

高维空间网络告警智能关联分析方法

High-dimensional Intelligent Approach to Network Alarm Association Analysis

匡立伟¹, 赵良², 杨韬³, 张贺²(1. 烽火通信科技股份有限公司, 湖北 武汉 430073; 2. 中国联通网络技术研究院, 北京 100048; 3. 中国联通江苏分公司, 江苏 南京 210029)

Kuang Liwei¹, Zhao Liang², Yang Tao³, Zhang He²(1. FiberHome Telecommunication Technologies Co., Ltd., Wuhan 430073, China; 2. China Unicom Network Technology Research Institute, Beijing 100048, China; 3. China Unicom Jiangsu Branch, Nanjing 210029, China)

摘要:

网络告警关联分析是通信网络运维的关键,传统基于专家规则的关联分析方法效果较差,基于人工智能技术实现网络告警精准关联是当前研究热点。针对网络告警字段复杂、数据量大、智能化程度弱三大特点,提出高维空间网络告警智能关联分析方法,基于人工智能技术从大规模告警数据中准确提取衍生告警与根源告警的关联规则,准确定位故障发生点。分析了智能关联分析方法的应用场景,对网络告警智能关联分析发展趋势进行展望。

关键词:

网络告警关联分析;人工智能技术;高维空间模型;张量模型

doi: 10.12045/j.issn.1007-3043.2018.12.003

中图分类号: TP181

文献标识码: A

文章编号: 1007-3043(2018)12-0012-05

Abstract:

Association analysis of network alarm is one of the fundamental process for telecommunication network management and maintenance. As the traditional rule-based method can not efficiently extract the association rules from network alarm, many researchers has been performing studies on artificial intelligence based association analysis of network alarm. According to the three features of network alarm, namely variety, volume, and low intelligence, it proposes a novel high-dimensional intelligent approach that can appropriately extract the root alarm and find the associated fault point of the network equipment. The application scenario of intelligent association analysis method is analyzed, and the development trend of intelligent association analysis of network alarm is prospected.

Keywords:

Association analysis of network alarm; Artificial intelligence technology; High-dimensional space model; Tensor model

引用格式: 匡立伟, 赵良, 杨韬, 等. 高维空间网络告警智能关联分析方法[J]. 邮电设计技术, 2018(12): 12-16.

0 引言

随着通信网络持续发展和演进,网络设备不断增多、网络规模日趋复杂,网络中的软硬件时刻产生海量网络告警,这些网络告警数据大、告警字段复杂。为了保证通信网络的正常运维,要求实时完成告警关联分析,从海量衍生告警中提取根源告警,压缩或过滤重复告警、衍生告警和不重要的告警,准确定位故障隐患点,从而保障网络健康运行,为用户提供高质

量通信业务。

传统网络告警监控和处理依靠人工专家完成,基于告警监控获取当前活动的告警信息,然后通知维护人员查看可能的故障点,汇报相应的检测结果。人工方式能够在一定程度上发现和处理故障,但人工方式一般是通过排查可疑故障点的处理模式,不仅消耗了大量的人力物力,增加网络的运营维护成本,并且处理过程非常耗时,在发生大量故障告警时难以满足告警处理的实时性要求。另外,采用人工处理的方式很有可能会忽略某些关键告警而导致整个故障不能及时得到解决,极大影响通信网络的质量。

收稿日期: 2018-11-22

随着机器学习、人工智能等技术的蓬勃发展,将新兴技术引入网络告警相关性分析应用中,构建智能化、自动化的网络故障处理监控系统和分析平台,从而为运维人员的故障处理提供更为高效的方式成为研究热点。针对当前网络告警告警字段复杂、数据量大、智能化程度弱三大特点,本论文提出高维空间网络告警智能关联分析方法,基于人工智能和数据挖掘技术从海量复杂告警数据中实时提取衍生告警与根源告警的关联规则,准确定位故障发生点。

