Enhanced dynamic and static graph attention network for power time series prediction

Yun Dong, Xuhua Ai, Zhaoli Chen, Yuan Yin, Xixiang Zhang*
Information Center, China Southern Power Grid Guangxi Power Grid Co., Ltd., Nanning, Guangxi, China

ABSTRACT

Accurate power time series prediction is crucial for stable operation and optimized scheduling of power grids rapidly developing and integrating increasing renewable energy sources. Traditional prediction models often neglect the spatiotemporal characteristics of power grid data, resulting in inadequate accuracy. To address this, we propose an enhanced dynamic and static graph attention network model for power time series prediction. Experimental results on the Solar Energy and Electricity datasets demonstrate superior fitness values of 99.01 and 99.88 after 30 and 16 iterations, respectively. The model achieves an RMSE value of 2.14% within 100 seconds on the Solar Energy dataset and a CORR value of 0.982 after 30 cycles. In practical application, the method consistently exhibits a low RSE value (within a fluctuation range from 0.021 to 0.035) as the Layer parameter increases. The proposed method offers high prediction accuracy, providing valuable insights for power system management, operation, and scheduling, thereby enhancing the safety, stability, and economic operation of power systems.

Keywords: Power time series prediction, enhanced dynamic and static graph attention networks, dual view deep neural network

1. INTRODUCTION

With the continuous growth of global energy demand and the widespread access to renewable energy, the stability and efficiency of power systems are facing unprecedented challenges. As a key link in the operation and management of power systems, power time series forecasting is of great significance for optimizing power resource allocation, improving grid stability, and reducing operating costs. Power threats may come from a variety of factors such as extreme weather events, equipment aging, system failures, or malicious attacks, all of which may lead to power outages or reduced grid efficiency[1,2]. Therefore, establishing a power time series prediction model that can effectively predict and respond to these threats has great practical application value for taking measures in advance and preventing them. Traditional forecasting methods, such as linear regression and time series analysis, often fail to capture the complexity and nonlinear characteristics of power data, resulting in insufficient forecasting accuracy. In recent years, with the rapid development of artificial intelligence and machine learning technologies, data-driven prediction models have been widely used in power systems[3]. In particular, graph neural networks (GNNs) have demonstrated excellent performance in power system state estimation, load forecasting, and fault detection due to their natural advantages in processing graph-structured data[4,5]. However, existing deep learning models still have some limitations in dealing with the dynamics and complexity of power systems, especially in the specific task of predicting power threats. For example, CNN has limited ability to capture long-distance dependencies, while RNN and LSTM are prone to gradient vanishing or exploding problems when processing long sequence data. In addition, existing models often ignore the interrelationships between different nodes in the power system and the multidimensional characteristics of power data. In order to solve these problems, many scholars have studied power time series prediction models.

To improve the prediction accuracy of short-term position marginal prices in the electricity market, Yang et al. proposed a price prediction model that combines a spectral convolution network and a three-branch network structure. The attention mechanism is used to allocate important weights between different nodes and time slots. It was found that the accuracy and robustness of the model were significantly superior[6]. VMD is used to decompose complex time series; GRU processes and learns the time domain features extracted by CNN to obtain the final prediction result. Finally, it was proved that the comprehensive performance of the model is significantly better, and it has higher feasibility and

*zhang_xx.xt@gx.csg.cn

International Conference on Optics, Electronics, and Communication Engineering (OECE 2024), edited by Yang Yue, Proc. of SPIE Vol. 13395, 133952N · © 2024 SPIE · 0277-786X · Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.3049156

accuracy[7]. Scholars such as He proposed an end-to-end model based on a cross-attention tree perception network to achieve accurate prediction of multivariate time series data. In the process, a tree structure and embedding method are first established to extract and learn the cross-features between different time series. Through experiments on different data sets, it was found that the prediction accuracy of this model is high, and it has very significant effectiveness and robustness[8]. To accurately predict vehicle movement flow within the city, He et al. proposed a traffic prediction method based on graph attention spatiotemporal network (GASTN). In this framework, a structural recurrent neural network is introduced to model global spatial relationships. The results show that the prediction performance of the model is significantly better, and the introduced global collaborative local learning strategy can significantly improve the model prediction performance[9]. Although deep learning provides a new perspective for power time series forecasting, existing research still faces some challenges. First, the high dimensionality and complexity of power data require the model to capture multi-scale features and patterns. Second, the dynamics of the power system require the model to be able to adapt to rapidly changing environments and conditions. In addition, the existing models still need to be improved in terms of interpretability and generalization capabilities. The existence of these problems limits the application effect of deep learning models in actual power systems.

To overcome these challenges, a power time series prediction model based on an improved dynamic-static graph attention network (DS-GAT) is proposed experimentally. By introducing the dynamic and static graph attention network, the model can capture the relationship between different nodes in the power system; at the same time, the attention mechanism is used to enhance the recognition of key features. At the same time, by combining the dual-view deep neural network to analyze the dynamic changes and static patterns of the time series, the model can more comprehensively understand the characteristics of power data and improve the accuracy of prediction; ultimately, the interpretability of the model is improved, making the prediction results easier to understand and trust.

The innovations of the proposed model can be divided into two points: First, the model can capture the static correlation between nodes within the power system. Through the dynamic graph attention network, the model can update and pay attention to the data at important time points in real-time, thereby improving the response speed to emergencies. Second, it can adapt to the dynamic characteristics of the power system that change over time. Through the static graph attention network, the topological structure of the power grid can be deeply analyzed to identify key nodes and potential weak links. It is expected that the method constructed through the experiment will provide more accurate prediction results for the stable operation of the power system.

