

文章编号: 2095-2163(2019)02-0078-05

中图分类号: TP311.13

文献标志码: A

不同聚类算法在 Wi-Fi 定位中的研究

陈 蕾

(浙江农林大学 信息工程学院, 浙江 临安 311300)

摘要: 本文比较分析了 K-means 聚类、蚁群聚类和 DBSCAN 聚类三种聚类算法在室内定位系统中的应用, 通过比较 3 种算法分别与基于 Wi-Fi 信号接收强度的支持向量机回归算法室内定位模型相结合定位的仿真实验, 发现 K-means 和 DBSCAN 聚类算法均优于蚁群聚类算法, 且两者定位结果相差不大。考虑到 DBSCAN 聚类算法可以去除噪音点, 将 DBSCAN 聚类与 K-means 聚类算法混合分析后, 再与定位模型结合, 获得更好的定位效果。

关键词: K-means 聚类; DBSCAN 聚类; 室内定位

Research on different clustering algorithms in Wi-Fi location

CHEN Lei

(College of Information Engineering, Zhejiang A&F University, Lin'an Zhejiang 311300, China)

[Abstract] This paper compares and analyzes three clustering algorithm such as the K-means clustering, ant clustering and DBSCAN clustering in the application of indoor location system. The three algorithms are compared respectively combined with the indoor location model of the support vector machine (SVM) regression based on Wi-Fi and receiving signal strength. The simulation experiment found that K-means and DBSCAN clustering algorithm are better than the ant clustering algorithm, and the location results of the aboved two are not much different. Considering that DBSCAN clustering algorithm can remove noise points, DBSCAN clustering is mixed with K-means clustering algorithm for analysis, and then combines with location model, which could obtain better location effect.

[Key words] K-means clustering; DBSCAN clustering; indoor location

0 引言

随着信息技术的高速发展, 基于位置服务的应用越来越受到学界关注。以卫星导航等传统方法为代表的定位技术由于 GPS 等卫星信号到达地面会有所减弱, 难以穿透建筑物, 因此在室内很难使用卫星定位系统获得准确的定位信息, 精度也不高。但是人在室内活动的平均时间约占整体的 80% 左右, 因此室内定位的应用已逐渐成为研究焦点。

时下, 通讯技术的不断进步, 即使得智能移动设备日趋成为社会生活中的时尚新宠儿。根据媒体市场研究公司 Zenith 对 52 个国家和地区的调查报告研究显示, 在 2017 年底全球智能手机普及率达 63%, 其中在一些领先的国家和地区, 智能手机使用的普及率都已经超过了 90%。智能移动设备的普及, 为室内定位的发展和应用带来了更广阔前景, 但同时也提出了更高的要求。

室内定位是指在室内环境中获取位置定位, 常见的主要有采用无线通讯、基站定位、惯导定位等多种技术集成形成一套室内位置定位体系, 从而实现

人员、物体等在室内空间中的位置监控。室内定位作为国内外研究的热点, 其研究成果在商场购物、交通引导、景区导览、人员设备管理、公共服务、紧急救援等方面都可见到典型应用。目前堪称主流的室内定位技术有: 红外线、超声波、超宽带、蓝牙、Wi-Fi、基站、RFID 技术, 此外还有比较新颖的可见光通信、地磁技术、视觉定位等。不同的定位技术均有各自的优缺点, 结合应用场景, 选择合适的定位技术至关重要。研究可知, 基于 Wi-Fi 技术的室内定位是现下最常见的室内定位技术, 虽然在众多定位技术中 Wi-Fi 技术的定位精度不是很高, 但由于现代生活中 Wi-Fi 的使用已随处可见, 各类环境中 Wi-Fi 设施布置完备, 技术相对简单, 又减少了硬件设备架构的成本, 因而使其在各类定位方法中仍然具有鲜明的优势。对此, 本文拟展开研究论述如下。

