International Journal of Agriculture Innovations and Research
Volume 3, Issue 2, ISSN (Online) 2319-1473
Manuscript Processing Details (dd/mm/yyyy) :
Received : 25/08/2014 | Accepted on : 04/09/2014 | Published : 11/09/2014

An Overview of Data Mining Techniques Applied to Agricultural Soil Data

Manisha Sahane
Faculty of Management Science, Dr. B.A.M. University, Aurangabad (M.S.), India 431004

Balaji Aglave
Florida Ag Research, 3001 N. Kingsway Road, Thonotossassa, FL 33592 USA

Razaullah Khan
Faculty of Commerce & Management Science, Maulana Azad College, Aurangabad (M.S.), India 431004

Sanjay Sirsat
Faculty of Management Science, Dr. B.A.M. University, Aurangabad (M.S.), India 431004

Abstract **– Data Mining is emerging research field in agriculture soil data analysis. Different data mining approaches are in use, such as Nearest Neighbor, Regression, Neural Networks, Support Vector Machines, and Bayesian for areas like online advertising, health care, news publishing, brain science, bioinformatics, analysis of social network study of business, consumer and social insights, agriculture field etc. This review paper describes the discussion on the application of data mining predictive and descriptive techniques in the field of agriculture. Data mining tool will help appropriate amount of plant nutrients in the soil. This information will help to select type of fertilizer, time of application and overall selection of cropping pattern and cropping system. By analyzing the soil data one can come to know that there is use of fertilizers more than the requirement of soil. Excess use of fertilizers is a major agricultural problem that remains to be solved based on the available soil data. The problem can be solved by applying different data mining techniques. Application of non-recommended rates or the wrong kinds of fertilizers can harm the crop and decrease crop production and productivity of the soil. The analysis of agriculture soil dataset with various data mining techniques may yield useful outcomes to farmers in order to make right decisions with respect to crop planning, cropping pattern, fertilizer management, and to maintain the fertility of soil. In this research review study, we describe an overview of data mining techniques applied to agricultural soil data for different purposes.**

Keywords **– Agriculture, Data Mining Technique, Knowledge Discovery, Soil Classification.**

I. INTRODUCTION

Data mining in Agriculture is a relatively a modern and synchronous research domain. Data mining is the process of discovering potentially useful, interesting, and previously unknown patterns from a large collection of data (Sahu*et al.*2011). Data mining has been used to analyze large data sets and establish a useful classification and patterns in the datasets (P. Revathi and M. Hemalatha. 2011). Agricultural and biological research studies have used various techniques of data analysis, including, natural trees, statistical machine learning and other analysis methods (Cunningham and Holmes1999). Data mining is a rapidly growing area of research that has incorporated many techniques from several domains, including statistics, machine learning, database, data warehouse system, pattern recognition, visualization, algorithms, information retrieval and many other (J. Han *et al.* 2011). Data mining is a general term which refers to a set of various different techniques and can be used to classify and clustered data. These techniques differ, but they all exploit the idea that existing data contain information that can be used in the future. To implement data mining techniques, there are various data mining tools available such as RapidMiner, R, KNIME, WEKA, SAS, MATLAB, IBM SPSS Statistics, IBM SPSS Modeler, and Microsoft SQL Server, etc. streams, sequence, text, spatial, temporal, multimedia, web, networks etc. are mining complex types of data.

The soil plays the most important part of farming, but most of the farmers are unaware of soil testing. They do not have pertinent knowledge about their own soil. Fertilizer utilization is a very important agricultural problem. Any farmer is interested in knowing how much fertilizers he is about to use. Today, farmers use fertilizers by considering their past experience on the particular field and the crop. In any of data mining procedures the training data is to be collected for some time back to the past and the gathered data is used in terms of training to learn how to classify future fertilizer recommendations (D. Ramesh and B. Vishnu Vardhan 2013).Depending upon the soil fertile, the sound knowledge experts determine which crops should be taken on the particular soil, how much water should be used and which fertilizers should be used for the same soil area. Farmers should be literate in making right decisions about amending and fertilizing soil. Recommendation of appropriate usage for the given soil sample increases crop productivity.

