ISSN 2320 – 2602
Volume 4 No.4, April 2015
International Journal of Advances in Computer Science and Technology
Available Online at http://www.warse.org/ijacst/static/pdf/file/ijacst13442015.pdf

A Significant Review of Different Drought Indices for Predicting Agricultural Droughts

R.KALPANA
PG Student (CSE)
Nandha Engineering College
Erode
rm.kalpana@gmail.com

Dr.S.ARUMUGAM
CEO
Nandha Educational Institutions
Erode
ceo@nandhaengg.org

ABSTRACT

Drought affects a large number of people and cause more losses to society compared to other natural disasters. India is a drought disaster-prone country. The frequent occurrences of drought possess an increasingly severe threat to the Indian agricultural production. Drought a very complex phenomenon and it is difficult to accurately quantify it. In the existing system, implement the ISDI model construction for evaluating the accuracy and the effectiveness. The ISDI model using a variety of methods and data, there is still need some work to be done in our future research because of its complex spatial and temporal characteristics of drought. To overcome limitation, the performance of the drought can be measured by using the Spatial and temporal characteristics of information. We collect the dataset from different regions and also collect the time varying information. In proposed system, we predict the drought conditions by using the supervised learning mechanism. It can be implemented by using the Bayesian supervised machine learning algorithm. Through this algorithm we can achieve the accuracy and performance, and also improve the effectiveness of predicting various drought conditions.

Keywords: ISDI, Bayesian Algorithm, Vegetation condition, Yield Estimation

1. INTRODUCTION

Data mining in agriculture is a very recent research topic. It consists in the application of data mining techniques to agriculture. Recent technologies are nowadays able to provide a lot of information on agricultural-related activities, which can then be analyzed in order to find important information. Carrying out effective and sustainable agriculture has become an important issue in recent years. Agricultural production has to keep up with an ever-increasing population. A key to this is the usage of modern technologies such as GPS (for precision agriculture) and data mining techniques to take advantage of the soil's heterogeneity. The large amounts of data that are nowadays virtually harvested along with the crops have to be analyzed and should be used to their full extent - this is clearly a data mining task [1]. Data mining allows to extract the most important information from such vast data and to uncover previously unknown patterns that may be relevant to current agricultural problems, thereby helping farmers and managing organizations to transform data into business decisions. Several Data mining techniques used in agriculture study area. We are discussed the few techniques here. Some of the data mining techniques are related to weather conditions and forecasts. For example, the K-Means algorithm is used to perform forecast of the pollution in the atmosphere, the K Nearest Neighbor (KNN) is applied for simulating daily precipitations and other weather variables, and different possible changes of the weather scenarios are analyzed using SVMs [3]. Also K means method is used to forward the pollution in atmosphere. Different changes of weather are analyzed using SVM. K means approach is used to classify the soil and plants. Wine fermentation process monitored using Data mining techniques. By using Multilayer Perception model of Neural Networks the researchers trained to predict wheat yield by considering sensor input and fertilizers as parameters. SVMs for detecting weed and nitrogen stress in corn. Data Mining techniques are often used to study soil characteristics. The K-Means approach is used for classifying soils in combination with GPS-based technologies. Apples are checked using different approaches before sending them to the market, and a neural network is trained for discriminating between good and bad apples [6]. Apply a supervised biclustering technique to a dataset of wine fermentations with the aim of selecting and discovering the features that are responsible for the problematic fermentations and also exploit the selected features for predicting the quality of new fermentations. Taste sensors are used to obtain data from the fermentation process to be classified using ANNs. Similarly, sensors are used to smell milk that is classified using SVMs [1].

2. PROBLEM FORMULATION

DATA MINING PROCESS MODEL

In the course of this project we have analyzed over 50 real-world data sets, primarily agricultural data sets provided by research institutes and businesses in New Zealand. From this experience we have developed a process model for applying data mining techniques to data, with the goal of incorporating the induced domain information into a software module. The key points of this model are

•A Two-way interaction between the provider of the data and the data mining expert

Both work together to transform the raw data into the final data set(s) input to the machine learning algorithms — with the domain expert providing information about data semantics and 'legal' transformations that can be applied to the data, and the data mining expert guiding the process so as to improve the intelligibility and accuracy of the results [10].

