基于人工智能的网络告警关联分析处理的应用

Application of Network Alarm Association Analysis Processing Based on Artificial Intelligence

陆　斌[1]，华　楠[1]，郑小平[1]，陈文军[2]（1. 清华大学电子工程系，北京 100084；2. 烽火通信科技股份有限公司，湖北 武汉 430074）

Lu Bin[1]，Hua Nan[1]，Zheng Xiaoping[1]，Chen Wenjun[2]（1. Tsinghua University，Beijing 100084，China；2. FiberHome Telecommunication Technologies Co.，Ltd.，Wuhan 430074，China）

摘　要：
随着网络快速发展，网络告警日趋复杂，规模急剧增大，传统处理模式难以满足需求。提出了基于人工智能的网络告警分析方法，使用聚类和数据挖掘进行规则挖掘，基于告警规则进行网络告警信息关联分析处理。

关键词：
网络告警；关联分析；人工智能
doi：10.12045/j.issn.1007-3043.2018.12.001
中图分类号：TP389.1
文献标识码：A
文章编号：1007-3043（2018）12-0001-06

Abstract：
With the rapid development of the network，network alarms are becoming more and more complex，and the scale is rapidly increasing. The traditional processing mode is difficult to meet the demand. A network intelligence analysis method based on artificial intelligence is proposed，which uses clustering and data mining for rule mining and carries out network alarm information correlation analysis processing based on alarm rules.

Keywords：
Network alarm；Association analysis；Artificial intelligence

引用格式：陆斌，华楠，郑小平，等. 基于人工智能的网络告警关联分析处理的应用[J]. 邮电设计技术，2018（12）：1-6.

1 概述

随着通信网络近些年的快速发展，其规模已经相当庞大，在网络中每天都会产生告警信息，并且这些信息数据量庞大、突发故障多，当网络设备出现故障并引发告警时，与它关联的设备也会引发相应的故障，并在短时间内产生大量告警信息[1-2]。由于一个故障的产生往往会引发多个告警事件，与故障相关的设备以及业务过程都会发出相关的告警信息，同时多个故障引发的告警信息会叠加到一起，把真正的告警信息淹没在里面，导致故障识别十分困难。当前网络告警的监控和管理主要依靠人工完成，网络运营维护成本高昂，处理过程十分耗时，在发生大量故障告警时基本不能满足告警处理的实时性要求。

告警相关性分析是网络故障诊断的重要方法之一，告警相关性分析采用的方法有很多，例如基于规则的告警相关性分析、基于事例的告警相关性分析、基于因果模型的相关性分析、基于神经网络的相关性分析等。但是这些方法都存在一定的缺点，例如基于事例的方法对于网络变化处理反映不敏感，这是因为它由特定应用领域决定，而不存在一个各个领域通用的事例，基于规则的方法难以适应大规模和技术复杂

收稿日期：2018-10-31

的通信网络模型,这是因为它需要人工来维护大量告警规则。目前我国的综合网络管理市场上,大部分相关产品,都提出了告警相关性分析的支持功能,也有部分公司表明其产品实现了告警相关性的分析,但其实只是实现了一些比较基础的告警过滤、告警规避等功能,规则的生成有待进一步实验研究[3-7]。

本文提出一种使用规则挖掘及基于规则的关联方法[8],主要是将当前告警系统领域的告警知识包含在一组规则集合中,通过对检测到的告警进行判定,并使用相应的推理规则来分类一个或者多个告警的发生是否符合某一个规则,进而确定具体的故障类型[9]。其工作原理主要是依赖于规则库和推理引擎。在规则库中将很多实践中获取的知识通过适当编码形成IF-THEN式的规则,递归向下匹配规则,定位最终的故障源。当发生新的告警时,系统将启用推理引擎来对告警进行处理。这种方法表现形式单一、直观,所以不需要长时间的培训学习,也不需要了解网络的底层架构,就可以定位网络中发生的故障。

2 网络告警分析处理系统设计

告警分析处理分为2个阶段:告警规则挖掘阶段、告警分析处理阶段。告警规则挖掘阶段是为了实现基于历史告警数据的大数据分析,从历史数据中获得告警之间的关联规则,形成规则数据库;告警分析处理阶段目的是基于所获得的规则数据库中的关联规则,对网络中的当前告警进行分析和处理,获得当前告警中的根源告警及衍生告警。

