

基于时空特征融合和图卷积神经网络的告警压缩技术研究

Research on Alarm Compression Technology Based on Time-space Feature Fusion and Graph Convolution Neural Network

黄兵明¹, 乔治², 黄剑锋³, 赵慧英¹, 马瑞涛¹ (1. 中国联通研究院, 北京 100048; 2. 中国联合网络通信集团有限公司, 北京 100045; 3. 北京神州泰岳软件股份有限公司, 北京 100012)

Huang Bingming¹, Qiao Zhi², Huang Jianfeng³, Zhao Huiying¹, Ma Ruitao¹ (1. China Unicom Research Institute, Beijing 100048, China; 2. China United Network Communications Group Co., Ltd., Beijing 100045, China; 3. Beijing Ultrapower Software Co., Ltd., Beijing 100012, China)

摘要:

分析了当前电信网络告警关联和压缩技术发展趋势和面临的问题, 提出了一种基于时空特征融合和图卷积神经网络(GCN)的网络告警根因分析及告警压缩技术方案。通过提取告警时间和空间特征, 精准剔除无效噪声数据, 并利用 GCN 和 XGBoost 的融合模型, 实现了网络告警精准根因分析和高效压缩。最后, 通过分析网络告警压缩技术应用实例成效, 验证了技术方案的有效性和可靠性。

关键词:

时空特征融合; 根因分析; 告警关联; 告警压缩

doi: 10.12045/j.issn.1007-3043.2023.07.016

文章编号: 1007-3043(2023)07-0087-06

中图分类号: TN914

文献标识码: A

开放科学(资源服务)标识码(OSID):



Abstract:

The development trend and problems of current telecommunication network alarm correlation technology are analyzed. The technical scheme of alarm correlation in telecommunication network based on time-space feature fusion and root cause analysis is proposed. By extracting the time and space features of alarms, effectively eliminating invalid noise data, and using the fusion model of GCN and XGBoost algorithm, accurate root cause analysis and efficient compression of network alarms are achieved. Finally, the effectiveness and reliability of the technical scheme are verified by analyzing the application results of the network alarm compression technology.

Keywords:

Space-time feature fusion; Root cause analysis; Alarm correlation; Alarm compression

引用格式: 黄兵明, 乔治, 黄剑锋, 等. 基于时空特征融合和图卷积神经网络的告警压缩技术研究[J]. 邮电设计技术, 2023(7): 87-92.

1 概述

与4G网络相比,一方面5G网络设备具备云化架构的技术特征,软硬件模块及接口更加复杂,网络系统的操作和维护难度大幅度提升。另一方面,微服务、网络切片等新技术的引入使得5G网络管理对象数

量暴增,新的网络形态和网络规模为网络告警分析带来了极大的挑战,告警分析处理工作量逐渐呈几何级数增加。

在当前的日常运维工作中,运维人员在分析各类告警时一般会采用聚类关联算法,对不同的告警数据进行聚类和压缩。告警处理系统中典型的算法有Apriori^[1-2]和FP-Growth算法等,还有从统计角度和可信度角度对关联规则挖掘算法进行的研究。Aprior算

收稿日期: 2023-05-08

法执行过程中需要对告警事务数据库反复进行全量扫描,并且会产生大量的候选项集,算法执行效率较低。FP-Growth算法通过构造频繁模式树这种比较紧凑的数据结构,将频繁模式信息进行压缩,本质上是一种深度优先搜索算法。基于上述典型的聚类算法开发,业内进行了很多电信告警关联规则的挖掘应用,也有很多系统基于大数据技术进行电信告警关联规则的挖掘实现。但以上告警关联分析技术中都或多或少存在如下2个问题。

a) 由于在告警关联分析过程中缺乏对全量的原始告警数据时空特征的融合分析,重要的告警信息被噪声告警淹没会导致漏报,大量噪声告警数据的混杂同样会导致误报。很多在时间、空间维度上都毫无关联的原始告警数据堆积在一起,导致真实的告警关联关系很难被洞察或无法被算法挖掘出来,尤其是对聚类算法的计算效率和收敛性带来很大的影响,直接导致告警关联分析结果的准确性降低。

b) 当前对告警数据关联的分析过程中,缺乏对告警根因分析的高效结合。通常的聚类算法只简单反应了一般性数据的关联性,没有考虑告警间的业务关系特点,例如网络不同层级的各类告警,从业务拓朴来看源自于同一个根告警。频繁项集包含过多次生告警关系项,干扰了关联关系的判断。上述研究中提到的算法方案很难分析出告警事件的关联根因,无法实现告警数据的进一步高效压缩。所以,研究如何通过各层级告警的根因智能分析定位,实现对不同层级的告警高效收敛和压缩具有重要意义。

