

基于大数据的风电变电站与光伏电站故障预测模型研究

蔡和坤

河北省张家口市康保县国华投资哈路风电场 河北 张家口 075000

摘要: 随着风电和光伏等新能源产业的迅猛发展,风电变电站与光伏电站的规模持续扩大。然而,设备故障不仅会导致发电量损失,还可能带来高昂的维修成本。本文基于大数据技术,深入研究风电变电站与光伏电站的故障预测模型。通过大数据预处理与特征提取、模型构建与优化、大数据处理与存储等技术的综合运用,提升故障预测的准确性与时效性。同时,对模型的实时性与可扩展性进行分析,并通过实际应用评估模型的有效性,旨在为新能源电站的稳定运行与智能化管理提供有力支持。

关键词: 大数据;风电变电站;光伏电站;故障预测模型

引言: 风能和太阳能作为清洁、可再生的能源,在全球能源结构中的占比日益增加。国际能源署(IEA)的数据显示,过去十年间,全球风电和光伏装机容量年复合增长率分别达到15%和20%。我国也在积极推动新能源产业的发展,截至2024年底,风电和光伏装机容量已突破10亿千瓦。但风电和光伏电站的设备运行环境复杂,易受自然因素和设备老化等影响,导致故障频发。以某大型风电场为例,一年因设备故障导致的发电量损失高达5%,维修成本超过1000万元。传统的故障检测方法多为事后处理,难以提前预防故障。而基于大数据的故障预测模型能够对设备运行数据进行实时分析,提前发现潜在故障隐患,降低故障发生率,提高电站的运维效率和经济效益。

1 大数据预处理与特征提取技术

1.1 数据来源与整合

风电变电站与光伏电站的运行依赖多种设备协同作业,数据来源极为广泛,主要涵盖各类传感器、监控系统、设备日志等。传感器作为数据采集的关键设备,可获取丰富的环境参数,如风速、风向、光照强度、温度、湿度等,同时还能监测设备的电压、电流、功率等运行参数。以风电机组为例,每台机组配备超20个传感器,每分钟便能产生数百条数据。监控系统对设备的实时状态进行全天候记录,设备日志则详细记载设备的操作记录和故障信息。为搭建统一的数据集,通常采用ETL(Extract, Transform, Load)工具,先从不同数据源抽取数据,由于各数据源数据格式不尽相同,需要进行格式转换,剔除无效数据,统一数据格式。再对数据进行

清洗,去除重复和错误数据,最后将处理后的数据加载到数据仓库中,为后续的深入分析筑牢基础。

1.2 数据清洗与去噪

在实际采集过程中,原始数据常受到各类因素干扰,存在噪声、异常值和缺失值。噪声会模糊数据特征,干扰分析结果;异常值可能由传感器故障或数据传输错误引发;缺失值则会导致数据不完整,影响模型训练效果。为消除噪声干扰,可采用均值滤波、中值滤波等方法。均值滤波计算邻域内数据的平均值,以此替换当前数据点;中值滤波则取邻域内数据的中值,抑制噪声的同时保留数据细节。对于异常值,可依据数据的物理特性和运行规律设定合理阈值进行检测和修正。以光伏组件的电流数据为例,正常工作时电流处于一定区间,超出该区间的数据可判定为异常值。针对缺失值,线性插值法通过相邻数据点的线性关系估算缺失值,均值填充法则使用该列数据的平均值进行填补,确保数据的完整性与准确性,为后续分析提供可靠数据。

1.3 特征选择与提取

从海量原始数据中挑选与设备故障相关的特征,是提升模型训练效率和预测准确性的关键。相关性分析通过计算特征与故障标签之间的相关系数,筛选出相关性高的特征,使模型聚焦关键信息。主成分分析(PCA)则通过线性变换,将多个特征转换为少数几个主成分,这些主成分能保留原始数据的大部分信息,有效降低数据维度。此外,对原始特征进行组合和变换,能够挖掘数据的潜在信息。例如,将风速和风向组合成风矢量,风矢量不仅包含风速、风向信息,还能反映风的总体特性;将功率和电压计算出电阻值,电阻值的变化可直观反映设备的运行状态,这些新特征的生成有助于提升模型的预测能力。

