PERFORMANCE ANALYSIS AND EVALUATION OF MACHINE LEARNING ALGORITHMS IN RAINFALL PREDICTION

[1]AMARAKONDA PREETHI, [2]Dr. P. VELAYUTHAM
[1]MCA Student, [2]Assistant Professor
DEPARTMENT OF MCA
SREE CHAITANYA COLLEGE OF ENGINEERING, KARIMNAGAR

ABSTRACT

Rainfall is one of the most significant parameters in a hydrological model. Several models have been developed to analyze and predict the rainfall forecast. In recent years, wavelet techniques have been widely applied to various water resources research because of their time frequency representation. In this paper an attempt has been made to find an alternative method for rainfall prediction by combining the wavelet technique with Artificial Neural Network (ANN). The wavelet and ANN models have been applied to monthly rainfall data of Darjeeling rain gauge station. The calibration and validation performance of the models is evaluated with appropriate statistical methods. The results of monthly rainfall series modeling indicate that the performances of wavelet neural network models are more effective than the ANN models

1. INTRODUCTION

Rainfall is a complex atmospheric process, which is space and time dependent and it is not easy to predict. Due to the apparent random characteristics of rainfall series, they are often described by a stochastic process (Chinchorkar et al. 2012). For water resources planning purposes, a long-term rainfall series is required in hydrological and simulation models (Tantanee et al. 2005). There have been many attempts to find the most appropriate method for rainfall prediction for example, coupling physical, marine, and meteorological or satellite data with a forecasting model, or even applying several techniques such as the artificial neural network or fuzzy logic as a forecasting approach (Hsu et al. 1995; Dawson and Wilby 2001;

Hettiarachchi et al. 2005). In recent years, several numerical weather forecasts have been proposed for weather prediction but most of these models are limited to short period forecasts. This paper introduces a new approach for prediction of rainfall series. Several time series models have been proposed for modeling monthly rainfall series (Bhakar et al. 2006) and annual rainfall series such as the autoregressive model (AR) (Yevjevich 1972), the fractional Guassian noise model (Matalas and Wallis 1971), autoregressive moving-average models (ARMA) (Carlson et al. 1970) and the disaggregation multivariate model (Valencia and Schaake 1973). Moustris et al. (2011) examine the possibility of long term precipitation forecast (four consecutive months) by the application of ANNs, using long monthly precipitation time series of four meteorological stations in Greece. In the past decade, wavelet theory has been introduced to signal processing analysis. In recent years, the wavelet transform has been successfully applied to wave data analysis and other ocean engineering applications (Massel 2001; Teisseire et al. 2002; Huang 2004). The time-frequency character of long-term climatic data is investigated using the continuous wavelet transform technique (Lau and Weng 1995; Torrence and Compo 1997; Mallat 1998) and

wavelet analysis of wind wave measurements obtained from a coastal observation tower (Huang 2004). Chou (2011) used wavelet denoising method in linear perturbation models (LPMs) and simple linear models (SLMs) for rainfall and runoff time series data. Wang and Li (2011) used a new wavelet transform method for developing the synthetic generation of daily stream flow sequences. Wu et al. (2004) used a combination of neural networks and wavelet methods to predict underground water levels. Dynamical Recurrent Neural Network (DRNN) on each resolution scale of the sunspot time series resulting from the wavelet decomposed series with the Temporal Recurrent Back propagation (TRBP) algorithm (Aussem and Murtagh 1997). There are some appreciable studies of wavelet transform based neural network models (Wang and Ding 2003; Anctil and Tape 2004; Cannas et al. 2006; Kisi 2008; Wang et al. 2009). The wavelet transform is also integrated with multiple linear regression (Kucuk and Agiralioğlu 2006; Kisi 2009, 2010) and support vector machine approach (Kisi and Cimen 2011). Adamowski and Sun (2010) compared the relative performance of the coupled wavelet-neural network models (WA–ANN) and regular artificial neural networks (ANN) for flow forecasting at lead times of 1 and 3 days for two different non-perennial rivers in semiarid watersheds of Cyprus. Kisi (2011) investigated the performance of the wavelet regression (WR) technique in daily river stage forecasting and determined the WR model was improved combining two methods, discrete wavelet transform and a linear regression model. Sang (2013), developed a method for discrete wavelet decomposition and an improved wavelet modeling framework, WMF for short was proposed for hydrologic time series forecasting. By coupling the wavelet method with the traditional AR model, the Wavelet-Autoregressive model (WARM) is developed for annual rainfall prediction (Tantanee et al. 2005). Partal and Kisi (2007) used a conjunction model (wavelet-neuro-fuzzy) to forecast the Turkey daily precipitation. The observed daily precipitations are decomposed to some sub series by using Discrete Wavelet Transform (DWT) and then appropriate sub series are used as inputs to neuro-fuzzy models for forecasting of daily precipitations. Each of these studies showed that different black box models trained or calibrated with decomposed data resulted in higher accuracy than the single models that were calibrated with an unrecompensed and noisy time series. In this paper, a Wavelet Neural Network (WNN) model, which is the combination of wavelet analysis and ANN, has been proposed for rainfall forecast Darjeeling station, India.

