

DOI:10.19651/j.cnki.emt.2519826

# 多特征融合的基于双塔 Transformer 模型 风电功率预测方法<sup>\*</sup>

唐俊萍 暴琳 梁勇

(江苏科技大学自动化学院 镇江 212100)

**摘要:** 为应对环境污染和能源危机,各国大力发展绿色风力发电,但受风速、风向等不可控因素影响,风力发电具有波动性和间歇性,故需对风电功率进行精准预测,保障综合能源系统安全运行。为此,本文提出多特征融合的基于双塔 Transformer 模型风电功率预测方法,该方法深入挖掘含气象信息的风电功率数据,全面抽取特征信息。引入时间维度注意力、经验分布和 K-means 聚类,构建基于时间关联和分布特征的气象模式解耦 Transformer 模型。同时,引入特征维度注意力、希尔伯特变换和季节趋势分解,构建基于特征关联和频域特征的季节趋势分解 Transformer 模型。最后,设计加权平均融合模块,融合上述两个 Transformer 模型预测结果,精准预测风电功率。通过与近 5 年的先进算法在通用公有数据集和企业专有数据集上的大量实验对比,验证了所提方法在预测综合能源系统风电功率时预测准确性和模型解释能力具有显著优势,与次优算法相比,均方误差(MSE)和平均绝对误差(MAE)最大幅度降低了 15.34% 和 6.04%,决定系数  $R^2$  最高提升了 39.33%。

**关键词:** 风电功率预测;特征融合;Transformer 模型;分布特征;频域特征

**中图分类号:** TN91;TM614 **文献标识码:** A **国家标准学科分类代码:** 470.4047

## Wind power prediction model based on a dual-tower transformer model with multi-feature fusion

Tang Junping Bao Lin Liang Yong

(College of Automation, Jiangsu University of Science and Technology, Zhenjiang 212100, China)

**Abstract:** In response to environmental pollution and the energy crisis, countries worldwide are actively advancing green wind power. However, owing to uncontrollable variables such as wind speed and direction, wind power generation is characterized by volatility and intermittency, necessitating accurate forecasting of wind power output to ensure the secure operation of integrated energy systems. To mitigate this challenge, this paper proposes a wind power forecasting method based on a dual-tower Transformer model, which integrates multiple features. This approach thoroughly investigates wind power data enriched with meteorological information, extracting feature information comprehensively. The method incorporates temporal attention, empirical distribution, and K-means clustering to construct a Transformer model that decouples meteorological patterns by considering temporal dependencies and distributional traits. Additionally, feature-level attention, Hilbert transform, and seasonal trend decomposition are integrated to build a Transformer model for seasonal trend decomposition that leverages feature associations and frequency-domain characteristics. Finally, a weighted average fusion module is proposed to combine the predictions of the two aforementioned Transformer models, thereby enhancing the accuracy of wind power forecasting. Extensive experimental comparisons with advanced algorithms from the past five years, conducted on both general public datasets and proprietary datasets from enterprises, reveal that the proposed method markedly outperforms existing techniques in terms of both prediction accuracy and model interpretability. In comparison with the runner-up algorithms, the proposed method reduces the mean squared error (MSE) and the mean absolute error (MAE) by up to 15.34% and 6.04%, respectively, while enhancing the coefficient of determination  $R^2$  by as much as 39.33%.

**Keywords:** wind power forecasting; feature fusion; Transformer model; distribution characteristics; frequency domain characteristics

## 0 引言

能源是社会经济发展的基础。自工业革命以来,化石

燃料一直是主要动力来源,但其需求激增导致能源危机并加剧环境污染<sup>[1]</sup>。为应对资源枯竭与气候变化,各国在 2016 年《巴黎协定》签署下承诺减少温室气体排放,推动可

收稿日期:2025-09-10

<sup>\*</sup> 基金项目:国家自然科学基金(62133015)项目资助

再生能源发展<sup>[2]</sup>。目前,新能源主要包含太阳能、风能、波浪能、潮汐能和地热能等,因其清洁性、可再生性、环境友好性及可持续性,新能源被视为解决能源供应和环境问题的重要途径与方案<sup>[3]</sup>。其中,风能因其低成本、资源分布广泛及可利用程度高而备受关注<sup>[4]</sup>。根据 2025 年 4 月全球风能理事会(Global Wind Energy Council, GWEC)发布的《2025 全球风能报告》显示,2024 年是新增装机创纪录的一年,全球新增风电并网装机容量为 117 GW<sup>[5]</sup>。然而,由于风速、风向和空气湿度等不可控因素影响,风力发电具有较大的波动性和间歇性,若缺乏合理调度及准确预测,容易导致弃风限电现象,造成大量风能无法有效利用<sup>[6]</sup>。因此,提高风电功率预测精度,保障电网调峰调频安全施行是当前综合能源系统面临且急需迫切解决的关键性问题<sup>[7]</sup>。

传统的风电功率预测方法,如:物理模型、统计模型等,难以应对风电功率因天气条件多变而带来的高度波动性,其预测结果往往受限于模型假设的简单性或计算复杂度的制约,在实际应用中面临严峻挑战<sup>[8]</sup>。随着信息技术和计算能力的迅猛发展,深度学习方法具备卓越的数据处理能力和对非线性特征的提取优势,为综合能源系统中风电功率预测任务注入了新的活力<sup>[9]</sup>。

近些年,随着深度学习的突飞猛进,许多基于深度学习的方法被大量用于风电功率预测。Jonkers 等<sup>[10]</sup>提出基于深度卷积神经网络(convolutional neural networks, CNN)的区域风电功率日前预测框架,该框架结合上下文感知风机分布图,并使用保形分位数回归(conformal quantile regression, CQR)生成具有有效置信覆盖的分位数预测。Chen 等<sup>[11]</sup>提出基于二次样条分位数函数的改进时间卷积网络(temporal convolutional network, TCN),该网络利用最大信息系数、高斯相似度和自适应重采样,改进 TCN 构建多步时间序列预测模型 MITCN,同时结合二次样条分位数函数。苏连成等<sup>[12]</sup>提出基于 TCN 残差校正的短期风电功率预测,采取自适应噪声完备集经验模态分解算法(complementary ensemble empirical mode decomposition with adaptive noise, CEEMDAN)分离出风电功率的局部特征信息,以网格搜索与交叉验证算法优化的支持向量回归模型对各分量进行预测,然后构建 TCN 残差预测模型,并使用灰色关联度分析方法选择输入特征,对支持向量回归预测结果进行校正。上述工作主要是基于 CNN 的时间序列预测方法,传统 CNN 在建模时间序列数据方面存在结构上的天然劣势,尽管部分研究如改进 TCN 试图通过引入因果卷积或残差机制加以缓解,但其对时间序列的长期依赖建模能力仍不及循环神经网络与注意力机制。

包广斌等<sup>[13]</sup>提出基于麻雀搜索算法(sparrow search algorithm, SSA)优化的 Transformer-BiGRU(bidirectional gate recurrent unit, BiGRU)组合模型,采用 CEEMDAN 将原始序列分解为多个模态分量和残差分量,结合 Transformer 的自注意力机制与 BiGRU 的双向时序建模

能力,同时引入 SSA 对超参数进行优化以提高风电功率预测精度。Lang 等<sup>[14]</sup>提出两阶段短期风电区间预测方法,搭建基于最小门控内存(minimal gated memory, MGM)网络的子序列数据点模型,结合改进的区间宽度自适应调整策略,获得风电的最终预测区间。Xu 等<sup>[15]</sup>提出基于经验模态分解(empirical mode decomposition, EMD)和深度学习的短期风电功率预测方法,利用 EMD 将历史数据分解为二维张量,通过 CNN 提取数据局部特征,将局部特征输入双向长短期记忆(bidirectional long short-term memory, BiLSTM)网络进行预测。贾睿等<sup>[16]</sup>提出 CNN-LSTM&GRU 多模型组合短期风电功率预测方法,利用 CNN 提取数据局部特征,结合 LSTM 构造出融合局部特征预提取模块的 CNN-LSTM 网络结构,与 GRU 并行。Gong 等<sup>[17]</sup>提出基于 TCN 和 Informer 的混合预测模型,使用 TCN 提取隐藏时间特征,通过 Informer 模型预测风电功率。Xu 等<sup>[18]</sup>提出 FITS 模型,基于傅立叶变换和低通滤波,在复频域内进行插值操作时间序列,结合时域和频域优势。Lin 等<sup>[19]</sup>提出 SparseTSF 模型,采用交叉周期稀疏预测技术,通过下采样原始序列提取周期特征,同时最小化模型的复杂性和参数数量。张登攀等<sup>[20]</sup>提出分段三次 Hermite 插值法(piecewise cubic hermite interpolating, PCHIP)与变分模态分解(variational mode decomposition, VMD)相结合的数据预处理方法,以及 SSA 与 LSTM 相结合的风电功率预测模型。首先采用 PCHIP 法对异常值进行数据修复,接着结合 VMD 法将数据分解为 4 个内在模态分量,最后输入 SSA-LSTM 风电功率预测模型,得到风电功率预测结果。蒲晓云等<sup>[21]</sup>提出结合数据分解技术的 IZOA-Transformer-BiGRU 组合预测模型,利用能量差值法确定 VMD 的子模态数,采用融合 Singer 混沌映射、透镜折射反向学习和单纯形法的改进斑马优化算法(improved zebra optimization algorithm, IZOA)对 Transformer-BiGRU 模型的关键超参数进行优化,最后对分解后的各子序列进行预测,叠加重构得到最终预测结果。上述研究成果主要是基于 LSTM、GRU 和 Transformer 的时间序列预测方法,它们围绕时间序列数据特性构建模型,优化特征提取算法与模块组织方式。然而,其特征维度考虑不够全面,未深度挖掘风电功率数据区别于其他时间序列数据的特性,难以准确拟合相关变量与风电功率之间的映射关系。同时,特征抽取模式单调,注意力机制使用方式单一,模型特征表达能力和功率预测精度有待进一步提高。

Qiu 等<sup>[22]</sup>提出 DUET 模型,在时间和通道维度引入双重聚类增强多元时间序列预测,时间聚类模块将时间序列聚类成细粒度的分布以处理异构的时间模式,通道聚类模块通过度量学习在频域中捕获通道之间的关系,并应用稀疏化减轻噪声通道的不利影响。Zhang 等<sup>[23]</sup>提出 Crossformer 模型,利用跨维度依赖性来进行多元时间序列预测,通过维度-分段嵌入将输入序列嵌入为二维向量数

