



研究与开发

基于移动蜂窝网的机器学习室外指纹定位方案

周志超, 冯毅, 夏小涵, 冯瑜瑶, 蔡超, 邱佳慧, 杨立辉, 乌云霄

(中国联合网络通信有限公司智网创新中心, 北京 100048)

摘 要: 基于移动蜂窝网络技术的定位方案是提供网络优化、紧急救援、公安巡警和位置服务等应用的重要技术途径之一。传统的基于小区基站位置信息的定位方案定位精度低、定位误差大, 无法满足某些定位应用需求。基于指纹定位的方案能够在基于小区粗定位方案基础上大幅度提升定位精度、节约计算成本、增强适用性, 成为定位研究的热点。针对室外指纹定位的业务需求, 深入研究分析了两种基于机器学习的栅格化和非栅格化室外指纹定位方案。通过参数加权、数据拟合等方法对于大规模指纹数据进行了清洗, 提高数据源的有效性。通过划定研究区域、栅格化、构建指纹数据库、训练模型、修正模型、非栅格化、粗定位耦合、匹配参数、训练参数等子模块的实现, 分析和优化了算法的运行效率和定位精度, 确定了影响算法性能的关键指标。进而结合仿真结果, 分析了两种基于指纹的定位方案的性能。最后介绍了基于机器学习的指纹定位方案在实际应用中的典型场景。

关键词: 指纹定位; 移动蜂窝网络; 机器学习; 栅格化; 典型应用场景

中图分类号: TN913

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2021201

Outdoor location scheme with fingerprinting based on machine learning of mobile cellular network

ZHOU Zhichao, FENG Yi, XIA Xiaohan, FENG Yuyao, CAI Chao, QIU Jiahui, YANG Lihui, WU Yunxiao

Center of Smart Network of China United Network Communication Co., Ltd., Beijing 100048, China

Abstract: The positioning scheme based on mobile cellular network technology is one of the important technical approaches to provide network optimization, emergency rescue, police patrol and location services. The traditional positioning scheme based on cell base station location information has low positioning accuracy and large positioning error, so it cannot meet the requirements of some positioning applications. The scheme based on fingerprint location can greatly improve the location accuracy, save computational cost and enhance the usability based on the coarse location scheme of the cell and become the hotspot of the research. Rasterization and non-rasterization of outdoor fingerprint location scheme based on machine learning were studied and analyzed to meet the business requirements of outdoor fingerprint location. By means of parameter weighting, data fitting and other methods, large-scale fingerprint data were cleaned to improve the effectiveness of data sources. Through the realization of sub-modules such as demarcating research area, rasterizing, constructing fingerprint database, training model, correcting model, non-rasterizing, rough positioning coupling, matching parameter and training parameter, the operation efficiency and



positioning accuracy of the algorithm were analyzed and optimized, and the key indexes affecting the algorithm performance were determined. Then, the performance of two fingerprint-based localization schemes was analyzed based on the simulation results. Finally, the typical scenarios of the fingerprint location scheme based on machine learning in practical application were presented.

Key words: fingerprint positioning, mobile cellular network, machine learning, grid, typical application scenario

1 引言

随着 5G 移动通信技术的商用,基于无线通信的位置服务(location based service, LBS)越来越受关注。运营商根据投诉用户的位置,可以准确地查找到相应基站的覆盖情况,从而有针对性地优化网络性能,并且能够为特定用户提供更加精准的服务推送^[1]。在基础设施遭到破坏的灾区,利用定位能力可以实施及时、精准的救援。

基于全球卫星导航系统(global navigation satellite system, GNSS)的定位精度较高,但是在遮挡严重的区域信号较弱,无法实现定位测算。因此,基于覆盖效果好的移动蜂窝网络的定位方案被不断深入研究,作为重要的 GNSS 补充定位方案。目前,基于无线移动通信网络的室外定位技术主要分为两类,分别为基于测距模型的定位技术与基于位置指纹的定位技术^[2]。测距定位技术包括基于信号到达时间(time of arrival, TOA)、基于到达时间差(time difference of arrival, TDOA)、基于到达角度(angle of arrival, AOA)和基于信号离开角度(angle of departure, AOD)的三角定位方案以及基于 cellID 的定位方案。TOA 和 TDOA 需要已知至少 3 个确定的位置信息,才能够计算待测用户的位置,而且对时间同步要求高。基于 AOA 的定位方案对信号角度的要求较高,需在接收机配置较大的天线阵列,极大地增加了硬件成本^[3]。阵列中天线数量越大计算复杂度就越高。并且,这类基于移动通信信号测距的定位方案,定位精度稍差,无法满足特定场景的定位精度需求。

基于指纹的定位技术不依赖信号到达时间和经验公式推算,通过终端上报的最小化路测

(minimized driving test, MDT)信息训练指纹定位模型,通过训练模型预测待测终端的位置。此方案相比于信号和公式推算的方案具有更高的定位精度和可实施性^[4]。指纹定位方案完全复用移动蜂窝网络通信过程中的信号强度和终端接入小区属性特征,对网络和终端设备没有特定的额外需求,相比于基于测距的方案,节省硬件成本且易于推广应用。基于指纹的定位方案可以根据不同的定位需求,差异化地划分指纹栅格尺寸,能够动态地适配不同需求场景的定位精度,因此,基于指纹信息的定位方案成为研究热点。随着机器学习算法的不断深入优化,模型训练阶段的耗时也不断得到优化,指纹定位与机器学习算法结合,以期降低位置测算的复杂度。基于机器学习的指纹定位方案分为模型训练阶段和位置预测阶段。其中阶段一是离线的,只在阶段二对终端位置预测时,是在线的过程。所以,此类定位方案的预测运算复杂度低、定位实时性高。基于移动蜂窝网络的机器学习指纹定位方案具有精度高、成本低、实时性强、与当前网络兼容性好的特点,因此成为当前的研究热点。

