Journal of Advances in Computer Research
Quarterly pISSN: 2345-606x eISSN: 2345-6078
Sari Branch, Islamic Azad University, Sari, I.R.Iran
(Vol. 5, No. 3, August 2014), Pages: 13-34
www.jacr.iausari.ac.ir

Improving Short-Term Wind Power Prediction with Neural Network and ICA Algorithm and Input Feature Selection

Elham Imaie[1]✉, Abdolreza Sheikholeslami[1,2], Roya Ahmadi Ahangar[2]

(1) Mazandaran University of Sciences Technology, Babol, Iran
(2) Noushirvani University of Technology, Babol, Iran

elham.imaie@yahoo.com ; a_sheikh@ustmb.ac.ir ; r.ahmadi@ieee.org

Received: 2013/12/26; Accepted: 2014/02/03

Abstract

According to this fact that wind is now a part of global energy portfolio and due to the unreliable and discontinuous production of wind energy; prediction of wind power value is proposed as a main necessity. In recent years, various methods have been proposed for wind power prediction. In this paper the prediction structure involves feature selection and use of Artificial Neural Network (ANN). In this paper, feature selection tool is applied in filtering of inappropriate and irrelevant inputs of neural network and is performed on the biases of mutual information. After determining appropriate inputs, the wind power value for the next 24-hours is predicted using neural network in which BP algorithm and PSO and ICA evolutionary algorithms are used as training algorithm. With investigation and compare numerical results, better performance of PSO and ICA evolutionary algorithm is deduced with respect to BP algorithm. More accurate survey will result in more proper efficiency of imperialist competitive algorithm (ICA) in comparison to swarm particle algorithm. Thus, in this paper; accuracy of the wind power prediction for the next 24-hours is improved considerably using mutual information and providing an irrelevancy filter for reducing the input dimension by eliminating the irrelevant candidates and more effectively using Imperialist competitive evolutionary algorithm for training the neural network.

Keywords: *Neural Networks, Wind Power Prediction, PSO Algorithm, ICA Algorithm, Feature Selection, Mutual Information*

1. Introduction

Nowadays, there is an obvious perspective of applying wind station in various countries according to the fast improvement in the wind turbine design technology, existence of rich wind energy resources at the most area and the tendency of using nonpolluting energy sources. Thus, investing at wind energy production is necessary and important and is absolute returning of the fund. Also, using estimation tools is very important [1]. Uncertain value of wind in a wind farm causes the produced power and energy of wind parts to have irregular and unpredictable fluctuations. Thus, in wind stations; prediction of wind speed, wind direction, amount of produced wind section power and … gets the most importance. Thereby, in spite of many advantages of wind energy, its variety and variable value at each moment cause it to be a non-dispatching source which can threaten the reliability of system. So, according to the complexity of prediction which has been investigated in [2] and [3], accurate prediction of wind power is gained a high importance. Nowadays, intelligent algorithms have been found to be promising techniques in forecasting of parameters. According to the good performance of neural networks in pattern recognition, they can be applied in prediction of wind speed and wind station power. In [4] ,wind power and speed are estimated using artificial neural networks. Due to the complexity of atmospheric processes, time series model of wind is more preferred among other physical models proposed for wind prediction. Also, two common types of artificial neural networks namely feed-forward network and Elman Network are used for predicting the wind speed and power. Performances of these models are compared based on the effective wind speed and power error. In [5], ARMA series model is used for wind speed and atmospheric pressure prediction using the weather information of wind farm. Also, Radial Basis Function (RBF) neural network method is used for wind power prediction. In [6], wind power prediction problem is investigated in Portugal country using neural network and the Mean Absolute percent Error (MAPE) criteria is applied for analyzing the accuracy of the prediction which is evaluated to 7.26% and denotes good performance of the proposed neural network. In [7], multi-layer neural network is used for 1-hour and 24-hours wind power prediction, meanwhile temperature and wind speed are chosen as the net required inputs. In [8], artificial neural network is applied following with Markov chain for short-term forecasting of wind speed so that the artificial neural network forecasts short-term value of the wind speed and the results are modified according to the long-term pattern achieved from Markov chain. In [9], two methods are introduced for wind power prediction based on RBF neural network and fuzzy logic technique in which one method is based on time series information of the wind power and the other is based on Numerical Weather Prediction (NWP) information. It can be seen through investigation of the numerical results and comparison with those obtained using Persistent method that the two proposed method have good performance. In [10], Multi-Layer Perceptron (MLP) neural network is applied for wind speed and power prediction. It can be seen by analyzing the mean value of the square error that the method has better performance in comparison with ARMA and Persistent method. In [11], a two hidden-layer neural network is gained for forecasting the output wind power. The data are gathered from a wind farm in south of Italy. The power value is estimated through using the BP training algorithm in neural network in which the tangent hyperbolic transfer function (tanh) is used in the first hidden layer and Sigmoid transfer function in the second hidden layer. In [12], adaptive wavelet is used by feed-forward neural network for forecasting the wind power. The forecasting method involves two procedures. First, wavelet decomposition is performed then adaptive wavelet neural network is gained to turn each decomposed signal back and estimate the wind speed for about 30 hours. In the second procedure, feed-forward neural network is applied for nonlinear mapping between the output wind speed and power to estimate the wind power according to the predicted wind speed. The effectiveness of this method in comparison with Persistent and ARMA model denotes the acceptable performance of this model. However, the problem involved with this statistic method is that the value of the error increases fast as the estimation time increases. In classical (or common or conventional) methods of training neural networks, BP algorithm or other gradient based algorithms are used for training neural networks. However, in cases where the function is complex and nonlinear, the performance is not good. Thus, using evolutionary algorithms can improve the performance of neural networks. For example, in [13] artificial neural network is used following with Adaptive Bayesian learning and Gaussian approximation for predicting the produced wind power. Among advantages of Bayesian method; controlling the network complexity (very complex networks have automatically errors), investigating the uncertainty in prediction and its flexibility can be inferred. In [14], short-term power estimation of wind turbine is performed based on neural network which is trained by Genetic algorithm and the technique is shown to have an acceptable accuracy. In [15], complex-valued recurrent neural network (CVNN) and real valued neural network (RVNN) are used for power value prediction and [16] proposed an enhanced particle swarm optimization (EPSO) based hybrid forecasting for short-term wind power forecasting that combined the Persistence method, the back propagation neural network, and the radial basis function(RBF) neural network, and compared them with those individual forecasting methods.

The wind power is proportional with air density, the area that winds are blowing from and cubic function of wind speed. The air density is an exponential function of the air pressure and temperature [2], thus, the air power is affected by the weather condition and the parameters such as wind speed, wind direction and temperature. Moreover, the wind power has a close relation with its correspondent past values at each time interval [17]. Thereby, due to the large dimension of neural networks inputs and preventing the Curse dimensionality, feature selection problem is proposed. The feature selection problem refers to identifying an effective subset of feature among initial information that leads to the reduction in the input dimension and making the calculations easier following with more accuracy in the neural network. In this paper, wind power prediction is performed using neural network and evolutionary algorithms imperialist competitive algorithm (ICA) and particle swarm optimization (PSO) for training the network and applying an effective technique for feature selection of neural network inputs which is based on mutual information (MI). In MI technique, more effective information is selected for neural network. In this technique, each input set can be processed by filtering the irrelevant properties, then filtered input set are used as neural network input [18]. Thus, time series information and numerical weather prediction (NWP) is used for predicting the power and the investigated information are correspondent to the one year period of recorded wind in a station at Semnan-Iran. The advantage of this method with respect to other previous prediction methods that used neural network is at applying an irrelevant filter in order to select effective input following with an evolutionary algorithm for training the neural network. Combination of these two techniques significantly improves the prediction speed and the accuracy of the result. Also, this configuration is effective at reducing the prediction error

In the second part of this paper, a brief description of Multi-Layer Perceptron(MLP) and back propagation training algorithms, PSO and ICA algorithms and Mutual Information.

