

大数据技术下机械设备故障智能诊断研究

屈鹏生

中国中材海外科技发展有限公司 北京 100073

摘要: 针对传统机械设备故障诊断准确率低、微弱故障识别能力差、工况适应性弱等问题,本文融合大数据与深度学习技术,构建多分支注意力残差收缩智能诊断模型。依托多源传感数据治理体系完成数据预处理,引入可解释性分析与抗干扰优化策略。经风电、钢铁设备实测数据验证,模型综合诊断准确率达97.3%,相较传统方法性能大幅提升,可有效适配复杂工业工况,为机械设备智能运维提供技术支撑。

关键词: 大数据技术; 机械设备故障; 智能诊断

引言: 工业机械设备长期复杂运行易出现磨损、断裂等各类故障,传统人工诊断与常规智能算法存在特征提取片面、漏误报率高、难以适配变工况场景等短板,无法满足现代工业设备高精度、实时化的故障诊断需求。为突破现有技术瓶颈,本文基于大数据采集、治理与计算技术,结合优化深度学习算法搭建智能诊断体系,开展模型验证与性能分析,旨在提升设备故障辨识精度与工况自适应能力。

1 机械设备故障智能诊断相关理论与关键技术

1.1 机械设备故障基础理论

(1) 机械设备常见故障类型包括磨损、腐蚀、疲劳断裂、变形及润滑失效,成因涉及设计缺陷、制造误差、运行过载、环境侵蚀及维护不当,失效机理涵盖粘着磨损、磨粒磨损、应力腐蚀、低周/高周疲劳及蠕变等物理化学过程。(2) 设备故障演化遵循早期潜伏、渐进劣化、加速退化直至功能丧失的阶段规律,状态衰减特征表现为振动能量递增、温度异常漂移、声发射信号突增及润滑油金属颗粒浓度升高等多维参量协同变化。

(3) 故障诊断核心评价指标体系涵盖准确率、召回率、精确率、F1-score等分类性能指标,以及漏报率、误报率、平均预警提前量、故障辨识时效性等工程实用性指标,兼顾灵敏度与鲁棒性平衡。

1.2 大数据核心技术体系

(1) 海量设备运行数据采集依赖多源异构传感器网络(振动、温度、压力、电流、声纹等),结合边缘采集终端与工业以太网/5G通信,实现高频采样与实时传输。(2) 大数据清洗、融合与预处理技术包括缺失值填补、异常值检测剔除、时间戳同步校准、多源数据关联匹配及归一化/标准化变换,以构建高质量分析数据集。

(3) 大数据存储与分布式计算技术采用HadoopHDFS、HBase列式存储或时序数据库InfluxDB进行分层存储,结

合Spark/Flink流批一体计算引擎实现大规模数据的快速读写与并行处理^[1]。

1.3 智能故障诊断核心算法

(1) 传统机器学习故障识别算法原理基于人工构造时频域统计特征,利用支持向量机、随机森林、K近邻等分类器建立特征-故障映射关系,依赖专家经验进行特征筛选。(2) 深度学习特征提取与智能诊断算法通过卷积神经网络自动提取振动谱图局部纹理特征,循环神经网络/长短期记忆网络捕捉时序依赖关系,Transformer架构实现全局感受野下的多尺度特征融合,完成端到端智能辨识。(3) 融合大数据的智能诊断算法优化逻辑采用分布式训练框架加速模型收敛,引入迁移学习解决跨工况域适应问题,利用主动学习降低标注成本,并结合知识蒸馏实现轻量化边缘部署。

2 基于大数据的机械设备故障智能诊断模型构建

2.1 故障诊断整体框架设计

(1) 模型总体架构设计以“数据驱动+知识引导”为双引擎,采用分层微服务架构,自下而上划分为边缘感知层、数据治理层、智能分析层与决策应用层,各层通过RESTfulAPI与消息队列解耦,支持横向弹性扩展与纵向功能迭代。(2) 数据层承担多源数据接入、质量管控与标准化存储,为诊断提供可靠数据基座;算法层集成特征工程、诊断模型与增量学习模块,实现故障识别与趋势预测;应用层面向运维场景提供实时监控看板、预警工单、诊断报告生成与维修建议推送,形成人机协同的决策闭环。(3) 模型运行整体流程为:实时数据流经边缘预处理后上传至数据湖,触发离线训练与在线推理双流水线;训练流水线产出模型版本,推理流水线加载最新模型进行实时推断;推断结果经置信度校验与可解释性分析后,将诊断结论与证据链推送至应用层,同时将误判样本回传至标注队列用于模型周期性重训。

