International Journal of Research in Computer Science
eISSN 2249-8265 Volume 3 Issue 1 (2013) pp. 11-18
www.ijorcs.org, A Unit of White Globe Publications
doi: 10.7815/ijorcs. 31.2013.056

PREDICTION OF ATMOSPHERIC PRESSURE AT GROUND LEVEL USING ARTIFICIAL NEURAL NETWORK

Angshuman Ray[1], Sourav Mukhopadhyay[2], Bimal Datta[3], Srimanta Pal[4]

[1 2]Institute of Engineering and Management,Salt Lake Electronics Complex, INDIA
Email: [1]angshuman.ray@iemcal.com, [2]saurabh.mukherji@gmail.com

[3]Regent Education and Research Foundation Group of Institutions, Barrackpore, Kolkata, INDIA
Email: bkd_hetc@yahoo.co.in

[4]Indian Statistical Institute, Kolkata, INDIA
Email: srimanta@isical.ac.in

Abstract: *Prediction of Atmospheric Pressure is one important and challenging task that needs lot of attention and study for analyzing atmospheric conditions. Advent of digital computers and development of data driven artificial intelligence approaches like Artificial Neural Networks (ANN) have helped in numerical prediction of pressure. However, very few works have been done till now in this area. The present study developed an ANN model based on the past observations of several meteorological parameters like temperature, humidity, air pressure and vapour pressure as an input for training the model. The novel architecture of the proposed model contains several multilayer perceptron network (MLP) to realize better performance. The model is enriched by analysis of alternative hybrid model of k-means clustering and MLP. The improvement of the performance in the prediction accuracy has been demonstrated by the automatic selection of the appropriate cluster.*

Keywords: *Artificial neural networks, backpropagation, data clustering, multi-layer perceptron, pressure.*

I. INTRODUCTION

The short term prediction of atmospheric pressure is very important to know any kind of changes in weather condition at a particular place. Accurate information about weather and proper prediction of air pressure is often useful for warning about natural disasters caused by abrupt change in climatic conditions. With the prediction of pressure it is known in advance whether severe or dangerous storm is coming. If the air pressure becomes low suddenly, then it indicates that there is a possibility of bad weather. The fishermen will be given warning about the possibility of bad weather condition so that they can return to the sea shore from the mid of the sea. This warning also helps the concerned authority whether flights can be allowed to take-off from the airport. It is also important to note that the atmospheric pressure is just one of many factors that affect fish feeding habits.

Presently, weather predictions are made by collecting quantitative data about the current state of the atmosphere and using scientific understanding of the atmospheric processes to project how the atmosphere will evolve. The patterns of atmospheric pressure in Kolkata are shown in Figure 1a and Figure 1b.

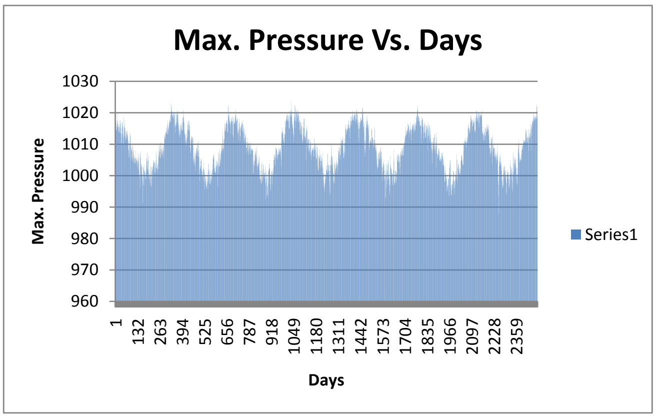

Figure 1(a): Actual max. pressure at Dumdum, Calcutta (year 1989 – 95)

