International Journal of Scientific Research & Engineering Trends
Volume 3, Issue 6, Nov-Dec.-2017, ISSN (Online): 2395-566X

An Unsupervised TLBO Based Drought Prediction By Utilizing Various Features

Shikha Ranjan Patel
Dept. of Computer Science and Engineering
Bansal Institute of Science and Technology,
Bhopal, India
shikharanjanpatel@gmail.com

Prof. Priyanka Verma
Dept. of Computer Science and Engineering
Bansal Institute of Science and Technology,
Bhopal, India

***Abstract* – Agricultural vulnerability is generally referred to as the degree to which agricultural systems are likely to experience harm due to a stress. In this work, an existing analytical method to quantify vulnerability was adopted to assess the magnitude as well as the spatial pattern of agricultural vulnerability to varying drought conditions. Based on the standardized precipitation index (SPI) was used as a measure of drought severity. A number of features including normalized difference vegetation index (NDVI), vegetation condition index (VCI), and SPI will be use for classification. Here proposed modal use Teacher Learning Based Optimization genetic approach for classify the different location present in geospatial dataset. By use of this TLBO approach prior knowledge is not required. Experiment results shows that proposed work is better as compare to previous work**.

***Keywords* – Drought condition, Rainfall Estimation, Prediction, Precipitation.**

I. INTRODUCTION

A drought should not be confused with arid climate which is a general feature of a low rainfall area, or with heat waves which usually are on the time scale of days to weeks. Due to the importance of the availability of water for human development droughts have played an important role in the. raise or downfall of civilizations. Current population growth demands an increased intensity in the use of agricultural resources which then leads to an increased demand in water and makes the system generally more susceptible to variations in the water supply (Mishra & Singh, 2010). Thus, it becomes more and more important to understand the occurrence of droughts, recognize patterns, and to enhance their predictability.
Droughts occur when one element of the water cycle starts to behave abnormally compared to the others. Cold temperatures in the mountains can delay the runoff of melted snow which may be needed for the growing season in the flatland. A heat wave can increase evapotranspiration and dry out lakes. Reduced precipitation can let a groundwater reservoir run dry if it is continuously.

By using climate simulation, which is done in this thesis. Advancements in recent years make it possible to realistically run a model world over long time periods and use this data as tested for theoretical experiments. A main interest lies in severe droughts that occur every few hundreds of years and thus, cannot be studied using observations. A climate simulation over more than 1000 years makes these events accessible for investigations and risk analysis. The data produced by climate models cannot be compared directly to real measurements. On short time scales the climate system is dominated by internal variability and chaotic behavior. A system following the same physics, and starting with the same boundary conditions would not show the same temporal development, i.e., the system is not deterministic.

Fig. 1: Left: A farm in the Australian Outback at the end of a decade-long drought. A field of dry cornstalks shows during the drought of August 2013 in Texas.

On longer time scales the internal variability averages out, and hence, the mean, variability and geographical patterns of an observable are the interesting quantities determined by physical processes that can be compared to simulations.

II. Related Work

In [15] a dry part of the world and extensively suffers from drought. Drought is a natural, temporary, and iterative phenomenon that is caused by shortage in rainfall, which affects people's health and well-being adversely as well as impacting the society's economy and politics with far-reaching consequences. Information on intensity, duration, and spatial coverage of drought can help decision makers to reduce the vulnerability of the drought-affected areas, and therefore, lessen the risks associated with drought episodes. One of the major challenges of modeling drought is unavailability of long-term meteorological data for many parts of the country. Satellite-based remote sensing data—that are freely available—give information on vegetation conditions and land cover. In this work, classification of drought conditions is done by utilizing the standardized precipitation index (SPI) was used as a measure of drought severity. A number of features including normalized difference vegetation index (NDVI), vegetation condition index (VCI), and temperature condition index (TCI). So this classification required large amount of data for classification. So here one fast an efficient approach is required to make multiclass classification of different geographical locations.

Jaclyn F. Brown et al [7] (2014) "The Vegetation Drought Response Index (VegDRI): A New Integrated Approach for Monitoring Drought Stress in Vegetation" The development of new tools that provide timely, detailed spatial-resolution drought information is essential for improving drought preparedness and response. This paper presents a new method for monitoring drought-induced vegetation stress called the Vegetation Drought Response Index (VegDRI). VegDRI integrates traditional climate based drought indicators and satellite-derived vegetation index metrics with other biophysical information to produce a 1 km map of drought conditions that can be produced in near-real time.

