Computers and Mathematics with Applications 57 (2009) 1883–1889

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

Middle-long power load forecasting based on particle swarm optimization☆

Dongxiao Niu[a], Jinchao Li[a,*], Jinying Li[b], Da Liu[a]

[a] School of Business Administration, North China Electric Power University, Beijing, 102206, China

[b] Department of Economic Administration, North China Electric Power University, Baoding, 071000, China

ARTICLE INFO

Keywords:
Power load forecasting
Error index
Entropy
Particle swarm optimization

ABSTRACT

Middle-long forecasting of electric power load is crucial to electric investment, which is the guarantee of the healthy development of electric industry. In this paper, the particle swarm optimization (PSO) is used as a training algorithm to obtain the weights of the single forecasting method to form the combined forecasting method. Firstly, several forecasting methods are used to do middle-long power load forecasting. Then the upper forecasting methods are measured by several indices and the entropy method is used to form a comprehensive forecasting method evaluation index, following which the PSO is used to attain a combined forecasting method (PSOCF) with the best objective function value. We then obtain the final result by adding all the results of every single forecasting method. Taking actual load data of a power grid company in North China as a sample, the results show that PSOCF model improves the forecasting precision compared to the traditional models.

1. Introduction

Since middle-long electric power load forecasting plays a key role in the decision of the power system investment and power system planning and operation, the researchers have to pay more and more attention to the research of middle-long term electric power load forecasting.

During recent decades, numerous investigations have been proposed to improve the accuracy of electricity load forecasting, such as expert systems [1], fuzzy inference [2], artificial neural networks (ANN) [3] and so on. All of the load forecasting methods can be sorted in three categories approximately. (1) The traditional methods such as the method of analogy, the method of proportional, the regression calculation model, time series prediction model and grey forecasting model; (2) New load forecasting algorithms, such as artificial neural network, support vector machine and so on. Artificial neural network does not need the expression of human experience. It aims to establish a network between the input data set and the observed output data set. It is good at dealing with the nonlinear relationship between the load and its relative factors, but the shortcoming lies in overfitting and long training time [4]. (3) Combined forecasting methods: In recent years, the forecasting method about middle-long load forecasting has come from the single forecasting method to the combined forecasting method. The combined forecasting method mostly obtained the weight of the single forecasting which is used for the combined method based on some single error index [5–7]; the combined results cannot improve the forecasting result completely. In this paper, the mean absolute error (MAE), the mean absolute percentage error (MAPE) and the relative degree are used to evaluate the forecasting method (the method of analogy, the method of proportional, the regression calculation

☆ The project was supported by the National Natural Science Foundation of China (Grant Nos. 70671039).

* Corresponding author.
E-mail addresses: niudx@126.com (D. Niu), gsyljch@163.com (J. Li).

0898-1221/$ – see front matter
doi:10.1016/j.camwa.2008.10.044

model, time series prediction model, grey forecasting model and BP forecasting method [8]) from the angle of nicety, trend and so on. Then the entropy method is used to calculate the weight of the three indices to form a comprehensive evaluation index about the single forecasting method, following which the objective function for the particle swarm optimization can be described by the comprehensive evaluation index. At last, the particle swarm optimization is used to calculate the weights of the single forecasting methods with the best objective function result and the improvement of the upper three evaluation indices.

2. Combined forecasting model

The combined forecasting method is that with the proper selected weights, several forecasting methods' results are added up. Its mathematics' model is shown here.

$$\hat{y}_t = \sum_{i=1}^{k} w_i y_{it} \, (t = 1, 2, \ldots, n), \quad \sum_{i=1}^{k} w_i = 1 \tag{2.1}$$

here, $y_{it} \, (i = 1, 2, \ldots, k;\ t = 1, 2, \ldots, n)$ is the forecasting value of the ith forecasting method at t time. k is the numbers of the single forecasting method. w_i is the weight of the ith forecasting method. The foundation of the good combined forecasting is the properly selected forecasting methods which have different information and are representational basic forecasting methods. In this paper, the method of proportional (MP), the regression calculation model (RC), the grey forecasting model (GM) and BP neural network (BP) are selected.

2.1. The foundation of the forecasting model

2.1.1. The method of proportional

$$K = \sqrt[m-n]{\frac{A_m}{A_n}} - 1, \qquad A_l = A_n(1+K)^{l-n} \tag{2.2}$$

here, A_m is the load amount of the *mth* year, A_n is the load amount of the *nth* year.