组,以保留时间和维度信息,其次提出双阶段注意力,用于高效捕获跨时间和跨维度的依赖性。本文所提方法与二者同样运用双塔思想,不同之处在于本文在“双塔”框架下进一步引入时间维度注意力与特征维度注意力的协同机制,综合建模分布特征与频域特征,并从气象模式与季节趋势两个角度对时序进行结构化解耦,以实现更强的特征表达能力与更高的预测精度。

基于此,本文考虑风电功率的波动性和随机性,从气象信息、特征抽取、精细预测和集成学习等多角度切入,深度挖掘众多相关变量及其内在关联,提出了多特征融合<sup>[24]</sup>的基于双塔 Transformer 模型风电功率预测方法。首先,获取含有风速、风向、湿度和降水量等气象信息的风电功率数据集;然后,建立基于时间关联和分布特征的气象模式解耦 Transformer 模型;同时,构建基于特征关联和频域特征的季节趋势分解 Transformer 模型;最后,通过加权平均融合策略,得到风电功率预测结果。基于公有数据集和专有数据集,与其他先进方法进行对比分析,验

证了所提方法在特征表达、预测精度和模型解释能力方面的显著提高。

本文所提方法的贡献点主要包含以下 3 点:

- 1)分别基于经验分布函数<sup>[25]</sup>和希尔伯特变换<sup>[26]</sup>,抽取整体分布特征和瞬时频域特征,增强模型特征表达能力;
- 2)分别解耦气象模式和分解季节趋势成分<sup>[27]</sup>后预测,增强模型预测精度;
- 3)分别建模时间点之间和特征变量之间的相关性,采用集成学习思想,综合考虑两个模型的预测结果,增强模型解释能力和泛化能力。

1 多特征融合的基于双塔 Transformer 模型风电功率预测算法

本文提出的多特征融合的基于双塔 Transformer 模型风电功率预测算法(wind power prediction model based on a dual-tower Transformer model with multi-feature fusion, DTTransFMF),其基本框图如图 1 所示。

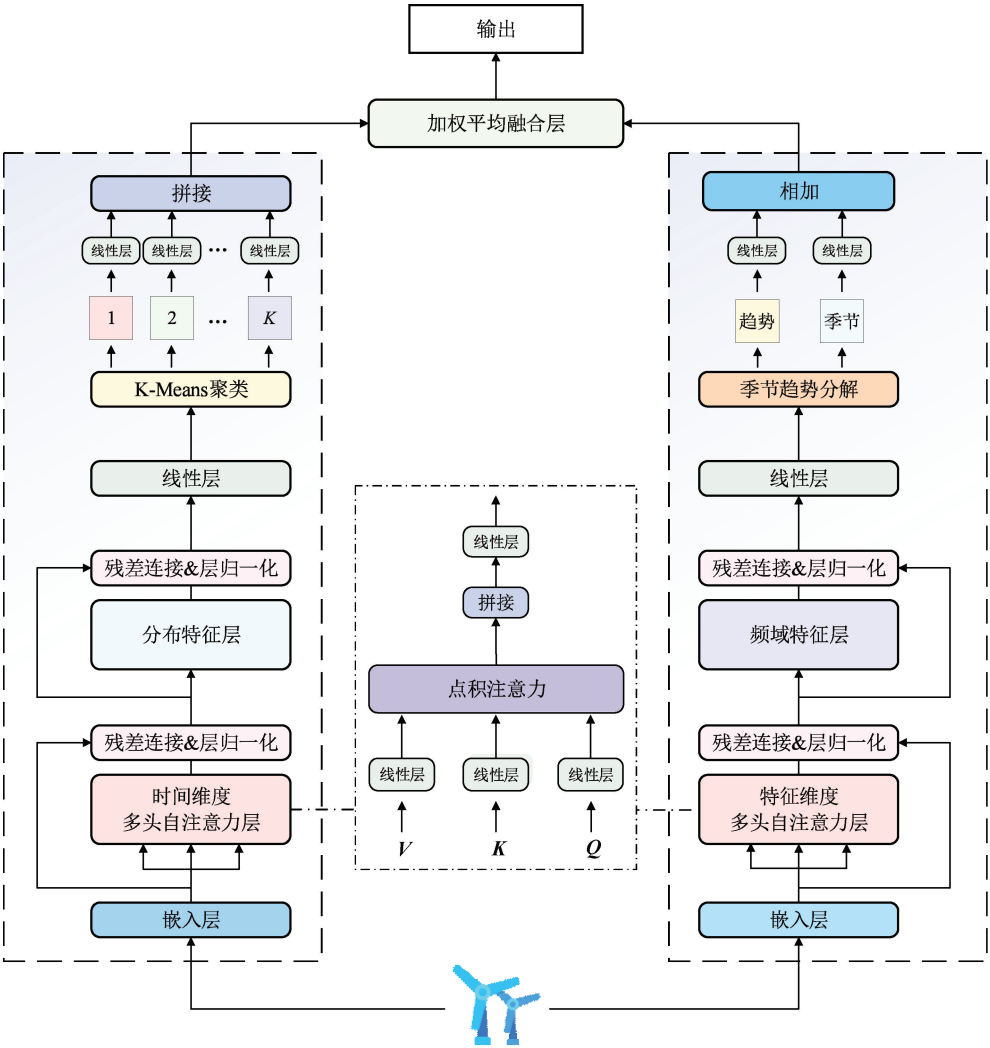


图 1 DTTransFMF 模型框架  
Fig. 1 DTTransFMF model framework

所提方法主要包含如下 3 部分:

1) 基于时间关联和分布特征的气象模式解耦 Transformer 模型:通过多头自注意力机制<sup>[28]</sup>提取时间维度关联关系,其次抽取经验分布特征,然后通过线性层进行初步预测,最后用 K-means 方法<sup>[29]</sup>解耦气象模式并精细预测。

2) 基于特征关联和频域特征的季节趋势分解 Transformer 模型:通过多头自注意力机制提取特征维度关联关系,其次以希尔伯特变换抽取频域特征,然后通过线性层进行初步预测,最后分解趋势成分与季节成分并精细预测。

3) 加权平均融合模块:两个 Transformer 的预测结果通过加权平均融合策略汇聚得到最终的风电功率预测结果。

### 1.1 基于时间关联和分布特征的气象模式解耦 Transformer 模型

#### 1) 嵌入层

基于时间关联和分布特征的气象模式解耦 Transformer 模型的嵌入层输入为多变量时间序列数据  $X \in \mathbf{R}^{S \times N}$ :

$$X = [x_1, x_2, \dots, x_N] \quad (1)$$

其中,  $x_i \in \mathbf{R}^S, i \in \{1, 2, \dots, N\}$ ;  $S$  表示回望窗口长度,即模型每次预测时可参考的历史数据长度;  $N$  表示特征变量数目,其中第一个特征变量为装机容量,最后一个特征变量为风电功率,其余特征变量为风速、风向等气象变量。

嵌入层将输入数据  $X \in \mathbf{R}^{S \times N}$  的特征维度  $N$  经线性变换映射到抽象特征维度  $D$ , 得到嵌入层输出  $X_{L1} \in \mathbf{R}^{S \times D}$ 。

#### 2) 时间维度多头自注意力层

嵌入层的输出  $X_{L1}$  送入时间维度多头自注意力层,捕获各时间步之间的关联关系,进行时间维度的信息交互,抽取更加抽象的特征信息。

首先,通过线性变换,将  $X_{L1}$  投影到 3 个不同的子空间,生成 3 个矩阵  $Q, K, V \in \mathbf{R}^{S \times D}$ :

$$Q = X_{L1} W^Q \quad (2)$$

$$K = X_{L1} W^K \quad (3)$$

$$V = X_{L1} W^V \quad (4)$$

其中,  $W^Q, W^K, W^V \in \mathbf{R}^{D \times D}$  是可学习的投影矩阵。

随后,将特征维度  $D$  分割为  $H$  个注意力头,各注意力头的特征维度为  $d = D \div H$ , 由此得到  $Q_h, K_h, V_h \in \mathbf{R}^{S \times H \times d}$ 。

对于单个注意力头,计算缩放点积注意力,其中注意力分数通过  $Q_h$  和  $K_h$  的点积计算,并除以缩放因子  $\sqrt{d}$  缓解梯度消失问题,单个注意力头的注意力输出  $\text{Att}_h \in \mathbf{R}^{S \times d}$  计算为:

$$\text{Att}_h = \text{softmax}\left(\frac{Q_h K_h^T}{\sqrt{d}}\right) V_h \quad (5)$$

其中,  $h \in \{1, 2, \dots, H\}$ 。

对于各注意力头都进行上述计算,产生  $H$  个注意力头的输出表示,并将这些单个注意力头的注意力输出拼接起来,得到含有时间步之间多头自注意力信息的时间注意力矩阵  $X_{MHATT} \in \mathbf{R}^{S \times D}$ :

$$X_{MHATT} = \text{Concat}(\text{Att}_1, \text{Att}_2, \dots, \text{Att}_H) W^O \quad (6)$$

其中,  $W^O \in \mathbf{R}^{D \times D}$  是时间注意力矩阵的线性投影矩阵,用于将拼接后的时间注意力矩阵映射回原来的  $D$  维空间。

最后,时间注意力矩阵  $X_{MHATT}$  与时间维度多头自注意力层输入  $X_{L1}$  做残差连接后再进行层归一化,得到融合了时间步之间多头自注意力信息的时间维度自注意力层输出  $X_{L2} \in \mathbf{R}^{S \times D}$ :

$$X_{L2} = \text{LayerNorm}(X_{L1} + X_{MHATT}) \quad (7)$$

经过上述时间维度多头自注意力层运算,获取时间步之间的关联关系,充分挖掘了风电功率数据的时间维度信息。

#### 3) 分布特征层

时间维度多头自注意力层的输出  $X_{L2}$  作为分布特征层的输入,对于各特征分别进行经验分布特征计算,抽取时序数据的分布特征。

抽取  $X_{L2}$  的特征维度,获得单个特征的时间序列  $x_{L2} \in \mathbf{R}^S$ , 根据数值对序列进行排序,得到序列  $x_{L2}(\pi(1)) \leq x_{L2}(\pi(2)) \leq \dots \leq x_{L2}(\pi(S))$ , 其中,  $\pi(s) \in \mathbf{R}^S, s \in \{1, 2, \dots, S\}$  是升序排序映射函数,表示从小到大排序后第  $s$  个的数值其在  $x_{L2}$  中的下标索引值。

