Academic Science

International Journal of Innovations & Advancement in Computer Science
IJIACS
ISSN 2347 – 8616
Volume 4, Special Issue
May 2015

E-commerce Recommendation System usingAssociation Rule Mining and Clustering

Prof. Vishal Parikh
Department of Computer Science and Engineering
Institute of TechnologyNirma University
Ahmedabad India-382481

Parth Shah
Department of Computer Science and Engineering
Institute of Technology Nirma University
Ahmedabad India-382481

Abstract
This paper analyses content based recommendation for e-commerce site. Recommendation system use to generate recommendation of the product that customer may want to buy. This system increase the sale of vendor and easy to find product from available product. Association rule mining and clustering technique use to make real time recommendation system. From the user's transaction data-set we can generate rules for customer buying tendency. Based on customer purchased product and customer profile, we can generate recommendation using association rule mining technique. Association rule mining is very time consuming process for large data-set. So, it is not feasible for real time recommendation system. To overcome this problem clustering technique is used. Using hierarchical clustering we can make partition of whole large data-set in to tree of clusters. It decrease the time for real time recommendation system.
Keywords—e-commerce recommendation system; association rule mining;cluster analysis;

1 Introduction

With the increase the usage of Internet worldwide, The growth of e-commerce business is very high, and buying products from the ecommerce website has become a new trend in modern culture. While vendors and purchaser become happy and feel convenient brought product from e-commerce website. There is large number of products available in market and it increasing day by day. It is very difficult to find the required product from available large set of product. If customers face difficulties then vendor have risk to lose the customer. In this case, the E-commerce recommendation system becomes very useful for customers as well as vendors. Earlier recommendation system gives same recommendation to every user based on available information. With the development of e-commerce business we want to meet the each and every customer's needs individually [12].Recommendation systems enhance e-commerce sales in three ways as below.

A. Converting visitors into purchaser

It is general trend that customers are just visit the site and get information about product and forgot it. Recommendation system stores the customers visited product list and give recommendation whenever customer visit the site. It help to customers by finding the product which they wish to purchase.

B. Increasing Cross-sell

Recommender systems use to increase cross-sell of product by recommending products to the customer. After purchasing some product, recommendation system give list of products that may be require for customer. This increase the sell of vendors.

C. Building Loyalty

Recommendation system use to build relationship between vendor and purchasers. Recommendation system use to learn the customer behavior while purchasing the products. Customer's feedback on recommended product is use to for built the loyalty.

E-commerce recommendation system gives recommendation to customer for purchasing product based on the customer's profile, history and other customer's previous transactions. This system access the customer's data such as age, gender, location, social networking site friend list etc. Based on this data recommendation system suggest list of product that customer may be required. Association rule mining is use to find the correlation between customer and product. In this paper multidimensional association rule mining technique is use for find more accurate recommendation. It is also important that each customer should get recommendation in less time [9].
We are using large set of customer transaction data for find the association rules. It requires much time to find the recommendation using association rule mining. To reduce the time complexity clustering techniques is used. In this paper hierarchical clustering technique is used. We create the tree structured cluster that decrease time complexity.

2 ASSOCIATION RULE MINING ANALYSIS

Association rule mining is a well researched methodology to discovering interesting relations between attribute in available large dataset. The intention of association rule mining is to discover the strong rule using different measurement. Based on the concept of strong rules we can identify recommended product.
Let I = {I1,I2,I3,....In} is the set of the available unique items in the dataset. C={C1,C2,C3,....Cn} is a set of customers that are signed in the site. T={I1,I3...Ik} is the transaction of each customer purchase. Each transaction is uniquely identify by the transaction_id(TID). Each transaction_id(TID) is correlated with customer_id(CID).Each transaction has a subset of item set. ($T \subset I$) . Let A and B is a set of item-set in Transaction T. Here $A \subset T$ and $B \subset T$ and $A \cap B = \Phi$. A association rule $A \rightarrow B$ implies that A and B both are present in transaction set D($D \subset T$) with support count S and confidence C. Here support count is percentage of transaction(T) where transaction contains item-set A. And confidence C is percentage of transaction where A and B both are present to the transaction that contains item-set A[1].
Support $(A \rightarrow B) = P(A \cup B)$.
Confidence$(A \rightarrow B) = P(B \mid A) = \text{support}(A \cup B) / \text{support}(A)$
i.e. There is 10000 transaction T. Out of 10000 transaction T there are 4000 transaction contains item-set A. so, support count s = 40% . Out of this 4000 transaction that contains A there are 1000 transaction that contains item set B. So, confidence = 25%.

