ANALYSIS OF SUPER MARKET USING ASSOCIATION RULE MINING

Daljeet Kaur
Student, Department of Computer Engineering
Punjabi University, Patiala

Jagroop Kaur
Department of Computer Engineering
Punjabi University, Patiala

Abstract: Supermarket analysis is a process to analyze the buyer's habits to discover the correlations between the different items in their shopping cart. The findings of these correlations can help the retailers to establish a profitable sales strategy by considering frequently purchased items together by customers. Association rule mining is one of the famous data mining techniques used to discover the correlations between one item to another. Association rule mining technique has number of algorithms, but this research focuses on the effectiveness of the combination of the two association rule mining algorithms that are apriori algorithm and eclat algorithm for supermarket analysis. The collaboration of both the algorithms revealed that both methods use the same concept with different criteria of processing the association rules, but the rules itself remains the same.

Keywords: Data Mining, Supermarket Analysis, Association Rule Mining, Apriori, Eclat.

I INTRODUCTION

Data mining is a process which is defined as to extract some useful information from huge volume of database to make crucial business decisions [1]. It is also concerned with the analysis of data and the use of software techniques for finding out hidden and unexpected patterns and correlations between sets of data [2]. So, the data mining mainly focuses on discovery of hidden and unexpected patterns. Data mining and artificial intelligence techniques are used to find out correlations between two or more variables [3].

Association rule mining is one of an important technique of data mining for knowledge discovery. The knowledge of the correlation between the items in the data transaction can use association rule mining [4]. It is a very famous technique for discovering correlations between variables in the huge databases [5]. It is a kind of supervised learning. Supervised learning found that the knowledge is based on the generated rules. The system of the correlation from each item can be knowledge to identify the future strategy obligation [6].

II Supermarket Analysis

Supermarket analysis is one of an application area of association rule mining technique [7]. Supermarket analysis is used to discover the system of correlation between one item to another. The possible percentage of the correlation of combined items gives the new knowledge. Therefore, it is a very helpful for determiner to take the decisions [8, 9].

For more details can be viewed at the following example: suppose that at supermarket, to answer the question "What are the items that may be frequently purchased together?"

To answer this question, supermarket analysis can be done on the purchase transaction data in the store. The result can be applied to sales planning such items frequently purchases together are placed close together [10]. There are two basic algorithms of association rule mining to find frequent itemset that are apriori algorithm and eclat algorithm.

III ASSOCIATION RULE MINING

An association rule mining is introduced in data mining to find out hidden patterns in large data sets and drawing inferences on how a subset of items impact the presence of another subset [11]. An association rule is one of the forms $A \rightarrow B$, where A is an "antecedent" (if part) and B is the "consequent" (then part). Here variables A and B are the item sets and the rule $(A \rightarrow B)$ means that customer who purchase an item set A are expected to purchase an item set B with the probability %c, where c is called confidence[12]. Interestingness measures of association rules are support and confidence.

Support(S): It is defined as the ratio of occurrence of two items and total number of transactions.

$$S\,(A, B) = \frac{\text{No. of Transactions containing both A and B}}{\text{Total No. of Transactions}}$$

Confidence(C): It is defined as ratio that how many instances satisfy the rule of an antecedent.

$$C\left(\frac{B}{A}\right) = \frac{\text{No. of Transactions containing both A and B}}{\text{No. of Transactions containing A}}$$

Let us consider the following example:

Customer ID	Item Purchased	Item Purchased
1	Burger	Coke
2	Bread	Butter
3	Burger	Mineral Water
4	Snacks	Coke

Now, if A is "Purchased Burger" and B is "Purchased Coke", then support and confidence are calculated as:

$$\text{Support} = S \text{ (A and B)} = 1/4$$
$$\text{Confidence} = C \text{ (B/A)} = 1/2$$

Item sets that satisfy minimum support and minimum confidence are known as strong association rules [13].

IV   APRIORI ALGORITHM

This algorithm has been often utilized for mining of frequent item sets and to find associations. The real distinction in Apriori is the less hopeful item sets it produces for testing in every database pass. The quest for association guidelines is guided by two parameters: support and confidence. Apriori gives back an association guideline on the off chance that its support and confidence qualities are above client characterized limit values. It is a breath first search algorithm. The yield is requested by confidence. On the off chance that few principles have the same certainty then they are requested by support [14]. In this manner Apriori supports more certain tenets and portrays these guidelines as additionally intriguing. The working of Apriori calculation is decently relies on the Apriori property which expresses that" All nonempty subsets of a regular item sets must be frequent".Apriori essentially traverses the tree in BFS (breadth first search), it implies it first checks for the item set of size 1 and after that further for the item arrangement of size 2 and so on. Apriori surveys the support of item sets may be by checking each of item set which the exchanges contains inside, or by navigating for an exchange each subset of the most as of late handled size and expanding the related item set counters [15].

