Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 5, May 2015, pg.491 – 495

REVIEW ARTICLE

A REVIEW ON CLASSIFICATION TECHNIQUES OVER AGRICULTURAL DATA

[A]Dr. S.Hari Ganesh, [B]D.Pritty Cindrella, [C]A.Joy Christy

[1]Assistant Professor, [2]Mphil.Scholar, [3]Research Scholar

Department of Computer Science, Bishop Heber College (Autonomous), Tiruchirapalli, India

[1] hariganesh17@gmail.com, [2] prittycindrella91@gmail.com, [3] joychristy001@gmail.com

Abstract: Data mining is the technique of extracting useful information from a large dataset. It is the process of discovering previously unknown and potentially interesting pattern in a large database. Nowadays the information technology plays a vital role in our day today life this is especially true for agriculture. This paper describes the role of data mining in agriculture. The agricultural mining is the process of identifying the hidden patterns from a large of agricultural data. Data mining in the field of agriculture is quite a novel research field. This paper gives the survey of some data mining techniques and the techniques used in agricultural data mining.

Key-words: Data mining, clustering, agricultural mining, J48

I. INTRODUCTION

Data mining is the technique of extracting useful information from large amount data set. It is the process of discovering meaningful new correlation, patterns and trends by shifting through large amount of data using Patten recognition technology as well as statistical and mathematical techniques. Data mining and knowledge discovery in database (KDD) are concerned with extracting patterns and models of interest from huge databases. KDD says that "knowledge" is the necessary end product of given data-driven discovery. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. The patterns, associations, or relationships among all this data can provide information.

Clustering is the classification of objects into different groups, the partitioning of a data set into subsets (clusters), so that the data in each subset shares some common features 4 according to some defined distance measure. Clustering plays an important role in agricultural mining, since we live in a world full of data where we encounter a large amount of information. One of the vital means in dealing with these data is to classify or group them into a set of categories or clusters. Clustering finds application in many fields. For example, data clustering is a common technique for statistical data analysis, which is used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics. Also, clustering is used to discover relevance knowledge in agricultural data. This paper is organized as follows. Section II describes literature survey. Section III explains mining techniques. Section IV concludes the paper.

II. LITERATURE SURVEY

Yethiraj .N.G et al.,[1] reviewed the applications of data mining techniques and found out that there are several algorithms and techniques being applied in agricultural domain. Similarly, Barghavi .P et al., [2], reviewed that the data mining techniques could be applied to characterize soil data and found that data mining depends on the amount of data used in the process. An increase in dataset size improves accuracy, which may improve the verification of valid patterns compared to standard statistical analysis.

Yethiraj .N.G et al., [3] made an attempt to review the research studies on application of data mining techniques in the field of agriculture. Some of the techniques, such as, ANN, ID3, the k-means, and the k-NN and support vector machines applied in the field of agriculture. Data mining in application in agriculture is a relatively new approach for forecasting orpredicting of animals, agricultural management[3].

Narsi Reddy Gayam in his research study examines the assumption of normality of crop yields using data collected from INDIA involving sugarcane and Soybean. The null hypothesis (Crop yield are normally distributed) was tested using the Lilliefore method combined with intensive qualitative analysis of the data[4].

Raoranne A. A., et al.,[5], discussed how data mining can bridge knowledge of the data to crop yield estimation. The study assessed new data mining techniques and was applied to various variables to establish if meaningful relationships can be found. He observed that efficient techniques can be developed and analyzed using appropriate data to solve complex agricultural problems using data mining techniques[5].

Ramar .K., et al., [6]. says Data mining classification techniques applied to soil database can be successful in establishing meaningful relationships from the data.

Genetic Algorithm (GA) is an effective tool to use in data mining and pattern recognition. However, GA has problems with premature convergence which inhibit diversity in the population and prevent exploration of the whole search space. To address this problem, the work of A. Hassani, et at., [7].suggested tweaking the GA to a specific problem and correctly set all parameters. The negative selection method is used by L. Na-Na, G. Jun-Hua, and L. Bo-Ying [8] to show promising results.

