ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

A Survey on Association Rule Mining using Apriori Algorithm

[1] Sneha R. M, [2] B. S. Varsha, [3] G. K. Meghana, [4] Anusha, [5] S Sahitya
[1][2][3][4][5] Dept. of CSE, RYMEC, Bellary, Karnataka (India).

Abstract: **- Data mining refers to the process of extracting useful and interesting knowledge from large set of data. There are many aspects like clustering techniques, classification, association rule mining etc. Among these, association rule mining is the important aspect for data mining. The algorithm's performance directly effects the efficiency , integrity and effectiveness of ultimate data mining results. Apriori algorithm is one of the influential algorithm for mining association rules. The basic idea of this Apriori algorithm is to identify all the frequent itemsets.**

Keywords: **- Apriori algorithm, association rule.**

I. INTRODUCTION

Here is discussing about Association rule mining using Apriori algorithm. To find association rules, there are several algorithms. One of the algorithm is the Apriori algorithm. Apriori algorithm is the basic and well known algorithm. It is proposed by R. Agrawal and R. Srikant in 1994. Apriori algorithm is used for mining frequent item set. It works based on level –wise searches in huge database or data repository. The limitation of original Apriori algorithm is that it searches whole or entire database, data repository for frequent item sets. Association is a data mining technique that discovers probability of the co-occurrence of the items in a database. Association rules are often used to analyze in business for transaction purpose. Now a days the size of data is gradually increasing. Here the problem arises, apriori algorithm takes a lot of time to scan the whole databases and generate large number of candidate itemset. Hence the efficiency of the system will be low. Some of the researchers improved the apriori algorithm that reduces the scanning period when compared to original Apriori algorithm. The consumed time for scan will be reduced up to 67.38%. Apriori algorithm is an iterative way to find frequent itemset. It works on the principle "Any subset of a frequent itemset must be frequent". Association rules must satisfy both minimum support and confidence.

II. LITERATURE SURVEY

Mohammed Al-Maolegi, Bassam Arkok [1] : describes that using data mining technique extract the exact information for business people's to achieve their target goals.

Sheila A.Abaya [2]: describes that using , an efficient approach is the use of weight factor and utility for mining high utility patterns.

Charanjeet kaur[3] : describes that predict the occurrence of an item based on the occurrence of the transaction.

ASSOCIATION RULE

An association rule defines an implication expression A->B , where A and B are two disjoint itemsets, A intersection B=null. The Association rule can be measured in terms of support and confidence. Support defines how often rule applicable to the dataset. Confidence defines how frequently items in B appears in transactions that contain A. Association rule mining consists of two process are as follows-

Find all frequent itemsets : An item sets can occur frequent as pre-determined minimum Support (min_sup).

Generate strong association rules from the frequent itemsets: It defines as it satisfy minimum support and minimum confidence.

Association rule mining problem: The set transaction should have support >=minsup and confidence >= minconf ,considering these as threshold.

APRIORI ALGORITHM:

$C_1 \leftarrow$ init-pass(T)
$F_1 \leftarrow \{f \mid f \in C_1, f.count / n >= minsup\}$;
//number of transaction T
for($k=2; F_{k-1} \neq \Phi; k++$)
  $C_k \leftarrow$ candidate-gen(F_{k-1});
  for each transaction $t \in T$ do
    for each candidate $c \in C_k$ do
      if c is contained in t then

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

c.count++;
end
end
$F_k \leftarrow \{ c \in C_k \mid c.count / >= minsup\}$
end

return $F \leftarrow \underline{\cup_k F_k};$

Algorithm for frequent generation itemset.

The working of the algorithm to find frequent itemset using apriori algorithm is shown below:

- The algorithm initially makes a single pass over the data set to determine the support of each item.
- The algorithm iteratively generate a new candidate k-itemsets using candidate-gen (Fk-1).
- After analyzing the support and confidence ,the algorithm eliminates all candidate itemsets whose count less than minsup.
- The algorithm terminates when there are no new frequent itemset generated.

1. An improved Apriori Algorithm for Association Rules of Mining:

An improved apriori algorithm takes less time to scan the whole databases very efficiently and effectively. It is used to extract frequent itemsets from association rule for discovering the knowledge.
Improvements of apriori algorithm are as follows
1. Adopts new database mapping and avoids repeatedly scanning the database or data repository.
2. Eliminates candidate itemset in order to improve join efficiency.
3. Adopts overlap technique to count support at high efficiency.

The improvement of algorithm as follows-

L1 = find_frequent_1_itemsets (T);
For (k = 2; Lk-1 ≠Φ; k++)
{
Ck = candidates generated from Lk-1;
x = Get _item_min_sup(Ck, L1);
Tgt = get_Transaction_ID(x);
For each transaction t in Tgt Do
Increment the count of all items in Ck that are found in Tgt;
Lk= items in Ck ≥ min_support;
end
}

2. Example for improvement Apriori algorithm

Transaction ID	ITEMSET
T1	{I1,I3,I4}
T2	{I2,I3,I5}
T3	{I1,I2,I3,I5}
T4	{I2,I5}

First candidate Generation

ITEMSET	COUNT
I1	2
I2	3
I3	3
I4	1
I5	3

After a first candidate itemset generation min_sup=2.
Eliminated I4.

Second candidate Generation

ITEMSET	COUNT
{I1,I2}	1
{I1,I3}	2
{I1,I5}	1
{I2,I3}	2
{I2,I5}	3
{I3,I5}	2

After the second candidate itemset generation min_sup=2.
Eliminated [{I1,I2} and {I1,I5}]

Third candidate Generation

ITEMSET	COUNT
{I2,I3,I5}	2

Support{I1,I3}=2 or 50% or 2 of 4 customers.
Conf({I2,I3}?=
Sup({I2})=3
Sup ({I3})=3
Sup({I2,I3})=2
Then conf({I2,I3})=2/3=0.66

III. CONCLUSION

Association rule using Apriori algorithm is very useful to find the frequent itemset. The Business people uses this data mining technique to reach their target. An improved version of Apriori algorithm increases high efficiency

and consumes less time to search the whole database or data repository.

REFERENCES

1. Mohammed Al-Maolegi1, Bassam Arkok2 AN IMPROVED APRIORI ALGORITHM FOR ASSOCIATION RULES.

2. Sheila A. Abaya Association Rule Mining based on Apriori Algorithm in Minimizing Candidate Generation.

3. Charanjeet Kaur Association Rule Mining using Apriori Algorithm: A Survey