Climate Changes Prediction Using Simple Linear Regression

Article *in* Journal of Computational and Theoretical Nanoscience · February 2019

DOI: 10.1166/jctn.2019.7785

2 authors:

E. Sreehari
VIT University
10 PUBLICATIONS **76** CITATIONS
SEE PROFILE

G S pradeep Ghantasala
Galgotias University
67 PUBLICATIONS **417** CITATIONS
SEE PROFILE

Journal of
Computational and Theoretical Nanoscience
Vol. 16, 1–4, 2019

Climate Changes Prediction Using Simple Linear Regression

E. Sreehari* and G. S. Pradeep Ghantasala

School of Computing Science and Engineering, Galgotias University, Greater Noida 201301, Uttar Pradesh, India

The rise in global temperatures, frequent natural disasters and rising sea levels, reducing Polar Regions have made the problem of understanding and predicting these global climate phenomena. Prediction is a matter of prime importance and they are run as computer simulations to predict climate variables such as temperature, precipitation, rainfall and etc. The agricultural country called India in which 60% of the people depending upon the agriculture. Rain fall prediction is the most important task for predicting early prediction of rainfall May helps to peasant's as well as for the people because most of the people in India can be depends upon the agriculture. The paper represents simple linear regression technique for the early prediction of rainfall. It can helps to farmers for taking appropriate decisions on crop yielding. As usually at the same time there may be a scope to analyze the occurrence of floods or droughts. The simple linear regression analysis methodology applied on the dataset collected over six years of Coonor in Nilagris district from Tamil Nadu state. The experiment and our simple linear regression methodology exploit the appropriate results for the rain fall.

Keywords: Regression, Simple Linear Regression, Prediction, Correlation Coefficient.

RESEARCH ARTICLE

1. INTRODUCTION

Regression is the popular technique used for prediction in areas like climate prediction and other areas. The climate can be defined as the average state of atmosphere over a longer period similarly weather can be changes strongly day by day [1]. The natural disaster which was occurred during the year 2015 November–December south India floods occurred and affected in the regions of Andhra Pradesh and Tamil Nadu. In which it may results in the loss of property approximately 100,000 cores of money, reducing natural resources and more than 500 people were killed. Similarly in the year 2014 a strong tropical cyclone storm occurred in the regions of Vishakhapatnam and caused 60,000 cores of wealth damaged as well as the death rate 125 people were occurred.

On other side most of the people in India can be survived upon the cultivation. There may be a rainfall required for doing cultivation. We are implementing the simple regression model for predicting the future rainfall [2–3]. By these there may be a scope to taking decision on crop yielding for the farmers. If we use the concept of prediction there may be a scope to analyze future occurrence of rainfall or drought, floods can be easily determined. Generally droughts occurred due to low precipitation. By this we can mitigate losses and including property and damage of the public as well as for the peasants. Weather prediction may be at regional or national levels. Generally, two approaches used for predicting rainfall [4–6]. One is Empirical approach and other is Dynamical approach. Empirical based on historical data to be collected and its relationship to various atmospheric variables. Dynamical approach, defines physical models based on systems of equation for prediction and can be implemented by using numerical rainfall forecasting method.

The most widely used empirical approaches which are used for climate prediction [9–11], they are regression, artificial neural network, fuzzy logic and group method of data handling. Support vector machines are a set of supervised learning methods that create a decision maker system which tries to predict new values. A simple climate forecasting can be done by regression techniques. These regression techniques are discussed further. In the recent years, the use of data mining [11–13] in the field of hydrology is increasing. The studies have been performed using data mining process in many areas. Integrated evaporation model, using DM process for three lakes in Turkey by Keskin. Now a day's Artificial Intelligent methods used in the estimation of rainfall [14].

*Author to whom correspondence should be addressed.

2. RELATED WORK

Nikhil Sethi and kanwal Garg implemented a multiple linear regression using climate parameters by considering Udaipur city of Rajasthan climate data set. He can use the linear regression coefficient formula for calculating coefficient values.

