IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

A Study on Data Mining Approaches for Agricultural Intelligence

V.Kaleeswaran[1], Dr.R.Rathipriya [2]

Research Scholar, Department of Computer Science, PeriyarUniversity, Salem, India[1]

Assistant Professor, Department of Computer Science, PeriyarUniversity, Salem, India[2]

Abstract: Agricultural intelligence is a specific and emerging field of intelligence dedicated to an enhanced understanding of cultivation, productivity of crop, and minimized risk associated agriculture. Crop prediction is an important agricultural problem. To address this problem, crop prediction technique is used. It is the one of the most commonly used intelligent technique based on Data Mining (DM) concepts to predict the crop yield for maximizing the crop productivity. This paper studies and records the various data mining techniques available in the literature for better crop productivity.

Keywords*:* Data mining, Crop prediction, k-means, k-nearest neibour, Fuzzy sets, Regression, Classification, Neural network Association Rule.

1. INTRODUCTION

The vision of meeting world demands for the increasing the crop yield for population throughout the world is becoming more important in the recent years. Crop model and decision tools are increasingly used in agricultural fields to improve production efficiency. The combination of advanced technology and agriculture used to improve the production of crop yield.

The data mining techniques like classification, neural networks and regression are required to apply on the realistic data sets for analyses and make the prediction on the agriculture crop yield.

The data mining tasks could be classified in two categories.

- Descriptive data mining.
- Predictive data mining.

Descriptive data mining: Descriptive modeling is a mathematical process that describes real-world events and the relationships between factors responsible for them. The process is used by consumer-driven organizations to help them target their marketing and advertising efforts.

The main aspects of descriptive modeling include:

- Customer segmentation: Partitions a customer base into groups with various impacts on marketing and service.
- Value-based segmentation: Identifies and quantifies the value of a customer to the organization.
- Behavior-based segmentation: Analyzes customer product usage and purchasing patterns.
- Needs-based segmentation: Identifies ways to capitalize on motives that drive customer behavior.

Predictive data mining: Predictive analytics is the process of extracting information from large sets in order to make prediction and estimates about future outcomes. The list of yield prediction models that more of them have been generally classified in two groups methods.

- Traditional Approach
- Artificial Intelligence

Traditional Approach: Most of the farmers were relied on their long-terms experiences in the field on particular crops to expect a higher yield in the next harvesting period.

Artificial Intelligence: The development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision making, and translation between languages.

2.APPLICATIONS OF DATA MINING TECHNIQUES IN AGRICULTURE

This section discusses about the various data mining methodologies commonly used for crop prediction. Crop prediction is the one of the agriculture intelligence techniques. It can be developed by using following methods.

Association Rule : Association rule mining technique is one of the most efficient techniques of data mining to search unseen or desired pattern among the vast amount of data. In this method, the focus is on finding relationships between the different items in a transactional database. [3] Association rules are used to find out elements that co-occur repeatedly within a dataset consisting of many independent selections of elements (such as purchasing transactions), and to discover rules. The simple problem statement is: Given a set of transactions, where each transaction is a set of literals, an association rule is a

phrase of the form X => Y, where X and Y are sets of objects. The instinctive meaning of such a rule is that transactions of the database which contain X tend to contain Y.[4] An application of the association rules mining is the market basket analysis, customer segmentation, store layout, catalog design, and telecommunication alarm prediction.[5]

Classification : Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. It is a process in which a model learns to predict a class label from a set of training data which can then be used to predict discrete class labels on new samples. To maximize the predictive accuracy obtained by the classification model when classifying examples in the test set unseen during training is one of the major goals of classification algorithm.[20] Data mining classification algorithms can follow three different learning approaches: semi-supervised learning, supervised learning and unsupervised learning. The different classification techniques for discovering knowledge are Rule Based Classifiers, Bayesian Networks (BN), Decision Tree (DT), Nearest Neighbor(NN), Artificial Neural Network(ANN), Support Vector Machine (SVM), Rough Sets, Fuzzy Logic, and Genetic Algorithms. [6]

Clustering: In clustering, the focus is on finding a partition of data records into clusters such that the points within each cluster are close to one another. Clustering groups the data instances into subsets in such a manner that similar instances are assembled together, while dissimilar instances belong to diverse groups. [7] Since the aim of clustering is to find out a new set of categories, the latest groups are of interest in themselves, and their assessment is intrinsic. There is no prior knowledge about data. The different clustering methods are Hierarchical Methods(HM), Partitioning Methods (PM), Density-based Methods(DBM), Model-based Clustering Methods(MBCM), Grid-based Methods and Soft-computing Methods [fuzzy, neural network based], Squared Error—Based Clustering (Vector Quantization), network data and Clustering graph.[8]

Regression : Regression is learning a function that maps a data item to a real-valued prediction variable. The different applications of regression are predicting the amount of biomass present in a forest, estimating the probability of patient will survive or not on the set of his diagnostic tests, predicting consumer demand for a new product. [9] Here the model is trained to predict a continuous target. Regression tasks are often treated as classification tasks with quantitative class tag. The methods for prediction are Nonlinear Regression (NLR) and Linear Regression (LR). [30]