针对以上问题和需求,本文提出了一套网络告警

智能分析和压缩的综合解决方案。方案利用时空特征数据预处理技术,高效过滤掉无关联的噪声告警数据,并在此基础上提出应用GCN和XGBoost这2种算法的融合模型进行根因智能分析定位。根据根因分析结果发现和清除衍生关联告警集,实现各层级各类告警数据的智能关联压缩,从而有效减小告警派单量,提升告警处理质量和效率。

2 基于时空特征融合和GCN算法的网络告警压缩技术方案

2.1 总体技术方案

本文提出的网络告警智能分析和压缩的综合解决方案整体流程如图1所示,方案关键步骤如下。

a) 数据采集和预处理,包括全量的告警数据、故障数据、资源数据的采集和预处理。预处理包括数据清洗、数据转换、数据补充、数据规范化处理等。

b) 时空特征融合处理,包括对告警、故障数据进行时间特征分析预处理、资源空间关联特征分析预处理,实现对无关噪声数据的精准剔除,提升数据质量。

c) 多维数据特征提取,包括告警、故障数据的时间特征、语义特征、网元资源位置特征、资源关联特征等的提取。

d) 多维数据向量生成,基于提取的多维数据特征,统一编排生成多维数据向量。

e) 多维告警根因定位,调用GCN算法模型,对多维告警和故障数据进行根因定位分析,输出故障告警根因。

f) 根因告警关联分析,对GCN算法模型输出的故

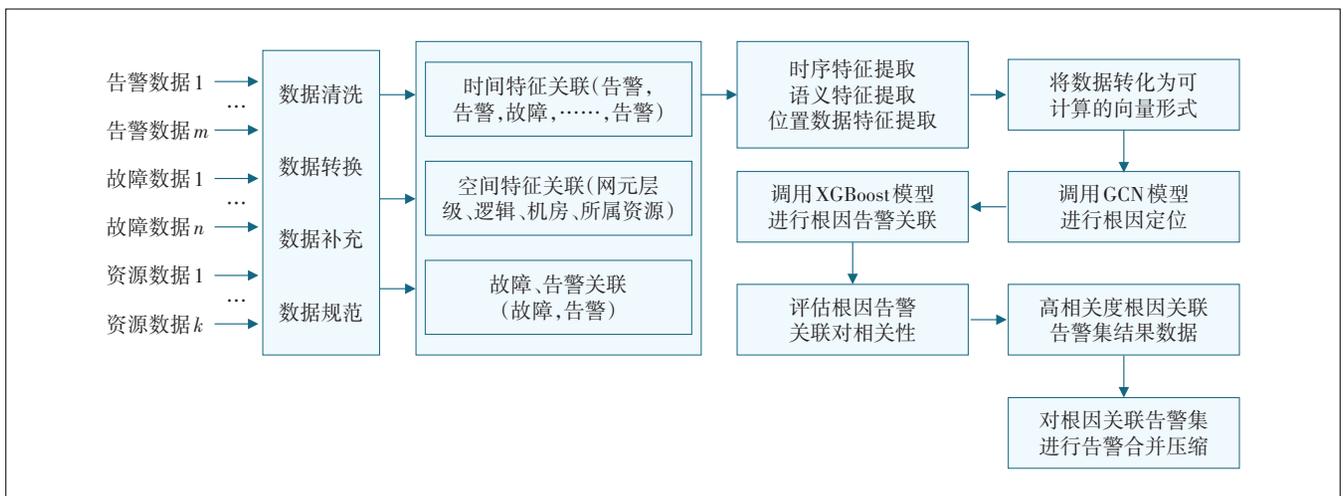


图1 综合告警智能分析总体解决方案流程

障告警根因,调用XGBoost算法模型,对所有的告警数据进行关联分析,输出根因关联告警集。

g) 根因关联告警评估,进一步分析告警根因和管理告警集的相关性,输出高相关度根因告警-关联告警集。

h) 根因关联告警压缩,基于根因高相关度关联告警集对非根因告警进行高效压缩,对同一根因导致的各层级告警进行汇聚,压缩为同一根因故障告警进行派单。

2.2 基于时空特征融合的数据预处理

本文提出的非关联数据剥离技术主要由2个关键步骤组成:告警数据时间关联性(时间特征)分析和告警数据资源关联性(空间特征)分析,精准剔除无效噪声数据,提升数据质量,最大限度提升告警关联分析算法的运算效率和准确性,其关键流程如图2所示。

