

# 基于最小二乘法和 BP 神经网络的 TOA 定位算法

浦佳祺<sup>1,2</sup>, 陈德旺<sup>1,2</sup>

(1.福州大学 数学与计算机科学学院 福建 福州 350108;

2.福州大学 轨道交通研究院 福建 福州 350108)

**摘 要:** 针对室内复杂环境而导致低精度定位的问题,提出了一种新的基于信号到达时间的 TOA 定位算法。对最小二乘法引入了线性误差分析项,并且使用 BP 神经网络的训练方法取代传统方案进行测距模型的建立,以消除对环境的过度经验依赖,提高算法针对不同环境的普适性。选取均方根误差作为性能评价指标,将该方法和具有代表性的两种算法,即传统的最小二乘法和经典的 Chan 算法进行了比较,利用实测的五组场景下的总计 6 000 组定位数据进行了训练与验证。实验结果表明,该算法能取得比 Chan 算法与最小二乘估计法这些经典算法优异的性能,使得定位精度大大增加。并且通过交叉实验,也说明了在不同场景下该模型具有一定的通用性。

**关键词:** 到达时间; 最小二乘法; BP 神经网络; 交叉实验

中图分类号: TP301.6

文献标识码: A

文章编号: 1673-629X(2018)05-0005-04

doi: 10.3969/j.issn.1673-629X.2018.05.002

## A TOA Positioning Algorithm Based on Least Square Method and BP Neural Network

PU Jia-qi<sup>1,2</sup>, CHEN De-wang<sup>1,2</sup>

(1.School of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China;

2.Academy of Rail Transport, Fuzhou University, Fuzhou 350108, China)

**Abstract:** In view of the problem of localization with low accuracy caused by complex indoor environment, we propose a new localization algorithm based on time of arrival (TOA). The linear error term is introduced in the least square method, and the BP neural network training is used to create the ranging model instead of the traditional scheme, so as to eliminate the over-experience dependence on the environment, improve the universality of the algorithm for different environments. In this paper, we select the root mean square error as the performance evaluation criterion, by which we perform a comparative testing of the proposed method and the two representative algorithms, traditional least squares method and classical Chan algorithm. By testing more than 6000 records in five different environments, the results show that the improved algorithm is able to obtain the better performance than the Chan algorithm and the least squares estimation method, and enhances the localization accuracy greatly. In addition, cross-experiments show that the models in different scenarios is universal.

**Key words:** TOA; least square estimation; BP neural network; cross experiment

### 1 概 述

在户外的应用场景中, GPS 定位技术已经得到了广泛的使用<sup>[1-2]</sup>。而到了室内, 由于室内遮挡物对信号的影响, 另一方面室内定位的精度要求比室外高, 因此无法满足室内定位的需求。所以室内无线定位技术受到了重视并有了长足进展, 例如: RFID 定位<sup>[3]</sup>、超声波定位<sup>[4]</sup>、WIFI 定位<sup>[5-6]</sup>等。

从技术角度来看, 现代商用通信网络对于三维定位的需求, 是使用尽可能少的基站完成对终端设备的定位、算法收敛速度快、对干扰和噪声具有鲁棒性等。相比于 GPS 等商用卫星定位系统, 基于通信基站的定位问题具有如下特殊性: 通信基站的目标区域是 GPS 等卫星定位系统无法实现定位的场景。在高楼林立的城区、建筑物内部、地下停车场等区域, GPS 等系统是

收稿日期: 2017-05-19

修回日期: 2017-09-20

网络出版时间: 2018-02-08

基金项目: 福建省闽江学者科研启动项目(510146); 工信部倍增计划(00101522); 福州大学人才引进科研启动项目(0030-510206)

作者简介: 浦佳祺(1994-), 男, 硕士研究生, 研究方向为城市智能等; 陈德旺, 教授, 博导, 从事交通信息工程及控制、机器学习、计算智能、最优化及其应用和智能交通系统等研究。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20180207.1906.048.html>

无法满足定位需求的。而这些应用场景基站、终端密集,是基站定位可以实现突破的地方。

常用的定位技术有到达时间(TOA)、到达时间差(TDOA)、角度到达(AOA)和接受信号强度(RSSI)等<sup>[7]</sup>。其中对 TOA 模型的解法有最小二乘法<sup>[8]</sup>等经典算法。而对 TDOA 模型,有 Chan 算法<sup>[9]</sup>等经典算法进行求解。对 RSSI,则有 BP 神经网络对其进行研究<sup>[10-11]</sup>。

