Application and Scope of Data Mining in Agriculture

Kaushik Bhagawati, Amit Sen, Kshitiz Kumar Shukla, Rupankar Bhagawati

Indian Council of Agricultural Research, ICAR Research Complex for NEH Region, Arunachal Pradesh Centre, Basar, Arunchal Pradesh, India

Abstract— *Making agriculture sustainable and resilient to the ongoing change in climate and social structure is a major challenge for the scientists and researchers across the globe. Agricultural system demands transition and a multidisciplinary approach. Intelligent and precision agricultural approaches were given due importance for increasing production and productivity from the very same limited resources. The approach needs information from various sources and efficient use of them in relevant field. This need lead to growing interest in knowledge discovery from vast piles of data generated out of various research and survey works. The emergence of Data Mining techniques revolutionized the field of information generation and pattern recognition. Though Data Mining is an emerging science, it finds a wide application in agriculture and allied sectors, and has a wide future prospect.*

Keywords— Data Mining, Agriculture, Sustainability, Information, Knowledge Discovery.

I. INTRODUCTION

Agriculture is experiencing a transition stage driven by population pressure and climate change. More production and productivity are being expected from limited resources. New intensive research is being done to explore ways to increase production with optimum use of resources maintaining sustainability. This lead to the use of modern sophisticated computer assisted technologies in agricultural research. Due to widespread use of computer and affordable storage facilities, there is an enormous wealth of data embedded in huge databases of different agri-allied enterprises. The process leads to the generation of megabytes, gigabytes, and terabytes of data that are piling up in the electronic vaults of different organizations, institutions and companies. With the advent of internet and World Wide Web, the accessibility of data increase thousand times. These massive databases coevolving with new research methodologies in agriculture with widespread application of information technology could be a precious repository of information for the decision makers, right from policy makers and researchers to the farmers. Beside the four economic factor of production *viz.* Land, Labor, Capital and Entrepreneurial ability, '*information*' emerges as fifth vital factor [1]. Particularly in agriculture and allied sectors, the role of information is ever increasing. Information regarding weather, soil, disease, insects, seed, fertilizer, market *etc*. becomes important input for economic and sustainable development of these sectors.

The tremendous amounts of data generated out of these processes have unexplored potential for improving the efficiency of the related sectors. We need techniques and technologies to derive information and knowledge from such gigantic data set. In the domain of scientific computing, the major problem is to infer valuable information from observed data, especially from those areas that generate enormous amount of data each day, like satellite remote sensing. No much attention was given on these huge repositories of information until 1990s due to lack of efficient methods and techniques and these vast accumulated data during processes of daily activities was dumped in archival files.

The recent advancement in analysis techniques and advent of faster analysis tools that can help filter and analyze the stockpiles of data, turning up valuable and often surprising information attracted the attention of researchers across the globe. The fact leads to the evolution of a new discipline in computer science – Data Mining, which involves exploration and analysis of large data sets, in order to discover meaningful patterns and rules [2]. The major focus and intension are to use existing data to invent new facts and to uncover new relationships previously unknown even to the experts. The basic idea is –"when same data analyzed in different context, new context based information and knowledge is generated". Data mining takes the evolutionary process beyond retrospective data to access and navigation to prospective and proactive information

delivery and is supported by three technologies: *massive data collection*, *high performance computing* and *data mining algorithms*. Data mining involves multi-disciplinary approaches and is component of wider process called knowledge discovery from databases. It draws work from areas including database technology, machine learning, statistics, pattern recognition, information retrieval, neural networks, knowledge-based systems, artificial intelligence, and high performance computing and data visualization. It presents techniques for discovery of patterns hidden in large data sets, focusing on issues relating to their feasibility, usefulness, effectiveness, and scalability.

The objective of this review is to introduce briefly the techniques of Data Mining and to outline its use in agriculture and allied sectors. Being much more superior to the conventional data analysis techniques used in agricultural research, data mining can open a new avenue for research and development in agriculture and associated ventures.

II. TASK OF DATA MINING

The main tasks of data mining include Classification, Estimation, Prediction, Association rules, Clustering, and Description & Visualization. Among these tasks, the first three – Classification, Estimation and Prediction are all examples of directed data mining or supervised learning in which the target is known in advance and the aim is to describe the target attribute(s) in terms of rest of the attributes. The next three tasks – Association rules, Clustering and Description are examples of undirected data mining in which new relationship is aimed to establish among the attributes.