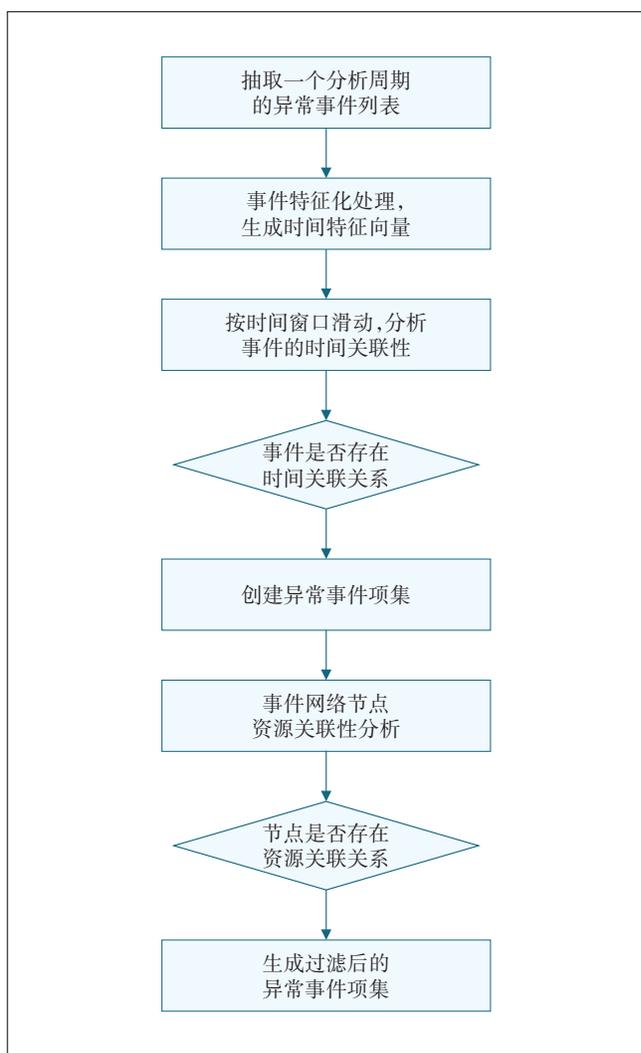


图2 时空特征融合处理关键流程

首先,抽取一个分析周期(如可以定义一个分析周期为1个月)的告警数据和故障数据,利用按需定义的特征化函数对上述数据进行特征提取,生成告警和故障时间特征向量。

然后设计一个滑动时间窗口,滑动时间窗口大小和时间步长可以根据不同的事件类型进行调整。将一个窗口周期内的异常特征事件定义为一个异常事件项集,这样能减少对不存在时间关联关系的告警、故障事件数据的无效分析,提升告警关联分析的效率和准确性。

最后进行物理和虚拟化空间中资源关联关系分析,即对上一步筛选得到的告警、故障事件数据项集再基于资源的物理、逻辑关联关系,进行下一步的数据清洗。基于资源拓扑进行关联切片处理,把相关联的资源形成切片集,获取同时具备时间关联性和切片集资源相关性的告警数据,形成资源关联告警项集。

将非资源关联异常事件数据从项集中剔除,从而避免对不存在资源关联关系异常事件的无效分析,进一步提升告警关联分析的效率和准确性。

2.3 融合GCN和XGBoost的智能告警压缩技术

本节重点描述了一种基于人工智能(AI)的智能根因告警压缩技术,即通过构建的AI算法模型,将通过2.2小节处理后的告警数据收敛到同一故障根因,减少实际派单数量并进行优化策略派单,实现故障派单的压缩合并,减少重复派单。智能根因告警压缩技术的算法模型训练方案如图3所示。

本文提出的智能根因定位AI算法模型采用GCN算法和XGBoost算法级联的融合算法方案。GCN算法定位同一时间窗口内、满足资源空间相关性的各层级告警数据的根因类型和故障节点,XGBoost算法根据GCN输出的根因类型和故障节点计算出根因关联告警集。

