

一种基于神经网络的RSSI定位算法

宗传林 高作相 贺育茹

(长春理工大学,吉林 长春 130022)

摘要:针对商场、地下室等室内环境GPS室内信号较弱、定位误差较大的问题,本文提出了一种基于神经网络的RSSI定位算法,通过对基本的BP算法加以改进,优化其功能,最后进行实验验证。

关键词:神经网络;指纹库;RSSI;K聚类算法

中图分类号:TP212.9

文献标识码:A

文章编号:1003-5168(2020)13-0024-03

An RSSI Localization Algorithm Based on Neural Network

ZONG Chuanlin GAO Zuoxiang HE Yuru

(Changchun University of Science and Technology Changchun, Jilin 130022)

Abstract: Aiming at the problem of weak indoor signal and large positioning error of GPS in shopping malls, basements and other indoor environments, this paper put forward a RSSI positioning algorithm based on neural network, improved the basic BP algorithm, optimizes its function, and finally carried out experimental verification.

Keywords: neural network; fingerprint library; RSSI; K clustering algorithm

随着互联网和物联网的高速发展及无线局域网的普及,人们越来越看重信息传递的准确性和即时性。在该背景下,产生了一个新型的服务种类,即基于位置服务(Location Based Services, LBS),为人们提供了巨大的便利。LBS在日常生活中主要体现在室外定位的应用上,如利用全球定位系统(Global Positioning System, GPS)、北斗卫星的道路导航、辅助定位服务等;但在室内,因障碍物过多、过厚,导致卫星信号无法穿透,加上卫星的多径效应等因素,严重影响了室内定位的精准性。通常,GPS室外定位精度能达到3~5 m,甚至可以更精确,但在结构较为复杂的室内,误差能够达到10 m。目前,室内的定位技术主要有蓝牙定位技术、射频定位技术、无线保真(Wireless Fidelity, WiFi)定位技术、红外定位技术以及超声波定位技术等^[1]。其中,WiFi定位技术又包括信道定位技术和信号强度(Received Signal Strength Indication, RSSI)定位技术。在我国,通信技术十分发达,在室内外公共场所,有着极高的WiFi覆盖率,使得WiFi室内定位具有良好的优势;同时,WiFi定位技术不需要额外的传感器,与GPS相比,速度提高了10倍以上^[2],且通过软件开

发就能实现定位,因此WiFi定位技术发展较快。

本文结合梯度下降法原理和神经网络编写了一套定位算法,利用无线接入点(Access Point, AP)与用户之间线性与非线性的关系,解决了在传统AP定位中用三个AP求解三角形法无法解决的多径效应,也解决了传播过程中依赖传播衰减公式带来的较大误差。同时,给出了对模型的奖励与惩罚方法,并引入了基本K聚类算法,以提高模型精度。在离线阶段给出了不同数据量情况下选择不同的查找方法以提高查找效率,最后通过在特定场景下的简要实验,得出此场景下的选择参数,其最小误差为1.42 m。

1 定位系统流程和相关原理

1.1 定位系统流程

定位系统流程图如图1所示。从图1可以看出,离线阶段:初始程序,读取各点RSSI值并进行迭代,将读取的值保存到指纹库中,直至迭代完成;在线阶段:通过将读取的RSSI值与指纹库中的指纹对比找出对应位置,并再次依次迭代。

收稿日期:2020-04-10

基金项目:吉林省大学生创新创业训练大赛“室内导盲机器人”(201910186088)。

作者简介:高作相(1998—),男,本科在读,研究方向:单片机控制;贺育茹(1999—),女,本科在读,研究方向:信号的检测与处理。

通信作者:宗传林(1998—),男,本科在读,研究方向:机器学习在通信系统中的应用。

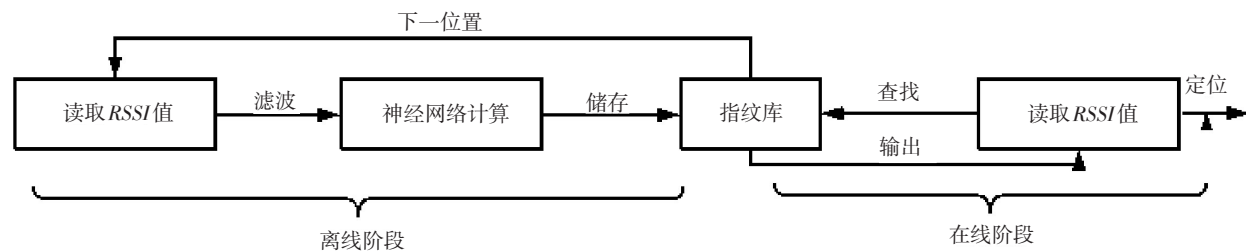


图1 系统流程图

1.2 基本原理

通常情况下,人工神经网络(Artificial Neural Networks, ANNS)主要用来处理非线性问题,例如,在信号传播过程中涉及多种噪声、衰减等非线性问题。因此,采用ANNS中的误差反向传播算法(Error Back Propagation, BP)具有很大的优势。

