

基于OTT数据的 5G端网协同智能优化研究与实践

Research and Practice of 5G Terminal-network Collaborative
Intelligent Optimization Based on OTT Data

陈 锋,李张铮,连 慧(中国联通福州分公司,福建 福州 350000)

Chen Feng,Li Zhangzheng,Lian Hui(China Unicom Fuzhou Branch,Fuzhou 350000,China)

摘 要:

针对当前5G MR不携带用户位置信息无法实现用户网络感知精细化分析的问题,引入OTT大数据通过使用神经网络算法、DBSCAN算法等机器学习算法,建立5G用户位置预测和5G弱覆盖小区分布聚类模型,开展精细化的5G用户感知保障。研究表明,基于OTT数据的5G端网协同智能优化能有效提升5G优化效率,节省网络运营成本。

关键词:

5G端网协同;OTT数据;用户位置预测;小区聚类;机器学习

doi:10.12045/j.issn.1007-3043.2022.10.007

文章编号:1007-3043(2022)10-0032-06

中图分类号:TN919

文献标识码:A

开放科学(资源服务)标识码(OSID):



Abstract:

Aiming at the problem that the current 5G MR can not carry user location information to achieve user network awareness fine analysis,by introducing OTT big data and using machine learning algorithms such as neural network algorithm and DBSCAN algorithm,5G user location prediction and 5G weak coverage cell distribution clustering model are established to carry out refined 5G user perception guarantee. The results show that 5G termina-network collaborative intelligent optimization based on OTT data can effectively improve the efficiency of 5G optimization and save network operating costs.

Keywords:

5G Terminal-network collaboration;OTT data;User location prediction;Cell clustering;Machine learning

引用格式:陈锋,李张铮,连慧. 基于OTT数据的5G端网协同智能优化研究与实践[J]. 邮电设计技术,2022(10):32-37.

1 概述

在目前5G网络大规模部署背景下,要拉动5G业务收入增长,运营商面临的首要关键问题是5G用户的常驻区域的网络质量。保障最容易出问题的覆盖地点,是5G网络端网协同优化的重中之重。传统网络保障通过路测数据(DT)或者测量数据(MR)分析发现网络缺陷,其中DT数据只能反映点线上的问题,无法做到全网精细化评估,而MR数据虽然全面但是5G R15版本终端暂不支持带位置信息的测量数据,即使后续版本支持也存在部分终端不上报MR位置的问题,对5G网络质量的评估造成影响。

机器学习技术作为人工智能的重要组成部分,是国家发展战略重点扶持的目标^[1]和当下各行业关注应用的焦点。本文通过使用用户终端APP上产生的用户无线网络数据(OTT)来补偿5G MR的缺陷,利用神经网络算法^[2]和DBSCAN聚类算法^[3]构建5G用户位置模型和弱覆盖小区分布,为5G用户网络感知精细化保障奠定良好基础。

2 目前5G端网协同优化方法存在的不足

在当下5G端网协同优化中,5G用户感知精细化保障始终是一个难点。由于5G MR不携带用户经纬度信息,网络分析严重依赖于人的优化经验等方面的特点,急需引入大数据和机器学习的优化方式,实现5G端网协同优化的智能化。

收稿日期:2022-08-26

2.1 MR经纬度缺失,网络精细化分析困难

MR是UE在网络中上报的无线网络测量数据,包含小区电平、小区质量和小区覆盖范围等无线环境指标。在4G中MR还包含了测量报告发生地点的经纬度信息,网优人员可以利用经纬度信息来评估特定地点的网络覆盖情况,对网络感知进行精细化保障。目前5G用户级MR文件不携带经纬度数据,MR平台仅有天级的小区MR指标,对感知差的特定时间、特定位置和特定用户无法进行精细化网络指标分析,造成网络优化的瓶颈。

