International Journal on Applications in Information and communication Engineering
Volume 7 : Issue 4 : November 2021, pp 5 – 8 www.aetsjournal.com ISSN (Online) : 2394 – 6237

SPATIAL HYBRID MODEL FOR RAINFALL PREDICTION

PRABAKARAN. S , RAMU . M , PALANISAMY . M

Abstract— Rainfall forecasting is an essential aspect of weather forecasting. Rainfall anticipation is a main factor for some divisions in this area, which has a few seasons per year. Rainfall data is a phase arrangement of data that vary over time as the weather and season shift. Furthermore, since rainfall is determined by a range of variables such as wind speed, temperature, and so on, producing a 100 percent reliable forecast is exceedingly difficult. This paper shows how a robust rainfall expectation approach based on Linear Regression and SVMcalculations are used to forecast rain shower. The gathered crude information wasn't set up for utilizing as contribution of calculation, hence it had been pre prepared physically to suit into the calculation, at that point nourished to the calculation. In terms of precision and process running time, the anticipated approach outperforms the ordinary systems, according to the evaluation results of the investigation based on the evidence. The proposed method produced the maximum reasonable expectation of 99.92 percent.

I. INTRODUCTION

Predicting rain is difficult, particularly when environmental conditions are changing as a result of global warming. It is a complex multiplex phenomenon that necessitates knowledge across multiple domains. As a result of this imbalance, dangerous problems emerge in a number of industries, with agriculture suffering the most. Some related issues have already been listed. Precipitation forecasting is a crucial feature of climate forecasting. Predicting accurate and timely precipitation is critical for a variety of activities, including water asset planning and maintenance, surge warnings, building drills, and flight tasks. The aim of this project is to predict rain a few weeks in advance and to describes the key factors that affect rain. With the development of technology and research,several evolving techniques, such as Support vector Machine and Artificial Neural Network, are commonly used in rainfall prediction.

A non-parametric model that eliminates empirical risk is an artificial neural network (ANN). Without any previous data or input space assumptions, this model will approximate any non-linear function to a certain degree and to any arbitrary precision. Also, without a clear knowledge of natural processes, a neural network's capacity to model the properties of rainfall prediction is remarkable.

Prabakaran. S , Faculty of Information Technology, Dhanalakshmi Srinivasan Engineering College, Tamil Nadu, India

Ramu . M , Faculty of Information Technology, Dhanalakshmi Srinivasan Engineering College, Tamil Nadu, India

Palanisamy . M , Faculty of Information Technology, Dhanalakshmi Srinivasan Engineering College, Tamil Nadu, India

II. LITERATURESURVEY

These were some of the papers taken as a reference to build our system.

1) Jorge Mendez Astudillo and Lawrence Lau present the analysis of Zenith Troposphere Delay Calculations in Specific Point Positioning (PPP) Services and PPP Program Packages. Global Navigation Satellite System (GNSS) signals undergo a troposphere delay while travelling through the troposphere due to a drop in the medium refractive index. The PPP technique can achieve precise placement of centimeters/millimeter with just one GNSS receiver. Zenith Tropospheric Delay (ZTD) is measured using unknown PPP location. Estimated ZTD can be especially helpful in meteorological applications, such as calculating amount of ambient water vapor. Using various algorithms and models, PPP is integrated into web repositories and product packages. This paper provides an evaluation of findings that includes review of ZTD from three PPP online platforms and three commodity packages. This paper's key contribution is to illustrate ZTD approximation precision in PPP. The thesis explains the dependence and effect of the processing algorithm on GNSS users and researchers. Observation results from mostly rainfall zones were included in this study based on the PPP ZTD calculation. PPP ZTD

measurements are related to ZTD calculated using the same-day tropospheric IGS product. Two of the three online PPP providers expect strong alignment (1 cm) with IGS ZTD values in northern stations. The results gives that online PPP platforms at all stations outperform selected PPP app bundles.

