Automatic Recommendation of Web Pages in Web Usage Mining

C.P. Sumathi,
Department of Computer Science,
SDNB Vaishnav College for
Women, Chennai, Tamil Nadu, India

R. Padmaja Valli
Research Scholar, Mother Teresa
Women's University Kodaikanal,
Tamil Nadu, India

T. Santhanam
Department of Computer Science,
DG Vaishnav College,
Chennai, Tamil Nadu, India

Abstract—With the rising growth of Web users, Web-based organizations are keen to analyze the on-line browsing behavior of the users in their web site and learn (identify) 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. However, the resulting patterns obtained through data mining techniques did not perform well in predicting the future browsing patterns due to the low matching rate of the resulting rules and users' browsing behavior. This 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. The experimental results performed on real usage data from a commercial web site show a significant improvement in the recommendation effectiveness of the proposed system.

Keywords-Clustering,web usage mining, recommendations

I. INTRODUCTION

The wealth of information on the 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 [7, 15, 1] 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.

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 [7, 15, 1, 3, 14]. The patterns / profiles are discovered by applying common data mining techniques to the preprocessed data and provide input to the recommendation engine that recommends appropriate pages based on the intelligence gained from the usage profiles.

II. RELATED WORK

Recommendation systems have been implemented using various approaches [4, 8, 9, 10].Collaborative filtering approach using kNN(k-Nearest Neighbor) technique is widely used in e-commerce systems. This technique requires explicit feedback provided by the user or user ratings on items. The current user's interest is matched with online clustering of users with "similar interest" to provide recommendations. This leads to severe limitations such as scalability and performance [5] due to the lack of sufficient user information. To overcome these limitations, recent research has focused on Web Usage Mining approach for Web Personalization [15]. This type of approach discovers patterns or usage profiles from implicit feedback such as page visits of users. The pattern discovery phase, using various data mining techniques, is performed offline to improve the scalability of collaborative filtering. The discovered patterns or aggregate usage profiles can be used to provide dynamic recommendations based on the user's short-term interest.

Recent researchers have proposed various recommender systems for online personalization through web usage mining. In [2], a model has been developed for deriving usage profiles using k-means clustering followed by classification for recommender systems to predict the future navigations. Using this approach, the prediction accuracy was improved. In [11], researchers have proposed a novel approach using Longest Common Subsequence algorithm for classifying user navigation patterns for recommendations which improves the quality of the system for predictions. In [3] usage-based personalization using various data mining techniques have been discussed. In [16], a lot of research was focussed on the discovery of user's interest in a session using model-based clustering approach. The resulting clusters are used to recommend pages to the user. The paper uses different methods such as Poisson parameters and entropy to determine the recommendation scores. An usage-based Web Personalization system called WebPersonalizer using Web mining techniques to provide dynamic recommendations was proposed in [5]. In [4], researchers have experimentally evaluated two different techniques such as PACT based on the clustering of user transactions and Association Rule Hypergraph Partitioning based on the clustering of clustering of pageviews for the discovery of usage profiles. In [17], Formal Concept Analysis approach is used to discover user access patterns represented as association rules from web logs which can then be used for personalization and recommendation. An improved Web page prediction accuracy by using a novel approach that involves integrating clustering, association rules and Markov models based on certain constraints has been presented in [9].

However, the quality of recommendations in the existing systems does not satisfy users particularly in large web sites. In this paper, a novel approach is used for classifying user navigation patterns through web usage mining system and thereby provides online recommendation effectively. The proposed approach is tested on msnbc.com dataset. The results indicate that the approach can improve the quality of the system for recommendations.

III. MATERIALS AND METHODS

Identification of the current interests of the user based on the short-term navigational patterns instead of explicit user information has proved to be one of the potential sources for recommendation of pages which may be of interest to the user. This would help organizations in various analyses such as web site improvement.