告警规则挖掘阶段采用离线处理的方式,对历史数据进行分析和挖掘,不要求实时性。初次部署时,获取大量网络历史告警,进行规则挖掘初始化,形成规则数据库,在网络中部署后,采用定期挖掘规则的形式,对规则数据库进行增量更新和补充。

告警分析处理阶段采用在线处理的方式,对当前告警进行处理,要求实时性。在软件部署后,便通过网管后台接口与网管进行通信,实时对网络告警进行处理。

图1给出了告警分析处理系统总体架构。

2.1 告警数据统计分析

本文对网络告警类型出现频次做了初步统计,如图2所示。从分布图可以看出,告警信息集中在少数几种类型中,例如PK_LOS,RCONTEXT_PACKET_LOS,VP_RDI,RCONTEXT_PACKET_LOS,E1_AIS等告警类型占据了约90%以上的信息,而M_BCFGRDIF,VP_MMG,MANUAL_SWITCH等告警类型所占比例远远小于1%,告警类型不均匀分布给后期的分析处理带来了极大的挑战。

2.2 告警规则挖掘阶段

告警规则挖掘阶段的示意图如图3所示。

a)由IPRAN的网管系统导出历史告警数据文件,作为规则挖掘所学习的数据。

b)数据预处理,读取历史告警数据后,检测所有数据的有效性,筛除其中无效数据,并对告警数据进行编码,导入到告警数据库中;告警数据库中同时导入网络拓扑、业务信息、告警层次信息等。

