

DOI:10.11992/tis. 201704031

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20170703.1854.014.html>

一种层次 Levenshtein 距离的无指纹校准的室内定位方法

何富贵^{1,3}, 杨铮², 吴陈沐², 赵姝³, 周先存¹

(1. 皖西学院 电子与信息工程学院, 安徽 六安 237012; 2. 清华大学 软件学院可信网络与系统研究所, 北京 100084; 3. 安徽大学 智能计算与知识工程研究所, 安徽 合肥 230039)

摘要:随着移动计算领域的兴起, 基于位置的服务越来越受青睐。目前各种室内定位的方法层出不穷, 由于室内广泛部署了无线基础设施, 基于 WiFi 指纹信息的室内定位技术是其主流方法。设备异构和室内环境变化是影响定位精度的主要因素。本文针对以上两个问题, 提出一种层次 Levenshtein 距离 (HLD) 的 WiFi 指纹距离计算算法, 实现异构设备的指纹无校准比对。将不同移动设备采集的 RSSI 信息转化为 AP 序列, 根据 AP 对应的 RSSI 值的差异性计算其层次能级, 结合 Levenshtein 距离计算 WiFi 指纹之间的距离。对于需定位的 WiFi 指纹 RSSI 信息, 利用 HLD 算法获取 K 个近邻, 采用 WKNN 算法进行预测定位。实验中, 为了验证算法的鲁棒性和有效性, 在 3 种不同类型的室内环境中采用 5 种不同的移动设备来采集 WiFi 的 RSSI 信息, 其定位的平均精度达 1.5 m。

关键词:室内定位; WiFi 指纹; 设备异构; 无指纹校准; Levenshtein 距离

中图分类号:TP181 **文献标志码:**A **文章编号:**1673-4785(2017)03-0422-08

中文引用格式:何富贵, 杨铮, 吴陈沐, 等. 一种层次 Levenshtein 距离的无指纹校准的室内定位方法[J]. 智能系统学报, 2017, 12(3): 422-429.

英文引用格式:HE Fugui, YANG Zheng, WU Chenshu, et al. An fingerprint calibrations-free indoor localization method based on hierarchical Levenshtein distance[J]. CAAI transactions on intelligent systems, 2017, 12(3): 422-429.

An fingerprint calibrations-free indoor localization method based on hierarchical Levenshtein distance

HE Fugui^{1,3}, YANG Zheng², WU Chenshu², ZHAO Shu³, ZHOU Xiancun¹

(1. School of Electronics and Information Engineering, West Anhui University, Lu'an 237012, China; 2. Institute of Trustworthy Network and System, School of Software, Tsinghua University, Beijing 100084, China; 3. Institute of Intelligent Computing and Knowledge Engineering, Anhui University, Hefei 230039, China)

Abstract: In the era of mobile computing, location-based services have become extremely important for a wide range of applications, and various wireless indoor localization techniques have been emerging. Amongst these techniques, WiFi fingerprint-based indoor localization is one of the most attractive because of the wide deployment and availability of WiFi infrastructure. The accuracy of indoor localization is affected by two main factors: equipment heterogeneity and environmental dynamics. To solve the above two problems, an algorithm based on hierarchical Levenshtein distance (HLD) was proposed to realize calibration-free fingerprint comparison of heterogeneous devices. Received signal strength indication (RSSI) information collected via different mobile devices was transformed into an AP sequence. The difference in the Received signal strength indication RSSI values was used to calculate the hierarchical energy level of each access point (AP). Next, the distance between the WiFi fingerprints was calculated using the Levenshtein distance. To locate WiFi fingerprint RSSI information, the HLD algorithm was used to obtain K neighbors and the weighted K nearest neighbor (WKNN) algorithm was used to predict its position. Five different mobile devices were used to collect WiFi RSSI information in three different types of indoor environments to verify the robustness and effectiveness of the algorithm. The average localization accuracy was 1.5 m.

Keywords: indoor localization; WiFi fingerprint; heterogeneous device; fingerprint calibration-free; Levenshtein distance

基于位置的服务, 如路径导航、广告推荐和邻

近社交网络等, 随着智能手机的普及已变得尤为重要。在室内环境中 GPS 对卫星信号的接收较弱无法满足用户的需求, 目前已有许多室内定位技术^[1-3]涌现。随着 IEEE802.11 (WiFi) 普及^[4], 基于 WiFi 指纹室内定位技术已成为研究的热点。现有

收稿日期: 2017-04-23. 网络出版日期: 2017-07-03.

基金项目: 国家自然科学基金项目 (61572366, 61303209, 61522110, 61402006, 61673020); 2016 年安徽省高校优秀中青年骨干人才国内外访学研修重点项目 (gxfkZD2016190); 安徽大学信息保障技术协同创新中心 2015 年度开放课题 (ADXXBZ201504).

通信作者: 何富贵. E-mail: fuguihe@163.com.