S.Veenadhari, et al., [9] observed the research studies on application of data mining techniques in the field of agriculture. Some of the techniques, such as ANN, ID3, the k-means, the k-NN and support vector machines applied in the field of agriculture were presented. Data mining in application in agriculture is a relatively new approach for forecasting or predicting of agricultural crop or animal management. This article explores the applications of data mining techniques in the field of agriculture and allied sciences.Thesupply chain operation of companies engaged in industries that use agricultural produce as raw material is important for Historical crop yield information. Animal feed, seed, chemical, poultry, fertilizer pesticides, seed, paper and many other industries use agricultural products as intergradient in their production processes. An accurate estimate of crop size and risk helps these companies in planning supply chain decision like production scheduling. Business such as seed, fertilizer, agrochemical and agricultural machinery industries plan production and marketing activities based on crop production estimates.

yashovardhankelkar, et al,.[10] surveyed and says that data selection is the data relevant to the analysis is decided and retrieved from the various data locations. Data preprocessing is the process of data cleaning and data integration is done. Data cleaning is also known as data cleansing; in this phase noise data and irrelevant data are removed from the collected data. In Data transformation the selected data is transformed into forms appropriate for the mining procedure. It is the crucial step in which clever techniques are applied to extract potentially useful patterns. The discovered knowledge is visually presented to the user[11]Association rule mining is applied to remote sensed imagery (RSI) data composed mainly of images and ground data mainly from the field of agriculture. In most of the cases applying existing algorithms on RSI data for generating association rules can consume a reasonable amount of time. Keeping that in mind an efficient algorithm has been devised for spatial data using Peano count tree (P-tree) structure [12]. P-tree structure provides a lossless and compressed representation of images. Based on P-trees, an efficient association rule mining algorithm PARM with fast support calculation and significant pruning techniques is introduced to improve the efficiency of the rule mining process. According to Qin Ding[12]Experimental results showed that PARM is more efficient than FP- growth and A Priori algorithms when applied on RSI spatial data [12s].

An Association rule is an implication of the form $P \Rightarrow Q$, where $P \cap Q = \Phi$ and P & Q are subsets of all item set I. There are two measures of rule interestingness; they are Support (σ) and Confidence (T). They reflect the usefulness and certainty of the rules. The rule $P \Rightarrow Q$ (support $\sigma = 10\%$, confidence $T = 80\%$) shows that 10% of all the transactions under analysis shows the simultaneous purchase of items P and Q by customers and 80% of confidence shows that 80% of customers who purchased item P also bought item Q [13].Association rules can be classified on the basis Single dimension or Multidimensional and based on level of abstractions involved, i.e. Single level association rules or Multilevel association rules. Variousalgorithms have been proposed for mining the association rules and can be decomposed in two phases.

- Find all the item sets whose support and confidence are greater than the user specified minimum σ and minimum T respectively. Such items are called frequent item sets.
- Frequent items are used to find desired association rules. These rules must satisfy minimum σ and T.

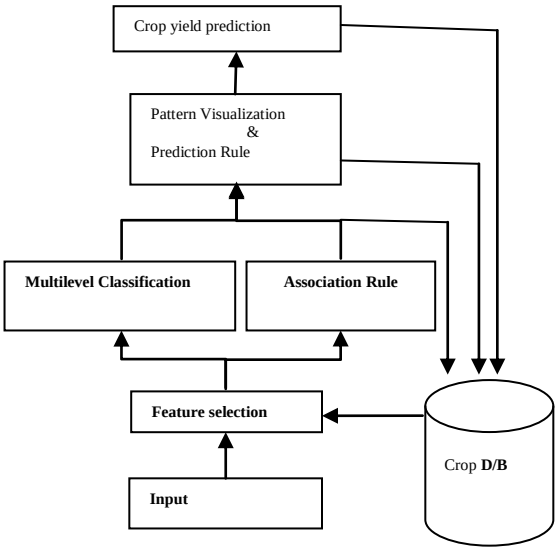

Fig. 1 The Architecture of Crop Prediction Model

III. MINING TECHNIQUES

Data warehouses can be prepared to hold agriculture data, which makes transaction management, information retrieval and data analysis much easier. On Line Analytical Processing (OLAP) can easily answer multidimensional queries it can be used for applications such as forecasting or prediction in agriculture. It also provides an opportunity of viewing agriculture data from different points of view to discover data characterization, data discrimination and association analysis [14].

J48 is associate degree open supply Java implementation of the C4.5 algorithmic data processing tool. C4.5 may be a program that makes a call tree supported a collection of labeled input file. This algorithmic rule was developed by Ross Quinlan. The choice trees generated by C4.5 will be used for classification, and for this reason, C4.5 is usually noted as a applied math classifier.

