基于AI的配置稽核系统在5G回传网中的应用

Application of AI Based Configuration Audit System in 5G Backhaul Network

刘惜吾[1]，马丹丹[1]，叶晓斌[1]，李亚梦[2]（1. 中国联通广东分公司，广东 广州 510627；2. 中国联通研究院，北京 100176）
Liu Xiwu[1], Ma Dandan[1], Ye Xiaobin[1], Li Yameng[2]（1. China Unicom Guangdong Branch, Guangzhou 510627, China; 2. China Unicom Research Institute, Beijing 100176, China）

摘　要：
介绍了广东联通在AI创新领域进行的一项网络智能化运维实践，提出一种基于AI关联分析的网络设备配置异常监测方法，创造性地以AI关联分析中得到的弱关联规则构建异常特征模型，从海量配置数据中快速发现配置异常。该算法具备自学习、自挖掘能力，具备良好的泛化性和应用推广性，能有效应对5G时代大量网络新建和网络改造带来的配置稽核工作量爆发式增长。

关键词：
AI；关联规则；异常特征模型；5G；配置稽核
doi：10.12045/j.issn.1007-3043.2021.08.004
文章编号：1007-3043（2021）08-0015-05
中图分类号：TN913
文献标识码：A
开放科学（资源服务）标识码（OSID）：

Abstract：
It introduces a network intelligent operation and maintenance practice of Guangdong Unicom in the field of AI innovation, and proposes a network device configuration anomaly monitoring method based on AI association analysis, which creatively uses weak association rules obtained from AI association analysis to build anomaly feature model, and quickly finds configuration anomalies from massive configuration data. The algorithm has the ability of self-learning and self mining, has good generalization and application promotion, and can effectively cope with the explosive growth of configuration audit workload caused by a large number of network construction and network transformation in 5G era.

Keywords：
AI; Association rules; Anomaly feature model; 5G; Configuration audit

引用格式：刘惜吾，马丹丹，叶晓斌，等. 基于AI的配置稽核系统在5G回传网中的应用[J]. 邮电设计技术，2021（8）：15-19.

1 概述

5G网络运营商面临网络新建和升级改造，网络规模和业务容量极速增长，网络结构呈现多维度复杂性，行业应用需求呈现多样化个性化，与此同时，用户对服务交付的质量和高效性的期许值也逐年提升。新旧网络的交织、客户市场需求的变化对基础维护工作提出更细致的要求和更高的挑战。

另一方面，从经验看网络配置引发的问题尤为突出，在配置下发过程中可能由于各种原因，如业务人员的技术水平、操作规范性等导致漏配、错配等问题。2020年广东省某地（市）一起故障，4个接入环和汇聚ASG设备间互联链路同时发生中断，经核实故障原因是由于备用平面相关环路中断站点二三层联动漏配，主用平面中断后网络切换不成功导致业务中断。为解决上述问题及挑战，广东联通积极探索5G时代网络发展的新模式，积极推进网络运维智能化进程，将大数据分析和AI技术引用到网络设备配置稽核领域，创新性提出基于AI的关联分析异常检查方法，学习建立异常配置模型，对全网设备配置进行全面高效的核查，充分发挥AI算法分析与决策能力，将运维人员从繁琐的重复性工作中解放出来，规避人为误操作，提

收稿日期：2021-06-16

升网络运维效率和网络质量可用性等级，在用户感知之前解决故障，降低网络故障率，提升用户满意度。

2 网络配置稽核现状

传统的运营商网络多采用OEM厂家提供的软硬件一体整体解决方案，如广东联通169城域网、IP承载网以华为、思科2个厂家为主，IPRAN承载网以华为、中兴、烽火3个厂家为主，这些OEM厂家以其专有的软硬件和私有协议、封闭的系统等控制行业生态。

以IPRAN承载网为例，CSG、ASG、RSG等设备的上线和业务开通过程通常需要完成大量的配置，包括一些基础配置（用户信息、AAA设置等）、端口配置（物理端口、VLAN端口等）、协议配置（ISIS、MPLS、BGP等）以及各类的业务配置（Tunnel、PW、L2VPN、L3VPN等）。现有的网络配置核查方案由厂家“分而治之”，依赖大量的人工，对不同厂家设备、同一厂家不同型号、同一型号不同版本定制化处理，维护效率低下，运营成本高，存在较多弊端。

