

DOI:10.19651/j.cnki.emt.2519754

基于动态频域特征解耦的中长期电力负荷预测*

罗鹏阳 朱文忠 王 文

(四川轻化工大学计算机科学与工程学院 宜宾 644000)

摘要: 中长期电力负荷预测是保障电力系统规划与运行稳定性、经济性的核心环节。一些研究通过傅里叶变换将输入数据转换到频域,以此来获得不同的信号分量,从而减轻噪声的干扰。但现有研究往往不加区分地处理全部的频域信号,使关键频域分量和无关频域分量混合,导致模型难以完全捕捉频域信号中蕴含的特征。因此,提出了一种融合频域分析与注意力机制的多变量长时序预测模型 FTAformer。该模型集成了时域和频域信息,协同建模以提高模型对全局特征的捕捉能力。首先,利用快速傅里叶变换将输入序列转换为频域信号,采用层级滤波和隔离策略,隔离出关键频域分量并抑制噪声。接着通过多头注意力机制在时域上捕捉不同变量间的相关性,并利用层归一化和前馈网络模块建模序列的全局表示。实验结果表明,在两个公开电力负荷数据集上,该模型的预测性能显著高于其他基准模型。相较于现有最优模型 iTransformer,所提方法的均方误差和平均绝对误差在多步预测场景下分别降低 15.26% 和 8.76%,充分验证了频域分析与多头注意力机制协同建模在中长期电力负荷预测中的有效性与优越性。

关键词: 频域;快速傅里叶变换;特征融合;iTransformer 模型;负荷预测;多步预测

中图分类号: TN911;TM715 **文献标识码:** A **国家标准学科分类代码:** 510.40

Medium and long-term power load forecasting based on
dynamic frequency domain feature decoupling

Luo Pengyang Zhu Wenzhong Wang Wen

(School of Computer Science and Engineering, Sichuan University of Science and Engineering, Yibin 644000, China)

Abstract: Medium and long-term power load forecasting is a core link to ensure the stability and economy of power system planning and operation. Some studies convert the input data to the frequency domain through Fourier transform to obtain different signal components, thereby reducing the interference of noise. However, existing studies often indiscriminately handle all frequency-domain signals, causing the key frequency-domain components and irrelevant frequency-domain components to mix, which makes it difficult for the model to fully capture the features contained in the frequency-domain signals. Therefore, a multivariable long-term prediction model FTAformer that integrates frequency-domain analysis and attention mechanism is proposed. This model integrates time-domain and frequency-domain information and conducts collaborative modeling to enhance the model's ability to capture global features. Firstly, the input sequence is transformed into a frequency-domain signal by using the fast Fourier transform. A hierarchical filtering and isolation strategy is adopted to isolate the key frequency-domain components and suppress the noise. Then, the correlations among different variables are captured in the time domain through the multi-head attention mechanism, and the global representation of the sequence is modeled by using layer normalization and the feedforward network module. The experimental results show that on two public power load datasets, the predictive performance of this model is significantly higher than that of other benchmark models. Compared with the existing optimal model iTransformer, the mean square error and mean absolute error of the proposed method are reduced by 15.26% and 8.76% respectively in the multi-step prediction scenario, fully verifying the effectiveness and superiority of the collaborative modeling of frequency domain analysis and multi-head attention mechanism in medium and long-term power load forecasting.

Keywords: frequency domain; fast Fourier transform; feature fusion; iTransformer model; load forecasting; multi-step prediction

0 引 言

在现代电力系统向高弹性、低碳化转型的背景下,中

长期电力负荷预测作为电网规划与能源战略布局的关键基础技术,其重要性日益凸显^[1]。“双碳”目标驱动下,分布式光伏、电动汽车集群等非线性负荷的大规模并网,使负荷时序

收稿日期:2025-09-01

* 基金项目:四川省科技计划重点研发项目(2023YFS0371)、企业信息化与物联网测控技术四川省高校重点实验室开放基金(2024WYJ03)、四川省智慧旅游研究基地项目(ZHYJ24-01)资助

呈现出多尺度周期性嵌套与强非平稳性交织的复杂动态特征^[2]。

目前,电力负荷预测方法包括统计模型方法、机器学习方法和深度学习方法。传统的统计方法,如卡尔曼滤波^[3]和ARIMA^[4]通常假设负荷数据是平稳且线性的,但现实数据往往是复杂、异构且非理想分布的,所以此类方法难以对电力负荷数据中的非线性依赖关系和时间序列特征进行建模,导致预测精度不高且波动较大;随着人工智能的发展,支持向量机^[5]和随机森林^[6]等机器学习方法开始应用于负荷预测领域。但此类机器学习方法往往需要依赖于专家知识进行特征筛选,且面对激增的电力负荷数据量时,难以从海量数据中捕捉复杂非线性关系。

随着算力资源的突破与技术革新,深度学习方法逐渐成为研究主流。众多研究围绕卷积神经网络(convolutional neural network, CNN)^[7]、循环神经网络(recurrent neural network, RNN)^[8]及其变体长短时记忆网络(long short-term memory, LSTM)^[9]等网络进行展开。例如,文献[10]提出一种基于生成对抗网络(generative adversarial network, GAN)的短期电力负荷预测方法,通过经验模态分解(empirical mode decomposition, EMD)分解原始信号,引入改进麻雀搜寻算法(improved sparrow search algorithm, ISSA)优化LSTM超参数,并利用GAN生成新数据样本,以提升负荷预测精度和模型泛化能力。文献[11]通过引入非支配排序遗传算法对LSTM模型进行超参数优化,用以提升住宅建筑短期供暖负荷预测的精准度。以上的方法虽然都在负荷预测上取得了一定的效果,但这些方法都存在着长期依赖和梯度消失等问题^[12]。Transformer^[13]通过其独特的自注意力机制能够很好地捕获长期依赖关系,极大地提高了模型在时序预测任务上的准确率,在回归预测领域应用较为广泛。例如,文献[14]结合图神经网络(graph neural network, GNN)和引入频域指数滑动平均注意力的Transformer,构建分量感知动态图Transformer模型,用以提升短期电力负荷预测的精准度。但由于Transformer模型复杂的结构导致其面对长时序数据时计算量会随着序列长度平方级增长,使计算成本显著增高。虽然,基于Transformer模型改进的如Informer^[15]、Autoformer^[16]、FEDformer^[17]和Crossformer^[18]等模型在一定程度上提高了Transformer的性能,但这种以序列为核心将整个时间步的多个变量嵌入成一个Token的做法忽略了变量间的内在关联。因此,文献[19]提出了一种倒置的思想,将同一变量的整个序列映射成高维特征表示,使注意力机制更好的捕捉变量间的相关性。由于每个变量时间序列被单独嵌入,注意力机制能够更有效地挖掘变量间的关系,而不像传统Transformer模型将同一时间戳多变量嵌入成一个Token可能导致信息丢失。最后,通过层归一化和前馈网络模块建模序列的全局表示。

