

基于知识图谱的智能故障诊断研究

Research on Intelligent Fault Diagnosis Based on Knowledge Graph

刘瑞宏,谢国强,苑宗港,宋文婷,王高虎(华为技术有限公司,广东 东莞 523808)

Liu Ruihong, Xie Guoqiang, Yuan Zonggang, Song Wenting, Wang Gaohu (Huawei Technologies Co., Ltd., Dongguan 523808, China)

摘要:

介绍了一种基于知识图谱的智能故障诊断系统,通过构建电信领域知识图谱,将电信网络领域零散的专家经验及庞大的产品、案例知识和故障数据有效关联,并应用知识图谱推理技术,进行网络故障智能诊断,辅助解决网络运维领域的故障问题。该系统利用知识图谱解决了故障知识与故障数据自动对接、自动推理生成故障传播链、按图索骥查询故障传播可解释性知识等问题,经过现网真实故障场景验证,可以极大提高故障根因定位分析全过程的处理效率。

关键词:

知识图谱;故障诊断;知识推理

doi: 10.12045/j.issn.1007-3043.2020.10.007

文章编号: 1007-3043(2020)10-0030-06

中图分类号: TN915

文献标识码: A

开放科学(资源服务)标识码(OSID):



Abstract:

A knowledge graph based intelligent fault diagnosis system is introduced. By constructing the knowledge graph of telecommunication domain, the scattered expert experience and huge amount of product and case knowledge and fault data in telecommunication network can be effectively correlated. The reasoning technology of knowledge graph is applied to carry out intelligent fault diagnosis of network, and assist to solve the fault problems in the field of network operation and maintenance. The system uses knowledge graph to solve the problems of automatic connection between fault knowledge and fault data, automatic inference to generate fault propagation chain, and query interpretable knowledge of fault propagation accordingly. The system can greatly improve the processing efficiency of the whole process of fault root cause location analysis through the verification of real fault scenarios in the existing network.

Keywords:

Knowledge graph; Faults diagnosis; Knowledge inference

引用格式: 刘瑞宏,谢国强,苑宗港,等. 基于知识图谱的智能故障诊断研究[J]. 邮电设计技术, 2020(10): 30-35.

1 概述

电信产业一直在探索数字化、自动化和智能化,从转型前期聚焦客户服务、产品业务层,逐步延伸到内部管理运营层,再到网络层。早在2011年,电信产业就希望通过利用SDN、NFV和云技术,提升业务和网络敏捷性的同时,降低成本和复杂性。但从结果来看,基于SDN/NFV技术的网络自动化仍无法完全解决未来各种应用大规模部署、网络新技术引入与扩张带来的问题。如何大规模、全流程地提升效率,并持续

快速迭代地引入新技术仍然是产业共同面临的难题。

自动驾驶网络正是诞生于这一背景下,通过应用多种智能技术,发挥融合优势驱动电信行业从数字化迈向智能化,将对电信产业的生产方式、运营模式、思维模式和人员技能等全方位带来深远影响。

