



基于AC-GAN数据重构的风电机组主轴承温度监测方法

尹诗, 侯国莲, 胡晓东, 周继威

引用本文:

尹诗, 侯国莲, 胡晓东, 等. 基于AC-GAN数据重构的风电机组主轴承温度监测方法[J]. *智能系统学报*, 2021, 16(6): 1106–1116.

YIN Shi, HOU Guolian, HU Xiaodong, et al. Temperature monitoring method of the main bearing of wind turbine based on AC-GAN data reconstruction[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(6): 1106–1116.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202009020>

您可能感兴趣的其他文章

一种改进的深度学习的路标识别算法

An improved deep learning algorithm for road traffic identification

智能系统学报. 2020, 15(6): 1121–1130 <https://dx.doi.org/10.11992/tis.201811009>

基于增强AlexNet的音乐流派识别研究

Music genre recognition research based on enhanced AlexNet

智能系统学报. 2020, 15(4): 750–757 <https://dx.doi.org/10.11992/tis.201909032>

基于生成式对抗网络的道路交通模糊图像增强

Enhancement of blurred road-traffic images based on generative adversarial network

智能系统学报. 2020, 15(3): 491–498 <https://dx.doi.org/10.11992/tis.201903041>

生成对抗网络辅助学习的舰船目标精细识别

Fine-grained inshore ship recognition assisted by deep-learning generative adversarial networks

智能系统学报. 2020, 15(2): 296–301 <https://dx.doi.org/10.11992/tis.201901004>

基于生成对抗网络的机载遥感图像超分辨率重建

Super-resolution reconstruction of airborne remote sensing images based on the generative adversarial networks

智能系统学报. 2020, 15(1): 74–83 <https://dx.doi.org/10.11992/tis.202002002>

基于T-S型模糊加权的多模软切换的风电机组变桨控制

Multi-mode soft-switch variable-pitch control of wind turbines based on T-S fuzzy weighting

智能系统学报. 2018, 13(4): 625–632 <https://dx.doi.org/10.11992/tis.201710009>

微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.202009020

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.TP.20210831.1251.004.html>

基于 AC-GAN 数据重构的风电机组 主轴承温度监测方法

尹诗^{1,2}, 侯国莲², 胡晓东², 周继威¹

(1. 中能电力科技开发有限公司, 北京 100034; 2. 华北电力大学 控制与计算机工程学院, 北京 102206)

摘要: 为更好地识别风电机组主轴承运行状态, 提出了一种基于辅助分类生成对抗网络 (auxiliary classifier generative adversarial networks, AC-GAN) 的数据重构算法对风电机组主轴承温度进行监测。首先, 利用采集与监视控制系统 (supervisory control and data acquisition, SCADA) 时序数据建立基于轻型梯度增强学习器 (light gradient boosting machine, LightGBM) 的主轴承温度预测模型, 并计算其残差特征。其次, 利用统计过程控制 (statistical process control, SPC) 方法对主轴承温度异常残差在控制线范围内进行筛选, 并利用 AC-GAN 算法对残差进行重构。最后, 分别提取主轴承温度正常和异常的残差特征, 建立基于自然梯度提升 (natural gradient boosting, NGBoost) 的主轴承状态监测模型。实验结果表明, 该方法对主轴承运行状态判断准确度高达 87.5%, 能够有效地监测风电机组轴承类运行状态。

关键词: 轻型梯度增强学习器; 辅助分类生成对抗网络; 自然梯度提升; 风电机组; 主轴承; 状态监测; 数据重构; 温度残差

中图分类号: TP8; TK83 文献标志码: A 文章编号: 1673-4785(2021)06-1106-11

中文引用格式: 尹诗, 侯国莲, 胡晓东, 等. 基于 AC-GAN 数据重构的风电机组主轴承温度监测方法 [J]. 智能系统学报, 2021, 16(6): 1106-1116.

英文引用格式: YIN Shi, HOU Guolian, HU Xiaodong, et al. Temperature monitoring method of the main bearing of wind turbine based on AC-GAN data reconstruction[J]. CAAI transactions on intelligent systems, 2021, 16(6): 1106-1116.

Temperature monitoring method of the main bearing of wind turbine based on AC-GAN data reconstruction

YIN Shi^{1,2}, HOU Guolian², HU Xiaodong², ZHOU Jiwei¹

(1. ZhongNeng Power-Tech Development Co., LTD, Beijing 100034, China; 2. College of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China)

Abstract: To better identify the operating status of the main bearing of wind turbines, a data reconstruction algorithm based on auxiliary classifier generative adversarial networks (AC-GAN) is proposed to monitor the temperature of the main bearing of the wind turbine. First, this work uses the time series data of the supervisory control and data acquisition system to establish the wind turbine's main bearing temperature prediction model based on the light gradient boosting machine and calculates its residual characteristics. Second, the statistical process control (SPC) method is used to screen abnormal temperature residuals of the main bearing within the control line, and the AC-GAN algorithm is used to reconstruct the residual sequence. Finally, normal and abnormal temperature residual characteristics of the main bearing are extracted, and the main bearing status monitoring model based on the natural gradient boosting algorithm is established. Experimental results show that the accuracy of the method for judging the operating state of the main bearing is as high as 87.5%, for which the algorithm can effectively monitor the running state of wind turbine bearings.

Keywords: light gradient boosting machine; auxiliary classifier generative adversarial networks; natural gradient boosting; wind turbines; main bearing; condition monitoring; data reconstruction; temperature residual

电机组主轴承连接着轮毂与齿轮箱, 作为重要的机械传动部件之一, 其可靠性要求较高, 但主轴承内部结构和受力较复杂, 且常常运行在重负荷、强冲击的工作状态下, 容易发生磨损、不对中、不平衡等问题^[1]。风电机组主轴承一旦损坏, 受限於维修过程的复杂, 其维修费用高、周期长, 严重影响风电场的经济效益。

风电机组故障诊断研究目前主要集中在振动信号分析方法、SCADA 数据分析方法、视频图像检测方法、润滑油检测方法、声发射信号检测方法、应变传感信号检测方法等^[2]。数据挖掘方法能够在大量数据中发现隐含的知识或潜在规律, 因此近年来在各行业中具有广泛的研究和应用。随着数据挖掘算法的不断发展, 一些算法逐渐被应用到故障预警和故障辨识中^[3]。目前基于数据挖掘的风电机组的状态监测研究主要基于采集与监视控制系统 (supervisory control and data acquisition, SCADA) 时序数据, 利用相关智能学习算法建立设备部件的正常运行模型, 通过分析正常模型预测值与实际观测值之间的残差进行状态监测。文献 [4] 利用深度置信网络建立发电机同步定子故障预警模型, 对残差设定故障阈值进行状态监测。文献 [5] 利用 BOX-COX 变换和相对熵对残差进行分析, 对齿轮箱进行状态监测。文献 [6] 利用数据分类重建和提取衰退指标的方法对齿轮箱进行状态监测。文献 [7] 将 DS 证据理论应用于 SCADA 警报分析对风电机组进行故障诊断。文献 [8] 通过提取风电机组 SCADA 系统中的实际运行数据, 采用双向递归神经网络建立风电机组运行预测模型, 根据滑动窗口的实际值与实测值之间的残差, 利用莱特准则实现故障预警。

