Volume 6, Issue 1, January 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Cross Category Product Recommendation System Using Clickstream Mining

[1]Amarjeet Bathija, [2]Jitesh Budhwani, [3]Surabhi Kanth, [4]Deepika Rohra, [5]Manisha Gahirwal
[1, 2, 3, 4] Student, Computer Engineering Department, VES Institute of Technology, Mumbai, Maharashtra, India
[5,] Assistant Professor, Computer Engineering Department, VES Institute of Technology, Mumbai, Maharashtra, India

Abstract— Due to the availability of Internet, people can search for all the information they need and buy whatever they want on the web. In the age of E-Commerce, it is difficult to provide support for customers to find the most valuable products that match their heterogeneous needs. Traditional approaches adopt pre-defined formats to describe the customer requirements, leading to inaccurate recommendations. In this paper, we propose a recommendation system using sequential pattern association rule mining and the item clustering to produce the recommendations. The recommendation will be based just not only on the basis of customers browsing history, but also the system will take into consideration the preferences, shopping behavior and similarity of interest between users to provide recommendations within similar category as well as cross category of products.

Keywords—sequential pattern, clustering, association rule mining, similarity of interests, cross category

I. INTRODUCTION

In recent times, there's been a huge increase in preference for online trading over the traditional approaches. This sudden increase can be attributed to the ease of use, without giving any regard to distance and time and at the same time online trade reduces the cost of operations significantly.

As a result, a large number of data has become available on internet [7]. So as easy it is to gather data, it has become a tedious task to sift through and get relevant information. Our system proposes a methodology which will filter and give recommendations to a user, of an e-commerce platform, and help him/her in making satisfactory choices. Our system comes as an application of web usage mining which itself is a consequence of web mining. Web mining can be defined as discovery and analysis of useful data on World Wide Web. It is classified in three categories and they are web content mining, web structure mining and web usage mining. Web content mining involves analysis of content i.e. text, audio, video, images etc. Web structure mining focuses on analysis of hyperlinks and structured organization.

Web usage mining is further based on application level logs, which are generated by an application regarding its usage, application server logs, which are generated due to interactions between applications and server and web server logs which are generated by client side browser or server itself. Clickstream mining is an extension of web usage mining. Clickstream mining can be defined as extraction and analysis of clicks made by a user on a
website during certain sessions. Thus, number of clicks and their patterns give us the inclinations of customers for certain products.

This paper proposes recommender systems for products that uses a sequential pattern mining method, association rule mining and clustering method. By combining these methods, we build a system with a higher accuracy of predictions. The proposed methods are based on combination of the above mentioned algorithms. The traditional approach considers only the either the navigational behaviour of each user to provide recommendation based on his past preferences or group users with a similarity attribute and provide recommendations on the basis of similar taste preferences. So it can collect irrelevant preferences of user. This assumption ignores the fact that a user's interests may be changed over time. Then it can recommend irrelevant items in inappropriate time.

Therefore, we propose a new combination of approaches, to take user's interest according to past and present preferences into account.

II. MOTIVATION

Recommender system helps users to process huge data and find the information or items of their interest. It has been used in communities where items like products, movies, music, books, news, etc. are suggested to users. A recommender system is a system to help customers to find what they want easily and help them to discover products they like but are not looking for. The traditional techniques of sequential pattern mining for generating recommendations suffered from increased computation complexity and storage maintenance Main goal of our recommender system is to suggest items in diverse categories which reflect changes. The user should be able to relate his preferences with the recommended ones. It should reflect the similarity of interest between different users with respect to different products. The sequence mining should generate patterns that are in relevance to the timestamp value in addition to clustering and association such that complexity is reduced and maintenance is efficient.

III. PROBLEM DEFINITION

The main objective of the project is to propose a model for product recommendation to different customers with similar interest patterns as well as products of different categories to same customer based on their browsing history and shopping preferences with respect to similarity with other users. Most of the existing models of recommendation system algorithms provide recommendations to customers either within the one specific category or only depend on their browsing history. For example if A user purchases shoes on an e-commerce website gets recommendation of shoes of other or similar brands only. Such assumptions provide irrelevant and limited recommendations .The proposed model will take into consideration the association between user-to-user as well as user-to-product by examining the various attributes of users and commodities involved.

IV. RELATED RESEARCH

Mohsin Riaz, Ansif Arooj,Malik Tahir Hassan,and Jeong-Bae Kim proposed a recommendation model for US based online retail store [4]. The model segregated data set and made the clusters of the similar attribute data. For example they segregated customers in the data set on the basis of age and gender. At first, they made two clusters on the basis of gender. The values in the field of gender in customer table were checked and two clusters C1 and C2 of males and females respectively were made. Then three clusters were made on the basis of age as follows C1 consisting of age 1-25, C2 consisting of 25-40 and C3 of above 40. Hence on the basis of clusters sequential web access patterns were observed and association rules were designed. For example, gender M and gender group 25-40 buys Gillete mark and sensitive mouth wash and confidence level is 95%. Hence, users in cluster of gender M and age 25-40 should be recommended other product from that pair if one is purchased by him.

