

基于余弦相似度的加权 K 近邻室内定位算法

黄运稳 陈光 叶建芳

(东华大学信息科学与技术学院 上海 201600)

摘要 基于位置指纹的室内定位系统能够实现较高精度的定位,其中 KNN(K-nearest neighbor)和余弦相似度定位算法原理简单且易于实现。但每种算法仅从单一限制条件进行匹配,导致定位精度不高。针对此问题,提出基于余弦相似度的加权 KNN 算法,并通过实验测试算法的定位性能。实测结果表明,该算法的定位精度高于传统定位算法。当 AP 数量为 5 时,平均误差为 1.67 m。定位精度优于 1 m 的置信概率为 42%,优于 2 m 的置信概率为 88%,最大定位误差为 4.3 m。

关键词 室内定位 Wi-Fi 指纹 奇异点 K 近邻法 余弦相似度

中图分类号 TP3 **文献标识码** A **DOI**:10.3969/j.issn.1000-386x.2019.02.029

WEIGHTED K NEAREST NEIGHBOR INDOOR LOCATION ALGORITHM BASED ON COSINE SIMILARITY

Huang Yunwen Chen Guang Ye Jianfang

(College of Information Science and Technology, Donghua University, Shanghai 201600, China)

Abstract Indoor positioning system based on position fingerprint can achieve positioning with higher precision. KNN and cosine similarity positioning algorithms are simple and easy to implement, but each algorithm only matches from a single constraint condition, resulting in low positioning accuracy. To solve this problem, this paper proposed a weighted KNN algorithm based on cosine similarity and tested the positioning performance of the algorithm through experiments. The results show that the positioning accuracy of this algorithm is higher than that of the traditional positioning algorithm. When the number of APs is 5, the average error is 1.67 m. The confidence probability of positioning accuracy better than 1m is 42%, and the confidence probability better than 2m is 88%. The maximum positioning error is 4.3 m.

Keywords Indoor positioning Wi-Fi fingerprint Singular point K nearest neighbor Cosine similarity

0 引言

随着社会的进步与发展,人们对室内位置服务 LBS(Location Based Services)的需求日益强烈^[1]。将全球定位系统 GPS 应用于室内定位时,由于卫星信号受建筑物环境的影响很大,定位的效率较低,定位的精度也较差,因此不能达到实时和准确的定位要求。

随着无线通信技术的应用发展,基于接收信号强度的 Wi-Fi 室内定位成为目前研究的热点^[2-3]。Wi-Fi

信号不受视距传播的影响,信号的覆盖范围较大,而且不易受到噪声的干扰,适合于复杂的室内环境定位。尤其是,Wi-Fi 指纹定位不用添加其他任何硬件,利用现有的 WLAN,通过软件编程就可以在移动智能终端上实现定位^[4-6]。

有鉴于此,国内外学者近年来对 Wi-Fi 指纹定位的匹配算法做了大量深入的研究,其中 KNN^[7]、WKNN^[8-10](K Weighted Nearest Neighbor)以及余弦相似度^[11-12]算法由于计算简单、易于实现而得到广泛应用。以上算法的核心在于通过 RSS 进行欧氏距离或相似度的匹配。然而一方面,由于接收信号强度自身

的不稳定性与环境的多变性,导致接收信号强度不能完全准确反映客观物理位置。另一方面,欧氏距离体现的是接收信号强度数值的绝对差异,而余弦相似度是从方向上区分接收信号强度的差异,以上因素导致各算法在定位过程中容易引入奇异点^[14]。针对上述问题,本文对 K 最近邻和余弦相似度的组合策略进行了分析研究与实验比较,给出了定位精度更高的优化组合算法。

1 Wi-Fi 指纹定位匹配算法

1.1 KNN 匹配算法

KNN 算法通过测量两个向量之间欧式距离来度量它们之间的相似度。该算法将待定位点采集到的信号强度 RSS 向量 $[s_1, s_2, \dots, s_n]$ 与指纹数据库中的信号强度 RSS 均值矩阵 $[S_{i1}, S_{i2}, \dots, S_{in}]$ 相匹配。设定位区域有 m 个参考点,共有 n 个 AP, s_j 为待测点收到第 j 个 AP 的信号强度, S_{ij} 为指纹库中第 i 个参考点采集到第 j 个 AP 的 RSS 均值, $i = 1, 2, \dots, m, j = 1, 2, 3, \dots, n$ 。距离的定义公式如下:

$$L_i = \sqrt{\sum_{j=1}^n (s_j - S_{ij})^2} \quad (1)$$

KNN 算法在欧氏距离中从小到大依次选取 K 个参考点,然后以该 K 个参考点的质心作为估计位置。

$$(\hat{x}, \hat{y}) = \frac{1}{K} \sum_{i=1}^K (x_i, y_i) \quad (2)$$

KNN 定位算法原理相对简单,以信号强度来反映物理位置关系,有利于定位系统的实现。

1.2 余弦相似度匹配算法

余弦相似度通过测量两个向量内积空间夹角的余弦值来度量它们之间的相似性。余弦值越接近 1,表示两个向量的夹角越接近 0 度,向量的方向越相近,即表示两个向量越相似。如 A 和 B 的余弦相似度的计算公式如下:

$$\cos\theta = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (3)$$

基于余弦相似度算法理论,定位点采集的 RSS 矩阵 $s = [s_1, s_2, \dots, s_n]$ 与指纹数据库中第 i 个参考点的 RSS 均值矩阵 $S_i = [S_{i1}, S_{i2}, \dots, S_{in}]$ 相匹配,通过式(4)计算余弦值,以余弦值从大到小依次选取 K 个参考点,并以式(2)计算得出估计位置。

$$\cos\theta_i = \frac{\sum_{j=1}^n s_j \times S_{ij}}{\sqrt{\sum_{j=1}^n (s_j)^2} \times \sqrt{\sum_{j=1}^n (S_{ij})^2}} \quad (4)$$

通过余弦相似度算法匹配,考虑了 RSS 向量的内在联系。余弦相似度使用两个向量夹角的余弦值作为衡量两个向量间差异的大小。相比欧氏距离,余弦相似度更加注重两个向量在方向上的差异。

2 基于余弦相似度的加权 KNN 算法

2.1 算法提出依据

针对以上对两种定位算法的介绍,通过三维坐标来进一步分析欧式距离和余弦相似度的区别,如图 1 所示。

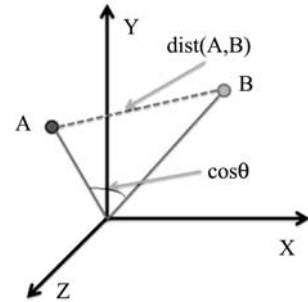


图 1 KNN 与余弦相似度区别

由图 1 可以看出,欧氏距离衡量的是空间各点的绝对距离,即与 RSS 向量各分量的大小直接相关;而余弦相似度衡量的是空间向量的夹角,体现的是 RSS 向量在方向上的差异。KNN 和余弦相似度算法各自有不同的计算方式和衡量特征,KNN 算法能够体现 RSS 向量的绝对差异,余弦相似度是从方向上区分差异,而对绝对的数值不敏感。

针对以上分析,本文提出了基于 KNN 和余弦相似度的组合算法,旨在弥补单一区分方式的不足,从而提高定位精度。

2.2 算法步骤

在 KNN 算法筛选出 K 个近邻点的条件下,计算 K 个近邻点的余弦值。由 1.2 节分析可知,余弦值越大,表示两个向量越相似,在欧氏距离相近的前提下,余弦值大的近邻点定位精度越高。本文采用线下加权的方式来确定权值,从而包含位置指纹全部数据信息,进而提升定位精度。算法步骤如下:

1) 筛选近邻点。基于式(1)计算出定位点 RSS 向量与指纹库中各向量的欧氏距离 L_i ,在 L_i 中从小到大顺次筛选出 K 个参考点,坐标为 $(x_i, y_i), i = 1, 2,$

..., K 。

2) 计算余弦值。基于式(4)分别计算待定位点与 K 个参考点之间的余弦值,余弦值的大小分别为 S_1, S_2, \dots, S_K 。

3) 确定权重。 S_1, S_2, \dots, S_K 中,余弦值越大,在位置估算时所作贡献越大,那么第 i 个参考点的权重值 ω_i 可表示为:

$$\omega_i = \frac{S_i}{\sum_{j=1}^K S_j} \quad (5)$$

4) 位置估计。利用权重值对 K 个参考点估算定位点的位置坐标,即:

$$(\bar{x}, \bar{y}) = \sum_{i=1}^K \omega_i (x_i, y_i) \quad (6)$$

3 实验分析

3.1 实验环境与系统

为了检验组合算法的定位性能,进行如下实验。实验环境为东华大学2号学院楼第六层,平面图如图2所示。

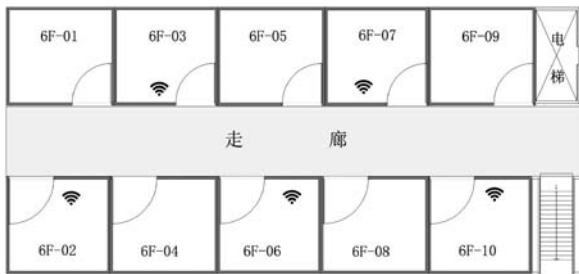


图2 实验环境平面图

文献[15]得出结论,当 $K=5$ 或 $K=6$ 时,系统定位的效果最佳。因此本文 K 值选择5对参考点坐标进行筛选。其中数据库信号采集与定位软件界面如图3所示。



(a) 离线信号采集 (b) 在线定位

图3 软件界面截图

离线采集:选用魅族 M578C 建立离线指纹数据库,并将数据库信息存储在服务器。离线采样间隔为 1 m ,为了减少数据库建立的误差,每个离线采样点采集20组数据,以20组信号强度均值作为采样点的最终样本值。其中,数据库中的数据包括采样点坐标、扫描到每个AP的ID和接收到每个AP信号强度。

在线定位:选取30个定位点测试,分别包括每个实验室选取2个参考点,走廊选取10个参考点。移动终端向服务器发送连接请求,并向服务器发送当前RSS向量,服务器接受RSS向量后通过组合算法进行匹配,估算定位点的坐标,并将坐标信息发送到移动终端。

3.2 性能分析

图4为走廊中间位置的定位点连续进行30次RSS采集,参与定位AP数量为5个时RSS的变化。由图可知,室内环境复杂多变,引起AP信号强度起伏变化。因此单一的传统算法极易造成匹配出现偏差,引入奇异点。

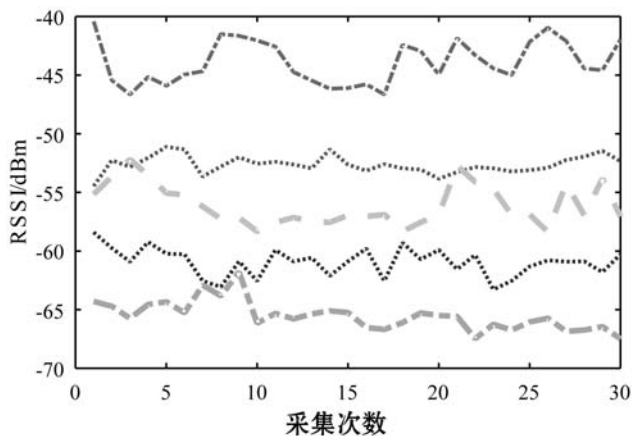


图4 不同定位点信号强度变化曲线

首先在实验环境中选2个AP参与在线定位阶段,分别是6F-03和6F-07中的AP,采集30个参考点接收到参与定位的2个AP的数据,与数据库中的数据相匹配,不参与定位的AP不匹配,计算此条件下,组合算法及传统算法定位的平均误差。然后每次增加一个AP参与定位,以相同方式算出不同算法的定位平均误差。实验中第六个AP是校园公共网络。不同算法在AP个数不同情况下,定位平均误差的结果如图5所示。

