

# 基于大数据+AI的 5G网络运营与保障

## 5G Network Operation and Guarantee Based on Big Data and AI

陈 锋, 连 慧, 李张铮, 洪林梦涵, 张雪平 (中国联通福州分公司, 福建 福州 350001)

Chen Feng, Lian Hui, Li Zhangzheng, Hong Linnmenghan, Zhang Xueping (China Unicom Fuzhou Branch, Fuzhou 350001, China)

### 摘 要:

传统5G用户营销方式精准度低,5G用户位置信息准确度不高,大多数MRO数据的用户位置信息缺失,无法针对性进行5G网络部署;传统网络感知分析手段有限,无法实现用户级评估;站点规划工作量大、人员水平有限的问题导致站址规划效果不佳。主要围绕5G用户是谁、5G用户在哪、5G用户网络感知如何这3个核心问题来开展研究,通过引入GBDT算法、随机森林算法和DBSCAN算法等AI智能算法,实现5G用户挖掘、5G用户位置预测、5G智能规划和5G网络感知保障等功能,完成基于大数据+AI的5G网络运营与保障。

### 关键词:

5G用户画像;5G端网协同;大数据;机器学习;网络优化

doi:10.12045/j.issn.1007-3043.2022.01.007

文章编号:1007-3043(2022)01-0036-06

中图分类号:TN915

文献标识码:A

开放科学(资源服务)标识码(OSID):



### Abstract:

The traditional marketing method is low precision, 5G user location information accuracy is not high, most of the MRO data user location information is missing, and can not be targeted for 5G network deployment. The traditional network awareness analysis means are limited, and can not achieve user-level evaluation. The problems of site planning workload and limited staff level lead to poor site planning. It focuses on the three core issues of who the 5G user is, where the 5G user is, and how the 5G user network perceives. By introducing AI intelligent algorithms such as GBDT algorithm, random forest algorithm and DBSCAN algorithm, the functions of 5G user mining, 5G user location prediction, 5G intelligent planning and 5G network awareness guarantee are realized, and the 5G network operation and guarantee based on big data and AI is completed.

### Keywords:

5G user portrait; 5G end-network collaboration; Big data; Machine learning; Network optimization

引用格式:陈锋,连慧,李张铮,等. 基于大数据+AI的5G网络运营与保障[J]. 邮电设计技术,2022(1):36-41.

## 1 概述

中国经济正处于新基建时代,移动通信日益改变大众工作生活方式,当前层出不穷的虚拟现实、增强现实、超高清视频等超大带宽移动互联网应用对移动宽带网速提出越来越高的要求<sup>[1]</sup>,5G网络应运而生。在目前5G网络大规模部署背景下,要拉动5G业务收入增长,运营商面临着两大首要问题:哪部分存量3G/

4G用户是潜在5G用户和这些潜在5G用户的常驻区域的网络质量。发展最可能成为5G用户的传统用户和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。传统营销方式精准度低,无法快速实现5G用户规模化;5G用户位置信息准确度不高,无法针对性进行5G网络部署,目前用户终端上报的MR信息中,支持A-GPS功能终端占比低,大多数MRO数据的用户位置信息缺失;传统网络感知分析手段有限,无法实现用户级评估;5G站点规划工作量大、人员水平有限的问题导致站址规划效果不佳。如何在提

收稿日期:2021-12-06

升5G用户营销成功率的同时保障5G用户级网络感知,提升5G站点规划效能,是运营商亟待解决的问题。

## 2 基于大数据+AI的5G网络运营与保障背景

基于传统方式的5G网络运营无法对哪些用户是5G用户、这些用户使用的网络位置在哪、该位置的用户网络感知如何等进行系统性分析,导致5G网络运营耗时耗力。主要表现在如下4个方面。

a) 移动网络传统用户营销方法存在诸多短板,比如营销策略模糊、目标用户存在盲目性、营销成效与人员水平相关等,严重影响5G用户营销效率,同时也造成5G网络用户保障的盲点。因此需要建立用户特征模型,实现潜在5G用户精准挖掘。