然而在实际预测场景中,负荷数据往往呈现出显著的波动性与周期性,不仅包含大量噪声,还易受天气变化、节假日等因素影响而出现突发波动。而电力负荷的周期性本质上体现为其频域特性^[20],但传统深度学习模型多在时域直接处理数据,缺乏对这些频率成分的显式分离机制——即无法将趋势、周期等不同频率特征有效分离并针对性建模,这导致模型对频域信息的利用不充分。同时,由于噪声以及无关信号的干扰导致关键频域特征的表达能力也受到限制,频域分析方法为解决这一问题提供了新路径。

近年来,频域特征解耦与动态滤波技术凭借对时序数据多尺度特性的精准捕捉能力,成为时序预测领域的研究热点,其核心思路在于通过显式分离频率成分、动态抑制噪声干扰以提升模型对关键特征的建模能力。在频域特征解耦方面,现有研究围绕“频率分离-针对性建模”展开深度探索。例如,FEDformer提出傅里叶增强块与小波增强块,通过频域变换将时序信号分解为不同频率分量,利用注意力机制对高频细节与低频趋势进行差异化建模,在电力负荷等长时序预测任务中实现精度与效率的双重提升;TFformer^[21]提出以最大重叠离散小波变换(maximal overlap discrete wavelet transform, MODWT)为核心的频域分解方式,通过MODWT将时序信号解耦为低频趋势成分与高频周期成分,避免子采样导致的信息丢失并提升粗尺度分辨率;同时设计序列级频率注意力机制强化高频成分中的关键周期信息,再通过时序提取块单独建模低频趋势特征,最终融合两类成分预测结果,既实现不同频率特征的有效解耦,又降低了时间复杂度。动态滤波技术则聚焦于自适应噪声抑制与关键特征强化,其发展呈现从静态滤波到动态调制的演进趋势。例如,FilterNet^[22]设计通用频率核与自适应滤波单元,通过输入信号动态调整滤波参数,但存在对强突发干扰响应滞后的问题;对此,AdaWaveNet^[23]提出以可学习卷积核为核心设计实现的动态滤波单元AdaWave Block,结合通道注意力聚焦关键频率,在非平稳信号处理中能更高效抑制突发干扰。频域方法的优势在于能够清晰揭示时间序列数据中存在的周期性模式,但现有的这些方法通常会不加区分地处理所有频率分量,这会导致关键信号与无关信号之间产生干扰和混合,最终会限制模型对关键周期性特征的捕捉以及抑制噪声的能力^[24]。

针对上述问题,本文提出一种将频域分析与注意力机制深度融合的负荷预测模型——FTAformer。与现有方法不同,FTAformer整合了时域和频域信息进行协同建模。首先,针对现有模型对频域信息利用不足的问题,本文设计了一种时频嵌入层,可将变量的整个序列从时域和频域两个维度映射至高维空间。该嵌入层在保留以“变量为核心”的基础上,通过整合变量的整个序列上时域的动态局部特征与频域的全局趋势,将两种维度的特征融合为统一的变量Token,为后续注意力机制有效捕捉变量间的跨维度相

关性奠定了基础。此外,为避免无关频段信号及噪声的干扰,模型采用层级化滤波与掩码策略对频率信号进行区分:先将其分层划分为多个子带,再通过掩码机制实现子带隔离,从而精准提取关键频段特征,有效规避了无关信号和噪声的影响。最后,通过注意力机制在时域上提取长期依赖特征与变量相关性特征。本研究的主要贡献为:

1)提出一种时频嵌入方法。现有嵌入层仅能将变量全序列从时域映射至高维空间,而该方法实现时域与频域双维度映射,通过融合两维度特征生成统一变量 Token,为后续注意力机制精准捕捉变量间跨维度相关性提供了有效支撑,显著提升了特征表征的完整性与关联性;

2)设计了一种可实现频域信号区分的 FreBandEncoder 编码器。该编码器采用层级化滤波与掩码策略,能够依据数据特性划分出数量可变的分层子带,每个子带对应特定的频率波段;同时为各子带构建二进制掩码矩阵,通过选择

性保留目标子带的频域信号,使模型得以聚焦于最具相关性的频域信息,在有效抑制无关信号及噪声干扰的同时,显著增强了频域特征提取的针对性与鲁棒性;

3)构建了一种动态加权平滑损失函数(dynamic weighted smooth loss, DWSLoss)。该损失函数通过动态权重调节机制,实现对负荷预测中不同误差关注度的自适应平衡:一方面采用基于误差绝对值的指数衰减权重策略,强化对细微负荷波动的拟合精度;另一方面通过分段函数设计实现误差惩罚的平滑过渡,提升模型优化过程的稳定性。

1 相关工作

iTransformer 是基于 Transformer 架构改进的神经网络,专门针对时间序列预测任务进行了优化,在负荷预测等多变量场景中能更高效地发挥作用。如图 1 所示。