A. Multidimensional Association Rule Mining

If association rule is based on more than two attributes then it is called multidimensional association rule. i.e. suppose there is item-set in A and B then we find the probability of item-set present in C ($(A \wedge B) \rightarrow C$)[2].

B. Use of Association Rule Mining in E-Commerce Recommendation System

The dataset that we use for recommendation system is as bellow.
Product_detail P={product_id , product_detail}
Customer_detail C= {customer_id, name, birth_date, gender, location, social_networking_site_profile_id}
Transaction_detail T= {transaction_id, customer_id, purchased_product_id_1, purchased_product_id_2, purchased_product_id_3 …….. purchased_product_id_n }
After signed up in we store the customer profile detail in Customer_detail dataset. All product are added in Product_detail dataset with their product_id and product_detail. when customer purchased any product his customer_id and purchased_product_id is mapped with the unique transaction_id.

International Journal of Innovations & Advancement in Computer Science
IJIACS
ISSN 2347 – 8616
Volume 4, Special Issue
May 2015

1) Recommendation based on customers profile:

Whenever customer login to the site recommendation system should gives recommendation based on the customer's profile. At the time of login, system find the customer's age group , gender , location and social networking site profile detail. From this data, system find all list of customers that are belongs to the same group with signed customer. To find the group of the customer assosiation rule mining algorithm is used [7]. i.e.Customer X who login is 27 year old, male and his location is newyork.

Now multidimensional association rule is used as below.

age(X , [25,30)) ^ gender(X, "male") → location(X, "newyork")

This query returns all the list of the customers_id who belong to appropriate group. Now, from this customer_id we find the set of the transactions (D where D ⊂ T) that contains same customer_id. From this transaction set D we find itemset of the products that belongs to transaction set D. Now we sort this itemset according to descending order of the number of occurrence (support count) in transaction set D.

i.e. product_1 is 576 times in D and product_2 is 127 time in D then we predict that product_1 is more required to particular customer. So, by this way from the customers profile we generate the recommendation for each customer.

2) Recommendation based on customers social networking site profile:

At the time of signup to the site it ask for login using social networking site. OAuth2 protocol is use to access the customers information from the third party website. Customer permission is also require to access the customer's data from the third party website. After customer permission recommendation system find the set of the customer's friend on social networking site. Out of all the friends we find the list of the customer's friend who also use the same site for purchasing the products. Then we find the transaction set that belongs to the customer's friend list. Now from this transaction set recommendation system gives suggestion according to the same procedure used in previous section

3) Recommendation based on customer's cart:

Suppose customer X purchase some product from the available product set. We can also generate recommendation from the list of customer's purchased products. Suppose customer X purchase product_1 and product_2. Now recommendation system make transaction set D (D ⊂ T) that contains product_1 and product_2. Now system finds the confidence of product_1 and product_2 with respect to the available all the product in transaction set D.

buy(X,"product_1")^buy(X,"product_2")→buy(X,"product_X ")

Here we put all the other product that belongs to the set D in 'product_X' and find the confidence of product that may be require to the customer. According to the confidence system sort the recommendable product in descending order of confidence. So, by this way using customer purchased product recommendation system suggest product that may be required to the customers [8].

