International Journal of
Soft Computing And Software Engineering (JSCSE)
e-ISSN: 2251-7545
Vol.3,No.11, 2013
Published online: Nov 25, 2013
DOI: 10.7321/jscse.v3.n11.3

Estimating Collaborative Filtering Technique for Web Personalization

Faria Kanwal[1], Komal Bashir[2], Ayesha Haider Ali[3]
[1,2,3]LCWU, Lahore, Pakistan.
Email : [1]faria.kanwal@gmail.com, [2] ko_junaid@yahoo.com, [3]ayesha.iqbal@gmail.com

Abstract. **World Wide Web is rapidly growing in size and usability. Web personalization is the hub of many e-commerce websites and web portals. It is the process of getting and storing information about site users. Moreover, web personalization helps in analyzing the information and then delivering the required data to the user. A number of techniques are proposed to acquire the better results. Collaborative filtering technique among all technique is better in a way that it provides recommendation to new user, based on the preference of similar users. All personalization techniques suffer from sparsity and black box problem. Association retrieval technique helps in reducing the sparsity and black box problems causing the limited recommendations and results to the new users of the website. To overcome the problem in Collaborative filtering technique gain the better results through this technique by recommending the transitive dependencies of the user to the items selected. Recommender system problems are identified and estimated.**

Keywords*: World Wide Web, Web Personalization, Collaborative Filtering Technique, User Preferences*

* Corresponding Author:
Ayesha Haider Ali,
Faculty of Engineering and Technology,
Lahore College for Women University (LCWU), Lahore, Pakistan.
Email: ayesha.iqbal@gmail.com Tel:0092-307-4557949

1. Introduction

World Wide Web (WWW) is communication channel for access, retrieval and sharing of information using internet. Web implements new techniques of design and development for online services. Mainly, web is vast and complex network and sometime users fail to get the required information and receive imprecise results. Web user, categorize & analyze the collected data & perform the action on the collected data [3].

Compulsion for guessing or predicting user needs in turn improves the usability and user protection of a web site. Web personalization is an approach to determine or imagine users' needs, interests and behavior and provide suitable and relevant information according to their interests. Web personalization brings relevant information for the users [8, 9]. It allows users to see what information they want and when they want. Personalization technique can work on predefined rules or existing user preferences. The techniques recommend items on the basis of some parameters. Sometimes, these techniques provide the irrelevant and vague suggestions. These techniques suffer from sparsity and black box problems causing the extraction of irrelevant information of the users. A new concept of association of transitive dependency is suggested along with the direct relation of similar users in Collaborative filtering technique (CFT) that can maximize the relevant and useful recommendations to the users.

International Journal of
Soft Computing And Software Engineering (JSCSE)
e-ISSN: 2251-7545
Vol.3,No.11, 2013
Published online: Nov 25, 2013
DOI: 10.7321/jscse.v3.n11.3

Section II consists of related work and the literature is summarized in form of comparison tables of web personalization techniques and comparison of different website and their personalization techniques, while the Section III describes the proposed solution of the defined problems.

2. Related Work

There are four techniques used for web personalization, Rule Based Filtering Techniques, Content Based Filtering Technique, Web usage mining and CFT.

Rule based technique uses some predefined rules. Web system keeps each user history and suggests the item on the basis of user's previous interest. Web usage mining is used to analyze the usage data about a particular website. Usage data means the pattern of usage of web pages such as IP address, page reference and the date and time of access [3]. Content based technique has relationship between the content of the item and user preference [2, 11]. It provides recommendation to the user on the basis of specific user interest and behavior. Among all these, collaborative filtering technique is most efficient in terms of analyzing huge amount of users and recommends contents to new user based on the similarity of new user with the old one, who likes similar item in the past [1,4,6, 10]. Along with the efficiency of collaborative filtering technique it has some major issues like sparsity and black boxes. Sparsity further addresses two problems. The cold start in which systems are unable to give useful recommendation to new user because of lack of ratings of previous items & first rater in which systems are unable to recommend new items to user until they will get maximum ratings from users. CFT is used by e-commerce websites and web portals. As e-commerce demand is increasing day by day, there is need to enhance the technique to offer the maximum benefit to the visitors (users). CFT finds relationship between new user and other users on the basis of collected data to establish similarity between users and provide recommendations [1]. Collaborative filtering technique is advance and improved form of content base filtering technique.

