

基于大数据分析的火力发电厂电气设备故障预测技术

侯丕文

国能长源荆州热电有限公司 湖北荆州 434000

摘要: 火力发电厂电气设备的安全稳定运行对保障电力供应至关重要。本研究提出一种基于大数据分析的电气设备故障预测技术框架。该框架系统性地构建了涵盖发电机组、变压器及开关设备的多源异构数据采集体系,并深入探讨了数据预处理、特征提取与混合预测模型构建等关键技术。研究设计了基于边缘计算与云端平台协同的实时预测系统架构,实现了轻量化模型部署、实时数据处理与可视化预警功能。实际应用表明,该技术能够有效识别设备早期故障特征,显著提升预测准确率与运维响应效率,为火力发电厂智能化运维提供了可靠的技术支撑。

关键词: 火力发电厂; 电气设备; 故障预测; 大数据分析; LSTM 神经网络

引言

火力发电厂作为国家能源体系的基石,其电气设备(如发电机组、变压器、开关设备等)的可靠性直接影响电网安全与经济运行。传统定期检修与事后维修模式存在滞后性与资源浪费问题^[1]。大数据技术的迅猛发展为设备状态监测与故障预测提供了全新范式。海量设备运行数据蕴含丰富的状态信息,通过先进的数据采集、处理与分析手段,能够实现设备潜在故障的早期识别与精准预警。

1 火力发电厂电气设备故障特征与数据采集

1.1 关键电气设备故障模式分析

火力发电厂电气设备的故障模式具有显著的物理特性与演化规律^[2]。发电机组作为核心设备,绝缘老化构成其主要失效形式之一,该过程表现为绝缘材料在长期电、热、机械应力作用下的介电强度持续下降,最终可能导致绕组短路等严重故障,绕组短路故障通常由绝缘破损、导电杂质侵入或过电压冲击等因素触发,在故障发生前常伴随局部放电量异常增加或三相电流不平衡度显著上升等可监测征兆。变压器运行状态的关键表征参数包含油温与局部放电活动,油温异常升高往往指示内部存在过热故障点或冷却系统效能衰减,而局部放电信号的强度与模式变化则是绝缘材料劣化的灵敏指示器,持续的局部放电活动会加速绝缘分解并产生特征气体。

1.2 多源异构数据采集体系构建

构建完备的多源异构数据采集体系是实现精准故障预测的基础。传感器网络的合理部署构成物理层数据获取的核

心环节,针对发电机组振动监测需求,压电式加速度传感器因其宽频响应特性成为首选,安装位置需覆盖轴承座与机座关键测点;温度监测则需在定子绕组槽部、铁芯及冷却介质出口等关键位置布置铂电阻或热电偶传感器,传感器选型需考虑测温范围、精度及长期稳定性;压力传感器主要用于监测氢冷发电机密封油压或变压器油流状态。设备运行日志的数字化处理依托厂级监控信息系统实现,该系统实时采集并存储来自分布式控制系统的模拟量、开关量及报警信息,数据接入过程需解决通讯协议转换与数据格式标准化问题。

2 基于大数据的故障特征提取与建模

2.1 数据预处理关键技术

2.1.1 异常值检测与修正算法

异常值检测是确保数据质量的首要步骤,基于统计阈值的方法通过计算特征变量的均值与标准差,将超出 $\mu \pm 3\sigma$ 范围的数据点视为异常值,该方法计算高效但难以识别非高斯分布数据中的异常^[3]。机器学习算法如孤立森林通过递归分割数据空间识别稀疏区域的样本点,其优势在于无需假设数据分布且能处理高维数据。例如,大唐国际托克托电厂在发电机定子温度监测中,某 PT100 传感器因接线松动出现间歇性漂移,超出正常范围达 120°C 。系统采用孤立森林算法识别异常点,并通过前后有效时段数据的线性插值进行修正。修正后数据经频谱分析,成功捕捉到因冷却水阀故障导致的 0.1Hz 低频温升波动,避免了误报警。

2.1.2 多维度数据对齐方法

多源数据的时间同步依赖高精度时钟协议, IEEE1588

精密时间协议可实现亚微秒级同步，通过在局域网内部署主时钟设备，为 SCADA 系统、振动监测装置及局部放电检测仪提供统一时基。例如，华能上海外高桥第三发电厂在 8 号机组振动故障诊断中，整合了键相信号、轴承座振动及过程参数。系统采用 PTP 协议对齐所有数据时标，并依据机组三维模型将振动传感器坐标转换至转子中心线参考系。同步分析发现 2 号轴承垂直方向振动相位在 1800MW 负荷时超前键相 38° ，结合轴系模拟计算，诊断为对中不良，经激光校准后振动幅值下降 60%。