在处理数据的过程中,研究者采用离差标准化方法对数据进行归一化处理,函数形式为:

$$X_n = (X_n - X_{\min}) / (X_{\max} - X_{\min}) \tag{1}$$

式中, X_{\min} 为数据中最小值; X_{\max} 为数据中最大值。

BP算法中隐层节点数对定位精度具有很大的影响:隐层节点数过少,训练精度过低,模型欠拟合;隐层节点数过多,训练时间加长,模型过拟合也会导致精度降低。推荐隐层节点数可由经验公式(2)计算得出:

$$ne th = \sqrt{(in + out)} + \beta \tag{2}$$

式中, $ne th$ 为隐层节点数; in 为输入层节点数; out 为输出层节点数; β 为0~10的常数。

1.3 指纹库相关算法原理

指纹库相关算法包括快速排序、归并排序和堆排序。快速排序法:先从序列中取一个值作为标准值,将小于其的值放在左边,大于其的值放在右边,以此进行迭代至全部排序完成。归并排序法:采用分治法,其方法将序列分成两个序列,再将得到的两个子序列继续二分下去,直到不能再分,最后再两两合并,形成有序序列。堆排序:利用堆这种数据结构所涉及的一种排序算法,其原理为堆积是一个近似完全二叉树的结构,并满足堆积性质:子节点的键值或索引值总是大于(或者小于)父节点。此外,采用快速排序法平均只需要 $O(n \log n)$ 的复杂度。在相同数据规模下(见表1),快速排序法的效率较高。

表1 排序效率

数据量	效率/s		
	快速排序	归并排序	堆排序
1 000 万	0.73	1.26	3.58
5 000 万	3.68	6.24	26.58
1 亿	7.50	13.08	61.37

1.4 K聚类算法

K聚类算法是聚类算法的一种,其步骤为:第一,随机选取 K 个中心点;第二,计算所有数据点距 K 个中心

点的距离,然后将数据点分到与其最近的中心点,形成 K 个簇;第三,计算每个簇中所有数据的均值,重新定义中心点;第四,以此进行迭代,直至中心点位置变化小于阈值,划分结束。

2 神经网络训练技术指标

2.1 传递函数的选择

本文选取Sigmoid函数作为传递函数,其中:

$$f(x) = \frac{1}{1 + e^{-\alpha x}} \tag{3}$$

式中, $f(x)$ 为因变量; x 为自变量; α 为倾斜参数,常被默认为1。使用Sigmoid函数(见图2)作为传递函数的原因:其输出范围有限,Sigmoid函数的输出被固定在(0,1),数据在传递过程中不容易发散。Sigmoid函数在(0,1)中存在单调、处处可导等优良性质,在(0,0.5)处斜率最大,且中心对称,两头被抑制明显,因而适合数据在多层神经元之间传递时使用。同时,神经元中间激活,两头克制的表现也与神经科学中的激励与抑制行为十分相似,能够更好地体现神经网络的特性^[3]。

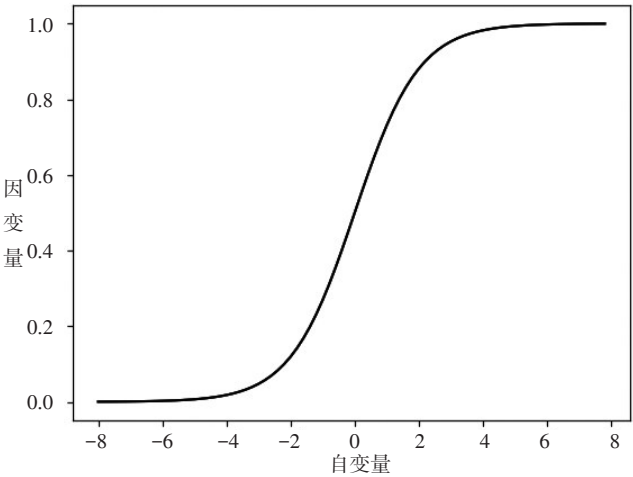


图2 Sigmoid函数

2.2 过拟合与欠拟合

在学习过程中模型会出现过拟合与欠拟合的问题。欠拟合指对样本训练程度不够,导致误差较大。由维度较低的特征或者模型过于简单引起,通常用增加特征维度与增加训练样本来解决;过拟合指训练程度过高,特征维度较高,模型过于复杂引起,常用正则化的方法来解决。

