Comparative Study of Different Weather Forecasting Models

Ramesh Medar
Dept. of Computer Science and Engineering
Gogte Institute of Technology
Belagavi , India
rameshcs.git@gmail.com

Prashant Y. Niranjan
Dept. of Computer Science and Engineering
Gogte Institute of Technology
Belagavi , India
pyniranjan@git.edu

Akshata B. Angadi
Dept. of Computer Science and Engineering
Gogte Institute of Technology
Belagavi , India
akshata_angadi@yahoo.co.in

Pushpa Tamase
Dept. of Information Science and Engineering
Gogte Institute of Technology
Belagavi , India
pstamase@git.edu

***Abstract*— Nowadays, Analysis of atmosphere conditions has become a vital process; where in prediction of state weather for a future time is made based on location. To improve the performance of system, collection of data related to current state of the atmosphere is important. Further using the understanding of atmospheric processes, determine or predict the atmosphere of future by applying ML techniques.**

The paper discusses application suitable Data Mining techniques, Regression approaches and Artificial Neural Network models to predict weather parameters. A study further ends by comparing different techniques for weather forecasting. The main moto of this study is to compare and identify a precise weather forecasting model. Weather prediction will be effective if, Input Data of years is taken instead of just 2 -3 days. Thus, if we train the system by considering huge data the performance will be more effective - "Better the training, better the result".

***Keywords*— Data Mining, Multiple Linear Regression, Autoregressive Integrated Moving Average, Artificial Neural Network.**

I. INTRODUCTION

The prediction of weather conditions can have significant impacts on various sectors of society in different parts of the country. Government and industries use the forecast data to protect life and property and to improve the efficiency of operations. The prime contribution of this paper is to compare the existing weather forecasting model and to select the precise model based on their predictive ability. The methodology consists of two stages for each study period data of weather parameters which are (1) Development of Predicting models and comparison of their aptitude. (2) Identification of precise and reliable weather forecasting model. These two components correspond to reducing the forecasting errors by relaxing certain assumptions of traditional forecasting techniques. Both the components are interlinked to each other.

978-1-5386-1887-5/17/$31.00 ©2017 IEEE

The different techniques used in prediction of weather are [3]:

1) Synoptic weather prediction: Its' a traditional approach that refers to the observation of different weather elements within the specific time of observation. Preparation of a meteorological center is done in order to keep track of records of atmosphere. Meteorological centers forms series of synoptic charts daily, which forms the very basic of weather forecasts. It engross vast collection and study of observational data obtained from thousands of weather stations.

2) Numerical weather prediction: [2] It is a sophisticated set of computer programs containing mathematical and physical equations/algorithms that describe the atmosphere and how it will change over time in order to produce meteorological forecasts. The Quality of Analysis depends on observations made and methods used. Observations related to land areas, atmosphere are all vital. If the initial stage of the weather is not completely known, the prediction will not be entirely accurate.

3) Statistical weather prediction: [24] It is a branch of objective weather forecasting. This prediction is commonly based on least square regression procedures. Statistical method use results of numerical models and numerical modelers recognize usefulness of properly applied statistical procedures. It uses the past records of weather data on the assumption that future will be a repetition of past weather. The main purpose is to identify the aspects of weather that are good indicators of the future events. Only the overall weather can be predicted in this way.

Data mining [4] is a process which finds useful patterns from large amount of data. Prediction is the most important technique of data mining which employs a set of pre-classified examples to develop a model that can classify the data and discover relationship between independent and dependent data.

Multiple Linear Regression is used when there is need to predict the value of one variable based on the value of two or more variables. An Autoregressive Integrated Moving

Average model predicts a value in a response time series as a linear combination of its own past values, past errors and past values of other time series. An Artificial Neural Network is a powerful data modelling tool that provides a methodology for solving many types of non-linear problems that are difficult to solve by traditional techniques. The advantage which ANN has over other weather forecasting method is that the ANN minimizes the error using various algorithms and gives us a predicted value which is nearly equal to the actual value.

