Short-Term Power Load Forecasting Method Based on GRU-Transformer Combined Neural Network Model

Weiwei Mao[1], Suping Yu[1] and Wenqing Chen[2]

[1]College of Computer and Information Engineering, Luoyang Institute of Science and Technology, Luoyang, China
[2]School of Electrical Engineering and Automation, Luoyang Institute of Technology, Luoyang, China

Load Forecast (LF) is an important task in the planning, control and application of public power systems. Accurate Short Term Load Forecast (STLF) is the premise of safe and economical operation of a power system. In the research of short-term power load forecasting, machine learning and deep learning are the most popular methods at present, but there still exists a problem that the single and simple structure of power load forecasting model leads to low accuracy of load forecasting. In order to improve the accuracy of STLF, a Gated Cycle Unit (GRU)-Transformer combined neural network model is proposed. Transformer encoder structure is used as feature extractor to mine the complex mapping relationships between the input features and load. The advantage of self-attention mechanism is used to solve the problem of information loss of long sequences in short-term power load forecasting. At the same time, the multivariate time series model of GRU is used for model training. The experimental results on the power load data set of a certain region in southwest China and Panama City show that the proposed combined model prediction method has higher accuracy than those proposed in other literatures, which further proves its feasibility and superiority.

ACM CCS (2012) Classification: Theory of computation → Design and analysis of algorithms

Keywords: short-term power load forecasting, self-attention mechanism, multivariate time series, gated recurrent unit, feature extractor

1. Introduction

With the rapid development of science and technology and social economy, smart grid has been developed. The power industry supports the development of the national economy, the stability of society and the convenience of people's life. It is the main driving force of modern production and plays an increasingly important role in the national economy [1, 2]. The role of the power industry is to meet the electricity demand of various industries of a country, and the power load forecasting is a crucial technology in the power industry. Accurate power load forecasting can promote the development and stable operation of the power industry. Power load forecasting is based on power load itself, temperature factors, weather factors, economic development factors, *etc.*Its goal is to study the dependance between historical load and load influencing factors, and to scientifically forecast future power load [3–5]. Accurate power load forecasting can alleviate the rift between power supply and demand, and is paramount to the reasonable arrangement of the operation mode and maintenance plan of the power system or microgrid, which not only reduces the operation cost, but also improves the efficiency of the power system or microgrid [6–9]. Inaccurate power load forecasting can cause power companies to bear huge financial burdens and additional costs [10].

Because it is difficult to store electric energy on a large scale, and furthermore, social electricity consumption is constantly changing, it is necessary for power system generation and social electricity consumption to balance dynamically. The power system provides stable, economic and uninterrupted high-quality electric energy [11–13] to support the development of the national economy and social stability. The main sources of electric energy are: hydropower, thermal power, nuclear power, wind power and others. With the proposal of "carbon neutrality", electricity is an important factor affecting carbon dioxide emissions [14]. In order to replace polluting power sources such as thermal power with clean power sources, it is increasingly necessary to accurately predict the power consumption required by society and make electric energy reserves [15]. Therefore, high-precision power load prediction is of great significance for the promotion of "carbon neutrality" [16, 17]. Power load forecasting can predict the load demand of each region, guide the power generation of power plants and the power supply of power supply companies, and ensure the stable operation of the power grid. Improving the accuracy of power load forecasting is not only of great significance for promotion of the development of "carbon neutrality", but also to reduce the unnecessary consumption of manpower and material resources in the development of power industry.

Power load forecasting is based on the changes of historical load data, time factors and meteorological factors, and the analysis of traditional or deep learning model methods is used to assess the demand load of the region [18]. This paper focuses on short-term load forecasting, which is to predict the load in the next few minutes to several weeks. There are two main steps in short-term power load forecasting: (1) Feature extraction of the original input data; (2) Construction a predictive model, using machine learning or deep learning methods, and finally using its fitting ability to establish the relationship between features and loads [19]. Before load forecasting, it is necessary to extract the features of the load to dig out the features contained in the power load, analyze the load characteristics to find out its impact factors, and extract the features contained in the impact factors [20]. In traditional feature extraction, existing features are processed to simplify features or add some new features. For example, Wang [21] uses sparse autoencoders to suppress the output of hidden layer neurons to achieve a sparse network effect. However, the trend of power load is relatively complex, and the trend of traditional statistical characteristics is relatively simple, which cannot be well fitted. Wu and Zhu [22] combined the time convolutional network with GRU, divided the data set into two parts, extracted the time series features of the time series data, and then combined the features of the non-time series to improve monotony and trend, so as to improve the accuracy of the model. Zhu *et al.* [23] used a convolutional neural network to select parameters with high correlation between input parameters and load as inputs, and built a load prediction model using BiLSTM. However, convolutional neural networks require a lot of data for fine-tuning in feature screening, and the lack of data makes the model easy to overfit. The LSTM algorithm is very dependent on the parameter setting, and the generalization of the model will be poor if the parameters are set by experience. Chen and Zhang [24] proposed nonlinear dynamic adjustment inertial weight particle swarm optimization algorithm to optimize the LSTM model, improve the global optimization ability, reduce the model's dependence on pre-set parameters, and improve the model generalization. Recurrent neural network (RNN) is used to solve the problem of power load forecasting due to its superiority in time series processing. Guo *et al.* [25] used the improved particle swarm optimization method to extract the features of weather factors in the input factors, and then used the RNN optimized by the simulated annealing algorithm to establish a short-term power load prediction model to solve the characteristics of weak global search ability of RNN, and the model achieved a good prediction effect. When dealing with time series, RNN acts in serial operation mode, which will restrict the running speed of the model. RNN will also have information loss when performing each recursion, and serious information loss will occur when the input sequence is long, resulting in memory degradation. This paper introduces Transformer model into the field of short-term power load forecasting to solve the above-mentioned problems. Transformer's multi-head attention mechanism can learn different behaviors based on the same attention mechanism. At the same time, its core self-attention mechanism seeks the dependency between input feature vectors based on matrix multiplication. This operation principle of self-attention realizes parallel computation, which greatly improves the running speed compared to RNN that can only be computed in a serial way [22]. The combined GRU-Transformer NN model proposed in this paper first uses Transformer's coding layer as a feature extractor to dig deeply into the characteristics of power load itself and search for relevant impact factors.That way it is possible to expresses and extract the features related to the impact factors to obtain the most significant features in the training data set. Then the extracted features are passed into the fully connected layer, and then the two-layer GRU model is used to predict, which can significantly improve the model learning ability compared with the single-layer GRU. Finally, a fully connected layer is used to fit the predicted value, and the short-term power load prediction is realized.

