

Research Article

Multi-Step-Ahead Combination Forecasting of Wind Speed Using Artificial Neural Networks

Dongfeng Wang, Fuqiang Wang and Xiaoyan Wang

Department of Automation, North China Electric Power University, Baoding 071003, China

Abstract: Wind speed plays a very important role in the scheduling of power systems and dynamic control of wind turbine. Wind speed forecasting has become one of the most important issue for wind energy conversion recently. Adaptive and reliable methods and techniques for wind speed forecasting are urgently needed in view of its stochastic nature that varies from time to time and from site to site. Back Propagation (BP) algorithm-based neural network, which is a commonly computational intelligence method, has been widely used in forecasting fields. But it does have some deficiencies and uncertainties, for example, the hidden nodes of BP directly affect the network's generalization ability and accuracy, but there is not yet an effective theory to determine the number of hidden nodes. In order to solve the problem of BP network, a combination forecasting model with differently weighed BP networks is proposed in this study. Wind speed data collected from a New Zealand wind power plant is used for experiment research. Simulations show that the results of combination forecasting method is better than those of only one BP network.

Keywords: Artificial neural networks, combination forecasting, wind speed

INTRODUCTION

In recent years, wind power technology has become one of the fastest renewable energy generation. The installed capacity grows about 30% every year. It was reported that wind power capacity will reach 30,000 MW in 2020 in China.

Scholars at home and abroad have researched on the wind power generation extensively and profoundly. Short-term wind speed forecasting plays a significant role in the dynamic control of the wind power system. Accurate wind speed forecasting is urgently needed for timely scheduling, capacity evaluation and determination of reasonable generation price (Li and Shi, 2010; Li *et al.*, 2011).

Wind speed is difficult to predict because it is affected by a variety of complex factors, such as pressure, temperature, earth's rotation, geomorphology and so on. Existing methods for short-term wind speed forecasting include *Neural Network* (NN), Kalman filter, wavelet analysis (Pan *et al.*, 2008), moving average method, spatial correlation model, fuzzy evaluation, linear prediction, discrete Ha Bote transformation and so on.

The Back Propagation (BP) NN has been widely used for prediction because its good nonlinear quality, high fitting accuracy, flexible and effective learning method, fully distributed storage structure and hierarchy quality of the model structure. However, it is

worth noting that its forecast accuracy is easily affected by some factors, such as network structure, learning rate, number of input-nodes and hidden-nodes, etc.

In BP algorithm, the error is propagated back to the input layer from the output layer, thus the more number of hidden layers, the less reliability of the error, especially near the input layer. If unreliable error is used to amend the weight, then the learning efficiency could be affected and finally result in slow convergence speed or even no convergence. Another disadvantage of BP algorithm is that the layer number, the hidden nodes number and the adjustment of weights are manually set and adjusted by trial, which may also increase the randomness of the algorithm. In order to improve forecasting accuracy of BP, methods like increasing the hidden layer number, the input nodes number or the sample size are used. But they also increase the system complexity and calculation amount, slow down the convergence speed and learning efficiency, reduce the generalization ability and prediction accuracy (Sfetsos, 2000; Sharm and Friedlander, 1984).

Systematic study on combination forecasting started from Bates and Granger and their research attracted the attention of scholars. The combination forecasting method gained further development in 1970s. In 1989, *Journal of Forecasting*, which is an international authoritative journal in the forecasting field, published a special issue on combination forecasting. Domestic scholars have also paid their

Corresponding Author: Dongfeng Wang, Department of Automation, North China Electric Power University, Baoding 071003, China

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

attention on this field and have obtained some results in recent decades. The method of combination forecasting is to combine different prediction models with appropriate weighted average form to derive a new model including all the information of various models (Chen, 2008; Andrawisa *et al.*, 2010). In order to effectively improve the prediction accuracy, the method of calculating weighted average coefficients become the key problem. It has been proved that combination forecasting is an effective way to increase prediction accuracy (Andrawisa *et al.*, 2010). The most significant characteristic of this method is that it can overcome the shortcoming of one single method among so many ones. It is indicated in many literatures that the combination forecasting is more effective than a single prediction scheme. Structure change and parameter drift over time in time series makes it hard to select one single best prediction method. The combination forecasting can decrease those adverse effects effectively.

