

文章编号:1000-5862(2020)02-0148-05

基于蚁群算法的支持向量机室内蓝牙标定定位

吴璇¹,薛峰²,余敏^{1*}

(1. 江西师范大学软件学院,江西 南昌 330022;2. 江西师范大学计算机信息工程学院,江西 南昌 330022)

摘要:针对不同型号的智能手机之间存在硬件差异,导致在使用不同的智能手机进行室内定位时,采集同一蓝牙信号强度观测值存在偏差而影响定位精度的问题,提出了一种基于蚁群算法的支持向量机室内定位蓝牙标定方法. 由于支持向量机参数的选取对其预测精度影响较大,因此利用蚁群算法避免人为盲目地选择支持向量机的参数,优化模型并提高预测精度. 实验结果表明:基于蚁群算法的支持向量机室内定位蓝牙标定方法相比标定前的精度提高了 37.3%,可以有效地进行室内定位.

关键词:软硬件异构;蓝牙标定;支持向量机;蚁群算法;标定模型

中图分类号:P 228 **文献标志码:**A **DOI:**10.16357/j.cnki.issn1000-5862.2020.02.06

0 引言

随着普适计算和无线传感器网络技术的发展,人们对于定位的需求逐渐增大,基于定位的服务进入飞速发展时期. 然而,随着时代不断发展,人们大部分时间处于室内环境,如地铁站、办公楼等. 由于在室内条件下用于定位的卫星信号被建筑物遮挡,人们无法获得精确的定位服务,因此,室内定位技术应运而生. 目前,科研人员提出了许多提供室内定位服务的方案,常用的定位技术有 ZigBee、蓝牙、wifi 等^[1]. 基于蓝牙信号强度(RSSI)的定位方法由于其布置简单、成本低廉等优势而获得广泛使用^[2]. 但是,随着电子设备软硬件更新换代速度越来越快,不同的电子设备之间获取到的蓝牙信号强度可能会存在较大差异,需要定位的设备若未在定位区域采集指纹数据库,则在使用其他设备采集的指纹库进行定位时会影响定位的精度. 但是针对每一款手机都采集指纹数据库也不切实际,会成倍增加工作量,这大大制约了基于蓝牙信号强度的室内定位技术的发展^[3]. 为了减小由于设备异构性造成的误差,就需要使用标定方法修正每款智能设备的蓝牙 RSSI 观测值,从而使得用于定位的设备和参考设备在相同位置上的 RSSI 观测值趋于一致.

对于蓝牙标定问题,文献[4]提出了采用人工校正的方法,在同一个采集点上同时记录 2 台不同型号移动设备采集的信号强度,然后利用最小二乘法进行数据拟合并对线性参数进行估计. 文献[5]提出使用 RSSI 观测值直方图来进行设备自动标定的方法. 虽然上述方法能减小由设备异构性所带来的误差,但是并不能得到一个完美的模型. 本文提出一种基于蚁群算法的支持向量机的蓝牙标定方法,支持向量机算法是利用 2 次型求最优化问题^[6]的全局最优解的一种算法,在小样本、非线性和高维模式识别等方面具有非常大的优势,但是它在应用上仍然存在一些需要改进的地方^[7],特别关键的是参数的选取,选取的参数是否准确将直接影响支持向量机的预测准确性,所以支持向量机的参数选择问题成为限制其发展的关键因素^[8,9]. 目前常用的参数优化方法有遗传算法、实验法、梯度下降法和蚁群算法等. 蚁群算法是一种自组织的算法,具有并行性、正反馈等特征,它对初始路线要求低,并且无需人工干预、不易陷入局部最优,易于得到优化问题的解^[10]. 通过对采集的数据进行训练,使用蚁群算法找到支持向量机预测模型最优的参数,得到优化后的支持向量机模型. 通过该模型可以对不同型号的设备接收到的蓝牙信号强度进行标定,从而大大减少由于设备异构性对定位精度造成的不利影响.

