

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

范士雄, 李东琦, 郭剑波, 王铁柱, 马士聪, 赵泽宁

(中国电力科学研究院有限公司, 北京市 海淀区 100192)

Ultra-short-term Load Forecasting Method Based on Time-varying Filter Empirical Mode Decomposition-reconstruction and iTransformer With Stand-alone Self-attention Mechanism

FAN Shixiong, LI Dongqi, GUO Jianbo, WANG Tiezhu, MA Shicong, ZHAO Zening

(China Electric Power Research Institute, Haidian District, Beijing 100192, China)

ABSTRACT: Accurate load forecasting is crucial for power systems' safe and stable operation. To further improve the precision of load forecasting, an ultra-short-term load forecasting method based on time-varying filter empirical mode decomposition (TVF-EMD)-reconstruction and transformer with stand-alone self-attention (SASA) mechanism are proposed by integrating data processing and model improvement methods. First, because of the non-stationary and nonlinear characteristics of ultra-short-term load data, the TVF-EMD is applied to decompose the load data, resulting in several intrinsic mode functions (IMF). These IMF components are then regrouped according to their entropy values using sample entropy (SE). Next, the iTransformer neural network is enhanced by introducing a SASA mechanism to replace the self-attention mechanism in the iTransformer encoder, effectively improving the model's ability to capture dependencies among different variables. Finally, the regrouped components are input into the iTransformer based on SASA mechanism for prediction, and the results are combined to obtain the final forecast value. Validation is performed using an actual active load dataset from the high-voltage side of a 220kV substation in a certain region of China, and comparisons are made with existing mainstream models. The results show that the forecasting method proposed in this paper offers superior predictive performance.

KEY WORDS: ultra-short-term load forecasting; time-varying filter empirical mode decomposition; sample entropy; iTransformer model; attention mechanism

摘要: 准确的负荷预测对电力系统安全稳定运行至关重要。为了进一步提高负荷预测的精准度,将数据处理和模型改进

的方法相融合,提出了一种基于时变滤波经验模态分解(time-varying filter empirical mode decomposition, TVF-EMD)-重构和独立自注意力(stand-alone self-attention, SASA)机制的 iTransformer 超短期负荷预测方法。首先,针对超短期负荷数据的非平稳和非线性特性,采用 TVF-EMD 对负荷数据进行分解,得到若干本征模态函数(intrinsic mode function, IMF),通过样本熵(sample entropy, SE)按熵值的大小将 IMF 分量进行重组;其次,对 iTransformer 神经网络进行改进,引入一种独立自注意力机制替换 iTransformer 编码器中的自注意力机制,有效提升了模型捕捉不同变量的依赖关系的能力;最后,将重组后的分量输入到基于独立自注意力机制的 iTransformer 中进行预测,将得到的结果进行叠加得到最终的预测值。以我国某地区 220kV 变电站高压侧的实际有功负荷数据集为例进行验证并与现有主流模型进行对比,结果表明该文采用的预测方法具有更好的预测性能。

关键词: 超短期负荷预测; 时变滤波经验模态分解; 样本熵; iTransformer 模型; 注意力机制

DOI: 10.13335/j.1000-3673.pst.2024.1970

0 引言

负荷预测是电力系统中至关重要的环节,其是通过历史负荷相关数据的分析和建模,预测未来某个时间段内电力系统的负荷需求,根据预测时间负荷预测可以分为超短期负荷预测、短期负荷预测、中期负荷预测和长期负荷预测^[1]。其中,超短期负荷预测是指预测时间范围为 1h 内的负荷需求,主要用于电力系统的实时调度和短期控制^[2]。由于电力负荷的波动性和不确定性,超短期负荷预测对于维持电力系统的平衡和稳定至关重要^[3-4]。

目前,负荷预测的方法可以分为物理方法,统计方法和人工智能的方法。物理方法主要基于物理

基金项目: 中国电科院青年科学家项目(52420023000Z)。

Project Supported by Youth Scientist Project of China Electric Power Research Institute (52420023000Z).

学原理和工程学知识,需要复杂的建模和计算,对于大规模系统或快速变化的负荷情况存在一定的局限性^[5];统计方法是通过构建回归模型来预测负荷,例如自回归积分滑动平均(auto regressive integrated moving average, ARIMA)模型^[6]、季节性自回归积分滑动平均(seasonal auto regressive integrated moving average model, SARIMA)模型^[7]等,尽管统计方法在处理周期性和趋势性数据方面具有一定优势,但在处理突发性和随机性事件时,其预测精度往往较低;人工智能方法由于其强大的数据挖掘能力和数据处理能力,在电力系统负荷预测中应用广泛^[8],传统的人工智能方法包括支持向量机^[9]和随机森林^[10]等,近年来,随着人工智能的不断发展,以深度学习为代表的人工智能模型得到了电力专家学者的青睐,如反向传播神经网络(back propagation neural network, BPNN)^[11],卷积神经网络(convolutional neural network, CNN)^[12]和长短时记忆网络(long short term memory, LSTM)^[13]等。文献[11]提出一种基于 Hadoop 架构的多重分布式 BPNN 的短期负荷预测方法,结果表明该方法比传统的 BPNN 具有更高的精度;文献[14]提出了一种 CNN 与 LSTM 相结合的方法,首先将负荷数据进行模态分解并分组,再利用 CNN 挖掘数据中的特征向量,最后将其输入到 LSTM 中进行预测,在实际负荷数据集上与其他模型进行对比,其方法预测精度更高;文献[15]提出一种以 LSTM 作为共享层的多任务学习负荷预测方法,通过共享层来模拟多元负荷间的耦合信息,取得了良好的预测效果。以上的方法虽然都在负荷预测上取得了一定的效果,但它们的模型都存在着长期依赖和梯度消失等问题。Transformer 模型^[16]通过其独特的自注意力机制能够很好的捕获长期依赖关系,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。文献[17]利用 XGBoost 方法对电力系统负荷数据进行规约,然后将数据输入到改进的 Transformer 模型中,不仅提高了模型的效率,还提高了模型的准确性;文献[18]在 Transformer 模型基础上进行改进,采用门控循环单元替换词嵌入,在多头注意力中加入交流机制,将门控单元加入到残差连接中,提高了 Transformer 模型的预测性能。但由于 Transformer 复杂的结构以及众多数量的超参数,增加了模型的计算成本和优化难度^[19],基于 Transformer 架构的改进模型如 Informer^[20]、Autoformer^[21]、FEDformer^[22]和 Crossformer^[23]等虽然都在一定程度上提升了 Transformer 的性能,但

