Climate Prediction and Forecasting Using Data mining

Blessy Beena Sunny
APJ Abdul Kalam Technological University Trivandrum, Kerala
Mount Zion College of Engineering, Kadammanitta, India

Abstract— The weather conditions are changing continuously and the entire world is suffers from the changing Clemet and their side effects. Therefore pattern on changing weather conditions are required to observe. With this aim the proposed work is intended to investigate about the weather condition pattern and their forecasting model. On the other hand data mining technique enables us to analyze the data and extract the valuable patterns from the data. Therefore in order to understand fluctuating patterns of the weather conditions the data mining based predictive model is reported in this work. The proposed data model analyze the historical weather data and identify the significant on the data. These identified patterns from the historical data enable us to approximate the upcoming weather conditions and their outcomes. To design and develop such an accurate data model a number of techniques are reviewed and most promising approaches are collected.

Keywords: Data Mining, Importance, Steps Using in Data Mining, Weather Training Data, Training Time

I. INTRODUCTION

Weather forecasting has been one of the most scientifically and technologically challenging problems around the world in the last century. This is due mainly to two factors: first, it's used for many human activities and secondly, due to the opportunism created by the various technological advances that are directly related to this concrete research field, like the evolution of computation and the improvement in measurement systems . To make an accurate prediction is one of the major challenges facing meteorologist all over the world. Since ancient times, weather prediction has been one of the most interesting and fascinating domain. Scientists have tried to forecast meteorological characteristics using a number of methods, some of these methods being more accurate than others . Weather forecasting entails predicting how the present state of the atmosphere will change. Present weather conditions are obtained by ground observations, observations from ships and aircraft, radiosondes, Doppler radar, and satellites. This information is sent to meteorological centers where the data are collected, analyzed, and made into a variety of charts, maps, and graphs.

Modern high-speed computers transfer the many thousands of observations onto surface and upper-air maps. Computers draw the lines on the maps with help from meteorologists, who correct for any errors. A final map is called an analysis. Computers not only draw the maps but predict how the maps will look sometime in the future. The forecasting of weather by computer is known as numerical weather prediction. To predict the weather by numerical means, meteorologists have developed atmospheric models that approximate the atmosphere by using mathematical equations to describe how atmospheric temperature, pressure, and moisture will change over time. The equations are programmed into a computer and data on the present atmospheric conditions are fed into the computer. The computer solves the equations to determine how the different atmospheric variables will change over the next few minutes. The computer repeats this procedure again and again using the output from one cycle as the input for the next cycle. For some desired time in the future the computer prints its calculated information. It then analyzes the data, drawing the lines for the projected position of the various pressure systems. The final computer-drawn forecast chart is called a prognostic chart, or prog. A forecaster uses the progs as a guide to predicting the weather. There are many atmospheric models that represent the atmosphere, with each one interpreting the atmosphere in a slightly different way.

Weather forecasting is mainly concerned with the prediction of weather condition in the given future time. Weather forecasts provide critical information about future weather. There are various approaches available in weather forecasting, from relatively simple observation of the sky to highly complex computerized mathematical models. The prediction of weather condition is essential for various applications. Some of them are climate monitoring, drought detection, severe weather prediction, agriculture and production, planning in energy industry, aviation industry, communication, pollution dispersal, and so forth. In military operations, there is a considerable historical record of instances when weather conditions have altered the course of battles. Accurate prediction of weather conditions is a difficult task due to the dynamic nature of atmosphere. The weather condition at any instance may be represented by some variables. Out of those variables, one found that the most significant are being selected to be involved in the process of prediction. The selection of variables is dependent on the location for which the prediction is to be made.

The variables and their range always vary from place to place. The weather condition of any day has some relationship with the weather condition existed in the same tenure of precious year and previous week. Rainfall is a form of precipitation. Its accurate forecasts can help to identify possible floods in future e and to plan for better water management. Weather forecasts can be categorized as: Now forecasts which is forecasts up to few hours, Short term forecasts which is mainly Rainfall forecasts is 1 to 3 days forecasts, Forecasts for 4 to 10 days are Medium range forecasts and Long term forecasts are for more than 10 days. Short range and Medium Range rainfall forecasts are important for flood forecasting and water resources management.

