E-Commerce Recommendations Using Multidimensional Association Rule Mining With Clustering A Review

AAKASH M SHAH

M.E Student, KITRC College, GTU, Ahmedabad, India.

ABSTRACT:

E commerce means trading in products, services using internet. Due to increasing popularity of internet e commerce is developing rapidly in last few years and purchasing products online has become social mainstream. On E commerce site customer are submerged with huge amount of information so in this situation it is not easy for customer to find the product they needed also merchant face the problem of losing customers.in this situation E commerce recommendation comes into exist. E commerce recommendation helps customers in what to buy and also help in their purchase process. An important topic is here to apply multidimensional association rules to E commerce database with clustering. This paper reviews some of the recent technology and methods used for E commerce recommendation. The aim is to illustrate what can be done and also to identify trends in these areas where further work is needed.

Keywords: — E commerce recommendations, data mining, multidimensional association rule, clustering, proposed system on E commerce recommendations.

I. INTRODUCTION

This paper presents a comparative study of the different data mining technology used for E commerce recommendations recently used. Earlier recommendation system of E commerce websites are almost static means they show same information to all the users. Personalized recommendation to the user based on their profile is not provided in early days of E commerce site. Personalized recommendations are based on user's interest, hobbies, habits, and correlation between aptitudes of different users to online recommended products. Personalized recommendations provide user friendly access and personalize services to the user and get a chance to customize the website for particular user. This review paper presents the how multidimensional association rule is applied on E commerce database what kind of association rules are available .Main advantage of using association rule to generate recommendations is it does not require specific domain knowledge in advance and it also find new point of interest for products. This paper is mainly focused on the E commerce personalized recommendation. How to get more accurate recommendations with minimum time and more specific to user.How clustering is applied with multidimensional association rules to generate recommendation. The main advantage of recommendation is beneficial for both seller and buyer. On the customer side it provide Better user experience as well as Customer Satisfaction and on the seller side Increased Sales ,Improve customer retention, Revenue,Increased Traffic/Page Views.

So main aim of any e commerce site to convert every visitor or viewer into buyer is get to be fulfilled by E commerce   recommendations.

II. E COMMERCE RECOMMENDATIONS AND TECHNOLOGY USED

In the past if any customer wanted to buy any product, item, services he had to go different stores to compare various prices and offers provides by different stores on different products of various brands. It is very time consuming and costly especially for that product that is not in neighborhoods location of customer and is newly arrival in market. Another difficulty with traditional method is also not suitable for old age peoples or those people who live in remote area from main markets & in extreme weather condition. Thus E commerce comes into picture.

As discussed earlier static recommendation and personalized recommendation are the two types for E commerce recommendations. Static recommendation shows same product recommendations to all users it is based on previous static transactional database. And personalized recommendations is show different recommendation for different particular user.it use customer profile, goods commodity relationship, hobby, preference of user, locality, interest, budget, history of customers buying pattern to give personalized recommendation.

The personalized recommendations of the particular e commerce website helpful for their sites to increasing marketing capacity and customer loyalty in three ways. (1)Transforming E commerce system's visitor into buyer. (2) Enhance cross marketing and selling of products on E commerce system. (3)keeping the customer for long duration .[1][2] on that site and preventing from information overload .some another advantages of E commerce recommendation is

increase traffic or page views on site, it useful for increase customer retention, it provide better user experience, increase revenue of site, it's provide better customer satisfaction, increasing sell. The e commerce recommendation system model used data mining Technology commerce application's database contains information about customer previous purchase similarly web server preserving information about customer log on message, customer visiting information, grading and voting information about different products. This data contains fruitful knowledge which plays an important role in taking business decision to the e commerce company. The basic knowledge for recommendation is to be found by different data mining technology. The rich knowledge is obtained from the raw data like customer purchasing pattern, customer hobbies, habits, profiles.

III. DATA MINING AND IT'S TASKS

Data Mining is defined as extracting the information from the huge set of data. In other words we can say that data mining is mining the knowledge from data. Data mining is extraction of useful patterns from data sources, e.g., databases, texts, web, image. [11]

❖ **Need of Data Mining**

In field of Information technology we have huge amount of data available that need to be turned into useful information.

This information further can be used for various applications such as market analysis, fraud detection, customer retention, production control, science exploration etc.