2. Control Structure

2.1 Multi-Layer Perceptron

Multi-Layer Perceptron neural network are the most common type of neural networks. In this network, there is one or more number of hidden layers where each layer involves one or more neuron. Neurons of each layer are attached to those in the next layer and the output of a neuron at one layer is only related to the neurons in the previous layer and the weights between them. The most important part of determining optimal structure of Multi-Layer Perceptron network is finding the number of hidden layer and the number of neurons at each hidden layer to achieve minimum error [19]. In Figure1, simple schematic view of the structure of a Multi-Layer Perceptron neural network is depicted in which input, hidden and output layers are illustrated. Also, activation functions are used to transfer the output of each layer to the next layer. There are various types of activation functions where the most important ones are linear function, sigmoid function, hyperbolic tangent function and … [20].

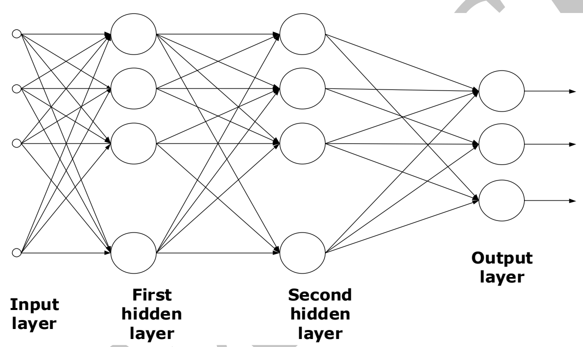


Figure1: Multi-Layer Perceptron artificial neural network

2.2 Back Propagation Algorithm

This algorithm which is proposed by David Rumelharthttp://www.google.com/search?biw=1280&bih=654&q=Rumelhart+and+Mack+Kaland%2Bneural+network&spell=1&sa=X&ei=gZ6xUtqfO-7T7Aazk4B4&ved=0CCYQvwUoAA and James McClelland in 1986 is used in feed-forward neural network. The term feed-forward refers to this fact that the artificial neurons are placed in successive layers and the output signal is sent as input. Also, the term back propagation refers to the fact that the errors are fed back to the network in order to modify the weights thereafter repeating its feed-forward input again. Back propagation method can be regarded as learning with teacher (or supervised learning method), namely the input data (or pattern) are labeled and their correspondent output is prescribed. Thus, the network output is compared with these ideal outputs and the network error is calculated. In this algorithm, first it is assumed that the network weights are chosen randomly. At each step, the network output is evaluated and according to the difference with the desired output, the weights are modified until the error is finally minimized [19] .

2.3 Particle Swarm Algorithm

PSO algorithm is a social search algorithm modeled from social behavior of birds flock. First, the algorithm was applied to find the patterns prevailing on the simultaneous flying of birds and their sudden changes of path and optimal transformation of the swarm. In PSO algorithm, particles are flown throughout the search space. The particles displacements in the search space are affected by the knowledge and experiences of themselves and their neighbors. Thus, the situation of other particles affects the search quality of a particle. The result of modeling such social behavior is providing a search process in which particles converge toward successful area. Particles learn from each other and moves toward their best neighbors based on achieved information. PSO is based on this fact that at each moment, each particle tunes its position in the search space according to the best position that has experienced by itself (pbset) and the best position exist in its whole neighborhood (gbest). After finding the low best value, the particle updates is velocity and positions with formulas (1) and (2):

$$V_i^{k+1} = \omega V_i^k + c_1 rand(...)(Pbest_i - X_i^k) + c_2 rand(...)(\mathrm{gbest}_i - X_i^k) \quad (1)$$

$$X_i^{k+1} = X_i^k + V_i^{k+1} \quad (2)$$

Where ω is the inertia weight, V_i^k is the particle velocity, X_i^k is the current particle, rand() is a random number between (0,1) and c_1,c_2 are learning factors[21], [22]. For training neural networks, networks weights and biases are considered as optimization variables.

2.4 Imperialist Competitive Algorithm

Imperialist competitive algorithm is an optimization method inspired from political competitions. Like other evolutionary algorithms, this algorithm starts with an initial population. Each member of this population is called a country. Some of the best countries (countries with the least cost function value) are determined as imperialist and the other as colonies. Each empire power is inversely proportional to its cost function value. After dividing all colonies between imperialists and forming an initial empire, the colony starts moving toward the empires. This motion denotes assimilation policy. Figure2 illustrates the motion of a colony toward an imperialist.

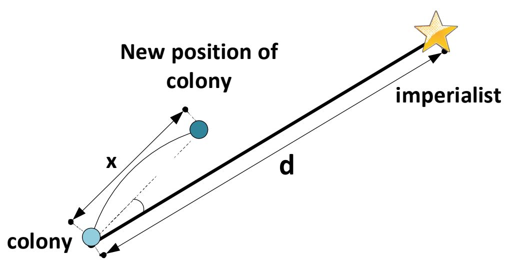


Figure2: Motion of colonies toward their relevant imperialist

In this motion θ and x are arbitrary number with uniform distribution and d denotes the distance between the imperialist and colony.

$$x \sim U(0, \beta \times d) \quad (3)$$

$$\theta \sim U(-\gamma, \gamma) \quad (4)$$

β and γ are random or arbitrary numbers that determine the search space around the imperialist. Coefficient β causes the colony to get closer to the imperialist from different directions during the motion. Imperialist competition between empires is as increasing and decreasing of the imperial power. The total power of an empire depends on both the power of imperialists country and the power of its colonies which is shown in formula(5).

$$T.C._n = Cost(imperialist_n) + \xi \,\mathrm{mean}\left\{Cost(colonoes \quad of \quad impire_n)\right\} \tag{5}$$

The empire will be removed if it cannot improve its power or its power is reduced. The procedure of removing weak empire and increasing the power of powerful empire continues until only one empire is remained. In this situation, other countries are colony of this empire and the algorithm is ended [23-25].

2.5 Mutual Information

In this section for reducing the calculation and according to this fact that neural network will encounter the problem of curse dimensionality incase using large computational calculations, MI method is applied for feature selection. In this method first the MI value between the candidates input Xm and the goal variable Y must be calculated. The method for calculating the MI value is presented in [26] in detail.
Candidate inputs can be classified according to their MI value by calculating mutual information between candidate input Xm and goal variable Y (such that the larger is the value of MI, the more relation there is between the input and the goal variable), and by using try and error method best number of neural network input is determined that involves less error. The MI value which is the amount of relation between the variables X and Y can be evaluated according to [18] and [26] and by calculating the entropy value (H) that in fact denotes the uncertainty value.

If the MI value between two variables is large the variables have close relation with each other and if the value of MI is zero it denotes that the two variables are completely independent of each other and they have no relation.The feature selection procedure must be repeated regularly at each time interval (1 hour at this paper) in order to remove irrelevant inputs. In [18] a simpler method is proposed for calculating MI that makes the calculations easier and less.

