Agricultural Land Soil Classification Using Data Mining Technique

Sindu Ja

Abstract- *Agriculture is the prime and traditional sector deals with the systematic division of soils based on distinguished characteristics as well as criteria. The practice of growing large scale for food and other purposes is known as agriculture .However, knowledge of using data mining to develop and discover the research area of soil classification. Soil is the essential element for the development of agriculture. Soil classified into group or classes or the basics of physical and chemical properties in their layers .The Large amount of data are harvested along the with the crop harvest in agriculture. It includes the knowledge from huge amount of data is virtually a difficult task in the current scenario. The advanced knowledge which were used in Agricultural research has been improved by technical advances in computation, automation and data mining. The Data Mining techniques applied on Agricultural data include GA Tree, k-nearest algorithm, CART and Naïve Bayes classifier data mining. In agricultural soil datasets is a relatively a contemporary research sector. This data mining algorithms are used for analyzing the soil datasets for classification purposes. The various methodologies were used in this data mining and it was used in this research. The study of soil classification is used to compare and analyze the classification of agriculture soil data.*

Keywords- Data mining, soil classification, GA Tree algorithm, k-nearest algorithm, CART, Navie Bayes.

I. INTRODUCTION

In day to day life,data Mining is used in different area.It is the process as discovering patterns in large data sets which describes the different the methods at the intersection of statistics and database system. it also may be organized in many different ways the logical or mathematical model of a particular organization of data is called data mining. The data appearing in our data structures are processed by means of certain operations.The traditional method of agriculture is slowly being replaced by scientific and technical methods.An algorithm can be defined as a step by step procedure that provides solution to a given problem.It is defined as a ranking of steps used for computational procedures,which usually starts with an input value and generates the desired output. soil classification deals with the regular categorization of soils based on discern characteristics as well as standard that conrol in use.soil classification of composite soils exclusively based on the particle size distribution is known as textural classification.the soils may be classified into particle size classification of soil,textural classification,Highway Research Board(HRB),Unified soil classification and IS Classification.By applying various sets of data we can introduce the data which were used in our data.It generates the methods to predict the value of database and dependent on the variable directly without first generating a model. data mining algorithm GATree, k-nearest algorithm , CART and Naive Bayes classifiers are used to collect the soil data from the given data sets. These algorithms were used to associate the data to execute the required output.

II. SOIL DATASETS

The soil datasets access is the web service and application purpose to requirements and requesting the soil survey and the tabular data.

soil datasets

Class no	Soil texture class	Class abbreviation	%sand	%silt	%clay
1	Sand	S	92	5	3
2	Sandy loam	SL	58	32	10
3	Loam	L	43	39	18
4	Silty clay	SIC	6	47	47
5	Clay	C	22	20	58
6	Water	W	0	0	0
7	other	O	0	0	0

III. CLASSIFICATION OF SOIL

The particle sizes encountered in soils is very large: with dimension of over 300 mm down to clay particles that are less than 0.002 mm. they requires three types of soil namely ABC. In this soil type the type C requires the least stable soil.the type C includes gravel and sand. Some clays particles contains less than 0.005 mm in size which behave as colloids, i.e. do not settle in water.

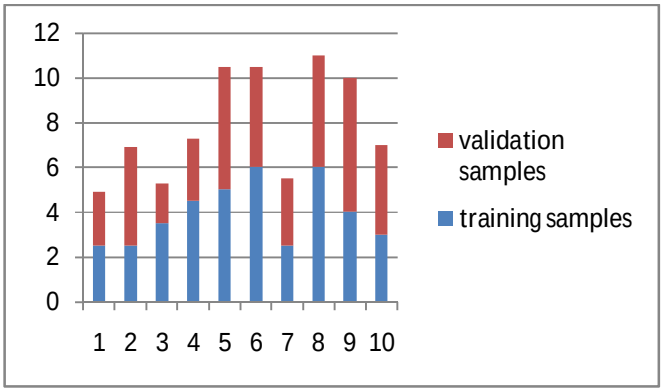
In the Indian Standard Soil Classification System (ISSCS), soils are classified into groups according to group, and the soils are further classified into medium, coarse and fine sub-groups.

