Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 214 (2022) 312–320

Procedia Computer Science

www.elsevier.com/locate/procedia

9th International Conference on Information Technology and Quantitative Management

Total electricity consumption forecasting based on Transformer time series models

Xuerong Li[a,d], Yiqiang Zhong[b], Wei Shang[a, d, *], Xun Zhang[a], Baoguo Shan[c], Xiang Wang[c]

[a]*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China*
[b] *Institute of Urban Environment,Chinese Academy of Sciences, Xiamen 361021, China*
[c]*State Grid Energy Research Institute CO., LTD, Beijing 102209, China*
[d] *MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation at UCAS*

Abstract

The total electricity consumption (TEC) reflects the operation condition of the national economy, and the prediction of the total electricity consumption can help track the economic development trend, as well as provide insights for macro policy making. Nowadays, novel neural networks provide a new perspective to predict the total electricity consumption. In this paper, a time series forecasting method based on Transformer model, Trans-T2V model, is proposed and applied to TEC forecasting. Transformer is the first network structure that completely relies on self-attention to calculate input and output. In this paper, the Time2vec method is used to improve the existing Transformer model, as embedding the month sequence more efficiently in the Transformer model. By comparing with the existing Transformer models and other intelligent algorithm models, the robustness and superiority of the proposed method framework are verified, and the highest accuracy reaches 97.36%. The method presented in this paper provides valuable insights in the field of time series prediction.

Keywords: Total electricity consumption; Transformer models; Forecasting framework; Time2vec embedding

* Corresponding author. Tel.: +010-82541284.
E-mail address: shangwei@amss.ac.cn
This work is supported by the science and technology project of State Grid Corporation of China (Project Code: 1400-202157207A-0-0-00).

10.1016/j.procs.2022.11.180

1. Introduction

Since electric power is closely related to industrial production, business activities and residents living, the electricity data could generally reflect the operation condition of the national economy. It is of great value for the government to formulate macro-control policies and promote governance capacity to look forward economic and social development through electricity data.

Among the statistical indicators of electricity, total electricity consumption (TEC) is one of the most comprehensive and basic indicators to reflect the electricity consumption situation of a country or region. Specifically, TEC comprises the primary, secondary and tertiary industries, and other fields of electricity consumption, including industrial electricity, agricultural electricity, commercial electricity, residential electricity, public facilities electricity, etc. The important value of TEC lies in that it can reflect the operation condition of the national economy. Accurate prediction of TEC can help track the trend of economic development and provide insights for macro-control policy making.

However, the prediction of TEC is a difficult task, and there are few researches in related fields. TEC includes various fields of electricity consumption, and the patterns of production electricity consumption and residential electricity consumption are different. Therefore, it is difficult to distinguish the complex factors influencing each other during forecasting, which adds uncertainty to the prediction results. With the development of artificial intelligence, novel neural networks that can capture the latent sequential patterns of TEC provides a new idea for the prediction of TEC. At present, the existing researches in related fields mostly focus on the prediction of electricity load [1-3], but there are still few models that can effectively predict TEC by using frontier intelligent models.

In this paper, a time series prediction method based on Transformer, Trans-T2V model, is proposed and applied to TEC prediction, which achieves satisficed performances. As a cutting-edge deep learning model, Transformer adopts self-attention mechanism to consider the position information between input sequences, and learns the underlying trend of time series through encoder and decoder modules, as to obtain high prediction accuracy. Transformer is the first network model that fully relies on self-attention to compute representations of inputs and outputs without using sequential alignment of recurrent neural networks (RNN) or swift model of convolutional neural networks (CNN). The proposed Trans-T2V model integrates Time2vec method to improve the existing Transformer model, embedding the month sequence into the Transformer model more efficiently and achieving better forecasting performance. In the out-of-sample test period, the highest accuracy has reached 97.36%. The robustness and superiority of the proposed framework are further verified by comparing with more benchmark models and multiple time windows.

The main contributions of this paper are in the following aspects: First, it is the first attempt to apply Transformer model to the forecasting of macro electricity indicators such as TEC, and provide empirical results for forecasting performance through comparative experiments. This paper verifies that the Transformer-based models have better ability to learn the trend of TEC series compared with traditional intelligent models. Second, we extend existing Transformer model specific to time series forecasting. The Time2VEC method is embedded in the network structure of the existing Transformer model to incorporate dynamic month sequence, thus improve the prediction ability for time series. The empirical results also verify the validity of the model framework, which enlightens on the practical applications of time series prediction.