以上方法对残差进行分析时需要人为设定故障预警阈值, 所述方法适用于某一特定风电场, 其泛化性有待提升。文献 [9] 利用 SCADA 数据提出了基于工况辨识的 Bi-RNN 神经网络, 建立预警模型对风电机组主轴承运行状态进行监测, 该方法在故障决策方面引入随机森林算法避免人为设定故障阈值, 但由于故障发生前 SCADA 数据不全是表征故障的数据, 所建立的状态决策模型精度有待提升。

上述研究在不同程度上对风电机组关键核心部件的状态监测和故障预警起到了推动作用, 但是普遍存在泛化性弱, 故障决策受主观因素影响、缺乏理论支撑等问题, 限制了状态监测模型的工程实用性。

因此, 本文以风电机组主轴承温度为研究对

象, 提出了一种基于辅助分类生成对抗网络 (auxiliary classifier generative adversarial networks, AC-GAN) 数据重构的风电机组主轴承状态监测方法对其运行状态进行监测。

1 建模方法设计

温度是风电机组运行数据中较为重要的观测指标, 具有很强的抗干扰性, 不会轻易因环境或工况变化产生剧烈跳跃变化。正常情况下, 轴承温度随着轴承开始运行缓慢上升, 后续达到稳定运行状态。风电机组主轴承温度随着热容量、散热速度、转速和负载而发生变化。但机械传动设备在运行过程中产生的磨损、润滑不良、屏蔽不良等问题往往会导致温度数据异常, 如果主轴承长期在高温下运行, 其运行寿命将会大大缩短, 甚至会引起更为严重的故障事故^[10]。因此, 本文重点分析风电机组主轴承温度参数的变化, 监测并实时掌握主轴承运行状态, 发现其潜在隐患。风电机组主轴承温度预测模型的准确度、泛化能力以及状态决策模型的准确度决定着主轴承状态监测的精准度。建模流程如图 1 所示。

首先, 利用 SCADA 时序数据建立基于轻型梯度增强学习器 (light gradient boosting machine, LightGBM) 的主轴承温度预测模型, 定义模型预测输出温度与实际观测温度之差为残差。相较于其他算法, LightGBM 算法无需通过计算所有样本信息增益, 其内置的特征降维技术具有较高的预测精度和较快的训练速度, 比较适合于工程实现。

其次, 通过滑动窗口和统计过程控制 (statistical process control, SPC) 方法将异常主轴承残差中的正常残差和异常残差进行有效区分。利用 AC-GAN 辅助分类生成对抗网络生成与主轴承异常残差分布相似的残差数据集, 用来替换异常主轴承残差分布中的正常残差数据集。由于风电机组异常主轴承的残差特征不全表征为异常状态, 无法统一进行标记。因此, 利用 AC-GAN 将温度残差特征进行数据重构, 得到异常主轴承下的残差特征, 从根本上解决了异常样本数据的标记问题, 进而提高了后续主轴承状态决策模型的预测精度。

最后, 建立基于自然梯度提升 (natural gradient boosting, NGBoost) 的状态决策模型对风电机组主轴承状态进行判断。NGBoost 算法利用自然梯度进行概率预测, 解决了传统状态决策方法中采用单一固定阈值或人为主观设定阈值进行风电机组运行状态监测的问题, 提高了状态决策模型的预测精度和泛化性。

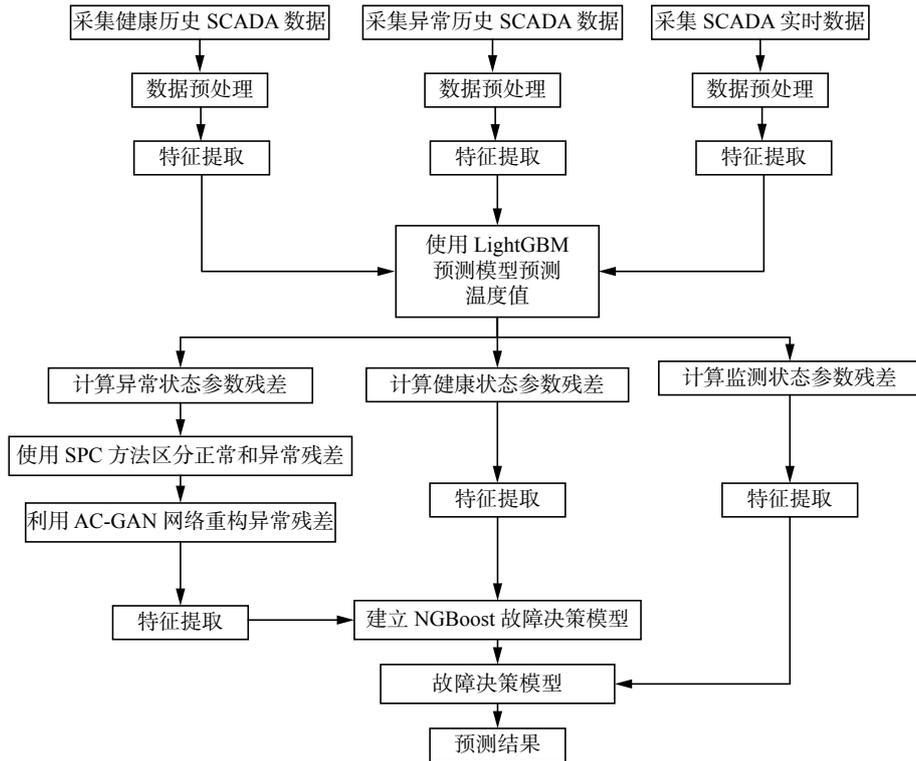


图 1 建模方法流程

Fig. 1 Flow chart of the modeling method

2 数据预处理

2.1 SCADA 数据说明

本文所采用的 SCADA 数据是河北某风电场 1.5 MW 双馈式风力发电机组的运行数据。该风电场的风电机组切入风速为 3 m/s, 切出数据为 25 m/s。SCADA 数据每 10 min 记录一条, 为了消除偶发性的故障对主轴温度预测模型的影响, 本文共选取该风电场 1.5 MW 机组共 36 台机组, 包括 23 台主轴承正常机组, 13 台主轴承异常机组。这 36 台风电机组生产厂家和型号相同, 并且都是同一个风电场的风电机组, 因此所利用的风资源和地理环境相似, 能够表征该风电场的所有运行工况。SCADA 时序数据包括时间、风速、风向、有功功率、发电机转速、叶轮转速、偏航角度、环境温度、齿轮箱油温等百余个有效观测数据。表 1 为部分 SCADA 有效数据。

2.2 数据清洗

由于风电机组本身和 SCADA 系统在运行过程中掉电、传感器损坏、系统宕机、通信设备故障等因素造成 SCADA 数据中夹杂着很多异常噪声数据。在建模前首先对数据进行清洗, 其过程如下:

- 1) 采用分区间方法按风速 0.5 m/s 划分子工况区间;
- 2) 将小于切入风速、大于切出风速、有功功

率小于或等于 0 的数据剔除;

