ISSN:0975-9646

Survey on Online Recommendation Using Web Usage Mining

Neeraj Iyer, Alex Dcunha, Akshay Desai, Kavita Jain

Department Of Computer Engineering
Xavier Institute Of Engineering,
Mumbai, India

***Abstract*— With the rising growth of Web users, Web-based organizations are keen to analyse the on-line browsing behaviour of the users in their web site and learn their interest instantly in a session. The analysis of the user's current interest based on the navigational behaviour may help the organizations to guide the users in their browsing activity and obtain relevant information in a shorter span of time. Since a user has a specific goal when searching for information, personalized search may provide the results that accurately satisfy user's specific goal and intent for the search. Personalization of web search is to retrieve information according to user's interests which may be inferred from user's action, browsed documents or past query history etc. The Paper focuses on recommender systems based on the user's navigational patterns and provides suitable recommendations to cater to the current needs of the user. This paper contains the concept of Web Usage Mining which is the discovery and analysis of user access patterns through mining of log files and associated data from a particular website. The data most accessed by user will be stored in log files.**

***Keywords*— Recommender Systems, Web Usage Mining, Collaborative Filtering , Content Based Filtering.**

I. INTRODUCTION

The wealth of information on World Wide Web has lured users to seek and retrieve information from the Internet .However, this plethora often creates its own set of problems with users being unable to retrieve useful and relevant information. One of the potential approaches to deal with this problem is to analyse navigational patterns of users interacting with one or more web sites. Analysis of the user's browsing patterns can help organizations to provide personalized recommendations of web pages according to the current interests of the user. Usage-based Personalized Recommendation has aroused interest in researchers as it has greatly contributed to solving this problem. Recommendation systems lessen information overload by suggesting pages that meet the user's requirement. Of late .Web usage mining has gained much attention as it is found to fulfill the needs of web personalization [5]. Web Usage Mining deals with the discovery and analysis of "interesting" patterns from click-stream and associated data collected during the interactions with Web server on one or more Web sites [8]-[9]. The pattern profiles are discovered by applying common data mining techniques to the pre-processed data [1] and provide input to the recommendation engine that recommends appropriate pages based on the intelligence gained from the usage profiles. The motivation of our study is how to shorten the difference between shopping on line and shopping in the reality using almost mature technologies. We believed that it would be more real and attractive that if the customers could get personalized help on time when they shopping on the websites, like playing an on-line game. Therefore, our thought is to provide service by collecting apparel dressing knowledge and apparel enterprise core knowledge database such as apparel style, material and detail database, combining with the customer personal information, which can compose an apparel recommendation intelligent system [2]-[4].

II. RELATED WORK

Online Recommendation exists and is being used by many shopping websites. Many algorithms, permutations and combinations, Associations will be used to create online recommendation. There are many different ways through which a system can be created.
Some algorithms used in general are:
1. Collaborative filtering:
   Item to Item.
   User to User.
2. Content Based Filtering:

A. Collaborative Filtering Based Recommender System :

Recommender systems [1]-[2] apply data analysis techniques to the problem of helping users find the items they would like to purchase at E-Commerce sites by producing a predicted likeliness score or a list of top-N recommended items for a given user. Item recommendations can be made using different methods. Recommendations can be based on demographics of the users, overall top selling items, or past buying habit of users as a predictor of future items. Collaborative Filtering (CF) is the most successful recommendation technique to date. The basic idea of CF-based algorithms is to provide item recommendations or predictions based on the opinions of other like-minded users. The opinions of users can be obtained explicitly from the users or by using some implicit measures.

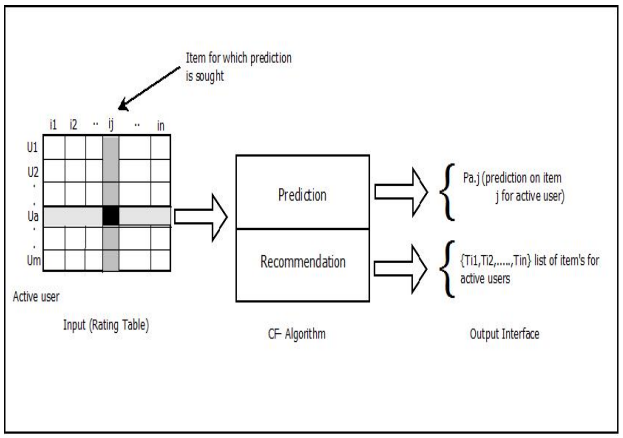

Fig.1 .Collaborative Filtering Process

Item-Based Collaborative Filtering Algorithm:

In this section we study a class of item-based recommendation algorithms for producing predictions to users. Unlike the user-based collaborative filtering algorithm, the item-based approach looks into the set of items the target user has rated and computes how similar they are to the target item i and then selects k most similar items {i1, i2 ,….ik}. At the same time their corresponding similarities {S_1, S2,….S_{ik}} are also computed. Once the most similar items are found, the prediction is then computed by taking a weighted average of the target user's ratings on these similar items. We describe these two aspects, namely, the similarity computation and the prediction generation in details here [3].

Fig.1.Item-based Collaborative Filtering Process

B. Content Based Filtering :

A content-based filtering system selects items based on the correlation between the content of the items and the user's preferences as opposed to a collaborative filtering system that chooses items based on the correlation between people with similar preferences. PRES is a content-based filtering system. It makes recommendations by comparing a user profile with the content of each document in the collection. The content of a document can be represented with a set of terms. Terms are extracted from documents by running through a number of parsing steps. First all HTML tags and stop words (words that occur very often and cannot be used as discriminators) are removed. The remaining words are reduced to their stem by removing prefixes and suffixes [Porter 1980]. For instance the words "computer", "computers" and "computing" could all be reduced to "compute" [6]-[8]. The user profile is represented with the same terms and built up by analysing the content of documents that the user found interesting. Which documents the user found interesting can be determined by using either explicit or implicit feedback. Explicit feedback requires the user to evaluate examined documents on a scale. In implicit feedback the user's interests are inferred by observing the user's actions, which is more convenient for the user but more difficult to implement [3]-[4].

