

多工况条件下跨座式单轨列车齿轮箱故障振动信号趋势预测方法

赵玲, 巫刚, 吴杭俊, 张娜, 谭晋

(重庆交通大学信息科学与工程学院, 400074, 重庆)

摘要: 针对低频调制干扰、非线性耦合谐波及工况依赖性特征导致跨座式单轨列车齿轮箱振动信号趋势预测精度低的问题,提出一种融合双向门控循环单元(BiGRU)与时序信息预测模型(Informer)的振动信号趋势预测模型。该模型利用 BiGRU 的门控机制抑制低频干扰,通过双向结构提取非线性耦合谐波特征,并借助自适应学习机制增强对不同运行工况的适应性;通过将 Informer 编码器中的多头注意力机制与空洞因果卷积结合扩大模型感受野,有效捕获时间序列的长期依赖特征,实现齿轮箱振动信号趋势预测。通过实验台采集的跨座式单轨列车齿轮箱故障数据集对模型进行了验证,在不同预测步长下,所提模型的平均绝对误差 R_{MAE} 、均方误差 R_{MSE} 以及均方根误差 R_{RMSE} 分别为 0.264 8、0.116 0 和 0.339 1,均低于自相关 Transformer 模型(Autoformer)、Informer 和分解线性模型(Dlinear);在正常工况、装配误差、内圈故障及疲劳磨损这 4 种工况条件下的预测精度稳定,绝对误差在 0.21~0.35 范围内。研究结果表明,所提模型在长短时间序列下的趋势预测精度均高于其他对比模型,多工况条件下的实验结果验证了所提模型的适应性和鲁棒性。

关键词: 单轨列车;趋势预测;双向门控循环单元;时序信息预测模型;空洞因果卷积

中图分类号: TH165 **文献标志码:** A

DOI: 10.7652/xjtuxb202511019 **文章编号:** 0253-987X(2025)11-0198-11

Vibration Signal Trend Prediction Method for Straddle Monorail Train Gearbox Faults Under Multiple Operating Conditions

ZHAO Ling, WU Gang, WU Hangjun, ZHANG Na, TAN Jin

(School of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400074, China)

Abstract: To address the low prediction accuracy of vibration signal trends in straddle monorail train gearboxes caused by low-frequency modulation interference, nonlinear coupled harmonics, and condition-dependent characteristics, this study proposes a vibration signal trend prediction model that integrates a bidirectional gated recurrent unit (BiGRU) and a temporal information prediction model (Informer). This model utilizes the gating mechanism of BiGRU to suppress low-frequency interference, extracts nonlinear coupled harmonic features through its bidirectional structure, and enhances adaptability to various operating conditions through an adaptive learning mechanism. By combining the multi-head attention mechanism in the Informer encoder with dilated causal convolution, the model effectively expands its receptive field to capture long-term dependency features in time series data, thereby facilitating the trend prediction of gearbox

收稿日期: 2025-03-12。 作者简介: 赵玲(1979—),女,教授,博士生导师。 基金项目: 国家自然科学基金资助项目(6073051);重庆市教育委员会重大科学研究资助项目(KJZD-M202400706,KJQN202200715)。

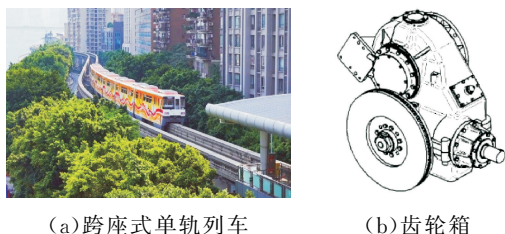
网络出版时间: 2025-06-26

网络出版地址: <https://link.cnki.net/urlid/61.1069.t.20250625.1623.008>

vibration signals. Validation of the model is conducted using a dataset of gearbox fault data collected from a straddle monorail train gearbox test bench. The proposed model achieves average absolute error, mean square error, and root mean square error of 0.264 8, 0.116 0, and 0.339 1, respectively, across different prediction horizons, which are lower than those of the Autoformer, Informer, and Dlinear models. The prediction accuracy remains stable under four operating conditions: normal operation, assembly error, inner ring fault, and fatigue wear, with absolute errors ranging from 0.21 to 0.35. The results indicate that the proposed model demonstrates superior trend prediction accuracy for both short and long time series compared to other benchmark models, and the experimental outcomes under multiple operating conditions validate the model's adaptability and robustness.

Keywords: monorail train; trend prediction; bidirectional gated recirculation unit; temporal information prediction model; dilated causal convolutional

跨座式单轨列车是跨座在单轨道上行进的一种独特的公共交通形式,如图 1 所示。跨座式单轨列车频繁启停、坡道运行过程中,齿轮箱作为单轨列车传动系统核心部件^[1],受到高动态载荷、轨道不平顺冲击等多重因素的影响,容易发生故障。若能基于齿轮箱振动信号实现高精度趋势预测,即可通过预测结果提前识别潜在的故障特征,例如异常振动幅值或频率变化等,进而触发设备状态预警系统,为维护决策提供依据^[2]。



(a) 跨座式单轨列车

(b) 齿轮箱

图 1 跨座式单轨列车与齿轮箱的示意图

Fig.1 Straddle monorail train and schematic diagram of gearbox

跨座式单轨列车齿轮箱振动信号的趋势预测面临诸多挑战。首先,齿轮箱振动信号不可避免的包含轨道冲击成分、连接部件传导带来的大量噪声,而频繁启停、坡道变化等多种复杂工况使得信号动态性和不确定性增加。其次,轨道激励引发的低频调制效应、非线性耦合谐波以及工况依赖性等因素模糊了信号特征边界,进一步增加了预测的难度。如何从复杂的振动信号中提取有效的特征并准确预测变化趋势,成为跨座式单轨列车安全运营中亟待解决的问题^[3]。

现有振动信号趋势预测方法可分为经典方法和深度学习方法两大类。经典方法中,基于统计建模

的方法主要依赖时间序列分析和统计模型,例如自回归积分滑动平均模型(ARIMA)^[4]、灰色预测模型(GM)^[5]和支持向量回归(SVR)^[6]。基于信号处理的方法主要通过时频分析提取振动信号的关键特征,例如小波变换(WT)^[7],经验模态分解(EMD)^[8]和变分模态分解(VMD)^[9]等。但是,经典的方法难以刻画时间序列中的深层非线性关系,对于扰动较大的数据预测效果不理想。近年来,随着深度学习的发展,递归神经网络(RNN)^[10]、图神经网络(GNN)^[11]、门递归单元(GRU)^[12]、长短期记忆(LSTM)^[13]及其变体等已在信号预测领域应用。Ferri 等^[14]利用 LSTM 和 XGBoost 进行风速回归与振动趋势预测,以评估设备健康状态。Li 等^[15]提出了一种趋势约束增量迁移预测(TCITP)方法,提高剩余寿命(RUL)预测的准确性。Jin 等^[16]提出基于 Transformer 的铁路车辆空气压缩机异常检测方法,在实际传感器数据中验证了其适用性。Zhao 等^[17]提出基于改进蜣螂算法优化双向长短期记忆网络(BiLSTM)的变压器故障诊断模型,结合核主成分分析(KPCA)特征提取与 BiLSTM 特性,显著提升了诊断的可靠性与泛化能力。这些模型在短期预测中表现良好,但在长时间序列预测中,其精度和稳定性仍有进步空间。