表 1 SCADA 数据示例
Table 1 Examples of SCADA data

采样时刻	风速/(m·s ⁻¹)	风向/(°)	发电机转速/(r·min ⁻¹)	有功功率/kW
2019/3/1 20:38	3.47	1.81	1130.52	240.26
2019/3/1 20:48	3.09	-3.91	1132.56	215.71
2019/9/2 10:28	4.89	-1.51	1424.43	420.00
2019/9/2 10:38	4.85	-1.45	1461.12	450.56

3) 采用统计学中的四分位原理^[11]对每个子工况区间的 SCADA 数据进行清洗。

为说明去除异常噪声数据的效果, 本文对 2.1 节提到的风电机组历史 SCADA 数据进行了异常噪声数据去除, 如图 2 所示。图 2(a) 为该风电机组历史 SCADA 数据过滤前的风速功率曲线图, 图 2(b) 为过滤后的风速功率曲线图。

2.3 特征数据提取

风电机组 SCADA 数据中并非所有的时序数据均与风电机组主轴承温度相关, 为提高风电机组主轴承温度预测模型精度, 同时降低模型训练时长, 选取与风电机组主轴承运行状态相关的特征子集。传统的特征筛选方式为利用皮尔森相关系数或根据工程师相关经验进行确定, 皮尔森相关系数对数据的要求必须服从正态分布, 但风电

机组由于弃风、限电等运行工况的变化导致 SCADA 数据并不符合正态分布, 且皮尔森相关系数受到数据异常值的影响较大, 仅适用于某些特定场合下的风电场或风电机组。鉴于此, 本文选取相关系数收敛快、可解释性好且对数据分布没有特殊要求的斯皮尔曼相关性分析方法提取直接或者间接反映风电机组主轴承温度特征的特征集^[12], 斯皮尔曼相关系数的计算步骤为:

1) SCADA 数据特征中主轴承温度定义为 Y , 其他特征定义为 X_i , 将 X_i 和 Y 列所对应的数据转换为各自列向量的排名, 记为 $R(X_i)$ 和 $R(Y)$ 。

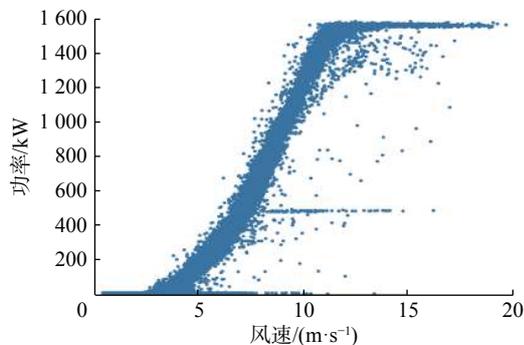
2) 两个列向量中对应数据 $R(X_i)$ 和 $R(Y)$ 之间的差异 d 为

$$d = \sum_{i=1}^N |R(X_i) - R(Y)|^2$$

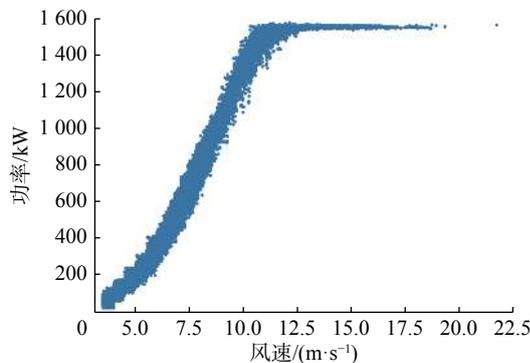
两个列向量之间的相关性 R_s 为

$$R_s = 1 - \frac{6 \times d}{N(N^2 - 1)}$$

式中: i 为每一列 SCADA 特征数据; N 为 SCADA 数据特征的长度。斯皮尔曼系数高于 0.5 时特征之间的相关性为强相关, 因此, 通过斯皮尔曼相关性系数分析, 得到 SCADA 数据中与主轴承温度相关性较高的特征, 见表 2 所示。因此, 选取发电机转速、叶轮转速、机舱温度等 10 个特征。



(a) 数据过滤前风速功率曲线



(b) 数据过滤后风速功率曲线

图 2 SCADA 数据过滤前后风功率对比图

Fig. 2 Contrast chart of the wind-power before and after SCADA data filtering

表 2 斯皮尔曼相关性系数

Table 2 Spearman correlation coefficient

选取的特征	斯皮尔曼相关性系数
发电机转速	0.880
叶轮转速	0.848
机舱温度	0.823
有功功率	0.808
环境温度	0.804
风速	0.780
风向	0.694
桨叶角度	0.648
A相电流	0.532
偏航角度	0.514

提取的特征数据中往往具有不同的量纲和量纲单位, 为了消除特征数据之间的量纲影响, 将数据归一化处理, 计算公式为

$$x_n = \frac{x - x_{\text{mean}}}{x_{\text{max}} - x_{\text{min}}}$$

式中: x 为每一个 SCADA 特征数据; x_{mean} 为特征数据的均值; x_{max} 为特征数据的最大值; x_{min} 为特征数据的最小值; x_n 为归一化以后的特征数据。

3 基于 LightGBM 的风电机组主轴承温度预测

3.1 LightGBM 算法

极限梯度提升 (eXtreme gradient boosting, XGBoost) 算法是在自适应增强 (adaptive boosting, adaboost) 算法和梯度提升迭代决策树 (gradient boosting decision tree, GBDT) 算法基础上优化形成的算法^[13], 具有良好的预测精度和分类准确率, 在众多领域得到了广泛的应用。但 XGBoost 算法普遍存在训练耗时长、内存占比大等缺点。针对这些缺点, Ke 等^[14]做了相应的改进, 并在 2017 年提出了 LightGBM 算法。LightGBM 算法无需计算所有样本信息增益, 具有训练效率高、低内存、高准确率、并行化学习等优势, 较适合于工程实践。

LightGBM 使用直方图算法替换了 GBDT 的预排序, 能够在不损害准确率的前提下加快 GBDT 模型的训练速度^[15]。为保证结果准确性, 算法使用梯度单边采样技术过滤大部分小梯度数据, 在计算信息增益的时候只利用具有高梯度的数据信息; 为大幅度减少占用内存, 采用独立特征合并技术实现互斥特征的捆绑, 减少样本特征数据^[16]。

假设训练一个具有 T 棵树的 LightGBM 模型, 给定数据集为 $D = \{(x_i, y_i) | i = 1, 2, \dots, n, x_i \in \mathbf{R}^m, y_i \in \mathbf{R}\}$, 其中给定的数据集共有 n 个样本, 每个样本 x_i 对应 m 个特征和一个标签值 y_i 。LightGBM

算法在迭代过程中,假设在前一轮迭代中得到的强学习器是 $f_{t-1}(\mathbf{x})$, 损失函数为 $L(\mathbf{y}, f_{t-1}(\mathbf{x}))$, 为了让本轮迭代的损失函数最小, 本轮迭代的目的是找到分类回归树模型的弱学习器 $h_t(\mathbf{x})$, 如式 (1) 所示:

$$L(\mathbf{y}, f_t(\mathbf{x})) = L(\mathbf{y}, f_{t-1}(\mathbf{x}) + h_t(\mathbf{x})) \quad (1)$$

利用损失函数负梯度拟合本轮损失函数近似值, 从而拟合一个树模型。第 t 轮的 第 i 个样本的损失函数 $L(y_i, f_t(x_i))$ 的负梯度 r_{ti} 为

$$r_{ti} = - \left. \frac{\partial L(y_i, f_{t-1}(x_i))}{\partial f(x_i)} \right|_{f(x)=f_{t-1}(x)}$$

