

Data Mining Techniques to Predict Weather: A Survey

M. Viswambari, Dr. R. Anbu Selvi

Department of Computer Science, Bishop Heber College (Autonomous),
Tiruchirappalli, Tamilnadu 620017, India

Abstract

Weather forecasting is one of the major applications in current technology. Data mining uses various technologies to forecast weather, predict rainfall, wind pressure, humidity, etc. Classification in data mining differentiates the parameters to view the clear information. The prediction of weather must be accurate and also the weather should be forecasted earlier at least a month before which will be helpful for many applications like agriculture, military, etc. This paper surveys the various techniques implemented in data mining to predict weather.

Keywords: *Data mining, Classification, Back Propagation, Weather forecasting.*

1. Introduction

1.1 Weather Forecasting

Weather forecasts are made by collecting quantitative data about the current state of the atmosphere on a given place and using scientific understanding of atmospheric processes to project how the atmosphere will evolve on that place. Weather warnings are important forecasts because they are used to protect life and property. Forecast based on temperature and prediction are important to agriculture and therefore to traders within commodity markets. Temperature forecasts are used by utility companies to estimate demand over coming days. Outdoor activities are severely curtailed by heavy rain, snow and the wind chill, forecasts can be used to plan activities around these events and to plan ahead and survive them.

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 increases. The use of ensembles and model consensus help narrow the error and pick the most likely outcome.

1.2 Data Mining Techniques

Data mining focuses on the discovery of unknown properties in the data. Classification is the problem of identifying to which of a set of categories a new observation belongs, on the basis of a training set of data containing observations whose category membership is known. The individual observations are analyzed into a set of quantifiable properties, known as various explanatory variables, features, etc. probabilistic classifiers can be more effectively incorporated into larger machine-learning tasks, in a way that partially or completely avoids the problem of error propagation.

Back propagation, is the "backward propagation of errors", is a common method of training artificial neural networks. From a desired output, the network learns from many inputs. It is a supervised learning method, and is a generalization of the delta rule. It requires a dataset of the desired output for many inputs, making up the training set. It is most useful for feed-forward Back propagation requires that the activation function used by the nodes be differentiable. The goal of any supervised learning algorithm is to find a function that best maps a set of inputs to its correct output. The back propagation algorithm for calculating a gradient has been rediscovered a number of times, and is a special case of a more general technique called automatic differentiation in the reverse accumulation mode.

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes. The data collection component of research is common to all fields of study including physical and social sciences, humanities, business, etc. While methods vary by discipline, the emphasis on ensuring accurate and honest collection remains the same. The goal for all data collection is to capture quality evidence that then translates to rich data analysis and allows the building of a convincing and credible answer to questions that have been posed.

2. Related Works

Sanjay D.Sawaitul et al., [1] focused on the information about weather and are observed and stored. The recorded parameters are used to forecast weather. If there can be a change in any one of the recorded parameters like wind direction, wind speed, temperature, rainfall, humidity, then the upcoming climatic condition can be predicted using artificial neural networks, back propagation techniques. The increase in signal range will work in large areas too.

Meghali A.kalyankar et al., [2] discussed that in a particular place the past data is used to predict the climate using knowledge discovery. History of data should be maintained for weather prediction. The system must be adaptive so that, if there is any sudden change in the data, it can easily acceptable. For this dynamic data mining methods are used. These methods must match the sudden change in weather conditions. Radar, lidar, satellites are used to record data.

Folorunsho olaiya et al., [3] explored data mining techniques and found weather forecasting parameters like temperature, rainfall, evaporation, wind speed are maximum . The performance can be improved by using Neuro-Fuzzy models to predict the change in weather.

Zahoor Jan et al., [4] the past data like rainfall, wind speed, dew point, temperature is used to predict weather using k nearest neighbor algorithm. It generates accurate result in advance in prediction of weather. The results are in Boolean attributes and numeric values. The changes can be recognized by using the patterns. Pattern recognition can be used.

A.R.Chaudhari et al., [5] using data mining techniques meteorological data has been predicted and decision had made. The Weather forecasting parameter's relationships are found using data mining techniques. Since meteorological data are vast and time constrained, it not only need to modify by traditional data mining . but also can be modified using some other techniques.

