

一种模糊聚类 KNN 位置指纹定位算法*

都伊林

(浙江警官职业学院, 浙江 杭州 310018)

摘要: 阐述了位置指纹定位算法在室内 WLAN 环境中的应用, 分析了 KNN 定位算法存在的不足, 提出一种模糊聚类 KNN 位置指纹定位算法。该算法首先选取与空间相关性较好的 4 个信号参数, 构成多径信号数据库; 然后应用主分量分析法 (PCA) 对原始信号数据库作降维运算, 滤除奇异性接入点 (AP); 最后用模糊 C 均值聚类算法 (FCM) 处理数据, 进一步滤除奇异性参考点 (RP), 实现提高定位算法效率与精度的目的。实验表明, 改进后的定位算法产生的定位误差明显减小。

关键词: 位置指纹; 室内定位; 模糊聚类; KNN 定位算法; 信号数据库

中图分类号: TP391

文献标识码: A

文章编号: 1674-7720(2012)23-0055-04

A fuzzy clustering KNN location fingerprint positioning algorithm

Du Yilin

(Zhejiang Police Vocational Academy, Hangzhou 310018, China)

Abstract: The paper describes the application of location fingerprint positioning algorithm in indoor WLAN environment, and analyzes the shortcomings of the KNN positioning algorithm, and proposes a fuzzy clustering KNN location fingerprint positioning algorithm. Firstly, it selects four signal parameters with good spatial correlation, and constitutes a database of multipath pattern signal. Secondly, it does dimensionality reduction operations to the original signal database by application of principal component analysis (PCA), to filter out the the singularity access point (AP). Then, by application of fuzzy C-means clustering algorithm (FCM), it processes the data to further filter the singular reference point (RP), to achieve the purpose of improving the efficiency and accuracy of the positioning algorithm. The experiments show that the positioning errors generated by the improved positioning algorithm is significantly reduced.

Key words: location fingerprint; indoor positioning; fuzzy clustering; KNN positioning algorithm; signal database

随着现代通信技术和网络技术的发展, 人们可携带计算设备的广泛应用, 以及国内城市开展无线城市试点工作, 用户可以通过计算设备随时随地接入互联网, 由此, 基于位置的服务 LBS (Location-based Services) 也受到了社会越来越多的关注。与此同时, 基于卫星导航定位技术的全球定位系统 GPS (Global Position System) 已在众多领域得到普及。

然而, 在室内环境下, 由于卫星信号被物体阻挡, 无线信号不能正常传输, GPS 的导航功能无法正常实现, 且无线传感定位系统要有专用传感器和网络支持, 需化费较多人力和财力。因此, 应用室内区域的 WLAN 网络 (如 Wi-Fi) 进行移动目标的定位管理是一个较适宜的

解决方案。

1 定位算法

基于 Wi-Fi 网络的室内定位系统大多数是利用接收信号强度 (RSS) 的均值, 其方法一般分为信号传输模型法和位置指纹识别法两类。前者是利用待测点接收至少 3 个接入点之间的距离信息, 由一定算法估计待测点的位置; 后者是通过待测点多径信号特征指纹信息与数据库预存参考点多径信息进行比对分析, 系统需运行大量数据, 用一定算法估计待测点坐标。

1.1 信号传输模型定位算法

信号传输模型定位算法可以分为测距与定位两个阶段。首先, 待测点接收来自 3 个不同已知位置接入点 AP 的信号强度值, 通过中值滤波技术提取均值; 然后,

* 基金项目: 浙江警官职业学院科研项目 (201008)

网络与通信 Network and Communication

根据无线信号的室内传输损耗模型,将接收信号强度转换为待测点与相应 AP 的距离;最后,应用三角形定位算法估算。

无线信号的室内传播模型^[1,9],一般简化为:

$$P_i(d_i) = P(d_0) + 10n \lg\left(\frac{d_i}{d_0}\right) + X_\sigma \quad (1)$$

其中, d_i 是待测点到接入点的实际距离; d_0 是参考点到接入点的距离; n 是信道衰减指数,一般取值为 2~4; X_σ 为均值为零,方差为 σ 的高斯分布的随机噪声变量; $P(d_0)$ 是参考点处的信号强度; $P_i(d_i)$ 是待测点处的信号强度。

