Rainfall Prediction And Analysis

Aravind V R

MCA, Student

SCOPE VITCC

aravindvalsarajan@gmail.com

Dr Shyamala L

Associate Professor

(SCOPE)

VITCC

shyamalal@vit.ac.in

Abstract

Rainfall is one of the important aspects of the environmental system that needs to be analysed in order to find the climate changes and also the environmental health of a particular geographical area. Sincere approaches have to be taken into consideration in the prediction and analysis of rainfall. Rainfall forecasting helps fields such as agriculture, whether prediction. Prediction and analysis aid in diagnosing the different nature of rainfall at different times so that suitable countermeasures against climate change can be performed. Numerous approaches aid in determining the most appropriate data diagnosis and that helps to do the necessary precautions. There are several modes for analysing and predicting but every method does not provide good results. This study investigates the methods used to summaries data based on the techniques used. The requirements are gathered and examined in order to provide an accurate rainfall forecast model that can be compared to other models. The paper compares various methodologies such as WNN, HWNN, ANN, and Deep Learning and proposes the best model among them with better accuracy by summarizing the past and current rainfall data to refrain the mode.

Keywords: WNN, HWNN, ANN, Deep Learning

1. Introduction

Rainfall is not easy to predict, due to its varying nature which is dependent on time and space. There is always an interest in forecasting the rainfall and all other natural habitat and forecasting nearly started around the start of 650 BC. This paper summarizes various methods used and proposes the best among them to adapt and also discusses the various reasons and strategies that are used to find the best method. The methodologies discussed are changed from years to year, the methodologies such as WNN mentioned have changed recently in to various deep learning and machine learning techniques. Rainfall prediction has much importance in the daily life of Indians, it either affects India in a manner of disaster or as a blessing in agriculture areas, there were several floods in last years and many of them was due to heavy rainfall.

The process of analyzing and proposing methodologies require comparison of these methodologies in aspect of various measures such as RMSE, NSE and R etc. The best one, proposed should be able to do the prediction effectively and also must be able to analyze the data.

There are significant changes in each approach that results unique results. The HWNN results better in prediction when compared with WNN [1], even though it can't be used in wet and dry months or the calculated value of HWNN on these months can't be taken for analysis.

Guided by: Shyamala P

2. Methodologies

There are several constraints those are validated with respect to a methodology like RMSE, coefficient of correlation, NSE etc, this is discussed here.

A. Deep learning

A hybrid DL approach is proposed which is a combined of one-dimensional Convolutional Neural Network (Conv1D) and Multi-Layer Perceptron (MLP). There are also other DL approaches like Multi-Layered Perceptron (deep MLP) and machine learning approach called Support Vector Regression (SVR). Deep Learning (DL) is a popularized AI approach that has numerous multi-dimensional benefits. The hybrid model takes nine meteorological variables as inputs, all of which are strongly related to daily rainfall variation. A General Circulation Model(GCM) is used to obtain the causative variables.GCM simulation of meteorological variables outperforms rainfall estimates, because observational data of meteorological variables are scarce and not completely unavailable in many places.

The Conv1D layer or also the main building block of CNN, and which is here combined with MLP to do deep learning.

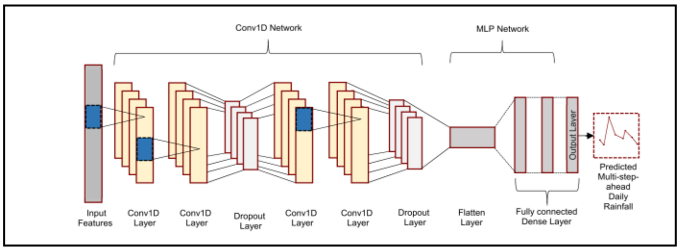

Fig. 1. Picture of of conv1D+MLP network

Conv1D layers are the 1st and 2nd layer added to the model in Fig 1. The convolutional layers aid in the detection of patterns and the extraction of hidden information from the input data.

ESN/DeepESN model can be used to forecast rainfall by actual testing and verification. The DeepESN model does ESN and and some Neural Network techniques (.Back-propagation network and Support Vector Regression, respectively). As a result, the DeepESN outperforms the other models here [7] in terms of rainfall prediction. Finally, investigating the impact of each input parameter by selecting an alternate input parameter. DeepESN is a mathematical model like all other models that we found. It demonstrates that rainfall, pressure, and humidity are the most important variables, and model's rainfall forecast performance is heavily influenced.

