

A Survey of Mining Association Rule Techniques

M. Sumitha^{1*}, P.S. Suganya²

^{1,2}BCA Department, Sri Akilandeswari Women's College, Vandavasi, Tiruvannamalai, Tamil Nadu, India

*Corresponding Author: msumi1990@gmail.com

Available online at: www.ijcseonline.org

Abstract— Association rule is a data mining techniques which is used to mine several algorithms of applying association rules from the given databases. One of the most important algorithm is Apriori that is used to extract frequent itemsets from large database and association rules are help to discovering the knowledge. Association rules are if/then statements that help uncover relationships between unrelated data in a relational database or other information repository. In this paper we mention different techniques for mining association rules from relational databases and their usage in different areas.

Keywords— Data mining, Apriori algorithm, Database, Association rules.

I. INTRODUCTION

Data mining

Data mining refers to extracting information from large amounts of data, and transforming that information into an understandable and meaningful structure for further use. Data mining is an essential step in the process of knowledge discovery from data (or KDD) [1]. Data mining tools are used to predict the pattern and behaviour of the data, thus helping the organizations to make decisions. Data mining is becoming popular because of its success in different areas. It has its applications in areas such as healthcare, finance, telecommunication, business, education and many more [2]. In data mining include we have different techniques are classifications, regression, summarization, clustering, association rule.

Association rule mining

Association rule mining is a popular and well researched method for discovering interesting relations between variables in large databases. Association rules are applied on large amount of data. It is intended to identify strong rules discovered in databases using different measures of interestingness. Association rules for discovering regularities between products in large-scale transaction data recorded by point-of-sale (POS) systems in supermarkets [3]. Association rules are applied on large amount of data. In business field, association rules are used to discover the frequently occurring pattern which will help in marketing and decision making. Also association rules are applicable to many other fields, and these include medical, market basket analysis, library etc [4]. For an example: “if a customer buys a conditioner, then he also probably buys shampoo (in the same transaction)”, the rule can be written as: {conditioner} → {shampoo} [5].

The Association Rule definition:

A set of items is referred as an itemset. A itemset that contains k items is a k-itemset.

The support s of an itemset X is the percentage of transactions in the transaction database D that contain X.

The support of the rule $X \square Y$ in the transaction database D is the support of the items set $X \square Y$ in D.

The confidence of the rule $X \square Y$ in the transaction database D is the ratio of the number of transactions in D that contain $X \square Y$ to the number of transactions that contain X in D [6].

Application of association rule mining

Association rule mining is used in different applications as it is very efficiency from data analysis. Some of the area that has adopted association rule mining are discussed below.

Market Basket Analysis

One of the most important typical examples of association rule mining is market basket analysis. This is a perfect example for illustrating association rule mining. This market basket analysis system will help the managers to understand about the sets of items are customers likely to purchase. This analysis may be carried out on all the retail stores data of customer transactions. These results will guide them to plan marketing or advertising approach. For example, market basket analysis will also help managers to propose new way of arrangement in store layouts. Based on this analysis, items that are regularly purchased together can be placed in close

proximity with the purpose of further promote the sale of such items together [7].

Medical Healthcare

In many healthcare settings, patients visit healthcare specialists occasionally and report multiple medical illnesses, or symptoms, at each encounter. A statistical modelling technique, called the Hierarchical Association Rule Model (HARM) developed by that forecasts a patient's possible future symptoms given the patient's present and past history of testified symptoms. This technique is a association model for selecting predictive association rules (such as "symptom 1 and symptom 2 \rightarrow symptom 3") from a large set of candidate rules. Because this method "borrows strength" using the symptoms of many similar patients, it is able to provide predictions specialized to any given patient, even when little information about the patient's history of symptoms is available [8].

Saptial census data

The information is related to the economic and population census and thus can be used in planning public services and business [2]. Censuses make a huge variety of general statistical information on society available to both researchers and the general public. Population and economic census information is of great value in planning public services (education, funds allocation, public transportation) as well as in private businesses (locating new factories, shopping malls, or banks, as well as marketing particular products). The application of data mining techniques to census data, and more generally, to official data, has great potential in supporting good public policy and in underpinning the effective functioning of a democratic society [9,2].

