ANALYSIS ON APRIORI ALGORITHM BY APPLYING CORRELATION THRESHOLD ON MARKET BASKET

P INDUMATHI, K.SOWJANYA
Mtech Scholar, Assistant Professor
Department of of CSE
ISTS Women's Engineering College, Rajanagaram, Rajamahendravaram, A.P, India.

ABSTRACT: -

Frequent item set mining becomes a challenging job in the field of data mining where each algorithm introduced in the past has their own advantages and disadvantages. Frequent item set and creating association rules becomes important in transactional data where we are going to represent data that contains a set of entries where each entry composes of items, from where based on the support and confidence parameters we need to create strong association rules which can be recommended to the higher officials for better decision making of the organization. We propose a new approach to classical Apriori algorithm by adding correlation threshold approach which is capable of producing more frequent item sets in lesser time by having a single scan to the entire transaction data by eliminating infrequent items from the data we are able to produce even strong association rules as compared to existing Apriori algorithm. By using probabilistic arrays along with Correlation approach, we prove our proposed approach is better than the classical approach of finding frequent item sets and finally to find most interesting items we are going to apply association rule mining.

1. INTRODUCTION

What is Data Mining?

In everyday life, information is collected almost everywhere. For example, at supermarket checkouts, information about customer purchases is recorded. When payback or discount cards are used, information about customer purchasing behavior and personal details can be linked. Evaluation of this information can help retailers devise more efficient and modified marketing strategies. The majority of the recognized organizations have accumulated masses of information from their customers for decades. With the e-commerce applications growing quickly, the organizations will have a vast quantity of data in months not in years. Data Mining, also called as Knowledge Discovery in Databases, is to determine the trends, patterns, correlations and anomalies in these databases that can assist to create precise future decisions. Physical analysis of these huge amount of information stored in modern databases is very difficult. Data mining provides tools to reveal unknown information in large databases which are already stored. A well-known data mining technique is Association Rule Mining. It is able to discover all the interesting relationships which are called as associations in a database. Association rules are very efficient in revealing all the interesting relationships in a relatively large database with a huge amount of data The large quantity of information collected through the set of association rules can be used not only for illustrating the relationships in the database, but also for differentiating between different kinds of classes in a database. Association rule mining identifies the remarkable association or relationship between a large set of data items. Data mining is the principle of sorting through large amounts of data and picking out relevant information. It is usually used by business intelligence organizations, and financial analysts, but it is increasingly used in the sciences to extract information from the enormous data sets generated by modern experimental and observational methods. It has been described as "the nontrivial extraction of implicit, previously unknown, and potentially useful information from data" and "the science of extracting useful information from large data sets or databases". A pattern represents knowledge if it is easily understood by humans; valid on test data with some degree of certainty; Data mining systems can be classified according to the kinds of databases mined, the kinds of knowledge mined, the techniques used or the applications adapted. In general data mining tasks can be classified as descriptive and predictive. Metadata, or data about a given data set, are often expressed in a condensed data mine-able format, or one that facilitates the practice of data mining. Common examples include executive summaries and scientific abstract

1.1 HOW DOES DATA MINING WORK?

It analyzes relationships and patterns in stored transaction data based. Any of four types of relationships are sought.

Classes: Stored data is used to locate data in predetermined groups.

Clusters: Data items are grouped according to logical relationships.

Associations: Data can be mined to identify associations

Sequential patterns: Data is mined to anticipate behavior patterns and trends

1.2 DATA MINING MODELS

There are several different types of models

Predictive Models: These types of models predict how likely an event is.

Summary Models: These models summarize data.

Network Models: To represent data by nodes and links.

Association Models: To find and characterize item set co-occurrences.

2. DATA MINING ARCHITECTURE

To best apply this advanced technique, it must be fully integrated with a data warehouse and interactive analysis tools. Many data mining tools currently operate outside of the warehouse, requiring extra steps for extracting, importing, and analyzing the data. Refer to Fig 1.1. The resulting analytic data warehouse can improve business processes throughout the organization. The design represents a fundamental shift from conventional decision support systems. Rather than simply delivering data to the end user through query and reporting software, the Advanced Analysis Server applies users' business models directly to the warehouse and returns a proactive analysis of the information.

