Computing and Informatics, Vol. 43, 2024, 561–582, doi: 10.31577/cai_2024_3_561

LOAD FORECASTING OF SPARROW SEARCH ALGORITHM OPTIMIZATION DOUBLE BIGRU

Di Wu, Lijun Yang, Wenli Ma

Department of Information and Electronic Engineering
Hebei University of Engineering
Handan, Hebei, China
e-mail: wudiwudi@hebeu.edu.cn

Abstract. In this paper, a PCA-SSA-DBIGRU-Attention multi-factor short-term power load forecasting model is proposed. Taking a complete account of the influence of meteorological factors, principal components analysis (PCA) is used to analyze the meteorological factors of daily minimum, maximum, daily average temperature, relative humidity, daily precipitation and power load data at the same time. The realization of original load data is dimensioned down. The complexity of power load forecasting models is reduced. Then, the Attention Double Bidirectional Gating Recurrent Unit (DBIGRU) model is constructed to calculate the different weights of the hidden layer states of the two-layer BIGRU. The hidden layer states are assigned different weights. The Sparrow Search Algorithm (SSA) is incorporated into the DBIGRU-Attention. The SSA-DBIGRU-Attention network model is constructed to optimize the learning rate, the number of iterations and the four hyperparameters of the first and second hidden layer neurons. The extracted principal components are input into SSA-DBIGRU-Attention to realize multi-factor short-term power load forecasting. Experimental results show that the prediction accuracy of the proposed model is improved, and the prediction time is reduced. Compared to the VMD-BILSTM, PCA-DBILSTM, CNN-GRU-Attention and CNN-BIGRU-Attention model, the four aspects of MAPE, MAE, RMSE and time are reduced by 29.55 %, 36.42 %, 32.34 % and 12.22 %, respectively, the R^2 is improved by 3.09 %.

Keywords: Power load forecasting, principal component analysis, sparrow search algorithm, two-layer BIGRU, attention mechanism

1 INTRODUCTION

In recent years, short-term power load forecasting has received much attention from experts in the power industry. It accurately predicts the electricity load for a future period based on data such as historical electricity load and meteorological indicators. A machine learning model with a long short-term memory and factor analysis was proposed by Veeramsetty et al. [1]. A Long Short Term Memory (LSTM) based power demand forecasting model was submitted by Roy et al. [2]. A multi-layer GA-LSTM model was proposed by Kumari et al. [3] for energy prediction. Forecasting of power load using optimal hybrid kernel functions was offered by Liang et al. [4]. Multi-timescale-based integrated long and short-term dual memory (MTS-LSTDM) for power load forecasting models in smart grids was proposed by Lou et al. [5].

In the above literature, power load data is modeled. However, the impact of meteorological factors on power load is not adequately considered. The literature [6] introduced meteorological factors. In [7], the EMD decomposition was used to obtain the correlation between the variation pattern of each component and the candidate influencing factors. Then, short-term load forecasting was achieved. The maximum deviation similarity criterion clustering algorithm with BP networks for predicting short-term power load was proposed by Luo et al. [8]. For factor screening and ultra-short-term power load forecasting, a data processing layer and a load prediction layer based on a two-layer XGBoost algorithm were constructed by Sun et al. [9]. An empirical approach that emphasizes the impact of data updates, climate events, power outages, human activities and public holidays on the overall performance of the model was presented by Mir et al. [10]. An adaptive error trend quadratic learning (A-SLET) for the adaptive trend effects was submitted by Elahe et al. [11].

Meanwhile, the correlation between the multivariate and the predicted results vary from period to period. A prediction model that introduces attention to CNN-GRU was offered by Zhao et al. [12]. A CNN-BIGRU-Attention-based network prediction model was presented by Fang et al. [13].

In response to the problem of hyperparameters having an enormous impact on the accuracy of traditional prediction models, many scholars have used population intelligence algorithms to optimize the hyperparameters. A Particle Swarm Optimization (PSO) optimized neural network forecasting model was presented by Shafiei Chafi and Afrakhte [14]. A short-term power load forecasting model based on improved PSO and neural networks was submitted by Duan et al. [15]. An evolutionary algorithm-based STLF model for intelligent machine learning power systems (IMLEA-STLF) was proposed by Mehedi et al. [16]. The sparrow search algorithm (SSA) with better global and local exploitation capability was proposed by Xue and Shen [17]. Hyperparametric optimization of the BIGRU model using SSA was submitted by Li et al. [18].

Traditional prediction models have high input data dimensionality and training complexity. In [19], the PCA reduces the multiple correlations between data sets. By using optimal support vector machines, Elman neural networks and their combination, prediction models are built. A multi-factor short-term load forecasting model based on PCA-DBILSTM was proposed by Li et al. [20].

Traditional short-term power load forecasting models suffer from inadequate consideration of meteorological influences, high input data dimensionality and difficulties in hyperparameters tuning. A PCA-SSA-DBIGRU-Attention short-term power load forecasting model is proposed. The PCA is used to extract the principal components of multivariate time series. The attention mechanism calculates different weights of the two-layer BIGRU hidden layer states. To optimize multiple hyperparameters of DBIGRU-Attention, the SSA algorithm is incorporated into the DBIGRU-Attention forecasting model.

The rest of the paper is organized as follows. The implementation process of the PCA-SSA-DBIGRU-Attention is introduced in detail in Section 2. Experimental results and analysis are presented in Section 3. Section 4 concludes the paper.

2 PCA-SSA-DBIGRU-ATTENTION MULTI-FACTOR SHORT-TERM FORECASTING MODEL

In this paper, the five meteorological factors of daily minimum, maximum and average temperatures, relative humidity and daily precipitation are fully considered. Therefore a short-term power load forecasting model based on the PCA-SSA-DBIGRU-Attention network is proposed. The meteorological data such as historical power load, daily minimum and maximum temperature are supplemented with missing values and normalized. To achieve the dimensionality reduction of the original data, the PCA is used to extract the principal components of multivariate time series. Then, a two-layer BIGRU network is constructed to thoroughly learn the changing pattern of the data. To calculate different weights for the hidden layer states of the two-layer BIGRU, the attention mechanism is added after the two-layer BIGRU. By using SSA, the hyperparameters in the DBIGRU-Attention are optimized. The prediction results are output through the fully connected layer. The framework of PCA-SSA-DBIGRU-Attention power load forecasting model is shown in Figure 1.

