RESEARCH ARTICLE OPEN ACCESS

Data Mining Techniques Using To Weather Prediction

P.Kalaiyarasi [1], Mrs.A.Kalaiselvi [2]
Research Scholar [1], Assistant Professor [2]
Navarasam Arts and Science College for Women, Arachaur
India

ABSTRACT

Agriculture is the strength of our Indian economy. Farmer only depends upon monsoon to be their cultivation. The good crop productivity needs good soil, fertilizer and also good climate. Weather forecasting is the very important requirement of the each farmer. Due to the sudden changes in climate/weather, The people are suffered economically and physically. Weather prediction is one of the challenging problems in current state. The main motivation of this paper to predict the weather using various data mining techniques. Such as classification, clustering, decision tree and also neural networks. Weather related information is also called the meteorological data. In this paper the most commonly used weather parameters are rainfall, wind speed, temperature and cold.

Keywords:- Data Mining, Weather Prediction

I. INTRODUCTION

Meteorological department worldwide is putting a great effort in the research areas of weather prediction. Since, India is an agriculture based country most of the people are dependent on the weather conditions. Large amount of India population depend on monsoon. It is always a matter of concern for the scientists to exactly forecast the weather conditions. Weather forecast includes prediction of rain, fog, winds, clouds, lightening, storm etc. One the biggest challenge in weather forecasting is its unpredictable and dynamic climate data sets, which can frequently change according to global climatic changes. Many techniques have been applied and suggested out of which data mining is considered as the most feasible approach towards Weather forecasting [1]. Data Mining has the capability to mine hidden patterns, relations and provide verification/validation of data sets based on certain input conditions.

II. BACKGROUND STUDY

A. Method for Weather Prediction

In Weather prediction, there are three methods available [4].

- Synoptic Weather prediction
- Numerical Weather prediction
- Statistical Weather prediction

1).Synoptic Weather Prediction

In metrological center, they provide synoptic chart for every day. Within a Specific time, different weather parameters are observed. Different data collection and the study of observational data observed from thousands of weather stations.

2) Numerical Weather Prediction

The capability of computer to predict the weather is known as Numerical Weather Prediction. If the initial stage of the weather is not totally known, the prediction will not be completely accurate.

3) Statistical Weather Prediction

Pure Statistical methods are used to predict the weather, along with numerical methods are also available. It used past records of weather parameters to predict the future occurrence.

B. Data Mining

Data Mining [5] is the process of discovering interesting patterns and knowledge from large amount of data. The data sources can include database, data warehouse, the web and other information repositories. Data mining is also known as knowledge discovery process. It is also an iterative sequence data.

Data Mining tasks can be classified in to two types: Descriptive and Predicative.

- Descriptive Mining tasks characterize properties of the data in a target data set.
- Predictive Mining tasks perform induction on the current data in order to make prediction

The most commonly used Data mining technique are classification, clustering, Decision Trees which is shown in Table I. Figure 1 shows the flow of machine learning sequence in data mining techniques.

C. Classification

Classification is the process of finding a model that describes and distinguishes data classes or concepts for the purpose of being able to use the model to predict the class of objects whose class label is unknown[6].

D. Clustering

Clustering analyses data objects without consulting a known class label. The unsupervised learning technique of clustering is a useful method for ascertaining trends and patterns in data, when there are no pre-defined classes [6].

Table I. Data Mining Techniques and Algorithm

Data Mining Techniques	Algorithm
classification	Back propagation, KNN and genetic algorithm
clustering	K-mean and k-medoid
Decision tree	ID3,C4.5 AND CART

E. Decision Tree

Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and ID3 [7].

F. Artificial Neural Network (ANN)

An Artificial Neural Network [8] is information processing that can be inspired by the nature of human nervous system that is brain process information. The common ANN applications through a learning process are pattern recognition (or) data classification.

There are three basic elements of neuron

1. Synapses connecting links obtained weight
2. Added input signals weighted by synapses
3. An activation function for limiting the amplitude of the output neuron

III. DATA MINING TECHNIQUES IN WEATHER FORECASTING

A. Data Collection

The data used for this work was collected from meteorologist's centre. The case data covered the period of 2012 to 2015. The following procedures were adopted at this stage of the research: Data Cleaning, Data Selection, Data Transformation and Data Mining.

B. Data Cleaning

In this stage, a consistent format for the data model was developed which is search missing data, finding duplicated data, and weeding out of bad data. Finally system cleaned data were transformed into a format suitable for data mining.

C. Data Selection

At this stage, data relevant to the analysis like decision tree was decided on and retrieved from the dataset. The Meteorological dataset had ten attributes in that were using two attributes for future prediction. Due to the nature of the Cloud Form data where all the values are the same and the high percentage of missing values in the sunshine data both were not used in the analysis.