2. Basic Principles

2.1. Transformer Feature Extractor

The transformer is introduced into this prediction model, using only the encoder part as a feature extractor. The encoder consists of an input layer, a positional coding layer and four coding layers, which are arranged together in a stacked manner. The input layer is primarily a fully connected neural network that maps the time series to a preset model dimension. Position coding injects relative or absolute position information into sequence vectors. The model can learn the position information of time series. The time series is then passed into four coding layers with the same structure, and the data is maintained in a complete and standard distribution through residual concatenation and layer normalization. The Transformer coding layer structure is shown in Figure 1.

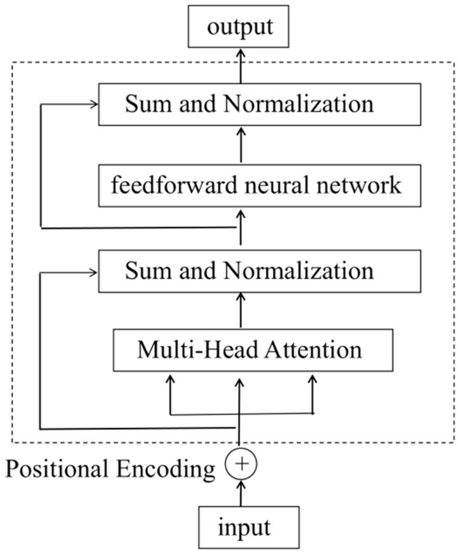

Figure 1. Transformer code layer structure diagram.

2.2. Self-attention Mechanism

The self-attention mechanism is used to embed the influence factors before and after the time series by calculating and adjusting the weight of the sample in the input vector. Figure 2 shows the structure of a single encoder and self-attention mechanism.

Suppose the input vector is $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_t, \boldsymbol{x}_{t+1}, ...\}$, t is the time series $\{t \mid t = 1, 2, ..., T\}$, maps each input vector to three different Spaces, generating a query vector (Q), a key vector (K), and a value vector (V). And the external linear mapping to three matrices (W^Q, W^K, W^V), the steps are as follows.

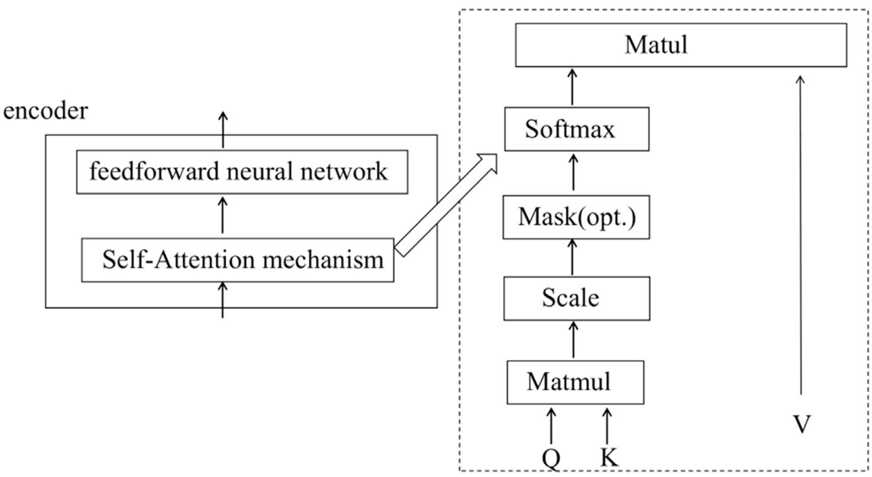

Figure 2. Single encoder and self-attentional mechanism.

a) Sample generation vector is obtained, the formula is:

$$\begin{Bmatrix} \boldsymbol{Q} = \boldsymbol{X} \cdot \boldsymbol{W}^{Q} \\ \boldsymbol{K} = \boldsymbol{X} \cdot \boldsymbol{W}^{K} \\ \boldsymbol{V} = \boldsymbol{X} \cdot \boldsymbol{W}^{V} \end{Bmatrix} \tag{1}$$

b) Use the scaled dot product as the attention scoring function, the formula is:

$$Attention(Q, K, V) = Softmax\left(\frac{\boldsymbol{Q}_i \cdot \boldsymbol{K}_i^T}{\sqrt{D_k}}\right) \cdot V \tag{2}$$

where, $\boldsymbol{Q_i} \cdot \boldsymbol{K_i}^T$ is the score of the sample in each vector, $\sqrt{D_k}$ is the optimization of the training gradient, and *Softmax* is the function of normalization by column.

c) Through the residual network normalization processing, the result is used as the input of the feedforward neural network.

2.3. Multi-head Attention Mechanism

Each self-attention module is called a head, and the focus of each self-attention head is different. For example, some heads pay attention to the local information of the time series, while some heads pay attention to the global information. The multi-head attention composed of multiple heads can enable each head to perform its own duties without interfering with each other, so as to carry out complex tasks. Formulas (3) and (4) are the splicing methods of multi-head attention

$$Multi(Q, K, V) = Concat\left(head_1, head_h\right)\boldsymbol{W}^{0} \tag{3}$$

$$head_i = Attention\left(Q\boldsymbol{W}_i^{Q}, K\boldsymbol{W}_i^{K}, V\boldsymbol{W}_i^{V}\right) \tag{4}$$

2.4. Gated Cycle Unit

The GRU is a variant of the LSTM with only update and reset gates. On the basis of retaining LSTM to solve the long- and short-term dependence problems, such as gradient disappearance and gradient explosion, the internal structure of neurons is optimized to reduce the calculation amount of cell state and solve the problem that LSTM has as it needs to set too many model parameters. GRU has a faster convergence rate than LSTM during training. Figure 3 shows the GRU network structure.