C. Limitation of Association Rule Mining Method for Recommendation System

In real time system there are lacs of transaction available for generate the recommendation. For generating the rules, system has to read all the transaction T and generate the subset of the transaction D. In set D suppose system find n number of unique product from transaction d. So to find the rule system read set D for n*d times. For real time application this procedure high time consuming and it also require high computation. To overcome this problem clustering technique is used.

International Journal of Innovations & Advancement in Computer Science
IJIACS
ISSN 2347 – 8616
Volume 4, Special Issue
May 2015

3. CLUSTER ANALYSIS

Cluster is a collection of similar objects that belong to the same class. In other words the similar objects are collected in one cluster and different are collected in other cluster. However classification techniques are use to differentiate the groups and its object. For creating the cluster each object should be unique classify and assign the unique label of the cluster. It is very costly and time consuming process. Assigning the label is based on the training tuples that we provide for clustering. This type of clustering is called supervised learning. In unsupervised approach first divide the set of data into collections based on characteristics of the dataset and then assign the unique labels to each the data set of groups. In this approach training data set is not requires for creating the cluster. This approach is requires less computation compare to supper vice approach [3].

The following are typical properties of clustering in data mining.

i. Scalability
ii. Ability to deal with different types of attributes
iii. Discovery of clusters with arbitrary shape
iv. Incremental clustering and insensitivity
v. Ability to deal with noisy data
vi. Constraint-based clustering
vii. Interpretability and usability
viii. High dimensionality

A. Types of Clustering Methods

The various types of clustering method is as below.

i. Partitioning methods
ii. Hierarchical methods
iii. Density-based methods
iv. Grid-based methods
v. Model-based methods

B. Hierarchical Clustering method

A hierarchical clustering method creates a hierarchical fragmentation of data from the available dataset. There are two techniques for hierarchical clustering.

i. Top-Down approach
ii. Bottom-Up approach

In top-down approach whole dataset is classify into no. of predefined clusters. This method initiate clustering process with all of the objects in the one cluster. After each repetitive iteration, a cluster dataset is divides into smaller clusters, until each single object be a cluster, or until a terminating condition is occurs. Termination condition is use to make the restriction on the clustering method.

In bottom-up approach first collect individual object from the dataset and create individual cluster for particular object. After creating the cluster if similar object is found than this new object is added to the cluster. This method is used when cluster in not predefined and it create dynamically[4].

C. Use of Heirarchical cluster for E –Commerce Recommendation System

For this recommendation system we use both top-down and bottom up approach for hierarchical clustering. As shown in Fig.1 we create cluster based on the customer's profile. First we split transaction dataset into two predefined group based on the gender so, here we use top-down approach. Then we create sub-cluster based on the customer's age group. After we split each sub cluster in to customer's location. we use bottom-up approach for partition data using location because each sub cluster have different location so, location may be different for each sub-cluster[5].

Now each sub-cluster is splits into divides into sub-cluster based on the transaction. We use bottom-up approach for implement this method. System pick-up each transaction and find it's each object and create different cluster for each object. Then it add same transaction in to each cluster according to the object in each transaction. Termination condition is use to make restriction on this clustering method [11]. i.e. suppose we make terminating condition that if count(cluster_transaction) >= 1000 then apply clustering technique. So, each end cluster node contain transaction less than 1000.

4. Recommendation System using Association rule mining and clustering

In previous section we discuss that how recommendation system works using multidimensional association rule mining algorithm but it is very time consuming process to generate the recommendation from the large set of the data. So, for real time application recommendation system only based on association rule mining as impractical. Hierarchical clustering technique partitioning data set into small group according to the properties of the dataset. Now we apply association rule mining technique to particularluster only based on customer profile and purchased item [6].