所谓三角形定位算法,是通过以已知位置的 3 个 AP 为圆心,以各 AP 到待测点的距离为半径作圆,所得 3 个圆的交点即为待测点位置;用数学方法联解 3 个方程,可以得到待测点的坐标。

此外,由式(1)可知,信号强度值具有较强的空间相关性,当建筑物室内布局发生变化,或人员频繁走动和阻挡时,信号强度将会随之发生波动。因此,实际上信号传输损耗模型无法满足实际环境的变化情况,特别是室内信号多径衰落因素会导致定位算法有较大的误差。为了有效地克服信号传输模型的上述缺点,人们提出了位置指纹识别算法^[2]。

1.2 位置指纹定位算法

基于位置指纹定位的方法一般分为离线训练阶段与在线定位阶段。离线训练阶段是建立一个位置指纹数据库,在无线信号覆盖区设置网格化的参考点 RP,测量参考点的多径信号特征参数,大多数文献采用均值预存在数据库^[3],即构成位置指纹地图(Radio Map)。在线定位阶段是用待测点在某一位置的信号特征参数,通过相应的匹配算法,将实测数据与预存数据进行比较分析,选取与实测数据相匹配的预存数据,从而可以估计出待测点坐标。在常用的匹配算法中,KNN 定位算法因运算简单与定位精度较高,而应用较为广泛。

1.3 KNN 定位算法

首先,计算待测点所测量到的 RSS 均值矩阵 $[s_1, s_2, \dots, s_n]$ 与数据库中参考点 RP 的 RSS 均值矩阵 $[s_1, s_2, \dots, s_N]$ 之间的距离,其中 $n \leq N$ 。因为,通过扫描待测点周围的 AP 收到它们的 RSS 值,设置一个阈值(如 -90 dBm)并按大小排序,选择其中若干实测数据,由宏地址确定具体 AP。所以,待测点所收到信号的 AP 数要小于整个区域布置的 AP 数。设数据库存有 m 个参考点 RP,有 N 个接入点 AP, S_{ij} 为第 i 个 RP 处的第 j 个 AP 的 RSS 均值, s_j 为测量到的第 j 个 AP 的 RSS 均值, $i=1, 2, \dots, m; j=1, 2, \dots, n$ 。则,各向量之间的距离定义如下:

$$L_{qi} = \left(\sum_{j=1}^n |s_j - S_{ij}|^q \right)^{\frac{1}{q}} \quad (2)$$

式(2)可作为实测数据与预存数据相似度的判据。当 $q=1$ 时, L_{qi} 称为曼哈顿距离;当 $q=2$ 时, L_{qi} 称为欧几

里德距离。KNN(K Nearest Neighbors)算法^[4-5]是在上面的 L_{qi} 中,从小到大选择 k 个参考点 RP($K \geq 2$),实际上确定了待测点周边的 RP。

其次,可利用式(3)计算其平均坐标作为待测点的估计位置。

$$(\hat{x}, \hat{y}) = \frac{1}{k} \sum_{i=1}^k (x_i, y_i) \quad (3)$$

其中, (x_i, y_i) 表示 k 个参考点中的第 i 个的坐标,选取合适的 k 值,以确保较好的定位精度; (\hat{x}, \hat{y}) 为待测点的估计坐标。当 $k=1$ 时,算法退化为最近邻(NN)算法,即用最近的一个参考点估计待测点的坐标。

KNN 定位算法的特点是:在计算实测数据与预存数据的距离后,根据式(2)选择相似度最近的个近邻参考点进行位置估计。若实测数据与预存数据之间距离越近,则估计坐标的精度越高。此算法不足之处在于:选择值是固定的,会影响某些位置的定位精度;在计算距离时,会丢掉许多位置信息,也将导致精度下降。因此,本文提出一种模糊聚类 KNN 定位算法,目的是提高定位精度。

1.4 模糊 C 均值聚类算法

模糊 C 均值聚类算法^[6](FCM)是将 n 个数据 $X = \{x_1, x_2, \dots, x_n\}$ 模糊划分为 C 个数据组 X_1, X_2, \dots, X_C ($X_1 \cup X_2 \cup \dots \cup X_C = X$),求每组的聚类中心,使得非相似性指标的目标函数达到最小化,其可表示为:

$$J_{FCM}^m(U, C, X) = \sum_{i=1}^C \sum_{j=1}^n \mu_{ij}^m d_{ij}^2 \quad (4)$$

其中, C ($C > 1$) 是分组数目; m 是模糊度,控制算法的柔性参数。若 m 过大,则聚类效果会很差;若 m 过小,则算法会接近 C 均值聚类算法,通常取值为 2。 $\mu_{ij} \in [0, 1]$ 表示 x_j 属于 X_i 的程度,即隶属度,且必须满足 $\sum_{i=1}^C \mu_{ij} = 1, \forall j = 1, 2, \dots, n; d_{ij}$ 是衡量第 i 个聚类中心与第 j 个数据间的非相似性指标。