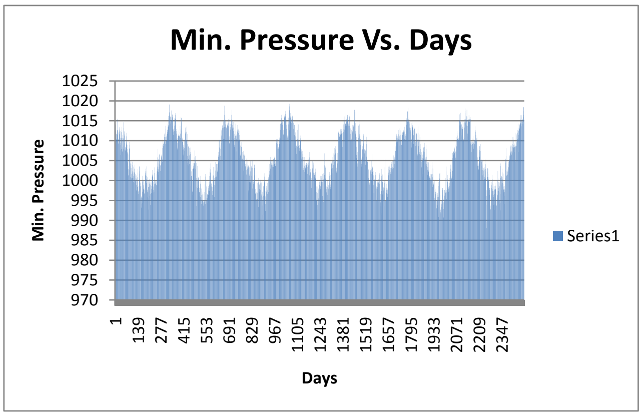

Figure 1(b): Actual min pressure at Dum Dum, Calcutta (year 1989 – 95)

Artificial Neural Networks (ANN) performs nonlinear mapping between the inputs and outputs without detailed consideration of the internal structure of the physical processes associated with pressure. This approach is essentially a data driven approach. ANN emulates the parallel distributed processing of the human nervous system and are parallel computational models, comprising closely interconnected adaptive processing units. The adaptive nature of neural networks adopts artificial intelligence (AI) learning techniques like supervised and unsupervised learning. ANN model has already proved to be very powerful in dealing with complex problems like function approximation, pattern recognition and has been applied for weather prediction, stock market prediction etc.

A number of studies have been reported that have used ANN to model complex nonlinear relation of input and output for weather forecasting [6][7]. However, very few works have used ANN-based connectionist methods to forecast air pressure. Again, all of these works are restricted to feed forward ANN models with back propagation and that uses either linear or nonlinear time series data only.

This paper is an outcome of an ANN based air pressure prediction model developed, trained and tested with continuous (daily) ground level air pressure data as input over a period of 7 years [1]. Here two distinct alternative models, namely MLP and Hybrid Kmeans-MLP have been studied and analyzed. However, these 2 models, MLP and Hybrid Kmeans-MLP were designed and tested separately with different number of hidden nodes. The results produced by each of the models were compared and the suitability of the models were justified. The model has been applied to justify that ANN is an appropriate predictor for air pressure forecasting. The prediction is based on the past observations of several meteorological parameters like temperature, humidity, air pressure and vapor pressure. The data was collected daily by the meteorological department of Dumdum Airport.

II. ANN APPROACH

In this section the basics of the 2 models as referred in Section 1 are discussed. This theoretical basis of the models has been applied during the design and implementation of the same.

A. Multilayer Perceptron (MLP)

MLP is one of the most widely used neural network architectures. It consists of several layers of neurons of which the first layer is known as the *input layer*, last layer is known as the *output layer* and the remaining layers are called as *hidden layer*. Every node in the hidden layers and the output layer computes the weighted sum of its inputs and apply a sigmoidal activation function to compute its output, which is then transmitted to the nodes of the next layer as input [3]. The main objective of MLP learning is to set the connection weights in such a way the error between the network output and the target output is minimized. A typical MLP network is shown in Figure 2. According to [2] under a fairly general assumption a single hidden layer is sufficient for multilayer perceptron to compute an uniform approximation of a given training set of input and output. So, the present study is restricted to three-layer network i.e. one hidden layer.

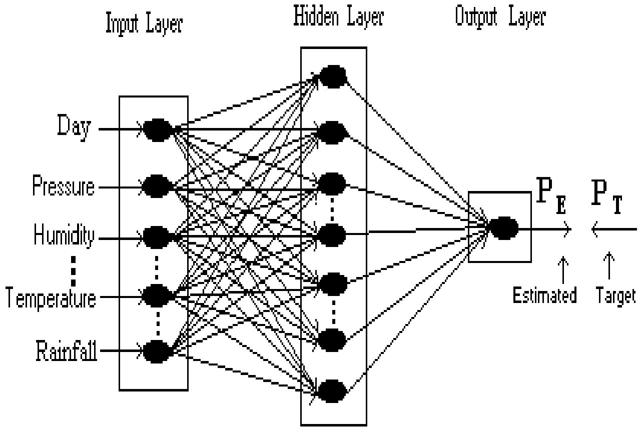

Figure 2: A typical MLP for air pressure prediction

B. Clustering

Cluster analysis or clustering is the assignment of a set of observations into subsets (called clusters) so that observations in the same cluster are similar in some sense. It is used to operate on a large data-set to discover hidden pattern and relationship helps to make decision quickly and efficiently. Clustering is a method of unsupervised learning, and a common technique for statistical data analysis used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics.

