

基于 K-Means 和 FCM 的增强型 Wi-Fi 指纹定位策略^①

陈 英^{1,2}, 单文杰¹, 杨丰玉¹

¹(南昌航空大学 软件学院, 南昌 330063)

²(南昌航空大学 物联网技术研究所, 南昌 330063)

摘要: 研究了通过数据处理算法以提高 Wi-Fi 指纹库室内定位性能的问题. 首先采集 Wi-Fi 指纹样本, 将其放入 MySQL 数据库中和 R 工程; 其次将 Wi-Fi 指纹库分成若干个簇, 使用 K-均值聚类(K-Means)和模糊 C-均值聚类(FCM)对待定位的 Wi-Fi 指纹进行聚类分析; 最后, 提出增强型的聚类策略(ECS)应用于 Wi-Fi 指纹匹配定位中. 实验结果表明, ECS 较仅使用 FCM 算法, 其定位耗时缩短约 50%-80%, 且定位精度上有所改善; ECS 较仅使用 K-Means 算法, 其定位精度提高约 20%-40%, 且定位稳定性较强并自动更新 Wi-Fi 指纹库.

关键词: Wi-Fi 指纹; K-均值聚类; 模糊 C-均值聚类; 增强型定位策略

Enhanced Positioning Strategy Based on K-Means and FCM for Wi-Fi Fingerprint

CHEN Ying^{1,2}, SHAN Wen-Jie¹, YANG Feng-Yu¹

¹(School of Software, Nanchang Hangkong University, Nanchang 330063, China)

²(Internet of Things Technology Institute, Nanchang Hangkong University, Nanchang 330063, China)

Abstract: The data processing algorithm is studied to improve Wi-Fi fingerprint indoor positioning performance. Firstly, Wi-Fi fingerprint samples are collected and then are put into MySQL database and R project. Secondly, the Wi-Fi fingerprint data is divided into several clusters, and the K-mean clustering (K-Means) and fuzzy C-means clustering (FCM) are used to cluster the Wi-Fi fingerprint respectively. Finally, an enhanced clustering strategy (ECS) is proposed to for Wi-Fi fingerprint matching. Experimental results show that ECS reduces the positioning time-consuming about 50%-80% than that consumed by only using FCM and the positioning accuracy is also improved; ECS improves about 20%-40% than that obtained by only using K-Means in terms of positioning accuracy and it proves positioning stability and can automatically update the Wi-Fi fingerprint database.

Key words: Wi-Fi fingerprint; K-Means; FCM; enhanced positioning strategy

随着物联网技术的发展和移动互联网的普及, 无处不在的位置服务也越来越被人们需要. 据诺基亚公司的数据表明, 人们 87%-90%的时间在室内度过. 但目前室内定位还处于初级阶段, 无论在定位精度、速度、鲁棒性等方面还都不能够满足日益增长的智能位置服务的需求. 基于 Wi-Fi 指纹匹配的无线室内定位技术^[1]一定程度上解决了定位精度问题, 且基础设备应用广泛, 成本较低, 深受研究人员的关注. 但由于 Wi-Fi 无线设备大多应用于人类活动最为频繁的室内环境, 无线室内定位易受室内布局变动、无线信号的

多径效应等因素影响, 其无线信号具有较强的波动性, 严重影响定位性能. 而基于 Wi-Fi 指纹匹配的定位方法首先通过不同位置上多次采集 Wi-Fi RSSI 信号建立指纹模型, 然后利用匹配算法估计待定位点的具体位置, 此方法避免了有信号特征转化为距离带来的误差, 故基于 Wi-Fi 指纹匹配的定位算法成为近年来室内定位的研究热点.

