International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 4 (2019) pp. 925-929

Prediction of Extreme Weather Events using Machine Learning Technique

Dr.C.S.KanimozhiSelvi
Associate Professor,
Dept. of Computer Science and Engineering,
Kongu Engineering College, Perundurai,
Tamil Nadu, India.

G.Sowmiya
PG Scholar,
Dept. of Computer Science and Engineering,
Kongu Engineering College, Perundurai,
Tamil Nadu, India.

Abstract

In climate applications, an event is an instance in time when a significantand persistent change occurs. Spatiotemporal Data Mining is used in the department of meteorology for the prediction of all kinds of weather data and global changes. The main problem is prediction of extreme weather events over a particular region. This system will predict theweather based on parameters such as air temperature, relative humidity, specific humidity, U-wind, V-wind and omega.

Here Anomaly frequency method (AFM) technique is used for extracting the anomalous weather events over a particular region. The AFM is an efficient technique in extracting the features which discriminate extreme events and non-extreme events. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is used to find the co-variation among different climate extremes and their relationship to other climate phenomena. DBSCAN and K-Means clustering algorithmsare performed on weather parameters and the extreme and non-extreme weather events are visualized.

The analysis techniques like homogeneity, completeness, V-measure, Adjusted Rand Index (ARI), Adjusted Mutual Information (AMI) are used to calculate the accuracy for all the parameters. Also the results are validated with the realtime weather events like BOB, Thane and Vardah.

Keywords: Extreme weatherevents; Anomaly frequency method; Density-Based Spatial Clustering of Application with Noise

INTRODUCTION

India is an agricultural country. The economy and a lot of its growth strictly depend on the vagaries of the weather and specifically the extraordinary weather events.Developing countries are liable to extremes of environmental condition variability, and natural process is to increase frequency and magnitude of some extreme weather events and disasters. Extreme weather events are sudden, unusual, unpredictable events that vary from past weather events.Weather event prediction is vital in climate forecasting. Accurate and timely weather event classification is essential for the planning and management of water resources andissue warningsrealted to cyclone, storm etc.

Weather prediction is the vital application in the field of metrology.People all over the world are affected by several extreme weather events like rainfall, droughts, high pressure, stirring storms, tornados etc. Nowadays prognostication is entirely supported numerical weather prediction techniques.The various methods used in prediction of weather are Synoptic weather prediction, Numerical weather prediction, Statistical weather prediction.Among these strategies the statistical weather prediction has several advantages in predicting weather by using numerical information. It uses the past records of weather data and predict the future weather data.

Spatio-Temporal Data Mining (STDM) is a subdivision of data mining which is used for the evaluation of large spatio-temporal databases. Data-mining methods have been proven to be of significant value for spatio-temporal functions. It is a user-centric, bilateral operations, where data-mining professionals and domain specialist work closely together to improve insight on a given issue.

In this paper various weather parameters are analyzed using machine learning technique. The data is preprocessed using Z-score normalization. Then AFM technique is used to extract features of extreme events and the class label is assigned. The DBSCAN and K-means clustering algorithm is used to group the extreme and non-extreme weather events. The performance of the algorithms are evaluated and validated with various real time weather events like BOB, Thane and Vardah.

RELATEDWORK

Piruthevi, C., and CS KanimozhiSelvi proposed a topic Anomaly Frequency Method (AFM) is used for extracting the anomalous weather events over a particular region. AFM method identifies the extreme deviation in the weather parameters. From the obtained anomalous weather events, association rules are found using Apriori, Predictive Apriori and Generalized Sequential Pattern algorithm for the generation of different weather patterns. Data summarization about the extreme weather events is made which results in providing the information about their occurrences based on the deviations found in the parameters used.

Cheng Tang and Claire Monteleoni (2016) proposed a topic-model-based approach to define and detect patterns corresponding to extreme climate-related events over different regions around the globe from the time series data of various climate variables. Inference from this model can be used to construct climate extreme indices, predict disastrous extreme events such as drought and floods, and understand the influence of climate change on climate extremes. Here the discovered topics are correlated with some known climate phenomena or suggest new patterns related to the extreme event.

