

北京交通大学

硕士专业学位论文

基于机器学习的室内射频指纹定位技术研究

Indoor RF Fingerprint Positioning Based on Machine Learning

作者：胡安民

导师：张立军

北京交通大学


2019年6月

## 学位论文版权使用授权书

本学位论文作者完全了解北京交通大学有关保留、使用学位论文的规定。特授权北京交通大学可以将学位论文的全部或部分内容编入有关数据库进行检索，提供阅览服务，并采用影印、缩印或扫描等复制手段保存、汇编以供查阅和借阅。同意学校向国家有关部门或机构送交论文的复印件和磁盘。学校可以为存在馆际合作关系的兄弟高校用户提供文献传递服务和交换服务。

(保密的学位论文在解密后适用本授权说明)

学位论文作者签名：胡安民

导师签名：

签字日期：2019年6月3日

签字日期：2019年6月3日

学校代码：10004

密级：公开

# 北京交通大学

## 硕士专业学位论文

基于机器学习的室内射频指纹定位技术研究

Indoor RF Fingerprint Positioning Based on Machine Learning

作者姓名：胡安民

学 号：17125027

导师姓名：张立军

职 称：副教授

工程硕士专业领域：电子与通信工程

学位级别：硕士

北京交通大学

2019年6月

## 致谢

首先，谨向我的导师张立军教授表示最真诚的谢意，在读研期间，张老师严谨的治学态度，孜孜不倦的科研精神和求真务实，积极进取的人生品格都深深的影响了我。在科研方面，无论是论文选题还是定稿修改，张老师都倾注大量心血给予了我耐心的指导和建议。在未来发展方面，张老师不仅给予我足够的指点和支 持，更给予我自由发展的空间，这些帮助为我的工作 和未来发展起到了决定作用。在生活方面，张老师更是给予了我无微不至的关怀，在艰难的时候安慰、鼓励我，使我能走出人生困境。

感谢实验室的师兄师姐和其他同学：程蕾、郑一帆、周伟、张大富、刘一建、高志朋、曾显询、陈泽龙、薛会、赵红娜、苏迪等，感谢你们在整个研究生期间对我学习工作给予的帮助和支持。感谢你们的陪伴，愿我们的友谊天长地久。

感谢我的父亲，给我健康的身体，完整的人格，教会我坚强的性格和遇到困难不屈服的个性。

感谢我的母亲，姐姐，姐夫，女朋友感谢你们在最艰难的时候的陪伴，因为有你们在我才能一直坚持下去，顺利的完成毕业论文。

最后，感谢为评阅本论文付出辛苦的各位专家和学者们。

## 摘要

近年来,无线局域网的广泛部署和移动智能设备的迅速普及为无线室内定位提供了便利条件。在众多纷纭的定位技术中,射频指纹技术由于具有精度高、成本低的优点,吸引了越来越多的研究人员关注。但与此同时,室内电磁环境复杂,加之随机因素过多,无疑会给射频指纹定位的精度带来影响,因此针对此问题,本文将机器学习引入射频指纹定位中,尝试提出新的算法以提高定位精度。

在传统的射频指纹定位系统中,由于射频信号受到多种因素的影响,尤其是在真实的大规模复杂室内环境中,指纹信号和地理位置之间对应的的数学模型难以构建。针对该问题,我们提出了一种基于七层神经网络的射频指纹方法。该方法先通过将原始的指纹信号输入到四层的深度神经网络中挖掘出射频指纹中的潜在特征,然后再将该潜在特征输入到一个三层的神经网络中进行定位。由于挖掘的特征对于不同的物理空间具有更好的适应性,该射频指纹定位系统在 UJIINDOORLOC 公开数据集上可以实现 99%的定位准确率。

其次,实验结果表明,当选取的测试集采集时间比训练集晚四个月时,定位准确度下降到 72%,这说明,在基于机器学习的射频指纹定位中,随着时间的推移,定位精度快速下降。通过分析以 RSSI 为代表的射频指纹随时间变化的统计特征,发现在不同时间段内,射频信号具有不同的数据分布,导致了定位系统的时效性较差。针对该问题,我们提出了两种基于迁移机器学习的射频指纹算法。一种是基于特征的迁移机器学习,该方法是通过挖掘离线指纹数据和在线指纹数据的潜在指纹特征来实现精确定位,因而具有更好的时间适应性。另一种则基于样本的迁移机器学习,通过引入小部分新的指纹数据来改变原始指纹数据分布,从而提高定位精度。在 UJIINDOORLOC 公开数据集上的测试结果表明,当选取的测试集采集时间比训练集晚四个月时,基于特征迁移机器学习的射频指纹定位系统,可以实现 82%的定位准确率;基于样本迁移机器学习的射频指纹定位系统,可以实现 90%的定位准确率,均比直接训练的模型的准确率有了显著的提高。

**关键词:** 射频指纹定位; 机器学习; 迁移学习; 数据分布

## ABSTRACT

In recent years, with the widespread deployment of wireless local area networks and the rapid spread of mobile smart devices, the prior art provides convenient conditions for wireless indoor positioning. Among the many positioning technologies, RF fingerprint technology has the characteristics of high precision and low cost, which attracts more and more researchers. However, due to the complex electromagnetic environment in the room and the excessive random factors, it will undoubtedly affect the accuracy of RF fingerprint location. Therefore, this paper introduces machine learning into RF fingerprint location and attempts to propose a new algorithm to improve the positioning accuracy.

In the traditional RF fingerprint positioning system, since the RF signal is affected by many factors, especially in a real large-scale complex indoor environment, the mathematical model corresponding to the fingerprint signal and the geographical location is difficult to construct. To solve this problem, we propose a radio frequency fingerprinting method based on machine learning. The method firstly extracts the potential features in the radio frequency fingerprint by inputting the original fingerprint signal into the four-layer deep neural network, and then inputs the feature into a three-layer neural network for positioning. Since the mining features have better adaptability to different physical spaces, when the test set and the training set are the fingerprint data collected during the same period, and the sample quantities are 8000 and 3000 respectively, the machine learning-based RF fingerprint positioning system is disclosed in UJIINDOORLOC. 99% positioning accuracy on the data set.

Secondly, the experimental results show that when the selected test set acquisition time is four months later than the training set, the positioning accuracy drops to 72%, which indicates that in the machine-based RF fingerprinting, the positioning is over time. The accuracy drops rapidly. By analyzing the statistical characteristics of the RF fingerprint represented by RSSI over time, it is found that the RF signals have different data distributions in different time periods, resulting in poor timeliness of the positioning system. In response to this problem, we propose two RF fingerprinting systems based on migration machine learning. One is feature-based migration machine learning, which achieves accurate positioning by mining potential fingerprint features of different time periods, and thus has better time adaptability. The other is based on sample-based migration machine learning, which introduces a small amount of new fingerprint data to change the original fingerprint data distribution, so that the model can adapt to RF signals

with different data distribution. The test results on the UJIINDOORLOC public dataset show that when the selected test set acquisition time is four months later than the training set, and the sample size is 10000 and 1000 respectively, the RF fingerprinting system based on feature migration machine learning can be realized. 82% of the positioning accuracy; based on the sample migration machine learning RF fingerprint positioning system, can achieve 90% of the positioning accuracy, which is significantly higher than the accuracy of the directly trained model.

**KEYWORDS :** RF fingerprint positioning; weighted KNN algorithm; coordinate clustering; point clustering

## 目录

摘要	iii
ABSTRACT	iv
1 引言	1
1.1 研究背景及意义	1
1.2 国内外研究现状	2
1.2.1 室内射频定位技术	2
1.2.2 基于机器学习的射频指纹定位技术	5
1.3 研究内容	8
2 浅析 WLAN 的射频指纹定位	10
2.1 射频指纹定位原理和流程	10
2.1.1 定位原理	10
2.1.2 定位流程	11
2.2 射频指纹点的选取和分析	13
2.2.1 射频指纹的选取	13
2.2.2 射频指纹的分析	13
2.3 本章小结	16
3 基于机器学习的射频指纹定位技术	18
3.1 数据集介绍	18
3.1.1 实验数据集	18
3.1.2 公开数据集	19
3.2 机器学习相关理论分析	20
3.2.1 近邻法 (Nearest Neighbor)	20
3.2.2 随机森林(Random Forest)	21
3.2.2 梯度提升算法(Gradient Boosting Decision Tree , GBDT)	22
3.3 基于机器学习的射频指纹定位	22
3.3.1 基于 KNN 的指纹定位	23
3.3.2 基于随机森林的指纹定位	24
3.3.3 基于梯度提升算法的指纹定位	25
3.3.4 三种算法总结	26
3.4 基于深度学习的射频指纹定位	27
3.4.1 深度学习理论介绍	28
3.4.2 基于四层神经网络的指纹定位系统	28



3.4.3 实验分析 .....	30
3.5 本章小结 .....	31
4 基于迁移学习的射频指纹定位 .....	32
4.1 迁移学习的理论介绍 .....	32
4.1.1 特征迁移学习 .....	32
4.1.2 样本迁移学习 .....	33
4.2 基于特征迁移的射频指纹定位 .....	34
4.2.1 算法介绍 .....	34
4.2.2 算法可行性分析 .....	35
4.2.3 实验分析 .....	36
4.3 基于样本迁移的射频指纹定位 .....	37
4.3.1 算法介绍 .....	37
4.3.2 算法可行性分析 .....	38
4.3.3 实验分析 .....	39
4.4 本章小结 .....	39
5 结论 .....	41
参考文献 .....	43
作者简历及攻读硕士/博士学位期间取得的研究成果 .....	47
独创性声明 .....	48
学位论文数据集 .....	49

# 1 引言

## 1.1 研究背景及意义

近年来,随着移动互联网的快速发展和智能手机终端的迅速普及,基于位置的服务(location based service, LBS)在社会生产生活中给人们带来了很大的便利。在室外环境下,人们通过定位可以方便的寻找餐厅、银行、外卖、出行,甚至交友等等。在室内环境中,人们通过定位可以方便进行室内导航,商业推广,物品追踪、机器人调度以及人员救援等<sup>[1]</sup>。目前,室外定位服务已经大规模商用,但室内定位服务仍处在研究阶段。由于GPS导航信号穿过墙体后迅速衰减,导致无法在室内使用卫星导航<sup>[2,3]</sup>。因此各种室内定位的方法与技术被相继提出,其中射频指纹定位技术因能够利用现有的广泛部署的无线局域网和移动终端进行定位,而具备定位成本低、定位精度高、环境适应能力强等优越性能,受到众多研究人员的关注,成为室内定位的主流技术。

射频指纹定位首先需要利用在待定位区域采集的射频信号的强度值来构建出与真实物理环境相对应的信号无线地图,然后将在待定位点测得的信号强度值输入到建立的无线地图中进行匹配来获得该点的实际位置<sup>[4]</sup>。然而,由于室内环境复杂、人员随意流动、无线信号在传输过程的衰减和多径效应导致信号无线地图建立费时费力、定位精度不高、定位系统不稳定等问题<sup>[5]</sup>。因此如何提高射频指纹定位的准确性和稳定性,具有十分重要的意义。

近年来,随着人工智能不断的在各个领域取得重大突破,将机器学习和室内射频指纹定位结合成为了一个新的研究热点。如何利用机器学习能够在大规模数据中挖掘出潜在特征建立模型的优势来解决室内射频指纹定位遇到的难题,成为了一个重要的研究内容。机器学习是一门横跨多个领域的交叉学科,通过在大量的数据中让计算机能够自发的学习到数据之间的逻辑关系,挖掘出影响模型结果的潜在因素,建立起客观的数学模型,进而解决实际问题。机器学习经过几十年的发展,已出现多种并行研究方向,例如逻辑回归、支持向量机、聚类、神经网络、以及学习理论等,且均已取得了众多有价值的成果<sup>[6]</sup>。

基于机器学习的室内指纹定位是可行的,一方面,近年来,机器学习技术在诸多领域得到了十分成功的应用与发展,如语音,文本,视频等<sup>[7]</sup>。机器学习可以在任何能够积累经验数据的地方发挥强大的作用,目前,指纹定位已与机器学习结合催生出了许多表现优秀的定位算法。另一方面,传统射频指纹建立的单一数学模

型难以适应室内定位环境的复杂多变，在大规模的多物理空间的定位环境中由于物理空间的差异性导致定位精度变差，而机器学习能够从大量的真实数据中挖掘出不同的室内环境中共有的潜在特征，使得建立的模型对于不同的室内环境有更好的适应性，因此基于机器学习和射频指纹定位的结合具有十分重要的意义。

目前，研究者在探索将机器学习和室内射频定位进行结合的时候，大多是在研究射频指纹数据特征的处理，没有考虑射频指纹信号分布跟随时间的推移发生变化。因此在前人的研究基础上，本文通过搭建实验平台，采集实验数据，分析指纹信号的统计特征，首先研究了基于机器学习的室内射频指纹定位在实际环境的应用，提出了一种基于七层自适应神经网络的室内定位系统。同时在实验中发现了射频信号分布的变化导致定位精度变差的问题，然后提出了基于迁移学习的解决方案，并设计了两种基于迁移机器学习的指纹定位算法，最后对本文提出的系统和算法进行了性能评估。

## 1.2 国内外研究现状

本节分为两部分，第一部分是对室内射频指纹定位技术的简要介绍。在该部分中通过介绍室内射频定位的技术分类和几种经典的室内射频指纹定位系统的基本原理和定位性能，来概述室内定位技术的起源和发展。第二部分是对基于机器学习的射频指纹定位技术的简要介绍。在该部分中，将从离线阶段的指纹库建立和在线阶段的位置计算两个方面，分别介绍机器学习在射频指纹定位过程中的指纹预处理和定位算法上的应用，以此概述机器学习在射频指纹定位上的发展过程。

### 1.2.1 室内射频定位技术

在 20 世纪 90 年代，在室内定位提出之后，发展了众多的定位技术，包括红外线(infrared)定位技术、蓝牙(blue tooth)定位技术、超宽带(ultra-wideband, UWB)的定位技术、ZigBee 定位技术以及射频定位(radio frequency, RF)技术<sup>[8-11]</sup>。其中射频技术因具有能利用现有的局域网和移动终端而具备成本低、精度高的优越性能而成为室内定位的主流技术，因此本文主要探讨射频定位技术。Wi-Fi 是一种支持 IEEE 802.11 标准的无线局域网技术<sup>[12-14]</sup>，Wi-Fi 信号因其在室内场景的普及性而成为射频定位技术首选的射频信号。基于 Wi-Fi 的室内定位被提出后，受到了研究者的广泛研究，发展了众多的定位方法<sup>[15,16]</sup>，这些方法主要有两种方式，一种是基于几何定位法，一种是基于射频指纹法。其中基于射频指纹定位技术受到了众多学者的研究，并开发了许多基于性能优越的射频指纹定位系统。在该小节中，本文将通过介绍射频定位的两种方式和几种经典的室内射频指纹定位系统的发展过程来

阐述射频指纹定位技术。

### (1) 射频定位的两种方式

#### 1) 几何定位法

基于几何定位法是先测量出待测点与周围的三个或三个以上的已知接入点 (access point, AP) 的相对位置, 然后运用三角几何原理推测出待测点位置的一种方法<sup>[17]</sup>。其中相对位置的测量有四种计算方式, 分别是信号到达角测量法 (angle of arrival, AOA), 信号到达时间测量法 (time of arrival, TOA), 信号到达时间差测量法 (time difference of arrival, TDOA) 和接收信号强度测量法 (received signal strength, RSS)<sup>[18]</sup>。

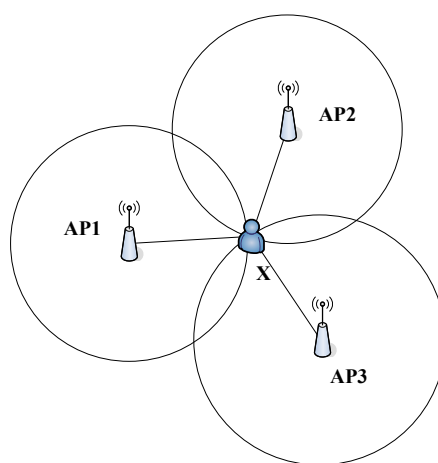


图 1-1 使用几何法计算出移动设备的位置

Fig.1-1 Use geometry to calculate the location of your mobile device

几何定位法因其计算简单而受到很多研究者的青睐, 但是由于其定位需要三个以上的 AP, 且定位过程需要保证时间严格同步和角度测量精准, 因此对于定位设备具有较高的时间同步性要求, 定位成本较高, 且定位精度容易因为各个 AP 之间的时间同步的偏差而受到较大的影响。

