

通信网络日志数据的质量优化与故障诊断适配研究

张栋栋 李浩然

上海航天电子技术研究所 上海 201109

摘要: 本文针对通信网络日志数据存在的噪声多、缺失率高、时效性差等问题展开研究。首先分析各类质量问题的表现与影响,随后构建包含完整性、一致性、时效性、信噪比的质量评估体系及量化模型,设计基于GAN的缺失日志修复算法、“规则引擎+NLP”的冗余日志清洗算法及其他质量优化策略,最后通过4G/5G基站故障数据集验证优化效果,建立“数据质量-诊断精度”关联模型。研究形成完整数据预处理方案,为提升通信网络故障诊断效率提供支撑。

关键词: 通信网络日志数据;质量优化;故障诊断;生成式对抗网络;量化评估模型

引言: 通信网络稳定运行依赖日志数据支撑故障诊断,但当前日志数据质量堪忧。噪声多干扰信息提取,缺失率高破坏数据完整性,时效性差延误故障响应。这些问题致使故障诊断模型精度受限,难以精准定位与处理故障。为解决此困境,需深入分析日志数据质量问题,构建质量评估体系,设计修复优化算法,验证其对诊断的适配效果,提升通信网络运维水平。

1 通信网络日志数据质量问题分析

1.1 噪声多问题

冗余日志常见表现形式包括重复记录相同设备状态信息、生成与故障诊断无关的冗余字段,这类日志多因设备配置参数设置不当,导致同一事件被多次触发记录,或日志采集范围未精准界定,纳入非关键运行数据^[1]。格式错误涵盖字段缺失、数据类型不匹配、字符乱码等类型,字段缺失表现为日志关键信息项空白,数据类型不匹配如将数值型参数记录为文本格式,字符乱码多由数据传输编码协议不一致引发,此类错误会导致日志数据无法被诊断系统正常解析,干扰对设备运行状态的准确判断。

1.2 缺失率高问题

网络中断是导致日志传输不完整的主要原因,通信基站与核心网之间的传输链路若出现信号中断、带宽不足等情况,正在传输的日志数据会出现部分丢失,尤其在恶劣天气或网络负载高峰期,中断频率增加会导致大量日志片段缺失。缺失日志会使故障诊断系统无法获取设备故障发生前后的完整运行数据,难以还原故障演变过程,例如设备故障前的异常参数变化日志缺失,会导致诊断模型无法精准定位故障触发点,增加故障误判或漏判的概率,阻碍诊断工作高效推进。

1.3 时效性差问题

日志上报延迟的原因包括设备端日志缓存机制设置

不合理,日志数据在设备本地缓存时间过长未及时上传,或传输网络拥堵导致数据排队等待,此外日志处理服务器负载过高时,接收与转发日志的效率下降也会引发延迟。时效性差会影响诊断模型的及时性,故障发生后若相关日志未能及时上报,诊断模型无法实时获取故障信息,导致故障识别与响应滞后,错过最佳故障处理时机,尤其在通信网络故障可能快速扩散的场景中,延迟会加剧故障对网络运行的影响范围。

2 数据质量评估体系构建

2.1 通信日志数据质量指标体系设计

完整性指标聚焦日志数据是否完整覆盖通信设备运行的关键信息,衡量方式以日志记录的缺失程度为核心,具体参数包括关键字段缺失数量、重要事件记录遗漏频次、数据传输中断导致的片段缺失长度,通过这些参数判断日志是否能完整反映设备运行状态与事件发生过程^[2]。一致性指标强调日志数据在格式、内容逻辑上的统一程度,其重要性体现在避免因数据矛盾干扰故障诊断判断,评估方法采用格式校验与逻辑比对,具体指标涵盖字段格式统一度、同一事件多设备记录的内容匹配度、时间戳连续性,以此排查格式错误与逻辑冲突数据。时效性指标关注日志从生成到上传至分析系统的时间间隔,评估标准以日志生成时间与接收时间的差值为核心,具体指标包括平均上报延迟时长、延迟日志占比、关键事件日志的实时性达标率,通过这些指标保障日志能及时支撑故障的快速检测。信噪比指标体现有效日志数据与噪声数据的比例关系,在通信日志中含义为有效故障信息、设备状态记录等有用数据占总日志量的比重,计算方法通过识别冗余日志、无效代码、重复记录等噪声数据并统计其数量,评估指标包括噪声数据占比、有效数据提取效率,以此量化日志数据的有效程度。