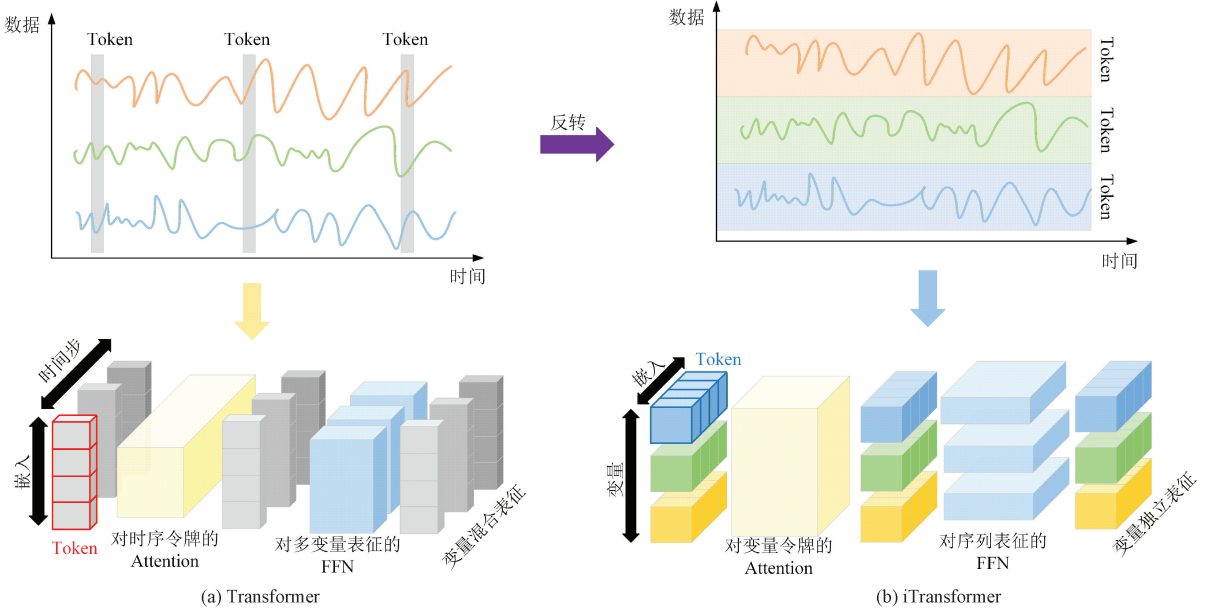


图 1 Transformer 和 iTransformer 对比图

Fig. 1 Comparison chart of Transformer and iTransformer

传统 Transformer 模型处理多变量时间序列时存在明显局限:它将同一时间步的所有变量数据聚合为一个 Token,这种方式难以充分建模变量间的内在关联。在负荷预测中,气温、节假日、历史负荷等因素均会影响用电负荷且彼此存在交互作用,但传统模型因将多变量压缩为单时间步 Token,无法区分各变量的独特属性,导致难以精准捕捉变量间的依赖关系,这直接影响了其在复杂多变量负荷预测中的表现与泛化能力。为解决这一问题,iTransformer 采用了“倒置”的核心设计思路:其框架包含嵌入层、多个可堆叠的 TrmBlock 模块和投影层,具体执行流程为:嵌入层对每个变量的完整时间序列进行单独处理——如历史负荷、气温等变量各自生成专属的高维特征序列即变量 Token;随后通过转置操作调整维

度,使模型在后续模块 TrmBlock 中聚焦于变量维度的注意力计算,即分析不同变量在同一时间步的关联及跨时间步的相关性;经层归一化和前馈网络处理后,最终由投影层映射出对应的预测结果。整个过程如式(1)~(3)所示。

$$\mathbf{H}^0 = \text{Embedding}(\mathbf{X}^T) \tag{1}$$

$$\mathbf{H}^l = \text{TrmBlock}(\mathbf{H}^{l-1}), l = 1, \dots, L \tag{2}$$

$$\mathbf{Y} = \text{Projection}(\mathbf{H}^L) \tag{3}$$

式中: \mathbf{X} 表示长度为 T 和 N 个变量的多维时序序列; $\mathbf{H} \in \mathbf{R}^{N \times D}$ 表示一个包含 N 个变量 Token 的矩阵,每个变量 Token 都是 D 维的, \mathbf{H} 的上标表示层索引; L 为 TrmBlock 的个数; $\mathbf{Y} \in \mathbf{R}^{N \times S}$ 表示时间长度为 S ,变量个数为 N 的最终预测结果。

2 模型的建立与分析

本文提出了一种基于频域分析与注意力机制的中长期电力负荷预测模型 FTAformer,图 2 展示了 FTAformer 的整体体系结构,主要由时频嵌入层、FreBandEncoder 及投影层构成。时频嵌入层从时域与频域双维度将多变量的完整序列映射至高维空间,通过协同建模时域与频域信息,提升特征表征的完整性与关联性;FreBandEncoder 编码器采用层级化滤波与掩码策略:首先通过快速傅里叶变换(fast Fourier transform,FFT)将输入数据转换至频域,随后按频率信号强度由高到低排序并进行等宽分区,每个

分区对应一个子带;接着为各子带生成二进制掩码矩阵,通过选择性保留目标子带的频谱成分,使模型能聚焦于关键频率成分,同时抑制无关频率成分的干扰;之后将各子带经逆快速傅里叶变换(inverse fast Fourier transform,IFFT)转回时域,再通过与子带数量匹配的独立线性层,完成不同子带特征的融合;最后通过注意力机制在时域中提取长期依赖特征与变量间相关性特征。该编码器通过这种层级递进方式,实现对负荷信号的滤波处理,逐步筛选并强化有效频率成分,抑制噪声及无关信号。最终,经投影层映射得到相应的预测结果。模型算法描述如算法 1 所示。

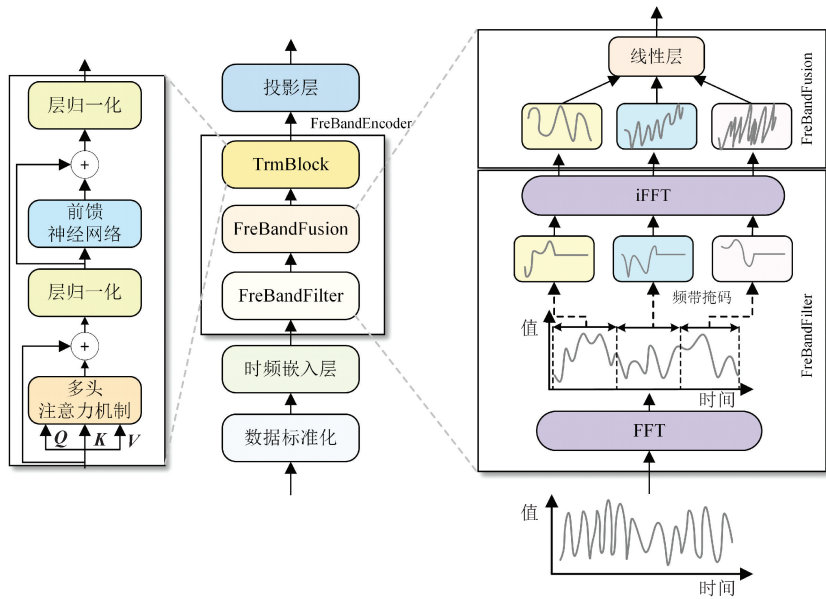


图 2 模型整体结构图

Fig. 2 Overall structure diagram of the model

2.1 时频嵌入层

时频嵌入层通过并行双分支架构实现数据多域特征的深度嵌入,核心设计思路是深度挖掘频域全局规律与时域局部动态的互补特性,以构建更完备的特征表征空间,具体结构如图 3 所示,通过双分支的功能分化设计,可针对性强化频域全局规律与时域局部动态的表征效能,为后续跨维度特征关联挖掘奠定基础。

频域处理分支,自适应傅里叶变换模块(adaptive transform Fourier block, ATFBlock)以傅里叶变换为核心,通过自适应频谱调制机制显式增强全局趋势特征的特征能力,具体是先通过 FFT 将输入信号映射至频域空间,再引入可学习复数权重矩阵对频域成分的相位特性与幅值强度进行自适应调制,该权重矩阵会为各样本频谱成分构建双通道调制机制。相位调制用于优化不同频率成分间的时序对齐精度,幅值调制则聚焦增强与负荷周期特性强相关的关键频率分量,调制后的频谱经 IFFT 转回时域以显式强化数据在频域层面的全局规律表征能力,最终通

过取实部操作提取频域增强特征,有效显化数据中隐含的全局周期性模式。

时域处理分支,稀疏加权偏移模块(sparse weighted shift block, SWSBlock)聚焦精准捕捉局部时序依赖关系,核心是结合结构化时移操作与稀疏特征映射,首先通过时移变换生成输入序列的 k 阶延迟版本,即对原始序列执行 k 次时间偏移,让每个时间步能同时感知当前及前 $k-1$ 时刻的局部信息,为短期动态特征提取提供结构化输入,同时为降低高维扩展带来的特征冗余,会引入可学习稀疏变换矩阵进行降维映射,该过程通过权重矩阵的自适应学习实现隐式稀疏约束,使模型自动弱化无关时移特征的权重占比,重点保留对负荷波动敏感的局部依赖信息。