首先，厂家配置巡检工具算法逻辑简单、稽核效率低。广东现网3个厂家虽已经配置稽核工具，但巡检逻辑单一，稽核效果不理想。如现网中兴设备使用的巡检工具ZXSEM/TIM400，通过编辑脚本定制巡检任务，通过网管对设备下发show命令，查询对应配置信息，不仅稽核时间长，还容易因为大量的任务处理导致死机，单地（市）均有上千台承载设备，配置命令总数达到200万行，例如某地（市）超过2 000台设备稽核40多小时之后死机。

其次，现有工具通常基于检测规则或者专家系统，无法适应网络设备版本的更新迭代。专家系统做配置巡检有其固有的优势，但是缺乏通用性和灵活性，如中兴的TIM400系统、华为的NCE系统均存在这样的问题，使用于某个地区或某个运营商网络的巡检工具在其他地区或其他运营商的网络上就不适用了，更无法应对5G时代大量网络新建和网络改造带来的爆发式增长的工作量，无法适配网络技术更新迭代的需求。

此外，现有工具的巡检对象往往是单台设备，没有学习能力，无法实现网络级冲突检测、隐患核查，存在较大的盲区，对于未知的配置错误大概率会出现漏检。5G新网络的运维也面临着设备种类繁多、数量庞大，客户业务多样等挑战，专业运维知识不可避免存在缺失，一些隐性的配置隐患，用传统的单台设备级视角或专家经验是很难发现的。

3 基于AI关联分析的配置稽核

配置稽核的目的是发现配置数据中的错误、隐患，从数据的角度看，就是要找到配置数据中的异常项。异常检测是机器学习应用的一个研究热点，神经网络、SVM、孤立森林、聚类等机器学习算法在网络流量、性能异常方面有大量应用。

基于AI的关联分析是传统的机器学习方法，也是强有力的数据挖掘工具，可以在海量数据中快速发现数据、事件之间的依赖关系或者因果关系，例如apriori（入选数据挖掘领域十大经典算法）、FpGrowth等，能够从大量的数据中自动搜索隐藏于其中的有着特殊关系性的信息。因此本文通过数据挖掘方法将配置文件中的关联关系挖掘出来，从而代替人工实现配置规则自挖掘、自学习。

通过上述现网配置稽核痛点分析及AI算法的研究，本文提出将AI关联分析用于网络设备配置稽核，融合了大数据分析、AI关联挖掘，借助统计分析对设备进行配置基线识别，对设备在网络中承担的角色功能进行层次化关联分析，如图1所示，按不同粒度分层检测，发现配置中的漏配、错配、冲突、冗余等配置异常，结束异厂家分而治之的局面，实现统一运维及配置稽核规则自挖掘、自学习，适应网络动态发展。

