ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

Implementation of Association Rule Mining for different soil types in Agriculture

M.C.S.Geetha

Assistant Professor, Department of Computer Applications, Kumaraguru College of Technology, Coimbatore, India

Abstract: Agriculture sector is the mainstay and backbone of the Indian economy. Despite the focus on industrialisation, agriculture remains a main sector of the Indian economy both in terms of contribution to gross domestic product (GDP) as well as a source of employment to millions across the country. The total Share of Agriculture & Allied Sectors (Including livestock, agriculture, forestry and fishery sub sectors) in terms of percentage of GDP is 13.9 percent during 2013-2014 at 2004-2005 prices. Agricultural exports constitute a fifth of the total exports of the country. At 157.35 million hectares, India holds the second leading agricultural land worldwide. All the 15 major climates are initiated in India and the country also possess 45 of the 60 soil types in the earth. India is the biggest producer of milk, tea, pulses, cashew and jute, and the second biggest producer of rice, wheat, fruits and vegetables, cotton, sugarcane and oilseeds [1]. In agricultural decision production process, both weather and soil characteristics plays a vital role. This research aimed to assess a variety of association techniques of data mining and apply them to a soil science database to establish if meaningful relationships can be created. A huge data set of soil database is extracted from the Soil Science India. This paper provides the data mining association techniques used in agriculture which includes Apriori.

Keywords: association, apriori algorithm, agriculture, soil types.

I. INTRODUCTION

Data mining is the process that results in the discovery of new patterns in huge data sets. The goal of the data mining process is to mine knowledge from an existing data set and change it into a human logical formation for advance use. It is the process of analyzing data from different perspectives and summarizing it into useful information. There is no restriction to the type of data that can be analyzed by data mining.

Data mining tasks can be classified into two categories: Descriptive data mining and Predictive data mining. Descriptive data mining tasks characterize the general properties of the data in the database while predictive data mining is used to predict explicit values based on patterns determined from known results. Prediction involves using some variables or fields in the database to predict unknown or future values of other variables of interest[2]. The paper is organized as follows: Chapter 2 discusses the methods. Chapter 3 discusses the implementation of data mining association technique. Chapter 4 discusses the conclusion.

II. METHODS

The main techniques for data mining include Association rule mining, Classification, Clustering and Regression. The different data mining techniques used for solving different agricultural problem has been discussed [3].

2.1 Association Rule Mining

Association rule mining technique is one of the most efficient techniques of data mining to search unseen or desired pattern among the vast quantity of data. In this method, the focal point is on finding relationships between the different items in a transactional database. Association rule mining are used to discover out elements that co-occur repeatedly within a dataset consisting of many independent selections of elements (such as purchasing transactions), and to discover rules. The simple problem statement is: A set of transactions given, where each transaction is a set of literals, an association rule is a phrase of the form A => B, where A and B are sets of objects.

The inherent meaning of such a rule is that transactions of the database which contain A tend to contain B.[4] An application of the association rule mining is the customer segmentation, market basket analysis, catalog design, store layout and telecommunication alarm prediction.

The different association rule mining algorithm are Partition, Apriori Algorithm(AA), Dynamic Itemset Counting(DIC), FP Growth(FPG), Dynamic Hashing and Pruning(DHP), SEAR, Spear, Eclat & Declat, MaxEclat.[5]

2.2 Classification

Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict upcoming data trends. It is a method in which a model learns to predict a class label from a set of training data which can then be used to predict discrete class labels on new samples.

To maximize the predictive accuracy obtained by the classification model when classifying examples in the test set unseen during training is one of the major goals of classification algorithm [6].

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

2.3 Clustering

In clustering, the focus is on finding a partition of data records into clusters such that the points within each cluster are close to each other. Clustering groups the data instances into subsets in such a manner that related instances are assembled together, while different instances belong to varied groups. Since the aim of clustering is to discover out a new set of categories, the newest groups are of interest in themselves, and their assessment is essential [7].

2.4 Regression

Regression is learning a function that maps a data item to a real-valued prediction variable. The different applications of regression are predicting the amount of biomass there in a forest, estimating the probability of patient will survive or not on the set of his diagnostic tests, expecting consumer demand for a new product [8].

III. IMPLEMENTATION OF DATA MINING ASSOCIATION TECHNIQUES

Apriori Algorithm:

Apriori is an algorithm for frequent item set mining and association rule learning over transactional record. It proceeds by categorizing the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database [9].

Apriori algorithm is easy to execute and very simple, is used to extract all frequent item sets in database. The algorithm makes much investigation in database to find frequent item sets where k item sets are used to generate k+1-itemsets. Each k-item set must be greater than or equal to minimum support threshold to be frequency. Otherwise, it is called candidate item sets.

In the first, the algorithm scan database to find frequency of 1-itemsets that contains only one item by counting each item in database. The frequency of 1-itemsets is used to find the item sets in 2-itemsets which in turn is used to find 3-itemsets and so on until there are not any more k-item sets.

