

基于高斯混合模型的无线局域网定位算法

程远国, 耿伯英

(海军工程大学电子工程学院, 武汉 430033)

摘 要: 目标定位是无线局域网提供位置服务的基础。该文围绕无线局域网目标定位中的定位精度和实时性 2 个 QoS 指标, 对基于 RSSI 的定位技术进行研究, 提出一种基于高斯混合模型的无线局域网定位算法。该算法由离线训练和在线定位 2 个阶段组成, 采用 GMM 对 RSSI 进行建模, 降低了系统定位误差, 并减少了定位阶段的运算量, 提高了定位的实时性。仿真实验结果表明, 该算法具有较好的定位精度和实时性。

关键词: 无线局域网; 接收信号强度指示; 高斯混合模型; 定位算法

WLAN Location Algorithm Based on Gauss Mixture Model

CHENG Yuan-guo, GENG Bo-ying

(College of Electrical Engineering, Navy Engineering University, Wuhan 430033)

【Abstract】 User location estimation is the precondition of the services in Wireless Local Area Network(WLAN). Aiming to improve both QoS aspects including accuracy and real time of WLAN, this paper conducts the research on localization technology based on Received Signal Strength Indicator(RSSI), and proposes an algorithm for WLAN location estimation based on Gauss Mixture Model(GMM), which consists of an offline training phase and a real time localization phase. The RSSI is also modeled by GMM, which decreases the error and the computation, and improves the real time of WLAN user localization. Simulation experimental results show this algorithm has better performance in both accuracy and real time.

【Key words】 Wireless Local Area Network(WLAN); Received Signal Strength Indicator(RSSI); Gauss Mixture Model(GMM); location algorithm

1 概述

自从 802.11 无线局域网(Wireless Local Area Network, WLAN)标准问世以来, WLAN 的各种应用得到迅速发展, 很多应用都基于 WLAN 提供的位置服务^[1]。WLAN 提供位置服务的一个基本前提是必须通过无线定位技术获得移动用户(终端)的当前地理位置信息。现有无线定位技术主要基于 TOA(Time of Arrival), TDOA(Time Difference of Arrival), AOA (Angle of Arrival)和 Received Signal Strength Indicator (RSSI)等定位技术^[2]。TOA 和 TDOA 要求系统能够精确同步, 且由于 WLAN 中 AP 覆盖范围有限, RF 信号传输时延可以忽略不计, 因此无法采用 TOA 或 TDOA 的方法定位。AOA 方法受限于非视线(No Line of Sight, NLOS) 问题对信号传播的影响, 且需要增加额外的硬件设备用于角度测量。基于 RSSI 的方法不需要额外的硬件, 适用范围广、成本低。可通过建立比较准确的 RSS 模型克服定位误差较大的缺点, 不失为一种可行的 WLAN 定位方法。

在基于 RSSI 的 WLAN 定位方法中, 微软的 Radar 系统利用信号传播的经验模型粗略估算基站与使用者的位置, 系统定位简单^[3]。尽管使用 k -NN(k -Nearest Neighbor)方法提高定位精度, 但总体说来该系统定位精度较低。为提高定位精度, 文献[4]采用高斯概率方法对 RSSI 建模, 并将位置估计过程看成机器学习问题。该方法不需要额外的硬件支持, 能够在不同环境下使用, 但在 AP 数量较少时, 系统定位误差较大, 当 AP 数目增加时, 系统的计算工作量较大, 实时性较差。在对接收信号进行概率分析时, 该方案采用一维高斯模型(Single Variable Gauss Model, SVGM), 即为每个 AP 单独建立模型。为提高定位精度, 在对接收信号进行概率分析时,

文献[5]提出一种多元高斯概率模型(Multivariable Gauss Model, MVGM), 即针对某一训练位置, 在多个 AP 之间以多元高斯概率分布对接收信号建模。实验结果表明, 基于 MVGM 的 RSSI 定位技术在定位精度方面优于 SVGM 技术。但多元高斯概率分布对 RSSI 建模大大提高了训练阶段和定位阶段的运算量, 增加了系统时延, 而对于 WLAN 目标定位应用来说, 实时性是位置服务 QoS 的一个主要性能指标。