1 国内外研究现状

基于指纹匹配的无线室内定位技术^[2]是利用物理

① 基金项目:江西省自然科学基金(20161BAB212034);南昌航空大学博士启动基金(EA201520009)

收稿时间:2016-09-01;收到修改稿时间:2016-10-12 [doi:10.15888/j.cnki.csa.005771]

空间内不同位置具有不同 RSSI 的特征, 作为唯一识别此位置的方法, 一般分为离线采样和在线定位这两个阶段. 离线采样时通过采集在多个参考点(reference points, RP)的不同接入点的信号强度值, 形成关于位置的指纹数据库(RSSI 与位置的映射关系). 在线定位阶段, 采集待定位点信息, 利用相关匹配算法与指纹库中指纹进行匹配计算, 来估计待定位点的最佳位置的坐标点.

然而, 室内环境的复杂且多变导致了信号的非视距传播, 研究发现基于传播模型的定位算法精度, 一般为 10 米以上, 而基于指纹匹配的定位方法通过获取更多的样本信息, 在 80%-90%的可信区间内, 其定位精度可到 5-8 米左右. 2000 年微软公司开发了基于 Wi-Fi 网络的 RADAR 定位系统^[3], 此系统利用 Wi-Fi 指纹匹配的方法, 使用 K-NN 算法, 取最近 k 个邻居的坐标平均值为位置估计值, 并基于 RSSI 信号的统计特性, 采用贝叶斯原理, 通过计算目标位置的后验概率分布进行定位. Zhou 提出了基于本地信息熵的方法来实现 Wi-Fi 指纹匹配定位, 研究表明增加接入点, 可提高定位的精度^[4]. 王凤使用聚类的方法对射频的接收信号强度进行时间、位置、设备异构等因素进行聚类分析, 并使用朴素贝叶斯模型, 定位精度达 4m, 准确率达 80%, 定位速度提高 50%^[5]. 张勇提出一种 SVM 和加权质心相结合的算法, 定位精度和准确度较高, 但定位时间较长^[6]. 王超使用 KNN 方法实现基于指纹匹配的室内定位, 定位平均误差达 3.25m^[7]. 周瑞提出将支持向量机(SVM)分类与回归分析相结合的 Wi-Fi 指纹定位算法, 以提高定位精度^[8]. 田增山采用径向基函数插值的方法, 利用一部分 RSS 被重新测量的参考点, 拟合出接收信号强度曲面, 估计出邻近未知参考点 RSS 值, 从而更新指纹数据库^[9]. Joaquín 针对基于距离函数、RSS 值的数据表现方式以及阈值策略的 Wi-Fi 指纹的机器学习和专家系统进行了探讨^[10].

综上所述, 目前国内外学者的持续关注和深入研究运用机器学习方法来提高室内定位精度、鲁棒性、定位速度. 但由于无线信号的波动性给高精度鲁棒性定位带来较大的挑战, 基于 Wi-Fi 指纹的定位技术还需解决的核心问题包括数据的采集和数据的预处理, 所以本文的主要研究内容将包括: 采集和搭建基于 Wi-Fi RSSI 的室内定位指纹库以及设计增强型的聚类算法对 Wi-Fi 指纹定位库进行聚类分析, 旨在使定位

精度和定位速度达到平衡.

2 基于Wi-Fi 指纹库的定位算法研究

2.1 研究架构

本文研究分为三个阶段, 分别是 Wi-Fi 指纹样本采集并存储、机器学习算法应用和实验结果及定位性能分析. 本文的研究架构如图 1 所示.

