

doi:10.19306/j.cnki.2095-8110.2019.06.015

基于自适应修正曼哈顿距离的室内定位方法

陈亦奇,周蓉,滕婧,周洪波,栾泉中

(华北电力大学控制与计算机工程学院,北京 102206)

摘要:在室内 WiFi 环境下,针对常见指纹匹配算法所忽略的信号波动问题,提出了一种基于自适应修正曼哈顿距离和 AP 选择的指纹匹配算法,并结合加权 K 近邻方法实现定位。首先采用 AP 选择算法获取部分受干扰程度小和出现频率高的 AP,在指纹匹配时仅使用该部分 AP 的接收信号强度进行计算;在分析 WiFi 信号传播衰减公式和信号波动的基础上,提出了将自适应修正曼哈顿距离作为指纹匹配的度量距离,使用该距离旨在平滑信号波动对指纹相似度计算的影响;最后采用加权 K 近邻方法估计测试点的坐标。实验结果表明,在加权 K 近邻方法的框架下,基于自适应修正曼哈顿距离的定位算法在定位精度上优于基于欧氏距离、曼哈顿距离、余弦距离和 Sorensen 距离的定位算法。

关键词:室内定位;WiFi 指纹;相似度度量;信号波动;曼哈顿距离

中图分类号:TP393

文献标志码:A

开放科学(资源服务)标识码(OSID):

文章编号:2095-8110(2019)06-0094-09



Indoor Positioning Method Based on Adaptive Correction Manhattan Distance

CHEN Yi-qi, ZHOU Rong, TENG Jing, ZHOU Hong-bo, LUAN Quan-zhong

(School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China)

Abstract: In the indoor WiFi environment, a fingerprint matching algorithm based on the adaptive correction Manhattan distance (ACMD) and access point (AP) selection is proposed for the signal fluctuation problem neglected by the common fingerprint matching algorithms, and the weighted K nearest neighbor (WKNN) method is used to estimate the position. First, the AP selection algorithm is used to obtain reliable APs, and only received signal strengths (RSSs) from reliable APs are used for fingerprint matching. Second, after the WiFi signal propagation attenuation formula and signal fluctuation phenomenon are analyzed, the ACMD is proposed as a similarity metric, which is designed to smooth the effect of signal fluctuations on the calculation of fingerprint similarity. Finally, WKNN is used to estimate the coordinates of the Test Point. The experimental results show that in WKNN method, the proposed algorithm is better in positioning accuracy than other positioning algorithms using Euclidean distance, Manhattan distance, cosine distance or Sorensen distance.

Key words: Indoor positioning; WiFi fingerprint; Similarity metric; Signal fluctuation; Manhattan distance

收稿日期:2019-05-20;修订日期:2019-07-04

基金项目:国家自然科学基金(61503137,61871181);中央高校基本科研业务费专项资金(2017MS035)

作者简介:陈亦奇(1995-),男,硕士研究生,主要从事基于 WiFi 的室内定位方面的研究。E-mail:358118276@qq.com

通信作者:周蓉(1975-),女,讲师,硕导,博士,主要从事粒子滤波理论与应用、统计信号处理、目标跟踪与检测和室内定位方向的研究。E-mail:zhourong@ncepu.edu.cn

0 引言

无线通信和移动设备的激增为室内定位服务 (Location-Based Services, LBS) 带来了广泛的应用前景。然而,与已经通过全球定位系统 (Global Positioning System, GPS) 完美解决的室外定位问题不同,室内定位的挑战在于非视距 (Non-Line of Sight, NLOS) 问题,即由于屋顶、墙壁和一些其他障碍物的存在, GPS 信号容易衰减或受到屏蔽使得 GPS 定位不再可靠。目前,实现高精度的室内定位已经成为定位领域的研究热点,在现有的诸多室内定位方法中,由于 WiFi 接入点 (Access Point, AP) 的广泛部署和无处不在的智能设备 (如智能手机、平板和笔记本电脑),基于 WiFi 指纹的室内定位方法成为了最易于实现的室内定位方法^[1]。除此之外,指纹定位方法还具有以下优点:无需额外的硬件辅助,定位精度高,定位成本低,适用于不同的室内环境^[2]。指纹定位算法包括 2 个阶段:离线训练阶段和在线定位阶段。主要思想是将在线定位阶段采集到的测试点 (Test Point, TP) 指纹与离线训练阶段生成的指纹库中各参考点 (Reference Point, RP) 指纹进行匹配,根据指纹匹配结果预测 TP 的位置坐标。目前,这 2 个阶段的研究方向主要在于:更可靠的指纹特征值、更便捷的指纹库构建和更准确的指纹匹配^[3],本文着重于实现最后一项。

常见的指纹匹配算法通过计算指纹在信号空间内的欧氏距离来实现,欧氏距离越小,则指纹相似度越高,指纹在物理空间中的距离越近。一般情况下,指纹间的所有接收信号强度 (Received Signal Strength, RSS) 差值,无论大小均会被用于计算欧氏距离。但在实际环境下,即使在相同条件下,在同一位置,间隔很短一段时间采集的信号也具有波动性。这种波动性体现在 RSS 的持续变化上,通常由于采集时间、采集设备和环境的变化而导致。因此指纹间部分较小的 RSS 差值并不一定能反映指纹的不同,反而可能是信号波动性的结果。通常的处理方式是对 RSS 做均值滤波,但均值滤波后的 RSS 会损失一部分的信息,从而对定位精度产生影响。Fang 给出的处理方案是将信号波动转化为在信号强度对数域中的一个附加随机变量,但实际上的信号波动并不是非常契合该模型,处理后的定位精度不佳^[4]。Zhuang 提出了基于拥有不变性的信号强度阶代替 RSS 作为指纹匹配的特征,从而规避