2.1.2. The method of the line regression calculation

$$y^{(1)}(k+1) = \widehat{a} + \widehat{b}x(k+1). \tag{2.3}$$

2.1.3. The method of grey forecasting model [9]

General modeling is to set up difference equation with initial series directly, but grey-modeling needs generating operation (AGO) of the initial series. Noise pollution makes messy series known as grey series or grey processes. So the model of grey process is termed the grey model (GM).

GM (1, 1) is one kind of grey models used most frequently. The formula is shown following:

$$y^{(1)}(k+1) = (\mathrm{e}^{-\widehat{a}} - 1)\left[x^{(0)}(1) - \frac{\widehat{u}}{\widehat{a}}\right]\mathrm{e}^{-\widehat{a}k}. \tag{2.4}$$

2.1.4. BP neural network forecasting method [10,11]

The learning algorithm of BP Network is as the following: The first step: Set the initial parameter ω and θ (ω is the initial weight, θ is the critical value, randomly let both of them be fairly small number). The second step: Input the known sample to the Network and calculate the output value, using

$$y_j = \left[1 + \mathrm{e}^{-(\sum \omega_{ij} x_i - \theta_j)}\right]^{-1} \tag{2.5}$$

where x_i is the input of that junction ($i = 1, \ldots, m$); w_{ij} is the connection weight from i to j ($i = 1, \ldots, m;\ j = 1, \ldots, n$), let the initial weight be fairly small number within[0, 1]; θ_j is the critical value; y_i is the calculated value. The third step: Adjust the weight coefficient ω on the basis of the difference ($d_j - y_j$) between the known output value and the calculated one.The adjustment is calculated, using

$$\Delta\omega_{ij} = \eta \delta_i x_j \tag{2.6}$$

where η is the ration coefficient (learning rate), x_j is the input, d_j is the actual output of the sample, δ_j is output deviation. Regarding to η, it is a small number within [0, 1]. Under the presupposition that oscillation is not stirred and a fairly good precision is guaranteed, the value of η can be increased step by step until a satisfactory training speed is reached. Regarding to x_j, it is the Network input to the junctions in the intermediate layer, but to junctions in output layer, it is the intermediate junctions' output. Regarding to δ_j, it is a value related to the output deviation. To the junctions in the output layer, it is calculated using

$$\delta_j = \eta_j(1 - y_i)(d_j - y_j). \tag{2.7}$$

To the junctions in the intermediate layer whose output are hard to compare, its value can be acquired by counter calculation using

$$\delta_j = x_j(1 - x_j) \sum \delta_k \omega_k \tag{2.8}$$

where k mean all the junctions in the output layer should be taken into account, the deviation δ_j is got by reverse calculation from the output. After being adjusted, the weight of the neuron in each layer is as the following:

$$\omega_{ij}(t) = \omega_{ij}(t - 1) + \Delta\omega_{ij} \tag{2.9}$$

where t is the learning time.

This algorithm is an interaction process in which all the values of ω are adjusted in each round. Such interaction is repeated until the output deviation is less than an acceptable value, then a good Network is trained successfully. It is the essence of the BP algorithms to turn the first grade sample input question into a nonlinear optimize question. Gradient decreasing method used in the BP algorithms is one of the most ordinary methods in optimization technology, while calculating the weight value by interactive computation is equals to the learning memory question.

2.2. The introduction of the evaluation indices about the forecasting method

2.2.1. Mean absolute error

Because there are negative and positive forecasting errors, so the mean absolute error is used as one of the evaluation index in order to forbid the counteraction of the positive and negative forecasting errors.

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |y_t - \widehat{y_t}| . \tag{2.10}$$

2.2.2. Mean absolute percentage error

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left|\frac{y_t - \widehat{y}_t}{y_t}\right| . \tag{2.11}$$

2.2.3. Relativity degree

This index shows the similar degree between two curves, in other words, the more similar the two curves in the geometry figure, the more the likely their development trend and the smaller the fitness error.

$$r_i = \frac{1}{n}\sum_{k=1}^{n} \xi_i(k) \tag{2.12}$$

$$\xi_i(k) = \frac{\text{minmin}\,|y_0(k) - y_i(k)| + \rho\text{maxmax}\,|y_0(k) - y_i(k)|}{|y_0(k) - y_i(k)| + \rho\text{maxmax}\,|y_0(k) - y_i(k)|} \tag{2.13}$$

where, ρ is the differentiation coefficient, it is a number between 0 and 1. In general, $\rho = 0.5$.

