

doi: 10.3969/j.issn.1672-0792.2020.09.006

基于 LSTM 网络的风机齿轮箱轴承故障预警

王 超, 李大中

(华北电力大学 控制与计算机工程学院, 河北 保定 071003)

摘 要:提出一种将互信息特征选择、深度学习和残差滑动窗口分析相结合的风机齿轮箱轴承故障预警方法。首先,对风电场 SCADA 数据进行预处理,采用互信息法筛选与齿轮箱轴承温度关联度高的特征。在此基础上,依据所选特征建立长短期记忆神经网络(LSTM)深度学习模型对齿轮箱轴承温度进行预测,并通过滑动窗口对预测残差进行分析处理,设定合适的报警阈值和规则。以西北某风电场现场数据对提出的方法进行验证,结果表明该方法可以对风机齿轮箱轴承故障进行有效预警。

关键词: SCADA 数据;深度学习;LSTM 网络;齿轮箱轴承;故障预警

中图分类号: TH17 文献标识码: A 文章编号: 1672-0792(2020)09-0040-06

Fault Early Warning of Fan Gearbox Bearing Based on LSTM Network

WANG Chao, LI Dazhong

(School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, China)

Abstract: A fault warning method for fan gearbox bearing is proposed, which combines mutual information feature selection, deep learning and sliding window residual analysis. First of all, the wind farm SCADA data are preprocessed, and the mutual information method is used to screen the characteristics with high correlation with gearbox bearing temperature. On this basis, according to the selected characteristics, the long short term memory neural network (LSTM) deep learning model is established to predict the gearbox bearing temperature, the prediction residual is analyzed and processed through the sliding window, and the appropriate alarm threshold and rules are set. The field data of a wind farm in northwest China are used to verify the proposed method. The results show that the method can effectively warn the fan gearbox bearing fault.

Key words: SCADA data; deep learning; LSTM network; gearbox bearing; fault early warning

收稿日期: 2020-07-07

基金项目: 河北省自然科学基金资助项目(F2011502001)

作者简介: 王 超(1996—),男,硕士研究生,研究方向为风电机组故障预测;

李大中(1961—),男,教授,研究方向为新能源发电系统控制及优化、智能优化理论及应用。

通信作者: 王 超

0 引言

齿轮箱作为风电机组重要组成部分之一,主要由齿轮和滚动轴承等零件组成,常见故障有磨损、点蚀、断裂等,其健康状况对机组的正常运行有着极大的影响。目前,风机故障报警大多依赖于SCADA系统设定的上下限阈值,但是阈值设定通常比较宽泛,从而导致无法及时报警。文献[1]利用主成分分析法选取对齿轮箱轴承温度影响较大的参数,建立BP神经网络预测模型,使用改进的粒子群算法对模型进行优化,最终实现轴承故障的有效预警。文献[2]首先对机组运行工况进行辨识,在每个工况中建立相应的双向循环神经网络预测模型,采用随机森林对模型预测残差进行分析,仿真结果表明该方法对机组轴承预警具有较高的准确率。文献[3]提出了一种基于SCADA数据的逐层编码网络学习模型,采用指数加权移动平均值对模型误差进行分析,实现轴承故障预测。文献[4]使用滑动窗口对模型残差进行统计,从而设定了相应的报警阈值。文献[5]结合信息熵和核极限学习机对机组异常工况进行预警。文献[6]建立自编码网络,提出了一种使用极值理论确定自适应报警阈值的方法,能够实现早期故障预警。文献[7]建立深度置信网络,与传统模型算法(朴素贝叶斯、随机森林、决策树)进行比较,结果表明该方法在机组故障诊断中具有较高的鲁棒性。文献[8]提出一种递归神经网络和自编码器相结合的算法,用于预测轴承的剩余使用寿命。文献[9]对机组轴承的振动信号进行特征提取,利用逻辑回归算法对轴承故障进行诊断,方法简单省时。文献[10]利用遗传算法对相关向量机进行优化,故障诊断结果表明该方法优于BP神经网络和支持向量机。文献[11]提出了一种W型结构元素特征提取方法,能够从原始信号中获得更多的特征信息,提升模型诊断结果的准确性。文献[12]利用卷积神经网络对机组运行数据进行特征提取,Soft-Max分类器用于诊断不同类型的齿轮箱故障,与传统算法相比,该方法效率较高。文献[13]采集机组的定子和转子电流信号,使用支持向量机模型对齿轮箱故障进行有效检测。