在现有的 Transformer 视角下,每个 token 的嵌入聚合了代表不同物理量的多个变量,这使得模型难以将变量作为中心,忽略了多变量之间的相关性。文献[24]提出了一种倒置的 Transformer 模型 iTransformer,从时间序列的反向视角出发,将每个变量的整条时间序列作为一个 token,通过这样倒置,嵌入的 token 聚集了序列的全局表征,将变量作为中心,更好地利用自注意力机制进行多变量关联。iTransformer 是基于原 Transformer 的组件,采用自注意力层来处理不同变量之间的依赖关系,本文为了提升 iTransformer 捕获依赖关系的能力,将一种独立自注意力(stand-alone self-attention, SASA)机制^[25]引入到 iTransformer 中,替换其原有的自注意力机制。

短期的负荷数据通常具有强烈的波动性,是典型的非平稳数据,需要利用经验模态分解的方法对数据进行预处理来提高模型预测精度。目前经验模态分解的方法主要包括集成经验模态分解(ensemble empirical mode decomposition, EEMD)^[26],完全自适应噪声集合经验模态分解(complete ensemble empirical mode decomposition with adaptive noise, CEEMDAN)^[27],变分模态分解(variational mode decomposition, VMD)^[28]等,上述方法虽然都有着各自的优势,但是这些传统分解方法在处理非平稳信号时,往往受到固定滤波器、不准确的瞬时频率估计以及模式混叠问题的限制,缺乏对信号局部特性和动态变化的适应能力,导致分解结果可能包含伪模式或丢失信号的关键特征。基于时变滤波器的经验模态分解(time-varying filter empirical mode decomposition, TVF-EMD)方法^[29]是一种完全自适应的信号分解方法,其通过时变滤波器动态调整分解过程,针对的是时域信号的非线性、非平稳特性,相较于传统的分解方法分解性能更优,因此适用于本文中波动较大的超短期负荷预测。该方法分解产生过多的本征模态函数分量(intrinsic mode function, IMF)会给模型带来数据量过大的问题,为了节约预测时间以及提高模型预测精度,有必要对 TVF-EMD 分解后的 IMF 分量通过样本熵(sample entropy, SE)算法进行重组,以提高模型的计算效率。

综上所述,本文提出了基于 TVF-EMD 分解-SE 重构和独立自注意力机制的 iTransformer 预测方法,将数据处理的改进方法进行融合,进一步提升负荷预测的效果。首先,针对负荷数据集的复杂性和多样性,将原始负荷数据集采用 TVF-

EMD 算法进行经验模态分解, 对数据进行了滤波, 来有效应对超短期负荷数据的波动性; 采用 SE 算法计算经验模态分解得到的本征模态分量的熵值, 根据熵值大小将本征模态分量重构为新的分量, 减少后续模型的训练时间。其次, 在 iTransformer 神经网络的基础上引入了一种独立自注意力机制替换原 iTransformer 中的自注意力机制, 增强了模型对于超短期负荷的预测性能。最后, 将重构的 IMF 分量分别输入到基于独立自注意力机制的 iTransformer 中, 得到的所有结果进行重构以实现对未来 1h 负荷数据的精确预测。采用我国某地区变电站高压侧的实际有功负荷数据集进行算例验证分析, 结果表明本文采用的预测方法可有效提升负荷预测的精度。

1 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 模型构建

本文提出了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 模型对负荷功率进行预测, 预测流程如图 1 所示。首先, 将负荷

原始数据集中的有功负荷数据进行 TVF-EMD 分解, 得到 n 个 IMF 分量, 并采用 SE 算法计算所有 IMF 分量的样本熵, 将熵值相近的 IMF 分量相加, 得到 m 个新的 IMF 分量, 然后将这些 IMF 分量分别输入到基于独立自注意力机制的 iTransformer 中进行预测。图 1 中基于独立自注意力机制的 iTransformer 只有一个编码器架构, 输入的 IMF 分量首先经过一个嵌入层, 它将不同变量的整条序列当作一个 token 嵌入, 然后输入到具有 L 层的 Transformer 块中; SASA 机制应用于 Transformer 块的注意力机制中, 它将查询向量与键向量进行 softmax 操作, 再与值向量相乘得到注意力分数, 从而获取变量之间的相关性; 层归一化应用于每个变量的特征表示, 让所有变量的特征通道都处于相对统一的分布下。前馈网络形成同一变量的整条序列的词向量, 使得模型提高了未见过的序列上的泛化能力, 在经过前馈层之后再经过一个层归一化, 通过映射层投影出最终的预测结果, 求得所有 IMF 分量预测结果之和即为最终的负荷功率的预测值。

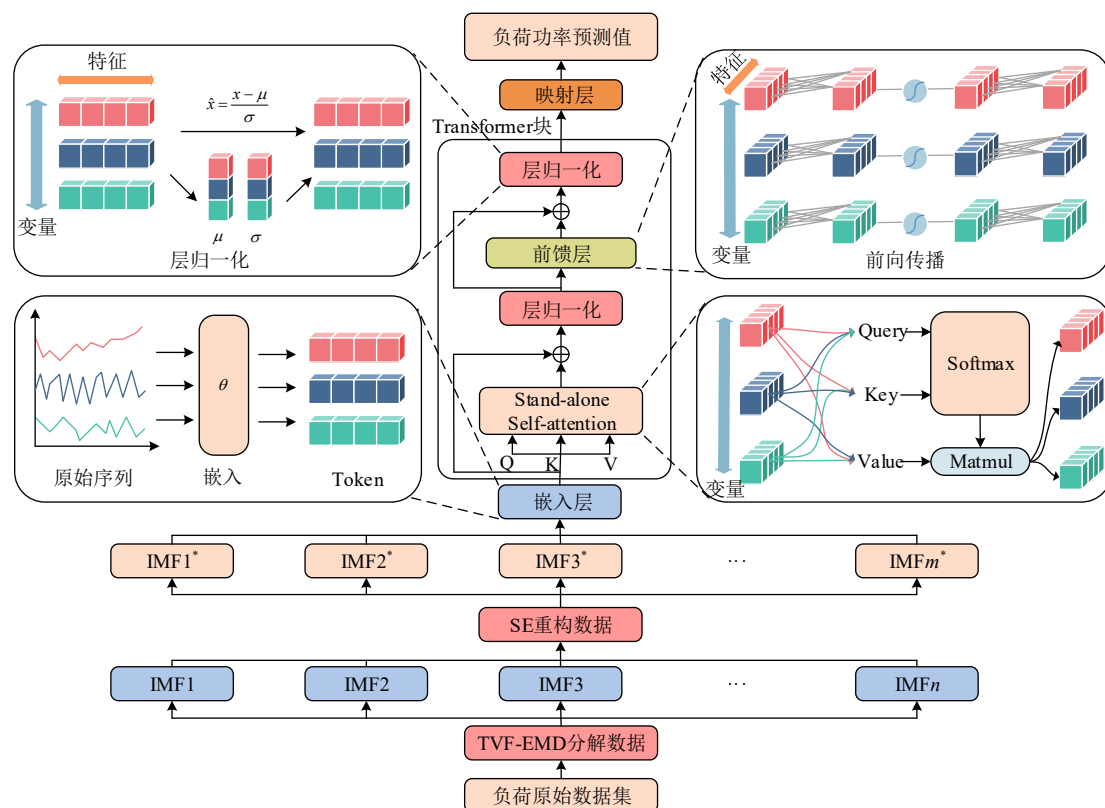


图 1 基于时变滤波经验模态分解-重构和独立自注意力的 iTransformer 模型
Fig. 1 iTransformer model based on TVF-EMD-reconstruction and SASA mechanism