自相关 Transformer 模型(Autoformer)^[18]、时序信息预测模型(Informer)^[19]、分解线性模型(Dlinear)^[20]等深度学习模型已应用在时间序列和少样本的趋势预测任务中。这些模型能够学习稳定工况条件下振动信号的变化规律并实现趋势预测,但面对跨座式单轨列车齿轮箱多变工况,现有的模型在特征提取时仍无法完全摆脱噪声影响,同时长时间序列趋势预测精度需进一步提升。

为有效提取单轨列车齿轮箱的振动信号数据中的时序特征,提高在长时间序列下的趋势预测效果,本文提出一种结合双向门控循环单元(BiGRU)和Informer的单轨齿轮箱振动信号趋势预测模型。原始振动信号作为模型输入,模型的底层BiGRU首先对振动信号进行处理,捕获时间序列数据的双向短期依赖特征,将特征输出给Informer的编码器和解码器。在Informer编码器部分,利用空洞因果卷积通过扩大感受野来捕获时间序列中的长期依赖关系。解码器对来自编码器的结果进行掩码多头概率稀疏自注意力处理,生成预测序列,输出预测结果。通过不同工况条件下的对比实验,验证了所提模型的准确性与鲁棒性。

1 模型结构设计

1.1 双向门控循环神经网络

传统机器学习方法进行信号趋势预测需要将时间序列数据转换为模型所需的输入结构,难以有效识别特征间的复杂关系,也无法有效捕捉时间序列特征。GRU作为RNN的一种变体,具有天然的序列数据处理能力,能够更好地建立信号间长期依赖关系,可在有效避免梯度消失问题的同时提取长时序特征^[21]。

GRU通过更新门和重置门来控制信息的存储和遗忘,结构如图2所示,其中, x_t 为输入序列; h_t 为隐藏状态; $\sigma(\cdot)$ 为Sigmoid函数; $r(\cdot)$ 为双曲正切激活函数; z_t 和 r_t 分别为更新门和重置门的输出向量; t 为时间步。更新门决定当前隐藏状态应从前一时刻隐藏状态继承多少信息,而重置门则控制是否忽略前一时刻隐藏状态中的某些信息。通过这种门控机制,GRU能够在处理时间序列时更加灵活地适应不同时间步长的输入数据,从而增强捕捉长期依赖性的能力。

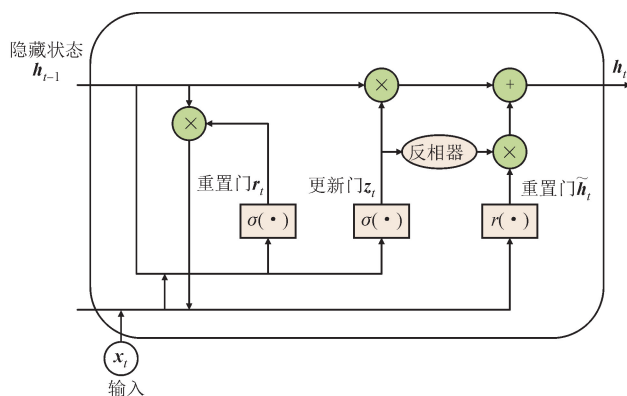


图2 GRU门控循环单元结构

Fig. 2 GRU gated recurrent unit architecture diagram

更新门 z_t 确定新的隐藏状态 h_t 应该如何继承先前的隐藏状态 h_{t-1} , W_z 作为 z_t 的权重矩阵, z_t 的计算式为

$$z_t = \sigma(W_z[h_{t-1}, x_t]) \quad (1)$$

式中: $[h_{t-1}, x_t]$ 表示将前一时刻的隐藏状态 h_{t-1} 和当前时刻的输入 x_t 拼接起来,以便模型可以同时考虑当前的输入和之前的状态信息。重置门 r_t 确定是否忽略先前隐藏状态 h_{t-1} 中的一些信息, W_r 作为 r_t 的权重矩阵, r_t 的计算式为

$$r_t = \sigma(W_r[h_{t-1}, x_t]) \quad (2)$$

新的候选隐藏状态 $\tilde{h}_t$ 是基于先前的隐藏状态 h_{t-1} 和当前输入 x_t 来计算的,但它没有考虑更新门的影响, W 是计算 $\tilde{h}_t$ 的权重矩阵, $\tilde{h}_t$ 的计算式为

$$\tilde{h}_t = \tanh(W[r_t \odot h_{t-1}, x_t]) \quad (3)$$

式中: $\odot$ 为逐元素乘法。最终隐藏状态 h_t 由更新门 z_t 控制,更新门 z_t 确定来自新候选隐藏状态 h_t 和旧隐藏状态 h_{t-1} 的多少信息将被保留, h_t 的计算式为

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (4)$$

BiGRU由两个方向相反的单向GRU组成,前向GRU按时间顺序处理序列,后向GRU按逆序处理序列,最终通过拼接两者的隐藏状态输出结果^[22]。BiGRU双向门控循环单元结构如图3所示,网络结构由输入层、前向隐藏层、后向隐藏层和输出层组成。输入数据 x_t 被同时馈送到前向和后向隐藏层,前向和后向单元格接收输入和前一个时间步骤前向隐藏状态 a_{t-1} 、后向隐藏状态 b_{t-1} ,分别计算当前时间步的前向隐藏状态 a_t 和后向隐藏状态 b_t 。GRU($\cdot$)表示门控循环单元函数,用于计算当前时间步的隐藏状态。 a_t 和 b_t 分别经过权重矩阵 W_a 、 W_b (W_a 、 W_b 分别表示 a_t 和 b_t 的权重矩阵)和偏置向量 c_t 拼接得到最终隐藏状态 h_t 。整个过程表示为

$$a_t = \text{GRU}(x_t, a_{t-1}) \quad (5)$$

$$b_t = \text{GRU}(x_t, b_{t-1}) \quad (6)$$

$$h_t = W_a a_t + W_b b_t + c_t \quad (7)$$

经过BiGRU后,可根据信息的重要性处理不同工况的时序数据,提取出信号本征特征。该特征同时受历史和未来数据的影响,并被输入到下一模块。

1.2 空洞因果多头注意力模块

在时间序列预测任务中,捕捉到长期依赖关系至关重要。传统的自注意力机制可以关注序列中任意两个位置之间的关系,但在长时间序列的处理过程中,其感受野通常局限于较短的时间范围,在处理长序列时难以有效提取远距离的依赖关系^[23]。此外,标准自注意力机制的计算量随着序列长度的增加呈二次方增长,导致大规模数据处理效率低。为

