

基于支持向量机多分类的室内定位系统

朱宇佳¹ 邓中亮^{1,2} 刘文龙¹ 徐连明¹ 方灵¹

(北京邮电大学电子工程学院 北京 100876)¹

(北京邮电大学信息光子学与光通信国家重点实验室 北京 100876)²

摘要 为解决室内实时定位中定位精度不高、显示效果来回跳动的问题,提出了一种基于支持向量机(SVM)多分类的室内定位算法。针对传统基于采样点的匹配算法处理非线性问题的不足以及实时定位时信号采集时间较短、变化幅度较大等问题引入网格定位的概念,将定位匹配设计成多分类问题,利用 SVM 得到目标最有可能所属的 K 个网格;利用实时定位中前、后两个位置的相关性剔除这 K 个网格中可能性较小的网格,最终所属网格坐标加权后得到估算位置坐标,并利用卡尔曼滤波算法对估算位置坐标进行滤波处理。实验结果表明,算法的定位精度与传统 SVM 的精度相比有明显的提高。

关键词 支持向量机(SVM), 网格, 室内实时定位, 接收信号强度(RSSI), 卡尔曼滤波

中图分类号 TP393 文献标识码 A

Multi-classification Algorithm for Indoor Positioning Based on Support Vector Machine

ZHU Yu-jia¹ DENG Zhong-liang^{1,2} LIU Wen-long¹ XU Lian-ming¹ FANG Ling¹

(School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China)¹

(State Key Laboratory of Information Photonics and Optical Communications, Beijing University of

Posts and Telecommunications, Beijing 100876, China)²

Abstract A multi-classification algorithm for indoor positioning based on SVM was proposed to tackle the problem of low precision and fluttering results faced in many real-time location systems. Traditional matching algorithms based on sampling points are always deficient in dealing with nonlinear problem and jumping results in a short time. In handing this limitation, object location process was considered as a multi-classification problem by introducing grid concept. K candidate grids were obtained using SVM first. These candidates were then refined by previous location results, and ultimate accuracy result was achieved through a Kalman filter. Temporal information was utilized in the matching process to make the object movement more stable and smooth. Experiments show the superiority of our method over naive SVM method.

Keywords Support vector machine(SVM), Grid, Real-time indoor location, Received signal strength indication(RSSI), Kalman filter

1 引言

随着物联网应用的不断扩大和深入, LBS(location based services, 基于位置的服务)显得越来越重要。离开位置信息, 感知数据也失去了应用价值。

RSSI(received signal strength index, 接收信号强度指示)无需额外的硬件装置, 成本低, 能满足大规模应用要求, 目前室内定位大多采用基于 RSSI 的定位方式。根据是否需要通过物理手段直接测量节点间距离, RSSI 定位技术可用于基于测距和非测距两类算法。基于测距的算法需要获取节点间的距离^[1]。但是, 室内环境相对复杂, 测距过程很容易受其影响, 而产生测距误差, 且由此带来的定位误差容易造成累积, 影响定位精度^[2]。非测距方式基于位置指纹匹配, 分为离线

采样和在线定位两个阶段。离线采样阶段, 操作人员在被定位环境里确定若干采样点, 然后遍历所有采样点, 记录下在每个采样点测量的无线信号特征, 建立信号分布图。在线定位阶段, 当用户移动到某一位置时, 根据实时收到的信号强度信息, 利用定位算法将其与位置指纹数据库中的信息匹配, 计算出该用户的位置。非测距方式无需考虑测距误差, 在成本和功耗方面与前者相比具有显著优势。

现有利用位置指纹进行匹配的算法主要有 K 最邻近算法、神经网络、概率算法和支持向量回归算法。 K 最近邻法^[3]根据信号分布图找出信号强度样本最接近的一个或多个样本, 将它们对应的采样点或多个采样点的平均作为估计的用户位置。文献[4-7]讨论了多层感知机网络、径向基函数网络在基于位置指纹的定位技术中的应用。文献[8]给出了两种

到稿日期: 2011-05-23 返修日期: 2011-08-27 本文受国家 863 计划(2009AA12Z324)资助。

朱宇佳(1984-), 女, 博士生, 主要研究方向为无线定位、LBS、物联网, E-mail: zyjbupt@yahoo.com.cn; 邓中亮(1965-), 男, 教授, 博士生导师, 主要研究方向为无线传感器网络、卫星导航定位、多媒体通信、微电子设计、MEMS 等。