II. SYSTEM ANALYSIS

EXISTING SYSTEM

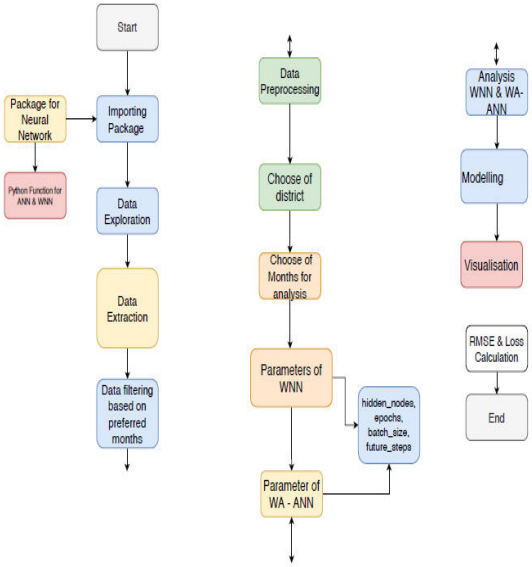
Prediction of the rainfall using other independent parameters (temperature, humidity, pressure, wind speed etc.) is attempted in many studies showing the comparison of different machine learning techniques and claiming the higher ac-curacy with categorizing rainfall in two to three categories, but most of them have not attempted the forecasting of rainfall for next season using machine learning techniques. In Few papers forecasting of the rainfall as well as different weather parameters like temperature, relative humidity, number of rainy days etc. is attempted. The result shows forecasting rainfall individually gives less accurate result compared to other weather parameters.

As forecasting rainfall individually using forecasting techniques gives less accuracy and prediction of rainfall with the help of different weather parameter using machine learning techniques gives higher accuracy it is necessary to design the fusion model.

PROPOSED SYSTEM:

By coupling the wavelet method with the traditional AR model, the Wavelet-Autoregressive model (WARM) is developed for annual rainfall prediction (Tantanee et al. 2005). Partal and Kisi (2007) used a conjunction model (wavelet-neuro-fuzzy) to forecast the Turkey daily precipitation. The observed daily precipitations are decomposed to some sub series by using Discrete Wavelet Transform (DWT) and then appropriate sub series are used as inputs to neuro-fuzzy models for forecasting of daily precipitations. Each of these studies showed that different black box models trained or calibrated with decomposed data resulted in higher accuracy than the single models that were calibrated with an unrecompensed and noisy time series. In this paper, a Wavelet Neural Network (WNN) model, which is the combination of wavelet analysis and ANN, has been proposed for rainfall forecast Darjeeling station, India.

ARCHITECTURE DIAGRAM



III. **IMPLEMENTATION**

MODULES:

Import packages

Here has to upload the following packages from to built the project python-graphviz-0.14.2, numpy , pandas , seaborn , matplotlib, plotly , keras , tensorflow , sklearn

Data exploration

Data exploration is a key aspect of data analysis and model building. Without spending significant time on understanding the data and its patterns one cannot expect to build efficient predictive models. Data exploration takes major chunk of time in a data science project comprising of data cleaning and preprocessing.

Data extractions

Data extraction is the process of retrieving data from data sources for further data processing or storage.

Data extraction is perhaps the most important part of the extract/translate/load (etl) process because it inherently includes the decision making on which data is most valuable for achieving the business goal driving the overall etl.

Sometimes data is relatively easy to extract because it exists in a structured data store such as a relational database management system (rdbms) in which case, there is a well defined standardized structured query language (sql) that is very powerful for doing targeted extracts of exact data. Sql may also be used to do some level of translation/transformation making it even more powerful.

In other cases, the data exists in a non-sql based data store or even spread across many different digital, and potentially non-digital, formats. In this case, more exotic tools or bespoke code is required. Unstructured data extraction generally makes the projects longer so the general rule of

understanding the value of the data you are going to extract is even more important.

Data filtering based on preferred months

Here we can filter the data based on the months, rain falls, etc..

IV. CONCLUSION

Rainfall plays a vital role in all aspects. It is better to predict future rainfall and take if any necessary steps need to be taken to minimize the amount of life loss. In our proposed work, we have proposed a model that could predict or forecast the rainfall of any particular location where the classifier is trained with numerous previous weather information. The classification was performed using Neural Networks using Linear Regression algorithms. The proposed model had good accuracy. The algorithm extracted the features of the dataset from the training set and was able to accurately predict the weather well in advance. Various other parameters such as efficiency and computational time for the process was also considered which was found to be better than the traditional methods. The future works may include the use of other deep learning techniques and also working on the security concerns of the model.

REFERENCES

1. Rahman, M. M., Bhattacharya, P., & Desai, B. C. (2007). A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. IEEE Transactions on Information Technology in Biomedicine, 11(1), 58-69.

2. Morales, M., Tapia, L., Pearce, R., Rodriguez, S., & Amato, N. M. (2004). A machine learning approach for featuresensitive motion planning. In Algorithmic Foundations of Robotics VI(pp. 361-376). Springer, Berlin, Heidelberg.

3. Ireland, G., Volpi, M., & Petropoulos, G. P. (2015). Examining the capability of supervised machine learning classifiers in extracting flooded areas from Landsat TM imagery: a case study from a Mediterranean flood. Remote sensing, 7(3), 3372-3399.

4. Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. International Journal of remote sensing, 23(4), 725-749.

5. Kotsiantis, S. B., Zaharakis, I., &Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering, 160, 3-24.