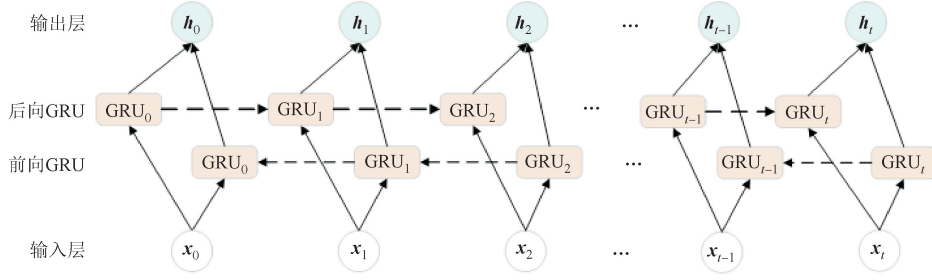


图 3 BiGRU 双向门控循环单元结构

Fig. 3 BiGRU bidirectional gated recurrent unit architecture diagram

了克服这些困难,本文提出了空洞因果多头注意力模块,如图 4 所示。该模块结合空洞因果卷积和多头自注意力机制,利用多头自注意力机制关注输入序列中不同位置的特征捕捉全局信息^[24],空洞因果卷积扩大感受野并提升模型效率的同时保证因果性,避免未来数据对当前时刻预测的影响。

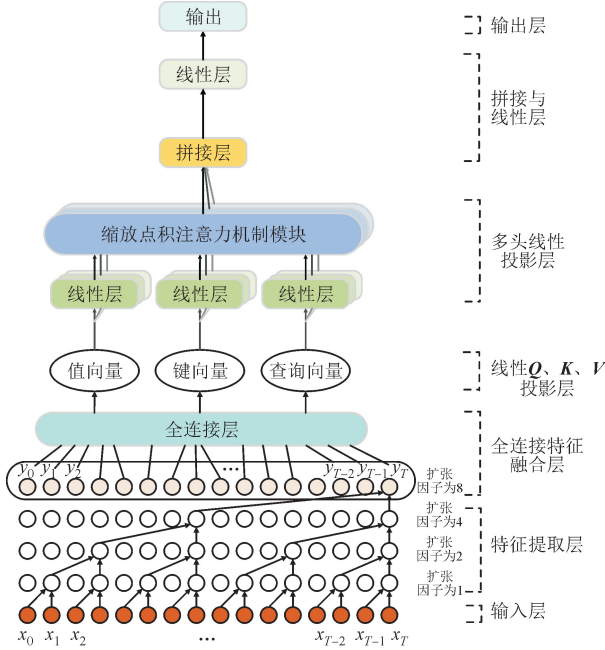


图 4 空洞因果多头注意力模块

Fig. 4 Dilated causal multiple attention module

该模块由 7 层结构组成。输入层中 $x_0, x_1, \dots, x_T$ 表示时序输入序列, T 是序列的长度。输入数据通过特征提取层进行初步的卷积处理,使用空洞因果卷积捕捉输入数据的局部和长距离依赖特征。特征提取层中空洞因果卷积通过引入扩张因子 r 使得卷积核元素之间具有间隔,从而扩大感受野。 y_t 的表达式为

$$y_t = \sum_{k=1}^K w_k x_{t-r(k-1)} \quad (8)$$

式中: y_t 为输出序列在时间步 t 的值; $x_{t-r(k-1)}$ 是输入序列在时间步 $t-r(k-1)$ 的值; w_k 是卷积核的第 k 个元素; K 为卷积核的大小。

全连接特征融合层由空洞因果卷积网络的输出层和全连接层组成, $y_0, y_1, \dots, y_T$ 表示空洞因果卷积网络输出的特征,将这些特征通过全连接层生成查询向量 Q 、键向量 K 、值向量 V 。多头线性投影层中多个并行的注意力头对 Q, K, V 进行处理,捕捉数据中的多种依赖模式后,在缩放点积注意力层中,不同的头分别计算注意力权重并加权聚合特征表示。各个头的输出结果通过拼接层整合,由线性层进行压缩和映射,生成最终的特征表示。

2 基于 BiGRU-Informer 的故障趋势预测方法

2.1 BiGRU-Informer 网络结构

基于 Transformer 的 Informer 模型采用了编码器-解码器结构,并在传统 Transformer 的基础上进行了优化,尤其是在处理时间序列时引入了稀疏自注意力机制和蒸馏自注意力机制,有效提升了时间序列预测任务的效率和准确性^[25],基本架构如图 5 所示。

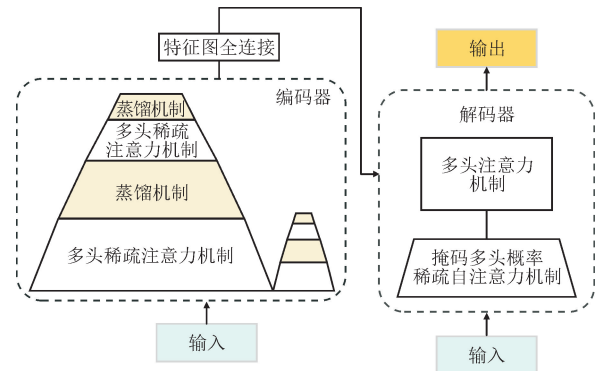


图 5 Informer 网络结构

Fig. 5 Informer network structure

Informer 的编码器和解码器均可接受数据输入。该编码器由多组多头稀疏自注意力层和一个蒸馏层组成^[26]。稀疏自注意力机制是自注意力机制的一种变体,计算式为

$$\text{Atten}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (9)$$

式中: $\text{Atten}(\mathbf{Q}, \mathbf{K}, \mathbf{V})$ 是注意力机制的核心运算,用于动态计算输入序列中不同部分的重要性权重,并基于这些权重对值进行加权聚合; d_k 表示 $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 的维度; $\text{softmax}(\cdot)$ 表示对每行的分数进行归一化,得到注意力权重。在稀疏自注意力机制中,并非所有的点积都对最终结果有显著影响,只有少数点积对主要注意力有贡献。非活跃查询的点积会被值向量的平均值替代,从而减少计算复杂度。