2.2 量化评估模型设计

评估模型整体架构采用分层设计,底层为数据采集层负责获取原始日志数据,中间层为指标计算层针对各质量指标进行参数运算,顶层为综合评估层整合各指标结果输出整体质量评分。各指标权重分配方法基于通信日志在故障诊断中的应用需求,通过专家打分结合故障诊断场景下各指标对诊断结果的影响程度确定,例如完整性与信噪比对故障信息提取影响更大,权重分配相对更高。量化评估模型计算流程先对原始日志数据进行预处理,去除明显无效数据后,分别代入各指标的计算公式得到单项指标得分,再根据预设权重对单项得分进行加权求和,得到日志数据整体质量评分,实现方式通过编程工具开发自动化计算模块,将指标计算逻辑与权重分配规则嵌入模块,支持对不同类型通信日志数据进行批量评估。

3 数据修复优化算法设计

3.1 针对缺失日志的修复算法

生成式对抗网络(GAN)通过生成器与判别器的对抗训练实现数据生成,生成器负责学习日志数据特征并生成模拟缺失数据,判别器负责区分真实日志与生成日志,二者持续迭代优化直至生成器产出接近真实分布的日志数据^[3]。基于GAN修复缺失日志需经过多步具体实现流程,数据预处理阶段需对原始日志进行格式统一、异常值剔除,再通过特征提取筛选设备运行参数、时间序列等关键特征,将其转化为模型可处理的向量形式;GAN模型构建时需设计适配日志数据的生成器与判别器网络结构,生成器采用全连接层与卷积层结合的架构捕捉数据特征,判别器通过多层感知机实现分类判断,训练过程中以日志数据分布差异为损失函数,不断调整模型参数;缺失日志生成阶段利用训练完成的生成器,根据缺失日志周边的完整数据特征生成补充内容,最后通过验证环节比对生成数据与真实数据的特征相似度,确保修复数据的合理性。

3.2 针对冗余日志的清洗算法

规则引擎设计需遵循准确性与灵活性原则,准确性要求制定的规则能精准匹配冗余日志特征,灵活性要求规则可根据日志类型变化进行动态调整,规则制定方法需结合通信日志冗余模式,从重复记录频率、字段相似度等维度提炼规则。清洗冗余日志的规则类型包括重复日志删除规则、无关字段过滤规则,重复日志删除规则依据日志时间戳、设备标识、参数值的一致性判断重复项,无关字段过滤规则剔除与故障诊断无关的冗余字段。规则引擎工作流程为首先导入原始日志数据,其次通过规则匹配模块筛选冗余日志,最后执行清洗操作去除冗

余信息,执行机制采用优先级排序,确保关键规则优先触发。自然语言处理(NLP)技术通过文本分词、语义分析识别冗余日志,对日志中的文本描述字段进行语义相似度计算,区分有效信息与重复表述;结合规则引擎和NLP的清洗算法实现时,先通过NLP技术分析日志文本语义筛选潜在冗余数据,再利用规则引擎对潜在冗余数据进行精准匹配与删除,提升冗余清洗的全面性与准确性。

3.3 其他质量问题的优化策略

针对格式错误的日志修复方法需先通过格式校验模块识别错误类型,该模块通过比对日志数据与预设格式模板的差异定位问题,包括字段长度不符、数据类型错误、编码异常等,再根据错误类型执行对应修复操作,字段长度不符时按标准长度截断或补全,数据类型错误时进行类型转换,编码异常时采用统一编码格式重新解析。提高日志时效性的优化措施可从传输与处理两方面入手,传输环节优化日志传输协议,采用轻量级传输格式减少数据传输量,同时设置传输优先级确保关键日志优先上传;处理环节优化日志服务器架构,采用分布式处理模式将日志数据分配至多个节点并行处理,提升日志接收与存储效率,减少日志在服务器端的滞留时间,确保日志快速进入诊断系统。

4 诊断适配验证与研究成果总结

4.1 4G/5G基站故障数据集介绍

数据集来源于实际通信网络运维过程中积累的4G/5G基站运行日志与故障记录,覆盖国内多个省份的城市与郊区基站,包含华为、中兴等主流品牌的多型号基站设备,涵盖不同网络负载与环境条件下的日志数据,能真实反映基站运行状态。数据集中包含的故障类型以“基站退服”“传输链路中断”为主,涵盖设备硬件故障、软件异常等其他常见故障,相关参数包括基站运行电压、信号强度、传输速率、日志生成时间戳等,这些参数为故障诊断与数据质量分析提供充足信息支撑。

