ISSN No. 0976-5697

Volume 2, No. 6, Nov-Dec 2011

International Journal of Advanced Research in Computer Science

REVIEW ARTICLE

Available Online at www.ijarcs.info

Data Mining Technique for Agriculture and Related Areas

Mamta Rajshree*
Assistant Professor
School of Computer Engineering and Information,
Technology, Shobhit University, Meerut, (UP), India
mamta.rajshree@rediffmail.com

Shwetank
Assistant Professor
Department of Computer Science, Faculty of Technology
Gurukul Kangri University, Haridwar, (Utrakhandh), India
shwetank_arya@rediffmail.com

R.P.Agarwal
Professor
School of Computer Engineering and Information Technology
Shobhit University, Meerut (UP), India

Abstract: Socio-economic factors are the social and economic experiences and realities that help to change one's attitude, personality, perception and life style. Examples of socio-economic factors are education, experience, age, income, wealth, occupation etc. Farmers' socio economy is a neglected area due to the emphasis on the level of agricultural productivity. The paper deals with the role of data mining techniques in the field of agriculture and related areas. Some of the most used data mining techniques such as classification; clustering and statistical techniques along with their applications in various domains of agriculture have been discussed. Data mining techniques in the field of agricultural decision making for forecasting agricultural production, quality of production and management are new research areas. This review is conducted to explore the application of data mining in the domains of agriculture for the purpose of finding some efficient technique to improve the socio-economy of the farmers in order to improve the country economy.

Keywords: K-Means clustering, K-Nearest neighbor clustering, Support Vector Machine, Decision Tree.

I. INTRODUCTION

Data mining is the task of discovering meaningful patterns from large data set. Data mining techniques can be categorized into three main classes: statistical techniques, classification techniques and clustering techniques.

A. Statistical Techniques:

There are various statistical concepts that are the basis of data mining such as point estimation, Bayes theorem, hypothesis testing, regression and correlation and graph based methods etc.

B. Classification Techniques:

Classification technique is supervised learning. These algorithms are based on decision tree, neural network, fuzzy set theory, rough set theory, genetic algorithm, k-nearest neighbor, rules based and grid based classification etc.

C. Clustering Techniques:

A process of grouping a set of physical or abstract object into a class of similar objects is called clustering [25]. Clustering Technique is unsupervised learning. Clustering is of two types hierarchical clustering and partition clustering. For example in farmers' data set with respect to their income three data clusters can be represented in the following manner [Figure 1].

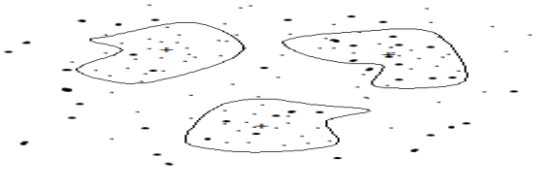

Figure 1. Farmer data with respect to their income, showing 3 data clusters. Each data cluster center is represented with +.

II. STATISTICAL TECHNIQUES

A study is carry out to assess the various Statistical classification techniques of data mining such as Bayes Bayes Net, Bayes Naive Bayes, Bayes Naive Bayes updatable, J48 and Random forest. Naïve Bayes 100% classified the instances of the data set [1]. Another study was done to compare Bayesian method, fuzzy method and hybrids approaches to extract knowledge from a soil map for classification, and resulted that the learning accuracy of the Bayesian network is lower than that of the fuzzy membership functions method, but higher than Naïve Bayes. On the basis of t-test result, under the confidence level of 95%, the Bayesian network algorithm performs significantly worse than the fuzzy method, but better than the Naïve. Bayes and the fuzzy method are significantly better than Naïve Bayes [2].

III. CLUSTERING TECHNIQUES

A. Hierarchical Clustering:

Hierarchical clustering technique on herbal plants was done, using SPSS tool, to groups the data, and to find out meaningful pattern like what type of habit of herbal plant species exist in which location of India [3]. In a study pest management framework model was developed using the combined techniques of hierarchical clustering and association rule mining [23].

B. Recursive noise Removal Techniques:

In a study Recursive Noise Removal Technique was used to find the circumstances that lead the farmers to an excessive pesticide usage [4].