的 WiFi 指纹室内定位主流方案由两阶段组成:现场勘查和指纹匹配。

RADAR^[5]在未考虑随机信号前提下对 WiFi 指纹利用欧氏距离和 K 近邻算法进行室内目标定位, Horus^[6]在此基础上考虑 RP 信号概率分布进行定位估计。建立和更新指纹库是决定基于 WiFi 指纹定位精度高低的关键。通过专业人士来收集 WiFi 成本昂贵,基于众包模式^[7]可明显降低指纹地图构建和维护成本,但会带来一些新的问题。

由于无线信号在传播过程存在多径效应,对环境的变化十分敏感,不同时间采集的数据都存在差异。环境动态对室内定位系统精度的影响是与生俱来的。对于设备异构的室内定位问题^[8-9],同一位置的异构设备检测到的 RSSI 值通常有不同的值,这严重降低了定位精度^[10-11]。为了处理基于 WiFi 指纹识别的室内定位系统所遇到的设备异质性问题,已有不同的解决方案^[8,10-15]。

由于用不同设备获取 RSSI 存在差异^[10,16],直接用 WiFi 信号强度进行预测定位无法达到在现场勘查和指纹匹配的过程中用相同的设备采集 WiFi 指纹的米级定位精度^[5-6,17-18]。因此,仅依赖于绝对信号强度测量来实现异构设备的定位是不可行的,迫切需要开发一种替代绝对 RSSI 的鲁棒指纹定位技术。在文献[8,15]中提出了一些无校准方法,以避免每个测试设备使用繁琐的手动 RSSI 校准程序。协作映射通过训练在线测量的 RSSI 值来估计线性映射函数^[15]。无监督的学习方法(如在线回归和期望最大化)已被用来学习映射函数^[8]。然而,以上无校准的指纹定位方法都需要耗时的在线处理过程。解决设备异构问题的另一种方式是定义和使用替代位置指纹,而不是绝对 RSSI 值。在文献[10]中提出信号强度差(SSD),Yang^[19]和 Jiang^[20]提出将 RSSI 测量向量转换为相对的 RSS 排序作为指纹,FreeLoc^[19]使用 Key-Value 机制构造指纹。Jiang^[20]提出使用长度为 $n < N$ 的有序 AP 索引向量的子序列作为房间指纹来实现房间级定位。GIFT^[21]引入 RSSI gradient 替代 RSSI 信息。文献[22]利用 Procrustes 分析法将绝对 RSSI 值进行标准化来消除设备异构的影响。针对动态环境下 WiFi 指纹定位,以适应短期和长期指纹受环境变化,HED^[23]提出基于 RSSI 的 AP 索引向量的容错序列匹配方法。但

该方法未考虑不同 AP 对位置定位的不同程度的影响力。

通过对 WiFi 指纹信息分析得知:1)对于同一位置,在容忍一定差异情况下,不同设备检测的 AP 的 RSSI 值形成的变化表现出很高的相似性;2)对于不同位置,在定位匹配过程中不同 AP 对其准确定位的贡献不同。不同 AP 在同一位置测量得到的 RSSI 值存在差异。其中一些 AP 对应的 RSSI 值较大,这些 AP 对定位的准确性影响较大。

为了实现无指纹校准的室内定位,考虑设备异构和环境动态的影响,本文提出基于层次 Levenshtein 距离的 WiFi 指纹距离计算算法。为了刻画同一位置不同设备检测的 RSSI 值的变化相似性,将不同移动设备采集的 RSSI 信息按照从大到小进行降序排列索引,将绝对 RSSI 信息转化为相对的 AP 序列,以实现不同设备采集的数据量纲的一致性。同时,对于各数据采集点根据各 AP 对应的 RSSI 值的差异性计算其层次能级,将各个 AP 映射到不同层次能级上,以描述各 AP 对位置定位的影响等级。联合各 AP 层次能级,利用 Levenshtein 距离来计算 AP 序列间的距离,实现异构设备的指纹无校准比对。对于需定位的 WiFi 指纹 RSSI 信息,利用 HLD 算法获取 K 个近邻,采用 WKNN 算法进行预测定位。

1 问题描述

对于室内定位系统, K 个发射点(access point, AP) $\{AP_1, AP_2, \dots, AP_K\}$,对于每个接收点(receive point, RP)接收到信号强度为 $RP_i = (RSSI_1^i, RSSI_2^i, \dots, RSSI_K^i)$,其空间位置为 $f_i = (x_i, y_i, z_i)$ 。对于 RP 来说,不能获取到所有的 AP 信号强度,故对未获取到的 RSSI 值设定为最小值,即 -95 dBm(分贝毫瓦)。在定位过程中,对于需要定位的位置 f_i 通过距离函数 $Dis(RP_i, RP_j)$ 在指纹数据库中匹配与其最近邻的 N 个指纹 $N(i)$, N 近邻对应的位置为 $\{Y_{N(1)}^i, Y_{N(2)}^i, \dots, Y_{N(N)}^i\}$ 。 f_i 的近邻权重 W_{ij} 与 $Dis(RP_i, RP_j)$ 相关。对于所有需定位位置,实际的定位位置 f_i 与通过 N 近邻预测的位置 $\hat{Y}_b = \sum_{j=1}^N W_{ij} Y_{N(j)}^i$ 差异越小,表明其定位系统的性能越好。

$$\begin{aligned} & \text{Minimize } \sum_i \left| f_i - \sum_{j=1}^N W_{ij} Y_{N(j)}^i \right| \\ & \text{Subject to } \sum_{j=1}^N W_{ij} = 1 \end{aligned} \quad (1)$$

通常情况下,距离函数 $\text{Dis}(\text{RP}_i, \text{RP}_j)$ 可通过欧氏距离、余弦距离或 Procrustes 氏距离等距离函数来计算。由于设备异构、环境的动态性,无法直接用绝对 RSSI 值来计算。本文用相对 RSSI 排序的 AP 序列替代 RSSI 绝对的信息,先对 $\text{RP}_i = (\text{RSSI}_1^i, \text{RSSI}_2^i, \dots, \text{RSSI}_K^i)$ 按照 RSSI 值进行降序,记录对应的 AP 序列 $\text{Seq}_{\text{RP}_i} = (N_1^{\text{RP}_i}, N_2^{\text{RP}_i}, \dots, N_K^{\text{RP}_i})$, 其中

$$d(\mathbf{a}_i, \mathbf{b}_j) = \begin{cases} 0, & i = 0, j = 0, \\ \min(d(\mathbf{a}_{i-1}, \mathbf{b}_j) + 1, d(\mathbf{a}_i, \mathbf{b}_{j-1}) + 1, d(\mathbf{a}_{i-1}, \mathbf{b}_{j-1})), & a_i = b_j \\ \min(d(\mathbf{a}_{i-1}, \mathbf{b}_j) + 1, d(\mathbf{a}_i, \mathbf{b}_{j-1}) + 1, d(\mathbf{a}_{i-1}, \mathbf{b}_{j-1}) + 1), & a_i \neq b_j \end{cases} \quad (2)$$