利用 (\mathbf{x}_i, r_{ti}) 拟合一个 CART 回归树, 进而得到 t 棵回归树所对应的叶子节点的范围为 \mathbf{R}_{ti} , $i = 1, 2, \dots, J$ 。其中 J 为叶子节点的样本, 当损失函数最小时拟合叶子节点输出值 c_{tj} 为

$$c_{tj} = \underset{c}{\operatorname{arg\,min}} \sum_{\mathbf{x}_i \in \mathbf{R}_{tj}} L(y_i, f_{t-1}(\mathbf{x}_i) + c)$$

式中: c 是损失函数最小化时的常数值; $\mathbf{x}_i \in \mathbf{R}_{tj}$ 表示样本 \mathbf{x}_i 属于第 t 棵树下的第 j 个叶子节点。本轮的决策树拟合函数为

$$h_t(\mathbf{x}) = \sum_{j=1}^J c_{tj} I(\mathbf{x}_i \in \mathbf{R}_{tj})$$

式中: $I(\mathbf{x}_i \in \mathbf{R}_{tj})$ 是指示函数, 当 $\mathbf{x}_i \in \mathbf{R}_{tj}$ 时, 指示函数的值为 1, 反之为 0^[17]。进而本轮最终得到的强学习器的表达式为

$$f_t(\mathbf{x}) = f_{t-1}(\mathbf{x}) + \sum_{j=1}^J c_{tj} I, \quad \mathbf{x}_i \in \mathbf{R}_{tj}$$

3.2 温度预测模型建立

选用 10 台主轴承正常机组的历史 SCADA 数据共 38955 组, 按上述方法进行预处理后得到共 25946 组数据。将数据集 80% 作为训练集, 数据集 20% 作为测试集。本文所有试验运行环境均为: 操作系统为 Windows10、python 版本为 3.7.1、集成开发运行环境为 anaconda3, LightGBM 算法、XGBoost 算法和随机森林算法调用 sklearn 的 API。后续使用的 AC-GAN 生成对抗神经网络调用 kears 深度学习框架 API, NGBoost 算法调用斯坦福的 NGBoost 框架。

对比分析 LightGBM 算法、XGBoost 算法、CatBoost(categorical boosting) 算法在风电机组训练主轴承温度预测模型的精度, 如表 3 所示。

采用模型训练时间、均方根误差 RMSE 和决定系数 r^2 指标对建模精度进行评价, 计算公式为

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

$$r^2 = 1 - \frac{\left[\sum_{i=1}^n (y_i - \hat{y}_i)^2 \right]}{\left[\sum_{i=1}^n (y_i - \bar{y})^2 \right]}$$

式中: y_i 为第 i 个主轴承温度的真实测量值; \hat{y}_i 为第 i 个主轴承温度的预测值; \bar{y} 为主轴承温度的真实测量值的均值。LightGBM 算法在均方根误差 RMSE、决定系数 r^2 指标和训练时间上均优于 XGBoost 算法和 CatBoost 算法。表 4 为 LightGBM、XGBoost 和 CatBoost 这 3 种算法在测试集样本中残差特征对比。LightGBM 算法在测试集上的残差最大值为 0.129, 残差均值为 0.022, 基于 LightGBM 的风电机组主轴承温度预测在测试集上具有较高的预测精度。

表 3 LightGBM、XGBoost 和 CatBoost 建模性能比较
Table 3 LightGBM, XGBoost, and CatBoost modeling performance comparison

建模方法	RMSE	r^2	时间/s
XGBoost	0.021	0.978	4.9
LightGBM	0.012	0.992	1.6
CatBoost	0.035	0.954	5.2

表 4 3 种算法测试集残差特征对比

Table 4 Comparison of residual characteristics of three algorithms in test data

建模方法	残差最大值	残差平均值
XGBoost	0.209	0.035
LightGBM	0.129	0.022
CatBoost	0.186	0.030

基于 LightGBM 算法的主轴承温度预测模型在测试集上的残差见图 3 所示。

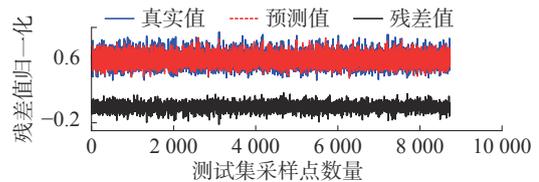


图 3 主轴承温度模型测试集残差

Fig. 3 Residual error of the main bearing temperature model in the test data

4 基于 AC-GAN 的主轴承温度残差重构

基于 AC-GAN 的主轴承温度残差重构方法具体步骤为: 首先, 采用 SPC 方法将主轴承异常机组残差在控制范围内的正常残差数据剔除; 其次, 将控制范围之外的异常残差数据作为训练数据, 采用 AC-GAN 生成对抗网络生成与真实数据分布相似的数据替换被剔除的数据。

4.1 基于 SPC 的残差特征提取

SPC 方法最初主要用来监测生产产品中的质量问题, 如果生产过程中出现随机质量问题说明

此过程处于统计过程当中, 如果出现故障问题说明此过程处于失控状态^[18]。

假设某产品在生产过程中的质量特征没有显著的非正态因素, 而是由很多随机原因导致, 则认为该质量特性服从正态分布^[19]。其质量特征 X 服从均值为 μ , 标准差为 σ 的正态分布, 则概率密度分布函数为

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$P(\mu - 3\sigma \leq x \leq \mu + 3\sigma) = \int_{\mu-3\sigma}^{\mu+3\sigma} f(x)dx = 0.9973$$

质量特征在 $[\mu - 3\sigma, \mu + 3\sigma]$ 的概率为 0.9973, 则认为在该范围内的数据是可控的, 范围之外为不可控。因此, 本文主轴异常机组温度残差中设定 $[\mu - 3\sigma, \mu + 3\sigma]$ 的范围, 范围内的残差为正常残差, 范围之外为异常残差, 其中 μ 和 σ 为主轴正常机组测试集温度残差的均值和标准差。利用 SPC 方法可以提取主轴异常机组残差中的特征, 将残差中正常残差进行剔除。

由文献 [20] 研究得知, 轴承类故障在发生故障前一个月会有明显的劣化趋势, 因此选取风电机组主轴故障发生前一个月的 SCADA 数据进行试验。图 4 为某机组主轴故障发生前一个月的温度残差, 红线为 $\mu - 3\sigma$ 下控制线, 蓝线为 $\mu + 3\sigma$ 上控制线, 由图可知故障发生前并不是所有的残差都超出控制线范围。采用 AC-GAN 生成对抗网络重构主轴异常机组温度残差序列, 得到异常轴承下的残差特征, 解决异常样本数据的标记问题, 进而提高了后续轴承状态决策模型的预测精度。

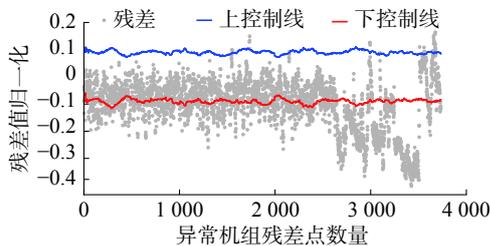


图 4 某机组故障发生前一个月温度残差

Fig. 4 Residual diagram of one month before a unit failure

4.2 AC-GAN 算法介绍

生成对抗网络 (generative adversarial networks, GAN) 是一种无监督学习方法, 包含生成器 G (generator) 和判别器 D (discriminator) 两部分。

