

## A Review on Associative Classification Data Mining Approach in Agricultural Soil Land

Annu Kumari Mishra<sup>1</sup>, Pooja Sharma<sup>2</sup>

<sup>1</sup> Computer Science Engineering, Barkatullah University Institute of Technology, Bhopal, India

<sup>2</sup> Computer Science Engineering, Barkatullah University Institute of Technology, Bhopal, India

---

**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. This research aimed to assess the various classification techniques of data mining and apply them to a soil science database to establish if meaningful relationships can be found. A large data set of soil database is extracted from the Soil Science & Agricultural department, Bhopal M.P and National Informatics Centre, The application of data mining techniques has never been conducted for Bhopal soil data sets. The research compares the different classifiers and the outcome of this research could improve the management and systems of soil uses throughout a large number of fields that include agriculture, horticulture, environmental and land use management.

**Keywords** - Data mining, Soil profiles, Agriculture, Classification techniques.

---

### I INTRODUCTION

Data mining has been used to analyze large data sets and establish useful classification and patterns in the data sets. “Agricultural and biological research studies have used various techniques of data analysis including, natural trees, statistical machine learning and other analysis methods [5]. The analysis of agricultural data sets with various data mining techniques may yield outcomes useful to researchers in the Agricultural field. Data mining software applications includes various methodologies that have been developed by both commercial and research centers. These techniques have been used for industrial, commercial and scientific purposes. Agricultural and biological research studies have used various techniques of data analysis including natural trees, statistical machine learning and other analysis tools. This research determined whether data mining techniques could also be used to classify soils that analyze large soil profile experimental datasets. The research aimed to establish if data mining techniques can be used to analyze different classification methods by determining whether meaningful pattern exists across various soil profiles characterized at various research sites [4]. The data set has been assembled from soil surveys at various agricultural areas located in Bhopal M.P, India. The research has utilized

Existing data collected from seven commonly occurring soil types in order to classify soils and correlations between a numbers of soil properties. The soil studies which have been conducted by the Department of Soil Science & Agricultural department, Bhopal M.P provide a vast amount of information on the classification of soil profiles and chemical characteristics. Classification is one of the most important tasks in data mining. Researchers are focusing on designing classification algorithms to build accurate and efficient classifiers for large data sets [6].

## II Associative Classification

Being a new classification method that integrates association rule mining into classification problems, associative classification achieves high classification accuracy, its rules are interpretable and it provides confidence probability when classifying objects which can be used to solve classification problem of uncertainty. Therefore, it becomes a hot theme in recent year. Building Classifier phase tries to remove the redundant rules, organize the useful ones in a reasonable order to form the classifier and the unlabeled data will be classified in the third step. Some experiments done over associative classification algorithms such as CBA [1], CMAR [2], MCAR [3] and GARC [4] state that the associative classification methods share the features of being more accurate and providing more classification rules.

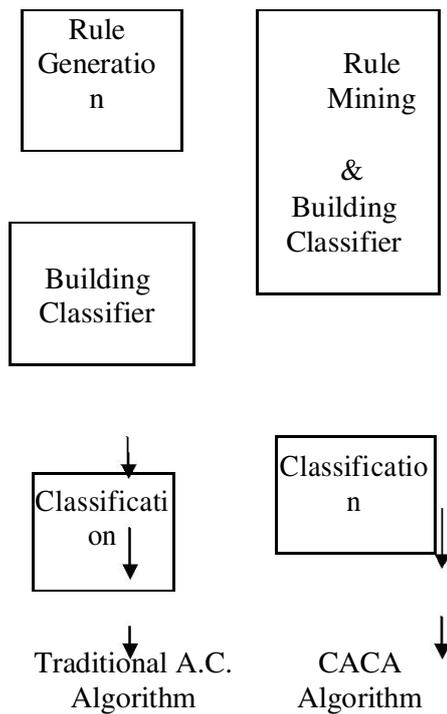
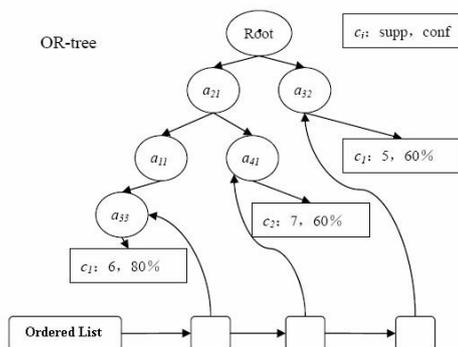


Fig.1 Procedures of A.C. Algorithms



The problem of the knowledge acquisition and efficient knowledge exploitation is very popular

also in agriculture area. One of the methods for knowledge acquisition from the existing agricultural databases is the methods of classification. In agricultural decision making process, weather and soil characteristics are play an important role.

