

REVIEW ON WEATHER FORECASTING USING ANN AND DATA MINING TECHNIQUES

Nowsheena yousf¹

¹(Scholar, computer science and engineering, jamia hamdard university, New delhi)

Abstract:

Weather is a practice to forecast the conditions of atmosphere for coming times for particularly given regions. Accurate weather forecasting is important in agricultural sectors, industrial sectors and many other sectors. Because of Non linearity in climatic data prediction of weather becomes very difficult task. However ANN is being put forward more to predict weather because of its ability to approximate large class of functions.. The paper presents survey of different methodologies utilized by various researchers to use ANN for weather prediction.

Keywords — Artificial neural network (ANN), Back Propagation, Weather Forecasting, Feed Forward Network

I. INTRODUCTION

Weather can be defined as the hustle and bustle of precipitation and day to day actions. Weather is one of the important factor of our lives and it can be referred as the one that can't be controlled where as weather more often controls how and what people do that is where we live, what we wear and even what we eat. Weather is composed of the parameters like wind, visibility, precipitation (rain, snow), temperature, cloud, pressure and humidity. Weather phenomena mostly take place in the lower part of atmosphere that is troposphere [1]. Weather primarily occurs due to the air pressure, temperature and moisture differences between one to another. These differences are caused seen because of the angle of sun that varies with the latitude. The polar and tropical areas vary widely in temperature and this wide temperature contrast led to the atmospheric circulations. Depending on the period forecasts remain valid is categorized into four types that are given below:

- **Short range forecasts:**-It is the prediction of weather conditions on daily basis that approximately predicts for one to seven days. Predictions are made on the basis of satellite

images, maps weather charts. For short range weather forecasts persistent and continuity methods are used. In daily life short range weather forecasting plays an important role in transportation and for fishermen.

- **Medium range forecasts:**-Prediction of weather conditions around one to four weeks in advance. Medium range weather forecasts are made by the calculations done by considering average of past and present weather conditions. Medium range forecasts are made by the global atmospheric models based on deterministic methods. Agro Advisor Service provided to the farmers by National centre for medium range weather forecasting in India so that the farmers can plan their agricultural activities accordingly.
- **Long range forecasts:**-Broaden the range for forecasting and cover periods between one month and a year in advance. Long range forecasts are made from fortnight to season of a year in advance [1]. Long range weather forecasting doesn't contain detailed information and has minimum accuracy. It has utility during heat and cold waves and during droughts. Forecasting is made by the ensemble models.
- **Hazardous weather forecasting:**-Weather hazards are threatening weather occurrences that are dangerous to life and properties. Weather hazards can be like tornadoes, lightning, thunderstorm, hails, flash flooding etc. National weather service has termed hazardous weather outlook a type of statement that is issued in order to give information of hazardous or severe weather events coming up within a week.

Since weather is multidimensional, continuous, chaotic, data intensive, dynamic process making prediction difficult task. In this paper the overall literature is presented using different techniques.

WEATHER FORECASTING USING DATA MINING TECHNIQUES

Data mining techniques can be considered as efficient techniques for finding patterns hidden in data. A number of models are available in literature based on classification and clustering.

F.dell et al. 2000 proposed an approach for identification of varying wind speed [2]. The classification techniques were used on the meteorological data to obtain prediction.

James N.K.Liu et al. 2001 proposed improved Naive Bayesian classification by doing variations in the weather condition data[3]. The number of models of genetic algorithm and Improved Naive Bayesian classification was compared and an appreciable amount of accuracy was determined.

Zohreh Nazeri et al. 2002 used the data provided by National Center for Atmosphere Research in terms of Airline Service Quality Performance (ASQP), Enhanced Traffic Management System (EIMS) and National Convective Weather Forecast (NCWF) to represent severe weather impacts on National Airspace System performance by using data mining applications[4]. Their work was divided into three phases first phase involved data preparations, second was related to feature extraction and third phase included data mining. Feature extraction consists of image segmentation, weather feature extraction, air traffic feature extraction and representation conversion phases. The authors used segmentation algorithm for identifying severe weather areas. Correlation analysis was used to calculate weather and air traffic features with performance. K means clustering algorithm was also used. Performance measure was divided by authors into three parts for the purpose of classifying whether the severe weather impact on NAS performance is bad, medium or good. Data mining classification techniques were used for this purpose.

Sanjay Chakarborty et al. 2011 proposed a methodology for time series forecasting through clustering. Incremental K- means clustering was used for this methodology[5]. Air pollution data was collected from "West Bengal air pollution control board". Four air pollution elements were

present in data i.e CO₂, SO₂, oxides of nitrogen, repairable particulate matter.

Kavita.P et al. 2012 used weather data set of Bangladesh [6]. To present a model for predicting the weather Using K- means clustering by analyzing temperature at atmospheric pressures with 400 hpa, 500 hpa, and 700 hpa.