将噪声信号 z 映射到样本空间, 通过生成器 G 得到生成样本数据 $X_{\text{fake}} = G(z)$, 将生成样本 X_{fake} 或真实样本数据 X_{real} 输入判别器 D 进而判定概率 $P(S|X) = D(X)$, 表示判别样本 X 属于 S 的概

率。由于 X_{real} 是真实数据, 所以判别器判定概率接近于 1, X_{fake} 是经生成器生成的数据并经过判别器判定是否真实。GAN 的目标函数为

$$\min_G \max_D V(D, G) = E[\ln P(S = \text{real} | X_{\text{real}})] + E[\ln P(S = \text{fake} | X_{\text{fake}})]$$

目标函数包含生成器目标函数和判别器目标函数, 其中生成器目标函数为

$$\min_G V(G) = E[\ln P(S = \text{fake} | X_{\text{fake}})]$$

判别器的目标函数为

$$\max_D V(D) = E[\ln P(S = \text{real} | X_{\text{real}})] + E[\ln P(S = \text{fake} | X_{\text{fake}})]$$

GAN 训练过程中, G 和 D 二者交替训练, 相互博弈, 最终使 G 生成的样本符合真实样本概率分布, 达到纳什均衡。GAN 在训练过程中无需先验概率即能学习真实样本的分布, 但同时 GAN 对初始参数极其敏感, 输入 G 的随机噪声信号无约束, 导致生成数据概率分布与真实数据差异大, 训练过程难以收敛, 使整个训练过程出现震荡, 发生模式崩溃^[21-22]。为了解决上述问题, 文献 [23] 提出了带标签辅助分类器的生成对抗网络 AC-GAN, 在传统 GAN 基础上增加了噪声数据对应的标签 c , 使用两者来生成 $X_{\text{fake}} = G(c, z)$, 判别器计算生成数据的概率分布 $P(S|X)$ 和类标签上的概率分布 $P(c|X)$ 分别为

$$\begin{cases} P(S|X) = D(X) \\ P(c|X) = D(X) \end{cases}$$

式中: $c = \{0, 1, 2, \dots, n\}$, n 表示样本类数。真实样本数据标记为正常温度残差 ($c = 0$) 和异常温度残差 ($c = 1$), AC-GAN 通过内部博弈, 最终实现主轴异常机组温度残差的重构。

4.3 基于 AC-GAN 的温度残差重构

AC-GAN 生成器和判别器均采用 RNN 神经网络, 输入层为 32 个神经元, 隐含层有 3 层神经元, 输出层为 1 个神经元, 激活函数为 sigmoid, 损伤函数为均方误差。AC-GAN 生成对抗网络的基本框架见图 5。

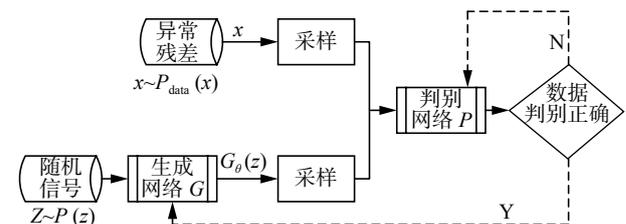


图 5 AC-GAN 对抗神经网络生成残差的基本框架

Fig. 5 Basic framework for the AC-GAN to generate residuals

温度残差重构具体步骤为:

1) 将主轴承发生故障前一个月的 SCADA 特征数据输入至基于 LightGBM 的主轴承温度预测模型, 计算得到一个月内的主轴承异常机组温度残差。

2) 采用 SPC 控制图将主轴承异常机组残差在控制范围内的正常温度残差数据剔除。

3) 定义控制线之外的残差数据为训练样本数据, 采用 AC-GAN 算法生成与真实样本数据分布相似的残差数据, 用来重构剔除的正常温度残差 2) 中剔除的正常温度残差。

5 基于 NGBoost 的主轴承运行状态决策

NGBoost 算法由斯坦福吴恩达团队在 2019 年 10 月提出的, 最初是用于预测不确定性估计的天气预测和医疗保健等领域^[24]。该算法使用自然梯度的 Boosting 方法, 这种方法可直接在输出空间中得到全概率分布, 用于预测量化不确定性。NGBoost 算法区别于其他 Boosting 算法是因为该算法可以返回每个预测的概率分布, NGBoost 通过预测参数 θ , 产生概率密度为 $P_\theta(y|x)$ 的概率预测。NGBoost 使用自然梯度来学习参数, 使得优化问题不受参数化的影响, 随后在 GBM 的框架下让每个基学习器去拟合自然梯度, 最后经过放缩和加权组合, 得到一个集成模型的参数, 由此可以学习最终条件分布参数, 从而达到概率预测的目的。NGBoost 作为一种梯度提升算法, 使用自然梯度 (natural gradient) 解决现有梯度提升方法难以处理的通用概率预测的技术难题, NGBoost 在不确定性估计和传统指标上的预测能力具有相当大的优势^[25]。

对于自然梯度, 首先定义得分规则为 S , 这个得分规则与概率分布 P 和输出值 y 相关, 记为 $S(P, y)$ 。对于正确的概率分布, 并期望取得最佳值, 概率预测越准确, 损失越小, 故有

$$E_{y \sim Q}[S(Q, y)] \leq E_{y \sim Q}[S(P, y)], \quad \forall P, Q$$

式中: Q 为正确的概率分布; P 为预测的概率分布。定义得分规则的散度:

$$D_S(Q||P) = E_{y \sim Q}[S(P, y)] - E_{y \sim Q}[S(Q, y)]$$

得分规则定义需要使得散度非负, 并且能够测量分布间的距离。如果得分规则定义为 MIF, 则

$$S = L(\theta, y) = -\ln P_\theta(y)$$

最大似然估计函数 $\sum \ln(P_\theta(y))$ 取得最大值, 同时使得散度变为 KL 散度, 即

$$D_L(Q||P) = E_{y \sim Q} \left[\ln \frac{Q(y)}{P(y)} \right] = D_{KL}(Q||P)$$

如果得分规则定义为 CRPS(连续概率排位分数), 则

$$C(\theta, y) = \int_{-\infty}^y F_\theta(z)^2 dz + \int_y^{\infty} (1 - F_\theta(z))^2 dz$$

式中 F 为累计概率分布函数。本来定义 CRPS 为

$$C^{\text{real}}(\theta, y) = \int_{-\infty}^{\infty} (F_\theta(z) - F_\theta^Q(z))^2 dz$$

式中: $F_\theta^Q(z)$ 为真实的累计概率分布, 当预测分布与真实分布完全一致时, CRPS 为零。预测分布过于集中, 过于分散, 或者偏离观测值太远时都会导致 CRPS 增大。但是, 真实概率分布未知, 只能通过一系列观察数据的累计分布来定义, 故有 $C(\theta, y)$, 对应的散度为

$$D_C(\theta, y) = E_{y \sim Q}[C(P, y)] - E_{y \sim Q}[C(Q, y)] = \int_{-\infty}^{\infty} (F_\theta(x) - F_P(x))^2 dx$$