University

The association rules mining techniques to handle university admission system. The current system is modeled as a relational database that require some preprocessing to be suitable for mining algorithms of association rules. We aim to help the admission office better understand the nature of the system and the parameters that affect the decision making policy for accepting or rejecting the incoming students" applications [10].

II. METHODS OF ASSOCIATION RULE MINING

There exist we are using two different method and algorithms which are used to generate the association rules in data analysis. These are include

- Apriori alorithm.

- If/Then Rule.

Apriori algorithm

The Apriori algorithm is an algorithm for mining frequent itemsets for Boolean association rules. Apriori uses a "bottom up" approach where frequent subsets are extended one item at a time which is known as candidate generation, and group of candidates are tested against the data [11].

Frequent Itemsets : All the sets which contain the item with the minimum support(denoted by L_i for i th itemsets)

Apriori Property: Any subset frequent itemset must be frequent.

Join Operation: To find L_k a set of candidate k -itemsets generated by joining L_{k-1} with itself.

Pseudo-code

C_k : Candidate itemset of size k

L_k : frequent itemset of size k , $L_1 = \{ \text{frequent items} \}$; for ($k = 1$; $L_k \neq \emptyset$; $k++$) do

- $C_{k+1} = \text{candidates generated from } L_k$;

- for each transaction t in database do

Increment the count of all candidates in C_{k+1} that are contained in t ; endfor;

- $L_{k+1} = \text{candidates in } C_{k+1} \text{ with } \text{min_support}$

endfor;

return $\bigcup_k L_k$;

Apriori works in the principle of support and confidence. The algorithm recursively generates candidate itemsets for each n -itemset. Apriori is capable of generating a very large set of risk factors.