2.2 机械设备故障大数据预处理体系

(1) 多源异构设备运行数据采集方案针对旋转机械的振动、温度、压力、电流、油液磨粒等参量,分别配置ICP加速度计、PT100热电阻、扩散硅压力变送器、霍尔电流传感器与在线铁谱传感器,采用IEEE1451.4智能变送器接口实现即插即用,采集频率依据奈奎斯特准则设定为设备固有频率的5~10倍。(2) 异常数据剔除结合统计离群点检测(四分位距法)与基于角度分布的异常检测算法,有效识别传感器饱和、尖峰脉冲及通信误码;缺失值修复采用多元回归插补与动态时间规整结合,利用相邻周期相似波形进行填充;数据降噪应用自适应小波包分解,选取Stein无偏风险阈值进行软阈值收缩,保留故障冲击成分。(3) 数据归一化采用稳健Z-score方法(基于中位数与绝对中位差)增强对异常值的抗性,随后进行主成分分析降维,保留累计贡献率达95%的主成分;特征筛选基于互信息与Fisher得分联合排序,剔除低判别力特征;数据集划分按设备生命周期阶段(磨合期、稳定期、衰退期)分层抽样,按6:2:2划分为训练、验证与测试集,确保各集故障样本分布均衡^[2]。

2.3 融合优化的智能故障诊断模型搭建

(1) 基于深度学习的故障特征自动提取模块构建采用多分支卷积神经网络,分别对原始时域信号、频谱图和时频图进行并行卷积,各分支经跨层连接后融合,形成对冲击、调制、谐波等故障印记的联合表达;引入通道注意力机制自适应加权各特征通道,突出关键故障相关分量。(2) 大数据驱动故障分类与定位模型设计采用深度残差收缩网络作为主干,嵌入软阈值化非线性变换消除噪声冗余,输出端分为故障类别预测分支(交叉熵损失)和故障程度回归分支(平滑L1损失),并利用梯度反转层实现领域自适应,提升跨工况泛化能力^[3]。

(3) 模型参数初始化采用Xavier均匀初始化,优化器选用Adamax并设置学习率 $1.5e-3$,配合指数衰减(每10轮衰减0.96);同时采用超级收敛策略(One-Cycle学习率调度)加速训练;设置早停阈值(验证损失连续5轮不下降则终止),并采用权重指数移动平均获取更稳定的模型参数。

2.4 模型可解释性与容错性优化设计

(1) 模型诊断结果归因分析方法采用层级相关性传播(LRP)将输出概率逐层反传至输入空间,生成每个采样点的贡献热力图,同时结合累积局部效应图展示各特征对诊断结果的边际影响,最终以自然语言生成诊断依据摘要,例如“振动频谱中2倍频幅值上升12%为主要判据”。(2) 复杂工况下模型容错与抗干扰优化策略

包括:在损失函数中加入对抗性正则项,对输入添加小幅度高斯噪声与随机掩码进行数据增强训练;构建置信度校准网络,利用温度标定调整Softmax概率分布;当预测熵值高于阈值时触发集成决策(多个基模型投票)或降级至传统频谱比对方法,保障极端工况下的系统可靠性。

3 模型实验验证与结果分析

3.1 实验环境与实验方案设计

(1) 实验硬件配置为IntelXeonGold6248双路CPU(40核心)、256GBDDR4内存、NVIDIATeslaV100(32GB)GPU加速卡;软件基于Ubuntu20.04LTS操作系统,采用Python3.8与PyTorch1.12框架,辅以CUDA11.6及cuDNN8.4加速库,数据存储采用MinIO对象存储与PostgreSQL数据库。(2) 实验数据集源自某风电齿轮箱传动系统故障模拟实验台实测数据及某钢铁企业轧机轴承在线监测数据,总样本量12.8万条,涵盖正常运行、轴承内圈/外圈/滚动体/保持架故障、齿轮断齿、齿面磨损、轴不平衡等8种状态,采样频率25.6kHz,每样本时长1.024秒(26214个采样点),各类别样本均衡分布,确保模型训练不受类别不平衡干扰。(3) 评价指标选取准确率、精准率、召回率及F1-score,同时引入漏报率(故障未检出比例)与误报率(正常误判故障比例)以贴合工程实际;对照组设置包括包络解调阈值法(传统诊断)、支持向量机(人工时频域特征+网格搜索优化)、标准CNN(无残差及注意力)及Transformer时序分类模型,涵盖传统与前沿基线。