了信号波动的影响^[5]。其他 WiFi 指纹定位算法都忽略了信号的波动性及其对定位精度的影响。

针对该问题,本文提出了自适应修正算法来处理信号波动,并根据自适应修正后的曼哈顿距离进行指纹匹配。首先,根据 AP 选择算法降低了指纹维度,仅选用可靠 AP 对应的 RSS;然后,在理论分析信号传播衰减公式和信号波动后,提出了基于自适应修正曼哈顿距离 (Adaptive Correction Manhattan Distance, ACMD) 计算指纹相似度的匹配算法;最后,结合加权 K 近邻 (Weighted K Nearest Neighbor, WKNN) 方法实现定位。在开放室内定位数据集 Zenodo^[6] 上进行定位实验,结果表明,在 WKNN 框架下,基于 ACMD 的指纹匹配算法在定位精度上优于基于欧氏距离 (Euclidean Distance, ED)、曼哈顿距离 (Manhattan Distance, MD)、余弦距离 (Cosine Distance, COSD) 和 Sorensen 距离 (Sorensen Distance, SD) 的指纹匹配算法。

1 相关工作

1.1 指纹定位原理

指纹定位算法通过计算 TP 和 RP 指纹之间的相似度,得到一组与 TP 相似度较高的 RP 集合,然后基于集合中 RP 的坐标加权估计 TP 的坐标。如图 1 所示,该算法分为 2 个阶段:离线训练阶段和在线定位阶段。

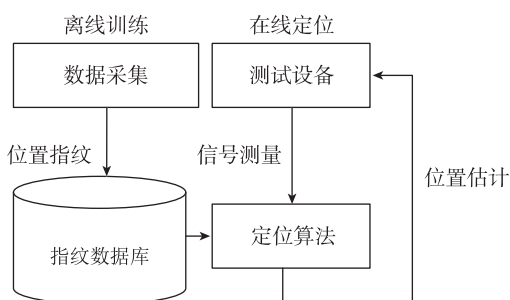


图 1 指纹定位算法流程图

Fig. 1 Flow chart of fingerprint positioning algorithm

指纹由一组采集到的不同 AP 的 RSS 组成,根据采集阶段的不同,指纹分为 TP 指纹和 RP 指纹。TP 指纹在在线定位阶段测得,用于估计 TP 的位置坐标;而 RP 指纹在离线训练阶段测得,存储在数据库中,与坐标一起构成指纹数据库,起预定义参照标准的作用。

离线训练阶段的任务是构建指纹数据库,指纹数

数据库由预设 RP 点物理空间坐标和该 RP 点的指纹组成。RSS 易受环境变化、行人走动或障碍物的影响,因此在训练阶段,通常在各个 RP 点进行多次 RSS 的采集以构建鲁棒的指纹库。假设 $RSS_{AP_i}(RP_j)$ 代表设备在 RP_j 处采集到 AP_i 的 RSS, AP 总数为 n , RP 总数为 m , 指纹库 FPDB 以矩阵形式保存, 如式(1)所示

$$FPDB = \begin{bmatrix} RSS_{AP_1}(RP_1) & RSS_{AP_2}(RP_1) & \cdots & RSS_{AP_n}(RP_1) & x_1 & y_1 \\ RSS_{AP_1}(RP_2) & RSS_{AP_2}(RP_2) & \cdots & RSS_{AP_n}(RP_2) & x_2 & y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ RSS_{AP_1}(RP_m) & RSS_{AP_2}(RP_m) & \cdots & RSS_{AP_n}(RP_m) & x_m & y_m \end{bmatrix} \quad (1)$$

在线定位阶段的任务是将 TP 点采集的指纹与指纹库中不同 RP 点的指纹进行匹配, 并根据匹配结果估计 TP 点位置以实现定位。在线定位阶段在 TP 点采集的指纹如式(2)所示

$$TP = [RSS_{AP_1} \quad RSS_{AP_2} \quad \cdots \quad RSS_{AP_n}] \quad (2)$$

1.2 指纹匹配算法

常见的指纹匹配算法大多是基于 KNN 类的确定性算法, 这类算法使用 RSS 均值或最大值作为信号特征。KNN 算法由 Bahl 和 Padmanabhan 首次引入到室内定位中, 通过在指纹空间中找到与 TP 欧氏距离最小的 K 个 RP, 计算 K 个 RP 的平均坐标作为 TP 的预测坐标^[7]。WKNN 方法是对 KNN 算法的改进, 基于 K 个 RP 与 TP 的欧氏距离分别赋予这 K 个 RP 相应的权重, 并根据各 RP 的权重对坐标加权实现定位。其中权重通常取欧氏距离的倒数, Chen 基于方差的加权距离改进 WKNN 算法, 改善了该算法在信号波动幅度较大时的定位精度^[8]。KNN 类的匹配算法实现简单, 定位精度较高, 因此其应用范围很广。但是也有着如下几种问题: 首先, 算法使用 RSS 均值或最大值来表示 WiFi 的信号强度, 这种简化会导致误差; 其次, 在信号空间内 K 个拥有最小欧氏距离的 RP 并不一定就是在物理空间内与 TP 距离最小的 K 个 RP。

除了基于 KNN 类的确定性匹配算法之外, 还有基于信号强度分布的概率性匹配算法, 这类方法将信号强度的概率分布作为特征。Zhou 将指纹数据库与物理近邻数据库相结合, 根据 2 次使用贝叶斯推论选择拥有最大先验概率的 RP 实现定位^[9]。Campos 通过组合无监督聚类和多表决反向传播神经网络构建室内定位模型, 提高了在跨越楼层的室