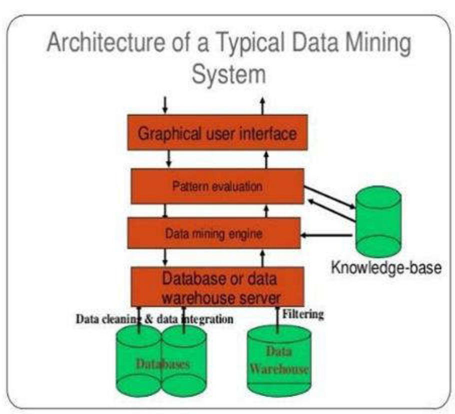

Fig: Data Mining Architecture

2.1 USES OF DATA MINING

Businesses can identity customer groups that are more profitable to companies and would start to direct all of their effects into making products for only that target market. Data mining can also be helpful to Human Resources in identifying the characteristics of their most successful employees. Information obtained, such as universities attended by highly successful employees, can help HR focus recruiting efforts accordingly. Another example of data mining, often called the market basket analysis, relates to its use in retail sales. If a clothing store records the purchases of customers, a data mining system could identify those customers who favor silk shirts over cotton ones.

3. ASSOCIATION RULES

Association rule mining finds interesting associations and/or correlation relationships among large set of data items. Association rules show attributes value conditions that occur frequently together in a given dataset. Association rules provide information of this type in the form of "if-then" statements. These rules are computed from the data and, unlike the if-then rules of logic, association rules are probabilistic in nature. This application of association rule learners is also known as market basket analysis. As with most data mining techniques, the task is to reduce a potentially huge amount of information to a small, understandable set of statistically supported statements. Questions such as "if a customer purchases product A, how likely is he to purchase product B?" are answered by association-finding algorithms For example: 60% of those who buy comprehensive motor insurance also buy health insurance; Probability that particular items are purchased together. X □Y where X n Y = 0

3.1 RULES INVOLVING MORE THAN ONE DIMENSIONS OR PREDICATES

• **(Single dimensional)**

buys (X, "IBM Laptop Computer")□buys (X, "HP Inkjet Printer")

• **(Multi Dimensional- Inter dimension Association Rule)**

age (X,"20..25") and occupation (X, "student")□buys (X, "HP Printer")

• **(Multi Dimensional- Hybrid dimension Association Rule)**

age (X, "20 _25") & (X, "IBM Laptop Computer") □buys (X, "HP Printer")

3.2. ASSOCIATION RULES ARE USED IN DIFFERENT FIELDS,SUCH AS

•Identify unexpected shopping patterns in supermarkets. Making appropriate offers to each visitor.
• Optimize web site profitability by making appropriate offers to each visitor
• Predict customer response rates in marketing campaigns.
•Define groups for marketing purposes profitable and unprofitable customers.
• Predicts which customers are likely to switch to an alternative supplier in future
• Improve yields in complex production by finding unexpected relationships.

3.3. ASSOCIATION RULES TERMINOLOGY

Item Set

Set of distinct items in a given set .Let I= (i1, i2,..., in) be a set of n distinct items. Each i, is an item .I represents it set

Item Set Length

It is the number of items in an Item Set.

Item Sets with k distinct elements are referred to as kitem sets.

Candidate Item Set

Set of unique items, which are to be analyzed

Transaction (T)
It may be defined as particular operation in which items are involved **Transaction Database (D)**
A database is a set of transactions.
Transaction Size
The unique number of items present in a transaction
Support Count
It is the total frequency of appearance of a particular pattern in a database
If T is a transaction in database D and X is an item set then T is said to support X, if X ⊆ T, hence support of
X is the fraction of transactions that support X.
Let D={T 1,T2,T3 }
Let T 1= {bread, jam, sugar} T2=(bread, jam}
T3,={jam, sugar}
 Candidate item set = {bread, jam, sugars}
 Number of transactions =3
 Bread support = 2/3 , jam support =3/3, sugar support -1 /3
Frequent Item set.
An item set whose support count is greater than or equal to the minimum support threshold specified by the user.
Infrequent Item set: An item set whose support count is less than the minimum support threshold specified by the user
Pruning:
Removal of the unwanted data item sets /infrequent item sets
Downward Closure Property (Basis for Top-down Search):
"If an item set is frequent then all its subsets must be frequent."
Upward Closure Property (Basis for Bottom-up Search):
If an item set is infrequent then all its supersets must be infrequent_"
Maximum Frequent Set
An item Set which is the set of all Maximal Frequent sets.
Association Rule: is the rule of the form R : X 3 Y, where X and Y are two non-empty and nonintersecting item sets.
Confidence is defined as Support of X∪Y / Support of X. age (X, 'under 12') and gender(X, 'male') buys(X, 'comic book')
Maximum Frequent Candidate Set (MFCS): is a minimum cardinality set of item sets such that the union of all the subsets of elements contains all frequent item sets but does not contain any infrequent item set, i.e. it is a minimum cardinality set satisfying it conditions. FREQUENT⊆ {2X I X € MFCS} INFREQUENT ⊆{2X | X € MFCS} Where FREQUENT and INFREQUENT', stand respectively for all frequent and Infrequent items

