International Journal For Technological Research In Engineering
Volume 4, Issue 2, October-2016 *ISSN (Online): 2347 - 4718*

WEB PERSONALIZATION RECOMMENDATION SYSTEM BASED ON CLUSTERING AND ASSOCIATION RULE

Mr. Somnath Lomate
Zeal College of Engineering and Research, Pune

ABSTRACT: ***The main problem faced by the users of web search today is the quality and the amount of the results they get back. The results frustrate a user and consume his precious time. The objective of a web personalization system is to provide users with the information they want or need, without expecting from them to ask for it explicitly. A web recommender system is a web-based interactive software agent. A WRS attempts to predict user preferences from user data and/or user access data for the purpose of facilitating and personalizing users' experience on-line by providing them with recommendation lists of suggested items. This research proposes a new personalized recommendation system integrating clustering and association rule technique. This system improves the recommendation quality of system and save time of recommendation process. It also overcomes the drawbacks of traditional recommendation system.***

Keywords: ***web personalization recommendation system; association rules; Clustering, K-medoids, filtering, web mining.***

I. INTRODUCTION

The main problem faced by the users of web search today is the quality and the amount of the results they get back. The results frustrate a user and consume his precious time. The objective of a web personalization system is to provide users with the information they want or need, without expecting from them to ask for it explicitly. Web personalization is the process of customizing a Web site to the needs of specific users, taking Advantage of the knowledge acquired from the analysis of the user's navigational behavior (usage data) in correlation with other information collected in the Web context, namely structure, content and user profile data. Due to the explosive growth of the Web, the domain of Web personalization has gained great momentum both in the research and the commercial area. (Magdalini Eirinaki, Michalis Vazirgiannis)[29]. A web recommender system is a web-based interactive software agent. A WRS attempts to predict user preferences from user data and/or user access data for the purpose of facilitating and personalizing users' experience on-line by providing them with recommendation lists of suggested items. In the context of personalized recommendation, resources (web pages, products, advertisements, etc.) Are recommended to a user according To the inner-established knowledge model that anticipates the user's needs. In this method, we the access patterns constructed by analyzing user navigation information (Fu, Budzik, & Hammond, 2000). In the WWW context, web sites are generating a great amount of web usage data that contain useful information about users' behavior. The term 'Web Usage Mining' (Cooley, Mobasher, & Srivastava, 1997) was introduced by Cooley et al., in 1997, in which they define web usage mining as the 'automatic discovery of user access patterns from Web Servers'. Web usage mining has gained much attention in the literature as a potential approach to fulfill the requirement of web personalization (Cooley et al., Eirinaki & Vazirgiannis, 2003; Fu et al., 2000; Gery & Haddad, 2003; Mobasher)

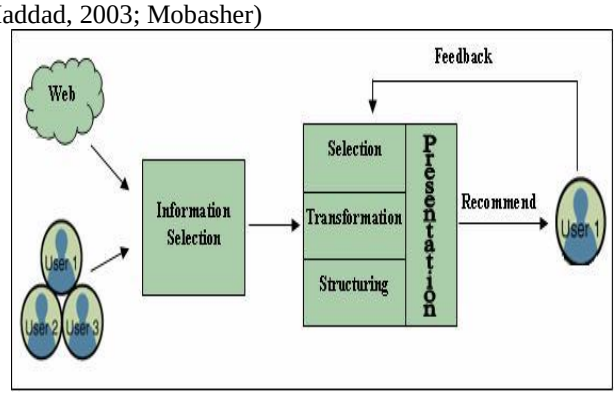


Fig 1.o Recommendation Process

Web personalization process viewed as an application of data mining requiring support for all the phases of a typical data mining cycle. These phases include data collection and preprocessing, Pattern discovery and evaluation, and finally applying the discovered knowledge in real-time to mediate between the user and the Web. We consider a number of classes of data mining algorithms used particularly for Web personalization, including techniques based on clustering, association rule discovery, sequential pattern mining, Markov models, and probabilistic mixture and hidden (latent) variable models.(Bamshad Mobasher) . The ability of a personalization system to tailor content and recommend items implies that it must be able to infer what a user requires based on previous or current interactions with that user, and possibly other users.

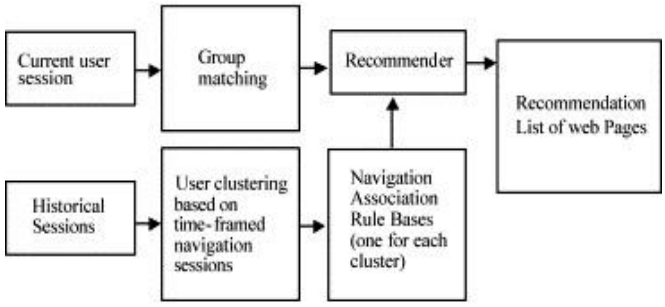


Fig.1.1 The personalized recommendation mechanism.

