An Improved Load Forecasting Method Based on Optimal Weighted Combination

Chun Zhang
Electrical Engineering College
Anhui Polytechnic University
Wuhu, China
e-mail: chzhang@ahpu.edu.cn

Feng Li
Electrical Engineering College
Anhui Polytechnic University
Wuhu, China
e-mail: 1558593713@qq.com

Hao Mei
Electrical Engineering College
Anhui Polytechnic University
Wuhu, China
e-mail: 539841834@qq.com

Abstract—Load forecasting is the basis for the planning and safe operation of electric power. To satisfy all forecasting conditions as much as possible and improve the results of load forecasting, the combination forecasting method is proposed. The new method integrates the advantages of residual gray theoretical model and back propagation (BP for short) neural network model with additional momentum-adaptive learning rate. And the weighted coefficient optimization process is introduced in the study. Basing on the electricity consumption of 2001-2011of a power supply company, the optimal weighted combination method are applied to forecast and analyze. The simulation results show that the mean square error of optimal weighted combination forecasting method is smaller than that of either a single forecasting method, which can improve the prediction precision significantly.

Keywords-load forecasting; grey model; BP neural network; additional momentum-adaptive learning rate; optimal weighted combination

I. INTRODUCTION

With the development of power market, the demand of forecasting precision is higher than before. At home and abroad, power system workers have accumulated a lot of experiences which has formed many economical and practical conventional forecasting methods. But typically, influenced by various uncertainties, load forecasting data will produce large deviation.

At present, traditional forecasting methods have not been widely accepted in engineering application because of many shortcomings such as massive history data, low forecasting precision. On the contrary, back propagation neural network and grey theory become new trend of load forecasting which can overcome all above shortcomings. But both of them have deficiencies. Some studies have pointed out that exponential solution of differential equation of grey model (GM for short) was suitable for load index with exponential growth trend[1,2]. For other growth trends, it was sometimes more difficult to improve forecasting precision because of large fitting grey. Reference [3] concluded many deficiencies of BP neural network which includes low convergence speed in learning process, easy vibrating in training process and resulting in the network non-convergence. To solve these problems, in recent years, many scholars have thorough research and put forward lots of improved algorithms. As to grey theory, same dimension new information replacement forecasting and grey hierarchical forecasting methods are detailed presented [4,5]. Specific to BP neural network, improved additional momentum [6], elastic BP algorithm [7], adaptive learning rate method [8] and parallel quasi-Newton method [9] are represented in succession. But because of uncertainty of load change, there is no one way to guarantee obtaining satisfactory prediction results in any case. Furthermore, when rashly discards those methods which owned large prediction error, it will lose some useful information.

So, the weighted combination method has been employed so as to improve the forecasting results. Reference [10] has proved that it is workable by using combined model of BP neural network theory and grey model to predict demand of electric power per person in China. Reference [11] introduced a combined forecasting method and proved that the error of the combined model is less than that of any single model. However, many studies only focus on the combination of several traditional forecasting model.

To obtain best forecasting results, the residual grey theory and improved BP network model built by additional momentum-adaptive learning rate (AMALR for short) are introduced to combine in the study. The combination forecasting model with optimal weighted coefficient can use various prediction sample information at utmost. Compared with single forecasting model, it is more systematically, more comprehensive and can effectively reduce the influence of random factors, which can result in high forecasting precision and lower forecasting risk.

II. PRINCIPLE OF WEIGHTED COMBINATION FORECASTING

Assuming the actual values for a forecasting problem are $f_t(t=1,2,\ldots,n)$ over a period. There are m forecasting methods for this problem, and the forecasting value of the ith method is f_{it} at t time (i=1,2,... , m). The corresponding forecasting absolute error is $\varepsilon_{it}=f_t-f_{it}$. Assuming the weights of various forecasting methods are $w_1, w_2, \ldots, w_m$, then the weighted combination forecasting model can be expressed as:

$$\hat{f}_t = \sum_{i=1}^{m} w_i f_{it} \qquad (t=1,2,...,n) \tag{1}$$

Where, $\sum_{i=1}^{m} w_i = 1$.