4. MARKET BASKET ANALYSIS

Affinity analysis is a data analysis and data mining technique that discovers co-occurrence relationships among activities performed by (or recorded about) specific individuals, groups, or in general an object identifier in a computer science context. An intuitive and well known example of affinity analysis is market basket analysis in the retail business. Market basket analysis identifies customers purchasing habits. It provides insight into the combination of products within a customer's

'basket'. The term 'basket' normally applies to a single order. However, the analysis can be applied to other variations. We often compare all orders associated with a single customer.
Ultimately, the purchasing insights provide the potential to create cross sell propositions:
which product combinations are bought, when they are purchased, and in what sequence. Developing this understanding enables businesses to promote their most profitable products. It can also encourage customers to buy items that might have otherwise been overlooked or missed. In this case retailers use it to understand the purchase behavior of groups of customers, and use it for cross-selling, store design, discount plans and promotions. A widely used example of cross selling on the internet with market basket analysis is Amazon.com's use of suggestions of the type Customers who bought book A also bought book B". In the case of retailers with stores, market basket information enables the retailer to understand the buyer's needs and rewrite the store's layout accordingly, develop cross-promotional programs, or even capture new buyers (much like the cross-selling concept). This even helps retailers avoid discounts, e.g. knowing that people who buy more than 12 cans of Pepsi also usually buy 12 cans of ?up during the same store trip, allows them not to discount both drinks at the same time, for the sale of one item usually leads to the sale of another.

5. APRIORI ALGORITHM

In computer science and data mining, Apriori is a classic algorithm for learning association rules. Apriori is designed to operate on databases containing transactions (for example, collections of items bought by customers. Or details of a website frequentation). Other algorithms are designed for finding association rules in dat having no transactions or having no timestamps (DNA sequencing). As it is common in association rule mining, given a set of item sets (for instance, sets of retail transa" each listing individual items purchased), the algorithm attempts to find subsets which are common to at least a minimum number C (the cutoff, or confidence threshold) of the item sets. Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation), and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found. Apriori uses breadth-first search and a hash tree structure to count candidate item sets efficiently. It generates candidate item sets of length k from item sets of length k - 1. Then it prunes the candidates which have an infrequent sub pattern. According to the downward closure lemma, the candidate set contains all frequent k-length item sets. After that, it scans the transaction database to determine frequent item sets among the candidates. For determining frequent items quickly, the algorithm uses a hash tree to store candidate item sets. This hash tree has item sets at the leaves and hash tables at internal nodes Note that this is not the same kind of hash tree used in for instance p2p systems Apriori, while historically significant, suffers from a number of inefficiencies or trade-offs, which have spawned other algorithms.

5.1 APRIORI ALGORTHM PSEUDO CODE

Apriori algorithm basically comprises of two passes

 frequent itemset is an itemset whose support is greater than some user-specified minimum support (denoted , where k is the size of the itemset)

 A candidate itemset is a potentially frequent itemset (denoted , where k is the size of the itemset)

Pass 1

1. Generate the candidate itemsets in 1,
2. Save the frequent itemsets in 1

Pass k.

1. Generate the candidate itemsets in Ck from the frequent itemsets in −1
1. Join −1
with −1 ,as
follows:
Insert into
select p.1, . 1,......, . −1, . −1
from −1 , −1
where p.
2. Generate all (k-1)- subsets from the candidate itemsets in
3. Prune all candidate itemsets from where some (k-1)- subset of the candidate itemsets is not in the frequent itemset −1
3. Scan the transaction database to determine the support for each candidate Itemset in
4. Save the frequent itemsets in

Example:

Assume the user-specified minimum support is 40%, and then generate all frequent item sets.
Given the transaction database shown below

Table 3.1 Sample transaction table

TID	A	B	C	D	E
1	1	1	1	0	0
2	1	1	1	1	1
3	1	0	1	1	0
4	1	0	1	1	1
5	1	1	1	1	0

Table 3.2 Pass 1(Candidate generation c1)

Pass 1

Itemset X	supp(X)
A	?
B	?
C	?
D	?
E	?

