

基于相似度的 K 阶临近定位算法^①

马文丽, 李世宝, 张志刚, 杨喜鹏, 王升志, 张 鑫

(中国石油大学(华东) 计算机与通信工程学院, 青岛 266580)

摘 要: 基于 WIFI 位置指纹的定位系统能实现较高精度的室内定位, 其中基于接收信号强度指示(RSSI)的近邻选择算法在进行室内定位时容易引入奇异点, 导致定位精度降低. 针对该问题, 本文提出了一种基于相似度的 K 阶临近定位算法(SKNN). 该算法借鉴二部分网络中求解节点相似性的思想, 建立位置指纹与 AP 之间的二部分网络, 并提出一个相似度参数, 用该参数去修正 K 阶临近定位算法. 实验结果表明, 本文提出的 SKNN 算法可以有效的降低奇异点对定位结果的影响, 提高定位精度, 80% 的定位误差均在 2 m 以内, 且在大场景中效果明显.

关键词: 室内定位; 位置指纹; 近邻选择算法; 二部分网络; 相似度

引用格式: 马文丽, 李世宝, 张志刚, 杨喜鹏, 王升志, 张鑫. 基于相似度的 K 阶临近定位算法. 计算机系统应用, 2017, 26(9): 165-169. <http://www.c-s-a.org.cn/1003-3254/5982.html>

Similarity-Based K-Nearest Neighborhood Location Algorithm

MA Wen-Li, LI Shi-Bao, ZHANG Zhi-Gang, YANG Xi-Peng, WANG Sheng-Zhi, ZHANG Xin

(Department of Computer and Communication Engineering, China University of Petroleum (East China), Qingdao 266580, China)

Abstract: The positioning system based on WIFI location fingerprint can achieve high precision indoor location. The neighbor selection algorithm based on Received Signal Strength Indicator (RSSI) is easy to introduce singular points when locating indoors, which leads to the decrease of positioning accuracy. To solve this problem, this paper proposes a Similarity-based K-Nearest Neighborhood Location Algorithm (SKNN). Referring to the idea used to solve the problem of similarity of nodes in bipartite networks, this algorithm builds a bipartite network between the location fingerprint and the AP. It proposes a similarity parameter which can be used to modify the K-Nearest Neighborhood localization algorithm. The experimental results show that the SKNN algorithm proposed in this paper can effectively reduce the influence of singular points on the positioning results and improve the positioning accuracy, with 80% of the positioning errors within 2m, and the effect is obvious in the large scene.

Key words: indoor localization; location fingerprint; nearest neighbor selection algorithm; bipartite network; similarity

随着情景感知、环境智能等应用需求的增加, 人们对用户位置信息精度的要求也在不断提高. 室外环境中, 全球定位系统(Global Positioning System, GPS)、网络辅助全球卫星定位系统(Assisted Global Positioning System, A-GPS)和蜂窝网定位系统等可以满足绝大部分的定位需求. 但是, 在室内环境中, 由于建筑物的遮挡, 使得 GPS 等已有的定位系统不再适用. 然而, 在

现实生活中, 人们对室内定位服务需求越来越强烈, 例如在商场导购、博物馆导航、地下车库导航、导盲等方面, 均开始热切关注室内定位的问题, 因此, 研究一种适用于室内环境的定位系统是很有必要的.

基于 WIFI 信号的室内定位方法^[1]因其成本低廉、硬件易于实现、精度高、定位过程简单成为研究的热点. 按照定位原理可以划分为基于测距^[2]的定位方

① 基金项目: 中国自然科学基金青年基金(61402433); 山东省自然科学基金面向项目(ZR2014FM017); 中央高校基本科研业务费专项资金(15CX05025A)
收稿时间: 2016-12-29; 采用时间: 2017-02-13

法和基于指纹的定位方法. 室内环境中无线信号的 NLOS 现象导致基于测距的定位方法在进行估算距离时易产生误差. 基于指纹的定位方法利用了 WIFI 信号空间位置差异性进行定位, 有效的减小 NLOS 现象对定位结果的影响.

