

一种基于多终端动态协同的室内定位方法*

杨 华^{1,2}, 刘军发², 陈益强^{1,2}

(1. 湘潭大学 信息工程学院, 湖南 湘潭 411105; 2. 中国科学院计算技术研究所, 北京 100190)

摘 要: 由于 Wi-Fi 环境的高动态性, 当标定样本类稀少或者逐渐失去其原始价值时, 将严重影响定位精度; 若重新标定样本, 又会带来大量人力物力投入, 不利于实际应用。针对这一问题, 提出了一种基于多终端动态协同的室内定位方法——OC-ELM。该方法利用环境中多个终端提供的 Wi-Fi 信息动态协同定位, 并在每轮定位结束后及时更新样本库。实验结果表明此方法不仅能实现高精度定位, 更重要的是避免了重复的标定工作, 提高了实际应用能力。

关键词: 高动态环境; Wi-Fi; 室内定位; 动态协同; 实时更新

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1001-3695(2012)07-2457-04

doi:10.3969/j.issn.1001-3695.2012.07.015

Dynamic collaborative indoor locating method based on multiple terminals

YANG Hua^{1,2}, LIU Jun-fa², CHEN Yi-qiang^{1,2}

(1. College of Information Engineering, Xiangtan University, Xiangtan Hunan 411105, China; 2. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Due to high dynamic of Wi-Fi environment, positioning accuracy will be affected under the situation of fewer samples of calibration or their original value lost. It will aggravate workload and be disadvantage of practical application if get new calibration samples. In order to solve this problem, this paper proposed a dynamic collaborative indoor locating method based on multiple terminals. The method used Wi-Fi information which provided by terminals for co-location of dynamic, and updated sample after locating. Experimental results show that the method not only realize high precision positioning, but also can avoid repetitive calibration, beneficial to actual application.

Key words: high dynamic environment; Wi-Fi; indoor positioning; dynamic collaborative; real-time update

0 引言

在现代信息化生活中, 人们对室内定位的需求日益增多, 定位应用也随着移动互联网时代的到来越发普及, 但由于环境的多样性与多变性, 对系统定位精度以及实际应用能力要求也就越来越高。近十年来, 涌现出了许多定位系统方案, 如 GPS 定位^[1]、GSM 定位^[2]、红外线定位^[3]、超声波定位^[4]、RFID 定位^[4]、蓝牙定位^[5]等, 这些系统都存在一定的局限。GPS、GSM 只适用于室外; 蓝牙、RFID 探测范围较小; 超声波与红外线方式易受介质遮挡等。目前各类移动终端中逐渐开始内置无线网卡, 同时无线的城市理念使得环境中出现足够多的无线接入点(AP), 这都为 Wi-Fi 定位技术提供了便利。基于 Wi-Fi 的定位方法从技术上主要分为基于信号到达时间(TOA)^[6]、基于信号到达时间差(TDOA)^[7]、基于信号到达角度(AOA)^[8]和基于信号强度的变化(RSS)等。目前最广泛使用的是基于 RSS 的方法, 与基于 TOA 和 AOA 定位相比, 它不需要添加额外的设备来进行精确的时间同步和角度测量, 能充分利用现有的无线网设施来降低成本。从方法上主要分为基于传播模型和基于指纹模型两种。基于传播模型的方法^[9]计算简单, 但

是由于室内环境复杂, 难以准确建模, 因而定位精度较低; 基于指纹模型的方法, 常见的有 KNN^[10]、SVM^[11]、MR^[12]等, 虽能实现较高定位精度, 但都没有过多地考虑实际应用, 忽略了样本集的重利用率。