本论文组织结构如下:第1章介绍网络告警关联分析、人工智能和深度学习技术国内外研究现状。第2章分析网络告警智能关联分析面临的挑战,提出高维空间智能关联分析方法框架。第3章详细阐述高维空间网络告警统一表示方法、高质量核心网络告警提取技术、训练样本构建技术,以及网络告警深度学习神经网络模型。第4章探讨如何将论文提出的方法在具体的业务场景中落地,分析智能关联分析系统与现网系统的对接方案。第5章对论文内容进行总结,并对人工智能技术在网络告警关联分析中的发展趋势进行展望。

1 国内外研究现状分析

1.1 网络告警关联分析研究现状

通信网络告警数据具有内在相关性,同一故障往往在网络层面上引发若干告警。如何借助告警间的相关性分析,从大量的告警中找到根源告警,压缩或过滤重复告警、衍生告警和不重要的告警是急需解决的关键问题。传统方法采用专家经验和数据挖掘方法,根据已有网络理论知识及积累的业务相关性经验总结出相关规则指导告警的关联溯源,从而梳理告警信息、提炼根源告警、确定故障隐患点。论文[1]构建一种告警关联分析系统 ARM-ACAS 用于提取关联规则,并通过优化的带权频繁模式树提高挖掘效率。为了提高关联规则挖掘效率,论文[2]提出一种增量式关联规则提取方法,在增量数据进行关联分析,并将结果与历史分析结果合并。针对传统告警关联方法需要复杂推理的挑战,论文[3]提出基于神经网络的告警关联分析方法。论文[4]采用大数据方法进行电信网络告警关联分析。这些传统方法能够提取衍生告警与根源告警关联规则,但往往会忽略或漏报根源告警,关联根源告警以及定位故障发生点的准确率都不高。

1.2 人工智能技术研究现状分析

人工智能^[5](Artificial Intelligence)是一门综合性技术,涉及计算机科学、信息论、控制论、神经生理学、语言学等多个领域,是在机器上模拟人类的思维方式和工作方法,实现和扩展人类智能的一门科学和技术。随着神经网络^[6]、机器学习、深度学习^[7]、大数据等各种技术方法的研究,人工智能新方法新思路不断出现,深度学习更是被评为2013年十大突破性技术的首位技术。深度学习是一种机器学习方法,尝试采用包含复杂结构多重非线性变换的多个处理对数据进行抽象处理。深度学习网络中,每一个节点在前一层输出的基础上学习识别一组特定的特征。深度学习是机器学习领域近年来一个成功的研究方向,目前深度学习技术正与各行各业结合,解决各行业的各类问题。人工智能深度学习技术在处理过程中需要大规模优质的样本进行训练,通过训练样本学习,得到数据中隐含的内在规律和特征。张量^[8-11]是一种数学模型,广泛用于处理分析大规模复杂数据,论文[12]提出了基于张量的大数据统一表示方法,在高维空间中实现结构化、半结构化、非结构化数据的统一表示。论文[13]、[14]和[15]采用云计算、分布式、增量式方法实现大规模数据处理。

网络告警关联分析、人工智能深度学习、分布式增量数据处理领域,各自都有较为成熟的技术和方法。但是,综合运用关联分析、深度学习、大数据处理技术,构建高效、智能的数学模型,研发智能化、自动化告警关联分析和故障定位平台,为通信网络运维人员故障处理提供更为高效的方式,学术界和产业界尚处于研究起步阶段。本论文提出高维空间智能告警关联分析方法,将大规模复杂告警数据统一描述在高维张量空间中,采用分布式增量方法从大规模告警数据中提取高质量核心数据,构建深度学习模型挖掘衍生告警与根源告警的关联关系,准确定位网络故障发生点,从而实现智能、自动排查故障,保障通信网络正常运维。

2 问题描述和方案框架

2.1 告警关联分析面临的三大问题

基于人工智能技术实现网络告警关联分析和故障定位,目前面临网络告警字段复杂、网络告警数据量大、告警关联分析智能化程度弱三大困难,本论文将这三大困难归结为3个问题,如下所述。