2.3. The weights calculated by entropy method

Entropy comes from thermodynamics, which is the quotient of variation of heat energy that cannot work divided by temperature. According to the information theory, entropy is a relevant measure of order and disorder in a dynamic system. It is used in many fields such as management and engineering. Assume the set of stations reflecting system is $A = a_1, a_2, \ldots, a_n$, the probability of a_j is p_j, $(j = 1, 2, \ldots, n)$, and this system has entropy:

$$E = -\sum p_j \ln p_j \tag{2.14}$$

when $p_j = \frac{1}{n}$, E reaches the maximum. $E_{\max} = \ln n$, we evaluate m single forecasting methods with n criteria and get the evaluation matrix . . .

$$X = (x_{ij})_{m\times n} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \tag{2.15}$$

x_{ij} is the values of error indices i with j. According to the feature of indicators, to find out the primary evaluation:

$$x_j^* = \begin{cases} \max(x_{ij}), & \text{for benefit criterion,} \\ \min(x_{ij}), & \text{for cost criterion.} \end{cases} \tag{2.16}$$

Then the distance between x_{ij} and the ideal value x_i^* is

$$D_{ij} = \begin{cases} \dfrac{x_{ij}}{x_j^*}, & x_j^* = \max(x_{ij}), \\ \dfrac{x_j^*}{x_{ij}}, & x_j^* = \min(x_{ij}). \end{cases} \tag{2.17}$$

Normalize it with

$$d_{ij} = D_{ij} \Big/ \sum_{j=1}^{n} \sum_{i=1}^{m} D_{ij}. \tag{2.18}$$

And the relative importance of criterion j to can be measured by entropy E:

$$E = -\sum_{i=1}^{m} \frac{d_{ij}}{d_j} \ln \frac{d_{ij}}{d_j}, \quad d_j = \sum_{i=1}^{m} d_{ij}. \tag{2.19}$$

Normalize E and get the importance entropy of criterion j:

$$e(d_j) = -\frac{1}{\ln m} \sum_{i=1}^{m} \frac{d_{ij}}{d_j} \ln \frac{d_{ij}}{d_j}. \tag{2.20}$$

According to $e(d_j)$, we get the weight of criterion j:

$$\theta_j = \left[1 - e(d_j)/(n - E_e)\right], \quad E_e = \sum e(d_j). \tag{2.21}$$

2.4. The introduction of the particle swarm optimization algorithm [12]

In the D dimension space, number of particle is m, the position of the particle i is $\vec{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iD})$, $i = 1, 2, \ldots, m$, the speed is $\vec{v}_i = (v_{i1}, v_{i2}, \ldots, v_{iD})$. Putting $\vec{x}$ into object function will calculate adaptive value. The best position for the particle i is $\vec{p}_i = (p_{i1}, p_{i2}, \ldots, p_{iD})$. The best position of the all is $\vec{p}_g = (p_{g1}, p_{g2}, \ldots, p_{gD})$.The updating operation is:

$$v_{id} = \omega v_{id} + c_1 r_1 (p_{id} - x_{id}) + c_2 r_2 (p_{gd} - x_{gd}) \tag{2.22}$$

$$x_{id} = x_{id} + v_{id} \tag{2.23}$$

where $i = 1, 2, \ldots, m$, $d = 1, 2, \ldots, D$; ω is no-negative constant is called inertia factor and it may reduced with linearly; learning factor c_1 and c_2 is no-negative constant; r_1 and r_2 is [0,1] randomly; $v_{id} \in [-v_{\max}, v_{\max}]$; $v_{\max}$ is constant.

The basic process of PSO algorithm can be described as following:

(1) Initialized particle swarm, set initialized position x and initialized speed v;
(2) Calculate adaptability of each particle;
(3) Compared its sufficiency with the sufficiency of the best position p_{id} ever, leaving the better one;
(4) For each particle, compared its adaptability with adaptability of the best position p_{id} the swarm ever, leaving the better one;
(5) According to the upper two formulas adjust the speed and position of the particle; The process end when the end off condition is fulfilled. If not, turn to step (2).