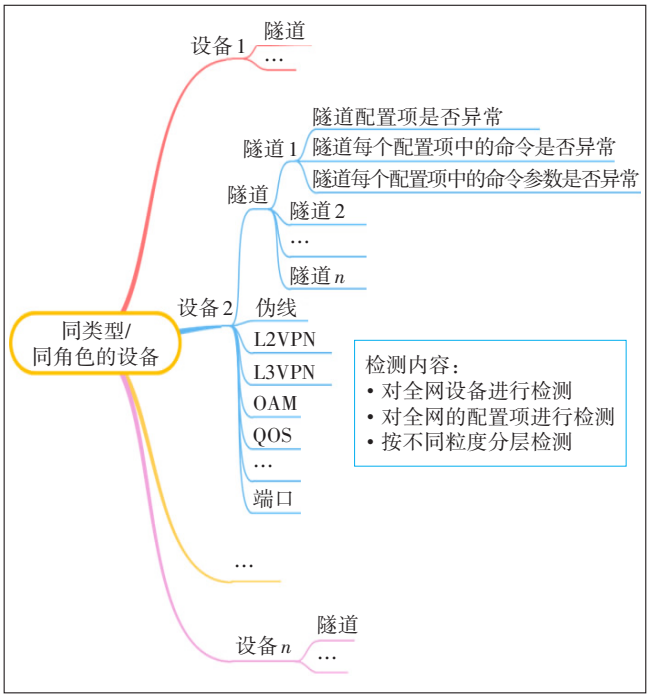

图1 基于AI的配置稽核系统逻辑架构图

基于AI关联分析的配置稽核系统分为数据采集、关联分析、人工标注、自动标注、告警通知5个模块：数据采集部分负责制定定期任务，收集基础网络设备配置文件；关联分析模块利用AI数据统计技术对配置文件进行异常检测；人工标注模块提供专业技术人员对异常列表进行标注的接口；自动标注模块收集人工标注数据集进行自动标注模型训练，标注之后的结果通过消息推送方式通知专业维护人员。系统架构图如图2所示。

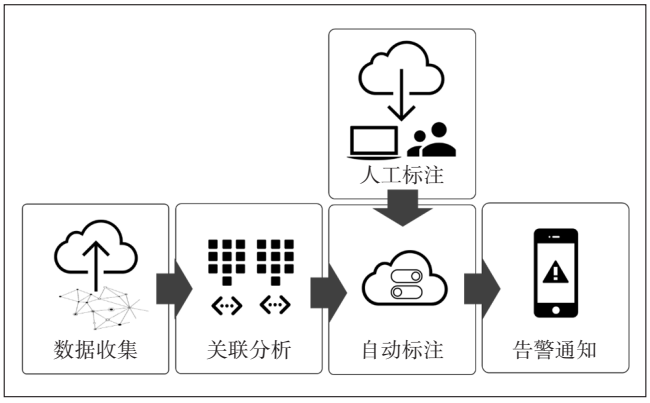

图2　基于AI的配置稽核系统架构图

4 配置稽核系统在现网的部署及验证

AI关联分析配置稽核系统部署方案如图3所示，由网络数据中台统一完成数据采集、处理，依托广东联通AI孵化平台AI框架及算力，部署AI关联分析算法，完成配置基线学习、数据挖掘关联分析、系统流程控制、用户管理、权限控制等功能，训练异常配置稽核模型。

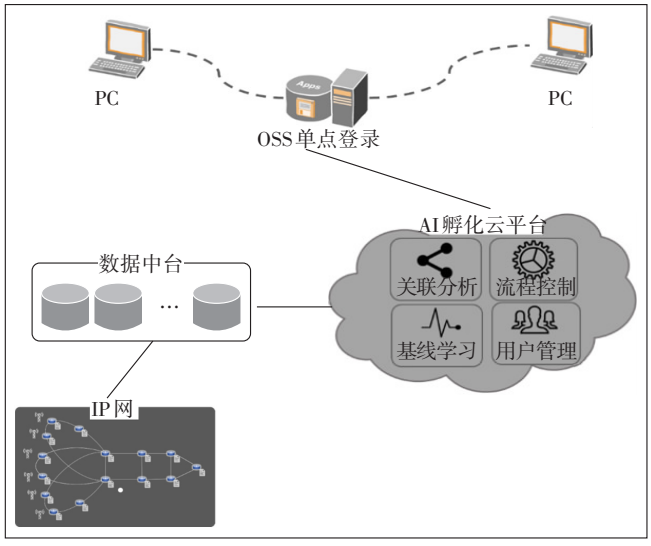

图3　配置异常检查系统部署方案

4.1 AI数据挖掘关联分析

常见的配置错误包括多配、少配、错配和冲突等，其中多配是指在某个场景或者业务的配置中出现了多余的命令或者参数，而少配则是缺少了必须的配置命令或者参数，错配通常表现为将一条配置命令配置成了和它相似的另一条命令，冲突则是配置数据中出现了2条或者多条不能同时配置的命令或者参数。对于每一种类型的设备，配置特征具备一定的共性，通过统计分析可以发现一些大致的规律，形成配置文件的基线。

