

An Advanced Recommendation System for E-commerce Users

Ramkumari Tyagi

Dept. of Computer Science Engg.
Shri Ram College of Engg & Management
Banmore, India
ramkumarityagigwl@gmail.com

Anand jawdekar (*Asst.Professor*)

Dept. of Computer Science Engg.
Shri Ram College of Engg & Management
Banmore, India
anand.cs2007@gmail.com

Abstract—With over 200 million customers online to the world-wide web, electronic commerce now account owed for a developing percentage of world trade. The internet business model, which provides accessibility directly to suppliers and new stages of efficiency with much less assets and minimum overheads, is being eagerly investigated by means of important organizations. The development in electronic commerce has result in a substantially multiplied demand for understanding and communications technology (ICT). This paper analyzes the influence of e-commerce on markets where situated businesses face competitors from web-established entrants with targeted choices. In specific, we study the retail sector where the development of online procedures and the provision of more than a few resources. We have developed a recommender procedure to analyze the have an effect on of competitors on the objects. The discounted items are supplied to the users in an effort to expand the client's curiosity.

Keywords—Data Mining; Web Mining; Web Usage Mining; Recommendation System; E-commerce.

I. INTRODUCTION

In these days Internet becomes the powerful and vast source of information. Internet act as a container, contain huge amount of information. It is difficult for the web master of an company is to extract information from this containers and to match with the needs of user and provide effective or correct result. At the point when searching and skimming a web site, users are frequently overwhelmed by tremendous amount of information and are faced with the huge challenge of discovering the wanted information in the correct time [1].

For information extraction from the internet Web usage mining play an important role. Web usage mining is that the method to fetch data from the history of user's interaction from the web. It is one amongst the necessary applications of Data mining. Recommender system is one among the vital applications of Web usage Mining. The main Objective of the Recommender system is to present a system with the feature of Automatic Recommendation on the basis of user's navigation patterns on the web or we can say that a system capable for guiding users towards their desired choices and services[2].

When a user interacts with the web, their navigation histories save in an area like a container called as web log. Web logs

contain information of user's navigation pattern or user interaction information from the web. In this paper, we tend to propose a recommendation approach that recommends a listing of pages based mostly upon client's historic pattern (recorded within the web log). This approach brings the advance within the accuracy of the pages displayed to the client or users.

II. WEB USAGE MINING

Web is the complex system of interconnected elements and mining is procedure to eliminate knowledge or data, Means web mining is the extract data procedure from this interconnected system[3]. Mining of web is one amongst all the application of mining of data. Mining of data essentially deals with the organized form of data, while mining of web deals with the unorganized and partially organized form of data. Mining of web is divided into three categories i.e. web content mining, WSM and WUM [4]. There is 3 types of web mining.

A. Web content mining

Web content mining means that to mine the info from record set or weblog. Website mining is additionally called Text mining. In this technique content like text, graphs and picture, are parsed or scanned to find out the significance of the content to the requested query. Text mining is coordinated against particular information, gave by the client or user, in search engines.

B. Web structure mining

Web structure mining means to mine the information from the linkage structure of the Webpage then this linkage information is used to capture the list of interesting patterns. It is the Procedure using in graph theory to explore node and web site connection structure.

C. Maintaining the Integrity of the Specifications

Web usage mining means to mine the information from the web access behavior of the customers at the same time as they are skimming and investigating through the Web [5]. The basic point of comprehension the route inclination of the guests is to improve the nature of electronic trade administrations (e-business), to customize the Web entrances,

to personalize the Web portals or to improve the Web structure and Web server performance.

III. WEB MINING IN E-COMMERCE

In the severe competition in an E-commerce market, any information related to consumer behavior is awfully precious to merchant. A most important challenge of E-commerce is to understand customer's wants, love and value orientation as much as possible, in order to make sure competitiveness in e-commerce era. Web mining can be used to find obvious data which have potential value.

