Load Forecast of Rail Transit Power System based on Dropout Echo State Network

1st Xiaohui Mu
department of automation
Tsinghua University
Beijing, China
muxiaohui1991@163.com

2nd Junwei Cao
Beijing National Research Center for Information Science and Technology
Tsinghua University
Beijing, China
jcao@tsinghua.edu.cn

***Abstract*—The power supply and peak shaving capacity of traditional rail transit are easily limited. Ensuring the stable operation of the power system and high-precision load forecasting of the power system are very important for the operation and scheduling of the rail transit power system. This paper uses Dropout echo state network(ESN) to predict short-term load forecasting of rail transit power systems. This method avoids the problems of slow convergence speed and easy falling into local optima encountered by traditional neural networks in load forecasting, solves overfitting problems of traditional neural networks, ensuring the stable operation of the rail transit power system. This paper compares the performance of original ESN and Dropout ESN in load forecasting. The simulation results show that Dropout ESN have higher prediction accuracy in load forecasting of rail transit power system.**

Keywords— artificial intelligence, Dropout echo state network, rail transit, load forecast, power system

I. Introduction

The rail transit power system is a complex system. The power grid regulation system is used for its safe and reliable operation. However, as the power grid structure continues to expand, the power big data presents the characteristics of massive multi-source heterogeneity [1]. The power supply and peak shaving capacity of rail transit are easily limited [2]. In order to ensure the stable operation of the power system, high precision power system load forecasting is crucial for operation scheduling [3]. With the development of artificial intelligence technology, neural networks have made significant progress in feature prediction of high-dimensional complex data. This paper will use neural network technology to predict the massive load data of the rail transit power system.

There are many studies using artificial intelligence technology for load forecasting. Lu et al. used convolutional neural networks for feature extraction, combined with short-term and short-term memory network method for prediction [4]. Chen et al. based on the gate controlled cyclic unit method and adopts an improved sparrow search method to improve search ability [5]. Xu et al. adopted the extended random forest algorithm to achieve controllable load state prediction, and combines it with cloud platforms to achieve precise load regulation [6]. Yin et al. combined the chaotic crossbar particle swarm optimization algorithm with the learning machine method to improve the generalization ability and accuracy of prediction [7]. Miao et al. proposed a wind speed error correction model for numerical weather forecasting based on time series correlation testing and residual channel attention network, and established a power prediction model [8]. Yang et al. proposed a naive Bayesian based probability interval prediction method for wind power combination [9]. Li et al. combined and improved ant colony clustering algorithm and support vector machine method to achieve power interval prediction [10]. Bi et al. combined with neural networks prediction methods for short term power prediction [11-16].

When combining algorithms for prediction [17-23], it is easy to lack sufficient research on the advantages and disadvantages of each algorithm, which increases the computational workload. In a large number of studies, neural network technology is used to predict the load of the power system, while there is little work on the load prediction of the rail transit power system. Traditional neural networks tend to converge slowly and fall into local optima during load forecasting. Based on the above analysis, in order to predict the load of the rail transit power system, improve the problems encountered in traditional neural network load forecasting, and the overfitting problem of neural networks. We predict the load of the rail transit power system based on Dropout echo state network. Echo state network (ESN) is a recursive neural network that differs from traditional neural networks in the form of backpropagation for learning. Reservoir has good performance. Multiple parameters can easily lead to overfitting in neural network learning. Dropout is a commonly used method to solve overfitting problems.

II. Background Introduction

A. Dropout method

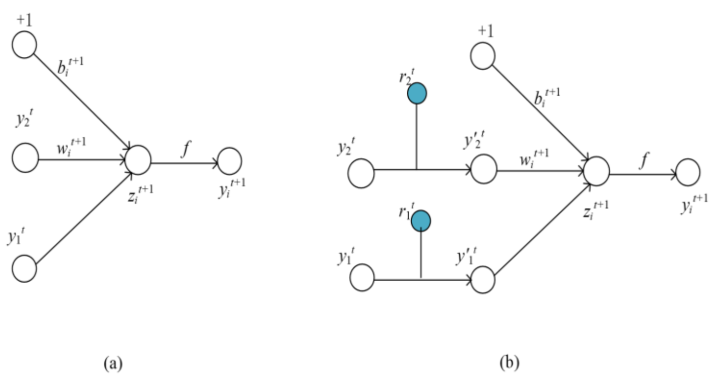

Fig.1. (a)neural network (b)dropout network

The current Dropout method is widely used in fully connected networks, causing the activation values of certain neurons to stop working with a certain probability. Overall, experiments need to be conducted based on specific networks. The following explains how the Dropout method causes the activation values of certain neurons to stop working with a certain probability. Firstly, in the training model stage, a probability process needs to be added to each neuron of the training network, as shown in Fig.1. Secondly, when predicting the model during the testing phase, the weight importance of each neuron is multiplied by the probability.