In order to increase the accuracy and generalization ability of the wind prediction model and overcome the difficulty in determining the hidden nodes of BP, a combination forecasting wind model is proposed in this study. Firstly, some BP models with different hidden nodes are developed. Then a multi-step combination forecasting model is established using corresponding theory. Results show that the model is more accurate and reliable.

BP NEURAL NETWORK

Principle of BP neural network: BP network is a multi-layer feed-forward network. It systematically solves the learning problem of connection weights between hidden layers and has become the most widely used method of neural network learning.

The central idea of BP algorithm is to adjust the weights to minimize the total network's error and ulteriorly minimize the mean of squared error between the actual output and the desired output. The learning process is actually the process of weights adjustment during the propagation of error.

The learning process of a multi-layer BP network includes both forward propagation and back propagation. During the process of forward propagation, the input information is first conducted from the input layer, then handled by the hidden layer and finally to the output layer. The neuron state of each layer only affects its next layer's neuron state. If the expected output cannot be obtained in the output layer, the error will be propagated back along the original connection channel. During the process of back propagation, the weights will be adjusted to minimize the error. Typical structure of BP neural network is shown in Fig. 1.

For learning sample p , the input of the i -th neuron in hidden layer is:

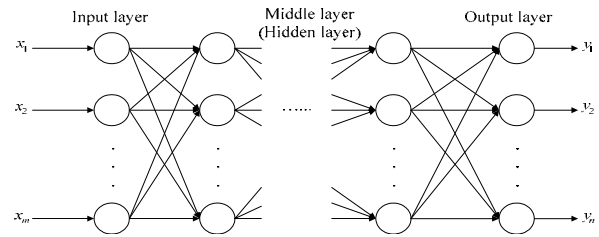


Fig. 1: BP network with hidden layers

$$net_i^p = \sum_{j=1}^M w_{ij} o_j^p - \theta_i = \sum_{j=1}^M w_{ij} x_j^p - \theta_i \quad (i = 1, 2, \dots, q) \quad (1)$$

where x_j^p and o_j^p denote the input and output of the j -th node, respectively. w_{ij} is the connection weight between the j -th neuron in input layer and the i -th neuron in hidden layer. θ_i is the threshold of the i -th neuron in hidden layer. M is the node number of input layer.

The output of the i -th neuron in hidden layer is:

$$o_j^p = g(net_i^p) \quad (i = 1, 2, \dots, q) \quad (2)$$

where $g(\cdot)$ is activation function.

The total input of the k -th neuron in output layer is:

$$net_k^p = \sum_{i=1}^q w_{ki} o_i^p - \theta_k \quad (k = 1, 2, \dots, L) \quad (3)$$

where,

w_{ki} = The connection weight between the i -th neuron in hidden layer and the k -th neuron in output layer

θ_k = The threshold of the k -th neuron in output layer

q = The node number of hidden layer

The actual output of the k -th neuron in output layer is:

$$o_k^p = g(net_k^p) \quad (k = 1, 2, \dots, L) \quad (4)$$

Parameters selection of BPNN: BP neural network is made up of multi-layer nodes. Some parameters, i.e., the node number of input layer, the layer number and the node number of hidden layer, the node number of output layer and so on are need to be selected in advance (Zhang *et al.*, 1998).

- **Selection of layer number and the node number of hidden layer:** Horni *et al* have proved that if linear transfer function (such as purelin function) is used between the input layer and the output layer and the Sigmoid transfer function is used in the hidden layer, then the Multi-Layer Perceptron (MLP) network including just one hidden layer can approximate any rational function with arbitrary precision. A three-layer BP network can reduce the

error by increasing the nodes number in hidden layer and it is easier to be trained than a BP network with more hidden layers, so a typical three-layer network is used in this study.

There is not a general and reliable method to determine the neuron number in hidden layer up to now. The neuron number in hidden layer is relevant to many factors, such as the node number in input and output layer, the complexity of the problem to be solved, the transfer function and the characteristics of the sample data. Existing methods of calculating the number of hidden nodes includes Kolmogorov theorem, one-way gradual change method and two-way method and so on. However, most of them are designed for arbitrary number of training samples and are aimed at the most adverse circumstances or samples containing noise. In fact, the hidden nodes number obtained using these formulas differs significantly sometimes (Wu *et al.*, 1998). Selection of the hidden nodes number is rather contradictory: on one hand, increase of the hidden nodes number can improve prediction accuracy; on the other hand, too much hidden node could result in excessive similar phenomena and reduce the generalization ability of the network