Hua Xie et al [8] (2014) "Droughts in Pakistan: a spatiotemporal variability analysis using the Standardized Precipitation Index" They investigated the spatiotemporal variability of drought incidence in Pakistan during 1960–2007 by calculating Standardized Precipitation Index fields for 3-, 6- and 12- month scales using gridded precipitation data. Principal component analysis revealed that droughts are wide-spread and often occur simultaneously over large areas. Furthermore, spectral analysis identified a 16-year drought recurrence period.

In [3] 2015 develop a conceptual prediction model of seasonal drought processes based on atmospheric/oceanic Standardized Anomalies (SA). Empirical Orthogonal Function (EOF) analysis was firstly applied to drought-related SA of 200 hPa/500 hPa geo-potential height (HGT) and sea surface temperature (SST), respectively. Subsequently, SA-based predictors were built based on the spatial configuration of the first EOF modes. This drought prediction model is essentially the synchronous statistical relationship between 90-day-accumulated atmospheric/oceanic SA-based predictors and 3-month SPI 15 (SPI3), calibrated by the simple method of stepwise regression. It is forced by seasonal climate forecast models like the NCEP Climate Forecast System Version 2 (CFSv2). It can make seamless drought prediction for operational use after being calibrated year-by-year.

III. Proposed Work

In order to make a general model which works on various available data Indices a Teacher Learning Based Optimization (TLBO) genetic algorithm is use is required to be modal. Here different indices values or data act as the input vector for the TLBO. So whole work will focus to collect data then pre-process whole information and classify the data by using the TLBO.

In this work, we used a number of well-known vegetation indices, which are described in the following.

3.1. Normalized Difference Vegetation Index: NDVI is the most commonly used vegetation index accounting the amount of vegetation cover in the land. NDVI was first suggested as an index of vegetation health and density. It is calculated as

$$N = \frac{b_{NIR} - b_{RED}}{b_{NIR} + b_{RED}}$$

where N is the NDVI and bNIR and bRED are the reflectance in the NIR and red bands, respectively.

3.2. Vegetation Condition Index: VCI was suggested by Kogan [39], which shows how close the NDVI of the current month is to the minimum NDVI calculated from the long-term record. It is calculated as

$$V_j = \left(\frac{N_j - N_{\min}}{N_{\max} - N_{\min}}\right) \times 100,$$

where Vj is the VCI value of month j, and Nmax and Nmin are, respectively, the maximum and the minimum values of NDVI that are calculated from a long-term record for that month (or week) and j is the index of the current month (week). The condition/health of the ground vegetation presented by VCI is measured in percent and may serve as an approximate measure of how dry the current month.

3.3. Pre-Processing: Data preprocessing reduces the size of the input data significantly. It involves activities like cleaning of data and convert in required environment format.

3.4. Create Vector and Generate Population: The pre-processed data is use for collecting feature of that index. Now feature vector is create by the use different index where each position in the vector is fix for various index. Here each vector act as a individual chromosome and group of these is population.

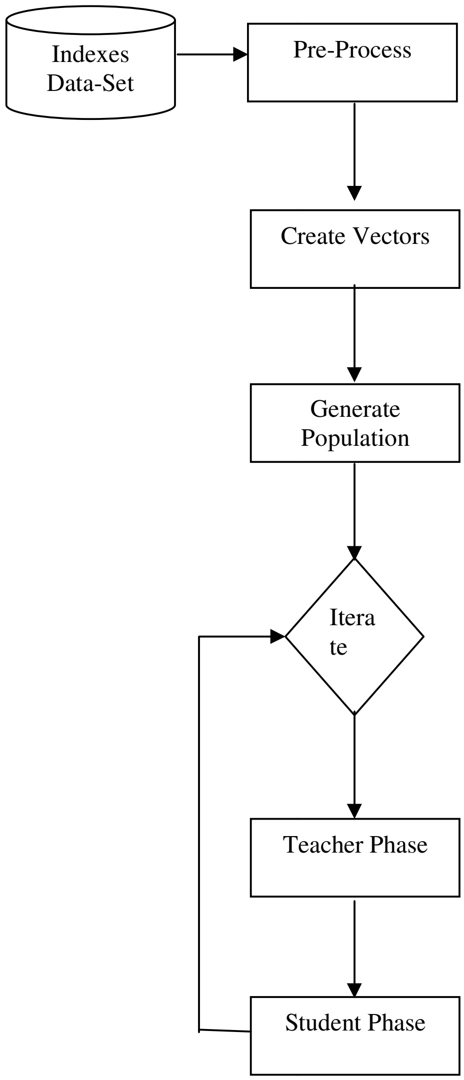


Fig.2 Represent Block Diagram of Proposed Work.

