



RESEARCH ARTICLE

**SURVEY PAPER ON COMPARISON OF VARIOUS WEATHER FORECASTING
TECHNIQUE FOR PREDICTIVE ANALYSES**

1,*Pallavi Gulati, 2Ulfat Jan and 3Dr. Anurag Dixit

^{1,2}M.tech student, GU, Greater Noida

³SCSE, Galgotias University

ARTICLE INFO

Article History:

Received 27th February, 2017
Received in revised form
29th March, 2017
Accepted 07th April, 2017
Published online 23rd May, 2017

Key words:

Datamining,
Prediction,
Atmospheric Conditions,
Weather Forecasting.

Copyright©2017, Pallavi Gulati et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Weather Forecasting is one of the emerging applications for Data mining technology. It is used to predict the atmospheric conditions based on humidity, temperature, rainfall, wind etc. This paper focus on comparing various data mining technique which were used previously for predicting the weather conditions. Predicting the weather is helpful in preparing for the best and the worst condition of the climate. Datamining plays a very important role in determining the conditions relating to the weather. Various technique for predicting weather include Regression, Artificial neural network, frequent pattern growth algorithm, decision tree and back propagation algorithm .The paper provide comparison of various techniques for predicting the weather.

Citation: Pallavi Gulati, Ulfat Jan and Dr. Anurag Dixit, 2017. "Survey paper on comparison of various Weather Forecasting Technique for Predictive Analyses", *International Journal of Current Research*, 9, (05), 50276-50279.

INTRODUCTION

Weather forecasting is consider as one of the most challenging problem which is face by the world from the last decades. Weather forecasting command in predicting how the current state of the atmosphere will change with the change of climate. [5] For making the forecasting accurate a user must have the data collected from the past. A user must understand the processes which are occurring in the atmosphere which are needed to generate the current weather conditions which is later on considered with the past data .Thus weather forecasting is considered as one of the important features for determining various day to day activities. Data mining is also defined as the process which is used to extract the useful information from large amount of incomplete data which is used for practical applications. Throughout these decade an attempt was made for forecasting the weather based on change in climate thus weather forecasting is consider as one of the interesting domains.

Weather forecasts are of following type [7]:-

- **Now casting**:-is defined as the method in which information about the current weather forecasting up to few hours ahead is given.

- **Short range forecast**:-is defined as the method in which information about forecasting of weather is predicted up to 3 days. This method has a duration upto 3 days. This type of forecasting is basically based on the latest weather charts.
- **Medium range forecast**:-This method of weather forecasting contain information which is less accurate then short range forecasting .They have a duration from 4 to 10 days.
- **Long range forecast**:-There is no definition for this type of forecasting but it contain information for more than 10 days which can be range from a month to a seasonal forecast.

Weather contain following parameters [17] like air temperature, atmospheric pressure, humidity, precipitation, solar radiation and wind. Based on this parameters event like rainfall, fog can be measured. There are several techniques [4] used in data mining including association [16], regression and so on. Among all classification technique are mostly used

Tools

There are several number of tools used in Data mining techniques include Classification, Prediction, Association, Clustering, Visualization.

**Corresponding author: Pallavi Gulati,*
M.tech student, GU, Greater Noida.

Some of the tools are

Rapid Miner [8], is an open source tool with software platform that implements the environments like analytics and mining. This tool also validates the process for prediction automatically.

Weka: is a graphical user interface that act as a data mining tool. This tool acquire the techniques as classification, association, visualization and so on.

Orange Canvas, is used to manage the unsupervised data. it is also one of the type of data mining tool. It include methods like classification, regression and type of regression. As this method contain regression thus it is used to evaluate the data in the form of matrices whicht help in predictions.

Methodologies Used

Basic steps for implementing various algorithms are:-

Data Collection:-refers to the process of gathering the information in an organized form, which is then used to evaluate the outcomes.

Data preprocessing:-refers to the process of transforming the raw data into an understandable form. It is one of the proven methods for resolving problems like incompleteness and inconsistency.

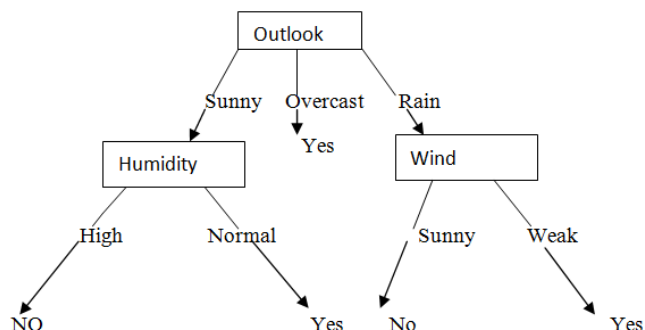
Data Transformation:-refers to the transformation of data into an suitable form

Applying various classification algorithms:-This step involves a method for classifying a data and based on this classification an appropriate algorithm is applied.