#### 2) 射频指纹定位法

射频指纹定位的基本原理是通过收集待测区域各点的射频指纹信号, 建立起射频指纹和坐标位置之间的无线地图, 然后将在待测点收集得到的射频指纹通过无线地图采用定位算法进行匹配获得其坐标位置。基于指纹射频指纹法, 根据采用的定位算法不同, 定位分为两种类型, 一种是在定位阶段采用确定性算法实现定位, 另一种是在定位阶段采用概率性算法实现定位。

a. 确定性算法是, 是在 2000 年 Microsoft 公司的 Bahl 发布 Randar 定位系统时提出的。是通过指纹向量之间的欧氏距离 (Euclidean distance) 来比较待测点和指纹库中已知点之前的距离, 其定位思想是指纹空间中欧式距离相近的两个点, 其在

实际的地理位置中有是相近的。常见的确定性定位算法近邻算法 (nearest neighbors, NN) 和 K 近邻算法 (k-nearest neighbors, KNN) 算法等。

b) 概率性算法是, 是在 2001 年由 Castor 发布 Nibble 室内射频定位系统时提出的。通过计算指纹库中的指纹统计特征, 拟合出指纹的数据分布, 测量到待测点指纹后根据拟合的数据分布再采用贝叶斯算法计算出待测点的可能位置。常见的概率性定位算法有最大后验概率和最大似然概率两种方式。

射频指纹定位技术无需额外的设备, 通过室内空间的无线局域网和智能移动设备即可完成位置的确定, 因此具有成本低的优点。同时相比其他算法, 射频指纹定位精度较高, 因此成为了射频定位中一种主流的定位技术;

## (2) 几种射频指纹定位系统

射频指纹定位技术被提出后, 国内外众多的研究学者开发了许多定位效果较好的定位系统, 本文通过介绍这些定位系统来阐述指纹定位技术在近年来的发展历程。

2000 年, 一款名为“Randar”的室内射频指纹定位系统发布, 该系统是由微软 (Microsoft) 公司的 Bahl P 带领的团队设计完成<sup>[19]</sup>。该系统定位分成两个阶段, 在离线阶段, 先对采集区域进行网格划分, 然后在将在每个网格采集的接收信号强度 (received signal strength index, RSSI,) 和该网格实际坐标位置存储成指纹库。在线阶段, 在待测点收集到 RSSI 后, 根据指纹数据库的数据, 采用最近邻算法 (nearest neighborhood, NN) 和 K 近邻算法 (k-nearest neighbors, KNN) 计算出待测点的实际坐标。此外, Randar 系统通过对比指纹匹配法和模型传播法, 证明在相同条件下, 指纹法的定位精度要高于模型传播法<sup>[20]</sup>。但由于在建库阶段和定位阶段 Randar 均采用了较为简单的处理方式, 如: 离线建库阶段, 取多次测量的 RSSI 的平均值作为该点的指纹值, 在定位阶段, 待定位点的坐标取指纹空间最相近的一个点的坐标或是 K 个点的坐标平均值, 因此定位误差较大。

2001 年, 另一款名为“Nibble”的室内射频指纹定位系统发布, 该系统是由 Castor 带领的团队设计完成<sup>[21]</sup>。和 Randar 相比, Nibble 有两个不同之处, 一个是在离线阶段, 不同与 Randar 系统采用 RSSI 作为指纹信号, Nibble 采用了射频信号接收信噪比 (signal-to-noise ratio, SNR) 作为位置指纹信号。另一个是在定位阶段, 不同与 Randar 系统采用 KNN 或 NN 的确定性定位算法, Nibble 采用了概率性算法。其在离线阶段, 在待测区域收集 SNR, 将指纹坐标点和收集的 SNR 的进行存储。在线阶段, 在待测点收集到 SNR 后, 通过指纹库中的数据根据最大后验概率算法计算出最有可能的定位点坐标。

2004 年, 一款名为“Horus”的室内射频指纹定位系统发布, 该系统是由 Youssef 带领的团队设计完成<sup>[22]</sup>。在位置指纹的选取上, Horus 同 Randar 系统一样, 采用

的是 RSSI 作为指纹特征。在离线定位算法的选取上, Horus 和 Nibble 一样, 采用的是概率定位算法。在离线阶段, 先在待测区域收集 RSSI, 然后根据信号分布拟合高斯分布图并存入指纹库中, 在线阶段, 先是在待测点收集到 RSSI 后, 然后根据指纹库中的数据, 采用贝叶斯算法获得待测待点的坐标位置。此外, 不同于 Randar 和 Nibble 系统离线阶段和在线阶段对于指纹特征的简单处理, Horus 在离线阶段, 引入了 AP 选择算法, 通过选择最优的 AP 来减少不稳定的 RSSI 值进而提高定位精度。其次, Horus 引入了聚类算法, 先是在离线建指纹库阶段对所有的指纹进行聚类, 然后在定位阶段, 首先判断是在哪个聚类区域内, 然后再进行精确定位, 提高了定位效率。

2006 年, 一款名为“Weyes”的室内射频指纹定位系统发布, 该系统是由北航师生设计完成。在位置指纹上的选取上, Weyes 选择 RSSI 的差值作为指纹信号。在离线定位的算法选取上, Weyes 选择了概率定位算法。在离线阶段, 先在待测区域收集各点的 RSSI, 然后对所有 RSSI 和其平均值进行做差, 将做差结果和指纹点坐标存入到指纹库中。在定位阶段, 先在待测点收集 RSSI 并做差得到差值后, 根据指纹库中数据, 采用概率算法来得到最有可能的定位点坐标。由于 Weyes 是将 RSSI 的差值序列作为指纹数据, 因此有效的解决了不同设备导致定位误差。其次, 在选择最有可能的 N 个点时, Weyes 系统采用的是动态的 N 值, 并将 N 个点的加权平均值作为待测点坐标位置, 提高了定位精度<sup>[23]</sup>。

### 1.2.2 基于机器学习的射频指纹定位技术

随着人工智能在各个领域不断取得突破, 将机器学习和射频指纹定位进行结合成为目前的一个研究热点。由于机器学习具有在大规模的数据中挖掘出潜在特质的能力, 因此近年来, 很多研究者开始在射频指纹定位中离线阶段和在线阶段引入机器学习来解决现有问题以此提供定位精度。

#### (1) 离线指纹库的建立

射频指纹定位中的指纹库的建立是通过在待测区采集接收信号强度值并同采集点的坐标一起存储在库中建立的。在离线建库阶段, 一个问题是指纹信号的采集耗时耗力, 传统的指纹采集工作是先在待测区域确定众多的参考点, 然后在每个参考点采集一段时间的射频信号接收强度值并存储在指纹库中。当采集的区域比较大的时候, 指纹库的建立将耗费巨大的人力物力。另一个问题是指纹信号的预处理, 由于在大范围的室内空间采集的指纹信号比较稀疏, 且又因为射频信号在室内传播受到衰减、多径效应的影响具有很大的不稳定性, 因此直接采用收集的指纹进行定位会造成定位精度的下降。

研究者为解决建库成本高的问题, 提出了众多的方案。早期的时候, 有研究者

提出了差值拟合的办法来减少采样点的个数,减少建库成本,在指纹采集阶段只需要对采集的少量的指纹进行差值拟合即可得到足够的指纹数量<sup>[24]</sup>。比较常用的差值办法有:加权距离反转法<sup>[25]</sup>。但随着机器学习在各个领域的快速突破,越来越多的研究者们考虑使用机器学习来解决这个问题。一种方式是半监督学习法,半监督学习是通过利用已经标记位置的射频指纹和一部分没有标记位置的射频指纹,来构建指纹库<sup>[26]</sup>。通过这种方式不仅可以大大的减少采样的成本而且同采集全部的指纹建库的方法相比定位精度并没有很大下降。研究者 Pan 在 2007 年提出一种新的建库方法,即半监督学习的建库方法,其在离线阶段利用一部分标记位置的 RSS 指纹和一部分没有标记位置的 RSS 指纹通过搭建的流形模型建立指纹库,在线阶段通过基于图的半监督学习方法和加权 KNN 的方式进行位置的计算,最终实现了移动设备和接入点的协同定位<sup>[27]</sup>。另一种方式是小样本学习法,小样本学习法主要是解决的是当样本数量比较少的时候如何能够学习数据中的特征和关系。目前的一种常见的解决办法是,根据少量的已经标记位置的样本中学习数据中的潜在特征,然后将其应用到大量没有标记位置的数据中。这种从已知的数据中学习知识运用到未知的数据模型中方法也可以认为是一个迁移学习的问题<sup>[28,29]</sup>。小样本学习法使得通过很少的人力物力成本就建立能够适应不同设备下的射频指纹库。

除此之外,近年来有研究者提出了基于众包的指纹库的构建方式,在这种方式下,用户只需要手持智能手机终端在测量区域内进行自由走动,手机客户端便会定时的采集射频指纹信号<sup>[30]</sup>,而收集指纹点的具体位置可以根据行人位置推算系统进行估计。用户在采集区域走动时间越长,客户端采集到的射频指纹越多。指纹收集完成后,还需利用半监督模型和数据挖掘技术对于这些无标识位置的数据进行处理,然后在存储到指纹库中。采用众包方式的指纹库建立方法不仅能够有效的减少建库成本,而且由于众包程序是长期运行的,因此能够及时更新指纹库和射频指纹定位系统。

## (2) 射频指纹数据的预处理

由于射频信号在室内传播过程中会受到多径效应、人员走动干扰、信号的反射和折射等因素,采集的射频指纹会受到很高的噪声干扰,又由于测量区域较大范围的时候,收集的射频指纹具有很高的稀疏性而导致定位的精度不高。在早期的研究中,研究者大多对射频指纹进行简单的处理,如在一个点测量多次后取平均值等。随着机器学习的迅速发展,有很多研究者开始将机器学习的理论应用到了射频指纹的预处理上。预处理主要有两种形式,一种是 AP 选择算法,主要是为了解决大范围的定位空间内指纹比较稀疏和部分 AP 不稳定导致定位精度下降的问题。另一种是指纹特征的提取,主要是利用机器学习能够挖掘出潜在特质的能力来提取出对定位模型比较关键的特质,以此提高定位模型的准确性。

在 2004 年, 在 Youssef 等人发布 Horus 室内射频指纹定位系统时就提出了 AP 选择算法<sup>[31]</sup>, 在 Horus 定位系统中, 每次在不同的定位区域中, 先根据信号的强度值进行排序, 然后根据信号的稳定性排序, 最后据此选择 AP 的 RSS 作为射频指纹。有效的提高了定位系统的稳定性和准确性。在 2006 年, chen 等人提出了一种基于信息值的 AP 选择算法<sup>[32]</sup>, 在这个算法中结合了聚类算法和决策树算法, 有效的对 AP 的判别力进行了评判, 通过该算法不仅能够有效的提高了定位系统的稳定性和准确性, 而且在相同的定位精度下, 能够减少客户端的功率消耗。

另一种射频指纹的预处理是指纹特征的提取, 不同与 AP 选择算法直接丢弃不稳定的 AP, 特征提取利用了机器学习擅长在大规模数据挖掘具有强判别力特征的优势, 通过对高维的指纹特征进行降维, 提取高判别力的特征, 有效的提高了射频指纹定位系统的准确性和稳定性。在 2006 年, Fang 等人提出了信号投影法的指纹特征提取法<sup>[33, 34]</sup>, 通过将采集的指纹信号投影到相关的物理空间上, 来减少了 AP 之间的相关性, 在文献实验部分采用了主成分分析 PCA 的方式来对比传统的 AP 选择算法, 证明了通过对指纹信号进行降维, 能够极大的缩减定位过程中的计算量并且保留了指纹特征中主要信息。近年来, 还有研究者开始通过将自编码器和核函数引入到射频指纹中来提取射频指纹的有效特征, 提高了定位系统的准确性和稳定性<sup>[35]</sup>。

### (3) 射频指纹的聚类

在线定位阶段, 是通过将在待定位点测量的指纹采用定位算法和指纹库中的指纹进行匹配计算以获得待定位点的位置坐标。在这个阶段, 一个问题是由于定位空间较大时, 定位过程的计算复杂度较高。另一问题是在线阶段的定位算法是决定射频指纹定位的关键因素, 如何选择合适的定位算法十分重要。

近年来, 很多研究者开始尝试用机器学习的理论去解决这些问题。为解决在线阶段的计算量太大的问题, 有研究者提出了聚类算法。具体做法是先指纹库的建立完成后, 将指纹库划根据聚类算法划分成几个区域, 然后在定位阶段, 获得待定位点的指纹向量后, 将待定位点确定在某个区域内, 最后在这个区域内进行位置的精度确定。这种方式极大的减少了在大规模定位环境中的定位复杂度, 有效的提高了定位效率。在 2004 年, Youssef 发布 Horous 定位系统时候, 提出了通过聚类算法来减少在线阶段的计算复杂度, 提高射频定位系统的效率<sup>[36]</sup>。在 2006 年, chen 等人又提出了一种新的聚类方式, 通过在建库中采用 K-means 后处理方式, 较大的提高了定位效率, 但这种算法容易在 k-means 计算过程中陷入局部极值。在 2012 年, 又有研究者提出了一种解决 k-means 容易陷入局部极值的方法, 实验证明这种改进后的新的聚类方式有效提高了定位效率<sup>[39]</sup>。在 2015 年, 有研究者提出了一种分块加权的聚类方法, 与之前的聚类算法相比, 这种聚类考虑了划分区域之间的不



同的权重，作者在实验中证明，这种新的聚类方法对于改善定位效率具有很好的效果<sup>[40]</sup>。

#### （4）在线定位算法

定位算法在射频指纹定位具有十分重要的作用，如何选择合适的定位算法十分关键。如上节所述，定位算法有两类，一类是根据指纹空间的欧式距离来衡量两点在实际位置的远近关系的确定性算法。另一类是根据贝叶斯算法计算的最大后验概率作为定位点选择依据的概率性算法。在早期的时候，定位阶段是采用简单的匹配算法，这种算法计算简单但定位精度不高。近年来，有研究学者将定位过程中应用了机器学习算法。在 2002 年，Brunato M 等人就开始研究统计机器学习算法在室内射频指纹定位上的应用。在文献中作者应用了支持向量机（support vector machine, SVM）作为定位算法，实验证明了该算法具有计算复杂度低，定位精度高的优点<sup>[41]</sup>。在 2006 年，Wu 等人提出了将支持向量回归引入到移动设备的定位系统中，并取得了不错的定位效果。近年来，又有研究者将神经网络等深度学习技术引入到了射频指纹中，通过搭建多层神经网络能够挖掘出大规模数据中潜在具有高判别性的特征，然后实现精确定位<sup>[42,43]</sup>。

尽管不断的有研究者提出新的算法，但射频指纹定位依然有问题没有解决。目前面临的问题有两个，一个是射频指纹信号的具有很大的不确定性，如何挖掘出不同环境中的射频指纹信号的潜在特质十分关键。另一个由于射频指纹信号的时变性，前期测量的指纹库由于环境变化导致定位失效，指纹库的建立、更新和维护需要花费较大的人力物力成本。本文将通过研究射频指纹信号的特征，为解决这两个问题提出可行的解决方案，并且基于实测数据和公共数据集进行验证分析。

### 1.3 研究内容

本文研究了基于机器学习的室内射频指纹定位技术在实际环境中的运用情况。在深入了解室内射频指纹定位技术的发展过程和基于机器学习指纹定位技术的研究现状后，针对传统的指纹定位中建立模型无法适应复杂的大规模室内环境导致定位精度不高和射频指纹信号在实际环境中的时变性导致定位精度下降的问题，从离线建库和在线定位两个方面开展研究工作。提出了一种基于四层神经网络的射频指纹定位算法和一种基于迁移学习的射频指纹定位算法，以解决传统定位模型无法适应复杂环境导致定位精度下降和射频指纹时变性问题，达到提高室内射频指纹定位的稳定性和精确性的研究目标。

本文的研究内容如下：

1. 针对传统指纹定位模型无法适应复杂的室内环境导致定位精度不高的问题，

提出了一种基于四层神经网络的定位模型。分析由于室内定位环境复杂，人员随意走动，射频信号在室内传播过程中受到了多径效应等影响，导致定位精度不高，为此通过设计了一种四层的神经网络来挖掘出适应不同室内环境的潜在指纹数据特征，然后搭建了四层神经网络定位模型进行位置确定。实验结果证明，此算法能有效的提高了定位精度。