The remaining parts of this paper are arranged as follows: Section 2 reviews relevant literature on electricity data prediction methods and Transformer models; Section 3 introduces the proposed framework of TEC prediction and experimental design; Section 4 shows the data description statistics and empirical results; Section 5 is the conclusion and discussion.

2. Literature review

In terms of electricity consumption analysis and prediction, there are many forecasting methods proposed by scholars worldwide, which can be roughly divided into classical forecasting methods, traditional forecasting methods and modern intelligent forecasting methods. Among them, the classical prediction method includes elastic coefficient method, the calculation of the capacity of the expansion of the industry, etc. The data frequency of traditional prediction methods is mainly annual and monthly. The commonly used models include time series models [1], regression models [2] gray prediction models [3], etc. With the development of data processing ability, modern intelligent models have been widely used in electricity consumption forecasting with monthly and daily basis data. Scholars have employed various models including neural network prediction method [4], support vector machine [5], chaos theory prediction method [6], also include other combination forecast method, etc.

In recent years, big data technology has gradually been applied in the research of electricity consumption prediction. Guo et al. (2020) develop a short-term load forecasting model of multi-scale CNN-LSTM hybrid neural network considering the real-time electricity price [7]. The model of [8] found that for every 1 degree increase in temperature, peak electricity consumption would increase by 0.45% to 4.6%. Also using deep learning models, Bedi and Toshniwal (2020) propose a deep learning based hybrid approach which firstly implements Variational Mode Decomposition (VMD) and Autoencoder models to extract meaningful sub-signals/features from the data [9]. Ayub et al. (2020) applied the GRU-CNN model to predict the daily electricity consumption of ISO-NE data set, which improved the prediction accuracy by 7% compared with the SOTA benchmark model [10].

Transformer model, proposed by Google in 2017[11] has achieved excellent performance in many tasks in natural language processing and computer vision, which has also attracted great interests in the time series field [12]. In the following, we mainly review its application and improvement in time series forecasting. [13] proposes convolutional self-attention by using causal convolution to generate queries and keys in the self-attention layer. Informer uses a ProbSparse self-attention mechanism, which reduces time complexity and memory usage to O(log), and effectively captures the individual long-term correlation of inputs and outputs of long series and improves the forecasting performance [14]. Autoformer devises a simple seasonal-trend decomposition architecture with an auto-correlation mechanism working as an attention module. The auto-correlation module is not a traditional attention module [15]. Instead, it measures the time-delay similarity between inputs signal and aggregate the top-k similar sub-series to produce the output with a reduced complexity of O(log).

In summary, the existing literature applies various intelligent algorithms and forecasting electricity data using historical datasets. At present, attention mechanism-based neural network models, especially Transformer models, are hardly used for TEC prediction. This paper fills this gap and further extend existing method. We put forward an improved method of Transformer model for time series prediction, and achieve better forecasting performance.

3. Methodology

This section introduces the proposed framework of time series method based on Transformer. Firstly, several Transformer models used in this paper are introduced, then the improved model Trans-T2V proposed in this paper is presented, and then the experimental design is introduced.

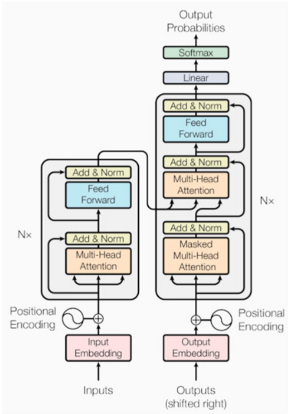
3.1. Transformer-based models

Transformer is a model for sequence-to-sequence tasks, which was first proposed in the paper [11] in 2017. Unlike recursive neural network or convolutional neural network, Transformer is the first representation that fully relies on self-attention to compute input and output. The network structure of Transformer is divided into Encoder and Decoder, as shown in Fig.1. (a). In order to account for oositional information between input sequences, positional encoding is used, which can be mapped using sine and cosine functions at different frequencies. The obtained location information is added with the embedding matrix to obtain the sequence matrix with location information.

