

# 大数据处理技术在风电机组故障诊断预警中的应用

桑逢文

(国家能源集团国源电力有限公司内蒙古分公司, 内蒙古 呼和浩特 010010)

**摘要** 面对日趋严峻的碳减排形势, 为早日实现碳达峰和碳中和目标, 我国风电装机规模继续保持稳步增长。本文以大数据处理技术作为研究方向, 分析了风电机组故障诊断预警系统架构、关键技术与故障诊断方法, 并阐述了大数据处理技术在风电机组故障诊断预警领域中的应用路径, 旨在为提高诊断预警精准度和改善成果利用条件、保证故障诊断与预测精度提供借鉴。

**关键词** 风电机组; 故障诊断预警; 大数据处理技术

中图分类号: TP31; TM3

文献标志码: A

DOI:10.3969/j.issn.2097-3365.2025.33.011

## 0 引言

近年来, 我国风力发电事业得到快速发展, 大规模投入运行风电机组, 并把大数据处理技术引入风电机组故障处理领域, 由大数据平台高效处理海量信息, 精准判断故障性质, 提醒工作人员及时解决故障问题, 这对缩短故障平均持续时间、提高风电机组可靠系数起到至关重要的作用。与此同时, 风电机组在线监测与故障诊断技术对于及时了解机组的运行状态、对机组故障进行预警、有针对性地开展设备维护、减少突发性故障和提高设备利用率都具有重要的现实意义。

## 1 基于大数据处理技术的风电机组故障诊断预警系统

### 1.1 系统架构

目前, 风电机组故障诊断预警系统普遍采取分布式架构, 建立故障诊断模型、预警模型, 持续向故障模型输入状态监测数据, 同步输出故障诊断结果, 发送预警信号。模型层次架构即为系统架构, 由数据采集、数据存储、数据计算、数据分析四大核心功能模块组成, 这也是大数据处理流程。其中, 数据采集模块负责实时采集多源异构数据, 以风电机组在线监测系统和地理信息系统 (GIS) 作为主要数据来源, 采取大数据连接器技术, 抽取独立系统相关数据, 按照统一标准, 转换数据格式, 提交到分布式文件系统, 也可选择提交到分布式数据库<sup>[1]</sup>。数据存储模块以分布式文件系统 / 数据库作为数据存储空间, 凭借高容错率与高吞吐量特征, 足以满足海量数据存储需求, 并把数据成本控制在较低水准, 也有利于提高风电机组故障

分析效率。数据计算模块依托大数据模型来完成复杂运算任务, 组合采取 Spark 内存计算、Storm 流计算等多项算法, 根据实际计算需求挑选算法, 如在数据计算任务时效性要求严格时, 采取 Storm 流计算方法。数据分析模块负责从数据计算结果内提取故障特征值, 对比故障知识库内样本案例的相似度, 判定故障性质, 在系统界面上可视化呈现故障诊断信息, 同步发送预警信号。

### 1.2 关键技术

新一代风电机组故障诊断预警系统以强化数据处理能力、提高故障诊断精度作为升级方向, 组合应用到多项大数据处理技术, 核心技术手段包括 Storm 集群处理、Spark 内存批处理、数据挖掘、HDFS 分布式文件系统等, 技术原理及功能定位有所不同。例如: 对于 Storm 集群处理技术, 采取集群数据处理架构, 基于大规模集群, 高效处理海量数据, 从原始数据内提取足够数量的风电机组故障特征信息, 有着不丢失数据内容、具备复杂数据处理逻辑、同步完成诸多指标计算和不同类型数据聚合任务的优点, 根据实际数据处理需求来控制节点数量<sup>[2]</sup>。而对于 Spark 内存批处理技术, 属于一项分布式内存计算方法, 利用内存完成数据处理任务, 支持多种语言, 具备二次开发与定制化设计条件, 有利于提高在线监测数据处理效率, 在更短时间内发现故障隐患。

### 1.3 故障诊断方法

以朴素贝叶斯分类算法为主, 相比其他故障诊断算法, 此项算法有着易于实现、诊断效率高、准确率

高的显著优势，更加满足大规模数据分类处理需求。大数据处理技术以风电机组状态监测数据作为来源，基于数据逻辑关系，相互流动形成逻辑层面的拓扑结构，向 Spout1 组件发送状态监测数据，保存在分布式存储系统，从原始数据内提取包含故障特征参量的记录信息，再把故障特征信息提交给诊断算法进行判定。同时，自动保存故障记录，补充到故障知识库内，作为故障诊断模型和故障预警模型的学习样本。故障特征信息提取完毕后，利用朴素贝叶斯分类算法，求解各类故障问题与特征属性的条件概率，把故障诊断问题转换为故障计数问题，把条件概率相对最高的故障问题作为输出结果<sup>[3]</sup>。

