

基于机器学习的5G室内定位方法

5G Indoor Positioning Method Based on Machine Learning

张千坤¹, 陈任翔¹, 钟志刚¹, 周国栋² (1. 中讯邮电咨询设计院有限公司, 北京 100048; 2. 中讯邮电咨询设计院有限公司郑州分公司, 河南 郑州 450007)

Zhang Qiankun¹, Chen Renxiang¹, Zhong Zhigang¹, Zhou Guodong² (1. China Information Technology Designing & Consulting Institute Co., Ltd., Beijing 100048, China; 2. China Information Technology Designing & Consulting Institute Co., Ltd. Zhengzhou Branch, Zhengzhou 450007, China)

摘要:

提出了一种适用于5G数字化室内分布式系统的通用室内定位方法。该方法首先采集数字化室分天线与终端(UE)之间的信道状态信息(Channel State Information, CSI),其次在时域和频域对CSI进行预处理,得到有效的CSI幅值,然后为减少CSI的重复采样和计算复杂度,构建灰色预测模型GM(1.1)对采集到的CSI样本值进行预测,生成新的CSI序列,最后采用机器学习算法构建室内信号传播模型,即CSI与距离之间的非线性关系,并结合多个天线完成定位。

关键词:

5G; 机器学习; CSI; 数字分布; 路径损耗模型
doi:10.12045/j.issn.1007-3043.2022.07.010
文章编号:1007-3043(2022)07-0050-06
中图分类号:TN929.5
文献标识码:A
开放科学(资源服务)标识码(OSID):



Abstract:

It proposes a general indoor positioning method suitable for 5G digital indoor division. This method first collects the Channel State Information (CSI) between the digital indoor antenna and the terminal (UE), and then preprocesses the CSI in the time domain and frequency domain to obtain an effective CSI amplitude, and then in order to reduce repeated sampling and computational complexity of CSI, the gray prediction model GM (1.1) is constructed to predict the collected CSI sample values, and a new CSI sequence is generated. Finally, the indoor signal propagation model is constructed by using machine learning algorithm, which describes the non-linear relationship between CSI and distance, and combined with multiple antennas, the positioning is completed.

Keywords:

5G; Machine learning; CSI; Indoor digital system; Path loss model

引用格式:张千坤,陈任翔,钟志刚,等. 基于机器学习的5G室内定位方法[J]. 邮电设计技术,2022(7):50-55.

1 概述

根据定位环境,无线定位系统可分为2种:室内无线定位系统和室外无线定位系统,室外无线定位系统,主要采用的是GPS定位^[1-2],而室内定位主要采用的是无线网^[3-5]、蓝牙^[6]、超宽带(UWB)^[7]等技术。5G网络的大带宽、低时延、高可靠特性,为业务提供必要的网络基础,推动业务的不断丰富。业务应用场所有

80%以上在室内,而室内定位由于GPS信号弱无法满足室内业务位置服务的需求。随着业务的发展,室内定位的需求越来越强烈。由于室内环境复杂,如物体位置的随机移动、多路径散射、电磁干扰等,在室内进行精确的目标定位相比室外来说要困难得多。因此5G室内场景急需一种定位精度高、复杂度低,且对设备以及终端没有要求的定位方法。

本文首先将采集到的CSI^[8]通过逆傅里叶变换转换到频域,在频域进行滤波,消除多径的影响,然后再将滤波后的CSI转到频域,在频域进行加权求和处理,

收稿日期:2022-05-20

减少频率选择性衰落的影响,得到有效的CSI,同时,为得到更加准确的CSI序列,本文还采用灰色预测模型GM(1.1)^[9],对频域的CSI序列进行预测,构建新的CSI序列,减少了采集CSI的数量,降低了计算复杂度,然后用新的CSI序列带入路径损耗模型进行距离估计。室内定位的难点主要在于室内路径损耗模型的建立,传统的路径损耗模型主要是基于经验或者确定性方法^[10],经验性模型比如对数-正态模型,这些模型的参数都是基于特定的环境,在比较复杂的室内环境,效果比较差,确定性模型比如射线跟踪模型,一般来说可以较好地还原室内环境的传播特性,但是缺乏计算效率,同时需要室内环境的几何信息以及材质信息,一旦传播环境发生变化就需要重新计算。因此本文基于机器学习方法,构建新的路径损耗模型,在室内不同的采样点采集CSI并计算UE与天线之间的距离,并将这2种信息输入到决策树构建新的路径损耗模型,用于距离估计。最后结合多个天线并采用最小二乘法实现用户的准确定位(见图1)。

