IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2020.Doi Number

Short-Term Load Forecasting of Power System based on Neural Network Intelligent Algorithm

Xiaoqiang Zheng[1,a], Xinyu Ran[2,b], Mingxin Cai[3,c*]

[1]School of Economics and Management, Southwest Petroleum University, Chengdu 610500, Sichuan, China
[2]Electrical Engineering and Automation, Southeast University, Nanjing 210018, Jiangsu, China
[3]College of Computer Science and Engineering, Northeastern University, Shenyang 110169, Liaoning, China

[a]E-mail: zhengxiaoqiang@163.com
[b]E-mail: ran_xinyu1@qq.com

*Corresponding author: Mingxin Cai ([c]E-mail: 1801700@stu.neu.edu.cn)

This work was supported by the Social Science Foundation of Southwest Petroleum University.

ABSTRACT Short-term load forecasting of power systems is an important part of the daily dispatch of the power sector. The accuracy of short-term load forecasting directly affects the safety, reliability and economy of power system operation. Therefore, the research on short-term load forecasting methods has always been the focus of scholars at home and abroad. In recent years, artificial neural networks have been widely studied as an intelligent algorithm and applied to the field of short-term power load forecasting. The network structure and learning law of artificial neural network are introduced. The mathematical model of short-term load forecasting based on BP neural network and Elman neural network is established respectively. Based on the basic structure of the BP neural network, a subsequent layer is added to store the internal state, so that the system can adapt to the time-varying characteristics. In the neural network modeling process, the model is optimized from the aspects of network algorithm, excitation function and network structure to improve the convergence speed and prediction accuracy of the network. In the simulation process, the above two mathematical models use the same data samples for short-term power load forecasting. In the modeling process, in order to ensure the stability of the system, a small learning rate is set, and the selection range is between 0.01-0.8. During the modeling process, the error between different learning rates is compared. Finally, the learning rate of 0.43 is selected. By comparing the prediction error with the prediction accuracy, the prediction effect of the Elman neural network model is better than that of the BP neural network model.

INDEX TERMS Machine Learning, Neural Network, Short-Term Load Test, Power System

I. INTRODUCTION

The power industry is the lifeblood of the national economy. The economy needs to develop and electricity is the first. Electricity plays a vital role in China's economic construction, national security, social stability, and quality of life. With the continuous development of the economy, customers' requirements for power supply quality have gradually increased, and power grid management has become more and more modern. Power load forecasting research has become particularly important and has become an important research area in the power industry [1]. The power load forecasting data is the basic information that the dispatch planning department should have. With the continuous development of technology and the continuous improvement of the economic level, China's power industry has ushered in an unprecedented development momentum. Accurate load forecasting can guide the power grid dispatching department to formulate an economic and reasonable dispatching plan, which is an effective measure to save energy, improve economic and social benefits, and provide important guarantee and development for normal social life and sustainable and healthy economic development [2]. As a special energy source, electric energy is difficult to store in large quantities, and the time difference between production, transmission and consumption is small. Therefore, from the perspective of economy and safety, it is necessary to keep the power generation of the power plant in line with the power consumption and maintain a dynamic balance. Otherwise, when the amount of power generation is greater than the amount of consumption, energy is wasted. When the power generation is less than the consumption, it can not meet the needs of users, and even negatively affect the safety and stability of the entire power grid. In order to ensure the safe, stable and reasonable operation of the power system, it is necessary to deeply understand the changing characteristics of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3021064, IEEE Access

different load types and analyze the trend of future power load, so that the power load becomes a reality. Forecasting has become an important research content of the power industry [3]. The significance of this research is to improve the accuracy of power load forecasting, which has a very important impact on the safety and stability of the power system and the interests of the power sector. Accurate prediction of power load allows the dispatcher to reasonably schedule the start and stop of the generator set to ensure safe and stable operation of the power system. At the same time, reasonable inspection and maintenance plans can be arranged to ensure that high quality power is available to customers at all times. Accurately judging the future trend of power load, we can also arrange a reasonable dispatch plan for the power grid to reduce power generation costs, and positively affect social benefits while improving economic efficiency. Power load forecasting is a core component of energy management and distribution management systems, and is the basic information that many departments of the power system need to master [4]. It is applied to the planning and design of power systems. The efficiency, economy, safety and stability of power grid operation are also important research topics in the power sector. Especially in today's unprecedented development of the power industry, the development model of power companies has gradually changed from a planned model to a market model. Improving the accuracy of power load forecasting has become one of the main tasks facing the power industry and one of the major issues facing the power industry. Survival of the fittest in standard electric power industry [5].

