

Critique on the Association Rule in Data Mining

A.KALAIARASAI

Bishop Heber college

Trichy-17

Mr. K.MOHAMED AMANULLA

Asst. Professor

Bishop Heber college

Trichy-17

Abstract-- Data mining techniques are used to extract frequent patterns, from massive amount of data in a form of data ware house. Apriori algorithm has been vital algorithm in association rule mining. Apriori algorithm is mainly used to find a frequent itemset in a large amount of datasets. Apriori algorithm basically works in two steps. In first step candidate itemset is generated using linking process and in next step frequent itemset from those candidate itemset is found based on minimum support count scanning database.

Index Term: Apriori algorithm, Association Rule Mining, Frequent pattern, FP-Growth.

1.Introduction

Data Mining is the process of analyzing data from different perspective and summarizing it into useful information. It allows users to analyze data from many different dimensions. Data Mining is used to extract the information from any system by analyzing the present in the form of data. Rest of the paper is organized as follow: Section 2 gives the review of literature section 3 the author will discuss the different association rule advantage and disadvantage section 4 author will conclude the result.

2. Literature Review

In this section the biographer has discussed some papers which had been previously undertaken in the field of association rule mining, Apriori algorithm. The Improved Apriori mining algorithm in bank customer segmentation to compare the traditional Apriori algorithm and Improved algorithm and conclude the result [3]. Education data mining using Improved Apriori algorithm to analysis the student performance whether they are very good, good, average, below average, bad[1]. Mining Efficient association rule through Apriori algorithm using attributes and comparative analysis of various association rule and conclude the result is the Apriori algorithm is perform better than the another association rule algorithm[2]. Credit Assessment of bank customer by Fuzzy Export system based on Rules Extracted from Association Rules to using the Apriori algorithm and CRISM-DM methodology analysis good and bad risk [4]. Execution of Apriori Algorithm of Data Mining Directed towards Tumultuous Crimes Concerning Women in this paper compare the Apriori and predictive Apriori and finally conclude the Apriori perform better than the predictive Apriori algorithm [6]. Advanced Apriori

Algorithm in this paper the author compare Improved Apriori algorithm, feature based association rule mining and Optimized Distributed Association Rule Mining Algorithm finally the author conclude the Feature Based Association Rule Mining Algorithm works best for the large database[8]. Survey on several improved Apriori algorithms in this paper the author discuss the different Apriori algorithm and their advantage and disadvantage [12].

3. Association Rule

Association rule mining finds interesting association or correlation relationships among a large set of data items with massive amount of data. Association rule are helpful for analyzing customer behavior in Banking, Marketing analysis. Association rule has two parts ‘Antecedent’ and ‘Consequent’.

Antecedent is the item that found in database.

Consequent is the item that found in combined with the first. Association rule has an two criteria that is

3.1 Support

$$(A \Rightarrow B) = P(A \cup B).$$

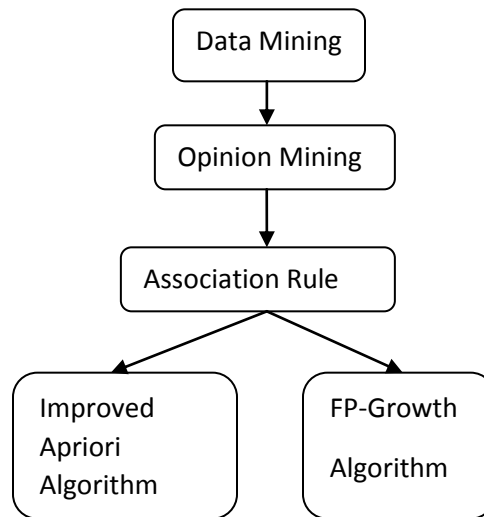
It is an indication of item how frequently it occurs in database. For a rule $A \Rightarrow B$, its support is the percentage of transaction in database that contain $A \cup B$ (means both A and B).