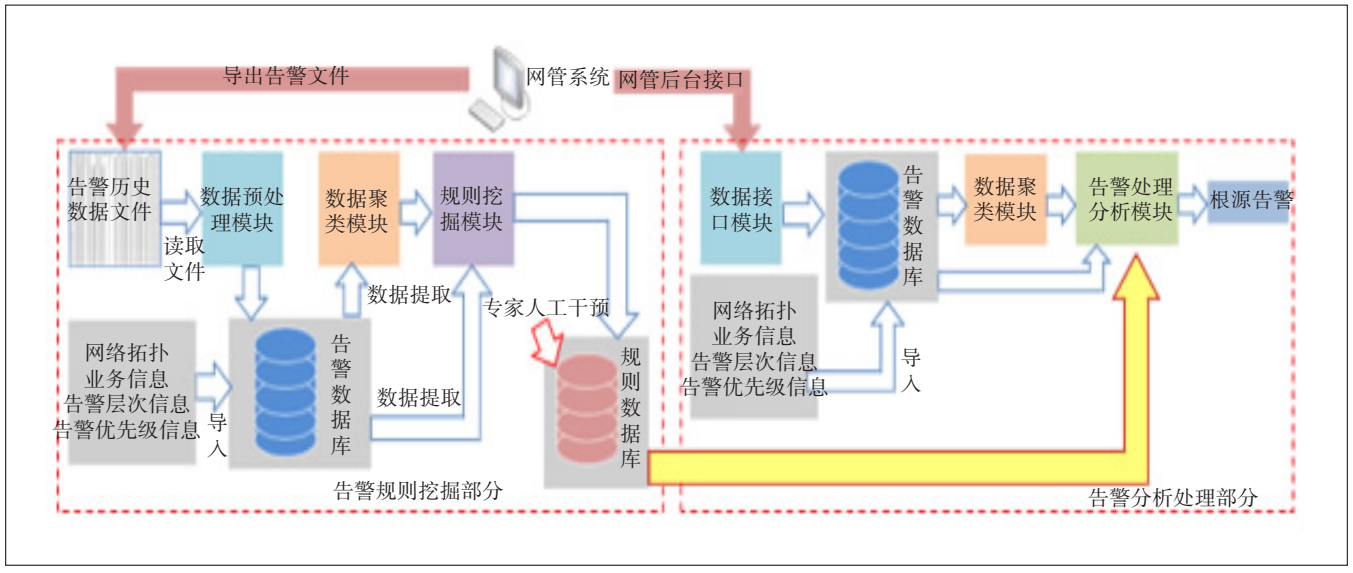

图1 告警分析处理系统总体架构

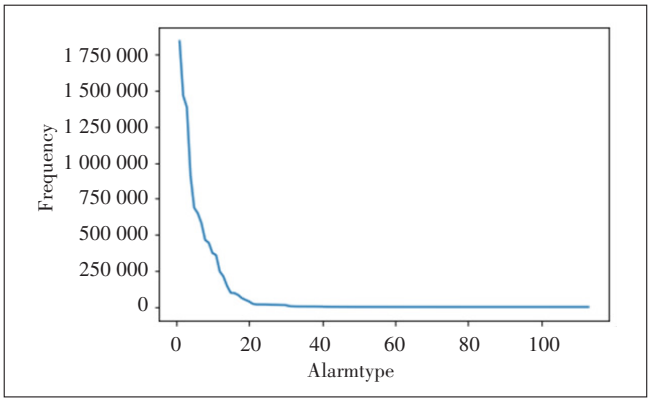

图2　告警类型频次分布图

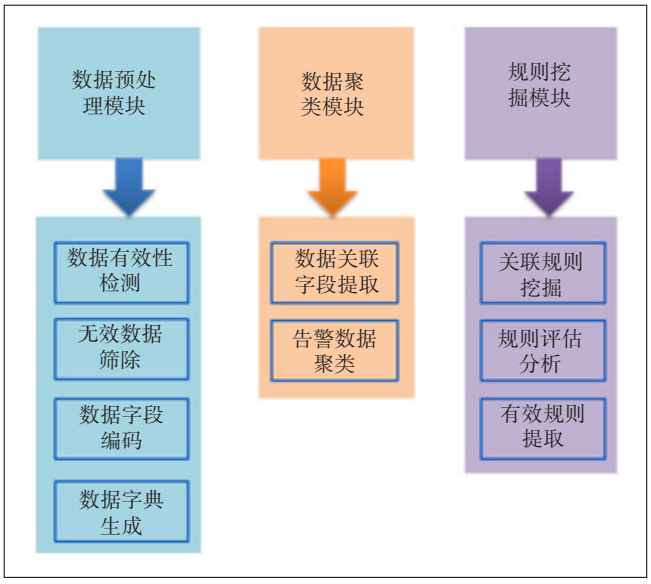

图3　告警规则挖掘

c）数据聚类，从告警数据库中提取出聚类所需数据关键字段，进一步实现对告警数据的聚类，将数据在时域和地理位置进行划分。

d）规则挖掘，获取聚类结果，同时从告警数据库中提取告警数据，对每一簇告警数据进行关联分析，实现规则挖掘。

e）将挖掘出的规则导入到告警数据库中，经过专家的人工干预，实现有效规则的筛选。

2.3 告警分析处理阶段

图4给出了告警分析处理。

a）由IPRAN的网管系统经过后台接口，将数据传入数据接口中。

b）数据接口读取到当前告警数据后，经相应处理导入到告警数据库中；告警数据库中同时导入网络拓扑、业务信息、告警层次信息等。

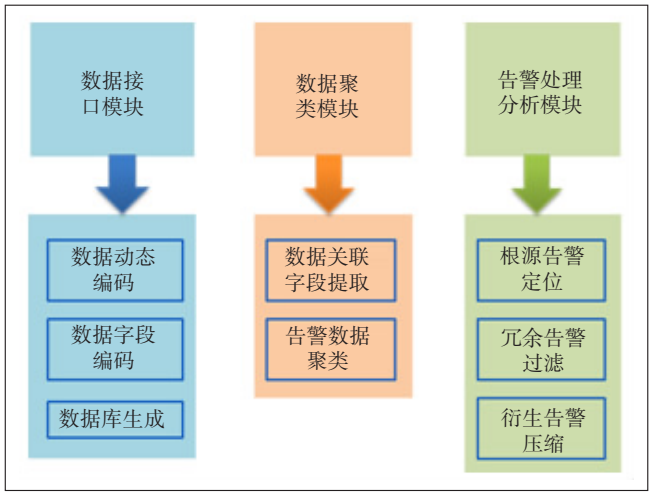

图4　告警分析处理

c）数据聚类从告警数据库中提取出带分析告警数据的关键字段，进一步实现对当前告警数据的聚类，将数据在时域和地理位置进行划分。

d）告警处理分析，获取聚类结果，同时从告警数据库中提取告警数据。从规则数据库中遍历所有告警规则，对每一簇告警数据进行分析，得到根源告警，实现告警压缩。

3 网络告警关联分析处理方法

3.1 数据聚类

聚类属于无监督的机器学习方式。聚类根据未知标签样本的数据集内部的数据特征，将数据集划分成多个不同的类，使得同一类的数据样本尽可能地相似，不同类的数据样本之间相似度尽可能地小。传统的关联规则分析在统计告警信息时，往往是用时间硬滑窗之后再进行统计，但是时间硬滑窗不能充分利用信息，有可能把过多的告警放入一个类，或者把本来属于同一个故障的告警切成了不同的类，这样就会把不同根源告警及其衍生告警混淆到一起，统计结果精确度不够，所以提出先对告警信息聚类，根据告警信息的数据属性把不同的根源告警及其衍生告警区分开来，即每一类代表一个根源告警及其衍生告警，然后再做关联规则分析，精确度能提高。实验中主要做了聚类方案，基于地点和时间的信息聚类（见图5）。