估计条件概率分布函数或似然函数的方法:核方法和直方图法。文献[11]将无线接入点的信号强度与二维位置坐标 (x, y) 之间的非线性关系列为回归问题,并指出对于回归问题,支持向量机与加权 K 最近邻法的性能非常接近。

上述基于位置指纹匹配的定位算法,往往将定位问题描述为在位置指纹数据库中找到与在线定位阶段采集到的无线信号最相近的点,离线阶段信号强度只在某些指定的采样点采集,记录的是采样点的坐标以及在该采样点采集到的信号强度的均值^[13-15]。即离线阶段时,设 S_{ij} 表示第 i 个采样点接收到第 j 个信标节点的信号强度平均值,在各个采样点处测量并记录下位置指纹数据,即 $(x_i, y_i, S_{i1}, S_{i2}, \dots, S_{iN}), i=1, 2, \dots, n, n$ 表示采样点的总数, N 表示信标节点的总数。

在线阶段时,设 S_j 表示接收到的来自第 j 个信标节点的信号强度平均值,由公式 $d_i = \{(\sum_{j=1}^N |S_j - S_{ij}|)^p\}^{\frac{1}{p}}$ 计算出 d_i 值最小的 k 个采样点并取它们位置的平均坐标作为定位结果 (\hat{x}, \hat{y}) ,此处 $(\hat{x}, \hat{y}) = \frac{1}{k} \sum_{i=1}^k (x_i, y_i)$ 。可以发现,这种基于采样点的匹配算法原理简单、实现方便。但是这种基于采样点的定位方法存在如下问题:

(1)接收信号强度本身是非线性的,但在计算 d_i 的时候是线性的;

(2)采样点的指纹数据是离线阶段接收到各信标节点的信号强度均值,在线阶段信号采集时间较短、信号变化幅度较大,与各采样点的信号强度均值进行匹配时容易造成误差。

本文根据室内定位的实际情况,提出了一种基于SVM的多分类室内定位算法,针对基于采样点的匹配算法存在的问题引入网格定位的概念,将定位匹配设计成多分类问题,并且针对实际情况下发定位请求的用户往往处于移动状态,手持移动终端的用户前、后位置具有一定相关性,即根据上一点的位置可以将下一点的位置限定在一定的范围内,算法将移动中用户前后位置的关联性融入位置指纹的匹配过程中,使用户位置的移动更加平缓,且提高了室内定位系统的定位性能和稳定性。

2 基于SVM分类的室内定位系统

基于SVM分类的室内定位系统主要包括离线采样和在线定位两个阶段。如图1所示,离线采样阶段,根据上节所述划分大小合适的网格,将每个网格作为一类的类别。在每个网格内收集各信标节点的接收信号强度(RSS),建立用于各网格内位置指纹的信号分布图,即收集学习机器的训练样本,利用SVM多分类算法对训练样本进行训练。在线定位阶段,用户实时测量RSS,结合上一时刻位置坐标,输入已构建好的支持向量机分类模型,将所得分类结果转换成物理位置坐标。

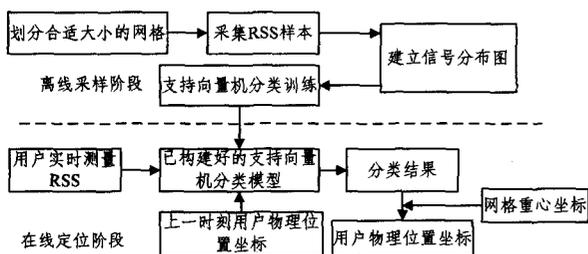


图1 室内定位系统框图

3 支持向量机基本理论

设样本集及其所属类别表示为 $(x_i, y_i), i=1, \dots, n, x \in R^d, y \in \{-1, +1\}$ 是类别标号。在线性可分情况下,支持向量分类器试图找到一个使间隔最大的最优分类超平面 $w^T \cdot x + b = 0$ 。

要找到这个超平面,需要求解下面的二次规划问题^[12]:

$$\min \frac{1}{2} \|w\|^2 \quad (1)$$

$$\text{s. t. } y_i [(w^T \cdot x_i) + b] \geq 1, i=1, 2, \dots, n$$

上述问题的解可以通过求解如下的二次规划问题得到:

$$\max W(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (2)$$

$$\text{s. t. } \sum_{i=1}^n \alpha_i y_i = 0, 0 \leq \alpha_i, i=1, 2, \dots, n$$