D. Data Transformation

"This is also known as data consolidation". It is the stage in which the selected data is transformed into forms appropriate for data mining. The data file was saved in Commas Separated Value (CVS) file format and the datasets were normalized to reduce the effect of scaling on the data.

E. Data Mining Stage

The data mining stage was divided into three phases. At each phase all the algorithms were used to analyse the meteorological datasets. The testing method adopted for this research was percentage split that train on a percentage of the dataset, cross validate on it and test on the remaining percentage. There after interesting patterns representing knowledge were identified.

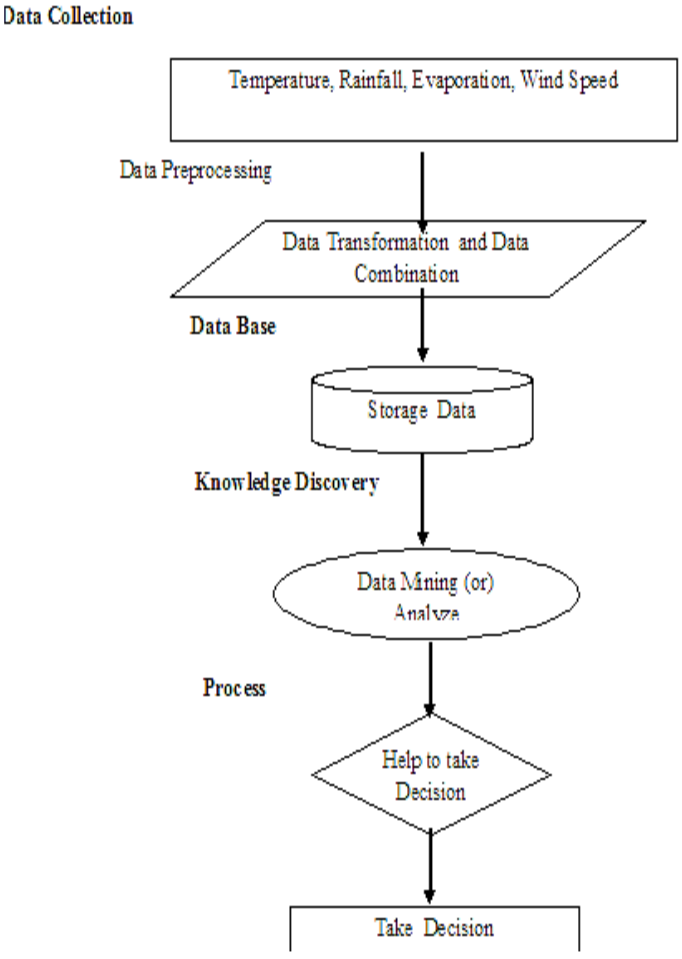

Figure 1. Technique Sequence Machine Learning

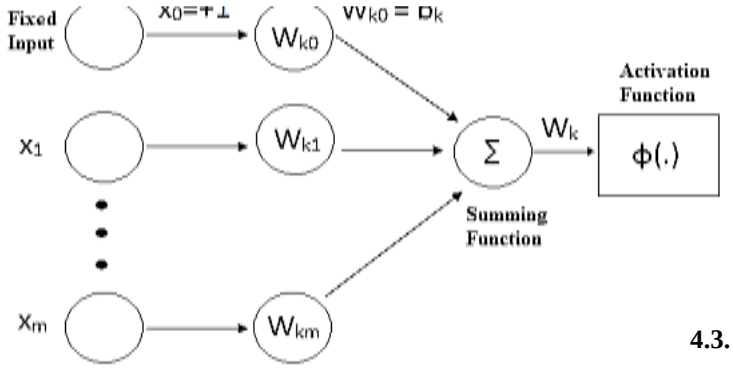

Figure 2. Non Linear Model ANN

IV. LITERATURE SURVEY

Data mining, a branch of computer science, is the process of extracting patterns from large datasets by combining methods from statistics and artificial intelligence with database management. Data mining is seen as an increasingly important tool by modern business to transform data into business intelligence giving an informational advantage. It is currently used in a wide range of profiling practices, such as marketing, surveillance, fraud detection, and scientific discovery.

The related terms data dredging, data fishing and data snooping refer to the use of data mining techniques to sample portions of the larger population data set that are (or may be) too small for reliable statistical inferences to be made about the validity of any patterns discovered. These techniques can, however, be used in the creation of new hypotheses to test against the larger data populations.

If there is much irrelevant and redundant information present or noisy and unreliable data, then knowledge discovery during the training phase is more difficult. Data preparation and filtering steps can take considerable amount of processing time. Data pre-processing includes cleaning, normalization, transformation, feature extraction and selection, etc. The product of data preprocessing is the final training set.

