Indian Journal of Science and Technology, Vol 9(10), DOI: 10.17485/ijst/2016/v9i10/88976, March 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Storm Analysis with Raw Rainfall Dataset by using Artificial Neural Network and Min-Max Algorithms

Anjana Joseph[1] and M. Lakshmi[2]

[1]Department of Computer Science and Engineering, Sathyabama University,Chennai-600119, Tamil Nadu, India; 4anjanaz@gmail.com
[2]Faculty of Computing, Sathyabama University, Chennai-600119,Tamil Nadu, India; laks@icadsindia.com

Abstract

Objectives: The main objective of this paper is to propose a methodology to develop a Storm analysis model from Raw Rainfall dataset using techniques such as Artificial Neural Network and Min-Max Algorithm. Storm analysis model aims at predicting the occurrence and strength of a storm by analyzing the rainfall data of that region. **Methods:** In the proposed methodology the raw rainfall dataset is being trained by Artificial Neural network based on the three layers -Input, Hidden, and Output layers. The trained dataset is then summarized into a model which performs the prediction of storm centric characteristics. Neural network training is implemented in Hadoop framework. We obtain a considerable improvement in the total performance of the system by employing Artificial Neural Network. Min-Max algorithm is also used in the system for predicting the intensity of storm. The dataset used for training and prediction consists of daily rainfall data of Cherrapunjee area collected by The Meteorological Department of India. **Findings:** In the existing system, the raw rainfall dataset is collected and stored in a relational database and then map-reduce based techniques are applied for storm analysis. The major disadvantages associated with this technique are the performance and accuracy rate get reduced with increase in data size. In the proposed methodology as the raw rainfall dataset is being trained by Artificial Neural network the performance and accuracy rate got improved. Also, the training process is done on multi-node hadoop cluster by considering large raw rainfall dataset. With multi-node hadoop cluster there was a large reduction in the total training time. Storm depth of a particular region is calculated by applying MIN-MAX algorithm. This improved the total efficiency of the storm intensity prediction. **Applications/Improvement:** The performance of the system can be further improved by reducing the training time by adding more nodes while implementing the process in multi node hadoop cluster. Also higher prediction accuracy can be obtained by combining various suitable fuzzy inference models[5] with the proposed neural network mode.

Keywords: Artificial Neural Network, Hadoop, MIN-MAX Algorithm, Rainfall Data, Storm Analysis

1. Introduction

Raw rainfall data is capable of determining and predicting the characteristics of a storm accompanied with it. Big raw rainfall numerical dataset is collected and stored in a text format. It is difficult to analyze and mine a dataset stored in a text format due to several reasons. First, the amount of data is huge. Second, it is difficult to identify storm from text formatted dataset. A solution to the above problem is to encapsulate raw rainfall data into a storm centric model which simplifies storm analysis and prediction. Storm centric model consists of only those data which are relevant for storm analysis. Thus the total data size will get reduced. The total volume of storm data is less than 1% of the total data.

Storm analysis and prediction is a detection problem which helps to predict whether a storm will occur or not based on the Raw Rainfall dataset. Initially raw rainfall dataset is preprocessed and trained to obtain storm data. The obtained Storm Result in text format is again preprocessed and trained. The trained dataset is used for storm analysis and prediction. The system will produce an error rate if there is no storm otherwise it will generate an alert for the storm with location and time. An Apache Hadoop framework is used for implementing Storm Analysis and Prediction problem. Apache Hadoop framework is open

*Author for correspondence

source software used for storing and processing of large scale dataset. Min-Max algorithm is also used for improving the efficiency of storm prediction.

In the existing works Various Storm identification algorithms[1] were proposed for extracting storm centric data from rainfall dataset. Algorithm classified identified storm data into three concepts: 1) Local storm 2) Hourly storm and 3) Overall storm. Identification algorithm is based on recursion and depth first search which are time consuming and will also leads to multiple retrieval of same data. In order to overcome these problems ,map reduce based identification algorithms were introduced namely Local storm identification algorithm[2], for identifying local storm out of rainfall data, Hourly storm identification algorithm[2], for identifying Hourly storm out of rainfall data, Overall storm identification algorithm[3], for identifying overall storm out of Hourly storm data. With the introduction of map reduce based techniques the total performance of the system got greatly improved. But in these techniques there is no preprocessing of dataset .Also Dataset in text format is directly applied as input.