3. System under study

For making an artificial neural network, sufficient information must be provided. This information must be large enough in order to prevent the network to fall in trouble for the selection process and also must be accurate enough and validate. In this paper wind stations information is used for making the artificial neural network. This station has the capacity of about 400 MW and involves 200 turbines with the capacity of 2MW. The correspondent information is large enough and contains the information of one successive year from 2007 to 2008 with one hour time interval. The information used in this paper involves: wind speed, wind direction, humidity, air temperature and the wind power. 60% of this information is used for training the neural network and the remaining 40% information is used for the test and verification of the network. It must be noted that the information used for training and test must be different. The information profile that is used for prediction is presented in Table 1:

Table 1: General characteristics of the studied plant information

	Wind Speed (m/s)	Wind Direction(º)	Temperature (ºC)	Humidity (%)
Mean	10.29	184.23	19.3	34.44
Max	32.91	355.83	45.9	100
Min	0.166	0	-14.2	4

4. The proposed method

In this paper multi-layer neural network is used for the wind power prediction for the next 24-hours. In order to select appropriate inputs for network, mutual information method is used to prevent the curse dimensionality and providing better and faster performance of the neural network. Also, evolutionary algorithm is used for training the neural network. Using both these methods simultaneously in neural network leads to a very good prediction performance. The wind power is a nonlinear function of some weather variables and their previous values. These variables involve wind speed, wind direction, air temperature and humidity that can be evaluated in the wind farm or correspondent weather station. In the system under study in this paper, the value of the wind speed, wind direction, humidity, air temperature and the wind power in a one year time interval is available and this information in the following set form is considered as input for forecasting the power for the next one year.

$$I(t) = \begin{Bmatrix} p(t-1),\ldots,p(t-N_p), \\ \mathrm{S}(\mathrm{t}-1),\ldots,\mathrm{S}(\mathrm{t}-\mathrm{N}_s), D(t), \\ D(t-1),\ldots,D(t-N_D),T(t) \\ ,T(t-1),\ldots,T(t-N_T), \\ H(t),H(t-1),\ldots,H(t-N_H) \end{Bmatrix} \tag{6}$$

I(t) is the candidate as the input of neural network. Where P(t) ,S(t), D(t),T(t) ,H(t) denote wind power, wind speed, wind direction, air temperature and humidity in time t respectively. In this paper, one year information of a station in Semnan-Iran is used with a one hour time interval and the goal is predicting the wind power for the next 24-hours with one hour time interval. The value of P(t-1), …, P(t- NP) is correspondent to the power value from the past one hour to past NP hour. and S(t-1), … , S(t-NS) is correspondent to the wind speed from the past one hour to the past NS hours and the same is repeated in I(t) for wind direction, air temperature and humidity. And the goal is predicting the power value for the next one hour, P(t); using this information. It was noted that N refers to the back shift of the desired candidates. The value of the back shift must be chosen large enough in order to prevent losing effective information. However, an agreement must be met for preventing the creation of a high dimensional set which is usually achieved with try and error. For example, if the value of back shift is considered even for 24 hours before, number of candidate inputs of neural network is 123 that may lead to curse dimensionality due to the large dimension of neural network and as a result the prediction procedure may be troublesome. To prevent such problem a method is needed for feature selection in a way that a smaller set of I(t) is achieved that involves properties with more information and related to the objective function while filtering unrelated properties. In the proposed method, MI technique is used to satisfy the goal. Now neural network must be trained. The goal of training is determining the value of weights and biases such that the error of trained data is minimized, thus various training algorithms are used for training the neural network. In common (or conventional or classical) training methods, back propagation and other gradient based algorithms are used. These methods reveal their weakness and inefficiency when the function is nonlinear and complex. Thereby, using evolutionary algorithms in the training process of neural network improves the performance of prediction procedure. In this paper, two PSO and ICA evolutionary algorithms are applied for training the neural network in order to increase the accuracy of neural network in prediction. The structure of the proposed method is depicted in Figure3 for forecasting the wind power:

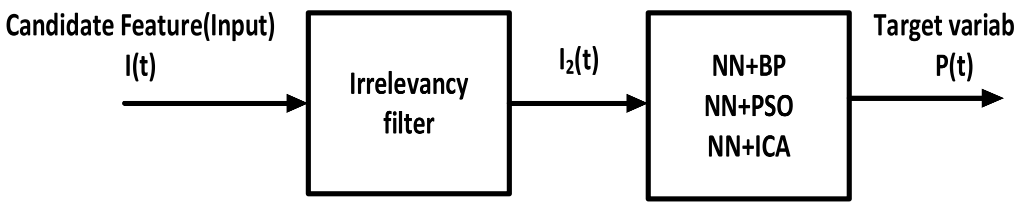


Figure3: Structure of proposed method

It can be seen from Figure3 that the proposed method involves two procedures. In the first procedure feature selection from the considered information by (or through) transferring (or passing) from irrelevancy filter is dealt with and in the second procedure wind power prediction for the next 1-hour in the neural network is performed using candidate inputs form the previous procedure by applying training algorithms BP, PSO and ICA that were described in previous sections. The prediction results are then analyzed according to the evaluation criteria that are defined in the next section. Flowcharts correspondent to training method of neural network using PSO and ICA algorithm after passing from their irrelevancy filter are depicted in Figure4 and Figure5. The fitness function in both algorithms is RMASE function according to formula(7):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (Y_r - Y_P)^2} \tag{7}$$

Where Y_r is the measured real value of power, Y_p is the predicted value of the power and N denotes the number of patterns for investigation[27].

In PSO algorithm, the inertia weight value selected is 0.9, the values of both c_1 and c_2 are taken as 2. The fitness function for each particle is obtained by updating the weights of ANN as specified by the variable of the particle and finding the root mean squared error obtained in ANN training as formula(7). Similarly the fitness function of all particles in the population are determined. The particle having lowest fitness function is the gbest particle and the fitness function of the gbest particle in compared with specified accuracy. If the required accuracy is obtained, then the training is stopped. Otherwise the velocity and new position of the particles are updated as formulas (1) and (2). The same process is repeated until the specified accuracy is reached. Similar to PSO algorithm for training neural network with ICA algorithm, we considered the weights of network training phase as the variable of an optimization problem. The root mean square error used as a cost function (fitness function) in ICA algorithm. The goal in the proposed algorithm is minimizing this cost function. The neural network weights initialized randomly and saved as initial population for the ICA algorithm to optimize the cost function. After training the neural network, mean absolute percentage error (MAPE) and regression parameter are also used for closer investigation of the predicted results. formula (8) is presented in the following for determining MAPE parameter:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left| \frac{Y_r - Y_P}{Y_r} \right| \tag{8}$$

