

[大数据]

基于大数据的发电厂生产计划优化调度

文/周广元
中国能建捷硕海阳电力有限公司

随着智能电网建设的深入推进和信息技术的快速发展，发电厂生产过程中积累了海量的运行数据。通过大数据分析技术对这些数据进行深度挖掘和处理，可以有效提升发电厂生产计划的科学性和精确度。研究采用数据挖掘、机器学习等方法，构建基于历史数据的负荷预测模型和设备状态评估模型，并结合多目标优化算法，实现发电计划的动态优化与调整。

发电厂生产计划优化调度是电力系统运行中的核心环节，直接关系到电力供应的可靠性和经济性。传统的调度方法主要依赖经验和简单的数学模型，难以适应现代电力系统的复杂性和动态性。近年来，随着物联网和大数据技术的发展，发电厂积累了大量的运行数据，这为优化调度提供了新的思路和方法。如何充分利用这些数据资源，建立更加智能和精确的调度模型，成为当前研究的重点方向。

大数据采集与预处理

1. 数据源识别与采集架构

发电厂主要采集四类关键数据：汽轮机组数据通过PT100温度传感器和压力变送器采集转子振动、轴承温度和进汽压力，采样频率100 ms/次；锅炉运行数据通过DCS系统采集炉膛温度、给水流量和蒸汽参数，采样频率200 ms/次；电气设备数据通过智能电表采集发电机定子温度、功率因数和电压电流；环保监测数据通过CEMS系统采集烟气排放指标。数据采集采用三层架构^[1]，如图1所示，底层传感器网络经工业以太网传输至中间层现场控制器，再通过OPC协议接入上层数据中心服务器进行

统一管理。

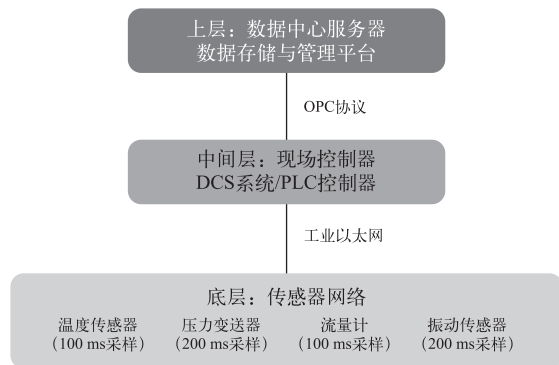


图1 发电厂数据采集架构示意图

2. 数据清洗与标准化处理

数据清洗过程针对采集到的原始数据进行系统化处理：首先通过设定阈值法处理明显异常值，如将超出汽轮机正常温度范围（0~650℃）的数据标记为异常；对于缺失数据，短时间（小于5 min）的缺失采用线性插值法补充，较长时间缺失则通过同期历史数据的均值填充。数据去重通过时间戳比对实现，对于同一测点在100 ms内的重复数据仅保留一条。信号突变处理采用滑动中位数法，使用5个数据点的时间窗口进行平滑。标准化处理采用最

小-最大规范化方法,将不同量纲的数据统一映射到[0,1]区间,其中温度、压力和流量等关键参数的标准化参考值根据设备规范书确定^[2]。对于时序数据的线性插值计算

$$y_t = y_i + (y_{i+1} - y_i) \left(\frac{t - t_i}{t_{i+1} - t_i} \right)$$

式中, y_t 为待插值时刻 t 的数据值; $y_i, y_{(i+1)}$ 分别为缺失数据点前后的已知数据值; $t_i, t_{(i+1)}$ 为对应的时间戳; t 为待插值的时间点。