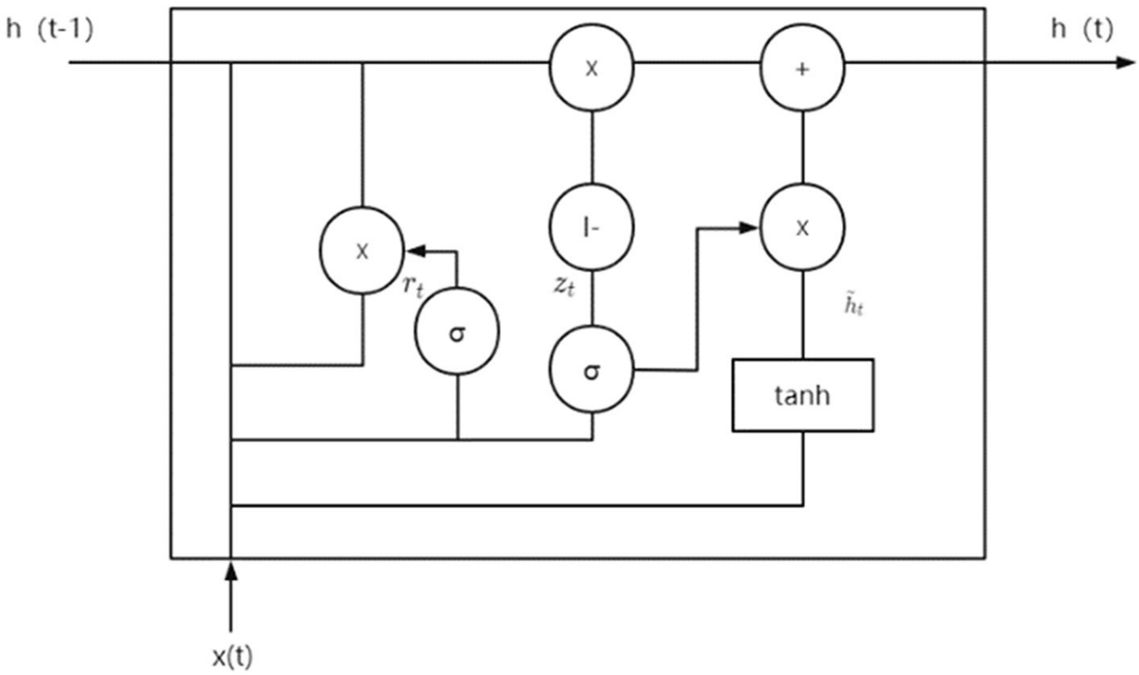

Figure 3. Schematic diagram of GRU network structure.

The forward update formula of GRU network is as follows:

$$z_t = \sigma\left(\boldsymbol{W_Z}\, x_t + \boldsymbol{U_Z} h_{t-1} + \boldsymbol{b}_z\right) \tag{5}$$

$$r_t = \sigma\left(\boldsymbol{W_r}\, x_t + \boldsymbol{U_r} h_{t-1} + \boldsymbol{b_r}\right) \tag{6}$$

$$\tilde{h}_t = \tanh\left(\boldsymbol{W_h}\, x_t + \boldsymbol{U_h}\, r_t \cdot h_{t-1} + \boldsymbol{b_h}\right) \tag{7}$$

$$h_t = \left(1 - z_t\right) \cdot h_t + z_t \cdot \tilde{h}_t \tag{8}$$

where, r_t is the reset gate, z_t is the update gate, x_t is the input value of the current time load, h_t the output value of the current time load, h_{t-1} is the hidden layer information of the previous node, b, W and U are the weight coefficient matrix, used to calculate the output of the reset gate and the update gate.

3. GRU-Transformer Combined Neural Network Prediction Model

3.1. Model Description

As a typical time series forecasting problem, short-term power load forecasting needs to consider various factors manually when applying the traditional time series forecasting model. Models based on deep neural networks have excellent effects in solving the above problems, among which transformer model is the most typical. The self-attention mechanism of a transformer can extract low - and high-dimensional features of input data when facing too long time series, which gives the transformer model an advantage in processing long-term information and capturing long-term dependencies.

However, transformer's self-attention mechanism will increase the cost and memory consumption due to the secondary calculation of the model caused by too long input samples, especially when the model is stacked with several self-attention blocks. To solve this problem, only the encoder part of Transformer is used to extract high-latitude and low-dimensional features of power load through the encoder part, so as to express power load impact factors in a deeper manner. The extracted impact factors and external influence factors are input into the fully connected layer, and then forecast is performed through the two-layer GRU network. Finally, the model output value is input to the output of fully connected layer in order to fit it to the predicted value.

The coding layer of transformer is applied to search for the main load impact factors, and the characteristics of the impact factors are expressed and extracted. Based on GRU network, the GRU-Transformer combined neural network prediction model is established.

3.2. Structure of Prediction Model

The coding layer of transformer is applied to search for major load impact factors, express and extract the characteristics of impact factors, and establish a combined prediction model of GRU-Transformer based on GRU network, as shown in Figure 4. It consists of an input layer, transformer feature extraction layer, GRU multiple time series network layer and an output layer. The model is described as follows.

Input layer: The load history data is combined with relevant factor data, and the normalization process is carried out as the input of the model. Assuming that the length of the data is N, the data is represented as $X = [x_1, x_2, ..., x_n]$.

Encoder feature extraction layer: This layer is composed of a location coding layer, multi-head attention mechanism layer and feedforward neural network layer. Position coding is generated by sine and cosine functions of different frequencies. Each sample of normalized data is marked with position information to represent different semantic information. The formula is shown in (9) and (10).

$$PE_{(pos,\,2i)} = \sin(pos/10\,000^{2i/d_{moder}}) \tag{9}$$

$$PE_{(pos,\,2i+1)} = \cos(pos/10\,000^{2i/d_{moder}}) \tag{10}$$

The position of the input sequence is pos, i is the dimension, and d_{moder} represents the size of the vector dimension. Multiple attention is stacked together to form multi-head attention, and the multi-head attention mechanism is used to make the model pay attention to different input features. Then, multiple attention heads are spliced together to obtain the multi-head attention matrix, as shown in (3) and (4). The formula of a feedforward neural network is shown in (11).

$$F_{FN}(Z) = \max(0, Z\boldsymbol{W_1} + b_1)\,\boldsymbol{W_2} + b_2 \quad (11)$$

where, b_2 is offset. After the multi-head attention mechanism and feedforward neural network, residual connection and layer normalization are used. The layer normalization formula is shown as (12) and (13), LN represents the layer normalization, u_L is the mean, and $\sigma_L{}^2$ is the variance.

$$s_{out} = LN\left(x + s_{out}(x)\right) \quad (12)$$

$$LN\left(x_i\right) = a \times \frac{x_i - u_L}{\sqrt{\sigma_L^2 + \varepsilon}} + \beta \quad (13)$$

GRU multiple time series network layer: The input of this layer is the feature factor and load data extracted after transformer feature extraction. The layer is constructed of a fully connected layer, two serial GRU layers and output fully connected layer. The input full connection layer combines the feature factor with the collected historical load data, whose formula is shown in (14).

$$H = ReLU(LN + d) \quad (14)$$

The combined data is entered into the GRU cell layer as a newly generated feature. Compared with the single GRU network, the two-layer GRU network can enhance the learning ability of the model. Let the output of the GRU layer be h, and its formula is shown in (15).

$$h = GRU(h_{n-1} + y_n) \quad (15)$$

The function of the output fully connected layer is to collect the influence of historical data in different time periods on the predicted points, and after synthesizing the influence, the real value of the final output is used as the prediction result of the model.

