BACKPROPAGATION NEURAL NETWORK APPROACH FOR MEAN TEMPERATURE PREDICTION

Manal A. Ashour[1,*], Somia A. ElZahaby[2] & Mahmoud I. Abdalla[3]
[1,2] Al-Azher University, Faculty of Science, Math. Department
[3] Zagazig University, Faculty of Engineering, Electronics and communication Department

ABSTRACT

Temperature is one of the basic components of the weather. In this paper, mean temperature have been forecasted using Artificial Neural Network (ANN). The design of the ANN based on four weather parameters. The ANN design has been applied for Cairo city, the capital of Egypt. The training and testing used meteorological data for twenty years (1996- 2016). In this study we predict the mean temperature by using the artificial neural network ANN model and the multiple linear regression MLR model. This study provides a neural network model based on backpropagation to predict the mean temperature and to compare the obtained results with the results obtained by the multiple linear regression MLR model. The different performance evaluation criteria are introduced to compare the results obtained by the neural network model and the results obtained by the multiple linear regression. Results show improved performance of the neural network model over the multiple linear regression model.

Keywords: *artificial neural network model ANN, backpropagation BPN, multiple linear regression model MLR, mean temperature forecasting.*

1. INTRODUCTION

Predict the future is one of the key issues used in many fields. Time series analysis is one of the common statistical methods used to predict and which are used widely in many applications of statistical and economic terms in which the behavior of the dependent variable prediction based on its behaviour in the past. On the other hand, there is a modern method more accurate and effective in forecasting which can use logic in their operations rather than the idea of the fixed relationship between variables known as artificial neural networks. Neural networks are suitable way in the representation of relationships between variables which are different from the traditional ways in such that they are arithmetic system consists of a set of simple elements and associated with each other to run the data dynamically in response to external input. Neural networks consider as a data processing system which have certain performance characteristics in a manner simulates biological nerve system.

The artificial neural network ANN is a contemporary advanced methodology; it attracted the attention of many researchers and scientists in various fields and disciplines including medical, engineering, statistical research operations, information technology and others. In the last years artificial neural networks have gained an increasing importance in processing and analysis of time series and future forecasts calculations, due to the advantage of its great flexibility compared to the known conventional methods which approved in this area as well as its ability to self-learn and adapt to any model.

2. ARTIFICIAL NEURAL NETWORK

Artificial neural network is a computational technique which is designed to simulate the way in which the human brain process a specific task by a huge distributed parallel processing. Neural network is made up of simple processing units. These units are called neurons. Neurons are mathematical elements which have a nervous property in that they store practical knowledge and empirical information to make it available to the user and that by adjusting the weights.

It is a matter of surprise that the speed of the computer exceeds the speed of the nerve cells by ten billion times. In spite of that, someone can recognize a familiar face in the tenth of a second using nerve cells with speed do not exceed 1/1000 of a second. The maximum steps followed by the cells conclude no more than (1000/10) 100 steps in any way. Scientists did not reach a convincing and logical explanation for how low speed cells can reach solutions with a high speed. Neuron cells process data in parallel that gives it high speed. This was enough to entice many researchers to mimic human neural networks using computer simulation.

It was possible to simplify the cell components. In principle, it was sufficient to take some of its functions into account and use a small number of them and then represent it mathematically to get the artificial neurons. Artificial neuron is a processing unit. Figure 1 shows a simplest representation of the artificial neural cell.

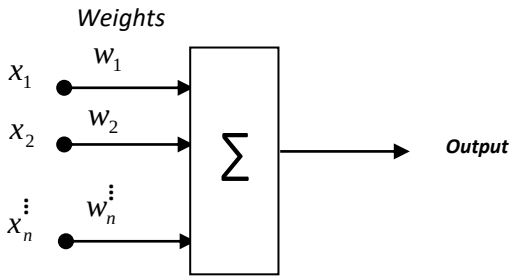

Figure 1. ANN model

Multiple network types have been proposed. Feed forward networks are the most important proposed network. In this structure, the signal transmit forward only from input to output. The output of any neuron in a layer affects only on the next layer and there is no any connection between neurons in the same layer. Most of the networks that follow this pattern consist of inputs layer and output layer. In addition to these two layers of the network, at least one hidden layer. Each of these tree layers consists of a number of neurons.