2.1 Input Layer

In the input layer, power load, daily maximum, minimum, average temperature, relative humidity and daily precipitation are the inputs for each time step. The time step sliding window size is T. The input sequence is represented as shown in Equation (1), (2).

$$X = \left[X_{t-T+1}, X_{t-T+2}, \ldots, X_{t-T+i}, X_t\right], \tag{1}$$

$$X_t = load_{t0}, load_{t1}, \ldots, load_{t23}, temp_{maxt}, temp_{mint}, temp_{avert}, rh_t, prcp_t, \tag{2}$$

where $load_{t0}, load_{t1}, load_{t23}$ represent the load at 0, 1 and 23 o'clock on day t, respectively. $temp_{maxt}, temp_{mint}, temp_{avert}, rh_t, prcp_t$ denote the daily maximum, min-

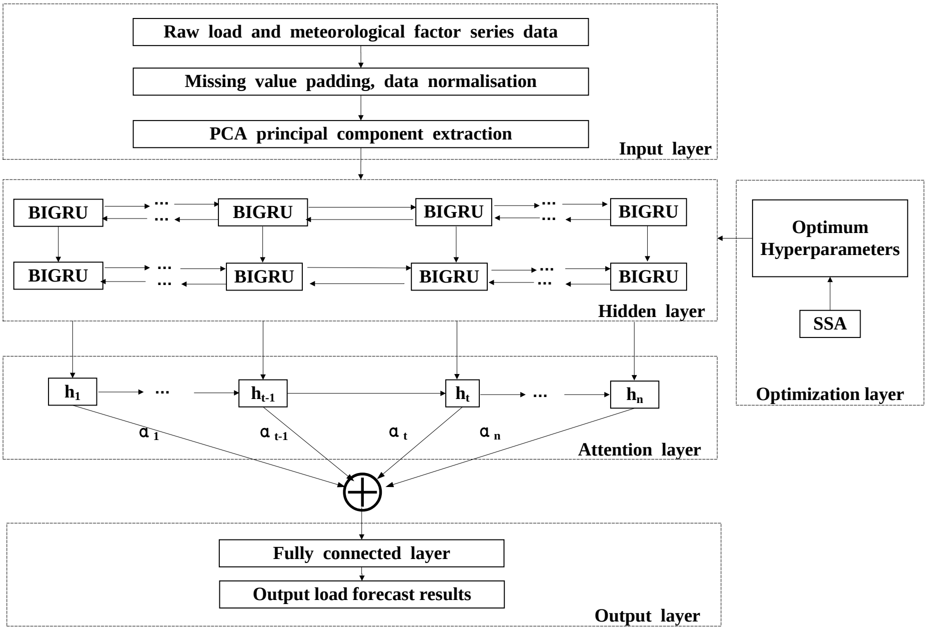

Figure 1. The framework of PCA-SSA-DBIGRU-Attention power load forecasting model

imum, average temperature, relative humidity and daily precipitation on day t. After the missing values are supplemented and normalized, the data is input into the PCA. The dimensionality is reduced from the actual data $X\left[x_1,\ldots,x_{t-1},x_t,\ldots,x_n\right]^T$ to $Y\left[y_1,\ldots,y_{t-1},y_t,\ldots,y_k\right]^T$ $(n>k)$. Then the reduced dimensional information is input into the prediction model.

1) Missing Value Addition. Missing values are clustering, grouping, deletion or truncation of data due to missing information in rough data. If the original data contains missing values, noisy data is produced. Then, the prediction accuracy is affected. Therefore, before the original information is input into the prediction model, the missing values are screened. To remove noisy data, missing values are filled directly with zero.

2) Data Normalization. Due to large fluctuations in time series data such as load, daily maximum temperature and daily precipitation, the training time is long and prediction efficiency is low in the prediction models. So, the data is normalized according to the standard $min-max$. The normalisation process is shown in Equation (3).

$$x^*=\frac{x-min}{max-min}, \tag{3}$$

where x is the original value, min and max represent the minimum and maximum value of the original data, and x^* is the normalized value.

3) PCA Principal Component Extraction. Multivariate time series input to short-term power load forecasting models is somewhat correlated. Therefore, the PCA method extracts the principal components of the multivariate time series [21]. By using new variables instead of the original multiple variables, the dimensionality of the input data to the prediction model is reduced. The effect of superimposed information on the prediction results is eliminated. The complexity of model training is reduced. The prediction accuracy is not affected.

Predictive model training complexity is influenced by the dimensionality of the input data. Therefore, the PCA is used to extract principal component information from the input historical power load, daily maximum, minimum, average temperature and precipitation data, while ensuring the accuracy of the model. Then, the dimensionality of the minimum input variable is determined. The eigenvalues, contribution rates and cumulative contribution rates of each principal component are obtained.

To reduce the dimensionality of the data on power load and meteorological influences, PCA is used. The specific implementation process is as follows:

Step 1: Assume that there are n historical load data in the normalized data $X = [X_1, X_2, \ldots, X_n]$. Each data has P power load and meteorological influence factor data $X_i = [x_1, x_2, \ldots, x_p]$. Calculate the covariance matrix $Z_{P \times P} = \sum_{i,j=1}^{p} Q_{i,j}$ of the normalized data. $Q_{i,j} = \frac{\sum_{k=1}^{n}(x_{ki}-\overline{x_i})\left(x_{kj}-\overline{x_j}\right)}{n-1}$, $\overline{x}$ is the mean of x.

Step 2: Calculate the eigenvalues of the covariance matrix and the corresponding orthogonalized unit eigenvectors. The i^{th} principal component of the original variable is $F_i = \alpha_i X$. $\alpha_i = \frac{\lambda_i}{\sum_{i=1}^{m} \lambda_i}$.

Step 3: The number of principal components m is determined by calculating the variance contribution $\beta_i = \frac{100\,\%\,\lambda_i}{\sum_{i=1}^{p} \lambda_i}$ and the cumulative variance contribution. Then, the scores $F_i = \alpha_{1i}X_1 + \alpha_{2i}X_2 + \cdots + \alpha_{pi}X_p$ of power load and meteorological data on the m principal components are calculated. α_{ij} represents the eigenvector corresponding to the eigenvalues of the covariance matrix.

2.2 Hidden Layer

The hidden layer is composed of two BIGRU power load forecasting network units. The GRU is a unidirectional neural network that always propagates in an orderly sequence from front to back. The GRU is described in detail in the literature [22]. The BIGRU can be considered structurally as a combination of forward and backward propagating GRU. The implicit state h_t at the current moment is determined by a combination of three components. The three sections are the implicit output $\overleftarrow{h_{t-1}}$ at the moment $t-1$ along the time forward propagation, the hidden layer output $\overrightarrow{h_{t-1}}$ at the moment $t-1$ along the time backward propagation and the input x_t at the current moment.

At moment t, the process of calculating the implicit state of each layer of the BIGRU is shown in Equation (4).

$$\begin{cases} \overrightarrow{h_t} = GRU\left(x_t, \overrightarrow{h_{t-1}}\right), \\ \overleftarrow{h_t} = GRU\left(x_t, \overleftarrow{h_{t-1}}\right), \\ h_t = \alpha_t \overrightarrow{h_t} + \beta_t \overleftarrow{h_t} + b_t, \end{cases} \tag{4}$$

where α_t and β_t are the output weight of the forward and backward propagating GRU unit implicit layer at the moment t, h_t represents the implicit layer state at the moment t and b_t denotes the h_t corresponding bias amount.

