

DOI: 10.19783/j.cnki.pspc.221697

基于 GRU-TGTransformer 的综合能源系统多元负荷短期预测

李云松, 张智晟

(青岛大学电气工程学院, 山东 青岛 266071)

摘要: 综合能源系统的多元负荷短期预测, 对系统的优化调度和经济运行至关重要。多元负荷之间耦合关系紧密, Transformer 作为一种完全建立在自注意力机制上的模型, 能很好地分析多元负荷之间的内在联系。传统 Transformer 模型针对自然语言类问题而设计, 难以直接应用于多元负荷预测。为此, 提出一种 GRU-TGTransformer (GRU-Talkinghead-Gated residuals-Transformer)模型。该模型采用门控循环单元代替原有的词嵌入及位置编码环节, 对输入数据进行特征融合, 取得具备相对位置信息的高维特征数据。通过在多头自注意力环节引入交流机制, 提高多头自注意力的表达效果。为进一步强化网络结构, 在残差连接中引入门控单元, 提高模型在时序预测问题上的稳定性。以美国亚利桑那州立大学坦佩校区的综合能源系统为算例, 通过对所提出模型与传统模型之间进行对比分析, 证明所提出的模型具有更高的预测精度。

关键词: 综合能源系统; 多元负荷预测; 短期; 多头交流自注意力; 门控残差

Multi load short-term forecasting of an integrated energy system based on a GRU-TGTransformer

LI Yunsong, ZHANG Zhisheng

(College of Electrical Engineering, Qingdao University, Qingdao 266071, China)

Abstract: The multi load short-term forecast is very important for the optimal scheduling and economic operation of an integrated energy system. There are strong coupling relationships among the multi loads, and the model structure of the transformer is completely based on the self-attention mechanism. This can better analyze the internal relationship between multi loads. It is difficult to directly apply the traditional transformer in multi load forecasting because it is designed for natural language processing problems. For this reason, this paper proposes a GRU-Talkinghead-Gated residuals-Transformer (GRU-TGTransformer) model, which uses gated recurrent units, instead of the original word embedding and position coding links, to extract the input features of input data and obtain high-dimensional feature data with relative position information. By introducing a communication mechanism in the multi-head self-attention, the self-attention expression effect is improved. A gate unit is introduced into the residual connection to improve the stability of the model in time series prediction. This paper uses the integrated energy system of Arizona State University's Tempe campus as an example to prove that the proposed model has higher prediction accuracy than the traditional model.

This work is supported by the National Natural Science Foundation of China (No. 52077108).

Key words: integrated energy system; multi load forecasting; short-term; talking-head self-attention; gated residuals

0 引言

我国的能源行业现已进入了追求高效、低碳的发展新阶段。传统供能系统的各项负荷之间相互独立, 缺乏统一调度, 导致能源利用效率低下, 且严重依赖非可再生能源, 造成了较严重的环境污染问

题^[1]。各类新兴可再生能源并网, 也为电力系统的可靠运行带来了新的挑战。综合能源系统(integrated energy system, IES)作为新一代能源系统, 注重对多项高品质能源的开发与利用^[2], 采用先进的信息管理技术, 协调利用多种类型的负荷, 在显著提高能源利用效率以及系统供能可靠性的同时, 积极响应了低碳、可持续的发展需求^[3-5]。

多元负荷短期预测主要预测未来几小时至几

基金项目: 国家自然科学基金项目资助(52077108)

天的负荷情况,可以依据此提前制定能源配置方案,通过对 IES 的优化、调度,提高能源利用效率以及系统运行可靠性^[6],已成为当前的研究热点。近年来,负荷预测主要采用以指数平滑^[7]、线性回归^[8]为代表的统计学方法,以及以支持向量机^[9-10]、模糊逻辑^[11]、XGBoost(Extreme Gradient Boosting)^[12-13]、深度神经网络^[14-15]、人工神经网络^[16]为代表的机器学习方法。机器学习方法依靠数据驱动,学习数据间的内在关联性,在 IES 负荷预测领域得到了广泛的应用。以往研究中所采用的深度神经网络预测方法各有侧重,差别主要体现在输入特征、特征处理、模型类型、是否进行多元负荷联合预测。文献[17]以历史负荷数据为输入特征,采用长短期记忆网络(long-short term memory, LSTM)进行多元负荷预测。文献[18]以气象、日类型和历史负荷数据为输入特征,采用循环神经网络(recurrent neural networks, RNN)进行预测。文献[19]采用卷积神经网络(convolutional neural network, CNN)进行特征处理,并以历史负荷为输入特征,采用 LSTM 模型以及多元负荷联合预测方式进行负荷预测。

传统时序循环神经网络采用顺序输入结构,该结构虽然可以较好地捕获输入数据的时序特征,但不善于发掘各输入数据之间的长期依赖效应。综合能源系统的各负荷之间耦合性较强,输入数据间的长期依赖效应较为显著,如果直接使用传统时序循环神经网络将难以获得良好的预测精度。注意力机制作为一种应对长期依赖效应的针对性改进,其应用目前在负荷预测领域取得了较好的效果^[20-23],Google 团队舍弃了传统递归模型与卷积模型结构,提出了完全基于自注意力机制的 Transformer 模型^[24],该模型不仅能够实现数据的并行输入,克服传统时序循环神经网络难以发掘长期依赖效应的缺点,同时还具备多头自注意力机制,可以在学习过程中从多个角度来分析输入数据之间的关联性。