The remainder of this article is organized as follows. We review the related work in Section 2. Section 3 describes the methods proposed in this paper. Section 4 reports the experimental results. Finally, we conclude the paper in Section 5.

2. RELATED WORK

Power time series prediction is an important foundation for the operation and management of power systems and is of great significance for the safety, stability, and economic operation of power systems. In recent years, with the rapid development of deep learning technology, various neural network models have become a hot research topic. Simeuno et al. proposed a station photovoltaic prediction model on the grounds of graph convolutional long short-term memory network to accurately predict the power generation of solar energy and ensure the safe operation of the power grid. This method starts from production data as the basic data and predicts virtual weather stations by using photovoltaic systems. The results showed that the predictive performance of this model is significantly better than traditional models, with a higher accuracy[10]. Cui et al. proposed a framework on the grounds of multi-scale GNNs to predict the correlation between multivariate time series. In addition, to achieve information dissemination between different time scales, an improved sampling strategy was designed to adjust the steps and effectively fuse information. It was found through testing on different datasets that the prediction accuracy of the model was significantly improved by 20%[11]. Peng's team proposed a method on the grounds of recursive neural networks integrating multiple graph operators for effective prediction of multivariate time series. Meanwhile, to perfectly depict the relationships between different multivariate time series, an ensemble model is introduced to capture information. The results showed that the MAE, MAPE, and RMSE values of the model are significantly better than traditional models, and can achieve accurate prediction of complex multivariate time series[12].

Meanwhile, some scholars have proposed to combine other methods with neural network models and conduct research. Chen and Jiang proposed a multivariate time series prediction model on the grounds of a dynamic spatiotemporal graph attention network to effectively predict and monitor sintering temperature. During the experiment, the attention module of the spatiotemporal map was used to extract temperature features. The application results showed that the model has high prediction accuracy for temperature and can achieve accurate control[13]. Researchers such as Dong and Han proposed a joint network capture method that integrates nonlinear graph attention and temporal attractiveness to accurately capture the correlation between geosensing time series. It evaluated the performance of the model using three real datasets from different fields. It was found that the performance of this model is significantly better than the other eight models, and it can have high prediction accuracy for ground sensing time series[14]. The teams of Wang and Feng proposed a spatiotemporal feature extraction model on the grounds of auxiliary information deep transformation network to address the incompleteness of feature extraction in space for multivariate time series data. The model had added three gate control units internally and captured the amount of information through auxiliary information gate control. Experiments on two datasets have shown that the performance of this model is far superior to traditional models[15].

In summary, the relevant models of graph attention networks have shown more significant advantages in many fields and have achieved good results in various application scenarios. However, there are still some challenges and shortcomings in existing research, such as the effective extraction of dynamic correlations and static features of power systems by models. To further improve the performance of the model, a power time series prediction model on the grounds of an improved dynamic and static GNN is proposed in the experiment. During the experiment, a dual perspective deep neural network and a dynamic static GNN will be combined to provide more accurate prediction results for the scheduling, operation, and management of the power system through continuous in-depth research.

3. POWER TIME SERIES PREDICTION MODEL ON THE IMPROVED DS-GAT

3.1 A power time series prediction model on the grounds of TVDNN algorithm

A Two-View Deep Neural Network (TVDNN) based power time series prediction model is proposed in the experiment. TVDNN may refer to processing data from two different perspectives or dimensions at the same time, such as the dynamic characteristics and static characteristics of time series. In power time series forecasting, such a network may predict future power consumption or active power loss by learning patterns in historical data. In addition, TVDNN may also include some regularization techniques or data enhancement strategies to ensure the stability and reliability of the prediction and reduce the uncertainty caused by incomplete data. Compared with traditional methods, TVDNN models are usually able to capture more complex nonlinear relationships, which is difficult to achieve in traditional linear or simple nonlinear models. The TVDNN model uses real induction motor proportional data as an example during operation. The model mainly consists of three modules: power timing missing processing module, power timing data step-by-step aggregation calculation module, and dual view deep neural network prediction module. The specific architecture of the TVDNN model is shown in Figure 1.

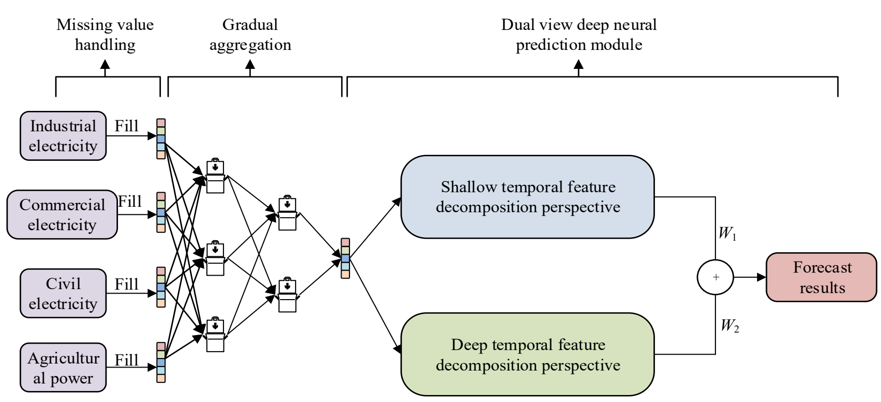


Figure 1. Specific architecture of TVDNN model.