II. SOIL CLASSIFICATION

It deals with the systematic categorization of soils based on distinguishing characteristics as well as criteria that dictate choices in use. The differentiating characters are the soil properties that can observe in the field or can infer in the field some of them are soil surface and sub surfaces, soil moisture regimes, soil temperature regimes, physical, and chemical properties of soils for classifying soils (P. Bhargavi and Dr. S. Jyothi 2010, 2011). Soil testing is the process of analysis of a soil sample to decide macro and micro nutrients, physical properties and chemical properties. Soil tests are required to calculate soil fertility and to know the deficiencies of some properties. The soil classification system is an essential for the identification of soil properties; therefore reliable automated system for soil classification is needed to make better utilization of time.

International Journal of Agriculture Innovations and Research
Volume 3, Issue 2, ISSN (Online) 2319-1473

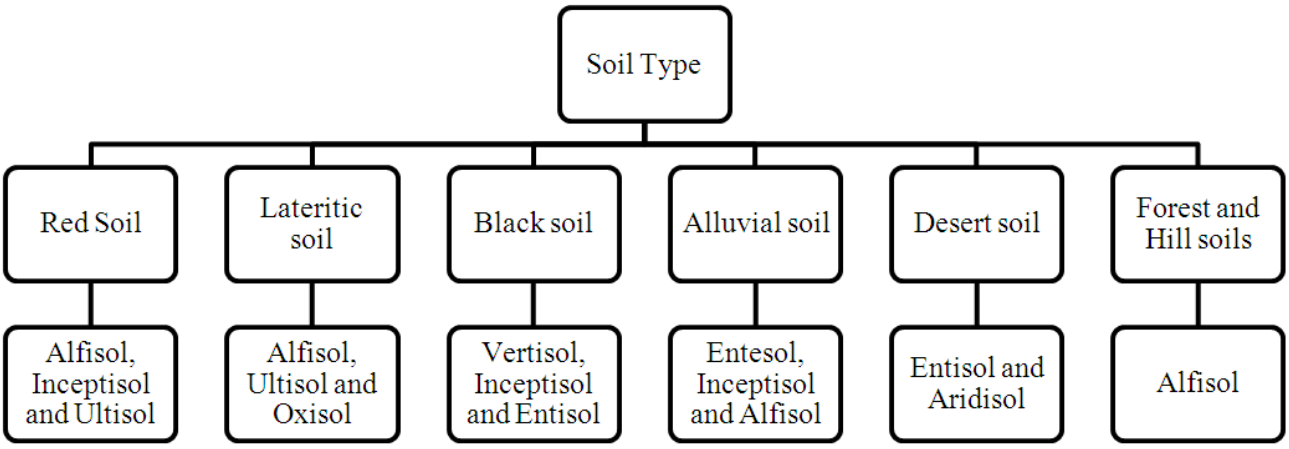

Fig.1 Taxonomically soil classification (Source: Department of Agriculture and Cooperation Ministry of Agriculture Government of India 2011)

Fig.1 shows some dominant groups of Indian soil, classified according to soil taxonomy. Plants need certain essential nutrient elements, some are obtained from air, water and the remaining 13 elements are obtained largely from the soil out of that Nitrogen (N), Phosphorous (P), Potassium (K), Calcium (Ca), Magnesium (Mg), and sulfur (S) are required in large amounts and are referred to as macronutrients. The elements are required considerably in smaller amounts are called micronutrients. It contains Iron (Fe), Zinc (Zn), Copper (Cu), chlorine (Cl), Boron (B), Manganese (Mn), Molybdenum (Mo) (Foth 1990).

III. Data Mining Process

The fast growing, tremendous amount of data stored in data repositories cannot be comprehended by human beings, so to handle that, different powerful mining tools are required. Data mining is about automating the process of searching for trends and hidden patterns in the data. Data mining is a crucial task within Knowledge Discovery in Databases. Knowledge Discovery from Data (KDD) may be defined as "The non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" (Frawley, W. *et al.* 1992). Below steps are in the process of knowledge discovery.

1. *Data Cleaning:* removes corrupted data or empty records.
2. *Data Integration:* means data to be collected and integrated into a single formatted structure.
3. *Data Selection:* Choose only such data that are relevant to the task.
4. *Data Transformation:* Transform Data into a data mining accepted format using smoothing, aggregation or normalization.
5. *Data Mining:* Discover potential knowledge which is hidden within the data.
6. *Pattern Evaluation:* identifies interesting patterns using data mining result and redundant patterns are therefore removed.
7. *Knowledge Presentation:* visualizes knowledge in the most understandable form.

