International Journal of Engineering & Technology, 7 (2.32) (2018) 206-208

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



A Survey on Analysis of Online Consumer Behaviour Using Association Rules

Dr.M.R. Narasingha Rao[1], K.V.L Sita Ratnam[2], M.D.S. Prasanth[3], P. Lakshmi Bhavani[4]

[1,2,3,4]*Department of Computer Science, Koneru Lakshmaiah Educational Foundation, Guntur, India.*
**Corresponding author E-mail: narasimharao@kluniversity.in*

Abstract

In vogue the competitive world is searching for customer oriented approach. In order to accomplish that customer relation management (CRM) provides the services which meet the customer requirement. The requirements of a customer are collected in the form of data which is to be analyzed for obtaining customer behavior. Therefore inquiry is made to determine the characteristic features of a customer by using the data which is collected from different sources of all the online shopping sites. Data mining is a process of extracting useful information from the historical data. In this analysis the data of a online shopping website which is take from mechanical sector is tested by using some data mining algorithms like Apriori, FP-growth and Eclat algorithm of association rule mining. As a result of the study it is aimed to obtain frequent items purchased by the consumer by using the association results.

Keywords:*Datamining, association, customer relationship management (CRM), Apriori, FP growth, Eclat..*

1. Introduction

In previous days shopping in online is done by the limited people , But in the running society it is difficult for a customer to buy the products in traditional shopping .Traditional shopping requires more time for purchasing the required items of a customer . To reduce such complications customers are going for online shopping for purchasing different goods like electronic gadgets, vegetables, furniture and more. Online shopping reduces the time to buy the required items for a customer. The same items are present in different online shopping sites with different cost. To know the best price of an item customers visit different sites for will buy the item which has high quality and less expensive. To increase the efficiency of online shopping the online heirs will take an interest for finding the promising customers based on their transaction history. Extracting useful information from the history is done by using data mining techniques.

Online shopping provides the customer to purchase the goods in a smooth way. Online shopping provides enormous combination of data that provides knowledge about various products present in different sites. In online shopping system we have stored the product ids purchased by the customer with their ids and the month and year and quantity purchased by the customer in the form a database. Analyzing them theoretically leads to more time complexity and memory wastage in order to run over this problem we have used data mining techniques such as Apriori ,FP growth and Eclat algorithms .These functions finish transactions in shopping database automatically in minor time. These tasks help in analyzing the customer behaviors which leads to improve the quality of customer service and provide good transportation facilities.

Data mining is the process of collection, analysis and modeling of large databases in order to determine models and patterns that are not known in previous .Most of the organizations in various sectors are using to take advantages over their challengers. The data mining process must be stable. Few applications of data mining are manufacturing engineering, market basket analysis, fraud detection, CRM, Lie detection, research analysis, corporate surveillance, financial banking, and customer segmentation.

Customer Relationship Management (CRM) is used for collecting and preserving customers that improves customer loyalty, gaining customer vision and implementing the customer related strategies. To maintain the relationship with the customer a business requires collecting the suitable information about its customer and arranging that information for decent analysis and actions. It needs to maintain the information up to date and make it available to all the employees.

.

Association rule mining is first discovered by R. Agarwal in 1993[2]. .Association rule mining is a rule based machine learning technique which discovers interesting patterns between the items in a tremendous data bases. Association rules mining is a data mining technique used for finding the frequent item sets.

If an association rule indicates that {Munch, Kit-Kat} ➔ {5 Star} gives a meaning that if a customer buys both Munch and Kit- Kat then he buys 5 Star too. Such intelligence can be used by marketing systems , medicine and also in various sectors.