1 基于 Wi-Fi 的室内定位

基于 Wi-Fi 技术的定位系统作为现下最常见的室内定位技术, 相关的室内定位算法非常多, 大体可以分为基于几何方法的定位、极大似然估计法定

作者简介: 陈 蕾(1989-), 女, 硕士研究生, 主要研究方向: 农林物联网、室内定位。

收稿日期: 2018-12-25

位和基于接收信号强度(RSSI)建立指纹库的定位。这里,对此可做探讨分述如下。

(1)几何方法。是通过测量与用户相关的几何参数来获得用户的位置。几何方法多可分为 2 种。一种是根据信号到达时间或到达时间差、到达角度来测算待测点与信号发生源之间的距离。另一种是利用信号传播模型来测距,常根据一些无线信道的半统计或者经验模型,利用在线测得无线信号强度估计距离。

(2)极大似然估计法。也是基于接收信号强度的 Wi-Fi 定位中一种主要的定位方法,极大似然估计法是建立在极大似然原理基础上的一个统计方法,极大似然定位算法则是在几何三边定位算法中加入极大似然估计原理,与几何方法有所区别,且具有较大的实用性。

(3)基于接收信号强度的指纹法定位。由于前文的几何方法受非视距传播路径影响较大,且测距成本不低,因而有研究者提出了该种定位方法。指纹法是在离线阶段将接收信号强度的某些特征与环境中的地理位置建立对应关系,构成指纹库。而后在线阶段则实时提取接收信号强度的信号特征,在指纹库中寻找对应数据,最终实现定位。相对而言,指纹法研究起来比较简单,有一定的实用研究价值。

2 聚类算法

聚类分析起源于分类学,但在此后科学技术的飞速发展,人类对分类的实现也已从仅主要依靠经验和专业知识而衍生转变为逐步融入了数学工具的定量分析,形成数值分类学,又引入了多元分析形成了聚类分析。聚类分析内容丰富,其算法可以分为划分法、层次法、基于密度的方法、基于网格的方法、基于模型的方法等。研究内容详见如下。

2.1 K 均值聚类算法

K 均值(K-means)聚类算法属于划分法,该方法需要输入聚类个数 k , 然后对数据对象进行聚类,输出满足方差最小标准的 k 个聚类,同一聚类中的对象相似度较高,不同聚类中的对象相似度较低。K-means 聚类是根据数据对象之间的相似度来间接聚类的,是无监督的学习方法。

2.2 蚁群聚类

蚁群聚类算法是 1991 年由 Deneubourg 提出的,而 Lumer 和 Faieta 将蚁群聚类算法模型应用到了数据分析的领域,所以蚁群聚类算法尚且还是一种比较新颖的算法。蚁群聚类算法是结合蚂蚁选择

觅食路径的行为和蚁群在蚁穴中将散落的蚂蚁尸体堆积成堆来清理的行为实现聚类的。蚁穴清理的行为是蚂蚁堆越大,越吸引工蚁搬运过去,因此,数据的空间分布直接影响聚类结果。

2.3 DBSCAN 聚类

DBSCAN (Density-Based Spatial Clustering of Application with Noise) 聚类算法是一种基于密度的聚类算法。该算法是在 1996 年由 Ester 等提出的一种简单又有效的聚类算法。算法可描述为:在某些空间中给定一组点,它将位置紧密靠在一起的点(与许多邻近邻域的点)组合在一起,标记为异常值点,单独存在低密度区域(其与最近的邻域点距离太远)。DBSCAN 算法是对数据样本进行划分的聚类算法,且事先并不需要知道数据样本的标签,是一种非监督的聚类算法。核心点(Core Point)是指在给定的聚类半径 ϵ 中,若是含有超过设定参数 $MinPts$ 数量的点,则该点被称为核心点。边界点(Border Point)是指在给定的聚类半径 ϵ 中,点的数量小于所设定的参数 $MinPts$, 那么就不属于核心点,但是该点又落在核心点的邻域之中,则称该点为边界点。噪音点(Noise Point)是指在给定的聚类半径 ϵ 中,点的数量小于所设定的参数 $MinPts$, 那么就不属于核心点,但是该点又不落在任一核心点的邻域之中,那么该点被称为噪音点。

