Indian Journal of Engineering

ANALYSIS

ISSN 2319 – 7757 EISSN 2319 –7765

Cyclone storm prediction using KNN algorithm

Publication History
Received: 04 August 2015
Accepted: 21 August 2015
Published: 1 September 2015 (Special issue)

Citation
Sri Priya PV, Geetha A. Cyclone storm prediction using KNN algorithm. *Indian Journal of Engineering*, 2015, 12(30), 350-354

CYCLONE STORM PREDICTION USING KNN ALGORITHM

P.V. Sri Priya
M.Phil Research Scholar,
Chikkanna Government Arts College, Tirupur
E-Mail: mscsripriya35@gmail.com

A. Geetha,
Assistant Professor,
Chikkanna Government Arts College, Tirupur.
E-Mail: geethagac07@gmail.com

***Abstract*: Nowadays Data Mining and Machine Learning Techniques are used in different areas in daily life. It is also used in the meteorological data processing and daily weather predictions. In this paper we used K-Nearest Neighbor algorithm to predict the Cyclone Storm. K-Nearest Neighbor is a good Classifier which classifies the data into different stages of cyclone storm. Here we have used three parameters Estimated Central Pressure, Maximum Sustained Surface Wind and pressure drop to decide the class of the storm .Here I used five years of storm data starting from 2001 to 2005 for the prediction. It classified the data of five classes Depression, Deep Depression, Cyclone Storm, Severe Cyclone Storm, and Very Severe Cyclone Storm and given the results with 88% accuracy, and 12% we have the misinterpreted or misclassified data.**
General Terms: Classifier, Cyclone Storm, RSRW

Keywords: Data Mining, K-Nearest Neighbor, Cyclone Storm,

1. INTRODUCTION

Data Mining is one of the vast areas which is used in different domains like Medical, Games, Business, Entertainment and so on. Here in this paper we have used data mining technique to weather prediction. Weather Prediction includes daily weather parameters like rainfall, temperature, wind speed etc, and also severe weather prediction like storm and flood prediction. By this we can avoid the economical and human lose in severe weather like disaster, flood and storm.

Tropical Cyclones (TCs) lead to potentially severe coastal flooding through wind surge and also through rainfall Runoff processes. There is growing interest in modeling these processes simultaneously.[1] Increasing intensity of hurricanes, extreme droughts, floods, and rising sea levels might occur due to rising temperature during the warming phase of the climate change. Especially considering that our planet will reach 9 billion inhabitants by mid-century, the associated social, economic, and environmental impacts are enormous. Chaunté W [2] tried to predict the Tropical Cyclones (TCs) using change in the Cloud Clusters (CCs). Prior studies have attempted to predict tropical cyclogenesis (TCG) using numerical weather prediction models, satellite and radar data but this is not a prediction study. The goal of this research is to objectively obtain actual locations of CCs, extract features to provide more information regarding each CC, and distinguish between developing and non-developing CCs based on the extracted features.

In this present work, K- nearest neighbor (K-NN) model is used for the prediction purpose. K-NN is an important pattern recognition technique in soft computing. Sharma et al. [8] forecasted storms using soft computing method. Chakrabarty 1 et al. [4] nowcasted severe storms using K-NN models having different values of K. K-nearest neighbor (K-NN) is one of the best data mining algorithms for classification, which is used in different applications. K-NN algorithm was originally suggested by Cover in 1968. This algorithm operation is based on comparing a given testing data point with training data points and finding the training data points (neighbors) that are similar to it, and then predict the class label of these neighbors [10]. K-nearest neighbor algorithm is a non-parametric method for classifying objects based on closest training examples in the feature space. It is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. K-NN technique is applied by Li et al. [11], to forecast solar flare. Brath et al. [12] and Jayawardena et al. [13] applied K-NN method for flood forecasting. Jan et al. [14] used data mining technique for the seasonal to Inter- Annual Climate prediction. Bankert and Tag [2002] used a set of characteristic features to define a TC in an SSM/I (Special Sensor Microwave/Imager) image and then used K-Nearest Neighbor (K-NN) algorithm to match these features with historical images of tropical cyclones to estimate the intensity. [15]