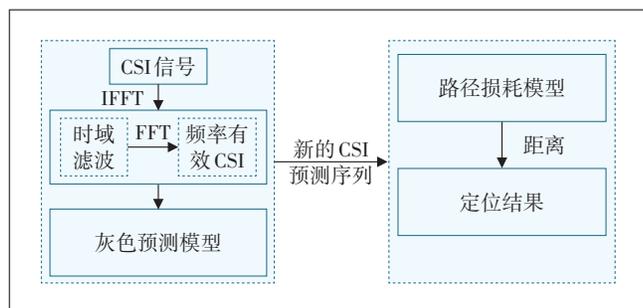


图1 定位架构

2 CSI预处理

2.1 CSI幅值信息预处理

本文采集到的原始CSI数据表征的是频域上的值,为信道频率响应(CFR),OFDM通信系统中不同的子载波的CSI可以表示为:

$$CSI[f_n] = \sum_{k=1}^D \alpha_k e^{-j2\pi(f_n + n\Delta f)\tau_k} \quad (1)$$

式中:

D ——信号的路径

α_k ——第 k 条路径的信号衰减

τ_k ——第 k 条路径的信号传播时间

f_n ——载波频率, $f_n = f_0 + n\Delta f$

Δf ——相邻2个子载波之间的频率差

CSI可以通过快速傅里叶逆变换(IFFT)变换到时

域,表示形式为信道脉冲响应(CIR):

$$h(t) = \sum_{i=1}^L a_i e^{-j\varphi_i \delta(t - \tau_i)} \quad (2)$$

式中:

φ_i 和 τ_i ——对应的相位和时延

如果可以得到时域信号的信道脉冲响应,便可以区分不同路径的信号。本文根据CSI的这种特性,对信号进行分离得到主径信号,提高距离估计的精度。

本文选择CIR最大幅值的50%作为阈值,将低于50%的部分滤除从而减少多径衰减的影响,然后利用快速傅里叶变换(FFT)将时域CSI转换到频域。滤波前和滤波后的CSI时域幅值如图2(a)和2(b)所示。

