

基于自回归嵌入的风电机组传动系统双阈值状态监测^{*}

曹利宵¹, 贾晓冰¹, 李 宁¹, 常雯博¹, 苏保中²

(1. 燕山大学电气工程学院 秦皇岛 066000; 2. 中海油能源发展装备技术有限公司 天津 300457)

摘要:风电机组是一种将风能转换为电能的复杂机电混合系统,其传动系统包含齿轮、轴承等众多高精度的机械部件,这些部件在长期运转过程中,容易出现疲劳、磨损等问题,导致传动系统易于发生各种故障,严重影响机组运行的安全与发电效率。为保证风电机组整机运行的安全性,需要高效准确地检测出传动系统的异常运行状态。针对现有研究复杂运行状态捕捉能力弱以及部件维护、剔除白噪声干扰准确率低的问题,构建了一种基于自回归嵌入的可学习分解与双注意力网络的状态监测模型。首先,对数据采集与监测系统(SCADA)收集到的数据进行预处理,并且通过相关性分析选择出与故障演化强相关的温度等特征参数。其次,引入自回归嵌入模块,使用动态令牌更好地捕捉多维时间序列特征,进行传动系统相关温度变量的预测,实现故障特征的动态建模。然后,提出双阈值判别的监测网络,结合相关变量的残差与信息熵确定双重健康阈值,通过阈值的双重机制,监测过程能够进一步剔除白噪声等异常信号,提供准确的预警时间。最后,通过两个实际风电机组传动系统的故障案例验证了所构建模型的有效性,当风电机组发生故障时,相较于传统的数据采集与监测系统(SCADA),可以提前约6~10天对故障信号进行早期预警。

关键词:数据采集与监测系统;传动系统;状态监测;自回归嵌入;双阈值判别

中图分类号: TH17 **文献标识码:** A **国家标准学科分类代码:** 460. 99

Dual-threshold condition monitoring of wind turbine drive train system based on autoregressive embedding

Cao Lixiao¹, Jia Xiaobing¹, Li Ning¹, Chang Wenbo¹, Su Baozhong²

(1. School of Electrical Engineering, Yanshan University, Qinhuangdao 066000, China; 2. CNOOC EnerTech Equipment Technology Co., Ltd., Tianjin 300457, China)

Abstract: Wind turbines are complex electromechanical hybrid systems that convert wind energy into electrical power. Their transmission systems include numerous high-precision mechanical components such as gears and bearings. These components are prone to fatigue and wear during prolonged operation, making the transmission system vulnerable to various failures, which significantly impacts both the safety of the units and their power generation efficiency. To ensure the safety of wind turbines, it is necessary to efficiently and accurately detect abnormal states in the transmission system. This paper addresses the limitations of existing research in capturing complex operational states and the low accuracy in component maintenance and white noise rejection. A state monitoring model based on a learnable decomposition and dual-attention network using autoregressive embeddings is proposed. First, the data collected by the supervisory control and data acquisition (SCADA) system is preprocessed, and correlation analysis is performed to select feature parameters such as temperature that are strongly correlated with fault evolution. Next, an autoregressive embedding module is introduced, utilizing dynamic tokens to better capture multidimensional time-series features. This allows for the prediction of relevant temperature variables in the transmission system, achieving dynamic modelling of fault characteristics. Then, a dual-threshold discrimination monitoring network is proposed, combining the residuals and information entropy of relevant variables to determine the

dual health thresholds, further eliminates abnormal signals such as white noise and providing accurate warning times. Finally, the effectiveness of the proposed model is verified through two actual wind turbine transmission system failure cases. Compared to traditional SCADA systems, early warning signals can be detected approximately 6 to 10 days in advance when a fault occurs in the wind turbine.

Keywords: data acquisition and monitoring system; drive train system; condition monitoring; autoregressive embedding; dual threshold discriminat

0 引言

风能作为一种极具发展前景的环保型可持续能源,已成为应对环境污染和全球能源危机的传统化石燃料的关键替代品。《2024年全球风能报告》显示,全球风电新增装机容量再创历史新高,达到121.6 GW,总装机容量突破1100 GW大关,同比增长约13%^[1-2]。然而,风力涡轮机往往位于偏远地带,再加上恶劣的运行条件和复杂的结构,使得它们故障频发。因此,对风力涡轮机进行连续监测以及实现早期故障检测显得尤为重要。风力发电机由复杂的电气系统和机械系统组成,一个子系统故障可能会影响其他子系统以及整个风电机组。而传动系统是关键的风机子系统,传动系统由2个基本部件组成:齿轮箱和发电机,齿轮箱负责在主轴与发电机之间构建能量传递通道,而发电机负责将变桨系统的机械能转换为电能^[3]。

监控与数据采集系统(supervisory control and data acquisition system, SCADA)作为风电场的核心监测平台,是广泛安装在风电场的数据采集系统,它配备了大量的传感器来采集涡轮机运行和环境参数的各种实时数据,能够同步获取机组功率输出、温度、振动等运行数据,并结合风速、风向、温度等环境参数构建全维度监测模型,为风电场的智能运维与能效优化提供精准的数据支撑^[4]。这些参数包含了风力发电机组的各种运行特性和丰富的健康状态信息,是状态监测和故障诊断的关键要素。因此,基于SCADA数据的风电机组状态监测或故障检测越来越受到学者们的关注^[5]。

近年来,随着人工智能的发展,深度学习技术凭借强大的建模能力,在解析海量非线性数据方面展现出显著优势,而使用深度学习的数据驱动方法在基于SCADA数据的状态监测和故障检测中也受到了更多关注。Jia等^[6]提出了利用SCADA数据,基于深度卷积神经网络(convolutional neural network, CNN)和引入去噪自编码器(denoising autoencoder, DAE)构建性能预测(performance forecasting, PF)框架进行监测。金晓航等^[7]提出了一种通过用K均值聚类算法(K-means clustering algorithm, K-means)划分工况,根据差分进化算法优化回声状态网络模型的方法,监测SCADA数据中的风功率进行早期故障预警。马同旭等^[8]提出了一种基于

两阶段多维数据生成动态阈值判定与实时健康指数评估的方法,监测齿轮箱运行状态。何群等^[9]提出了使用长短时记忆网络(long short-term memory, LSTM)以及时空注意力联合(gated recurrent unit, GRU)网络提取时间特征,通过监测残差值实现早期预警。考虑将SCADA数据进行分割为固定时间窗口长度的样本,网络进一步提取样本的时间特征,进行实时更新监测,效果明显优于不考虑数据时序特征的方法。

