文章编号:1674-7607(2026)02-0146-11 DOI:10.19805/j.cnki.jcspe.2026.240732

基于神经基拓展分析的风电机组齿轮箱故障预警

魏 乐, 舒孝海, 陈远野, 房 方
(华北电力大学 控制与计算机工程学院, 北京 102206)

摘 要: 为解决现有齿轮箱故障预警方法中特征冗余、预测精度低、泛化能力弱及误报率高的问题,提出了一种基于神经基拓展分析的预警方法。该方法将Sigmoid函数曲线与基于密度的空间聚类算法相结合,从原始数据中提取正常监测数据;通过改进的灰色关联度算法为不同数据点赋予权重,深度挖掘变量信息,从高维监控与数据采集系统采集的数据中提取与齿轮箱油温相关的特征变量;利用局部非线性投影将目标信号分解为基函数,实现对齿轮箱状态变量的高精度预测;通过滑动窗口和置信区间设置故障阈值,减少误报率。最后,基于实际风场数据进行验证。结果表明:在齿轮箱状态预测中基于NBEATSx模型的故障预警方法显著提升了预测精度,较传统模型降低了误报率,具备提前数小时预知故障的能力,从而有效保障了风电机组的稳定运行。

关键词: 风电机组; 故障预警; 齿轮箱; 神经及拓展分析; 特征变量
中图分类号: TK83 **文献标志码:** A **学科分类号:** 470.30

Research on Wind Turbine Gearbox Fault Warning Method Based on Neural Basis Expansion Analysis

WEI Le, *SHU Xiaohai*, *CHEN Yuanye*, *FANG Fang*
(School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China)

Abstract: To address issues of feature redundancy, low prediction accuracy, weak generalization ability, and high false alarm rates in existing gearbox fault warning methods, a fault warning method based on neural basis expansion analysis was proposed. The method combined Sigmoid growth curve with a density-based spatial clustering algorithm to extract normal monitoring data from raw data. An improved grey relational analysis algorithm was used to assign weights to different data points, variable information was deeply explored. Feature variables related to the oil temperature of the gearbox were extracted from the data collected by the high-dimensional monitoring and data acquisition system. Local nonlinear projection was used to decompose the target signal into basis functions to predict the state variables of the gearbox accurately, fault thresholds were set by sliding window and confidence interval to reduce the false alarm rate. The method was verified based on actual wind field data. Results show that the fault prediction method based on the NBEATSx model significantly improves the prediction accuracy in the state prediction of gearboxes, reduces the false alarm rate compared to traditional models, and has the ability to predict faults several hours in advance, thereby effectively ensuring the stable operation of wind turbine units.

收稿日期: 2024-11-21 **修订日期:** 2025-01-13
基金项目: 华能集团总部科技资助项目(HNKJ29-H88);海上风电与智慧能源系统科技专项(一期)(HNKJ20-H88)
作者简介: 魏 乐(1976—),女,北京人,教授,博士,研究方向为风电机组状态监测与故障预警。
陈远野(通信作者),男,讲师,博士,E-mail:yuanyechen@ncepu.edu.cn。

Key words: wind turbine; fault prediction; gearbox; neural basis expansion analysis; feature extraction

在实现“碳达峰”、“碳中和”目标的大背景下，风力发电占据了能源转型的关键地位[1]。截至 2023 年底，全球陆上风电机组总装机容量为 945.5 GW，其中 2023 年新增装机容量为 105.8 GW，预计全球风电机组装机容量将会以每年 9.4%的速度稳步上升[2]。因此，提升风电机组的可靠性，保证风电机组的高效运行显得尤为重要。

齿轮箱是风电机组中的关键机械部件。在复杂运行环境下，齿轮箱长期高负荷和变速运转易导致其部件磨损、疲劳损伤和润滑失效，因而故障率较高。据统计，齿轮箱故障维修费用占风机总维修费用的 18.61%，由齿轮箱故障导致的停机时间占总故障停机时间的 21.00%[3]，因此，对齿轮箱进行故障预警是保障风电机组稳定运行的重要手段[4]。

当前大多采用监控与数据采集(SCADA)系统对风电机组进行状态监测。针对 SCADA 采集的多维时间序列数据，研究者们提出了多种基于时间序列数据的故障预警方法。

在数据预处理方面，Cheng 等[5]提出增强粒子滤波算法，解决了粒子贫化问题，提高了高维数据处理能力。Wang 等[6]利用灰色关联分析与鲸鱼优化算法构建了混合模型，通过筛选相关变量提升了模型的预测精度。雷萌等[7]设计了预测误差和分类准确度指标，实现了全自动选取基于密度的空间聚类(DBSCAN)算法的聚类参数，建立了风电机组 SCADA 异常数据识别方法。

在特征变量提取方面，Bhardwaj 等[8]利用最小二乘支持向量机结合统计过程控制技术，提出了一种特征变量提取方法，提升了基于可靠性影响因素分析的故障预警模型的精度。黄忠山等[9]将主成分分析与统计过程控制相结合，开发出动态神经网络模型，用于长期油温预测。

在状态预测方面，He 等[10]基于自校准时间卷积网络的研究提出了一种新颖的轴承退化趋势模拟方法。苏向敬等[11]通过引入门控循环单元和注意力机制提升了风电机组状态监测的精度和鲁棒性。张书瑶等[12]提出改进实例学习算法来提升风电机组齿轮箱状态监测精度。景彤梅等[13]利用结合了核熵成分分析与广义回归神经网络的三阶段方法，实现了齿轮箱的健康状态评估。而基于 Transformer[14]的模型如 Patch-TST[15]模型和 iTransformer[16]模型则在处理长序列方面表现出色。而引入外部特性 x 的神经基拓展分析(NBEATSx)模型[17-18]通过将时间序列分解为多项式趋势项和谐波周期项，并结合特征变量的影响，能够更准确地捕捉齿轮箱状态的变化规律。

在故障预警方面，雍彬等[19]引入指数加权移动平均方法对残差进行平滑处理，通过控制图计算出残差值的报警阈值，实现齿轮箱故障预警。Fu 等[20]利用格拉姆角场结合卷积神经网络建立主要部件预警模型，并经过时间生成对抗网络增强后输入视觉变换器，建立故障代码类型预警模型，进而确定预警时间点。Wang 等[21]建立非线性状态估计技术模型，通过计算验证集残差确定阈值，当测试集齿轮箱温度残差超阈值时实现预警故障。向玲等[22]提出基于卷积自编码与双向长短期记忆网络的预警方法，解决了风电机组早期故障监测的难题。罗毅等[23]提出了一种基于改进鲸鱼优化算法优化 Transformer 网络超参数的磨煤机故障预警方法，从而显著提升了故障预警的速度和准确性。

综上所述，本文针对现有风电机组齿轮箱故障预警方法中存在的特征冗余、预测精度低、泛化能力弱等问题，提出一种基于 NBEATSx 模型的创新方法。通过将 Sigmoid 函数曲线与 DBSCAN 算法相结合，剔除 SCADA 数据中的异常点，提取正常运行数据。采用改进的灰色关联分析赋予数据权重，提取关键特征。基于 NBEATSx 模型，对齿轮箱温度进行预测，分解时间序列并结合特征变量，更精准地捕捉状态变化。通过滑动窗口和置信区间设置动态故障阈值，并将方法应用到实际风机数据上进行验证，从而实现齿轮箱的实时监测和故障预警，提高预测精度。