基于接收信号强度^[3](Received signal Strength Indicator, RSSI)的指纹定位技术因其硬件成本和计算开销低而被广泛使用, 常用的匹配算法有概率法、神经网络法、支持向量机法和近邻选择算法. 其中概率法、神经网络法和支持向量机法的算法复杂难以实现, 并且计算量大, 不能满足用户对实时性的要求. 因此本文在后面章节将主要针对近邻选择算法进行研究和改进.

1 相关研究

近邻选择算法^[4]由于计算简单、易于实现而得到广泛应用, 其核心在于通过计算欧氏距离寻找与待测点距离最近的一个或多个采样点. 常见的近邻选择算法有最近邻法(Nearest Neighborhood, NN), K 阶临近法(K-Nearest Neighborhood, KNN), 加权 K 阶临近法(Weighted K-Nearest Neighborhood, WKNN), 聚类过滤法(Cluster Filtered KNN, CFK)等.

NN 算法是所有基于 RSSI 距离近邻选择算法中最简单的一种. 将移动端采集到的 RSSI 向量与指纹数据库中所有指纹点进行匹配, 计算欧氏距离 D_{ist} , 距离最小的点对应的位置信息 (x_i, y_i) 即为对待测点的位置估计. RADAR 系统^[5]使用 NN 算法进行指纹的匹配与计算, 实现了 2~5 米的定位精度. KNN 算法^[6]是对 NN 算法的改进, 用 NN 算法中计算距离的方法计算得到具有最小距离的个指纹点 $(x_i, y_i), i \in (1, K)$, 那么待测点的位置估计就是这 K 个指纹点位置的质心 (\bar{x}, \bar{y}) , 有效解决偶然性导致较大定位误差的问题. 为了提高定位精度, 文献^[7]提出 WKNN 算法, 在选出 K 个相近的指纹后, 不是直接计算 K 个点的质心, 而是先根据每个指纹点的 RSSI 值计算它对待测点的贡献度, 将贡献度作为权值分配给各个指纹点, 然后求 K 个指纹点的加权质心, 质心的位置即为最终的位置估计, 能有效提高定位精度. 由于指纹数据库中存在着 RSSI 向量相似但实际位置相距较远的指纹点, 将与待测点的 RSSI 向量值相似但实际位置相距较远的指纹点称为奇异点, 上述近邻选择算法在进行指纹匹配时容易引入奇异点, 从而导致定位精度降低.

文献^[8]针对指纹数据库中奇异点, 在求解信号欧氏距离时进行加权处理. Jun MA 等人提出聚类过滤 KNN 算法^[9], 该算法针对利用聚类算法滤除奇异点. 王跃等人提出一种基于模拟退火聚类的室内定位算法^[10]. 该算法采用模拟退火聚类的方法消除了具有一定特征相似性的奇异点. 文献^[11]提出采用模糊 K-均值聚类的方法对 KNN 方法进行改进, 利用 K-均值聚类奇异点进行筛选.

通过聚类算法滤除奇异点的方法时间开销是非常大的, 在具体应用上不能满足用户对实时性的要求. 聚类算法也不能保证达到最优的聚类效果, 两种采用 K-均值聚类的算法在聚类的时结果受初始值的影响大, 并且容易出现局部最优值的问题. 虽然一定程度上减小了奇异点的影响, 但并没有提高匹配定位的效率. 针对上述问题, 本文借鉴二部分网络中判断节点之间相似度的思想, 将指纹点和 AP 建成二部分网络, 考虑指纹点之间共同连接的 AP 数目对定位的影响, 提出基于相似度的 KNN 算法(Similarity-based K-Nearest Neighborhood, SKNN).

2 SKNN 算法

针对奇异点影响近邻选择算法定位精度的问题, 本文对原有的阶临近定位算法进行改进, 提出了 SKNN 算法. 在算法设计过程中, 需要解决的难点问题是在分析奇异点影响定位精度的根本原因、建立指纹点和 AP 的二部分网络和相似度的计算. 针对以上问题, 将该算法的设计分为以下四个步骤: 网络的建立、相似度的计算、基于相似度的加权欧氏距离计算和真实位置估计. 算法流程图如图 1 所示.