3.3. Loss Function

The mean square error (MSE) is used as the error formula and the loss function is defined as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \acute{y}_i\right)^2 \quad (16)$$

In the command, y_i and $\acute{y}_i$ are label values and output values respectively.

4. Simulation Experiment

4.1. Experimental Environment

Anaconda Navigator software was used in Windows 10, and experiments were conducted in Python 3.7, using the TensorFlow framework.

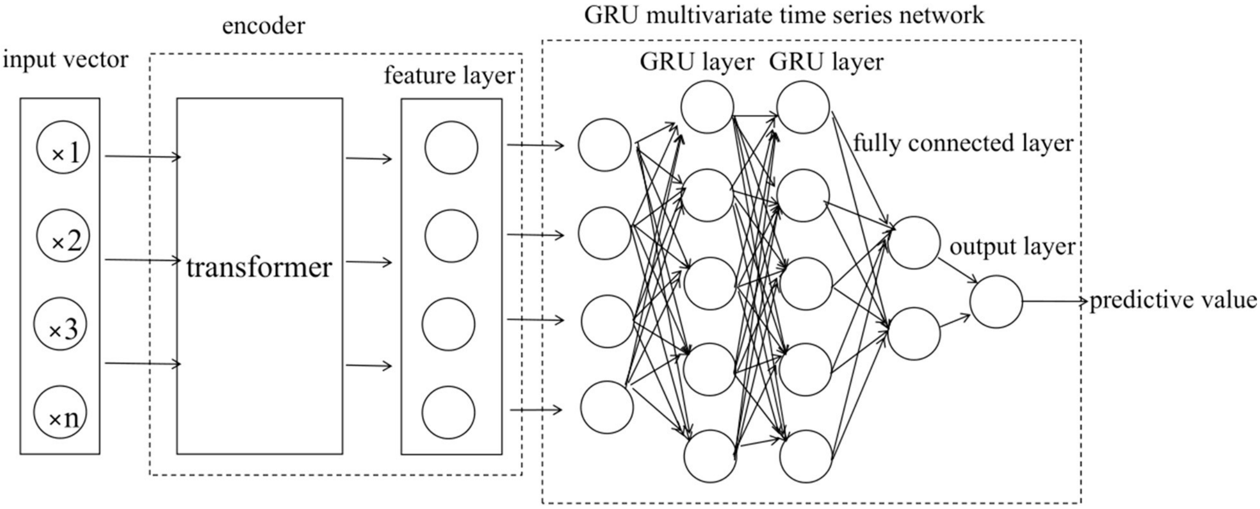

Figure 4. Network structure of GRU-Transformer-NN combined model.

4.2. Acquisition of Experimental Data

From January 2020 to February 2021, loads in a region in southwest China were collected once every 15 minutes every day, 96 times a day. By using the crawler method, weather data and temperature data at the corresponding time were crawled from the meteorological website, and data labels were obtained as shown in Table 1.

4.3. Analysis of Influencing Factors

1. Influence of meteorological factors

Meteorological factors are important factors affecting the change of power load. Figure 5 shows the thermal map of the correlation distribution between the collected meteorological factors and the corresponding load in the current time period. It can be seen from the figure that the power load is strongly correlated with temperature factors and weakly correlated with humidity and wind speed. It is not even related to the weather type, and it can reflect that the temperature, humidity, wind speed and other factors are positively correlated with the load. It can be concluded that with the increase of temperature, humidity and wind speed in the region, the power load in the region will also increase.

2. Influence of time factor

As shown in Figure 6, the power load changes in a single day in January 2020. A total of 96 points are collected, and each point represents the load every 15 minutes. According to the analysis of the figure, the daily power load can be divided into three parts: morning, afternoon and evening. In the morning, people start to work, and industrial and commercial electricity consumption increases slowly. At noon, people enter the lunch break, and the load curve declines. In the afternoon, the work load curve begins to rise again.

Table 1. Data sets used in the experiment.

Attribute	**Meaning**	**Attribute**	**Meaning**
YMD	Date	Weather	Type of weather
max_temp	Maximum temperature	Type of week	Type of week
min_temp	Minimum temperature	Wind speed	Wind speed
ave_temp	Average temperature	Season	Season
Relative humidity	Relative humidity	Last moment load	Last moment load
Rainfall	Rainfall	Current time load	Current time load