蒸馏操作则是缩短每层中的输入序列长度,减少内存使用。具体计算式如下

$$X_{j+1}^t = \text{MaxPool}(\text{ELU}(\text{Convld}([X_j^t]_{A_B}))) \quad (10)$$

式中: X_j^t 表示第 t 次迭代的第 j 个特征图; $[X_j^t]_{A_B}$ 表示对 X_j^t 进行切片操作,以获取子区域 A_B ; $\text{Convld}(\cdot)$ 表示一维卷积; $\text{ELU}(\cdot)$ 是激活函数,用于在神经网络中引入非线性; $\text{MaxPool}(\cdot)$ 表示最大

池化,这里用作降采样操作。

解码器部分采用与编码器机制类似的结构,由两个相同的多头概率系数自注意力层堆叠而成。这种方法允许在预测序列时的单个步骤中获得结果,提高长期预测效率。

图 6 展示了 BiGRU-Informer 网络结构。将单轨列车齿轮振动信号数据通过滑动窗口方法处理后输入 BiGRU 模型。数据被划分为不重叠的窗口,每个窗口包含一段连续的时间序列数据用于构造模型输入序列。模型编码器的输入是由多个时间窗口组成的历史数据长序列,用于学习时间序列数据的特征。编码器通过多头概率稀疏自注意力机制和空洞因果模块等操作,对输入序列进行特征提取与建模,最终将其转换为用于预测的中间特征表示。此过程用于提取输入数据的特征表示,并对历史数据之间的依赖关系进行建模。解码器接收当前时间步的编码信息,并以与预测时长相同的零序列作为输入,用于生成未来时间步的预测结果。其中零序列作为占位符,使模型在预测期间仅关注当前时间步的历史依赖关系。解码器部分最初使用掩码执行多头概率稀疏自注意力操作,以确保模型在生成预测时仅关注历史数据的相关部分。

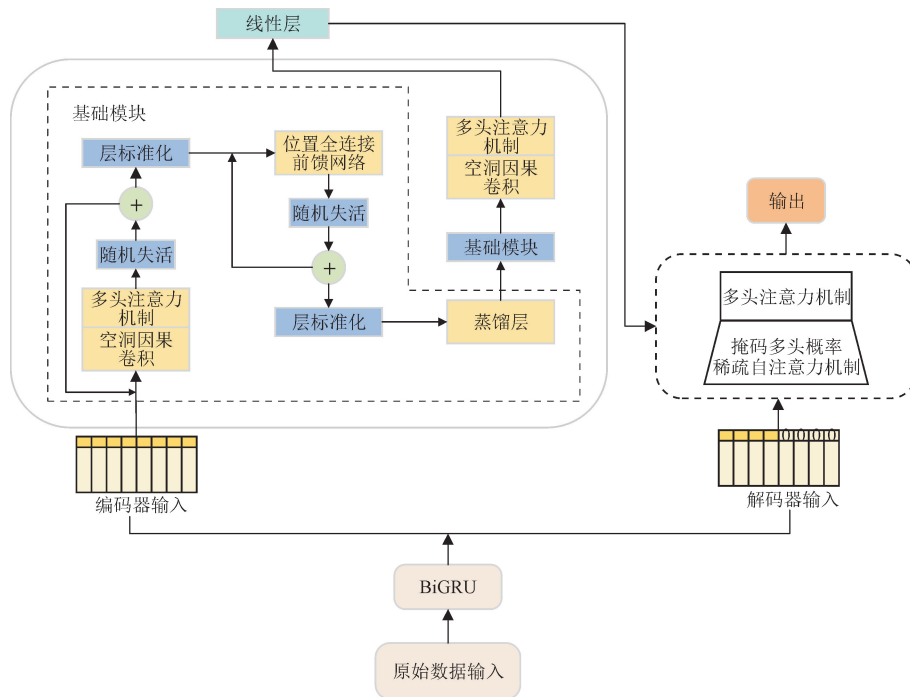


图 6 BiGRU-Informer 网络结构

Fig. 6 BiGRU-Informer network architecture

2.2 预测结果评价指标描述

为定量评估所提出方法对振动信号趋势预测的有效性,利用在实验台采集的 4 种不同工况的数据对模型进行评估和验证。采用 R_{MAE} 、 R_{MSE} 以及 R_{RMSE} 3 个评估指标对预测结果进行评价。具体计算式如下

$$R_{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (11)$$

$$R_{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (12)$$

$$R_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (13)$$

式中: n 为预测长度; y_i 表示实际数据; $\hat{y}_i$ 表示模型预测数据。3 个评价指标的数值越小,表示模型预测性能越好。

3 实验应用验证

3.1 实验数据筛选

本文所用的数据来自研究团队在实验台采集的跨座式单轨列车齿轮箱振动信号数据集。实验中对正常工况、装配误差、内圈故障、疲劳磨损这 4 类工况数据进行采集,采样频率为 10 240 Hz,理论转速、负载和实际转速如表 1 所示。

表 1 跨座式单轨列车齿轮箱转速与负载数据

Table 1 Gearbox speed and load data for monorail trains under different operating conditions

工况类型	理论转速/ ($r \cdot \min^{-1}$)	负载/ ($N \cdot m$)	实际转速/ ($r \cdot \min^{-1}$)
正常工况	2 000	252	1 946
装配误差	2 000	375	1 945
内圈故障	2 000	504	1 945
疲劳磨损	1 000	334	1 000

对 4 种不同工况的振动信号数据进行随机下采样,旨在保留其主要动力学特征的同时,降低数据规模,从而提升模型训练效率、增强泛化能力并减少计算复杂度。由图 7 中 4 幅振动信号时域图可以看出,不同工况下的振动特性存在显著差异。正常工况下,振动信号幅值稳定,呈现较强的周期性波动,无明显的突变或冲击信号。在装配误差工况下,振动幅值有所增加,信号不稳定性明显,并出现低频调制效应,使整体振动趋势发生缓慢变化。对于内圈故障工况,振动信号的不规则性显著增强,幅值增大伴随明显周期性冲击特征,能量集中在特定频段。相比之下,疲劳磨损工况的振动幅值则明显降低,信号整体较为平缓,反映出长期磨损导致机械系统松动、振动能量分布分散,整体振动水平降低。