由此计算该特征下所有时间步的经验分布函数值,获得该特征的经验分布特征向量  $x_{\text{edf}} \in \mathbf{R}^S$ :

$$x_{\text{edf}} = x_{L2}(\pi(s)) = \frac{s}{S} \quad (8)$$

重复上述操作,遍历各特征后,聚合各分布特征向量  $x_{\text{edf}}$ , 得到经验分布特征矩阵  $X_{\text{EDF}} \in \mathbf{R}^{S \times D}$ 。

通过可学习的权重矩阵和偏置向量,对  $X_{\text{EDF}}$  的特征维度进行变换,学习更抽象的深层次特征,获得经过特征维度线性变换的特征加强经验分布矩阵  $X'_{\text{EDF}} \in \mathbf{R}^{S \times D}$ :

$$X'_{\text{EDF}} = \text{GELU}(X_{\text{EDF}} W_D + b_D) \quad (9)$$

$$\text{GELU}(x) = 0.5x \left[ 1 + \tanh\left(\sqrt{\frac{2}{\pi}}(x + 0.047715x^3)\right) \right] \quad (10)$$

其中,  $W_D \in \mathbf{R}^{D \times D}$  和  $b_D \in \mathbf{R}^D$  分别是特征维度的权重矩阵和偏置向量;GELU 函数是激活函数。

再次,对  $X'_{\text{EDF}}$  的时间维度进行变换,捕获更抽象的深层次时间信息,得到经过时间维度线性变换的时间加强经验分布矩阵  $X''_{\text{EDF}} \in \mathbf{R}^{S \times D}$ :

$$X''_{\text{EDF}} = ((X'_{\text{EDF}})^T W_S + b_S)^T \quad (11)$$

其中,  $W_S \in \mathbf{R}^{S \times S}$  和  $b_S \in \mathbf{R}^S$  是时间维度的权重矩阵和偏置向量。



最后,将  $\mathbf{X}_{\text{EDF}}'' \in \mathbf{R}^{S \times D}$  与分布特征层输入  $\mathbf{X}_{L2}$  做残差连接后再进行层归一化,获得融合分布特征的时序数据  $\mathbf{X}_{L3} \in \mathbf{R}^{S \times D}$ :

$$\mathbf{X}_{L3} = \text{LayerNorm}(\mathbf{X}_{L2} + \mathbf{X}_{\text{EDF}}'') \quad (12)$$

经过上述分布特征层运算,捕获时间序列数据的经验分布特征,充分挖掘数据的分布特性。

#### 4) 初步预测层

分布特征层的输出  $\mathbf{X}_{L3}$  送入初步预测层,通过线性投影层进行粗略预测,将特征维度和时间维度映射到预测所需维度,获得初步预测层的输出  $\mathbf{X}_{L4} \in \mathbf{R}^{L \times N}$ :

$$\mathbf{X}_{L4} = [(\mathbf{X}_{L3}\mathbf{W}_1 + \mathbf{b}_1)^T \mathbf{W}_2 + \mathbf{b}_2]^T \quad (13)$$

其中,  $\mathbf{W}_1 \in \mathbf{R}^{D \times N}$  和  $\mathbf{b}_1 \in \mathbf{R}^N$  分别是特征维度的权重矩阵和偏置向量,用于将数据的特征维度从  $D$  映射到实际特征维度  $N$ ;  $\mathbf{W}_2 \in \mathbf{R}^{S \times L}$  和  $\mathbf{b}_2 \in \mathbf{R}^L$  分别是时间维度的权重矩阵和偏置向量,用于将数据的时间维度从  $S$  映射到所需预测的时间长度  $L$ 。

经过上述初步预测层运算,获得风电功率初步预测结果。

#### 5) 气象模式解耦预测层

初步预测层的输出  $\mathbf{X}_{L4}$  作为气象模式解耦预测层的输入,利用 K-means 聚类算法,根据气象特征对输入时序数据的不同时间步进行聚类,通过各时间步的特征将其分配到  $K$  个气象模式中,最小化目标函数  $g(x)$ ,找到聚类中心  $\mathbf{C}_k$  并完成各时间步的聚类:

$$g(x) = \sum_{i=1}^L \|\mathbf{x}_{(i)}^{L4} - \mathbf{C}_{k(i)}\|_2^2 \quad (14)$$

其中,  $\mathbf{C}_k$  表示第  $k$  个聚类中心,  $k \in \{1, 2, \dots, K\}$ ;  $k(i)$  是数据点  $\mathbf{x}_{(i)}$  所属的聚类类别索引。

由此得到表示各时间步对应气象模式的标签向量  $\text{labels} = \{l_1, l_2, \dots, l_L\} \in \mathbf{R}^L$ , 根据标签向量将  $\mathbf{X}_{L4}$  分成  $K$  个子矩阵:

$$\mathbf{X}_k = \mathbf{X}_{L4}[\text{labels}_k], k \in \{1, 2, \dots, K\} \quad (15)$$

其中,  $\mathbf{X}_{L4} = \bigcup_{k=1}^K \mathbf{X}_k$ 。

这里为各气象模式  $k$  定义一个线性预测层  $f_k$ , 用于对属于气象模式  $k$  的子矩阵进行预测,预测函数为:

$$f_k(\mathbf{X}_k) = \mathbf{W}_k \mathbf{X}_k + \mathbf{b}_k \quad (16)$$

其中,  $\mathbf{W}_k \in \mathbf{R}^{N \times N}$  是气象模式  $k$  的线性变换权重;  $\mathbf{b}_k \in \mathbf{R}^N$  是气象模式  $k$  的偏置向量。

由于不同气象模式的数据具备不同特性,上述设计使得各气象模式拥有独立参数,提升风电功率预测的准确性。汇聚各气象模式  $k$  的预测结果,得到气象模式解耦预测层的输出  $\mathbf{Y}_L \in \mathbf{R}^{L \times N}$ :

$$\mathbf{Y}_L = F_{\text{Weather}}(\mathbf{X}_{L4}) = \bigcup_{k=1}^K f_k(\mathbf{X}_k) \quad (17)$$

经过基于时间关联和分布特征的气象模式解耦 Transformer 模型,抽取数据的时间关联关系和经验分布特征,解耦不同气象模式,获得此 Transformer 模型的风电

功率输出结果。

## 1.2 基于特征关联和频域特征的季节趋势分解 Transformer 模型

### 1) 嵌入层

基于特征关联和频域特征的季节趋势分解 Transformer 模型的嵌入层将输入数据  $\mathbf{X} \in \mathbf{R}^{S \times N}$  实际特征维度  $N$  后面拼接上时间戳特征  $\mathbf{X}_{\text{Emb}} \in \mathbf{R}^{S \times E}$ , 同时将数据的实际时间维度  $S$  经线性变化映射到抽象时间维度  $G$ , 转置后得到  $\mathbf{X}_{R1} \in \mathbf{R}^{N' \times G}$ , 其中  $N' = N + E$ 。

### 2) 特征维度多头自注意力层

嵌入层的输出  $\mathbf{X}_{R1}$  送入特征维度多头自注意力层,捕获各特征变量之间的关联关系,进行特征变量维度的信息交互。本层运算过程与 1.1 的第 2) 节类似,故不再赘述,最后得到特征维度自注意力层输出  $\mathbf{X}_{R2} \in \mathbf{R}^{N' \times G}$ , 其融合了特征变量之间的自注意力信息。本层获取特征变量之间的关联关系,充分挖掘了风电功率数据的特征变量维度信息。

### 3) 频域特征层

特征维度多头自注意力层的输出  $\mathbf{X}_{R2}$  送入频域特征层,对于各特征分别进行频域特征计算,抽取时序数据的频域特征。

抽取  $\mathbf{X}_{R2}$  的特征维度,获得单个特征的时间序列  $\mathbf{x}_{R2}(t) \in \mathbf{R}^G$ , 其中  $t \in \{1, 2, \dots, G\}$ , 进行离散时间信号的希尔伯特变换,得到  $\mathbf{H}\{\mathbf{x}_{R2}(t)\}$ :

$$\mathbf{H}\{\mathbf{x}_{R2}(t)\} = \frac{2}{\pi} \sum_{\tau=-\infty}^{+\infty} \frac{\mathbf{x}_{R2}(t - 2\tau - 1)}{2\tau + 1} \quad (18)$$

引入虚数求得  $\mathbf{x}_{R2}(t)$  的解析信号  $\mathbf{z}(t)$ :

$$\mathbf{z}(t) = \mathbf{x}_{R2}(t) + j\mathbf{H}\{\mathbf{x}_{R2}(t)\} \quad (19)$$

进一步,获得信号的瞬时幅度  $\mathbf{A}(t) \in \mathbf{R}^G$  和瞬时相位  $\Phi(t) \in \mathbf{R}^G$ :

$$\mathbf{A}(t) = |\mathbf{z}(t)| = \sqrt{\mathbf{x}_{R2}(t)^2 + \mathbf{H}\{\mathbf{x}_{R2}(t)\}^2} \quad (20)$$

$$\Phi(t) = \arg(\mathbf{z}(t)) = \arctan\left(\frac{\mathbf{H}\{\mathbf{x}_{R2}(t)\}}{\mathbf{x}_{R2}(t)}\right) \quad (21)$$

重复上述操作,遍历各特征后,得到频域特征矩阵,分别为瞬时幅度特征矩阵  $\mathbf{A} \in \mathbf{R}^{N' \times G}$  和瞬时相位特征矩阵  $\Phi \in \mathbf{R}^{N' \times G}$ 。

通过可学习的权重矩阵和偏置向量,分别对频域特征矩阵进行变换,得到经过时间维度线性变换的时间自适应瞬时幅度特征矩阵  $\mathbf{A}' \in \mathbf{R}^{N' \times G}$  和时间自适应瞬时相位特征矩阵  $\Phi' \in \mathbf{R}^{N' \times G}$ :

$$\mathbf{A}' = \text{GELU}(\mathbf{A}\mathbf{W}_G + \mathbf{b}_G) \quad (22)$$

$$\Phi' = \text{GELU}(\Phi\mathbf{W}_G + \mathbf{b}_G) \quad (23)$$

其中,  $\mathbf{W}_G \in \mathbf{R}^{G \times G}$  和  $\mathbf{b}_G \in \mathbf{R}^G$  分别是时间维度的权重矩阵和偏置向量; GELU 函数是激活函数。

随后,进行特征维度线性变换,得到特征自适应瞬时幅度特征矩阵  $\mathbf{A}'' \in \mathbf{R}^{N' \times G}$  和特征自适应瞬时相位特征矩阵  $\Phi'' \in \mathbf{R}^{N' \times G}$ :

$$\mathbf{A}'' = ((\mathbf{A}')^T \mathbf{W}_{N'} + \mathbf{b}_{N'})^T \quad (24)$$

$$\Phi'' = ((\Phi')^T \mathbf{W}_{N'} + \mathbf{b}_{N'})^T \quad (25)$$

其中,  $\mathbf{W}_{N'} \in \mathbf{R}^{N' \times N'}$  和  $\mathbf{b}_{N'} \in \mathbf{R}^{N'}$  分别是特征维度的权重矩阵和偏置向量。