综上所述,本文针对传统 Transformer 模型进行了改进,提出一种 GRU-TGTransformer 的 IES 多元负荷短期预测模型。通过门控循环单元(gate recurrent unit, GRU)进行特征处理,然后将特征信息输入到具备多头交流自注意力机制与门控残差环节的 Transformer 模型内,进行电、冷、热负荷的多元负荷联合预测。在算例分析环节,将本文所提出的模型与其他传统预测模型进行对比分析,证明本文所提出模型在该领域的高效性与准确性。

1 数据处理与分析

本文选用美国亚利桑那州立大学坦佩校区 IES

的公开数据集作为研究对象,负荷数据类型包括电、冷、热负荷。气象数据从当地气象站点获得,数据类型包括气温、气压、风强度、露点温度、云量、风向、湿度。本文对历史数据进行了相关性分析,确保充分发挥各模型的预测潜能。

1.1 数据预处理

传感器采集的数据会因外界干扰而出现一定程度的异常。为防止异常数据对相关性及模型训练产生干扰,本文采用拉依达准则对输入数据进行分析处理,如式(1)所示。

$$|x_k - \bar{x}| = |v_k| > 3\sigma \quad (1)$$

式中: x_k 为输入数据; \bar{x} 为 x_k 平均值; v_k 为剩余误差; σ 为标准差。当 v_k 绝对值大于 3σ 时,认为该数据出现异常,进行删除处理。

为消除数据间因量纲不同而导致的数量级差异,本文对数据进行了线性归一化处理,如式(2)所示。

$$x'_k = \frac{x_k - x_{\min}}{x_{\max} - x_{\min}} \quad (2)$$

式中: x'_k 为 x_k 归一化后数值,取值范围为[0,1]; x_{\max} 、 x_{\min} 分别为输入向量中的最大值、最小值。

IES 中,用户的用能情况与多种因素相关,如果盲目选取与输出目标低相关性的数据作为输入特征,将徒增模型计算量,影响模型的训练效率及预测精度^[25]。为了能对各负荷及气候之间的相关性有更明确的认识,本文采用 Pearson 相关系数来分析各数据之间的相关性,计算公式为

$$\rho_{x,y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \quad (3)$$

式中: X 、 Y 为特征向量; $\rho_{x,y}$ 为 Pearson 相关系数,取值范围为[-1,1], $\rho_{x,y}$ 的绝对值越接近 1,表明 X 、 Y 两者之间线性相关性越强; $\text{cov}(X,Y)$ 为 X 、 Y 的协方差; σ_X 、 σ_Y 分别为 X 、 Y 的标准差。

通过计算获得电、冷、热负荷以及气象数据之间的 Pearson 相关系数并取绝对值,可得到相关性热力图如图 1 所示。

考虑到相关系数绝对值达 0.5 及以上的数据之间关联性较高,从气象因素中选取气温、气压、露点温度作为输入特征。

从图 1 中可以看出,多元负荷之间具备明显的关联性,这表明了多元负荷联合预测潜在的优越性,以及从耦合关系出发,进一步改进模型多头自注意力机制这一思路的正确性。

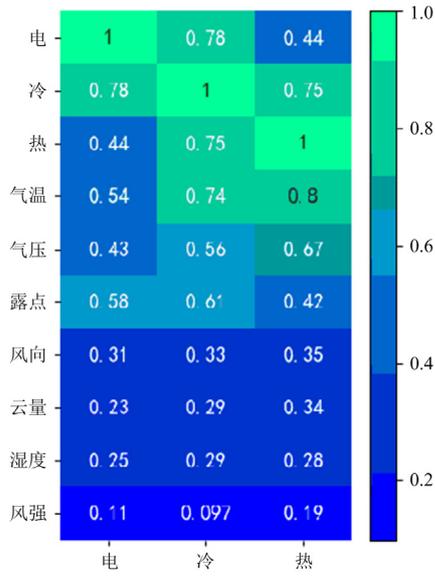


图 1 输入数据相关性分析

Fig. 1 Input data correlation analysis

1.2 多元负荷自相关性分析

IES 用户用能具有较强的周期性规律, 通过分析历史负荷数据, 可以很好地预测负荷的未来走向。本文依据相关性分析对历史负荷数据进行了筛选, 意在确保模型的计算效率与精度, 让模型更好地分析用户未来的用能走向。本文选取了一周内关键时间节点的历史负荷数据进行自相关性分析, 计算公式为

$$\gamma_k = \frac{\text{cov}(X_k, X_{k-t})}{\sigma_X^2} \quad (4)$$

式中: γ_k 为 k 时间点负荷与历史数据的自相关系数; X_k 为在 k 时间点的负荷数据; X_{k-t} 为与 k 距离 t 时间长度的负荷数据; $\text{cov}(X_k, X_{k-t})$ 为 X_k 与 X_{k-t} 的协方差; σ_X^2 为输入向量的方差。自相关性分析结果如表 1 所示。

表 1 历史负荷数据自相关性分析

Table 1 Autocorrelation analysis of historical load

时间距离/h	相关性/%		
	电负荷	冷负荷	热负荷
1	97.93	99.41	98.79
2	95.39	98.01	96.64
23	90.59	97.50	93.77
24	91.46	97.90	94.21
25	90.09	97.40	93.47
48	85.17	96.46	90.55
72	82.70	95.46	87.83
96	80.62	94.13	85.24
168	87.72	93.47	83.62