In this paper we have proposed a paper to predict the cyclone storm and its class by using Data Mining technique. We have used 5 years of cyclone storm data in Bay of Bengal, Indian Ocean, and Arabic Sea range. All the data has been collected from the Indian Meteorological Department. Here we used data from 2001 to 2005 in these three ranges. For the prediction we have used K-Nearest Neighbor algorithm and got the results with 88% accuracy.

2. DATA

2.1 DATA COLLECTION

In this paper all the data has been collected from Indian Meteorological Department, Govt. of India. We have used 5 years of Cyclone Storm data starting from 2001 to 2005. It includes three regions of cyclone storm spaces Bay of Bengal, Arabic Sea and Indian Ocean. And all the data has been collected from stating time of the cyclone storm depression and the various states of cyclone storm intensities and ends with the weakened state of the cyclone storm to a depression. We have totally 703 records for the prediction.

2.2 DATA DESCRIPTION

In the collected data set there are totally eleven parameters available including Name of the Storm, Region, Latitude, Longitude, Time, Cl No, Estimated Central Pressure, Maximum Sustained Surface Wind, Pressure Drop and Grade of the Cyclone Storm. With this data set we can conform the place time and intensity of the storm. And the Grade class is considered as the target of the prediction process which means we are going to predict the Grade of the Cyclone Storm by giving the other parameters of the storm data. For the prediction process only those parameters which affect the results will be considered, thus we only take three parameters out of eleven for the prediction process. Those are Estimated Central Pressure, Maximum Sustained Surface Wind and Pressure Drop.

These three parameters are recorded from the beginning of the storm depression and then calculated every three hours till the cyclone storm loses its intensity and weaken into the small depression. The Estimated Central Pressure is calculated in the measurement called hPa (Hecto Pascal). Maximum Sustained Surface Wind is the value of the wind speed measured in the surface level in the measurement kt(Knots). Pressure Drop is the level of pressure level from the surface level which also measured in hPa (Hecto Pascal). We have totally 703 records in the total data set from year 2001 to 2005. And also we took three region measurements like Indian Ocean, Bay of Bengal and Arabic Sea. Target Class or the Grade of the Cyclone Storm has five categories included. A Cyclone Storm stars with the Depression (D), and then turn into Deep Depression (DD), then the actual Cyclone Storm (CS) will start, next level will be Severe Cyclone Storm (SCS), and the last grade is Cyclone Storm at its at most level Very Severe Cyclone Storm (VSCS).

3. METHOD

Here K-Nearest Neighbor algorithm is used to predict the class of the Cyclone Storm, and reports which class the particular Cyclone Storm belongs to using the following methodology.

3.1. K-Nearest Neighbor (K-NN)

Yakowitz extended the K-nearest neighbor method constructing a robust theoretical base for it and introduced it into the successful forecast in the hydrological research. K-nearest neighbor method is applied to recognize the Grade of the Cyclone Storm in this paper. The total data set is divided into five classes and these are training dataset and test dataset. The total number of dataset in the training class is 500 records, and we have 101 records from Depression class, 121 from Deep Depression class, 173 records of Cyclone Storm, 62 records of Severe Cyclone Storm and lastly 43 records of Very Severe Cyclone Storm. In the test data set we have 203 records available. Test data set includes 56 records of Depression class, 76 records of Deep Depression class, 51 records of Cyclone Storm, five records of Severe Cyclone Storm and 15 records of Very Severe Cyclone Storm. The training data set is arranged consecutively by D, DD, CS, SCS, and VSCS class data vector. The similarity measure has been taken between each data vector of test set with each data vector of training set. Similarity between training and test observation vectors say,
$p = (p1, p2, \ldots\ldots, p\gamma)$, $q = (q1, q2, \ldots\ldots, q\gamma)$ is defined as

$$\frac{\sum_{i=1}^{\gamma} P_i q_i}{\sqrt{[\sum_{i=1}^{\gamma} P_i^2 \sum_{i=1}^{\gamma} q_i^2]}}$$

