

基于 MR 大数据的 4G 网络深度共建共享创新实践

王磊

中国电信股份有限公司，新疆维吾尔自治区分公司，新疆 乌鲁木齐 830000

[摘要]新疆电信和新疆联通已经逐步进入 4G 网络深度、全面共建共享合作阶段，文中阐述两项应用新疆电联深度共建共享的实践创新技术：基于空间数据计算高效实现深度共建共享中 TAC 区对齐及 MR 大数据分析共建共享后网络覆盖评估的应用。以解析后 MR 数据为基础，结合 Geopandas 地理空间大数据计算，利用构建轻量化高效的共建共享网络评估方法，并应用于疫情期间通信保障，及敏捷开发支撑政企市场需求的应用案例。

[关键词]：TAC 对齐；cKDtree；MR (MDT)；地理空间计算；共建共享

DOI: 10.33142/sca.v5i5.7342

中图分类号：TN929.5

文献标识码：A

4G Network Deep Co-construction and Sharing Innovation Practice Based on MR Big Data

WANG Lei

Xinjiang Uygur Autonomous Region Branch of China Telecom Co., Ltd., Urumqi, Xinjiang, 830000, China

Abstract: Xinjiang Telecom and Xinjiang Unicom have gradually entered the stage of 4G network depth and comprehensive joint construction and sharing cooperation. This paper describes two practical innovation technologies that apply Xinjiang ITU to in-depth joint construction and sharing: The application of TAC area alignment in in-depth joint construction and sharing based on spatial data computing and the application of network coverage evaluation after joint construction and sharing based on MR big data analysis. Based on the analyzed MR data, combined with Geopandas geospatial big data calculation, the evaluation method of building a lightweight and efficient co construction and sharing network is used to ensure communication during the epidemic, and the application case of agile development to support the market demand of government and enterprises.

Keywords: TAC alignment; cKDtree; MR(MDT); geospatial computing; construction and sharing together

引言

为贯彻党中央、国务院“资源整合，共建共享”的指示精神，落实网络强国战略，中国电信与中国联通以 5G 共建共享为契机，积极深入推进各场景 4G 网络的共建共享，进一步提升网络竞争力，降低网络整体 TCO。伴随着双方共建共享工作的升入，也逐步暴露出一下阻碍其进制的技术难题。第一，由于中国电信和中国联通在推进共建共享前已完成了 4G 网络的基本建设组网，双方基站编号规则，如 eNodeB ID、Cell ID、TAC ID 都已按最初组网规划各自分配，双方各省规划基本不同。其中 4G 跟踪区 (TAC)划分有较大差异，同一区域跟踪区编码互不一致，TAC ID 的设置对共享载波方式影响较大，造成了共享载波方式下双方跟踪区编码插花问题，这一问题严重制约了双方 4G 共建共享工作的推进。

第二，双方如何高效找到 4G 网络的覆盖薄弱点来指导共建共享建设？如何客观评价已经共建共享完毕的区域的共享效果？

面对上述网络问题，传统的网络规划和评估手段不仅耗费大量人力物力，而且双方数据也有可能不全面，同时也难以满足实时高效的要求。

本项目旨在构建深度 4G 共享网络运营优化体系，对深度共享网络运营优化的关键问题开展研究，基于大数据

结合地理空间并行计算，促进 4G 共享网络联合优化效率和用户感知双提升，打造无线网、网管、核心网端到端智能化的 4G 共享网络运营优化体系，从规建维优营网络生命周期全流程流为共享网络保驾护航。

1 基于空间计算的共建共享 TAC 对齐方法

1.1 TAC 插花问题

4G 共建共享分为独立载波和共享载波两种方式，通过表 1 两种共享方式的差异可以看出，共享载波方式具有投资较小的优势，适合在低业务量场景低成本快速部署，符合目前联通和电信共享场景的需求。