从自相关性分析中可以看出, 当前时刻负荷与历史负荷具有较强的关联性。总体而言, 当前时刻的冷、热负荷数据与历史数据关联性随时间递减。当前时刻的电负荷数据除了与短时间距离的电负荷数据关联性较强之外, 还与上一周的历史数据具有较强的关联性, 该关联性甚至高于距离两天时间点的历史数据。上述规律较好地符合了现实情况中 IES 用户的用能周期性规律。

综上所述, 本文选取与输出目标相关度较高的气象以及历史负荷数据, 共同组成输入向量特征集, 其结构如图 2 所示。

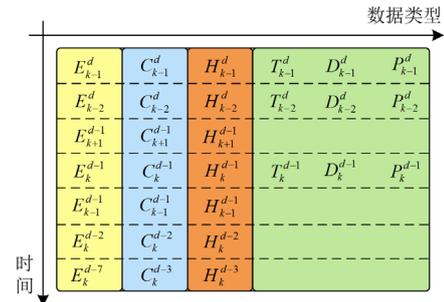


图 2 输入特征集合结构

Fig. 2 Input feature set structure

图 2 中: d 为数据所处天数; k 为数据所处时间点; E 、 C 、 H 分别为电负荷、冷负荷、热负荷; T 、 D 、 P 分别为气温、露点温度、气压。

2 GRU-TGTransformer 网络模型

本文提出了一种具备特征处理环节的 GRU-TGTransformer 多元负荷联合预测模型, 模型结构如图 3 所示。该模型由输入层、 N 个 TGTransformer 编码层、输出层组成。传统 Transformer 的输入层由词嵌入与位置编码环节组成, 其目的是更好地处理自然语言类问题。本文采用 GRU 代替了传统 Transformer 模型在输入层的词嵌入与位置编码环节, 对输入向量进行特征处理, 以更好地针对 IES 多元负荷短期预测问题。TGTransformer 编码层用于计算输入特征向量的自注意力表达, 输出层负责输出模型的预测结果。

2.1 输入层

传统的 Transformer 模型不采用时序输入结构, 这导致了模型无法获取输入负荷以及气象因素的相对位置关系。传统 Transformer 的解决方案是通过在输入层设置正弦绝对位置编码, 引入输入数据之间的位置信息, 在经过线性变换和点积等运算后, 该方法易导致负荷与气象的相对位置信息出现丢

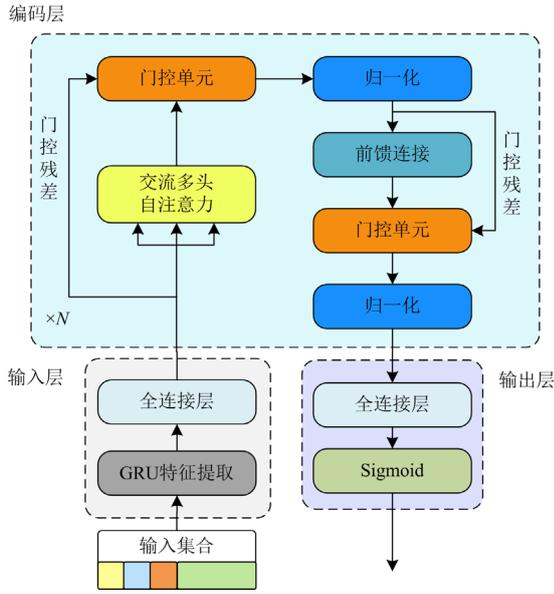


图3 GRU-TGTransformer 模型结构

Fig. 3 GRU-TGTransformer model structure

失。本文采用特征融合处理方式来对此环节进行改进，常用于特征融合处理的神经网络为 CNN 以及基于 RNN 改进后的 LSTM 以及 GRU 模型^[23]。CNN 模型广泛应用于图像处理领域，通过对输入数据的卷积处理，提取数据隐含的关键信息。LSTM、GRU 通过记忆门结构，控制对输入信息及其相对位置的更新与提取，不仅更适用于学习 IES 多元负荷预测这类时序问题，同时也能对 Transformer 的时序位置信息丢失问题作出较好的改进。GRU 相对 LSTM 网络结构更加简洁，待优化参数更少，具有更快的收敛速度，在对 IES 多元负荷数据的复杂耦合信息提取过程中，表现出了更高的融合效率。

综上，本文采用 GRU 及全连接层代替传统 Transformer 模型中的词嵌入与位置编码环节，获取关于负荷与气象的更高纬度特征，以具备更详尽位置信息的高维特征数据作为编码层的输入，弥补传统 Transformer 的固有结构缺陷，提高模型的最终预测效果，计算公式如式(5)。

$$X = W^i F_{GRU}(x) + b^i \quad (5)$$

式中： x 为负荷与气象数据构成的输入向量； F_{GRU} 为 GRU 门控循环计算函数； W^i 、 b^i 分别为全连接层的参数矩阵与偏置向量； X 为输入层输出的高维度负荷与气象数据向量，用作编码器的输入。

