

DOI:10.19322/j.cnki.issn.1006-4710.2021.02.008

# 神经网络定位中多点多映射概率样本增强方法研究

杨 静<sup>1</sup>, 曹秀伟<sup>2</sup>

(1. 西安理工大学 机械与精密仪器工程学院, 陕西 西安 710048;

2. 中建科技集团有限公司深圳分公司, 广东 深圳 518000)

**摘要:** 神经网络室内定位算法的定位精度取决于训练样本的数量和样本对数据特征的代表性, 针对样本采集工作量大的问题, 本文提出了一种多点多映射概率样本的增强方法。该方法首先根据无线信号的传播特性以及空间相关性, 利用少量实测点数据, 快速构建多点反应环境影响的特征, 然后利用概率法生成表征无线信号波动的多映射正态分布样本数据。实验结果表明, 利用多点多映射概率样本增强方法生成的训练样本集, 有效数据量大, 信号表征性好, 且人工采集工作量低, 由此训练的神经网络定位算法定位精度高。

**关键词:** 神经网络; 空间相关; 环境修正; 数据特征; 多映射概率样本

中图分类号: TP393

文献标志码: A

文章编号: 1006-4710(2021)02-0209-06

## Research on construction method for training set in neural network location technology

YANG Jing<sup>1</sup>, CAO Xiuwei<sup>2</sup>

(1. School of Mechanical and Precision Instrument Engineering, Xi'an University of Technology, Xi'an 710048, China;

2. Shenzhenbranch of China Construction Science &amp; Technology Group Co., Ltd., Shenzhen 518000, China)

**Abstract:** The precision of the indoor positioning algorithm basing on the neural network depends on the number and the characterization of training samples. This paper proposes a method for enhancing samples with multi-point and multi-mapping probability feature to simplify the collection work of samples. First, according to the propagation characteristics and spatial correlation of wireless signals, the characteristics of multi-point concerning environmental influences are quickly constructed by small amount of measured data. Then, multi-mapping normal distribution samples indicating fluctuations characterize of wireless signal are generated using probabilistic methods. The experimental results show that the training samples generated by this enhancement method has many advantages such as large amount of effective data, good signal representation and low manual collection workload. Using those training samples to train the neural network can improve the positioning precision.

**Key words:** neural network; spatial correlation; environmental correction; data features; multiple mapping probability samples

随着无线网络、大数据等技术的不断发展, 基于位置服务的应用需求快速增长, 普及到了社会生活和生产的各个领域<sup>[1]</sup>。基于无线信号 RSSI(强度接收指示, received signal strength indication)的定位技术, 由于成本低, 使用灵活等优点得到广泛的应用<sup>[2]</sup>。基于 RSSI 信号定位精度的关键在于如何建立尽可能多的有效的信号强度指示与空间物理位置的映射关系。过去研究者们通常采用 KNN<sup>[3]</sup>等匹

配算法进行定位, 但是却无法保证实时性与精度同时提升。而神经网络凭借自身的易用性、鲁棒性以及强大的非线性拟合能力, 近几年在定位应用领域逐渐受到关注<sup>[4]</sup>。神经网络算法的性能很大程度上取决于训练样本, 在定位算法中, 只有当训练样本足以表征不同位置信号 RSSI 特征, 才能保证网络学习模型其具有泛化能力。根据无线信号传播特点, 在复杂环境中, 人员、墙体以及其他障碍物, 都会给

收稿日期: 2020-07-13; 网络出版日期: 2021-01-05

网络出版地址: <https://kns.cnki.net/kcms/detail/61.1294.N.20210104.1610.004.html>

基金项目: 西安市科技计划资助项目(2017080CG/RCO43(XALU036))

