

基于 Prophet 时序算法的 无线网络突变小区识别方法研究

Research on Recognition Method of Abrupt Cells in Wireless
Network Based on Prophet Sequence Algorithm

杨飞虎¹,许国平²,刘贤松¹,张硕伟¹,郑敏之¹,施亚洲³(1. 中国联通网络 AI 中心,上海 200050;2. 中国联合网络通信集团有限公司,北京 100033;3. 上海联通无线网运营中心,上海 200050)

Yang Feihu¹, Xu Guoping², Liu Xiansong¹, Zhang Shuowei¹, Zheng Minzhi¹, Shi Yazhou³(1. China Unicom Network AI Center, Shanghai 200050, China; 2. China United Network Communications Group Co., Ltd., Beijing 100033, China; 3. China Unicom Shanghai Branch, Shanghai 200050, China)

摘要:

KPI 作为衡量无线网络质量的重要指标,准确及时检测 KPI 指标突变情况非常重要。提出了一种基于 Prophet 时序算法的无线网络突变小区识别方法,利用模型的预测指标得到动态门限,并在动态门限和绝对门限之间设置相对门限,通过实际指标与不同门限的对比,使运维人员更精准地掌握网络性能,降低误报和错报异常,提升网络优化效率和质量。

关键词:

Prophet; 时序算法; KPI; 突变小区

doi: 10.12045/j.issn.1007-3043.2021.02.005

文章编号: 1007-3043(2021)02-025-06

中图分类号: TN929.5

文献标识码: A

开放科学(资源服务)标识码(OSID):



Abstract:

KPI is an important index to measure wireless mobile network. It is especially important to detect the mutation of KPI accurately and timely. It proposes a method of identifying abrupt cell in wireless network based on Prophet timing algorithm. The dynamic threshold is obtained by using the prediction index of the model, and the relative threshold is set between the dynamic threshold and the absolute threshold. Through the comparison of actual indicators with different thresholds, the operation and maintenance staff are able to accurately master the network performance and reduce false alarms and abnormal false alarms, which improves the efficiency and quality of network optimization.

Keywords:

Prophet; Timing algorithm; KPI; Abrupt cells

引用格式: 杨飞虎,许国平,刘贤松,等. 基于 Prophet 时序算法的无线网络突变小区识别方法研究[J]. 邮电设计技术, 2021(2): 25-30.

1 概述

随着 5G 和物联网的引入^[1],移动网络越来越复杂多变,但人们对无线网络的质量要求越来越高,这对运营商的无线网络优化提出了新的挑战^[2]。KPI 指标直接反映了无线网络的性能和质量,传统的小区 KPI 突变检测方法是先采集 PM 和配置等数据,网络优化工程师再根据专业知识和经验,对各个 KPI 指标分别设置阈值,观察 KPI 指标是否连续若干个小时超出阈

值,来判断 KPI 指标是否存在异常^[3]。但是网络制式多样化,用户行为不可控,总会有部分属于正常性突变或者偶发性波动的 KPI 指标作为异常被识别出来,这种方法存在资源投入大、阈值不准确、优化不及时等问题。

2 传统无线网络 KPI 突变小区检测方法

小区 KPI 突变检测是无线网络优化的一个底层核心技术,是改善网络覆盖质量的重要手段,因此小区 KPI 突变检测在无线网络优化工作中具有重要的指导意义^[4]。传统的小区 KPI 突变检测流程如图 1 所示。