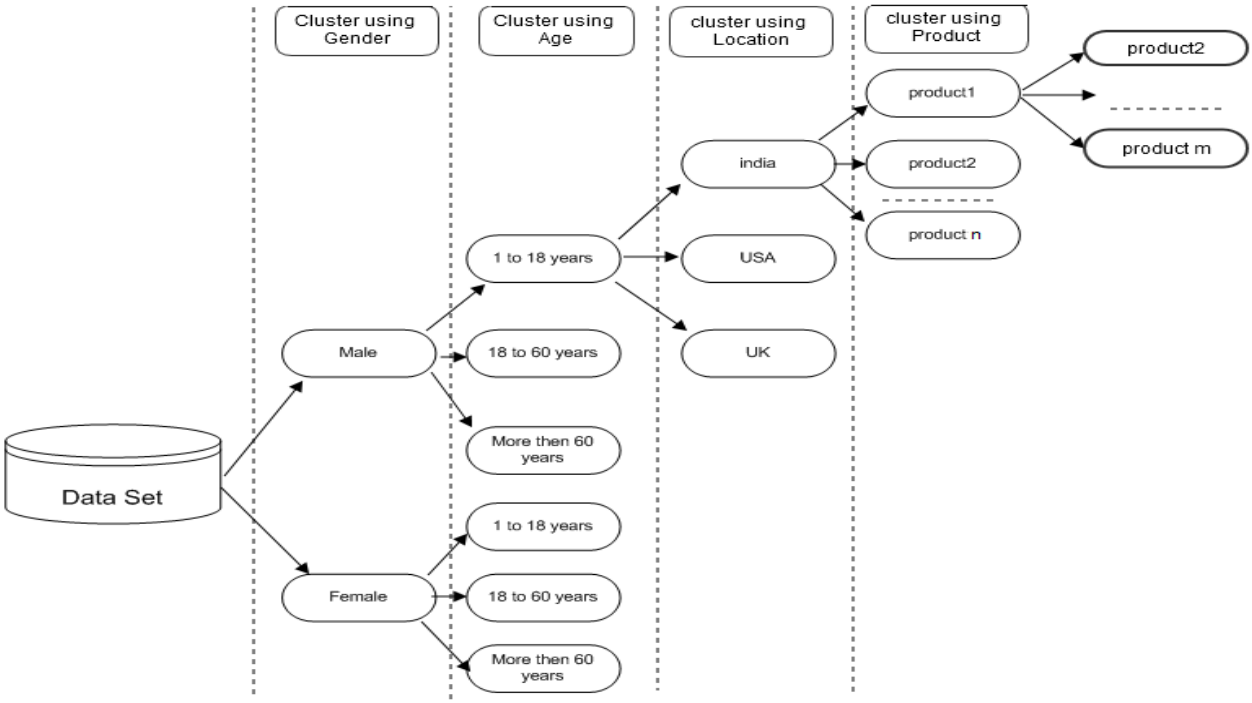

Fig. 1. Hierarchical clustering for e-commerce recommendation system

Generating cluster is very time consuming process. For hierarchical clustering time require to create clusters is totally depends on no. of level of clustering that we want to create. And no. of level is depends on maximum no. of object in each cluster. If data in the cluster grater then the threshold value (maximum no of object) than it again partition in to sub-cluster. If threshold value is low than we want to it create more cluster and it takes more time for clustering. If no. of object in the cluster is low than finding association rule is quicker. So, as we increase the threshold value creating cluster time is decrease but recommendation generation time is increase [10].

For experiment we are using customer transaction dataset that contains 10000 transaction. For different threshold value we generate the results for creating cluster time and generate recommendation time.

International Journal of Innovations & Advancement in Computer Science
IJIACS
ISSN 2347 – 8616
Volume 4, Special Issue
May 2015

Here Table 1 and Fig.2 shows time to create cluster with respect to maximum size of the cluster.