特征融合阶段则通过残差连接机制整合双分支输出,得到蕴含双域信息的统一变量 Token,这种设计既保留了双域特征的互补性,又通过残差连接缓解了深层网络的梯度衰减问题,最终输出的变量 Token 同时涵盖经频谱调制的全局趋势信息与经稀疏约束的局部时序依赖,可让后续注意力机制高

算法 1

```

输入数据:  $X_t \in \mathbf{R}^{L \times N}, X_s \in \mathbf{R}^{L \times S}$ 
输出数据:  $Y \in \mathbf{R}^{L \times 1}$ 
# 应用时频嵌入层
enc_out = TF_embedding( $X_t, X_s$ )
for e in e_layers do
    # 应用频带滤波器
     $X_{fused} = FreBandFilter(enc\_out)$ 
    # 应用频带特征融合
     $X_{features} = FreBandFusion(X_{fused})$ 
    # 计算查询 Q、键 K、值 V 的相似度矩阵 S
     $S = Scale * (Q * K^T)$     $Scale = \frac{1}{\sqrt{D_h}}$ 
    # 应用因果掩码计算注意力权重 A
     $A = Softmax(S * Mask)$ 
    # 加权聚合值向量得到 attn_out
    attn_out = A * V
    # 前馈神经网络
    fnn_out = Conv1D(ReLU(Conv1D( $X_{features}$ )))
    # 残差与归一化
    enc_out += Dropout(attn_out + fnn_out)
    enc_out = LayerNorm(enc_out)
endfor
# 投影层
Y = projection(enc_out)

```

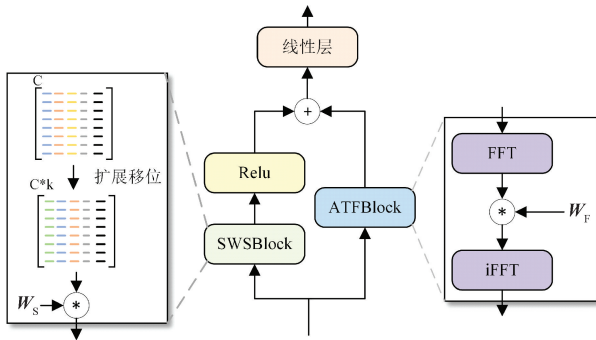


图 3 时频嵌入层

Fig. 3 Time-Frequency embedding layer

效挖掘变量间跨维度的相关性,提升模型对复杂负荷数据的特征表征能力,其具体计算公式如式(4)~(8)所示。

$$X_F = W_F * FFT(X) \quad (4)$$

$$X_F = iFFT(X_F), real \quad (5)$$

$$X_S = \Theta_K(X) \quad (6)$$

$$X_S = W_S * \Phi(X) \quad (7)$$

$$H_n = Linear(X_F + Relu(X_S)) \quad (8)$$

式中: W_F 和 W_S 代表两个可学习的权重矩阵, Θ 代表扩展矩阵操作, Φ 代表矩阵移位操作, $*$ 代表矩阵乘法。

2.2 FreBandEncoder

在时间序列分析中,全局特征信息反映了序列的整体趋势,为把握数据的总体走向提供了重要依据。频域分析是捕捉这类全局特征的有效手段,但传统方法通常对整个频域进行统一的特征提取,容易导致不同频率成分间的相互干扰,进而影响全局特征提取的精准性。

为解决这一问题, FreBandEncoder 编码器通过层级化滤波和掩码策略将频域数据划分为层级化子频带,并对各子带实施选择性特征提取,借助这种层级化处理减少频段间的干扰,使模型能够更聚焦于最相关的频率信号,从而精准捕捉各频率成分的细粒度特征。该编码器由 FreBandFilter、FreBandFusion 和 TrmBlock 3 个组件构成。其中,通过频带滤波器 FreBandFilter 进行频带划分,并通过二进制掩码选择性地提取目标频带的特征。然后,通过频带特征融合模块 FreBandFusion 整合多个变量每个独立分支的特征,获取更精确的全局特征。最终,通过 TrmBlock 在时域提取长期依赖特征与变量相关性特征。

1) FreBandFilter

为捕捉不同层级全局特征,与传统的按频率大小进行划分方法不同,该模块可根据频率信号幅度由高到低排序后进行等宽分区,分区的数量可根据数据特性自主设定,每个分区对应一个特定子带。例如,当分区数量为 3 时,幅度排前 1/3 的分配给高频波段,中间 1/3 分配给中频波段,剩下 1/3 分配给低频波段。随后,为每个子带构建一个二进制掩码矩阵,其每个元素值如式(9)所示。

$$Mask_i(b, c, f) = \begin{cases} 1, & f \in B_i \\ 0, & \text{其他} \end{cases} \quad (9)$$

式中: b 为样本索引; c 为特征索引; f 为频率索引; B_i 是按幅度排序后第 i 频段内的频率。

每个掩码矩阵只标记按幅度排序后,属于第 i 个频段范围的原始频率索引,1 表示该频率属于当前频段,0 表示不属于。通过掩码矩阵与目标频带逐元素相乘的方式,得到被掩码的频域数据。然后,对每个被屏蔽的频域数据进行快速傅里叶逆变换,以获得相应的时域分支。

根据幅度排序动态划分的子频带,通过上述描述的掩码机制,使用掩码选择性地保留目标频带的频谱成分。这种方法使后续的注意力机制模块能更专注于最相关的频率成分,同时抑制不相关的频率成分,从而增强模型对频域特征的提取能力。

为直观揭示该模块与传统划分方法的本质差异,本研究从数据集 1 中选取前 5 000 条样本数据开展对比实验,具体划分结果如图 4、5 所示。由图 4 可知,传统划分方法以频率数值为唯一依据,采用固定阈值将频谱机械划分为低频段、中频段与高频段,该方式未考虑信号幅度这一关键数据特征,存在划分逻辑与实际数据特性脱节的局限。而本研究提出的 FreBandFilter 方法,首先对全频段信号的幅度值进行排序,再以幅度分布特征为核心依据完成频带