Power load forecasting is not only an indispensable component of power dispatching and development planning, but also a basic information that many power system departments should master. Improving the accuracy of power load forecasting not only guides power generation, transmission and distribution. Improving the prediction accuracy of load forecasting is conducive to formulating a reasonable scheduling plan, maintaining the dynamic capacity and power consumption of the power plant, achieving the goal of balance between supply and demand, and formulating a reasonable power grid construction plan to reduce the grid load [6]. Waste energy improvement to solve the traditional method of short-term power load forecasting. The problems of domestic technology in power load forecasting include time series method, gray theory method, regression analysis method and other classical methods, including artificial neural network and fuzzy logic method, support vector machine (SVM) and wavelet analysis and prediction methods, intelligent prediction methods of chaotic methods [7]. Compared with foreign scholars' research on neural networks, China's research on neural networks started late [8]. The first informal meeting was held very early in Beijing: the 1990 National Neural Network, the first academic conference jointly sponsored by the China Artificial Intelligence Society and the Chinese Institute of Electronics, the China Institute of Physics and other eight Chinese learning neural networks. Chinese scholars have begun to pay attention to the development of neural network technology, and took a historic first step; in 1991, the second neural network conference was held in Nanjing, through the establishment of the neural network society [9]. Under the efforts of many researchers, the research of neural networks has gradually attracted the attention and development of domestic academic circles. This provides a good environment and conditions for the development of neural network technology, and urges China to continue research on neural networks. Progress has narrowed the gap with foreign advanced levels [10].

In this paper, the machine learning-neural network intelligent algorithm is used to establish the mathematical model, and the optimization is improved from the aspects of data preprocessing, network structure selection and learning algorithm. The basic theory of BP neural network is introduced. The short-term load forecasting model is established based on BP neural network. The modeling process is described in detail, including data processing, network structure and parameter selection. Finally, the model is simulated and predicted. And analyze the output. Finally, the model is simulated by Matlab software to analyze the problems in the model.

II. PROPOSED METHOD

A. FACTORS OF SHORT-TERM POWER LOAD FORECASTING

In short-term power load forecasting, it is important to analyze and obtain the characteristic factors that affect short-term power load. After analysis and reference to a large number of literatures, the following characteristics were selected as key factors for load measurement: 1) Time factor (particle size accurate to 15 minutes). The magnitude of the power load will vary significantly with peak and valley values of power consumption time. Therefore, it belongs to time series data, which is why LSTM algorithm can predict with high precision. It has excellent performance in processing time series data. 0 represents a working day, 1 represents a holiday, and 2 represents a holiday factor. For users and businesses, holidays are very different in terms of electricity consumption. When dealing with them, we will discretize them, 3 represents the temperature coefficient, temperature changes will also significantly affect the amount of power generation, choose the highest and lowest temperature of the day as a key feature, 4 represents precipitation, humidity to a large extent will lead to the power plant adjusts its output power, and 5 represents extreme weather. When extreme weather such as thunderstorms occur, it affects the power source and is therefore an important indicator in the model. 1.2 data preprocessing in the power load forecasting, a large amount of historical load data will be acquired, in which, due to machine or human reasons, there are some deviations and missing data from the data groups in the data, and the existence of these “bad data” is usually will greatly affect the accuracy of the forecast, so we first test and deal with. For abnormal data, the following methods are used for processing: 1) Statistics of non-abnormal data are performed using probability and statistics methods, and abnormal data is identified and processed through statistical confidence intervals. 2) The curve replacement method is used to delete and replace some daily loads that are significantly different from the normal load curve. 3) From an empirical point of view, the empirical correction method finds the data value of a certain period by comparing the normal load curve. 4) The average fill method replaces the average of adjacent data with the missing data. In addition, because the dimensions of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3021064, IEEE Access

each feature are different, if the data of different dimensions is directly imported into the model for training, it is easy to cause the model to not converge. Therefore, the data should be normalized:

$$X_{norm} = \frac{X - X_{\min}}{X_{\max}} \tag{1}$$

Among them, for normalized data, X is the original data, and X_{max} and X_{min} are the maximum and minimum values of the original data, respectively.

B. CHARACTERISTICS OF POWER LOAD FORECASTING

Based on the history of load and current fluctuations in the power system, the future trends of load fluctuations and possible operating conditions of the system are predicted [11]. Scientific and reasonable prediction algorithms and models are important technical foundations for the real-time, reliability and accuracy of the entire load forecasting process. The power system is affected by many factors, and all variables have nonlinear time-varying characteristics. Therefore, the entire load forecasting process must be characterized by uncertainty, conditionality and regional effects [12].

(1) The uncertainty of short-term load forecasting is defined by Zadeh very early, pointing out its ambiguity, uncertainty and other problems. The power load is a time variable of development and is affected by the technical transformation and expansion of the grid system. In order to improve the social and economic benefits of dispatching and dispatching of power companies at all levels, it is necessary to estimate the trend of system load in advance. However, in the forecasting process, it is difficult to completely consider some hidden dangers systematically, and some temporary situations will inevitably make the prediction results different from the actual results, that is, the load prediction results are definitely not correct, or the structure is considered to be incomplete. Therefore, there must be an error between the results of short-term load forecasting and the actual situation [13-14]. The main purpose of model improvement and optimization research is to effectively control the error within the allowable range through relevant algorithms, thereby improving the comprehensive social and economic benefits of scheduling operations.