Various techniques are employed for achieving personalized recommendation. This paper employs web usage mining techniques for determining the interest of “similar” users, technique for classifying and matching an online user based on his browsing interests. A novel approach for recommendations of unvisited pages has been employed. The complete process for recommendation, represented in the architecture [3] broadly consists of two components: offline component and online component. The offline component involves Data Preprocessing, Pattern Discovery and Pattern Analysis. The outcome of the offline component is the derivation of aggregate usage profiles using web usage mining techniques. The online component is responsible for matching the current user’s profile to the aggregate usage profiles. The scope of this paper is to match an online user’s navigational activity with the aggregate usage profiles obtained through mining tasks and provide suitable page recommendations which may be of interest to the user.

A. *Data Preprocessing and Usage Mining*

Data preprocessing and model-based session clustering has been discussed in detail [12, 13]. Data preprocessing is a pre-requisite phase before the data can be mined to obtain useful and interesting patterns. A session file is created which consists of a sequence of user’s request for pages P= $\{p_1, p_2, p_3,\ldots,p_n\}$ and a set of m sessions, S = $\{s_1, s_2, s_3,\ldots,s_m\}$ where each s_i € S is a subset of P.

Example: 5 4 5 6 1 represents a session consisting of a sequence of page requests.

A session-pageview matrix is obtained. Each row represents a session and each column represents a frequency of occurrence of the page view in the session. The above session of page visits is represented in Table I where the first row represents the page id.

TABLE I. SESSION-PAGEVIEW

1	2	3	4	5	6
1	0	0	1	2	1

In the proposed approach, the weight of the pageview is further determined by evaluating the importance of a page in terms of the ratio of the frequency of visits to the page with respect to the overall page visits in a session. A numerical weight is assigned to each pageview visited with the purpose of "measuring" its relative importance/interest within the session. If the page has not been visited, the weight of the page is assigned 0. The page visits repeated consecutively have been treated as a single visit to that page. The weights have been normalized to account for variances.

The session file is represented using the vector space model. Each session s_i is modeled as a vector over the n-dimensional space of pageviews. Each session s_i is represented as

$$s_i = \{pf_1, pf_2, pf_3, \ldots, pf_n\}$$

where each pf_j is the **relative** frequency of pageview j in session i,. This is represented in Table II. This type of weight normalization is referred to as transaction normalization which is beneficial since it captures the relative importance/interest of the pageview in a session.

TABLE II. WEIGHTED SESSION-PAGEVIEW

1	2	3	4	5	6
0.2	0	0	0.2	0.4	0.2

B. *Pattern Discovery - Model Based Session Clustering*

The next step in the offline task is to determine sessions with “similar” navigation patterns/interest from the user session file. A model-based Expectation Maximization clustering technique is employed to determine the session clusters. Each cluster represents several sessions of “similar” usage patterns and does not represent an effective method of depicting the aggregate view of common user patterns. Hence, the profile interest is learnt by determining an aggregate usage profile using the formula

$$wt(pg, up_c) = \frac{1}{nc}\sum_{s\in c} w_{pg}^{s} \longrightarrow \quad (1)$$

where w^{s}_{pg} represents the weight of the page in session $s \in c$ and nc represents the number of sessions in cluster c. Table III shows the aggregate usage profiles.

C. Pattern Analysis

In addition to the discovery of aggregate usage profiles, a learning task is involved to represent the significance of user's (session) interest in the cluster. If the aggregate usage profiles consist of n clusters and k pages, then the significance in the cluster can be determined as follows:

$$\max_{i=0}^{n}(wt(pg_j, up_i)) \;,\; 1 \le j \le 17 = M_{I(j)} \longrightarrow \quad (2)$$

where I(j) is the index of the maximum value in each page and $M_{I(j)}$ represents the maximum value.

For instance, cluster0 predominantly represents users/ sessions interested in weather while cluster 3 represents users interested in Sports and Bulletin and not much interested in weather. This maximization function is used later to recommend pages to users belonging to a profile / cluster.

