

融合多尺度特征的短期电力负荷预测模型^①



陈万志¹, 詹美玉¹, 林春洋²

¹(辽宁工程技术大学 软件学院, 葫芦岛 125105)

²(国网辽宁省电力有限公司 营口供电公司, 营口 115002)

通信作者: 陈万志, E-mail: chenwanzhi@lntu.edu.cn

摘 要: 为提高短期电力负荷预测精度, 提出一种融合多尺度特征的短期电力负荷预测模型. 首先, 采用皮尔逊相关系数和最大信息系数对电力负荷数据进行特征优选后, 借助动态频谱滤波模块去除数据高噪成分. 其次, 将数据输入多尺度时间卷积模块进行多尺度的时序提取. 再次, 通过变量维度的多头注意力机制和前馈神经网络, 挖掘数据间的相关性并进行特征编码. 最后, 构建线性层级融合模块, 与投影层分量同时并行处理已编码的特征, 二者加权组合实现数据特征深度融合的预测结果. 在数据集 1 上的实验结果表明, 本文模型在对比实验和消融实验均取得较好效果, MAE 和 RMSE 较次优模型提升了 5.2% 和 6.8%. 数据集 2 中对比实验结果显示, 本文模型各项指标均优于其他模型, MAPE 和 R^2 值达到 0.467、0.949. 所提模型有助于提高短期电力负荷预测精度, 并具有良好的泛化能力.

关键词: 短期电力负荷预测; iTransformer 模型; 多头注意力机制; 多尺度时间卷积; 多尺度特征融合

引用格式: 陈万志, 詹美玉, 林春洋. 融合多尺度特征的短期电力负荷预测模型. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/10247.html>

Short-term Power Load Forecasting Model Using Multi-scale Feature

CHEN Wan-Zhi¹, ZHAN Mei-Yu¹, LIN Chun-Yang²

¹(Software College, Liaoning Technical University, Huludao 125105, China)

²(Yingkou Power Supply Company, State Grid Liaoning Electric Power Supply Co. Ltd., Yingkou 115002, China)

Abstract: To improve the accuracy of short-term power load forecasting, this study proposes a model for short-term power load forecasting using multi-scale features. First, feature selection is performed on the power load data using the Pearson correlation coefficient and the maximal information coefficient, followed by dynamic spectral filtering to remove high-noise components. Second, the data are fed into a multi-scale temporal convolution module for multi-scale temporal feature extraction. Third, correlations in the data are mined, and features are encoded using a variable-dimensional multi-head attention mechanism combined with feed-forward neural networks. Finally, a linear hierarchical fusion module is developed to process the encoded features in parallel with a projection layer. The weighted combination of these two components yields prediction results through the deep fusion of data features. Experimental results on Dataset 1 show that the proposed model achieves good results in both comparative and ablation experiments, outperforming the next-best model by 5.2% and 6.8% in MAE and RMSE, respectively. On Dataset 2, the proposed model surpasses baseline models across all evaluation metrics, achieving an MAPE of 0.467 and an R^2 of 0.949. The proposed model enhances short-term power load forecasting accuracy and exhibits strong generalization capabilities.

Key words: short-term power load forecasting; iTransformer model; multi-head attention mechanism; multi-scale temporal convolution; multi-scale feature fusion

① 基金项目: 国网辽宁省电力有限公司科技项目 (2024-02-20); 辽工程 GPU 资源支持项目 (2026-01)

收稿时间: 2026-02-06; 修改时间: 2026-02-27; 采用时间: 2026-03-09; csa 在线出版时间: 2026-07-10

2020年9月,我国在第75届联合国大会上宣布于2030年前及2060年前分别实现碳达峰与碳中和的目标^[1].电力行业是我国碳排放主要来源,是实现低碳发展的重要前提^[2].电力负荷预测是电力系统规划和运行的关键任务,其中短期电力负荷预测对实时调度、备用容量配置等具有决定性作用,精准的预测结果对整个电力系统调度、检修以及控制具有重要意义.

针对电力系统负荷预测方法大致可分为3类:统计方法、机器学习法和深度学习法.统计方法包括回归分析、卡尔曼滤波、自回归整合移动平均等.此类方法有操作简单,计算成本低的优点,但此方法缺少鲁棒性,对非线性数据的预测能力有限.机器学习常用的方法有随机森林、人工神经网络、支持向量机(support vector machine, SVM).相较于传统模型,机器学习法能对复杂且大规模的非线性时间序列预测提供更准确的结果.随着计算机性能的增强,深度学习方法因具有强大的非线性能力和高效数据表征被应用于电力负荷预测领域,主要包括时间卷积网络(temporal convolutional network, TCN)^[3]、卷积神经网络(convolutional neural network, CNN)^[4]、门控循环单元(gated recurrent unit, GRU)^[5]、长短期记忆(long short-term memory, LSTM)网络^[6].

基于上述方法,学者们在短期电力负荷预测领域展开大量研究.文献[7]结合卷积神经网络、双向门控循环单元以及注意力机制构建短期电力负荷预测模型,利用卷积神经网络提取负荷数据局部特征.然而,在实际应用中负荷数据具有周期性和复杂动态变化,单一尺度建模可能无法捕捉不同时间尺度的特征,并且会限制模型泛化能力^[8].文献[9]提出基于变分模态分解的LSTM与轻量级梯度提升机预测模型,以解决多特征预测提取特征信息差的问题,但LSTM模型受限于递归架构固有缺陷,存在长期依赖和梯度消失等问题.近年来,Transformer^[10]及其变体模型因能捕获成对依赖关系和交互作用已广泛应用于短期电力负荷预测领域.文献[11]将原始数据经过两次分解并进行相关性分析后输入到Informer模型中,对预测结果进行组合重构生成最终预测值.TFT模型^[12]因平等分配历史数据注意力权重,降低对重要时间节点及突变特征的识别效能,对于负荷关键特征提取能力受限.TimeDiT^[13]模型因其固有的迭代采样特性导致推理速度受限,无法满足短期电力负荷预测的实时响应需求.而iTransformer

模型^[14]将变量维度的序列化表示作为模型核心输入单位,使电力负荷数据中每个特征维度能够单独参与学习,获得更精确的特征表示,提高预测准确度.文献[15]结合时变滤波经验模态分解-重构算法、独立自注意力机制和iTransformer模型预测负荷功率,该预测方法具有较好的预测性能.