There are different algorithms in association rule mining like Apriori , FP- Growth, Eclat. In this paper the online customer behavior is predicted based on association rules. By using these rules the frequent items purchased by the customer is known.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

2. Literature Survey

Jeetesh Kumar Jain [1] has studied about two different methods in association rule mining .Those are Apriori and FP-Growth methods .The generation of frequent item sets is done by using the two methods. The advantages and disadvantages of those methods is discovered. According to the author in Apriori algorithm at first the frequent item sets are generated and then pruning is done on those item sets. The main advantages of Apriori algorithm are that it is applied for large data sets, easily parallelized and implementation is easy. The disadvantages of apriori algorithm are to obtain frequent item set the data base must be scanned for many times and candidate generation is more expensive in both space and time. In FP-Growth algorithm the FP-tree is generated. The main advantage of FP-growth algorithm is there is no candidate generation required and less scans of data base for obtaining frequent item sets . FP-Growth algorithm is much faster than the Apriori algorithm. The FP- tree generated may not fit in the memory and construction of FP- Tree is more expensive are the disadvantages of FP-Growth algorithm. In these paper applications of association rule mining like market basket analysis is discussed.

Rana Alaa El-Deen Ahmed [3] has studied about eleven classification techniques which are comparatively tested to know the perfect classifier which describes the behavior of consumer towards online shopping. The algorithms studied by the author are Bayes Net , Naïve Bayes, K Star , Classification via clustering , Filtered classifier ,END, JRIP , Ridor, Decision table, J48 , Simple cart. The analysis of the algorithms is done based on the TP rate, FP rate , recall , precision, ROC curves-measures ,accuracy and execution time to construct the model . The results have shown that decision table will give more accuracy than all other algorithms and classification via clustering has lowest accuracy. These results will help all the online shopping sites for predicting the consumer behavior.

Moushumi Sharma [4] has studied about association rule mining techniques. The author has divided the association algorithms into two ways. One with candidate generation i.e, Apriori algorithm and another without candidate generation i.e, FP-Growth algorithm. In that paper a brief description of the most recent work done by association rule mining using Apriori and FP-Growth algorithms is presented. The results has shown that increase in Apriori algorithm is more suitable than FP-Growth algorithm.

The rest of the paper illustrates various association rule mining methods which are proposed for generating frequent item sets and concludes the paper with references.

3. Apriori Algorithm:

The name itself indicates that it uses prior knowledge of frequent item sets . Apriori algorithm uses breadth first search technique for obtaining frequent items . Apriori engages in an iterative approach noted as level wise search

technique ,where $-item sets are used to explore ($+1) – item sets. At first the first set of frequent item set is recognized by scanning the data base to generate the count for each item in the data set and collecting those items which satisfy the minimum support value into a set – I. Then Set – II frequent items are found by using Set – I items . This process continues till there are no frequent items found in the new item set.

Consider a sample data set having minimum support of 50 percent:

Table I: Sample Data Set

Customer ID	Product ID
C1	i , iii , iv
C2	ii, iv
C3	i , ii , iii, v
C4	ii , iv, vi

Table II: The items which satisfy the minimum support

Item set($)	Supp($)
(i, iii)	50%
(ii, iv)	50%

From the above table we can find the confidence of each item set. After calculating we found that the confidence of ‘i’ is 100 percent with ‘iii’ and vice verse .But the confidence of ‘ii’ is 60 percent with ‘iv’ , with this we can conclude that the customers who buy item ‘I’ will compulsorily buy ‘iii’ and the customers who will buy ‘ii’ there is a chance of buying ‘iv’ but not for sure.

4. Fp-Growth Algorithm:

FP-Growth algorithm is used for generating frequent item sets without using the candidate item set generation. FP-Growth algorithm uses Bottom-Up approach .There are two steps to find frequent items from a data set. The first step consists of constructing FP-tree using two passes over the data set. The extraction of frequent item sets directly from the tree is done in the second step. The scanning of data base and finding the support of each item, after removing the item sets which does not satisfy the minimum support value sort the obtained item sets in decreasing order based on their support values is done in the first step. The FP –Tree is constructed using the sorted order. The frequent item sets are extracted using the tree.