内环境下的定位精度^[10]。Calderoni 结合 RFID 技术和随机森林分类, 提出了一种能够抵御环境因素影响的定位系统, 已经在部分印度医院投入使用^[11]。概率性匹配算法相比确定性匹配算法避免了使用 RSS 均值导致的误差, 但也存在模型训练时间长、需要样本数据量大的缺点^[12]。

在基于 KNN 类的大多数指纹匹配算法中仅出现一些常见的度量距离, 如欧氏距离、曼哈顿距离和马氏距离。Xie 根据在同一位置采集的 RSS 大小排序基本稳定这一现象, 使用斯皮尔曼距离匹配指纹^[13]。Liu 提出的 M-WKNN 方法是基于聚类指纹及改进 WKNN 的方法, 采用一种指数函数方式衡量指纹间的相似度及指纹的权重^[14]。Han 针对信号采集设备不同这一问题, 提出了基于余弦距离匹配指纹^[15]。苗云龙等将指纹转换为包含 32 位 16 进制表示的 MD5 序列, 在指纹匹配时直接匹配 MD5 序列^[16]。Torres-Sospedra 同样对不同的指纹匹配距离进行了比较, 并在 KNN 框架下进行了定位比较, 实验结果表明, 欧氏距离并非指纹匹配的最佳距离, 基于 Sorensen 距离的匹配算法定位精度最高^[17]。

2 算法描述

2.1 数据预处理

为了建立鲁棒的指纹数据库, 需要先对 RSS 组进行预处理, 剔除异常数据和粗大误差。

假设向量 r_{ss} 是在 RP_j 上收集到来自 AP_i 的连续 p 次测量的 RSS, 如式(3)所示

$$r_{ss} = [r_{ss_{AP_i}^1}(RP_j), r_{ss_{AP_i}^2}(RP_j), \dots, r_{ss_{AP_i}^p}(RP_j)] \quad (3)$$

定义残差 r_q 用来表示 r_{ss} 向量中 $r_{ss_{AP_i}^q}(RP_j)$ 与均值 $\overline{r_{ss}}$ 的偏差程度, 如式(4)和式(5)所示

$$r_q = r_{ss_{AP_i}^q}(RP_j) - \overline{r_{ss}}, \quad q \in [1, p] \quad (4)$$

$$\overline{r_{ss}} = \frac{1}{p} \sum_{q=1}^p r_{ss_{AP_i}^q}(RP_j) \quad (5)$$

再计算残差的均方根误差 σ , 如式(6)所示

$$\sigma = \sqrt{\left(\frac{1}{p-1} \sum_{q=1}^p r_q^2 \right)} \quad (6)$$

WiFi 信号接收强度的分布可近似看成一种正态分布, 然后根据拉依达准则剔除粗大误差。若残差 $|r_q| > 3\sigma$, 则对应的 $r_{ss_{AP_i}^q}(RP_j)$ 将被作为粗大误差从 r_{ss} 中剔除, 最后对预处理后的 r_{ss} 求均值得到最终的 $RSS_{(AP_i)}(RP_j)$ 。

2.2 AP 选择

在理想环境下,指纹定位的精度会随着定位所使用 AP 数量的增加而持续升高。然而,在实际环境下,任何 AP 发出的信号都会受到障碍物和多径效应的影响,不经筛选地使用所有 AP 不仅不会提高定位精度反而还会增加额外的计算负担。因此,为了提高定位精度的同时减少计算负担,挑选出受干扰程度小和出现频率高的 AP 参与指纹匹配是一种可行的方案。

在稳定的室内环境下,使用 RSS 均值做信号特征的定位精度不如使用 RSS 最大值做信号特征的定位精度^[18]。同时较大的 RSS 也意味着信号接收装置与 AP 间的距离较小,信号受到多径效应的干扰程度也较小。所以信号的受干扰程度指标 $M(AP_i)_{RP_j}$ 与在 RP_j 处接收到 AP_i 的最大 RSS 值 $\max_{AP_i}(RP_j)$ 相关,如式(7)所示

$$M(AP_i)_{RP_j} = \frac{U - \max_{AP_i}(RP_j)}{\max_{AP_i}(RP_j)} \quad (7)$$

其中, U 是未接收到信号对应的预设 RSS,通常取 -105dBm , $M(AP_i)_{RP_j}$ 越大, RP_j 处的 AP_i 就越可靠。

在离线训练阶段的一个采样周期中,假设在 RP_j 处某一 AP 的信号出现频率较高,则在在线定位阶段中在 RP_j 处也会有较大概率接收到该 AP 的信号。如式(8)所示,信号的出现频率指标 $P(AP_i)_{RP_j}$ 用于表示在离线阶段的一个采样周期中,在 RP_j 处接收到 AP_i 信号的出现频率

$$P(AP_i)_{RP_j} = \log_2 \left(\frac{S_{RP_j}}{S_{RP_j} - p_{AP_i}(RP_j) + \epsilon} + 1 \right) \quad (8)$$

对于一个采样周期来说, S_{RP_j} 是在 RP_j 处信号采集的总次数, $p_{AP_i}(RP_j)$ 是在 RP_j 处接收到 AP_i 信号的次数, ϵ 是一个极小的正数,避免分母为 0。 $P(AP_i)_{RP_j}$ 越大,对应的 AP_i 在 RP_j 处出现的频率越高,也相应越可靠。

AP 选择算法的标准 $R(AP_i)_{RP_j}$ 由上述 2 个指标综合而得,反映了 AP_i 在 RP_j 处的可靠程度,如式(9)所示

$$R(AP_i)_{RP_j} = M(AP_i)_{RP_j} \times P(AP_i)_{RP_j} \quad (9)$$

基于 AP 的可靠程度对 RP_j 处所有 AP(下标为 $1 \sim n$) 进行排序,保存可靠度最高的前 L 个 AP 的集合。在 TP 指纹与 RP_j 指纹进行匹配时,仅使用 RP_j 对应的 L 个可靠 AP(下标为 $L1 \sim LL$),过程如图 2 所示。与普通的 AP 选择算法所不同的