2.针对射频指纹信号时变性导致定位精度下降的问题，提出了基于迁移学习的射频指纹定位算法。首先分析射频指纹信号的统计特征，发现随着时间的推移，射频指纹信号的数据分布也在发生变化，使得根据之前指纹数据建立的模型对于新的指纹数据不适应，从而导致了定位精度的下降。然后根据数据分布的相似性，提出了一种基于特征迁移的室内射频指纹定位算法和一种基于样本的室内射频指纹定位算法。

本文目录安排：

本文共分为五章，具体的结构安排如下：

第一章首先概述了室内射频定位的研究意义和背景，然后分别从室内定位技术分类，射频指纹定位原理和经典的室内定位系统三个方面介绍了该技术的发展过程和研究现状。

第二章首先概述射频指纹定位技术的基本流程和定位原理，然后从离线建库和在线定位两个阶段讨论了射频指纹定位技术。最后分析了射频指纹信号的时间特性和空间特性，证明了射频指纹信号在时间上具有不确定性。

第三章首先介绍了本文中用到的模拟、实测、和公开数据集，为下文中的实验部分进行铺垫。然后阐述了本文用到的机器学习的相关技术理论。最后详细介绍了本文提出了基于机器学习的射频指纹定位算法和实验过程，同时发现了射频指纹定位性能随着时间推移下降的问题。

第四章首先介绍了传统的机器学习射频指纹定位技术面临的问题，即射频指纹数据分布的变化导致了室内定位精度的下降。然后介绍了本文需要用到的迁移学习的相关理论知识。最后，详细阐述了本文提出了基于迁移学习的射频指纹定位算法，即通过学习原来指纹数据分布的潜在特征并运用到新的指纹数据上的基于特征迁移的指纹定位算法和通过引入一部分新数据改变原来指纹数据分布的基于样本迁移的指纹定位算法，并详细介绍了算法的实验过程。

第五章总结了本文的研究内容和解决的问题，并对本文的不足之处做了说明，最后对射频指纹定位技术做了展望。

## 2 浅析 WLAN 的射频指纹定位

在众多的室内定位技术中，射频指纹室内定位技术由于具有精度高、成本低的优越性能，因而成为室内定位的最佳解决方案而被广泛应用。本章将首先介绍了指纹定位技术的基本流程和定位原理，从离线建库和在线定位两个阶段讨论射频指纹定位技术，然后分析射频指纹信号的时间特性和空间特性，最后对本章内容做了总结。

### 2.1 射频指纹定位原理和流程

#### 2.1.1 定位原理

射频指纹定位是通过建立指纹信号空间和地理位置空间之间的对应关系来实现位置确定。定位过程分为两个阶段：指纹库建库的离线阶段和位置定位的在线阶段，如图 2-1 所示。离线阶段，建立存储射频指纹信号和对应位置坐标的指纹库。在线阶段，将收到的待定位点指纹信号特征和指纹库中的指纹进行匹配，获得匹配指纹后，即可认为匹配指纹的位置坐标即为待测点的位置坐标。

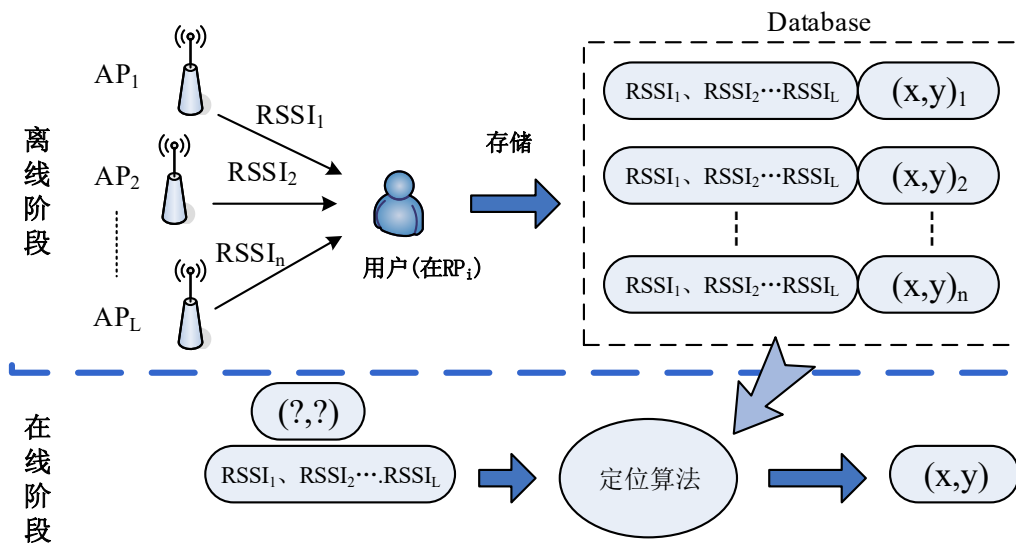


图 2-1 射频指纹定位原理

Fig.2-1 Principle of RF fingerprinting positioning

传统的射频指纹过程，可以看成是一个匹配问题，在离线建库的时候，先在待测区域选取多个参考点，然后将每个参考点的射频指纹特征和位置坐标存储在数据库中，在线定位的时候，在待测点测得射频指纹信号后，采用匹配算法和库中的指

纹进行匹配，获匹配指纹后，即可将该匹配指纹的位置作为待测点的定位位置。

在这个过程中，匹配算法设计是决定定位精度的另一个关键因素。匹配算法根据射频指纹空间中两个指纹特征之间的近似程度来衡量两个指纹对应坐标位置的远近程度。不同的匹配算法，选择了不同的指纹近似程度。近年来，研究者们采用信号指纹空间中的距离度量来作为这种近似程度，有：欧氏距离(Euclidean distance)、曼哈顿距离(Manhattan distance)等。在众多的算法中，利用欧式距离来衡量指纹信号之间的距离占据了主流地位，近年来，又有研究者提出了一种基于切比雪夫距离(Chebyshev distance)的匹配算法，实验证明较其他算法定位精度有了明显改善。

与传统的射频指纹定位不同，基于机器学习的射频指纹定位通过在大规模的指纹数据中学习得到射频指纹向量和位置坐标之间的关系函数来实现定位。定位的过程同样分为两个阶段：指纹库建库阶段和位置定位阶段。离线建库阶段，同样是首先需要将多个参考点的指纹特征和位置坐标存入到数据库中。但不同于传统的指纹定位，在这个过程中，基于机器学习的射频指纹定位会对存入到的指纹库中指纹进行预处理。预处理的过程包括了 AP 选择和特征选择。然后根据这些处理好的射频指纹信号数据采用合适的机器学习算法去学习指纹特征和位置坐标之间的非线性关系，最后在线定位阶段，在待定位点收集到射频指纹后，将射频指纹带入到学习到非线性函数中即可得出待定位点的位置坐标。

### 2.1.2 定位流程

如上节所述，基于机器学习的射频指纹定位过程，同传统的射频指纹定位一样，也分为离线建库阶段和在线定位阶段两个过程。不同之处在于，为了提高射频指纹数据质量，离线建库阶段包括了 AP 选择和特征选择两个过程，为了提高射频指纹定位的计算速度，在线定位阶段包括了区域定位和精确定位两个过程。

#### (1) AP 选择过程

在实际定位环境中，由于室内空间环境复杂，人员随意走动，射频信号在传播过程受到多径效应，反射，折射等多个因素的影响。导致 AP 发出的射频指纹具有较强的不稳定性，因此即使在同一个位置，在不同的时刻能够采集到的 AP 也可能不相同，从而导致同一个位置在离线指纹库中指纹特征和在线阶段收集到射频指纹定位也不同，因此造成了定位精度的下降。

为解决这个问题，近年来，不少研究者提出了多种不同的 AP 选择算法<sup>[31]</sup>，有 AP 强度排序法，即对每个参考点的接收的 AP 射频指纹信号进行排序，选择平均信号强度最高的几个 AP。有方差选择法，即对每个参考点接收到的 AP 射频信号一段时间的信号强度值的方差进行排序，选择方差最小的几个 AP 值。有综合选择法，即先对参考点的接收到的 AP 射频指纹信号根据平均强度值进行排序，然后再

根据各个 AP 的射频信号的方差值选择最稳定的几个 AP 值<sup>[32]</sup>。

### (2) 特征选择过程

在实际大范围的定位环境中，由于布置了多个 AP，而每个 AP 的辐射范围是有限的，因此在每个参考点处只能接收到一部分 AP 值，导致存入的指纹库中的指纹具有很强的稀疏性。这些稀疏指纹数据在定位过程中导致了定位精度的下降。

为了解决这个问题，近年来，研究者们提出了多种不同的指纹特征提出算法。有线性降维的方法，如主成分分析（principal components analysis, PCA），通过对指纹信号进行线性变换，将高维度的稀疏指纹数据转变成低维度的互不相关的向量，有效的解决了指纹库中指纹稀疏性的问题，但由于线性降维的局限性，导致定位精度的下降<sup>[34]</sup>。在文献[35]中，研究者提出了一种非线性的降维方式即自适应编码器 auto-encoder 的方法。通过构建一个多层的神经网络结构，不仅能够将高维度的稀疏指纹向量转换成一个低维度的指纹向量，而且可以有效的学习非线性特征，提高了模型的定位性能。

### (3) 区域定位过程

在较大范围的实际室内环境中，由于室内选取的参考点较多，导致了在线定位阶段计算复杂度较高，定位速度较慢。

为了解决这个问题，在线定位之前，先对指纹库中的指纹进行聚类操作，将指纹聚类成几部分，根据指纹特征和位置坐标的对应关系，聚类在一起的指纹对应了实际环境中一个特定区域。在线定位的时候，首先根据采集到射频指纹和每个聚类区域中的指纹数据根据相似度进行匹配，确定在哪个聚类区域内后，再根据这个区域内的指纹特征和坐标位置之间的非线性关系得出具体的位置。在这个过程中，衡量聚类区域指纹和待测指纹向量近似程度可以通过对区域指纹取平均值和中心值，然后比较待测指纹和各个区域内平均值或中心值之间的欧式距离。

在聚类方法中，近年来，有研究者提出了基于 k-means 的聚类方法，即根据指纹特征之间的欧式距离将多个指纹聚成几个不同的簇<sup>[36]</sup>。另一种聚类方式，是采用 DB-Scan 的聚类方式，这种聚类方式根据指纹特征的密度将指纹特征聚成多个簇。由于这种聚类方式对于环状数据具有较好的聚类效果，因此比较适合室内射频指纹的聚类。

### (4) 精确定位过程

在对指纹数据中的指纹进行聚类处理后，根据每个簇中的指纹数据，采用机器学习算法，学习指纹特征和坐标位置之间的非线性关系。经过区域定位后，可以确定待测指纹在具体某个区域内，然后在这个区域内进行精确位置确定，只需要根据这个区域内的非线性函数关系，可以得出具体位置坐标。

在定位算法的选择上，有两种方式，一种是基于回归的机器学习算法，如线性

回归, 将预测坐标和实际坐标之间的平方差作为损失函数, 求解出指纹特征和位置坐标之间的影响因子。一种是基于分类的机器学习算法, 如逻辑回归, 随机森林等, 通过离线指纹库中的指纹特征和其位置坐标学习出合适的分类函数。近年来, 有研究者提出了基于深度学习射频指纹定位算法, 通过搭建多层的神经网络, 实现对位置坐标的精确计算, 取得了较好的定位效果。

## 2.2 射频指纹点的选取和分析

### 2.2.1 射频指纹的选取

在指纹定位中, 指纹是指可以被测量且能够反映室内空间位置唯一性的信号特征向量, 因其具有唯一性的特点, 称为指纹。而指纹的类型, 可以是声音, 图像, 或者电信号。在射频指纹定位中, 射频指纹是指能够反映空间唯一性的无线电信号。实际上, 一个好的射频指纹应该具有以下几种特性, 一是空间唯一性, 由于射频定位是通过建立指纹空间和位置坐标之间的对应关系来实现定位的, 因此在室内空间中, 每个参考点处的射频指纹应各不相同, 不存在具有同一个射频指纹的多个参考点。二是时间稳定性, 即室内空间中, 一个参考点在不同时刻测量得到的射频指纹应该是一样的, 或应该在一段时间内保持较好的稳定性, 不会出现较大的波动。三是噪声鲁棒性, 由于在实际环境中, 人员的随意走动和室内温度的变化都会造成室内环境发生变化, 因此射频指纹应对于环境中的微小变化应该有一定的适应性, 射频指纹对于环境变化产生的噪声有一定的抗干扰性<sup>[44, 45]</sup>。

在众多可测量的无线电信号中, 由于接收信号强度 (receive signal strength, RSS) 能够利用现有广泛部署的无线设施和大部分的移动智能终端无需添加额外的设备即可接收 RSS, 因此作为射频指纹定位中最广泛使用的射频信号。

除此之外, 无线信号空间谱特征, 信道状态信息(channel state information, CSI)、信道的冲击响应 (channel impulse response, CIR), 信道的频率响应 (channel frequency response, CFR)、信号信噪比 (signal noise ratio, SNR), 也可以作为射频指纹<sup>[46-48]</sup>。在文献[49]中, 研究者通过将选取了 CIR 作为射频指纹, 提高了射频指纹的定位精度, 证明了物理层的信号作为的射频指纹的优越性。但是这些具有底层物理信息的指纹必须依靠专门的软硬件才能接收, 增加额外的硬件导致了成本的上升, 因此极大的限制了这些指纹的发展。因此本文选择 RSS 作为射频指纹定位中指纹。

### 2.2.2 射频指纹的分析

RSS 作为指纹也有一定的问题，一、RSS 只是记录了信号传播到监测点的强度值，保留的信息量很少。二、由于 RSS 信号在室内空间的传播受到多径效应、信号反射、信号折射、信号衍射以及人员在室内随意走动导致环境的变化等因素的影响，在同一个位置不同时间段测量的 RSS 也可能有些差别，导致了 RSS 具有了较强的不确定性。三、RSS 作为射频定位中的指纹，是基于 RSS 和实际室内空间中的距离存在着一个非线性关系，即服从对数路径损耗模型，但这个模型中的多个参数会在不同的室内空间中具有不同的取值。以上的三个因素导致了 RSS 在作为射频指纹定位中的指纹的时候，会出现较大的定位的误差，是造成定位精度下降的主要因素。

为了建立具有较好定位可靠性的离线指纹库，降低定位误差，提高定位精度，本节将全面、客观的了解、分析 RSS 的传播特性和在室内传播受到的各个因素的影响。本节，将根据实测数据从影响 RSS 传播的两个因素时间和空间分别讨论 RSS 的传播特性和统计特性。

#### (1) RSS 的空间特性

实验一：

RSS 作为射频指纹是基于 RSS 在不同的空间位置具有不同的取值，能够建立起信号空间和位置空间的对应关系。在室内空间中，RSS 在信号传播过程中服从路径损耗模型，即 RSS 跟随信号传播的距离成指数衰减。为了验证这个模型，我们设计了实验一，选择同某个 AP 距离逐渐增加的多个参考点来分析 RSS 和距离之间的对应关系。在实际环境下，选择一个 AP 作为测试 AP，保持 AP 的天线方向不变，在路由器的正前方每隔 0.6 米取一个点，在每个参考点的测试了 30 次并取平均值。

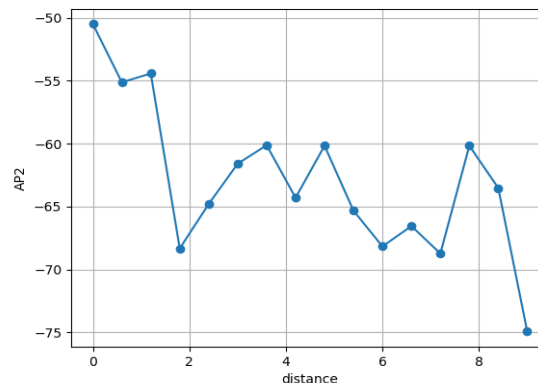


图 2-2 RSS 和传输距离之间的关系

Fig.2-2 The relationship of fingerprint received signal strength and distance

实验的结果，如上图所示。显示测试 AP 在距离其不同距离的参考点 RSS 值尽管存在波动，但是具有随着距离的增大，接收的 RSS 逐渐减少。

### 实验二:

射频指纹的定位是通过建立信号空间和位置空间的对应关系实现的, 为了验证这种对应关系, 本文在实验环境中设置了 9 个相隔 0.6 米的网格点, 作为地理位置空间, 同时选两个 AP 作为测试 AP, 与实验 1 的相同, 在每个参考点测试 30 次来自两个 AP 的信号强度值。