2.2 端网协同优化手段缺乏,对技术要求高

传统端网协同优化主要通过路测或者用户投诉发现无线网络覆盖问题,是目前主流的网络优化方法,其能在一定程度上解决无线覆盖问题,但对于覆盖问题不能第一时间解决,往往在问题发生后才能进行排查;解决问题阶段,目前主流分析方法为人工分析、排查收集到的质差数据,判断弱覆盖类型、弱覆盖原因并预估解决方案,分析人员的主观判断在这个过程中起主要作用,容易出现误判且消耗大量人力^[4]。

3 基于OTT数据的5G端网协同智能优化

5G终端MR经纬度位置信息的缺失直接影响到5G端网协同精细化分析,只能实现小区级的网络质量分析,另外传统根据MR采样点人工搜寻弱覆盖区域的优化方法费时费力。面对5G网络不断壮大的规模,亟待MR精细化智能化应用的出现。

本文通过研究OTT数据和MR数据的关联性,利用机器学习算法基于OTT数据建立5G用户位置预测模型,再迁移到MR数据上进行应用,预测出5G用户MR发生位置。计算出MR的位置后,对MR的主小区电平值进行聚类分析,得到弱覆盖小区分布。该算法可在网络指标优化、RF优化、用户投诉分析等网优日常5G优化工作中起到积极作用。

3.1 OTT及MR数据概述

3.1.1 OTT数据

OTT数据来自于真实的用户在真实的应用场景下产生的真实网络状况,而且覆盖范围广,包含异网运营商甚至Wi-Fi运营商的信息。

OTT数据采集的主要内容如下。

- a) 数据覆盖数百个手机APP,包括视频、直播、新闻、支付等。
- b) 数据覆盖中国的安卓用户,包含三大运营商及

2G/3G/4G/5G网络用户。

c) 数据包含无线网质量、Wi-Fi、定位、时间、基站、终端及APP等信息。

具体地,OTT大数据采集信息包括用户一些特定的APP上报的终端型号、运营商标识、网络制式、网络设备标识、网络指标(如RSRP、RSRQ及SINR)、邻区PCI、邻区电平、经纬度、室内外标识、检测最强Wi-Fi信息、Wi-Fi信号强度等(见表1)。

表1 OTT数据采集字段信息

字段分类	字段	说明	
5G基础参数及网络测量	基础参数	CELL_ID	小区ID
		ENBID	基站ID
		NR_PCI	5G小区PCI
		NR_ARFCN	5G小区频点
	服务小区测量值	NR_SSB_RSRP	5G下行广播信道信号强度测量值
		NR_SSB_RSRQ	5G小区参考信号质量
		NR_SSB_SINR	5G小区下行广播信道信噪比
	邻区测量值	NR_Neighbor_RSRP_list	邻区RSRP,并且降序排列
		NR_Neighbor_PCI_list	邻区PCI,对应顺序同NR.Neighbor_RSRP_list
		NR_Neighbor_EARFCN_list	邻区频点,对应顺序同NR.Neighbor_RSRP_list
位置信息	LGT	经度longitude(WGS84)	
	LTT	纬度latitude(WGS84)	

3.1.2 MR数据

MR数据主要来自终端对周边无线网络环境进行测量时产生的报告。MR数据包含终端当前位置服务小区/邻区ID、频点、PCI、RSRP、RSRQ和SINR等无线测量信息,目前R15版本不支持经纬度位置信息的上报。5G MR数据和OTT数据的对应关系如表2所示。

表2 OTT字段和MR数据对应关系

OTT字段	MR字段
NR_PCI	MR.NRScPei
NR_ARFCN	MR.NRScArfcn
NR_SSB_RSRP	MR.NRScSSRSRP
NR_SSB_RSRQ	MR.NRScSSRSRQ
NR_SSB_SINR	MR.NRScSSSINR
NR_Neighbor_RSRP_list	MR.NRNeSSRSRP
NR_Neighbor_PCI_list	MR.NRNePei
NR_Neighbor_EARFCN_list	MR.NRNeArfcn