Predicting the data:-based on classification algorithm suitable prediction can be made.

This are the basic steps involve in each algorithm. Methodologies are given below

Decision Tree



Diagrammatically representation of a Decision Tree in weather forecasting

Decision Tree [2] is one of the method of data mining that uses a tree like graph which is used to make further predictions. There are different kind of decision tree algorithm such as C4.5,CART (classification and regression tree), CHAID (chi-squared automatic interaction detection) and many more.

In Decision Tree making attribute selection measures are used to select the attribute that divide the data instances into different classes. Therefore Decision Tree Algorithm is one of the method which is used to analyze the past data and can predict the future data.

In decision tree making, attributes which are used for selecting the procedures are used to select the attribute that divide the different instances of data into different classes. C4.5 uses gain ratio as an attribute for selecting the procedures whereas CART uses gini index which can be consider as a binary split for each attribute.

Naïve Bayesian Classifications

Bayesian Classification is based on Bayes Theorem. [9] The Basic idea of Bayes Theorem is that an event can be predicted based on the evidences that can be observed. Bayesian Classifier [11] is basically the Statistical Classifiers that can be used to predict the probabilities. Basic Formula for Bayes

Theorem is

$$P(H/X) = P(X/H) P(H)/P(X)$$

Where P(X) is constant.

Bayesian Classifier work in following manner

- Firstly G is taken as a training set of tuples or data sets with their associated class labels, tuple $A=(a_1,a_2...a_n)$, which is used to depict n measurements made on the tuples from n attributes, respectively, $D_1,D_2,...D_n$.
- If there were m classes, $B_1,B_2...B_m$. For the given tuple A, the classifier will predict that A belongs to a class having the higher posterior probability. Thus Naïve Bayesian Classifier predict the tuple A belong to a class B_1 if it satisfy the following condition

$$P(B_1/A) > P(B_k/A)$$

- Thus, the value of $P(B_1/A)$ should be maximize. This is known as maximum posteriori hypothesis. Thus by Bayes Theorem it can be depicted that

$$P(B_1/A) = p(A/B_1)p(B_1)/p(A)$$

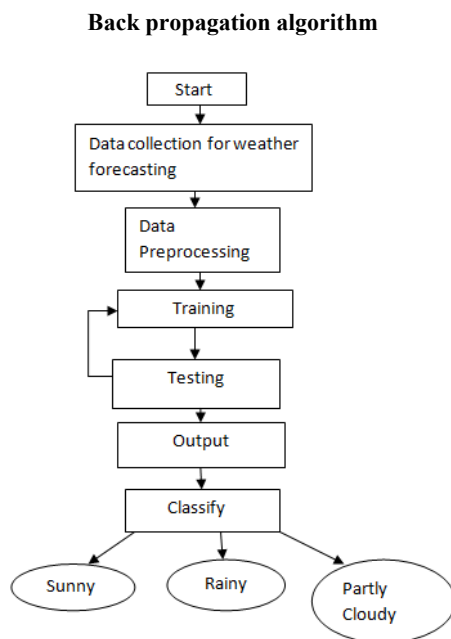
- In the above equation $p(A)$ is considered as constant for all classes, therefore $p(A/B_1)p(B_1)$ needs to be maximized. if the probability is not known, then it can be assumed that all the classes are equal which means $P(B_1)=P(B_2)=...P(B_m)$ therefore the value of $P(A/B_1)$ needs to be maximized.
- Prior probabilities of classes can be estimated by $P(B_1)=|B_{1,g}|/|G|$, here $|B_{1,g}|$ is the number of training instances of class B_1 in G.
- Thus the naïve assumption for class conditional independence is made. This denotes that the value of the attributes are conditionally independent of one another.
- In the below mention equation A_j refers to the value of attributes D_j . Therefore

$$P(A/B_1) = P(A_1/B_1) \times P(A_2/B_1) \times P(A_3/B_1) \dots \times P(A_n/B_1)$$

- Thus Bayesian Classification algorithm uses past dataset and current dataset in order to predict [14] the values.

Back Propagation

Back Propagation [1] is based on neural network. Neural Network is refers to the process of connecting the input or the output which has the weight associated with them. By adjusting their weight it can able to predict the future. Back Propagation [6] is based on supervised learning technique. The weight of each connection in back propagation can be adjusted in order to reduce the error value and it can be repeated until the value of an error is relatively very small. With the help of this, [13] the value of parameters can be predicted by changing the value of any one parameter



Flowchart [21] for weather classification using Back Propagation Algorithm In Back Propagation a network can be trained by changing the weight of a network in order to get the output.