GCN AI算法可以利用拓扑连接性和节点特征处理图结构数据,即基于图神经网络的算法模型经过大数据训练后可以准确捕获拓扑信息。对于目标电信网络,本文使用属性图 $G=(V,E)$ 编码它的拓扑信息。其中, $v_i \in V$ 是图中的第*i*个节点,节点 v_i 具备属性 X_{v_i} ,节点对 $(u,v_i) \in E$ 表示图中2个节点之间的边,边 (u,v_i) 具备属性 $e_{u,v_i} \in E$ 。对于电信网络,GCN中的一个节点可以是一个物理设备,也可以是引发告警数据的故障根因的逻辑节点。GCN中相邻节点之间的物理连

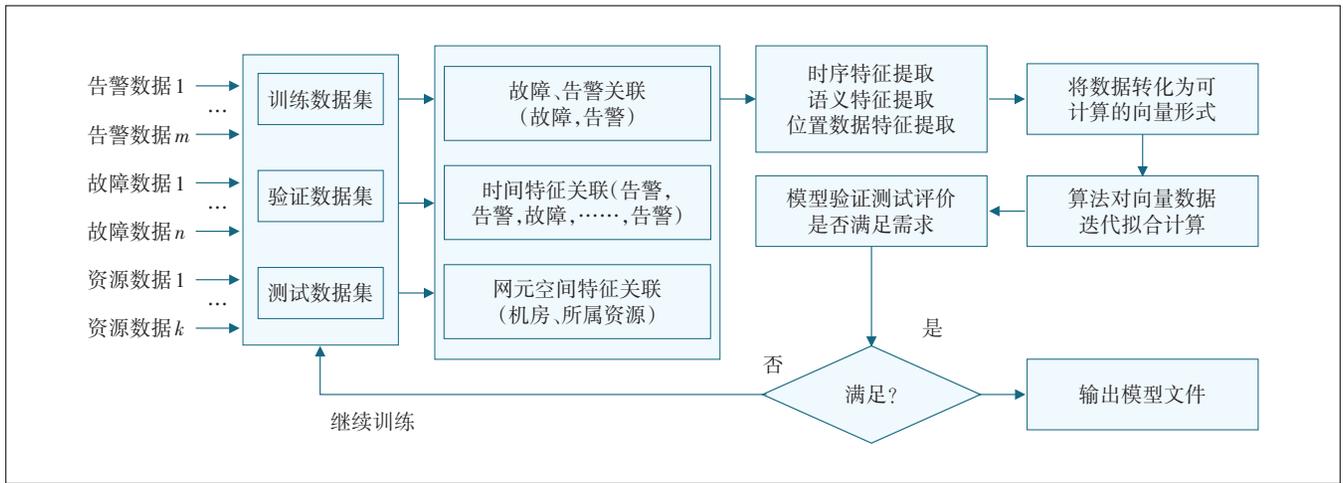


图3 告警压缩算法模型训练方案

接或逻辑连接用图形的边进行表征。本文采用GCN算法智能判断每个节点是否发生了故障问题,并判断故障问题的具体根因类型。

图卷积神经网络中节点 v_i 的表征信息通过聚合其相邻节点的表征信息进行迭代更新,GCN节点 v_i 的表征可以在 n 次迭代后捕获其 n 阶邻居的信息。本文根因定位GCN算法使用的逐层传播规则如下:

$$H^{(l+1)}\sigma = \tilde{D} - \frac{1}{2}\tilde{A}\tilde{D} - \frac{1}{2}H^{(l+1)}W^{(l+1)} \quad (1)$$

其中, $\tilde{A} = A + I_N$,为G的邻接矩阵; I_N 为单位矩阵
 $\tilde{D}_i = \sum_j \tilde{A}_{ij}$, $W^{(l)}$ 为层与层之间不可训练权重矩阵; σ 为激活函数; $H^{(l)} \in R^{n \times d}$ 为第 l 层的表示矩阵,且有 $H^{(0)}=X$ 。

在GCN定位出根因结果的基础上,调用XGBoost分类算法进行进一步的根因关联告警集智能分析,挖掘根因告警和其他各层级告警之间的强关联关系。

首先,基于根因告警和关联层级告警样例数据进行XGBoost建模训练,得到最优的XGBoost算法模型参数集。建模过程中可以学习到对应告警数据集的特征重要度信息,并保存在模型的Feature_importances_属性中。通过网络搜索方法进行交叉验证实验评估,选出XGBoost在候选参数中的最优超参数。