第一作者: 杨静, 女, 博士, 副教授, 研究方向为机电系统检测与控制。E-mail: yjzhd@163.com

RSSI 信号带来非视距以及多径效应的不确定性<sup>[5]</sup>。大量研究表明,利用概率论与数理统计的方法处理 RSSI 信号的不确定性<sup>[6]</sup>,具有较好的抗噪声性能,可极大地去除信号中的噪声,有效提高定位精度。然而,现有的神经网络定位算法利用大量实测 RSSI 数据,通过人工提取数据的均值、方差等特征样本对神经网络进行训练<sup>[7-8]</sup>。这种方法不仅人工成本高,而且不能有效表征无线信号的不确定性。文献[9]提出了一种多信息融合的位置指纹采集方法,利用航位推算与步态分析生成指纹,降低了人工采集工作量,但指纹数据精度较低。2020 年成意等<sup>[10]</sup>提出了群智指纹库建立方法,利用聚类算法指纹数据进行扩展,建立了 RSSI 信号的 Rice 模型,信号的特征性较好,但模型参数受位置影响,运算复杂不利于工程实践推广。本文以推进指纹法在工程实际中的应用为出发点,利用无线信号的传播特性、空间相关性以及信号的 Gauss 分布模型对有限的实测数据进行扩充,从而构建多点多映射概率特征样本。利用扩充样本训练 RSSI 神经网络位置模型,相比少样本的训练模型,可以提高蓝牙 RSSI 的室内定位精度。

## 1 RSSI 信号传播模型与特征

无线信号从发射端经过无线信道到达接收端,根据环境不同,传播过程存在大尺度与小尺度两种衰减特性。其中,大尺度衰减又包括路径损耗和阴影衰落。路径损耗是由信号传播距离以及信号载频导致;阴影衰落主要是由于建筑物、高山等对信号的阻挡导致。小尺度衰减是由于建筑墙体、地板或家具等导致的无线信号多径传播效应。大尺度衰减特性描述信号在大距离(室外环境是 10~100 m,室内更小)上信号强度变化。小尺度衰减特性描述信号在小距离(波长数量级)上信号强度变化,因此,大尺度衰减是影响 RSSI 定位的主要因素。在无线信号传播中,由于障碍物的反射、折射导致信号随机变化,实验结果表明<sup>[11-12]</sup>,在一定空间位置上的 RSSI 信号呈现随机波动,信号强度服从正态分布的分布,其概率密度函数表示为:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (1)$$

式中: $f(x)$ 表示信号强度为  $x$  的概率; $\mu$  为信号均值,与无线信号传播距离以及环境有关; $\sigma$  为信号标准差(方差),描述了信号波动。自由空间(弱干扰空间)无线信号阴影传播模型为:

$$\mu = A + 10n\lg d \quad (2)$$

式中: $d$  为无线信号接收点与发射点之间的距离; $n$  为受环境影响的路径损耗因子; $A$  为单位距离(1 m)处的信号强度 RSSI 值; $\mu$  为  $d$  米处接收信号的 RSSI 值。模型(2)反映了自由空间传播距离  $d$  以及信号波长等无线硬件特性对接收信号均值的影响,但在复杂传播环境中,无线信号均值  $\mu$  以及方差  $\sigma$  都会受环境影响。

## 2 多映射概率特征样本增强算法描述

在神经网络的训练中,样本数量以及样本对数据特征的代表性,直接影响网络的性能。通常研究者们会采用插值算法<sup>[13]</sup>对指纹数据进行扩充,但该方法无法表征环境影响与数据特征。本文提出了对有限实测数据的增强算法,其增强数据可以有效体现受实际环境的影响数据的特征,为神经网络训练提供大量有效样本。首先在自由空间拟合无线信号阴影传播模型(式(2)),并生成多点特征。然后,利用无线信号的空间相关性,对多点特征进行环境修正。最后,构建多点多映射概率特征样本,反应信号的波动特征。本文提出的多点多映射样本增强算法流程见下。

第一步:自由空间阴影传播模型拟合。

固定一个锚节点,获取不同距离  $d$  的信号强度 RSSI,通过 Kalman 滤波<sup>[14]</sup>处理得到均值  $\mu$ ,再利用最小二乘法拟合无线信号阴影传播模型(2)中的参数  $n$ 。

第二步:生成实测样本特征集  $z_1$ 。

根据实际定位环境的干扰情况,选取  $m$  个能够代表整个定位空间环境因素的实测采样点,读取每个实测采样点的 RSSI 样本数据,通过 Kalman 滤波可得实测样本数据的均值  $\mu_1$  以及方差  $\sigma_1$ ,由  $m$  个实测样本点均值构成的实测样本特征集  $z_1$ 。以图 1 三个锚节点  $A, B, C$  构成的三角形定位空间为例,选图中  $\odot$  标记的  $m$  个实测点( $g_1, g_2, \dots, g_m$ )为实测样本集。