A brief comparison of various web personalization techniques is shown in Table 1. Table 1 depicts that CFT matches the search items on the basis of similar users' preferences. Moreover, performance and consistency of CFT is better than all other techniques.

A comparison of different websites is summarized in Table 2 that exhibits that CFT is used by some commonly used websites like amazon.com [5], google.com, igoogle.com, netflix.com, digg.com, photoree.com, movielens.com, barnes and nobles.com, CDNow.com, Launch.com and all of these are recommender systems [12]. Recommender systems retrieve information on the bases of user's preferences and interests. Most of them not provide best recommendations because they are unable to handle sparsity. Most of them provide trust worthy explanations behind the recommendations but some of them not. Association retrieval technique generally uses to handle sparsity, it identifies transitive association between users and items and provides recommendations but it reduces sparsity to minimum extent [7]. A new concept is suggested with the association retrieval technique that can maximum reduces sparsity and can provide effective and useful recommendations.

3. Need of Work

In [13] author has mentioned the objective of recommender systems that is, by using the information about users and user pro-files, to calculate the utility or relevance of a particular item, thus providing personalized recommendations about a user or the items. Recommender systems have verified to be useful in domain such as e-commerce, and they have a promising future in many other perspectives like Web search engines and digital TV program recommenders.

The recommender system predicts the rating to the item according to the user and is used to recommend a list of items to the user. In this case, the system often finds good items, but the recommender systems have not yet attracted much interest among researchers.

The technique of collaborative filtering that recommends items based on the opinions of other users is very popular, especially in e-commerce. In recent years, numerous enhancements are brought into account to overcome the sparsity and black box problems.

4. Proposed Work

In recommender systems, matrix is maintained which is composed of users and items, as shown below. In this matrix users are denoted by rows and items are denoted by columns. 0, 1 are Boolean

International Journal of
Soft Computing And Software Engineering (JSCSE)
e-ISSN: 2251-7545
Vol.3,No.11, 2013
Published online: Nov 25, 2013
DOI: 10.7321/jscse.v3n11.3

numbers. Boolean number 1 shows that rating/purchasing has taken place and 0 show no rating/purchasing has taken place. The matrix is called user-item interaction matrix. Large amount of transactions record in the matrix, the matrix become sparse when there exist few items whose value is 1. Due to sparsity it is extremely possible that similarity between two users become 0 which means no similarity exists between users.

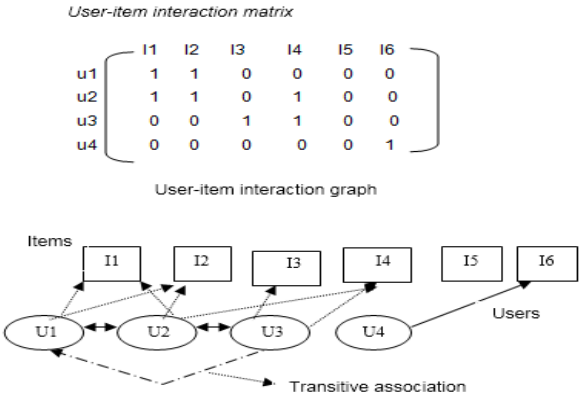

Figure 1: Matrix showing user-item interactions.

Association retrieval technique is generally used to handle sparsity problem. This technique discovers transitive association between user and items. Transitive association will be explained with the help of graph as shown in above figure; the graph is based on user-item interaction matrix. If users u1 rate item I1 and I2, and u2 rate item I1, I2 and I4, u3 rate item I3 and I4, all these items belong to same category. Suppose user u1 is new user and want to view recommendation, so standard collaborative filtering technique will associate u1 with u2, u2 with u3 but not u1 with u3, but association retrieval technique identify the transitive association of u1 with u3.This technique find association between u1 and u3.This technique not completely recommend those items that are rated by minimum users, as shown in matrix, item I3 which is rated by only user u3 when transitive association identified between users u1 and u3 item I3 recommended to u1 but item I6 which is rated by only user u4, this item will not recommended to user u1 because association retrieval technique unable to identify association between users u4 and u1.Association retrieval technique not recommend those items which is not rated by any user for example item I5 in the above matrix.