划分,实现了划分逻辑与数据内在属性的深度耦合。

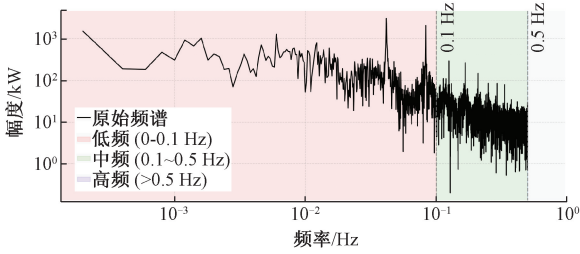


图4 频率划分方法

Fig. 4 Frequency division method

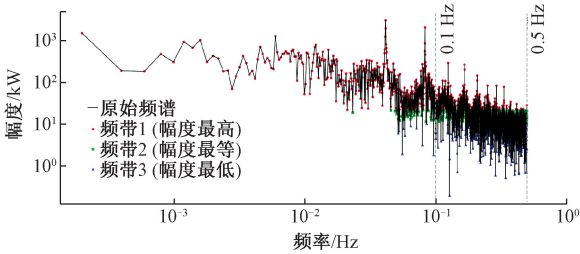


图5 幅度划分方法

Fig. 5 Amplitude division method

进一步结合图5分析可知, *FreBandFilter* 方法突破了传统固定频率区间的约束:无论信号原始频率处于低频、中频还是高频范围,只要其幅度值位列全部信号幅度的前1/3,均会被归类至频带1。这种以数据固有特征,即以幅度分布为导向的划分策略,有效解决了传统方法因频率区间固定而无法适配复杂数据模式的问题。针对电力负荷数据的典型特征——具有显著的多尺度时间模式,既包含居用电呈现的日、周等周期性波动,又涵盖工业生产过程中随机产生的高频冲击信号, *FreBandFilter* 方法能够通过动态适配数据幅度特征,灵活捕捉电力负荷在不同时间尺度下的全局特征,为后续负荷分析与预测提供更精准的数据支撑。

2) *FreBandFusion*

对于每个变量的分支,采用独立线性层提取特征。随后,通过多频融合机制,有效整合各频段的全局信息,形成了更全面、更精确的全局特征表示。具体来说,每个分支的全局特征计算公式为:

$$GlobalBranch_i = Linear(Branch_i) \quad (10)$$

$$GlobalFeature = \sum_{i=1}^m GlobalBranch_i \quad (11)$$

3) *TrmBlock*

TrmBlock 模块主要作用是捕捉多变量序列数据间的相关性与长期依赖特性。它通过多头自注意力机制让不同变量标记间进行信息交互,获取变量间相关性;利用层归一化统一不同变量的测量单位与特征分布,减少变量取值差异影响;借助共享的前馈网络对序列进行独立处理,学习序列的时间特性,从而更好地提取序列的全局特征。

其涉及到的计算公式如式(12)~(13)所示。

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (12)$$

$$LayerNorm(H) = \left\{ \frac{h_n - Mean(h_n)}{\sqrt{Var(h_n)}} \mid n = 1, \dots, N \right\} \quad (13)$$

2.3 DWSLoss

DWSLoss 在 *L1loss* 的基础上引入了自适应加权机制,以增强对不同程度误差的灵活处理能力。该损失函数通过两个关键参数进行调控: β 设定误差从平方损失过渡到线性损失的阈值,确保对小误差平滑优化而对大误差保持鲁棒性; α 则控制一个基于误差绝对值的动态权重,使得较小的误差获得更高的相对权重,而较大的误差权重逐渐衰减,从而让模型更关注难以拟合的样本。其具体计算公式如式(14)所示。

$$L = \frac{1}{N} \sum_{i=1}^N (1 + \alpha \times e^{-|y_i - \hat{y}_i|}) \begin{cases} \frac{1}{2} \frac{(y_i - \hat{y}_i)^2}{b}, & |y_i - \hat{y}_i| < \beta \\ |y_i - \hat{y}_i| - \frac{1}{2}b, & |y_i - \hat{y}_i| \geq \beta \end{cases} \quad (14)$$

式中: N 代表样本总数, y_i 和 \hat{y}_i 分别代表第 i 个样本的真实值和预测值, α 和 β 是两个超参数。

3 实验及结果分析

3.1 数据来源与预处理

本文选用两个公开的电力负荷数据集来验证模型的有效性。数据集1为Kaggle平台上公开的西班牙地区2015年1月1日~2018年12月31日的电力数据集,采样频率为1h,共包括35065条数据。数据集2的数据集为澳大利亚地区2006年~2010年间的负荷数据,采样频率为1h,共43826条数据。上述两个数据集不仅包含了该地区的负荷数据,同时还包含温度、湿度等气象数据和电价数据。气象因素对电力负荷的变化有直接影响。例如,当气温上升时,居民可能会增加风扇、空调等大功率电器的使用,进而导致电力负荷的增加。此外,电价的调整同样也对电力负荷需求产生影响,因为电价的变动可能会改变消费者的用电习惯。

本文对上述两个数据集进行数据预处理,首先通过数据可视化的方式检查数据集中是否存在异常值和缺失值,通过线性插补法对这些异常数据进行处理。同时,为了消除特征尺度差异对模型训练的干扰,本文对数据进行标准化的处理。通过统一不同样本及特征的尺度,有效均衡了各特征在模型训练中的贡献权重,减少了因特征量级差异导致的模型对高尺度特征的过度依赖。最后,将两个数据集进行划分,将数据集的70%用于训练集,10%用于验证集,20%用于测试集。

3.2 实验环境及参数设置

本文实验环境配置为 AMD Ryzen7 4800H 处理器与 NVIDIA GeForce GTX 1650 显卡;模型开发基于 Python 编程语言与 PyTorch 2.0.0 深度学习框架。

深度学习模型的参数选择与调优对模型最终精度及资源消耗具有显著影响。本研究中,所提算法的核心参数集中于特征提取模块,具体包括模型维度 d_{model} 与划分频带数 n_{bands} 。为探究模型对特征提取参数的敏感性,本研究在数据集 1 上结合变量控制法与网格搜索法进行实验验证。

图 6 展示了不同 d_{model} 与 n_{bands} 组合下验证集的 MSE 表现。实验结果表明,模型性能对 d_{model} 与 n_{bands} 的选择具有显著敏感性,且并非维度越高、频带数越多性能越优。最终确定 $d_{\text{model}}=512$ 、 $n_{\text{bands}}=2$ 为最优参数配置,该配置在预测精度与计算成本间实现了最优平衡。

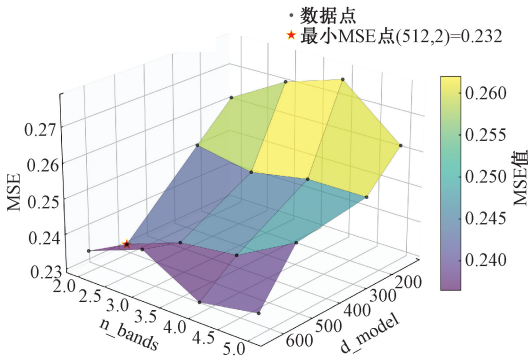


图 6 参数敏感性分析

Fig. 6 Parameter sensitivity analysis

模型超参数中,学习率采用动态调整策略,每训练一轮衰减 50% 的阶梯式指数衰减;其余超参数均采用时序预测模型领域的常用默认值。具体超参数设置如表 1 所示。

表 1 参数设置

Table 1 Parameter settings

参数名称	参数设置
输入序列长度	96
注意力头数	8
激活函数	gelu
隐藏层维度	512
编码器层数	2
丢失率	0.1
学习率	0.000 1
批次大小	32
训练轮数	10