1. Implementation of K-Means Clustering Algorithm:

K-Means is one of the simplest unsupervised learning algorithms used for clustering. K-means partitions n observations into k clusters in which each observation belongs to one of the clusters whose centre is nearest.This algorithm aims at minimizing an objective function, in this case a squared error function.

Initially we have only the raw data. So, it is clustered around a single point. If the cluster number K is fixed then we need to cluster around that point. If the cluster is not fixed then it is continued until the centered is not changed. Initially the whole data is in a same group. But when K-means clustering is applied on it then it clusters the whole data into four major categories.

The following is the algorithm:

Step 1. The algorithm arbitrarily selects k points as the initial cluster centers ("means").

Step 2. Each point in the dataset is assigned to the closed cluster, based upon the Euclidean distance between each point and each cluster center.

Step 3. Each cluster center is recomputed as the average of the points in that cluster.

Step 4. Repeat Step 2 and Step 3 until the centroids no longer move. This produces a separation of the objects into different clusters.

III. DATA SETS AND EXPERIMENTS

The present study developed an ANN model based on the past observations of several meteorological parameters like temperature, humidity, air pressure and vapor pressure as an input for training the model. The developed model overcomes the difficulties in training ANN models with continuous data. The architecture of the proposed model contains several multilayer perceptron network (MLP). The model is enriched by analysis of alternative hybrid model of k-means clustering and MLP for better prediction. The improvement of the performance in the prediction accuracy has been demonstrated by the online selection of the appropriate cluster.

The experiments were carried out in the following sequence. First, the effectiveness of multilayer perceptron networks was studied for prediction of air pressure. Next, in Hybrid model cluster is selected online while producing good prediction.

A. Data Acquisition

The meteorological data were captured by the Dumdum meteorological center of Kolkata. The parameters of the data acquisition are:

i. Minimum Temperature *(Min. Temp(t))*
ii. Maximum Temperature *(Max. Temp(t))*
iii. Minimum Relative Humidity *(Min. RH(t))*
iv. Maximum Relative Humidity *(Max. RH(t))*
v. Minimum Air Pressure *(Min. Press.(t))*
vi. Maximum Air Pressure *(Max. Press.(t))*
vii. Minimum Vapour Pressure *(Min. VP(t))*
viii. Maximum Vapour Pressure *(Max. VP(t))*
ix. Rainfall *(Rain(t))*

This information is stored in an input file. The file contains data of seven years. So, there is an observation of 9 variables on a particular day, say *t*. In the MLP model the air pressure for the next (t^{th}) day is determined by the atmospheric parameters for the current day i.e. day *(t-1)*. To enable the selection of the best model, the training data set should cover air pressure at different seasons. So the data for entire years were chosen as the training data sets. The data is pre-processed before training. The detail of the pre-processing is discussed in the next section.

Our model has the following functional relations.

Max. Pressure (t) =

$$f\big(Min.Temp(t-1), Max.Temp(t-1), Min.RH(t-1), Max.RH(t-1), Min.Press(t-1), Max.Press(t-1), Min\,VP(t-1), Max.VP(t-1), Rain(t-1)\big)$$

Max. Pressure (t) =

$$f\big(Min.Temp(t-1), Max.Temp(t-1), Min.RH(t-1), Max.RH(t-1), Min.Press(t-1), Max.Press(t-1), Min\,VP(t-1), Max.VP(t-1), Rain(t-1)\big)$$

This model is non-deterministic model and therefore Artificial Neural Network approach is used to predict maximum and minimum pressure at a particular day on the basis of past observations.

The functional relation between the max. pressure and other parameters is non-linear. The ANN model is a non-linear model. That is why we are implementing the prediction of max. pressure and min. pressure by using ANN approach.