3 实验仿真

本文实验是通过对接收到的 Wi-Fi 信号强度先进行聚类分析,根据分析后的特征值再输入支持向量机的回归模型,实现待测点的位置定位。文中对此可得研究详述如下。

3.1 数据来源及处理

本文实验环境为 Window10 操作系统, Matlab 2016a(9.0.0.341360)版本,实验数据采集自浙江农林大学 B10 学生宿舍楼的第二层,过程中选择在该楼层采集 35 个点的所有无线接入点(AP)的信号强度。待测位置 RSSI 值采集的设备采用 IphoneX,该设备中的开发 App: AirPort 工具可以采集该位置所接收到的信号强度。如图 1 所示。

在 35 个采集点停留 1 min 左右,进行信号强度采集,共采集 AP 名为 ChinaNet 的信号强度数据 2 931 条,根据不同的时刻将数据进行分类,分类后数据按照 AP 的硬件地址进行排序,最终获得实验数据 330 条。



图1 AirPort工具采集Wi-Fi的RSSI

Fig. 1 AirPort tool collects the RSSI of Wi-Fi

在330条实验数据中选取300条作为训练数据集,其余30条数据作为测试数据集,将训练数据集作为聚类算法的数据对象,进行聚类分析。再将聚类结果作为SVM的输入参数,对应的位置坐标 (x,y) 中的横坐标和纵坐标作为输出参数,分别进行训练,建立SVM模型。将测试数据集作为输入参数,分别输入训练好的SVM模型,获得横坐标和纵坐标的实验数据集,输出实验结果,即为定位结果。

