

机器学习在网络性能预测中的应用

Application of Machine Learning in Network Performance Prediction

张勉知,叶晓斌,程亚锋,马丹丹,姚丽红,刘惜吾(中国联通广东分公司,广东 广州 510627)

Zhang Mianzhi, Ye Xiaobin, Cheng Yafeng, Ma Dandan, Yao Lihong, Liu Xiwu (China Unicom Guangdong Branch, Guangzhou 510627, China)

摘要:

广东联通在承载网、核心网、无线网、主机等网络中,针对设备流量、DNS、RA-DIUS、RRC等网络和业务指标进行大数据分析,指导运维人员实现重要故障的提前发现并及时对网络中的异常和隐患实施优化和应急措施。介绍并对比3种经典的时间序列预测模型,结合不同的阈值设定方法,实现现网性能指标的精准预测和异常告警。基于机器学习进行网络性能预测的方案在落地项目中展现出极好的赋能效果。

Abstract:

Guangdong Unicom conducts big data analysis on equipment traffic, DNS, radius, RRC and other network and business indicators in the bearer network, core network, wireless network, host and other networks, guides the operation and maintenance personnel to find important faults in advance, and timely implement optimization and emergency measures for abnormalities and hidden dangers in the network. It introduces and compares three classical time series prediction models, combined with different threshold setting methods, to realize the accurate prediction of current network performance indicators and abnormal alarm. The scheme of network performance prediction based on machine learning shows an excellent enabling effect in the landing project.

Keywords:

Performance prediction; Abnormal alarm; Machine learning; Big data

关键词:

性能预测;异常告警;机器学习;大数据

doi:10.12045/j.issn.1007-3043.2021.08.006

文章编号:1007-3043(2021)08-0025-04

中图分类号:TP391

文献标识码:A

开放科学(资源服务)标识码(OSID):



引用格式: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.

1 概述

近年来,AI技术蓬勃发展,5G网络建设不断推进,如何利用前瞻技术协同各专业实现运维的智能转型成为运营商面临的重大挑战之一。面对复杂的四代共生网络和海量设备,广东联通致力于推进多个AI创新项目的研究及落地应用,成功为市场线提供快速支撑。

本文构建了不同的机器学习模型,对承载网、核心网、无线网、主机等网络的多个黄金性能指标进行

自学习自监控,能够指导运维人员提前发现特大隐患及故障,打破专业壁垒,并辅助网络变更等工作顺利开展。本文阐述了广东联通利用机器学习在网络性能预测中的应用进展,重点就3种时间序列预测模型和相应的阈值设定方法进行介绍。

2 研究现状

传统的运营与维护手段耗时耗力,依靠人工分析的模式已无法应对复杂网络和海量设备的性能监控要求。另一方面,网管只能显示网络上的参数和运行异常,并不能告知当前的业务状态,一线以及专业间的沟通壁垒无法打通。随着业务向多样化、差异化的

收稿日期:2021-06-04

方向不断推进,网络扩容、设备入网等场景的网络变更需求与日俱增。网络性能指标的异常监控在网络、业务及新需求的主动感知和预测中起着越来越重要的作用。如何对海量网络性能指标进行实时监控、减少人工参与并以更高效率实现故障的提前发现,是运营商应对行业竞争、实现运维转型过程中亟需解决的关键问题。

近年来,机器学习算法的理论基础已日趋完备。其中,时间序列预测模型受到众多研究人员的关注,已被广泛应用于工程技术、医学工程、经济学和网络通信等多个领域中,并取得了不错的成果。传统的建模方法包括线性回归^[1-2]、差分自回归移动平均模型(ARIMA)^[3-4]、三次指数平滑法(Holt-Winters)和卡尔曼滤波等,这些模型概念清晰,发展比较成熟,国内外已有许多预测实例。随着人工智能技术的发展,基于神经网络的时间序列预测方法得到快速发展。

3 基于机器学习的网络性能预测

3.1 模型框架

网络性能指标(即时间序列)具备稳定性或规则性,过去态势会顺延到未来。基于这一核心思想,为实现各指标的实时监控,本文综合考虑数据特性、建模复杂性、预测精度及应用场景等,构建合适的算法框架以挖掘指标变化规律,通过对历史数据的特征学习,对网络性能指标进行精准预测,并选取合适的阈值设定方法,最终实现异常事件的提前告警。基于机器学习进行网络性能预测的主要步骤如下。

a) 数据采集及清洗:实现各性能指标的历史数据采集,并进行必要的预处理操作,如数据缺失值、异常值剔除或填补等。

b) 模型训练:用历史数据训练时间序列预测模型,实现数据的精准拟合。

c) 指标预测:利用训练好的模型实现未来一段时间的指标预测。

d) 异常告警:选取阈值设定方法,依据预测值与实际值的差异,进行实时的异常告警。

广东联通在多个AI创新项目中分别采用经典的线性回归、ARIMA+Boosting模型及Holt-Winters算法,并分别选取 3σ 、动态阈值、静态阈值方法,实现对现网指标的时间序列预测、异常告警功能,取得了不错的应用效果。下面对3套解决方案进行详细介绍。