CRPS 对于真实概率, 其值最小, 散度非负, 度量分布距离符合规则的定义要求。对于选择的得分规则 $S(\theta)$, 满足 $E[\nabla S(\theta)] = 0$, 对 L 关于 θ 求导有

$$\nabla S(\theta, y) + E[\nabla^2 S(\theta)] = 0$$

最终的梯度为

$$\tilde{\nabla} S(\theta, y) = (E[\nabla^2 S])^{-1} \nabla S(\theta)$$

对于 MIF, 有

$$\tilde{\nabla} L(\theta, y) = E[-\nabla_\theta^2 L(\theta, y)]^{-1} \nabla S(\theta, y)$$

对于 CRPS, 有

$$\tilde{\nabla} C(\theta, y) = \left[2 \int_{-\infty}^{\infty} \nabla_\theta F_\theta(z) \nabla_\theta F_\theta(z)^T dz \right]^{-1} \nabla C(\theta, y)$$

风电机组的主轴承温度残差同样是一种不确定估计, 对主轴承状态监测实质也是一种故障发生不确定性的概率预测, 因此可利用该算法建立主轴承运行状态决策模型。

为使降低故障决策中单点误报, 获得较高的预测准确度, 本文按天提取主轴承温度残差的最大值、最小值、均值、偏度、峰度、中位数、方差、标准差 8 个特征, 将主轴承正常机组的温度残差标签设置为 0, 主轴承异常机组重构的温度残差标签设置为 1。

6 实例应用验证

本实验所采用的 SCADA 数据为河北某风电场的历史数据。选取机组编号为 07#、09#、23#、31#、32#、41#、14#、11#、71#、81#、10 台主轴承正常机组, 选取编号为 69#、37#、84#、88#、99#、15#、96#、86#、13#、70# 的 10 台主轴承异常机组, 共计 20 台机组。通过基于 LightGBM 的主轴承温

度预测模型计算各机组主轴承温度残差, 对于异常机组通过 AC-GAN 算法重构残差值。提取主轴承正常机组和主轴承异常机组 8 个温度残差特征作为状态决策模型输入, 预测的主轴承正常或异常的概率值作为状态决策模型输出。由于主轴承正常机组标签设置为 0, 异常机组标签设置为 1, 因此, 输出概率在 0.5 以上可判断为异常状态, 0.5 以下判断为正常状态。温度残差特征 80% 用于训练状态决策模型, 剩余 20% 用于测试模型的准确度。按上述流程提取残差特征后共得到 938 组

数据。对比分析 NGBoost、XGBoost 和随机森林 3 种算法在残差重构前后测试样本中的准确性。

NGBoost 算法在测试集上的准确为 0.875, 混淆矩阵见图 6(a)。采用同样的特征提取方法提取没有经过残差重构的特征, 在测试集上的准确度为 0.660, 混淆矩阵见图 6(b)。XGBoost 算法在测试集上的准确为 0.843, 混淆矩阵见图 7(a)。采用同样的特征提取方法提取没有经过残差重构的特征, 在测试集上的准确度为 0.651, 混淆矩阵见图 7(b)。

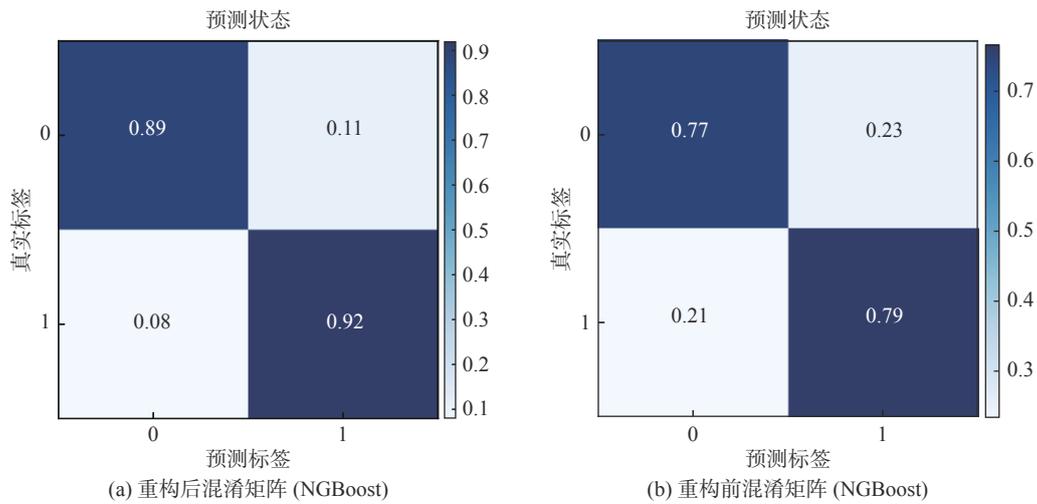


图 6 残差重构前后故障决策模型混淆矩阵 (NGBoost 算法)

Fig. 6 Confusion matrix of the fault decision model before and after residual reconstruction(NGBoost algorithm)

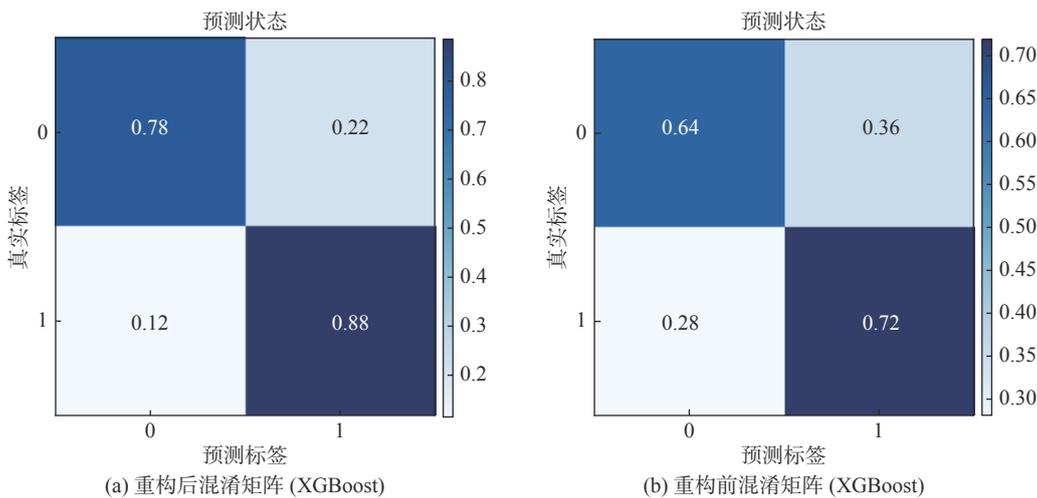


图 7 残差重构前后故障决策模型混淆矩阵 (XGBoost 算法)

Fig. 7 Confusion matrix of the fault decision model before and after residual reconstruction(XGBoost algorithm)

随机森林算法在测试集上的准确为 0.750, 混淆矩阵见图 8(a)。采用同样的特征提取方法提取没有经过残差重构的特征, 在测试集上的准确度为 0.642, 混淆矩阵见图 8(b)。

