International Journal of Research in Engineering, Science and Management
Volume-2, Issue-3, March-2019
www.ijresm.com | ISSN (Online): 2581-5792

Rainfall Prediction using Deep Learning on Highly Non-Linear Data

M. Selva Balan[1], Jerin Paul Selvan[2], Himanshu Rajeev Bisht[3], Yash Abhijit Gadgil[4], Indrajeet Raghunath Khaladkar[5], Vina M. Lomte[6]
[1]Scientisit-D,CWPRS, Pune, India
[2,3,4,5]Student, Department of Computer Engineering, RMD Sinhgad School of Engineering, Pune, India
[6]Professor & HoD, Department of Computer Engineering, RMD Sinhgad School of Engineering, Pune, India

***Abstract*: Machine learning has become one of the most desired techniques to solve many real time problems. Artificial Neural Network (ANN) and Deep Learning are computational algorithms used in Machine Learning. Artificial Neural Network and Deep Learning make use of historical data to train the systems. The accuracy of the system depends upon the algorithms used to train the system and also on the degree of normalized data. This paper proposes few statistical analysis techniques and the use of Artificial Neural Network to predict rainfall.**

***Keywords*: Artificial Neural Network (ANN), back-propagation, Deep Learning, Epochs, Feed-forward, Normalization, Non-linear, Regression.**

1. Introduction

Flood is one of the most severe disasters which affect both life and property. Heavy rainfall, cloud bursts and untimely opening of flood gates are the reasons for flood. Various regions are affected by floods in India every year, which indirectly affects the whole ecosystem. As per statistics 10 percent of the country is affected by floods. One of the main reasons behind flood is the lack of efficient weather forecasting and management.

Prediction of rainfall is difficult as it depends on many atmospheric parameters like humidity, pressure, wind speed, wind direction and temperature. Weather data is highly non-linear. The non-linearity nature of the data makes it difficult to use regression and other classification algorithms. By using statistical analysis of the dataset it is possible to conclude the convergence of data.

Artificial Neural Network (ANN) is one of the most trending techniques of Machine Learning (ML). ANN is designed by using neurons and hidden layers by applying different activation functions and loss functions. It is efficient to predict regression and categorical values by learning from existing dataset and the relationship obtained between the highly non-linear weather parameters. The accuracy of an ANN model can be improved by preprocessing and normalizing the data. The correlation between the attributes gives an estimate as to how influential that parameter is to prediction.

2. Developed system

In the system developed, a model is designed for rainfall prediction which is used for flood disaster prediction. The input parameters for model are Temperature (T), Maximum Temperature (MT), Minimum Temperature (Tm), Pressure (SLP), Relative Humidity (HUM), Maximum Wind Speed (VV), Minimum Wind Speed (Vm), Average Wind Speed (V), Precipitation (PP) and the like. The model uses ANN techniques. The ANN is network of layers which are interconnected by neurons. The neuron works in similar way as our human brain does. Fig. 1 shows basic ANN architecture which consists of different layers mainly input, hidden and output. Each layer is interconnected to other. Each neuron is initialized with random weights and bias values. The value of each weight and bias is updated at every epoch, to minimize loss. To convert a input signal of a node to output signal of a node activation function is used. For mapping Non-Linear function activation function used in this model is ReLU (Rectified Linear Units).

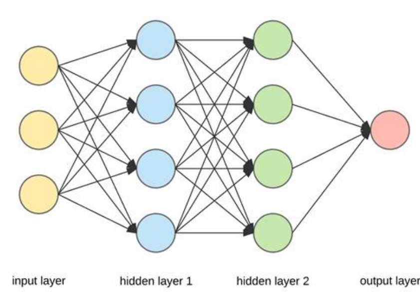

Fig. 1. Architecture of artificial neural network

There are several methods to implement an ANN. The Multi-Layer Feed Forward Network with back-propagation is one of the most promising methods of them all. The Back-propagation is used to reduce the error at each epoch to minimize loss. One feed-forward and back-propagation is called an epoch. The dataset used is of Thiruvananthapuram region of the period 1978-2017 and it consists of 11688 samples and 9 features.

The ANN can be used to predict regression values and categorical values. The attributes must be normalized and pre-processed before importing it into the ANN model. Pre-processing involves the removal of those attributes that do not contribute to the prediction of the result. The best method to find out the contributing attributes is by calculating the correlation between the attributes and the parameter that is to be predicted. Fig. 2 shows the scatter plot of the atmospheric parameters against precipitation. The table 1 shows the correlation and the Pearson's correlation value for the atmospheric parameters and precipitation.