收稿日期:2019-10-20
基金项目:国家重点研发计划课题(2016YFB0502204)资助项目.
通信作者:余 敏(1964-),女,江西南昌人,教授,博士生导师,主要从事无线传感器网络与技术等研究. E-mail:myu821@163.com

1 支持向量机回归预测模型

支持向量机是由 Vapnik 及其团队提出的一种新型机器学习算法^[11],它以统计学理论为基础.支持向量机的核心在于核函数,在通常情况下样本空间中向量机较难线性区分,而若将低维线性不可分情况映射为高维空间线性可分需要使用核函数,并可大大降低计算复杂度.支持向量机算法支持分类和回归问题,本文提出的蓝牙标定属于回归问题.设训练样本集为 $\{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbf{R}^d, y_i \in \{-1, 1\}, i = 1, 2, \dots, n\}$,其中 \mathbf{x}_i 为样本集的输入变量, y_i 为输出结果, n 为样本集的样本数目.支持向量机通过使用非线性映射 φ 将样本数据映射到高维特征空间,并在该高维空间中完成最优超平面分类,即

$$g(\mathbf{x}) = \mathbf{w}\varphi(\mathbf{x}) + \mathbf{b}, \tag{1}$$

其中 \mathbf{w} 为高维空间中的矢量权重, \mathbf{b} 为阈值.由 \mathbf{w} 和 \mathbf{b} 可以得到一个最优分类平面的位置.结构风险最小化是该公式需遵循的准则,所以(1)式可以通过求解

$$f(\mathbf{w}, \xi_i, \xi_i^*) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \tag{2}$$

的最小值获得,其中 C 是惩罚参数且大于 0,它的取值表示对误差的容忍度.惩罚参数 C 的值越大代表可容忍的误差越小, C 值越小代表可容忍的误差越大,它权衡了模型复杂性和样本的误差值.并加入松弛变量 ξ_i, ξ_i^* ,且 ξ_i, ξ_i^* 皆非负,可以将(2)式转化为等价最优化问题.(2)式的限制条件为

$$\begin{cases} g(\mathbf{x}) - y_i \leq \xi_i^* + \varepsilon, \\ y_i - g(\mathbf{x}) \leq \xi_i + \varepsilon, \\ \xi_i \geq 0, \\ \xi_i^* \geq 0. \end{cases}$$

上述问题属于凸二次优化问题,通过引入拉格朗日函数可转化为对偶问题并求解该最优化问题^[12],即

$$\max w(a, a^*) = \sum_{i=1}^n y_i(a_i - a_i^*) - \sum_{i=1}^n \varepsilon(a_i + a_i^*) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (a_i - a_i^*)(a_j - a_j^*)(\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)),$$
其约束条件为

$$\begin{cases} \sum_{i=1}^n a_i - \sum_{i=1}^n a_i^* = 0, \\ 0 \leq a_i^* \leq C, \\ 0 \leq a_i \leq C. \end{cases}$$

核函数 $(\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j))$ 确定了从低维到高维的映射方式,可以将低维线性不可分问题转化为高

维线性可分问题.目前被广泛使用的核函数有径向基核函数(RBF)、线性核函数、Sigmoid 核函数、多项式核函数等.在本文使用的标定方法中,采用径向基核函数最为合适^[12],故使用径向基核函数建立支持向量机模型^[13],其数学模型为

$$K(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2 / (2\sigma^2)),$$

其中 σ 为核宽度,它与学习样本的输入空间范围或宽度有关,样本输入空间范围越小, σ 取值越小.由此可以获得最终的回归表达式为

$$f(\mathbf{x}) = \bar{\mathbf{b}} + K(\mathbf{x}_i, \mathbf{x}_j) \sum_{i=1}^n (a_i - a_i^*).$$

通过实验表明,该模型的预测精度较大程度上取决于惩罚参数 C 和核宽度 σ 的取值.而常规方法是经过人工大量的实验来寻找最优的参数,使用该方法难以找到最优的参数,并且耗时耗力.为解决该问题,本文采用蚁群算法在一定范围内进行搜索最优的参数匹配,从而提高模型预测的精度.

2 使用蚁群算法对支持 SVM 模型参数优化

蚁群算法模仿蚂蚁觅食的过程.每只蚂蚁都会在寻找食物的路径上留下“信息素”,其他蚂蚁会沿着“信息素”浓度高的路径行走并继续留下“信息素”,经过一段时间,蚁群就可以找到距离食物最近的路径.蚁群算法是一种自组织的算法,具有并行性、正反馈、启发式搜索等特征,它对初始路线要求低,并且无需人工干预、不易陷入局部最优等优点,易于得到优化问题的解^[13].因此,本文使用蚁群算法对支持向量机进行优化^[14],把支持向量机的参数选取视为参数的组合优化,建立目标函数处理参数的组合,然后在该函数中使用蚁群算法即可搜索出最优的参数值^[15].具体流程如图 1 所示.

