Farmer Buddy-Weather Prediction and Crop Suggestion using Artificial Neural Network on Map-Reduce Framework

Manali Joshi
Sr.no 9/1/1 Vadgoan,
Narhe road, Pune

Saminabano Shaikh
A/P-Tisgaon Tal- pathardi,
Ahmednagar

Prachi Waghmode
Tushar park,Khadakwasla,
Pune

Padma Mali
Lahan maliwada, shivaji
nagar, Nandurbar.

ABSTRACT
Weather condition analysis and its prediction has always been a task of great efforts. It is tedious to predict the weather accurately. India being a country with agriculture as the primary occupation, there is an extreme need to predict data with improved accuracy. Big data comes with solution since it becomes easy and comparatively less expensive to store tremendous amount of data. This paper proposes an approach of using Hadoop for processing such Big volumes, variety and velocity of weather data. It includes application of Artificial Neural Network which is a convenient approach. ANN is implemented on Map-Reduce framework for short term rainfall prediction. Additionally, method will detect soil and Regional analysis which can also detect crop depending on user's location. Crop data along with feasibility check of soil condition is provided so as to help farmer/users in decisive situation.

Keywords
Data mining, Crop Analysis, Artificial Neural Network, Map-Reduce Framework, Rainfall Forecasting

1. INTRODUCTION
As far as India is concerned, rainfall forecasting is one of the most important weather prediction task because India's economy is highly dependent on its agriculture product. Weather prediction is a complicated and tedious task. Previously used methods such as deterministic and statistical are time consuming and has low performance. Various statistical methods for weather predictions are used for which data is collected over a large time period for statistical analysis. Relationship between various deciding factors is established by using the future conditions. Machine Learning is the subject of computer science that gives computers the ability to learn without being explicitly programmed. Machine learning explores the study and construction of algorithms that can learn from and make prediction on data, such algorithms overcome following strictly static program instructions making data driven prediction or decision. Weather data since being huge and have the great velocity machine learning approach is most compatible. Since, data analysis will be the major task in weather forecasting within the field of data analysis, Machine learning is method used to device complex models and algorithms that lend themselves to predictions.

Machine Learning aims a system that learns from its experiences. Weather is truly complex in nature and designing as deterministic algorithm gives the process results. Paper thus tries to integrate the two concepts in order to improve the efficiency of the work. Firstly treating weather data as huge data or big data, processing it by using Hadoop framework. Secondly using ANN as a learning approach and also for prediction of the result. ANN algorithm tends to build a system that learns on its own from that training data and improves its efficiency of predictions by using trial and error approaches. Hadoop is used for processing the huge weather data in a distributed approach. This framework is to build for processing the huge amount of data, so as to make its analytical decision leading to better analytical decision. All the modules in Hadoop are designed with a fundamental assumption that hardware failures are common occurrences and should be automatically handled by the framework. Hadoop consists of a storage part, known as Hadoop Distributed File system (HDFS) and a processing part called Map-Reduce. The Hadoop framework itself is mostly written in the java programming language, with some native code in C and command line utilities written as shell scripts. Hadoop thus allows the detect to be processed faster and more efficiently than it would be in a more conventional supercomputer architecture that relies on a parallel file system, where computation and data are distributed via high speed.

2. RELATED WORK
1.Weather prediction is a major problem. This paper makes use of classification and clustering. techniques. Classification/regression models were used to obtain the crop yield prediction result. Linear regressions, K-NN, Neural network are the models used for purpose of prediction. Clustering algorithms named k-means is used to group the districts into distinct clusters such that districts belonging to similar values of relevant attributes should be in same cluster.

2.Parallelizing the machine learning algorithms has become a crucial task in order to speed up the application relevant to machine learning. This paper shows that the algorithms that are suitable to the statically query model can be written in a certain "Summation form" so they can be easily parallelized on multicore computers. It introduces the use of Google's map-reduce techniques on a variety of learning algorithms. It shows the implementation of Navies Bayes, K-means, Neural Network, and Logistic Regression on the multicore map-reduce framework.

