

# 基于大模型的电厂环保数据智能分析与决策支持研究

王 波

内蒙古京泰发电有限责任公司 内蒙古 鄂尔多斯 010030

**【摘要】**本文提出融合轻量化大模型的环保数据智能分析与决策支持体系。首先，梳理电厂环保核心数据维度及行业标准要求；构建“数据预处理-轻量化大模型训练-智能分析-决策输出”四级架构，采用剪枝、量化等技术优化大模型部署效率；最后，通过某火电厂实际数据集验证系统性能。实验结果表明，该系统对 SO<sub>2</sub>、NO<sub>x</sub>等污染物排放预测准确率达 94.7%，异常检测响应时间≤0.3 秒，决策建议合规性符合《火电厂大气污染物排放标准》（GB13223-2011）要求，较传统数据分析方法效率提升 40% 以上。研究为电厂环保治理数字化转型提供了新路径，具有重要的工程应用价值。

**【关键词】**大模型；电厂环保数据；智能分析；决策支持；模型轻量化；环境治理

DOI:10.12417/2705-0998.25.19.071

## 1 研究背景与意义

在“双碳”目标与环保法规日趋严格的背景下，电厂作为能源生产核心主体，面临着污染物排放控制、能耗优化、合规监管的多重压力。据《中国电力行业环保发展报告（2024）》统计，我国火电行业 SO<sub>2</sub>、NO<sub>x</sub>、颗粒物年排放量占工业总排放量的 35% 以上，环保数据的精准分析与高效决策成为电厂绿色转型的关键支撑。大模型凭借其强大的特征提取与复杂推理能力，在工业数据处理领域展现出巨大潜力。然而，原生大模型存在参数量大、推理延迟高、部署成本高的问题，难以直接适配电厂边缘计算环境。因此，研究轻量化大模型在电厂环保数据中的应用，构建“分析-决策-执行”闭环体系，对提升电厂环保治理水平、降低合规风险具有重要的理论与实践意义。本文首先梳理电厂环保数据特征与大模型相关理论；其次设计轻量化大模型架构与数据预处理流程；然后开发智能分析与决策支持模块；最后通过某火电厂数据集进行性能测试，优化系统参数并验证应用价值。

## 2 基于大模型的电厂环保数据智能分析与决策支持系统总体设计

### 2.1 系统架构设计

本文设计的基于大模型的电厂环保数据智能分析与决策支持系统采用分层架构，分为数据层、模型层、分析层与应用层，具体如下。

#### 2.1.1 数据层

负责环保数据的采集、存储与预处理，是整个系统的数据基础与质量保障。数据采集阶段通过标准化 API 接口与电厂现有 CEMS 在线监测系统、设备传感器、ERP 管理系统及环保监管平台实现无缝对接，覆盖气态污染物、水质指标、设备运行参数、能耗数据等多维度信息，采集频率根据数据类型动态调整，确保数据时效性与完整性。存储方案采用“时序数据库+关系型数据库”的混合架构，其中 InfluxDB 专为时序数据设计，具备高写入性能与高效查询能力，可支撑海量监测数据的实时

存储与快速检索；MySQL 用于存储环保法规标准、排污许可证限、设备额定参数等静态数据，便于结构化查询与管理，两者协同实现“动静分离”的高效存储。数据预处理环节针对多源异构数据的特点，采用孤立森林算法识别并剔除传感器漂移、传输异常导致的噪声数据，通过 Z-Score 标准化将不同量级的数据统一映射至 [0,1] 区间，消除量纲影响；采用熵权法确定各数据源权重，通过加权融合技术整合多渠道监测数据，提升数据可靠性，为后续模型训练与智能分析提供高质量数据输入。

