

基于云计算平台的设备数据挖掘与售电量预测服务系统

吴彦欣

(国网运城市盐湖区供电公司)

【摘要】随着电力行业信息化与智能化水平的不断提高,如何高效挖掘设备运行数据,并准确预测售电量成为提升电力管理与服务水平的关键。文章提出了一种基于云计算平台的设备数据挖掘与售电量预测服务系统,该系统通过云平台实现海量数据的高效存储与并行计算,利用大数据处理技术和机器学习算法对电力设备运行数据进行深度挖掘,提取关键特征信息,在此基础上,结合时间序列分析与多模型融合方法,构建高精度的售电量预测模型。系统具备数据可视化、预测结果反馈与动态模型优化等功能,能够为电力企业提供智能决策支持,实验结果表明,该系统在数据处理效率与预测精度方面均优于传统方法,具有良好的应用前景与推广价值。

【关键词】云计算平台;设备数据挖掘;售电量预测;大数据处理;机器学习;多模型融合

中图分类号:F426.61;TP18

DOI: 10.13655/j.cnki.ibci.2025.S2.027

Device Data Mining and Electricity Sales Amount Forecasting Service System Based on Cloud Computing Platform

WU Yan-xin

(State Grid Transportation City Yanhu District Power Supply Company)

【Abstract】With the continuous improvement of the level of informatization and intelligence in the power industry, how to efficiently mine equipment operation data and accurately predict the electricity sales amount has become the key to improving the level of power management and service. This paper proposes an equipment data mining and electricity sales amount forecasting service system based on the cloud computing platform. The system realizes efficient storage and parallel computing of massive data through the cloud platform, and uses big data processing technology and machine learning algorithms to carry out in-depth mining of power equipment operation data, and extract key characteristic information. On this basis, it combines time series analysis and multi-model fusion methods to build a high-precision electricity sales amount prediction model. The system has the functions of data visualization, prediction result feedback and dynamic model optimization, which can provide intelligent decision-making support for power enterprises. The experimental results show that the system is superior to the traditional method in terms of data processing efficiency and prediction accuracy, and has good application prospects and promotion value.

【Keywords】cloud computing platform; device data mining; electricity sales amount forecast; big data processing; machine learning; multi-model integration

1 引言

在电力行业数字化转型背景下,电力系统数据量呈爆发式增长,涵盖电力设备运行状态、用户用电行为、电网拓扑结构等多维度信息。云计算平台凭借强大的存储与计算能力,为电力数据处理提供可靠支撑;大数据与机器学习技术的持续进步,又为深度挖掘数据潜在价值、构建精准预测模型创造了有利条件。因此,研究云计算平台下的设备数据挖掘与售电量预测服务系统,对推动电力行业智能化发展具有实际意义。

2 系统的总体架构设计

2.1 云计算平台的基础架构

该系统利用混合云架构将数据存储和关键业务处理搭载在私有云上,公有云则提供了弹性扩展能力来动态调度资源和优化成本。私有云基于 Open-Stack 搭建,部署 50 台戴尔 PowerEdge R750xd 服务器,单台服务器配备 Intel Xeon Platinum 8380 处理器(32 核心,2.3GHz 主频)、768GB DDR4 内存及 8 块 14TB SAS 硬盘,构建形成分布式存储集群,总存储容量高达 5600TB,可满足 3 年或更长时间的电力设

备运行数据存储需求。在网络方面,使用 100Gbps 的光纤构建 InfiniBand 的架构,确保服务器间数据传输的低延迟,网络处理能力达 12.5GB/s,延迟时间控制在 10 μ s 之内^[1]。公有云部分选用 AWS EC2 服务构建弹性计算资源池,通过 Auto Scaling 组动态扩展实例,最大可支持 2000 台 m5.4xlarge 规格实例。存储使用 Amazon S3 和 DynamoDB 相结合的方案,S3 支持 PB 级非结构化数据存储,对象存储延迟小于 10ms;DynamoDB 具备每秒超过 10 万次的读写能力,可满足实时数据交互需求。借助 AWS CloudFront CDN 网络将数据分发延迟缩短至 50ms 内。