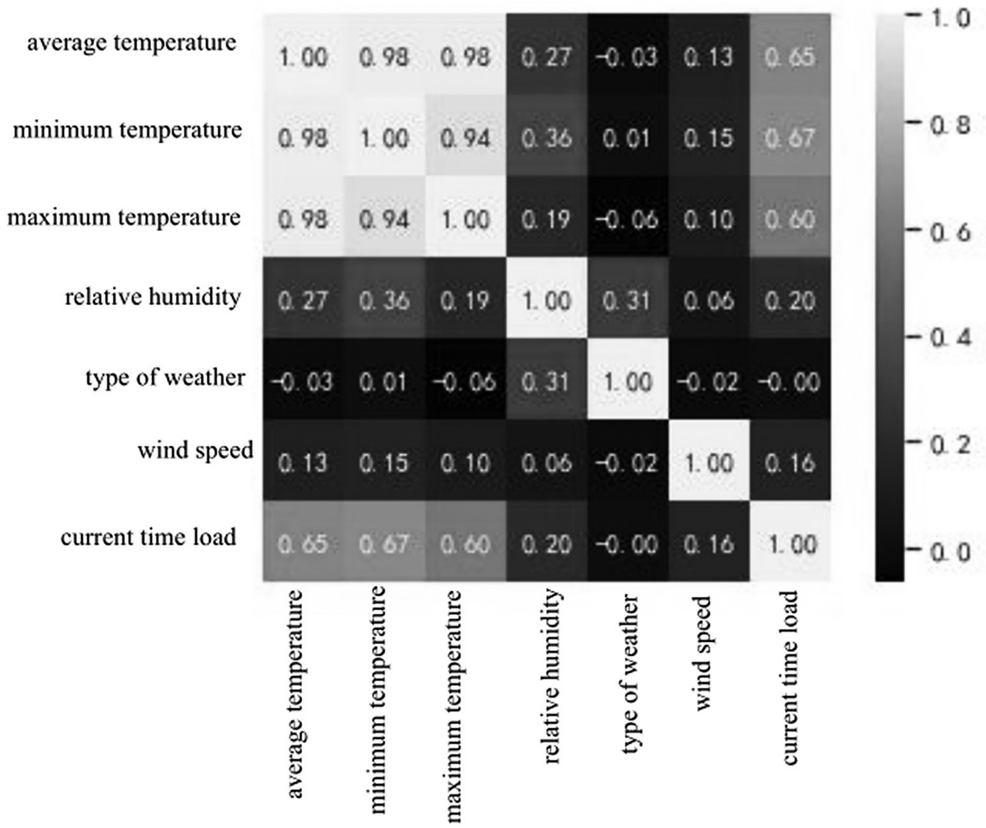

Figure 5. Thermal diagram of correlation between meteorological factors and power load.

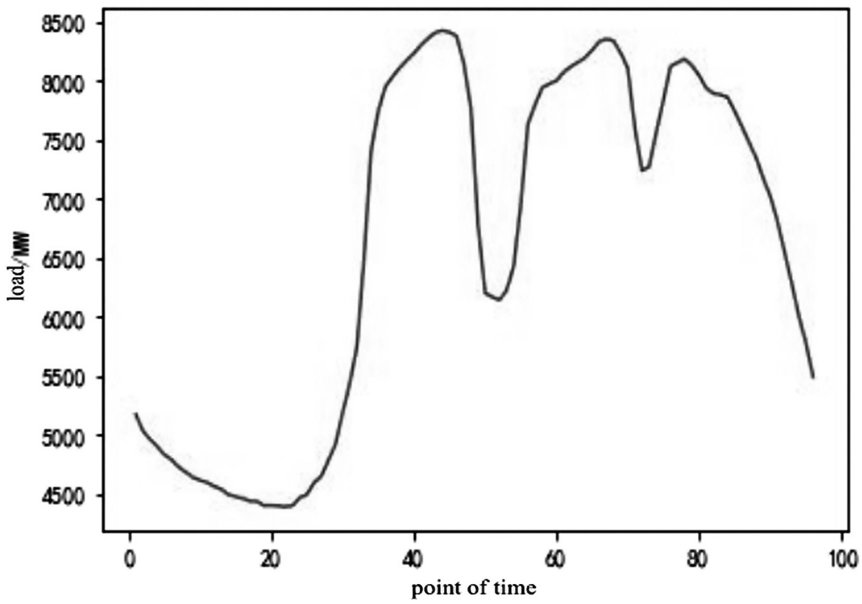

Figure 6. Load variation curve within a day.

4.4. Experimental Data Preprocessing

There will be data anomalies in the data obtained by crawling, including missing data and distorted data. The input of these abnormal data will affect the model prediction accurecy, so it is necessary to set the distorted data to null values, and then fill the null value and missing data with a interpolation filling method.

We normalized the data to [0, 1], the formula is as follows

$$x_{scale} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{17}$$

$x_{\min}$ and $x_{\max}$ are the minimum and maximum values in the data, x is the unprocessed data, and x_{sclae} is the normalized value.

4.5. Experimental Evaluation Criteria

The root mean square error (RMSE), mean absolute percentage error (MAPE) and R Squared are used as the evaluation metrics of the model, and the calculation formula is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \acute{y}_i\right)^2} \tag{18}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \acute{y}_i}{y_i}\right| \tag{19}$$

$$R^2\left(y, \acute{y}\right) = 1 - \frac{SS_{res}}{SS_{tot}} \tag{20}$$

where, $SS_{tot} = \sum_{i=1}^{n}(y_i - \ddot{y}_i)^2$, $SS_{res} = \sum_{i=1}^{n}(y_i - \acute{y}_i)^2$. y_i indicates the actual load, $\acute{y}_i$ indicates the predicted load, and $\ddot{y}_i$ indicates the average actual load.

4.6. Comparative Experiment and Result Analysis

1. Selecting network model parameters

Transformer is used as a feature extractor and GRU multiple time series model to forecast power load. The add function is used to stack multiple network layers, and the Adam optimization method is used in the network training phase, which has the advantage of non-stationary targets.

2. Network training

Based on the power load of a region in southwest China from January 2020 to February 2021, 35798 data points were used, among which the training set and test set were 8: The test set data is input into the prediction model, and the loss of the model decreases as shown in Figure 7. It can be seen that in both the training set and the test set, the loss function gradually decreases and approaches 0. The closer the loss value is to 0, the more suitable the prediction model is for the training set. Therefore, the selected parameters are the optimal parameters of the model.

3. Model prediction results

A total of 192 data points within two days of the forecast day were selected to compare the real value and predicted value of the load data in the form of a line chart, as shown in Figure 8. It can be seen that the predicted load trend is close to the real load trend, indicating that the model proposed in this paper can effectively fits the load change trend and predicts the load data well. However, there is a deviation at point 28 and 48, one of which is the peak value and the other is the valley value. More specifically, this time is 7 o'clock in the morning and 12 o'clock in the noon, respectively. At those times, the disturbance caused by the increase of civil electricity consumption will be the direction of further research.

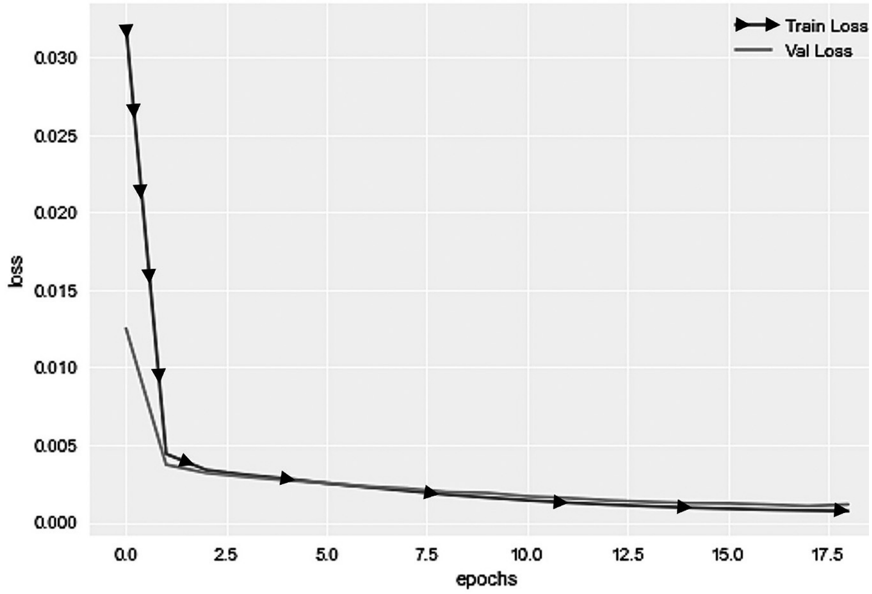

Figure 7. Model loss change curves of training set and test set.