3.2 线性回归在性能指标预测中的应用

随着业务向多样化、差异化的方向发展,网络变更(割接)已成为运营商的日常操作,用于应对中继扩容、设备入网等多个场景需求。广东联通现已研发并推出了AI网络无人驾驶系统,解决割接任务耗时久、风险高、效率低的问题。本方案采用基于机器学习算法的自学习自检测模块判别割接是否成功。本方案采用简单线性回归模型对过去7天、5 min颗粒度的割接设备上联端口流量进行拟合,输出未来一天数据的预测结果。

在对流量异常判定的过程中,本方案采用基于 3σ 准则的动态阈值方式实现告警的触发。 3σ 准则又称为拉依达准则,它是先假设一组检测数据只含有随机误差,对其进行计算处理得到标准偏差,按一定概率确定一个区间,认为凡超过这个区间的误差,就不属于随机误差而是粗大误差。定义误差率 $\Delta = |\text{预测值} - \text{实际值}| / \text{预测值} \times 100\%$ 。通过对历史流量数据误差率的计算,得到误差率的平均值 μ 和方差 σ ,若当前时刻的误差率 Δ 满足 $|\Delta - \mu| \geq 3\sigma$,则认为此时流量为异常状态,触发割接验证失败。图1为使用线性回归对IDC设备某客户业务流量的预测结果,并基于 3σ 准则进行异常触发,箭头处发现流量异常。可以看出,该方法能够较好地拟合指标变化趋势,并能有效检测出异常情况,实现割接后健康度决策。

3.3 ARIMA在性能指标预测中的应用

广东联通通过大数据+AI能力,进行流量建模,帮助网络部门对网络和业务流量进行预测,指导网络精准扩容。本文运用ARIMA算法实现流量数据回归分析,并构建Boosting模型来提高时间序列的预测精度。利用过去一周的历史数据进行模型训练,对未来24 h的流量速率进行趋势预测。预测值= $0.5 \times \text{ARIMA}$ 预测值 + $0.5 \times \text{Boosting}$ 预测值。广东联通针对4地(市)的移动网络出口流量进行了预测,表1对比了ARIMA及ARIMA+Boosting模型的准确率,从结果可以看出,Boosting极大优化了预测性能,准确率有近20%的提升。

本文采用静态阈值实现流量异常告警的逻辑是计算误差率,当误差率大于设定阈值时发出告警。误差率 $\Delta = |\text{预测值} - \text{实际值}| / \text{预测值}$,设定当误差率大于一个固定阈值时发出告警。这种告警方法有一个很大的弊端,即可能出现预测曲线波峰区域实际值与预测值偏差大却未有告警,而波谷区域实际值与预测

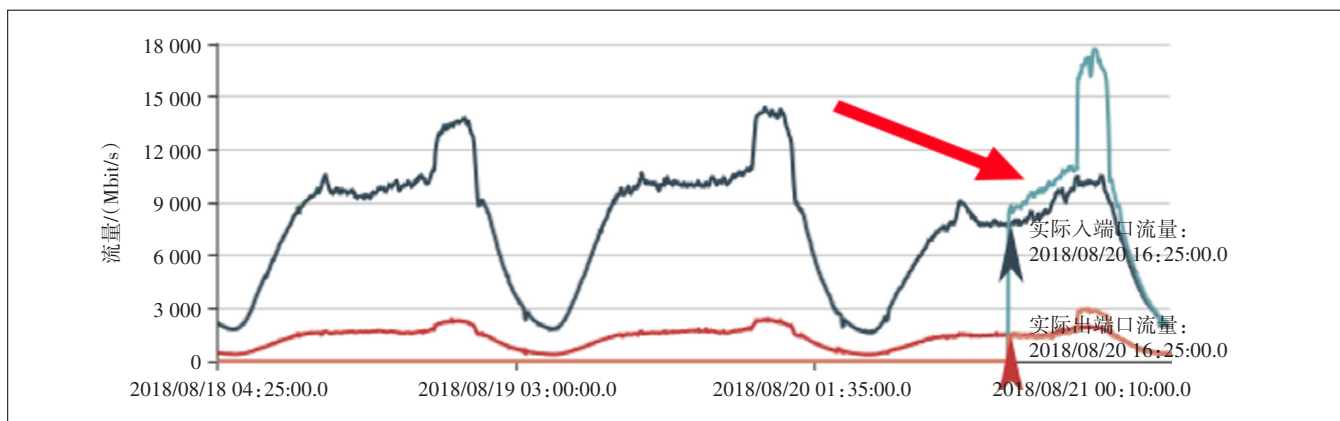


图1 基于线性回归的IDC设备业务流量预测结果

表1 ARIMA、ARIMA+Boosting 预测准确率对比

地(市)	入速率准确率/%		出速率准确率/%	
	ARIMA	ARIMA+Boosting	ARIMA	ARIMA+Boosting
A	72.22	92.36	71.53	97.22
B	70.86	93.4	80.9	98.96
C	87.5	96.53	77.78	89.93
D	81.6	96.18	73.61	94.44