目前,为提高数据集质量,通常在模型训练前剔除不相关特征^[16].然而,现有方法大多局限于线性或非线性单一角度的相关性评估,这种方法可能不足以捕捉到特征与负荷之间的交互效应,从而导致特征选择的局限性.同时,电力负荷数据具有时变特性,通常会伴随噪声干扰,多数研究采用分解算法进行数据预处理.文献[17]采用变分模态分解技术对负荷数据进行分解,识别并合并受噪声影响较大的模态分量,以降低原始数据非平稳性并提高预测模型的精度.文献[18]提出一种基于优化变分模态分解和双向长短时记忆网络的混合模型,在变分模态分解中优化分解效果.然而,此类分解算法对原始数据进行处理可能会泄露部分所要预测的目标信息^[19],导致模型在测试集上的预测精度与实际精度存在偏差.此外,电力负荷预测模型训练追求高稳定性和准确度,单一模型通过调参难以有效解决固有缺陷^[20].文献[21]提出一种基于极端梯度提升决策树与改进Informer模型相结合的短期电力负荷预测方法,虽考虑负荷数据特征因素并对位置编码开展优化设计,但上述方法在预测过程中未充分整合已提取的特征信息,限制了模型性能的提升.

综上所述,现有短期电力负荷预测研究面临以下局限:(1)仅从线性或非线性单一角度进行相关性分析;(2)未结合负荷数据的实际特性,对于多尺度特征提取与高噪声干扰处理方面考虑不足;(3)多数研究忽视预测过程中可能出现的序列特征丢失现象,导致已提取特征信息未能得到有效利用^[22].

针对上述问题,提出一种基于iTransformer的短期电力负荷预测模型.首先,通过计算皮尔逊相关系数和最大信息系数对电力负荷数据进行双指标相关性分析,再利用动态频谱滤波模块,自适应调节噪声抑制幅度,去除噪声干扰;其次,采用多尺度时间卷积模块,设计与电力负荷周期性特征相匹配的卷积核,以捕捉复杂时间层次依赖关系;再次,在变量维度上应用多头注意力机制和前馈神经网络,以挖掘不同特征间的多元相关性并进行特征编码;最后,在预测阶段并行加入线

性层级融合模块充分优化模型对所挖掘特征的利用效能,并将其与投影层所得预测结果加权获取整体预测结果。

1 模型概述

1.1 模型总体框架

本文模型结构如图1所示,主要由4部分组成。

(1) 特征提取: 嵌入层将各负荷变量进行独立嵌入,结合动态频谱滤波模块自适应抑制噪声干扰,并利用多尺度时间卷积模块深度挖掘不同尺度下的负

荷特征。

(2) 多头注意力机制: 对负荷特征相关性进行建模,以挖掘全局依赖关系。

(3) 前馈神经网络: 全连接式特征编码,通过全连接层、激活函数和Dropout操作学习非线性表示。

(4) 预测阶段: 编码特征分别输入线性层级融合模块与投影层生成预测结果,并通过加权协同输出得到最终预测值。将投影层全局映射能力与线性层级融合模块对特征信息的整合利用能力相结合,通过并行支路的决策互补显著提升预测精度。

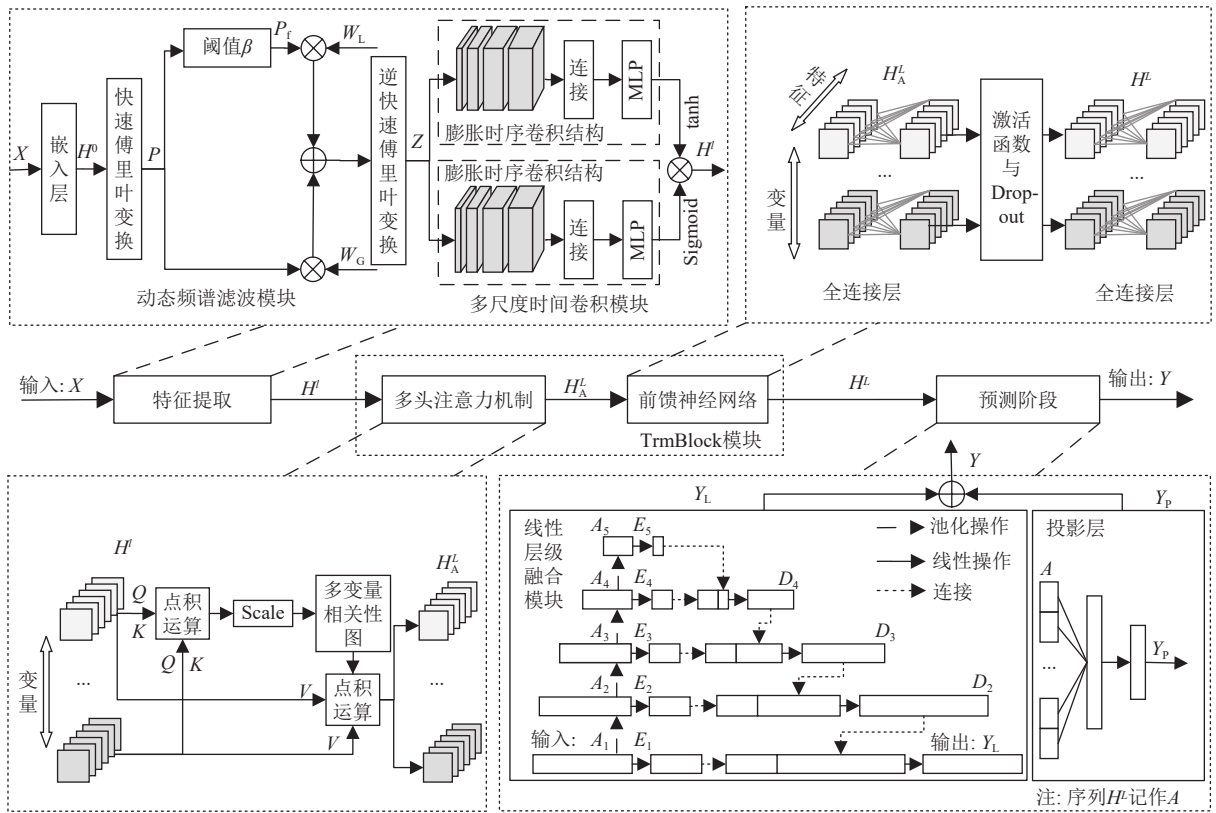


图1 本文模型框架图

1.2 特征提取

1.2.1 嵌入层

嵌入层将单个电力负荷变量独立地嵌入为一个标记 h_n^0 . 输入的电力负荷序列的顺序已经隐含在前馈神经网络的神经元排列顺序中,因此不需要额外引入位置编码,计算公式为:

$$h_n^0 = f_E(X) \quad (1)$$

其中, $f_E(\cdot)$ 为嵌入操作, X 为输入矩阵。

1.2.2 动态频谱滤波模块

电力负荷数据在采集与传输过程中常伴有噪声干

扰,掩盖真实用电模式,影响模型对负荷突变与微小波动识别能力^[23],削弱预测精确度与稳定性. 为此,本文在嵌入层后引入动态频谱滤波模块(dynamic spectral filtering module, DSFM)^[24]. 通过可学习阈值机制动态调节高频成分抑制强度,在保留主要趋势信息的同时有效削弱噪声干扰,并避免固定阈值带来适应性不足问题. 同时,使用快速傅里叶变换处理每个输入数据仅依赖当前和历史数据,滤波操作不涉及未来时间点信息,避免因未来信息泄露而产生的潜在问题。