是,该算法会在各 RP 处给出相对应的优选 AP 集合,根据 RP 对应的 AP 集合进行指纹匹配,匹配结果更为可靠。

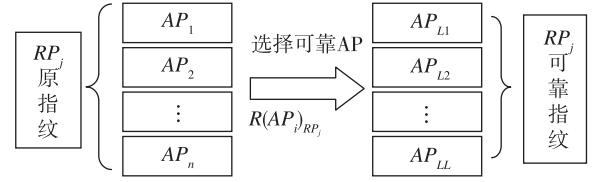


图 2 基于 AP 选择处理指纹的示意图

Fig. 2 Schematic diagram of modified fingerprint based on AP selection

2.3 相似度度量

在指纹匹配算法介绍中提到,常见算法的相似度度量距离是欧氏距离,但欧氏距离并非用于指纹匹配的最佳度量距离。WiFi 信号强度衰减模型如式(10)所示^[19]

$$RSS_{(d_i)} = RSS_{(d_0)} - \eta \log \left(\frac{d_i}{d_0} \right) + X \quad (10)$$

其中, RSS_d 是距离 WiFi 接入点距离为 d 的信号采集点处采集到的 RSS, η 是路径损耗参数, X 代表由采集设备、时间或者环境不同引起的 RSS 波动。由于 d_0 、 $RSS_{(d_0)}$ 、 η 均为预设模型参数,在采集到 $RSS_{(d_i)}$ 后,根据式(10)可以计算出 WiFi 接入点到信号采集点之间的未知距离 d_i ,如式(11)所示

$$d_i = d_0 10^{\left(\frac{RSS_{d_0} - RSS_{d_i} + X}{10\eta} \right)} \quad (11)$$

对于同一个 AP、RP 和 TP 之间的物理距离 Δd 与 RP 和 TP 采集到的 RSS 差值成正相关,如式(12)所示

$$\begin{aligned} \Delta d &= d(RP)_{(AP_i)} - d(TP)_{(AP_i)} \\ &= d_0 \left(10^{\left(\frac{RSS_{d_0} - RSS_{d(RP)_{(AP_i)}} + X_1}{10\eta} \right)} - 10^{\left(\frac{RSS_{d_0} - RSS_{d(TP)_{(AP_i)}} + X_2}{10\eta} \right)} \right) \\ &\propto RSS_{d(RP)_{(AP_i)}} - RSS_{d(TP)_{(AP_i)}} + X(X_1, X_2) \quad (12) \end{aligned}$$

由于 WiFi 信号差值与它们之间的物理距离正相关,因此,可以基于同一 AP 的 RP 和 TP 之间的 RSS 差来估计它们之间的物理距离,这是使用曼哈顿距离来进行指纹匹配的理论基础。 $X(X_1, X_2)$ 代表联合信号波动,下面引入自适应修正对信号波动进行处理。

2.4 自适应修正

在常见的指纹匹配方法中,指纹间的所有 RSS 差值无论大小均会被用于计算指纹的相似度。然而,即使是在相同位置连续采集的 RSS 也会持续变

化,本文将这种 RSS 的持续变化称为 RSS 波动。因此在指纹匹配时,指纹间的 RSS 差值应包含实际的 RSS 差异和 RSS 波动,部分较小的 RSS 差值可能完全是 RSS 波动的结果。RSS 波动如图 3 所示,若没有接收到某 AP 的信号,则将 AP 对应的 RSS 记作 -105dBm 。

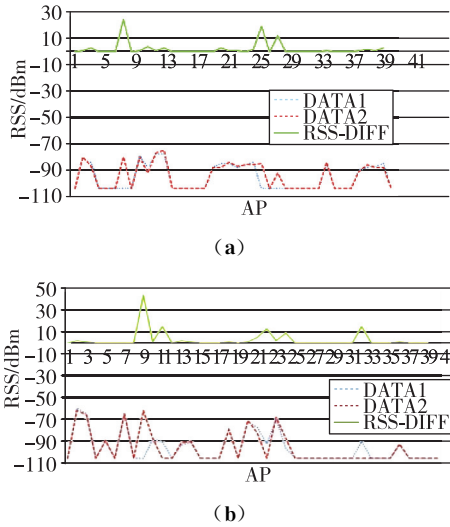


图 3 部分 AP(AP₁~AP₄₀)的 RSS 波动
Fig. 3 RSS fluctuation of part AP (AP₁~AP₄₀)

图 3(a)和图 3(b)分别代表了 2 个不同位置的情况,DATA1 代表从某一位置采集到 40 个 AP 的 RSS,DATA2 代表间隔 2s 后在与 DATA1 相同条件下采集到的 RSS,绘制在每张图顶部的 RSS-DIFF 代表 DATA1 与 DATA2 的 RSS 绝对差值,用于反映 RSS 的波动。根据图 3 可以观察到,即使在相同位置间隔极短时间采集的 RSS 也会有最大 10dBm 左右的波动,随着信号强度的衰减,波动的频率和最大值也逐渐增大。该现象的原因是 RSS 越小意味着采集设备距离 AP 越远,RSS 受到环境内障碍物和信号多径传播效应的影响也越大,若对该种波动不加以处理会影响指纹匹配的结果。

通过上面的分析,本文提出了使用自适应修正的方式处理在不同 RSS 大小下的 RSS 波动,下文提及的 RSS 差值均指差值的绝对值。修正的含义是给予指纹匹配时的 RSS 差值一个波动上限 T ,当 RSS 差值小于等于波动上限时,认为该 RSS 差值是 RSS 波动的结果,并将该 RSS 差值置为 0;当 RSS 差值大于波动上限时,认为该 RSS 差值代表了指纹的差异,将该 RSS 差值减去波动上限后保留。具体的规则如式(13)所示

$$c_diff = \begin{cases} diff - T, & diff > T \\ 0, & diff \leq T \end{cases} \quad (13)$$