基于地点和时间信息聚类：利用准确的地点信息（例如网元），对告警数据进行“硬划分”；利用告警的开始时间及结束时间，使用DBSCAN算法[10-11]在时间维度进行聚类。

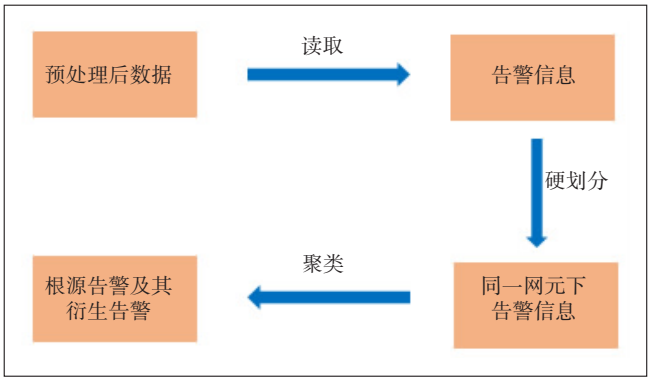

图5 聚类示意图

3.2 规则挖掘

规则挖掘的方案设计如图6所示。

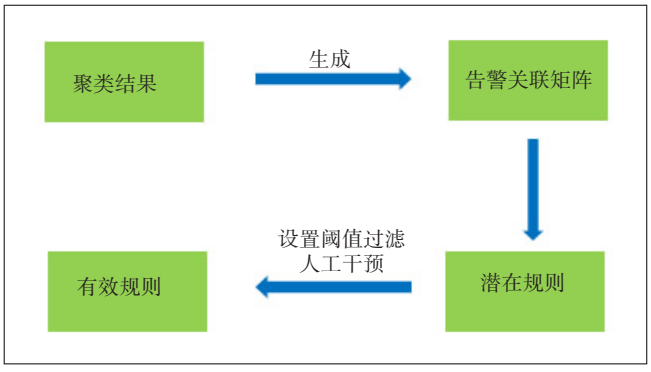

图6 规则挖掘示意图

告警规则挖掘主要分为3个过程,首先是由聚类结果生成告警关联矩阵(ACM——Alarm Correlation Matrix)[13-15],然后从告警关联矩阵中挖掘出潜在规则,通过设置阈值过滤或者人工干预的手段,得到有效的告警规则。下面针对上述3个过程进行详细描述。

3.2.1 聚类结果生成告警关联矩阵

对海量告警信息进行聚类之后,在聚类结果中挖掘告警信息之间的关系。挖掘2个告警之间的关系时,使用2个评价指标:支持度和置信度。支持度指的是有序告警对(a→b)在聚类结果中出现的次数,即关联频次。置信度指的是在告警a出现的前提下,告警b接着出现的条件概率。同时,还定义了后件置信度,后件置信度的提出是为了解决置信度忽略规则后件中项集的支持度,前件置信度a→b统计结果100%,则说明有a就一定有b,后件置信度为100%,就说明b前面一定有a,当前件置信度很低,但是后件置信度很高时,认为规则a→b也是有效的。所以引入了后件置信度[12],即在告警b被发现的前提下,告警b由告警a导致的条件概率。支持度可以用来衡量有序告警对(a→b)出现是否频繁,而置信度和后件置信度则用来说明有序告警对(a→b)之间的关联强度。置信度和后件置信度越高,说明有序告警对(a→b)之间的关联强度越大。