Table III: Consider a sample data set with minimum support of 30 percent:

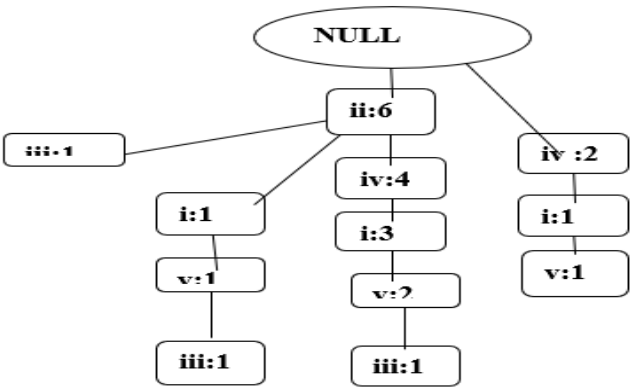
Customer ID	Product ID
C1	v, i, iv, ii
C2	vi, i, iii, v, ii
C3	iii, i, ii, v
C4	ii, i, iv
C5	iv
C6	iv, ii
C7	i, iv, v
C8	ii, iii

The resultant table after finding the minimum support value for each item and removing the items which does not satisfy the minimum support value and sorting the obtained items is as follows :

Table IV: Items satisfying minimum support

Customer ID	Product ID	Ordered Items
C1	v, i, iv, ii	ii, iv, i, v
C2	vi, i, iii, v, ii	ii, iv, i, v, iii
C3	iii, i, ii, v	ii, i, v, iii
C4	ii, i, iv	ii, iv, i
C5	iv	iv
C6	iv, ii	ii, iv
C7	i, iv, v	iv, i, v
C8	ii, iii	ii, iii

FP –Tree construction:



All Fp-Trees will have ‘null’ node as a root node.The root node is drawn first followed by the items in the rows one by one respectively based on table 4

5. Eclat Algorithm:

Eclat algorithm is used for generating frequent item sets based on minimum support and high confidence. Eclat algorithm uses depth first search technique for generating frequent item sets. In éclat the data is represented in vertical format. The frequent item sets generated based on the minimum support value as in apriori algorithm. The process continues until there are no frequent items generated further. Depth first tree reduces the memory complexity.

Table V: Sample Dataset

Transaction ID	Product Id
T1	i, iii, iv
T2	ii, iii, v
T3	i, ii, iii, v
T4	ii, v

Table VI: Eclat algorithm constructs the vertical data set

i	ii	iii	iv	v
T1	T2	T1	T1	T2
T3	T3	T2		T3
	T4	T3		T4

After generating the vertical data set the item which does not satisfy the minimum support value are removed and forms the frequent item set as follows:

Table VII: Frequent Item Set

Transaction ID	Product ID
i, iii	T1,T3

By observing the above table we can easily know that the customer who buys item ‘i’ will buy item ‘iii’ and vice versa.

6. Conclusion

In this paper, three association rule mining techniques are examined include Apriori , FP-Growth and Eclat algorithm. Apriori algorithm requires more data base scans for obtaining frequent item sets where as FP- Growth algorithm requires less data base scans . In Apriori algorithm candidate generation is done whereas in FP-Growth algorithm there is no candidate generation. Eclat algorithm is used for generating frequent item sets for very small data sets. Apriori algorithm is applicable for large data sets. The results have shown that FP-Growth algorithm is much faster than Apriori algorithm but in FP-Growth algorithm the construction of FP-Tree is more expensive and may not fit in the memory. Therefore, Apriori algorithm is more suitable for obtaining customer behavior towards online shopping

References

[1] Jeetesh Kumar Jain, Nirupama Tiwari, Manoj ya ,International Journal of Engineering Research and Applications,”A Survey : On Assoiation Rule Mining”, Vol. 3, Issue 1, January -February 2013, pp.2065-2069.

[2] Rana Alaa El-Deen Ahmed, M.Elemam.Shehab, Shereen Morsy, Nermeen Mekawie,” Performance study of classification algorithms for consumer online shopping attitudes and behaviour using data mining”, Fifth International Conference on Communication Systems and Network Technologies,2015.

[3] Moushumi Sharma, Ajit Das, Nibedita Roy, “A Complete Survey on Association Rule Mining and Its Improvement”, International Journal of Innovative Research in Computer and Communication Engineering , Vol. 4, Issue 5, May 2016.