

# Application of Data Mining Technology in Fault Diagnosis of Wind Turbine Generators

Shengyang Li

Shaanxi Deyuan Fugu Energy Co., Ltd., Yulin, Shaanxi, 719000, China

## Abstract

The status of wind energy and new energy is continuing to improve, and the installed capacity of wind turbines is also increasing year by year, making it gradually become one of the important sources of power supply. However, with the impact of the harsh operating environment, the maintenance cost of wind turbines has been high, and the traditional responsive maintenance scheme has been unable to meet the actual demand. In this context, preventative maintenance also faces technical challenges. To solve these problems, this paper proposes an innovative maintenance scheme, based on wind field big data and fan history data. Through the application of data mining technology, the deep mining of big data and the establishment of optimized fault prediction model can make better use of the value of big data of wind power generation and improve the timeliness and accuracy of fault prediction.

## Keywords

data mining technology; wind turbine; fault diagnosis

## 数据挖掘技术在风力发电机组故障诊断中的应用

李升阳

陕西德源府谷能源有限公司, 中国·陕西 榆林 719000

## 摘要

风能新能源的地位在持续提升, 风力发电机组的装机量也在逐年增加, 使其逐渐成了电力供应的重要来源之一。然而, 随着恶劣运行环境的影响, 风力发电机的维修成本一直居高不下, 传统响应式维护方案已经无法满足实际需求。在此背景下, 预防式维护也面临着技术方面的挑战。为解决这些问题, 论文提出一种创新的维护方案, 基于风场大数据和风机历史数据。通过数据挖掘技术的运用, 对大数据进行深入挖掘, 并建立优化故障预测模型, 可以更好地利用风力发电大数据的价值, 提高故障预测的及时性和准确性。

## 关键词

数据挖掘技术; 风力发电机组; 故障诊断

## 1 引言

全球范围内的风能作为一种绿色可再生能源受到了广泛的关注, 风电装机量以 20%~30% 的速度持续增长。2016 年, 风能装机量达到了历史最高水平, 累计并网装机容量高达 1.48 亿千瓦, 占全部发电装机容量的 9.8%, 年增长率高达 23%。政策的支持和经济的推动为风力发电的发展带来了积极的环境。然而, 风力发电机运行成本高昂, 同时位置偏远、环境恶劣也增加了维护成本。据统计, 故障费用主要来自发电机、齿轮箱以及主轴等部件。因此, 发展故障诊断技术已成为降低运作成本、减少安全事故的重要途径。

## 2 风力发电机组构成

齿轮箱: 齿轮箱是风力发电机的一个重要组成部分,

它可以将低速的风力转化为高速的旋转动力, 从而能够驱动发电机。

发电机: 发电机是风力发电机的核心部分, 它可以通过齿轮箱传递的旋转动力产生电能。

叶片: 叶片是风力发电机捕捉风能的关键部件, 它将通过旋转产生动力, 同时将风能转化为机械能。

轴承: 轴承起着支撑和连接的作用, 它连接着发电机和齿轮箱, 使得整个系统得以运转。

机舱: 机舱是风力发电机的主体部分, 所有的主要部件都安装在机舱内部。

台架: 台架是用来支撑和固定风力发电机的结构, 它通常由钢材或者混凝土制成。

## 3 风力发电机组常见故障

根据需要进一步了解各类故障的具体表现和原因, 以便更好地维护和保养风力发电机。

【作者简介】李升阳(1993-), 男, 中国山西朔州人, 硕士, 助理工程师, 从事电力信息化研究。

### 3.1 齿轮箱

风力发电机组的齿轮箱在运转过程中，通过精密的齿轮咬合滚动，承担着传递动力的重任。其结构设计精巧且稳健，虽然故障率较低，但维修过程复杂且成本较高。在齿轮箱中，轴承高速旋转，带动齿轮咬合转动，以此传递动力。如果出现接触面残留颗粒物较多或摩擦系数较大导致齿轮无法正常运行的情况，轴承就会呈现出非正常运行状态<sup>[1]</sup>。

### 3.2 发电机

风力发电机与小型汽轮发电机在原理上有相似之处。然而，风力发电机在轴承、发电机绕组以及相关电气传输设备的故障上出现的频率较高。振动是由发电机转子在转动过程中产生的，主要原因包括转动时能量的分配不平衡、刚度不对称以及电气传输上存在的缺陷，这些因素导致线圈产生不均匀的电磁力。轴承或相关部件的问题往往与其电气传输特性有关，比如电压、电流等系数的变化。

### 3.3 主轴

主轴作为齿轮箱与叶片、轮毂之间的关键连接部件，具有稳固的结构，能够承受巨大的压力。然而，由于转速相对较低，主轴容易受到转速突变的影响，导致弯曲变形。为了确保主轴在更高压力下的稳定表现，我们建议从主轴承的材料和抗负荷性两方面进行严格把控。主轴长期处于剧烈的转速变化环境中，这使得轴承的老化速度加快。此外，如果轴承衔接口的润滑不足，摩擦过大，可能会引发脱落故障。为了防止此类问题，需要密切关注轴承的润滑情况，确保其稳定运行<sup>[2]</sup>。