b) 用户级网络感知保障面临复杂多变的无线环境,不同的地点存在不同的网络质量,要对用户进行精细化保障就必须知道MR发生的真实位置,这样才能保证用户的网络质量问题可在地理空间上进行回溯,但是目前支持A-GPS功能终端占比低,大多数MRO数据的用户位置信息缺失。需要建立AI位置预测模型,预测无经纬度的5G用户位置信息。

c) 传统站点规划主要通过路测或者用户投诉发现无线网络覆盖问题,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;另外采用仿真手段,通过设定基站类型和网络参数输入选址仿真软件,输出基站建设位置,此方案成本较高且与实际网络环境存在偏差。需要在5G用户弱覆盖位置信息进行聚类,有效指导5G规划。

d) 传统网络感知保障仅限于小区级的KPI/KQI指标,无法对用户级无线环境质量进行保障,这就导致无线网络多数覆盖面上指标好,少数点上无线质量差的表里不一现象,无法发现实际存在的问题点,即精细化用户级感知保障缺失,给运营商网络口碑造成巨大压力。需要实现MR数据与SEQ数据的关联,同时依靠地理可视化工具,直观呈现用户级网络感知,提升网络优化效率。

## 3 基于大数据+AI的5G网络运营与保障方案

### 3.1 项目核心功能模块及创新点

本项目围绕Who、Where和How,即5G用户是谁、5G用户在哪、5G用户网络感知如何这3个核心问题来

开展研究,包含5G用户画像、5G用户位置预测、5G智能规划和5G网络感知保障四大核心模块。

a) 5G用户画像:要识别5G用户是谁主要解决如何找到5G用户的问题,传统运营商在制定5G用户营销策略时往往是一刀切,没有考虑用户个体差异性,5G用户发展效率低效果差,本项目针对这一短板利用机器学习分类算法,建立用户特征模型,实现潜在5G用户精准挖掘,提升营销效率节省成本。

b) MR用户位置预测:要知道5G用户在哪主要是获取用户位置信息,目前5G用户主要是NSA用户,用户同时附着于4G和5G网络上,可以利用4G MR获取用户位置信息,但仅有A-GPS终端能实现位置信息获取。目前A-GPS终端占比低,导致用户位置信息获取不完整。针对这个问题项目采用MRO多维数据进行AI建模,预测无经纬度的5G用户位置。

c) 5G网络感知保障:在知道用户是谁用户在哪后,就需要了解这些5G用户的网络体验如何,传统的优化手段通过SEQ与MR平台进行分析,但2个平台缺乏关联性,比如对于单用户速率慢问题,根据SEQ无法了解到当时用户位置的无线环境情况,根因定位困难,针对该问题,本项目组开发工具自动关联SEQ与MR用户数据,实现5G用户级感知监控。

d) 5G智能规划:基于5G用户的位置信息和感知就可以在5G建网初期进行更精准的站址规划与部署,但传统规划方法更多采用人工审核,找出规划站点过程中面临工作量大、人员水平参差不齐的问题,导致站址规划效果不佳,项目利用机器学习算法对5G用户覆盖差的位置进行智能聚类,输出5G网络覆盖亟待加强的区域,指导5G规划。

本项目主要创新点如下。

a) 5G用户画像:使用出账数据、上网行为、感知指标等维度数据对多种机器学习算法进行评估,选择最优的GBDT算法进行潜在5G用户建模,输出潜在5G用户清单。

b) 5G用户位置预测:结合MR数据的电平、质量、网络工参的小区经纬度、方位角、覆盖类型等多维度数据对多种机器学习算法进行模型选择,最后通过随机森林算法学习带经纬度标签的用户级MR特征数据,同时针对全网统一建模存在的定位误差不均衡问题,建立差异化潜在5G用户位置预测模型,输出潜在5G用户经纬度数据。