Grover et al (2015) studied specially the power of making predictions via a hybrid approach that combines discriminatively trained predictive models with a deep neural network that models the joint statistics of a set of weather-related variables. They performed temporal analysis using short- and longer-term features within a gradient-tree based learner. They derived an efficient learning and inference procedure that allows for large scale optimization of the model parameters. They proposed an architecture which combines a bottom-up predictor for each individual variable with a top-down deep belief network that models the joint statistical relationships.

Liu et al (2014) presented some initial investigations on applying DNN to deal with the time series problem in meteorology field which represents the features of the raw weather data layer by layer. The obtained features are employed to predict the weather change in the next 24 hours. The results show that the DNN is able to provide a better feature space for weather data sets, and DNN is also a potential tool for the feature fusion of time series problems.

Munir and Ghosh (2013) proposed the anomaly frequency method (AFM) where the predictors for SVM are identified with the frequency of high anomaly values of weather variables at different pressure levels, which are present before extreme events, but absent for non-extreme conditions. The two-phase SVM model describes that the weather patterns before the extreme rainfall events are different during the daytime and at nighttime. This model predicts all the extreme events well in advance but produces many false alarms.

DATA

The weather information has been taken from the website NCEP/NCAR. The reanalysis data has a spatial grid resolution of 2.5 degrees * 2.5 degrees.The region taken for weather prediction is from latitude of 7.5°N to 15°N to a longitude of 75°E to 90° E which is the land area that covers the Tamil Nadu and the sea region of Bay of Bengal. The atmospheric variables like air temperature, vertical velocity (omega), relative humidity, specific humidity, u-wind and v-wind are taken from 1 January 1948 to 31 December 2017 are downloaded in the format of NetCDF. On daily basis, the data for all the parameters have been taken.

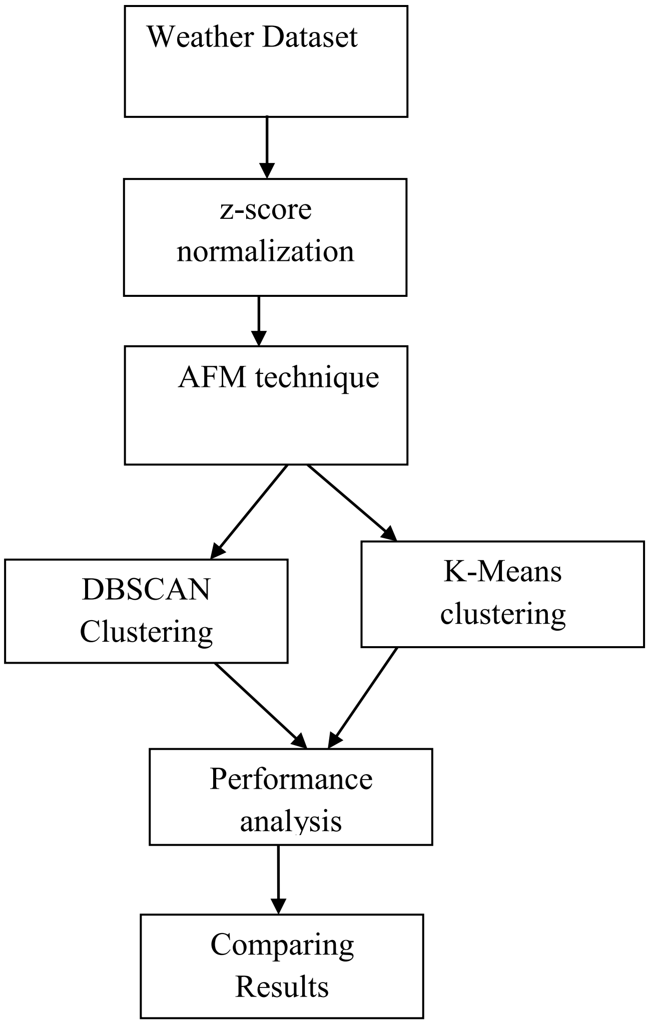
EXTREME WEATHER EVENT PREDICTION

The downloaded data is preprocessed for earlier calculation. This preprocessing involves data retrieval and data normalization. In data retrieval step, the extracted data is in 4D form with latitude, longitude, day and its corresponding value as four co-ordinates. Then this data is converted into 2D data with day and the value of each grid and latitude and longitude as its co-ordinates for easier access of data.The converted data is arranged in the order ie the set of values for a given day for every year are grouped together.