### 3.4 故障产生原因分析

风力发电机组是一种复杂的机械设备，其常见故障可以由多种原因导致。以下是一些主要的故障原因：

①部件磨损：风力发电机组的各个部件在长期运行过程中，会因摩擦、冲蚀等原因而逐渐磨损。如主轴、齿轮、轴承等部件的磨损都可能导致设备性能下降，甚至引发故障。

②轴承过热：风力发电机在运行过程中，如果轴承润滑不足或承载过大，便可能产生过热现象。轴承过热会对其正常运转产生影响，严重时甚至可能导致轴承破裂或失效。

③产品寿命到期：风力发电机组的各部件都有其设计使用寿命，如轴承、齿轮等。当这些部件达到其设计寿命时，便可能出现性能下降、故障率增加等问题。

④开裂失效：某些关键部件，如主轴、齿轮等，在受到材料质量、制造工艺、运行环境等因素影响时，可能会出现开裂现象，从而导致设备失效。

## 4 数据挖掘技术在风力发电机组故障诊断中的应用分析

### 4.1 数据挖掘技术要点分析

数据挖掘技术具有较为广泛的应用面，针对不同数据

具有不同类型的处理方式，一般分为分类、关联以及回归三类。

#### 4.1.1 分类

在采用分类算法时，其主要核心就是构建一个分类函数或者是分类器，该类型模型可将相关数据进行集合，然后将其映射到某一给定类别。分类算法作为一种具有较高准确性、安全性的监督学习法，一般用于预测所测数据类别以及数据关系规则中。

#### 4.1.2 关联

在大数据挖掘中，简单关系、时序关系和因果关系这三种关联关系形式因其各自的特性具有重要价值。以故障诊断为例，利用关联法进行故障规则的挖掘是十分有效的。通过匹配故障类型，可以形成一系列的关联规则。具体来说，每一次故障的特征和结果，都被视为一条独立的关联规则。有经验的技术专家对这些规则进行不断完善，以添加到故障关联规则库中。这种方式有助于对未知特征匹配进行判定，以确定是否存在故障或故障类型<sup>[3]</sup>。

关联规则的思路非常清晰。首先设定一个元素集合，再针对 X 的所有记录进行集合设置。每一条记录包含若干个属于特定元素的项。关联规则可以表示为  $X \Rightarrow Y$ ，其中 X、Y 属于 I 且  $X \cap Y = \emptyset$ 。为了控制关联规则的质量，我们设定了两个度量标准：支持度和可信度。这两个参数能够确保关联规则的有效性和准确性<sup>[4]</sup>。

#### 4.1.3 回归

通常情况下，评估建造和使用模型的未知值一般是预测，或者通过评估给定样本，发现尽可能具有的值空间或者是属性，通过构建相应模型，即可估计预测值，一般由时间序列、回归分析两种模型。

①时间序列。在建立时间序列模型过程中，主要是以数据时序为基础，预测固定数量时间点的具体数值，该方法一般用于短期预测中，如下公式（1）所示，自回归模型为  $q=0, P \neq 0$ ；移动平均模型则为  $q \neq 0, p=0$ 。

通常使用自回归移动平均模型，即 ARMA，并依据最小信息准则分析公式（1）中的 q、p 数值，因为受到时间序列原理影响，预测点数量一般较少，以便更好地确定固定长度预测时间点数。本次研究主要使用数据集，用以验证该模型。

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \dots + \theta_p X_{t-p} - \Theta_1 \varepsilon_{t-1} - \Theta_2 \varepsilon_{t-2} - \dots - \Theta_q \varepsilon_{t-q} \quad (1)$$

②线性回归。如公式（2）所示，为某线性回归公式：

$$Y = \alpha + \beta X \quad (2)$$

在公式（2）回归系数为  $\beta$  和  $\alpha$ ，依据已经确定好的数据点，通过相关数值分析法，如最大似然法、最小二乘法可得公式（3），公式（4）：

$$\alpha = \bar{y} - \beta \bar{x} \quad (3)$$

$$\beta = \frac{\sum_{i=1}^S (X_i - \bar{x})(Y_i - \bar{y})}{\sum_{i=1}^S (X_i - \bar{x})^2} \quad (4)$$

针对已经构建的回归模型来说, 还需进行显著性检验。在回归分析算法中, 还存在一部分非线性拟合图形, 非线性拟合图具有一定优势, 但一般存在拟合现象。一般在线性回归分析算法中, 线性回归使用场景要比非线性回归场景要多。

## 4.2 选择数据挖掘工具

### 4.2.1 提取 SQL server 数据

SQL server 是一款功能全面的关系型数据库管理系统, 由 Microsoft 推出。在全球范围内, 它被誉为顶尖的数据库专家, 在企业生产中发挥着无可替代的作用。其显著优点包括使用方便、伸缩性优良、高度集成的相关软件、可移植性强、兼容多个平台以及支持集成的商业智能 (BI) 工具 [5]。

作为企业级应用的首选, SQLServer 在最新版本中添加了数据仓库和数据分析等功能, 为企业提供了强大的支持。这些新功能简化了数据的获取和整合, 并扩展了传统用户对数据的使用范围。

