

基于随机森林的海上风电机组发电机轴承 异常状态监测方法

饶 雷^{1,2}, 冉 军², 陶建权², 胡号朋², 吴 沁², 熊圣新²

(1. 南方海洋科学与工程广东省实验室(湛江), 广东湛江 524013; 2. 中国船舶重工集团海装风电股份有限公司, 重庆 401120)

摘 要: 发电机轴承是海上风电机组中重要的零部件, 其状态直接影响了海上风电机组的运行状态和发电量。文章通过结合机理分析和数据驱动提出了基于随机森林的海上风电机组发电机轴承异常状态监测方法, 该方法先通过机理分析选取变量、清洗数据和标定样本状态, 然后通过数据驱动的方法对海上风电机组发电机轴承的状态进行预测。通过现场实际的海上风电机组 SCADA 数据对该模型进行验证, 所述方法能够有效预测海上风电机组发电机轴承的状态, 并且能有效避免对异常状态的误报和漏报。

关键词: 海上风电机组; 发电机轴承; 机理分析; 数据驱动; 随机森林

中图分类号: TK83 **文献标志码:** A **【DOI】** 10.13788/j.cnki.cbge.2022.S2.06

Abnormal Condition Monitoring Method for Generator Bearing of Offshore Wind Turbine Based on Random Forest

RAO Lei^{1,2}, RAN Jun², TAO Jianquan², HU Haopeng², WU Qin², XIONG Shengxin²

(1. Southern Marine Science and Engineering Guangdong Laboratory (Zhanjiang), Zhanjiang 524013, Guangdong, China; 2. CSIC Haizhuang Windpower Co., Ltd., Chongqing 401120, China)

Abstract: Generator bearing is an important part of offshore wind turbine, its state directly affects the operation state and power generation of offshore wind turbine. By combining mechanism analysis and data driving, this paper proposes a method for monitoring the abnormal state of offshore wind turbine generator bearing based on random forest. This method first selects variables, cleaning data and calibration sample state through mechanism analysis, and then forecasts the state of offshore wind turbine generator bearing through data driving method. The model is verified by on-site actual offshore wind turbine SCADA data, and the method can effectively predict the state of offshore wind turbine generator bearing, and can effectively avoid false alarm and missing alarm of abnormal state.

Key words: offshore wind turbines; generator bearing; mechanism analysis; data driven; random forest

0 引言

“减少碳排放”现在已成为全世界的共同目标, 风电机组的装机量在逐年上升。截止 2021 年, 全球新增装机量 99.2 GW, 其中中国新增装机量达到 55.8 GW, 占全球新增装机量的 57%。风场复杂的风况和恶劣的环境容易导致风电机组的零部件出现提前失效的情况^[1], 发电机轴承作为风力发电机中重要的机械部件直接影响风力发电机的正常运行。一旦出现

异常将影响风电机组的正常运行, 造成发电量的损失, 严重时还会影响现场人员的生命财产安全, 因此研究出能够准确识别出风电机组发电机轴承状态的方法有着很高的实际意义和应用价值^[2]。

目前, 针对风电机组发电机轴承的状态监测的方法研究已有较多的研究, 但是主要以试验室数据进行模型的构建和验证, 缺乏对现场实际 SCADA 数据情况的考虑与分析, 因此难以应用于现场实际部署。因

收稿日期: 2022-09-13; 修回日期: 2022-10-24

基金项目: 国家重点研发计划(2020YFB1506600)

作者简介: 饶 雷(1994—), 男, 硕士、助理工程师。研究方向: 大数据建模。

此本文考虑海上风电场实际的 SCADA 数据情况, 通过机理分析与数据驱动相结合提出了基于随机森林的海上风电机组发电机轴承异常状态监测模型。

1 随机森林

随机森林 (Random Forest, RF) 是一种集成学习算法, 是将多个弱分类器进行组合然后经过投票或求均值确定最后的结果。该算法由 BREIMAN 提出^[3], 综合了其提出的 Bagging 算法^[4]和 HO 提出的随机子空间方法^[5], 这可整体提高模型的精度和泛化能力。随机森林算法中的基本分类器是 CART 决策树, 又被称为分类回归树 (Classification and Regression Tree, CART)。