3.2 基于OTT数据的5G用户位置预测

受限于R15 5G终端MR不包含经纬度,如何利用AI手段学习已有的带位置信息的数据来预测不带位

置信息的5G MR样本成为网优智能化的重要方向。本文对获取到的5G用户OTT数据,尝试使用神经网络算法对数据样本进行训练生成用户位置预测模型,鉴于全网统一建模的效果不理想,将数据分地域进行差异化建模,取得了较好的预测效果(见图1)。

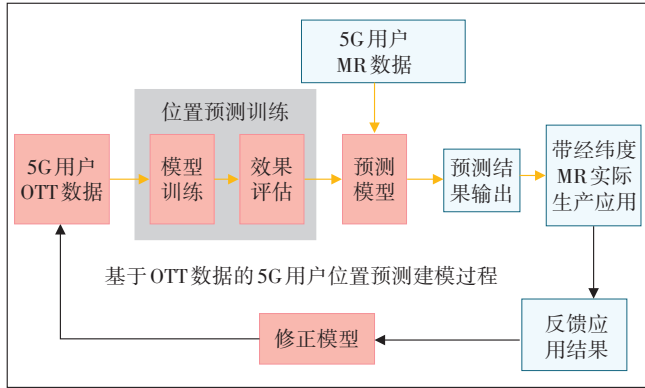


图1 基于OTT数据的5G用户位置预测流程

3.2.1 数据集准备

根据获取到的某市联通5G用户OTT数据,结合小区工参中经纬度、区县、方位角、下倾角和覆盖类型,共生成39个特征维度和2个标签维度的数据,将数据按7:3的比例切分为训练集和测试集,如表3所示。

3.2.2 数据预处理

3.2.2.1 高基数特征处理

基数(cardinality)指的是类别特征所有可能的不同值的数量,对于高基数的类别特征,直接使用Label Encoding和One-Hot Encoding的数据预处理的方法往往得不到令人满意的结果^[5]。分析此处enbid、pci类别特征可以发现存在高基数问题,如果直接进行One-Hot编码会形成上百列的稀疏矩阵,易消耗大量内存和训练时间;同时映射之间完全独立,并不能表示出不同类别之间的关系。本文基于神经网络算法使用Embedding层来处理高基数特征,Embedding技术的一个非常普遍的应用就是实现高维稀疏特征向量向低维稠密特征向量的转换,也就是把离散特征经过独热编码后的稀疏向量表达转化成稠密的特征向量表达。训练中可以通过优化网络的参数和权重来减少损失以改善embedding表示。

3.2.2.2 室分小区方位角的表示

通常工参里覆盖类型为室分的小区方位角都是0,这与实际室分小区为全向覆盖不符,故室分小区的方位角需修正。修正方法如下:若室分小区与宏站邻小区同经纬度,则室分小区取宏站邻小区的方位角;

表3 训练集/测试集特征标签

属性	训练/测试集字段	含义	来源
样本特征	s_cellid/n1_cellid/n2_cellid	5G服务小区/第1邻区/第2邻区id	OTT数据CELL_ID和小区工参
	s_enbid/n1_enbid/n2_enbid	5G服务基站/第1邻区基站/第2邻区基站id	OTT数据ENBID和小区工参
	s_pci/n1_pci/n2_pci	5G服务小区/第1邻区/第2邻区PCI	OTT数据NR_PCI和NR_Neighbor_PCI_list
	s_arfcn/n1_arfcn/n2_arfcn	5G服务小区/第1邻区/第2邻区频点	OTT数据NR_ARFCN和NR_Neighbor_EARFCN_list
	s_long/n1_long/n2_long	5G服务小区/第1邻区/第2邻区经度	小区工参
	s_lat/n1_lat/n2_lat	5G服务小区/第1邻区/第2邻区纬度	小区工参
	s_SSB_RSRP/n1_SSB_RSRP/n2_SSB_RSRP	5G服务小区/第1邻区/第2邻区电平	OTT数据NR_SSB_RSRP和NR_Neighbor_RSRP_list
	s_SSB_RSRQ	5G服务小区信号质量	OTT数据NR_SSB_RSRQ
	s_SSB_SINR	5G服务小区信噪比	OTT数据NR_SSB_SINR
	s_region/n1_region/n2_region	5G服务小区/第1邻区/第2邻区归属区县	小区工参
	s_azi/n1_azi/n2_azi	5G服务小区/第1邻区/第2邻区方位角	小区工参
	s_tilt/n1_tilt/n2_tilt	5G服务小区/第1邻区/第2邻区下倾角	小区工参
	s_covtype/n1_covtype/n2_covtype	5G服务小区/第1邻区/第2邻区覆盖类型	小区工参
样本标签	OTT_long	OTT发生经度	OTT数据LGT
	OTT_lat	OTT发生纬度	OTT数据LTT

