



Data Mining and Agribusiness

Dr. Rahul G. Thakkar

Assistant Professor, ASPEE Agribusiness Management Institute, Navsari Agricultural University, Navsari

Prof. Bhavesh D. Patel

Assistant Professor, ASPEE Agribusiness Management Institute, Navsari Agricultural University, Navsari

Dr. Bankim L. Radadiya

Director IT, Navsari Agricultural University, Navsari

ABSTRACT

Data mining in agriculture is a very recent research topic. It consists in the application of data mining techniques to agriculture. Recent technologies are nowadays able to provide a lot of information on agricultural-related activities, which can then be analyzed in order to find important information. A related, but not equivalent term is precision agriculture. The wide availability of huge amounts of agriculture data has generated an urgent need for the research of data mining. Generating rules with higher accuracy for Agriculture databases can be done using different techniques of data mining. As the analysis of agriculture dataset is usually a complex work.

KEYWORDS : Association Rule Mining, Decision Tree and Agriculture

II. INTRODUCTION

Data Mining provides the Enterprise with intelligence. Using records of data mining technique gives some economical beneficial information to the organization. To best apply these advanced techniques, they must be fully integrated with a data warehouse as well as flexible interactive business analysis tools. Many data mining tools currently operate outside of the warehouse, requiring extra steps for extracting, importing, and analyzing the data. Furthermore, when new insights require operational implementation, integration with the warehouse simplifies the application of results from data mining. An OLAP (On-Line Analytical Processing) server enables a more sophisticated end-user business model to be applied when navigating the data warehouse. The multidimensional structures allow the user to analyze the data as they want to view their business – summarizing by product line, region, and other key perspectives of their business. The Data Mining Server must be integrated with the data warehouse and the OLAP server to embed ROI-focused business analysis directly into this infrastructure. An advanced, process-centric metadata template defines the data mining objectives for specific business issues like campaign management, prospecting, and promotion optimization. Integration with the data warehouse enables operational decisions to be directly implemented and tracked. As the warehouse grows with new decisions and results, the organization can continually mine the best practices and apply them to future decisions.

This design represents a fundamental shift from conventional decision support systems. Rather than simply delivering data to the end user through query and reporting software, the Advanced Analysis Server applies users' business models directly to the warehouse and returns a proactive analysis of the most relevant information. These results enhance the metadata in the OLAP Server by providing a dynamic metadata layer that represents a distilled view of the data. Reporting, visualization, and other analysis tools can then be applied to plan future actions and confirm the impact of those plans.

III. TECHNIQUES USED IN DATA MINING FOR AGRI-BUSINESS

- **Artificial neural networks:** Non-linear predictive models that learn through training and resemble biological neural networks in structure.
- **Association Rules:** through the Supports and confidence associate the different products for the economical benefit.
- **Decision trees:** Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and

Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID).

- **Genetic algorithms:** Optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution.
- **Nearest neighbor method:** A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where $k \geq 1$). Sometimes called the k -nearest neighbor technique.
- **Rule induction:** The extraction of useful if-then rules from data based on statistical significance.

IV. SOME OF SIGNIFICANT WORKS DONE IN AGRICULTURE

Recent studies by agriculture researchers in Pakistan (one of the top four cotton producers of the world) showed that attempts of cotton crop yield maximization through pro-pesticide state policies have led to a dangerously high pesticide use.

- To monitor cotton growth, different government departments and agencies in Pakistan have been recording pest scouting, agriculture and metrological data for decades.
- Cotton pest scouting data recorded stands at around 1.5 million records
- Pesticides sprayed at the wrong time, Right pesticides used for the right reasons and relationship between pesticide usage and day of the week

On Line Analytical Processing (OLAP)

- OLAP is an approach to swiftly answer multi-dimensional analytical queries. Data mining is a part of OLAP with application such as forecasting or prediction in agriculture.
- It provides an opportunity of viewing agriculture data from different points of view to better understand what that data means OLAP has been used extensively for analysis of Soil physical characteristics.
- The recent advances in data base technology and data warehouses, the multi dimensional data base, OLAP and data mining technologies are being successfully applied to the management of Agriculture resources.