因此, 目前室内 Wi-Fi 定位技术主要面临两方面的挑战: a) Wi-Fi 环境的多样与多变性, 直接影响着定位精度; b) 尽管大部分指纹定位模型方法可以很大程度上提高定位精度, 但时间迁移性很差, 即环境随时间推移而变化后, 定位模型性能将急速下降, 无法真正投入到实际应用中。本文针对这一问题, 提出了一种基于多终端动态协同的室内定位方法——OC-ELM, 该方法充分利用环境中多个终端提供的 Wi-Fi 信息, 辅助已标定样本类对用户进行位置跟踪, 并针对当前的定位响应结果和其对应的 Wi-Fi 信息组合成新样本类更新到样本库中。

1 ELM 算法

ELM(extreme learning machine)算法^[13]是神经网络的一部分, 确切地说, 它是一种单隐层的前馈神经网络。对于一个输入向量 $\mathbf{x} \in \mathbb{R}^n$, 其输出定义如式(1)所示, 其中 N 是隐层节点

收稿日期: 2011-12-15; **修回日期:** 2012-02-09 **基金项目:** 国家“863”计划资助项目(2007AA01Z305); 国家自然科学基金面上项目(61173066)

作者简介: 杨华(1986-), 女, 湖南常德人, 硕士, 主要研究方向为室内定位、机器学习(yanghua860709@163.com); 刘军发(1973-), 男, 助理研究员, 博士, 主要研究方向为普适计算、人机交互; 陈益强(1973-), 男, 研究员, 博士, 主要研究方向为普适计算、人机交互。

数, $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ 是联系隐层节点和输出节点的权值向量; $G(a_i, b_i, x)$ 是第 i 个隐层节点的输出; a_i 是联系输入节点和第 i 个隐层节点的权值向量; b_i 是第 i 个隐层节点的阈值。

$$\hat{f}_N(x) = \sum_{i=1}^N \beta_i G(a_i, b_i, x) \quad a_i \in R^n, \beta_i \in R^m \quad (1)$$

对于训练数据集 $\{(x_i, t_i) | i = 1, 2, \dots, N\}$, $x_i \in R^n$ 是输入信号的特征向量, $t_i \in R^m$ 是目标的物理位置。期望获得 $\hat{N}$ 个隐层的单隐层前馈神经网络零误差逼近这 N 个样本, 也就是说对于 β_i, a_i, b_i, N 个样本满足 $H\beta = T$, 其中:

$$H = \begin{bmatrix} G(a_1, b_1, x_1) & \dots & G(a_{\hat{N}}, b_{\hat{N}}, x_1) \\ \vdots & & \vdots \\ G(a_1, b_1, x_N) & \dots & G(a_{\hat{N}}, b_{\hat{N}}, x_N) \end{bmatrix}_{N \times \hat{N}} \quad (2)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\hat{N}}^T \end{bmatrix}_{\hat{N} \times m} \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m}$$

针对式(2)作进一步优化, 即等价于解决以下问题:

$$\arg \min_{\beta} (\|H\beta - T\|) \quad (3)$$

为了有效地利用非标定样本, 引入图的拉普拉斯算子, 如式(4)所示, 其中第一项为标定样本带来的误差, 第二项为基于图的拉普拉斯算子复杂模型带来的误差。

$$\arg \min_{\beta} \frac{1}{2} (\|f - T\|^2 + \lambda f^T L f) \quad (4)$$

再由 $f = H\beta$ 可得出:

$$\arg \min_{\beta} l(\beta) = \arg \min_{\beta} \frac{1}{2} \{ \|JH\beta - T\|^2 + \lambda (H\beta)^T L H\beta \} \quad (5)$$

H 可由式(2)求得, 令 $J = \text{diag}(1, 1, \dots, 0, 0)$, L 个 1, 剩下的都是 0。为了求解式(5)的最优解, 对其求导:

$$\frac{\partial l}{\partial \beta} = 0 \Rightarrow (JH\beta - T)^T JH + \lambda (H\beta)^T L H = 0 \quad (6)$$