#### 1.4 故障预警方法

提前统计风电机组故障问题，确定故障类型，以及各类故障的形成机理、变化规律及外在征兆，针对性建立多种故障预警模型，以此来保证故障预警信号的时效性、准确性。例如：面向风电机组齿轮箱轴承超温故障，可选择采取 BP 神经网络技术来建立轴温预警模型，高度逼近任意精度的非线性连续函数，模型结构由输出层、隐含层以及输出层组成。有关轴承超温故障的运行参数作为输入层神经元，包括轴承工作温度、风速值、风电机组有功功率等，并把预测前后的轴承温度故障特征作为输出层神经元，神经元相互联系，正向传播输入信号，反向传播输出误差。

### 2 大数据处理技术在风电机组故障诊断预警系统中的应用策略

#### 2.1 选择故障诊断预警特征

大数据处理技术以强大数据处理能力作为核心优势，相比数据处理能力，决策分析能力并不出众，缺少详细规则的前提下，故障诊断精准度缺乏保证。对此，工作人员需要提前在系统内设定各类故障问题的特征信息，作为故障诊断规则，方可精准判定故障性质。从现实层面来看，要求工作人员全面收集整理风电机组历史数据、故障案例，以及其他同类项目的故障案例，统计分析齿轮箱、塔筒结构、传感器、发电机、叶片、液压系统等组成部件的故障结构，把出现频率相对较高、危害程度严重的失效故障作为典型故障。逐一分析各类典型故障问题形成机理与外在征兆，故障机理设定为故障诊断规则，故障征兆设定为故障预警规则，通过增强故障诊断预警机制和故障规律的关联程度，从而提高故障诊断精准度与增强故障预警信号时效性，无需大数据处理技术完成故障规律总结在内的繁琐数据处理任务<sup>[4]</sup>。例如：对于发电机部件，典型故障问

题包括异常振动、发电机过热、绝缘损坏、转子 / 定子线圈短路、转子断条。对于齿轮箱部件，典型故障问题包括齿面擦伤、齿面接触疲劳、断齿、轴体不平衡、轴体疲劳剥落等。同时，故障诊断预警特征并不局限于部件运行参数，把电力参数、风力参数、风电机组状态参数连同其他反馈信号一同作为故障诊断预警模型的输入特征向量。

#### 2.2 状态监测评估

风电机组故障问题具备随机性、突然性的鲜明特征，不同时刻的机组状态和故障出现概率有着显著区别，唯有把实时运行数据作为模型输入值，才能准确判定风电机组当前时刻的运行状态、健康程度和辨别故障隐患。因此，需要把状态监测系统、故障诊断系统和故障预警系统一同接入大数据平台。状态监测数据评估环节，实时采集整理风电机组监测数据，利用故障分类概率矩阵来构成分类器，把预处理完毕的监测数据导入分类器，从中计算各类故障问题的特征值，把出现概率相对最高的故障特征值作为输出结果，用于辨别风电机组当前实质存在的故障问题，同步输出故障诊断报告并发送预警信号。如果要求大数据处理技术可以提前发现早期故障征兆，在问题实质性发生前，提醒工作人员消除故障隐患，则在分类器内植入神经网络模型，把实时监测数据作为输入内容，预测后续时刻故障特征，比对警戒阀值和预测数据，推断故障问题未来变化情况，即可准确判定风电机组早期故障征兆。

#### 2.3 模型训练

故障诊断预警是一项高度复杂的数据处理任务，同时承受准确率、时间双重压力，输出准确结果需要消耗大量时间，需短时间迅速出具故障诊断报告。故障诊断准确率和数据处理规模有着密切联系，数据规模越大，诊断结果可信程度越高，但却会降低故障诊断效率。因此，在大数据处理技术落地应用期间，需要长期组织进行模型训练，以故障案例作为训练样本，故障诊断预警模型通过不断总结完善故障规律，在输入、处理少量故障信息的前提下，即可准确判定故障性质，无需输入过大规模的故障数据。从现实层面来看，当前主要采取 RBM 自训练方法和 DBNS 训练方法。以 RBM 自训练方法为例，可视层神经元与隐含层神经元构成随即神经网络，层内神经元并未相互连接，层间保持全连接状态，向底层神经元输入原始特征向量，向上传递期间，故障特征向量变化成抽象特征向量，经过持续训练学习，更加全面、准确地掌握风电机组各类型失效故障的形成原因、变化规律。