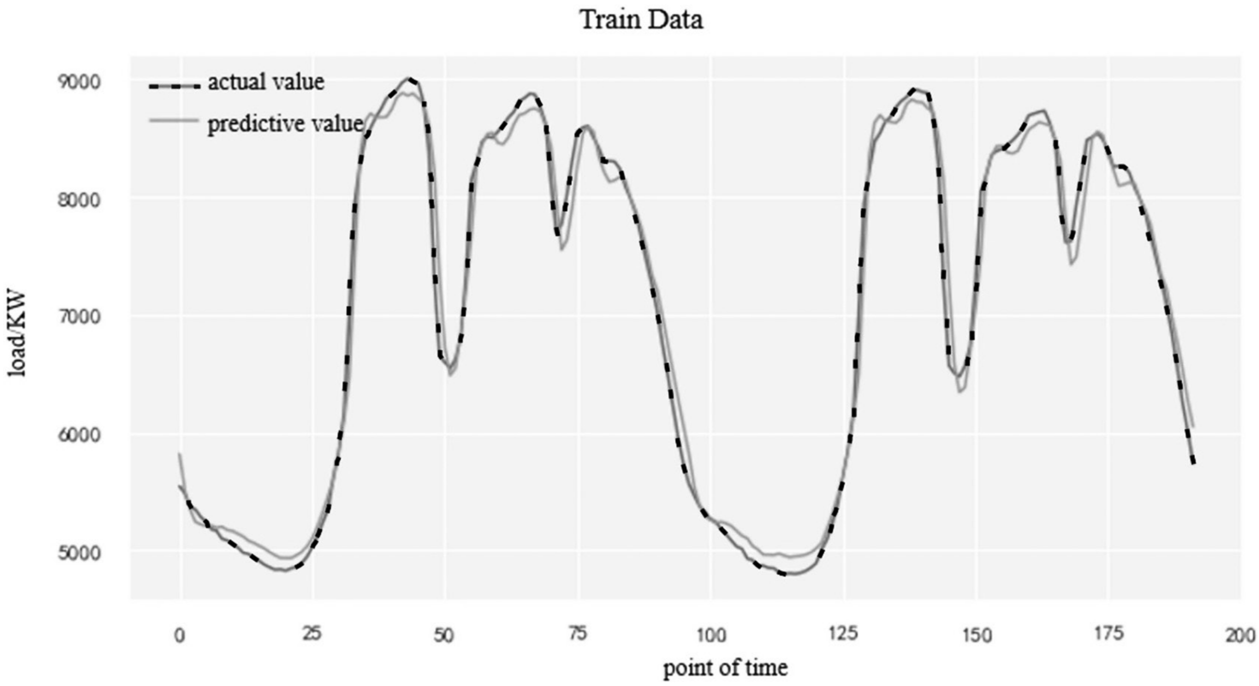

Figure 8. Model prediction result curve.

4. Comparison of prediction models

In order to display the prediction effect of the model more intuitively and accurately, LSTM, GRU, Transformer and other neural network models are used to forecast the same power load data points, and corresponding RMSE, MAPE and R square are calculated as the model performance evaluation indicators. The comparison results are shown in Figure 9. As can be seen from the comparison in Figure 9, LSTM model and GRU model can only predict

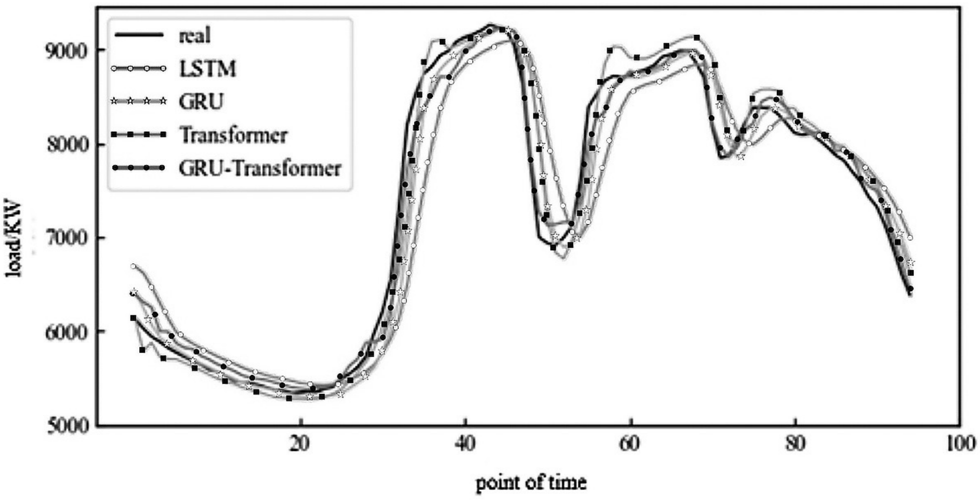

Figure 9. Four model prediction result curves.

the trend of load change, while the transformer model has a large prediction error in peak and valley values. The load trend predicted by the proposed model in this paper is close to the real load trend, and it fits the real data well in any time period. Compared with single LSTM model, GRU model and Transformer model, the fitting effect is also improved. The corresponding RMSE, MAPE and R square were calculated as the performance evaluation indexes of the model, and the comparison results were shown in Table 2. The MAPE of the model proposed in this paper is 1.639%, and the error is 2.581% and 1.133% lower than that of the single GRU and Transformer models, respectively.

4.7. Model Verification

The Panama City power consumption data set (2015-2020) from the Kaggle competition data set was used as the validation data set of the model, and the data set was divided 8:2 to get 38,440 training data points and 9610 test data points. GRU neural network, support vector machine (SVM), transformer and the model proposed in this paper are used in the Panama City power consumption data set for model construction and prediction, and the model prediction error is shown in Table 3. The experimental results show that the proposed model has lower error rates than the other models.

Table 2. Comparison of model prediction accuracy.