B. WNN combined with ANN

The neural network was fed decomposed details (D) and approximation (A) as inputs. There were several algorithms that was used to train the neural network structure to achieve the ideal weights (parameters). The conclusion of these studies is a traditional MLP with a logarithmic sigmoidal transfer function(because here we need the output in the form of probability) for the hidden layer, use a linear transfer function, and for the output layer, use a linear transfer function. Trial and error was used to determine the number of concealed nodes.. The output node will be one step ahead of the original value. In Fig 2 ,the layers are attached to connected weights and then they are sent to the hidden layers then in the output they are merged altogether.

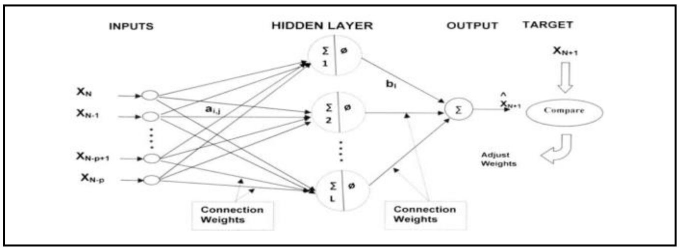

Fig. 2. WNN combined with ANN

C.WNN Used

Wavelet analysis is a type of signal processing that is commonly used in optical engineering and communications and is a pretty ancient one to deal with. It is a non-linear process that can be utilized to analyses and preserve non-linear phenomena. There is always a comparison between Fourier transforms and wavelets and the difference turned out to be that the wavelets can pinpoint the exact location of a change in a sequence's dynamical patterns, while the Fourier transforms can only provide its frequency. Wavelet analysis can be defined as breaking up a signal in to ascended version of the original wavelet. The ultimate result will be a signal representation in time and frequency with various resolutions. There will be a collection of signals that gives light in to multiresolution analysis. The variability in all case will be understandable for a person who decomposes this timeline.

Wavelets is a powerful tool in the case of long memory processes for analyzing and synthesizing of the data.

The fundamental goal of wavelet analysis is to determine a signal's frequency (or scale) content, then examine and determine the temporal fluctuation of that frequency content. This trait is diametrically opposed to Fourier analysis, which can determine the frequency content of a signal but cannot determine frequency-time dependency.

When signals are characterized by localized excessive frequency occurrences or when signals are defined by a large number of scale variable processes, the wavelet transform is the tool of choice.

There can be discrete as well as continuous wavelet transform. The decomposition of signal x(t) in terms of elementary contribution is turned out to be wavelets, CWT is the tool which allows this decomposition. In real-world applications, input data f(t), yearly run-off time series, monthly precipitation time series, and water quality time series are examples of these types of time series., is often discretely sampled.

$$\boldsymbol{f(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\, \varphi\left(\frac{t-b}{a}\right) dt} \quad (1)$$

The equation for wavelet transformations of a time series (1) is as follows. Here, 'a' stands for the characteristic frequency, is determined by a scale or dilation factor. And 'b' stands for the temporal translation that depicts the wavelet's 'sliding' across f(t). When |a| is less than 1,the wavelet will be shrunken and compressed along with higher frequency bands. Whenever |a| is greater than 1 then there will be low frequencies mostly. It's better to have a larger scale when we need a general view of the signal and small scales help to get a more detailed view

In the case of continuous wavelet transform (CWT) they can be described as (2).

$$\varphi_{a,b}\,(t) = \frac{1}{a}\varphi(\frac{t-b}{a}) \quad (2)$$

Here also 'b' is the translation time and 'a' is the factor which represents dilation.

D. ANN Used

An artificial neural network (ANN) is a system or mathematical model that consists of a large number of disordered artificial neurons that function in parallel. ANN models are the so called 'black box' models which does have specific qualities that are well suited in simulating dynamic nonlinear systems. The most well-known ANN design in hydrologic modelling is the multilayer perceptron (MLP) trained with the BP technique. A multilayer perceptron network consists of an input layer, one or more hidden layers of computation nodes, and an output layer. The nature of the real input and output determines the number of input and output nodes. The number of hidden nodes, is always selected by the modeler, frequently by trial and error, depending on the mathematical structure of the issue. The information provided Layer after layer, the signal travels through the network in a forward manner. Each and every hidden and output node processes its input by multiplying each value by a weight and adding the results to obtain a result, and then putting the sum through a nonlinear transfer function. . This process is referred to as training. It comes to a halt when the number of errors is reduced to a minimum or another norm is reached. The Back Propagation Neural Network (BPNN) is a type of neural network that may be written as

$$Y = f\left(\sum WX - \theta\right) \quad (3)$$

In the equation (3) , where X is the input or hidden node's value; Y is the hidden or output node's output value; f () is the transfer function; W is the weights connecting the input and hidden nodes; and is the bias of each node (or threshold).