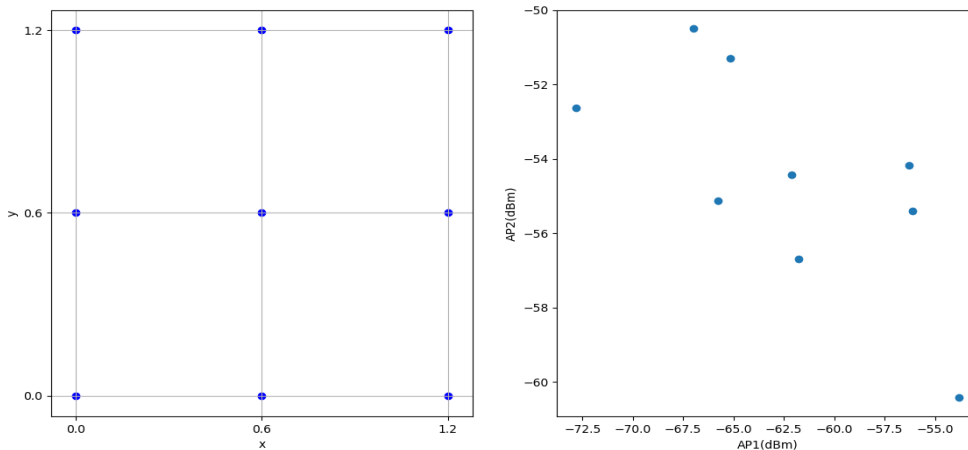


图 2-3 位置空间和指纹空间对应关系

Fig.2-3 Correspondence of Location space and fingerprint space

实验的数据结果, 如下图所示, 显示位置空间中均匀分布的 9 个参考点, 其在指纹空间的对应的 RSS 指纹分布没有规律。两个空间中不存在明显的线性关系。其次, 可以发现在指纹空间中, 指纹向量的分布非常不均匀, 有些分布比较稀疏, 有些分布比较紧密。根据指纹定位的原理可知, 在指纹分布稀疏的定位, 定位较准确, 在指纹分布稠密的地方, 定位容易出现误定位, 导致定位精度较低。

通过以上实验, 首先可以看出 RSS 在实际的室内环境下, 大致符合衰减模型, 为射频指纹定位提供了理论支持。其次, 通过实验发现位置空间和指纹信号空间不存在明确的线性关系, 且指纹空间中指纹的可区分度受到多个因素的影响。因此射频指纹定位要实现较好的定位效果, 仍面临着很大的挑战。

### (2) RSS 的时间特性

由于信号在传播过程中受到多径效应、信号反射、信号折射以及人员在室内的随意走动造成室内环境的变化, 使得在同一个位置不同的时间段收集到的信号强度也会发生变化, 即接收信号强度随着时间具有不确定性。

### 实验三:

为了验证 RSS 值跟随时间不确定性, 在实验环境中选择了一个参考点作为测试参考点, 在测试参考点 300 s 接收来自一个 AP 的信号强度。



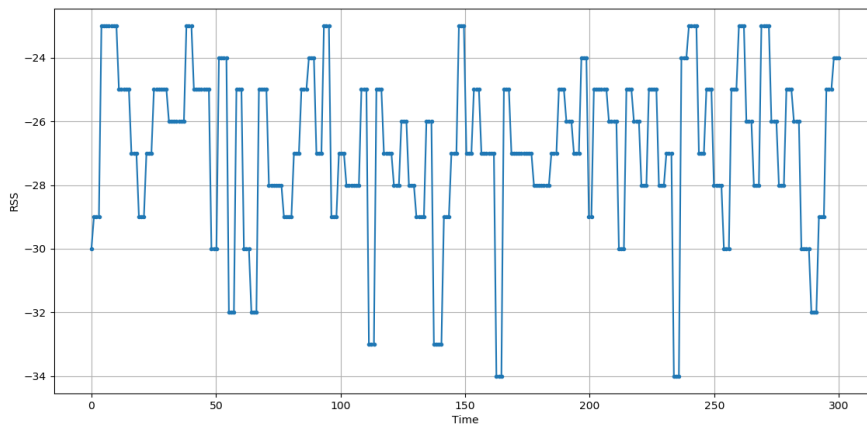


图 2-4 射频指纹信号随时间的不确定性

Fig.2-4 Uncertainty of RF fingerprint signal over time

实验结果如上图，在同一位置，连续时间段获取的信号强度值虽然总体上保持不变，但是会在一个值上下波动。实验证明了，信号空间中的指纹向量跟随时间具有较强的不确定性。

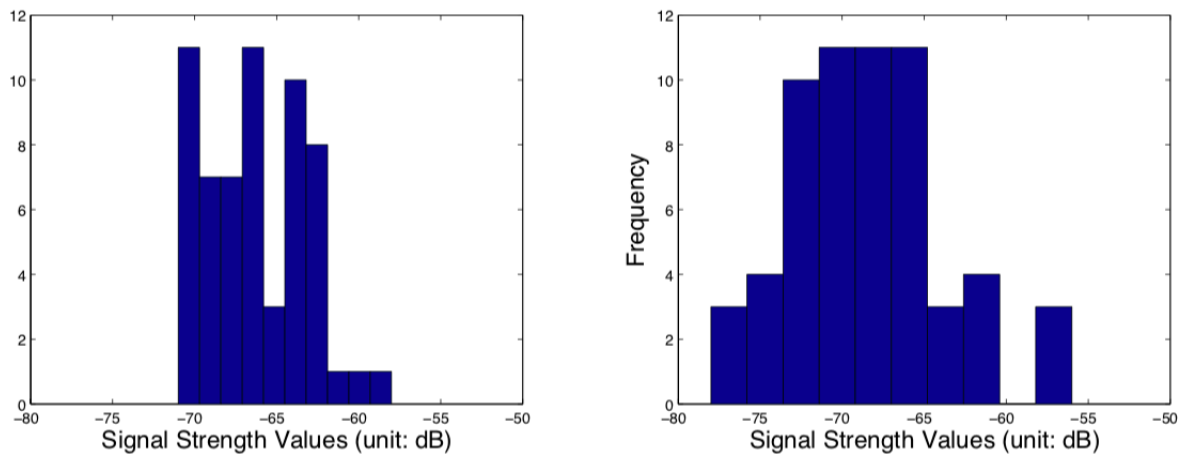


图 2-5 不同时间段收集的指纹数据分布图

Fig.2-5 Distribution of fingerprint data collected at different time periods

综上所述，在时间特征上，RSS 具有不确定性，且在文献[61]中也证明了同一室内环境不同时间段收集的指纹信号具有不同的数据分布。在空间特性上，RSS 指纹向量和位置空间具有非线性关系，信号空间和指纹空间并不是严格的一对一的关系。因此导致了指纹在信号空间中的可区分度不高，发生严重的定位误差，使得定位精度的下降。

## 2.3 本章小结

在本章中，首先概述了射频指纹定位技术的基本流程和定位原理，并从离线建

库和在线定位两个阶段讨论了射频指纹定位技术。然后分析了射频指纹中指纹的选择标准，并选择了 RSS 作为本文中的指纹。最后通过设计实验分析了 RSS 射频指纹信号的时间特性和空间特性，证明了射频指纹信号在时间上具有不确定性，为下文的基于迁移学习的定位系统提供了理论支持。

### 3 基于机器学习的射频指纹定位技术

机器学习和射频指纹定位的结合成为了近年来一个研究热点，本章首先将介绍实验中用到的数据集和常规机器学习算法的理论知识。然后通过实验来验证三种常见的机器学习算法在射频指纹定位上的应用效果，以证明机器学习和射频指纹定位的结合的可行性。最后，针对常规机器学习算法无法适应复杂的大规模的室内环境导致定位精度不高的问题，本文结合机器学习方法对指纹定位进行了创新，提出了一种基于七层深度神经网络的定位算法，并在大规模室内环境数据集 UJIINDOORLOC 上进行验证。

#### 3.1 数据集介绍

本文为验证基于机器学习的射频指纹定位系统在多种室内环境下的定位性能，在实验过程中使用了两个数据集，一个数据集是在真实的小型室内环境中采集得到的实验数据集，另一个是在大规模多建筑物多楼层的室内空间中采集的公开数据集，现在分别介绍两个数据集。

##### 3.1.1 实验数据集

本文为了验证在小规模的实际环境中射频指纹定位的性能，通过搭建真实的实验平台，收集真实室内环境下二维空间的射频指纹，建立离线数据指纹库，作为评估定位算法的真实数据。

###### 1) 实验场地

本文选定的实验场地为北京交通大学某教学楼三楼的一间教室，教室的大小长为 20m 宽为 15m，选择路由器作为无线接入器(access point, AP)，为了保证 Wi-Fi 信号能够覆盖到整个教室，在该教室布置了 6 个 AP，参考点网格间隔为 0.6m。

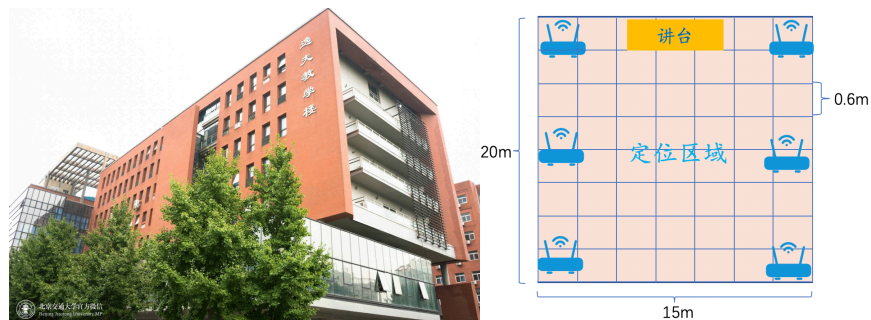


图 3-1 实验环境

Fig.3-1 Experiment configuration

## 2) 实验平台

本文设计了一个基于 Android 系统的一个信号测量系统。该系统由三个模块组成，一个是信号采集模块，一个信号处理模块，一个是指纹库建立模块。在数据采集模块中，移动智能设备通过手机自带的 Wi-Fi 连接功能连接到某一个无线热点后，信号收集模块开始采集接收到的无线信号，每次测量次数为 30 次。在这个模块中记录的信号信息包含，采用时间、AP 名称、接收信号强度、连接速度。在数据处理模块中，由于 RSS 的波动性对于定位精度有很大的影响，需要对在一个参考点接收的多次的 RSS 进行平均化处理，即对在某个参考点多次收集的指纹取平均值。在指纹库建立阶段，将建立 RSS 和位置空间的对应关系。在这个模块中，每个参考点在位置空间上由一个二维坐标表示，在 RSS 信号空间中由一个六维的 RSS 信息表示。在本文实验中，一共选取了 225 个参考点，每个参考点对应的信号维度为 8 维，其中前 6 维是来自六个 AP 的接收信号强度，第 7、8 维是参考点在室内的位置坐标。

表 3-1 记录的无线信号信息列表

Table 3-1 List of information about received radio signal

名称	描述	数据类型	举例
Time	采样时间	datetime	2019-01-11 12:48:27
SSID	AP 名称	char	TP-LINK_6274
RSSI	接收信号强度	int	-55
LinkSpeed	连接速度	int	65

### 3.1.2 公开数据集

为了验证本文提出算法的在大规模建筑物室内环境下的适用性，本文使用了 UJINDOORLOC 公开数据集<sup>[51]</sup>。该数据集是在西班牙 Jaume I 的大学的多栋教学楼测量的,是一个跨楼跨层的大型数据集，数据集定位区域的总面积为 108703 平方米，包含三栋教学楼，每栋教学楼有四层或五层，在这些区域共计布置了 520 个 AP 且布置的位置大多比较隐蔽，都布置在了天花板或墙壁上。是一个标准的用于实际建筑物和楼层识别分类的数据集。数据集的收集采用了众包技术，数据来自于 20 个用户的 25 个移动设备的采集。数据集包含了 933 个 Wi-Fi 参考点的，有 21049 条记录。其中分为 19937 的训练集和 1111 的测试集。且测试数据集是比训练数据集晚四个月采集的。

该数据集有 529 个维度，其中前 520 维是接收强度值，由于 AP 辐射范围有限，因此数据具有很强的稀疏性，接收信号强度为-100dBn 表示信号非常微弱，而信号强度为 0dB 表示信号非常强。因此前 520 维信号取值为-100dB 到 0dB 之间。

521 到 523 分别对应了该参考点在所在的经度、维度、楼层号。524 维所在楼号，525 是对应的房间类型，526 是标志是否在室内采集，527 是采集者身份标识，528 是移动设备类型，529 是采样时间点标识。

## 3.2 机器学习相关理论分析

机器学习是一门横跨多个领域的交叉学科，通过在大量的数据中让计算机能够自发的学习到数据之间的逻辑关系，挖掘出影响模型的潜在因素，建立起客观的数学模型，进而解决实际问题。机器学习经过几十年的发展，已出现多种并行研究方向，例如随机森林、梯度提升算法、聚类、神经网络、以及学习理论等，且均已取得了众多有价值的成果。

本节对于三种机器学习算法原理的阐述将通过与射频指纹定位的具体场景进行结合来介绍，为下一节的应用提供理论支持。在定位区域内，假设在大规模环境中部署了  $m$  个 AP，共选取了  $n$  个参考点。接收信号强度 (RSS)  $X = (x^1, x^2, \dots, x^j, \dots, x^n)$  被定义为信号矢量，其中  $x^j$  代表在  $j_{th}$  位置收集到的指纹， $x^j = (x_1^j, x_2^j, \dots, x_i^j, \dots, x_m^j) \in R^m$ ， $x_i^j$  代表 RSS 值，表示从第  $j_{th}$  个位置收集来自  $AP_i$  的值。 $Y = (y^1, y^2, \dots, y^j, \dots, y^n)$  表示位置标签向量，其中  $y^i$  表示第  $i$  个位置的坐标，其中  $y^i = (y_1^i, y_2^i)$  表示了该位置的横纵坐标。然后， $n$  个参考点的位置向量和射频指纹向量组成了离线指纹库。我们使用这些数据来训练可以将指纹映射到位置的模型。

### 3.2.1 近邻法 (Nearest Neighbor)

近邻法是一种非监督机器学习算法，同时也是一种经典的在线定位算法<sup>[52]</sup>，其定位的基本思想是，在已经建立的离线指纹库中，比较在待定位点测量的指纹向量和库中所有指纹向量的相似程度，选择相似程度最高的指纹向量对应的坐标作为待定位点的坐标。其中相似程度可用向量直接的欧式距离、曼哈顿距离等衡量，且向量之间的距离越近，相似程度越高，其中欧式距离的计算如式(3-1)。

$$d_{(x,y)} = \sqrt{(x^1 - y^1)^2 + (x^2 - y^2)^2 + \dots + (x^n - y^n)^2} \quad (3-1)$$

近邻算法由于是直接取库中一个指纹向量对应坐标作为待定位点的坐标，而在实际情况中，待测点很大可能不在离线指纹库点参考点上，因此定位过程中会有很大的误差。因此 K 近邻 (K nearest neighbors) 作为近邻算法的改进算法，在选择和待定位点的指纹向量相似指纹向量的时候，选择最相似的 K 个指纹向量。然后将 K 个指纹对应的坐标进行平均即得到待定位点的坐标向量，计算公式如式(3-2)。

$$(x, y) = \left( \frac{x_1 + x_2 + \dots + x_k}{k}, \frac{y_1 + y_2 + \dots + y_k}{k} \right) \quad (3-2)$$

K 近邻由于是对 K 个指纹向量对应的坐标直接求平均，而 K 个指纹和待定位点的指纹向量相近程度确不相同，因此造成了定位误差。因此加权 K 近邻 (Weight Nearest Neighbor) 作为 K 近邻的改进算法，在 K 个向量的坐标进行综合的时候，不再是直接平均求和，而是将待定位点的指纹向量和 K 个指纹向量的相近程度求反比作为坐标权重，对坐标进行加权求和即为待定位点坐标，使得预测的坐标更加准确，计算公式如式(3-3)。

$$(x, y) = \left( \frac{w_1 * x_1 + w_2 * x_2 + \dots + w_k * x_k}{k}, \frac{w_1 * y_1 + w_2 * y_2 + \dots + w_k * y_k}{k} \right) \quad (3-3)$$

$$w_i = \sum_i^k d_{(0,i)} / d_{(0,i)} \quad (3-4)$$

### 3.2.2 随机森林(Random Forest)

随机森林是一种经典的机器学习集成模型，同时也是一种常用的在线定位算法<sup>[53]</sup>。随机森林可以适用于分类和回归问题，当将定位看作一个回归问题时，即建立一个指纹向量和位置坐标之间的一个非线性回归函数，其定位的基本思想是，根据离线指纹数据库中的指纹向量和位置坐标构建多个回归决策树 (decision)，其中每个回归决策树的训练样本和样本特征是通过随机抽样得到的，且每个回归树的训练是并行的，然后将多个回归树预测结果按照集合规则进行组合。在这个过程中，常用的集合规则有取平均和加权平均两种方式。