The forecasting absolute error and relative error of weighted combination forecasting model can be expressed as respectively:

$$\varepsilon_t = f_t - \hat{f}_t \tag{2}$$

$$\eta_t = \frac{\varepsilon_t}{f_t} \times 100\% \tag{3}$$

The key problem of weighted combination forecasting method is to determine the weight coefficients of various forecasting models. Often it requires ε_t and η_t under all weight coefficients to be as small as possible. So, the norm performance indices are introduced:

$$\min J_1 = (\sum_{t=1}^{n} |\varepsilon_t|^p)^{\frac{1}{p}} \tag{4}$$

$$\min J_2 = (\sum_{t=1}^{n} |\eta_t|^p)^{\frac{1}{p}} \tag{5}$$

Supposed the absolute error ε_t as the optimization criterion, the optimal weighted combination(OWC for short) forecasting problem can be turn into the conditional extreme problem, like:

$$\begin{cases} \varepsilon_t = f_t - \hat{f}_t \\ \min J_1 = (\sum_{t=1}^{n} |\varepsilon_t|^p)^{\frac{1}{p}} \\ \sum_{i=1}^{m} w_i = 1 \\ w_i \ge 0 \quad (i=1,2,....m) \end{cases} \tag{6}$$

III. OWC FORECASTING MODEL

A. Residual grey theory model

Let residual sequence be $\varepsilon^{(0)}$={ $\varepsilon^{(0)}(1)$, $\varepsilon^{(0)}(2)$, $\varepsilon^{(0)}(3)$,…, $\varepsilon^{(0)}(n)$ }. If there is k_0 satisfied:

(1) $\forall$ $k \ge k_0$, sign of $\varepsilon^{(0)}(k)$ consistency;

(2) n-$k_0 \ge 4$.

Then, $\varepsilon^{(0)}$={ $\varepsilon^{(0)}(k_0)$, $\varepsilon^{(0)}(k_0+1)$,…, $\varepsilon(0)(n)$ } is named as residual tail section which can build.

As to original sequence $x^0(k)$, forecasting sequence can be written as follows after building GM(1,1) model and counting down:

$$x'^{(0)}(k+1) = (1-e^{a'})[x^{(0)}(1) - \frac{u'}{a'}]e^{-a'k} \tag{7}$$

Where, k=0, 1, 2…n.

Residual sequence can be obtained by difference between original and forecasting sequence and expressed as:

$$\begin{aligned} \varepsilon^{(0)}(k) &= x^{(0)}(k) - x'^{(0)}(k) \\ &= \{\varepsilon^{(0)}(1), \varepsilon^{(0)}(2), ..., \varepsilon^{(0)}(n)\} \end{aligned} \tag{8}$$

Take part of sequence of $\varepsilon^{(0)}(k)$ and write as:

$$\varepsilon^{(0)}(k') = \{\varepsilon^{(0)}(1'), \varepsilon^{(0)}(2'), ..., \varepsilon^{(0)}(n')\} \tag{9}$$

After building GM(1,1) model of $\varepsilon^{(0)}(k')$ and counting down, forecasting sequence is expressed as:

$$\varepsilon'^{(0)}(k'+1) = (1-e^{a''})[\varepsilon^{(0)}(1') - \frac{u''}{a''}]e^{-a''k'} \tag{10}$$

Taking $\varepsilon'^{(0)}(k'+1)$ as modified model of $x'^{(0)}(k+1)$, then formula (7) changes to following form:

$$\begin{aligned} x'^{(0)}(k+1) = (1-e^{a'})[x^{(0)}(1) - \frac{u'}{a'}]e^{-a'k} + ... \\ + \delta(k-i)(1-e^{a''})[\varepsilon^{(0)}(1) - \frac{u''}{a''}]e^{-a''k'} \end{aligned} \tag{11}$$

Where, $\delta(k-i) = \begin{cases} 1 & k \ge i \\ 0 & k < i \end{cases}, \quad i = n - n'$.

The parameters including a', u', a'' and u'' which present in formula (7) to (11) represent whiten differential equation parameters.

