

文章编号: 2095-2163(2021)07-0102-05

中图分类号: TP393

文献标志码: A

# 基于 Wi-Fi 指纹和随机森林的定位算法

韩学法, 吴 飞, 时瑶佳, 胡 锐, 聂大惟

(上海工程技术大学 电子电气工程学院, 上海 201620)

**摘要:** 针对 Wi-Fi 信号强度的相似性对室内定位的影响, 本文提出一种基于 Wi-Fi 指纹和随机森林的室内定位算法。该算法采用 Wi-Fi 作为信号源, 以接收信号强度指示和基本服务集标识符来构建 Wi-Fi 指纹库, 从而建立随机森林模型用于室内位置感知。仿真实验表明, 该算法的定位误差约为 2.26 m, 与同类算法相比, 在执行时间和定位精度上具有较好的优越性, 算法精度提高约 3.2%。

**关键词:** 室内定位; Wi-Fi 指纹; 指纹数据库; 随机森林; 位置感知

## Location algorithm based on Wi-Fi fingerprint and Random Forest

HAN Xuefa, WU Fei, SHI Yaojia, HU Rui, NIE Dawei

(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

**【Abstract】** Aiming at the influence of the similarity of Wi-Fi signal strength on indoor positioning, this paper proposes an indoor positioning algorithm based on Wi-Fi fingerprint and Random Forest. The algorithm uses Wi-Fi as a signal source to construct the Wi-Fi fingerprint library by receiving signal strength indication and basic service set identifiers, thereby establishing a Random Forest model for indoor location sensing. The simulation results show that the positioning error of the algorithm is about 2.26 m. Compared with similar algorithms, it has better performance in execution time and positioning accuracy, and the accuracy of the algorithm is improved by about 3.2%.

**【Key words】** indoor positioning; Wi-Fi fingerprint; fingerprint database; Random Forest; location perception

## 0 引言

随着智能手机的普及, 基于位置服务 (Location Based Services, LBS) 的应用在现实生活中备受关注, 高精度的室内定位技术已经成为现阶段的研究热点<sup>[1]</sup>。室内定位技术主要包括蓝牙技术<sup>[2]</sup>、Wi-Fi 技术<sup>[3]</sup>、UWB<sup>[4]</sup>技术及 RFID<sup>[5]</sup>技术等。

基于 Wi-Fi 的定位技术因具有硬件部署成本低、计算开销小、定位精度相对较高的优点<sup>[6]</sup>, 目前被广泛用于室内定位。国内外学者对室内 Wi-Fi 指纹定位技术进行了大量的研究。文献[7]提出一种基于密度聚类的随机森林的室内定位算法, 采用网格搜索和交叉验证技术, 降低了定位误差; 文献<sup>[8]</sup>提出一种基优化随机森林模型的改进算法, 提高定位算法的预测能力; 文献[9]提出一种结合 Wi-Fi 指纹和无线接入点 (Wireless access point, AP) 的选择的室内定位方法, 有效地降低多径效应的对

定位结果的影响。

然而, 上述方法只是单一考虑接收信号强度指示值 (Received Signal Strength Indication, RSSI), 但定位结果会受到 Wi-Fi 信号强度相似性的影响<sup>[10]</sup>。因此, 针对这一问题, 本文提出一种基于 Wi-Fi 指纹和随机森林的室内定位算法。与同类算法相比, 该算法在执行时间和定位精度具有较好的优越性, 且定位准确率提高了约 3.2%。

## 1 Wi-Fi 指纹定位技术

Wi-Fi 指纹定位技术是利用空间位置 Wi-Fi 信号的唯一性进行位置指纹匹配的定位方法, 从而得到定位点的位置信息。指纹匹配定位是一种常用的室内定位方法, 包括离线阶段和在线阶段<sup>[11]</sup>。

离线阶段的任务是采集各个无线接入点 (Wireless access point, AP) 的 RSSI 值, 将参考点的 RSSI 值与位置信息结合起来建立指纹库; 在线阶段

**基金项目:** 上海市高校研究生创新项目 (19KY0216); 上海市科技学术委员会重点项目 (18511101600); 国家自然科学基金青年基金项目 (61902237); 上海市科委青年科技英才“扬帆计划”项目 (19YF1418200)。