3.3 模型评价指标

本文使用了均方误差 (mean squared error, MSE) 和

平均绝对误差 (mean absolute error, MAE) 两种误差评价指标衡量预测模型的准确性。平均绝对误差用于衡量预测值与真实值之间的平均绝对差异,表示预测值与真实值之间的平均偏差程度;均方误差用于衡量预测值与真实值之间差值平方的平均值,表示预测值与真实值之间的平均平方偏差程度,通过对误差进行平方运算,会放大较大误差的影响,从而更敏感地反映预测值与真实值的偏离情况。其计算公式为:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (15)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (16)$$

其中, n 表示测试集的样本总数, y_i 表示真实值, \hat{y}_i 表示预测值。

3.4 预测结果对比分析

为全面评估所提模型在中长期预测任务中的性能表现,本研究在统一的实验环境和相同的超参数设置下与 iTransformer、WDformer^[25]、FilterTS^[26]、FilterNet 和 FreTS^[27] 5 个代表性模型进行了系统对比。其中, iTransformer 作为基于标准注意力架构的当前性能标杆,其对比结果可直观体现本文方法融合频域学习与多头注意力机制的优势;WDformer、FilterTS、FilterNet 和 FreTS 作为频域学习模型,则与本文频域处理模块构成方法学层面的对比分析。

实验数据如表 2 所示,在预测步长为 96、192、336、720 的全场景下,本文模型的 MSE、MAE 指标均显著低于对比模型。具体而言,相较于次优模型 iTransformer,在数据集 1 中,本文模型不同预测步长下的 MSE 平均降低 17.08%、MAE 平均降低 10.76%;数据集 2 中 MSE 平均降低 13.07%、MAE 平均降低 6.47%。表中加粗的数据标示了对比模型中表现最出色的预测结果,实验结果充分验证了本文模型在多步长预测任务中的准确性与稳定性优势。

3.5 可视化分析

为更直观地验证 FTAformer 模型的预测性能,本文对部分预测结果进行可视化处理,具体结果如图 7 所示。图 7 展示了各对比模型在数据集 1、预测步长为 96 步场景下的预测结果;横轴表示时间步长,纵轴为标准化后的负荷值,实线代表预测值,虚线代表真实值。由图中曲线对比可知,FTAformer 模型的预测曲线与真实值曲线的拟合程度显著优于其他模型,表明该模型可提供更精准的预测结果。

3.6 不同回溯步长下模型稳定性分析

为验证模型在不同回溯步长下的稳定性与适应性,本研究将本文模型与 iTransformer、WDformer、FilterTS、FilterNet、FreTS 等对比模型置于相同实验环境以及超参数下,在数据集 1 上选取回溯窗口长度为 48、96、192、336 开展对比实验。图 8 结果表明,本文模型在全回溯窗口范

表2 各模型预测结果对比

Table 2 Comparison of prediction results of each model

数据集	预测步长	本文		iTransformer		WDformer		FilterTS		FilterNet		FreTs	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
数据集 1	96	0.232	0.323	0.314	0.389	0.305	0.384	0.389	0.469	0.378	0.466	0.322	0.421
	192	0.283	0.371	0.330	0.407	0.350	0.422	0.438	0.502	0.434	0.503	0.377	0.464
	336	0.314	0.400	0.371	0.444	0.385	0.454	0.476	0.531	0.476	0.536	0.404	0.490
	720	0.331	0.423	0.384	0.460	0.385	0.464	0.478	0.53	0.482	0.542	0.401	0.491
	Avg	0.290	0.379	0.349	0.425	0.356	0.431	0.445	0.508	0.442	0.511	0.376	0.466
数据集 2	96	0.104	0.216	0.123	0.237	0.126	0.241	0.189	0.310	0.180	0.304	0.149	0.268
	192	0.123	0.240	0.138	0.256	0.151	0.270	0.202	0.318	0.198	0.318	0.160	0.283
	336	0.121	0.244	0.144	0.263	0.146	0.268	0.214	0.331	0.210	0.333	0.172	0.300
	720	0.122	0.251	0.138	0.264	0.144	0.270	0.245	0.365	0.243	0.369	0.202	0.337
	Avg	0.118	0.238	0.135	0.255	0.141	0.262	0.212	0.331	0.207	0.331	0.170	0.297

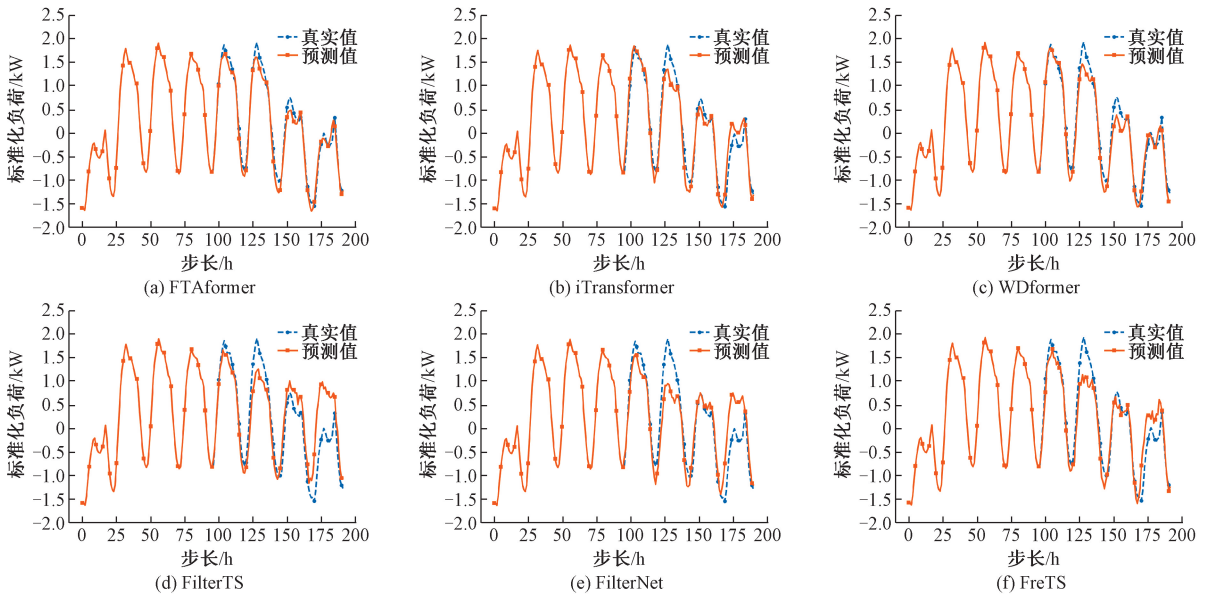


图7 预测结果可视化对比图

Fig. 7 Visual comparison chart of the prediction results

围内的 MSE 指标均显著低于其他模型,体现出对不同历史信息长度更强的适应性与稳定性。进一步分析发现,各模型在回溯步长为 96 时普遍趋近最优精度,基于此,本文实验均采用回溯步长 96 的设置以平衡计算效率与预测性能。