Supervised neural network need a teacher during the training phase to monitor to show the desired output for each input to eventually reach the correct results and hence there is no need to the external supervision teacher. This operation is carried out using several methods and algorithms. The most important method is the backpropagation method. This process starts by computing the error between the desired output and the calculated output. Then return this error back from the output layer to the hidden layers and finally to the input layer. During the backpropagation process, weights are changing in the direction that pushes the error to decrease to be zero. This training method used by the feed forward networks. Feed forward naming due to the network structure and backpropagation naming due to the training method used by the network. In this study the sigmoid function has been used as the activation function because it is very easy to be derived. Backpropagation enjoy with many benefits, the most important benefits are: the minimum square error guarantee, its ability to deal with confusing data and its ability to deal with systems of linear and non linear separable functions.

Training of the network consists of six basic steps; (1) give random weights to the network, (2) supply the network with inputs which are prepared for training, (3) apply feed forward process to compute the network outputs, (4) compare the actual output with desired output and determine the error value, (5) return the error over the network and correct weights in the direction that the error value decrease, (6) decrease the total error for each used input in training. Training process is required to be repeated many several times to get a lower value for the error. It should pay attention to some important issues when dealing with the network especially during network designing and network training. The lack of attention to this issues leads to incomplete or ineffective neural network. Some of them are; over fitting training, under fitting training, neural network size, normalization and learning rate. To avoid the problems of training, the data divided into three sets; training set, validation set and testing set.

The choice of the appropriate size of the network considers the toughest challenges ever in the artificial neural network design. In addition to the many available choices of the activation function for each neuron, in each layer, there is the problem of the choice of the appropriate number of neurons in the layer and the choice of the number of the hidden layer itself in the network. Without any doubt, all of these choices must be made before starting in training process. The conciliator choice of the network size leads to unacceptable results. The easiest and widely used method in choosing the size of the network is the "trail and error" method. The designer should to try out a number of networks and choose the best. This experimentation should to be systematically somewhat so as not to take a long time. The designer can start with a simple network and then increase its size bit by bit by adding layers or neurons until it reaches to acceptable results. It can also begin a complex network and works on simplifying gradually until it reaches to acceptable in terms of network complexity and performance [4,5,6]. One of the choices that the designer must be identified during the training process is the learning rate. This variable determines the speed of weights adapting. If the learning rate has a small value then the training process takes a long time. However, if the learning rate has a large value the weights may fluctuate and move away from the required weights taking the training process to an instability case. In practice, the designer can begin with a randomly learning rate

value ($\leq$ 1) and change it little by little to reach a suitable choice that combines between the speed of training and maintaining its stability.

3. ARTIFICIAL NEURAL NETWORK FOR WEATHER FORECASTING

After the first world war, Lewis Richardson in 1922 [12] supposed that he can use mathematics to predict the weather because atmosphere follows the laws of physics, but the equations were very complex process and the computing process consumed considerable time in such that the aerodynamic front go before the forecasters could complete their computation. In addition, Richardson used weather information taken every six hours, but as noted by the french meteorology specialist Rene Shabu that to reach the minimum value of accuracy in weather forecasting, it requires that measurements are taken every thirty minutes at most.

Weather forecasting includes the study of many different weather phenomena, such as temperature, barometric pressure and so many things and phenomena related to the atmosphere. Study weather phenomena were not someday luxury without interest but are the most important science that should be largely provided in all countries and regions due to the number of large applications that rely on this type of science, and the study of weather that could save the lives of people from dying as a result of abnormalities in the weather could lead to many disasters on the country and the whole region. For this reason, the first benefit from the study of weather is to save the lives of as many people before it is too late. Weather forecasting has many applications and other benefits, which are many and varied. Air traffic relies mainly on the weather. In addition, the study of changes in the atmosphere and the weather conditions are very useful in maintaining the vegetation in a specific area of the risk of ice formation which leads to death if the necessary precautions and measures are not taking. The Meteorology is very important in the study and analysis of the water situation in a given area and the amount of expected rain in a given area, and more areas affected by these rains, which helps a good deal with them and not to be wasted. Also, the science of meteorology is very important for maritime traffic; where meteorological working to determine whether or not the movement of ships, and states that these applications have many sections in the science of weather.