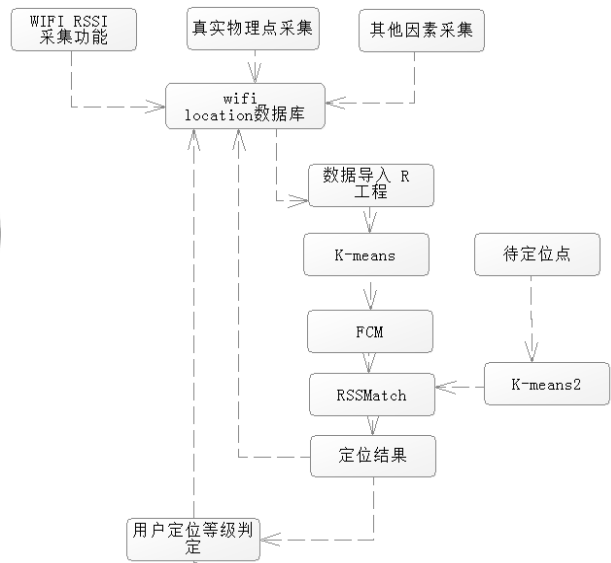


图 1 架构流程图

其中 Wi-Fi 指纹样本采集存储阶段主要功能为 Wi-Fi RSSI 信息的采集, 其他因素的采集和与真实物理点建立映射关系, 并将数据存于 wifi_location 数据库中. 定位阶段主要功能是使用 K-Means、FCM 对 Wi-Fi 指纹定位库进行聚类分析, 并使用自定义的匹配算法 RSSMatch, 对经过 K-Means 的待定位点进行指纹匹配, 得出最佳位置估计值. 实验结果及性能分析阶段主要是通过对比定位结果的分析, 得到定位性能各类指标, 并自动更新 wifi_location 数据库.

2.2 样本采集及搭建数据库

数据采集的实际场景的部分平面图如图 2 所示. 接入点(AP)覆盖广且密集, AP 信号易受其他信号干扰 (zigbee、蓝牙等), 网络环境复杂.

如图 2 所示, 黄色方格为 Wi-Fi 的 AP, 每层约有 18 个左右的 AP. 实验采用工具 Xirrus Wi-Fi Inspector 对实验场地内 Wi-Fi 信号的进行抓包处理, 每个采样位置点进行约 10-20 次采样, 其中每个方向按早中晚三个时间段进行 3 次采样, 得到如表 1 所示数据.

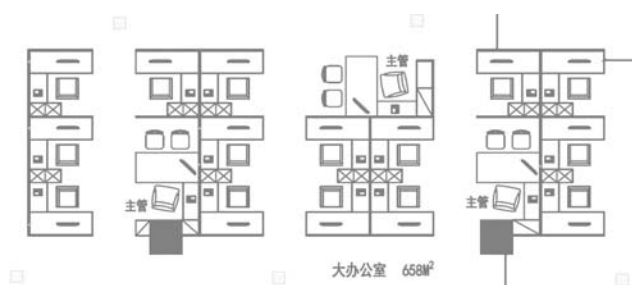


图2 数据采集的实际场景的部分平面图

表1 部分 Wi-Fi 信息

连接	SSID	强度	信道	频率	网络类型
FALSE	dotx-Wi-Fi	-25	11	2462	AP
FALSE	lierda	-28	1	2412	AP
TRUE	lierda	-31	11	2462	AP
FALSE	dotx-Wi-Fi	-32	6	2437	AP
FALSE	lierda	-33	6	2437	AP
FALSE	A92752	-34	6	2437	AP

其中“连接”为是否连接 AP; SSID 为 AP 的名称; “强度”为接收信号值, 其单位为 dbm.

从采集后的数据分析可知, Wi-Fi RSSI 在单个位置上存在约-10dbm 的数据波动和某个短时间的数据跳变现象. 但长时间内在室内环境无较大变化情况下, 各 AP 的 RSSI 在某位置保持相对稳定. 于是将当前此地点的各 AP 的 Wi-Fi RSSI 进行保存, 生成 Wi-Fi_network_report.csv 文件作为 Wi-Fi 信号原始数据. 并通过 wifi_location 进行缺损值及非法变量进行过滤, 得较完整的原始数据.

通过 Node.js 和 Express 搭建 web 服务. 此 web 服务为用户提供输入当前坐标值、方向、运动速率、是否遮蔽和是否为热点路径的数据接口, 并使用 MySQL 数据库对数据进行存储, 并过滤非法数据. 将 Wi-Fi 的_network_report.csv 数据和通过 web 服务得到的数据加载在基于 MySQL 的 wifi_location 数据库中. 使用 RODBC 组件, 将数据库 wifi_location 中数据导入到 R 工程中, 数据库数据导入 R 工程.