The relation between output (O) and input (**x**) is given by

$$\left[O = F(x) = \frac{1}{1+e^{-w.x}}\right]$$

Where **x** =

$$Min.Temp(t-1), Max.Temp(t-1), Min.RH(t-1), Max.RH(t-1), Min.Press(t-1), Max.Press(t-1), Min\,VP(t-1), Max.VP(t-1), Rain(t-1)$$

B. Data Preprocessing

Neural network training can be made more efficient if certain preprocessing steps are performed on the network inputs and targets. Before training, it is often useful to scale the inputs and targets so that they always fall within a specified range. The input metrics were *normalized* using min-max normalization. Min-max normalization performs a linear transformation on the original data. Suppose that minA and maxA are the minimum and maximum values of an attribute A. It maps value v of A to v' in the range -1 to 1 using the formula below:

$$y' = \frac{v - min_A}{max_A - max_A}$$

The trained network is simulated with the normalized input, and then the network output is converted back into the original units.

C. Methodologies

The data was organized for training and testing. The records of the input file were partitioned into two separate files. One input file containing 60% of the total records was used for training the ANN. This is considered as 'training data set'. The other input file containing rest 40% records was considered as 'test data set'. This was used for testing the network. The resulting output predictions $y_j(t)$ are compared with a corresponding desired or actual output, $d_j(t)$. The mean squared error at any time t, $E(t)$, is calculated using the formula.

$$MSE(t) = \frac{1}{2}\sum (y_j(t)d_j(t))^2 \ for\ j = 1\ to\ n$$

For the MLP model, the transfer function is the well-known sigmoid function. There was a single hidden layer and several runs of MLP were made with different number of hidden nodes in the hidden layer. The number of nodes in hidden layer was taken as 3, 4, 5, 6 and 7.

In case of hybrid Kmeans-MLP model, the model itself partitions the training data into homogenous subgroups. It trains all the clusters (subgroups) as different networks. While testing, the model read a set of test data first and it selects the proper network for this data set by clustering method and then it triggers that network for the computation of the output.

The following is the diagrammatical representation for the use of hybrid K means-MLP network.

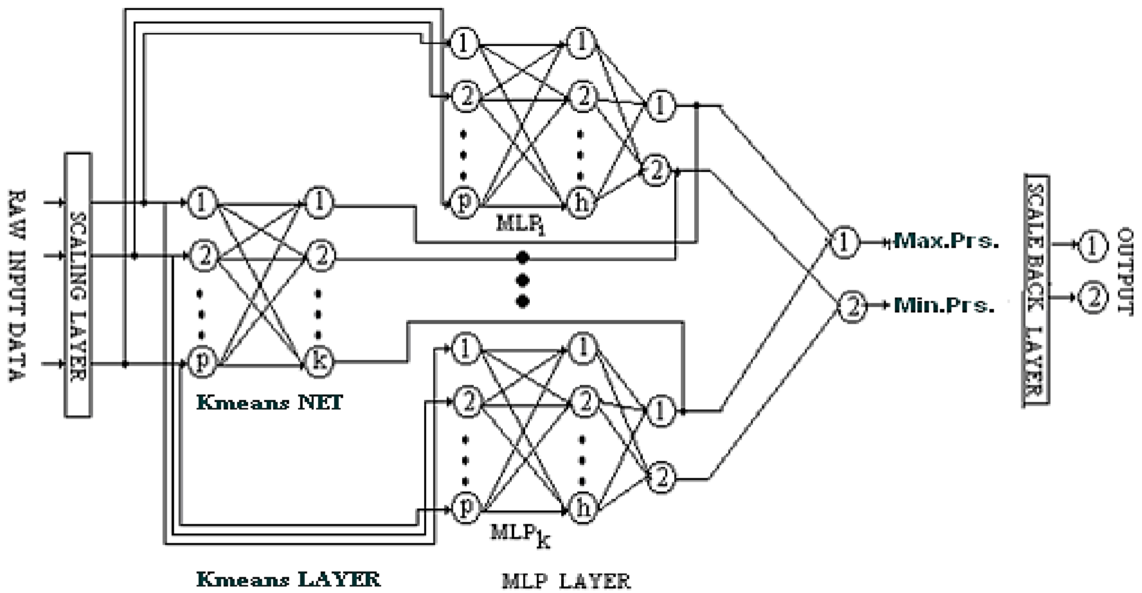

Figure 3: Hybrid Neural Network for Pressure prediction.