In Figure 1, the TVDNN model mainly includes shallow temporal feature decomposition perspectives and deep temporal feature mining perspectives. Due to the non-repeatability of the data collected in the smart grid, using the missing power data for analysis when there is insufficient data redundancy may lead to significant deviations between the conclusions obtained and the actual patterns. Therefore, the experiment uses the power timing data missing processing module in the TVDNN model to correct and fill in the relevant data of active power missing, to ensure the integrity of the power timing data[16]. The experiment sets the size of the date window for historical data to and uses the daytime data moving average method to fill in missing data by using relevant data from historical dates meanwhile in the date window. This can fit the current missing data and the fluctuation trend of historical data at some time points. The calculation formula is shown in equation (1).

$$val_f = \frac{\sum_{h=1}^{T_h} val(h)}{T_h} \tag{1}$$

In equation (1), val_f represents the filling value. $val(h)$ represents the historical data of the same time two days before the current date. In addition, for the power time series data of active power, at time T, the power data value has a strong dependence on the data from the previous time. The use of horizontal short-term month on month anomaly detection methods can obtain the periodic characteristics followed by the data within the nearest time window T. The experiment sets T to n and compares the fill value from the previous moment with the value from the past $n-1$ moments. If it is greater than the threshold, it will add 1 to $count$. Otherwise, if $count$ is set to $nums$, the data point can be considered as an outlier. The comparison formula is shown in equation (2).

$$H = \begin{cases} 0, count\left(\sum_{i=0}^{n}\left(val_n - val_i\right) > t_d\right) < nums \\ 1, count\left(\sum_{i=0}^{n}\left(val_n - val_i\right) > t_d\right) \geq nums \end{cases}, t_d = \min(\max - avg, avg - \min) \tag{2}$$

In equation (2), H represents the short-term month on month horizontal anomaly detection result (0 represents no anomaly. 1 represents anomaly). val_i represents the data at time i ($i \in 1, \cdots, n-1$). t_d represents the dynamic threshold. After supplementing the missing active power, due to the different power loads generated by different industries, the experiment weighted and averaged the active power of different industries at the substation level, city level, and province level. This is to gradually aggregate and calculate the proportion of induction motors. The proportion of induction motors that aggregate electricity consumption from different industries to substations, from different substations to various prefecture level cities, and from various prefecture level cities to provinces is calculated using equation (3).

$$\begin{cases} km_s = \dfrac{k_1 P_1 + k_2 P_2 + \cdots + k_n P_n}{P_1 + P_2 + \cdots + P_n} \\ km_c = \dfrac{km_{s_1} + km_{s_2} + \cdots + km_{s_n}}{n} \\ km_p = \dfrac{km_{c_1} + km_{c_2} + + km_{c_m}}{m} \end{cases} \tag{3}$$

In equation (3), km_s represents the proportion of induction motors in the substation. n represents the type of industry electricity consumption under the substation. k_n represents the corresponding industry weight. P_n represents the active power of the corresponding industry. km_c represents the proportion of induction motors in prefecture level cities. km_p represents the proportion of induction motors in the province. Due to the trend, periodicity, and potential impact of holidays and long-term historical data on the proportion data of induction motors, a dual perspective deep neural network model for induction motor proportion prediction was designed in the experiment. The prediction results and the calculation of the proportion of induction motors under the perspective of shallow temporal feature decomposition are shown in equation (4).

$$\begin{cases} Y(t) = P(t)w_1 + L(t)w_2 \\ w_1 + w_2 = 1 \\ P(t) = g(t) + s(t) + h(t) + \varepsilon(t) \end{cases} \tag{4}$$

In equation (4), t represents the time. $P(t)$ and $L(t)$ represent predicted values. w_1 and w_2 represent weights. $Y(t)$ represents the final prediction result. $g(t)$ represents the logistic regression growth module. $s(t)$ represents periodic modules. $h(t)$ represents the irregular fluctuation module. To explore the nonlinear growth trend changes hidden in the power time series, the segmented logistic regression growth algorithm was used in the experiment. $g(t)$ represents trend changes, $s(t)$ represents periodic changes, and the specific calculation is shown in equation (5).

$$\begin{cases} g(t) = \dfrac{C(t)}{1+e^{\left(-\left(k+a(t)^T\right)\delta\right)\left(t-(m+a(t)^T)Y\right)}} \\ a(t) = a_j(t) = \begin{cases} 1, t \geq s_j \\ 0, othersize \end{cases} \\ s(t) = \sum_{n=1}^{N}\left(a_n \cos(\dfrac{2\pi nt}{P}) + b_n \sin(\dfrac{2\pi nt}{P}) \right) \end{cases} \tag{5}$$

In equation (5), $C(t)$ represents the saturation value of the curve. $\left(k+a(t)^T\right)$ represents the curve growth rate. $\delta = \delta_j$ represents the change at time t_j . $m+a(t)^T$ represents the midpoint of the curve. γ represents the offset of smoothing processing at time s_j . s_j represents the moment of the j -th change point. P represents the period. a_n, b_n represent the smoothing parameter. N represents the Fourier order. Then, the experiment used Long Short-Term Memory (LSTM) network to achieve deep temporal feature mining perspective, adding forget gates, input gates, and output gates in the traditional structure to select information and preserve important information. The improved structure of the LSTM network is shown in Figure 2.

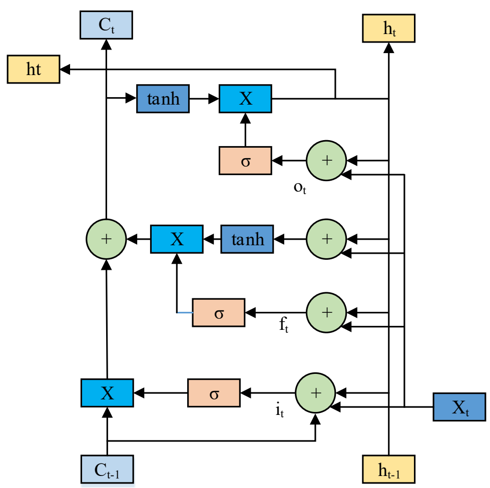


Figure 2. The improved structure of LSTM network.