Back Propagation include the following steps

- Firstly, the initialization of network can be done by setting its weight to a random number.
- After setting the weight an input pattern is applied in order to get the output.
- Calculate the error. The value of an error can be changed in order to minimize it.
- Repeat the above steps in order to reduce the value of an error.

Thus by the changing the value of any one parameter prediction can be made.

Comparison

Thus weather forecasting can be done by any one of the above mention method like Decision tree, Bayesian classification and Back propagation algorithm. The Decision Tree considers the past data and generates a tree that provides the attributes which can be considered for future prediction. Bayesian classifier is used to classify the data based on the evidences that can be used to predict the future outcome.

In Back Propagation weather forecasting can be predicted by changing the value of any one parameter.

Conclusion

Weather Forecasting is having a predominant importance in our day to day life. This survey paper includes some of the important techniques which can be used for weather forecasting. Thus Data Mining is one of the most emerging technologies which can be used to predict the real time applications.

REFERENCES

1. Classification and Prediction of Future Weather by using Back Propagation Algorithm-An Approach by Sanjay D. Sawaitul, Prof. K. P. Wagh, Dr.P. N. Chatur Government College of Engineering, Amravati, Maharashtra, India.
2. Decision Tree for the Weather Forecasting by Rajesh Kumar, Ph.D Asst. Prof., Dept. of ECS Dronacharya College of Engineering, Gurgaon, India.
3. Rainfall Prediction using data mining by Sangari.R.S, Dr.M.Balamurugan School of CSE, Bharathidasan University, Trichy, India.
4. Data Mining Techniques for Weather Prediction by Divya Chauhan, Jawahar Thakur Dept. of Computer Science, Himachal Pradesh University Shimla , India.
5. Towards a Self-Configurable Weather Research and Forecasting System by Khalid Saleem, S. Masoud Sadjadi, Shu-Ching Chen ,School of Computing and Information Sciences, Florida International University, Miami FL.
6. An Efficient Weather Forecasting System using Artificial Neural Network by Dr. S. Santhosh Baboo and I.Kadar Shereef.
7. Convective weather forecast accuracy analysis at center and sector levels by yao wang and banavar sridhar, nasa ames research center, moffett field, California.
8. Artificial Neural Networks' Application in Weather Forecasting – Using RapidMiner by A Geetha, G M Nasira, Mother Teresa Women's University, Kodaikanal.
9. Alleviating Naive Bayes Attribute Independence Assumption by Attribute Weighting by Nayyar A. Zaidi, Jesus Cerquides, Mark J. Carman Geoffrey I. Webb, Monash University VIC 3800, Australia.
10. Locally Weighted Naive Bayes by Eibe Frank, Mark Hall, and Bernhard P fahringer Department of Computer Science, University of Waikato Hamilton, New Zealand.
11. A Tutorial on Naive Bayes Classification by Choochart Haruechaiyasak.
12. Prediction of rainfall using Data mining technique over Assam by Pinky saikia dutta, hitesh tahbilder, Guwahati University ,Gauhati, Assam, India.
13. Prediction of Severe Thunderstorms applying Neural Network using RSRW Data by Himadri Chakrabarty, Sonia Bhattacharya Panihati Mahavidyalaya Barasat State University Kolkata, India.
14. Heart Disease Prediction System using Naive Bayes by Dhanashree S. Medhekar, Mayur P. Bote, Shruti D. Deshmukh.
15. Air Temperature Forecasting using Radial Basis Functional Artificial Neural Networks I.El-feghi, Zakaria Suliman zubi, A. Abozgay, University of Tripoli, Tripoli-Libya.
16. Efficient Mining of Intertransaction Association Rules, Anthony K.H. Tung, Member, IEEE, Hongjun Lu,

- Member, IEEE, Jiawei Han, Member, IEEE, and Ling Feng, Member, IEEE.
17. Accuweather.com, <http://www.accuweather.com>
18. <http://thenewstack.io/six-of-the-best-open-source-data-mining-tools>.
19. Neural Networks and Back Propagation Algorithm Mirza Cilimkovic, Institute of Technology Blanchardstown Blanchardstown Road North Dublin 15 Ireland.
20. Temperature and Humidity Data Analysis for Future Value Prediction using Clustering Technique: An Approach by Badhiye S. S., Dr. Chatur P. N. Wakode B. V. Government College of Engineering, Amravati, Maharashtra, India.
21. <http://www.ijsrp.org/research-paper-1212/ijsrp-p1211.pdf>
