Hybrid Techniques for Predicting the Rainfall System with Regression Model

[1]B. Venkata Sai Kiran, [2]K. Malathi

[1]UG Scholar, [2]Associate Professor

[1,2]Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India

[1]bvsaikiran98@gmail.com, [2]malathi@saveetha.com

Abstract

In Science and technology Weather forecasting is the important application is used to the weather condition based on the given input parameters and attributes. Many systems are already implemented for the weather prediction using statistical methods. These systems are failed to give the accurate prediction. Using the multi linear regression concept proposed technique can produce better results than the existed systems.

*Keywords***:** *Science & Technology, Rainfall System, Hybrid Techniques*

1. Introduction

In metrology weather prediction is a very complex thing. Previously for the weather prediction relied on the nature events like sunset color and nature odor. Rainfall prediction will help in many fields like Transportation, aircraft, disaster management, defense etc.,

Predict the future is the key issues used in many fields. Time series analysis is one of the common statistical methods used to predict and which are used widely in many applications of statistical and economic terms in which the behavior of the dependent variable prediction based on its behavior in the past. On the other hand, there is a modern method more accurate and effective in forecasting which can use logic in their operations rather than the idea of the fixed relationship between attributes known as Neural Networks.

2. Related Works

In paper 1 authors used k-means clustering algorithm to present the model. For weather predictions implemented density function algorithm for the generation of numerical results. Authors developed a special construction model for knowing the probability by the k means clustering technique. Authors obtained a very good results and accuracy will be shown in the table 1 [1].

In paper 2 authors used SLIQ and proposed a decision tree method for the precipitation model. They find out that the decision tree method archives a very near argument between actual and predicted rainfall. A greater accuracy rate is given by the SLIQ method when compared to the other weather prediction methods [2].

In paper 3 authors are proposed annual average rainfall prediction model on the basis of the BP network combined with discriminant step wise method and Bayesian statistical method used to improve the network's generalization ability and prediction model accuracy, but entire performance has to be developed [3].

In paper 4 authors proposed a time series of analysis which are used as a prediction algorithm. In these authors research on the both rainfall and crop management. For agriculture management decision support system is used by prediction algorithm a system is aimed to develop trend of the rainfall and the result of the accuracy is shown in the table 1 [4].

In paper 5 authors proposed an autoregressive integrated moving average for the rain data forecasting. In this results are more accurate in the individual method other than the ensemble methods [5]. In paper 6 authors proposed an incremental k-means clustering technique for the weather forecasting. In this paper authors taken the air pollution data for the weather prediction based on the atmospheric gas percentage getting the weather prediction values [6].

In paper 7 authors proposed picture fuzzy clustering and spatiotemporal methodology for the weather casting from satellite images by sequencing them. In this paper mainly clusters are used for the pixels which are acquired by the satellite images for this fuzzy clustering is used. For filtering of non predictable values used fourier transform method [7].

In paper 8 authors proposed a genetic algorithm concept on the evolution of the natural genetic mechanisms working on giving the solutions to population contrast. In this paper authors observed the. smoothness of the function which is optimized get a very good performance [8].

Linear regression method proves that accuracy of the prediction model will be high when compared to the other techniques. In this project a hybrid technique is used namely Multi linear regression and neural networking.

Data collection

Data is necessary for the processing of project. In this project data is collected from the Indian Metrological department for the Uttar Pradesh region data is taken for total 4,50,000 records from the year 2011 to the year 2016.

Uttar Pradesh area has recorded for the various climatic changes from o° to 46°, sample data has shown in the table 1.

