The forecast of household power load based on genetic algorithm optimizing BP neural network

Hengchao Liu
School of International Education, North China Electric Power University, Baoding, Hebei, 071003, China
E-mail: liuhc@ncepu.edu.cn

Abstract. In view of the fact that power load forecasting is increasingly becoming an important part of power system planning and power dispatching, this paper proposed a forecasting method based on genetic algorithm optimizing BP neural network and applied it to household power load forecasting. First of all, the original data were pre-processed by factor analysis to eliminate the correlation between variables, and then the household load from January 1st to March 22nd was predicted by using BP and GA-BP models respectively. The author obtained the predicted results in the next 5 days summing to 120 hours and utilized MAPE and RMSE to compare their prediction errors quantitatively. The predicted results of the calculation example are in good agreement with actual values. Therefore it is demonstrated that the BP neural network optimized by genetic algorithm has better forecasting effects, which can provide reference for residential and municipal short-term load forecasting.

1. Introduction
Load rate is an important quality index of power grid management, and load forecasting is also crucial work of power dispatching department. Accurate forecasting of power load is of great significance to ensure the stable operation and economic benefits of the power system.

In recent years, there are numerous literature and methods about power load forecasting, such as traditional time series and new intelligent algorithms represented by neural network and support vector machine. At present, the research on prediction algorithms has been very in-depth, and the achievements are also very considerable. Most of the literature focuses on the improvement of the algorithm. Such as literature[1] proposed a combined forecasting method of power load, in which SARIMA and GRNN are weighted respectively, put into SVM, and the output results are obtained. Literature[2] combined the Prophet model used for data forecasting with LSTM neural network algorithm to achieve more accurate load forecasting. Literature[3] extracted the best feature subsets of the data sets and gained the best prediction results by using Bagging algorithm with REPTree. In addition, many algorithms such as Random Forest (RF) and Markov Chains have been applied to power load forecasting and achieved great results[4-5].

The following figure shows the developing timeline of main models and algorithms for power load forecasting.
Due to its characteristics of self-learning and approximation to arbitrary continuous functions, the neural network is widely used in data prediction and text classification, and it is the most attractive method in power load forecasting in the past decade[6]. This paper proposes a BP neural network model improved by genetic algorithm. The weights and thresholds of BP neural network are optimized through genetic algorithm and the updated weights and thresholds are parameters to be replaced in the BP model. After several times of iterative training, the network can output prediction results with smaller errors and better effects.

2. Construction of Forecasting Model

BP neural network, also known as Error Back Propagation Model, is trained according to error gradient descent method. The structure of BP network usually consists of the input layer, hidden layer and output layer. In the process of training, the weights and thresholds of the network are constantly updated according to the prediction error until the mean square error between the network output and actual output is reduced to the minimum. The following is the structure diagram of BP neural network.

In the topology, $X_1, X_2, ..., X_6$ are the input variables, $Y$ is the output variable, $w_{ij}$ and $w_{jk}$ are the weights of the neural network. The network expresses the function mapping relation from 6 independent variables to 1 dependent variable.

The weight adjusting equations of BP network are as follows.

Figure 1. Developing timeline of main forecasting algorithms for power load.

Figure 2. The structure diagram of BP neural network.
\[
\begin{align*}
\Delta w_i &= -\eta \frac{\partial E}{\partial w_i} = \eta \sum_{k=1}^{m} (\delta_k w_{jk}) x_i f'(\text{net}_j) \quad i = 1, 2, \ldots, n; j = 1, 2, \ldots, l \\
\Delta w_{jk} &= -\eta \frac{\partial E}{\partial w_{jk}} = \eta (R_k - L_k) y_j f'(\text{net}_k) \quad j = 1, 2, \ldots, l; k = 1, 2, \ldots, m
\end{align*}
\]

Where \( \text{net}_j \) is a linear combination of input variables and threshold of hidden layer, \( f(\text{net}_i) \) is activation function of hidden layer. We use the Sigmoid function as the activation function because it is continuous and smooth and is the most widely used activation function in BP neural network[7].

