

A Review on Statistical Modeling for Rainfall prediction using Data mining Technique

Dhananjay P. Atole¹, Prof. Rushi Longadge², Prof. Deepak Kapgate³

Department of Computer Science and Engineering
G. H. Raison Academy of Engineering and Technology, 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. In this paper focuses on several data mining techniques like empirical method and dynamic method which is essential to predict the rainfall. The empirical method predicts rainfall using regression, ANN, Fuzzy Logic and group method of data handling. The regression is Statistical empirical technique which is used for prediction. In this paper several methods are discussed, that explore which methods will be better for prediction of rainfall.

Keywords- Empirical method, Dynamic method, K-Nearest Neighbour algorithm, decision tree, neural network, MPR, MLR.

I. 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. DM technique analyzing the data which is present in dataset and solving problem easily. For extracting the data pattern in order applied intelligent methods in data mining.

Data mining technique involve following main component:

- The Data Mining techniques transform, extract data into data warehouse.
- The Data Mining techniques are useful for store data and manage data in multilayered dataset.
- It is provide data approach to information technology professionals.
- In data mining application software is used for analysis of data.

Data mining utility are used to identify the type of patterns to be present in data mining work. Data mining work can be divided into two types –descriptive mining and predictive mining. Descriptive mining work describe the general properties of the data in dataset. Predictive mining work achieve inference on the current data in order to make predictions.

II. CURRENT APPROACHES

Empirical Method

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 [2].

Dynamic Method

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.

Statistical Modeling

A collection of probability distribution functions or probability density functions it is called statistical modeling. The statistical modeling is relationship between the variable in form of Mathematical equations. The statistical models are used for how random variables are related to other variable. Basically statistical empirical approach is type of regression is used for relationship between two or more computable variables on observational database. In empirical statically modeling involve regression technique like MLR, MPR, fuzzy logic, ANN, stochastic and group method of data handling. In statistical modeling regression are used for prediction.

Model Output Statistics

The model output statistic is other method used for weather forecasting it is also known as MOS technique. Model output statistical methods determine the statistical connection between predict and variable from numerical model. In this method weather related statistic are determine in numerical model.

III. REGRESSION TECHNIQUES

In data mining regression is statistical empirical technique, in which there are mainly two techniques Simple and multiple linear regression technique. It is widely used in biological sciences, atmosphere changes and climate prediction. In simple terms regressions means predicting one variable from another 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).

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

Where Y is the predicted rainfall value, the predictor coefficient are a_1, a_2, a_3, a_4, a_5 and the predictor variable are x_1, x_2, x_3, x_4 . Multiple linear regressions assume that y is directly related to a linear combination of explanatory variables the term called 'Linear'. MLR model consists 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].

Clustering approach

In the clustering technique data are divided into different cluster such as called group of similar object. By using this technique form necessary cluster and achieve simplification of data. In data mining technique clustering used for Mathematics, statistic, numerical analysis. K-means, KNN, Classification algorithm these are algorithm is used for clustering in data mining.

IV. Related Work

In this paper some data mining techniques are discussed. 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] has present 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] predict 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] has 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] it describes all algorithms for used in dada 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 its compared each other and analyze which algorithm gives better prediction. M.Kannan, et al. [5] has 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] has describe 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.

V. Discussion

Linear regression, Multivariable polynomial regression, Empirical approach, Model output statistics are used for prediction of weather forecasting in data mining. Linear regression is used for prediction it shows the relationship between independent and dependent variable. In statistical approach MLP is used for prediction of rainfall. With the help of multivariable linear regression shows relationship between two or more quantative variable on observational dataset so output variable can predicted from the other. Empirical method can implemented easily and provide accuracy to result in this method regression, artificial neural network, fuzzy logic are used for climate prediction. Model Output Statistics (MOS) is determining Statistical relationship between predict and variable but this method does not consider general attributes.

VI. CONCLUSION

In this paper we study much technique of data mining which is use for rainfall prediction. The empirical method, dynamic method and statistical modeling each method is used for predicting the rainfall. Empirical method analyzed historical data, dynamic method analyzed few years given data and statistical method is based on Mathematical equation. Comparatively the statistical modeling provide higher accuracy and gives better result than other methods.

REFERENCES

- [1] 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.
- [2] WintThindaZaw, Thinn Thu Naing, "Empirical Statistical Modeling of Rainfall Prediction over Myanmar", World Academy of science, Engineering and Technology, vol:2 2008-10-27.
- [3] Imran Ahmed, ShrutiMenon, Nikitha K B, "Rainfall Prediction Using Multiple Regression Technique", International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 3 Issue 5 may, 2014 Page No. 6206
- [4] Sangari.R.S, Dr.M.Balamurugan, " A survey on Rainfall Prediction using Data Mining", International Journal of Computer Science and Mobile Applications, Vol.2 Issue. 2, February- 2014, pg. 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 and 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, July 2011," Implementation of Data Mining Techniques for Meteorological DataAnalysis".
- [9] Valmik B Nikam and B.B.Meshram, "Modeling Rainfall Prediction Using Data Mining Method", Fifth 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", Fourth International Conference on Computational Intelligence and Communication Networks, 2012.
- [11] Soo-YeonJi, SharadSharma ,Byunggu Yu and Dong Hyun Jeong, "Designing a Rule-Based Hourly Rainfall Prediction Model, IEEE IRI, August 8-19, 2012.