这是一个不等式约束下的二次函数寻优问题,存在唯一解。容易证明,解中只有少部分拉格朗日乘子 α_i 不为零,对应的样本就是支持向量。最终得到的最优分类超平面是 $f(x) = \sum_{i=1}^n \alpha_i^* y_i x_i^T x + b^*$,其中 b^* 可以通过任一支持向量求得。

对于非线性可分问题,原始训练样本通过一个非线性映射 $\Phi: R^d \rightarrow F$ 映射到一个高维特征空间 F ,这样在原空间的非线性分类问题就变为了特征空间中的线性分类问题。若引入核函数 $K(x_i, x_j) = \Phi(x_i) \Phi(x_j)$,在实际求解时不需要知道 $\Phi(\cdot)$ 的具体形式。此时相应的分类超平面变为 $f(x) = \sum_{i=1}^n \alpha_i^* y_i K(x_i, x_j) + b^*$ 。

支持向量分类方法最初是为了解决二类分类问题提出的,在解决多类模式识别问题时,常采用一对一或者一对多策略^[4]。

4 基于网格的多分类定位问题描述

本文将待定位点与采样点之间的匹配(见图2)转换为多分类的问题,根据室内建筑结构,划分合适大小的网格(见图3),将每一个网格作为一类的类别,以网格为单位采集信号强度并记录网格号,网格内不同形状的点表示在该网格内采集到的样本;对于非线性分类问题,SVM通过核函数将非线性变换映射到高维特征空间,然后在高维特征空间中进行线性分类。

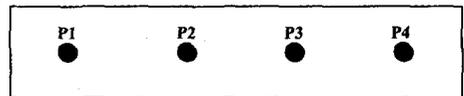


图2 基于采样点的定位问题描述示意图

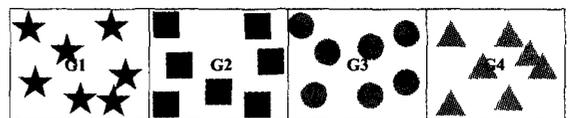


图3 基于网格分类的定位问题描述示意图

本文考虑将由 N 个信标节点组成的无线网络 $\{l_1, l_2, \dots, l_N\}$ 部署在二维区域 C 中, $C \subseteq R^2$ 。二维区域 C 被划分为 M 个网格,网格编号分别为 $1, 2, \dots, M$ 。令集合 $M = \{1, 2, \dots, M\}$ 表示网格的集合,网格 $G_i (i \in M)$ 的重心真实坐标为 $g_i (g_i \in R^2)$,并令 $G_a = [g_1, g_2, \dots, g_M]^T$ 表示网格重心的坐标矩阵。定位的目标是计算移动终端位置的估计值 \tilde{x} ,使得 \tilde{x} 尽

可能地逼近未知节点的真实坐标 x 。

5 基于 SVM 的多分类定位模型

离线采样阶段,每个网格采集 Q_{G_i} ($i \in M$) 条包含各信标节点接收信号强度和信标节点编号的位置指纹 $S_j \{G_i \{ \langle r_1, l_1 \rangle, \langle r_2, l_2 \rangle, \dots, \langle r_q, l_q \rangle \} \}$, $i=1, 2, \dots, M, j=1, 2, \dots, Q_{G_i}$ 作为训练样本。

任取其中两个网格 G_i 和 G_j 的样本,作为正负两类进行二分类问题的建模。由于训练样本非线性可分,相应的分类超平面即为

$$f(x) = \sum_{i=1}^n a_i^* y_i K(x_i, x_j) + b^*$$

要确定在线阶段采集到的位置指纹 $S_x \{ \langle r_1, l_1 \rangle, \langle r_2, l_2 \rangle, \dots, \langle r_q, l_q \rangle \}$ 属于哪个网格,需要通过二值分类器进行决策。若决策函数

$$f(x) = \text{sgn} \{ w^* \cdot \Phi(x) + b^* \} \\ = \text{sgn} \left\{ \sum_{i=1}^n a_i^* y_i K(x_i, x_j) + b^* \right\} \quad (3)$$