由于通信基站所处的电磁信号环境较之 GPS 等系统更加复杂,以室内环境为例,无线电信号的传播过程中会经过墙面的多次反射、室内物体的折射和吸收等。这些物理因素会导致通信基站测量得到的诸如距离、角度等信息存在噪声而产生非视距的环境。基于这些有噪声的测量,得到对于位置信息的准确估计,也是通信基站实现对终端定位需要解决的问题。对于 TDOA 的实现算法主要有 Chan 算法、Fang 算法<sup>[12]</sup>和 Taylor 算法<sup>[13]</sup>。而这些算法的不足之处在于,Fang 算法不能充分利用多个基站提供的冗余信息;Taylor 算法需要一个与实际位置接近的初始值,并且运算量较大;Chan 算法虽然计算快捷,但是在非视距(NLOS)环境下定位性能显著下降。而在改进方法中,有一种方法是基于 Elman 神经网络,对 NLOS 误差进行修正,再利用 Chan 算法定位<sup>[14]</sup>。

对此,文中提出了一种在非视距环境下的改进算法。先通过最小二乘估计来减小 NLOS 环境中数据的噪声,并且获得一个与实际位置接近的初始值。之后运用 BP 神经网络通过训练学习,再次优化所得到的结果,从而获得最终结果。

## 2 算法说明

该算法考虑从减少 TOA 的噪声方面入手,主要由两部分组成。第一部分通过最小二乘法先去除部分 NLOS 环境下的噪声,第二部分通过 BP 神经网络训练对最后的模型进行修正。

### 2.1 误差项估计

考虑在非视距环境下,在极大似然估计法中基站  $BS_i$  到移动端 MS 的距离  $d_i$  存在多种噪声误差,包括时间同步噪声、非视距环境噪声等,但是其主要因素应该为非视距环境噪声。TDOA/TOA 测量误差的正均值通常具有随移动台与基站之间距离线性增加的趋势<sup>[15]</sup>。所以考虑在极大似然估计法中引入一个线性误差项:

$$d_i = d_{i, \text{real}} + M \cdot d_i + N \quad (1)$$

其中,  $d_{i, \text{real}}$  表示第  $i$  个基站  $BS_i$  到移动端 MS 的真实距离;  $M \cdot d_i + N$  为一个线性误差项。则最小二乘法的公式转化为如下形式:

$$\begin{cases} (x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2 = (d_1 - Md_1 - N)^2 \\ (x - x_2)^2 + (y - y_2)^2 + (z - z_2)^2 = (d_2 - Md_2 - N)^2 \\ \vdots \\ (x - x_n)^2 + (y - y_n)^2 + (z - z_n)^2 = (d_n - Md_n - N)^2 \end{cases} \quad (2)$$

对应的方程组变为:

$$AX = b' \quad (3)$$

其中

$$A = \begin{pmatrix} 2(x_1 - x_2) & 2(y_1 - y_2) & 2(z_1 - z_2) \\ 2(x_1 - x_3) & 2(y_1 - y_3) & 2(z_1 - z_3) \\ \vdots & \vdots & \vdots \\ 2(x_1 - x_n) & 2(y_1 - y_n) & 2(z_1 - z_n) \end{pmatrix} \quad (4)$$

$$X = (x \ y \ z)^T \quad (5)$$

$$b' = \begin{pmatrix} K_1 - K_2 + D_2^2 - D_1^2 \\ K_1 - K_3 + D_3^2 - D_1^2 \\ \vdots \\ K_1 - K_n + D_n^2 - D_1^2 \end{pmatrix} \quad \begin{matrix} K_i = x_i^2 + y_i^2 + z_i^2 (i = 1, \\ 2, \dots, n) \end{matrix} \quad \begin{matrix} D_i = d_i - Md_i - N (i = 1, 2, \dots, n) \end{matrix} \quad (6)$$

