

Drought Prediction and Management using Big Data Analytics

Himani Shah
Department of Computer
Engineering
VESIT, Mumbai

Vinita Rane
Department of Computer
Engineering
VESIT, Mumbai

Jayesh Nainani
Department of Computer
Engineering
VESIT, Mumbai

Benita Jeyakumar
Department of
Computer Engineering
VESIT, Mumbai

Nupur Giri, PhD
Head of Department
Department of
Computer Engineering
VESIT, Mumbai

ABSTRACT

The prediction of occurrence of droughts has been a challenging task. However, it is necessary that prediction is done with at most accuracy to prevent loss of life and property. Based on the previous year's rainfall, temperature and evapotranspiration data, DDI will be calculated which will be based on SPI, SPEI, PDSI, PHDI and ZIND indices. This proposed index will be trained using random forest algorithm and the output will help to predict the severity of drought for the upcoming years. Also, round robin algorithm with dynamic quantum size is used for resource allocation for the victims of drought affected areas.

General Terms

SPI, SPEI, PDSI, ZIND, PHDI, GAHP, Round Robin

Keywords

Drought prediction, rainfall, drought indices, prediction models, resource allocation

1. INTRODUCTION

Big data analytics is useful for advanced analytic techniques against very large, diverse data sets that include different types of data such as structured/ unstructured and streaming/batch. Better and faster decisions can be made by analysts, researchers, and business users by analyzing big data. Using these techniques, precise drought prediction and resource allocation to drought prone areas is possible.

Disasters involve widespread human, material, economic or environmental impacts, which exceed the ability of the affected community or society to cope using its own resources. Effective management of crises and disasters is a global challenge. All communities are vulnerable to crisis, both natural and induced by human activities. Disaster management system includes:

Disaster prevention

It is designed to provide permanent protection from disasters. Few Man-Made disasters can be prevented by adopting various prevention methods but most of the natural disasters cannot be prevented. With smart evacuation plans, strategic planning and design standards for environment, the risk of loss of life can be reduced.

Disaster preparedness

Being prepared mitigates the impact disasters can have on our lives. This can reduce the loss of life and property by removing people from threatened location. Also, it includes facilitating people living in disaster prone area with effective rescue, relief and rehabilitation.

Disaster relief

It is a coordinated effort to help disaster affected victims by providing activities like rescue, taking care of health and hygiene, providing food, shelter and water, preventing epidemics, disease breakouts and disability, repairing vital services such as providing temporary shelter houses and emergency first aid.

Disaster recovery

After the emergency needs have been fulfilled and the initial crisis is over, disaster recovery phase begins which include rebuilding infrastructure, health care and rehabilitation.

This paper focuses on disaster preparedness and disaster relief modules for drought management system.

Severe drought conditions and extreme heat have caused acute health problems in various part of the world. Also, prolonged drought conditions have led to total loss of crops in some regions. Drought prediction system will help government to pacify the effect of disaster. Some of the benefits are that farmers will be able to predict which crop must be grown in that area, government can provide pre-disaster aid, water can be drawn to these regions. The modules include:

A. Disaster preparedness (Prediction)

From the available dataset of previous years' rainfall records in an area, prediction of the severity of occurrence of drought in that area has been done.

B. Disaster relief (Resource Allocation)

After getting an approximate count of the affected people in such drought prone areas, relief packets are distributed.

2. LITERATURE REVIEW

2.1 Drought Indices

There are various drought indices that have been compared to find out the most accurate and reliable index to be used for drought prediction.

[6] states that Palmer Drought Severity Index (PDSI) uses readily available temperature and precipitation data to estimate relative dryness. It is effective in determining long-term drought, especially over low and middle latitudes but lacks multi-timescale features of indices, making it difficult to correlate with specific water sources like runoff, snowpack, reservoir storage, etc. [7] Describes variants of palmer indices and overcomes PDSI's drawbacks. PHDI requires monthly temperature and precipitation information. ZIND uses soil data and total water balance methodology. They predict short-term conditions better than PDSI and are calculated for much shorter timescales, enabling it to identify rapidly developing drought condition. It allows the total water system to be considered because of its water balance approach. [3] analyses normalized difference vegetation index (NDVI) which is a simple graphical indicator that can be used to analyze the remote sensing measurements, from space platform, and assesses whether the target being observed contains live green vegetation. [1] states that SPI is based on the probability of precipitation for any of the time scales. It takes only precipitation data into consideration. It is simple and versatile as it can be calculated on different time scales. [6] is a variant of SPI which considers evapotranspiration data along with rainfall data. Its other features like multi timescale, and versatility are similar to SPI.