表 1 独立载波和共享载波差异点

差异	独立载波方式	共享载波方式
载波配置	联通、电信各自配置一个独立载波	联通、电信配置一个共享的载波
组网方式	联通、电信各自使用自己的频点，各自网络不存在异频切换	联通、电信使用同一个频点，共享方将引入异频切换
参数设置	独立的小区级特性，共享双方可按各自需求进行参数配置	相同的小区级特性，仅部分参数可按照 PLMN 设置
资源分配	联通和电信用户在各自独立的载波作业，不需要考虑空口资源分配和控制算法	联通和电信用户在同一载波下进行业务，需要考虑空口资源分配策略
投资差异	需要扩容第二载波，投诉较大	不需要扩容第二载波，投诉较小

共享载波方式共享双方使用同一个载波,大部分参数无法独立设置,其中就包括跟踪区编码(Tracking Area Code, TAC)。跟踪区编码与 PLMN 共同组成跟踪区标识(Tracking Area Identity, TAI),用于全网唯一标识1个跟踪区(Tracking Area, TA)。为了确定 UE 的位置,LTE 网络覆盖区划分为许多跟踪区,其功能与 3G 的位置区(Location Area, LA)和路由区(Routing Area, RA)类似,是 LTE 系统中位置更新和寻呼的基本单位。当 UE 从一个跟踪区进入另一个跟踪区时,必须新的跟踪区上重新进行位置登记,以通知网络更改它所存储的 UE 位置信息,这个过程就是跟踪区更新(Tracking Area Update, TAU)。共享载波方式一个小区只能设置一个 TAC,如果联通和电信共享区域的 TAC 不同,共享方将出现 TAC 插花问题,造成频繁 TAU,见图 1。



图1 共享双方因 TAC 插花引起 TAU

1.2 电联双方 TAC 调整原则

联通和电信双方网络存在一定差异,TAC 调整中有不同需求,为有效指导 TAC 调整工作方案制定,根据电联双方集团公司全面 4G 共建共享的工作需要,区公司云网运营中心与电信无线局双方经探讨协商确定了 TAC 调整工作总体调整原则:

- (A) TAC 调整方案应尽量减少双方调整工作量;
- (B) 联通 TA 和 LA 有绑定关系,保持联通 LA 不变,通过 4G TA 调整实现;
- (C) 联通启用 Talist 功能,对 TA 进行整合;
- (D) 为避免一个 TA 对应多个 LA,调整后的 TA/Talist 不能跨联通 LA;
- (E) 为实现特服分区功能,TA 不跨行政区边界;
- (F) 避免出现过小 TA 浪费 TAC 资源;
- (G) 考虑共享站点接收双方寻呼带来的寻呼量增加因素,单 Talist 覆盖下建议站点数量不超过 300 个基站(方案 TAL 下的一方自有站+另一方共享站总量不超过 300 个基站)。

1.3 TAC 调整方案分析

以新疆某地市某县为例,联通现网存在 2 个 TAC 区域,电信现网存在 4 个 TAC 区域;如果为了今后双方全量共建共享,而不至于产生大量插花及 TAU 负荷巨增,目前按照与新疆电信无线网优中心商讨决定,为减少对现网影响,建议双方采用方式一进行 TAC 调整工作。具体就是联通 TAC 边界与电信 TAL 区域对齐(电信 TAL 不跨联通 LAC),

对齐方案和效果图如图 2 和图 3;