本文提出两种基于机器学习的指纹定位方案,通过利用机器学习框架极大地简化了定位过程,并结合贝叶斯算法以及欧几里得算法准则,根据网络侧大量用户上报的终端场强数据,通过训练成熟的模型可以准确、快速地获取用户的位置信息,不仅能够满足灾区探测、紧急救援、公安巡警等场景较高精度定位的需求,还可以为网络优化和管理提供更具有针对性的指导建议。并且在保证精度的前提下能够降低实现成本,提高了算法的鲁棒性^[5]。

2 室外定位业务需求

我国位置服务的市场规模增长迅速，截至2019年已经达到万亿元量级。目前，国内市场日定位请求次数达到千亿，全球定位服务及设备市场规模已达到万亿级别，定位服务已成为各类线上、线下应用的底层能力，预期仍将快速发展^[6]。

随着社会的进步和科技的发展，室外定位在技术手段、定位精度、可用性等方面均取得了质的飞跃，并且逐步从航天、交通、水利、测绘、军事、农业、自然灾害预防等领域渗透到社会生活的方方面面，成为人们日常中不可或缺的重要应用，比如人员搜寻、位置查找、交通管理、车辆导航与路线规划等。用户期望得到的定位服务功能如图1所示。

室外指纹定位技术对三打一查、医疗救援、消防安全及物流运输等行业做出了很大贡献。仅110全国接警量约为3.4亿次/年，潜在市场规模约3 400万元/年，并且在第一批推进20个较大城市110报警定位试点中，预期短时间内实现收入约220万元/年，需求规模约9 000万次/年^[7]。北京红十字会999紧急救援中心曾迫切提出，要求在车辆熄火并无主动打开终端定位功能时，仍能迅速找到救援点；儿童和老人走丢、走失事件时有发生，因此亟须精准主动定位技术；对于110、120等，移动通信技术的定位方案显得尤为突出，根据通信终端的信号信息可以快速地锁定关注对象的位置，从而可以快速制定合理的精准施救方案。典型行业的室外指纹定位应用需求见表1。

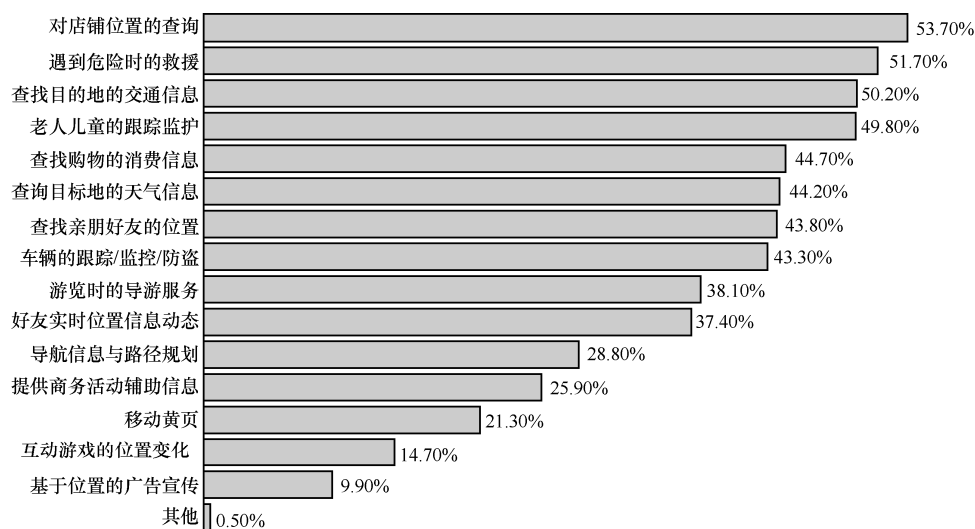


图1 用户期望得到的定位服务功能

表1 室外指纹定位不同应用场景下需求指标

应用分类	应用名称	市场容量	定位精度要求	时间要求
重点行业	三打一查	>3.5 亿次	30~50 m	<30 s
	医院急救	>2 亿次	5~10 m	<60 s
	紧急救援	2020 年规模超过 3 800 万元	5~10 m	<2 min
其他行业	物流运输	全国规模超 3 000 万元	2 cm	<1 min
	老人儿童关爱定位	保守估计约 1 050 万元	5~10 m	<2 min



3 基于机器学习的指纹定位关键技术

基于指纹的定位方案是在基于小区 cellID 粗定位基础上对定位进行精细化的结果,在基于小区 cellID 的定位方案中,终端的接入小区基站的位置近似为终端的位置,但是 LTE 中基站的覆盖范围在几百米,所以以基站的位置作为终端位置的替代方案,无法满足部分需求的定位精度要求。因此,在小区 cellID 粗定位的基础上,通过指纹定位的方案获取终端位置较为准确的结果。本节将对基于指纹的 3 种定位算法方案进行详细说明。

机器学习由来已久,但是因为机器性能不能实现大数据量的处理,所以一直没能得到充分的发展和应用。近年来随着机器计算和处理性能的大幅提升,机器学习成为模型化处理大数据的主要方案。本文在深入研究机器学习算法框架的基础上,将模型训练的优势与指纹定位场景相结合,为指纹定位方案提供强有力的算法支撑。