Data normalization means transforming all variables in the data to a specific range. Normalization is a pre processing stage. It can be helpful for the prediction or forecasting purpose**.** The arranged values are then normalized using z-score normalization technique using the formula as shown in equation (1).The standard value for Mean is zero and standard deviation is one

$$Z = \frac{X - \mu}{\sigma}$$

Where Z represent the normalized Z-score value; X represents the corresponding value to be normalized; μ and σ are mean and standard deviation of the set of daily values.



Filtering Extreme Events

The anomalous weather events are extracted by using the Anomaly Frequency Method (AFM). This technique, based on the understanding from figure printing technique the anomaly is defined as deviation of more than 3 times the climatologically standard deviation from climatologicallymean. The positive anomalous threshold and negative anomaly threshold is calculated is shown in equation (2)&(3).

$$\delta^{+} = \bar{X} + 3\overline{SD}$$

$$\delta^{-} = \bar{X} - 3\overline{SD}$$

Where$\bar{X}$represents the climatological mean of a weather variable and $\overline{SD}$represents the climatologically standard deviation of the variable in a particular grid at a specific time instant.The normalized values are then compared with the positive and negative thresholds to spot the anomalous events.The values that exceed the positive threshold are called positive anomalous behavior and which are below the negative anomaly threshold are called negative anomalous behavior.

DBSCAN

DBSCAN is a density-based clustering algorithm.There are two parameters eps and minimum number of points. If the eps is one then draw a circle with the radius of 1 unit and if eps is two then the radius will be slightly large.ε-Neighborhood of an object contains at least MinPts of objects then it is known as high density region.

DBSCAN finds extreme and non extreme weather events. The purposes of eps is used finding the neighborhood of each point, and determine the corepoints with quite min Pts as neighbors. The connected element of core points in the neighbor graph, negate all non-core points. Assign every noncore point to acloseby cluster if the cluster is an ε (eps) neighbor, otherwise assign it to noise.

K-MEANS

K-MEANS algorithm is a well known cluster analysis algorithm in data mining.It is a partitioned clustering approach. Each cluster is associated with centroids.Each point is assigned to the cluster with the closest centroid.Number of clusters K must be specified.

Initial centroids are often chosen randomly that is Clusters produced vary from one run to another. The mean of the points in the cluster is the centroid. Most of the convergence happens in the first few iterations. It has ability to process large datasets.

In K-MEANS algorithm K denotes the number of clusters.The main goal is to Predicting the cluster membership of a new instance. The procedure must be consistent with the modeling approach.After executing the dataset the output will be A label for each data point and the center for each label.

RESULTS AND DISCUSSION

The extreme weather events for a particular region are identified using the anomaly frequency method. The name AFM comes from the frequencies of anomalies it is used to extract the features of extreme events. Thus, AFM is used here for the filtering the extreme weather events. The extreme weather events are labeled as 1 and the non extreme events are labeled as 0.

DBSCAN clustering algorithms is used for visualization of extreme and non extreme weather events from the atmospheric parameters over 69 years. In figures three clusters are estimated and three color are denoted. Black color denotes noise. White and purple color denotes the extreme and non extreme weather events. The values for eps and min points are 0.3 for eps and 20 for min points. Here DBSCAN for air temperature, omega, relative humidity, specific humidity, u-wind, v-wind are visualized along with the noise.