(2) Conditional short-term load forecasting is to compare the load fluctuation trend in the future period with the recursive process, that is, in the forecasting process, external constraints have a greater impact on the forecast results. If there is no basis, these constraints cannot be assumed, but need to be measured by corresponding technical means. In short-term load forecasting, we first need to determine a target, define the constraint with the exact number of concepts, and then predict the daily weather conditions, working day types, holidays and other external conditions to set conditional constraints load forecasting, in order to predict consistent data of daily load fluctuations, High-precision load forecasting is conducive to the development of efficient and reliable operation of the relevant power sector.

(3) Regional effects of short-term load forecasting: The level of social and economic development in different regions is different, and the proportion of load components is also different. Therefore, the conditional factors that affect load and fluctuation will also vary. For large power grids, due to its main transmission function, its load fluctuations have a strong regularity. However, for a regional distribution network, due to its complex network structure and uncertain power usage time, there is no regular change in load capacity. In other words, there is a regional effect in grid load forecasting. For large grid load forecasting, high precision forecast data is usually available. However, the load forecasting accuracy of small and complex regional distribution networks is relatively low. Therefore, this paper chooses large-scale power grid as the research object.

C. METHOD FOR SHORT-TERM LOAD FORECASTING OF POWER SYSTEMS

Using computer and other optimization algorithms to construct power load forecasting model, improve the processing rate and accuracy of load forecasting model, and provide real and reliable load estimation data information for power system dispatching operation, which has important engineering practical significance. The power system traditional load forecasting method generally includes: unit energy consumption method, proportional elastic coefficient method, one-way regression parameter estimation method, gray (GM) prediction method, and the like. The above methods improve the accuracy of short-term load forecast to a certain extent, but most of them need to construct accurate mathematical or physical models as the basis of theoretical analysis. In order to obtain high-precision load fluctuation data information that can truly reflect system load fluctuations, many power research scholars have summarized the above basic algorithms, combined with various optimization techniques such as error compensation and acceleration convergence to form a project. Improved intelligent algorithms for short-term load forecasting, such as: expert system method, neural network prediction method, fuzzy control method, and neural-fuzzy hybrid prediction method. This chapter will compare and analyze the methods commonly used in the field of short-term load forecasting engineering, and reasonably select the basic arithmetic method theory that is consistent with this paper.

1) TRADITIONAL FORECASTING METHODS:

(1) Unit energy consumption method: The unit energy consumption method is a statistical analysis method based on the energy required to produce a unit product. It is the total production capacity of the products and products obtained by multiplying the average electric energy required by the unit. The specific expression is as follows:

$$S = PQ \tag{2}$$

Where: S consumes the electricity required to produce the product; P produces the average unit volume of the product; Q is the total production volume of the additional product. In the process of power load forecasting, because of certain production processes, technical measures, and the level of operation of workers, the above three variables are all variable, so the relative amount can be used to convert, in the formula: The use of dispatching and producing such products is found in practical work applications. The unit income method is simple and effective in the short-term load forecasting process, but the single-consumption method requires a lot of detailed statistical investigation work, which is in a certain production. Due to the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3021064, IEEE Access

influence of production process, production machinery efficiency, etc., it is impossible to accurately count the average efficiency of a unit product, and the entire statistical workload is quite large. It is very difficult to accurately calculate the load demand of a certain event in the modern complex power grid, it can be said that it is almost impossible. Proportional Elasticity Coefficient Method: The proportional elastic coefficient method is actually looking for some internal binding proportional relationship between the rate of change of electricity consumption and the rate of change of load affecting another non-variable of load fluctuation. The specific expression is:

$$\varepsilon_{yx} = \frac{d_x}{y_x} \tag{3}$$

Where: the proportional elastic coefficient of the variable y to the variable x; the average rate of change of the variable y for the variable x in a certain period of time. The proportional elastic coefficient ε_{yx} is used to measure the trend of the variable y. When the power system is running, the load data may be missing due to equipment communication failures and other reasons. In order to accurately predict the load, it is necessary to complete the missing load data, and make the data of the supplement basically follow the law of load change. In the case of a small amount of data loss, the data can be completed based on the actual experience of the power system dispatcher, based on the missing data completion or the previous one or more similar load data, plus the trend increment of the load change. When there are multiple consecutive missing load data for a certain purpose, the data interpolation can be performed by linear interpolation.

