

2020AIIA 杯人工智能 5G 网络应用大赛

暨 ITU AI/ML in 5G 挑战赛中国区比赛

中国联通-华为联合赛站

复赛数据及规则说明

➤ 赛题名称

基于 MEC 平台的网络设备故障定因

➤ 复赛说明

本赛题复赛阶段仍使用初赛阶段的数据，但是提交结果要求与评审规则有所改变，

请仔细阅读以下说明。

另外，广东赛区复赛评审将采用人工评审，NAIE 平台及初赛数据用作选手训练模

型、优化算法，NAIE 复赛评分排行榜仅用作选手参考，不再用做复赛最终评审办法。

➤ 赛题描述

MEC 属于通信云的一种，他将行业用户的个性化应用需求从中心云以虚拟机或容器的方式推到边缘云上，通过边缘计算来降低时延、加速敏感型应用计算，如工业图像识别等，对于用户属于移动通信能力在边缘的定制化开放。此项目的目的在于，通过人工智能算法分析 MEC 客户、运营商 IT 及 CT 基础设备的 KPI 性能及日志，从而进行故障预测及决策。来提升 MEC 运营能力及可提供给客户的 AI 能力，如快速发现、定位、预测、自愈等。本赛题主办方提供了业内不同厂商的 CT 设备日志及部分流量 KPI 数据。参赛选手需要根据提供的训练数据来进行算法模型训练，并使用模型对故障测试数据进行分析，快速预测故障趋势或定位故障根因并复现故障发生过程。

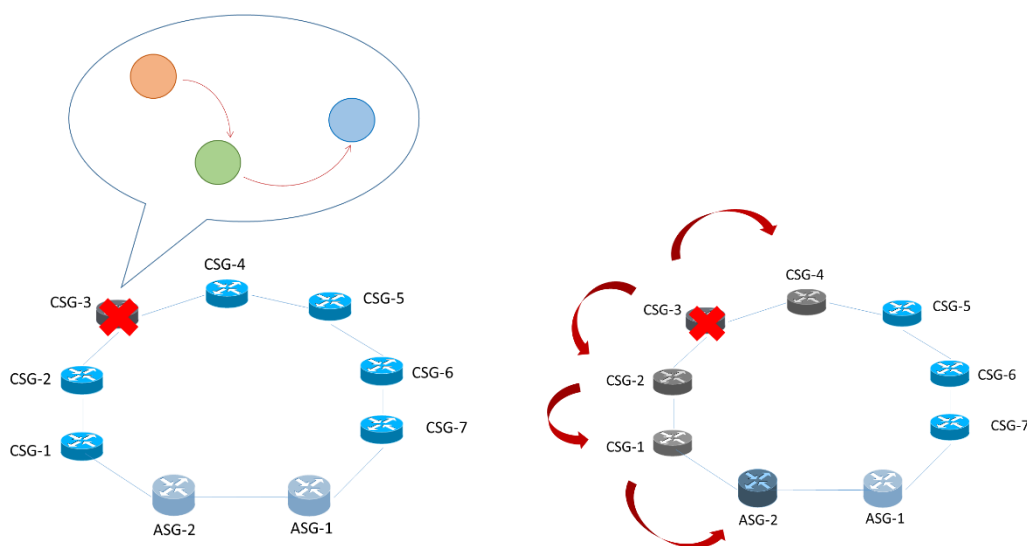
➤ 攻关方向

本题目希望参赛者重点关注提供的日志数据特征分析，以发现问题、找到根因并复现问题为目的，结合提供的网络拓扑信息分别分析网络设备日志描述的关联关系、提取

日志模板、预测关键日志、关键字搜索关联等，找出影响网络正常运行的故障点、对故障进行定因。复赛主要考察对故障传播的分析，即实现网络故障事件回放。

网络故障的传播包括设备内部的故障传播和设备之间的故障传播。在设备内部，如果设备的一个端口发生 down 故障，则基于这个端口的协议或业务也将发生状态变化，这就是在设备内部的端口 down 引起的协议、业务状态变化的一种故障传播；在设备之间，如果某台设备的某个接口或协议发生了异常，相关联的设备如果依赖于这台设备发生异常的接口或协议的话，则也会产生相应的异常信息（日志），这就是设备之间的故障传播情况。

如下图中左边所示，为设备内部的故障传播示意图，右边所示为设备之间的故障传播示意图。



➤ 复赛提交结果

复赛：选手需要提交三部分：一是提交传播关系算法模型以及分析结果（算法模型包括算法训练方式说明及训练超参数值和分析结果以.csv 格式提交）；二是带注释的核心代码以及代码设计说明文档（核心代码使用.zip 压缩包提交，说明文档以.pdf 文件格式提交）；三是故障传播分析结果（以.csv 格式提交）。最终把所有的文件打包压缩成 zip 文件通过邮箱 Alguangdong1@163.com 进行提交。