CART 算法是 BREIMAN 和 STONE 共同提出, 该算法在信息熵的基础上并依据特定的准则构造得到的决策树。该算法与 ID3 决策树算法和 C4.5 决策树算法不同的是在节点分裂选择特征属性的时候采用了基

尼指数 (Gini Index) 最小准则来进行选择, 通过基尼指数可以选择特征属性并且通过基尼值可以反映样本的纯度。数据集 D 的纯度定义为

$$Gini(D) = \sum_{k=1}^{|y|} \sum_{k' \neq k} p_k p_{k'} = 1 - \sum_{k=1}^{|y|} p_k^2 \quad (1)$$

式中: p_k 为样本点属于第 k 类的概率; y 表示数据集 D 的类别。

基尼指数定义为

$$Gini_{\text{index}}(D, \alpha) = \sum_{v=1}^{|y|} \frac{|D^v|}{|D|} Gini(D^v) \quad (2)$$

式中: α 为特征条件; v 表示数据集 D 中满足特征条件 α 的类别。

随机森林算法属于 Bagging 算法的一种, 因此两种算法的算法流程基本上是一致的, 其具体流程见图 1。

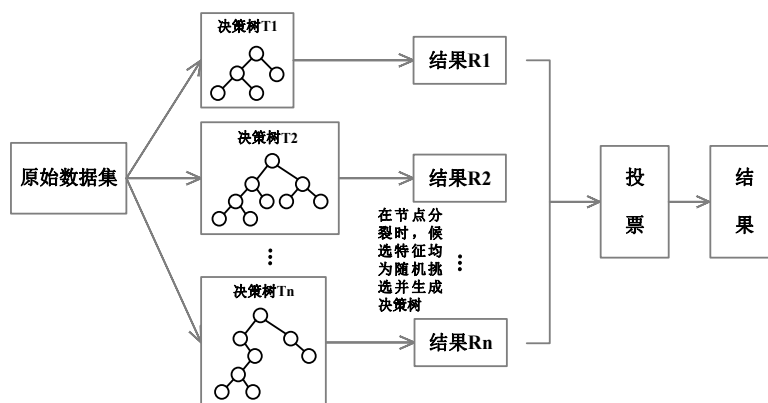


图 1 随机森林算法流程

1) 在原始样本数据集中进行 k 次 Bootstrap 重采样, 每次采集固定个数的样本且每采集一个样本都将样本放回, 采样结束后即可得到 k 个子样本集。

2) 利用 CART 算法构建针对于每个子样本集的决策树。假设从第 i 个子样本集中选取出的特征 f_i 中包含了 C 个类别, 则其 $Gini$ 值为

$$Gini(f_i) = \sum_{j=1}^C \sum_{j' \neq j} p_j p_{j'} = 1 - \sum_{j=1}^C p_j^2 \quad (3)$$

式中: p_j 为类别 C_j 出现的概率。

$Gini$ 值越小样本纯度越高。因此, 决策树节点的分裂特征即为该决策树中所有特征 $Gini$ 最小的特征。

3) 每个子样本集按照步骤 2 产生一颗决策树, 所有样本子集的决策树共同构成随机森林。每棵决策树在生长时, 随机从含有 M 个特征的特征集合 T 中抽取 \sqrt{M} 个特征作为一个特征子集, 将抽取出的特征子集

作为决策树的划分属性, 按照 $Gini$ 值最小准则生长决策树。

4) 利用多数投票算法对所有决策树的结果进行分析和投票, 最终的投票结果即为随机森林的结果。

2 基于随机森林的海上风电机组发电机轴承异常状态监测模型

海上风电机组发电机轴承工作环境恶劣, 经常需要承受高温和高载荷, 而且受到外部海洋环境带来的复杂风况也对发电机轴承的工作情况产生巨大的影响, 随机森林算法具有较强的鲁棒性, 对数据的敏感程度较低, 所以不会因为个别异常样本对模型的结果造成影响^[6]。因此本问基于随机森林算法构建了海上风电机组发电机轴承异常状态监测模型 (见图 2)。