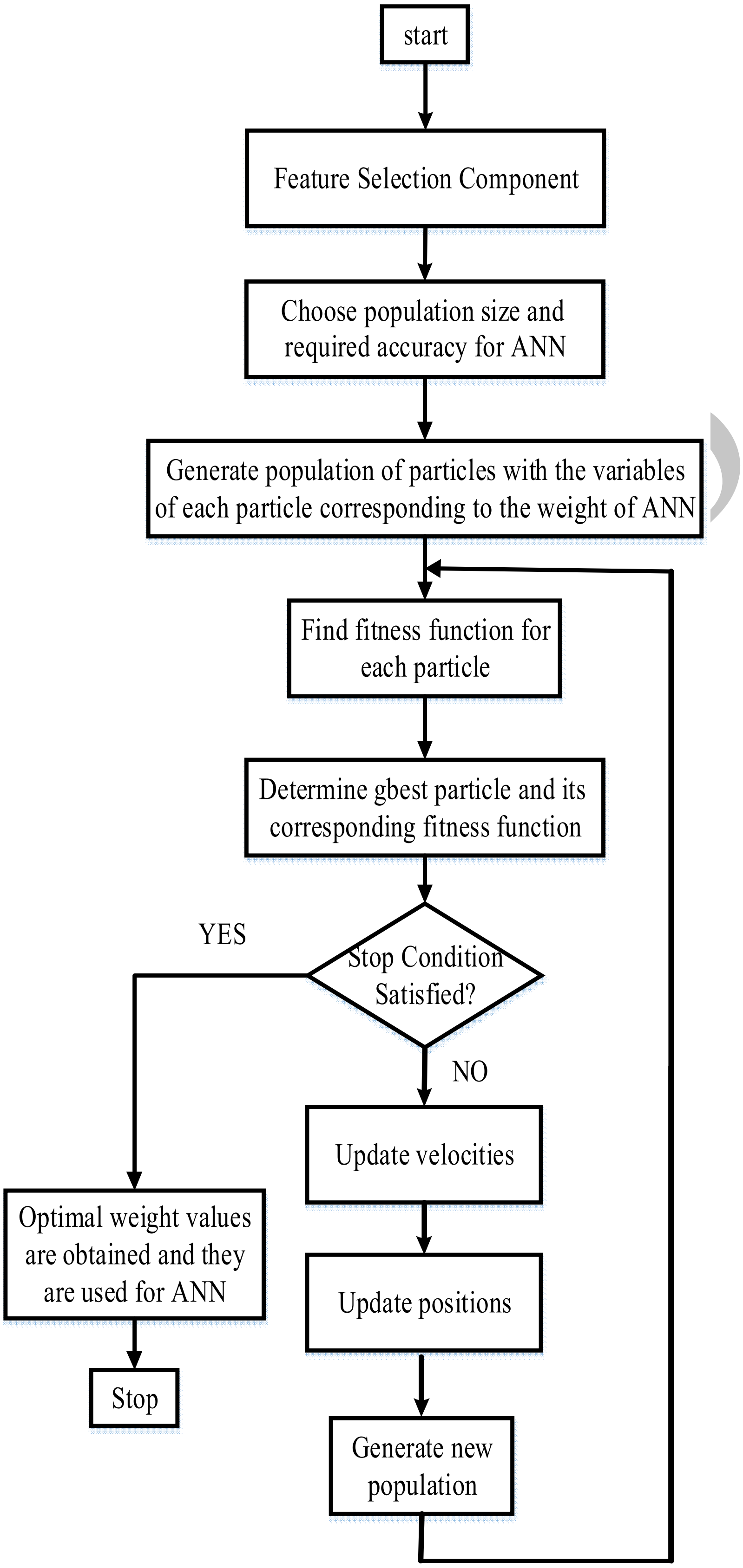


Figure4: Flowchart of proposed method with PSO as training algorithm

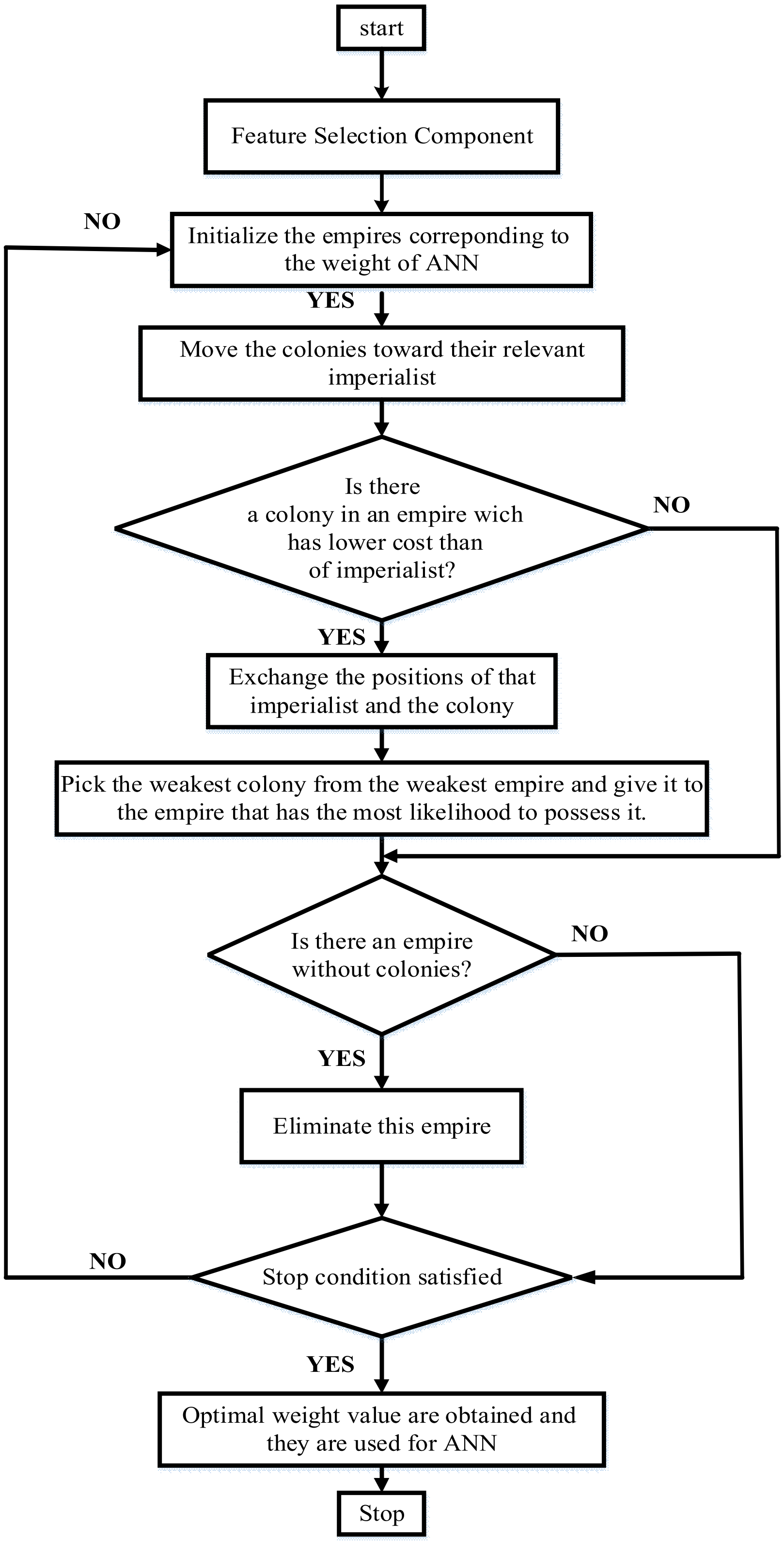


Figure5: Flowchart of proposed method with ICA as training algorithm

The neural network used in this paper has three layers in which the activation function in the first and second layer is tangent sigmoid and in the output layer is of linear type. Number of Neurons in the first and second layer is 10 and 5 respectively and it is chosen one for the outpour layer. Parameters specifications of PSO and ICA evolutionary algorithm as training algorithm for neural network are presented in Appendix.

5. The proposed method

For making a neural network, sufficient and appropriate information is required in order the network to be accurate enough. In this paper, information of a wind station in Semnan-Iran is used. This station contains 299 turbines with 2MW capacity and the measured data are for a one successive year of this station. These data involves wind speed, wind direction, air temperature, humidity and the wind power. The time interval between the data is one-hour and time series method is used for estimating the power value for the next 24 hours with one-hour time interval.