D. Recommendation Process

The recommendation process is an online phase and consists of two sub-phases: matching profile and recommendation. Initially, the new, partial session is compared to the already existing aggregate usage profiles to obtain matching cluster(s). Since the user's interests are dynamic in nature, a short-term history of navigational behaviour is taken into consideration. This is represented by the last 'n' pages visited by the user in the session and is termed as the active session. A window of fixed size 'n' consisting of the pages in the active session is called the working-set window or sliding window.

The similarity of the active session with each of the discovered aggregate profiles is determined using the well-know cosine similarity measure. If an active session is represented as s_i and cluster as c_k, then their similarity can be measured as follows:

$$\text{sim}(s_j, c_k) = \frac{\sum_{i=1}^{n} w_{i,j} \bullet w_{i,k}}{\sqrt{\sum_{i=1}^{n} w_{i,j}^2} \bullet \sqrt{\sum_{i=1}^{n} w_{i,k}^2}} \longrightarrow \quad (3)$$

where $w_{i,j}$ represents weight of page i in active session j and $w_{i,k}$ represents weight of page i in cluster k.

While the user navigates through the web site, the window also slides and the active session window is compared to the aggregate usage profiles. Profiles having a similarity greater than a threshold value μ_c are selected as matching clusters. Hence, the current interest of the user is determined for each active session. These matching clusters can be used for recommending pages instantaneously which have not been visited by the user. In the proposed approach, each profile determines the significant pages given by Equation (2).

The identification of matching clusters using Equation (3) would help to determine the significant pages in the cluster for recommendation. This would help the user to browse through the recommended pages also which may be of interest to him. The pages representing maximum values in each cluster and greater than the threshold value μ_r are considered as significant pages for recommendation.

E. Experimental Design

The web log files of msnbc.com web site have been used for this research. This data set is publicly available through the UCI KDD Archive (2005) at the University of California. The web site includes the page visits of users who visited the "msnbc.com" web site on 28/9/99. The visits are recorded at the level of URL category (for example sports, news and so on) and are recorded in time order. It includes visits to 17 categories (i.e., 17 distinct pageviews). The data is obtained from IIS logs for msnbc.com and news-related portions of msn.com. The client-side data is not available in the web log files. Each sequence in the dataset corresponds to a user's request for a page.

The 17 categories are:

Id	Category	Id	Category
1	Frontpage	10	Living
2	News	11	Business
3	Tech	12	Sports
4	Local	13	Summary
5	Opinion	14	BBS
6	On-air	15	Travel
7	Misc	16	Msn-news
8	Weather	17	Msn-sports
9	Health		

The offline component has been discussed in detail in [12, 13. A clustering model is estimated using 10,000 samples approximately in the dataset which are generally split into training and testing sets. The model is first designed using training samples and then it is evaluated based on the performance on the test samples. In the proposed approach, the dataset has been partitioned into 60% of training data and the remaining 40% as test data. The Expectation Maximization clustering algorithm has been applied. The experiment was performed within 10 iterations resulting in 9 clusters with a

TABLE III. AGGREGATE USAGE PROFILES

C	F-Page	News	Tech	Local	Opin	On-air	Misc	W'ther	Health	Living	Buz	Sports	Sumary	B'tin	Travel	msn-N	msn-S
C0	0.024	0.003	0.002	0.002	0.001	0.005	0.000	0.946	0.001	0.006	0.006	0.002	0.000	0.001	0.000	0.000	0.001
C1	0.146	0.075	0.010	0.008	0.084	0.028	0.003	0.001	0.002	0.238	0.221	0.009	0.003	0.002	0.112	0.001	0.058
C2	0.000	0.362	0.040	0.028	0.000	0.107	0.047	0.004	0.000	0.001	0.011	0.000	0.321	0.061	0.017	0.000	0.001
C3	0.004	0.000	0.000	0.000	0.001	0.003	0.001	0.007	0.000	0.005	0.000	0.472	0.004	0.502	0.001	0.000	0.000
C4	0.094	0.015	0.736	0.047	0.003	0.022	0.007	0.000	0.009	0.047	0.005	0.008	0.000	0.000	0.007	0.000	0.000
C5	0.061	0.008	0.044	0.065	0.035	0.021	0.037	0.035	0.563	0.000	0.018	0.054	0.035	0.009	0.013	0.000	0.000
C6	0.653	0.089	0.016	0.021	0.000	0.020	0.021	0.000	0.022	0.004	0.030	0.067	0.009	0.049	0.000	0.000	0.000
C7	0.028	0.012	0.011	0.000	0.000	0.884	0.000	0.037	0.018	0.002	0.000	0.007	0.000	0.002	0.000	0.000	0.000
C8	0.045	0.029	0.006	0.635	0.000	0.038	0.080	0.076	0.000	0.012	0.000	0.009	0.016	0.053	0.000	0.000	0.003

