PREDICTING FLOOD USING ARTIFICIAL NEURAL NETWORKS
ANASWARA ASOKAN¹, NAKULRAJ K.R²

Abstract

Flood is a natural disaster that cause loss in life and property. An accurate and reliable flood water level prediction can significantly improve public safety and mitigate economic damages caused by the flood. The objective is to build a model that predicts whether there is a chance for flood or not. A warning can be given to people living in different Places and can prevent disasters caused due to flood. A system is designed to predict future floods based on the previous precipitation data. The method used is Artificial Neural Network. The steps taken are the search and data collection of rainfall of different states in India that will be used as training data and test data. Artificial Neural Networks are a complex structure developed based on the activity of neurons in the human brain. Three layers in ANN are input, hidden and output layers. The prediction process is done by comparing the original or actual target data with the neural network output data. Error calculation is done, that gives the measurement of how the network can learn well so that the comparison of new patterns will be easily recognized. The error in the network output is the difference between the current output value and the desired output value. After error correction, predicted result is obtained from ANN.

1. Introduction

Flood disaster is one among the natural disasters that end in economic paralysis of 40 percent in a country. The accuracy of flood prediction with certain parameters is incredibly important as a future reference in predicting flood disasters. A flood is defined as a covering by water of land not normally covered by water. Differing kinds of floods has been observed over the years, river flood, urban flood, coastal flood etc. Flooding may occur as an overflow of water from water bodies, like a river, lake, or ocean, within which the water overtops or breaks levees, leading to a number of that water escaping its usual boundaries, or it should occur because of an accumulation of rainwater on saturated ground in an aerial flood. While the scale of a lake or other body of water will vary with seasonal changes in precipitation and snow melt, these changes in size are unlikely to be considered significant unless they flood property or drown stock. Flood mitigation system has been created to attenuate the aftereffects of floods. Flood forecasting is defined as a process of estimating and predicting the magnitude, timing and duration of flooding supported known characteristics of a river basin, with the aim to stop damages to human life, to properties, and to the environment. Flood forecasting is one element of flood mitigation system. If people are made aware of chances of a flood occurring in near time, the impact of a flood is reduced. A flood warning system could be a way of detecting threatening events beforehand. this permits the public to be warned so actions is taken to cut back the adverse effects of the event. As such, the first objective of a flood warning system is to cut back exposure to coastal flooding. Machine learning has become quite popular because it is solving diverse real-world problems. because the number of information increases, there’s a chance of applying these methods to solve various problems. as an example, during this problem to predict the probability of a flood, methods of ML is used. These methods are data-driven algorithms. It gives an ability to computers to be told from data without being explicitly programmed. These algorithms attempt to find a pattern in given data between input and output and whenever a brand-new data appears, it predicts the corresponding output. Performance of the algorithm must be considerably good to be used for solving an actual problem. for every method, a training is finished which fits the pattern between input and output.
Machine learning is often divided into categories as Supervised Learning, Unsupervised Learning, Reinforcement Learning, Regression, SVM (Support Vector Machines), ANN (Artificial Neural Network) are few methods of ML. Artificial neural networks are a family of models inspired by biological neural networks (the central nervous systems of animals, particularly, the brain). Like other ML methods, ANN is employed to estimate or approximate functions which will depend on an outsized number of inputs. ANN consists of a stack of neurons arranged in layers and might be viewed as a bipartite graph. Each neuron represents a function and every edge has a weight which tells the contribution of the neuron’s output towards the ultimate objective.

2. Related Work

In this section, different approaches of predicting flood are discussed on the basis of literatures by various authors. Wireless Sensor Networks for Flash-Flood Alerting [Castillo2004] discusses the utilization of wireless sensor networks (WSN) within the Andean region of Venezuela. WSN is used for monitoring the environment and tracking the disaster while it evolves. The study conducted on the matter of flood stage forecasting using support vector machines (SVM) in the region of Dhaka, Bangladesh [Liong2002]. In this paper, data from 8 water stations are taken into consideration and also the next water level has been predicted using SVM. Results obtained from both SVM (Support Vector Machine) and ANN (Artificial Neural Network) are compared.

In year 2013 proposed an ANN modeling for flood water level prediction for early warning system [Adnan2013]. The strategy uses BPNN with NN Inverse Model placed at the output for performance improvement. The Back-Propagation algorithm was applied on dataset obtained from the Department of Irrigation and Drainage Malaysia. The study conducted on the performance of prediction by evaluating differing types of ANN [Baltas2004]. The case study presented shows that ANN are often used for flood fore-casting due to the property of universal approximation, neural networks are used widely within the field of hydrology.

3. Dataset Description

The flood dataset is obtained from the KAGGLE data repository. Dataset contains monthly rainfall detail of 36 meteorological sub-divisions of India. Main feature taken is the rainfall level. Contain the rainfall details from the time period of 1901 to 2015. Monthly rainfall data is also included in mm unit. The following are the attribute fields in the dataset.

Subdivision: Includes different subdivisions in India. The 36 meteorological subdivisions are included.
Year: Containing year wise data from 1901 to 2015 of all 36 subdivisions.
Monthly Rainfall: Rainfall data of 12 months in corresponding years. Rainfall level of combined months are also included.
Annual Rainfall: Annual rainfall in every year from 1901 to 2015 is included.