在保证基站个数大于 6 的情况下,求解上面的式子可以得到初始值  $(x \ y \ z)$  和线性误差项的系数  $M, N$ ,这样就获得了所需要的数据。

### 2.2 基于 BP 神经网络的定位误差修正

BP 神经网络的多层前馈网络应用最为普遍的是单隐层网络,已经证明单隐层的神经网络可以模拟任何连续函数。考虑定位模型的复杂度并不是太大,并且经过实验验证后,选取单隐层的三层神经网络结构作为建立的模型。

基于 BP 神经网络的定位误差测量值修正模型,输入层 3 个神经元的输入为前面一部分用带误差项的极大似然估计得到的定位的三维坐标,输出层也由 3 个神经元构成,为修正后得到的最终定位的三维坐标。其中有一个单隐层,  $w_{ij}$  为输入层与隐藏层之间的权值矩阵,  $w_{jk}$  为隐藏层与输出层之间的权值矩阵。之后通过训练来修改对应的权值,以逼近想要的连续函数。

## 3 数据集和算法参数设置

### 3.1 实验平台及数据集说明

文中使用的软件为 Matlab R2014a。数据集来自华为公司提供的 TOA 定位数据。每个场景下的数据集包括:该场景下基站的个数,每个基站的三维坐标,每个终端在该场景中每个基站记录的 TOA 时间。

测试数据包括 5 组不同场景下的 TOA 数据,给定的基站的  $x \ y$  轴的范围是  $-400 \sim 400 \text{ m}$ ,  $z$  轴的范围为

0~6 m。分别给 20, 30, 40, 50, 60 个基站的 5 组数据, 对应每个场景分别包含 1 000, 1 100, 1 200, 1 300, 1 400 个移动端定位信息。

### 3.2 定位流程

首先对给定的数据进行预处理, 用极大似然法, 获得一个初步定位后的三维坐标。然后输入 BP 神经网络进行训练, 获得最终的定位模型。

BP 神经网络的实验利用 Matlab 的神经网络工具箱, 在测试大量的参数(如隐层的节点数、隐层的个数等)后, 实验参数设置为: 隐层激活函数 sigmoid, 隐层节点个数 10, 单隐层。

对于 5 组不同场景下的数据, 把这些数据抽样随机划分为两个独立的数据子集, 一部分作为训练数据集, 占总数据的 70%, 另一部分作为验证数据集, 占 30%。

## 4 实验结果分析与比较

为了验证实验结果, 利用统计学中的均方根误差来检验模型的有效性。

RMSE =

$$\sqrt{\frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (z_i - \hat{z}_i)^2}{n}} \quad (7)$$

其中,  $x_i, y_i, z_i$  为最后每个定位点计算定位得到的三维坐标;  $\hat{x}_i, \hat{y}_i, \hat{z}_i$  为该定位点的实际三维坐标;  $n$  为总的测量个数。

首先对加入误差项的极大似然估计法与未加入误差项的极大似然估计法之间的均方根误差进行对比, 如表 1 所示。

表 1 极大似然估计法误差对比

	不加入误差项/m	加入误差项/m
第一组	6 193.303 9	15.050 9
第二组	474.690 3	31.367 8
第三组	1 589.451 4	20.117 5
第四组	381.678 7	19.068 7
第五组	3 374.461 1	41.951 2

从表 1 可以看出, 加入线性误差项后的定位精度大大增加。不过定位误差还是超过 10 m 以上, 仍不满足一般情况下的定位要求, 所以有进一步减小误差的可能。

之后对上一步得到的结果随机划分为 70% 的训练集和 30% 的测试集, 每个场景分别对这些训练集进行 10 次训练, 训练结果的均方根误差如图 1 所示。

可以看到, 对于训练结果, 除了第五组的场景, 其他场景的训练误差都小于 0.5 m。而就算是训练结果最差的第五组, 其训练误差也不超过 1 m, 说明训练效

果较好。并且对每个场景而言, 10 次训练结果的波动范围不大, 说明训练有较好的稳定性。同时通过对场景一中训练集的第 1 次测试结果的误差分布可知, 场景一中的测试集的点数总计为 770 个, 其中有超过半数的三维坐标误差平方和小于 0.5 m, 大致有 600 个点的误差平方和小于 1, 说明 BP 神经网络对于误差的训练十分之有效。