3.1 基于栅格的指纹定位方案

本节对基于地图栅格的指纹定位方案进行详细的分析和说明。在该定位方案中分为区域划定、数据清洗、区域栅格化处理、指纹数据库构建、分类模型训练、对测试集数据进行预测、模型参数优化 7 个功能子模块,各个子模块之间紧密联系,对应的逻辑关系如图 2 所示。

(1) 区域划定

区域划定是基于栅格的指纹定位的前提,划定的区域一般以 1~2 km 为边长的方形区域,因为基站覆盖的范围是不规则的扇形区域,所以划定的多个方形研究区域会存在边缘重叠的现象,即方形区域 A 与相邻的方形区域 B 之间会有边缘的重叠部分。

(2) 数据清洗

训练模型的准确性依赖于数据质量,所以数据的清洗和分析是至关重要的子模块。对数据的清洗分为 3 部分:源数据应有明确的经度、纬度

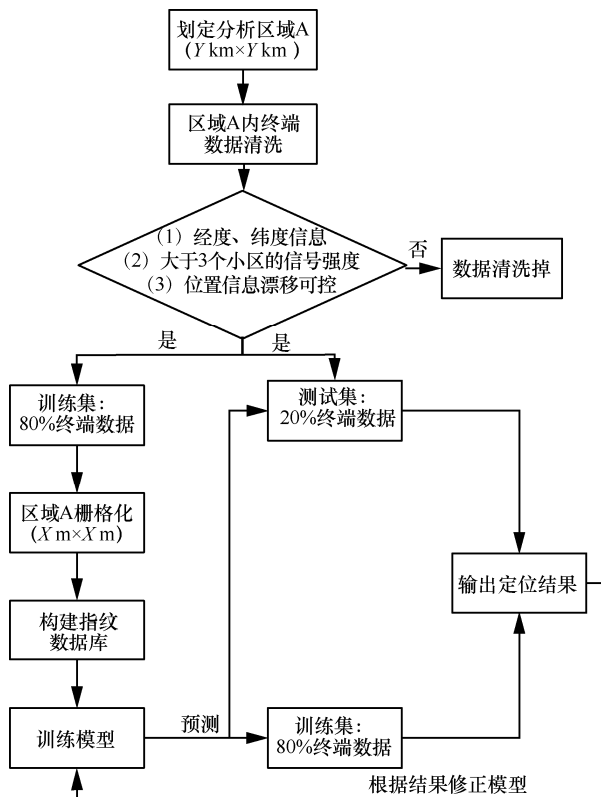


图 2 基于栅格的定位方案流程

信息;每条源数据应有大于 3 个小区的信号强度信息;每条源数据的位置信息漂移可控。上述为 3 个必要条件。经过数据处理得到满足上述条件的源数据后,将其进行随机分类,80%数据源用来训练模型,20%数据源用来测试模型。

(3) 区域栅格化处理

根据划定区域的相对距离,将其转化为全球绝对的经度、纬度信息,从而将划定的区域按 $X\text{ m} \times X\text{ m}$ 的规格进行栅格的划分,根据方形区域的面积把整个区域均匀地划分为均大小栅格,并对栅格进行编号。将转化的全球绝对经度、纬度信息与栅格的编号进行对应,从而使每个栅格都有独一无二的经度、纬度属性和编号属性。

(4) 指纹数据库构建

根据训练集终端数据的经度、纬度信息将其与栅格的经度、纬度信息对应,在栅格的编号属性与经度、纬度属性对应的基础上,将对应的源数据点按位置信息映射到相应的栅格内。形成指

纹数据库, 后续将根据该数据库的对应关系进行模型的训练。终端与栅格的映射关系如图 3 所示。

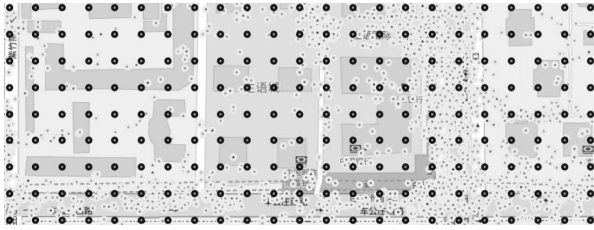


图 3 栅格与终端点的对应映射关系

(5) 分类模型训练

根据步骤(4)中构建的指纹数据库, 数据库中每条数据的属性如 $\langle \text{cellID}, \text{RSRP1}, \text{RSRP2}, \dots, \text{gridID}, \text{feature1}, \text{feature2} \dots \rangle$, 其中 cellID 为接收服务小区的 ID, RSRP 表示该终端接入小区的信号强度, gridID 表示为栅格的编号, feature1 和 feature2 表示其他类型的属性特征。通过调用机器学习算法模型框架实现模型的训练, 在本文中采用机器学习自回归算法, 采用 KNN (K -nearest neighbor) 和 WKNN (weight- K -nearest neighbor) 训练原则^[8-9], 可分别表示如下。

KNN:

$$\begin{aligned} \min_{N, \text{gridID}} & \sqrt{\sum_{i=1}^N (\text{RSRP}_i - \text{RSRP}_{\text{gridID}})^2} \\ \text{s.t. } & 3 \leq N, 0 \leq \text{gridID} \leq \max(\text{gridID}) \end{aligned} \quad (1)$$