Principal elements of Web personalization include (a) the categorization and pre processing of Web data, (b) the extraction of correlations between and across different kinds

of such data and (c) the Determination of the actions that should be recommended by such a personalization system.

II. LITERATURE SURVEY

A. Content-based filtering systems

Content-based filtering examines the relationship between resources and users. It takes advantage of the similarity of information and users' interests to filter information. The weakness of the approach is that it is difficult to discern the quality and form of the information sought in current user sessions and it can only recommend information similar to the user's previously-identified interests, rather than advising on newly-identified resource needs arising from concurrent sessions.Content-based filtering establishes user profiles according to each user are visiting history and content searches and then classifies website content on this basis. When a user visits the website and seeks resources that match the established profile, the related information is retrieved. It is a system that recommends an item to a user based upon a description of the item and a profile of the user's interests. Content-based recommendation systems may be used in a variety of domains ranging from recommending web pages, news articles, restaurants, television programs, and items for sale. Although the details of various systems differ, content-based recommendation systems share in common a means for describing the items that may be recommended, a means for creating a profile of the user that describes the types of items the user likes, and a means of comparing items to the user profile to determine what to recommend.(Michael J. Pazzani and Daniel Billsus) Content-based filtering techniques are based on content analysis of target items. For examples, the technique of term frequency analysis for text document and its relation to the user's preferences is a well-known content analysis method. In content-based filtering systems, recommendations are provided for a user based solely on a profile built up by analyzing the content of items that the user has rated in the past and/or user's personal

B. Collaborative filtering systems

In collaborative filtering, items are recommended to a particular user when other similar users also prefer them. The definition of 'similarity' between users depends on applications. For example, it may be defined as users having similar ratings of items or users having similar navigation behavior. This kind of recommendation systems is the first one that uses the artificial intelligence technique to do the personalized job (Riecken, 2000). A collaborative filtering system collects all information about users' activities on the web site and calculates the similarity among the users. If some users have similar behavior, they will be categorized to the same user group. When a user logins into the web site again, the system will first compute the group most similar to the user using methods like the k-nearest neighborhood, and then recommend items that the members of the group prefer to the user. A pure collaborative filtering system also has several shortcomings and critical issues, including that the coverage of item ratings could be very sparse, hence yielding poor recommendation efficiency; and that it is difficult to provide services for users who have unusual tastes, and the user clustering and classification problems for users with changing and/or evolving preferences (Konstan et al., 1997).
In a general way, collaborative filtering recommendation systems work as follows: first the system collects and maintains information about the user. This information includes specific interest of users in certain items and it is stored in separated profiles. Once all the profiles have been collected the system compares all the profiles in order to determine similarities between them. The way similarity is computed depends on the algorithm used and can vary from system to system. Finally, to produce recommendations to a user the system creates a set with the most similar profiles and use the information contained in this set to do the recommendations. One of the biggest drawbacks of collaborative filtering is that it is highly dependent on the information provided by the users. If the user does not provide reliable information the performance of the system will decrease considerably. In the past years some work has been done to overcome these situations. For instance, the work presented in [03] attempts to create clusters related with music artists by crawling the web. This kind of information can be used later for testing the reliability of the information provided by users in the system. Furthermore, the work presented in [20] attempts to retrieve collections of lists of related music from the web. Then it uses these lists as pseudo users for collaborative filtering systems. The experiments presented in these works prove that methods for recollecting information automatically are nearly as effective as data provided by real users.

C. Hybrid recommender system

Recent research has demonstrated that a hybrid approach, combining collaborative filtering and content-based filtering could be more effective in some cases. Hybrid approaches can be implemented in several ways: by making content-based and collaborative-based predictions separately and then combining them; by adding content-based capabilities to a collaborative-based approach (and vice versa); or by unifying the approaches into one model. Several studies empirically compare the performance of the hybrid with the pure collaborative and content-based methods and demonstrate that the hybrid methods can provide more accurate recommendations than pure approaches. It can not recommend new items to the users and completely denies any information that could be extracted from contents of item [11]. On the other hand, content-based methods fail in providing as good recommendations as collaborative filtering does. The reason for this is that it is hard to extract really high level meaningful features of music from the audio signals. Hybrid recommendations systems are developed in the recent years as an attempt of overcome the weakness of pure content-based or pure collaborative methods. The main idea behind hybrid recommendation techniques, as stated in [11], is that "a combination of algorithms can provide more accurate recommendations than a single algorithm and disadvantages of one algorithm can be overcome by other algorithms". According to [10] incorporating content into collaborative filtering systems allows increasing the quality of a recommendation system. Besides, when data is too

sparse additional content information is a need in order to fit global probabilistic models. The work presented in [9] explains that a method that integrates both ratings and content data enables more accurate recommendations with a richer variety than pure content-based or pure collaborative filtering techniques.