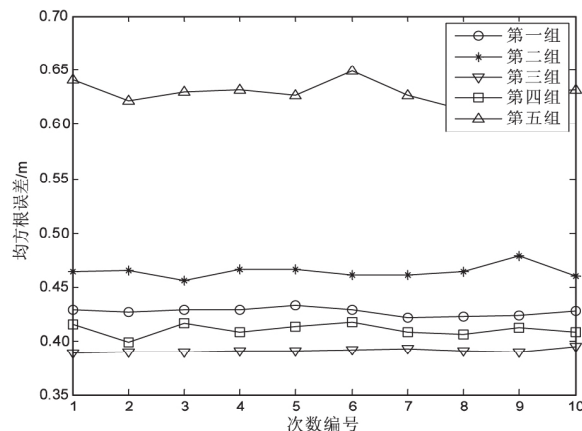


图 1 各个场景下 BP 神经网络训练后的误差

对于每个场景, 每组实验得到的测试结果如图 2 所示。可以看到, 测试结果与训练结果基本上相近。除了第五组的测试误差较大, 其他各个场景的测试误差都比较接近, 最大的测试误差也小于 1 m, 可见 BP 神经网络有效修正了初次定位后的误差。

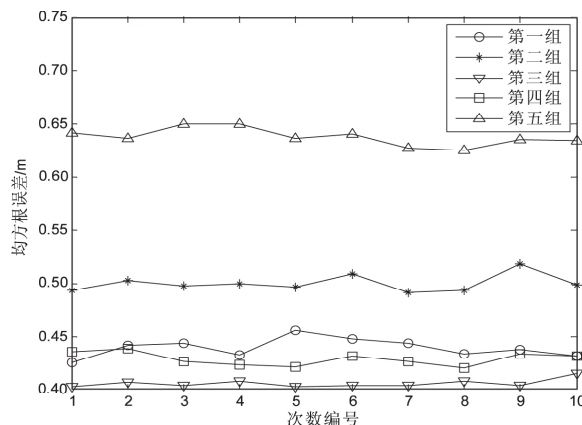


图 2 各个场景下 BP 神经网络测试后的误差

同时通过对场景一中测试集的第 10 次测试结果的误差分布, 场景一中的测试集的点数总计为 330 个, 其中有超过半数的三维坐标误差平方和小于 0.5 m, 大致有 300 个点的误差平方和小于 1。可以看出, BP 神经网络对于误差的训练十分有效。其他各个场景和每次实验的测试数据的误差平方和也与之类似, 大部分的误差值都小于 1 m, 个别测量点的误差大于 1 m 以上, 所以就不一一展示了。

然后将定位算法的实验结果分别与 Chan 算法和加入误差项的极大似然估计法进行了比较。由于 BP 神经网络的随机性, 文中算法的均方根误差使用之前

10 次实验的平均值作为其最后的误差平方和。具体结果如图 3 所示。

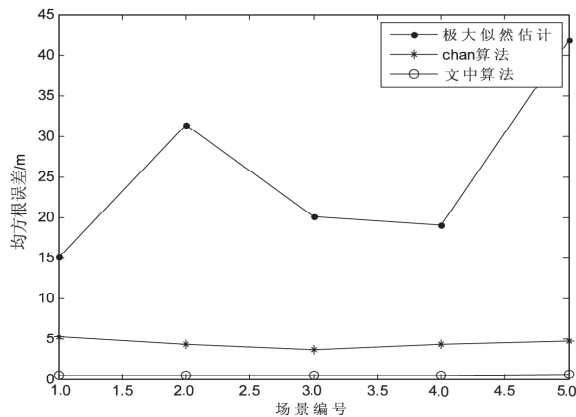


图 3 三种方法的误差比较

三种方法在各个场景下具体误差值的比较如表 2 所示。

表 2 三种定位方法的均方根误差值

	极大似然估计/m	Chan 算法/m	文中算法/m
第一组	15.050 8	5.111 8	0.444 6
第二组	31.367 9	4.329 4	0.500 2
第三组	20.117 4	3.619 7	0.405 7
第四组	19.068 7	4.251 0	0.428 4
第五组	41.951 2	4.588 1	0.638 1

很明显可以看出,文中算法相比经典算法,效果有着显著的提升,能够有效地把误差缩减到 1 m 甚至更低。这应该主要归功于 BP 神经网络强大的连续函数逼近能力,通过训练有效地消除极大似然定位中的误差。