本文提出一种基于高斯混合模型的 RSSI 定位方法, 采用高斯混合模型(Gaussian Mixture Model, GMM)对 RSSI 建模, 提高训练阶段 RSSI 建模的准确性, 从而降低系统定位误差; 同时在定位阶段避免了由于使用多元高斯模型带来的运算量大的问题, 减少定位时延, 提高定位的实时性能。实验评估结果表明, 该方法具有较好的定位精度和实时性。

2 高斯混合模型及其 EM 求解算法

GMM 是单一高斯概率密度函数的延伸, 能平滑地逼近任意形状的密度分布^[6]。

假设 N 个 RSSI 集合, 记为 X , $X = \{x_1, x_2, \dots, x_N\}$, 对于 X 中的单个采样 x_i , 其高斯混合分布的密度函数为

$$P(x_i | \Theta_j) = \sum_{j=1}^M \rho_j g(x_i | \mu_j, \Sigma_j) = \sum_{j=1}^M \frac{\rho_j}{\sqrt{2\pi} |\Sigma_j|} \exp\left[-\frac{1}{2}(x_i - \mu_j) \Sigma_j^{-1} (x_i - \mu_j)^T\right]$$

基金项目: 国家“863”计划基金资助项目(2007AA01Z309); 湖北省自然科学基金资助项目(2007ABA299)

作者简介: 程远国(1970 -), 男, 讲师、博士, 主研方向: 无线通信与网络; 耿伯英, 副教授、博士

收稿日期: 2008-09-10 **E-mail:** farcountry@163.com

其中, $i=1,2,\dots,N; j=1,2,\dots,M; \Theta=\{\rho_1,\rho_2,\dots,\rho_k, \mu_1,\mu_2,\dots,\mu_k, \Sigma_1, \Sigma_2,\dots,\Sigma_k\}$ 为各混合成分的参数矢量; $\Theta=\{\rho_j,\mu_j,\Sigma_j\}$ 表示第 j 个高斯分布的参数; ρ_j 是混合权重系数, 表示各混合成分的先验概率; μ_j 和 Σ_j 是高斯分布的参数, 即均值和方差, 记为 $C_j=(\mu_j, \Sigma_j)$ 。记 $p(X|\Theta)=\prod_{i=1}^N p(x_i|\Theta_i)$, 参数 Θ 的极大似然估计量 $\hat{\Theta}$ 为

$$\hat{\Theta} = \arg \max_{\Theta} p(X|\Theta) = \arg \max_{\Theta} \left[\prod_{i=1}^N p(x_i|\Theta_i) \right] = \arg \max_{\Theta} \left[\prod_{i=1}^N \sum_{j=1}^M \rho_j g(x_i|C_j) \right] \quad (1)$$

用 EM 算法求解式(1)的极大值。EM 算法分为 4 个步骤:

(1)初始化

假设有 M 个高斯成分, 初始参数为 $\Theta_j^0 = \{\rho_j^0, \mu_j^0, \Sigma_j^0\}$, $j=1,2,\dots,M$, 其中, ρ_j^0 为个成分初始权重, 表示第 j 个成分出现的概率, 且 $\rho_j^0 = \frac{1}{M}$; 利用 K -Means 方法计算出的各聚类中心点作为 μ_j^0 ; Σ_j^0 为协方差阵, 可取

$$\Sigma_j^0 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu_j^0)(x_i - \mu_j^0)^T.$$

(2)E-Step

引进数学符号 $E(Z_{ij})=p(C_j|X_i)$, 则 $E(Z_{ij})$ 可以看作是以下事件的概率: 在观测 RSSI 是 x_i 时, x_i 是由第 j 个高斯密度函数所产生的, 若用条件概率常用的表达式, 后验概率 $E(Z_{ij})$ 可写为

$$E(Z_{ij}) = p(C_j|x_i) = \frac{p(C_j)p(x_i|C_j)}{p(x_i)} = \frac{p(C_j)p(x_i|C_j)}{\sum_{j=1}^M p(C_j)p(x_i|C_j)} = \frac{\rho_j p(x_i|C_j)}{\sum_{j=1}^M \rho_j p(x_i|C_j)} \quad (2)$$