2) As suggested by Shilpa Manandhar, GPS Precipitable Water Vapor (PWV) PWV is a reasonable source for monitoring precipitation. It is defined as absolute water vapour in the atmosphere. Past radiosondes and microwave radiometers were used to measure PWV. In the other hand, they have low temporal resolution and high operating costs. GPS pulse delay is increasingly used in these applications. This paper's main purpose is to investigate the association between GPS-derived PWV and precipitation. We provide an analysis demonstrating that PWV increases until a rainfall happens and then decreases. We remove a threshold PWV that detects rainfall before PWV crosses the threshold. PWV and rainfall data from June 2010 and 2011 validated the model.

3) Storms Prediction: Anne Ruiz-Gazen and Nathalie Villa proposed Logistic Regression Vs Random Forest for Unbalanced Data to contrast two supervised classification methods on a critical meteorological subject. Data comes from satellite cloud formation observations known as convective or non-convective. Thunderstorm monitoring and warning are strongly reliant on detecting convective cloud patterns, contributing to lightning. As the condition is too unbalanced, we embrace many performance parameters and solutions. In a data mining advanced course, this case study will be used to demonstrate how to use logistic regression and random forest on real-world data selection with unbalanced groups.

III. EXISTING SYSTEM

The data set is stored and processed in structured database like MySQL, Oracle etc. The time consumed to process volumes of data to pre-process and classify were huge. The model is utilized to generate the prediction of future rainfall based on the dataset. MySQL is a flexible computing framework suitable for small applications. Therefore, MySQL was used as simple computing and storage platform.

1) Disadvantages:

For such a large dataset, the machine learning process was relatively slow.

For a given task, more memory use is needed

IV. PROPOSED SYSTEM

To overcome the fallback in the existing system we propose a machine learning based system to increase the efficiency and accuracy. To handle voluminous data,weare using Hadoop to store and retrieve data from the distributed hadoop file system (hdfs). Hadoop allows the user to load data into a cluster. Random forest algorithm is to be implemented for forecastingrainfall.

1) ADVANTAGES:

Random Forest has the advantage of being able to execute both regression and classification functions, and the relative value it gives to input features is readily apparent. Random Forest is well-known as a useful prediction result. The amount of hyper parameters isn't overwhelming, and they're straightforward to understand. A Random forest qualified model is used to predict rainfall for a given month. The projection is for a few months' time.

V. ARCHITECTURE

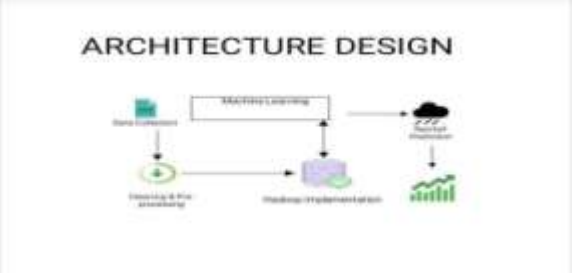

Fig 1. System Architecture

1) List of Modules

- Data collection andpre-processing
- Random Forest model generation

2)DATA COLLECTION AND PRE-PROCESSING

Weather dataset is collected in a comma separated values (CSV) file for the last three or four years of

rainfall information. The dataset contains the month wise aggregation. The datasetincludes null values, negative values, or an error. During pre-processing, the dataset is washed. The pre-processing methods include the elimination of incomplete documents. When we have a clean dataset, we must prepare it for use by the machine learning algorithm.

3)FOREST MODEL GENERATION AT RANDOM

Random forestry, which is also known as random decision-making woods, is a method of learning for the grouping, regression, and other activities by training a large number of decision-making trees, and then generating the class that reflects the average predictor of the individual trees. Random forest generates and merges several decision trees to obtain a forecast that is more accurate and reliable.Random forest has the advantage of being able to overcome classification and regression problems, which are the most common types of machine learning problems. To train our system and create a model for future prediction, we use a dataset.

Fig 2 Rainfall prediction in December

The graph portrays the rapid increase in temperature over the last few decades. According to NASA (National Aeronautics and Space Administration), which monitors any change in environment, the climate will deteriorate if these factors are not addressed.