Maximum Likelihood estimate of 24.95867. Each cluster represents several sessions of navigational patterns representing "similar" interest in the web pages or the usage profile. Hence an aggregate usage profile is determined using Equation (1).

The online component is developed using Java connected to Microsoft Oracle through JDBC. During the online phase, the pages visited in a session are stored in a user session file. After each page visit, the relative frequency of pageviews in the active session is determined. An active session with sliding window size 'n' consists of the current page visit and the most recent n-1 pages visited. As the user browses through various pages, the window slides. In the proposed approach, the size of the sliding window is taken as 5 since it is the average number of pagevisits in the dataset.

The active session is matched with the aggregate usage profiles and matching cluster(s), greater than the threshold value, are obtained using the cosine similarity measure. These matching cluster(s) are used for recommendation purposes for recommending pages exceeding the recommendation threshold that have not been visited by the user.

F. Results and Discussion

For example, consider the following 3 sessions consisting of page visits

1 6 1

1 6 11

1 11 1 11 1 14 1

For example, consider session 3. The most recent 'n' pages (n=5) constitute the sliding window. The sliding window consists of the pages 1 11 1 11 1 in the fifth page visit. Table IV represents the page visits in the sliding window. The frequency of visited pages is shown in Table V. Figure 1 represents the pages visited in each session.

From Figure 1, the frequently visited pages in sessions and the page which is repeatedly visited within a session can be identified.The weight of the pageview is further determined by evaluating the importance of a page in terms of the ratio of the frequency of visits to the page with respect to the overall page visits in the active session. This is represented in Table VI.

The page(s) visited in the active session is matched with the aggregate usage profiles using cosine similarity measure. Cluster(s) greater than the threshold value, µc are chosen to be matching clusters. Table VII represents the matching clusters.

In comparison with the aggregate usage profiles and the maximization function (Equation 2), it has been found that when the user visits page 1 (window size 1), the appropriate clusters, exceeding the threshold value are cluster 6 and cluster 1. Similarly when the user visits page 6 subsequent to page visit 1(window size 2), the appropriate matching clusters are cluster 7, cluster 6 and cluster 1. As the window size increases to the fixed size limit (n=5), correspondingly, the matching clusters for the visited page(s) in the active session and the recommendations are dynamic in nature.

A graphical representation for cluster similarity for page 1 (window size 1) of Table VII and its comparison with the aggregate usage profiles (Table III) is shown in Figure 2, Figure 3 and Figure 4.

From Figure 2, page 1 is largely associated to cluster 6 and, to a certain extent, to cluster 1. The aggregate usage profiles for cluster 6 and cluster 1 has been shown graphically in Figure 3 and Figure 4. It has been found that users who are greatly interested in page1 (Front-page) are associated to cluster 6 and cluster 1 in varying extents.

G. Recommendations

The matching cluster(s) are used for recommendation purposes for recommending pages, exceeding the recommendation threshold, that have not been visited by the user. For recommendation of pages, the significant pages in the appropriate matching cluster taken into consideration. From Figures 3 and 4, the pages not visited by the user can be recommended. It is seen that pages 5, 10, 11, 15, 17 can be recommended from cluster 1. Since cluster 6 is fully concentrated in page 1, there are no other pages to recommend from cluster 1. Table VIII shows the recommended set of pages for 3 sessions.