2.3 基于 K-Means 的定位过程

由于 Wi-Fi 指纹库中数据量庞大, 直接使用指纹匹配算法会导致计算量激增, 耗时严重, 并且定位精度低. 本文首先基于 K-Means 方法对 Wi-Fi 指纹库进行聚类分析, 具体实现步骤为:

(1) 对当前 Wi-Fi 指纹库提取指纹信息, 包括 RSSI、位置、AP 名称、AP 的 MAC、方向、是否遮蔽等信息, 加载于矩阵 wifio 中;

(2) 使用 R 中 K-Means 算法, 以欧式距离作为距离函数模型, 分别选取不同的聚类数对 RSSI、位置、AP 名称、AP 的 MAC 等进行聚类分析处理;

(3) 比较得到不同的聚类中心, 选择最优的聚类中心保存于 wifidata 中;

(4) 使用 RSSMatch 算法将聚类中心与待定位点的指纹进行匹配, 得到位置的估计值;

(5) 与真实的位置点进行比较, 评估定位性能.

在定位阶段, 对当前待测点的 Wi-Fi RSSI 和其他因素使用 K-Means 方法, 进一步减少计算量, 加快定位时间, 并在聚类后, 剔除部分异常数据, 提高定位精度. 具体实现如下:

(1) 对待测点进行每 0.25s 采集一次样本, 采集 2s, 共 4 次样本;

(2) 将样本导入 R 工程中, 进行 K-Means 聚类;

(3) 提取各聚类中心中子集数最多的 2 个簇中心, 作为当前此位置 Wi-Fi 指纹;

(4) 使用自定义匹配算法 RSSMatch, 与 Wi-Fi 指纹库中的已完成的聚类中心进行匹配, 估计当前位置坐标值;

(5) 与该地点的真实坐标值进行比较, 得出定位结果、精度误差和定位时间.

其中 RSSMatch 算法主要流程为:

(1) 将 wifidata 中的 RSSI 进行归一化处理;

(2) 将指纹库中的聚类中心、当前 Wi-Fi 指纹和递归步长输入, 一般为 0.1;

(3) 通过阈值计算, 分别用 Wi-Fi 指纹的各因素与聚类中心进行比较;

(4) 得到 5 个及以上的位置估计点, 则转(6);

(5) 以步长进行递归操作, 直到得到 5 个及以上的位置估计点;

(6) 得到的位置估计点, 再次进行数据过滤, 得到真正的位置估计点, 输出.

2.4 基于 FCM 的定位过程

由于 K-Means 存在不能发现大小差别很大的簇或非凸形状的簇, 对离群点和噪声比较敏感等缺点, 本文接着基于 FCM 算法对 Wi-Fi 指纹定位库进行聚类分析. 具体实现步骤为:

(1) 对当前 Wi-Fi 指纹库提取指纹信息, 包括 RSSI、位置、AP 名称、AP 的 MAC、方向、是否遮蔽等信息, 加载于矩阵 wifio 中;

(2) 使用 FCM 算法, 选择欧式距离作为距离模型, 设置单一簇最大子集数, 一般为 100, 设置初始簇数. 分别选取不同的聚类数对 RSSI、位置、AP 名称、AP 的 MAC、方向等进行聚类分析处理;

(3) 分别得到不同的聚类中心, 并进行比较, 选择最优的聚类中心保存于 wifidata 中;

(4) 使用 RSSMatch 算法将聚类中心与待定位点的指纹进行匹配, 得到位置的估计值;

(5) 与真实的位置点进行比较, 评估定位性能.