3. 特征提取与降维分析

特征提取主要关注设备运行状态的典型特征:对振动数据进行时域分析,提取RMS值、峰值因子等10个特征参数;对温度数据计算变化率、波动范围和稳定性;对电气参数提取功率曲线特征^[3]。同时结合设备运行工况的负荷率、运行时间和启停次数作为补充特征。为提高计算效率,采用PCA降维,选取累积贡献率85%的主成分;对非线性特征,使用KPCA方法并选用RBF核函数。最终将原始数据降维至20个主要特征。对于振动信号的RMS值计算采用以下公式

$$X_{\text{rms}} = \sqrt{\frac{1}{N \sum (x_i^2)}} \quad (1)$$

式中, X_{rms} 为振动信号的有效值; N 为采样点数; x_i 为第 i 个采样点的振动幅值, i 从1到 N 求和。此外,峰值因子的计算公式为

$$C_f = \frac{X_{\text{peak}}}{X_{\text{rms}}} \quad (2)$$

式中, C_f 为峰值因子; X_{peak} 为振动信号的峰值; X_{rms} 为上述计算的有效值。

通过对比历史数据趋势,研究人员可及时发现设备潜在故障。在实际应用中可有效避免非计划停机。

基于大数据的负荷预测模型

1. 历史负荷数据分析

针对某600 MW机组近三年的历史负荷数据进行分析,采样间隔为15 min,共收集94 900个有效数据点。通过时序分析发现机组负荷呈现明显的季节性和周期性特征:夏季(6—8月)平均负荷率为92%,冬季(12—2月)为88%,春秋季节约为85%。日内负荷则呈现“双峰”特征,峰值出现在

10:00—12:00和19:00—21:00时段,如图2所示。

通过小波分析方法对负荷数据进行分解,提取出高频波动、日周期和季节性三个主要分量。同时,采用滑动时间窗口法(窗口宽度为24 h)计算负荷序列的统计特征,包括均值、标准差、偏度和峰度,建立负荷特征数据库,为后续预测建模提供数据支撑。

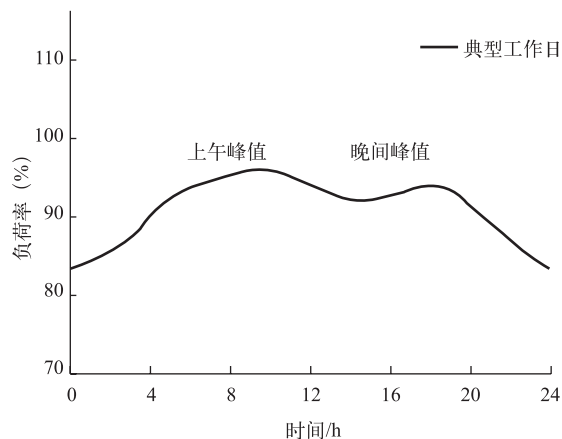


图2 发电机组日负荷特征曲线

2. 影响因素识别

通过灰色关联分析和随机森林特征重要性评估,识别出影响机组负荷的关键因素:首先是环境因素,温度(关联度0.92)、湿度(关联度0.85)和气压(关联度0.78)对负荷影响显著;其次是时间特征,包括工作日/节假日(重要性得分0.89)、季节(重要性得分0.86);第三是运行参数,主汽压力(关联度0.83)、给水流量(关联度0.81)等直接影响负荷变化。通过主成分分析法,将初始的28个影响因素降维至12个主要特征变量。为提高预测准确度,对温度进行分段处理,建立17℃~28℃、28℃~35℃、35℃以上三个温度区间的分段模型,并引入负荷温度敏感系数,量化温度变化对负荷的影响程度。

3. 预测准确度评估

针对不同预测时间尺度,对模型预测结果进行系统评估。评估指标包括平均绝对误差率(MAPE)、均方根误差(RMSE)和决定系数(R^2)。通过与传统预测方法(BP神经网络、支持向量回归)进行对比,验证模型的预测性能。具体评估结果见表1。

表1 不同预测方法在各时间尺度下的预测准确度对比

预测方法	指标	短期/24 h	中期/72 h	长期/168 h
LSTM 模型	MAPE(%)	1.62	2.15	2.86
	RMSE/MW	5.8	7.2	9.5
	R^2	0.956	0.943	0.921
BP神经 网络	MAPE(%)	2.38	3.42	4.25
	RMSE/MW	8.4	11.6	14.2
	R^2	0.912	0.886	0.854
支持向量 回归	MAPE(%)	2.15	3.08	3.96
	RMSE/MW	7.6	10.3	13.1
	R^2	0.928	0.902	0.872