TABLE IV. PAGE VISITS IN THE SLIDING WINDOW

Session	Order of Visit	Window	Active Session (Page Visited)				
1	1	1	1	0	0	0	0
1	2	2	1	6	0	0	0
1	3	3	1	6	1	0	0
2	1	1	1	0	0	0	0
2	2	2	1	6	0	0	0
2	3	3	1	6	11	0	0
3	1	1	1	0	0	0	0
3	2	2	1	11	0	0	0
3	3	3	1	11	1	0	0
3	4	4	1	11	1	11	0
3	5	5	1	11	1	11	1
3	6	6	11	1	11	1	14
3	7	7	1	11	1	14	1

TABLE V. FREQUENCY OF PAGE VISITED

Session	Order of Visit	Front-Page	News	Tech	Local	Opinion	On-air	Misc	Weather	Health	Living	Business	Sports	Summary	B'tin	Travel	msn-News	msn-Summar
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	2	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	3	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
2	3	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
3	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	2	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
3	3	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
3	4	2	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
3	5	3	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
3	6	2	0	0	0	0	0	0	0	0	0	2	0	0	1	0	0	0
3	7	3	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0

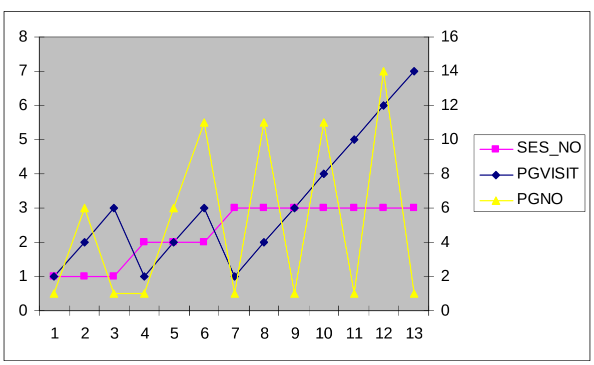

Figure 1. Pages visited in each session

TABLE VI. WEIGHTED PAGEVIEW

Session	Order of Visit	Front-Page	News	Tech	Local	Opinion	On-air	Misc	Weather	Health	Living	Business	Sports	Summary	B'tin	Travel	msn-News	msn-Summary
1	1	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	2	0.500	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	3	0.667	0.000	0.000	0.000	0.000	0.333	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	1	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	2	0.500	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	3	0.333	0.000	0.000	0.000	0.000	0.333	0.000	0.000	0.000	0.000	0.333	0.000	0.000	0.000	0.000	0.000	0.000
3	1	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	2	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000
3	3	0.667	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.333	0.000	0.000	0.000	0.000	0.000	0.000
3	4	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000
3	5	0.600	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.400	0.000	0.000	0.000	0.000	0.000	0.000
3	6	0.400	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.400	0.000	0.000	0.200	0.000	0.000	0.000
3	7	0.600	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.200	0.000	0.000	0.200	0.000	0.000	0.000

TABLE VII. MATCHING CLUSTERS

Session	Order of Visit	Page Visited	Window	C0	C1	C2	C3	C4	C5	C6	C7	C8
1	1	1	1	0.000	0.098	0.000	0.000	0.000	0.000	0.306	0.000	0.000
1	2	6	2	0.000	0.075	0.000	0.000	0.000	0.000	0.187	0.230	0.000
1	3	1	3	0.000	0.089	0.000	0.000	0.000	0.000	0.240	0.155	0.000
2	1	1	1	0.000	0.098	0.000	0.000	0.000	0.000	0.306	0.000	0.000
2	2	6	2	0.000	0.075	0.000	0.000	0.000	0.000	0.187	0.230	0.000
2	3	11	3	0.000	0.131	0.000	0.000	0.000	0.000	0.141	0.166	0.000
3	1	1	1	0.000	0.098	0.000	0.000	0.000	0.000	0.306	0.000	0.000
3	2	11	2	0.000	0.159	0.000	0.000	0.000	0.000	0.190	0.000	0.000
3	3	1	3	0.000	0.143	0.000	0.000	0.000	0.000	0.242	0.000	0.000
3	4	11	4	0.000	0.159	0.000	0.000	0.000	0.000	0.190	0.000	0.000
3	5	1	5	0.000	0.150	0.000	0.000	0.000	0.000	0.222	0.000	0.000
3	6	14	6	0.000	0.143	0.000	0.065	0.000	0.000	0.168	0.000	0.000
3	7	1	7	0.000	0.120	0.000	0.063	0.000	0.000	0.232	0.000	0.000