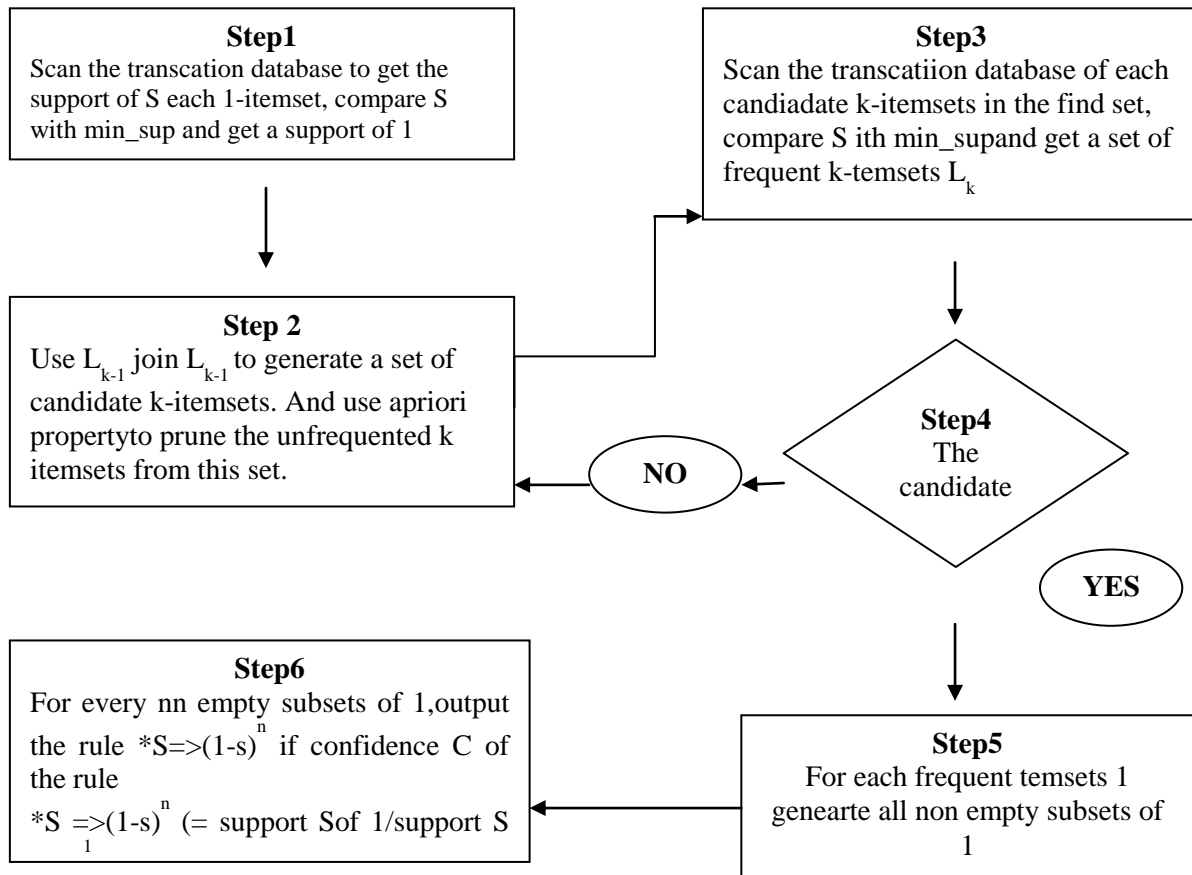


Fig.1: Steps to perform apriori algorithm [12]

IF THEN Rule

Classification is the process finding a model(or function) that describe distinguishes data classes or concepts. The model are derived based on the analysis f a set of training data(i.e., data objects for which is the class label is known). The model is used to predict the class label of objects for which the class label is unknown.

Rule-based classifier makes use of a set of IF-THEN rules for classification. We can express a rule in the following from.

IF condition THEN conclusion

Let us consider a rule R1,

and these can be used in prediction. Apriori is a recursive algorithm which extracts only one risk factor at a time and group them to produce the rule for prediction.It generates a very large set of ruleset. The limitation is that it generates candidate itemset at each level of processing [13].

Example of Apriori algorithm:

Let the minimum support be 20%

R1: IF age = youth AND student = yes

THEN buy_computer = yes

- The IF part of the rule is called **rule antecedent** or **precondition**.
- The THEN part of the rule is called **rule consequent**.
- The antecedent part the condition consist of one or more attribute tests and these tests are logically ANDed.

The consequent part consists of class prediction [15,16]

TID	List of item_IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

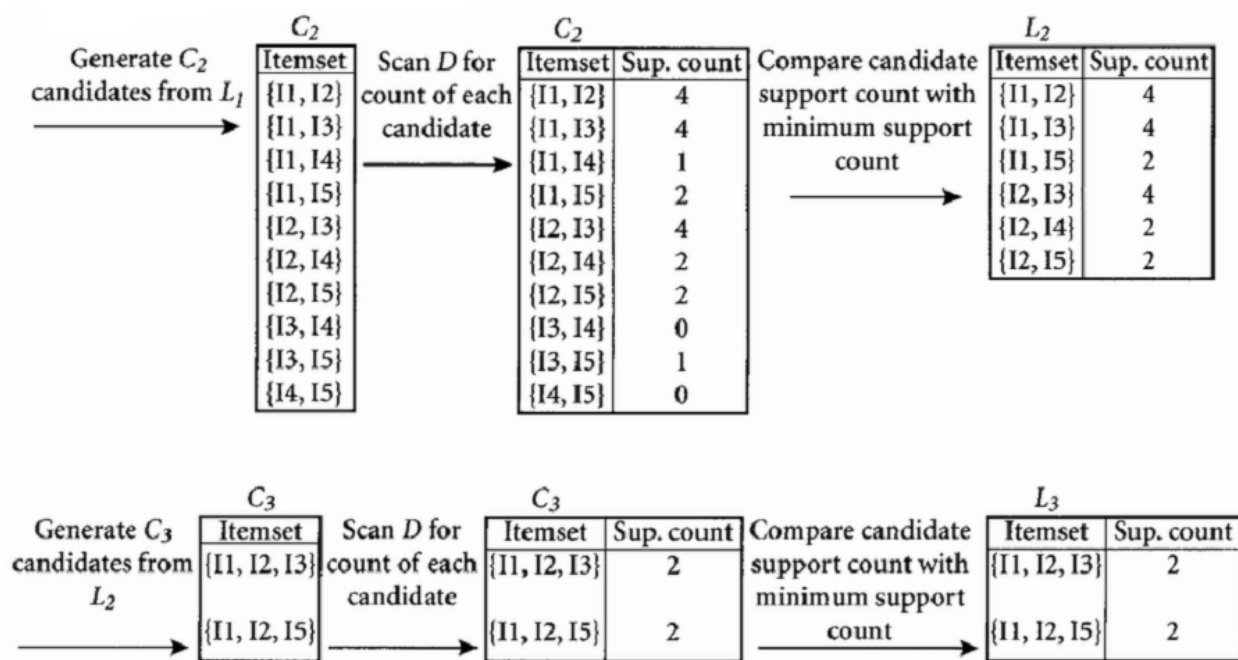
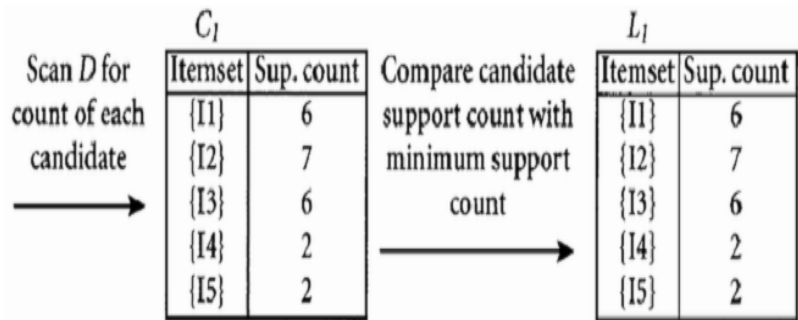