以上文献的研究方法推动了风机齿轮箱轴承

故障预警方面的研究,但依然存在一些研究盲点,比如未能考虑运行数据时间前后的联系;报警阈值设定不合理,泛化能力较差,模型复杂,参数量较大,参数寻优困难等。本文以风机齿轮箱轴承故障作为研究对象,首先利用互信息法从高维度的原始数据中提取出有用的特征信息,然后使用长短期记忆神经网络建立轴承温度预测模型,该模型能够对运行数据时间前后的关系进行深度特征提取。为了防止模型产生误报警,采用滑动窗口对预测误差变化趋势进行分析处理,设定相应的报警阈值和报警规则,以确保风机齿轮箱轴承故障预警的准确性和有效性。

1 风机运行数据处理

1.1 数据预处理

本文数据源自西北某风电场SCADA系统记录的2013年8月1日至10月2日的机组运行数据,报警记录显示齿轮箱轴承损坏。该数据详细地记录了机组的运行状态参数,包括时间、风速、总发电量、环境温度、发电机转速、风向角等,采样间隔为1 min。但是原始数据包含了一些无效数据,如机组启停、空值记录等,为了得到好的模型预测效果,需对原始数据进行预处理,按照以下原则剔除无效数据:

发电机输出功率小于零的点;风速小于切入风速,以及大于切出风速的点;发电机转速小于零的点。

1.2 特征选择

原始数据共包含数据特征45项,而这些高维度特征中存在着一些冗余和无关特征。数据决定了模型预测效果的上限,当模型输入维度超过某一界限时,会对模型的预测效果产生一定的影响,降低模型运算效率。因此有必要对原始数据进行特征选择,提取出对研究目标有用的数据信息。本文采用互信息法进行特征筛选,其公式为^[14]:

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \quad (1)$$

式中: $p(x)$ 、 $p(y)$ 为变量 x 、 y 的边缘分布概率; $p(x,y)$ 为联合分布概率。

经过特征选择后得到与齿轮箱轴承温度关联度较高的特征:风速、发电机输出、10 min 平均

风速、1 min 平均风速、1 min 发电机平均输出、机舱温度、环境温度、齿轮油温度、发电机定子温度、风向角。可以发现所选特征中并非全部都与齿轮箱轴承运行状态有直接关联,一些特征间接地影响着轴承温度的变化,如果忽略这些特征,模型预测效果可能会出现一定的偏差。

2 基于 LSTM 网络的预测模型建立

2.1 LSTM 原理

LSTM 模型是循环神经网络 (RNN) 的一种特殊形式,能够挖掘运行数据时间前后的联系,解决训练过程中出现的梯度消失和爆炸的问题。如图 1 所示,LSTM 模型通过输入门、遗忘门、输出门控制记忆单元的信息输入和输出^[15-16],计算过程如式 (2)~(9) 所示。

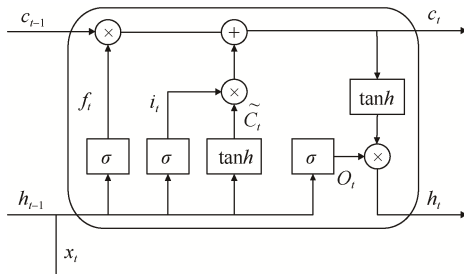


图 1 LSTM 网络结构

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (2)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (3)$$

$$C_t = C_{t-1} \otimes f_t + i_t \otimes \tilde{C}_t \quad (4)$$

$$\tilde{C}_t = \tan h(W_c[h_{t-1}, x_t] + b_c) \quad (5)$$

$$O_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = O_t \otimes \tan h(C_t) \quad (7)$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (8)$$

$$\tan h(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (9)$$

式中: f_t 为遗忘门; i_t 为输入门; O_t 为输出门; C_t 为细胞状态; h_t 为输出值; W 为权重值; b 为偏置; \otimes 表示逐点相乘。