实验结果表明: 在同等条件下, NGBoost 算法在风电机组主轴承状态决策模型中优于 XG-

Boost 算法和随机森林算法。而且 NGBoost、XGBoost 和随机森林 3 种算法经过残差重构的状态决策模型的准确度分别提高了 21.5%、19.2%、10.8%, 说明基于 AG-GAN 的数据重构对风电机组主轴承运行状态具有良好的预测准确度。

下一步, 选取未参与模型训练和测试的 6 台

机组用来验证基于残差重构和 NGBoost 算法下的准确性和泛化性。选取主轴正常机组编号为

08#、65#、66#，主轴异常机组编号为 10#、85#、91#。

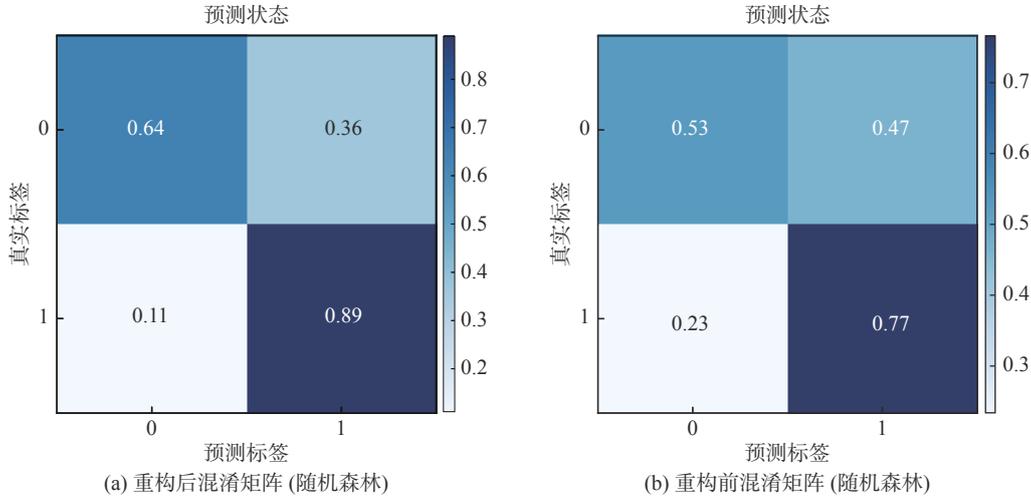


图 8 残差重构前后故障决策模型混淆矩阵 (随机森林算法)

Fig. 8 Confusion matrix of the fault decision model before and after residual reconstruction (random forest algorithm)

10#机组在 2019 年 3 月 28 号巡检时发现该机组主轴振动过大, 需要更换主轴, 选取 2019 年 3 月 1 号到 2019 年 3 月 30 号的数据。85#机组在 2019 年 4 月 1 号巡检时有异常响声, 经厂家检查发现该机组主轴电腐蚀严重, 因振动导致平衡环境存在轻微开裂迹象, 选取 2019 年 3 月 8 号到 4 月 4 号的数据。91#机组在 2019 年 10 月 13 号发生主轴开裂故障, 选取 2019 年 9 月 16 号到 10 月 16 号的数据。正常机组随机选取一个月的数据。

对比分析表 5、6, 正常机组编号为 08#的机组在经过残差重构后预测的准确率提升了 28%, 65#机组的准确率提升了 3%, 66#机组的准确率提升了 1%。对比分析表 7、8, 异常机组编号为 10#的机组在经过残差重构后预测的准确率提升了 17%, 85#机组的准确率提升了 15%, 91#机组的准确率提升了 7%。

表 5 正常机组未经过残差重构时的概率预测值

Table 5 Probability prediction value of the normal wind turbine without residual reconstruction

风机编号	预测为0(正常)	预测为1(异常)
08#	0.64	0.36
65#	0.58	0.42
66#	0.64	0.36

表 6 正常机组经过残差重构时的概率预测值

Table 6 Probability prediction value of the normal wind turbine after residual reconstruction

风机编号	预测为0(正常)	预测为1(异常)
08#	0.92	0.08
65#	0.61	0.39
66#	0.65	0.35

表 7 异常机组未经过残差重构时的概率预测值

Table 7 Probability prediction value of the abnormal wind turbine without residual reconstruction

风机编号	预测为0(正常)	预测为1(异常)
10#	0.34	0.66
85#	0.43	0.57
91#	0.46	0.54

表 8 异常机组经过残差重构时的概率预测值

Table 8 Probability prediction value of the abnormal wind turbine after residual reconstruction

风机编号	预测为0(正常)	预测为1(异常)
10#	0.17	0.83
85#	0.28	0.72
91#	0.39	0.61

7 结束语

本文以风电机组主轴为研究对象, 针对状态监测和故障预警中人为设定阈值的相关问题, 提出了基于 AC-GAN 数据重构的风电机组主轴状态监测方法, 得到如下结论:

1) 采用 LightGBM 算法建立主轴温度残差预测模型, 并将 XGBoost 算法、CatBoost 算法与之对比分析, 在同等条件下, LightGBM 算法在主轴温度建模中综合性能优于 XGBoost 算法和 CatBoost 算法。

2) 采用滑动窗口提取主轴异常机组残差, 利用 SPC 方法对主轴异常温度残差在控制线范围内进行筛选, 并利用 AC-GAN 算法对残差序列

进行重构,解决了人为设定阈值的相关问题,提升了主轴承异常和正常数据标签标注的准确率。

3) 在同等条件下,NGBoost算法在风电机组主轴承状态决策模型中优于XGBoost算法和随机森林算法。而且,NGBoost、XGBoost和随机森林3种算法经过残差重构的状态决策模型的准确度分别提高了21.5%、19.2%、10.8%,选择6台机组进行测试分析(3台主轴承正常机组,3台主轴承异常机组),均能够判断正确。基于NGBoost的状态决策模型的平均准确率从60.5%(无残差序列重构)提升至72.3%(利用残差数据重构)。

参考文献:

- [1] 曾军,陈艳峰,杨苹,等.大型风力发电机组故障诊断综述[J].电网技术,2018,42(3):849-860.
ZENG Jun, CHEN Yanfeng, YANG Ping, et al. Review of fault diagnosis methods of large-scale wind turbines[J]. Power system technology, 2018, 42(3): 849-860.
- [2] 金晓航,孙毅,单继宏,等.风力发电机组故障诊断与预测技术研究综述[J].仪器仪表学报,2017,38(5):1041-1053.
JIN Xiaohang, SUN Yi, SHAN Jihong, et al. Fault diagnosis and prognosis for wind turbines: an overview[J]. Chinese journal of scientific instrument, 2017, 38(5): 1041-1053.
- [3] 刘帅.基于实时监测数据挖掘的风电机组故障预警方法研究[D].北京:华北电力大学(北京),2019.
LIU Shuai. Research on fault warning method of wind turbine based on real-time monitoring data mining[D]. Beijing: North China Electric Power University (Beijing), 2019.
- [4] 李俊卿,李斯璇,陈雅婷,等.同步发电机定子故障预警模型[J].电力科学与工程,2020,36(5):7-14.
LI Junqing, LI Sixuan, CHEN Yating, et al. Synchronous generator stator fault prediction model[J]. Electric power science and engineering, 2020, 36(5): 7-14.
- [5] 王梓齐,刘长良.基于Box-Cox变换和相对熵残差分析的风电机组齿轮箱状态监测[J].中国电机工程学报,2020,40(13):4210-4218.
WANG Ziqi, LIU Changliang. Wind turbine gearbox condition monitoring based on Box-Cox transformation and relative entropy residual analysis[J]. Proceedings of the CSEE, 2020, 40(13): 4210-4218.
- [6] 刘帅,刘长良,甄成刚.基于数据分类重建的风电机组故障预警方法[J].仪器仪表学报,2019,40(8):1-11.
LIU Shuai, LIU Changliang, ZHEN Chenggang. Fault warning method for wind turbine based on classified data reconstruction[J]. Chinese journal of scientific instrument, 2019, 40(8): 1-11.
- [7] QIU Yingning, FENG Yanhui, INFIELD D. Fault diagnosis of wind turbine with SCADA alarms based multi-dimensional information processing method[J]. Renewable energy, 2020, 145: 1923-1931.
- [8] LIANG Tao, QIAN Siqi, MENG Zhaochao, et al. Early fault warning of wind turbine based on BRNN and large sliding window[J]. Journal of intelligent & fuzzy systems, 2020, 38(3): 3389-3401.
- [9] 尹诗,侯国莲,于晓东,等.基于Bi-RNN的风电机组主轴承温度预警方法研究[J].郑州大学学报(工学版),2019,40(5):44-50.
YIN Shi, HOU Guolian, YU Xiaodong, et al. Research on temperature prediction method for main bearing of wind turbine based on Bi-RNN[J]. Journal of Zhengzhou University (engineering science edition), 2019, 40(5): 44-50.
- [10] 陈雪峰,郭艳婕,许才彬,等.风电装备故障诊断与健康监测研究综述[J].中国机械工程,2020,31(2):175-189.
CHEN Xuefeng, GUO Yanjie, XU Caibin, et al. Review of fault diagnosis and health monitoring for wind power equipment[J]. China mechanical engineering, 2020, 31(2): 175-189.
- [11] 沈小军,付雪姣,周冲成,等.风电机组风速-功率异常运行数据特征及清洗方法[J].电工技术学报,2018,33(14):3353-3361.
SHEN Xiaojun, FU Xuejiao, ZHOU Chongcheng, et al. Characteristics of outliers in wind speed-power operation data of wind turbines and its cleaning method[J]. Transactions of China electrotechnical society, 2018, 33(14): 3353-3361.
- [12] 贾科,杨哲,魏超,等.基于斯皮尔曼等级相关系数的新能源送出线路纵联保护[J].电力系统自动化,2020,44(15):103-111.
JIA Ke, YANG Zhe, WEI Chao, et al. Pilot protection based on spearman rank correlation coefficient for transmission line connected to renewable energy source[J]. Automation of electric power systems, 2020, 44(15): 103-111.
- [13] CHEN Tianqi, GUESTRIN C. XGBoost: a scalable tree boosting system[C]//22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, USA, 2016: 785-794.
- [14] KE Guolin, MENG Qi, FINLEY T, et al. LightGBM: a highly efficient gradient boosting decision tree[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA, 2017: 3149-3157.
- [15] 黄伟,李阳.基于MCS-MIFS与LightGBM的燃气轮机功率预测方法[J].电力科学与工程,2020,36(5):23-31.
HUANG Wei, LI Yang. Gas turbine power forecasting

- based on MCS-MIFS and LightGBM[J]. *Electric power science and engineering*, 2020, 36(5): 23–31.
- [16] 周挺, 杨军, 周强明, 等. 基于改进 LightGBM 的电力系统暂态稳定评估方法 [J]. *电网技术*, 2019, 43(6): 1931–1940.
ZHOU Ting, YANG Jun, ZHOU Qiangming, et al. Power system transient stability assessment method based on modified LightGBM[J]. *Power system technology*, 2019, 43(6): 1931–1940.
- [17] 李楠, 李保罗, 朱建华, 等. 计及样本不平衡与重叠的暂态稳定评估方法 [J]. *电力系统自动化*, 2020, 44(21): 64–71.
LI Nan, LI Baoluo, ZHU Jianhua, et al. Transient stability assessment method considering sample imbalance and overlap[J]. *Automation of electric power systems*, 2020, 44(21): 64–71.
- [18] 王熠琛, 郑宏, 张明聚, 等. 统计过程控制在地铁穿越工程沉降监测分析中的应用 [J]. *岩石力学与工程学报*, 2019, 38(S1): 3238–3245.
WANG Yichen, ZHENG Hong, ZHANG Mingju, et al. Settlement monitoring and analysis for subway under engineering by using statistical process control[J]. *Chinese journal of rock mechanics and engineering*, 2019, 38(S1): 3238–3245.
- [19] 胡姚刚. 大功率风电机组关键部件健康状态监测与评估方法研究 [D]. 重庆: 重庆大学, 2017.
HU Yaogang. Study of health condition monitoring and assessment for critical components of large wind turbine generator systems[D]. Chongqing: Chongqing University, 2017.
- [20] 于广滨, 卓识, 于军, 等. 基于 InfoLSGAN 和 AC 算法的滚动轴承剩余寿命预测 [J]. *航空动力学报*, 2020, 35(6): 1212–1221.
YU Guangbin, ZHUO Shi, YU Jun, et al. Remaining useful life prediction of rolling bearings using InfoLSGAN and AC algorithm[J]. *Journal of aerospace power*, 2020, 35(6): 1212–1221.
- [21] 曹锦纲, 李金华, 郑顾平. 基于生成式对抗网络的道路交通模糊图像增强 [J]. *智能系统学报*, 2020, 15(3): 491–498.
CAO Jin'gang, LI Jinhua, ZHENG Guping. Enhancement of blurred road-traffic images based on generative adversarial network[J]. *CAAI transactions on intelligent systems*, 2020, 15(3): 491–498.
- [22] 毕晓君, 潘梦迪. 基于生成对抗网络的机载遥感图像超分辨率重建 [J]. *智能系统学报*, 2020, 15(1): 74–83.
BI Xiaojun, PAN Mengdi. Super-resolution reconstruction of airborne remote sensing images based on the generative adversarial networks[J]. *CAAI transactions on intelligent systems*, 2020, 15(1): 74–83.
- [23] ODENA A, OLAH C, SHLENS J. Conditional image synthesis with auxiliary classifier GANs[C]//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia, 2017: 2642–2651.
- [24] DUAN T, ANAND A, DING D Y, et al. NGBoost: natural gradient boosting for probabilistic prediction [C]//Proceedings of the 37th International Conference on Machine Learning. Vienna, Austria, 2020: 2690–2700.
- [25] TRUSHEIM F, CONDURACHE A, MERTINS A. Boosting black-box variational inference by incorporating the natural gradient[C]//2018 24th International Conference on Pattern Recognition (ICPR). Beijing, China, 2018: 19–24.

作者简介:



尹诗, 博士研究生, 主要研究方向为风电机组故障预警、新能源状态监测。



侯国莲, 教授, 博士生导师, 主要研究方向为风电故障诊断、复杂系统优化控制、新能源技术及控制策略。



胡晓东, 硕士研究生, 主要研究方向为风电机组故障诊断和预警。