The Bernoulli function is used to generate probability vectors.

The network calculation formula without the Dropout method is as follows:

$$y_i^{t+1} = f(z_i^{t+1}) \tag{1}$$

$$z_i^{t+1} = w_i^{t+1} y^t + b_i^{t+1} \tag{2}$$

The network calculation formula without the Dropout method is as follows:

$$r_j^t \sim Bernoulli(p) \tag{3}$$

$$y'^t = r^t * y^t \tag{4}$$

$$z_i^{t+1} = w_i^{t+1} y'^t + b_i^{t+1} \tag{5}$$

$$y_i^{t+1} = f(z_i^{t+1}) \tag{6}$$

During the training phase, the network causes a neuron to stop working with probability p, which means that its activation function output value becomes 0 with probability p. For example, if the number of neurons in a certain layer of network is 1000 and the output value of its activation function is $y_1, y_2, \ldots\ldots, y_{1000}$, the probability p after using the Dropout method is 0.4. After using the Dropout method, approximately 400 neurons in this layer will have their values set to 0. After shielding certain neurons to make their activation value 0, it is necessary to scale the vector $y_1, y_2, \ldots\ldots, y_{1000}$ line by $1/(1-p)$. If $y_1, y_2, \ldots\ldots, y_{1000}$ is not scaled after being set to 0 during training, then the weight needs to be scaled during testing. During the testing phase of prediction, the weight of each neuron is multiplied by the probability*(1-p)*.

As shown in Fig.2, circles represent neurons, arrows represent synapses, and W represents the synaptic weight matrix.

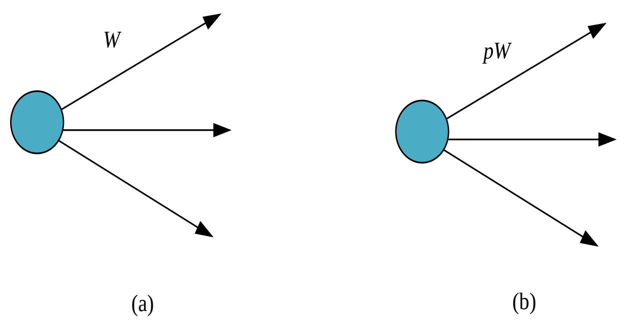

Fig.2. Dropout method

B. ROC curve

ROC curve is a commonly used evaluation indicator for binary classification in machine learning. When the distribution of positive and negative samples in the test set changes, the ROC curve can remain unchanged. Because there may be class imbalance in the dataset. That is, there are far more negative samples than positive samples (or vice versa). In the test data, the distribution of positive and negative samples may also constantly change over time.

In binary classification problems, instances are divided into positive or negative examples. However, in actual classification, the following four situations usually occur. The first scenario is that an instance is positive and predicted to be positive, it is a true positive (TP) instance. The second scenario is that an instance is a positive example but is predicted to become a negative example. This is the False Negative (FN). The third scenario is that an instance is a counterexample but is predicted to be a positive example. This is the false positive (FP) example. The fourth scenario is that an instance is a counterexample, predicted to be a counterexample. This is the true negative (TN).

In Fig.3, the y-axis represents the True Positive Rate (TPR), which represents the proportion of samples that are predicted to be positive and actually positive to all positive samples. How much did you guess correctly, which is actually 1. The abscissa is the False Positive Rate (FPR), which represents the proportion of samples predicted to be positive but actually negative to all negative samples. That is, the actual number is 0. How many guesses did you make. In Fig.3, (0, 1) indicates that this is a perfect classifier that correctly classifies all samples. (1, 0) indicates that this is the worst classifier because it successfully avoids all correct answers. (0, 0) indicates that the classifier predicts that all samples are negative. (1, 1) indicates that the classifier actually predicts that all samples are positive.