2 算法原理

2.1 TVF-EMD 算法

TVF-EMD 算法是一种改进的经验模态分解方

法, 常用于处理非平稳和非线性信号。该算法是通过设计一组时间变化的滤波器来分解信号, 这些滤波器的频率响应随时间变化, 能够更好地适应信号的局部特性, 从而得到一系列本征模态函数, 克服

了传统 EMD 在处理复杂信号时带来的模态混叠和端点效应等问题。

TVF-EMD 算法充分利用了瞬时振幅和频率信息,且采用了非均匀 b 样条近似作为时变滤波器,具有完全自适应的特性,其具体步骤如下所示:

1) 假设负荷序列为 $X(t)$, 经过希尔伯特变换后表示为 $\hat{X}(t)$, 其瞬时幅值 $A(t)$ 和瞬时频率 $\varphi'(t)$ 表示为

$$A(t) = \sqrt{X(t)^2 + \hat{X}(t)^2} \quad (1)$$

$$\varphi'(t) = d(\arctan[\hat{X}(t) / X(t)]) / dt \quad (2)$$

2) 通过比较瞬时幅值 $A(t)$ 的大小, 确定该值的局部极大值序列和局部极小值序列, 分别表示为 $A(\{t_{\max}\})$ 与 $A(\{t_{\min}\})$ 。

3) 计算瞬时均值 $a_1(t)$ 和瞬时包络 $a_2(t)$ 。

$$a_1(t) = \frac{\beta_1(t) + \beta_2(t)}{2} \quad (3)$$

$$a_2(t) = \frac{\beta_1(t) - \beta_2(t)}{2} \quad (4)$$

式中: $\beta_1(t)$ 与 $\beta_2(t)$ 分别为对 $A(\{t_{\max}\})$ 与 $A(\{t_{\min}\})$ 进行插值得到。

4) 分别对数值序列 $A^2(\{t_{\min}\})\varphi'(\{t_{\min}\})$ 与 $A^2(\{t_{\max}\})\varphi'(\{t_{\max}\})$ 进行插值, 得到 $\eta_1(t)$ 与 $\eta_2(t)$, 瞬时频率分量 $\varphi'_1(t)$ 与 $\varphi'_2(t)$ 可以表示为

$$\varphi'_1(t) = \frac{\eta_1(t)}{2a_1^2(t) - 2a_1(t)a_2(t)} + \frac{\eta_2(t)}{2a_1^2(t) + 2a_1(t)a_2(t)} \quad (5)$$

$$\varphi'_2(t) = \frac{\eta_1(t)}{2a_2^2(t) - 2a_1(t)a_2(t)} + \frac{\eta_2(t)}{2a_2^2(t) + 2a_1(t)a_2(t)} \quad (6)$$

5) 计算局部截止频率 $\varphi'_{\text{bis}}(t)$ 。

$$\varphi'_{\text{bis}}(t) = \frac{\varphi'_1(t) + \varphi'_2(t)}{2} \quad (7)$$

6) 为防止序列中的间歇成分对局部截止频率产生影响, 对 $\varphi'_{\text{bis}}(t)$ 进行重构, 计算表达式为

$$f(t) = \cos\left[\int \varphi'_{\text{bis}}(t) dt\right] \quad (8)$$

7) 将 $f(t)$ 的极值点作为时变滤波器的节点, 用 B 样条插值方法对负荷序列 $X(t)$ 逼近。

8) 设置停止信号 $\theta(t)$ 。

$$\theta(t) = \frac{B_{\text{Loughlin}}(t)}{\varphi_{\text{avg}}(t)} \quad (9)$$

式中: $B_{\text{Loughlin}}(t)$ 为 Loughlin 瞬时带宽; $\varphi_{\text{avg}}(t)$ 为加权平均瞬时频率。

若 $\theta(t)$ 小于阈值 ξ , 将 $X(t)$ 作为一个 IMF 分量, 若不满足, 则令 $x_1(t) = x_t - m(t)$, 式中 $m(t)$ 为瞬时包络局部平均值, 重新执行式(1)~(9)。

上述步骤即 TVF-EMD 算法的计算原理, 通过

不断重复上述步骤, 可以得到模态分解后的所有的 IMF 分量。

这些 IMF 分量反映了负荷序列的不同成分。对于高频 IMF 分量, 其反映的是负荷序列中快速变化的部分, 代表负荷信号中的短期随机波动或高频扰动, 可能是由负荷的短时切换(如设备启动或停机)引起的瞬态效应, 或者是测量过程中的传感器噪声或环境干扰。对于中频 IMF 分量, 其与电力负荷中的周期性变化有关, 例如日负荷波动和周期性市场行为或特定用电规律。对于低频 IMF 分量, 其通常反映信号的长期趋势或缓慢变化, 例如季节性负荷变化和社会经济发展导致的中长期用电需求趋势。对于残余项, 其代表信号中最慢的变化或整体趋势, 它通常反映负荷数据的长期变化趋势。

2.2 SE 算法

SE 算法是一种在数据科学和信息理论中广泛应用的概念, 用于描述数据集中的复杂度和不确定性。它通常通过计算样本中不同类别或特征的分布来确定来衡量样本中信息的混乱程度或多样性, SE 算法的计算不依赖数据长且具有更好的一致性。TVF-EMD 算法分解出的 IMF 分量数量较多时, 容易造成计算量过大的问题。SE 算法可以用来评估 IMF 分量的纯度和频率大小, SE 值相似的 IMF 分量具有类似的物理意义, 可以重构为一个新的分量, 避免了大量 IMF 分量造成的计算复杂度过高的问题, 从而减少了模型的训练时间。