c) 5G智能规划:分不同场景基于DBSCAN聚类

分析对潜在5G用户位置进行聚类发现无线环境差的地点,指导5G智能规划。

d) 5G网络感知保障:融合时空维度无线网络质量数据和用户经纬度信息,可视化呈现监控5G用户网

络感知,提升5G网络口碑。

图1给出了基于大数据+AI的5G网络运营与保障项目框架。

### 3.2 项目具体实现算法与应用

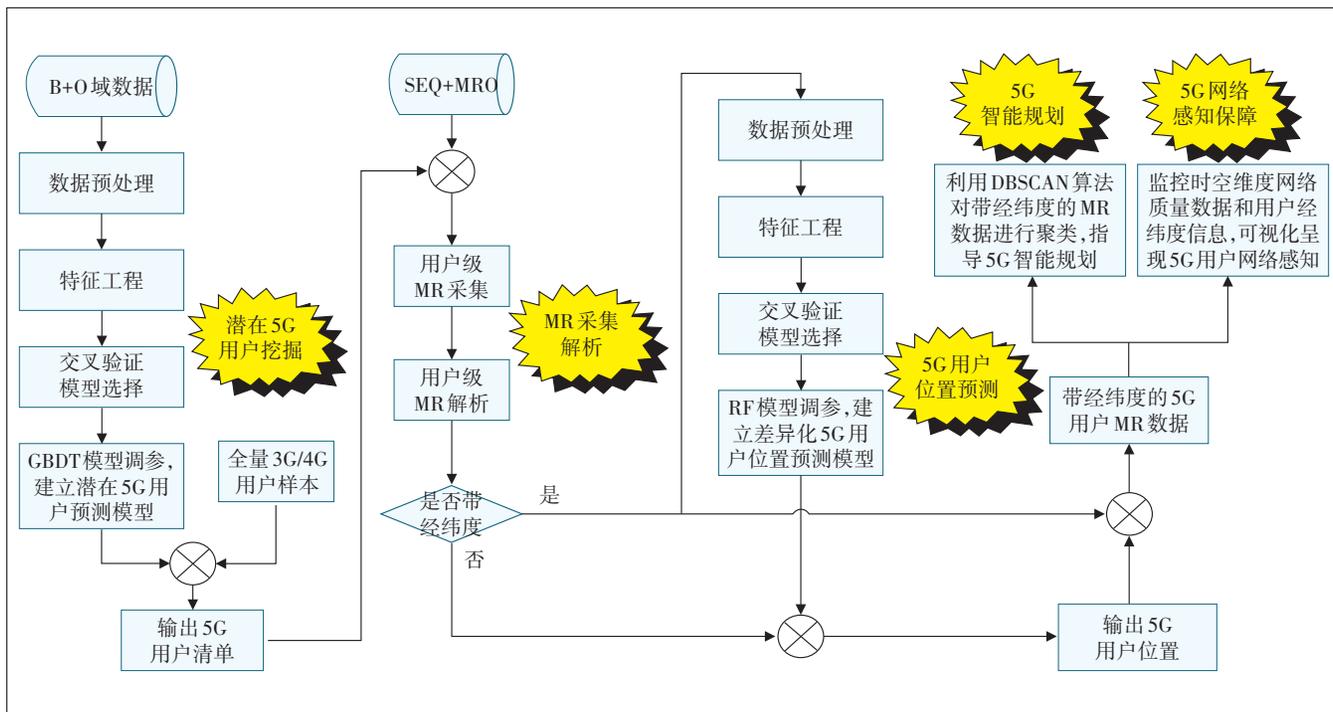


图1 基于大数据+AI的5G网络运营与保障项目框架

对每个上述核心功能模块,本项目严格按照机器学习建模的标准流程,从数据源获取、数据集拆分、数据预处理、特征工程、模型选择和超参数调优等步骤进行模型开发<sup>[2]</sup>,最终实现基于大数据+AI的5G网络运营与保障。

#### 3.2.1 基于AI算法的5G用户智能预测

本项目通过梯度提升GBDT机器学习算法学习5G用户正负样本历史上的B域出账数据和O域网络数据,建立5G用户分类预测模型,预测出传统移动网络用户是否是潜在5G用户。该模型可在5G用户营销

