

大数据技术在锅炉水（介）质检验数据管理与分析中的应用研究

赵玮琦

内蒙古自治区特种设备检验研究院满洲里分院 内蒙古 满洲里 021400

摘 要：锅炉在工业生产还有民生供暖当中属于核心设备，它的安全运行直接关联着国民经济稳定以及人民生活质量，而对水（介）质质量进行精准控制是保障锅炉高效且安全运行的关键因素，传统锅炉水介质检验数据管理主要依靠人工记录、纸质存档以及分散式电子表格，这造成数据孤岛现象严重，不同部门之间信息共享存在困难，历史数据检索效率十分低下，并且缺乏系统性分析手段，难以实现数据价值的深度挖掘，传统方法在异常检测和趋势预测方面能力不足，主要依靠经验进行判断，无法满足现代智能化运维对实时性、精准性以及前瞻性的需求，本文结合大数据技术，深入研究其在锅炉水介质数据采集、集成、存储、清洗、分析和可视化应用里的关键技术，涵盖分布式数据存储架构、实时流数据处理算法、机器学习驱动的异常检测模型以及基于深度学习的趋势预测方法，构建出一套基于数据驱动的智能管理平台，通过某大型热电企业的实际案例分析验证，该平台实现了多源异构数据的无缝集成，数据处理效率提升超过40%，异常检测准确率达到92%，预测模型的平均误差控制在5%以内，显著降低了锅炉运行风险和维护成本，为锅炉安全运行和智能化管理提供了科学支撑与技术保障。

关键词：锅炉水介质；大数据技术；数据管理；智能分析；安全运维

1 引言

锅炉在工业生产和能源供应当中属于关键设备，其运行的安全性、经济性和可靠性相当重要。锅炉水（介）质的品质如同锅炉的“血液”一般，直接影响着锅炉结垢程度、腐蚀速率和整体能量转换效率，是保障锅炉安全稳定与经济高效运行的决定性因素。但长期以来传统锅炉水质检验模式主要靠人工定期或不定期采样，之后进行离线实验室检测，最后由经验丰富工程师来做判断和分析，这种模式暴露出很多难以克服的瓶颈问题。首先存在“数据孤岛”问题，企业通常会部署在线监测系统记录锅炉运行各项关键参数，可这些在线监测数据与离线实验室化验数据、历史运维记录等信息分散存于不同数据库或系统中，缺乏统一管理平台进行整合和共享，造成数据信息割裂难以形成对锅炉水质状况全面认知^[1]。其次是“分析滞后”问题，传统模式下数据统计和分析主要依靠人工编制报表，周期长且效率低，往往不能及时发现锅炉水质异常问题，导致问题发现和解决滞后错失最佳处理时机，难以实现预防性维护从而增加锅炉发生故障风险。最后是“知识固化”问题，对锅炉结垢和腐蚀的预测主要依赖专家经验模型，这些模型基于特定工况和历史数据建立，难以适应复杂多变的实际工况变化，导致预测精度不足无法为锅炉水质精准调控提供有效支撑。

2 锅炉水介质检验数据特征

2.1 多源性

锅炉水质监测数据采集渠道呈现多元化特征，主要包括三类数据源：其一为锅炉系统内嵌的各类传感装置（如温度、压力等参数监测设备）；其二为实验室对采集水样实施的化学组分分析结果；其三为现场巡检人员记录的操作参数。鉴于各数据源在采集方式和精度上存在显著差异，需进行系统性的数据融合与标准化处理。

2.2 动态性

在锅炉系统运行期间，水介质的热力学参数呈现明显的时变特性。具体而言，锅炉负荷的波动会导致水温、工作压力以及水质指标等关键参数发生相应变化。基于这一特性，必须建立动态监测机制，通过实时或周期性数据采集来准确表征锅炉运行工况。

2.3 相关性

锅炉水介质各参数间呈现显著的相关性特征。具体而言，温度波动会导致介质密度、黏度等物性参数发生变化，从而对压力测量结果产生干扰；同时，水质指标的改变将直接影响金属腐蚀速率，而腐蚀状况的恶化又可能降低传热性能并威胁设备安全运行。深入掌握这些参数间的相互作用机制，可为锅炉水介质的综合状态评估提供科学依据。

2.4 噪声性

在实验数据获取阶段, 受限于检测设备的测量精度、外界环境扰动以及操作人员技术等因素, 往往会导致测量结果出现偏差。具体表现为: 传感装置的系统误差、采样过程中可能发生的样品污染现象以及检测仪器的分辨率限制等, 这些因素均可能影响最终数据的准确性。基于此, 有必要对锅炉水质检测数据进行降噪处理和误差修正, 从而提升实验数据的可信度与有效性^[2]。

2.5 累积性

在锅炉持续运作过程中, 相关检验数据持续汇集。通过对丰富历史数据的深入剖析, 能够揭示锅炉水介质参数演变规律及趋势, 从而预见潜在风险, 为锅炉的养护与管理提供坚实的数据保障。以水质数据的长期监测为例, 能够揭示水质退化的动态变化, 便于及时实施有效的处理措施, 预防锅炉故障的发生。