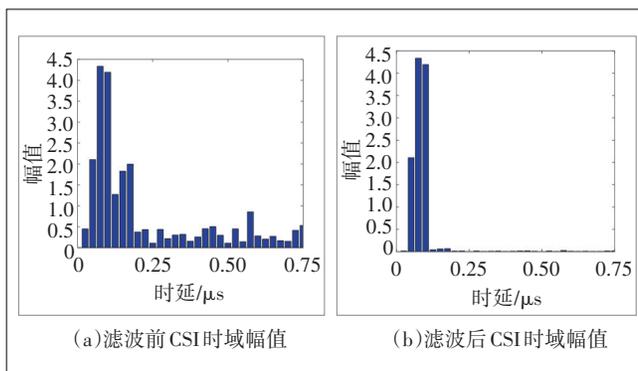


图2 滤波前和滤波后的CSI时域幅值

从图2可以看出,经过滤波以后,保留了直射径以及信号比较强的反射路径部分,信号比较弱的路径在时域受到了抑制。

将滤波后的时域CSI进行FFT变换得到的频域CSI幅值如图3所示。

从图3中可以看出,CSI在进行时域滤波以后幅值

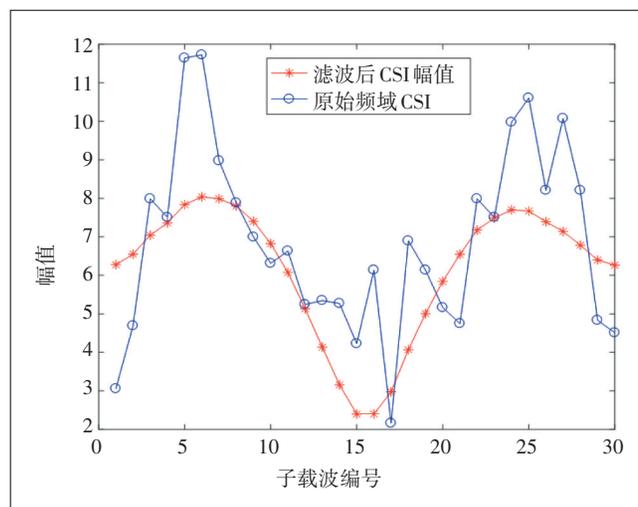


图3 原始频域CSI以及滤波后的CSI幅值

相比滤波前更加稳定,可以提高距离估计的精度。

2.1.2 频域幅值衰落补偿

由于在5G NR系统中,带宽最小为100 MHz,在市区特别是室内环境下,系统带宽大于相干带宽,采集到的CSI数据会受到频率选择性衰落的影响,如图4所示。

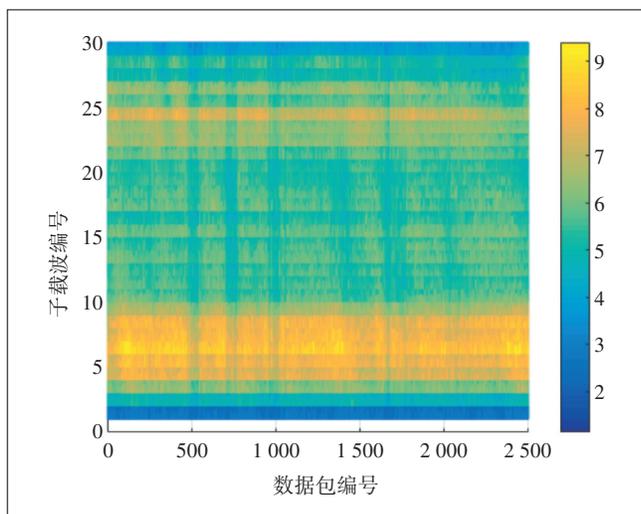


图4 CSI幅值信息

图4为在同一地点采集到CSI,从图4中可以看出不同子载波的幅值相差比较大,如果只选择最大的CSI幅值带入距离估计模型进行计算则会导致较大的距离估计误差。因此本文选择联合所有的子载波的幅值一起进行距离估计,本文将所有的子载波都赋予不同的权值,并对加权以后的所有子载波进行求均值计算得到一个有效的CSI,可以表示为:

$$|\text{CSI}_{\text{eff}}| = \frac{1}{K} \sum_{k=1}^K \frac{f_k}{f_0} \times |H_k| \quad (3)$$

最终得到的CSI可以表示为:

$$\text{CSI}_{\text{eff},i} = e^{-j\phi_i} \times |\text{CSI}_{\text{eff}}| \quad (4)$$

式中:

f_k ——第 k 个子载波的载波频率

K ——子载波的个数

f_0 ——信号的中心频率

$e^{-j\phi_i}$ ——第 i 个子载波的相位

$\text{CSI}_{\text{eff},i}$ ——第 i 个子载波的有效CSI

将图4中采集到的CSI的幅值按照式(4)进行处理得到的有效CSI幅值如图5所示,相比图4,图5的幅值跨度在1以内,要稳定的多。

2.2 灰度预测模型生成新的CSI序列

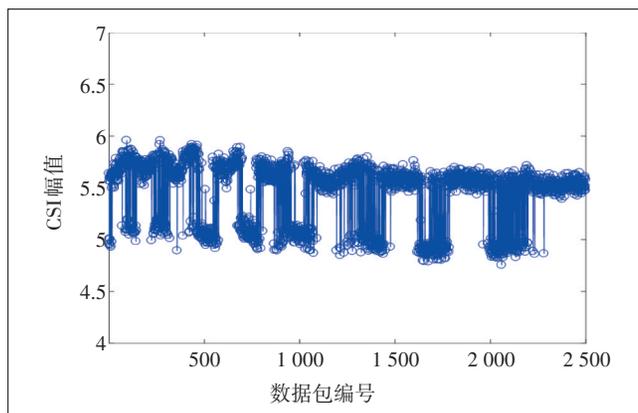


图5 有效CSI幅值

灰色系统理论是关于信息不完全或不确定的系统的控制理论,为了预测发展趋势,可以有效提高建模序列的平滑性。

灰色预测的基础是采用灰色预测模型,本文采用GM(1,1)模型,设置原始的有效CSI数据序列,表示为:

$$\begin{cases} x^{(0)}(1) = \text{CSI}_{\text{eff}}^i \\ x^{(0)}(2) = \text{CSI}_{\text{eff}}^{i+1} \\ x^{(0)}(3) = \text{CSI}_{\text{eff}}^{i+2} \\ \dots \\ x^{(0)}(n) = \text{CSI}_{\text{eff}}^{i+n-1} \end{cases} \quad (5)$$

