The Weather Forecast Using Data Mining Research Based on Cloud Computing.

Zhanjie Wang and A. B. M. Mazharul Mujib

Electronics Information & Electrical Engineering, Dalian University of Technology, Liaoning, China
Email: wangzhi@dlut.edu.cn, mazharul.mujib@gmail.com

Abstract. Weather forecasting has been an important application in meteorology and one of the most scientifically and technologically challenging problem around the world. In my study, we have analyzed the use of data mining techniques in forecasting weather. This paper proposes a modern method to develop a service oriented architecture for the weather information systems which forecast weather using these data mining techniques. This can be carried out by using Artificial Neural Network and Decision tree Algorithms and meteorological data collected in Specific time. Algorithm has presented the best results to generate classification rules for the mean weather variables. The results showed that these data mining techniques can be enough for weather forecasting.

1. Introduction
Weather Prediction is the application of science and technology to predict atmospheric conditions ahead of time for a particular region. Prediction is one of the basic goals of Data Mining. Data Mining is to dig out knowledge and rules, which are hidden and unknown. User may be interested in or has potential value for decision-making from the large amounts of data. Such potential knowledge and rules can reveal the laws between the data. There are many kinds of technical methods of data mining, which mainly include: association rule mining algorithm, decision tree classification algorithm, clustering algorithm and time series mining algorithm, etc. [1]. How to store, manage and use these massive meteorological data, discover and understand the law and knowledge of the data, to contribute to weather forecasting completely and effectively has attracted more and more Data Mining researcher’s attention[2]. This article constructs the Weather Forecasting platform, using data mining for meteorological forecast and the forecast results are analyzed.

2. Weather Forecasting
Weather forecasting plays a significant role in meteorology [3]. Weather forecasting has remained a formidable challenge because of its data intensive and frenzied nature. Generally, two methods are used to forecast weather: a) the empirical approach and b) the dynamical approach. The first approach is based on the occurrence of analogues and often referred to as analogue forecasting. This approach is useful in predicting local scale weather if recorded cases are plentiful. The second case is based upon equations and forward simulations of the atmosphere and often referred to as computer modeling. Most weather prediction systems use a combination of both of these techniques.
This framework as a service (FAAS) has selected seven common forecasting methods. These are Regression (R), Logistic Regression, Time Series, Artificial Neural Network, Random Forest, Support Vector Machine and Multivariate Adaptive Regression Splines (MARS). For instance, Regression may encounter them collinearity among variables. Logistic Regression could only deal with the dataset where the dependent variable is nominal.

3. Weather Prediction Architecture
Artificial Neural Networks (ANN) and Decision Trees (DT) were used to analyze meteorological data, gathered in-order to develop classification rules for the Application of Data Mining Techniques in Weather Prediction. Artificial Neural Networks (ANN) has received special attentions among different forecasting methods in recent years [4, 5]. Main reason for the popularity of ANN is its capability of supervised learning from complex relations using non-linear functions [6]. This algorithm combines with both of the time series and regresional approaches. Weather parameters [7] over the study period use available historical data for the prediction of future weather conditions. The targets for the prediction are those weather changes that affect our daily life e.g. changes in minimum and maximum temperature, rainfall, evaporation and wind speed. These techniques are often more powerful, flexible, and efficient for exploratory analysis than the statistical techniques. The most commonly used techniques in data mining are: artificial neural networks, logistic regression, discriminant analysis and decision trees. By this model, temperature (T), rainfall (R) and wind (W) speed can easily be predicted. Now Prediction method that only single parameter for example some researchers [8], [9] used wind speed and other researchers used wind Power for prediction [10], [11]. We have provided more accurate prediction model data and it has more parameters and efficiency. (Figure.2)

There are three basic elements of a neuron model. Figure.3 shows the basic elements of neuron model with the help of a perceptron model, which are, (i) a set of synapses, connecting links, each of which is considered by a weight/strength of its own (ii) an adder, for summing the input signals, weighted by respective neuron’s synapses (iii) an activation function, for limiting the amplitude of neuron’s output. A typical input-output relation can be expressed as shown in Equation 1.
Figure 3 Model of a perceptron

\[
\text{net}_j = \sum_{i=1}^{n} W_{ij}X_i + b_j
\]

\[
O_i = f_i(\text{net}_i)
\]

Where \(X\) = inputs to node in input, \(W\) = weight between input node and hidden node, \(b\) = bias at node, \(\text{net}\) = adder, \(f\) = activation function.