所有文件的命名格式统一以**赛题名+团队名称**命名文件，例如：“基于 MEC 平台

的环路网络设备故障定因+中国联通网研院.zip”。

提交文件评分说明：

文件夹	文件及格式	分值
传播关系算法模型以及分析结果	模型 (.csv)	20
	分析结果 (.csv)	20
带注释的核心代码以及代码设计说明文档	核心代码 (.zip)	20
	代码设计说明文档 (.pdf)	20
故障传播分析结果	传播分析结果 (.csv)	20

➤ 复赛评价指标

复赛评价指标为相关方案结果综合评分，分 3 部分进行综合评分，第一部分传播关系算法模型以及分析结果 F1；第二部分带注释的核心代码以及代码设计说明文档 F2；第三部分故障传播分析结果 F3。

第一部分传播关系算法模型以及分析结果，其中传播关系分析模型占 20 分，分析结果占 20 分，如果模型和分析结果完整，则 $F1=40$ ；第二部分带注释的核心代码以及代码设计说明文档，其中核心代码占 20 分，设计说明文档 20 分，如果代码清晰，说明文档完整，则 $F2=40$ ；第三部分故障传播分析结果，传播分析结果中包含故障相关设备的设备内部传播过程以及设备间的传播过程，如果故障回放过程分析正确则 $F=20$ 分。

分析处理数据对象一共分为 A、B 两份，分别对两份数据的 test data 分析结果进行打分 $F_A = F1_A + F2_A + F3_A$ ， $F_B = F1_B + F2_B + F3_B$ 。

最终得分： $F_{\text{复赛}} = 0.5 * (F_A + F2_B)$ 。

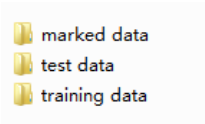
➤ 数据说明

复赛数据仍然沿用初赛数据，不再提供新的数据。

本赛题一共提供 A、B 两份数据，这两份数据是由不同厂家的网络设备产生的，数据结构会略有不同，

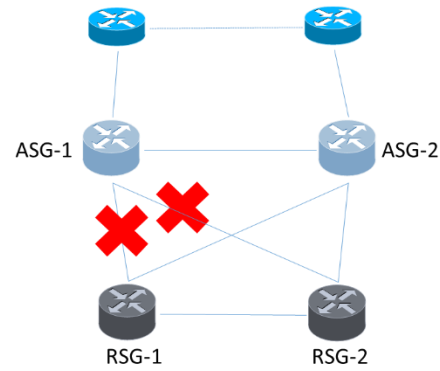
A 数据说明

1. 文件夹结构



2. marked data

(1) 拓扑结构



(2) 数据文件列表

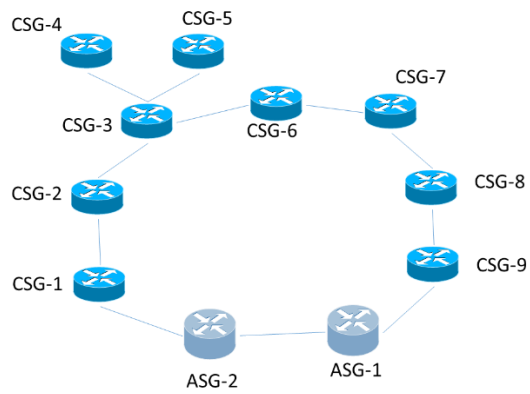
FILE LIST	CONTENT
ASG-1	logs
ASG-2	logs
RSG-1	logs
RSG-2	logs

(3) 故障说明

故障原因：2 月 19 日 22:08 在 ASG-1 至 RSG-1，2 月 20 日 05:23 在 ASG-1 至 RSG-2，汇聚环路双断导致下挂业务中断。

3. training data

(1) 拓扑结构

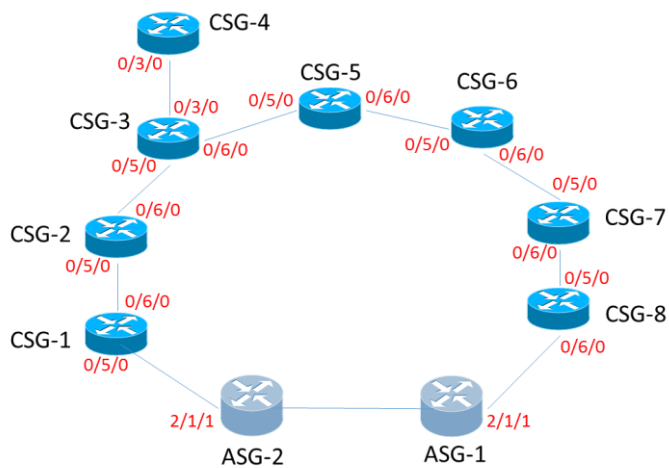


(2) 数据文件列表

FILE LIST	CONTENT
ASG-1	null
ASG-2	null
CSG-1	logs
CSG-2	logs
CSG-3	logs
CSG-4	logs
CSG-5	logs
CSG-6	logs
CSG-7	logs
CSG-8	logs
CSG-9	logs
interface flow rate	flow rate