In Fig.3, the curve above the diagonal is the ROC curve. The points on the diagonal indicate that the classifier guesses half of the samples as positive and the other half as negative. The closer the ROC curve is to the upper left corner, the better the performance of the classifier. When the ROC curve of one model is enveloped by another model, the latter performs better than the former.

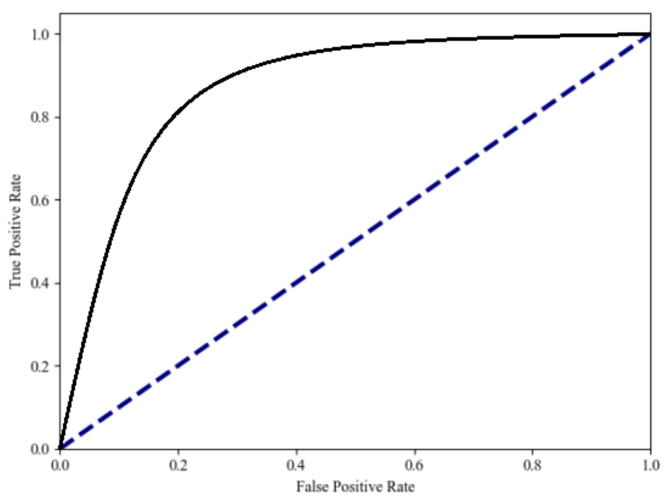

Fig.3. ROC curve

C. echo state network

Introducing the theory of neural networks into the study of nonlinear time series, neural networks can use gradient descent algorithms when their scale is slightly small. Moreover, if the training time of neural networks is too long, the prediction results of the network are prone to falling into local optima. To avoid such problems, dynamic neural networks can achieve better performance in predicting nonlinear time series. In 2004, Jaeger et al. proposed an artificial recurrent neural network algorithm called Echo State Network (ESN) [24]. The echo state network uses a simple single training algorithm to shorten the training time of the network and solve the long-standing problem of long and slightly complex training time in neural network research. The dynamic reservoir of the echo state network can maintain dynamic activity, even if there is no input, it can be continuously activated. Therefore, compared to traditional neural networks, the stability of the network can be guaranteed during the training process, effectively avoiding the problem of local optima. For the prediction of nonlinear time series, the echo state network has better performance.

Echo state network is a new type of recurrent neural network composed of front-end input layer, reservoir, and output layer, as shown in the Fig.4. The reservoir is composed of randomly sparsely connected neurons on a large scale, which has the characteristics of containing a relatively large number of neurons, randomly generating connection relationships between neurons, and sparsity of connections between neurons. The reservoir accepts input from two directions. One from the input layer. Another output from the previous state of the reservoir. The weights of the input layer and the state feedback weights do not require training and are determined by the random initial state. Therefore, the weights are represented as a large sparse matrix, with non-zero elements in the matrix connecting the activated neurons in the reservoir. The reservoir and output layer have a linear connection relationship, and in the actual training process, the weights of the linear connection need to be trained.

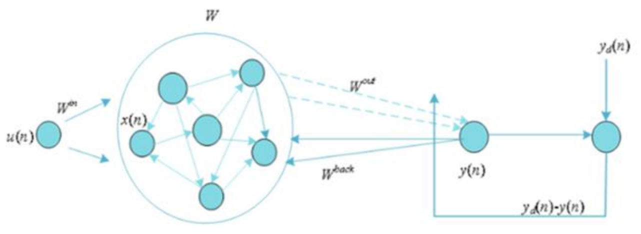

Fig.4. echo state network(ESN)

D. Dropout echo state network

The core structure of ESN is a randomly generated and invariant reservoir. The internal dynamic reservoir contains a large number of sparsely connected neurons, the operational status of the system, and has short-term memory function. The size of the reservoir is related to the number of neurons. Random connections through elements can affect the predictive performance of the network.

Dropout method has been added to the reservoir. Applying Dropout ESN in load forecasting of rail transit power system. The neurons in the reservoir with Dropout ESN are set to stop working with different probabilities. The prediction performance of Dropout ESN and ESN were compared and analyzed in load forecasting of rail transit power system.

Figure 5 shows Dropout ESN. Among them, W^{in}, W^{out}, and W^{back} are the input, output, and feedback connection weight matrices, respectively. The dashed circle represents the randomly screened neurons in the reservoir. Unlike original ESN, this network randomly blocks out some neurons in the reservoir, and the neurons stop working with a certain probability.