进而得到:

$$\beta = ((J + \lambda L^T)H)^+ J T \quad (7)$$

其中, λ 为 0 表示非标记数据被忽略。假如标定数据 $\{(x_i, t_i) | x_i \in R^n, t_i \in R^m, i = 1, 2, \dots, l\}$, 非标定数据 $\{x_i | x_i \in R^n, i = 1, 2, \dots, u\}$, 隐层节点数 $\hat{N}$, 权值 λ , $G(a_i, b_i, x) = g(b_i \| x - a_i \|)$, 则整个过程有五步: a) 随机分配 a_i, b_i 和隐层节点数 $\hat{N}$; b) 计算 H (维度为 $m \times \hat{N}$), 其中 $m = 1 + u$; c) 计算图的拉普拉斯算子 L , $L = D - W$, $D_{ii} = \sum_{j=1}^m W_{ij}$, $W_{ij} = e^{-\frac{\|x_i - x_j\|^2}{2\delta^2}}$ ($i, j = 1, 2, \dots, m$), L, D, W 的维数为 $(m \times m)$; d) 计算 β ; e) 利用模型 $f = H\beta$ 预测。

2 OC-ELM 算法

2.1 框架介绍

本文提出方法的具体原理如图 1 所示。

在无位置指纹的环境中新标定一批样本, 和当前环境中其他终端提供的非标定样本一起构成训练数据库, 进行相应的数据预处理工作后, 训练位置指纹预测模型。当用户进入该环境后, 请求位置服务的同时提交测试样本, 结合之前训练的位置预测模型计算用户当前的位置信息。并将当前返回的位置和对应的测试样本数据一起组合成新的标定样本更新到样本库

中。随着时间的推移, 训练数据库也会逐渐淘汰掉失效样本。相比以往传统的指纹模型方法, 充分利用了环境中动态终端提供的信息, 进一步提高定位精度。另外及时更新训练数据库, 免去了重复标定工作。

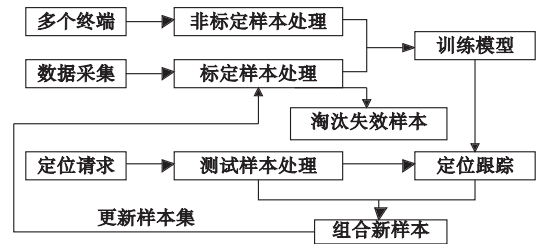


图1 方法原理

2.2 定位过程

OC-ELM 算法主要分为离线训练和在线定位两个阶段。

1) 离线训练阶段

(1) 准备训练样本。假如环境中存在 n 个 AP, 训练数据来源有三个。

① 将此环境区域规律划分为 m 个网格, 在每个网格内进行数据采集工作, 构成 m 个样本类, 每个样本类里存在 p 条样本数据, 这一批数据属于标定样本。存储格式为 $(\text{time}, \text{MAC_APi}, \text{RSSI_APi}, X_label, Y_label)$ ($i = 1, \dots, n$)。其中, time 为当前时间, MAC_APi 表示第 i 个 AP 的 MAC 地址, RSSI_APi 表示第 i 个 AP 的信号强度值, X_label 表示参考点的 X 坐标, Y_label 表示参考点的 Y 坐标。

② 来自环境中的其他终端, 他们请求位置服务时, 提交的当前环境中的 Wi-Fi 信息。存储格式为 $(\text{time}, \text{MAC_APi}, \text{RSSI_APi})$, 此类数据为非标定数据。

③ 测试阶段组合的新样本, 用来更新训练样本库, 存储格式为 $(\text{time}, \text{MAC_APi}, \text{RSSI_APi}, X_label, Y_label)$ 。

(2) 淘汰失效样本。随着时间推移、环境不断的变化, 有些样本类将会慢慢失去原有的价值。及时淘汰掉失效样本可以提高定位精度。主要根据 time 的值来决定, 离目前越久的样本价值越小, 离现在越近的样本价值越大。

(3) 数据预处理。首先按照 MAC 地址将每个样本类的 p 条序列按相同的规律排列整齐; 然后, 由于高动态环境下噪声数据时有出现, 直接影响着定位精度, 为此利用滑动滤波求均值的方法对数据进行了噪声处理。