收稿日期: 2020-12-25

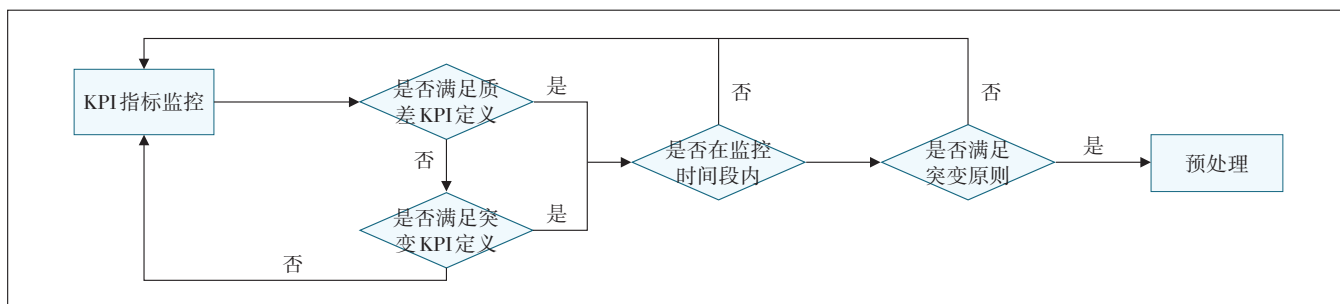


图1 传统的小区KPI突变检测流程图

传统的小区KPI突变检测是对全网各项指标进行日常性监控,并对有问题的小区进行预处理和派单。在监控KPI指标时,除了满足质差小区定义及突变小区定义外,还需要同时满足监控时间条件,如不满足监控时间条件则继续监控指标。通常在各项指标日常监控中,监控时间段为:09:00—24:00(15 h),并且质差或突变条件触发后要同时满足“6 h及以上或连续4 h”,才能触发告警派单。如果人为设定规则过于苛刻,监控系统预处理与派单有严重的滞后性,网络运维人员不能及时处理网络质量或性能问题^[5];如果规则过于宽松,可能出现因为非无线原因、正常潮汐、特殊事件等造成的正常性突变或者偶发性波动被当作KPI指标异常处理的情况。

随着整个移动网络系统中数据规模的指数式增加,传统的基于“人为设置单一规则”的自动化小区KPI突变检测显得力不从心^[6]。现阶段的网络运维优化迫切需要一种能够从海量运行数据中依据人工智能方法自动地获取和更新更加可靠的规则,以实现无线网络质量提供强有力的保障^[7-8]。

3 Prophet时序算法介绍

Prophet算法是2017年Facebook发布的开源时序预测框架,它主要研究时序数据特征和时序变化规律,并对未来走势进行预测^[9-10]。该算法不仅能够弥补传统时序模型对时序数据过于局限、缺失值需要填充、模型取法灵活性等不足,而且相较于传统的时序模型,在模型准确率以及使用者之间的互动方面具有更好的效果^[11]。Prophet算法是一个可分解的时间序列模型,它对时序数据各个特征进行分析,例如对时序数据的长期趋势性、季节周期性、节假日效应等特征分析。Prophet算法通过将这几项特征拟合,累加起来得到时间序列的预测值。Prophet模型的构成如下:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon t \quad (1)$$

式中:

$g(t)$ ——趋势项,用来拟合时间序列的趋势性变化

$s(t)$ ——周期性变化,一般来说是以周或者年为单位

$h(t)$ ——节假日因素造成的影响,数据中是否存在节假日

εt ——误差项,通常表示突发事件的影响

3.1 趋势项

在Prophet算法中,趋势增长类似于种族增长。Facebook采用了2种增长函数,一个是基于逻辑回归函数,另一个是基于分段线性函数。

基于分段线性函数的趋势:

$$g(t) = (k + a(t)\delta) \times t + (m + a(t)^T\gamma) \quad (2)$$

式中:

k ——增长率

δ ——增长率的变化量(适应率)

m ——偏移量参数

γ ——让函数连续

基于逻辑回归函数的非线性趋势:

$$g(t) = \frac{C(t)}{1 + \exp\{-[k + a(t)^T\delta] \times [t - (m + a(t)^T\gamma)]\}} \quad (3)$$

其中, $C(t)$, $k(t)$, $m(t)$ 是随着时间 t 变化的函数, $a(t) = (a_1(t), \dots, a_s(t))^T$, $\delta = (\delta_1, \dots, \delta_s)^T$, $\gamma = (\gamma_1, \dots, \gamma_s)^T$ 。

