Drought Prediction using Data Mining Techniques and Effective Aid Delivery System

Abhijeet Deshpande , Vinayak Dixit , Ankur Dalvi , Mrunal Ghanwat and Prof. Deepa Abin
Department of Computer Engineering
Pimpri Chinchwad College of Engineering , Nigdi - 411044

***Abstract*—Drought affect a massive area and has a long lasting impact as compared to any other natural disaster. The frequent occurrence of drought affects the agricultural production at a larger rate. The techniques used for drought prediction at current stage are by using satellite imaging, hyper-spectral imagery and many other techniques. In the proposed system we are increasing the granularity for addressing this problem. The rainfall statistics and local catchment area can provide the statistical data for prediction. The prediction techniques in Data Mining can be used accordingly to predict whether the area will be affected by droughts and accordingly mobile techniques can be implemented for delivery of aid and commodities provided by the administrative services.**

I. Introduction

The Indian subcontinent has itself a wide variety in its climate. In some areas the climate is arid where as in some of the areas it is cloudy.This variety in the climate affects the human life at a larger scale and has beneficial as well as adverse effect on the people living in that area. It also affects the flora and fauna in the area affecting the vegetation at a larger scale.The proposed system described here mainly deals with the problem faced by the community due to the drought. The Deccan plateau is one of the most affected regions by the drought causing the people here suffer at a larger extent for mere surviving. The proposed system doesn't deal with the impact of the drought as it is the natures force which is uncontrollable but the proposed system is predicting the key to reduce the force of impact. Hence the proposed system will be predicting whether the specific area is affected or not by the drought. Later system deals with the delivery of the allocated aid to the victim.The proposed system will work as a continuous flow of execution as the phases which will be carried are in a sequential manner.The completion of one phase will only trigger the initialization of the next phase. Proposed system will not only predict the drought for given area but also tackle the delivery to the needy once. Thus the system is a full fledged representation for an alternate plan for post impact of drought reducing the severity of the calamity.

The delivery system will consist of integrated finger print scanner and commodity storage. Finger print scanner will be used to authenticate the identity of people those have been sanctioned for the aid. Commodity storage will have the function to render aid. In the system the data of the people eligible to get a sufficient amount of aid will be collected in the initial stage and it will be stored on the centralized system. According to the data collected, the amount of aid will be decided based on the number of persons in the family, financial status and assets they possess.

The prediction system will take input statistics are rainfall, population statistics, area categorization and catchment area to monitor and categorize the areas provided to it as an input in either drought affected or non-affected.

II. Methodologies

There are many different prediction methods available to solve all kinds of problems. They can be divided into short-term and long-term forecasting. By resulting properties, methods can be classified into qualitative and quantitative prediction method. Qualitative prediction technology is used for forecasting the change trend, the possibility and extent of anomalies while quantitative prediction technology is to make predictions by looking for the quantity characteristics of the prediction objects, according to sufficient historical data [5].

At current stage, most commonly methods for drought prediction are satellite imaging, hyper-spectral imagery and some Data Mining techniques. Proposed system uses Data Mining technique for prediction of drought. Below subsections contains the discussion about mentioned techniques for drought prediction.

A. Satellite Imaging

Satellites measure energy intensities (radiances) at several wavelengths of the electromagnetic spectrum. This information is useful because everything the ground, the oceans, the atmosphere, clouds, rain, vegetation, cities, people, etc. absorbs energy at certain wavelengths and emits energy at other wavelengths. You may be familiar with some of these: visible satellite imagery of clouds showing the movement and strength of storms and fronts; infrared imagery which measures the temperature of clouds and weather systems; water vapor maps generated from a spectroscopic analysis of satellite data.

B. Hyper-spectral imagery

Spectral reflectance, measured by hyperspectral imaging equipment, is the amount of reflected light from a surface. Hyperspectral imaging is the process by which images are taken and numerical values (spectral radiance) assigned to each pixel, utilizing a range of wavelengths across the electromagnetic spectrum, including visible and infrared regions.

Through the use of specialized software and statistical analysis, these pixels are sorted and characterized to distinguish between groups of pixels or in the case of precision agriculture, plant characteristics and environmental conditions. Earlier remote sensing technology, in particular multispectral imaging, collects data at a few widely-spaced wavelengths.The data from each wave-length band is assembled into a three-dimensional hyperspectral data cube for processing and analysis. Each layer of the cube represents data at a specific wavelength.

C. Data Mining Techniques

The main goals of data mining are prediction and description identifying patterns and relationships in data, and generates new knowledge about the data set. The basic data mining tasks are divided as follows: Predictive: Classification, Regression, Time series analysis, Prediction
Descriptive: Clustering, Summarization, Association rules, Sequence discovery (sequential analysis), Estimating the model

As large collections of historical data are essential in identifying relationships among different climatic and meteorological parameters, proper handling of data and quick processing becomes necessary. This highlights that efficient data management and information extraction is important in drought research. The difficulty of handling huge amount of data, understanding the complex results derived from atmospheric phenomenon and extracting the spatial and temporal pattern of drought can be resolved with data mining techniques.