2.3 iTransformer 模型

传统的 Transformer 模型在自然语言处理和计算机视觉领域取得了巨大的成功, 但在解决时间序列预测任务中面临着一些挑战, 特别是当回顾窗口较大时, Transformer 的性能会显著下降, 同时计算的复杂度也会迅速增加。此外, 传统 Transformer 将多个变量的同一时间点嵌入到一个时间标记中, 可能会导致不同变量之间的相关性被忽略^[24]。iTransformer 是通过对 Transformer 结构进行反转的模型, 可有效解决 Transformer 所面临的上述问题。图 2 展示了传统 Transformer 与 iTransformer 嵌入方式的比较, 传统 Transformer 嵌入的 token 包含每个时间步长的多元表示, 同一个时间步长中包含了完全不同的物理意义, 难以揭示有益的信息。iTransformer 将每个序列独立地嵌入到 token 中, 通过倒置, 嵌入的 token 聚合了序列的全局表示。这些表示能够聚焦于输入的变量, 并通过模型自身的注意力机制有效的描述多元变量的相关性。

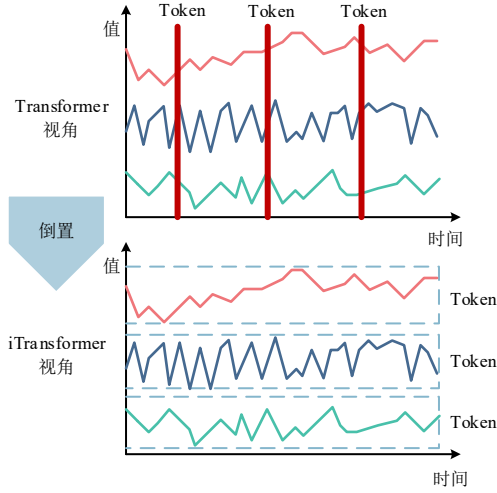


图2 Transformer 与 iTransformer 嵌入方式的比较
Fig. 2 Comparison between Transformer and iTransformer embedding methods

不同于以往 Transformer 预测模型使用的较为复杂的 encoder-decoder 结构, iTransformer 仅使用 encoder, 它包括嵌入层 (Embedding), 投影层 (Projection) 和 L 个可堆叠的 Transformer 块 (TrmBlock)。对于一个时间长度为 T , 变量数为 N 的多维时间序列 $X \in \mathbb{R}^{T \times N}$, 使用 $X_{t,:}$ 表示表示为 t 步同时记录的时间点, 用 $X_{:,n}$ 表示为由 n 索引的每个变量的整个时间序列。由于 $X_{:,n}$ 比 $X_{t,:}$ 具有更强的语义与相对一致的测量单位, 与传统对 $X_{t,:}$ 进行特征嵌入的方式相比, iTransformer 使用嵌入层对每个 $X_{:,n}$ 独立进行特征映射, 获得 N 个维度为 D 的变量的特征表示 $H = \{h_1, \dots, h_N\} \in \mathbb{R}^{N \times D}$, 其中 $h_i \in \mathbb{R}^D$ 表示变量在历史时间内的变化。这些特征表示在 Transformer 块中, 首先通过注意力机制进行变量之间的信息交互, 使用层归一化统一不同变量的特征分布, 在前馈网络中进行全连接式的特征编码, 最终通过投影层映射为预测结果。整个模型实现过程可以表示为

$$h_n^0 = \text{Embedding}(X_{:,n}) \quad (10)$$

$$H^{l+1} = \text{TrmBlock}(H^l), l = 0, \dots, L-1 \quad (11)$$

$$\hat{Y}_{:,n} = \text{Projection}(h_n^L) \quad (12)$$

式中: $\hat{Y}_{:,n}$ 为每个变量的预测结果。

2.4 独立自注意力

iTransformer 虽然通过自注意力机制考虑了所有的输入向量, 但对于捕获不同向量间潜在的相关性的能力欠佳。本文采用了一种独立自注意力机制, 将位置信息添加到了自注意力运算中, 还通过空间感知自注意力来解决自注意力无法包含空间信息的问题, 提升了模型捕获不同变量关系的能力。

独立自注意力机制的计算过程如图 3 所示, 输入是一个特征图, 将图中的一个像素 $x_{ij} \in \mathbb{R}^d$, 和以 x_{ij} 为中心提取的位置为 $a, b \in N_k(i, j)$ 像素局部区域 x_{ab} , 利用权重矩阵 W_Q 、 W_K 和 W_V 计算该目标区域的 query、keys 和 values 的值。然后将每一个 query 单元与 keys 中的各个 key 单元相乘并求和, 再经过 softmax 层后与 values 对应像素位置相乘后加权得到最终的输出单元 y_{ij} , 其计算式如下式所示:

$$y_{ij} = \sum_{a,b \in N_k(i,j)} \text{softmax}_{ab}(q_{ij}^T k_{ab}) v_{ab} \quad (13)$$

式中: q_{ij} 为查询向量; k_{ab} 为键向量; v_{ab} 是值向量; 它们是位置 (i, j) 及附近像素点经过线性变换得到的; softmax_{ab} 表示在 ij 附近将输入映射到 0-1 的范围内。

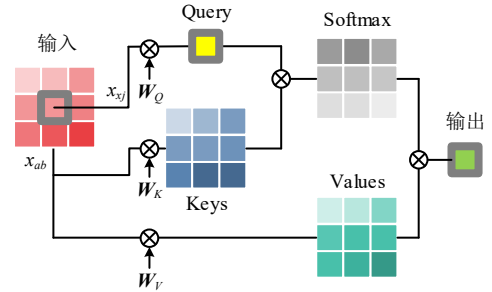


图3 独立自注意力计算方式
Fig. 3 SASA calculation method

为了获得更好的预测精度, 引入位置信息并将其编码在注意力中, 首先定义 (i, j) 到每个位置的 $a, b \in N_k(i, j)$, 每个元素接受两个距离: 行偏移量 $a-i$ 与列偏移量 $b-i$, 行偏移量与列偏移量分别与嵌入 r_{a-i} 和 r_{b-i} 相关联, 两者被连接起来形成 $r_{a-i, b-i}$, 这种空间相对注意力表示为

$$y_{ij} = \sum_{a,b \in N_k(i,j)} \text{softmax}_{ab}(q_{ij}^T k_{ab} + q_{ij}^T r_{a-i, b-i}) v_{ab} \quad (14)$$

该注意力在视觉领域取得了良好的效果, 对其进行一定的维度变换, 本文将该注意力机制应用到 iTransformer 中并替换原有的自注意力机制, 提高捕获不同变量依赖关系的能力。

3 算例分析

本文算例仿真在 Python3.9 版本和 Pytorch1.9.1 版本的运行环境下运行。采用的设备为 Intel(R) CoreTM i5-9300H CPU @ 2.40GHz 处理器, 运行内存为 16GB, 显卡为 NVIDIA GeForce GTX 1050, 显存为 3GB。

3.1 实验数据集

本文采用我国某地区 220kV 变电站高压侧的实际有功负荷数据集和该地区的天气信息进行实验。数据集选取了 2019 年 1 月 1 日—2019 年 12 月 26 日