H. Hasani presented a model for forecasting the growth of height with based on polynomial regression. This model can be useful for the studying growth of the children.

H. Nguyen describes a model in which it can incorporates artificial neural networks and regression for predicting sales of industry using historical sales and economic indicator.

Imran Ahmad et al., presented a model for predicting rainfall using multiple regressions in the Coonor region of Nilgris district, Tamil Nadu. He can proposed new algorithm for calculating regression coefficient values.

Ismail implemented the model for forecasting to predict price of gold using linear regression. Whatever he can implement using factors inflation and money and it can conclude that MLR perform better than prediction method called Naïve. The forecasting strategies can be analyzed by the data mining scientists and others [15–20].

The rest of the paper is organized as follows. In Section 2, we give a brief overview of the related work. In Section 3, we explain our methodology and solution to the climate prediction using regression [7]. In Section 4, we define the architecture represented other several baselines for regression. Result can be defined in Section 5. In Chapter 6 we define the conclusion and future work. Acknowledgement and references concluded at the end.

3. SIMPLE LINEAR REGRESSION

In the concept of statistical solving simple linear regression is an empirical approach and it can solves the tasks by considering the historical data set of the climate values or parameters. It can be only consists of single dependent variable and independent variable. In the simple linear regression model can be exists only two variables.

The representation of simple linear regression will be like

$$Y = \alpha + \beta X \tag{1}$$

In which $Y =$ dependent variable, $X =$ independent variable, α, $\beta =$ regression coefficients.

In simple linear regression we will implement by calculating Slope and intercept because it will be like mathematical equation of slope, intercept line. The strength and direction of the association between the two variables can be estimated by using the regression coefficient formula. Similarly there are various correlation coefficient formulas can also be available in the mathematical and statistical evolution processing.

The mathematical for r is given as

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \tag{2}$$

The coefficient determination measures how well data can be represented in the regression line. It can define strength and direction of relationship between the dependent and independent variables.

3.1. Error Calculation

The error can be calculated after calculating the predicted values and the difference can be calculated using the actual and predicted values. The formula for calculating error is

$$\text{RMSE} = \sqrt{\frac{(y_i - \bar{y}_l)^2}{n}} \tag{3}$$

$Y_i =$ Actual value, $Y_i =$ Predicted value, $n =$ Total number of samples.

RMSE can be known as root mean square error or root mean square deviation. It can be used measure of difference between sample and population values.

4. RAINFALL PREDICTION USING SLR

The architecture shown in Figure 1 will helps you regarding how to accomplish the task of rainfall prediction in a sequential manner.

The Algorithm for Simple Linear Regression

Step 1: Function (CC, VP, A, B, X, Y, F1)

Step 2: Collect all the data in to excel.

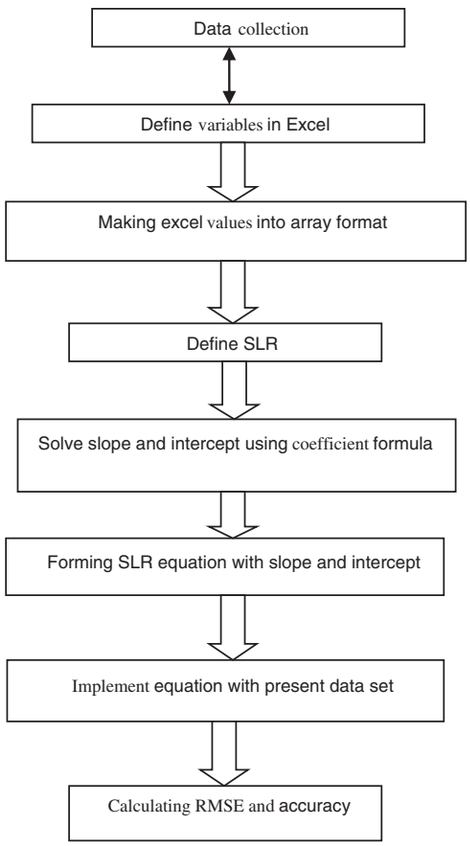

Fig. 1. Architecture to implement SLR.