Pass 1

Itemset X	supp(X)
A	100%
B	60%
C	100%
D	80%
E	40%

Table 3.3 Pass 2 (Candidate generation set c2)

C2		L2		L2 after saving only the frequent item sets	
Item set X	**supp (X)**	**Item set X**	**supp (X)**	**Item set**	**supp (X)**
A,B	?	A,B	60%	A,B	60%
A,C	?	A,C	100%	A,C	100%
A,D	?	A,D	80%	A,D	80%
A,E	?	A,E	40%	A,E	40%
B,C	?	B,C	60%	B,C	60%
B,D	?	B,D	40%	B,D	40%
B,E	?	B,E	20%	C,D	80%
C,D	?	C,D	80%	C,E	40%
C,E	?	C,E	40%	D,E	40%
D,E	?	D,E	40%		

Candidate generationc2 Pruned set in pass2Nothing pruned since all subsets of these item sets are infrequent

Table 3.4 Pass 3 (Candidate generation c3)

To create C3 only look at items that have the same first item (in pass k, the first k-2 items must match)

	Itemset X	**supp(X)**
join AB with AC	A,B,C	?
join AB with AD	A,B,D	?
join AB with AE	A,B,E	?
join AC with AD	A,C,D	?
join AC with AE	A,C,E	?
join AD with AE	A,D,E	?
join BC with BD	B,C,D	?
join CD with CE	C,D,E	?

Candidate generation c3

Itemset X	**supp(X)**
A,B,C	?
A,B,D	?
A,C,D	?
A,C,E	?
A,D,E	?
B,C,D	?
C,D,E	?

Itemset X	**supp(X)**
A,B,C	60%
A,B,D	40%
A,C,D	80%
A,C,E	40%
A,D,E	40%
B,C,D	40%
C,D,E	40%

Pruning eliminates ABE since BE is not frequent

Pass 4 First k-2=2 items must match in pass k = 4

Table 3.5 Pass 4 (Candidate generation c4)

	Item set X	supp(X)
combine ABC with ABD	A,B,C,D	?
combine ACD with ACE	A,C,D,E	?

Pruning:

1) For ABCD we check whether ABC, ABD, ACD, BCD are frequent. They are in all cases, so we do not prune ABCD.
2) For ACDE we check whether ACD, ACE, ADE, CDE are frequent. Yes, in all cases, so we do not prune ACDE

Table 3.6 Final frequent item set

Item set X	supp(X)
A,B,C,D	40%
A,C,D,E	40%

Final frequent item set Both are frequent

Pass 5: For pass 5 we can't form any candidates because there aren't two frequent 4-itemsets beginning with the same 3 items.

6. CORRELATION THRESHOLD ON APRORI ALGORITHM

Correlation threshold finds its application in candidate item set generation. In Modified Apriori, we incorporate correlation threshold for finding strong Association rules between the itemsets. The correlation threshold is a value Between 0 and 1. If the value is 1, then the attributes are highly related to each Other. While a value close to zero shows the dataset as independent. This Correlation confirms the presence of all itemset appearing in traditional Apriori in proposed algorithm. Algorithm begins by scanning the data, D. Suppose there are n elements in D, algorithm initializes a probabilistic array, PA[n]. After the first scan, probability of occurrence of each 1 itemsets appearing in the transaction is entered into PA[n]. From the probabilistic array, correlation threshold is found out using the equation (1). This acts as the minimum support threshold. Those itemsets whose threshold is below the correlation value is eliminated. This step repeats iteratively and 2-itemsets are generated from PA[n] by calculating new correlation threshold. Repeated transaction data scan is being avoided as the candidate itemset generation is done directly from the transaction data and not by continuous scanning of D. The process is continued till all attributes in the transaction data is scanned.

