

基于动态多尺度与双重注意力的短期电力负荷预测

朱莉¹, 高靖凯^{1*}, 朱春强^{2,3}, 邓凡¹

(1. 西安科技大学计算机科学与技术学院, 陕西 西安 710054; 2. 西安交通大学计算机科学与技术学院, 陕西 西安 710049;
3. 国网陕西省电力公司培训中心, 陕西 西安 710032)

摘要: 短期电力负荷预测在电力系统的优化调度和安全运行中具有至关重要的作用。电力负荷数据具有多周期特性, 在不同时间尺度上表现出不同的模式和趋势, 准确提取尺度大小有助于识别和分离这些特征。目前方法通过使用一个或一组固定的 patch 长度作为步长, 将称之为 patches 的片段来编码时间序列, 但其无法适应现实世界负荷序列数据的复杂的动态变化。为此, 提出一种基于动态多尺度与双重注意力的预测模型(MDAT)。首先, 利用逐次变分模态分解(SVMD)分离负荷序列不同的时间模式, 通过快速傅里叶变换(FFT)提取出每个模式的显著周期。其次, 根据检测到的显著周期, 将负荷序列以不同大小的 patch 划分为不同的时间分辨率, 使用 Transformer 的多个分支同时建模不同尺度分割序列的依赖关系。然后, 对这些 patches 进行双重注意力, 以捕获全局相关性和局部细节。最后, 对每个分支的输出进行非线性特征融合, 通过堆叠多层 Transformer 模块得到最终的负荷预测结果。在两个公开数据集上的实验结果表明, 该模型在预测精度指标上表现良好, 相比最新的基于 Transformer 及多层感知器(MLP)的模型, 在 Australia 数据集和 Morocco 数据集上平均绝对误差(MAE)分别降低了 10.26%~17.06% 和 9.08%~70.25%。

关键词: 短期负荷预测; 逐次变分模态分解; 多尺度特征; 双重注意力; Transformer 模块

中图分类号: TP391

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0070067

Short-term Power Load Forecasting Based on Dynamic Multi-Scale and Dual Attention Mechanisms

ZHU Li¹, GAO Jingkai^{1*}, ZHU Chunqiang^{2,3}, DENG Fan¹

(1. School of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an 710054, Shaanxi, China;
2. School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China;
3. State Grid Shaanxi Electric Power Company Training Center, Xi'an 710032, Shaanxi, China)

【Abstract】 Short-term power load forecasting plays a crucial role in the optimal scheduling and safe operation of power systems. Power load data exhibit multiperiod characteristics, showing different patterns and trends at various time scales. Accurately extracting the scale size helps identify and separate these features. Current methods use a fixed patch length or a set of fixed patch lengths as steps and encode time series into segments called patches. However, these methods cannot adapt to the complex dynamic changes in real-world load series data. Therefore, this paper proposes a prediction model based on a dynamic Multi-scale and Dual Attention Transformer (MDAT). First, Successive Variational Mode Decomposition (SVMD) is used to separate different time patterns in the load series, and Fast Fourier Transform (FFT) is performed to extract the significant period of each pattern. Subsequently, based on the detected significant periods, the load series is divided into different time resolutions using patches of varying sizes, and multiple branches of a transformer are used to simultaneously model the dependencies of the sequences segmented at different scales. Next, dual attention is applied to these patches to capture the global correlations and local details. Finally, nonlinear feature fusion is performed on the outputs of each branch, and the final load prediction results are obtained by stacking multiple transformer modules. Experimental results on two public datasets demonstrate that the proposed model performs well in terms of prediction accuracy. Compared to the latest models based on Transformer and Multilayer Perceptron (MLP), the Mean Absolute Error (MAE) on the Australia and Morocco datasets is reduced by 10.26%-17.06% and 9.08%-70.25%, respectively.

【Key words】 short-term load forecasting; Successive Variational Mode Decomposition (SVMD); multi-scale features; dual attention; Transformer module

收稿日期: 2024-07-02 修回日期: 2024-08-14

基金项目: 国网陕西省电力有限公司科技项目(5226PX240003); 国网陕西省电力有限公司数字化项目(B326PX230001, B326PX230000); 陕西省自然科学基金基础研究项目(2022JM317)。

通信作者 E-mail: * 669675342@qq.com

0 引言

在现代电力系统中,准确的短期电力负荷预测对保障电力供应的稳定性和提高电网运营效率至关重要。随着社会经济的发展和水平的提高,电力需求变得愈发复杂和多变。可再生能源和分布式发电技术的普及,进一步增加了电力负荷的波动性和不确定性^[1]。电力负荷预测不仅影响电力系统的安全运行和经济效益,还对能源资源的合理配置和环境保护具有重要意义^[2]。

最早期的预测方法主要依赖于时间序列分析技术,如自回归移动平均模型(ARMA)^[3]、指数平滑法(ETS)^[4]等。这些方法在一定程度上能够捕捉电力负荷的周期性和趋势性变化,但对非线性和高波动性的负荷数据预测效果有限。随着计算技术和数据挖掘技术的发展,人工神经网络(ANN)^[5]、支持向量机(SVM)^[6]和随机森林(RF)^[7]等机器学习方法被引入到电力负荷预测中。这些方法能够处理非线性关系,并具有更强的泛化能力。INDIRA 等^[8]通过自适应藤壶交配优化器自动调整 ANN 的权重,从而实现更快的收敛速度和更高的预测精度。然而,机器学习方法也存在一些问题,如特征选择和参数优化复杂,模型容易过拟合。

近年来,深度学习方法因其自动特征提取和捕捉长期依赖关系的能力,在负荷预测中得到了广泛应用。卷积神经网络(CNN)^[9]最初主要用于图像处理,但后来被应用于处理时间序列数据,特别是在提取局部特征和模式方面表现出色。随后,循环神经网络(RNN)^[10]被引入用于处理序列数据。SMYL 等^[11]提出一种配备新型扩展循环单元的多层 RNN,通过有效模拟时间序列中的短期和长期依赖性,提高模型在非线性随机预测问题上的表达能力。然而,由于 RNN 存在梯度消失和梯度爆炸的问题,长短期记忆(LSTM)网络^[12]作为其改进版本,通过其独特的门控机制,能够有效地捕捉和处理时间序列数据中的长期依赖关系。李丽芬等^[13]提出一种时间卷积 LSTM 网络的负荷预测方法,通过 LSTM 层挖掘时间序列的长期依赖关系,提高了预测性能。然而,这些方法在捕捉复杂的长距离依赖和上下文信息时可能不够充分。在处理多尺度、多模式的数据时,其表达能力可能受到限制。

在深度学习模型中,Transformer 在序列处理领域取得了显著成功^[14]。通过自注意力机制,Transformer 可以直接关注输入序列中的任意位置,在捕捉长距离依赖关系方面更加有效。通过多

层堆叠和多头注意力机制,Transformer 具有更强的表达能力,能够更好地捕捉序列中的复杂模式和关系^[15]。DU 等^[16]提出时间序列数据通常具有很强的局部性,因为其值一般是连续变化的,邻近时刻的数值通常非常接近,而非突变。因此,直接采用自然语言处理领域中的点状注意力进行时间序列建模并不合适。更为合适的做法是使用段状注意力,即将时间序列分段后,以段为单位进行注意力计算,利用矩阵点积来计算两个段之间的相似度。NIE 等^[17]使用分 patch 的方式将时间序列划分为多个 patch,将每个 patch 视为一个 token 输入到 Transformer 中进行预测,显著提高了预测精度,并且降低了计算复杂度。YU 等^[18]提出在 patch 上应用自注意力建模会固地导致信息丢失,因为每个 patch 中仅封装了原始序列的一部分,这种不完整的表示可能会妨碍模型捕捉数据完整时间动态的能力。CHEN 等^[19]提出双重注意力机制,通过结合 patch 内注意力和 patch 间注意力,使得局部和全局时间依赖性都得到保留,从而缓解了基于 patch 的方法固有的信息丢失问题。因此,本文采用基于 patch 的 Transformer 注意力建模方式,并引入双重注意力机制来确保模型捕捉数据完整的时间依赖关系。