\[
\text{net}_j = \sum_{i=1}^{n} w_i x_i - a_j \quad j = 1, 2, \ldots, l
\]

\[
f(\text{net}_j) = \frac{1}{1 + e^{-x}}
\]

\( E \) is the error function of network output, denoted as follows.

\[
E = \frac{1}{2} \sum_{k=1}^{m} (R_k - L_k)^2
\]

The equations for updating weights are as follows.

\[
\begin{align*}
\Delta w_i (t+1) &= \Delta w_i (t) + \alpha [\Delta w_i (t) - \Delta w_i (t-1)] + \eta \delta_i x_i \\
\Delta w_{jk} (t+1) &= \Delta w_{jk} (t) + \alpha [\Delta w_{jk} (t) - \Delta w_{jk} (t-1)] + \eta \delta_j y_j
\end{align*}
\]

The following is the flow chart of genetic algorithm optimizing BP neural network.

![Flow Chart of Algorithms](image)

Figure 3. The flow chart of the algorithms.
3. The Calculation Example and Analysis

3.1. Data Pre-processing
The household load changes dramatically for days in the year and has strong seasonal changing traits[8-9]. The investigation and study of household load’s features are significant for power load forecasting, power grid planning and power marketing[10].

This paper selects the real electricity load data of a family in France from 2006 to 2010 as the original data sets. The data sets include the data of 7 input variables, such as active power, reactive power, voltage and so on. Because the original data is in minutely granularity, this paper uses SPSS to calculate the total of variables like power and load intensity in unit hour, to average the voltage and to remove the missing values and abnormal data at the same time. Finally, the author obtains the family’s load data in hourly granularity and selects the data from 1st January 2006 to 22nd March 2006 as the prediction data sets with a total of 1271 data points.

3.2. Correlation Analysis
The correlation coefficient matrix and scatter plot of the original 7 variables are as follows:

![Correlation Coefficient Matrix]

![Figure 4. The correlation coefficient matrix.]

![Figure 5. The corresponding scatter plot.]
Where $H_{our}$ is time(h), $Re_P$ is reactive power(W), $AVG_U$ is average voltage(V), $Me_1$, $Me_2$ and $Me_3$ are the power indicators of three instruments respectively with unit(W), $Ac_P$ is active power(W).

From the correlation coefficient scatter plot, we can clearly see that only the average voltage has negative correlation with all the other variables, while all the other variables have positive correlation with one another.

3.3. Factor Analysis

Factor analysis is a model studying the correlation coefficient matrix and adopts specific methods to reduce the initial variables into fewer factors, and these factors could explain the main information as much as possible. So it also can be called dimensionality reduction[11].

Before factor analysis, KMO-Bartlett spherical test should be taken to determine whether the data are suitable for factor analysis. SPSS shows that the KMO test statistic is 0.741 and the p-value of Bartlett statistic is 0.000, which is less than the significance level. That means it is suitable for factor analysis to a certain extent.

Through factor analysis, we extract six principal components, calculate the coefficients of factor score functions in the Anderson-Rubin method, and then obtain the value of each factor as the data of input variables of BP model.

3.4. BP and GA-BP Analysis

The author used the six factors as input variables and the corresponding load intensity as output variables. By running and debugging BP and GA-BP codes respectively through Matlab, the forecast results in the next 5 days with a total of 120 hours are as follows.

It is obvious that both the prediction results of the two algorithms are good, and the BP optimized by genetic algorithm output curve is more consistent with the actual output curve, which has better effects than a single BP neural network.

The following picture shows the GA-BP fitness curve. The algorithm converges after 6 iterations, and the optimal fitness is 107.801. The convergent speed is rapid and searching efficiency is high, which reflects the advantages of genetic algorithm optimization.
Figure 7. GA-BP fitness curve.

