Improved Crop Yield prediction Using Neural Network

Mrs.Prajakta Prashant Bhangale[1], Prof. Yogesh S. Patil[2], Prof. Dinesh D. Patil[3]

[1]M.E. Student, Dept. of Computer Science and Engineering, SSGBCOET, Bhusawal, Maharashtra, India
[2]Assistant Professor, Computer Science and Engineering, SSGBCOET, Bhusawal, Maharashtra, India
[3]HOD, Computer Science and Engineering, 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.

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 modeling 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.*

---

1. 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. Data mining with a view of its socking business vision are now becoming a data library and information strategy-making in the field of agriculture research [1], [2]. The research of agriculture field in data mining such as discovering wine fermentation and predicting yield using sensor data [3], a classification system for sorting mushrooms by grade [4], rainfall forecasting for crop growth [5] and highest humidity prediction [6] are developed. In order to do such data analysis in a clustering plays an important role for finding data information and pattern recognition in data mining[7,8].

Recommender System

- The goal of a Recommender System is to generate meaningful recommendations to a collection of users for items or products that might interest them.
  Examples:

- Suggestions for books on Amazon, or movies on Netflix.
- The design of such recommendation system depends on the domain and the particular characteristics of the available data.
- Collaborative Filtering systems analyze historical interactions alone, while Content-based Filtering systems are based on profile attributes; and Hybrid techniques attempt to combine both of these designs.

The architecture of recommender systems and their evaluation on real-world problems is an active area of research.

2. LITERATURE SURVEY

To understand the advancement of clustering techniques, it is essential to briefly examine their history. In data mining, clustering plays an important role for finding data information and pattern recognition. Clustering is comes under the descriptive model that identifies patterns or relationship in data. Clustering is the process of grouping the data into classes or clusters, so that objects within a cluster have high similarity in agreement to one another but are very dissimilar to objects in option clusters. Hierarchical micro clustering algorithm, constrained k-Means algorithm, SWK k-Means algorithm, expectation maximization Algorithm, and improved kMeans algorithm is clustering techniques. [4], [5], [6], [7].

A. Hierarchical Micro-Clustering Algorithm

Hwanjo Yu et.al presented clustering based support vector machine (CB-SVM) designed for handling very large data sets in 2003. Basically SVM is data classification method whose training complexity highly depends on size of data. So it is not worked for large dataset. In order to work authors have designed CB-SVM for handling large dataset. CB-SVM is a hierarchical micro clustering algorithm that scan entire data set only once to provide an SVM with high quality of sample that carry statistical summaries of the data. Hierarchical micro clustering algorithm introduces the clustering feature and clustering feature tree. Clustering feature is triple which summarize the information about cluster of objects such as number of data points, linear sum of data points and square sum of data points. Clustering Feature tree is a height balanced tree that stores the clustering features for a hierarchical clustering. The algorithm starts by scanning the database to build CF tree then SVM boundary function is trained from the centroids of the root entries of CF tree .The low margin clusters is determined that are close to the boundary and thus needs to be declustered into the finer level. Hierarchical micro-clustering algorithm shows good quality of cluster for large dataset but it is expensive to update and store the cluster and also splitting and merging the data degrades performance. [4]

B. Constrained k-Means Algorithm

KiriWagstaff et.al developed HARVIST (Heterogeneous Agricultural Research via Interactive, Scalable Technology) graphical interface that allows user to interactively run auto-matic classification and clustering algorithm in 2005. The used constrained k- means clustering algorithm for pixel clustering which merge the concept of constraint-based and partitioning methods. In k-means instance alongside constraint are useful way to express a Prior knowledge almost which instance should or should not be grouped together based on Must link and cannot-link constraints. It shows good quality of clusters for huge datasets and additionally gives better activity than hierarchical clustering, but it has drawbacks such at the time that local optima problem, cannot find arbitrary shape cluster, backbreaking to gain initial value of cluster feelings, affected by to noise. [5]