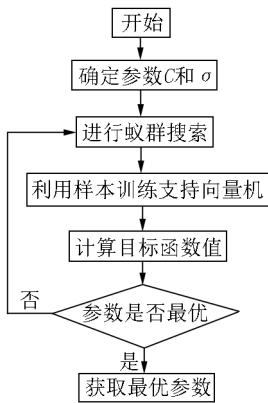


图 1 基于蚁群算法的支持向量机参数优化流程图

2.1 初始化相关参数

由于本文使用径向基核函数建立支持向量机模型,预测精度较为依赖惩罚参数和核宽度的取值,需要给蚁群中每只蚂蚁赋予随机的惩罚参数 C 和核宽度 σ ,并设定蚁群大小 M 、循环终止次数 K 、信息素挥发损失系数 ρ 等.

2.2 路径选择

对于任意一只蚂蚁,为其定义一个 1 维数组 K_{nots} 用于存储所有经过的节点,其他蚂蚁在选择节点时将不会选择这些节点. 蚂蚁在第 i 个节点选择第 j 个节点时将使用轮盘赌的方式,并将第 j 个节点存入 K_{nots} 中. 第 s 只蚂蚁选择第 j 个节点的概率为

$$P_{ij}^s(t) = \begin{cases} \frac{(\tau_{ij}(t))^\alpha (\eta_{ij}(t))^\beta}{\sum_{g \in a_{lloved}(k)}^n (\tau_{ig}(t))^\alpha (\eta_{ig}(t))^\beta}, & j \in a_{lloved}(k), \\ 0, & \text{其他,} \end{cases}$$

其中 τ_{ij} 是第 i 个节点到第 j 个节点路径上的信号素强度,该条路径走过的蚂蚁越多则信号素强度越高. α 、 β 分别为信息素启发因子和期望启发式因子, α 的值越小,蚂蚁寻找新的路径的概率越大,重复走之前路径的概率越小; β 的值越小,蚂蚁选择局部最短路径的概率越小,也越不易陷入局部最优. $a_{lloved}(k)$ 表示蚂蚁在接下来可选择的且未经过节点的集合.

2.3 信息素更新

蚂蚁在走过某一条路径后,将会在这条路径上留下信息素 τ_{ij} ,它反映了先前蚁群经过这条路径的频率. 信息素会随着时间的挥发损失,其更新公式为

$$\tau_{ij} = (1 - \rho)\tau_{ij} + \Delta\tau_{ij},$$

其中 ρ 中为信息素挥发损失系数,取值在 0 和 1 之间. $\Delta\tau_{ij}$ 为这条路径上信息素的增加量,其表达式为

$$\Delta\tau_{ij} = \begin{cases} Q/L_k, & \text{第 } k \text{ 只蚂蚁在此次觅食过程中经过} \\ & \text{路径}(i,j), \\ 0, & \text{其他,} \end{cases}$$

其中 Q 为普通数, L_k 表示第 k 只蚂蚁在本次觅食过程中走过的距离.

2.4 终止条件

若迭代次数达到初始设置的循环终止次数 K 且蚁群尚未收敛到同一条路径,则将蚁群重置于起点处并重新开始迭代. 若迭代次数达到初始设置的循环终止次数 K ,或者迭代次数未达到循环终止次数 K 但蚁群会收敛到某一相同的路径,则停止搜索并得到蚁群的这条路径即为最优路径,并输出最优参数.

3 标定实验与分析

3.1 实验方法

本文分别使用华为 Mate10pro 手机作为参考设备、华为 Mate8 手机作为用户设备采集蓝牙 iBeacon 发出的蓝牙信号,实验场地选择江西师范大学室内定位实验室. 将室内环境作为采集区域,并分为若干场景,在每个场景中分别挑选若干个采集点,将这些采样点进行编号,并使用每款手机分别在每个采样点对能搜索到的各个 AP 的 RSSI 值进行采样. 由于蓝牙信号在传输中可能会存在波动以及受到环境的影响,因此本文在每个距离点分别采样 100 次,并记录下每个采样点的平均信号强度以及位置用于模型训练. 对于手机在采样点 P 采样的数据,其格式为

$$\{P_i, \{ (AP_1, RSSI_1), \cdots, (AP_k, RSSI_1), \cdots, (AP_m, RSSI_1) \}, \cdots, \{ (AP_1, RSSI_j), \cdots, (AP_k, RSSI_j), \cdots, (AP_m, RSSI_j) \}, \cdots, \{ (AP_1, RSSI_n), \cdots, (AP_k, RSSI_n), \cdots, (AP_m, RSSI_n) \} \}.$$

3.2 参数设置

蚁群算法的各初始值如下: 蚁群规模 $M = 1\,000$, 信息素挥发损失系数 $\rho = 0.5$. 支持向量机的各参数初始值范围如下: $C \in (0, 1\,000)$, $\sigma \in (0.1, 100)$.