3.2 3种不同聚类算法的实验仿真

K-means聚类算法的实验中 k 取35,同坐标点数量一致,迭代5次,实验仿真结果如图2和图3所示,定位结果分析见表1。

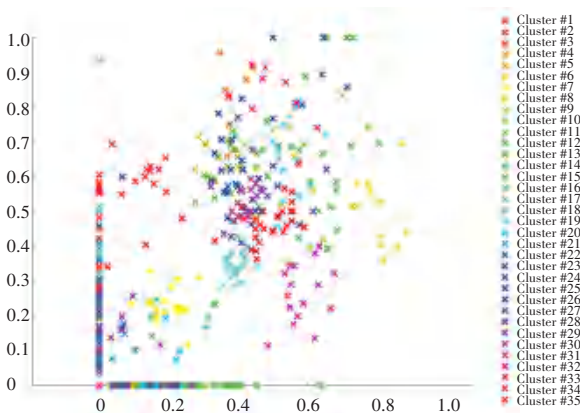


图2 K-means聚类结果

Fig. 2 K-means clustering results

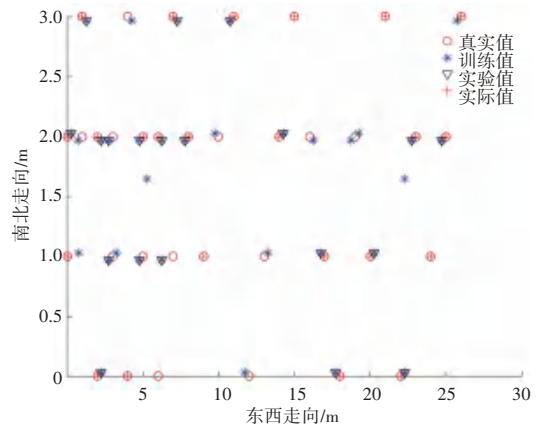


图3 K-means-SVM定位结果

Fig. 3 K-means-SVM location results

表1 K-means-SVM定位结果分析

Tab. 1 Analysis of K-means-SVM location results m

	最大偏差	最小偏差	平均偏差
训练数据	9.23	0.25	1.24
测试数据	4.77	0.25	0.97

根据实验结果可得,使用K-means聚类分析输入数据后,获得较好的定位结果。

蚁群聚类算法的实验中 k 取35,同坐标点数量一致,迭代5次,实验仿真结果如图4和图5所示,定位结果分析见表2。

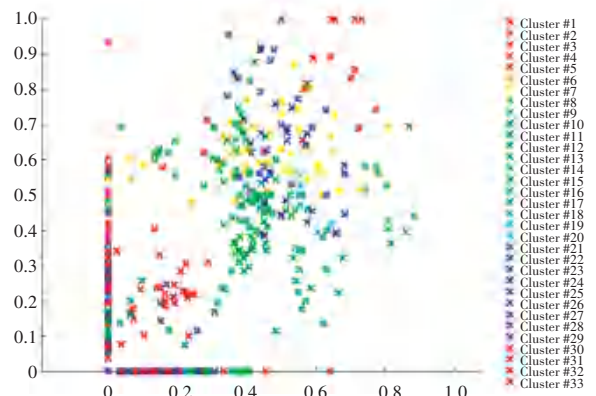


图4 蚁群算法聚类结果

Fig. 4 Ant colony algorithm clustering results

表2 蚁群算法SVM定位结果分析

Tab. 2 Analysis of ant colony algorithm SVM location results m

	最大偏差	最小偏差	平均偏差
训练数据	9.23	0.25	1.55
测试数据	8.80	0.25	1.68

根据实验结果可得,使用蚁群聚类分析输入数据后,获得定位结果比K-means差。

DBSCAN聚类实验经过多次实验,选择参数邻域半径 $\varepsilon = 0.2$,邻域样本数阈值 $MinPts = 3$,实验仿真如图6和图7所示,定位结果分析见表3。

