Research on load forecasting method of large Power Grid based on Deep confidence Network

Banghan Xu¹, Tianci Wu¹, Xue Wang², Ying Zhu¹, Nengxing Guo², Zhengying Cai¹,*

¹School of Computer and Information, China Three Gorges University, Yichang, Hubei, 443002
²College of Electrical Engineering and New Energy, China Three Gorges University, Yichang, Hubei, 443002

*Corresponding author: zhengyingcai@ctgu.edu.cn

Abstract. In view of the existing electric power load forecasting models exist in the measuring precision is not high and the disadvantage of low stability, this paper puts forward a deep belief network - based extreme learning machine (Deep Belief Network – Extreme Learning Machine, DBN - ELM) of power load forecasting model. On the basis of traditional DBN, this model introduces ELM as the regression layer, and combines DBN's advantages in feature extraction with ELM's strong generalization ability to improve the accuracy of power load prediction. The experiment shows that compared with the traditional DBN neural network, the model can better fit the variation trend of power load data time series, and has higher prediction accuracy.

1. Introduction
Power load prediction plays an important role in the power system, which is the basis to guarantee the stable operation of the power system and the premise to realize the maximum economic benefit of the power market [1]. On the one hand, accurate power load forecasting can provide a basis for the production planning of producers, so as to ensure the balance of power supply and demand and realize the optimal allocation of resources, which can not only improve the stability of the power system, but also reduce the production cost for enterprises [2]. Since electricity cannot be stored in large quantities, inaccurate load forecasting may lead to a series of problems, such as power failure due to insufficient supply or waste of resources due to oversupply [3]. On the other hand, with the deepening of power market reform, all players in the power market make scientific decisions based on electricity price prediction [4], and electricity price prediction is based on load prediction, and the accuracy of load prediction is an important index for electricity price estimation in the power market. Therefore, power load forecasting is of great significance to the power system and the power market, and is also a hot topic in the field of power load research.

The power load data is collected by smart electricity meters and then transmitted to the power center by wireless transmission. In the whole collection process, due to the influence of environment, equipment, communication conditions and other factors, noise, deviation and missing values are easily generated in the power data time series. At the same time, the power data itself has the characteristics...
of nonlinear, volatility and randomness, which makes the power load prediction a very challenging problem.

At present, machine learning has become the mainstream method in data prediction, especially artificial neural network. For example, Bin et al. proposed a power load prediction theory based on BP neural network. Zhang et al. used Elman neural network to conduct power load prediction research. Guan et al. proposed a wavelet neural network method, whose core idea is to apply peak value filtering technology to detect and correct the peak value in power load data. However, because the shallow neural network is difficult to fully extract the characteristics of the original load data, it greatly limits the accuracy of load prediction. With the development of deep learning, deep neural network is also applied in power load forecasting. The deep neural network can transform the input data for many times before the output, and features can be mapped layer by layer. The upper layer can learn more abstract features from the lower layer, and the most representative features obtained can be used for nonlinear approximation. Therefore, compared with the shallow model, the deep neural network can learn more complex patterns and has better nonlinear approximation performance. DeepBeliefNetwork (DBN) is a typical representative of deep neural network model and has been applied in the field of power load prediction.

As is known to all, DBN is stacked up by multiple restricted Boltzmann machines, and preprocesses data through unsupervised learning. The traditional DBN's local weight fine-tuning is accomplished through back-propagation (BP) based on gradient descent, so it is easy to fall into local optimization in the training process. At the same time, a large number of iterative computations tend to cause too slow convergence speed. In view of this, the nonlinear approximation performance of DBN in the prediction field is not satisfactory.

In order to solve the above problems, this paper proposes a power load forecasting model based on DBN-ELM. DBN models on the feature extraction of superior performance and extreme learning machine (Extreme Learning Machine, ELM) fast learning speed, shi combine good performance characteristics. The experimental results show that this method has better prediction accuracy in power load data prediction than the traditional DBN method.

2. The network structure

2.1. Deep confidence network - Ultimate learning machine model

Generally speaking, DBN-ELM is composed of DBN neural network which is responsible for feature extraction and ELM neural network which is the regression layer. It combines DBN's feature extraction ability with ELM's advantages of fast learning speed and good generalization ability, thus improving the prediction performance. Its structure diagram is shown in Figure 1.

Assuming that a DBN contains N hidden layers, the first N-1 layer is initialized with greedy training method, and the bias and weight of the N-1 hidden layer to the NTH hidden layer and the NTH hidden layer to the output layer are determined by the ELM algorithm.