式中 $\mathbf{a}_i, \mathbf{b}_j$ 分别为长度为 i, j 字符串序列; a_i, b_j 分别为 $\mathbf{a}_i, \mathbf{b}_j$ 字符串中第 i, j 序号。

对 RP_i 和 RP_j 按照 RSSI 值进行降序排列后得到 AP 序列 $\text{Seq}_{\text{RP}_i} = (N_1^{\text{RP}_i}, N_2^{\text{RP}_i}, \dots, N_K^{\text{RP}_i})$, $\text{Seq}_{\text{RP}_j} = (N_1^{\text{RP}_j}, N_2^{\text{RP}_j}, \dots, N_K^{\text{RP}_j})$ 。利用 Levenshtein 距离计算 $\text{Dis}(\text{Seq}_{\text{RP}_i}, \text{Seq}_{\text{RP}_j})$ 。

对于 $\forall k, \exists \text{RSSI}_{N_k^{\text{RP}_i}}^i \geq \text{RSSI}_{N_{k+1}^{\text{RP}_i}}^i$, 表示 RP_i 从 $\text{AP}_{N_k^{\text{RP}_i}}$ 获得的 RSSI 值要大于等于从 $\text{AP}_{N_{k+1}^{\text{RP}_i}}$ 获得的。在定位匹配过程中发现,不同 AP 对位置准确定位的贡献不同, RSSI 值越大对该位置定位的精度影响越大,通过层级能级来刻画各 AP 对 RP 的贡献。从理论上证明,在一定噪声范围内,接收信号强度之差小于 ω , 区域内 AP 序列顺序无差异。

为了区别 RSSI 值差异性,对 $\text{Seq}_{\text{RP}_i} = (N_1^{\text{RP}_i}, N_2^{\text{RP}_i}, \dots, N_K^{\text{RP}_i})$ 依据 RSSI 值差别大小分成不同层级能级,记为 $\text{level}_{\text{RP}_i} = (l_1^{\text{RP}_i}, l_2^{\text{RP}_i}, \dots, l_K^{\text{RP}_i})$ 。规定两个相邻 AP 对应的 RSSI 值 $\text{RSSI}_{N_{j-1}^{\text{RP}_i}}^i$ 与 $\text{RSSI}_{N_j^{\text{RP}_i}}^i$ 之差小于阈值 θ , 其层级能级相同。考虑到连续出现相邻 RSSI 值差小于阈值 θ , 最大与最小值差很大,故对最大与最小值差大于 2θ 的情况,最小值的 AP 的层级能级加 1。两个相邻 AP 对应的 RSSI 值 $\text{RSSI}_{N_{j-1}^{\text{RP}_i}}^i$ 与 $\text{RSSI}_{N_j^{\text{RP}_i}}^i$ 之差在阈值 $\theta \sim 2\theta$ 之间, $\text{RSSI}_{N_j^{\text{RP}_i}}^i$ 层级能级为 $\text{RSSI}_{N_{j-1}^{\text{RP}_i}}^i$ 加 1; 以此类推,当两个相邻 AP 对应的 RSSI 值 $\text{RSSI}_{N_{j-1}^{\text{RP}_i}}^i$ 与 $\text{RSSI}_{N_j^{\text{RP}_i}}^i$ 之差大于阈值 4θ , $\text{RSSI}_{N_j^{\text{RP}_i}}^i$ 层级能级为 $\text{RSSI}_{N_{j-1}^{\text{RP}_i}}^i$ 加 4。对于接收点

$\text{RSSI}_{N_j^{\text{RP}_i}}^i \geq \text{RSSI}_{N_{j+1}^{\text{RP}_i}}^i$ 。则有距离函数 $\text{Dis}(\text{RP}_i, \text{RP}_j)$ 调整为 $\text{Dis}(\text{Seq}_{\text{RP}_i}, \text{Seq}_{\text{RP}_j})$ 来求解式 (1); $\text{Dis}(\text{Seq}_{\text{RP}_i}, \text{Seq}_{\text{RP}_j})$ 的性能将决定定位精度,在第 2 部分介绍如何计算两个 AP 序列的距离。

2 基于层次 Levenshtein 距离的距离算法

Levenshtein 距离这个概念由 Vladimir Levenshtein 在 1965 年提出,通过最少编辑(修改、删除和插入)操作次数来计算字符串之间的相似度 $d(a_i, b_j)$ 。