Here assume some cluster set that are the combination of different index value. This is generate by the random function which select any vector as a centroid of the cluster. This can be understand as let the number of centroid be Cn then one of the possible solution is {C1, C2, Cn}. In the similar fashion other possible solutions are prepared which can be utilize for creating initial population.

3.5. Teacher Phase: For finding difference two functions are use first is Eludician Distance formula other is cosine similarity function The Euclidean distance d between two solutions X and Y is calculated by

d = [SUM ((X-Y).^2)]^ 0.5

The Cosine distance d between two vectors X and Y is

d = [1 – (X*Y' / √(X*X')* (Y*Y'))]

Following Step will find distance between the documents

```
Loop x = 1:S T      // This is for finding the teacher in the Initial Population
    Loop n = 1: N
        D [n, x] = Dist (Ds[n], x) // Here Dist is a Distance function Like Cosine/ Eludician
    End Loop
End Loop
```

So the matrix D contain all the values of the centriod distance from the vectors then find the minimum distance which will evaluate specify best possible solution.

```
S←Sum (D)        // Sum matrix row wise
[V I]←Sort(S)    // Sort matrix in increasing order
```

Top possible solution after sorting will act as the teacher for other possible solutions. Now selected teacher will teach other possible solution by replacing fix number of centroid as present in teacher solution. By this all possible solution which acts as student will learn from best solutions which act as teacher.

A good teacher is one who brings his or her learners up to his or her level in terms of knowledge. But in practice this is not possible and a teacher can only move the mean of a class up to some extent depending on the capability of the class. This follows a random process depending on many factors. Let Mi be the mean at any iteration i. The teacher will try to move mean Mi towards his/her own level so the new mean will be designated as Mnew. The solution is updated according to the difference between the existing and the new mean given:

Difference Meani = ri * (Mnew – Tf *Mi)

Where Tf = teaching factor.
Teaching factor (Tf) decides the value of mean to be changed, and ri is a random number in the range [0, 1]. Tf is not a parameter of the algorithm and its value is not given as an input to the algorithm. The value of TF can be either 1 or 2, which is a heuristic step and decided randomly with equal probability as,

TF = round [1 + rand(0, 1)]

This difference modifies the existing solution according to the following expression
Xnew,i = Xold, i + Difference Meani,i

Where Xnew,i is the updated value of Xold,i. Accept Xnew,i if it gives better function value.

International Journal of Scientific Research & Engineering Trends
Volume 3, Issue 6, Nov-Dec.-2017, ISSN (Online): 2395-566X

3.6 Student Phase: In this phase all possible solution after teacher phase are group for self learning from each other. This can be understand as let group contain two student then each student who is best as compare to other will teach other solution. Teaching is similar as done in teacher phase, here replacing fix number of centroid is done which is similar as in best student of the group.

For i = 1: Pn // Randomly select two learners Xi and Xj, where i is not equal to j
If f (Xi) < f (Xj)
X new, i = Xold, i + ri (Xi − Xj) (for a minimization problem)
Else
Xnew, i = Xold, i + ri (Xj − Xi)
End If
End For

Accept Xnew if it gives a better function value. Once student phase is over then check for the maximum iteration for the teaching if iteration not reach to the maximum value then GOTO step of teacher phase else stop learning and the best solution from the available population is consider as the final centroid of the work. Now documents are cluster as per centroid.

IV. Experiment and Results

In order to conduct experiment and measure evaluation results MATLAB 2012a version software is use. This section of paper show experimental setup and results. The tests were performed on an 2.27 GHz Intel Core i3 machine, equipped with 4 GB of RAM, and running under Windows 7 Professional.

4.1. Data set: Experiment is done on real dataset having from the NASA website: http://nassgeodata.gmu.edu/VegScape/ where VCI and SPI values for the feature extraction are taken. While NDVI value is taken from the http://drought.unl.edu/MonitoringTools/ClimateDivisionSPI/ArchivedSPIMaps.aspx

4.2. Evaluation Parameter: Comparison of proposed work is done on the basis of following parameters.

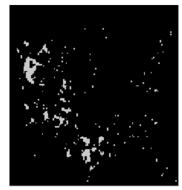
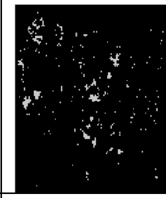
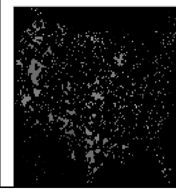
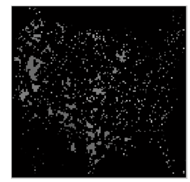
Precision = TP / (TP +FP)

Recall = TP / (TP +TN)

F_ score = 2 *Precision *Recall / (Precision +Recall)

Here TP : True Positive
TN: True Negative
FP : False Positive

4.3 Results

Classified Image comparison		
	Proposed Work	**Previous Work**
Set1		
Set2		

Results are compare with the previous work (SGM) Sequential Generative Model in [1] which is term as previous work in this paper.

