
Statistical Modeling for Rainfall Prediction using Data Mining Technique

Dhananjay P. Atole*

Department of CSE, GHRAET, Nagpur
RTM Nagpur University, Nagpur, India

Deepak Kavgate

Department of CSE, GHRAET, Nagpur
RTM Nagpur University, Nagpur, India

Rushi Longadge

Department of CSE, GHRAET, Nagpur
RTM Nagpur University, Nagpur, India

Abstract

India is an agricultural country, so crop productivity is highly depends upon rainfall. For the better crop productivity rainfall prediction is necessary and required. There are several techniques of rainfall prediction. Data mining technique is used to calculate or analyze the rainfall prediction. For better result used statistical modeling for rainfall prediction. In statistical modeling clustering algorithm are used for cluster like K-means, k-medoid and for prediction used regression technique Multi Linear Regression (MLR), Multi Polynomial Regression (MPR) and gives better result.

Keywords: *Multiple Linear Regression, Empirical Approach, Rainfall Forecasting..*

****Author for correspondence*** dhananjaypatil47@gmail.com

1. Introduction

Rainfall is one of several important factors affecting farming countries like India. Studies have shown the rainfall is basic issue in farming sector. The rainfall prediction it is complicated job in last centuries, for solving this problem many techniques and methods are developed. One of the techniques is data mining technique. It is used for prediction of rainfall. Data mining is used to extract the important information from the large amount of data. The goal of data mining is to clipping the data and converts them into useful data for future information. The data mining is connection between a several method like machine learning, database, artificial intelligence and statistics. The empirical method 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 use empirical approaches used for climate prediction are regression, artificial neural network, fuzzy logic and group method of data handling [1]. In dynamical approaches is implemented using numerical weather forecasting method. The main objective of dynamical approaches is analyzing the few years of rainfall data. Physical models are used to generate the prediction of global climate system. The models are based on systems of equations which predict the climatic changes around the world.

For rainfall prediction first collect data over five cities Nagpur, Pune, Mumbai, Chennai and Delhi. On collected data clustering algorithm are applied and get different cluster on basis of cities and for prediction used regression technique. Finally predicted and actual result is compared.

2. Related Work

Rainfall prediction is very crucial task. It requires the intelligence and facts according to that it analyzes the predicted values with the help of this the better results can be obtained.

Andrew Kusiak, Xiupeng et al. [1] presented modeling and prediction of rainfall using radar reflectivity data. A data-mining approach is applied to predict rainfall in Oxford, Iowa City with the help of radar reflectivity data and tipping-bucket (TB) data. Five data mining algorithm are used to predict rainfall there are neural network, random forest, classification and regression tree, support vector machine, and k-nearest neighbor. Basically, three models are used in model I constructed from radar data covering Oxford. In model II actual prediction is done it predict rainfall from radar data and TB data at collected Oxford. In model III is constructed using radar and TB data at South Amana and Iowa City. WintThidaZaw et al. [2] predicted the rainfall over Myanmar. For prediction over Myanmar used Multi Linear Regression (MLR) technique and also used second order Multi variables polynomial regression (MPR). The MLR and MPR is way to describe the complex nonlinear input output relationships that why outcome variable can be predicted from the other or others. In this paper polynomial regression equation is used to predict rainfall over Myanmar. Imran Ahmed et al. [3] have described the empirical method and dynamic method. By using multiple linear regression methodology on collected six year of historical data over Coonor, Nilgris district, Tamil Nadu state and gives rainfall prediction over Tamil Nadu state.

Sangari, R. S. et al. [4] described all algorithms used in data mining technique. Basically these algorithms are important factor of better result. Naive Bayes, K- Nearest Neighbour algorithm, Decision Tree, Neural Network and fuzzy logic these are used for prediction of rainfall and compared each other and analyze which algorithm gives better prediction. M. Kannan et al. [5] have described the Karl Pearson coefficient, multiple linear regressions. The empirical and dynamic method also used for prediction and describe some of data mining technique. Harry R. Glahn et al. [6] have described statistical modeling and Model output statistic (MOS) method. The MOS determination of weather related statistic of numerical model. This method screening regression has been applied to prediction of probability of precipitation, maximum temperature, cloud amount, conditional probability frozen precipitation.

3. Data Analysis

Analysis of the rainfall prediction is done by using the methods K-means, K-mediod, regression technique Multi Linear Regression (MLR), Multi Polynomial Regression (MPR) and gives better result. In Data Mining, some of the techniques are used i.e. classification, clustering, regression or prediction etc. Using the classification we classified what is the reason for rainfall in the ground level. Using clustering technique, we grouped the element that is particular area occupied by the rainfall region. In prediction, we have to predict the rainfall occurs in the particular region [2]. We collected rainfall data of five different cities i.e. Nagpur, Pune, Mumbai, Chennai and Delhi with the help of radar station and open government data. This collected different cities data

are arranged and create only one data set. Gather all data into one data set. We took seven years data and predicted the rainfall using regression and compared it with the actual data.

4. Methodology

Regression technique is statistical empirical technique of data mining which is wide range of field used for climate prediction, biological science etc. in regression technique for single variable used simple regression and multiple variable used multi linear regression technique.

Simple regression

Simple linear regression equation is $Y=a_1+a_2x_1\dots\dots Eq. (1)$

Where Y is the predicted response, the predictor coefficient is a1 and a2, x1 is predictor variable. Simple linear regression is used for only one variable.