若室分小区与室分邻小区同经纬度,则室分小区方位角取值-1;若室分小区与邻小区不同经纬度,则室分小区方位角取室分小区与邻小区连线与正北方向的顺时针夹角 r (见图2)。

设室分小区经纬度 (X_1, Y_1) ,邻小区经纬度 (X_2, Y_2) ,具体小区连线夹角 r 计算公式如下:

$$\begin{cases} x_1 = X_1 \times \frac{\pi}{180}, y_1 = Y_1 \times \frac{\pi}{180} \\ x_2 = X_2 \times \frac{\pi}{180}, y_2 = Y_2 \times \frac{\pi}{180} \end{cases} \quad (1)$$

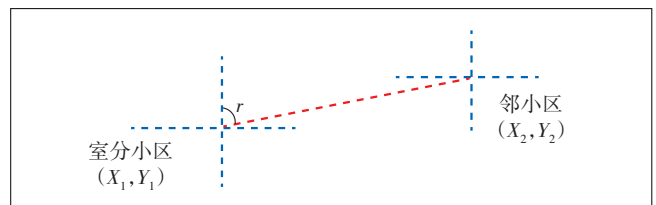


图2 室分小区方位角定义

$$r = \begin{cases} 180, & x_1 = x_2, y_1 > y_2 \\ 0, & x_1 = x_2, y_1 < y_2 \end{cases} \quad (2)$$

令

$$A = \frac{\sqrt{4 \left(\sin \frac{y_1 - y_2}{2} \right)^2 - \sin \frac{x_1 - x_2}{2} \times (\cos y_1 - \cos y_2)^2}}{\sin \frac{|x_1 - x_2|}{2} \times (\cos y_1 + \cos y_2)}$$

则

$$r = \begin{cases} 270 - \frac{\tan^{-1} A}{\pi} \times 180, & x_1 > x_2, y_1 > y_2 \\ 270 + \frac{\tan^{-1} A}{\pi} \times 180, & x_1 > x_2, \text{其他} \\ 90 + \frac{\tan^{-1} A}{\pi} \times 180, & x_1 < x_2, y_1 > y_2 \\ 90 - \frac{\tan^{-1} A}{\pi} \times 180, & x_1 < x_2, \text{其他} \end{cases} \quad (3)$$