Let m be the number of neurons in the NTH hidden layer, and L be the number of neurons in the NTH hidden layer, then this network can be expressed as:

$$\sum_{i=1}^{m} \beta_i g (W_i \bullet \text{O}_{n-1} + b_i) = y_j, j = 1, \ldots, l$$

(1)

The optimal predicted result is the smallest output error, which can be expressed as:

$$\sum_{j=1}^{m} \| y_j - t_j \| = 0$$

(2)
And, get special $\beta_i$, make the following formula true, namely

$$\sum_{i=1}^{m} \beta_i O_{n,j} = t_j, j = 1, \ldots, l$$

(3)

The above equation can be converted to:

$$O_n \beta = H$$

(4)

The details can be expressed in the following form:

$$O_n(W_1, \ldots, W_m, b_1, \ldots, b_m, O_{n-1,1}, \ldots, O_{n-1,l}) = \begin{bmatrix}
g(W_1 \cdot O_{n-1,1} + b_1) & \cdots & g(W_m \cdot O_{n-1,1} + b_m) \\
\vdots & \ddots & \vdots \\
g(W_1 \cdot O_{n-1,l} + b_1) & \cdots & g(W_m \cdot O_{n-1,l} + b_m)
\end{bmatrix}$$

(5)

$$\beta = [\beta_1^T, H_1^T, \ldots, \beta_m^T, H_m^T]$$

(6)

$$\parallel O_m(\hat{W_i}, \hat{b_i})\beta - H \parallel = \min_{W,b,\beta} \parallel O_n(W_i, b_i)\beta - H \parallel$$

(7)

Here, the last hidden layer and the output layer of DBN can be regarded as a single hidden layer neural network, and ELM algorithm is still used to train it. Similarly, the parameters $W_i$ and $B_i$ of the network model in the last layer are selected by random determination. After the parameters of these two models are determined, the output matrix $H$ is uniquely determined. At this point, the output weight of the entire DBN model can be solved by the following formula:

$$\hat{\beta} = O_m^T H$$

(8)
2.2. Deep Belief Networks

DBN limited by multiple boltzmann machine (Restricted Boltzmann Machines RBMs) stack and become, by the greed of unsupervised method step by step to learn. After the initial weight is given, the RBM of the first layer is trained and its output is taken as the input of the next layer, and so on. The output of the lower layer is always taken as the input of the upper layer. In this case, we assume that the neurons of visible layer V and hidden layer H of RBM are binary and random, that is, only set at 0 or 1. When given \((v, h)\), the energy function can be defined as follows:

\[
E(v, h \mid \theta) = -\sum_{i=1}^{m} \sum_{j=1}^{n} v_i w_{ij} h_j - \sum_{i=1}^{m} a_i v_i - \sum_{j=1}^{n} b_j h_j
\]  

Through the energy function, the joint probability distribution function of the visible element and the hidden element of the model can be obtained, namely:

\[
p(v, h \mid \theta) = \frac{1}{Z_\theta} e^{-E(v, h \mid \theta)}
\]

\[
Z_\theta = \sum_v \sum_h e^{-E(v, h \mid \theta)}
\]

In RBM, neurons of the same layer are not connected, while neurons of adjacent layers are all connected. Therefore, when the state of visible layer node is given, the activation state of each hidden layer node is independent of each other. Therefore, conditional probability can be deduced from the principle of Bayesian formula and expressed by the following logical function:

\[
p(v_i = 1 \mid h) = \sigma(a_i + \sum_j w_{ij} h_j)
\]

\[
p(h_j = 1 \mid v) = \sigma(b_j + w_{ij} v_i)
\]

\[
\sigma(x) = \frac{1}{1 + e^{-x}}
\]

In order to calculate the update equation of each parameter, we adopted the contrast divergence (CD) algorithm proposed by Hinton. First, the visible layer is initialized according to the first input data. Then the conditional probability of hidden neuron is calculated according to the value of visible layer and conditional probability formula. Finally, a sample is extracted from the calculated probability using Gibbs sampling, and the visible layer is reconstructed with this sample. By repeating the process, the update rules for the relevant parameters are expressed as follows:

\[
\Delta W_{ij} = \varepsilon (\langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{recon}}) 
\]

\[
\Delta a_i = \varepsilon (\langle v_i \rangle_{\text{data}} - \langle v_i \rangle_{\text{recon}})
\]

\[
\Delta b_j = \varepsilon (\langle h_j \rangle_{\text{data}} - \langle h_j \rangle_{\text{recon}})
\]

With this guideline, we can get an appropriate weight and use the same method until all RBM weights have been updated.
2.3. Ultimate learning machine

ELM is a kind of based on feedforward neural networks (Feed forward Neural Network, FNN) building a machine learning method, the network structures include input layer, hidden layer and output layer, can be the input weights and bias random initialization, and then get the output of the corresponding weight. Here, assuming M samples, a neural network consisting of L hidden layer nodes can be represented by the following formula:

$$\sum_{i=1}^{L} \beta_i g(W_i \cdot X_j + b_i) = y_j, j = 1, \ldots, M$$  \hspace{1cm} (18)