如表1所示，某地（市）500台网络设备配置文件，约75万行的统计分析中间结果，从表1中可以看出某些配置出现机率非常大，说明这些配置是普遍存在的高频配置。同样从表1中也可以分析出某些配置命令只出现在某些设备中，说明某些配置具有个性化的统计特征，某些配置命令只出现在或更多地出现在某一类设备中。

表1　数据挖掘中间数据

配置命令	出现频次	机率	分类项集
interface Virtual-Ethernet4/0/0.2004	30 645	高	A
interface GigabitEthernet0/0/0	15 078	高	A
undo snmp-agent trap enable feature-name OSPF trap-name ospfIfAuthFailure	14 437	高	A
bfd 10026-eth bind pw interface Virtual-Ethernet4/0/0.2089 remote-peer ……	13 059	高	A
interface Ethernet0/2/21	9 256	中	B
bfd 101355-eth bind pw interface GigabitEthernet0/3/0.100 remote-peer ……	8 577	中	B
bfd 89852-eth bind pw interface GigabitEthernet0/2/1.99 remote-peer ……	8 270	中	B
controller E1 0/2/1	8 207	中	B
bfd 10020-tdm bind pw interface Serial6/0/1/1:0	8 004	低	C
interface Serial3/0/0/1:0	5 465	低	C
interface Trunk-Serial1/2:0	4 635	低	C

本方案利用大数据统计分析进行配置脚本的基线学习，在海量配置数据中进行数据预处理，去除干扰项，完成数据清洗从而得出分类项集，进一步用于关联分析所需的训练集。

4.2 基于关联规则的设备配置异常检查

在基线学习完成的分类项集中，本文认为配置错误应该是稀少的、偶现的，否则现网设备就不可能正常运行了。基于AI的配置稽核系统采用AI关联分析挖掘数据集中的关联规则，用绝对出现次数的占比作

为配置正确的支持度参数，配置脚本中出现的配置命令频次越大，即表示其上下文呈强关联性，配置越接近标准，配置异常的可能性越小；出现频次越低的配置与上下文呈弱关联性，异常的机率越大。

如图4所示，配置数据由一系列CLI命令构成，每条CLI命令包含一定数量的参数，CLI命令之间可能存在一些特定的标识符和分隔符，用于指示特定场景或者业务配置的起始和终结。其中左侧配置样式中的L11出现次数是1次，右侧配置样式中的L10和L11出现次数是1 731次，则配置稽核系统认为左侧配置样式中的L11是错误的，并且可以根据右侧的配置样式进行修改调整。

出现次数1

```
L1  !
L2  interface gei_*/*
L3  out_index*
L4  hybrid-attribute fiber
L5  negotiation auto
L6  jumbo-frame enable
L7  switchport mode trunk
L8  switchport trunk native vlan *
L9  switchport trunk vlan *
L10 switchport qinq normal
L11 sd enable block
L12 !
```

出现次数1731

```
L1  !
L2  interface xgei_*/*
L3  description*******_***_*/*
L4  out_index*
L5  jumbo-frame enable
L6  switchport mode trunk
L7  switchport trunk native vlan *
L8  switchport trunk vlan *
L9  switchport qinq normal
L10 sd enable
L11  port-delay-up***
L12 !
```

图4　关联规则检测异常配置

基于上述规则对采集到的海量设备配置数据进行关联性分析学习，从中挖掘弱关联规则构建异常配置模型，基于训练得出的异常配置模型对设备配置数据进行扫描，发现其中的可疑配置并上报运维人员进行处理。

4.3 异常标注

通过AI关联关系分析检测出的异常配置项，需要经过标注进行异常分类，系统最初是采用人工标注，人工标注的内容包括异常类型、严重程度、异常说明、标注者。

经过标注之后的检查结果就可以用于配置异常的修改，但是每次扫描问题列表都要经过人工分析是不现实的，会给专业人员造成更多的工作负担，因此系统设计了自动标注方式，将人工标注的历史数据作为训练数据，按照异常情况进行聚类分析并且一一映射对应处理方案，使自动标注模型学会自动识别异常问题类型。

