Rainfall Forecasting Using Data Mining Technique

[1]M.Kannan, [2]S.Prabhakaran, [3]P.Ramachandran
[1]Assistant Professor CSA Dept., [2]Research Scholar CSA Dept., [3]Research Scholar CSA Dept.
[1]SCSVMV University, Enathur, Kanchipuram - 631 561.
[1] saikannan1999@rediffmail.com,[2]prabhakaran.babu07@gmail.com,[3]rkmvc.rc@gmail.com

Abstract

Rainfall is important for food production plan, water resource management and all activity plans in the nature. The occurrence of prolonged dry period or heavy rain at the critical stages of the crop growth and development may lead to significant reduce crop yield. India is an agricultural country and its economy is largely based upon crop productivity. Thus rainfall prediction becomes a significant factor in agricultural countries like India. Rainfall forecasting has been one of the most scientifically and technologically challenging problems around the world in the last century.

Keywords: Rainfall forecasting, Karl Pearson coefficient, multiple linear regressions.

1. Introduction

India is basically an agricultural country and the success or failure of the harvest and water scarcity in any year is always considered with the greatest concern. The term monsoon seems to have been derived either from the Arabic mausin or from the Malayan monsin. As first used it was applied to southern Asia and the adjacent waters, where it referred to the seasonal surface air streams which reverse their directions between winter and Summer, southwest in summer and north east in winter in this area. During the summer the continent is heated, leading to rising motion and lower pressure. This induces airflow from sea to land at low elevations.

A wide range of rainfall forecast methods are employed in weather forecasting at regional and national levels. Fundamentally, there are two approaches to predict rainfall. They are Empirical method and dynamical methods.

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 use empirical approaches used for climate prediction are regression, artificial neural network, fuzzy logic and group method of data handling. 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. The Dynamical approaches are implemented using numerical rainfall forecasting method. This paper describes empirical method technique belongs to the regression approach which try to make a short-term forecast of rainfalls over specified region in our state. The objective analyses the three months rainfall data of particular region for five years because these months are winter season for our state. Multiple linear regressions are used to predict the rainfall using the previous year's data from the corresponding time period.

2. Related Work

Accurate and timely weather forecasting is a major challenge for the scientific community. Rainfall prediction modeling involves a combination of computer models, observation and knowledge of trends and patterns. Using these methods, reasonably accurate forecasts can be made up. Several recent research studies have developed rainfall prediction using different weather and climate forecasting methods. Regression is a statistical empirical technique and is widely used in business, the social and behavioral sciences, the biological sciences, climate prediction, and many other areas.

N. Sen. [1] has presented long-range summer monsoon rainfall forecast model based on power regression technique with the use of Ei Nino, Eurasian snow cover, north west Europe temperature, Europe pressure gradient Wind pattern, Arabian sea SST, east Asia pressure and south Indian ocean temperature in previous year. The experimental results showed that the model error was 4%.

S. Nkrintra, [2] described the development of a statistical forecasting method for SMR over Thailand using multiple linear regression and local polynomial-based nonparametric approaches. SST, sea level pressure (SLP), wind speed, EiNino Southern Oscillation Index (ENSO), IOD was chosen as predictors. The experiments indicated that the correlation between observed and forecast rainfall.

T. Sohn, [3] has developed a prediction model for the occurrence of heavy rain in South Korea using multiple linear and logistics regression, decision tree and artificial neural network.

M. T. Mebrhatu [4] modeled for prediction categories of rainfall (below, above, normal) in the highlands of Eritrea. The most influential predictor of rainfall amount was the

southern Indian Ocean SST. Experimental results showed that the hit rate for the model was 70%.

H. Hasani [5] proposed human height prediction model based on multiple polynomial regression that was used successfully to forecast the growth potentials of height with precision and was helpful in children growth study.

Vaccari [6] modeled plant motion timeseries and Nutrient recovery data for advanced life support using multi variable polynomial regression.

