

本文算法在离线阶段不需要进行定位模型的训练，而是直接根据在线接收的 RSS 向量进行目标定位。在粗定位和随机子集抽样过程中，涉及到多个参数，如：粗定位时的候选 RP 数  $N$ 、随机抽样子集数  $M$ 、子集中的 AP 数  $l$ 。为研究各参数对定位性能的影响，通过上述公共数据集进行相应实验，实验参数与上文一致。

a. 候选 RP 数  $N$

在粗定位阶段，候选 RP 数对于确定目标可能所在局部空间的大小起着重要作用。实验初始抽取 10 个子集，每个子集中包含随机抽取的 15 个 AP，改变粗定位阶段的候选 RP 数，观察定位误差变化，实验结果如图 9 所示。在一定范围内，随着候选 RP 数的增加，平均定位误差逐渐减小，当候选 RP 数达到一定值后，其改变对定位精度的影响不大。在实际应用中，过多的候选 RP 会增加算法的计算复杂度，还可能包含一些距离目标较远的 RP，从而降低定位精度。因此，需要在合理范围内选择适当数量的候选 RP 来平衡定位精度和计算复杂度。

b. 抽样子集数  $M$

实验设置粗定位阶段选取 40 个候选 RP，在后续子集抽样中，每个子集包含 15 个随机抽取的 AP。改变子集的个数  $M$  进行实验，所得结果如图 10 所示。在一定范围内，随着子集数量的增多，定位误差逐渐减小，当子集个数达到一定数量后，其对定位精度的影响逐渐减小。由于子集抽样过程中的 AP 选择是随机的，可能导致抽取的子集中

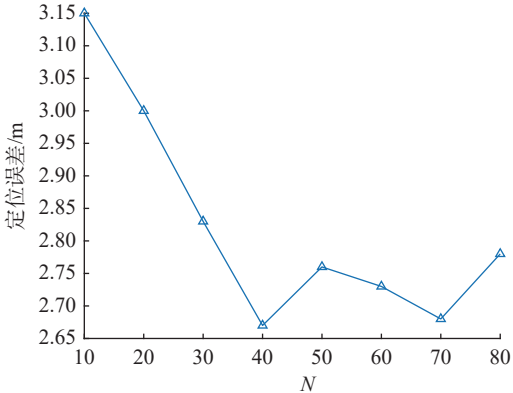


图 9 候选 RP 数对定位误差的影响

Fig.9 Effect of the number of candidate RP on positioning error

包含大量信号值为 0 的 AP，RSS 信息量不足会导致后续算法的定位结果产生偏差，选择较多数量的子集可以有效降低以上特殊情况对定位精度的影响。在实际应用中，为降低算法的复杂度，需要在合理范围内选择适当数量的子集。

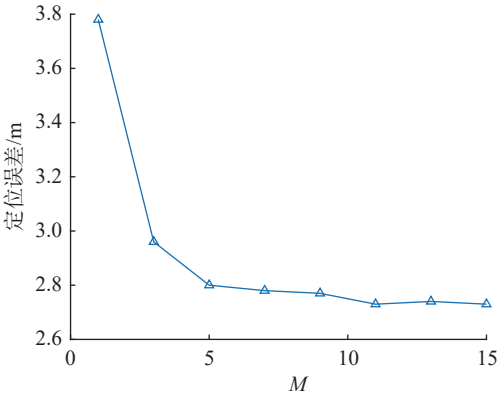


图 10 抽样子集数对定位误差的影响

Fig.10 Effect of the number of subsets on positioning error

c. 子集中的 AP 数  $l$

实验设置粗定位阶段选取 40 个候选 RP，抽取 10 个子集，改变子集中的 AP 数  $l$ ，实验结果如图 11 所示。随着所选 AP 数量的增加，定位精度不断提高。充分利用 AP 信息，有利于更加准确地实现目标的定位，实际应用中可根据整个定位区域 AP 数量进行选取。

2.5 与已有方法的比较

为进一步验证本文算法的性能，与已有的 WKNN、WKNN-Kmeans、PCA-KRR 算法<sup>[20]</sup>进行对比实验。其中：WKNN-Kmeans 在离线阶段利用 Kmeans 算法对指纹聚类，在线阶段通过簇匹配实现目标粗定位，再在簇内利用 WKNN 实现定位；PCA-KRR 算法是对指纹库中数量众多的 AP 信号

