International Journal of Business Intelligent
Volume: 01 Issue: 02 December 2012,Pages No.40-42
ISSN: 2278-2400

Applying Data Mining Techniques in the Field of Agriculture and Allied Sciences

N G Yethiraj
Assistant Professor, Department of Computer Science,Maharani's Science College for Women, Bangalore, India

Abstract-In this paper an attempt has been made to review the research studies on application of data mining techniques in the field of agriculture. Some of the techniques, such asID3 algorithms, the k-means, the k nearest neighbour, artificial neural networks 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 / predicting of agricultural crop/animal management. This article explores the applications of data mining techniques in the field of agriculture and allied sciences.

Keywords-Data mining, K-means algorithm, crop productivity, ID3 algorithm, rough sets, k nearest neighbour.

I. INTRODUCTION

The major reason that data mining has attracted a great deal of attention in information industry in recent years is due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge. The information and knowledge gained can be used for applications ranging from business management, production control, and market analysis, to engineering design and science exploration.Data mining can be viewed as a result of the natural evolution of information technology. An evolutionary path has been witnessed in the database industry in the development of the following functionalities: data collection and database creation, data management (including data storage and retrieval, and database transaction processing), and data analysis and understanding (involving data warehousing and data mining). For instance, the early development of data collection and database creation mechanisms served as a prerequisite for later development of effective mechanisms for data storage and retrieval, and query and transaction processing. Become the next target systems opening query and transaction processing as common practice, data analysis and understanding has naturallyData mining is the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviours, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analysis offered by data mining move beyond the analysis of past events provided by retrospective tools typical of decision support systems. Agriculture and allied activities constitute the single largest component of India's gross domestic product, contributing nearly 25% of the total and nearly 60% of Indian population depends on this profession. Due to vagaries of climate factors the agricultural productivities in India are continuously decreasing over a decade. The reasons for this were studied mostly using regression analysis. In this paper an attempt has been made to compile the research findings of different researchers who used data.

II. APPLICATION OF DATA MINING TECHNIQUES IN AGRICULTURE

Many techniques for learning rules and relationships automatically from diverse data sets were developed [14], to simplify the often tedious and error-prone process of acquiring knowledge from empirical data. While these techniques are plausible, theoretically well-founded, and perform well on more or less artificial test data sets, they depend on their ability to make sense of real-world data. This paper describes a project that is applying a range of machine learning strategies to problems in agriculture and horticulture. They briefly surveyed some of the techniques emerging from machine learning research, describe a software workbench for experimenting with a variety of techniques on real-world data sets, and describe a case study of dairy herd management in which culling rules were inferred from a medium-sized database of herd information. They also described a range of machine learning strategies to problems in agriculture and horticulture. They briefly surveyed some of the techniques emerging from machine learning research, described a software workbench for experimenting with a variety of techniques on real-world data sets, and described a case study of dairy herd management in which culling rules were inferred from a medium-sized database of herd information.

In recent years, several models for the simulation of soil dynamics have been developed. Some examples are DSSAT [11], CROPSYST [17], and GLEAMS [13], to name a few. Such models are able to simulate the dynamics in a soil, using some soil parameters that need to be specified. Three are the most used parameters, referred to as LL, DUL, and PEWS. LL is the lower limit of plant water availability; DUL is the drained upper limit; PESW is the plant extractable soil water. Unfortunately, these parameters are usually unknown. The available information about the soils usually regards their texture, such as the percentage of clay, silt, sand and organic carbon in the soil. While the texture of a soil is usually known, the LL, DUL and PEWS parameters are difficult to estimate.
Data mining techniques are often used to study soil characteristics. As an example, the k-means approach is used for classifying soils in combination with GPS-based technologies [19], k-means approach [15] to classify soils and plants and SVMs to classify crops [6].Independent component analysis techniques for mining spatio-temporal data has been applied to mine for patterns in weather data using the North Atlantic Oscillation (NAO) as a specific example [5]. They