The type of transfer/activation function affects the size of the steps taken in weight and space [12]. ANN’s architecture needs determination of number of the connecting weights and the way in which the information flows through this network is carried out via the number of layers, nodes number in each layer, and their connectivity. The output nodes numbers are fixed, according to the estimated quantities. The input nodes numbers are dependent on the existing problem under consideration, and the modeler’s choice to utilize knowledge of domain. The neurons in the hidden layer are enhanced gradually, and the network performance in the form an error is examined.

4. Weather Forecast In Cloud Computing

Before Cloud computing has improved the efficiency of data storage, delivery, and dissemination across multiple platforms and applications, allowing easier collaboration and data sharing, including data processing and distribution systems that disseminate key weather forecasting, severe weather warning, and climate information. Data mining techniques and forecasting applications are very much needed in the cloud computing paradigm. In this study, data mining in Cloud Computing allows weather forecasting and data storage, with assurance of efficient, reliable and secure services for their users. The implementation of data mining techniques through Cloud computing will allow the users to retrieve meaningful information from virtually integrated data warehouse that will reduce the costs of infrastructure and storage.

5. Implement And Process

5.1. Forecasting Model

Our prediction model is based on a Multi-Layer Perceptron ANN. The central, innovative point of the method is the construction of the data set and the choice of input data unlike the other works [13][14]. Along with the historical weather data of the site, chosen for the prediction, nearby sites are also taken into account for the evolution of the weather fronts. Therefore, for the identification of the site where to make the wind forecast (called WFS), I have chosen four cardinal directions and the four intermediate directions, two points for each direction, one more far away (1000 km) and second one halfway (500 km) from the place identified. For sites at maximum distance, weather data referred to the instant \(t-2\), for sites at minimum distance, weather data refer to the instant \(t-1\). Because, weather data was referred to instant \(t0\) for the WFS. The implemented MLP, has one hidden layer using transfer function, \(F(x)\), the logistic function:

\[
\int F(x)dx = \frac{1}{(1+e^{-x})}.
\]
The MLP network is trained through the back-propagation learning algorithm. This training algorithm is to minimize the cost function (E) of the network, expressed as mean square error amongst all the outputs of the network and the respective desired values defined in the training set. The mean square error is calculated by using the following expression:

5.2. Input Data
The data used in this study for each point are temperature, rainfall and wind. The data used are the daily average values. The input Dataset variable are (T), (R) and (W) used in this Model and this model is similar to Fig.2. We have collected the data of Dalian weather forecasting and input the data in the model for computing and got the results. X=(1, 2, 3) used are site-dependent. The prediction of weather forecasting, with a resolution of one day, requires the knowledge of the weather conditions in previous time instants to the one chosen for the prediction. Earlier, Mesoscale Meteorological Modeling [15], Numerical Weather Prediction Models [16] and other methods [17] were used for prediction. Unlike these traditional methods, we have improved this model, using different time instants for weather forecasting.

5.3. Simulation Results
In this section we have introduce a case study; the test site is a real weather forecast located in Dalian Liaoning Province, China. In this study of data analysis for Dalian weather forecasting comes from the Dalian Meteorological Bureau in the last five years. In our research, we have set up a new model with more parameters and finished the simulation later. Our provided model has shown more accurate results than the old model. In this prediction model, we have used ANN and FAAS. The figures of merit used for analyzing the performance. The forecasting data are compared with the real data recorded in the site. In this work, we have used last five years data analysis, the variables including in the highest temperature, the lowest temperature and wind speed. Variables of analysis are described in following chart:

![Figure 4 Weather Forecasting Temperature for Last Five Years (2011-2015) Highest and Lowest Value](image1)

![Figure 5 Results of wind speed prediction Value in Last Five Years. (2011-2015)](image2)
The results have shown that concerning Wind Speed and Temperature highest and lowest Value displayed the simulation results (figure 4 and 5) in Intervals of last five years. While using more input data values, the execution time can be as longer as expected.