最后, 将特征自适应瞬时幅度特征矩阵  $\mathbf{A}''$  及特征自适应瞬时相位特征矩阵  $\Phi''$  与本层输入  $\mathbf{X}_{R2}$  做残差连接后再进行层归一化, 得到融合频域特征的时序数据  $\mathbf{X}_{R3} \in \mathbf{R}^{N' \times G}$ :

$$\mathbf{X}_{R3} = \text{LayerNorm}(\mathbf{X}_{R2} + \mathbf{A}'' + \Phi'') \quad (26)$$

经过上述频域特征层运算, 捕获时间序列数据的瞬时幅度特征和瞬时相位特征, 增强了对于局部瞬时变化的捕捉精度。

#### 4) 初步预测层

频域特征层的输出  $\mathbf{X}_{R3}$  作为初步预测层的输入, 将数据的时间维度从  $G$  映射到需要预测的时间长度  $L$ , 转置后在特征维度保留前  $N$  维, 得到本层经过初步预测的结果  $\mathbf{X}_{R4} \in \mathbf{R}^{L \times N}$ :

$$\mathbf{X}_{R4} = (\mathbf{X}_{R3} \mathbf{W}_1 + \mathbf{b}_1)^T[:, N] \quad (27)$$

其中,  $\mathbf{W}_1 \in \mathbf{R}^{G \times L}$  和  $\mathbf{b}_1 \in \mathbf{R}^L$  分别是时间维度的权重矩阵和偏置向量。

#### 5) 季节趋势分解预测层

初步预测层的输出  $\mathbf{X}_{R4}$  送入季节趋势分解预测层, 进行季节趋势分解预测。为避免池化操作中滑动窗口计算过程丢失边界信息, 对时间序列数据两端进行对称填充, 定义滑动窗口大小为  $M$ , 取  $P = (M-1)/2$ , 前端填充矩阵  $\mathbf{X}_{\text{front}} \in \mathbf{R}^{P \times N}$  对  $\mathbf{X}_{R4}$  时间维度首元素进行  $P$  次复制, 后端填充矩阵  $\mathbf{X}_{\text{end}} \in \mathbf{R}^{P \times N}$  对  $\mathbf{X}_{R4}$  时间维度末元素进行  $P$  次复制, 合并填充后得到数据  $\mathbf{X}_{\text{padded}} = \text{Concat}(\mathbf{X}_{\text{front}}, \mathbf{X}_{R4}, \mathbf{X}_{\text{end}}) \in \mathbf{R}^{(L+2P) \times N}$ , 填充过程示意图如图 2 所示。

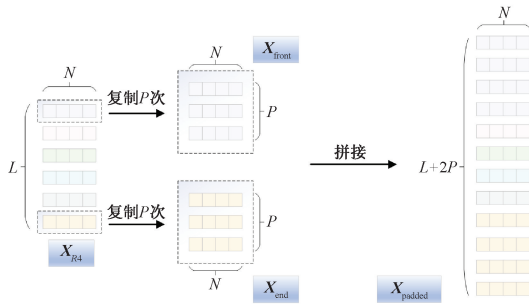


图 2 对称填充

Fig. 2 Symmetrical filling

对填充后的时间序列数据  $\mathbf{X}_{\text{padded}}$  使用一维平均池化操作计算局部平均值, 得到平滑后的时间序列, 即趋势成分  $\mathbf{X}_{\text{trend}} \in \mathbf{R}^{L \times N}$ :

$$\mathbf{X}_{\text{trend}} = \bigcup_{t=0}^L \mathbf{X}_{\text{trend}}[t, :] \quad (28)$$

$$\mathbf{X}_{\text{trend}}[t, :] = \frac{1}{M} \sum_{i=0}^{M-1} \mathbf{X}_{\text{padded}}[t+i, :] \quad (29)$$

进而得到季节成分  $\mathbf{X}_{\text{seasonal}} = \mathbf{X}_{R4} - \mathbf{X}_{\text{trend}} \in \mathbf{R}^{L \times N}$ , 对季节成分  $\mathbf{X}_{\text{seasonal}}$  和趋势成分  $\mathbf{X}_{\text{trend}}$  分别利用线性层进行线性变换, 得到预测季节成分  $\mathbf{X}'_{\text{seasonal}} \in \mathbf{R}^{L \times N}$  和预测趋势成分  $\mathbf{X}'_{\text{trend}} \in \mathbf{R}^{L \times N}$ :

$$\mathbf{X}'_{\text{seasonal}} = (\mathbf{W}_s \mathbf{X}_{\text{seasonal}} + \mathbf{b}_s)^T \quad (30)$$

$$\mathbf{X}'_{\text{trend}} = (\mathbf{W}_t \mathbf{X}_{\text{trend}} + \mathbf{b}_t)^T \quad (31)$$

其中,  $\mathbf{W}_s \in \mathbf{R}^{L \times L}$  和  $\mathbf{b}_s \in \mathbf{R}^L$  分别是季节成分的线性变换权重和偏置项;  $\mathbf{W}_t \in \mathbf{R}^{L \times L}$  和  $\mathbf{b}_t \in \mathbf{R}^L$  分别是趋势成分的线性变换权重和偏置项。

将预测季节成分  $\mathbf{X}'_{\text{seasonal}}$  和预测趋势成分  $\mathbf{X}'_{\text{trend}}$  相加, 得到季节趋势分解预测层的输出结果  $\mathbf{Y}_R \in \mathbf{R}^{L \times N}$ :

$$\mathbf{Y}_R = \mathbf{X}'_{\text{seasonal}} + \mathbf{X}'_{\text{trend}} \quad (32)$$

经过基于特征关联和频域特征的季节趋势分解 Transformer 模型, 抽取数据的特征变量关联关系和瞬时频域特征, 分解季节成分与趋势成分, 获得此 Transformer 模型的风电功率输出结果。

### 1.3 加权平均融合模块

设置两个可学习的权重参数  $\alpha$  和  $\beta$ , 对 1.1 和 1.2 节获得的预测结果  $\mathbf{Y}_L$  和  $\mathbf{Y}_R$ , 采用加权平均计算得到最终预测结果  $\mathbf{Y} \in \mathbf{R}^{L \times N}$ ,  $\mathbf{Y}$  的最后一列数据即为本模型预测的风电功率数据:

$$\mathbf{Y} = \frac{\alpha}{\alpha + \beta} \times \mathbf{Y}_L + \frac{\beta}{\alpha + \beta} \times \mathbf{Y}_R \quad (33)$$

综合考虑时间点间关联关系、特征变量间关联关系、分布特征、频域特征、气象模式和季节趋势等要素, 获得了风电功率的最优预测结果, 为综合能源系统稳定运行及新能源的发展保驾护航。

## 2 实验结果与分析

### 2.1 数据集与实验环境

为了展现本文所提 DTTransFMF 算法的综合性能, 采用通用公有数据集和企业专有数据集, 包括 2024 年数字中国创新大赛中的福建省某风电场风电功率数据集和江苏省东台市某企业数据集。各数据集基本统计信息描述如表 1 所示。

表 1 数据集基本统计信息

Table 1 Basic statistical information of the dataset

| 公有/专有 | 数据集         | 样本数目   | 时间频率/变量数目 |    |
|-------|-------------|--------|-----------|----|
|       |             |        | min       |    |
| 公有    | Spring      | 8 832  | 15        | 12 |
|       | Summer      | 8 832  | 15        | 12 |
|       | Autumn      | 8 736  | 15        | 12 |
|       | Winter      | 8 640  | 15        | 12 |
| 专有    | PrivateData | 11 112 | 60        | 13 |

图 3 是 4 个公有数据集 Spring、Summer、Autumn 和 Winter 中风电功率变量的小提琴图, 小提琴图融合了核密

度估计与箱线图的结构,用于同时呈现数据的概率密度、集中趋势以及异常值,图中灰色圆点代表异常值。

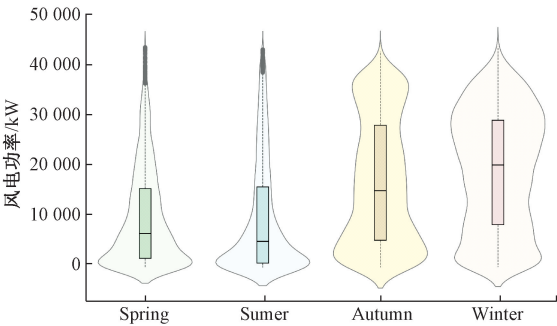


图 3 4 个公有数据集的风电功率小提琴图

Fig. 3 Violin plots of wind power from four public datasets

观察图 3 可以得出:Spring 数据集异常值最多,其灰色圆点集中于上尾部高功率区域,表示大量高于正常范围的极端值,反映了春季风电功率输出的不稳定性极高,存在大量“跳跃”现象;Summer 数据集异常值较少但仍存在,主要集中在上尾部高功率区域,其风电功率的中位数最低;Autumn 数据集无异常值标记,分布对称且核密度平滑,其风电功率的连续性强、稳定性高;Winter 数据集无异常值标记,分布稳定,其风电功率的中位数最高。

实验环境是 Intel(R) Xeon(R) CPU E52683 v4 2.10 GHz 和 8 GB P4,实验平台使用 Python 3.10 开发。实验中采用 4 种评价指标:均方误差 (mean squared error, MSE)<sup>[30]</sup>、平均绝对误差 (mean absolute error, MAE)<sup>[31]</sup>、决定系数 (coefficient of determination,  $R^2$ )<sup>[32]</sup> 和分位数损失 (pinball loss)<sup>[33]</sup>,评估时间序列预测算法的综合性能。

2.2 实验结果与实验分析

为了验证本文所提方法的可行性、有效性及优越性,本节设计了 4 组实验,分别是对比实验、消融实验、参数研究以及计算复杂度分析。实验过程中,所有模型均使用相

同的训练集、验证集和测试集。

1) 对比实验

为了客观比较本文所提方法的综合性能,选择 Informer (2021)<sup>[34]</sup>、FiLM (2022)<sup>[35]</sup>、iTransformer (2024)<sup>[36]</sup>、FITS (2024)<sup>[18]</sup>、SparseTSF(2024)<sup>[19]</sup> 5 个先进预测算法进行对比实验。各对比算法简短描述如下:

Informer:利用改进自注意力机制和生成型解码器,捕捉时间序列中的远程依赖关系。

FiLM:是一种频率改进的 Legendre 记忆模型架构,用于稳健的多尺度时间序列特征提取。

iTransformer:将独立时间序列视为符元,利用注意力机制显式建模各变量间的相关性,提高模型的时间序列表征能力。

FITS:以复值线性层为核心,学习振幅缩放和相移,是一种轻量级的时间序列分析模型。

SparseTSF:通过跨周期稀疏预测技术,对时间序列按照固定周期降采样,聚焦于跨周期的趋势预测。

本文所提 DTTransFMF 算法的实验参数设置如表 2 所示。

表 2 实验参数设置

Table 2 Experimental parameter settings

| 参数              | 参数值 |
|-----------------|-----|
| 多头自注意力头数 $H$    | 4   |
| 天气模式数目 $K$      | 6   |
| 滑动窗口长度 $M$      | 13  |
| Trans1 维度大小 $D$ | 8   |
| Trans2 维度大小 $G$ | 128 |

在关于预测误差的对比实验中,回望窗口长度  $S$  为 96,预测时间长度  $L \in \{48, 96\}$ ,各时间序列预测算法独立运行 10 次,统计平均实验结果如表 3 所示。

表 3 预测误差对比实验结果

Table 3 Comparison experiment results of prediction errors

| 算法           | 评价指标 | Spring       | Sumer        | Autumn       | Winter       | PrivateData  |
|--------------|------|--------------|--------------|--------------|--------------|--------------|
| Informer     | MSE  | 1.363        | 0.704        | 1.500        | 1.066        | 0.726        |
|              | MAE  | 0.974        | 0.571        | 1.050        | 0.824        | 0.672        |
| FiLM         | MSE  | 1.096        | 0.732        | 0.682        | 0.875        | 0.547        |
|              | MAE  | 0.715        | 0.570        | 0.611        | 0.736        | 0.543        |
| iTransformer | MSE  | 0.787        | 0.556        | 0.503        | 0.697        | 0.389        |
|              | MAE  | 0.579        | 0.467        | 0.501        | 0.634        | 0.445        |
| FITS         | MSE  | 0.815        | 0.574        | 0.531        | 0.701        | 0.352        |
|              | MAE  | 0.591        | 0.486        | 0.529        | 0.645        | 0.414        |
| SparseTSF    | MSE  | 0.896        | 0.602        | 0.554        | 0.779        | 0.397        |
|              | MAE  | 0.626        | 0.503        | 0.538        | 0.683        | 0.436        |
| DTTransFMF   | MSE  | <b>0.710</b> | <b>0.540</b> | <b>0.481</b> | <b>0.678</b> | <b>0.298</b> |
|              | MAE  | <b>0.544</b> | <b>0.452</b> | <b>0.489</b> | <b>0.628</b> | <b>0.391</b> |
| Imp.         | MSE  | +0.077       | +0.016       | +0.022       | +0.019       | +0.054       |
|              | MAE  | +0.035       | +0.015       | +0.012       | +0.006       | +0.023       |

表 3 中,最优结果用粗体标注,次优结果用下划线标注,Imp. 表示所提 DTTransFMF 算法与次优算法相比性能指标的提升程度。

观察表 3,可以得出以下结论:

(1) 在所有实验数据集的对比实验中,本文所提 DTTransFMF 算法均取得最优预测准确性和预测精度。在大部分数据集上,iTransformer 算法取得次优结果,这主要是因为含气象信息的数据集各特征变量之间相关性较强,而 iTransformer 算法显式建模各变量之间的关联关系,因而其预测结果相比其他对比预测算法较优。关于均方误差 MSE 评价指标,DTTransFMF 算法在 5 个数据集上相比次优算法分别降低了 9.78%、2.87%、4.37%、2.72%和 15.34%;而对于平均绝对误差 MAE 评价指标,DTTransFMF 算法在 5 个数据集上相比次优算法分别降低了 6.04%、3.21%、2.39%、0.94%和 5.55%。其中,DTTransFMF 算法在 Spring 和 PrivateData 数据集上的 MSE 和 MAE 评估指标的优势显著高于其他 3 个数据集,展现了所提算法对于波动性较强的数据集具备更强建模能力及适应性。

(2) 综合能源系统中的风电功率数据具有明显的非线性和非高斯分布特性,其中春季及其转换阶段天气剧烈变化时的波动性更强。本文所提 DTTransFMF 算法引入经验分布特征模块,增强模型对高偏态与极端值的表达能力,可显著降低预测误差,提高预测精度。同时,所提算法的频域特征提取模块能够有效捕捉风速的多尺度周期性变化,适用于春季日内、周内风况的复杂波动模式;气象模式解耦模块实现风电功率在不同天气状态下的细粒度建模,缓解了跨类别样本对预测性能的干扰。另外,季节转换阶段由冷到热的变化趋势尤为明显,若不加以区分,极易混淆短期波动与长期趋势,从而影响建模效果。所提算法中的季节趋势分解模块将时间序列分解为季节项和趋势项,显式分离季节性周期变化与长期趋势,进而取得了优良的预测效果。所提 DTTransFMF 算法中的各模块设

计共同助力于提高风电功率预测算法的预测准确性和预测精度,提升了对于具有高波动性、复杂周期性和季节转换特征等复杂环境的适应能力。

(3) 实验中,使用了某企业的专有数据集 PrivateData 验证所提算法的实际预测效果及适应性。PrivateData 数据集具有较强的日内周期、显著的季节性波动,以及由气象变化导致的局部突发波动等特性,体现出典型的中频电力负荷场景特征。在 PrivateData 数据集实验中,本文所提 DTTransFMF 算法显著提升了预测准确性,其主要原因可能是模型引入双重多头注意力机制与数据特征之间形成良好适配。其中,所提算法设计了时间点之间的多头注意力机制,动态识别时间序列中不同历史时刻对于当前预测结果的影响强度,捕捉温度、降水量等气象变量的延迟效应与非平稳波动。PrivateData 数据集覆盖周期长,时间序列包含多种时间尺度的信息,如:日内、周内、季节性波动等,多头注意力机制能够在不同尺度下并行建模时间依赖关联关系,极大增强了预测算法对复杂时间序列的建模能力。此外,所提算法引入特征变量之间的多头注意力机制,根据预测时刻的上下文信息自适应分配各气象变量的注意力权重,强化关键变量、削弱冗余特征,有效提升模型的特征表达能力。

综上所述,本文所提 DTTransFMF 算法协同建模两种多头注意力机制,结合频域信息建模、气象模式解耦等结构设计,使得预测模型在应对实际复杂场景时具备更强的感知能力和预测准确性,提高了综合能源系统风电功率预测方法的预测效果和适应性。

为了进一步展示所提 DTTransFMF 算法的优越性,在 5 个对比算法中选取 3 个较为优秀的算法,与 DTTransFMF 算法从解释能力方面即  $R^2$  评价指标进行实验结果对比分析。在关于解释能力的对比实验中,回望窗口长度  $S$  为 96,预测时间长度  $L$  为 48,各预测算法独立运行 10 次,统计平均实验结果,实验结果如表 4 所示,其中最优结果用粗体表示,次优结果用下划线标注。

表 4 解释能力对比实验结果

Table 4 Comparison experiment results of explanation ability

| 算法           | Spring       | Summer       | Autumn       | Winter       | PrivateData  |
|--------------|--------------|--------------|--------------|--------------|--------------|
| iTransformer | <u>0.536</u> | <u>0.211</u> | <u>0.310</u> | <u>0.329</u> | 0.856        |
| FITS         | 0.532        | 0.189        | 0.247        | 0.322        | <u>0.875</u> |
| SparseTSF    | 0.453        | 0.127        | 0.192        | 0.209        | 0.846        |
| DTTransFMF   | <b>0.583</b> | <b>0.294</b> | <b>0.354</b> | <b>0.346</b> | <b>0.889</b> |
| Imp.         | +0.047       | +0.083       | +0.044       | +0.017       | +0.014       |

观察表 4,可以得出以下结论:

(1) 在所有公有数据集和专有数据集实验中,所提 DTTransFMF 算法的决定系数  $R^2$  均取得了最优预测结果,相比次优算法分别提高了 8.76%、39.33%、14.19%、

5.16%和 1.60%,其中,在 Summer 和 Autumn 数据集上所表现出的性能提升尤为显著,展现了所提算法对不同季节数据集良好的泛化能力和结构优势,同时对于风电功率数据的解释能力较强,拟合能力优秀。



(2)公有数据集中 Summer 与 Autumn 阶段对应了风电功率波动性较强与季节过度阶段,通常伴随着较为复杂的气象条件、多变量耦合关系增强,以及时间序列中趋势项与周期项的交织,因而对建模时间序列数据提出了更高要求。本文所提 DTTransFMF 算法设计了频域特征提取模块,增强了对风速、风向等变量中潜在周期结构的建模能力,适用于日内变化显著的夏季风况;给出经验分布特征模块,提升了算法对于输出变量分布特性的刻画,缓解了异常值对于模型回归性能的干扰;提出气象模式解耦模块,通过分类建模方式减弱不同天气状态混杂带来的误差积累;设计了季节趋势分解模块,显式拆解长期趋势与季节波动,有助于提升预测算法对目标变量波动的解释能力。

为了展示本文所提 DTTransFMF 算法在不同分位数上的预测性能优越性,以 Spring 数据集为例,回望窗口长度  $S$  为 96,预测时间长度  $L$  为 96,各预测算法独立运行 10 次,统计平均实验结果,实验结果如图 4 和 5 所示。

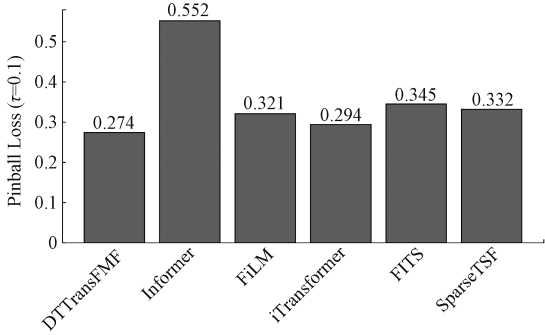


图 4 分位数参数取 0.1 时的 Pinball Loss

Fig. 4 The Pinball Loss when the quantile parameter is set to 0.1

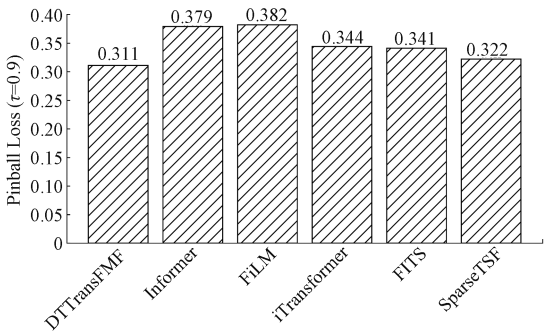


图 5 分位数参数取 0.9 时的 Pinball Loss

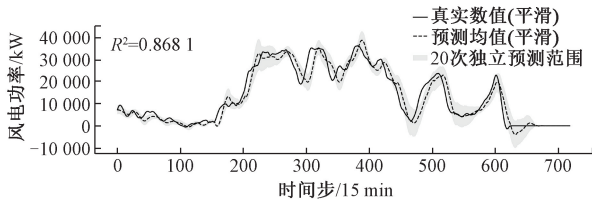
Fig. 5 The Pinball Loss when the quantile parameter is set to 0.9