The similarity measures between two vectors reflect the cosine of the angles between them. The similarity is more if the angle is smaller. The similarity measure indicates vicinity between the two vectors (one test vector and one training vector) with each other. The cosine angle for each of the test data vector with each of the training data vector is determined. These cosine angles are arranged in the decreasing order. As the number of training data vectors is 500, the numbers of cosine angles are also 500. Half numbers of cosine angles are considered for analysis at first. For each of the Grade Depression vector, if maximum number of Depression data appears within half of the set of cosine angles then it is to be considered as properly classified as Depression class. Similar thing happens for other classes as well. Here the value of k is 26.

4. RESULT

Chart – 1:

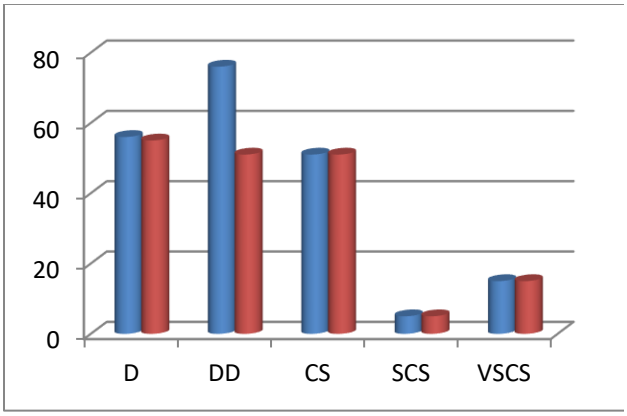

Table – 1:

	Results from the prediction using KNN					
Actual Class of the test data		CS	D	DD	SCS	VSCS
	CS	51	0	0	0	0
	D	0	55	25	0	0
	DD	.0	1	51	0	0
	SCS	0	0	0	5	0
	VSCS	0	0	0	0	15

The above results have been obtained from using KNN algorithm in the R platform. R is a programming language and software environment for statistical computing and graphics. R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is currently developed by the R Development Core Team, of which Chambers is a member. R is named partly after the first names of the first two R authors and partly as a play on the name of S. R is a GNU project. The source code for the R software environment is written primarily in C, Fortran, and R. R is freely available under the GNU General Public License, and pre-compiled binary versions are provided for various operating systems. R uses a command line interface; there are also several graphical front-ends for it.

Form the observation of results from KNN algorithm more than 88% of test data have been classified correctly and the other 12% of misclassification is happen in the Depression and Deep Depression class only and the other classes have the perfect classification without any misclassified data set. And in this paper we have taken the value of k=26 which is the square root of the total data records available. By doing this we are obtaining the most accurate results compared to the other values to the k. Chakrabarty1 et. al., in 2013 applied K-nn technique for the prediction of severe thunderstorm having around 12 hours lead time. They used two types of weather parameters such as moisture difference and dry adiabatic lapse rate. These parameters are considered from the surface level up to the five different geopotential layers of the upper air. So there are 10 weather parameters. They got only 55.55% of the “squall storm“ days which are properly classified. When they applied modified KNN technique (where k = 3) they obtained more than 87% accurate classification of the “squall storm“ days and more than 71% accurate classification for “no storm“ days. But in this paper we have used the KNN method (where k=26) are used on the three weather variables Estimated Central Pressure, Maximum Sustained Surface Wind and Pressure Drop.

5. DISCUSSION AND CONCLUSION

The challenge that has been undertaken for this forecasting work is the proper selection of the machine learning technique to get accurate prediction using only the three types of input weather variables: Estimated Central Pressure, Maximum Sustained Surface Wind and Pressure Drop. The results of the model shows the result of nearly 88% of data to be classified correctly and it only have the error rate of

the 12% which is only occur in Deep Depression and Depression classes only which is the starting states of the cyclone storm. And we can clearly predict the cyclone storm intensity which can be used to warn the people before to avoid the destructive events.