WKNN:

$$\begin{aligned} \min_{w, N, \text{gridID}} & \sqrt{\sum_{i=1}^N w_i (\text{RSRP}_i - \text{RSRP}_{\text{gridID}})^2} \\ \text{s.t. } & 3 \leq N, 0 \leq \text{gridID} \leq \max(\text{gridID}), \\ & w_1 > w_2 > \dots > w_N \end{aligned} \quad (2)$$

其中, N 表示同一个终端接收到信号的小区个数, $\text{RSRP}_{\text{gridID}}$ 表示栅格编号为 gridID 对应的信号强度, w_i 表示第 i 个信号强度的加权值。通过上述优化问题的求解, 获得的最优解即模型预测的结果, 并输出。

(6) 对测试集数据进行预测

应用训练的预测模型对 20% 的测试数据进行

终端位置预测, 然后将终端所落在的栅格 gridID 转换为地理经度、纬度信息, 因为测试集中每个终端的信息中包括信号强度和经度、纬度位置信息, 所以可以将预测的结果与实际的终端位置信息进行比较, 得到模型预测准确度的量化结果。

(7) 模型参数优化

根据对终端位置预测的位置结果, 闭环修正训练模型, 调整模型的训练参数, 从而强化模型的预测能力。因为修订的模型与预测的结果是闭环反馈的关系, 所以在该闭环中能够有效地完成模型的校正, 提升模型的预测准确度。

3.2 基于非栅格的指纹定位方案

非栅格化方案的模型训练也需要目标区域终端的位置与信号强度信息, 并通过这些信息构建位置-信号强度的关系。通过对于位置-信号强度库的直接训练来形成指纹库, 并通过此模型直接预测输出待定位目标的实际位置。与栅格化方案不同之处在于非栅格的方案中不需要对区域进行栅格化处理, 也不需要栅格与终端经度、纬度的对应转换关系。

非栅格方案数据集与 3.1 节中基于栅格的指纹定位方案的数据集是一致的。

非栅格的指纹库生成: 创建一个完整的 RSRP 与经度、纬度对应的指纹库。在实际的环境中, 特定位置的 RSRP 指纹会不断受到周围环境变化的影响。为了保持定位精度, 非栅格指纹库要以更高的频率维护与更新, 以保护模型的鲁棒性和精确度。

非栅格指纹数据库有多个不规则点 n 。指纹库的坐标标记为:

$$I = \{I_i\}_{i=1}^n \quad (3)$$

其中, 第 i 个训练点的签名可记为:

$$R_i = \{r_i, P_i\} \quad (4)$$

在每个训练点的签名中, $r_i = [r_{i,1}, r_{i,2}, r_{i,3}, \dots, r_{i,k}]^T$ 代表终端获得的主服务小区和相邻小区 ID 的一维向量, $P_i = [p_{i,1}, p_{i,2}, p_{i,3}, \dots, p_{i,k}]^T$ 训练点在每个



小区的 RSRP 分量。由于训练点可能相互叠加, P_i 设定为规定时间的均值。由于终端的初始位置随意, 以及各种衰落和切换等原因造成小区信息缺失。训练点的签名集合表示为:

$$\{R_{i,c}\}_{i=1}^n = \{r_{i,c}, p_{i,c}\}_{i=1}^n \quad (5)$$

其中, c 表示小区信息, 用于提高指纹定位精度。不考虑 c 时指纹库则完全由 RSRP 分量训练。

签名集合会和各个点的位置联合, 然后采用卡尔曼滤波进行清洗。建立的联合签名集合为:

$$R_i^L = \{r_i, P_i, l_i\} \quad (6)$$

本文采用签名相似性比较作为位置估计, 选择对比了 K 近邻法和加权 K 近邻算法。大致描述如下: KNN 的思想为通过计算测试签名和所有训练签名的相似度距离, 获得相似度距离最小的 k 个训练签名, 并以 k 个训练签名的坐标平均值作为估计位置。相似度度量采用欧氏距离和曼哈顿距离。加权 K 近邻算法则是考虑到 KNN 算法签名的可靠性和优先级, 增加了 KNN 的分量权重。

将所有训练签名记为 $R_i = \{r_i, P_i\}$, 测试点签名记为 $R^E = \{r^E, p^E\}$, 取最近邻点 k 个, 则 KNN 的相似度距离对比为:

$$D(R^E, R_i^E) = \|d(P^E) - d(P_i)\|^2 \quad (7)$$

获得的 k 个距离记为 $a_k = \{R_1, R_2, \dots, R_k\}$, 对应的位置则为 $L = \{L_1, L_2, \dots, L_k\}$, 于是测试估计位置为:

$$L^E = \frac{1}{k} \sum_{i=1}^k L_i \quad (8)$$

考虑到小区 RSRP 权重不同, 在 WKNN 算法中, 本文在指纹库中加入权重, 设定加权系数为 $w = f(P_1, P_2, \dots, P_c)$ 。同时在位置估算中为了减少数值影响, 将距离加权系数归一化, 由此, 计算 WKNN 的测试估算距离为:

$$L^E = \sum_{i=1}^k w_i^c L_i \quad (9)$$

3.3 基于栅格和非栅格的指纹定位方案对比

基于栅格的指纹定位方案在模型训练前要对划定区域的栅格进行划分, 将位置的相对距离信息转化为地球表面绝对的经度、纬度信息, 在该转换过程中会有精度的折损。另外, 在用模型预测终端过程中, 是将终端所在栅格的编号信息返回来, 然后根据编号再转换成经度、纬度信息, 在此转换中也折损了预测精度, 所以基于栅格的指纹定位方案比基于非栅格的方案在精度上有部分的精度折损。但是以栅格为依托标准, 可以减少终端位置漂移带来的影响, 在非栅格的指纹定位方案中, 可能会存在较大的位置预测偏移。