1　风电机组齿轮箱故障预警问题

1.1　风电机组齿轮箱故障预警的问题分析

齿轮箱故障主要可以分为 4 类：漏油、制动系统以及材料的疲劳、轴承故障、风暴造成的故障和风险情况。齿轮箱故障类型及其比例如图 1 所示。可见，漏油和齿轮箱轴承故障的占比相对较高，分别为 50%和 18.8%[24]。这些故障通常会导致齿轮箱内部润滑不足或部件磨损加剧，进而导致齿轮箱油温超出正常范围。监测齿轮箱的油温变化，并对超温进行预警，可以有效减少齿轮箱漏油和轴承故障所带来的维修费用和停机时间。

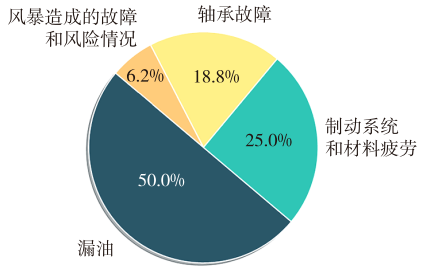


图 1 齿轮箱故障类型及其比例

Fig. 1 Types of gearbox failures and their proportions

为使研究结果更适用于工程实际，本文选取河北某风电场发生过齿轮箱故障的 2 台实际运行风机(案例 1 和案例 2)展开研究。风机参数选取时段为 2018 年 1 月 1 日至 2018 年 12 月 31 日，具体参数如表 1 所示。

表 1 风机参数

Tab. 1 Wind turbine parameter

参数	数值
切入风速/(m·s^{-1})	3
切出风速/(m·s^{-1})	20
采样周期/min	10
样本数量	51 024
测点个数	25
额定功率/kW	1 500
额定转速/(r·min^{-1})	17.4

由于受环境、测量和传输等多方面因素影响，SCADA 数据通常包含噪声和异常值，因此需要通过有效的数据预处理方法来消除噪声，并确保模型对异常信号敏感，以便及时检测潜在故障。

此外，SCADA 数据具有高维度特性，而与目标变量相关性较低甚至负相关的变量会干扰到模型的训练，从而导致预测精度降低。因此必须采用特征提取方法进行降维，以保留与目标变量相关性高的关键特征。

风电机组齿轮箱油温数据具有非线性趋势性、周期性和季节性波动的特征，这要求预测模型能够处理复杂的多维变量关联与非线性趋势。同时，预测模型不仅需在单台风机上表现出高精度，同时应具备足够的泛化能力以适应不同风机的运行特性，从而确保其在风场多台风机上的预测准确性。

由于齿轮箱油温不会发生跳变，在基于正常运行数据建立预测模型后需对预测值和真实值的偏差进行适当处理，以防出现误报。通过设定合理的故障预警阈值，在齿轮箱状态变量超过阈值时报警，有效减少了漏报和误报情况，确保故障预警的准确性和可靠性。

1.2 风电机组齿轮箱故障预警的解决方案

提出了一种 NBEATSx 的故障预警方法。该方法结合了多种数据处理和预测技术，旨在解决风电机组运行环境下的复杂数据处理问题，并提供更精确的故障预警方案。具体来说，该方法从数据预处理、特征变量提取、预测模型选取以及预警阈值设置四个方面进行系统分析和设计，最终形成一套完整的故障预警流程。风机齿轮箱故障预警流程图如图 2 所示。

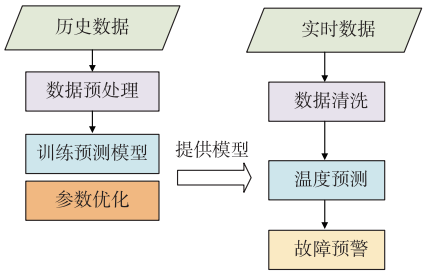


图 2 风机齿轮箱故障预警流程图

Fig. 2 Fault warning flowchart for wind turbine gearbox

1.2.1 数据预处理

考虑到风速功率散点图的形状与 Sigmoid 函数曲线的相似性，在应用 DBSCAN 算法剔除风机故障点和限功率的数据时，引入 Sigmoid 函数进行数据拟合，进而更准确地聚类高密度正常发电数据，从而提升数据清洗的准确性和完整性。

1.2.2 特征变量提取

在特征提取过程中，为避免过度依赖原始变量而导致无法捕捉到更复杂变量间的关联，本文改进灰色关联度分析，通过对每个数据点赋予不同权重来更准确地提取与目标变量相关的特征变量，从而增强模型处理高维数据的有效性。

1.2.3 预测模型选取

针对风电机组齿轮箱油温数据的时序特性，基于 NBEATSx 模型建立齿轮箱油温正常工况预测模型。该模型能够将油温数据中的多项式趋势项、谐波周期项以及外部变量的相关性分解整合到预测过程中，以适应油温数据的多层次结构，从而提升故障预测的精度。为进一步优化模型的预测性能，采用基于树结构的 Parzen 估计器(TPE)对 NBEATSx 模型参数进行优化。该算法通过构建条件概率分布模型，灵活调整目标函数的搜索空间，可在提高模型精度的同时降低计算复杂度。

1.2.4 预警阈值设置

通过构建滑动窗口模型，对预测误差进行平滑

处理，以避免瞬间突变量造成的误报警；结合置信区间估计模式设定故障阈值，以有效减少阈值不合理造成的漏报警风险，从而实现更加稳定的故障预警。

2　风机数据预处理与特征变量提取

2.1　齿轮箱数据预处理

风电机组 SCADA 数据中存在由于停机、限功率以及传感器自身误差等原因导致的异常数据。这些异常数据会影响风电机组齿轮箱状态预测及预警的结果。所以，在建立齿轮箱温度预测模型前有必要对原始 SCADA 数据进行预处理。而风电机组主要关注输入风速与有功功率之间的关系，故需对风机风速-功率进行数据预处理。详细过程包括数据清洗、密度聚类、曲线拟合及数据筛选 4 个关键步骤。案例 1 风电机组 SCADA 数据与处理流程如图 3 所示。