As was mentioned in formula (6), information involves wind speed, wind direction, air temperature and humidity that their present value and 48-hours before value are used in this paper, also the wind power value for the 48-hours before is used. Namely, 243 inputs at each step that makes the neural network to encounter with the problem of curse dimensionality and make the calculations impossible. Thus, the goal of introducing MI feature selection method is to choose more effective and efficient inputs as the inputs of neural network (I2 in figure3). Now, by choosing I2 as the input of neural network and the power value as the objective function, the power value for the next hours can be estimated.

5.1 Determining MI

In this section using the formulas for calculating MI that are presented in [18], MI value is determined for inputs and the inputs are classified according to this MI such that the larger is the value of MI, the higher is the dimension (or ranking). However, for determining selected candidate (NI) based on more effective inputs; iterative search procedure that is proposed in [18] or the try and error method can be used. In this paper, best number for chosen inputs is achieved based on MI value of 27 through using iterative search procedure. Now, appropriate inputs for neural network can be chosen by calculating the MI value and determining the value of NH and according to the classification of the information based on their MI value. For example the candidate inputs for the first and second day of train and their rankings based on the MI value are presented in Tables2 and 3.

Table 2: Selected feature for the first train data

Selected feature	Rank	Selected feature	Rank
P(t-1)	1	T(t-4)	15
P(t-3)	2	P(t-7)	16
P(t-6)	3	P(t-11)	17
P(t-15)	4	S(t-5)	18
S(t-1)	5	P(t-9)	19
P(t-2)	6	P(t-10)	20
S(t-6)	7	D(t)	21
D(t-1)	8	S(t-4)	22
P(t-18)	9	P(t-5)	23
S(t-2)	10	S(t-3)	24
H(t-1)	11	D(t-5)	25
T(t-1)	12	P(t-14)	26
P(t-16)	13	D(t-11)	27
S(t-9)	14		

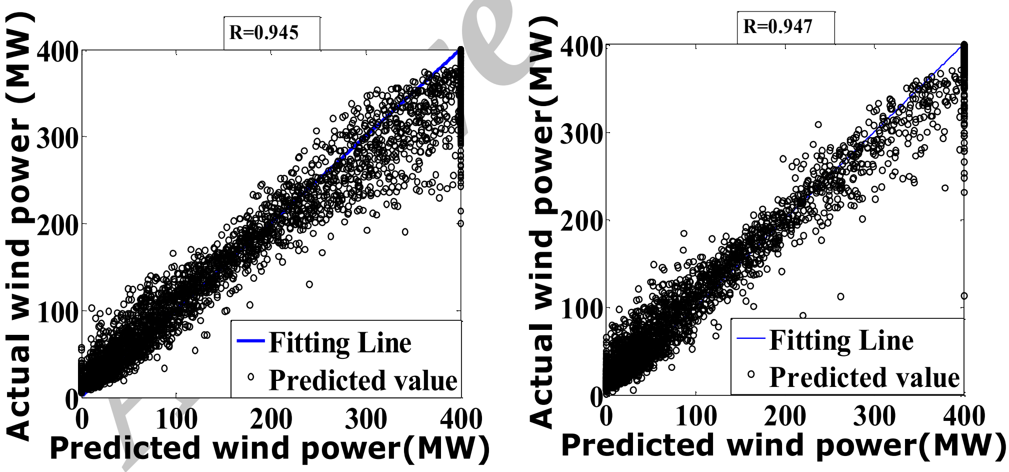
Table 3: Selected feature for the second train data

Selected feature	Rank	Selected feature	Rank
P(t-1)	1	S(t-8)	15
P(t-3)	2	P(t-23)	16
P(t-6)	3	T(t)	17
P(t-15)	4	D(t-3)	18
S(t-1)	5	S(t-9)	19
P(t-2)	6	S(t-5)	20
S(t-6)	7	P(t-8)	21
S(t-11)	8	P(t-9)	22
D(t)	9	D(t-2)	23
P(t-5)	10	S(t-4)	24
S(t-16)	11	P(t-14)	25
T(t-1)	12	P(t-13)	26
H(t-1)	13	D(t-1)	27
P(t-10)	14		

It can be seen from Table2 and 3 that by using MI method the number of input data to the neural network is significantly reduced and more relevant data are chosen thereby the computational calculation is reduced and curse dimensionality is prevented and the prediction is performed with best performance.

5.2 Power prediction using neural network and evolutionary algorithms

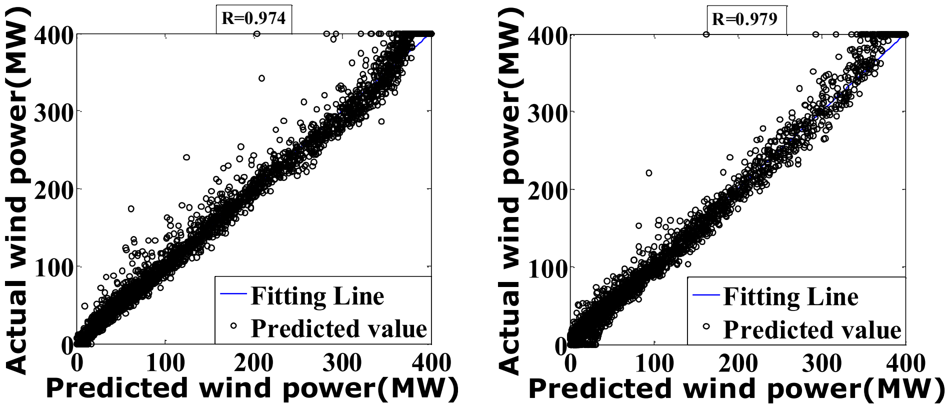
In this section using MI relation for feature selection, dimension of neural network input is reduced and then is ready to enter in to the neural network. The proposed neural network involves two hidden layers and one output layer and the goal is predicting the wind power for the next 24 hours with 1-hour time interval. Information is divided into two parts, the first part is used for training the neural network and the second part is used for testing the network after training and verification of the model. After the training and test procedure and evaluating the criteria that were mentioned in previous sections, if the neural network is accurate enough; the power estimation problem can be investigated. Also, it must be noted that the test and the training data must be different. Here, 60 % of the data is used for train and the rest is used for test. The goal is using the training algorithm in neural network for determining the best value of the weights and biases such that the prediction error is minimized. In this paper after feature selection, neural network is trained using BP, PSO and ICA algorithm. The prediction results are then investigated. One of the evaluation methods is calculating the linear regression. Regression is a measurement of the relationship between the net output and the real output data and a regression denotes an accurate relationship between neural network output and real output. Thus, as this value is closer to 1, prediction is more accurate. In Figures 6, 7 and 8, regression correspondent to the test and training information for all three algorithms of back propagation, PSO and ICA is depicted.



Training(ANN-BP) : R=0.945 **Test(ANN-BP) : R=0.947**

Figure6: Regression diagram correspondent to the test and train data using BP algorithm for wind power prediction

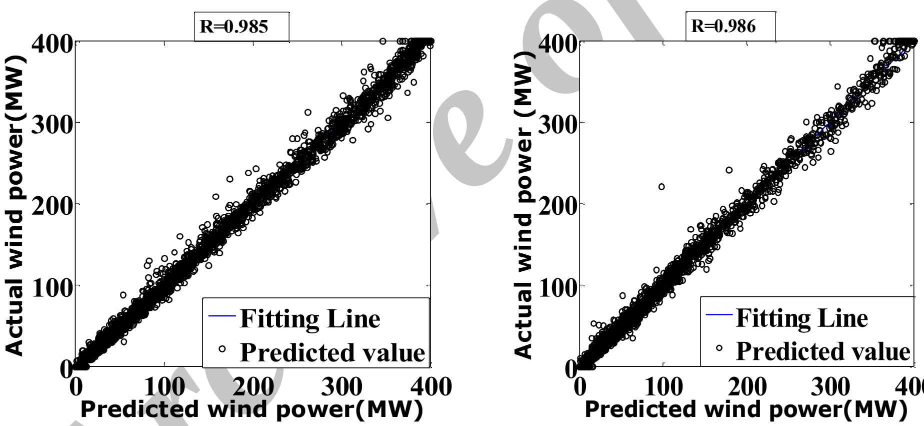
Figure6 is corresponding to the regression diagram of BP algorithm. It can be seen from the figure that the vertical axis denotes the real value of the wind power for training and test data and the horizontal axis denotes the predicted wind power value. In the above figure, according to distance between the data and fitting line, there is a great difference between the real and predicted value that can be reduced using evolutionary algorithms and the regression value can also be made closer to unity.