SCADA数据是典型的时间序列数据,考虑到风机的运行状态数据往往呈现出高度的非线性特征和时间相关性^[10],其时间序列特征的分析对于风电机组的状态监测至关重要。陈海燕等^[11]提出一种基于时-频特征融合的多变量时间序列预测模型,在时域提取长短期和周期特性,从而提高时序预测的精度。朱昶胜等^[12]提出了一种基于多层感知器(multi-layer perceptron, MLP)的时间序列预测模型,利用MLP在低频段的拟合能力,通过自适应序列分解模块等应对通道冗余问题。Zhou等^[13]提出通过使用移动平均核(moving average kernel, MOV)将变换器(transformer, TF)与季节-趋势分解结合,通过分解捕获全局轮廓。然而,在风电机组多变量时间序列数据中,每个组成变量的时间序列往往在其趋势中表现出独特的变化。Liu等^[14]提出现实世界的时间序列数据会不断受其趋势演变所引起的分布变化影响——这是一个独特的属性,增加了对风机数据动态序列内变化模式建模的复杂性。虽然上述学者提出将时间序列数据进行分解以便捕获全局轮廓,但其分解的过程具有不可训练性且未对动态序列内变化建模的复杂性进行讨论与剖析,在风电机组的状态监测方向,其具体应用也尚未被有效扩展。

1 基于自回归嵌入的可学习分解与自注意力网络的状态监测方法

针对风电机组的传动系统,本研究提出一种新的基于自回归嵌入的可学习分解与自注意力网络(autoregressive-based learnable decomposition and self-attention module, AR-Ledsam)的状态监测方法,主要包括4个阶段:数据准备、模型训练、双阈值计算和在线测试阶段,其整体框架如图1所示。数据准备阶段,对历史正常SCADA数据进行预处理,包括局部离群因子(local outlier factor, LOF)和数据归一化。然后,通过皮尔逊相关系数

选择与目标运行参数相关度较高的变量。模型准备阶段,构造 AR-Ledsam 模型,主要包括可学习分解和自回归嵌入部分,以捕获 SCADA 数据中固有的复杂时间特征。绿色部分的双阈值计算阶段,分别计算预测值的残差有效值 R_m 与残差序列信息熵 E_n ,将健康状态下运行的最大值 H_{R_m} 与 H_{E_n} 设置为阈值。在线测试阶段,将预处理后的 SCADA 数据送入训练好的 AR-Ledsam 模型中,利用预测结果计算 R_m 与 E_n 这 2 个指标。最后,将指标与双阈值进行比较,跟踪系统运行状况,当同时满足两个阈值的故障条件时,检测传动系统为异常状态。

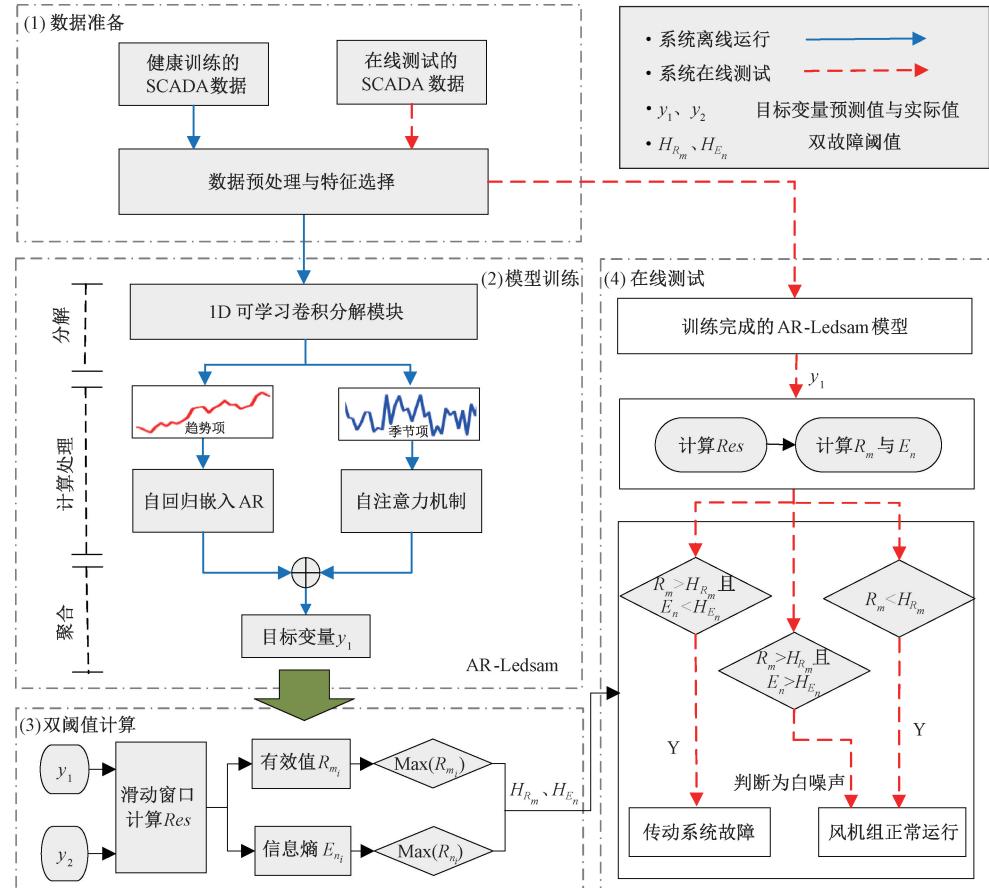


图 1 状态监测方法整体框架

Fig. 1 Overall framework diagram of the condition monitoring method

2 数据准备

风电机组 SCADA 数据往往具有海量规模和高维度的属性,且常有缺失值、异常值和噪声的干扰。为了提升训练效率、增强预测性能,引入较为关键的数据准备阶段。数据准备主要包括数据预处理和参数选择两个部分。

2.1 风电机组数据预处理

数据预处理主要包括:1)对健康数据中的缺失值、功率零值负值、 >25 m/s 或 <4 m/s 的异常风速值等进行剔除;2)由于实际风电场存在受通信干扰、极端气象等现象的影响,采用 LOF 算法对风功率图像中的离群值进行剔除。LOF 是一种基于密度的无监督异常检测

算法^[15],通过比较数据点局部密度与邻域密度的相对差异来识别离群点,即计算每个数据点的 LOF 值,量化离群程度,有效识别并剔除异常的风功率值。风电机组 SCADA 数据预处理前后的对比图,如图 2 所示。

2.2 风机组数据特征选择

SCADA 数据包含了大量与运行特性相关的变量,如风速、电流和温度等。然而,过多的参数变量导致模型拟合不足,影响模型预测精度和异常检测速度。因此,为了减轻模型的计算负担、提升运行效率,使用皮尔逊相关系数法选择出与运行状态高度相关的变量,排除了相关性较低的变量^[16]。为了增强模型训练的稳定性并减轻梯度消失或爆炸问题,采用最小-最大归一化来映射提取的变量。皮尔逊相关系数法和归一化的计算为:

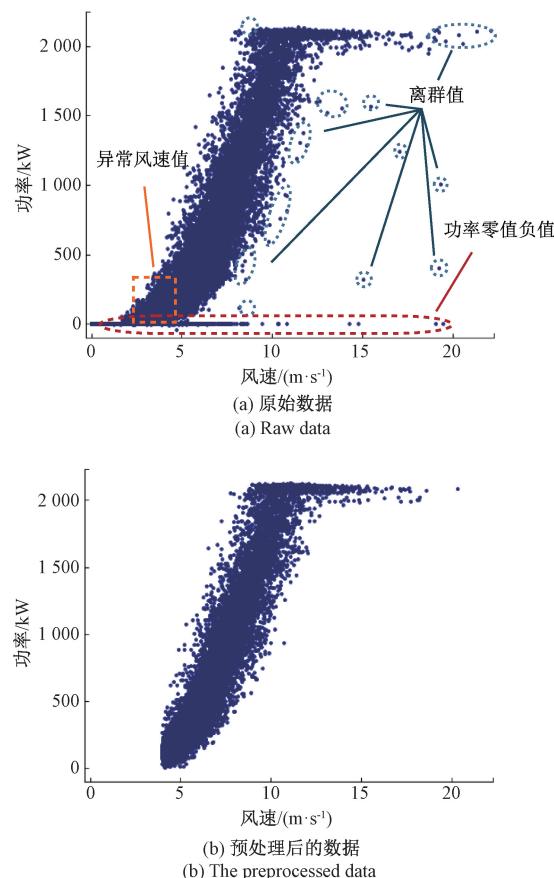


图2 风机SCADA数据预处理前后对比

Fig. 2 Comparison chart of wind turbine SCADA data before and after preprocessing

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (1)$$

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2)$$

式中: r 表示变量 X 与 Y 之间的相关性,越接近1,表示二者的相关性越强; \bar{X} 和 \bar{Y} 分别表示变量 X 、 Y 的平均值; n 表示抽样数; x_{norm} 表示归一化后变量的值; x_{min} 和 x_{max} 分别表示变量的最小值和最大值。对于风电机组的状态监测,应选择具有较大相关系数的SCADA变量。

3 模型训练

传统的时间序列分解模块采用移动平均核的方法提取趋势性部分,然而,基本的平均核对窗口内所有数据点均匀加权,并不适合捕获时序中的非平稳趋势^[17]。趋势性部分捕获后,要进一步理解序列内在的时间变化。由于传统点对点机制的限制^[18],使其忽略了数据中的波动

和潜在的不确定性,相比补丁式,容易导致次优结果,而将时序分解为一系列补丁也不可避免地会引入信息丢失。

为了解决以上问题,本研究通过一维可学习卷积分解模块提取趋势性部分,使得卷积层在识别特定时间特征时对中心位置更加敏感。同时,将自回归嵌入(autoregressive, AR)模块加入到提取的趋势性部分中,这种自回归令牌的生成方法可以动态模拟时间序列内的时间变化,最大限度地保留原始时间序列的信息,以提取多维的时间特征。此外,引入了基于残差有效值(R_m)和信息熵(E_n)的双阈值监测,有效排除噪声干扰,提高了模型的鲁棒性和泛化能力。

3.1 一维可学习卷积分解模块

投影与位置嵌入。将数据准备阶段完成后的多变量时间序列数据 $X \in R^{N \times T}$ 输入,绘制从原空间到新空间的投影,随后合并位置编码 $Pos \in R^{N \times D}$,有 $X_{embed} = (XW + b) + Pos$, $X_{embed} \in R^{N \times D}$,其中权重 $W \in R^{T \times D}$, $b \in R^{1 \times D}$, N 表示通道的维数,对应时间序列中变量的个数, D 表示层的维数,对应位置编码和时间编码后嵌入的维数。

可学习的卷积分解核。首先,定义 $S=1$ 的步幅和 $K=25$ 的核大小,权重的初始化是利用高斯分布进行的,假设权值 $\omega \in R^{1 \times 1 \times K}$,超参数 $\sigma \in R$,设置 $\sigma=1.0$,权值 ω 的计算公式为:

$$\begin{cases} \mathbf{U} \in R^{1 \times 1 \times K} \\ \mathbf{U}[0,0,i] = \exp\left(-\frac{\left(i - \frac{K}{2}\right)^2}{2\sigma^2}\right), \quad i = 1, 2, \dots, K \\ \omega = \text{Softmax}(\mathbf{U}, \text{dim} = -1) \end{cases} \quad (3)$$

权重初始化导致卷积核中心位置的权值最大,而核边缘位置的权值较小^[19]。这通常有利于卷积层在识别特定特征时对中心位置更加敏感。而投影与位置嵌入完成后,为了保证多变量时间序列 $X_{embed} \in R^{N \times D}$ 在卷积前后的序列长度相等,使用终端值来进行填充,因此得到 $X_{padded} \in R^{N \times (D+K-1)}$ 。将其分解为 N 个独立的时间序列 $x_i \in R^{1 \times (D+K-1)}$, $i = 1, 2, \dots, N$,对于每个 x_i ,应用一个具有共享权值的1D卷积核提取其趋势分量,记为 $\hat{x}_i \in R^{1 \times D}$ 。随后,将所有 \hat{x}_i 的卷积输出进行级联,形成结果矩阵 $X_{Trend} \in R^{N \times D}$ 。通过 $X_{Seasonal} = X_{embed} - X_{Trend}$,得到季节部分 $X_{Seasonal} \in R^{N \times D}$ 。整个过程概括计算为:

$$\begin{cases} X_{Trend} = LD(Padding(X_{embed})) \\ X_{Seasonal} = X_{embed} - X_{Trend} \end{cases} \quad (4)$$

趋势项部分。考虑到趋势项部分对序列内变化建模的适用,将AR加入到提取的趋势性部分 $X_{Trend} \in R^{N \times D}$ 当中,在保留原始信息的同时,更好的动态提取

时间特征。

季节项部分。添加自注意力机制,更好的模拟序列

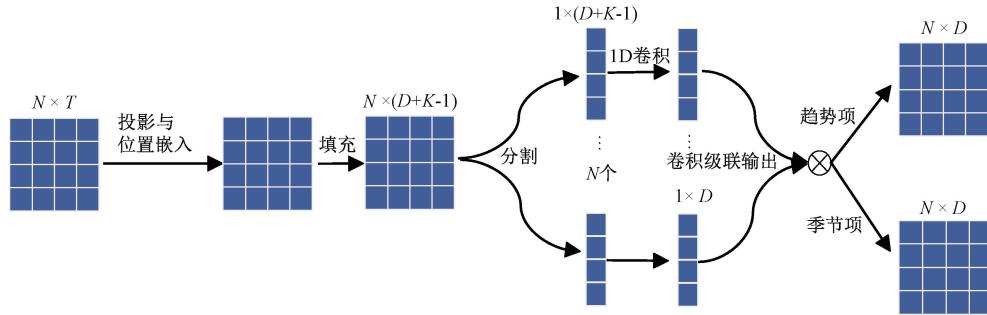


图 3 可学习卷积分解模块流程

Fig. 3 The flowchart of the convolutional decomposition module that can be learned