比较上述两种方案的运算复杂度, 在方案一中基于栅格的指纹定位方案, 需要将区域栅格化处理, 另外在划定的区域内划分完栅格后, 需要构建指纹数据库, 即将终端的信号强度、位置信息与栅格的属性对应的过程。如果源数据量比较大, 构建指纹数据库的过程需要的算力较高。在方案二基于非栅格的指纹定位方案中, 模型训练的数据源就是终端的信号强度, 模型的输出即是终端的位置经度、纬度信息, 整个处理过程单纯、直接。所以, 方案二的运算复杂度远小于方案一的运算复杂度。从大规模可实现角度分析, 仅以计算复杂度为判断标准, 方案二更具有大规模实施的可能性。但是基于指纹的定位方案和基于 cellID 定位的方案相比, 在实现上会对移动网络网的网元功能有新的需求定义^[10-11], 在实施过程中, 根据实际场景的需要调整网络组成。两种指纹定位方案见表 2。

表 2 方案一和方案二指纹定位方案的比较

指纹定位方案	计算复杂度	性能	鲁棒性	实施性
方案一	较高	有折损	较好	对服务器集群要求较高
方案二	较低	无折损	稍差	对服务器集群要求较低

3.4 用 KNN 方法进行定位结果选取的分析

基于机器学习框架的指纹定位以项目实现为主,理论分析文献较少。而不同公司使用的机器学习框架多不相同,原始的机器学习算法应用到指纹定位场景中,都需要一系列的参数调优,寻找到最优的模型参数集后,才能取得较好的定位效果。

但是模型调参的过程没有明确的行业标准,通过不断地尝试直到输出结果满足需求为止。在仿真实验中,尝试了 XGBoost 方案和深度学习方案,前者需要不断地调整模型参数,但是一直没有取得超过 KNN 算法的模型。深度学习方法存在复杂度和结果精度之间的折中,当神经网络层数少时,性能较差;当通过增加层数提高定位精度时,复杂度会极速上升,不能保证结果的实时输出。

基于上述原因,本文仿真中取 KNN 算法作为定位预测的标准,采用定位精度和运算复杂度的折中方案。

4 实验结果与分析

对方案一和方案二指纹定位算法进行仿真比对,仿真参数说明见表 3。

表 3 仿真实验参数

仿真参数名称	参数值
划定区域 Y	1.5 km
统计小区个数	5 个
栅格尺寸 X	50 m
终端源数据量	1 048 574 条
训练集占比	80%
测试集占比	20%
机器学习框架	sklearn.KNeighborsRegressor、sklearn.KNeighborsClassifier
学习和预测规则	KNN、WKNN
终端接收的 RSRP 个数	3 个
经度、纬度精度	$1.0 \times 10^{-12}^\circ$
相对距离精度	1 m
方案一	基于栅格划分的指纹定位算法
方案二	基于非栅格的指纹定位算法

实验中使用的数据集来自中国联通数据中台累积的实时手机终端上报的 MR/MDT 数据信息,因此,所有应用数据都是真实且随机分布在道路、公园等室外真实场景中的,所以数据具有普适性。

方案一基于栅格的指纹定位算法的模型在训练集上的预测结果如图 4 和图 5 所示。图 4 为随机取 2 000 个数据源用于测试模型的预测准确度,其中方框点表示训练模型对随机手机终端预测的位置结果,“×”点表示被预测终端实际的位置,由图 4 可知模型的预测结果与实际结果相合度是比较高的,结果的符合度高达 $95.75\% < 1\,916, 2\,000 >$ 。图 5 为预测值与实际值的 CDF 曲线,由该曲线也可以看出训练模型的预测结果在训练集上的准确度是比较高的。

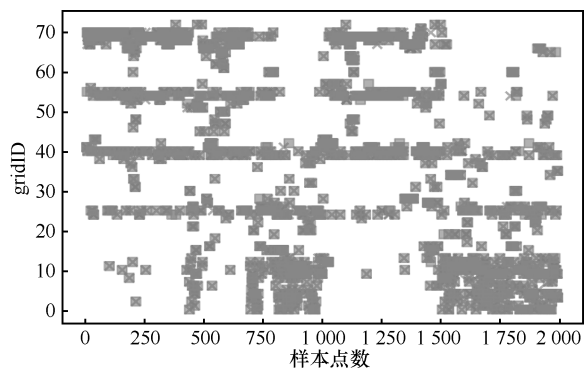


图 4 方案一模型在训练集上的预测与实际结果比对

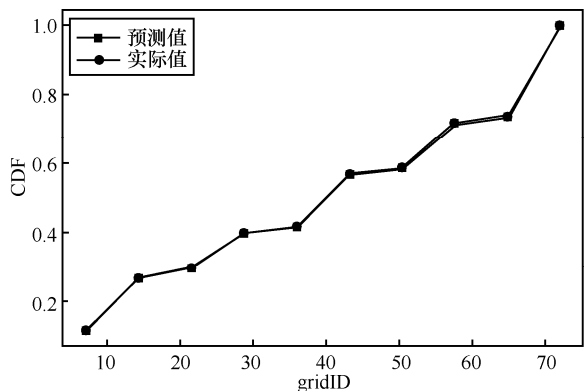


图 5 方案一模型在训练集上的预测与实际结果 CDF 曲线

方案一基于栅格的指纹定位算法模型在测试集上的预测结果如图 6 和图 7 所示。图 6 为随机取 600 个数据源用于测试模型的预测准确度,其



中方框点为训练模型预测的结果,“×”点是终端实际的位置所在的栅格 gridID,由图 6 可知模型的预测结果与实际结果相合度在数据点密度比较大的位置准确度是比较高的,但是在数据源本身比较散的情况下,预测的结果比较差。因为模型的训练依赖于已知数据源点数,在只有 1 个或者 2 个散点的情况下,是无法训练成熟的预测模型的,此结论从图 6 中可以明显地显示。