The given network was trained using the Levenberg-Marquardt method (LM). Because of its high computational and memory requirements, the Levenberg-Marquardt method algorithm (LMA) can only be employed in small networks. Increasing the number of training patterns gives you more information about the solution surface's geometry. Allowing the network to reach a higher level of accuracy. As a result, there's There is no certainty that adding more training patterns will make a difference. The amount of time it takes to train a network grows as the number of patterns in the training set grows. When there was no substantial gain in efficiency, the training was ended, and the model's generalization properties were examined

RESULTS

The results make us understand about the efficiency of methodologies as well as approaches, in some way we could combine methodologies and results also examines whether those combined results give better accuracy or not.

When we consider a chronology, ANN and WNN are pretty old ways to do so. The new model techniques include machine learning as well as Deep Learning [4]. The Deep Learning has several techniques such as Multi-Layer Perception (MLP), Convolutional Neural Network(CNN), Recurrent Neural Network(RNN), Long-Short Term Memory (LSTM). While saying WNN and ANN are old, it is important to mention that these technologies still have the same importance they had and also in many cases ANN is still in use and WNN stands out better than many technologies such as HWNN.

The techniques in deep learning as well as machine learning are used in several fields such as image recognition, speech recognition etc. The hybrid model takes nine meteorological variables as inputs, all of which are strongly related to daily rainfall variation. A General Circulation Model is used to obtain the causative variables (GCM). A one-dimensional Convolutional Neural Network (Conv1D) and a Multi-Layer Perceptron (MLP) combination is studied [4]. The suggested hybrid approach can be compared to Multi-Layered Perceptron (deep MLP), a deep learning approach, and Support Vector Regression, a machine learning approach (SVR). The deep MLP has three layers: an input layer for feeding data, for computing, a lot of hidden layers, and a prediction a value or a vector of values layer. In the hidden and output layers of MLPs, an activation function transfers the the weighted input multiplied by the neuron's output. The proposed architecture is educated on a collection of to discover the link between input and output data after the layers have been configured. SVR is a Support Vector Machine-based supervised machine learning technique. It aids in the linear separation of input data by translating it to a three-dimensional environment using a function of nonlinear mapping. SVR was optimized using two regularization parameters, Gamma () and the cost function (C) of the radial basis function (RBF). Root Mean Squared Error (RMSE), coefficient of correlation (r), and Nash–Sutcliffe Efficiency are three statistical measures used to compare the performance of the Conv1D-MLP hybrid model to that of SVR, and deep MLP models (NSE). The correlation coefficient (4) is a metric for how well two variables are related linearly.

$$r = \frac{\sum_{t=1}^{n} \left(Y_t - \bar{Y}\right)\left(Y_t' - \bar{Y}'\right)}{\sum_{t=1}^{n} \left(Y_t - \bar{Y}\right) \sum_{t=1}^{n} \left(Y_t' - \bar{Y}'\right)}$$

(4) Correlation coefficient

The range of r is [1, 1], with 1 denoting a complete negative linear relationship and 0 denoting no linear association. The model's performance improves as the value of r increases. which is the same case in[2]. The root mean square error (RMSE) is a commonly used metric for determining the distinction between actual and anticipated values. It's mostely a positive number, and a lower RMSE means the model is doing better. The equation for the RMSE measure is (5) , where n depicts the anticipated value.

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} \left(Y_t - Y_t'\right)^2}{n}}$$

(5) Root mean square value

The suggested model's efficiency is measured using NSE. The NSE ranges from (-∞, 1]. A value greater than 0 implies that the model is more efficient, whilst a value of 0 shows that the projected values are as excellent as they can be as the average of values,Values that have been seen of the NSE values less than that are less than zero shows model's performance is unacceptable. In some other cases as discussed [table 1] the lower RMSE and higher NSE higher value of R indicate better performance of the model

Whenever artificial neural network-based technique to classifying rainfall patterns was examined. To pick features that will expose the maximum accuracy for monthly rainfall, a particle swarm optimization approach was used.

For daily rainfall prediction, a hybrid model was created by Support vector regression and artificial neural networks are being combined. [6]. In the modelling process, raw rainfall data was broken down using singular spectrum analysis. The prediction uses Clustering, or fuzzy c, is a technique for dividing a training set into subgroups. The methodologies such as Singular spectrum analysis (SSA), Artificial neural networks (ANN), Support vector regression (SVR), Fuzzy C-means (FCM) are discussed in this. While considering each one of them, Then SSA can decompose daily rainfall data into a number of additive components known as 'trend' components (which may not exist). When the SSA is utilized as the signal filter technique, the noise and/or high-frequency oscillatory component can be filtered outs. The SSA technique is explained in four phases using this methodology, which is based on a univariate rainfall time series. The noise and/or high-frequency oscillatory component can be filtered out when the SSA is used as the signal filter technique.