支撑、5G网络感知保障等网优日常工作中起到积极作用。

a) 数据准备。收集5G用户和非5G用户在历史3G/4G网络上的字段数据作为样本,这些原始字段包含B域的用户基础信息和用户消费信息、O域的用户上网行为和用户网络感知KQI指标(见表1)。

b) 数据预处理。主要是对原始数据的缺失值或异常字符进行校正替换,高维度类别特征进行降维处理,类别特征不平衡字段进行重新归并。

c) 特征工程。对样本数据分别采用特征构造、特

表1 5G用户正负样本原始字段

用户基础信息								
用户号码	账期	年龄	性别	归属城市	套餐名称	入网时间	终端厂家	单双卡终端
用户消费信息								
出账收入	语音通话时长	数据流量	语音通话时长2G	语音通话时长3G	流量2G	流量3G	流量4G	
用户上网行为								
最大使用APP协议大类			最大使用APP流量占比			前5名的APP流量占比		
用户网络感知KQI指标								
页面响应成功率			视频流媒体初始播放成功率			视频流媒体有效下载速率		

征抽取和特征选择进行处理, 举例如下:

(a) 特征构造。通过设置标杆时间将原始字段的“入网时间”转换成在网月数。

(b) 特征抽取。使用PCA算法对KQI数据进行降维, 使用降维后的PCA分量与标签做相关性分析, 保留最相关的分量作为最终特征进行模型训练。

(c) 特征选择。采用相关系数法进行特征筛选, 为了避免部分特征间的相关性过高, 导致多重共线性, 影响模型效果, 删除相关性高的特征。

d) 模型选择与优化。使用多种机器学习模型进行交叉验证打分, 选择分数最高的GBDT模型进行建模训练(见图2)。

选定模型后利用随机搜索算法来调整GBDT算法

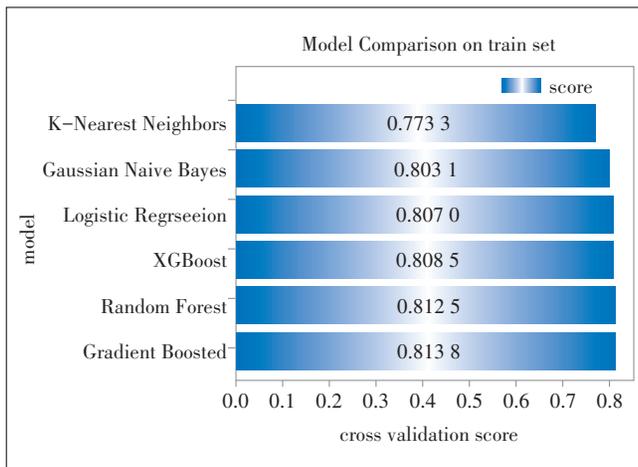


图2 基于交叉验证的分类模型选择

的超参数, 最终搜索得到最佳超参数组合。在测试集上进行评估, 分类正确率 accuracy 为 0.858, 召回率为 0.732。对比二分类随机判别 50% 的正确性, 本模型有较好的预测性能。

### 3.2.2 基于MR的AI智能用户位置定位算法

受限于终端不具备A-GPS功能, 目前MRO携带位置信息的MR数据占比较低, 如何利用AI手段学习已有的带位置信息的MR数据来预测不带位置信息的MR样本成为网络优化智能化的重要方向。本项目使用自主采集解析的MRO多维用户MR数据, 尝试多种AI算法对MR样本进行训练, 最终确定使用随机森林算法分地域差异化建模, 取得了较好的预测效果<sup>[3]</sup>。

a) 数据准备: 利用前面自主开发的MRO采集解析工具对当日1h的福州联通全网MRO数据进行采集解析, 得到11个维度的MR数据, 结合小区工参中经纬度、区县、方位角和覆盖类型, 共生成16个维度的训练