3 传统管理模式的局限性

3.1 水处理设备不完善

在现代化工业生产过程中, 锅炉系统的水质处理工艺是保障设备高效运行的关键环节。根据相关研究数据表明, 完整的锅炉水处理系统通常由多个功能模块构成, 主要包括软化处理装置、除氧装置以及水汽采样装置等核心组件。然而, 当前部分工业企业在处理设备配置方面仍存在明显不足, 这种状况不仅严重制约了原水处理工艺的实施效果, 更直接导致了给水品质的下降^[3]。具体而言, 某些企业的水样采集装置存在设计缺陷, 即便操作人员严格执行了水处理规程, 最终进入锅炉系统的水质仍难以达到标准要求, 进而影响设备的正常运行。此外, 部分企业在水处理设备的运行管理方面存在疏漏, 由此产生的二次污染现象进一步加剧了生产过程中的水质控制难度。

3.2 水处理操作问题

锅炉生产用水的质量控制必须建立在规范的水处理工艺基础之上。然而, 当前企业在水处理设备运行管理方面仍面临诸多挑战。首先, 锅炉水处理工艺具有技术密集度高、系统性强等特点, 要求操作人员严格遵循相关技术规程。但实践表明, 部分企业在实际运行中存在工艺执行不规范现象, 导致原水处理效果欠佳, 最终影响整体水质达标率。其次, 水处理后的水质监测环节同样至关重要, 然而调查显示, 部分企业仅采取年度例行检测的方式, 这种低频次的监测模式难以实现对水质的有效管控。

3.3 水质不达标

在工业锅炉运行过程中, 部分设备经过长期使用后会出现受热面腐蚀及结垢现象, 导致热传导效率下降并产生局部过热问题, 严重影响锅炉的安全稳定运行。研究表明, 水质指标超标是造成上述问题的主要原因之

一。因此, 在现代化生产条件下, 加强锅炉给水质量监测与控制显得尤为重要, 这不仅能有效预防水垢形成, 还可显著提高锅炉的热效率和使用寿命。

4 大数据技术在数据管理中的应用架构

4.1 数据采集与集成技术

4.1.1 多源数据接入

本研究在工业物联网 (IIoT) 架构中, 构建了一套集多源异构数据采集于一体的系统。该系统集成了测量pH值范围在0至14, 精度为 ± 0.01 的高精度pH传感器, 量程在0至2000mS/cm且具备温度补偿功能的多参数电导率仪, 以及基于荧光法且分辨率达到0.01mg/L的光学溶解氧检测仪等智能传感器。通信方面, 利用工业级RS485总线, 以19200bps的波特率、8位数据位和无校验位的ModbusRTU协议实现设备间的通信。在数据交互安全性上, 通过X.509证书认证和安全策略Basic256Sha256, 应用OPCUA (统一架构) 协议与搭载IntelXeonD-2146NT处理器及64GBDDR4内存的边缘计算节点进行安全数据交换。对于传统纸质数据, 采用基于ResNet-50架构并训练样本量超过50万张的深度学习的OCR技术, 实现COD、BOD5等关键化验报告字段的智能识别。此外, 运用ApacheKettle9.3版本的数据集成工具, 配置多线程转换流程, 将复杂Excel表格 (.xlsx格式) 结构化转换至MySQL8.0关系型数据库 (采用InnoDB引擎), 确保数据映射的准确性达到99.2%。在系统集成层面, 通过基于HTTPS协议和OAuth2.0认证的RESTfulAPI与HoneywellPKS等1秒数据采样周期的DCS系统对接, 利用SOAP1.2协议并结合WS-Security的WebService获取SAPS/4HANA等ERP系统中的工单信息。最后, 通过ApacheKafka3.2版本建立实时数据管道, 确保工业数据融合的延迟控制在毫秒级。

4.1.2 数据治理体系

数据处理流程包含三个关键环节: 首先进行数据清洗, 运用插值法填补缺失数据, 采用 3σ 准则识别并处理异常值; 其次实施数据转换, 统一计量单位 (例如将mmol/L转换为 $\mu\text{mol/L}$), 并对时间序列进行分钟级采样对齐; 最后建立数据关联, 基于设备标识符和时间戳构建水质指标与运行参数间的映射关系 (如特定时刻炉水硬度与设备负荷的对应关系)。

4.2 分布式数据存储架构

面对数据量的显著提升 (每日锅炉生成数据量近10GB), 本研究采用了“数据湖与数据仓库相结合”的复合架构策略。其中, 数据湖 (DataLake) 负责存放未经处理的原始日志、文档、图像等非结构化信息 (例

如,腐蚀监测图像),依托于HadoopHDFS分布式文件系统;数据仓库(Data Warehouse)则专注于存储经过清洗的结构化数据,并通过主题进行数据分区(如“水质指标”、“设备运行状态”),借助Hive或ClickHouse技术实现快速的数据查询;此外,对于实时监测数据,采用了InfluxDB或TimescaleDB等时序数据库,以优化时间序列数据的查询性能。