问题1:网络告警数据字段包括告警类型、告警名称、网元类型、网元名称、告警级别、开始时间、结束时间、清除时间、流水号、操作员等。不同厂商的告警数据往往有所不同。采用高效的表示方法对不同厂商网络告警数据进行统一表示,是告警关联分析需要解决的第1个问题。

问题2:网络告警数据规模大,但海量告警数据包含大量不完全、不一致、重复冗余的告警数据,如何采用分布式、增量式处理技术,从海量告警数据中提取核心告警,构建高质量训练样本,是告警关联分析需要解决的第2个问题。

问题3:网络告警关联分析和故障定位是一个动态持续的过程,需要对已有关联规则进行学习、分析,在关联分析和故障定位过程中不断学习,持续优化,生成新的、更有效的关联规则,去除旧的、低效的关联规则。如何让智能关联分析系统在分析过程中变得更加智能,是告警关联分析需要解决的第3个问题。

2.2 网络告警智能关联分析技术框架

针对上述三大问题,本论文在高维空间模型中统一表示网络告警数据,基于张量高阶奇异值分解获取高质量核心告警数据,借助分布式、增量式技术提高核心数据提取能力,构建深度学习神经网络模型准确提取衍生告警与根源告警之间的关联规则,准确定位故障发生点。智能关联分析方法总体方案如图1所示。

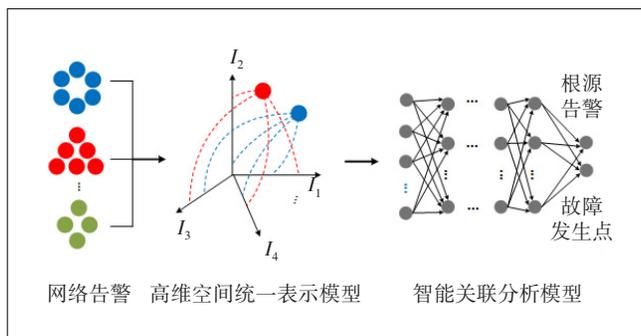


图1 高维空间网络告警智能关联分析方法总体架构

图1左边蓝色、红色、绿色圆圈表示通信网络中不同类型的网络告警。图1中间,网络告警在高维张量空间中统一表示,张量的每一个阶对应网络告警的一个字段。图1右边,将网络告警训练样本输入到深度学习神经网络模型,挖掘样本数据中隐藏的关联模式,获取衍生告警与根源告警关联规则,定位故障发生点。

3 高维空间智能关联方法

3.1 网络告警数据统一表示方法

通信网络告警数据字段包括告警类型、告警名称、网元类型、网元名称、告警级别、开始时间、结束时间、清除时间、流水号、操作员等,本论文在高维张量空间模型中统一表示网络告警数据各个字段,张量模型定义如下:

$$T \in R^{I_1 \times I_2 \times I_3 \times \dots \times I_N} \quad (1)$$

式中:

T ——张量模型

R ——实数域

$I_1, I_2, I_3, \dots, I_N$ ——张量模型的第1、第2、第3、第 N 阶。张量的每个阶对应网络告警的一个字段,例如张量第1阶 I_1 对应告警类型,第2阶 I_2 对应告警名称,第3阶 I_3 对应网元类型,以此类推。在进行网络告警统一表示的时候,告警名称、网元类型、网元名称等都进行统一编号,每一个编号对应一个特定的告警名称,或者对应特定的网元类型。通过编号将具体的告警名称、网元类型、网元名称等数据信息统一映射到张量空间的各个阶上。

3.2 核心告警数据分布式增量计算方法

基于张量实现网络告警统一表示以后,通过高阶奇异值分解公式^[12]求得核心张量和近似张量。

$$S = T \times_1 U_1^T \times_2 U_2^T \dots \times_p U_p^T, \quad (2)$$

$$\hat{T} = S \times_1 U_1 \times_2 U_2 \dots \times_p U_p.$$

式中:

S ——核心张量

U_1, U_2, \dots, U_p ——原始张量的各个左奇异矩阵

\hat{T} ——近似张量,包含高质量告警数据

本论文采用分布式增量方法计算核心张量。

图2是分布式增量网络告警核心数据处理方法示意图。图2左边是一个三阶张量模型,通过高阶奇异值分解求得核心张量和3个截断的左奇异矩阵。当有新的网络告警数据到达时,采用增量式高阶奇异值分解方法,分别求得核心张量的增量以及3个截断左奇异矩阵的增量,然后合并求得更新后的核心张量和近似张量。为了提高计算效率,如图2右下所示,本文提出在云平台中采用分布式方法实现核心张量 S 、截断奇异矩阵、近似张量 \hat{T} 的计算。

3.3 关联分析告警训练样本构建方法

得到高质量核心告警数据张量模型后,沿着张量

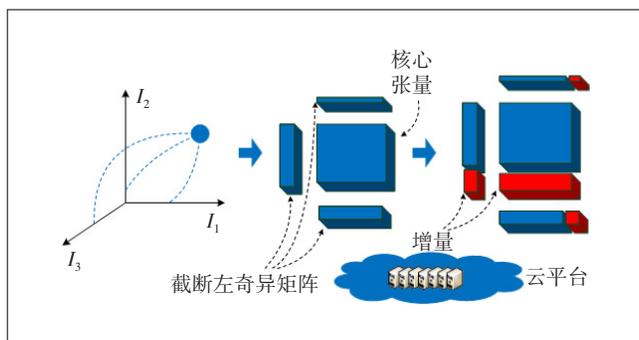


图2 分布式增量核心告警数据处理方法

的不同阶进行展开求得模展开矩阵。将张量沿第 p 阶展开得到矩阵 $T(p)$ 定义如下:

$$T_{(p)} \in R^{I_p \times (I_1 \times I_2 \times \dots \times I_{p-1} \times I_{p+1} \times \dots \times I_n)} \quad (3)$$

公式(3)中的具体展开方式可以参考论文[12]。高质量网络告警张量模型展开为矩阵后,可以再次转换为向量,然后输入智能关联分析深度学习网络中进行学习和训练。在实际应用中,可以根据场景要求,自行选择沿着哪一阶展开。

为了基于人工智能深度学习神经网络模型去分析衍生告警与根源告警的关联关系,通过根源告警定位故障发生点,需要构建一批高质量的训练学习样本,将样本数据输入深度学习神经网络模型中,通过函数学习关联规则所隐含的内在特征规律。训练样本来自于传统的专家基于经验生成的关联规则库,也可以基于数据挖掘关联分析提取的告警关联规则库。本论文用 $A_1 \rightarrow A_2$ 表示一条告警关联规则,其中 A_1 是根源告警, A_2 是衍生告警。本论文用 $F_1 \rightarrow A_2$ 表示故障发生点与衍生告警的关联规则,其中 F_1 是根源告警, A_2 是衍生告警。本文深度学习神经网络训练样本的输入为衍生告警的各个字段值,输出为根源告警名称对应的编号,以及网络故障发生点名称对应的编号,训练样本形式描述为:

$$e = (v_1, v_2, v_3, \dots, v_r, v_f) \quad (4)$$

式中:

e ——训练样本的个体

v_1, v_2, v_3 ——衍生告警的各个字段

v_r 和 v_f ——根源告警和故障发生点对应编号

3.4 网络告警智能关联分析模型

网络告警训练样本构建完毕后,输入智能关联分析深度学习模型,在神经网络中学习真实通信网络告警样本数据中隐藏的关联规律和内在特征,通过高质量核心样本数据中的输入和输出,不断调整优化神经

网络各层节点的权重参数,通过上述高质量代表性的告警样本数据,训练出一个能识别内在关联模式的神经网络。本论文采用大规模、复杂字段网络告警样本数据对神经网络进行训练,样本中已经去除重复、冗余、不一致数据,样本数据的概率特征与真实网络告警样本特征相同,保证深度学习网络能寻找到内在的告警关联规律。