$$L_{k+j} = L_k + \frac{L_{k+i} - L_k}{i} \bullet j, 0 < j < i \tag{4}$$

As shown in the following formula, the input gate is i, the forgotten gate is f, the output gate is 0, the input gate determines the current time system input, the forgetting gate determines the forgotten information, and the output gate determines the final $h_t = o_t \tanh(c_t)$ of the final data output:

$$i_t = \sigma\left(W_{xi}x_t + W_{hi}h_{t-i} + W_{ci}C_{t-1} + b_i\right) \tag{5}$$

$$ft = \sigma\left(W_{xf}x_t + W_{hi}h_{t-i} + W_{ci}C_{t-1} + b_i\right) \tag{6}$$

$$C_t = f_t C_{t-1} + i_t \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right) \tag{7}$$

$$0_t = \sigma\left(W_{xo}x_t + W_{ho}h_{t-i} + W_{ct-1} + b_0\right) \tag{8}$$

$$h_t = o_t \tanh(c_t) \tag{9}$$

(2) Evaluation index of the model: When evaluating the regression prediction model, the general root mean square error (RMSE) is used to describe it. The calculation formula is as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i - x_i}{x_i}\right| \times 100\% \tag{10}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - x_i}{x_i}\right)^2} \times 100\% \tag{11}$$

Where x_i is the predicted value corresponding to the actual value of the load at the i-th time.

D. LSTM CYCLIC NEURAL NETWORK

(1) Tensor Flow-based LSTM cyclic neural network algorithm: Tensor FIow deep learning tool. Tensor Flow is the second generation artificial intelligence learning system developed by Google. It is an open source software library that relies on data flow graph for numerical calculation [15-17]. It can be used for research in all aspects including machine learning and deep learning. Due to its high flexibility, excellent portability and multi-language support, it has become the most popular deep learning tool. The LSTM algorithm of this paper is also based on the Tensor Flow tool for experimental development [18]. The LSTM Cyclic Neural Network Recurrent Neural Network (RNN) is a neural network that processes and predicts sequence data, and can fully analyze the time series information and semantic information in the mining data. The RNN takes the output data of the hidden layer at the last few moments as the input of its own layer, so that the time dimension information is retained. Module A inputs Xi and outputs a value Hi, which loops the information from the current step to the next step [19]. The structure of the RNN is shown in Figure 1.

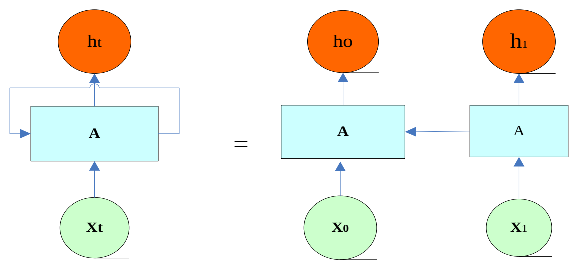

FIGURE 1. RNN structure diagram

(2)The Long and Short Time Memory (LSTM) type RNN model is an improvement of the traditional RNN. It mainly solves the problem of gradient explosion and gradient dispersion of the RNN model. The LSTM accepts the previous moment output, the current system state and the current system input, updates the system status through the input gate, the forgetting gate and the output gate and outputs the final result as shown in Figure 2.

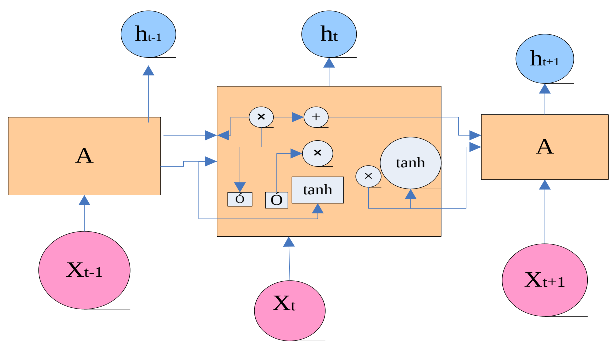

FIGURE 2. LSTM structure diagram

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3021064, IEEE Access

III. Experiments

A. Experimental Settings

1) EXPERIMENTAL DATA

In this paper, when using fuzzy C-means algorithm for cluster analysis, although it has the advantage of stable algorithm results, the obvious disadvantage is that cluster variable C needs to be determined in advance. In fact, for most clustering problems, the number of clusters is difficult to determine in advance. The ISODATA algorithm is improved on the basis of the FCM algorithm, and the dynamic search strategy of clustering numbers is approximated. The algorithm automatically merges and splits the cluster centers to ensure the minimum cluster loss function value. However, no restrictions or requirements are required before adopting this algorithm. Instead, many parameters need to be predetermined, such as the expected number of clusters, the small number of samples in the cluster, the merge parameters, and so on. In the case of many clustering of these parameters, such as clustering of electrical load prediction, the loss function should not only increase the total distance deviation, but should also include the absolute loss term of the average relative deviation and the absolute loss term of the maximum relative deviation. Therefore, this paper redefines the loss function and proposes a dynamic fuzzy C clustering algorithm. In the clustering method, the loss function is defined as:

$$J_b = \sum_{jml}^{c}\sum_{iml}^{n}\left[\mu\ (x_i)\right]^b \left\|x_i - mj\right\|^2 + M\sum_{jml}^{c}\sum_{jml}^{n}\left[\mu(x_i)\right]^b PX_{ij}^2 + M_2\sum_{jml}^{c}\sum_{iml}^{n}\left[\mu_j(xl)\right]^b \left\|x_i - m_j\right\|^2 \tag{13}$$