Badhiye S.S.et al. 2012 provided an approach using clustering techniques for future prediction of Temperature and humidity [7]. The clustering analysis was used to partition data by taking the similarity of objects into consideration. K-Nearest Neighbor algorithm was used. Absolute distance and Euclidean distance measuring was used to obtain distances between datasets in K- Nearest Neighbor algorithm.

Valmik B et al. 2013 proposed a model taking parameters like vapor, humidity, wind pressure and wind speed into consideration to predict weather data using classification techniques [8]. Relatively good accuracy was obtained by correlating the above parameters.

Menhat.W et al. 2013 proposed that K-nearest neighbor classification can be used to estimate wind speed and also used three parameters air temperature, atmospheric pressure and wind direction to analyze the wind speed [9].

WEATHER FORECASTING USING ANN

Vamsidhar et al. 2010 proposed an ANN model for prediction of rainfall in India using BPNN [10]. As ANN can easily capture complex relationships in linear and non linear systems and there is no need for assumptions as it was required in traditional approaches. Humidity, dew points and pressure were taken as parameters for predicting the rainfall. The training phase of neural network required two-third of the data and one-third of data was used for testing purposes. This paper presented that ANN using back propagation showed accuracy in predicting rainfall. 99.79% of accuracy was obtained in training phase and 94.28% of accuracy was obtained during testing phase. Thus this method was used for predicting rainfall in India.

Shareef et al. 2011 presented weather forecasting techniques using BPNN. Forecasts about temperature were made by collecting quantitative data about conditions of atmosphere. Neural

networks can capture complex relationships among data. Neural networks comprises of a wide range of learning algorithms. one among them is back propagation algorithm that can easily estimate large class of functions. This paper uses modified levenberg marquardt algorithm for more prediction accuracy. The number of oscillations in learning procedure is also reduced because of the modified levenberg marquardt algorithm. For this parameter inputs were taken as atmospheric pressure, atmospheric temperature, relative humidity, wind velocity and wind direction. The dataset was evaluated for Madras Minambak, India [11]. Data of 2010 was taken for this purpose. The available dataset was divided into four different seasons I.e winter, pre-monsoon, south-west monsoon and north-east monsoon. MSE, minimum and maximum error, and prediction accuracy were chosen as performance parameters. In this research the proposed algorithm improved convergence and damped the oscillations.

Mekanik et al. 2012 used ANN in order to find non linear relationship in the weather data [12] Researchers found that Australian rainfall also get affected by the key modes of complex climate variables. As it is known that rainfall is complex atmospheric phenomenon which can't be handled by using the linear methods to know its characteristics. Researchers made few trials to establish the combined effect of these indices on rainfall in order to have knowledge about forecasting system. It was observed ANN proved to have greater correlations to forecast rainfall. Three different stations were used for this purpose Hirsham, Melbourne and orbost and showed 99%, 98% and 43% increase in model correlation respectively.

Litta et al. 2013 provided a research article about ANN for predicting pre monsoon thunderstorms. Thunderstorm known to all as global weather phenomena can occur at any time and at any place in whole world [13] it is a powerful convective activity. Although thunderstorm doesn't remain for large time but can be seen as very harmful to human life and properties. In India thunderstorms are oftenly seen during pre monsoon season I.e march to may. Thunderstorm forecasting has been shown as the most effortful tasks because of the

presence of nonlinearity in their dynamics and physics and small spatial and temporal extension. ANN has been proved as the satisfactory for the prediction of thunderstorms because it can easily capture complex relationships between data. In this research ANN was used for the prediction of severe thunderstorms over Kolkata. This research involved the estimation of temperature and relative humidity. The module for designing ANN comprises of number of procedures. First one involves collection of data. T he input parameters provided to ANN were mean sea level pressure, relative humidity and wind speed that were obtained from IMD Kolkata. The second step for designing of ANN involved pre-processing of data i.e. finding the missing values in data and provides data in normalized form. The missing values were obtained by doing averaging of neighbouring values. Data normalization is required because convergence problems occur if we input original data to neural network. Thus data is normalized by transforming all values in range of -1 and 1 by dividing difference of actual and minimum values by the difference of max and min values. Data is denormalized at end in order to get desired output. Two different ANN's were formed for predicting temperature and relative humidity respectively. The third step involved training of ANN and 67% of data was used for training of data. 33% was used for testing purposes. The frequently used MLPN is back propagation learning algorithm. Gradient descent one of the variant of BP algorithm suffers the problems in finding global minimum error function as it gets stuck in local minima. Gradient descent suffers long training session problems. For this research six different learning algorithms (Step, Momentum, DBD, QKP, conjugate gradient,LM) performance was evaluated for the surface temperature and relative humidity. Also comparisons for different advanced predictions were made for 1h, 3h, 6h, 12h and 24h ahead for the thunderstorm days. First case occurred on may 5 which was a severe thunderstorm that lasted for only few minutes with the maximum speed of 0.12Kmph. The second case was severe thunderstorms that occur on may 11, 2009 with max. Speed of 84Kmph. The third case reported on may 15, 2009 with the maximum wind speed 68.4Kmph.