4.1.K-means Algorithm:

K-means clustering is a data mining/machine learning algorithm used to cluster observations into groups of related observations without any prior knowledge of those relationships.

The k-means algorithm is one of the simplest clustering techniques and it is commonly used in medical imaging, biometrics and related fields.

4.2. Working of k means algorithm

1. Place K points into the space represented by the objects that are being clustered.
2. These points represent initial group centroids. Assign each object to the group that has the closest centroid.
3. When all objects have been assigned, recalculate the positions of the K centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

4.3. Decision tree:

Decision tree learning uses a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value. More descriptive names for such tree models Computer Science & Information Technology (CS & IT) 91are classification trees or regression trees. In these tree structures, leaves represent class labelsand branches represent conjunctions of features that lead to those class labels.

In data mining, a decision tree describes data but not decisions; rather the resulting classification tree can be an input for decision making. J48 are the improved versions of C4.5 algorithms or can be called as optimized implementation of the C4.5. The output of J48 is the Decision tree. A Decision tree is similar to the tree structure having root node, intermediate nodes and leaf node. Each node in the tree consist a decision and that decision leads to our result. Decision tree divide the input space of a data set into mutually exclusive reas, each area having a label, a value or an action to describe its data points. Splitting criterion is used to calculate which attribute is the best to split that portion tree of the training data that reaches a particular node.

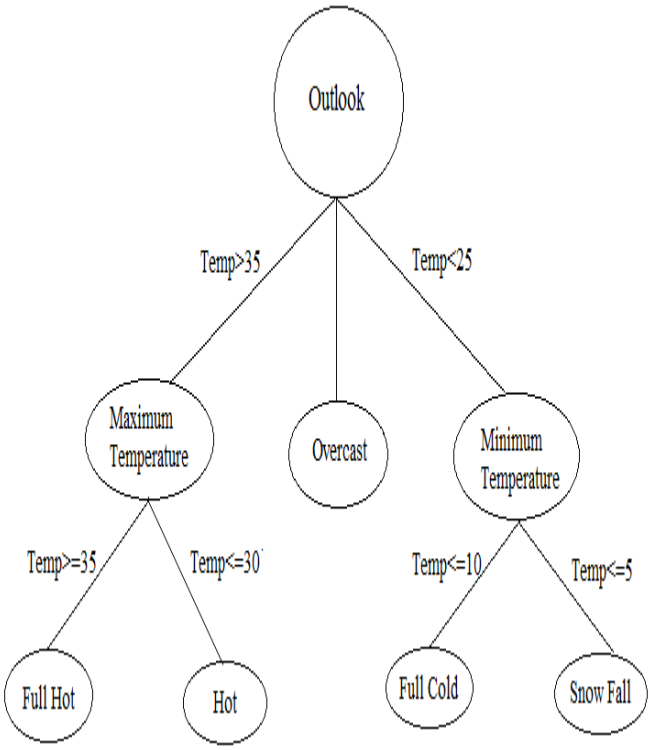

Figure3. Decision Tree generated by Training data sets

V. RESEARCH APPROACH

Weather prediction has always been a fascinating domain of research as changes in climatic conditions directly impact the population. [2] has classified the approach of weather prediction in two category, Empirical approach and Numerical approach. Empirical approach collects present weather conditions through ground observations i.e. observations from ships, satellites etc. This present climatic ondition observations taken are then forwarded to meteorological centres where it is analysed and is converted into multidimensional maps through the use of computers. Scientist predicts the changes in the map that will occur over the certain time period. Second approach is dynamical approach or numerical approach that uses mathematical equations over climatic variables to solve the prediction

VI .COMPARATIVE ANALYSIS : DATA MINING APPROACH IN WEATHER FORECASTING

Table 2 compare and analyses various data mining techniques used in whether forecasting on the basis of various parameters. Weather prediction approaches are challenged by complex weather phenomena with limited observations and past data.

From the below comparison based on various parameters, ANN comes out to be the most powerful technique that can give accurate and timely result. It can handle complex real and non real values, has greatest computation speed, provide more accurate results, and also provide better results for multidimensional data sets.