RP_i , 具体层级能级计算方法见式 (3)。阈值 θ 取值大小通过实验来验证。

$$j = 1; l_j^{\text{RP}_i} = 1; j = 2, 3, \dots, K;$$

$$\text{RSSI}_{N_{j-1}^{\text{RP}_i}}^i - \text{RSSI}_{N_j^{\text{RP}_i}}^i \leq \theta, \max_{k \neq j} \text{RSSI}_{N_k^{\text{RP}_i}}^i - \text{RSSI}_{N_j^{\text{RP}_i}}^i \leq$$

$$2\theta, l_j^{\text{RP}_i} = l_{j-1}^{\text{RP}_i};$$

$$\text{RSSI}_{N_{j-1}^{\text{RP}_i}}^i - \text{RSSI}_{N_j^{\text{RP}_i}}^i \leq \theta, \max_{k \neq j} \text{RSSI}_{N_k^{\text{RP}_i}}^i - \text{RSSI}_{N_j^{\text{RP}_i}}^i >$$

$$2\theta, l_j^{\text{RP}_i} = l_{j-1}^{\text{RP}_i} + 1;$$

$$\theta < \text{RSSI}_{N_{j-1}^{\text{RP}_i}}^i - \text{RSSI}_{N_j^{\text{RP}_i}}^i \leq 2\theta, l_j^{\text{RP}_i} = l_{j-1}^{\text{RP}_i} + 1;$$

$$2\theta < \text{RSSI}_{N_{j-1}^{\text{RP}_i}}^i - \text{RSSI}_{N_j^{\text{RP}_i}}^i \leq 3\theta, l_j^{\text{RP}_i} = l_{j-1}^{\text{RP}_i} + 2;$$

$$3\theta < \text{RSSI}_{N_{j-1}^{\text{RP}_i}}^i - \text{RSSI}_{N_j^{\text{RP}_i}}^i \leq 4\theta, l_j^{\text{RP}_i} = l_{j-1}^{\text{RP}_i} + 3;$$

$$\text{RSSI}_{N_{j-1}^{\text{RP}_i}}^i - \text{RSSI}_{N_j^{\text{RP}_i}}^i > 4\theta, l_j^{\text{RP}_i} = l_{j-1}^{\text{RP}_i} + 4. \quad (3)$$

结论 在一定噪声范围内,接收信号强度之差小于 ω , 区域内 AP 序列顺序无差异。

证明 以实际环境中常用无线信号传输模型——Shadowing 模型来论证。

$$P(d) |_{\text{dBm}} = P(d_0) |_{\text{dBm}} - 10\beta \lg\left(\frac{d}{d_0}\right) + X_{\text{dB}} \quad (4)$$

式中: $P(d)$ 表示接收端到发射机距离为 d 时接收到的信号强度; $P(d_0)$ 表示接收端到发射机参考距离为 d_0 时接收到的信号能量; β 为路径损耗系数, 与环境和建筑物的类型相关联; $X_{\text{dB}} \in [-\text{Max}_{\text{noise}}, \text{Max}_{\text{noise}}]$ 是服从均值为 0 的高斯分布变量。

不失一般性,现有接收端到发射机 A 和 B 距离分别为 d_A 和 d_B , 则有

$$P_A(d_A) \big|_{\text{dBm}} = P_A(d_0) \big|_{\text{dBm}} - 10\beta_A \lg\left(\frac{d_A}{d_0}\right) + X_{\text{dB}}^A \quad (5)$$

$$P_B(d_B) \big|_{\text{dBm}} = P_B(d_0) \big|_{\text{dBm}} - 10\beta_B \lg\left(\frac{d_B}{d_0}\right) + X_{\text{dB}}^B \quad (6)$$

对于同一区域的近邻位置, $\beta_A \approx \beta_B = \beta$, 将式(5)与式(6)相减, 有

$$\left(\frac{P_A(d_A)}{P_A(d_0)}\right)_{\text{dB}} - \left(\frac{P_B(d_B)}{P_B(d_0)}\right)_{\text{dB}} - (X_{\text{dB}}^A - X_{\text{dB}}^B) \approx -10\beta \lg\left(\frac{d_A}{d_B}\right) < 0 \quad (7)$$

则有

$$\left|\left(\frac{P_A(d_A)}{P_A(d_0)}\right)_{\text{dB}} - \left(\frac{P_B(d_B)}{P_B(d_0)}\right)_{\text{dB}}\right| > 2\text{Max}_{\text{noise}} \quad (8)$$

由式(8)可知:在噪声存在的情况下,两接收信号强度差值大于阈值 ω (其大小与环境噪声有关)时,才能区别 d_A 和 d_B 大小。故有两个接收信号强度之差小于 ω , 那么认为由这两个接收信号强度的排序得到的 AP 序列顺序无差异。证毕。

由于 AP 序列中各 AP 在 WiFi 指纹中的能级等级不同,删除和插入操作需要考虑 AP 的能级等级。由结论 1 可知,接收信号强度之差越小,AP 序列 Seq_{RP_i} 顺序差异越小。修改操作距离通过修改前后两个 AP 的 RSSI 变化差异性来表现。对 Levenshtein 距离进行修正,得到基于 Levenshtein 距离的距离函数 $\text{Dis}(\text{Seq}_{\text{RP}_i}, \text{Seq}_{\text{RP}_j})$ 。

对于 $\text{Seq}_{\text{RP}_i} = (N_1^{\text{RP}_i}, N_2^{\text{RP}_i}, \dots, N_K^{\text{RP}_i})$, $\text{Seq}_{\text{RP}_j} = (N_1^{\text{RP}_j}, N_2^{\text{RP}_j}, \dots, N_K^{\text{RP}_j})$:

当 $|\text{Seq}_{\text{RP}_i}| = 0$ 和 $|\text{Seq}_{\text{RP}_j}| = 0$ 时, $\text{Dis}(\text{Seq}_{\text{RP}_i}, \text{Seq}_{\text{RP}_j}) = 0$;