**作者简介:** 韩学法 (1994-), 男, 硕士研究生, 主要研究方向: 室内定位与导航; 吴 飞 (1968-), 男, 博士, 教授, 主要研究方向: 计算机网络与计算机能耗优化。

**通讯作者:** 韩学法 Email: hxf\_edu@edu.com

收稿日期: 2020-12-14

的任务是在定位点采集各个 AP 的 RSSI 值,通过匹配算法从指纹库中匹配出最相似的指纹并得到定位点的位置信息<sup>[12]</sup>。Wi-Fi 指纹定位技术流程如图 1 所示。

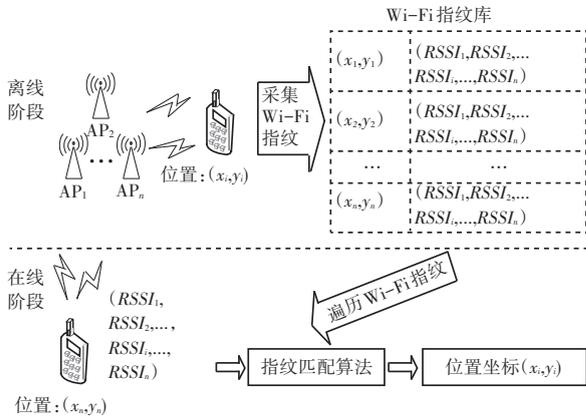


图 1 Wi-Fi 指纹定位技术的流程图

Fig. 1 Flow chart of Wi-Fi fingerprint positioning technology

指纹匹配算法是室内 Wi-Fi 指纹定位过程中最为重要的环节,常用的算法包括 NN 算法、KNN 算法、WKNN 算法和统计概率算法<sup>[12]</sup>。

## 2 基于 Wi-Fi 指纹和随机森林的定位算法

考虑到 Wi-Fi 信号强度的相似性,本文以 RSSI 值和基本服务集标识符 (BSSID) 来构建指纹数据库,提出一种基于随机森林和 Wi-Fi 指纹的定位算法。

### 2.1 随机森林

随机森林(Random Forest, RF)是在 2001 年提出的一种基于决策树分类器的机器学习算法<sup>[13]</sup>,采用 Bagging 方法<sup>[14]</sup>在训练样本集  $D$  中抽取  $N$  个自助样本集,为每一个样本集训练一个决策树。RF 的生成步骤下如图 1 所示<sup>[15]</sup>,算法具体过程描述如下:

**Step 1** 在样本集  $D$  选取  $N$  个子训练样本集  $D_1, D_2, \dots, D_N$ ,并预建  $N$  棵决策树。

**Step 2** 在决策树的子节点上随机地从  $n$  个指标中选取  $m$  个,选取最优分割指标进行分割。

**Step 3** 重复步骤 Step 2 遍历预建的  $N$  棵分类树。

**Step 4** 由  $N$  棵分类树形成 RF。

RF 生成步骤如图 2 所示。由图 2 可知,随机森林生成过程中关键是确定  $N$  值, $N$  值直接影响投票决定的最优结果。

### 2.2 定位算法

传统的室内 Wi-Fi 指纹定位算法,采用随机森林模型时只考虑采样室内空间网格中的 RSSI 值,这样难以依据环境因素反映每个网格位置的信号强度

并加以区分。因此,本文以 RSSI 值和 BSSID 共同作用来区分网格位置的信号强度,构建包括位置坐标、RSSI 和 BSSID 的 Wi-Fi 指纹库。

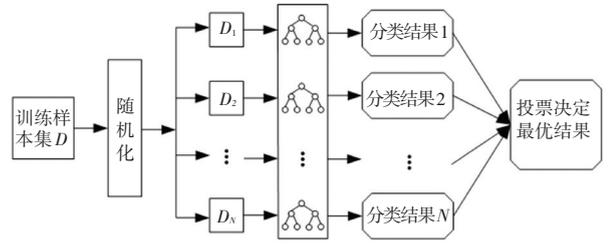


图 2 RF 生成步骤

Fig. 2 RF generation steps

由于在每个位置采集到所有 AP 的 BSSID (MAC 地址)是相同的,但其强度大小不同,所以可通过 RSSI 值和 BSSID 共同作用过滤掉强度差的 AP。从而减少后期在线阶段的匹配次数,提高执行效率。本文所提出的定位算法与指纹匹配定位算法原理相似,其流程如图 3 所示。