另外,本论文采用反馈机制去优化、更新智能关联分析深度学习模型。在网络运维过程中,评估挖掘出的各项关联规则,并根据评估结果构建新的训练样本,然后输入深度学习神经网络,调整优化权重参数,让神经网络模型不断优化,不断学习,不断提高告警关联分析的智能化水平,提高分析的准确率。

4 应用场景分析

本论文提出基于高维空间的网络告警智能关联分析方法,基于论文中的方法开发关联分析平台,部署到通信网络中,对接现网统一管控平台,可以实现告警数据上报、衍生告警过滤、网络故障排除等功能,如图3所示。

图3最下边是通信网络设备,将网络告警上报至统一管控平台。统一管控平台包含传统网管系统中的告警管理、性能管理、配置管理,以及软件定义网络中的控制面相关功能。统一管控平台对接基于本论文所提方法开发的网络告警智能关联分析平台,对网络告警数据进行统一表示,获取核心网络告警数据,构建训练样本,通过深度学习神经网络模型全面挖掘告警关联模式,获得关联规则,然后基于关联规则对实时上报的网络衍生告警进行过滤,定位网络故障发生点,然后通过统一网管系统或其他网络管理平台,排除网络故障。

5 总结与展望

本论文提出基于高维空间的网络告警关联分析方法,针对网络告警网络告警字段复杂、数据量大、智能化程度弱三大特点,在高维张量模型中对网络告警数据进行统一表示,采用分布式增量方法实时获取高质量网络告警核心数据,将网络告警学习转换为学习训练样本,然后构建深度学习神经网络模型匹配、逼近网络告警关联模式,最终学习到隐藏在数据内部的告警关联规律,精确高效地挖掘告警关联规则。论文分析了高维空间网络告警关联分析模型与现有系统