2.5 基于 ECS 的定位过程

实验结果表明, 由于 FCM 在定位精度优于 K-Means, 但在消耗时间方面, 却明显劣于 K-Means. 本文结合 FCM 和 K-Means 各自的优势, 提出增强型的聚类算法 (ECS) 应用于 Wi-Fi 指纹匹配定位中, 旨在使定位精度和定位速度达到平衡. 具体实现步骤如下:

(1) 对 Wi-Fi 指纹定位库进行 K-Means 分析, 其中簇数 N_k 为 300-5000, 得簇中心点 $\{Rk_1, Rk_2...Rk_n\}$, 其中 $k_n=N_k$;

(2) 对簇中心 $\{Rk_1, Rk_2...Rk_n\}$ 使用 FCM 分析, 其中簇数 N_f 为 20-100, 得簇中心 $\{Rf_1, Rf_2...Rf_n\}$, 其中 $f_n=N_f$;

(3) 使用自定义匹配算法 RSSMatch 与簇中心 $\{Rf_1, Rf_2...Rf_n\}$, 其中 $f_n=N_f$, 进行匹配, 得到定位估计值, 并评估定位性能.

3 实验结果分析

使用 R 绘制的聚类效果如图 3 所示. 从该图可知, FCM 的聚类效果要好于 K-Means 的聚类效果, ECS 的聚类效果最佳.

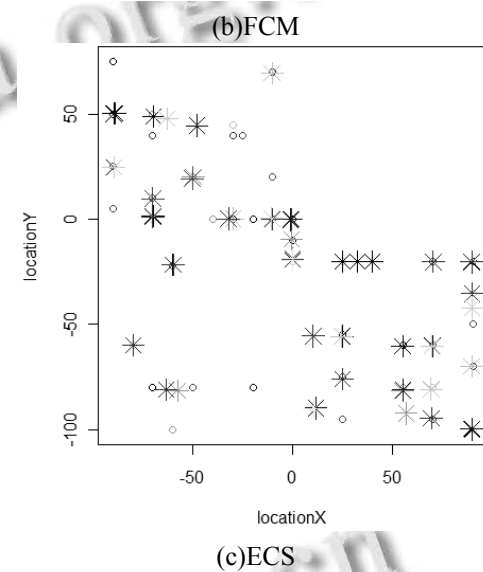
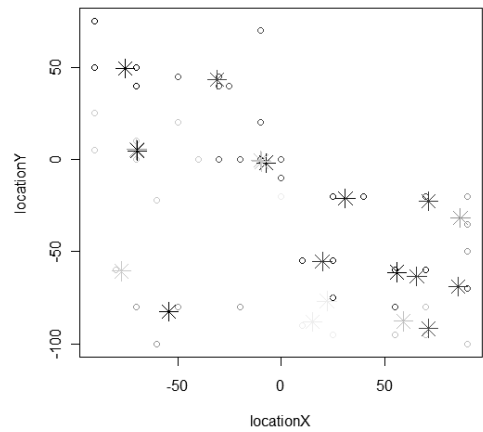
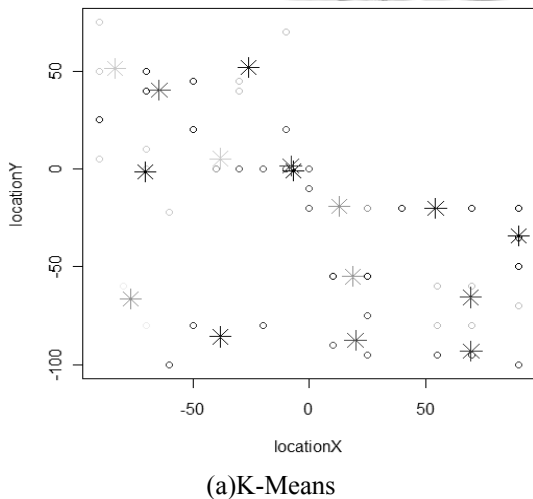


图 3 Wi-Fi 指纹聚类中心图

当 Wi-Fi 指纹定位库聚类数分别为 10, 20, 50, 75, 100, 150, 200, 300 时, 表 2 和表 3 分别为基于 K-Means 和 FCM 得到定位性能表. 如表可知, 当指纹库的聚类数为 50-100 时均得到较好的定位效果, 定位精度和定位速度均较好. 在定位精度上 FCM 总体优于 K-Means, 但消耗时间要明显劣于 K-Means 算法. 在聚类数为 75 时, 估计点为 1 个, 精度误差为 0.9966m, 为最佳定位效果.