2.1 网络的建立

二部分网络是一种简单无向网络 $G(V, E)$, 其中 V 代表节点的集合, E 代表边的集合, 至少存在一对节点集合 X 和 Y , 并满足 $X \cap Y = \emptyset$; $X \cup Y = V$; E 中的任意边一定恰有一个顶点在集合 X 中, 另一个顶点在 Y 中. 如图 2 “用户-商品”二部分网络.

鉴于上述思想, 本文将指纹匹配的问题用二部分网络中求解节点相似性的思想解决, 首先将指纹数据库中的每个指纹点位置作为 X 集合, AP 点的集合作为 Y 集合, 指纹点 X_i 中收到 AP_j 的信号即为 X_i 与 Y_j 之间产生一条连边 E_{ij} . 将实际网络表示成二部分网络, 如图 3, 用不同指纹点之间共同连接的 AP 的数目来衡量两个指纹点之间的相似性. 本文引入一个“邻

居”的概念: 在“AP-指纹”网络中, 与指纹点有连边的 AP 叫做该指纹点的邻居节点, 如果两个指纹节点都跟同一个 AP 相连接, 则将这个 AP 称为两个指纹节点的共同邻居。

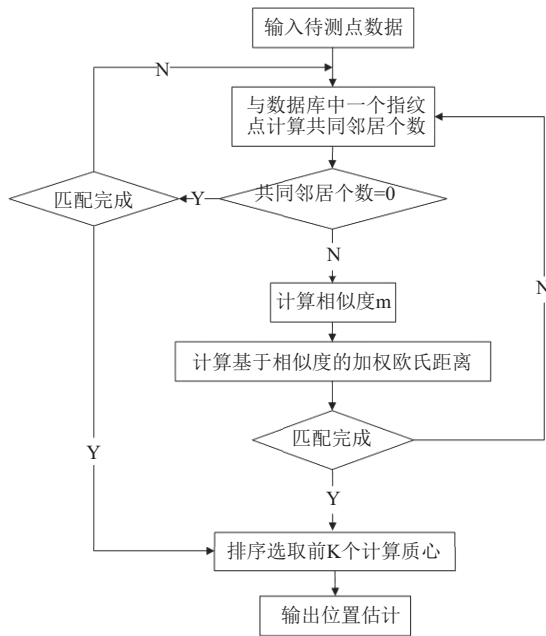


图1 SKNN 算法流程图

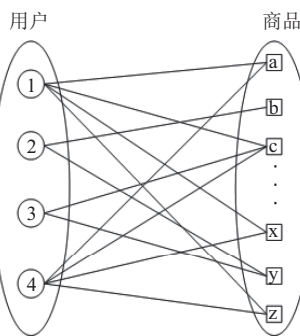
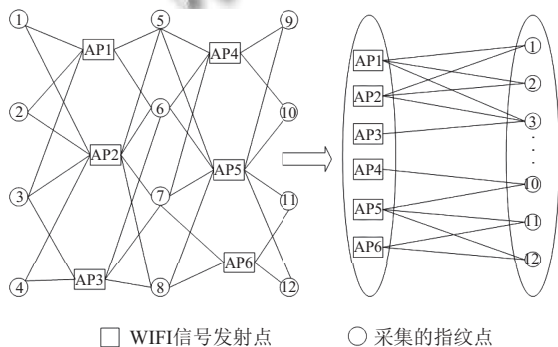


图2 “用户-商品”二部分图



□ WIFI信号发射点 ○ 采集的指纹点

图3 “AP-指纹”网络

2.2 相似度的计算

在理想的定位环境中, 位置越接近的点 RSSI 向量值就越相似. 但在实际室内环境中, 信号强度不一定完全由物理位置的远近造成, 也可能由信号强度自身的波动或反射折射等因素造成, 通过欧式距离筛选出来的点并不一定全都是距离待测点实际位置接近的点, 通常也会包含 RSSI 值相似但实际位置却相距很远的指纹点, 如图 4 所示, 待测点为 Z 点, 运用 KNN 算法定位时, 根据欧式距离的大小筛选出 A、B、C 三个点, 则三个点的质心 M 即为最终的位置估计. 指纹数据库中的 D 点与 C 点的 RSSI 值相似但与 C 的实际位置相距较远, 且有比 C 点更小的欧氏距离, 运用 KNN(K=3) 算法定位时, D 点就作为一个奇异点被引入, A、B、D 三点的质心 N 为最终定位结果, 比较 M 和 N 的位置可知: 奇异点的引入导致了较大的定位误差。