REFERENCE

[1] Francesco Cioffi, Federico Conticello, Thimoty Hall, Upmanu Lall, and Philip Orton, "A Statistical forecast model for Tropical Cyclone Rainfall and flood events for the Hadson River", 2014, Geophysical Research Abstracts, Vol. 16, EGU2014-3568, 2014.

[2] Chaunté W. Lacewell, Abdollah Homaifar and Yuh-Lang Lin, "innovative approach to the identification of Cloud Clusters developing into Tropical Cyclone", 2013, The Third International Workshop on Climate Informatics

[3] K. K. Han. text categorization using weight adjusted knearest neighbour classification. Technical report, Dept. of CS, University of Minnesota, 1999.

[4] Chakrabarty1 Himadri, Murthy C. A., and Das Gupta Ashish, "Application of pattern recognition techniques to predict severe thunderstorms", 2013, International Journal of Computer Theory and Engineering (IJCTE), Vol. 5, No. 6, pp. 850-855, ISSN: 1793-8201

[5] Q. B. Gao, Z. Z. Wang, ―Center Based Nearest Neighbor Class‖, Pattern Recognition, 2007, pp 346-349.

[6] Y. C. Liaw, M. L. Leou, ―Fast Exact k Nearest Neighbor Search using Orthogonal Search Tree‖, Pattern Recognition 43 No. 6, pp 2351-2358.

[7] J.Mcname, ―Fast Nearest Neighbor Algorithm based on Principal Axis Search Tree‖, IEEE Trans on Pattern Analysis and Machine Intelligence, Vol 23, pp 964-976.

[8] Sharma Sanjay, Dutta Devajyoti, Das J, and Gariola R.M., "Nowcasting of severe storms at a station by using the Soft Computing Techniques to the Radar Imagery", 5thEuropean Conference on Severe Storms, Landshut-Germany, 2009.

[9] R. R. Lee and J. E. Passner, "The Development and Verification of TIPS: An Expert System to Forecast Thunderstorm Occurrence, *Weather and Forecasting*, vol. 8, pp. 271-280, 1993.

[10] Moradian, Mehdi and Baraani Ahmad, "KNNBA: K-Nearest-Neighbor Based Association Algorithm, 2009, Journal of Theoretical and Applied Information Technology", Vol. 6, No.1, pp. 123-129.

[11] Rong Li, Wang Hua-Ning, He Han, Cui Yan-Mei and Du Zhan-Le, "Support Vector Machine combined with K-Nearest Neighbors for Solar Flare Forecasting", Chinese Journal of Astronomy and Astrophysics,,2007 Vol. 7, pp. 441-447.

[12] Brath, A., Montanari A and Toth E, 2002, "Neural networks and non-parametric methods for improving real-time flood forecasting through conceptual hydrological models", Hydrology and Earth System Sciences, Vol. 6 (4), pp.-627-640.

[13] Jayawardena, A.W., Fernando D.A.K. and Zhou M.C., 1997, "Comparison of Multilayer Perceptron and Radial Basis Function networks as tools for flood forecasting", Proceedings of the Conference Water-Caused Natural Disasters, their Abatement and Control, held at Anaheim, California, Publ. no. 239.

[14] Zahoor Jan, Abrar Muhammad, Bashir Shariq and Mirza Anwar M., 2008, "Seasonal to Inter-Annual Climate Prediction Using Data Mining KNN Techniques", Communications and Computer and Information Science, Vol. 20, ISSN: 1865-0929, PP. 40-51.

[15] Bankert, R. L., and P. M. Tag, "An automated method to estimate tropical cyclone intensity using SSM/I imagery", 2002, J. Appl. Meteorol., 41, 461–472.