The use of neural networks to predict the various parameters of the weather as a new systematic science has proved as a successful technique in forecasting. Weather is a continuous, data intensive, multi dimensions, dynamic and chaotic process. These properties make weather forecasting a major challenge. Due to the non linear nature of weather's data, the focus has shifted towards the non linear prediction of the weather data. Since the weather data is non linear and follows a very irregular trend, artificial neural network has evolved out to be a better technique to bring out the structural relationship between the various entities. Artificial neural network is non linear flexible function that does not require the restricted assumptions on the relationship between the independent and dependent variables. Neural Networks treats well with nonparametric data or small data size in addition to its accuracy with parametric data. Many studies have shown that neural networks offer better predict levels compared with other traditional statistical methods as well as neural networks can manage analytical and predictive tasks in a way more speed which will positively reflect on the element of time.

The ANN model has been designed by using four basic numbers of procedures: (1) collecting data, (2) processing data, (3) building the network, (4) training, validating and testing network (figure 2). This study used daily data. The data was collected from Cairo Airport International Station (HECA) 30.11 N°, 31.41 E° through the interval 1996-2016 [2]. The mean temperature, mean dew point, mean humidity, mean sea level pressure and the mean wind speed are recorded in Excel data file consists of 20 data sheets represents the mean temperature, mean dew point, mean humidity, mean sea level pressure and mean wind speed belongs to January (1996-2016). After the data collection, the recorded data has been reprocessing. Data reprocessing procedures are conducted to train the ANN more efficiently. The missing data are replaced by the average of neighbouring values.

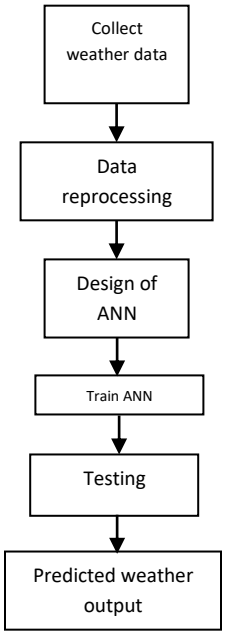

Figure 2. The block diagram of training procedure

In artificial neural network software, all inputs and outputs are normalized between 0 and 1. The use of the original data as input to neural network may cause convergence problem [3]. Neural networks provide improved performance with the normalized data. Data normalizing is the process of scaling data to fall within a smaller range. Normalizing help in speeding up the training phase. This step is very important when dealing with parameters of different units and scales. Therefore, all parameters should have the same scale for a fair comprise between them. Two methods are usually well known for rescaling data. Normalization scales all numeric variables in the range [0, 1] [13]. One possible formula is given below:

$$x_i^{'} = \frac{x_i - \min}{\max - \min}$$

On the other hand, you can use standardization on your data set. It will then transform it to have zero mean and unit variance [14, 15], for example using the equation below:

$$x_i^{'} = \frac{x_i - \mu}{\sigma}$$

where μ, σ are the mean and the standard deviation of the data set. It is worth mentioning that, both of these methods have their drawbacks. If you have outliers in your data set, normalizing your data will certainly scale the "normal" data to a very small interval. And generally, most of data sets have outlines. When using standardization, your new data are not bounded (unlike normalization). Normalization will help the input and output parameters to correlate with each other. The outputs then were denormalized into the original data format for achieving the desired results.

