Improving Recommendation in E-Commerce Using Apriori Algorithm

Ranjan S G[1], Sandesh A Hegde[2], Sujay N[3], Swaroop S Rao[4], Padmini M S[5]

[1234]*BE, Department of CSE, NIE Mysuru, Karnataka, India*
[5]*Assistant Professor, Department of CSE, NIE Mysuru, Karnataka, India*

---***---

Abstract-*The project aims at improving recommendation in E-Commerce websites. For that there will be User Interface development which acts as a kind of input module to the project. The existing E-Commerce websites aims at providing recommendation based on the transaction history of the user and sometimes they recommend latest products and also highly rated products. In our project we aim at recommending products to the user based on the transaction history of other users who has same characteristics as this user. Thus we aim at collecting details like Age, Gender, Education, Marital status, Salary etc. So there requires the data mining techniques like clustering. Apriori Algorithm is the main algorithm used in our project. Apriori Algorithm is the general algorithm which can be used by developers according to their need and implement it in their projects.*

Key Words: E-Commerce, Apriori, Association rule mining, Market Segmentation, Data Mining, Recommendation.

1.INTRODUCTION

Association rule mining is a major factor and a method of data mining . Apriori algorithm is a tool in association rule mining used to implement it. Data mining helps using the huge amount of data in an efficient way, extract the required data and predict the future based on customer behavior. Data mining is one of the steps in the knowledge discovery process. It is the process of extracting interesting patterns from large amounts of data and in the proposed system it's used to provide recommendations to the customers based on their profiles. Data mining uses the already available data i.e., datasets to solve a problem or to find the patterns. The two types of data mining tasks: descriptive and predictive. Predictive data mining or predictive analysis uses modelling, statistical techniques to study the current and historical data to predict the future. Descriptive data mining or descriptive analysis uses data aggregation to provide an insight into the past. Data mining is widely used by different companies and organizations to extract information, predict the future like predicting stock market,sales,profits. It helps the companies to gain information about customer satisfaction. Using data mining, organizations can increase their profit by providing better services customer and minimizing the risk. In the proposed system, Association rule mining is used to extract the customer transaction data from a e-commerce organization. Apriori algorithm is used to analyze the customer behavior based on his/her profile and provide recommendation to the customer by extracting and finding matching patterns.

1.1 Data Mining Tasks

The initial form of data is the data present in a database. The first step of data cleaning process is the removal of invalid data, the NULLs, the transactions aborted mid-way (3). The next step of refinement is removal of the data which is not necessary for the analysis, say if we are analyzing the products bought in India, there is no point in considering the data of the products bought in Germany. The next step is clustering, grouping of data which are similar. There are a bunch of parameters set for any client or a transaction, the transactions for which most or all of these properties match are grouped as similar. The end result of such a clustering is that when a new transaction or client registers with the same properties as set before, the output of the analysis can be utilized on that client. The final step in the process is transforming the data which we have refined till now into a format the pattern searching algorithm expects.

1.2 Transaction Set for Apriori Algorithm

1. CUSTOMER_DETAILS

Email_ID	Name	Password	Gender	DateOfBirth	MartialStatus	Education
de.1728@gmail.com	Ganesh	ganesh	Male	28/02/1996	Single	Diploma
kjajith@gmail.com	ajith	ajith	Male	1994	Single	Degree
manjasarpa@gm...	Manjunath	manjunath	Male	12/09/1995	Married	PHD
mysujay@gmail....	Sujay N	sujay	Male	13/03/1995	Single	PU
ranjansg3@gmai...	Ranjan	ranjan	Male	1/1/1996	Single	Degree
ranjansg5@gmai...	ranjan	abc	Male	1901	Single	Degree
sangeeta.epatr...	Sangeeta	sangeeta	Female	25/02/1995	Single	Master degre
shegde95@gmai...	Sandesh	sandesh	Male	31/3/1995	Single	Degree
spatikahim@gma...	Spatika	spatika	Female	23/04/1995	Married	10TH
swaroopsrinivas...	Swaroop	swaroop	Male	12/04/1996	Single	Degree

ils | LENOVO-PC\SQLEX...mer_Transactions | SQLQuery1.sql - L...VO-PC\Ranjan (53)) |

Occupation	Income	Address	ContactNo	City
Business	10k - 20k	Sirsi	5678901234	Mandya
Student	30k - 40k	Mysuru	8886783434	NULL
Bank	20k - 30k	Perla	7890123456	Coorg
Student	5k - 10k	Nanjangudu	4567890123	Mysore
IT	40k - 50k	Mysuru	9876543210	Mysore
Student	5k - 10k	mysore	9448602258	NULL
IT	30k - 40k	Sagara	6789012345	Shivmoga
Bank	10k - 20k	Davangere	2345678910	Bengaluru
Student	5k - 10k	Mangalore	8901234567	Coorg
IT	40k - 50k	Mysuru	1234567890	Mysore
NULL	NULL	NULL	NULL	NULL

2. CUSTOMER_TRANSACTIONS

LENOVO-PC\SQLEX...mer_Transactions | SQLQuery1.sql - L...VO-PC\Ranjan (53)) | LENOVO-PC\SQLEX...nsaction_Details

Transaction_ID	Email_ID	Transaction_Date	Dispatched_Date	Status
1	ranjansg3@gmai...	2017-03-27 14:...	NULL	Pending
2	shegde95@gmai...	2017-03-27 14:...	NULL	Pending
3	mysujay@gmail...	2017-03-27 14:...	NULL	Pending
4	swaroopsrinivas...	2017-03-27 15:...	NULL	Pending
5	sangeeta.epatr...	2017-03-27 15:...	NULL	Pending
6	spatikahim@gma...	2017-03-27 15:...	NULL	Pending
7	ranjansg5@gmai...	2017-03-28 13:...	NULL	Pending