The symbols of the various soils are as: Gravel(G), Sand(S), Silt or Silty(M), clay or clayey(C), peat(P), well graded(W) and poorly graded(P).

USCS	ISSCS
High plasticity (if liquid limit > 50%)	High plasticity(if liquid limit > 50%)
Low plasticity (if liquid limit <50%)	Intermediate plasticity(if liquid limit is in between 35% to 50%)
	Low plasticity(if liquid limit < 35%)

GATree algorithm

Genetic algorithm is used in computer science and operations research. It is the process of natural selection and that depends to the larger scale of algorithm which is known as Evolutionary Algorithm(EA). It gives a approximate solutions to all the required problems. A classic genetic algorithm requires: [1] a genetic representation of the solution domain [2] a fitness function to evaluate the solution domain. Genetic operator is also a combination of crossover it is also called as recombination and mutation. The soil classification is used the Genetic Algorithm which is used under the process of decision tree. The idea which is applied to this paper is simple and intention to this paper. There are some methods to build up the data of Genetic Algorithm using some modules of selection and crossover. The Genetic Algorithm is used under the process of Decision tree by using this it optimize some values of Genome Score of 0.34.



k-nearest algorithm

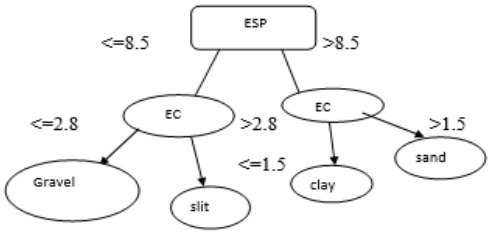
K nearest algorithm is used to store all available cases and classifies based on similarity measures. It is a computer classification method and it is also used to predict the data values. It is very simple algorithm and it is wored under the distance queried to the training samples of KNN algorithm. It is an unsupervised learning and the algorithms are used to make a group or sets of data.

Compute the distance

$$a(x) = ||x'-y||/||x-y||$$

CART(Classification and Regression Tree)

CART data mining is a process of identifiying the useful information from large amount of random data. It is mostly used with the decision tree algorithm. The main target of the algorithm is to get the input values. It is the general method of input values and a mixture of variables. CART algorithm is constructed via an algorithm to identify the ways to execute a data set which based on different values.



Navie Bayes

It is a classification technique which is used assume the prediction values which gets the required output from the desired values. It is easy to build and particularly it is used to store the data sets. Navie bayes has some calculating way to use the large data sets of the given values. Navie bayes is used to classify the soil and to analyze the agriculture techniques. This paper is based on different algorithm to solve and predict the soil classification datasets. There are various sets of data to improve the soil in the agriculture fields. This Navie bayes algorithm is mainly establish the relationship of soil datasets.

IV. CONCLUSION

In this paper , we have discussed about the algorithms which were used in the agriculture soil classification. This paper proposed the analyse the data which was stored in the datasets. The result of this paper showed that verification in soil parameter of major contribution which was yield by the soil fertilizes and its classification of soil. We have also discussed about the several algorithm which was used in this data mining technique of soil classification.

REFERENCES

[1] M. Kumari & S. Godara, (2011), “Comparative Study of Data Mining Classification Methods.

[2] V. Ramesh and K. Ramar, 2011. Classification of Agricultural Land Soils: A Data Mining Approach. Agricultural Journal, 6: 82-86.
[3] Verheyen, K., Adriaens, D., Hermy, M., & Deckers, S. (2001). High-resolution continuous soil classification using soil profile descriptions. 101(3), 31-48.
[4] Rajagopalan, B., & Lall, U. (1999). k–nearest-neighbor simulator for soil classification RESEARCH,35(10), 3089-3101.
[5] Megala, S., & Hemalatha, M. (2011). A Novel Datamining Approach to Determine the Vanished Agricultural Land in Tamilnadu International Journal of Computer Applications, 23.
[6] D Ramesh, B Vishnu Vardhan, (2013). Data Mining Techniques and Applications to Agricultural Yield
[7] J. Han & M. Kamber, (2006), "Data Mining: Concepts and Techniques" Second Edition, San Francisco: Morgan Kaufman Publishers,