C. SWK k-Means Algorithm

A MajidAwan et.al has developed a software system in 2007 for predicting Oil-Palm Yield from climate and plantation data. It used unsupervised partitioning of data for finding spatio temporal patterns using kernel method. By using only k-means partitioning method it is burdensome to deal with abstract data so authors have incorporated kernel method. Kernel exercise implicitly defines a non-linear transformation that maps the data from their original space to a high dimensional space where data are expected to be more. A weighted kernel kmeans clustering algorithm is used which incorporated spatial constraints to spare spatial neighbourhood information in order to handle noise. This algorithm requires added information such as kernel matrix, no of clusters, weight for each point, stopping criterion, penalty term parameter. It shows good quality of clusters for huge datasets and along with give better performance than hierarchical clustering, but it has global optima problem. [6]

D. Expectation Maximization (EM) Algorithm

Sun Kim et.al proposed model for theme-based clustering algorithm that capture probabilistic for text documents in 2011.Probabilistic clustering comes under model-based clustering methods in which data are generated by mixture of probability distributions. Given text, subject terms are extracted and used for clustering document in a probabilistic framework. An EM algorithm is used for learning the proposed model in order to ensure annals are assigned to correct themes. EM algorithm is an iterative refinement algorithm used to estimation the parameters of the probability distribution so at the time that to best firm the data. It starts with initial estimates.

Then, it iteratively refines the parameters based on Expectation and Maximization step. It is good in handling with real world dataset but becomes sensitive to noise and along with highly abstract in nature. [7]

E. Improved k-means Algorithm

DUAN weing-ying et.al proposed improved k means clustering algorithm with weighted based on density in 2012. It proposed a solution to search initial central points and combines it with a distance measure with weight. This improved algorithm decreases the level of interference brought by the isolated points to the kmeans algorithm, and makes the clustering analysis more effective and objective. An Improved k-Means clustering algorithm requires add-on parameter such as density, threshold and number of cluster. It is efficient in local as well as global optima and additionally reduces impact of noise data. [8]

F. BeeHive Algorithm

M. Gunasundari et.al suggested crop yield prediction model which is used to predict crop yield from historical crop data set in 2013. A relational cluster Bee Hive algorithm is proposed for extracting yield patterns across multiple data sets. The outcome helps in identification of and investigates areas of unusually high or low yield. The BeeHive Algorithm is an optimization algorithm inspired by the natural foraging behaviour of honey bees to find the optimal solution. It searches for appropriate cluster centres such that the clustering metric is minimized. It requires the specific parameters to set such as scout bees, number of sites selected from neighbourhood search, qualified sites, best site, and number of bees recruited for the remaining selected sites and size of patch. The BeeHive algorithm is good in handling large dataset and efficient in local as well as global optima. [1]

G. Neural network:

Work on artificial neural networks, commonly referred to as neural networks, has been motivated right from its inception by the recognition that the brain computes in an entirely different way from the conventional digital computer. The struggle to understand the brain owes much to the pioneering work of RamónyCajál (1911), who introduced the idea of neurons as structural constituents of the brain. Typically, neurons are five to six orders of magnitude slower than silicon logic gates; events in a silicon chip happen in the nanosecond (10-9 s) range, whereas neural events happen in the millisecond (10-3 s) range. However the brain makes up for the relatively slow rate of operation of a neuron by having a truly staggering number of neurons (nerve cells) with massive interconnections between the brain is a highly complex, nonlinear, and parallel information-processing system. It has the capability of organizing neurons so as to perform certain computations (e.g. pattern recognition, perception, and motor control) many times faster than the fastest digital computer. It resembles the brain in two respects: Knowledge is acquired by the network through a learning process. Interneuron connection strengths known as synaptic weights are used to store the Knowledge. Benefits of neural networks: A neural network derives its computing power through, first, its massively parallel distributed structure and, second, its ability to learn and, therefore, generalize. The use of neural networks offers the following useful properties and capabilities: [5]

1. Nonlinearity: A neuron is basically a nonlinear device. Consequently, a neural network, made up of an interconnection of neurons, is itself nonlinear. Moreover, the nonlinearity is of a special kind in the sense that it is distributed throughout the network.

2. Adaptively: Neural networks have a built-in capability to adapt their synaptic weights to changes in the surrounding environment.

3. Contextual information: Knowledge is represented by the very structure and activation state of a neural network.

4. Fault tolerance: A neural network, implemented in hardware form, has the potential to be inherently fault tolerant in the sense that its performance is degraded gracefully under adverse operating.

5. VLSI implements ability: The massively parallel nature of a neural network makes it potentially fast for the computation of certain tasks.

