

基于机器学习的风电机组发电机故障预警

胡梦浩 吴宗明 刘 闯

华北水利水电大学电气工程学院 河南郑州 450045

摘要: 风电机组发电机的故障运维占据风电场运维的大部分成本, 发电发电机运行状态的精准预测是保障风电机组健康稳定运行的基础。因此, 本文提出一种基于 WOA-ELM 的风电机组发电机故障预警方法。首先利用四分位数算法, 检测风电机组 SCADA 数据中的异常数据点, 对 SCADA 数据进行预处理; 之后构建基于 WOA-ELM 的发电机油箱温度预警模型; 利用滑动窗口法结合预测误差计算发电机的运行状态监测指标; 最后利用非参数核密度估计进行报警阈值设置, 实现发电机运行状态故障的动态预警。

关键词: 四分位法, WOA-ELM, 状态监测指标, 非参数核密度估计

引言

风力发电作为清洁能源的代表之一, 在全球范围内得到了广泛的关注和发展。根据 2022 年的世界风能报告, 全球风力发电新增装机容量在 2021 年超过 93.6 千兆瓦, 总装机容量达到了 837 千兆瓦, 相比去年增长了 14%^[1]。

由于风电机组运行环境的恶劣, 发电机作为风力发电的核心装置, 容易发生各类故障, 影响风电机组的正常运行。陆上风电机组的运行和维护成本能占到总发电收入的 15%~20%^[2-3], 因此, 风电机组的故障预警是保障风电机组安全稳定运行的关键。

风电机组发电机故障的传统方法通常依赖于设备的物理监测和人工检查, 如振动分析、温度监测和声音诊断等^[4-5], 这些方法需要在风电机组上加装新的传感器或者人工采集数据, 效率低下, 且对于早期故障的检测能力有限。相比之下, 基于监控和数据采集 (SCADA) 系统的智能检测方法, 为风电机组发电机故障的预测和诊断提供了新的方案^[6-7]。这些方法通过分析收集到的大量数据, 能够更有效地识别故障模式, 在故障发生前实现预警。当前最典型的风电机组故障预警方法主要分为两类: 统计方法和机器学习方法^[8]。

基于统计的故障预警方法, 如 Chow 检验, 概率估计等。Phong B.Dao^[9]采用 chow 检验法对 SCADA 数据结构突变检测模型中回归系统的稳定型进行评估, 通过对各关键部件系统的稳定性监测, 来对风电机组运行状态进行监测。Zeng 在文献^[10]提出了一种基于 SCADA 数据的变速箱油温异常监测方法, 根据历史数据估计给定置信水平下油温的可能变化范

围, 再利用 HT 检测温度的实际值来实现齿轮箱运行状态的异常检测。

经典机器学习方法, 如决策树、支持向量机、人工神经网络等, 苏连成等^[11]在 BP 神经网络和非线性状态估计法的基础上, 提出了一种基于组合模型的风电机组状态建模和故障识别方法。邓子豪^[12]利用特征选择 (ReliefF) 算法和核密度均值提取反应齿偏航轮箱运行工况的风电机组数据特征, 进行齿轮箱故障诊断。

本文提出一种基于 WOA-ELM 的风电机组发电机故障预警模型, 首先通过四分位法对风电机组 SCADA 数据进行清洗; 然后利用鲸鱼优化算法 (WOA) 优化极限学习机的输入权值和隐含层阈值, 提高 ELM 模型的收敛速度和预测精度; 最后根据预测误差, 采用滑动窗口模型计算发电机状态监测指标, 确定正常运行状态的阈值; 最后采用该风机的历史故障数据对本文所提出的预警模型进行验证。

1. 数据清洗

四分位算法是统计学的一种分析方法, 可根据数据分布的四分位数进行风电机组 SCADA 数据中的异常数据清洗。算法具体的计算过程如下^[13]。

将风电机组 SCADA 数据按照风速划分区间后, 对于各区间内排列好顺序的数据样本平均分为 4 部分。于是, 得到数据点下四分位数 Q_1 、中位数 Q_2 和上四分位数 Q_3 。

