Rainfall Forecasting using Regression Techniques

Author: Narendra Khare

Co-authors: Himanshu Raut, Kashmira Savdavkar, Prof. Priyanka Bhilare

*Department of Computer Engineering, Rajiv Gandhi Institute of Technology, Juhu-Versova link road
Mumbai-400053*

---***---

Abstract: Rainfall is one of the most important factor 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 significantly reduce crop yield. India is an agricultural country and its economy is largely based upon crop cultivation. It is also seen that India has a lot of regions lying under the Low Rainfall Zone(LRZ), where activities like food production is really a very tedious task. Thus, rainfall prediction becomes an important factor in agricultural countries like India. Rainfall forecasting has been one of the most scientifically and technologically challenging problems around the world for the last century.

Index terms: Karl Pearson coefficient, Multiple linear regressions, Rainfall forecasting.

I. INTRODUCTION

India is known as an agricultural country and the success or failure of the harvest and water scarcity in any year is always considered with the greatest concern. During summer, the continent is heated, leading to rising motion and lower pressure. This induces airflow from sea to land at lower elevations.

A wide range of rainfall forecasting methods are employed in weather forecasting not only at regional but also at national levels. Fundamentally, there are two approaches in order to predict rainfall. They are Empirical method and Dynamical methods. The Empirical approach is based on various 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 for climate predictions are regression, artificial neural network, fuzzy logic and group method of data handling. In Dynamical approach, predictions are generated by using physical models based on sets of equations which helps to predict the evolution of the global climate of system in response to the atmospheric conditions present till date.

The Dynamical approaches are implemented using numerical rainfall forecasting method describes empirical method technique belongs to the regression approach which try to generate a short-term forecast report of rainfalls over specified region in our state. The objective analysis uses the specified three months rainfall data of a particular region for the previous five years because these months are winter season for our state. Multiple linear regressions are used to take multiple data in account to predict the rainfall using the previous year's data for the corresponding time period.

For previous a couple of years the effect of global warming on food security, rainfall and temperature patterns all over the world has received great attention from policy makers and academics. Food insecurity is caused by many factors such as lack of proper planning and knowledge, inappropriate farming methods and techniques, inadequate rainfall longer lead - time prediction models. In India droughts occur more frequently affecting large numbers of people and causing tremendous economic losses, environmental damage and social hardships. Majority of the India depend on rain fed agriculture which is the mainstay of the economies. The prime cause of drought is the occurrence of below normal precipitation, which gets affected by various natural phenomena.

Research has provided evidence of relationships between meteorological variables (Webster, 1981; Rocha, 1992; Ropelewski Halpert, 1987). Manatsa et al (2007) used correlation analysis to identify the period lags for which Southeastern Oscillation Index (SOI) and Darwin pressure anomalies are significantly correlated with the Summer Precipitation Index. The authors concluded that progressive lagged four months averaged Darwin pressure anomalies Research has provided evidence of relationships among meteorological variables (Webster, 1981; Rocha, 1992; Ropelewski Halpert, 1987). Manatsa et al (2007) used correlation anal- ysis to identify the period lags for which Southeastern Oscillation Index (SOI) and Darwin pressure anomalies are significantly correlated with the Summer Precipitation Index. The authors concluded that progressive lagged four months averaged with respect to Darwin pressure anomalies.

II. LITERATURE SURVEY

In literature [1] Spectroscopic Determination of Leaf Biochemistry, Stepwise linear regression was used to select wavelengths in the broad absorption features. These methods, including empirical equations and plant reflectance models, relate spectral reflectance measurements to the concentrations of plant biochemical

constituents. Independent applications of linear regression analysis using these wavelengths to chemical concentrations of seven sites were prepared accurately. Predictions using normalization increased the accuracy. The method was consistent over independent Data set and a wide diversity of species

The authors in [2] Spatial sampling algorithm suitable for multiple linear regression model identification and estimation introduced a spatial site selection algorithm designed to identify a minimal number of calibration sites for multiple linear regression model estimation. Also, two statistical criteria that are useful for selecting optimal MLR variable combinations describe a technique for identifying faulty signal data.

The aim of study [3] was to predict next day hourly ozone concentrations through a new methodology based on feedforward artificial neural networks using principal components as inputs.

In survey [4] Activation function is used to obtain the output value of the node was studied from multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. In this study the average values were overestimated and was not viable for this project.

The authors in [5] analyzed 'Visual World' eye tracking data MLR framework overcomes some of the problems with conventional analyses, making it possible to incorporate time as a continuous variable along with gazing location as a categorical dependent variable.