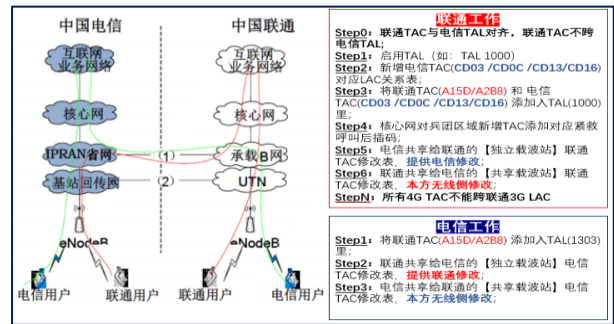


图2 联通 TAC 与电信 TAL 边界对齐方案

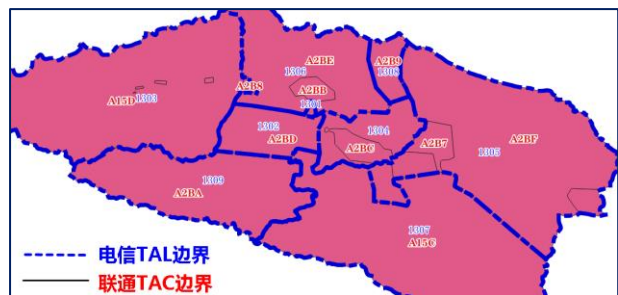


图3 某区域内联通 TAC 边界与电信 TAL 区域对齐效果图

结合地理空间计算的 TAC 效验原理

随着电联双方无线网络侧基站工参数数据的交换,如何高效准确分析挖掘哪些基站的 TAC 对应 LAC 关系,并且需要将双方 TAC 与 TAC LIST 的关系也进行校验核对。为提高共建共享效率,就需要借地理信息系统(GIS)建立、呈现、分析带有地理空间属性的数据。对于 GIS 里最基本的三个元素:点要素/线要素/面要素,其在无线网优对象中也可以对应,GIS 元素及无线优化对象主要对应关系及方法如表 2。

表2 地理空间原始与无线对象关系及方法

无线网优对象	对应GIS元素	Geopandas方法	优化应用案例
基站小区/MDT	point(点属性)	gpd.points_from_xy(df1.lon, df1.lat)	最近相邻基站/最大站距
道路/铁路	line(线属性)	line = LineString(xyList)	线场景内覆盖小区搜寻
地州/市/县/场景	polygon(面属性)	polygon = Polygon(xyList)	面场景内基站归属/场景区域内MR/MDT计算

上述点/线/面元素其实都带有 GPS 经纬信息,默认也是 WGS-84(标准地心坐标系)。我们就可以使用 geopandas 函数对其进行高效地理坐标化处理。

我们引入机器学习算法,使用 cKDTree 算法,快速计算采样点距离及地理空间归属问题。

表3 不同算法计算 3G 站点最近的 4G 站信息耗时对比

使用函数算法	3G基站数	4G基站数	运算量级(亿)	耗时
MySQL的ST_DISTANCE_SPHERE()函数				15分39秒
python的shapely库nearest_points()函数	31952	28201	9.01078352	4分27秒
python的scipy.spatial库cKDTree()函数				1.79秒

如表 3,不同算法计算,全网 3G 基站点最近的 1 个 4G 站信息耗时对比,使用 cKDTree 算法后,比 MySQL 加索引优化后仍提升 524 倍,比 shapely 库循环计算 nearest_points() 提升 149 倍;效率提升明显。

2 基于地理空间技术+MR 的共享网络评估方式

MR 的数据采集和 MR 大数据计算都已有完善的流程和现成系统平台,完全解决了 MR 原始数据的统计计算问题。由于各场景 MRS 文件均采样 XML 格式,对应单个站点或少量 MRS 的数据解析,我们也尝试使用 python 的 Xml.etree.ElementTree 库进行解析读取。但只作为验证数据自动的一个补充手段。

本文主要介绍新疆分公司网络优化专业利用“MR 分析系统”的小区级 MR 实时数据或利用新疆云化 FAST 平台输出经过 OTT 定位+指纹算法定位 MDT 数据进行网络评估分析。

2.1 评估主要参数

MDT (minimization of Drive-Test), 简称最小化路测,是 3GPP 在 R10 版本中引入的一种通过网络配置对用户终端进行测量数据采集、上报的自动化路测技术。其关键点在于用户终端开启 GNSS (如 GPS) 并支持 MDT 功能,终端就能向基站上报包含用户位置信息的测量数据。MDT 的数据采集可用于大数据分析以减少路测开销,缩短优化周期,分析用户感知和网络容量等问题,从而降低移动通信运营商网络优化和维护成本^[2]。