配置稽核系统对广东某地（市）数据进行扫描之后，检测到异常192项，经过自动标注，一共标注了170项，其中高中风险有4项，没有被标注的22项异常是因为自动标注模型中没有学习到对应的异常情况，经过不断的数据积累，无法自动标注的情况会越来越少。图5是自动标注的结果。

4.4 结果验证

查准率和查全率是评价机器学习模型有效性最常用的2个指标。从整个AI配置稽核系统来看，查准率是算法找出的错误配置中到底有多少是错的，而查全率就是在所有的错误中，算法找到了多少错误。

首先考虑查准率的评估。对于配置异常模型扫描得到的配置异常，由运维人员进行标注确认，本文采集了广东省内3个地（市）的设备配置数据，各地（市）的配置数据量大小如表2所示。表2中第4列的

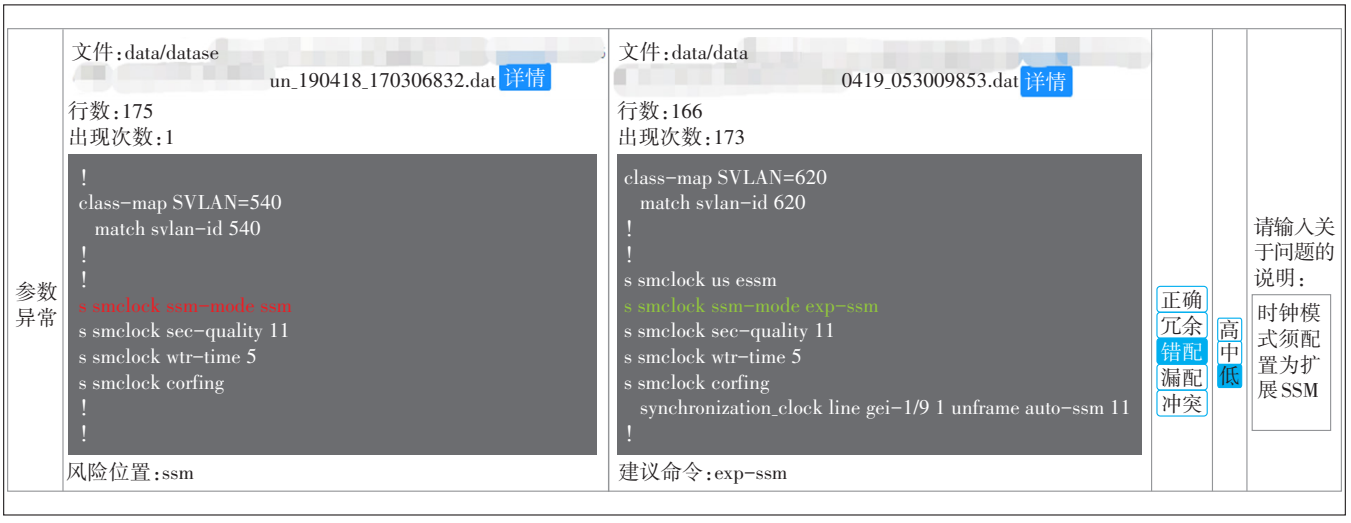

图5　配置稽核系统的自动标注结果

数值是AI配置异常模型扫描出来的可疑问题数量，第5列是人工标注确认后的问题数量，可以发现，3个地（市）的查准率都超过了80%，其中A市的查准率接近90%。另外，3个地（市）的扫描耗时都在分钟级，检查效率非常高。

表2 3个地（市）的配置数据量和扫描结果

地（市）	设备数	配置命令行数	AI配置异常模型检测结果	人工标注结果	查准率/%	系统耗时/s
A	1 079	1 993 643	192	170	88.54	58
B	1 067	5 139 324	246	213	86.59	166
C	2 612	3 820 026	417	344	82.49	109