Data Mining deals with different kind of patterns to be mined. On the basis of what kind of be mined there are two kind of functions involved in Data Mining, that are listed below.[11][12]

❖ **Descriptive:**

‣ Class/Concept Description**:**

‣ Class/concept description refers the data associated with class/concepts. For example, in a company classes of items for sale include smartphones and covers, and concepts of customers include big spenders and budget spenders. Such descriptions of a class or a concept are called class/concept descriptions.

　o This can be derived by two ways Data Characterization - This refers to summarizing data of class under study. This class under study is called as Target Class; Data Discrimination-It refers to mapping or classification class with some predefined group or class.

‣ Mining of Frequent Patterns**:**

　o Frequent Item Set - It refers to set of items that frequently appear together in database for example milk and biscuits,

　o Frequent Subsequence- A sequence of patterns that occur frequently in database such as purchasing a laptop is followed by earphone, Frequent Sub Structure - Substructure refers to different structural forms, such as graphs, trees, or lattices, which may be combined with different item sets or subsequence.

‣ Mining of Associations: Association mining is used in retail sales to identify patterns that are frequently purchased together .This process relates to uncovering relationship among data and determining association rules .For example a retailer generates association rule that show that 80% of time milk is sold with bread and only 20% of times biscuits are sold with bread.

‣ Mining of Correlations: It is used for finding additional analysis performed to uncover interesting statistical correlations between associated-attribute- value pairs or between two item Sets to analyze that if they have positive, negative or no effect on each other.

‣ Mining of Clusters: Cluster mining refers to a group of similar kind of objects. Cluster analysis refers to creating group of objects that are very similar to each other but are highly different from the objects in other clusters.

❖ **Prediction**

‣ Classification and Prediction: classification is the process that describes data classes and concepts the main purpose is to use this model to predict the class of objects whose class label is unknown. This derived model is based on analysis of set of training data. The derived model can be presented in the following forms

‣ Classification (IF-THEN) Rules:

‣ Decision Tree

‣ Mathematical Formulae

‣ Neural Networks:

Here is the list of functions involved

• Classification: It predicts the unknown class label's objects. It finds a derived model that describes and distinguishes data classes or concepts. The derived model is based on analysis set of training data. The data object whose class label is known

• Prediction - missing or unavailable numerical data values can be find using these techniques without using class labels. Regression Analysis is generally used for prediction. Prediction can also be used for identification of distribution trends based on available data.

IV. Association rule mining and Cluster analysis

▶ Association rule mining means Finding frequent patterns, associations, correlations, or causal structures among sets of items or objects in transaction databases, relational databases, and other information repositories.

▶ Given:

(1) Database of transactions,

(2) Each transaction is a list of items purchased by a customer in a visit

Find: <u>all</u> rules that correlate the presence of one set of items with that of another set of items

• E.g., *98% of people who purchase milk and bread also get biscuits.*

• Applications: CRM of credit card business, Protein sequences. Banking: loan/credit card approval Customer relationship management, Targeted marketing, Fraud detection: telecommunications, financial transactions, Manufacturing and production, Medicine: disease outcome, effectiveness of treatments[12]

Different types of association rules[12]

(1) Single dimensional Association rules: Association rules that imply a single predicate that is in example given here, the predicate *buys*. For instance, in mining our *All Electronics* database, we may discover the Boolean association rule. It refers to single dimensional or intra dimensional association rule.it contains single distinct predicates with multiple occurrences.

Buys(X, "Digital Camera")$\Rightarrow$Buys(x,"Hp printer")

(2) Multidimensional Association rules: Association rules that involve two or more dimensions or predicates can be referred to as multidimensional association rules. Rule below contains three predicates (*age, occupation*, and *buys*), each of which occurs *only once* in the rule.

Age(X, "20….29") $\wedge$Occupation(X, "Student")

$\Rightarrow$Buys(X, laptop)

(3) Hybrid Association rules**:** mining multidimensional association rules with repeated predicates, which contain multiple occurrences of some predicates, is called hybrid-dimensional association rules. An example of such a rule is the following, where the predicate *buys* is repeated: These rules are called hybrid-dimensional association rules

Age(X,"20….29")$\wedge$Buys(X,"laptop")$\Rightarrow$Buys(X,"Hp printer")

What is Clustering?

Custer is the group of objects that belongs to same class. Means similar group of objects belongs to one cluster and dissimilar group of objects belongs to another cluster. Clustering is the process of making group of abstract objects into classes of similar objects. [12]

Data object's clusters can be treated as one group. While performing cluster analysis first portion set of data into groups. And based on similarity assign labels to that group. Main advantage of cluster is adaptable to change that helpful to find out useful features that distinguishes different groups.