3. Case study

In this paper, we set up a combined forecasting method based on PSO (PSOCF). Then we use the load data of Baoding City which in Hebei province in the north China to test the method PSOCF. The load data are from January 2004 to January 2006.

The objective function of PSO is the following.

$$\min F = \sum_{j=1}^{m} w_{ej} \times g_j \left(\sum_{i=1}^{n} (w_i \times f_i) \right) \tag{3.1}$$

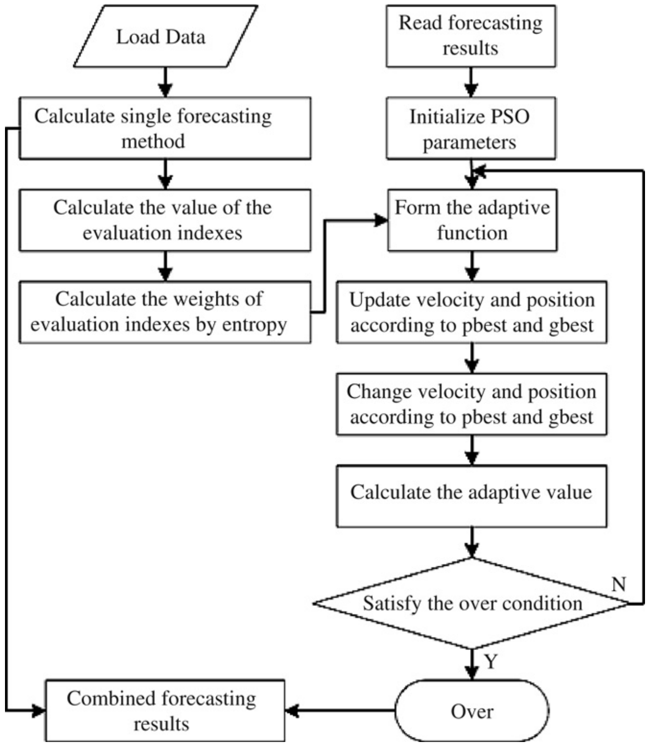

Fig. 1. The processes of the PSOCF.

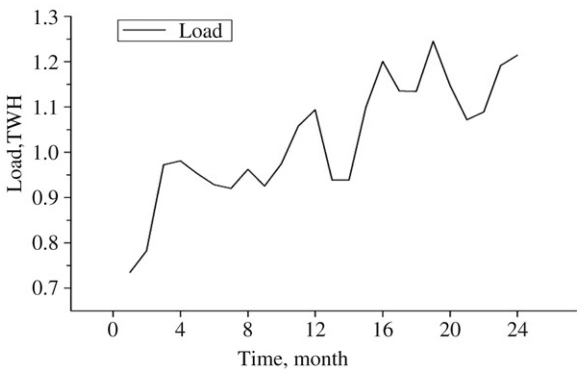

Fig. 2. Month load data between Jan 04 and Jan 06 of Baoding City.

here, w_{ej} are the weights of the upper evaluation indices which are got by the entropy method, w_j are the weights of the upper selected forecasting methods which got by PSO method. f_i delegate the forecasting methods. g_j delegate the evaluation indices.

The processes of the upper method are shown in Fig. 1.

The results are shown in Fig. 2.

The Method of Proportional (MP)′, method of the line regression calculation (RC)′, method of grey forecasting model (GM)′ and BP neural networks (BP)′ forecasting results are shown in Fig. 3.

The evaluation indices (MAE, MAPE, RD)′ calculating results are shown in Table 1.

The weights of the evaluation indices are 0.331, 0.3343, 0.3347, which calculated by the entropy method. At last, we use the Matlab.7.0 software and PSO toolbox to realize the improved PSO method, the calculated results are 0.0013, 0.0006, 0.0432, 0.9549 which are the weights for the single forecasting methods to combine. The last forecasting results are shown in Fig. 4.