的电力负荷数据, 采样间隔为 5min, 共 103680 条数据, 除有功功率负荷外, 还包含最高气温、最低气温、天气、风向与风速 5 种天气特征。采用 7:1:2 的比例将实验数据集分为训练集、验证集与测试集。

3.2 模型评价指标

本文采用平均绝对误差(mean absolute error, MAE) y_{MAE} 和平均绝对百分比误差(mean absolute percentage error, MAPE) y_{MAPE} 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; 平均绝对百分比误差衡量预测值和实际值之间差异的平均绝对百分比, 以百分比形式表达误差。平均绝对误差与平均绝对百分比误差的公式表示为

$$y_{MAE} = \frac{1}{n} \sum_{i=1}^n |y_{act}(i) - y_{pred}(i)| \quad (15)$$

$$y_{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_{act}(i) - y_{pred}(i)}{y_{act}(i)} \right| \quad (16)$$

式中: n 表示负荷预测结果的数量; $y_{act}(i)$ 和 $y_{pred}(i)$ 为第 i 条数据中的负荷的真实值与预测值。

3.3 模型参数设置

本文构建的预测模型具有多个关键参数: 如 TVF-EMD 中的带宽阈值、分量个数和 b 样条阶, iTransformer 中的训练回合、输入序列长度、输出序列长度、编码器层数、批尺寸、激活函数、损失函数、模型学习率、丢弃率等等, 其具体设置见表 1。

表 1 模型参数设置 Table 1 Model parameter settings		
模型	参数	参数值
TVF-EMD	带宽阈值	0.25
	分量个数	12
	b 样条阶	50
iTransformer	训练回合	10
	输入序列长度	96
	输出序列长度	12
	编码器层数	2
	批尺寸	32
	激活函数	gelu
	损失函数	mse
	丢弃率	0.1
	学习率	0.0001
SASA	步幅	1
	组数	1

本文采用的对比模型 CNN、LSTM 和 Transformer 的参数设置见附录 A。本文采用滚动预测方式, 输入为 96 个已知负荷数据, 输出为 12 个未来负荷数据, 每个负荷数据的时间间隔为 5min, 即采用已知 8h 负荷数据预测未来 1h 的负荷数据。

本文选择 gelu 作为激活函数, 相较于 sigmoid、relu 等常用激活函数更加平滑, 有助于提高模型训练过程的收敛速度和性能, 同时本文还选择了较小的学习率以及较小的批尺寸, 通过上述方法可以有效应对“梯度爆炸”问题。

3.4 原始负荷数据分解及重构

本文使用 TVF-EMD 算法将原始负荷数据分解为 12 个 IMF 分量, 如图 4 所示。不同的 IMF 分量表示了原始负荷功率序列在不同时间尺度上的特征。

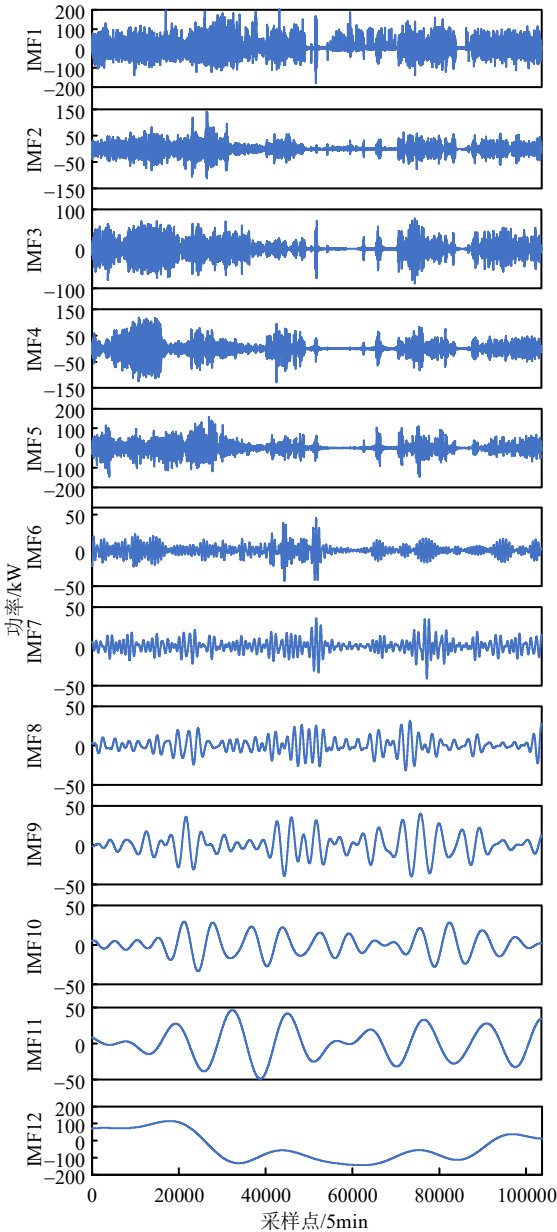


图 4 TVF-EMD 分解后的分量
Fig. 4 TVF-EMD decomposed components

若直接将 12 个 IMF 分量输入到基于独立自注意力机制的 iTransformer 进行预测, 模型训练的时间会大大增加, 而且此外, 对于频率较高的分量, 由于其高复杂度导致模型的预测结果较差^[30]。为了节约模型训练时间, 有必要使用 SE 算法对 12 个 IMF 分量进

行重构。TVF-EMD 分解后各 IMF 分量的 SE 值如表 2 所示。

根据表 2 中样本熵的大小,采用 K-means 算法将相似的分量进行聚类,重构为新的分量。为了验证重构分量对于减少模型训练时间的作用,分别将未重构的 12 个分量、重构为 8 个分量和重构为 4 个分量时分别输入到基于独立自注意力机制的 iTransformer 中进行训练,所用时间如表 3 所示。可见,当重构为 4 个分量时的时间较短,本文选取训练时间最短的 4 个分量作为模型的输入来进行预测。

表 2 IMF1-IMF12 的 SE 值 Table 2 SE value of IMF1-IMF12			
分量	SE 值	分量	SE 值
IMF1	0.67421	IMF7	0.40045
IMF2	0.68003	IMF8	0.40082
IMF3	0.66758	IMF9	0.39587
IMF4	0.62270	IMF10	0.38202
IMF5	0.54136	IMF11	0.36593
IMF6	0.40888	IMF12	0.35393

表 3 不同分量数的模型训练时间 Table 3 Performance comparison of different models	
分量数	模型训练时间/s
12	13223
8	9466
4	4009

3.5 预测结果及分析

为更好地验证本文提出的基于时变滤波经验模态分解-重构和独立自注意力的 iTransformer 模型的预测能力,分别采用当前主流 CNN、LSTM、Transformer 与 iTransformer 模型对负荷功率进行预测,将得到的结果与本文所提的模型进行比较,如表 4 所示。