3.2 A power time series and threat prediction model integrating TVDNN and IDSGAT

When predicting and analyzing the relationship between power time series data, it was found that many models did not explicitly model the dependency relationship between time series data that threatened the power system. Meanwhile, existing explicit modeling methods only consider the static relationship between power time series data, and do not take into account the dynamic changes between time series data and the threats they face[17]. To this end, the experiment proposed a power time series prediction model based on dynamic and static graph attention networks (Improving Dynamic and Static Graph Attention Networks, IDSGAT). The overall framework of the DSGAT model is shown in Figure 3.

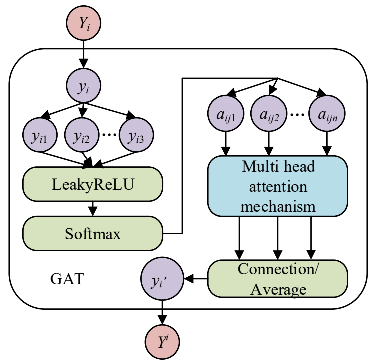
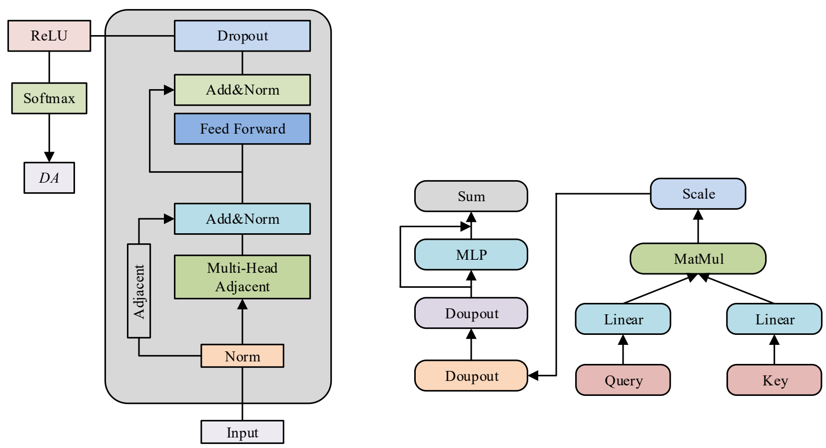


Figure 3. The specific architecture of the DS-GAT model.

In Figure 3, when the DS-GAT model is applied to reality, the time series data not only contains long-term features but also has constantly changing short-term features. In view of this, the experiment introduces a dynamic and static graph learning module, and adaptively learns the mapping function of the dynamic graph and the adjacency matrix of the static graph through the dynamic and static graph learning layer to capture the hidden relationship between different data. The structure of the dynamic graph learning layer is shown in Figure 4.



(a) Dynamic Graph Learning Layer (b) Multi-Head Adjacent

Figure 4. Dynamic graph learning layer structure.

In Figure 4, in multivariate time series data, the data between variables exhibits dynamic dependencies. The dynamic graph learning layer considers both node information and captures the dynamic spatiotemporal correlations between different variables. In addition, to explore more information between different variables, the experiment combined the idea of Transformer based multi head attention to construct the dynamic matrix. The expression of attention matrix is shown in equation (6).

$$Attention(Q,K) = dropout\left(\frac{QK^T}{\sqrt{d_k}}\right) \tag{6}$$

In equation (6), $Attention(Q,K)$ represents the attention matrix. It utilizes the $dropout$-layer to improve the generalization performance of the model. Additionally, a linear layer is used for projection to obtain $head_i$, and the multi head attention matrix is summed to obtain $MultiHead \in R^{N \times N}$. The calculation is shown in equation (7).

$$\begin{cases} MultiHead = Sum(head_1, \cdots, head_n) \\ head_i = W_h Attention(QW_i^Q, KW_i^K) \end{cases} \tag{7}$$

In equation (7), W_h, W_i^Q, and W_i^K all represent parameter matrices. Then, $E_r \in R^{N\times d_k}$ is used for dot product as residual connection, and multi-layer perceptrons are used for feature projection to obtain dynamic images. The calculation is shown in equation (8).

$$\begin{cases} A_{MH} = LN(MultiHead) + LN(E_r, E_r^T) \\ DA = SoftMax(\mathrm{Re}\,lu(\max(0, A_{MH}W_1 + b_1)W_2 + b_2)) \end{cases} \tag{8}$$

In equation (8), $E_r = W_r X, LN$ represents the normalization of dimension d . W_r, W_1, W_2 both represent parameter matrices. b_1, b_2 represent bias parameters. To improve the convergence and interpretability of static graphs, an information fusion layer was designed on the grounds of gated recurrent units in the experiment. The process of information fusion is shown in equation (9).

$$\begin{cases} r_T = \delta(W_r \bullet E + U_r \bullet X_T) \\ z_T = \delta(W_z \bullet E + U_z \bullet X_T) \\ \hat{h}_T = \tanh(W_h \bullet X_T + r_T(U_h \bullet E)) \\ h_T = (1 - z_T) \bullet E + z_T \bullet \hat{h}_T \end{cases} \tag{9}$$

In equation (9), $E \in R^{N\times d}$ represents the given node embedding. X represents power timing input data. X_T represents dimension. W, U both represent weight matrices. $\hat{E} = h_T \in R^{N\times d}$ represents a new node embedding. Then it calculates the static dependency matrix SA between different nodes, as shown in equation (10).

