



Efficient Execution of Apriori Algorithm Using WEKA

Dr. Neeraj Bhargava

Associate Professor: Dept. of Computer Science,
School of Engineering and System Sciences,
MDS University, Ajmer, India
drneerajbhargava@yahoo.co.in

Nidhi Gupta

Dept. of Computer Science,
School of Engineering and System Sciences,
MDS University, Ajmer, India
nidhigupta.research11@gmail.com

Ajay Singh Gaur

Director: TJS College, Kota
Kota, India
ajaytjs@gmail.com

Abstract: In these days people are depending on computer and electronic devices. These devices require some specific information to be store like customer interest in marketing. Sometimes this stored information is not able to help other system such as: frequent products demand. So, dataware house is used to store information. But still, it requires an approach to access data and operate. Data Mining is the collection of operation that used to relate and manipulate data according to requirement. A significant approach which is used to associate data defines by Association Rule. WEKA provides an implementation of Association Rule using Apriori Algorithm. This research paper presents the efficient use of WEKA for execution of Apriori Algorithm on data.

Keywords: DataMining; AssociationRule; AprioriAlgorithm; WEKA

I. INTRODUCTION

With the advent of DBMS (Database Management System), it is very easy to store large data into managed form. The collection of Historical database from different sources (known as dataware house). The science of extracting useful information from large data sets or databases (known as Data Mining). The challenging task is to extracting useful information from the large collection of data in Dataware house and data. Around the world lot of research is underway to discover the knowledge from the large collection of data in data warehouse. There are several major data mining techniques have been developed and used in data mining projects recently including association, classification, clustering, prediction and sequential patterns. Concept wise, Association rule uses before a long time. Association rules are used for knowledge discovery and to take useful managerial decision in the organization based on the results of associations among data stepping toward to make a smarter system. For Data Mining technique a free GUI software is available that is WEKA. This introduced as a machine learning free software after 1997. It was written in Java which makes it platform independent. It has easy to use GUI and ability to perform many operations [1].

The actual data mining task is using large quantities of data to extract previously unknown interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining).

II. DATA MINING

Data Mining is a Knowledge-Discovery in Database. A Data Mining is process access useful information From Large Databas. Data Mining uses following technology First Statistics this uses (Collect data, classify data, summarized data, present data, draw a conclusion) Second this concerned with the

development of techniques which allow computer to learn, this concerned with algorithmic complexity of computational implements This taking a Raw Data and action based on Category of data. This aims to classify data patterns based on statistical information. A Data Mining uses following techniques Association, Classification, Prediction, Sequential pattern. Using these technology Data Mining process is uses so user could access data in understandable format [1] [2].

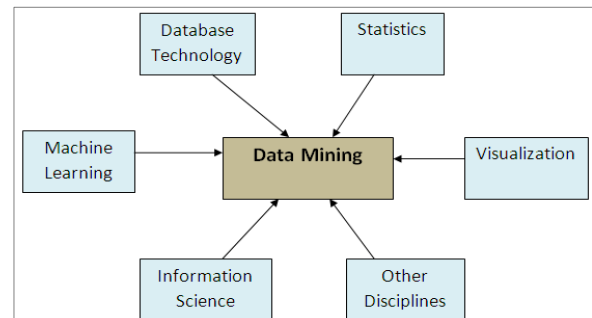


Figure1:- Fields of Data Mining

III. WEKA

A Data Mining technology is implements though a freely available GUI software is called WEKA. WEKA is open source and platform independent Software. This is easily understandable of those uses that are not friendlier with software environment. WEKA Contain tools for Datapre-processing, Classification, and Regression, Clustering, Association rule and Visualization. One can access the WEKA class library from our own Java program, and implement new machine learning Algorithm. WEKA uses the six tabs for the processing. First Preprocess this tan is used to chose the data file to be used in application. Second Classify this tab is use to test and train different learning schemes on preprocessed data file under experimentation Third Cluster this tab is used to

apply different tools that identify cluster within a Data files Fourth Association Rule this tab is used to apply different rule to the file and find frequent item sets Fifth Select Attributes used to apply Different Rules to reveal changes based on selected attributes inclusion or exclusion from the experiments Sixth Visualize this is used to see what the various manipulation produced on the dataset in 2d format, in scatter.