最后对不同场景下测得的模型是否具有通用性进行了测试。对每个场景的模型分别用了另外 4 组场景的测试集进行测试,并且做了对比,选取场景五下的误差,具体如图 4 所示。

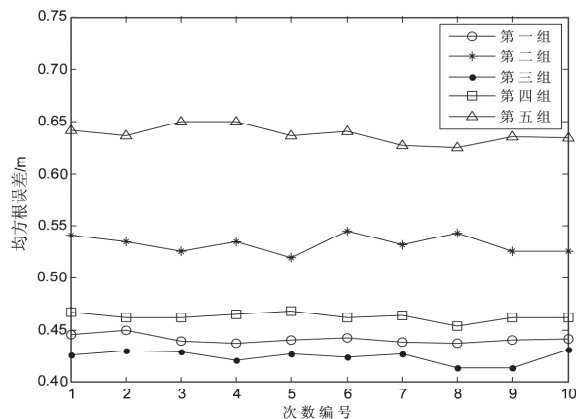


图 4 场景五模型的测试集误差

从图 4 可以看出,在其他场景模型下,即使是不同场景的测试数据最后的定位误差都小于 2 m,与其他两种算法相比,精确度也较高。在通过 BP 神经网络

的训练后,提高了算法针对不同环境的普适性。所以可以通过这种方式来忽略场景的影响。

## 5 结束语

提出了一种基于最小二乘估计和 BP 神经网络的混合定位算法,该算法利用神经网络的强大非线性映射能力,修正在 NLOS 环境下最小二乘法的定位误差,增大了定位精度。实验结果表明,该算法定位精度高,可靠性好,定位性能好于最小二乘法和 Chan 算法,并且具有较强的抗 NLOS 误差能力,提高了针对不同环境的普适性。

## 参考文献:

- [1] BAJAJ R, RANANWEERA S L, AGRAWAL D P. GPS: location-tracking technology [J]. Computer, 2002, 35(4): 92-94.
- [2] PIKANDER M, ESKELINEN P. Differential GPS dynamic location experiments at sea [J]. IEEE Aerospace and Electronic Systems Magazine, 2004, 19(4): 36-39.
- [3] JO M, YOUNG H Y. Intelligent recognition of RFID tag position [J]. Electronics Letters, 2008, 44(4): 308-309.
- [4] 王凡, 彭勇. 基于 TDOA 的室内超声波定位方法的改进 [J]. 计算机技术与发展, 2014, 24(6): 250-252.
- [5] 肖金红, 刘智, 杨阳, 等. 基于 CDMA-TDOA 的室内超声波定位系统 [J]. 吉林大学学报: 信息科学版, 2012, 30(2): 213-217.
- [6] 黄华晋, 秦建息. 基于三角定位算法的 WiFi 室内定位系统设计 [J]. 广西科学院学报, 2016, 32(1): 59-61.
- [7] 范平志, 邓平, 刘林. 蜂窝网无线定位 [M]. 成都: 电子工业出版社, 2002: 80-91.
- [8] 孙瑜. 射频识别 (RFID) 室内定位算法研究 [D]. 成都: 西南交通大学, 2005.
- [9] CHAN Y T, HO K C. A simple and efficient estimator for hyperbolic location [J]. IEEE Transactions on Signal Processing, 1994, 42(8): 1905-1915.
- [10] 王千, 金光, 钮俊. 一种基于 RSSI 的混合定位算法 [J]. 传感技术学报, 2015, 28(12): 1823-1829.
- [11] 兰雪梅, 朱健, 黄承明, 等. BP 网络的 MATLAB 实现 [J]. 微型电脑应用, 2003, 19(1): 6-8.
- [12] FANG B T. Simple solutions for hyperbolic and related position fixes [J]. IEEE Transactions on Aerospace & Electronic Systems, 1990, 26(5): 748-753.
- [13] FOY W H. Position-location solutions by Taylor-series estimation [J]. IEEE Transactions on Aerospace & Electronic Systems, 1976, 12(2): 187-194.
- [14] 吴燕红, 关维国, 王艳峰. 基于 Elman 神经网络的 TDOA 定位算法 [J]. 计算机应用, 2011, 31(3): 629-631.
- [15] 杨洪娟. 蜂窝网络无线定位技术的研究 [D]. 哈尔滨: 哈尔滨工业大学, 2009.