(4) 特征提取。对处理好的数据一共提取了两种类型特征: ① 各参考点对应的所有 AP 信号强度; ② 相邻 AP 信号强度差值。格式为 $(\text{PRO_RSSI_APi}, \text{PRO_RSSI_APi} - \text{PRO_RSSI_APi} - 1, X_label, Y_label)$ 、 $(\text{PRO_RSSI_APi}, \text{PRO_RSSI_APi} - \text{PRO_RSSI_APi} - 1)$ 。

(5) 训练模型。前面工作完成后, 用上面介绍的 ELM 方法将标定样本和非标定样本组合在一起训练位置模型, 供在线定位时使用。

2) 在线定位阶段

(1) 准备测试数据。用户请求位置服务时, 会自动扫描当前环境中的 Wi-Fi 信息, 并按照训练数据的格式保存并提交。

(2) 数据预处理与特征提取。根据训练数据的规律排列整齐并用滑动滤波求均值的方法对数据进行噪声处理。预处理完成后, 按照训练数据提取特征的方法对测试数据进行相应

的特征提取。

(3)定位跟踪。由第一阶段训练出的位置模型可知,模型和位置之间存在一一对应的映射关系,将测试序列输入此模型后,便可快速准确地计算出终端的位置。

(4)组合新样本。每轮定位结束后,测试数据序列与系统返回给终端的位置组合成新的标定样本,以此更新训练样本库。

3 实验与分析

3.1 数据描述

为了验证与评价 OC-ELM 算法的有效性,本文选择了中国科学院计算技术研究所一个开放的工作区进行实验,环境区域大小为 $15\text{ m} \times 24\text{ m}$,如图2所示。首先对整个区域进行网格划分,大小为 $1\text{ m} \times 1\text{ m}$;然后在每个网格内进行训练数据采集,每个点采集500次,数据采集的频率为 5 s/次 。存储的内容及格式为:当前时间、搜索到的AP数目、所有AP信息(其中每个AP存储它的MAC地址、信号强度和AP名称)以及坐标位置。由于环境的多变与多样性,每个位置一般可探测到8~15个AP不等。用户测试时,环境中其他终端提交的Wi-Fi信息归一化到训练库中,作为OC-ELM算法的输入进行位置预测模型的训练。每进行一次定位请求与响应后,及时更新一次训练数据库。

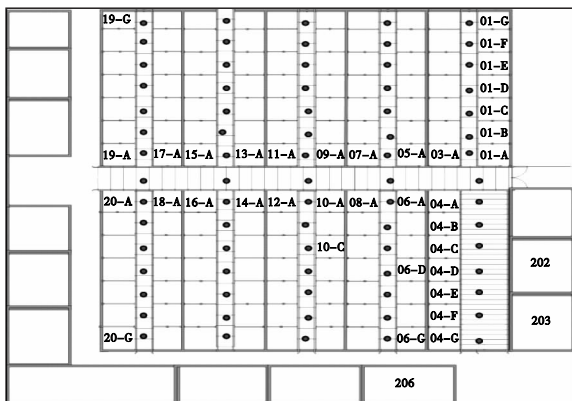


图2 实验场景描述

实验中的服务器为安装服务端程序的PC机,搭建在中科院计算所大楼八层。客户端是基于Windows Mobile 6.1系统的MOTO 3100智能手机;Wi-Fi信号源是环境中现存的。