- **Selection of node number of input layer:** Because input nodes contain important information of time series structure, the number selection for them is very important. Many scholars have studied on this over the past decades. However, there is not a method that can be better than the others in all cases. One widely used method is AIC (Akaike Information Criterion) and it is still under extensive discussion.
- **Selection of node number of output layer:** The choice of output nodes is relatively simple. For time series prediction, the number of output nodes is usually associated with the forecasting area. There are two main forecasting ways at present (Taieb *et al.*, 2012): rolling prediction (only one output node is used) and MIMO(multiple input multiple output) prediction. These two methods have been already applied to multi-step prediction in some literature. For the rolling prediction, the prediction result is taken as an input of the prediction model for the next step prediction. There are many output nodes in the MIMO prediction method and each step prediction can be directly achieved in the model with only one time calculation. It has been pointed out that the rolling prediction method is obviously better than the MIMO prediction method when used in the sunspot forecasting (Weigend *et al.*, 1992).

COMBINATION FORECASTING MODEL

Principle of combination forecasting: Combination forecasting model combines different individual

prediction methods together in consideration of the characteristic of each individual prediction method. For example, suppose that there is a prediction method with large error but containing independent information of the system, if it is combined with another prediction method with relatively smaller error, then the prediction performance of the system can also be assured (Chen, 2008).

Let $\hat{x}_t = l_1x_{1t} + l_2x_{2t} + \dots + l_mx_{mt}$ be the combinational predictive value of x_t and l_1, l_2, \dots, l_m are weighting coefficients and they are constrained by Eq.(5):

$$\sum_{i=1}^m l_i = 1 \quad l_1, l_2, \dots, l_m \geq 0 \quad (5)$$

If m_i ($i = 1, 2, \dots, m$) are denoted as prediction index of all individual prediction models and m_c is denoted as prediction index of the combined model, then we have the minimal and maximal prediction indexes as follows:

$$m_{\min} = \min\{m_i, i = 1, 2, \dots, m\}$$

$$m_{\max} = \max\{m_i, i = 1, 2, \dots, m\}$$

It is clear that smaller m_c means better combination forecasting method.

Definition 1: If $m_{\min} \leq m_c \leq m_{\max}$, the model is called non-inferior combination forecasting. If $m_c < m_{\min}$, the model is called optimum combination forecasting. If $m_c > m_{\max}$, the model is called inferior combination forecasting.

Definition 2: If m_c cannot be decreased by adding a single prediction model to the combination model, then the single prediction model is called redundant prediction model. That is to say, the optimal weight of the single prediction method is zero, which indicates that it only provides redundant information.

Inference 1: Simple average combination forecasting method is at least a non-inferior combination forecasting.

Linear combination forecasting model with Sum Of Squared Error (SSE): It is well known that SSE is one of the most important indexes to reflect the prediction accuracy. For the combination forecasting model in which the weight coefficients are limited by nonnegative constraints, its optimal solution possesses concrete mathematical expression, thus it can be calculated by using the formula directly. This kind of combination forecasting model has been widely applied in practical prediction fields currently. Let e_t be the forecasting error at time t , then we have:

$$e_t = x_t - \hat{x}_t = \sum_{i=1}^m l_i e_{it}$$

where $e_{it} = x_{it} - \hat{x}_{it}$ is the forecasting error of the i -th forecasting method at time t .

Then the combination forecasting model with minimum SSE can be built by Eq. (6):

$$\min J_1 = L^T E L, \text{ s.t. } R_n^T L = 1 \text{ and } L \geq 0 \quad (6)$$

where, $L = (l_1, l_2, \dots, l_m)^T$, $R_n = (1, 1, \dots, 1)_{n \times 1}^T$, $e_i = (e_{i1} e_{i2}, \dots, e_{iN})^T$, $E_{ij} = e_i^T e_j = \sum_{t=1}^N e_{it} e_{jt}$, $i, j = 1, 2, \dots, m$, $E = (E_{ij})_{m \times m}$

Notice that there are nonnegative constraints on L , thus the original problem is a nonlinear programming problem.