III. PROPOSED SYSTEM

Online recommendation is already in use. But Different ways and algorithms are used for building it. We have tried to make a comparative study of the techniques which were used in the previous systems. We have combined algorithms and made it efficient and other measures of efficiency such as high speed , minimum memory usage [7]-[9].

Our Proposed System works in 5 phases:

Phase 1: Sort Phase

The database (D) is sorted, with customer-id as the major key and transaction-time as the minor key. This step implicitly converts the original transaction database into a database of customer
sequences.

Phase 2: Large Item-set Phase

In this phase we find the set of all L-itemsets. We are also simultaneously finding the set of all large 1-sequences, since this set is just $\{(l)|\ l \in L\}$
.

Phase 3: Transformation Phase

In a transformed customer sequence, each transaction is replaced by the set of all L-itemsets contained in that transaction. If a transaction does not contain any l-itemset, it is not retained in the transformed sequence. If a customer sequence does not contain any l-itemset, this sequence is dropped from the transformed database. However, it still contributes to the count of total number of customers.

Phase 4: Sequence Phase

Use the set of litemsets to find the desired sequences. Algorithms for this phase are:

1. Apriori_All

The AprioriAll algorithm uses each pass to find large sequences from the previous pass to generate the candidate sequences and then measure their support by making a pass over the database.
At the end of the pass, the support of the candidates is used to determine the large sequences.

In the first pass, the output of the L itemset phase is used to initialize the set of large 1-sequences.
The candidates are stored in hash-tree to quickly find all candidates contained in a customer sequence.

2. Apriori_Some

1. Two of the proposed algorithms, AprioriSome and Apriori-All, have comparable performance, albeit AprioriSome performs a little better when the minimum number of customers that must support a sequential pattern is low.
2. The major advantage of AprioriSome over Apriori-All is that it avoids counting many non-maximal sequences.

ALGORITHM:
Input: Database D of transactions; minimum support threshold; min_sup
Output: L, frequent itemset in D

Step 1: (Join Step)
To find L_k, a set of candidate k-itemsets is generated by joining L_{k-1} with itself. This set of candidate is denoted by C_k.

Step 2: (Prune step)
$C_k$is a superset of L_k, that is, its members may or may not be frequent, but all of the frequent k-itemset are included C_k.

C_k: Candidate itemset of size k
L_k: Frequent itemset of size k
L_1 = {frequent items};

for(k=1;L_k != Ø ; k++) do begin

C_{k+1} = Candidates generated from L_k;

For each transaction t in database do
increment the count of all candidates in C_{k+1} that are contained in t.

L_{k+1} = candidates in C_{k+1} with min_support
End
return$U_k L_k$;

Phase 5: Maximal Phase(Output Phase)

Find the maximal sequences among the set of large sequences. In some algorithms, this phase is combined with the sequence phase to reduce the time wasted in counting non-maximal sequences.

IV. CONCLUSION

This paper describes the Online Recommendation System for a shopping cart, an application that enables the user to search for and receive recommendations to find the product based on the previously searched products. The search result comprises a large number of products. Hence the recommendation system cuts down the burden on the user by recommending the products that best fits his preferences by placing them at the top. This recommendation is retrieved from the log files [1]-[3]-[9]. Explicit feedback leads to an increase in the user's frustration level. The recommendation process uses implicit feedback. The feedback is obtained by making some observations on the user's purchase history. The online recommendation system for shopping cart makes successful recommendations by using the technique of assigning feature weights and user specific preference based recommendations.

REFERENCES

[1] Rakesh Agrawal , Ramakrishnan Srikant , "Mining Sequential Pattern"IBM Almaden Research Center,650 Harry Road, San Jose, CA 95120
[2] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, " Analysis of Recommendation Algorithms for Ecommerce " GroupLens Research Group / Army HPC Research Centre, Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455
[3] Rahul Neve, K.P Adhiya, " Comparative study of Web mining algorithms for web page prediction in recommendation system " International Journal of Advanced Research in Computer and Communication Engineering, Vol 2,Issue 1, January 2013
[4] C.P.Sumathi, R.Padmaja Valli,T.Santhanam,"Automatic Recommendation of web pages in Web Usage Mining", (IJCSE)International Journal of Computer Science and Engineering,Vol 02,No 09,2010,3046-3052
[5] Zhu Xinjuan,Huang junfang,Shi Meihang," An Intelligent on-line recommendation system in B2c apparel e-commerce ", Department of Computer Science, Xi'an Polytechnic, University, International Conference on E-Business and E-government,2010
[6] Guandong Xu, " Web Mining Techniques for Recommendation and Personalization " , Dissertation submitted to the School of Computer Science and Mathematics,Victoria University,Australia,March 2008
[7]http://scholarworks.sjsu.edu/cgi/viewcontent.cgi?article=1111&context=etd_projects
[8] Dipa Dixit,Jayant Gadge," Automatic Recommendation for online users using Web Usage Mining", International Journal of Management Information Technology(IJMIT),Vol 02,No 03, August 2010
[9] http://en.wikipedia.org/wiki/Web_mining
[10] Jaideep Srivastava,Robert Cooley,Mukund Deshpande, " Web Usage Mining- Discovery and Application of Usage Pattern ", Dept of Computer Science and Engineering,University of Minnesota,MN55455