#### 2.1.2 模型层

核心为轻量化大模型，是系统智能分析能力的核心载体，包含预训练模块与微调模块，兼顾模型性能与工业场景部署适配性。基础模型选用 LLaMA-27B 开源模型进行二次开发，该模型基于 Transformer 架构，参数量适中，既具备强大的特征提取与复杂推理能力，又避免了超大规模模型的高算力依赖，天然适配电厂边缘计算环境。模型轻量化优化采用“剪枝+量化”的组合方案：剪枝环节基于 L1 正则化结构化剪枝算法，通过分析模型各层参数的贡献度，自动识别并移除 20% 冗余参数，剪枝后模型参数量降至 5.6 亿，且保持网络结构完整性，确保与现有部署框架兼容；量化环节采用 INT8 整型量化技术，将传统 32 位浮点数参数转换为 8 位整数，通过最小化量化误差算法将精度损失控制在 2% 以内，最终使模型存储体积从 28GB 压缩至 7GB，推理速度提升 3 倍。模型微调阶段基于某火电厂 3 年历史数据进行针对性训练，优化模型注意力机制，增强对电厂特定工况的适配性，使模型能精准捕捉行业数据的独特规律，避免通用大模型的泛化性不足问题。

#### 2.1.3 分析层

实现三大核心分析功能，是连接数据层与应用层的关键枢纽，通过模型推理将原始数据转化为有价值的数据结果。排放预测模块基于历史排放数据与设备运行参数，采用时间序列注意力机制聚焦关键影响因素，构建多维度预测模型，可精准预

测未来 24 小时内每 15 分钟的污染物排放浓度，并输出预测置信区间，为电厂提前调整运行策略、规避超标风险提供科学依据。异常检测模块通过双重对比机制识别异常：一方面对比实时数据与模型预测值，当偏差超过 3 倍置信区间时触发数据异常警报；另一方面对比实时数据与环保合规标准值，当污染物浓度接近或超过限值时触发排放异常警报，同时结合设备运行日志分析异常成因，为故障排查提供方向。结合排污许可证要求，实时计算当前运营状态的合规指数，生成低、中、高三级合规风险等级，并标注风险点，形成结构化合规评估报告，替代传统人工审核流程。

#### 2.1.4 应用层

提供决策支持与可视化展示，基于分析结果，生成针对性优化建议，并估算建议实施后的环保效益与成本；采用 Web 端 dashboard 展示实时排放数据、预测结果、异常警报、决策建议等，支持数据导出与报表生成；预留与电厂 ERP 系统、环保监管平台的对接接口，实现决策建议的落地执行与数据上报。

### 2.2 核心算法流程

#### 2.2.1 数据预处理流程

多源数据实时采集，采样频率为 1 分钟/次；采用孤立森林算法检测异常值，去除误差超过  $\pm 10\%$  的数据；采用 Z-Score 标准化公式：

$$x' = \sigma x - \mu$$

其中  $\mu$  为均值， $\sigma$  为标准差；基于熵权法确定各数据源权重，加权融合多渠道监测数据。

#### 2.2.2 大模型推理流程

输入预处理后的环保数据与运行参数，通过 Transformer 编码器提取数据特征；轻量化大模型进行排放预测、异常检测与合规评估；生成分析结果与决策建议，推理延迟  $\leq 0.3$  秒。

## 4 实验验证与结果分析

### 4.1 实验环境

#### 4.1.1 硬件环境

实验硬件选型充分考虑电厂工业场景的部署适配性与实时处理需求，CPU 采用 Intel Xeon 8375C，具备高并发数据处理能力，可支撑多源环保数据的实时接入与预处理；GPU 选用 NVIDIA A100，其张量核心技术可加速大模型推理计算，满足复杂数据分析的算力需求；内存配置 128GB DDR4-3200，保障海量时序数据的高速缓存与读写；存储采用 2TB SSD 固态硬盘，读写速度达 3500MB/s，可高效存储 1.2TB 实验数据集及模型文件，避免存储 IO 瓶颈影响系统响应速度。该硬件组合既适配实验室验证场景，也与电厂常用边缘计算节点的硬件规格兼容，为后续工程化部署奠定基础。

#### 4.1.2 数据集

实验数据来源于某大型燃煤火电厂 2021 年 1 月至 2023 年 12 月的全量环保相关数据，数据集总规模 1.2TB，数据完整性达 99.2%，经预处理后异常值占比仅 1.8%，数据质量满足模型训练与验证要求。具体数据构成如下。