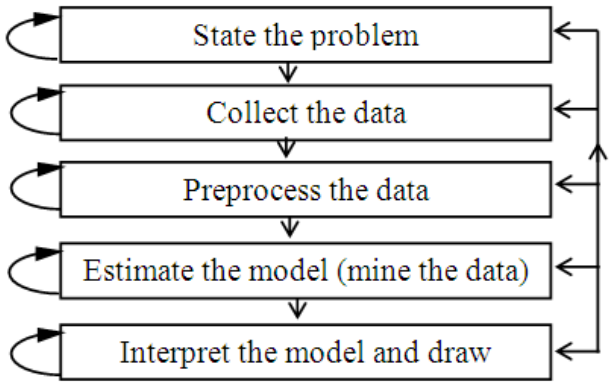

Fig. 2 The data mining process (Source: Kantardzic, M. 2011)

Data mining has become a growing area of significance of the recent increasing demand for knowledge discovery, knowledge acquisition, data visualization and high performance computing. In the process of knowledge discovery, data mining is an essential step. The data mining process is conducted in accordance with the results of the statistical analysis.

Fig. 2 shows the general experimental procedure adapted to data mining problems. To state the meaningful problem and formulate the hypothesis requires domain specific knowledge and experience. Data is usually collected from the databases, data warehouses, and data marts. The collected data are used for model estimation and later used for testing, applying a model. Data preprocessing involves detection of outlier (unusual data elements) and scaling, encoding, and selecting features. By specific data mining techniques that are applied to perform a successful learning process from data and to develop an appropriate model. Data mining tasks are expected to yield highly accurate outcomes using developed model.

IV. Data Mining Techniques

The Ultimate goal of data mining is prediction. *Predictive data mining techniques* predicts future

unknown values of other variables. A Predictive model makes a prediction about values of data using known results found from different data. It contains classification, prediction, regression, Time series analysis. Among Predictive models, Classification is probably the best understood of all data mining approaches. *Descriptive data mining techniques* use to reveal patterns in data, easily interpreted by the farmer. A descriptive model identifies patterns or relationships in data and serves as a way to explore the properties of the data examined, not to predict new properties. It involves clustering, summarization, association rule, and sequence analysis.