现实世界的时间序列数据具有多周期特性和复杂的动态变化,单一尺度的建模可能无法捕捉不同时间尺度上的特征,并且会限制模型的泛化能力^[20-21]。多尺度的 patch 建模能够同时处理短期和长期特征^[22],通过不同尺度的 patch 捕捉这些复杂的时间依赖性,通过多尺度的信息融合使得模型能够学习到更丰富和多样化的特征表示,进而提高预测的准确性^[23]。MO 等^[24]通过多尺度 patch 将时间序列转换成不同长度尺度的二维 patches,便于感知时间序列的局域性和长期相关性,通过整合来自不同尺度的信息获得更准确的预测。ZHANG 等^[25]提出一个多分辨率时间序列 Transformer 框架(MTST),通过构造多组不同 patch 大小的 token,使用 MTST 层对不同分辨率的时间模式进行多分支自注意力建模。然而,在这些方法中,通常选用一组预先设定好的 patch 大小,缺乏自适应学习能力,在面对不同周期性和波动性的时间序列时,可能无法有效捕捉所有相关特征。这需要对时间序列数据进行精确的动态分析以确定关键尺度。因此,本文提出一种动态多尺度建模方法,可以根据不同数据集和不同类型的时间序列数据,自适应地调整 patch 的大小和尺度。

基于上述分析,本文提出一种基于动态多尺度与双重注意力的短期电力负荷预测模型(MDAT)。该模型利用显著周期检测模块识别电力负荷数据中的主要周期,自适应地确定 patch 长度。在此基础上,使用多尺度 Transformer 模块的多个分支处理不同 patch 大小分割的时间序列数据,并通过双重注意力机制融合 patch 间和 patch 内的时间依赖性。然后,使用多分支聚合模块进行非线性融合得到综合的特征表示,最终通过堆叠多层 Transformer 模块得到负荷预测结果。本文的创新点与主要贡献如下:

1)提出一种显著周期检测方法。通过精确识别电力负荷序列的不同的模式和趋势,自适应地确定多尺度建模的关键尺度信息。

2)提出一种基于 Transformer 的动态多尺度建模方法。通过多个分支同时学习不同尺度的特征,并通过非线性地聚合不同尺度的特征,提高负荷预测精度。

3)将双重注意力机制引入基于 patch 的 Transformer 建模中,缓解了 patch 内部信息丢失的问题,并结合相对位置编码来更有效地捕捉序列内局部相互依赖的特征。

1 MDAT 整体架构

本文设计的 MDAT 模型的整体架构如图 1 所示。该模型由显著周期检测模块、多尺度 Transformer 模块、多分支聚合模块和线性预测头输出模块构成。

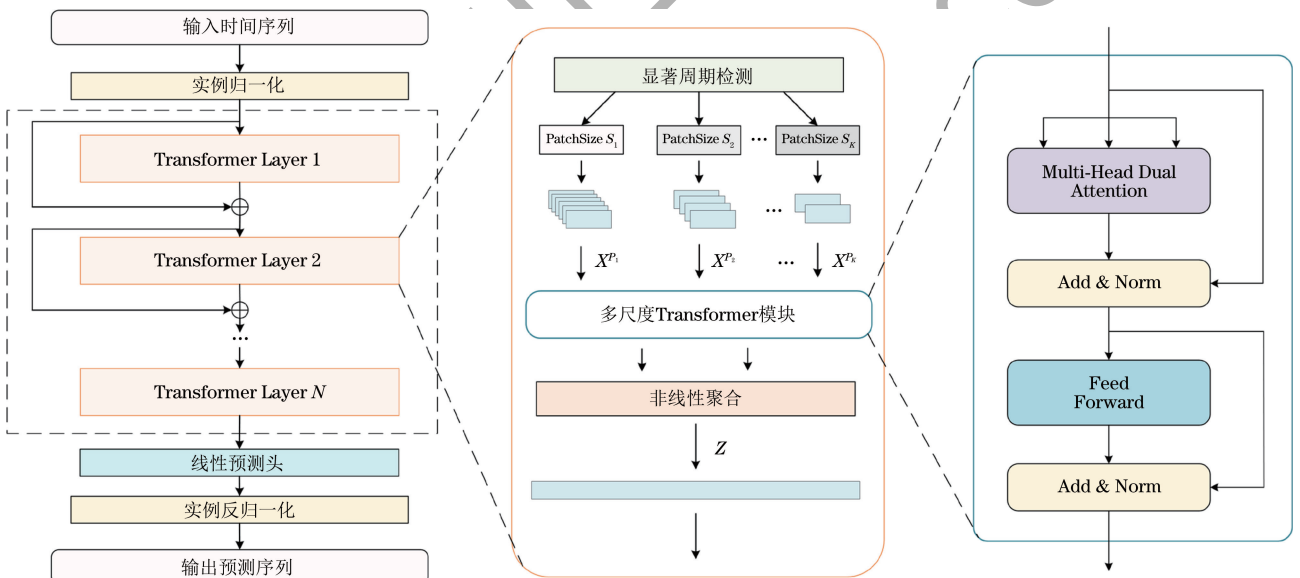


图 1 MDAT 模型的整体结构

Fig.1 Overall architecture of MDAT model

MDAT 模型的主干部分由 N 个 Transformer 层堆叠组成,用于挖掘深层多尺度特征。每层包含多个分支,每个分支通过多尺度 Transformer 模块处理基于不同大小 patch 分割的输入。通过 Transformer 中的双重注意力机制挖掘 patch 内部和 patch 之间的依赖关系,并将双重注意力的输出通过残差连接、归一化和前馈网络的组合进行处理。线性预测头^[26]利用全连接神经网络将输出线性变换映射到预测的目标窗口大小。通过多分支聚合模块来组合这些多尺度特征,将分支输出非线性融合在一起,形成一个单一的嵌入,并馈送到下一层。

2 MDAT 模型组成

2.1 显著周期检测模块

在显著周期检测模块中,使用逐次变分模态分

解(SVMD)^[27]可以有效地分离复杂周期结构信号的不同频率成分,减少模式混叠,提高周期检测的精确性。结合快速傅里叶变换(FFT)^[28],可以准确提取每个成分的主频率,并根据主频率计算出对应的周期,从而准确识别电力负荷数据中的显著周期变化。具体的显著周期检测流程如图 2 所示。

2.1.1 SVMD

SVMD 是一种高效的信号处理方法,它将复杂信号分解为特定频率的模态函数,无需预先设置模态函数的个数 K 。通过逐次迭代,SVMD 每次提取并剔除一个主要频率成分,直至所有频率成分被完全分解。这种方法能够更精确地捕捉和重构信号中的各个频率成分。

SVMD 的目标是将输入信号 $f(t)$ 分解成一系列本征模态函数(IMF) $u_L(t)$ 和残余信号 $f_r(t)$,

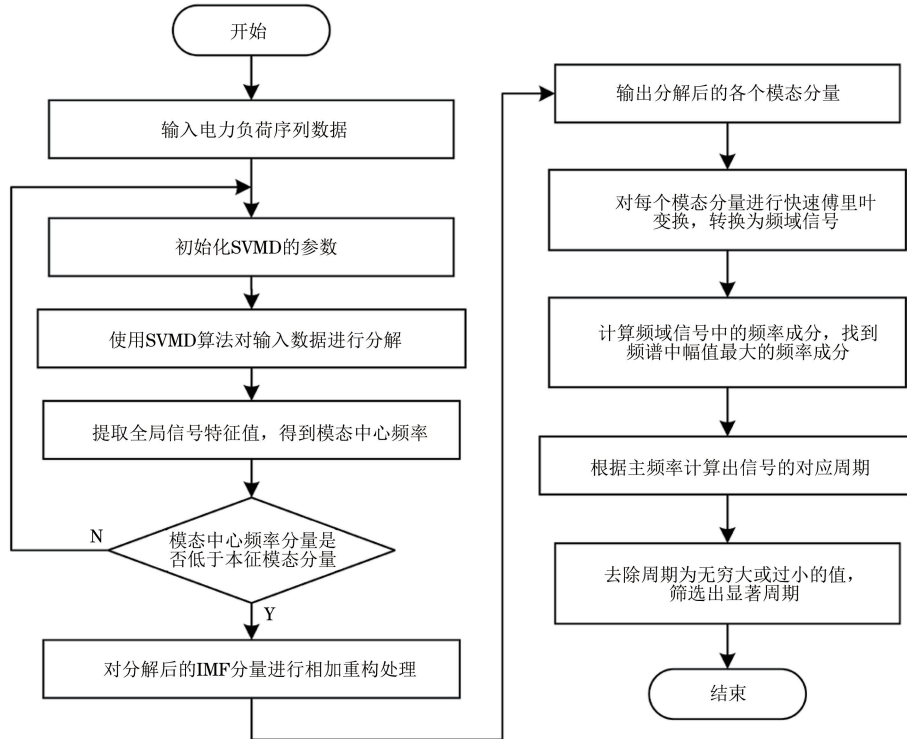


图 2 显著周期检测流程

Fig.2 Significant period detection procedure

计算公式如下:

$$f(t) = u_L(t) + f_r(t) \quad (1)$$

式中:残差信号 $f_r(t)$ 包含之前获得的模式和未处理部分的信号。 $f_r(t)$ 的计算公式如下:

$$f_r(t) = \sum_{i=1}^{L-1} u_i(t) + f_u(t) \quad (2)$$

为了找到第 L 个模式,SVMD 通过求解以下优化问题来找到具有最大紧致谱的信号 $u_L(t)$, 使得它与已提取模式之和的重建误差最小:

$$\begin{aligned} \min_{u_L, \omega_L, f_r} \{ \alpha J_1 + J_2 + J_3 \} \\ \text{s. t. } u_L(t) + f_r(t) = f(t) \end{aligned} \quad (3)$$

式中: α 是平衡参数; ω_L 为 L 阶模态的中心频率; J_1 、 J_2 和 J_3 为不同约束条件的代价函数。

代价函数 J_1 用于最小化每个模式 $u_L(t)$ 的带宽。计算公式如下:

$$J_1 = \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_L(t) \right] e^{-j\omega_L t} \right\|_2^2 \quad (4)$$

式中: ∂_t 表示对时间 t 的导数; $*$ 表示卷积运算; $\delta(t)$ 为狄拉克函数。

代价函数 J_2 用于最小化原始信号与模态之和的重构误差 $f_r(t)$ 。计算公式如下:

$$J_2 = \|\beta_L(t) * f_r(t)\|_2^2 \quad (5)$$

式中: $\beta_L(t)$ 是滤波器,其频率响应为 $\hat{\beta}_L(\omega) = 1/[\alpha(\omega - \omega_L)^2]$ 。

通过上述两个约束条件,无法辨别 L 阶模态与 $L-1$ 阶模态两种模态,可以利用 J_2 约束的思路,所用滤波器的频率响应为:

$$\hat{\beta}_i(\omega) = \frac{1}{\alpha(\omega - \omega_i)^2}, i = 1, 2, \dots, L-1 \quad (6)$$

因此,代价函数 J_3 所建立的约束为:

$$J_3 = \sum_{i=1}^{L-1} \|\beta_i(t) * u_L(t)\|_2^2 \quad (7)$$

为了改进收敛性并在噪声存在时保证重建的准确性,使用了拉格朗日乘子法和二次惩罚项的组合来建立增广拉格朗日函数:

$$\begin{aligned} \mathcal{L}(u_L, \omega_L, \lambda) = \alpha J_1 + J_2 + J_3 + \\ \|f(t) - R(t)\|_2^2 + \langle \lambda(t), f(t) - R(t) \rangle \end{aligned} \quad (8)$$

式中: $\lambda(t)$ 是拉格朗日乘子。 $R(t)$ 定义为:

$$R(t) = u_L(t) + f_r(t) + \sum_{i=1}^{L-1} u_i(t) \quad (9)$$

使用交替方向乘子法(ADMM)迭代求解该优化问题,迭代更新 u_L 、 ω_L 、 λ 。

通过以上步骤,SVMD 逐次提取信号中的模式,直到所有模式都被提取或重建误差低于预设阈值。

2.1.2 FFT

FFT 是一种高效计算离散傅里叶变换(DFT)及其逆变换的方法。FFT 通过分治法减少了计算量,能够快速计算信号在不同频率成分上的幅度和

相位,广泛应用于信号处理等领域。通过 FFT 提取显著周期的过程如下:

对 SVM D 分解后产生的每一个模态分量 $u_k(t)$ 和其离散采样值 $u_k[n]$, 应用 FFT 计算:

$$U_k[m] = \text{FFT}(u_k[n]) \quad (10)$$

式中: m 表示频率索引; 对应的频率为 f_m 。

找到频谱 $U_k[m]$ 的最大值对应的频率索引 $m_k = \arg\max_m |U_k[m]|$, 对应的主频率 $f_k = m_k \cdot f_s / N$, 其中, f_s 是采样频率, N 是 FFT 的点数。最后计算对应的周期 T_k :

$$T_k = \frac{1}{f_k} = \frac{N}{m_k \cdot f_s} \quad (11)$$

2.2 多尺度 Transformer 模块

MDAT 中的多尺度 Transformer 模块通过在不同分支中使用不同的 patch 片段大小, 允许对多个尺度的时间模式进行建模。较小的片段大小聚焦于短期、高频率的时间特征, 而较大的片段大小则有助于学习长期、低频率的季节性和趋势。通过同时学习不同尺度的时间序列表示, 可以有效捕捉时间序列数据中的短期和长期依赖关系。

2.2.1 输入表示

根据显著周期设置原始 patch 长度 $\{P_1, P_2, \dots, P_K\}$, 对于形状为 $B \times C \times L$ 的输入 X 采用不重叠的方式进行 patch 分割。如果 X 的长度不能被 P 整除, 则通过重复 X 的最后一个值来添加填充。经过 patch 操作后转化为 $B \times C \times P \times N$ 的四维张量, 其中, C 为通道数, $N = \lfloor L/P \rfloor$ 。通过将该张量重构为大小为 $(B \cdot C) \times P \times N$ 的三维张量, 避免不同通道的特征混合。每个 patch 大小 P 对应一个不同的分辨率的视图, 对 X 进行不同尺度的分割后生成一个集合 $\{X^{P_1}, X^{P_2}, \dots, X^{P_K}\}$, 然后以多分支并行的方式处理不同尺度的时间序列表示。

2.2.2 相对位置编码

相对位置编码 (RPE)^[25] 通过引入相对位置信息, 能够更好地捕捉序列中元素之间的相对距离和相对顺序。相较于绝对位置编码, 它可以更好地处理长序列并捕捉元素之间的相对关系。具体过程如下:

引入相对位置矩阵 R , 并对每个相对位置 $r_{ij} = j - i$ 进行编码。使用正弦和余弦函数生成位置编码向量:

$$P_{PE}(i, 2j) = \sin\left(\frac{i}{10000^{2j/d_{\text{model}}}}\right) \quad (12)$$

$$P_{PE}(i, 2j+1) = \cos\left(\frac{i}{10000^{2j/d_{\text{model}}}}\right) \quad (13)$$

式中: $j \in \{1, 2, \dots, d_{\text{model}}/2\}$, d_{model} 是嵌入维度。

加入相对位置编码的影响, 修正注意力得分:

$$a_{ij} = \frac{(q_i \cdot k_j) + (q_i \cdot e_{r_{ij}})}{\sqrt{d_k}} \quad (14)$$

式中: $e_{r_{ij}}$ 是相对位置 r_{ij} 的编码向量。

对所有位置对 (i, j) 计算修正后的注意力得分矩阵为:

$$A = \text{Softmax}\left(\frac{QK^T + QE_{\text{rel}}}{\sqrt{d_k}}\right) \quad (15)$$

通过将相对位置编码的值和索引用于注意力得分的修正, 以增强模型对序列中元素相对位置的感知能力。

2.2.3 双重注意力机制

双重注意力机制包括每个被划分的 patch 的 patch 内注意力和不同 patch 间的 patch 间注意力。通过整合全局和局部信息, 这种机制避免了 patch 内部信息丢失的问题。双重注意力的机制如图 3 所示。

patch 内注意力用于捕捉每个补丁内的时间步长之间的关系。其计算过程如下:

对于输入的序列 X 的第 i 个 patch 嵌入 $X_i^{\text{intra}} \in \mathbb{R}^{B \times S \times C \times d}$, 其中 S 为 patch 的大小, 利用一个可训练的查询矩阵 $Q_i^{\text{intra}} \in \mathbb{R}^{B \times 1 \times C \times d}$ 来融合 patch 的上下文, 并计算局部细节的交叉注意力:

$$A_{\text{Attn}_i^{\text{intra}}} = \text{Softmax}\left(\frac{Q_i^{\text{intra}}(K_i^{\text{intra}})^T}{\sqrt{d_m}}\right) V_i^{\text{intra}} \quad (16)$$