当 $N_m^{\text{RP}_i} = N_n^{\text{RP}_j} (1 \leq m, n \leq K)$ 时,

$$\begin{aligned} \text{Dis}(\text{Seq}_{\text{RP}_i}(m), \text{Seq}_{\text{RP}_j}(n)) &= \\ \min(\text{Dis}(\text{Seq}_{\text{RP}_i}(m-1), \text{Seq}_{\text{RP}_j}(n)) + \alpha, \\ \text{Dis}(\text{Seq}_{\text{RP}_i}(m), \text{Seq}_{\text{RP}_j}(n-1)) + \beta, \end{aligned}$$

$$\text{Dis}(\text{Seq}_{\text{RP}_i}(m-1), \text{Seq}_{\text{RP}_j}(n-1))) ; \quad (9)$$

当 $N_m^{\text{RP}_i} \neq N_n^{\text{RP}_j} (1 \leq m, n \leq K)$ 时,

$$\begin{aligned} \text{Dis}(\text{Seq}_{\text{RP}_i}, \text{Seq}_{\text{RP}_j}) &= \min(\text{Dis}(\text{Seq}_{\text{RP}_i}(m-1), \\ \text{Seq}_{\text{RP}_j}(n)) + \alpha, \text{Dis}(\text{Seq}_{\text{RP}_i}(m), \text{Seq}_{\text{RP}_j}(n-1)) + \beta, \\ \text{Dis}(\text{Seq}_{\text{RP}_i}(m-1), \text{Seq}_{\text{RP}_j}(n-1)) + \delta) \quad (10) \end{aligned}$$

式中: α 受序号 $N_n^{\text{RP}_j}$ 的 AP 约束, 根据 $\text{level}_{\text{RP}_j}$ 来计算,

$\alpha = \frac{1}{l_n^{\text{RP}_j}}$; β 受序号 $N_m^{\text{RP}_i}$ 的 AP 约束, 根据 $\text{level}_{\text{RP}_i}$ 来计

$$\text{算}, \beta = \frac{1}{l_m^{\text{RP}_i}}; \delta = \begin{cases} \gamma, & 0 \leq \gamma < 1 \\ 1, & \gamma \leq 1 \end{cases}$$

$$\gamma = \frac{|\text{RSSI}_{N_m^{\text{RP}_i}}^i - \text{RSSI}_{N_n^{\text{RP}_j}}^j|}{|\text{Avg}(\text{RSSI}_{N_m^{\text{RP}_i} \sim N_n^{\text{RP}_j}}^i) - \text{Avg}(\text{RSSI}_{N_m^{\text{RP}_i} \sim N_n^{\text{RP}_j}}^j)|} \quad (11)$$

式中: $\text{RSSI}_{N_m^{\text{RP}_i}}^i$ 表示在 RP_i 中序号为 $N_m^{\text{RP}_i}$ 的 AP 的 RSSI 值; $\text{RSSI}_{N_n^{\text{RP}_j}}^j$ 表示在 RP_j 中序号为 $N_n^{\text{RP}_j}$ 的 AP 的 RSSI 值; $\text{Avg}(\text{RSSI}_{N_m^{\text{RP}_i} \sim N_n^{\text{RP}_j}}^i)$ 为在 RP_i 的 RSSI 值降序序列中 RP_i 中序号为 $N_m^{\text{RP}_i}$ 的 AP 到与 RP_j 中序号为 $N_n^{\text{RP}_j}$ 的 AP 之间的 RSSI 值的平均值; $\text{Avg}(\text{RSSI}_{N_m^{\text{RP}_i} \sim N_n^{\text{RP}_j}}^j)$ 为在 RP_j 的 RSSI 值降序序列中 RP_j 中序号为 $N_m^{\text{RP}_i}$ 的 AP 到与 RP_j 中序号为 $N_n^{\text{RP}_j}$ 的 AP 之间的 RSSI 值的平均值; γ 表示在 RP_i 、 RP_j 中序号分别为 $N_m^{\text{RP}_i}$ 、 $N_n^{\text{RP}_j}$ 的 AP 对应的 RSSI 值的相对差异量。

3 性能评价

3.1 实验准备

我们在 5 种不同的安卓设备上测试算法性能, 这 5 种设备分别为 Huawei G9、Huawei M2、Samsung s6、Samsung note3、Nexus 5。分别对 3 种不同室内场景(实验室、图书馆和教室)进行实验。实验室的布局分布如图 1(a), 总体布局为 15 m×45 m; 图书馆的布局分布如图 1(b), 总体布局为 78 m×95 m; 教室的各层布局分布如图 1(c), 平面总体布局为 20 m×120 m, 共 3 层。各实验场景的 WiFi 是事先已布置的, 本次实验没有额外布置 AP。考虑环境变化和不同实验操作者对采集信号的影响, 5 个志愿

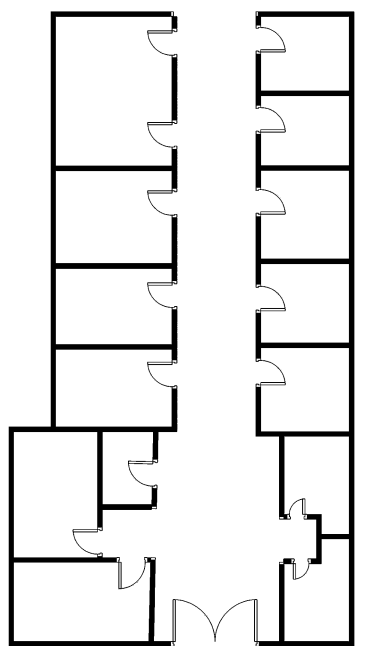
者在4个时间段(间隔一个星期、1个月和3个月)内手持上述5种不同的安卓设备各自独立在不同场景下采集WiFi的RSSI信息,并记录采集点的地理位置。各个采集点采集时间间隔不等,其跨度在10~25 s。为了客观评估定位性能,现场勘查的指纹