2.2 TGTransformer 编码层

Transformer 的多头自注意力机制可通过突出重点特征高效捕捉输入特征向量之间的依赖关系，对于具有强耦合性的 IES 负荷具有很好的针对性。

为了进一步发挥这一优势，本文在多个自注意力头之间引入相互映射，实现了自注意力头间的相互交流，强化自注意力机制的输出效果。

为进一步提高 Transformer 模型在时序预测问题上的稳定性，本文在残差连接环节加入了门控单元，提升了模型的参数优化效率以及预测精度。

2.2.1 多头自注意力机制

注意力机制模拟了人类在面对大量信息时，忽略低价值信息，集中注意力关注重点信息的行为。Transformer 的自注意力机制通过对负荷与气象数据进行注意力计算，进而筛选出其中用于负荷预测的关键信息，计算公式为

$$X \cdot [W^q, W^k, W^v] = [Q, K, V] \quad (6)$$

$$A(Q, K, V) = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right) \cdot V \quad (7)$$

式中： W^q 、 W^k 、 W^v 为参数矩阵，用于计算 Q 、 K 、 V 向量， Q 、 K 、 V 分别对应了查询向量、键向量以及值向量，通过计算查询值与键值的点积 $Q \cdot K^T$ 来反映负荷与气象数据的内在关联性以及相互关联性，并在经过指数归一化后，与值向量 V 相乘进而得到 X 的注意力表达 $A(Q, K, V)$ ； softmax 为指数归一化函数，用于计算自注意力权重值； d_k 为 K 的维度数， $Q \cdot K^T$ 通过除以 $\sqrt{d_k}$ 实现缩放计算，防止 $Q \cdot K^T$ 因数值过大而导致 softmax 出现梯度消失情况。

上述自注意力机制仅创造对输入向量 X 的一组映射，为单头自注意力机制。多头自注意力机制则在此基础上更进一步，通过创造 h 组 X 的相互独立的映射，建立多方面的自注意力表达，计算公式为

$$A_i = A(Q_i, K_i, V_i) \quad (8)$$

$$A_{MH}(X) = W^{mh} \cdot [A_1, \dots, A_h] \quad (9)$$

式中： i 为自注意力头编号； A_i 为第 i 号的自注意力表达； Q_i 、 K_i 、 V_i 为 X 的第 i 号自注意力头的三组映射，其计算方式同式(6)； W^{mh} 为多头参数矩阵； h 为自注意力头数； $A_{MH}(X)$ 为 X 多头自注意力表达。

2.2.2 交流多头自注意力机制

多头自注意力结构中的各头之间相互独立，不存在信息交流，不利于进一步分析 IES 多元负荷的复杂耦合特性以及负荷及气象数据之间潜在的关联性。本文在 softmax 指数归一化操作前后加入了多个注意力头之间的线性映射，分别用于完成对注意力键值对、自注意力权重值的多头相互交流，计算公式及过程如式(10)~式(14)。

$$[j_1, \dots, j_h] = \mathbf{W}^j \left[\frac{\mathbf{Q}_1 \cdot \mathbf{K}_1^T}{\sqrt{d_k}}, \dots, \frac{\mathbf{Q}_h \cdot \mathbf{K}_h^T}{\sqrt{d_k}} \right] \quad (10)$$

$$J_i = \text{softmax}(j_i(\mathbf{Q}_i, \mathbf{K}_i)) \quad (11)$$

$$[L_1, \dots, L_h] = \mathbf{W}^l [J_1, \dots, J_h] \quad (12)$$

$$A_{Ti} = L_i \cdot \mathbf{V}_i \quad (13)$$

$$A_{TMH}(X) = \mathbf{W}^{mh} \cdot [A_{T1}, \dots, A_{Th}] \quad (14)$$

式中: i 为自注意力头编号; $\mathbf{Q}_i \cdot \mathbf{K}_i^T / \sqrt{d_k}$ 构成键值对 j_i , 对负荷与气象数据关联性的进行初步计算; \mathbf{W}^j 为交流参数矩阵; J_i 为自注意力权重, 由 j 经指数归一化后获得, J_i 为模型从不同的多个角度理解输入数据的重要性观点, 在通过权重交流参数矩阵 \mathbf{W}^l 计算后, 可以完成对多个角度的交流认识, 交流后的自注意力权重为 L_i ; 为了强化自注意力表达效果, 将 L_i 与值向量 \mathbf{V}_i 进行点积运算, 进而得到具备交流机制的自注意力表达结果 A_{Ti} 并通过汇总矩阵 \mathbf{W}^{mh} 计算后完成对注意力结论的输出; A_{TMH} 为具备交流机制的多头自注意力计算结果, 计算原理同式(9)。

2.2.3 门控残差

残差连接通过将输出更改为输入向量与输入向量非线性变换的叠加, 有效改善了深层神经网络训练过程中出现的梯度消失和网络退化等问题。Transformer 通过引入残差连接保证了模型的深度学习能力, 但这也对整体网络结构造成了一定程度的负面影响, 不利于模型在时序预测领域的应用。如果能对残差环节加以控制, 就可以在提高 Transformer 时序适用能力的同时, 确保模型的深度学习能力, 从而提高模型的整体预测效果。