Storm Identification and Forecasting were also done using High-Resolution Images of Short-Range X-Band Radar[4]. In this work initially they selected a threshold for storm identification, then isolated loosely-connected storm and finally identified sub storm within large storm. Two automatic methods which are based on the histogram of the image were tested for storm identification. They proposed a new technique called SALdEdA for storm tracking. SALdEdA stands for structure-amplitude-location-eccentricity-circularity. Also, constructed two simple tables for ensuring the accuracy of prediction.

Hybrid approaches were also used in detection problems. Thunderstorms and lightning Detection System using Hybrid Approach used satellite images for predicting whether the cloud image produces thunderstorms or not. Clustering and wavelet transform techniques were also used here. Under this topic they have demonstrated comparison between different models used in then detection of thunderstorms such as STP model, MOM model, CG model, LM model, QKP model and DBD model. Here they considered the four basic performance measures i.e. sensitivity, specificity, accuracy and precision for comparison purpose. The proposed method produced an average accuracy of 89.23% in thunderstorm prediction. Storm intensity prediction can also be done using Artificial Neural Network. In an existing work[5] they implemented a storm intensity prediction model for classifying 3 major type of storm that is Hurricane, Tropical cyclones, Tropical depression with artificial neural network.

2. Materials and Methods

2.1 Artificial Neural Network

Artificial Neural Networks are processing models inspired from animal central nervous system. It can be explained as a network of interconnected processing elements known as neurons. ANN is widely used in developing prediction models as it is highly nonlinear, flexible and possess good ability in determining complex relationships between variables. In Storm prediction model, ANN is used for establishing a simple relationship between input (rainfall dataset) which is the cause and output (storm) which is the effect.

In an ANN the trained dataset is obtained by adjusting the connections between the elements (the weights). In theory, this adjustment process can be viewed as a form of learning. The particular type of ANN learning used in this study is a supervised one, where an output vector (target) is specified, and the ANN is trained to minimize the error between the output and input vectors, thus resulting in an optimal solution.

Figure 1 represents the proposed methodology and ANN architecture for storm analysis over rainfall data. ANN architecture consists of three layers–input layer, hidden layer and output layer. Input data (Rainfall data) is supplied to the input layer. The input layer firstly weighted the data with random values and produces it to the hidden layer. Hidden layer train the input data using supervised

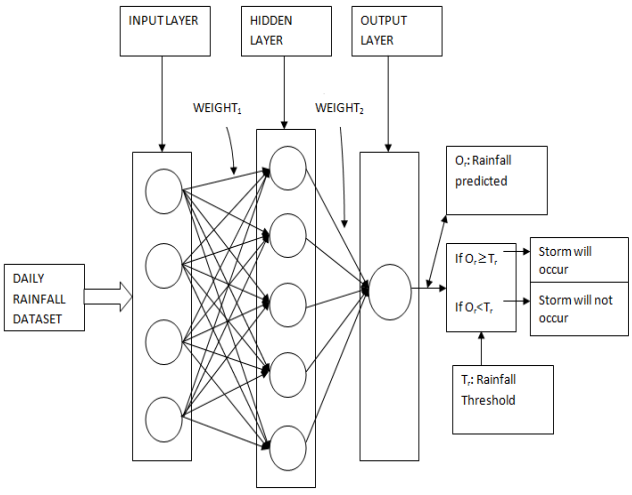

Figure 1. Proposed Methodology.

training method. An average rainfall percentage per day is also computed .This is set as the threshold value. The hidden layer output could get feed directly to output layer and produced as output. The produced output is then compared with the predetermined threshold value in order to examine whether a storm will occur or not. Min max algorithm is also employed in order to find out the minimum and maximum range of storm intensity values.