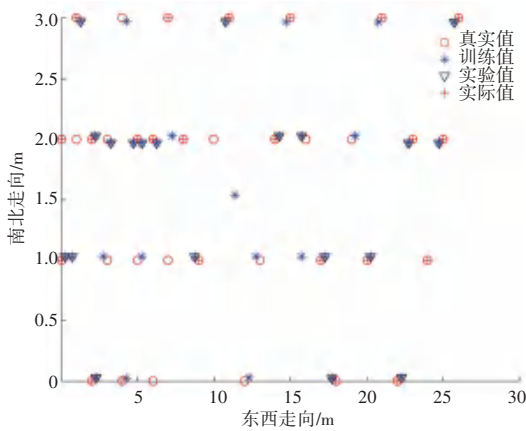


图 5 蚁群算法 SVM 定位结果

Fig. 5 Ant colony algorithm SVM location results

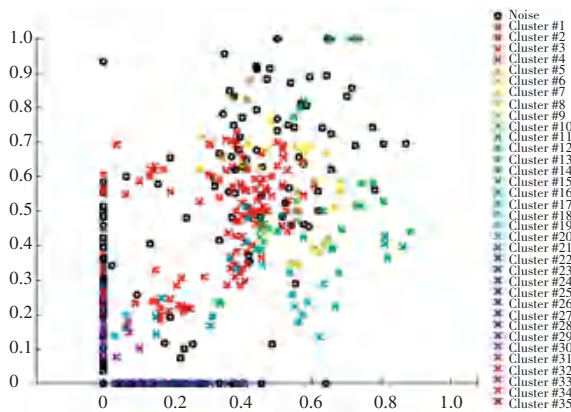


图 6 DBSCAN 算法聚类结果

Fig. 6 DBSCAN clustering results

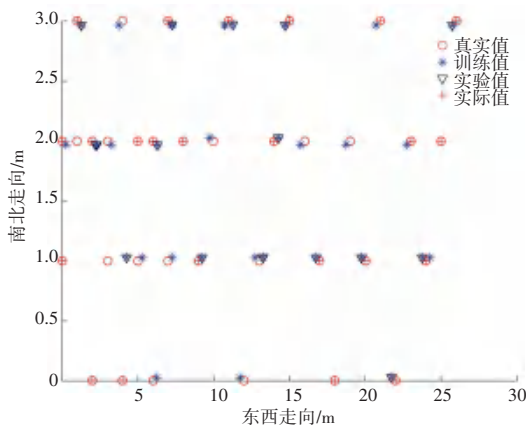


图 7 DBSCAN-SVM 定位结果

Fig. 7 DBSCAN-SVM location results

表 3 DBSCAN-SVM 定位结果分析

Tab. 3 Analysis of DBSCAN-SVM location results m

	最大偏差	最小偏差	平均偏差
训练数据	7.73	0.25	1.07
测试数据	5.75	0.25	1.04

由于 DBSCAN 聚类能自动标识噪音点, 去除噪音点后能获得较好的定位结果。但是 K-means 聚类虽然没有去除噪音点, 定位结果仍然是 3 种聚类算法中最好的。

3.3 DBSCAN-Kmeans 混合聚类实验仿真

研究可知 K-means 聚类定位结果最好, 但却不能自主去除噪音点, 因此会影响到实验结果。基于此, 研究将 DBSCAN 聚类算法与 K-means 聚类算法相结合, 先利用 DBSCAN 算法去除噪音点, 再使用 K-means 聚类对去除噪音点的数据进行聚类分析。最后将聚类结果通过支持向量机的回归模型研究处理后, 获得定位结果。实验仿真如图 8 和图 9 所示, 定位结果分析见表 4。

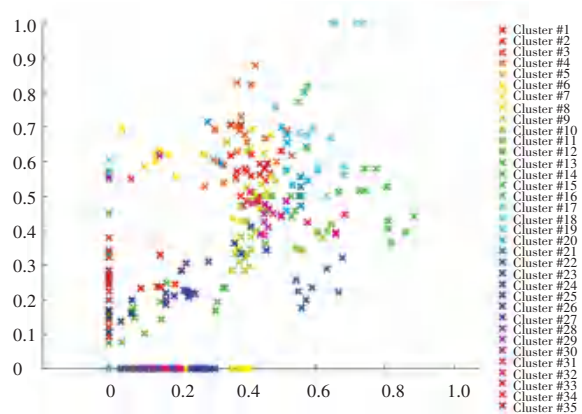


图 8 去除噪音点的 K-means 算法聚类结果

Fig. 8 K-means algorithm with noise point removal clustering results

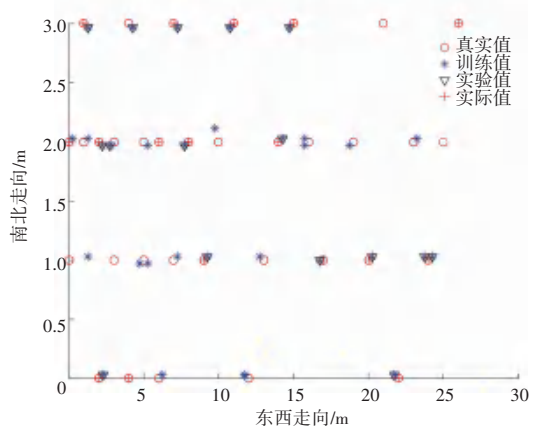


图 9 DBSCAN-K-means-SVM 定位结果

Fig. 9 DBSCAN-K-means-SVM location results

表 4 DBSCAN-K-means-SVM 定位结果分析

Tab. 4 Analysis of DBSCAN-K-means-SVM location results m

	最大偏差	最小偏差	平均偏差
训练数据	5.80	0.25	0.70
测试数据	4.23	0.25	0.78

根据实验结果可得, 使用 DBSCAN 聚类分析输入数据后, 获得定位结果与 K-means 相差不大。

(下转第 88 页)