Personalized Recommendation System for Onscreen Clients Using Web Usage Mining

R. Vigneshwaran[1], S. Sobitha Ahila[2]

[1]Department of Computer Science and Engineering, Easwari Engineering College Anna University, Chennai, India-600021
[2]Easwari Engineering College Anna University, Chennai, India-600089
Email address: [1]smart3vicky@gmail.com

Abstract— E-commerce organizations are developing exponentially with time in terms of both business and data. Many organizations rely on these websites to attract new clients and maintain the existing ones. Client interacts with web portal and click stream data is maintained in raw log files can be used that records client access patterns. Multiple preprocessing and data cleaning task are performed to extract valuable information from raw log files and transform it in structured form. Cleaned log data is further used to discover patterns, hidden rules and provide product recommendation to all users of E-commerce site. This paper focuses on providing personalized recommendation to all the visitors of the website irrespective of been registered or unregistered users. Results prove that the proposed method can be improved for other non e-commerce site and social network via using natural language processing and web link relationship.

Keywords— E-Commerce; lexical patterns; personalized recommendation; weblog files; web usage mining.

I. OVERVIEW

E Commerce, the type of personalization described is applicable to any Web browsing activity. Web personalization can be described, as any action that makes the Web experience of a user personalized to the user's taste. Customers are overloaded with multiple choices for a specific product, which results in a confused and lost state. It has become trivial for the webmasters to evaluate whether the products and services provided are catering to the needs of the customers or not. Therefore it is important to devise new marketing strategies such as one-to-one marketing and customer relationship management. One effective solution to handle this issue is to provide personalized recommendation to individual user i.e. providing the customer with the type of product recommendation list he or she is interested in.

A promising solution to overcome this issue is recommendation system which provides and guides the customer with type of product he or she is interested in buying/purchasing. Till date a wide variety of recommendation system have been devised and implemented. Recommendation systems can be broadly classified into two: Content-based and Collaborative filtering systems. Content based system takes into consideration the product attributes to generate recommendation. Collaborative filtering system makes use of customer - product interaction and ignores the other facts to provide recommendations

II. PROBLEM STATEMENT

As number of consumers and products increases in number it slows down neighbor selection per second. Another issue that can be pinpointed is most of recommendation systems make use of binary transaction (click stream) data. I.e., whether a specific item is purchased or not. However many a times they are unable to exploit intrinsic characteristics of these data that can be used to provide better recommendation.

Another drawback of conventional recommendation system is retargeting. I.e., providing the customer with the same product that is already purchased. Studies in indicate that web usage mining can act as an effective solution to overcome the limitation of traditional recommendation systems.

III. OBJECTIVE

To generate recommended list for E-Commerce users by accessing user preferences on search patterns are stored in web log files. Data cleaning tasks are performed and transformed it in structured form. Cleaned log data is further used to discover patterns, hidden rules and provide product recommendation to all users of E-commerce site. Further the data in social network sites like (Face book) can be used their comments on E-commerce sites.

IV. WEB USAGE MINING

Web usage mining focuses on predicting users' preferences and behavior by analyzing web logs with help of traditional data mining techniques. Customer's click stream data can act as a very rich source of information. Click stream indicates user's path through a website. Click stream data is captured and maintained in web log files. Strategic use of navigational data can be very helpful in providing effective recommendation. Good quality recommendation systems will not only help in satisfying customers preferences for a product but also in improving sales and attracting new consumers. Indigent quality of recommendation, results in two types of peculiar errors, false negatives: these are the items not recommended even though the customer likes it. False positive these are the items recommended even though the customer dislikes it.

In an E-commerce domain the most important errors that need to be handled and circumvented are false positives errors, which can result in unsatisfied customers and minimize their possibility to revisit the site once again.