数据库选择前3次采集的数据,对同一个位置的多条数据只随机保留一条记录,且保证各数据之间的地理位置距离大于等于1 m;定位测试数据选择第3、4次采集的数据。3个实验场景下4种不同时间的AP数量变化、指纹数据库和测试数据数量如表1所示。

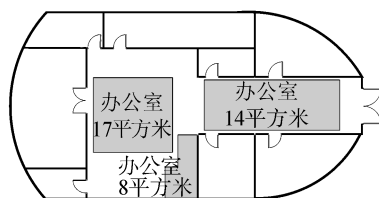
表1 3个实验场景在4种不同时间的AP数量、指纹数据库和测试数据数量

Table 1 Number of AP in three experiment scenarios at four different times and number of fingerprint database, and test data

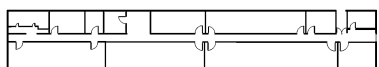
实验场景	第1次 AP 数量	第2次 AP 数量	第3次 AP 数量	第4次 AP 数量	指纹数据库数量	测试数据数量
实验室	19	21	19	22	370	50
图书馆	18	17	18	18	1450	100
教室	26	21	26	24	6500	300



(a) 实验室



(b) 图书馆



(c) 教室

图1 实验场景

Fig.1 Experiment building

3.2 实验结果

本文的定位方法过程是:对于每个接收点通过不同类型的移动设备接收到 $(RSSI_1^i, RSSI_2^i, \dots, RSSI_K^i)$,并标记其空间位置 (x_i, y_i, z_i) 。用HLD算法,对各 RP_i 的RSSI信息降序排列得到AP序列 Seq_{RP_i} ,计算层次能级 $level_{RP_i}$ 。对于定位预测节点 RP_j ,计算 $Dis(Seq_{RP_i}, Seq_{RP_j})$ 后获取 K 个近邻,采用WKNN算法进行预测定位。WKNN算法中的权重设置为 $1/Dis(Seq_{RP_i}, Seq_{RP_j})$ 。其定位方法过程涉及两个参数:能量层级参数 θ 和WKNN算法的近邻数目 K 。

在实验中,本文提出的方法与3种经典室内定位方法进行对比。第一种方法是文献[5]提出的RADAR算法。因为不同设备获取的RSSI差异较大,无法用欧氏距离来计算各RP的相似度,在实验中将RADAR算法中欧氏距离替换为余弦距离计算各RP的相似度;第二种方法是文献[19]提出的FreeLoc算法,该算法的基本思想是通过Key-Value机制构造指纹来容忍诸如设备多样性和信号变化的环境动态;第三种方法是文献[23]提出的HED算法,该算法利用容错序列匹配策略来实现动态环境下WiFi指纹定位。采用平均定位误差和累计定位误差两种误差度量来对比4种室内定位方法的定位性能优越程度。

1) 能量层级参数 θ 、近邻数目 K 选择

为了测试能量层级参数 θ 、近邻数目 K 对定位精度的影响,分别设定 $\theta = 3, 4, 4.5, 5, 5.5, 6, 7$ 和 $K =$

2,3,4,5,6,对比在三种场景下的定位测试。对于不同场景,通过两个参数交叉验证。当 $\theta = 5$ 时,算法达到最优。 $K=5$ 时参数 θ 对预测定位的影响如图2所示。结合分析图3~5的累计定位误差和表2的平均定位误差: $K=5$ 时,在实验室和图书馆场景下,其算法达到最优; $K=4$ 时,在教室场景下,其算法达到最优。其原因是,实验室和图书馆场景是二维位置数据信息,教室的数据是立体地理位置。

表2 HLD 算法在3种实验场景下不同 K 值情况的平均定位误差
Table 2 Average error of HLD algorithm under different K in three experiment scenarios m

场景	$K=2$	$K=3$	$K=4$	$K=5$	$K=6$
实验室	3.08	2.40	1.88	1.46	3.14
图书馆	2.37	2.31	2.01	1.35	2.57
教室	2.74	2.27	1.45	1.58	2.86

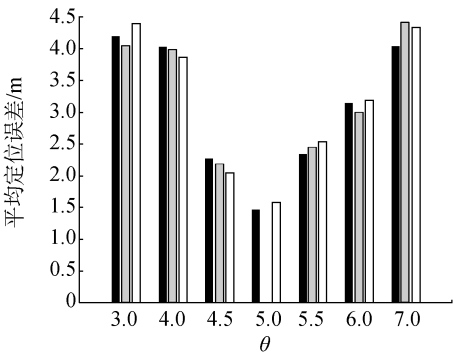


图2 $K=5$ 时不同参数 θ 平均定位误差对比

Fig.2 The average error comparison under different θ parameters in $K=5$

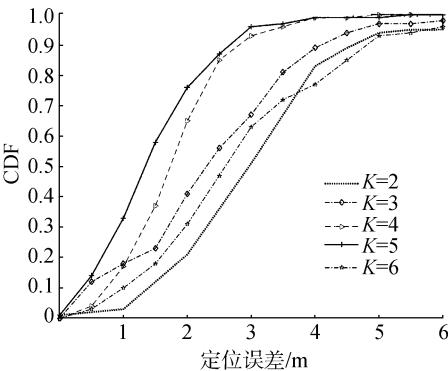


图3 HLD 算法在实验室环境下的不同 K 值的累计定位误差比较

Fig.3 Cumulative localization error comparison of HLD algorithm under different K values in laboratory

