Research Paper   Open Access

Study of Various Rainfall Estimation & Prediction Techniques Using Data Mining

Mr. M. S. Chaudhari[1], Dr. N. K. Choudhari[2]

[1]Assistant Professor, PBCOE Nagpur
[2]Principal, PBCOE, Nagpur

ABSTRACT: *It is important to estimate accurately rainfall for effective use of water resources and optimal planning of water structures and availability. For this purpose, the various models and techniques are developed to estimate rainfall in various researches using data mining techniques. The accurate and exact estimation of rainfall prediction and estimation of precipitation is not possible though many techniques are available. The use of data mining techniques to predict rainfall and its consequences may prove significance in the prediction of accurate rainfall that will help in growth of agriculture sector and the farmers can take their decisions accordingly. This paper studies various techniques of rainfall prediction and estimation and their results with the actual rainfall value.*
Keywords*: Rainfall Estimation, Prediction, Precipitation*

I.   INTRODUCTION

India and Indian subcontinent is dependent on agriculture sector that directly depends on the rainfall of the area. The Indian economy is mainly dependent on the growth and prosperity of the agriculture sector. But this entire agriculture field relies on rainfall due to very less and improper arrangement of irrigation and sources of water. But since last few years the pattern of rainfall has been changed in the entire region and it is unpredictable. The accurate prediction of rainfall may help in taking decision by the farmers for which lots of techniques are available ranging from MLR, Artificial neural networks etc.

Literature Survey:

There are different techniques used for the prediction of rainfall such as Regression analysis, clustering and Artificial Neural Networks (ANN). Fundamentally, two approaches are used for predicting rainfall. One is Empirical approach and the other is Dynamical approach. The empirical approach is based on analysis of historical data of the rainfall and its relationship to a variety of atmospheric and oceanic variables over different parts of the world. The most widely used empirical approaches, which are used for climate prediction, are regression, artificial neural network, fuzzy logic and group method of data handling. On the other hand in dynamical approach, predictions are generated by physical models based on systems of equations that predict the evolution of the global climate system in response to initial atmospheric conditions [1].

The different rainfall estimation models were developed by Ozlem Terzi [2] by using the monthly rainfall data of Isparta, Senirkent, Uluborlu, Egirdir, and Yalvac stations of Turki . Rainfall estimation models were built using Decision Table, KStar, Multilinear Regression, M5'Rules, Multilayer Perceptron, RBF Network, Random Subspace, and Simple Linear Regression algorithms and quality of these models were tested using chosen coefficient of determination (R^2) and root mean-squared error (RMSE) which are the most well known and the commonly used performance criteria. Using different combinations of Input given to above developed Models, he has generated the MLR model that gives the best results to estimate rainfall over Isparta region. J.M. Spate et al [3] has prepared a model to measure a stream flow from the measured and estimated /interpolated rainfall. K-medoid algorithm on clustering has been discussed to clustering shapes/peaks. The paper has discussed the various classification and association rule extraction methods. Instead, they have selected all those catchments in their region of interest where high-intensity rainfall data does exist for at least some temporal interval. Then they applied some simple criteria to the high-intensity data; for example so much rain must fall in such a small time interval on a given day for that fall to be flagged as an intense event. Having generated a Boolean series with 1's on every day with an intense event and 0's elsewhere, they used   data mining to automatically extract those combinations of daily data characteristics which tend to occur on a day with 1 in the Boolean series.

Pratap Singh Solanki et al [4] reviewed the studies related to use of data mining techniques in the field of water resource sector for Water Management. Presently, the Water Resource Management has become most challenging, interesting and fascinating domain around the world since last many years. Scientist tries to predict the Rainfall, Flood Warning, Water Inflow, Water Availability and Requirements etc. based on huge available metadata using various methods. In this article, they tried to search the use of data mining techniques for predicting the inflow, drought possibility, weather report, rainfall, evaporation, temperature, wind speed etc. This paper provides the survey of some literature and work done by the researchers using various algorithms and modeling method viz. Associations rules, Classification, Clustering, Decision Tree, and Artificial Neural Network etc.