假设 $\Theta_j = \{\rho_j, \mu_j, \Sigma_j\}$ 已经求出, 则由式(2)得出:

$$E(Z_{ij}) = \frac{\rho_j p(x_i|C_j)}{\sum_{j=1}^M \rho_j p(x_i|C_j)}$$

(3)M-Step

将 E-Step 结果代入式(3)~式(5), 求 Θ_j^{n+1} :

$$\rho_j^{n+1} = \frac{1}{N} \sum_{i=1}^N E(Z_{ij}) \quad (3)$$

$$\mu_j^{n+1} = \frac{1}{N \rho_j^{n+1}} \sum_{i=1}^N E(Z_{ij}) x_i \quad (4)$$

$$\Sigma_j^{n+1} = \frac{1}{N \rho_j^{n+1}} \sum_{i=1}^N E(Z_{ij}) [(x_i - \mu_j^{n+1})(x_i - \mu_j^{n+1})^T] \quad (5)$$

(4)若 Θ_j^{n+1} 满足式(6)的收敛条件, 则迭代运算停止, 否则令 $\Theta_j = \Theta_j^{n+1}$, 并跳回 E-Step, 其中, ε 为设定的极小门限值。

$$\|\Theta_j^{n+1} - \Theta_j\| < \varepsilon, \quad \sum_{j=1}^M \|\Theta_j^{n+1} - \Theta_j\| < \varepsilon \quad (6)$$

3 基于 RSSI 的 GMM 定位方法

定位过程实际上是模式匹配问题, 分为 2 个阶段:

(1)训练阶段

在 WLAN 覆盖的区域内选择若干训练位置, 针对每一训练位置, 以固定功率发射大量的探测信号, 一定数目的无线接入点(Access Point, AP)接收这些信号, 形成训练位置的 RSSI 样本数据集, 各 AP 对收集到的这些 RSSI 样本数据集按照上述 GMM 方法进行概率分析, 建立信号强度概率分布数据库, 数据库中记录格式为 $\langle AP^{(i)}, X_i^{(j)}, \Theta_i^{(j)} \rangle$, 其中, $AP^{(i)}$ 表示该条记录属于第 i 个 AP; $X_i^{(j)}$ 表示第 i 个 AP 记录的第 j 个训练位置的坐标; $\Theta_i^{(j)}$ 表示第 i 个 AP 训练出的第 j 个训练位置的 GMM 参数。

(2)定位阶段

定位时使用者发出探测信号, 各 AP 接收的信号强度模

式为 $\langle x^{(1)}, x^{(2)}, \dots, x^{(P)} \rangle$, 其中, P 为满足一定 RSSI 门限的 AP 个数; $x^{(i)}(i=1,2,\dots,P)$ 为第 i 个 AP 收集的 RSSI, 按照记录的 GMM 参数计算出使用者可能出现位置的概率, 位置估计 $x^{(i)}$ 结果取最大概率对应的位置坐标, 则第 i 个 AP 估计的坐标 $x^{(i)}$ 为

$$X^{(i)} = \arg \max_{X_i^{(i)}} p(x^{(i)} | \Theta_i^{(j)})$$

取所有 AP 中最大概率对应的位置作为最终定位结果 X :

$$X = \arg \max_{X_i^{(i)}} \{ \max \{ p(x^{(i)} | \Theta_i^{(j)}) \} \}$$

实验结果发现, GMM 概率最大的定位结果往往分布在真实位置的四周, 可取前 p 个具有最大概率值的点的坐标平均值作为最终定位结果 X :

$$X = \frac{1}{P} \sum_{i=1}^P X^{(i)}$$

4 实验评估

实验环境为某单位 802.11b 无线局域网, 楼层面积为 400 m², 实验区域被划分为 100 个 2 m×2 m 的小方格, 把这些小方格的 121 个交点位置作为训练位置。在任意一个训练位置进行实验, 临近的一个 AP 收集 RSSI 值, 并考察这些 RSSI 分布。