通过上下四分位数可以得到该数据分布的四分位距, 具体的计算方法如式 (1)

$$I_{QR} = Q_3 - Q_1 \quad (1)$$

本文取 1.5 倍的四分位距作为正常值的范围，当数据超出这个界限我将其作为异常值处理，界限具体计算如式 (2) 所示。

$$[Q_L, Q_U] = [Q_1 - 1.5 * I_{QR}, Q_3 + 1.5 * I_{QR}] \quad (2)$$

2.WOA-LSTM 模型

2.1 鲸鱼优化算法

鲸鱼优化算法 (WOA) 具有良好的优化效果，该方法主要由包围猎物、螺旋更新，收缩包围和随机搜索三个部分构成 [14]。

利用鲸鱼算法优化极限学习机的输入权值和隐含层阈值，提高极限学习机的收敛速度和预测精度，其计算步骤如下：

$$D = |C \cdot X^*(t) - X_i(t)| \quad (3)$$

$$X_i(t+1) = X^*(t) - A \cdot D \quad (4)$$

其中 D 为当前位置与种群全局最优位置的距离， $X_i(t)$ 为第 i 只鲸鱼在第 t 次迭代时的位置， $X^*(t)$ 为第 t 次迭代时鲸鱼种群的全局最优位置。A 为收敛因子，C 为摆动因子，可根据式 (3) 和式 (4) 对鲸鱼位置的进行迭代更新，从而找到最优参数。

2.2 极限学习机

极限学习机 (ELM) 模型采用随机的输入层权值和偏差，对于输出层权重则通过广义逆矩阵理论计算得到 [15]。极限学习机的网络结构如图 1 所示：

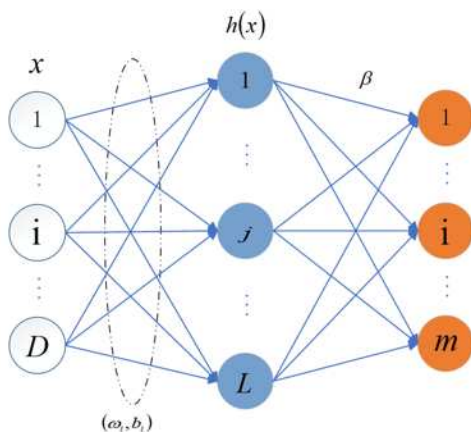


图 1 极限学习机网络拓补图

隐藏层的输出是乘上对应的权重加上偏差，再通过一个非线性函数的所有节点结果求和结果得到。计算公式如下：

$$h_i(x) = g(\omega_i, b_i, x) = g(\omega_i x + b_i) \quad (5)$$

$$\omega_i \in R^D, b_i \in R$$

式 (5) 中，g 是 Sigmoid 激活函数， ω_i 为模型隐含层权值，x 为输入， b_i 为隐含层偏置项。

ELM 算法的预测精度与其输入权值 W 和隐含层阈值 b 息息相关；为此，本文利用 WOA 算法优化 ELM 算法的输入权值 W 和隐含层阈值 b，以 ELM 算法最优的输入权值 W 和隐含层阈值 b 进行建模预测。

3. 预警机制

3.1 滑动窗口

本文引入滑动窗口通过窗口内误差数据的平均来对建立状态监测指标，本文所设置的滑动窗口宽度 h 为 10 分钟，移动时间增量 q 为 1 分钟。

3.2 状态监测指标及阈值计算

(1) 状态监测指标

本文通过分析每个滑动窗口中模型预测误差的分布特性，采用式 (6) 计算风电机组发电机的状态检测指标。

$$V(t, t) = \sqrt{\sum_{i=t-h}^{t_q} (T_{true} - T_{pre})^2 / \sum_{i=t-h}^{t_q} V^2} \quad (6)$$

式中， T_{true} 、 T_{pre} 为窗口输出参数的实际值和预测值。

(2) 阈值设置

本文采用非参数核密度估计对正常运行状态下的发电机油箱温度预测误差进行分布特性分析，确定置信水平 $1-\alpha$ 下的置信区间上界，作为发电机油温的报警阈值，当状态监测指标超过阈值时进行系统报警；反之，认为该机组正常运行。