嵌入标记矩阵 $H^0 = \{h_1^0, \dots, h_n^0\}$ 沿空间维度执行快

速傅里叶变换, 得到综合频域, 其包含输入负荷数据在所有变量维度的频谱特征, 计算公式为:

$$P = f_F(H^0) \quad (2)$$

其中, P 为频域表示, $f_F(\cdot)$ 为快速傅里叶变换操作.

去除高频噪声成分关键在于自适应地从功率谱 G 中滤除高频成分. 可学习阈值 β 在初始化阶段设置为 $[0, 1]$ 区间内随机值, 根据负荷数据频谱特征进行动态调整, 在训练期间通过反向传播进行优化. 高于阈值的频率被保留, 其他频率被过滤掉, 计算公式为:

$$G = |P|^2 \quad (3)$$

$$P_f = P \otimes (G > \beta) \quad (4)$$

其中, $\otimes$ 为逐元素相乘, $(G > \beta)$ 为二元掩码.

随后, 采用全局滤波器从频域数据 P 中学习, 局部滤波器从自适应滤波后数据 P_f 中学习, 并对滤波特征进行整合得到 P_s . 应用逆快速傅里叶变换将频域数据转换回时域, 计算公式为:

$$P_s = (W_G \otimes P) + (W_L \otimes P_f) \quad (5)$$

$$Z = f_I(P_s) \quad (6)$$

其中, W_G 为全局权重, W_L 为局部权重, Z 为时域数据, $f_I(\cdot)$ 为逆快速傅里叶变换操作.

1.2.3 多尺度时间卷积模块

传统方法通常基于单一时间尺度建模, 捕捉负荷数据周期性变化和其他形式时间层次依赖关系可能存在局限. 因此, 本文在文献[25]基础上进行改进, 在多头注意力机制前, 设计多尺度时间卷积模块 (multi-scale temporal convolution module, MTCM). 解决原始 iTransformer 模型全局注意力机制对局部特征感知力弱的局限, 并克服了单一感受野无法适配电力负荷多维周期性特征的不足, 实现多尺度特征提取.

该模块采用两层膨胀时序卷积结构 (dilated temporal convolution structure, DTCS), 第 1 层后接 $\tanh$ 激活函数, 捕获负荷数据复杂非线性特征. 第 2 层配备 Sigmoid 激活函数, 作为门控动态特征调控. 每层 DTCS 采用 1×2 、 1×3 、 1×6 和 1×7 这 4 种与电力负荷周期性特征相匹配的卷积核尺寸捕捉负荷序列在不同时间尺度下的变化规律. 以第 1 个 DTCS 层为例: 首先, 时域数据 Z 通过 4 种不同尺寸卷积核提取多尺度时间依赖性; 其次, 以时间维数最少特征为标准截取相同长度使时间

维数一致; 最后, 在通道维度上连接, 恢复和输入相同通道数量, 计算公式为:

$$Z * O_{1 \times b}(t) = \sum_{s=0}^{b-1} O_{1 \times b}(s) Z(t - d \times s) \quad (7)$$

$$\hat{Z}_1 = f_C(Z * O_{1 \times 2}, Z * O_{1 \times 3}, Z * O_{1 \times 6}, Z * O_{1 \times 7}) \quad (8)$$

其中, $O_{1 \times b}$ 为 $1 \times b$ 大小的卷积核, t 为当前时间步, d 为膨胀因子, s 为卷积核在滑动窗口中的位置, $\hat{Z}_1$ 为经 4 种不同尺寸的卷积核处理后的输出. $f_C(\cdot)$ 表示拼接操作, $*$ 表示卷积操作.

通过多层感知机 (multi-layer perceptron, MLP) 进行线性映射, 先升维再降维, 恢复与输入相同的时间维数, 计算公式为:

$$Z_1 = f_M(\hat{Z}_1) \quad (9)$$

$$f_M(\hat{Z}_1) = f_2(f_R(f_1(\hat{Z}_1))) \quad (10)$$

其中, $f_M(\cdot)$ 为 MLP 操作, $f_1(\cdot)$ 为负责升维的第 1 层线性映射, $f_R(\cdot)$ 为激活函数, $f_2(\cdot)$ 为负责恢复维度的第 2 层线性映射. Z_1 为第 1 层 DTCS 输出.

采用与第 1 层 DTCS 相同处理方式得到输出序列 Z_2 , 并对两层输出分别处理, 计算公式为:

$$H^l = \tanh(Z_1) \otimes \sigma(Z_2) \quad (11)$$

其中, H^l 为经过两层 DTCS 的输出, $\sigma(\cdot)$ 为 Sigmoid 激活函数.

1.3 多头注意力机制

多头注意力机制计算注意力分布值, 挖掘数据之间相关性, 计算公式为:

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

其中, Q 、 K 、 V 分别为查询向量、关键向量、值向量, $Softmax$ 为激活函数, d_k 为输入 K 的维度.

1.4 前馈神经网络

前馈神经网络对特征表示 H_A^L 进行编码. 每层 Trm-Block 包括多头注意力机制、层归一化以及前馈神经网络, 计算公式为:

$$H^{l+1} = f_T(H^l), l = 0, \dots, L-1 \quad (13)$$

其中, $f_T(\cdot)$ 为每层 TrmBlock 内操作, L 为层数.

1.5 预测阶段

1.5.1 线性层级融合模块

实际应用中, 负荷受多因素耦合影响且不同场景

下特征具有异质性. 若在预测阶段仅依赖单一线性投影层, 模型无法充分利用已提取特征, 从而限制预测性能与跨场景适应能力. 为解决这一问题, 引入线性层级融合模块 (linear hierarchical fusion module, LHFM)^[26]. 该模块采用由深至浅的层级化整合机制, 将更高深度负荷特征逐层整合到更低级别负荷特征中, 并与投影层构成并行结构, 从而更有效利用多层次特征信息, 提升模型对短期负荷的综合预测能力.