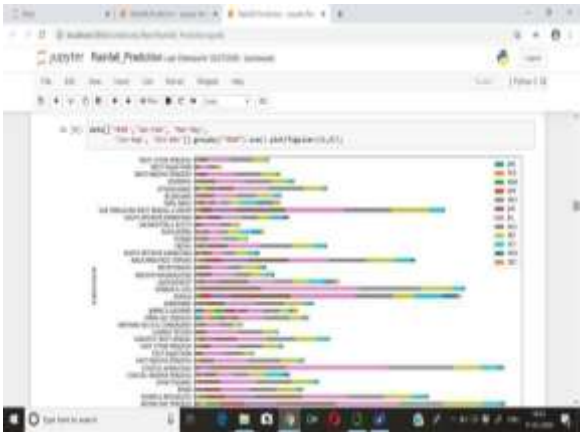

Fig 3. Monthly Rainfall Prediction in Different states in India

VI. ALGORITHM SPECIFICATION

Random forest is a supervised learning method that generates a classification tree. This algorithm uses an input data vector to identify a new object from input feature vectors in each tree of the forest. Each tree in the forest gives each other "votes," and the trees with the most "votes" are ranked. Pseudocode is a form of code that isn't really code. From a total of "m" features, choose "k" features at random. Break the product into several components using the strongest split.

Using data from the weather station the RF model's ability to now cast rain is investigated (RPG-HATPRO). Seven frequencies in the 22–31.4 GHz band and seven frequencies in the 51.26–58 GHz band are available on the radiometer. The raw data is saved on a disc in binary format, then translated to ASCII using radiometer tools. Because of their exposure to water vapour and temperature variations in the environment, these frequency bands were selected. The 22–31,4 GHz band is associated with the weak water vapour belt extended by friction, which can be used to obtain the water vapour profile.

At wavelengths, the atmosphere becomes gone down from the oxygen absorption axis, which allows researchers to calculate brightness temperatures in various atmospheric layers. As the ratio of mixing and temperature addiction for oxygen absorption have been established, the well-defined weighting feature of the 51.26–58 GHz band can be used to attain the air temperature profile. This radiometer is adjusted using three

ways. In contrast to absolute calibration once a year, calibrating the calibration curve between antenna temperature and receiver voltages is carried out whenever a test is performed. Noise injection calibration is accomplished.

VII. CONCLUSION

I have done a weather forecasting model that makes forecasts based on the combined impact of main weather variables. We created a data-centric kernel and showed how GPR would efficiently interpolate across space when accounting for weather phenomena such as turbulence. We used short- and long-term features to conduct temporal analysis inside a gradient-tree dependent learner. To model the relationships between weather variables, we applied a deep belief network to the system and modified the parameters. A collection of experiments utilising real-world evidence suggest that the most current solution outperforms NOAA benchmarks, as well as recent analysis that shows changes over the benchmarks.

REFERENCES

[1] Xiong, Lihua, and Kieran M. OConnor. "An empirical method to improve the predictionlimitsoftheGLUEmethodologyinrainfallrunoffmodeling ."Journalof Hydrology 349.1-2(2008):115-124.
[2] Schmitz, G. H., and J.Cullmann. "PAI-OFF: A new proposal for online flood forecasting in flash flood prone catchments." Journal of hydrology 360.1-4 (2008): 1-14.
[3] Riordan,Denis,andBjarneK.Hansen."Afuzzycase-basedsystemforweather prediction." Engineering Intelligent Systems for Electrical Engineering and Communications10.3 (2002): 139-146.
[4] Guhathakurta,P."Long-rangemonsoonrainfallpredictionof2005forthedistricts and sub-division Kerala with artificial neural network." Current Science 90.6 (2006): 773-779.
[5] Pilgrim, D. H., T. G. Chapman, and D. G. Doran. Problems of rainfall-runoff modelling in arid and semiarid regions." Hydrological Sciences Journal 33.4(1988): 379-400.*9th* International Conference on Cloud Computing, Data Science &Engineering(Confluence)*395*
[6] Lee, Sunyoung, Sungzoon Cho, and Patrick M. Wong. "Rainfall prediction using artificial neural networks." Journal of geographic information and Decision Analysis 2.2 (1998):233-242.
[7] French, Mark N., Witold F. Krajewski, and Robert R. Cuykendall. "Rainfall forecastinginspaceandtimeusinganeuralnetwork."Journalofhydrology137.1-4 (1992):1-31.
[8] Charaniya,NizarAli,andSanjayV.Dudul."Committeeofartificialneuralnetworks for monthly rainfall prediction using wavelet transform."International Conference on. IEEE,2011.