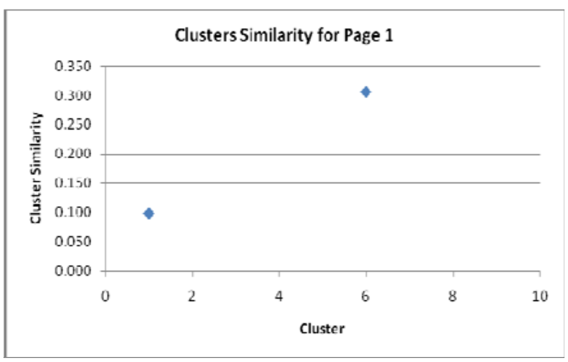

Figure 2. Cluster Similarity for page1

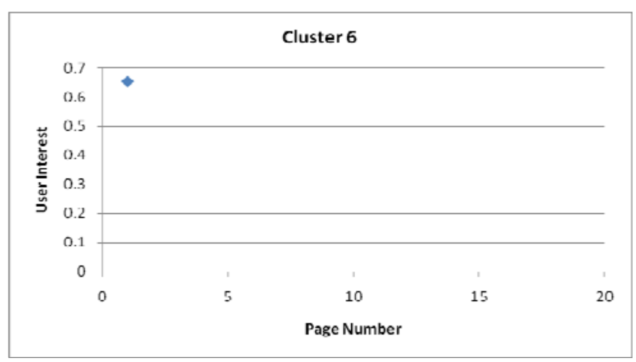

Figure 3. Aggregate Usage profile of Cluster 6

Figure 4. Aggregate Usge Profile of Cluster 1

TABLE VIII. RECOMMENDATION SET FOR SESSION1, SESSION 2 AND SESSION 3

Session	Order of Visit	Active Session Window (Pages)	Matching Cluster(s)	Recommended Pages
1	1	1	1,6	5, 10, 11, 15, 17
1	2	1-->6	1,6,7	5, 10, 11, 15, 17
1	3	1-->6-->1	1, 6, 7	5, 10, 11, 15, 17
2	1	1	1, 6	5, 10, 11, 15, 17
2	2	1-->6	1,6	5, 10, 11, 15, 17
2	3	1-->6-->11	1, 6, 7	5, 10, 15, 17
3	1	1	1,6	5, 10, 11, 15, 17
3	2	1-->11	1, 6	5, 10, 15, 17
3	3	1-->11-->1	1, 6	5, 10, 15, 17
3	4	1-->11-->1-->11	1, 6	5, 10, 15, 17
3	5	1-->11-->1-->11-->1	1, 6	5, 10, 15, 17
3	6	11-->1-->11-->1-->14	1,3,6	5,10, 12, 15, 17
3	7	1-->11-->1-->14-->1	1, 3, 6	5, 10, 12, 15, 17

IV. CONCLUSION AND FUTURE WORKS

Identification of the current interests of the user based on the short-term navigational patterns instead of explicit user information has proved to be one of the potential sources for recommendation of pages which may be of interest to the user. In this work, we classify and match an online user based on his browsing interests using statistical techniques.

A novel approach for recommendations of unvisited pages has been suggested in this work. An offline data preprocessing and clustering approach is used to determine groups of users with similar browsing patterns. Future work may involve the use of optimization techniques to assess the quality of recommendations.