表 2 K-Means 定位性能表

指纹库聚类数	10	20	50	75	100	150	200	300
估计点个数	0	1	4	3	5	7	8	13
平均精度(m)	无	9.131	5.483	4.672	6.8220	7.890	10.3700	9.7590
最小精度(m)	无	9.131	1.693	0.7171	0.7276	1.411	0.5503	0.200
最大精度(m)	无	9.131	7.849	8.004	13.7600	14.410	33.2500	17.990
中位数精度(m)	无	9.131	6.906	7.7430	6.0960	6.989	8.7300	9.7590
运行时间(s)	0.01	0.03	0.04	0.06	0.39	0.43	0.44	0.5

表 3 FCM 的定位性能表

指纹库聚类数	10	20	50	75	100	150	200	300
估计点个数	0	0	2	1	3	8	6	6
平均精度(m)	无	无	4.961	0.996	3.9150	7.9380	6.998	5.529
最小精度(m)	无	无	0.965	0.996	0.9915	0.8225	1.072	1.022
最大精度(m)	无	无	8.957	0.996	6.0920	14.2500	11.730	11.71
中位数精度(m)	无	无	4.961	0.996	4.6620	8.8840	6.118	5.673
运行时间(s)	3.14	15.69	38.25	64.21	87.42	140.14	201.52	273.89

ECS 的定位性能如表 4 所示, 从该表的数据可知, 随着聚类数的增加, 运行时间也增加, 但整体提高了定位性能. 较仅使用 FCM 算法, ECS 算法的时间缩短约 50%-80%, 且定位精度上有所改善; 在定位精度方面, 较仅使用 K-Means 算法, ECS 算法的定位精度提高约 20%-40%, 且定位稳定性较强.

表 4 ECS 的定位性能表

K-Means 聚类数	300	300	500	500	500	500	1000	1000
FCM 聚类数	50	100	20	50	75	100	20	50
估计位置个数	1	3	0	3	2	3	0	3
平均精度(m)	0.68843.4920	无	7.989	8.8470	5.9070	无	7.1780	
最小精度(m)	0.68840.4872	无	0.770	0.8502	0.7734	无	0.8605	
最大精度(m)	0.68845.2910	无	14.12	14.2308	6.560	无	11.310	
中位数精度(m)	0.68844.6980	无	9.080	11.4608	8.2920	无	9.3630	
运行时间(s)	1.38	1.51	1.48	2.49	4.14	4.23	2.35	4.78

为了进一步确定 K-Means 和 FCM 选取的初始最佳聚类数, 本文将估计位置个数少于 4, 平均精度小于 8.000m, 最小精度小于 1.5000m, 最大精度小于 10.000m, 进行数据整理, 得到的数据如表 5 所示.