2.2 齿轮箱轴承温度预测模型

与传统的神经网络相似,预测模型的网络结构由输入层、隐含层、输出层构成,模型的输入

为经过互信息法筛选的特征参数,模型的输出为齿轮箱轴承温度,通过设定合适的时间步长,深度提取前一时刻与当前时刻运行数据的特征信息。为了防止模型出现过拟合现象,需要在隐含层后加入 dropout 层,按照一定比例放弃隐含层输出值。根据模型的评价标准,经过多次实验来最终确定模型的隐含层神经元个数,迭代次数 (epochs) 和每次处理样本数量 (batch size) 等参数。

3 模型预测残差滑动窗口分析

选取机组正常状态下的运行数据训练 LSTM 模型,当齿轮箱轴承处于正常工作状态时,机组的运行数据相对比较平稳,LSTM 模型的预测误差相对较小。当轴承出现潜在故障时,随着运行时间的增加,故障程度也会加剧,轴承温度特性会偏离正常工作区间,从而导致 LSTM 模型的预测误差变大。为了避免误报警,连续地分析预测残差特性,本文采用滑动窗口分析方法对预测残差进行处理,可以有效地消除随机因素对残差变化的影响。

取一个宽度为 N 的滑动窗口,计算该窗口下所包含残差的均值和标准差,公式如下:

$$\begin{cases} \bar{X} = \frac{1}{N} \sum_{i=1}^N e_i \\ S = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (e_i - \bar{X})^2} \end{cases} \quad (10)$$

式中: e_i 为窗口中的第 i 个残差。

利用滑动窗口对训练集的残差进行分析处理,计算统计得到均值绝对值的最大值和标准差的最大值,记为 \bar{X}_{\max} 和 S_{\max} ,设定故障报警阈值:

$$\begin{cases} \bar{X}_T = \pm k_1 \bar{X}_{\max} \\ S_T = k_2 S_{\max} \end{cases} \quad (11)$$

式中通常取 k_1 、 k_2 为 2,当均值和标准差都超过报警阈值时发出故障报警信号^[17]。

4 仿真验证

4.1 模型参数优化

选取正常工作状态下的数据作为模型的训练集,共 12 800 条。选取故障报警时刻前一段时间

内的数据作为模型的测试集,共5000条。将数据进行归一化处理,建立长短期记忆神经网络预测模型,模型包含输入层、单隐含层和输出层,设置输入层神经元为10个,输出层神经元为1个,模型输出为齿轮箱轴承温度值。初步设定模型参数,如隐含层神经元个数、激活函数、迭代次数、每次处理样本数量、dropout系数、优化方法等,采用平均绝对误差(MAE)和均方根误差(RMSE)作为模型的评价标准^[18],优化模型参数使预测效果达到最优。

$$\begin{cases} MAE = \frac{1}{n} \sum_{i=1}^n |y(i) - y'(i)| \\ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y(i) - y'(i))^2} \end{cases} \quad (12)$$

式中: n 为采样点总数; $y(i)$ 为第 i 个采样点的实际温度值; $y'(i)$ 为第 i 个采样点的模型输出值。

表1截取了一部分参数优化的结果。从表中可以看出,当其它参数不变时,在一定范围内增加隐含层神经元个数,模型效果会有所提升。随着迭代次数的增加,模型效果也会有一定程度的改善,但是程序运行时间会变长。综合多方面因素最终确定隐含层神经元个数为65个,迭代次数为550,每次处理样本数量为60。

表1 不同参数下的模型预测效果

隐含层 神经元	epochs	batch size	训练集		测试集		运行 时间/s
			MAE	RMSE	MAE	RMSE	
10	100	20	2.185	2.596	6.295	8.164	117.44
20	100	20	1.782	2.088	4.740	6.066	122.96
30	100	20	1.526	1.827	4.515	5.935	128.25
40	200	40	1.097	1.284	4.039	5.785	148.43
50	200	40	0.886	1.072	3.606	5.263	180.36
60	300	40	0.690	0.855	3.564	5.455	269.20
70	300	60	0.813	0.977	3.463	5.187	201.93
80	400	60	0.998	1.142	3.352	4.854	283.59
80	500	60	0.653	0.801	3.261	5.156	350.13
80	600	70	0.745	0.917	3.581	5.516	395.05

4.2 正常状态下模型预测结果分析

随机选取训练集数据的25%作为验证集,设置模型损失值为MAE,得到训练集和验证集的损失曲线,如图2所示:

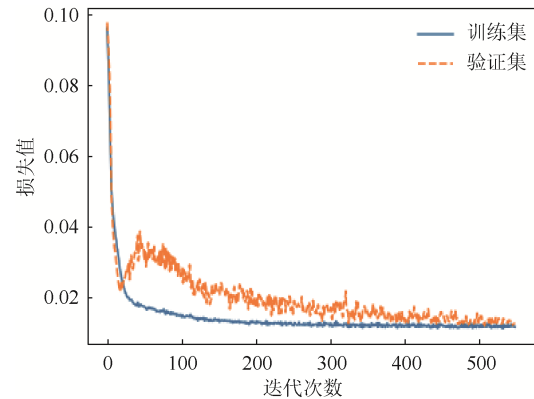


图2 LSTM模型损失曲线

从图中可以看出,训练集的收敛速度要快于验证集。随着迭代次数的增加,模型损失值不断减小,最终趋于稳定。证明网络结构设计得当,模型参数设置合理。训练集预测结果及残差变化趋势如图3、图4所示。

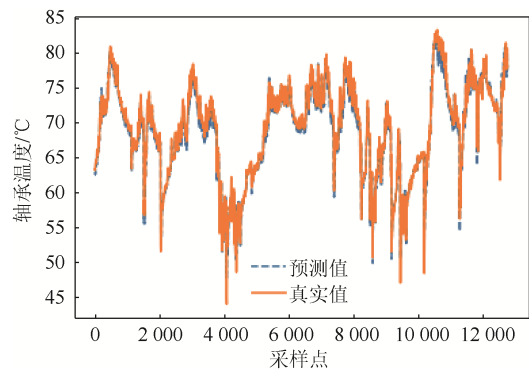


图3 训练集模型预测值与实际温度值

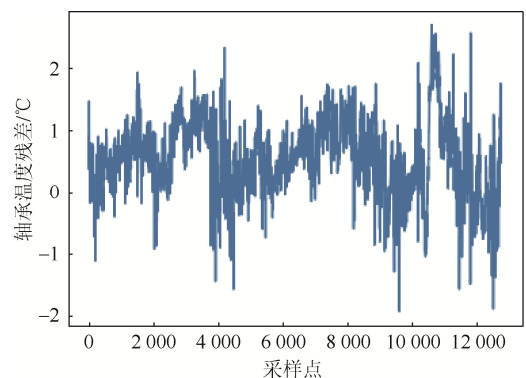


图4 训练集残差

从图中可以看出训练集预测效果较好,实际温度值和预测值的差值大部分集中在-1至1之间。确定滑动窗口宽度为150,对残差进行分析处理,计算窗口内的残差均值和标准差。随着窗口的移动,变化趋势如图5、图6所示。

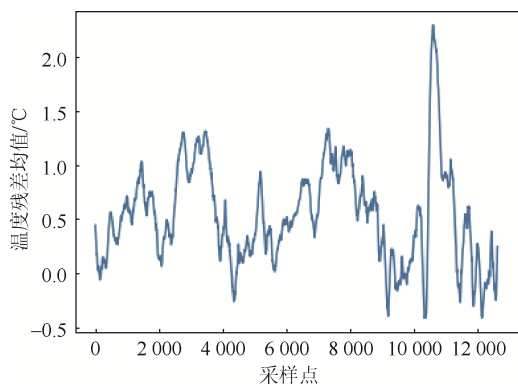


图5 残差均值变化趋势

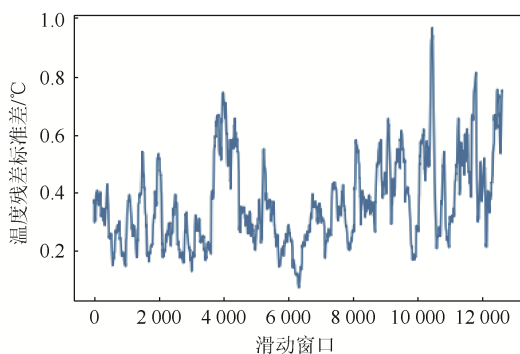


图6 残差标准差变化趋势

从图中可以看出,风机正常状态下,模型预测残差的均值处于 $-0.408 \sim 2.292$ 之间,绝对值最大为 2.292。残差的标准差处于 0.072 到 0.966 之间,最大值为 0.990。根据式(8)设定均值的报警阈值为 ± 4.584 ;标准差的报警阈值为 1.932。