其中 n 为有效CSI的数量。

灰色预测模型的过程如下,首先生成新的数据列:

$$x^{(1)} = [x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)] \quad (6)$$

其中

$$\begin{aligned} x^1(t) &= \sum_{k=1}^t x^{(0)}(k), t = 1, 2, \dots, n \\ x^1(t+1) &= \sum_{k=1}^{t+1} x^{(0)}(k), t = 1, 2, \dots, n \end{aligned}$$

然后建立一个一阶线性方程 $x^{(1)}$:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u \quad (7)$$

其中 a, u 为待解系数,分别称为发展系数和灰色作用量。

对累加生成数据做均值生成 B 与常数项向量:

$$B = \begin{Bmatrix} \frac{1}{2}[x^{(1)}(1) + x^{(1)}(2)] \\ \frac{1}{2}[x^{(1)}(2) + x^{(1)}(3)] \\ \dots \\ \frac{1}{2}[x^{(1)}(n-1) + x^{(1)}(n)] \end{Bmatrix} \quad (8)$$

$$Y_n = [x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)]^T \quad (9)$$

利用最小二乘法求解灰参数 \hat{a} ,

$$\hat{a} = \begin{pmatrix} a \\ u \end{pmatrix} = (B^T B)^{-1} Y_n \quad (10)$$

将灰参数 \hat{a} 带入式(7),得到

$$\hat{x}^{(1)}(t+1) = \left(x^{(0)}(1) - \frac{u}{a}\right)e^{-\alpha} + \frac{u}{a} \quad (11)$$

将上述结果累减还原,即可得到预测值

$$\hat{x}^{(0)}(t+1) = \hat{x}^{(1)}(t+1) - \hat{x}^{(1)}(t) \quad (12)$$

根据灰色预测模型的式(6)~式(12),可以得到预测的CSI值为:

$$CSI_{\text{pred},i} = \hat{x}^{(0)}(n) \quad (13)$$

得到预测CSI值以后,本文采用加权预测方法,即:

$$CSI_{\text{pred}(\text{weight})} = w_1 CSI_{\text{pred},i} + w_2 CSI_{\text{eff},i-1} \quad (14)$$

其中,这里 w_1 和 w_2 根据测试结果选取0.5时,精度最高。

3 基于机器学习的路径损耗模型及定位

3.1 路径损耗模型训练

基于机器学习的路径损耗模型主要包括以下几个步骤。

a) 在室内环境下的采集不同的采样点的CSI数据,按照CSI的预处理方法进行处理,由于是在室内环境下进行定位,因此本文选择天线距离地面的高度、处理后的CSI数据作为输入数据,终端和天线之间的水平距离作为输出数据。

b) 对这些输入输出数据进行归一化处理,具体如下:

$$x_N = \frac{2(x - x_{\min})}{x_{\max} - x_{\min}} - 1 \quad (15)$$

其中 x 表示的是需要归一化的特征值, x_N 表示的是归一化后的特征值。

c) 模型选择,本文选择随机森林对模型进行训练,如图6所示,在室内空间设置 N 个采样点,采集天线到终端的距离、天线距离地面的高度以及CSI作为样本值 D ,表示为:

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\} \quad (16)$$

式中:

x_i ——第 i 个采样点的 $CSI_{\text{pred}(\text{weight})}$ 和天线高度, $x_i = [x_i^{(1)}, x_i^{(2)}]$

y_i ——第 i 个采样点与天线的水平距离

采用有放回的方式每次随机抽取 m 个样本,抽取 T 次,则可以得到 T 个子集,记为 $D_i, i = 1, 2, \dots, T$ 。

在训练数据集所在的输入空间中,递归地将每个区域划分为2个子区域并决定每个子区域上的输出值,构建二叉决策树,具体如下:

选择最优切分变量 j 与切分点 s ,求解:

$$\min_{j,s} \left[\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right] \quad (17)$$

其中 $R_1(j,s) = \{x | x^{(j)} \leq s\}$, c_1 和 c_2 为划分后2个区域内固定的输出值;