2) EXPERIMENTAL PROCESS

We evaluated the actual effect of the algorithm and experimented with TensoFlow 1.3.0 under the Ubuntu 16.04 system. The actual computer settings are as follows: tercore i7-3770@3.40GHz processor, DDR3800MHz, 8GB memory. The experimental data is the power generation load of a power plant in a certain period of time in a certain city. For comparison purposes, we also use the same data and use the current high-precision traditional learning random forest algorithm and logistic regression algorithm to predict the load, so that we can compare and test the advantages of the algorithm.The percentage of prediction error for different algorithms for the same amount of data is shown in Table 1.

TABLE I
PERCENTAGE OF PREDICTION ERROR FOR DIFFERENT ALGORITHMS FOR THE SAME AMOUNT OF DATA

Algorithm	Random forest	SVM	LSTM
Parameter	3.32	4.24	1.81
MAPE%	4×8	3×7	5×9

Compared with the traditional random forest algorithm and logistic regression algorithm, the proposed LSTM algorithm has higher prediction accuracy. In the process of training massive data, the deep learning algorithm can automatically learn the hidden information in the data and make it more robust to errors and outliers. As the amount of data increases, the prediction accuracy of the LSTM algorithm will be further improved. Through multiple iterations, a high-precision prediction model was finally trained.

Similarly, due to communication transmission anomalies, accidents, major political events, etc., historical load data may also be abnormal. In order to detect and process bad data, the following methods can be used: 1) The data lateral comparison method compares the load at a certain moment with the load at the time before and after. If the difference is greater than a certain threshold, it is considered to be a load glitch; the correction formula is as follows:

If $\left|y(i,t)\right|y(i) > \delta(t)$, then:

$$y(i,t) = \begin{cases} y() - \delta \\ y() - \delta \end{cases} \begin{matrix} y \\ y \end{matrix} (i,t) > \bar{y}(i) \tag{14}$$

The data longitudinal comparison method compares the load value at a certain moment with the load value at the same time of the previous day and 2 days, and if the deviation is greater than a certain threshold, the average value is replaced; the correction formula is as follows:

If $\left|y(i,t) - \bar{y}(i)\right| > \varepsilon(t)$,then:

$$y(i,t) = \begin{cases} \bar{y}(i) + \varepsilon, y(i,t) > \bar{y} \\ \bar{y}(i) - \varepsilon, y(i,t) < \bar{y} \end{cases} (i) \tag{15}$$

In the formula, y(i, t) is the load at the time t at the i-th day; it is the average value of the load at the time f in the last few days; the threshold can be artificially set according to the actual situation. The simulation of the algorithm is mainly for selecting one-fifth of the training data for generating the initial sub-network group h, and the steps of setting and updating the sub-network structure are the same as before. For the case where the number of data contained in the designed data block is taken as 10, 5 and 1, respectively, the verification method of the two sets of standard data is also 10 times of five-fold cross-validation in five different orders. Through experiments, it is found that when the data block is relatively large, the prediction error will be relatively large, but the acquisition will be small, and the prediction error will also increase. Therefore, the selection of the data block capacity needs to be specifically analyzed and determined according to the specific problem to select the optimal value. It can be seen that the online regression algorithm based on boosting has better prediction results than the improved online bagging algorithm, especially for time series prediction. At the same time, the comparison between the offline algorithm of the single BP network and the online boosting integration algorithm for the 24-hour load normalization value on the same day is given. The average relative errors are 3.2% and 2.7%, respectively. The comparison curves between the predicted result and the true value (load normalized value) are respectively shown, where y1 and y2 represent the online integrated algorithm and the prediction result of a single network, respectively, and y represents the true value.

It can be seen that the prediction of the online integration algorithm can better reflect the trend and be more accurate. The experimental results also show that Boosting algorithm has better learning prediction effect for individual learning with both bias and variance.

V. DISCUSSION

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3021064, IEEE Access

A. SELECTING LEARNING PARAMETERS

1) TAKE THE CONNECTION WEIGHT

The selection of the connection weight of the neural network will have a direct impact on the duration of the model, the convergence speed and the training results. If the initial weight and threshold are chosen too large, the convergence process will be slowed down, and even the stagnation will occur. The initial connection weight is usually set according to experience, but after the training starts, the initial connection weight begins to change, and the model affects the prediction accuracy by continuously adjusting the connection weight. The value with the highest prediction accuracy is selected as the initial weight as shown in Table 2.