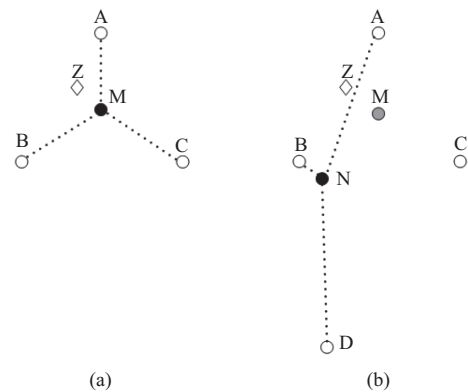


图4 奇异点示意图

在“用户-商品”二部分网络中, 如果两个用户的共同邻居越多, 这两个用户的相似性越高, 基于这种相似性为用户推荐更多的有用信息. 同理, 基于 RSSI 的指纹匹配算法的目的是在指纹数据库中找到与待测点相似的 K 个指纹点. 对“AP-指纹”网络进行分析可知: 在欧氏距离相同的情况下, 实际距离越相近的两个指纹点, 它们的共同邻居节点数占邻居节点总数的比例越高. 本文定义了一个“相似度”参数:

$$m_{xy} = \frac{|\tau_x \cup \tau_y|}{|\tau_x \cap \tau_y|} = \frac{S_\beta}{S_\alpha} \quad (1)$$

即邻居节点总数与共同邻居节点数的比值, 用 X 表示待测点, Y 表示指纹数据库中的指纹点. τ_x 、 τ_y 分别表示与 X、Y 相连的 AP, S_α 表示 X 与 Y 共同邻居数目, S_β 表示 X 与 Y 各自邻居节点的总和. 式(1)中 m 是

两个集合节点数目的比值,表示网络中的 X 与 Y 节点连接的 AP 的匹配程度. $m \in (1, \infty)$, m 的值越接近于 1, 证明两个节点的相似度越高; 反之 m 的值越大, 节点相似度越低.

引入相似度之后可以有效的减小奇异点对最终定位结果的影响. 在奇异点与临近点有相同欧氏距离的情况下, 奇异点会乘以一个远远大于 1 的相似度, 会将原先得到的欧氏距离放大很多倍; 临近点则会乘以一个接近于 1 的相似度, 欧氏距离基本不变, 由于算法需要将欧式距离最小的 K 个点的质心作为最终的位置估计, 所以相似度的引入会大幅降低奇异点对定位结果的影响.

2.3 基于相似度的加权欧式距离

定义当前时刻移动终端在待测点接收到的各个 AP 的 RSSI 向量为 R_t , $R_t = (r_1, r_2, r_3 \dots r_n)$, r_n 是移动终端接收到的 AP _{n} 的 RSSI 值; 指纹数据库中的向量 $F_i = (f_{i1}, f_{i2}, f_{i3} \dots f_{in})$ 进行匹配, 其中 f_{in} 表示在第 i 个参考位置处接收到的 AP _{n} 的指纹信息. 则当前待测点与指纹数据库中第 i 个指纹点的欧氏距离表示为公式(2):

$$D_{\text{ist}}(R_t, F_i) = R_t - F_i^2 = \sum_{j=1}^n (r_j - f_{ij})^2 \quad (2)$$

$$D'_{\text{ist}}(R_t, F_i) = m \sum_{j=1}^n (r_j - f_{ij})^2 \quad (3)$$

$$D'_{\text{ist}}(R_t, F_i) = \frac{|\tau_x \cup \tau_y|}{|\tau_x \cap \tau_y|} \sum_{j=1}^n (r_j - f_{ij})^2 \quad (4)$$

基于相似度的加权欧氏距离需要在原 $D_{\text{ist}}(R_t, F_i)$ 基础上乘相似度 m , 即公式(4). 然后依次计算当前测量向量 R_t 与指纹数据库中的每个参考向量 F_i 的距离.