其次,基于最优XGBoost分类算法模型,通过在寻优范围 $[0, N]$ 内设置参数num_class(分类的类别个数),使Soft max(z_i) = $e^{z_i} / \sum_{c=1}^C e^{z_c}$ 目标函数(损失函数)最小,得到根因告警关联的各层级告警集。目标函数如下:

$$\text{Obj} = -\sum_j \frac{G_j^2}{H_j + \lambda} + 3\gamma \quad (2)$$

其中, G_j 为所有属于第 j 个叶子结点的样本的 g_i 总和, H_j 为所有属于第 j 个叶子结点的样本的 h_i 总和。

通过XGBoost分类算法,进行根因关联告警集的相关性分析,输出最优相关性的根因关联告警集。将得到的根因告警集进行合并,进一步剔除其余高相关性的各层级告警,实现对根因关联告警的有效压缩。

3 智能网络告警压缩技术应用实例

3.1 实例背景和目标

基于本文的技术创新研究成果,中国联通某省分公司在5G云网动环告警场景中对本文提出的电信网络告警智能压缩技术进行了试点验证。本次试点旨在实现对冗余告警的压缩和高效精准的故障根因定位,减少根因故障关联告警工单派发量,提升故障告警处理效率,减轻运维人员的故障工单处理负担,提效降本。

3.2 告警压缩处理流程

本节通过实例重点分析5G云网动环告警场景中,如何对数据时空特征的处理、基于GCN根因定位及根因关联告警集的生成以及算法的实现方式等方面进行适配和改进,从而更有效地解决告警关联压缩分析问题。实例对省分公司现网环境共4周的5G云网动环告警数据进行了根因关联分析和关联告警压缩,主要为以下几个过程。

3.2.1 告警原始数据时间特征分析

通过现网一级网管和二级网管采集了4周的5G云网故障数据、全量告警数据以及云网资源关联数据,并进行了数据清洗、数据转换、数据补充、数据规

范化等处理。例如剔除关键字段(如告警编码、告警标题、告警设备等)有缺失的告警数据,并根据业务层级和物理意义进行字段定义及格式对齐、数值归一化处理等。

在此基础上进一步进行5G云网动环全量告警数

据的时间特征分析和聚类,剔除关联的聚类项非时间窗口以外的噪声数据,如图4所示。基于工程实践数据分析出相关告警的发散期为 n min,则时间窗口为 n min(n 一般在10左右), i =时段宽度/ n ;当 $i>2$ 时,针对各个时间段进行时间滑动窗口聚类。