4. test data

(1) 拓扑结构



(2) 数据文件列表

FILE LIST	CONTENT
ASG-1_10.24.91.28	logs
ASG-2_10.24.91.27	logs
CSG-1_10.25.95.128	logs
CSG-2_10.25.95.129	logs
CSG-3_10.25.95.131	logs
CSG-4_10.25.95.135	logs
CSG-5_10.25.95.130	logs
CSG-6_10.25.95.132	logs
CSG-7_10.25.95.133	logs
CSG-8_10.25.95.134	logs
ASG-1_interface flow rate	flow rate
ASG-2_interface flow rate	flow rate

5. 日志说明

(1) 文本结构

日志随时间打印的每一条文本都有固定的结构, 如下图中日志结构与真实日志进行对比可以看出, 这该日志产生时间为 UTC 时间 2020 年 4 月 10 日 06 点 07 分 32 秒 (+08:00 无需关注), 主机名为 CSG, 该日志由 NTP 模块发出, 级别为 4, 日志描述的内容是源参考时钟的系统层数由 16 变为 10。

TimeStamp HostName %% dd ModuleName/Severity/Brief(I)[Count]:Description
1 2 3 4 5 6 7 8 9 10



Apr 10 2020 06:07:32+08:00 CSG %%01NTP/4/STRATUM_CHANGE(I)[23279]:System stratum changes from 16 to 10.

(2) 日志等级

日志级别用于标识一条日志的严重程度，日志级别共分 8 级，如下表所示。

0	Emergency
1	Alert
2	Critical
3	Errors
4	Warnings
5	Notifications
6	Informational
7	Debug

(3) 模块


该字段表示日志是由哪个模块产生的。网络设备的模块名有很多种，但日常打印日志的时候只会打印设备运行进程中涉及的模块日志。如常见模块：AAA、ADDR、ARP、ATM、BFD、BGP、CMD、DHCP、DRV、IGMP、ISIS、L2V、L3VPN、LLDP、LSPM、NTP、OAM、OSPF、OSPFV3、PIA、PPP、QINQ、QoS、RIP、RM、RSVP、SHELL、SOCKET、SRM、SSH、TRUNK、TUNNEL、VRRP。


(4) 详细信息

详细信息是各个模块实际向信息中心输出的字符串信息，由各个模块在每次输出时填充，详细描述该日志的具体内容。此部分内容一般是非结构化的，但相同类型的日志详细信息结构是一样的，只有参数的差别。因此在对日志分析的过程中，详细内容的分析是一个难点，可能需要借助 NLP 等技术才能进行分析。