核密度估计法的原理如式 (7) 所示：

$$\begin{cases} f_{nk}(V) = \frac{1}{nh_k} \sum_{i=1}^n k(\frac{V-V_i}{h_k}) & i=1,2,\dots,n \\ K(V) > 0 & \int_{-\infty}^{\infty} K(V)dV=1 \end{cases} \quad (7)$$

式中： V_i 为状态监测指标识别结果，为在状态监测指标 V 的总体密度函数 f(V) 的核密度估计； $K(\cdot)$ 为核函数， h_k 为核密度估计的宽度。

阈值的计算方法如式 (8) 所示：

$$\begin{cases} F(v) = \int_{-\infty}^v f_{nk}(v)dv \\ F(v_{upper}) = 1 - \alpha \end{cases} \quad (8)$$

式中： $F(v)$ 为概率分布函数， v_{upper} 为置信度为 $1-\alpha$ 的

置信上限，也即状态监测指标的报警阈值。

4. 实例验证

本文所使用的数据来自中国中部某省的风电场站，装机容量为 1.5MW，采集的 SCADA 数据的时间分辨率为 1min，每日具有 1440 个数据点，特征包括发电机油箱温度，齿轮箱高速轴承温度，齿轮箱低速轴温度，齿轮箱油温，U1 绕组温度，发电机低速轴温度，发电机转速，转子转速，发电机 U3、U2、U1 绕组温度。其中以发电机油箱温度作为预警目标变量，数据集为 2016 年 11 月 1 日 0 时至 2017 年 10 月 31 日 24 时的完整年数据，

4.1 数据预处理技术

风电机组因为运行环境恶劣，经常会出现数据错误的问题，此类异常的数据会造成模型偏差较大，预测精度下降。因此，对风电场原始 SCADA 数据的预处理是保障预测模型效果的关键。

(1) 风电数据筛选和异常缺失值处理

根据风机报告可知，该风电场风机的切入切出风速分别为 3m/s 和 20m/s，该类停机数据会影响模型的精度，因此将风速低于 3m/s 和高于 20m/s 的数据去除。若风电机组 SCADA 数据存在缺失和异常值，则对其使用前值进行填充替换。

(2) 数据标准化

为了减少模型训练时的学习误差，提高学习速度，在数据输入模型前需要对它进行归一化处理，统一不同特征数据间的量纲。本文标准化采用了法，其标准化计算公式为式 (9)：

$$x_{norm} = \frac{x - \min}{\max - \min} \quad (9)$$

其中 x_{norm} 为数据归一化后的值；max 表示数据中最大值；min 表示最小值。

4.2 发电机油温预测分析

本文选择使用发电机油箱温度作为发电机故障预警的目标预测变量，其余变量作为模型输入，本文选择了 WOA-ELM, PSO-ELM, PSO-LSSVM, WOA-LSSVM 模型 对于 2016 年 2 月 25 日发电机油温分别进行了预测，预测效果如下所示：

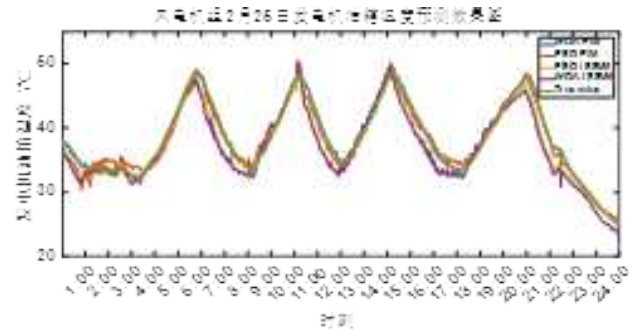


图 2 发电机 2 月 25 油箱温度预测图

图中，蓝色、红色、黄色和紫色曲线分别为 WOA-ELM 模型、PSO-ELM 模型、PSO-LSSVM 模型和 WOA-LSSVM 模型预测结果，绿色曲线为真实温度值曲线，表 1 是 4 种模型在以上两日预测中的预测结果。