The single-layer BIGRU can learn the impact of past and future data on the current power load. So, deep-level features of load data are beneficial to be extracted. However, historical load and meteorological data cannot be fully learned by a single-layer of BIGRU. Meanwhile, complex time series type data is processed with unsatisfactory performance. Therefore, a two-layer BIGRU power load forecasting model is constructed. The structure of DBIGRU is shown in Figure 2.

2.3 Attention Layer

To assign weights to the output of the hidden layer, the attention layer uses the attention mechanism. The attention mechanism mimics the resource allocation mechanism of human visual attention. In terms of network structure, the attention mechanism learns a distribution of weights on data features by focusing on the data input. Then, the learned weights are applied to the original data features. A salient feature impact is provided for subsequent processing of the data so that the focused features receive greater attention.

To dig deeper into the historical load curve features, probabilistic assignment weights are used instead of random assignment weights. To highlight the role of important information, different weights are given to the implicit layer of DBIGRU through the attention mechanism. Attention is used to calculate the different weights of the implicit layer states of a two-layer BIGRU network. Time series state features of historical load data are efficiently learned. Historical time series status weights are obtained. Critical information is provided to forecast. Therefore, the accuracy of short-term load forecasting for power is improved. The structure of the power load feature weight assignment incorporating attention is shown in Figure 3.

X_t represents the load data after dimensionality reduction at the moment t. h_t denotes the implicit layer state of the hidden layer at the moment t. At the moment t, the attention weight $\alpha_t = \frac{exp(e_t)}{\sum_{k=1}^{t} exp(e_k}$ of the implicit layer state h_t is calculated by dot product form. e_t denotes the value of the attention probability distribution determined by h_t at the moment t. $e_t = u \tanh(\omega h_t + b)$. The input of the attention layer is the implicit layer state h_t and the attention weights α_t of the hidden layer DBIGRU network. The output of the attention layer is $R_t = \sum_{t=1}^{i} \alpha_t h_t$. u and ω are the weight coefficients. b denotes the bias coefficient.

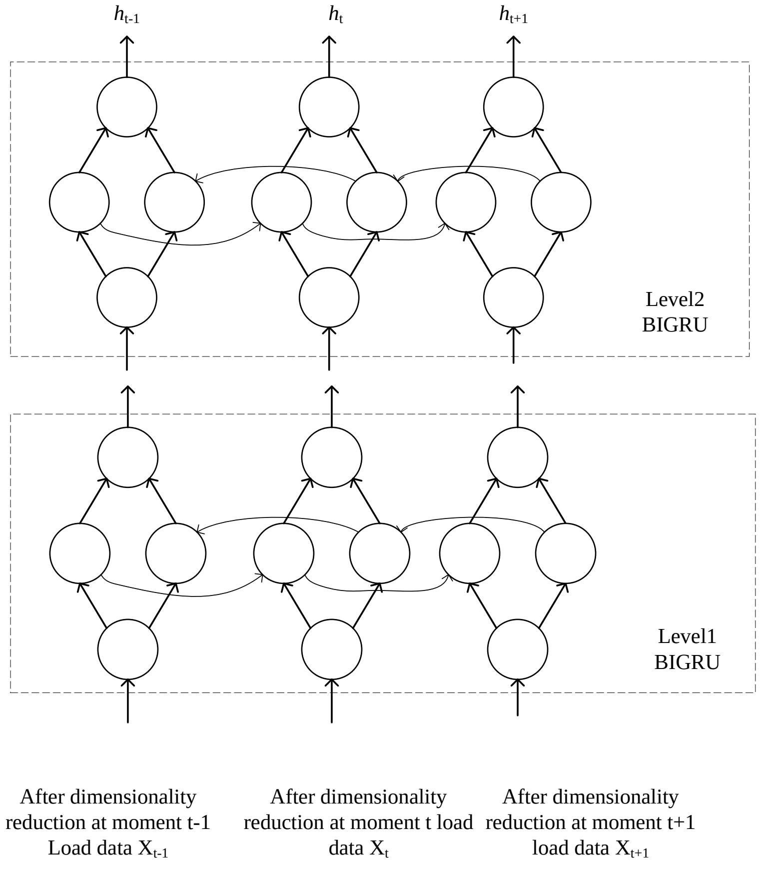

Figure 2. Structure of DBIGRU power load forecasting model

2.4 Optimization Layer

In the optimization layer, the hyperparameters in the two-layer BIGRU are optimized by SSA. In neural networks, parameters such as the number of hidden layers, units, iterations and batch size are called hyperparameters. Structural parameters of BIGRU networks are essential for training and predicting samples. Optimization of essential hyperparameters can improve the accuracy of prediction models.

Recently, SSA is a population intelligence optimization algorithm that simulates the process of sparrow flock foraging. Sparrow populations are divided into three types depending on their duties. Finders are responsible for searching for food. Followers are accountable for following the finders to forage. Scouts are in charge of vigilance detection.

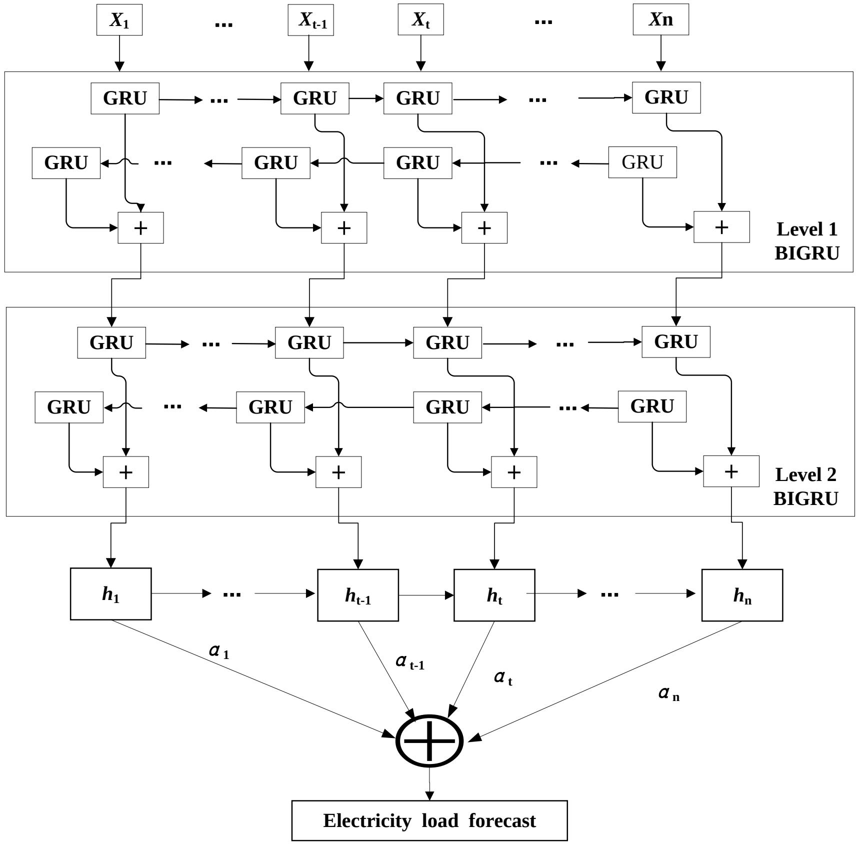

Figure 3. Structure of Attention power load feature weight distribution

The steps of SSA are in Algorithm 1.