To forecast the temperature for the next day, we need to construct an abstract model that reflects the reality and actually of the real system. In order to keep the simplicity of the modelling structure, only one hidden layer with 15 neurons is considered. This number was arrived by trail and error method. Performance of the network was evaluated by increasing the number of hidden neurons. This number was arrived after analyzing 5, 10, 15, 20, 25 and 30 neurons in the hidden layer. The architecture with 5 and10 neurons in the hidden layer was faster in computation but the convergence rate is very slow. The architecture with 20, 25 and 30 neurons in the hidden layer was converging equally well as that with 15 neurons. Therefore, the architecture 15 neurons in the hidden layer were selected (table 1). After finding hidden neurons, epochs increase till we find the suitable epochs [9]. The number of input neurons is four representing the chosen weather parameters. The parameters chosen in this study for the prediction are the mean temperature, mean dew point, mean relative humidity, mean sea level pressure and mean wind speed.

Table 1. Meteorological variables

Parameter no.	Meteorological variables	units
1	Mean temperature	C°
2	Relative humidity	%
3	Dew point	C°
4	Sea level pressure (SLP)	hpa
5	Wind speed	Km/h

There is no reason behind this choice of weather parameters. The choice is made just to predict the mean temperature. The number of hidden neurons is 15 for processing and the number of outputs is 1 representing the weather variable to be forecasted. All the weather data sets were therefore, transformed into values between 0 and 1. Finally outputs were denormalized into the original data format for achieving the desired result. Training goal for the network was set to 10^{-2}. Network was trained for a fixed number of epochs (table 2). The network is first initialized by setting up all its weights to be small uniformly random numbers between -1 and +1 i.e. weights ~U (-1, 1). Next the input pattern is applied and the output calculated (forward pass). The calculation gives an output which is completely different to what you want (the target), since all the weights are random. We then calculated the error of each neuron, which is essentially*: predicted output – actual output*. This error is then used mathematically to change the weights in such a way that the error will get smaller. In other words, the output of each neuron will get closer to its target (reverse or back pass). The process is repeated again and again until the error is minimal.

Table 2 ANN model structure

Number of hidden layers	1
No. hidden neurons	15
No. of epochs	2000
Activation function used in input layer	linear
Activation function used in hidden layer	Sigmoid
Activation function used in output layer	Sigmoid
Learning Method	Supervised

4. MULTIPLE LINEAR REGRESSION MODEL (MLR)

Forecasting model based on time series data are being developed for prediction of the different variables. Regression is a statistical empirical technique and is widely used in business. The multiple regression linear model is fitted to predict the weather parameters [10,11]. The multiple regression linear model is fitted to predict the daily mean temperature as dependent parameter taking the other daily independent parameters as mean dew point, mean relative humidity, mean sea level pressure and mean wind speed. The most significantly contributed parameters are selected using enter regression analysis based on 31 data set of 20 years and the best fit multiple regression model is given below:

$$y = 1.003x_1 - 0.268x_2 - 0.05x_3 + 0.001x_4 + 75.331$$

where, mean temperature y as the dependent parameter having 91% Contribution of the significant parameters mean dew point x_1, mean relative humidity x_2, mean sea level pressure x_3 and mean wind speed x_4. The multiple linear regression MLR procedure for selecting the significant parameters for the mean temperature parameter is mentioned in Appendix.

5. PERFORMANCE EVALUATION CRITERIA

Many analytical methods have been proposed for the evaluation and inter comparison of different models, which can be evaluated in terms of graphical representation and numerical computations. The graphical performance criterion involves a linear scale plot of the actual and the predicted weather parameters for training and testing data for all the models. The numerical performance criteria involves

Mean error (BAIS):

$\frac{1}{N}\sum_{i=1}^{N}(X_i - Y_i)$

Mean absolute error (MAE):

$\frac{1}{N}\sum_{i=1}^{N}|X_i - Y_i|$

Root mean square error (RMSE):

$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - Y_i)^2}$

Prediction error (PE):

$\frac{\langle |X_i - Y_i| \rangle}{\langle Y_i \rangle}$

Correlation Coefficient (r): this is obtained by performing a regression between the predicted values and the actual values.

$$r = \frac{\sum_{i=1}^{N}\left((Y_i - \overline{Y}_i)(X_i - \overline{X}_i)\right)}{\sqrt{\sum_{i=1}^{N}(Y_i - \overline{Y}_i)^2 \sum_{i=1}^{N}(X_i - \overline{X}_i)^2}}$$

Where, $\langle\ \rangle$ implies the average over the whole test set, N is the total number of forecast outputs. Y_i and X_i are the actual and the predicted values respectively for $i = 1,2,...,N$, $\overline{Y_i}$ and $\overline{X}_i$ are the mean values of the actual and the predicted values respectively.