In the analysis between ANN and WNN The observed and modelled values during calibration and validation of ANN and WNN models are taken and analyzed. It was discovered that values predicted using a The WNN model accurately matched the values observed, whereas the ANN model overestimated the observed values In terms of rainfall prediction, the WNN model beat the ANN and AR models, according to this study. In the graph of observed and modelled rainfall of ANN there is significant difference in the points where the modelled and observed are marked. In the graph of the observed and modelled rainfall of WNN there is a visible similarity between these marked points. It's worth noting that the WNN model outperformed the regular ANN and AR models in terms of rainfall forecasting.

Figures and Tables

TABLE 1: COMPARISON BETWEEN DEEP LEARNING AND MACHINE LEARNING APPROACHES

MODELS	RMSE	NSE	R
Conv1D-MLP	POSITIVE AND LOWER	GRATER THAN ZERO	HIGH
Deep MLP	HIGHER (comparatively)	LOWER THAN Conv1D-MLP HIGHER THAN SVR	LOWER THAN Conv1D-MLP
SVR	HIGHER THAN BOTH Conv1D-MLP AND Deep MLP	LOWER (NEAR TO ZERO)	

TABLE 2: COMPARISON BETWEEN NEURAL NETWORK MODELS

MODELS	RMSE	R	COE
WNN	LOWER	HIGHER	HIGHER
ANN	HIGHER	LOWER	LOWER
AR	HIGHER	LOWER	LOWER

In deep learning and machine learning approaches whenever deep learning has incorporated with Neural networks then it results far better than many machine learning approaches such as SVR from TABLE 1. In the case of Neural Networks WNN performs better than ANN and even HWNN whenever we consider all circumstances such as prediction and analysis on wet and dry surfaces. It is clear from TABLE 2 that WNN outperforms all other including AR.S

Conclusion

In the comparison and understanding of various methodologies, it’s visible that the methodology involving deep learning stands out and it is better to predict and analyze the data. The various approaches of deep learning show more of a good performance when we compare the performances of the methodologies. In many cases considered deep convolution network stands out as a best methodology that is to be used for.

WNN and all aged methodologies could also be used in order to understand the difference and only in scenarios where people want to compare them. The focusing of a particular area while considering the methodologies and prediction would also give the result after the comparison that deep learning technique stands out and provides a far better performance.

REFERENCES

[1] Zhang, S., Chang, T. and Lin, D., 2018. A preliminary study on a hybrid wavelet neural network model for forecasting monthly rainfall. EURASIA Journal of Mathematics, Science and Technology Education, 14(5), pp.1747-1757.

[2] Tran Anh, Duong, Thanh Duc Dang, and Song Pham Van. "Improved rainfall prediction using combined pre-processing methods and feed-forward neural networks." J 2.1 (2019): 65-83.

[3] Mishra, Neelam, et al. "Development and Analysis of Artificial Neural Network Models for Rainfall Prediction by Using Time-Series Data." International Journal of Intelligent Systems & Applications 10.1 (2018).

[4] Khan, Mohd Imran, and Rajib Maity. "Hybrid deep learning approach for multi-step-ahead daily rainfall prediction using GCM simulations." IEEE Access 8 (2020): 52774-52784.

[5] Hettiarachchi, P., Hall, M. J., & Minns, A. W. (2005). The extrapolation of artificial neural networks for the modelling of rainfall—runoff relationships. Journal of Hydroinformatics, 7(4), 291-296.

[6] Chau, K. W., & Wu, C. L. (2010). A hybrid model coupled with singular spectrum analysis for daily rainfall prediction. Journal of Hydroinformatics, 12(4), 458-473

[7] Haidar, Ali, and Brijesh Verma. "Monthly rainfall forecasting using one-dimensional deep convolutional neural network." IEEE Access 6 (2018): 69053-69063.

[8] Yen, Meng-Hua, et al. "Application of the deep learning for the prediction of rainfall in Southern Taiwan." Scientific reports 9.1 (2019): 1-9.

[9] T.Kashiwao, K. Nakayama, S. Ando, K. Ikeda, M. Lee, and A. Bahadori, ''A neural network-based local rainfall prediction system using meteoro logical data on the Internet: A case study using data from the Japan Mete orological Agency,'' Appl. Soft Comput., vol. 56, pp. 317–330, Ju

[10] Haidar and B. Verma, ''Monthly rainfall categorization based on opti mized features and neural network,'' in Proc. 30th Australas. Joint Conf. Artif. Intell., Melbourne, VIC, Australia, 2017, pp. 208–220. Monthly Rainfall Categorization Based on Optimized Features and Neural Network | SpringerLink