得到了聚类结果之后,在聚类结果中分别统计支持度、置信度、后件置信度。统计完成之后,可以得到3个告警关联矩阵,在这个方阵中,元素的行代号代表在前的告警,元素的列代号代表在后的告警。如表1红色数字所示,代表的是04告警→02告警的支持度。

表1 告警关联矩阵示意图

	01	02	03	04	05
01	15 066	110	20	280	0
02	119	31 941	9 608	708	44
03	22	8 936	8 455	150	12
04	455	930	239	1 244	0
05	0	51	15	0	40

3.2.2 告警关联矩阵挖掘潜在规则

告警关联矩阵中信息较多,为了筛选出潜在的规则,定义2个参数:衍生强度derive和后件衍生强度bderive,用来衡量有序告警对(a→b)的衍生强度。衍生强度和后件衍生强度的公式如下。

$$\mathrm{derive}(a \to b) = \frac{\mathrm{conf}(a \to b)}{\mathrm{conf}(b \to a)} \quad (1)$$

$$\mathrm{bderive}(a \to b) = \frac{\mathrm{bconf}(a \to b)}{\mathrm{bconf}(b \to a)} \quad (2)$$

衍生强度和后件衍生强度基于的假设是:告警之间不能两两互推,如果存在(a→b),就不存在(b→a)。如果衍生强度或者后件衍生强度大于1,则a→b要比b→a更加可信,更加符合统计规则。

从式(1)和(2)中可以看出,告警之间的自推是没有的,因为derivea→b和bderivea→b都会等于1,会被过滤掉。根据上述的原则,可以得到潜在的告警关联规则。置信度包括后件置信度说明的是a→b的关联强度,值越大说明关联强度越大。但并不能完全说明a能推导出b,因为在这种情况下b→a的置信度包括后件置信度也有可能很高。为了避免一部分有效的规则被过滤掉,在由告警关联矩阵挖掘潜在高级规则时,算法当中的一些阈值可以设定得低一些,那么得到的潜在规则就会相应地多一些。

3.2.3 设置阈值过滤及人工干预得到有效规则

在得到潜在规则之后，通过进一步设置阈值过滤或者人工干预得到有效的规则。通过人工检查可以将一部分错误的告警规则剔除，进一步提高告警规则的准确度。

在实际应用中，潜在的告警规则规模可能会比较大，人工的检查工作量很大。为了降低人工检查的工作量，可以将阈值提高，进一步缩小潜在的告警规则规模。但是阈值提高越多，被剔除的有效规则也就越多，需要对两者进行权衡。

3.3 告警处理分析

告警处理分析采用先聚类后过滤的处理方法，具体的处理分析步骤分为：

a）根据有效的关联规则，形成告警关系层级和根源衍生告警关系。

b）根据聚类后的告警数据，对每一条告警判断其是否与其根源告警并存，若有则该告警被排除，若不是则该告警保留，直至遍历所有告警数据后得到根源告警数据集。

c）将被排除的衍生告警添加标记，将根源告警添加标记后上报网管实现进一步过滤。

得到了一系列的有效规则之后，按照有序的顺序把这些告警标出来，就会得到一个有向图。

3.4 告警处理分析示例

告警分析处理，输入是一系列的告警，这些告警经过了聚类的处理，将同一个故障引起的告警尽可能地放在一个聚类中。在仿真过程中使用的样本数据如表2所示。

根据告警信息的时间属性进行聚类，得到如表3所示结果。

表2　告警数据样本

SeqNo	AlarmName	StartTime	SeqNo	AlarmName	StartTime
1	VC_LOC	2018-01-03 10:51:06	7	VC_LOC	2018-01-03 09:22:55
2	VP_LOC	2018-01-03 10:51:08	8	VP_LOC	2018-01-03 09:22:51
3	VC_LOC	2018-01-03 10:51:07	9	VC_LOC	2018-01-03 09:22:50
4	VP_LOC	2018-01-03 10:51:08	10	VP_LOC	2018-01-03 09:22:49
5	R_LOS	2018-01-03 10:51:09	11	R_LOS	2018-01-03 09:22:49
6	LASER_TF	2018-01-03 10:51:06	12	OTRX_ABSENT	2018-01-03 09:22:47