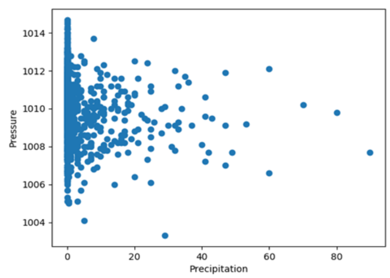

Fig. 2(a). Precipitation against pressure

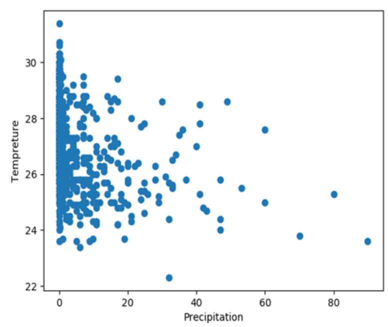

Fig. 2(b). Precipitation against humidity

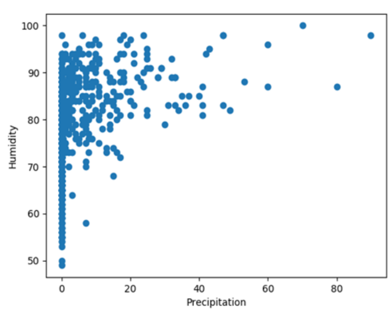

Fig. 2(c). Precipitation against temperature

Table 1
Correlation and Pearson values

Graph Name	Correlation Value	Pearson Value
PP vs T	-0.331825092610193	1.564591768110133e-282
PP vs HUM	0.378491076883631	0.0
PP vs SLP	-0.11384929293409	3.0755531759304078e-33
PP vs TM	-0.330756246249492	1.2774082399112478e-280
PP vs Tm	-0.151020385740237	1.9511285362760716e-57
PP vs VV	-0.214502739064552	2.639729980917295e-115
PP vs V	-0.025161333867961	0.008133204806042705
PP vs Vm	-0.0164937255633134	0.08279985101460159

The correlation value determines the relationship between attributes. Correlation of the attributes is mentioned in the table 1.

The formula of Pearson's Correlation and the normal correlation is mentioned below.

$$r = \frac{1}{n-1} \Sigma \left(\frac{x-\overline{x}}{s_x}\right)\left(\frac{y-\overline{y}}{s_y}\right)$$

$$r = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2 - (\Sigma x)^2][N\Sigma y^2 - (\Sigma y)^2]}}$$

Where:
N = number of pairs of scores
Σxy = sum of the products of paired scores
Σx = sum of x scores
Σy = sum of y scores
Σx^2 = sum of squared x scores
Σy^2 = sum of squared y scores

Artificial Neural Network works with both categorical and regression values. Datasets can be converted into different categories by making use of histogram. With regards to the histogram in Fig. 4 the dataset can be categorized into 6 categories namely category 1 precipitation from 0-10, category 2 precipitation from 10-20, category 3 precipitation from 20-30, category 4 precipitation from 30-40, category 5 precipitation from 40-50 and category 6 precipitation 50 and above.

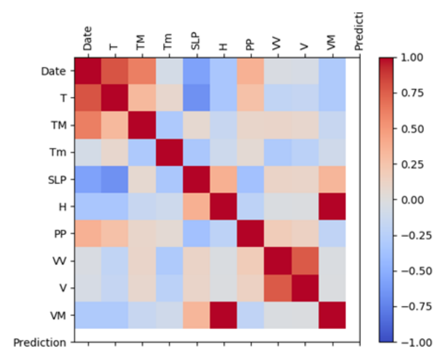

Fig. 3. Correlation of attributes

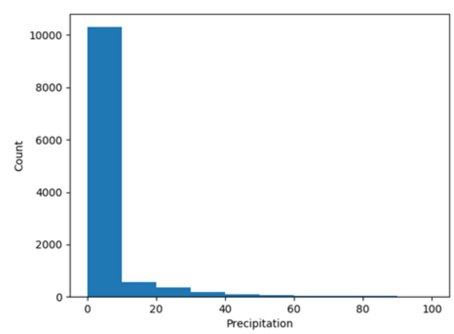

Fig. 4. Histogram of the dataset

The data is divided into 70% training, 10% validation and 20% testing. The missing values are filled using mean strategy. The feature values are normalized using standard normalization. The training data is given to ANN which feed-forwards the data to other layer and calculates the error function (difference between actual and target). Dropout is a technique where randomly selected neurons are ignored while training. Dropout of 20% is used to avoid over fitting of network. The error function used is MSE (Mean Squared Error). The optimizer used is Adam with learning rate 0.01% to evaluate performance of model.