大于零,则样本属于第 i 类,否则属于第 j 类。

本文在进行多分类时采用的是一对一策略。分类过程如下:待分类位置指纹 S_x 代入二值分类器决策函数,如果结果大于零,就对 S_x 属于第 i 类的可能性投一票,否则就对属于第 j 类的可能性投一票。如此使用每个分类器对待分类指纹 S_x 进行求值、投票,最后对投票数进行排序。投票越多,说明 S_x 越有可能属于这个类,即 S_x 越有可能属于这个网格。

实时跟踪状态下的定位,由于前一时间刻的位置与后一时间刻的位置存在相关性,因此可以结合卡尔曼滤波算法,根据上一点点的分类结果判定下一点位置出现的范围。

卡尔曼滤波在数学上是一种统计估算方法,它通过处理一系列带有误差的实际量测数据而得到物理参数的最佳估算,根据前一个估计值和最近一个观测值来估计信号的当前值,不需要全部的观测值^[16]。通过卡尔曼滤波可以对用户位置坐标进行滤波处理,以进一步提高室内定位系统的定位精度。

卡尔曼滤波的过程方程和观测方程如下:

$$x_k = F_k x_{k-1} + w_k, w_k \sim N(0, Q) \\ x(0) \sim N(X(0), V(0)) \quad (4) \\ z_k = H_k x_k + v_k, v_k \sim N(0, R)$$

系统中各个量的含义和矩阵大小如表 1 所列。假设过程噪声(process noise)和观测噪声(measurement noise)是相互独立的,分别用协方差矩阵 Q 和 R 来表示。

表 1 系统中各个量的含义和矩阵大小

变量符号	名称	矩阵大小
x_k	系统状态	$n \times 1$
F	状态矩阵	$n \times n$
w_k	过程噪声	$n \times 1$
z_k	观测值	$m \times 1$
H	观测矩阵	$m \times n$
v_k	测量噪声	$m \times 1$
P_k	均方误差矩阵	$n \times n$
K_k	修正矩阵	$n \times m$
I	单位矩阵	可变

卡尔曼滤波的时间更新方程和测量更新方程分别为:

①时间更新方程

$$\hat{x}_k = F \hat{x}_{k-1}, P_k = F P_{k-1} F^T + Q \quad (5)$$

②测量更新方程

$$\begin{cases} K_k = P_k H^T (H P_k H^T + R)^{-1} \\ \hat{x}_k = \hat{x}_k + K_k (z_k - H \hat{x}_k) \\ P_k = (I - K_k H) P_k \end{cases} \quad (6)$$

已知 $k-1$ 时刻 Kalman 滤波器输出为 \hat{x}_k , 假设 k 时刻 SVM 多分类投票结果为 $V_k \{v_1, v_2, \dots, v_M\}$, 从 V_k 中选取 K 个投票数最大的类别, 记为 $\{M_1(V_k), M_2(V_k), \dots, M_K(V_k)\}$, 对应的网格重心的坐标为 $[g_{M_1(V_k)}, g_{M_2(V_k)}, \dots, g_{M_K(V_k)}]^T$ 。对这 K 个网格进行选择, 设阈值为 $\xi (\xi > 0)$, 若 $\|g_{M_i(V_k)} - \hat{x}_k\|_2 < \xi$, $i=1, 2, \dots, K$, 则此网格作为此次分类可能的结果之一保留。反之, 此网格 $M_i(V_k)$ 被剔除。其中阈值 ξ 根据人移动的速度设定, 而网格数 K 则根据定位区域大小及网格划分大小确定。此点位置观测值输出为

$$z_k = \sum_{i=1}^K \left(\frac{\frac{1}{\eta(g_{M_i(V_k)}) \cdot x_i}}{\sum_{i=1}^K \frac{1}{\eta(g_{M_i(V_k)}) \cdot x_i}} \right) \cdot g_i^T \quad (7)$$

式中, $\eta(g_{M_i(V_k)}) = \begin{cases} 1, & \|g_{M_i(V_k)} - \hat{x}_k\|_2 < \xi \\ 0, & \text{其它} \end{cases}, i=1, 2, \dots, K$ 。

第 k 个时刻卡尔曼滤波输出为 $\hat{x}_k = F \hat{x}_{k-1} + K_k (z_k - H F \hat{x}_{k-1})$, \hat{x}_k 为第 k 个时刻状态的卡尔曼滤波输出, \hat{x}_{k-1} 为第 $k-1$ 个时刻状态的卡尔曼滤波输出, K_k 为第 k 个时刻的卡尔曼滤波系数, z_k 为第 k 个时刻的观测状态。由式(6)可以求得第 k 个时刻的方差、第 k 步卡尔曼增益系数。