2.1.3 特征降维与选择策略

高维数据易引发维度灾难，主成分分析通过正交变换将相关变量转换为线性无关的主成分，保留最大方差方向的信息，适用于温度场、振动频谱等具有空间相关性的数据。LASSO 回归则通过 L1 正则化惩罚系数，将不重要特征的系数压缩至零，实现特征选择与预测模型构建同步完成，特别适用于故障预测中的关键变量筛选。例如，国家能源集团陡壁电厂在变压器故障预测项目中，初始特征集包含油色谱、电气量及环境参数。应用 PCA 将油色谱数据降至 3 个主成分，再结合 LASSO 回归筛选出 H_2 浓度、 C_2H_2/H_2 比值及顶层油温 3 个核心特征。基于精简特征训练的模型在测试集上 F1 分数达 0.91，较原始特征提升 12%。

2.2 混合预测模型构建

2.2.1 时序特征建模方法

长短期记忆网络因其门控机制能有效捕捉时间序列中的长期依赖关系，适用于设备退化趋势建模^[4]。优化策略包含引入注意力机制聚焦关键时间步、使用双向 LSTM 融合前后向信息，以及结合卷积层提取局部时序特征。例如，浙能嘉兴电厂在 1 号主变压器热点温度预测中，构建了包含环境温度、负载电流、冷却器状态及历史温度的时序数据集。采用 Attention-BiLSTM 模型，输入窗口为 72 小时，预测未来 24 小时温度。模型在夏季高峰负荷期间，提前 6 小时预警 1 相热点温度将超限（实测 98°C ，限值 100°C ），运维人员及时启动备用冷却器，成功将温度控制在 95°C 。

2.2.2 设备关联性分析模型

图神经网络通过消息传递机制聚合邻居节点信息，天然适用于建模设备间的物理或逻辑关联。将电厂电气系统抽象为图结构：节点代表设备（发电机、变压器、断路器等），边代表电气连接或功能耦合关系（如馈线连接、冷却

系统共享）。例如，广东粤电靖海电厂构建了包含 56 个节点（含 4 台机组、12 台变压器、40 台开关柜）的电气拓扑图。GNN 模型基于开关柜温度、变压器油色谱及机组振动数据，检测到 3 机组出口开关柜（节点 A）异常升温与相邻 2 主变（节点 B）乙炔含量上升的关联性。系统定位故障为 A 柜触头烧蚀导致 B 变承受谐波过电压，提前 2 周安排停运检修，发现 A 柜触头灼伤面积达 30%。

2.2.3 多模型融合预测框架

单一模型存在局限，集成学习通过结合多个基学习器提升鲁棒性与精度。加权投票根据各模型历史表现分配权重；Stacking 方法将基模型预测结果作为新特征训练元模型（如逻辑回归、梯度提升树）。例如，国电泰州电厂在发电机绝缘故障预测中，集成 LSTM（时序特征）、GNN（拓扑关联）及 XGBoost（统计特征）三个基模型。采用 Stacking 策略，以基模型输出训练 LightGBM 元模型。该系统在准确预警 4 发电机定子绕组绝缘劣化，经离线介损测试验证 $\tan \delta$ 值达 0.025（阈值 0.02），融合模型的 AUC 为 0.94，显著高于最优基模型（LSTM-AUC0.89）。

3 实时故障预测系统实现方案

3.1 边缘计算层设计

3.1.1 轻量化模型部署

TensorRT 通过层融合、精度校准（FP16/INT8）及内核自动调优优化神经网络推理效率。在英伟达 JetsonAGXOrin 等边缘设备上，可将 LSTM 推理延迟降至毫秒级，满足实时性要求。例如，华电邹县电厂在输煤皮带电机状态监测边缘节点部署 TensorRT 优化的轴承故障诊断模型（原 TensorFlow 模型大小 150MB，优化后 28MB）。在运行中，边缘节点实时检测到 7B 电机驱动端轴承内圈故障特征频率（BPFI）幅值突增，触发本地报警并上传诊断报告至云端。模型推理耗时从原 35ms 降至 8ms，保障了每秒 100+ 振动样本的实时处理。