6. Vi. Discussion and Conclusions
In this Study Forecasting Models algorithm are used for classifying weather parameters such as maximum temperature, minimum temperature, and wind speed in terms of the month and years. The implementation of data mining approach to solve the wind forecasting problems for wind farm production, in particular, for predicting wind speed. The data mining prediction algorithm-ARIMA time series prediction algorithm is also integrated into the system. The platform has the ability of mass storage of meteorological data, efficient query and analysis, weather forecasting and other functions. In this study we also adapted the method of Artificial Neural Networks, it can detect the relationships between the input variables and generate outputs based on the observed patterns inherent in the data without any need for programming or developing complex equations to model these relationships. An artificial neural network (multi-layer perceptron) was applied and several simulations have been conducted for comparison purposes. ANN’s can detect the relationships between weather parameters and use these for future prediction. Weather conditions are important to climatic change studies because the variation in weather conditions in term of temperature and wind speed can be studied using these data mining techniques. ANNs are implemented, in order to compare their effectiveness in changing the network topology and the training mode. The results obtained from real data are based on time series of meteorological data provided by the Dalian Meteorological Bureau. The test cases pointed out that the proposed approach gives a very interesting performance of the implemented network and shows good performance in term of MSE. For future perspective, there is still significant potential for improvement in weather forecasting by using ANN model, through introducing climate change and global warming variables, in order to forecast more realistic weather parameters.

7. Acknowledgment
This research was supported by the National Natural Science Foundation of China (Grant No. 61672126) and (Grant No. 51579041). We would like to thank the anonymous reviewers for their attentive reading and for their constructive comments that have helped to further strengthen this paper.

8. References
[1] M. Zhang, “Application of Data Mining Technology in Digital Library”, Journal of Computers, vol. 6, no. 4, (2011) April, pp. 761-768.
[2] Z. Danping and D. Jin, “The Data Mining of the Human Resources Data Warehouse in University Based on Association Rule”, Journal of Computers, vol. 6, no. 1, (2011) January, pp. 139-146.
[3] Introduction to Data Mining and Knowledge Discovery, Third Edition, Two Crowds Corporation, http://www.twocrows.com/introdm.pdf, accessed on 12 April 2009.
[4] L. M. Saini and M. K. Soni, “Artificial neural network-based peak load forecasting using conjugate gradient methods,” IEEE Transactions on Power Systems, vol. 12, no. 3, pp. 907–912, . 2002.
[5] S. Fan, C. X. Mao, and L. N. Chen, “Peak load forecasting using the self-organizing map.” in Advances in Neural Network-ISNN 2005. New York: Springer-Verlag, 2005, pt. III, pp. 640–649.
[6] Kourentzes, N., “Intermittent demand forecasts with neural networks”, International Journal of Production Economics, Volume 143, Number 1, pages 198-206, 2013.
[7] Elia G. P., 2009, “A Decision Tree for Weather Prediction”, Universitatea Petrol-Gaze din Ploiești, Bd. București 39, Ploiești, Catedra de Informatică, Vol. LXI, No. 1.
[8] A. R. Finamore; V. Calderaro; V. Galdi; A. Piccolo; G. Conio; S. Grasso, “A day-ahead wind speed forecasting using data-mining model– a feed-forward NN algorithm”, IEEE International Conference on Renewable Energy Research and Applications, 2015, pp. 1230-1235.
[9] E. Erdem, J. Shi, “ARMA based approaches for forecasting the tuple of wind speed and direction”, Applied Energy 88, ELSEVIER, 2011, pp. 1405–1414.
[10] N. Chen, Z. Qian, I. T. Nabney, and X. Meng, “Wind Power Forecasts Using Gaussian Processes and Numerical Weather Prediction,” IEEE Transaction on Power Systems, vol. 29, no. 2, 2014.
[11] EWEA Wind in power: European Statistics. 2013.
[12] Martin T. H., Howard B. D, Mark B., 2002, Neural Network Design, Shanghai: Thomson Asia PTE LTD and China Machine Press.
[13] N. Chen, Z. Qian, I. T. Nabney, and X. Meng, “Wind Power Forecasts Using Gaussian Processes and Numerical Weather Prediction,” IEEE Transaction on Power Systems, vol. 29, no. 2, 2014.
[14] R. R. B. de Aquino, H. T. V. Gouveia, M. M. S. Lira, A. A. Ferreira, O.N. Neto, M. A. Carvalho Jr., “Wind Forecasting and Wind Power Generation: Looking for the Best Model Based on Artificial Intelligence”, IEEE World Congress on Computational Intelligence, 2012.
[15] Roger A. Pielke Sr., “Mesoscale Meteorological Modelling”, International Geophysics Series, Volume 98, oct, 2013.
[16] David J. Stensrud, “Parameterization schemes – Keys To Understanding Numerical Weather Prediction Models”, Cambridge University Press, 2007.
[17] Renewables – Global Status Report- REN21, 2015.