式中: $K_i^{\text{intra}}, V_i^{\text{intra}}$ 为 X_i^{intra} 经过线性变换得到的矩阵。

然后, 将所有 patch 的 patch 内注意力结果进行拼接, 生成最终的内部注意力输出:

$$A_{\text{Attn}^{\text{intra}}} = \text{Concat}(A_{\text{Attn}_1^{\text{intra}}}, A_{\text{Attn}_2^{\text{intra}}}, \dots, A_{\text{Attn}_P^{\text{intra}}}) \quad (17)$$

式中: $A_{\text{Attn}_i^{\text{intra}}} \in \mathbb{R}^{B \times N \times C \times d}$; N 为 patch 的数量。

patch 间注意力用于捕捉不同 patch 之间的关系, 以捕获全局关联。其计算过程如下:

对于经过输入处理后的序列 X , 将 patch 大小 P 和特征嵌入维度 d 结合起来, 进行新的特征投影得到 $X_i^{\text{inter}} \in \mathbb{R}^{(B \times C) \times N \times d'}$, 其中 $d'_m = P \cdot d$ 。然后, 计算自注意力, 表示时间序列的全局相关性:

$$A_{\text{Attn}^{\text{inter}}} = \text{Softmax}\left(\frac{Q^{\text{inter}}(K^{\text{inter}})^T}{\sqrt{d'_m}}\right) V^{\text{inter}} \quad (18)$$

式中: $Q^{\text{inter}}, K^{\text{inter}}, V^{\text{inter}}$ 为 X^{inter} 经过线性变换得到的矩阵。为了融合通过双重注意力机制捕捉到的全局相关性和局部细节, 首先对 patch 内注意力的输出 $A_{\text{Attn}_i^{\text{intra}}} \in \mathbb{R}^{B \times N \times C \times d}$ 进行线性变换以匹配 patch 间注意力的输出形状 $A_{\text{Attn}_i^{\text{intra}}} \in \mathbb{R}^{B \times L \times C \times d}$, 其中 L 为

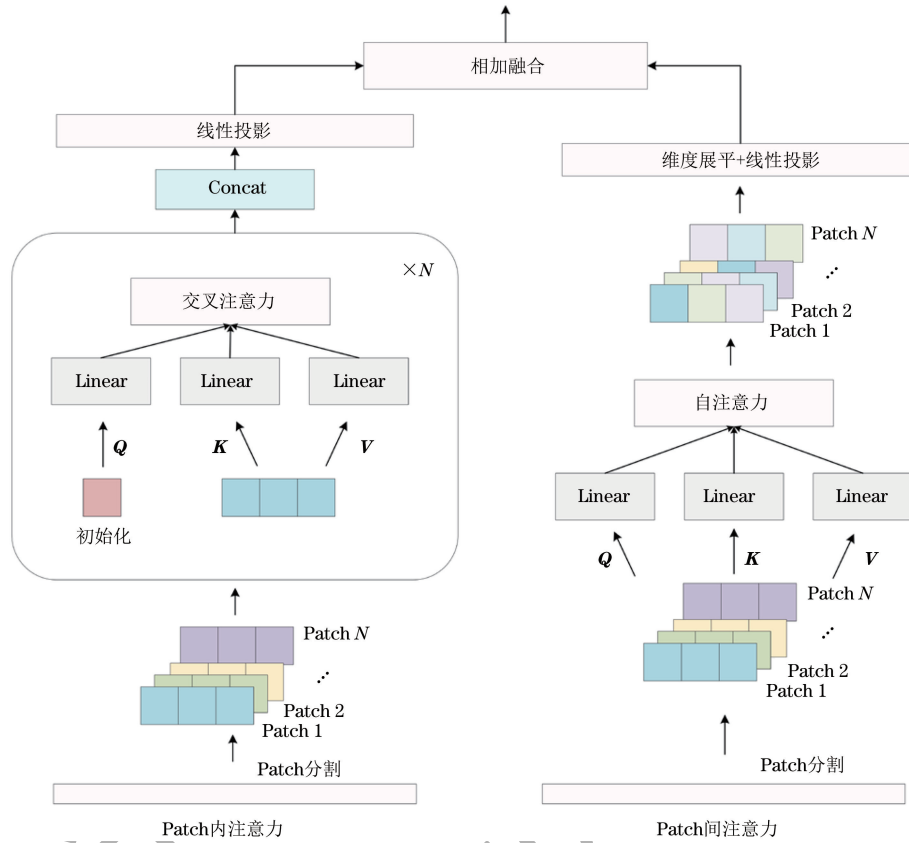


图 3 双重注意力机制

Fig.3 Dual attention mechanism

序列的长度。通过线性变换和加和操作,得到双重注意力的最终输出:

$$A_{\text{Attn}} = \text{Linear}(A_{\text{Attn}^{\text{intra}}}) + A_{\text{Attn}^{\text{inter}}} \quad (19)$$

2.3 多分支聚合模块

经过多尺度 Transformer 模块每一个分支的输出 $Z_i \in \mathbb{R}^{B \times C \times L \times d}$, 将它们聚合并执行加权和以产生融合后的特征表示。权重基于相应周期的归一化幅度:

$$\{w_1, w_2, \dots, w_K\} = \text{Softmax}(\{P_1, P_2, \dots, P_K\}) \quad (20)$$

经过多层 Transformer 层堆叠后,通过平坦化后两个维度形成新的向量 $\text{Flatten}(Z_i) \in \mathbb{R}^{B \times C \times (L \times d)}$, 对每一个通道 C 应用线性变换映射到预测的目标窗口大小 H , 得到最终输出的表示 $\tilde{Z}_i \in \mathbb{R}^{B \times C \times H}$ 。

2.4 损失函数

均方误差(MSE)损失函数对异常值高度敏感,导致模型容易对这些异常值过度拟合,从而降低整体泛化能力。本文引入自适应损失函数^[29],通过鲁棒性参数 α , 能够自适应调整对异常值的敏感度,提高模型的鲁棒性和泛化性能,并且无需手动调参。广义损失函数公式如下:

$$f(z, \alpha, m) = \frac{|\alpha - 2|}{\alpha} \left(\left(\frac{(z/m)^2}{|\alpha - 2|} + 1 \right)^{\alpha/2} - 1 \right) \quad (21)$$

式中: z 是真实值与预测值之间的差值; m 是控制损失函数在 $x=0$ 附近的二次函数曲率的比例因子; α 是控制损失函数鲁棒性的形状参数,在训练过程中以端到端方式学习。如图 4 所示,自适应损失函数形状随 α 的变化而变化。

3 实验结果与分析

3.1 实验数据及平台

为验证本文模型的性能,本文采用了两个真实的电力负荷数据集进行实验,分别是 Australia 电价与电力负荷数据集^[30]以及 Morocco 得土安市电力负荷数据集^[31]。这两个数据集的相关统计信息如表 1 所示。

表 1 电力负荷数据集详细信息
Table 1 Electricity load dataset details

数据集	采样总数/个	变量数/个	采样间隔/min
Australia	35 040	6	30
Morocco	52 416	6	10

Australia 数据集选取 24 个月的干球温度、露点湿度、湿球温度、湿度、电价信息和电力负荷数据,

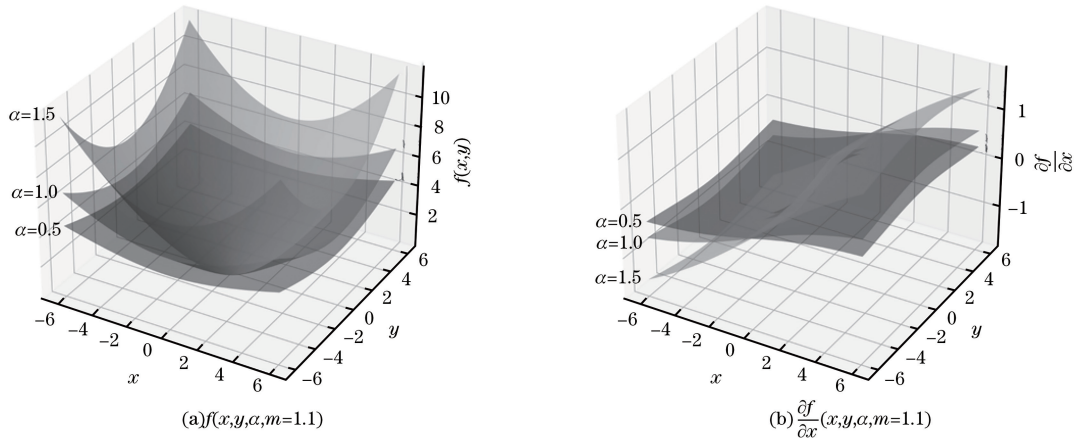


图 4 自适应损失函数

Fig. 4 Adaptive loss function

时间范围为 2009 年 1 月 1 日—2010 年 12 月 31 日。Morocco 数据集选取 12 个月的温度、湿度、风速、扩散流和区域 3 的电力负荷数据,时间范围为 2017 年 1 月 1 日—2017 年 12 月 30 日。所有实验均使用 PyTorch 进行,并在 NVIDIA RTX 3080 10 GB GPU 上运行。

