Available at: www.researchpublications.org

Use of Artificial Neural Network in Data Mining For Weather Forecasting

Gaurav J. Sawale[#], Dr. Sunil R. Gupta[*]

[#]*Department Computer Science & Engineering,*
P.R.M.I.T& R, Badnera.
[1]gaurav.sawale@yahoo.co.in
[*] *P.R.M.I.T& R, Badnera*
[2]sunilguptacse@gmail.com

Abstract— **Knowledge of weather data or climate data in a region is essential for business, society, agriculture and energy applications. The main aim of this paper is to overview on Data mining Process for weather data and to study on weather data using data mining technique like clustering technique. By using this technique we can acquire.Weather data and can find the hidden patterns inside the large dataset so as to transfer the retrieved information into usable knowledge for classification and prediction of climate condition. We discussed how to use a data mining technique to analyze the Metrological data like Weather data.. A variety of data mining tools and techniques are available in the industry, but they have been used in a very limited way for meteorological data. In this paper, a neural network-based algorithm for predicting the atmosphere for a future time and a given location is presented. We have used Back Propagation Neural (BPN) Network for initial modelling. The results obtained by BPN model are fed to a Hopfield Network. The performance of our proposed ANN-based method (BPN and Hopfield Network based combined approach) tested on 3 years weather data set comprising 15000 records containing attributes like temperature, humidity and wind speed. The prediction error is found to be very less and the learning converges very sharply. The main focus of this paper is based on predictive data mining by which we can extract interesting (non-trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of meteorological data.**

Keywords— **Artificial Neural Network (ANN), Back Propagation Neural (BPN) Network, Hopfield Network, Meteorological data, Data Mining.**

I. INTRODUCTION

Data mining [4] is the extraction of hidden predictive information from large databases. It is a powerful new technology with great potential to analyze important information in databases and data warehouses. Weather data has Synoptic data or climate data are the two classifications. Climate data is the official data record,usually provided after some quality control is performed on it. Synoptic data is the real-time data provided for use in aviation safety and forecast modelling. We know the Climate and weather affects the human society in all the possible ways. For example: Crop production in agriculture, the most important factor for water resources i.e. Rain, an element of weather, and the proportion of these elements increases or decreases due to change in climate. The effect of frost on the growth and quality of crops is leading potentially to total harvest failure. Meteorology is one of such domains,where data mining can improve the productivity of its analyststremendously by transforming their voluminous,unmanageable and prone to ignorance information into usable pieces of knowledge.

Weather forecasting was done previously by checking barometric pressure, current weather conditions, and sky condition with manual calculation. But forecast models based on data mining, soft computing are now used to determine future conditions. Human input is still required to pick thebest possible forecast model to base the forecast upon, which involves pattern recognition skills, teleconnections,knowledge of model performance, and knowledge of model biases.The chaotic nature of the atmosphere, the massive computational power required to solve the equations that describe the atmosphere, error involved in measuring the initial conditions, and an incomplete understanding of atmospheric processes mean that forecasts become less accurate as the difference in current time and the time for which the forecast is being made (the range of the forecast) increases.Data mining is of two types, Descriptive data mining, and Predictive data mining. Descriptive mining describes concepts or task-relevant data sets in concise, summarized, informative,discriminative forms, and Predictive mining based on data and analysis, constructs models for the database, and predicts the trend and properties of unknown data. Here we use Predictive mining, by which we can predict exact weather conditions.Initially we have to build up a model from the training data set using Artificial Neural Network (ANN) and after that we have to test the accuracy of that model by using test data

Available at: www.researchpublications.org

test.Before building the model, some pre-processing steps are applied to the data to help improve the accuracy,efficiency,and scalability of the classification or prediction process. After collecting the whole data set, data cleaning is one of the important parts. Data cleaning is a pre-processing of data in order to reduce noise and handle missing values. After data cleaning, the next step is relevance analysis or feature selection which removes the irrelevant or redundant attributes.And finally Data transformation is applied on the data set to normalize the data (such as -1.0 to 1.0, or 0.0 to 1.0). Then we analyse data with the help of Predictive data mining concept.After creating the whole model, applying all those processes,we can definitely predict some good results for weather.

