

# AI智能运维在5GC SA网络中的应用研究

## Application Research of AI Technology in 5GC SA Network Operation and Maintenance

张勉知,刘惜吾,叶晓斌,姚丽红,程亚锋,马丹丹(中国联通广东分公司,广东广州510627)  
Zhang Mianzhi, Liu Xiwu, Ye Xiaobin, Yao Lihong, Cheng Yafeng, Ma Dandan (China Unicom Guangdong Branch, Guangzhou 510627, China)

### 摘要:

5G带来通信网络能力提升的同时,全新的网络形态和网络规模也给运维工作带来了极大的挑战,而变化最大的5G核心网,由传统CT网络演变成NFV+SDN网络,传统的OSS工具已难以支撑。随着大数据和人工智能的发展,采用AI技术实现智能运维成为解决5GC网络监控、隐患发现、故障定位/修复的主要手段。重点探讨AI算法在5GC网络指标异常检测、KPI-告警关联中的应用。

### 关键词:

智能运维; AI技术; 5GC; 异常检测; KPI-告警关联

doi: 10.12045/j.issn.1007-3043.2020.10.010

文章编号: 1007-3043(2020)10-0047-04

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:

While 5G brings about the improvement of communication network capability, the brand new network form and network scale also bring great challenges to the operation and maintenance work. 5G core network, which has changed the most, has evolved from the traditional CT network to the NFV+SDN network, and the traditional OSS tools are no longer able to support it. With the development of big data and artificial intelligence, adopting AI technology to realize intelligent operation and maintenance has become the main means to solve 5GC network monitoring, hidden trouble discovery, and fault location/repair. It will focus on the application research of AI algorithm in 5GC network indicator anomaly detection and KPI-alarm association.

### Keywords:

Intelligence operations; AI technology; 5GC; Anomaly detection; KPI- Alarm correlation

引用格式: 张勉知,刘惜吾,叶晓斌,等. AI智能运维在5GC SA网络中的应用研究[J]. 邮电设计技术, 2020(10): 47-50.

## 1 概述

5G网络正式商用,可提供大容量、可靠性、稳定性、灵活性业务。由于5G通信云采用微服务+云化架构,业务能力空前敏捷,但同时IT+CT的架构带来维护难度提升。传统人工运维模式面临网络监控/管理困难、故障定界定位复杂、运维效率低下、无法有效避免隐患发生等多个难题。

针对上述运维难点,广东联通聚焦5GC智能运维解决方案的研究,采用智能化手段提高网络自治能

力,更及时、准确和自动地发现网络的异常问题。本文重点阐述了基于多种AI算法的5GC告警、性能、配置数据自学习自挖掘,从而实现指标异常检测、KPI-告警关联分析,最终辅助运维人员实现网络侧的自动驾驶。

## 2 5G运维痛点

从网络规模上看,根据全球移动通信协会发布的报告显示,到2025年,全球5G用户规模预计达到13.6亿,其中,中国5G用户总量4.54亿,全球居首。无论从投资端、需求端还是产值来看,5G通信产业将迎来爆发式增长,三大运营商今年早些时候公布的投资计划

收稿日期: 2020-09-01

显示,中国移动、中国联通和中国电信2020年计划建设50万+个5G基站。

从网络架构上看,5G网络是全面虚拟化、云化的网络,在通用硬件之上的虚拟化层提供了灵活的资源编排和调度能力,为上层应用的部署和功能扩展提供了极大的灵活性。为了支撑这些重要特性的实现,5G网络也引入了更多的网元和接口。

可以看出,全新的网络形态和网络规模为5G网络运维工作带来了极大的挑战。随着管理对象的日益增多,网络可能产生的故障大幅增加,人工监控手段已无法满足运维需求。近几年随着大数据和人工智能的发展,AI辅助的智能运维成为解决5G网络时代运维工作努力的方向。通过引入AI技术,实现5GC网络黄金指标异常检测、KPI-告警关联,可帮助运维人员及时发现潜在的隐患、规避网络故障的发生,并在故障发生时,快速定界定位,缩短修复时间。

