

基于大数据分析的电力监控系统故障预警研究

吴少宇¹ 杨晓飞² 黄谊斌² 柳 岩¹ 张 旭²

1. 中国三峡新能源（集团）股份有限公司宁夏分公司 宁夏 银川 750000

2. 三峡新能源红寺堡发电有限公司 宁夏 吴忠 751900

摘 要：随着电力系统智能化发展，海量数据处理与精准故障预警成为关键挑战。大数据分析技术凭借分布式计算、机器学习等优势，可深度挖掘电力监控系统数据价值。研究剖析传统预警系统在数据处理、模型准确性和实时性方面的不足，提出基于大数据分析的故障预警方法，涵盖数据采集预处理、特征提取选择、模型构建及评估。通过LSTM、XGBoost等算法优化模型，多维度指标评估性能，为电力系统安全稳定运行提供技术支撑。

关键词：大数据分析；电力监控系统；故障预警；电力系统

1 大数据分析技术在电力监控系统中的应用基础

1.1 大数据分析技术概述

大数据分析技术是一种融合统计学、计算机科学、机器学习等多学科知识的综合性技术，旨在从海量、复杂的数据中挖掘有价值的信息。随着信息技术的飞速发展，大数据呈现出“5V”特征，即Volume（大量）、Velocity（高速）、Variety（多样）、Value（低价值密度）和Veracity（真实性）。在电力监控系统应用场景下，大数据分析技术主要通过数据采集、清洗、存储、分析等流程，实现对电力系统运行状态的精准洞察。传统的数据处理技术在面对海量电力数据时往往力不从心，而大数据分析技术通过分布式计算框架（如Hadoop、Spark），能够高效处理PB级甚至更高量级的数据。机器学习算法（如决策树、神经网络）与深度学习模型（如卷积神经网络、循环神经网络）的引入，进一步提升了数据分析的自动化和智能化水平。这些技术不仅可以处理结构化数据（如设备运行参数、计量数据），还能对非结构化数据（如设备日志、维护记录）进行有效分析，为电力监控系统提供全方位的数据支持。

1.2 电力监控系统数据结构及特点

电力监控系统的数据来源广泛，涵盖发电、输电、变电、配电、用电等多个环节，数据结构复杂多样。从数据类型来看，包含实时采集的遥测数据（如电压、电流、功率）、遥信数据（设备开关状态、告警信号）等结构化数据，以及设备巡检报告、故障处理文档等半结构化和非结构化数据^[1]。这些数据在时间维度上具有高频采样的特点，例如智能电表每15分钟甚至更短时间采集一次用电数据；在空间维度上，覆盖从变电站、输电线路到用户终端的庞大网络。电力监控系统数据具有明显的特点：一是实时性强，电力系统运行状态瞬息万变，

对数据采集和处理的时效性要求极高；二是动态性显著，负荷变化、设备老化等因素导致数据不断波动；三是关联性复杂，一个节点的异常可能引发整个系统的连锁反应，数据之间存在着复杂的因果关系和耦合特性。这些特点决定了传统的数据处理方法难以满足电力监控系统的需求，必须借助大数据分析技术实现数据价值的深度挖掘。

2 电力监控系统故障预警问题分析

2.1 数据处理能力不足

随着智能电表、传感器等设备的广泛部署，电力监控系统产生的数据量呈指数级增长。传统的数据处理平台在面对海量数据时，常出现数据处理速度慢、存储容量不足等问题。例如，部分地区的电力公司仍采用关系型数据库存储数据，当数据量超过数据库处理能力时，查询响应时间显著延长，甚至出现系统崩溃的情况。此外，传统平台在处理非结构化数据时能力有限，无法充分利用设备日志、文本报告等蕴含的丰富信息，导致数据价值流失。

2.2 预警模型准确性低

现有的电力监控系统故障预警模型多基于阈值判断和简单的统计分析，难以适应电力系统复杂多变的运行环境。阈值设定通常依据历史经验或行业标准，但电力系统受天气、负荷变化等因素影响较大，固定的阈值无法准确反映设备的真实运行状态，容易产生误报和漏报。例如，在高温天气下，电力设备的正常运行温度可能超过常规阈值，导致预警系统频繁误报；而一些早期故障迹象可能因未达到阈值而被忽略，造成漏报，延误故障处理时机^[2]。

2.3 实时性与适应性差

传统故障预警系统在数据处理和分析过程中存在明

显的延迟,无法满足电力系统对实时性的严格要求。从数据采集到预警发布,中间经过多个环节,数据传输、存储和处理的时间累积可能导致预警滞后,错过最佳故障处理时机,电力系统运行工况复杂多变,传统模型一旦建立,难以快速适应新的运行模式和设备状态变化。