II. RELATED WORK

In this section we are presenting the related work done in the area of weather prediction system and BPN network.Y.Radhika and M.Shashi [3] presented an application of Support Vector Machines (SVMs) for weather prediction.They used time series data of daily maximum temperature at location to predict the maximum temperature of the next day at that location based on the daily maximum temperatures for a span of previous n days referred to as order of the input.Performance of the system is observed for various spans of 2 to 10 days by using optimal values of the kernel. Mohsen Hayati et. al, [5] used the Multi Layer Perceptron on data set of ten years meteorological data. The data set was for year 1996 to 2006. The results show that MLP network has the minimum forecasting error and can be considered as a good method to model the short-term temperature forecasting [STTF] systems.Brian A. Smith et. al, [6] focused on developing ANN models with reduced average prediction error by increasing the number of distinct observations used in training, adding additional input terms that describe the date of an observation,increasing the duration of prior weather data included in each observation, and reexamining the number of hidden nodes used in the network. Models were created to forecast air temperature at hourly intervals from one to 12 hours ahead.Each ANN model, having network architecture and set of associated parameters, was evaluated by instantiating and training 30 networks and calculating the mean absolute error(MAE) of the resulting networks for some set of input patterns.Mike O'Neill [7] focus on two major practical considerations: the relationship between the amounts of training data and error rate (corresponding to the effort to collect training data to build a model with given maximum error rate) and the transferability of models' expertise between different data sets (corresponding to the usefulness for general handwritten digit recognition).S. Santhosh Baboo and I. Kadar Shereef [10] present an application of Back Propagation Neural (BPN) Network forweather prediction. Their proposed idea is tested using the real time dataset. The results are compared with practical working of meteorological department and these results confirm that the real time processing of weather data indicate that the BPN based weather forecast have shown improvement not only over guidance forecasts from numerical models, but overofficial local weather service forecasts as well.

S. Kotsiantis et al. [13] focused on investigating the efficiency of data mining techniques in estimating minimum,maximum and mean temperature values. A number of experiments have been conducted by them with well-known regression algorithms using temperature data from the city of Patras in Greece. The performance of their algorithms has been evaluated using standard statistical indicators, such as Correlation Coefficient, Root Mean Squared Error, etc.

III. ARTIFICIAL NEURAL NETWORK (ANN)

Artificial Neural Network [1, 2] is a powerful data modelling tool that is able to represent complex input /output relationships.The motivation for the development of neural network technology stemmed from the desire to implement an artificial system that could perform intelligent tasks similar to those performed by the human brain. A **Back Propagation Network** [9, 12] consists of at least three layers (multi layer perception): an input layer, at least one intermediate hidden layer, and an output layer. Typically, input units are connected in a feed-forward fashion with input units fully connected to units in the hidden layer and hidden units fully connected to units in the output layer. An input pattern is propagated forward to the output units all the way through the intervening input-to-hidden and hidden-to-output weights when a Back Propagation network is cycled. As the algorithm's name gives a meaning, the errors (and therefore the learning) propagate backwards from the output nodes to the inner nodes. The multilayer neural network shown in Figure 1.

Available at: www.researchpublications.org

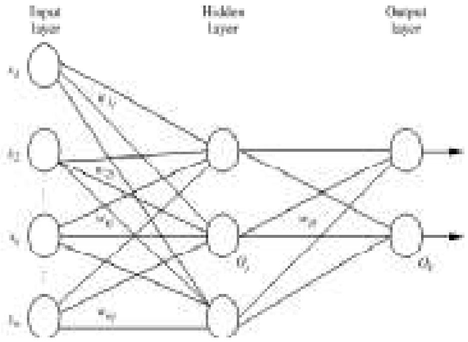

Fig 1: A multi layer feed-forward neural network.

A **Hopfield Network** [11] is a form of recurrent artificial neural network invented by John Hopfield. Hopfield networks serve as content-addressable memory systems with binary threshold units. They are guaranteed to converge to a local minimum, but convergence to one of the stored patterns is not guaranteed. A neuron of Hopfield networks is presented in Fig 2.