遍历变量 j , 对固定的切分变量 j 扫描切分点 s , 选择使式(17)达到最小值的 (j,s) ;

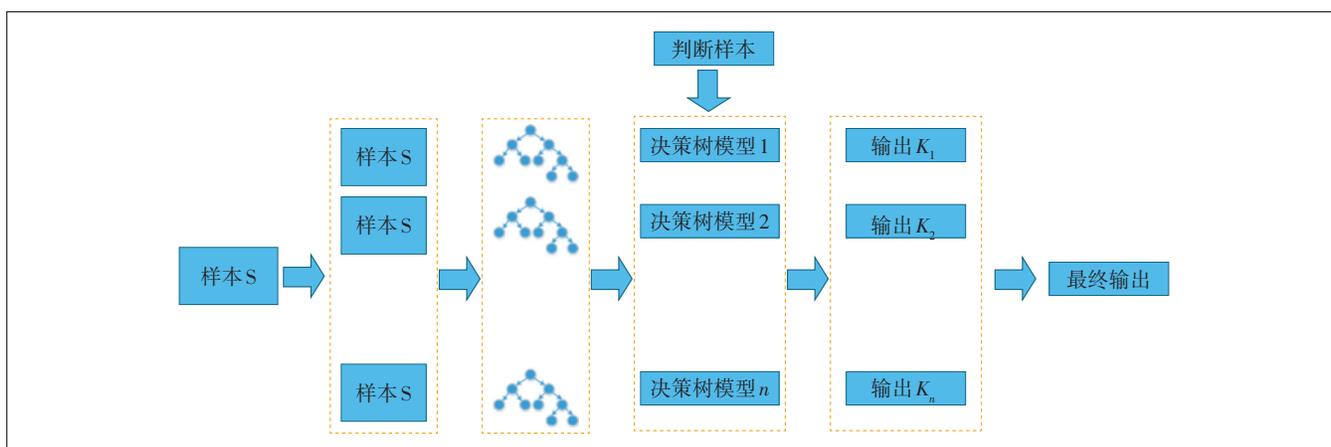


图6 随机森林流程图

用选定的 (j, s) 划分区域并决定相应的输出值:

$$\hat{c}_m = \frac{1}{N_m} \sum_{x_i \in R_m(j,s)} y_i, x \in R_m, m = 1, 2 \quad (18)$$

继续对2个子区域调用式(17)和式(18),直至满足停止条件。

将输入空间划分成 M 个区域 R_1, R_2, \dots, R_M ,生成决策树:

$$h(x) = \sum_{m=1}^M \hat{c}_m I(x \in R_m) \quad (19)$$

其中 I 为指标函数。训练得到决策树后即可得到室内环境下的路径损耗模型。

新的预测值可以对 T 个决策树的预测取平均,表示为:

$$y' = \frac{1}{T} \sum_{i=1}^T h_i(x) \quad (20)$$

其中 T 为决策树的数量。

3.2 模型误差分析

用构建的随机森林模型进行预测后,需要建立恰当的评价指标来验证模型的准确性。在模型验证时,可以用均方根误差(RMSE)、平均绝对百分比误差(MAPE)、平均绝对误差(MAE)来反映误差的大小,计算值越小说明模型训练的效果越好。

$$RMSE = \sqrt{\frac{1}{Q} \sum_{q=1}^Q (PL_q - PL'_q)^2} \quad (21)$$

$$MAPE = \frac{100}{Q} \sum_{q=1}^Q \left| \frac{PL_q - PL'_q}{PL_q} \right| \quad (22)$$

$$MAE = \frac{1}{Q} \sum_{q=1}^Q |PL_q - PL'_q| \quad (23)$$

其中 PL_q 为测量值, PL'_q 为预测值,本文采用RMSE和MAE来表征模型预测的效果。

3.3 位置估计

利用路径损耗模型,可得到用户与天线之间的水平距离之间的关系,在室内某一位置采集到CSI以后,可得到用户与天线之间的距离,联合多个天线与用户之间的距离,对用户进行定位,假设有3台基站,位置分别为 $(x_a, y_a), (x_b, y_b), (x_c, y_c)$,根据几何关系得到:

$$\begin{aligned} \sqrt{(x_i - x_a)^2 + (y_i - y_a)^2} - \sqrt{(x_i - x_b)^2 + (y_i - y_b)^2} &= \frac{1}{2} d_{ab} \\ \sqrt{(x_i - x_b)^2 + (y_i - y_b)^2} - \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} &= \frac{1}{2} d_{bc} \\ \sqrt{(x_i - x_a)^2 + (y_i - y_a)^2} - \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} &= \frac{1}{2} d_{ac} \end{aligned} \quad (24)$$