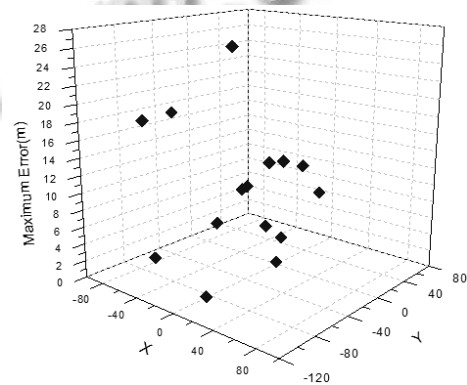
表 5 ECS 的较佳定位性能表

K-Means 聚类数	300	300	500	2000	3500	3500	5000	5000
FCM 聚类数	50	100	100	50	50	75	20	50
估计位置个数	1	3	3	2	3	4	1	2
平均精度	0.68843.4920	5.9070	4.703	2.0270	3.0840	0.66310.9391		
最小精度(m)	0.68840.4872	0.7734	0.705	0.6576	0.6324	0.66310.9159		
最大精度(m)	0.68845.2910	8.6560	8.702	4.6850	5.3590	0.66310.9622		
中位数精度(m)	0.68844.6980	8.2920	4.703	0.7373	3.1730	0.66310.9391		
运行时间(s)	1.38	1.51	4.23	9.16	14.37	22.71	8.66	20.28

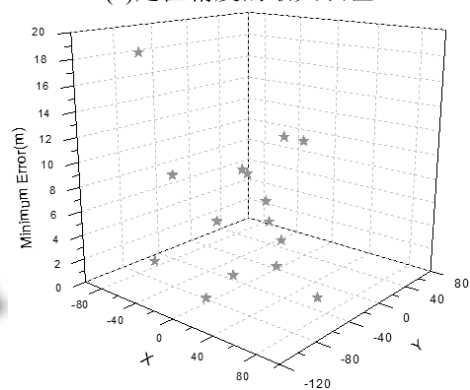
由表 5 的数据可知, 在 K-Means 聚类数较小时, 定位精度较高且定位速度较快. 而 FCM 的聚类数一般保持在 50 左右, 定位效果最佳. 其中当 K-Means 的聚类数为 300, FCM 的聚类数为 50 时, 估计位置个数为 1

个, 平均精度为 0.6884m, 运行时间为 1.38s. 为了达到定位精度和定位速度的平衡, 本文在结合 K-Means 和 FCM 优势的实验过程中, 以 K-Means 为主, FCM 为辅的策略, 一般选择 K-Means 的聚类数为 300-1000, FCM 聚类数为 50-75. 这策略的平均精度可达 4m 以内, 最小精度误差达 1m 以内, 并且估计位置数较少, 运行时间较少.

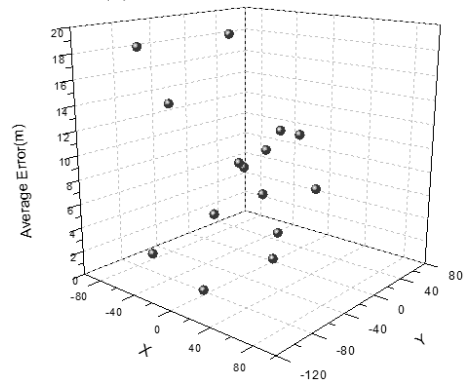
本文使用 K-Means 聚类数为 300, FCM 聚类数为 75, 在实际场景中, 随机定位 50 次, 验证此策略的准确性. 图 4 为部分待定位点的定位性能表.



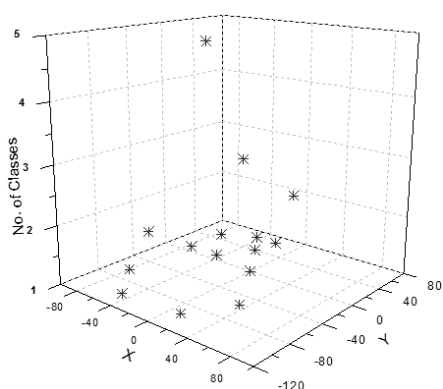
(a) 定位精度的最大误差



(b) 定位精度的最小误差



(c) 定位精度的平均误差



(d)定位点个数

图 4 部分待定位点的定位性能分析

由实验可知,各待定位点的定位效果不尽相同,84%的定位点的平均误差小于 8m,71%的定位点的最小误差小于 5m,28%的定位点的平均误差小于 2m,平均聚类数为 1.53 个.实验发现,各待定位点的定位性能与该待定位点在 Wi-Fi 指纹定位库中的指纹的稀疏程度有关,如果待定位点附近的指纹越密集,定位效果越好,如待定位点(-27,40)和(0,0);反之,如果待定位点附近指纹稀疏,则定位效果较差,如待定位点(-36,-82)和(10,-65).