2019年5月,电信管理论坛TMF联合英国电信、中国移动、法国Orange、澳大利亚Telstra、华为和爱立信等成员,合作发布了业界第一部自动驾驶网络白皮书。

图1提出了“单域自治、跨域协同”的3层框架与4个闭环,给产业提供了运营商数字化转型的架构蓝图,给产业各方的实践与合作提供顶层架构参考,并促进产业高效合作。

收稿日期: 2020-08-18

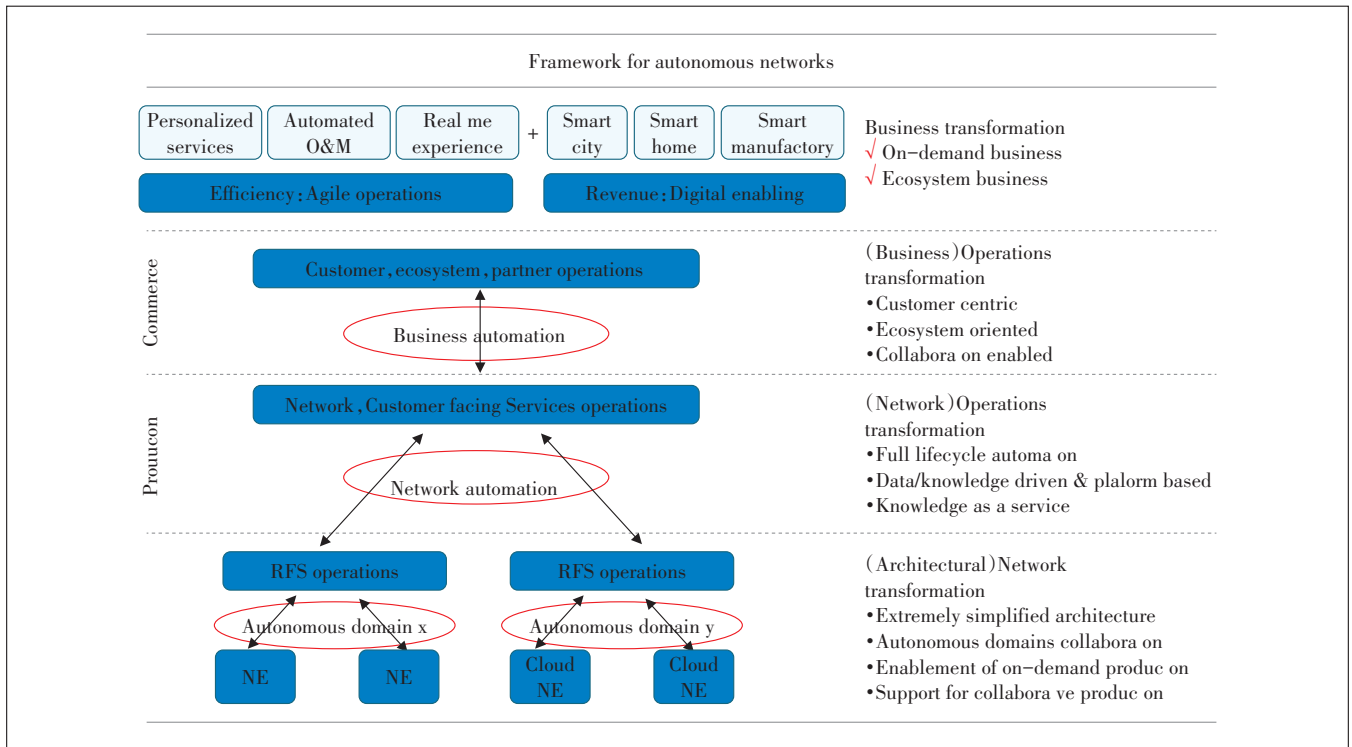


图1 TMF自动驾驶网络目标架构

与此同时,该白皮书还进一步定义了自动驾驶网络L1到L5的高阶分级标准(见图2),为产业的逐级递进给出了高阶参考。

L0- 手工运维:系统提供辅助监控能力,所有动态任务都需要手动执行。

L1-辅助运维:系统根据预先配置,执行某个重复的子任务,提高执行效率。

L2-部分自动驾驶网络:系统基于AI模型,针对确定的外部环境,系统内部分单元实现闭环运维。

L3-有条件自动驾驶网络:在L2的能力基础上,系统能实时感知环境变化,在特定网络领域,能根据外部环境进行自我优化和调整,实现基于意图的闭环自治。

L4-高度自动驾驶网络:在L3层的能力基础上,在更复杂的跨多网络领域环境下分析和决策,系统面向业务和客户体验,实现预测式或主动式的闭环自治。

L5-完全自动驾驶网络:这一层次是电信网络演进的终极目标,具备跨多业务、多领域、全生命周期的全场景闭环自治能力。

Autonomous networks levels						
Level Definition	L0: Manual Operation & Maintenance	L1: Assisted Operation & Maintenance	L2: Partial Autonomous Network	L3: Conditional Autonomous Network	L4: High Autonomous Network	L5: Full Autonomous Network
Execution	P	P/S	S	S	S	S
Awareness	P	P	P/S	S	S	S
Analysis	P	P	P	P/S	S	S
Decision	P	P	P	P/S	S	S
Intent/ Experience	P	P	P	P	P/S	S
Applicability	N/A	Select scenarios				All scenarios

图2 TMF自动驾驶网络5级等级定义

自动驾驶网络的终极目标是通过数据与知识驱动,实现网络自动、自优、自愈,敏捷使能新业务并实现极致客户体验,实现资源和能源利用效率最高,在自动、自优、自愈基础上,网络自我演进实现自治。

知识图谱是自动驾驶网络的关键使能技术之一,在电信领域知识图谱该如何应用呢?