3.7 不同损失函数下对比分析

为验证本文所提 DWSLoss 的有效性,在相同实验环境以及超参数下,将其与 MSELoss、L1Loss 展开对比分析。由图 9 可知,DWSLoss 在各对比模型上均有效提升了预测精度;在不同预测步长下,相较于 MSELoss 与 L1Loss,其对应的 MSE 指标分别平均降低 1.23% 和 1.93%。

3.8 模型效率分析

为深入剖析各模型在效率与精度间的平衡表现,本研

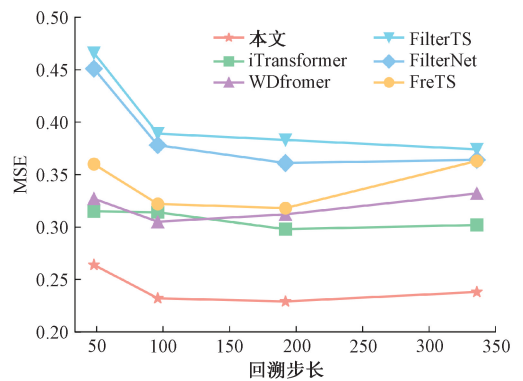


图8 不同回溯步长模型效果对比图

Fig. 8 Comparison chart of the effects of models with different backtracking step sizes

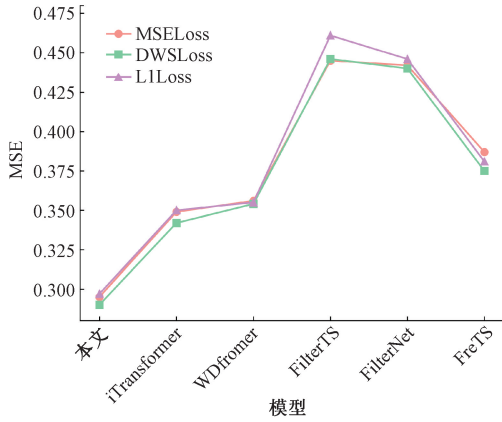


图 9 损失函数对比图

Fig. 9 Comparison chart of loss functions

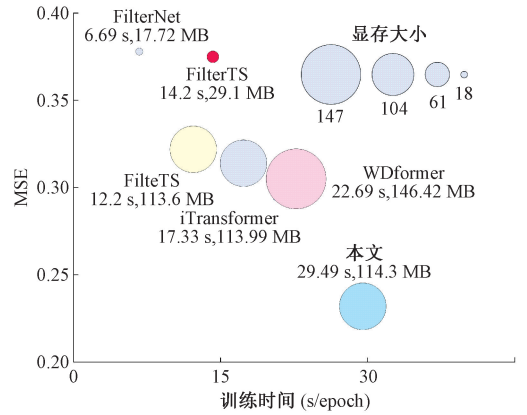


图 10 模型效率分析图

Fig. 10 Model efficiency analysis diagram

究在相同实验环境以及超参数下,在数据集 1 上开展效率分析实验,通过量化训练时间、显存占用等计算资源指标并结合预测误差进行多维度对比。实验结果如图 10 所示,采用多头注意力机制的模型如:本文模型、iTransformer 和 WDformer 模型显存占用显著高于线性模型 FilterNet 和 FilterTS,其中本文模型显存占用为 114.3 MB,与 iTransformer 和 FreTS 接近,明显低于 WDformer。尽管本文模型训练时间略长,但 MSE 指标显著优于所有对比模型,预测值与真实值偏差最小,适用于高精度需求场景。相比之下,FilterNet 和 FilterTS 虽显存占用低但预测误差偏高,iTransformer 虽显存消耗接近本文模型但精度有明显下降。综合来看,本文模型在保证预测精度的前提下,将显存占用与训练时间控制在与主流模型相当的水平,实现了效率与性能的更优均衡。

3.9 消融实验

为验证模型各模块的有效性,本研究移除关键模块分别在两个数据集上进行消融实验。具体来说,重点研究了两个核心模块:时频嵌入层和 FreBandEncoder。考虑了 4 种实验设置,具体实验设置如下: M1 为整体模型 FTAformer; M2 在移除了时频嵌入层; M3 移除了频域处理部分只采用 iTransformer 原本的 TrmBlock 结构; M4 是不含时频嵌入层和不含频域处理部分。实验结果如表 3 所示, M4 的 MSE 和 MAE 指标相较于 M1、M2、M3 均显著升高:在两个数据集中, M4 较 M1、M2、M3 的 MSE 分别平均升高了 14.10%、8.57%、5.47%。结果表明,无论是单一模块的引入及组合使用均能显著提升预测精度,充分验证了各模块在模型架构中的必要性与有效性。其中当去除 FreBandEncoder 模块是预测精度下降最大,这表明该模块对模型的性能贡献最大。

表 3 消融实验结果

Table 3 Results of ablation experiment

数据集	预测步长	M1		M2		M3		M4	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
数据集 1	96	0.232	0.323	0.237	0.332	0.264	0.345	0.287	0.363
	192	0.283	0.371	0.293	0.382	0.304	0.382	0.333	0.402
	336	0.314	0.400	0.337	0.417	0.345	0.420	0.367	0.436
	720	0.331	0.423	0.383	0.457	0.375	0.443	0.384	0.453
	Avg	0.290	0.379	0.312	0.397	0.322	0.397	0.342	0.417
数据集 2	96	0.106	0.219	0.106	0.218	0.109	0.222	0.119	0.232
	192	0.123	0.240	0.131	0.246	0.130	0.246	0.134	0.251
	336	0.121	0.244	0.131	0.252	0.137	0.257	0.143	0.260
	720	0.122	0.251	0.119	0.245	0.132	0.258	0.133	0.256
	Avg	0.118	0.238	0.121	0.240	0.127	0.245	0.132	0.249

4 结 论

由于中长期电力负荷预测面临的多尺度周期性、强非平稳性及频域信息利用不充分等核心问题,针对传统模型难以兼顾时域-频域协同建模、关键频域特征易受噪声干扰,提出融合频域分析与多头注意力机制的 FTAformer 模型,主要研究结论包括:

通过时频嵌入层实现变量序列的时频域双维度高维映射,生成的统一变量 Token 为注意力机制捕捉跨维度相关性提供支撑,弥补了传统嵌入层仅关注时域的局限;