•An Iterative Approach

Machine learning is an exploratory process; it generally takes several cycles through the process model to find a good "fit" between a representation of the data and a data mining algorithm. In addition, distinct attribute combinations run through different schemes can produce wildly different data models, even though the predictive accuracy of the results may be equivalent. These alternative views may provide valuable insights into patterns covering different subsets of the data. In the model presented in Figure, activity flows in a clockwise direction [2].

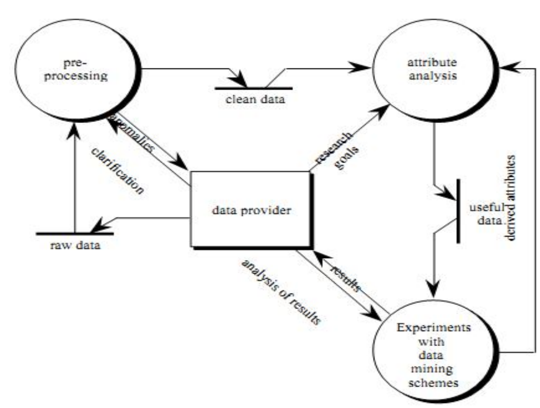

Figure.1. Process Model for a Machine Learning Application

In the pre-processing stage, the raw data is firstly represented as a single table, as required by the data mining algorithms included in WEKA. This table is translated into the ARFF format, an attribute/value table representation that includes header information on the attributes' data types. The data may also require considerable 'cleansing', to remove outliers, handle missing values, detect erroneous values, and so forth. At this point the data provider (domain expert) and the data mining expert collaborate to transform the cleansed data into a form that will produce a readable, accurate data model when processed by a data mining algorithm. These two analysts may, for example, hypothesize that one or more attributes are irrelevant, and set aside these extraneous columns. Attributes may be manipulated mathematically, for example to convert all columns containing temperature measurements to a common scale, to normalize values in a given column, or to combine two or more columns into a single derived attribute [7].

A. DATASET SELECTION

Input Selection is the first process. In this, first have to browse and select the input for the process. Input of the process is dataset. Most commonly a data set corresponds to the contents of a single database table, or a single statistical data matrix, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set in question. The data set lists values for each of the variables, such as height and weight of an object, for each member of the data set. Each value is known as a datum [8]. The data set may comprise data for one or more members, corresponding to the number of rows. The term data set may also be used more loosely, to refer to the data in a collection of closely related tables, corresponding to a particular experiment or event. In our process Dataset is collected from the website or collected by manually.

B. DATASET PREPROCESSING

Normally Dataset preprocessing is the method for cleaning the dataset. Normally data contains incomplete: lacking attribute values, lacking certain attributes of interest, or containing only aggregate data noisy: containing errors or outliers inconsistent: containing discrepancies in codes or names. In this we are going to eliminate this type of occurring in the dataset. Eliminating the unwanted value or symbols or characters in the dataset.

C. ATTRIBUTE CLASSIFICATION

Classification is a way of categorizing the data (records) for an attribute. The choice of classification system is critical to information displayed by a map. Classification can be used to enhance the information or to deliberately mislead. Attributes can use different classifications for the same data to change the nature of the display; this can be achieved based on the attributes in the dataset. Attributes in the dataset are the scarcity of precipitation, actual evapotranspiration at only a small fraction of the potential evapotranspiration rate, and the shortage of soil moisture) impact on agriculture (e.g., yield reduction) [7]. By using this type of information's, we classify the dataset.

D. PROBABILITY ESTIMATION

In our implementation process, we implement the drought research by using the supervised classification algorithm called Naïve Bayes Classification algorithm. A Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem (from Bayesian statistics) with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". Depending on the precise nature of the probability model, naive Bayes classifiers can be trained very efficiently in a supervised learning setting [5]. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood.

An advantage of the naive Bayes classifier is that it requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix. In our implementation, naïve bayes is used for estimating and predict the drought affected regions in less time execution. It is estimated by using the neighborhood relationship between the data in the dataset.