决策树是一种简单而经典的机器学习算法，其实现定位的基本思想是，在指纹向量空间中学习一个超平面，使得通过超平面划分的每个指纹空间中的指纹向量对应同一个位置坐标向量。超平面实际上是有多个指纹特征分类边界组成，即每次根据一个 AP 的取值选择一个划分边界。其中在边界划分的过程中，划分方式包括：信息增益、信息增益比、基尼系数，其中式(3-5)为信息增益，式(3-6)为基尼系数。适用三种划分方式的决策树分别称为：ID3、C4.5、CART。其中在指纹定位算法中，使用的是 CRAT 回归树。

$$g(D, AP) = H(D) - H(D|AP) \quad (3-5)$$

其中，特征 AP 对于指纹数据集 D 的信息增益为  $g(D, AP)$ ,  $H(D)$  为指纹数据的经验熵， $H(D|AP)$  为特征 AP 在给定条件下的条件熵。

$$\text{Gini}(p) = \sum_{k=1}^K p_k(1 - p_k) \quad (3-6)$$

其中， $p_k$  为样本点属于第 k 类的概率。

在定位过程中，一颗回归决策树对应了指纹空间中的一个划分以及在划分单

元的位置坐标输出。在选择划分单元过程中，在 CRRT 回归树中采用的是启发式的方法。假设指纹空间是由  $n$  个 AP 的接收信号强度组成的，每个 AP 的可能值在  $-100\text{dB}$  到  $0\text{dB}$  之间，需要遍历所有 AP 值选择一个 AP，然后遍历该 AP 的所有取值，对指纹空间进行划分，直到取到第  $m$  个 AP 的取值  $s$ ，使得损失函数最小，即找到了一个分界点，其中损失函数见式(3-7)。

$$\min [\min (\text{Loss}(y_i, c_1) + \min (\text{Loss}(y_i, c_2)))] \quad (3-7)$$

在将指纹空间划分成了  $M$  个分类区域后， $R_1, R_2, \dots, R_m$ ，那么每个区域的输出值就是  $C_m = \text{ave}(y_i | x_i \in R_m)$ ，即该分类区域内所有位置坐标的平均值。在线定位过程中，在待定位点收集到来自于  $m$  个 AP 的接收信号强度后，根据离线阶段学习的超平面将该指纹向量划分到具体的某个分类区域内，该区域对应的位置坐标即为待定位点坐标。

### 3.2.2 梯度提升算法(Gradient Boosting Decision Tree, GBDT)

梯度提升算法是一种经典的可用于回归、分类问题的机器学习集成模型，同时也是一种重要的在线定位算法<sup>[54]</sup>。

其定位的基本思想是，根据离线指纹数据库中的指纹向量和位置坐标构建多个回归决策树 (Decision Tree)，然后将多颗树按照某种集成规则组合成一个关于指纹向量和位置坐标之间的非线性函数。但是不同于随机森林的并行的集成方式，在梯度提升算法中，每颗决策树的训练是为了弥补已有模型的不足，即回归决策树的训练是串行的。在定位过程中，输入向量是指纹特征，输出向量是对应的位置坐标，根据指纹库中的数据进行训练，学习出指纹和位置之间的非线性函数。在这个过程中，目标函数是损失函数最小。

$$L(y, F(x)) = \frac{1}{2} (y - F(x))^2 \quad (3-8)$$

其中  $y$  为坐标位置， $F(x)$  是模型学习的指纹向量和位置坐标之间的非线性函数；在求解过程中，梯度提升方法采用了数值优化的思维，通过最速下降法求解损失函数。最终得到指纹和位置之间的非线性函数。

## 3.3 基于机器学习的射频指纹定位

传统的射频指纹定位采用的是基于指纹关系匹配的完成的，但是由于射频信号在实际环境中会受到多径效应、衰减、行人干扰等诸多因素的影响，导致实际定位环境下的定位精度远低于理论定位精度。因此针对目前该领域传统定位方法精度不高的问题，本文结合机器学习方法对指纹定位进行了创新，利用常规的机器学

习算法来实现位置的确定, 这些算法能够在具有较低成本的情况下, 有效的实现位置的确定, 具有一定的创新性。本节将介绍这些算法, 并在实测的二维的实验数据集上验证三种机器学习算法在射频指纹定位上应用效果, 并分析每种算法的有效性、复杂度和影响因素。

### 3.3.1 基于 KNN 的指纹定位

基于 KNN 的射频指纹定位, 因其具有定位算法简单, 定位效果较好的优点, 而成为了目前应用最广泛的一种算法。在定位过程中, KNN 通过调整选取合理的 K 的取值来尽量减少指纹受到干扰造成的影响, 以达到较好定位效果的目的。在基于 KNN 的射频指纹的定位算法实现过程中, 在离线阶段我们根据实验环境实测的指纹数据建立指纹库, 在线阶段采用 KNN 算法进行定位。

(1) 算法原理:

离线阶段将每个位置的指纹向量  $X = (x^1, x^2, \dots, x^j, \dots, x^n)$  和位置坐标  $Y = (y^1, y^2, \dots, y^j, \dots, y^n)$  存储在离线指纹库中, 在线阶段将采集到的指纹向量  $x^j = (x_1^j, x_2^j, \dots, x_i^j, \dots, x_m^j) \in R^m$  和指纹库中指纹向量求欧式距离, 选择欧式距离最小的 K 个指纹, 并对 k 个指纹对应的位置坐标取平均值  $(y_1, y_2) = (\frac{y_1^1 + y_1^2 + \dots + y_1^k}{k}, \frac{y_2^1 + y_2^2 + \dots + y_2^k}{k})$  作为预测的位置坐标。

(2) 算法分析:

在使用 KNN 进行位置预测的过程中, K 的取值是影响定位精度的主要因素, 因此, 文本通过选择不同的 K 值来比较定位精度, 来确定最佳的 K 值;

(3) 实验分析:

为了检验基于 KNN 的射频指纹定位系统在实际定位环境中的定位效果, 在该实验中, 本文采用了实测数据集作为实验数据, 该数据集共有 225 个参考点, 每个参考点相距 0.6m, 数据集按照 7:3 划分训练集和测试集, 其中测试集上的定位精度如下所示:



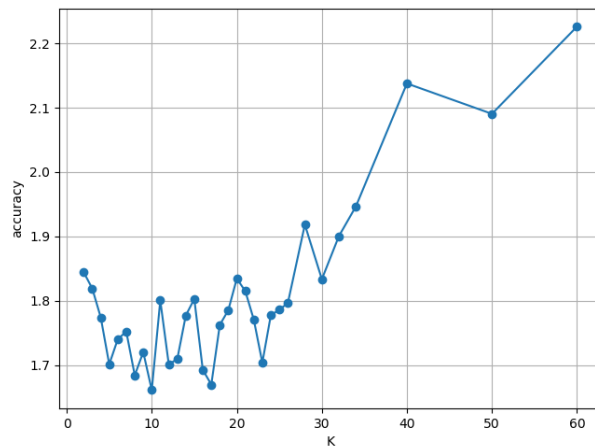


图 3-2 回归 KNN 定位性能

Fig.3-2 回归 KNN positioning performance

根据上图结果，开始的时候，随着  $K$  值的增大，定位精度逐渐增大，定位误差开始降低，当  $K$  值取得 10 的时候，定位精度达到最大值，定位误差为 1.65m，再之后，随着  $K$  的增大，定位精度开始下降；因此根据实验结论，不同的  $K$  值对于定位精度有着显著影响， $K$  值偏大和  $K$  值偏小都会导致定位精度的增大。分析原因，在于当  $K$  值取大的时候，预测的位置坐标会受到不相关的位置点的影响，当  $K$  值取值过小的时候，预测的位置会受到噪声的影响而定位精度；

### 3.3.2 基于随机森林的指纹定位

基于 Random Forest 的射频指纹定位，因具有定位效果较好的优点，而成为了目前应用较为广泛的一种算法。在定位过程中，Random Forest 通过建立多颗决策树，并在定位阶段投票选择最好的定位位置，尽量减少指纹受到干扰造成的影响。在基于随机森林的指纹定位算法实现过程中，离线阶段根据实验环境实测的指纹数据建立指纹库，在线阶段采用 Random Forest 算法实现位置的确立，来分析基于 Random Forest 的射频指纹定位技术的定位效果。

#### (1) 算法原理：

随机森林通过建立多个回归决策树实现位置的确定，离线阶段每个回归决策树会根据指纹向量  $X = (x^1, x^2, \dots, x^j, \dots, x^n)$  和位置坐标  $Y = (y^1, y^2, \dots, y^j, \dots, y^n)$  建立回归函数，在线阶段每颗决策树会根据收集到的指纹向量确定一个坐标  $(y_1, y_2)$ ，并对  $k$  颗树的位置坐标进行平均作为预测位置  $(y_1, y_2) = (\frac{y_1^1 + y_1^2 + \dots + y_1^k}{k}, \frac{y_2^1 + y_2^2 + \dots + y_2^k}{k})$ 。

#### (2) 算法分析：

在使用随机森林进行位置坐标的预测过程中，决策树的数量是决定定位精度一个重要因素，因此，在本文通过选择不同数量的决策树来比较定位精度，来确定最佳的树个数。

### (3) 实验分析:

为了检验基于随机森林的射频指纹定位系统在实际定位环境中的定位效果,在该实验中,本文采用了本文实测数据集作为实验数据,该数据集共有 225 个参考点,每个参考点相距 0.6m,数据集按照 7:3 划分训练集和测试集,其中测试集上的定位精度如下所示:

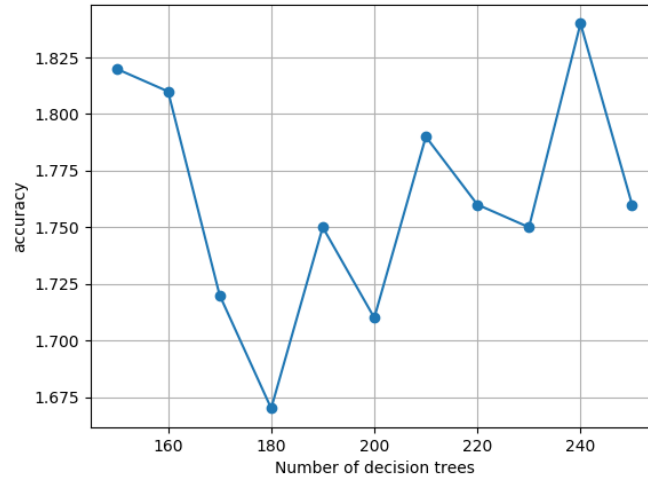


图 3-3 回归 Random Forest 定位性能

Fig.3-3 回归 Random Forest positioning performance

根据上图结果,开始的时候,随着决策树数量的增大,定位精度逐渐增大,定位误差开始降低,当决策树数量取 180 的时候,定位精度达到最大值,定位误差为 1.675 m,再之后,随着决策树数量的增大,定位精度开始下降;因此根据实验结论,不同的决策树数量对于定位精度有着显著影响,数量偏大和数量偏小都会导致定位精度的增大。分析原因,在于当决策树数量过大的时候,预测的位置坐标因为模型过拟合而导致定位精度下降,当决策树数量过小的时候,预测的位置坐标因为模型欠拟合而导致定位精度下降;

### 3.3.3 基于梯度提升算法的指纹定位

基于 GBDT 的射频指纹定位,因具有定位效果较好的优点,而成为了目前应用较为广泛的一种算法。在定位过程中,GBDT 通过串行建立多颗决策树,不断优化指纹数据和位置坐标之间的非线性函数,以尽量减少指纹受到干扰造成的影响。在本小节中,将通过在离线阶段对在实验环境实测的指纹数据建立指纹库,在线阶段采用 GBDT 算法位置确立来,分析基于 GBDT 的射频指纹定位技术的定位效果。

#### (1) 算法原理

同随机森林类似,梯度提升算法通过建立多个回归决策树来实现定位,但每颗

决策树的建立依赖于之前所有决策树组成的决策函数的预测效果，即每颗决策树都是在弥补所有之前的回归函数的定位误差,最终多颗树的决定的回归函数作为最终的位置回归函数。

### (2) 算法分析

在使用梯度提升算法进行位置坐标的预测过程中，决策树的数量是决定定位精度一个重要因素，因此，在本文通过选择不同数量的决策树来比较定位精度，来确定最佳的树个数

### (3) 实验分析

为了检验基于 KNN 的射频指纹定位系统在实际定位环境中的定位效果，在该实验中，本文采用了本文实测数据集作为实验数据，该数据集共有 225 个参考点，每个参考点相距 0.6m，数据集按照 7:3 划分训练集和测试集，其中测试集上的定位精度如下所示：

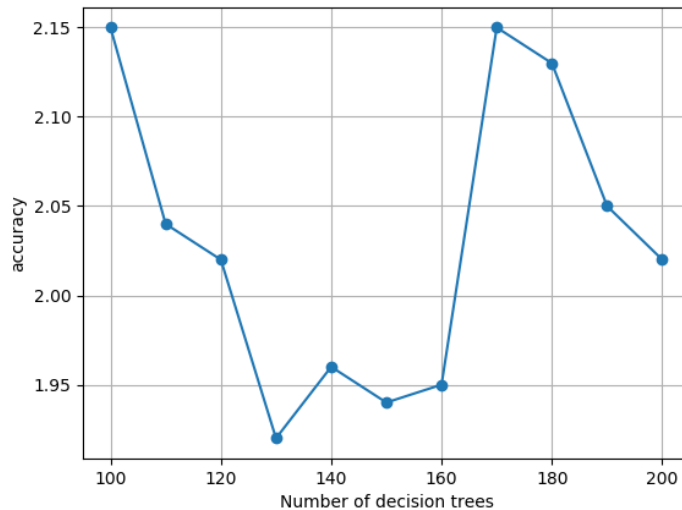


图 3-4 回归 GBDT 定位性能

Fig.3-4 Random Forest positioning performance

根据上图结果，开始的时候，随着决策树数量的增大，定位精度逐渐增大，定位误差开始降低，当决策树数量取 130 的时候，定位精度达到最大值，定位误差为 1.93 m，再之后，随着决策树数量的增大，定位精度开始下降；因此根据实验结论，不同的决策树数量对于定位精度有着显著影响，数量偏大和数量偏小都会导致定位精度的增大。分析原因，在于当决策树数量过大的时候，预测的位置坐标因为模型过拟合而导致定位精度下降，当决策树数量过小的时候，预测的位置坐标因为模型欠拟合而导致定位精度下降；

### 3.3.4 三种算法总结

综上实验，可知基于机器学习的指纹定位误差都在 2m 以内，且最好的定位精

度可以达到 1.65m，相比传统的射频定位方法，具有较好的定位效果，因此机器学习和指纹定位的结合是可行的。

同时本文对三种较常用的算法：KNN，Random Forest 和 GBDT 算法的定位精度、算法复杂度及定位成本做了分析比较，分析结果见表 3-2。通过对比发现，KNN 算法从定位精度和算法复杂度上均具有较高优势，因此成为了一种较为广泛使用的射频指纹定位算法。

本文以射频指纹定位方法为研究对象进行了较为系统的分析和研究。在国内外众多学者的研究基础上，针对目前该领域传统定位方法精度不高的问题，结合机器学习方法进行了创新。较传统的匹配定位算法，本文对提出的基于传统机器学习算法的指纹定位技术能够在具有较低成本的条件，有效的实现位置的确定，具有一定的创新性。

表 3-2 算法复杂度分析

Table 3-2 Algorithm complexity analysis I

算法	KNN	Random Forest	GBDT
定位精度	1.62m	1.675m	1.93m
定位成本	$O(n*k)$	$O((m \log n))$	$O((n \log n))$
算法复杂度	低	高	高
主要参数	K=10	N=180	N=150