Standard CFT makes recommendations on the bases of user's direct similarity. Suppose u1 is new user and want recommendation. Standard CFT associate u1 with u2, because they rate common items I1 and I2 this means their exist similarity between them, so it will recommend different items of u2 to u1 which is I4. Standard CFT find association between user u1 and u2 by determining the presence and length of paths of u1 linking to the different item of u2. So there are two paths of length three that are linking u1 with I4.

Now how similarity found between u1 and u3.U1 rate I1 and I2 but u3 rate I3 and I4 there rated items are totally different, there exist an indirect similarity that standard CFT cannot identify .When there is no direct similarity, I3 can't recommend to u1. By using association retrieval technique and including transitive association of u1 and u3. Association retrieval technique discovers that there exist two paths of length five that link u1 with I3. There linking paths discovered this means similarity exist between u1 and u3, when similarity determined so I3 can be recommended to u1.

It is analyzed that Association retrieval technique not recommends I5 and I6 to u1 because this technique is not expert in establishing association of u1 with I5 and I6. I5 is that item which is not rated by any user and I6 is that which is rated by only user u4, u4 has neither direct nor indirect similarity with u1. For this, it is suggested that a new concept can be added in association retrieval technique with transitive association. Concept is that, association retrieval technique also matches new user rated items with all items.

International Journal of
Soft Computing And Software Engineering (JSCSE)
e-ISSN: 2251-7545
Vol.3,No.11, 2013
Published online: Nov 25, 2013
DOI: 10.7321/jscse.v3.n11.3

With this matching it is highly possible that the items which are less rated or not rated by the users like I5 and I6 is not shown to the new user. There should be two groups of item, one fetch from the association retrieval technique and second should consist of those items that is not ranked by many users. The items shown to the user on the basis of some parameters like the same geographic area, high ranked items and recommend the new items to the user. This concept can reduce the sparsity and can provide effective, useful and important recommendation to user.

5. Conclusion

Different techniques are used for web personalization, but CFT is most successful technique for web personalization. CFT also needs improvement because of sparsity and black box problem. Different Websites are analyzed, it concludes from the analyses of websites that sparsity exists in maximum websites and black boxes exist in minimum websites. Generally association retrieval technique use transitive association to reduce the sparsity but this technique reduces sparsity to minimum extent. So a new concept is suggested by which sparsity can be maximum reduced.

6. Future Work

This research has the importance to personalize the web results according to the users' or item preferences. The proposed solution can be implemented in e-commerce to facilitate the users in selecting items in which they are interested. Based on this attempt, we will implement the idea on some web based application and extend this research to provide the better solutions for the other web personalization techniques.

References

[1] J. L. Herlocker, J. A. Konstan, and J. Riedl, "Explaining Collaborative Filtering Recommendations", Proceedings of the 2000 ACM conference on Computer supported cooperative work.
[2] R. van Meteren and M. van Someren, "Using Content-Based Filtering for Recommendation", Computer and Information Science 2006, Vol 4203, pp 47-56
[3] M. Eirinaki, M. Vazirgiannis,"Web mining for web personalization", ACM Transactions on Internet Technology (TOIT), Volume 3 Issue 1, February 2003
[4] F. L. Wang, "Improvements to Collaborative Filtering Systems", Computational and Information Science, 2005.
[5] G. Linden, B. Smith, and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering", Proceeding of the Internet Computing, IEEE, 2003
[6] Good, N. Schafer, J. B. Konstan, et.al, "Combining Collaborative Filtering with Personal Agents for Better Recommendations", Proceedings of The National Conference on Artificial Intelligence, 1999
[7] Z. Huang, H. Chen, D. Zeng,"Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collaborative Filtering", ACM Transactions on Information Systems (TOIS), Vol 22(1)
[8] S.S.Anand, B.Mobasher, "Introduction to Intelligent techniques for Web Personalization", ACM Transactions
on Internet Technology (TOIT), Vol 7(4)
[9] B. Mobasher, H. Dai, T. Luo et.al, "Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization"
[10] B.Sarwar, G.Karypis, J.Konstan, et.al, "Item-based collaborative filtering recommendation algorithms", Proceedings of the 10th international conference on World Wide Web, 2001
[11] P. Shoval, V. Maidel, B. Shapira, "An Ontology- Content-Based Filtering Method",International Journal ITA 2008, Vol 15(4)
[12] Oemarjadi, C.S.; Maulidevi, N.U. "Web personalization in used cars ecommerce site" Electrical Engineering and Informatics (ICEEI), 2011