训练阶段移动终端以固定功率在每个训练位置每秒钟发射一次探测信号, 每个训练位置共计 1 500 次, 高斯成分为 3 个。5 台带有无线网卡的笔记本电脑运行 Windows XP 操作系统和定位服务程序, 定位服务程序需利用操作系统的 ndisui(ndis user i/o)服务来采集 RSSI 值。根据不同实验目的, 在实验区域内选取若干训练位置。1 台 PC 机作服务器运行服务器程序和后台数据库, 使用 Microsoft Visual C++ 6.0 开发应用程序进行以下实验。

(1)定位误差(Localization Error, LE)与目标位置, 考察 GMM 算法的定位效果。定位结果如图 1 所示。

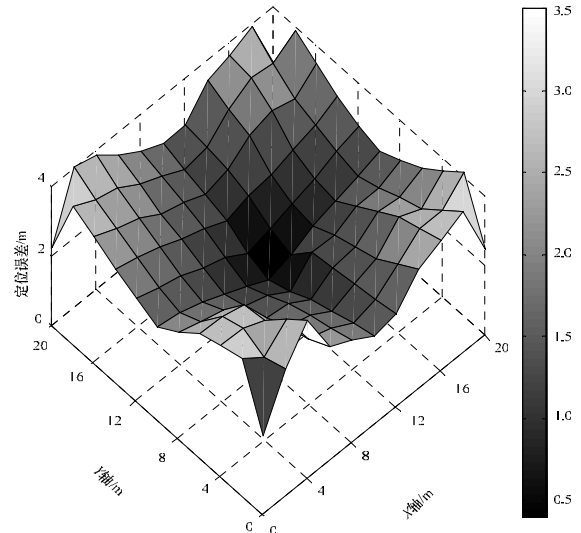


图 1 定位误差与训练位置

从图 1 可以看出, 在设定训练位置数目、AP 数量和高斯成分个数前提下, LE 是目标位置的函数, 目标在实验区域中心时 LE 最小, 目标向实验区域边界移动时, LE 增加。设定实验环境条件下平均定位误差是 1.97 m, 最小定位误差 0.1 m, 最大定位误差 3.2 m, 总体定位效果较好。

(2)SVGD, MVGD 和 GMM 性能比较。主要考察 2 个指标:

平均定位误差(Mean Localization Error, MLE),用于比较不同算法的定位精度;定位响应时间(Localization Respond Time, LRT),用来评估算法的实时性能。

由于 AP 数量对 3 种算法影响的区别较大,这里评估的是在不同 AP 数量条件下 3 种算法的 MLE 和 LRT,实验结果如图 2 所示。

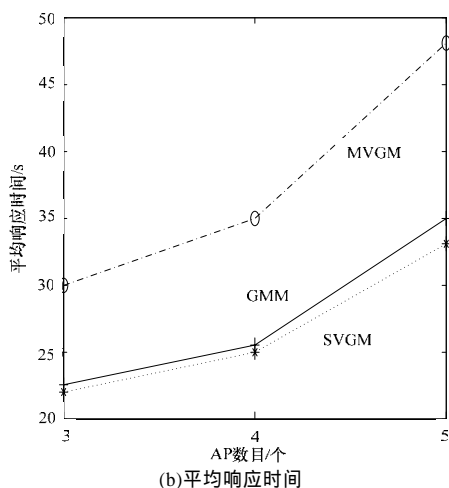
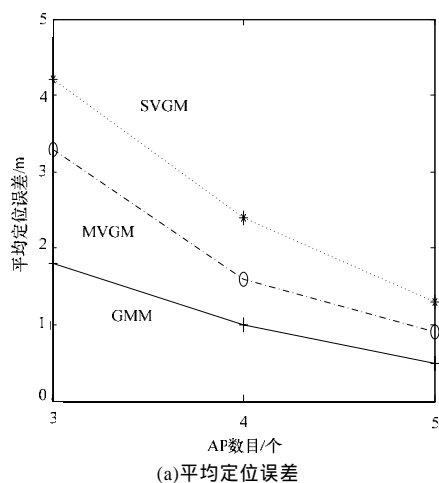


图 2 3 种方法平均定位误差和平均响应时间比较

(上接第 24 页)