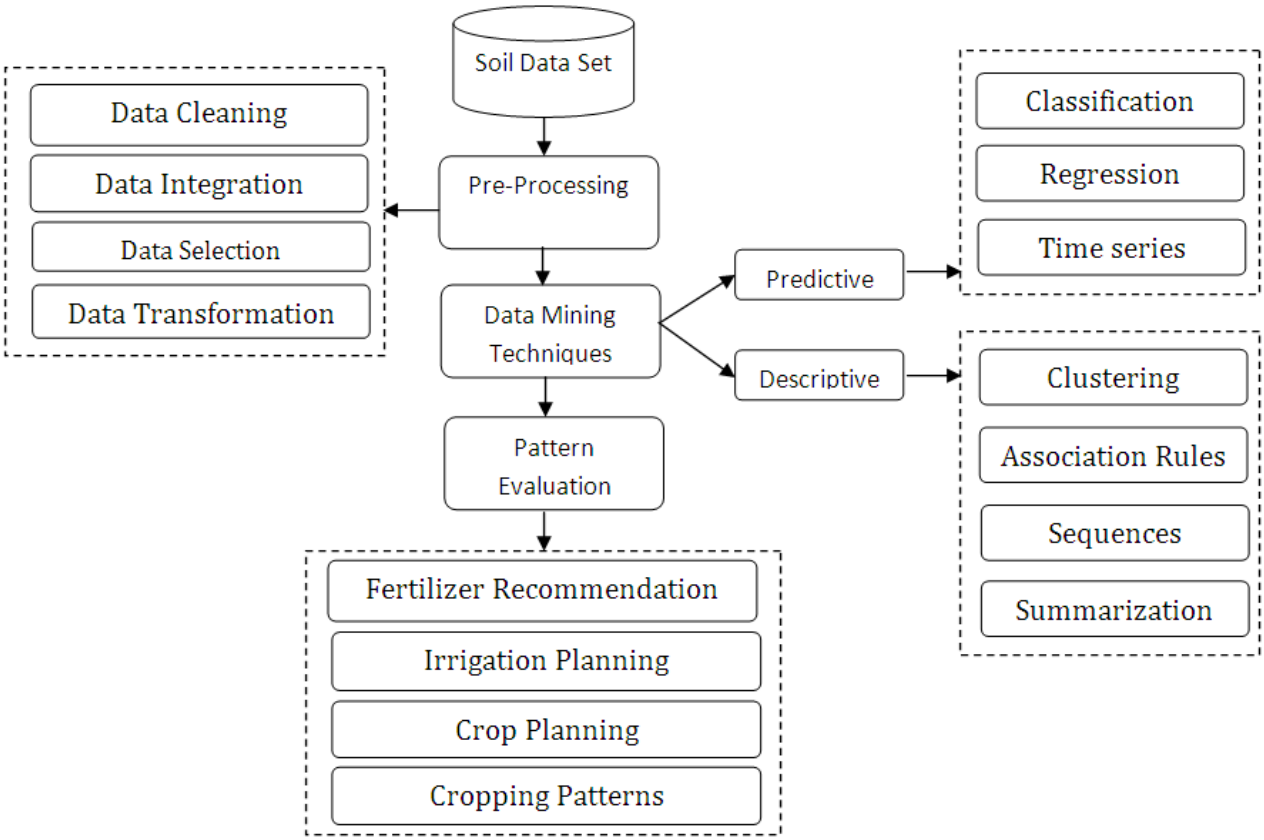

Fig.3 Major steps of soil data set mining

Fig. 3 shows most relevant data mining tasks are classification, clustering, association, sequence or path analysis, regression, visualization.Broadly, there are predictive and descriptive data mining types and in agriculture area mainly predictive techniques are used up till now. One of the methods for knowledge acquisition from the existing agricultural soil databases is the methods of classification. For example, from soil nutrient information, try to predict the fertilizer decisions for farmers. Data classification is a two-step process, consisting of a learning step or training phase (where a classification model is constructed) and a classification step (where the model is used to predict class labels for giving data) (Han, J. *et al.* 2011).Classification can be represented in various forms like a decision tree, neural network and IF-THEN rules. Several data mining techniques are in agriculture and allied area. Few of techniques are discussed here. GATree, Fuzzy classification rules, and fuzzy C-Means algorithms can be used for soil dataset classification (P. Bhargavi and Dr. S. Jyothi 2011). Different data mining classification approaches are in use, such as Nearest Neighbor, Regression, Neural Networks, Support Vector Machines, and Bayesian techniques to classify soil type, soil properties, and soil texture. (Das Diamond and Dandapat Snigdha Jyotsna 2011) attempted to develop a neural network model of the final pH affected by initial pH, pulp density of red mud and time period of incubation. K-means and otsu threshold clustering algorithm is used for feature extraction to identify leaf diseases where K-means is more accurate than otsu threshold clustering technique (Badnakhe and Deshmukh 2012). DM techniques when applied to an agricultural soil profile, may improve the verification of valid soil profile (Ramesh Vamanan *et al.* 2011). Applying genetic algorithm helps one to classify soil texture based on soil properties effectively, which influences fertility, drainage, water holding capacity, aeration, tillage, and bearing strength of soils and also helps in knowing the accuracy of a decision tree along accurate rules (P. Bhargavi and Dr. S. Jyothi 2010).

V. Advantages Of Use Of Data Mining In Agriculture

Different kinds of soils in India indicate that the soil diversity is quite large because of variability of several factors of soil formation (Bhattacharyya *et al.* 2013). Soil fertility is the concept of soil as a medium for plant growth

is an ancient concept. Soil fertility is the ability of a soil to supply essential elements or nutrients in adequate amounts to facilitate optimum plant growth without a toxic concentration of any nutrient. Soil Productivity is for a soil to produce high crop production; it must be fertile for the crop grown. Soil productivity is the soil's capacity to produce a certain yield of crops with optimum planning. For example, productivity of soil for tomato production is commonly expressed as kilos per hectare. High yield depends on optimum management. Sometimes the weather changes effect the less production, but it can increase in the future. Fertilizers detection depends on no of applications, time interval, rate, requirement of micro and macro nutrients, etc. A study of soil profile supplemented by physical, chemical and biological properties of the soil will give a full picture of soil fertility and productivity. (Department of Agriculture and Cooperation Ministry of Agriculture Government of India 2011).