作者简介: 蔡和坤(1999年3月23日—),男,汉族,山西省大同市人,本科,助理工程师,研究方向:输变电。

2 故障预测模型构建与优化

2.1 模型选择与设计

风电变电站和光伏电站的设备运行环境复杂,其产生的数据不仅变量众多,且具备显著的时序特性。在挑选故障预测模型时,要对各类模型的特性进行深入剖析。支持向量机(SVM)基于结构风险最小化原则,在处理小样本、非线性问题时,能够有效避免过拟合,实现精准分类。人工神经网络(ANN)模仿人类大脑神经元结构,具备强大的模式识别与复杂函数逼近能力,对复杂模式有着出色的识别效果。而长短期记忆网络(LSTM)作为深度学习模型的典型代表,其特有的遗忘门、输入门和输出门结构,能够对设备运行数据中的长期依赖关系进行有效捕捉,尤其适用于处理时间序列数据。因此,本文选用LSTM模型开展故障预测工作。在设计模型结构时,输入层节点数取决于前期精心提取的特征数量;隐藏层通常设置2-3层,每层节点数在64-256间进行灵活调节,以此在模型复杂度与训练效率之间找到最佳平衡;输出层节点数与电站可能出现的故障类别数量一致,保证模型能够准确输出故障预测结果。

2.2 模型训练与验证

将经过严格预处理的数据按照7:3的比例,科学划分为训练集和验证集。在训练阶段,运用训练集对LSTM模型进行反复训练。反向传播算法作为模型训练的核心算法,通过计算预测值与真实值之间的误差,并将误差从输出层反向传播至输入层,不断对模型的权重和偏置参数进行调整,促使模型预测结果与实际值之间的误差达到最小化。在此过程中,选择均方误差(MSE)作为损失函数,这是因为MSE能够直观、准确地反映预测值与真实值之间的偏差程度。同时,采用Adam优化器来动态调整学习率,Adam优化器结合了Adagrad和Adadelat两种优化算法的优点,不仅能够加快模型的收敛速度,还能自适应地调整每个参数的学习率。训练完成后,使用验证集对模型进行全面验证,通过计算准确率、召回率、F1值等指标,对模型性能进行综合评估,确保模型在面对新数据时,具备良好的泛化能力,有效避免过拟合现象。

2.3 模型优化与调参

为进一步提升模型性能,对模型进行优化和调参十分必要。引入正则化方法是防止模型过拟合的有效手段,L1正则化通过在损失函数中添加参数的绝对值之和,使得部分对模型贡献较小的参数变为0,从而实现特征选择;L2正则化则通过添加参数的平方和,限制参数的大小,避免参数过大导致模型过拟合。此外,对模型的超参数,如学习率、迭代次数、隐藏层节点数等进行

细致调整。借助网格搜索、随机搜索等方法,在一定范围内寻找最优的超参数组合。以学习率为例,在0.001-0.1的区间内进行搜索,每次调整幅度为0.001,在大量的实验和计算中,筛选出最适合模型的学习率,进而提高模型的预测准确性。

3 大数据处理与存储技术

3.1 分布式数据处理框架

随着风电光伏电站规模不断扩大,设备运行产生的数据量呈指数级增长,传统单机处理方式在速度与效率上都难以满足需求。基于内存计算的分布式处理框架,通过将数据和计算任务分散到多个节点,减少磁盘读写操作,极大提升了数据处理速度。它提供丰富接口,可完成数据清洗、转换与分析全流程操作。其中,类SQL语法降低了数据查询的技术门槛,让普通开发者也能快速上手;实时处理功能则能对设备运行数据进行不间断监测,一旦出现电流突变、温度异常等情况,可立即捕捉并预警,为故障预测提供及时且准确的数据支持,有力保障电站稳定运行。

3.2 数据存储架构

电站数据存储需综合考量容量、性能和安全性。分布式文件系统采用分块多节点存储策略,每个数据块设置多个副本,即使部分节点故障,也能确保数据完整可用,这种高容错特性使其成为存储设备日志、传感器原始数据等非结构化数据的理想选择。而分布式数据库采用列式存储结构,针对设备参数、故障记录等结构化数据,查询时无需扫描整行,仅读取所需列族数据,大幅提升查询效率。在光伏电站中,运维人员通过它能快速调取光伏组件的电压、电流历史数据,直观分析组件性能变化趋势,为设备维护提供数据依据。

4 模型实时性与可扩展性研究

4.1 实时数据处理体系

为实现设备故障的实时预测,需要构建一套完整的实时数据处理体系。在数据采集与传输阶段,引入高性能消息队列系统。这类系统采用分布式的分区副本机制,将数据分散存储在多个服务器节点上,大大提高了数据传输的可靠性和吞吐量。凭借零拷贝技术与批量处理机制,消息队列系统能够在低延迟的情况下,每秒处理数百万条消息,因此成为风电和光伏电站实时数据传输的理想选择。与此同时,借助实时计算框架,对采集到的实时数据进行深度分析。实时计算框架以微批次的方式,将实时数据流按照预设的时间片,比如500毫秒或1秒进行分割,转化为一系列的数据集。利用数据集中丰富的操作算子,例如对数据进行格式转换、过滤异常数