Pinky Saikia Dutta [5] in her Project, Rainfall prediction is implemented with the use of empirical statistical technique. She used 6 years (2007-2012) datasets such as minimum temperature, maximum temperature ,pressure, wind direction ,relative humidity etc and performed prediction of Rainfall using Multiple Linear Regression(MLR) . This model forecasts monthly rainfall amount in summer monsoon season (in mm). Regression is a statistical empirical technique that utilizes the relation between two or more quantitative variables on observational database so that outcome variable can be predicted from the others. One of the purposes of a regression model is to find out to what extent the outcome (dependent variable) can be predicted by the independent variables. Predictors selected for the model are minimum temperature, maximum temperature, mean sea level pressure, wind speed and rainfall.

Jyothis Joseph [6] described empirical method technique belonging to clustering and classification approach. ANNs are used to implement these techniques. He used *Relative Humidity, Pressure, Temperature, Precipitable Water, Wind Speed.* In this paper subtractive clustering is used. Subtractive clustering is a fast, one-pass algorithm for estimating the number of clusters and the cluster centers in a set of data. Applying subtractive clustering, the optimum numbers of clusters are obtained .The rainfall values are categorized as low, medium & heavy. The classifier model has been evaluated against a confusion matrix and the results have been obtained. This paper applies neural network for rainfall prediction. In this paper two methods such as classification and clustering are implemented. The neural network Bayesian regularization has been applied in the implementation.

K. Poorani , K Brindha in [7] has used Principal Component Analysis method for forecasting of rainfall. The proposed PCA method is used when there is vital inter-correlation between the predictors. The PCA model avoids the inter-correlation and support to reduce the degrees of liberty by controlling the number of predictors. Their experiment studies therefore suggest that PCA has some more benefits over ANN in analyzing climatic time series such as rainfall, particularly with regards to the interpretability of the extracted signals.

E. Sreehari et al [8] has also discussed the various regression techniques of prediction. It includes the regression structure, multiple linear regression, Matrix formulation of Multiple Linear Regression, Steps to calculate the coefficients (parameters), Regression Equation using Covariance matrix method, Multi Structure Regression Equation.

Narasimha Prasad et al **[9]** proposes a need of the models for improving accuracy in the precipitation prediction using Supervised Learning In Quest (SLIQ) decision tree using Gini index for the prediction of the precipitation .This paper proposes to employ SLIQ decision tree using gain ratio that improves the accuracy using attributes humidity, temperature, pressure, wind speed and dew point.
For every attribute they found split point using the attribute and its corresponding class label pair wherever there is change in the class label. For every split point identified find the midpoint for the changed class labels and proceed until it reaches the end of the data .
Now, compare all the split points' gain ratio values and the value which is maximum is the best split point for that attribute . The gain value obtained for the attribute is to be divided by split info value of class label, in order to obtain the gain ratio value for that attribute as shown in equation

Gain Ratio (V) = Gain (V) / Split info (V).

Weather prediction can be simplified in A Geetha, G.M.Nasira *[10]* by using the artificial neural networks (ANN) with back propagation for supervised learning using the data collected at a particular station at a specified period. After training the model, they used to predict the weather conditions. As an experimental method, the model is made known to predict the values as unknown values. The model is compared with the Rapid Miner tool and the results are found more satisfactory.

Neha Khandelwal et al [11] have used multiple regression approach on the data set. From this approach they predicted rainfall in any one of the future's year by using climatic factors. After calculating predictable rain then they applied statistical analysis on that data for finding drought possibility. For finding drought possibility they used standard deviation, variance of coefficient, drought indices, drought perception etc. Now they apply statistical analysis [11] on the resulting equation's data for drought condition.

Sharma Vishal et al [12] have used average mean method, Linear Prediction filtering, Forward Linear Prediction. Results show that calculations get closer data than that of the IMD prediction. Here they also take care of the temperature difference of Tahiti and Darwin to get the information about El-Nino or La-Nino.

T.B. Trafalis et al [13] paper proposes to use a multiparametric approach by using Z, V, and W for precipitation prediction from the Norman, Oklahoma WSR-88D and applied data mining techniques to Z, V and W in order to understand the naturally occurring interrelationships and signatures of these data when rain is detected. It uses linear models and ANNs for precipitation prediction.