II. LITERATURE SURVEY

There is a wide applicability of data mining techniques for weather prediction. E. G. Petre [5] presented a small application of CART decision tree algorithm for weather prediction. The data collected is registered over Hong Kong. The data is recorded between 2002 and 2005. The data used for creating the dataset includes parameters year, month, average pressure, relative humidity, clouds quantity, precipitation and average temperature. WEKA, open source data mining software, is used for the implementation of CART decision tree algorithm. The decision tree, results and statistical information about the data are used to generate the decision model for prediction of weather. The way the data is stored about past events is highlighted. The data transformation is required according to the decision tree algorithm in order to be used by WEKA efficiently for weather prediction. In case of multiple regression models Goulden (1962) found the relationship between monthly average of weather parameters and crop yield using multiple regression technique. Ramchandran (1967) made an analysis of the normal rainfall of 167 observatory station distributed over India and the neighbourhood country, using regression equation representing monthly and annually rainfall as a linear function of latitude, longitude and elevation above sea level. Iqbal et al. (2005) made the study on ARIMA to forecast the area and production of wheat in Pakistan. Further, suggesting that the scope of higher area and production lies in adequate availability of inputs, educating and training the farming community, soil conservation and reclamation, and especially the supportive government policies regarding wheat cultivation in the country. Zhou and Hu (2008) proposed a hybrid modelling and forecasting approach based on the grey and the Box–Jenkins autoregressive moving average (ARMA) models to forecast the gyro drift concluding that the hybrid method has a higher forecasting precision to the complex problems than the single method. Kal et al. (2010) developed a framework to determine the optimal inventory policy under the environment that the lead-time demand is generated by the ARIMA process. An Artificial Neural Network is a powerful data modelling tool that provides a methodology for solving many types of non-linear problems that are difficult to solve by traditional techniques. Neural Network makes very few assumptions as opposed to normality assumptions commonly found in statistical methods several works has been done and different artificial neural networks (ANN) models have been tested. Kaur[4] and Maqsood[5] describes a model that predicts the hourly temperature, wind speed and relative humidity 24 hour ahead. The authors have made a comparison of Multilayer Perceptron Networks (MLP), [20, 21] Elman Recurrent Neural Network (ERNN), Radial Basis Function Network (RBFN) and the Hopfield Model (HFM) and ensembles of these networks. MLP was trained by back propagation. The authors have suggested one hidden layer and 72 neurons for the MLP network and 2 hidden layers with 180 neurons for RBFN as the optimal architecture. The log-sigmoid is the activation function for the hidden layer unit of MLP network. RBFN has the best performance. The work described by Sanjay Mathur[8] focuses on maximum and minimum temperature forecasting and relative humidity prediction using time series analysis. The network model used is a Multilayer feed forward ANN with back propagation learning. A fully connected, feed forward 3 layer MLP network for temperature prediction is also presented by Santhosh Babu [9]. The error is said to be "very less". The set of input differs. Atmospheric pressure, atmospheric temperature, relative humidity, wind velocity and wind direction are chosen. The training is done by back propagation. Another short term temperature forecasting system is described by Hayati[10]. A three layer MLP network with 6 hidden neurons, a sigmoid transfer function for the hidden layer and a pure linear function for the output layer was found to yield the best performance.

III. FORECASTING MODELS

There are various weather forecasting models; however few are discussed in this paper.

A. Data Mining:

Data mining using clustering is a powerful way to analyze data and gives prediction.

Proposed Algorithm:

Step-1: Collect data (NO2, O3, CO2, SO2) in every one hour and store in the original database.

Step-2: Using the Convex Hull Technique, after every two hours data received is converted into structural data. All extreme data are also included. Then store it in the modified structural air pollution database.

Step-3: The database splits into four sub-databases on the basis of weather partition.

Step-4: Apply K-Means clustering of the structural data, where initial cluster centers are guessed using well known genetic algorithm (GA).