有监督学习算法训练的目的是最小化损失函数,本文引入均方误差损失函数,其预测值与样本真实值的误差平方和为:

$$L(\theta)=\frac{1}{2l}\sum_{i=1}^l[f_{\theta}(x_i)-y_i]^2$$
 (4)

其中, y_i 是样本标签值; $f_{\theta}(x_i)$ 是预测函数的输出值; θ 是模型参数; l 为样本数。

为了防止过拟合,可以对复杂的模型进行惩罚,加上一个惩罚项,使模型参数尽可能小。加入惩罚项后的损失函数为:

$$L(\theta)=\frac{1}{2l}\sum_{i=1}^l[f_{\theta}(x_i)-y_i]^2+\lambda R(\theta)$$
 (5)

式中, λ 为惩罚项系数; $R(\theta)$ 为惩罚函数,后半部分为正则化项,正则化项可以使用L2范数,L2范数相对于L1范数,在求解最优化问题有着更好的数学性质。

并且惩罚项中L2范数正则化项的梯度为:

$$\lambda \theta$$
 (6)

3 解决的问题与实验简要分析

3.1 面临的问题

随着WiFi的普及以及机器学习的高速发展,室内定位技术如雨后春笋般层出不穷,如可见光定位技术、无线定位技术等。但是,都面临着如何解决定位精度的问题,本文只对自己所编写BP算法和聚类算法进行阈值的选取。

3.2 样点的选择

在一个长10 m、宽8 m的空间内等位置放置6个AP信号(见图3),在长、宽分别为0.05 m的地方选取32 000个点,将其每点作为输入层,输入神经网络中,并将数据保存到指纹库中;另随机再选取32 000个点做同样的处理,对其在线阶段测得位置与其最相近的点取均方差,训练次数2 000次。采样方式比较如表2所示。

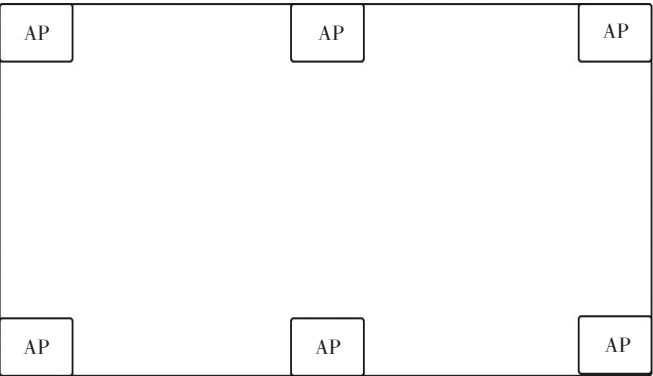


图3 实验布局

在采样方式上,等距取样随着点数的增加误差逐渐

减小,但运算时间相对于非等距取样时间要长(见表2)。等间距取样更有利于定位准确性,因此后续实验采用等距取样的方法。

表2 采样方式比较

取样方式	误差	时间/s	最大距离占比/%
等距	0.018 925	1 268	75.7
非等距	0.106 856	968	87.5

3.3 实验与分析

再次选取32 000个取样点做测试,在离线阶段测得各个点的RSSI值,将其保存在指纹库内。在线阶段时再对每个点RSSI依次运算做比较,结果为:正确点数为29 865个,错误点数为2 185个。

依次取不同样本点数,如30 000、25 000、15 000、5 000、2 000再次进行测量,测得不同误差下的概率如表3所示。

表3 采样误差率

平均定位误差	误差在1 m内概率/%	误差在4 m内概率/%	最大误差
1.82	33.6	93.4	4.05

误差的大小随采样点数的减少而增大,而较多的采样点数会增加计算与查找时间。为此,研究者需要确定一个合适的点数来满足工程中一定的条件,而时间大部分用于程序的训练。当采样点数为5 000次时,训练次数分别为1 000、2 000、3 000、4 000,比较运行时间与误差均值的大小可知,在4 000次下有最小的误差。以4 000次训练次数为准,分别取20 000、15 000、10 000、8 000、6 000、4 000、2 000个点。经过多次模拟可知,当采样点数为6 000时具有最小的误差,其值为1.42 m。

4 结论

本文对实验所需的参数,如采样点数、采样方式、学习次数以及机器参数等,作出一个合理的选择,得出良好的定位效果;但因实验装备精度、运算能力及环境等因素的影响,无法推广到所有的应用场合中。在后续的研究中,重点放在AP取点位置、K聚类算法的优化、对RSSI滤波的最优化等问题的研究上。

参考文献:

[1]王建林.基于WIFI和多用户组群的室内定位系统[D].西安:西安电子科技大学,2017.
[2]郑文翰.基于WLAN的室内定位技术研究 with 实现[D].成都:电子科技大学,2016.
[3]王亚敏.全卷积神经网络在眼底图像结构中的应用研究[D].成都:电子科技大学,2019.