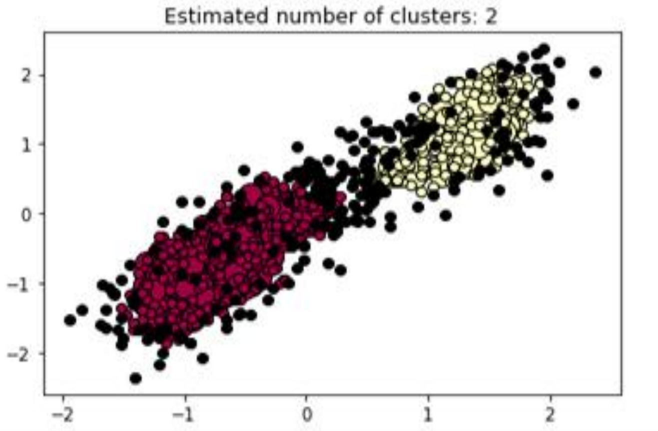


Figure 1. DBSCAN clustering for all parameters

K-means algorithm is also employed to visualize the extreme and non extreme events and to compare the performance with the DBSCAN algorithm.

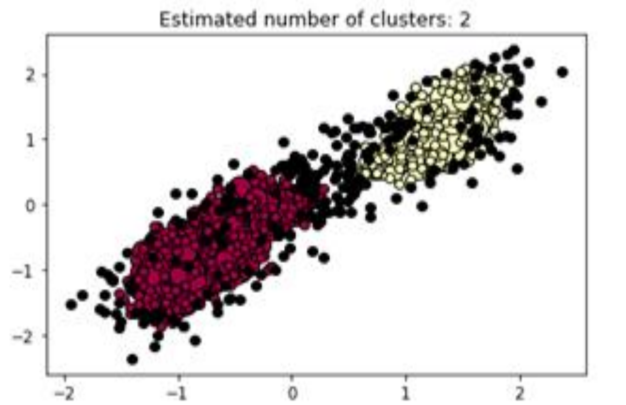


Figure 2. K-MEANS clustering for all parameters

VALIDATING THE EXTREME WEATHER EVENTS

Cyclonic storm BOB 06(08B) (1996)

Cyclonic storm BOB 06 (08B) was a very severe cyclone storm occurred from Nov 26 to Dec 7 in 1996.The intensity is 967 hpa.It was arise from the monsoon still,a specific circulation continued on November 26 near the Andaman and Nicobar Island, followed by a poorly defined area of deportation. On November 30 an increasing ridge to the north turned the storm back to west. Then it is aggravate into a storm on December 2 and on December 3 the relative humidity increases and cause high rainfall in most of the region of Tamil Nadu.On December 6 storm immediately weakened and scattered. The atmospheric values for the year 1996 is extracted from the dataset which includes the cyclone period.DBSCAN and K-means clustering is performed.

The validation is done by using different validation techniques which includes homogeneity, completeness, V Measure, Adjusted Rand Index (ARI) and Adjusted Mutual Information (AMI). The results show that the values obtained after validation are greater than 0.5 which implies that the clusters formed are valid clusters.

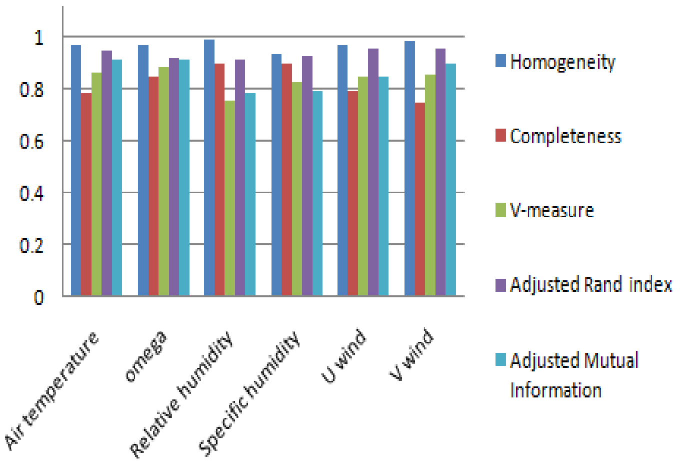


Figure 3. Comparison between DBSCAN and K-MEANS for Cyclonic storm BOB 06 (08B)