为清晰表示, 将输入序列 H^L 记作 A . 采样参数设置为: 卷积核大小 g 为 3, 步长 c 为 2, 填充 p 为 1. 以 5 层数据为例, 计算公式为:

$$A_1 = A \quad (14)$$

$$A_i = f_A(A_{i-1}) \quad (15)$$

$$R(T_i) = \left\lceil \frac{T_{i-1} + 2 \times p - g}{c} + 1 \right\rceil \quad (16)$$

其中, i 为索引 ($1 < i \leq 5$), $R(T_i)$ 表示输出的时间维度, $f_A(\cdot)$ 为池化操作.

为充分利用这些特征, 采用多个线性层将其映射到中间特征 E . 线性层将时间维数 $T_i + T_{i-1}$ 映射回 T_{i-1} . 第 i 层特征与第 $i-1$ 层特征在时间维度上进行拼接, 并利用线性层生成第 $i-1$ 层的输出特征 D , 最终得到输出 Y_L , 计算公式为:

$$D_{i-1} = f_L(f_C(D_i, E_{i-1})), 1 < i \leq 5 \quad (17)$$

其中, $i = 5$ 时, $D_4 = f_L(f_C(D_5, E_4))$; $i = 2$ 时, $Y_L = D_1$ 为第 1 层输出.

1.5.2 预测决策协同机制

前馈神经网络编码后的特征分别输入线性层级融合模块与投影层生成预测结果, 并通过加权方式得到最终输出, 计算公式为:

$$Y_P = f_P(H^L) \quad (18)$$

$$Y = Y_L \oplus Y_P \quad (19)$$

其中, Y_P 为投影层输出, $f_P(\cdot)$ 为投影层操作, h_n^L 为 H^L 中第 L 个序列, Y 为模型最终输出.

2 算例及结果分析

2.1 实验数据选取

为全面评估本文模型预测性能, 选取 3 个不同地理范围、气候的公开数据集. (1) 数据集 1 选取 Kaggle

平台巴拿马某城区电力负荷数据集, 时间范围为 2015 年 1 月 31 日–2019 年 4 月 10 日. 时间粒度为 1 h, 输入长度为 168, 共 36 720 条数据. (2) 数据集 2 选用“2018 年第 6 届国际可再生与可持续能源会议”发布的非洲北部得土安市 2017 年整年电力负荷数据集. 时间粒度为 10 min, 输入长度为 144, 共 52 416 条数据. (3) 数据集 3 选用福建省泉州市 2018 年整年电力负荷数据集, 电力负荷数据来源于中国国家电网, 气象数据来源于中国气象局. 时间粒度为 15 min, 输入长度为 96, 共 35 041 条数据. 3 个数据集均依据 7:2:1 比例, 将数据划分为训练集、测试集与验证集.

2.2 数据预处理

数据集不包含任何缺失值、损失值及离群值, 因此无需进行缺失值填补或其他数据清洗操作. 为消除电力负荷数据与不同影响因素之间的量纲差异, 对数据进行 Min-Max 归一化处理. 将数值范围统一到 $[0, 1]$ 区间. 最后对预测值以及误差进行反归一化操作, 还原到实际数据值范围, 计算公式为:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (20)$$

$$x^* = x'(x_{\max} - x_{\min}) + x_{\min} \quad (21)$$

其中, x' 为归一化后的负荷数据, x 为原始负荷数据, $x_{\max}$ 和 $x_{\min}$ 分别为该数据的最大值和最小值, x^* 为反归一化后的负荷预测值.

2.3 评价指标

为了客观地量化模型在短期电力负荷预测任务中的表现, 本文选取平均绝对误差 (MAE)、均方根误差 (RMSE)、平均绝对百分比误差 (MAPE) 和决定系数 (R^2) 作为模型评价指标, 计算公式为:

$$e_{\text{MAE}} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (22)$$

$$e_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (23)$$

$$e_{\text{MAPE}} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (24)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (25)$$

其中, $\hat{y}_i$ 为电力负荷预测值, y_i 为电力负荷真实值, $\bar{y}$ 为电力负荷真实值的平均值, n 为预测点数目。

2.4 双指标相关性分析

为最大限度提高模型特征集构建能力, 从线性和非线性角度, 采用皮尔逊相关系数和最大信息系数 (maximal information coefficient, *MIC*) 进行双指标相关性分析, 计算公式为:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (26)$$

其中, r 为皮尔逊相关系数, n 为样本总数, X_i 和 Y_i 为变量 X 和 Y 第 i 个对应值, $\bar{X}$ 和 $\bar{Y}$ 为样本均值。

$$MIC(X, Y) = \max_{G_x G_y < B(n)} \left\{ \frac{I(X, Y)}{\log \min\{G_x, G_y\}} \right\} \quad (27)$$

其中, G_x 和 G_y 为在 X 和 Y 上的划分个数; $I(X, Y)$ 为当前划分下的最大互信息; $\log \min\{G_x, G_y\}$ 为归一化项, 使 MIC 值始终落在 $[0, 1]$ 之间, G_x 与 G_y 的乘积需小于 $B(n)$, 一般 $B(n) = n^{0.6}$ 。

皮尔逊相关系数可分析变量间线性相关性, $|r|$ 越大, 说明相关性越强。 MIC 可衡量变量之间的非线性相关性。 MIC 值越大, 则相关性越强。 以数据集 1 为例进行双指标相关性分析, 具体结果见表 1。

表 1 数据集 1 特征双指标相关性分析

特征名称	特征含义	皮尔逊相关系数	<i>MIC</i>
week_X-2	预测前第2周的负荷滞后	0.88	0.63
week_X-3	预测前第3周的负荷滞后	0.88	0.63
week_X-4	预测前第4周的负荷滞后	0.88	0.64
MA_X-4	从预测前第1-4周的负荷移动平均值	0.92	0.73
dayOfWeek	从星期六开始计算, 一周中的某一天	0.23	0.14
weekend	周末的二元指标	-0.29	0.17
holiday	节日的二元指标	-0.18	0.07
hourOfDay	一天中的小时	0.46	0.59
T2M_toc	温度	0.68	0.41
QV2M_toc	相对湿度	-0.06	0.04
TQL_toc	降水量	0.08	0.04
W2M_toc	风速	0.1	0.04