This algorithmic rule is employed for generating a call tree with naive Bayes classifiers at the leaves (Kohavi R., 1991)

Table 1.Prediction Accuracy of Reviewed Articles

Classifier	NB Tree	Simple Cart
Correctly Classified Instances	1700	1824
-Incorrectly Classified Instances	288	164

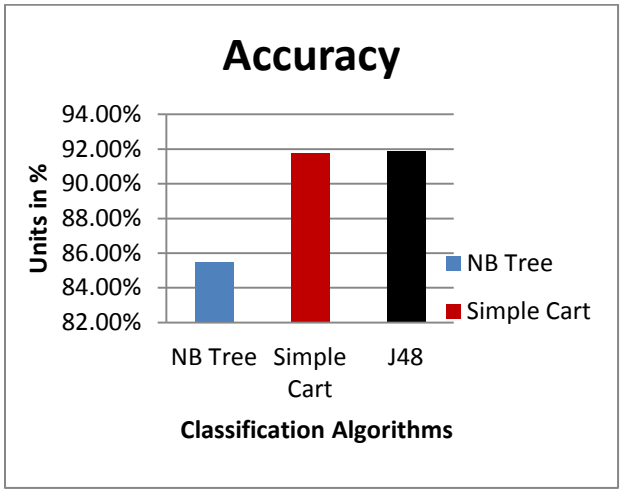

Figure 2 Accuracy Analysis of Reviewed Articles

It's a non-parametric call tree learning technique that produces either classification or regression trees, looking on whether or not the variable is categorical or numeric, severally. It is used for implementing stripped-down cost-complexity pruning (Breiman L. et al. 1984) during this paper, 3 call tree techniques (J48 (C4.5), NBTree and Simple Cart) in data mining were evaluated and compared on basis of accuracy and Error Rate. Denary cross-validation was employed in the experiment. Our studies showed that J48 (C4.5) model clothed to be best classifier for soil samples. The comparsion are tabulated in table 1 and the results are ploted in graph as shown in figure 2

IV. CONCLUSION

This paper is an attempt to provide on over view of some previous researches and studies of data mining and data mining technique in agriculture. Crop prediction model in agricultural mining is to estimate agricultural production as a function of weather and soil conditions as well as crop management. The data mining techniques are used in agricultural field to increase the income of the farmer, reduce the transport cost and to predict the climate change using previously stored data set. This study concludes that the agricultural mining technique become highly active research area in data mining research.

REFERENCES

[1]N. G. Yethiraj, "Applying Data Mining Techniques in the field of agriculture and allied sciences" International Journal of Business Intelligentsvol. 1, no. 2, December 2012.

[2]P. Barghavi and S. Jyothi, "Applying naïve bayesdata mining technique for classification of agricultural land soils," International Journal of Computer Science and Network Security, vol. 9, no. 8, pp 117-122, August 2009.

[3] Yethiraj N G. "Applying Data Mining Techniques IN THE FIELD OF AGRICULTURE AND ALLIED SCIENCES", International Journal of Business Intelligents ISSN: 2278-2400, Vol 01, Issue 02, December 2012.

[4]Roberto Benedetti A, Remo Catenaro A, Federica Piersimoni B, "GENERALIZED SOFTWARE TOOLS FOR CROP AREA ESTIMATES AND YIELD FORECAST "2010.

[5] A. A. Raoranne and R. V. Kulkarni, "Data Mining: An effective tool for estimation in the agricultural sector," International Journal of Emerging Trends and Technologyin Computer Science, vol. 1, no. 2, pp. 75-79, July-August 2012.

[6] R. Vamanan and K. Ramar, "Classification of agricultural land soils a data mining approach," International Journal of Computer Science and Engineering, vol. 3, no. 1, pp. 379-384, 2011.

[7] A. Hassani and J. Treijis"Overview of standard and parallel genetic algorithms", in Proc. IDT Workshop on Interesting Results in Computer Science and Engineering (IRCSE '09) , Mälardalen University, Sweden, October 30, 2009.

[8] L. N. Na, G. J. Hua and L. B. Ying, "A new genetic algorithm based on negative selection", in Proc. 2006 International Conference on Machine Learning and Cybernetics, pp. 4297-4299, 2006.

[9] S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh, "Data mining Techniques for Predicting Crop Productivity – A review article", International Journal of Computer Science and technology, march 2011.

[10]D.Rajesh, International Journal of Computer Applications ,Volume 15, February 2011.
[11]International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 2, February 2012) 275 "Survey on Data Mining"VibhaMaduskar and Prof. yashovardhankelkar
[12] Qin Ding Dept of Comp. Sci. East Carolina Univ,Greenville,NC,Systems,Man and Cybemetics, IEEE Transactions(Volume 38, Issue :6)
[13] "Association Rules Mining",SotirisKotsiantis, DimitrisKanellopoulos.
[14] RameshVamanan and K.Ramar International Journal on Computer Science and Engineering (IJCSE).