C. K-Means clustering:

The k-means algorithm partitions a set of data into k clusters by finding inherent pattern in the given set. K is predetermined value. If Sj represents one of these clusters, then the following condition must be satisfied:

$$X= \bigcup_{j=1}^{K} Sj,\ \ Sj \prod Sl = 0 \quad l \le j \ne l \le k$$

Each cluster is subset of X. In this technique, the representative of a cluster is the mean of the cluster. K-Means clustering was used to analyze the agricultural meteorology for the enhancement of crop yields and the reduction of crop loses [7], classifying soils in combination with GPS-based technologies [11] and to classify soils and plants [13]. In another study k-Means method was used to forecast pollution in the atmosphere [8]. K-means approach was used to analyze color images of fruits as they run on conveyor belts [14] and X-ray images of apples were used to monitor the presence of water cores [15].

IV. CLASSIFICATION TECHNIQUE

A. Decision Tree:

In a study Decision tree was used to test if Data mining technique is capable to determine the crop yield as a function of physical-chemical soil properties, in order to correct the low yield [5]. Another study was done to simulate the farmers' decision rule in a catchment using the decision tree classification [6].

B. J4.8 Classifier:

J4.8 Classifier and wrapper method was used for mushroom grading [22].

C. K-nearest neighbor:

K-Nearest neighbor is a common classification scheme based on the use of distance measures. For classifying a new element, its distance to each item in the training set is determined and only the k closest entries in the training set are considered. The new item is then placed in the class that contains the maximum elements from this set of k closest elements [26]. K-nearest neighbor tries to classify an unknown sample based on the known classification of its neighbor [24]. K-nearest neighbor was used for simulating daily precipitation and other weather variables [9]. In another study k nearest neighbor approach was used to evaluate forest inventories and to estimate forest variables analyzing satellite imagery [19].

D. Support Vector Machine:

Support vector machine are supervised learning method used for classification [24]. Support Vector Machines technique was used to analyze the different possible changes in the weather scenarios [10]. In a study SVMs was used to classify crops [12], and for detecting weed and nitrogen stress in corn [21]. A study was done to classify the milk, where sensors were used to smell milk [17]. Data mining techniques were used in the sound recognition problems. Birds' sounds were classified using SVMs [18].

E. Neural Network:

Neural network is a set of connected input output units in which each connection has a weight associated with it. In the learning stage, network learns through adjusting its weights, so as to be able to predict the correct classification of input output [25]. For example farmers' production level is classified as high, low and medium. We have two input nodes: education and experience and we want to classify them into three classes therefore we use three output nodes. One hidden layer is assumed. Each node is labeled with function. At the input layer f1 and f2 are used which take the value of education and experience and process it as output. During processing, functions at each node are applied to the input to produce the output. The output of node f3 is processed using the following way:

f3 (w13ed + w23ex) where ed is education and ex is experience [figure 2].

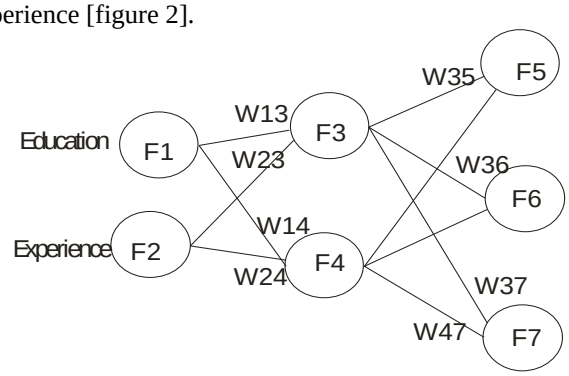

Figure 2.Neural Network for farmer data set.

In a study neural network was trained for discriminating the good and bad apples [15] and fertile and non-fertile eggs [20]. In another study to monitor the wine fermentation process, taste sensors were used to obtain data and data was classified using neural network [16].

V. CONCLUSION

Farmers' decisions with regard to agriculture production and technology use are strongly influenced by their socio-economic environment. In the agriculture the role of age, education, farming experience, land size, risk willingness etc. is important. Using the appropriate technique of data mining, important existing socio-economic pattern in the farmers' data set can be identified; to improve their socio-economic conditions consequently the agricultural productivity and quality can be improved.

VI. REFERENCES

[1] Ramesh V. and Ramar K., "Classification of Agricultural Land Soils: A Data Mining Approach", Agricultural Journal, 6(3), 82-86, 2011.

[2] Qi Feng, "Fuzzyness vs. Probability in a Data Mining Application for Soil Classification", Seventh International Conference on Fuzzy Systems and Knowledge Discovery, 2010 IEEE.