去除皮尔逊相关系数和 MIC 绝对值均低于 0.1 变量。 筛选后气象特征包括温度和风速: 温度与负荷之间呈高度相关性, 尤其在空调制冷或采暖季, 气温变化对负荷有直接驱动作用; 风速虽对负荷影响较弱, 但在一定

程度上会影响温度, 可能间接提升用电量。 时间特征与用电特征主要包括历史负荷数据与时间属性: 历史负荷数据与负荷之间呈高度相关性, 能够反映用户行为时间滞后效应与负荷长期趋势变化; 时间属性包括时段划分和节假日类型, 将周末和节假日均归为非工作日, 其与电力负荷有较强相关性; 故虽 MIC 绝对值低于 0.1, 也将其保留。 工作日负荷呈明显周期性, 如凌晨低谷、早晚高峰等; 非工作日电量显著高于工作日, 体现出时间属性对用电行为规律性影响。

2.5 实验设置

本文实验基于 PyTorch 2.3.1 深度学习框架进行实现, 采用 Python 3.9 编程语言。 实验运行环境: 单个 NVIDIA RTX 3090 显卡, 配备 24 GB 显存, 操作系统为 CentOS7。 为防止训练模型过拟合, 采用“早停法”对模型进行训练, 早停轮数设置为 5, 模型超参数设置见表 2, 对比模型参数设置见表 3。

表 2 模型超参数设置

超参数	参数值
预测周期	周、日
初始学习率	0.0001
损失函数	MSE
丢弃率	0.1
输入长度	168、144、96
编码器维度	11
编码器层数	3
前馈网络维度	2048
训练轮数	100
激活函数	GELU

表 3 对比模型参数设置

模型	模型参数设置
SVM	核函数: 径向基函数核, 惩罚参数: 1
GRU	隐藏层维度: 64, 丢弃率: 0.1
LSTM	隐藏层维度: 64, 丢弃率: 0.1, 批次大小: 32
ES-dRNN	初始学习率: 0.003, Batch: 16, 隐藏层维度: 40, 层数: 3
Crossformer	前馈网络隐藏层维度: 512, 学习率: 0.0001, 多头注意力头数: 4, 编码器层数: 3, 批次大小: 32
TimesNet	前馈网络隐藏层维度: 64, 学习率: 0.0001, 重要频率数目: 5, 卷积核数量: 6
iTransformer	前馈网络隐藏层维度: 2048, 学习率: 0.0001, 批次大小: 32, 编码器层数: 2, 多头注意力头数: 8

2.6 对比实验结果与分析

为评估本文所提模型预测效果, 选取领域主流模型 LSTM、GRU、Crossformer^[27]、TimesNet^[28], 经典机器学习模型 SVM 以及结合扩张递归神经网络、GRU

和注意力机制的短期电力负荷预测模型 ES-dRNN^[29] 进行对比实验, 实验结果见表 4.

表 4 数据集 1 不同模型预测结果

模型	MAE (MW)	RMSE (MW)	MAPE (%)	R^2
SVM	86.406	104.126	6.81	0.654
GRU	63.475	74.877	5.41	0.805
LSTM	57.018	72.381	4.82	0.827
ES-dRNN	31.825	54.837	2.64	0.880
Crossformer	36.985	43.838	2.95	0.921
TimesNet	31.629	37.380	2.58	0.958
iTransformer	25.889	36.667	2.11	0.966
本文模型	24.540	34.163	1.84	0.977

注: 加粗数据表示最优结果

由表 4 可知, 在数据集 1 下本文模型评价指标均为最优值. 相较于 SVM、GRU 和 LSTM 此类通常采用单尺度特征映射方式的基础模型, MAE 分别降低 71.6%、61.3% 和 56.9%, RMSE 分别降低 67.2%、54.4% 和 52.8%, MAPE 分别降低 72.9%、65.9% 和 61.8%, R^2 提升 49.4%、21.4% 和 18.1%, 说明本文模型可捕捉电力负荷在多时间尺度下的复杂依赖关系, 在提取长期周期性与短期波动趋势特征时展现出更优越性能, 从而提升短期电力负荷预测精度. 相较于 ES-dRNN 模型, MAE、RMSE 和 MAPE 分别降低 22.9%、37.7% 和 30.3%, R^2 上升 11.0%. ES-dRNN 和本文模型均为组合模型, 但本文模型在评价指标上均表现更优. 说明本文模型具有更强的非平稳信号处理能力与预测稳定性, 降低由负荷随机波动引起的预测偏离, 提升电网调度所需的预测可靠性. 相较于 Crossformer、TimesNet 和 iTransformer, MAE 分别降低 33.6%、22.4% 和 5.2%, RMSE 分别降低 22.1%、8.6% 和 6.8%, MAPE 分别降低 37.6%、28.7% 和 12.8%, R^2 分别上升 6.1%、2.0% 和 1.1%. 说明本文模型在预测阶段能够更有效整合负荷时间依赖信息及非线性特征, 避免高维特征在传递过程中的信息丢失, 从而增强模型综合预测能力, 实现更高的短期电力负荷预测精度与稳定性. 为更直观地展示本文提出方法的准确性, 在预测结果中, 选取某周的结果进行可视化对比, 如图 2 所示.

由图 2(b) 分析可得, SVM 和 GRU 模型在负荷剧烈变化阶段存在明显低估, 预测曲线与真实曲线间偏差较大. LSTM 模型相较于 SVM 和 GRU 模型, 能够更准确地反映负荷快速上升和下降趋势, 但仍存在低估问题. ES-dRNN 模型整体预测曲线较为平滑, 能够跟踪主要趋势, 但在峰值与谷值处拟合精度相对不

足. Crossformer 模型在负荷上升阶段能够较好跟随真实曲线的变化, 但在负荷下降过程中存在明显滞后, 且曲线在快速变化区间响应不够灵敏. TimesNet 模型对上升趋势响应较为及时, 但对下降趋势响应滞后, 曲线表现过于平滑, 未能充分反映负荷瞬时扰动细节差异. iTransformer 模型能够较为准确地跟踪负荷上升与下降趋势, 在峰值和谷值区域表现较为优异. 本文模型在峰谷拟合上表现最佳, 显示出对负荷波动的强适应能力, 在总体趋势上也更准确贴合真实数据. 同时, 通过图 2(a) 也可知, 本文模型取得最优预测效果.