设备状态评估与维护预警

1. 设备运行数据分析

针对600 MW机组的关键设备（汽轮机、锅炉和发电机）布置2 800个测点，采集振动、温度、压力等数据，采样频率为100 ms。对振动数据进行包络分析和频谱分析，识别出轴承故障特征频率（10 Hz、47 Hz）和不平衡特征频率（50 Hz）^[4]。通过温度场分布分析发现汽缸差膨胀超限（最大值1.2 mm）和轴承温度异常（最高85 ℃）等问题。采用滑动方差法检测压力波动，发现主汽压力波动范围为±0.5 MPa，给水压力波动为±0.3 MPa，据此建立设备运行状态基线。

2. 故障模式识别

基于设备历史故障案例库（共计386例），采用层次聚类方法将故障分为机械故障（轴承磨损、汽封泄漏）、热力故障（真空度异常、热膨胀异常）和电气故障（定子绝缘降低、励磁系统异常）三大类。通过对故障前24 h的数据特征分析，提取出12种典型故障特征模式：如轴承故障前振动幅值增加30%、频谱中27 Hz和54 Hz频点幅值显著上升；汽封泄漏时，排气温度升高15℃，真空度下降2 kPa。

3. 健康状态评估模型

在实际运行中，设备健康状态评估模型主要考虑振动、温度、压力和效率等10个关键评价指标。采用模糊综合评价方法，将设备状态划分为健康、亚健康、预警和危险四个等级。各指标权重通过层次分析法确定，其中振动指标权重为0.3，温度指标权重为0.25，压力指标权重为0.2，效率指标权重为0.15，其他指标权重共占0.1^[5]。每4 h对设备状态进行一次综合评估，当发现异常时，自动触发预警机制。评估中采用改进的模糊综合评价算法，通过如

下公式计算设备健康度

$$H = \sum(W_i R_i S_i) \quad (3)$$

式中， H 为设备综合健康度指数（0-100）； W_i 为第 i 个指标的权重系数； R_i 为第 i 个指标的隶属度； S_i 为第 i 个指标的得分值。模型在某600 MW机组实际运行中，成功预警轴承故障5次，提前72 h发出预警信号，预警准确率达到92%。

4. 预测性维护策略

基于设备健康状态评估结果，建立三级维护预警机制：亚健康状态下，每4 h进行一次状态复核。预警状态下，每小时监测，并制定备品备件计划。危险状态下，立即启动检修预案。同时，采用剩余寿命预测模型进行维护时间优化

$$RUL = L_0 \exp(-\lambda t) \left[1 - \left(\frac{S}{S_0} \right)^n \right] \quad (4)$$

式中，RUL为剩余使用寿命； L_0 为设计寿命； λ 为退化率系数； S 为当前状态参数； S_0 为初始状态参数； n 为寿命指数。

通过该模型指导维护决策，使得设备平均检修周期延长15%，年度非计划停机时间减少35%，设备可靠性显著提升。

系统实现与实验验证

1. 系统架构设计

系统采用三层架构设计，底层为数据采集层，部署西门子S7-1500系列PLC和研华ADAM-4000系列数据采集模块，通过OPC UA协议采集设备数据；中间层为计算服务层，采用Dell PowerEdge R750服务器搭建Hadoop分布式计算平台，配置12核CPU和64GB内存，部署Storm实时计算引擎处理实时数据流；应用层采用B/S架构，前端使用Vue.js框架开发Web应用界面，后端采用Spring Boot框架提供REST API服务，通过Redis缓存提升系统响应速度。