3.1.2 实时数据流处理

ApacheKafka 作为分布式消息队列，提供高吞吐、低延迟的数据管道。生产者（传感器网关）将数据发布至 Topic，消费者（预测模型、存储服务）并行订阅处理。分区机制与副本策略保障数据可靠性与水平扩展能力^[5]。例如，中电投常熟发电厂构建 Kafka 集群处理全厂 3 万+ 测点数据（峰值吞吐量 12 万条/秒）。数据流经三个主 Topic：

sensor-raw` (原始数据)、`feature-engineered` (处理后的特征)、`model-output` (预测结果)。3 机组励磁系统数据突增期间, Kafka 通过动态增加分区与消费者实例, 维持端到端延迟稳定在 200ms 内, 确保控制系统实时获取预测结果。

3.1.3 本地预警规则引擎

规则引擎基于 Drools 等框架实现, 支持阈值比较 (如油温 $>85^{\circ}\text{C}$)、趋势分析 (如温升速率 $>5^{\circ}\text{C/h}$) 及复合逻辑 (如振动超标且相位突变)。本地执行避免云端通信中断导致的漏报。例如, 国能汉川电厂在给水泵电机边缘节点配置规则: IF 轴向振动速度 $>4.5\text{mm/s}$ AND 3 倍频幅值周环比增长 $>30\%$ THEN 预警等级 = High。该规则触发 1A 给水泵电机预警, 现场检查发现基础螺栓松动。规则引擎在局域网故障期间仍持续运行, 保障了关键设备监控不中断。

3.2 云端分析平台构建

3.2.1 大规模数据存储方案

HadoopHDFS 提供 EB 级分布式存储, HBase 基于列族存储实现高速随机读写, 适用于设备时序数据。数据按设备 ID 与时间戳联合索引, 支持毫秒级范围查询。例如, 华润海丰电厂采用 HBase 存储全厂 5 年历史数据 (约 30TB), 表设计为: 行键 = `设备 ID|年月日`, 列族 = `cf_metrics` (温度、振动等)、`cf_events` (告警、操作记录)。查询某变压器 (ID=TR-201) 的顶层油温数据, 响应时间 $<50\text{ms}$, 满足分析平台交互需求。

3.2.2 预测模型在线更新机制

增量学习通过新样本微调模型参数, 避免全量重训练开销。迁移学习将预训练模型 (如通用轴承故障模型) 适配至特定设备, 缩短训练周期。例如, 大唐三门峡电厂基于 FEDAVG 框架实现风机齿轮箱模型联邦增量更新。边缘节点每月上传模型梯度至云端, 聚合更新全局模型。针对 33 风机新出现的调制边频特征, 增量训练后模型召回率从 70% 提升至 88%, 全量重训练需 72 小时而增量更新仅需 2 小时。

3.2.3 可视化决策支持系统

基于 WebGL 的 3D 设备建模实现关键参数空间映射 (如发电机绕组温度场渲染)。故障热力图叠加在厂区平面图上, 直观展示风险分布。例如, 粤电沙角 C 电厂平台集成 Unity 引擎构建全厂 3D 模型。系统生成故障热力图显示 2 机组汽端轴承区域风险值达 0.87 (阈值 >0.8), 同时 3D 模型高亮显示该轴承振动频谱中 2.5 倍频异常。运维团队据此制定针对性点检计划, 发现轴承座共振问题。

4 总结

本研究构建的基于大数据分析的火力发电厂电气设备故障预测技术体系, 实现了从多源数据采集、智能特征提取到混合模型预测的全流程闭环。通过融合时序建模、图神经网络与集成学习技术, 系统显著提升故障早期识别能力。边缘-云协同架构保障了预测的实时性与可扩展性, 可视化平台强化了决策支持效能。未来研究将探索联邦学习框架下的跨厂区知识共享, 并融合物理机理模型以增强预测可解释性。

参考文献:

- [1] 刘紫木. 火力发电厂电气运行中故障原因及应对方法[J]. 中国设备工程, 2024,(08):167-169.
- [2] 王玲. 浅析发电厂电气设备的运行管理及维护措施[J]. 中国设备工程, 2023,(22):57-59.
- [3] 尹响. 深度学习算法在垃圾焚烧发电厂电气设备故障诊断中的应用研究[J]. 中国机械, 2023,(22):77-80.
- [4] 韩云. 基于大数据加速分析的水电站电气设备运行监测技术[J]. 电器工业, 2023,(07):54-57.
- [5] 贾超. 火力发电厂电气运行中故障的原因及对策分析[J]. 中国金属通报, 2021,(10):243-244.

作者简介:

侯丕文 (1982-03), 男, 汉, 湖北荆州, 本科, 助理工程师, 研究方向: 电气设备运行与检修。