### 3 AI技术在5GC智能运维中的应用研究

#### 3.1 性能指标异常检测

图1为5G核心网指标异常检测的整体流程,主要分为数据清洗和异常检测两大块。

**数据清洗:**输入的指标数据往往存在粒度不一致,数据缺失等问题,为了保障后续算法的可靠性,需要对数据做预处理。

**异常检测:**基于数理统计将指标波形做划分,再针对不同波形采取相应的异常点检测方法,保障准确率和实效性。

##### 3.1.1 数据清洗

由于目前是5G建网初期,原始数据存在粒度不

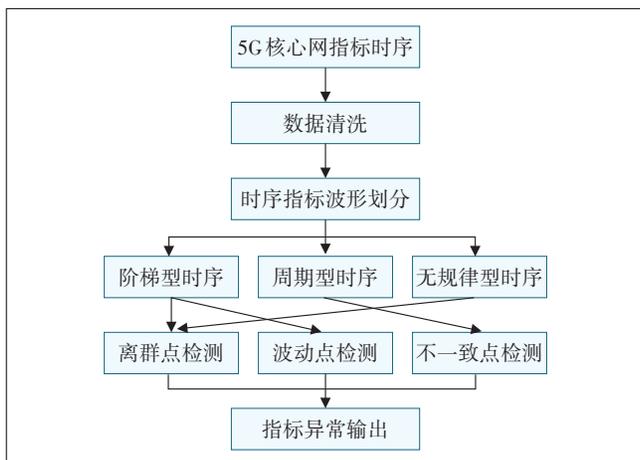


图1 5G核心网指标异常检测

一,粒度缺失等数据问题,所以在应用算法之前需要对数据进行有效的清洗预处理,保证数据质量,排除干扰。主要是对输入数据“时间列”和“值列”分别处理的过程,具体流程如图2所示。

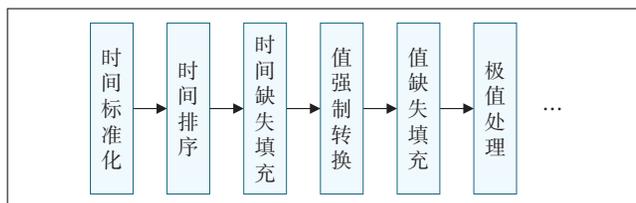


图2 指标数据清洗

a) 时间标准化:将时间列统一转换为“yyyy-mm-dd HH:MM:SS”格式。

b) 时间排序:将输入数据按照时间列从小到大排序。

c) 时间缺失填充:时间缺失判断(相邻2条数据时间列做差值,差值结果不一样,则存在时间缺失);时间缺失是否需要填充,不同异常检测算法要求不同。

d) 值强制转换:将输入数据值列强制转换为数值型,有缺失或字符的统一转换为NA。

e) 值缺失填充:是否需要填充,不同异常检测算法要求不同,基于统计的方法可以不用填充,基于时序拟合的算法需要填充;对于无规律的波形,使用前一个非NA值填充;对于周期型的波形,可以使用同时刻的均值或中位数填充。

f) 极值处理:是否需要极值处理,与算法有关,基于统计的方法不需要极值处理步骤,基于时序拟合的算法需要极值处理步骤;极值处理方法可基于IQR方法设置正常值范围 $[Q_1-3\times IQR, Q_2+3\times IQR]$ ,该范围之外的分别使用 $Q_1-3\times IQR$ 或 $Q_2+3\times IQR$ 进行替换。