B 数据说明

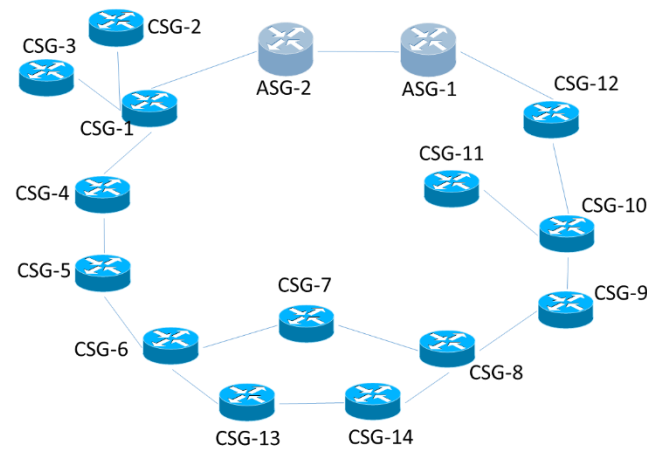
1. 文件夹结构

 test data

 training data

2. training data

(1) 拓扑结构



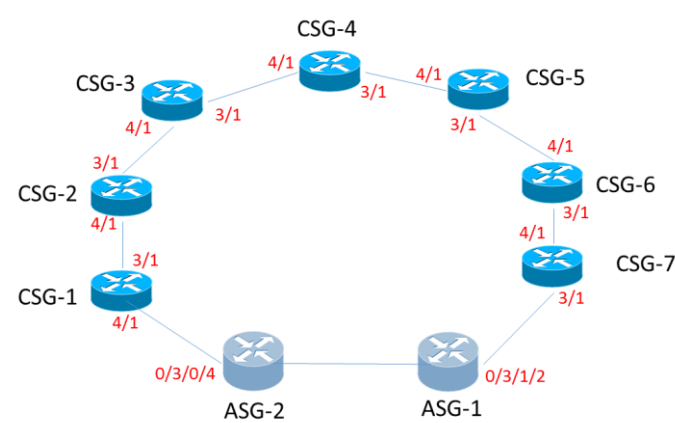
(2) 数据文件列表

FILE LIST	CONTENT
ASG-1_10.24.168.25	logs
ASG-2_10.24.168.24	logs
CSG-1_10.28.202.116	logs
CSG-2_10.28.202.117	logs
CSG-3_10.28.202.122	logs
CSG-4_10.28.202.115	logs
CSG-5_10.28.202.118	logs
CSG-6_10.28.202.114	logs
CSG-7_10.28.202.113	logs
CSG-8_10.28.202.112	logs
CSG-9_10.28.202.111	logs
CSG-10_10.28.202.110	logs
CSG-11_10.28.202.125	logs
CSG-12_10.28.202.109	logs
CSG-13_10.28.202.126	logs

CSG-14_10.28.202.124	logs
interface flow rate	flow rate

2. test data

(1) 拓扑结构



(2) 数据文件列表

FILE LIST	CONTENT
ASG-1_10.24.155.7	logs
ASG-2_10.24.155.8	logs
CSG-1_10.26.230.40	logs
CSG-2_10.26.230.39	logs
CSG-3_10.26.230.37	logs
CSG-4_10.26.230.36	logs
CSG-5_10.26.230.35	logs
CSG-6_10.26.230.34	logs
CSG-7_10.26.230.42	logs

4. 日志说明

(1) 文本结构

日志随时间打印的每一条文本都有固定的结构,如下图中日志结构与真实日志进行对比可以看出,这该日志产生时间为 UTC 时间 2019 年 8 月 6 日 20 点 31 分 56 秒,主机名为 ASG-2,由 IP 模块发出告警,级别为 1,日志描述的内容是接口 Down。

<type><code> ID <flow-id> level <level>{occurred | cleared} at <time> sent by <position> %<module>% <description>



An alarm 150101 ID 4546756 level 1 occurred at 20:31:56 08-06-2019 sent by ASG-2 %IP% Interface status The interface(index=10356,name='te_tunnel59') turned into protocol DOWN

(2) 日志等级

日志级别用于标识一条日志的严重程度，日志级别共分 8 级，如下表所示。

1	EMERGENCIES (严重告警)
2	ALERTS (重大告警)
3	CRITICAL (紧急告警)
4	ERRORS (错误告警)
5	WARNINGS (警告告警)
6	NOTIFICATIONS (通告信息)
7	INFORMATIONAL (报告信息)
8	DEBUGGING (调试信息)

(3) 模块

该字段表示日志是由哪个模块产生的。网络设备的模块名有很多种，但日常打印日志的时候只会打印设备运行进程中涉及的模块日志。如常见模块：MAC、STP、PPP、ARP、UDLD、CFM、LACP、IP、ACL、PIM、BGP、OSPF、ISIS、RIP、L2VPN、L3VPN、TUNNEL、RSVP、VRRP、DHCP、NTP、SNMP、BFD。

(4) 详细信息

详细信息是各个模块实际向信息中心输出的字符串信息，由各个模块在每次输出时填充，详细描述该日志的具体内容。此部分内容一般是非结构化的，但相同类型的日志详细信息结构是一样的，只有参数的差别。因此在对日志分析的过程中，详细内容的分析是一个难点，可能需要借助 NLP 等技术才能进行分析。

传播分析说明

1. 设备内部传播，在日志分析的过程中，是指设备发生故障时，故障相关日志的先后打印顺序的特征；设备之间传播，是指从根故障设备向相关设备传递异常的顺序特征。

2. 传播分析结果提交，可参考如下格式进行提交。

字段名称	字段描述
Device1	故障相关设备名称
Spread2device1	传递下级设备名称（数量<3）
SpreadParaForD1	传递给下一个设备的传递参数
Log1	传播链根日志
Spread2log1	传递下级日志（数量<5）
SpreadParaForL1	传递给下一条日志的传递参数
Log2	传播链根日志
Spread2log2	传递下级日志
SpreadParaForL2	传递给下一条日志的传递参数
.....	设备内传播日志条数最多 20 条
Device2	故障相关设备名称
Spread2device2	传递下级设备名称（数量<3）
SpreadParaForD2	传递给下一个设备的传递参数
Log1	传播链根日志
Spread2log1	传递下级日志（数量<5）
SpreadParaForL1	传递给下一条日志的传递参数
Log2	传播链根日志
Spread2log2	传递下级日志
SpreadParaForL2	传递给下一条日志的传递参数
.....	设备内传播日志条数最多 20 条
Device3	故障相关设备名称
.....	设备间传播设备数最多 10 台
.....	
.....	