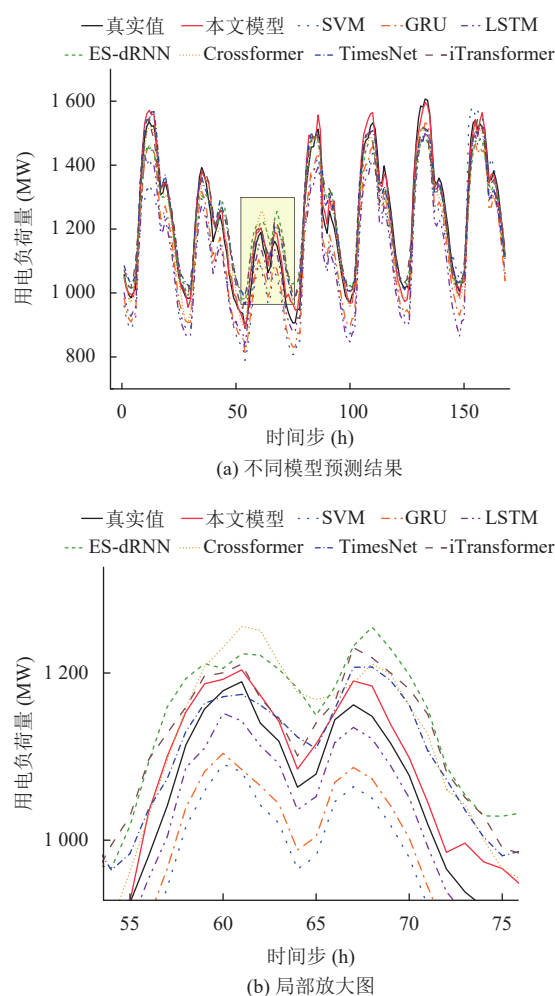


图 2 数据集 1 中某周的不同模型预测结果

2.7 模型泛化性分析

为验证所提模型的泛化能力, 选用与数据集 1 不同地区的数据集 2 进行实验分析, 以检验模型在多样场景与条件下适用性, 并保证其对于未见过数据能够持续输出可靠且精确的预测结果. 对数据集 2 进行双

指标相关性分析,与数据集1筛选标准相同,所有特征对于实际预测目标均有影响,数据集2特征详见表5.

表5 数据集2特征含义

特征名称	特征含义
Temperature	温度
Humidity	湿度
Wind Speed	风速
General diffuse flows	一般扩散流
Diffuse flows	扩散流

扩散流描述在某些气象条件下,风速和风向大范围、缓慢变化.这种渐进变化对风能发电和区域电力负荷分布有显著影响.风速逐渐变化会影响依赖风能的区域的电力负荷消耗.温度对电力负荷有直接影响,高温天气因降温需求增加用电,低温天气则因取暖需求提升电力需求.湿度对电力负荷的影响较小,主要通过影响舒适度间接调节电力需求.扩散流和风速变化直接导致发电与负荷分布的不稳定,而温度和湿度则通过影响设备运行需求,间接调节整体电力消耗.

选用表3提到的7种模型在同一环境下进行实验,实验结果见表6.为更直观地展示各模型预测效果,选取某天的预测结果进行可视化对比,结果如图3所示.

表6 数据集2不同模型预测结果

模型	MAE (kW)	RMSE (kW)	MAPE (%)	R^2
SVM	950.390	1126.212	7.58	0.804
GRU	908.296	1089.390	6.98	0.837
LSTM	866.938	963.476	6.04	0.860
ES-dRNN	738.690	925.977	5.96	0.891
Crossformer	934.492	1043.710	7.23	0.811
TimesNet	660.798	752.401	4.86	0.929
iTransformer	605.340	780.389	4.90	0.923
本文模型	567.667	748.005	4.67	0.949

注:加粗数据表示最优结果

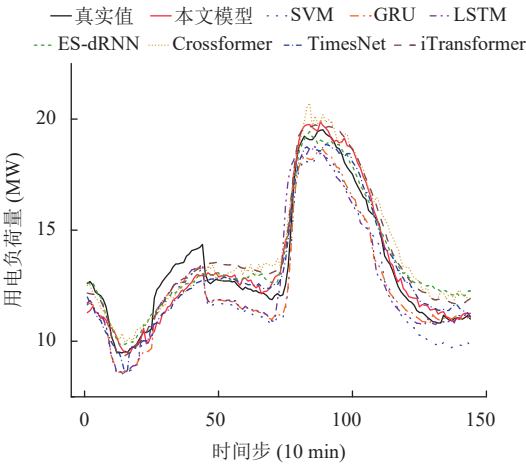


图3 数据集2中某天的不同模型预测结果

对数据集2实验结果分析可知,本文模型在与7种对比模型性能对比中,MAE、RMSE、MAPE和 R^2 分别为567.667、748.005、4.67%和0.949,均取得最优结果且误差显著低于其他模型,验证所提模型在跨地区、负荷模式差异显著的复杂场景下的优越性能.预测曲线对比进一步表明,所提模型不仅能够准确捕捉负荷变化的整体趋势,还在峰值、低谷以及快速波动区域表现出高度贴近实际值的预测能力,验证本文模型在短期负荷预测任务中的泛化能力.

2.8 消融实验结果与分析

本节共设计6组实验,均在数据集1上进行训练,以验证各模块有效性及协同作用.实验结果见表7,实验设置如下.

- (1) 实验1: 本文模型,用于作为性能对比基准.
- (2) 实验2: 将DSFM模块可学习阈值 β 设置为固定值0.3.
- (3) 实验3: 将MTCM模块中 1×3 和 1×7 卷积核替换为 1×4 和 1×8 .
- (4) 实验4: 仅移除特征提取层的DSFM模块,将未经去噪处理的负荷数据直接输入MTCM模块.
- (5) 实验5: 仅移除特征提取层的MTCM模块,将经过去噪处理而未进一步特征提取的负荷数据直接输入多头注意力机制中.
- (6) 实验6: 仅移除LHFM模块,未融合多尺度特征信息,在预测与融合层仅使用投影层.