3.Paper utilizes clustering algorithm for obtaining clustered regions according to the climate. Five major types of classifier are used in the experiment such as Linear regression, Non-

Linear regression, Regression tree, Ensemble learning and Artificial Neural Network. The errors obtained by using these methods are compared by means of root mean square error (RMSE). The error between the expected output and the actual output is squared. It is then averaged and square root is found. This is the procedure is applied for 3 types of error i.e. training error, testing error, and cross validation error

3. METHODOLOGY

Let S be the system defined as:

S= {I/P, O/P, SS, SU, FS, A, DD, NDD}

S= System.

I/P= Username, Password.

O/P= Display decision

 (weather prediction, crop suggestion, and crop Information.)

SS= System State.

Login State, Weather data fetching state, data processing state, Prediction state, Result state.}

DD=Deterministic Data.

User data (user_name, user_id, user_phn, user_tal, user_dist).}

NDD=Non Deterministic Data.

  {Weather data attributes (humidity, pH, rainfall amount, MAX temp, MIN temp.).}

A=Set of Functions. = {F1, F2, F3}

F1=Preprocessing

  { Map-Reduce framework}

F2= Applying ANN algorithm

  {Multi-layer Perception: (linear activation function, back propagation learning techniques)}

F3= Display Result.

SU= Successful State.

  {Weather prediction will be done accurately based on user's locations, along with crop suggestion and information.}

FS= Failure State.

 {System fails to predict the weather condition and crop suggestion.}

4. NEURAL NETWORK ON MAP-REDUCE

In map-reduce framework, mappers will do parallel processing of data using a function defined in it and reducers will summarize the result. It's a kind of embedding one neural network on each map node. Feed forwarding the training samples and back propagation of the corresponding error will be done by mappers. Weights will not update in this phase. Instead, mappers pass these weight updates to reducer. Arc identical like (Ii and i) will from the intermediate keys and the weight update Δw will be passed as value. Thus a batch update of the weight parameters will be done with this averaged Δw values. This map-reduce based NN was implemented and pre-processed climates data was provided as input. It has been used for forecasting the rain on the next

days, using rain and temperature values of a span of seven days.

5. WEATHER PREDICTION ALGORITHM

5.1 ANN

The idea of ANN is based on the belief that working of human brain by making the right connections can be imitated using silicon and wires as living neurons and dendrites. The human brain is composed of 100 billion nerve cell called as **neurons.**(see Figure 1)
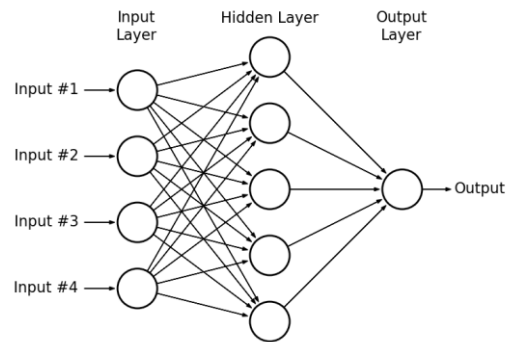

Fig 1: Artificial neural network.

5.1.1 Back Propagation algorithm

It is the training data set or learning algorithm. It learns by examples. If you submit to the algorithm the example of what you want the network to do, it changes the network weights so that it can produce desired output for a particular input on finishing the training. Back propagation networks are idea for simple Patterns Recognition and Mapping tasks.

5.2 Multi Layer Perception

A multi-layer perception (MLP) is a feed forward ANN model that maps sets of input data onto a set of appropriate outputs. A MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Expect for the input nodes, each node is a neuron with a nonlinear activation function. MLP utilizes a supervised learning technique called back propagation for training the network. MLP is a modification of the standard linear perception and can distinguish data that are not linearly separable.

5.2.1 Activation function

If a multi-layer perception has a linear activation function in all neurons, that is a linear function that maps the weighted inputs to the outputs of each neuron, then it easily proved with linear algebra that any number of layers can be reduced to the standard two-layer-input-output-model.

$$y(vi) = tanh(vi) \wedge y(vi)$$
$$= (1 + e) - (vi) - 1$$

5.2.2 Layers

The multi-layer perception consists of three or more layers of nonlinear activating nodes and is thus considered a deep neural network.

5.2.3 Learning through back propagation

Learning occurs in the perception by changing connection weights after each piece of data is processed, based on the amount of error in the output compared to the expected result.