II. IMPORTANCE OF DATAMINING IN PREDICTION

Data mining, also called knowledge discovery in databases, the process of discovering interesting and useful patterns and relationships in large volumes of data. ie knowledge mining from data, knowledge extraction or data or pattern

analysis .Data mining is knowledge discovery from data(KDD process).

A. Steps in Data Mining

- Data Cleaning
- Data Integration
- Data Selection
- Data Transformation
- Data Mining
- Pattern Evaluation
- Knowledge Presentation
- Data Cleaning: This process is used to remove noise and inconsistent data.
- Data Integration: Data integration is used to combine multiple set of data sources.
- Data Selection:data selection is used to retrieve the relevant data's for analysis from databases or other repositories.
- Data Transformation: Data are transformed or consolidated into forms appropriate for mining by performing some summary or aggregation operations.
- Data mining: its an intelligent methods to extract different data patterns.
- Pattern evaluation: identifying the truly interesting patterns.
- Knowledge presentation: ie, visualizaton and knowledge presentation techniques used to present the mined knowledge.

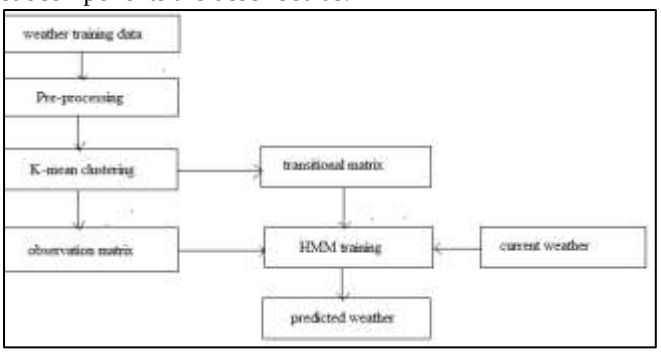
III. RELATED WORK

A. K-Means

The K-Means clustering algorithm is a partition-based cluster analysis method. According to the algorithm we firstly select k objects as initial cluster centers, then calculate the distance between each object and each cluster center and assign it to the nearest cluster, update the averages of all clusters, repeat this process until the criterion function converged Square error criterion for clustering.

B. Proposed Work

The proposed data model is provided using the figure 1. In this diagram the different components of predictive data model is demonstrated. These components are used for organizing the complete system therefore the detailed subcomponents are described as:-



1) Weather Training Data

The supervised algorithms are working on labeled data. Such kind of data most of the time found in structured format. This structured data has some pre-defined class labels that are representing the outcomes of the combination of attributes. Thus using the training data the algorithm learned on pre-defined patterns. In this presented work the Bhopal weather forecasting data for last five years are used. That training dataset contains the different weather attributes observations and the class labels as the weather conditions.

2) Pre-Processing

The pre-processing is a technique by which the data is refined, transformed and cleaned for improving the quality of the training data on which the data model is prepared for decision making or prediction. In this given technique the three major contributions is placed in this phase.

1) Removal of attributes that are not fluctuating with the other patterns
   In this phase those attributes are removed from the training samples which are not fluctuating with the different data patterns. Thus this technique reduces the dimension of the dataset by which the memory consumption of the data analysis is reduces.
2) Removal of attributes that uniquely defined the instance data
   In this phase the data set is evaluated for removal of data that are performing the identity representation for the dataset objects.
3) Missing data handling
   In this phase the data is analyzed for finding the missing attributes in the data set. That also improves the quality of data for representation of accurate data model. Therefore those dataset objects are removed which are not completed or having the missing attributes.

3) K-Means Clustering

In this phase the well refined and defined data is used for prepare the groups of the data which are simulating the similar behavioral patterns. Therefore the entire data is clustered according to the similar attributes in these grouping of data the data attributes are considered.