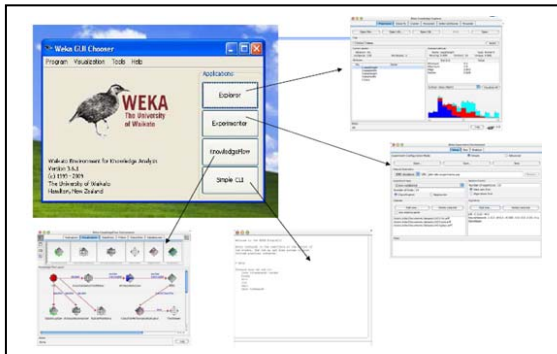


Figure2:-WEKA Graphical User Interface

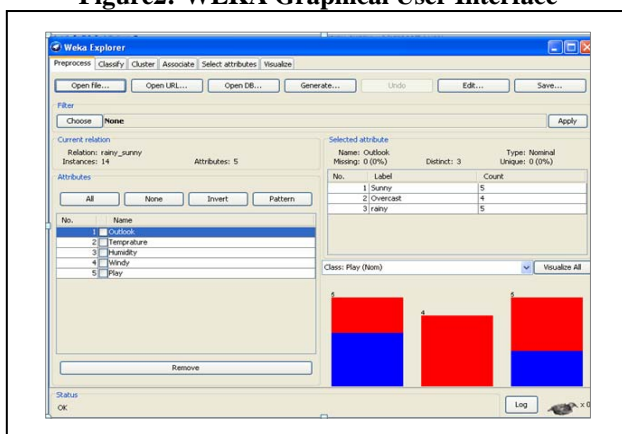


Figure3:-WEKA Environment

proceedings, and not as an independent document. Please do not revise any of the current designations [3] [4].

IV. ASSOCIATION RULE

Mining for Association among items in a large database of sales transaction is an important database mining function. Association rule is to identify items that are brought together by sufficiently many customers. A Market Basket Analysis to find how items purchased by customer are related. For example, the information that a customer who purchases a keyboard is tends to buy a mouse at the same time is represents this transaction can consider as a Association rule. A Apriori Algorithm finds the frequent itemsets. These itemsets that have Minimum Support. In Apriori Algorithm a Subset of frequent itemset must also be frequent itemset. Eg :-{ A, B } frequent itemset both {A} and {B} should be frequent itemset. This find frequent itemset with Cardinality from 1 to k (k-itemset). These frequent itemsets uses to generate Association Rule [2].