采用层级化滤波与掩码策略,可动态划分频域子带并隔离关键分量,显著抑制无关信号与噪声干扰,解决了现有频域方法无差别处理信号导致的特征混淆问题。

未来的工作将探索该方法在其他场景负荷预测中应用,如在极端天气、突发事件等异常工况下的预测表现与适应性,并进一步优化模型以适应更广泛的实际需求。

参考文献

- [1] 马恒瑞,袁傲添,王波,等. 基于深度学习的负荷预测研究综述与展望[J]. 高电压技术,2025,51(3):1233-1250.
MA H R, YUAN A O T, WANG B, et al. Review and prospect of load forecasting based on deep learning[J]. High Voltage Technology, 2025, 51(3): 1233-1250.
- [2] 韩富佳,王晓辉,乔骥,等. 基于人工智能技术的新型电力系统负荷预测研究综述[J]. 中国电机工程学报,2023,43(22):8569-8592.
HAN F J, WANG X H, QIAO J, et al. Review on artificial intelligence based load forecasting research for the new-type power system[J]. Proceedings of the CSEE, 2023, 43(22):8569-8592.
- [3] 刘鑫蕊,常鹏,孙秋野. 基于 XGBoost 和无迹卡尔曼滤波自适应混合预测的电网虚假数据注入攻击检测[J]. 中国电机工程学报,2021,41(16):5462-5476.
LIU X R, CHANG P, SUN Q Y. Grid false data injection attacks detection based on XGBoost and unscented kalman filter adaptive hybrid prediction[J]. Proceedings of the CSEE, 2021, 41(16): 5462-5476.
- [4] 刘士进,孙立华,郭鹏. 基于 EEDM-Arima 算法的某地区电力系统负荷预测及算例分析[J]. 电子测量技术,2020,43(7):185-188.
LIU S H J, SUN L H, GUO P. Load forecasting and analysis of power system based on EEDM-Arima algorithm[J]. Electronic Measurement Technology, 2020, 43(7): 185-188.
- [5] 杨邓,杨俊杰,胡晨阳,等. 基于改进 LSSVM 的短期电力负荷预测[J]. 电子测量技术,2021,44(18):47-53.
YANG D, YANG J J, HU C H Y, et al. Short-term power load forecasting based on improved LSSVM[J].

Electronic Measurement Technology, 2021, 44(18): 47-53.

- [6] 张羽姗,周亚同. 基于 VMD-SSA-RF 算法的短期电力负荷预测模型优化[J]. 电工电能新技术,2024,43(12):30-39.
ZHANG Y S H, ZHOU Y T. Optimization of short-term power load prediction model based on VMD-SSA-RF algorithm[J]. Advanced Technology of Electrical Engineering and Energy, 2024, 43(12): 30-39.
- [7] 惠永永,孙凯文,脱奔奔,等. 基于 CNN-STA-DLSTM 模型的间歇过程质量预测[J]. 电子测量与仪器学报,2024,38(11):168-181.
HUI Y Y, SUN K W, TUO B B, et al. Batch process quality prediction based on CNN-STA-DLSTM model[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(11):168-181.
- [8] 程换新,黄震. 基于改进 PSO 优化 RNN 的短期电力负荷预测模型[J]. 电子测量技术,2019,42(20):94-98.
CHENG H X, HUANG Z H. Short-term electric load forecasting model based on improved PSO optimized RNN[J]. Electronic Measurement Technology, 2019, 42(20): 94-98.
- [9] 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 H, 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.
- [10] 曾进辉,苏旨音,肖锋,等. 基于生成对抗网络和 EMD-ISSA-LSTM 的短期电力负荷预测[J]. 电子测量技术,2024,47(20):92-100.
ZENG J H, SU Z H Y, XIAO F, et al. Short-term power load forecasting based on generative adversarial networks and EMD-ISSA-LSTM [J]. Electronic Measurement Technology, 2024, 47(20): 92-100.
- [11] DEKA P, CORDEIRO-COSTAS M, PÉREZ-OROZCO R, et al. Novel NSGA-II optimized LSTM and GRU models for short-term forecasting of residential heating load[J]. Energy and Buildings, 2025: 115999.
- [12] 范士雄,李东琦,郭剑波,等. 基于时变滤波经验模态分解-重构和独立自注意力机制的 iTransformer 超短期负荷预测方法[J]. 电网技术,2025,49(6):2436-2445.
FAN S H X, LI D Q, GUO J B, et al. Ultra-short-term load forecasting method based on time-varying filter empirical mode decomposition-reconstruction and

- iTransformer with stand-alone self-attention mechanism [J]. Power System Technology, 2020, 49(6): 2436-2445.
- [13] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. ArXiv preprint arXiv: 1706.03762, 2017.
- [14] 朱莉, 高靖凯, 朱春强, 等. 基于分量感知动态图 Transformer 的短期电力负荷预测 [J]. 计算机应用研究, 2025, 42(2): 381-390.
ZHU L, GAO J K, ZHU CH Q, et al. Short-term power load forecasting based on component-aware dynamic graph Transformer [J]. Application Research of Computer, 2020, 42(2): 381-390.
- [15] ZHOU H Y, ZHANG SH 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.
- [16] WU H X, XU J H, WANG J M, et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting [J]. Advances in Neural Information Processing Systems, 2021, 34: 22419-22430.
- [17] 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. PMLR, 2022: 27268-27286.
- [18] ZHANG Y H, YAN J CH. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting [C]. The Eleventh International Conference on Learning Representations, 2023.
- [19] LIU Y, HU T G, ZHANG H R, et al. ITransformer: Inverted transformers are effective for time series forecasting [J]. ArXiv preprint arXiv: 2310.06625, 2023.
- [20] ZHANG Q, MA Y, LI G L, et al. Short-Term load forecasting based on frequency domain decomposition and deep learning [J]. Mathematical Problems in Engineering, 2020, 2020(1): 7240320.
- [21] ZHAO T L, FANG L X, MA X, et al. TFformer: A time-frequency domain bidirectional sequence-level attention based transformer for interpretable long-term sequence forecasting [J]. Pattern Recognition, 2025, 158: 110994.
- [22] YI K, FEI J R, ZHANG Q, et al. FiltNet: Harnessing frequency filters for time series forecasting [J]. Advances in Neural Information Processing Systems, 2024, 37: 55115-55140.
- [23] YU H, GUO P K, SANO A. AdaWaveNet: Adaptive wavelet network for time series analysis [J]. ArXiv preprint arXiv:2405.11124, 2024.
- [24] LIU H H, DIAO Y N, SUN K, et al. AMSFormer: A transformer with adaptive multi-scale partitioning and multi-level spectral filtering for time-series forecasting [J]. Neurocomputing, 2025, 637: 130067.
- [25] WANG X J, ZHANG CH L, ZHENG ZH L, et al. WDformer: A Wavelet-based differential transformer model for time series forecasting [J]. ArXiv preprint arXiv:2509.25231, 2025.
- [26] WANG Y L, LIU Y SH, DUAN X Y, et al. Filterts: Comprehensive frequency filtering for multivariate time series forecasting [C]. AAAI Conference on Artificial Intelligence, 2025, 39(20): 21375-21383.
- [27] YI K, ZHANG Q, FAN W, et al. Frequency-domain MLPs are more effective learners in time series forecasting [J]. Advances in Neural Information Processing Systems, 2023, 36: 76656-76679.

作者简介

罗鹏阳, 硕士, 主要研究方向为电力负荷预测。

E-mail: 2714250951@qq.com

朱文忠(通信作者), 硕士, 教授, 主要研究方向为深度学习、智能信息处理。

E-mail: zwz@suse.edu.cn

王文, 硕士, 主要研究方向为新能源出力预测。

E-mail: 2510083129@qq.com