## 2.4 诊断预警界面可视化设计

在早期风电机组故障诊断预警系统，诊断预警结果以二维平面格式进行输出，如数据图表、风电机组参数变化曲线，诊断预警结果不够直观，工作人员查阅图表内容后，才能初步了解故障情况，再采取应对措施，间接造成延长故障持续时间、加剧故障受损程度的后果。为增强故障处理活动时效性，应对故障诊断预警界面进行可视化设计，通过三维化、立体化形式来输出诊断预警结果。例如：建立风电机组总体模型，以及齿轮箱、叶轮、塔筒结构等核心部件的三维模型，模型上标注不同颜色来反映故障位置和受损程度。模型颜色为绿色时，表明风电机组运行工况平稳、正常，无故障隐患。模型颜色为橙色/黄色时，表明风电机组存在故障隐患，或是存在不影响机组运行工况的故障问题，短时发送预警信号，择机开展维修作业来消除故障。模型颜色为红色时，表明风电机组存在重大故障问题，危及机组运行安全，或导致机组停机，持续发送预警信号，直至故障问题得到解决，结束故障预警任务，取消预警信号<sup>[5]</sup>。

## 2.5 丰富系统功能

现阶段，大数据处理技术功能效用并未得到充分发挥，应用场景单一，风电机组故障诊断预警系统仅实现故障辨别、故障诊断、故障预警等少数几项功能，未能全面展示风电机组实时运行状况，也没有深入挖掘更深层次的故障信息。相比传统故障诊断预警方法，对大数据处理技术的应用，虽然在一定程度上提高故障诊断精度、缩短故障持续时间，但总体故障处理能力的提升幅度没有达到预期水准。对此，应把丰富系统功能作为技术价值深入挖掘方向，结合故障处理需求，增设其他使用功能。从实操角度来看，可选择增设寿命预测、性能下降预警、自适应报警等功能，重点发挥系统强大数据处理能力。对于寿命预测功能，建立寿命模型，以风电机组核心部件的平均故障时间、平均失效时间、故障率作为指标，定期评价风电机组健康程度，预测剩余使用寿命。对于性能下降预警功能，不再以故障是否存在作为评价风电机组安全系数的唯一标准，确定风电机组运行参数在短时间内剧烈震荡后，同样判定机组工况不稳，发送预警信号。对于自适应报警功能，重点分析风电机组在不同运行工况下的故障特征，不再按照工作人员先期设定规则来诊断故障问题，而是考虑到风电机组运行条件与故障问题的内在规律，自行生成故障判定阈值，定时重新设定全新判定值，覆盖历史判定值，以此来解决误报警问题。

## 2.6 建立故障知识库与专家系统

利用大数据处理技术，收集、整理和分析大量风电机组故障数据，建立完善的故障知识库。该知识库不仅包含各类故障的特征、机理、征兆等信息，还应记录故障的处理过程和解决方案。同时，结合机器学习算法，对知识库中的数据进行挖掘和学习，不断更新和完善知识内容。

基于故障知识库，构建专家系统。专家系统可以模拟领域专家的思维方式，根据实时监测数据和故障特征，快速准确地诊断故障，并提供相应的维修建议和解决方案。例如：当系统检测到发电机过热故障时，专家系统可以根据知识库中的相关信息，分析故障原因，如冷却系统故障、负载过大等，并给出具体的检修步骤和注意事项。此外，专家系统还可以与运维人员进行交互，解答他们在故障处理过程中遇到的问题，为风电机组的维护提供有力的技术支持。通过建立故障知识库与专家系统，能够充分利用大数据处理技术的优势，提高风电机组故障诊断预警系统的智能化水平和故障处理能力。

## 3 结束语

大数据处理技术为风电机组故障诊断预警系统指明了升级方向，有助于进一步强化故障处理能力，更长时间维持风电机组良好运行工况。工作人员必须加强对大数据处理技术的应用力度，基于故障处理实际需求，着手建设新一代风电机组故障诊断预警系统，落实设定故障诊断预警特征、状态监测评估、模型训练等策略，从而对机组设备进行异动检测，实现运维辅助决策功能。

## 参考文献：

- [1] 赵龙. 大数据处理技术在风电机组故障诊断预警中的应用 [J]. 电子技术, 2024,53(09):112-113.
- [2] 张少敏, 毛冬, 王保义. 大数据处理技术在风电机组齿轮箱故障诊断与预警中的应用 [J]. 电力系统自动化, 2016,40(14):129-134.
- [3] 郑天甫. 数据处理技术在风电机组故障诊断及预警中的应用研究 [J]. 江西电力职业技术学院学报, 2022,35(01): 15-17,26.
- [4] 吉庆昌, 邱英杰, 阴兆武, 等. 大数据处理技术在风电机组故障诊断及预警中的应用 [J]. 中小企业管理与科技 (中旬刊), 2021(10):179-181.
- [5] 申云乔, 江政儒, 叶新红. 大数据技术在风电机组运行状态评估及故障诊断中的应用分析 [J]. 电工技术, 2023 (20):38-40,44.