随着通信网络近些年的快速发展,网络连接数和网络数据的迅猛增长,网络结构日益复杂,故障也越来越多。传统故障处理模式难以满足日益增长的网络故障,主要表现在以下3个方面。

a) 传统运维依赖运维专家经验,但运维专家经验差异大,想要准确推导出故障根因,对运维人员的要求比较高。

b) 成千上万运维人员处理大量重复性问题,运维经验积累慢,处理效率低。

c) 运维经验相关文档分散,命名不规范,不便于运维人员查找。

由于这些问题的存在,自动化和智能化的网络故障诊断技术得到了广大研究者的关注,并致力于降低门槛,提高运维人员工作效率。针对这个问题,本文设计了一种智能化的故障诊断系统,并命名为基于知识图谱的智能故障诊断系统,融合电信领域专家经验及文档知识,构建专属电信领域的知识图谱,利用知识图谱的表示、推理、人机交互技术实现智能化根因推导,构建电信网络领域故障诊断通用能力,降低运维门槛,提高运维人员工作效率。

2 方案原理

该系统参照医生对患者的疾病诊疗过程,应用网络知识图谱技术对网络故障进行诊断,分析、定位网络故障根因。接下来从3个方面进行阐述:系统原理、构建知识图谱、应用知识图谱诊断网络故障。

2.1 系统原理

图3示出的是医生对患者疾病诊疗过程示意图。

医生基于经验,根据病人基本信息和症状,逐步分析、检验、诊断出病因,实施治疗方案。如果基于经验和检查信息无法确定病因或诊断不足以确诊,就需要增加检验手段,甚至查阅相关材料(如疾病文献),来分析病症,找到解决办法。

图4示出的是网络故障智能化辅助诊断过程示意图。

网络系统相比人体,可监测性、可度量性更好,可

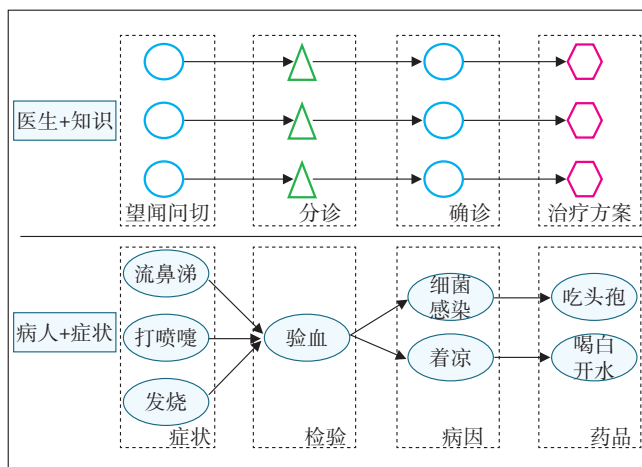


图3 医生对患者疾病诊疗过程示意图

获得的检验信息更全面,智能故障运维辅助系统就像给机器看病的医生,根据网络基本信息和故障数据,分析故障现场、查看故障现象与趋势、推理故障根因、提供故障处理建议。如果基于有限的故障信息无法推断根因,就需要增补采集故障信息,补充故障知识,甚至查阅故障相关案例,来辅助判定分析根因,找到解决方案。

2.2 构建知识图谱

网络故障知识图谱的构建过程如图5所示,主要包括4个阶段。

阶段1,在网络故障领域专家的指导下,进行网络领域概念建模,包括概念、概念属性、概念间的层次关系。建模过程中需要表达网元基础信息、网络运行信息、网络故障事件信息、故障维护知识的核心概念和关系,输出Schema定义。

阶段2,基于阶段1定义的Schema,需要结合结构化、半结构化、非结构化的网络数据和产品文档、故障案例文档等语料分析数据源,进行知识抽取。

阶段3,故障知识融合阶段对前期抽取结果进行实体对齐、类型对齐、属性对齐,建立关联关系,为了保障知识融合的正确性,将电信领域术语表、规范作为知识抽取的同义概念和同义关系参照标准,使得抽取结果具备统一表示、统一理解的基础。