$$SA = SoftMax(\mathrm{Re}\,lu(\hat{E} \bullet \hat{E}^T)) \tag{10}$$

In equation (10), $\mathrm{Re}\,LU$ and $Soft\max$ both represent activation functions. The former is responsible for filtering operations, while the latter is responsible for normalizing the adaptive static matrix. Then, to explore the periodic characteristics of power time series data variables, the experiment was improved on the grounds of time convolutional networks, and a multi-scale time convolution module was designed on the grounds of this. The tanh activation function is used as the feature filter, and the $sigmoid$ activation function is used as the filter to control the information transferred to the next module of the process, as shown in equation (11).

$$\begin{cases} Z = cancaat(S * f_{1\times2}, S * f_{1\times3}, S * f_{1\times6}, S * f_{1\times7}) \\ \hat{Z} = AvgPool(\tanh(Z) + sigmoid(Z)) \end{cases} \tag{11}$$

In equation (11), $S \in R^T$ represents one-dimensional temporal input. $f_{1\times2} \in R^2, f_{1\times3} \in R^3, f_{1\times6} \in R^6, f_{1\times7} \in R^7$ represent filters of different scales. The purpose of the graph attention module is to weigh and fuse the information of different nodes with the information of neighboring nodes. The architecture of the attention module is shown in Figure 5.

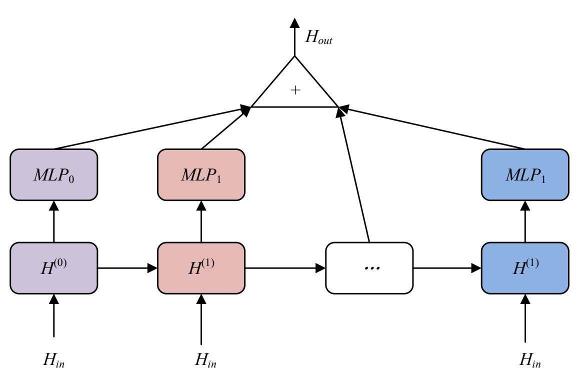


Figure 5. The architecture of the attention module.

However, the different steps of information dissemination can lead to the problem of over smoothing in graph attention networks. In response to this issue, the experiment retained some of the original states of nodes during the propagation process to achieve multi-faceted properties. The calculation of information dissemination steps is shown in equation (12).

$$H^{(k)} = \beta H_{in} + (1-\beta)\overline{A}H^{(k-1)} \tag{12}$$

Then, to preserve some lost node information to prevent extreme situations where there is no space, information selection steps are introduced to filter out important information. The calculation is shown in equation (13).

$$H_{out} = \sum_{i=0}^{K} H^{(k)} W^{(k)} \tag{13}$$

In equation (13), β represents a hyperparameter. K represents the depth of propagation. $W^{(k)}$ represents the parameter matrix, which is treated as a feature selector in the experiment. H_{in} represents the hidden state of the previous layer's output. H_{out} represents the hidden state of the current convolutional layer output. To model the interaction between long-term and short-term patterns, two graph attention modules were used in the experiment to aggregate node information on dynamic and static graphs. The final aggregated output calculation is shown in equation (14).

$$H_f = H_{static} + H_{dynamic} \tag{14}$$

Finally, there is a skip connection layer that standardizes the aggregated information from the graph attention module to obtain the same sequence length and pass it to the output module. The output module consists of a convolutional neural network with two filters, which converts the channel size of the input into the desired output size. The experiment uses multivariable power time series data to carry out single-step prediction, so the overall framework of the proposed method is finally obtained as shown in Figure 6.

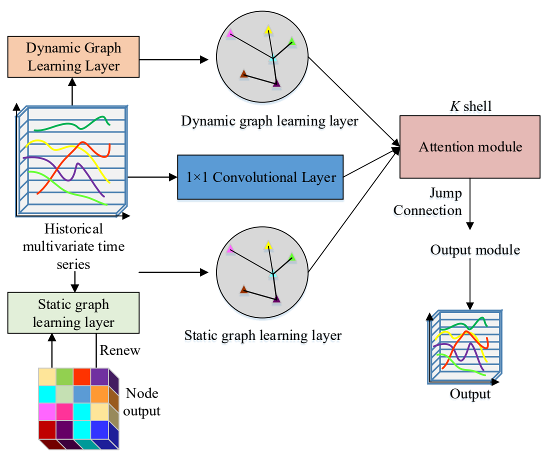


Figure 6. The overall framework of the constructed model.

In Figure 6, to mine the long-term and short-term correlation patterns implied between power time series metrics, the DSGAT model uses two graph learning layers to learn to generate dynamic and static adjacency matrices respectively. The temporal convolution module uses a gating mechanism to extract time series correlations and is mainly composed of two parallel multi-scale convolutional neural network layers to process power time series data. In the graph attention module, DSGAT uses two independent graph attention networks to summarize the information of dynamic and static matrices to optimize the time series features.

4. PERFORMANCE TESTING AND APPLICATION EFFECTS

The rapid increase in power timing data in smart grids makes research on power timing prediction crucial for the safe and stable operation of the entire power grid. To analyze the performance and practical application effects of the improved power time series prediction model constructed in the experiment, three other methods are selected for comparison.

4.1 Analysis of performance test results

It is necessary to verify the actual performance of the power time series prediction model (IDS-GAT) on the grounds of improved dynamic and static graph attention networks constructed in the experiment. Therefore, this experiment selects a wind power energy prediction model on the grounds of Graph Attention Network and Bidirectional Deep Learning Long-Short Term Memory (GAN-LSTM), an improved EMD based wind power effective prediction model (IEMD), a big data and time series based ocean environmental information prediction model (IClustering ANN), and the method constructed by the research institute for comparison[18-20]. Meanwhile, to avoid accidental errors in the data during the experiment, it is necessary to ensure that all experimental parameters are set the same. The experimental simulation environment and parameter settings are shown in Table 1.