其中, $diff$ 代表初始 RSS 差值, T 代表波动上限, c_diff 代表修正后的 RSS 差值。

之前的研究表明,RSS 波动更容易在距离 AP 较远的位置出现,且波动的范围也更大。因此,引入自适应函数 A 对各 RSS 值对应的波动上限 T 做相应的调整,该过程在离线训练阶段实现,不会影响定位的实时性,具体如式(14)所示

$$AT_{AP_i}(RP_j) = A(RSS_{AP_i}(RP_j)) \cdot T \quad (14)$$

其中, $AT_{AP_i}(RP_j)$ 是指纹库中信号强度 $RSS_{AP_i}(RP_j)$ 对应的波动上限,自适应函数 A 的实现如式(15)和式(16)

$$r_{SS_{AP_i}}(RP_j) = \frac{U - RSS_{AP_i}(RP_j)}{RSS_{AP_i}(RP_j)} \quad (15)$$

$$= \frac{U}{RSS_{AP_i}(RP_j)} - 1$$

$$A(RSS_{AP_i}(RP_j)) = \alpha^{\frac{r_{SS_{AP_i}}(RP_j)}{\text{MAX}(all_rss)}} \quad (16)$$

其中, U 是未接收到信号对应的信号强度,通常取 -105dBm , all_rss 以矩阵形式存储了经过式(15)变换后的所有 r_{ss} ,底数 α 用于改变自适应系数的变化范围。某位置的 RSS 值越大,意味着该位置距离 AP 越近,出现波动的可能性和波动的变化范围越小,因此波动上限 T 也越小,对 RSS 波动的界定也越严格,修正程度也更小。自适应修正 RSS 差值的规则与修正 RSS 差值的规则类似,如式(17)所示

$$c_diff = \begin{cases} diff - AT, & diff > AT \\ 0, & diff \leq AT \end{cases} \quad (17)$$

其中, AT 代表自适应的波动上限。

此外,RSS 的数值实际上表示接收信号功率 $Power$ 与 1mW 参考功率之间的比例,如式(18)所示

$$RSS = 10 \cdot \log_{10} \left(\frac{Power_{\text{mW}}}{1\text{mW}} \right) \quad (18)$$

在不同的信号强度下,信号强度差与信号功率差不适用线性关系来转化。举例来说,信号强度 -50dBm 与 -60dBm 之间的信号功率差是 $9 \times 10^{-6}\text{mW}$,信号强度 -90dBm 与 -100dBm 之间的信号功率差是 $9 \times 10^{-10}\text{mW}$;然而在现有的指纹定位算法中,上面 2 组信号的 RSS 差值被认为是相同的 10dBm,实际上 2 组信号的功率差相差整整 10^4 倍,仅靠 RSS 差值计算指纹相似度无疑会引入误差。因此,设置 RSS 差值的调整函数 M ,基于信号强度给予 RSS 差值相应的调整系数, M 函数通过

式(19)获得

$$M(RSS_{AP_i}(RP_j)) = \frac{1}{A(RSS_{AP_i}(RP_j))} \quad (19)$$

通过 M 函数调整 RSS 差值,使得较大信号强度对应的 RSS 差值在指纹相似度计算的过程中具有更大的系数,但调整系数的变化范围同样不宜设置过大,避免破坏指纹数据的真实性,影响指纹匹配的结果。

综上所述,首先采用自适应修正的方式解决指纹匹配中的信号波动问题,然后基于信号强度的大小对相应的 RSS 差值作出调整。在进行 RP_j 指纹和 TP 指纹的匹配时,指纹间的自适应修正曼哈顿距离如式(20)所示,即 L 个优选 AP(下标 $L1 \sim LL$)对应的修正 RSS 差值的总和

$$ACMD_{RP_j} = \sum_{i=L1}^{LL} (c_diff_{AP_i} \cdot M(RSS_{AP_i}(RP_j))) \quad (20)$$

2.5 基于 ACMD-WKNN 的定位算法

基于 ACMD-WKNN 的定位算法的思想是:在完成数据预处理和 AP 选择之后,基于 ACMD 进行指纹匹配,用 WKNN 方法实现位置估计,具体流程图如图 4 所示。

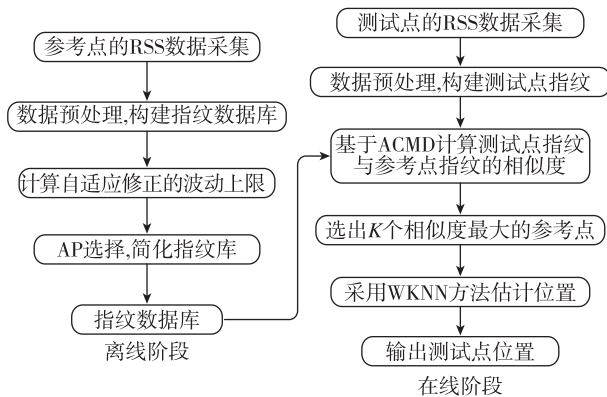


图 4 基于 ACMD-WKNN 的定位算法流程图

Fig. 4 Flow chart of positioning algorithm based on ACMD-WKNN

假设计算得到 K 个相似度最大的参考点,下标 j 的变化范围是 $[1, K]$, TP 指纹与 RP_j 指纹的 ACMD 是 $ACMD_{RP_j}$,对应的权重 W_{RP_j} 如式(21)所示, M 为一较大数,这里暂取 100

$$W_{RP_j} = \begin{cases} \frac{1}{ACMD_{RP_j}}, & ACMD_{RP_j} \neq 0 \\ M, & ACMD_{RP_j} = 0 \end{cases} \quad (21)$$