4.3 异常状态下模型预测结果分析

将测试集数据输入到 LSTM 模型中,验证模型预警的有效性。测试集预测结果及残差变化趋势如图 7、图 8 所示。

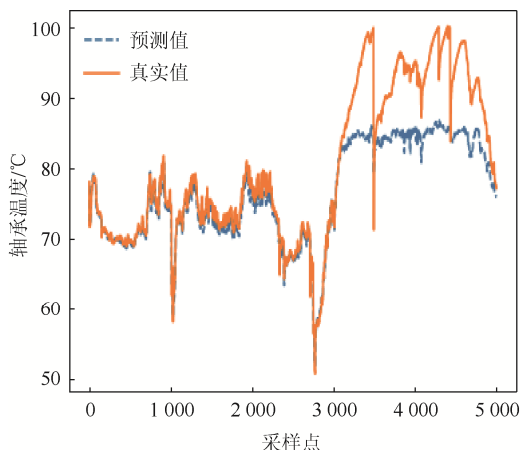


图7 测试集模型预测值及实际温度值

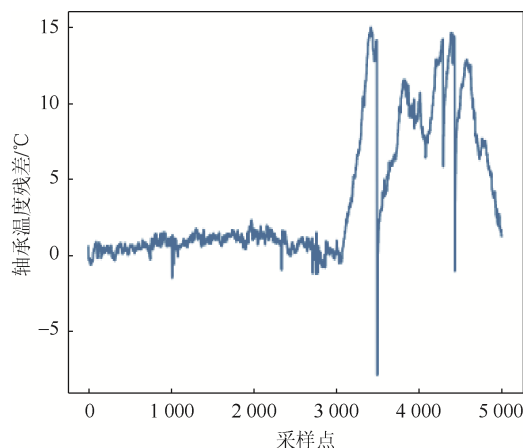


图8 测试集残差

模型前半段预测误差较小,相对比较稳定。大约从第 3 000 点开始,残差逐渐增大,并在一定范围内波动,齿轮箱轴承可能出现了潜在故障。仍然取滑动窗口宽度为 150,分析测试集残差的均值和标准差变化情况,如图 9、图 10 所示。

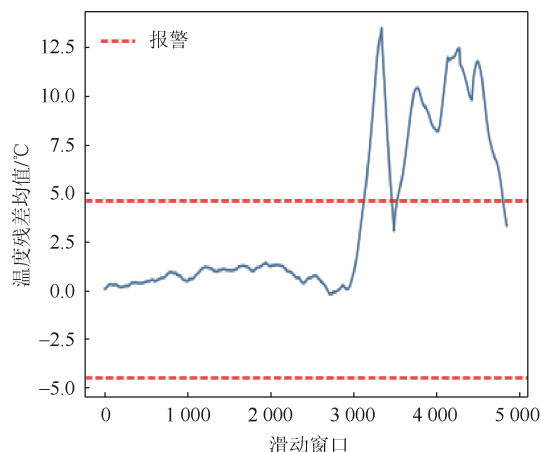


图9 残差均值变化情况

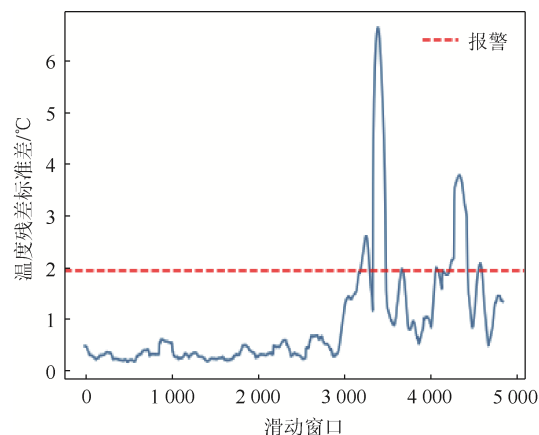


图10 残差标准差变化情况

观察图9可以看出,残差均值从第3130个滑动窗口开始超过均值的报警阈值,在第3463个窗口时回落到报警线以下,但是到第3531个窗口时又超过了报警线,之后处于报警线以上。观察图10发现标准差在第3197个滑动窗口时超过了标准差的报警阈值。该风机实际报警信息是第5000个采样点发出的,显示风机齿轮箱轴承故障。根据本文方法,当均值和标准差同时超过报警阈值时发出报警信号,所以在第3197个滑动窗口,即第3347个采样点时就会做出报警,报警时间明显提前(大约提前27.3h),避免了齿轮箱轴承故障进一步加重。