阶段4,融合的知识结果按照网络领域和网络应用场景分类组织、存储、发布,就可以形成各领域网络故障知识库,支撑各领域网络故障诊断应用。

2.2.1 知识表示及建模

在电信领域定义“实体-属性-属性值”“概念-关系-概念”以事件为核心的知识图谱。图谱表示技术

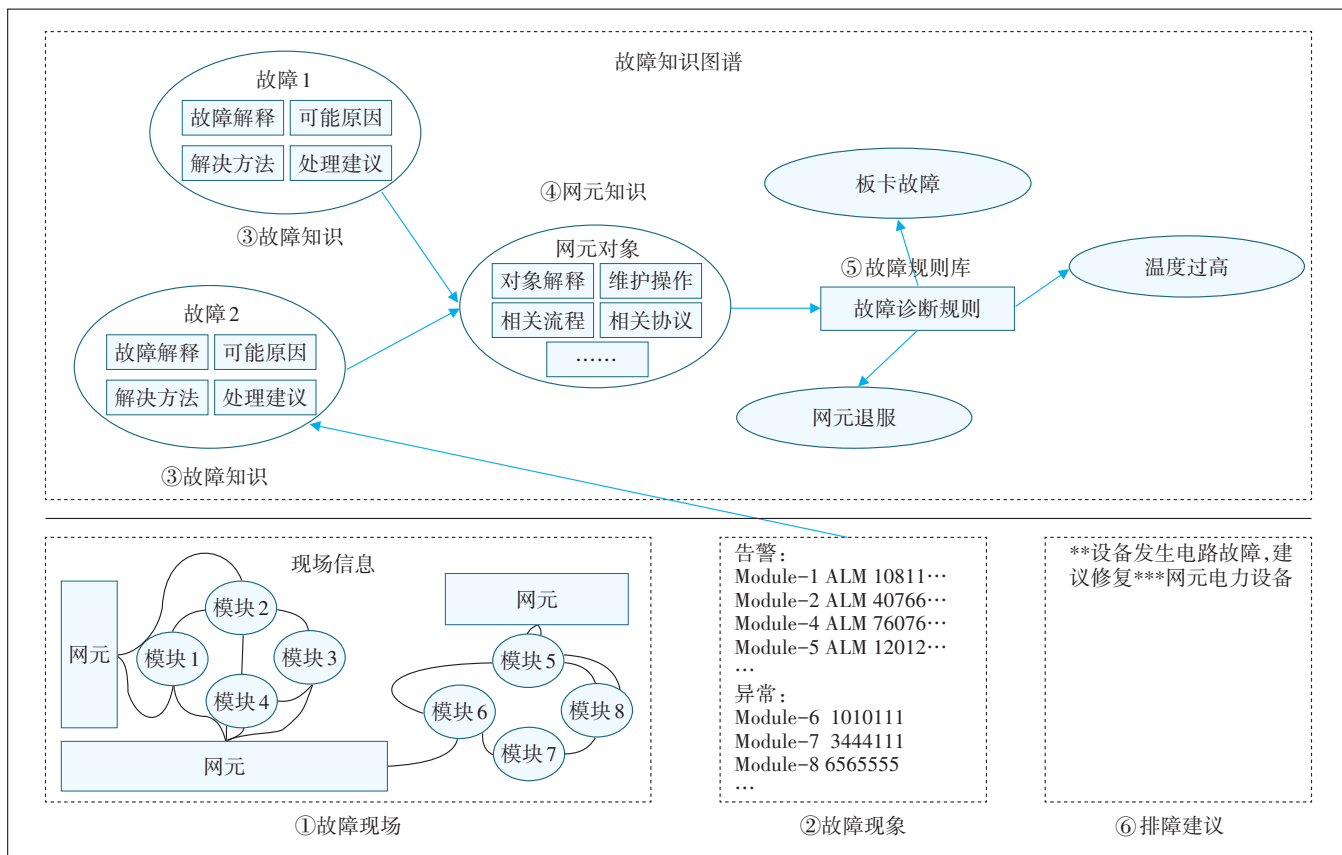


图4 网络故障智能化辅助诊断过程示意图

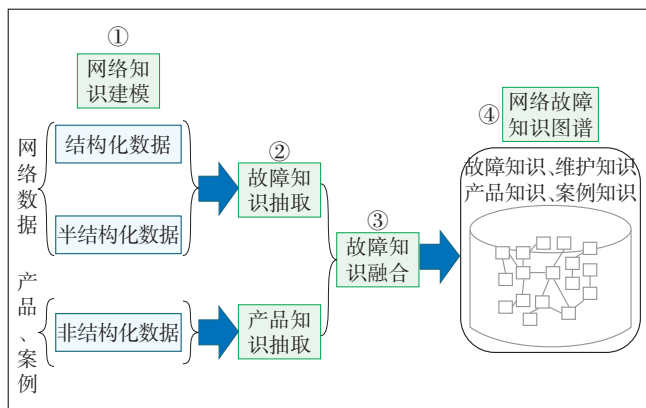


图5 网络故障知识图谱构建过程示意图

实现了对网络概念、实体、属性的表示,还可以对动态时序数据(如通信流程、故障处理工序、故障事件组合等领域知识)进行表示。

通过“名词+动词”的知识表示方法来描述故障事件。如图6所示,“SIP TCP链路拥塞”事件中,名词实体对象为SIP TCP,动词为拥塞;“链路所属的局向部分链路故障”导致“SIP TCP链路拥塞”形成故障事件组合。

2.2.2 知识抽取

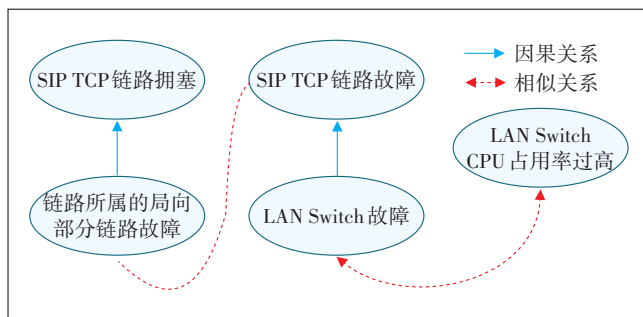


图6 故障模式

电信领域中包含了大量领域文献及专有术语,这些语料大致分为结构化、半结构化和非结构化3种类型。结构化数据抽取比较成熟,可以根据元数据与图谱Schema进行数据映射直接抽取。对于半结构化和非结构化需要根据语料内容和结构,设计相应的抽取算法和包装器,知识抽取的整体框架如图7所示。