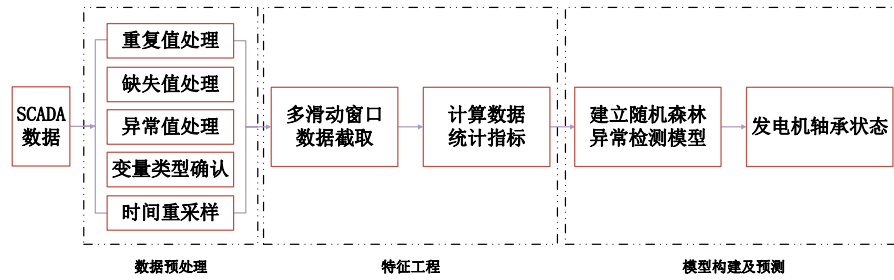


图2 海上风电机组发电机轴承异常监测流程

图2展示了本文提出的基于随机森林的海上风电机组发电机轴承异常状态监测模型的算法流程图,该流程分为3个主要的部分:数据预处理、特征工程和模型构建及预测。

1)数据预处理包含5个子部分:重复值处理、缺失值处理、异常值处理、变量类型确认和时间重采样。

(1)重复值处理。SCADA数据会因为采集和存储等因素造成变量数据重复或者时间戳重复,本文针对重复值处理的方式是将时间戳相同的数据进行合并,对时间戳重复的数据进行“并集”操作,最终达到所有时间戳去重,并保留其对应数据的唯一性。

(2)缺失值处理。同样是因为采集和存储的原因,SCADA数据会出现数据缺失的情况,本文针对缺失值处理的方式是用上一个时刻的数据填充下一个时刻的缺失数据。

(3)异常值处理。缺失值数据一般是由传感器等器件的松动损坏造成了数据异常,这部分异常数据会影响后面模型的构建,本文针对异常值的处理方法是通过对机理理解对各个变量中严重偏离正常值所对应的时间戳样本进行删除。

(4)变量类型确认。SCADA分钟级数据在存储时,状态量和模拟量的存储方式不同,状态量是变化存储,当出现变化时会直接插入到数据中,带有具体的秒级时间戳,且其他变量会保持缺失,因此需要进行分情况处理。

(5)时间重采样。由于SCADA数据可能存在某时间段数据缺失,所以本文针对时间重采样采取如下操作:首先模拟量变量重采样为完整规律的1min采样周期,若发生某段时间没有采集数据,也给予时间戳,变量值标记缺失;状态量变量通过适当的方式向前填充,补充到每1min的规整时刻中。填充方法考虑下一次状态量变化时刻,以及模拟量全部缺失的时间段;将状态量变量通过独热编码的形式,构造只包含0、1值的变量,并统计在每1min规整时刻该变量在过去1min的和值。

2)特征工程部分包含2个子部分:多滑动窗口数据截取和数据统计指标计算。首先采用多滑动窗口数据截取对输入的数据在时间维度上按照不同尺寸的滑动窗口进行截取,每个尺寸的滑动窗口会对输入的数

据进行完整的截取,然后对所有不同尺寸窗口内的数据计算他们的统计指标,最后将这些统计指标作为风电机组发电机异常监测模型的特征。

3)模型构建及预测共包含建立随机森林异常监测模型和发电机轴承状态监测。将通过步骤2得到的特征向量输入到基于随机森林构建的海上风电机组发电机轴承异常监测模型中进行模型训练,然后将训练好的风电机组发电机轴承异常监测模型应用到实际工况中监测并给出当前风电机组发电机轴承的状态。