Table 1: Cumulative percentage frequency for MLP Networks

Range in mb	**% Frequency of pressure for test data**									
	n_h **3**		n_h **4**		n_h **5**		n_h **6**		n_h **7**	
	max	**min**	**max**	**min**	**max**	**min**	**max**	**min**	**max**	**min**
±0.5	7.8	12.0	7.8	12.7	9.3	13.8	7.2	13.0	7.3	13.0
±1.0	16.0	25.5	15.5	26.2	19.0	28.0	16.0	29.8	16.7	26.8
±1.5	24.0	40.8	23.8	41.7	29.5	39.8	26.7	43.5	24.3	40.7
±2.0	33.3	51.5	32.8	51.7	39.5	52.0	36.0	56.0	33.8	52.5
±2.5	41.7	62.2	41.2	62.7	48.2	62.5	46.8	67.3	43.3	63.7
±3.0	51.7	72.0	50.7	73.7	59.0	72.2	55.7	76.2	52.2	74.0
±3.5	61.8	80.7	62.3	81.2	68.5	80.5	67.3	82.8	63.2	81.8
±4.0	72.8	85.8	73.5	86.2	77.0	86.2	77.0	87.8	74.5	87.0
Max Dev	11.1	9.0	12.9	9.1	11.7	9.1	14.0	9.2	13.7	9.0
Avg Dev	2.9	2.2	2.9	2.2	2.7	2.2	2.8	2.1	2.9	2.2

IV. RESULTS AND OBSERVATIONS

After the input file is prepared, the training is done taking into consideration all the parameters such as:

$Min\ Temp(t-1), Max\ Temp(t-1), Min\ Vpr\ Prs(t-1), Max\ Vpr\ Prs(t-1), Max\ Prs(t-1), Min\ Prs(t-1), Min\ Rel\ Humidity(t-1), Max\ Rel\ Humidity(t-1), rainfall(t-1)$

After the training the testing is done. The result is shown in the previous table (Table 1). The graphical representation of the target and computed air pressure is shown in the following Figure 4a and Figure 4b.

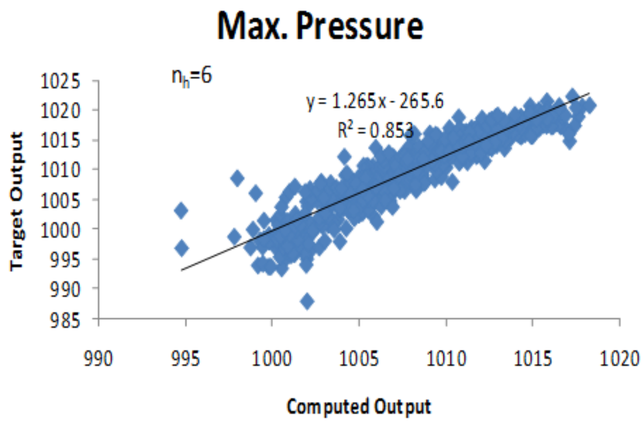

Figure 4(a): Graphical representation of target and computed max pressure for MLP Network.

Figure 4a implies that the square of the measure of goodness of fit R^2 is 0.853 on the basis of the MLP. Similarly from Figure 4b it is found that the square of the measure of goodness of fit R^2 is 0.880.

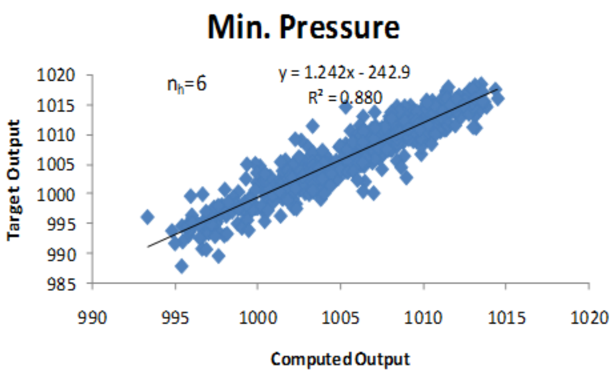

Figure 4(b): Graphical representation of target and computed min pressure for MLP Network.

The results obtained from MLP are satisfactory. But these results are not so good. One possible reason for this can be the presence of seasonality. This can be improved further. So we now propose a hybrid network of Kmeans-MLP which can account for seasonality of data. Our basic philosophy would be as follows.