4.2 测试方案设计

不同数据质量优化策略的组合和实施方式分为单一优化策略与综合优化策略,单一优化策略包括单独使用GAN修复缺失日志、单独采用“规则引擎+NLP”清洗冗余日志,综合优化策略则整合缺失修复、冗余清洗、格式修正与时效性提升四种措施,每种策略均设置三次重复测试以减少偶然误差^[4]。测试环境的搭建基于服务器集群构建模拟通信网络运维系统,配置与实际基站运维一致的日志采集、存储与诊断模块,同时模拟恶劣网络环境下的日志传输场景,确保测试环境与真实场景的一致性。测试指标的确定以故障诊断准确率为核心,同时纳

入诊断耗时、误判率等辅助指标,评估方法采用对照实验,将未优化数据与不同优化策略处理后的数据分别输入同一故障诊断模型,对比输出结果差异。

4.3 故障诊断准确率提升效果测试

在未进行数据质量优化时,因日志噪声多、缺失率高,“基站退服”“传输链路中断”等故障的诊断准确率较低,存在较多误判与漏判情况,尤其在负载网络环境下诊断误差更为明显。进行单一优化策略后,采用GAN修复缺失日志能提升诊断准确率,减少因数据缺失导致的漏判;采用“规则引擎+NLP”清洗冗余日志可降低误判率,提升诊断精度,两种单一策略对不同故障类型的准确率提升幅度存在细微差异。实施综合优化策略后,诊断准确率较未优化状态有显著提升,且优于单一优化策略效果。分析不同优化策略对诊断准确率的影响程度发现,缺失日志修复与冗余日志清洗对准确率提升贡献更大,格式修正与时效性提升的影响相对较小,整体呈现优化措施越全面,准确率提升越明显的趋势。

4.4 “数据质量-诊断精度”关联模型建立

关联模型的构建思路以数据质量指标为输入变量,以故障诊断准确率为输出变量,其中数据质量指标包括完整性、一致性、时效性、信噪比,通过机器学习算法挖掘二者间的映射关系。构建方法采用多元线性回归与神经网络结合的方式,先通过线性回归初步确定指标与准确率的线性关联,再通过神经网络拟合非线性关系。模型参数的确定基于测试数据进行迭代训练,通过调整权重系数与激活函数参数,使模型预测值与实际测试值的误差最小化,通过交叉验证避免模型过拟合。关联模型的验证采用新的基站故障数据集进行测试,验证模型预测结果的准确性;应用方向包括通过数据质量指标预测诊断精度,或根据目标诊断精度反推所需数据质量标准,为数据预处理提供量化依据。

4.5 研究成果总结与创新点阐述

研究成果总结方面,在数据质量评估领域,构建了包含四大核心指标的通信日志数据质量评估体系与量化

模型,实现对日志数据质量的系统评估;在数据修复优化领域,设计了基于GAN的缺失日志修复算法与“规则引擎+NLP”的冗余日志清洗算法,同时提出格式错误修复与时效性提升策略,形成完整的数据优化方案;在诊断适配验证领域,通过测试验证了不同优化策略对诊断准确率的提升效果,建立了“数据质量-诊断精度”关联模型。创新点方面,首次针对通信日志数据的特性构建质量优化体系,该体系充分考虑通信日志冗余多、缺失率高的特点,区别于通用数据质量优化方案,对通信系统数据预处理具有更强针对性,能有效解决实际运维中的数据质量问题;量化数据质量对故障诊断的影响,通过关联模型明确各质量指标与诊断精度的量化关系,改变以往仅定性分析数据质量重要性的现状;为通信系统数据预处理提供标准化方案,从质量评估、修复优化到适配验证形成完整流程,可直接应用于4G/5G基站等通信设备的日志数据处理,推动通信网络运维数据预处理的标准化发展。

结束语

电气工程及其自动化在电气安全生产中发挥着不可替代的作用,从风险预防到应急响应,从操作规范到数据决策,全方位提升了安全生产水平。然而,技术实施过程中系统可靠性、人机协作及网络安全等问题仍需重视。未来,应持续优化技术,完善管理策略,加强人员培训,以更好地应对挑战,推动电气安全生产向更高水平发展,保障生产活动稳定有序进行。

参考文献

- [1]罗鹏,李景文.面向大数据平台调度优化的通信网络故障诊断[J].微型电脑应用,2024,40(03):198-201+213.
- [2]韩茂玲.网络优化调度算法在船舶大数据通信中的应用[J].舰船科学技术,2021,43(06):169-171.
- [3]叶占坤.通信网络故障诊断与维护在通信工程中的应用[J].科技新时代,2023(9):75-78.
- [4]万强.通信网络质量监测与故障诊断方法的研究与应用[J].电力设备,2023(21):46-48.