The optimization process of SSA for the DBIGRU-Attention network is shown in Figure 4. The optimization parameters include the learning rate, the number of iterations, neurons in the first and second hidden layers.

The SSA parameters are initialized. The number of hyperparameters of the BIGRU network is taken as the dimension of the sparrow search. The BIGRU network hyperparameters $pop = (pop_0, pop_1, pop_2, pop_3)$ are encoded as the initial position vector of individual sparrows $X_{i,j} = (x_{i1}, x_{i2}, x_{i3}, x_{i4})$. $pop_0 - pop_3$ correspond to the learning rate of BIGRU, the number of iterations and neurons in the first and second hidden layers. $X_{i,j}$ denotes the j^{th} position of the i^{th} sparrow in the four-dimensional space. The fitness value of the sparrow is calculated. The individual part of the sparrow is updated and compared with the historical optimum. If the stopping condition is satisfied, the iteration is terminated. Otherwise, the iteration continues.

Algorithm 1: SSA

Input: Initial values for learning rate, number of iterations, neurons in the two BIGRU hidden layers of the optimal upper and lower bounds (l_r, M, h_{n1}, h_{n2})

Output: Optimal values for learning rate, number of iterations and the first, second hidden layer neurons

Step 1: Set the maximum number of iterations, the number of sparrow populations, discoverers, detectors and warning values, form a search space matrix and initialize the relevant hyperparameters.

Step 2: Initial adaptation calculation, find the best and worst adapted individual.

Step 3: Update the location of finders, followers and scouts.

Step 4: Compare fitness values and update sparrow global optimal position.

Step 5: Determine if the maximum number of iterations is reached and the optimal parameters are output, otherwise repeat Steps 2–4.

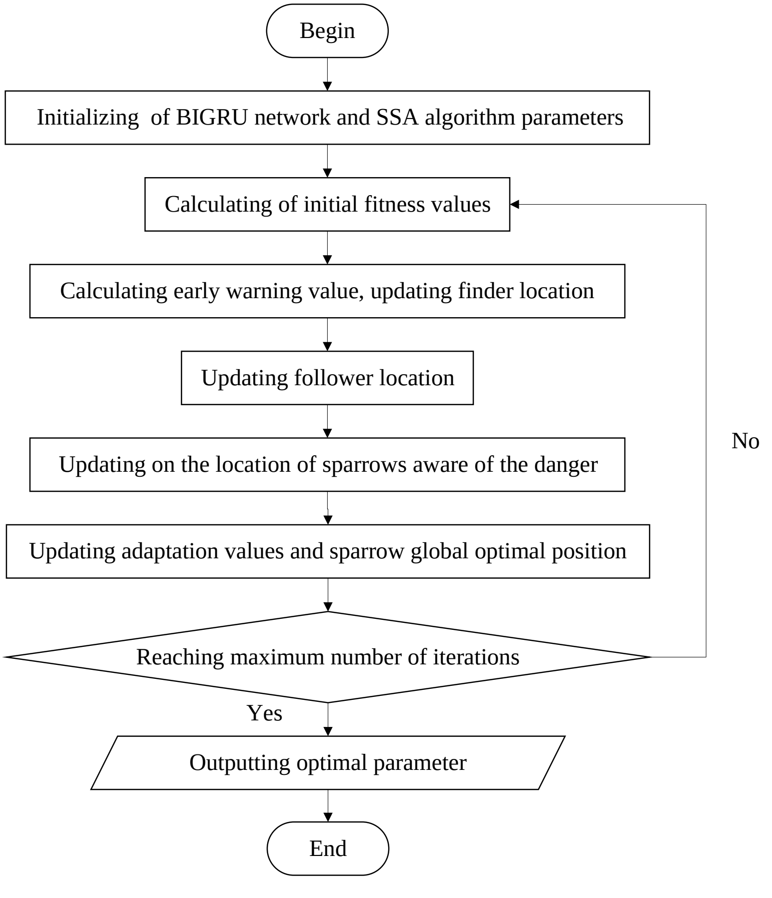

Figure 4. Flow chart of SSA optimised DBIGRU-Attention network

After several iterations, the global optimal point is searched. To obtain the optimal hyperparameters of the BIGRU network, the optimal solution after SSA optimization is decoded and transformed. Then, the optimal hyperparameters are assigned to the BIGRU network. The optimized BIGRU network is used to train and predict the input vectors.

2.5 Output Layer

The output of the attention layer is used as the input of the Output layer. The output layer calculates the prediction result $Y = \{y_1, y_2, \ldots, y_n\}^T$ through the fully connected layer. The fully connected layer selects the sigmoid as the activation function. Then, the power load value y_t at the moment $t+1$ is output. The y_t is calculated as shown in Equation (5).

$$y_t = sigmod\,(\omega R_t + \alpha), \tag{5}$$

where y_t is the predicted output value at time t. ω denotes the weight matrix. α represents the deviation vector.

The steps of the PCA-SSA-DBIGRU-Attention model are as follows:

Step 1: Fill and normalize the data with missing values. The data includes historical daily whole hours load for the previous N days $x = x_1, x_2, \ldots, x_{23}$, influencing factors (daily maximum, minimum, average temperature, relative humidity, daily precipitation). So that the data is distributed in the interval $[0, 1]$ for subsequent model training.

Step 2: Extract the principal components of the normalized data.

Step 3: Build a two-layer BIGRU neural network model. To learn the temporal state features of the historical data, attention mechanism is added to the model. The historical material state weights are obtained.

Step 4: Incorporate the SSA into DBIGRU-Attention. Therefore, the necessary hyperparameters are optimized and the optimal hyperparameters are obtained.

Step 5: Output the inverse normalized prediction results. Whole hours load for day $N+1$ $y = y_1, y_2, \ldots, y_{23}$.

A PCA-SSA-DBIGRU-Attention short-term power load forecasting model is proposed in this paper. In the input layer, the PCA is used to perform principal component extraction on multivariate time series. The complexity of the power load forecasting model is reduced. The hidden layer consists of BIGRU power load forecasting network units. In the Attention layer, the Attention mechanism is used to compute different weights for the hidden layer states of the two-layer BIGRU. Different weights are assigned to the hidden layer states. In the Optimisation layer, SSA is incorporated into the DBIGRU-Attention prediction model and DBIGRU-Attention multiple hyperparameters are optimised. In the fully-connected layer, the prediction value at each moment is calculated by the sigmoid activation function.