International Journal of Business Intelligent
Volume: 01 Issue: 02 December 2012,Pages No.40-42
ISSN: 2278-2400

found that the strongest independent components match the observed synoptic weather patterns corresponding to the NAO. They validated the results by matching the independent component activities with the NAO index.Analytical exploration of vast amount of agricultural data can best be supported by an appropriate application. [1] applied data warehousing and Online Analytical Processing (OLAP) technologies for appropriate utility of agricultural data. A data warehouse provides a flexible yet efficient and reliable storage structure for vast amount of data while OLAP techniques provide mechanisms for ad hoc and in depth analysis of this data. Traditional analytical tools and database techniques may not succeed here due their rigid nature. Techniques used in their work are equally applicable at any geographic location provided that related data is available.

A case study of interpreting paddy distributions of three counties on Northern Taiwan during two crop seasons on year 2000 using multi-temporal imageries together with cadastre GIS by Bayesian posteriori probability classifier was studied [7]. In order to integrating Bayesian conditional probability, priori probabilities of paddy's attributes were estimated from photogrammettric interpretation results provided by the Food Bureau, and the spectrum reflectance from different growth stages was used. Due to the spatial heterogenous of paddy's distribution, classifier parameters were established individually on each map-quadrangle. Temporal change of NDVI from different growth stages pass through rice's life cycle has been measured and we find two-stage images make significant improvement on classification results. Results of the study help us to evaluate the accuracy of the classifier. Imagery classification results were compared with aerial photo's interpreting results for assessing accuracy. Overall accuracy of first crop of Tao-yuan, Hsin-chu, and Miao-li were 89.93% 92.83% 95.33% respectively. Bayesin classifier has advantages including easy-to-adjusted and easy-to-computed rules and comparative stable results when limited SPOT satellite imageries available. Bayesin method also provides results with probability that help the operator to assess the places having least confidence. These advantages allow us to suggest Bayesian method be used in paddy-area investigation in Taiwan.Studies conducted by agricultural researchers in Pakistan have shown that attempts of crop yield maximization through propesticide state policies have led to a dangerously high pesticide usage. These studies have reported a negative correlation between pesticide usage and crop yield. Hence excessive use of pesticides is harming the farmers with adverse financial, environmental and social impacts. Study [2] had shown that how data mining integrated agricultural data including pest scouting, pesticide usage and meteorological recordings is useful for optimization of pesticide usage. Unsupervised clustering of the data was performed first through Recursive Noise Removal (RNR). These clusters reveal interesting patterns of farmer practices along with pesticide usage dynamics and hence help identify the reasons for this pesticide abuse

A mechanism of performing the mapping from nominal to numeric values(actually ranking) based on the transmittance as well as the statistical properties of the plants was proposed [3]. Spectral analysis (using chemical means) is a tedious and time consuming process, thus difficult to repeat, each and every time, for classification of (numerically) unclassified cotton varieties. A supporting statistical method was also proposed based on linear regression curve fitting using normalized nominal attributes. Subsequently a rank is assigned to the variety based on its R2 value and slope of the plot. This rank thus becomes the numeric equivalent of the nominal alphanumeric name of the variety being considered. Spectral analysis of 12 cotton varieties in the visible and near infra red regions was also performed. A 60% classification accuracy was achieved i.e., correspondence in order of ranking generated through regression, as compared with the spectral rank order.A process model for analyzing data, and describes the support that Weka to Environment for Knowledge Analysis (WEKA) provides for this model [8]. The domain model learned by thedata mining algorithm can then be readily incorporated into a software application. This WEKA based analysis and application construction process was illustrated through a case study in the agricultural domain i.e., in mushroom grading.Effect of pesticides on humans can't be directly checked because of the poisonous nature of pesticides, therefore the usage of pesticides on cotton crop has been taken [16] into consideration for the purpose. The COF Clustering Tool cannot only be used for pesticide data, but also possesses the flexibility to deal with any numeric data.