GRU 解决了 RNN 模型中的梯度消失问题, 同时还保留了 RNN 的长期记忆能力, 其实现原理是通过引入门控单元来对上一时刻的输入加以控制。受到 GRU 原理的启发, 本文在 Transformer 的两个残差环节引入门控单元来控制残差连接, 模型多头自注意力层的门控残差实现方式如式(15)一式(18)。

$$z = \sigma(\mathbf{W}_z \mathbf{X} + \mathbf{U}_z A_{MH}) \quad (15)$$

$$r = \sigma(\mathbf{W}_r \mathbf{X} + \mathbf{U}_r A_{MH}) \quad (16)$$

$$\tilde{A}_{MH} = \tanh(\mathbf{W} \mathbf{X} + r \odot \mathbf{U} A_{MH}) \quad (17)$$

$$A'_{MH} = \tilde{A}_{MH} \odot (1 - z) + z \odot A_{MH} \quad (18)$$

式中: σ 为 sigmoid 激活函数; \odot 为 Hadamard 乘积, 定义为矩阵之间对应位置元素相乘; \mathbf{W}_z 、 \mathbf{U}_z 为计算更新门的参数矩阵; \mathbf{W}_r 、 \mathbf{U}_r 为计算重置门的参数矩阵; \mathbf{W} 、 \mathbf{U} 为计算门控结论的参数矩阵, 上述参数矩阵随每次训练更新寻优; z 、 r 分别为更新

门与重置门, 更新门决定有多少残差连接前的自注意力数据需要被输入到残差连接之后, 重置门则决定要遗忘多少残差连接前的注意力数据; \tilde{A}_{MH} 为 A_{MH} 新的记忆表达, 使用了重置门存储的 A_{MH} 信息, 以及残差连接所需的 \mathbf{X} 信息; A'_{MH} 为最终输出的门控残差结果。后续的前馈连接门控残差计算结构与该结构相同。

2.3 输出层

输出层由一个全连接层以及 sigmoid 压缩函数组成, 计算公式为

$$Y = \text{sigmoid}(\mathbf{W}^o X_D + b^o) \quad (19)$$

式中: Y 为输出层输出结果; X_D 为编码层输出结果; \mathbf{W}^o 、 b^o 为全连接层参数矩阵与偏置量。

3 算例分析

3.1 实验设置

本文采用 Python 3.9.7 作为编程语言, 使用 Pytorch 1.10.0 为框架搭建模型。硬件配置为 Intel Core i5-8265U CPU, 8 G 内存。

数据结构层面, 选用美国亚利桑那州立大学坦佩校区 IES 数据, 以 2020 年 1 月 1 日到 2020 年 9 月 21 日的数据为训练集, 2020 年 9 月 22 日到 2020 年 9 月 28 日的数据为测试集。考虑到 IES 多元负荷短期预测所需数据的主要时间单位为 1 h, 此处的多元负荷以及气象数据分辨率均设为 1 h。

本文提出的 GRU-TGTransformer 模型包含输入层、编码层、输出层。输入层中的 GRU 计算环节的隐藏层维度为 525, 全连接层维度为 525; 编码层层数 N 为 3, 交谈多头自注意力层维度为 525, 自注意力头数为 8, 前馈连接维度为 2048; 输出层全连接层维度为 525。在每层添加 Dropout 环节, 防止出现过拟合, 数值为 0.1。实验采用分批次输入, 批次大小为 24, 训练轮次为 175 次。优化算法采用 Adam(adaptive moment estimation)算法, 学习率为 1×10^{-5} 。损失函数采用 Huber 损失函数。

3.2 仿真评价标准

本文采用绝对百分比误差(mean absolute percentage error, MAPE)、加权平均绝对百分比误差(weighted mean absolute percentage error, WMPAE)、均方根误差(root mean square error, RMSE)、训练时间为评价指标, 并采用误差概率密度分布统计分析模型的预测性能。MAPE 从百分比角度描述误差均值, WMAPE 用于反映多项负荷预测结果的总体误差, RMSE 用于分析预测中极端异常数据的影响。误差概率密度分布可统计预测结果的误差分布区间, 反映模型的预测稳定性。式(20)一式(22)为相关

计算表达式。

$$E_{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \quad (20)$$

$$E_{WMAPE} = \alpha_e E_{MAPE_e} + \alpha_c E_{MAPE_c} + \alpha_h E_{MAPE_h} \quad (21)$$

$$E_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (22)$$

式中： E_{MAPE} 、 E_{WMAPE} 、 E_{RMSE} 分别为绝对百分比误差、加权平均绝对百分比误差、均方根误差； y_i 为实际值； \hat{y}_i 为预测值； n 为样本数； α_e 、 α_c 、 α_h 分别为电、冷、热负荷 WMAPE 求和权重。从负荷历史均值来看，电、冷、热负荷分别占总历史负荷的 40%、40%、20%。为凸显高占比负荷在整个 IES 内的重要性，此处的 α_e 、 α_c 、 α_h 分别取 0.4、0.4、0.2。

3.3 仿真结果分析

3.3.1 多元负荷联合预测与单负荷预测对比分析

采用 GRU-TGTransformer 模型，分别进行多元负荷联合预测与单负荷预测，对多元负荷联合预测的潜在优势进行验证。相关数据结果见表 2。

表 2 多元负荷联合预测与单负荷预测对比分析

Table 2 Comparative analysis of multi-load joint forecasting and single load forecasting

预测模型	MAPE/%			训练用时/s
	电负荷	冷负荷	热负荷	
多元负荷联合预测	1.71	1.36	2.34	4121.42
单负荷预测	1.75	1.40	2.39	11 950.64

从表 2 中可以看出，多元负荷联合预测方法的误差略小于单负荷预测，且训练用时降低了 65.51%，明显低于单负荷预测。主要原因是单负荷预测需要对不同类型的负荷分别建模求解，而多元负荷联合预测能一次性输出多元负荷数据，很大程度上提高了计算效率。对多元负荷之间内在联系的分析利用也在一定程度上提高了预测精度。这证实了多元负荷联合预测的高效性，在下述对比分析环节，各模型将均采用多元负荷联合预测方式。