TABLE II
RELATIONSHIP BETWEEN INITIAL WEIGHTS AND NETWORK PREDICTION ACCURACY

Initial weight	Network prediction average error（%）	Network prediction accuracy (%)
-0.5	5.34	96.64
-0.3	6.12	95.46
0.2	4.92	95.98
0.3	2.11	96.22
0.3	1.98	97.66
0.4	3.29	95.48
0.6	4.97	94.23
0.7	5.36	95.37

For the same network model, the initial weights are different, and the output values will have different results. Therefore, the reasonable selection of the initial connection weight has an important impact on the training duration and convergence of the whole model. Generally, the initial weight is selected within the range. In the selection process, according to the prediction results obtained by different values, the initial weight with the highest prediction accuracy is 0.2 as shown in Figure 3.

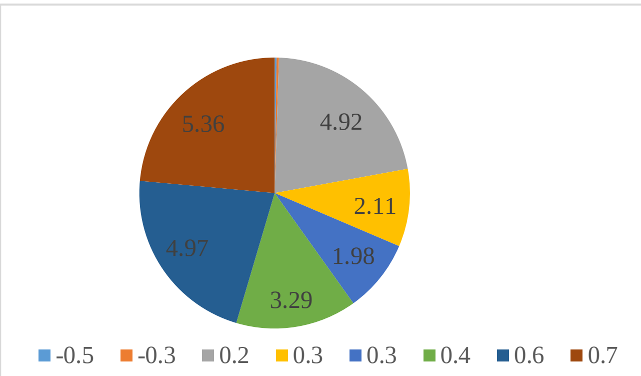

FIGURE 3. Comparison of initial and output values for network prediction accuracy

During the network training process, the learning rate will have a certain impact on the change of the connection weight, which plays a very important role in the overall performance of the whole model. If the value of the learning rate is too large, it will adversely affect the stability of the model. If the value is too small, the model training and learning time will be too long, and the convergence will be worse. Therefore, it depends on the specific model. The structure selects a reasonable learning rate. The smaller learning rate is selected, and the training time of the network is shorter, but the effect of not sinking into the local minimum during the convergence process can be achieved, so that the network error is continuously reduced. Increasing the learning rate can effectively reduce the time spent on training, but when it exceeds a certain range, it will have a greater impact on the prediction results. Therefore, different models should choose the learning rate that suits them. In the modeling process, in order to ensure the stability of the system, a small learning rate is generally set, and the selection range is between 0.01 and 0.8. During the modeling process, the error between different learning rates is compared. Finally, the learning rate of 0.43 is selected. The relationship between learning rate and network prediction accuracy is shown in Figure 4.

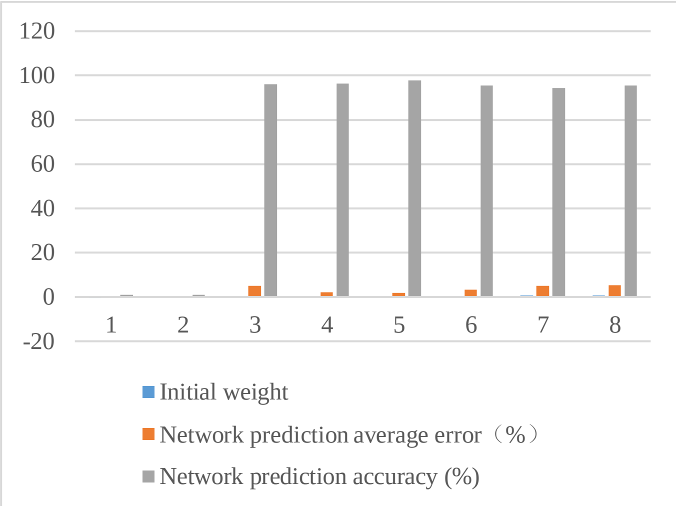

FIGURE 4. Relationship between learning rate and network prediction accuracy

2) SELECT THE MOMENTUM FACTOR

The higher the learning rate, the faster the convergence speed, but too large will cause the model to be unstable. The method of additional momentum factor can usually be adopted, which can effectively improve the transmission performance.

$$\Delta\omega_{ji}(n) = a\Delta\omega_{ji}(n-1) + \eta\delta_j(n)v \tag{16}$$

After the samples are sequentially added, the above formula can be rewritten as a time series form, and the time is from 0 to n. The above equation can be regarded as the $\Delta\omega_{ji}(n)$ first-order difference equation, so there are:

$$\Delta\omega_{ji}(n) = \eta\sum_{i=0}^{n} a^{n-1}\frac{\partial E(t)}{\partial\omega_{ji}(t)} \tag{17}$$

For offline Boosting, the online Boosting algorithm with 5 data blocks and the running time of the improved online algorithm are compared. The results show that the online regression algorithm based on Boosting also has advantages in running time, and it is not significant. The calculation amount is increased, which is suitable for practical applications. The

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3021064, IEEE Access

comparison between online adaptive Boosting and offline Boositng and improved Online-Bag is shown in Table 3 and Figure 5, respectively.