FCM 算法是一个简单的迭代与优化过程:用值在 0~1 间的随机数初始化隶属矩阵 U ,或初始化聚类中心,通过反复的迭代运算,逐步降低目标函数的误差值,当目标函数值收敛时,得到最终聚类结果。该算法适合于正态分布的数据聚类,对于奇异性孤立点数据有敏感性,因此,可用于基于信号强度的定位算法。

FCM 算法因算法简单,收敛速度快,且能处理大批量数据,解决应用性问题广。本文应用 FCM 算法可以实现从待测点角度对相关数据进行有效的处理;聚类处理后可以滤除奇异性 RP。聚类处理前后数据比较如图 1 所示。实际上这类 RP 距离待测点较远,会引起较大的定位误差。因此,FCM 算法能起到“数学聚焦”的功能,有助于提高定位精度。所谓奇异性 RP 是指影响定位精度较大的参考点。

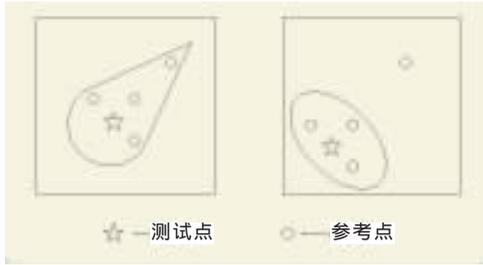


图1 聚类处理前后数据比较

2 模糊聚类 KNN 定位算法

2.1 信号指纹数据库

为了克服信号强度 RSS 值对信道传输模型的依赖性,如多径效应、墙壁阻挡和环境条件的变化等因素,提出了使用位置指纹定位算法。本文分析了信号强度各类特征参数与空间相关性的关系,认为选取 4 种参数构成数据库较为合适,能更多地保留空间信道的相关信息,即信号强度的均值、中值、最大值和最小值为特征参数。

由于室内多径信号传播对环境有很大的依赖性,某一位置上信道的多径结构理论上是唯一的,终端无线信号经过反射和折射传输,产生与周围环境密切相关的特定模式的多径信号,这种多径特征的信号可认为是某位置上的“信号纹”,因此选取网格化参考点 RP,构建了多径信号数据库,进行离线训练学习过程,用主分量分析法(PCA)降维和优化数据库。由此建立了信号特征参数与空间位置的内在对应关系,为后续比对分析提供保障,并以主分量数据进行存储,可以节省容量和提高运算效率。 $4 \times k$ 维的多径信号数据库结构如图 2 所示。

图2 $4 \times k$ 维的多径信号数据库示意图

2.2 定位算法

2.2.1 主分量分析法 PCA

首先,从宏观上,用主分量分析法(PCA)处理高维度数据库,在不损失主要信息的基础上,以低维度线性组合的数据进行运算,采用正交变换矩阵和拉格朗日乘子法^[7],根据估计坐标与参考坐标的均方差最小化原则,选取参考点 RP 对应的合适接入点 AP 和参数构成信号纹数据库。其次,从微观上,用模糊 C 均值聚类算法(FCM)处理待测点数据,通过设置相应的隶属度和相似性的阈值,进行数学筛选;将大于隶属度阈值及小于相似性阈值的奇异性数据识别与挖掘出来,滤除奇异性参考点 RP;在数学运算上,通过模糊分类矩阵实现目标

函数的最小化,以确保聚类的数据具有较好的相似性,以提高定位精度。

PCA 法具体步骤为:设原始高维度的具有一定相关性的数据为 $(x_1, x_2, \dots, x_{4k})$,重新线性组合成一组较低维度的互不相关的综合数据为 $(y_1, y_2, \dots, y_{k_0})$,其中 $k_0 < 4k$,则:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1,4k}x_{4k} \\ y_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2,4k}x_{4k} \\ \vdots \\ y_{k_0} = a_{k_0 1}x_1 + a_{k_0 2}x_2 + \dots + a_{k_0,4k}x_{4k} \end{cases} \quad (5)$$