A. Personalized Service

When user browses Web sites, to meet each user's browsing interest, personalized service constantly adjusts to adapt to the user's browsing interests change, so that each user feel he/she is a Web site unique user. To achieve personalized service, it has to acquire and gather information on clients to grasp spending habits, interests, psychology, etc. of the consumer. To find out customer spending behavior patterns, the classical marketing method is very difficult, but it can be done applying Web mining method.

B. Improve the e-commerce website design

Use Attractiveness of the site depends on its sensible content, design and structure of organizational. Web mining can endow with user behavior details, giving designers decision making basis to advance the design of the site.



Fig. 1. Top Motivator factors for shopping online

C. Advertising effectiveness evaluation

The effectiveness of advertising can be evaluated by analyzing the number of consumer behavior patterns by Web mining. Several combinations of commodity schemes are designed to help the goods in the sales process.

Top Motivators for Shopping Online Times of India (February 12, 2013) has published that top motivators for shopping online which include cash back guarantee, cash on delivery, fast delivery, substantial discounts compared to

retail, and access to branded products, while barriers include inability to touch and try products before purchase, fear of faulty products, apprehension of posting personal and financial details online and inability to bargain.

IV. LITERATURE REVIEW

As of late, many Web Usage Mining systems have been proposed to forecast navigation patterns of users and their inclination. In this section we review some of the research work related to web recommendation.

WANG Xiao-Gang and LI Yue[6] work on Web personalization, one of the applications of web usage mining for detecting user Access patterns. For this they used Sequential pattern mining algorithm for finding frequently access patterns. Frequently access patterns are then stored in a Tree structure each node of the tree are represented by one of the item discovered by sequential pattern mining algorithm. And in last step this tree is used for matching and generating recommendation rules for the web personalization.

Mohit Rajput, Rohit Agrawal, Dilip kumar Sharma[7] proposed a comprehensive study about the different methods of the web personalization in various areas with their merits and demerits. And they presented a detailed comparison between these various areas of web personalization. They also focus on different issues and challenges of these methods like Security and privacy issue, Fair Dealing and Integrity, cost etc.

Ms.Dipa Dixit Mr. Jayant Gadge[8] Present a approach for recommendation in which they capture the list of user navigation behavior and provide recommendations in the form of list. They propose two tier architecture. First tier is concern with the offline component of web usage mining in which a knowledge base is created (with the analysis of web log data) which is accurate and noise free. And second tier is related with the online component of web usage mining , recommend page views to the user whenever he /she comes online for the next time.

Pragya Rajput, Joy Bhattacharjee, Roopali Soni[9] proposed work helps personalize the website according to users, country being its analytical parameter(" web personalization is a action that tailors the web experience to several or any individual client").IP addresses are provided to every internet user by the Internet Service Provider (ISP) which is unique. Every region has some specific ISP. For example, in India AIRTEL is the major ISP. Through IP Address we can identify the country of the user and thereafter we can modify the contents of web pages according to the country and prediction of its citizens' requirements based on their previous surfed history.

Khan and Mahapatra[10] remarked that methodology perform a vital role in improving the QoS provided through the business units. One of the methodology which actually brought information revolution in the society is IT and is rightly regarded as the third revolution wave after agricultural

and industrial revolution. The cutting edge for business today is e-Commerce.

Waghmare[11] defined that Ecommerce, generally known as ecommerce or eCommerce, consists of the selling and buying of services or products over e system such as internet and different computer network. Intent is the methodology for e-commerce as it offers easier ways to the access companies and individuals at most low cost in order to carry out currently business transactions. SEM is a form of web advertising that companies use to products promote and SERPs. SEM is focused on the search engine advertisements efficient use (a.k.a., sponsored results, sponsored links) that SERP appear. SEM which permit firms to endpoint consumers through placing ads on search engines has proven to be an effective audience acquisition approach. Unlike classical online advertising, advertisers pay only when users actually click on an ad when successfully implemented, SEM can create steady traffic stage and tremendous return on investment (ROI).