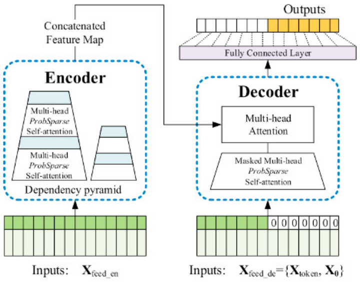
Informer proposes three improvements based on Transformer: ProbSparse self-attention mechanism with O(log) time complexity; The self-attention distillation mechanism was proposed to shorten the input sequence length of each layer, and the calculation amount and storage would naturally decrease when the sequence length was short. The self-attention mechanism highlights the dominant attention by cutting the input of cascade layer in half, and effectively processed the excessively long input sequence. A generative decoder mechanism is proposed to obtain the result in one step instead of step-by-step when predicting the sequence (including inference phase), and the prediction time complexity is directly reduced from O(N) to O(1). The network structure of Informer is shown in Fig.1. (b).

Time series decomposition is a common method of time series analysis, which can decompose time series into several underlying time patterns, such as periodic component and trend component. In the prediction task, due to the unknowability of the future, the input is usually decomposed first, and then each component is predicted separately.

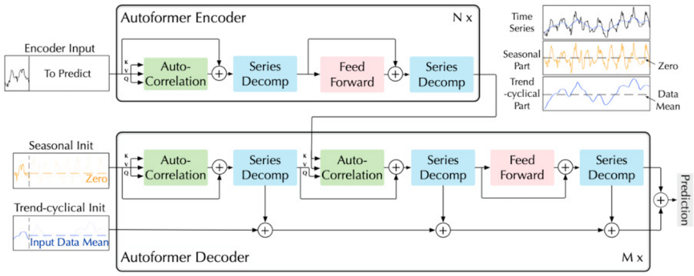
But this limits the forecast to decomposed series and ignores the interactions between components in the long-term future. To solve the above problems, Autoformer adopts the deep decomposition architecture. In the process of prediction, it separates the trend component and the period component from the hidden variables gradually, and realizes the progressive decomposition. The network structure of Autoformer is shown in Fig.1. (c). The model alternates the prediction result optimization and sequence decomposition, which can promote each other.



(a) Transformer model architecture
(Li et al., 2019)



(b) Informer model architecture
(Zhou et al., 2020)



(c) Autoformer model architecture
(Wu et al., 2021)

Fig. 1. Transformer-based model architectures

3.2. Trans-T2V forecasting model

The network structure of the TEC prediction model of Trans-T2V proposed in this paper is shown in Fig.2. Among them, the TEC sequence and month sequence are merged as the input of encoder module, and the month sequence is embedded by Time2VEC. The month sequence is embedded using Time2vec as input to the decoder module. Position embedding adopts the method in Transformer. Value embedding adopts a fully connected linear layer network.

Time2vec method are proposed in [16], which transforms dynamic data into static embedding vector. The calculation method is as follows:

$$T2V(\tau)[i] = \begin{cases} \omega_i \tau + \varphi_i \, , if \; i = 0 \\ F(\omega_i \tau + \varphi_i), if \; 1 \le i \le k \end{cases} \tag{1}$$

where k is Time2vec dimension, τ is original time series variable, F are periodic activation functions, ω and φ are a group of parameters to be learned, that is, the weight coefficients in the Time2vec embedding layer. In the experiment, F is selected as a sinusoidal function in order to enable the algorithm to capture the periodic behaviour in the dataset. Meanwhile, the linear term (corresponding to i =0) represents the aperiodic process of time and is used to capture aperiodic patterns in the time input.