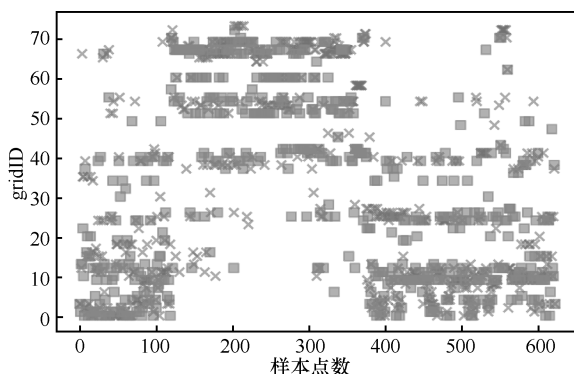


图 6 方案一模型在测试集上的预测与实际结果比对

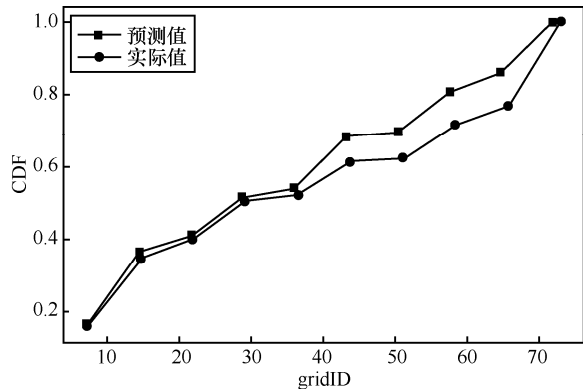


图 7 方案一模型在测试集上的预测与实际结果 CDF 曲线

图 7 为方案一的训练模型在同一个 600 点的测试集上的预测结果与实际终端位置的 CDF 曲线。从趋势上来看,预测值和实际值是有比较一致的趋势的,预测结果的符合度可高于 51%。因为数据源的获取是在高楼林立的城区,据统计终端在室内与在室外的比例为 8:2,所以,测试数据大部分在室内的环境,所以预测结果准确度与空旷区域的数据源相比有较大的下降。另外,由于实际终端的分布规律无法把握,所以源数据本身

也存在位置漂移的现象,此现象将极大地弱化位置预测的准确度。划定一块空旷的区域并获得高质量的数据源是比较困难的,但是为了验证预测模型的准确性,后续将努力划定一块空旷的区域,并获取在此区域的通信源数据,进一步验证、提升基于机器学习的模型定位精度。

非栅格模式下考虑粗定位和不考虑粗定位的预测结果分别如图 8 和图 9 所示,从两图中可以看出,在考虑到利用 cellID 的粗定位结果之后,图 8 出现了很好的聚集性,对应左上角的点簇。而不考虑粗定位情况下,如图 9 所示,离散点很容易出现误判的情形,甚至预测误差比较大。



图 8 考虑粗定位耦合的非栅格 KNN 方案预测结果

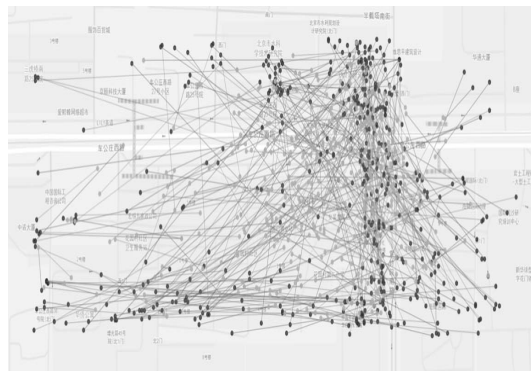


图 9 不考虑粗定位耦合的非栅格 KNN 方案预测结果

考虑粗定位耦合情况下,对 WKNN 的归一化参数进行了调整,即在不同的权值情况下的位置预测准确性,结果如图 10 所示。 w 是邻小区 RSRP 相对主小区 RSRP 的权重。在实际测量中,邻小区的 RSRP 值比主服务小区的 RSRP 值低 5~6 dB,

根据主服务小区与相邻小区内信号强度 RSRP 的关系, 设置与之对应的加权系数 w , 实验结果表明当 w 在 0.2~0.3 时, 定位的精度结果最优。

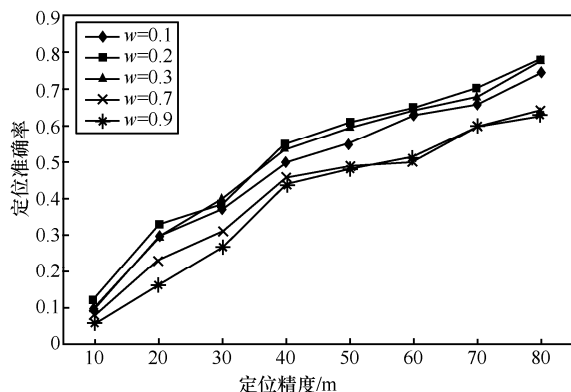


图 10 考虑粗定位耦合的 WKNN 算法不同权重预测准确率结果

通过上述两种方案的仿真分析可以得出如下结论: 基于非栅格的指纹定位方案比基于栅格化的指纹定位方案的运算复杂度要小, 方案二在实施中对服务器集群的要求远低于方案一基于栅格的指纹定位方案。在相同数据集的前提下, 方案一和方案二的模型预测准确度相当。但是方案二相较于方案一, 可能存在预测位置漂移的现象, 鲁棒性低于方案一基于栅格的指纹定位方案。在实际的应用过程中, 依据不同场景的实际需求可以选用不同的方案算法。