基于文本的“实体-关系-实体”联合抽取,目前有2类方法:一种是使用流水线的方法(Pipelined Method)进行抽取:输入一个句子,首先进行命名实体识别,然后对识别出来的实体进行组合,再进行关系分类,输出“实体-关系-实体”三元组。流水线的方法主要存

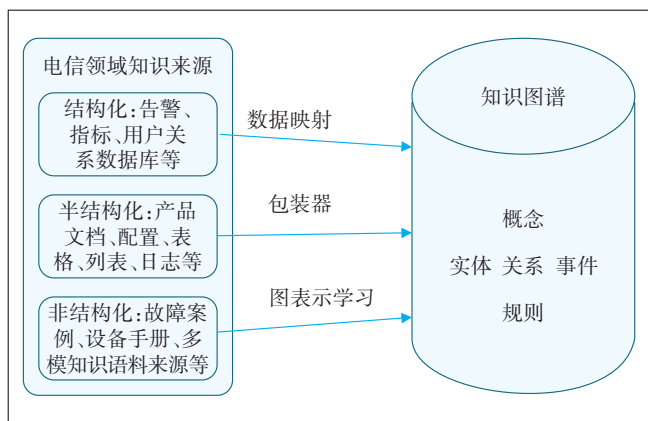


图7 网络故障知识图谱抽取框架

在以下缺点。

a) 错误传播,实体识别模块的错误会影响到下面的关系分类模块准确性。

b) 没有利用抽取子任务之间存在的语义关系,如果在图上存在 Alarm-Object 关系,那么可以知道 Alarm 必然属于 Object,后一个实体属于 Object 类型,流水线的方法没法利用这样的信息。

c) 产生了没必要的冗余信息,由于对识别出来的实体进行配对,然后再进行关系分类,那些没有关系的实体对就会带来多余信息,错误率高。

本系统应用的联合抽取算法:输入一个句子,通过实体和关系联合抽取算法模型,直接得到“实体-关系-实体”三元组。这种方法可以克服上面流水线方法的缺点。

知识抽取算法模型采用预训练语言模型,使用电信领域语料进行重训练对领域文本语句进行分布式表示,进行非对称核化关系分类,同时抽取“实体-关系-实体”,相对于通用预训练语言模型,抽取准确率提升10%以上。