(1) 平均覆盖电平 AVG_RSRP

$$AVG_RSRP = \frac{\sum_{k=0}^n RSRP_k}{n} \quad (n \text{ 为总采样点数}) \quad (1)$$

(2) volte 覆盖率 RSRP ≥ -112 占比

$$\text{基于 } RSRP \geq -112 \text{ 占比} = \frac{\sum_{k=0}^n f(RSRQ_k \geq -112)}{n} \quad (n \text{ 为}$$

总采样点数, $f(RSRQ_k \geq -112)$ 当 $\geq -112\text{dB}$ 时返回计数 1, $\geq -112\text{dB}$ 采样点数)

(3) 重叠覆盖率

重叠覆盖率

$$AVG_GRID110_OVERLAPDU6 = \frac{\text{重叠覆盖度} \geq 3 \text{ 的采样点数}}{\text{总采样点数}} \quad (2)$$

(其中重叠覆盖度:服务小区 RSRP 大于 -110dBm,主服务小区和相邻小区电平差值小于 6dB 的相邻小区个数);

(4) MOD3 干扰比例

$$MOD3rate = \frac{MOD3 \text{ 干扰采样点数}}{\text{总采样点数}} \quad (3)$$

(只要有一个满足以上条件的邻区的 PCI 和服务小区的 PCI 模 3 结果相等,即记为一个 MOD3 干扰采样点);

2.2 数据分析处理

全疆 MR 数据的每天约产生 280G 的文件, MRO/MRS 的 XML 文件的解析目前由现有大数据平台完成处理(解析的方法这里不做阐述)。为结合集团场景化管控及政企市场需要灵活区域分析的需要,新疆网优团队利用 QGIS 对全疆地州/市/县级/及口碑/重点/一般等 78 类场景存储入

PostgreSQL 数据库管理,利用 PostGIS 空间数据引擎,进行空间对象存储维护管理。

由于城市化推进,城区非城区及场景边框的更新需要及时调整,而目前沃运维中“MR 分析系统”,暂时无法及时更新,另外对省以下的地州及划小单元分析处理不理想。为解决以上问题,特引入地理空间计算,来对小区归属,区域 MDT 统计计算,为基层划小单元的网络数据进行区分提供了解决方案。如图 4 所示,对网络侧大数据进行经纬度关联后,生成对应的 geometry 属性,并通过 Gepandas 的 clip() 函数和 overlay() 函数(内置了 scipy 的 cKDtree 算法),可以快捷高效的计算对应的地理区域内外的归属和非归属点信息。

```
gpd.clip(gdf_urumqiMRCcell_poi, gdf_cityX_polygon)
gpd.overlay(gdf_urumqiMRCcell_poi, gdf_cityX_polygon, how='symmetric_difference')
```

图 4 Geopandas 地理区域归属实现命令

对 MDT 进行边框区域 polygon 归属后,即可以利用 PYTHON 进行统计数据计算,输出“区域 MDT 样本结合地理空间计算后 LTE 覆盖可视化效果”。主要实现代码如图 5 所示。

```
## 插入覆盖百分比图
left, bottom, width, height = 0.10, -0.05, 1, 1
ax15 = inset_axes(ax, width='10%', height='20%', loc='upper left',
                 bbox_to_anchor=(left, bottom, width, height),
                 bbox_transform=ax.transAxes)
df_value_counts = llius_mr_tools().hbarRSRP_addColorInfo(gdf, colName_str)
## 提取“汇总排序后的RSRP信息”df
y = df_value_counts['labels']
x = df_value_counts['ratio']
for n, m in zip(y, x):
    ax15.text(m, n, '{}%'.format(round(m * 100, 2)), fontproperties=zhfont1, ha='left', va='center', fontsize=15, color='k')

ax15.barh(y, x, color=df_value_counts['color'].tolist())
ax15.set_title("栅格RSRP[单位:dB]占比分布", fontsize=20, fontproperties=zhfont1)
#####
gdf_zoneurumqi_polygon.plot(ax=ax, facecolor='f0f0f4', edgecolor='k', linewidth=2, alpha=0.5)
gdf_cityroad_polyline_clip.plot(ax=ax, edgecolor='k', linewidth=0.1)
gdf.plot(ax=ax, column=colName_str, markersize=0.5, cmap=cmap, norm=norm)
```