注：其中 k 是 KNN 算法中 k 值，n 是样本数目，m 是特征数。

以上算法的不足在于，实验是在单一的室内环境下实现的定位，但在真实生活中的定位环境中，现在许多的定位场景都是大规模的室内定位环境，而传统的机器学习算法在复杂的大规模多空间的室内定位环境环境中会由于各个空间的物理环境的差异性导致定位效果变差。因此如何解决在大规模复杂室内环境中的定位问题，再下一节我们将提出新的算法来解决。

### 3.4 基于深度学习的射频指纹定位

如上节所述，传统的机器学习算法很难适应大规模多空间的室内环境，而在真实的定位环境中，定位场景大多是大规模的多物理空间的定位环境，在这种定位环境下，传统指纹定位由于难以适应这种多物理空间的差异性，导致定位精度下降。而随着人工神经网络不断取的新的突破，深度学习由于其更好的学习性能引起了很多关注，因此本节将尝试通过神经网络和射频指纹定位进行结合来解决这个问题。为此，本文提出了一种改进的基于七层神经网络模型的定位算法，本节中将介

绍该定位算法，并在大规模的公开数据集 UJINDOORLOC 上对该算法进行验证。

### 3.4.1 深度学习理论介绍

深度学习是一种由多层神经网络发展的学习模型<sup>[55,56]</sup>。传统的机器学习算法对于特征工程的依赖性比较大，而通过深度学习模型可以通过特征提取或分层特征提取来代替人工获取特征，来达到更容易从实例中学习任务效果。

神经网络，在 80 年代开始成为了人工智能中的研究热点，其通过对人脑神经进行抽象建立数学模型，由多层的神经元和神经元之间的不同的连接方式构成网络。神经网络在其他领域不断取得突破的同时，也成为了一种重要的指纹定位算法。其定位的基本原理是利用多层神经元之间的连接，根据指纹库中数据学习出指纹向量和位置坐标之间的非线性函数。在定位过程中，输入的向量为射频指纹向量，即  $X = [x_1, x_2, \dots, x_n]$ ，输出向量为位置坐标  $Y = [y_1, y_2, \dots, y_n]$ 。各层之间的激活函数采用非线性激活函数，在训练过程中，通过反向传播算法求解出各层神经元之间的权重，得到权重即可得到指纹向量和位置坐标之间的非线性函数

### 3.4.2 基于四层神经网络的指纹定位系统

本文提出了一种基于七层神经网络的射频指纹定位系统，具体的系统结构如下图所示 3-5 所示。在离线阶段，收集到的射频指纹将通过一个四层的自编码神经网络 (Auto-encoder)<sup>[59]</sup>，然后得到指纹向量降维后的潜在特征。在线阶段，指纹向量线通过已经搭建好的自编码网络获得降维后的特征，然后通过一个四层神经网络实现位置的确定。在此系统中，高维度的初始指纹向量通过 Auto-encoder 后转化成低纬度的深层次特征向量。由于该特征向量能够完整的恢复出原始的指纹向量，因此保留了指纹向量中最深层次的特征。由于在大规模的室内环境中，AP 的辐射范围有限，因此在参考点收集的指纹数据具有很强的稀疏性。因此经过此网络，我们提取出射频指纹中的潜在特征实现定位，提高了定位精度。

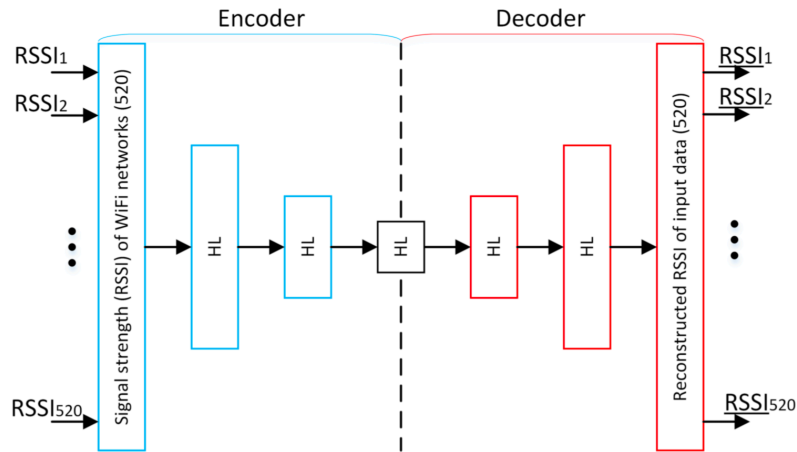


图 3-5 离线阶段框架

Fig.3-5 Offline stage framework

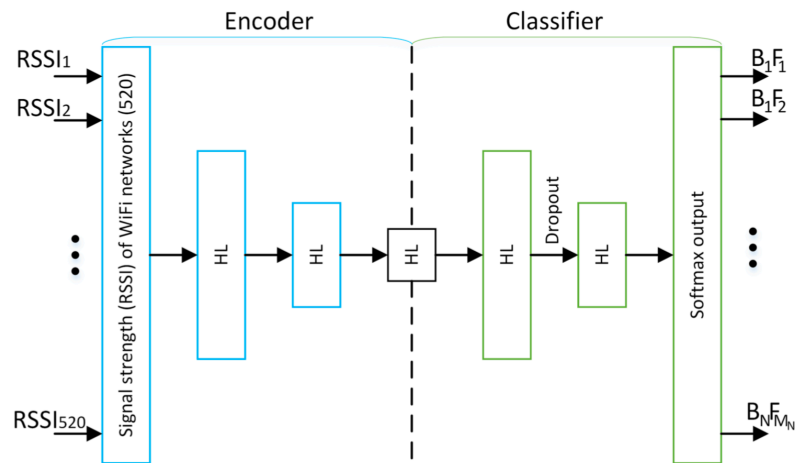


图 3-6 在线阶段框架

Fig.3-6 Online stage framework

在离线部分的网络框架中，Encoder 部分是深度学习的一部分，用于通过无监督训练期间学习原始数据，完成对高维数据进行降维处理，以提取指纹潜在的特征，Decoder 部分的目的是恢复出输入的指纹数据。如此训练 Encoder-Decoder 目的是能够学习出输入指纹数据的潜在特征。

在线部分的网络框架中，Encoder 是用来完成对指纹数据的潜在特征提取。Classifier 是将提取出的指纹数据的潜在特征作为输入，最终输出层是 SoftMax 层，其输出属于预测当前样本所属位置分类的概率。同时在 Classifier 的隐藏层之间使用了随机丢弃，通过在训练阶段随机丢失层之间的连接，以强制网络学习冗余表示，从而实现更好的泛化并避免过度拟合。

### 3.4.3 实验分析

为了检验提出的定位系统在大规模多建筑物室内环境中的定位效果，在该实验中，本文采用了 UJINDOORLOC 公开数据集作为实验数据。该数据集包含了 3 座建筑物共计 13 个楼层的数百个教室的实验数据，每个教室的物理环境均有差别，收集到指纹数据具有很强的稀疏性，因此在此大规模的室内环境中进行定位具有很大的挑战性。又因为该数据集的标签是建筑物号和楼层号，因此在该实验中定位问题转化成分类问题。同时为了验证提出的算法和传统的机器学习算法之间的区别，同时设计了在上节中提到的算法作为对照实验。具体实验结果如下：

表 3-1 机器学习和深度学习对比算法的实验结果

Table 3-1 Experimental results of machine learning and deep learning comparison algorithms

Algorithm	KNN	Random Forest	GBDT	Auto-encoder
训练集	92%	90%	89%	99%
测试集	64%	62%	61%	72%

Epoch	loss	Train accuracy	Test accuracy
1	1.4230	0.7831	0.5661
6	0.1449	0.9824	0.7281
11	0.0906	0.9961	0.7443
16	0.0546	0.9968	0.7335
20	0.0517	0.9971	0.7335

图 3-7 基于七层神经网络的定位系统的训练过程

Fig 3-7 Positioning performance of positioning system based on deep neural network

在该实验中，我们发现，基于七层神经网络的定位系统随着迭代次数的增加，定位准确率逐渐增加，最终在训练集达到了 99% 的准确率，而在测试集达到了 72%，相比传统的机器学习算法在训练数据集和测试数据集上的准确率都有了较好的改善，证明里了提出算法的可行性。

但是根据实验结果，基于神经网络的定位系统在测试集中的准确性明显低于了训练集中的准确性，通过分析，我们发现了实验数据集中测试集和训练集的采集时间相差了 4 个月，由第二章的实验结果可知 RSS 指纹信号随时间具有较强的不确定性，又由于在四个月的时间里多个教室的室内物理环境可能发生了较大的变化，

因此导致了两批数据的分布发生了变化。而测试集数据分布变化，使得训练集学习的模型不在适用于测试集数据，造成了定位性能在测试阶段的下降。在下一章中，将会通过引入新的算法来解决这个问题。

### 3.5 本章小结

首先介绍了本文中用到的模拟、实测、和公开数据集，为下文中的实验部分进行铺垫。然后阐述了本文用到的机器学习的相关技术理论。最后详细介绍了本文提出了基于机器学习的射频指纹定位算法和实验过程。

同时本文在国内外众多学者的研究基础上，针对常规的机器学习算法无法解决大规模复杂室内定位环境的问题，结合深度神经网络的方法进行了创新。较传统的定位算法，本文提出的基于七层神经网络的指纹定位技术能够在具有较低成本的条件下，有效的实现在大规模的室内环境下位置的确定，具有一定的创新性。



## 4 基于迁移学习的射频指纹定位

如上一章所述,在进行室内定位的过程中,室内环境的变化将会导致射频指纹信号强度的分布发生变化,造成定位性能的下降。与此同时,在大规模的环境中重新收集新的数据又将花费巨大的时间成本。由于迁移学习能够解决训练集和测试集的数据具有不同的数据分布的学习问题<sup>[60-62]</sup>,因此本文尝试用迁移学习去解决这个问题,提出了两种基于迁移学习的射频指纹定位算法,并在公开室内数据集 UJIINDOORLOC 进行验证算法。

### 4.1 迁移学习的理论介绍

近年来,机器学习在各个领域取得很大的进展,但机器学习算法的成功应用都有一个重要假设,即训练集和测试集具有相同的数据分布。而在实际应用的很多方面,这个假设很有可能并不成立。在实际环境中,随着时间的推移,数据的分布也在发生变化,而当训练数据集和验证数据集具有不同的数据分布的时候,大多数的统计模型需要重新收集数据并进行建模,射频指纹定位便是其中的一个例子。

迁移学习正是用来解决验证集数据分布不同于训练集数据分布导致的学习性能变差的一种技术<sup>[63]</sup>。通过迁移学习,将避免大量的重复采集并标注数据的工作,极大的提高学习性能。在迁移学习中,已有的数据称为源域,将要应用的数据称为目标域,通过将源域学习到的模型或学习到的特征迁移到目标域来实现学习性能的提高。本节将探讨迁移学习的两种类型。一种是基于样本的方法,通过改变目标域的样本权重来解决目标域和源域特征空间不一致的问题。另一种是基于特征的方法,它通过在低维空间中找到源域和目标域的共同特征来解决这个问题。

在迁移学习中,数据域由两部分组成,源域 $D_S$ 和目标域 $D_T$ ,且两个域具有不同的数据分布。学习任务有两个,源任务和目标任务,两个任务可以相同。目的是将源域 $D_S$ 在源任务学习的知识迁移到目标域 $D_T$ 在目标任务上来。

#### 4.1.1 特征迁移学习

基于特征的迁移学习,基本假设是源域 $D_S$ 和目标域 $D_T$ 存在公共特征,这个公共特征可以是在当前的特征空间下,也可以是在潜在特征空间下。基本思想是寻找源域 $D_S$ 和目标域 $D_T$ 的潜在公共特征,使得源域和目标域之间的差异能够最小化,同时能够减少分类回归任务的误差。通过将源域和目标域映射到相同的空间,使得

在这个低维度的空间下，源域和目标域的差距较小<sup>[64, 65]</sup>，或通过对源域 $D_S$ 和目标域 $D_T$ 的特征进行特征变换实现特征迁移。近年来，有研究者又不断的提出新的方法，在文献中提出了以最大均值差异(Maximum Mean Discrepancy, MMD)作为特征度量标准，以源域 $D_S$ 和目标域 $D_T$ 特征分布差异最小为目标函数，实现特征迁移，该方法称为迁移成分分析方法(transfer component analysis, TCA)。文献提出在进行特征变换的时候，可以将某一个空间中独有的特征变化映射到其他所有的空间上<sup>[66]</sup>。文献提出，在进行特征迁移实现分布差异最小化的同时，将基于样本迁移的选择方法同基于特征的迁移方法进行结合<sup>[67]</sup>。同时近年来，很多研究者又开始将神经网络和传统的基于特征迁移的方法进行结合，实现在神经网络的训练中将源域 $D_S$ 和目标域 $D_T$ 的特征进行了迁移<sup>[68]</sup>。

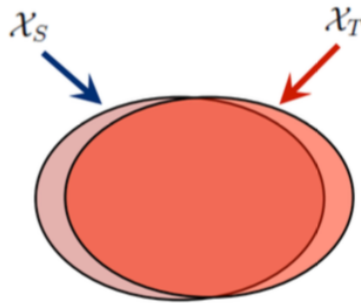


图 4-1 源域 $D_S$ 和目标域 $D_T$ 中公共特征

Fig.4-1 Common features in source domain  $D_S$  and target domain  $D_T$

#### 4.1.2 样本迁移学习

基于样本表示的迁移，基本思想是通过改变源域的数据样本权重达到新的数据分布和目标域的数据分布更相似的目的。即目标域和源域之间在高维度虽然数据分布不同，但是映射到低维度后相同的数据分布，然后再采用机器学习算法进行学习。

在文献中<sup>[69]</sup>，提出了基于 AdaBoost 的改进算法，通过降低源域中的分错样本权重，增加对样本权重，实现对源域中数据重用以调整数据分布的目的。文献中<sup>[70]</sup>，作者提出了核方法均值匹配的方法(Kernel Mean Matching, KMM)，通过引入核函数达到加权后的源域和目标域的分布之间的差值尽可能的小。在文献[71]中，作者提出了传递式的迁移学习方法(Transitive Transfer Learning, TTL)，在不同的领域内通过迁移学习进行知识迁移共享。基于样本的迁移学习在性能上和基于特征的迁移学习相比，具有理论性强，泛化上界易于证明的优点，同时也有只有在数据域中分布具有较小差异的时候才适用的缺点。

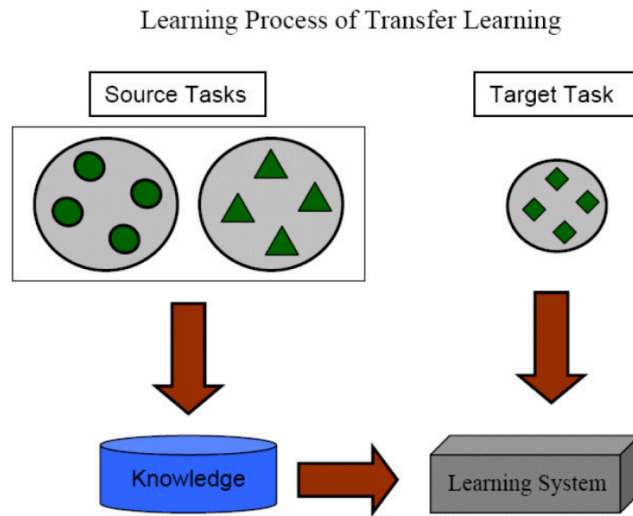


图 4-2 迁移学习框架

Fig.4-2 The framework of transfer learning

## 4.2 基于特征迁移的射频指纹定位

目前大多数的指纹定位系统在进行位置的确定的过程中，没有考虑指纹数据的分布会随着时间发生变化，因此导致了定位系统随着时间的推移定位精度不断下降。本文结合特征迁移学习方法对指纹定位进行了创新，提出了一种基于特征迁移学习指纹定位算法，该算法能够在真实的定位环境中，克服指纹数据分布变化导致定位精度下降的问题，本节将介绍该算法并在 UJIINDOORLOC 公开数据集上进行验证。

### 4.2.1 算法介绍

本文为验证基于特征迁移的指纹定位性能，选择训练集中的所有数据作为源域 $D_S$ ，选择验证集中的部分数据作为目标域 $D_T$ ，根据源域 $D_S$ 和目标域 $D_T$ 中的数据建立一个射频指纹定位模型，然后用验证集中的另一批数据来验证该定位模型