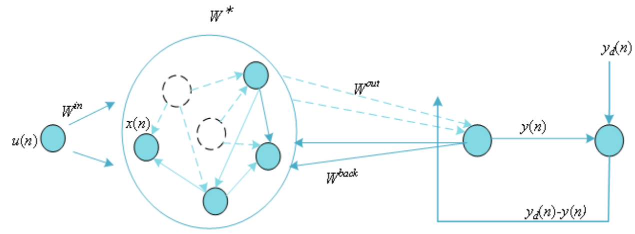

Fig.5. Dropout echo state network

At sampling time n, the update equation and output equation of Dropout ESN are as follows.

$$x(n+1) = \tanh(W^* x(n) + W^{back} y(n) + v(n)) \tag{7}$$

$$y(n+1) = W^{out} x(n+1) \tag{8}$$

Among them, W^* is the internal weight matrix of the reservoir during training. $W^* = pW$. $v(n)$ is the noise vector. $\tanh$ is the activation function.

E. Introduction to rail transit simulation system

The configuration structure and parameters of the rail transit simulation model in the main simulation validation scenario. The rail transit simulation verification scenario is configured based on the actual operation scenario of Qingdao Metro Line 11.

1) Train traction characteristic curve

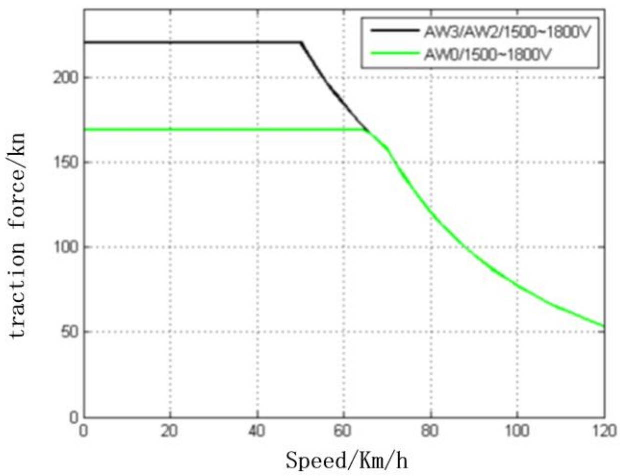

Fig.6 wheel rim traction-speed curve(overall train)

In the Fig.6 and Fig.7, AW0 represents empty load, AW2 represents full capacity, and AW3 represents overcrowding.

2) Power supply system AC side circuit configuration

The power supply system of rail transit is divided into two parts: AC power supply system and DC traction power supply system. The following mainly elaborates on the configuration file information of the AC power supply system. The specific configuration of the line parameters used is shown in the table 1.

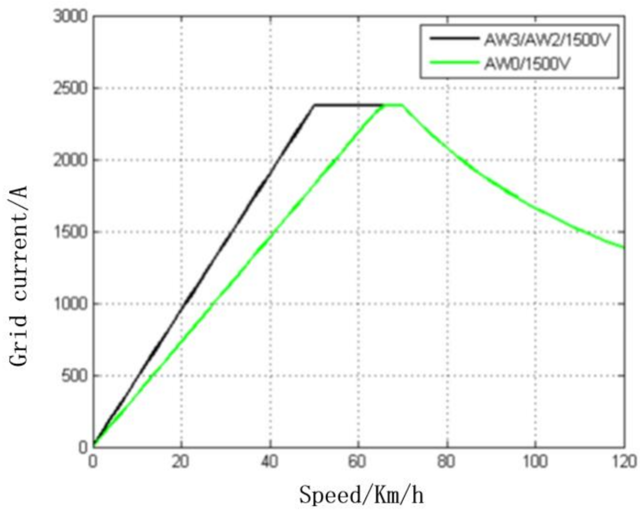

Fig.7 (Overall train) grid current- speed curve

TABLE1 LINE PARAMETERS

Line specifications/cross-sectional area	Resistance	reactance	Number of circuits
150	0.210	0.373	1
240	0.132	0.358	1
400	0.080	0.373	2

III. TEMPLATE NUMERICAL SIMULATION EXPERIMENTS

A. *experimental setup*

The simulation tool required for the experiment is JetBrains PyCharm 2019.3. The experiment used rail transit simulation power load data for simulation analysis, with a total of 15000 data points. The data diagram of the rail transit power system is shown in Fig.8.