3.2 问题有效性验证

a)列举了如下可能的高动态环境,分别对其中比较典型的情况进行了实验分析。

(a)终端离AP距离不同,Wi-Fi的信号强度变化如图3(a)所示,离AP越远,信号强度越小;离AP越近,信号强度越大。

(b)随时间等间隔推移,Wi-Fi信号强度的变化如图3(b)所示,同一个地点扫描到的同一个AP的信号也存在波动,特别是刚开始20s的数据稳定性较差。

(c)终端在AP的不同方向上,Wi-Fi信号强度的变化如图3(c)所示。正对着AP信号强度最大,背对着AP信号强度最小。

(d)不同温度、湿度时Wi-Fi信号强度的变化如图3(d)所示。低温对信号干扰小于高温。

b)环境中Wi-Fi也是动态变化的(如AP更换、个数加减、位置变更等)。表1记录的是数据库中部分AP的信号变化情况

况,如-100表示此刻没有搜到该AP, T 表示时间值。随着时间的推移,不断有新的AP加入或者旧的AP去除。

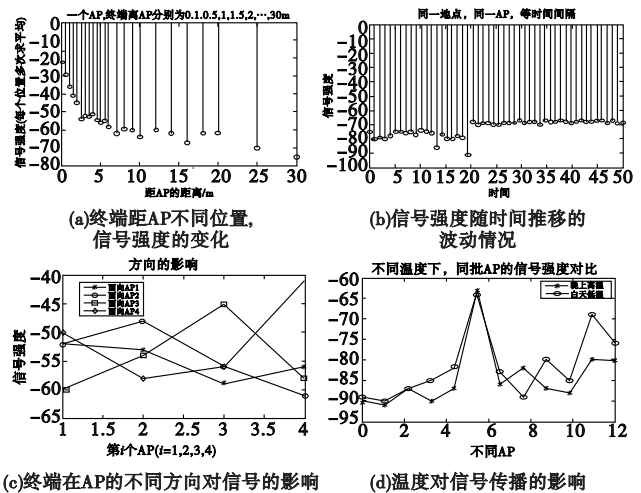


图3 高动态环境对信号传播的影响

表1 AP随时间推移的变化

T	AP ₁	AP ₂	AP ₃	AP ₄	AP ₅	AP ₆	AP ₇	...
t_1	-60	-65	-54	-71	-100	-100	-100	...
t_2	-65	-59	-55	-70	-100	-100	-100	...
t_3	-61	-62	-57	-100	-100	-100	-100	...
t_4	-60	-64	-55	-100	-50	-100	-100	...
t_5	-60	-64	-100	-100	-68	-100	-100	...
t_6	-61	-58	-100	-100	-65	-71	-100	...
t_7	-100	-57	-100	-100	-66	-70	-100	...
t_8	-100	-56	-100	-100	-66	-69	-55	...
...	...	...	...	...	...	...	...	...

结论:高动态环境严重影响Wi-Fi信号的传播,最终波及到定位精度;Wi-Fi的不定期变化限制了样本库的可复用性。所以展开对高动态环境下高精度定位方法的研究。

3.3 实验结果比较

在真实的无线环境中,高动态无处不在,严重影响定位精度以及样本库的复用性,这一直是定位技术投入实际应用中面临的挑战,通过本文提出的方法很好地解决了此问题。

为了验证本文方法的实际性能,将其与SVM方法进行了比较。同样的环境,同批数据源,性能相当的终端,按照两种方法进行了对比模拟实验,实验结果如图4所示,由于OC-ELM方法引入了环境中其他终端提供的Wi-Fi信息,对原有的定位模型进行了优化逼近,所以定位精度优于传统的SVM方法,且5m误差范围内能实现91.5%的精度。

如图5为随时间推移,当环境不断发生变化后两种方法在定位性能上的比较。由于Wi-Fi不定期的变化,样本库中原有样本慢慢失去了其原有价值,若重新标定样本,不利于定位技术的实际操作;若不重新标定样本,实验结果显示:传统的SVM方法定位精度必受影响。由于本文的OC-ELM方法采用了数据集更新机制,定位精度随时间推移不断提高,最终达到稳定。

4 结束语

为克服传统方法的不足,本文提出了一种基于多终端动态

协同的室内定位方法——OC-ELM,该方法利用终端提供的动态非标定样本信息,辅助已标定样本类对用户进行位置跟踪,并实时更新样本库。实验结果表明,此方法不仅能适应高动态环境,提高定位精度,还能实现样本库的重复利用,易于实际应用。