Cyclone Thane(2011)

A severe storm cyclone thane was occurred from December 26 to December 31. It was the strongest tropical cyclone of 2011 in North Indian Ocean. In the beginning Thane was developed as a tropical disturbance within the monsoon through the west of Indonesia and it is gently moves toward the northwest.On December 27 there is a rapid decrease in the value of parameter omega at Bay of Bengal.It causes storm with 969hpa it is a lowest pressure. On December 28 in many regions the cyclone started to moves towards west due to raise in V wind and cause heavy rain. The cyclone start depressed on December 30.

In figure 4 the atmospheric values for the year 2011 along with the values for cyclone thane are extracted from the dataset and clustered using DBSCAN and K-means algorithm. The results show that the values for the validation parameter obtained are greater than 0.5 and the clusters obtained are valid and meaningful.

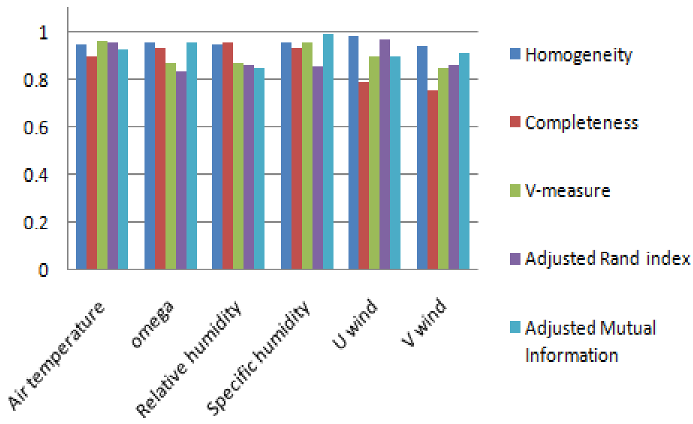


Figure 4. Comparison between DBSCAN and K-MEANS for Cyclone Thane

Cyclone Vardah (2016)

The severe storm vardah was occurred from December 6 to December 19.Vardah was the fourth tropical cyclonic storm of 2016 in North Indian Ocean. This storm struck the Andaman and Nicobar Island and South India. On December 3 the low pressure area near Malay Peninsula and get depress on December 6.This cyclone was similar to the Cyclone Bob06.On December 10 it cause heavy storm and vardah reaches its peak on December 11.On December 12 due to severe storm it causes landfall over the east coast of India.On December 14 the storm entered the Arabian sea.Due to the the warm sea surface temperature the cyclone start depressed on December 19. In figure5 theatmospheric values of the year 2016 along with the cyclone Vardah are extracted and the clustering techniques DBSCAN and K-means are applied. The results show that the correct clusters have been formed and they clearly distinguishes the Cyclone from other period.