上述两种方案的性能比较见表 4, 在数据集数据有限的情况下, 方案能够达到 57%@50 m 的定位精度, 方案二的定位精度为 46%@50 m, 结合图 10 可知, 方案二在终端密度较高的区域性能会有较大的改善, 在 50 m 时性能出现明显的转折。因为在方案一中有 50 m 指纹栅格的框定, 当预测终端在某个栅格内时, 位置结果基本不会有太大的漂移, 而方案二中预测的结果是针对单个终端的位置预测, 会存在位置的漂移。当定位的精度目标定得较大时, 位置的漂移将被包含在预测结果内, 所以大于 50 m 标准时, 方案二表现出较大的优势, 而方案一仍然是 50 m 的栅格折损, 性能提升不如方案二明显。

表 4 方案一和方案二指纹定位方案的性能比对

指纹定位方案	定位准确性	性能折损性	鲁棒性	适合场景
方案一	57%@50 m	50 m	较好	终端密度小的场景
方案二	46%@50 m	0 m	稍差	终端密度大的场景

方案一中将栅格的中心位置作为此终端的位置预测, 精度折损在 50 m, 当终端密度较大时, 一个栅格中可能有多个终端分布的情况, 此时多个终端的预测位置将是一个值, 所以方案一更适用于终端密度较小的场景。另外, 方案一有栅格的框定, 在栅格预测正确的情况下最大位置漂移不会超过栅格折损值, 所以鲁棒性更强。方案二预测的目标就是每个终端, 所以在终端密度较大的场景中更加适用, 当位置漂移较大的情况出现时, 预测的鲁棒性较低。

MLP 神经网络算法、分类预测算法、回归算法和 SVM 算法在测试集上对待测点位置的预测结果如图 11 所示, 图 11 中显示利用回归算法模型和分类算法模型得到的结果是最优的, 而且相差较小, 分类的方案略优于回归的方法。本文重点采用的是分类的算法方案, SVM 算法的预测准确率相比于前两者较差, 而且预测正确率堆积在 20% 以下, 因此 SVM 方案不适合本文的定位场景预测。在 MLP 方案中, 第一隐含层神经元数量设定为 100, 第二隐含层设定神经元数量为 50, 用此方案训练模型后对测试数据进行结果预测时出现两点震荡的现象, 调整参数后震荡的位置有所变化, 但是仍然得不到预期的结果, 所以 MLP 方案也不能很好地适应本文的应用场景。

5 基于指纹定位的典型应用场景

根据被定位的终端设备在定位过程中是否主动寻求定位, 把定位方案分为主动定位和被动定位方案。其中, 如果设备主动请求自己的位置信

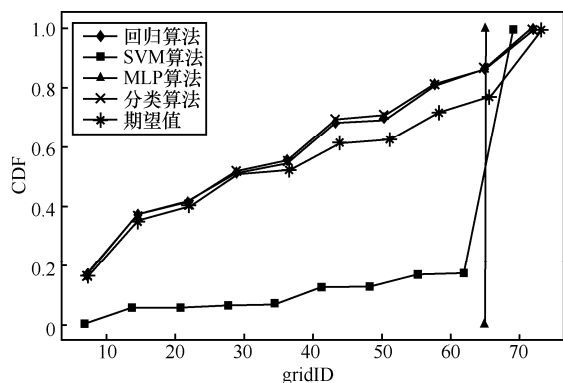


图 11 文献中经典的机器学习算法测试结果的 CDF 曲线

息并上报，称之为主动定位，比如 GPS 定位；如果设备提前约定或者在规则认可的情况下，不经过设备请求操作而被直接定位，称之为被动定位。

主动定位适用于商业场景，比如地图、导航、社交、游戏等；但有些场景因为直接影响人身及财产安全，并且不具备主动定位条件，这种场景需要被动定位的支持，比如接警中心对报警人的定位、地震等自然灾害中对受灾群众的定位、公安机关在抓捕行动中对嫌疑人的定位等。

(1) 应用场景：110、122、12122 等接警中心

目前，一般由报警人自行提供，这个过程中存在多种问题：警情发生时，一般情况比较紧急，报警人提供位置信息耽误宝贵时间；警情发生时，报警人处于紧张状态，可能无法清晰描述自己的位置；在陌生环境下，报警人不知道自己的位置；有报假警的情况存在；突发状况中断报警电话，导致报警人处于危险之中。

通过指纹定位系统，可以与接警中心联动，在受理接警电话的过程中，对报警人进行定位。根据警情，接警的工作人员可以通过问答等方式，快速对报警人定位，或者直接出警。

(2) 应用场景：嫌疑人的精准锁定及抓捕

公安机关对嫌疑人进行抓捕的过程中，需要获知嫌疑人的位置，由于嫌疑人不可能主动上报位置，所以这个过程只能通过被动定位来实现。指纹定位技术能够相对精确地提供嫌疑人的位置区域，方便警方监视、布控、抓捕。