Applications of Cluster Analysis**:** market research, pattern recognition, data analysis, image processing, characterizing customer groups based on purchasing pattern, classifying documents on the web for information discovery, detection of credit card fraud. In field of biology it can be used to derive plant and animal taxonomies

Requirements of Clustering in Data Mining**:** Scalability, Ability to deal with different kind of attributes, Discovery of clusters with attribute shape High dimensionality Ability to deal with noisy data Interpretability

❖ **Clustering Methods: The clustering methods can be classified into following categories:**

▶ Partitioning Method: Suppose we are given a database of M objects, the partitioning method construct K partition of data. Each partition will represent a cluster and k≤m. It means that it will classify the data into k groups, which satisfy the following requirements. First each group contain at least one object and second each object must belong to exactly one group.[11]

    ▶ Hierarchical Method: this method hierarchical decomposition of the given set of data objects how the hierarchical decomposition is formed based on that we can classify into two methods

    ▶ Agglomerative Approach**:** This approach is also known as bottom-up approach. In this we start with each object forming a separate group. It keeps on merging the objects or groups that are close to one another. It keeps on doing so until all of the groups are merged into one or until the termination condition holds.

    ▶ Divisive Approach: This approach is also known as top-down approach. In this we start with all of the objects in the same cluster. In the continuous iteration, a cluster is split up into smaller clusters. It is down until each object in one cluster or the termination condition holds.

    ▶ Density-based Method: This method is based on the notion of density. The basic idea is to continue growing the given cluster as long as the density in the neighbourhood exceeds some threshold i.e. for each data point within a given cluster the radius of a given cluster has to contain at least a minimum number of points.

    ▶ Grid-Based Method: In this the objects together from a grid. The object space is quantized into finite number of cells that form a grid structure. The major advantage of this method is fast processing time. And it is dependent only on the number of cells in each dimension in the quantized space.

  ▶ Model-Based Method: In this method a model is hypothesize for each cluster and find the best fit of data to the given model. This method locates the clusters by clustering the density function. This reflects spatial distribution

of the data points. This method also serve a way of automatically determining number of clusters based on standard statistics, taking outlier or noise into account. It therefore yields robust clustering methods.

Constraint-based Method: In this method the clustering is performed by incorporation of user or application oriented constraints. The constraint refers to the user expectation or the properties of desired clustering results. The constraint gives us the interactive way of communication with the clustering process. The constraint can be specified by the user or the application requirement.

V. .LITERATURE REVIEW ON DIFFERENT E COMMERCE RECOMMENDATIONS TECHNIQUES

These tables shows comparative analysis of different recommendation technology used what are the input parameters how process are performed to generate recommendations result.

Recommender technology	Input	process
Collaborative filtering	User U ratings to item I	Get users in U similar to target U and produce recommender based on ratings of i
Content-based filtering	Targeted users U ratings to item i	Get a classification that fits Target u on items I
Mining association rules	The records of user u buy item i	Get association rules and produce recommender
Demography based filtering	Use demography information	Get users in U demography Similar to target U and produce recommender based on ratings of demography
Knowledge-based filtering	Users U needs of interest	Get a match between target U and others

Table 1. Different E commerece recommadation Technology used[1]

Content based filtering: content based filtering of E commerce recommendation is based on comparing items similarity between users not based on similar laity between functions. Content based filtering simply ignores customer's buying behavior it only focus on goods and commodity relationships. The advantage is simply and effective. Modeling and measurements of similarity between goods done offline so recommendation response time is also less. Disadvantage it is difficult to distinguish between different goods information, quality style. And it does not found information for that customers they are interested in buying new products.

Collaborative based filtering: collaborative filtering used similarity between users so this filtering based on user's expressed interest, preferences, and test of the particular user.[1]

Collaborative filtering algorithms generally require (1) users' active participation, (2) an easy way to represent users' interests to the system, and (3) algorithms that are able to match people with similar interests.

1. A user expresses its interest by rating items (e.g. books, movies or CDs) of the system. These ratings can provide approximate representation corresponding domain.
2. The system matches this user's ratings against other users' and finds the people with most "similar" preferences.
3. With similar users, the system recommends items that the similar users have rated highly but not yet being rated by the new user because absence of rating is often considered as the unfamiliarity of an item
4.

Demography based filtering: demography based techniques used in this recommdation.it provides recommendation based on demography information and personal characteristics of the user. In addition, some systems use machine learning to get classifier based on demographic information. , its advantage is that users do not need to evaluate historical data; however, the sensitivity of personal information the user is to obtain demographic information barriers.