3 试验验证

为了验证本文所提出的海上风电机组发电机轴承异常监测模型的有效性,以我国南方某海上风场的风电机组的SCADA系统在线监测数据为例进行验证分析。

3.1 变量选取及数据标签确定

通过对风电机组发电机有关变量的相关性分析,并结合机理分析筛选出其他能够直接或间接影响风电机组发电机轴承状态的有关变量,具体见表1。

表1 风电机组发电机轴承异常状态监测需求变量

序号	变量名	类型
1	10 min 平均有功功率	Single
2	10 s 平均有功功率	Single
3	60 s 平均有功功率	Single
4	发电机冷却水温度	Single
5	发电机前轴承温度	Single
6	发电机后轴承温度	Single
7	发电机有功功率	Single
8	发电机滑环室温度	Single
9	发电机绕组 u1 温度	Single
10	发电机绕组 u2 温度	Single
11	发电机绕组 v1 温度	Single
12	发电机绕组 v2 温度	Single
13	发电机绕组 w1 温度	Single
14	发电机绕组 w2 温度	Single
15	发电机绕组最高温度	Single
16	发电机转矩	Single
17	发电机转速	Single
18	机舱前端温度	Single

在本文的验证试验中，通过结合故障报警时间来确定样本的标签，因此将故障报警时刻之前 30 天的数据标签定为 1，即状态异常样本；考虑到现场的实际检修和更换的情况，为了能合理准确的确定正常样本数据，所以报警时刻之后 30 天的数据标签定为-1，即不进行状态评价；若两个故障报警时刻相隔天数小于 30 天，则两个时刻之间的数据同样定为状态异常样本；其他时段的数据则视为状态正常样本，标签设定为 0。

3.2 有效性分析及验证

本次验证试验对模型的评价选用了机器学习算法常见评价指标，即精确率(Precision)、召回率(Recall)、F1 分数(F1_Score)、ROC_AUC。为了进一步提升模型的精度和鲁棒性，试验中对模型进行了 3 折交叉验证，具体的数据划分如表 2 所示。其中， K_0 、 K_1 和 K_2 用于交叉验证， K_3 用于测试。

表 2 风电机组发电机轴承异常状态监测模型 3 折交叉验证数据划分

K 折编号	训练集包含的风机号	验证集包含的风机号
K_0	3, 4, 7, 8, 10, 13, 18, 19, 20, 21, 22, 24	14, 6, 11, 17, 5, 15
K_1	5, 6, 7, 8, 10, 11, 13, 14, 15, 17, 20, 22	21, 19, 4, 24, 18, 3
K_2	3, 4, 5, 6, 11, 14, 15, 17, 18, 19, 21, 24	20, 8, 10, 22, 7, 13
K_3	3, 4, 5, 6, 7, 8, 10, 11, 13, 14, 15, 17, 18, 19, 20, 21, 22, 25	1, 2, 9, 12, 16, 23, 24

经过建模和交叉验证结果的分析，最终得到表 3 所示的评价结果。从表中可以看出风电机组发电机轴承异常状态监测模型经过 3 折交叉验证后，其 ROC_AUC 的均值达到了 0.762 6 明显高于 0.5，说明

该模型对发电机轴承的异常状态具有较好的预测能力。而且通过 K_3 的训练集和验证，模型的 ROC_AUC 指标达到 0.950 8，准确率(Accuracy)达到 0.895 2，因此模型已经具有良好的预测能力。

表 3 风电机组发电机轴承异常状态检测模型评价结果

K 折编号	ROC_AUC	准确率	精确率	召回率	F1 分数
K_0	0.645 3	0.870 2	0.093 9	0.033 3	0.049 2
K_1	0.836 3	0.695 1	0.886 0	0.041 8	0.079 8
K_2	0.806 1	0.720 7	0.293 5	0.816 7	0.431 8
Mean(K_0 , K_1 , K_2)	0.762 6	0.762 0	0.424 4	0.297 3	0.186 9
K_3	0.950 8	0.895 2	0.676 1	0.536 7	0.598 4