Model	RMSE/MW	MAPE%	R^2
LSTM	548.62	5.507	0.858
GRU	411.40	4.220	0.920
Transformer	566.926	2.772	0.968
GRU-Transformer-NN	143.144	1.639	0.984

Table 3. Comparison of model prediction accuracy.

Model	RMSE/MW	MAPE%	R square
GRU	37.961	2.748	0.902
SVM	52.225	4.144	0.814
Transformer	57.61	2.871	0.958
GRU-Transformer-NN	17.589	1.63	0.977

Figure 10 shows the line chart of the actual and predicted load data of Panama City for two days in the middle of the forecast date. It can be seen from the chart that the model method proposed in this paper has good forecasting performance, and there is little deviation between the predicted value and the actual value at the peak and valley values.

5. Conclusion

In order to improve the accuracy of STLF, based on GRU and transformer models, this paper introduces gated cycle unit for short-term power load prediction that utilizes a two-layer gated cycle unit to enhance the learning ability, and uses an input fully connected layer and output fully connected layer to fit the load data.

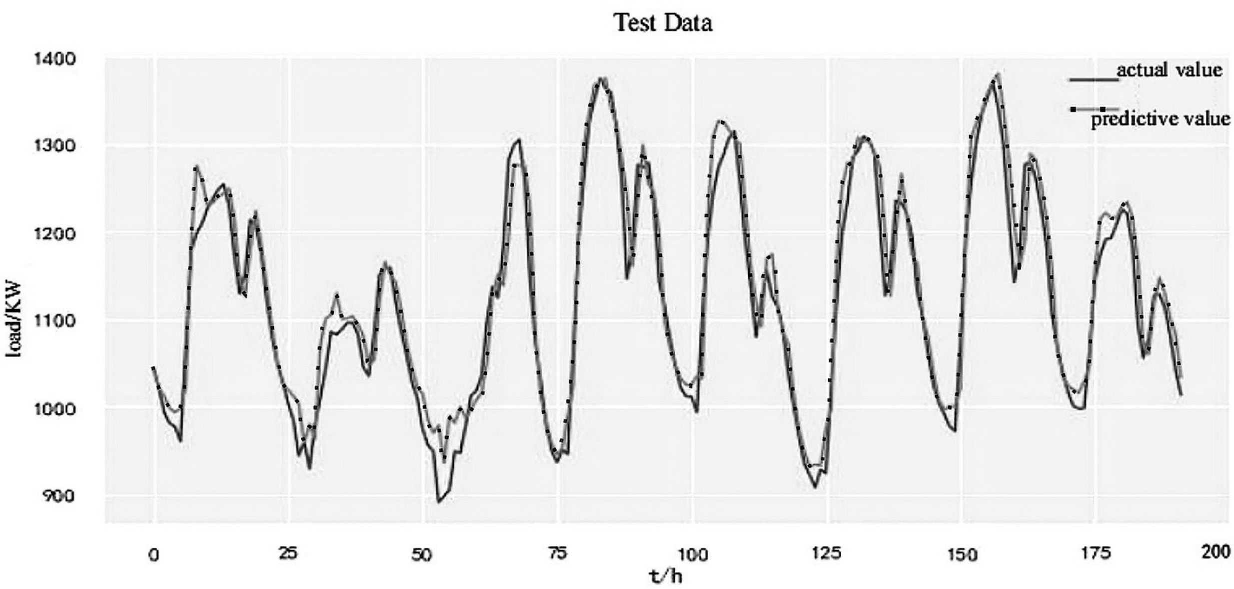

Figure 10. Prediction result curve of the model.

A combined GRU-transformer neural network model is proposed. The GRU module of the combined GRU-transformer neural network is dedicated to processing time series data, and the transformer module extracts features from the variable data affecting the power system. The model proposed in this paper is compared with LSTM, GRU and transformer models in the actual experiment of power load forecasting in a region in southwest China and Panama City. The results show that the proposed model can effectively extract the hidden features of the data set and process the time series data well. The GRU-Transformer combined neural network model has the lowest MAPE and RMSE values, which indicates that the proposed model achieves the best performance among the competing models.

However, when using the GRU Transformer model for short-term power load forecasting, the selected hyperparameters such as learning rate, batch size, and number of neurons in the GRU layer are based on human experience. Hyperparameters chosen is such manner are not likely to achieve the best prediction results, and they may cause overfitting or underfitting during the training process. To prevent this phenomenon from occurring, future research will introduce the sparrow search algorithm and improve it by optimizing hyperparameters to further enhance the prediction accuracy of the model.

Acknowledgement

This paper is supported by the Key Science and Research Projects of Henan Higher Education (Grant No.: 24B510008) and Henan Province Science and technology Development Plan Project (Grant No.: 232300420157, 232102210091 and 232102220069).

References

[1] I. Goodfellow *et al.*, "Deep Learning", MIT Press, 2016.

[2] I. M. Naguib and S. A. Ragheb, "Achieving Sustainability in Smart Cities & Its Impact on Citizen", *International Journal of Sustainable Development and Planning*, vol. 17, no. 8, pp. 2621–2630, 2022. https://doi.org/10.18280/ijsdp.170831

[3] T. Ahmad and H. Chen, "Deep Learning for Multi-scale Smart Energy Forecasting", *Energy*, vol. 175, pp. 98–112, 2019. https://doi.org/10.1016/j.energy.2019.03.080

[4] K. V. B. Saraswathi Devi and M. Srivenkatesh, "An Advanced Hybrid Meta-heuristic Model for Solar Power Generation Forecasting via Ensemble Deep Learning", *Ingénierie des Systèmes d'Information*, vol. 28, no. 5, pp. 1395–1407, 2023. https://doi.org/10.18280/isi.280528

[5] W. Bendali *et al.*, "Optimization of Deep Reservoir Computing with Binary Genetic Algorithm for Multi-time Horizon Forecasting of Power Consumption", *Journal Européen des Systèmes Automatisés*, vol. 55, no. 6, pp. 701–713, 2022. https://doi.org/10.18280/jesa.550602

[6] G. W. Chang and H. J. Lu, "Integrating Gray Data Preprocessor and Deep Belief Network for Day-ahead PV Power Output Forecast", *IEEE Transactions on Sustainable Energy*, vol. 11, no. 1, pp. 185–194, 2018. https://doi.org/10.1109/TSTE.2018.2888548