The hybrid Kmeans-MLP model will group the data, X, into a set of homogeneous subgroups. Then for each subgroup it trains a separate feed forward network. In this prediction, the model will choose the appropriate trained MLP and then apply the test input to that net to get the prediction. The partitioning of the training data will be done using a K-means clustering.

The following tables (Table 2 and Table 3) and Figure (Figure 5) shows that the hybrid model gives the better result.

Table 2: Cumulative percentage frequency for hybrid Kmeans-MLP Networks

Range in mb	**% Frequency of pressure for test data**									
	n_h **3**		n_h **4**		n_h **5**		n_h **6**		n_h **7**	
	max	**min**	**max**	**min**	**max**	**min**	**max**	**min**	**max**	**min**
±0.5	12.3	16.0	12.5	17.3	14.0	16.7	15.3	18.7	15.3	17.8
±1.0	25.0	33.0	26.3	33.5	27.0	34.2	28.5	33.8	28.2	33.7
±1.5	37.3	47.7	37.0	49.3	39.3	49.3	41.7	48.7	40.7	48.8
±2.0	47.7	60.3	49.3	61.8	49.8	62.8	52.2	63.3	52.0	64.5
±2.5	55.7	69.5	58.3	71.8	59.0	71.2	63.7	72.0	62.8	72.3
±3.0	64.5	79.7	67.2	78.7	68.3	80.5	74.0	79.2	74.5	79.3
±3.5	74.8	86.0	76.8	86.8	77.0	87.7	82.5	87.5	82.3	87.7
±4.0	84.0	90.8	84.0	91.0	84.8	92.2	87.3	91.3	87.5	92.2
Max Dev	21.6	8.2	12.6	8.5	11.4	8.2	10.6	8.3	24.5	8.2
Avg Dev	2.5	1.9	2.3	1.9	2.3	1.8	2.1	1.8	2.2	1.8

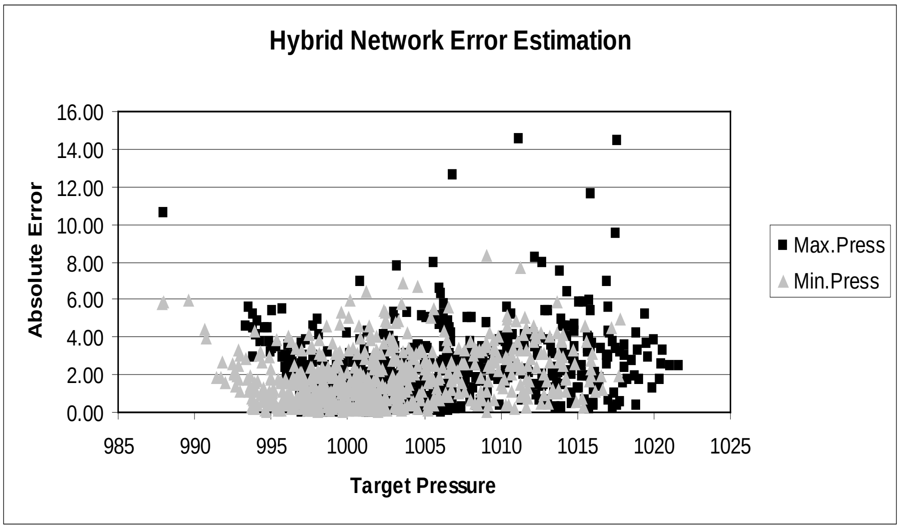

Figure 5: Graphical representation of absolute error and target air pressure after clustering.

Table 3: Air Pressure estimation with hybrid Kmeans-MLP Networks.