Neural networks: It focuses the information about weather and are observed and stored. The recorded parameters are used to forecast weather. If there is a change in any one of the recorded parameters like wind speed, [23] wind direction, temperature, rainfall, humidity, then the upcoming climatic condition can be predicted using artificial neural networks, back propagation techniques. The increase in signal range will work in large areas as well. [10]

Clustering: Data Mining is the process of discovering meaningful patterns and trends by shifting through huge amount of data, using pattern detection technologies as well as statistical and mathematical techniques. Data mining techniques are often used to studied soil characteristics. As an example, the K-Mean approach is used for classifying soils in combination with GPS based techniques. [8]

Fuzzy set : It describes applications to agricultural related areas. Such as Yield prediction is a very important agricultural problem. Any farmer might be interested in knowing how much yield is expected. In the past, [26] yield prediction was achieved by considering farmer's experience on particular field, crop and climate condition. We have discussed additional information about data like probability in probability theory, grade of membership in fuzzy set theory. [19]

Decision tree and Bayesian classification :
The findings of the study revealed that the decision tree analysis indicated that the productivity of soybean crop was mostly influenced by comparative humidity followed by temperature and rainfall. The decision tree analysis shows that the productivity of paddy crop was mostly inclined by Rainfall followed by comparative Evaporation and humidity. [4] For Wheat crop, the analysis shows that the productivity is mostly influenced by Temperature followed by relative humidity and rainfall. The result of decision tree was confirmed from Bayesian classification. The rules formed from the decision tree are useful for identifying the conditions intended for high or low crop productivity. [5]

Bayesian network is a powerful tool and broadly used in agriculture datasets. The model developed for agriculture application based on the Bayesian network learning method.[11] The results show that Bayesian Networks are feasible and efficient.

Table 1.1: Data Mining Methodologies

S.NO	Author	Data mining methodologies
11	D Ramesh. B Vishnu Vardhan.	Density based clustering, multiple linear regression.
22	Luke Bornn. James v. Zidek.	Bayesian, Spatial correlation, Data smoothing.
33	Askar Choudhury. James Jonea.	Data smoothing.

44	Miss.SnehalS.Dahikar Dr.SandeepV.Rode.	Artificial neural network
55	Anup K Prasad. Lim Chai. Ramesh P Singh. MenasKafatos.	Quasi Newton method
66	Aorance A.A Kulkarni R.V	Regression analysis
77	D Ramesh B Vishnu Vardhan	k-means nearest neighbor Artificial neural network Support vector machines
88	Ch.Mallikarjunarao Dr.A.Ananda Rao	Clustering Classification
19	K. Verheyan D.Adriaens M.Hermy ,S.Deckers	Fuzzy sets

Bayesian approach improves hydro geological site characterization even when using low-resolution resistivity surveys.

K-nearest neighbor: A number of studies have been carried out on the application of data mining techniques for agricultural data sets. For example, the K-Nearest Neighbor is applied for simulating daily precipitations and other weather variables. [19]

Support Vector Machine: It is applied to future climate predictions from the second generation Coupled Global Climate Model to obtain future projections of precipitation. The results are then analyzed to assess the crash of climate change on rainfall over India. [15] It is shown that SVMs provide a promising alternative to conventional artificial neural networks for statistical downscaling, and are appropriate for conducting climate impact studies. [22]. The table 1.1 shows the methodologies used in crop yield prediction.

3.CONCLUSION

This paper has been presented research possibilities for the application of data mining methodologies or approaches to the problem of yield prediction. From the study, it has been observed that there are a huge number of applications of data mining techniques in agriculture. It is clearly known that there are several Data Mining approaches in the literature to enhance the crop productivity. For this, many researchers have implemented data mining approaches like K-Means algorithm, K Nearest Neighbor, Support Vector Machines, Neural Network, Regression techniques like liner and multiple regression techniques. It is sure that global development through the power of agricultural intelligence is possible by reducing the information gap in agriculture.