Multiple Linear Regression (MLR)

The main goal of MLR is to model the relationship between explanatory and response variables. MLR is used to model the linear relationship between the dependent variable and independent variable (one or more variable).

$$Y=a_1+a_2x_1+a_3x_2+a_4x_3+a_5x_4\dots Eq. (2)$$

Where Y is the predicted rainfall value, the predictor coefficient are a1,a2,a3,a4,a5 and the predictor variable are x1,x2,x3,x4. Multiple linear regressions assume that y is directly related to a linear combination of explanatory variables the term called ‘Linear’. MLR model consist of predictors expressing the first, second and third powers of the predictors so it is called as third order polynomial model with predictor variable. Thus, the statistical relationship between rainfall amount and other climate data each search with the use of second order MLR equation [3]. By using the second equation predict rainfall. Multiple Linear Regression technique gives more efficient result than other technique. MLR model involve predictors expressing the first and second and third powers of the predictors, and then this polynomial model is called a third-order polynomial model with predictor variable. Thus, the statistical relationship between rainfall amount and other climatic data is searched with the use of second order MLR equation which contains added terms and nonlinear cross product interaction of n predictors expressing the equation [3].

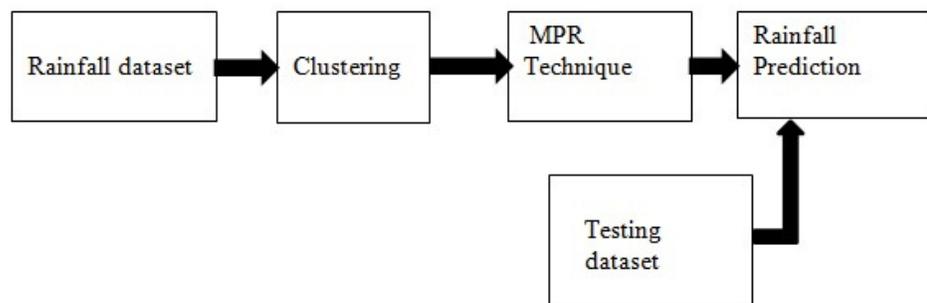


Fig.: Working of Model

Working of Model

First we gather data and create dataset then apply on this data clustering algorithm like K-medoid. After clustering, apply Multi variables polynomial regression (MPR) technique for better prediction after calculation. We have predicted rainfall and compare with testing data check efficiency of model to existing model.

5. Result

We computed the rainfall prediction day wise, week wise, month wise, using seven years data of five cities in India i.e. Nagpur, Pune, Mumbai, Chennai and Delhi. By using Multi variables polynomial regression (MPR) technique, we can predict the rainfall easily and prediction value is not accurate but approximate to the existing system. In the above graph, year wise prediction of actual result and predicted result has been displayed. The actual result values are near about to the predicted result. Every city has different accuracy and result value (shown in mm). Accuracy of the result is shown in percentage and comparative actual result and predicted result is shown in figures.

Rain Fall Prediction Yearwise on City



City wise Data

Fig 2: Accuracy Graph of Result

6. Conclusion & Future Enhancement

Prediction of rainfall is very complex topic and can't be predicted easily. We predict rainfall using seven year data and shows there minimum temperature, humidity, wind speed and also show there event this day is sunny or rainy. Multi variables polynomial regression (MPR) & Multi linear regression technique is reliable for prediction. So using this technique we can predict whether forecast and other application in future.

References

- [1] WintThindaZaw, Thinn Thu Naing, "Empirical Statistical Modeling of Rainfall Prediction over Myanmar", World Academy of science, Engineering and Technology, vol:2 2008-10-27
- [2] Andrew Kusiak, Xiupeng Wei, AnoopPrakashVerma, Evan Roz, "Modeling and Prediction of Rainfall Using Radar Reflectivity Data: A Data-Mining Approach", IEEE Transaction on geoscience and remote sensing ,vol. 51, no. 4, April 2013

- [3] Imran Ahmed, Shruti Menon, Nikitha “Rainfall Prediction Using Multiple Regression Technique” International Journal of Applied Engineering Research, ISSN 0973-4562, Vol. 8, No. 19.
- [4] Sangari, R. S, Balamurugan, M., “A survey on Rainfall Prediction using Data Mining”, International Journal of Computer Science and Mobile Applications, Vol.2 Issue. 2, 2014, pp. 84-88.
- [5] M.Kannan, S.Prabhakaran, P.Ramachandran, “Rainfall Forecasting using Data Mining Technique”, International Journal of Engineering and Technology Vol.2 (6), 2010, 397-401.
- [6] Harry R. Glahn, & Dale A. Lowry, “The Use of Model Output Statistics (MOS) in Objective Weather Forecasting”, Technique development laboratory, National weather service, NOAA, Silver Spring, Md. 20910.
- [7] Berkhin P, 2002, CA, Tech. Rep, “Survey of clustering data mining techniques, Accrue Software, San Jose”.
- [8] Sarah N. Kohail, Alaa M. El-Halees, “Implementation of Data Mining Techniques for Meteorological DataAnalysis”. July, 2011.
- [9] Valmik B Nikam and B.B.Meshram, “Modeling Rainfall Prediction Using Data Mining Method”, 5th International Conference on Computational Intelligence, Modeling and Simulation, 2013.
- [10] N.A.Charaniya and S.V.Dudul, “Focused Time delay neural network model for Rainfall Prediction Using Indian Ocean Dipole Index”, 4th International Conference on Computational Intelligence and Communication Networks, 2012.
- [11] Soo-Yeon Ji, Sharad Sharma, Byunggu Yu and Dong Hyun Jeong, “Designing a Rule-Based Hourly Rainfall Prediction Model, IEEE IRI, August 8-19, 2012.