Implementation of Improved Crop Yield Prediction using Neural Network

Mrs. Prajakta Prashant Bhangale[1] Prof. Yogesh S. Patil[2] Prof. Dinesh D. Patil[3]
[1]ME Student [2]Assistant Professor [3]Head of Dept.
[1,2,3]Department of Computer Science & Engineering
[1,2,3]SSGBCOET, Bhusawal, Maharashtra India

Abstract— *An Agricultural sector is in need for well-organized system to predict and improve the crop over the world. The complexity of predicting the best crops is high due to unavailability of proper knowledge discovery in crop knowledgebase which affects the quality of prediction. In data mining, classification of data is a crucial step in mining useful information. The classification techniques such as neural network, SVM, Naïve Bayes, KNN which make this task complicated due to problem like random selection of initial parameters. Here we propose crop prediction system using neural network and compare the results of the proposed system with various classifiers. Artificial neural networks have been demonstrated to be powerful tools for modelling and prediction, to increase their effectiveness. Crop prediction methodology is used to predict the suitable crop by sensing various parameter of soil and also parameter related to atmosphere. Parameters like type of soil, PH, nitrogen, phosphate, potassium, organic carbon, calcium, magnesium, sulphur, manganese, copper, iron, depth, temperature, rainfall, humidity. For that purpose we are used artificial neural network (ANN). Agricultural researchers over the world insist on the need for an efficient mechanism to predict and improve the crop growth. The need for an integrated crop growth control with accurate predictive yield management methodology is highly felt among farming community. The complexity of predicting the crop yield is highly due to multi-dimensional variable metrics and unavailability of predictive modelling approach, which leads to loss in crop yield. It works on an adaptive cluster approach over dynamically updated historical crop data set to predict the crop yield and improve the decision making in precision agriculture.*

Keywords— Crop Yield Prediction, Data Mining Technique, Crop Growth, Precision Agriculture

I. INTRODUCTION

Data mining is a ground-breaking technology, developing with database and artificial intelligence. It is a processing overture of action of extracting trustworthy, novel, useful and understandable patterns from database. Achieving maximum crop yield at minimum cost is one of the goals of agricultural production. Early detection and management of problems associated with crop yield indicators can help increase yield and subsequent profit. By influencing regional weather patterns, large-scale meteorological phenomena can have a significant impact on agricultural production. Predictions could be used by crop managers to minimize losses when unfavorable conditions may occur. Additionally, these predictions could be used to maximize crop prediction when potential exists for favorable growing conditions [2].

Nowadays, there are a lot of yield prediction models, that more of them have been generally classified in two groups: a) Statistical Models, b) Crop Simulation Models. Recently, application of Artificial Intelligence (AI), such as Artificial Neural Networks (ANNs), Fuzzy Systems and Genetic Algorithm has shown more efficiency in dissolving the problem. Application of them can make models easier and more accuracy from complex natural systems with many inputs. In this it has been tried to develop a various crop yield prediction model using ANNs [13]. To handle such a situation, an extremely versatile approach of "Artificial neural networks" (ANNs) is developing rapidly. Most widely used ANN is feed forward back propagation artificial neural network. As an illustration, the methodology has been applied for modeling and forecasting of various crop yield on the basis of various predictor variables, viz. type of soil, PH, nitrogen, phosphate, potassium, organic carbon, calcium, magnesium, manganese, copper, iron, depth, temperature, rainfall, humidity. ANN with zero, one, and two hidden layers has been considered [13].

This work will make four contributions,

- Obtaining/ gathering the data regarding farm knowledge such as crop types, soil types, soil-PH value, crop disease and pesticides, seasonal parameters such as kharif, rabi and summer crops. The knowledge-base also consists of zonal as well as district information, environmental parameter such as maximum and minimum temperature value and average rainfall,
- Normalization of the data gathered.
- Evaluating the performance of the system using various classification algorithms.
- Selecting appropriate classifier.
- Building a simulation of the crop yield prediction over the course of different kinds of clustering algorithms considering various cluster centers.