Vellingiri, J., T.Parthiban,S. Kaliraj, S. Satheeshkumar[12] present a novel approach for the user navigation pattern discovery and for the analysis of these patterns. For this they used 3 steps of web usage mining. First is Preprocessing for the removal of noisy and unclear data. Other is the pattern detection which is concerned with the behavior of user's. Behavior of user's is structured into a set of clusters by one of the clustering algorithms, which consists of "similar" data items based on the user behavior and navigation patterns for the use of pattern discovery. Last is related to the analysis of these patterns.

Neha Goel, C.K. jha [13] presents a method for analyzing behavior of user with the help of one log analyzer tool among the variety of available tools. They used Web log expert tool. Log analyzer tools are available for mapping the user behavior means to find out the navigation pattern, means in which part of the website the users are mostly concerned.

V. PROPOSED WORK

In the proposed system user interacts with the web portal. Data preprocessing and data cleaning tasks are performed to extract valuable information. The cleaned data is further used to discover patterns, hidden rules and provide top recommendation to all the users making use of the E-commerce.

A. Data Acquisition

In mining of Web data, Web log files on the Web server are the main source of data. Web log files contain the history of the visitor's browsing behavior. Web log files conclude the server log, client log and agent log.

B. Data Preprocessing

The actual data collected have certain features such as redundancy, ambiguity and incomplete. In order to eliminate the information more efficiently, pre-processing the data collected is essential. Preprocessing can give accurate, concise

data for mining of data. Data preprocessing, conclude cleaning of data, user identification, user session identification, access way supplement and transaction identification.

a) The main task of cleaning of data is to eliminate the Web log redundant information which is not associated with the valuable information, narrowing the data objects scope.

b) Determining the single user must be completely after cleaning of data. The user identification purpose is to classify the user's unique. User identification can be completed through cookie technology means, user registration methods and also heuristic rules.

c) User session identification should be complete on the user identification basis. The purpose is to divide all user's access knowledge into various separate session procedures. The simplest way is to use time-out estimation method, that is, when the time interval between page requests exceeds the provide value, namely, that the user has started a novel session.

d) Because the widespread use of the page caching methodology and the proxy servers, the access way recorded through the Web server access logs may not be the done access users path. Incomplete access log does not accurately reflect the user's access patterns, so it is necessary to add access way. Path supplement can be achieved applying the Web site topology to create the analysis of page.

e) The identification of transaction is based on user's session recognition, and its purpose is to divide or combine transactions according to the data mining tasks demand in order to create it appropriate for demand of analysis of data mining.

C. Recommendation System

A recommendation system is produced that works for both type of users i.e. registered and unregistered. Also we have made the recommendation based on the items that have been added in the list whether they are purchased, added in wishlist or not purchased at all. In the proposed technique, we are recommending items based on navigational details of the user. The whole process of the proposed algorithm can be explained with the help of a pseudo-code. The pseudo-code helps us understand the proposed process in a better way.

D. Quality Evaluation

TABLE I. RECOMMENDATION MATRIX

	Recommend Items by the System	Item not Recommended by the System
Expected Item	True Positive (TP)	False Negative (FN)
Not an Expected item	False Positive (FP)	True Negative (TN)

Based on the recommendation matrix we calculate recall, precision and accuracy as shown in equation 4 to 6.

Recall can be defined as a fraction of all relevant items that are recommended by the system.

$$Recall = \frac{True_Positive(TP)}{True_Positive(TP) + False_Negative(FN)}$$

Precision is a fraction of all the recommended products that are relevant.

$$Precision = \frac{True_Positive(TP)}{True_Positive(TP) + False_Negative(FN)}$$

The accuracy is ratio of true positive to sum of all the item recommended.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FN}$$

E. Clustering

In this paper we use the user browsing path clustering algorithm. The user browsing path clustering algorithm is a clustering method of user paths, it's a path clustering according to the path similarity matrix. In this paper we use it for user clustering, we found that users browsing paths clustering algorithm can be implemented on user clustering and has good effect when we use similarity matrix replace the path similarity matrix.