3.2 实验设置和评价指标

在 Australia 数据集上,预测长度 H 为 336 个时间步,在 Morocco 数据集上预测长度 H 为 720 个时间步。回看窗口 L 都是 96 个时间步。通过步长为 1 的时间窗口 L 逐步移动整个序列,创建多个子序列以训练和预测。其他具体实验参数见表 2。

表 2 预测模型超参数设置

Table 2 Prediction model hyperparameter settings

超参数	参数值
Encoder 层数	2
Decoder 层数	1
Batch size	32
Num of heads	4
Dimension of model	516
Learning rate	0.000 5
Dropout	0.1
Epochs	10
Activation function	GELU
Loss function	MSE
Optimizer	Adam

为衡量模型的预测性能,本文采用平均绝对误差(MAE, E_{MAE})、均方根误差(RMSE, E_{RMSE})及平均绝对百分比误差(MAPE, E_{MAPE})作为评价指标,对预测结果进行误差分析。其计算公式如下:

$$E_{MAE} = \frac{1}{T} \sum_{t=1}^T |A_t - F_t| \quad (22)$$

$$E_{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (A_t - F_t)^2} \quad (23)$$

$$E_{MAPE} = \frac{1}{T} \sum_{t=1}^T \left| \frac{A_t - F_t}{A_t} \right| \times 100\% \quad (24)$$

式中: A_t 表示时间点 t 的实际值; F_t 表示时间点 t 的预测值; T 表示总的时间点数。

3.3 数据预处理

经过 SVMDF-FFT 进行显著周期检测后, Australia 数据集上的周期为 $\{48, 24, 16\}$, Morocco 数据集上的周期为 $\{144, 48, 18\}$, 单位为时间步。图 5 展示了 Australia 数据集经 SVMDF 分解的部分结果及其频谱图,从图 5 中可以看出,每个分量在频谱图中都显示出单一显著的频率峰值,对应负荷序列的一个周期特性。

对输入序列进行 Z-score 标准化处理,使数据具有均值为 0、标准差为 1 的分布,从而帮助模型更快地收敛并提高预测的稳定性。

$$X'_i = \frac{X_i - \mu_i}{\sqrt{\frac{1}{T} \sum_{t=1}^T (X_{it} - \mu_i)^2 + \epsilon}} \quad (25)$$

式中: X 为输入张量; $\mu_i = \frac{1}{T} \sum_{t=1}^T X_{it}$ 为时间步维度上的均值; ϵ 是一个很小的值防止除以 0。标准化后的数据按照 7:2:1 的比例划分为训练集、测试集和验证集。

3.4 结果分析

3.4.1 消融实验分析

在 Australia 数据集上进行消融研究,以测量多尺度、双重注意力、相对位置编码和自适应损失函数对 MDAT 预测精度的影响。表 3 列出了组件消融的实验结果,整体上最好的方法以粗体显示。从结果可以看出,当去掉不同的模型组件时,模型性能出

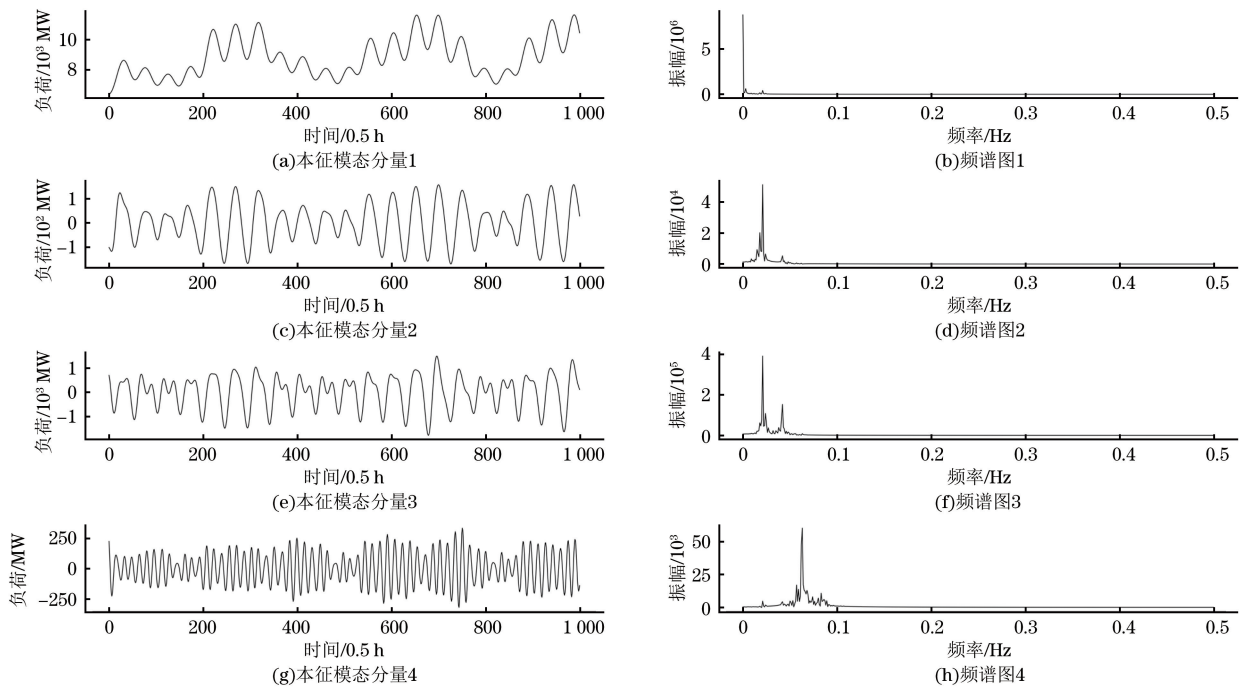


图 5 SVM 算法分解结果及其频谱图

Fig.5 SVM algorithm decomposition results and their spectrograms

现了不同程度的下降: MAE 增大了 2.29%~7.14%, RMSE 增大了 2.26%~5.35%, MAPE 增大了 2.33%~8.13%。其中,多尺度组件对模型的性能提升最为显著,忽略它会导致模型的 MAE、RMSE 和 MAPE 分别增大 7.14%、5.35% 和 8.13%。这表明捕捉数据中不同时间尺度上的特征信息,可以有效提高预测精度。其次是双重注意力,这说明融合 patch 内和 patch 间的注意力信息可以进一步提升序列建模能力。然后是相对位置编码和鲁棒损失函数,这表明捕捉多周期的复杂数据中元素的相对关系的重要性,并且自适应地调整损失函数的形状可以更有效地学习到真实的数据波动。完整的 MDAT 模型在 3 个指标上均表现最好。具体来说,MDAT 模型的 MAE 为 0.35 GW, RMSE 为

0.486 GW, MAPE 为 4.121%。表明了各组件的协同对于模型的整体性能提升的作用。

表 3 组件消融实验

Table 3 Component ablation experiment

模型组件	MAE/GW	RMSE/GW	MAPE/%
MDAT	0.350	0.486	4.121
无多尺度	0.375	0.512	4.456
无双重注意力	0.365	0.502	4.314
无相对位置编码	0.362	0.499	4.259
无自适应损失函数	0.358	0.497	4.217