式中:

d_{ab}, d_{bc}, d_{ac} ——目标到第1个AP与到第2个AP的距离之差、目标到第2个AP与到第3个AP的距离之差、目标到第1个AP与到第3个AP的距离之差

(x_i, y_i) 表示目标的位置点,如图7所示。

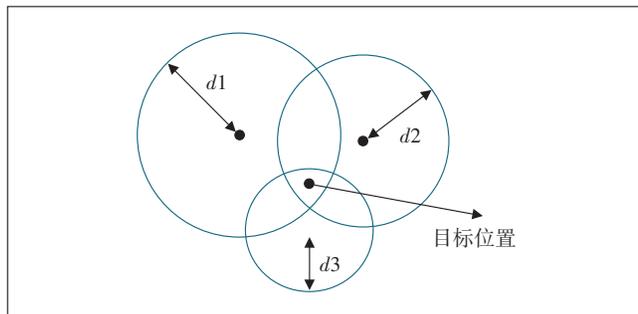


图7 多基站定位模型

4 误差分析

本文采用数据集来验证路径损耗模型的效果,本文将基于随机森林的训练模型与基于ANN的训练模型进行对比,得到的结果如图8所示。

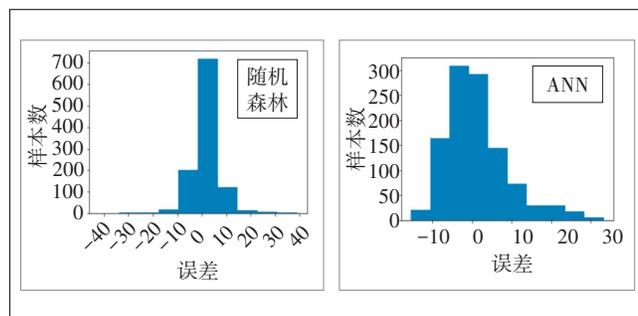


图8 误差分布

从图8可以看出,基于随机森林训练的模型要优于人工神经网络,本文计算这2种模型的MSE和MAE并与对数-正态模型(log-Distance)进行对比,得到的结果如表1所示。

本文分别将这3种模型的预测值与真实值进行对比,得到的结果如图9所示。

表1 3种模型的MSE和MAE对比

模型	MSE	MAE
随机森林(RF)	1.93	1.62
人工神经网络(ANN)	65.89	5.88
对数-正态模型(log-Distance)	102.83	8.23

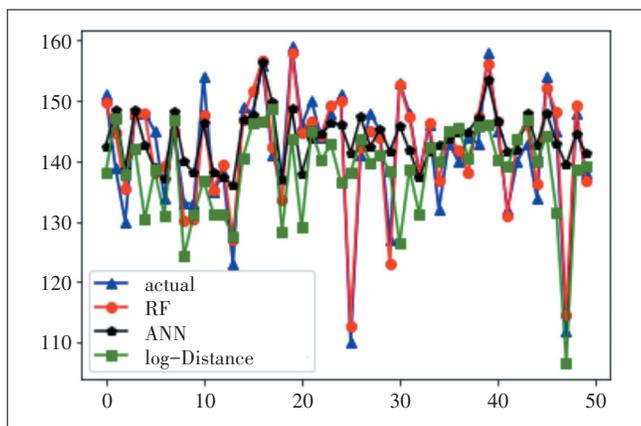


图9 预测值与真实值的对比

从图9中可以看出,基于随机森林的预测结果与真实值最为接近,而基于对数-正态模型的预测结果最差,这是因为对数正态模型只是一个经验值,并不能代表室内信号传播的真实情况,误差相对较大。将预处理后的CSI输入到训练好的模型即可得到距离信息,然后联合多个天线的水平距离信息即可得到用户的位置,定位结果如图10所示。