6 实验结果与分析

6.1 网格大小仿真

在实验之前本文对网格大小进行了仿真, 模拟 6.2 节所述实验场地长廊型环境, 6 个信标节点呈 Z 字型分布于两侧 2m 高处, 信标节点间距 20m。信号强度通过常用的对数距离路径损耗模型(式(8))模拟, 其中 ζ 是标准偏差为 5 的正态随机变量, d_0 为 1m 处信号强度, 取为经验值 -45dbm, 路径损耗指数 n 取经验值 2.326。

$$p = p_0 - 10 \cdot n \cdot \lg(d/d_0) + \zeta \quad (8)$$

网格宽 2m, 用户位置样本数为 3400 个, 均匀分布于仿真区域, 未滤波的 SVM 分类算法在不同网格长度下定位精度的仿真结果如图 4 所示, 网格长 1m 与 2m 定位精度差异不大。网格长度大于 2m 时, 网格粒度越小, 精度越高, 与文献[9]参考点设定得越密定位精度越高的论述一致。

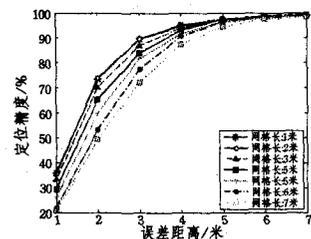


图 4 不同大小网格定位精度仿真

6.2 实验环境

为验证本文提出的动态定位算法的有效性, 笔者在北京

邮电大学教四四楼东侧走廊、一间办公室和一间卫生间进行了多次实验。根据仿真结果,将东侧走廊均分为 $1.8\text{m} \times 2\text{m}$ 的网格,办公室和卫生间由于面积较小,分别算作一个网格,共 35 个网格。

离线阶段,在每个网格内缓慢移动 (0.5m/s),保证网格内各方向都能采集到样本。采集频率设置为 500ms ,连续采集 100 条位置指纹样本。信标节点分为无线接入点 (AP, Access Point) 和无线传感器网络节点两种。无线接入点主要分布在走廊区域,无线传感器网络节点主要置于房间内。采集设备为实验室自主研发的、能够同时采集无线局域网信号强度和无线传感器网络信号强度的智能终端,在采集过程中,智能终端共探测到了 7 个无线接入点和 8 个无线传感器网络节点。设置接收信号强度阈值为 -75dbm ,即信号强度小于 -75dbm 的信标节点不予记录。在线定位的采集频率和接收信号强度阈值与离线采集相同,系统每 0.5s 输出一次定位结果。

SVM 分类训练采用台湾大学林智仁博士等人开发的开源、易用的支持向量工具 Libsvm (<http://www.csie.ntu.edu.tw/~cjlin/libsvm>),并且利用 Libsvm 提供的工具进行网格搜索,获取参数 C 和 g 的值。核函数使用高斯径向基核。

6.3 实验结果和精度分析

针对上述实验环境,实验比较了本文提出的算法、文献 [11] 中提出的 SVM 回归算法的实验结果与真实坐标。实验场景为 x 轴方向上的长廊型区域,图 5 为该方向上的坐标对比结果。式 (7) 中 K 值定为 3,阈值 ξ 设置为 45。

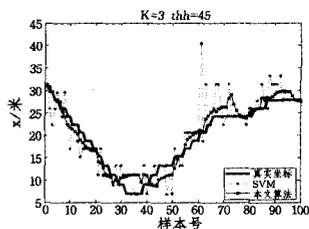


图 5 x 轴坐标的定位精度比较

从图 5 可以看出,SVM 回归算法得到的定位结果抖动较大,而本文提出的算法明显平滑,与真实位置也更为接近。

图 6 表明本文提出的算法在定位精度上与 SVM 回归算法相比,有明显的提高,2m 以内的定位精度为 66.9%,提高了 15.8%;3m 以内的定位精度为 86.3%,提高了 23.8%;5 米内的定位精度为 95.7%,提高了 14.4%;新的定位算法平均误差为 1.93m,比 SVM 改进了 1.65m。

为了验证信标节点数对定位精度的影响,本文还进行了信标节点数对定位精度影响的对比实验。受实验场地的限制,只改变了走廊内信标节点布局,观察到的定位精度变化结果如图 7 所示。