V. MODULAR DESIGN

A. Data Acquisition

In this phase the entire navigational data which includes all the web pages visited is collected and stored. The proposed work makes use of common log file format to maintain the data, important attributes namely IP address, timestamp, status code, URL, method (GET and POST), user agent and Referrer URL are recorded and used for further analysis. This data obtained is highly unstructured and inconsistent in nature. And therefore has to be preprocessed for further analysis.

B. Data Preprocessing

Good quality input data needs to be served for better analysis. In this phase inconsistent, redundant data is eliminated using following steps as depicted in field separation stage focuses on distinguishing one attribute from another by making use of separator character such as space. In data cleaning stage we filter out outliers data. We check for URL suffixes. Log entries having filename suffixes such as gif, jpeg, tiff, jpg are discarded.

All records having failed http status code are removed i.e. status code greater than 200 and less than 299 are eliminated. In user differentiation phase we assign unique user ID, to each IP address and registered users to differentiate one customer from other. Finally we construct session in session identification phase. In session clustering phase we group together session belonging to unique user. Session provide us with complete set of activities done by the user in specific time period. Finally in data formatting stage we place the data in tabular form.
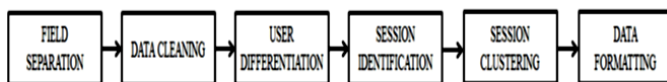

Fig. Data preprocessing steps.

C. Recommendation System

Two different recommendation systems have been proposed. Product based recommendation technique works specifically for unregistered users. Dependent on users IP address session are constructed. Based on session details recommendation list is generated. Combination of session is taken to generate final recommendation list.

User based technique is used for registered users. In this technique based on user's navigational details appropriate recommendations are provided. A detailed description, for all the two approaches customer inclination analysis is done dependent on three parameters: clicked through, basket placement and purchased product.

D. Recommendation Generation Method

In this technique based on user's category (whether registered or unregistered) corresponding product recommendation is made. For unregistered user IP address is the deciding factor and for registered user unique user id issued. In this technique we fetch session based on most recent time stamp (ts).

User Interest Measure (μ): It implies whether a user is actually interested in a specific product. μ is calculated using page stay time.

$$\text{User Interest measure}(\mu)\begin{cases}1, \text{if } tsp \geq 0.05 \\ 0, \text{if } tsp < 0.05\end{cases}$$

Based on above parameters we construct initial recommendation list and then fetch related products based on category and manufacturing details to produce final list of products. We also calculate similarity measure to determine which lexical string is to be retrieved in case if multiple session have same product.

The working of entire technique is given in below fig. 4. In this approach dependent on the type of customer (whether registered or unregistered) corresponding technique is involved. Depending upon the current traversal pattern recommendation list is generated. Different parameters namely page stay time, count, wish list buffer content and similarity measure corresponding product recommendation string is generated. Dynamic recommendation list is generated as per changing behavior of the user in the current session.

VI. RELATED WORK

Cooley and et.al defined the term web usage mining for the first time and aims on predicting user's preferences and behavior. The entire web usage mining process is divided into three phases namely data preparation, pattern discovery and pattern analysis are discussed. Most of the data needed for web log analysis resides on web servers, proxy servers, enterprise logs, web clients etc. Many a times these data is ambiguous in nature and needs to be cleaned for further analysis. In order to yield better recommendation results good quality of data should be served as an input. Evaluation conducted in show that 80% of time spends in data mining is consumed in log data preprocessing.

Web log data has peculiar characteristics and therefore detailed description about various steps followed is given in. The pattern discovery tasks include the discovery of association rules, sequential patterns, user classifications etc. In explains Nave Bayesian techniques for dynamic mining of user's interest navigation pattern.

However this technique is time consuming. In explain agglomerative hierarchical clustering method that makes use of Euclid and distance to calculate similarity measure and clusters the users having similar browsing feature. In makes use of association rule mining, product taxonomy and web usage mining to provide personalized recommendation. It tries to overcome certain limitations of collaborative filtering. We propose a system that will try to overcome limitations of traditional recommendation systems.