表 4 不同模型性能比较 Table 4 Performance comparison of different models		
模型	y_{MAE}	y_{MAPE}
CNN	0.03290	0.07150
LSTM	0.03233	0.06886
Transformer	0.02338	0.04918
iTransformer	0.01647	0.03141
本文方法	0.01470	0.02828

根据表 4 中的结果,可以得出以下结论:

1) 在上述 5 种不同的方法中,本文的方法取得了最好的效果,验证了本文提出的方法对于负荷预测的有效性与优越性。

2) 本文提出的方法相较于 iTransformer 模型预测效果有一定的提升,MAE 降低了 10.75%,MAPE 降低了 9.96%,本文提出的模型通过将原始波动较

大的负荷功率分解为一系列频率不同的 IMF 分量并进行重组,降低了负荷数据的复杂性,又引入独立自注意力机制引入到 iTransformer 中,提升了模型捕获长距离依赖能力。本文提出的方法相较于 CNN、LSTM 与 Transformer 三个对比模型,MAE 指标分别降低了 55.32%, 54.53%, 37.13%, MAPE 指标分别降低了 60.45%, 58.93%, 42.50%, 本文提出的方法在 MAE, MAPE 指标上相较于传统方法大幅降低,进一步验证了本文提出方法的有效性。

为了更直观地展示本文提出方法的准确性,在预测结果中,选取 2019 年 11 月 2 日的结果进行可视化对比,如图 5 所示。

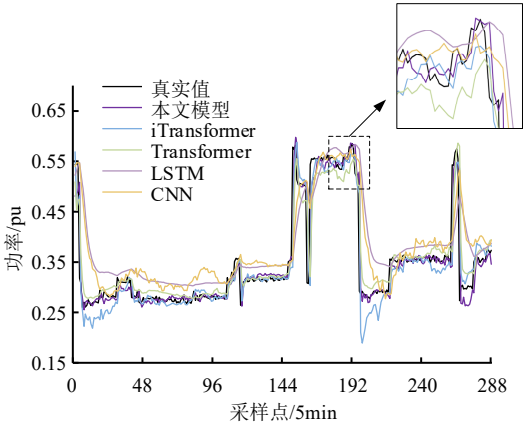


图 5 2019 年 11 月 2 日预测结果对比
Fig. 5 Comparison of forecast results on 2019/11/2

由图 5 可以看出, CNN 模型与 LSTM 模型没有很好地学习到负荷功率曲线的变化,与真实值有着较大的差异。Transformer 模型很好地预测到了负荷功率曲线的规律,但在一些平缓曲线上效果欠佳。iTransformer 模型在数据波动不大时预测效果良好,但在数据变化剧烈时与实际负荷功率曲线有着一定的误差。本文提出的方法在一定程度上克服了上述模型中存在的问题,取得了最优的预测效果,即使是在数据波动较大时,如在采样点 1、144 与 192 附近负荷发生剧烈波动,本文模型与负荷功率曲线拟合度依然较高,体现了该模型对于负荷数据急剧变化具有较好的处理能力。

3.6 消融实验

为验证时变滤波经验模态分解-重构算法和独立自注意力机制对 iTransformer 性能的影响,设置以下几组对照实验:第一组为原始的 iTransformer 模型,第二组在原始的 iTransformer 模型上加入时变滤波经验模态分解-重构算法,第三组在原始的 iTransformer 中使用独立自注意力机制替换自注意力机制,第四组为本文提出的时变滤波经验模态分

解-重构和独立自注意力机制的 iTransformer 模型。实验结果如表 5 所示。

表 5 消融实验结果 Table 5 Results of ablation experiment		
对照组	YMAE	YMAPE
第一组	0.01647	0.03141
第二组	0.01492	0.02909
第三组	0.01517	0.02910
第四组	0.01470	0.02828

结果表明,时变滤波经验模态分解-重构算法与独立自注意力机制替换自注意力的方法都能使得 iTransformer 的预测性能得到提升。其中,时变滤波经验模态分解-重构算法使得原始模型的 MAE 指标降低了 9.41%,独立自注意力机制替换自注意力的方法使得原始模型 MAE 指标降低了 7.89%,而本文提出的模型结合了以上这两种方法,使得 MAE 指标降低了 10.75%,取得了最优的预测效果。

3.7 不同分解方法和注意力机制下 iTransformer 的性能对比

为验证本文中时变滤波经验模态分解算法与独立自注意力机制对与 iTransformer 模型预测能力提升的优越性,本文选取了几种典型的自注意力机制和经验模态分解算法分别用在原始的 iTransformer 模型中作对比试验。

3.7.1 不同自注意力机制的对比

本文选取了 ProbSparse attention^[20]、LSH attention^[31]、Flow attention^[32]3 种自注意力机制与独立自注意力机制对比,来分析不同注意力机制给 iTransformer 模型带来的效果。上述自注意力机制分别具有不同特点,如 ProbSparse attention 主要是通过概率方法选择最重要的一部分注意力权重进行计算,而忽略那些对结果影响较小的权重,其能够显著降低计算复杂度,同时保持较高的模型性能。LSH attention 通过哈希函数将相似的数据点映射到相同的哈希桶中,从而在低维空间中保留高维空间的相似性,这样可以大幅减少注意力计算的复杂度。Flow attention 是通过守恒性质将竞争引入注意力机制,确保了各个 token 之间的竞争和分配,同时保持线性复杂度。

表 6 为不同注意力机制下的实验结果。根据对比结果可知,相较于以上 3 种作注意力机制中,本文采用的独立自注意力机制通过添加位置信息与空间感知自注意力,使得模型取得了最好的预测结果。值得注意的是,ProbSparse attention 的结果表现不如 iTransformer 中自注意力机制,可能是由于其稀疏性忽略了负荷的部分信息所导致。

表 6 不同注意力实验结果比较 Table 6 Comparison of different attention experiment results		
注意力机制	YMAE	YMAPE
ProbSparse attention	0.01733	0.03376
LSH attention	0.01550	0.02987
Flow attention	0.01595	0.03064
SASA	0.01517	0.02910

3.7.2 不同经验模态分解方法对比

本文选取了典型的 EMD 算法与 CEEMDAN 算法与 TVF-EMD 算法进行对比,来分析不同经验模态分解方法给 iTransformer 模型带来的效果。通过这两种对比算法与本文 SE 算法结合,分别得到了 4 个 IMF 分量,最终预测结果见表 7。根据对比结果可知,EMD 算法由于存在着模态混叠和端点效应等问题,使得模型预测精度下降。CEEMDAN 算法引入自适应噪声和多次迭代的方式,相较于 EMD 算法有一定的提升,但效果不如 TVF-EMD 算法。TVF-EMD 算法引入时变滤波器,相较于 EMD 算法和 CEEMDAN 算法拥有更好的分解能力,进而提升了负荷预测的效果。