International Journal of
Soft Computing And Software Engineering (JSCSE)
e-ISSN: 2251-7545
Vol.3,No.11, 2013
Published online: Nov 25, 2013
DOI: 10.7321/jscse.v3.n11.3

[13] Cacheda F., Carneiro V., Fern´ Andez D,and Formoso V., "Comparison of Collaborative Filtering Algorithms: Limitations of Current Techniques and Proposals for scalable, High-Performance Recommender Systems" ACM Transactions on the Web, 2011, Vol. 5(1)

Table 1. Project selection matrix rules.

Issues	**Collaborative filtering technique**	**Web usage mining technique**	**Rule based filtering technique**	**Content based Filtering technique**
Definition	An approach to provide recommendations based on the similar users preferences	Process of analyzing web log data and applies data mining techniques on it to discover usage patterns from web data	Based on "if this, then that" rule processing	An approach to provide recommendations on the bases of item profile and user profile
Functionality	Analyze the user data and extract useful information for future prediction	Analyze user's web log data and discover useful pattern that provide relevant information	Websites define some set of rules, so users have to done searching and performed transactions according to those set of "rules"	Match and measure similarity between user and item profile. The matching denoted as a ranking score indicate similarity between the user and each item profile
Privacy	User covers his/her personal data and then send to data collector, because data collector can't receive truthful information about user's private information	Users use privacy enhancing technologies when they provide personal information online and check privacy policies on websites	No privacy issue	Password protection and encryption technique is use to stop unauthorized access to the user private information
Usability	Facilitates users by making recommendations on the bases of similar user's profile	Discovery of usage patterns facilitate user to view relevant results	Users perform searching and transaction on the bases of rules provided by the website	Facilitates users by making recommendations on the bases of specific user profile
Users	Selects items on the bases of relationship between users with similar preferences	Discovers usage patterns on the bases of users log data	Systems do not have to predict out of the users need. Administrator keeps the system up to date for all users	Selects items based on the correlation between the content of the items and the user's

Table2. Comparison of different web sites

Features	**Igoogle.com**	**Amazon.com**	**Netflix.com**	**Photoree.com**	**Digg.com**	**Barnes and Nobles.com**	**Movielens.org**
Definition	Brings all google function and its contents from whole web on single page	Facilitate user to do online shopping of books and other products	Largest online movie DVD rental service website	System with huge collection of photos	User are allowed to submit articles, contents, news, images and online videos	Facilitate user to do online shopping of Books, DVD's and music	Provide user with rating services on movies through rating scale
Personalization	User can personalize gadgets, themes and tabs on igoogle home page	Provide *Wishlist* that contain personalized list of books which users like to buy	Provide *Queue* that contain personalized list of movies which users like to borrow	Provide *Update galley* contain personalized list of products which users like to buy	No personalized queue	Provide *Favorites* that contain personalized list of products which users like to buy	Provide *Wishlist* that contain personalized list of movies which users like
Are they recommender system?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Membership	Compulsory	Compulsory	Compulsory	compulsory	compulsory	Compulsory	Compulsory
Personalization Technique	Initially content base, now CFT	Initially content base, now CFT	CFT	CFT	CFT	CFT	CFT
Maintain User browsing History	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Recommendation	Provide user content recommendation when users add tabs	Provide recommendation for particular user feature	Provide recommendation feature	Provide recommendation at run time	Provide more recent recommendation feature	Provide suggestion for user feature	Provide top picks for user feature
Similarity Statistics	Not Provided	Not Provided	Not Provided	Provided	Provided	Not Provided	Provided
User Rating	1-5 stars	1-5 stars	1-5 stars	*up* and *down rate*	*digg* feature	1-5 stars	1-5 stars

Maintains User Profile		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sparsity	Sparse items	Not Recommended	Not Recommended	Not Recommended	Not Recommended	Recommended	Not Recommended	Recommended
	Cold start	Exist	Exist	Exist	Exist	Reduce	Exist	Reduce
	First rater	Exist	Exist	Exist	Exist	Reduce	Exist	Reduce
Black Boxes		Not exist	Not exist	Not exist	Exist	Not exist	Exist	Exist