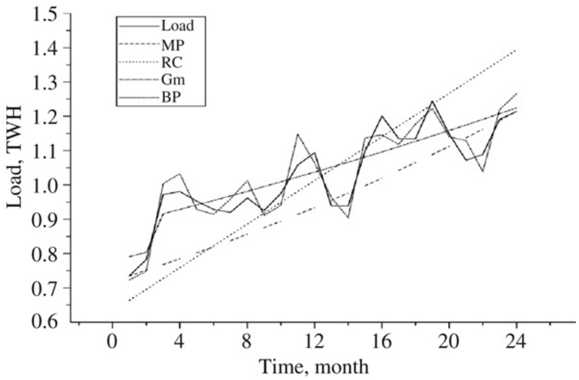

Fig. 3. The results of several forecasting methods.

Table 1
The evaluation results of forecasting method.

Method	MAE	MAPE	RD
MP	89 449.72	0.09	0.60
RC	108 082.1	0.11	0.55
GM	51 006.88	0.05	0.72
BP	35 544.76	0.03	0.78

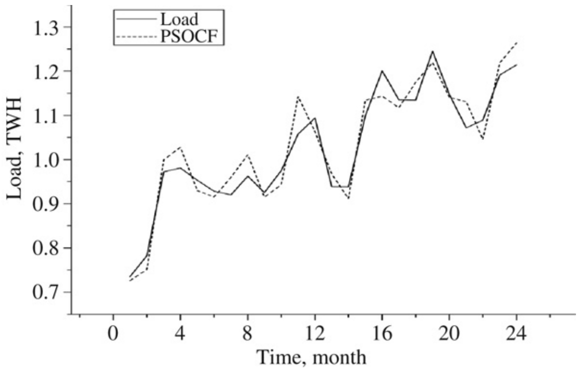

Fig. 4. The comparison between combined forecasting results and real load.

4. Conclusions

In this paper, a combined forecasting method based on particle swarm optimization method is set up. Firstly, several forecasting methods which are successful and practical in the middle-long load forecasting are selected. Secondly, we use mean absolute error, mean absolute percentage error and relativity degree to evaluate the single forecasting method. Thirdly, we use the entropy method to calculate the weights of the upper three evaluation indices. At last, the particle swarm optimization method is used to calculate the weights of the single forecasting methods to form the combined forecasting results. This method offsets the shortage of the single forecasting and evaluation by a single error index. The combined forecasting method in this paper improved the forecasting stability and reliability.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China (NSFC) (70671039).

References

[1] K.J. Hwan, G.W. Kim, A short-term load forecasting expert system, in: Proceedings of the Fifth Russian-Korean International Symposium on Science and Technology, 1 (1), 2001, pp. 112–116.
[2] A.A. Desouky, M.M. Elkateb, Hybrid adaptive techniques for electric load forecast using ANN and ARIMA, IEEE Proceedings Generation, Transmission and Distribution 147 (4) (2000) 213–217.
[3] K.H. Kim, H.A. Youn, Y.C. Kang, Short-term load forecasting for special days in anomalous load conditions using neural networks and fuzzy inference method, IEEE Transactions on Power Systems 15 (2) (2000) 559–565.

[4] J. Yang, J. Stenzel, Short-term load forecasting with increment regression tree, Electric Power Systems Research 76 (2006) 880–888.
[5] Jiangang Yao, Jian Zhang, Electric Market Analysis, Higher Education Press, Beijing, 1996.
[6] M.S. Chen, J.W. Han, P.S. Yu, Data mining: An overview from a database perspective, IEEE Transaction on Knowledge and Data Engineering 8 (8) (1996) 1–41.
[7] L.D. Chen, S. Toru, Data mining methods, applications, and tools, Information System Management 17 (1) (2000) 65–70.
[8] D. Niu, S. Cao, L. Zhao, Power Load Forecasting Technology and its Application, China Electric Power Press, Beijing, 2001.
[9] Y. Luo, F. Luo, Accident Predicted by the Gray Dynamic Model, China Offshore Oil and Gas Engineering 14 (2) (2002) 56–59.
[10] T. Senjyu, One-hour-ahead load forecasting using neural network, IEEE Transactions on Power Systems 17 (1) (2002) 113–118.
[11] H. Xie, H. Cheng, G. Zhang, D. Niu, J. Yang, Applying rough set theory to establish artificial neural networks for short term load forecasting, Proceedings of the Csee 23 (11) (2003) 1–4.
[12] D.W. Boeringer, D.H. Werner, A comparison of particle swarm optimization and genetic algorithms for a phased array synthesis problem, Antennas and Propagation Society International Symposium 1 (1) (2003) 181–184.