Prediction of Indian Monsoon using Machine Learning

Abstract

Accurate prediction of Indian monsoon rainfall is highly desirable; however, it is one of the toughest challenges the scientific community is facing today. Among the various scientific approaches, statistical, numerical, and empirical modeling are popular and widely used in recent times for monsoon prediction, whereas machine learning has not gained any such attention, which although a potential approach, may be able to solve this problem. Rainfall is a product of several atmospheric and oceanic processes and its prediction remains a daunting task. While physically based numerical modeling has progressed considerably over the last half-century, due to the inherent complexities of the underlying physical processes, alternative approaches are explored to make the predictions accurate. Although the dynamical models are capable of providing temporal and spatial distribution of rainfall using cause-effect association among several atmospheric processes, the prediction accuracy by the climate models is not yet at the desired stage and hence, statistical models are still preferred for the rainfall prediction. The statistical models are based on the association of the Indian monsoon with various climate predictors, but, the method of finding a good predictor is quite complex as the monsoon rainfall depends on several weather-related phenomena happening around a year or beyond several years. The other approach besides the aforesaid numerical and statistical based approaches is the empirical approach which uses the past rainfall values to predict the future rainfall values. The conventional models of empirical modeling were based on simple regression based optimization methods; thus failed to produce a higher degree of accuracy in the prediction of an unknown parameter.
In the empirical approach, the rainfall time series is supposed to carry the imprint of all causes and weather phenomena in itself. Thus, it is necessary to analyze the rainfall time series and extract important patterns hidden inside the time series data, and for this, the explorative data analysis method (Empirical Mode Decomposition- Detrended Fluctuation Analysis (EMD-DFA) was applied in this study. The Indian monsoon rainfall datasets were decomposed into a finite number of empirical modes called Intrinsic Mode Functions (IMF). The obtained IMFs play an important role in rainfall prediction and are also providing a physical basis to relate monsoon rainfall with different meteorological parameters, which show a similar period of occurrences as a particular IMF. The EMD-DFA approach helps us to find out that the first few empirical modes are nonlinear and the rest are linear. Among the different modes of IMFs, the last mode represents the residual, which shows the long-term climatic average and always produces constant values for each corresponding year. According to the current study, the non-linear modes can explain more than 50% of inter-annual variability. The first few IMFs exhibit high-frequency modes. The linear IMFs are slower as compared to the non-linear IMFs and are with less randomization. There is a requirement for analyzing both linear and non-linear frameworks for modeling and predicting the Indian monsoon rainfall. Thus, in this study, both parts were analyzed, and the impact of combining both parts has also experimented.

Presently, machine learning techniques are gaining popularity in other contexts and their lack of application from a climate prediction perspective is a need of concern. The primary goal of inherent non-linear dynamics in meteorology is accurate predictions that need careful model selection, tuning, and validation. The present study is an attempt in this endeavour wherein the applications of machine learning techniques for Indian monsoon rainfall prediction are focused and based on the research outcome and learning experience, attempts were made to
improve it. In this study, we have used Artificial Neural Network (ANN), Extreme Learning Machine (ELM) and for the first time applied Regularized Online Sequential Random Vector Function Link (ROS-RVFL), EMD-DFA and deep Long Short-Term Memory (LSTM) neural network (EMD-LSTM) techniques for the rainfall prediction purpose and subsequently examined their relative performances. Seasonal prediction of rainfall is made a year in advance using the previous year’s data. It is found that the deep EMD-LSTM neural network model is producing more accurate prediction outcomes as compared to other techniques used in this study. Based on the computed Central Periods of IMFs, the possible association of the Indian monsoon with the influencing climate predictors has been suggested. The method of decomposing the monsoon rainfall into IMFs and later application of DFA and deep learning techniques is a new approach to Indian monsoon rainfall prediction. The proposed method is efficacious in predicting rainfall statistically as verification is done on an independent test dataset. This study is novel, unique, and expected to give a new direction for the Indian monsoon rainfall prediction.