参考点 RP_j 的坐标是 (X_j, Y_j) , TP 的位置

(X, Y) 可通过式(22)估计获得

$$X = \frac{\sum_{j=1}^K (X_j \times W_{RP_j})}{\sum_{j=1}^K W_{RP_j}}, Y = \frac{\sum_{j=1}^K (Y_j \times W_{RP_j})}{\sum_{j=1}^K W_{RP_j}} \quad (22)$$

3 定位实验

3.1 实验环境与步骤

本文使用学者 Mendoza-Silva G、Richter P 及 Torres-Sospedra J 等提供的开放室内定位数据集 10.5281/zenodo.1066040 作为实验数据集。该数据集从一幢图书馆的 3 层与 5 层分别采集获得,采集区域为各层面积为 125.4m^2 的书架区域,实验数据集包含总共 15 个月的测量数据,该数据集存储了共 448 个以 MAC 地址为唯一区分的 AP(单路由器上有多 AP),在每层 24 个共 48 个 RP 处的信号强度值。此外,采集时会预先给各 RP 设置序号,并根据序号进行采集,为了避免偶然性带来的误差,在各采集点上均进行 6 次采集。每月的数据集通常包含 1 组训练集与 5 组测试集,数据采集点的分布如图 5 所示。

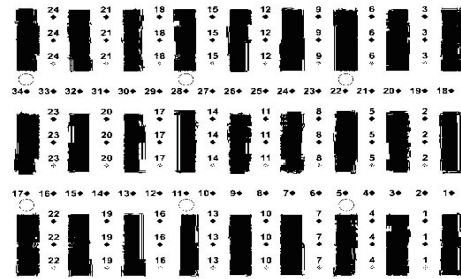


图 5 数据采集点的位置及分布情况

Fig. 5 Position and distribution of data collection points

图 5 中,黑色矩形是书柜,横向来看,每 2 排书柜间有 3 个采集点,从上至下依次属于测试集 2、训练集(测试集 1、测试集 5)和测试集 3;纵向来看,每 2 排书架间的采集点属于测试集 4,按照序号升序和降序的顺序先后各进行 1 次测量。选择该数据集进行实验的原因有二点:1)该数据集提供长达 15 个月的信号强度数据,大数据量的实验结果保证了定位算法的可靠性与鲁棒性;2)在该数据集中包含有大量可用于模拟用户持续移动的测试点,如测试集 2 和测试集 4 对应模拟用户在书架间行走和走道上行走的不同状态,方便后续展开基于目标跟踪的研究。在本次实验中选用第 3 层全部月份所有训练集

和测试集的指纹数据进行定位实验。

3.2 波动上限 T 的影响

自适应修正算法中波动上限 T 的变化会导致修正后的 RSS 发生变化,进而影响指纹匹配结果。若 T 过大,会导致对 RSS 波动的过度修正,将较多相似度并不高的指纹保留下来;若 T 过小,算法会退化成无修正算法,无法实现对 RSS 波动的平滑。在实验中取 T 为 0dBm、5dBm 和 10dBm,结果如图 6 所示。

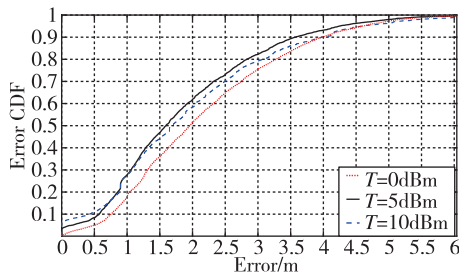


图 6 基于不同波动上限 T 的误差 CDF

Fig. 6 Error CDF based on different fluctuation upper limit T

图 6 以误差累计概率分布函数 (Cumulative Distribution Function, CDF) 的形式给出结果,当 $T=0$ dBm 时,即基于无修正曼哈顿距离的定位精度并不如 $T=5$ dBm 或 10dBm 的定位精度。因此,基于自适应修正曼哈顿距离的算法相比无修正的算法的确能够提升定位精度。除此之外, $T=10$ dBm 时的定位精度并不如 $T=5$ dBm 时的定位精度,即定位精度并不会随着 T 的增加而不断提高。根据图 6 基本确定了最佳 T 在 5dBm 附近,进一步缩小了 T 的测试范围到 3~7dBm,基于不同 T 的定位误差如表 1 所示。由于 T 相近时 CDF 图也较为接近,故不再依靠误差 CDF 图分析实验结果。

表 1 基于不同 $T(3\sim 7\text{dBm})$ 的定位误差

Tab. 1 Positioning error based on different $T(3\sim 7\text{dBm})$

T/dBm	3	4	5	6	7
平均误差/m	1.91	1.89	1.86	1.85	1.86
75% 概率误差/m	2.64	2.62	2.59	2.56	2.58

通过表 1 可以得到,当 $T=3\sim 7$ dBm 时,平均误差及 75% 概率误差均在 $T=6$ dBm 时取到最小,故在本文的实验中取 $T=6$ dBm,实验结果也从侧面证明了处理信号波动能够提高定位精度。

3.3 底数 α 的影响

底数 α 用于生成自适应的波动上限 T ,波动上限 T 随 RSS 值的大小自适应发生改变,图 7 是基于不同的底数 α 的值所绘制的 RSS 与自适应波动上限 AT 的关系图,其中初始波动上限 T 取 6dBm, e 为自然对数。