3. Data Analysis

We analysis the rainfall forecasting is done by using the methods artificial intelligence, neural network, fuzzy sets and data mining in some journals. Artificial intelligence and neural network are more difficult compared to data mining because artificial intelligence involves some algorithms.

In Data Mining, some of the functionalities are used i.e. classification, clustering, regression or prediction etc. Using the classification we classify what is the reason for rainfall fall in the ground level. Using clustering technique, we grouping the element that is particular area occupied by the rainfall region. In Prediction we have to predict the rainfall occurs in the particular region.

Finally, we take the prediction methods in data mining because rainfall occupied in the region done by the regression approach. In regression we use Karl Pearson correlation Coefficient for finding how many centimeters rainfall fall in the particular region. We have to predict the rainfall fall in the future years by using the multiple linear regression approach.

Data used in the present study are collected from the Statistical department of Tamil Nadu, Chennai. We take five years of data during three months September, October, November are explored because these three months are the winter Season for our state. Using this data, we are computed for rainfall fall in the ground level using Pearson correlations coefficient.

The Pearson correlations between the data pertaining to different months are computed and are displayed in graph. The graph shows that the months are not significantly correlated with respect to monthly rainfall of the winter season.

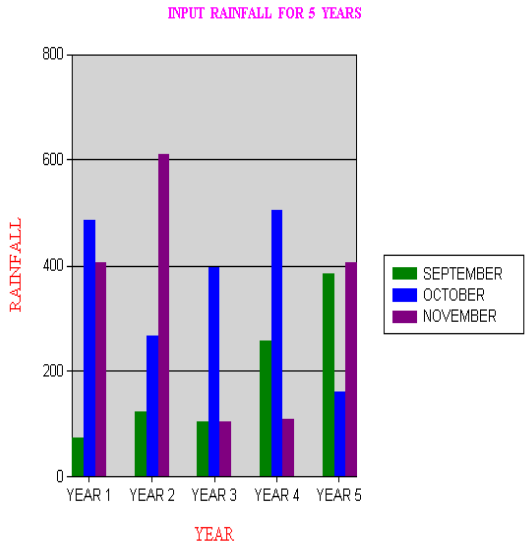
Pearson correlation coefficient between two variables is defined as covariance of two variables divided by the product of their standard deviations. It is denoted by r.

Pearson correlation coefficient is a measure of the strength of the association between two variables. It ranges from -1 to +1.

We take the five years data are computed using Pearson coefficient and then compare with predicted data using regression approach.

Rainfall fall in earth will be measured by Millimeter (mm.). Measurement and years plot in the graph using x and y axis. Using the input data, we give output as soon as possible one. Therefore we predict the output as future three years rainfall fall in the ground level.

Output will be approximate because we predict the rainfall for future three years by using regression approach. In this approach we using some predictor variables and then only this is useful to predict for three months.



4. Regression

Regression is a statistical empirical technique that utilizes the relation between two or more quantitative variables on observational database so that an outcome variable can be predicted from the others.

Regression use two methods Simple linear regression and multiple linear regression models. Regression produces a polynomial describing the relationship between any set of inputs and corresponding output.

Simple linear regression model is of the form:

$$Y = w0+w1x$$

Where x = predictor variable
w0, w1 is predictor variable

Regression model which contains more than two predictor variables is called MLR. If a MLR model which contains n predictors expressing with the first and second and third powers of the predictors, 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

4. 1 Multiple Linear Regression Model

We predict the rainfall by using multiple linear regression (MLR) method is adopted to predict the average summer monsoon rainfall in a given year using the monthly rainfall data of the summer-monsoon of the previous year.