2.2 Prediction Models to Determine Values of SPI

In the papers that have been reviewed and analyzed, many prediction models have been studied.

[3,8,9] did analysis of CART model. It has been concluded that CART is nonparametric (no probabilistic assumptions), it automatically performs variable selection by eliminating the non-significant variables and focuses on the more significant variables. But the model is a step function and does not give continuous values. The decision trees formed are unstable, that is, arbitrary data fluctuations could result in entirely different trees. Most importantly, CART does a poor job of modelling linear structure.

[2] compared AR and GAHP Models, by checking the performance of each model. AR model computes values faster and is a simple model to work but SPI-1 time series could not be predicted using this model as it had a white noise power spectrum. Even though AR model predicted precipitation and then calculated SPI values, it failed as it is observed that AR just extracts the seasonal cycle, thus preventing to have a good knowledge of large deviations from it. The accuracy of this model was lower than GAHP model which overcame the shortcomings of AR model. It predicted the precipitation for a month as the value that was most probable which was described by the probability density function of the precipitation for that month. [2] states that in the research they performed, GAHP method appeared to perform a better

prediction when seasonal variability and deviation are both considered.

[8,9] compared random forest algorithm with many other algorithms like CART, k-means, k-Nearest Neighborhood (kNN), support vector machines (SVMs), gradient boosted model (GBM), Naive Bayes classification etc. [9] inferred that the tree based algorithms are better than logistic or any line fitting techniques. [8] found that random forest is the best algorithm among all these algorithms and gives the best accuracy.

2.3 Resource Allocation Algorithms:

[10] compared various resource allocation algorithms and it was found that round robin was the best algorithm. Other algorithms were simpler but they suffered either from starvation problem or resource allocation was done inefficiently in those algorithms. The only drawback of round robin was that the appropriate quantum size must be selected for proper resource allocation. If the quantum size chosen is small then the time complexity of algorithm increases and if the quantum size is large then the allocation is not done optimally.

[4, 5] have proposed dynamic resource allocation technique by calculating quantum size dynamically. [4] calculates quantum size by taking the mean of all the required resources and allocate resources accordingly. [5] selects quantum size dynamically with help of median and highest resource allocation request size. These algorithms reduce the number of context switches, average waiting time and average turnaround time.

3. PROPOSED MODEL

In this model, mainly two functions are being carried it out:

3.1 Analysis and prediction of data to determine occurrence of drought

The analyst analyses several different parameters like temperature, precipitation and potential evapotranspiration to compute the Drought Determining Index(DDI). The DDI is then used to determine occurrence of drought. Based on previous years' data for 100 years, the values for current year is computed leading to prediction of drought. Various graphs like Resources allocated to people vs total number of people in an area, Year vs DDI, etc. are also generated.

3.2 Resource donation and allocation

A registered donor logs in using his credentials (username, password) and enters the amount he wishes to donate. Resources are accepted in the form of cash. These resources are allocated when admin of any area requests resources after the occurrence of drought. Resources will also be allocated to districts based on population and drought intensity.