3.2 季节周期项

时间序列通常随着天、周、月、年等周期性的变化而呈现周期性的变化,在数学分析中,区间内的周期性函数是可以通过正弦和余弦函数来表示。在Prophet算法中,使用傅里叶级数来模拟时间序列的周期性。傅里叶级数的形式如下:

$$s(t) = \sum_{n=1}^N \left[a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right] \quad (4)$$

式中:

P ——时间序列的周期,如 $P=365.25$ 表示以年为周期, $P=7$ 表示以周为周期

Prophet 的周期函数如下:

$$s(t) = X(t) \beta \quad (5)$$

其中,

$$X(t) = \left\{ \cos\left[\frac{2\pi(1)t}{P}\right], \sin\left[\frac{2\pi(1)t}{P}\right], \dots, \cos\left[\frac{2\pi(N)t}{P}\right], \sin\left[\frac{2\pi(N)t}{P}\right] \right\}$$

参数向量 β 初始化为 $\beta \sim \text{Normal}(0, \sigma^2)$ 。

4 基于 Prophet 时序算法的突变小区检测方法

通过 Prophet 训练模型的时序分解方法,将 KPI 指标分解为 3 项:趋势项、周期项和误差项,针对趋势、周期特征分量,建立 Prophet 模型进行预测。流程如图 2 所示。

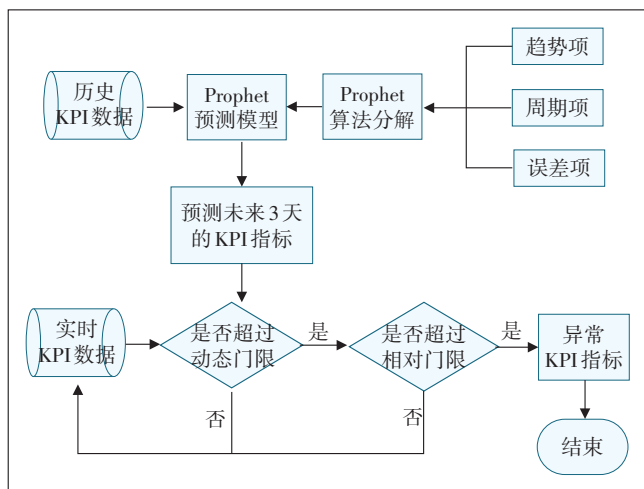


图2 基于 Prophet 时序算法的小区 KPI 突变检测流程

a) 将实际 KPI 指标值与各分量的预测值相减获得随机序列,随机序列的标准差和各分量的预测值合成得到 KPI 指标的动态门限。

b) 在绝对门限和动态门限之间设置相对门限,来设定异常 KPI 指标等级。

c) 通过判断 KPI 数值是否超过相对门限,来判定该 KPI 指标是不是发生异常。

4.1 Prophet 预测

Prophet 算法在预测过程中,能够通过调整转折点

平台参数自动探测到转折点。对历史 KPI 指标发生异常的数据进行调整,减少对 KPI 指标历史规律的影响,这样可以使预测结果更加灵活、平滑。

a) 首先采用 Prophet 模型训练的时序数据分解方法,将历史 KPI 指标数据 S_t 分解成趋势项 trend_t 、周期项 ($\text{daily}_t, \text{weekly}_t$) 和误差项 ε 。

b) 将分解出的趋势 trend_t 和周期项 ($\text{daily}_t, \text{weekly}_t$),采用 Prophet 模型预测获得第 $t+3$ 天趋势预测值 trend_{t+3} 和周期预测值 ($\text{daily}_{t+3}, \text{weekly}_{t+3}$)。

c) 将上述趋势预测值 trend_{t+3} 、周期预测值 ($\text{daily}_{t+3}, \text{weekly}_{t+3}$) 和误差项 ε 集成得到第 $t+3$ 天小区 KPI 指标的预测值 S_{t+3} , $S_{t+3} = \text{trend}_{t+3} + \text{daily}_{t+3} + \text{weekly}_{t+3} + \varepsilon$ 。