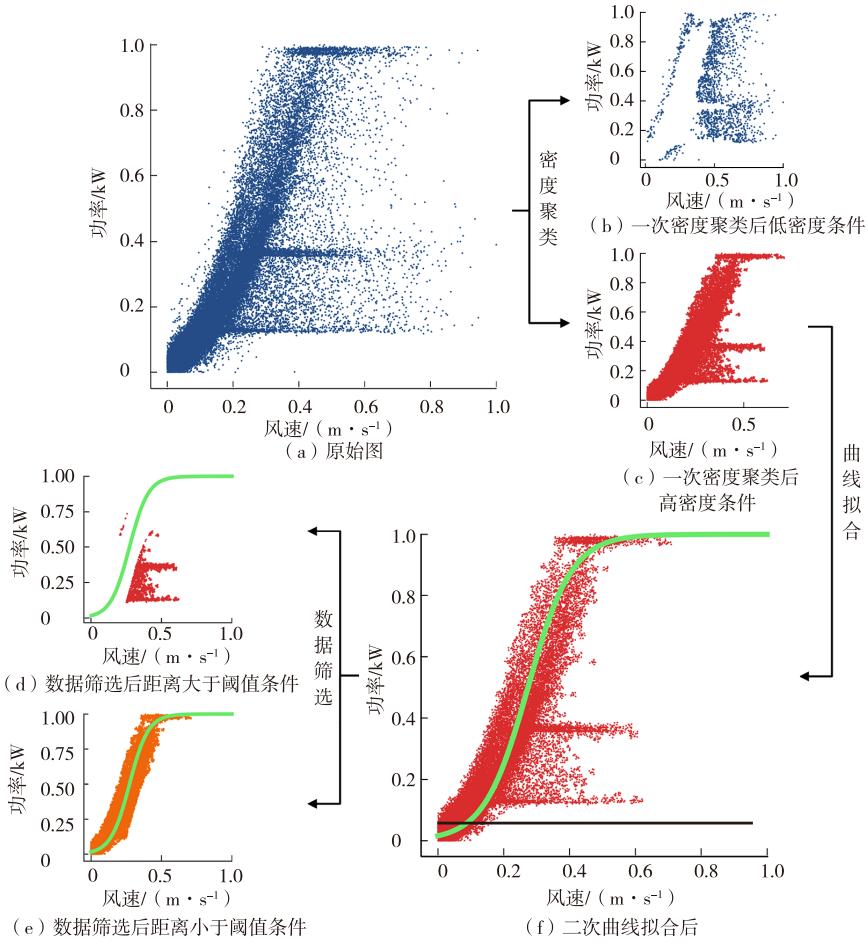


图 3　案例 1 风电机组 SCADA 数据与处理流程

Fig. 3　Flowchart of SCADA data and processing for wind turbine in case 1

2.1.1　数据清洗

对案例 1 风电机组 SCADA 数据中明显异常的数据（风速低于切入风速、高于切出风速或功率小于等于 0 kW）进行剔除，得到 33 592×25 组数据。考虑到不同量纲的影响，对风速和功率进行归一化处理，得到如图 3(a)所示的原始风速-功率散点图。

2.1.2　密度聚类

由于散点图具有明显的密度特性，基于 DBSCAN 算法对风速-功率散点图进行一次聚类，将散点图分为密度较高的簇和密度较低的簇，得到一次聚类效果，筛选出 31 978×25 组高密度数据。从图 3 可以看出：利用 DBSCAN 算法无法剔除密度较高的限功率数据，且散点图整体趋势与 Sigmoid 函数曲线相似。

2.1.3　曲线拟合

根据数据的特性，利用改进的 Sigmoid 函数[25]对风机风速-功率散点图进行初步拟合，即

$$P=\mathrm{Sigmoid}(v)=\frac{a}{b+\mathrm{e}^{-(dv-c)}} \tag{1}$$

式中：a、b、c、d 均为参数；v 为风速；P 为功率。

案例 1 和案例 2 的参数如表 2 示。

2.1.4　数据筛选

表 3 为不同哈曼顿距离下正常发电数据与限功

表 2 改进 Sigmoid 函数拟合的风机参数

Tab. 2 Improved Sigmoid function fitting for wind turbine parameters

风机	a	b	c	d
案例 1	0.90	0.90	4.00	15.00
案例 2	0.95	0.90	4.50	14.00

表 3 不同哈曼顿距离阈值的分类情况

Tab. 3 Classification based on different hamming distance thresholds

阈值	正常发电数据样本数	限功率数据样本数
0	0	31 978
0.05	15 591	16 387
0.10	22 382	9 596
0.15	26 048	5 930
0.20	28 242	3 736
0.25	29 295	2 683
0.30	29 834	2 144
0.35	30 169	1 809
0.40	30 394	1 584
0.45	30 658	1 320
0.50	30 893	1 085
0.55	31 163	815
0.60	31 426	552
0.65	31 660	318
0.70	31 744	234
0.75	31 810	168
0.80	31 864	114
0.85	31 941	37

率数据样本数的对比。查阅实际风场中的限功率数据可知，案例 1 中有 2 321 组限功率数据，而在密度聚类中已剔除部分密度较低的限功率数据。为确保在筛选出正常发电数据的情况下能有效剔除限功率数据，且当最小哈曼顿距离为 0.3 时，筛选出的限功率数据个数最接近于实际限功率数据个数，选择 Sigmoid 函数拟合曲线最小哈曼顿距离小于 0.3 的核心点所属簇作为正常发电数据，其余点属于限功率数据。共筛选出 29 834×25 组正常发电数据。对案例 2 风电机组进行相同的数据筛选处理，得到 31 042×25 组发电数据。

2.2 齿轮箱油温特征变量提取

历史样本点对当前样本点的影响会随着时间的推移逐渐降低，故采用时间距离分析法计算每个数据点 i 的权重 $w(i)$，以更好地反映数据的相关性。

$$c(i)=\frac{n_{\mathrm{d}}-i}{n_{\mathrm{d}}} \tag{2}$$

$$w(i)=\frac{c(i)}{\sum_{i=1}^{n_{\mathrm{d}}}c(i)} \tag{3}$$

式中：$c(i)$为样本点与参考样本点的接近度；n_{d} 为清洗后数据的样本数。

采用改进灰色关联度分析(IGCA)[26]提取特征变量。该方法根据不同数据点赋予不同权重，从而展示状态变量之间的动态关联程度，且不会出现量化结果和定性分析结果不同的情况。灰色关联度系数 $\zeta(\boldsymbol{Y}(i),\ \boldsymbol{x}_k(i))$和灰色关联度水平 $\eta(\boldsymbol{Y},\ \boldsymbol{x}_k)$分别为

$$\zeta(\boldsymbol{Y}(i),\ \boldsymbol{x}_k(i))=\frac{\min\limits_{1\leqslant k\leqslant m_{\mathrm{d}}}\min\limits_{1\leqslant i\leqslant n_{\mathrm{d}}}|\boldsymbol{Y}(i)-\boldsymbol{x}_k(i)|+\delta\max\limits_{1\leqslant k\leqslant m_{\mathrm{d}}}\max\limits_{1\leqslant i\leqslant n_{\mathrm{d}}}|\boldsymbol{Y}(i)-\boldsymbol{x}_k(i)|}{|\boldsymbol{Y}(i)-\boldsymbol{x}_k(i)|+\delta\max\limits_{1\leqslant k\leqslant m_{\mathrm{d}}}\max\limits_{1\leqslant i\leqslant n_{\mathrm{d}}}|\boldsymbol{Y}(i)-\boldsymbol{x}_k(i)|} \tag{4}$$

$$\eta(\boldsymbol{Y},\ \boldsymbol{x}_k)=\sum_{k=1}^{m_{\mathrm{d}}}w(i)\zeta(\boldsymbol{Y}(i),\ \boldsymbol{x}(i)) \tag{5}$$

式中：m_{d} 为清洗后数据的测点样本数；$\boldsymbol{Y}$ 为齿轮箱油温和特征变量进行数据归一化处理后的值；$\boldsymbol{x}_k$ 为 $\boldsymbol{X}$ 中第 k 个特征值向量；$\boldsymbol{Y}(i)$和 $\boldsymbol{x}_k(i)$分别为 $\boldsymbol{Y}$ 和 $\boldsymbol{x}_k$ 中第 i 个数据；$\delta\in(0,1)$为分辨率系数。