6.1 EXPLANATION

In the Co Apriori algorithm we will calculate the ItemProbability Of each n-itemset. And then we will add the all ItemProbabilities of the n-itemsets and divide the obtained total value with the number of itemsets.And the obtained value is the new Support value which is generated dynamically .And the remaining itemsets are pruned with the new Support value .And further the Co-Apriori calculation is dine for the remaining n-Itemsets. And finally the detailed
report of the pruned and existing items in the given transaction data is obtained.

6.2 CALCULATIONS

ItemCount : It is the total count of an item in the total transaction data.

Item Probability: It is the ratio of the itemcount to the total transactions in the transaction data.

Support Value: It is calculated as the ratio of the total sum of Item Probabilities of individual item of each nitemset
to the total number of itemset according to the nth level.

Example:

Transaction Data:

Tid	TData
1	A,B,C,E,F
2	A,C,G
3	E,G
4	A,C,D,E,G
5	A,C,E,F,G
6	F,G
7	A,B,C,E,F,G
8	A,C,D
9	A,C,E,G,H
10	A,C,E,F,H

n-Itemsets generations for the given Transaction data.

1-Itemset:

Frequent Itemsets	IC	IP
H	2	0.2
G	7	0.7
F	5	0.5
E	7	0.7
D	2	0.2
C	8	0.8
B	2	0.2
A	8	0.8

2-Itemset:

Frequent Itemsets	IC	IP
D,H	0	0.0
C,H	2	0.2
C,D	2	0.2
B,H	0	0.0
B,D	0	0.0
B,C	2	0.2
A,H	2	0.2
A,D	2	0.2
A,C	8	0.8
A,B	2	0.2

3-Itemset:

Frequent Itemsets	IC	IP
C,D,H	0	0.0
B,D,H	0	0.0
B,C,H	0	0.0
B,C,D	0	0.0
A,D,H	0	0.0
A,C,H	2	0.2
A,C,D	2	0.2
A,B,H	0	0.0
A,B,D	0	0.0
A,B,C	2	0.2

4-Itemset:

Frequent Itemsets	IC	IP
B,C,D,H	0	0.0
A,C,D,H	0	0.0
A,B,D,H	0	0.0
A,B,C,H	0	0.0
A,B,C,D	0	0.0

5-Itemset:

Frequent Itemsets	IC	IP
A,B,C,D,H	0	0.0

6.3. DETAILED REPORT

*********** Itemset Generation Using Correlation Apriori **********

Actual Itemsets of the File:
H#G#F#E#D#C#B#A#
[Support Factor Value 0.6375000000000001]
Level-1
H#D#C#B#A#
[Support Factor Value 0.0]
Level-1
D,H#C,H#C,D#B,H#B,D#B,C#A,H#A,D#A,C#A,B#
[Support Factor Value 0.64]
Level-2
C,H#C,D#B,C#A,H#A,D#A,C#A,B#
[Support Factor Value 0.0]
Level-2
C,D,H#B,D,H#B,C,H#B,C,D#A,D,H#A,C,H#A,C,D#
A,B,H#A,B,D#A,B,C#
[Support Factor Value 0.56]
Level-3
A,C,H#A,C,D#A,B,C#
[Support Factor Value 0.0]
Level-3
B,C,D,H#A,C,D,H#A,B,D,H#A,B,C,H#A,B,C,D#
[Support Factor Value 0.0]
Level-4
A,B,C,D,H#

7. EXISTING SYSTEM

Apriori Algorithm is found to be better for association rule mining. Among the methods discussed for data mining, Apriori Algorithm is found to be better for frequent itemset and association rule mining. Still there are various difficulties faced by Apriori Algorithm.
In Apriori Algorithm Support value is same for all nitemsets and also the Time complexity increases. Still
there are various difficulties faced by Apriori Algorithm.

7.1 DEMERITS

It scans the data a number of times. Every time additional choices will be created during the scanning process and creates additional work for the data to search. Time complexity increases, so the algorithm will not result in better result. Therefore, it is required to improve, or re-design algorithms.

8. PROPOSED SYSTEM

To overcome disadvantages of Apriori Algorithm we proposed a modified Algorithm called Correlation threshold on Apriori Algorithm and also Association Rule Mining For the itemsets. It produces more frequent item sets in lesser time to reduce the time complexity of item sets by having a single scan to the entire data, by eliminating infrequent items from the database to produce even strong association rules as compared to existing Apriori algorithm.