4. Methodology

The method used is Artificial Neural Network (ANN). An artificial neuron network (ANN) could be a computational model based on the structure and functions of biological neural networks. Information that flows through the network affects the structure of the ANN because a neural network changes or learns, in an exceedingly sense supported that input and output. ANNs are considered nonlinear statistical data modeling tools where the complex relationships between inputs and outputs are modeled or patterns are found. The architecture contains the different processes distributed within the project. The important things distributed swollen prediction are the info collection, data preprocessing, Artificial Neural Network process, Error calculation and Result Prediction.

A. Data Collection

Data collection is a process of collecting information from all the relevant sources to find answers to the research problem, test the hypothesis and evaluate the outcomes. Data is collected from Kaggle data repository.

B. Data Preprocessing

Data cleaning and data visualization are the 2 processes carried out in data preprocessing. Data cleansing or data cleaning is that the process of detecting and correcting (or removing) corrupt or inaccurate
records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the info so replacing, modifying, or deleting the dirty or coarse data. Data cleansing is also performed interactively with data wrangling tools, or as execution through scripting. After cleansing, a data set should be in line with other similar data sets in the system. The inconsistencies detected or removed may be originally caused by user entry errors, by corruption in transmission or storage, or by different data dictionary definitions of comparable entities in several stores. Data cleaning differs from data validation in this validation almost invariably means data is rejected from the system at entry and is performed at the time of entry, rather than on batches of knowledge. Data visualization is that the graphical representation of knowledge and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

C. Artificial Neural Network Process

Artificial Neural Networks are a posh structure developed supported the activity of neurons within the human brain. Such systems “learn” to perform tasks by considering examples, generally without being programmed with task specific rules. The original goal of the ANN approach was to solve problems within the same way that somebody’s brain would. However, over time, attention moved to performing specific tasks, resulting in deviations from biology. ANNs are used on a spread of tasks, including computer vision, speech recognition, artificial intelligence, social network filtering, playing board and video games, diagnosis and even in activities that have traditionally been considered as reserved to humans, like painting. In general, Artificial Neural Networks have three layers, namely, input layer, hidden layer, and output layer.

1. **Input Layer**: All the inputs are fed in the model through this layer. Contain those artificial neurons which are to receive input from the outside world.

2. **Hidden Layers**: Hidden layers are mentioned hidden in between input layers and the output layers. Transform the input into something meaningful that the output layer or unit can use in some way. There can be more than one hidden layer which are used for processing the inputs received from the input layers. Each of the hidden layers is individually connected to the neurons in its input layer and also to its output layer leaving nothing to hang in the air. This makes it possible for a complete learning process and also learning occurs to the maximum when the weights inside the artificial neural network get updated after each iteration.

3. **Output Layer**: The data after processing is made available at the output layer. Contain units that respond to the information that is fed into the system and also whether it learned any task or not.
D. Error Calculation

Error calculation is a measurement of how the network can learn well so that the comparison of new patterns will be easily recognized. Error in the network output is the difference between the current output value and the desired output value. Commonly used equations for error calculations are:

1. **MSE (Mean Squared Error)**

Mean Squared Error (MSE) or Mean Squared Deviation (MSD) of an estimator measures the average of the squares of the errors that is, the average squared difference between the estimated values and the actual value. The equation to find MSE is,

\[
MSE = \frac{\sum_{t=1}^{N} (Y_t - T_t)^2}{N}
\]

2. **RMSE (Root Mean Squared Error)**

Root Mean Square Error (RMSE) is the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are. RMSE is a measure of how spread out these residuals are. The equation to find RMSE is,

\[
RMSE = \sqrt{\frac{\sum_{t=1}^{N} (Y_t - T_t)^2}{N}}
\]

3. **MAPE (Mean Absolute Percentage Error)**

The Mean Absolute Percentage Error (MAPE), also known as Mean Absolute Percentage Deviation (MAPD), is a measure of prediction accuracy of a forecasting method in statistics, for example in trend estimation, also used as a loss function for regression problems in machine learning. The equation to find MAPE is,

\[
MAPE = \frac{\sum_{t=1}^{N} |(Y_t - T_t)|}{N}
\]
E. Result Prediction

The final result is predicted. The result will be the percentage chance of occurring the flood. Predict whether the flood will occur in a subdivision or not.

5. Conclusion and Future Work

Floods may remain for few hours and might cause considerable damage and a possible threat to life. Flood prediction requires quantitative knowledge about infiltration and runoff dynamics, which is usually gained at the local scale. Among all disasters that occur in India, floods and drought are commonly occurs due to irregularity in monsoon. Highest death rate and damages are reported due to flood. So necessarily need some solution to give prior indication of flood. Artificial Neural Networks are used here to predict flood efficiently supported rainfall and water level status. Fault tolerance, adaptive learning, real-time operation and self-organization are the important characteristics of ANN other over models. The model uses more attribute values like annual rainfall and monthly rainfall data of 115 years. The system harnesses the power of latest technologies in predicting flood risks. Improve public safety and mitigate economic damages caused due to flood. Warning are often given to people living in several places. Predicts the arrival of flood as accurate as possible. System can reduce flood risk and permit emergency response personnel to be better prepared for and mitigate damages. In future, Artificial neural network is often wont to predict other natural disasters like landslide, earthquake, fire etc.

References