In summary, Principal Components Analysis Algorithm, Hyper Parameter Optimisation Algorithm Sparrow Search Algorithm and Attention Mechanism are incorporated into the DBIGRU Algorithm in order to improve the power load forecasting accuracy, speed and fitting effectiveness.

3 EXPERIMENTAL RESULTS AND ANALYSIS

3.1 Dataset and Experimental Environment Setup

The experiments are conducted in Win 10 with Python programming language and Kear's framework. In this paper, the short-term load dataset of the power system in Group A of the 9^{th} Electrical Property Modelling Competition test in 2016 is used. The sampling start time is from January 1, 2012, to January 10, 2015, with more than one thousand data. The ratio of training and test set in the data is 7:3. The original data contains the total 0-24 o'clock power load value of a day, daily maximum, minimum, average temperature, relative humidity and daily precipitation. The input for the time step is ten days of 290-dimensional data. The output is the load data of the next day. The dataset attributes and some of the dataset are shown in Tables 1 and 2.

Attribute Serial Number	Attribute Name
1–24	0–23 hour load
25	Maximum temperature
26	Minimum temperature
27	Average temperature
28	Relative humidity
29	Precipitation

Table 1. Dataset attribute table

YMD	Maximum Temperature [°C]	Minimum Temperature [°C]	Average Temperature [°C]	Relative Humidity [%]	Precipitation [mm]
20120104	14.9	9.3	10.9	62	0
20120105	9.2	5.1	6.9	78	2.9
20120106	11	6.3	8.2	92	3.3
20120107	12.4	9.4	10.5	80	0

Table 2. Partial dataset table

3.2 Evaluation Indicators

In this paper, four prediction performance evaluation metrics are selected. The four are absolute percentage error (MAPE), mean absolute error (MAE), root mean square error (RMSE), and goodness of fit (R^2). They are calculated as shown in Equations (7), (8), (8) and (9).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100\,\%, \tag{6}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{7}$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n}, \tag{8}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\bar{y}_i - y_i)^2}. \tag{9}$$

The smaller the MAPE, MAE and RMSE, the higher the prediction accuracy. The closer R^2 to 1, the more accurate the prediction. y_i and $\hat{y}_i$ are the actual and predicted load at the i^{th} sampling point on the prediction day. n is the total number of predicted sampling points.

3.3 Multi-Factor Short-Term Forecasting Model Results and Analysis

The data input to the prediction model is highly dimensional. The data are supplemented with missing values and normalized. Then, the covariance matrix of the normalized data, the eigenvalues, contribution rates and cumulative contribution rates of the covariance matrix are calculated. The principal components are extracted based on the cumulative contribution rates ($\geq 98\,\%$). The calculation results are shown in Table 3 and Figure 5.

Principal Components	Eigenvalue	Contribution of Variance [%]	Cumulative Contribution Rate [%]
0	2.8413	71.996	71.996
1	0.2733	6.9264	78.9224
2	0.2242	5.6819	84.6043
3	0.1867	4.7331	89.3374
4	0.0979	2.4823	91.8196
5	0.0944	2.3925	94.2121
6	0.0678	1.718	95.9302
7	0.0446	1.1302	97.0604
8	0.0423	1.0719	98.1323

Table 3. Sample variable eigenvalue and variance contribution rate

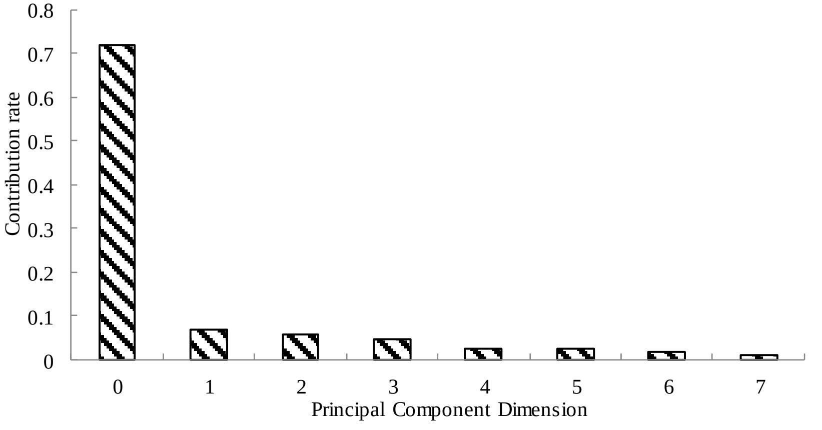

Figure 5. Principal components and contribution rate

As shown in Table 3 and Figure 5, the first eight principal components can be used as the input of the prediction model instead of the original data. Therefore, the first eight principal components are selected to replace the original input data for SSA-DBIGRU-Attention network model training. Using the PCA method, the dimensionality of data is reduced while ensuring maximum information retention. Prediction model training complexity and prediction time are reduced.

By using the PCA method, the original load data is dimensionally reduced to obtain principal components. Then, the principal components are input into the SSA-DBIGRU-Attention model for prediction. The single-layer BIGRU prediction model does not sufficiently learn the feature patterns of the data. Therefore, a two-layer BIGRU network is selected. In the sparrow search algorithm, T is set to 0.8, the learning rate l_r, the number of iterations M and two BIGRU layer nodes h_{n1}, h_{n2} of the optimal upper and lower bounds are set to $[0.001, 1, 1, 1]$ and $[0.01, 50, 100, 100]$, respectively. To find the optimal hyperparameters efficiently and improve the prediction accurately, the SSA is incorporated into the DBIGRU-Attention network model. The optimal hyperparameters are obtained, as shown in Table 4.

Serial Number	Hyperparameters	Hyperparameter Values
1	Learning Rate	0.0049
2	Number of iterations	48
3	First hidden layer neurons	28
4	Second hidden layer neurons	72

Table 4. Sample variable eigenvalue and variance contribution rate

The optimization process of PCA-DBIGRU-Attention network hyperparameters is shown in Figures 6, 7, 8 and 9. Four hyperparameters are optimized and T is

Figure 6. Curve of learning rate with iteration times

Figure 7. The process of finding the optimal iteration value

set to 0.8. Figure 6 shows the optimal learning rate search process. The learning rate is stable and constant at first, then increase gradually. The second iteration reaches a plateau, the sixth iteration starts to gradually increase and the seventh iteration reaches stability with a learning rate of 0.0049. Figure 7 shows the optimal search process of iterative value. At the start of the seventh iteration, a plateau is reached and the optimal iteration value is 48. Figure 8 shows the first hidden layer node search process. The number of nodes in the first hidden layer reaches stability at the beginning of the seventh iteration. The optimal number of neurons is 28. Figure 9 shows the second hidden layer node search process. The number of nodes in the second hidden layer is stable from the seventh iteration. The optimal number of neurons is 72.