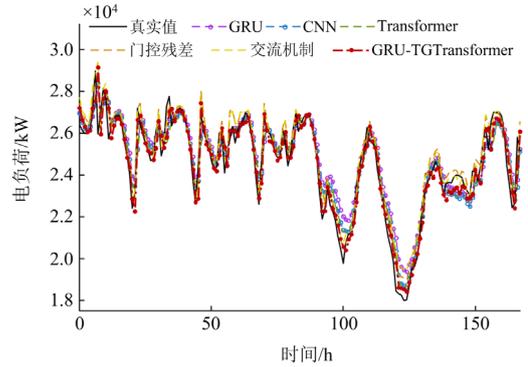
3.3.2 模型预测结果对比分析

为验证 GRU-TGTransformer 在 IES 多元负荷短期预测问题上的优越性，本文采用在时序预测领域应用成熟的 GRU、CNN 模型进行对比分析。同时，为验证所采取改进方案的有效性，本文将传统 Transformer 模型、仅具有多头交流自注意力机制的 Transformer 模型、仅具有门控残差机制的 Transformer 模型与 GRU-TGTransformer 模型的预测结果进行对比分析，所得到的各模型的预测结果如图 4 所示，

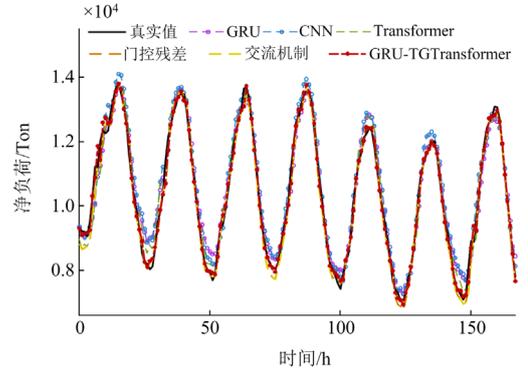
评估数据如表 3 所示。电、冷、热负荷的单位关系为 $1 \text{ kW} = 0.284 \text{ Ton} = 3.4 \times 10^{-3} \text{ MMBtu/h}$ 。

从统计结果中可以看出，对比于传统的 GRU、CNN 模型，Transformer 具有更高的预测精度，体现了注意力机制在分析 IES 多元负荷内在耦合联系方面的优势，模型的 WMAPE 分别降低了 17.86%、12.15%。

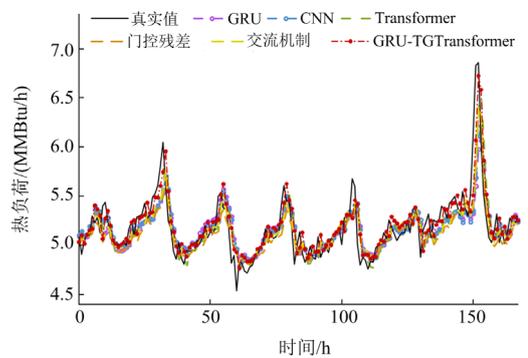
门控残差单元的引入使得 Transformer 模型的预测精度有了进一步的提高，WMAPE 降低了 17.00%，验证了残差控制环节对网络结构改进的有效性。多



(a) 电负荷预测结果



(b) 冷负荷预测结果



(c) 热负荷预测结果

图 4 预测结果对比

Fig. 4 Comparison of forecasting results

表 3 不同模型预测性能对比

Table 3 Comparison of forecasting performance of different models

预测模型	MAPE/(电/冷/热)	WMAPE/%	RMSE(电/冷/热)	训练用时/s
GRU	2.95/3.53/2.46	3.08	2312.13/1098.83/0.53	3131.11
CNN	2.35/3.65/2.43	2.88	1913.39/1120.27/0.51	3390.63
Transformer	2.38/2.72/2.45	2.53	1950.59/937.00/0.49	4352.18
门控残差	2.16/1.87/2.42	2.10	1718.47/640.01/0.50	5584.69
交流机制	2.07/1.77/2.29	1.99	1770.34/628.93/0.46	4634.59
GRU-TGTransformer	1.71/1.36/2.30	1.70	1566.66/509.13/0.44	3901.42

头注意力之间引入交流机制,也提高了模型的预测能力,通过强化注意力机制的输出表达效果,使模型的 WMAPE 降低了 21.34%,证明了该改进方法的有效性。

同时采用了交流机制以及门控残差单元的 GRU-TGTransformer 模型的电负荷与冷负荷的预测性能有了较为明显的提升,原因在于该地区为热带沙漠气候,热负荷需求率低,模型将注意力集中在了数量级更大的电负荷与冷负荷上,进而提高模型的总体预测精度。对比传统 Transformer 模型,电负荷和冷负荷的 MAPE 分别降低了 28.15%、50.00%,RMSE 分别降低了 19.68%、45.66%,WMAPE 降低了 32.81%,总体预测精度提升较为显著。GRU-TGTransformer 的网络结构更加复杂,单次训练所需时间也更长,但交流机制与门控残差环节分别从自注意力与网络结构层面改进了模型,显著提高了训练效率,从而使模型能更快到达训练最优值,缩短了总的训练时长。相对于传统 Transformer 模型,其训练用时降低了 10.36%,处于可接受范围内,实现了模型学习效率与预测性能两方面的平衡。