表 1 多模型发电机油箱温度预测误差对比分析

日期	评价指标	WOA-ELM	PSO-ELM	PSO-LSSVM	WOA-LSSVM
2/25	MAE	0.1713	0.7311	1.3359	1.5669
	RMSE	0.2226	0.9192	1.7166	2.0068
	MAPE	0.4563	1.9501	3.4179	4.0738
	R2	0.9992	0.9864	0.9732	0.9627

由表 1 可知，WOA-ELM 模型在 2 月 25 日发电机油温预测效果均优于其他模型，可知，WOA-ELM 模型具有预测精度高和泛化能力强的优点。

4.3 发电机故障预警

本文根据对该机组在 2 月份正常运行状态下的发电机油温预测误差进行分析，依据非参数核密度估计结果设置 95% 置信度下的置信水平上限为报警阈值，对该机组 2 月 18 日的发电机油温温度的状态检测指标进行预警，预警结果如图所示：

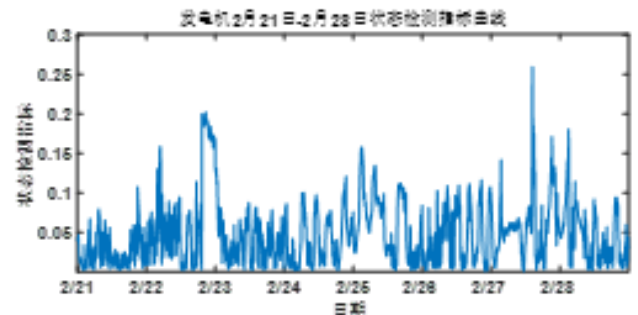


图 3 发电机 2 月 21 日 -2 月 28 日状态监测指标曲线

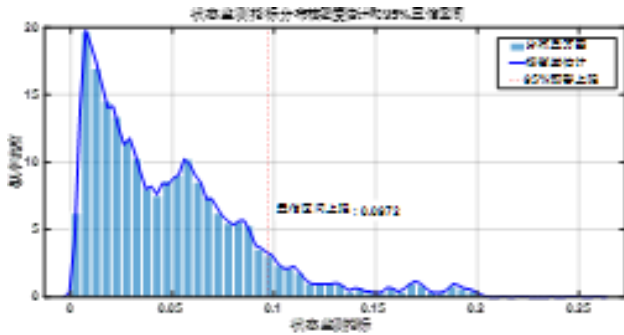


图 4 状态监测指标曲线分布核密度估计和 95% 置信区间上限

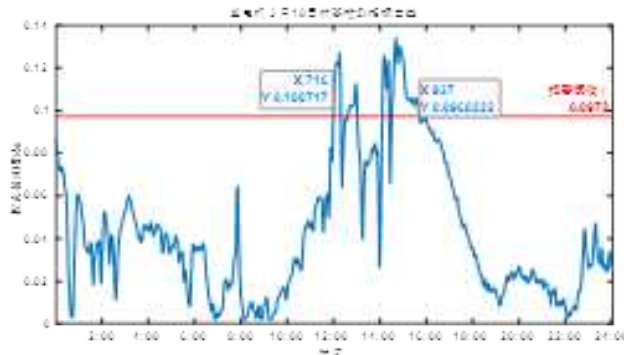


图 5 发电机 3 月 18 日状态监测指标预警曲线

图 3 所示为发电机 2 月 21 日 -2 月 28 日状态监测指标曲线图, 根据该机组运行报告显示, 该机组在此时间段内并未发生报警, 认为该时间段为正常运行状态, 利用该时间段的状态监测指标确定其概率分布如图 4 所示, 利用非参数核密度估计出其概率分布函数, 并计算出 95% 置信水平下的置信上限作为该机组发电机油温的预警阈值, 进而对该机组 3 月 18 日的状态监测指标进行预警。

图 5 反映了此种方法的预警结果曲线, 可以发现, 该方法在第 716 个数据点 (11:56) 开始报警, 在第 937 个数据点 (15:37) 结束报警, 运行报告中指出该机组的发电机油温过高报警时间为 14:21-18:26, 可见此种方法可以提前 2 个小时 25 分钟进行报警, 能够避免发电机故障的进一步加剧, 对发电机的早期维修, 减少风电机组运维成本有着重要意义。