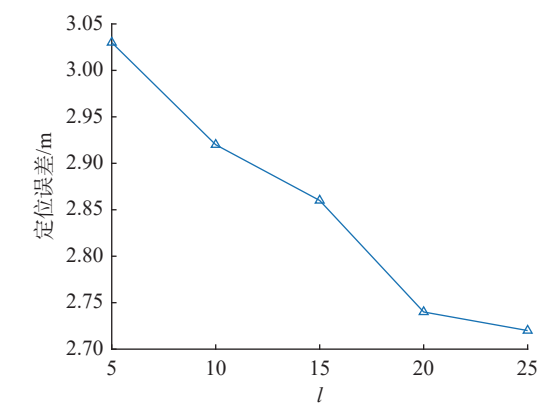


图 11 子集中 AP 数对定位误差的影响

Fig.11 Effect of the number of AP in the subset on the positioning error

提取主成分，实现数据降维，再利用核岭回归算法建立位置与 RSS 信号的模型。在本实验中，已将各对比算法参数调至最优，所得定位误差的累积分布函数如图 12 所示。

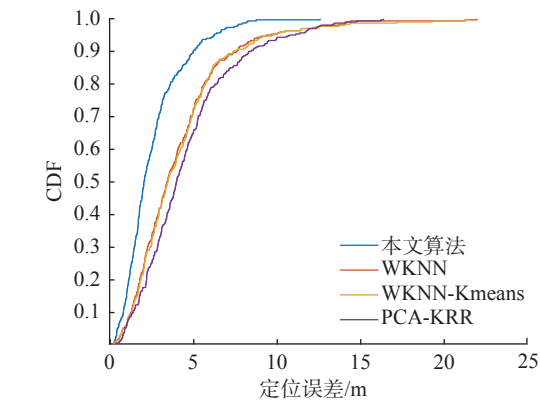


图 12 公开数据集定位误差累积分布函数

Fig.12 Cumulative distribution function of positioning error in public datasets

权衡本文算法的计算复杂度和定位精度，将参数设置为：粗定位阶段选取 40 个候选 RP，抽取 10 个子集，每个子集包含 15 个随机抽取的 AP。由图 12 可知，本文算法中：90% 的定位误差小于 5 m；WKNN 算法和 WKNN-Kmeans 算法中，70% 的定位误差小于 5 m；PCA-KRR 算法中，60% 的定位误差小于 5 m。本文算法、WKNN、WKNN-Kmeans、PCA-KRR 算法的平均定位误差分别为 2.47、4.06、4.20、4.54 m。本文算法有效降低了平均定位误差，对比 WKNN、WKNN-Kmeans、PCA-KRR 算法，本文算法的定位精度分别提升了 39%、41%、46%。

2.6 AP 异常对定位精度的影响

在实际应用中，可能存在某些 AP 异常的情

况，如目标在线采集信号时，少数 AP 被关闭或被交换位置，导致信号接收出现偏差，从而对定位结果产生影响。实验对比了两种 AP 异常情况下本文算法与 WKNN、WKNN-Kmeans、PCA-KRR 算法定位误差的差异，结果如表 3 和表 4 所示。

表 3 AP 交换情况下不同算法的定位误差

Tab.3 The positioning error of different algorithms under AP handover conditions

| 交换AP个数 | 定位误差/m |      |             |         |
|--------|--------|------|-------------|---------|
|        | 本文算法   | WKNN | WKNN-Kmeans | PCA-KRR |
| 0      | 2.47   | 4.06 | 4.20        | 4.54    |
| 2      | 3.18   | 4.59 | 4.59        | 5.04    |
| 3      | 3.16   | 4.67 | 4.78        | 5.47    |
| 4      | 3.12   | 5.04 | 5.30        | 5.49    |
| 5      | 4.05   | 5.12 | 5.03        | 5.76    |

表 4 AP 关闭情况下不同算法的定位误差

Tab.4 The positioning error of different algorithms under AP shutdown conditions

| 关闭AP个数 | 定位误差/m |      |             |         |
|--------|--------|------|-------------|---------|
|        | 本文算法   | WKNN | WKNN-Kmeans | PCA-KRR |
| 0      | 2.47   | 4.06 | 4.20        | 4.54    |
| 2      | 2.94   | 4.46 | 4.53        | 4.91    |
| 3      | 3.04   | 4.56 | 4.71        | 4.99    |
| 4      | 3.69   | 4.82 | 4.88        | 5.11    |
| 5      | 3.74   | 5.13 | 5.19        | 5.06    |

随着交换 AP、关闭 AP 个数的逐渐增多，4 种方法的定位性能均逐渐下降，但本文算法有效抑制了指纹偏差引入的大定位误差。对比于 WKNN、WKNN-Kmeans、PCA-KRR 算法，本文算法的定位精度提升了至少 20%。在实际应用中，本文算法具有更好的鲁棒性，面对 AP 异常时，能够实现更好的定位性能。

3 结 论

针对 RSS 信号受传播环境突发噪声影响以及 AP 信号存在异常的实际情况，提出了一种基于随机子集抽样的高效室内指纹定位算法。该算法在线阶段利用 AP 覆盖向量完成目标粗定位，可快速将目标定位到局部区域，然后采用随机子集抽样实现了目标位置的二次精确估计。此外，从不同粗定位方法、不同的参数选择、AP 异常等方面对