Spatial data mining methods to extract interesting and regular knowledge from large spatial databases of agriculture were studied [12] aiming at discerning trends in agriculture production with reference to the availability of inputs. The predicted and real vs. Counter graph illustrated how closely the poly analyst prediction follows the actual value of the attribute over the range of the dataset. Applying the data mining techniques to agriculture the target for different food grains can be achieved. Their study demonstrated the scope for application of spatial mining tools for a utility study and analysis. The specific application of Polyanalyst gave a clear scope for evaluation and comparison of predicted and real values.Influence of climatic factors on major kharif and rabi crops production in Bhopal District of Madhya Pradesh State was studied [18]. The findings of the study revealed that the decision tree analysis indicated that the productivity of soybean crop was mostly influenced by Relative humidity followed by rainfall and temperature. The decision tree analysis indicated that the productivity of paddy crop was mostly influenced by Rainfall followed by Relative humidity and Evaporation. For Wheat crop the analysis indicated that the productivity is mostly influenced by Temperature followed by Relative humidity and Rainfall. The findings of decision tree were confirmed from Bayesian classification. The decision tree in the study area fast to execute and much to be desired as representations of knowledge interpretations. The rules formed from the decision tree are helpful in identifying the conditions responsible for the high or low crop productivity.

Powdery Mildew of Mango a devastating disease of mango was predicted [9] using Decision Tree induction, Rough Sets (RS) and hybridized Rough Set based Decision Tree Induction (RDT) in comparison with the standard Logistic Regression (LR) method. The induction algorithms shown better performance over logistic regression.

International Journal of Business Intelligent
Volume: 01 Issue: 02 December 2012,Pages No.40-42
ISSN: 2278-2400

A web based expert information system based on ID3 algorithm was studied [4] in which an expert system provides advisory services to Tomato growers regarding pests, diseases and their control measures. The web based system has also provision for the growers to interact with other growers on the management practices of tomato crop cultivation.An advanced version of decision-making tree algorithm IBLE that it mainly uses in the information theory [20]. The channel capacity concept to take chooses the important characteristic to the entity in the measure. Combines the rule with many characteristics the point to distinguish the example can effectively the correct distinction. They applied this algorithm in the oral cavity disease diagnosis, the experimental result indicated this algorithm has the very strong recognition capability to agriculture case diagnosis to very good assistance diagnosis function.The application of information technology in agriculture accelerates the digitization of agriculture information [10] presented a new improved CA algorithm based on traditional decision tree method. It introduces a pre-treatment theory about double dimension reduction which can deal with large and high-scale datasets. By using CA algorithm in maize seed breeding, they analyzed the potential rules and found out useful information from it for direct growth of maize. Their experiment showed the improved CA algorithm can obtain more intuitive and efficient information.

III. CONCLUSIONS

There is a growing number of applications of data mining techniques in agriculture and a growing amount of data that are currently available from many resources. This is relatively a novel research field and it is expected to grow in the future. There is a lot of work to be done on this emerging and interesting research field. The multidisciplinary approach of integrating computer science with agriculture will help in forecasting/managing agricultural crops effectively.

REFERENCES

[1] Jiawei and Micheline Kamber. Simon Fraser University “Data Mining Concepts & Techniques” 2000.

[2] Abdullah, A., Brobst, S., M.Umer M. 2004. "The case for an agri data ware house: Enabling analytical exploration of integrated agricultural data". Proc. of IASTED International Conference on Databases and Applications. Austria. Feb

[3] Abdullah, A., Brobst, S, Pervaiz.I., Umer M.,A.Nisar. 2004. "Learning dynamics of pesticide abuse through data mining". Proc. of Australian Workshop on Data Mining and Web Intelligence, New Zealand, January.

[4] Using Data Mining to Discover Patterns in Autonomic Storage Systems. Zhenmin Li, Sudarshan M. Srinivasan, Zhifeng Chen, Yuanyuan Zhou, Peter Tzvetkov, Xifeng Yan, and Jiawei Han. 1st Workshop on Algorithms and Architectures for Self- Managing Systems in conjunction with ISCA and SIGMETRICS, June 2003.

[5] Abdullah, A., Bulbul.R., Tahir Mehmood. 2005. "Mapping nominal values to numbers by data mining spectral properties of leaves". Proc. of 3rd International Symposiumm on Intelligent Information Technology in Agriculture. Beijing, China. Oct, 2005.