图3给出了室分小区方位角特征校正代码。

3.2.3 神经网络模型构建及训练

本文神经网络模型设计采用 Embedding层+全连

```
for idx, df_idx in enumerate(df_indoor_cells.index.to_list()):
    if (long_arc_from[df_idx] == long_arc_to[df_idx]) & (lat_arc_from[df_idx] > lat_arc_to[df_idx]):
        df_indoor_cells.loc[df_idx, df_indoor_cells.columns[4]] = 180
    elif (long_arc_from[df_idx] == long_arc_to[df_idx]) & (lat_arc_from[df_idx] < lat_arc_to[df_idx]):
        df_indoor_cells.loc[df_idx, df_indoor_cells.columns[4]] = 0
    elif (long_arc_from[df_idx] == long_arc_to[df_idx]) & (lat_arc_from[df_idx] == lat_arc_to[df_idx]):
        if (df_indoor_cells.loc[df_idx, df_indoor_cells.columns[5]] == '室外'):
            df_indoor_cells.loc[df_idx, df_indoor_cells.columns[4]] = df_indoor_cells.loc[df_idx, df_indoor_cells.columns[6]]
        else:
            df_indoor_cells.loc[df_idx, df_indoor_cells.columns[4]] = 368 #若同一点上邻区为室分小区,则用368表示该邻区
    elif (long_arc_from[df_idx] > long_arc_to[df_idx]) & (lat_arc_from[df_idx] > lat_arc_to[df_idx]):
        df_indoor_cells.loc[df_idx, df_indoor_cells.columns[4]] = 270 - np.arctan((sub1[idx] - sub2[idx]) / (mul1[idx] * mul2[idx])) * 180 /
    elif (long_arc_from[df_idx] > long_arc_to[df_idx]) & (lat_arc_from[df_idx] <= lat_arc_to[df_idx]):
        df_indoor_cells.loc[df_idx, df_indoor_cells.columns[4]] = 270 + np.arctan((sub1[idx] - sub2[idx]) / (mul1[idx] * mul2[idx])) * 180 /
    elif (long_arc_from[df_idx] < long_arc_to[df_idx]) & (lat_arc_from[df_idx] > lat_arc_to[df_idx]):
        df_indoor_cells.loc[df_idx, df_indoor_cells.columns[4]] = 90 + np.arctan((sub1[idx] - sub2[idx]) / (mul1[idx] * mul2[idx])) * 180 /
    elif (long_arc_from[df_idx] < long_arc_to[df_idx]) & (lat_arc_from[df_idx] <= lat_arc_to[df_idx]):
        df_indoor_cells.loc[df_idx, df_indoor_cells.columns[4]] = 90 - np.arctan((sub1[idx] - sub2[idx]) / (mul1[idx] * mul2[idx])) * 180 /
```

图3 室分小区方位角特征校正代码

接层的构建方式,模型结构如图4所示。模型共包括4个隐藏层,第一隐藏层包含3×2个Embedding层输入s_enbid/n1_enbid/n2_enbid和s_pci/n1_pci/n2_pci高基数特征和1个Dense层输入其他数值特征;后面连续连接3个Dense层,最后输出层有2个神经元分别对应经

纬度。建模框架使用 tensorflow2.0 的 keras 库,主要使用的网络层有 Embedding、Dense,另外还涉及优化器、损失函数等训练参数的配置。训练中的训练集和验证集损失函数变化如图5所示,验证集 loss 从 0.017 9 降低到 0.007 2,达到较好的训练效果。