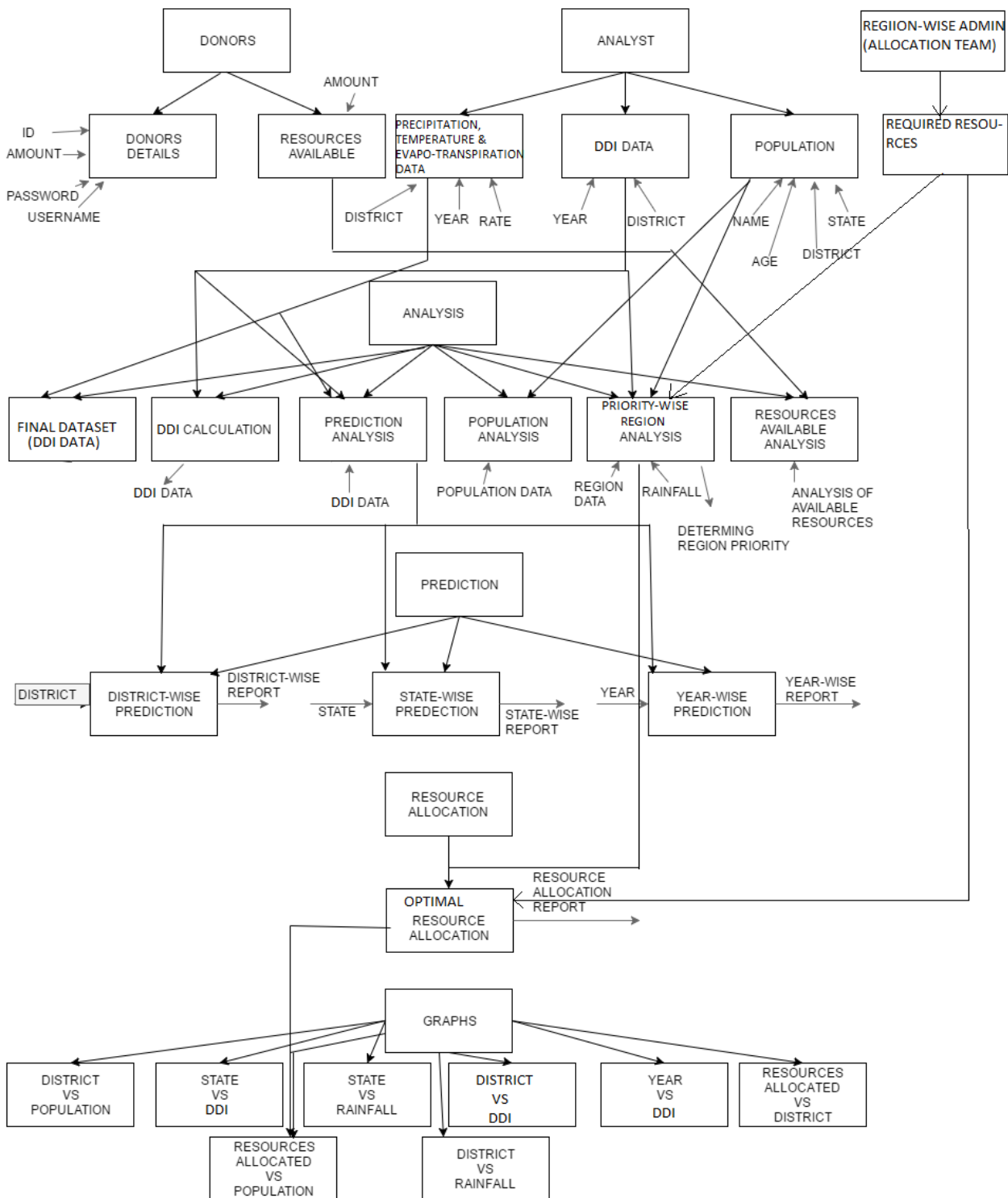


Fig 1: System block diagram

4. PROPOSED METHODOLOGY

After careful study and analysis of the above-mentioned papers, DDI as the drought index for determining occurrence of drought, Round robin algorithm for resource allocation and Random forest algorithm has been proposed for predicting DDI values owing to their higher accuracy.

4.1 Drought Determining Index (DDI)

DDI index will be a combination of Standardized Precipitation Index (SPI), Palmer drought severity Index (PDSI)[6], Standardized Precipitation Evapotranspiration

Index(SPEI), Palmer Hydrological Drought Index(PHDI) and Palmer Z-index(ZIND).