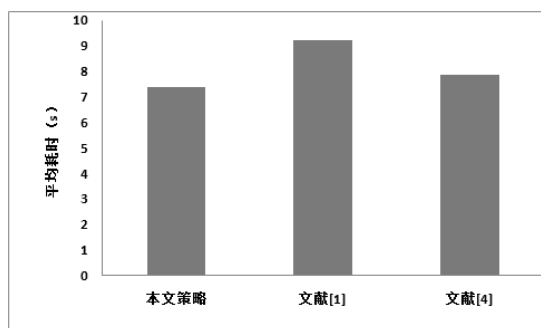


图 5 平均耗时比较

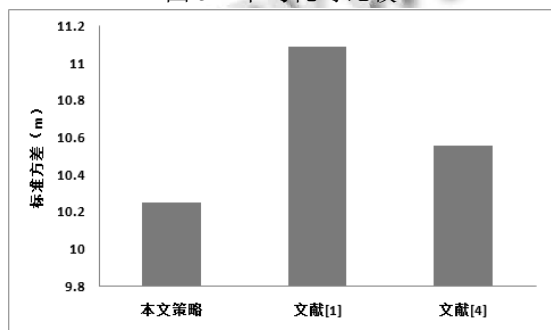


图 6 定位点误差的标准方差比较

同时,在机器配置为 Intel core i3 3.4G CPU 和 4G

内存的情况下,把本文的策略和已发表算法就定位处理的平均耗时和定位点误差的标准方差进行算法的性能比较,得到如图 5 和图 6 所示的结果,从图中可以看出,相对于其它的算法,本文所提出的策略具有更有的定能性能.

4 总结

本文基于 Wi-Fi 指纹库,研究了基于 Wi-Fi 指纹匹配方法的室内定位技术.此研究提高了室内定位的精度,减少定位的时间,增强定位的稳定性,达到了预期的效果.但随着采样数据的增加,指纹库不断变化和越发庞大,使用 K-Means 和 FCM 都需要提前确定簇数,如何动态自适应选择簇数,来尽可能实现更好的聚类分析,提供可靠的位置估计值,提高定位性能,是下一阶段的研究重点.同时,如何较均匀地采集 Wi-Fi 指纹,以进一步提高定位性能,也是以后研究的方向.

参考文献

- 颜俊杰.基于 Wi-Fi 的室内定位技术研究[硕士学位论文].广州:华南理工大学,2013.
- 其宽.基于位置指纹的无线室内定位技术研究[硕士学位论文].秦皇岛:燕山大学,2015.
- Bahl P, Padmanabham V. RADAR: An in-buliding RF-based user location and tracking system. Institute of Electrical & Electronics Engineers Inc, 2000, 2: 775-784
- Zhou M, Tian Z, Xu K, Yu X, Wu H. Theoretical entropy assessment of fingerprint-based Wi-Fi localization accuracy. Expert Systems with Applications, 2013, 40(15): 6136-6149.
- 王凤.基于聚类的射频指纹定位技术研究[硕士学位论文].北京:北京邮电大学,2014.
- 张勇,徐小龙,徐科宇.基于加权质心法的 WLAN 室内定位系统.电子测量与仪器学报,2015,29(7):1-4.
- 王超.基于 WLAN 室内定位的分类算法研究与实现[硕士学位论文].北京:北京邮电大学,2015.
- 周瑞,袁兴中,黄一鸣.基于卡尔曼滤波的 WiFi-PDR 融合室内定位.电子科技大学学报,2016,45(3):399-404.
- 田增山,代海鹏.基于曲面拟合的 WiFi 指纹数据库更新.计算机应用,2016,36(5):1192-1195.
- Joaquín TS, Raúl M, Sergio T. Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems. Expert Systems with Applications, 2015, 42(23): 9263-9278.