表 7 不同经验模态分解实验结果比较 Table 7 Comparison of different empirical mode decomposition experiment results		
分解算法	YMAE	YMAPE
EMD	0.01850	0.03529
CEEMDAN	0.01547	0.02999
TVF-EMD	0.01493	0.02909

4 结论

为提高负荷预测的预测精度,本文提出了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 超短期负荷预测模型,该模型结合了时变滤波经验模态分解-重构算法、独立自注意力机制和 iTransformer 模型来预测负荷功率。其中时变滤波经验模态分解算法、SE 算法和独立自注意力机制分别起到了数据滤波、加快训练速度和捕获变量依赖关系等作用,通过将以上这些方法应用到新颖的 iTransformer 模型中,提高了模型的预测精度。本文使用我国某地区 220kV 变电站高压侧的实际有功负荷数据集进行验证,与 CNN、LSTM、Transformer 与 iTransformer 模型这些主流模型进行对比实验,结果表明,本文的方法在 MAE 与 MAPE 两个评价指标上均优于其他的基准模型。

在未来的研究中,应重点关注新型的人工智能模型,深度神经网络等各种前沿技术在处理复杂时间序列数据方面展现出巨大的潜力,通过利用这些先进模型,结合数据和模型增强的方法显著提高负

荷预测的精度和可靠性, 及时应对电力需求的动态变化, 为新型电力系统的优化和运行提供强有力的技术支撑。

附录见本刊网络版(<http://www.dwjs.com.cn/CN/1000-3673/current.shtml>)。

参考文献

- [1] 康重庆, 夏清, 刘梅. 电力系统负荷预测[M]. 2版. 北京: 中国电力出版社, 2017: 7-8.
- [2] 范士雄, 刘幸蔚, 於益军, 等. 基于多源数据和模型融合的超短期母线负荷预测方法[J]. 电网技术, 2021, 45(1): 243-250.
FAN Shixiong, LIU Xingwei, YU Yijun, et al. Multi-source data and hybrid neural network based ultra-short-term bus load forecasting[J]. Power System Technology, 2021, 45(1): 243-250(in Chinese).
- [3] 臧海祥, 陈玉伟, 程礼临, 等. 基于多尺度分量特征学习的用户级超短期负荷预测[J]. 电网技术, 2024, 48(6): 2584-2592.
ZANG Haixiang, CHEN Yuwei, CHENG Lilin, et al. User level ultra-short-term load forecasting based on multi-scale component feature learning[J]. Power System Technology, 2024, 48(6): 2584-2592(in Chinese).
- [4] 窦真兰, 张春雁, 许一洲, 等. 基于多变量相空间重构和径向基函数神经网络的综合能源系统电冷热超短期负荷预测[J]. 电网技术, 2024, 48(1): 121-128.
DOU Zhenlan, ZHANG Chunyan, XU Yizhou, et al. Ultra-short-term load forecasting of electrical, cooling and heating for integrated energy system based on multivariate phase space reconstruction and radial basis function neural network[J]. Power System Technology, 2024, 48(1): 121-128(in Chinese).
- [5] 廖婉焕, 胡智宏, 马莹莹, 等. 电力系统短期负荷预测方法综述[J]. 电力系统保护与控制, 2011, 39(1): 147-152.
LIAO Nihuan, HU Zhihong, MA Yingying, et al. Review of the short-term load forecasting methods of electric power system[J]. Power System Protection and Control, 2011, 39(1): 147-152(in Chinese).
- [6] 吴峰, 王飞, 顾康慧, 等. 基于 MEEMD-ARIMA 模型的波浪能发电系统输出功率预测[J]. 电力系统自动化, 2021, 45(1): 65-70.
WU Feng, WANG Fei, GU Kanghui, et al. Output power prediction of wave energy generation system based on modified ensemble empirical mode decomposition-autoregressive integrated moving average model[J]. Automation of Electric Power Systems, 2021, 45(1): 65-70(in Chinese).
- [7] 徐晶, 迟福建, 葛磊蛟, 等. 基于 SARIMA-GRNN-SVM 的短期商业电力负荷组合预测方法[J]. 电力系统及其自动化学报, 2020, 32(2): 85-91.
XU Jing, CHI Fujian, GE Leijiao, et al. Short-term combined commercial load forecasting method based on SARIMA-GRNN-SVM[J]. Proceedings of the CSU-EPSA, 2020, 32(2): 85-91(in Chinese).
- [8] FAN Shixiong, GUO Jianbo, MA Shicong, et al. Enabling artificial intelligence-based scenario application in new type power systems[J]. Energy Internet, 2024: 1(1): 9-13.
- [9] 肖白, 赵晓宁, 姜卓, 等. 利用模糊信息粒化与支持向量机的空间负荷预测方法[J]. 电网技术, 2021, 45(1): 251-258.
XIAO Bai, ZHAO Xiaoning, JIANG Zhuo, et al. Spatial load forecasting method using fuzzy information granulation and support vector machine[J]. Power System Technology, 2021, 45(1): 251-258(in Chinese).
- [10] 史佳琪, 张建华. 基于多模型融合 Stacking 集成学习方式的负荷预测方法[J]. 中国电机工程学报, 2019, 39(14): 4032-4041.
SHI Jiaqi, ZHANG Jianhua. Load forecasting based on multi-model by stacking ensemble learning[J]. Proceedings of the CSEE, 2019, 39(14): 4032-4041(in Chinese).
- [11] 苏学能, 刘天琪, 曹鸿谦, 等. 基于 Hadoop 架构的多重分布式 BP 神经网络的短期负荷预测方法[J]. 中国电机工程学报, 2017, 37(17): 4966-4973.
SU Xueneng, LIU Tianqi, CAO Hongqian, et al. A multiple distributed BP neural networks approach for short-term load forecasting based on Hadoop framework[J]. Proceedings of the CSEE, 2017, 37(17): 4966-4973(in Chinese).
- [12] 臧海祥, 许瑞琦, 刘璟璇, 等. 基于多维融合特征和卷积神经网络的多任务用户短期负荷预测[J]. 电力系统自动化, 2023, 47(13): 69-77.
ZANG Haixiang, XU Ruiqi, LIU Jingxuan, et al. Short-term load forecasting for multi-task consumers based on multi-dimensional fusion feature and convolutional neural network[J]. Automation of Electric Power Systems, 2023, 47(13): 69-77(in Chinese).
- [13] 王琛, 王颖, 郑涛, 等. 基于 ResNet-LSTM 网络和注意力机制的综合能源系统多元负荷预测[J]. 电工技术学报, 2022, 37(7): 1789-1799.
WANG Chen, WANG Ying, ZHENG Tao, et al. Multi-energy load forecasting in integrated energy system based on ResNet-LSTM network and attention mechanism[J]. Transactions of China Electrotechnical Society, 2022, 37(7): 1789-1799(in Chinese).
- [14] 刘亚琛, 赵倩. 基于聚类经验模态分解的 CNN-LSTM 超短期电力负荷预测[J]. 电网技术, 2021, 45(11): 4444-4451.
LIU Yahui, ZHAO Qian. Ultra-short-term power load forecasting based on cluster empirical mode decomposition of CNN-LSTM[J]. Power System Technology, 2021, 45(11): 4444-4451(in Chinese).
- [15] 孙庆凯, 王小君, 张义志, 等. 基于 LSTM 和多任务学习的综合能源系统多元负荷预测[J]. 电力系统自动化, 2021, 45(5): 63-70.
SUN Qingkai, WANG Xiaojun, ZHANG Yizhi, et al. Multiple load prediction of integrated energy system based on long short-term memory and multi-task learning[J]. Automation of Electric Power Systems, 2021, 45(5): 63-70(in Chinese).
- [16] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017: 6000-6010.
- [17] 董家富, 万雄, 王岩, 等. 基于 XGB-Transformer 模型的短期电力负荷预测[J]. 电力信息与通信技术, 2023, 21(1): 9-18.
DONG Jiafu, WAN Xiong, WANG Yan, et al. Short-term power load forecasting based on XGB-Transformer model[J]. Electric Power Information and Communication Technology, 2023, 21(1): 9-18(in Chinese).
- [18] 李云松, 张智晟. 基于 GRU-TGTransformer 的综合能源系统多元负荷短期预测[J]. 电力系统保护与控制, 2023, 51(15): 33-41.
LI Yunsong, ZHANG Zhisheng. Multi load short-term forecasting of an integrated energy system based on a GRU-TGTransformer[J]. Power System Protection and Control, 2023, 51(15): 33-41(in Chinese).
- [19] 李练兵, 高国强, 吴伟强, 等. 考虑特征重组与改进 Transformer 的风电功率短期日前预测方法[J]. 电网技术, 2024, 48(4): 1466-1476.
LI Lianbing, GAO Guoqiang, WU Weiqiang, et al. Short-term day-ahead wind power prediction considering feature recombination and improved transformer[J]. Power System Technology, 2024, 48(4): 1466-1476(in Chinese).
- [20] ZHOU Haoyi, ZHANG Shanghang, PENG Jieqi, et al. Informer:

- beyond efficient transformer for long sequence time-series forecasting[C]//Proceedings of the 35th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI, 2021: 11106-11115.
- [21] WU Haixu, XU Jiehui, WANG Jianmin, et al. Autoformer: decomposition transformers with auto-correlation for long-term series forecasting[C]//Proceedings of the 35th International Conference on Neural Information Processing Systems. Curran Associates Inc., 2021: 1717.
- [22] ZHOU Tian, MA Ziqing, WEN Qingsong, et al. FEDformer: frequency enhanced decomposed transformer for long-term series forecasting[C]//Proceedings of the 39th International Conference on Machine Learning. Baltimore, USA: PMLR, 2022: 27268-27286.
- [23] ZHANG Yunhao, YAN Junchi. Crossformer: transformer utilizing cross-dimension dependency for multivariate time series forecasting[C]//The Eleventh International Conference on Learning Representations. Kigali, Rwanda: OpenReview.net, 2023: 10.
- [24] LIU Yong, HU Tengge, ZHANG Haoran, et al. iTransformer: inverted transformers are effective for time series forecasting[C]//The Twelfth International Conference on Learning Representations. Vienna, Austria: OpenReview.net, 2024: 7-11.
- [25] RAMACHANDRAN P, PARMAR N, VASWANI A, et al. Stand-alone self-attention in vision models[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019: 7.
- [26] WU Zhaohua, HUANG N E. Ensemble empirical mode decomposition: a noise-assisted data analysis method[J]. Advances in Adaptive Data Analysis, 2009, 1(1): 1-41.
- [27] TORRES M E, COLOMINAS M A, SCHLOTTHAUER G, et al. A complete ensemble empirical mode decomposition with adaptive noise[C]//2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Prague, Czech Republic: IEEE, 2011: 4144-4147.
- [28] DRAGOMIRETSKIY K, ZOSSO D. Variational mode decomposition [J]. IEEE Transactions on Signal Processing, 2014, 62(3): 531-544.
- [29] LI Heng, LI Zhi, MO Wei. A time varying filter approach for empirical mode decomposition[J]. Signal Processing, 2017, 138: 146-158.
- [30] 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[J]. 电网技术, 2024, 48(6): 2574-2583.
- SHI Zhuojian, RAN Qiwu, XU Fucong. Short-term load forecasting based on aggregated secondary decomposition and informer[J]. Power System Technology, 2024, 48(6): 2574-2583(in Chinese).
- [31] KITAIEV N, KAISER L, LEVSKAYA A. Reformer: the efficient transformer[C]//The 8th International Conference on Learning Representations. Addis Ababa, Ethiopia: OpenReview.net, 2020: 1-11.
- [32] WU Haixu, WU Jialong, XU Jiehui, et al. Flowformer: linearizing transformers with conservation flows[C]//Proceedings of the 39th International Conference on Machine Learning. Baltimore, USA: PMLR, 2022: 24226-24242.



范士雄

在线出版日期: 2025-04-27。

收稿日期: 2024-11-05。

作者简介:

范士雄(1984), 男, 通信作者, 博士, 教授级高级工程师, 硕士生导师, 研究方向为电网调控和人工智能应用技术, E-mail: fanshixiong@epri.sgcc.com.cn;

李东琦(2001), 男, 硕士研究生, 研究方向为人工智能在电力系统中的应用, E-mail: 15726347446@163.com。

(责任编辑 王金芝)

附录 A

表 A1 CNN 主要参数设置
Table A1 Main parameter settings of CNN

参数	$\mathcal{J}^{MAE}$
输入序列长度	96
输出序列长度	12
输入特征数量	5
批大小	32
损失函数	mse
网络层数	2
卷积核个数	16, 8
卷积核大小	3×3, 3×3
滑动步长	1, 1

表 A2 LSTM 主要参数设置
Table A2 Main parameter settings of LSTM

参数	$\mathcal{J}^{MAE}$
输入序列长度	96
输出序列长度	12
输入特征数量	5
批大小	32
损失函数	mse
网络层数	2
隐层大小	128, 128

表 A3 Transformer 主要参数设置
Table A3 Main parameter settings of Transformer

参数	$\mathcal{J}^{MAE}$
输入序列长度	96
输出序列长度	12
输入特征数量	5
批大小	32
损失函数	mse
编码器层数	2
解码器层数	1
注意力头数	4
学习率	0.0001
丢弃率	0.1