[7] Y. Chen *et al.*, "Icing Load and Risk Forecasting for Power Transmission Line Based on Multi-scale Time Series Phase-space Reconstruction and Regression", *International Journal of Safety and Security Engineering*, vol. 11, no. 1, pp. 79–90, 2021. https://doi.org/10.18280/ijsse.110109

[8] M. Imani, "Electrical Load-temperature CNN for Residential Load Forecasting", *Energy*, vol. 227, 2021. https://doi.org/10.1016/j.energy.2021.120480

[9] Y. I. D. Kobibi *et al.*, "Continuation Power Flow Analysis of Power System Voltage Stability with Unified Power Flow Controller", *Journal of Intelligent Systems and Control*, vol. 1, no. 1, pp. 60–67, 2022. https://doi.org/10.56578/jisc010106

[10] C. Yang *et al.*, "Time Series Data Classification Based on Dual Path CNN-RNN Cascade Network", *IEEE Access*, vol. 7, pp. 155304–155312, 2019. https://doi.org/10.1109/ACCESS.2019.2949287

[11] R. Liu *et al.*, "State Assessment and Fault Prediction Method of Distribution Terminal Based on SDAE and Hierarchical Bayesian", in *Proc. of the 2019 IEEE Sustainable Power and Energy Conference (iSPEC), Beijing, China*, 2019, pp. 2783–2787. https://doi.org/10.1109/iSPEC48194.2019.8975050

[12] S. Sun *et al.*, "Remaining Life Prediction of Conventional Low-voltage Circuit Breaker Contact System Based on Effective Vibration Signal Segment Detection and MCCAE-LSTM", *IEEE Sensors Journal*, vol. 21, no. 19, pp. 21862–21871, 2021. https://doi.org/10.1109/JSEN.2021.3104290

[13] M. A. Djehaf *et al.*, "Transient Stability Improvement of Multi Machine Power System Including DFIG Wind Farm Using HVDC Link", *Journal Européen des Systèmes Automatisés*, vol. 55, no. 4, pp. 485–493, 2022. https://doi.org/10.18280/jesa.550407

[14] A. K. A. Peñaloza *et al.*, "Review of Deep Learning Application for Short-term Household Load Forecasting", in *Proc. of the 2020 IEEE PES Transmission & Distribution Conference and Exhibition - Latin America (T&D LA), Montevideo, Uruguay*, 2020, pp. 1–6. https://doi.org/10.1109/TDLA47668.2020.9326148

[15] T. Y. Kim and S. B. Cho, "Predicting Residential Energy Consumption Using CNN-LSTM Neural Networks", *Energy*, vol. 182, pp. 72–81, 2019. https://doi.org/10.1016/j.energy.2019.05.230

[16] H. Hua *et al.*, "An Ensemble Framework for Short-term Load Forecasting Based on Parallel CNN and GRU with Improved ResNet", *Electric Power Systems Research*, vol. 216, 2023. https://doi.org/10.1016/j.epsr.2022.109057

[17] Y. Du *et al.*, "Optimization of Magnetically Coupled Resonant Wireless Power Transfer Based on Improved Whale Optimization Algorithm", *Journal of Industrial Intelligence*, vol. 1, no. 1, pp. 63–74, 2023. https://doi.org/10.56578/jii010105

[18] H. Choi *et al.*, "Short-term Load Forecasting Based on Resnet and LSTM", in *Proc. of the 2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), Aalborg, Denmark*, 2018, pp. 1–6. https://doi.org/10.1109/SmartGridComm.2018.8587554

[19] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning", in *Proceedings of the 33rd International Conference on Machine Learning, New York, USA*, 2016, pp. 1050–1059.

[20] M. Khodayar and J. Wang, "Spatio-temporal Graph Deep Neural Network for Short-term Wind Speed Forecasting", *IEEE Transactions on Sustainable Energy*, vol. 10, no. 2, pp. 670–681, 2019. https://doi.org/10.1109/TSTE.2018.2844102

[21] T. Wang, "Power Load Forecasting Based on LSTM Deep Network", Master's thesis, Shanxi University, 2020.

[22] H. Wu and X. Zhu, "Short-term Electric Load Forecasting Model Based on PSO-BP", in *Proc. of the 2023 4th International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering, Hangzhou, China*, 2023, pp. 224–228. https://doi.org/10.1109/ICBAIE59714.2023.10281261

[23] L. J. Zhu *et al.*, "Short-term Power Load Forecasting Based on CNN-BiLSTM", *Power System Technology*, vol. 45, no. 11, pp. 4532–4539, 2021.

[24] J. Chen and J. Zhang, "A Dual Attention-based CNN-GRU Model for Short-term Electric Load Forecasting", *Frontier Academic Forum of Electrical Engineering*, pp. 715–725, 2022. https://doi.org/ 10.1007/978-981-99-3404-1_63

[25] X. Guo *et al.*, "A Short-term Load Forecasting Model of Multi-scale CNN-LSTM Hybrid Neural Network Considering the Real-time Electricity Price", *Energy Reports*, vol. 6, pp. 1046–1053, 2020. https://doi.org/10.1016/j.egyr.2020.11.078

Received: January 2024
Revised: -
Accepted: March 2024

Contact addresses:
Weiwei Mao
College of Computer and Information Engineering
Luoyang Institute of Science and Technology
Luoyang
China
e-mail: mw116@lit.edu.cn

Suping Yu*
College of Computer and Information Engineering
Luoyang Institute of Science and Technology
Luoyang
China
e-mail: badkid@126.com
*Corresponding author

Wenqing Chen
School of Electrical Engineering and Automation
Luoyang Institute of Technology
Luoyang
China
e-mail: cwqlit@lit.edu.cn

WEIWEI MAO was awarded his Master's degree from Sichuan University in Chengdu, China, in 2009. Currently, he is dedicated to teaching and researching in the field of computer science at Luoyang Institute of Technology. He has led numerous scientific research projects at the provincial level. His primary areas of expertise lie in computer application technology and communication technology.

SUPING YU was awarded her Master's degree from Henan University in Kaifeng, China, in 2008. As an Associate Professor, she currently serves in the School of Computer and Information Engineering at Luoyang Institute of Technology, where she is primarily involved in teaching and research within the field of computer science. Her main research focus encompasses graphic imaging and network technology.

WENQING CHEN obtained his PhD from Huazhong University of Science and Technology in Wuhan, China, in 2011. Presently, he holds a professorship at the School of Electrical Engineering and Automation at Luoyang Institute of Technology. His responsibilities include lecturing on automation among other subjects, and his research is principally focused on intelligent control technology and fault diagnosis.