### **III MATERIALS AND METHOD**

#### **3.1 Soil Classification**

The classification of the soil was considered critical to the study because the soil types must be the same in all locations across the study are for the results to be accurate. Soil classification deals with the systematic categorization soils based on distinguishing characteristics as well as criteria that dictate choices in use. Soil classification is a dynamic subject, from the structure of the system itself, to the definitions of classes, and finally in the application in the field. Soil classification can be approached from the perspective of soil as a material and soil as a resource. Engineers, typically Geotechnical engineers, classify soils according to their engineering properties as they relate to use for foundation support or building material. Modern engineering classification systems are designed to allow an easy transition from field observations to basic predictions of soil engineering properties and behaviors.

#### **3.2 The USDA Soil Taxonomy**

The Soil Taxonomy developed since the early 1950's are the most comprehensive soil classification system in the world, developed with international cooperation it is sometimes described as the best system so far. However, for use with the soils of the tropics, the system would need continuous improvement.

#### **3.3 The FAO/UNESCO System.**

The FAO/UNESCO system was devised more as a tool for the preparation of a small-scale soil map of the world than a comprehensive system of soil classification. The map shows only the presence of major soils, being associations of many soils combined in general units. The legend of the soil map of the world lists 106 units classified into 26 groupings. The soil units correspond roughly to great groups from the USDA Soil Taxonomy,

#### **3.5 Classification of soil in Bhopal M.P District**

A set of soil properties are diagnostic for differentiation of pedons. The differentiating characters are the soil properties that can be observed in the field or measured in the laboratory or can be inferred in the field. Some diagnostic soil horizons, both surface and sub surfaces, soil moisture regimes, soil temperature regimes and physical, physical chemical and chemical properties of soils determined were used as criteria for classifying soils.

#### **3.6 Classification in Data Mining**

Techniques used in Data mining can be divided in to two big groups. The first group contains techniques that are represented by a set of instructions or sub-tasks to carry out in order to perform a certain task. In this view, a technique can be seen as a sort of recipe to follow, which must be clear and unambiguous for the executor. If the task is to "cook pasta with tomatoes"

the recipe may be : heat water to the boiling point and then throw the pasta in and check

whether the pasta has reached the point of being at dente : drain the pasta and add preheated tomato sauce and cheese. Even a novice chef would be able to achieve the result following this receipt. Moreover, note that another way to learn how to cook pasta is to use previous cooking experience and try to generalize this experience and find a solution for the current problem.

#### **IV DATA MINING PROCESS**

The data mining process was conducted in accordance with the results of the statistical analysis. The following steps are a general outline of the procedure that allowed a cluster analysis to be conducted on the dataset.

##### **4.1 Data Collection cleaning and checking**

Relevant data was selected from a subset of the soil science database. The soil samples collected from the various regions of Bhopal M.P District

##### **4.2 Data formatting**

The data was formatted into an Excel format from the Access database, based on the ten soil types and relevant related fields. The data was then copied into a single Excel spread sheet. The Excel spread sheet was then formatted to replace any null or missing values in the soil data set to allow coding for the file in the next phase.

##### **4.3 Data Coding**

The soil data set was then converted into a comma delimited (CSV) format file for the Excel Spread Sheet. This file was then saved and opened using a text editor. The text editor was used to format and code the data into the type that will allow the data mining techniques and programs to be applied to it. The coding was formatted so that the input will recognize names of the attributes, the type of value of each attribute and the range of all attributes. Coding was then conducted to allow the machine learning algorithms to be applied to the soil data to provide relevant outcomes that were required in the research.

#### **V RESULTS**

The research activities involved a process to establish if classification could be found in the data. These processes involved the statistical manipulation of the data set in Excel. The aim of the research was to determine if a relationship or correlation can be established with soil data set. The process involved the creation of analysis tools and charting the data so that the classification of soils is displayed and experts can interpret the findings.

#### **VI CONCLUSION**

The experiments conducted analyzed small number of traits contained within the dataset to determine their effectiveness when compared with standard statistical techniques. The agriculture soil profiles that are used in this research were selected for completeness and for classification of soils. The recommendations arising from this research implies that data mining

techniques may be applied in the field of soil research in the future as they will provide

research tools for the comparison of large amount of data. Data mining techniques when applied to an agricultural soil profile, may improve the verification of valid soil profile, may improve the verification of valid patterns and profile classification when compared to standard statistical analysis techniques.

## REFERENCES

- [1] Campus-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: 134 – 141.
- [2] Cunningham, S.J. and Holmes, G., *The Proceedings of the Southeast Asia regional computer confederation conference*, 1999.
- [3] Jorquera H, Perez R, Cipriano A, Acuna G(2001) Short term forecasting of air pollution episodes. In. Zannetti P (eds) *Environmental Modeling 4*. WIT Press, UK.
- [4] Meyer GE, Neto JC, Jones DD, Hindman TW (2004) Intensified fuzzy clusters for classifying plant, soil and residue regions of interest from color images. *Comput Electronics Agric* 42: 161 – 180.
- [5] Rajagopalan B. Lall U (1999) A k-nearest-neighbour for daily precipitation and other weather variables. *Wat REs Res* 35(10): 3089 – 3101.
- [6] Tripathi S, Srinivas VV, Nanjudiah RS (2006) downscaling of precipitation for climate change scenarios: a support vector machine approach, *J. Hydrol* 3 621 – 640.