图 6 显示了在具有强周期性模式的合成数据集上,通过绝对位置编码和相对位置编码得到的注意力矩阵热力图。在图 6(a)中,绝对位置编码的注意力矩阵热力图显示出较为分散的注意力权重分布,

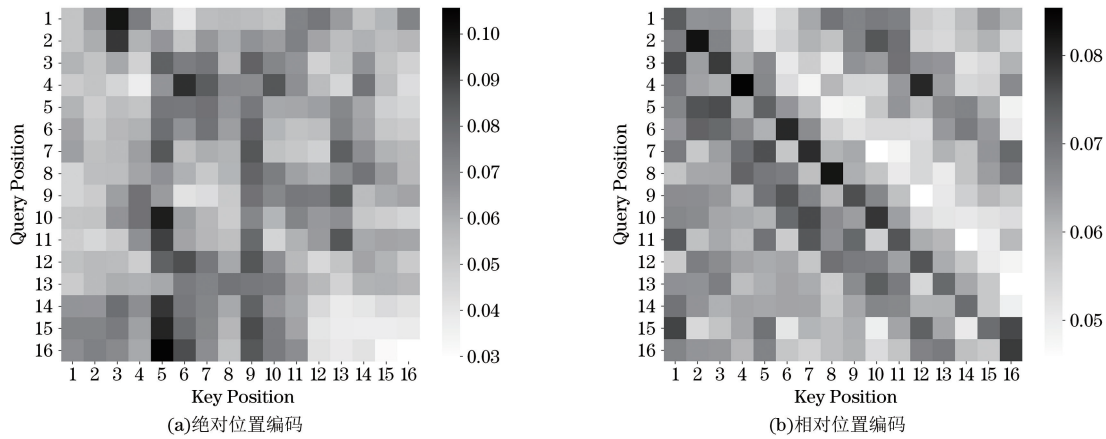


图 6 绝对和相对位置编码的注意力得分热图

Fig.6 Heat maps of attention scores encoded by absolute and relative positions

某些区域的权重较高。这可能表明模型在没有相对位置编码的情况下,更难以捕捉到序列数据中的局部相互依赖关系,而是较为均匀地关注整个序列。在图 6(b)中,相对位置编码的注意力矩阵热力图显示出某些对角线或相邻位置的权重较高,表明模型在关注 Query 和 Key 的相对位置时,更倾向于关注与自身位置接近的 Key。这种模式有助于捕捉序列数据中局部相互依赖的特征。

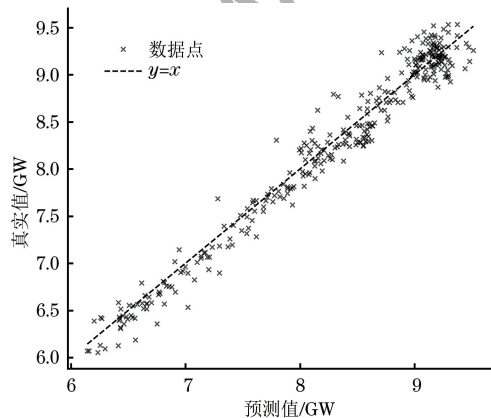
为了验证所提出的基于 SVM-D-FFT 的周期检测的有效性,选取快速傅里叶变换(FFT)、小波变换(WT)、自相关(ACF)和功率谱密度(PSD)对负荷序列的周期分析进行对比,提取前 5 个峰值频率对应的周期,具体结果见表 4。并且将周期检测结果应用于 MDAT 模型,得到的不同检测方法的对比结果如表 5 所示。

表 4 周期检测结果

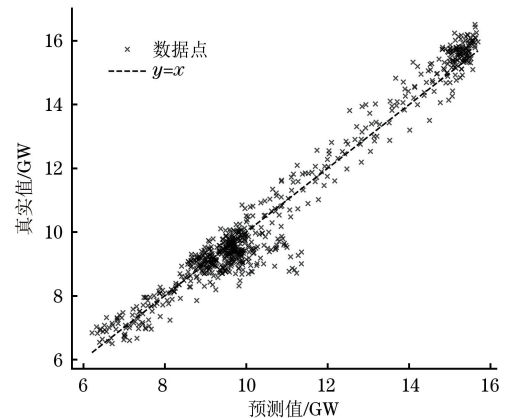
Table 4 Periodic detection results

周期检测方法	显著周期/0.5 h
SVMD-FFT	48, 24, 16
FFT	48, 24
WT	190, 152, 162, 26, 22
ACF	336, 192, 144, 96, 48
PSD	48, 24

从表 5 可以看出,通过不同检测方法获得的周期作为多尺度 Transformer 模块的 patch 尺寸进行预测,SVMD-FFT 取得了最好的效果。这说明 SVMD 可以有效地分离复杂多周期结构负荷数据的不同频率成分,并且结合 FFT 快速识别到的显著周期,可以实现更有效的多尺度建模。



(a)Australia数据集7天的预测数据



(b)Morocco数据集5天的预测数据

图 7 MDAT 预测结果散点图

Fig.7 Scatter plot of MDAT prediction results

在 Australia 数据集上的各模型预测误差定量评价结果对比如表 6 所示,其中,最优数据用加粗表示,次优数据用下划线表示。

表 5 不同检测结果的预测对比
Table 5 Comparison of predictions with different detection results

周期检测方法	MAE/GW	RMSE/GW	MAPE/%
SVMD-FFT	0.350	0.486	4.121
FFT	0.356	0.492	4.206
WT	0.375	0.512	4.456
ACF	0.351	0.488	4.103
PSD	0.356	0.492	4.206

3.4.2 对比实验分析

为评估本文所提出模型的预测效果,本文选择了 5 种广泛应用于时间序列预测的模型,从不同的角度进行对比分析。在对比模型中,PatchTST^[17]通过基于单尺度 patch 和 Transformer 架构进行预测。MTST^[25]则由多分支架构组成,每个分支处理不同 patch 尺度的输入,用于在不同尺度下同时建模不同的时间模式。Pyraformer^[32]模型基于金字塔注意力的 Transformer,在不同尺度上对不同范围的时间依赖性进行建模。FEDformer^[33]是基于周期趋势项分解和频域注意力的 Transformer。LightTS^[34]则是使用轻量采样的 MLP 结构网络进行预测。

为直观地展示 MDAT 模型的准确性,本文使用散点图图 7 可视化 Australia 数据集和 Morocco 数据集上的预测结果。回归线 $y = x$ 表示预测值与真实值相等的地方。通过观察图 7 可知,绝大多数点聚集在回归线附近,显示出不错的预测准确性。

由表 6 可知:基于 patch 的模型 PatchTST 和 MTST 取得了不错的预测效果,其中,MTST 采用了多尺度 patch 建模的预测方式,相较于单一尺度的

PachTST 预测效果更好。Pyraformer 是基于多尺度特征金字塔进行建模,相较于 FEDformer 和 LightTS,也取得了较好的预测效果。相较于其他模型,MDAT 模型在所有指标上均表现最佳,MAE 为 0.35 GW, RMSE 为 0.486 GW, MAPE 为 4.121%, 其中, MAE 下降了 10.26%~17.06%, RMSE 下降了 7.25%~12.12%, MAPE 下降了 10.74%~18.40%。充分证明了本文所提出模型的有效性。

表 6 不同模型在 Australia 数据集上的预测效果对比

Table 6 Comparison of prediction performance of different models on the Australia dataset

模型	MAE/GW	RMSE/GW	MAPE/%
MDAT	0.350	0.486	4.121
PatchTST	0.393	0.540	4.642
MTST	<u>0.390</u>	<u>0.524</u>	<u>4.617</u>
Pyraformer	0.405	0.550	4.799
FEDformer	0.422	0.552	4.985
LightTS	0.421	0.553	5.050

图 8 是基于这 3 个误差指标建立的气泡图,图中的横轴表示 MAE,纵轴表示 RMSE,气泡的颜色深度代表 MAPE。从图 8 中可以看出:MDAT 模型在各个指标均表现出了最优的预测效果。