If we ignore the nonnegative constraints, then the problem become a linear programming problem:

$$\min J_1 = L^T E L, \text{ s.t. } R_n^T L = 1 \quad (7)$$

The optimal solution of Eq. (7) can be obtained by using the Lagrange method:

$$L = \frac{E^{-1} R_n}{R_n^T E^{-1} R_n} \quad (8)$$

COMBINATION FORECASTING STEP WITH BP NETWORK

Prediction procedure: The combination forecasting with BP network is used for multi-step wind speed prediction in this study. Purelin function is used between the input layer and the output layer. The Sigmoid transfer function is used in the hidden layer. Compared with other training methods, LM (Levenberg-Marquardt) training algorithm has fast convergence speed and high training accuracy, so it is used to train BP network. The specific step are shown in Fig. 2.

Evaluation Indexes: Prediction accuracy is closely related to the prediction error. In order to reflect the effect of the combination forecasting, the RMSE (root mean square error) and MAE (mean absolute error) are used as evaluation indexes:

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (x_t - \hat{x}_t)^2} \quad (9)$$

$$MAE = \frac{1}{N} \sum_{t=1}^N |x_t - \hat{x}_t| \quad (10)$$

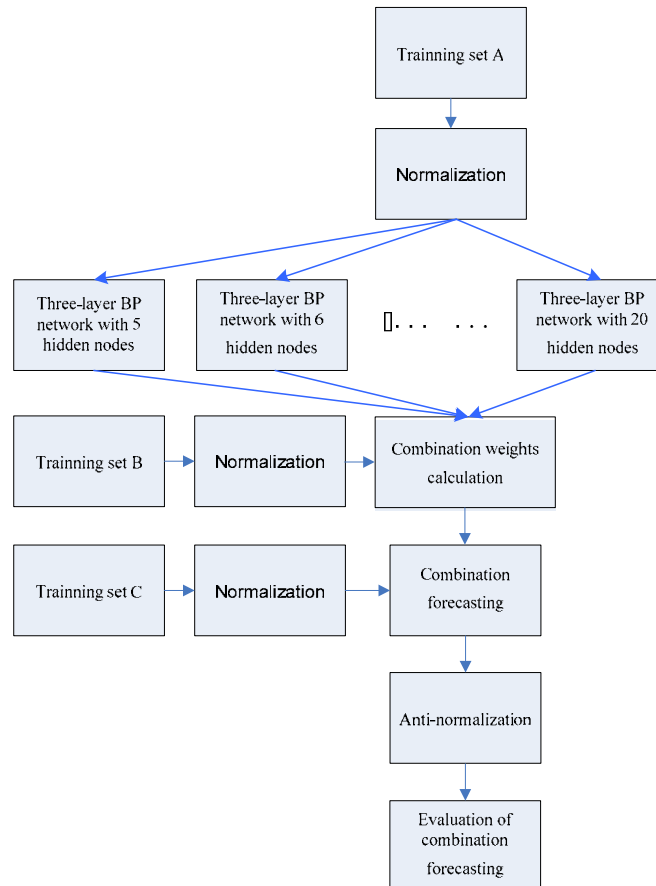


Fig. 2: Flowchart of wind speed forecasting

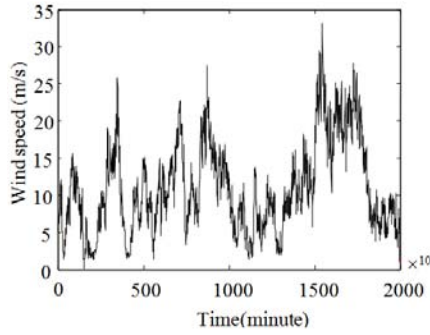


Fig. 3: Wind speed series with a sample interval of 10 min

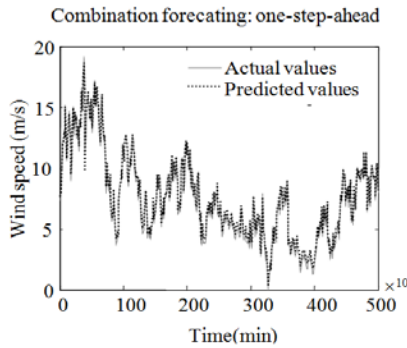


Fig. 4: One-step (10 minutes)-ahead prediction of combination forecasting method

EXAMPLE

The experiment data is collected from a New Zealand wind power plant from the year of 2002 to 2003 (sampling time is 10 min) and the first 2000 of the data is demonstrated in Fig. 3. The input dimension is determined as five by using the method of phase space reconstruction. The number of hidden nodes is selected from 5 to 20. The prediction step is from one to six, respectively. The initial weights are chosen as simple average combination weights according to Inference 1

Table 1 demonstrates the prediction results of five different models with prediction step from one to step six. The five models respectively are:

- **Persistence method. In this model:** The measured value of last step is taken as the prediction value at the present step.
- **Poorest BPNN:** The model has the poorest prediction effect among BP networks with different number of hidden nodes (from 5-20).
- **Best BPNN:** The model has the best prediction effect among BPNNs with different number of hidden nodes (from 5-20).
- **Combination model with the constraint of non-negative weights:** It refers to the linear combination of prediction model with the minimum RMSE under the constraint of non-negative weights.