III. PROPOSED CLUSTERING ALGORITHM

The proposed clustering algorithm consists of five steps, firstly set the minimal intra cluster similarity then calculate the medoid after that calculate the average intra-cluster similarity then apply the association mining rules.

I) set the minimal intra cluster similarity: δ it is a user specified parameter. Initially there is only one cluster consisting of all objects.

II) Compute medoid: calculate the average similarity between tentative medoid and the other objects within the cluster. Then apply algorithm of swapping medoids in k-medoids algorithm, and find the new medoid the results in the maximal average similarity. Repeat same process, until no new medoid can found.

III) Calculate the average intra-cluster similarity: For each cluster i, calculate the average intra-cluster similarity si, where

$$\text{Si}= \underset{o}{Avg} \{Sim(G_i, p)\}$$

Where Gi is the methoid of cluster I, p is an object in the cluster I, and Sim () is a similarity function.

IV) If si< δ apply the 2-medoid algorithm to divide cluster i into two sub-cluster and repeat Steps 2 to 4 , otherwise stop.

V) Apply association mining rules.

IV. RECOMMENDATION PROCESS

A. Calculating the similarity:

Calculate the similarity between items in the recommendation list.si and sj are items in the recommendation list, so similarity can be calculated by using formula.

$$Sim(S_{is}, S_{jt}) = \frac{|S_{is} \cap S_{jt}|}{|S_{is} \cup S_{jt}|}$$

$1 \leq s \leq n$, $1 \leq t \leq m$. similarity is calculated between two or different items in the recommendation item set.

Users' navigation sessions are divided into frames of navigation sessions according to a pre-specified time interval. Given two users Ui and Uj; and one of their time-framed navigation sessions, as shown below, respectively,

$$\left\langle U_i : TF_u(U_i) = \{S_{i1}, S_{i2}, \ldots S_{in}\} \right\rangle$$

$$\left\langle U_j : TF_v(U_j) = \{S_{j1}, S_{j2}, \ldots S_{jm}\} \right\rangle$$

Is the uth time-framed sessions.

The vth time –framed sessions. Where session S_k is a collection of web pages that the users have visited during a session. U_i And U_j are two users.

b. Weighted precision rates:

$$AWP = \frac{\sum_1^{\min(|A_i|,|R_i|)} W_j}{\sum_{i=1,\ldots,|R_i|} W_j}$$

Where AWP is a average weighted precision rate.

c. Weighted recall rates:

$$WR_i^{\max} = \frac{\sum_1^{\min(|A_i|,|R_i|)} W_j}{\sum_{j=1,\ldots,|A_i|} W_j}$$

Is used to calculate the Average weighted recall rate.

$$AWR = \frac{\sum_{i=1}^{s} WR_i}{\sum_{i=1}^{s} WR_i^{\max}}$$

Is the average weighted recall rate.

d. Data collection and Preprocessing.

V. EXPERIMENTS

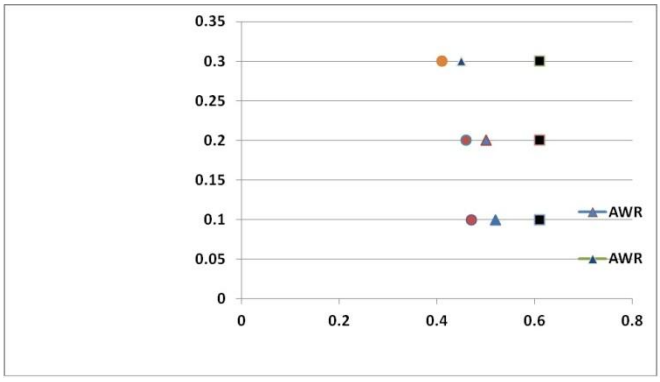
Performance evaluations of clustering

In order to check the performance of the proposed clustering algorithm, the algorithm is first applied to web based virtual Classroom in Ming chuan University (http://www.eduplanet.mcu.edu.tw/) in which learning activities include the browsing of course syllabus, material, learning sheet and work sheet, online testing ,GD,BBS, chatting room and so on.

Frame size	Class	Number of framed sessions	Number of clusters
semester	Half of A	35	5
	A	69	7
Week	Half of A	152	26
	A	315	29

Table 1. Result of clustering.