Fig.2: Example of Apriori Algorithm [1]

Decision Tree Induction

A decision tree is a flowchart like tree structure, where each internal node denote a test. On the attribute, each branch represents an outcome of the test, and each leaf node holds a Class label. The topmost node in a tree is the root node.

Example of IF THEN rule and generating the Decision tree induction

- IF age (X, "youth") AND income (X, "high")
- THEN class (X, "A")
- IF age (X, "youth") AND income (Y, "low") THEN class (X, "B")
- IF age (X, "middle_aged") THEN class (X, "C")
- IF age (X, "senior") THEN class (X, "C")

It is the concept of buys_computer, indicating whether an AllElectronics customer is likely to purchase a computer. Each internal node represent a test on an attribute. Each leaf node represent a class (either buys_computer= yes or no).

Algorithm: Generate_decision_tree

Generate a decision tree from the training tuples of data partition D

Input:

- Data partition, D, which is a set of training tuples and their associated class labels;
- Attribute_list, the set of candidate attributes;

□ Attribute_selection_method, a procedure to determine the splitting criterion that "best" partitions the data tuples into individuals classes. This criterion consists of spitting_attribute and possibly, either a split point or splitting subset.

Output: *A Decision tree*

Method:

```

Create a node N;
If tuples in D are all of the same class, C then
    return N as a leaf node labeled with the
class C;
if attribute_list is empty then
    return N as a leaf node labeled with
majority class in D://majority voting
apply attribute_selection_method (D, attribute_list)
to find the "best" splitting_criterion;
label node N with splitting_criterion;
if splitting_attribute is discrete-valued and
multiway splits allowed then
    attribute_list ← attribute_list – splitting
_attribute;
for each outcome j for splitting_criterion
    let Dj be the set of data tuples in D satisfying
outcome;
if Dj is empty then
    attach a leaf labeled with the majority
class in D to node N;
else
    attach the node returned by
Generate_decision_tree (Dj, attribute_list) to node N;
end for
return N;[17]

```

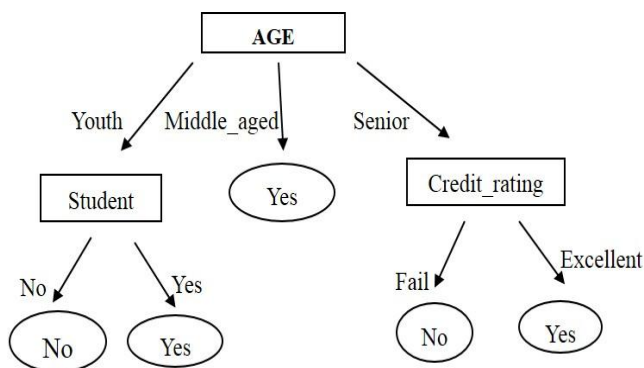


Fig.3 : Decision Tree Induction [16]

III. CONCLUSION

In this paper we have presented survey of various algorithm for association rule mining techniques are used for different applications. The main task of association rule used to find out the discovering pattern interestingness and correlation among different itemsets of database. Apriori, IF-THEN rule, Decision tree induction are some important of association rule mining techniques. Apriori algorithm used to extract the frequent itemset from large database. It is used for transaction large databases of transaction like supermarket, medical healthcare, protein sequence etc. IF-THEN rule is a process of finding model rule is used to predict the data object which for the class label is either known or unknown. It is useful for the uncover the relationships between unrelated data from relational database or other information repository. The paper conclude that give a basic idea to the mining association rule techniques about the algorithm which would yield better results.