V. WEKA SOFTWARE FOR DATA MINING

Weka supports several standard data mining tasks, more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection. All of Weka's techniques are predicated on the

assumption that the data is available as a single flat file or relation, where each data point is described by a fixed number of attributes (normally, numeric or nominal attributes, but some other attribute types are also supported). Weka provides access to SQL databases using Java Database Connectivity and can process the result returned by a database query. It is not capable of multi-relational data mining, but there is separate software for converting a collection of linked database tables into a single table that is suitable for processing using Weka. Another important area that is currently not covered by the algorithms included in the Weka distribution is sequence modeling.

- Machine learning/data mining software written in Java, Used for research, education, and applications .
- Its developed by University of waikato, New Zealand.
- Open Source Software.
- Main features: Comprehensive set of data pre-processing tools, learning algorithms and evaluation methods.
- Graphical user interfaces (incl. data visualization).

VI. OTHER AVAILABLE SOFTWARE TOOLS FOR DATA MINING

- **DataDetective:-** The powerful yet easy to use data mining platform and the crime analysis software of choice for the Dutch police.
- **DataLab (more focus on marketing):-** A complete and powerful data mining tool with a unique data exploration process, with a focus on marketing and interoperability with SAS.
- **GhostMiner:-** Complete data mining suite, including k-nearest neighbors, neural nets, decision tree, neurofuzzy, SVM, PCA, clustering, and visualization.
- **WITNESS Miner:-** A graphical data mining tool with decision trees, clustering, discretisation, feature subset selection, and more.
- **Advanced Miner Professional:-** Provides a wide range of tools for data transformations, Data Mining models, data analysis and reporting.

- **Angoss Knowledge Studio:-** A comprehensive suite of data mining and predictive modeling tools; interoperability with SAS and other major statistical tools.

GainSmarts:- Uses predictive modeling technology that can analyze past purchase, demographic, and lifestyle data , to predict the likelihood of response and develop an understanding of consumer characteristics.

XLMiner:- Data Mining Add-In For Excel.

VII. 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 purpose.

Data mining has achieved tremendous success. Many new problems have emerged and have been solved by data mining researchers. However, there is still a lack of timely exchange of important topics in the community as a whole.

Agriculture has traditionally been maintained by families and communities where the passing on and sharing of knowledge is regarded as very important. The accumulation and sharing of knowledge has resulted in better overall efficiency and productivity. Agriculture is the embodiment of a large amount of ancient knowledge.

REFERENCES

- [1] Abdullah, A. and Brobst, S. (2003): Clustering by recursive noise removal. Proc. Atlantic Symposium on Computational Biology and Genome Informatics, USA, pp. 973-977. | [2] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, pages 144–152. ACM Press, 1992. | [3] Jiawei Han
- Micheline Kamber, Data Mining Concepts And techniques. | [4] Larry Greenfield's Data Warehouse Information Center <http://pwp.starnetinc.com/larryg/> | [5] Larry Greenfield's Data Warehouse Information Center <http://pwp.starnetinc.com/larryg/> | [6] <http://www.olapcouncil.com/> | [7] en.wikipedia.org | [8] www.ijabpt.com/pdf/47029-Lakshmi%20Velide.pdf | [9] storm.cis.fordham.edu/~gweiss/data-mining/weka-explorer-tutorial.pdf | [10] www.slideshare.net/butest/microsoft-powerpoint-weka-readonly | [11] www.kdnuggets.com/software/suites.html | [12] ijirce.com/upload/2014/nscl14/21_es03.pdf | [13] www.ijcst.com/vol21/veenadhari.pdf | [14] www.rmmnaidu.com | [15] www.academia.edu/8068006/Agricultural_Development_in_IRAN_Base_on_Cloud_Computing_Theory | [16] www.thearling.com/text/dmwhite/dmwhite.htm | [17] www.emory.edu/BUSINESS/et/datamine/ | [18] www.ijird.com/index.php/ijird/article/viewFile/36260/29381 | [19] www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.html |