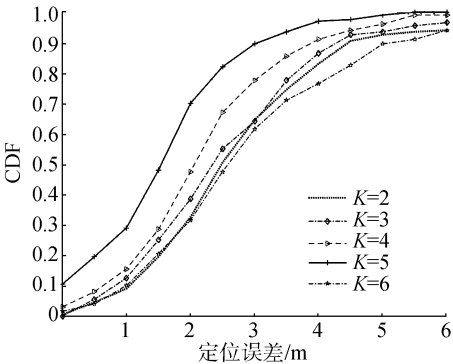


图4 HLD 算法在图书馆环境下的不同 K 值的累计定位误差比较

Fig.4 Cumulative localization error comparison of HLD algorithm under different K values in library

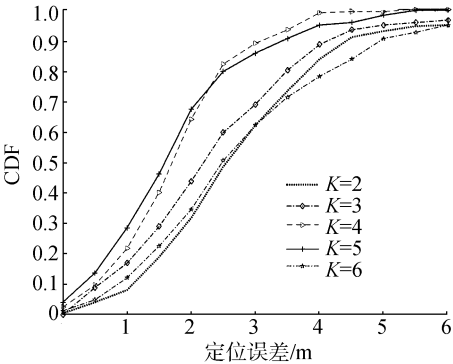


图5 HLD 算法在3层教室环境下的不同 K 值的累计定位误差比较

Fig.5 Cumulative localization error comparison of HLD algorithm under different K values in three floors classroom

2) 与 RADAR、FreeLoc 和 HED 算法性能比较

在实验室和图书馆场景下,HED 算法的 K 取值 5;在教室场景下,HED 算法的 K 取值 4。RADAR 算法中 K 值和 HED 算法相同。在 3 种实验场景下 4 种方法累计定位误差对比如图 6~8 所示,表 3 描述了 3 种实验场景下 4 种方法平均定位误差对比。RADAR 对于平面环境下的定位误差在 4~6 m,2 m 内误差准确度在 20% 以下,基本达不到定位的效果;对三维环境下的定位效果更差。FreeLoc、HED 对于平面环境下的定位误差在 2~4 m,3 m 内误差准确度在 50%~60%,但在三维环境下的定位效果 3 m 内误差准确精度下降到 40% 左右。HLD 算法整体的平均误差为 1.5 m,2 m 内误差准确度为 70%~80%,对空间维度敏感度较低。

在定位过程中,与 FreeLoc、HED 算法执行效率进行比较,HLD 算法时空复杂度比 FreeLoc 低 40%,比 HED 空间复杂度高 10%,时间复杂度相当。

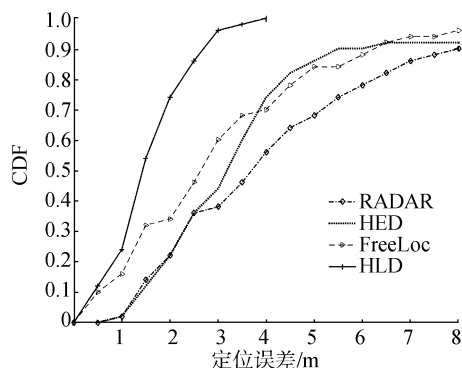


图6 实验室环境下的4种方法累计定位误差比较

Fig.6 Cumulative localization error comparison of four methods in laboratory

表3 3种实验场景下4种方法的平均误差对比

Tab 3 The average error comparison of four methods in three experiment scenarios m

场景	RADAR	FreeLoc	HED	HLD
实验室	4.25	3.11	3.57	1.46
图书馆	5.70	2.97	3.27	1.35
教室	9.56	5.02	6.78	1.45

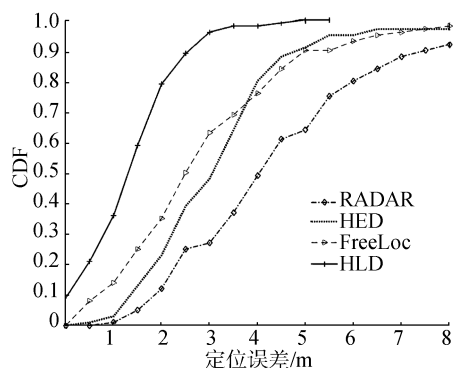


图7 图书馆环境下的4种方法累计定位误差比较

Fig.7 Cumulative localization error comparison of four methods in library

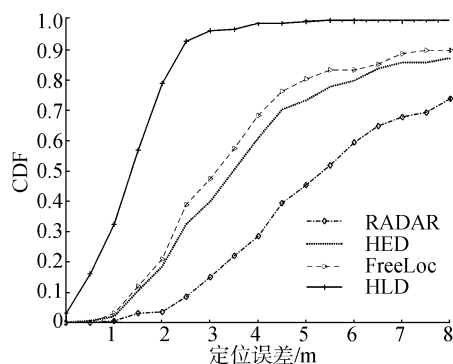


图8 3层教室环境下的4种方法累计定位误差比较

Fig.8 Cumulative localization error comparison of four methods in three floors classroom

4 总结

本文讨论了设备异构和室内环境变化情形下的室内定位问题,提出了基于层次 Levenshtein 距离的 WiFi 指纹距离计算算法。在无校准情形下,将不同移动设备采集的 RSSI 信息转化为 AP 序列,以解决不同设备获取的 RSSI 信息的量纲一致性问题。根据 AP 对应的 RSSI 值的差异性计算其层次能级,结合 Levenshtein 距离计算 WiFi 指纹之间的距离,提供一种适应异构设备和动态环境下计算各 RP 间指纹距离的鲁棒性方法。对于需定位的 WiFi 指纹 RSSI 信息,采用 WKNN 算法进行预测定位。实验中,为了验证算法的鲁棒性和有效性,在 3 种不同类型的室内环境中用 5 种不同的移动设备来采集 WiFi 的 RSSI 信息,其定位的平均精度达 1.5 m,2 m 内误差准确度在 70%~80%,对空间维度敏感度较低。