II. LIMITATIONS OF THE EXISTING SYSTEM

Although a lot of research has been conducted for developing the decision support system for farmers, the most of the research focus on the crop management, crop disease management and crop yield forecasting. But the farmer's crop selection at the earlier stage is one of the most important factors since appropriate crop selection at the earlier stage will help farmers to improve crop management and crop yield.

A crop prediction is a widespread problem that occurs. During the rising season, a farmer had curiosity in knowing how much yield he is about to expect. In the earlier period, this yield prediction become a matter of fact relied on farmer's long-term experience for specific yield, crops and climatic conditions. Farmer directly goes for yield prediction rather than concerning on crop prediction with the existing system. Unless the correct crop is predicted how the yield will be better and additionally with existing

systems pesticides, environmental and meteorological parameter related to crop is not considered.

The management of agricultural production operations can be complex and daunting. A manager who is faced with a decision confronts many factors that need to be simultaneously considered. In addition to facing a management decision that will potentially improve the immediate operation, the manager must ultimately be accountable to society and to the environment. The impacts of decisions go far beyond the farm. Agricultural decision support provides managers with recommendations for specific situations and help with analysing choices.

The project will assist the farmers for selecting the crop properly at the earlier stage according to the climatic conditions of his farm. The proper selection of the crop at earlier stage will ultimately improve the crop yield and help to reduce the depression of the farmer as well as there will be no need of re-cropping.

III. PROPOSED WORK

A. Artificial Neural Network

An ANN is typically defined by three types of parameters:

1) The interconnection pattern between different layers of neurons
2) The learning process for updating the weights of the interconnections
3) The activation function that converts a neuron's weighted input to its output activation.

One type of network sees the nodes as "artificial neurons". These are called artificial neural networks (ANNs). The back propagation algorithm is used in layered feed-forward ANNs. This means that the artificial neurons are organized in layers, and send their signals "forward", and then the errors are propagated backwards. The network receives inputs by neurons in the input layer, and the output of the network is given by the neurons on an output layer. There may be one or more intermediate hidden layers. We shall examine one of the most common neural network architectures, the feed forward back propagation neural network. This neural network architecture is very popular, because it can be applied to many different tasks. The first term, "feed forward" describes how this neural network processes and recalls patterns. In a feed forward neural network, neurons are only connected foreword. Each layer of the neural network contains connections to the next layer (for example, from the input to the hidden layer), but there are no connections back. The term "back propagation" describes how this type of neural network is trained. Back propagation is a form of supervised training. When using a supervised training method, the network must be provided with both sample inputs and anticipated outputs. The anticipated outputs are compared against the actual outputs for given input. Using the anticipated outputs, the back propagation training algorithm then takes a calculated error and adjusts the weights of the various layers backwards from the output layer to the input layer.

B. Supervised Learning

Which incorporates an external teacher; so that each output unit is told what its desired response to input signals ought to be during the learning process global information may be required. An important issue concerning supervised learning is the problem of error convergence, i.e. the minimization of error between the desired and computed unit values as shown in figure1 the aim is to determine a set of weights which minimizes the error. One well-known method, which is common to many learning paradigms, is the least mean square (LMS) convergence.

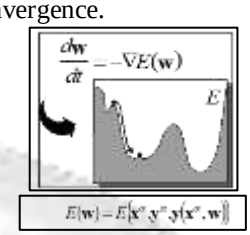

Fig. 1: Supervised learning

C. Transfer Function

The behavior of an ANN (Artificial Neural Network) depends on both the weights and the input-output function (transfer function) that is specified for the units. This function typically falls into one of three categories:

- linear (or ramp)
- threshold
- sigmoid $f(a) = 1/(1+\exp(-a))$

For linear units, the output activity is proportional to the total weighted output.

For threshold units, the output are set at one of two levels, depending on whether the total input is greater than or less than some threshold value.