2.4 位置估计

将上一步得到的距离 $D'_{\text{ist}}(R_t, F_i)$ 进行排序, 选出距离最小的前 K ($K=1, 2 \dots n$) 个指纹点, 其坐标分别为 (x_i, y_i) ($i=1, 2 \dots K$), 用公式(5)计算 K 个参考位置的质心, 质心就是最终的位置估计.

$$(\hat{x}, \hat{y}) = \frac{1}{k} \sum_{i=1}^k (x_i, y_i) \quad (5)$$

3 实验及结果分析

为验证 SKNN 算法的性能, 在一个大型商场进行定位实验, 商场的平面图如图 5 所示. 利用各个商家已

经部署的 AP(在整个商场中共搜索到 83 个可用的 AP), 移动采集终端选择 OPPOr9 手机.

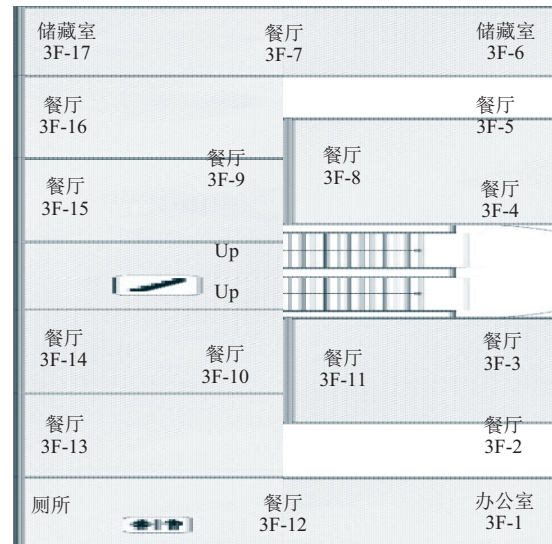


图 5 商场平面图

3.1 离线阶段

离线阶段即指纹采集阶段. 指纹采集时, 将图 5 所示的整个区域划分成多个 $1 \text{ m} \times 1 \text{ m}$ 的网格, 每个网格采集一个指纹样本, 共采集 900 个指纹数据. 为了减小 RSSI 时变带来的影响, 在各个网格多次采集各个 AP 的信号强度. 通过大量实测数据分析, 如图 6 所示在确定的网格内测得特定 AP 的 RSSI 值会在 56 dB 附近波动, 所以本文离线阶段采用高斯滤波技术, 舍弃概率比较低的 RSSI 值, 对概率高的 RSSI 值求平均, 从而降低信号的随机误差.

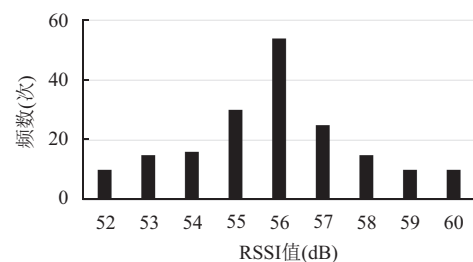


图 6 某一 AP 的 RSSI 值分布图

3.2 在线定位阶段及结果分析

在线定位阶段, 用户手持移动终端在定位区域进行定位, 采集当前位置各个接入点的 RSSI 值, 组成测试集, 再将测试集与指纹数据库中的训练集进行搜索匹配, 匹配算法采用经典 KNN 算法和 SKNN 算法.

由于本文提出的 SKNN 算法适用于较大区域,为了验证其在不同大小的区域中的效果,在同一个实验场景中选择三个大小不同的区域进行了三次实验,三个区域为 map1、map2 和 map3,面积分别为 40 m^2 、 80 m^2 和 200 m^2 。实验误差及分布率示意图如图 8 所示,SKNN 算法大场景中的定位效果优于小场景。