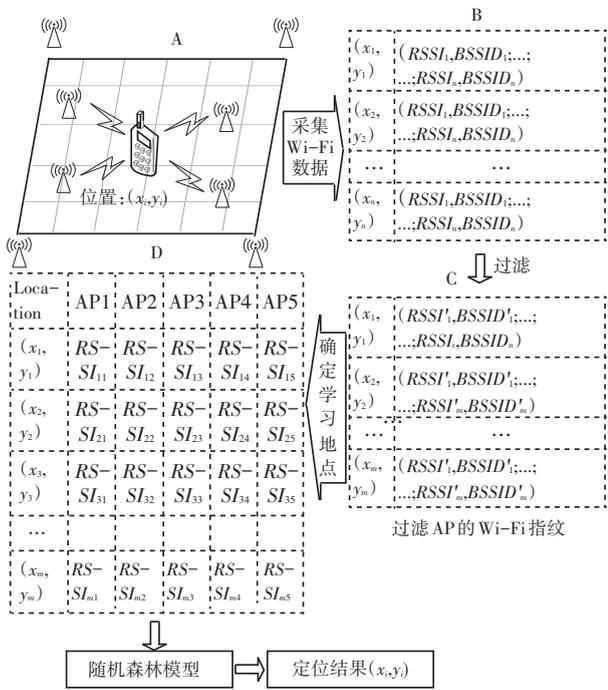


图 3 本文算法流程图

Fig. 3 Algorithm flow chart in this paper

在上述流程图中,A 为采集 Wi-Fi 数据的室内区域缩略图,B 为过滤前采集的 Wi-Fi 指纹数据,C 为过滤后采集的 Wi-Fi 指纹数据,D 为用于随机森林进行学习训练的 Wi-Fi 指纹库。本文所提出的基于 Wi-Fi 指纹和 Wi-Fi 指纹的定位算法的步骤描述如下:

**Step 1** 执行算法,采集 A 区 Wi-Fi 实测数据,

将其记为  $B$ 。 $B$  中包括位置坐标信息、RSSI 和 BSSID。

**Step 2** 位置筛选,将每个位置的 BSSID 序列和  $B$  中进行对比。若一致,则以 RSSI 值降序方式保存 RSSI 和 BSSID 到  $C$  中,否则删除不一致的数据。

**Step 3** 选取强度最大 5 个 AP,构建 Wi-Fi 指纹,将其作为随机森林进行学习的样本集  $D$ 。

**Step 4** 采用 2.1 节所提出的随机森林算法构建随机森林,进行学习训练。

**Step 5** 定位阶段将保存的 Wi-Fi 指纹库  $D$  作为样本集输入随机森林模型进行预测。

**Step 6** 得出预测的定位结果  $(x_i, y_i)$ 。

### 3 仿真实验与分析

为了验证本文所提的基于随机森林和 Wi-Fi 指纹的定位算法的可行性以及与同类算法相比的优势,进行仿真实验与分析。

#### 3.1 实验设计

实验硬件设备为 TinkPad T480 笔记本电脑一台,以 MATLAB R2018a 为仿真实验平台构建一个长 20 m、宽 15 m、高 4 m 的开放的室内空间,将其划分为大小为 1 m 的方格进行仿真实验。通过一次仿真实验得到指纹样本集  $C$  的部分信息见表 1。

在表 1 中的样本集  $C$  中选取信号强度最大 5 个 AP 构建 Wi-Fi 指纹库  $D$ ,并将其作为随机森林模型的样本集,样本集  $D$  的具体信息见表 2。

表 2 中存储参考节点(待定位点)处信号强度最强的 5 个 AP 的 RSSI 值,作为随机森林进行学习的样本集。

表 1 样本集  $C$  的部分信息

Tab. 1 Part of the information of sample set  $C$

Location	RSSI_BSSID
$(x_1, y_1)$	$(-49, 70; 3a; 0e; 06; 95; b2; -52, 70; 3a; 0e; 06; 95; b1; \dots)$
$(x_2, y_2)$	$(-43, c4; ca; d9; 25; 8a; 62; -48, c4; ca; d9; 25; 8a; 60; \dots)$
$(x_3, y_3)$	$(-46, 70; 3a; 0e; 06; 95; b1; -50, 70; 3a; 0e; 06; 95; b0; \dots)$
...	...
$(x_m, y_m)$	$(-38, 70; 3a; 0e; 06; 95; b0; -45, 70; 3a; 0e; 06; 95; b1; \dots)$