III. DATA MINING TECHNIQUES

Data mining is an emerging science and new techniques are developing specific to task. The techniques of data mining are broadly classified under following subjects - Statistics, Machine Learning, Fuzzy Logic and Rough sets techniques. The elaboration and mathematics of techniques are available in standard Data Mining text books.

IV. APPLICATION OF DATA MINING TECHNIQUES IN AGRICULTURE

Data mining techniques find wide application in agriculture and allied sectors. Lee *et al.*[3] used the knowledge discovery life cycle (KDLC) model for study dealing with crop yield and visualization using Geographic Information System. In the study, the significance of the multi-strategy knowledge discovery and visualization process in analyzing the classifications and learned rules has been empirically verified in KDLC. Bajwa *et al* [4] has evaluated five methods of band selection (three unsupervised and two supervised) were compared for selecting signature bands from hyperspectral imagery for characterizing soil EC_a and canopy density in agricultural fields. In main focus of the work of [5], was to study the functioning model of a detritic aquifer undergoing overexploitation and nitrate excess input coming from strawberry and citrus intensive crops in its recharge zone using the Data Mining technique of Fuzzy Logic. By studying the large dataset, several simulation models of soil dynamics have been developed *viz.* DSSAT [6], CROPSYST [7] and GLEAMS [8] to name a few. Jain *et al* [6] presents the potential of three machine-learning techniques *viz.* Decision Tree induction using C4.5, Rough Sets and hybridized rough set based decision tree induction over the traditional regression analysis for disease forecasting. The forecasting based on machine-learning was found to be accurate compared to the traditional methods. In their studies using fuzzy logic, Meyer et al [7] found that fuzzy techniques using ZGK algorithm could be potentially useful for remote sensing, mapping, crop management, weed, and pest control for precision agriculture. Data mining played vital role in classification of soil. Stockle *et al* [9] used technique of K-means Approaches for classifying soils in combination with GPS-based technologies. A technique like Support Vector Machine (SVM) was found to have important application in classification of crops [10] and yield prediction [11]. The yield prediction is further tuned by advanced recent concept of spatial autocorrelation [12]. Artificial Neural Network (ANN) plays very important role in development of precise forecasting and forewarning models of plant diseases [13]. To identify patterns of weather data, the technique like Independent Component Analysis found to be very effective [14]. Integration of agricultural data that includes pest scouting, pesticide usage and meteorological recording is found to be useful for optimization of pesticide usages [15]. Automatic Data Mining techniques have been recently used for recognizing and grading fruits [16]. In China, the relation between climate change, water resources and agriculture was undertaken using the technique of Data Mining [17]. Data mining is recognized as the most advance concept for prediction of market fluctuation and price variability. Ding *et al* [18] used the technique of Decision Tree for prediction of market price of pig in China. It also finds application in prediction of food borne disease outbreaks [19] and the forecast of water consumption in agriculture [20]. Fuzzy set and interpolation techniques are

applied for land suitability evaluation for maize in Northern Ghana [21]. Thus Data Mining has proved to have surprisingly broad application in every issue related to agriculture and allied field.

V. CONCLUSION

Economic and efficient use of resources requires timely and sophisticated analysis on an integrated view of the data. The growing gap between more powerful storage systems and the users' ability to effectively analyze and act on the information they contain is being minimized by the techniques of data mining. Data mining has importance regarding pattern recognition, forecasting, discovery of knowledge *etc.*, in different business domains. Data mining has wide application domain almost in every industry where the data is generated that's why data mining is considered one of the most important frontiers in database and information systems and one of the most promising interdisciplinary developments in Information Technology. The technology and its application in Agri-allied sector are still in its initial stage and have tremendous future prospect.

REFERENCES

[1] K. Bhagawati, P.K. Malhotra, R.N. Sahoo, "GIS Assisted Farm Management Information System". Lap Lambert Academic Publishing, Germany, pp.148, 2012.

[2] A.K. Pujari, "Data Mining Techniques", Universities Press, pp-288, 2001.

[3] S. W. Lee, L. Kerschberg, "Methodology and life cycle model for data mining and knowledge discovery in precision agriculture", Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Vol. **3**, pp. 2882–2887,1998.

[4] S. G. Bajwa, P. Bajcsy, P. Groves, L.F. Tian, "Hyperspectral image data mining for band selection in agricultural applications", Transactions of the American Society of Agricultural Engineers, Vol. 47, Issue 3, 895–907, 2004.

[5] J.M, Andujar, J. Aroba, *et al*., "Contrast of evolution models for agricultural contaminants in ground waters by means of fuzzy logic and data mining" Environmental Geology, Vol. 49, Issue 3, pp. 458–466, 2006.