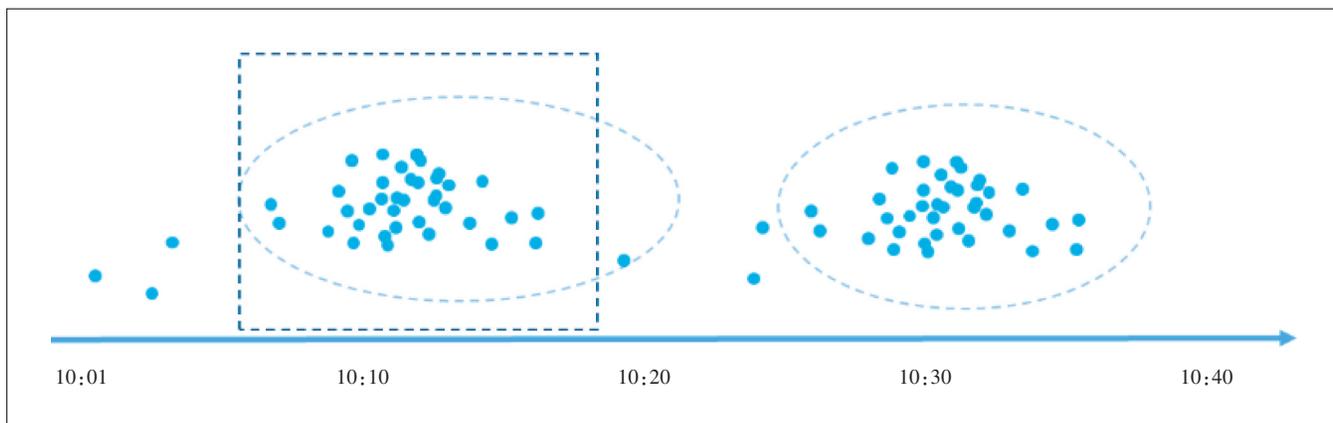


图4 5G云网动环全量告警数据时间特征分析

3.2.2 告警原始数据空间特征分析

将不同索引来源的告警和故障数据根据资源关联关系进行整合和筛选,基于资源拓扑进行切片处理,把相关联的资源形成切片集获取切片集相关资源的告警数据,形成告警库。使用K-Means算法结合滑动时间窗口算法,对告警库的告警数据进行聚类处理,形成时空特征融合告警事务集。5G云网动环全量告警数据空间特征分析如图5所示。

3.2.3 通过GCN和XGBoost算法完成告警根因定位

对于基于故障场景数据及通过时间特征和空间特征汇聚后的向量化多层次告警数据,通过调用已构

建的GCN算法模型对数据进行分析,输出告警数据的根因类型和所在故障节点。

通过XGBoost算法分析出告警根因强关联的各层级告警集,并在此基础上计算根因故障与关联告警的关联度,筛选出根因高相关度(建议关联度大于0.85)的关联告警集。部分根因关联告警集及关联度样例数据如表1所示。

3.2.4 通过根因告警压缩关联告警集派单

基于根因高相关度关联告警集对非根因告警进行高效压缩,对同一根因导致的各层级告警进行汇聚,压缩为同一根因故障告警,统一合并派发根因告警处理工单。

3.3 应用效果

在某省分公司5G动环专业一个月内挖掘有效根因汇聚压缩规则9+条,通过基于根因分析故障告警汇聚大大压缩了符合派单条件的告警数量,工单量压缩率(符合派单条件的告警数量/实际派单数量)达到31倍以上,大幅减少了人工逐级排查的工作量,缩短了排障时间,提高了工作效能,降低了工作成本,大大减轻了工单处理部门的工作压力,应用成效显著。部分根因关联告警集压缩的效果如表2所示。

4 结论

实际应用成效表明,本文提出的融合时空特征的

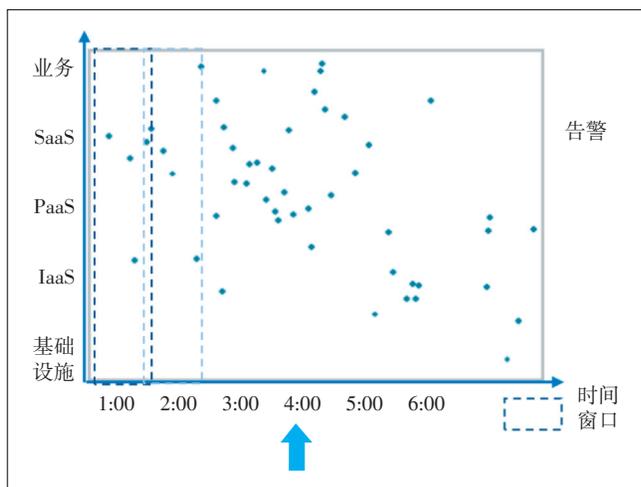


图5 5G云网动环全量告警数据空间特征分析

表1 根因关联告警集及关联度样例数据

专业	根因告警	关联告警集	关联度
动环	整流模块XX风扇告警_开关电源	整流模块XX过温告警_开关电源	0.958 6
动环	整流模块XX关机告警_开关电源	整流模块XX故障告警_开关电源	0.946 6
动环	交流输入故障告警_开关电源	B相电压过低告警_低压交流配电、A相电压过低告警_低压交流配电、停电告警_低压交流配电、交流输入XX电压过低告警_开关电源、交流缺相告警_低压交流配电、C相电压过低告警_低压交流配电、输入断电告警_低压交流配电、交流输入频率告警_开关电源、交流缺相告警_开关电源	0.939 2
动环	市电停电_智能电表	输入断电告警_低压交流配电、市电缺相_智能电表	0.939 3
动环	设备状态告警_智能电表	智能电表XX通信状态告警_监控设备	0.899 8
动环	停电告警_开关电源	交流输入故障告警_开关电源、A相电压过低告警_低压交流配电、C相电压过低告警_低压交流配电、输入断电告警_低压交流配电、B相电压过低告警_低压交流配电	0.887 3