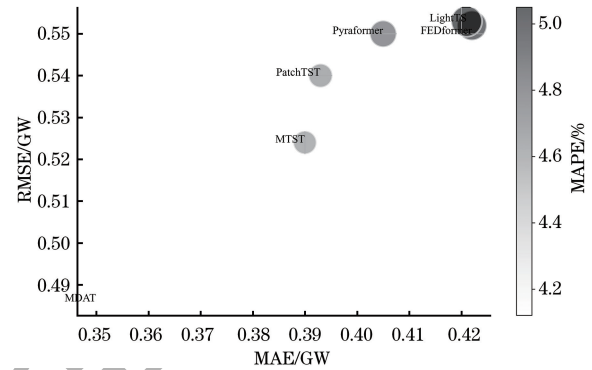


图 8 预测性能对比气泡图

Fig. 8 Prediction performance comparison bubble chart

图 9 所示为各模型的 7 天的短期负荷预测值以及真实值,并将局部位置的值放大,以便对预测结果进行深入分析。由图 9 可以看出:不同模型的预测结果在整体趋势上基本一致,都能较好地跟随实际负荷的波动。但在负荷急剧变化的区域, FEDformer 和 LightTS 模型呈现出较大的偏差,具体体现在峰值和波谷的幅度与真实值拟合较差。而 PatchTST 和 MTST 取得了较好的拟合效果,但在负载的上升和下降期间会出现一些滞后或超前。相比之下,本文中 MDAT 模型能更准确地捕捉负荷波动规律,从而提高预测精度。

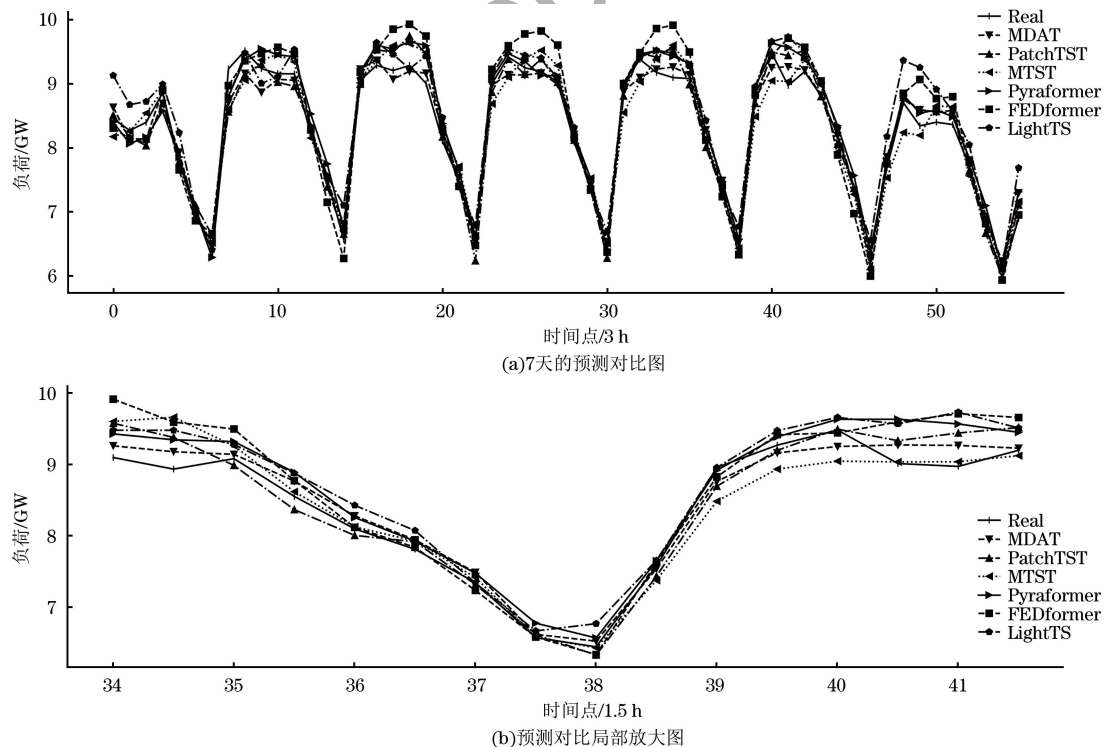


图 9 不同模型在 Australia 数据集上的预测曲线对比

Fig. 9 Comparison of prediction curves of different models on the Australia dataset

在 Morocco 数据集上的各模型预测误差定量评价结果对比如表 7 所示,其中,最优数据用加粗表示,次优数据用下划线表示。

由表 7 可知:预测步长翻倍后,各模型的预测误差均有较大提高,但 MDAT 仍表现出了最佳的预测效果,其 MAE、RMSE 和 MAPE 分别为

1.031 MW、1.678 MW 和 8.617%，均为最低值，其中，MAE 下降了 9.08%~70.25%，RMSE 下降了 2.61%~55.01%，MAPE 下降了 10.89%~73.88%。验证了本文模型的稳定性。在对比模型中，PatchTST 和 MTST 的表现次之，MTST 略优于 PatchTST。Pyraformer 和 LightTS 的误差最高，而 FEDformer 的性能介于上述两类之间。体现出了基于 patch 的时间序列建模的优越性和良好的泛化能力。

表 7 不同模型在 Morocco 数据集上的预测效果对比

Table 7 Comparison of prediction effects of different models on the Morocco dataset

模型	MAE/MW	RMSE/MW	MAPE/%
MDAT	1.031	1.678	8.617
PatchTST	1.237	1.813	10.588
MTST	1.134	1.723	9.670
Pyraformer	3.392	3.730	32.986
FEDformer	1.952	2.476	16.956
LightTS	3.466	3.715	31.705

3.4.3 复杂度分析

作为一个基于 Transformer 的模型，MDAT 在第 n 层运算的复杂度主要与参与注意力机制计算的 token 数有关，在双重注意力中，假设序列长度为 L ，每个 patch 包含 S 个 token，那么一共有 $N=L/S$ 个 patch，对于 patch 内注意力，因为每个 patch 内有 S 个 token，两两之间需要计算注意力分数。总共有 N 个 patch，因此 patch 内注意力的总复杂度为： $O(N \times S^2)=O((L/S) \times S^2)=O(L \times S)$ 。patch 间注意力的计算复杂度取决于 patch 之间的注意力计算，因为每个 patch 的嵌入可以看作是一个 token，所以 patch 间注意力的计算复杂度为： $O(N^2)=O((L/S)^2)=O(L^2/S^2)$ 。因此，双重注意力总计算复杂度是 $O(L \times S)+O(L^2/S^2)$ ，与 patch 大小 S 相关，由于实际选择时 $S < L$ ，因此其比标准的点积运算的复杂度 $O(L^2)$ 要低，由于多尺度 Transformer 模块采用多分支并行的方式计算，模型 MDAT 最终的计算复杂度取决于复杂度最高的分支。

4 结束语

本文提出一种基于动态多尺度与双重注意力的短期电力负荷预测模型 MDAT。针对电力负荷数据的多周期特性，设计了显著周期检测模块，通过 SVM 准确分离复杂周期结构信号的不同频率成分，并结合 FFT 准确识别电力负荷数据中的显著周

期。针对单一或固定尺度的模型无法有效捕捉时间序列复杂动态变化的问题，提出多尺度 Transformer 模块，根据识别到的周期利用 patch 建模不同尺度的时间序列表示，同时结合双重注意力机制来融合 patch 内和 patch 间的注意力信息。通过非线性聚合每个尺度的特征，从而获得更丰富的特征表示。实验结果表明，所提出的方法与当前主流的时间序列预测方法相比具有更高的精度。下一步将考虑在训练中生成自适应权重，根据权值的大小对分支进行开启和关闭，以降低计算开销。

参考文献

[1] EREN Y, KÜÇÜKDEMİR İ. A comprehensive review on deep learning approaches for short-term load forecasting[J]. Renewable and Sustainable Energy Reviews, 2024, 189: 114031.

[2] WAZIRALI R, YAGHOUBI E, ABUJAZAR M S S, et al. State-of-the-art review on energy and load forecasting in microgrids using artificial neural networks, machine learning, and deep learning techniques[J]. Electric Power Systems Research, 2023, 225: 109792.

[3] TARMANINI C, SARMA N, GEZEGIN C, et al. Short term load forecasting based on ARIMA and ANN approaches[J]. Energy Reports, 2023, 9: 550-557.

[4] SMYL S, DUDEK G, PELKA P. Contextually enhanced ES-dRNN with dynamic attention for short-term load forecasting[J]. Neural Networks, 2024, 169: 660-672.