Step-5: Further, New Data inserted will use K-Means clustering.

Step-6: Finally, find the resulting clusters.

Step-7: Result obtained in diverse years (max 3-4 years) can be predicted by priority based protocol.

Step-8: Standard Threshold temperature value ranges can be considered to decide the probable weather condition.

For the models discussed below seven weather parameters are considered like Rainfall, maximum temperature, minimum temperature, Relative humidity at

7AM, Relative humidity at 2 PM, Pan evaporation, Bright sunshine.

B. Multiple (multivariate) linear regression(MLR):

This is the most popular technique for forecasting. It is used when there is a need to predict the value of one variable based on the value of two or more variables. The variable whose value is to be predicted is called dependent variable and the other variables are called independent variables. The equation fitted for weekly weather parameters using Multiple Linear Regression uses one parameter as dependent and other six as independent. The generalized form is,

$$Y = \beta 0 + \beta 1\ X\ 1 + \beta 2\ X\ 2 + \ldots\ldots\ldots\ldots + \beta 6\ X6 + \varepsilon \quad \text{--- (1)}$$

where:

β 0 = Intercept,

β i = regression coefficient of ith independent parameters,

(i = 1,2,…, 6),

ε = error term,

Xi = ith weather parameter.

Stepwise Regression analysis is used to predict the dependent weather parameter based on six independent weather parameters. The process starts with building of a simple regression model where extremely correlated one independent parameter is incorporated in the company of dependent parameter. Process repeats till the initial model includes all contributing parameters.

C. Autoregressive Integrated Moving Average Model(ARIMA):

[Wiki]The ARIMA Models are often referred to as Box-Jenkins models. In ARIMA (p,d,q) modelling, the first step is to check if the time series data is stationary or non-stationary. If it is non-stationary then it is transformed to stationary time series data by applying appropriate degree of differencing by selecting suitable value of 'd'. Further appropriate values of 'p' and 'q' are chosen. In the ARIMA model the Autoregressive (AR) term indicates that the evolving variable of interest is regressed on its own lagged values. The Integrated (I) term, indicates that the data values have been replaced with the difference between their values and previous values. The Moving Average (MA) term, indicates that the regression error is actually a linear combination of error terms whose values occurred contemporaneously and at various times in past.Non-seasonal ARIMA models are generally denoted as ARIMA(p,d,q) where parameters p, d, q are non-negative integers, 'p' is the order of autoregressive model, 'd' is the degree of differencing and 'q' is the order of moving average of moving model.

D. Artificial Neural Network Model(ANN):

Artificial Neural Network provides methodology for solving many types of non-linear problems that are difficult to solve by traditional techniques. A typical feed forward artificial neural Network with back propagation network should have at least three layers- an input layer, a hidden layer, and an output layer. Appropriate selection of number of hidden layers and the number of neurons in each of them needs experimentation. The main aim is to create and train a network that can predict the individual weather components for e.g. maximum temperature, minimum temperature, wind speed etc. for a particular station and a particular day given the weather for the previous day (target data) and the historical 10 year data of that particular day(input data). There are many neural network models, but the basic structure involves a system of layered, interconnected nodes and neurons are presented in figure 4.1. The nodes are arranged to form an input layer, with neurons in each hidden layer connected to all neurons in neighbouring layers. The input layer supplies data to the hidden layer and does not contain activation or transfer functions. A typical feed-forward network might use a dot-product activation function that, for each neuron Bj (j = 1, 2, …..,n) in the hidden layer, is computed as:

$$Bj = \sum_{i=1}^{m} w\ ij\ Ai + w0j\ A0 \quad \text{---- (2)}$$

with input nodes Ai (i = 1, 2,…., m) and weights Wij between nodes Ai and neurons Bj. The bias node (Ao) typically has a constant input of 1, with a matching weight Woj. A similar calculation is made for each neuron Ck (k = 1, 2, …., o) in the output layer (o = 1 for the example in figure 4.1), using weights Wjk between neurons Bj and Ck (with Wok and Bo for the bias). Each neuron value is subsequently passed through a transfer function, which may be linear or nonlinear (Zurada, 1992).