3.2 自回归嵌入

为了捕捉趋势项部分时间序列内的变化,提出一种 AR 的方法。趋势项的初始序列首先经过自回归处理,生成令

牌,令牌在细致保留全部原始信息的同时,部分地仿真序列中的动态变化。最终通过注意力机制、归一化等步骤将令牌传送至编码器中。整体流程如图 4 所示。

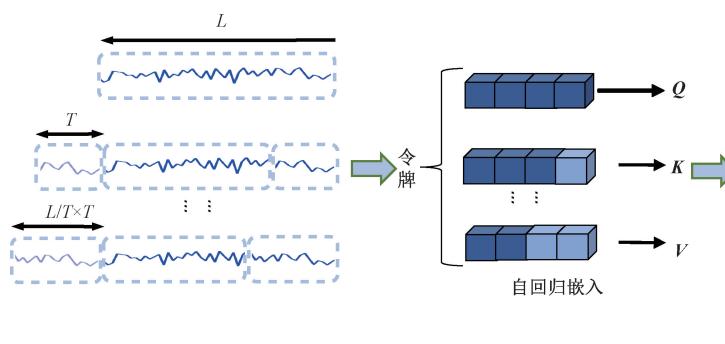


图 4 自回归嵌入动态令牌生成作用图

Fig. 4 Autoregressive embedding of dynamic token generation action diagrams

首先,将趋势项 X_{Trend} 拆分为 N 个独立的时间序列 $\mathbf{x}_t^i \in R^{1 \times D}, i = 1, 2, \dots, N$ 。对于每个 \mathbf{x}_t^i , 给定一个长度 L , 通过从序列的开始切割给定的长度并连接到时序的末尾来生成 $\mathbf{T}^i \in R^{D/L \times D}$ 个令牌, 计算公式为:

$$\begin{cases} \mathbf{T}^i[j, :] = \mathbf{x}_t^i[j \cdot L : D] \parallel \mathbf{x}_t^i[0 : j \cdot L] \\ j \in \left\{0, 1, \dots, \left\lfloor \frac{D}{L} \right\rfloor - 1\right\} \end{cases} \quad (5)$$

此过程中生成的令牌即为自回归嵌入。自回归令牌生成的方法能够动态模拟时间序列中的时间变化。与将时间序列划分为补丁和通过采样策略获得子序列相比, 该方法最大限度地保留了原始序列的时间信息。随后, 使用一个 Transformer 编码器, 它的权重在所有通道之间共享, 以对序列内的变化进行建模, 但考虑到主要关心原始序列的时间信息, 将原始序列 \mathbf{x}_t^i 指定为 Q , 同时将整个序列 \mathbf{T}^i 作为 K 和 V , 计算公式为:

$$\begin{cases} \mathbf{Q} = \mathbf{x}_t^i \mathbf{W} + \mathbf{b} \\ \mathbf{K} = \mathbf{T}^i \mathbf{W} + \mathbf{b} \\ \mathbf{V} = \mathbf{T}^i \mathbf{W} + \mathbf{b} \\ \mathbf{Attn} = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right) \mathbf{V} \\ \mathbf{H} = \text{LayerNorm}(\mathbf{x}_t^i + \mathbf{Attn}) \\ \mathbf{X}_{In}^i = \text{LayerNorm}(\text{FFN}(\mathbf{H}) + \mathbf{H}) \\ \mathbf{X}_{In} = \mathbf{X}_{In}^1 \parallel \mathbf{X}_{In}^2 \parallel \dots \parallel \mathbf{X}_{In}^n \end{cases} \quad (6)$$

所有通道输出的 \mathbf{X}_{In}^i 级联为 $\mathbf{X}_{In} \in R^{N \times D}$, 趋势项部分的最终输出结果计算公式为:

$$\begin{cases} \mathbf{W} \in R^{D \times F} \\ \mathbf{b} \in R^{1 \times F} \\ \mathbf{X}_{Out} = \mathbf{X}_{In} \mathbf{W} + \mathbf{b} \end{cases} \quad (7)$$

3.3 自注意力机制

为了模拟季节项部分时间序列间的依赖关系,加入通道自注意力机制。将季节项不同通道间的序列作为令牌,直接发送到用于学习的Transformer编码器中,如图5所示,其中不同通道所在的变量,与令牌的生成一一对应。与补丁式相比,全序列的嵌入方式保存了变量的所有语义信息,更适合对序列间的依赖性进行建模。

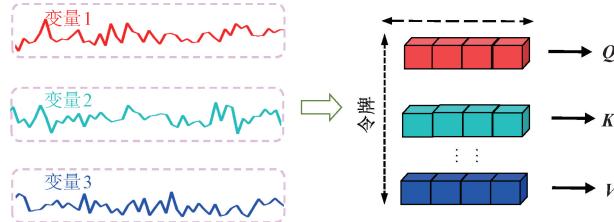


图5 序列间依赖关系作用图

Fig. 5 Diagram of inter-sequence dependency relationships

首先,将季节项 $X_{Seasonal}$ 不同通道的时间序列 $X_{Seasonal}[i, :] \in R^{1 \times D}, i = 1, 2, \dots, N$ 作为令牌。随后,将所有令牌发送到编码器中进行学习,获得 $X_{inter} \in R^{N \times D}$, 计算公式为:

$$\begin{cases} Q = X_{Seasonal} W + b \\ K = X_{Seasonal} W + b \\ V = X_{Seasonal} W + b \\ W \in R^{D \times D} \\ b \in R^{1 \times D} \end{cases} \quad (8)$$

$$Attn = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V$$

$$H = \text{LayerNorm}(X_{Seasonal} + Attn)$$

$$X_{inter} = \text{LayerNorm}(FFN(H) + H)$$

季节项部分的最终输出结果计算公式为:

$$\begin{cases} W \in R^{D \times F} \\ b \in R^{1 \times F} \\ X_{Sout} = X_{inter} W + b \end{cases} \quad (9)$$

最后,通过 $\hat{Y} = X_{Sout} + X_{Tout}$ 对趋势项部分和季节项部分进行叠加,获得最终的预测结果。

4 双阈值计算

故障阈值的确定是故障管理中的一个关键环节。传统的故障判别往往使用较为单一的判别方式,将固定指标与故障阈值进行比较进而判定故障,忽略了现实风电场中噪声等不可控因素的影响^[20]。本研究提出了一种基于双阈值叠加判定的故障监测方法,通过残差序列有效值和信息熵双重判定的方式,有效剔除干扰信号,减少