The echo state network sets the number of neurons in the reservoir to 1000, the spectral radius to 0.9, trains 8000 steps, and tests 1000 steps. Set the noise to 0.1.

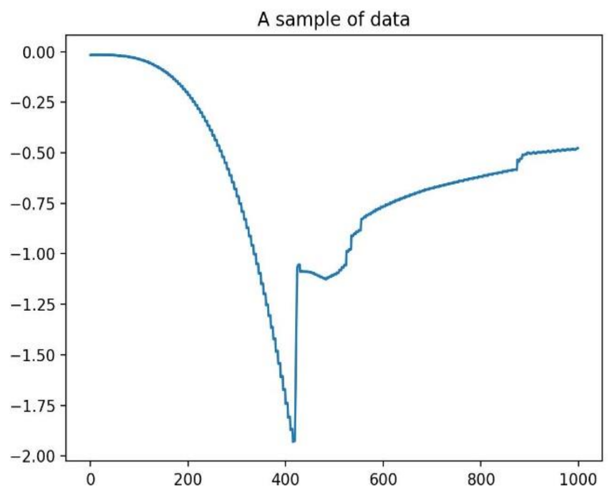

Fig.8. Power Data Diagram of Rail Transit Power System

B. *experimental results*

The Dropout ESN was applied to predict the load of the rail transit power system. By adjusting the connection probability of neurons in the reservoir, the predictive ability of Dropout ESN was analyzed from two aspects: predictive performance and ROC curve analysis.

1) *predictive performance analysis*

Figure 9 shows the comparison between original ESN and Dropout ESN in predicting rail transit load data. Among them, the red line is the Dropout ESN prediction curve for rail transit load data. The blue line is the prediction curve of original ESN for rail transit load data. The green line is the load data curve of rail transit. When the prediction range is 700 steps, the Dropout ESN predicted curve is closer to the real rail transit load data curve. The experiment shows that the prediction accuracy of Dropout ESN is higher in the load forecasting of rail transit power system.

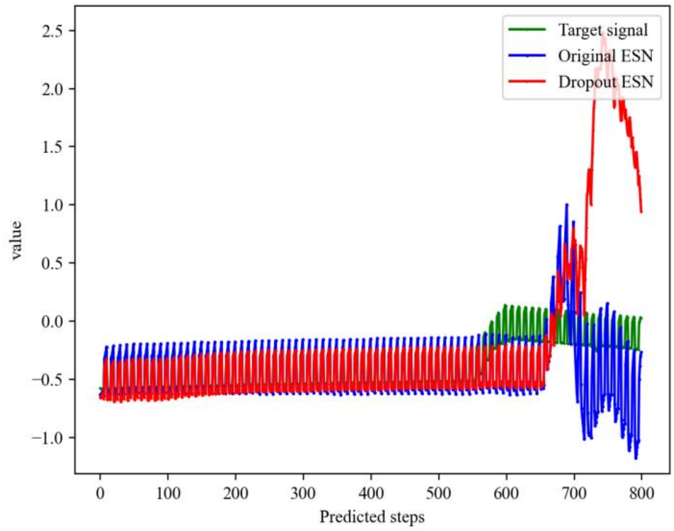

Fig.9. predicted steps

2) *ROC curve analysis*

In neural network prediction, whether the predicted value is within the allowable error range is considered as a binary situation. Below, the prediction performance of Dropout ESN will be analyzed using the commonly used evaluation indicator ROC curve for binary classification. The closer the ROC curve is to the upper left corner, the better the performance of the network, the higher prediction accuracy.

The ROC curves of original ESN and Dropout ESN are shown in Fig.10. In Fig.10, the orange line represents the predictive performance of Dropout ESN. The blue line represents the predictive performance of original ESN. These two curves are very close to the upper left corner. The ROC curve of Dropout ESN is closer to the upper left corner than that of original ESN. This indicates that compared to original ESN, Dropout ESN can classify numerical values more accurately, generate data that is closer to the sample data during prediction, and have higher prediction accuracy.