表 2 样本集  $D$  的部分信息

Tab. 2 Part of the information of sample set  $D$

Location	$AP_1$	$AP_2$	$AP_3$	$AP_4$	$AP_5$
$(x_1, y_1)$	-46	-49	-52	-53	-56
$(x_2, y_2)$	-43	-47	-48	-52	-55
$(x_3, y_3)$	-45	-48	-50	-54	-57
...	...	...	...	...	...
$(x_m, y_m)$	-39	-42	-44	-48	-53

#### 3.2 实验分析

由在 2.1 节中随机森林生成算法可知,确定  $N$  值尤为重要。每个定位点采用欧式距离计算误差,其公式为:

$$err_i = \sqrt{(x_{0i} - x_i)^2 + (y_{0i} - y_i)^2}, i = 1, 2, \dots, k \quad (1)$$

其中,  $(x_{0i}, y_{0i})$  为参考点的实测坐标,  $(x_i, y_i)$  为各定位算法解算的参考点坐标。然而,本文算法的误差由每颗决策树的均值误差共同作用,其计算公式为:

$$err_N = \frac{\sum_{i=1}^k err_i}{N}, i = 1, 2, \dots, k; N = 5, 10, \dots, 50 \quad (2)$$

确定合适的  $N$  值。为了确定合适的  $N$  值进行 3 次实验,具体信息见表 3。

表 3 确定  $N$  值的实验记录

Tab. 3 Experimental records to determine the value of  $N$

算法	决策树棵数	次数 1			次数 2			次数 3		
		定位误差/m	训练时间/s	预测时间/ms	定位误差/m	训练时间/s	预测时间/ms	定位误差/m	训练时间/s	预测时间/ms
基于随机森林和 Wi-Fi 指纹的定位算法	5	2.467	1.89	5.45	2.459	1.94	5.99	2.477	1.97	5.23
	10	2.373	3.91	9.50	2.364	3.98	9.61	2.337	3.98	6.72
	15	2.322	5.89	10.10	2.340	5.94	13.30	2.329	5.95	12.50
	20	2.293	7.96	15.60	2.276	7.77	17.50	2.283	8.00	14.70
	25	2.284	9.96	18.50	2.269	9.93	24.10	2.269	9.91	19.90
30	2.268	11.90	25.50	2.256	11.70	23.40	2.270	11.90	23.10	
35	2.265	13.90	28.20	2.268	13.70	26.30	2.266	13.80	20.10	
40	2.263	15.90	21.60	2.243	15.60	32.30	2.273	15.80	27.40	
45	2.257	17.90	41.50	2.256	17.90	33.00	2.244	17.90	36.70	
50	2.246	19.50	44.30	2.259	19.70	38.60	2.241	19.90	40.10	

由表 3 可知,随着  $N$  值的不断增大,模型的训练时间和预测时间也会增加。对于定位效果来说,实时性是至关重要,减少误差的同时不能忽略执行时间。为了更加直观,对 3 次实验的误差取均值作图,如图 4 所示。

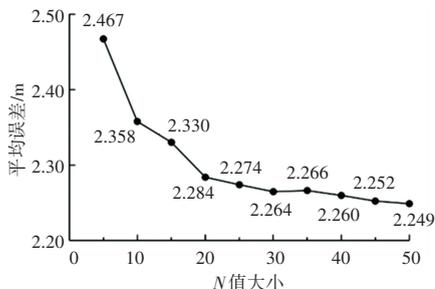


图 4 确定  $N$  值的效果图

Fig. 4 The effect of determining the  $N$  value

由图 4 可知,当  $N$  为 30 ~ 40 间的误差趋于稳定, $N$  大于 40 时,误差有所减低但执行时间也增加了。对于定位来说,要考虑时效性,故  $N$  值取 30 是比较合适的。

