

基于数据挖掘的5GC网络质差根因分析定位系统的设计与实现

Design and Implementation of Root Cause Analysis and Location System for 5GC Network Poor Quality Based on Data Mining

杨士军,于平苹,谢绍富(中国联通山东分公司,山东 济南 250101)

Yang Shijun, Yu Pingping, Xie Shaofu (China Unicom Shandong Branch, Jinan 250101, China)

摘要:

为打造中国联通5G网络差异化竞争优势,推进大区集约化网络运营能力,北部大区运营中心自主设计开发5GC网络质差根因分析定位系统,挖掘设备网管、专业分析工具、信令分析系统数据,通过对相关数据的聚类分析,实现网络KPI波动告警和劣化指标异常原因定界定位,结合经验库将携带定位原因和处理建议的告警精准发送给一线维护人员,及时高效完成指标优化,不断提升用户感知和网络质量。

关键词:

5GC;数据挖掘;聚类分析;KPI波动;定界定位
doi:10.12045/j.issn.1007-3043.2022.01.014
文章编号:1007-3043(2022)01-0073-04
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:

In order to build the differentiated competitive advantage of China Unicom 5G network and accelerate the regional intensive network operation capability, the northern regional operation center independently designs and develops 5GC network poor quality root cause analysis and positioning system, which mines the data of equipment network management, professional analysis tools and signaling analysis system. Through the cluster analysis of relevant data, the demarcation and positioning of abnormal causes of network KPI fluctuation alarm and deterioration index is realized, and the alarm with positioning causes and treatment suggestions is accurately sent to the front-line maintenance personnel in combination with the experience base. So the index optimization is completed timely and efficiently, and the user perception and network quality is improved continuously.

Keywords:

5GC; Data mining; Clustering analysis; KPI fluctuations; Delimitation and positioning

引用格式:杨士军,于平苹,谢绍富. 基于数据挖掘的5GC网络质差根因分析定位系统的设计与实现[J]. 邮电设计技术,2022(1):73-76.

1 概述

网络KPI直接关系着网络质量和用户感知,如何快速发现网络指标异常波动并充分利用各方数据和经验库,高效对质差根因进行精准定位和及时处理是网络维护和优化人员一直面临的挑战。基于数据挖掘的5GC网络质差根因分析定位系统首先实时跟踪5GC和IMS域的关键性能KPI,通过完善的算法及时发现质差指标,然后结合各类关联数据进行聚类分析

并匹配经验库完成根因定位,最后派发一线维护人员实施优化。该系统可提高运营效率,是推动核心网数字化转型工作落地的有效途径。

首批确定指标及其定义如表1所示。

2 系统设计与实现

2.1 系统架构

整个系统分为数据挖掘、KPI异常检测、关联数据聚类分析、经验库匹配、与其他系统交互等模块。不但实现指标波动告警、质差根因定位,而且与现网生产系统打通,全面实现告警派单和经验库优化等全流

收稿日期:2021-11-24

表1 关注指标及其定义

网络类型	指标名称	指标定义
5GC	AMF 初始注册成功率 (剔除用户原因)	$[(\text{初始注册成功次数} + \text{初始注册失败次数}_{\text{非法用户}} + \text{初始注册失败次数}_{\text{非法设备}} + \text{初始注册失败次数}_{\text{PEI 不允许}} + \text{初始注册失败次数}_{\text{5GS 服务不允许}}_{\text{用户原因}} + \text{初始注册失败次数}_{\text{跟踪区内无合适小区}}_{\text{用户原因}} + \text{初始注册失败次数}_{\text{N1 模式不允许}}) / \text{初始注册请求次数}] \times 100\%$
5GC	5G SA 会话建立成功率	$[\text{PDU 会话建立成功次数} / \text{PDU 会话建立请求次数}] \times 100\%$
5GC	EPS FB 成功率	$[\text{EPS Fallback 流程成功次数} / \text{EPS Fallback 流程请求次数}] \times 100\%$
VIMS	VoLTE 网络接通率 (剔除用户原因)	$[(\text{S-CSCF LTE 接入主叫应答次数} + \text{S-CSCF LTE 接入用户原因未应答次数}) / (\text{S-CSCF LTE 接入主叫试呼次数})] \times 100\%$, 其中用户原因包括主叫早释、主叫振铃早释、主叫 403 请求禁止、主叫 404 未找到、主叫 408 请求超时、主叫 480 久叫不应、主叫 484 Request-URI 不完整、主叫 486 用户忙、主叫 487 请求终止、主叫 600 用户忙、主叫 603 用户拒接、主叫 604 用户信息不存在