图5结果表明,当不同个数的AP参与定位时,组合算法的平均误差均低于传统算法,由此排除了组合算法在特定AP数量下提升定位性能的可能性。

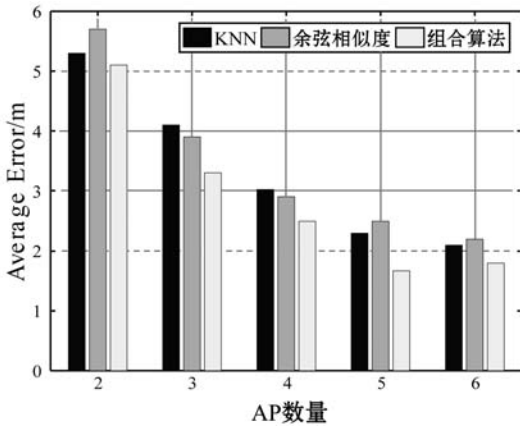


图5 参与定位的 AP 个数对定位结果影响

根据性能分析中不同 AP 个数对定位精度影响,我们选用 5 个 AP 参与定位,进一步比较不同算法的累加误差概率,如图 6 所示。

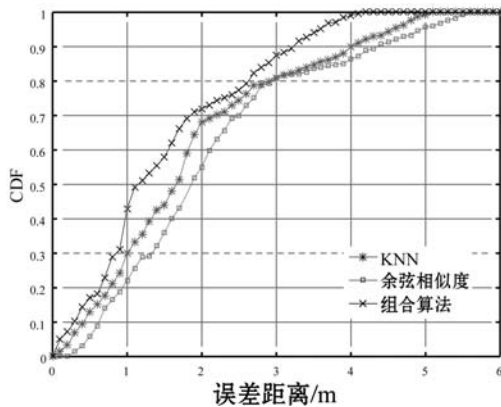


图6 不同算法的 CDF 曲线

由图 6 可知,本文提出的组合算法定位精度明显优于单一的传统算法。组合算法定位精度优于 1 m 的置信概率为 42%,优于 2 m 的置信概率为 72%,最大误差为 4.3 m,相对传统算法均有改善。实验结果表明,文中提出的基于余弦相似度的加权 KNN 算法能够有效地提高室内定位的精度。

4 结 语

为了提高室内定位的精度,考虑到 KNN 和余弦相似度匹配算法在定位特点的互补性,提出了基于余弦相似度的加权 KNN 算法,从而弥补单一传统算法的不足。实验结果表明,基于余弦相似度的加权 KNN 算法在 5 个 AP 参与定位时,平均误差减少到 1.67 m,定位精度明显提高。

随着室内定位技术的发展,基于余弦相似度的加权 KNN 算法需要在更大面积的环境、更复杂的干扰因素、更大量的指纹样本条件下进一步测试和优化,使得该算法更加完善。

参 考 文 献

- [1] 林志, 茆云霞. 一种基于 WiFi 信号指纹的室内定位技术[J]. 信息通信, 2017(7): 153-156.
- [2] 席瑞, 李玉军, 侯孟书. 室内定位方法综述[J]. 计算机科学, 2016, 43(4): 1-6.
- [3] 姜莉. 基于 WiFi 室内定位关键技术的研究[D]. 大连: 大连理工大学, 2010.
- [4] Liu H, Gan Y, Yang J, et al. Push the limit of WiFi based localization for smartphones [C]//International Conference on Mobile Computing and NETWORKING. ACM, 2012: 305-316.
- [5] Helen M. Using calibration in RSSI-based location tracking system [C]//Proceedings of the 5th World Multiconference on Circuits, Systems, Communications & Computers, 2001.
- [6] Shu Y, Huang Y, Zhang J, et al. Gradient-based fingerprinting for indoor localization and tracking [J]. IEEE Transactions on Industrial Electronics, 2016, 63(4): 2424-2433.
- [7] Ma J, Li X, Tao X, et al. Cluster filtered KNN: A WLAN-based indoor positioning scheme [C]//Proceedings of the 2008 International Symposium on a World of Wireless, Mobile and Multimedia Networks. IEEE Computer Society, 2008: 1-8.
- [8] 于洋, 陈光, 单志勇, 等. 基于中心矩改进的 WiFi 室内定位算法 [J]. 计算机应用与软件, 2018, 35(2): 130-133, 160.
- [9] 毕京学, 甄杰, 汪云甲, 等. 高斯函数定权的改进 KNN 室内定位方法 [J]. 测绘通报, 2017(6): 9-12.
- [10] 马文丽, 李世宝, 张志刚, 等. 基于相似度的 K 阶临近定位算法 [J]. 计算机系统应用, 2017, 26(9): 165-169.
- [11] 蔡敏敏. 基于 WiFi 指纹的室内定位系统中采样和匹配算法研究 [D]. 南京: 南京邮电大学, 2016.
- [12] 吴赟, 蒋新华, 齐雁飞, 等. 基于加权余弦相似度的 WiFi 指纹室内定位 [J]. 上海师范大学学报(自然科学版), 2017, 46(1): 124-128.
- [13] Castro P, Chiu P, Kremenek T, et al. A probabilistic room location service for wireless networked environments [C]//International Conference on Ubiquitous Computing. Springer-Verlag, 2001: 18-34.
- [14] Ge X, Qu Z. Optimization WIFI indoor positioning KNN algorithm location-based fingerprint [C]//IEEE International Conference on Software Engineering and Service Science. IEEE, 2017: 135-137.
- [15] 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 [C]//2009 国际信息技与应用论坛. 2009.