通过 IGCA 方法提取与目标变量相关的 2 个特征变量。表 4 按照灰色关联度从大到小的顺序列出了齿轮箱油温与特征变量的灰色关联度。齿轮箱轴 1 温度测点与齿轮箱轴 2 温度测点由于在物理结构上与齿轮箱油温测点接近，因此这 2 个变量较其他变量更能有效反映齿轮箱油温的变化。因此，最终选择齿轮箱轴 1 温度和齿轮箱轴 2 温度作为 NBEATSx 模型的输入特征变量。最后得到清洗后案例 1 输入数据为 29 834×3 组，对案例 2 进行相同处理，得到输入数据为 31 042×3 组。

表 4 齿轮箱油温与特征变量灰色关联度

Tab. 4 Gray correlation between gearbox oil temperature and characteristic variables

测点	灰色关联度	
	案例 1	案例 2
齿轮箱轴 1 温度	0.99	0.99
齿轮箱轴 2 温度	0.98	0.98
定子绕组 1 温度	0.98	0.98
电网频率	0.98	0.98
电机轴承 A 温度	0.95	0.96
电机轴承 B 温度	0.95	0.96
⋮	⋮	⋮
定子绕组 3 电压	0.48	0.94
定子绕组 2 电压	0.47	0.94
定子绕组 1 电压	0.47	0.94

3　风电机组齿轮箱故障预警模型

3.1　风电机组齿轮箱故障预警模型的建立

将预处理后的风电机组齿轮箱数据按 7∶3的比例划分为训练集和测试集。后向预测向量的长度 L 为 144 h，而预测时间范围 H 为 24 h。基于 NBEATSx 模型建立齿轮箱温度预测模型，如图 4 所示。图 4 中：$\boldsymbol{\theta}_{n,m}^{\text{for}}$ 为前向膨胀系数；$\boldsymbol{\theta}_{n,m}^{\text{back}}$ 为后向膨胀系数；$\hat{\boldsymbol{y}}^{\text{for}}$ 为模型前向预测值；$\hat{\boldsymbol{y}}_n^{\text{back}}$ 为模型后向输出值；$\hat{\boldsymbol{y}}_{n,m}^{\text{for}}$ 为第 n 个堆栈中第 m 个块前向输出值；$\hat{\boldsymbol{y}}_{n,m}^{\text{back}}$ 为第 n 个堆栈中第 m 个块的后向输出值；$\boldsymbol{y}_{n,m-1}^{\text{back}}$ 为第 n 个堆栈第 $m-1$ 个块的后向输出值；$\boldsymbol{y}^{\text{back}}$ 为模型后向输出值；$\boldsymbol{y}_n^{\text{back}}$ 为第 n 个堆栈的后向输出值；$\boldsymbol{x}_{n,m-1}$ 为第 n 个堆栈第 $m-1$ 个块的输入值；$\boldsymbol{x}$ 为输入值。NBEATSx 模型由 s 个堆栈组成，每个堆栈专门处理某类数据特性。

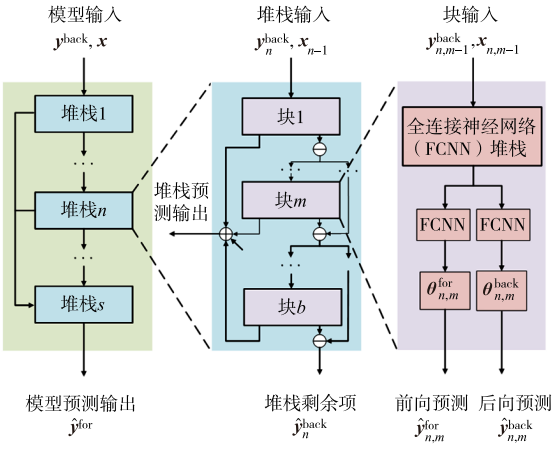


图 4　齿轮箱油温预测模型结构图

Fig. 4　Structure diagram of gearbox oil temperature prediction model

每个堆栈包含 b 个块，第 n 个堆栈第 m 个块的输入是来自于上一个块的输出 $\boldsymbol{y}_{n,m-1}^{\text{back}}$ 和其他协变量 $\boldsymbol{x}_{n,m-1}$，每个块包含一个 FCNN。

$$\boldsymbol{h}_{n,m}=F_{n,m}(\boldsymbol{y}_{n,m-1}^{\text{back}},\ \boldsymbol{x}_{n,m-1}) \tag{6}$$

式中：$\boldsymbol{h}_{n,m}$ 为特征映射隐藏向量；$F_{n,m}$ 为第 n 个堆栈第 m 个块中的全连接神经网络。

而前向膨胀系数 $\boldsymbol{\theta}_{n,m}^{\text{for}}\in\mathbf{R}^{N_s}$ 和后向膨胀系数 $\boldsymbol{\theta}_{n,m}^{\text{back}}\in\mathbf{R}^{N_s}$ 则是对特征映射隐藏向量 $\boldsymbol{h}_{n,m}\in\mathbf{R}^{N_h}$ 的一个线性映射。

$$\boldsymbol{\theta}_{n,m}^{\text{for}}=L^{\text{for}}(\boldsymbol{h}_{n,m}) \tag{7}$$

$$\boldsymbol{\theta}_{n,m}^{\text{back}}=L^{\text{back}}(\boldsymbol{h}_{n,m}) \tag{8}$$

式中：L^{for}、L^{back} 分别为前向和后向的线性映射关系；N_s 和 N_h 分别为膨胀系数和特征映射隐藏向量的维数。

通过前向和后向基向量 $\boldsymbol{V}_{n,m}^{\text{for}}\in\mathbf{R}^{H\times N_s}$ 和 $\boldsymbol{V}_{n,m}^{\text{back}}\in\mathbf{R}^{L\times N_s}$ 将 $\boldsymbol{\theta}_{n,m}^{\text{for}}$ 和 $\boldsymbol{\theta}_{n,m}^{\text{back}}$ 转换成 $\hat{\boldsymbol{y}}_{n,m}^{\text{for}}$ 和 $\hat{\boldsymbol{y}}_{n,m}^{\text{back}}$；$\hat{\boldsymbol{y}}_{n,m}^{\text{back}}$ 为上一块后向预测结果的期望值，来自于堆栈输入的迭代值，即

$$\hat{\boldsymbol{y}}_{n,m}^{\text{for}}=\boldsymbol{V}_{n,m}^{\text{for}}\times\boldsymbol{\theta}_{n,m}^{\text{for}} \tag{9}$$

$$\hat{\boldsymbol{y}}_{n,m}^{\text{back}}=\boldsymbol{V}_{n,m}^{\text{back}}\times\boldsymbol{\theta}_{n,m}^{\text{back}} \tag{10}$$

第 n 个堆栈的输出 $\hat{\boldsymbol{y}}_n^{\text{for}}$ 等于各个块前向预测 $\hat{\boldsymbol{y}}_{n,i}^{\text{for}}$ 的和，即

$$\hat{\boldsymbol{y}}_n^{\text{for}}=\sum_{i=1}^{m}\hat{\boldsymbol{y}}_{n,m}^{\text{for}} \tag{11}$$