Table 1. The experimental simulation environment and parameter settings.

Project	Parameter
Processor	Intel(R)Xeon(R)CPUE7-4809V3@2.00GHz
Experimental environment	COREi7
Data regression analysis platform	SPSS 26.0
Data storage	MySQL
Simulation tools	Simulink
Simulation software	Matlab R2018b
Memory	16GB
Operating system	Windows10
Network architecture	Pytorch v1.2.0

The experiment selected two publicly available benchmark datasets, Solar Energy dataset and Electricity dataset, for power time series prediction tasks. The data in the Solar-Energy dataset usually comes from the solar power generation monitoring system and may include real-time data from solar power plants in multiple geographic locations. It is usually stored in the form of time series in the format of CSV or similar structured data formats. The dimensions include timestamp, solar radiation intensity, temperature, power generation, etc. The solar radiation intensity and temperature in the Solar-Energy dataset directly affect the efficiency of solar power generation. The Electricity dataset comes from the smart grid system of the power company and covers detailed data on power consumption, production and distribution. The data format is also stored in formats such as CSV or JSON, and the dimensions include timestamp, power consumption, voltage, current, power factor, etc. The power consumption and voltage in the Electricity dataset are key indicators for power grid load management and power distribution. Data incompleteness, noise, outliers, and interdependence between multiple variables are the main challenges in power system prediction tasks.

To ensure the representativeness and generalization of the dataset, the experiment considered different geographical locations, different time periods, and different weather conditions when selecting data; and did not divide the dataset into validation sets and training sets, and only trained on the two datasets. The higher the fitness value, the faster the model converges to a stable state. Therefore, the experiment first compares the fitness of the four algorithms running on different datasets, as shown in Figure 7.

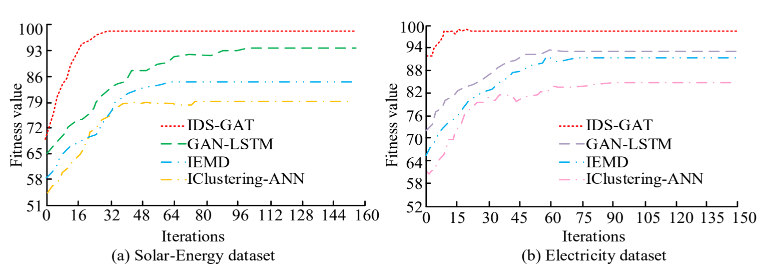


Figure 7. Comparison of convergence speed.

Figure 7a shows the convergence rate changes of four methods on the Solar Energy dataset. This indicates that when the system iteration reaches the 30th iteration, the fitness value of the IDS-GAT model reaches a stable state and has the highest value, up to 99.01. When the system was iterated 104 times, 58 times, and 42 times respectively, the fitness values of the GAN-LSTM model, IEMD model, and IClustering ANN model reached their maximum values of 93.58, 85.97, and 79.01, respectively. Figure 7b shows the convergence rate changes of different models on the Electricity dataset. When the system iterates to the 16th iteration, the fitness value of the IDS-GAT model begins to maintain a stable state, with a value as high as 99.88. At this time, the fitness values of the other three algorithms are still constantly changing. When the number of iterations of the system is about 60, the fitness values of the GAN-LSTM model, IEMD model, and IClustering-ANN model begin to stabilize. By comparison, it can be seen that the IDS-GAT model has a faster convergence speed and can achieve stable operation more quickly when monitoring and predicting power time series data. Next, the average relative error values obtained by running different algorithms on two datasets are compared, and the specific results are shown in Figure 8.

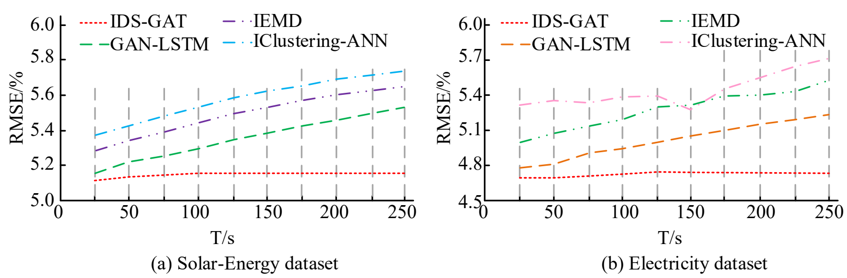


Figure 8. Changes in root mean square difference (RMSE) between two datasets.

Figure 8a shows the root mean square difference changes of four methods on the Solar Energy dataset. As time goes on, the root mean square difference of all algorithms shows varying degrees of increase. When the time is 100 seconds, the RMSE value of the IDS-GAT model is 2.14% and begins to stabilize. However, throughout the entire process of time, the RMSE values of the GAN-LSTM model, IEMD model, and IClustering ANN model did not show stable values, and remained in a constantly increasing and changing state, with the RMSE values increasing. Figure 8b shows the root mean square difference variation of different models on the Electricity dataset. The RMSE value of the IDS-GAT model is 4.66% and has remained at its minimum since then. The RMSE values of the other three algorithms have been consistently increasing, far exceeding those of the IDS-GAT model. Overall, it can be seen that the performance of the IDS-GAT model is continuously optimized on both datasets, and the prediction accuracy is gradually improving. It then compares the changes in the CORR values on the two datasets, as shown in Figure 9.