We represent the error in outputs node j in the nth data

points by $ej(n) = dj(n) - yj(n)$, where d is the target value and y is the value produced by the perception. We then corrections to the weight of the nodes based on those corrections which minimize the error in the entire output,

$$\varepsilon(n) = \frac{1}{2} \Sigma x_{ej\,2(n)}$$

The derivation to be calculated depends on the induced local field vj which it varies. It is easy to prove that for an output node this derivation can be simplified to $\frac{\partial \varepsilon(n)}{\partial vj\,(n)} = ej(n)\phi\big(vj(n)\big)$

Where ϕ is the derivation of the activation function.

6. PROPOSED SYSTEM ARCHITECTURE

1. 1.Weather prediction: ANN is used to predict weather condition for following seven days of weather data.

2. Map-Reduce Framework: For performing analysis on the data generated, Map-reduce framework is used with ANN algorithms for distributed and faster processing.

3. Soil and Regional Analysis for Decisive Crop Planting: Depending on user's location, crop data along with feasibility check of soil condition is provided so as to help user/farmer in decisive situation.(see Figure 2)

4. Fetch Weather Data: Weather data for user's location is stored into database so that it can be used for processing and prediction. Data is fetched from forecast.io using its developer API. A generated unique key is used for this.

5. Parse Weather data: Data is received into form of a JSON which must be parsed for relevant data. JSON is lightweight data-interchange format. It is easy for human to read and to write. It is easy for machines to parse and generate.

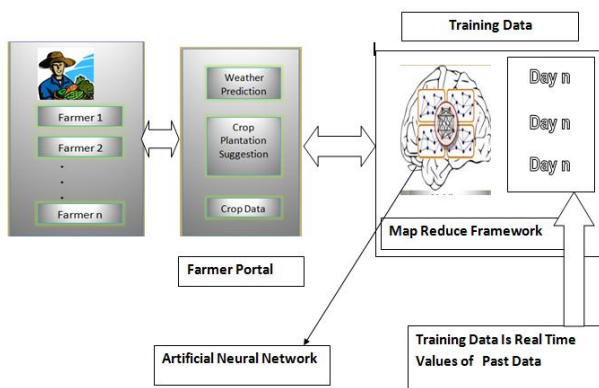Weather prediction system architecture is as follows:

Fig 2: Weather Prediction Architecture.

7. CONCLUSION

Weather data is highly dynamic and chaotic in nature and has large size of data. This paper is an attempt to see how big data solution can be utilized in the field of weather prediction. Artificial Neural Network (ANN) on Map-Reduce framework is implemented for rainfall prediction and crop suggestion. Here ANN is used for prediction next seven day's rainfall. In our paper we propose the use of back propagation technique which is an error correction technique which leads to high accuracy in the predicted results. An attempt is made to suggest best suitable crop and its information based on farmer's location and weather condition to increase the yield of crops. Implementing this solution on Hadoop makes it faster and scalable.

8. ACKNOWLEDGEMENT

9. REFERENCES

[1] A.T.M Shaikh Ahamed, Navid Tanzeem Mahmod, Rasedur M Rahman, "Applying Data Mining Techniques to predict annual yield of major crop and recommended planting different crops in different Districts in Bangladesh. ", 2015.

[2] Cheng-Tao chu, Sang kyun kim, Yi-An Lin, YaunYaun Yu, Gary Bradski, Andrew Y.Ng, "Map-Reduce for machine learning on multi core",2014.

[3] Mohammad Motiur Rahman, Naheeena Haq, Rasedur M Rahman, "Machine Learning facilitated Rice Prediction in Bangladesh. ",2014

[4] Namitha K,Jayapriya A,G Santhosh Kumar,"Rainfall Prediction using Artificial Neural Network on Map-Reduce Framework",2015 IEEE.ISBN 978-1-4503-3361-0/15/08.

[5] Max A. Little , Patrick E. McSharry and James W. Taylor Generalised Linear Models for Site Specific Density Forecasting of UK Daily Rainfall journals.ametsoc.org, Monthly Weather Review Volume 137, Issue3 (March 2009).

[6] D Ramesh , B Vishnu Vardhan. Data Mining Techniques and Appli-cations to Agricultural Yield Data. International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue9, September 2013,pp.3477-3480.

[7] P. Guhathakurta "Long-range monsoon rainfallprediction of 2005 for the districts and sub-divisionKerala with artificial neural network"CURRENTSCIENCE, VOL. 90, NO. 6, 25 MARCH 2006.