2.2.3 知识融合

对于从多个语料来源抽取的知识,需要进行知识融合,以减少知识库冗余。知识融合的基础是实体对齐,进行实体对齐、类型对齐、属性对齐,建立关联关系,可以保障知识融合的正确性。当前通用的知识融合技术存在如下2个痛点。

a) 计算复杂度高:2个知识库,为发现所有匹配对,需要将一个知识库中的所有实体和另一个知识库中的所有实体逐一对比。

b) 图谱表示差异大:相同实体有不同名字,相同名字指代不同实体,实体定义的粒度不同,相同属性在

不同的知识库中有不同的判别能力,相同实体在不同知识库中的属性数量不一致。

该系统提出了一种面向电信领域的知识融合框架,如图8所示,基于图谱的表示学习及实体、属性、关系的迭代联合学习,可以自动学习到更深层次的知识特征,减少语义漂移,提升实体对齐准确率。

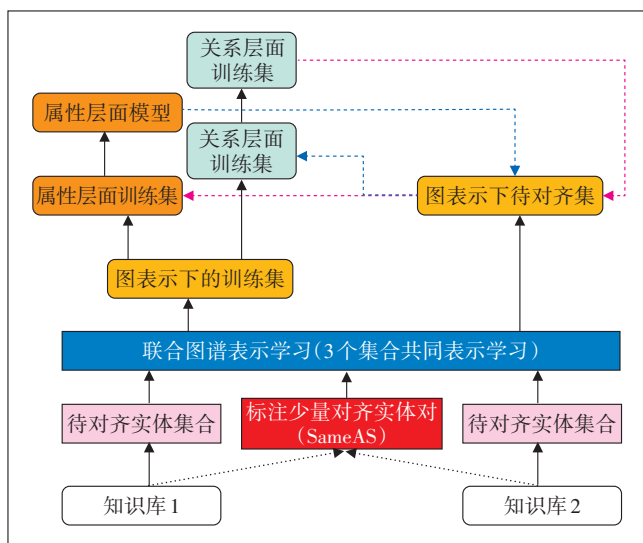


图8 知识融合框架

2.3 应用知识图谱

基于构建好的知识库,可以支持上层应用的设计开发。结合网络领域故障诊断场景的应用诉求,应用系统的重点能力主要包括两大功能:知识推理和知识查询。

2.3.1 知识推理

图9示出的是知识图谱推理框架。

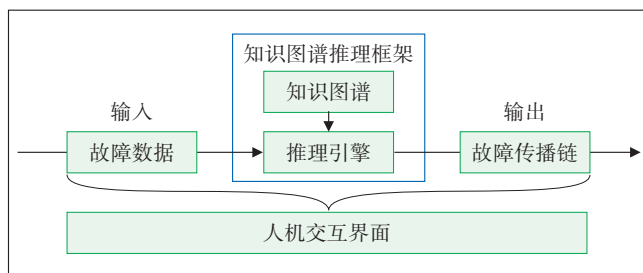


图9 知识图谱推理框架

现网发生故障时产生海量的故障事件信息,如异常日志、告警等,知识推理框架从故障数据中自动分析故障间传播规律,生成推理结果。在此过程中,领域运维专家经验是重要的知识来源,需要充分利用、有效转化;知识推理利用图存储中的产品知识和规则知识,按

策略编排自动执行;规则知识的自动学习发现也对专家经验起到了有效补充作用,进一步提升了知识推理框架的能力,支撑更丰富的故障场景的自动适配;利用NLP知识自动抽取技术抽取产品文档、案例文档等不同语料中的产品知识,用于知识查询,进一步增强了针对故障传播关系知识查询的可解释性。