The aim in [9] was studying about computational tools for probing interactions in multiple linear regression, multilevel modelling and latent curve analysis was the one which provided a better knowledge by being less prone to errors and with better plotting capabilities.

III.   PROPOSED SYSTEM

The input to the rainfall prediction system can be an excel file or a CSV file. The input data is of tabular form containing sub-regions, years, months and annual rainfall in mm as columns. The rows contain name of sub-regions, no. of years, and rainfall in mm. The file is first pre-processed to eliminate all the null values and fill them with '0'. After pre-processing the finally obtained table is classified using the KNN classification algorithm. The KNN classification classifies the data into 4 regions low, medium, high and very high. The classification labels are then converted into numerical values since machine learning algorithms deal with only numerical values.

After pre-processing and transforming the data, Multiple Linear Regression is applied to predict the rainfall. The correctness of the output i.e. predicted rainfall is checked by Mean Squared Error method. Mean Squared Error method is a risk function that gives us the corresponding expected value of square error loss.

IV. METHODOLOGY
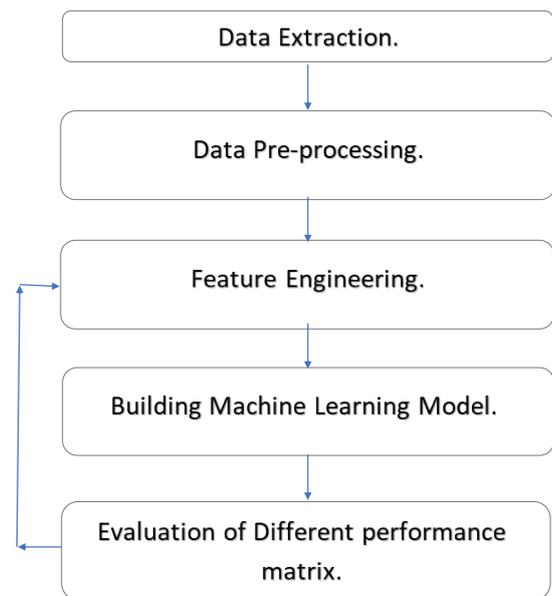
Figure 1: Flow of Data

Data used in the present study are collected from the Statistical Department of Mum- bai. We take five years of data during three months of September, October, November are explored because these three months are the winter season for our state. Using this data, we computed rainfall in the ground level using Pearson's Co-efficient of Co-relation.

Methodologies that are used are Regression and Pearson's Co-efficient of Co-relation.

1. Pearson's Co-efficient of Co-relation.

2. Multiple Linear Regression.

Pearson co-relation co-efficient is a measure of the strength of the association between two variables. Using Pearson's Co-efficient of Co-relation, we calculate the co-relation co-efficient between the different months. Using this co-relation, a graph is then plotted. The graph represents the input rainfall.

Multiple Linear Regression technique is then used on the same input data to calculate the predicted rainfall values and plot them on a graph. This graph represents predicted rainfall.

The input rainfall graph and the predicted rainfall graph can be compared to check for the accuracy and efficiency of the forecasted rainfall.

V.   IMPLEMENTATION

Simple Linear Regression formula:

$$y = a + b * x$$

Where: a, b = regression coefficients

Pearson's Co-relation Co-efficient formula:

$$r = \frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sqrt{\sum(X-\overline{X})^2}\sqrt{(Y-\overline{Y})^2}}$$

Where, $\overline{X}$ = mean of X variable

$\overline{Y}$ = mean of Y variable
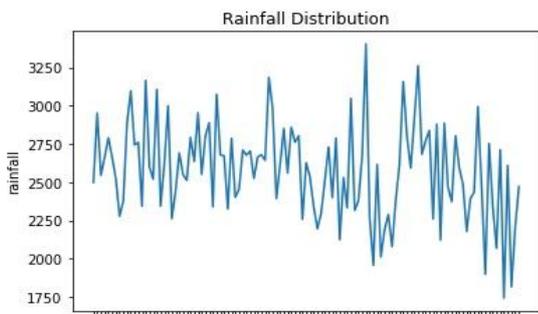
Figure 2: Rainfall distribution of a particular region
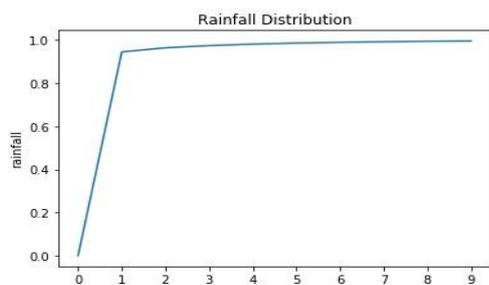
Figure 3: Principal Component Analysis