Knowledge based filtering: knowledge based filtering recommendations is based on user's needs, preferences. It does not dependent upon commodity rate customer data. Access to the document summary can be recommended based on knowledge of technology is divided into three categories: knowledge discovery in databases, case-based reasoning and knowledge inference, in which the main and traditional KDD recommended technologies are used.

Association rule mining: In this Transaction database is used for mining association rules and produce recommendation. statistical purchased goods set A, what percentage of users purchase the product set B. main advantage is it does not require domain knowledge and find new points of interest. If the confidence of rule is more than the required threshold we can set A to buy goods in commodity of user's recommendations product set Y the user has not purchase yet.

Recommender technology	Advantage	Disadvantage
Collaborative filtering	Without domain knowledge, performance improvement over time	Cold start problems, sparse problems, new user Issue,etc..
Content based filtering	Easy to explain, without domain knowledge	Complex attribute poor treatment, Sufficient data classification structure.
Mining association rules	Find new points of Interest, Do not domain knowledge	Association rule extraction is difficult, time consuming; Low degree of personalization
Demography based filtering	No new user issue, not use Domain knowledge	D information valuable to the statistical information
Knowledge based filtering	Consider non product Attributes, To map user needs to products;	Hard to get Knowledge, Recommended is the static

Table 2 Advatages and disadvantages of different recommadation technology[1]

RESEARCH PAPER	TECHNOLOGY USED
E-commerce Recommendation System Based on CBR and Web Log Mining[2]	CBR (Case-based Reasoning) and web log mining
A Rough Set-based Clustering Collaborative Filtering Algorithm in E-commerce Recommendation System[3]	a rough set-based clustering collaborative filtering algorithm, classifier model based on rough set for the pre- classification to items and give realization of clustering
Clustering technology application in e-commerce recommendation system[4]	Content-Based Filtering and Collaborative Filtering.
Application and Research of Collaborative Filtering in E-commerce Recommendation System[5]	a improved collaborative filtering algo cluster. solve the scalability problem , better effect in the case of sparse data
Research on Personalized Recommendation System for e-Commerce based on Web Log Mining and User Browsing Behaviour[6]	introduces the UserID- URL associated matrix according to log information, We cluster users into user groups.

Table 3 Survey of technical parameters used for E commerce recommendations

VI. **Future scope**:

In future we can use customer profile also in order to increase accuracy of recommended products. Analyze and understand all the literature. I conclude that by applying Multidimensional association rule mining with clustering on E commerce database to generate recommendation we can increase the accuracy of personalized recommendation and another benefit is it also generate results in less time duration. This one will be my proposed system. The main problem with association rule mining is solved by using clustering. This one is important factor for any commerce site because if we display more relevant product according to customer preference it would results into increase revenue of the site &customer loyalty.

Reference

1. Minchin Research on Recommender Technology in Ecommerce Recommendation System 2010 2nd International Conference on Education Technology and Computer (ICETC)

2. Ya-min WANG, Xue-ling HAN, Xiao-wei LIU E-commerce Recommendation System Based on CBR and Web Log Mining 2011 IEEE
3. Yongjian Fan Jianying Mai, XiaofeiRen A Rough Setbased Clustering Collaborative Filtering Algorithm in Ecommerce Recommendation System 2009 International Conference on Information Management, Innovation Management and Industrial EngineeringInternational Journal of Trend in Research and Development Volume 1(2) www.ijtrd.com IJTRD | Nov-Dec 2014 Available Online@www.ijtrd.com 6
4. Wu JingHui Liu QiangLuoSiWenClustering technology application in e-commerce recommendation system International Conference on Management of e-Commerce and e-Government2008 IEEE
5. Hu JimningApplication and Research of Collaborative Filtering in E-commerce Recommendation System 2010 IEEE
6. Xia Min-jie Zhang Jin-ge Research on Personalized Recommendation System for e-Commerce based on Web Log Mining and User Browsing Behaviours2010 International Conference on Computer Application and System Modelling (ICCASM 2010)
7. LuoYa E-commerce Personalized Recommendation Technology Based on Expert System
8. Wei Zhang Na zhao Research on the FP Growth Algorithm about Association Rule Mining 2008 International Seminar on Business and Information Management
9. AnimeshTripathy,Subhalaxmi Das Prashanta, Kumar Patra An Improved Design Approach in Spatial Databases Using Frequent Association Rule Mining Algorithm 2010 IEEE
10. A Bayesian Association Rule Mining Algorithm 2013 IEEE International Conference on Systems, Man, and Cybernetics
11. http://www.tutorialspoint.com/data_mining/dm_tasks.html
12. DM Concepts & Techniques _ Han&Kamber