Maximum size of cluster	Time to create cluster(in second)
100	1174
200	711
500	415
1000	298

Table 1.Maximum cluster size to Time for create cluster

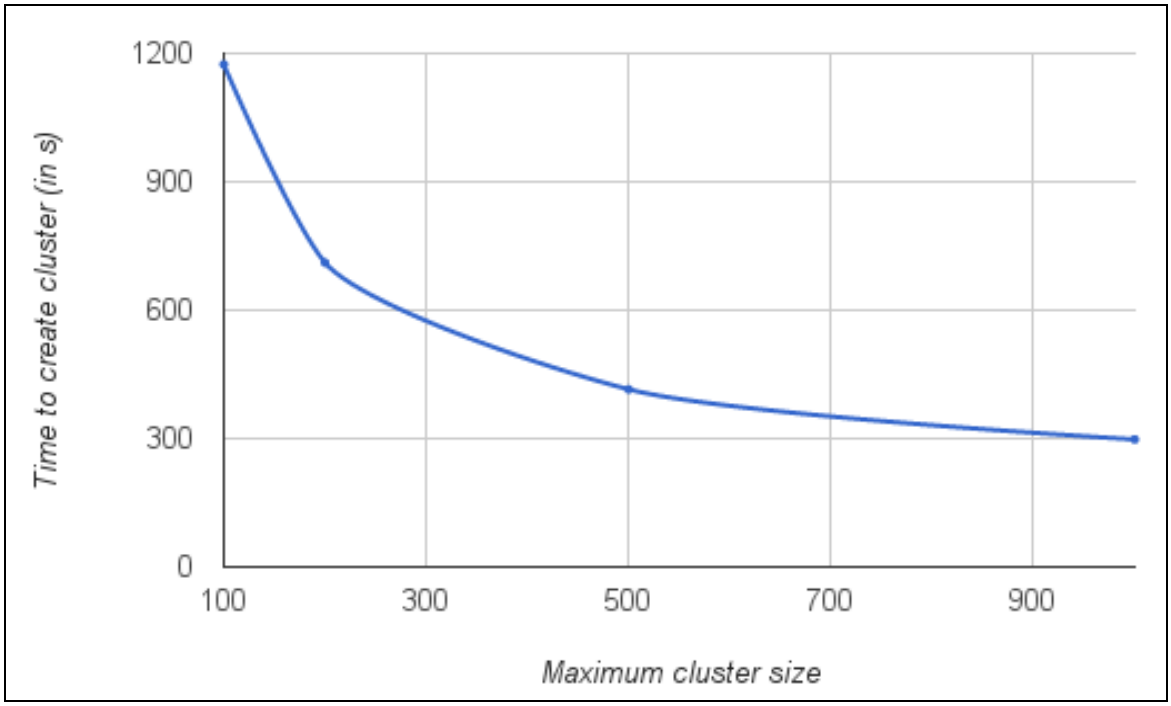

Fig.2. Maximum cluster size to Time for create cluste chart

Table 2 and Fig.3 shows time to generate recommendation with respect to maximum size of the cluster.

Maximum size of cluster	Time for generate recommendation(in milliseconds)
100	98
200	268
500	842
1000	1468
10000(without clustering)	6121