程闭环管理。

2.2 逻辑设计

质差根因分析系统整体设计逻辑如图1所示,所有数据均以网元为单位每15 min提取分析,KPI检测规则支持灵活设置,检测时间可分时段,时间粒度可以为15 min或1 h。

2.2.1 数据采集入库

数据源包括设备专业网管中原始性能统计数据,厂家专业工具(华为Discovery等)中协议失败原因、次数及用户明细,端到端系统中的无线小区和IMEI等信息。如图2所示,定位系统通过专用接口完成这3类数据的采集,然后解析入库做进一步的汇总分析。

为便于后续及时精准分析,对采集入库的数据有如下要求。

- a) 原始数据必须齐全:指标涉及尝试、成功次数,

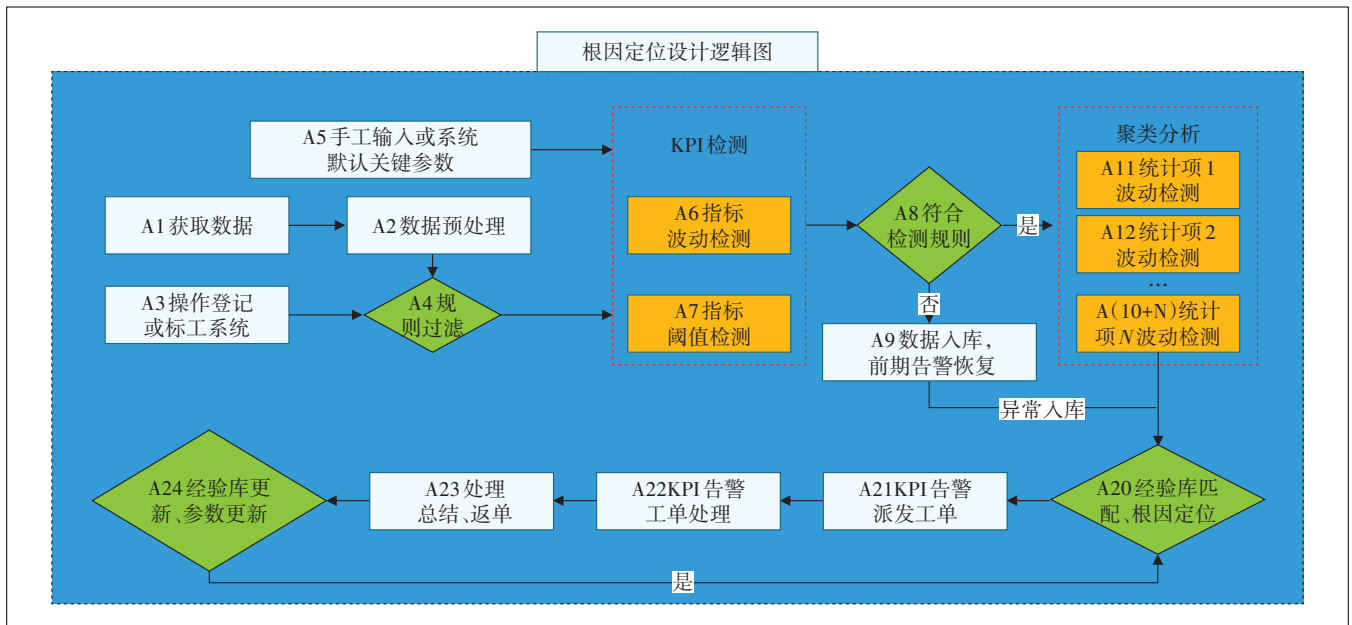


图1 根因定位系统设计逻辑图

各类原因失败次数等原始统计项均需上报,以便对各类统计项进行波动分析,可以有针对性地发现问题。

b) 测量对象最小原则:网管上有关KPI指标的测量对象齐全,并且细分统计到支持的最小粒度(如TAC),以便进行精准定位。

c) 专业工具和端到端系统数据根据分析要求提供。

2.2.2 数据预处理

获取到原始数据后的处理规则如下。

- a) 数据完整性检查:如有网元、指标项统计数据

缺失,系统会在告警界面发出告警。

b) 异常数据剔除:该系统会删除测量结果为0的统计项。

2.2.3 操作网元数据处理

系统与网元操作登记或标工系统对接,获取网元和操作时间,将操作时间段内对应网元的性能统计数据剔除,以免影响指标波动判断。

2.2.4 KPI检测参数

手动输入用于指标检测的阈值,如果该环节不输入数据,则使用系统默认值,系统默认将表1中指标的

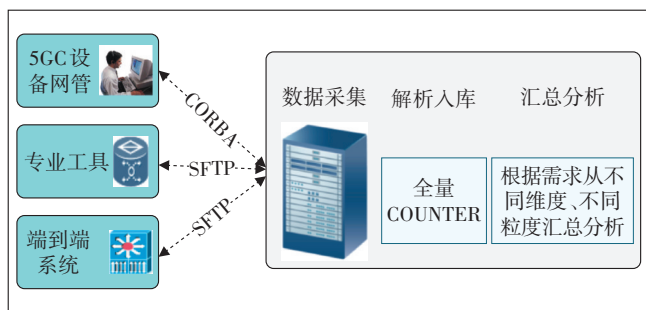


图2 数据采集入库

检测阈值设置为99%。

2.2.5 KPI检测