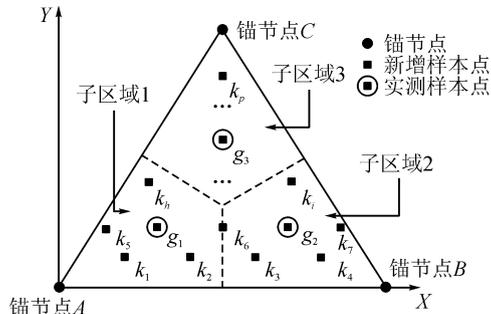


图1 三锚节点定位空间的样本点与子区域  
Fig. 1 Sample points and sub-regions of the three-anchor node positioning space

第三步:生成特征的环境影响量  $\sigma$  以及  $\Delta\mu$ 。

考虑到环境对 RSSI 信号的影响,本文定义环境对 RSSI 信号的环境影响量为自由空间的阴影传播模型(2)计算的采样点 RSSI 特征与实测 RSSI 特征的差。因此,根据自由空间的阴影传播模型(2)计算第二步中  $m$  个实测采样点的计算均值  $\mu_0$ , 以及结合实测样本的均值  $\mu_1$  可以得到  $m$  个采样点特征的环境影响量:  $\Delta\mu = \mu_1 - \mu_0$ ;  $\sigma_1 = \sigma_0$ 。

第四步:构建新增样本点特征值并修正。

根据无线信号传播的空间相关性,可以认为在一定半径范围区域内,环境对无线信号的影响相同<sup>[15]</sup>。因此,以实测样本点为聚类中心,将定位空间划分成若干个子区域,每个子区域包含一个实测点,在同一区域内新增样本点与实测点受环境影响是相同的。

在图 1 中,  $m$  个实测点将 ABC 三角区域分为  $m$  个子区域,图 1 中用虚线分割。在三角区域 ABC 中随机新增  $p$  个样本点 ( $k_1, k_2, \dots, k_p$ ) (图中 ■ 标记位置)。首先利用每个新增样本点的位置信息,通过无线信号阴影传播模型(2)可以得到新增样本点 RSSI 计算均值  $\mu_{p0}$ , 然后,根据每个新增样本点所在子区域,对新增样本的特征进行环境修正,修正均值为  $\mu_p = \mu_{p0} + \Delta\mu$ , 修正方差为  $\sigma_p = \sigma_1$ 。

第五步:生成增强样本特征集  $z_2$ 。

实测样本点特征 ( $\mu_1, \sigma_1$ ) 与新增样本点特征 ( $\mu_p, \sigma_p$ ) 共同组成具有  $(m+p)$  个点的增强样本特征集  $z_2$ 。

第六步:生成多点多映射概率分布样本集  $z_3$ 。

根据 RSSI 信号分布模型(式(1)),由增强样本点的均值与方差,按正态分布在每个样本点生成  $N$  个多映射样本。即由  $m$  个实测样本,增强为  $(p+m)N$  个 Gauss 概率分布的多点多映射样本集  $z_3$ , 该增强样本集更好地反应了环境对无线信号的衰减以及波动影响。

利用上述方法我们可以根据定位区域环境干扰程度不同,自适应构建不同定位区域的多点多映射概率增强样本集。

### 3 实验结果与分析

#### 3.1 实验环境与过程

图 2 是某地下车库的实验场景,该场地为半开放式复杂场景。图 3 为该地下车库的俯视图,车库长约 38.8 m,宽约 27.5 m,高约 5 m。车库为双层金属结构车库,车库中有通道 1、2、3,图 3 中灰色 ■ 部分为混凝土障碍立柱,黄色部分是停车区,在通道 3 两侧以及车库四周有混凝土墙,车库顶部以及停车区内有大量金属材质障碍物。该实验环境复杂,实验结论具有说服力。