RESEARCH ARTICLE

Step 3: Read the values from the excel sheet $f1 =$ new file (D:\\sample.xls).
Step 4: CC = [j] = (a. get contents ( )).
Step 5: VP = [j] = (b. get contents ( )).
Step 6: To calculate the correlation coefficient value.
Step 7: Y = A + B x // here B is the correlation coefficient value.
Step 8: we have to find the A = Y–Bx. // here we can get A value.
Step 9: With "A" and "B" values we can form SLR equation.
Step 10: By using this equation with considered data set we can obtained the predicted value which is closest to dependent variable.
Step 11: we have calculate the RMSE
Step 12: we can calculate accuracy by using confusion matrices.

4.1. Initially Consideration of Requirements for SLR

- Data set which is in the form of climate variables with numerical values.
- Data base maintenance by Excel software.
- JDK software.
- The IDE software called Net Beans.
- The package of java Jxl.jar.
- Weka software.

4.2. Steps Regarding the Implementation of a Project

- Defining the data set variables in excel.
- Installing the NetBeans IDE along with JDK if JDK is not installed in your system.
- Creation of project in NetBeans framework.
- Importing the jxl.jar package into the project library folder.
- Declaring the variables and source file path in the by creating a file.
- Implementing correlation coefficient calculation using programming.
- Define simple regression equation.
- Calculating slope value from using simple regression equation and coefficient value.
- Final forming the predicted equation with slope and coefficient values.
- Implementing the equation with the climate parameters and comparing the actual values with the predicted values.
- Calculate the RMSE using past data set and obtained values.
- Install the WEKA software.
- Obtaining confusion matrix using WEKA.
- Calculating specificity and sensitivity.
- Finally implementing the accuracy formula using specificity and sensitivity.

5. RESULTS

In this section we present the experimental results for the regression methodologies simple regression and multiple regression, we have introduced briefly in the previous sections regarding regression concepts. The concept of regression can be implemented by calculating coefficient, slope and the considered climate data set either day, monthly or annual wise. As well as the performance of predicting the future values can be calculated by using simple linear regression algorithm. By using Net beans and weka framework we implemented our project. We can form the regression equation by calculating regression coefficient will be like in this form

$$Y = 20.35 + \text{cloud cover}(-0.005) \quad (4)$$

By substituting the considered data set in the regression equation, we can obtain and carried out the experimental values as shown in Table I.

The final graph representing after implementing my proposed system called simple linear regression and existing system multiple linear regression.

The graph we can obtained like in this way.

Table I represents the predicted values and the values can be estimated with the help of the computed equation. Here the table constructed with the help of simple and multiple regression strategies.

In Figure 2 we compare the both the regression methodologies simple linear regression and multiple linear regressions by considering the same data set. Here we consider the same climate variables for predicting the future values. The existing methodology called multiple linear regressions can require and uses more climate parameters and

Table I. The results of SLR and MLR.

Actual values	SLR values	MLR values
19.28	19.24	19.43
21.38	20.28	19.53
19.24	19.14	19.39
20.488	20.487	18.41
20.444	20.488	20.59
20.487	19.38	18.44

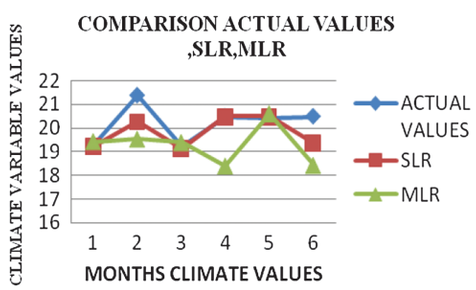

Fig. 2. Comparing of the SLR and MLR values graph.

RESEARCH ARTICLE

it is difficult to implement this regression. In the other way simple linear regression can be easy to implement and analyze.