E. Hybrid Model:

In recent times, the concept of combined model instead of single time series model is being prepared for prediction purpose. The combination of MLR with ARIMA and MLR with ANN is proposed in the present study. The hybrid of multiple linear regression with ARIMA and ANN techniques to analyse the weekly weather parameters of all the seven parameter studied and included in the comparative study to identify the best precise weather forecasting model.

Hybrid Model of Multiple Linear Regression and Autoregressive Integrated Moving Average (MLR_ARIMA) [8, 22, 23] The composition of a multiple linear regression with autoregressive integrated moving average model is proposed to develop a new hybrid model. It is assumed that the predictive performance improves by integrating two single models. For this purpose the selected significantly contributed parameters obtained through stepwise regression analysis are used to develop the MLR_ARIMA model and their performance is compared with all other models.

Hybrid Model of Multiple Linear Regression and Artificial Neural Network (MLR_ANN)

It has been observed in the current researches that a single model may not be sufficient to identify all the characteristics of the time series data. The hybrid models decompose a time series into linear and non-linear form and prove to be better approach in comparison to single model. The hybrid model of multiple linear regression with neural network approach is proposed to yield more accurate results. Similar to previous model, the significantly contributed parameters selected through stepwise regression analysis in multiple linear regression model are used to develop the hybrid MLR_ANN model.

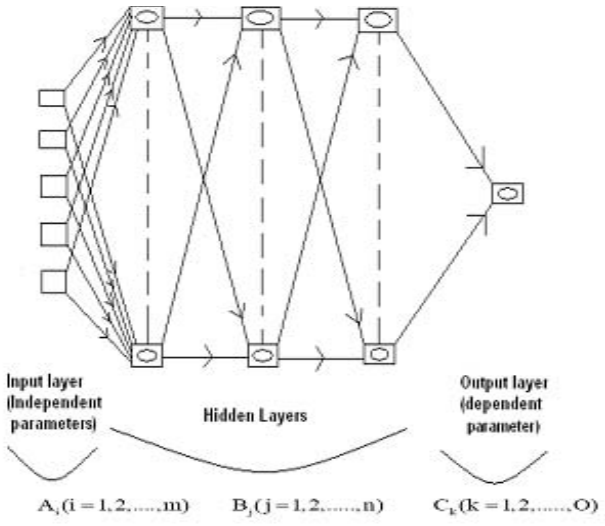

Fig. 1. An (m x n x o) artificial neural network structure, sowing a multilayer perceptron.

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 criteria involve:

A linear scale plot of the predicted and observed weather parameters for training and testing data sets for all the models. The numerical performance criterion involves:
Mean error(BIAS), Mean absolute error(MAE), Root mean square error(RMSE), Prediction error(PE) and Correlation Coefficient(r).

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 observed and predicted values. The recital of weather forecasting models had been evaluated on the basis of Mat lab 7.0.1 version, students' academic SAS version and Microsoft Excel.

IV. COMPARISION OF WEATHER FORECASTING MODELS

The role of statistical techniques for providing reliable predictions of weather parameters is considered to be most important in the field of metrology all over the world. These predictions influence the agricultural as well as the industrial strategies. The comparison among the above discussed models shows the trend, based on previous years but the actual values were fluctuating. The proposed hybrid MLR_ANN model was observed as precise model in comparison to MLR, ARIMA, ANN and hybrid MLR_ARIMA models. In view of the fact that, all the prediction models are consistent but the finest model is the lone having least mean absolute error and root mean square error, prediction error and high correlation coefficient, as observed in hybrid MLR_ANN model. It was observed that the ANN model is also precise weather forecasting model as compared to MLR and ARIMA models which coincides with the previous findings. At last, the study reveal that hybrid MLR_ANN model maintained can be used as an appropriate forecasting interest to estimate the weather parameters. The following table shows a comparison of various models developed so far for weather forecasting.