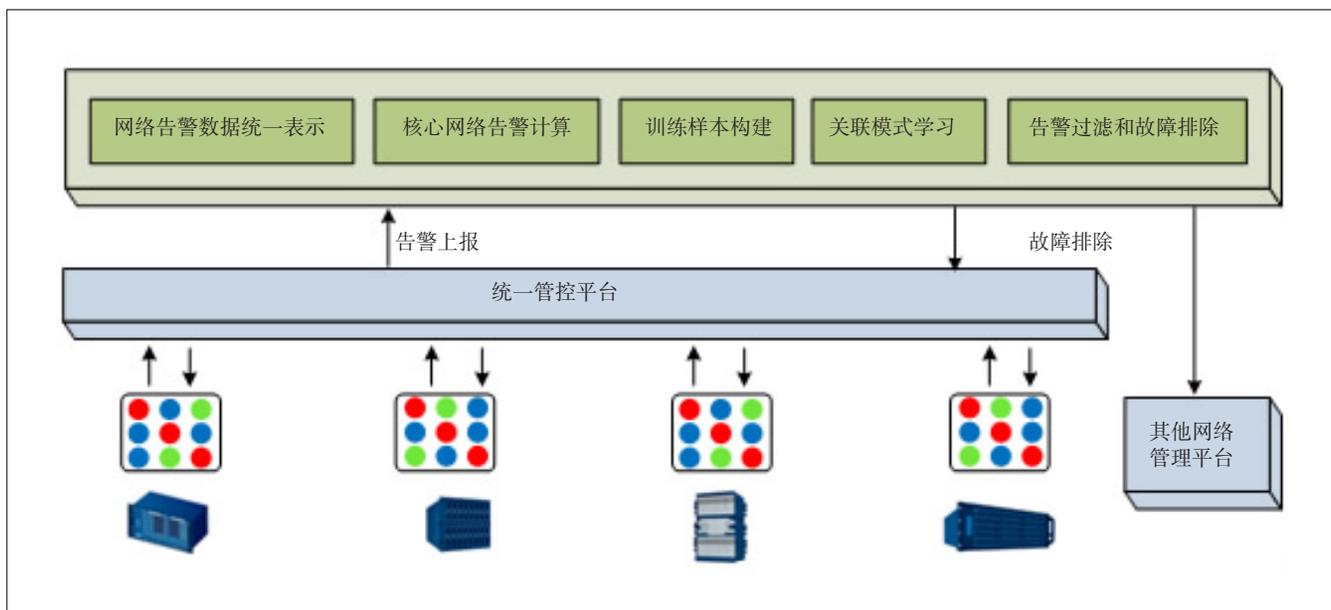


图3 智能关联分析方法应用场景

的对接模式,通过应用场景分析给出高维空间智能关联分析方法在通信网络的落地方案。

在后续研究过程中,可以引入非对抗网络等前沿技术,通过生成系统与判定系统的博弈不断优化告警关联规则,提高关联规则的准确率,提高深度学习网络模型的智能化水平。另外,需要进一步研究网络告警样本数据的概率分布规律,优化神经网络模型去逼近最真实的网络告警关联函数,综合运用有监督学习和无监督学习方法,提升智能关联分析方法的性能和效果。

参考文献:

- [1] LI T, LI X. Novel alarm correlation analysis system based on association rules mining in telecommunication networks[J]. Information Sciences, 2010, 180(16): 2960-2978.
- [2] LI J F, WANG H B. Network Fault Alarm Correlation Analysis Based on Association Rule[J]. Computer Engineering, 2012, 38(5): 44-46.
- [3] 刘洪波, 陈刚, 宫钦. 基于神经网络的通信网络告警关联分析及应用[J]. 电信技术, 2018(5).
- [4] 张永华. 基于大数据技术的电信网络告警关联分析与实现[J]. 电信工程技术与标准化, 2016, 29(4): 18-23.
- [5] 钟义信. 人工智能的突破与科学方法的创新[J]. 模式识别与人工智能, 2012, 25(3): 456-461.
- [6] HOPFIELD J J. Neural Networks and Physical Systems with Emergent Collective Computational Abilities [J]. Proc of the National Academy of Sciences of the USA, 1982, 79(8): 2554-2558.
- [7] 余凯, 贾磊, 陈雨强, 等. 深度学习的昨天、今天和明天, 计算机研究与发展[J]. 2013, 50(9): 1799-1804.
- [8] LATHAUWER L D, MOOR B D, VANDEWALLE J. A Multilinear Singular Value Decomposition [M]. Society for Industrial and Applied Mathematics, 2000: 1253 - 1278.
- [9] KOLDA T G, BADER B W. Tensor Decompositions and Applications [J]. Siam Review, 2009, 51(3): 455-500.
- [10] BRAZELL M, LI N, NAVASCA C, et al. Solving Multilinear Systems via Tensor Inversion [J]. SIAM Journal on Matrix Analysis and Applications, 2013, 34(2): 542 - 570.
- [11] KUANG L, YANG L T, WANG X, et al. A tensor-based big data model for QoS improvement in software defined networks [J]. IEEE Network, 2016, 30(1): 30-35.
- [12] KUANG L, HAO F, YANG L T, et al. A Tensor-Based Approach for Big Data Representation and Dimensionality Reduction [J]. IEEE Transactions on Emerging Topics in Computing, 2014, 2(3): 280-291.
- [13] DANNAEUS A. Big Data and Cloud Computing: Current State and Future Opportunities [C]// International Conference on Extending Database Technology. ACM, 2011.
- [14] WANG X, YANG L T, XIE X, et al. A Cloud-Edge Computing Framework for Cyber-Physical-Social Services [J]. IEEE Communications Magazine, 2017, 55(11): 80-85.
- [15] ZHANG Q, ZHU C, YANG L T, et al. An Incremental CFS Algorithm for Clustering Large Data in Industrial Internet of Things [J]. IEEE Transactions on Industrial Informatics, 2017: 1-1.

作者简介:

匡立伟, 毕业于华中科技大学计算机科学与技术学院, 博士, 主要研究方向包括人工智能、机器学习、数据挖掘、网络功能虚拟化、云计算等; 赵良, 工程师, 主要从事传输承载和人工智能相关的技术研究工作; 杨韬, 高级工程师, 主要从事光网络相关的技术研究工作; 张贺, 教授级高工, 主要从事传输承载相关的技术研究工作。