Standardized Precipitation Index (SPI) will be used as it is a standard method used by all the countries and it gives the most accurate results as compared to other models, and also calculates index for different timescales, but it only takes precipitation data into consideration. SPEI is a variation of SPI and also considers potential evapotranspiration. It can determine resistance of a region to drought. PDSI and its variants PHDI and ZIND shall be used for calculating long

term droughts, it uses both precipitation and temperature data for predicting drought.

4.2 Round Robin resource allocation technique

The data obtained from region administrator will be sorted based on priority (population and severity of drought). Then round robin technique will be used for resource allocation as this technique is better than all the above-mentioned techniques. In this technique, the resource allocation is done optimally as all the districts will receive at least some resources. The drawback of this technique will be overcoming by using finest time quantum which will calculate quantum size dynamically with the help of median and highest resource allocation request size. The formula for calculating Quantum size (number of resources allocated to the districts at any time):

Quantum size = Ceil (sqrt (median * highest resource allocation request value)) [5]

As the value of quantum size is dynamically calculated, the allocation of resources is done optimally.

4.3 Random Forest algorithm

Random forest is a supervised learning technique. It can perform both regression and classification tasks. It can work with missing values and can also handle outliers.

In this method, multiple trees are generated, and during classification each tree receives votes. The tree which receives the most votes is chosen at the end by this algorithm. And in case of regression it takes average of outputs of different trees.

As it is a tree based algorithm, the prediction is highly accurate, stable and easy to interpret.

5. CONCLUSION

In this paper, a system for drought prediction and resource allocation system has been proposed. Comparison between various drought indices and prediction models has been done by stating their advantages and disadvantages and finally DDI drought index which is a combination of several other drought indices has been proposed and random forest prediction model has been chosen for this system. Resource allocation algorithms have been compared and round robin algorithm with dynamic quantum size has been chosen. The overall model design has been proposed. Successful implementation of this model will be helpful for people to placate the effects of droughts and to optimally allocate resources to benefit the drought prone victims.

This project can be helpful for farmers as based on the severity of droughts the farmers can decide which crops to grow. Also the overall GDP will grow as agriculture is the most important contributor to GDP in most of the countries. If any regions drought type is changing from mild to severe, then precautions can be taken well in advance by growing trees in those areas. Plan can be made by the government to send more water to drought affected areas.

6. REFERENCES

- [1] Book on "Standardized Precipitation Index", WMO-No. 1090 © World Meteorological Organization, 2012
- [2] Isabella Bordi, Marcello Petitta and Alfonso Sutera, "Methods for predicting drought occurrences", Department of Physics, University of Rome "La Sapienza", P.le Aldo Moro 2 00185 Rome, Italy
- [3] Getachew Berhan, Shawndra Hill, Tsegaye Tadesse, and Solomon Atnafu, "Drought Prediction System for Improved Climate Change Mitigation", IEEE Transactions on Geoscience and Remote sensing, Vol. 52, No. 7, July 2014
- [4] IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 1, May 2011, ISSN (Online): 1694-0814
- [5] International Journal of Latest Trends in Engineering and Technology (IJLTET), ISSN: 2278-621X
- [6] Amin Zargar, Rehan Sadiq, Bahman Naser, and Faisal I. Khan, "A review of drought indices", University of Delaware on 6, October 2014
- [7] Lampros Vasiliades and Athanasios Loukas, "Hydrological response to meteorological drought using the Palmer drought indices in Thessaly, Greece", Article in Desalination, February 2009
- [8] D. Zakirov, A. Bondarev and N. Momtselidze, "A comparison of data mining techniques in evaluating retail credit scoring using R programming," 2015 *Twelve International Conference on Electronics Computer and Computation (ICECCO)*, Almaty, 2015, pp. 1-4
- [9] Richard Derrig, Ph.D. and Louise Francis, FCAS, MAAA, "Distinguishing the Forest from the TREES: A Comparison of Tree Based Data Mining Methods" Casualty Actuarial Society Forum, Winter 2006
- [10] Imran Qureshi, "CPU Scheduling Algorithms: A Survey", International Journal of Advanced Networking and Applications (IJANA) Volume: 05, Issue: 04, Pages:1968-1973 (2014) ISSN : 0975-0290