了误判现象,进而高效识别故障即将发生的态势并发出预警。

目标变量的异常判别可以通过分析残差来确定。由于目标变量的异常会使残差序列的值发生实质性的变化,并使残差序列的信息熵发生反转。为此,提出了一种残差有效值比较和信息熵比较相结合的双阈值监测方法。该方法使用 R_m 和 E_n 这 2 个指标来监测残差是否异常, R_m 值越大, 检测样本异常的概率越高。 E_n 用于检测样本异常的不确定度^[21], E_n 值越小, 检测样本异常的确定性越强。为了提高检测精度,将目标变量的长时间序列划分为多个短时间段,通过滑动窗口采样进行每个短周期的异常检测,参数计算公式为:

$$\begin{cases} R_m = \sqrt{\frac{1}{N_T} \sum_{i=1}^{N_T} (y_i - y_i^*)^2} \\ E_n = -\sum_{i=1}^{N_S} p_i \ln p_i \\ p_i = \frac{d_i T_{inv}}{T_r}, T_r \geq T_{inv} \end{cases} \quad (10)$$

与此同时,设定了 3 个统计区间来计算每个检测周期的信息熵。这些统计区间由基于验证数据集获得的目标变量残差的箱形图确定,如图 6 所示。

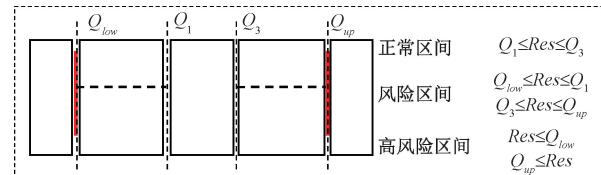


图6 统计区间的划分图

Fig. 6 The division graph of statistical intervals

3 个统计区间从左至右分别为高风险区间、风险区间和正常区间。箱形图的 4 个关键值,即: Q_{low} 、 Q_1 、 Q_3 和 Q_{up} , 它们分别用于划分不同的间隔。其中, Q_1 和 Q_3 分别为残差序列的 0.25 和 0.75 的四分位数; Q_{up} 和 Q_{low} 是箱形图的上限和下限,计算公式为:

$$\begin{cases} Q_{up} = Q_3 + 1.5(Q_3 - Q_1) \\ Q_{low} = Q_1 - 1.5(Q_3 - Q_1) \end{cases} \quad (11)$$

观察 3 个统计区间中包含的样本比例 $p_i (i=1, 2, 3)$ 与对应 E_n 之间的关系,可以发现,当大比例的残差落在同一区域时, E_n 较小,反映了事件发生的强确定性。图 7 表示出了异常检测过程,包括 2 个判断步骤。1) 确定检测时段的 R_m 是否超过阈值 H_{R_m} 。不超过则未发生异常,否则,需要基于另一判断 E_n 来确定由偶发噪声还是真实故障引起。2) 若 E_n 小于给定阈值 H_{E_n} , 则 R_m 超过 H_{R_m} 判定为由真实故障引起。 H_{R_m} 和 H_{E_n} 的值则是基于正常操作数据获得的。首先,根据验证数据集计算目标变量

的残差。其次,通过滑动窗口采样将目标变量时间序列划分为不同的检测周期。然后,计算每个检测周期 R_m 和 E_n 的值。最后,将所有检测周期中的最大 R_m 和 E_n 分别作为 H_{R_m} 和 H_{E_n} 的值。

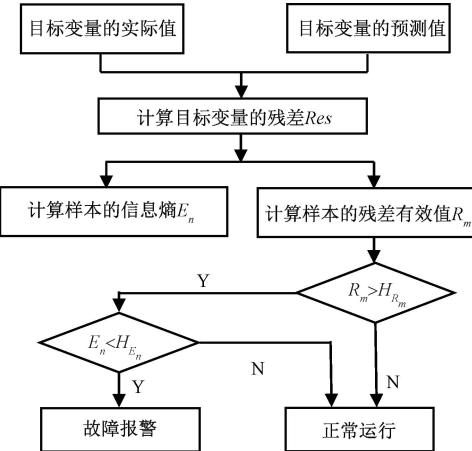


图 7 异常检测的流程

Fig. 7 Flowchart of anomaly detection

5 风电机组案例分析与讨论

本文使用中国南方某风电场的真实故障案例,通过 6#、14#、16#这 3 台风电机组的 SCADA 数据,验证了提出方法的合理性和优越性。该风电场共有 17 台变速变桨距风电机组,机组的额定功率为 2.0 MW,切入风速为 3 m/s,切出风速为 25 m/s。SCADA 数据是在 2016 年 4 月至 2017 年 7 月期间收集的,用于稳态运行条件,数据采集系统共有 254 个变量,其中非状态指标参数 73 个,采样间隔为 10 min。根据风电场运行日志,选择无故障的自然月作为健康数据,选择传动系统不同的故障数据验证模型的有效性。风机传动系统的详细故障信息如表 1 所示,训练、验证、测试数据集的划分信息如表 2 所示。

表 1 风机传动系统故障信息

Table 1 Fault information of the WT drive train system

风机故障系统	故障描述	故障时间
14#风机的传动系统	发电机转子侧集电环烧毁	2017-05-15T14:29—2017-05-18T12:45
6#风机的传动系统	齿轮箱油箱温度过高	2017-02-15T00:05—2017-02-15T12:36

5.1 健康数据集的验证

为了评估所提出的方法的预测性能,本节使用了风机正常运行下的健康 SCADA 数据。选取 16# 风机 2017 年 3 月 7 日到 3 月 16 日约 1 335 个数据样本作为验证数据集。根据风机相关领域知识^[22-23],与传动系统相

关的 SCADA 变量如表 3 所示,这些变量的皮尔逊相关热力图如图 8 所示。

表 2 训练、验证和测试数据集的信息

Table 2 The information of training, validation and testing datasets

数据集	时间	数据样本个数	采样频率/min
训练集 验证集	14#: 2016-04, 2016-06, 2016-08, 2016-10 16#: 2016-05, 2016-06, 2016-08, 2016-11, 2016-12, 2017-02, 2017-03, 2017-04	25 867	10
案例 1 的测试集	14#: 2017-04-25—2017-05-20	3 691	10
案例 2 的测试集	6#: 2017-01-27—2017-02-20	3 457	10

表 3 传动系统相关的变量

Table 3 Variables related to the drive train system

序号	传动系统相关的变量	单位
1	发电机功率实时值	kW
2	发电机转速	rpm
3	桨叶 1 角度	°
4	桨叶 2 角度	°
5	桨叶 3 角度	°
6	风速	m/s
7	L1 电流	A
8	L2 电流	A
9	L3 电流	A
10	轮毂温度	℃
11	主轴承 A 温度	℃
12	主轴承 B 温度	℃
13	齿轮箱油温	℃
14	齿轮箱驱动端轴承温度	℃
15	齿轮箱非驱动端轴承温度	℃
16	发电机驱动端轴承温度	℃
17	发电机非驱动端轴承温度	℃
18	发电机冷风温度	℃