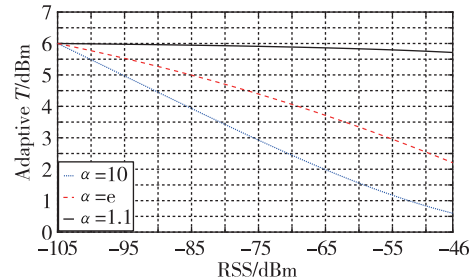


图 7 基于不同 α 的 RSS 与 AT 的关系图

Fig. 7 Relationship between RSS and AT based on different α

从图 7 中可以看出,基于不同底数 α 的 RSS-AT 曲线在 RSS 最小时均取到最大波动上限,但随着底数 α 的增大,AT 的变化范围也逐渐增大,在 RSS 取到最大值 -46 dBm 时,底数 α 取 1.1、 e 和 10 对应的波动上限分别为 5.405dBm、2.207dBm 和 0.645dBm。根据 3.2 节的结论,过小的波动上限无法实现对 RSS 波动的平滑,且波动上限 T 在 6dBm 附近取得最小的定位误差,因此自适应的波动上限 AT 也应在 6dBm 附近变化。当 α 取 1.1 时能够满足上述要求,此时波动上限 T 的变化范围是 $[5.405, 6]$,在给予大 RSS 值小波动上限的同时保证了对全部 RSS 的修正效果。

3.4 AP 选择数量 L 的影响

AP 选择算法的目标是在选出更可靠、受干扰更小的 AP 的同时减轻计算负担。因此,确定 AP 选择算法所选 AP 数量 L 的重要性不言自明,实验中将选择 AP 数量 L 设置为 10~70,用于分析 L 对定位精度的影响;同时与无 AP 选择的算法进行对比,用于验证 AP 选择算法对于定位精度的改进效果,结果如图 8 所示。实验中 AP 的总数为 448 个,横轴上的“NO-Selection”代表无 AP 选择 ($L=448$) 的情况。

根据图 8 可知,当 L 小于 50 时,随着用于定位的 AP 数量 L 的增大,定位精度也随之提升;当 L 达到 50 后,定位精度最佳,且优于无 AP 选择算法的定位精度;当 L 超过 50 后,算法选择的 L 个 AP 中不可靠

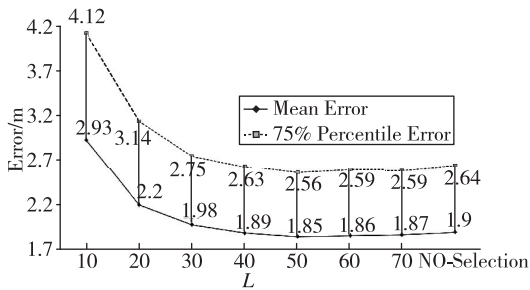


图 8 基于不同 AP 选择数量 L 的误差

Fig. 8 Error based on different AP selection number L

AP 的数量增多,定位精度反而有所下滑。因此,AP 选择算法能够在减少指纹匹配计算量的同时提高定位精度,本文设置 AP 选择数量 L 为 50。

3.5 基于不同距离的匹配算法比较

实验中实现了基于 ACMD 和基于欧氏距离、曼哈顿距离、余弦距离及 Sorensen 距离的指纹匹配算法,并在 WKNN 框架下进行了定位实验。根据定位结果进行比较和分析,基于不同距离的匹配算法对应的定位精度如图 9 所示,WKNN 方法中的 K 取 3。

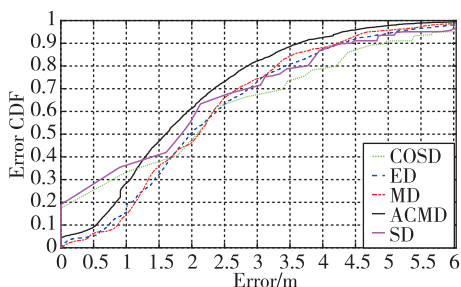


图 9 基于不同距离的匹配算法定位误差

Fig. 9 Positioning error based on matching algorithm of different distance

首先,基于不同距离的匹配算法从平均定位误差、67%累积概率误差、95%累积概率误差、定位误差标准差和精度提升比 5 个方面进行了比较,结果如表 2 所示。

表 2 基于不同距离的定位误差

Tab. 2 Positioning error based on different distance

距离	ACMD	COSD	ED	MD	SD
平均误差/m	1.85	2.50	2.30	2.25	2.18
67%概率误差/m	2.20	2.92	2.72	2.55	2.54
95%概率误差/m	4.24	5.69	5.09	4.67	5.21
误差标准差/m	1.26	2.02	1.45	1.35	1.81
提升比(平均误差)	无	26.0%	19.6%	17.8%	15.1%

其次,基于不同距离的匹配算法在固定精度要求下的误差累积概率如表 3 所示。

表 3 在固定精度要求下的误差累积概率表

Tab. 3 Accumulated probability of error under fixed accuracy requirements

距离/m	1	2	3
ACMD	28.41%	62.41%	83.23%
COSD	33.31%	48.10%	67.44%
ED	17.67%	50.67%	73.28%
MD	14.59%	46.33%	74.41%
SD	37.41%	56.06%	71.02%

综上所述,基于 ACMD 的匹配算法相对于其他距离的匹配算法在定位精度上有显著提升,且定位误差的标准差更小,定位误差在 2m 和 3m 内的概率分别达到了 62.41%和 83.23%。

4 结论

本文针对常见的指纹匹配算法中所忽略的信号波动问题,提出了一种基于自适应修正曼哈顿距离的 WiFi 指纹匹配算法,并结合 AP 选择算法简化了指纹匹配过程,最后基于 WKNN 方法实现定位。算法分析与实验结果表明:

1)在实际环境下,由环境因素、采集设备或采集时间差异导致的信号波动总是存在,并会对定位结果产生影响。本文所提出的自适应修正算法在指纹匹配时,根据信号强度自适应地对指纹间的信号强度差值进行修正,减少信号波动对指纹匹配的影响,为后续研究处理信号的波动性问题提供了新的思路。修正过程的自适应参数在离线阶段获得,不会对匹配和定位带来额外的计算负担。