4.2 动态门限

根据中心极限定理,如果一个事物受到多种因素的影响,不管每个因素本身是什么分布,它们叠加后结果的平均值呈现正态分布。因此实际值与预测值的差值即随机序列呈现正态分布。正态分布的标准差反映了组内个体间的离散程度。如一个较大的标准差,表示大部分的数值和其平均值之间差异较大;一个较小的标准差,代表这些数值较接近平均值。

a) 将历史 KPI 指标数据 S_t 与趋势预测值 trend_t 、周期预测值 ($\text{daily}_t, \text{weekly}_t$) 相减得到随机序列 R_t , R_t 呈现正态分布, $R_t = S_t - \text{trend}_t - \text{daily}_t - \text{weekly}_t$ 。

b) 随机序列 R_t 的标准差反映了实际 KPI 指标与 KPI 指标预测的离散程度,计算随机序列 R_t 的标准差

$$\sigma, \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}。$$

c) 将第 $t+3$ 天小区 KPI 指标的预测值 S_{t+3} 与随机序列 R_t 的标准差 σ 进行集成得到第 $t+3$ 天小区 KPI 指标的动态门限 d_thres_{t+3} , $d_thres_{t+3} = S_{t+3} + \sigma$ 。

4.3 相对门限和绝对门限

a) 公司对各个 KPI 指标的健康度阈值有明确的要求,如 4G 接通率不能低于 90%,4G 掉话率不能高于 5% 等等。根据 KPI 指标的健康度阈值,设置绝对门限 a_thres 。

b) 落在动态门限和绝对门限之间的数据不一定很差。通过设置相对门限 r_thres ,来控制派单量以及 KPI 指标异常程度。相对门限设置 10 个等级,等级越大,相对门限越大,落在相对门限和绝对门限之间的

概率越小, KPI 指标越接近于绝对门限, KPI 指标异常越严重; 等级越小, 相对门限越小, 落在相对门限和绝对门限之间的概率越大, 派单量也会越多。设相对门限等级为 β 。

$$r_thres_{\text{下}} = d_thres_{t+3} + \beta \frac{d_thres_{t+3} - a_thres}{10}$$

$$r_thres_{\text{上}} = d_thres_{t+3} + \beta \frac{a_thres - d_thres_{t+3}}{10}$$

5 实验分析

5.1 原始数据

本文所采用的数据集来自某省网优某小区过去连续 28 天的 KPI 指标, 数据集中包含日期、4G 接通率等信息。在对原始数据分析的过程中, 发现数据中存在格式的不一致性、空缺值和无用信息等噪声数据, 这些噪声数据将会不利于模型的训练, 需要对原始数据进行预处理。

5.2 数据预测处理

数据格式。原始数据集中包含很多 KPI 指标, 本文实验只是使用 4G 接通率, 因此需要对其他 KPI 指标进行过滤。在保证原始数据完整性和准确性的同时, 对原始数据格式进行转换, 以满足算法模型对数据的要求。

空缺值处理。原始数据中存在空缺值, 如 4G 接通率在某一天的某个时刻没有数值。本文采用上下均值替代法, 即用空缺值时刻的上个时刻数据和下个时刻数据的均值来填补缺失的 KPI 指标数据。经过一系列对原始数据的分析与处理, 得到适合模型训练的标准、连续的数据集。

5.3 参数说明

Prophet 本质上是一种可分解的加法回归模型, 即按时序特征可将时序数据分解为非周期性变化的趋势项、天或周的季节周期项和节假日效应等。因此, Prophet 模型可以灵活地对各时序特征的参数进行设置, 这些参数值的大小分别表示各个特征对模型预测结果的贡献度。通过分析模型的训练结果, 可以调节各特征的参数, 以便进一步提升预测精确度。本文使用的 Prophet 算法的部分参数说明如表 1 所示。