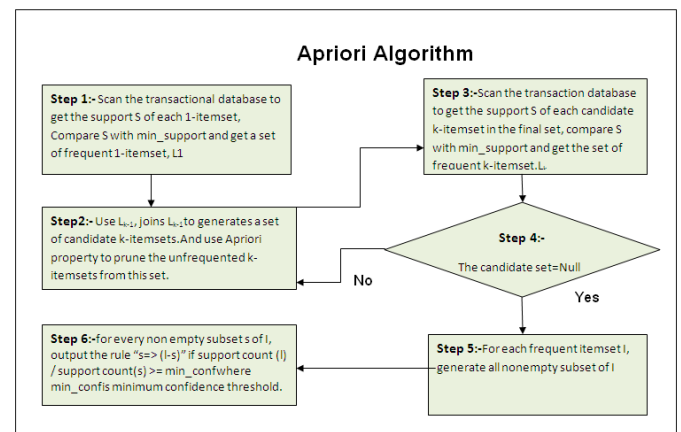


Figure 4:-The Steps of Apriori Algorithm

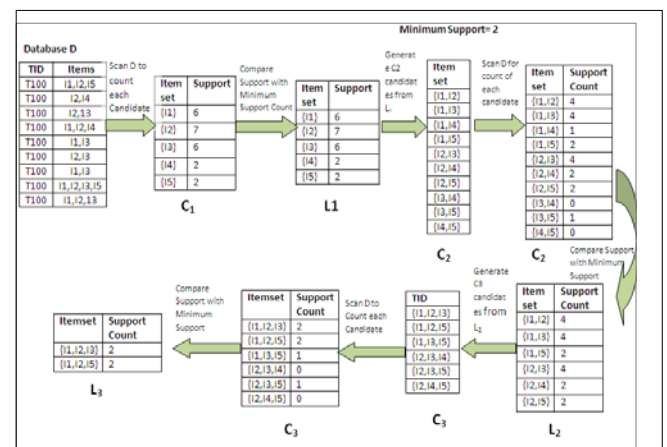


Figure 5:-Example of Apriori Algorithm

The algorithm uses $L_3 \text{ Join } L_3$ to generate a candidate set of 4-itemsets, C_4 . Although the join results in $\{\{I1, I2, I3, I5\}\}$, this itemset is pruned since its subset $\{\{I2, I3, I5\}\}$ is not frequent. Thus $C_4 = \text{NULL}$ and Algorithm terminate having found our frequent itemsets our Algorithm is complete. Now our next step is using frequent itemsets generates Strong Association Rule. Back to the Example we had $L = \{\{I1\}, \{I2\}, \{I3\}, \{I4\}, \{I5\}, \{I1, I2\}, \{I1, I3\}, \{I1, I5\}, \{I2, I3\}, \{I2, I4\}, \{I2, I5\}, \{I1, I2, I3\}, \{I1, I2, I5\}\}$. $I = \{I1, I2, I5\}$ Its all nonempty subsets are $\{I1, I2\}, \{I1, I5\}, \{I2, I5\}, \{I1\}, \{I2\}, \{I5\}$. Let minimum confidence thresholds say 70%. The resulting association rules are shown below, each listed with its confidence. $R1: I1 \wedge I2 \Rightarrow I5$ Confidence = $\text{sc}\{I1, I2, I5\} / \text{sc}\{I1, I2\} = 2/4 = 50\%$ $R1$ is Rejected. $R2: I1 \wedge I5 \Rightarrow I2$ Confidence = $\text{sc}\{I1, I2, I5\} / \text{sc}\{I1, I5\} = 2/2 = 100\%$ $R2$ is Selected. $R3: I2 \wedge I5 \Rightarrow I1$ Confidence = $\text{sc}\{I1, I2, I5\} / \text{sc}\{I2, I5\} = 2/2 = 100\%$ $R3$ is selected. Using such steps we creates Association rule.

V. RESULT

As WEKA help to use Association rule, the first step to open a MSeExcel to create a table. Now, add records in a table and save table with .CSV (Comma Separate Value) format. By using WEKA, one can open a .CSV file and save this file in .ARFF format. WEKA support .ARFF version of file to apply association rules. The following diagrams below, describes the GUI interaction.

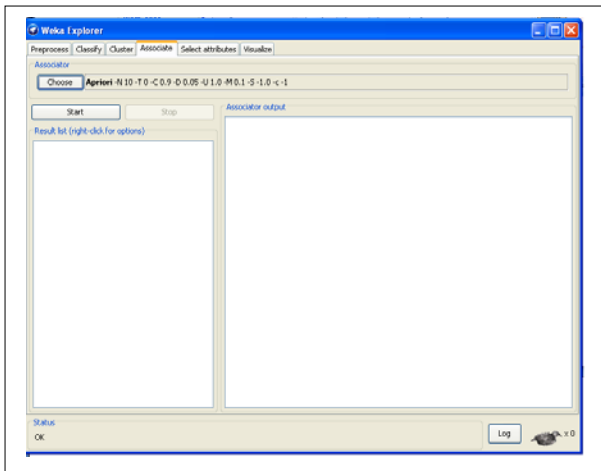


Figure 6:- Select on the Association Rule Tab

Minimum Support can also be established according to requirements.