TABLE III
COMPARISON OF ONLINE ADAPTIVE BOOSTING WITH OFFLINE BOOSITNG AND IMPROVED ONLINE-BAG

Data Set	Bosting	Online Boosting -1	Online Boosting -5	Online Boosting -10	Improved Online-Bag
Bosting Housing	8.9	13.6	10.7	11.4	11.5
Abalone	5.3	6.2	6.0	6.1	5.9
Electrical Load(%)	2.6	4.2	2.7	2.9	7.3

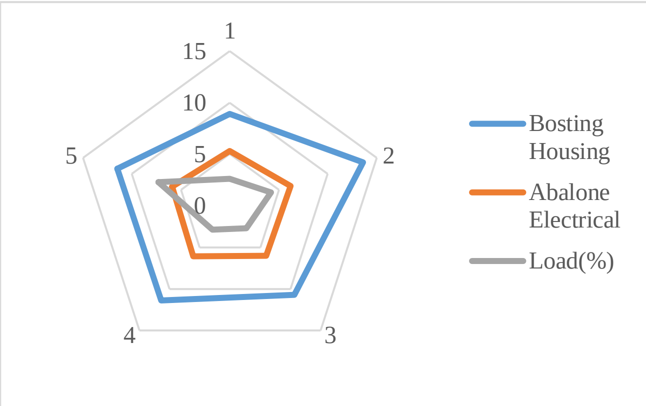

FIGURE 2. Comparison of Online adaptive Boosting and Batch algorithm

VI. CONCLUSIONS

(1) When implementing various machine learning applications, quality control results not only play an important role in building confidence in the algorithm, but also play an important role in solving two skeptical aspects: 1) the machine will replace the artificial concerns; 2) concerns about the "black box" type of algorithm. Although these two issues are effective in abstraction, the concepts of machine learning and machine learning help to solve these problems in real life. This makes machine learning algorithms the enabler, helping practitioners to accomplish tasks more quickly, and can also help improve human interpretation and/or highlight human error. Various examples are shown here. In fact, these things are almost always observed. Eventually, machine-to-human interaction is achieved to optimize the mixing results.

(2) Short-term power load forecasting data is important information that the power sector needs to master. The improvement of forecasting accuracy has an important impact on the safety, stability and economy of power grid operation. In this paper, based on some problems existing in the traditional prediction method, a mathematical model based on BP neural network and Elman neural network is established, and the model is optimized from the aspects of learning algorithm and network structure. Finally, the simulation experiment is carried out by Matlab software, and the short-term power load forecasting effect based on the improved BP neural network and Elman neural network model is compared. It is concluded that the Elman neural network has higher ability to deal with dynamic problems than BP neural network.

(3) Periodic time series. If the sample subsequences with the same change pattern are separately trained, the accuracy can be improved and the calculation can be simplified. According to this idea, fuzzy clustering is integrated into multi-learning machine learning, and combined with integrated algorithm, a multi-neural network model based on improved dynamic fuzzy C-means clustering is proposed and applied to ultra short term load forecasting with periodic characteristics. The experiment proves that the fuzzy clustering samples are separately trained and learned, and the new samples are subordinated and predicted. The prediction accuracy is much better than that of the single network model, and it has higher precision than the online integration algorithm. (Super) short-term prediction of energy (electricity) load is very important for energy system scheduling and optimizing. At present, artificial neural networks have been successfully applied to power load forecasting, but they also have certain limitations, such as over-learning and complex systems.

References

[1] Rosario Cammarota, Indranil Banerjee, &Ofer Rosenberg. (2018) “Machine Learning”,Conference on Artificial Intelligence Research & Development: International Conference of the Catalan Association for Artificial Intelligence Vl.1, pp.5-10.

[2] Helma, C., Cramer, T., Kramer, S., & De, R. L. “Data Mining and Machine Learning Techniques for The Ldentification of Mutagenicity Inducing Substructures and Structure Activity Relationships of Noncongeneric Compounds,”*J Chem Inf Comput Sci*, vol. 35, no. 39, pp. 1402-1411, 2018.

[3] Giusti, A., Guzzi, J., Dan, C. C., He, F. L., Rodríguez, J. P., & Fontana, F. “A Machine Learning Approach to Visual Perception of Forest Trails for Mobile Robots,” *IEEE Robotics & Automation Letters*, vol. 1, no. 2, pp. 661-667, 2017.

[4] Erdogan, B., Hekimoglu, S., Durdag, U. M., & Ocalan, T. “Empirical Estimation of the Power of Test in Outlier Detection Problem”, *Studia Geophysica et Geodaetica*, vol. 63, no. 1, pp. 1-16, 2019.

[5] Chen, X., Huang, L., Fu, P., Gao, G., Song, Z., & Xu, L. “System realization of caships dc high power test facility”, *CPSS Transactions on Power Electronics and Applications*, vol. 4, no. 2, pp. 101-108, 2019.