We use Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) to analyse the prediction accuracy of the algorithms.

\[
E_{\text{mape}} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - \hat{y}_n}{\hat{y}_n} \right| \times 100\% \\
E_{\text{rmse}} = \left[ \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - \hat{y}_n)^2 \right]^{1/2} \times 100\%
\]

The comparison of the average MAPE and RMSE between the two prediction algorithms is shown in the following table:

| Algorithm type | BP       | GA-BP    |
|----------------|----------|----------|
| MAPE/%         | 2.9801   | 1.1668   |
| RMSE/%         | 4.7419   | 2.2922   |

According to the table, both BP and GA-BP are great in forecasting household load, while the MAPE and RMSE of the latter are smaller, thus the forecasting accuracy of GA-BP is higher.

4. Conclusion

With the development of national economy and improvement of people's life, the proportion of family load is increasing day by day, and it has gradually become an important part of the urban load. Load forecasting is a very important task in power system. Energy transmission, maintenance scheduling as well as expansion planning all depend on accurate and efficient load forecasting[12].

This paper applied BP and GA-BP algorithms to household load forecasting. Firstly, the author utilized factor analysis on the original data to eliminate the correlation interference and then proved that BP optimized by the genetic algorithm has better prediction effects than a single BP through a practical example of household load. Such a combined model has a fast convergent speed and high algorithm efficiency. Therefore, for the forecast of households' load, it is prioritized to consider neural network and relevant improved algorithms.
Acknowledgments
Above all, I would like to thank my alma mater for teaching and nurturing me for such a long time. When staying there, I met a lot of meticulous, diligent teachers and like-minded, ambitious classmates and friends. Without their influences on my study bit by bit, it would be hard for me to get what I want, and hard for me to become what I want to become.

Secondly, I have to thank my parents. Without their encouragement and support, it would be much more difficult for me to keep going on the road of scientific research all the time.

Of course, all the other people I met in my life should not be forgotten. your help, intentionally or unintentionally, was an indispensable part of my growing up.

References
[1] Xu Jing, Chi Fujian, GE Leijiao, et al. (2020) Short-term Combined Commercial Load Forecasting Method Based on SARIMA-GRNN-SVM[J]. Proceedings of the CSU-EPSA, 32(2): 85-91.
[2] Peng Pai, Liu Min. Short-term load forecasting method based on Prophet-LSTM combined model[J]. Proceedings of the CSU-EPSA, 1-9.
[3] Tüfekci, P. (2014) Prediction of Full Load Electrical Power Output of a Base Load Operated Combined Cycle Power Plant Using Machine Learning Methods[J]. International Journal of Electrical Power and Energy Systems, 60.
[4] Chen Lvpeng, Yin Linfei, Yu Tao, et al. (2018) Short-term Power Load Forecasting Based on Deep Forest Algorithm[J]. Electric Power Construction, 39(11): 42-50.
[5] Shao Chuanyong. (2020) Family Load Forecasting Method Based on Markov Chains[D]. South China University of Technology.
[6] Li Yuancheng, Fang Tingjian, Yu Erkeng. (2003) STUDY OF SUPPORT VECTOR MACHINES FOR SHORT-TERM LOAD FORECASTING[J]. Proceedings of the CSEE, (6): 55-59.
[7] Han, J., Moraga, C. (1995) The influence of the sigmoid function parameters on the speed of backpropagation learning. In: International Workshop on Artificial Neural Networks. Springer, Berlin, Heidelberg. pp. 195-201.
[8] Yang Jianping. (2002) Analysis and Study of Nanjing Residential Load Profile[J]. Power Demand Side Management, (2): 24-27.
[9] Dai Qi. (2005) Load Modeling of Power System Based on Load Composition[D]. HoHai University.
[10] Wu Zhiqiang, Wu Zhihua, Song Xiaohui, et al. (2006) Study on Urban Residential Load Characteristics[J]. Power System Technology, (S2): 659-662.
[11] Yong A.G., Pearce S. (2013) A Beginner’s Guide to Factor Analysis: Focusing on Exploratory Factor Analysis[J]. Tutorials in quantitative methods for psychology, 9(2): 79-94.
[12] Gupta, S., Singh, V., Mittal, A. P., Rani, A. (2016) Weekly load prediction using wavelet neural network approach. In: 2016 Second International Conference on Computational Intelligence & Communication Technology (CICT). IEEE. pp. 174-179.