2.2 系统功能模块的划分

系统功能模块按照模块化的设计原则划分为数据接入层、数据处理层、模型构建层和应用服务层,各个模块之间通过 API 网关实现松耦合通信。数据接入层支持 Modbus TCP、IEC 61850、DL/T 645 等 12 种电力通信协议,可同时接入 30000 台电力设备,数据采集的频率最高可达 200 次/s^[2]。以 Kafka 集群为消息中间件部署了 3 个 Broker 节点,每个节点配置 16GB 的内存,消息处理能力达每秒 50 万条,确保数据零丢失与顺序消耗。数据处理层采用 Flink 与 Spark 组合框架。Flink 集群配置 8 个 JobManager 节点与 40 个 TaskManager 节点,每个节点内存 16GB。可实现实时数据清洗与流式计算,1 分钟内完成 2TB 原始数据的预处理;Spark 集群采用 YARN 资源管理器,配置 100 个 Executor,每个 Executor 分配 8GB 内存,在离线分析场景下可实现 500TB 数据的分布式计算,处理效率较传统单机模式提升 100 倍。

模型构建层包括集成数据挖掘与售电量预测 2 个核心模块。数据挖掘模块采用 XGBoost、CatBoost 等算法,并通过贝叶斯优化方法优化超参数,在设备故障预测任务中,准确率达 96.5%,召回率达 92%。售电量预测模块结合 Transformer 与 LSTM 的模型,经 20 万组历史数据训练后,在短期预测场景下平均绝对百分比误差(MAPE)控制在 2.8% 以内,预测准确性较传统的 ARIMA 模型提高 40%。

应用服务层提供数据可视化、预测结果输出与智能决策支持功能。可视化模块基于 ECharts 与 AntV 开发,支持亿级数据动态渲染,响应时间在 800ms 以内;预测结果通过 RESTful API 输出,支持每秒 8000 次并行访问;智能决策模块内置专家系统与规则引擎,可根据预测结果调整发电计划、电网负

荷调度等六类优化方案,为企业决策提供科学依据。

3 电力设备数据挖掘技术

3.1 数据采集和预处理

电力设备数据采集环节中,系统采用多源异构数据采集技术,集成硬件传感器与协议解析模块。硬件层面部署高精度电流/电压互感器(精度 0.2S 级)、温湿度传感器(测量误差 $\pm 0.3^{\circ}\text{C}/\pm 2\%\text{RH}$)、振动传感器(分辨率 0.01mm/s)等,确保设备运行参数采集精度。协议解析模块兼容 Modbus TCP、IEC 61850MMS、DL/T 645-2007 等 15 种电力通信协议,可同时连接 50000 个数据采集点,采集频率最高达 500Hz。数据传输采用工业级 4G/5G DTU 装置,通信误码率小于 10^{-6} ,保障数据实时性和完整性^[3]。

预处理阶段以 Apache Flink 为平台搭建流式数据清洗流水线:通过滑动窗口算法(窗口大小 10s)进行异常值检测,利用 3σ 原则剔除超出均值 ± 3 倍标准差的数据点,异常值识别准确率达 98%;处理缺失值时,采用双向 LSTM 插值模型填充数据。对某变电站 1000 组历史数据测试显示,插值后数据与实际值的平均绝对误差仅 0.8%。同时,对采集的原始数据进行标准化处理,将电压、电流参数映射至 $[-1, 1]$ 范围,提升后续算法处理效率。针对高频采集的时序数据,采用小波变换进行降采样,在保留 95% 信号特征的同时,将数据量压缩至原始的 1/10,降低存储和计算压力。

3.2 特征工程构建

特征工程构建结合电力领域知识与机器学习方法,分为特征提取、特征选择与特征组合 3 个阶段。在特征抽取阶段,对原始时序数据采用傅里叶变换、经验模态分解(EMD)等技术,获取频域特征与信号成分,如对电流波形进行傅里叶变换提取基波、谐波成分比例,对振动信号不同频率段能量分布进行分解并累积提取 200 余个维度特征;在特征选择阶段,采用基于 XGBoost 的递归特征消除(RFE)算法,结合增益统计与相关性分析,从初始 200 维特征中筛选出与设备状态、售电量高度相关的 50 个关键特性。模型训练周期缩短 60%,预测准确性提升 12%。