Table 1: Prediction results of different forecasting methods from step 1 to step 6

Step	Method	RMSE	MAE
1	Persistence method	0.9278	0.9784
	Poorest BP	0.9223	0.9673
	Best BP	0.8656	0.8854
	Nonnegative weights combination	0.7486	0.7968
	General combination	0.7185	0.7328
2	Persistence method	1.2413	1.3241
	Poorest BP	1.2561	1.2873
	Best BP	1.1365	1.1782
	Nonnegative weights combination	1.0465	1.0629
	General combination	0.9433	0.9625
3	Persistence method	1.5085	1.5187
	Poorest BP	1.4651	1.4764
	Best BP	1.2803	1.3429
	Nonnegative weights combination	1.1758	1.2011
	General combination	0.9675	1.0251
4	Persistence method	1.5568	1.6701
	Poorest BP	1.6285	1.6582
	Best BP	1.5422	1.5981
	Nonnegative weights combination	1.2995	1.3486
	General combination	1.1602	1.1680
5	Persistence method	1.7202	1.7685
	Poorest BP	1.7179	1.7651
	Best BP	1.6746	1.6912
	Nonnegative weights combination	1.4826	1.5031
	General combination	1.3157	1.3559
6	Persistence method	1.7848	1.8931
	Poorest BP	1.7962	1.8971
	Best BP	1.7995	1.8139
	Nonnegative weights combination	1.4853	1.5937
	General combination	1.4203	1.4729

- **Combination model without the constraint of non-negative weights:** It refers to the linear combination of prediction model with the minimum RMSE but without the constraint of non-negative weights.

Figure 4 to 9 show the comparison of the combination forecasting model output and the actual value with prediction step from 1 to step 6, respectively.

In the process of multi-step prediction, the persistence model is the worst one and the combination forecasting model is the best one. Although the actual meaning of negative weights is still under discussion in academia, it provides the best predictive results in this example, while the combination forecasting model with nonnegative weights provides the next best predictive results in this example. Experimental results show that, in the process of multistep wind speed forecasting, the combination forecasting model effectively improves the reliability and accuracy of the prediction results. BP networks with zero weight in the combination model indicate that those models just provide redundant information and do not contribute to the improvement of accuracy and reliability in the process of combination forecasting.

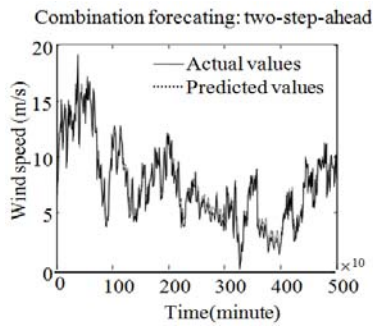


Fig. 5: Two-step (20 minutes)-ahead prediction of combination forecasting method

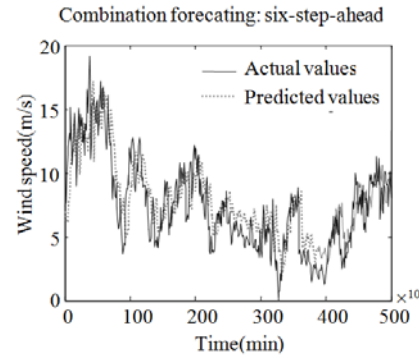


Fig. 9: Six-step (60 minutes)-ahead prediction of combination forecasting method

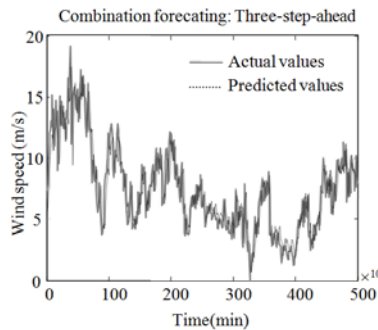


Fig. 6: Three-step (30 minutes)-ahead prediction of combination forecasting method