5. RESULT AND DISCUSSION

For the best prediction, the BIAS, MAE, and RMSE values should be small and PE should be sufficiently small i.e., close to 0. But r should be found closer to 1 (between 0-1) for indicating better agreement between actual and predicted values. The recital of weather forecasting models had been evaluated on the basis of SPSS version 14. It can be seen from table 3 that the bias for testing data is 0.0968 and 0.097 for NN and MLR models respectively. The bias is lesser for artificial neural network than the values that are obtained from MLR model. The MAE for NN is 2.351 which are lesser than the MAE for MLR model. The RMSE is also found to be 0.6887 for NN which is less than that of the MLR model. The artificial neural network model also shows a smaller value for PE against to this of MLR model. Further, the correlation coefficient is observed to be highest for artificial neural network model as 0.971677 in comparison to 0.942955 for multiple linear regression MLR model. Thus, the graphical representation (figure 3) as well as the numerical estimates both favored.

Table 3. Comparison of the performance of forecasting models for mean temperature

Model	ANN	MLR
Bias	0.0968	0.097
MAE	0.5566	0.559
RMSE	0.688799	0.689659
PE	1.006845	1.008509
CC	0.971677	0.942955

6. CONCLUSION

Artificial neural network is more accurate and efficient in predicting than the traditional statistical methods. ANN model with the learning rate of 0.7 and with momentum 0.3 is a preferred performance in comparison to MLR model, concluding that this ANN can be used as an effective mean temperature forecasting tool.