Training(ANN-PSO) : R=0.974 Test(ANN_PSO) : R=0.979

Figure7: regression diagram correspondent to the test and train data using PSO algorithm for wind power prediction

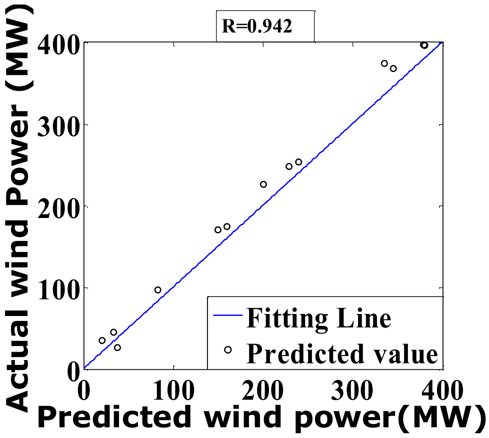
In Figure7 regression diagram using PSO as training algorithm is illustrated. According to the diagram and regression value that is increased 3% for the test and train situation it is obvious that in situation of using PSO as training algorithm the difference between the real and predicted value is decreased in comparison with the previous case and the prediction is more accurate.



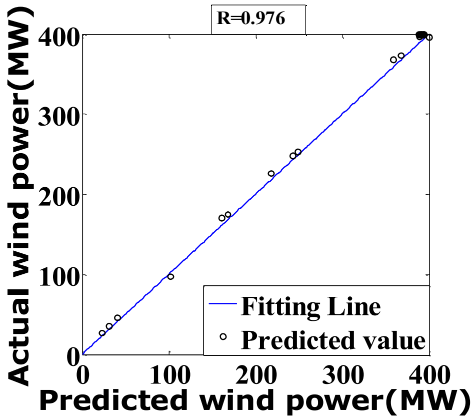
Training(ANN-ICA) : R=0.985 Test(ANN-ICA) : R=0.986

Figure8: regression diagram correspondent to the test and train data using ICA algorithm for wind power prediction

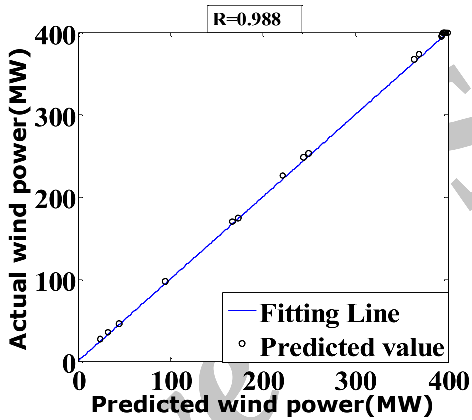
Regression diagram in Figure8 is correspondent to the proposed method of this paper. Namely the situation in which the ICA training algorithm is used for neural network. It can be seen that the ICA algorithm is more accurate in comparison with the two previous algorithms and the regression value is increased 1% and 4.3% over the situation that PSO and BP were used as training algorithm for neural network respectively. Also, in Figure9, regression diagram correspondent to the predicted value for the next 24-hours is depicted.



A: Validation(ANN-BP): R=0.942



B:Validation(ANN-PSO) : R=0.976



C: Validation(ANN-ICA): R=0.988

Figure9: regression diagram of the results achieved from network validation for the next 24-houres using A: BP as training algorithm, B: PSO as training algorithm, C: ICA as training algorithm

In these Figures the regression value of the predicted data for the next 24-hours in case of using BP, PSO and ICA as training algorithm are calculated 0.942, 0.976 and 0.988 respectively that shows the efficiency of the proposed method. In addition to comparison and investigation of predicted results using regression diagram, in Figure10, 11 and 12 the real value of the wind power is presented in the test, train and validation situation following with the predicted value using neural network with PSO and ICA and BP as training algorithm. It is obvious form the figures that evolutionary algorithms are more accurate than back propagation algorithm because of the weakness of back propagation and gradient based algorithms at solving complex nonlinear problems. In test and train diagram, only 50 first patterns are presented to make the comparison easier from the figure.

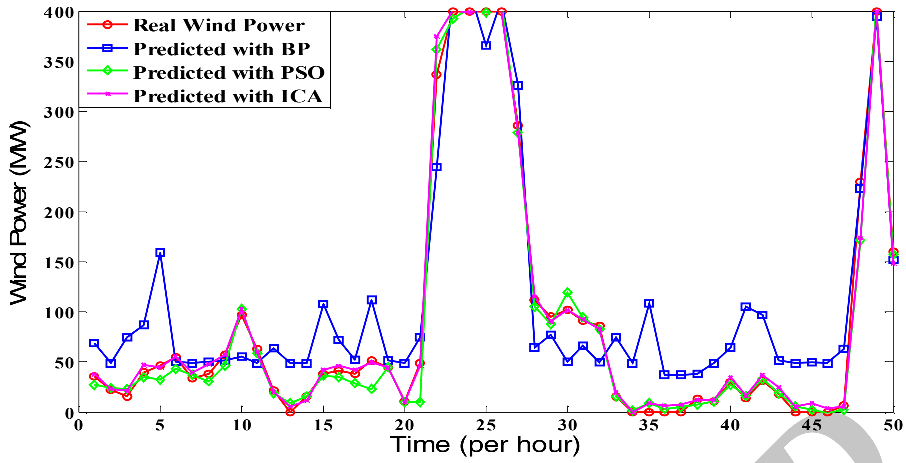


Figure10: comparison of real and predicted value of train data (50 first patterns) using back propagation, PSO and ICA algorithms

Figure10 is correspondent with the real and predicted value in train procedure. For easier investigation of the above figure, 50 first patterns are used. It can be seen that the predicted results have a considerable difference with their real value without using evolutionary algorithms. In case where PSO and ICA evolutionary algorithms are used as training algorithm for neural network the difference is very small and between these two methods, the predicted results using ICA method almost matches with their real value.