[6] Babu, MSP., Ramana Murthy, NV, SVNL Narayana, 2010. "A web based tomato crop expert information system based on artificial intelligence and machine learning algorithms". Int. J. of Comp. Sci., and Information Technologies. Vol. 1(1) . pp. 6-15.

[7] Basak J., Sudharshan, A., Trivedi D., M.S.Santhanam. 2004. "Weather Data Mining Using Independent Component Analysis". J. of Machine Learning Research 5: pp. 239-253.

[8] Camps-Valls G, Gomez-Chova L, Calpe-Maravilla J, Soria-Olivas E, Martin-Guerrero JD, Moreno J., 2003, "Support vector machines for crop classification using hyperspectral data". Lect Notes Comp Sci 2652: pp. 134–141

[9] Chi-Chung LAU, Kuo-Hsin HSIAO, 2005. "Bayesian Classification For Rice Paddy interpretation". Paper presented in Conference on data mining held at China Tapei. December, 2005

[10] Cunningham S.J., G. Holmes. 2005. "Developing innovative applications in agriculture using data mining". Proc. Of 3rd International Symposium on Intelligent Information Technology in Agriculture. Beijing, China. Oct, 2005.

[11] Jain Rajni, Minz, S., V. Rama Subramaniam. 2009. "Machine learning for forewarning crop diseases". J. Ind. Soc. Agri. Stat. 63(1): pp. 97-107.

[12] Jianlin Ji Dan, Qiu Chen, Jianping Chen, Li He Peng , 2010. "An improved decision tree algorithm and its application in maize seed breeding". Sixth Internation Conference on Natural Computation, held at Yantai, Shandon 10-12th January. pp. 117-121.

[13] Jones JW, Tsuji GY, Hoogenboom G, Hunt LA, Thornton PK, Wilkens PW, Imamura DT, Bowen WT, Singh U., (1998), "Decision support system for agrotechnology transfer: DSSAT v3". In: Tsuji GY, Hoogenboom G, Thornton PK (eds) , "Understanding options for agricultural production". Kluwer Academic Publishers, Dordrecht, pp 157–177

[14] Kiran Mai, C., Murali Krishna, I.V., A.Venugopal Reddy, 2006. "Data Mining of Geo-spatial Database for Agriculture Related Application". Proc. of Map India. New Delhi.

[15] Leonard RA, Knisel WG, Still DA., 1987, GLEAMS: groundwater-loading effects of agricultural management systems. Trans Am Soc Agric Eng 30(5): pp. 1403–1418

[16] McQueen Robert J, Garner S.R.,Nevill-Manning C.G. , Ian H. Witten, 1995. "Applying machine learning to agricultural data". Comptuers and Electronics in Agriculture. Vol. 12:pp. 275-293.

[17] Meyer GE, Neto JC, Jones DD, Hindman TW, 2004, "Intensified fuzzy clusters for classifying plant, soil, and residue regions of interest from color images". Computer Electronics Agric Vol. 42: pp. 161–180.

[18] Rabia Imitiaz, Malik Sikandar Hayat Khiyal, Shahid Khalil , Ahsan Abdullah, 2005, "Effect of pesticides on human life through visual data mining". Journal of Theoretical and Applied Information Technology. pp. 104-109.

[19] Stockle CO, Martin SA, Campbell GS, 1994, "CropSyst, a cropping systems model: water/nitrogen budgets and crop yield". Agric Syst Vol. 46(3): pp. 335–359.

[20] Verheyen K, Adriaens D, Hermy M, Deckers S., 2001, "High-resolution continuous soil classification using morphological soil profile descriptions". Geoderma Vol. 101: pp. 31–48

[21] Yue Jin Hai, Song Kai, 2010. "IBLE Algorithm in agricultural disease diagnosis". In third International Conference on Intelligent Networks and Intelligent Systems held at Shenyang, Liaoning China during November 01-November 2003.

[22] Olivia Parr Rud : “Data Mining, Modeling data for marketing risk, and Customer Relationship Management”, Wiley Publications 2003.