表2 根因关联告警压缩效果分析

专业	根因告警	符合派单条件 关联告警数量	实际派 单数量	工单压 缩率/倍
动环	整流模块XX关机告警	425	18	23
动环	整流模块XX故障告警	3 560	61	59
动环	整流模块XX风扇告警	19	2	13
动环	整流模块XX过温告警	35	2	17
动环	设备状态告警	429	27	16
动环	智能电表XX通信状态告警	769	59	13
	合计	5 238	168	31

智能根因分析告警压缩技术,可有效实现故障精准定界和告警压缩、派单压缩,提升电信网络运维的自动化、智能化运维保障能力,同时显著降低人力投入成本,具有重要的工程应用价值和社会经济价值。但现阶段仍存在问题,例如算法获得关联规则可解释性较差等。为解决该问题,未来可考虑引入知识图谱等技术,整合更多的数据,包括网络性能信息、业务状态信息、拓扑信息、工单信息等,将规则进一步扩展沉淀为完整闭环的知识图谱,全面关联网络告警,准确定位网络故障。

参考文献:

[1] AGRAWAL R, IMIELŃSKI T, SWAMI A. Mining association rules between sets of items in large databases [J]. ACM SIGMOD Record, 1993, 22(2): 207-216.
 [2] AL-DHARHANI G S, OTHMAN Z A, BAKAR A A. A graph-based ant colony optimization for association rule mining [J]. Arabian Journal for Science and Engineering, 2014, 39(6): 4651-4665.
 [3] JIAO M H, YAN P, JIANG H Y. Research and application on Web information retrieval based on improved FP-growth algorithm [J]. Wuhan University Journal of Natural Sciences, 2006, 11(5): 1065-1068.
 [4] YUAN J B, DING S L. Research and improvement on association rule algorithm based on FP-growth [C]//Web information systems and

mining. Berlin, Heidelberg: Springer, 2012: 306-313.

[5] 徐前方,肖波,郭军.一种基于相关度统计的告警关联规则挖掘算法[J].北京邮电大学学报,2007,30(1):66-70.
 [6] 肖波.可信关联规则挖掘算法研究[D].北京:北京邮电大学,2009.
 [7] YUAN Y B, HUANG T Z. A matrix algorithm for mining association rules [C]//Advances in Intelligent Computing. Berlin, Heidelberg: Springer, 2005: 370-379.
 [8] 张雄.基于关联规则的电信网告警相关性分析[D].南京:东南大学,2016.
 [9] 张永华.基于大数据技术的电信网络告警关联分析设计与实现[J].电信工程技术与标准化,2016,29(4):18-23.
 [10] 杨磊.基于FPgrowth机器学习的影响用户感知无线根因问题的快速定位方法研究[J].江苏通信,2019,35(2):56-62.
 [11] 蒲元芳.关联规则在电信告警系统中的研究与应用[D].广州:广东工业大学,2010.
 [12] 岳海涛.基于事件关联和数据挖掘的网络故障管理技术的研究[D].长沙:中南大学,2010.
 [13] 徐前方.基于数据挖掘的网络故障告警相关性研究[D].北京:北京邮电大学,2007.
 [14] 李文栋.基于Spark的大数据挖掘技术的研究与实现[D].济南:山东大学,2015.
 [15] 李川.基于Spark的电信网络告警大数据关联规则算法研究与实现[D].北京:北京邮电大学,2017.
 [16] 常铮,马少伟,毛斌宏.大数据场景下基于机器学习的5G云网告警关联分析[J].邮电设计技术,2022(6):71-76.

作者简介:

黄兵明,毕业于北京化工大学,工程师,硕士,主要从事人工智能在网络运维中的应用、网络智能运维的实现等工作;乔治,毕业于北京工业大学,高级工程师,硕士,主要从事业务平台的规划、设计、建设与运营工作;黄剑锋,毕业北京邮电大学,博士,主要研究方向为宽带移动通信新技术、算力网络新技术、大数据人工智能技术等领域;赵慧英,毕业于北京交通大学,中国联通集团企业博士后,博士,主要从事网络智能化中图神经网络与知识图谱的关键技术研究工作;马瑞涛,毕业于北京邮电大学,教授级高级工程师,硕士,主要负责人工智能、区块链及新型数字技术的创新应用工作。