3 基于大数据分析的电力监控系统故障预警方法

3.1 数据采集与预处理

数据采集作为电力监控系统故障预警的根基,其完整性、准确性和实时性直接关乎预警的可靠性。在当前智能电网建设进程中,电力系统广泛部署各类智能传感器与智能电表,这些设备如同“神经末梢”,能够以毫秒级甚至更短的时间间隔,对电压、电流、功率等关键运行参数进行高频采集。例如,在大型变电站内,智能传感器可实时监测变压器油温、绕组温度等数据,每秒采集数十次,为系统运行状态分析提供海量基础数据。同时,物联网技术凭借其强大的互联互通能力,打破了数据来源在类型和地域上的限制,将分布在发电侧、输电线路、配电台区以及用户终端的各类数据源有机整合,构建起统一的数据采集平台,实现数据的集中管理与高效传输。然而,原始采集的数据往往存在诸多质量问题。数据噪声会干扰真实信号,缺失值可能导致关键信息丢失,异常值则可能误导分析结果。因此数据预处理不可或缺。数据清洗通过数据查重算法识别并删除重复记录,利用数据校验规则修正错误数据,采用均值填充、回归填充等方法处理缺失值;数据转换借助标准化、归一化等操作,将多样的数据格式统一为便于分析的标准格式;数据集成运用数据映射、数据融合技术,消除不同数据源间的语义差异与结构冲突。

3.2 特征提取与选择

在电力监控系统庞大的原始数据中,充斥着大量与故障预警关联性较弱的冗余信息,如设备的出厂编号、安装位置等基础参数,以及环境温湿度等外部因素数据。若直接将这些数据用于模型训练,不仅会增加计算负担,还可能降低模型性能。因此特征提取与选择成为挖掘数据核心价值的关键步骤。其核心目标在于从海量原始数据中筛选出最能表征设备运行状态、与故障发生紧密相关的关键特征,通过去除冗余信息,降低数据维度,从而提升模型训练效率与预测准确性。在特征提取方面,主成分分析(PCA)通过线性变换将原始变量转换为少数几个互不相关的主成分,保留数据主要方差信息;线性判别分析(LDA)则基于类别信息,寻找最能区分不同类别的投影方向。傅里叶变换和小波变换作为信号处理领域的重要工具,能够将时域信号转换为频域特征,

揭示信号的频率成分与变化规律。在特征选择环节,通过计算皮尔逊相关系数、互信息等指标,量化评估各特征与故障之间的关联程度,进而筛选出最具代表性的特征组合。例如,在某地区电网的故障预警实践中,通过特征提取与选择,将原始数百个特征压缩至20个关键特征,模型训练时间缩短40%,且预测准确率提高15%。

3.3 故障预警模型构建

故障预警模型构建是电力监控系统实现精准故障预判的核心所在。鉴于电力系统运行数据具有时序性、非线性和复杂性等特点,单一的传统模型难以满足实际需求,需结合大数据分析优势,选用合适的机器学习与深度学习算法。对于具有明显时间序列特性的电力数据,如电网负荷曲线、设备运行参数的时间变化序列,循环神经网络(RNN)及其变体LSTM、GRU凭借独特的记忆单元和门控机制,能够有效捕捉数据中的长期依赖关系,准确预测未来短时间内设备状态变化趋势,实现短期故障预警;而对于设备红外热成像图、局部放电图像等可视化数据,卷积神经网络(CNN)利用卷积层和池化层,可自动提取图像中的纹理、边缘等关键特征,判断设备是否存在过热、放电等潜在故障^[1]。在实际模型构建过程中,需紧密贴合电力系统复杂的运行场景进行算法优化。针对电力数据中故障样本稀缺、正常样本占比过高的不平衡问题,除了采用SMOTE等过采样技术,还可结合欠采样和代价敏感学习策略,调整模型对不同类别样本的学习权重。为增强模型的可解释性,便于电力运维人员理解故障预警依据,随机森林和梯度提升树等集成学习算法通过组合多个基学习器,不仅提高了模型预测性能,还能通过特征重要性排序,直观展示各特征对故障预测的贡献程度,为故障诊断和运维决策提供有力支持。