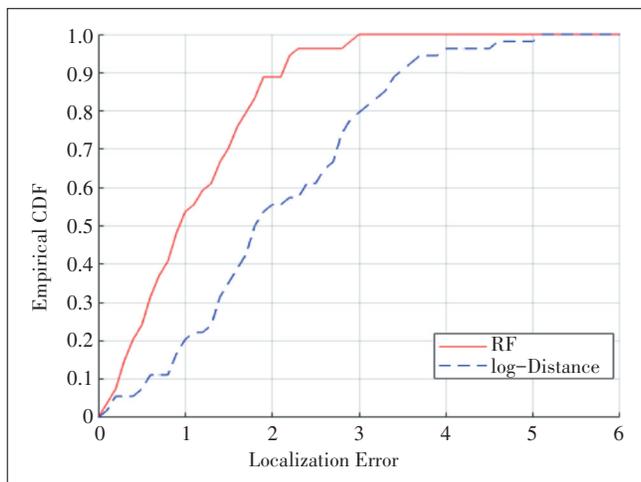


图10 定位结果

从图10中可以看出,本文提出的基于机器学习的定位效果要明显优于基于对数-正态模型,其中对数正态模型是在室内进行校准过,所使用的参数为最优参数。

5 结论

随着5G网络的开发和部署,网络规划对路径损耗预测的准确性、复杂度和通用性提出了更高的要求。同时路径损失模型在室内定位方面也起着重要的作

用,建立距离与能量之间的对应关系对定位的精度有很大的影响。基于机器学习的方法,特别是监督学习,可以建立复杂的非线性关系,提高路径损耗预测的准确性。本文提出一种基于机器学习的适用于数字化室分的室内定位方法,采集信道状态信息数据,并通过时频域处理以及灰度预测模型提高数据采集的准确性、稳定性、减少数据采集的数量,并根据机器学习算法在室内训练路径损耗模型,更加真实地反映室内的信号传播特性,仿真结果表明,该方法相比传统的对数-正态模型预测精度有很大的提升,将此模型用于距离估计以及定位将极大地提高定位的精度,同时可以解决目前5G定位中需要设备之间同步的问题。

参考文献:

- [1] EL-RABBANY A. Introduction to GPS: the global positioning system [M]. Boston, MA: Artech House, 2006.
- [2] RYCROFT M J. Understanding GPS. Principles and applications [J]. Journal of Atmospheric and Solar-Terrestrial Physics, 1997, 59(5): 598-599.
- [3] 田增山,张千坤,周牧,等. 基于CFR的角度/距离联合单站定位[J]. 电子科技大学学报, 2019, 48(5): 698-705.
- [4] 田增山,张千坤,周牧,等. 基于CFR虚拟阵列天线的AOA室内定位[J]. 电子学报, 2018, 46(6): 1468-1474.
- [5] 何维,岳紫颖,周牧,等. 基于载波相位差值的室内Wi-Fi定位方法[J]. 北京邮电大学学报, 2021, 44(4): 19-25.
- [6] 周牧,王斌,田增山,等. 室内BLE/MEMS跨楼层融合定位算法[J]. 通信学报, 2017, 38(5): 1-10.
- [7] 何成文,袁运斌,谭冰峰. 一种基于UWB TDOA定位模式的迭代最小二乘算法[J]. 大地测量与地球动力学, 2021, 41(8): 806-809.
- [8] HALPERIN D, HU W J, SHETH A, et al. Tool release: gathering 802.11n traces with channel state information [J]. ACM SIGCOMM Computer Communication Review, 2011, 41(1): 53.
- [9] 徐宁,丁松,公彦德. 灰色GM(1,1)预测模型及拓展研究进展[J]. 数学的实践与认识, 2021, 51(13): 52-59.
- [10] ÖSTLIN E, ZEPERNICK H J, SUZUKI H. Macrocell path-loss prediction using artificial neural networks [J]. IEEE Transactions on Vehicular Technology, 2010, 59(6): 2735-2747.

作者简介:

张千坤,助理工程师,硕士,主要从事移动增值业务相关咨询设计工作;陈任翔,高级工程师,学士,主要从事移动增值业务相关咨询设计工作;钟志刚,中讯邮电咨询设计院有限公司总工程师,教授级高级工程师,先后在微波通信、卫星通信和移动通信领域开展了工程咨询、规划、设计和研究等工作;周国栋,高级工程师,学士,主要从事移动增值业务相关咨询设计工作。