For sigmoid units, the output varies continuously but not linearly as the input changes. Sigmoid units bear a greater resemblance to real neurons than do linear or threshold units, but all three must be considered rough approximations. There is a three-layer network to perform a particular task by using the following procedure: The network with training examples, which consist of a pattern of activities for the input units together with the desired pattern of activities for the output units. That determine how closely the actual output of the network matches the desired output. It changes the weight of each connection so that the network produces a better approximation of the desired output.

D. Train a neural network

Train a neural network to perform some task, it has to adjust the weights of each unit in such a way that the error between the desired output and the actual output is reduced. This process requires that the neural network compute the error derivative of the weights. In other words, it must calculate how the error changes as each weight is increased or decreased slightly. ANN is set up using feed forward network and has been trained with back propagation algorithms. It is well known that a feed forward network with one hidden layer and enough neurons in the hidden layer is capable of fitting any input-output data. Hence, in this network it used one hidden layer and varies the number of neurons to obtain the optimum performance. To train ANN model using the training set and validate the trained model with the testing data set. The use of testing set for validation ensures that are able to accurately measure the generalization performance of trained model. As shown in figure 2 first to select zone, district and season then it will

ask for input which should be enter as parameters such as soil ph, maximum and minimum temperature, then get test data in which open pre-processed data set then train neural network

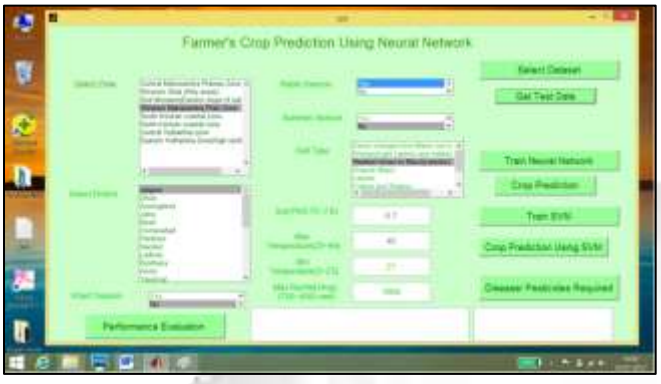

Fig. 2: Input parameters such as soil ph, min temp, and max temp

Matlab provides high-level network creation functions, like newlin (create a linear layer), newp (create a perceptron) or newff (create a feed-forward backpropagation network) to allow an easy construction of. As an example we construct a perceptron with two inputs ranging from -2 to 2:

>> net = newp([-2 2;-2 2],1)

First the architecture parameters and the subobject structures

subobject structures:

- inputs: {1x1 cell} of inputs
- layers: {1x1 cell} of layers
- outputs: {1x1 cell} containing 1 output
- targets: {1x1 cell} containing 1 target
- biases: {1x1 cell} containing 1 bias
- inputWeights: {1x1 cell} containing 1 input weight
- layerWeights: {1x1 cell} containing no layer weights are shown.

The latter contains information about the individual objects of the network. Each layer consists of neurons with the same transfer function net. transferFcn and net input function net.netInputFcn, which are in the case of perceptrons hardlim and netsum. If neurons should have different transfer functions then they have to be arranged in different layers. The parameters net.inputWeights and net.layerWeights specify among other things the applied learning functions and their parameters.

During training, the training tool window opens and displays the progress. Training details such as the algorithm, the performance criteria, the type of error considered, etc. are shown. This gives the prediction for crop that which crop is suitable as per the given parameters, through which the farmers can get the maximum crop production as shown in figure 3,

E. Training an SVM Classifier

Train an SVM classifier with the svmtrain function. The most common syntax is:

SVMstruct = svmtrain(data,groups,'Kernel_Function','rbf');

The inputs are:

Data-Matrix of data points, where each row is one observation, and each column is one feature.

- Groups — Column vector with each row corresponding to the value of the corresponding row in data. Groups should have only two types of entries. So groups can have logical entries, or can be a double vector or cell array with two values.
- Kernel Function — the default value of 'linear' separates the data by a hyperplane.