图 2 地下车库实验实景图

Fig. 2 Experimenton the underground garage

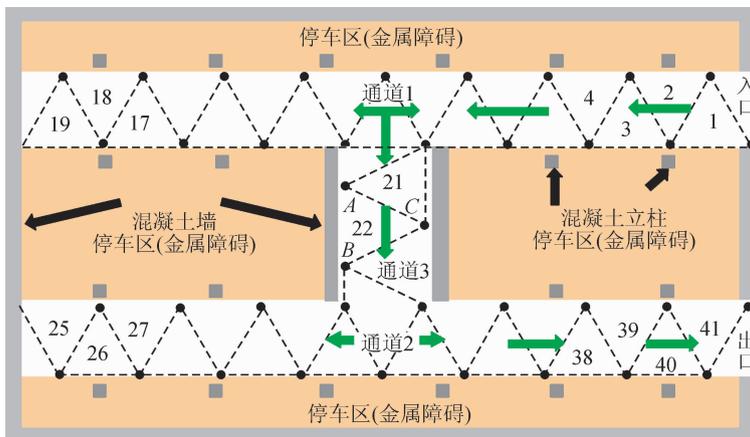


图 3 实验场地锚节点布局

Fig. 3 Layout of anchor nodes in the experimental site

本实验采用蓝牙技术<sup>[16]</sup>对该地下车库中人员进行定位。实验采用蓝牙信号采集模块(锚节点)为深圳瑞迪莱科技公司的 NRF52832;蓝牙信号发射模块(信标)为北京四月兄弟科技的 April Beacon N04。根据文献<sup>[17]</sup>可知,当 3 个锚节点以最佳测量半径为边长布局为等边三角形时,待定位点在三角形区域内的定位精度最高,故本实验在待定位区域内锚节点布局见图 3(图中●为锚节点)。每 3 个锚节点构成一个待定位区域,相关运算在图中 22 号定位区域内进行。

首先,在自由空间对所使用的蓝牙锚节点以及信标进行不同距离处 RSSI 的测量,拟合得到蓝牙模块的阴影传播模型:

$$\mu = -35.7 + 8.8 \lg d \quad (3)$$

其次,在定位区域中选取 7 个实测样本点,如图 4 中⊙所示。从三个锚节点分别读取每个实测点信标 RSSI 样本数据(连续的 100 个 RSSI 值),生成 7 个实测点的特征样本集  $\approx 1$ 。

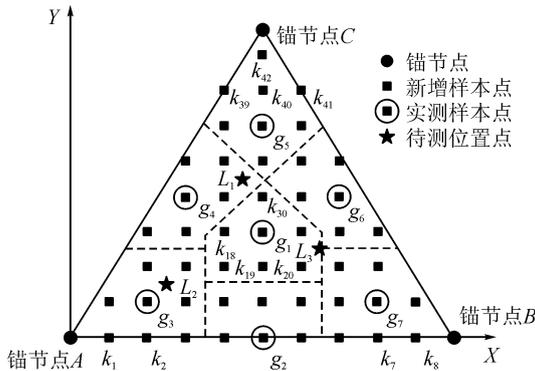


图 4 22 号定位区域样本点布局图

Fig. 4 Layout of sample points in No. 22 positioning area

图 5 中虚线为三个锚节点读取  $g_1$  点信标的 RSSI 实测样本数据,实线为滤波后的数据,可见 RSSI 滤波后的数据仍然呈现随机波动,利用滤波后数据可得均值  $\mu_1$ 。图 6 为滤波后 RSSI 的概率分布直方图,可见接收的 RSSI 信号基本呈正态分布,计算可得方差  $\sigma_1$ 。