E. EVALUATION

Predicting the drought affected region by using the naïve bayesian algorithm and evaluate the performance by using the parameters of the process. Then, we evaluate the graph based on the extreme, mid, low, normal droughts etc., like that. These are estimated by using the probability values in the dataset. We can evaluate the graph by using this type of parameters form probability estimations.

3. PROPOSED FRAMEWORK

The proposed framework is divided into four major modules. They are

- Metrological Statics data Analyzer
- Vegetation index for the Agro-Drought Monitoring
- Indicator of Agro-Drought Severity.
- Sensational Historical Variation region Distribution

A. METROLOGICAL STATICS DATA ANALYZER

Agriculture is an important economic sector in Guangdong province of China, where precipitation distributes unevenly in both spatial and temporal dimensions. More than 60% of annual rainfall drops in the short period from June to August, leading to frequent occurrence of agro-drought in other months of the year. Seasonal droughts usually occur in winter and spring to shape a significant impact on agricultural harvest. Drought monitoring has thus been a necessary effort to alleviate the impact. Generally drought can be monitored through either ground observation or remote sensing. Satellite-based drought indices have obvious advantages compared to station-based meteorological drought indices in spatial resolution.

The basic meteorological data set used in the ISDI contains 50 years (1960–2009) long daily series of air temperature and atmospheric precipitation at 130 weather stations in mid-eastern China [10]. All the data set are derived from the China Meteorological Data Sharing Service System.

B. VEGETATION INDEX FOR THE AGRO-DROUGHT MONITORING

Vegetation index was an important parameter in calculating VSWI for drought monitoring. Since green vegetation had strong absorption of spectrum in red region and high reflectance in infrared region, vegetation index was thus generally formulated as various combinations of red and infrared bands. Over 20 indices developed as vegetation index and the most famous was the normalized difference vegetation index (NDVI), hence would also be used in this study for computation of VSWI. We try to integrate multi-source information to achieve the purpose of accurately monitoring agricultural drought [4]. The information used to generate the new drought index (ISDI) includes climate-based drought index information, satellite-based drought indices related to vegetation condition and soil moisture content, and several biophysical characteristics used to reflect regional difference of drought. We collected all the involved observed meteorological data at the meteorological stations and the Moderate Resolution Imaging Spectrodiometer (MODIS) Terra data as the data source of ISDI inputs.

C. INDICATOR OF AGRO-DROUGHT SEVERITY

Growing crops need continuous supply of soil water to ensure harvest. Rainfall and irrigation were main sources of soil water in agricultural fields. When soil water supply was sufficient for crop growing, evapotranspiration from agricultural fields would be high, leading to low surface temperature observed in satellite remote sensing images. During the drought period, soil water supply was in shortage to meet the normal demand of crop growing.

Consequently stoma on crop leaves tends to close in order to decrease water lost from canopy, leading to apparent increase of temperature in the fields. Therefore, using the relationship between canopy temperature change and soil water supply in the fields, we were able to develop an approach for drought monitoring [5]. SPI provides a measure of precipitation deficit compared to historical precipitation record. PDSI accounts for the effect of both precipitation and temperature and their combine effect on soil water available to drought conditions.

D. SENSATIONAL HISTORICAL VARIATION REGION DISTRIBUTION

However, seasonal changes were also obvious as to the degrees of favorable condition. The highest value of drought index was observed in the periods of late-May and early-June in Guangzhou, Shaoguan and Shantou. As a contrast, the lowest value of the index was also seen in this period in Xuwen. Since the higher the index, the wetter the ground surface is. Therefore, we could observe a sharp difference in drought severity in the province during this season. The ground was generally in normal condition for agriculture during the growing seasons ranging from May to September [8]. The slightly worse conditions were observed in the periods between late-June and early-July, as well as between mid-August and early-September. As a contrast, the general better conditions appeared in the period between mid-July and early-August. Since the drought index of the four regions located.