Target Pressure		Computed Pressure		Absolute Error	
max	**min**	**max**	**min**	**max**	**min**
1003.50	1001.80	1005.69	1001.71	2.19	0.09
1003.30	1001.00	1001.28	999.08	2.02	1.92
1003.10	999.90	1001.13	998.58	1.97	1.32
1003.10	998.80	1002.51	999.55	0.59	0.75
1006.50	1001.40	1003.68	999.74	2.82	1.66
1008.70	1006.80	1008.06	1005.06	0.64	1.74
1006.30	1005.20	1006.10	1003.60	0.20	1.60
1005.90	1004.00	1004.13	1001.61	1.77	2.39
1005.60	1002.50	997.61	997.66	7.99	4.84
1006.10	1003.50	1004.87	1002.26	1.23	1.24
1004.70	1003.10	1004.19	1002.00	0.51	1.10
1003.40	1000.50	1001.85	1000.37	1.55	0.13
1000.70	997.80	1001.39	998.31	0.69	0.51
1000.80	997.80	1001.65	997.37	0.85	0.43
1001.40	998.50	1001.43	997.53	0.03	0.97
998.50	997.00	1000.58	997.93	2.08	0.93
997.60	994.30	998.73	995.24	1.13	0.94
997.60	995.70	999.23	995.09	1.63	0.61
998.10	994.70	998.05	994.66	0.05	0.04
1003.60	998.40	1000.74	997.04	2.86	1.36
1003.70	1000.50	1003.63	1000.35	0.07	0.15
1002.00	1000.90	1002.78	999.49	0.78	1.41
1001.00	998.20	1000.31	997.38	0.69	0.82
1001.40	997.80	1001.43	997.55	0.03	0.25
1002.40	998.80	1000.85	998.41	1.55	0.39
1002.90	998.80	1001.85	998.73	1.05	0.07
1003.70	1001.20	1005.76	1001.87	2.06	0.67
1003.30	999.40	1002.88	999.78	0.42	0.38
1003.00	1000.10	1003.94	1000.00	0.94	0.10
1000.70	998.40	1003.14	999.16	2.44	0.76

After clustering technique the test result improves. The result is quite good. The graphical representation of the comparative study of the MLP and hybrid Kmeans-MLP is shown in the following Figure 6a and Figure 6b.

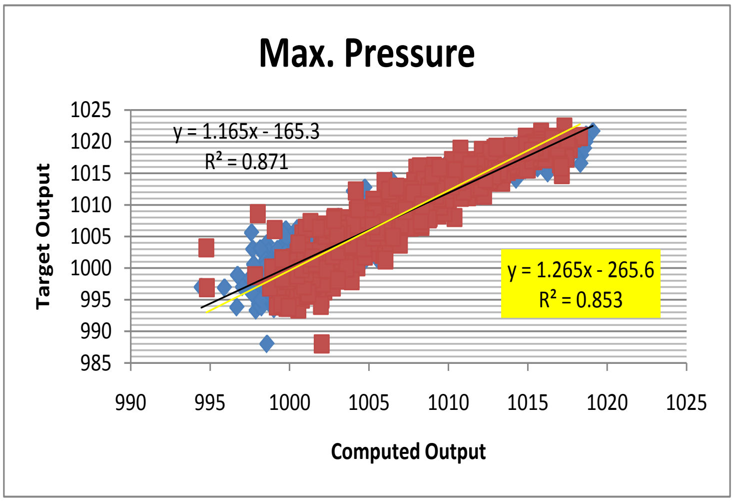

Figure 6(a): Graphical representation of target and computed max. pressure.

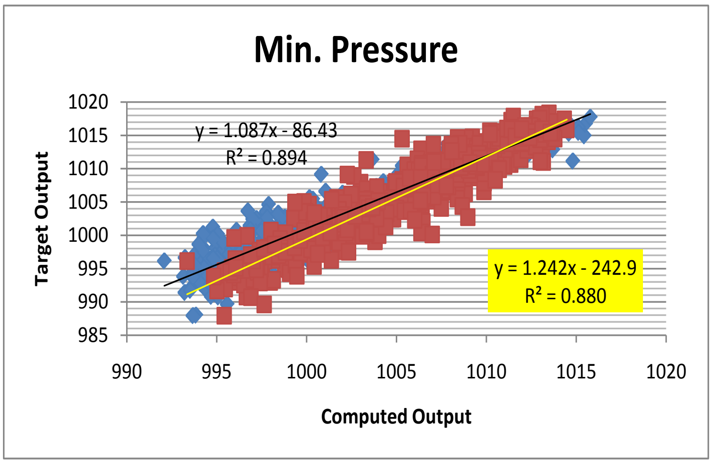

Figure 6(b): Graphical representation of target and computed min. pressure.