最后将风电机组的样本数据都输入到本文提出的基于随机森林的海上风电机组发电机轴承异常状态检测模型中进行测试，并将模型的输出结果与风电机组的实际状态情况进行对比并统计，统计结果见表 4。表 5 为模型在经过测试后的评价结果，表 5 中计算了 5 个指标：正确率、错误率、故障准确率、故障

漏报率、故障误报率，最终经过计算，这 5 个指标分别达到了 93.39%、6.10%、90%、0%和 10%。这 5 个指标的结果反应出，该模型能够很好的对风电机组发电机轴承的状态进行预测，对异常状态有较高的识别率，且异常状态的误报率和漏报率都较低。

表 4 模型输出结果统计

风机号	报警次数 FN+TP	预警次数 FP+TP	漏报次数 FN	误报次数 FP	报准次数 TP	无报警、无预警次数 TN
1	0	0	0	0	0	1
2	1	1	0	0	1	0
3	0	0	0	0	0	1
4	1	1	0	0	1	0
5	0	0	0	0	0	1
6	1	1	0	0	1	0
7	0	0	0	0	0	1
8	1	1	0	0	1	0
9	0	0	0	0	0	1
10	1	1	0	0	1	0
11	2	3	0	1	2	0

表 4 模型输出结果统计（续）

风机号	报警次数 FN+TP	预警次数 FP+TP	漏报次数 FN	误报次数 FP	报准次数 TP	无报警、无预警次数 TN
12	1	2	0	1	1	0
13	0	0	0	0	0	1
14	1	1	0	0	1	0
15	0	0	0	0	0	1
16	0	0	0	0	0	1
17	0	0	0	0	0	1
18	0	0	0	0	0	1
19	2	2	0	0	2	0
20	2	2	0	0	2	0
21	4	4	0	0	4	0
22	0	0	0	0	0	1
23	1	1	0	0	1	0
24	0	0	0	0	0	1
25	0	0	0	0	0	1
总计	18	20	0	2	18	13

表 5 评价指标结果

指标名称	计算公式	数值
漏报	FN	0
误报	FP	2
报准	TP	18
无报警、无预警	TN	13
正确率	$ACC_{correct} = (TP+TN)/(TP+TN+FP+FN) \times 100\%$	93.9%
错误率	$ACC_{error} = (FP+FN)/(TP+TN+FP+FN) \times 100\%$	6.1%
故障准确率	$F_{error} = TP/(TP+FN+FP) \times 100\%$	90.0%
故障漏报率	$F_{error} = FN/(TP+FN+FP) \times 100\%$	0%
故障误报率	$F_{miss} = FP/(TP+FN+FP) \times 100\%$	10.0%

4 结论

本文提出了一种基于随机森林的风电机组发电机轴承异常状态监测模型，并且以我国南方某海上风电场的风电机组的 SCADA 系统监测数据对该方法进行了验证。通过验证结果可以表明本文提出的模型方法在实际的海上风电机组数据环境中能够很好的对风电机组发电机轴承异常状态进行预测，并且能够准确识别出异常状态同时能够有效避免对异常状态的误报和漏报。

参考文献：

- [1] 黄宏臣, 郭四洲, 王子彦, 等. 加速度包络解调方法在风力发电机滚动轴承早期故障诊断应用研究[J]. 机械设计与制造, 2022(3): 251-253.
- [2] 吕明珠. 风电轴承状态监测与智能维护策略研究[J]. 电气开关, 2021, 59(3): 35-40.
- [3] BREIMAN L. Random Forest[J]. Machine Learning, 2001, 45(1): 5-32.
- [4] LEO B. Bagging Predictors[J]. Machine Learning, 1996, 24(2): 123-140.
- [5] HO T K. The Random Subspace Method for Constructing Decision Forests[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(8): 832-844.
- [6] ZHANG Y, CHEN J, WANG X F, et al. Application of Random Forest on Rolling Element Bearings Fault Diagnosis[J]. Computer Engineering & Applications, 2018, 15: 281-293.