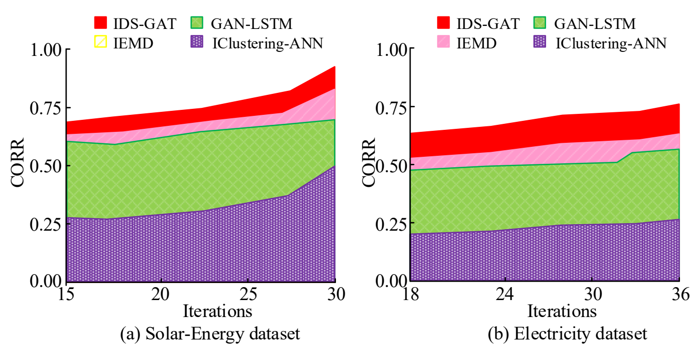


Figure 9. CORR values (linear relationship between true and predicted values).

Figures 9a and 9b show the changes in CORR values for the four methods on the Solar Energy dataset and the Electricity dataset, respectively. This indicates that as the system iteration increases, the CORR values of all four algorithms begin to continuously increase. On the Solar Energy dataset, when the system is iterated 30 times, the CORR value of the IDS-GAT model is the highest, reaching 0.982. On the Electricity dataset, when the system iterates 36 times, the CORR value of the IDS-GAT model reaches 0.756. The ARR values of the IDS-GAT model have consistently been significantly higher on two datasets, far exceeding those of the other three algorithms. This indicates that the IDS-GAT model is more accurate in predicting power time series and has a more stable operating state during the prediction process.

4.2 Comparison of application effects

Finally, the constructed model is applied to the power system of a company in the United States. Then it compares the actual threats faced by the power timing with the threat predictions obtained by the model, as shown in Figure 10.

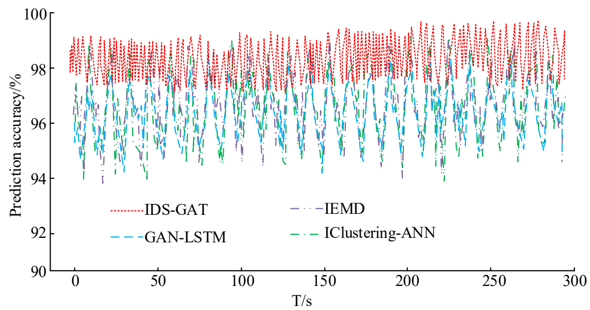


Figure 10. Comparison of accuracy of threat prediction using different methods.

Figure 10 shows that over time, the accuracy of the four methods in predicting threats to the power system shows a fluctuating state. When the time is 200s, the IEMD algorithm has a minimum prediction accuracy for threats, approaching 94%, and the fluctuation range of prediction accuracy throughout the entire change process is 94% to 99%. Meanwhile, the accuracy of threat prediction using the GAN-LSTM model and IClustering-ANN model fluctuates between 94% and 99%, but the range of accuracy fluctuations is smaller. The IDS-GAT model has a denser accuracy in predicting threats, with fluctuations concentrated around 98% and approaching 99%. By comparison, it can be seen that the IDS-GAT model has a higher accuracy in predicting threats to the power system and can be used as the main method for threat prediction. Then, when the parameter Layer changes, the RSE values predicted by the four models for the threat to the power system are compared, as shown in Figure 11.

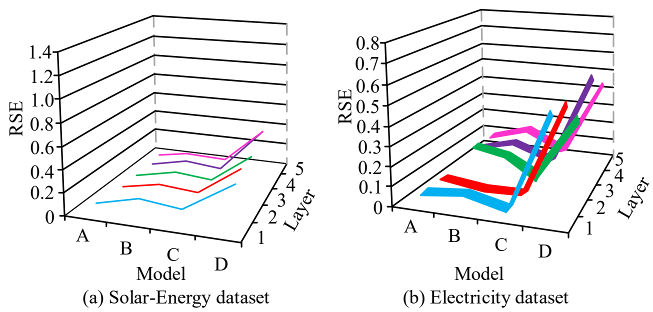


Figure 11. Comparison of RSE values of different models on two datasets when the parameter Layer changes.

In Figure 11, the IDS-GAT method is set to A, the GAN-LSTM method is set to B, the IEMD method is set to C, and the IClustering ANN method is set to D. Figure 11a shows the RSE value changes of four methods on the Solar Energy dataset. This indicates that as the parameter Layer increases, the RSE values of the four methods exhibit the same trend of fluctuation. But among them, the RSE value of the IDS-GAT method has always been the smallest, with a range of numerical fluctuations between 0.021 and 0.035. The RSE values of GAN-LSTM method, IEMD method, and IClustering ANN method are significantly higher than those of IDS-GAT method. The RSE value of the IClustering ANN method has always been the highest, reaching as high as 0.385. Figure 11b shows the variation of RSE values for different models on the Electricity dataset. This indicates that the RSE value of the IDS-GAT method has been fluctuating between 0.050 and 0.10, and is significantly smaller than the other three algorithms. The above results indicate that using the IDS-GAT method to predict threats to the power system has the smallest RSE value and higher accuracy, which can achieve accurate prediction of threats to the power system.