表7 数据集1消融实验结果

实验	MAE (MW)	RMSE (MW)	MAPE (%)	R^2
1	24.540	34.163	1.84	0.977
2	24.750	34.564	1.95	0.972
3	24.796	34.471	1.92	0.974
4	24.994	35.414	2.01	0.971
5	24.919	34.968	1.98	0.972
6	25.036	35.430	1.95	0.969

实验2将DSFM模块可学习阈值设置为固定值0.3,MAE、RMSE和MAPE分别提升0.86%、1.17%和5.98%, R^2 降低0.51%.说明固定阈值限制模型对频谱特征自适应去噪能力,从而验证DSFM模块可学习阈值设计在抑制噪声干扰的必要性.实验3卷积核数量不变,改变其尺度范围与感受野,MAE、RMSE和MAPE分别上升1.04%、0.90%和4.35%, R^2 降低0.31%.说明负荷波动建模能力受限,原始卷积核配置更贴近负荷数据时间结构特性.实验4仅移除DSFM模块,存

在高噪声成分干扰的情况下, MAE、RMSE 和 MAPE 分别上升 1.85%、3.66% 和 9.24%, R^2 降低 0.61%。说明去除噪声抑制后, 负荷数据中高频干扰未被过滤, 导致模型后续难以准确建模负荷数据本质趋势, 验证 DSFM 模块在应对负荷数据波动性与非平稳性重要作用。实验 5 仅移除 MTCM 模块后, 模型特征提取能力受限, MAE、RMSE 和 MAPE 分别提升 1.54%、2.36% 和 7.61%, R^2 降低 0.51%, 说明缺乏 MTCM 模块影响对负荷波动趋势建模能力。而将 DSFM 模块去噪处理后的数据输入 MTCM 模块进行特征提取, 更能准确反映负荷波动趋势, 解决原始 iTransformer 模型全局注意力在局部特征捕捉方面不足, 从而增强模型适应复杂时序变化能力。实验 6 仅移除 LHFM 模块, 未充分利用提取到的特征信息, MAE 与 RMSE 达到最高值, MAE、RMSE 和 MAPE 分别提升 2.02%、3.71% 和 5.98%, R^2 降低 0.82%。说明仅依赖单一预测结构, 多尺度特征信息难以实现有效融合, 部分负荷特征在预测过程中发生丢失, 进而影响模型拟合能力与泛化性能。

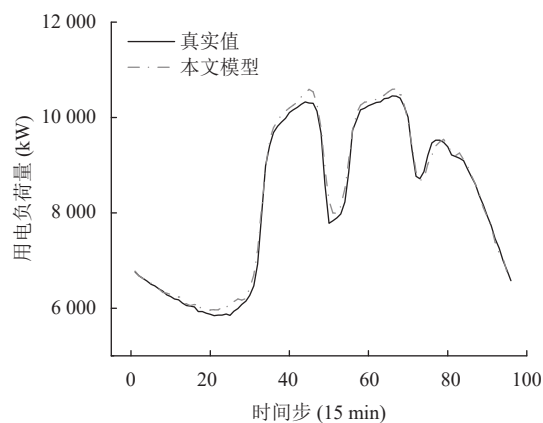
综上所述, DSFM 模块为后续特征建模提供更加稳定的输入表示, 降低噪声干扰对特征学习过程的影响。MTCM 模块通过多尺度卷积操作, 对经过 DSFM 模块去噪处理的数据进行进一步特征提取, 从而增强模型适应复杂时序变化的能力。LHFM 模块通过特征融合机制实现多尺度信息有效整合, 使多尺度特征在预测阶段得到充分利用, 提升模型对复杂模式的理解能力, 从而显著提高最终预测的精度。本文模型最优表现验证各模块协同设计的有效性与合理性及其在短期电力负荷预测中的重要性。

2.9 模型适用性分析

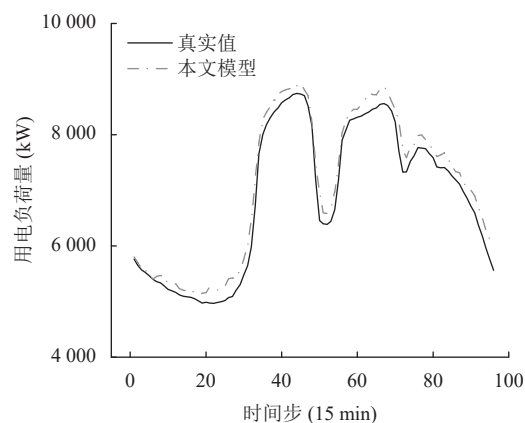
本节在数据集 1 上进行对比实验和消融实验, 以初步验证本文模型, 在数据集 2 上评估模型泛化能力。由于工作日负荷具有规律性和明显峰谷特征, 而非工作日负荷呈现更随机的波动模式。为验证模型在不同日类型下的适用性, 数据集 3 中随机选择工作日 (2018 年 10 月 19 日) 和非工作日 (2018 年 10 月 22 日) 进行分析。数据集 1 和 2 已通过实验验证模型相对优势, 因此数据集 3 不再重复验证相对优势。与数据集 1 筛选标准相同, 对数据集 3 进行双指标相关性分析, 筛选后特征包括最高温度、最低温度、平均温度和相对湿度。使用相同的实验流程和环境, 以确保结果可靠性, 实验结果见表 8, 可视化结果如图 4 所示。

表 8 数据集 3 工作日与非工作日实验结果

类型	MAE (kW)	RMSE (kW)	MAPE (%)	R^2
工作日	112.615	144.045	1.38	0.992
非工作日	338.965	394.423	4.83	0.912



(a) 工作日 (2018年10月19日)



(b) 非工作日 (2018年10月22日)

图 4 数据集 3 中工作日与非工作日预测结果

由实验结果和可视化结果可知, 本文模型在不同日类型下均表现出较高预测精度。预测曲线在整体周期趋势和局部波动细节上与真实负荷高度一致, 尤其在峰值与谷值区域, 模型能够准确跟踪负荷的快速变化, 未出现明显滞后或振荡现象。这表明, 模型在结构差异明显的工作日与非工作日负荷下均具备稳定的短期预测能力, 体现出较强的鲁棒性与适应性。

3 结论

本文针对特征提取不足、高噪声成分干扰以及单一模型在预测方面时序特征损失等问题, 提出一种融合多尺度特征的短期电力负荷预测模型。通过双指标相关性分析, 选取相关性较高的特征作为输入, 提高模型特征集构建能力; 采用动态频谱滤波模块通过频域

与时域转换和自适应阈值,在避免信息泄露的基础上,减少噪声在预测中的影响,为模型建立高质量的数据输入基础;多尺度时间卷积模块能够考虑不同时间尺度下电力负荷数据的时序特征信息,实现对电力负荷数据复杂规律及依赖关系的捕捉,使模型的预测性能有所提升;引入线性层级融合模块借鉴逐层特征融合机制思想,有效整合多尺度特征,并通过与投影层输出的决策互补,充分挖掘特征在负荷预测过程中的潜力,提高模型预测能力。