(1) 排放数据。涵盖  $\text{SO}_2$ 、 $\text{NO}_x$ 、 $\text{CO}_2$ 、颗粒物等气态污染物浓度，废水 pH 值、化学需氧量、氨氮含量等水质指标，所有数据均通过 CEMS 在线监测系统采集，符合《固定污染源烟气排放连续监测技术规范》(HJ75-2017) 要求；(2) 运行数据。包含锅炉负荷、脱硫剂用量、脱硝催化剂温度、风机转速、燃煤热值等 5 类核心设备的 23 项关键参数。(3) 合规数据。涵盖《火电厂大气污染物排放标准》(GB13223-2011) 等 3 类国家及地方环保标准、电厂排污许可证核定的排放限值、近 3 年历史合规记录。

数据集按 7:3 比例划分为训练集 (0.84TB) 与测试集 (0.36TB)，训练集用于大模型微调与参数优化，测试集用于系统性能验证，划分过程中保持数据的时序连续性与工况覆盖完整性，避免样本分布偏差影响验证结果。

### 4.2 评价指标

#### 4.2.1 分析精度指标

排放预测准确率采用平均绝对误差与均方根误差评估；异常检测准确率计算精确率、召回率与 F1 分数；对比系统评估结果与人工审核结果的一致性。

#### 4.2.2 系统性能指标

推理延迟从数据输入到结果输出的总时间；模型存储体积轻量化后模型的占用空间；资源消耗 CPU/GPU 使用率、内存占用率。

#### 4.2.3 决策有效性指标

电厂技术人员对决策建议的采纳比例；实施建议后污染物减排量、能耗降低率；合规风险等级从高/中降至低的比例。

### 4.3 实验结果与分析

#### 4.3.1 分析精度结果

结果表明，本文系统的分析精度显著优于传统机器学习方法。 $\text{SO}_2$  排放预测 MAE 仅  $3.2\text{mg}/\text{m}^3$ ，较传统 LSTM 模型降低 28%， $\text{NO}_x$  排放预测 RMSE 为  $4.1\text{mg}/\text{m}^3$ ，降低 31%，预测误差均控制在行业监测精度要求 ( $\pm 5\%$ ) 以内，可精准支撑排放管控决策。模型通过时间序列注意力机制，有效捕捉了锅炉负荷波动、脱硫剂用量调整等因素与污染物排放的非线性关联，在机组负荷骤变（如从 50% 升至 80%）等复杂工况下，预测偏差仍小于  $5\text{mg}/\text{m}^3$ 。

F1 分数达 0.92，其中精确率 0.90、召回率 0.94，既能有效识别真实异常，又能减少虚假警报。相比传统方法，系统对低

浓度缓慢超标、间歇性异常等隐蔽问题的识别能力显著提升，召回率提升 14.6%，为故障早发现、早处置提供了有力支撑。

在 1200 个验证时段中，1144 个时段的评估结果与人工审核一致，准确率达 95.3%。系统可自动解读环保法规中的豁免条款，准确区分合规边界模糊场景，避免误判误报，较传统方法的评估效率提升 3 倍以上，且一致性更高。

表 1 精度结果

评估指标	实验结果
SO <sub>2</sub> 排放预测 MAE	3.2mg/m <sup>3</sup>
NO <sub>x</sub> 排放预测 RMSE	4.1mg/m <sup>3</sup>
异常检测 F1 分数	0.92
合规评估准确率	95.3%

#### 4.3.2 系统性能结果

轻量化优化效果显著，系统性能完全满足电厂实时监测与边缘部署需求。模型部署效率大幅提升，通过 L1 正则化结构化剪枝，移除 20%冗余参数，结合 INT8 量化技术，将存储体积从 28GB 压缩至 7GB，便于在电厂边缘服务器中部署，无需依赖云端算力支持；推理速度提升 3 倍，从 0.85 秒缩短至 0.28 秒，远低于电厂环保数据实时分析的阈值要求，可实现“数据采集-分析-决策”的实时闭环；资源消耗控制合理，满负载运行时 CPU 平均使用率 38%、GPU 峰值使用率 65%、内存占用率 42%，适配电厂现有边缘计算硬件，无需额外购置高端设备，降低了工程实施成本。系统连续运行 30 天无故障、无内存泄漏，稳定性满足工业级应用要求。