Zhang Xizheng proposed personalized recommendation system using Association rule-based mining [2]. In this system, semantic analysis were fed to the system as an input and then input was processed in four stages namely 1) Requirement extraction.2) Association Rule generation.3) Classification module.4) Performance validation. In requirement extraction, requirements stated by users in natural language were analysed semantically. Then semantic analysis were utilized in association rule generation module to build classifiers. Apriori Algorithm was adopted to acquire the classifiers. Then the generated classifiers were pruned and only good quality classifiers were kept for recommendations. When new information came, the system identified the corresponding class labels using multiple classifiers. Finally the performance of the system was validated to evaluate how accurately the system gives recommendations.

Arthur Pitman and Markus Zanker proposed a recommendation model based on sequential pattern mining using click-stream data. The model divided the process into three main stages-sequential pattern mining stage, rule construction stage and rule application stage [6] whereas WANG Xiao-Gang explained the detailed model of sequential pattern mining [5]. The model consisted of following three stages-sequential pattern mining, pattern tree construction and recommendation rules generation. In the first stage web log database is collected from web server which consists of transactions of items. Then this list is ordered with respect to timestamp and checked for access sequences. Minimum support that is number of data sequences containing that pattern is decided and on the basis of those frequent sequential patterns is determined. Then in second stage the obtained sequential patterns are stored in the tree known as pattern tree which was further used in rule generation stage for generating rules which gives list of products to be recommended.

The proposed design used CS miner for sequential pattern mining.

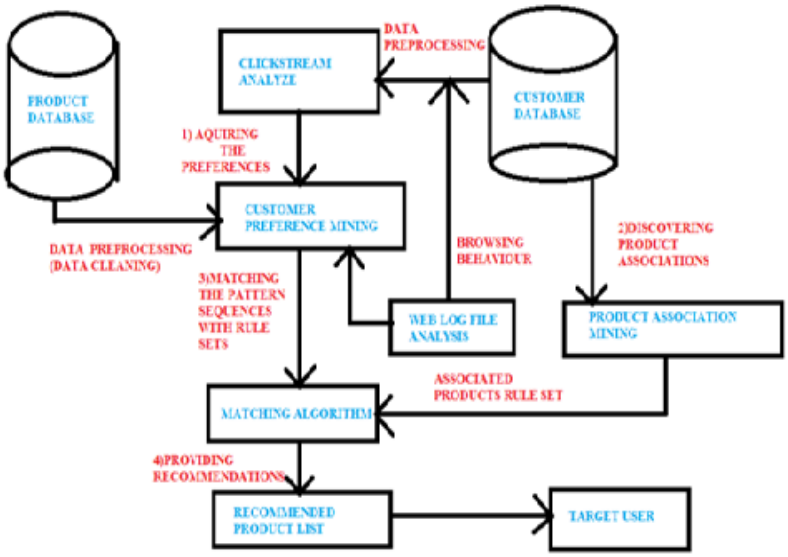

Fig 1. Framework of the proposed recommender system

V. PROPOSED MODEL

We propose a system which will combine all approaches used till now and make use of best of them.

So we're going to use association mining which will be a consequence of sequential pattern mining and clustering. Thus, for example, if a user A buys item I then according to his/her browsing history item J will be suggested and item K will be suggested as consequence of similarity between other users sharing same interests [1].

There are three basic steps involved in the proposed recommendation system:

1) Data pre-processing which involves data cleansing and transformation.
2) Customer preferences mining on the basis of segmentation of customers into similar groups (clustering) and analyzing web log files for browsing history.
3) Product association mining which includes formation of relationships between products from the sequence patterns.

Lastly, based on the above customer and product matching patterns, a list of recommended products for each user is generated.

VI. METHODOLOGY

Web usage mining attempts to discover useful information obtained from the interactions of the users with the Web [7]. The raw dataset consist of customers profile information, product description, customer's shopping behaviour integrated with the web log file which consist of URL of the webpage, access time, session ID and timestamp value.
The different phases involved are:

A. Data Pre-processing

In this phase, the dataset completeness is achieved by ensuring that all the customer and product related information is present which will be needed for our project. The web log files are also checked
for all the required fields like sequences in which web pages are visited, access time of users on web pages and other history related information.
Elimination of redundancy is ensured by removing extraneous information such as not required products, user's that have unusual browsing behaviour and other attributes that are not necessary with respect to this system.
Further, irrelevant fields are removed in order to ensure complete consistency of data present in the dataset.
The tool which can be used for data pre-processing is WEKA 3.7.11 so that the dataset is filtered from unrequited and redundant information.