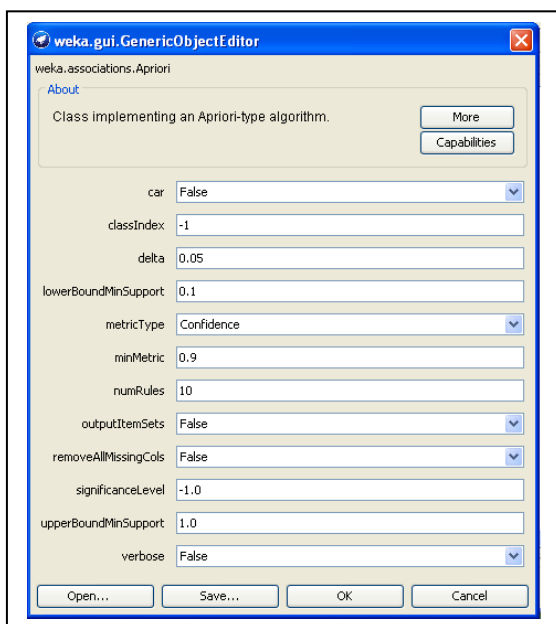


Figure 7:-Change the Minimum Support Values

Now after execution a possible item sets will appear as a result which will uses for answer. Given the result below the Association rules is defined according to set Minimum support. In result possible item sets of data is displayed as T3,T2,T5 a possible combination of item sets is display also with confidence on datasets In a large item sets there is all possible combination is calculates. A unique ID identifies a Dataset through which a possible combination is display. In market strategies uses such types of information. In this calculates easy to access Association rule.

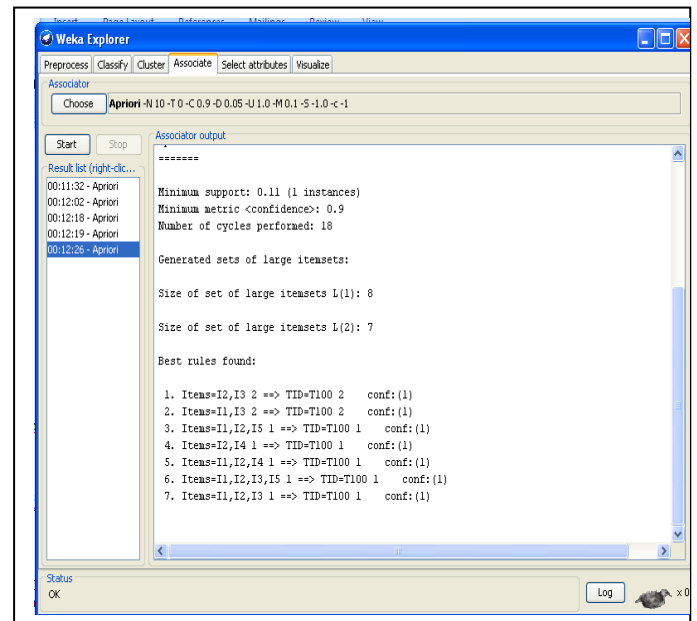


Figure 8:-Result After Association Rule

VI. CONCLUSION

This research paper provides a basic tutorial to learn about Association Rule using WEKA. By using WEKA, one can calculate Apriori Algorithm automatically, without needed any calculation by the user. WEKA provides nearly all features of Data Mining. Association rule provides Market strategies for improving product quantities. Considering Association rule a number of products could easily generates. WEKA predicts the number of item set to generate frequencies. This paper helps to easily generates the list of the possible item sets using WEKA.

VII. REFERENCES

- [1] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and ex Aher, Sunita B., and LOBO L.M.R.J.. "Data Mining in Educational System using WEKA."Proceedings published by International Journal of Computer Applications (IJCA)(2011): 6. Web. 9 Apr. 2013.
- [2] Han, Jiawei. "Data Mining: Concepts and Techniques."www.cs.uiuc.edu/~hanj(2006): 30. Web. 10 Apr. 2013
- [3] Serasiya, Shilpa, and Neeraj Chaudhary. "Simulation of Various Classifications Results using WEKA. "International Journal of Recent Technology and Engineering (IJRTE) 1.3 (August 2012): 8. Web. 10 Apr. 2013.
- [4] Scuse, David, and Peter Reutemann. "WEKA Experimenter Tutorial for Version 3-5-8."c 2002-2008 David Scuse and University of Waikato(July 14, 2008): 40. Web. 10 Apr. 2013.
- [5] Hall, Mark, Eibe Frank, and Geoffrey Holmes. "The WEKA Data Mining Software: An Update." 11.1 9. Web. 10 Apr. 2013.