[6] Chubov, S. V., & Soldatov, A. I. “Electronic Load as Part of the Test Complex of the Power Processing Unit of Electric and Plasma Propulsion,” *Materials Science & Engineering Conference Series*, vol. 177, pp.012021, 2017.

[7] Lu, H., Setiono, R., & Liu, H. “Effective Data Mining Using Neural Networks,” *Knowledge & Data Engineering IEEE Transactions on*, 957-961, 2016.

[8] Chen, S. Z., Xu, L. I., Peng, W., & Zhang, D. H. “Motor Power Model of Cold Rolling Based on Numerical Integration and Power Loss Test,” *Journal of Northeastern University*, vol. 38, no. 3, pp. 361-365, 2017.

[9] Tighilt, Y., Bouttout, F., & Khellaf, A. “Modeling and Design of Printed Antennas Using Ueural Networks,” *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 21, no. 2, pp. 228-233, 2015.

[10] Fernandes, John F.T., Lamb, Kevin L., & Twist, Craig. “A Comparison of Load-Velocity and Load-Power Relationships between Well-Trained Young and Middle-Aged Males During Three Popular Resistance Exercises,” *The Journal of Strength & Conditioning Research*,vol. 32, pp. 1, 2018.

[11] Wysokinski, T. W., Renier, M., Suortti, P., Belev, G., & Chapman, L. M. “High-Power-Load Dclm Monochromator for a Computed Tomography Program at Bmit at Energies of 25–150 Kev,” *Journal of Synchrotron Radiation*, vol. 25, no. 5, pp. 128, 2018.

[12] Y. Bai, G. Zhao, J. Dou, G. Huang, & Z. Liu. "Short-Term Power Load Combination Forecasting Based on Parallel Membrane Computing," *Dianli Xitong Baohu Yu Kongzhi/power System Protection & Control*, vol. 45, no. 7, pp. 35-42, 2017.

[13] S. Lin, E. Tian, Y. Fu, X. Tang, & Q. Wang. "Power Load Classification Method Based on Information Entropy Piecewise Aggregate Approximation and Spectral Clustering," *Zhongguo Dianji Gongcheng Xuebao/Proceedings of the Chinese Society of Electrical Engineering*, vol. 37, no. 8, pp. 2242-2252, 2017.

[14] Liu Yuan, Jianzhong Zhou, Zijun Mai, & Yuanzheng Li. "Random Fuzzy Optimization Model for Short-Term Hydropower Scheduling Considering Uncertainty of Power Load," *Water Resources Management*,vol. 31, no. 9, pp. 2713-2728, 2017.

[15] Chao Song, Zhixian Chen, Xiaozhi Qi, Baoliang Zhao, & Ying Hu. "Human Trajectory Prediction for Automatic Guided Vehicle with Recurrent Neural Network," *The Journal of Engineering*, vol. 2018, no. 6, pp. 1574-1578, 2018.

[16] Yifan Wang, Rui Xu, & Miaolei Zhou. "Prandtl-ishlinskii Modeling for Giant Magnetostrictive Actuator Based on Internal Time-Delay Recurrent Neural Network,"*IEEE Transactions on Magnetics*, on. 99, pp. 1-4, 2018.

[17] YeongHyeon Park, & Il Dong Yun. "Arrhythmia Detection in Electrocardiogram Based on Recurrent neural Network Encoder–Decoder with Lyapunov Exponent," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 14, no. 2, pp. 54, 2019.

[18] Jitendra Kumar, Rimsha Goomer, & Ashutosh Kumar Singh. "Long Short Term Memory Recurrent Neural Network (lstm-rnn) Based Workload Forecasting Model for Cloud Datacenters," *Procedia Computer Science*, vol. 12, no. 5, pp. 676-682, 2018.

[19] Wanquan Liu, & Lixin Xu. "A New Recurrent Neural Network Adaptive Approach for Host-Gate Way Rate Control Protocol within Intranets Using ATM abr Service," *Journal of Industrial and Management Optimization (JIMO)*, vol. 1, no. 13, pp. 389-404, 2017.

XIAOQIANG ZHENG was born in Changning, Sichuan. P.R. China, in 1981. He received PhD from Lanzhou University, P.R. China. Now, he is an associate professor at the School of Economics and Management, Southwest Petroleum University. His research interest include energy economics and sustainable development.
E-mail: zhengxiaoqiang@163.com

XINYU RAN was born in Guiyang City, Guizhou Province. P.R. China, in 1999. Now, he studies in College of Electrical Engineering and Automation, Southeast University. His research interest include power electronics and machine learning.
E-mail: ran_xinyu1@qq.com

MINGXIN CAI received the bachelor's degree from Northeastern University, in 2018 where she is currently pursuing the master's degree. Her research interests include time series data processing and machine learning.
E-mail: 1801700@stu.neu.edu.cn