[5] SAYED H A, WILLIAM A, SAID A M. Smart electricity meter load prediction in Dubai using MLR, ANN, RF, and ARIMA[J]. Electronics, 2023, 12(2): 389.

[6] AN W H, GAO B, LIU J H, et al. Predicting hourly heating load in residential buildings using a hybrid SSA-CNN-SVM approach[J]. Case Studies in Thermal Engineering, 2024, 59: 104516.

[7] WANG F, CHEN C, ZHANG H, et al. Short-term load forecasting based on variational mode decomposition and chaotic grey wolf optimization improved random forest algorithm[J]. Journal of Applied Science and Engineering, 2022, 26(1): 69-78.

[8] INDIRA G, BHAVANI M, BRINDA R, et al. Electricity load demand prediction for microgrid energy management system using hybrid adaptive barnacle-mating optimizer with artificial neural network algorithm[J]. Energy Technology, 2024, 12(5): 2301091.

[9] LU Y T, WANG G C, HUANG X F, et al. Probabilistic load forecasting based on quantile regression parallel CNN and BiGRU networks[J]. Applied Intelligence, 2024, 54(15): 7439-7460.

[10] 陆心怡, 关艳, 高曦莹, 等. 面向工业用户的混合 DWT-DE-RNN 电力负荷预测[J]. 机械设计与制造, 2024, 404(10): 73-78.

LUX Y, GUAN Y, GAO X Y, et al. Hybrid DWT-DE-RNN power load forecasting for industrial users[J]. Machinery Design & Manufacture, 2024, 404(10): 73-78. (in Chinese)

[11] SMYL S, DUDEK G, PELKA P. ES-dRNN: a hybrid exponential smoothing and dilated recurrent neural network model for short-term load forecasting[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(8): 11346-11358.

- [12] BARETH R, YADAV A, GUPTA S, et al. Daily average load demand forecasting using LSTM model based on historical load trends[J]. IET Generation, Transmission & Distribution, 2024, 18(5): 952-962.
- [13] 李丽芬, 张近月, 曹旺斌, 等. 基于双重注意力时间卷积长短期记忆网络的短期负荷预测[J]. 系统仿真学报, 2024: 1-12[2024-06-29]. <https://doi.org/10.16182/j.issn1004731x.joss.24-0282>.
- LI L F, ZHANG J Y, CAO W B, et al. Short-term load prediction based on dual-attention temporal convolutional long short-term memory network[J/OL]. Journal of System Simulation, 2024: 1-12[2024-06-29]. <https://doi.org/10.16182/j.issn1004731x.joss.24-0282>. (in Chinese)
- [14] 李新, 张旭, 余乐安, 等. 基于改进 Transformer 模型的景区短时客流预测研究[J]. 中国管理科学, 2005(2): 13-20.
- LI X, ZHANG X, YU A L, et al. Enhancing short-term tourist flow forecasting and evaluation using an improved Transformer framework[J]. Chinese Journal of Management Science, 2005(2): 13-20. (in Chinese)
- [15] 李浩阳, 贺小伟, 王宾, 等. 基于改进 Informer 的云计算资源负载预测[J]. 计算机工程, 2024, 50(2): 43-50.
- LI H Y, HE X W, WANG B, et al. Cloud computing resource load prediction based on improved informer[J]. Computer Engineering, 2024, 50(2): 43-50. (in Chinese)
- [16] DU D Z, SU B, WEI Z W. Preformer: predictive transformer with multi-scale segment-wise correlations for long-term time series forecasting[C]//Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. Washington D. C., USA: IEEE Press, 2023: 1-5.
- [17] NIE Y, NGUYEN N H, SINTHONG P, et al. A time series is worth 64 words: long-term forecasting with Transformers[C]//Proceedings of the 12th International Conference on Learning Representations. Washington D. C., USA: IEEE Press, 2021: 157-168.
- [18] YU G, ZOU J, HU X, et al. Revitalizing multivariate time series forecasting: learnable decomposition with inter-series dependencies and intra-series variations modeling[C]//Proceedings of the 41st International Conference on Machine Learning. Washington D. C., USA: IEEE Press, 2020: 457-468.
- [19] CHEN P, ZHANG Y Y, CHENG Y Y, et al. Pathformer: multi-scale transformers with adaptive pathways for time series forecasting[EB/OL]. [2024-06-01]. <https://arxiv.org/abs/2402.05956v5>.
- [20] 付文龙, 章轩瑞, 张海荣, 等. 多尺度特征提取与非线性融合的综合能源系统多元负荷短期预测[J]. 电力系统及其自动化学报, 2023, 35(12): 89-99.
- FU W L, ZHANG X R, ZHANG H R, et al. Short-term multivariate load forecasting of integrated energy system based on multiscale feature extraction and nonlinear fusion[J]. Proceedings of the CSU-EPSA, 2023, 35(12): 89-99. (in Chinese)
- [21] 孟衡, 张涛, 王金, 等. 基于多尺度时空图卷积网络与 Transformer 融合的多节点短期电力负荷预测方法[J]. 电网技术, 2024, 48(10): 4297-4305.
- MENG H, ZHANG T, WANG J, et al. Multi-node short-term power load forecasting method based on multi-scale spatiotemporal graph convolution network and Transformer[J]. Power System Technology, 2024, 48(10): 4297-4305. (in Chinese)
- [22] 张鹏飞, 胡博, 胡展硕, 等. 基于 STD-ST-Former 的现货电价长步时空预测[J]. 中国电机工程学报, 2024, 44(7): 2732-2743.
- ZHANG P F, HU B, LUO Z S, et al. Long step spatial-temporal prediction of spot electricity price using dual channel ST-Former framework based on seasonal trend decomposition[J]. Proceedings of the CSEE, 2024, 44(7): 2732-2743. (in Chinese)
- [23] DU L F, XIN J, LABACH A, et al. MultiResFormer: transformer with adaptive multi-resolution modeling for general time series forecasting[EB/OL]. [2024-06-01]. <https://arxiv.org/abs/2311.18780v2>.
- [24] MO S T, WANG H X, LI B X, et al. TimeSQL: improving multivariate time series forecasting with multi-scale patching and smooth quadratic loss[J]. Information Sciences, 2024, 671: 120652.
- [25] ZHANG Y T, MA L H, PAL S, et al. Multi-resolution time-series transformer for long-term forecasting[EB/OL]. [2024-06-01]. <https://arxiv.org/abs/2311.04147v2>.
- [26] ZENG A L, CHEN M X, ZHANG L, et al. Are transformers effective for time series forecasting? [J]. Artificial Intelligence, 2023, 37(9): 11121-11128.
- [27] NAZARI M, SAKHAEI S M. Successive variational mode decomposition[J]. Signal Processing, 2020, 174: 107610.
- [28] LUO X Y, WANG H, HAN T, et al. FFT-trans: enhancing robustness in mechanical fault diagnosis with Fourier transform-based transformer under noisy conditions[J]. IEEE Transactions on Instrumentation Measurement, 2024, 73: 3381688.
- [29] BARRON J T. A general and adaptive robust loss function[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2019: 4331-4339.
- [30] ZHANG J L, ZHANG Y J, LI D Z, et al. Forecasting day-ahead electricity prices using a new integrated model[J]. International Journal of Electrical Power & Energy Systems, 2019, 105: 541-548.
- [31] SALAM A, EL HIBAOU A. Comparison of machine learning algorithms for the power consumption prediction[C]//Proceedings of the 6th International Renewable and Sustainable Energy Conference. Washington D. C., USA: IEEE Press, 2018: 1-5.
- [32] LIU S, YU H, LIAO C, et al. Pyraformer: low-complexity pyramidal attention for long-range time series modeling and forecasting[C]//Proceedings of International Conference on Learning Representations. Washington D. C., USA: IEEE Press, 2021: 256-267.
- [33] ZHOU T, MA Z Q, WEN Q S, et al. FEDformer: frequency enhanced decomposed transformer for long-term series forecasting[EB/OL]. [2024-06-01]. <https://arxiv.org/abs/2201.12740v3>.
- [34] ZHANG T P, ZHANG Y Z, CAO W, et al. Less is more: fast multivariate time series forecasting with light sampling-oriented MLP structures[EB/OL]. [2024-06-01]. <https://arxiv.org/abs/2207.01186v1>.

编辑 索志书