4) Transition Matrix

As discussed previously the output of k-means clustering is organized in two matrixes the first observation matrix is given using the table 2 which includes the probability distribution of the weather conditions in the observations.

5) HMM Training

The hidden Markov model is responsible to accepting these two matrixes as input and producing the learned model for prediction. The Hidden Markov model is trained in this phase for the given observational patterns and the transitional patterns.

6) Current Weather

In order to predict the next weather condition or upcoming weather condition the system required to take input the just patterns of the weather conditions, based on the observation and transitional patterns the system generate the next possible pattern of weather condition.

7) Predicted weather

That is the final outcome of the proposed data model described.

Input: training dataset D, current weather conditions C
Output: predicted weather condition P
Process:
1. O= Read_Training_data(D)

2. [ ] = Kmeans_Clustering(O, 12)
3. T = HMM.Train( )
4. P = HMM.Predict (T, C)
5. return P

C. Training time

The amount of time consumed during the training of the system is termed as the training time of the algorithm. That is also termed as the training time complexity for the algorithm.

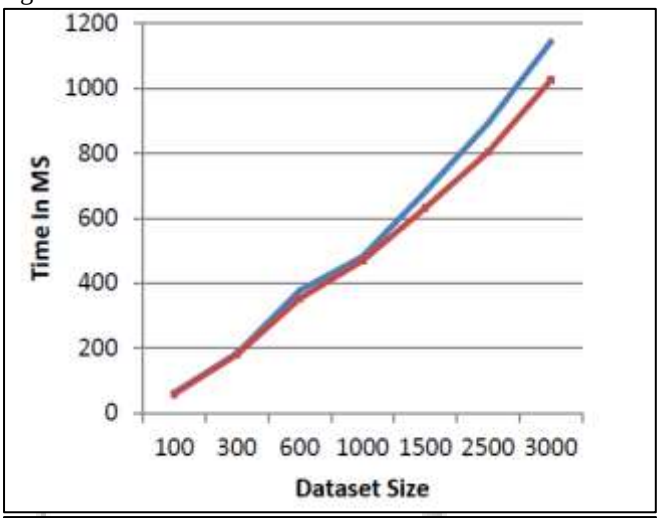


Figure 2 training time

Table 4 training time

Dataset Size	Proposed technique	ID3 decision tree
100	63.4	59.21
300	184.7	180.43
600	378.1	353.15
1000	482.32	469.41
1500	681.45	631.52
2500	891.38	803.04
3000	1142.43	1024.41

Figure 2 and the table 4 shows the training time of the algorithms in terms of milliseconds. Additionally during the experimentation, the performance of algorithms with the size of dataset is increases or decreases are reported. In order to represent the performance of system the X axis of diagram contains the dataset size and the Y axis shows the estimated time in milliseconds.

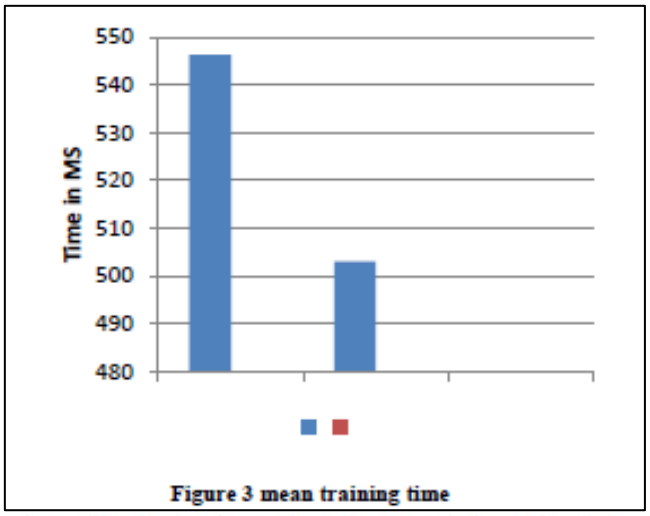


Figure 3 mean training time

According to the obtained results the traditional algorithm namely ID3 decision tree algorithm consumes less time as compared to the proposed algorithm. This because for improving the accuracy in prediction of the system needs to calculate the additional parameters as compared too traditional method. Thus the time complexity of proposed algorithm is higher as compared to traditional system. To clearer view of the data modelling the mean training time of both the techniques are evaluated and compared using figure 3. In this diagram the X axis contains the methods