4. EXISTING SYSTEM

Agricultural drought indices use methods of mathematical modeling to transform the factors affecting crop growth and development into a single number and give a comprehensive drought description for decision making in the agricultural sector. Since the beginning of 20th century, a lot of drought indices have been developed for monitoring the occurrence and variation of agricultural drought. The first developed indices are meteorological drought indices such as Standard precipitation Index (SPI) and Palmer Drought Severity Index (PDSI). The SPI was one of the drought indices been widely used worldwide, which is designed to be a spatially invariant indicator (spatially and temporally comparable) only based on in-situ precipitation data. The PDSI is calculated using historical temperature and precipitation, and information of the available water content of the soil based on a soil moisture/water balance equation. Although the meteorological drought indices can get more accurate and spatially and temporally comparable drought conditions, their utilization is enslaved to the density and distribution of the station network. This type of indices also can not reflect the vegetation condition induced by the water deficit [10].

Table 1 Regression Precision Evaluation of VEGDRI and ISDI

The three seasonal model		Average error	Relative error
Spring	VegDRI	0.3688	0.24
	ISDI	**0.3569**	**0.23**
Summer	VegDRI	0.7152	0.42
	ISDI	**0.7064**	**0.42**
Autumn	VegDRI	0.3984	0.20
	ISDI	**0.4105**	**0.22**

A new drought index named ISDI was established based on the concept of VegDRI. Through analysis, we improved the original index, adding satellite-derived temperature information and elevation to the input variables of ISDI. The improved model attempts to characterize drought in a more accurate and comprehensive way. Cross-validation analysis proved that the regression accuracy is very high and the drought condition can be accurately portrayed using the ISDI input variables. We also found that the influence of independent variables to the model's judgment condition and linear regression equation results are different in three phases of the model. This indicated that the ISDI model using data mining technology can accurately reflect the temporal difference of drought characteristics. Compared to the drought news report and field observation of drought condition, the ISDI results can correctly reflect the regional and local scale spatial distribution of drought [8].

DRAWBACKS OF EXISTING SYSTEM

The following drawbacks are identified from the existing system.

- However, the above methods still has its own limitation and need for further research and improvement.
- The 1 km spatial resolution should be improved to 500 m, 250 m or higher using the other remote sensing products to provide more localized drought information.
- Other indices, such as the Enhanced Vegetation Index (EVI) or Leaf Area Index (LAI) should be evaluated as new inputs for the ISDI to avoid the limitation of NDVI saturation at values near 0.8.
- Currently, only 10 years historical record training database was obtained to establish the various models. In the future, a longer historical record of remote sensed data and meteorological data are required to establish a more stable regression rules.

5. PROPOSED SYSTEM

In this project, we try to integrate multi-source information to achieve the purpose of accurately monitoring agricultural drought. Agricultural drought generally occurred after a period of time since the cessation of rain, when the available stored water will support the actual evapotranspiration. It involves a variety of meteorological drought characteristics (e.g., the scarcity of precipitation, actual evapotranspiration at only a small fraction of the potential evapotranspiration rate, and the shortage of soil moisture) impact on agriculture (e.g., yield reduction). Therefore, the agricultural drought condition is affected by many factors, such as precipitation, soil moisture, temperature, vegetation type, soil type and phenology. Based on the defined drought criteria, the intensity, temporal and spatial distribution of agricultural drought can be monitored. By using this type of attributes, we are going to predict the drought conditions by using the supervised machine learning methods.

Advantages

- It is more accurate and efficient when compared with the existing system.
- By using the spatial and temporal characteristics of attributes, we can easily identify the drought affected regions.
- It takes less time execution when compared with the existing system.

6 .RESULTS & DISCUSSION

A. METEOROLOGICAL DATA BASED DROUGHT ANALYSIS

The computation of rainfall departure was carried out in various region using annual and monthly rainfall data. Annual rainfall departure (using equation 2) analysis indicated that the annual rainfall deficiency during the drought year had been varied from 38% in 1923 to 28% in 2011. It is seen from the above analysis that none of the districts were severely affected by droughts, however, maximum number of years are lying in the range of mild to moderate drought.