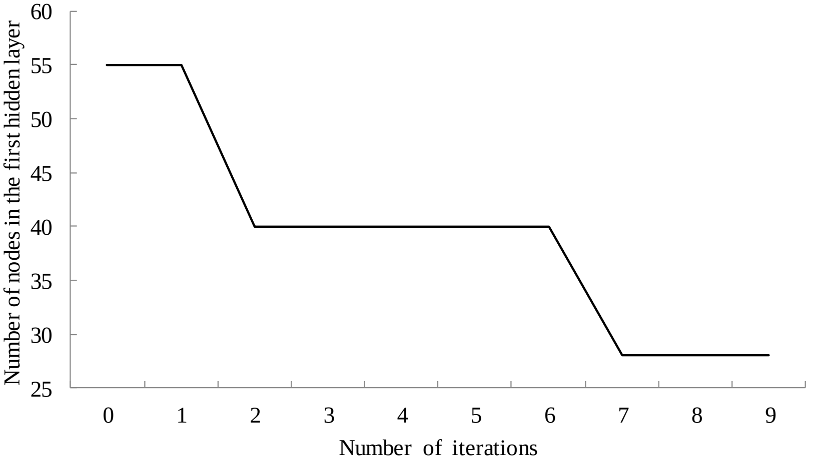

Figure 8. The process of finding the optimal number of nodes in the first hidden layer

Figure 9. The process of finding the optimal number

As shown in Figure 10, the difference between the prediction effect of this model and the real value is slight. And it has a high degree of match. The above experiments verify that the proposed model has good applicability for short-term power load prediction.

3.4 Model Comparison

To verify the effectiveness of the proposed model in terms of forecasting accuracy and time, the comparison experiments are conducted with the following four existing models.

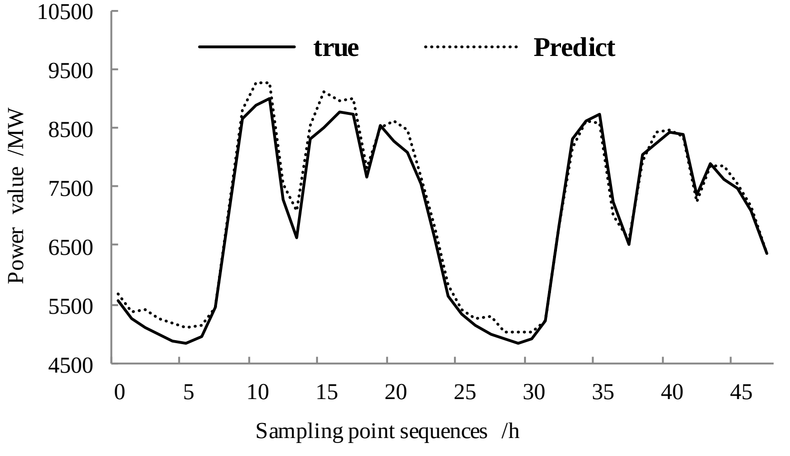

Figure 10. The result of the PCA-SSA-DBIGRU-Attention power load forecasting model

1) **VMD-BILSTM [20]:** By using the VMD, the load sequence is decomposed into a set of sub-series components. A BILSTM-based time series prediction model is constructed for each sub-series. Bayesian theory is used to optimize sub-series correlation hyperparameters. The load predictions are obtained by superimposing the forecasts of each subseries.
2) **PCA-DBILSTM [5]:** Multi-layer BILSTM in PCA-DBILSTM is set to two layers. PCA-DBILSTM uses PCA to extract principal components from a time series consisting of original multidimensional input variables. Initial load dimensionality reduction is achieved. Using the DBILSTM algorithm, the non-linear relationship between the extracted principal component and the actual output sequence of the load is modeled for network prediction.
3) **CNN-GRU-Attention [12]:** CNN is used to extract high-dimensional features. The proposed feature vectors are input into the GRU network. To give different weights to the implied states of the GRU, the attention mechanism is introduced. The loss of historical information is reduced and the impact of important information is enhanced. The short-term load prediction is completed.
4) **CNN-BIGRU-Attention [13]:** CNN is used for load and weather data features extraction. The latent timing patterns are extracted using BIGRU. To highlight key features, the Attention mechanism is incorporated into the CNN-BIGRU model. The prediction results are output.

The MAPE, MAE, RMSE, R^2 and time results of the prediction model comparison experiments are shown in Table 5.

From the analysis of the prediction results, the proposed model has the highest prediction accuracy and the shortest prediction time compared to the other four models. Compared to the other four models, MAPE is reduced by 35.19 %, 50.36 %,

Models	MAPE [%]	RMSE	MAE	R^2	Time [ms]
VMD-BILSTM	3.85	320.817	281.892	0.9506	923
PCA-DBILSTM	5.18	455.527	368.011	0.9004	90
CNN-GRU-Attention	3.64	345.028	274.269	0.9428	355
CNN-BIGRU-Attention	3.35	326.788	242.417	0.9487	557
PCA-SSA-DBIGRU-Attention	2.36	203.968	164.023	0.9800	79

Table 5. Comparison of prediction performance indicators of different models

19.30 %, 21.55 %. RMSE is reduced by 36.42 %, 55.22 %, 40.88 %, 37.58 %. MAE is reduced by 41.81 %, 55.43 %, 40.20 %, 32.34 %, respectively. The R^2 is closest to 1. In contrast to the VMD-BILSTM and PCA-DBILSTM models, the model proposed in this paper utilises the Attention mechanism to calculate different weights for the hidden layer states of the BIGRU network. Larger weights are given to the variables that have a greater impact on the prediction results. The model prediction accuracy is effectively improved. In contrast to the CNN-GRU-Attention model, the two-layer BIGRU network is adopted by the PCA-SSA-DBIGRU-Attention model. Data feature laws are fully learnt. In contrast to the CNN-GRU-Attention and CNN-BIGRU-Attention models, hemp SSA is incorporated into the PCA-SSA-DBIGRU-Attention model. The hyperparameters of the two-layer BIGRU model are optimised. The optimal value is found. Model prediction accuracy is effectively improved. The time is reduced by 91.44 %, 12.22 %, 77.75 %, and 85.82 % compared to the other four models, mainly due to the fact that PCA is employed in this model before the data is fed into the prediction model. The original high dimensional data was reduced to low dimensions. The prediction time of the prediction model was effectively reduced.

The reasons for their generation are described below in terms of both prediction time and accuracy. The experimental results are shown in Table 5 and Figure 11.