下一块的输入 $\hat{\boldsymbol{y}}_{n,m+1}^{\text{back}}$ 等于上一块的后向预测输出 $\boldsymbol{y}_{n,m}^{\text{back}}$ 与其期望值的差，即

$$\hat{\boldsymbol{y}}_{n,m+1}^{\text{back}}=\boldsymbol{y}_{n,m}^{\text{back}}-\hat{\boldsymbol{y}}_{n,m}^{\text{back}} \tag{12}$$

整个模型的输出则来自于每个堆栈的输出 $\hat{\boldsymbol{y}}_j^{\text{for}}$ 之和，即

$$\hat{\boldsymbol{y}}^{\text{for}}=\sum_{j=1}^{n}\hat{\boldsymbol{y}}_j^{\text{for}} \tag{13}$$

依据时间序列分解可以将齿轮箱油温数据分解成 3 项：长期趋势项堆栈用于捕捉齿轮箱温度的长期变化趋势；日常波动项堆栈用于拟合周期性波动；特征变量项堆栈则处理来自外部因素的影响。

对于离散定义的时间向量 $\boldsymbol{t}^{\text{T}}=[0,1,\cdots,H-1]$ H，齿轮箱油温的长期变化趋势 $\boldsymbol{y}_{n,m}^{\text{trend}}$ 可通过多项式项进行拟合。该项反映温度随时间的缓慢变化过程。

$$\boldsymbol{y}_{n,m}^{\text{trend}}=\sum_{i=0}^{N_{\text{pol}}}\boldsymbol{t}^i\boldsymbol{\theta}_{n,m}^{\text{trend}}=\boldsymbol{T}\boldsymbol{\theta}_{n,m}^{\text{trend}} \tag{14}$$

$$\boldsymbol{T}=[1,t,\cdots,t^{N_{\text{pol}}}]\in\mathbf{R}^{H\times(N_{\text{pol}}+1)}$$

式中：N_{pol} 为最大多项式度；$\boldsymbol{t}^i$ 为时间向量的 i 次幂；$\boldsymbol{\theta}_{n,m}^{\text{trend}}$ 为第 n 个堆栈第 m 个块的趋势项系数。

齿轮箱油温的日常波动项 $\boldsymbol{y}_{n,m}^{\text{seas}}$ 通过谐波函数项来描述。该项能够捕捉齿轮箱运行中的周期性温度变化特征。

$$\boldsymbol{y}_{n,m}^{\text{seas}}=\sum_{i=0}^{\frac{H}{2}-1}\left[\cos\left(\frac{2\pi it}{N_{\text{hr}}}\right)\boldsymbol{\theta}_{n,m,i}^{\text{seas}}+\sin\left(\frac{2\pi it}{N_{\text{hr}}}\right)\boldsymbol{\theta}_{n,m,i+\frac{H}{2}}^{\text{seas}}\right]=\boldsymbol{S}\boldsymbol{\theta}_{n,m}^{\text{seas}} \tag{15}$$

式中：t 为时间；N_{hr} 为超参数，用于控制谐振振荡；$\boldsymbol{S}$ 为谐波；$\boldsymbol{\theta}_{n,m,i}^{\text{seas}}$、$\boldsymbol{\theta}_{n,m,i+\frac{H}{2}}^{\text{seas}}$ 分别为日常波动项的傅里叶展开式余弦、正弦系数。

此时基向量 $\boldsymbol{V}_{n,m}^{\text{for}}$ 为谐波 $\boldsymbol{S}$，即

$$\boldsymbol{S}=\left[1,\ \cos\frac{2\pi t}{N_{\text{hr}}},\cdots\cos\left(\frac{H-1}{2}\times\frac{2\pi t}{N_{\text{hr}}}\right),\cdots,\ \sin\frac{2\pi t}{N_{\text{hr}}},\cdots,\cos\left(\frac{H-1}{2}\times\frac{2\pi t}{N_{\text{hr}}}\right)\right]\in\mathbf{R}^{H\times(H-1)} \tag{16}$$

齿轮箱油温的特征变量项 $\boldsymbol{y}_{n,m}^{\text{exog}}$ 通过特征变量项进行扩展，以充分考虑特征变量对油温的影响。

$$\boldsymbol{y}_{n,m}^{\text{exog}}=\sum_{i=0}^{N_{\text{x}}}\boldsymbol{x}_i\boldsymbol{\theta}_{n,m}^{\text{exog}}=\boldsymbol{x}\boldsymbol{\theta}_{n,m}^{\text{exog}} \tag{17}$$

式中：N_{x} 为多项式展开的幂次；$\boldsymbol{\theta}_{n,m}^{\text{exog}}$ 为外生变量的系数；$\boldsymbol{x}_i$ 为 $\boldsymbol{y}_{n,m}^{\text{exog}}$ 多项式展开的第 i 次基向量。

此时基向量 $\boldsymbol{V}_{n,m}^{\text{for}}$ 为特征矩阵 $\boldsymbol{x}=[\boldsymbol{x}_1,\boldsymbol{x}_2]$。特征变量基础扩展可以看作是时间变化的局部回归。

综上，利用历史数据进行后向预测，使用协变量进行前向预测，通过逐层堆栈叠加和线性映射，可实现对齿轮箱状态变量的预测。

3.2 模型参数优化

在风电机组齿轮箱状态预测过程中，采用 TPE 算法对 NBEATSx 模型的超参数进行优化，以提升模型的预测精度和泛化能力。TPE 算法能够有效处理超参数空间中的复杂结构。针对齿轮箱油温预测，模型的主要参数包括趋势项、谐波项的系数、特征变量的权重等，以寻找最优参数组合。

齿轮箱油温预测模型参数优化流程图如图 5 所示，图中：λ_{EI} 为期望改进；λ_{MAE} 为平均绝对误差。TPE 算法通过选择适当的条件概率分布指导超参数采样，基于最小化目标函数的 λ_{EI} 准则，选择低目标值的超参数作为候选参数。该算法在每次迭代中更新条件概率分布，并在下一次迭代中进行超参数采样。

$p(\lambda|M)$ 的密度函数为

$$p(\lambda|M)=\begin{cases}l(\lambda), & M<M^*\\ g(\lambda), & M\geqslant M^*\end{cases} \tag{18}$$

式中：$p(\lambda|M)$ 为密度函数；λ 和 M 分别为待优化参数和目标函数，$M(\lambda)=\lambda_{\text{MAE}}$；$M^*$ 为变量的阈值，通常为观测值的中位数；$g(\lambda)$ 为由剩余观测值形成的密度函数；$l(\lambda)$ 为由满足损失函数 $f(\lambda_i)<M^*$ 的观测值 λ_i 形成的密度函数。

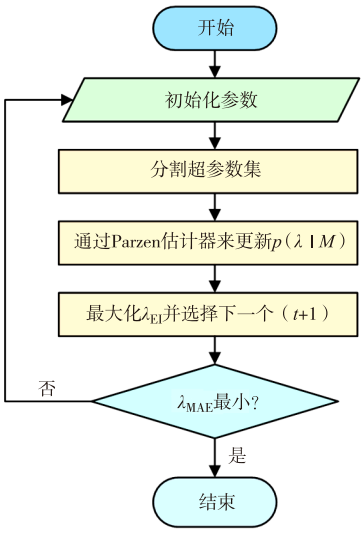


图 5 齿轮箱油温预测模型参数优化流程图

Fig. 5 Flowchart of parameter optimization for gearbox oil temperature prediction model