##### 3.1.2 异常检测

常用针对性能指标的异常检测算法主要有3大类。

a) 时序分析类异常检测:通过时间序列分析拟合输入的性能指标时序数据,分析拟合误差的分布,给出置信区间阈值,当实际数据超过阈值时,触发异常告警,传统的算法有ARIMA、Holt-winter等,基于深度学习的算法有LSTM及其变体、MTNet等。

b) 假设检验类异常检测:通过时间序列分析拟合输入的性能指标时序数据(可选),对拟合误差做假设检验来判断是否数据存在异常,比如ESD、Seasonal Hybrid ESD(S-H-ESD)等。

c) 机器学习类异常检测:基于异常点检测的机器学习算法来分析数据的空间分布,以此判定是否存在异常点,比如 Isolation Forest、one-class SVM 等。

上面的3类方法具有较好的通用性,有现成的算法包,可以快速在新的数据集上验证测试,取得较好的效果,但针对特定场景的优化,满足特定场景的异常检测准确率,需要在一定程度上做算法的改造以及组合。同时上述的三大类方案,在大规模数据场景下的应用,往往资源消耗较大,而且耗时较长。

针对核心网指标的异常检测,通过对数据特征的分析,采用统计分析的思路提出 OCDT 算法,在满足准确率的同时,可以很好地保障时效性。算法流程如下:

a) 基于统计分析对时序波形做划分,主要分为阶梯型、周期型和无规律型。

(a) 阶梯型:阶梯型数据特征为连续一段时间内取值相同,通过差分方式判定区分。阶梯型数据实例如图3所示。



图3 阶梯型数据实例

(b) 周期型:周期型数据特征为每天数据变化趋势具有较为明显的相似性,通过计算任意相邻2天子序列(每天数据作为一个子序列)的相似度大小,设置相似度阈值范围,实现周期型波形的判定区分。周期型数据实例如图4所示。

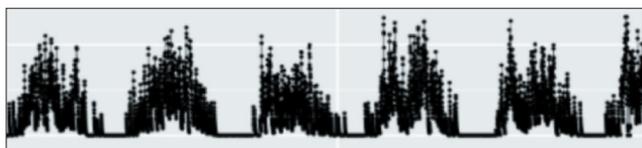


图4 周期型数据实例

(c) 无规律型:数据特征为随机变动,不具有规律性,通过排除法(非周期型或阶梯形)判定区分。

b) 针对不同的时序波形,采取对应的异常点检测方案。阶梯型时序采取离群点和波动点检测,周期型时序采取不一致点检测,无规律型采取离群点检测。

(a) 离群点检测:使用 IQR 和  $3\text{-}\sigma$  相结合的方式。计算  $Q_1$  分位数、mean 均值、 $Q_3$  分位数和方差 sd; 设置上限为  $\max(Q_3+3\times(Q_3-Q_1), \text{mean}+3\times\text{sd})$ , 下限为  $\max(Q_1-$

$3\times(Q_3-Q_1), 0)$ ; 超出上下限范围标记为异常点。

(b) 波动点检测:对任一点  $x$ , 计算  $x$  与之前  $n$  个点的平均变化量 ( $n$  为指定参数), 平均变化量 =  $(n \times X_t - X_{t-1} - X_{t-2} - \dots - X_{t-n}) / n$ 。计算所有点的平均变化量  $Q_1$ 、 $Q_3$  分位数。设置变化量上限 =  $\max(Q_3+3\times(Q_3-Q_1))$ , 超出上限标记为异常点。

(c) 不一致点检测:将所有数据按分钟分组统计, 计算每组数据的计算  $Q_1$  分位数、mean 均值、 $Q_3$  分位数和方差 sd。设置上限为  $\max(Q_3+3\times(Q_3-Q_1), \text{mean}+3\times\text{sd})$ , 下限为  $\max(Q_1-3\times(Q_3-Q_1), 0)$ , 超限标记为异常点。

### 3.2 KPI-告警关联

通过 AI 算法将系统历史告警数据和对应时间内的 KPI 性能指标异动情况进行关联挖掘, 自动发现故障和指标裂化之间的相关性, 当实时性能指标出现裂化点时, 及时对业务故障进行提前预警。告警关联分析的整体流程主要分为数据特征统计分析和告警关联分析两大块。