G (x) is the activation function of the hidden layer, Wi and are input and output weights respectively, and BI is the bias of the ith hidden layer. The ultimate purpose of the neural network with single hidden layer is to minimize the output error, namely:

$$\sum_{i=1}^{L} \beta_i g(W_i \cdot X_j + b_i) = t_j, j = 1, \ldots, M$$  \hspace{1cm} (19)

Here, O represents the output of the hidden layer node, and H represents the expected output, thus:

$$O \beta = H$$  \hspace{1cm} (20)

3. Experiment and analysis

3.1. Data and Its Preprocessing

This article chooses the PJM posted on its website in 2017 power load data, the data is collected every hour, a day to get 24 set of load data, selecting 1600 groups, 800 of which point in time as the training data set, another 800 time points as the validation data set, at the same time, in order to eliminate the initial transient, set up a forgotten, abandoned before 100 time points. Since the obtained data are from actual measurements and there are deviations or missing values, this experiment adopts means interpolation and min-max normalization to carry out linear transformation of the original data, so that the resulting values map to between [0,1]. MATLAB2019 was used as a computational tool to carry out simulation research, and the output results were averaged with 100 predicted values.

3.2. The evaluation index

In order to evaluate the performance of the model, three evaluation indexes, namely normalized root mean square error (NRMSE), directional change prediction (POCID) and determination coefficient (R2), were used in this experiment, and their expressions were as follows:

$$VRMSE = \sqrt{\frac{\sum_{t=1}^{L} (\hat{Y}_t - Y_t)^2}{ns^2}}$$  \hspace{1cm} (21)

$$R^2 = 1 - \frac{\sum_{t=1}^{L} (\hat{Y}_t - Y_t)^2}{\sum_{t=1}^{L} (Y_t - \bar{Y})^2}$$  \hspace{1cm} (22)

$$POCID = \frac{1}{L} \sum_{t=1}^{L} D_t$$

$$F_t = (\hat{Y}_{t-1} - \hat{Y}_{t-1}) (Y_t - Y_{t-1})$$  \hspace{1cm} (23)

3.3. Performance analysis

In Figure 2-3, the time step comparison curve obtained by DBN and DBN-ELM neural network model in power load prediction and the error amplitude curve between the actual value and the predicted value are presented respectively. As can be seen from the figure, the DBN-ELM model used in this experiment
can effectively fit the change trend of the real power load data, and the predicted error is significantly smaller than that of the DBN model. Further, Table 1 shows the performance evaluation indexes of DBN and DBN-ELM neural network models in power load prediction, including NRMSE, POCID and R2. These indexes can reflect the accuracy of model prediction. For NRMSE, the smaller the value is, the smaller the residual is, and therefore the better the prediction performance of the model. For POCID and R2, the higher these index values are, the better the model fitting is. Specifically, POCID measures the model's ability to predict trends in time series, while R2 reflects the model's ability to capture variability in the target time series. As can be seen from the table, the NRMSE value of DBN-ELM model we adopted is 0.1370, which is significantly smaller than the 0.3272 of DBN, while the POCID and R2 values are 0.8475 and 0.9812, respectively, which are larger than the 0.8250 and 0.8928 of DBN. It shows that the three performance evaluation indexes of DBN are better than DBN model, and it can predict the change trend of power load data more effectively and accurately.

![DBN-ELM output and margin of error](image)

**Fig 2. DBN-ELM output and margin of error**

### 4. Conclusion

This paper proposes a novel prediction method of electric power load data, the method using DBN-ELM for power load forecasting, the structure, DBN stack in the matter can be fully extracted data characteristics, these characteristics provides a good initial point, ELM makes ELM itself to fully embody the advantages of strong generalization ability, through the simulation experiment shows that the model in power load forecasting has good capability of nonlinear approximation, contrast, DBN model has higher precision of prediction, has wide development prospect in the field of power load forecasting.

### References

[1] X Xu, M Gong, L Huang, A Liu. High-Precision Power Load Forecasting Using Real-time Temperature Information and Deep Learning Method[J]. Electric Power Components and Systems, 2019, 47(16-17): 1574-1583.

[2] S Bouktif, A Fiaz, A Ouni, M Serhani. Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches[J]. Energies, 2018, 11(7): 1636 1655.

[3] Z Wang, T Hong, M Piette. Building Thermal Load Prediction Through Shallow Machine Learning and Deep Learning[J]. Applied Energy, 2020, 263: 114683-114696.

[4] B Monica, S Franco. On the Complexity of Neural Network Classifiers: A Comparison Between Shallow and Deep Architectures [J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 25(8):1553-1565.