B. Clustering

Clustering is a technique which is used to group items or users. Segmentation of the whole dataset into different smaller groups, on the basis of certain logical parameters specifically, homogeneity in an atomic fashion, called clustering, consists of objects that are similar among themselves and dissimilar to objects of other groups. This is an unsupervised learning technique and is very valuable tool for better understanding of data in the form of clusters.
In our study, we would perform segmentation of dataset with respect to 'gender' firstly such that if C is the customer database then dividing into two clusters 'male' C1 and 'female' C2 then C=C1UC2 [4].
Similarly, we would from clusters on two more parameters namely age i.e. one cluster would be of age less than 30, other greater than 30. The last parameter is location where customers are segmented on the basis of location.
After clustering, we perform sequential pattern mining, association rule mining for each cluster to provide recommendations to customers with respect to the cluster they belong

C. Sequential Pattern Mining

On the basis of the access time, session id, URL and timestamp [6] the sequence of each user's browsing behaviour within each cluster is recorded in a table such that it contains the list of sequences of products browsed. This sequence table will be used as the data resource for our association rule based algorithms.

D. Generating Association Rule Set

The list of products of each sequences from the above sequence table are collected to find the association or similarity relation between the products. All such product-to-product relations forms different rule sets. These rule sets are matched with the product list of each user in the sequence table. The derived patterns by constructing of pattern-tree. By construction of pattern-tree, the system has to scan all sequential patterns at once. Pattern-tree can be used for matching the current user interest and generating recommendation according to a proposed generation recommendation rules algorithm. If the pattern matched is more than or equal to predefined threshold support value, the products which are present in the rule set apart from the ones already present in the each sequence in the sequence table is added to the recommendation product list of each user.

E. Providing Recommendations

The recommended product list of each customer would consist of items that could have been added because of the following reasons:

1) The parameter such as age, gender and location cluster to which the customer belongs (similarity with respect to users).
2) The pattern match of the rule set with the sequence pattern (product association matching).
3) Browsing behaviour derived from sequential pattern.

The expected output consists of graphical analysis of the following:

a) The list of products recommended for each user.
b) The number of purchased products from the list of recommended products.

c) The most viewed or purchased product.
d) The acceptance ratio of the system.

VII. CONCLUSION

Recommender systems can help people to find interesting things and they are widely used in our life with the development of electronic commerce. Most of the already existing recommendation systems employ either one of techniques such as sequential pattern mining, collaborative filtering or product association mining .The resulting recommendations are limited and irrelevant to the user interest often.
Our approach of a recommendation consist of a combination of many important techniques of data mining, i.e., pruning, cleaning, clustering, sequential pattern mining and frequent item set mining, for obtaining strong association rules. The objective is to facilitate relevant product recommendations to customers for the quicker and optimum product selection.

The major strengths and weakness of the study are:

1. A precise and well-organized cross product reference method presented.
2. Time efficient model presented that uses lesser time to generate rules because of the prior data pruning on the basis of clusters.
3. Generating rules followed by careful data analysis helps in extracting strategically important buying patterns that can be used for email marketing and promotion packages.
4. Difficult to generate association rules for new products that have a small number of transactions in history to reach the required support threshold.
5. Difficult to generate web log file from real time processing.

Thus our proposed model of recommendation system will provide personalized recommendations on the basis of customer preferences and product-to-product recommendations on the basis of similarity with other customer's interest and product associations. In this way, our system is efficient and effective as it takes into account customers present and past preferences providing suitable recommendations.

REFERENCES

[1] SongJie Gong "A Collaborative Filtering Recommendation Algorithm Based on User Clustering and Item Clustering" in JOURNAL OF SOFTWARE, VOL. 5, NO. 7, JULY 2010
[2] Zhang Xizheng "Building Personalized Recommendation System in E-commerce Using Association Rule-based Mining and Classification" in Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing
[3] Su Sande Ko Ko and Rachsuda Jiamthapthaksin "A Categorized Item Recommender System Coping with User Interest Changes" in International Journal of Machine Learning and Computing, Vol. 4, No. 5, October 2014
[4] Mohsin Riaz, Ansif Arooj, Malik Tahir Hassan, and Jeong-Bae Kim "Clustering Based Association Rule Mining on Online Stores for Optimized Cross Product Recommendation" in The 2014 International Conference on Control, Automation and Information Sciences (ICCAIS 2014) December 2-5, 2014, Gwangju, South Korea
[5] WANG Xiao-Gang and LI Yue "Web Mining Based on User Access Patterns for Web Personalization" in 2009 ISECS International Colloquium on Computing, Communication, Control, and Management
[6] Arthur Pitman and Markus Zanker "Insights from Applying Sequential Pattern Mining To E-Commerce Click Stream Data" in 2010 IEEE International Conference on Data Mining Workshops
[7] Ketul B. Patel, Jignesh A. Chauhan, Jigar D. Patel "Web Mining in E-Commerce: Pattern Discovery, Issues and Applications" in International Journal of P2P Network Trends and Technology- Volume1Issue3- 2011.