B. BP neural network model with AMALR

Combining of additional momentum method and adaptive learning rate algorithm, it makes full use of their respective advantages. BP algorithm can find the global optimal solution using momentum method and the training time of BP network can be shorten while using adaptive learning rate. Thus combining the two algorithms, it will not only effectively inhibit the network into a local minimum but improve the efficiency of learning.

If BP network has L+1 layers, the first layer is named as input layer and the (L+1)th layer output layer. All of layers between input layer and output layer are called hidden layer. P denotes training sample number. Output variant of the jth of the first layer neuron uses the sign $O_{pj}^{(l)}$. $W_{ij}^{(l)}$ denotes the neuron weight between the jth of the lth layer neuron and the ith of the (l+1)th layer neuron. Therefore, the input/output relation of neurons can represent as follows:

$$net_{pi}(l+1) = \sum_{j=1}^{N_l} W_{ij}^{(l)} O_{pj}^{(l)} - \theta_i^{(l+1)} \tag{12}$$

$$O_{pj}^{(l+1)} = f_l(net_{pi}^{(l+1)}) \tag{13}$$

Where, f_l represents non-linear differentiable non-decreasing function. Generally, its form is S-form function:

$$f_l(x) = \frac{1}{1+e^{-x}} \tag{14}$$

Adjustment formulas introduced with additional momentum factor weights and thresholds are written as follows:

$$\Delta w_{ij}(k+1)=(1-mc)\eta\delta_i p_j + mc\Delta w_{ij}(k) \quad (15)$$

$$\Delta\theta_i(k+1)=(1-mc)\eta\delta_i + mc\Delta\theta_i(k) \quad (16)$$

Where, k represents training times and mc represents momentum factor which empirical value is about 0.95. η represents learning step. δ represents node error and p represents neuron input.

In training program design, determining conditions of momentum method is described as:

$$mc=\begin{cases} 0 & E(k)>1.04E(k-1) \\ 0.95 & E(k)<E(k-1) \\ mc & others \end{cases} \quad (17)$$

Where, $E(k)$ represents error square sum of the kth step.

Adjustment formula of adaptive learning rate is described as:

$$\eta(k+1)=\begin{cases} 1.05\eta(k) & E(k+1)<E(k) \\ 0.7\eta(k) & E(k+1)>1.04E(k) \\ \eta(k) & others \end{cases} \quad (18)$$

C. OWC forecasting model

Supposing that $f_1(x)$ is residual grey theory model forecasting value and $f_2(x)$ is BP neural network forecasting value with AMALR, the weighted combination forecasting model can be expressed as:

$$f(x)=w_1 f_1(x)+w_2 f_2(x) \quad (19)$$

Where, $f(x)$ represents combination forecasting value, w_1 and w_2 represent the weighted coefficients of $f_1(x)$ and $f_2(x)$ respectively.

In order to obtain the optimal weighted coefficient, w_1 is set a certain step increasing. Combination forecasting value in each kind of weighted coefficient can be calculated by Matlab. Then mean square error between combination forecasting value and the actual value will be computed in each weighted coefficient to measure the fitting degree of forecasting value and the actual value. It can obtain the optimal weighted coefficient when the mean square error reaches the minimum.

IV. SIMULATION EXPERIMENTS

The power consumption original data of a power supply company in 2001-2011 for load forecasting and analysis is presented in table 1. It is obviously increasing trend year by year. But the growth rate is different and has no rule to conclude. In order to predict the annual power consumption after the 2011, residual grey theory method, BP neural network method with AMALR and the optimal weighted combination forecasting method are used. First, the power consumption values of 2001-2008 are used to create the forecasting model. Then, the forecasting values of 2009-2011 are computed by the three methods respectively and compared with actual power consumption in the precision.

TABLE I. THE POWER CONSUMPTION OF A POWER SUPPLY COMPANY IN 2001-2011

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Power consumption (Million kilowatts)	17.6655	18.2664	18.9805	19.0963	19.4226	21.2219	23.5861	25.8091	27.7613	30.1056	32.74013

According to the formula (19), each weighted combination forecasting value can be calculated by f_1 and f_2 weighted combination which w_1 ranges from 0 to 1 and the step length is set at 0.05. Then the mean square error between combination forecasting values and the actual values will be computed to determine the regression level between forecast values and actual values in each corresponding weighted coefficient. Fig. 1 shows the mean square error curve in different weighted coefficients.