The syntax for classifying new data with a SVMstruct structure is:

newClasses = svmclassify(SVMstruct,newData)

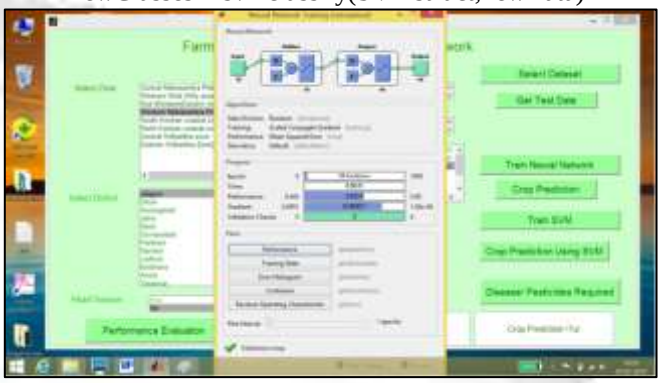

Fig. 3: Train Neural Network

Fig. 4: Train SVM

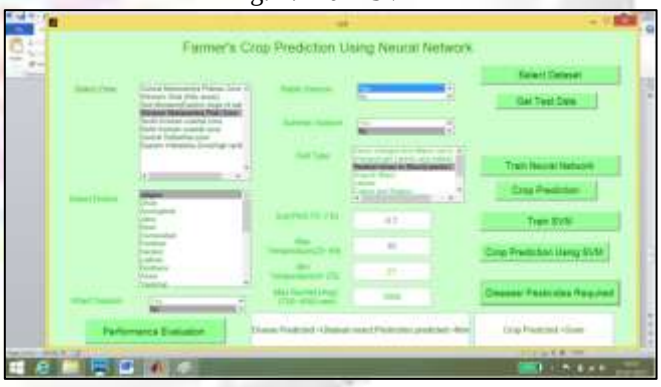

Fig. 5: Disease Predicted and Pesticides required

Figure 5 shows that which pesticides can be used according to the disease prediction,

Fig. 6: Performance Evolution

Figure 6 shows the performance evolution among neural network and SVM in which neural network shows more accuracy than SVM. The SVM classifier is slow to train as well as slow to use. It also has large memory requirements. The classifier performs better for small data sets but upon increasing the size of the training set the computational requirements get increased. Various techniques can be incorporated for increasing speed. Disadvantages of SVM are Time and space required for training and run is more, it is very sensitive to noisy data, prone to over fitting and thus bad generalization. The results are greatly influenced by the choice of kernel function and the manually setting of parameters than SVM.

IV. Conclusion

Agriculture is backbone to every country in the world either directly and indirectly. Mankind continues its journey with food produced by means of agriculture. From the review of literature it is understood that the state-of-the-art technology usage in agriculture has been disappointing. The adaptation of technology towards better productivity in agriculture has been slow and further research is needed to expedite it towards precision agriculture. Crop yield prediction is a significant component of national food security assessment and food policy making. Crop growth and yield data are critical for regulating agricultural cultivation system, and agricultural operation and management. The proposed crop yield prediction consists of three phases namely, pre-processing, feature reduction and prediction. Here the proposed method use input data as real world data. Real world data is often incomplete, inconsistent, and/or lacking in certain behaviour or trends, and is likely to contain many errors. Agriculture in the country can be improved with automatic prediction of crop yield based on the reliable variables. The system helps farmers to do right things at right time. The productivity gets improved in agriculture with sustained research in the field of spatial data mining to realize precision agriculture. Several Data mining techniques used in agriculture study area .Different changes of weather are analyzed using SVM. K means approach is used to classify the soil and plants. In this way concluded that ANN is beneficial tool for crop prediction. This includes the parameter of their regional soil parameter. Then it is analyze by using feed forward back propagation ANN. Analyzed in mat lab ANN approach to make it more efficient.