REFERENCES

- [1] S.Mihika, N.Sindhu. "A Survey of Data Mining Clustering Algorithms" International Journal of Computer Applications (0975 – 8887) Volume 128 – No.1, October 2015.
- [2] K.Jagmeet, M.Neena. "AssociationRuleMining: A Survey" International Journal of Hybrid Information Technology Vol.8, No.7 239-242, 2015.
- [3] C. Ila, A. Mari Kirthima. "A Survey on Association Rule Mining Algorithms" International Journal of Mathematics and Computer Research [Volume 1 issue 10, 270-272, 2013.
- [4] A. Rajak and M. K. Gupta. "Association Rule Mining: Applications in Various Areas", International Conference on Data Management.
- [5] S.Sonia, Dr. Jyoti. "Multi-Level Association Rule Mining: A Review" International Journal of Computer Trends and Technology (IJCTT) – Volume 6 no.3, 2013.
- [6] EvgueniSmirnov "AssociationRule" <https://project.dke.maastrichtuniversity.nl/datamining/2013-Slides/lecture-06.ppt>.
- [7] G.Savi, M.Roopal. "A Survey on Association Rule Mining in Market Basket Analysis" International Journal of Information and Computation Technology. Volume 4, Number 4, pp. 409-414, 2014.
- [8] M. Renuka Devi, A. Baby sarojini. "Applications of Association Rule Mining in different databases" Journal of Global Research in Computer science, Volume 3, No. 8, 2012.
- [9] M. Donato, Francesca A. Lisi, A. Annalisa, S. Francesco. "Mining Spatial Association Rules in Census Data: A Relational Approach"
- [10] M. Abdul Fattah, Mohammed M. Fouad, Philip S. Yu, Tarek F. Gharib "Discovery of Association rules from University admission System data" I.J. Modern Education and Computer Science, 2013, 4, 1-7, 2013.
- [11] M.E. Anuradha Bhatia. Computer Engineering, "Big Data Analytics-Apriori Algorithm" <http://www.anuradhabhatia.com>.
- [12] Apriori algorithm" <https://www.slideshare.net/INSOFE/apriori-algorithm-36054672>.
- [13] J. Omana, S. Monika, B. Deepika. "Survey on efficiency of Association Rule Mining Techniques" International Journal of Computer Science and Mobile Computing IJCSMC, Vol. 6, Issue. 4, 5–8, 2017.

- [14] "Mining Association Rules in large database"
[http://www1.pu.edu.tw/~ytwang/docs/DM/ Assoc1.ppt](http://www1.pu.edu.tw/~ytwang/docs/DM/Assoc1.ppt).
- [15] "Rule based data mining" https://www.tutorialspoint.com/data-mining/dm_rbc.htm.
- [16] H. Jiawei, K. Michrine, P. Jian. "Data Mining Concepts and Techniques.
- [17] "Decisiontreeinduction"
<https://www.researchgate.net>publication>.

Authors Profile

The author M. Sumitha is working as an Assistant Professor in the BCA Department, Sri Akilandeswari Women's college – Wandiwash. She having three years of teaching experience in the field of Computer Applications. She has completed her BCA from Sri Akilandeswari Women's college. M.C.A., from The Adhiparasakthi Engineering College, Melmaruvathur. M.Phil., from Sun Arts and Science College, Tiruvannamalai. She has published two papers from national journals. She has participated and presented various National and International conference.



The author P.S. Suganya is working as an Assistant Professor in the BCA Department, Sri Akilandeswari Women's college – Wandiwash. She having 5 years of teaching experience in the field of Computer Applications. She has completed her B.Sc from Theivanai Ammal College for Women, Villupuram, in 2006. M.Sc., from Government Arts College for Men Nandhanam, (Autonomous), Chennai in 2008. B.Ed from Vinayaka Mission's Faculty of Education, Puducherry, in 2009. M.Phil., from Theivanai Ammal College for Women, Villupuram, in 2010. M.Ed from National College of Education, Neively, in 2015. She has published One paper from national journals. She has participated and presented various National and International conference.