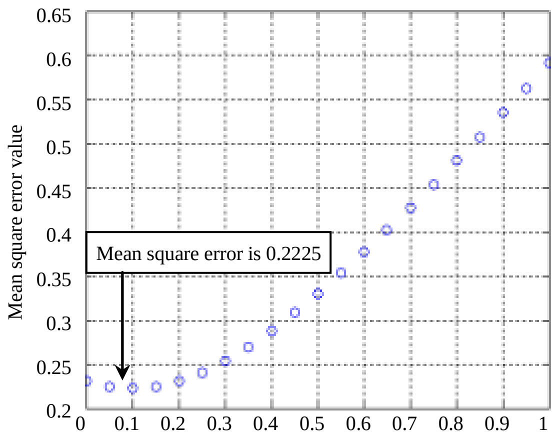

Figure 1. Mean square error in different weighted coefficients of OWC

Obviously, when w_1 equals 0.1, namely w_2 equals 0.9, the mean square error between combination forecasting values and the actual values is 0.2225, which reaches minimum. Fig. 2 and fig. 3 represent the comparison curves and their partial curves of forecasting values and the actual values under three forecasting methods respectively. From figures 2-3, it can be seen that the variation trends of the OWC forecasting values and the actual values are similar, and the fitting effect of OWC forecasting method is best. Table 2 gives the list of forecasting values, relative error and mean square error under three forecasting methods.

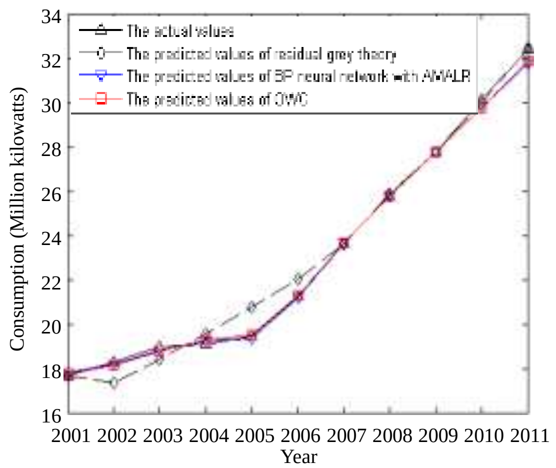

Figure 2. Comparison curves of forecasting values and actual values under three forecasting methods

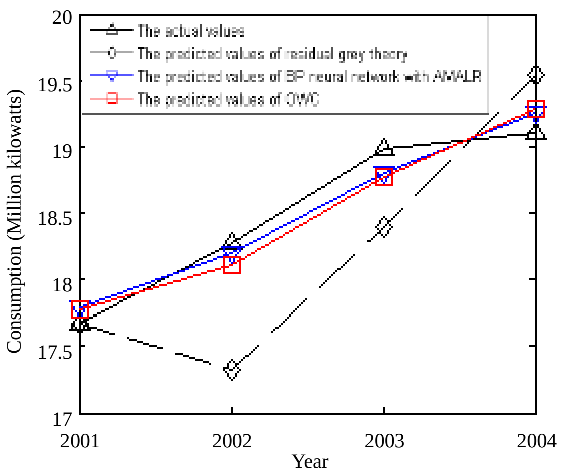

Figure 3. The partial comparison curves of 2001-2004

From table 2, it can be seen that the forecasting mean square error of BP neural network with AMALR is 0.2307, which is better than that of residual grey model (its mean square error is 0.5910) as single model. The mean square error of combination forecasting method is 0.2225, which is less than any single model. All above analysis indicates that combination forecasting model is feasible and better than other two single models. Compared with the simulation

V. CONCLUSION

Combination forecasting method is one of the hottest topics in current forecasting scientific research. The object of load optimization combination forecasting is taking full advantage of useful information of various load forecasting model and improving the forecasting precision as far as possible. This study improves traditional forecasting methods and combined improved grey theory and improved BP neural network with weighted. Simulation results show that the OWC forecasting method can largely use various forecasting sample information and achieve well results in load forecasting. Using this method, it also can accurately predict the power consumptions are 34.4129 million kilowatts, 37.2165 million kilowatts and 39.6010 million kilowatts during 2012 to 2014. It can provide reliably reference for electric power planning in the region.