观察图 4 和 5,可以得出以下结论:

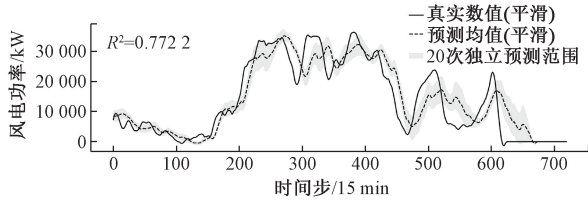
本文所提 DTTransFMF 算法在分位数参数取 0.1 和 0.9 时均表现最好,在低分位点和高分位点上均进行精准预测。这表明模型能有效捕捉含气象信息风电功率数据的非对称分布和异常值,在极端天气、机器故障、维检停机 etc 场景导致的风电功率极端情况下,能够稳定表现出较低

误差,具有较好的鲁棒性。

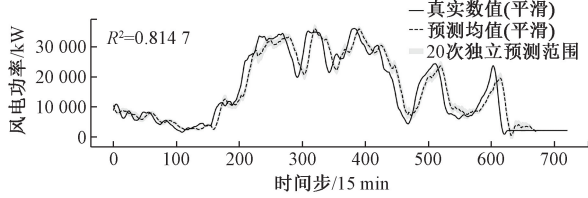
为了更直观地展示本文所提 DTTransFMF 算法的预测性能,以 Spring 数据集为例,回望窗口长度  $S$  为 48,预测时间长度  $L$  为 12,DTTransFMF、Informer 和 SparseTSF 各独立运行 20 次,统计平均实验结果,进行多步预测结果的可视化对比,如图 6 和 7 所示。



(a) DTTransFMF模型预测均值与真实数值的折线图  
(a) The line graph of the predicted mean values of the DTTransFMF model versus the true values



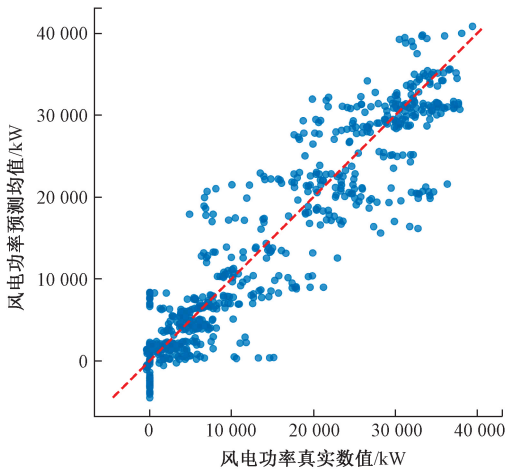
(b) Informer模型预测均值与真实数值的折线图  
(b) The line graph of the predicted mean values of the Informer model versus the true values



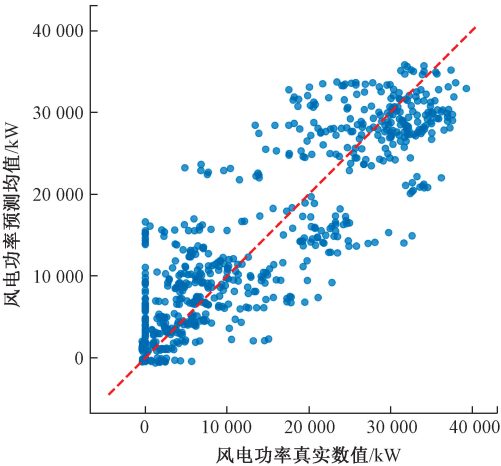
(c) SparseTSF模型预测均值与真实数值的折线图  
(c) The line graph of the predicted mean values of the SparseTSF model versus the true values

图 6 预测均值与真实数值的折线图

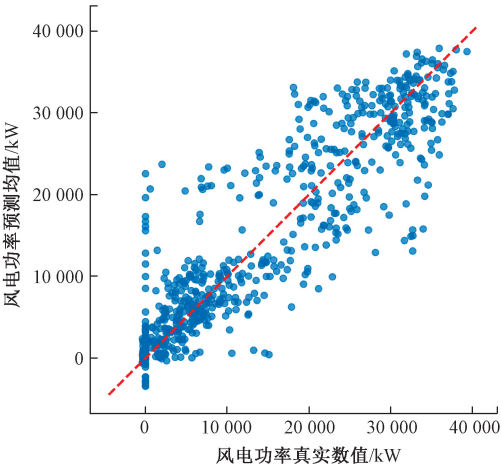
Fig. 6 The line graph of the predicted mean values versus the true values



(a) DTTransFMF模型预测均值与真实数值的散点图  
(a) The scatter plot of the predicted mean values of the DTTransFMF model versus the true values



(b) Informer模型预测均值与真实数值的散点图  
(b) The scatter plot of the predicted mean values of the Informer model versus the true values



(c) SparseTSF模型预测均值与真实数值的散点图  
(c) The scatter plot of the predicted mean values of the SparseTSF model versus the true values

图 7 预测均值与真实数值的折线图

观察图 6,可以发现本文所提模型的预测均值与真实数值拟合较好,真实数值大部分情况下落在 20 次独立预测范围内,并且  $R^2$  最高,模型拟合能力优秀。观察图 7,可以发现本文所提模型的散点图各数据点最为集中,集中分布在  $y = x$  对角线附近,模型预测效果较好。

综上所述,本文所提 DTTransFMF 算法在高波动数据拟合、多时间尺度建模、特征交互表达和趋势扰动剥离等方面表现出较强的适应性,从而在各实验数据集上展现出优秀的预测准确性和解释能力。

2)消融实验

为了验证本文所提 DTTransFMF 算法中各模块的有效性及其必要性,简化 DTTransFMF 算法得到 4 个所提算法变体模型:D\_EDF、D\_Hilbert、D\_Weather、D-Decomp,进行消融实验。其中,D\_EDF 是本文所提算法删除分布特征模块的模型,D\_Hilbert 是本文所提算法删除频域特征模块的模型,D\_Weather 是本文所提算法删除气象模式解耦模块的模型,D-Decomp 是本文所提算法删除季节趋势分解模块的模型。

消融实验中,回望窗口长度  $S$  为 96,预测时间长度  $L \in \{48,96\}$ ,所有预测算法分别独立运行 10 次,统计平均实验结果,消融实验结果如表 5 所示,其最优结果用粗体标注,最差结果用下划线标注。

观察表 5,可以得出以下结论:

(1)在 Spring 数据集上,贡献最大者为分布特征模块。春季冷热交替、风速变化异常,风电功率输出存在“跳跃”现象,所提算法中的分布特征模块能够更好地建模非常态分布和极端值带来的波动,从而提升风电功率预测精度。

(2)在 Summer 数据集上,贡献最大者为频域特征模块。夏季受局部对流、雷雨等因素影响,会有短期周期性大风,所提算法中的频域特征模块能够提取出雷雨天气引起的风速周期变化,有助于时间序列短期预测。

(3)在 Autumn 数据集上,贡献最大者为季节趋势分

表 5 消融实验结果 1

Table 5 Results 1 of the ablation experiment

| 模型         | 评价指标 | Spring       | Summer       | Autumn       | Winter       | PrivateData  |
|------------|------|--------------|--------------|--------------|--------------|--------------|
| D_EDF      | MSE  | <u>0.741</u> | 0.609        | 0.507        | <u>0.752</u> | 0.310        |
|            | MAE  | <u>0.556</u> | 0.496        | 0.505        | <u>0.660</u> | 0.400        |
| D_Hilbert  | MSE  | 0.733        | <u>0.612</u> | 0.497        | 0.731        | 0.354        |
|            | MAE  | 0.554        | <u>0.497</u> | 0.498        | 0.647        | 0.429        |
| D_Weather  | MSE  | 0.722        | 0.597        | 0.499        | 0.719        | <u>0.376</u> |
|            | MAE  | 0.549        | 0.483        | 0.502        | 0.646        | <u>0.446</u> |
| D-Decomp   | MSE  | 0.734        | 0.573        | <u>0.515</u> | 0.712        | 0.347        |
|            | MAE  | 0.554        | 0.475        | <u>0.509</u> | 0.642        | 0.421        |
| DTTransFMF | MSE  | <b>0.710</b> | <b>0.540</b> | <b>0.481</b> | <b>0.678</b> | <b>0.298</b> |
|            | MAE  | <b>0.544</b> | <b>0.452</b> | <b>0.489</b> | <b>0.628</b> | <b>0.391</b> |

解模块。秋季风速和风电功率的变化趋势稳定,所提算法中的季节趋势分解模块更易分解出平滑趋势与噪声项,剥离长期趋势和残差波动,减少预测算法过拟合现象。

(4)在 Winter 数据集上,贡献最大者为分布特征模块。冬季风电功率的输出较为稳定,极少受离群点干扰,所提算法中的分布特征模块其经验分布估计值更加可靠平滑且不受噪声影响,充分发挥经验分布函数的统计功效,有助于风电功率数据的预测。

(5)在 PrivateData 数据集上,贡献最大者为气象模式解耦模块。专有数据集横跨时间长,包含多种天气场景,所提算法中的气象模式解耦模块针对不同天气场景构建独立预测网络,提高了所提算法的准确性及精确度。

为了进一步验证本文所提 DTTransFMF 算法中“双塔”的必要性以及加权平均融合模块的作用,设计 3 个变体模型。D\_T1 变体模型只保留基于时间关联和分布特征的气象模式解耦 Transformer 模型,并在分布特征层后加上频域特征层,删去基于特征关联和频域特征的季节趋势分解 Transformer 模型;D\_T2 变体模型只保留基于特征关联和频域特征的季节趋势分解 Transformer 模型,并在频域特征层后加上分布特征层,删去基于时间关联和分布特征的气象模式解耦 Transformer 模型;D\_G 是本文所提算法中加权平均融合模块替换成门控机制融合模块的变体模型。以 Spring 数据集为例,回望窗口长度  $S$  为 96,预测时间长度  $L$  取 96,3 个变体模型分别独立运行 10 次,统计平均实验结果,消融实验结果如表 6 所示,其最优结果用粗体标注。