B. HYDROLOGICAL DATA BASED DROUGHT ANALYSIS

ISDI for the period of 2000-2009 is calculated for all the districts and the stream flow data was collected from different gauging stations of the study area. Keeping this in view, we carried out hydrological analysis of discharge data lying in four districts only. From the analysis it is clearly seen that 2009 and 2000 are the years when maximum number of stations are witnessed the drought situation according to SDI values.

C. REMOTE SENSING BASED DROUGHT ANALYSIS

Normalized Difference Vegetation Index (NDVI) a measure of the “greenness,” or vigor of vegetation. It is derived based on the known radiometric properties of plants, using visible (red) and near-infrared (NIR) radiation. NDVI images are analyzed for various region of different months of the year 2008 to 2010 and 2014.

7. CONCLUSION & FUTURE ENHANCEMENT

A. CONCLUSION

Our bayesian model attempts to characterize drought in a more accurate and comprehensive way. Cross-validation analysis proved that the regression accuracy is very high and the drought condition can be accurately portrayed using the input variables. This model application using a variety of methods and data, there is still some work to be done in our future research because of the complex spatial and temporal characteristics of drought. To overcome imitation, it assesses performance of measuring the drought by using the Spatial and temporal characteristics of information’s. We collect the dataset from different regions and also collect the time varying information’s like that. In this, we predict the drought conditions by using the supervised learning mechanism. It can be implemented by using the Bayesian supervised machine learning algorithm. Through this we can achieve the accuracy and performance, and also we can improve the performance and effectiveness.

B. FUTURE ENHANCEMENT

We will take more in-depth validation and assessment in our future research with the field observation data increasing and improvement. Multi-year, multi-seasonal meteorological and observed vegetation condition and soil moisture data at different sites and region will be used to assess the ISDI performance by using the different set of environmental characteristics.

REFERENCES

[1] Altannar Chinchulunn, Petros Xanthopoulos, Vera Tomaino, P.M.Pardalos, “ Data Mining Techniques in Agricultural and Environmental Sciences”, International Journal of Agricultural and Environmental Information Systems,1(1),26-40,January-June 2010

[2] J. Brown et al., “The Vegetation Drought Response Index (VegDRI): A new integrated approach for monitoring drought stress in vegetation,” GIS Remote Sens., Vol. 45, pp. 16–46, 2008.

[3] S.S.Baskar, L.Arockiam, V.Arul Kumar, L.Jeyasimman, “Brief Survey of Application of Data Mining Techniques to Agriculture”, Agricultural Journal 5(2): 116:118,2010 ISSN:1816-9155

[4] Prof. Chandrakanth. Biradar, Chatura S Nigudgi, “ An Statistical Based Agriculture Data Analysis”, International Journal of Emerging Technology and Advanced Engineering Vol 2, Issue 9, September 2012

[5] Chih-Hung Hsu, An-Yuan Chang, Hui-Ming Kuo, “Data Mining QFD for The Dynamic Forecasting of Life Cycle under Green Supply Chain”, WSEAS Transactions On Computers Issue 1, Vol 11, January 2012

[6] D. Diepeveen and L. Armstrong,” Identifying key crop performance traits using data mining”, IAALD AFITA WCCA2008 World Conference On Agricultural Information And IT

[7] Elodie Vintrou, Dino Ienco, Agnès Bégué, and Maguelonne Teisseire, “Data Mining, A Promising Tool for Large-Area Cropland Mapping”, IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing, Vol. 6, No. 5, October 2013

[8] Han-Wen Hsiao,Meng-Shu Tsai, And Shao-Chiangwang, “Spatial Data Mining of Colocation Patterns for Decision Support in Agriculture”, Asian Journal of Health and Information Sciences, Vol. 1, No. 1, pp. 61-72, 2006G. Holmes, D. Fletcher and P. Reutemann, “An application of data mining to fruit and vegetable sample identification using Gas Chromatography-Mass Spectrometry”, International Environmental Modelling and Software Society (iEMSs) 2012

[9] Monitoring: Theory, Validation, and Application in Mid-Eastern China”, IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing, Vol. 6, No. 3, June 2013