Figure 1: Project Study Area

Table 1: Sample data

LATITUDE	LONGITUE	ALTITUDE(	TIME(GMT	DATE(GM	TIME(IST)	DATE.IST.	AIR_TEMP	WIND_SPI	WIND_DIF	ATMO_PR	HUMIDITY	RAIN_FAL	SUN_SHIN	BATTERY_V
28.45	80.567	null	0	8/9/2011	5:30	8/9/2011	32.9	0.4	93.8	928.3	88	0	0:00	9953
28.45	80.567	null	1	8/9/2011	6:30	8/9/2011	32.9	0.3	59.1	928.6	90	0	0:00	9953
28.45	80.567	null	2	8/9/2011	7:30	8/9/2011	33.6	0.8	134.9	928.9	79	0	0:00	9953
28.45	80.567	null	3	8/9/2011	8:30	8/9/2011	34	1.1	64	929.2	77	0	0:00	9953
28.45	80.567	null	4	8/9/2011	9:30	8/9/2011	34.9	0.5	111.9	929.3	70	0	0:00	9953
28.45	80.567	null	5	8/9/2011	10:30	8/9/2011	35.6	1	130	929.1	63	0	0:00	9953
28.45	80.567	null	6	8/9/2011	11:30	8/9/2011	36	1.1	134.9	928.9	61	0	0:00	9953
28.45	80.567	null	7	8/9/2011	12:30	8/9/2011	36.2	0.7	176	928.6	61	0	0:00	9953
28.45	80.567	null	8	8/9/2011	13:30	8/9/2011	36.3	1	150	928.3	63	0	0:00	9953

Entire dataset is divided into two parts one is training data and another is the testing data. In this 70% of data is used for training purpose and 30% for testing purpose.
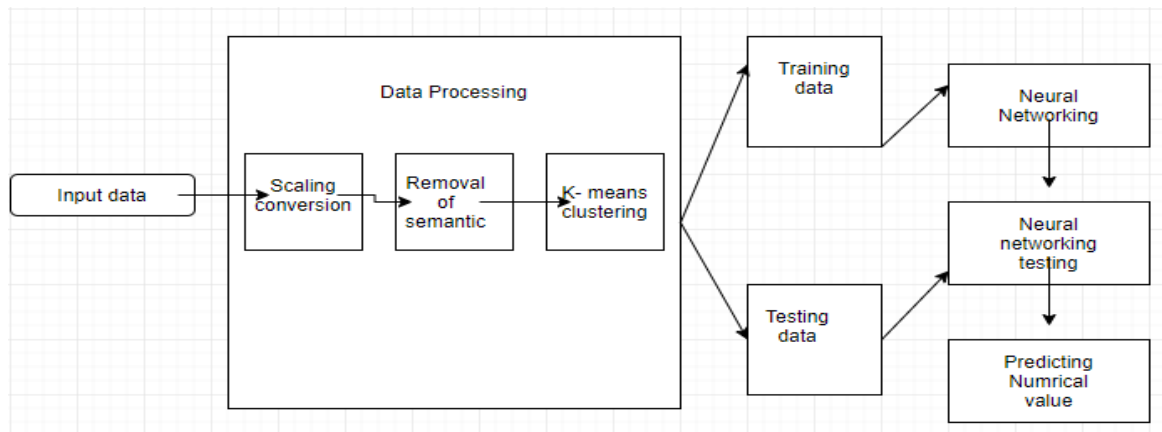
3. Methodology

Figure 2: Architecture of the Proposed Methodology

Figure 2. shows the architecture for the proposed methodology using the Neural Network Back propagations algorithm. Input data is collected from the Indian Metrological Department (IMD).

For the data preprocessing initially input data is taken from the given dataset. At the time of the preprocessing outliers and noisy data will be removed using clusters data will be grouped together using k-means clustering technique. As noted data is divided into two parts training and testing data by using the neural network functions by evaluating the training data with the testing data able to give the accurate value.

Algorithms

a) Multiple Linear Regression

b) Neural Network Back propagation

To form a good relation between the various parameters and rainfall in the dataset multi linear regression will be effective for the prediction of the rainfall value.

Equation for linear regression

$Y=b_0+\sum_{n=1}^{5}(b_n * X_n)$

Here,

y – dependent attribute

b_0 – Intercept of dependent attribute

X_n – Independent varaibles

With the help of this equation able to calculate the error value by comparing the values with testing data.

Training of neural network consists six basic steps

1. Give random weights to the network.
2. Supply the network with inputs which are prepared for training.
3. Apply feed forward process to compute the network outputs.
4. Compare the actual output with desired output and determine the error value.
5. Return the error over the network and correct weights in the direction that the error value decrease.
6. Decrease the total error for each used input in training.
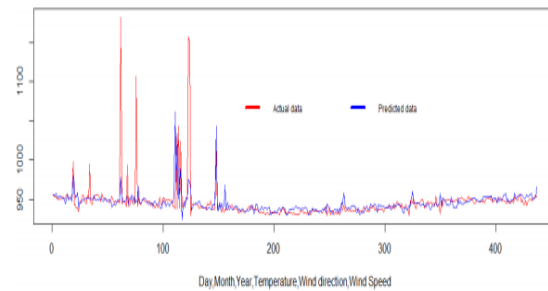
4. Results

Figure 3: Projection of actual and testing data

The above figure describes the projections of the actual data with the help of the test data and predicted values are the outcome of training data is 11.58.