6. Uniformity of analysis and design: Basically, neural networks enjoy universality as information processors.

7. Neurobiological analogy: The design of a neural network is motivated by analogy with the brain, which is a living proof that fault-tolerant parallel processing is not only physically possible but also fast and powerful.

3. WORKING OF THE SYSTEM

- 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.

The proposed agricultural DSS framework provides a means to predict the cropping information in advance from a set of inputs. It comprises of five major components including data input, data mining, and prediction and visualization module, association rule mining for analysing the relationship between soil type and crop production and crop management module. The framework for agricultural decision system is depicted in following figure.

The database will store agricultural and climatic data. The framework will provide the data input unit which will take users input via graphical user interface (GUI). User will select his agricultural zone and district. Accordingly from the database, soil, temperature and rail fall information will be retrieved from the database. The data mining will use both classification and association rule mining algorithms. The farmers will be suggested for the appropriate crop using the prediction functionality. Data visualization will be provided to display the information to user in the form of scatter plot or various graphical representations. The proposed agricultural DSS will provide a mean to assist farmer decision making for regional specific environment and cropping practices. The system is also extended for suggesting farmers with the probable diseases and fertilizers required.

1. Study Area and Data

The study area for this research will be the major states from INDIA. Accurate collection and recording of agricultural knowledge, not surprisingly, must reflect the complexity of the knowledge itself. Such collection is difficult and success, not surprisingly, seems to require methods appropriate for the knowledge. The data will be gathered from the publically available resources and agriculture website of the government. Each state is divided in to number of agricultural zones according to rainfall and geographic location. For example, Maharashtra state is divided into 9 agricultural zones as bellow

1. South Kokan Coastal Zone
2. North Kokan Coastal Zone
3. Western Ghat Zone
4. Transition Zone-1
5. Transition Zone -2
6. Scarcity Zone
7. Assured Rainfall Zone
8. Moderate Rainfall Zone
9. Eastern Vidharbha Zone

For each zone, the climatic and agricultural data will be collected such as soil type, soil Ph, Average rainfall, max., min. temperature, humidity etc. This information will be used for predicting the appropriate crop.

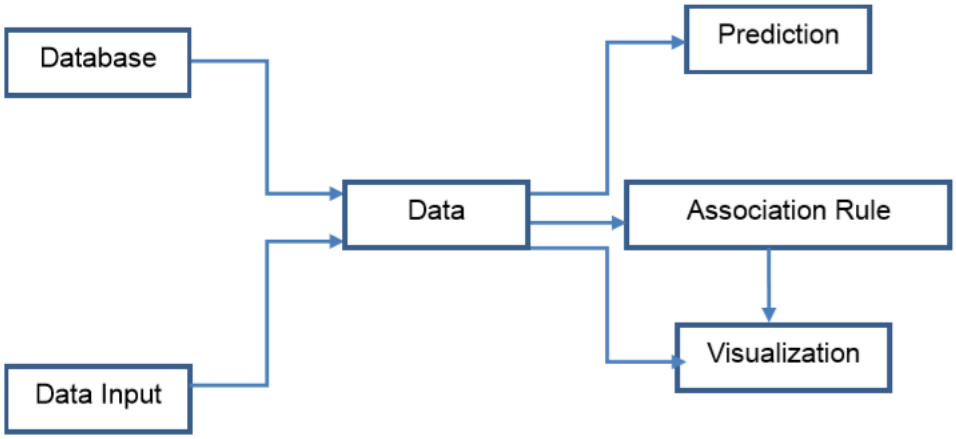

Figure 3.1The Framework for Agricultural Decision System

2. Methodology Used

The phases that will be used in this research are

A. Data Set Pre-processing

B. Data mining for crop Prediction

C. Association Rule mining for gathering more information

D. Crop Diseases and pesticides required.

E. Data Visualization

A. Data Set Preprocessing

The data sets will be integrated using Microsoft Excel. The raw data collected form public resources will require extensive pre-processing for handling missing values and other data inconsistencies. The data needs to be normalized for classification or association rule analysis.

B. Data mining for crop Prediction

Various Data mining classification algorithms will be evaluated for better performance such as Neural Network, SVM. These data mining classification techniques will be used for predicting the appropriate crop to the farmer according to his field's climatic and soil information available.

C. Association Rule mining for gathering more information

Association rule mining will be carried out on the dataset to generate different associations with crop yield, rainfall, soil type etc. These associations will be useful for deciding the appropriate crop.