III. LITERATURE SURVEY

K. Sriram and K. Suresh [2] has established a new method, named the Integrated Surface Drought Index (ISDI). ISDI integrates traditional meteorological data, remotely sensed indices and biophysical data and attempt to describe drought from a more comprehensive perspective. The evaluation results indicated that the construction models for three phases of growth season have very high regression accuracy. But it has large time taken for finding the results. It is more complicated to implement by using the spatial and temporal characteristics of data. For this apply the best machine learning mechanism for predicting the agricultural yields by using the time and spatial parameters.

Standardized precipitation index (SPI) has been used as a conventional tool to identify and monitor drought occurrences. However, to reduce and mitigate the adverse effects of drought impacts, effective forecasting of future droughts is necessary. To derive the SPI values for durations of 3 to 9 months, Daniel Hong and Kee An Hong [3], average long term monthly rainfall data for eight stations covering both the dry and wet seasons from Selangor river basin in Malaysia have been used. These drought indicators were used as time series for drought forecasting for the basin using the multi-layer artificial neural networks model. Results show that more accurate predictions are achieved using SPI of longer durations, i.e. 6 and 9 months.

Xin HUANG, Hong-liang LI and Lin QIU [4] proposed a system which uses fuzzy clustering iteration method to cluster the data of many years rainfall. Later authors have uses R/S analysis to establish a prediction model to research on time series data mining of rainfall, then predicting the years of occurrence of agricultural drought.

Xiaotian Gu and Ning Li [5] have used Extended Empirical Orthogonal Function technique to combine the spatial characteristics and temporal characteristics to make their Droughts and Floods predicting system more effective.

In a study by Xiaofan Liu, Liliang Ren, Fei Yuan, Bang Yang [6] paper, daily precipitation data was obtained from 52 rainfall stations spread across the catchment, and other meteorological data are obtained from 4 meteorological stations around the catchment. Based on DEM, the key meteorological variables (temperature, vapor pressure and wind velocity) were to graphically corrected based on the empirical relationships [6]. All meteorological data were interpolated over the study area using the inverse distance square method.

A Markov process is a process for which, if the present is given, the future and the past are independent of each other. A Markov chain used in this system considers the sample space of a problem (the set of all possible outcomes of a random process) to be finite. A Markov chain model deals with presentation of time as a discrete value. In this case all the variables of the problem become integers, the description is reduced to a matrix of transition probabilities, and matrix algebra becomes the basic tool for modeling a process. In this research, the time series in each grid cell were predicted by the Markov chain model.

IV. EXISTING SYSTEM

The system until today has defined for prediction of agricultural drought on the basis of multi-source information for prediction.

A. Using Naive Bayes classification:

Xin HUANG, Hong-liang LI, Lin QIU [4] proposed a system which uses Nave Bayes classification to predict the patterns for drought arrival. The input selection parameters are meteorological characteristics such as scarcity of precipitation, actual evaporation at only smaller fraction to the potential evaporation, shortage of moisture in the soil. These contents are considered for a single database table or single statistical data matrix. The data may be found incomplete: lacking attribute values, lacking certain attributes of interest or containing only aggregate data noise. The elimination of unwanted characteristics has to be carried out.

The system uses supervised classification algorithm called Nave Bayes for the implementation of drought predictive system.

B. Using Markov chain model:

The required meteorological data used in this system is daily mean, maximum and minimum air temperature; air vapor pressure; wind velocity; daylight duration; and precipitation.

Xiaofan Liu, Liliang Ren, Fei Yuan, Bang Yang [6] paper have used Markov process for predicting drought. A Markov process is a process for which, if the present is given, the future and the past are independent of each other. A Markov chain used in this system considers the sample space of a problem (the set of all possible outcomes of a random process) to be finite. A Markov chain model deals with presentation of time as a discrete value. In this case all the variables of the problem become integers, the description is reduced to a matrix of transition probabilities, and matrix algebra becomes the basic tool for modeling a process. In this research, the time series in each grid cell were predicted by the Markov chain model.

These are the similar systems defined for the prediction of the drought conditions. However, the finding out intensity of drought, collection of the data of people residing in such regions, identifying the attributes to provide the aid to people in affected region, considering the varying amount of aid as per need and delivery of the aid to the needy people is carried out conventionally by the government. The existing systems for delivery of the aid to affected region use the survey of affected areas to decide allocation of aid.

Initially, the areas affected by the drought are identified by weather forecasting officers and the data of people located in those areas are collected which may be in need to get an aid. This data is verified and accordingly central government allocates the aid to specific states. This aid is then transferred to various districts by the state government and it eventually reaches to the people located in areas affected by the drought.

V. Proposed System

Proposed System is divided into two subsystems, Prediction system and Effective Aid Delivery System. Proposed System uses Data Mining Technique for drought prediction and uses embedded system for implementation of an effective aid delivery system. The objective of this system is to provide an effective drought prediction giving a prior notice to inculcate appropriate measures beforehand and to ensure not only to delivery aid whenever necessary but also to the intended people in allocated quantity.