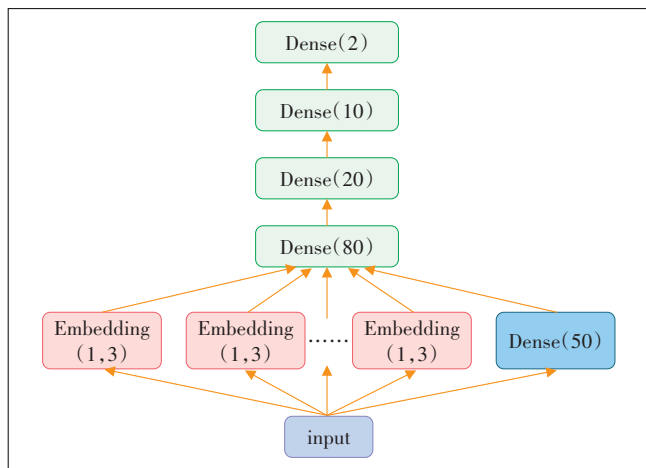


图4 基于OTT数据的5G用户位置预测神经网络模型结构

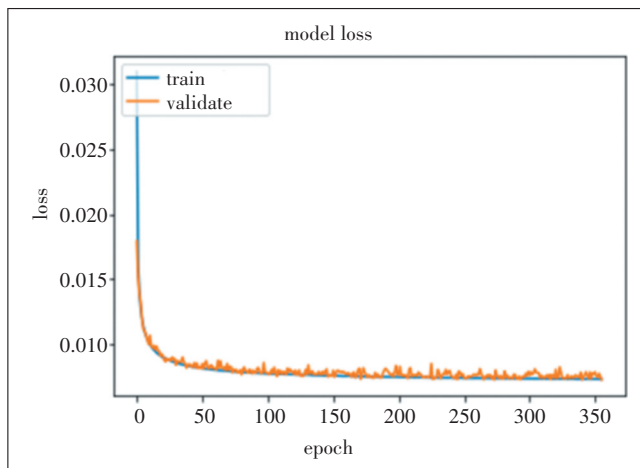


图5 神经网络模型训练集损失值和验证集损失值变化曲线

训练参数配置如下。

- a) 激活函数:ReLU 函数的变种 ELU。
- b) 优化器:Adma。
- c) 损失函数:MeanAbsoluteError。
- d) 超参数:训练轮数为 356,批大小为 200,使用 tf.keras.callbacks.ReduceLROnPlateau() 自动动态调整学习率。

3.2.4 用户位置预测结果及分析

使用获取到的某市全网 13 个区县所有 OTT 5G 用户数据进行训练,测试集上预测全网用户位置平均误差为 210 m,各误差段(以 100 m 为标准)的分布如图 6 所示。

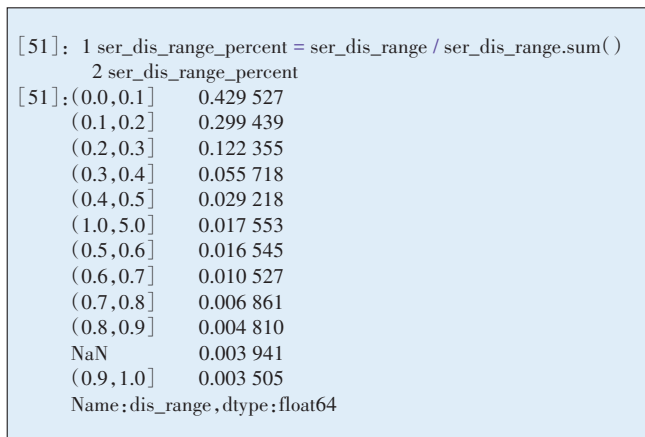


图 6 神经网络模型预测用户位置误差距离分布

观察预测结果误差分布,可以发现误差在不同区县间存在明显的差异(见图 7),原因可能是地域间网络结构的不同或者用户行为不同。其中该市全网平均预测误差为 210 m,有 4 个区优于平均水平;普通郊区误差大于 200 m,特别偏远地区误差最大,在 1 km 左右。

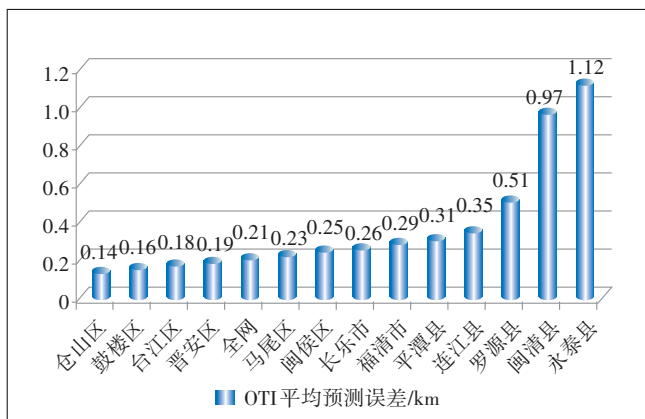


图 7 基于神经网络模型的 OTT 平均预测误差

受实际环境中不同区县间的网络发展不平衡和用户行为习惯不同等因素的影响,采用全网全量数据建立统一的模型存在如下缺陷:数据分布差异大,导致模型性能不佳;数据量巨大,模型训练时间长。因此后续考虑对每个区县分开建模,建立差异化模型,提升模型精确度。

3.2.5 差异化分地域神经网络建模结果分析

分别选取某市五城区 OTT 5G 用户数据样本作为训练集进行神经网络模型训练,并在测试集进行测试,平均误差距离为 74 m,相对全网统一建模神经网络算法精度有较大提升。改进后该市五城区的差异化建模测试效果如表 4 所示。