5 结论

本文基于风电场运行数据提出一种风机齿轮箱轴承故障预警方法,经过仿真验证,得出以下结论:

1)互信息法对预处理后的数据能够有效地进行特征选择,筛选出与目标对象(即齿轮箱轴承温度)关联度高的特征信息。

2)长短期记忆神经网络模型能够对运行数据时间前后的关系进行深度特征提取,并且基于关联变量对齿轮箱轴承温度进行准确有效地预测。

3)通过滑动窗口对预测残差的分析处理,可以设定合适的报警阈值及报警规则对风机齿轮箱轴承故障进行提前预警,避免故障进一步加重。

参考文献:

- [1] 林涛,刘刚,蔡睿琪,等.基于轴承温度的风机齿轮箱故障预警研究[J].可再生能源,2018,36(12):1877-1882.
- [2] 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):45-51.
- [3] 赵洪山,刘辉海.基于深度学习网络的风电机组主轴承故障检测[J].太阳能学报,2018,39(3):588-595.
- [4] 丁佳煜,许昌,葛立超,等.基于轴承温度模型的风电机组故障预测研究[J].可再生能源,2018,36(2):276-282.
- [5] 刘帅,刘长良,曾华清.基于核极限学习机的风电机组齿轮箱故障预警研究[J].中国测试,2019,45(2):121-127.
- [6] ZHAO H, LIU H, HU W, et al. Anomaly detection and fault analysis of wind turbine components based on deep learning network[J]. Renewable Energy, 2018, 127(11): 825-834.
- [7] WU Q H, YU D, LI M S, et al. A radically data-driven method for fault detection and diagnosis in wind turbines[J]. International Journal of Electrical Power and Energy Systems, 2018, 99(7): 577-584.
- [8] SENANAYAKA J S L, VAN KHANG H, ROBBERSMYR K G. Autoencoders and recurrent neural networks based algorithm for prognosis of bearing life[C]// 21st International Conference on Electrical Machines and Systems (ICEMS). IEEE, 2018: 537-542.
- [9] BODLA M K, MALIK S M, RASHEED M T, et al. Logistic regression and feature extraction based fault diagnosis of main bearing of wind turbines[C]// IEEE 11th Conference on Industrial Electronics and Applications (ICIEA). IEEE, 2016: 1628-1633.
- [10] 卢锦玲,绳菲菲,赵洪山.基于相关向量机的风机齿轮箱故障诊断方法[J].华北电力大学学报(自然科学版),2017,44(2):91-96.
- [11] 李永亨,齐咏生,杨苗,等.一种W型结构元素的自适应数学形态学及其在风电机组轴承故障诊断中的应用[J].太阳能学报,2020,41(1):192-200.
- [12] 李东东,王浩,杨帆,等.基于一维卷积神经网络和Soft-Max分类器的风电机组行星齿轮箱故障检测[J].电机与控制应用,2018,45(6):80-87.
- [13] CHENG F, PENG Y, QU L, et al. Current-based fault detection and identification for wind turbine drivetrain gearboxes[J]. IEEE Transactions on Industry Applications, 2016, 53(2): 878-887.
- [14] 李大中,张华英.基于风场大数据互信息关联的风电机组性能评估[J].科学技术与工程,2018,18(29):211-215.
- [15] LIU H, MI X, LI Y. Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network[J]. Energy conversion and management, 2018, 156(1): 498-514.
- [16] 曾慧洁,郭建胜.双向LSTM神经网络的航空发动机故障预测[J].空军工程大学学报(自然科学版),2019,20(4):26-32.
- [17] 许清.基于运行数据的风电机组故障诊断的研究[D].北京:华北电力大学,2019.
- [18] 王桂松,郭鹏,胥佳,等.基于XGBoost建模和Change-Point残差处理的风电机组齿轮箱温度预警[J].电力科学与工程,2018,34(9):46-53.