Table 6 memory consumption

Dataset Size	Proposed technique	Traditional technique
100	30992	31773
300	31648	32937
600	33827	34912
1000	35711	37811
1500	36837	38817
2500	38271	39721
3000	39813	41019

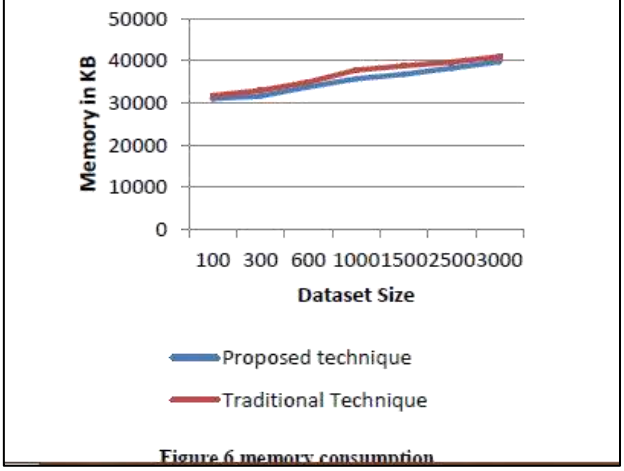


Figure 6 memory consumption

That is also known as the space complexity in algorithm study. The figure 6 and table 6 shows the memory

consumption of the system with increasing size of training dataset. The amount of data is given using the X axis and the Y axis shows the amount of consumed memory during experimentation with respective amount of data in terms of kilobytes. According to the experimented results the amount of memory is similar and not more fluctuating. But the respective comparison the proposed algorithm is efficient than the traditional approach of prediction. In order to investigate the difference in the memory consumption of the proposed and traditional technique the mean memory consumption is calculated and reported using the figure 7. In this diagram the X axis contains the methods name and the Y axis contains the mean memory consumption of the system.

IV. FUTURE ENHANCEMENT

Climate prediction using data mining is scope ful and easy to predict the weather.

So we do not bothered about the changes of climatic conditions. Also this is helpful in business valuable information that the business can use to make decisions about the future of the entire organizations.

The proposed work is intended to find the solution for accurate weather data modeling and prediction using the historical data. Therefore the data mining technique is studied for developing such kind of data model. The data mining techniques analyze the data of some pre-defined pattern and extract the significant on the data. Using the extracted patterns from the data the model takes training and prepared for classifying or predicting the similar patterns of associated class labels. The classification of the data supported by supervised technique of data mining. The key issue in weather prediction is to accomplish the relationship among the class labels and the attributes which are used for predicting the weather conditions. Therefore using the available accurate techniques a new data model is developed for weather forecasting.

The proposed weather forecasting data model utilizes the k-means unsupervised learning technique for performing the clustering on the entire training dataset. This clustering is performed for finding the pattern level pattern similarity among two instance data. The total number of 12 clusters is developed; these clusters are providing the observations of the data and their weather conditions. Using the extracted observations and available class labels the data is re-organized in terms of observation matrix and the transition matrix. These two different matrixes are provided into next phase namely Hidden Markov Model and the training are performed. The trained data model is used for prediction or the pattern recognition work.

REFERENCES

[1] David A. Dickey, N. Carolina State U., Raleigh, NC, "Introduction to Predictive Modeling with Examples", Statistics and Data Analysis, SAS Global Forum 2012
[2] https://www.google.co.in/search
[3] www.studymafia.org