表 4 某市五城区差异化模型用户位置预测效果

区县	统一平均误差/m	差异平均误差/m	精度提升比/%	最小误差/m	最大误差/m
1	156	62	60.26	0	131
2	181	77	57.46	0	148
3	141	59	58.16	0	125
4	193	81	58.03	0	155
5	229	94	58.95	0	160

建立差异化用户位置预测模型后,根据表 2 中 OTT 字段和 5G 用户 MR 数据对应关系,将预测模型输入的 OTT 特征替换为 MR 样本特征,预测出 5G MR 的经纬度。后续就可利用带经纬度的 5G 用户 MR 开展 5G 网络优化工作。

3.3 基于 OTT+MR 数据的 5G 覆盖智能优化

对于庞大复杂的 5G 无线网络,网络覆盖问题层出不穷,现有无线覆盖分析的局限性日趋明显,如何尽可能减少人工路测成本和分析成本、提高分析判断准确性成为重中之重^[6]。本文实现基于 DBSCAN 算法的 5G 智能覆盖优化,在 5G 用户 MR 位置可以预测的基础上,利用密度聚类算法定位问题,高效挖掘 MR 弱覆盖区域。

5G MR 数据携带终端接收电平值等反映无线网络覆盖的信息,再加上前面建立的 5G MR 位置预测模型得到 5G MR 所在经纬度信息,在空间上可以通过密度聚类算法对存在弱覆盖问题的样本点进行聚类分析,挖掘得到弱覆盖区域。对某产业园用户 5G MR 采样点进行筛选,定义服务小区 RSRP<-110 dBm 的采样点为弱覆盖点,共筛选出弱覆盖采样点 1 923 个,如图 8 所示。

对这些弱覆盖点开展模型参数搜索,设定 DB-

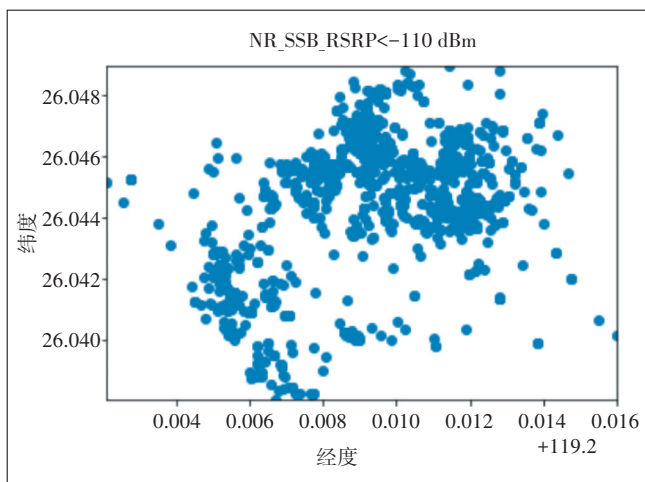


图8 某产业园5G弱覆盖点分布

SCAN 参数最少样本个数以 10 为步长,从 10 变化到 50,半径取 0.5。从计算结果可以发现,当最少样本个数为 10 时轮廓系数最高。此时聚类数量为 23 个,即发现该区域有 23 个弱覆盖场景需要 RF 优化(见图 9)。对比原始弱覆盖采样点的位置,可以发现 DBSCAN 算法精准地定位到了弱覆盖的位置,避免了人工搜索。

4 总结

传统 5G 端网协同优化是网优工作的重点和难点,人工优化方法费时费力。由于当前 5G 用户级 MR 不支持携带经纬度信息,优化中无法使用 MR 数据进行网络精细化分析。本文通过引入 OTT 数据建立 5G 用