Figure 3: Classified regions.
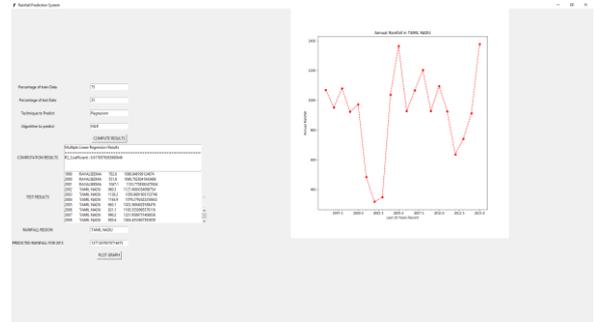
Figure 4: Multiple Linear Regression

VI. ACKNOWLEDGMENT

The authors would like to thank the Prof. Priyanka Bhilare of Rajiv Gandhi Institute of Technology for the guidance, encouragement, and support.

VII. CONCLUSION

Rainfall time series may be unfounded. The topic of monsoon-rainfall prediction data series is highly complex; the role of multiple linear regression in this topic might be useful for future researches—it appears, from the evidence here, not to be useful as a predictive model. Whether it might be useful for offering an approximate value for future monsoon rainfall remains to be seen.

REFERENCES

[1] Alessandro Armando, Gabriele Costa, Alessio Merlo, Bring Your Own Device, Securely, *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, ACM New York, NY, USA , pp. 1852-1858 March 2013.

[2] M Rebecca Copeland, Noel Crespi, Analyzing Consumerization - Should Enterprise Business Context Determine Session Policy, 2012 IEEE*Proceedings of 16th International Conference on Intelligence in Next Generation Networks (ICIN)*, pp.187-193 ,2012.

[3] Prashant Kumar Gajar, Arnab Ghosh, Shashikant Rai,BRING YOUR OWN DEVICE (BYOD): SECURITY RISKS AND MITIGATING STRATE- GIES*Proceedings of Journal of Global Research in Computer Science* , Volume 4, No.4, April 2013.

[4] Sean Chung, Sam Chung, Teresa Escrig, Yan Bai,Barbara Endicott- Popovsky,2TAC: Distributed Access Control Architecture for Bring Your Own Device Security ,' *proceeding of the International Conference on BioMedical Computing*, 2012.

[5] Dennis Titze, Philipp Stephanow, Julian Schutte,'A Configurable and Extensible Security Service Architecture

for Smartphones,' *IEEE Transactions on 27th International Conference on Advanced Information Networking and Applications Workshops*, 2013.

[6]   Dae Hyeob Kim, Ji Hoon Gong, Won Hyung Park, Neo Park, 'Vulnerability of Information Disclosure in Data Transfer Section for Safe Smartwork Infrasturucture',*IEEE transaction on International Conference on Information Science and Applications (ICISA), 2013* , pp. 1 - 3, 24-26 June 2013.

[7] Dennis Gessner, Joao Girao, Ghassan Karame, Wenting Li 'Towards a UserFriendly Security Enhancing BYOD Solution'*Proceedings of the NEC technical journal*, Vol.7 No.3/2013.

[8] Hanay, Y.S., Wolf, T.,'Techniques for Policy Enforcement on Encrypted Network Traffic', *Proceedings of the Military Communications Conference, 2009. MILCOM 2009. IEEE*, PP. 1-7,2009.

[9] Ioannis Koskosas,'A Short Literature Review in Information Systems Security Approaches,' *Proceeding of the IJAKECS*, Vol.1 No.1,pp.1-7, January- June 2013.

[10] Antonio Scarf, 'New security perspectives around BYOD' *IEEE Transactions Seventh International Conference on Broadband, Wireless Computing, Communication and Applications*, Page(s): 446-451,2012.

[11] Mohammad Nauman, Sohail Khan, Xinwen Zhang,'Apex: Extending Android Permission Model and Enforcement with User dened Runtime Constraints', *ASIACCS '10 Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security*, Pages 328-332,2010.

[12] Ionut Andronache, Claudiu Nisipasiu,'Web Single Sign On Implementation SimpleSAMLphp Application', *Journal of Mobile, Embedded and Distributed Systems*, vol. III, no. 1, 2011.