5.4 实验结果

5.4.1 Prophet 算法预测结果与分析

如图 3 所示, 黑点表示某小区 4G 接通率的真实数据, 蓝线表示预测值。从预测结果可以看出, 模型对

表 1 模型的参数

参数名称	类型	说明	参考数值
growth	Linear/logistic	模型趋势的增长方式	Linear
changepoint_prior_scale	double	模型的拟合程度	0.001
changepoint_range	double	寻找突变点的比例	0.95
weekly_prior_scale	double	表示模型的周季节性灵活度	auto
daily_prior_scale	double	表示模型的天季节性灵活度	auto
changepoints	double	指定潜在突变点	None

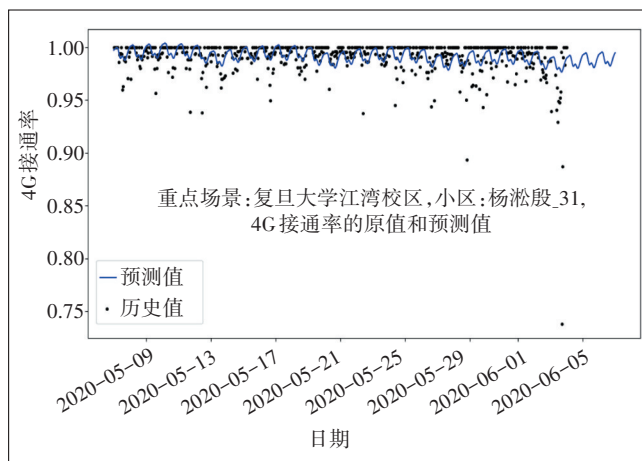


图 3 4G 接通率 KPI 指标预测值

历史数据有较好的拟合, 自动调整历史 4G 接通率数据中的异常点, 减小异常点对拟合曲线的影响, 预测结果较好地拟合了大部分历史数据的规律, 并对未来 3 天的数据进行预测, 可以发现模型对 KPI 指标历史规律的预测比较准确。而且相较于传统小区 KPI 突变检测方法, 能够对未来 3 天的 KPI 指标趋势进行预测。

5.4.2 动态门限结果分析

如图 4 所示, 浅蓝色部分表示某小区 4G 接通率预测结果的动态门限。动态门限包含了大部分的真实值, 在动态门限内的数值都属于正常 KPI 指标, 而超过动态门限的数值则属于异常 KPI 指标。相较于传统小区 KPI 突变检测方法, 动态门限更容易发现不符合历史规律的 KPI 数值, 值得运维人员重点关注。

5.4.3 相对门限与绝对门限结果分析

如图 5 所示, 较深蓝色部分表示某小区 4G 接通率预测结果的相对门限, 深蓝色部分表示预测结果的绝对门限。相对门限将动态门限和绝对门限之间的数据进一步筛选, 选取 1 级相对门限, 异常 KPI 数据减少了 20%。因此, 相对门限对异常 KPI 数据有进一步的筛选作用, 帮助运维人员关注 KPI 数据异常程度更高