本文使用3个不同公开数据集进行实验验证,实验结果表明,本文提出的模型在各评估指标上均达到最优表现,并且在泛化能力与预测效果方面取得显著优势。消融实验则验证本文模型各模块的有效性及其协同作用。未来研究可致力于探索更高效的优化算法,构建计算量更小且精度更高的短期电力负荷预测模型。此外,还可考虑电价及能源供需波动对电力负荷的影响,以使模型更好地适应实际应用场景,增强其在复杂和动态环境中适用性与鲁棒性。

参考文献

- 1 陈胜,卫志农,顾伟,等.碳中和目标下的能源系统转型与变革:多能流协同技术.电力自动化设备,2021,41(9):3–12. [doi: 10.16081/j.epae.202109037]
- 2 李响,牛赛.双碳目标下源-网-荷多层评价体系研究.中国电机工程学报,2021,41(S1):178–184. [doi: 10.13334/j.0258-8013.pcsee.211576]
- 3 李飞宏,肖迎群.基于EMD-TCN-ELM的短期电力负荷预测.计算机系统应用,2022,31(11):223–229. [doi: 10.15888/j.cnki.csa.008781]
- 4 吕志星,张斌,王沈征,等.基于K-Means和CNN的用户短期电力负荷预测.计算机系统应用,2020,29(3):161–166. [doi: 10.15888/j.cnki.csa.007287]
- 5 刘蓉晖,石炬烽,孙改平,等.基于MV-WC和门控循环单元的短期净负荷概率预测.南方电网技术,2025,19(6):152–161. [doi: 10.13648/j.cnki.issn1674-0629.2025.06.014]
- 6 朱继忠,苗雨旺,董朝阳,等.基于Attention-LSTM与多模型集成的短期负荷预测方法.电力工程技术,2023,42(5):138–147. [doi: 10.12158/j.2096-3203.2023.05.016]
- 7 任爽,杨凯,商继财,等.基于CNN-BiGRU-Attention的短期电力负荷预测.电气工程学报,2024,19(1):344–350. [doi: 10.11985/2024.01.037]
- 8 李科,潘庭龙,许德智.基于MSCNN-BiGRU-Attention的短期电力负荷预测.中国电力,2025,58(6):10–18. [doi: 10.11930/j.issn.1004-9649.202406098]
- 9 张未,余成波,王士彬,等.基于VMD-LSTM-LightGBM的多特征短期电力负荷预测.南方电网技术,2023,17(2):74–81. [doi: 10.13648/j.cnki.issn1674-0629.2023.02.009]
- 10 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- 11 石卓见,冉启武,徐福聪.基于聚合二次模态分解及Informer的短期负荷预测.电网技术,2024,48(6):2574–2583. [doi: 10.13335/j.1000-3673.pst.2023.1467]
- 12 Lim B, Arık SÖ, Loeff N, *et al.* Temporal fusion Transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 2021, 37(4):1748–1764. [doi: 10.1016/j.ijforecast.2021.03.012]
- 13 Cao DF, Ye W, Zhang YZ, *et al.* TimeDiT: General-purpose diffusion Transformers for time series foundation model. arXiv:2409.02322, 2024.
- 14 Liu Y, Hu TG, Zhang HR, *et al.* iTransformer: Inverted Transformers are effective for time series forecasting. Proceedings of the 12th International Conference on Learning Representations. Vienna: OpenReview.net, 2024. 7–11.
- 15 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.电网技术,2025,49(6):2436–2445. [doi: 10.13335/j.1000-3673.pst.2024.1970]
- 16 陈景文,黄羽倩,刘耀先,等.基于复合因子构造的KAN-BiLSTM电力负荷预测方法.中国电力,2025,58(12):178–189,198. [doi: 10.11930/j.issn.1004-9649.202504033]
- 17 张羽珊,周亚同.基于VMD-SSA-RF算法的短期电力负荷预测模型优化.电工电能新技术,2024,43(12):30–39. [doi: 10.12067/ATEEE2307021]
- 18 谢国民,陆子俊.基于优化VMD和BiLSTM的短期负荷预测.电力系统及其自动化学报,2025,37(4):30–39. [doi: 10.19635/j.cnki.csu-epsa.001508]
- 19 Niyogi D. A novel method combines moving fronts, data decomposition and deep learning to forecast intricate time series. arXiv:2303.06394, 2023.
- 20 Zhong B, Yang L, Li BR, *et al.* Short-term power grid load forecasting based on VMD-SE-Bilstm-Attention hybrid model. International Journal of Low-carbon Technologies, 2024, 19:1951–1958. [doi: 10.1093/ijlct/ctae143]
- 21 俞胜,孙可,蔡华,等.结合极端梯度提升决策树与改进Informer的短期电力负荷预测方法.中国电力,2025,58(10):195–205. [doi: 10.11930/j.issn.1004-9649.202412004]
- 22 杨佳泽,王灿,王增平.新型电力系统背景下的智能负荷预

- 测算法研究综述. 华北电力大学学报(自然科学版), 2025, 52(3): 54–67. [doi: [10.3969/j.issn.1007-2691.2025.03.06](https://doi.org/10.3969/j.issn.1007-2691.2025.03.06)]
- 23 Zhou LC, Zhou Y, Liu LL, *et al.* A comparative study of different deep learning methods for time-series probabilistic residential load power forecasting. *Frontiers in Energy Research*, 2024, 12: 1490152. [doi: [10.3389/fenrg.2024.1490152](https://doi.org/10.3389/fenrg.2024.1490152)]
- 24 Eldele E, Ragab M, Chen ZH, *et al.* TSLANet: Rethinking Transformers for time series representation learning. *Proceedings of the 41st International Conference on Machine Learning*. Vienna: OpenReview.net, 2024. 12409–12428.
- 25 Liu Z, Huang X, Zhang JY, *et al.* Multivariate time-series anomaly detection based on enhancing graph attention networks with topological analysis. *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. Boise: ACM, 2024. 1555–1564.
- 26 Li C, Xiao BJ, Yuan QP, *et al.* UnetTSF: A better performance linear complexity time series prediction model. *arXiv:2401.03001*, 2024.
- 27 Zhang YH, Yan JC. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. *Proceedings of the 11th International Conference on Learning Representations*. Kigali: OpenReview.net, 2023. 10.
- 28 Wu HX, Hu TG, Liu Y, *et al.* TimesNet: Temporal 2D-variation modeling for general time series analysis. *Proceedings of the Eleventh International Conference on Learning Representations*. Kigali: OpenReview.net, 2023.
- 29 Smyl S, Dudek G, Pelka P. ES-dRNN with dynamic attention for short-term load forecasting. *Proceedings of the 2022 International Joint Conference on Neural Networks (IJCNN)*. Padua: IEEE, 2022. 1–8.

(校对责编: 李慧鑫)