Table 2.Maximum cluster size to Time for generate recomandation

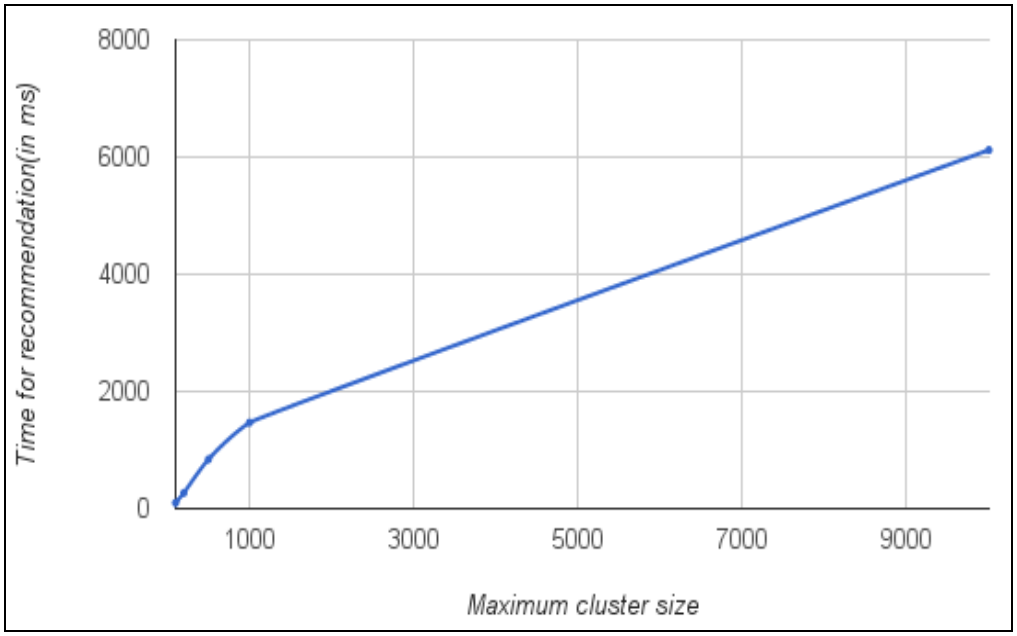

Fig.3.Maximum cluster size to Time for generate recommendation chart

5 Conclusion

Using E-Commerce Recommendation System we can find the frequent item-set for real time application that customer may want to buy. From customer's profile data like age, gender, location and other customer's transactions, system can find the behavior of customer for buying the products. Using association rule mining we can find rules for generate recommendation to buyers. For recommend accurate product we have to find the rule from large data-set. And for large dataset it takes more time for finding the rules which is infeasible. Clustering divides whole data-set into more than one group of data that belongs to same class . And it used to reduce the time for generating the rule. So, using association rule mining and clustering technique we can make recommendation system efficiently that generate recommendation in less time.

References

[1] Zhanlong Wang, Xiufeng Piao, Gang Liu, "Research on Mining Positive and Negative Association Rules Based on Dual Confidence" Harbin , China in Fifth International Conference on Internet Computing for Science and Engineering, 2010.
[2] Shingo Mabu, Nannan Lu, Tuo Wang and Kotaro Hirasawa," Efficient Hybrid Rule Pruning for Intrusion Detection using Multi-Dimensional Probability Distribution" Tokyo, Japan in SICE Annual Conference, 2011.
[3] Xinshun MA, Tongju Shi, "Web-based Software Development for Cluster Analysis Experiment",Hangzhou in International Conference on Multimedia Technology (ICMT), 2011.
[4] Sukumar Nandi, Bidyut Kr. Patra, P. Viswanath "Data summarization based fast hierarchical clustering method for large datasets" International Conference on Information Management and Engineering,2009.
[5] Xue-ling HAN, Ya-min WANG, Xiao-wei LIU, "E-commerce Recommendation System Based on CBR and Web Log Mining" IEEE 18Th International Conference on Industrial Engineering and Engineering Management (IE&EM),2011.
[6] Mingrui Chen, Zhao Qiu, Jun Huang "Design of Multi-mode E-commerce Recommendation

System" Third International Symposium on Intelligent Information Technology and Security Informatics,2010.

[7] Yongjian Fan, Jianying Mai, Xiaofei Ren "A Rough Set-based Clustering Collaborative Filtering Algorithm in E-commerce Recommendation System" International Conference on Information Management,2009.

[8] Ansel Y. Rodríguez-Gonzalez "Mining frequent patterns and association rules using similarities" Expert Systems with Applications,2013

[9] Hsiao-Fan Wang, Cheng-TingWu "A strategy-oriented operation module for recommender systems in E-commerce" Computers & Operations Research,2012

[10] Sunita B. Aher, L.M.R.J. Lobo "Combination of machine learning algorithms for recommendation of courses in E-Learning System based on historical data" Knowledge-Based Systems ,2013

[11] Unil Yun, Heungmo Ryang, Keun Ho Ryu "High utility itemset mining with techniques for reducing overestimated utilities and pruning candidates" Expert Systems with Applications,2014

[12] ZHANG Zheng, WANG Hui-wen "An Efficient Association Rule Mining Algorithm and Business Application", International conference on Communication,2007