选择具有高相关性的 7 个变量 $X_1 = \{X_6, X_7, X_{14}, X_{15}, X_{16}, X_{17}, X_{18}\}$ 作为模型的输入,并选择与其他变量相关度较高的齿轮箱油温作为模型输出。预测结果如图 9 所示,其中预测的齿轮箱油温由虚线描绘,实际的齿轮箱油温由实线描绘,图 9 底部示出了 RMSE 结果为 0.017 2, MAE 结果为 0.012 6, R^2 结果为 0.962 2。结果表明,该方法预测误差较小,具有良好的预测性能和鲁棒性。

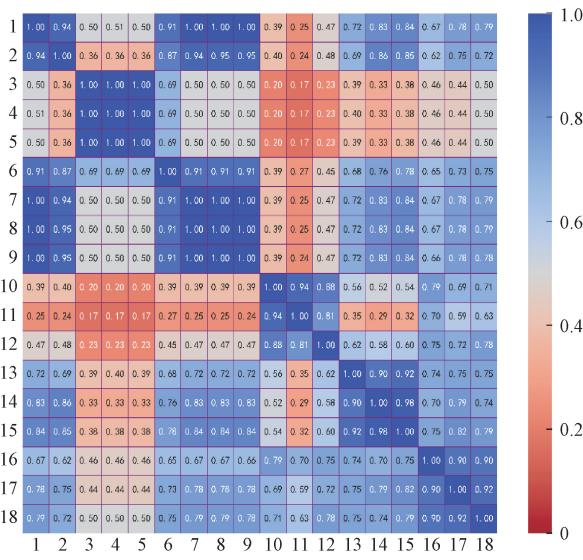


图8 传动系统相关系数矩阵热力图

Fig. 8 Heat map of correlation coefficient matrix of the drive train system

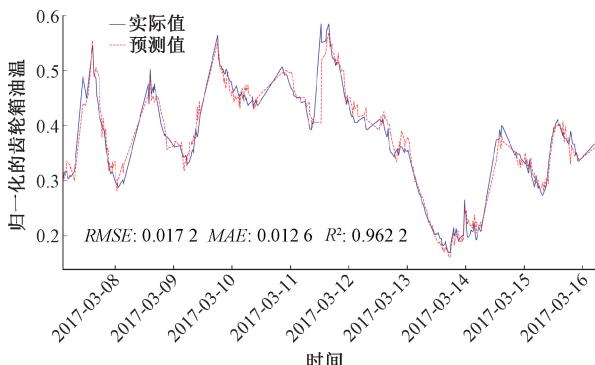


图9 所提出的方法对健康传动系统数据的预测结果

Fig. 9 The predictive results of the proposed method on healthy drive train system data

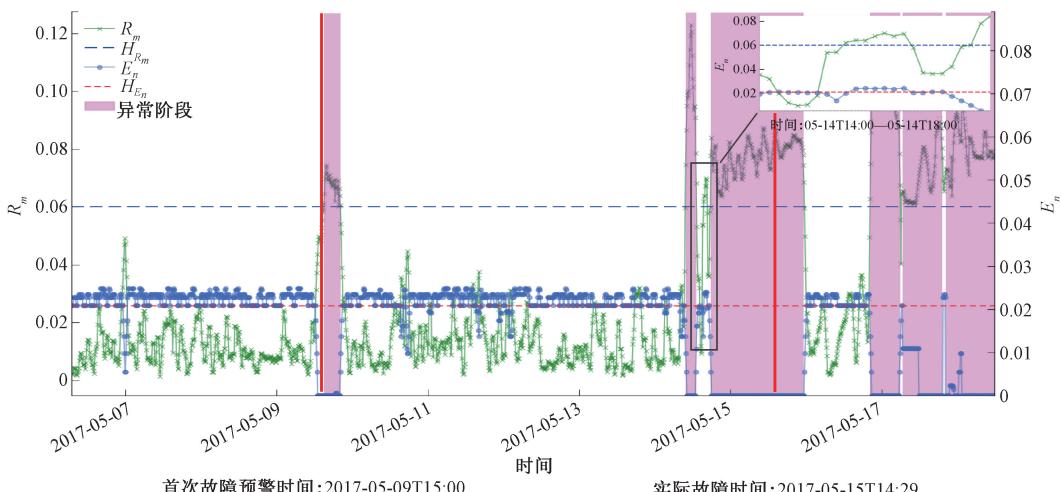


图10 案例1:16号风电机组传动系统故障监测结果

Fig. 10 Case 1: Fault monitoring results of the drive train system of No. 16 WT

5.2 案例1:发电机转子侧集电环烧毁引起传动系统故障

#14风电机组发电机转子侧集电环烧毁引起传动系统故障,严重影响了发电机与传动系统之间的电力传输和信号传递,导致发电机输出功率不稳定,传动效率大幅下降。根据人工维修记录,故障开始时间为2017年5月15日14时29分,如图10中右侧虚线竖线所示,叉号连接的实线和点连接的实线分别代表残差有效值 R_m^e 和残差序列信息熵 E_n^e ,上侧虚线和下侧横虚线分别代表残差的阈值 $H_{R_m}^e$ 和信息熵的阈值 $H_{E_n}^e$ 。当 R_m^e 超过阈值 $H_{R_m}^e$ 且 E_n^e 低于阈值 $H_{E_n}^e$ 时,检测为传动系统故障,即图10中方框区域为机组运行的异常阶段。当 R_m^e 超过阈值 $H_{R_m}^e$ 且 E_n^e 也超过阈值 $H_{E_n}^e$ 时,检测为噪声干扰,机组正常运行,如图10中部分曲线所示,即机组于5月14日14时至5月14日18时为正常运行状态。在本研究中,使用了连续触发机制来提高状态监测结果的可靠性。当6个连续的预测点超过阈值间隔时,识别为故障状态。如图10所示,最早故障报警(以左侧竖线表示)发生在2017年5月9日15时,比人工维护记录约早6天。预测结果证实了所提出的AR-Ledsam用于监测风电机组传动系统状态的有效性。

5.3 案例2:齿轮箱油温过高引起传动系统故障

齿轮箱油温过高引起润滑油性能下降和机械部件磨损加剧,影响齿轮箱内部零件的正常润滑和冷却效果,危害齿轮、轴承等关键部件的使用寿命和可靠性,最终导致传动系统故障。通过所提出的模型对#6风电机组的传动系统故障进行了测试,如图11所示。该故障于2017年2月5日19时50分首次检测到,比记录早了月9.17天。图中部分曲线所示,机组于2月4日16时30分至2月5日3时检测为噪声干扰,机组正常运行。故障区域显示故障区域持续3 h直到故障发生,