VI. RESULT



VII. CONCLUSION

This project describes Web personalized recommender

system that utilizes clustering of virtual class room database through proposed Clustering algorithm and provides the recommendations for the user with good quality rating using similarity measures. The proposed clustering algorithm has better accuracy than K-means and K-medoid clustering, which helps to improve the quality of rating. In traditional recommender system similarity is normally the only heuristic used in recommendation process where as in the proposed system, similarity is combined with density of the clusters. This helps in exploration of other clusters which have similarity closer to the active user and provide him/her with good set of recommendations.

REFERENCES

[1] Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions Knowledge Database Engineering, 17(6), 734–774.

[2] Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. Proceedings of ACM SIGMOD (pp. 207–216).

[3] Agrawal, R., & Srikant, R. (1994). Fast algorithm for mining association rules. Proceedings of the VLDB conference (pp. 487–499).

[4] Balabanovic, M., & Sholam, Y.

[5] (1997).Combining content-based And collaborative recommendation. Communications of ACM, 40(3), 46–61.

[6] Balabanovic, M., Shoham Y.: FAB: Content-based, Collaborative Recommendation. Communications of the Association for Computing Machinery 40(3) (1997) 66-72.

[7] Balabanovic, M. and Y. Shoham. Fab: Content-based, collaborative recommendation. Communications of the ACM, 40(3):66-72, 1997.

[8] Basu, C, Hirsh, H., and Cohen W: Recommendation as Classification: Using Social And Content- Based Information in Recommendation. In: Proceedings of the 15th National Conference on Artificial Intelligence, Madison, WI (1998) 714-720.

[9] Burke, R.: Hybrid Recommender Systems: Survey and Experiments. User Modeling And User-Adapted Interaction 12(4), 331–370 (2002).

[10] Cantador, I., Bellog´ın, A., Castells, P.: News hand: A Semantic Web Approach To Recommending News. In: W. Nejdl, J. Kay, P. Pu, E. Herder (Eds.)Adaptive Hypermedia and Adaptive Web-Based Systems, Lecture Notes in Computer Science, vol. 5149, pp. 279–283. Springer (2008).

[11] Dwyer, F. R. Customer Lifetime Valuation to Support Marketing Decision Making. Journal of Direct Marketing, Vol 3(4), 1989.

[12] Eirinaki, M., & Vazirgiannis, M. (2003). Web mining for Web personalization.ACM Transactions on Internet Technology, 3(1), 1–27.

[13] Fawcett, T., and F. Provost. Combining data mining and machine learning for efficient User profiling. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), 1996.

[14] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE Trans. Knowl. Data Eng., vol. 17, no. 6, pp. 734–749, Jun. 2005.

[15] Hae-Sang & Chi-Hyuck Jun (2009). A simple and fast algorithm for K-medoids clustering. Expert Systems with Applications, 3336–3341.

[16] Kim, B. M., & Li, Q. (2006). A new approach for combining content-based and collaborative filters. Journal of Intelligent Information Systems, 27, 79–91.

[17] Lee, C. H., Kim, Y. H., & Rhee, P. K. (2001). Web personalization expert with combining collaborative filtering and association rule mining technique. Expert Systems with Applications, 21, 131–137.

[18] Mobasher, B., Cooley, R., & Srivastava, J. (2000). Automatic personalization based on Web usage mining. Communications of the ACM, 43(8), 142–151.

[19] M. Balabanovic and M. Shoham, "FAB: Content-based collaborative recommendation,"Commun. ACM, vol. 40, no. 3, pp. 66–72, 1997.

[20] R. Burke, "Hybrid recommender systems: Survey and experiments," User Model. User-Adapted Interact, vol. 12, no. 4, pp. 331–370, 2002.

[21] P. Melville, R.J. Mooney, R. Nagarajan, Content-boosted collaborative filtering for improved recommendations, Proceedings of the 18th national conference on Artificial intelligence, 2002, pp. 187–192.

[22] Sarwar, S., and Karypis et al. (2000b). Item based Collaborative algorithm. In Proceedings of 10th international worldwide conference, Hong Kong, (pp. 285–295).

[23] Shinde, S. K., and Kulkarni, U. V. (2012). Hybrid personalized recommender System using centering-bunching based clustering algorithm. Expert systems With applications, (pp.1381–1387).

[24] Shinde, S. K., & Kulkarni, U. V. (2010). The hybrid web personalized recommendation Based on web usage mining. International Journal of Data Mining, Modeling and Management, 2(4), 315–333.

[25] Pazzani, M.J., Billsus, D.: Content-Based Recommendation Systems. In: P.Brusilovsky, A. Kobsa, W. Nejdl (eds.) The Adaptive Web, Lecture Notes in Computer Science, vol. 4321, pp. 325–341 (2007). ISBN 978-3-540-72078-2.