With the help of present model, the farmers can access interactive and flexible information for selecting the crop according their farm's climatic condition. The model will help the farmers in increasing their productivity by selecting the appropriate crop for their land and climatic conditions. Along with the crop choice, the farmers will be provided with the crop diseases and pesticides required information. Association rule mining will be performed for finding the associations between rainfall, soil type and crop produced etc. This mined information will be very much useful for the farmers in selecting the appropriate crop. The agricultural framework for crop selection at the earlier stage with all the ready information helps the farmers in a very useful manner. The farmers can get all the information at just a click of the mouse, and they need not to travel to Agricultural Universities for that.

Acknowledgement

Department of Computer Science and Engineering, Shri Sant Gadge Baba College of Engineering & Technology, Bhusawal, Maharashtra, helped to complete this work.

References

[1] M. Ananthara, T. Arunkumar, and R. Hemavathy, Cry: An improved crop yield prediction model using bee hive clustering approach for agricultural data sets," in Pattern Recognition, Informatics and Medical Engineering (PRIME), 2013 IEEE International Conference, February 21-22, pp. 473-478.

[2] U. Narkhede and K.P.Adhiya, "A Study of Clustering Techniques for Crop Prediction - A Survey", American International Journal of Research in Science, Technology, Engineering & Mathematics, February 2014, ISSN no: 2328-3491, pp. 45-48, 2014.

[3] A. Mucherino and G. Rub, "Recent developments in data mining and agriculture", Leipzig, Germany: IBaI Publishing, September 2011.pp.3-7.

[4] G. Nasrin Fathima, R.Geetha "Agriculture Crop Pattern Using Data Mining Techniques" Volume 4, Issue 5, May 2014, pp.781-783.

[5] S. J. Cunningham and G. Holmes, "Developing innovative applications in agriculture using data mining," 1999.pp.2-3.

[6] M.Kannan, S.Prabhakaran, and P.Ramachandran, "Rainfall forecasting using data mining technique," International Journal of Engineering and Technology, vol. 2, no. 0975-4024, , 2010,pp. 397-401.

[7] D Ramesh, B Vishnu Vardhan, "Data Mining Techniques and Applications to Agricultural Yield Data", Vol. 2, Issue 9, September 2013, pp. 3477-3479.

[8] Aakunuri Manjula, Dr.G.Narsimha," XCYPF: A Flexible and Extensible Framework for Agricultural Crop Yield Prediction", IEEE Sponsored 9th International Conference on Intelligent Systems and Control (ISCO) 2015, pp.1-2.

[9] R. A.A. and K. R.V., Review - role of data mining in agriculture," International Journal of Computer Science and Information Technologies (IJCSIT), 2013,vol. 4(2), no. 0975-9646 pp 270-272.

[10] J.Y.Hwanjo Yu and J.Han, in Classifying Large Data Sets Using SVMs with Hierarchical Clusters. ACM New York, NY, USA 2003, pp. 306-315.

[11] D. M. Kiri L. Wagstaff and and S. R. Sain, "Harvist: A system for agricultural and weather studies using advanced statistical methods," 2005.pp.44-46.
[12] A.Awan and M.Md.Sap, "A framework for predicting oil-palm yield from climate data", International Journal of Information and Mathematical Sciences, 2012.pp.43-46.
[13] Miss.Snehal S.Dahikar, Dr.Sandeep V.Rode, "Agricultural Crop Yield Prediction Using Artificial Neural Network Approach", Vol. 2, Issue 1, January 2014, pp-683-685.
[14] W.Duan, Qiu, L Duan, Q. Liu, and H.-q. Huan, "An improved rough k-means Algorithm with weighted distance measure," in Granular Computing (GrC), 2012 IEEE International Conference, pp. 97-101.
[15] Ramesh A. Medar,Vijay S.Rajpurohit, "A survey on Data Mining Techniques for Crop Yield Prediction ", Volume 2, Issue 9, September 2014,pp.59-61.
[16] Simon Haykin "Kalman Filtering and Neural Networks" Copyright © 2001 John Wiley & Sons, Inc. ISBNs: 0-471-36998-5 (Hardback); 0-471-22154-6 (Electronic), 2001, pp.5-21.
[17] The Neural network analysis data filters Alexey Minin, JASS 2006, and pp. 4-11.