The browsing path clustering algorithm is a kind of clustering algorithm with high practicability, can deal with high degree of data and process a large number of user web users paths.

F. Proposed Work

Algorithm Pseudo-code:

1. Load the web log and price data.
2. Performing the data Cleaning .
3. User and session identification of the registered and unregistered users.
4. Finding unique products based on product_id.
5. Counting the products in all the sessions.
6. Creating p_list as W,P and N
Where, P – purchased products
W – products added in wishlist
N – products not purchased.
7. Creating recommendation list of the unique items based on the priorities given
 - a. Items of wishlist having frequency more than or equals to 3.
 - b. Items of purchaselist having frequency more than or equals to 3.
 - c. Items not purchased but watched and having frequency more than or equals to 3.
8. Now price recommendation for the users who added products in the wishlist giving 10% discount to them so as to increase a way of selling the products.
9. Creating new recommendation list after discount.
10. End

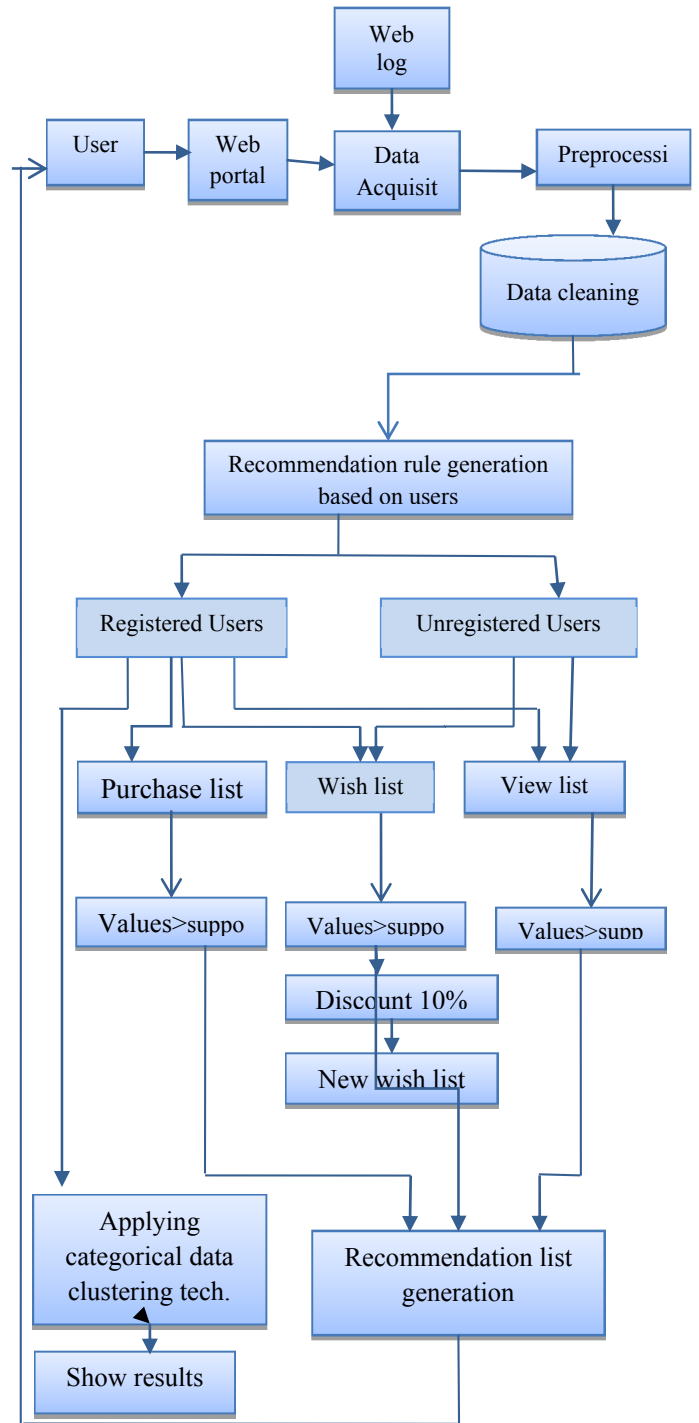
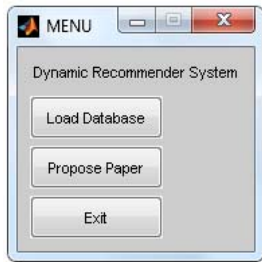