6. CONCLUSION

The natural incidents may not possible to stop and cannot estimate in a efficient and accurate manner. In general by using the concept of future estimation concept or events or values there may be a scope to minimize lot of problems. In this project we have implemented the simple regression methodology and it can predict the values in a appropriate manner and better than multiple linear regression. By considering rainfall value as a dependent variable with other values as independent we can successfully implemented the simple linear regression method. This is the concept of prediction but not in a accurate manner because we know that climate factors changes due to different reasons and impacts on it.

7. FUTURE WORK

As a future study the logistic regression we can also able to implement by considering the same data set and to accomplish the logistic regression the dependent variable must in be in binary format. Similarly we can obtain results that are called predicted values and the base line of multiple regression concepts can be considered for doing some mathematical calculations. So in future we can also implement the simple regression by using different correlation coefficient formulas with the same methodology. By using different correlation coefficient formulas the predicted equation also changes with different values. However, by implementing this concept, we can achieve the prediction in a better and appropriate manner.

References

1. Divya Chauahan and Jawhar Thakur, **2015**. Data mining techniques for weather prediction. *International Journal on Recent and Innovation Trends in Computing and Communication*.
2. Ozlemterzi Hindwani Monthly Rainfall Estimation by Data-Mining Process by Publishing Corporation **2012**.
3. Nikhil Sethi and KanwalGarg, **2014**. Exploiting data mining technique for rainfall prediction. *International Journal of Comuputer Science and Information Technologies*.
4. Imran Ahmed, SruthiMenon, and Nikitha, **2014**. Rainfall Prediction Using Multiple Regression Techniques.
5. Pinky SaikiaDutta and Hitesh Tahbilder, **2013**. Prediction of Rainfall by Datamining Technique in Assam.
6. Ismail, Z., et al., **2009**. Forecasting gold pieces using multiple linear regression method. *American Journal of Applied Sciences*.
7. Paras, et al., **2012**. A simple weather forecasting model using mathematical regression. *Indian Research Journal of Extension Education*, *1*.
8. http://indiawaterportal.org/met_data/.
9. Damle, C. and Yalcin, A., **2007**. *Journal of Hydrology Flood Prediction Using Time Series Data Mining*, *333*(2–4), pp.305–316.
10. Chau, K.W. and Muttil, N., **2007**. Data mining and multivariate statistical analysis for ecological system. *Coastal Waters Journal of Hydroinformatics*, *9*(4), pp.305–317.
11. Roz, E.P., **2011**. Water quality modeling and rainfall estimation: A data driven approach [M.S.thesis], University of Iowa, Iowa city, Iowa, USA.
12. Keskin, M.E. and Terziin, O., **2009**. *Data Mining Process for Integrate Devoparation Model*.
13. Radhika, Y. and Shashi, M., **2009**. *Atmospheric Temperature Prediction Using Support Vector Machines*.
14. Trafalis, T.B., Richman, M.B., White, A. and Santosain, B., **2002**. Data Mining Techniques for Improved WSR-88D Rain Fall Estimation.
15. Ball, J.E. and Sharma, A., **2001**. *An Application of Artificial Neural Networks for Rainfall Forecasting*.
16. Zhang, M., Scofield, A. and Fulcher, J., **1997**. Rainfall Estimation Using Artificial Neural Network Group.
17. Shoji, T. and Kitaura, H., **2006**. Statistical and Geostatistical Analysis of Rainfall in Central Japan.
18. Ramírez, M.C.V., Velho, H.F.C. and Ferreira, N.J., **2005**. Artificial Neural Network Technique for Rainfall Forecasting Applied to the Sao Paulo Region.
19. Teegavarapu, R.S.V. and Chandramouli, V. **2005**. Improved Weighting Methods, Deterministic and Stochastic Data-Driven Models for Estimation of Missing Precipitation Records.
20. Chiang, Y.-M., Chang, F.J., Jou, B.J.D. and Lin, P.F., **2007**. Dynamic ANN for Precipitation Estimation and Forecasting from Radar observations.

Received: 24 September 2018. Accepted: 20 October 2018.