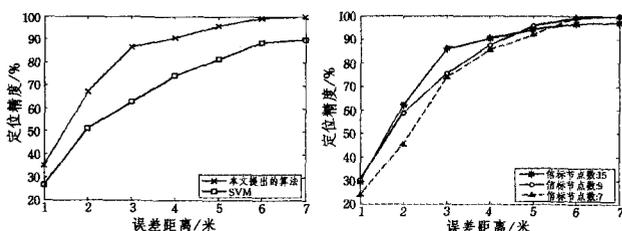


图 6 定位精度对比

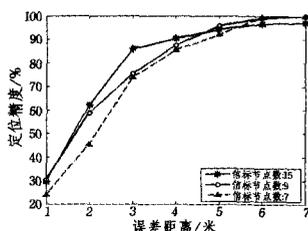


图 7 信标节点数对定位精度的影响

结束语 针对传统基于采样点的匹配算法在处理非线性问题的不足以及实时定位时信号采集时间较短、变化幅度较大的问题,本文将室内实时定位过程设计成多分类问题,建立了基于多类 SVM 的动态定位模型。另一方面,由于实时跟踪状态下的定位,前一时刻的位置与后一时刻的位置存在相关性,因此本文将卡尔曼滤波算法应用在基于 SVM 的多分类投票过程中。实验证明,本文提出的基于 SVM 的多分类定位算法与 SVM 回归算法相比,定位精度明显提高,运动轨迹更加平滑,用户体验更好。

此外,实验证明定位精度并非随着信标节点密度的增大而升高。如何根据场地形状和建筑结构合理布设信标节点,在保证精度的情况下进一步减小信标节点的密度,有待进一步研究。

参考文献

- [1] 周艳. 基于 RSSI 无线传感器网络空间定位算法[J]. 通信学报, 2009, 30(6): 75
- [2] Hara S, Anzai D. Comparison of Three Estimation Methods for RSSI-based Localization with Multiple Transmit Antennas[C]// MASS. 2007: 1
- [3] Bahl P, Padmanabhan V N. RADAR: an in-building RF-based user location and tracking system[C]// INFOCOM, 2000, 775(2)
- [4] Chih-Wei H, Chih-Jen L. A comparison of methods for multi-class support vector machines[J]. IEEE Transactions on Neural Networks, 2002, 13(2): 415
- [5] Nerguizian C, Despins C, Affès S. Indoor geolocation with received signal strength fingerprinting technique and neural networks[C]// Telecommunications and Networking-ICT. 2004: 866
- [6] Martínez E A, Cruz R, Favela J. Estimating user location in a WLAN using backpropagation neural networks[C]// Advances in Artificial Intelligence-IBERAMIA. 2004: 737
- [7] Krumm J, Platt J. Minimizing calibration efforts for an indoor 802. 11 device location measurement system [C] // Machine Learning Meets the User Interface. 2003
- [8] Kushki A, Plataniotis K N, Venetsanopoulos A N. Kernel-based positioning in wireless local area networks[J]. IEEE Transactions on Mobile Computing, 2007, 6(6): 689
- [9] Wassi G I, et al. Indoor location using received signal strength of IEEE 802. 11b access point[C]// Electrical and Computer Engineering. 2005
- [10] Saha S, et al. Location determination of a mobile device using IEEE 802. 11b access point signals[C]// WCNC. 2003
- [11] Brunato M, Battiti R. Statistical learning theory for location fingerprinting in wireless LANs[J]. Computer Networks, 2005, 47(6): 825
- [12] 邓乃扬, 田英杰. 数据挖掘中的新方法: 支持向量机[M]. 北京: 科学出版社, 2004
- [13] Li B, Wang Y, Lee H K. Method for yielding a database of location fingerprints in WLAN [J]. IEE Proceedings-Communications, 2005, 152(5): 580
- [14] 张明华. 基于 WLAN 的室内定位技术研究[D]. 上海: 上海交通大学, 2009
- [15] 赵永翔, 周怀北, 陈森, 等. 卡尔曼滤波在室内定位系统实时跟踪中的应用[J]. 武汉大学学报: 理学版, 2009, 55(6): 696
- [16] 万江文, 吴佳灵, 冯仁剑. 卡尔曼滤波在无线传感器网络节点定位中的应用[J]. 高技术通讯, 2009, 19(2): 151