3.3 采样点选择

分别使用每个采样点蓝牙的平均观测量数据建模,模型分别使用线性模型和支持向量机预测模型,并通过在实验室的其他区域采样观测值进行验证模型的有效性,得到较稳定的模型并确定最小可以建立模型的采样点数. 每种模型使用多个点建模的误差减小情况如表 1 所示.

表 1 2 种模型在使用不同采样点数下的误差减小统计

点数	模型/%	
	线性模型	支持向量机预测模型
2	31.70	45.50
3	48.50	55.10
4	78.30	82.70
5	80.40	85.40
6	82.10	87.70
7	83.80	88.20
8	84.70	88.50

由表1可知,在不同的采样点数量下标定效果随着采样点数的增多而提高.当采样点数小于4个时,标定模型对标定效果的提升效果一般.当采样点数大于5个时,随着点数的增加,标定效果提升很有限,而且随着采样点数的增加会使得工作量增加.当采样点数为6个左右时模型可减小82.1%和87.7%的误差,可大幅度提高标定效果.因此,本文在每个场景中分别使用6个采样点的RSSI平均值进行建模.

3.4 实验结果

将用户设备在某个采集点所有AP的RSSI平均值作为自变量 x_i ,并将参考设备在该采集点对应AP的RSSI平均值作为因变量 y_i ,并通过python代码建立模型.将与线性模型和支持向量机模型进行对比,结果如图2和图3所示.

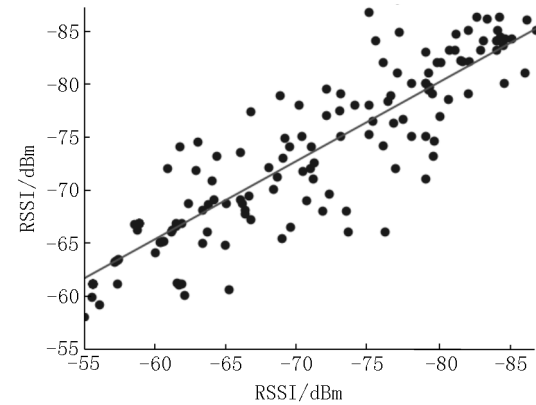


图2 线性标定模型

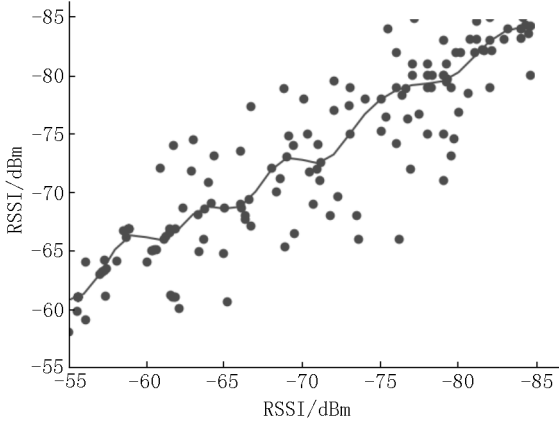


图3 支持向量机模型

图2为经线性模型标定后的信号强度与参考设备之间蓝牙RSSI观测值的函数关系,图3为经支持向量机预测模型标定后的华为Mate8手机与参考设备之间蓝牙RSSI观测值的函数关系.为验证模型的有效性,分别使用2种模型对不同品牌不同型号的智能手机在用户定位时的蓝牙RSSI观测值进行实

时校正,校正结果如表2所示.使用校正结果进行定位蓝牙指纹定位实验,定位误差结果如表3所示.

表2 蓝牙RSSI观测值校正结果表 dBm

设备	原始数据	线性方法校准	本文方法校准
	平均差值	后平均差值	后平均差值
M5 note	13.2	6.4	2.2
Honor8	8.4	6.6	2.0
Mate8	13.1	10.2	3.9

表3 蓝牙定位精度对比 m

	未标定	线性模型	SVM模型
平均误差	3.24	2.58	2.03

由表3可知,本文提出支持向量机模型标定方法相比未标定方法和线性模型标定方法平均误差分别减小了37.3%和21.3%.因此,使用本文提出的模型可以有效减小由于设备异构性造成的误差,从而提高室内定位精度.