Figure 6(a) implies that the square of the measure of goodness of fit R^2 is 0.871 on the basis of the hybrid Kmeans-MLP. Similarly from Figure 6(b) it is found that the square of the measure of goodness of fit R^2 is 0. 894.

By comparative study, the predicted result of maximum pressure and minimum pressure using hybrid Kmeans-MLP is better than MLP, as the goodness of fit R^2 in case of hybrid Kmeans-MLP is greater than the measure of goodness of fit R^2 in case of MLP.

A. Observation

If a comparative study is done between Table 1 and Table 2, then it is clearly visible that in Table 1, the efficiency level of the neural network system was low. The average error from the Table 1 for n_h =6 is found to be 2.1 in case of minimum pressure and 2.8 in case of maximum pressure. Again using clustering, the predicted result becomes quite good as shown in Table 2. The average error from the Table 2 for n_h =6 is 1.8 in case of minimum pressure and 2.1 in case of maximum pressure.

So clustering can be used to improve the predicted result of a neural network based prediction system. The accuracy level becomes high after incorporating clustering technique.

V. CONCLUSIONS

Artificial neural network model discussed here has been developed to predict air pressure for a particular day based on the data of previous day. The meteorological data of the year 1989-1995 were collected from Kolkata Meteorological center and used for study of the proposed model. Two alternative ANN models were tested to compute the output and this computed output was compared with the target output i.e. pressure. After testing these models, the following conclusions are made.

i. Hybrid model of K-means and MLP turns out to be an excellent tool that can predict the air pressure accurately by overcoming the seasonality effect on air pressure.

ii. The neural network models proposed here can be good alternatives for traditional meteorological approaches for weather forecasting.

In the future works, the combined use of Feature selection and hybrid Kmeans-MLP may result in an excellent paradigm for prediction of air pressure. Moreover, hybrid Kmeans-MLP set may be used for prediction of other atmospheric parameters.

VI. REFERENCES

[1] Dutta, B., Ray, A., Pal, S., Patranabis, D.C. 2009. "A connectionist model for rainfall prediction". Neural, Parallel and Scientific Computations, vol. 17, pp. 47-58.

[2] Haykin S.,. "Neural Networks : A Comprehensive Foundation, Second Edition" , Pearson Education, 2001.

[3] Kohonen, T., "The self-organizing map". Proceedings of the IEEE, vol.78, no.9, pp. 1464-1480. doi: doi: 10.1109/5.58325

[4] Pal N. R., Chintalapudi K., "A connectionist system for feature selection". Neural, Parallel and Scientific Computations, vol.5, pp. 359-381, 2007.

[5] Pal N. R., Pal S., Das J., Majumdar, K., "SOFM-MLP: A hybrid neural network for atmospheric temperature prediction". IEEE Transactions on Geoscience and Remote Sening,. vol. 41, pp. 2783-2791. doi: doi: 10.1109/TGRS.2003.817225

[6] Paras, Sanjay Mathur, Avinash Kumar, and Mahesh Chandra, "A Feature Based Neural Network Model for Weather Forecasting". World Academy of Science, Engineering and Technology, vol. 34, pp. 66-73, 2007.

[7] Raodknight C.M., Balls G.R., Mills G.E., Palmer-Brown D., "Modeling Complex Environmental Data", IEEE Transactions on Neural Networks, vol. 8, no. 4, pp. 852-861. 1997. doi: 10.1109/72.595883

[8] Sarma D.K., Konwar M., Das J., Pal S. Sharma S., "A soft computing approach for rainfall retrieval from the TRMM microwave imager". IEEE Transactions on Geoscience and Remote Sensing, vol. 43, pp. 2879-2885, December 2005. doi: 10.1109/TGRS. 2005.857910

How to cite

Angshuman Ray, Sourav Mukhopadhyay, Bimal Datta, Srimanta Pal, "Prediction of Atmospheric Pressure at Ground Level using Artificial Neural Network". *International Journal of Research in Computer Science*, 3 (1): pp. 11-18, January 2013. doi: 10.7815/ijorcs. 31.2013.056