可以看出,随着时间(移动对象更新次数)的变化,TPR*-tree 回答查询所需要的索引节点访问代价随之增加,但相比之下,采用本文构建和插入延迟更新算法的 TPR*-tree 索引具有较好的稳定性。

5 结束语

本文提出一种改进的 TPR*-tree 索引构建算法,通过在 TPR*-tree 索引根节点层引入溢出桶以缓存插入的移动对象记录,提出插入延迟更新的动态维护算法,提高 TPR*-tree 索引的查询和更新性能。实验结果表明,采用本文构建和插入延迟算法的 TPR*-tree 查询性能及动态更新性能均优于采用标准构建和插入算法的 TPR*-tree 索引。

参考文献

[1] Mohamed F, Mokbel T M. Spatio-temporal Access Methods[J].

从图 2(a)可以看出,3 种方法中 GMM 的 MLE 最小, MVGD 次之, SVGD 的 MLE 最大。而在 LRT 的比较中, MVGD 的 LRT 显然高于 GMM 和 SVGD, GMM 稍高于 SVGD, 如图 2(b)所示。这是因为在训练阶段, GMM 要进行迭代运算以求 GMM 参数,所以运行时间明显高于 SVGD。尽管 GMM 的 LRT 稍高于 SVGD,但能够用很小的时间延迟换来较高的定位精度,还是合算的。总之,采用本文提出的 GMM 定位技术,其总体性能要优于其他 2 种方法。

5 结束语

提出一种基于 GMM 的匹配方法,并进行一系列实验以验证和评估所提方法的性能,实验结果表明,该方法在定位的精度和实时性 2 个方面均有较好的结果。另外,在移动终端移动速度较快的情况下,该定位方法不稳定,表现在平均定位误差波动较大上。下一步工作是进一步完善该算法,以提高系统的实时性,并降低定位误差,同时增强定位系统的健壮性。

参考文献

[1] 余涛,黄书宝,葛昭攀,等. 无线局域网环境下的位置服务研究[J]. 计算机工程, 2007, 3(14): 122-124.
 [2] 董梅,杨曾,张健,等. 基于信号强度的无线局域网定位技术[J]. 计算机应用, 2004, 24(12): 49-52.
 [3] Bahl P, Padmanabhan V N. RADAR: An In-building RF-based User Location and Tracking System[C]//Proc. of 2000 Annual Joint Conference of the IEEE Computer and Communications Societies. [S. l.]: IEEE Computer Society, 2000.
 [4] Roos T, Myllymaki P, Tirri H, et al. A Probabilistic Approach to WLAN User Location Estimation[J]. International Journal of Wireless Information Networks, 2002, 9(3): 155-164.
 [5] Moustafa A. Multivariate Analysis for WLAN Location Determination Systems[C]//Proc. of the 2nd Annual Int'l Conf. on Mobile and Ubiquitous Systems: Networking and Services. San Diego, California, USA: ACM Press, 2005.
 [6] Dempster A P, Laird N M, Rubin D B. Maximum Likelihood from Incomplete Data via the EM Algorithm[J]. Journal of the Royal Statistical Society, 1977, 39(2): 1-38.

IEEE Data Engineering Bulletin, 2003, 26(2): 40-49.

[2] Simonas S, Christian S. Indexing the Positions of Continuously Moving Objects[C]//Proc. of the SIGMOD. Dallas, Texas, USA: [s. n.], 2000.
 [3] Papadias D. The TPR*-tree: An Optimized Spatio-temporal Access Method for Predictive Queries[C]//Proc. of the VLDB. Berlin, Germany: [s. n.], 2003.
 [4] Mohamed F. LUGrid: Update-tolerant Grid-based Indexing for Moving Objects[C]//Proc. of the MDM. Nara, Japan: [s. n.], 2006.
 [5] Lin Bin, Su Jianwen. On Bulk Loading TPR-tree[C]//Proc. of the MDM. Berkeley, California, USA: [s. n.], 2004.
 [6] Brinkhoff T. A Framework for Generating Network-based Moving Objects[J]. GeoInformatica, 2002, 6(2): 153-180.