3.2 Confidence

$$(A \Rightarrow B) = P(B|A).$$

It indicates the no of times the statements found to be true. Confidence of the rule given above is the percentage of transaction in database containing A that also contain B. The rule $A \Rightarrow B$ has support s in the transaction set Data Base if s percentage of the transactions in database contain

AUB. Confidence denotes the strength of implication and support indicates the frequency of the patterns occurring in the rule.



3.3 Apriori Algorithm

Apriori is an algorithm for frequent item set mining and association rule learning over transactional database. Apriori uses a “bottom up” approach where frequent subset are extended one item at a time. Apriori uses breadth-first search and a Hash tree structure to count candidate item set efficiently.

Improved Apriori Algorithm

```

H1=find_frequent_1-itemsets(D);
for(i=2; Ai-1≠∅; i++)
{
  Ei=apriori_gen(Ai-1, min_sup);
  for each transaction t∈D
  {
    Ct=subset(Ei,t);
    for each candidate c∈Ct c.count++;
  }
  Ai={ c∈Ei |c.count≥min_sup }
}
  
```

```

Answer=UiAi ;
Procedure apriori_gen(Ai-1:frequent(i-1)-itemsets) for
each itemset H1 ∈ Ai-1
{
for each itemset I2 ∈ Ai-1
{
if(H1 [1]= I2 [1])∧ (H1 [2]= I2 [2]) ∧...∧(H1 [i-2]= I2
[i-2]) ∧(H1 [i-1]< I2 [i-1]) then
{
c=H1∞ I2;
if infrequent_subset(c, Ai-1) then
delete c;
else
add c to Ei ;
} } }
return Ei;
Procedure infrequent_subset(c: candidate i-itemset;
Ai-1:frequent(i-1)-itemsets)
for each(i-1)-subset s of c
{
if s ∈ Ai-1 then
return true;
}
return false;

```

Pruning Apriori Algorithm

```

Gen-itemsets with the given Lk-1:
Ak=∅
For all itemset I1∈Lk-1 do
For all itemset I2∈Lk-1 do
If I1[1]=I2[1]^I1[2]=I2[2]^.....^I1[k-1]<I2[k-1] then
C=I1[1],I1[2],.....,I1[k-1],I2[k-1]
Ak= Ak U {C}
The pruning set eliminates the extension of (k-1)
item sets which are infrequent from the
counting support.
Prune(Ak)
For all c∈ Ak
For all(k-1) subsets d of c do
If d∈Lk-1
Then Ak=Ak{c}

```

A. Limitations

1)Apriori, while historically significant suffers from a number of in efficient or trade off.

2) Candidate generation generates large numbers of subsets.

B. Advantage

- 1) It is very easy and simple algorithm.
- 2) Easy to implement.

C. Disadvantage

- 1) It does multiple scan over the database to generate candidate set.
- 2) The number of database passes are equal to the max length of frequent item set.
- 3) For candidate generation process it takes more memory, space and time.

3.4 FP growth Algorithm:

Let us define one of the most important problems with scalability of the Apriori algorithm. The complexity of the computation increases, aggressive. That is only one of the several factors that influence the development of several new algorithms for association - rule mining. FP Growth method is an efficient way of mining frequent itemsets in large databases. The algorithm mines frequent itemsets without the time - consuming candidate - generation process that is essential for Apriori.

FP-Growth Algorithm

```

Procedure FP-growth(Tree, α)
{
if Tree contains a single prefix path
then
{
let R be the single prefix-path part of Tree;
let S be the multipath part with the top branching node
replaced by a null root;
for each combination (denoted as β) of the nodes in
the path R do
generate pattern β U α with support = minimum

```

```
support of nodes in  $\beta$ ;  
let frequent pattern set(R) be the set of patterns so  
generated;  
}  
else  
let S be Tree;  
for each item pi in S do  
{  
generate pattern  $\beta = pi \cup \alpha$  with support = pi .support;  
construct  $\beta$ 's conditional pattern-base and then  $\beta$ 's  
conditional FP-tree Tree $\beta$  ;  
if Tree $\beta = \emptyset$   
then call FP-growth(Tree $\beta$ ,  $\beta$ );  
let frequent pattern set(S) be the set of patterns so  
generated;  
}  
return(frequent pattern set(R)  $\cup$  frequent pattern  
set(S)  $\cup$  (frequent pattern set(R)  
 $\times$  frequent pattern set(S)))  
}
```

A. Feature

- No candidate generation, no candidate test.
- Use compact data structure.
- Eliminate repeated database scan.
- Basic operation is counting and FP-tree building.