KPI检测包括指标波动检测和指标阈值检测2类规则,2类规则并行运行,均对单网元指标进行检测。

2.2.5.1 指标波动检测

假设当前需要检测指标为 X_t ,取前31天与 X_t 对应相同时段的KPI时序样本,记为 $X=(x_1, x_2, x_3, \dots, x_n)$,如果前31天数据不足,则按天向前顺延,直到取到31个样本为止。

指标波动检测采用 3σ 算法,具体算法步骤如下。

a) 均值计算,均值 $\mu=(X_1+X_2+\dots+X_n)/n$ 。

b) 标准差计算,标准差 $\sigma=\sqrt{\frac{1}{n}[(x_1-\mu)^2+(x_2-\mu)^2+\dots+(x_n-\mu)^2]}$ 。

c) 异常判断,判断规则如图3所示。

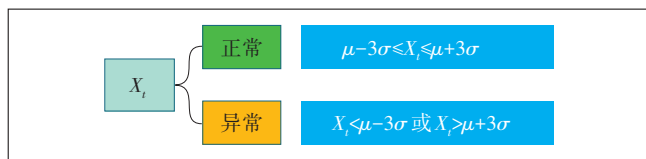


图3 3σ 算法检测规则

2.2.5.2 指标阈值检测

根据A5环节输入或系统默认值对指标进行检测,低于阈值即判断为异常。

2.2.6 符合检测规则

只要在KPI检测的A6或A7环节(见图1)被检测出异常,都将进入聚类分析环节,根据更多数据做进一步的分析和定位。

2.2.7 数据入库

KPI检测正常和异常的数据都将做入库处理,并且根据新入库数据对之前告警做是否恢复的判断。

a) 正常指标入库:A6、A7环节未被检测出异常,则入正常库,作为 X 样本在后续使用。入库字段包括但不限于省、网络类型、网元、时间、指标。

b) 异常指标入库:检测出异常,则入异常库,入库字段包括但不限于省、网络类型、网元、告警时间、指标、波动情况、统计子项波动情况。

c) 前期告警恢复:如果某指标入正常库,而异常库中有该指标之前时间的异常告警,则生成之前异常告警的恢复告警,输出字段包括省、网络类型、网元、指标、告警时间、指标波动情况、恢复时间、恢复后指标值等。