TABLE 2.COMPARATIVE ANALYSIS OF DATA MINING

parameter	**ANN**	**Decision tree**	**clustering**
performance	maximum	high	Moderate
Supervised learning	no	yes	No
Unsupervised learning	yes	no	No
Computation speed	fast	fastost	Slow
Cost effectiveness	yes	yes	Yes
Accuracy	More	less	More
Complexity level	complex	less	Moderate
Fault tolerance	yes	yes	Yes
Nature	analytical	analytical	Descriptive
Domain area	complex	simple	complex
Parallel computing	yes	yes	no

VII. RELATED WORKS

Table 3: Tabulates the various data mining techniques and ANN methods used for weather prediction with different set of weather parameters

author name	**prediction**	**techniq ue**	**Parameter**
folorunshoolaiya	Weather and climate	Data mining and artificial neural network	Temperature, rainfall, evaporation, wind speed
supriyBachaletal	Weather prediction for solar output of PV- cells	Data mining	Temperature, rainfall, cloud cover, humidity.
AbhishekSaxona et al,[8]	Weather forecasting	Artificial Neural network	Maximum and Minimum,

			Temperature , Rainfall, Cloud Conditions, Wind Stream and their Direction
ArpitTiwari [5]	Cloudburst	Data Mining	Year, Month, Average, Pressure, Relative Humidity, Clouds quantity, Precipitation and average temperature
Badhiye et al [6]	Temperature, humidity	Data Mining	Sea surface temperature.
DivyaChauhan etal,	weather	Data mining	Temperature , rainfall, pressure, and cloud condition.
Zahoor Jan [10]	Seasonal to Inter Annual climate Prediction	Data mining	Rain, wind speed, dew point temperature.
Rubu	rainfall	Data mining, neural network, fuzzy logic, and ANFISI	Surface pressure, heat, moisture.
P. Hemalatha [	Weather forecasting for ships using global positioning system.	Data mining	Climate(sunny ,cloudy, Rainy), Temperature , Humidity, Stromy
M.A.K. Nafeela et al, [12]	climate	Data mining	Temperature , humidity.

VIII. CONCLUSION

Weather forecasting is a meteorological work that easy to modify researcher work by applying the numerical weather prediction method. weather forecasted by using various data mining techniques especially classification clustering and neural network, decision tree. the key aim for improving the classification and prediction performance for the traditional; weather prediction model is designed and developed in this work. but some limitation of the model is also observed, thus in near future need to be review before use of the proposed technique. and also soil there are some issues and challenges in which better implement of data mining technique should be implemented in field of weather forecasting.

REFERENCES

[1] Folorunsha Olaiya, Adesesan Barnabas Adeyemo, Application of Data Mining Techniques In Weather Prediction And Climate Changes Studies, International Journal of Information Engineering and Electronic Buisness, pp.51-59,2012.

[2] Supriya Bachal, Shubhada Mone, Aparna Ghodke & Ujwala Ahirrao, Weather Forecasting Using Predictive Analysis For Determination Of Solar Output Of PV Cells, Imperial Journal Of Interdisciplinary Research, Vo l.2, No.4, 2016.

[3] Kavita Pabreja, Clustering Technique For Implementation Of Cloudburst Over Leh, CSI Journal Of Computing, Vol.1, 2012.

[4] Divya Chauhan, Jawarhar Thakur, Data Mining Techniques For Weather Prediction : A Review, International Journal on Recent And Innovation Trends In Computing And Communication, Vol.2, No.8, 2014

[5]] Arpit Tiwari, S. K. Verma, Cloudburst Predermination System, IOSR Journal Of computer Engineering, Vol.17, No.3, pp 47-56, 2015.

[6]] Badhiye S.S , Wakode B. V , Chatur P. N. , Analysis Of Temperature And Humidity Data For Future Value Prediction, International Journal Of Computer Science And Information Technologies , Vol.3, pp.3012-3014, 2012.

[7] Rupa, A Review Study Of Rainfall Prediction Using Neuro-Fuzzy Inference System, International Journal Of Engineering Science & Research Technology, 2015.

[8] Sarah N.Kohail, Alaa M.El-Halees "Implementation of Data Mining Techniques for Meteorological Data Analysis".
[9] M.Viswambari, Dr. R. Anbu Selvi "Data Mining Techniques to Predict Weather: A Survey".
[10] R.Samya, R. Rathipriya "Predictive Analysis for Weather Prediction using Data Mining with ANN: A Study".
[11] Folorunsho Olaiya, Adesesan Barnabas Adeyemo "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies".
[12] T V Rajini kanth, V V SSS Balaram and N.Rajasekhar "Analysis of Indian Weather Data Sets Using Data Mining Techniques".
[13] Siddharth S. Bhatkande, Roopa G. Hubballi, "Weather Prediction Based on Decision Tree Algorithm Using Data Mining Techniques".
[14] Rohit Kumar Yadav, Ravi Khatri "A Weather Forecasting Model using the Data Mining Technique".
[15] Sara khan, Mohd Muqueem, Nashra Javed "A Critical Review of Data Mining Techniques in Weather Forecasting".