Table 1. Comparison of weather forecasting models

AUTHOR	APPLICATION	TECHNIQUE	ATTRIBUTES	DATASET	REMARKS
1. Zhi-liang wang , Hui-hua sheng (2010)[11]	Prediction of annual Rainfall	Generalized Regression neural network	Rainfall	-	Simulation results and accuracy for annual rainfall is better than Back Propagation networks.
2.Kannan, prabhakaran and ramachandran (2012)[12]	Prediction of annual rainfall	Multiple Linear Regression	-	Five years data	Does not show accuracy but shows an approximate value.
3. Jesada, Kok and Chung(2012) [13]	Monthly Rainfall Prediction	Fuzzy inference system	-	-	FIS is good alternative method to predict accurately

4. Valmik and Meshram (2013)[14]	Prediction of rainfall	Bayesian model	Station level pressure, mean sea level pressure, temperature, relative humidity, vapour pressure, wind speed and rainfall.	-	Bayesian model for rainfall prediction provides good accuracy
5. Y.Radika And M. Shashi(2009)[15]	Prediction of Temperature	Support Vector Regression (SVR)	-	-	SVM performs better than MLP trained with back propagation algorithm
6.Ch.Jyosthna Devi , B.Syam Prasad Reddy, K.Vagdhan Kumar,b.Musala Reddy,n.Raja(2012)[16]	Prediction of Temperature	Back propagation algorithm	-	-	Predicts future temperature with less error.
7.Somia A. Asklany,Khaled Elhelow(2013)[17]	Prediction of rainfall	Fuzzy rule base and fuzzy logic	Relative humidity, total cloud cover, wind direction, temperature and surface pressure		The output percentage approximately around eighty percent is considered as success forecast.
8. Ankita Joshi, Bhagyashri et al(2015)[18]		Decision tree algorithm	Max temperature, rainfall, evaporation and wind speed	-	The performance of this algorithm would be compared with the standard performance metrics.
9. Ahmad Shahi, Rodziah Binti Atan, Md. Nasir Sulaiman(2009)[19]	Weather Forecasting	Fuzzy c- mean clustering and type-2 fuzzy logic			This proposed method is very effective on uncertainty data.

V. CONCLUSION AND FUTURE SCOPE

In this research work we took up a study to identify the precise and reliable weather forecasting model through comparison of several existing and proposed models. The study reveals that Hybrid MLR_ANN model is an appropriate forecasting interest to estimate the weather parameters, in disparity to the multiple linear regression, ARIMA, ANN and hybrid MLR_ARIMA models.

By being more precise in what is uncertain, the forecasts will become more reliable. A collection of models can be used instead of just one and by doing so we can better cover the uncertainties in the forecast.

REFERENCES

[1] Divya Chauhan and Jawahar Thakur "Data Mining Techniques for Weather Prediction: A Review ,August 2014.

[2] Heikki Pohjola,"The Vaisala Weather Radar WRM200 – Catching Storms and Insects at Long Ranges", Application Note, 2011.

[3] S. Kotsiantis and, "Using Data Mining Techniques for Estimating Minimum, Maximum and Average Daily Temperature Values", World Academy of Science, Engineering and Technology, 450-454, 2007.

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

[5] Elia Georgiana Petre "A Decision Tree for Weather Prediction", Buletinul, Vol. LXI No. 1, 77-82, 2009.

[6] Amanpreet Kaur, J K Sharma, and Sunil Agrawal.,"Artificial neural networks in forecasting maximum and minimum relative humidity", International Journal of Computer Science and Network Security, 11(5):197-199, May 2011.

[7] Imran Maqsood, Muhammad Riaz Khan, and Ajith Abraham, "An ensemble of neural networks for weather forecasting: , Neural Computing & Applications, pp. 112–122, 2004.