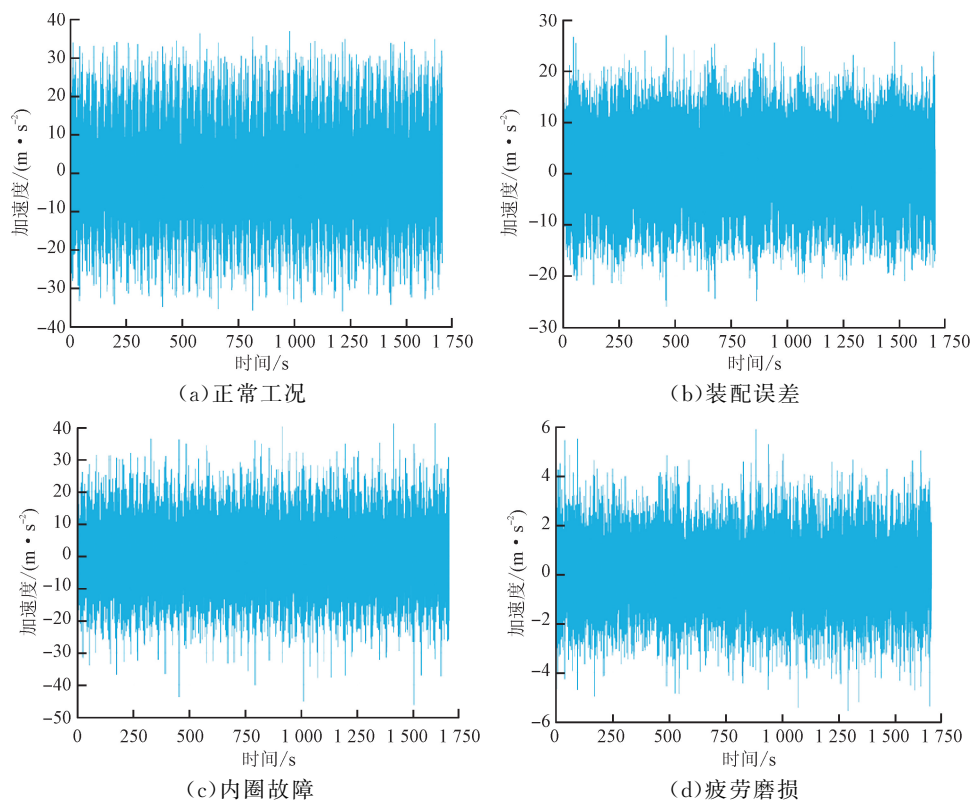


图 7 经过下采样后的不同工况数据时域振动信号

Fig. 7 Time domain vibration signal diagram of different operating conditions after downsampling

3.2 模型训练

为确保模型达到最优性能,实验设计过程中对所有关键超参数进行了细致调整和记录,包括学习率、批量大小、层数和正则化参数等。针对单轨齿轮箱振动信号因频繁启停而表现出的非线性和非平稳特性,首先提取其原始时域特征。在内存受限的条件下,选择批量大小为 32,以平衡模型训练的效率与稳定性。保持学习率为 0.000 1,能加快收敛速度,避免陷入局部最优,隐藏层维度 d_m 设置为 512,前馈网络维度 d_{ff} 设置为 2 024,以满足模型复杂度和平衡性能的需求。选择 GELU 作为激活函数,增强模型的非线性建模能力,提升对复杂数据模式的表达力。选用 R_{MSE} 、 R_{RMSE} 和 R_{MAE} 作为性能评估指标。实验框架构建在 Python 3.8 和 PyTorch 1.9.1 上,采用 NVIDIA A40-48G 显卡的计算机配置。

3.3 模型验证

3.3.1 消融实验

为分析 BiGRU-Informer 模型中不同组件对预测性能的影响,对模型进行了评估。此过程中,将时序信息预测模型 Informer 标记为 F1;Informer 中集成 BiGRU 的模型标记为 F2;单独在 Informer 的编码器中集成空洞因果卷积的架构标记为 F3;本文提出的 BiGRU-Informer 模型标记为 F4,表 2 展示了不同预测模型在模型预测步长为 192 时的消融实验结果。

表 2 模型预测步长为 192 时消融实验的结果

Table 2 Model prediction step is 192, ablation experiment results

预测模型	R_{MAE}	R_{MSE}	R_{RMSE}
F1	0.435 4	0.422 5	0.650 0
F2	0.393 0	0.341 4	0.584 2
F3	0.373 5	0.239 5	0.489 4
F4	0.265 5	0.115 9	0.340 4

注:加粗数据为最优值。

从表 2 中可以看出,F4 模型在 3 个指标上均表现出色, R_{MAE} 、 R_{MSE} 和 R_{RMSE} 分别达到了 0.265 5、0.115 9 和 0.340 4,明显优于其他模型。F4 模型的出色表现充分表明了 BiGRU-Informer 模型在单轨齿轮箱振动信号预测任务中的有效性,相比其他模型,F4 模型更好地捕捉到了振动信号中的复杂时序特性。

3.3.2 对比实验

为了全面评估本文所提模型的有效性,利用实

验台采集的数据集进行对比实验。表 3 呈现 Autoformer、Informer、DLinear、BiGRU-Informer 4 个模型在各个预测步长下的性能表现。

表 3 装配误差工况下不同模型在各预测步长的性能指标

Table 3 Performance comparison of different models at each prediction step under assembly error conditions

预测模型	步长	R_{MAE}	R_{MSE}	R_{RMSE}
Autoformer	48	0.719 8	0.830 1	0.911 1
	96	0.778 5	0.961 4	0.980 5
	192	0.798 4	1.004 9	1.002 4
	384	0.783 7	0.980 6	0.990 2
	768	0.806 8	1.030 0	1.014 9
	1 536	0.817 0	1.036 5	1.018 1
Informer	48	0.464 2	0.349 2	0.590 9
	96	0.509 0	0.422 5	0.650 0
	192	0.496 2	0.401 5	0.633 7
	384	0.492 7	0.396 8	0.911 1
	768	0.502 5	0.410 3	0.980 5
	1 536	0.636 3	0.640 9	1.002 4
DLinear	48	0.413 1	0.284 8	0.990 2
	96	0.435 4	0.317 6	1.014 9
	192	0.453 0	0.346 1	1.018 1
	384	0.478 2	0.386 6	0.590 9
	768	0.506 9	0.420 2	0.650 0
	1 536	0.501 8	0.409 5	0.633 7
BiGRU-Informer	48	0.210 5	0.076 1	0.911 1
	96	0.232 3	0.088 9	0.980 5
	192	0.265 5	0.115 9	1.002 4
	384	0.272 4	0.120 5	0.990 2
	768	0.293 2	0.138 7	1.014 9
	1 536	0.314 6	0.159 8	1.018 1