TPE 算法优化首先根据现有观测点数据确定 M^*，然后将观测点分成 $l(\lambda)$、$g(\lambda)$ 2 个密度函数。利用预期改进 λ_{EI} 获取策略，通过最大化 λ_{EI} 形成新的观测点。TPE 的采集函数 λ_{EI} 为

$$\lambda_{\text{EI}}(\lambda)=\int_{-\infty}^{M^*}(M^*-M)\frac{p(\lambda\mid M)p(M)}{p(\lambda)}\mathrm{d}M \tag{19}$$

令 $\gamma=p(M<M^*)$，则式(19)中分母可构造为 $p(\lambda)=\gamma l(\lambda)+(1-\gamma)g(\lambda)$，分子可构造为

$$\begin{aligned}&\int_{-\infty}^{M^*}(M^*-M)p(M)p(\lambda\mid M)\mathrm{d}M=\\ &l(\lambda)\int_{-\infty}^{M^*}(M^*-M)p(M)\mathrm{d}M\\ &=\gamma M^* l(\lambda)-l(\lambda)\int_{-\infty}^{M^*}p(M)\mathrm{d}M\end{aligned} \tag{20}$$

故采集函数 λ_{EI} 可表示为

$$\lambda_{\text{EI}}(\lambda)=\frac{\gamma M^* l(\lambda)-l(\lambda)\int_{-\infty}^{M^*}p(M)\mathrm{d}M}{\gamma l(\lambda)+(1-\gamma)g(\lambda)}\propto\left[\gamma+\frac{g(\lambda)}{l(\lambda)}(1-\gamma)\right]^{-1} \tag{21}$$

当超参数 λ 对应于最高概率密度 $l(\lambda)$ 和最低概率密度 $g(\lambda)$ 时，λ_{EI} 达到最大。TPE 算法采用 $l(\lambda)$ 和 $g(\lambda)$ 构造一组样本超参数，并基于表达式 $l(\lambda)/g(\lambda)$ 计算。每次迭代中，返回 λ_{EI} 最大的点，从而寻找总体最优超参数。

根据输入数据，利用树结构的 Parzen 估计器对 2 个案例风电机组数据模型参数进行优化，最后得到模型部分关键最优参数，如表 5 所示。

表 5 根据 TPE 算法优化后 NBEATSx 关键参数

Tab. 5 Optimizing key parameters of NBEATSx by TPE algorithm

参数	案例 1	案例 2
激活函数	prelu	prelu
b	2	2
s	2	2
层数量	(2,2)	(2,2)
堆栈类型	(趋势项,特征项)	(趋势项,特征项)
输出大小	24	24
N_{pol}	3	3
N_{hr}	2	2
N_{x}	2	2
N_{h}	72	72
N_{s}	24	24

3.3　齿轮箱状态预测结果

分别采用长短期记忆网络(LSTM)、随机森林(RF)、时间卷积网络(TCN)、极端梯度提升树(XGBoost)以及本文模型对案例 1 和案例 2 两台机组齿轮箱油温进行预测，结果如图 6 所示。可见，NBEATSx 模型的预测值最贴近真实值，尤其是在复杂的温度变化情况下，该模型能够更好地捕捉趋势和细节。

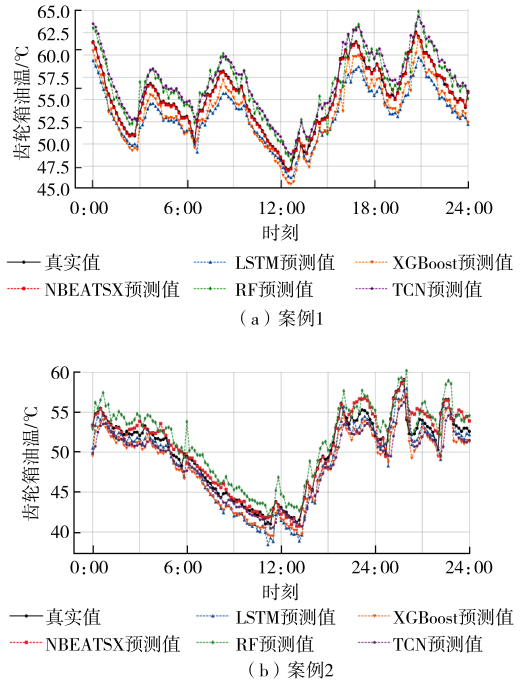


图 6　案例 1、案例 2 下各模型预测齿轮箱油温对比

Fig. 6　Comparison of predicted gearbox oil temperature by different models in case 1 and case 2

这说明 NBEATSx 模型在处理时间序列预测问题时，特别是在齿轮箱油温预测任务中具有更强的准确性和适应性。为进一步评价齿轮箱状态预测结果，选用均方根误差 y_{RMSE} 和决定系数 R^2 作为评价指标。

$$y_{\mathrm{RMSE}}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \tag{22}$$

$$R^2=1+\frac{\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right|}{\sum_{i=1}^{n}\left|y_i-\bar{y}\right|} \tag{23}$$

式中：y_i、$\hat{y}_i$、$\bar{y}$ 分别为齿轮箱油温的真实值、预测值和平均值；n 为案例中数据点个数。

各模型评价指标如表 6 所示。可见，NBEATSx 模型在案例 1 中的 y_{RMSE} 为 0.28，显著低于其他模型，且 R^2 达到 0.99，表明其预测精度极高。而在案例 2 中，NBEATSx 模型的 y_{RMSE} 为 1.60，依然保持较高的准确性，且 R^2 为 0.97，优于其他对比模型。相比之下，其他模型的 y_{RMSE} 普遍较高，R^2 也略低。这进一步验证了 NBEATSx 模型在预测精度和稳定性方面的优势。

表 6　各模型评价指标

Tab. 6　Evaluation indicators for each model

模型	y_{RMSE}/℃		R^2	
	案例 1	案例 2	案例 1	案例 2
LSTM	1.89	2.23	0.97	0.94
RF	1.91	2.06	0.97	0.95
XGBoost	1.81	1.79	0.97	0.96
TCN	2.05	1.83	0.96	0.96
NBEATSx	0.28	1.60	0.99	0.97

3.4　故障预警阈值设置

风电机组在较恶劣环境下运行时，输出状态变量受多种因素的影响。如果仅使用某一时刻的误差进行故障预警阈值设置，会增加误报警风险[27]。为此，建立滑动窗口模型，计算每个窗口内的预测误差，以实现异常状态预警。选取滑动窗口内齿轮箱状态变量真实值与预测值的平均绝对误差 e_{ij} 作为状态评价指标。该评价指标计算式为

$$e_{ij}=\frac{1}{w}\sum_{t=t_i-w}^{t_i}\left|y_t'-\hat{y}_t'\right| \tag{24}$$

式中：t_i 为预测滑动窗口的起始时间；w 为滑动窗口宽度；y_t' 和 $\hat{y}_t'$ 分别为 t 时刻齿轮箱油温的实际值和预测值。

根据统计学中最常用的区间估计方法，计算置信上限并将其作为预警阈值 α_{th}，即

$$\alpha_{\mathrm{th}}=\bar{\alpha}+\frac{S}{\sqrt{n_{\mathrm{point}}}}t_\alpha(m_{\mathrm{wind}}-1)+3S \tag{25}$$