3.3.3 模型预测误差结果分析

采用综合负荷预测误差的误差概率密度分布图,从预测误差的角度对模型的预测性能进行进一步分析。综合负荷预测误差通过电、冷、热负荷预测误差分别乘以各自求和权重后相加获得,统计结果如图 5 所示。

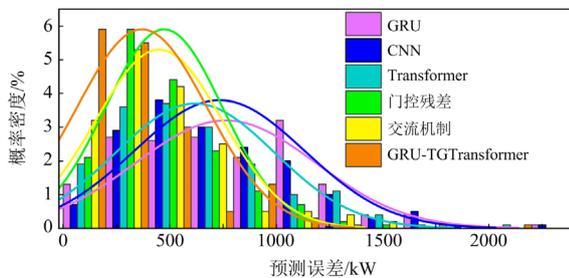


图 5 综合负荷预测误差分布对比

Fig. 5 Comparison of error distribution of comprehensive load forecasting

从各模型的误差分布对比中可以看出,对比其他模型,本文所提模型的预测误差主要分布在低误差区间范围内,且分布更加集中。这说明本文所提模型更难出现大误差的预测失误,其预测结果相较于其他模型更具合理性以及稳定性。

4 结论

本文针对 IES 多元负荷短期预测问题,提出了一种基于 GRU-TGTransformer 模型的预测方法。该模型通过 GRU 提取高维数据特征,采用多头交流自注意力机制提取输入特征的重点信息,以门控残差的方式稳定模型网络结构。通过算例分析得到以下结论:

1) 多元负荷联合预测方式在 IES 多元负荷短期预测问题上的表现优于单负荷预测,能在确保预测精度的同时显著提高运算效率。

2) 自注意力机制通过对 IES 多元负荷耦合关系的分析处理,有效发挥了 Transformer 模型的预测能力,获得了相较于传统时序预测模型更高的预测精度,但运算速度相对缓慢。

3) 多头交流自注意力机制以及门控残差方法均能在不同程度上提高 Transformer 模型在 IES 多元负荷短期预测问题上的预测能力,且两者共同作用能显著提高模型的预测精度、训练效率和预测稳定性。

参考文献

- [1] CHENG Y, ZHANG N, LU Z, et al. Planning multiple energy systems toward low-carbon society: a decentralized approach[J]. IEEE Transactions on Smart Grid, 2019, 10(5): 4859-4869.
- [2] 胡泉, 尚策, 陈东文, 等. 考虑能量品质的区域综合能源系统多目标规划方法[J]. 电力系统自动化, 2019, 43(19): 22-31, 139.
HU Xiao, SHANG Ce, CHEN Dongwen, et al. Multi-objective planning method for regional integrated energy systems considering energy quality[J]. Automation of Electric Power Systems, 2019, 43(19): 22-31, 139.