据、进行数据聚合等操作,深入挖掘数据背后隐藏的关联信息。例如,在某大型风电场,通过这套实时数据处理系统,及时捕捉到风机振动数据的异常变化,提前对风机叶片的故障风险发出预警,有效避免了设备停机造成的经济损失。

4.2 模型可扩展性研究

随着风电和光伏电站规模持续扩大,数据量呈爆发式增长,这对故障预测模型的可扩展性提出了很高的要求。借助分布式训练和模型并行化等技术,可以有效实现模型的横向扩展。以分布式计算框架为例,在分布式训练过程中,通过增加集群节点数量,将海量的训练数据均匀分配到各个节点。每个节点独立进行梯度计算,然后将计算结果汇总,统一更新模型参数,从而完成模型的分布式训练。在模型并行化方面,根据模型的结构特点,将不同的计算任务分配到多个节点。比如在多层神经网络模型中,把输入层、隐藏层和输出层的计算任务拆分到不同节点,极大地减轻了单个节点的计算负担。此外,采用模型压缩技术,例如去除不重要的连接、降低参数精度等方法,减少模型的参数量;运用增量学习技术,使模型能够持续学习新数据,在不增加过多计算资源的情况下,提升模型的计算效率,降低对硬件资源的依赖,确保模型在面对不断增长的数据量和复杂计算任务时,依然能够高效运行,适应电站规模进一步发展的需求。

5 故障预测模型评估与实际应用

5.1 模型评估指标

在对风电变电站与光伏电站故障预测模型进行评估时,仅依靠准确率、召回率、F1值等传统指标远远不够,还需借助均方根误差(RMSE)、平均绝对误差(MAE)等指标,对模型的预测误差展开深入分析。RMSE通过计算预测值与实际值差值的平方和的平均值的平方根,能够直观反映预测值与实际值之间的偏差程度,其值越小,表明模型预测越接近真实值。MAE则通过计算预测值与实际值差值的绝对值的平均值,衡量预测误差的平均大小,MAE不受异常值的过度影响,能更稳定地反映模型的预测准确性。此外,还可结合平均绝对百分比误差(MAPE),从相对误差的角度对模型进行评估。通过综合运用这些指标,构建多维度的评估体

系,从不同层面考察模型的性能,进而全面、客观地评价模型在风电和光伏电站故障预测场景中的表现。

5.2 实际应用与部署

将精心构建的故障预测模型应用于实际的风电变电站与光伏电站中,能显著提升电站的运维管理水平。以某风电场为例,在实际应用过程中,模型凭借对设备运行数据的深度分析,提前72小时精准预测到风机叶片故障,不仅有效避免了发电量损失,还节省了因设备突发故障导致的高额维修成本。在日常运行中,对模型进行实时监测,定期收集模型的预测结果与实际故障情况,通过对比分析,及时发现模型存在的问题。根据电站设备的更新、运行环境的变化等实际情况,对模型进行调整和优化。将模型与电站的运维管理系统深度融合,当模型预测到故障风险时,运维管理系统会自动发出预警信息,运维人员可依据预警,提前安排检修任务,实现智能运维。这一举措极大提高了电站的运行效率,保障了电站运行的可靠性。

结语

本文基于大数据技术,对风电变电站与光伏电站的故障预测模型进行了深入研究。通过大数据预处理、模型构建与优化、大数据处理与存储等技术的应用,提高了故障预测的准确性和时效性。同时,对模型的实时性和可扩展性进行了研究,为模型的实际应用提供了保障。在未来的研究中,可以进一步探索新的算法和技术,提高模型的性能,推动新能源电站向智能化、高效化方向发展。

参考文献

- [1]刘磊,周毅,陈芝屹.基于CNN-BiLSTM-Attention模型的变电站设备运维故障诊断与预测研究[J].电力设备管理,2024,(13):14-16.
- [2]代荣艳,徐册利.基于大数据技术的水电站设备故障诊断与预测研究[J].科技资讯,2024,22(05):36-38.
- [3]石文超,张彤阳.基于大数据分析的变电站故障诊断与维护优化研究[J].通信电源技术,2024,41(02):240-242.
- [4]唐健强.基于深度学习的变电站设备故障预测与运维研究[J].电脑编程技巧与维护,2024,(11):165-167.
- [5]魏潘,温和昌,方琦.基于大数据的变电站设备故障预测与维护策略[J].家电维修,2025,(01):110-112.