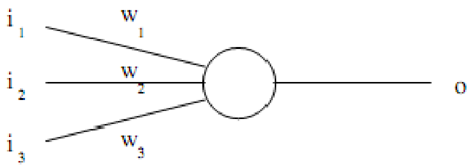

Fig 2: An artificial neuron as used in a Hopfield network.

These artificial neurons have N inputs. With each input i there is a weight wi associated. They also have an output. The state of the output is maintained, until the neuron is updated.

Updating the neuron entails the following operations:

• The value of each input, xi is determined and the weighted sum of all inputs, i wixi is calculated.

• The output state of the neuron is set to +1 if the weighted input sum is larger or equal to 0. It is set to -1 if the weighted input sum is smaller than 0.

• A neuron retains its output state until it is updated again.

Written as a formula:

$$o = \begin{cases} 1 & : \sum_i w_i x_i \geq 0 \\ -1 & : \sum_i w_i x_i < 0 \end{cases}$$

A Hopfield network is a network of N such artificial neurons, which are fully connected. The connection weight from neuron j to neuron i is given by a number wij.The collection of all such numbers is represented by the weight matrix W, whose components are wij. Now we determine the weight matrix. Here we impose two conditions on the weight matrix:

Symmetry : wi j = wj i

No self connections : wi I = 0

IV. METHODOLOGY

Weather forecasting is a crucial application in meteorology.Weather is a continuous, data-intensive, multidimensional,dynamic process that makes weather forecasting a formidable challenge. The surface stations on the Earth are shown below in Fig 3.

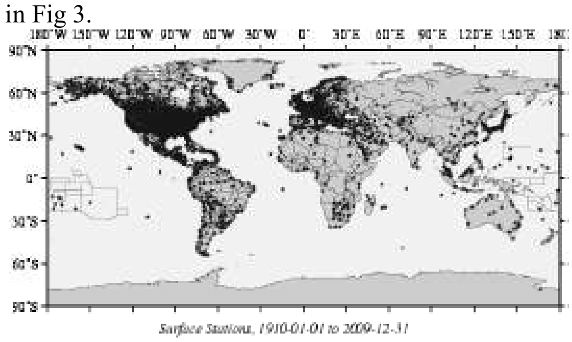

Fig 3: The surface stations on the Earth

After collecting whole data set, data cleaning is one of the important parts. The following preprocessing steps may be applied to the data to help improve the accuracy, efficiency, and scalability of the classification or prediction process.

Data cleaning: This refers to the preprocessing of data in order to remove or reduce noise (by applying smoothing techniques, for example) and the treatment of missing values (e.g., by replacing a missing value with the most commonly occurring value for that attribute, or with the most probable value based on statistics).

Relevance analysis: Many of the attributes in the data may be redundant attributes. Relevance analysis can be used to detect attributes that do not contribute to the classification or prediction task. Including such attributes may otherwise slow down, and possibly mislead, the learning step.

Data transformation and reduction: The data may be transformed by normalization, particularly when neural networks or methods involving distance measurements are used in the learning step. Normalization involves scaling all values for a given attribute so that they fall within a small specified range, such as -1.0 to 1.0, or 0.0 to 1.0.

The proposed algorithm is given below:

1. Initialize the earth (or globe) into equi-space regions(say 16 regions).

2. Each region is represented by a single node and thatnode has three input parameters Temperature, Wind Speed and Humidity respectively.

Available at: www.researchpublications.org

3. Connect the nodes with each other using links, i.e. build a mesh topological network structure. Links between any nodes are bi-directional as because Temperature, Wind Speed and Humidity can flow in both directions. Build a BPN for each pair of locations (including the self-loop). The input and output layer consists of 3 neurons where as the hidden layer has 5 neurons.

4. Each node I has an input Si and output Xi. It is very easy to represent Si as 3-tuple (T, WS, H). So Xi =1/1+e-S i where S i= W1WTXi1+ W2WwsXi2+ W3WHXi3 where W1, W2, W3 are scalars and Si is represented as a vector with Temperature, Wind Speed and Humidity are the components.