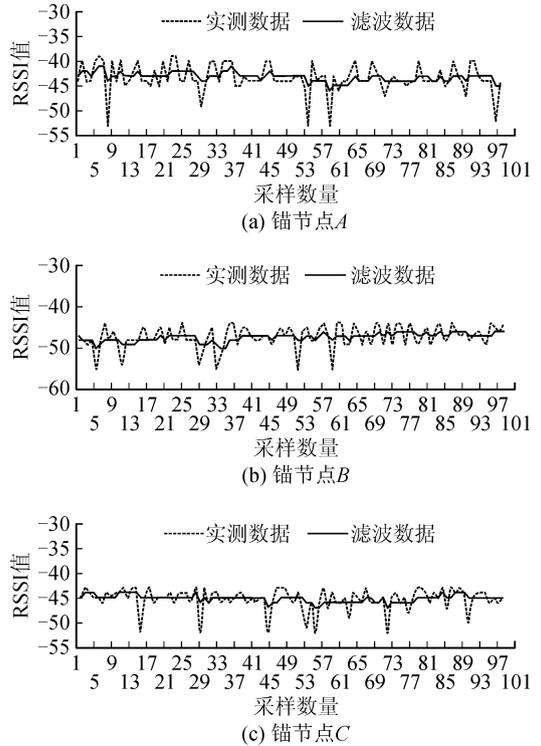


图 5 三个锚节点读取实测点  $g_1$  的 RSSI 值  
Fig. 5 Three anchor nodes read the RSSI value of the measured point  $g_1$

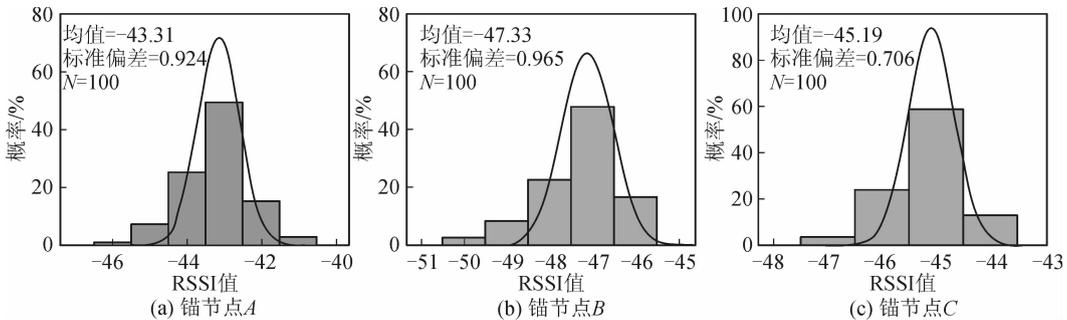


图 6 实测点  $g_1$  滤波后 RSSI 概率分布直方图

Fig. 6 Histogram of RSSI probability distribution after filtering of measured point  $g_1$

然后,将定位空间分为 7 个子区域,并新增 42 个样本点■,见图 4。利用模型(3)以及无线信号传播的空间相关性,可以得到增强样本的特征集  $\approx 2$ 。之后利用增强样本点的均值与方差,通过 MATLAB 函数 normrnd,在每个样本点各生成 100 个正态分布的样

本,得到 4 900 个多点多映射概率样本集  $\approx 3$ 。

### 3.2 实验结果分析

1) 增强样本与实测样本的 RSSI 误差分析  
为了对本文样本增强算法的合理性做评估,通过人工对上述实验中增强的 49 个采样点进行实测,

每个采样点测量 100 组 RSSI 值,对实测原始数据滤波后得到 49 个样本点的实测 RSSI 的均值  $\mu$  与方差  $\sigma$ 。表 1 给出了 49 个采样点对锚节点 A, B, C 的实测 RSSI 特征,与本文增强算法得到的样本特征的对比。由表 1 可见,本文增强方法得到的 RSSI

与实测 RSSI 值的均值  $\mu$  误差为 3.9%,方差  $\sigma$  的误差为 36.4%。其中,方差特征的误差比较大,是因为实验环境对实测数据的多径效应影响严重,导致测试数据波动大,增强数据更加平稳,消除了环境影响,有利于提高定位精度。

表 1 49 个实测样本点均值、方差对比

Tab. 1 Comparison of mean and variance of 49 measured sample points

实测样本点	多点多映射概率样本均值	测试样本均值	均值的误差	多点多映射概率样本方差	测试样本方差	方差的误差
$k_1$	(-34, -49, -55)	(-29, -49, -52)	(-5, 0, -3)	(0.2, 0.2, 0.3)	(0.1, 0.2, 0.4)	(0.1, 0, -0.1)
$k_2$	(-38, -50, -52)	(-36, -48, -50)	(-2, -2, -2)	(0.4, 0.3, 0.2)	(1.0, 0.8, 0.6)	(-0.6, -0.5, -0.4)
...	...	...	...	...	...	...
$k_{48}$	(-52, -47, -34)	(-50, -49, -38)	(-2, 2, 4)	(0.2, 0.2, 0.1)	(0.2, 0.5, 0.4)	(0, -0.3, -0.3)
$k_{49}$	(-56, -47, -26)	(-52, -49, -29)	(-4, 2, 3)	(0.3, 0.3, 0.2)	(0.4, 0.8, 0.2)	(-0.1, -0.5, 0)