5. CONCLUSION

To effectively predict the operational status and potential threats of the power system, a power timing and threat prediction model on the grounds of an improved dynamic and static graph attention network was proposed in the experiment. During the process, a dual perspective deep neural network was first used to correct and fill in the relevant data of active power loss, to ensure the integrity of power timing data. Next, it used the DSGAT model to learn the generation of dynamic and static adjacency matrices and ultimately completes the prediction of power time series. The data showed that on the Solar Energy dataset and the Electricity dataset, when the system iterated to the 30th and 16th times respectively, the fitness values of the IDS-GAT model were as high as 99.01 and 99.88, respectively. By using different methods to predict threats to the power system, the accuracy of the IDS-GAT model in predicting threats approached 99% when the operating time was 200s. In practical application, the increase of the parameter Layer did not cause significant changes in the RSE value of the IDS-GAT model. The above results indicated that the DS-GAT model can demonstrate superior predictive performance on different power time series datasets, with higher accuracy and better robustness, which was of great significance for risk management and emergency planning in power systems. However, the proposed model may still not fully adapt to all the characteristics of the data generated by the power system. Future research can consider collecting more regional and diverse types of power system data for cross-regional and cross-type empirical analysis.

ACKNOWLEDGEMENT

This research was supported by the Guangxi Power Grid Technology Project under Grant 046100KK52222001.

REFERENCES

[1] Liu, H., Liu, Z., Jia, W. and Lin, X., "Remaining useful life prediction using a novel feature-attention-based end-to-end approach," IEEE Transactions on Industrial Informatics, 17(2), 1197-1207 (2020).

[2] Wu, Y., Dai, H. N. and Tang, H., "Graph neural networks for anomaly detection in industrial Internet of Things," IEEE Internet of Things Journal, 9(12), 9214-9231 (2021).

[3] Li, F., Feng, J., Yan, H., Jin, G., Yang, F., Sun, F. and Li, Y., "Dynamic graph convolutional recurrent network for traffic prediction: Benchmark and solution," ACM Transactions on Knowledge Discovery from Data, 17(1), 1-21 (2023).

[4] Qin, Y., Chen, D., Xiang, S. and Zhu, C., "Gated dual attention unit neural networks for remaining useful life prediction of rolling bearings," IEEE Transactions on Industrial Informatics, 17(9), 6438-6447 (2020).

[5] Torres, J. F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F. and Troncoso, A., "Deep learning for time series forecasting: a survey," Big Data, 9(1), 3-21 (2021).

[6] Yang, Y., Tan, Z., Yang, H., Ruan, G., Zhong, H. and Liu, F., "Short-term electricity price forecasting based on graph convolution network and attention mechanism," IET Renewable Power Generation, 16(12), 2481-2492 (2022).

[7] Huang, C. J., Shen, Y., Chen, Y. H. and Chen, H. C., "A novel hybrid deep neural network model for short-term electricity price forecasting," International Journal of Energy Research, 45(2), 2511-2532 (2021).

[8] He, H., Zhang, Q., Bai, S., Yi, K. and Niu, Z., "CATN: Cross attentive tree-aware network for multivariate time series forecasting," Proceedings of the AAAI Conference on Artificial Intelligence," 36(4), 4030-4038 (2022).

[9] He, K., Chen, X., Wu, Q., Yu, S. and Zhou, Z., "Graph attention spatial-temporal network with collaborative global-local learning for citywide mobile traffic prediction," IEEE Transactions on mobile computing, 21(4), 1244-1256 (2020).

[10] Simeunović, J., Schubnel, B., Alet, P. J. and Carrillo, R. E., "Spatio-temporal graph neural networks for multi-site PV power forecasting," IEEE Transactions on Sustainable Energy, 13(2), 1210-1220 (2021).

[11] Cui, Y., Zheng, K., Cui, D., Xie, J., Deng, L., Huang, F. and Zhou, X., "METRO: a generic graph neural network framework for multivariate time series forecasting," Proceedings of the VLDB Endowment, 15(2), 224-236 (2021).

[12] Peng, B., Ding, Y., Xia, Q. and Yang, Y., "Recurrent neural networks integrate multiple graph operators for spatial time series prediction," Applied Intelligence, 53(21), 26067-26078 (2023).

[13] Chen, H., Jiang, Y., Zhang, X., Zhou, Y., Wang, L. and Wei, J., "Spatio-temporal graph attention network for sintering temperature long-range forecasting in rotary kilns," IEEE Transactions on Industrial Informatics, 19(2), 1923-1932 (2022).

[14] Dong, H., Han, S., Pang, J. and Yu, X., "A joint network of non-linear graph attention and temporal attraction force for geo-sensory time series prediction," Applied Intelligence, 53(14), 17346-17362 (2023).

[15] Wang, Y., Feng, S., Wang, B. and Ouyang, J., "Deep transition network with gating mechanism for multivariate time series forecasting," Applied Intelligence, 53(20), 24346-24359 (2023).
[16] Deng, A. and Hooi, B., "Graph neural network-based anomaly detection in multivariate time series," Proceedings of the AAAI Conference on Artificial Intelligence, 35(5), 4027-4035 (2021).
[17] Spadon, G., Hong, S., Brandoli, B., Matwin, S., Rodrigues-Jr, J. F. and Sun, J., "Pay attention to evolution: Time series forecasting with deep graph-evolution learning," IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(9), 5368-5384 (2021).
[18] Mansoury, I., El Bourakadi, D., Yahyaouy, A. and Boumhidi, J., "A novel wind power prediction model using graph attention networks and bi-directional deep learning long and short-term memory," International Journal of Electrical and Computer Engineering (IJECE), 13(6), 6847-6854 (2023).
[19] Abedinia, O., Lotfi, M., Bagheri, M., Sobhani, B., Shafie-Khah, M. and Catalão, J. P., "Improved EMD-based complex prediction model for wind power forecasting," IEEE Transactions on Sustainable Energy, 11(4), 2790-2802 (2020).
[20] Wen, J., Yang, J., Jiang, B., Song, H. and Wang, H., "Big data driven marine environment information forecasting: a time series prediction network," IEEE Transactions on Fuzzy Systems, 29(1), 4-18 (2020).