表 6 消融实验结果 2

Table 6 Results 2 of the ablation experiment

| 参数    | DTTransFMF | D_T1  | D_T2  | D_G   |
|-------|------------|-------|-------|-------|
| MSE   | 0.857      | 1.115 | 0.902 | 0.873 |
| MAE   | 0.604      | 0.711 | 0.623 | 0.611 |
| $R^2$ | 0.369      | 0.179 | 0.335 | 0.357 |

观察表 6,可以得出以下结论:

(1)与 3 个变体模型相比,本文所提 DTTransFMF 算法性能最为优秀,证明了“双塔”的必要性以及加权平均融合模块的作用。

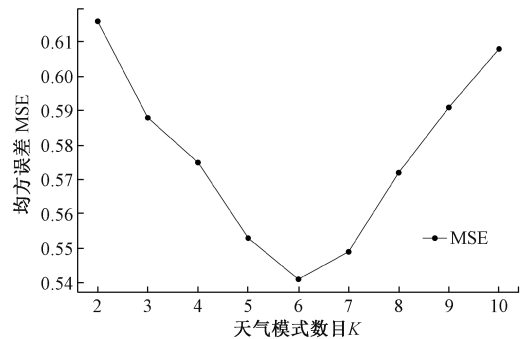
(2)本文所提 DTTransFMF 算法设计双塔的原因是两者关注不同注意力,一者关注时间步之间的依赖关系,即传统序列注意力;一者关注特征变量之间的依赖关系,即跨变量的注意力,基于此,分别抽取分布特征与频域特征。若将双塔变成单塔,可能导致注意力缺失和信息稀释,即缺失时间依赖或者缺失特征依赖,分布特征与频域特征相互稀释。双塔结构可以保持模块独立性,保留双维度独立表征能力,最后对各自结果有效融合形成最终预测,获得更好的效果。

(3)本文所提 DTTransFMF 算法中加权平均融合模块的两个可学习标量能够在训练中自适应地调整两个 Transformer 之间的重要性权重,从而更自然地保持不同特征空间的比例关系,而不引入额外的非线性偏差。相比之下,门控机制虽然灵活,但可能在特征间引入冗余耦合,导致信息融合过度或不均衡。

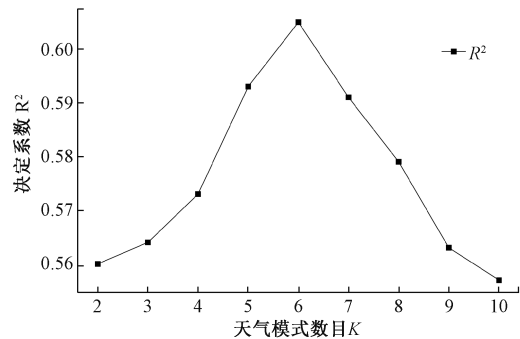
综上所述,所提 DTTransFMF 算法包含经验分布特征模块、瞬时频域特征模块、气象模式解耦模块和季节趋势分解模块,各模块都具备重要作用。所提 DTTransFMF 算法的双塔架构相比单塔能够更有效地捕获多维特征的互补信息,加权平均融合模块在双塔融合方面更为实用,该算法具备优秀的预测准确性和拟合能力。

3) 参数研究

对天气模式数目  $K$  进行研究,以 Spring 数据集为例。回望窗口长度  $S$  为 96,预测时间长度  $L$  取 48,滑动窗口长度  $M = 13$ ,取  $K \in \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$ 。不同  $K$  值的 DTTransFMF 模型在 Spring 数据集上分别独立运行 5 次,统计平均实验结果如图 8 所示。



(a) 不同  $K$  值下 DTTransFMF 模型的 MSE  
(a) MSE of DTTransFMF model under different  $K$



(b) 不同  $K$  值下 DTTransFMF 模型的  $R^2$   
(b)  $R^2$  of DTTransFMF model under different  $K$

图 8 不同  $K$  值下 DTTransFMF 模型的评价指标

Fig. 8 Evaluation indicators of DTTransFMF model under different  $K$

观察图 8,可以得出以下结论:

当天气模式数目  $K$  值设定为 6 时,DTTransFMF 模型在预测性能方面表现最优,均方误差 MSE 达到最低值,决定系数  $R^2$  达到最高值,表明该设定下模型的预测精度与解释能力均达到最优水平。结合所提算法分析,这一现

象可能源于天气模式数目  $K$  值的设定对气象场景解耦程度的影响。当  $K$  值设置过小时,模型无法充分区分不同的气象条件,导致具有显著差异的天气场景被混合建模,从而降低了模型对复杂气象变化的响应能力。相反,当  $K$  值设置过大时,虽然能更细致地刻画气象差异,但容易导致相似或相同的天气场景被冗余地划分为多个模式,造成天气场景的过度解耦。这种冗余划分不仅增加了模型复杂性,也可能削弱模型对同类气象条件下时间序列的整体学习能力。

因此,将天气模式数目  $K$  值设定为 6,在解耦粒度和模式冗余之间取得了较好的平衡。该取值既能有效分离气象条件差异显著的时间片,提升模型对异质天气场景的适应能力,又能避免相似天气场景被重复建模,增强了模型的整体预测精度。

对滑动窗口长度  $M$  进行研究,以 Spring 数据集为例。回望窗口长度  $S$  为 96,预测时间长度  $L$  取 48,天气模式数目  $K = 6$ ,取  $M \in \{7, 13, 25, 49\}$ 。不同  $M$  值的 DTTransFMF 模型在 Spring 数据集上分别独立运行 5 次,统计平均实验结果如图 9 所示。

值,决定系数  $R^2$  达到最大值,表明在该参数配置下,模型的预测精度与解释能力均达到了最优水平。结合所提算法设计分析其原因,滑动窗口长度  $M$  值在季节趋势分解模块中起着关键作用,对趋势项与季节项的提取效果具有直接影响。若  $M$  值设置过小,趋势分量对局部波动的响应将过于敏感,导致原本应当平滑的长期变化趋势被频繁扰动,从而失去对时间序列宏观结构的准确刻画。此外,短窗口还可能致模型将短期的随机扰动误识为趋势信息,进而引发对噪声的过拟合,削弱模型在测试集上的泛化能力。相反,若  $M$  值设置过大,趋势提取过程将过度平滑,不仅可能掩盖原序列中的中短期结构性变化,如突变点、趋势转折等,还会使模型对序列中快速变化的动态特征响应迟缓,影响对非平稳时间序列的建模效果。同时,由于大窗口在序列起始与结束位置难以完整展开,可能造成边界区域的趋势和季节性提取结果不准确,进而影响整体模型的预测性能。

因此,滑动窗口长度为 13 时,在局部扰动的抑制与结构变化的保留之间实现了良好的平衡,能够有效分离长期趋势与短期波动,增强了模型对时间序列内在规律的刻画能力,从而显著提升了预测精度。

4) 计算复杂度分析

为了验证本文所提算法实际落地应用的可行性,对模型的计算复杂度进行分析。以 Spring 数据集为例,回望窗口长度  $S$  为 96,预测时间长度  $L \in \{96, 192, 288, 384, 480, 576\}$ ,统计单次推理中模型的参数量、推理时间、聚类时间、显存峰值和 FLOPs,结果如表 7 所示。

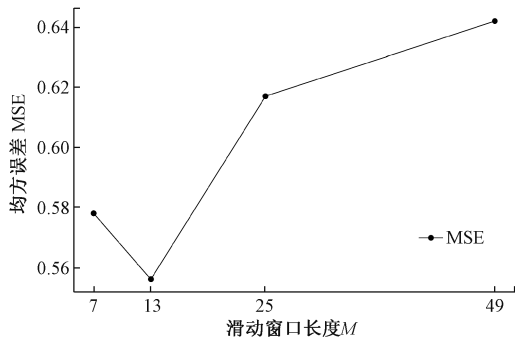
表 7 不同  $L$  值下的计算复杂度相关指标

Table 7 Indicators related to computational complexity under different  $L$  values

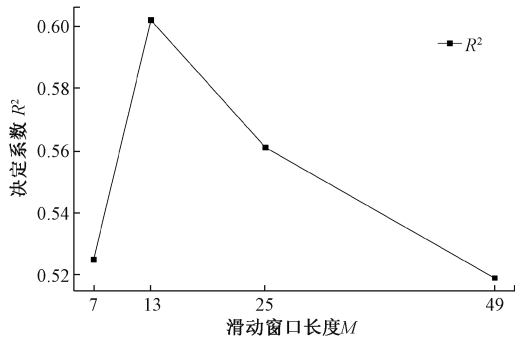
| $L$ 值 | 参数量     | 推理<br>时间/s | 聚类<br>时间/s | 显存<br>峰值/MB | FLOPs      |
|-------|---------|------------|------------|-------------|------------|
| 96    | 147 504 | 0.032      | 0.021      | 552.680     | 2 919 336  |
| 192   | 224 688 | 0.035      | 0.025      | 553.300     | 3 916 200  |
| 288   | 338 736 | 0.040      | 0.030      | 554.210     | 5 355 432  |
| 384   | 489 648 | 0.041      | 0.031      | 555.400     | 7 237 032  |
| 480   | 677 424 | 0.044      | 0.034      | 556.880     | 9 561 000  |
| 576   | 902 064 | 0.048      | 0.038      | 559.070     | 12 327 336 |

观察表 7,可以得出以下结论:

预测长度  $L$  从 96 增至 576,模型的参数量、推理时间、聚类时间、显存峰值和 FLOPs 分别增长了 511.552%、50%、80.952%、1.156%、322.265%, $L$  对参数量和 FLOPs 的影响最大。聚类时间占推理时间的比重较大,且随着  $L$  的增长,比重也逐渐增加。后续工作可着重考虑优化聚类算法,如采用更高效的距离度量方法、减少冗余计算、使用并行化技术等。综合分析,在深度学习模型中,本



(a) 不同  $M$  值下 DTTransFMF 模型的 MSE  
(a) MSE of DTTransFMF model under different  $M$



(b) 不同  $M$  值下 DTTransFMF 模型的  $R^2$   
(b)  $R^2$  of DTTransFMF model under different  $M$