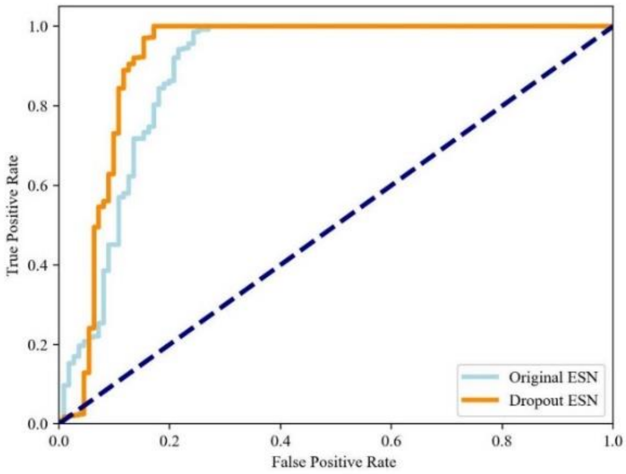

Fig.10 ROC curves

This paper applies Dropout ESN to the load forecasting task of the rail transit power system. Adjust the parameters in the reservoir of Dropout ESN. A comparative analysis was conducted on the load prediction ability of the rail transit power system using Dropout ESN and ESN. The experimental results indicate that when the neurons in the Dropout ESN reservoir stop working with a certain probability, the Dropout ESN has a higher accuracy in predicting the load of the rail transit power system.

The next step of work will combine the predicted results information to conduct a situation analysis of the rail transit power system, achieve perception of the operating status and trends of the power system, and make corresponding regulatory decisions based on this, ensuring the stable operation of the rail transit power system, and providing a more comprehensive optimization perspective for real-time operation scheduling of the rail transit power system.

ACKNOWLEDGMENT

This paper is sponsored by National Key Research and Development Program of China under Grant 2022YFB4300501.

REFERENCES

[1] J. Yin, J. Zhang, D. Li, T. Wang and K. Jing, "Big data cleaning model of smart grid based on Tensor Tucker decomposition," 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering(ICBASE), Bangkok, Thailand, 2020, pp. 166-169, doi: 10.1109/ICBASE51474.2020.00043.

[2] H. Guo, Y. Wang, S. Liu, L. Tang, L. Li and S. Mei, "Research on Capacity Allocation of Grid Side Energy Storage Participating in Power System Peak Shaving," 2022 7th International Conference on Power and Renewable Energy (ICPRE), Shanghai, China, 2022, pp. 1164-1169, doi: 10.1109/ICPRE55555.2022.9960433.

[3] Y. Gu, F. Wang, M. Li, L. Zhang and W. Gong, "A Digital Load Forecasting Method Based on Digital Twin and Improved GRU," 2022 Asian Conference on Frontiers of Power and Energy (ACFPE), Chengdu, China, 2022, pp. 462-466, doi: 10.1109/ACFPE56003.2022.9952254.

[4] L. Jixiang, Z. Qipei, Y. Zhihong, et al. Short term load forecasting method based on CNN-LSTM hybrid neural network model[J] . Automation of Electric Power Systems, 2019, 43(8)：131-137.

[5] CH. Zixing, J. Tao, ZH. Xidong, et al. Short-term power load forecasting based on a new date mapping method and an ISGU hybrid model[J]. Power System Protection and Control, 2022, 50 (15)：72-80.

[6] X. Peng, S. Yi, ZH. Jian, et al. Load situation awareness and control method based on artificial intelligence agent[J]. Automation of Electric Power Systems, 2019, 43(3)：178-186.

[7] Y. Hao, D. Zhen, M. Anbo. Short-term load forecasting based on extreme learning machine optimized by particle swarm optimization integrated with chaotic crisscross optimization[J]. Application Research of Computers, 2018, 35(7)：2088-2091.

[8] M. Changxin, W. Xia, L. Hao, et al. Day-ahead prediction of wind power based on NWP wind speed error correction[J]. Power System Technology, 2022, 46(9)：3455-3464.

[9] Y. Xiyun, ZH. Yanfeng, Y. Tianze, et al. Prediction of combination probability interval of wind power based on naive Bayes[J]. High Voltage Engineering, 2020, 46(3)：1099-1108.

[10] L. Jinghua, H. Yujin, H. Qian. Interval prediction method of wind power based on improved chaotic time series[J]. Electric Power Automation Equipment, 2019, 39(5)：53-58.

[11] B. Guihong, ZH. Xin, CH. Chenpeng, et al. Ultra short-term prediction of photovoltaic power generation based on multi-channel input and PCNN-BiLSTM[J]. Power System Technology, 2022, 46(9)：3463-3476.