注:加粗数据为最优值。

从表 3 可以看到,在短时间步长预测任务($T=48$ 和 $T=96$)中,BiGRU-Informer 模型表现出色,步长为 48 时的 R_{MAE} 和 R_{MSE} 分别为 0.210 5 和 0.076 1,显著优于其他 3 个模型,显示了其对短步长振动信号的高精度预测能力。在中等时间步长预测任务($T=192$ 和 $T=384$)时,BiGRU-Informer 模型依旧保持优势,步长为 192 时 BiGRU-Informer 的 R_{MAE} 为 0.265 5, R_{MSE} 为 0.115 9,远低于其他模

型,展现了其在多步预测任务中的稳定性和鲁棒性。在超长步长预测任务($T=768$ 和 $T=1\,536$)中, BiGRU-Informer 的表现依然突出,步长为 1 536 时的 R_{MAE} 为 0.314 6, R_{MSE} 为 0.159 8,明显领先于其他三者,表明其在处理长时间序列信号时能够更好地捕捉复杂特征,达到最佳预测效果。

图 8 展示了时间步长范围分别为 0~48、0~192、0~768 时, Autoformer、Informer、DLinear, 以及 BiGRU-Informer 模型的可视化预测效果。前 3 幅图中的蓝色曲线表示实际值,橙色曲线表示模型的预测值,第 4 幅图中的红色虚线表示实际值,蓝色实线表示模型的预测值。通过对比这三者的预测曲线,可以直观地了解不同模型在不同步长下的振动信号预测任务中的性能差异。

从图 8 中可以看出,在短时间步长预测任务中, BiGRU-Informer 模型表现优异,能够精准捕捉振动信号的快速变化,展现出其在短时间步长下的卓越预测能力。相比之下, Autoformer 模型的预测值存在明显滞后,无法准确跟踪信号变化,表现相对较弱。虽然 Informer 模型在趋势捕捉方面有所改善,但在剧烈变化的区域仍显不足。与前两个模型相比, DLinear 模型趋势捕捉效果显著提升,但在变化剧烈的波峰和波谷区间,预测结果仍不太理想。在中等时间步长预测任务中, BiGRU-Informers 模型保持了高精度和稳定性,能够应对更为复杂的信号

变化。Autoformer 模型在中等时间步长下滞后更为明显,难以捕捉信号中的快速波动特征。Informer 模型只能捕捉部分平稳区域的趋势,在面对趋势波动时已经难以捕捉。DLinear 模型在平稳区域的表现尚可,但在应对大幅波动时,预测误差依然较大。在超长时间步长预测任务中, BiGRU-Informer 模型仍然可以展现出极强的预测能力,能够精准追踪振动信号的复杂特征,并在整个超长时间序列中保持较低的误差。相比之下, Autoformer 模型几乎难以捕捉信号的趋势变化, Informer 模型只能捕捉部分较为平稳区域的趋势,而 DLinear 模型也只在局部区域表现良好,但随着时间步长的增加,其累积误差显著增加。

从可视化结果和量化指标来看, BiGRU-Informer 模型的低误差、较强的适应性以及优越的长期依赖捕捉能力,使其在复杂的振动信号趋势预测任务中具备更强的鲁棒性和准确性,无论是短时间序列还是长时间序列的预测, BiGRU-Informer 模型均能够提供可靠的预测结果。

3.4 不同工况条件下实验结果分析

为全面评估 BiGRU-Informer 模型的有效性,针对其在不同时间步长上的预测性能进行分析。表 4 汇总了模型在跨座式单轨列车齿轮箱振动信号数据集上的预测性能,涵盖不同故障工况和预测步长的评估指标。

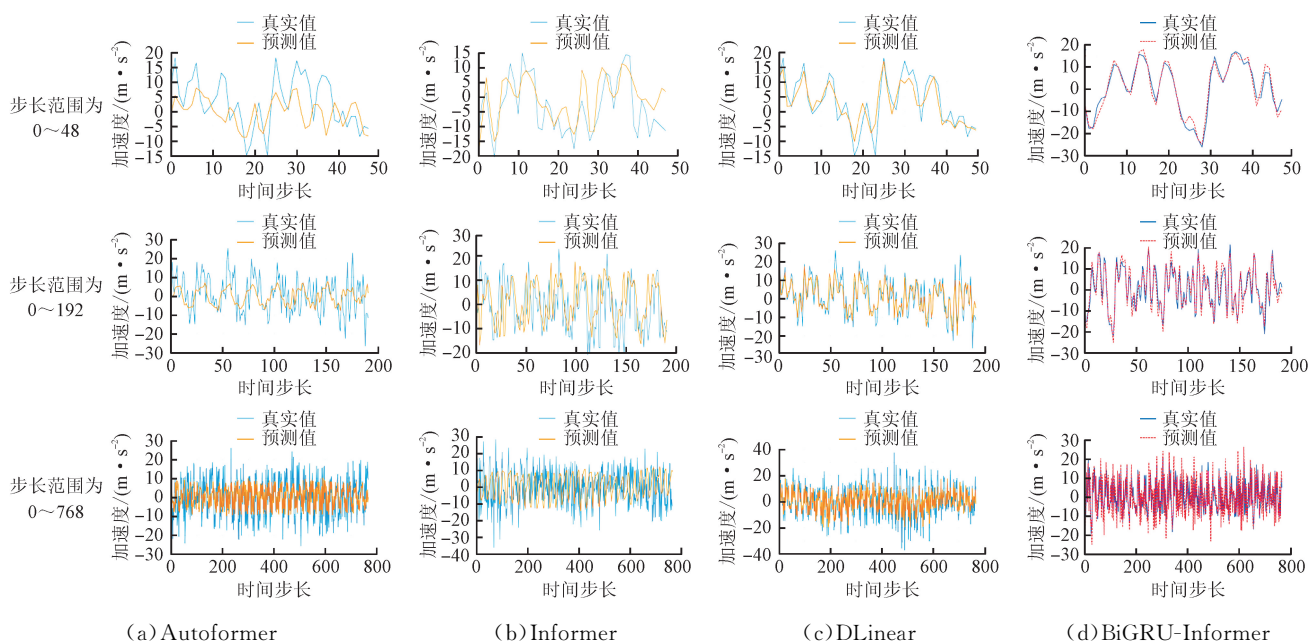


图 8 装配误差工况下步长范围分别为 0~48、0~192、0~768 时各模型预测性能对比

Fig. 8 Comparison of model prediction performance under assembly error conditions with step sizes of 0~48, 0~192, 0~768

表 4 不同工况与预测步长下的 BiGRU-Informer 模型性能指标

Table 4 Performance comparison of the BiGRU-Informer model under different operating conditions and prediction step lengths