After computation, the MLR equation is set as

$$y=aX1+bX2+cX3$$

Where, a, b, c are regression coefficient

X1= September rainfall of year Y

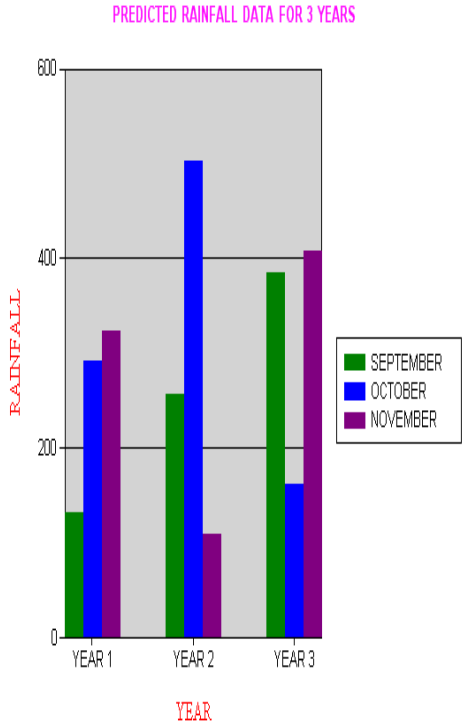
X2= October rainfall of year Y

X3= November rainfall of year Y

y= Average rainfall of year Y+1

Using this equation we predict the rainfall for future years. We use the regression coefficient is mean value of the September and October rainfall data.

Similar the process goes to an end. In multiple linear equations, we use some coefficients because take the mean value of the corresponding instances.



Since we predicted the rainfall forecasting for future years is compared to previous rainfall occur in the past years. The statistics are computed from the MLR equation and compared to Pearson correlation coefficient graphical values. The computed values are based upon the three months input value is a good predictor.

Once we get the value from the calculations, the value must lie between or below or above the values in the graphical notations. It is found that all the computed values fall below the Pearson correlation coefficient graphical values. This illustrates, in the sample studied, that none of the months is a good predictor of average monsoon rainfall of a given year.

Finally, we get the rainfall data having some approximate value not a predictor value. So we use some variety methods of data mining concepts in future years.

4.2 Result

We computed values for rainfall fall in the ground level using five years input data by Karl Pearson correlation coefficient and predicted for future years rainfall fall in ground level by multiple linear regression. Finally predicted values are lie below the computed values. So, it is not show an accurate but show an approximate value.

5. Conclusion

Rainfall time series may be unfounded. The topic of monsoon-rainfall data series is highly complex; the role that multiple linear regressions might play in this topic is one for future research—it appears, from the evidence here, not to be useful as a predictive model. Whether it might be useful for offering an approximate value of future monsoon rainfall remains to be seen. Using this regression method, we have to forecast rainfall for our state also.

5.1 Future Enhancements

Rainfall is most essential for our life. So, we predict that rainfall in the certain period. Therefore, we avoid flood, cyclone, forest fire detection, global warming etc. In future we predict the rainfall forecasting and other applications done by using the artificial intelligence, neural network and fuzzy sets etc. We do the research on public sectors and save the world.

References

[1] Klaush Juliseh, "Data mining for Intrusion Detection – A critical review", Applications of Data mining in computer security, Daniel Barbara, Sushil jajodia, Published by Springer.
[2] Guhathakurta, P (2005) "Long-range monsoon rainfall prediction of 2005 for the districts and sub-division Kerala with artificial neural network", Current Science, 90, 773-779
[3] Rajeevan, M (2001) "Prediction of Indian summer monsoon: Status, problems and prospects", Current Science.
[4] D. Mark, "Geographical information science; critical issues in an emerging cross – disciplinary Research domain", URISA Journal, vol. 12, February 1999.
[5] Tamil Nadu Meteorology Department, Chennai.
[6] European – Commission, "Forest Fires in Europe 2007, Technical report, Report No. 8, 2008.
[7] L.A. Zadeh, "Fuzzy sets", Information And control vol. 8, 1965.
[8] Nguyen T. Danh, Huyah N. Phien and Ashim D.Gupta 1999, "Neural network models for river flow forecasting, "Journal of Water SA, Vol.25.
[9] S. Lee, S. Cho and P.M. Wong 1998, "Rainfall prediction using Artificial Neural Networks," Journal of Geographic Information and Decision Analysis, Vol. 2, No. 2.
[10] Hastenrath, S(1988) "Prediction of India Monsoon Rainfall: Further Exploration", Journal of Climate.