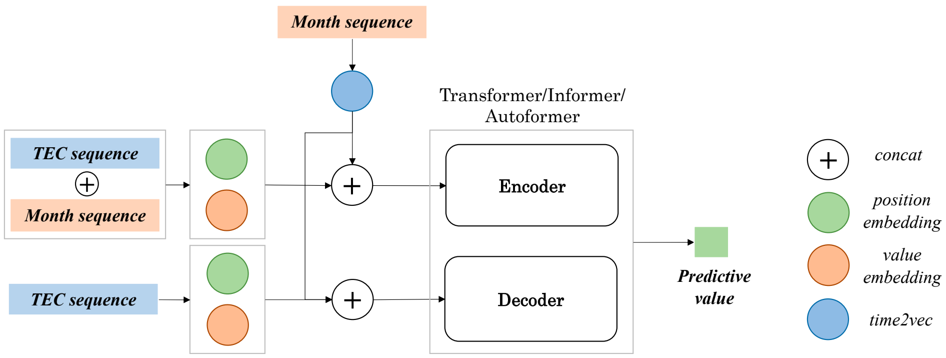


Fig. 2. Transformer-T2V model architecture

The encoder and decoder modules of the model can integrate the architecture of Transformer, Informer or Autoformer respectively. Since Autoformer can better deconstruct the seasonal and trend terms of time series and is more in accordance with the characteristics of TEC indicators this paper mainly uses the encoder and decoder architecture of Autoformer. In the comparative experiment, we also provide the comparisons of the forecasting performance of three Transformer-based models

3.3. Experimental design

In order to verify the prediction ability of the Trans-T2V model proposed in this paper, we use the monthly TEC in China from January 2009 to December 2020 for empirical study, and design a series of comparative experiments. We use encoder and decoder modules of Autoformer to verify the prediction ability of Trans-T2V model, and mainly carry out the following two types of comparison experiment: 1) Horizontal comparison of different models in the same training and testing period is conducted to compare the forecasting performance of Trans-T2V model, existing Transformer models and other intelligent algorithms. 2) Vertical comparison of different training and testing periods is conducted to verify the robustness of the performance of Trans-T2V model, and to exclude the comparison bias caused by random factors of the data set.

The models and parameter settings used in the experiment are shown in Table 1, where d_{model} represents the embedding dimensions of the input sequence; *dropout* represents the early stop probability of dropout. During model training, each neural unit is retained with probability p (dropout drop rate is 1-p). *d_ff* represents the number of channels in the middle layer of the convolution of encoder and decoder modules. *N_heads* indicates the number of heads within the multi-head attention mechanism; *e_layers* and *d_layers* denote the number of layers of encoder and decoder, respectively.

Table 1. Model specifications and parameters

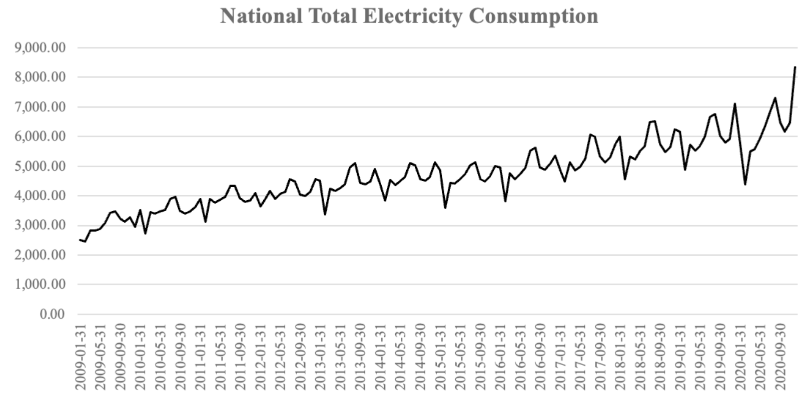
Label	Model specifications	Parameters
Transformer-T2V	Autofomer model embedding with time2vec	
Transformer	Transformer model without time2vec	Parameters of Transformer-based models: d_{model} =64; $dropout$ = 0.05; $d\_ff = 16$; $e\_layers = 1; d\_layers = 1; N\_heads = 1$
Informer	Informer model without time2vec	
Autoformer	Autoformer model without time2vec	

MLP	Neural network with multilayer perceptron	Parameters of MLP, SVR and XGBoost: default values of Python Scikit-learn library
SVR	Support vector regression	
XGBoost	Extreme gradient boosting	

4. Empirical results

4.1. Data description

The prediction target of the model is the monthly TEC in China, and the sample period is from January 2009 to December 2020. The line chart and descriptive statistics of the series are shown in Figure 3. As shown in the figure, TEC series has an obvious seasonal and increasing trend over the years, and the network structure of Auto former integrates the extraction of seasonal and trend terms, which can theoretically predict TEC more effectively.