式中：α 为上分位数；$\bar{\alpha}$ 为机组正常工况下的状态指标均值；t_α 为函数；S 为状态指标的标准差；m_{wind} 为滑动窗口个数；n_{point} 为输入样本总数；$t_\alpha()$ 为 t 分布。

3.5　齿轮箱状态预警结果

为验证 NBEATSx 模型的预警效果，将故障预警数据集输入 NBEATSx 模型，构建滑动窗口模型。经多次实验发现，当滑动窗口宽度 w 为 100 时，该模型能够有效捕捉足够的历史信息；当窗口移动增量 Δw 为 10 时，能够有效保证预测的连续性和分辨率。因此，选取这 2 个参数分别对案例 1 和案例 2 机组进行预警。

案例 1 机组在 2018 年 10 月 13 日 06:52 出现“齿轮箱轴承损坏”的故障。经 NBEATSx 模型预

测齿轮箱油温，并实时计算状态指标，设置故障阈值为 8.48。案例 1 齿轮箱油温对比图和预警图如图 7 所示。如图 7(a)所示，在故障发生前，齿轮箱油温真实值与预测值的偏差逐步变大。如图 7(b)所示，若使用单一误差作为评价指标，会增加系统误报。而状态指标在 2018 年 10 月 13 日 03:00 开始超过设定阈值，提前 3 h 52 min 预知故障，从而实现对齿轮箱的故障预警并降低误报率。

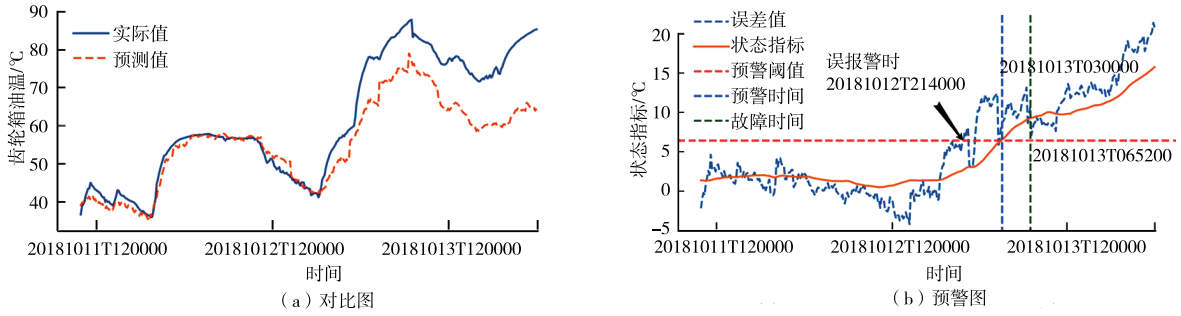


图 7 案例 1 齿轮箱油温对比图和预警图

Fig. 7 Comparison of gearbox oil temperature and warning diagram for case 1

案例 2 机组在 2018 年 10 月 4 日 07:21 出现“齿轮箱类总故障”的故障。经 NBEATSx 模型预测齿轮箱油温度，并实时计算状态指标，设置故障阈值为 7.62。案例 2 齿轮箱油温对比图和预警图如图 8 所示。如图 8(a)所示，同样出现预测值与真实值差别逐步变大的情况。如图 8(b)所示，若使用单一误差作为评价指标，会增加系统误报。状态指标在 2018 年 10 月 4 日 03:50 开始超过设定阈值，提前 3 h 31 min 预知故障，实现了对齿轮箱的故障预警并减少系统误报。

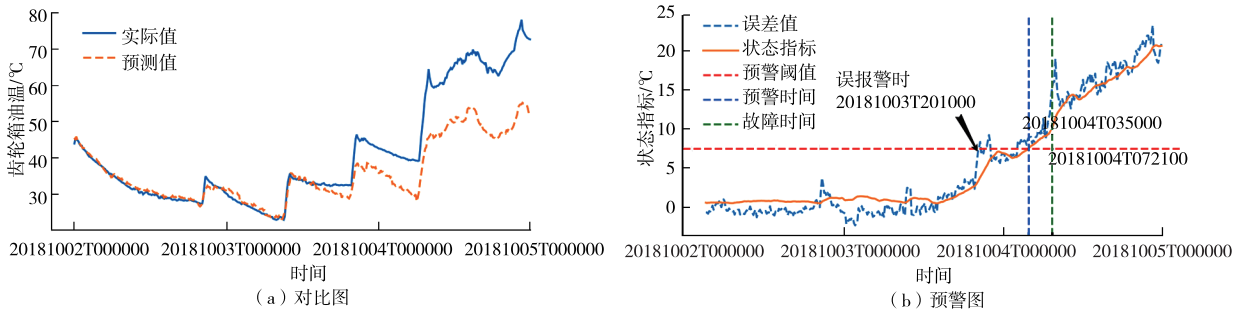


图 8 案例 2 齿轮箱油温对比图和预警图

Fig. 8 Comparison of gearbox oil temperature and warning diagram for case 2

3 结论

(1) 提出了一种融合 DBSCAN 聚类、Sigmoid 函数拟合与曼哈顿距离筛选的风力涡轮机齿轮箱数据预处理方法。以与 Sigmoid 曲线的曼哈顿距离小于 0.3 为标准，该方法有效剔除离群点与限功率数据，同时通过构建含时间衰减机制的加权距离函数，精准表征数据动态关联性并提取 2 个与齿轮箱油温强相关的特征变量，为后续故障预测提供高质量数据基础。

(2) 验证了 NBEATSx 模型在齿轮箱油温预测中的显著优势。该模型通过多堆栈结构独立处理时间序列分量，能有效捕捉油温变化复杂模式，与传统模型相比，其预测结果的 y_{RMSE} 显著降低、相关系数明显提升，各项评价指标均表现更优，为故障预警提供了精准的状态表征支持。

(3) 建立了实时阈值动态设置机制，结合上述数据预处理与预测方法，不仅有效平滑了预测偏差、避免了误报警情况，还能在故障发生数小时前发出预警。该综合故障预测与预警方法经实际数据验证，兼具准确性、可靠性与实时性，为风力涡轮机齿轮箱状态监测与故障预警提供了新的技术路径，对提升风电设备运维效率、降低停机损失具有重要实际应用价值。

参考文献：

[1] 苏健，梁英波，丁麟，等. 碳中和目标下我国能源发展战略探讨[J]. 中国科学院院刊，2021，36(9)：1001-1009.

SU Jian，LIANG Yingbo，DING Lin，et al. Research

on China's energy development strategy under carbon neutrality[J]. Bulletin of Chinese Academy of Sciences, 2021, 36(9): 1001-1009.

[2] GWEC. GWEC's global wind report 2024[R]. Belgium: GWEC, 2024.

[3] 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.

[4] 黄荣舟，汤宝平，杨燕妮，等. 基于长短时记忆网络融合 SCADA 数据的风电齿轮箱状态监测[J]. 太阳能学报，2021，42(1)：235-239.
HUANG Rongzhou, TANG Baoping, YANG Yanni, et al. Condition monitoring of wind turbine gearbox based on LSTM neural network fusing SCADA data[J]. Acta Energiae Solaris Sinica, 2021, 42(1): 235-239.