由式(5)可知, y_1 是 x_1, x_2, \dots, x_{4k} 的所有线性组合中方差最大的第一主成分, y_2 与 y_1 不相关, y_2 是第二主成分,依此类推, y_{k_0} 与 $y_1, y_2, \dots, y_{k_0-1}$ 都不相关,是 k_0 个主成分。

关于选择主成分数量和特征值,是依据方差累计贡献率^[8] $C_k = \sum_{i=1}^{k_0} \lambda_i / \sum_{j=1}^{4k} \lambda_j$ 来确定的。当 C_k 大于85%时,可认为能够反映原始数据的主要信息,对应可提取前 k_0 个主成分和相应特征值 λ_i 。

2.2.2 定位基本步骤

本定位算法可分为离线训练阶段和在线定位阶段。离线训练阶段的主要任务是建立位置信号指纹数据库。首先,选择网格化参考点 RP 的位置,设置一定信号阈值,即某一个 RP 接收周边若干 AP 的信号,记录 4 个信号特征参数;通过主分量分析法降维与优化处理,根据均方差最小化^[10]以及方差累计贡献率大于 85% 的原则,建立低维度主分量线性组合的,与空间相关的数据库,能保留起主要作用的 AP 和参数,由此存储隐含位置信息的信号数据库称为位置指纹地图。具体步骤如图 3 的上面两个方框所示。

在线定位阶段是收集待测点在某一位置的信号特征参数,也是其收集周边若干 AP 的信号及 AP 的宏地址,由待测点标签再发回至定位服务器;通过模糊聚类 KNN 算法,将实测数据与预存数据进行对比分析,根据目标函数最小化原则,保留起主要作用的 RP,提取待测点周边的 RP 预存值,即确定聚类范围;计算待测点与参考点的距离,选取相似性好的 RP,即以距离为依据选取 RP,从而估计待测点的实际坐标。具体步骤如图 3 的下面 3 个方框所示。

3 定位实验

3.1 实验平台

为了评估本文定位算法的实际性能,设置的实验环境是:警院安防科技园 3 楼 5 间 100 m² 的展示区,隔墙材料为轻质石膏板,层高为 3.3 m。将 12 只定位 AP 均

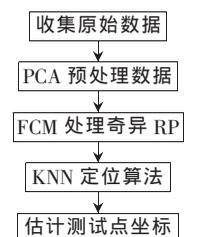


图3 定位算法方框

网络与通信 Network and Communication

匀排列,安装高度为 2.4 m,实现 Wi-Fi 无线信号全覆盖,定位主机设在第二展区内,用网线连接各 AP 至一台交换机,组成定位系统局域网。网格化参考点 RP 间隔为 2 m,呈方格排列^[11],待测点为双向有源标签。具体平面布局如图 4 所示。

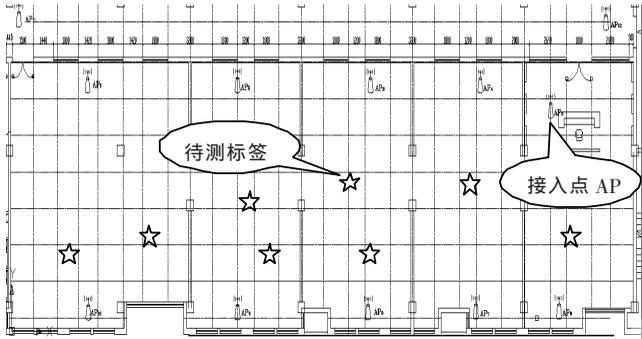


图 4 无线信号接入点 AP 布局平面图(图中五角星为双向标签)

实验采用定位服务器配置:CPU 为 4 核处理器,主频 2.4 GHz 以上,内存为 8 GB 以上,硬盘为 256 GB 以上。数据库服务器按以上标准另行配置。定位服务器软件运行环境为:操作系统为 Windows 2003/2008 Server 32 bit,数据库为 Microsoft SQL server 2005/2008,电子地图采用 JPG 格式。定位服务器软件实现与定位器(AP)和标签(Tag)之间的指令,以及相关数据的交互。根据标签发往 AP 的回传信号数据,由定位算法分析标签(测试点)与预存信息(参考点)的匹配关系,估算出标签的实际位置。AP 定位器主要指标为:2.4 GHz,IEEE802.11b/g,最高速率为 54 Mb/s,天线增益 2 dBi,同时扫描标签 128 个/s。胸卡式标签:2.4 GHz,速率为 1 Mb/s,双向通信,最大发射功率为 20 dBm,发射间隔为 1 s,48 bit 唯一 ID 号。