V. REFERENCES

[1] Agarwal, R. and R. Srikant, 1994. Fast algorithms for mining association rules in large database. Proceeding of the 20th Conference on Very Large Data Bases, Sept. 12-15, Morgan Kaufmann Publishers Inc., San Francisco, CA., USA., pp: 487-499. DOI: 10.1234/12345678

[2] AlMurtadha, Y.M., M.N.B. Sulaiman, N. Mustapha and N.I. Udzir, 2010. Mining web navigation profiles for recommendation system. Inform.Technol. J., 9: 790-796. DOI: 10.3923/itj.2010.790.796

[3] Bamshad Mobasher, Robert Cooley and Jaideep Srivatsava, 2000. Automatic personalization based on Web usagmining. Commun. ACM, 43:142-151. DOI:10.1145/345124.345169

[4] Bamshad Mobasher, Honghua Dai, Tao Luo, Miki Nakagawa, 2002. Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization. Commun. ACM, 6:61-82 DOI: 10.1023/A:1013232803866

[5] Bamshad Mobasher,H.Dai,T.Luo and M.Nakagawa,2001.Improving the effectiveness of collaborative filtering on anonymous Web usage data. In Proceedings of the IJCAI 2001 Workshop on Intelligent Techniques for Web Personalization (ITWP01),Seattle.

[6] Bamshad Mobasher,2001. WebPersonalizer: A Server-Side Recommender System Based on Web Usage Mining. Technical Report TR01-010, School of Computer Science, Telecommunications and Information Systems, DePaul University, Chicago, IL, USA

[7] Cooley, R., B. Mobasher and J. Srivatsava, 1997. Web mining: Information and pattern discovery on the world wide web. Proceeding of the 9th IEEE International Conference on Tools with Artificial Intelligence, Newport Beach, CA., pp: 558-567. DOI: 10.1109/TAI.1997.632303

[8] Eirinaki M. , M. Vazirgiannis, 2003.Web mining for Web personalization. ACM Transactions on Internet Technology, 3(1), pp. 1–27.http://doi.acm.org/10.1145/643477.

[9] Faten Khalil, Jiuyong Li and Hua Wang,2008. Integrating Recommendation Models for Improved Web Page Prediction Accuracy. Proceedings of the thirty-first Australasian conference on Computer science, 74:91-100

[10] Forsati R., M. R. Meybodi and A. Ghari Neiat, 2009. Web Page Personalization based on Weighted Association Rules. International Conference on Electronic Computer Technology, pp: 130-135 DOI: 10.1109/ICECT.2009.104

[11] Mehrdad Jalali, Norwati Mustapha, Ali Mamat, Md. Nasir B Sulaiman , 2009, A Recommender System for Online Personalization in the WUM Applications. Proceedings of the World Congress on Engineering and Computer Science 2009 Vol II, San Francisco, USA http://www.iaeng.org/publication/WCECS2009/WCECS2009_pp741-746.pdf

[12] Sumathi C.P, R. Padmaja Valli and T. Santhanam, 2010. An Application of Session Based Clustering to Analyze Web Pages of User Interest from Web Log Files, Journal of Computer Science,pp:785-793

[13] Suresh R.M. and R. Padmajavalli, 2006. Overview of data preprocessing in data and web usage mining. Proceedings of the IEEE International Conference on Digital Information Management, December 6-12, Bangalore,pp: 193-198

[14] Spiliopoulou, M. (2000). Web usage mining for Web site evaluation. Communications of ACM. 43(8), 127-134.

[15] Srivatsava, J., R. Cooley, M. Deshpande and P.N. Tan, 2000. Web usage mining: discovery and applications of usage patterns from Web data. ACM SIGKDD Explorat. Newsletter, 1: 12-23. DOI: 10.1145/846183.846188

[16] Şule Gündüz and M. T. Özsu,2003.A User Interest Model for Web Page Navigation. In Proceedings of International Workshop on Data Mining for Actionable Knowledge (DMAK), Seoul, Korea, pp 46-57.

[17] Vasumathi D. and A. Govardhan, 2005. Efficient Web Usage Mining Based On Formal Concept Analysis. Intelligent Information Processing II IFIP International Federation for Information Processing, Volume 163/2005, 437-441, DOI: 10.1007/0-387-23152-8_55