当  $N = 30$  时,将本文的算法和基于 RSSI 和随机森林的同类算法相比,对比效果如图 5 和图 6 所示。

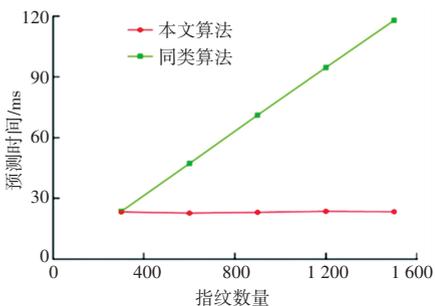


图 5 预测时间对比图

Fig. 5 Comparison of forecast time

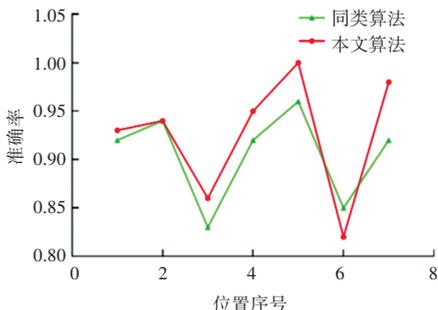


图 6 算法准确率对比图

Fig. 6 Algorithm accuracy comparison chart

由图 5 可知,本文算法的预测时间不会随着指纹数量的增多而增加,而同类算法则会随着指纹数量的增多而增加。

线性增长趋势。对于定位来说,实时性相当重要,所以本文所提的算法在实时性是满足要求的。

由图 6 可知,本文算法在不同位置的定位准确率高于同类算法约 3.2% 左右。因此,本文算法在不影响执行时间的情况下,提高了室内定位的准确率。

## 4 结束语

Wi-Fi 信号强度的相似性会对定位效果产生影响,同一点处采集 Wi-Fi 信号来都自于相同的 AP,单一考虑 Wi-Fi 信号的 RSSI 值是无法确定 Wi-Fi 信号强度对定位结果的影响。因此,本文提出一种基于 Wi-Fi 指纹和随机森林的定位算法,该算法以 RSSI 和 BSSID 共同作用来选取强度最大的 5 个 AP 用于构建指纹库。将 Wi-Fi 指纹库输入随机森林进行学习,从而构建随机森林模型。仿真实验表明,本文算法的定位误差约为 2.26 m,与同类算法相比,算法的执行时间不会随样本集大小而变化,且定位准确率提高约 3.2%。然而,由于 Wi-Fi 信号中含有一种受墙壁或障碍物影响而产生的噪声,因此未来研究需要解决噪声问题。

## 参考文献

- [1] 李华蓉,赵一. 室内定位导航地图一体化设计[J]. 测绘通报, 2015(9): 32-35.
- [2] KWAK J, SUNG Y. Beacon-based indoor location measurement method to enhanced common chord-based trilateration[J]. Journal of Information Processing Systems, 2017,13(6): 1640-1651.
- [3] MA Zixiang, WU Bang, POSLAD S. A WiFi RSSI ranking fingerprint positioning system and its application to indoor activities of daily living recognition[J]. International Journal of Distributed Sensor Networks, 2019,15(4): 155014771983791.
- [4] De ANGELIS G, MOSCHITTA A, CARBONE P. Positioning techniques in indoor environments based on stochastic modeling of UWB round-trip-time measurements[J]. IEEE Transactions on Intelligent Transportation Systems, 2016,17(8): 2272-2281.
- [5] MONTASER A, MOSELHI O. RFID indoor location identification for construction projects[J]. Automation in Construction, 2014, 39:167-179.
- [6] HU Xujian, WANG Hao. WIFI indoor location optimization method based on position fingerprint algorithm [C]//2017 International Conference on Smart Grid and Electrical Automation (ICSGEA).Changsha, China: IEEE,2017:585-588.
- [7] 张萌,吕艳,倪益华,等. 基于密度峰值聚类的随机森林室内定位[J]. 计算机工程与设计, 2018,39(05): 1490-1496.
- [8] 郭妍,陈晓,任晓晔. 一种优化随机森林模型的室内定位方法[J]. 激光杂志, 2018,39(10): 70-74.
- [9] LI Qiyue, LI Wei, SUN Wei, et al. Fingerprint and assistant nodes based Wi-Fi localization in complex indoor environment [J]. IEEE Access, 2016,4:2993-3004.