3.2 实验结果

在定位实验区,当处于离线训练学习阶段时,在地面打好 256 个网格化参考点位(即网格的交叉点上),用胸卡式标签采集及回传无线信号,由主机定位服务器进行处理与存储。当在线定位阶段,采用人员配带胸卡标签的方式进行测试,具体人数为 25 人,胸卡式标签为 200 只。为了评估本文算法与 KNN 算法性能的优劣,收集标签数量与平均定位误差比较图。采样点数据是通过电子地图上显示待测点位置与实际标签位置的比较计算得到的,每个点位采样为 20 次,取平均值,获得平均定位误差数据。在采样过程中,忽略了因人员走动时信号漂移等现象引起的明显误差。模糊聚类算法(改进算法)与 KNN 算法比较如图 5 所示,可知改进算法的定位精度有比较明显的改善,大约提高了 5% 左右。

在 KNN 定位算法的基础上,通过本算法可以有效地克服 KNN 运算中丢失位置信息的不足,从而提高定位

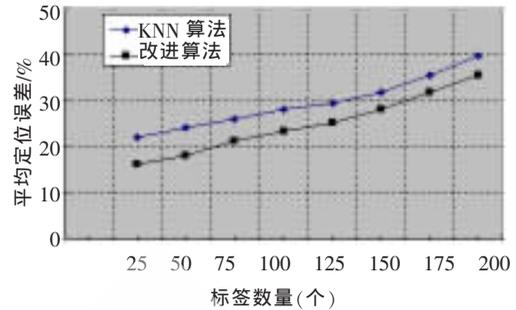


图 5 定位算法性能比较

算法的定位精度。实验表明,当室内人员较多且人员快速移动时,还会出现无线信号的漂移和时延显示等现象;这一类信号传输问题有待于今后进一步优化定位算法,实现能自适应物理环境变化的定位算法。

参考文献

- [1] 唐文胜,李姝,匡旺秋.RF 室内定位指纹库空间相关生成算法[J].计算机工程与应用,2008,44(23):226-229.
- [2] 卢恒惠,刘兴川,张超,等.基于三角形与位置指纹识别算法的 WiFi 定位比较[J].移动通信,2010(10):72-76.
- [3] 汤丽,徐玉滨,周牧,等.基于 K 近邻算法的 WLAN 室内定位技术研究[J].计算机科学,2009,36(4B):54-55,92.
- [4] 刘兴川,林孝康.基于聚类的快速 Wi-Fi 定位算法[J].计算机工程,2011,37(8):285-287.
- [5] 李文杰,李文明.基于 K-近邻算法的定位方法设计和仿真[J].计算机仿真,2009,26(4):194-196.
- [6] 潘玉奇,周劲,杨秀丽.基于模糊聚类分析的数据检索的应用[J].微电子学与计算机,2005,22(6):167-172.
- [7] Zhou Mu, Xu Yubin, Ma Lin. Radio-map establishment based on Fuzzy clustering for WLAN hybrid KNN/ANN indoor positioning[J]. China Communications,2010(7):64-79.
- [8] Yang Qiang, PAN S J, ZHENG V W. Estimating location using Wi-Fi[J]. IEEE Intelligent Systems, 2008,23(1):8-13.
- [9] 戴立伟,李向阳,程赟.无线传感器网络的 RSSI 定位技术研究[J].计算机工程与设计,2009,30(19):4395-4397.
- [10] TRAN Q, TANTRA J W, FOH C H, et al. Wireless indoor positioning system with enhanced nearest neighbors in signal space algorithm[C]. IEEE 64th Vehicular Technology Conference,2006:1-5.
- [11] 李昊.位置指纹定位技术[J].山西电子技术,2007(5):84-87.

(收稿日期:2012-07-14)

作者简介:

都伊林,男,1961 年生,本科,高级实验师,主要研究方向:电子电路与通信技术。