2.3.2 知识查询

在故障诊断过程中,会用到大量的机器数据、产品文档、案例文档等知识,为了能够更好地引导用户利用知识库进行故障诊断,需要具备便捷的查询能力。

该系统结合运维领域故障分析思路,总结出几个关键分析步骤,逐层深入解决故障问题(见图10)。

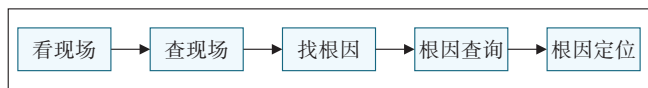


图10 网络故障知识查询流程

a) 看现场:当分析网络故障问题时,首先需要还原故障现场,看网元、版本,现场是否存在高危操作,是否有过重大升级,因为这些操作都有可能造成故障的根因。

b) 查现象:通过上一步的现场排查,接下来就要看故障的具体现象,主要检查是否有高危告警、异常性能指标等故障事件产生。

c) 找根因:确定重点关注的故障事件后,结合故障现场数据和故障知识,自动推导生成故障传播链,推导出故障的根因事件。

d) 根因查询:若当前的根因事件还不足以表达故障的真实根因,还可通过查询知识库内的相关基础知识、案例,发现可能原因,进一步获取故障的解决方法。

3 应用示例

当前该系统已在几个关键电信领域进行了测试验证,可针对关键业务性能指标劣化、资源故障异常等场景进行辅助诊断和定位。

以某局点“移动软交换业务异常”为例,应用该系统进行故障诊断,分4个步骤验证了该系统诊断定位故障根因的有效性。

a) 还原故障现场:导入故障数据,查看网元版本信息及配置操作,完成故障现场信息的还原,从全局了解故障现场情况。

b) 分析故障现象:结合故障趋势分析,选择故障发生时段,明确故障现象为“位置更新请求次数下降”。

c) 推导根因:结合故障数据和故障知识,自动推导出故障传播链,发现问题根因是“告警IP地址流量不均衡”。

d) 定位故障:通过知识查询,得出“告警IP地址流量不均衡”的故障处理步骤,逐一进行产品知识查询、故障案例查询,排除该网元故障的各种可能原因,最终验证确认该问题是由“IP承载网环路故障”引起的。

4 结束语

综上所述,基于知识图谱的智能故障诊断系统,可以综合运用图谱的知识表示、知识推理、知识查询等技术,将零散的人工经验及庞杂的文本知识有效进行关联,实现故障根因推理定位,辅助解决网络故障运维领域的问题。当前该方案仍存在一些不足和改进空间,主要包括:一是知识表示的范围上除故障知识,还可以针对规、建、维、优多环节知识进一步完善;二是领域故障知识还不完备,特别是关系知识不完备直接影响到故障根因推导的准确性。后续还需逐步完善,达到网络诊断故障自优自治自维护的自动网络驾驶目标。

参考文献:

- [1] DONG X, GABRILOVICH E, HEITZ G, et al. Knowledge vault: A web-scale approach to probabilistic knowledge fusion [J]. in SIG-KDD.ACM, 2014(3):601-610.
- [2] AHMED M, MAHMOOD A N, HU J. A survey of network anomaly detection techniques [J]. Journal of Network & Computer Applications, 2016, 60(2):19-31.
- [3] BOSNEAG A M, HANDURUKANDE S, WANG M. Fighting Fire with Fire: Survey of Strategies for Counteracting The Complexity of Future Networks Management [C]// International Icinn Conference-innovations in Clouds, 2016.
- [4] PATRICK J D, NGUYEN D H M, YEFENG W, et al. A knowledge discovery and reuse pipeline for information extraction in clinical notes [J]. J Am Med Inform Assoc, 2011(5):574-579.
- [5] 苏佳林, 王元卓, 靳小龙, 等. 融合语义和结构信息的知识图谱实体对齐 [J]. 山西大学学报, 2019, 42(1):23-30.

作者简介:

刘瑞宏,毕业于北京大学,华为网络P&S知识图谱TIG主任,硕士,主要从事网络AI知识图谱、数据湖、大数据、数据仓库等规划、设计工作;谢国强,毕业于西安电子科技大学,学士,主要从事自动驾驶网络智能、体验和人机交互领域的技术研究和规划工作;苑宗港,毕业于东南大学,ICT系统架构师,硕士,主要从事知识图谱的架构设计、电信领域知识表示&知识推理、大数据与知识工程等工作;宋文婷,毕业于西安电子科技大学,硕士,主要从事知识图谱关系抽取、意图理解等设计工作;王高虎,毕业于中国民航大学,工程师,硕士,主要从事电信领域自然语言及知识图谱分布式表示学习、知识抽取、图计算等算法研究及开发工作。