[8] Rosmina Bustami, Nabil Bessaih, Charles Bong, and Suhaila Suhaili. "Artificial neural network for precipitation and water level predictions of bedup river" IAENG International Journal of Computer Science, 34(2):228-233, 2007.

[9] Dr. S. Santhosh Baboo and I.Kadar Shereef. "An efficient weather forecasting system using Artificial neural network", International Journal of Environmental Science and Development, 1(4):321-326, 2010.

[10] Mohsen Hayati and Zahra Mohebi. "Application of artificial neural networks for temperatureforecasting" World Academy of Science, Engineering and Technology, 28:275-279, 2007.

[11] Zhi-liang Wang and Hui-hua Sheng, "Rainfall Prediction Using Generalized Regression Neural Network: Case study Zhengzhou", International Conference on computational and Information Sciences, 2010.

[12] M.Kannan, S.Prabhakaran and P.Ramachandran, "Rainfall Forecasting Using Data Mining Technique", International Journal of Engineering and Technology, Vol.2 (6), 397-401, 2010.

[13] Jesada Kajornrit, Kok Wai Wong and Chun Che Fung, "Rainfall Prediction in the Northeast Region of Thailand Using Modular Fuzzy Inference System", World Congress on Computational Intelligece, 10-15, June 2012.

[14] Valmik B Nikam and B.B.Meshram, "Modeling Rainfall Prediction Using Data Mining Method", Fifth International Conference on Computational Intelligence, Modeling and Simulation, 2013.

[15] Y.Radika And M. Shashi, "Atmospheric Temperature Prediction Using Support Vector Machines," International Journal Of Computer Theory And Engineering,vol.1,no.1,Apr 2009.

[16] Ch.Jyosthna Devi , B.Syam Prasad Reddy, K.Vagdhan Kumar,b.Musala Reddy,Raja "ANN Approach For Weather Prediction Using Back Propagation", International Journal Of Engineering Trends And Technology-Volume3, 2012.

[17] Somia A. Asklany , Khaled Elhelow , I.K. Youssef and M. Abd El-wahab ," Rainfall events prediction using rule-based fuzzy inference

system",Elsevier, Atmospheric Research 101 (2011) 228–236.

[18] Ankita Joshi, Bhagyashri Kamble, Vaibhavi Joshi, Komal Kajale, Nutan Dhange, "Weather Forecasting And Climate Changing Using Data Mining Application", International Journal Of Advanced Research In Computer And Communication Engineering,Vol. 4, Issue 3, March 2015.

[19] Ahmad Shahi, Rodziah Binti Atan, Md. Nasir Sulaiman "An Effective Fuzzy C-mean And Type-2 Fuzzy Logic For Weather Forecasting" Journal Of Theoretical And Applied Information Technology 2009.

[20] Kumar Abhisekh, M.P.Singh, Saswata Ghosh,Abhisekh Anand "Weather Forecasting Model using Artificial Neural Network", Elseveir, Procedia Technology, Vol 4 , pp 311-318, 2012.

[21] Pituk Bunnoon, Kusumal Chalermyanont, Chusak Limsakul, " Mid Term Load Forecasting of the Country Using Statistical Methodology: Case Study in Thailand", 2009 International Conference on Signal Processing Systems, published in IEEE, 2009.

[22] Mourycy Ciupak , Bogdan Ozga-Zielinski, Jon Adamowski, John Quilty, Bahaah Khalil "The application of Dynamic Linear Bayesian Models in hydrological forecasting: Varying Coefficient Regression and Discount Weighted Regression", Journal of Hydrology, Volume 530, Nov, pp. 762-784,2015.

https://doi.org/10.1016/j.jhydrol.2015.10.023.

[23] Sarıca, B., Eğrioğlu, E. & Aşıkgil, B. "A new hybrid method for time series forecasting: AR–ANFIS", Neural Computing & Applications, Springer, 2016. DOI: 10.1007/s00521-016-2475-5

[24] Harry R. Glahn, "Sattistical Weather Forcasting", ECMWF organization report, pp. 263-310.