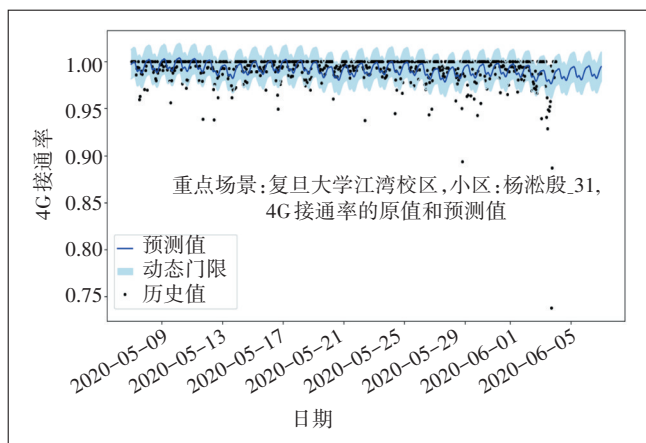


图4 动态门限

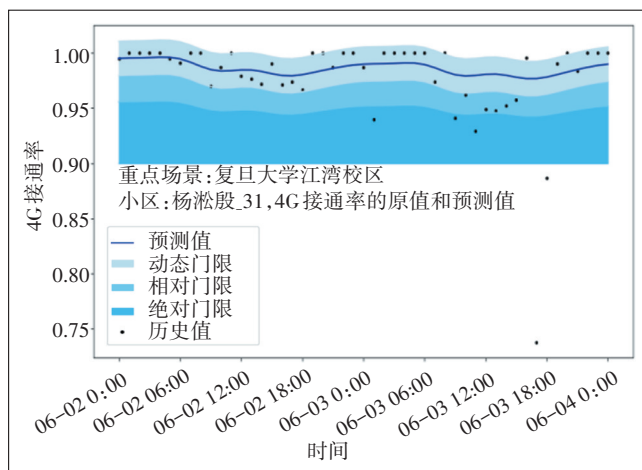


图7 等级3相对门限

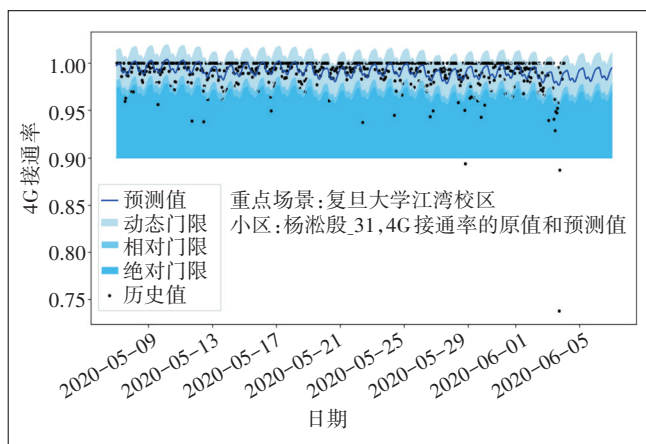


图5 相对门限与绝对门限

的时刻。

5.4.4 不同等级相对门限分析

如图6和图7所示,超过等级1相对门限的KPI数值有8个,而超过等级3相对门限的KPI数值只有5个。因此,不同等级的相对门限,可以人为控制KPI指

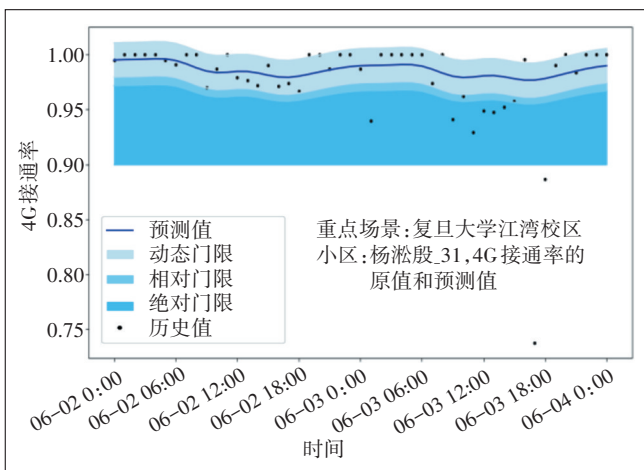


图6 等级1相对门限

标异常数量,减少预警次数。

如表2所示,某小区4G接通率从2020年6月3日9点开始发生异常,从11点到14点连续低于相对门限但未低于绝对门限,直到17点4G接通率突然恶化,直接低于绝对门限,降到74%。可以发现模型对KPI指标突然恶化有明显的预警作用,如果运维人员能够在2020年6月3日9点到14点找到4G接通率KPI指标异常原因,那么17点就不会发生4G接通率突然恶化,也不会造成用户投诉。