Table 1: Precision value comparison of previous work with proposed work different sets.

Values	**Precision Values**	
	Previous Work	**Proposed Work**
Set1	0.4244444	0.696049
Set2	0.5124	0.748889

It has been observed by table 1 that proposed work of prediction works well as compare to the previous method adopt in [1]. One more important factor is observed as the data size images vary training required in the previous work which is unsupervised process.

Table 2: Recall value comparison of previous work with proposed work different sets

Values	**Recall Values**	
	Previous Work	**Proposed Work**
Set1	0.2916667	0.69881
Set2	0.3245	0.417251

© 2017 IJSRET

It has been observed by table 1 that proposed work of prediction works well as compare to the previous method adopt in [1]. One more important factor is observed as the data size images vary training required in the previous work which is unsupervised process.

Table 3: F-Measure value comparison of previous work with proposed work different sets

Values	F-Measure Values	
	Previous Work	Proposed Work
Set1	0.3385965	0.697427
Set2	0.3651	0.535913

It has been observed by table 1 that proposed work of prediction works well as compare to the previous method adopt in [1]. One more important factor is observed as the data size images vary training required in the previous work which is unsupervised process.

Table 4: Accuracy value comparison of Previous work with proposed work different sets

Values	Accuracy Values	
	Previous Work	Proposed Work
Set1	36.79825	58.3821
Set2	26.356	37.6868

It has been observed by table 4 that proposed work of prediction works well as compare to the previous method adopt in [1]. One more important factor is observed as the data size images vary training required in the previous work which is unsupervised process.

V. Conclusion

With the drastic increase of the digital data on the servers, libraries it is important for researcher to work on it. Considering this fact work has focus on one of the issue of the drought geospatial classification which is build by the different indices such as TCI, VCI, etc. Here many researchers has already done lot of work based on neural network classification. In few work document classification are done on the basis of the background information, but this work overcome this dependency as well here it classify all the data without having prior knowledge. As propose work give an distance efficiency value which is quite impressive as well as in data classification it give effective results.

References

[1]. Amir Aghakouchak. "A Multivariate Approach For Persistence-Based Drought Prediction: Application To The 2010–2011 East Africa Drought". 0022-1694/ 2014 Elsevier.

[2]. Hua Xie *Et Al*, "Droughts In Pakistan: A Spatiotemporal Variability Analysis Using The Standardized Precipitation Index", 2014.

[3]. Zhenchen Liu1 , Guihua Lu1 , Hai He1 , Zhiyong Wu1 , Jian He. "A Conceptual Prediction Model Of Seasonal Drought Processes Using Atmospheric And Oceanic Standardized Anomalies: Application In Four Recent Severe Drought Events In China". Hydrol. Earth Syst. Sci. Discuss., Doi:10.5194/Hess-2017-136, 2017.

[4]. Albert J.Peters, Elizabeth A.Walter-Shea, Lei Ji, Andres Vina, Michael Hayes, Mark D.Svoboda, "Drought Monitoring With Ndvi Based Standardised Vegetation Index", 2002.

[5]. Hua Xie Et Al, "Droughts In Pakistan: A Spatiotemporal Variability Analysis Using The Standardized Precipitation Index", 2014.

[6]. Richard R., Heim Jr , "A Review Of Twentieth-Century Drought Indices Used In The United States", 2002.

[7]. Jesslyn F. Brown. "The Vegetation Drought Response Index (Vegdri): A New Integrated Approach For Monitoring Drought Stress In Vegetation" 2014.

[8]. Mishra, A. K. & Singh, V. P. (2010). A Review Of Drought Concepts. Journal Of Hydrology, 391, 202–216.

[9]. Hannaford, J., Lloyd-Hughes, B., Keef, C., Parry, S., & Prudhomme, C. (2011). Examining The Large-Scale Spatial Coherence Of European Drought Using Regional Indicators Of Precipitation And Streamflow Deficit. Hydrological Processes, 25, 1146–1162.

[10]. Shaman, J. & Tziperman, E. (2011). An Atmospheric Teleconnection Linking Enso And Southwestern European Precipitation. Journal Of Climate, 24, 124–139.

[11]. Mahdi Jalili, Member, Ieee, Joobin Gharibshah, Seyed Morsal Ghavami,Mohammadreza Beheshtifar, And Reza Farshi . "Nationwide Prediction Of Drought Conditions In Iran Based On Remote Sensing Data". Ieee Transactions On Computers, Vol. 63, No. 1, January 2014.

© 2017 IJSRET