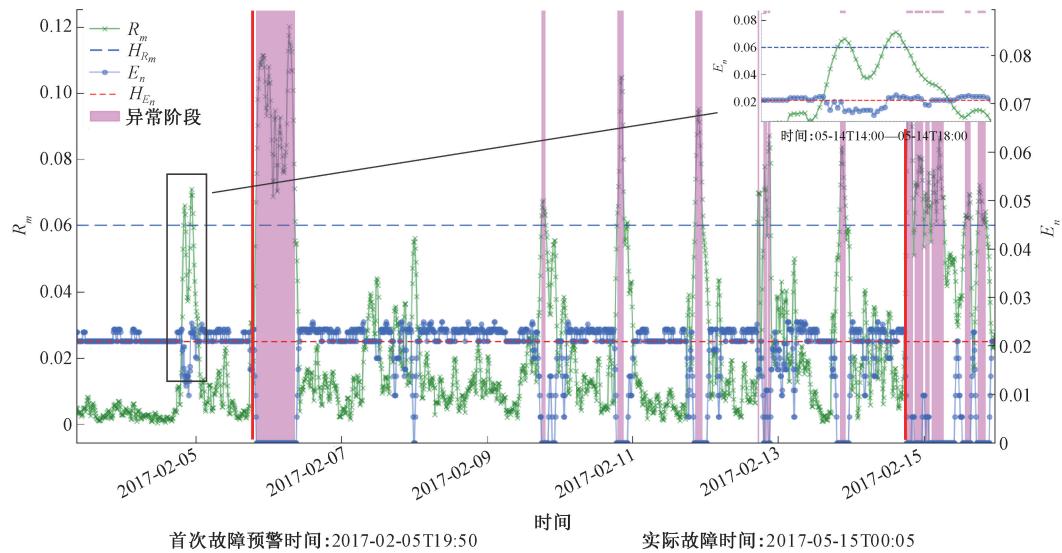


图 11 案例 2:6 号风机传动系统故障监测结果

Fig. 11 Case 2: Fault monitoring results of the drive train system of No. 6 WT

并在故障期间保持该状态。实验结果验证了所提出的 AR-Ledsam 在风电机组传动系统早期故障监测中的有效性。

5.4 不同监测方法对比

为进一步验证所提出方法的优势,本研究分别利用长短期记忆网络 LSTM^[24]、长序列预测的增强型自注意力变换器模型 (infomer, IF)^[25]、时间卷积网络 (temporal convolutional network, TCN)^[26]、稀疏卷积交互网络

(sparse convolutional interaction network, SCINet)^[27] 这 4 种时序预测方法构造温度预测模型, 针对 #14 和 #6 风电机组传动系统故障进行预警, 并基于首次故障预警时间展开对比分析, 如表 4 所示。通过表 4 可以得出: 针对 #14 和 #6 风电机组, 本研究提出的模型较 SCADA 系统分别提前了约 6 和 9.17 天, 相较于其余预测模型提前了约 1~3 天。

表 4 不同模型首次的故障预警时间对比

Table 4 Comparison of the first fault early warning time of different models

预测方法	故障风机	首次故障预警时间	提前预警时间/天
数据采集与监测系统数据	#14	2017-05-15T14:29	—
	#6	2017-02-15T00:05	
长短期记忆网络	#14	2017-05-11T08:20	4.26
	#6	2017-02-08T21:10	6.12
长序列预测的增强型自注意力变换器模型	#14	2017-05-10T13:40	5.03
	#6	2017-02-06T15:20	8.36
时间卷积网络	#14	2017-05-11T4:30	4.42
	#6	2017-02-08T12:10	6.50
稀疏卷积交互网络	#14	2017-05-10T20:30	4.75
	#6	2017-02-07T10:20	7.57
本文方法	#14	2017-05-09T15:00	5.98
	#6	2017-02-05T19:50	9.17

6 结 论

针对风电机组传动系统提升对复杂变化状态的捕捉

能力以及部件维护、剔除白噪声干扰准确率的问题, 提出了一种基于自回归嵌入的可学习分解与自注意力网络的状态监测方法, 其主要贡献在于:

1) 设计了基于动态 AR 的可训练卷积分解模块与自

注意力模型,该模型能够将时间序列数据动态分解为季节性与趋势性,有效捕获变化的动态趋势信息,提高对复杂SCADA数据的建模能力和预测精度。

2)为了提高状态监测的可靠性与鲁棒性,提出了基于残差有效值(R_m)与残差序列信息熵(E_n)相结合比较的双阈值区间监测方法,有效区分噪声与异常信号。

3)通过真实风电场传动系统中发电机和齿轮箱的两个故障案例,验证了所提方法的有效性。

通过实际风场数据验证表明,该方法能够有效识别并剔除白噪声等干扰信号,提前约6~10天检测出传动系统的异常状态,且相较于其他模型在机组故障时能够较早检测出异常。

参考文献

- [1] 高飞,赵开功,戴振冲,等.风电机组偏航断齿问题分析及对策研究[J].能源科技,2025,23(4):77-81.
- GAO F, ZHAO K G, DAI ZH CH, et al. Analysis and countermeasures for pitch gear tooth breakage in wind turbines [J]. Energy Science and Technology, 2025, 23(4): 77-81.
- [2] 刘杰,曹静,赵昕.基于OOB-GWO-SVR的风电机组齿轮箱故障预警[J].电子测量与仪器学报,2022,36(12):97-105.
- LIU J, CAO J, ZHAO X. Wind turbine gearbox fault warning based on OOB-GWO-SVR[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(12): 97-105.
- [3] 周金霞.风力发电装置动力传动系统关键技术研究[D].合肥:合肥工业大学,2018.
- ZHOU J X. Research on key technology of power transmission system of wind power generation device[D]. Hefei:Hefei University of Technology, 2018.
- [4] 柳源,李忠虎,王金明,等.风电机组SCADA“风速-功率”数据处理方法研究[J].太阳能学报,2025,46(7):353-360.
- LIU Y, LI ZH H, WANG J M, et al. Research on the “wind speed-power” data processing method of wind turbine SCADA[J]. Acta Energiae Solaris Sinica, 2025, 46(7): 353-360.
- [5] WANG Z Q, LIU CH L. Wind turbine condition monitoring based on a novel multivariate state estimation technique[J]. Measurement, 2021, 168: 108388.
- [6] JIA X J, HAN Y, LI Y J, et al. Condition monitoring and performance forecasting of wind turbines based on denoising autoencoder and novel convolutional neural networks[J]. Energy Reports, 2021, 7:6354-6365.
- [7] 金晓航,喻轩昂,关汉林.基于回声状态网络的风电机组运行状态监测[J].仪器仪表学报,2025,46(1):258-269.
- JIN X H, YU X ANG, GUAN H L. Wind turbine operation monitoring based on echo state network[J]. Chinese Journal of Scientific Instrument, 2025, 46(1): 258-269.
- [8] 马同旭,刘帅,刘卫亮,等.基于两阶段多维数据生成与实时健康指数的风机齿轮箱故障预警[J].仪器仪表学报,2024,45(11):266-276.
- MA T X, LIU SH, LIU W L, et al. Early warning of wind turbine gearbox failure based on two-stage multidimensional data generation and real-time health index[J]. Chinese Journal of Scientific Instrument, 2024, 45(11): 266-276.
- [9] 何群,尹飞飞,武鑫,等.基于长短期记忆网络的风电机组齿轮箱故障预测[J].计量学报,2020,41(10):1284-1290.
- HE Q, YIN F F, WU X, et al. Fault prediction of wind turbine gearboxes based on long-and short-term memory networks[J]. Acta Metrologica Sinica, 2020, 41(10): 1284-1290.
- [10] 李登峰,黄细霞,吴恩洁,等.基于L-DCGAN的风机齿轮箱时序温度数据生成模型[J].可再生能源,2023,41(2):215-220.
- LI D F, HUANG X X, WU EN J, et al. L-DCGAN-based model for generating time-series temperature data of wind turbine gearbox[J]. Renewable Energy Resources, 2023, 41(2):215-220.
- [11] 陈海燕,任宝民.一种基于时-频特征融合的多变量时间序列预测模型[J/OL].华中科技大学学报(自然科学版),1-13[2025-03-29].
- CHEN H Y, REN B M. A multivariate time series forecasting model based on time-frequency feature fusion[J/OL]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 1-13[2025-03-29].
- [12] 朱昶胜,杨琛,冯文芳,等.基于多层次感知器的高频增强型时间序列预测模型[J/OL].计算机应用,1-12