V   ECLAT ALGORITHM

Eclat creates less number of succession tables which sets aside less time for frequent accessed patterns to examples when contrasted with apriori. In apriori if huge data is their then it takes colossal time to create the successive frequent accessed patterns. Eclat execution speaks to the arrangement of exchanges as a bit network and meets columns give the backing of thing sets[16]. It takes after a profundity first traversal of a prefix tree. Eclat, explores the prefix tree top to bottom first demand, being backwards to apriori. It just implies that, it broadens a thing set prefix until and unless it reaches to the limit in the middle of the rare and continuous

item sets and afterward further backtracks to process the nearing prefix. Eclat figures the support of each and every item set by making the rundown of the considerable number of identifiers of exchanges that contain the itemsets. It utilizes the methodology of crossing two arrangements of exchange identifiers for two distinctive item sets just by a solitary item or together frame the thing set as of late prepared [17].

VI   EXPERIMENTAL SETUP & RESULTS

For implementing both apriori and eclat algorithms we've used the R Environment with various user defined and system defined libraries. We have used stocks dataset consisting of 200 records of various items for implementing both the algorithms.

User Libraries:
- arules
- arulesViz

System Libraries:
- matrix
- grid

Apriori algorithm is best for small datasets where the minimum number of candidate have been generated and Eclat algorithm is good for large databases where data is huge in numbers. Therefore, this will be the best idea to use both the algorithms collectively. This will make an easier and systematic approach for extracting knowledge.

A threshold or limit can be set for the transactions. If transactions exceed the limit then Eclat algorithm should be used, but if transaction does not exceed the limit then ultimately Apriori algorithm should use. At the end set of association rules has been generated.

A decision support system based on apriori and eclat algorithm will boost the efficiency of the decision making process based on the associations obtained.

- ➢ In this research work, we set the threshold on number of columns of stocks dataset. Here number of columns of the given dataset satisfy the given threshold, therefore apriori algorithm executes followed by the given below parameter specifications.

```
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target    ext
       0.8    0.1    1 none FALSE            TRUE         5    0.01      2      10 rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 2

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[444 item(s), 200 transaction(s)] done [0.00s].
sorting and recoding items ... [239 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 done [0.00s].
writing ... [61 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
      lhs                                     rhs                      support confidence
[1]  {I2=shopping bags}               => {I5=bottled water}       0.010    1
[2]  {I2=pork}                        => {I1=sausage}             0.010    1
[3]  {I2=cereals}                     => {I1=sugar}               0.010    1
[4]  {I6=cat food}                    => {I5=rolls/buns}          0.010    1
[5]  {I2=canned beer}                 => {I6=whole milk}          0.010    1
[6]  {I3=ham}                         => {I5=other vegetables}    0.010    1
[7]  {I3=ham}                         => {I6=whole milk}          0.010    1
[8]  {I4=grapes}                      => {I6=whole milk}          0.010    1
[9]  {I6=brown bread}                 => {I5=domestic eggs}       0.010    1
[10] {I2=frankfurter}                 => {I5=other vegetables}    0.010    1
[11] {I2=frankfurter}                 => {I6=whole milk}          0.010    1
[12] {I4=hamburger meat}              => {I1=tropical fruit}      0.010    1
[13] {I2=sliced cheese}              => {I3=rolls/buns}          0.010    1
[14] {I2=sliced cheese}              => {I1=sausage}             0.010    1
[15] {I5=dessert}                     => {I3=other vegetables}    0.010    1
[16] {I2=hamburger meat}              => {I3=citrus fruit}        0.010    1
[17] {I3=chocolate}                   => {I2=soda}                0.010    1
```

Fig.1 Resultant Apriori Association Rules

Here apriori algorithm executes and create 61 set of association rules. As first rule described as {I2=shopping bags} => {I5=bottled water} with high support (0.010), confidence (1).