参考文献:

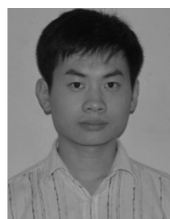
- [1] GU Y, LO A, NIEMEGEREERS I. A survey of indoor positioning systems for wireless personal networks [J]. IEEE commun. surveys and tutorials, 2009,11(1): 13-32.
- [2] HARLE R. A survey of indoor inertial positioning systems for pedestrians[J]. IEEE commun. surveys & tutorials, 2013, 15(3): 1281-1293.
- [3] SUBBU K P. Analysis and status quo of smart-phone-based indoor localization systems [J]. IEEE wireless commun, 2014,21(4): 106-112.
- [4] 石柯,陈洪生,张仁同.一种基于支持向量回归的 802.11 无线室内定位方法 [J]. 软件学报, 2014, 25 (11): 2636-2651.
SHI Ke, CHEN Hongsheng, ZHANG Rentong. Indoor location method based on support vector regression in 802.11 wireless environments[J]. Journal of software, 2014, 25 (11): 2636-2651.
- [5] BAHL P, PADMANABHAN V. RADAR: an in-building rf-based user location and tracking system[C]//Proc. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Tel Aviv, Israel, 2000:775-784.
- [6] YOUSSEF M, AGRAWALA A . The horus wlan location determination system[C]//Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services, Washington, USA, 2005: 205-218.
- [7] WANG B, CHEN Q, YANG L T, et al. Indoor smartphone localization via fingerprint crowdsourcing: challenges and approaches [J]. IEEE wireless communication, 2016(6):

- 82-89.
- [8] TSUI A W, CHUANG Y H, CHU H H. Unsupervised learning for solving rss hardware variance problem in wifi localization[J]. Mobile networks and applications, 2009, 14 (5): 677-691.
- [9] CHENG H, WANG F, TAO R, et al. Clustering algorithms research for device-clustering localization[C]//2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Sydney, Australia, 2012:1-7.
- [10] MAHTAB HOSSAIN A, JIN Y, SOH W S, et al. SSD: a robust RF location fingerprint addressing mobile devices heterogeneity[J]. IEEE transactions on mobile computing, 2013, 12(1): 65-77.
- [11] PARK J G, CURTIS D, TELLER S, et al. Implications of device diversity for organic localization[C]//The 30th IEEE International Conference on Computer Communications, Shanghai, China, 2011:3182-3190.
- [12] FIGUERA C, ROJO-LVAREZ J L, MORA-JIMNEZ I, et al. Time-space sampling and mobile device calibration for wifi indoor location systems [J]. IEEE transactions on mobile computing, 2011, 10(7): 913-926.
- [13] HAEBERLEN A, FLANNERY E, LADD A M, et al. Practical robust localization over large-scale 802.11 wireless networks[C]//Proceedings of the 10th Annual International Conference on Mobile Computing and Networking, Philadelphia, USA, 2004: 70-84.
- [14] KJIRGAARD M B. Indoor location fingerprinting with heterogeneous clients[J]. Pervasive and mobile computing, 2011, 7(1): 31-43.
- [15] DELLA ROSA F, LEPPAKOSKI H, BIANCULLO S, et al. Ad-hoc networks aiding indoor calibrations of heterogeneous devices for fingerprinting applications[C]//2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Zurich, Switzerland, 2010: 1-6.
- [16] CHEN L H, WU E H K, JIN M H, et al. Homogeneous features utilization to address the device heterogeneity problem in fingerprint localization [J]. IEEE sensors journal, 2014, 14(4): 998-1005.
- [17] ZOU H, LU X, JIANG H, et al. A fast and precise indoor localization algorithm based on an online sequential extreme learning machine[J]. Sensors, 2015, 15(1): 1804-1824.
- [18] LYMBEROPOULOS D, LIU J, YANG X, et al. A realistic evaluation and comparison of indoor location technologies: experiences and lessons learned[C] //Proceedings of the 14th International Conference on Information Processing in Sensor Networks, Catania, Italy, ACM, 2015: 178-189.
- [19] YANG S. Freeloc: calibration-free crowdsourced indoor localization[C]//The 32th IEEE International Conference on Computer Communications, Turin, Italy, 2013: 2481-2489.
- [20] JIANG Y. Ariel: automatic wi-fi based room fingerprinting for indoor localization[C]//Proc ACM Conf. Ubiquitous Computing, Pittsburgh, Pennsylvania, United States, 2012: 441-50.
- [21] 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.
- [22] ZOU H, HUANG B, LU X, et al. Standardizing location fingerprints across heterogeneous mobile devices for indoor localization [C]//IEEE Wireless Communications and Networking Conference (WCNC 2016). Doha, Qatar, 2016: 1-6.
- [23] GU Y, CHEN M, REN F, et al. HED: handling environmental dynamics in indoor wifi fingerprint localization[C]// IEEE Wireless Communications and Networking Conference (WCNC 2016), Doha, Qatar, 2016: 5-10.

作者简介:



何富贵,男,1982年生,副教授,主要研究方向为移动计算、室内定位和粒计算。发表学术论文10余篇。



杨铮,男,1983年生,副教授,博士生导师,研究方向为无线网络与移动计算,包括传感网、Mesh网络、室内定位、群智感知等。发表论文60余篇,其中CCF推荐A类论文40余篇;出版中、英文学术专著各1部。获得国家自然科学奖二等奖。



吴陈沐,男,1989年生,博士,研究方向为无线网络与移动计算,包括室内定位、群智感知等。