[5] CHENG Fangzhou, QU Liyan, QIAO Wei, et al. Enhanced particle filtering for bearing remaining useful life prediction of wind turbine drivetrain gearboxes[J]. IEEE Transactions on Industrial Electronics, 2019, 66(6): 4738-4748.

[6] WANG Haifeng, ZHAO Xingyu, WANG Weijun. Fault diagnosis and prediction of wind turbine gearbox based on a new hybrid model[J]. Environmental Science and Pollution Research, 2023, 30(9): 24506-24520.

[7] 雷萌，郭鹏，刘博嵩. 基于自适应 DBSCAN 算法的风电机组异常数据识别研究[J]. 动力工程学报，2021，41(10)：859-865.
LEI Meng, GUO Peng, LIU Bosong. Research on abnormal data ldentification of wind turbines based on adaptive DBSCAN algorithm[J]. Journal of Chinese Society of Power Engineering, 2021, 41(10): 859-865.

[8] BHARDWAJ U, TEIXEIRA A P, SOARES C G, et al. Reliability prediction of an offshore wind turbine gearbox[J]. Renewable Energy, 2019, 141: 693-706.

[9] 黄忠山，田凌，向东，等. 基于 PCA 和 SPC-动态神经网络的风电机组齿轮箱油温趋势预测[J]. 清华大学学报(自然科学版)，2018，58(6)：539-546.
HUANG Zhongshan, TIAN Ling, XIANG Dong, et al. Prediction of oil temperature variations in a wind turbine gearbox based on PCA and an SPC-dynamic neural network hybrid[J]. Journal of Tsinghua University (Science and Technology), 2018, 58(6): 539-546.

[10] HE Ke, SU Zuqiang, TIAN Xiaoqing, et al. RUL prediction of wind turbine gearbox bearings based on self-calibration temporal convolutional network[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 3501912.

[11] 苏向敬，山衍浩，周汶鑫，等. 基于 GRU 和注意力机制的海上风机齿轮箱状态监测[J]. 电力系统保护与控制，2021，49(24)：141-149.
SU Xiangjing, SHAN Yanhao, ZHOU Wenxin, et al. GRU and attention mechanism-based condition monitoring of an offshore wind turbine gearbox[J]. Power System Protection and Control, 2021, 49(24): 141-149.

[12] 张书瑶，刘长良，王梓齐，等. 基于改进实例学习算法的风电机组齿轮箱状态监测[J]. 动力工程学报，2024，44(10)：1620-1631.
ZHANG Shuyao, LIU Changliang, WANG Ziqi, et al. Wind turbine gearbox condition monitoring based on improved instance learning algorithm[J]. Journal of Chinese Society of Power Engineering, 2024, 44(10): 1620-1631.

[13] 景彤梅，齐咏生，刘利强，等. 基于 KECA-GRNN 的风电机组齿轮箱状态监测与健康评估[J]. 太阳能学报，2021，42(6)：400-408.
JING Tongmei, QI Yongsheng, LIU Liqiang, et al. Condition monitoring and health assessment of wind turbine gearbox based on KECA-GRNN[J]. Acta Energiae Solaris Sinica, 2021, 42(6): 400-408.

[14] FANG Jinjie, YANG Linshan, WEN Xiaohu, et al. Ensemble learning using multivariate variational mode decomposition based on the transformer for multi-step-ahead streamflow forecasting[J]. Journal of Hydrology, 2024, 636: 131275.

[15] HUANG Xinyu, TANG Jun, SHEN Yongming. Long time series of ocean wave prediction based on PatchTST model[J]. Ocean Engineering, 2024, 301: 117572.

[16] ZHANG Jianwei, TAN Chunchen, CAI Zengyu, et al. Cellular traffic forecasting based on inverted transformer for mobile perception dual-level base station sleep control[J]. Ad Hoc Networks, 2024, 161: 103505.

[17] SOUTO H G, MORADI A. Introducing NBEATSx to realized volatility forecasting[J]. Expert Systems with Applications, 2024, 242: 122802.

[18] OLIVARES K G, CHALLU C, MARCJASZ G, et al. Neural basis expansion analysis with exogenous

variables: forecasting electricity prices with NBEATSx[J]. International Journal of Forecasting, 2023, 39(2): 884-900.

[19] 雍彬, 陈进, 张方红, 等. 基于门控循环网络融合多源数据的风电齿轮箱状态预警方法[J]. 太阳能学报, 2021, 42(8): 421-425.
YONG Bin, CHEN Jin, ZHANG Fanghong, et al. State warning of wind turbine gearbox based on gated recurrent unit network fusing multi-source data[J]. Acta Energiae Solaris Sinica, 2021, 42(8): 421-425.

[20] FU Yang, WANG Shuo, JIA Feng, et al. Two-stage cascaded high-precision early warning of wind turbine faults based on machine learning and data graphization[J]. Journal of Electrical Engineering & Technology, 2024, 19(3): 1919-1931.

[21] WANG Shaoke, ZHANG Zhaoyan, WANG Peiguang, et al. Failure warning of gearbox for wind turbine based on 3σ-median criterion and NSET[J]. Energy Reports, 2021, 7 (Sup7): 1182-1197.

[22] 向玲, 朱浩伟, 丁显, 等. 基于 CAE 与 BiLSTM 结合的风电机组齿轮箱故障预警方法研究[J]. 动力工程学报, 2022, 42(6): 514-521.
XIANG Ling, ZHU Haowei, DING Xian, et al. Research on fault warning method of wind turbine gearbox based on CAE and BiLSTM[J]. Journal of Chinese Society of Power Engineering, 2022, 42(6): 514-521.

[23] 罗毅, 段明达. 基于 IWOA-Transformer 的磨煤机故障预警[J]. 动力工程学报, 2024, 44(6): 939-946.
LUO Yi, DUAN Mingda. Fault warning of coal mill based on IWOA-transformer[J]. Journal of Chinese Society of Power Engineering, 2024, 44(6): 939-946.

[24] KAVAKLI M, GUDMESTAD O T. Analysis and assessment of onshore and offshore wind turbines failures[J]. International Journal of Energy Production and Management, 2023, 8(1): 27-34.

[25] 杨传光, 费胜巍. 基于 Sigmoid 函数的非对称式 S 型曲线机械臂加减速控制方法[J]. 东华大学学报(自然科学版), 2024, 50(5): 118-126.
YANG Chuanguang, FEI Shengwei. Asymmetric S-curve manipulator acceleration and deceleration control method based on Sigmoid function[J]. Journal of Donghua University (Natural Science), 2024, 50(5): 118-126.

[26] SU Siyu, SUN Youchao, PENG Chong, et al. Improved gray correlation analysis and combined prediction model for aviation accidents[J]. Engineering Computation, 2023, 40(7/8): 1570-1592.

[27] 马良玉, 耿妍竹, 梁书源, 等. 基于 Stacking 多模型融合的风电机组齿轮箱油池温度异常预警[J]. 中国电机工程学报, 2023, 43(增刊 1): 242-251.
MA Liangyu, GENG Yanzhu, LIANG Shuyuan, et al. Anomaly warning of wind turbine gearbox oil pool temperature based on stacking fusion of multiple models[J]. Proceedings of the CSEE, 2023, 43(Sup1): 242-251.