在此, 技术人员采用了 SQLServer 2008 版本, 成功地将 SCADA 系统的数据传入数据库中。随后, 我们对这些数据进行清洗和简单聚合, 如计算日均功率输出、日均外界温度以及轴承不同部位的温差等。为了更有效地利用这些数据, 我们设计了一套简单且实用的维度表和事实表, 将多种有效数据分类存放。

### 4.2.2 分析 Rapid Miner 数据

随着数据的爆炸式增长, 数据处理挑战日益严峻。而在这个领域, Rapid Miner 以其卓越的技术实力和丰富的实践经验独树一帜, 为用户提供了解决复杂问题的有力武器。为了满足不同行业用户的个性化需求, Rapid Miner 量身定制了各种解决方案, 包括企业级大数据管理、云端数据仓库建设、实时数据流处理以及智能决策支持等。Rapid Miner 与市场主流数据库、数据平台、云服务商紧密合作, 实现了数据的无缝导入与导出, 为用户节省了大量宝贵的时间和精力。在数据科学领域, Rapid Miner 处于领先地位, 它的出色表现不仅在于帮助用户挖掘数据中的潜在价值, 还切实考虑到实际业务需求, 提供数据可视化、预测分析和大数据平台之间的连接。通过将复杂数据处理任务分解成易于理解的小任务, 用户可以迅速把握整体流程, 实时监控执行效果, 从而更好地应对业务挑战。面对日益严峻的数据处理挑战, Rapid Miner 凭借其卓越的技术实力和丰富的实践经验, 为解决复杂问题提供了有效工具。作为一款全球领先的数据挖掘解决方案, Rapid Miner 将继续引领数据创新领域的发展, 进一步推动各行业的数据应用创新, 让数据发挥出更大的价值, 助力企业迈向成功之路。

## 4.3 设计故障诊断应用体系

风力发电机组在运行过程中可能会发生各种故障。为

了能够及时发现并处理这些故障, 实时监测其运行状态至关重要。论文介绍了一种基于数据挖掘和统计分析的风力发电机组故障诊断方法。该方法主要分为三个步骤: 统计分析、数据预处理和故障预测。

我们通过对风力发电机组的各个部件工作状态、油温和转速等参数进行初步统计分析, 初步判定其运行状态。接着, 我们进行分时统计, 观察这些指标在各个时间段的变化趋势, 以找出潜在的异常。然后, 我们进行集群分析, 将多台风力发电机组的数据进行聚类分析, 将不同的数据状态分成不同的群组, 以进一步发现异常风力发电机组 [6]。

一旦发现异常风力发电机组, 我们会对其历史数据和状态进行深入分析。首先, 我们进行数据清洗, 以消除其中的噪声和异常值。然后, 我们选择合适的特征参数和机器学习算法对历史数据进行预处理, 建立预测模型。在故障预测环节, 我们使用已经建立好的预测模型对异常时间区间的指标进行预测, 以判断是否发生了故障。如果发生故障, 我们还需要进一步确定故障发生的地点等其他相关信息。

通过这种基于数据挖掘和统计分析的风力发电机组故障诊断方法, 可以提高其运行效率和可靠性, 确保风力发电系统的稳定运行。这种方法有助于及时发现并处理风力发电机组的故障, 避免潜在的损失和风险。

## 5 结语

总之, 风力发电机组大数据驱动故障预测方案, 具有十分卓越的灵活性和创新性, 有效提升预测的准确性和实时性。在进行机械传感器参数处理时, 必须充分考虑风力发电机组的运行方式、结构以及电气参数。针对工业大数据的独特特性, 我们已经开发出一套多元化的数据清洗、预处理和建模策略。这些策略已被成功应用于风力发电机故障预测, 并已成功地完善了故障知识库。通过利用关联规则等机器学习方法对历史故障数据进行深入分析, 我们能够对故障预测工作提供强大的辅助。

## 参考文献

- [1] 张李炜, 李孝忠. 基于改进CNN和BiGRU双通道特征融合的风电机组故障诊断模型[J]. 天津科技大学学报, 2023, 38(1): 55-60.
- [2] 辛鹏, 杨凯勋, 文孝强. 基于改进SE-CNN的风电机组故障诊断方法研究[J]. 吉林化工学院学报, 2023, 40(1): 34-40.
- [3] 李洪川, 王旭东, 王东明, 等. 基于故障树的风电机组变桨系统故障诊断研究[J]. 设备管理与维修, 2022(15): 168-169.
- [4] 安文杰, 陈长征, 田淼, 等. 基于MSCNNSA-BiGRU的变工况风电机组滚动轴承故障诊断研究[J]. 机电工程, 2022, 39(8): 1096-1103.
- [5] 尹晓伟, 江雪峰, 王龙福. 风电机组轴承故障诊断与疲劳寿命研究综述[J]. 轴承, 2022(5): 1-8.
- [6] 刘家瑞, 杨国田, 王孝伟. 基于孪生深度神经网络的风电机组故障诊断方法[J]. 系统仿真学报, 2022, 34(11): 2348-2358.