- Data mining techniques can be used to solve complex soil dataset to improve the effectiveness and accuracy of classification of the large soil datasets whereas in general, the statistical techniques are time consuming and highly expensive.
- Application of data mining techniques uses for automation and to develop a decision support system for taking strategic decisions on the agricultural production and protection.
- Data mining techniques are used for efficient knowledge exploration and knowledge acquisition to produce optimized results about farm cultivation.
- Prediction-based data mining models tell revenue and productivity estimation and reporting to aid in making decisions.

VI. Conclusion

The aim of this article is to explain use of data mining in Agriculture. This defines classification data mining techniques for soil data so that the students or researchers can use this for further research or discussion on the computational techniques. This paper reviews many data mining techniques and its experimental use. The analysis of agricultural data sets with various data mining techniques may yield outcomes useful to researchers in the Agricultural field. Data mining techniques can be applied in the field of soil science research for the analysis of the huge amount of data with high accuracy. The analysis and interpretation of classification is a time consuming process that requires a deep understanding of statistics. Data mining techniques when apply to an agricultural soil profile, may improve decision making about the right use of fertilizers, irrigation plans, crop planning, etc. so that soil fertility and soil productivity may get an increase.

References

[1] Armstrong, L. J., Diepeveen, D., and Maddern, R. (2007). The application of data mining techniques to characterize agriculture soil profiles. In *proceedings of 6th Australasian Data Mining Conference* (Australia; pp. 81-95).

[2] Badnakhe, M. R., and Deshmukh, P. R. (2012). Infected Leaf Analysis and Comparison by Otsu Threshold and k-Means Clustering. *International Journal of Advanced Research in Computer Science and Software Engineering*, *2 (3),* 449-452.

[3] Bhargavi, P.and Jyothi, S. (2011). Soil classification using data mining techniques: A comparative study. *International Journal of Engineering Trends and Technology*, 55-58.

[4] Bhargavi, P. and Jyothi, S. (2010). Soil classification using GATREE. *International journal of computer science and Information Technology, 2 (5),* 184-191.

[5] Bhattacharyya, T., Pal, D. K., Mandal, C., Chandran, P., Ray, S. K., Sarkar, D., Velmourougane, K., Srivastava, A., Sidhu, G. S., Singh, R. S., Sahoo, A. K., Dutta, D., Nair, K. M., Srivastava, R., Tiwary, P., Nagar, A. P. and Nimkhedkar, S. S. (2013). Soils of India: historical perspective, classification and recent advances. *Current Science, 104 (10),* 1308-1323.

[6] Cunningham, S. J., and Holmes, G. (1999). Developing innovative applications in agriculture using data mining. *In the Proceedings of the Southeast Asia Regional Computer Confederation Conference.*

[7] Das Diamond and Dandapat Snigdha Jyotsna. (2011). Artificial Neural Network Modelling for the Study of pH on the Fungal Treatment of Red mud. *Research Journal of Chemical Sciences, 1 (2),* 108-112.

[8] Department of Agriculture and Cooperation Ministry of Agriculture Government of India (2011). Methods Manual Soil Testing in India.

[9] D. Ramesh, and B. Vishnu Vardhan (2013). Data mining techniques and applications to agricultural yield data. *International Journal of Advanced Research in Computer and Communication Engineering, 2 (9*), 3477-3480.

[10] Foth, H. D. (1990). Fundamental of soil science. 8thEd. A John Wiley and sons, inc., publication.

[11] Frawley, W. Piatetsky-Shapiro, G. and Matheus, C. (1992). Knowledge Discovery in Databases: An Overview. *AI Magazine 14 (3),* 57–70.

[12] Han, J. and Kamber, M. (2012). Data mining concepts and techniques. 3rdEd. Boston: Morgan Kaufmann Publishers.

[13] Kantardzic, M. (2011). Data mining concepts, models, methods, and algorithms. 2nd.ed. A John Wiley and sons, inc., publication.

[14] P. Revathi, and Dr. M. Hemalatha (2011). Categorize the Quality of Cotton Seeds Based on the Different Germination of the Cotton Using Machine Knowledge Approach. *International Journal of Advanced Science and Technology, 36*, 9-14.

[15] Sahu, H., Shrma, S., and Gondhalakar, S. (2011). A brief overview on data mining survey. *International Journal of Computer Technology and Electronics Engineering, 1 (3),* 114-121.

[16] Vamanan, R. and Ramar, K. (2011). Classification of agricultural land soils a data mining approach. *International journal on computer science and engineering, 3 (1),* 379-384.