ACKNOWLEDGMENT

This study thanks the Anhui provincial natural science foundation (No. 1208085MF115) and national training programs of innovation and entrepreneurship for undergraduates (201310363048).

REFERENCES

[1] YANG Hua-fen. Short-term load forecasting in power system based on improved fuzzy neuro net. J.Changchun Inst.Tech. (Nat. Sci. Edi.), 2009, 10(1):68-71.

[2] MAO Wen-jin, ZHENG Shou-chun. Improved GM (1,1) model and application in natural gas demand forecasting of Sichuan province. Journal of Southwest Petroleum University (Social Sciences Edition) , 2010,3 (6): 19-21.

[3] XIE Li-chun. BP neural network algorithm improvements and convergence analysis. Computing Technology and Automation, 2007, 26(3):52-56.

[4] ZHOU De-qiang. Application of improved gray verhulst model in middle and long term load forecasting. Power System Technology,2009, 3(18): 124-127.

[5] HUANG Yuan-sheng, ZHANG Hui-juan. Application of grey relational grade based combined weights to power load forecasting.East China Electric Power, 2009,37(2):225-228.

[6] WANG Shu-sen , ZHAO Dong-ling. Improved BP algorithm based on the additional momentum method. Journal of Jiyuan Vocational and Technical College, 2012, 11(3): 9-13.

[7] LIU Hong, XIA Li-xian, LIAO Li-ping, ZHANG Zhi-guo. Application of the resilient back-propagation in prediction of the gold deposits in Guizhou. Progress in Geophysics, 2005,20(2):399-401.

[8] TIAN Xiu-mei. Research on the improvement and simulation of BP algorithm. Electronic Technology, 2011,(3):65-66.

[9] HUANG Xi-ping, LI Rui , LIU Jun. A novel predictive control method based on neural networks. Computer Simulation, 2006,23(4): 154-156.

[10] ZHANG jiang-yong, XU fei-yun. Application of combination forecasting for demand of electric power per person in China. Heilongjiang Science and Technology Information, 2008,(21):12-13.

[11] CHENG Jian-dong, DU Ji-gui. Application of combined method in power load forecastin. Jiangsu Electrical Engineering, 2011, 30(6):38-41.

TABLE II. THE LIST OF FORECASTING VALUES, RELATIVE ERROR AND MEAN SQUARE ERROR UNDER THREE FORECASTING METHODS

Serial number	Year	Residual grey theory		BP neural network with AMALR		Optimal weighted combination	
		Forecasting values (Million kilowatts)	relative error	Forecasting values (Million kilowatts)	relative error	Forecasting values (Million kilowatts)	relative error
1	2001	17.6655	0%	17.7953	-0.74%	17.7823	-0.66%
2	2002	17.3223	5.17%	18.1985	0.37%	18.1109	0.85%
3	2003	18.3945	3.09%	18.8067	0.92%	18.7655	1.13%
4	2004	19.5332	-2.29%	19.2577	-0.85%	19.2853	-0.99%
5	2005	20.7424	-6.79%	19.3602	0.32%	19.4984	-0.39%
6	2006	22.0263	-3.79%	21.2087	0.06%	21.2905	-0.32%
7	2007	23.5861	0%	23.6723	-0.37%	23.6637	-0.31%
8	2008	25.8570	-0.19%	25.7531	0.22%	25.7635	0.18%
9	2009	27.7828	-0.08%	27.7758	-0.05%	27.7765	-0.05%
10	2010	29.9516	0.51%	29.7713	1.11%	29.7893	1.05%
11	2011	32.4256	-0.08%	31.7838	1.91%	31.8480	1.71%
mean square error		0.5910		0.2307		0.2225	