3. ITEM_DETAILS

SQLQuery1.sql - L...VO-PC\Ranjan (53)) | LENOVO-PC\SQLEX...nsaction_Details | LENOVO-PC\SQLEX...bo.Item_Details

Item_ID	SubCategory_ID	Item_Name	Item_Cost	Item_Details	Item_Image	Quantity	Attachment
1	1	JohnPlayers	1000	Blue shirts	/Item_Photos/Jo...	21	NULL
2	2	Arrow	850	casual shirt	/Item_Photos/Ar...	28	NULL
3	3	Nike	850	Sportswear	/Item_Photos/Ni...	14	NULL
4	7	Peter England	300	Classic Ties	/Item_Photos/P...	42	NULL
5	4	Bata	1500	Bata Shoes	/Item_Photos/B...	29	NULL
6	5	Django	1000	Casual Shoes	/Item_Photos/Dj...	13	NULL
7	6	Reebok	2000	Sports Shoes	/Item_Photos/R...	21	NULL

4. TRANSACTION_DETAILS

SQLQuery1.sql - L...VO-PC\Ranjan (53)) | LENOVO-PC\SQLEX...nsaction_Details

Details_ID	Transaction_ID	Item_ID	Quantity
1	1	4	1
2	1	1	1
3	1	7	1
4	2	4	1
5	2	1	1
6	2	6	1
7	3	4	1
8	3	2	1
9	3	1	1
10	3	6	1
11	4	2	1
12	4	6	1
13	5	3	1
14	5	7	1
15	6	1	1
16	7	7	1
* NULL	NULL	NULL	NULL

2. APRIORI ALGORITHM

The steps for the algorithm:

- Scan the opinion data set and determine the support(s) of each item.
- Generate L1 (Frequent one item set).Use Lk-1, join Lk-1 to generate the set of candidate k - item set.
- Scan the candidate k item set and generate the support of each candidate k – item set.
- Add to frequent item set, until C=Null Set.
- For each item in the frequent item set generate all non empty subsets.
- For each non empty subset determine the confidence. If confidence is greater than or equal to this specified confidence .Then add to Strong Association Rule.

A->CLASSIC TIES

B->CASUAL SHIRT

C->BLUE SHIRT

D->SPORTS SHOES

E->CASUAL SHOES

INPUT DATASET:

TID	Items
1	A,C,D
2	A,C,E
3	A,B,C,E
4	B,E

Minimum Support = 50%

Minimum Confidence = 80%

Item set: ⮚ A, B, C, D, and E

STRONG ASSOCIATION RULE:

This is the result obtained.

1. {B}->{E}

2. {CE}-> {A}

3. {AE}->{C}

4. {A}-> {C}

5. {C}->{A}

3. CONCLUSIONS

The proposed system uses Apriori algorithm for finding item sets frequently bought together considering customer profile factors such as age, gender, income, education, marital status rather than providing product based recommendation. The customer spends more time due to better recommendation. This increases the profit for the organization and also increases the number of customers.

REFERENCES

[1] "SEGMENTING CUSTOMERS WITH DATA MINING TECHNIQUES" ,Sabri Serkan Güllüoğlu Computer Engineering Department Istanbul Arel University İstanbul, Turkey Türkoba Mah. Erguvan Sok. 26-K Tepekent/Büyükçekmece .

[2] "CUSTOMER DATA CLUSTERING USING DATA MINING TECHNIQUE",Dr. Sankar Rajagopal,Enterprise DW/BI Consultant Tata Consultancy Services, Newark, DE, USA e-mail:rajagopal.sankar@tcs.com

[3] "Using Data Mining Techniques in Customer Segmentation" ,Hasan Ziafat *, Majid Shakeri ** *(Department of Computer Science, Islamic Azad University Natanz branch, Natanz, Iran) ** (Department of Electronic science, Islamic Azad University Natanz branch, Natanz, Iran).

[4] https://www.quora.com/How-do-I-pick-appropriate-support-confidence-value-when-doing-basket-analysis-with-Apriori-algorithm

[5] https://www.quora.com/What-is-the-pseudocode-for-Apriori-algorithm

[6] "Extension of Local Association Rules Mining Algorithm Based on Apriori Algorithm", Zhang Chun-sheng College of Computer Science and Technology Inner Mongolia University For Nationalities, tongliao028043,CHINA zhangcs_817@163.com, Liyan College of Computer Science and Technology Inner Mongolia University ForNationalities,tongliao028043,CHINA liyanyx@163.com

[7] "Apriori Algorithm for Association Rule Mining in High Dimensional Data",Sandhya Harikumar∗, Divya Usha Dilipkumar∗, ∗Department of Computer Science and Engineering, Amrita School of Engineering, Amritapuri, Amrita Vishwa Vidyapeetham, Amrita University, India sandhyaharikumar@am.amrita.edu,divyadilip72@gmail.com

[8] "A Semi-Apriori Algorithm for Discovering the Frequent Itemsets",Sallam Osman Fageeri Department of Computer and Information Sciences Universiti Teknologi PETRONAS Bandar Seri Iskandar, 31750 Tronoh,Perak,Malaysia,Sall_G01825@utp.edu.my,Sallam_fageeri@hotmail.com, Rohiza Ahmad, Baharum B. Baharudin Department of Computer and Information Sciences Universiti Teknologi PETRONAS Bandar Seri Iskandar, 31750 Tronoh,Perak,Malaysia{ Rohiza_ahmad, Baharbh}@petronas.com.my