值偏差较小,却大量告警的现象。

针对上述问题,本方案采用一种动态阈值方法实现异常告警。在采用前一周的历史数据训练模型时,

通过统计计算数据集85分位(P_{85})的值,其中85分位是指将流量数据从小到大排序,落在数据集长度85%位置的流量速率值。流量异常判断依据如下。

误差率 $\Delta = | \text{预测值} - \text{实际值} | / H \geq \xi$, 其中 $H = (P_{85} + \text{预测值}) / 2$, ξ 为设定的阈值。

如图2(a)所示,ARIMA+Boosting模型能对周期性变动的流量实现准确的预测,但是对于数据随机波动的细节部分,拟合效果有待提高。从图2(b)可以看出,在预测曲线波谷区域,采用改进的动态阈值方法的误差率值比静态阈值时要小,可以减小波谷区域的误告警率。

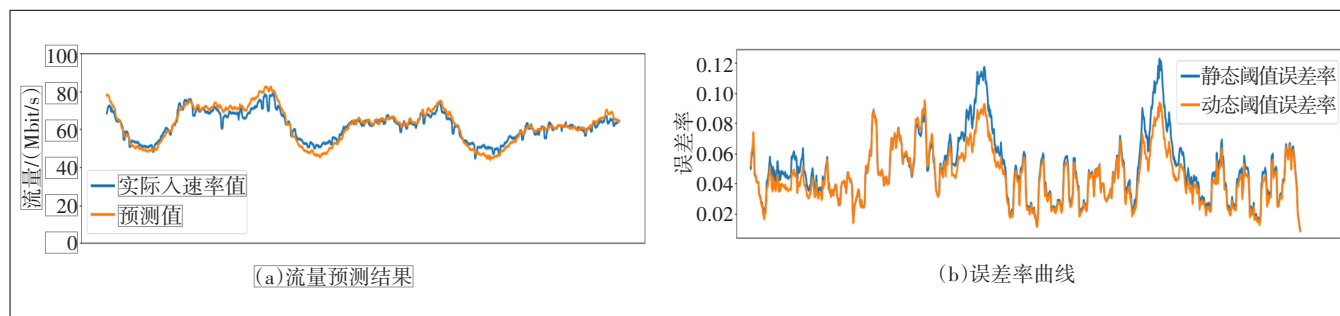


图2 基于ARIMA+Boosting+动态阈值模型的流量预测

3.4 Holt-Winters在性能指标预测中的应用

为向一线提供黄金网络指标轻触点,解决特大隐患、故障定位困难的问题,广东联通开展了大网黄金性能指标的监控、分析,研发并推出了“AI端到端分析指标墙”应用系统。该系统自2019年7月份试用以来,发现典型网络隐患多起,和广州、深圳、佛山等多个分公司建立了长期有效的联动机制。该系统已完成广东全省核心网、承载网、传送网、互联网主机系统的单专业及跨专业黄金网络指标收集及其趋势自学

习自监控,并制定了阈值告警规则触发异常告警。目前已上线微信机器人,实现业务告警的自动推送。

具体而言,该系统对城域网/承载网/物联网流量、分组网附着用户数及成功率、物联网/固网宽带用户数、DNS请求量等多个指标进行历史数据采集。本方案运用Holt-Winters算法,取过去一周的数据进行模型训练,实现未来一天的数据预测。为保证告警的准确性、有效性,该系统设定强告警规则,即当预测值、实际值和预测值的差值分别大于相应阈值才判定为