(3) 其他特殊场景

在地震等自然灾害中，可以通过指纹定位技术，反向查找受困群众；在矿井等地下作业场景中，由于没有 GPS 信号，通过指纹定位技术定位井下作业工人位置，提高生产安全水平；对社区矫正人员进行实时监控等。

6 结束语

本文提出了两种基于机器学习算法的指纹定位方案，并从理论和仿真角度比较了两种定位方案的复杂性和定位精确性。通过深入分析基于移动通信网络的机器学习室外指纹定位方案，优化了指纹定位算法方案的运算效率和定位精度。通过仿真结果分析，比较了两种基于机器学习的室外指纹定位算法方案的性能。目前采用的数据集为城区高楼环境，包括大量的室内用户终端，所以数据源本身的位置信息存在较严重的漂移现象，未来的工作将集中在对源数据的修复处理，提高数据源的质量，并通过大量的数据训练、优化机器学习的算法模型，提高指纹定位的精度。

参考文献:

- [1] 裴登科. 基于测量报告的 LTE 终端室外定位算法研究[D]. 合肥: 中国科学技术大学, 2019.
PEI D K. Research on outdoor localization algorithms of LTE terminal based on measurement report[D]. Hefei: University of Science and Technology of China, 2019.
- [2] 倪磊. 基于 LTE 信令数据的指纹定位方法研究[D]. 大连: 大连理工大学, 2018.
NI L. A fingerprint localization algorithm based on LTE signaling data[D]. Dalian: Dalian University of Technology, 2018.
- [3] 韩亚楠. LTE 中 RF 指纹定位算法研究[D]. 北京: 交通大学, 2016.
HAN Y N. Research on RF fingerprinting in LTE network[D]. Beijing: Beijing Jiaotong University, 2016.
- [4] 张玄黎, 杨东凯, 王延昭, 等. 基于机器学习的室内高精度指纹定位方法研究[C]//第十二届全国信号和智能信息处理与应用学术会议论文集. 杭州: 出版者不详, 2018: 518-524.
ZHANG X L, YANG D K, WANG Y Z, et al. Research on indoor high-precision fingerprint localization method based on machine learning[C]//Proceedings of the 12th National Conference on Signal and Intelligent Information Processing and Applications. Hangzhou:s.n., 2018: 518-524.

- [5] 卞智. 基于机器学习算法的指纹匹配定位技术研究[D]. 北京: 北京邮电大学, 2017.
BIAN Z. Research on fingerprinting positioning technology-based on machine learning[D]. Beijing: Beijing University of Posts and Telecommunications, 2017.
- [6] 中评协课题组. 2016 年度我国主板市场资产评估情况统计分析报告(上)[J]. 中国资产评估, 2018(2): 21-36, 1.
CAS Research Group. Statistical analysis report of asset appraisal in main board market of China, 2016(part one)[J]. Appraisal Journal of China, 2018(2): 21-36, 1.
- [7] 刘艳亮, 张海平, 徐彦田, 等. 全球卫星导航系统的现状与进展[J]. 导航定位学报, 2019, 7(1): 18-21, 27.
LIU Y L, ZHANG H P, XU Y T, et al. Development status and trend of global navigation satellite system[J]. Journal of Navigation and Positioning, 2019, 7(1): 18-21, 27.
- [8] 党小超, 马平川, 郝占军. 基于 CSI 的改进 KNN 室内定位方法[J]. 传感器与微系统, 2019, 38(10): 51-53.
DANG X C, MA P C, HAO Z J. Improved KNN indoor positioning method based on CSI[J]. Transducer and Microsystem Technologies, 2019, 38(10): 51-53.
- [9] 李连亮. 基于 LANDMARK 系统的 KNN 室内定位改进算法[J]. 山东工业技术, 2015(1): 213, 229.
LI L L. KNN indoor positioning improvement algorithm based on LANDMARK system[J]. Shandong Industrial Technology, 2015(1): 213, 229.
- [10] 3GPP. Technical specification group radio access network: lte positioning protocol (LPP)[S]. 2020.
- [11] 3GPP. Technical specification group radio access network: evolved universal terrestrial radio access (E-UTRA): LTE Positioning protocol A (LPPa)[S]. 2020.

[作者简介]



周志超(1989-), 男, 中国联合网络通信集团有限公司智网创新中心工程师, 主要研究方向为高精度定位、蜂窝移动通信、C-V2X 等。



冯毅(1976-), 男, 中国联合网络通信集团有限公司智网创新中心高级工程师, 主要研究方向为移动通信系统网络规划、认知无线电、异构无线网络传输技术、5G 垂直应用等。



夏小涵(1989-), 男, 中国联合网络通信集团有限公司智网创新中心 C-V2X 研发工程师, 主要研究方向为高精度定位、RSU、车路协同等。

冯瑜瑶(1995-), 女, 中国联合网络通信集团有限公司智网创新中心通信工程师, 主要研究方向为无线移动通信、高精度定位、C-V2X 及 MEC 解决方案等。

蔡超(1984-), 男, 中国联合网络通信集团有限公司智网创新中心总监, 主要研究方向为无线移动通信、高精度定位、数据网、安全网关等。

邱佳慧(1985-), 女, 中国联合网络通信集团有限公司智网创新中心车联网技术总监, 主要研究方向为车联网、5G 通信、高精度定位等。

杨立辉(1986-), 男, 中国联合网络通信集团有限公司智网创新中心解决方案经理, 主要研究方向为无线移动通信、高精度定位、视频编解码及传输等。

乌云霄(1985-), 女, 中国联合网络通信集团有限公司智网创新中心高级工程师, 主要研究方向为移动通信、移动网络 OSS。