In terms of prediction time, among the five prediction models, VMD-BILSTM has the longest prediction time. Firstly, the original sequence is decomposed by the VMD. Each dimensional data is raised to multiple dimensions. Then the subsequences are modeled separately. And the prediction results are overlaid and reconstructed. Hence the prediction time overhead is not ideal. When extracting high dimensional features reflecting complex dynamic changes of load, CNN-GRU-Attention and CNN-BIGRU-Attention models build a CNN architecture consisting of convolutional and pooling layers. Then, the time series is constructed by the proposed feature vector. And It is input into the GRU and BIGRU network prediction models. Although the forecast time overhead has improved in the two models above, the forecast time is relatively long. Without affecting the prediction effect, principal component analysis is used to reduce the original high-dimensional data to low-dimensional. The prediction time of the forecasting model is effectively reduced.

Regarding prediction accuracy, the model PCA-DBILSTM uses the PCA to achieve dimensionality reduction of the data. And, the prediction time is reduced. However, the BILSTM model considers all input data to be of equal importance to the prediction result. The impact of important data information is not highlighted. Thus, the prediction accuracy is not high. The proposed model uses the attention mechanism to calculate the different weights of the hidden layer states of the BIGRU network. The information entered is selectively focused. Larger weights are given to variables with more significant predictive impact. The model prediction accuracy is effectively improved. The other four comparison models have more hyperparameters. Therefore, it is more difficult to find the optimal value of hyperparameters manually. Thus, the prediction effect is compromised and the prediction accuracy is not high. The proposed model incorporates a sparrow optimization algorithm in the two-layer BIGRU prediction process. It simulates the foraging process of a flock of sparrows. The hyperparameters of the two-layer BIGRU model are optimized. Until the optimal value is found, the model prediction accuracy is effectively improved.

According to a comprehensive analysis, the indicators of MAPE, MAE, RMSE and time all decreased significantly, and the indicator of R^2 increased. It shows that the prediction error of the proposed model is small and the goodness of the fitting result is closer to 1. To display the load prediction results more clearly and intuitively, it compares the actual values and the short-term load prediction curves of different models in the last two days in Figure 11.

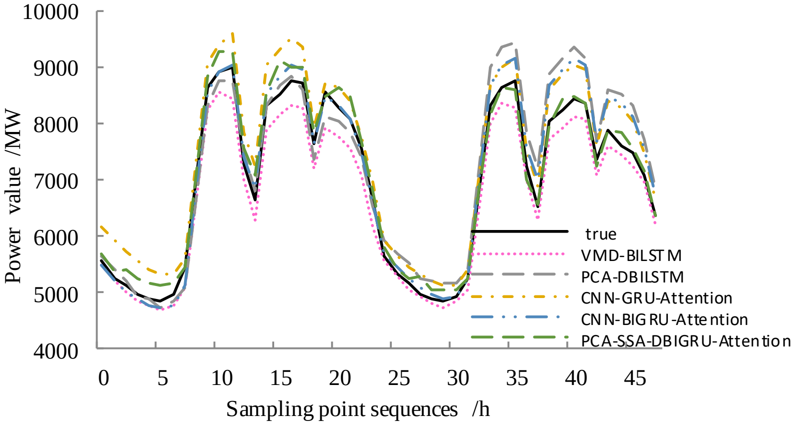

Figure 11. Comparison of results of different power load forecasting models

As it can be seen from Figure 11, the proposed model fitting has better results and a higher prediction accuracy. This dataset has a significant variation in daily load. In the morning, the load value changes significantly. Compared with other models, this model can predict the load value more accurately and smoothly now. It can also capture the load change pattern better near each peak and trough. By using the PCA, the original data is dimensioned down to obtain updated data. Then, the updated information is input into the DBIGRU forecasting model. To calculate the different weights of the implicit layer states of the BIGRU network, the attention mechanism is added after the DBIGRU. The information entered is selectively focused on. Larger weights are assigned to variables with a higher predictive impact. Then the sparrow search algorithm is incorporated into the DBIGRU-Attention to obtain the optimal hyperparameters. So the optimal forecasting results are obtained. As shown in Table 4 and Figure 11, compared with the other four forecasting models, the proposed model fitting results better, with higher prediction accuracy and shorter prediction time.

4 CONCLUSIONS

Traditional short-term power load forecasting models suffer from low accuracy and long forecasting times. A PCA-SSA-DBIGRU-Attention power load forecasting model is proposed. In the input layer, the PCA extracts principal components from the normalized data. The model training complexity is effectively reduced. The attention mechanism is used to assign weights to the DBIGRU output content. The more important the data, the more weight it will be given. In the output data, the richness of the properties and regularities are effectively learned. The SSA algorithm is incorporated into the DBIGRU-Attention model. Then, optimal hyperparameters of the model are obtained. In the fully connected layer, the predicted values are calculated at each moment through the sigmoid activation function. Power load forecasting accuracy and speed are improved. Better fitting results are obtained. Under the short-term load dataset of the power system in Group A of the 9^{th} Electrical Property Modelling Competition test in 2016, the PCA-SSA-DBIGRU-Attention model proposed in this paper shows higher accuracy in prediction error metrics compared to PCA-DBILSTM, VMD-BILSTM, CNN-GRU-Attention, and CNN-BIGRU-Attention power load forecasting models. The MAE, MAPE and RMSE are reduced by 32.34 %, 19.30 % and 36.42 % at the lowest, respectively. The MAE, MAPE and RMSE are reduced by 55.43 %, 50.36 % and 55.22 % at the highest, respectively. Compared with other models, The R^2 is closest to 1. The R^2 improved by a minimum of 3.09 % and a maximum of 8.84 %. The prediction time is reduced by a minimum of 12.22 % and a maximum of 91.44 %.

REFERENCES

[1] Veeramsetty, V.—Chandra, D. R.—Salkuti, S. R.: Short-Term Electric Power Load Forecasting Using Factor Analysis and Long Short-Term Memory for Smart Cities. International Journal of Circuit Theory and Applications, Vol. 49, 2021, No. 6, pp. 1678–1703, doi: 10.1002/cta.2928.

[2] Roy, K.—Ishmam, A.—Taher, K. A.: Demand Forecasting in Smart Grid Using Long Short-Term Memory. 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), 2021, pp. 1–5, doi: 10.1109/ACMI53878.2021.9528277.

[3] Kumari, S.—Kumar, N.—Rana, P. S.: Big Data Analytics for Energy Consumption Prediction in Smart Grid Using Genetic Algorithm and Long Short Term Memory. Computing and Informatics, Vol. 40, 2021, No. 1, pp. 29–56, doi: 10.31577/cai_2021_1_29.

[4] Liang, L.—Huang, W.—Dong, Z.—Zhao, J.—Li, P.—Lu, B.—Zhu, X.: Short-Term Power Load Forecasting Based on Combined Kernel Gaussian Process Hybrid Model. E3S Web of Conferences, Vol. 256, 2021, doi: 10.1051/e3sconf/202125601009.