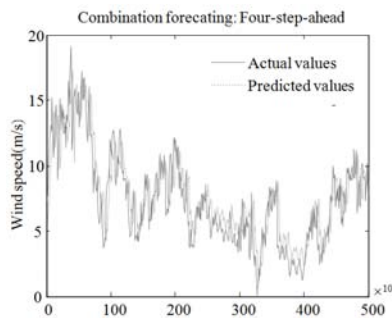


Fig. 7: Four-step (40 minutes)-ahead prediction of combination forecasting method

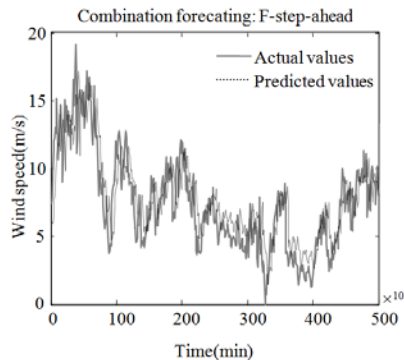


Fig. 8: Five-step (50 minutes)-ahead prediction of combination forecasting method

CONCLUSION

This study proposed a combination forecasting model using BPNNs with different number of hidden nodes, which is very helpful for wind power bidding strategy in short-term electricity market (Hu *et al.*, 2012; Varkani *et al.*, 2009). The main conclusions are as follows:

- The combination forecasting model can effectively avoid poor quality of BP neural network resulted from inappropriate selection of the number of hidden nodes.
- The combination forecasting model can improve the accuracy and reliability of multi-step ahead prediction.
- The method with nonlinear weights is needed to study further and it is also the main focus of our future work.

ACKNOWLEDGMENT

This study was supported by the National Natural Science Foundation of China (Grant No. 50677021) and the Fundamental Research Funds for the Central Universities (Grant No.11MG49)

REFERENCES

- Andrawisa, R.R., A.F. Atiyaa and H. El-Shishinyb, 2010. Combination of long term and short term forecasts: With application to tourism demand forecasting. *Int. J. Forecast.*, 27(3): 1-17.
- Chen, H., 2008. Validity of the Theory of Combination Fore-Casting Method and its Application. Science Press, Beijing.
- Hu, W.H., Z. Chen and B. Bak-Jensen, 2012. Stochastic optimal wind power bidding strategy in short-term electricity market. *Int. Rev. Elec. Eng.*, 7(1): 3380-3390.

- Li, G. and J. Shi, 2010. On comparing three artificial neural networks for wind speed forecasting. *Appl. Energy*, 87: 2313-2320.
- Li, G., J. Shi and J. Zhou, 2011. Bayesian adaptive combination of short-term wind speed forecasts from neural network models. *Renewab. Energy*, 36: 352-359.
- Pan, D., H. Liu and Y. Li, 2008. Optimization algorithm of short-term multi-step wind speed forecast. *Proc. CSEE*, 28(26): 87-91.
- Sfetsos, A., 2000. A comparison of various forecasting techniques applied to mean hourly wind speed time series. *Renewab. Energy*, 21(1): 23-35.
- Sharm, K.C. and B. Friedlander, 1984. Time-varying autoregressive modeling of a class nonstationary signals. *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 9: 227-230.
- Taieb, S.B., G. Botempi, A.F. Atiya and A. Sorjamaa, 2012. A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Exp. Syst. Appl.*, 39(8): 7067-7083.
- Varkani, A.K., H. Monsef and H.R. Baghaee, 2009. Strategy for participation of wind power in power market considering the uncertainty in production. *Int. Rev. Elec. Eng.*, 4(5): 1005-1014.
- Weigend, A.S., B.A. Huberman and D.E. Rumelhart, 1992. Predicting Sunspots and Exchange Rates with Connectionist Networks. In: Casdagli, M. and S. Eubank (Eds.), *Nonlinear Modeling and Forecasting*. Addison-Wesley, Redwood City, CA, pp: 395-432.
- Wu, C., L. Liu and B. Wang, 1998. The study of the method to determining the number of hidden units of three-layer BP neural networks. *J. Wuhan Tech. Univ., Surveying Mapp.*, 24(2): 177-179.
- Zhang, G., B.E. Patuwo and M.Y. Hu, 1998. Forecasting with artificial neural networks: The state of the art. *Int. J. Forecast.*, 14(1): 35-62.