表3　聚类结果

Cluster	SeqNo
C1	1,2,3,4,5,6
C2	7,8,9,10,11,12

然后根据规则树对不同类中的告警信息判断根源告警和衍生告警，规则树是在告警规则挖掘阶段根据正确规则产生的，每个节点表示告警类型，例如VP_LOC，VC_LOC等，节点之间的有向线段VP_LOC → VC_LOC表示告警VP_LOC可以导致告警VC_LOC产生，如图7所示。

根据规则树挖掘的最终结果如表4所示。

4 结束语

告警信息体量已经达到了大数据规模，处理方法也应该与时俱进。本文提出的聚类方法进行规则挖掘，基于规则的关联分析对网络告警信息的分析处理适应了告警信息体量剧增的现象。网络告警信息处理的问题已经无法单纯依靠人工来解决，必须结合人工智能的方法来处理，这也提出了更高要求。

参考文献：

[1] 邓歆，孟洛明．智能分布式通信网告警相关性模型及实现[J]．电子与信息学报，2006，28(10)：1902-1905
[2] 朱秋艳．基于关联规则挖掘的网络告警关联[D]．北京：北京邮电大学，2008.
[3] 于漫．电信网络智能化告警系统研究与实现[D]．长春：长春工业大学，2010.
[4] 黄宇．关联规则分析在电信告警系统中的研究与应用[D]．成都：电子科学技术大学，2007.
[5] 王仲佳．具有动态加权特性的关联规则算法及其在电信故障告警序列模式发掘中的应用[D]．长春：吉林大学，2005.
[6] 杨一兵．移动通信网络告警及其关联分析[D]．哈尔滨：哈尔滨工程大学，2008.
[7] 刘斌．移动通信网络故障告警关联分析方法与系统实现[D]．长沙：中南大学，2009.
[8] HARRISON K A. Event Correlation in Telecommunication Network Management[Z]. INCL HP Labs，1994.
[9] STERRITT R，BUSTARD D，MCCREA A . Autonomic computing correlation for fault management system evolution[C]// IEEE Interna-