集数据(见表2), 同样采集次日同时段的全网MRO数据作为测试集数据。

b) 数据预处理。基数(cardinality)指的是类别特

表2 16个维度的MR训练数据

特征名称	含义	特征类型	编码方式
MR.LteScCgi	小区唯一标识	类别	Embedding
MR.LteScRSRP	服务小区电平	数值	-
MR.LteNcRSRP	邻区电平	数值	-
MR.LteScRSRQ	服务小区质量	数值	-
MR.LteNcRSRQ	邻区质量	数值	-
MR.LteScTadv	服务小区TA	数值	-
MR.LteScSimrUL	服务小区上行信干比	数值	-
MR.LteScEarfcn	服务小区频点	类别	one-hot
MR.LteScPci	服务小区扰码	类别	Embedding
MR.LteNcEarfcn	邻区频点	类别	one-hot
MR.LteNcPci	邻区扰码	类别	Embedding
Longitude	服务小区经度	数值	-
Latitude	服务小区纬度	数值	-
县市	-	类别	one-hot
AZIMUTH	小区方位角	类别	one-hot
覆盖类型	室分/室外	类别	one-hot

征所有可能的不同值的数量, 在高基数的类别特征面前, 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。本项目的MR.LteScCgi、MR.LteScPci和MR.LteNcPci 3个类别特征可以发现存在高基数问题, 因此对所述的高基数特征采用特征Hash编码的方式映射到低维空间, 使用API为: sklearn.feature\_extraction.FeatureHasher。

c) 模型选择: 分别使用基于多目标回归的线性回归、XGBoost、GBDT、Ridge和随机森林算法进行5折交叉验证模型选择(见图3), 发现随机森林算法得分最高(见图4), 因此使用随机森林算法进行建模。

d) 测试集测试结果分析: 预测结果分区县来看, 存在非常明显的地域差异, 其中福州城区台江、鼓楼、仓山和晋安好于全网平均水平, 台江最好, 误差仅140m; 偏远区县闽清和永泰误差最大, 误差达到1km以上(见图5)。不同区域的基站和人群的分布不一, 全网建立一个通用的且符合预期精度的模型不具备实践性和可操作性。

e) 算法优化: 鉴于上述缺陷, 调优的主要思路是选取一个大小合适的区域进行建模和训练, 将算法选型和模型参数优化好之后, 再在其他区域进行推广建模, 建立差异化的位置定位算法模型。本项目对福州

```

model_scores_lst = []
for estimator in model_estimators_lst:
    model_score = cross_val_score(estimator, df_features, df_target, cv=5, verbose=2,
                                  scoring= 'neg_mean_absolute_error', n-jobs=-1)
    print( ' current model score is: {0}'.format(model_score) )
    model_scores_lst.append(model_score)

# Dataframe to hold the results
model_comparison = pd.DataFrame({' model ' : [ ' Linear Regrseccion ',
                                                ' Ridge ',
                                                ' Random Forest ',
                                                ' Gradient Boosted ',
                                                ' XGBoost ' ],
                                ' score ' : [model_scores_lst [0].mean(), model_scores_lst [1].mean(),
                                              model_scores_lst [2].mean(), model_scores_lst [3].mean(),
                                              model_scores_lst [5].mean(),
                                              ]})
    
```

图3 交叉验证模型选择代码

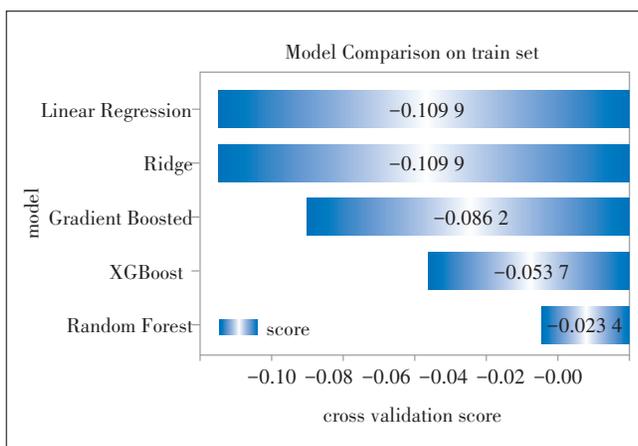


图4 交叉验证模型选择得分

联通 13 个区(县)MR 分别进行差异化建模,取次日五城区同时段 MR 数据作为测试集进行测试,福州五城区的差异化建模测试效果如表 3 所示,与统一建模相比,预测精度有较大提升。

### 3.2.3 基于 DBSCAN 算法的 5G 智能规划

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<sup>[4]</sup>。本项目提出的基于 DBSCAN 算法的 5G 智能规划在用户 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。对福州联通某区域用户 MR 采样点进行筛选,定