Mean	4697.453
Standard error	89.07936
Median	4565
Mode	4563
Std.	1068.952
Variance	1142659
Kurtosis	0.167377
Skewness	0.387706
Observations	144
Max.	8338
Min.	2461.499

Fig. 3. Descriptive statistics of National Total Electricity Consumption

4.2. Horizontal comparison of different models

In order to verify the forecasting performance of the Trans-T2V model proposed in this paper, the prediction accuracies of different models are firstly compared horizontally in the same training and testing period, and the results are shown in Table 2. The training period of all models is from January 2009 to December 2018, and the forecast period is from January 2019 to December 2020. Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE) are used for evaluation criteria in Table 3, and the calculations are as follows:

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{\hat{x}_t - x_t}{x_t}\right| \tag{2}$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|\hat{x}_t - x_t\right| \tag{3}$$

where $\hat{x}_t$ and x_t represent the predicted and true values of the forecasting model, respectively.

Table 2. Horizontal comparison of different models

	Transformer-T2V	Transformer	Informer	Autoformer	MLP	SVR	XGBoost
MAPE	**0.0560**	0.0589	0.0645	0.0645	0.099	0.2216	0.0700
MSE	**347.87**	393.51	395.37	395.37	612.17	1438.96	451.96

Note: The training period of the models in the table is from January 2009 to December 2018, and the forecast period is from January 2019 to December 2020; The model with the best performance is shown in bold

The results in the table show that the prediction accuracy of the proposed Trans-T2V model reaches 94.4%, which is higher than the three existing Transformer models and the classical intelligent algorithms such as MLP, SVR and XGBoost. In addition, the three Transformer models also achieve better accuracy than the classical intelligent algorithms. This result indicates that Transformer-based models have a good prediction ability for TEC, and the improvement strategy proposed in this paper can make accurate prediction according to the characteristics of TEC time series more effectively. The reason is that it captures the seasonal and trend terms of TEC series and incorporates them into dynamic position embedding.

4.3. Vertical comparison of different periods

In order to further verify the robustness of the Trans-T2V model and exclude the comparison bias caused by random factors in the data set, vertical comparison of the forecasting performance in different prediction periods and training periods is conducted in this paper. The results are shown in Table 3. The first column of the table lists the four forecast periods, the corresponding training period is from January 2009 to the one month before the start of the forecast period. MAPE and MAE are still used to evaluate the prediction accuracy. The last two columns are the average values of accuracies over all forecasting periods, reflecting the overall performance of the model.

Table 3. Vertical comparison of different periods

	2011.1～2011.4		2013.1～2013.9		2017.1～2018.4		2019.1～2020.12		Mean	
	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE
Transformer-T2V	**0.0264**	**97.66**	**0.0720**	**308.26**	**0.0286**	**154.32**	**0.0560**	**347.87**	**0.0457**	**227.02**
Transformer	0.0689	243.62	0.1002	437.87	0.0664	351.20	0.0589	393.51	0.0736	356.55
Informer	0.0587	203.39	0.1010	450.90	0.0433	235.76	0.0645	395.37	0.0668	321.35
Autoformer	0.0326	108.76	0.0735	320.18	0.0306	165.68	0.0645	395.37	0.0503	247.49
MLP	0.1017	381.06	0.082	363.33	0.0826	444.42	0.099	612.17	0.0913	450.24
SVR	0.0920	333.68	0.152	679.03	0.160	879.62	0.2226	1438.96	0.1564	832.82
XGBoost	0.0945	338.85	0.0871	381.59	0.0553	304.61	0.0700	451.96	0.0767	369.25

Note: The first column of the table is the forecast period, which corresponds the corresponding training period is from January 2009 to the one month before the start of the forecast period; the last two columns are the average of the accuracies of all forecast periods; the model with the best performance is shown in bold

The results in the table show that the Trans-T2V model achieves the highest prediction accuracy among all models in multiple training-prediction experiments. For the January-April 2011 forecast period, the Trans-T2V model achieves the highest accuracy of 97.36%. The average accuracy of Trans-T2V model in all forecast periods is 95.43%, which is highest among all existing Transformer models and classical intelligent algorithms. The results of the above comparative experiments show that the Tran-T2V model proposed in this paper has effectively captures the seasonal and trend terms of the TEC series, and this advantage is not easily affected by the randomness of the data set with good robustness.