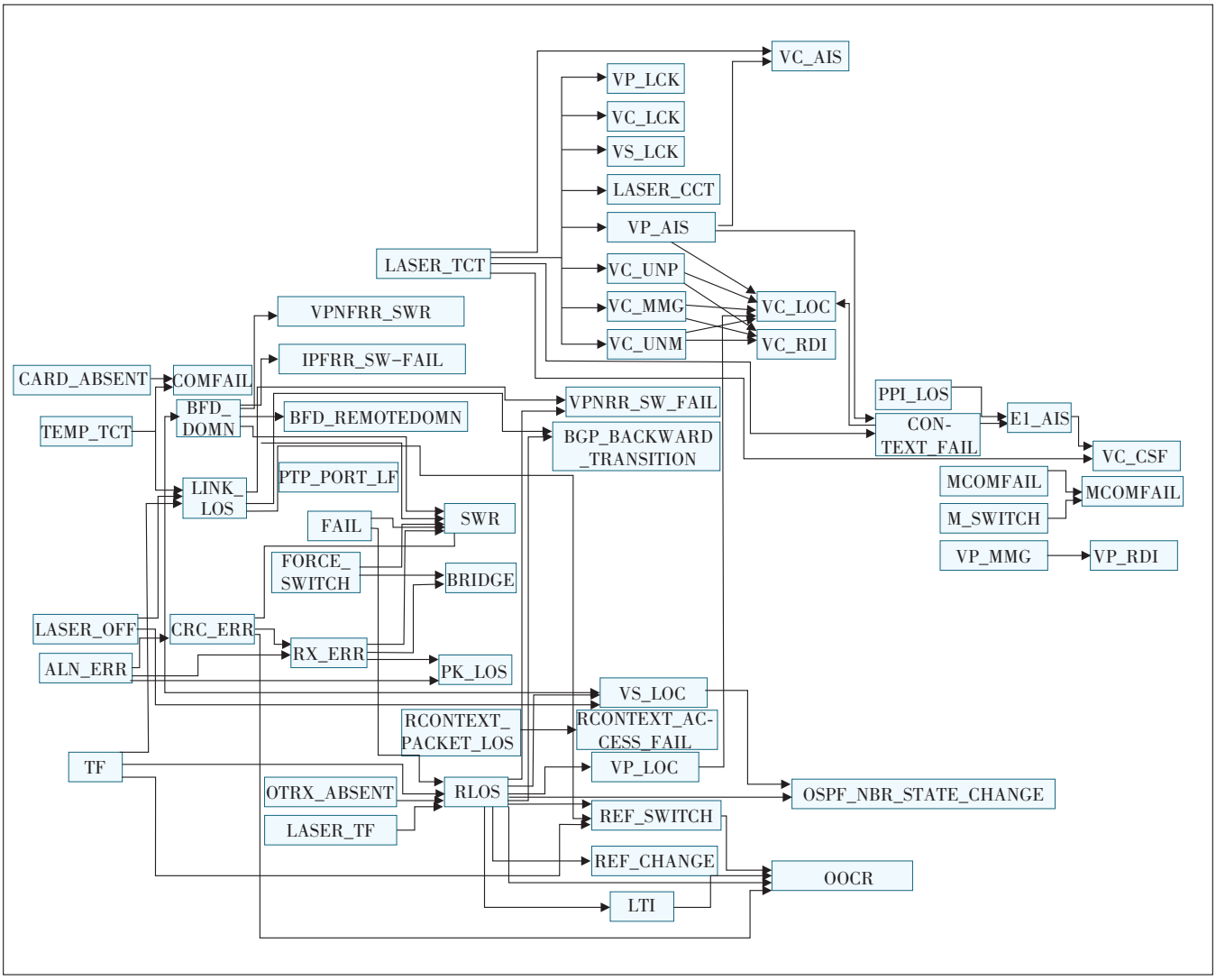

图7 根据正确规则描绘的规则树

表4 根据规则树挖掘的最终得到结果

类别	C1	C2
根源告警	LASER_TF	OTRX_ABSENT
衍生告警	R_LOS,VP_LOC,VC_LOC	R_LOS,VP_LOC,VC_LOC

tional Conference on Industrial Informatics. IEEE,2003.

[10] ESTER M,KRIEGEL H P,XU X. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise[C]// International Conference on Knowledge Discovery and Data Mining. AAAI Press,1996:226-231.

[11] OLIVEIRA D P D,JR J H G,SOIBELMAN L . A density-based spatial clustering approach for defining local indicators of drinking water distribution pipe breakage[J]. Advanced Engineering Informatics,2011,25(2):380-389.

[12] ZHU B,GHORBANI A A. Alert Correlation for Extracting Attack-Strategies[J]. International Journal of Network Security,2006,3(3):244-258.

[13] SKINNER K,VALDES A. Probabilistic Alert Correlation[J]. Proceedings of Recent Advances inIntrusion Detection,4th International Symposium,(RAID 2001),LNCS2212:54-68.

[14] A Toolkit for Intrusion Alert Analysis[EB/OL].[2018-08-11]. http://discovery.csc.ncsu.edu/software/correlator/ver0.4/index.html.

[15] NING P,CUI Y. An Intrusion Alert Correlator Based on Prerequisites of Intrusions[M]. North Carolina State University at Raleigh,2002.

作者简介:

陆斌,清华大学硕士在读,主要从事光网络方面的研究;华楠,毕业清华大学,副研究员,硕士生导师,主要从事智能光网络管控及交换方面的研究;郑小平,毕业于清华大学,教授,博士生导师,面向国家宽带信息网络发展的重大需求,长期致力全光通信网络与微波光子学的研究;陈文军,工程师,研究领域为IPRAN设备4G、5G业务承载方案,对于IPRAN网络业务部署、告警产生机制、设备业务告警关联分析有深入了解。