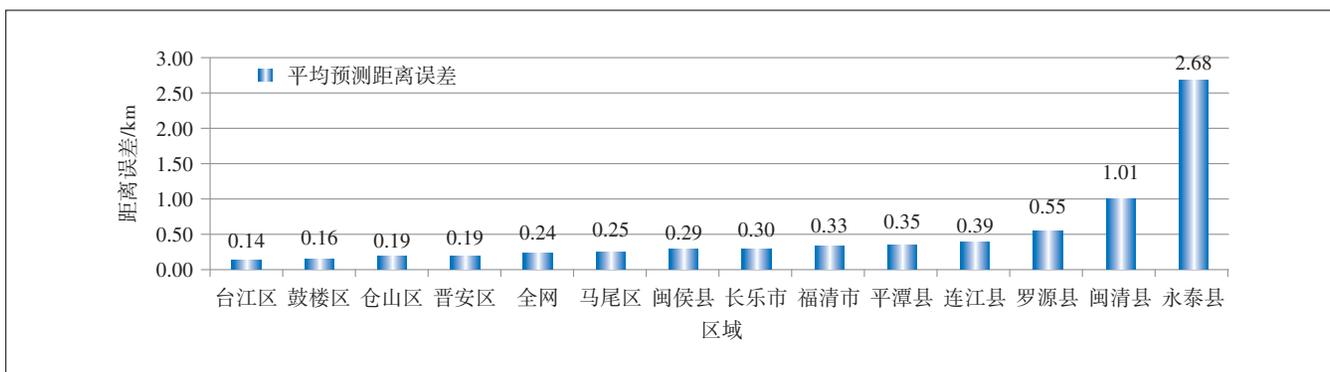


图5 基于全网统一模型的平均预测距离误差

义服务小区 RSRP 小于 -110 dBm 的采样点为弱覆盖点,共筛选出弱覆盖采样点 3 500 个<sup>[5]</sup>。

对这些弱覆盖点开展第一轮粗粒度的参数搜索,先设定 DBSCAN 参数最少样本个数以 10 为步长,从 10 变化到 100,半径始终 200 m,经过 DBSCAN 聚类得到的 10 种弱覆盖簇(见图 6)。从计算结果可以发现,最

少样本个数为 30 时,轮廓系数值最高 0.368 5。第 2 轮开展细粒度的参数搜索,设置最小样本个数以 30 为中心,步长为 1,从 25 变化到 34,半径始终 200 m,使用聚类算法分析当最少样本个数为 25 时轮廓系数最高为 0.388 7。此时聚类数量为 26 个,即该区域发现 26 个弱覆盖场景,需要规划建设。