2.2 Hadoop Framework

The neural network implementation is done in Hadoop framework. Hadoop can be viewed as open source software used for the storage and processing of data large scale. HDFS and map reduce are the core components of Hadoop. HDFS stands for Hadoop Distributed File System and performs data storage while map reduce performs processing of data. Both these are derived from Google's Map Reduce and Google File System (GFS) papers.

There are two parallelization approaches used for training data, 1) data parallelization 2) node

Parallelization. In this work data parallelization approach is followed in which the whole training data is divided into many small blocks of dataset and each node process some blocks. After all the nodes finish its training, results are collected and applied for batch update. In batch mode updating will takes place only after getting all the trained data.

2.3 Proposed Architecture

The proposed work is divided into three phases:

- Preparation of the storm dataset and Preprocessing.
- Normalization of the dataset by ANN Training.
- Storm analysis and detection.

Figure 2 represents the flow of work through different phases. The three different phases are explained below.

2.3.1 Preparation of the Storm Dataset and Preprocessing

This phase explains about the collecting and preprocessing of rainfall dataset. Raw daily rainfall dataset with location and date is collected. These dataset will be in numerical format. It is converted into binary values for reducing the complexity of processing data and to improve the total training time.

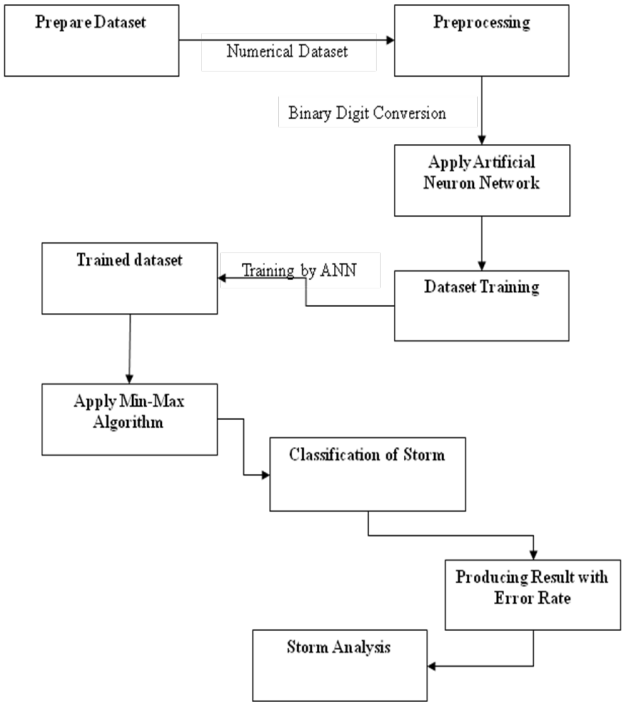

Figure 2. Flow Diagram.

2.3.2 Normalization of the Dataset by ANN Training

This is the phase in which both the normalization and training of dataset is performed. In ANN and other data mining approaches we need to normalize the inputs, otherwise the network will be ill conditioned. In essence, normalization is done to have the same range of values for each input to the ANN model. This can guarantee stable convergence of weights. In the training process the correct output for each input record is known and the output nodes are assigned with these correct values. The network's calculated values for the output nodes is compared to these "correct" values, and calculate an error term for each node .These error terms are then used to adjust the weights in the hidden layers so that during next iteration the output values will be closer to the "correct" values.

2.3.3 Storm Analysis and Detection

Under this phase the storm analysis is done. Storm analysis means, in order to obtain the storm centric characteristic such as intensity a min-max algorithm is applied over the trained dataset. This is to improve the total efficiency and correctness of the storm prediction. Finally a storm classification data is produced in text format saying whether

a storm will occur or not. Also an alert is generated for storm with location and date.

3. Results and Discussion

3.1 Data Source

The dataset considered for training is from Cherrapunjee[6,7] areas. The dataset consists of daily rainfall data of Cherrapunjee area collected by The Meteorological Department of India. In the dataset each row represents the precipitation value of that region in a particular date. Dataset available from 01-01-2009 to 07-08-2015.Dataset is represented in tabular format with 2424 rows and 3 columns namely DATE, DAILY RAINFALL and ANNUAL RAINFALL.