图 5 场景 MDT 覆盖评估样本实现主要代码

特别说明,我们对比了使用“算数平均”和“加权平均”两种方法计算全网平均 RSRP 和基于 RSRP > -112 占比值,结果不同,如表 4 所示。由于 MRS 统计特性,在小区级 MR 统计数据中,每个小区的平均 RSRP 是本小区所有采样点的平均值,如果要对某个区域进行计算统计,需要按照加权平均的方式,用每个小区的均值乘以小区采样点数,再除区域总采样点。根据表 4 数据分析,MDT 数据量仅站全网 MR 数据的 1.5%左右,故评估全网或较大区域覆盖效果时建议采样全量 MR (或已对 MR 进行定位处理的近似全量 MDT) 类型统计,并使用加权平均的方式计算统计相关指标。

表 4 区域内 MR/MDT 算数平均及加权平均计算对比

类型	MR 总采样点	总 MDT 采样点	平均 RSRP	基于 RSRP ≥ -112 占比
MR	5.66 亿	0.09 亿	-92.18	95.88%
			平均 RSRP _a	基于 RSRP ≥ -112 占比 _a
			-91.91	95.07%
类型	总 MDT 采样点		平均 RSRP _w	RSRP 大于等于 -112 占比 _w
MDT	0.10 亿		-86.21	97.91%
			平均 RSRP _a	RSRP 大于等于 -112 占比 _a
			-89.73	96.18%

由于原始 MDT 数量仅占整个 MR 样本量的 2%左右, 占比极低。而现有的“MR 分析系统”没有进行全量 MR 匹配经纬度算法, 故新疆联通采用诺基亚 FAST 平台 10*10m 栅格数据, 利用其指纹库的神经网络定位算法, 对全量 MR 样本进行经纬度定位。

MR 数据进行精准定位栅格统计后, 与客服投诉平台的投诉点 GIS 数据结合, 分析判断出投诉区域覆盖状况, 并给出处理建议。如图 6, 利用客服投诉数据锁定大致区域, 并将此区域的 MR 数据与 KPI 数据中 MOS 值, 丢包率进行关联, 从而定性判断问题是否由无线侧覆盖导致, 对覆盖问题区域可视化呈现, 有效支撑客户投诉分析, 促进问题解决闭环;

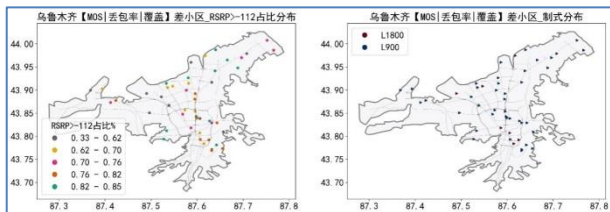


图 6 MR 数据与 KPI 数据进行关联搜寻问题小区

2.3 小区级栅格级指标和质差小区关联情况

将 MR 数据与 KPI 数据关联后, 定性为无线侧的覆盖问题导致。50 米栅格无法判断出小区“重叠覆盖/Mod3 覆盖”, 可以通过 FAST 小区级栅格导出单个基站“重叠覆盖/Mod3 覆盖”情况, 进行可视化判断。如下图 7 所示基站的三个扇区问题栅格可视图。