要准确评估查全率就需要提前知道数据集中到底有多少错误配置，本文采用一种基于抽样的近似检测方法。首先由运维专业人员挑选11个常见的、不同类型的错误配置，然后将这些人为制造的错误配置随机加入到A市的配置数据集中，再由配置异常模型进行扫描，最后统计扫描结果中识别出人为制造的错误配置，由此得出算法的查全率。表3列出了挑选的11个错误内容，可见错误类型即包含常见的CLI命令漏配、错配，也有命令参数的漏配错配等，比较有代表性。扫描结果显示，本文的算法可以发现其中的9个错误，查全率达到81.8%。

表3 人为制造的配置错误

序号	设备位置	错误内容	是否找到
1	10.24.154.xxx	隧道51漏配tunnel mpls traffic-eng autoroute metric 1	是
2	10.24.154.xxx	隧道50漏配tunnel mpls traffic-eng hot-standby protect 2 dynamic	是
3	10.24.154.xxx	隧道47漏配tunnel mpls traffic-eng hot-standby protect 2 dynamic	是
4	10.24.154.xxx	BGP配置漏配bgp frr和bgp frr wtr 5	是
5	10.24.154.xxx	ip vrf IPRAN_ZTE的mpls label mode错配为per-prefix	是
6	10.24.155.xxx	OSPF AREA 0.0.0.1 配错为stub	是
7	10.24.155.xxx	pw624未关联track会话TRACK_PW_BFD_624	否
8	10.24.155.xxx	xgei-0/2/0/2.1896接口mtu错配成9000	是
9	10.24.155.xxx	cip 35漏配 status track enable	否
10	10.24.155.xxx	l3access92.1234漏配ip proxy-arp和arp gratuitous-learn	是
11	10.24.155.xxx	PW bfd检测间隔错配为100	是

进一步分析发现，第7个错误配置未找到的主要原因是在A市数据集中这种错误非常多，导致异常配置模型未包含其特征，因此未能在扫描中识别出来。

本文通过分析大量现网验证数据发现，基于AI的配置稽核算法是基于配置错误是稀少的、偶现的这个假设，当某类错误配置频繁出现时，该算法可能不能准确获得这个错误特征导致未能检查出此类错误。后续需要结合更多的机器学习算法，进一步提高配置异常检查的查准率和查全率。

5 结束语

本文提出了一种基于AI关联分析的设备配置异常检测方法，该方案结合最前沿的AI技术与网络运维技术，创造性地改变了传统人工配置稽核方式，同时有别于以往的研究，创新性地采用逆向思维，将AI关联分析中的弱关联规则作为配置异常的特征，在此基础上从海量训练集中学习配置异常模型，进而利用配置异常模型完成配置异常稽核。从现网运行结果显示，此算法的查准率和查全率都大于80%，部分场景准确率达到90%，系统检测时间低至分钟级，有效提升了配置稽核效率与配置风险识别率。

此外，该创新方案采用的AI算法具备强大的自学习、自挖掘能力，可以无缝移植到设备配置巡检核查中，如城域网、承载网、分组网等，适配5G时代海量设备运维需求，具备良好的泛化能力，能有效应对网络的动态发展，具备广泛的实用性以及可推广性，实现传统运维的智能化变革。

参考文献：

[1] 马铮，朱常波. 网络设备安全基线配置核查分析系统设计与实现[J]. 邮电设计技术，2019(4)：6-11.
[2] 徐加祥. 数通设备配置分析系统的设计与实现[D]. 济南：山东大学，2017.
[3] 齐辉. 计算机网络组网形式与网络设备配置技术分析[J]. 信息与电脑（理论版），2017(21)：141-143.
[4] 王志文，刘康平，韩冬，等. CIMS网管的智能配置分析[J]. 小型微型计算机系统，2001，22(2)：141-144.
[5] 梁睿，周涛，肖凡，等. 局数据配置管理在网络运维中的应用研究[J]. 通信电源技术，2020，37(4)：64-66，69.
[6] 唐德权，王绪峰，朱林立，等. 一种快速挖掘频繁项集算法的研究[J]. 湖南科技学院学报，2006，27(5)：117-120.

作者简介：
刘惜吾，工程师，硕士，主要研究5G网络技术应用和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；叶晓斌，工程师，学士，主要研究AI在通信网络中的应用；李亚梦，工程师，硕士，主要研究AI在通信网络中的应用。