数据特征统计分析: 输入指标往往比较多, 告警只会和其中个别指标产生关联, 为了减少后续算法的噪声干扰, 通过数据特征统计分析检测指标和告警之间是否存在关联关系。

告警关联分析: 结合数据的量级和模型的特点, 基于机器学习中的 XGBoost 分类算法挖掘建立筛选后的数据指标和告警之间的关联关系。

#### 3.2.1 数据特征统计分析

数据特征统计分析是为了查出各个数据特征与告警有无关系, 如有, 大致判断有什么关系。下面从秩和检验量化分析、特征重要性2方面进行分析。

秩和检验通过将2组数据混合后排序, 将每个序列值在序列中的次序称为秩, 分别将两组数据的秩相加得到两组数据的秩和, 若两组数据同分布则秩和不应过大也不应过小。通过秩和检验来判断告警发生时刻周围数据与未发生告警时刻的数据分布是否相同, 取显著性水平为 0.05, 如果  $p$  值小于 0.05, 则说明故障发生时刻周围数据与未发生故障时刻数据分布不同, 即可将这个特征作为判断故障是否发生的特征之一。

特征重要性根据该特征每次作为分裂点时获得的增益总和除以所有特征的增益总和计算得出。将经过秩和检验筛选后的指标传入 XGBoost 模型进行初步建模后, 输出特征重要性, 根据指标重要性选取一

定量的指标进行再次建模。

$$\text{Gain} = \frac{1}{2} \left[ \frac{\left(\sum_{i \in L_L} g_i\right)^2}{\sum_{i \in L_L} h_i + \lambda} + \frac{\left(\sum_{i \in L_R} g_i\right)^2}{\sum_{i \in L_R} h_i + \lambda} - \frac{\left(\sum_{i \in L} g_i\right)^2}{\sum_{i \in L} h_i + \lambda} \right] - \gamma$$

### 3.2.2 告警关联性分析算法

该问题属于多标签问题,即每个样本数据可能对应多个标签,通过对每一个类告警做一个二分类,判断此类告警是否发生,分类方法通过多轮实验对比后选择XGBoost。XGBoost可以自动处理缺失值,且做了底层优化使得运行速度加快,是现在工程化分类问题上使用的主流算法。将数据预处理好后传入XGBoost,运用交叉验证确定其中一些模型参数,最后使用网格搜索进行调优。

XGBoost属于GBDT,是一种以决策树为基学习器的串行集成算法,目标函数同时考虑了损失函数与树的复杂度:

$$\text{Obj} = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k)$$

$$\text{其中: } \Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2$$

原目标函数经过泰勒公式二阶展开后,求得最优目标函数为:

$$\text{Obj} = -\frac{1}{2} \sum_{j=1}^T \frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T$$

每次分裂获得的增益为:

$$\text{Gain} = \frac{1}{2} \left[ \frac{\left(\sum_{i \in L_L} g_i\right)^2}{\sum_{i \in L_L} h_i + \lambda} + \frac{\left(\sum_{i \in L_R} g_i\right)^2}{\sum_{i \in L_R} h_i + \lambda} - \frac{\left(\sum_{i \in L} g_i\right)^2}{\sum_{i \in L} h_i + \lambda} \right] - \gamma$$

每次选取的分裂点为增益最大的点,即:

$$\max \left( \frac{\left(\sum_{i \in L_L} g_i\right)^2}{\sum_{i \in L_L} h_i + \lambda} + \frac{\left(\sum_{i \in L_R} g_i\right)^2}{\sum_{i \in L_R} h_i + \lambda} - \frac{\left(\sum_{i \in L} g_i\right)^2}{\sum_{i \in L} h_i + \lambda} \right)$$