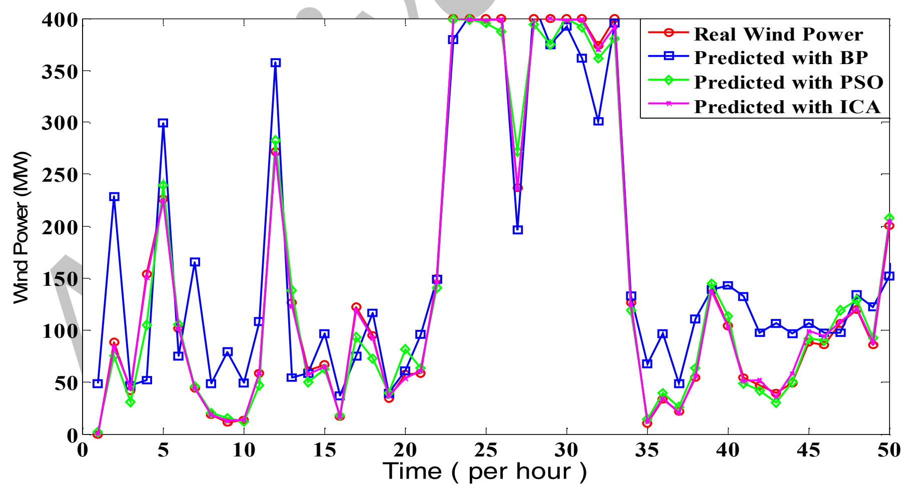


Figure11: comparison of real and predicted value of test data (50 first patterns) using back propagation, PSO and ICA algorithms

Figure 11 depicts the real and predicted value in the test procedure. In this section, for easier investigation only 50 first pattern is used and the same as Figure10, it can be seen that in the test procedure the difference between the real and predicted value while using PSO or ICA algorithm is less than the case where back propagation algorithm is used as training algorithm for neural network. Also, the predicted values using ICA as training algorithm for neural network almost matches the real value and the error is negligible.

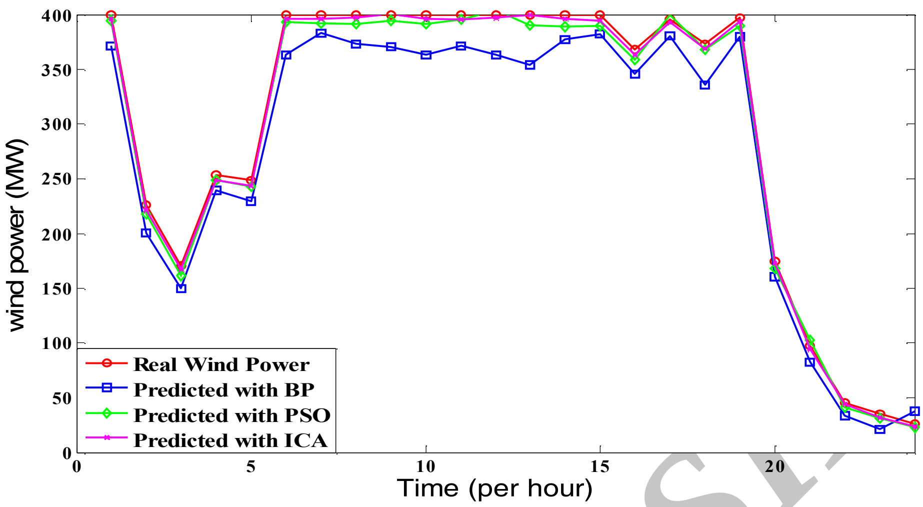


Figure12: comparison of real and predicted value of the power for the next 24-houres using neural network with back propagation, PSO and ICA algorithms as training algorithm

Figure12 depicts the predicted value of the power for the next 24-houres with 1-hour time interval. According to the test and train data achieved from the previous section it can be seen that the diagram correspondent to the predicted value of the wind power using ICA as training algorithm for neural network is very close to those from real value of the wind power and they roughly match with each other. It can be seen from the figure that this method has less error than the other two methods namely back propagation and PSO. For closer investigations, the error value can be determined from the RMSE and MAPE value defined in formulas (7) and (8). The Root mean absolute error (RMSE), mean absolute percentage error (MAPE) and regression for all three cases is presented in Table4.

Table4: Root mean absolute error (RMSE) , mean absolute percentage error (MAPE) and regression measurements of wind power in the construction of neural networks with proposed method

MI+ICA+NN			MI+PSO+NN			MI+BP+NN			
MAPE (%)	**R**	**RMSE (MW)**	**MAPE (%)**	**R**	**RMSE (MW)**	**MAPE (%)**	**R**	**RMSE (MW)**	
4.58	0.985	4.47	5.12	0.974	6.95	10.89	0.945	15.3	**Train**
5.01	0.986	4.45	5.79	0.979	6.98	11.34	0.947	15.37	**Test**
4.87	0.988	4.41	5.51	0.976	6.96	11.43	0.942	15.35	**Validation**

Same result as regression and comparison diagrams are achieved by closer investigation of Table4. According to the Table4, the minimum RMSE is relevant to the ICA, PSO and back propagation training algorithms respectively. The computed RMSE value using ICA training algorithm is reduced at about 35.49% over PSO algorithm and 70.8% over BP algorithm. Also, using PSO, the RMSE value is decreased to about 54.8% over back propagation algorithm. Thus, using two evolutionary algorithms for training the neural network is more effective. Also the computed MAPE value using ICA training algorithm is reduced about 11.88% Over PSO algorithm and 57.05% over BP algorithms. By comparing the MAPE results achieved from the proposed methods in this paper with those achieved from the methods proposed in other papers[12],[15] and [16] (Table5), performance of the proposed method in this paper with respect to the other methods proposed in this field before can be verified. In Table5, MAPE value is presented which is achieved from the prediction process of the proposed method in this paper and in the mentioned references.

Table5: comparison of MAPE results for wind power forecasting in this paper with [12] and [15] and [16]

Method	MAPE(%)
RVNN[15]	22.535
CVNN[15]	14.867
AWNN[12]	7.081
Proposed Method in[16]	11.52
RBF Neural Network[16]	12.96
Persistence Method[16]	21.56
BP Neural Network[16]	17.6
MI+NN+BP	11.43
MI+NN+PSO	5.51
MI+NN+ICA	4.87

According to MAPE results, it can be seen that the proposed method in this paper namely training with ICA algorithm has about 78.37%, 67.24% , 31.21% ,57.72%,62.42%,77.41% and 72.32% decrease with respect to RVNN, CVNN , AWNN, proposed method in[16],RBF neural network, Persistence and back propagation neural network methods respectively. Also, for the PSO algorithm the MAPE value has a decrease of about 75.5%, 62.7%, 22.1%, 52.1%, 57.48%, 74.44% and 68.69% respectively. The MAPE value for the case when BP is used as training algorithm has a decrease to about 49.2% , 23.1%, 0.7%, 11.8%, 46.99% and 35.05% with respect to RVNN , CVNN, proposed method in[16],RBF neural network, Persistence and back propagation neural network methods respectively. However, the AWNN method has less MAPE value (about 37.8%) with respect to BP algorithm.

6. Conclusion

In this paper a two procedure configuration is gained for short-term forecasting of wind power using weather information involving wind speed, wind direction, humidity and temperature. The first procedure proposes a technique called MI for feature selection of input data to network such that using irrelevancy filter lead to the reduction in the input dimension and provides more relevant and appropriate information for improving the network performance. The second procedure involves neural network trained by training algorithms of BP, PSO and ICA for wind power prediction in a perspective of 24-houres. By investigating the achieved results, it can be said that the proposed evolutionary algorithms can effectively estimate the goal function (wind power). Also, the results verify that the ICA algorithm proposes better performance with respect to PSO algorithm and these two algorithms both propose better characteristics with respect to BP algorithm. It must be mentioned that evolutionary algorithms such as PSO and ICA are still completing and using their combination configuration can yield to better results. In addition to the performance of evolutionary algorithms, it can be seen that using MI technique as a FS tool yields to a significant reduction in the computational calculations in neural network and prevents the curse dimensionality problem. Thus, by investigating the results achieved from this paper [12],[15] and [16], it can be concluded first that MI is an efficient and appropriate method for feature selection and eliminating irrelevant elements in order to reduce the input dimension of neural network and second that evolutionary algorithms have better and more effective efficiency in predicting of the wind power value.