[5] Lou, Y.—Huang, Y.—Xing, X.—Cao, Y.—Wang, H.: MTS-LSTDM: Multi-Time-Scale Long Short-Term Double Memory for Power Load Forecasting. Journal of Systems Architecture, Vol. 125, 2022, Art. No. 102443, doi: 10.1016/j.sysarc.2022.102443.

[6] Deng, A.—Jin, M.: Time-Scale Feature Extraction Based on EMD and Its Application on Short-Term Power Load Forecasting. Application Research of Computers, Vol. 35, 2018, No. 10, pp. 2952–2955 (in Chinese).

[7] Li, K.—Huang, W.—Hu, G.—Li, J.: Ultra-Short Term Power Load Forecasting Based on CEEMDAN-SE and LSTM Neural Network. Energy and Buildings, Vol. 279, 2023, Art. No. 112666, doi: 10.1016/j.enbuild.2022.112666.

[8] Luo, Y.—Cai, Y.—Qi, Y.—Huang, H.: Short-Term Power Load Forecasting Algorithm Based on Maximum Deviation Similarity Criterion BP Neural Network. Application Research of Computers, Vol. 36, 2019, No. 11, pp. 3269–3273, doi: 10.19734/j.issn.1001-3695.2018.05.0293 (in Chinese).

[9] Sun, C.—Lü, Q.—Zhu, S.—Zheng, W.—Cao, Y.—Wang, J.: Ultra-Short-Term Power Load Forecasting Based on Two-Layer XGBoost Algorithm Considering the Influence of Multiple Features. High Voltage Engineering, Vol. 47, 2021, No. 8, pp. 2885–2898, doi: 10.13336/j.1003-6520.hve.20210172 (in Chinese).

[10] Mir, A. A.—Khan, Z. A.—Altmimi, A.—Badar, M.—Ullah, K.—Imran, M.—Kazmi, S. A. A.: Systematic Development of Short-Term Load Forecasting Models for the Electric Power Utilities: The Case of Pakistan. IEEE Access, Vol. 9, 2021, pp. 140281–140297, doi: 10.1109/ACCESS.2021.3117951.

[11] Elahe, M. F.—Jin, M.—Zeng, P.: An Adaptive and Parallel Forecasting Strategy for Short-Term Power Load Based on Second Learning of Error Trend. IEEE Access, Vol. 8, 2020, pp. 201889–201899, doi: 10.1109/ACCESS.2020.3036096.

[12] Zhao, B.—Wang, Z.—Ji, W.—Gao, X.—Li, X.: A Short-Term Power Load Forecasting Method Based on Attention Mechanism of CNN-GRU. Power System Technology, Vol. 43, 2019, No. 12, pp. 4370–4376, doi: 10.13335/j.1000-3673.pst.2019.1524 (in Chinese).

[13] Fang, N.—Yu, J.—Li, J.—Wang, C.: Short-Term Power Load Forecasting Based on CNN-BIGRU-ATTENTION. Computer Simulation, Vol. 38, 2022, No. 2, pp. 40–44 (in Chinese).

[14] Shafiei Chafi, Z.—Afrakhte, H.: Short-Term Load Forecasting Using Neural Network and Particle Swarm Optimization (PSO) Algorithm. Mathematical Problems in Engineering, Vol. 2021, 2021, Art. No. 5598267, doi: 10.1155/2021/5598267.

[15] Duan, X.—Li, X.—Chen, W.: Improved Particles Swarm Optimization Algorithm-Based VMD-GRU Method for Short-Term Load Forecasting. Advanced Technology of Electrical Engineering and Energy, Vol. 41, 2022, No. 5, pp. 8–17 (in Chinese).

[16] Mehedi, I. M.—Bassi, H.—Rawa, M. J.—Ajour, M.—Abusorrah, A.—Vellingiri, M. T.—Salam, Z.—Abdullah, M. P. B.: Intelligent Machine Learning with Evolutionary Algorithm Based Short Term Load Forecasting in Power Systems. IEEE Access, Vol. 9, 2021, pp. 100113–100124, doi: 10.1109/ACCESS.2021.3096918.

[17] Xue, J.—Shen, B.: A Novel Swarm Intelligence Optimization Approach: Sparrow Search Algorithm. Systems Science and Control Engineering, Vol. 8, 2020, No. 1, pp. 22–34, doi: 10.1080/21642583.2019.1708830.

[18] Li, X.—Ma, X.—Xiao, F.—Xiao, C.—Wang, F.—Zhang, S.: Time-Series Production Forecasting Method Based on the Integration of Bidirectional Gated Recurrent Unit (Bi-GRU) Network and Sparrow Search Algorithm (SSA). Journal of Petroleum Science and Engineering, Vol. 208, 2022, Art. No. 109309, doi: 10.1016/j.petrol.2021.109309.

[19] Qian, K.—Wang, X.—Yuan, Y.: Research on Regional Short-Term Power Load Forecasting Model and Case Analysis. Processes, Vol. 9, 2021, No. 9, Art. No. 1617, doi: 10.3390/pr9091617.

[20] Li, Z. W.—Hu, R.—Liu, X.—Deng, Y.—Tang, P.—Wang, Y.: Multi-Factor Short-Term Load Prediction Model Based on PCA-DBILSTM. Proceedings of the CSU-EPSA, Vol. 32, 2020, No. 12, pp. 32–39, doi: 10.19635/j.cnki.csu-epsa.000501 (in Chinese).

[21] Wang, Y.—Xie, D.—Wang, X.—Li, G.—Zhu, M.—Zhang, Y.: Prediction of Interaction Between Grid and Wind Farms Based on PCA-LSTM Model. Proceedings of the CSEE, Vol. 39, 2019, No. 14, pp. 4070–4081, doi: 10.13334/j.0258-8013.pcsee.181221 (in Chinese).

[22] Zheng, C.—Wu, Y.—Chen, Z.—Wang, K.—Zhang, L.: A Load Forecasting Method of Power Grid Host Based on SARIMA-GRU Model. In: Cai, Z., Li, J., Zhang, J. (Eds.): Theoretical Computer Science (NCTCS 2021). Springer, Singapore, Communications in Computer and Information Science, Vol. 1494, 2021, pp. 135–153, doi: 10.1007/978-981-16-7443-3_9.

Di Wu received her B.Sc. and M.Sc. degrees in computer application technology from the Hebei University of Engineering, and her Ph.D. degree in computer application technology from the Yanshan University. She is currently Associate Professor with the School of Software Engineering, Hebei University of Engineering, China. Her research interests include data mining, natural language processing and information retrieval.

Lijun Yang is currently pursuing the Master's degree with the School of Information and Electronic Engineering, Hebei University of Engineering, China. Her research interests include data mining and natural language processing.

Wenli Ma is currently pursuing the Master's degree with the School of Information and Electronic Engineering, Hebei University of Engineering, China. Her research interests include data mining and natural language processing.