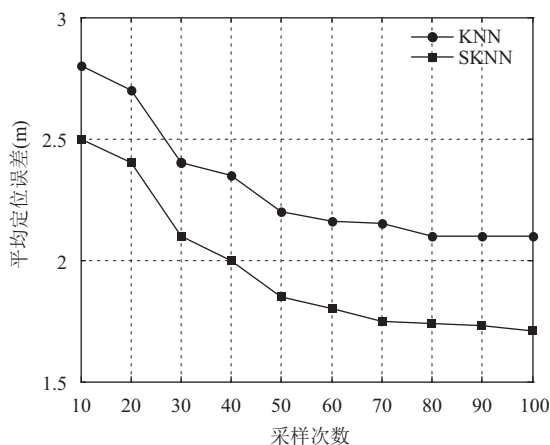


图7 KNN 和 SKNN 算法对比图

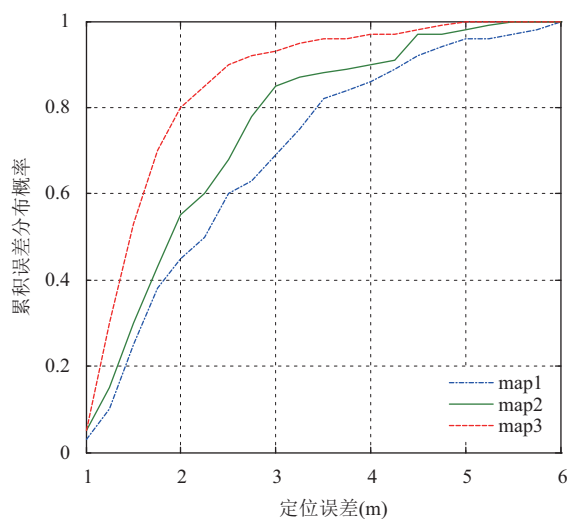


图8 三种区域下误差分布率

SKNN 算法与 KNN 算法定位误差比较如表 1 所示,从表 1 可以看出,用 SKNN 算法计算得到的坐标误差要小于 KNN 算法。

表1 定位误差比较

| 定位算法 | 最大误差 (m) | 平均误差 (m) |
|---------|----------|----------|
| KNN 算法 | >6 | 2.3 |
| SKNN 算法 | 5 | 1.8 |

4 结束语

本文针对基于 RSSI 的近邻选择算法在指纹匹配时引入奇异点导致定位精度降低的问题,建立指纹点与 AP 之间的二部分网络,基于二部分网络中求解节点相似性的思想提出相似性参数,用该参数修正经典 KNN 算法.该算法可以有效的减弱奇异点对定位结果的影响,减小定位误差,且在大场景中效果更明显。

参考文献

- Chintalapudi K, Padmanabha Iyer A, Padmanabhan VN. Indoor localization without the pain. Proc. of the 16th Annual International Conference on Mobile Computing and Networking. Chicago, Illinois, USA. 2010. 173-184.
- 席瑞,李玉军,侯孟书.室内定位方法综述.计算机科学,2016,43(4): 1-6, 32.
- 曹世华.室内定位技术和系统的研究进展.计算机系统应用,2013,22(9): 1-5.
- 邓志安.基于学习算法的 WLAN 室内定位技术研究[博士学位论文].黑龙江:哈尔滨工业大学,2012: 10-36.
- Bahl P, Padmanabhan VN. RADAR: An in-building RF-based user location and tracking system. Proc. of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies. Tel Aviv, Israel. 2000, 2: 775-784.
- Sun YX, Liu M, Meng MQH. WiFi signal strength-based robot indoor localization. Proc. of the 2014 IEEE International Conference on Information and Automation (ICIA). Hailar, China. 2014. 250-256.
- Brunato M, Battiti R. Statistical learning theory for location fingerprinting in wireless LANs. Computer Networks, 2005, 47(6): 825-845. [doi: 10.1016/j.comnet.2004.09.004]
- 蔡朝晖,夏溪,胡波,等.室内信号强度指纹定位算法改进.计算机科学,2014,41(11): 178-181. [doi: 10.11896/j.issn.1002-137X.2014.11.035]
- Ma J, Li XS, Tao XP, et al. Cluster filtered KNN: A WLAN-based indoor positioning scheme. Proc. of the 2008 International Symposium on a World of Wireless, Mobile and Multimedia Networks. Newport Beach, CA, USA. 2008. 1-8.
- 王跃,崔维嘉,王大鸣,等.基于模拟退火聚类的室内定位算法.信息工程大学学报,2016,17(2): 161-164.
- 都伊林.一种模糊聚类 KNN 位置指纹定位算法.微型机与应用,2012,31(23): 55-58. [doi: 10.3969/j.issn.1674-7720.2012.23.017]