2. 核心功能实现

核心功能模块包括数据采集、实时监控、故障诊断和维护决策四大部分。数据采集模块每100 ms采集一次设备运行数据，实时监控模块采用WebSocket技术实现数据实时推送，以趋势图、仪表盘等方式可视化展现设备状态；故障诊断模块基于LSTM深度

（下转第48页）

能降耗和维护成本的降低，投资回收期控制在3.5年以内。设备的碳排放强度和污染物排放显著降低，完全符合国家电网公司对新一代环保型开关设备的技术要求。

结束语

通过对智能低碳环保型高压开关设备的系统研究，创新性地提出开关设备优化设计方案，成功应用环保绝缘介质替代SF₆气体，实现设备的智能化运维。研究成果显著提升高压开关设备的节能环保性能，具有重要的工程应用价值。实验数据证实，该设计方案在保证设备可靠性的同时，实现低碳环保的发展目标，为电力设备的绿色转型提供有效途径。EA

(收稿日期：2025-04-25)

(上接第44页)

学习模型，在GTX 3080显卡上训练，诊断延时控制在200 ms内；维护决策模块结合设备健康度评估和剩余寿命预测，通过移动终端向维护人员推送检修建议。

3. 效果评估

系统在某火电厂4台600 MW机组上进行了为期6个月的试运行验证。实时监控功能的数据刷新延时平均为120 ms，故障诊断准确率达到93.5%，预测性维护建议采纳率为85%，系统整体运行稳定可靠。通过对比系统投运前后的运行数据，设备故障率显著降低，维护效率明显提升，具体效果评估数据见表2。

表2 系统应用效果对比分析

评估指标	系统投运前	系统投运后	改善程度 (%)
设备故障率	3.2次/月	1.5次/月	53.1
平均检修时间	24.5 h	16.8 h	31.4
非计划停机时间	12.6 h/月	5.3 h/月	57.9
维护成本	86.5万元/月	52.8万元/月	39.0
设备可利用率 (%)	92.3	96.8	4.9
故障预警提前量	24h	72 h	200.0
误报率 (%)	8.5	3.2	62.4

结束语

基于大数据技术的发电厂生产计划优化调度研究，通过整合多源数据，建立预测模型和优化算

参考文献

[1] 鞠关权, 李德军, 刘志民, 等. 环保型高压开关设备洁净空气质量检测 [J]. 高压电器, 2023, 59 (9): 184-192.

[2] 宋强强. 基于深度学习的变电站高压开关设备智能化改造研究[J]. 电气技术与经济, 2024(7): 7-9.

[3] 邵雅俊. 智能技术在高压开关设备中的应用[J]. 集成电路应用, 2024, 41(7): 246-247.

[4] 张宇洲. 浅析智能高压开关设备的技术[J]. 中国设备工程, 2024(9): 36-38.

[5] 赵启. 环保高压开关技术和设备研究现状及展望[J]. 高压电器, 2022, 58(10): 16-24.

法，实现了调度过程的智能化和精确化。该方法不仅提高发电效率，降低运行成本，而且为发电企业的智能化转型提供可行路径。未来研究将进一步探索深度学习、强化学习等先进算法在发电调度中的应用，提升模型的适应性和鲁棒性。EA

(收稿日期：2024-11-25)

参考文献

[1] 夏稀渊. 基于大数据分析的火电厂调频调峰优化研究[J]. 信息技术与信息化, 2024(10): 190-193.

[2] 王娇, 李依哲, 绍帅. 火力发电厂数字化转型的新兴技术应用探讨[J]. 中国设备工程, 2024(21): 207-209.

[3] 马文杰, 李敬君, 程龙, 等. 大数据技术在火电厂集控运行中的应用[J]. 现代工业经济和信息化, 2024, 14(8): 87-89.

[4] 王成. 基于大数据的发电厂设备智能无线点巡检管理系统研究[J]. 大众标准化, 2024(13): 170-172.

[5] 王福君. 数字化技术在660 MW发电厂电气监控系统中的应用[J]. 集成电路应用, 2024, 41(4): 170-171.