指标异常。针对不同的指标,各专业按照告警规则设定不同的阈值实现异常预警。基于 Holt-Winters+静态阈值方法的部分性能指标预测结果如图3所示,曲

线图中红色部分表明该时刻检测到指标异常。由图3可以看出,该框架能对多种指标进行精准预测,帮助运维人员提前发现故障。

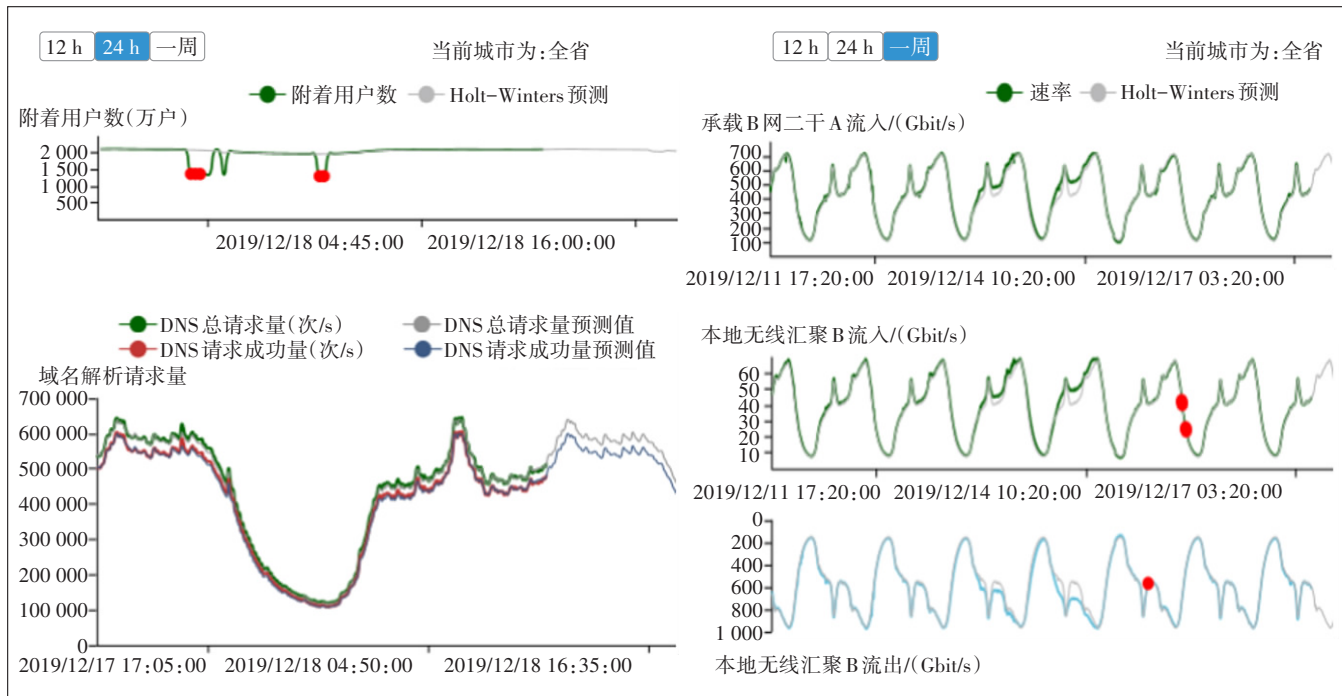


图3 跨专业、多指标的Holt-Winters预测结果图

4 结束语

本文从时间序列分析这一关键技术入手,介绍了广东联通针对现网问题提出的3套解决方案,方案均取得了较好的赋能结果并具有泛化、自适应能力。本文通过构建线性回归+ 3σ 、ARIMA+Boosting+动态阈值和 Holt-Winters+静态阈值模型,实现了网络性能指标预测及异常触发。上述3种时间序列预测模型能够准确拟合多种指标,值得一提的是,即使各指标呈现出非线性特性,简单线性回归模型也展现了不错的预测效果,且具有计算速度快的优势。根据实际应用经验,对于有递增或递减变化趋势的时间序列,建议采用 Holt-Winters 算法。静态阈值方法通过制定严格的告警规则实现指标异常触发,灵活性差,需人工设定多个阈值。而设定动态阈值具有灵活性好、误告警率低的优点。因此在设定阈值时,通常建议使用动态阈值实现异常检测。

综上所述,针对重要和特大故障提前发现及网络变更等运维问题,本文介绍了广东联通网络性能指标预测方案,借助机器学习算法,实现设备流量、DNS、

RADIUS、RRC等网络和业务指标的现网数据分析,切实解决了传统运维故障、性能看不见的问题,展现出较好的赋能效果。

参考文献:

- [1] RAO C R, TOUTENBURG H, SHALABH, et al. Linear models and generalizations[M]. 3rd ed. Beijing: Beijing World Publishing Corporation, 2015.
- [2] 王惠文, 孟洁. 多元线性回归的预测建模方法[J]. 北京航空航天大学学报, 2007, 33(4): 500-504.
- [3] CONTRERAS J, ESPINOLA R, NOGALES F J, et al. ARIMA models to predict next-day electricity prices [J]. IEEE Power Engineering Review, 2002, 22(9): 57-57.
- [4] 薛可, 李增智, 刘浏, 等. 基于 ARIMA 模型的网络流量预测[J]. 微电子学与计算机, 2004, 21(7): 84-87.

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;叶晓斌,工程师,学士,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;姚丽红,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。