It was observed that LM proved to be more efficient than other learning algorithm which showed best accuracy in predicting temperature and relative humidity. Also statistical analysis was performed and LM showed less RMSE, more correlation coefficient, more percent correct and less mean absolute error. Also it was observed that 1h prediction in advance proved to be more accurate than 3h and 24h showed good accuracy whereas 6h and 12h ahead prediction failed in capturing the entire pattern.

Saba et al. 2017 presented an article on weather forecasting based on hybrid neural model (MLP and RBF) to enhance the accuracy in weather forecasting [14]. Sample data was taken of Saudi Arabia with the input parameters average dew point, minimum temperature, maximum temperature, mean temperature, average relative moistness, precipitation, normal wind speed, high wind speed and average cloudiness. The output parameters were rainy and dry weathers. Correlation coefficients, RMSE and scatter index are the important measurements that are taken for accurate forecasting. 50% of data was used for training purpose, 30% for validation and 20% for testing purposes. In this paper MLP was proved to be better in accuracy than RBF at individual level.

CONCLUSION

In ancient times number of people tried their hands on forecasting then numerical methods came into existence. Even soft computing was also used for forecasting of weather. Data mining using classification and clustering techniques were used for forecasting of weather. However ANN proved to be best choice for prediction of weather. This paper has explained some of the works that were performed using data mining and Artificial neural network techniques.

REFERENCES

1. <https://en.m.wikipedia.org>
2. F.Dell Acqua and P.Gamba, "A simple Model Approach to the problem of meteorological object Tracking", Volume: 1, Issue No: 7803-6359, PP: 2152-2154, Italy, 2000.
3. James N.K.Liu, Bavy N.L.Li, and Tharam S.Dillon, "An Improved Naïve Bayesian Classifier Technique Coupled with a Novel Input Solution Method", IEEE Transactions on

Systems, Man and Cybernetics-Part C: Applications and Reviews, Vol.31, No.2, PP:249-256, May 2001.

4. Zohreh Nazari and Jianping Zhang " Mining Aviation Data to Understand Impacts of Severe Weather on Airspace System Performance", Proceedings of the International Conference on Information Technology: Coding and Computing[ITCC'OL], Volume-1, ISSN: 7695-1506, PP: 1-6, 2002.
5. Sanjay Chakraborty, Prof.N.K. Nagwani, Lopamudra dey, "Weather Forecasting using Incremental K-means Clustering", International Conference in High Performance Architecture and Grid Computing, Vol.169, Part-2, PP:338-341,2011.
6. Kavita Pabreja , Research scholar, " Clustering Technique to Interpret Numerical Weather Prediction Output Products for Forecast of Cloudburst", International Journal of Computer Science and Information Technologies (IJCSIT), vol-3(1), ISSN:0975-9646, PP. 2996-2999, 2012.
7. Badhiye S.S, Dr.Chatur P.N, Wakode B.V, "Temperature and Humidity Data Analysis for Future Value Prediction using clustering Technique: An Approach", International Journal of Emerging Technology and Advanced Engineering, ISSN:2250-2459, volume-2, Issue-1, PP- 88-91, January 2012.
8. Valmik B Nikam and B.B. Meshram, "Modeling Rainfall Prediction using DataMining Method", Fifth International Conference on Computational Intelligence, Modeling and Simulation, Issue No: 2166-8531, PP:132136, 2013.
9. MEHMET YESILLBUDAK, SEREF SAGIROGLU and ILHAMI COLAK, "a new approach to very short term wind speed prediction using K-nearest neighbor classification", energy conversion and management, volume- 69, PP: 77-86, 2013
- 10.Enireddy.Vamsidhar,K.V.S.R.P.Varma P.Sankara Rao, Ravikanth satapati, "Prediction of Rainfall Using Back propagation Neural Network Model", International Journal on Computer Science and Engineering, Volume 02, No. 04,pp 1119-1121,2010.
11. Kadar Shereef, Dr. S. Santhosh Baboo, "A New Weather Forecasting Technique using Back Propagation Neural Network with Modified Levenberg-Marquardt Algorithm for Learning", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 6, No 2, November 2011.
12. F. Mekanik and M. A. Imteaz, "A Multivariate Artificial Neural Network Approach for Rainfall Forecasting: Case Study of Victoria", Australia, Proceedings of the World Congress on Engineering and Computer Science 2012, Vol I WCECS 2012, October 2426, 2012, San Francisco, USA
- 13.A.J.Litta ,Sumam Mary Idicula, and U.C.Mohanty, "Artificial Neural Network Model in Prediction of Meteorological Parameters during Premonsoon Thunderstorms", Hindawi Publishing Corporation InternationalJournalofAtmosphericSciences Volume2013.
14. Tanzila Saba, Amjad Rehman, Jarallah S. AlGhamd "Weather forecasting based on hybrid neural model", Springerlink.com, 15 February 2017.