表 2 系统性能结果

性能指标	实验结果
推理延迟	0.28 秒
模型存储体积	7GB
峰值 GPU 使用率	65%
内存占用率	42%

#### 4.3.3 决策有效性结果

通过在某火电厂 1 个月的试运行，结果显示决策建议采纳率达 82%，主要集中在脱硫脱硝设备参数优化、燃料配比调整等方面；实施建议后，SO<sub>2</sub> 平均排放浓度降低 8.3%，NO<sub>x</sub>降低 7.6%，标煤耗降低 2.1%；合规风险降低率达 78%，未发生超标排放事件，合规检查时间从 2 小时/次缩短至 10 分钟/次。

#### 4.4 对比实验分析

将本文系统与传统数据分析方法（基于 LSTM 的预测模型

#### 参考文献：

- [1] 张正樵.大数据技术在发电厂环保对策中的应用[J].集成电路应用,2024,41(05):370-371.
- [2] 王智,赵海宝,钟剑锋,等.燃煤电厂炉后环保设备数据整合系统分析[J].环境工程,2023,41(S2):1293-1295+752.
- [3] 方鑫,陈良,陈耀峰.燃煤电厂环保数据传输系统智慧感知研究与应用[J].现代工业经济和信息化,2021,11(02):66-68.

+人工决策）进行对比，结果如下。

表 3 对比实验结果

对比指标	本文系统	传统方法	提升比例
排放预测准确率	94.7%	76.2%	18.5%
异常检测响应时间	0.28 秒	3.5 秒	92.0%
合规检查效率	10 分钟/次	2 小时/次	91.7%
污染物减排率	7.9%	3.2%	146.9%

对比结果表明，本文系统在多维度均显著优于传统方法。异常检测响应时间从 3.5 秒缩短至 0.28 秒，满足电厂环保数据实时监测的核心需求，可及时触发应急处置流程；合规检查效率提升 91.7%，解决了传统人工审核耗时久、易出错的痛点；污染物综合减排率达 7.9%，是传统方法的 2.47 倍，环保效益显著，助力电厂实现“双碳”目标。模型训练周期缩短 40%，人工干预频率降低 86.7%，部署硬件成本降低 42.9%，兼顾了技术先进性与经济可行性；系统无需依赖复杂硬件环境，可直接部署于电厂现有边缘节点，且建议具有强可操作性，与电厂实际运营流程高度契合，较传统方法更易落地推广。

#### 4.5 不足与展望

本文研究仍存在一定不足，模型对极端天气下的排放预测精度有待提升；决策建议未充分考虑设备寿命与维护成本；缺乏多电厂数据的泛化性验证。后续可引入气象数据与设备健康状态数据，优化模型预测算法，提升极端场景下的分析精度；融合设备生命周期成本模型，完善决策支持体系，实现环保效益与经济成本的平衡；扩大数据集范围，纳入不同类型电厂数据，提升模型泛化能力；探索大模型与数字孪生技术的融合应用，构建电厂环保数字孪生系统，实现全流程可视化与智能化管控。

#### 5 结论

本文提出一种基于轻量化大模型的电厂环保数据智能分析与决策支持系统，构建涵盖排放、运行、能耗、合规四类数据的电厂环保数据体系，提出了“清洗-标准化-融合”的预处理流程，解决了多源异构数据融合难题；设计基于剪枝与量化的轻量化大模型，将 LLaMA-27B 模型参数量压缩 20%、存储体积压缩 75%，推理延迟控制在 0.3 秒以内，实现了大模型在电厂边缘节点的高效部署；排放预测、异常检测、合规评估三大核心模块，构建了“分析-决策-执行”闭环体系，决策建议采纳率达 82%，污染物减排率超 7%，合规检查效率提升 91.7%；实验证明，系统在分析精度、运行性能、决策有效性方面均优于传统方法，可有效支撑电厂环保治理数字化转型。