➢ If the number of columns of the given dataset exceeds the given threshold, then eclat algorithm executes followed by the given below parameter specifications.

```
Eclat

parameter specification:
 tidLists support minlen maxlen              target   ext
    FALSE    0.01     2     10 frequent itemsets FALSE

algorithmic control:
 sparse sort verbose
      7   -2    TRUE

Absolute minimum support count: 2

create itemset ...
set transactions ...[444 item(s), 200 transaction(s)] done [0.00s].
sorting and recoding items ... [239 item(s)] done [0.00s].
creating sparse bit matrix ... [239 row(s), 200 column(s)] done [0.00s].
writing  ... [211 set(s)] done [0.00s].
Creating S4 object  ... done [0.00s].
> inspect(rules)
      items                                              support
[1]  {I2=canned beer,I6=whole milk}                      0.010
[2]  {I3=ham,I5=other vegetables,I6=whole milk}          0.010
[3]  {I3=ham,I6=whole milk}                              0.010
[4]  {I3=ham,I5=other vegetables}                        0.010
[5]  {I4=grapes,I6=whole milk}                           0.010
[6]  {I2=shopping bags,I5=bottled water}                 0.010
[7]  {I5=domestic eggs,I6=brown bread}                   0.010
[8]  {I1=sugar,I2=cereals}                               0.010
[9]  {I5=rolls/buns,I6=cat food}                         0.010
[10] {I2=frankfurter,I5=other vegetables,I6=whole milk}  0.010
[11] {I2=frankfurter,I6=whole milk}                      0.010
[12] {I2=frankfurter,I5=other vegetables}                0.010
[13] {I4=margarine,I5=sugar}                             0.010
[14] {I1=tropical fruit,I4=hamburger meat}               0.010
[15] {I2=hamburger meat,I3=citrus fruit}                 0.010
[16] {I2=soda,I3=chocolate}                              0.010
[17] {I2=photo/film,I6=soda}                             0.010
```

Fig.2 Resultant Eclat Association Rules

Here eclat algorithm executes and create 211 set of association rules. As first rule described as {I2=canned bear, I6=whole milk} with high support (0.010).

VII   CONCLUSION

In this paper, the findings of association rules from both algorithms concludes that both the algorithms collectively work better by giving best results as apriori algorithm serves better for small datasets by giving accurate results whereas eclat algorithm serves better for large datasets by taking less execution time and by creating less number of tables for dataset.

VIII. REFERENCES

[1] Smita,Priti Sharma. (2014). Use of Data Mining in Various Field:A Survey Paper. IOSR-Journal of Computer Engineering(IOSR-JCE).

[2] Aarti Sharma,Rahul Sharma,Vivek Kr.Sharma,Vishal Shrivastava. (2014). Application of Data Mining:A Survey Paper. (IJCSIT) International Journal of Computer Science and Information Technologies.

[3] Jiawei Han and Micheline Kamber. (2006). Data Mining: Concepts and Techniques. Diane Cerra.

[4] Jagmeet Kaur, Neena Madan. (2015). Association Rule Mining: A Survey. International Journal of Hybrid Information Technology.

[5] T. Karthikeyan and N. Ravikumar. (2014). A Survey on Association Rule Mining. Intenational Journal of Advanced Research in Computer and Communication Engineering.

[6] Nengsih, W. (2015). A Comparative Study on Market Basket Analysis and Apriori Association Technique. International Conference on Information and Communication Technology. IEEE.

[7] Phani Prasad J, Murlidher Mourya. (2013). A Study on Market Basket Analysis Using a Data Mining Algorithm. International Journal of Emerging Technology and Advanced Engineering.

[8] Yunkang Yue,Tiebao Zhang,Xingbing Xu. (2010). Analysis on the Effect of Data-mining to Supermarket. International Conference on Computer Science & Education Hefei, China. IEEE.

[9] Zhang Aiguo,Jiang Lanling,Song Ping. (2011). Application of Data Mining in Supermarket. International Conference on Consumer Electronics, Communications and Networks (CECNet). IEEE.

[10] Djoni Haryadi Setiabudi, Gregorius Satia Budhi, I Wayan Jatu Purnama, Agustinus Noertjahyana. (2011). Data Mining Market Basket Analysis' Using Hybrid-Dimension Association Rules, Case Study in Minimarket X. International Conference on Uncertainity Reasoning and Knowledge Engineering. IEEE.

[11] Jeetesh Kumar Jain,Nirupama Tiwari,Manoj Ramaiya. (2013). A Survey: On Association Rule Mining. International Journal of Engineering Research and Applications (IJERA).

[12] Surbhi K. Solanki,Jalpa T. Patel. (2015). A Survey on Association Rule Mining. International Conference on Advanced Computing & Communication Technologies. IEEE.

[13] Qiankun Zhao,Sourav S. Bhowmick. (2003). Association Rule Mining: A Survey.

[14] Muhammad Asif, Jamil Ahmed. (2015). Analysis of effectiveness of Apriori and Frquent Pattern Tree Algorithm in Software Engineering Data Mining. Internatonal Conference on Intelligent Systems, Modelling and Simulations. IEEE.

[15] Warnia Nengsih. (2015). A Comparative Study on Market Basket Analysis and Apriori Association Technique. International Conference on Information and Communication Technology (ICoICT). IEEE.

[16] Dunham, M.H. (2003). Data Mining: Introductory and Advanced Topics. Pearson Education

[17] Bina Kotiyal, Ankit Kumar, Bhaskar Pant, R.H.Goudar, Shivali Chauhan and Sonam Junee. (2013). User Behaviour Analysis in Web Log through Comparative study of Eclat and Apriori. IEEE International Conference on Intelligent System and Control(ISCO 2013).