Let's start with example,

Soil Types	Crops Grown
Alluvial	Rice, Wheat, sugarcane, Cotton, Jute.
Black	Rice, Wheat, sugarcane, Cotton, Tobacco, Vegetables,
Laterite	Cashew, Rubber, Coconut, Tea, Coffee
Mountain	Tea, Coffee, Spices, Tropical fruits
Red	Rice, Wheat, sugarcane, Tobacco, Vegetables, Ragi.

Original Table:

Soil Types	Crops Grown
Alluvial	{R,W,S,CN,J,}
Black	{R,W,S,CN,TC,V}
Laterite	{C,RU,CT,T,CF}
Mountain	{T,CF,SP,TF}
Red	{R,W,S,TC,V,RA}

Step 1: Count the number of transactions in which each item occurs, Note 'R=Rice' is crops grown 3 times in total,

Soil Types	Crops Grown
R	3
W	3
S	3
CN	2
J	1
TC	2
V	2
C	1
RU	1
CT	1
T	2
CF	2
SP	1
TF	1

Step 2: Now remember we said crops grown at least 3 soils. So in this step we remove all the crops that are planted less than 3 types of soil from the above table and we are left with

Soil Types	Crops Grown
R	3
W	3
S	3

Step 3: We start making pairs from the first item, like RW, RS and then we start with the second item like WS. We did not do WR because we already did RW. After making all the pairs we get,

Soil Types
RW
RS
WS

Step 4: Now we count how many times each pair is grown together. For example R and W is just grown together in {R, W, S, CN, J}

While R and S is grown together 3 soils in After doing that for all the pairs we get {R,W,S,CN,J}, {R,W,S,CN,TC,V} AND {R,W,S,TC,V,RA}

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

Soil Types	Crops Grown
RW	3
RS	3
WS	3

Step 5: Golden rule to the rescue. Eliminate all the item pairs with number of crops grown less than three and we are left with

Soil Types	Crops Grown
RW	3
RS	3
WS	3

Step 6: To make the set of three items, one more rule is needed (it's termed as self-join), It simply means, from the Item pairs in the above table, find two pairs with the same first Alphabet, so get.

- RS and RW, this gives RSW

Then find how many times R,S,W are crops grown together in the original table, so get the following table

Soil Types	Crops Grown
RSW	3

Step 7: So, again apply the golden rule, that is, the item set must be grown together at least 3 times which leaves with just RSW.

Thus the set of three items that are grown together most frequently are R, S, W.

IV. CONCLUSION

Agriculture is the most significant application area particularly in the developing countries like India. Use of information technology in agriculture can change the situation of decision making and farmers can yield in better way. Data mining helps in decision making on several issues related to agriculture field. This paper discusses about the role of data mining techniques in agriculture field and their related work by implementing association rule mining for different soil types in context to agriculture domain.

REFERENCES

[1]. http://www.ibef.org/industry/agriculture-india.aspx
[2]. Ramesh D, Vishnu Vardhan B., "Data Mining Techniques and Applications to Agricultural Yield Data", IJARCCE, Vol. 2, Issue 9, September 2013.
[3]. Mucherino, A., Papajorgji, P., & Pardalos, P. (2009), "Data mining in agriculture" (Vol. 34), Springer.
[4]. Srikant, R V Q & Agrawal, R (1997, August), "Mining Association Rules with Item Constraints. In KDD" (Vol. 97, pp. 67-73).
[5]. Zaki, M J (1999), "Parallel and distributed association mining: A survey". IEEE concurrency, 7(4), 14-25.
[6]. Beniwal, S. & Arora, J. (2012), "Classification and feature selection techniques in data mining", International Journal of Engineering Research & Technology (IJERT), 1(6).
[7]. Lior Rokach, Oded Maimon, "Clustering Methods", Chap-15
[8]. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996), "From data mining to knowledge discovery in databases". AI magazine, 17(3), 37.
[9]. http://en.wikipedia.org/wiki/Apriori_algorithm
[10]. Hunt, E.B., Marin. and Stone,P.J. (1966). Experiments in induction, Academic Press, New York
[11]. Shafer, J., Agrawal, R., and Mehta, M. (1996). Sprint: A scalable parallel classifier for data mining. Proceedings of the 22nd international conferenceon very large data base. Mumbai (Bombay), India
[12]. Quinlan, J. R. (1993). C45: Programs for Machine Learning. Morgan Kaufmann, San Mateo, CA.
[13]. Mehta, M., Agrawal, R., and Rissanen, J. (1995). MDL-based decision tree pruning. International conference on knowledge discovery in databases and data mining (KDD-95) Montreal, Canada

BIOGRAPHY

M.C.S.Geetha is an Assistant Professor in the Department of Computer Applications, Kumaraguru College of Technology, Coimbatore. She received Master of Computer Applications (MCA) degree in 2004 from P.S.G.R.Krishnnammal College for Women, Coimbatore, India. She received M.Phil (Computer Science) in 2006 from Bharathiyar University. Her research interest is Data Mining. She has published many papers in International Journal.