3.4 模型验证与评估

模型构建完成后,其性能优劣需经过严格的验证与评估,以确保在实际电力监控场景中能够稳定、准确地实现故障预警。交叉验证作为模型验证的常用方法,将数据集合理划分为训练集、验证集和测试集。训练集用于模型参数的学习与优化,验证集在模型训练过程中实时监控,用于调整超参数,避免模型过拟合;测试集则在模型训练完成后,对模型进行最终性能评估,衡量其在未知数据上的泛化能力。评估指标的选择直接影响对模型性能的判断,准确率、召回率和F1值从分类结果的角度,衡量模型正确识别故障的能力;均方误差(MSE)则从数值预测的角度,反映模型预测值与真实值之间的平均误差。在电力监控系统中,还需结合行业

实际需求引入特定指标。平均故障检测时间（MTTD）体现模型对故障的响应速度，直接关系到故障处理的及时性；误报率反映模型错误预警的频率，过高的误报率会干扰运维人员正常工作，降低预警系统可信度。通过多维度、全方位的评估，能够全面了解模型在不同场景下的性能表现，精准定位模型存在的问题，如预测偏差、响应延迟等，进而针对性地优化模型结构、调整算法参数或改进数据处理流程，确保模型在实际应用中发挥最大效能。

4 基于大数据分析的故障预警模型构建

4.1 模型选择与算法介绍

在电力监控系统故障预警模型构建中，模型选择至关重要。综合考虑电力数据的特点和故障预警需求，可选择多种算法进行建模。深度学习算法如LSTM网络，因其在处理时序数据方面的优势，能够有效捕捉电力设备运行状态的动态变化规律，适用于长期故障趋势预测；而XGBoost（极端梯度提升树）作为一种集成学习算法，具有训练速度快、泛化能力强、可解释性好等特点，适合处理结构化数据，可用于短期故障预警。LSTM网络通过引入记忆单元和门控机制，能够有效解决传统RNN网络存在的梯度消失和梯度爆炸问题，更好地学习数据中的长期依赖关系。XGBoost则基于梯度提升算法，通过不断迭代训练弱分类器，逐步减少预测误差，在提高模型准确性的同时，通过特征重要性分析提供模型的可解释性，便于电力运维人员理解故障预警的依据。

4.2 模型训练与优化

模型训练是通过大量的历史数据对选定的算法进行参数调整，使其能够准确拟合数据特征，实现故障预警功能。在训练过程中，首先将数据集划分为训练集、验证集和测试集，训练集用于模型参数学习，验证集用于调整超参数（如学习率、迭代次数、网络层数等），测试集用于评估模型的最终性能。为提高模型的训练效果，可采用多种优化策略。例如，针对数据不平衡问题，可使用SMOTE（合成少数类过采样技术）对少数类故障数据进行过采样，增加故障样本数量；在训练过程中，采用早停法避免模型过拟合，当验证集上的损失函

数不再下降时，停止训练。通过调整学习率、选择合适的优化器（如Adam、Adagrad）等方法，加快模型收敛速度，提高训练效率。

4.3 模型评估

模型评估是检验故障预警模型性能的关键环节。除了常用的准确率、召回率、F1值等指标外，在电力监控系统中，还需结合实际应用场景，引入更具针对性的评估指标^[4]。例如，平均故障检测时间（MTTD）反映模型从故障发生到发出预警的平均时间，该指标直接关系到故障处理的及时性；误报率则衡量模型错误发出预警的频率，过高的误报率会增加运维人员的工作负担，降低预警系统的可信度。通过对模型进行全面、客观的评估，不仅可以了解模型在不同性能指标上的表现，还能发现模型存在的不足，为进一步优化模型提供依据。例如，若模型的误报率较高，可通过调整分类阈值、改进特征选择方法或优化模型结构等方式进行改进；若平均故障检测时间过长，则需考虑优化数据处理流程或选择更高效的算法，提高模型的实时性。

结束语

本研究将大数据分析 with 电力监控系统故障预警结合，有效解决传统预警系统的缺陷，提升了故障预警的准确性与实时性。然而，电力系统运行环境复杂多变，新能源接入带来新挑战。未来需进一步探索更先进的算法与技术，增强模型适应性；深化跨领域数据融合分析，提升故障预警的智能化水平，推动电力系统向更安全、高效方向发展。

参考文献

- [1]李霄,刘思宇.基于大数据分析的电力系统故障诊断与预测研究[J].光源与照明,2024(06):90-92.
- [2]杨硕,范军太,卫伟,马夏野.基于大数据分析的电力系统远程运维及故障诊断[J].电力学报,2021,36(01):84-89.
- [3]李存斌,计丽妍,赵德福,李小鹏.基于数据驱动的电力系统灾变演化及防控研究与展望[J].电力系统自动化,2020,44(09):1-11.
- [4]胡海波.封卫国.大数据环境下电力系统故障预测与健康研究[J].电力需求侧管理,2020,22(3):54-59.