Elia Georgiana Petre et al., [6] by using decision trees the stored data in past are used to predict the upcoming climate. By using all parameters the prediction can be improved and perfect. And also the limit for prediction will not limit itself.

Godfrey C. Onwubolu et al., [7] employed self organising data mining technique called enhanced Group Method of Data Handling (e-GMDH) to predict and forecast weather. e- GMDH works efficiently when compared with older data mining techniques. Graphical

User Interface (GUI) which is partly placed in the algorithm must be updated to include the current functionalities.

Sarah N.Kohail et al., [8] described that data from a particular place is extracted by knowledge discovery. For this, the system must be adaptive so that if there is any change in weather it can also be acceptable. The changeable data mining methods like data cleaning and data processing are used.

R.Nagalakshmi et al., [9] MLPN, ERNN, HFM, RBFN algorithms are compared by choosing correct predictors by finding mean, median, range, standard deviation. Improving prediction by applying statistical data, weather can be forecasted effectively.

Seema Mahajan et al., [10] explored on correlation analysis which is applied to predict seasons at particular place. Correlation is found between the rainfall at particular place and the factors that affect rainfall in future. Wind stress is the main parameter used in this paper. The technique applied is accurate but still it remains challenging when it is applied for wider region.

3. Conclusion

The weather is predicted or forecasted by using various techniques especially back propagation, neural network, data collection. Still there are some issues and challenges in which better implement of data mining technique should be implemented in the field of weather forecasting.

References

- [1] Sanjay D. Sawaitul, Prof. K.P. Wagh, Dr. P.N. Chatur, "Classification and Prediction of Future Weather by using Back Propagation Algorithm-An Approach", International Journal of Emerging Technology and Advanced Engineering, Vol. 2, Issue 1, January 2012, pp. 110-113.
- [2] Meghali A. Kalyankar, Prof. S. J. Alaspurkar, "Data Mining Technique to Analuse the Metrological Data", International Journal of Advance Research in Computer science and Software Engineering", Vol. 3, Issue. 2, February 2013, pp. 114-118.
- [3] Folorunsho Olaiya, Adesesan Barnabas Adeyemo, "Application of Data Mining Technique in Weather Prediction and Climate Chandes Studies", I. J. Information Engineering and Electronic Business, MECS press, 2012, pp. 51-59.
- [4] Zahoor Jan, M. Abrar, Shariq Bashir, Anwar M. Mirza, "Seasonal to Inter-annual Climate Prediction Using Data Mining KNN TYechnique", Springer, 2008, pp. 40-51.
- [5] A. R. Chaudhari, D. P. Rana, R. G. Mehta, Data Mining with Meteorological Data", International Journal of Advanced

Computer Research, Vol. 3, No. 3, Issue 11, September 2013, pp. 25-29

- [6] Elia Georgiana Petre, “A Decision Tree for Weather Prediction”, Vol. LXI, No. 1/2009, 2009, pp. 77-82.
- [7] Godfrey C. Onwubolu, Petr Buryan, Sitaram Garimella, Visagaperuman Ramachandran, Viti Buadromo, Ajith Abraham, “Self-Organizing Data Mining for Weather Forecasting”, IADIS European Conference Data Mining, 2007, pp. 81-88.
- [8] Sarah N. Kohail, Alaa M. El-Halees, “Implementation of Data Mining Techniques for Meteorological Data Analysis”, International Journal of Information and Communication Technology Research, Vol. 1, No. 3, July 2011, pp. 96-99.
- [9] R. Nagalakshmi, M. Usha, RM. A. N. Ramanathan, “Application of Data Mining Techniques in Maximum Temperature Forecasting: A Comprehensive Literature Review”, International Journal of Advance Research in Computer Science and Management Studies, Special Issue, December 2013, pp. 1-9.
- [10] Seema Mahajan, Dr. S. K. Vij, “Modeling and Prediction of Rainfall Data using Data Mining”, International Journal of Engineering Science and Technology, Vol. 3, No. 7, July 2011, pp. 6799-6804.