- [2025-03-29].
- ZHU CH SH, YANG CH, FENG W F, et al. A high-frequency enhanced time series prediction model based on multilayer perceptron [J/OL]. Computer Applications, 1-12 [2025-03-29].
- [13] ZHOU T, MA Z Q, WEN Q S, et al. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting [C]. International Conference on Machine Learning, 2022: 27257-27275.
- [14] LIU Y, WU H X, WANG J M, et al. Non-stationary transformers: Exploring the stationarity in time series forecasting [J]. Advances in Neural Information Processing Systems, 2022, 35: 9881-9893.
- [15] 李宗朋,苏良立,赖鸿波,等. 基于局部离群因子的电力计量数据异常值自动化监测系统 [J]. 自动化与仪表, 2024, 39(11): 137-140.
- LI Z P, SU L L, LAI H B, et al. Automated monitoring system for power metering data outliers based on local outliers [J]. Automation & Instrumentation, 2024, 39(11): 137-140.
- [16] KONG Z Q, TANG B P, DENG L, et al. Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA data by convolutional neural networks and gated recurrent units [J]. Renewable Energy, 2020, 146: 760-768.
- [17] YU G Q, ZOU J, HU X W, et al. Revitalizing multivariate time series forecasting: Learnable decomposition with inter-series dependencies and intra-series variations modeling [J]. ArXiv preprint arXiv:2402.12694, 2024.
- [18] ZHOU H Y, ZHANG SH H, PENG J Q, et al. Informer: Beyond efficient transformer for long sequence-time-series forecasting [C]. 2021 AAAI Conference on Artificial Intelligence, 2021, 35(12): 11106-11115.
- [19] 孙晓永,孙备,郭润泽,等. 面向多无人机协同的多模态目标检测方法 [J]. 仪器仪表学报, 2025, 46(2): 209-220.
- SUN X Y, SUN B, GUO R Z, et al. A multimodal target detection method for multi-UAV collaboration [J]. Chinese Journal of Scientific Instrument, 2025, 46(2): 209-220.
- [20] 宋晔,吴一全. 基于无人机航拍的风力发电机叶片表面缺陷检测综述 [J]. 仪器仪表学报, 2024, 45(10): 1-25.
- SONG Y, WU Y Q. A review of wind turbine blade surface defect detection based on UAV aerial photography [J]. Chinese Journal of Scientific Instrument, 2024, 45(10): 1-25.
- [21] ZENG X J, YANG M, FENG CH, et al. Data-driven anomaly detection method based on similarities of multiple wind turbines [J]. Journal of Modern Power Systems and Clean Energy, 2023, 12(3): 803-818.
- [22] 晋成凤,曹玲燕,沈瑶宇,等. 嵌入式风机传动链状态监测系统研究 [J]. 电工技术, 2024(22): 9-11, 16.
- JIN CH F, CAO L Y, SHEN Y Y, et al. Research on embedded wind turbine drive chain condition monitoring system [J]. Electrotechnology, 2024(22): 9-11, 16.
- [23] 金晓航,杨宇辰,喻轩昂. 改进多层自组织映射网络驱动的风电机组状态监测 [J]. 仪器仪表学报, 2025, 46(3): 231-241.
- JIN X H, YANG Y CH, YU X ANG. Improved multi-layer self-organised mapping network-driven condition monitoring of wind turbines [J]. Chinese Journal of Scientific Instrument, 2025, 46(3): 231-241.
- [24] DELGADO I, FAHIM M. Wind turbine data analysis and LSTM-based prediction in SCADA system [J]. Energies, 2020, 14(1): 125.
- [25] CHEN W H, ZHOU H T, CHENG L SH, et al. Condition monitoring and anomaly detection of wind turbines using temporal convolutional informer and robust dynamic mahalanobis distance [J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 13329105.
- [26] 张佳,关启学,姜月秋. 基于改进 BiGRU-TCN 混合模型的风机轴承温度异常预警方法 [J]. 沈阳理工大学学报, 2025, 44(4): 13-20, 29.
- ZHANG J, GUAN Q X, JIANG Y Q. Early warning method for abnormal temperature of wind turbine bearing based on improved BiGRU-TCN hybrid model [J]. Journal of Shenyang University of Technology, 2025, 44(4): 13-20, 29.
- [27] CAO L X, ZHANG J, QIAN ZH, et al. Condition monitoring of wind turbine based on a novel spatio-temporal feature aggregation network integrated with adaptive threshold interval [J]. Advanced Engineering Informatics, 2024, 62: 102676.

作者简介



曹利宵(通信作者),2020年于北京航空航天大学获得博士学位,现为燕山大学讲师,主要研究方向为风电机组健康管理与智能运维、大型装船机故障智能识别、大数据挖掘与智能算法。

E-mail: caolixiao@ysu.edu.cn

Cao Lixiao (Corresponding author) received her Ph. D. degree from Beijing University of Aeronautics and Astronautics in 2020. She is currently a lecturer at Yanshan University. Her main research interests include wind turbine health management and intelligent operation and maintenance, intelligent identifica-

tion of faults in large shiploaders, and big data mining and intelligent algorithms.



苏保中,2014年燕山大学取得学士学位,现为中海油能源发展装备技术有限公司的资深电气设备运维工程师,主要研究方向为海缆运维、船用高压电气设备检测。

E-mail: subzh@cnooc.com.cn

Su Baozhong received his B. Sc. degree from Yanshan University in 2014. He is currently a senior electrical equipment operation and maintenance engineer at CNOOC EnerTech Equipment Technology Co., Ltd. His main research interests include submarine cable operation and maintenance, and marine high-voltage electrical equipment inspection.