## 4 总结

针对5G网络运维难题,本文介绍了广东联通基于AI技术的智能运维解决方案,实现性能指标的异常检测及KPI-告警关联分析,为中国联通各个大区的5G网络运维树立了示范标杆。在性能指标异常检测中,采用统计分析的思路提出OCDT算法,针对不同时序波形采用不同的检测手段,在满足准确率的同时,可以很好的保障时效性。在KPI-告警关联分析中,广

东联通对KPI与告警关联预测进行了算法验证,将告警与性能指标关联,通过对性能指标的精准预测,进而提前发现故障隐患,提前告知、提前处置,避免故障发生。

### 参考文献:

- [1] CONTRERAS J, ESPINOLA R, NOGALES F J, et al. ARIMA Models to Predict Next-Day Electricity Prices [J]. IEEE Transactions on Power Systems, 2003, 18(3): 1014-1020.
- [2] 田中大, 李树江, 王艳红, 等. 基于ARIMA补偿ELM的网络流量预测方法[J]. 信息与控制, 2014, 43(6): 705-710.
- [3] KAYS H M E, KARIM A N M, DAUD M R C. A responsive multiplicative Holt-Winters approach for enhanced forecasting accuracy [J]. International Journal of Industrial and Systems Engineering, 2020, 35.
- [4] GREFF K, SRIVASTAVA R K, JAN KOUTNÍK, et al. LSTM: A Search Space Odyssey [J]. IEEE Transactions on Neural Networks & Learning Systems, 2016, 28(10): 2222-2232.
- [5] 沈兰奔, 武志昊, 纪宇泽, 等. 结合注意力机制与双向LSTM的中文事件检测方法[J]. 中文信息学报, 2019, 33(9).
- [6] HUANG W, STOKES J W. MtNet: A Multi-Task Neural Network for Dynamic Malware Classification [J]. 2016.
- [7] CHANG Y Y, SUN F Y, WU Y H, et al. A Memory-Network Based Solution for Multivariate Time-Series Forecasting [J]. 2018.
- [8] 黄谦, 王震, 韦韬, 等. 基于One-class SVM的实时入侵检测系统 [J]. 计算机工程, 2006, 32(16): 127-129.
- [9] 余翔, 陈国洪, 李霆, 等. 基于孤立森林算法的用电数据异常检测研究 [J]. 信息技术, 2018, 42(12): 96-100.
- [10] 刘高. 基于Spark的孤立森林算法并行化研究 [D]. 武汉: 华中科技大学, 2019.
- [11] HOCHENBAUM J, VALLIS O S, KEJARIWAL A. Automatic Anomaly Detection in the Cloud Via Statistical Learning [J]. 2017.
- [12] LIU F T, TING K M, ZHOU Z H. Isolation-Based Anomaly Detection [J]. Acm Transactions on Knowledge Discovery from Data, 2012, 6(1): 1-39.
- [13] 叶倩怡. 基于Xgboost方法的实体零售业销售额预测研究 [D]. 南昌: 南昌大学, 2016.
- [14] 庞宇明, 任江涛. 一种基于Xgboost分类算法的文本分类方法: CN 201710060026 [P]. 2017-10-06.
- [15] CHEN T, GUESTRIN C. XGBoost: A Scalable Tree Boosting System [E/OL]. [2020-04-20]. <https://doi.org/10.1145/2939672.2939785>.

### 作者简介:

张勉知, 工程师, 硕士, 主要研究AI在通信网络中的应用; 刘惜吾, 工程师, 硕士, 主要研究AI在通信网络中的应用; 叶晓斌, 工程师, 学士, 主要研究AI在通信网络中的应用; 姚丽红, 工程师, 硕士, 主要研究AI在通信网络中的应用; 程亚锋, 工程师, 硕士, 主要研究AI在通信网络中的应用; 马丹丹, 工程师, 硕士, 主要研究AI在通信网络中的应用。