8.1. MERITS

 Since the Support values depends on previous phase, dynamically itemsets changes unlike Classical Apriori which is having constant value

 Time Complexity comes down to a great extent.

 Generate More Frequent Item sets and Association Rules as Compared to Apriori.

9. MODULES IN PROJECT

1. Uploading of Market Dataset:

Here we collect transactional data of a market for market basket analysis that contains a set of transactions done by different customers. And we upload the transaction data and save for further purpose.

2. Preprocessing:

The dataset that contains information from past may not be accurate and may contain some noisy data, which cannot be tested in further part of Data Mining techniques, hence we require to remove noisy from it by applying normalization process.

3. Application of Algorithms:

Here the data which is normalized will be tested with existing and proposed algorithms where for existing algorithm (apriori) we need to give support factor based on which frequent items will be pruned, and for proposed algorithm the support factor is a dynamic concept which changes at every level hence helping us in producing more frequent itemsets.

3.1 Apriori:

Apriori is an algorithm for frequent item set mining and association rule learning over transactional **databases**. It proceeds by identifying 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. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the **database** this has applications in domains such as **market basket analysis.**

3.2 Correlation Apriori:

Correlation apriori means applying correlation threshold on apriori algorithm. By applying the correlation threshold, the support values will change dynamically i.e., the values are changed based on the before final values. as in the simple apriori algorithm the threshold value is same for all the remaining data. After applying the correlation threshold, the value is dynamically changed. Which 0decreases the time complexity and also it generate More Frequent Item sets and Association Rules as Compared to Apriori.

4. Association rule Generation:

Here once frequent itemsets are generated we need to apply association rule mining which asks for confidence that specifies the relationship between one item and another item.

5. Comparative Results:

Here we will compare the number of itemsets of normal apriori and co apriori and also we show the graph presentation about no of itemsets generated between existing and proposed algorithms.Based on the result obtained the of the proposed algorithm will have higher efficiency than the existing system.This will help in analysis of the specific product sale in the market. And this will be an important application for the stores and companies to increase their sale.

10. CONCLUSION

The proposed algorithm suggests a correlation approach to the traditional Apriori. Method enhances the efficiency of algorithm by evaluating the number of candidate itemsets generated. Proposed scheme creates more number of frequent itemsets in lesser time. The time complexity was found to reduce from O(en) to O(n). Time complexity was reduced due to the abatement of the number of database scan. Through pruning the infrequent itemsets and by retaining the frequent ones strong rules are created. Database scan which was fully depended on the length of frequent itemset was supplanted by the introduction of probabilistic array. Results confirm that, with extended inter-transactional association, absolute and remarkable relations were able to mine. The traditional association rules are Intra-transactional since they only capture associations among items within the same transactions, where the notion of the transaction could be the items bought by the same customer, the atmospheric events that take place at the same time, and so on. However, an Inter-transactional association rule can represent not only the associations of items within transactions, but also the associations of items among different transactions along certain dimensions.

As the existing apriori algorithm considers only one support value for the generating frequent itemsets. Where the proposed Co-Apriori Algorithm generates the support value dynamically based on the Item Probability count and then prunes the frequent itemsets based on the obtained support value. And, finally the detailed report is obtained. In the last phase these Association Rule Mining is done and the Association rules are generated .And also the comparison for both algorithms are shown in the graphical representation.

REFERENCES

[1] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," In Proc. of VLDB, pp. 487– 499, 1994.

[2] L. Zeng, Q.He, and Z. Shi, "Parallel implementation of apriori algorithm based on mapreduce," In Proc. Of SNPD, pp.236-241, Aug 2012.

[3] H. Wu, Z. Lu, L. Pan, R. Xu, and W. Jiang, "An improved apriori-based algorithm for association rules mining," In Proc. of sixth international conference on fuzzy systems and knowledge discovery, pp. 51–55, 2009.

[4] S. Tao and P. Gupta, "Implementing improved algorithm over apriori data mining association rule algorithm," IJCST, vol. 3, pp. 489 – 493, Jan-Mar 2012.

[5] K. Suneetha and R. Krishnamoorti, "Advanced version of apriori algorithm," In Proc. of ICIIC, vol. 5, pp. 238–245, Aug 2010.

[6] L. Shi, J. niu Bai, and Y. lin Zhao, "Mining association rules based on apriori algorithm and application," In Proc. of IFCSTA, vol. 3, pp. 141– 145, Dec 2009.