2.2.8 聚类分析

针对某项指标从省分、网元和跟踪区维度对各类失败码、失败用户/终端、无线区域涉及失败次数进行占比分析,对各类失败次数逐项与前7个相同时段的平均值做比较,波动超过50%为异常。

通过多维度细化聚类分析和对比,实现质差问题定位。

2.2.9 经验库匹配

将原因明确或者经过回单验证过的根因和处理方法逐一汇总到经验库,并通过智能算法对回单结果进行分析处理,不断优化更新经验库,经验库格式如表2所示,后续不断补充完善。

表2 经验库格式

省分	网元类型	指标	异常情况	TAI失败情况	用户/终端失败情况	协议原因次数波动情况	根因定位
XX省	AMF等	AMF初始注册成功率(剔除用户原因)	下降	某TAI失败次数较前7日平均值增幅超阈值	个别用户失败次数较前7日平均值增幅超阈值	某类型协议失败原因次数较前7日平均值增幅超阈值	XXX无线区域拥塞

2.2.10 KPI异常波动告警处理

系统根据派单规则和根因定位情况,将KPI指标异常的告警派发到相关单位/人员处理,携带信息主要包括网络类型、网元名称、KPI名称、时间、异常情况、波动子项信息、根因定位、建议处理步骤等。接单人员进行KPI异常波动告警工单处理,直至告警恢复。告警恢复后,总结处理步骤,按照固定格式回单。

2.2.11 更新经验库

系统支持根据回单结果对经验库中的根因定位进行更新,包括对指标检测阈值进行调整优化,不断提升系统告警有效性和根因定位准确性。

a) 经验库更新:系统根据关键字段对回单内容进行机器学习建模,自动分析学习回单内容,并输出关键词得到具体原因和处理步骤,并更新经验库。系统支持根据每次的回单内容及分词结果评估情况进

行自优化。

b) 如果KPI属于正常波动,则回单时提出算法更新和告警阈值参数设置调整建议。

2.3 系统优势

该系统的优势如下。

a) 根据自定义算法实现指标波动告警,支持全国指标排名波动告警。

b) 根据TAI归属对指标进行细化,实现地(市)和TAI粒度指标统计。支持网元、用户号码、小区等多维度失败次数统计分析,精准发现网络中存在的局部问题,增强监控灵敏度。

c) 根据经验库实现网络质差根因定位,精准高效支撑网络优化、提升。

3 取得效果

该系统基于多维度对网络指标中的失败次数做波动分析,不但可以通过及时定位解决网元、无线区域甚至用户/终端级别的问题来优化指标,而且可以强化端网业匹配,提升网络运营效能。

a) 第一时间发现解决无线网络和异常终端等问题。某省AMF初始注册成功率出现突降,从协议原因失败次数上发现52+UE无响应占比最高,无法进行根因定位,如图4所示,再从TAI的维度做失败次数排列,则可以可以看出TAI 4600133XXXX明显异常,该TAI下的失败次数占到了总失败次数的65%,其他TAI变化不大,联系网优确认为该TAI无线拥塞引起。