A. Architecture of proposed system:

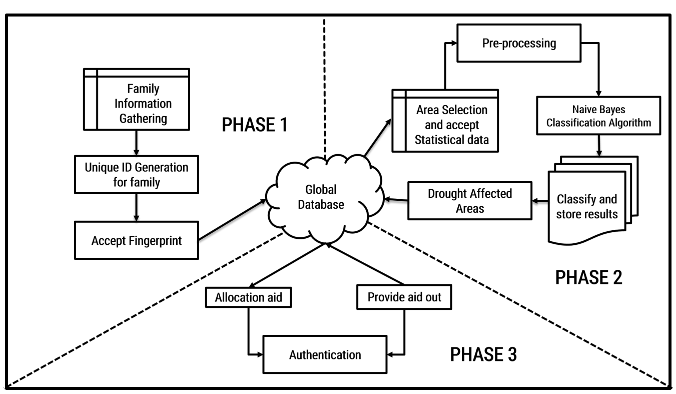

Fig. 1. Architecture of Proposed System

As shown in Fig [1], the proposed system is divided into three phases:

a) Dataset: This phase is dealing with acquiring the information of the people living in the specific area .This information consists of the family assets , the members of family and choosing the family head.

b) Prediction: This phase of the system will deal only with the drought prediction using advance Data Mining classification techniques. This phase starts with taking necessary statistical data of particular areas for which drought is to be predicted. Then this data goes through Pre-processing to make it compatible for algorithm that is going to used further i.e Naive Bayes. The output parameters are drought affected areas and non-affected areas. The result is then stored in Global database for further use.

c) Delivery: The prediction analysis result will be available beforehand being the rains at the end. Hence the civic authorities will have the prior knowledge of the count of the affected and hence the aid can be kept ready before the calamity hits. The delivery system deals with this phase. The system will have all computing allocation details to make sure appropriate amount of aid is provided to each victims according to their need. This phase starts with authentication. If authentication succeeds then aid will provided out from the system.

B. Algorithm:

In the prediction stage there will be the a classification algorithm based on supervised learning algorithm for prediction of the area in drought affected or non affected. The algorithm will be fed with a class label as the prediction set whether yes or no where as the data provided will be

- Rainfall Statistics for that particular area
- Catchment or storage area that the specified area has
- Population of the specific area
- Classification of the area as agricultural based or residential based

Bayes theorem provides a way of calculating the posterior probability, P(c—x), from P(c), P(x), and P(x—c). Naive Bayes classifier assume that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence.

$$\underbrace{P(c \mid x)}_{\text{Posterior Probability}} = \frac{\overbrace{P(x \mid c)}^{\text{Likelihood}}\,\overbrace{P(c)}^{\text{Class Prior Probability}}}{\underbrace{P(x)}_{\text{Predictor Prior Probability}}}$$

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

- P(c—x) is the posterior probability of class (target) given predictor (attribute).

- P(c) is the prior probability of class.
- P(x—c) is the likelihood which is the probability of predictor given class.
- P(x) is the prior probability of predictor.

VI. Comparative Study

Existing systems are defined for drought prediction techniques. Allocation of the aid and delivery is carried out by conventional method of delivery process by the government. Proposed system provides validation of data mining techniques through accuracy. Classification of area based on the weather conditions and drought affection will help to determine the allocation of the aid.

As the officers have personally visit the places and decide the amount of aid that should be allotted to people in affected region, the delivery process requires a lot of time. The prediction carried out in advance in proposed system will help government to define the probability of drought arrival thereby the making delivery of aid corresponding to need of people in lesser amount of period.

The transportability by the mobile van, easy access to the global database through the cloud system, secure transaction commitment will be effectively provided by the proposed system as compared to the existing system.

Availability of the data in proposed system will be persistent as database will be made available through central storage and updating of the data carried out corresponding to the aid delivered/undelivered.

VII. Conclusion

Proposed system can be developed for prediction of drought using Data Mining techniques. There is a need to adapt such system to predict the drought before it hits and leads to reduction of impact of drought on victims by providing the aid effectively through delivery system.

References

[1] Raj Kumar, Dr. Rajesh Verma, *Classification Algorithms for Data Mining: A Survey,* International Journal of Innovations in Engineering and Technology (IJIET), 2012

[2] K. Sriram and K. Suresh *Machine Learning Perspective for Predicting Agricultural Droughts Using Naive Bayes Algorithm,* Middle-East Journal of Scientific Research, 2016

[3] Daniel Hong, Kee An Hong *Drought Forecasting Using MLP Neural Networks,* Middle-East Journal of Scientific Research, 2015

[4] Xin HUANG, Hong-liang LI, Lin QIU *Research on Predicting Agricultural Drought Based on Fuzzy Set and R/S Analysis Model,* International Conference on Advanced Computer Theory and Engineering (ICACTE), 2010

[5] Xiaotian Gu, Ning Li *Study of Droughts and Floods Predicting System Based on Spatial Temporal Data Mining,*

[6] Xiaofan Liu, Liliang Ren, Fei Yuan, Bang Yang *Meteorological Drought Forecasting using Markov Chain Model,* International Conference on Environmental Science and Information Application Technology, 2009