3.2 Experimental Result

Here the ANN training produced a trained dataset which consists of only those data relevant for storm prediction. Along with that an additional value named annual rainfall percentage per day is computed for each date and added as a new column to the table. This will act as the threshold value. If the predicted rainfall value is greater than the threshold value then storm will occur otherwise not.

The training process is done on multi-node hadoop cluster by considering large raw rainfall dataset. With multi-node hadoop cluster there was a large reduction in the total training time. Storm depth of a particular region is calculated by applying MIN-MAX algorithm. Minimum and maximum ranges of storm strength are also computed with respect to the annual threshold percentage. This improved the total efficiency of the storm intensity prediction.

Figure 3 is presenting the monthly storm analysis with untrained dataset (Existing Technique), trained dataset (Min-Max technique) and the exact data. It shows that min-max is providing an exact prediction over the strength of storm in compare to existing technique.

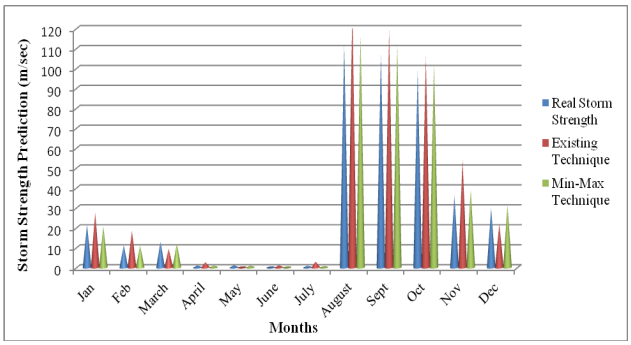

Figure 3. Monthly Storm Analysis

4. Conclusion

In this work, we are proposing enhanced Artificial Neural Network and Min-Max algorithm for making an error-less prediction of the storm. These proposed algorithms are producing a low error alert for storm, where Artificial Neural Network is performing the task of normalization for big raw rainfall dataset. The well trained dataset is producing a result over the storm within location and time. In other hand, the Enhanced Min-Max algorithm is performing the task of enhancement of the result. Proposed system is producing a combined prediction of the storm with location and time. And also achieved higher accuracy for the prediction of storm based on the big raw rainfall dataset.

The performance of the system can be further improved by reducing the training time by adding more nodes while implementing the process in multi node hadoop cluster. Also higher prediction accuracy can be obtained by combining various suitable fuzzy inference models[5] with the proposed neural network model.

5. References

1. Jitkajornwanich K, Elmasri R, Li C, McEnery J. Extracting Storm-Centric Characteristics from Raw Rainfall Data for Storm Analysis and Mining. Proceedings of the 1st ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data (ACM SIGSPATIAL BIGSPATIAL'12); 2012. p. 91–99.
2. Jitkajornwanich K, Gupta U, Elmasri R, Fegaras L, McEnery J.Using Map Reduce to Speed Up Storm Identification from Big Raw Rainfall Data. Proceedings of the 4th International Conference on Cloud Computing, GRIDs, and Virtualization (CLOUD COMPUTING'13); 2013. p. 49–55.
3. Jitkajornwanich K, Gupta U, Elmasri R, Fegaras L, McEnery J. Complete Storm Identification Algorithms from Big Raw Rainfall Data Using Map Reduce Framework. Proceedings of the 2013 IEEE International Conference on BigData (BigData 2013); Silicon Valley: CA. 2013. p. 13–20.
4. Shah S, Notarpietro R, Branca M. Storm Identification, Tracking and Forecasting Using High-Resolution Images

of Short-Range X-Band Radar. Proceedings of Atmosphere; 2015; 6(5): 579-606.
5. Pruthvi R. 'Web-Users' Browsing Behavior Prediction by Implementing Neural Network in Map Reduce. International Journal of Advance Foundation and Research in Computer (IJAFRC); 2014; 1(5):60–69.
6. Daily Weather Data 2009-2013.Available at: http://www.cherrapunjee.com/weather-info/daily-weather-data-2009-2013/. Date accessed: 31/07/2015.
7. Daily Weather Data 2014.Available at: http://www.cherrapunjee.com/weather-info/daily-weather-data-2014/. Date accessed: 31/07/2015.