基于特征迁移的指纹定位系统包含两部分，在离线阶段，根据离线指纹数据搭建一个深层神经网络学习指纹向量和位置直接的非线性关系，并将模型的前六层权重保留。由于神经网络能够从原始数据中学习到低维度的潜在特征，因此保留前六层权重即保留了潜在特征提取方式。在线阶段，使用一部分后来采集的数据来训练网络的最后一层，最后一层作为最终的位置确定。在这个模型中，离线阶段和在线阶段的原始的指纹数据具有不同的数分布，但是通过相同的特征提取过程

之后，两批在原始维度上存在分布差异的数据映射到另一个特征空间具有相似的数据分布，从而克服了数据分布随时间发生变化导致定位精度下降的问题。

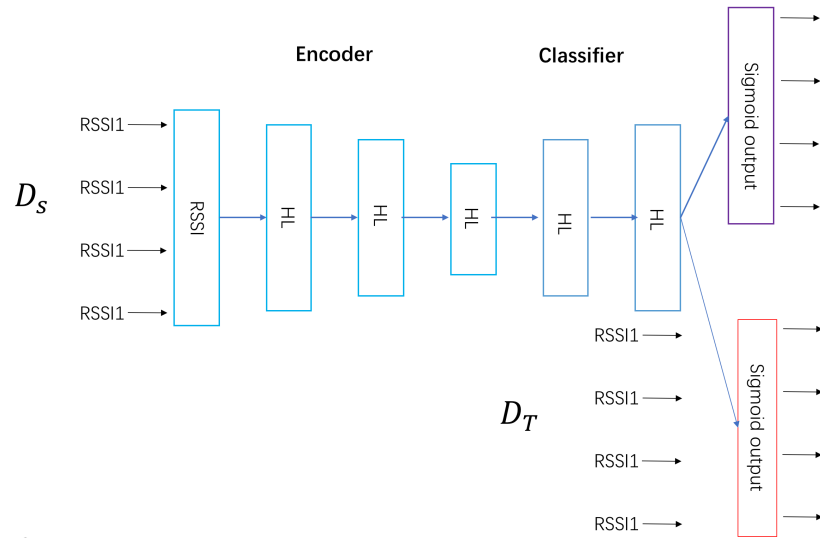


图 4-3 基于迁移学习的指纹定位系统框图

Fig.4-3 System diagram of the fingerprint positioning based on transfer learning

离线阶段的训练过程分为两个阶段，第一阶段是利用离线指纹库中的射频指纹训练出一个七层的神经网络，该七层神经网络同第三章提出的七层神经网络定位模型具有相同的结构，前四层是 Encoder 部分，该部分的目的是提取出原始指纹数据的潜在特征。后三层是 classifier 部分，该部分的目的是根据提取的潜在特征实现对位置坐标的预测，输出预测出的样本属于哪一类的位置的概率。第二个阶段是利用少量的已经标记位置的在线阶段的指纹数据，修改第一阶段学习的网络模型参数。在这个阶段中，保留了第一阶段学习网络的前六层网络连接和参数，根据少量的标记位置的在线阶段收集的指纹数据，修改最后一层模型参数。

在线阶段，指纹数据通过一个 Encoder 部分，提取出指纹数据的潜在特征，然后通过修改过最后一层网络的 classifier 部分，实现位置的确定，输出的同样是预测出的样本属于哪一类的位置的概率。

#### 4.2.2 算法可行性分析

在实际的室内环境中，一方面，在短时期内，射频指纹信号具有时间上的不确定性，即信号在传播过程中受到多径效应、反射、折射以及人员在室内的随意走动造成室内环境的变化，使得在同一个位置不同的时间段收集到的信号强度也会发生变化，即接收信号强度随着时间具有不确定性。另一方面，在长时期内，房间格局和设施的摆放都可能发生变化，导致室内环境会发生较大变化，使得在同一位置不同时间段收集到指纹信号也有较大差别。

但与此同时，由于室内放置 AP 的位置没有变化，参考点和 AP 位置之间的相

对位置也没有发生变化，因此具有在这种情况下，不同时间段在同一个室内环境收集的指纹数据尽管有不同的数据分布，但是两批数据仍有潜在联系，因此可以通过迁移学习来实现当室内环境发生变化时提高指纹定位性能。因此通过特征迁移学习，将离线阶段和在线阶段的指纹数据映射到相同维度上，使得在原本维度具有不同的分布的两批数据，在低维度上具有相同的数据分布，从而解决了由于离线阶段和在线阶段具有不同数据分布导致的定位精度下降的问题。

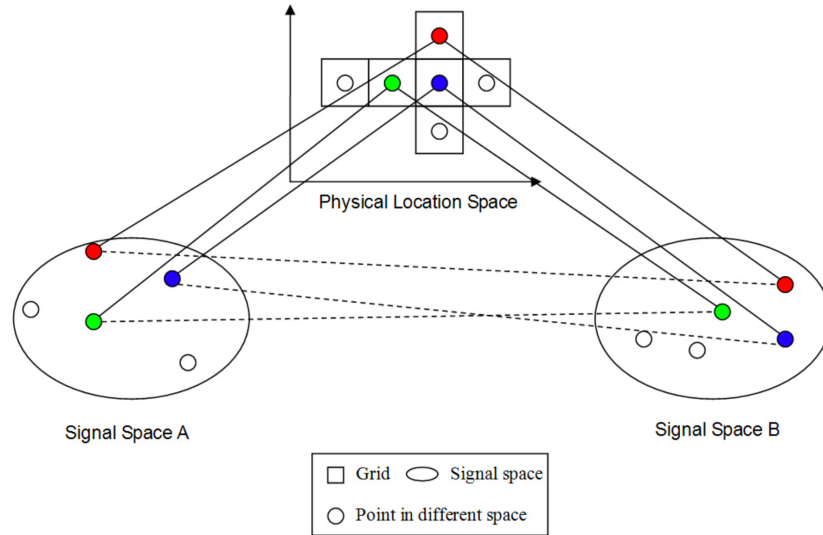


图 4-4 同一位置不同指纹空间之间的潜在联系

Fig.4-4 Potential links between different fingerprint spaces in the same location

### 4.2.3 实验分析

在该实验中，本文采用了 UJINDOORLOC 公开数据集作为实验数据。如上一章所述，由于 UJINDOORLOC 的验证集数据和训练集数据的采集时间相差 4 个月，导致验证集和训练集具有不同的数据分布，从而使得传统定位算法在测试集上的指纹定位性能较差。本文提出的用于解决该问题的算法结果如图所示，相比于直接用通过训练集学习的模型应用到验证集上的模型，基于特征迁移的射频指纹定位准确度为 84%，提高了 12 个百分点。

Epoch	Loss	Train accuracy	Test accuracy
1	0.9743	0.9919	0.7612
6	0.0980	0.9956	0.8129
11	0.0931	0.9955	0.8354
16	0.1067	0.9954	0.8451
20	0.0822	0.9956	0.8451

图 4-5 基于特征迁移的指纹定位系统性能

Fig 4-5 Performance of fingerprint positioning system based on feature transfer learning

本文在国内外众多学者的研究基础上，针对指纹数据的分布的时变性导致的定位精度下降的问题，结合迁移学习方法进行了创新。较常规的射频指纹定位算法，本文提出的基于样本迁移学习算法的指纹定位技术能够在具有较低成本的前提下，有效的解决定位精度随时间下降的问题，具有一定的创新性。

### 4.3 基于样本迁移的射频指纹定位

如上节所述，目前众多的研究学者对指纹定位进行研究的过程中，没有考虑指纹数据的分布会随着时间发生变化，因此导致了定位系统随着时间的推移定位精度不断下降。本文结合样本迁移学习方法对指纹定位进行了创新，提出一种基于样本迁移学习指纹定位算法，该算法能够在真实的定位环境中，克服指纹数据分布变化导致定位精度下降的问题。本节将介绍该算法并在公开数据集 UJIINDOORLOC 进行验证。

#### 4.3.1 算法介绍

本文为验证基于样本迁移的指纹定位性能，选择训练集中的所有数据作为源域 $D_S$ ，选择验证集中的部分数据作为目标域 $D_T$ ，根据源域 $D_S$ 和目标域 $D_T$ 中的数据建立一个射频指纹定位模型，然后用验证集中的另一批数据来验证该定位模型。在离线过程中，初始化源域 $D_S$ 和目标域 $D_T$ 的样本权重并训练多个弱分类器，对于来自源域的样本，如果分类正确则增大样本权重，如果分类错误则减少样本权重。对于来自目标域的样本，则样本权重的调整则刚好相反，如果分类正确则减少样本权重，如果分类错误则增大样本权重。其中权重的调整方式如式(4-1)。经过多次迭代，最终源域和目标域中的样本权重趋于稳定。然后根据源域中的样本权重，取较大权重的数据样本和目标域中样本组成训练集，建立定位模型。具体的算法过程如表 4-1 所示。

在模型中，由于更改了离线阶段的样本的权重，使得原本存在分布差异的两批数据在经过样本权重调整后具有相似的数据分布，从而克服了数据分布不同噪声的定位精度下降的问题。

$$W_i^{t+1} = \begin{cases} W_i^t \beta^{|f(x_i)|} & x_i \in X_{T_0} \\ W_i^t \beta^{-|f(x_i)|} & x_i \in X_{T_1} \end{cases} \quad (4-1)$$

$$f_t(x_i) = \begin{cases} 1 & \text{if } h_t(x_i) \neq y_{x_i} \\ 0 & \text{if } h_t(x_i) = y_{x_i} \end{cases} \quad (4-2)$$

$$\beta_{T_1} = \frac{\varepsilon_{T_1}}{1 - \varepsilon_{T_1}} \quad \beta = \frac{\varepsilon_{T_1}}{1 + \sqrt{\ln \frac{n}{N}}} \quad (4-3)$$

$$\varepsilon_{T_1} = \frac{\sum_{i=1}^k w_i^t \cdot f(x_i)}{\sum_{i=1}^k w_i^t} \quad (4-4)$$

$$F(x) = \sum_{i=1}^t \beta_{T_1}^i \cdot h_i(x_i) \quad (4-5)$$

表 4-1 基于样本迁移的指纹定位算法图

Table 4-1 Fingerprint localization algorithm based on sample migration

算法：样本权重调整算法
<p><b>输入：</b> 已经标记位置的旧数据集 <math>X_{T_0}</math>，少量标记位置的新数据集 <math>X_{T_1}</math>，选定的基础算法，最大迭代次数 <math>N</math>；</p> <p><b>输出：</b> 新数据分布的数据样本、定位模型；</p> <p><b>开始：</b></p> <ol style="list-style-type: none"> <li>1. 初始化两批数据集中的样本权重；</li> <li>2. 循环迭代（<math>t \leq N</math>） <ol style="list-style-type: none"> <li>a) 权重标准化 <math>P_t = \frac{W}{\sum_{i=1}^{n+m} w_i^t}</math></li> <li>b) 将 <math>X_{T_0}</math> 和 <math>X_{T_1}</math> 作为训练数据集。训练指纹向量和位置坐标之间的非线性函数 <math>h_t : X \rightarrow Y</math>；</li> <li>c) 计算 <math>h_t</math> 在数据集 <math>X_{T_0}</math> 和 <math>X_{T_1}</math> 的错误率；</li> <li>d) 按照(4-3)更新 <math>\beta_{T_1}</math>；</li> <li>e) 按照(4-1)更新权重；</li> </ol> </li> <li>3. 对于 <math>X_{T_0}</math> 中样本根据权重进行排序，选择样本权重最大的一批数据。</li> <li>4. 输出定位模型； <math display="block">F(x) = \sum \beta_{T_1}^t \cdot h_t(x_i)</math> </li> </ol> <p><b>结束</b></p>

### 4.3.2 算法可行性分析

如上节所述，指纹定位性能随着时间的推移不断下降，其中一个原因就是指纹向量的分布发生了变化。基于特征迁移的指纹定位思想是通过将两个数据映射到

低维空间中，使得在高维空间中分布不同的数据在低维度空间中具有相似分布。而基于样本迁移的指纹定位思想是通过改变训练集中样本的样本权重，对源域 $D_s$ 对样本根据权重进行重新采样，使得训练集中的数据分布和验证集中数据分布保持一致；

### 4.3.3 实验分析

为验证提出算法的可行性，我们设计了两个对照实验，实验一用原始训练集训练模型，用验证集验证模型。实验二，将原始验证集划分成两部分，一部分用于训练模型，另一部分用于验证模型。实验结果显示，实验一的建筑物和楼层定位在验证集的准确率为 74%，分析原因在于实验一的训练数据和验证数据由于采集时间相差四个月，两批数据的分布出现较大的差异，导致在验证集的定位效果较差。实验二的结果显示建筑物和楼层定位准确率为 78%，分析原因在于实验二的训练数据是验证集的其中一部分，因为数据量太小，导致定位性能较差。对比显示，基于特征迁移定位准确率为 90%，分析原因 16% 的准确率提高，在于通过样本迁移调整了数据分布，使得定位性能得到了改善。

表 4-2 基于样本迁移的指纹定位实验结果

Table 4-2 Fingerprint localization experiment results based on sample migration

Algorithm	实验一	实验二	本文提出算法
Accuracy	74%	78%	90%

本文以迁移学习和射频指纹定位方法为研究对象进行了较为系统的分析和研究，在国内外众多学者的研究基础上，针对指纹数据的分布的时变性导致的定位精度下降的问题，结合迁移学习方法进行了创新。较传统的匹配定位算法，本文对提出的基于特征迁移学习算法的指纹定位技术能够在具有较低成本条件下，有效的解决定位精度随时间下降的问题，具有一定的创新性。

## 4.4 本章小结

在本章中，首先介绍了传统的机器学习射频指纹定位技术面临的问题，即射频指纹数据分布的变化导致了室内定位精度的下降。然后介绍了本文需要用到的迁移学习的相关理论知识。最后，详细阐述了本文提出了两种算法基于特性迁移学习的射频指纹定位算法和基于样本迁移学习的射频指纹定位算法，即通过学习原来



指纹数据分布的潜在特征并运用到新的指纹数据上的基于特征迁移的指纹定位算法和通过引入一部分新数据改变原来指纹数据分布的基于样本迁移的指纹定位算法，并详细介绍了算法的实验过程。最后我们对两种算法的定位效果和算法复杂度做了对比，具体结果如表 4-2。根据对照结果，可知基于样本迁移的指纹定位系统定位效果好，算法复杂度低，因此具有较好的定位效果。

表 4-2 基于样本迁移的指纹定位和基于特征迁移度指纹定位对照

Table 4-2 Fingerprint localization based on sample migration and fingerprint migration based on feature mobility

算法	定位精度	算法复杂度
基于特征迁移的射频指纹定位	84%	复杂
基于样本迁移的射频指纹定位	90%	简单

## 5 结论

随着无线局域网络在全球范围内的广泛部署和移动智能设备的迅速普及，射频指纹室内定位技术由于具有精度高、成本低的优越性能，成为近年来一种主流的室内定位技术解决方案。但与此同时，因为室内环境复杂多变、指纹信号在时间上具有较高的不确定性，因此实现高精度的射频指纹定位仍具有很大的挑战。鉴于人工智能技术在各个领域不断取得突破，本文将机器学习和射频指纹定位进行结合，寻求具有高定位精度的射频指纹定位算法。主要的工作与成果总结如下：

1 搭建了一个射频指纹定位系统平台，通过在实际室内环境下部署的射频指纹定位系统，收集射频指纹实验数据，并分析指纹信号的统计特性，以实现射频指纹的精确定位。通过对指纹信号的时间和空间特性进行分析，发现：1) 射频指纹信号在时间特性上具有较高的不确定性，即在同一位置点不同的时间段收集的射频指纹信号具有不同的数据分布。这是由于室内环境的复杂多变，信号在传播过程中受到多径效应、折射、反射等因素等影响，因此不同时间段收集到的指纹数据分布存在差异。2) 射频指纹信号在信号传播过程中服从路径损耗模型，即 RSS 随信号传播的距离成指数衰减，实验数据表明，尽管测量 RSS 值存在波动，但随着传播距离的增加，RSS 整体呈指数下降趋势。3) 射频指纹空间和实际位置空间不存在明显的线性关系，即在实际环境中，位置坐标和指纹向量没有严格的一对一的关系，指纹向量和位置坐标的对应关系由于噪声等因素的影响具有较大的不确定性。这种不确定的非线性关系是影响定位精度的主要因素。4) 在指纹射频指纹空间中，指纹向量的分布非常不均匀，有些分布比较稀疏，有些分布比较紧密。根据指纹定位的原理可知，在指纹分布稀疏的定位，定位较准确，在指纹分布稠密的地方，定位容易出现误定位，导致定位精度较低。这种指纹向量之间的区分度不确定同样是造成定位误差的重要原因。