## 2) 定位精度分析

利用本次实验的 7 个实测点特征样本集  $z_1$ , 49 个增强样本特征集  $z_2$ 、多点多映射样本集  $z_3$ , 以及对 49 个全人工采集样本集  $z_4$ , 分别对同参数的 BP 神经网络进行训练, 得到 4 个不同的定位网络 net1, net2, net3 以及 net4。在图 4 中任意选取三个待测位置  $L_1, L_2, L_3$  (图中★标识), 利用图 4 中网络分别对每个位置进行 5 次定位并求均值, 得到结果见表 2。表中定位误差  $\Delta d$  为待测点理论位置与计算位置之间的欧氏距离。

与 49 点全人工采样数据训练的网络 net4 相比, 定位误差仅增加了 5.5%。但是从样本集的构建成本分析, 多点多映射概率样本增强方法生成的 49 点样本集仅需 7 个实测点即可, 与人工采集 49 点样本集相比, 人工采集工作量降低了 85.7%。综合分析, 本文提出的样本增强方法在基于 RSSI 的神经网络定位算法中应用, 人力成本与定位精度综合性能最优。

## 4 结 论

本文根据自由空间无线信号的传播特性, 由少量实测点数据生成定位空间多点特征, 并利用空间相关性进行特征的环境修正。最后, 利用信号波动的概率特征构建了多映射的正态分布数据样本。实验结果表明, 本文提出的多点多映射概率分布样本生成方法, 人工成本低, 生成的样本能更好地表征数据特征, 提高了神经网络的定位精度。但考虑环境对无线信号影响的不确定, 如何建立复杂环境的传播模型将是主要的研究方向。

### 参考文献:

- [1] 哈吉德玛. 基于位置服务 (LBS) 的应用研究[J]. 现代信息技术, 2019, 3(4): 61-62.  
Khajidmaa. Application research of location based service LBS[J]. Modern Information Technology, 2019, 3(4): 61-62.
- [2] 陶佳峰, 杨晓洪, 王剑平, 等. 基于 RSSI 的室内测距模型优化技术[J]. 传感器与微系统, 2017, 36(1): 24-26, 30.  
TAO Jiafeng, YANG Xiaohong, WANG Jianping, et al. Model of indoor distance measurement optimization techniques based on RSSI [J]. Transducer and Micr-

表 2 三种训练样本得到的神经网络定位误差  $\Delta d$ Tab. 2 Neural network positioning error  $\Delta d$  obtained by the three training samples

定位点	net1	net2	net3	net4
$L_1$	125.5	78.1	47.9	55.3
$L_2$	148.7	122.6	36.4	27.7
$L_3$	112.1	52.8	31.1	26.4
平均值	128.8	84.5	38.5	36.5

由表 2 可见, 由于 7 个实测点特征样本的数据量过少, 导致网络 net1 的泛化能力弱, 平均定位误差为 128 cm, 无法在实际工程中应用。网络 net2 采用了增强的 49 个特征样本, 样本数量增加, 且通过修正反应了环境在整个定位空间的影响, 网络定位精度提高 34.4%, 但由于单采集点只有一组均值, 数据量过少, 无法体现 RSSI 的概率分布情况, 故平均定位精度提高有限。网络 net3 采用了多点多映射正态分布样本, 样本数量多, 且更真实反应数据波动, 由此训练的网络 net3 模型定位精度较高, 平均定位误差 38.5cm, 相对网络 net2 精度提高了 50%,