5. Conclusions and discussions

The total electricity consumption reflects the operation condition of the national economy. The accurate prediction of TEC is of great significance for the government to track the trend of economic development and formulate macro-control policies. TEC is difficult to predict due to its various components, while big data and deep learning models provide new perspectives for predicting TEC. In this study, the existing Transformer model are improved based on the characteristics of TEC series. The Time2VEC method is used to capture the seasonal and trend terms of TEC, and the month sequence is embedded into the Transformer model more efficiently, which achieves better forecasting performance.

The empirical results show that the prediction accuracy of the proposed Trans-T2V model is higher than that of the three existing Transformer models (Transformer, Informer, Autoformer) and the classical intelligent algorithms such as MLP, SVR and XGBoost. In the out-of-sample test period, the highest prediction accuracy has reached 97.36%. Vertical comparative experiments in multiple time windows show that the forecasting superiority of Trans-T2V model is not easily affected by the randomness of the dataset, which further validates the robustness and superiority of the proposed method framework.

The main contribution of this paper is providing an early attempt to apply Transformer model in macro electricity indicators forecasting such as TEC, and verifying the performance through comparative experiments. In addition, this paper also proposes an effective strategy to improve the existing Transformer model. The Time2vec method is embedded in the network structure to improve the prediction ability of time series.

Limitations of this paper and possible future research directions include: the current Trans-T2V model only uses the historical data of TEC, and does not consider other exogenous variables that may affect TEC, such as industrial production, consumption, and other macroeconomic variables. Further research can be done to embed other exogenous variables and their month sequence more effectively in the Transformer model. In addition, high-frequency TEC data of weekly and daily basis can also be further integrated into Transformer model, which is expected to further improve the prediction accuracy.

Acknowledgements

This work is supported by the science and technology project of State Grid Corporation of China (Project Code: 1400-202157207A-0-0-00).

References

[1] Hussain A, Rahman M, Memon JA. Forecasting electricity consumption in Pakistan: The way forward. Energy Policy 2016;90:73-80.
[2] Feng Y, Ryan SM. Day-ahead hourly electricity load modeling by functional regression. Applied Energy 2016;170:455-465.
[3] Xu W, Gu R, Liu Y, Dai Y. Forecasting energy consumption using a new GM–ARMA model based on HP filter: The case of Guangdong province of China. Economic Modelling 2015;45:127-135.
[4] Kunwar, NY, et al. Area-load based pricing in DSM through ANN and heuristic scheduling. Smart Grid 2013; 4:1275-1281.
[5] Cheng Q, Yan Y, Liu S, et al. Particle filter-based electricity load prediction for grid-connected microgrid day-ahead scheduling. Energies 2020;13:6489.
[6] Liu ZJ, Yang HM and Lai MY, Electricity price forecasting model based on chaos theory. In: 2005 International Power Engineering Conference, 2005, p. 1-9.
[7] Guo X, Zhao Q, Zheng D, et al. A short-term load forecasting model of multi-scale CNN-LSTM hybrid neural network considering the real-time electricity price. Energy Reports 2020;6:1046-1053.
[8] Santamouris M, Cartalis C, Synnefa A, et al. On the impact of urban heat island and global warming on the power demand and electricity consumption of buildings—A review. Energy & Buildings 2015;98:119-124.
[9] Bedi J, Toshniwal D. Energy load time-series forecast using decomposition and autoencoder integrated memory network. Applied Soft Computing 2020;93:106390.
[10] Ayub N, Irfan M, Awais M, et al. Big data analytics for short and medium term electricity load forecasting using AI techniques ensembler. Energies 2020;13:5193.
[11] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). New York: Curran Associates Inc.; 2017, p. 6000–6010.
[12] Wu B, Wang L, Zeng Y R . Interpretable wind speed prediction with multivariate time series and temporal fusion transformers. Energy 2022;252:123990.
[13] Li SY, Jin XY, Xuan Y, et al. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. New York: Curran Associates Inc.; 2019, p. 5243–5253.

[14] Zhou H, Zhang S, Peng J, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting. 2020. DOI: https://doi.org/10.48550/arXiv.2012.07436
[15] Wu H, Xu J, Wang J, et al. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. 2021. DOI: https://doi.org/10.48550/arXiv.2106.13008.
[16] Kazemi SM, Goel R, Eghbali S, et al. Time2Vec: Learning a vector representation of time. 2019. DOI: https://doi.org/10.48550/arXiv.1907.05321.