图 9 不同  $M$  值下 DTTransFMF 模型的评价指标  
Fig. 9 Evaluation indicators of DTTransFMF model under different  $M$

观察图 9,可以得出以下结论:

当滑动窗口长度  $M$  值设置为 13 时,DTTransFMF 模型在预测性能方面表现最佳,其均方误差 MSE 达到最小



文所提模型参数量属于中等规模,占用显存较小,所需算力较低,推理时间短,可满足风电场功率预测实时要求。

### 3 结 论

本文面向综合能源系统的风电功率预测任务,充分利用并深入挖掘含气象等相关信息的风电功率时间序列数据,提出了多特征融合的基于双塔 Transformer 模型风电功率预测算法。通过引入时间点注意力与特征变量注意力机制,实现更加全面的特征提取及多特征融合;设计了分布特征模块、频域特征模块、气象模式解耦模块和季节趋势分解模块,强化了预测模型对于整体数据分布态势的感知能力和长期趋势的理解能力,提升了对于不同场景的适应能力与泛化能力;通过加权平均融合策略实现双塔 Transformer 模型集成,获得风电功率预测结果,取得预测精度和解释能力的显著优势。通过大量实验展现了本文所提方法在综合能源系统风电功率预测任务中的有效性、优越性及可行性,为融合多特征信息的时间序列建模提供了可供借鉴的方法及路径,丰富了综合能源系统风电功率预测的理论研究和实际应用,具有广阔的应用前景及实用价值。在今后的研究工作中,将进一步有效利用信号处理技术、气象模型搭建及统计数据分析等领域知识,深入研究高效的时间序列数据预测方法及其应用。同时,考虑系统安全应用和用户隐私保护需求,联合联邦学习技术建立更加安全可靠的综合能源系统风电功率预测模型,服务于能源行业。

### 参考文献

[1] PAN Y L, DONG F. Dynamic evolution and driving factors of new energy development: Fresh evidence from China[J]. Technological Forecasting and Social Change, 2022, 176: 121475.

[2] LIU W F, MCKIBBIN W J, MORRIS A C, et al. Global economic and environmental outcomes of the paris agreement [J]. Energy Economics, 2020, 90: 104838.

[3] ZOU C N, XIONG B, XUE H Q, et al. The role of new energy in carbon neutral [J]. Petroleum Exploration and Development, 2021, 48(2): 480-491.

[4] SADORSKY P. Wind energy for sustainable development: Driving factors and future outlook[J]. Journal of Cleaner Production, 2021, 289: 125779.

[5] GWEC. Global Wind Report 2025[R/OL]. [2025-06-11]. <https://www.gwec.net/reports/globalwindreport>.

[6] 张冬冬,单琳珂,刘天皓.人工智能技术在风力与光伏发电数据挖掘及功率预测中的应用综述[J].综合智慧能源,2025,47(3):32-46.

ZHANG D D, SHAN L K, LIU T H. Review on the application of artificial intelligence in data mining and

wind and photovoltaic power forecasting [J]. Integrated Intelligent Energy, 2025, 47(3): 32-46.

[7] 何志福,徐戈.面向风电功率预测的关键技术及应用综述[J].现代工业经济和信息化,2024,14(10): 129-130.

HE Z F, XU G. A review of key technology and applications for wind power prediction [J]. Modern Industrial Economy and Informationization, 2024, 14(10): 129-130.

[8] 黎静华,桑川川,甘一夫,等.风电功率预测技术研究综述[J].现代电力,2017,34(3):1-11.

LI J H, SANG C C, GAN Y F, et al. A review of researches on wind power forecasting technology [J]. Modern Electric Power, 2017, 34(3): 1-11.

[9] 刘谭,刘娜,刘贵平,等.深度学习方法在风电功率预测中的应用与研究方向概述[J].计算机科学与探索,2025,19(3):602-622.

LIU T, LIU N, LIU G P, et al. Overview of applications and research directions of deep learning methods for wind power prediction [J]. Journal of Frontiers of Computer Science and Technology, 2025, 19(3): 602-622.

[10] JONKERS J, AVENDANO D N, VAN W G, et al. A novel day-ahead regional and probabilistic wind power forecasting framework using deep CNNs and conformalized regression forests [J]. Applied Energy, 2024, 361: 122900.

[11] CHEN Y J, HE Y J, XIAO J W, et al. Hybrid model based on similar power extraction and improved temporal convolutional network for probabilistic wind power forecasting [J]. Energy, 2024, 304: 131966.

[12] 苏连成,朱娇娇,李英伟.基于时间卷积网络残差校正的短期风电功率预测[J].太阳能学报,2023,44(7): 427-435.

SU L C, ZHU J J, LI Y W. Short-term wind power prediction based on residual correction of temporal convolutional networks [J]. Acta Energiae Solaris Sinica, 2023, 44(7): 427-435.

[13] 包广斌,杨龙龙,范超林,等.基于 SSA 优化的 Transformer-BiGRU 短期风电功率预测[J].电子测量技术,2025,48(13):139-147.

BAO G B, YANG L L, FAN C L, et al. Short-term wind power prediction based on SSA-optimized Transformer-BiGRU [J]. Electronic Measurement Technology, 2025, 48(13): 139-147.

[14] LANG J X, PENG X S, LI W Z, et al. A novel two-stage interval prediction method based on minimal gated memory network for clustered wind power

- forecasting[J]. Wind Energy, 2021, 24(5): 450-464.
- [15] XU R, BAO G. Short-term wind power prediction based on empirical mode decomposition and deep learning[C]//The 40th Chinese Control Conference(CCC). IEEE, 2021: 8372-8379.
- [16] 贾睿, 杨国华, 郑豪丰, 等. 基于自适应权重的 CNN-LSTM&GRU 组合风电功率预测方法[J]. 中国电力, 2022, 55(5): 47-56, 110.
- JIA R, YANG G H, ZHENG H F, et al. Combined wind power prediction method based on CNN-LSTM&GRU with adaptive weights [J]. Electric Power, 2022, 55(5): 47-56, 110.
- [17] GONG M J, YAN C C, XU W, et al. Short-term wind power forecasting model based on temporal convolutional network and Informer [J]. Energy, 2023, 283: 129171.
- [18] XU Z J, ZENG A L, XU Q. FITS: Modeling time series with 10k parameters [C]//The 12th International Conference on Learning Representations (ICLR), 2024: 29143-29166.
- [19] LIN S S, LIN W W, WU W T, et al. SparseTSF: Modeling long-term time series forecasting with 1k parameters[C]//The 41st International Conference on Machine Learning(ICML), 2024.
- [20] 张登攀, 兰征, 杜怡衡. 基于 PCHIP-VMD 数据分析与 SSA-LSTM 模型的短期风电功率预测[J]. 电子测量与仪器学报, 2025, 39(5): 251-261.
- ZHANG D P, LAN Z, DU Y H. Short-term wind power forecasting based on PCHIP-VMD data analysis and SSA-LSTM model [J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(5): 251-261.
- [21] 蒲晓云, 杨靖, 杨兴, 等. 基于分解技术的 IZOA-Transformer-BiGRU 短期风电功率预测[J]. 电子测量技术, 2025, 48(2): 39-48.
- PU X Y, YANG J, YANG X, et al. IZOA-Transformer-BiGRU short-term wind power prediction based on decomposition technique [J]. Electronic Measurement Technology, 2025, 48(2): 39-48.
- [22] QIU X F, WU X J, LIN Y, et al. DUET: Dual clustering enhanced multivariate time series forecasting[C]//The 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2025: 1185-1196.
- [23] ZHANG Y H, YAN J C. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting[C]//The 11th International Conference on Learning Representations(ICLR), 2023.
- [24] 李季桐, 刘杰, 杨娜, 等. 基于层次化多尺度特征融合的金属缺陷分类模型[J]. 仪器仪表学报, 2025, 46(3): 206-218.
- LI J T, LIU J, YANG N, et al. Enhanced hierarchical multi-scale feature fusion model for metal defect classification[J]. Chinese Journal of Scientific Instrument, 2025, 46(3): 206-218.
- [25] HONG J C, YANG J S, LIANG F W, et al. An adaptive threshold strategy based on empirical distribution functions and information entropy for battery abnormal diagnosis and fault alarm [J]. Energy, 2025, 324: 135980.
- [26] LE T T H, HEO S, KIM H. Toward load identification based on the Hilbert transform and sequence to sequence long short-term memory [J]. IEEE Transactions on Smart Grid, 2021, 12(4): 3252-3264.
- [27] WU H X, XU J H, WANG J M, et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting [C]//The 35th Neural Information Processing Systems(NeurIPS), 2021, 34: 22419-22430.
- [28] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//The 31st Neural Information Processing Systems(NeurIPS), 2017, 30.
- [29] LUND B D, MA J X. A review of cluster analysis techniques and their uses in library and information science research: K-means and K-medoids clustering [J]. Performance Measurement and Metrics, 2021, 22(3): 161-173.
- [30] REN J W, ZHANG M Y, YU C J, et al. Balanced MSE for imbalanced visual regression [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 7926-7935.
- [31] HODSON T O. Root mean square error(RMSE) or mean absolute error(MAE): When to use them or not[J]. Geoscientific Model Development Discussions, 2022, 2022: 1-10.
- [32] CHICCO D, WARRENS M J, JURMAN G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation [J]. PeerJ Computer Science, 2021, 7: e623.
- [33] CHUNG Y, NEISWANGER W, CHAR I, et al. Beyond pinball loss: Quantile methods for calibrated uncertainty quantification [J]. Advances in Neural Information Processing Systems, 2021, 34: 10971-10984.
- [34] ZHOU H Y, ZHANG S H, PENG J Q, et al.

Informer: Beyond efficient transformer for long sequence time-series forecasting[C]//AAAI Conference on Artificial Intelligence,2021, 35(12): 11106-11115.

[35] ZHOU T, MA Z Q, WEN Q S, et al. FiLM: Frequency improved legendre memory model for long-term time series forecasting[C]//Advances in Neural Information Processing Systems(NeurIPS),2022, 35: 12677-12690.

[36] LIU Y, HU T G, ZHANG H R, et al. iTransformer: Inverted transformers are effective for time series forecasting [C]//The 12th International

Conference on Learning Representations(ICLR),2024: 4004-4028.

**作者简介**

唐俊萍,硕士研究生,主要研究方向为风电功率预测。  
E-mail:3168645015@qq.com

暴琳(通信作者),博士,副教授,主要研究方向为推荐算法。  
E-mail:baolin\_zj@163.com

梁勇,硕士研究生,主要研究方向为综合能源系统负荷预测。  
E-mail:1152023466@qq.com