REFERENCES

1. 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.
2. Ramesh D, Vishnu Vardhan B., "Data Mining Techniques and Applications to Agricultural Yield Data", IJARCCE, Vol. 2, Issue 9, September 2013.
3. Mucherino, A., Papajorgji, P., &Pardalos, P. (2009), "Data mining in agriculture" (Vol. 34), Springer.
4. Srikant, R V Q & Agrawal, R (1997, August), "Mining Association Rules with Item Constraints. In KDD" (Vol. 97, pp. 67-73).
5. Zaki, M J (1999), "Parallel and distributed association mining: A survey". IEEE concurrency, 7(4), 14-25. 6. Beniwal, S. & Arora, J. (2012), "Classification and feature selection techniques in data mining", International Journal of Engineering Research & Technology (IJERT), 1(6).
7. LiorRokach, OdedMaimon, "Clustering Methods", Chap-15
8. Xu, R &Wunsch, D (2005), "Survey of clustering algorithms", Neural Networks, IEEE Transactions on, 16(3), 645-678.
9. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996), "From data mining to knowledge discovery in databases". AI magazine, 17(3), 37.
10. Sanjay D. Sawaitul, Prof. K.P. Wagh, Dr. P.N. Chatur, "Classification and Prediction of Future Weather by using Back Propagation AlgorithmAn Approach", International Journal of Emerging Technology and Advanced Engineering, Vol. 2, Issue 1, January 2012, pp. 110-113.
11. K. SOMVANSHI, ET AL., "MODELING AND PREDICTION OF RAINFALL USING ARTIFICIAL NEURAL NETWORK AND ARIMA TECHNIQUES", J. IND. GEOPHYS. UNION, VOL. 10, NO. 2, PP. 141-151, 2006.
12. K. Verheyen, D. Adriaens, M. Hermy, and S. Deckers, "High resolution continuous soil classification using morphological soil profile descriptions", Geoderma, vol. 101, pp. 31-48, 2001.
13. Urtubia, A., Pérez-Correa, J. R., Soto, A., &Pszczolkowski, P. (2007, "Using data mining techniques to predict industrial wine problem fermentations", Food Control,18(12), 1512-1517.
14. I. Jagielska, C. Mattehews, T. Whitfort, "An investigation into the application of neural networks, fuzzy logic, genetic algorithms, and rough sets to automated knowledge acquisition for classification problems", Neurocomputing, Vol. 24, pp. 37-54, 1999.
15. Tellaeche, A., BurgosArtizzu, X. P., Pajares, G., & Ribeiro, A. (2007), "A vision-based hybrid classifier for weeds detection in precision agriculture through the Bayesian and Fuzzy k-Means paradigms", In Innovations in Hybrid Intelligent Systems (pp. 72-79). Springer Berlin Heidelberg.
16. Veenadhari, S. 2007, "Crop productivity mapping based on decision tree and Bayesian classification". Unpublished M.Tech Thesis submitted to MakhanlalChaturvedi National University of Journalism and Communication, Bhopal.
17. Shalvi D and De Claris N., "Unsupervised neural network approach to medical data mining techniques", in Proceedings of IEEE International Joint Conference on Neural Networks, (Alaska), pp. 171-176, May 1998.
18. AltannarChinchulunn, PetrosXanthopoulos, Vera Tomaino, P.M.Pardalos, "Data Mining Techniques in Agricultural and Environmental Sciences", International Journal of Agricultural and Environmental Information Systems,1(1),26-40,January-June 2010.
19. B. Rajagopalan and U. Lal, "A K-nearest neighbor simulator for daily precipitation and other weather variable", Water Resources, vol. 35, pp. 3089-3101, 1999.
20. 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 IJCST Vol. 2, Issue 1, March 2011.
21. Tripathi, S., Srinivas, V. V., &Nanjundiah, R. S. (2006), "Downscaling of precipitation for climate change scenarios: a support vector machine approach", Journal of Hydrology, 330(3), 621-640. BIOGR
22. Mr. Abhishek B. Mankar, Mr. Mayur S. Burange, "Data Mining - An Evolutionary View of Agriculture", International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 3, Issue 3, March 2014.

23. Raorane A.A., Kulkarni R.V, "Review- Role of Data Mining in Agriculture", International Journal of Computer Science and Information Technologies, Vol. 4 (2) , 2013, 270 – 272.
24. Lansigan, et.al., "Analysis of Climatic Risk and Coping Strategies in Two Major Corn Growing Areas in the Philippines" 2010.
25.. Georg Rub, Rudolf Kruse, Peter Wagner, and Martin Schneider. "Data mining with neural networks for wheat yield prediction. In Petra Perner, editor, Advances in Data Mining (Proc. ICDM 2008)", pages 47–56, Berlin, Heidelberg, July 2008, Springer Verlag.
26. Georg Ruß Data Mining of Agricultural Yield Data: A Comparison of Regression Models, ICDM'09,. Leipzig, Germany, July 2009.
27. Roberto Benedetti A, Remo Catenaro A, Federica Piersimoni B, "GENERALIZED SOFTWARE TOOLS FOR CROP AREA ESTIMATES AND YIELD FORECAST "2010.
28. Maria Rossana C. de Leon, Eugene Rex L. Jalaok, "A Prediction Model Framework for Crop Yield Prediction", Asia Pacific Industrial Engineering and Management System, 2013.
29. A. Mucherino, A. Urtubia, Consistent Biclustering and Applications to Agriculture, IbaI Conference Proceedings, Proceedings of the Industrial Conference on Data Mining (ICDM10), Workshop "Data Mining in Agriculture" (DMA10), Berlin, Germany, 105-113, 2010.
30. A. Mucherino, S. Cafieri, A New Heuristic for Feature Selection by Consistent Biclustering, arXiv e-print, arXiv:1003.3279v1, March 2010.