The modeling of monthly rainfall prediction over Myanmar is described in detail in [14] by applying the polynomial regression equation. The statistical relationship between rainfall amount and other climatic data is searched with the use of second order MPR equation which contains added terms and nonlinear cross product interaction of n predictors expressing with the first and second power of the predictors. Then the predictors which have high intercorrelation with others are reduced because the presence of many highly intercorrelated explanatory variables may substantially increase the sampling variation of the regression coefficients, and not improve, or even worsen the models' predictive ability. Experiments and graphs are reported for Pathein rain gauge station located in lower Myanmar and Magwe station located in upper Myanmar where rain fall prediction is needed more for agriculture planning and management. For experiments, regional rainfall amount taken from rain gauge stations over Myanmar and large scale data such as East India SST, SOI, ONI taken from various references are used as predictors.

II. CONCLUSIONS

In India, rainfall is a critical factor in farm management and water resource management. In this survey paper we found the use of various data mining techniques on the collected data set from the various resources may found useful in accurate prediction of rainfall. In this way Data mining offer us a much-needed opportunity to deliver scientific findings and information to stakeholders and decision makers for providing collective decision-making tools.It is observed from various studies that rainfall estimation and prediction varies from using MLR to SLIQ.

REFERENCES

[1]   Nikhil Sethi, Dr.Kanwal Garg," Exploiting Data Mining Technique for Rainfall Prediction",(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 3982-3984.

[2]   Ozlem Terzi,"Monthly Rainfall Estimation Using Data-Mining Process", Hindawi Publishing Corporation Applied Computational Intelligence and Soft Computing ,Volume 2012,Article ID 698071, 6 pages doi:10.1155/2012/698071.

[3]   J.M. Spate, B.F.W. Croke, A.J. Jakeman, "Data Mining in Hydrology" ,Department of Mathematics, The Australian National University, Canberra ACT 0200, Australia .

[4]   Pratap Singh Solanki , R. S. Thakur   '"A Review of Literature on Water Resource Management Using Data Mining Techniques",International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391 Volume 5 Issue 7, July 2016 www.ijsr.net Licensed Under Creative Commons Attribution CC BY

[5]   Pinky Saikia Dutta et.al,"Prediction Of Rainfall Using Datamining Technique Over Assam", Indian Journal of Computer Science and Engineering (IJCSE).

[6]   Jyothis Joseph, Ratheesh T K ," Rainfall Prediction using Data Mining Techniques ", International Journal of Computer Applications (0975 – 8887) Volume 83 – No 8, December 2013- 11 .

[7]   K Poorani ,K Brindha ," Data Mining Based on Principal Component Analysis for Rainfall Forecasting in India ",International Journal of Advanced Research in Computer Science and Software Engineering" Volume 3, Issue 9, September 2013 ISSN: 2277 128X , Research Paper Available online at: www.ijarcsse.com

[8]   E. Sreehari, J. Velmurugan, Dr. M. Venkatesan," A Survey Paper on Climate Changes Prediction Using Data mining ",International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 2, February 2016 Copyright to IJARCCE DOI 10.17148/IJARCCE.2016.5261 294 .

[9]   Narasimha Prasad LV, Naidu MM   ," An Efficient Decision Tree Classifier to Predict Precipitation Using Gain Ratio", The International Journal of Soft Computing and Software Engineering [JSCSE], Vol. 3, No. 3, Special Issue: The Proceeding of International Conference on Soft Computing and Software Engineering 2013.

[10]   A Geetha,G. M. Nasira," Artificial Neural Networks' Application in Weather Forecasting – Using RapidMiner ",International Journal of Computational Intelligence and Informatics, Vol. 4: No. 3, October - December 2014 ISSN: 2349-6363 177.

[11]   Neha Khandelwal, Ruchi Davey,"Climatic Assessment Of Rajasthan's Region For Drought With Concern Of Data Mining Techniques", International Journal Of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue 5, September- October 2012, pp.1695-1697, 1695 .

[12]   Sharma Vishal, Choudhary Sudesh,"Monsoon Rain Fall Prediction of Haryana 2016 on the Basis of Historical Data",International Journal of Recent trends in Engineering and Research, Volume 02, Issue 06; June - 2016 [ISSN: 2455-1457],pp608-612.

[13]   T.B. Trafalis, M.B. Richman et al ,"Data mining techniques for improved WSR-88D rainfall estimation", Computers & Industrial Engineering 43 (2002) 775–786.

[14]   Wint Thida Zaw , Thinn Thu Naing ," Empirical Statistical Modeling of Rainfall Prediction over Myanmar ",World Academy of Science, Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:2, No:10, 2008.