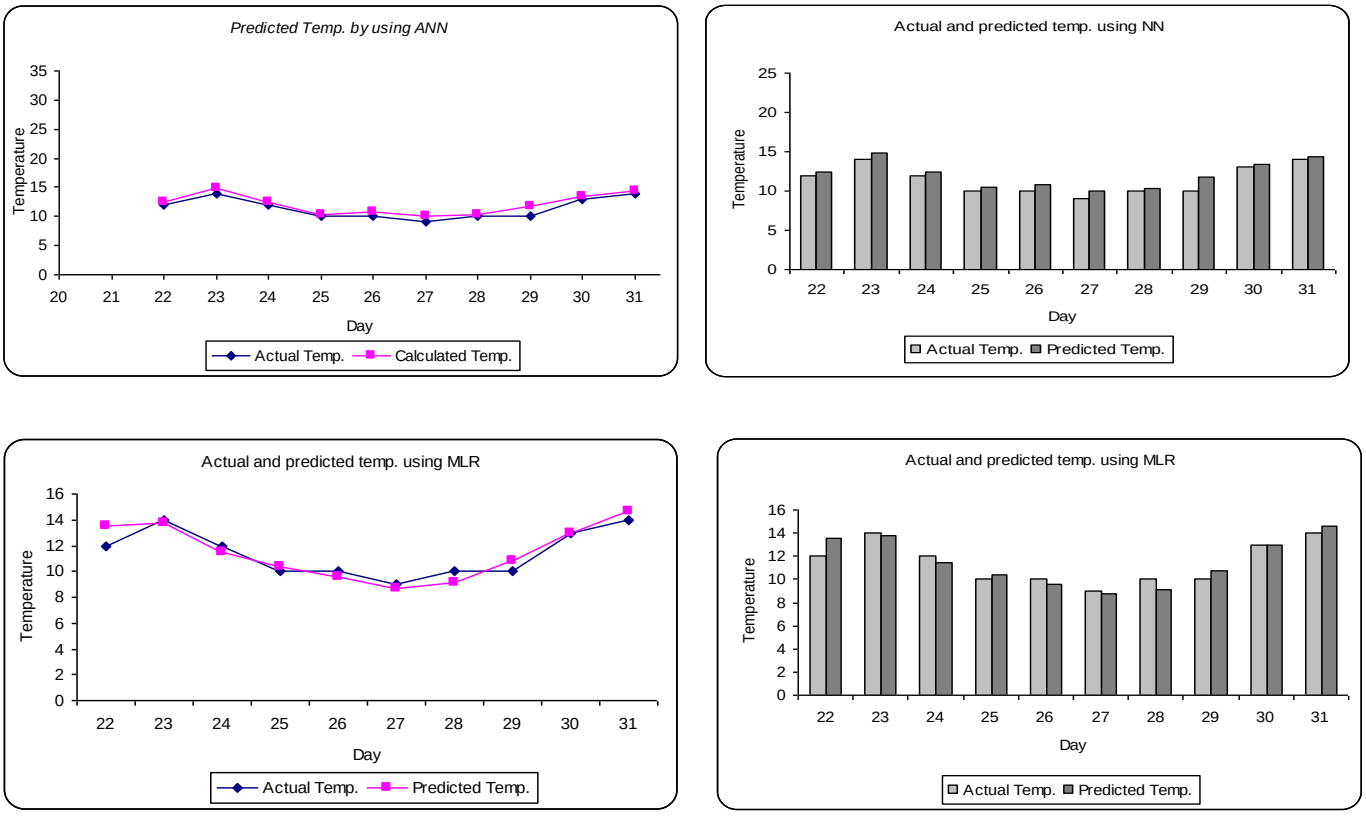

Figure 3 comparison of the actual and the predicted mean temperature using ANN and MLR models

REFERENCES

[1]. S. Mathur "A feature based neural network model for weather forecasting" World academy of science, engineering and technology 34, 2007.
[2]. https://www.wunderground.com/.
[3]. M.R.Khan and C.Ondrusek, "short- term electric demand prognosis using artificial neural networks", Journal of Electrical Engineering, Vol. 51, pp.296-300, 2000.
[4]. M.T Hagan, Demuth HB, Beale MH (1996) Neural Network design. PWS, Boston, MA.
[5]. M. Khan, Ondrusek C (2000) Short-term electric demand prognosis using artificial neural networks. J Elec Engg 51:296-300.
[6]. Y. Tan, Wang J, Zurada JM (2001) Nonlinear blind source separation using a radial basis function network. IEEE Transactions on Neural Networks 12:124-134.
[7]. R. Rojas, "The back propagation algorithm of Neural Networks – A Systematic Introduction, "chapter 7, ISBN 978-3540605058.
[8]. S. Chattopadhyay, "Multilayered feed forward Artificial Neural Network model to predict the average summer-monsoon rainfall in India, "2006.
[9]. Y. Inde, A. Buzo, R. Gray, An algorithm for vector quantizer design. IEEE Transactions, No. 28, 1980, pp. 84-95.
[10]. P. and Sanjay Mathur, "A simple weather forecasting model using Mathematical regression", Indian Research Journal ofExtension Education, special issue vol. 1,jan 2012.
[11]. O. Terzi, "Monthly Rainfall estimation using data mining process, applied computational intelligence and soft computing, vol. 2012.
[12]. P. Lynch, The Emergence of Numerical Weather Prediction. Quarterly journal of the royal meteorological society, Q. J. R. Meteorol. Soc. 133: 2143–2144, 2007
[13]. G. Shmueli, R. Nitin, C. Peter, Data Mining In Excel: Lecture Notes and Cases. Draft December 30, 2005. Distributed by: Resampling Stats, Inc. 612 N. Jackson St. Arlington, VA 22201, USA, info@xlminer.com, www.xlminer.com
[14]. G. Milligan, M. Cooper. 1988. A study of standardization of variables in cluster analysis. Journal of Classification, 5, 181-204.
[15]. O. Tongeren,. 1995. Cluster Analysis. Pages 174-212 in: R. H. G. Jongman, C. J. F. ter Braak, and O. F. R. van Tongeren, Eds. Data Analysis in Landscape and Community Ecology. (Cambridge & New York: Cambridge University Press).