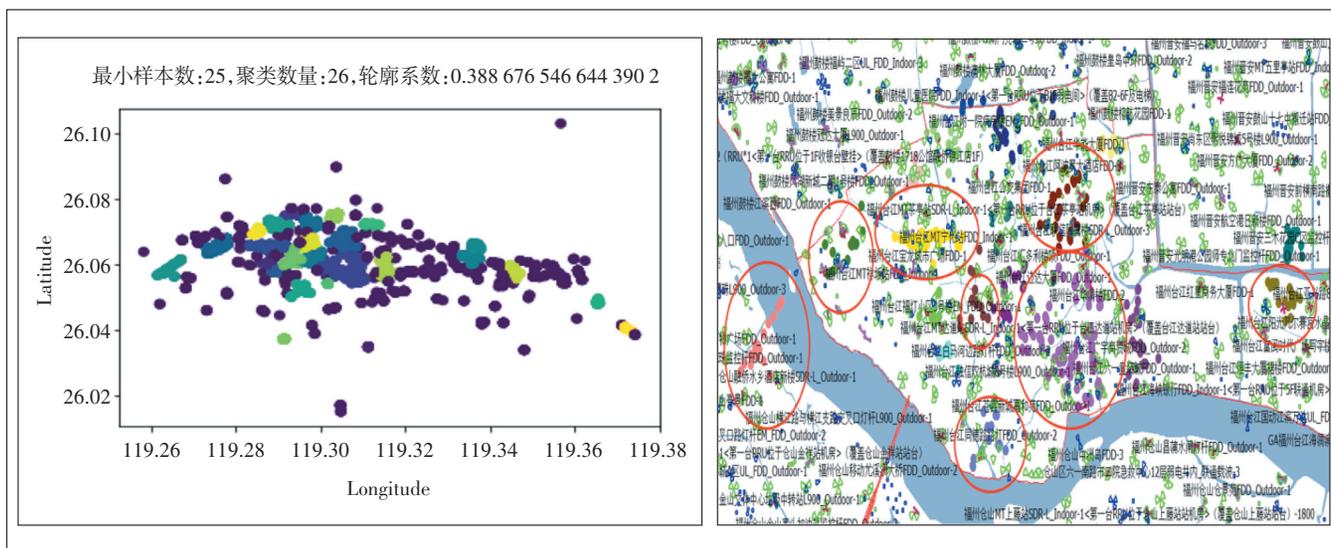


图6 福州联通某区域弱覆盖 DBSCAN 聚类结果

### 3.2.4 基于用户级MR的5G网络感知保障

传统网优平台仅有天级小区MR指标,无法针对用户投诉信息进行回溯。本项目将用户SEQ数据与MR信息进行关联,利用基站ID、用户S1AP ID和问题发生时间回溯用户当时特定地点的无线网络环境,分析解决用户网络感知问题<sup>[6]</sup>。此外,在用户级感知保障的基础上扩展实现区域级保障,将用户级MR汇总到相应区域,渲染出全区的MR感知分布图,应用于区域级网络优化。

### 3.3 项目应用情况

本项目的落地实施为某市联通5G市场营销和网络运营带来了巨大效益。在5G用户画像应用方面,某市联通2020年9月实际新增5G终端用户11311户,其中存量用户5276户,5276户中预测出的用户有2981户,命中率56.5%。2020年11月份的5G用户预测已向市场部门完成第2次用户推送,累计推送用户4.9万。在5G智能优化方面,基于精细化MR的5G用户感知获取模块落地后,用户级保障共分析4G/5G用户321人,其中发现问题63个,解决42个,闭环率67%;区域级保障共发现172个问题,解决117个,闭环率68%。问题平均解决时间由原来3天/人降低到当前0.5天/人,极大提升了问题解决效率。通过基于DBSCAN聚类分析网络质量模块,发现4G/5G现网MR弱覆盖地点97处,对于这些问题借助5G站点开通过程中的5G网络优化来解决,同时解决4G底层网覆盖问题;第1期已解决42个问题点,第2期计划完成剩余45个问题点。

## 4 总结

推进全面数字化转型是贯彻国家新基建发展战略的具体行动,站在数字化转型的新关口,福州联通充分利用运营商大数据优势,结合现有的AI应用能力,针对5G用户是谁、5G用户在哪、5G用户网络感知如何这3个核心问题来开展基于大数据+AI的5G网络运营与保障研究,取得了一系列创新成果,对加快福州地区5G新基建发展、加快5G应用、引领科技创新、助推产业升级意义重大。

### 参考文献:

- [1] 瞭望. 新基建十大战略方向[EB/OL]. [2021-09-21]. [http://www.xinhuanet.com/2020-04/29/c\\_139017002.htm](http://www.xinhuanet.com/2020-04/29/c_139017002.htm).
- [2] 周志华. 机器学习[M]. 北京:清华大学出版社,2016:225-266.
- [3] 李航. 统计学习方法[M]. 2版. 北京:清华大学出版社,2012:200-256.
- [4] 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,2020:219-224.
- [5] MCKINNEY W. 利用Python进行数据分析[M]. 2版. 北京:机械工业出版社,2018:199-201.
- [6] 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,2021:194-233.

### 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;连慧,毕业于福州农林大学,工程师,主要从事5G共建共享工作;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;洪林梦涵,毕业于集美大学,工程师,主要从事无线网络优化工作;张雪平,毕业于福建农林大学,工程师,主要从事无线网络优化工作。