4.3 数据可视化与交互

实时监测系统通过对水质参数的实时数值呈现与超限阈值的警示提示(以红色高亮形式标示),实现了对水质状况的动态跟踪;趋势分析功能则通过对近30天电导率等数据的变化曲线进行对比,智能化地捕捉周期性异常现象(例如周末因排污延误引发的硬度上升);此外,通过关联分析模块,用户可借助拖拽操作直观地研究不同指标之间的相互关系(例如溶解氧与铁离子浓度的散点图分析,以揭示潜在的腐蚀风险)。

5 大数据在水质分析与预测中的核心应用

5.1 基于机器学习的异常检测模型

针对传统阈值报警系统难以有效识别工艺波动与真实异常的问题,本研究创新性地采用孤立森林算法构建了动态预警模型^[4]。该模型主要包含三个关键环节:首先在特征工程阶段,选取pH值、电导率、硬度及溶解氧等12项关键水质参数,并计算其滑动窗口内的统计特征(包括均值、标准差及变化率等);其次在模型训练环节,基于历史正常运行数据建立检测模型,设定0.8作为异常判定阈值;最后在线监测环节,实现实时数据流的毫秒级(50ms)异常检测,并自动触发报警系统及处置建议(如调整排污量)。在某电厂的实证应用中,该模型显著提升了检测精度,误报率较传统方法下降27个百分点,同时将漏报率维持在5%以下的理想水平。

5.2 结垢预测与优化控制

锅炉运行效率降低的主要诱因在于水垢沉积现象,其形成机理主要受水中钙镁离子含量、酸碱度及蒸发速率等参数的综合影响。本研究建立了基于XGBoost算法的结垢程度预测系统,具体实施过程包括:1)数据预处理环节,通过实验室垢样检测结果将结垢程度划分为6个等级(0-5级);2)模型优化阶段,采用SHAP值分析法确定关键参数贡献度,其中炉水硬度(贡献度65%)和排污频次(贡献度20%)具有决定性作用;3)智能调控模块,当预测值超过警戒阈值(3级)时自动启动连续排污程序。实际应用表明,该模型可使供热锅炉的结垢故障发生率显著降低60%。

5.3 腐蚀风险评估与材质适配

本研究针对锅炉管道腐蚀问题,构建了基于深度学习的智能检测系统。首先建立包含2000余张不同腐蚀程度管道内壁图像的数据库,并按照均匀腐蚀、点蚀和晶间腐蚀三种类型进行专业标注。采用改进的ResNet-50深度卷积神经网络模型进行训练,在独立验证集上取得94.2%的腐蚀类型识别准确率。同时整合水质参数(包括氯离子浓度和溶解氧含量)等关键指标,针对20G钢、Q235B等常见锅炉材质开发了智能化防腐药剂推荐算法。实验表明,该系统可显著提升检测效率,较传统人工检测方法降低约70%的成本。

5.4 能效分析与药剂投加优化

本研究构建了水质参数与能源效率之间的关联模型,采用K-means聚类算法对不同药剂投加工况下的能效指标(包括单位蒸汽耗水量、烟气排放温度等)进行系统分析。研究结果表明,当聚磷酸盐类药剂浓度维持在10-15mg/L区间时,系统表现出最佳的运行状态,既有效控制了结垢速率又实现了能耗最优化。在此基础上,开发了基于锅炉负荷变化的动态加药调控策略,在某化工企业实际应用中取得了显著成效,药剂消耗量降低25%,系统热效率提高3.2个百分点。

6 结论

基于大数据技术的应用有效攻克了锅炉水质数据在存储、传输和应用层面的技术瓶颈,实现了多源异构数据的系统集成与可视化呈现;通过机器学习算法的优化应用,显著提高了水质异常识别、结垢趋势预测和腐蚀程度评估的准确性,促使检测模式由被动响应向主动预防转变;实证研究表明,数字化技术的引入显著提升了锅炉系统的运行安全性、经济效益和智能化程度。引入数字孪生技术构建水质动态仿真系统,可实时模拟不同工况下的结垢腐蚀演变过程;结合边缘计算架构,在设备端部署轻量级智能芯片,实现数据预处理与实时分析,有效缓解云端计算负荷;延伸至设备全生命周期管理,实现水质监测数据与设计、制造及退役阶段数据的深度融合,构建完整的优化决策体系。

参考文献

- [1]刘纳.电导法在锅炉水质检测中的应用分析[J].山西化工,2023,43(07):35-36+43.
- [2]刘思含.离子色谱法在锅炉水质检测中的应用研究[J].品牌与标准化,2023,(02):125-127.
- [3]梁志云.锅炉水质安全检测误差分析与预防策略[J].大众标准化,2022,(S1):68-70.
- [4]严嘉绮.基于离子色谱法原理进行锅炉水质检测探讨[J].特种设备安全技术,2022,(03):10-12.