[12] Ch. Han-Chieh, F. Lin, J. Pan, et al. Power Load Forecasting Based on VMD and Attention-LSTM[C]//International Association of Applied Science and Engineering.Proceedings of 2020 International Conference on Data Science and Information Technology(DSIT 2020).ACM,2020:7.DOI:10.26914/c.cnkihy.2020.060839.

[13] L. Yanzhen, G. Yinglei ,Sh. Pengfei , et al. Power Quality Prediction Based on BP Neural Network[C]//Institute of Management Science and Industrial Engineering.Proceedings of 2018 2nd International Conference on Computer Science and Intelligent Communication(CSIC 2018).Clausius Scientific Press,2018:7.

[14] X. Cai, J. Zhu, J. Liu, Z. Meng, Y. Huo and Y. Yu, "Short- Term Power Prediction Method for Photovolaic Power Generation Based on Elman Neural Network for Aspen Swarm Optimization," 2023 6th International Conference on Energy, Electrical and Power Engineering (CEEPE), Guangzhou, China, 2023, pp. 1063-1067, doi: 10.1109/CEEPE58418.2023.10166666.

[15] H. Huang et al., "Traffic Load Prediction of Power Communication Network Based on Graph Neural Network," 2023 International Conference on Information Technology Research and Innovation (ICITRI), Jakarta, Indonesia, 2023, pp. 109-113, doi: 10.1109/ICITRI59340.2023.10249883.

[16] S. Chao, W. Hui, M. Xiangping, J. Xiu and A. Yinping, "Load forecasting of high-rise buildings based on BP neural network," 2020 Chinese Automation Congress (CAC), Shanghai, China, 2020, pp. 3140-3144, doi: 10.1109/CAC51589.2020.9326465.

[17] R. Zheng, G. Li, K. Wang, B. Han, Z. Chen and M. Li, "Short-term Photovoltaic Power Prediction Based on Daily Feature Matrix and Deep Neural Network," 2021 6th Asia Conference on Power and Electrical Engineering (ACPEE), Chongqing, China, 2021, pp. 290-294, doi: 10.1109/ACPEE51499.2021.9436879.

[18] X. Xie and X. Lin, "Ultra-short-term Forecasting Method of Photovoltaic Power Generation Based on Combined Deep Neural Network," 2022 7th International Conference on Power and Renewable Energy (ICPRE), Shanghai, China, 2022, pp. 927-932, doi: 10.1109/ICPRE55555.2022.9960441.

[19] S. Yang, X. Zheng, M. Li and Y. Wei, "Offshore Wind Power Prediction Based on Variational Mode Decomposition and Long Short Term Memory Networks," 2021 6th International Conference on Power and Renewable Energy (ICPRE), Shanghai, China, 2021, pp. 1034-1039, doi: 10.1109/ICPRE52634.2021.9635584.

[20] Y. Yin, Z. Yu, X. Gou and J. Wang, "Photovoltaic power prediction model based on empirical mode decomposition-long-short memory neural network," 2021 International Conference on Intelligent Computing, Automation and Systems (ICICAS), Chongqing, China, 2021, pp. 349-353, doi: 10.1109/ICICAS53977.2021.00079.

[21] C. Zhu, Y. Wei, J. Zhu and C. Jiang, "Short-term Wind power Prediction Model Based on Cuckoo Algorithm and BP neural Network," 2020 Asia Energy and Electrical Engineering Symposium (AEEES), Chengdu, China, 2020, pp. 795-800, doi: 10.1109/AEEES48850.2020.9121389.

[22] Z. He and H. Yang, "A Combined Quartile-Regression-Interpolation Data Preprocessing Strategy for Short-Term Wind Power Prediction," 2023 6th International Conference on Energy, Electrical and Power Engineering (CEEPE), Guangzhou, China, 2023, pp. 1187-1191, doi: 10.1109/CEEPE58418.2023.10165706.

[23] Z. Zhou, S. Sun and Y. Gao, "Short-term wind power prediction based on EMD-LSTM," 2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA), Shenyang, China, 2023, pp. 802-807, doi: 10.1109/ICPECA56706.2023.10076173.

[24] H. Jaeger, H. Haas, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication[J]. Science, 2004, 304(5667): 78-80.