Table 2
ANN Architecture parameters

Architecture Parameters	Values
No of neurons in i/p layer	9
No of hidden layers	5
No of neurons in hidden layer	100
No of neurons in o/p layer	1
Learning rate	0.01
Optimizer	Adam
Maximum epochs	100
Error function	MSE

3. Results

The difference in the result of the model with and without normalization of the dataset is described in the figure given below. Normalization dataset appears to perform better than the other and has minimal loss at the optimal epoch.

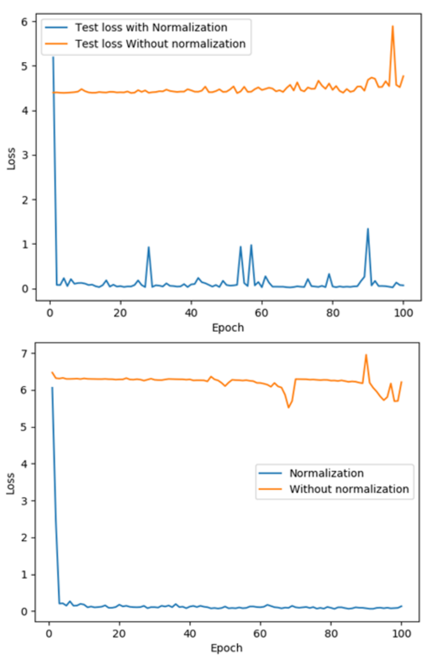

Fig. 5. Loss function of normalized and non-normalized data

4. Conclusion

The Artificial Neural Network works well with pre-processed and normalized data. The correlation between the attributes defines the influence on the prediction of any system. Attributes with no correlation can be removed as they do not contribute to the activation of the neuron. The loss functions, activation functions, number of neurons and the number of hidden layers also affect the accuracy of the system. This model can be expanded to a system that predicts rainfall a day in advance and to estimate the occurrence of flood. This will be beneficial in developing an alert system.

Acknowledgement

The authors would like to thank Central Water and Power Research Station, Pune for providing us with necessary data and guidance.

References

[1] Swapnil Bande, and Virendra V. Shete, "Smart flood disaster prediction system using IoT & Neural Networks", IEEE, International Conference On Smart Technology for Smart Nation, pp. 189, 2017
[2] Sulafa Hag Elsafi, "Artificial Neural Networks for flood forecasting at Dongola station in the River Nile, Sudan, Alexandria University", Alexandria Engineering Journal, Elsevier: pp 655-662, 2014.
[3] A. R. Ghumman, A. R. Sohail, K. Watanbe, Yousry M. Ghazaw, "Runoff forecasting by Artificial Neural Network and conventional model", Alexandria University, Alexandria Engineering Journal, Elsevier: pp 345-353, 2011.
[4] A. Stanley Raj, D. Hudson Oliver, J. Viswanath, Y. Srinivas, "Wavelet based analysis on rainfall and water table depth forecasting using Neural Networks in Kanyakumari district, Tamil Nadu, India", Groundwater for Sustainable Development, Elsevier: pp 178-186,2017
[5] Sruthi R, "Medium Access Control Protocols for Wireless Body Area Networks: A Survey, Global colloquium in recent advancements and effectual researches in engineering, science and technology", Elsevier: pp 621-628, 2016.
[6] Jang-Ping Sheu, Prasan Kumar Sahoo, "Design and Analysis of collision free MAC for Wireless Sensor Networks with or without data retransmission", Journal of Network and computer applications, Elsevvier: pp 10-21, 2017.
[7] A. Aitalieva, G. Celik, H. Celebi, M. Uysal, "A software defined radio based data link design for VHF band wireless sensor networks", Physical communication, Elsevier: pp 167-178, 2018.
[8] Ahlam Saud Althobaiti, Manal Abdullah, "Medium Access Control protocols for Wireless Sensor Networks Classification and Cross-Layering", Intenational Conference on communication, Management and Information Technology, pp 4-16, 2015.
[9] Yoshikazu Hayashi, "Confidence intervals of climatic signal", Journal of the atmospheric sciences, Volume 39 Number 9, September 1982.