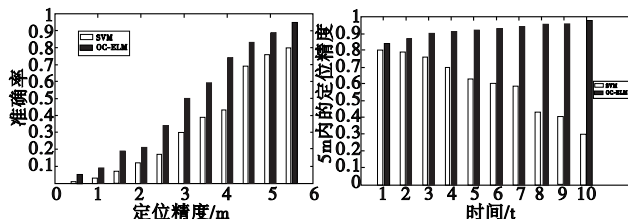


图4 不同距离定位精度比较

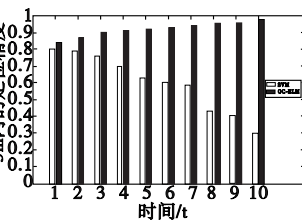


图5 不同时刻定位精度比较

但随着系统的长期使用,训练数据库规模不断扩大,所以下一步工作重点是:a)分析多终端 Wi-Fi 信息之间的相关性,尽量减少冗余信息对定位的负作用,使训练数据呈爆炸式增长时,降低在线更新模型定位复杂性并提高实时响应速率;b)当 Wi-Fi 环境比较稳定,而新标定数据过度更新时,解决在线更新定位模型可能导致的过学习问题。

参考文献:

- [1] SWANEPOEL L H, DALERRUM F, Van HOVEN W. Factors affecting location failure of GPS collars fitted to African leopards (panthera pardus) [J]. *South African Journal of Wildlife Research*, 2010, 40 (1): 10-15.
- [2] DELIGIANNIS N, LOUVROS S. Hybrid TOA-AOA location positioning techniques in GSM networks [J]. *Wireless Personal Communications*, 2010, 54 (2): 1-13.
- [3] KOYUNCU H, YANG Shang-hua. A survey of indoor positioning and object locating systems [J]. *International Journal of Computer Science and Network Security*, 2010, 10 (5): 121-128.
- [4] PRIYANTHA N B, CHAKRABORTY A, BALAKRISHNAN H. The cricket location-support system [C] // Proc of the 6th ACM/IEEE International Conference on Mobile Computing and Networking. New York: ACM Press, 2000: 32-43.
- [5] 顾永超, 一种基于蓝牙技术的室内定位子系统的设计与实现 [D]. 北京: 北京大学, 2009.
- [6] KANAAN M, PAHLAVAN K. A comparison of wireless geolocation algorithms in the indoor environment [C] // Proc of IEEE Wireless Communications and Networking Conference. 2004: 177-182.
- [7] YANG Le, HO K C. An approximately efficient TDOA localization algorithm in closed-form for locating multiple disjoint sources with erroneous sensor positions [J]. *IEEE Trans on Signal Processing*, 2009, 57 (12): 4598-4615.
- [8] NICULESCU D, NATH B. Ad hoc positioning system (ASP) using AOA [C] // Proc of the 22nd Annual Joint Conference on IEEE Computer and Communications. 2003: 1734-1743.
- [9] NI Wei, WANG Zong-xin. Indoor location algorithm based on the measurement of the received signal strength [J]. *Frontiers of Electrical and Electronic Engineering in China*, 2006, 1 (1): 48-5.
- [10] BAHL P, PADMANABHAN V N. Enhancements of the RADAR user location and tracking system, MSR-TR-2000-12 [R]. Cambridge: Microsoft Research, 2000.
- [11] BRUNATO M, BATTISTI R. Statistical learning theory for location fingerprinting in wireless LANs [J]. *Computer Networks*, 2005, 47 (6): 825-845.
- [12] PAN J J, YANG Qiang. Co-localization from labeled and unlabeled data using graph laplacian [C] // Proc of the 20th International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann Publishers, 2007: 2166-2171.
- [13] LIU Jun-fa, CHEN Yi-qiang, LIU Ming-jie, et al. SELM: semi-supervised ELM with application in sparse calibrated location estimation [J]. *Neurocomputing*, 2011, 74 (16): 2566-2572.