5. Initialize the scalars with some random values between -1.0 and 1.0. Build a Hopfield Network model with the help of training data set.

6. Now test the network model with test data set. The network must perform Temperature or Wind Speed orHumidity flow in order to establish equilibrium. This process will continue iteratively and in each iteration bias and weight values need to be updated until it converges.

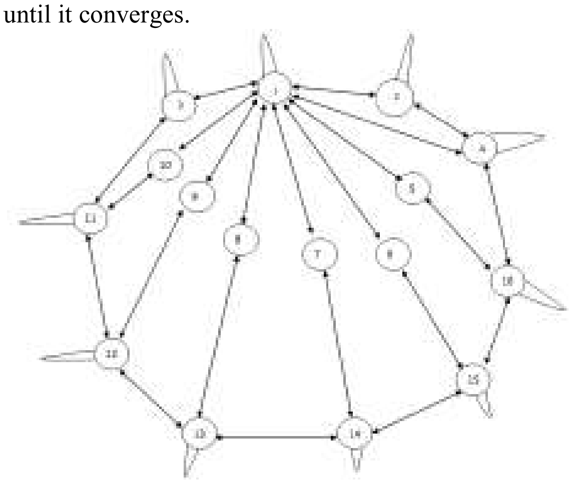

Fig 4: 16 Equi-space region of Earth (Globe)

.

VI. CONCLUSION

When tested on real adat set of 3 year period, the performance of the Back Propagation Network and Hopfield Network Model was satisfactory as there were not substantial number of errors in categorizing. Back Propagation Network and Hopfield Network based approach for weather forecasting is capable of yielding good results and can be considered as an alternative to traditional meteorological approaches. In this paper, Back Propagation Neural Network and Hopfield Network Model is used for predicting the atmospheric condition based on the training set provided to the neural network. This is the first approach of weather prediction which combines both Back Propagation Network (BPN) and Hopfield Network Model effectively.This approach is able to determine the non-linear relationship that exists between the historical data (temperature, wind speed,humidity, etc.,) supplied to the system during the training phase and on that basis, make a prediction of what the weather would be in future.

REFERENCES

[1] Xinghuo Yu, M. Onder Efe, and Okyay Kaynak, *A General Back propagation Algorithm for Feedforward Neural Networks Learning*. IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL.13, NO.1, JANUARY2002.

[2] R. Rojas: *Neural Networks*, Springer-Verlag, Berlin, 1996.

[3] Y.Radhika and M.Shashi, *Atmospheric Temperature Prediction using Support Vector Machines*, International Journal of Computer Theory and Engineering, Vol. 1, No. 1, April 2009 1793-8201.

[4] Han J., Kamber M. *Data Mining*: *Concepts & Techniques*, Morgan & Kaufmann, 2000.

[5] Mohsen Hayati, and Zahra Mohebi," Application of Artificial Neural Networks for Temperature Forecasting," World Academy of Science, Engineering and Technology 28 2007.

[6] Brian A. Smith, Ronald W. McClendon, and Gerrit Hoogenboom. *Improving Air Temperature Prediction with Artificial Neural Networks*. International Journal of Computational Intelligence 3;3 2007.

[7] Mike O'Neill, *Neural Network for Recognition of Handwritten Digits*, Standard Reference Data Program National Institute of Standards and Technology.

[8] http://www.wunderground.com/history

[9] Surajit Chattopadhyay, *Multilayered feed forward Artificial Neural Network model to predict the average summer-monsoon rainfall in India.* 2006.

[10] Dr. S. Santhosh Baboo and I.Kadar Shereef. *An Efficient Weather Forecasting System using Artificial Neural Network*. International Journal of Environmental Science and Development, Vol. 1, No. 4, October 2010 ISSN: 2010-0264.

[11]
http://www.comp.leeds.ac.uk/ai23/reading/Hopfield.pdf

[12] Gowri T. M. and Reddy V.V.C. 2008. *Load Forecasting by a Novel Technique using ANN*. ARPN Journal of Engineering and Applied Sciences. 3(2): 19-25.