VII. LITERATURE SURVEY

A. Application of Web Usage Mining and Product Taxonomy to Collaborative Recommendations in E-Commerce

The rapid growth of e-commerce has caused product overload where customers on the Web are no longer able to effectively choose the products they are exposed to. To

overcome the product overload of online shoppers, a variety of recommendation methods have been developed. Collaborative filtering (CF) is the most successful recommendation method, but its widespread use has exposed some well-known limitations, such as scarcity and scalability, which can lead to poor recommendations.

B. Mining Web Browsing Patterns for E-Commerce, Computers in Industry

Web user clustering, Web page clustering, and frequent access path recognition are important issues in E-commerce. They can be used for the purposes of marketing strategies and product offerings, mass customization and personalization, and Web site adaptation. In this paper, we view the topology of a Web site as a directed graph, and use a user's access information on all URLs of a Web site as features to characterize the user and use all users' access information on a URL as features to characterize the URL. The user clusters and Web page clusters are discovered by both vector analysis and fuzzy set theory based methods.

C. A Personalized Recommender System Based on Web Usage Mining and Decision Tree Induction

A personalized product recommendation is an enabling mechanism to overcome information overload occurred when shopping in an Internet marketplace. Collaborative filtering has been known to be one of the most successful recommendation methods, but its application to e-commerce has exposed well-known limitations such as scarcity and scalability, which would lead to poor recommendations. This paper suggests a personalized recommendation methodology by which we are able to get further effectiveness and quality of recommendations when applied to an Internet shopping mall. The suggested methodology is based on a variety of data mining techniques such as web usage mining, decision tree induction, association rule mining and the product taxonomy.

VIII. CONCLUSION

Work we focus on providing good quality product recommendations to all the users especially unregistered ones

of E-commerce site. The beauty of the proposed system is it dynamically provides recommendation as per changing users' behavior and traversal patterns by making use of web usage mining and constructing patterns from the historical data. The proposed recommendation system minimizes the false positive errors that occur frequently in traditional recommendation system. Also issue of binary ratings and cache memory are handled by the system thereby providing good quality recommendations.

Results prove that accuracy of approximately 80 to 85 percent is achieved for registered user and 65 to 70 percent for unregistered user in rational recommendation technique, which is better than product based technique and almost equivalent to user based approach. The recommendation system has the potential to attract new customers and retain existing ones. This technique can help the E-commerce organization have competitive edge in the market and can be helpful in forecasting demands and sales for a specific product. It would be interesting to evaluate the proposed technique with different conventional recommendation approaches and measure its accuracy.

REFERENCES

[1] Y. H. Cho and J. K. Kim, "Application of web usage mining and product taxonomy to collaborative recommendations in e-commerce," *Expert Systems with Applications*, vol. 26, issue 2, pp. 233-246, 2004.

[2] Q. Song and M. Shepperd, "Mining web browsing patterns for e-commerce," *Journal Computers in Industry*, vol. 57, issue 7, pp. 622-630, 2006.

[3] Y. H. Cho, J. K. Kim, and S. H. Kim, "A personalized recommender system based on web usage mining and decision tree induction," *Expert Systems with Applications*, vol. 23, issue 3, pp. 329-342, 2002.

[4] P. Lopes and B. Roy, "Recommendation system using web usage mining for users of E-commerce site," *International Journal of Engineering Research & Technology*, vol. 3, issue 7, 2014.

[5] C. R. Varnagar, N. N. Madhak, T. M. Kodinariya, and J. N. Rathod, "Web usage mining: A review on process, methods and techniques," *International Conference on Information Communication and Embedded Systems (ICICES)*, pp. 40-46, 2013.

[6] J. Lee, M. Sun, and G. Lebanon, "PREA: Personalized recommendation algorithm toolkit," *The Journal of Machine Learning Research*, vol. 13, pp. 2699-2703, 2012.