B. Advantage of FP growth

- 1) Reduce the overall size of all input data set.
- 2) It used an extended prefix tree structure for storing compressed and crucial information about frequent pattern.

4. Conclusion

After doing the survey the Association rule in Feature based association rule mining algorithm is perform better than the Improved Apriori algorithm, Predictive Apriori algorithm, Traditional Apriori algorithm, FB- Growth and the Tertius Association Rule algorithm. Apriori Algorithm analysis the data

at quickly in Banking sector, Education sector, Credit card and Debit card, Predicting heart disease and cancer.

5. Reference

1. Jayshree Jha, Leena Ragma. Education Data Mining using Improved APRIORI Algorithm: International Research Publication House (2013).
- MS Shweta, Dr Kanwal Garg. Mining Efficient association Rules through APRIORI Algorithm using attributes and comparative analysis of various association rule: International Journal of Advance Research in Computer Science and Software Engineering (2013).
- Yang Gong Xin, Zhengzhou, Henan. The research of improved APRIORI Mining Algorithm in bank customer segmentation: Atlantis Press (2013).
- Hamid Eslami Nosratabadi, Ahmad Nadali and Sanaz Pourdarab. Credit Assessment of bank customer by Fuzzy Expert system based on Rules Extracted from Association Rules: International Journal of Machine Learning and computing (2012).
- D. Gunaseelan, P. Uma. An Improved Frequent Pattern Algorithm for Mining Association rule: International Journal of Information and Communication Technology Research (2012).
- Divya Bansal, Lekha Bhambhu. Execution of APRIORI Algorithm of Data Mining Directed Towards Tumultuous Crimes Concerning Women: International Journal of Advanced Research in Computer Science and Software Engineering (2013).

7. Ms. Aarti Patil, Ms. Seem Kolkur, Ms. Deepali Patil. Advanced Apriori Algorithms: International Journal of Scientific & Engineering Research (2013).
8. Malathi. A, Dr. S. Santhosh Baboo. An Enhanced Algorithm to Predict a Future Crime using Data Mining: International Journal of Computer Applications (2011).
9. Supriya S. Borhade, Bipin B. Shinde. Privacy Preserving Data Mining Using Association Rule With Condensation Approach: International Journal of Computer Science and Information Technologies (2014).
10. Zhou Xin. Association Rules for Predicting Customer Lifetime Value in Retail Banking Context Based on RDB-MINER Algorithm: International Journal of Computer and Information Technology (2013).
11. Ms. Rina Raval, Indr Jeet Rajput, Vinitkumar Gupta. Survey on several improved Apriori algorithms: IOSR Journal of Computer Engineering (2013).
12. Jiao Yabing. Research of an Improved Apriori Algorithm in Data Mining Association Rules: International Journal of Computer and Communication Engineering (2013).
13. Aieman Quadir Siddique, Md. Saddam Hossain. Predicting Heart-disease from Medical Data by Applying Naïve Bayes and Apriori Algorithm: International Journal of Scientific & Engineering Research (2013).
14. Gitanjali J, C. Ranichandra, M. Pounambal. APRIORI algorithm based medical data mining for frequent disease identification: IPASJ International Journal of Information Technology (2014).
15. Shilpi Singla, Arun Malik. Survey on various improved Apriori Algorithms: International Journal of Advanced Research in Computer and Communication Engineering (2014).
16. Divya Jain, Sumanlata Gautam. Implementation of Apriori Algorithm in Health Care Sector: A Survey: International Journal of Computer Science and Communication Engineering (2013).