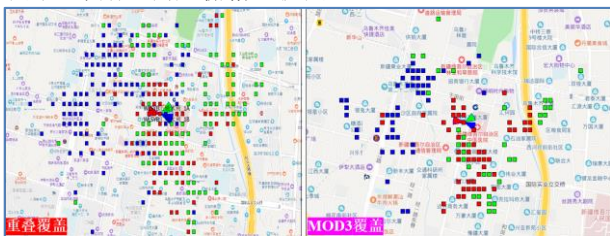


图 7 栅格重叠/Mod3 覆盖分布呈现示意图

重叠覆盖优化思路之一是突出主覆盖小区信号强度, 降低相邻小区干扰信号电平。结合栅格可视化图层若地图上呈现两个或多个小区直线距离为 100 米左右的, 且小区间电平相近的情况则可判断为站点过近, 容易出现小区间干扰; 若直线距离为 600 米以上的, 且小区间电平较差的情况则可判断为站点过远, 容易出现因站点稀疏而弱信号。以上情况可通过天馈调整或站点搬迁方法来处理。

2.4 效果评估

对区域内 4G 基站情况进行评估, 一是梳理四超站址(超远站、超近站、超高站、超重叠覆盖站)利用 MR 数据可视化, 通过天馈整改、共享友商站址等手段进行优化。二是优化 PCI, 重点优化存在 Mod3 干扰的重叠覆盖区域, 提升网络质量。三是梳理站点邻区关系, 核查邻区漏配、垃圾邻区等问题, 调整切换序列, 优化网络结构。通过

MR 数据和质差小区进行关联处理后, 分析质差小区和 MR 关联情况是 234, 通过 fast 建议解决质差小区是 217 个, 解决率是 92.27%。

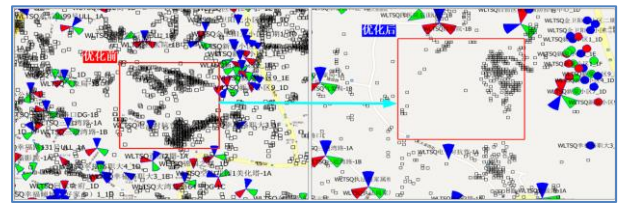


图 8 调整前后区域 Mod3 优化效果可视化呈现

Mod3 问题处理前后: 如图 8 所示, 在东大梁西街周边位置, 根据 MR 栅格分析发现, Mod3 干扰严重。解决措施: (1) WLTSQ 蓝波湾-1A 与 WLTSQ 蓝波湾-1B 的 PCI 互换, 并调整 WLTSQ 蓝波湾-1A 的机械下倾角。(2) 后台核查 WLTSQ 四医院-1C, WLTSQ 四医院-1B 的 PCI 配置是否合理, 建议更改 PCI。处理效果: 将 WLTSQ 蓝波湾-1A 与 WLTSQ 蓝波湾-1B 的 PCI 互换后, Mod3 干扰减少。

对比共建共享前后效果, 覆盖指标改善明显, 平均 RSRP 提高了 1.31dB, RSRP 大于-110 占比提升 5%。如表 5 所示:

表 5 共建共享前后基于 MR 大数据关联分析对比表

	4月第二周 (4G共享前)		5月第二周 (4G共享后)					
	全量采样 点数	占比	1800M采 样点数	占比	全量采样 点数	占比	1800M采 样点数	占比
总采样点数	1.27E+08		87896279		1.31E+08		89298662	
平均RSRP	-94.88		-93.23		-93.56		-92.77	
[-100,0)	82847515	65.28%	62490668	71.10%	93548964	71.50%	65654006	73.52%
[-105,-100)	17266683	13.61%	1082879	12.61%	18153419	13.88%	12305169	13.78%
[-110,-105)	3083025	10.31%	7674823	8.73%	16366689	8.99%	7862891	8.81%
[-112,-110)	3862531	3.04%	2024461	2.30%	2288177	1.75%	1182236	1.32%
[-115,-112)	4660585	3.67%	2380533	2.71%	2451738	1.87%	1201560	1.35%
[-120,-115)	4051437	3.19%	1903171	2.17%	2045989	1.56%	887301	0.99%
[-150,-120)	1130351	0.89%	339743	0.39%	705258	0.54%	205499	0.23%
重叠覆盖	2737540	2.16%	2036481	2.32%	1673032	1.28%	1315880	1.47%
Mod3干扰	7066498	5.57%	5776643	6.57%	6397309	4.89%	5262933	5.89%
MR覆盖率	1.13E+08	89.20%	81248371	92.44%	1.23E+08	94.27%	85822066	96.11%