2)基于自适应修正曼哈顿距离的指纹匹配算法相比基于其他几种常见距离的指纹匹配算法定位误差更小,定位更稳定,同时证实了欧氏距离并非指纹匹配时的最佳相似度度量距离。本文的算法实现了在面积为 125.4m² 的定位区域内,平均定位误差 1.85m,62.41%的定位误差在 2m 以内和 83.23%的定位误差在 3m 以内的定位效果。

3)本文提出的基于自适应修正曼哈顿距离的指纹匹配算法还缺乏与更多相似度度量距离的比较,因此在后续工作中会增加更多用于对比的相似度度量距离。目前仅在 Zenodo 数据集上进行了定位实验,后续将在更多知名的定位数据集和真实环

境中进行测试。

参考文献

- [1] 赵龙, 陶冶. 基于WiFi指纹库的室内定位研究进展和展望[J]. 导航定位与授时, 2018, 5(3): 6-13.
Zhao Long, Tao Ye. Research progress and prospect of indoor positioning based on WiFi fingerprint database[J]. Navigation Positioning and Timing, 2018, 5(3): 6-13(in Chinese).
- [2] He S, Chan S H G. Wi-Fi fingerprint-based indoor positioning: recent advances and comparisons[J]. IEEE Communications Surveys & Tutorials, 2015, 18(1): 466-490.
- [3] 邓中亮, 尹露, 唐诗浩, 等. 室内定位关键技术综述[J]. 导航定位与授时, 2018, 5(3): 18-23.
Deng Zhongliang, Yin Lu, Tang Shihao, et al. A survey of key technology for indoor positioning[J]. Navigation Positioning and Timing, 2018, 5(3): 18-23(in Chinese).
- [4] Fang X, Nan L, Jiang Z, et al. Multi-channel fingerprint localisation algorithm for wireless sensor network in multipath environment[J]. IET Communications, 2017, 11(15): 2253-2260.
- [5] Zhuang J, Zhang J, Zhou D, et al. An improved Wi-Fi indoor positioning method via signal strength order invariance[C]//2014 IEEE International Conference on Computer and Information Technology. IEEE, 2014: 3-6.
- [6] Mendoza-Silva G, Richter P, Torres-Sospedra J, et al. Long-term WiFi fingerprinting dataset for research on robust indoor positioning[J]. Data, 2018, 3(1): 3.
- [7] Bahl P, Padmanabhan V N, Bahl V, et al. RADAR: an in-building RF-based user location and tracking system[J]. Proceedings IEEE Infocom, 2000(2): 775-784.
- [8] 陈空, 宋春雷, 陈家斌, 等. 基于改进WKNN的位置指纹室内定位算法[J]. 导航定位与授时, 2016, 3(4): 58-64.
Chen Kong, Song Chunlei, Chen Jiabin, et al. An indoor location fingerprint algorithm based on improved WKNN[J]. Navigation Positioning and Timing, 2016, 3(4): 58-64(in Chinese).
- [9] Mu Z, Qiao Z, Tian Z, et al. Integrated location fingerprinting and physical neighborhood for WLAN probabilistic localization[C]// 5th International Conference on Computing, Communications and Networking Technologies(ICCNT). IEEE, 2014.
- [10] Campos R S, Lovisolò L, de Campos M L R. Wi-Fi multi-floor indoor positioning considering architectural aspects and controlled computational complexity[J]. Expert Systems with Applications, 2014, 41(14): 6211-6223.
- [11] Calderoni L, Ferrara M, Franco A, et al. Indoor localization in a hospital environment using random forest classifiers[J]. Expert Systems with Applications, 2015, 42(1): 125-134.
- [12] Torres-Sospedra J, Jiménez A, Knauth S, et al. The smartphone-based offline indoor location competition at IPIN 2016: analysis and future work[J]. Sensors, 2017, 17(3): 557.
- [13] Xie Y, Wang Y, Nallanathan A, et al. An improved K-nearest-neighbor indoor localization method based on spearman distance[J]. IEEE Signal Processing Letters, 2016, 23(3): 351-355.
- [14] Liu W, Fu X, Deng Z, et al. Smallest enclosing circle-based fingerprint clustering and modified-WKNN matching algorithm for indoor positioning[C]// 2016 International Conference on Indoor Positioning and Indoor Navigation(IPIN). IEEE, 2016: 1-6.
- [15] Han S, Zhao C, Meng W, et al. Cosine similarity based fingerprinting algorithm in WLAN indoor positioning against device diversity[C]// 2015 IEEE International Conference on Communications (ICC). IEEE, 2015: 2710-2714.
- [16] 苗云龙, 陆彦辉, 尹峰, 等. 基于MD5-KNN的Wi-Fi室内定位算法研究[J]. 计算机应用研究, 2019, 36(9).
Miao Yunlong, Lu Yanhui, Yin Feng, et al. Research on Wi-Fi indoor location algorithm based on MD5-KNN[J]. Application Research of Computers, 2019, 36(9)(in Chinese).
- [17] Torres-Sospedra J, Montoliu R, Trilles S, et al. Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems[J]. Expert Systems with Applications, 2015, 42(23): 9263-9278.
- [18] Xue W, Qiu W, Hua X, et al. Impact analysis of rssi characteristic values on wifi indoor positioning accuracy[J]. Journal of Geomatics, 2016, 41(4): 23-26.
- [19] Luo R C, Hsiao T J. Dynamic wireless indoor localization incorporating with an autonomous mobile robot based on an adaptive signal model fingerprinting approach[J]. IEEE Transactions on Industrial Electronics, 2018, 66(3): 1940-1951.