In the below figure displayed the relation between the attributes in the dataset. In this corrplot is ranging between the 1 to -1. Used parameters are air temperature, wind speed, wind direction, atmosphere pressure, humidity, rainfall measurement.
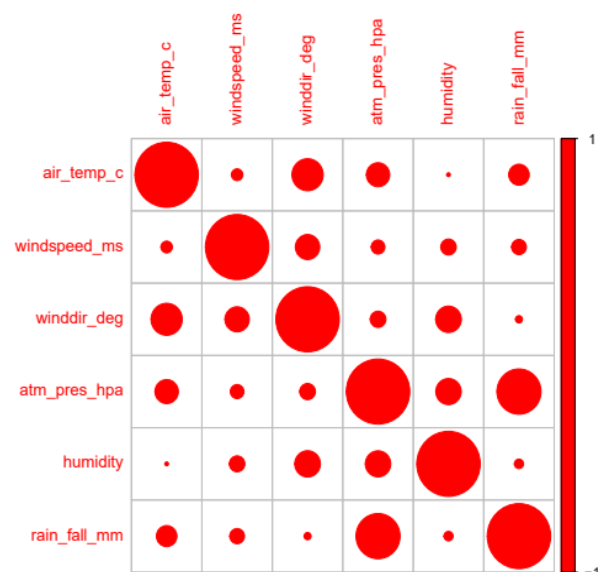

Figure 4: Attributes correlation

5. Conclusion

In this project with the help of the multi linear regression method followed by the neural networking functions able to calculate the accurate result which is far better than the other methods. In rainfall prediction system every single value is important based on that value only results are generated. Neural networking system in this project plays a major role with the help of the back-propagation method able to count and calculate every measurement in the attribute parameters. Data which is used in this project is highly processed and fit for testing and training. Above

generated results are clearly shown that this System is better than the previous systems.

References

[1] Abhay Kumar, Ramnish Sinha, Daya Shankar Verma, Vandhana Bhattacherjee, Satendra Singh (2012), "Modeling using K-Means Clustering Algorithm", First International Conference on Recent Advances in Information Technology, Vol:4, Issue-1, Issue No: 4577-0697, PP:1-5

[2] Narasimha Prasad, Prudhvi Kumar and Naidu MM, "An Approach to Prediction of Precipitation Using Gini Index in SLIQ Decision Tree", 4th International Conference on Intelligent Systems, Modeling and Simulation, 2013.

[3] Dingsheng Wan, Yaming Wang, Nan Gu and Yufeng Yu, "A Novel Approach to Extreme Rainfall Prediction Based on Data Mining", 2nd International Conference on Computer Science and Network Technology, 2012.

[4] Mark Ian Animas, Yung-Cheol Byun, Ma.Beth Concepcion and Bobby D. Gerardo, "Decision Support System for Agricultural Management Using Prediction Algorithm", 2013.

[5] Suhartono, Ria Faulina, Dwi Ayu Lusia, Bambang W. Otok, Sutikno and Heri Kuswanto, "Ensemble Method based on ANFISARIMA for Rainfall Prediction", 2013.

[6] Sanjay Chakraborty, Prof.N.K. Nagwani, Lopamudradey, "Weather Forecasting using Incremental K-means Clustering", International Conference in High Performance Architecture and Grid Computing, Vol.169, Part-2, PP:338-341,2011

[7] Nguyen Dinh Hoa, Pham Huy Thong, Le Hoang Son, "Weather now casting from satellite image sequences using picture fuzzy clustering and Spatio-Temporal regression", international symposium on geo informatics for spatial infrastructure development for earth and allied sciences, PP: 1-6, 2014.

[8] F. Herrera, J.L. Verdegay (Eds.), Genetic Algorithms & Soft Computing, Springer, New York, USA, 1996. [7] J.H. Holland, Outline for a logical theory of adaptive systems