D. Crop Diseases and Pesticides required

Once the farmer has been suggested with the appropriate crop, he needs to decide the expenses required throughout the season. The most influencing part here is the crop diseases and pesticides required. The farmers will be provided information about the probable crop diseases in his area and the pesticides required.

E. Data Visualization

In order to develop a general understanding of the dataset, a data visualization module will be designed. Visualization of climate and other data will be carried out. A visual comparison of rainfall trends across years and spatially across districts will be used for further analysis of crop prediction.

4. ADVANTAGES

The timely information to the farmers is closely linked to the Agricultural Development and well-being of the rural communities. Quick information transfer between the researchers and the farmers has specific importance. Hence this research work is focused on the problem of developing an agricultural framework for assisting farmers in selection of the crop based on various parameters. The present work will provide the platform for the farmers to make the use of technology to take important decision of crop selection. 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.
- If implemented at Village, District and State level, the model will provide valuable information to other agencies and panchayats in particular.
- The proposed system attempts to stimulate the farmer's behavior in selecting farming systems given relevant constraints and then aggregating up to the node. A large number of database queries can be generated according to Crop, Water Availability and Requirement, Socio-economic constraints and so on.
- 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.

For management problems in farms, intelligent DSSs in agriculture have been introduced to monitor and to assist farmers to make decisions in a timely manner. However, designing a DSS is quite complex; it requires knowledge from various multidisciplinary areas, such as crop agronomy, computer hardware and software, mathematics and statistics to analyze data. For example, to understand crop growth, it is necessary to know how each variable affects crop growth.

5. APPLICATIONS

1. Crop knowledge base:

The crop knowledge base consists of farm knowledge such as crop types, soil types, soil-ph value, crop disease and pesticides, seasonal parameter 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.

2. Feature Selection/Normalization of data:

The feature selection module is responsible for selection of attribute from crop knowledge-base for partitioning. The data collected needs to be normalised/ pre-processed in the specified range.

3. Classification Approaches:

Neural Network approach will be used for evaluation. The appropriate number of hidden selected. Various classification algorithms will be evaluated for crop prediction.

4. Sample Testing and Prediction:

There is need to provide input parameters such as zone, district, and selection of seasons, soil type, maximum temperature, minimum temperature and average rainfall for sample testing. Based on the output values of each clustering, the test data calculates the distance measure with clustering output and selects minimum distance as a predicted value.

6. CONCLUSION

With the improvement of data mining technologies, especially those without any premises or humans subjective, data mining can be applied in many areas. In this some data mining techniques were adopted in order to estimate crop yield analysis with existing data and their use in data mining. This presents new research possibilities for the application of modern classification methodologies to the problem of yield prediction. There are a growing number of applications of data mining techniques in agriculture and a growing amount of data that are currently available from many resources. It is observed that efficient technique can be developed and analyzed using the appropriate data, the data which is collected from the district to solve complex agricultural problems using Data mining techniques. 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 matlab ANN approach to make it more efficient.

7. REFERENCES

[1] M. Ananthara, T. Arunkumar, and R. Hemavathy, Cry: An improved crop yield pre-diction model using bee hive clustering approach for agricultural data sets," in Pattern Recognition, Informatics and Medical Engineering (PRIME), 2013 IEEE International Conference , pp. 473478.

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

[5]M.Kannan, S.Prabhakaran, and P.Ramachandran, "Rainfall forecasting using data mining technique," International Journal of Engineering and Technology, vol. 2, no. 0975-4024, pp. 397401, 2010.

[6] Kiruthika.V.G, Arutchudar.V, and S. Kumar.P, "Highest humidity prediction using data mining techniques," International Journal of Applied Engineering Research, vol. 9, no. 16, pp. 32593264, 2014.

[7] J. Han, M. Kamber, and J. Pei, Data Mining, Second Edition: Concepts and Techniques, ser. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science, 2006.pp.23-27

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] S. Kim and. Wilbur, “An em clustering algorithm which produces a dual representation," in Machine Learning and Applications and Workshops (ICMLA), 2011 IEEE 10th International Conference on, vol. 2, pp. 90-95.

[13] 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.

[14] D. Arthur and S. Vassilvitskii, “k-means++: The advantages of careful seeding," in SODA '07 Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms, 2007-13, pp. 1027-1035.