3 与传统方法比较

为提升精准优化水平和规划投资效率, 共建共享网络评估需结合高流量, 高用户, 高投诉等进行 MR 大数据关联分析, 实现高价值区域评估。而基于 MR 大数据分析紧跟规划变化, 及时跟踪分析已共享 4G 区域的网络覆盖情况。基于 MR 大数据与空间计算的 LTE 网络可视化评估应用, 其是利用开源原件搭建的轻量级应用服务, 使无线网络数据可视化, 处理耗时短, 区域分析灵活, 可拓展性好。

基于 AI 聚类算法+空间地理大数据计算, 助力电联深度共建共享: 投入 2 个月时间, 实现了自动化输出 TAC 调整基站数据, 自动核查 TAC 插花基站; 高效完成了共建共享 TAC 调整工作。彻底避免了共享载波方式 TAC 插花问题, TAC 拉齐的同时启用了 TAlist 功能, 割接后 TAU 次数有明显下降, 降低了信令负荷, TAU 成功率割接前后基本保持平稳。

4 结束语

在数字化互联网化转型大背景下, 新疆联通云网运营中心联合新疆电信无线网络优化中心探索基于 MR 大数据

与空间计算的 LTE 网络深度共建共享规划与评估应用,可以快速规划核查双方 TAC 对应关系,评估共建共享前后网络覆盖改善情况,也能满足政企市场客户快速评估小区域网络质量的要求,精准可视化网络问题,具有可推广价值。面对今后 5G 网络覆盖,要以高价值和精品网络提升为导向,对于快速精确打造重点区域竞争优势,有效提升网络质量与用户感知,为解决客观及可视化评估网络问题提供创新思路,为打造 VoLTE 精品网络和高价值 5G 网络提供更多可借鉴的手段。

[参考文献]

- [1]田桂宾,冯胜杰,姬刚.基于 MR 数据的 LTE 网络规划及结构优化[J].邮电设计技术,2019(3):52-55.
- [2]吕非彼,杨洁艳,吴浩然,等.图像识别技术在 5G 网络覆盖率评估中的应用探讨[J].邮电设计技术,2020(6):40-44.
- [3]3GPP. TR 25.912 V9.0.0 Feasibility study for

evolved Universal Terrestrial Radio Access (UTRA) and Universal Terrestrial Radio Access Network (UTRAN) (Rel-9) [S]. 2009-12. 44-49

- [4]范晶晶,王力,褚文博,等.基于 KDTree 树和欧式聚类的越野环境下行人识别的研究[J].汽车工程,2019(12):1.
 - [5]张艳芹,杨习贝,陈向坚.基于 kd 树的快速邻域分类方法[J].《江苏科技大学学报(自然科学版)》,2018(12):98.
 - [6]刘韬,吴涛,王斌.XDR 与 MR 数据关联与应用[J].电信科学,2019(4):120-130.
 - [7]官元峰,黄轶.基于大数据分析的室内深度覆盖优化方法研究[J].电信科学,2019(5):149-154.
 - [8]曹丽娟,程新洲,徐乐西,等.基于机器学习算法的 LTE 高投诉小区预判方法[J].邮电设计技术,2020(5):18-21.
- 作者简介:王磊(1983-10),男,新疆昌吉人,汉族,部门主任、大学本科。