Appendix

Table A: PSO parameters

Number of Particle	300
Number of Iteration	350
Cognitive Acceleration	2
Social Acceleration	2
Initial Inertia Weight	0.9
Final Inertia Weight	0.4

Table B: ICA parameters

Number of Countries	250
Number of Initial Imperialist	30
Number of Decades Generations	300
Revolution Rate (Mutation Rate)	0.3
Assimilation Coefficient(β)	2
Assimilation Angle Coefficient	0.5
Zeta	0.02
Damped Ratio	0.99

References

[1] Maria Grazia De Giorgi, Antonio Ficarella and Marco Tarantino, “Error Analysis of Short Term Wind Power Prediction Models,” Applied Energy, vol. 88, pp. 1298-1311, November 2011.

[2] Yuan-Kang Wu, Jing-Shan Hong, “A literature review of wind forecasting technology in the world,” presented at IEEE Lausanne Power Tech Conf., pp.504-509, 2007.

[3] S.Fan, J.R.Liao, R.Yokoyama, L.Chen and W.J.Lee, “Forecasting the Wind Generation Using a Two Stage Network Based on Meteorological Information,” Energy Conversion, IEEE Transactions ,IEEE Power & Energy Society, vol. 24, no. 2, pp. 474 - 482, 20 January 2009.

[4] S.Jayaraj, K.Padmakumari, E.Sreevalsan and Arun P, “Wind Speed and Power Prediction Using Artificial Neural Networks,” in European Wind Energy Conference 2004, London.uk, November 22-25 2004.

[5] Lingling Li, Minghui Wang, Fenfen Zhu and Chengshan wang, “Wind Power Forecasting Based on Time Series and Neural Network” in Proceedings of the Second Symposium International Computer Science and Computational Technology(ISCSCT ’09), China, 26-28 Dec 2009.

[6] Catalao, J.P.S., Pousinho, H.M.I. and Mendes, V.M.F., “An Artificial Neural Network Approach for Short-Term Wind Power Forecasting in Portugal,” in Intelligent System Applications to Power Systems, 2009. ISAP '09. 15th International Conference on , Curitiba , 8-12 Nov. 2009 .

[7] Mahmoudi, J., Jamil, M. and Balaghi, H., “Short and mid-term wind power plants forecasting with ANN,” in Renewable Energy and Distributed Generation (ICREDG), Tehran , 6-8 March 2012.

[8] Kani, S.A.P. and Riahy, G.H., “A new ANN-based methodology for very short-term wind speed prediction using Markov chain approach,” in Electric Power Conference ‚EPEC 2008. IEEE, Canada , 6-7 Oct. 2008 .

[9] G. Sideratos and N.Hatziargyriou, “using radial basis neural networks to estimate wind power production,” in Power Engineering Society General Meeting, 2007. IEEE , Tampa,FL, 24-28 June 2007.

[10] Barbounis TG and Theochair JB, “Locally recurrent neural networks for long-term wind speed and power prediction,” Neurocomputing, vol. 69, no. 4-6, pp. 466-496, January 2006.

[11] Giuseppe Grassi and Pietro Vecchio, “Wind energy prediction using a two-hidden layer neural network,” Communications in Nonlinear Science and Numerical Simulation, vol. 15, no. 9, pp. 2262-2266, SEptember 2010.

[12] Kanna Bhaskar and S.N Singh, “AWNN-Assisted Wind Power Forecasting Using Feed-Forward Neural Network,” Sustainable Energy, IEEE Transactions, vol. 3, no. 2, pp. 306 - 315, April 2012.

[13] Ruddy Blonbou, “Very Short-Term Wind Power Forecasting with Neural Network and Adaptive Bayesian Learning,” Renewable Energy, vol. 36, no. 3, pp. 1118-1124, March 2011.

[14] Xin Weidong, Liu Yibing and Li Xingpei, “Short-Term Forecasting of Wind Turbine Power Generation Based on Genetic Neural Network,” in 8th World Congress on Intelligent Control and Automation, Jinan,China, July 6-9 2010.

[15] Ziqiao Liu , Wenzhong Gao, Yih-Huei Wan and Eduard Muljadi, “Wind Power Plant Prediction by Using Neural Networks,” in Energy Conversion Congress and Exposition (ECCE), IEEE, Raleigh, NC, 15-20 Sept. 2012

[16] Wy Chang,“Short-Term Wind Power Forecasting Using the Enhanced Particle Swarm Optimization Based Hybrid Method” Energies,2013;6(9):4879-4896

[17] N.Amjady, F.Keynia and H.Zareipour, “A New Hybrid Method for Short-Term Wind Speed Forecasting,” European Transactions on Electrical Power, vol. 21, no. 1, pp. 581-595, January 2011.

[18] N.Amjady and F.Keynia, “Day-Ahead Price Forecasting of Electicity Market by Mutual Information Technique and Cascaded Neuro-Evolutionary Algorithm,” IEEE Trans.Power Syst, vol. 24, no. 1, pp. 306-318, Feb.2009.

[19] Gurney K, “An Introduction to Neural Networks,” August 5, 1997.

[20] Ahmed Ouammi, Hanane Dagdougui and Roberto Sacile, “Short Term Forecast of Wind Power by An Artificial Neural Network Approach,” in Systems Conference (SysCon), 2012 IEEE International, Vancouver, BC, 19-22 March 2012.

[21] Kennedy.J and Eberhart.R, “Particle Swarm Optimization,” in Neural Networks, 1995. Proceedings., IEEE International Conference on (Volume:4) , Perth, WA , 27 Nov 1995-01 Dec 1995 .
[22] Chia-Nan Ko, Ying-Pin Chang and Chia-Ju Wu, “A PSO Method with Nonlinear Time-Varying Evolution for Optimal Design of Harmonic Filters,” Power Systems, IEEE Transactions, vol. 24, no. 1, pp. 437 - 444 , Feb 2009.
[23] E.Atashpaz Gargari and Lucas.C, “Imperialist Competitive Algorithm:An Algorithm for Optimization Inspired by Imperialistic Competition,” in Evolutionary Computation\. CEC 2007. IEEE Congress, Singapore, 2007.
[24] Shahram Mollaiy Berneti and Mehdi Shahbazian, “An Imperialist Competitive Algorithm Artificial Neural Network Method to Predict Oil Flow Rate of the Wells,” International Journal of Computer Applications, vol. 26, no. 10, pp. 47-50, July 2011.
[25] Vahid Khorani, Nafiseh Forouzideh and Ali Motie Nasrabadi, “Artificial Neural Network Weights Optimization Using ICA, GA, ICA-GA and R-ICA-GA: Comparing Performances,” in Hybrid Intelligent Models And Applications (HIMA), 2011 IEEE Workshop On, Paris, 11-15 April 2011.
[26] N.Kwak and C.H.Choi, “Input Feature Selection for Classification Problems,” Neural Networks, IEEE Transactions, vol. 13, no. 1, pp. 143 - 159 , Jan 2002 .
[27] Aoife M.Foley, Paul G.Leahy, Antonio Marvuglia and Eamon J.Mckeogh, “Current Method and Advances in Forecasting of Wind Power Generation,” Renewable Energy, vol. 37, no. 1, pp. 1-8, January 2012.
[28] Ziqiao Liu , Wenzhong Gao, Yih-Huei Wan and Eduard Muljadi, “Wind Power Plant Prediction by Using Neural Networks,” in Energy Conversion Congress and Exposition (ECCE), IEEE, Raleigh, NC, 15-20 Sept. 2012.