算法进行了分析,并与其他室内定位算法进行了对比。实验结果表明,本文所提出的定位算法能有效降低指纹匹配的计算开销,具有良好的定位性能。

#### 参考文献:

- [1] HE S N, CHAN S H G. Wi-Fi fingerprint-based indoor positioning: recent advances and comparisons[J]. *Communications Surveys & Tutorials*, 2016, 18(1): 466–490.
- [2] RUAN L, ZHANG L, ZHOU T, et al. An improved bluetooth indoor positioning method using dynamic fingerprint window[J]. *Sensors (Basel)*, 2020, 20(24): 7269.
- [3] 许远航, 张鑫宇, 乐燕芬. 室内多用户在线协同指纹定位算法[J]. *小型微型计算机系统*, 2024, 45(3): 699–706.
- [4] 周军, 魏国亮, 田昕, 等. 融合 UWB 和 IMU 数据的新型室内定位算法[J]. *小型微型计算机系统*, 2021, 42(8): 1741–1746.
- [5] 高锐, 程良伦, 胡莘. 一种基于空间分割的无源 RFID 室内定位方法[J]. *计算机应用研究*, 2012, 29(1): 184–186, 210.
- [6] BAHL P, PADMANABHAN V N. RADAR: an in-building RF-based user location and tracking system[C]//*Proceedings of Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies*. Tel Aviv: IEEE, 2000: 775–784.
- [7] LU X M, QIU Y, YUAN W L, et al. An improved dynamic prediction fingerprint localization algorithm based on KNN[C]//*Proceedings of 2016 Sixth International Conference on Instrumentation & Measurement, Computer, Communication and Control (IMCCC)*. Harbin: IEEE, 2016: 289–292.
- [8] HU J S, LIU D W, YAN Z, et al. Experimental analysis on weight K-nearest neighbor indoor fingerprint positioning[J]. *IEEE Internet of Things Journal*, 2019, 6(1): 891–897.
- [9] HSIEH C H, CHEN J Y, NIEN B H. Deep learning-based indoor localization using received signal strength and channel state information[J]. *IEEE Access*, 2019, 7: 33256–33267.
- [10] LE Y F, ZHANG H N, SHI W B, et al. Received signal strength based indoor positioning algorithm using advanced clustering and kernel ridge regression[J]. *Frontiers of Information Technology & Electronic Engineering*, 2021, 22(6): 827–838.
- [11] YOUSSEF M, AGRAWALA A. The Horus WLAN location determination system[C]//*Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services*. Seattle: ACM, 2005: 205–218.
- [12] CHEN C, WANG Y J, ZHANG Y, et al. Indoor positioning algorithm based on nonlinear PLS integrated with RVM[J]. *IEEE Sensors Journal*, 2018, 18(2): 660–668.
- [13] PETRIC M, NESKOVIC A, NESKOVIC N, et al. Indoor localization using multi-operator public land mobile networks and support vector machine learning algorithms[J]. *Wireless Personal Communications*, 2019, 104(4): 1573–1597.
- [14] HE S N, LIN W B, CHAN S H G. Indoor localization and automatic fingerprint update with altered AP signals[J]. *IEEE Transactions on Mobile Computing*, 2017, 16(7): 1897–1910.
- [15] 乐燕芬, 许远航, 施伟斌. 基于 DPC 指纹子空间匹配的室内 WiFi 定位方法[J]. *仪器仪表学报*, 2021, 42(11): 106–114.
- [16] 田洪亮, 钱志鸿, 梁潇, 等. 离散度 WKNN 位置指纹 Wi-Fi 定位算法[J]. *哈尔滨工业大学学报*, 2017, 49(5): 94–99.
- [17] FENG C, AU W S A, VALAEE S, et al. Received signal strength based indoor positioning using compressive sensing[J]. *IEEE Transactions on Mobile Computing*, 2012, 11(12): 1983–1993.
- [18] KHALAJMEHRABADI A, GATSIS N, AKOPIAN D. Structured group sparsity: a novel indoor WLAN localization, outlier detection, and radio map interpolation scheme[J]. *IEEE Transactions on Vehicular Technology*, 2017, 66(7): 6498–6510.
- [19] TÓTH Z, TAMÁS J. Miskolc IIS hybrid IPs: dataset for hybrid indoor positioning[C]//*Proceedings of the 26th International Conference Radioelektronika*. Kosice: IEEE, 2016: 408–412.
- [20] 张贺娜, 乐燕芬, 施伟斌. 基于特征降维的核岭回归室内定位算法[J]. *仪器仪表学报*, 2020, 41(10): 83–91.

(编辑: 董 伟)