4 结束语

本文提出了基于蚁群算法的支持向量机蓝牙标定方法.支持向量机方法以统计学理论为基础,但是其预测精度较大程度上取决于惩罚参数 C 和核宽度 σ 的取值,因此考虑引入蚁群算法优化支持向量机的参数选取.通过采集数据,得到信号强度数据并训练模型,最后使用该模型对采集的信号强度结果进行校正.并且通过实验,对比该模型与未标定以及线性模型的标定结果,定位误差分别减小了37.3%和21.3%,这说明此模型精度更高,在蓝牙标定方面效果较好,可有效提高室内定位精度.

5 参考文献

[1] 陈锐志,陈亮. 基于智能手机的室内定位技术的发展现状和挑战[J]. 测绘学报,2017,46(10):1316-1326.

[2] 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):14-23.

[3] 宋斌斌,余敏,何肖娜,等. 一种BP神经网络的室内定位WiFi标定方法[J]. 导航定位学报,2019,7(1):43-47.

[4] Haeberlen A, Flannery E, Ladd A M, et al. Practical robust localization over large-scale 802.11 wireless networks[EB/OL]. [2019-09-13]. <https://dl.acm.org/citation.cfm?id=1023728>.

[5] Kjaergaard M B. Automatic mitigationof sensor variations for signal strength based location systems [EB/OL]. [2019-09-13]. https://dl.acm.org/doi/10.1007/11752967_3.

[6] 刘涛,王华. 传感器非线性校正的遗传支持向量机方法 [J]. 电子测量与仪器学报,2011,25(1):56-60.

[7] 张学工. 关于统计学习理论与支持向量机 [J]. 自动化学报,2000,26(1):36-46.

[8] 闫国华,朱永生. 支持向量机回归的参数选择方法 [J]. 计算机工程,2009,35(14):218-220.

[9] 梁礼明,陈明理,刘博文,等. 基于分形理论的支持向量机核函数选择 [J]. 江西师范大学学报:自然科学版,2019,43(3):309-313,319.

[10] 史恩秀,陈敏敏,李俊,等. 基于蚁群算法的移动机器人全局路径规划方法研究 [J]. 农业机械学报,2014,45(6):53-57.

[11] Vapnik Vladimir N. The nature of statistical learning theory [M]. 2nd ed. New York:Springer-Verlag,2000.

[12] 赵琳. 基于 RSSI 的核函数机器学习室内定位技术 [D]. 南京:南京邮电大学,2018.

[13] 林升梁,刘志. 基于 RBF 核函数的支持向量机参数选择 [J]. 浙江工业大学学报,2007,35(2):163-167.

[14] 庄严,白振林,许云峰. 基于蚁群算法的支持向量机参数选择方法研究 [J]. 计算机仿真,2011,28(5):216-219.

[15] Matthias, Wittlinger, Rüdiger, et al. The ant odometer: stepping on stilts and stumps [J]. Science, 2006, 312(5782):1965-1967.

The Indoor Positioning Method with Bluetooth Calibration of Supporting Vector Machine Based on Ant Colony Algorithm

WU Xuan¹, XUE Feng², YU Min^{1*}

(1. College of Software, Jiangxi Normal University, Nanchang Jiangxi 330022, China;
2. College of Computer and Information Engineering, Jiangxi Normal University, Nanchang Jiangxi 330022, China)

Abstract: For hardware differences between different models of smartphones, it can occur a deviation when smart-phone indoor positioning is used to collect the same bluetooth signal strength measurements, it may affect the position-ing accuracy. The indoor positioning bluetooth calibration method of support vector machine(SVM) based on ant colony algorithm is proposed in this paper. Since the selection of SVM parameters has a great influence on its pre-diction accuracy, ant colony algorithm can avoid blindly selecting SVM parameters, optimize the model and improve the prediction accuracy. Experiments show that the accuracy of the proposed method is 37.3% higher than that be-fore calibration, and the proposed method can effectively perform indoor positioning.

Key words: hardware and software heterogeneity; bluetooth calibration; support vector machine; ant colony algo-rithm; calibration model

(责任编辑:冉小晓)