[6] R. Jain, S. Minz, V. Rama Subramaniam, "Machine learning for forewarning crop diseases", J. Ind. Soc. Agri. Stat., Vol. 63, Issue 1, pp. 97-107, 2009.

[7] G. E. Meyer, J.C. Neto, D.D. Jones, T.W. Hindman, "Intensified fuzzy clusters for classifying plant, soil, and residue regions of interest from color images". Computer Electronics Agric, Vol. 42, pp. 161–180, 2004.

[8] J.W. Jones, G.Y. Tsuji, G. Hoogenboom, L.A. Hunt, P.K. Thornton, P.W. Wilkens, D.T. Imamura, W.T. Bowen, U. Singh, "Decision support system for agrotechnology transfer: DSSAT v3". In: G.Y. Tsuji, G. Hoogenboom, P.K. Thornton (eds), "Understanding options for agricultural production", Kluwer Academic Publishers, Dordrecht, pp. 157–177, 1998.

[9] C.O. Stockle, S.A. Martin, G.S. Campbell, "CropSyst, a cropping systems model: water/nitrogen budgets and crop yield", Agric Syst, Vol. 46, Issue 3, pp. 335–359, 1994.

[10] M.S.P. Babu, N.V. Ramana Murthy, S.V.N.L. Narayana, "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, Issue 1, pp. 6-15, 2010.

[11] G. Ruß, "Data Mining of Agricultural Yield Data: A Comparison of Regression Models", Conference Proceedings "Advances in Data Mining – Applications and Theoretical Aspects", P. Perner (Ed.), Lecture Notes in Artificial Intelligence 6171, Berlin, Heidelberg, pp. 24–37, Springer, 2009.

[12] G. Ruß, A. Brenning, "Data Mining in Precision Agriculture: Management of Spatial Information", Conference Proceedings, "Computational Intelligence for Knowledge- Based Systems Design", E. H. ullermeier, R. Kruse, and F. Hoffmann (Eds.), Lecture Notes in Artificial Intelligence 6178, Berlin, Heidelberg, pp. 350–359, Springer, 2010.

[13] R. Bhagawati, K. Bhagawati, A.K.K. Singh, R. Nongthombam, R. Sarmah, G. Bhagawati, "Artificial Neural Network Assisted Weather Based Plant Disease Forecasting System". International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 3, Issue 6, pp. 4168-4173, 2015.

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

[15] A. Abdullah, S. Brobst, I. Pervaiz, M. Umer, A. Nisar, "Learning dynamics of pesticide abuse through data mining", Proc. of Australian Workshop

on Data Mining and Web Intelligence, New Zealand, January, 2004.

[16] D. Unay, B. Gosselin, O. Kleynen, V. Leemans, M.F. Destain, O. Debeir, "Automatic Grading of Bi-colored Apples by Multispectral Machine Vision", Computers and Electronics in Agriculture, Vol. **75,** Issue 1, pp. 204–212, 2011.

[17] S. Piao, P. Ciais, Y. Huang, Z. Shen, S. Peng, J. Li, L. Zhou, H. Liu, Y. Ma, Y. Ding, P. Friedlingstein, C. Liu, K. Tan, Y. Yu, T. Zhang, J. Fang, "The Impacts of Climate Change on Water Resources and Agriculture in China", Nature, Vol. 467, pp. 43–51, 2010.

[18] L. Ding, J. Meng, Z. Yang, "An Early Warning System of Pork Price in China Based on Decision Tree", IEEE Conference Proceedings, International Conference on E-Product E-Service and E-Entertainment (ICEEE), Henan, China, pp. 1–6, 2010.

[19] M. Thakura, S. Olafssonb, J.S Leeb, C.R Hurburgha, "Data Mining for Recognizing Patterns in Foodborne Disease Outbreaks", Journal of Food Engineering Vol. 97, Issue 2, pp. 213– 227, 2010.

[20] S. Lu, Z.J. Cai, X.B Zhang, "Forecasting Agriculture Water Consumption based on PSO and SVM", IEEE Conference Proceedings, "2nd IEEE International Conference on Computer Science and Information Technology", Beijing, China, pp. 147–150, 2009.

[21] A. K. Braimoh, L. Vlek Paul, Stein Alfred, "Land Evaluation for Maize Based on Fuzzy Set and Interpolation", Environmental Management, Vol. 33, Issue 2, pp. 226–238, 2004.