表2 2020年6月3日4G接通率实际值与3个门限

时间	实际值/%	动态门限/%	相对门限/%	绝对门限/%
8:00:00	100	96.75	96.07	90
9:00:00	94.07	96.46	95.82	90
10:00:00	96.19	96.35	95.71	90
11:00:00	92.89	96.40	95.76	90
12:00:00	94.88	96.50	95.85	90
13:00:00	94.76	96.52	95.87	90
14:00:00	95.24	96.41	95.77	90
15:00:00	95.76	96.24	95.61	90
16:00:00	95.52	96.11	95.50	90
17:00:00	73.79(低于绝对门限)	96.11	95.50	90
18:00:00	88.69(低于绝对门限)	96.25	95.63	90

5.5 应用效果

目前该方法已经在某省联通无线网运营中心进行试点使用,根据“区域”模块可选择全网小区、重点场景、网格、区县分公司等指定区域,通过时间和相对门限等级选择,可以自动对比绝对门限筛选出不同KPI的TOP小区,便于后台指标监控人员实时掌握不同区域的KPI突变情况(见图8)。

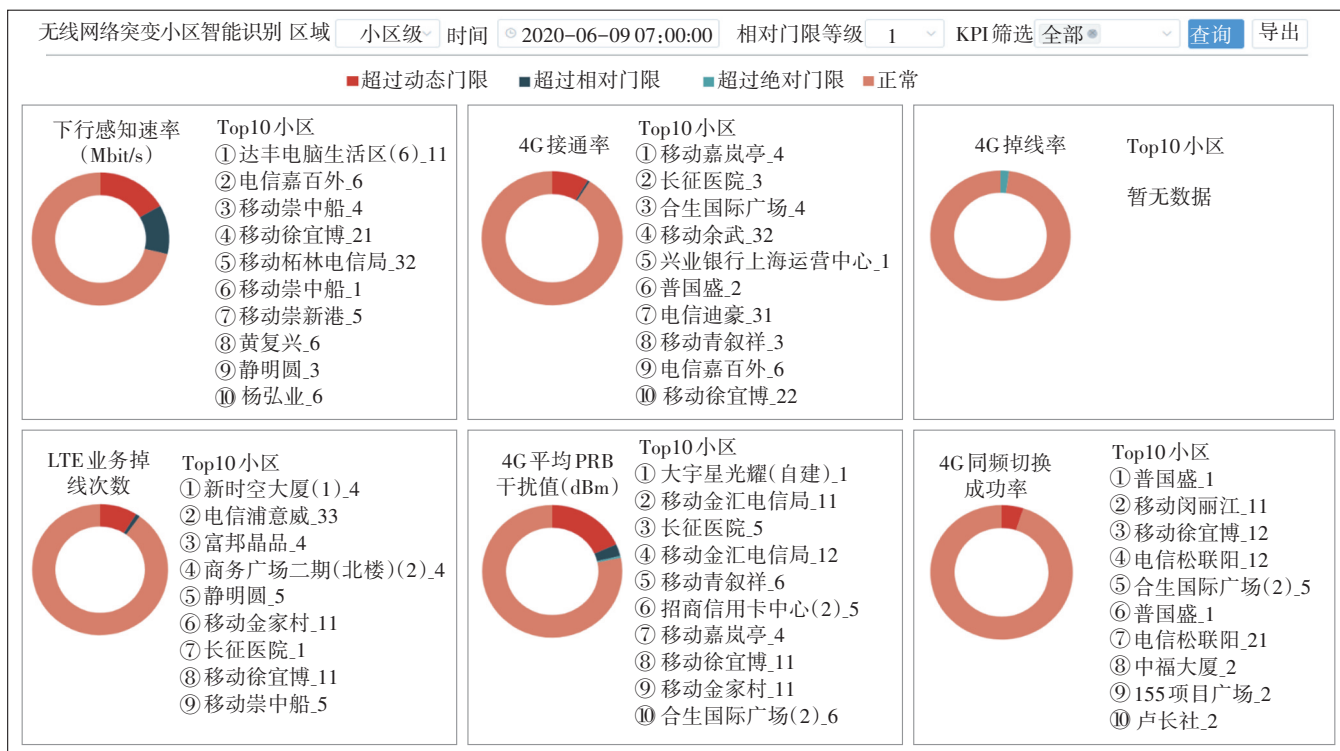


图8 无线网络突变小区智能识别应用

通过点击任意KPI指标表盘中的“超过动态门限”区域,可显示出超过动态门限的TOP小区具体分析结果及指标变化趋势图,随时掌握TOP小区指标突变时刻及历史趋势。

6 结束语

本文构建了一种基于 Prophet 时序算法的无线网络突变小区识别方法,对突变小区 KPI 指标进行识别,实验结果表明,该方法将 KPI 指标历史数据拟合,可发现历史 KPI 指标的分布规律,通过动态门限、相对门限和绝对门限精确地识别突变 KPI 指标,使运维人员及时掌握网络性能,降低误报和错报异常,提升网络优化效率和质量。目前,该方法是对 4G 网络 KPI 指标进行突变检测,将来还可以应用到对 5G 网络 KPI 指标突变检测。后续工作将考虑多维度 KPI 指标关联和 KPI 突变根因定位等,探索更加准确、适应范围更广的预测模型。

参考文献:

- [1] LI S, XU L D, ZHAO S. 5G Internet of Things: A survey[J]. Journal of Industrial Information Integration, 2018(10): 1-9.
- [2] 张平, 崔琪楣, 侯延昭, 等. 移动大数据时代: 无线网络的挑战与机遇[J]. 科学通报, 2015(5): 433-438.

- [3] 宋旭翊, 逢钰. 网络运维数据的异常检测[J]. 有线电视技术, 2018, 25(2): 92-94.
- [4] 李峻洋, 赵占强, 郭省力. LTE 无线网络优化关键性能指标研究[J]. 邮电设计技术, 2014(4): 83-86.
- [5] 徐海东, 刘立卫. 网络告警自动派单实现方法与应用效果分析[J]. 电信技术, 2009, 1(4): 32-34.
- [6] 赵诚雅. 一种基于粗糙集分类器的智能 KPI 异常检测方法[J]. 保险职业学院学报, 2020, 34(1): 62-64.
- [7] 姚荣欢. 基于小波神经网络的数据中心 KPI 预测[J]. 电子技术应用, 2019(6).
- [8] 纪勇. 基于频繁模式的 KPI 异常检测研究[J]. 无线互联科技, 2016(15): 115-118.
- [9] 彭志行, 陈峰, 陶红, 等. 时间序列分析在麻疹疫情预测预警中的应用研究[J]. 中国卫生统计, 2010, 27(5): 459-463.
- [10] 刘璐, 丁福利, 孙立民. 基于 SVM 的烟草销售量预测[J]. 软件导刊, 2016, 15(11): 134-137.
- [11] TAYLOR, SEAN J, LETHAM, et al. Forecasting at Scale[J]. American Statistician, 2018.

作者简介:

杨飞虎, 毕业于华东理工大学, 工程师, 硕士, 主要从事移动网络运维与优化工作; 许国平, 毕业于北京邮电大学, 高级工程师, 博士, 主要从事移动网络运维与优化工作; 刘贤松, 毕业于武汉大学, 高级工程师, 硕士, 主要从事移动网络运维与优化工作; 张硕伟, 毕业于上海师范大学, 工程师, 硕士, 主要从事机器学习、大数据移动网络分析工作; 郑敏之, 毕业于莫纳什大学, 工程师, 硕士, 主要从事移动网络运维与优化工作; 施亚洲, 毕业于中国民航大学, 工程师, 学士, 主要从事移动网络运维与优化工作。