Fig. 2. Clustering technique Used in Proposed Work

VI. RESULT ANALYSIS

1. The menu for the recommender system is created. On clicking the first load database button, the database is loaded.



2. On clicking the propose paper button, a new recommendation list is generated. The items recommended are:-

Final recommendation list

2 4 5 6 7 8 11 12 13 14
 15 16 17 18 19 20 21 22 23 24
 25 26 27 29 30 31 33 35 36 37
 40 41 42 43 44 46 47 48 50 9
 38

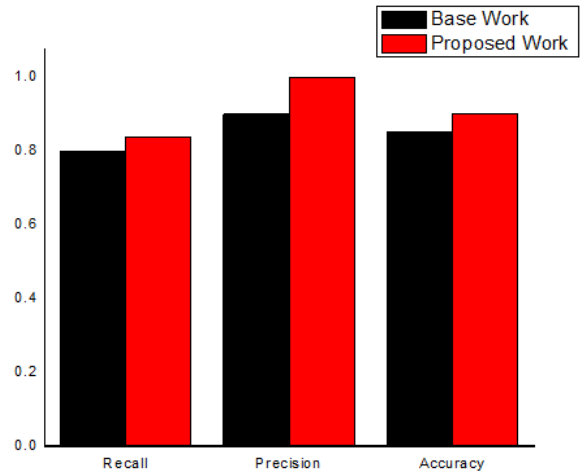
3. The discounted wishlist is shown:-

Product Id	Actual Price	Discounted Price
2	200.00	180.000
4	65.000	58.500
5	198.000	178.200
6	234.000	210.600
7	550.000	495.000
8	345.000	310.500
11	345.000	310.500
12	235.000	211.500
13	125.000	112.500
14	168.000	151.200
15	190.000	171.000
16	235.000	211.500
17	245.000	220.500
18	280.000	252.000
19	900.000	810.000
20	1000.000	900.000
21	239.000	215.100
22	342.000	307.800
23	788.000	709.200
24	987.000	888.300
25	567.000	510.300
26	876.000	788.400
27	456.000	410.400
29	230.000	207.000
30	905.000	814.500
31	987.000	888.300
33	654.000	588.600
35	432.000	388.800
36	321.000	288.900
37	200.000	180.000
40	670.000	603.000
41	879.000	791.100
42	546.000	491.400
43	657.000	591.300
44	876.000	788.400
46	123.000	110.700
47	234.000	210.600
48	432.000	388.800
50	300.000	270.000

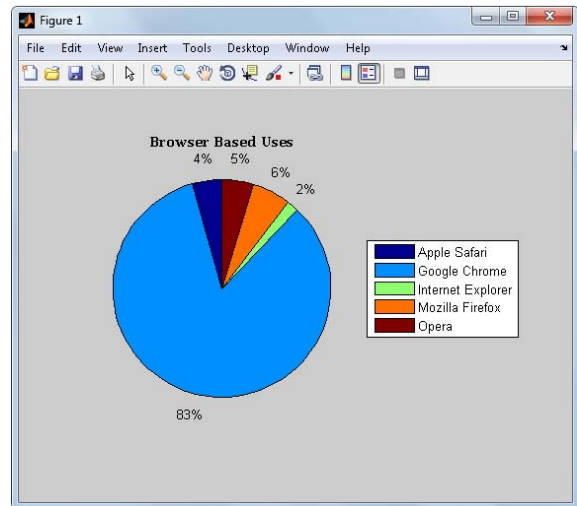
4. Based on the results, few factors are calculated on which the results depend. The factors are:-

- i. Recall = 0.838710
- ii. Precision = 1.000000
- iii. Accuracy = 0.900000

In graphical form, the results can be shown like this:-



5. The browsers used by the users are also shown. This shows that which browser is most popular amongst people. The results are shown with the help of pie-chart.