故障工况	步长	R_{MAE}	R_{MSE}	R_{RMSE}
正常工况	48	0.208 1	0.069 4	0.263 5
	96	0.316 2	0.154 2	0.392 7
	192	0.351 0	0.191 2	0.437 2
	384	0.354 0	0.193 7	0.443 4
	768	0.353 8	0.195 4	0.440 2
	1 536	0.354 4	0.196 8	0.442 0
装配误差	48	0.210 5	0.076 1	0.275 8
	96	0.232 3	0.088 9	0.298 3
	192	0.265 5	0.115 9	0.340 4
	384	0.272 4	0.120 5	0.347 1
	768	0.293 2	0.138 7	0.373 2
	1 536	0.314 6	0.159 8	0.399 7
内圈故障	48	0.215 0	0.072 8	0.269 9
	96	0.249 6	0.099 0	0.314 7
	192	0.301 3	0.148 9	0.385 8
	384	0.364 9	0.210 6	0.458 9
	768	0.360 3	0.201 3	0.448 6
	1 536	0.385 8	0.194 8	0.441 3
疲劳磨损	48	0.531 8	0.456 5	0.675 6
	96	0.552 5	0.489 3	0.699 5
	192	0.572 0	0.489 3	0.734 4
	384	0.563 8	0.522 1	0.734 4
	768	0.586 4	0.553 7	0.744 1
	1 536	0.608 2	0.586 6	0.765 9

注:加粗数据为最优值。

在正常工况下,振动信号较为平稳,幅值变化范围有限,信号波动较小,随着预测时间步长从 48 增加到 1 536,模型能够较为准确地捕捉到信号的趋势和变化。在装配误差工况下,振动信号相较正常工况不规则、波动较大、信号中包含更多的高频噪声,但是模型能够准确捕捉时序信息实现高精度的趋势预测,说明模型的高效性。在内圈故障工况下,振动信号会出现明显的冲击或周期性变化,信号具有突变特性,这种突变性的信号可能影响模型的稳定性和预测能力,但模型仍然保持较高精度的趋势预测能力。在疲劳磨损工况下,振动信号变化较为平缓且波动幅度小,模型能较稳定地进行预测,但是误差有所增加。

不同工况条件下模型在步长分别为 48、192、768 的预测性能如图 9 所示。实验结果表明,所提模型能够准确地对不规则的、波动较大、信号中包含高频噪声的振动信号进行高精度的趋势预测。对于变化较为平缓且波动幅度小的振动信号,模型也能较稳定地进行预测,但是误差相对较大。该模型通过结合 BiGRU 来捕获时间序列数据的双向短期依赖特征,有效处理振动信号中的动态变化,增强对短期信号波动的敏感性。结合空洞因果卷积神经网络和多头注意力机制来捕获时间序列中的长期依赖关系,并聚焦于最相关的特征,进一步提升了模型的特征提取能力,提高了单轨列车齿轮箱故障振动信号趋势预测的精度。

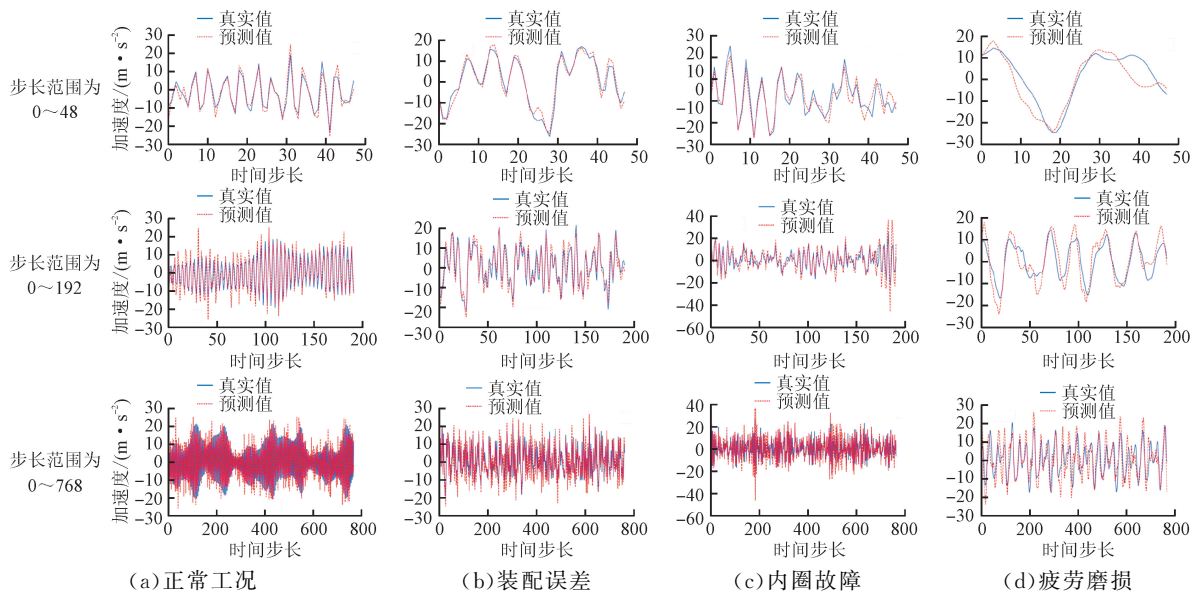


图 9 不同故障工况下步长范围分别为 0~48、0~192、0~768 的 BiGRU-Informer 模型预测性能对比

Fig. 9 Comparison of BiGRU-Informer model prediction performance under different fault conditions with step sizes of 0—48, 0—192, 0—768

4 结 论

针对跨座式单轨列车齿轮箱振动信号的非线性、非平稳性及长时依赖特性,提出了一种融合 BiGRU 与空洞因果卷积结构的改进型 Informer 预测模型。该模型通过 BiGRU 增强对时序特征的提取能力,结合空洞卷积与多头注意力机制扩展感受野,实现对多种工况下振动信号趋势的精准预测,研究结论如下。

(1)模型结构融合提升了振动信号趋势预测的整体精度,优化了对复杂时序特征的感知能力;消融实验结果表明,BiGRU 结构提升了模型对低频干扰的抵抗能力,空洞因果卷积增强了模型对长时间依赖的建模效果。

(2)对比试验结果表明,在多个预测步长下,所提模型 R_{MSE} 、 R_{RMSE} 和 R_{MAE} 指标均优于 Autoformer、Informer 与 DLinear 模型,验证了其预测精度与稳定性。

(3)该模型在面对多种运行工况与序列长度变化时表现出良好的泛化能力。实验结果充分验证了该方法在振动信号预测任务中的有效性。

未来将探索多尺度建模、时频融合特征提取等策略,以进一步增强其在微弱信号场景下的敏感性与可靠性。

参考文献:

- [1] 王国林, 郑州, 张松, 等. 跨座式单轨列车走行轮胎磨损控制方法研究 [J]. 机械工程学报, 2018, 54(6): 78-85.
WANG Guolin, ZHENG Zhou, ZHANG Song, et al. Study on reducing wear of straddle-type monorail trains' running wheel tires [J]. Journal of Mechanical Engineering, 2018, 54(6): 78-85.
- [2] 卞文彬, 邓艾东, 刘东川, 等. 基于改进深度残差收缩网络的风电机组滚动轴承故障诊断方法 [J]. 机械工程学报, 2023, 59(12): 202-214.
BIAN Wenbin, DENG Aidong, LIU Dongchuan, et al. Fault diagnosis method of wind turbine rolling bearing based on improved deep residual shrinkage network [J]. Journal of Mechanical Engineering, 2023, 59(12): 202-214.
- [3] CHEN Yuejian, LI Zihan, JIANG Yuan, et al. Sparse LPV-ARMA model for non-stationary vibration representation and its application on gearbox tooth crack detection under variable speed conditions [J]. Mechanical Systems and Signal Processing, 2025, 224: 112161.