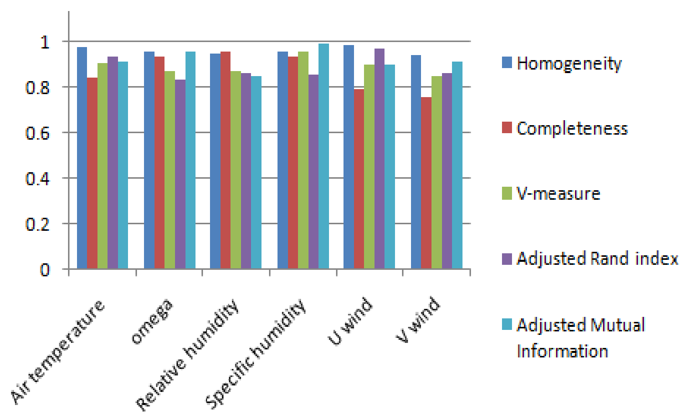


Figure 5. Comparison between DBSCAN and K-MEANS for Cyclone VARDAH

CONCLUSION

Climatalogical conditions need to be predicted to save the life of people which is a challenging problem. Machine learning techniques can be applied to predict the extreme weather events.This paper addresess the use of machine learning algorithms to filter and visualize the extreme weather event. In the proposed system, AFM technique is used to filter the events and based on the results the class labels are assigned.The extreme and non-extreme weather events are visualized using DBSCAN and K-means clustering algorithm. The past extreme events like BOB(06),Thane and vardah are validated and the results are verified by the parameters like homogeneity, completeness, V-measure, Adjusted Rand Index (ARI), Adjusted Mutual Information (AMI). The results show that above 90% of extreme and non-extreme events are correctly identified by the proposed method.

REFERENCES

[1] Piruthevi, C. and Selvi, C.K., 2017." Filtering of anomalous weather events over the region of Tamil Nadu."In Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), 2017 IEEE International Conference pp. 1-7.

[2] Cheng Tang and Claire Monteleoni.,2015."Detecting Extreme Events from Climate Time Series via Topic Modeling", Springer International Publishing Switzerland, pp. 207-215.

[3] Grover A, Kapoor A, Horvitz E.,2015."A deep hybrid model for weather forecasting."In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 379–386.

[4] Liu J.N, Hu Y, You J.J, Chan.,2014."Deep neural network based feature representation for weather forecasting", In: Proceedings on the International Conference on Artificial Intelligence (ICAI), p. 1. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing.

[5] Nayak M.A, Ghosh S 2013."Prediction of extreme rainfall event using weather pattern recognition and support vector machine classifier", In Theoret. Appl. Climatol. 114(3–4), pp. 583–603.

[6] Root B, Knight P, Young G, Greybush S, Grumm R, Holmes R, Ross J.,2007."A fingerprinting technique for major weather events", In: J. Appl. Meteorol. Climatol.46(7), pp. 1053–1066.

[7] Satya Prasad and Salwa Mohammed Nejres.,2015."Use of Data Mining Techniques for Weather DataIn Basra City", International Journal of Advanced Research in Computer Science and Software Engineering 5(12),pp.135-139.

[8] Sulagna Gope, Sudeshna Sarkar, Pabitra Mitra, and Subimal Ghosh.,2016."Early Prediction of Extreme Rainfall Events": A Deep Learning Approach' In: springer pp. 154-167.

[9] Hany Hasanien, Senior Member.,2015."Shuffled Frog Leaping Algorithmfor Photovoltaic Model Identification," IEEE Transactions on Sustainable Energy, vol. 6, pp. 509-515.

[10] Kevin Lee, ZhihongMan, Dianhui Wang, Zhenwei Cao.,2011."Classificationof bioinformatics dataset using finite impulse response extreme learningmachine for cancer diagnosis," Neural Computing and Applications,vol. 22, pp. 457-468.

[11] Yong Zhang, Dunwei Gong, Ying Hu, Wanqiu Zhang.,2015."Feature selectionalgorithm based on bare bones particle swarm optimization," Neurocomputing, vol. 148, pp. 150-157.

Authors:

C. S. Kanimozhi Selvi is a faculty member in the Department of Computer Science and Engineering of Kongu Engineering College, Perundurai, Erode, Tamil Nadu, India. She received her Bachelor's degree in Computer Science in 1994 from Bharathiar University and a Master's degree in Computer Applications from Bharathiar University in 1998. Then, she received her Master's degree M.E., in Computer Science and Engineering from Anna University, Chennai in 2004. She has completed her Ph.D. in 2011. She has been in the teaching profession for the past 19 years. Her areas of academic interest include data mining, database management systems and cloud computing. She has published 20 articles in international journals and more than 30 papers in international and national conferences.

G.Sowmiya is a student member in the Department of Computer Science and Engineering of Kongu Engineering College, Perundurai, Erode. She received her Bachelor Degree B.E., in Computer Science from Velelar College of Engineering and Technology in 2017. Her areas of interest is Machine learning.