[3] Rao N.S. and Das K.S "Classification of herbal gardens in India using Data mining", Publication of Little Lion Scientific R & D", Journal of Theoretical and Applied Information Technology, vol. 25, no 2, 2011.

[4] Abdullah A. Brobst S. and Pervaiz I.,"Learning Dynamics of Pesticide Abuse through Data Mining", The Australasian Workshop on Data Mining and Web Intelligence, 2004, New Zealand.

[5] Canteri M.G., Avila B.C., Santos E.L. dos, Sanches M.K., Kovalechyn D., Molin J.P. and Gimenez L.M.," Application of Data Mining in automatic description of yield behavior in agricultural areas, World congress on computers in agriculture and natural resources.

[6] Ekasingh B., Ngamsomsuke K., Letcher R.A., Spate J., "A data mining approach to simulating farmers' crop choices for integrated water resource management", Elsevier Ltd., 2005.

[7] Rajesh D.,"Application of Spatial Data Mining for Agriculture", International Journal of Computer Application, Volume 15-No 2, February 2011.

[8] Jorquera H., Perez R., Cipriano A. and Acuna G. "Short term forcasting of Air Pollution episodes in Environmental Modeling", Zannetti, P. (ED). WIT Press, UK, 2001.

[9] Rajagopalan b, Lall U., "A K-nearest neighbor simulator for daily precipitation and other weather variables", Water Resource. Res., 1999, 35:3089-3101.

[10] Tripathi, S., V.V. Srinivas and R.S. Najundiah,"Downscaling of precipitation for climate change scenarios: A support vector machine approach", J Hydrol., 2006, 330:621-640.

[11] Veheyen, K., Adriaens, D. Harmy M. and Deckers S., "High resolution continuous soil classification using morphological soil profile descriptions", Geoderma, 2001, 101:31-48.

[12] Camps-Valls G. Gomez-Chova L. Calpe-Maravilla J., Soria-Olivas E., Martin-Guerrero JD and Moreno J., « Support Vector Machine for crop classification using hyperspectral data », Lect Notes Comp Sci. 2003, 2652:134-141

[13] Mayer et.al., 2004, " Intensified fuzzy clusters for classifying plant, soil and residue regions of interest from color images", Computers and Electronics in Agriculture, Volume 42, Issue 3, March 2004, PP 161-180

[14] Leemans V, Destain MF, "A real time grading method of apples based on features extracted from defects", J Food Eng 61: , 2004, 83-89.

[15] Shahin MA, Tollner EW and McClendon RW, "Artificial Intelligence classifiers for sorting apples based on water core", J Agric Eng Res 79(3): 2001, 265-274.

[16] Riul A Jr. de Sousa HC Malmegrim RR, dos Santos DS Jr, Carvalho ACPLF, Fonseca FJ, Oliverira Jr ON and Mattoso LHC, « Wine classification by taste sensors made from ultra-thin films and using Neural Networks, Sens Actuators B98, 2004, :77-82.

[17] Brudzewski K. Osowski S. and Markiewiez T.," Classification of milk by means of an electronic nose and SVM neural network, Sens Actuators B98, 2004, PP 291-298.

[18] Fagerlund S, "Bird species recognition using support vector machines. EURASIP J Adv Signal Processing, 2007, Article ID 38637, 8p.

[19] Holmgren P. Thuresson T, "Satellite remote sensing for forestry planning: a review. Scand J for Res, 1998, 13(1):90-110.

[20] Das and Evans, "Detecting fertility of hatching eggs using machine vision H: Neural Network classifiers. Trans ASAE, 1992, 35(6):2035-2041.

[21] Du C-J, Sun D-W, "Pizza sauce spread classification using color vision and support vector machines", J Food Eng , 2005, 66: 137-145

[22] Cunningham, S.J. and G. Holmes, 1999. "Developing innovative applications in agriculture using data mining", Proceedings of the Southeast Asia Regional.

[23] Tripathy A.K.,Adinarayana J. and Sudharsan D., "Geospatial Data Mining for Agriculture Pest Management – a Framework", Geoinformatics, 2009.pp 1-6.

[24] Mucherino A., Papajorgji P.J. and Pardalos P.M., "Data Mining in Agriculture", Springer.

[25] Kamber M. and Han Jiawei, "Data Mining Concepts and Techniques, II edition, Elsevier.

[26] Dunham M.H., "Data Mining Introductory and Advanced topics", Pearson Education, 6th impression 2009.