6. The frequency chart is also shown in results which describes that which time period is most frequently visited by the users. With the help of the graph, we can predict the timings in which the sale can be maximized by the websites. The graph is shown below:-

CONCLUSION

The recommender system created for the user's behavior is shown in this paper. The system generated is of very much importance as it gives user's a chance to buy things that they are interested in. The discounts and the factors that affect the user's ideas thus lead to a better system. This paper gives a new proposed system based on user's interest in the field of e-commerce. As this field is increasing in an exponential manner. The results show a good diversion.

REFERENCES

- [1] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. Second edition Morgan Kaufmann Publishers.
- [2] R.Colley,B.Mobasher and J.Srivastava "Data Preparation for Mining World Wide Web Browsing Patterns" *Knowledge and Information Systems*,pp.5-32,1999.
- [3] J.Srivastava,R.Cooley,M.Despande and P.Tan "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data" *SIGKDD Explorations*,pp.12-23,2000.
- [4] Hussain T., Asghar S., Masood N "Web log cleaning for mining of web usage patterns" *IEEE ICCRD*,2011,pp.490-494.
- [5] Mei Li and Cheng Feng, Overview of WEB Mining Technology and Its Application in E-commerce, *IEEE 2nd International Conference on Computer Engineering and Technology* ,Volume 7. 2010 pp V7-277-V7-280.
- [6] LI Yue and WANG Xiao-Gang "Web Mining Based on User Access Patterns for Web Personalization" *IEEE ISECS International Colloquium on CCCM 2009* , PP.194-197
- [7] Rohit Agrawal , Mohit Rajput Dilip kumar Sharma "A STUDY & COMPARATIVE ANALYSIS OF WEB PERSONALIZATION TECHNIQUES", *IEEE International Conference on Information and Communication Technologies (ICICT)* 2013,PP. 1293-1298
- [8] Mr. Jayant Gadge, Ms.Dipa Dixit "Automatic Recommendations for Online Users Using WUM", *International Journal of Managing Information Technology (IJMIT)* Vol.2, No.3, August 2010 ,PP. 33-42
- [9] Roopali Soni , Pragma Rajput, Joy Bhattacharjee, "A Proposed Framework to Implicit Measures of User Interests through Country and Predicting Users Future Requests in World wide web", *IJSCE(International Journal of Soft Computing and Engineering)* Volume-3, Issue-1, PP. 340-343.
- [10] M. S. Khan and S. S. Mahapatra, "Service quality evaluation in internet banking: an empirical study in India", *Int. J. Indian Culture and Business Management*, vol. 2, no. 1, (2009), pp. 30-46.
- [11] G. T. Waghmare, "E-commerce; A Business Review and Future Prospects in Indian Business", *Internet Marketing in India: Indian Streams Research Journal*, vol. 2, no. IV, (2012), pp. 1-4.
- [12] Vellingiri, J., S. Kaliraj, S. Satheeshkumar and T.Parthiban, "A Novel Approach For User Navigation Pattern Discovery And Analysis For Web Usage Mining " *Journal of Computer Science* 11 PP: 372-382, 2015.
- [13] C.K. Jha, Neha Goel "Analyzing Users Behavior from Web Access Logs using Automated Log Analyzer Tool", *International Journal of Computer Applications* Volume 62– No.2, January 2013, PP. (0975 – 8887).