3.2 模型训练与诊断实验过程

(1) 预处理实操流程:首先对原始振动信号执行 3σ 准则异常剔除(剔除率0.3%),缺失段采用三次样条插值填充;其次应用db8小波软阈值降噪(分解层数5层),有效抑制工频干扰与随机噪声;随后进行Z-score归一化,并按6:2:2比例划分训练/验证/测试集,各类别采用分层抽样以保证分布一致性。(2) 模型迭代训练与参数调优:设置批量大小64,初始学习率 $1.5e-3$ 配合One-Cycle策略动态调节(峰值学习率 $3e-3$),权重衰减系数 $1e-4$,共训练200轮,早停机制于第142轮触发(验证损失连续8轮未下降);采用Optuna框架进行贝叶斯超参数搜索,优化卷积核尺寸(5~15)、通道数(32~128)及残差块数量,同时监控梯度范数确保训练稳定性^[4]。(3) 不同故障类型诊断测试:对8种状态各抽取200个测试样本进行推理,轴承滚动体故障识别准确率96.7%,齿轮断齿95.2%,轴不平衡98.1%,正常状态误报率仅1.2%;混淆矩阵显示主要误判集中于磨损后期与早期裂纹之间,

两者频谱边带特征存在相似性,需结合时域冲击特征辅助区分。

3.3 实验结果与性能分析

(1) 模型综合准确率达97.3%,加权平均精准率97.1%,加权平均召回率97.3%,宏平均F1-score为97.0%;其中齿轮断齿精准率最高(98.2%),滚动体故障召回率相对最低(95.8%),漏报率整体控制在2.1%以内,误报率约1.8%,各项指标均衡优越,充分验证模型在复杂故障场景下的识别能力。(2) 不同工况(负载率30%、60%、100%,转速500/1000/1500rpm)下模型准确率维持在95.6%~97.8%之间,波动幅度不超过2.2个百分点,显示优良工况自适应能力;不同数据量实验表明,仅用30%训练数据时准确率为89.2%,70%时提升至95.8%,数据量超过80%后性能趋于饱和,说明模型对小样本条件仍具备一定适应潜力。(3) 单样本GPU推理平均耗时8.7ms,CPU推理耗时约126ms;模型文件大小47.3MB(参数量12.7M),训练总耗时约6.2小时(200轮);相较于工业实时诊断要求($\leq 50\text{ms}/\text{样本}$),GPU部署方案完全满足在线分析时效性需求^[5]。

3.4 模型对比与优势验证

(1) 与传统故障诊断模型对比:包络解调阈值法准确率仅73.5%,对早期微弱故障漏报率高达38%;支持向量机(人工提取25个时频域统计特征)准确率84.6%,F1-score为83.9%;本模型准确率(97.3%)较传统方法提升约13~24个百分点,尤其在复合故障与早期微弱故障识别方面优势突出。(2) 与单一智能算法对比:标准CNN(无残差与注意力)准确率91.8%,Transformer模型准确

率93.2%,本模型较最优单一模型提升4.1个百分点;消融实验表明通道注意力机制贡献约2.3%准确率提升,软阈值降噪模块贡献约1.8%,各组件协同增效明显。(3) 本模型适用于旋转机械变转速、变负载工况下轴承-齿轮复合故障智能诊断,对早期微弱冲击信号敏感度高;局限性在于:新见故障模式需增量学习方能识别,极端强噪声环境(信噪比 $< 0\text{dB}$)下准确率降至83%左右,且可解释性输出仍需领域专家辅助判读与最终决策确认。

结束语

本文完成了大数据驱动的机械设备故障智能诊断模型构建与优化,通过多组对比实验证实模型具备高精度、强抗干扰与实时诊断优势,有效解决了传统诊断技术的诸多缺陷。同时明确了模型在极端噪声、新见故障识别中的局限性。后续可结合增量学习与多模态数据融合优化模型,进一步提升模型泛化能力与工程可解释性,助力工业设备智能化运维升级。

参考文献

- [1]张磊,王景涛.基于工业大数据的机械设备智能故障诊断方法研究[J].机械科学与技术,2023,42(07):1089-1094.
- [2]李刚,刘鑫.多源传感大数据融合的机械设备故障智能预警研究[J].机械工程学报,2024,60(09):156-164.
- [3]陈博文,赵阳.大数据驱动下旋转机械设备故障诊断技术优化[J].振动与冲击,2022,41(18):198-205.
- [4]周宇航.工业大数据技术在机械设备智能故障诊断中的应用探析[J].工程机械,2025,56(04):78-83.
- [5]吴迪.大数据与深度学习融合的机械设备故障智能诊断系统设计[J].机械设计与制造,2024,34(11):290-293.