- osystem Technologies, 2017, 36(1):24-26, 30.
- [3] GE Xingbin, QU Zhiyi. Optimization WIFI indoor positioning KNN algorithm location-based fingerprint[C]// 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), 2017: 135-137.
- [4] 吴超, 张磊, 张琨. 基于 BP 神经网络的 RFID 室内定位算法研究[J]. 计算机仿真, 2015, 32(7):323-326.  
WU Chao, ZHANG Lei, ZHANG Kun. Research on RFID indoor localization algorithm based on BP neural network [J]. Computer Simulation, 2015, 32 (7): 323-326.
- [5] 冯雪元. 无线信号传播衰弱浅析[J]. 科技视界, 2014 (36):90-91, 197.
- [6] SESHADRI V, ZÁRUBA G V, HUBER M. A Bayesian sampling approach to in-door localization of wireless devices using received signal strength indication[M]// Third IEEE International Conference on Pervasive Computing and Communications, 2005, 1:75-84.
- [7] 陈龙鹏, 叶宁, 王汝传. 基于双神经网络的 RFID 室内定位方法[J]. 计算机系统应用, 2019, 28(11):218-223.  
CHEN Longpeng, YE Ning, WANG Ruchuan. Indoor position method for RFID system based on dual neural network[J]. Computer Systems & Applications, 2019, 28(11):218-223.
- [8] 葛柳飞, 李克清, 戴欢. 基于自适应 GRNN 的无线室内定位算法[J]. 计算机工程, 2016, 42(6):81-85, 90.  
GE Liufei, LI Keqing, DAI Huan. Wireless indoor location algorithm based on adaptive GRNN[J]. Computer Engineering, 2016, 42(6):81-85, 90.
- [9] 李竞, 卢建洲. 一种基于图优化的室内定位指纹数据库建立方法[J]. 电子测量与仪器学报, 2020, 34(2): 29-35.  
LI Jing, LU Jianzhou. Method for radio fingerprint databased establishment using graph optimization [J]. Journal of Electronic Measurement and Instrument, 2020, 34(2):29-35.
- [10] 成意, 马永涛. 基于群智指纹的指纹库建立及优化算法[J]. 南开大学学报(自然科学版), 2020, 53(1):65-70.  
CHENG Yi, MA Yongtao. An optimized radio map construction algorithm based on crowdsourcing fingerprinting [J]. Acta Scientiarum Naturalium Universitatis Nankaiensis (Natural Science Edition), 2020, 53 (1):65-70.
- [11] KAEMARUNGSI K, KRISHNAMURTHY P. Analysis of WLAN's received signal strength indication for indoor location fingerprinting[J]. Pervasive and Mobile Computing, 2012, 8(2):292-316.
- [12] LUO Jiayou, ZHAN Xingqun. Characterization of smart phone received signal strength indication for WLAN indoor positioning accuracy improvement [J]. Journal of Networks, 2014, 9(3):739-746.
- [13] JAN S S, YE H S J, LIU Y W. Received signal strength database interpolation by Kriging for a Wi-Fi indoor positioning system [J]. Sensors, 2015, 15 (9): 21377-21393.
- [14] 吴蔚, 谭献海, 钱晓群, 等. 基于卡尔曼滤波的室内定位可信指纹库研究与实现[J]. 计算机应用研究, 2020, 37 (6):1835-1838.  
WU Wei, TAN Xianhai, QIAN Xiaoqun, et al. Research and implementation of trusted fingerprint library in indoor localization based on Kalman filter[J]. Application Research of Computers, 2020, 37(6):1835-1838.
- [15] 邹士娇. 室内短距离无线信道测试及其传播特性的研究[D]. 南京:南京信息工程大学, 2018.  
ZOU Shijiao. Research on indoor short-distance wireless channel testing and its propagation characteristics [D]. Nanjing: Nanjing University of Information Science and Technology, 2018.
- [16] PATERNA V C, AUGÉ A C, ASPAS J P, et al. A bluetooth low energy indoor positioning system with channel diversity, weighted trilateration and Kalman filtering[J]. Sensors, 2017, 17(12): 2927.
- [17] 刘旭明, 王伟, 景强, 等. 基于粒子群优化的 RSSI 等弧三边定位算法[J]. 传感器与微系统, 2018, 37(8): 134-136.  
LIU Xuwei, WANG Wei, JING Qiang, et al. RSSI-based equal-arc trilateral localization algorithm based on particle swarm optimization [J]. Transducer and Microsystem Technologies, 2018, 37(8):134-136.

(责任编辑 王绪迪)