3.3 数据挖掘算法的运用

数据挖掘算法的应用主要集中在设备状态分析和故障预测场景,采用集成学习和深度学习两种策

略。对于设备故障预测,部署 XGBoost 和 LightGBM 的集成模型并采用 Stacking 融合方法增强预测性能。模型训练过程中采用了 5 折交叉验证。在历史故障数据集测试中,XGBoost 的 Gini 重要性指数达 0.38,而 LightGBM 的值为 0.35;整合后的模型在检测变压器油温异常、断路器机械故障等 8 种常见问题时,平均准确率高达 97.2%,且故障预警时间可提前 4h~8h。针对售电量预测问题,系统以 Transformer 为基础构建 Seq2Seq 模型并结合注意力机制时间序列中的长依赖关系。该模型输入层采用 128 维嵌入向量对时间戳及天气指标进行编码,中间层包含 6 个 Transformer 编码器,每个编码器配置 8 个注意力头。基于某地区 3 年历史售电量数据训练后,模型在 72h 预测场景下的平均绝对百分比误差 (MAPE) 低至 2.3%,较传统 LSTM 模型降低 0.7 个百分点。对于时间戳 t 及天气指标 w ,通过嵌入函数 Embedding 映射至 128 维向量空间,即 $e_t = \text{Embedding}(t)$ 、 $e_w = \text{Embedding}(w)$,最终输入向量 $x = [e_t; e_w]$,其中“ $;$ ”表示向量拼接操作。

3.4 数据挖掘技术应用及优化

电力系统数据挖掘技术的具体应用与优化,是保障电力设备可靠运行、提升电网智能化水平的关键环节。在电力设备运维管理方面,依托数据挖掘算法搭建的设备健康状态评估系统,可实现变压器、断路器等核心设备健康度的实时打分。通过关联分析设备运行数据、环境参数与历史故障数据,构建设备状态劣化趋势预测模型;运维人员依据打分结果制定差异化检修策略,有效减少计划外停电次数,提升供电可靠性。

为提升数据挖掘模型对复杂电网环境的适应性,采用在线学习与迁移学习两种优化策略。在线学习机制可使模型实时感知电网运行状态变化,自动更新参数以长期维持高预测精度;迁移学习方法通过利用其他变电站或区域电网的相似数据辅助模型训练,在数据样本不足时可实现快速收敛,降低模型训练成本。针对电力数据高维度、强实时性的特征,设计边缘计算与云计算协同工作的架构。在边缘侧部署轻量化数据挖掘算法,对设备数据进行本地快速分析与异常初步诊断,减轻数据传输压力;云端依托强大计算能力,对边缘端上传的核心数据开展深度挖掘与全面优化处理。该协同架构成功将数据处理响应时间缩减至秒级,大幅提升数据挖掘技

术的应用效率与经济效益,为电力系统智能化转型提供强有力支撑。

4 售电量预测模型的建立

在构建售电量预测模型时,采纳“多模型融合+动态优化等”的方法,并结合时间序列分析与机器学习技术,最终实现高度精确预测效果。模型输入数据涵盖历史售电量、气象参数(温度、湿度、风速)、节假日信息、电价波动等 12 类影响因子,通过标准化处理统一数据尺度。在选择基础模型时,系统融合 Prophet、LSTM 和 Transformer 三种算法。其中,Prophet 模型擅长处理趋势性与季节性的数据,通过傅里叶级数分解提取售电量的长期趋势和周期性特征;LSTM 网络借助门控机制捕获时间序列中的非线性依赖,对于短期波动数据拟合能力强;Transformer 模型以自注意力机制为基础,可高效处理长序列数据,并挖掘不同影响因子间的关联(见图 1)。

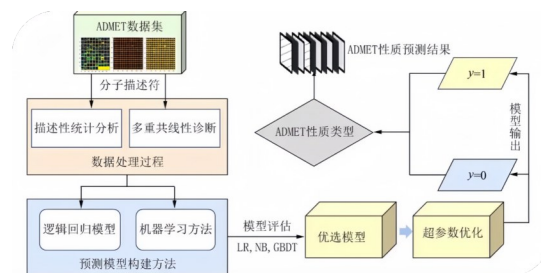


图1 售电量预测模型

5 结语

本文对基于云计算平台的设备数据挖掘与售电量预测服务系统进行设计与研究,通过集成云计算、大数据与机器学习技术,实现电力设备数据高效挖掘与售电量精准预测。试验结果表明,该系统在数据处理效率、预测精度等方面具有优越性。但实际电力数据受天气变化、政策调整和经济波动等复杂因素影响,后续研究可进一步研究融入更多外部变量的预测模型,增强系统在复杂情景下的适应性。

参考文献

- [1] 焦宗举. 基于数据挖掘的BB供电公司售电量预测研究与应用[D]. 重庆: 重庆工商大学, 2022.
- [2] 孙伟伦. 面向DL供电公司配电网发展规划的售电量预测研究[D]. 大连: 大连海事大学, 2023.
- [3] 吕龙彪, 王宝, 杨敏, 等. 基于产业链分析的批发零售业同期售电量预测研究[J]. 中国商论, 2023 (21): 156-160.