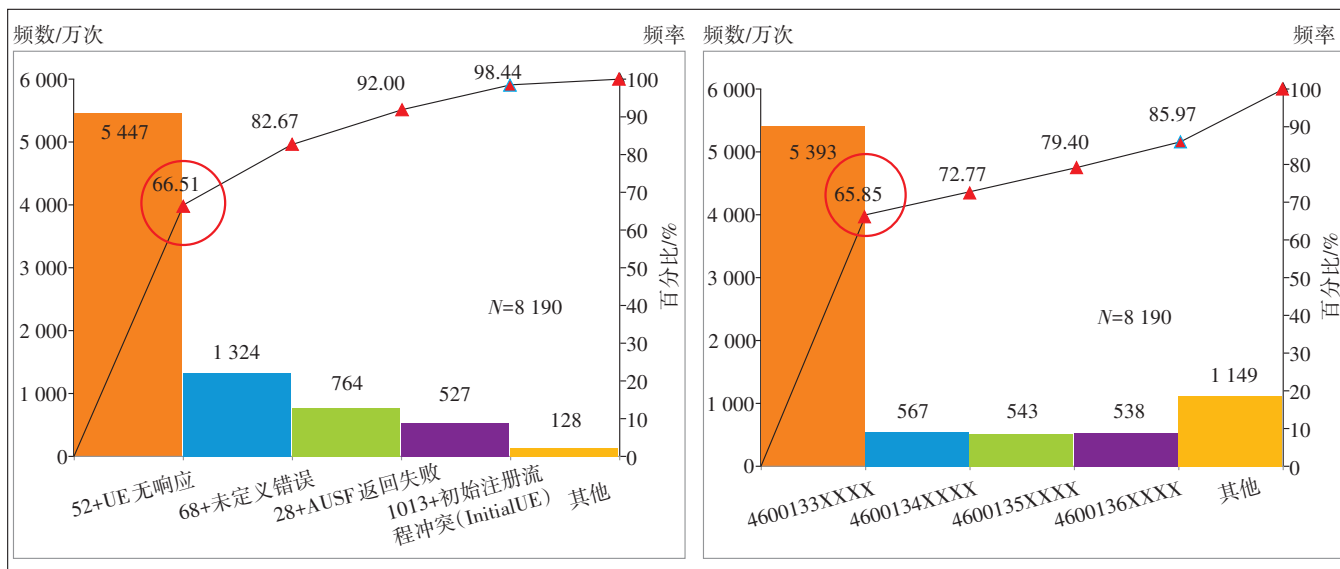


图4 协议原因失败次数和TAI失败次数排列图

b) 根据终端、无线区域用户分布和指标情况加强端网业分析,指导建网和无线优化。

4 结束语

基于数据挖掘的5GC网络质差根因分析定位系统给出了精准有效的指标波动预警算法,并通过对各类相关数据的聚类分析和派单等相关流程的设计,实现了KPI异常波动的精准定位和快速处理,提升了网络质量和用户感知。后续在逐步完善优化功能的基础上,对其他网元指标进行迭代,不断提升网络数字化运营能力。

参考文献:

[1] 张勉知,叶晓斌,程亚锋,等. 机器学习在网络性能预测中的应用

[J]. 邮电设计技术,2021(8):25-28.

[2] 任心怡,金鑫,郑圣,等. 基于机器学习的VoLTE核心网KPI异常检测方法[J]. 江苏通信,2021,37(4):31-35.
 [3] 郭正,郭宁,黄蕴思. 基于多维度数据挖掘的自学习故障根因定位系统[J]. 电子技术与软件工程,2021(15):146-149.
 [4] 曹志英,曹伟,张秀国. 基于多维度特征分析的KPI异常检测[J]. 计算机工程与设计,2021,42(5):1439-1447.
 [5] 张国光,赵占强,许国平,等. 5G网络维护自动化及优化智能化研究[J]. 邮电设计技术,2020(7):52-57.

作者简介:

杨士军,毕业于西安邮电学院,高级工程师,硕士,主要研究方向为移动核心网维护、大区支撑系统开发和网络优化;于平苹,毕业于聊城大学,高级工程师,主要研究方向为移动网端到端优化,大数据分析和项目开发;谢绍富,毕业于山东大学,高级工程师,硕士,主要研究方向为核心网和无线网。