- [3] 余晓丹, 徐宪东, 陈硕翼, 等. 综合能源系统与能源互联网综述[J]. 电工技术学报, 2016, 31(1): 1-13.
YU Xiaodan, XU Xiaodong, CHEN Shuoyi, et al. A brief review to integrated energy system and energy internet[J]. Transactions of China Electrotechnical Society, 2016, 31(1): 1-13.
- [4] 屈小云, 吴鸣, 李奇, 等. 多能互补综合能源系统综合评价研究进展综述[J]. 中国电力, 2021, 54(11): 153-163.
QU Xiaoyun, WU Ming, LI Qi, et al. Review on comprehensive evaluation of multi-energy complementary integrated energy systems[J]. Electric Power, 2021, 54(11): 153-163.
- [5] 何桂雄, 金璐, 李克成, 等. 基于改进 DaNN 的综合能源系统多能负荷预测[J]. 电力工程技术, 2021, 40(6): 25-33.
HE Guixiong, JIN Lu, LI Kecheng, et al. Multiple energy load forecasting of integrated energy system based on improved DaNN[J]. Electric Power Engineering Technology, 2021, 40(6): 25-33.
- [6] 王明军, 穆云飞, 孟宪君, 等. 考虑热能输运动态特性的电-热综合能源系统优化调度方法[J]. 电网技术, 2020, 44(1): 132-140.
WANG Mingjun, MU Yunfei, MENG Xianjun, et al. Optimal scheduling method for integrated electro-thermal energy system considering heat transmission dynamic characteristics[J]. Power System Technology, 2020, 44(1): 132-140.
- [7] LAOUAFI A, MORDJAOUI M, LAOUAFI F, et al. Daily peak electricity demand forecasting based on an adaptive hybrid two-stage methodology[J]. International Journal of Electrical Power & Energy Systems, 2016, 77: 136-144.
- [8] 严通煜, 杨迪珊, 项康利, 等. 基于时间分解技术的中远期逐时负荷预测模型[J]. 电力系统保护与控制, 2019, 47(6): 110-117.
YAN Tongyu, YANG Dishan, XIANG Kangli, et al. Mid-long term hourly load forecasting model based on time decomposition[J]. Power System Protection and Control, 2019, 47(6): 110-117.
- [9] 朱清智, 董泽, 马宁. 基于即时学习算法的短期负荷预测方法[J]. 电力系统保护与控制, 2020, 48(7): 92-98.
ZHU Qingzhi, DONG Ze, MA Ning. Forecasting of short-term power based on just-in-time learning[J]. Power System Protection and Control, 2020, 48(7): 92-98.
- [10] 杨海柱, 田馥铭, 张鹏, 等. 基于 CEEMD-FE 和 AOA-LSSVM 的短期电力负荷预测[J]. 电力系统保护与控制, 2022, 50(13): 126-133.
YANG Haizhu, TIAN Fuming, ZHANG Peng, et al. Short-term load forecasting based on CEEMD-FE-AOA-LSSVM[J]. Power System Protection and Control, 2022, 50(13): 126-133.
- [11] LI C, TANG M, ZHANG G, et al. A hybrid short-term building electrical load forecasting model combining the periodic pattern, fuzzy system, and wavelet transform[J]. International Journal of Fuzzy Systems, 2020, 22(1): 156-171.
- [12] 姜建, 刘海琼, 李衡, 等. 基于 XGBoost 的配电网线路峰值负荷预测方法[J]. 电力系统保护与控制, 2021, 49(16): 119-127.
JIANG Jian, LIU Haiqiong, LI Heng, et al. Peak load forecasting method of distribution network lines based on XGBoost[J]. Power System Protection and Control, 2021, 49(16): 119-127.
- [13] 孙超, 吕奇, 朱思瞳, 等. 基于双层 XGBoost 算法考虑多特征影响的超短期电力负荷预测[J]. 高电压技术, 2021, 47(8): 2885-2898.
SUN Chao, LÜ Qi, ZHU Sitong, et al. Ultra-short-term power load forecasting based on two-layer XGBoost algorithm considering the influence of multiple features[J]. High Voltage Engineering, 2021, 47(8): 2885-2898.
- [14] 彭曙蓉, 黄士峻, 李彬, 等. 基于深度学习分位数回归模型的充电桩负荷预测[J]. 电力系统保护与控制, 2020, 48(2): 44-50.
PENG Shurong, HUANG Shijun, LI Bin, et al. Charging pile load prediction based on deep learning quantile regression model[J]. Power System Protection and Control, 2020, 48(2): 44-50.
- [15] 王剑锋, 郑剑, 王旭东, 等. 基于改进深度信念网络的短期电力负荷预测方法[J]. 电力系统及其自动化学报, 2021, 33(10): 125-130.
WANG Jianfeng, ZHENG Jian, WANG Xudong, et al. Short-term power load forecasting method based on improved deep belief network[J]. Proceedings of the CSU-EPSA, 2021, 33(10): 125-130.
- [16] WANG L, LEE E W M, YUEN R K K. Novel dynamic forecasting model for building cooling loads combining an artificial neural network and an ensemble approach[J]. Applied Energy, 2018, 228: 1740-1753.
- [17] 朱刘柱, 王绪利, 马静, 等. 基于小波包分解与循环神经网络的综合能源系统短期负荷预测[J]. 电力建设, 2020, 41(12): 133-140.
ZHU Liuzhu, WANG Xuli, MA Jing, et al. Short-term load forecast of integrated energy system based on wavelet packet decomposition and recurrent neural network[J]. Electric Power Construction, 2020, 41(12): 133-140.

- [18] 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.
- [19] ZHU R, GUO W, GONG X. Short-term load forecasting for CCHP systems considering the correlation between heating, gas and electrical loads based on deep learning[J]. Energies, 2019, 12(17): 3308.
- [20] 任建吉, 位慧慧, 邹卓霖, 等. 基于 CNN-BiLSTM-Attention 的超短期电力负荷预测[J]. 电力系统保护与控制, 2022, 50(8): 108-116.
REN Jianji, WEI Huihui, ZOU Zhuolin, et al. Ultra-short-term power load forecasting based on CNN-BiLSTM-Attention[J]. Power System Protection and Control, 2022, 50(8): 108-116.
- [21] 田浩含, 张智晟, 于道林. 基于改进 LSTM 的区域综合能源系统多元负荷短期预测研究[J]. 电力系统及其自动化学报, 2021, 33(9): 130-137.
TIAN Haohan, ZHANG Zhisheng, YU Daolin. Research on multi-load short-term forecasting of regional integrated energy system based on improved LSTM[J]. Proceedings of the CSU-EPSA, 2021, 33(9): 130-137.
- [22] 乔石, 王磊, 张鹏超, 等. 基于模态分解及注意力机制长短期网络的短期负荷预测[J]. 电网技术, 2022, 46(10): 3940-3951.
QIAO Shi, WANG Lei, ZHANG Pengchao, et al. Short-term load forecasting by long- and short-term temporal networks with attention based on modal decomposition[J]. Power System Technology, 2022, 46(10): 3940-3951.
- [23] 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.
- [24] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C] // Advances in Neural Information Processing Systems, 2017, Long Beach, USA.
- [25] 朱继忠, 董瀚江, 李盛林, 等. 数据驱动的综合能源系统负荷预测综述[J]. 中国电机工程学报, 2021, 41(23): 7905-7923.
ZHU Jizhong, DONG Hanjiang, LI Shenglin, et al. Review of data-driven load forecasting for integrated energy system[J]. Proceedings of the CSEE, 2021, 41(23): 7905-7923.

收稿日期: 2022-10-25; 修回日期: 2023-01-19

作者简介:

李云松(1999—), 男, 硕士研究生, 研究方向为区域综合能源系统多元负荷预测; E-mail: 1012413565@qq.com

张智晟(1975—), 男, 通信作者, 博士, 教授, 研究方向为电力系统和综合能源系统负荷预测、经济调度等。
E-mail: slnzszs@126.com

(编辑 魏小丽)