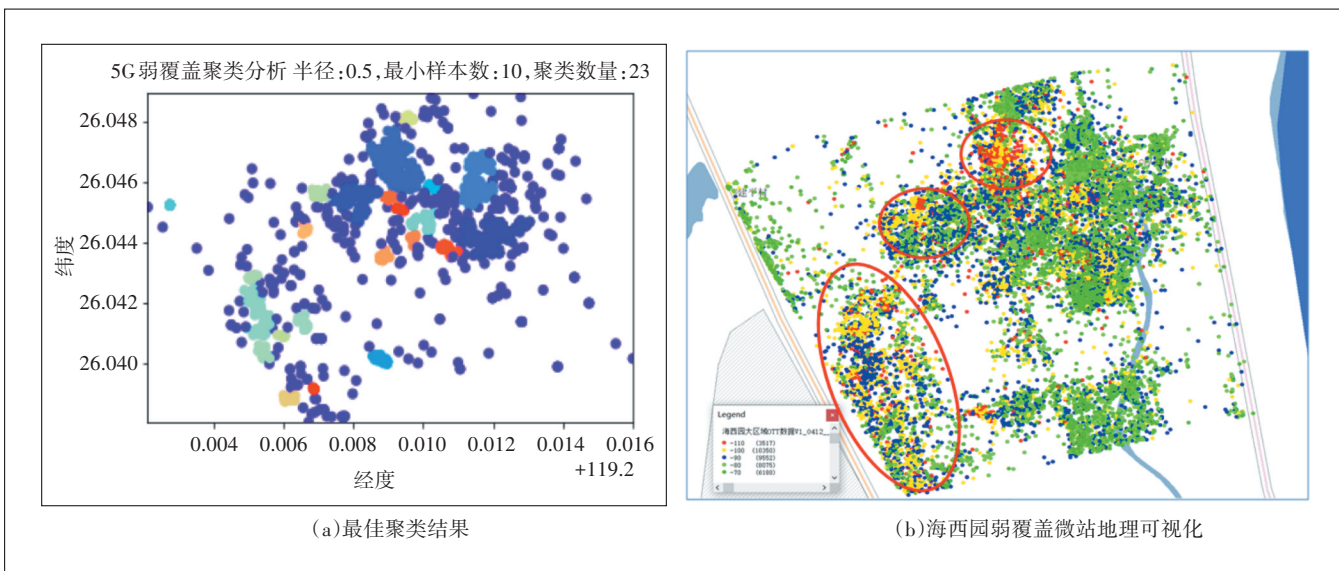


图9 基于OTT+MR数据的5G覆盖智能优化

户位置预测模型,利用 OTT 与 MR 数据的关联性获得 5G 用户级 MR 的经纬度信息。在此基础上,开展基于 DBSCAN 算法的 5G MR 覆盖智能分析,提升了优化效率,实现 5G 端网协同优化的智能化。

参考文献:

[1] 工业和信息化部发布《促进新一代人工智能产业发展三年行动计划(2018-2020年)》[EB/OL]. [2022-06-14]. https://www.miit.gov.cn/jgsj/kjs/jsxc/gjsfz/art/2020/art_291b5e6bc13f415494e84a0e9eac78f1.html.

[2] 李航. 统计学习方法[M]. 第2版. 北京:清华大学出版社,2019:211-278.

[3] 周志华. 机器学习[M]. 北京:清华大学出版社,2016:200-300.

[4] 张军民,金超,蒋伯章. 5G网络优化与实践进阶[M]. 北京:人民

邮电出版社,2021:200-267.

[5] MCKINNEY W. 利用Python进行数据分析[M]. 第2版. 北京:机械工业出版社,2018:101-279.

[6] 王强,刘海林,黄杰,等. 5G无线网络优化[M]. 北京:人民邮电出版社,2020:205-278.

[7] 左扬. 基于人工智能的5G无线网络智能规划和优化[J]. 电信科学,2020(S01):9.

[8] 王浩,赵伦. 人工智能在5G无线网络优化中的设计与实现[J]. 武汉工程职业技术学院学报,2021,33(4):5.

作者简介:

陈锋,毕业于福州农林大学,高级工程师,主要从事无线网络优化工作;李张铮,毕业于大连理工大学,工程师,主要从事无线网络优化工作;连慧,毕业于福州大学,工程师,主要从事5G共建共享工作。