5 结论

(1) 通过四分位法以风电机组 SCADA 数据的风速 - 功率之间的关系对该机组 SCADA 数据进行数据清洗, 替换了异常值, 提高了数据可利用率;

(2) 利用鲸鱼优化算法 (WOA) 优化极限学习机的输入权值和隐含层阈值, 可以提高 ELM 模型的收敛速度和预

测精度; 对比多个预测模型, 结果表明本文提出的模型具有优越性。

(3) 根据预测误差, 计算发电机状态监测指标, 确定报警阈值, 最后采用该风机的历史故障数据进行验证, 可以原始系统提前报警 2 个小时 25 分钟, 证明了本文所提预警机制的有效性。

参考文献

- [1] 姚钢, 杨浩猛, 周荔丹等. 大容量海上风电机组发展现状及关键技术 [J]. 电力系统自动化, 2021,45(21):33-47.
- [2] Zhang Guangyao, Li Yanting, et al. A novel fault diagnosis method for wind turbine based on adaptive multivariate time-series convolutional network using SCADA data[J]. Advanced Engineering Informatics, Volume 57, 2023, 102031
- [3] Liu Dongdong, Cui Lingli, et al. Fault diagnosis of wind turbines under nonstationary conditions based on a novel tachometerless generalized demodulation[J]. Renewable Energy, Volume 206, 2023, Pages 645-657
- [4] Ali Dibaj, Zhen Gao, et al. Fault detection of offshore wind turbine drivetrains in different environmental conditions through optimal selection of vibration measurements[J]. Renewable Energy, Volume 203, 2023, Pages 161-176
- [5] Omneya Attallah, Rania A. Ibrahim, et al. CAD system for inter-turn fault diagnosis of offshore wind turbines via multi-CNNs & feature selection[J]. Renewable Energy, Volume 203, 2023, Pages 870-880
- [6] Prince Waqas Khan, Chan Yeob Yeun, et al. Fault detection of wind turbines using SCADA data and genetic algorithm-based ensemble learning[J]. Engineering Failure Analysis, Volume 148, 2023, 107209
- [7] Xiang Ling, Xin Yang Xin, et al. Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks[J]. Applied Energy, Volume 305, 2022, 117925
- [8] Xu Xinhua, Huang Xinyu, et al. Total Process of Fault Diagnosis for Wind Turbine Gearbox, from the Perspective of Combination with Feature Extraction and Machine Learning: A Review[J]. Energy and AI, 2023, 100318
- [9] Phong B. Dao, Condition monitoring and fault diagnosis

of wind turbines based on structural break detection in SCADA data[J]. Renewable Energy, Volume 185, 2022, Pages 641–654

[10] Zeng X.J, Yang M., et al. Gearbox oil temperature anomaly detection for wind turbine based on sparse Bayesian probability estimation[J]. International Journal of Electrical Power & Energy Systems, Volume 123, 2020, 106233

[11] 苏连成, 邢美玲, 张慧. 基于组合预测模型的风电机组关键部位故障检测 [J]. 太阳能学报, 2021, 42(10): 220–225.

[12] 邓子豪, 李录平, 刘瑞等. 基于 SCADA 数据特征提取的风电机组偏航齿轮箱故障诊断方法研究 [J]. 动力工程学报, 2021, 41(01): 43–50.

[13] Han Shuang, Qiao Yanhui, et al. Wind turbine power curve modeling based on interval extreme probability density for

the integration of renewable energies and electric vehicles[J]. Renewable Energy, Volume 157, 2020, Pages 190–203.

[14] 慎慧强. 基于混合预测模型的风光短期功率预测及不确定性分析 [D]. 华北水利水电学, 2022. DOI: 10.27144/d.cnki.ghbsc.2022.000172.

[15] Meng Anbo, Zhu Zibin, et al. A novel wind power prediction approach using multivariate variational mode decomposition and multi-objective crisscross optimization based deep extreme learning machine[J]. Energy, Volume 260, 2022, 124957.

作者简介:

胡梦浩(1998.5–), 男, 汉族, 河南洛阳人, 硕士研究生, 研究方向: 风电机组运行状态分析。