2 针对传统射频指纹定位系统难以建立指纹信号和地理位置之间准确的数学模型的问题，本文先提出了基于机器学习的射频指纹定位算法，利用机器学习能够在大规模数据挖掘潜在特征的优越性能，建立客观的数学模型，收集室内射频指纹向量，实现位置的确定。接着在实际室内环境中，验证了 KNN, Random Forest 在内的多种机器学习回归和分类算法的定位性能，并对每种算法影响定位性能的因素进行了分析。最后，本文提出了一种基于深度学习的射频指纹方法，由于深度学习对于数据量较少的数据集处理性能较差，因此本文选择 UJIINDOORLOC 数据集作为实验数据。该方法先通过将原始的指纹信号输入到四层的深度神经网络中挖掘出射频指纹中的潜在特征，然后再将该特征输入到一个三层的神经网络中进行

定位。实验结果表明，对于同批采集的验证集，定位准确度为 96%，但对于不同批采集的验证集，准确度为 72%。原因在于随着时间的推移，后采集的验证集数据分布与训练集数据分布存在差异，导致了定位精度的下降。

3 在基于机器学习的射频指纹定位的实验过程中发现，指纹定位存在随着时间的推移定位精度快速下降的问题。在系统分析了影响定位精度的各个因素后，发现是射频信号的分布随着室内环境的变化而变化，导致定位系统时效性较差。针对这个问题，我们提出了两种基于迁移机器学习的射频指纹算法，一种是基于特征的迁移机器学习，该方法是通过挖掘不同时间段收集到指纹向量的潜在指纹特征来实现精确定位，由于该潜在特征具有更好的时间适应性，因此在 UJIINDOORLOC 数据集上进行验证表明定位准确度提高到 82%，相比传统算法提高了 10%。另一种是基于样本的迁移机器学习，这个方法是通过引入一小部分新的指纹数据来改变原始指纹数据分布，同样在 UJIINDOORLOC 数据集上进行实验验证，表明定位准确度提高到 90%，相比传统算法提高了 18%。

上述结果表明，在射频指纹定位系统中，采用机器学习寻找射频指纹与位置的对应关系，只要训练的数据集足够大，数据的时间一致性好，就能获得较高的定位准确率，不失为一种有效的定位方法。

## 参考文献

- [1] 邓中亮, 余彦培, 徐连明, 袁协. 室内外无线定位与导航[M]. 北京:北京邮电大学出版社, 2000
- [2] Leondes C T, Yonezawa K. Evaluation of geometric performance of global positioning system[J]. IEEE Transactions on Aerospace and Electronic Systems, 1978 (3): 533-539.
- [3] Misra P, Enge P. Global Positioning System: signals, measurements and performance second edition[J]. Massachusetts: Ganga-Jamuna Press, 2006.
- [4] Yang Z, Wu C, Liu Y. Locating in fingerprint space: wireless indoor localization with little human intervention[C]//Proceedings of the 18th annual international conference on Mobile computing and networking. ACM, 2012: 269-280.
- [5] So J, Lee J Y, Yoon C H, et al. An improved location estimation method for Wi-Fi fingerprint-based indoor localization[J]. International Journal of Software Engineering and Its Applications, 2013, 7(3): 77-86.
- [6] Michie D, Spiegelhalter D J, Taylor C C. Machine learning[J]. Neural and Statistical Classification, 1994, 13.
- [7] Dietterich T G. Ensemble methods in machine learning[C]//International workshop on multiple classifier systems. Springer, Berlin, Heidelberg, 2000: 1-15.
- [8] Liu CH, Lo JY. The study for the WLAN with ZigBee positioning system[C]//Networked Computing and Advanced Information Management(NCM), 2010 Sixth International Conference on. IEEE, 2010: 520-525.
- [9] Want R, Hopper A, Falcao V, et al. The active badge location system[J]. ACM Transactions on Office Information Systems(TOIS), 1992, 10(1):91-102.
- [10] Nakamori E, Tsukuda D, Fujimoto M et al. A new indoor position estimation method of RFID tags for continuous moving navigation systems[C]//2012 International Conference on Indoor Positioning and Indoor Navigation. IEEE, 2012.
- [11] Mahfouz MR, Fathy AE, kuhn MJ et al. Recent trends and advances in UWB positioning [C]//IEEE MTT-S International Microwave Workshop on Wireless Sensing, Local Positioning, and RFID (IMWS 2009 - Croatia). IEEE, 2009.
- [12] 赵培 李凯 张需博等. 室内无线通信技术原理与工程实践[M]. 北京邮电大学出版社, 2015.
- [13] 郎新培 许可 赵明. 基于无线局域网的位置定位技术研究和发发展[J]. 计算机科学, 2006, 33(6):21-24.
- [14] 方旭明. 移动 Ad Hoc 网络研究与发展现状. 数据通信, 2003(4):15-18.
- [15] Chen Y., Luo R.. Design and implementation of a WiFi-based local locating system[C]// Portable Information Devices 2007, IEEE International Conference on. IEEE, 2007: 1-5.
- [16] Liu H, Lo W, Tseng C. A wifi-based weighted screening method for indoor positioning systems[J]. Wireless Pers Commun, 2014, 7: 611-628.
- [17] Hartley R I, Sturm P. Triangulation[J]. Computer vision and image understanding, 1997, 68(2): 146-157.

- [18] Flick U. Triangulation in qualitative research[J]. *A companion to qualitative research*, 2004, 3: 178-183.
- [19] Bahl P, Padmanabhan V N, Bahl V, et al. RADAR: An in-building RF-based user location and tracking system[J]. 2000.
- [20] Bahl V, Padmanabhan V. Enhancements to the RADAR user location and tracking system[J]. 2000.
- [21] Castro P, Chiu P, Kremenek T, et al. A probabilistic room location service for wireless networked environments[C]//*International conference on ubiquitous computing*. Springer, Berlin, Heidelberg, 2001: 18-34.
- [22] Youssef M, Agrawala A. The Horus location determination system[J]. *Wireless Networks*, 2008, 14(3): 357-374.
- [23] 郎昕培, 许可, 赵明. 基于无线局域网的位置定位技术研究和发展的[J]. *计算机科学*, 2006, 33(6): 21-24
- [24] Li B, Wang Y, Lee H K, et al. Method for yielding a database of location fingerprints in WLAN[J]. *IEE Proceedings-Communications*, 2005, 152(5): 580-586.
- [25] Lee M, Han D. Voronoi tessellation based interpolation method for Wi-Fi radio map construction[J]. *IEEE Communications Letters*, 2012, 16(3): 404-407.
- [26] Pulkkinen T, Roos T, Myllymäki P. Semi-supervised learning for wlan positioning[C]//*International Conference on Artificial Neural Networks*. Springer, Berlin, Heidelberg, 2011: 355-362.
- [27] Pan J J, Yang Q. Co-localization from labeled and unlabeled data using graph laplacian[J]. *Tc*, 2007, 40: 70.
- [28] Pan S J, Yang Q. A survey on transfer learning[J]. *IEEE Transactions on knowledge and data engineering*, 2010, 22(10): 1345-1359.
- [29] Yang Q, Pan S J, Zheng V W. Estimating Location Using Wi-Fi[J]. *IEEE Intelligent Systems*, 2008, 23(1): 8-13.
- [30] Howe J. *Crowdsourcing: How the power of the crowd is driving the future of business*[M]. Random House, 2008.
- [31] Youssef M A, Agrawala A, Shankar A U. WLAN location determination via clustering and probability distributions[C]//*Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003.(PerCom 2003)*. IEEE, 2003: 143-150.
- [32] Chen Y, Yang Q, Yin J, et al. Power-efficient access-point selection for indoor location estimation[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2006, 18(7): 877-888.
- [33] Kushki A, Plataniotis K N, Venetsanopoulos A N. Intelligent dynamic radio tracking in indoor wireless local area networks[J]. *IEEE Transactions on Mobile Computing*, 2010, 9(3): 405-419.
- [34] Fang S H, Lin T N, Lin P C. Location fingerprinting in a decorrelated space[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2008, 20(5): 685-691.
- [35] Deng Z A, Xu Y B, Ma L. Indoor positioning via nonlinear discriminative feature extraction in wireless local area network[J]. *Computer Communications*, 2012, 35(6): 738-747.
- [36] Youssef M, Agrawala A. Location-clustering techniques for WLAN location determination systems[J]. *International Journal of Computers and Applications*, 2006, 28(3): 278-284.

- [37] Schloter C P, Aghajan H. Wireless symbolic positioning using support vector machines[C]//Proceedings of the 2006 international conference on Wireless communications and mobile computing. ACM, 2006: 1141-1146.
- [38] Feng C, Au W S A, Valaee S, et al. Received-signal-strength-based indoor positioning using compressive sensing[J]. IEEE Transactions on mobile computing, 2012, 11(12): 1983-1993.
- [39] Arya A, Godlewski P, Campedel M, et al. Radio database compression for accurate energy-efficient localization in fingerprinting systems[J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(6): 1368-1379.
- [40] 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): 689-705.
- [41] Brunato M, Battiti R. Statistical learning theory for location fingerprinting in wireless LANs[J]. Computer Networks, 2005, 47(6): 825-845.
- [42] Wu Z, Li C, Ng J K Y, et al. Location estimation via support vector regression[J]. IEEE Transactions on mobile computing, 2007, 6(3): 311-321.
- [43] Pan J J, Kwok J T, Yang Q, et al. Multidimensional vector regression for accurate and low-cost location estimation in pervasive computing[J]. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(9): 1181-1193.
- [44] 杨铮, 吴陈沐, 刘云浩. 位置计算:无线网络定位与可定位性[M].北京:清华大学出版社, 2014
- [45] 万群, 郭贤生, 陈章鑫. 室内定位理论、方法和应用[M].北京:电子工业出版社, 2014
- [46] Kleisouris K, Chen Y, Yang J, et al. Empirical evaluation of wireless localization when using multiple antennas[J]. IEEE Transactions on Parallel and Distributed Systems, 2010, 21(11): 1595-1610.
- [47] Sen S, Choudhury R R, Nelakuditi S. SpinLoc: Spin once to know your location[C]//Proceedings of the Twelfth Workshop on Mobile Computing Systems & Applications. ACM, 2012: 12.
- [48] Sen S, Radunovic B, Choudhury R R, et al. You are facing the Mona Lisa: Spot localization using PHY layer information[C]//Proceedings of the 10th international conference on Mobile systems, applications, and services. ACM, 2012: 183-196.
- [49] Nypan T, Gade K, Hallingstad O. Vehicle positioning by database comparison using the Box-Cox metric and Kalman filtering[C]//Vehicular Technology Conference. IEEE 55th Vehicular Technology Conference. VTC Spring 2002 (Cat. No. 02CH37367). IEEE, 2002, 4: 1650-1654.
- [50] Yang C F, Wu B C, Ko C J. A ray-tracing method for modeling indoor wave propagation and penetration[J]. IEEE transactions on Antennas and Propagation, 1998, 46(6): 907-919.
- [51] Torres-Sospedra J, Montoliu R, Martínez-Usó A, et al. UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems[C]//2014 international conference on indoor positioning and indoor navigation (IPIN). IEEE, 2014: 261-270
- [52] Zhang M L, Zhou Z H. ML-KNN: A lazy learning approach to multi-label learning[J]. Pattern recognition, 2007, 40(7): 2038-2048.
- [53] Liaw A, Wiener M. Classification and regression by randomForest[J]. R news, 2002, 2(3): 18-22.
- [54] Ke G, Meng Q, Finley T, et al. Lightgbm: A highly efficient gradient boosting decision tree[C]//Advances in Neural Information Processing Systems. 2017: 3146-

- [55] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers [J]. *Neural processing letters*, 1999, 9(3): 293-300.
- [56] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. *nature*, 2015, 521(7553): 436.
- [57] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: theory and applications[J]. *Neurocomputing*, 2006, 70(1-3): 489-501.
- [58] Park J, Sandberg I W. Universal approximation using radial-basis-function networks[J]. *Neural computation*, 1991, 3(2): 246-257.
- [59] Ng A. Sparse autoencoder[J]. *CS294A Lecture notes*, 2011, 72(2011): 1-19.
- [60] Zheng V W, Pan S J, Yang Q, et al. Transferring Multi-device Localization Models using Latent Multi-task Learning[C]//AAAI. 2008, 8: 1427-1432.
- [61] Hu D H, Yang Q. Transfer learning for activity recognition via sensor mapping[C]//Twenty-second international joint conference on artificial intelligence. 2011.
- [62] Pan S J, Tsang I W, Kwok J T, et al. Domain adaptation via transfer component analysis[J]. *IEEE Transactions on Neural Networks*, 2011, 22(2): 199-210.
- [63] Long M, Wang J, Ding G, et al. Transfer joint matching for unsupervised domain adaptation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 1410-1417.
- [64] Duan L, Tsang I W, Xu D. Domain transfer multiple kernel learning[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(3): 465-479.
- [65] Pan S J, Tsang I W, Kwok J T, et al. Domain adaptation via transfer component analysis[J]. *IEEE Transactions on Neural Networks*, 2011, 22(2): 199-210.
- [66] Blitzer J, McDonald R, Pereira F. Domain adaptation with structural correspondence learning[C]//Proceedings of the 2006 conference on empirical methods in natural language processing. Association for Computational Linguistics, 2006: 120-128.
- [67] Borgwardt K M, Gretton A, Rasch M J, et al. Integrating structured biological data by kernel maximum mean discrepancy[J]. *Bioinformatics*, 2006, 22(14): e49-e57.
- [68] Long M, Wang J, Ding G, et al. Transfer joint matching for unsupervised domain adaptation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 1410-1417.
- [69] Dai, W., Yang, Q., Xue, G.-R., and Yu, Y. (2007). Boosting for transfer learning. In *ICML*, pages 193–200.
- [70] Huang J, Gretton A, Borgwardt K, et al. Correcting sample selection bias by unlabeled data[C]//Advances in neural information processing systems. 2007: 601-608.
- [71] Tan B, Song Y, Zhong E, et al. Transitive transfer learning[C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015: 1155-1164.

## 作者简历及攻读硕士/博士学位期间取得的研究成果

### 一、作者简历

姓名：胡安民 性别：男 汉族：汉 出生年月：1994-07-01 籍贯：河南省信阳市  
2013-09——2017-07 河南师范大学计算机与信息工程学院  
2017-09——2019-07 北京交通大学电子信息工程学院

### 二、发表论文

- [1] Anmin Hu and Lijun Zhang Adaptive Indoor Localization with Wi-Fi Based on Transfer Learning [C]//2019 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW).

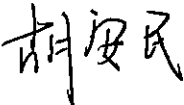
### 三、参与科研项目

- [1] 基于大数据的提案系统及其实施效果相关舆情演进分析  
.  
.  
.



## 独创性声明

本人声明所提交的学位论文是本人在导师指导下进行的研究工作和取得的研究成果，除了文中特别加以标注和致谢之处外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京交通大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名： 签字日期：2019 年 6 月 3 日

## 学位论文数据集

表 1.1: 数据集页

关键词*	密级*	中图分类号	UDC	论文资助
指纹定位、机器学习、迁移学习	公开			
学位授予单位名称*	学位授予单位代码*	学位类别*	学位级别*	
北京交通大学	10004	专业硕士	硕士	
论文题名*	并列题名			论文语种*
基于机器学习的室内射频指纹定位技术研究				中文
作者姓名*	胡安民	学号*	17125027	
培养单位名称*	培养单位代码*	培养单位地址	邮编	
北京交通大学	10004	北京市海淀区西直门外上园村 3 号	100044	
工程领域*	研究方向*	学制*	学位授予年*	
电子与通信工程	室内定位	两年制	2019 年	
论文提交日期*	2019 年 6 月 3 日			
导师姓名*	张立军	职称*	副教授	
评阅人	答辩委员会主席*	答辩委员会成员		
	沈波	王晓、陶丹、赵永祥、徐少毅		
电子版论文提交格式 文本 ( ) 图像 ( ) 视频 ( ) 音频 ( ) 多媒体 ( ) 其他 ( ) 推荐格式: application/msword; application/pdf				
电子版论文出版 (发布) 者	电子版论文出版 (发布) 地		权限声明	
论文总页数*	58 页			
共 33 项, 其中带*为必填数据, 为 21 项。				