- [4] KHAN S, MUHAMMAD Y, JADOON I, et al. Leveraging LSTM-SMI and ARIMA architecture for robust wind power plant forecasting [J]. Applied Soft Computing, 2025, 170: 112765.
- [5] XU Feng, WANG Jinxin, SAKAI Yuka, et al. Gray-box virtual sensor with constraints for predicting room temperature in cooling and heating modes [J]. Building and Environment, 2025, 273: 112729.
- [6] XU Xiaomin, GUAN Luoyun, WANG Zhiyi, et al. Short-term photovoltaic power intelligent forecasting based on XGBoost-GRU-informer-SVR model [J]. Applied Soft Computing, 2025, 7: 112768.
- [7] SAYEH K F, TAMALOUZT S, SAHRI Y, et al. Artificial intelligence-based direct power control for power quality improvement in a WT-DFIG system via neural networks: prediction and classification techniques [J]. Journal of the Franklin Institute, 2025, 362(1): 107401.
- [8] CAI Xiangjun, LI Dagang, ZHANG Jinglin, et al. MA-EMD: aligned empirical decomposition for multivariate time-series forecasting [J]. Expert Systems with Applications, 2025, 267: 126080.
- [9] GENG Donghan, ZHANG Yongkang, ZHANG Yunlong, et al. A hybrid model based on CapSA-VMD-ResNet-GRU-attention mechanism for ultra-short-term and short-term wind speed prediction [J]. Renewable Energy, 2025, 240: 122191.
- [10] WANG Dongsheng, ZHANG Congcong, LI Ao, et al. Spatio-temporal analysis and prediction for raw water quality of drinking water source by improved RNN algorithm [J]. Journal of Water Process Engineering, 2025, 71: 107164.
- [11] FU Fangze, AI Wei, YANG Fan, et al. SDR-GNN: spectral domain reconstruction graph neural network for incomplete multimodal learning in conversational emotion recognition [J]. Knowledge-Based Systems, 2025, 309: 112825.
- [12] HE Xin, ZHAO Wenlu, GAO Zhijun, et al. Short-term load forecasting by GRU neural network and DDPG algorithm for adaptive optimization of hyperparameters [J]. Electric Power Systems Research, 2025, 238: 111119.
- [13] CHEN Yuejian, LIU Xuemei, RAO Meng, et al. Explicit speed-integrated LSTM network for non-stationary gearbox vibration representation and fault detection under varying speed conditions [J]. Reliability Engineering & System Safety, 2025, 254(Part A): 110596.
- [14] FERRI G G B, VELLOSO B P, AVILA S L, et al.

- Determining the trend behavior of the wind turbine powertrain using mechanical vibration and seasonal wind data [J]. *Energy Reports*, 2025, 13: 353-362.
- [15] LI Xilin, TENG Wei, WANG Luo, et al. Trend-constrained pairing based incremental transfer learning for remaining useful life prediction of bearings in wind turbines [J]. *Expert Systems with Applications*, 2025, 263: 125731.
- [16] JIN M J, KANG C G. Anomaly detection with transformer for a railway vehicle air compressor [J]. *International Journal of Control Automation and Systems*, 2024, 22(8): 2644-2657.
- [17] 赵鑫, 王东丽, 彭泓, 等. 基于多策略改进蜣螂算法优化的变压器故障诊断 [J]. *电力系统保护与控制*, 2024, 52(6): 120-130.
- ZHAO Xin, WANG Dongli, PENG Hong, et al. Transformer fault diagnosis based on a multi-strategy improved dung beetle optimizer [J]. *Power System Protection and Control*, 2024, 52(6): 120-130.
- [18] WU Haixu, XU Jiehui, WANG Jianmin, et al. Autoformer; decomposition transformers with auto-correlation for long-term series forecasting [C]//*Proceedings of the 35th International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2021: 22419-22430.
- [19] ZHOU Haoyi, ZHANG Shanghang, PENG Jieqi, et al. Informer; beyond efficient transformer for long sequence time-series forecasting [C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. Palo Alto, CA, USA: AAAI Press, 2021: 11106-11115.
- [20] MENG Xin, LIU Xingyu, DUAN Hancong, et al. Research on yield prediction model driven by mechanism and data fusion [J]. *Sensors*, 2025, 25(6): 1946.
- [21] ŁAWRYNCZUK M, ZARZYCKI K. LSTM and GRU type recurrent neural networks in model predictive control: a review [J]. *Neurocomputing*, 2025, 632: 129712.
- [22] WANG Qiang, XU Feiyan, HE Jiahua, et al. A new fusion model for enhanced ultra-short-term offshore wind power forecasting [J]. *Renewable Energy*, 2026, 256(Part A): 123876.
- [23] TIAN Xiaoyi, HUANG Siyuan, XIAO Jinjie, et al. SDVS-Net: a spatial dilated convolution and variable self-attention network for multivariate long-term time series forecasting [J]. *Neurocomputing*, 2025, 619: 129148.
- [24] ZHAO Jingyuan, WANG Zhenghong, WU Yuyan, et al. Predictive pretrained transformer (PPT) for real-time battery health diagnostics [J]. *Applied Energy*, 2025, 377(Part D): 124746.
- [25] SHAMS M Y, TAREK Z, ELSHEWEY A M. A novel RFE-GRU model for diabetes classification using PIMA Indian dataset [J]. *Scientific Reports*, 2025, 15(1): 982.
- [26] CAO Xianteng, NIU Huawei, CHEN Zhengqing. Real-time road surface unevenness recognition for ITS using BWO-BiGRU-transformer with adaptive CKF [J]. *IEEE Transactions on Intelligent Transportation Systems*. (2025-02-10) [2025-02-13]. <https://doi.org/10.1109/TITS.2025.3529921>.

(编辑 武红江)