Advances in Computational Sciences and Technology
ISSN 0973-6107 Volume 10, Number 6 (2017) pp. 1953-1968

http://www.ripublication.com

Enhanced Web Personalization for Improved Browsing Experience

R. B. Wagh[1] and J. B. Patil[2]

[1]*Research Scholar, Dept. of Computer Engineering, R.C.P.I.T., Shirpur*

[2]*Principal and Professor, R.C.P.I.T., Shirpur*

Abstract

A Web Personalization system is the process of customizing the Website to the needs of individual user or a set of users. It makes use of various data mining techniques such as association rule mining, sequential pattern discovery, clustering, classification etc. for accurate prediction of user future movements. Recent techniques for web personalization lag in appropriate prediction of user interests. To provide effective suggestions, we have developed a novel web personalization technique. The proposed work is based on finding appropriate weights among the web pages of a website. We have used distance measure of visit relationship as well as occurrence frequency measure of web pages for this purpose. The approach uses enhanced graph based partitioning algorithm for clustering of web pages and classify the current user activities more accurately. A Threshold value is used to make a decision among web pages for recommendation purpose. Our experimental results get around 61% accuracy, 34 % coverage and 44.23 % F1 measure. It helps to improve browsing experience of user.

Keywords: Classification, Clustering, Recommender Systems, Web Personalization, Web Usage Mining

1. INTRODUCTION

Huge amount of data and information is being uploaded to the internet every day. Many websites such as news portals and other day to day life websites are updated frequently on hourly or daily basis. This makes it necessary to sort out relevant and irrelevant data. Also, most web structures are complicated and large in size. It may mislead the users with unnecessary and unambiguous information. It subsequently results in loss of valuable user time. It will be really better to get rid of this information overload problem and save the searching time of user. Hence it is required to predict the user needs to improve browsing experience and providing them with what they want. Web Personalization or Recommender System is the best available solution now days for this purpose.

It is the process of customizing a Website to the needs of specific users taking benefit of knowledge acquired from the analysis of Web information i.e. content, structure, user profile data along with users' navigational behavior i.e. usage Data. Web personalization is a broader area covering recommender systems, adaptive Web sites and customization. Web customization is the process of adjusting the site to each user's preference regarding its presentation and structure. Whenever the registered user logs in, his / her customized home page is loaded. This process is done manually or semi automatically whereas in Web personalization modifications in structure or content of Website are performed dynamically.

To achieve the objective of "Providing the users with the information they need", Web personalization system may expect few inputs from users explicitly or may do it implicitly by its own [1].

In literature, Web personalization is also defined as the process of providing useful links, items, and objects to the user to save valuable time. The provision is based either on explicit likes/choices of user or learned implicitly by the system. Users are analyzed based on historical browsing behavior, geographical locations, similar users, items, links, products etc. Various data mining techniques are used for user analysis and recommendation purpose. It helps in improving the business or user satisfaction of various E-Commerce websites [2].

Recommender systems mostly deal with rating based web sites where users rate the objects of web site. The object may be movie, music, book, joke or similar product. It is also a problem rich research area and used for many applications. One such example is amazon.com which recommends CDs, Books, Music, movies and other products. Other examples include News at VERSIFI technologies (formerly AdaptiveInfo.com), movies by Movie lens and many more [3].

In this paper, we give a brief overview of elements, steps and types of web personalization in Section 2. Section 3 discusses work related to web mining for web personalization. In Section 4 we focus on methodology whereas Section 5 describes system evaluation and experimental results. Finally we conclude our paper in Section 6 with conclusions and future directions.

2. ELEMENTS, STEPS AND TYPES OF WEB PERSONALIZATION

Personalization is a technique that learns preferences, habits and patterns. It is primarily used in systems which support E-Business. Personalization aims at achieving two objectives from E-business point of view, first is to improve the usability of website and another to retain the users of website.

2.1 Elements

The key elements of it include categorization and preprocessing of Web data, extraction of relationship between these data and finding the actions to be done by the system. Web data can be one of the following [1]:

***User Profile Data*:** It Provides information about users of a Web site. It contains name, age, sex, country, state, marital status, education, interest, etc. for each user of a Web site. It also contains information about user's preferences and interests. Such information is collected through questionnaires or registration forms or can be generated by analysis of Web server logs.

***Content Data*:** It can be simple text, images or structured data such as information retrieved from databases. The content data is presented to the end user.

***Structure Data*:** It refers to the way content is organized. They can be XML or HTML tags (data entities used within a Web page) or it can also be a hyperlink connecting one page to another (data entities to put a Web site together). It is an existence of links between various pages to restrict navigation performed by the user to predefined paths.

Usage *Data:* It represents a Web site's usage. It includes parameters like visitor's IP address, date and time of access, complete path (directories or files) accessed, and referrer's

2.2 Steps

Collection of data: Collecting Web server log files from various web servers

Preprocessing of data: It includes categorization and modeling of collected data

Analysis of collected data: It focuses on various ways of analyzing and mining web data to extract useful knowledge from it.

Prediction/Recommendation: Based on the analysis performed this step recommends the actions to be performed.

2.3 Types

The way in which analysis is carried out gives rise to following four types in general:

Content Based: Individual users' behavior, preferences and past are analyzed by the system and similar items are recommended to the user to achieve personalization.

Collaborative Filtering: This system work on the principle of neighbors or similar users .These are the users whose likes and dislikes matches with current user for whom recommendations are generated. These systems try to find usefulness of items to specific user based on previously rated items by similar users.

Rule Based Learning: A set of questionnaire is to be answered by user to get recommendations.

Web Usage Mining: It makes use of data mining methods and statistical methods to Web log data. As a result the user gets a useful pattern which indicates navigational behavior of user [3].

3. RELATED WORK

Many Techniques are used to personalize a web. Content based, Collaborative filtering, Rule based, Web usage mining are some variants. In this section we will review some recent work related to web personalization emphasizing on Web usage mining.

Z. Malik and C. Fyfe have focused a review of web personalization. The building blocks of web personalization viz. learning, matching and recommendations are discussed in detail with the recent trends. The matching phase and its types content based, collaborative filtering, rule based learning and hybrid approach are explained with benefits and limitations. The challenges like high scalability of data, lack of performance, black box filtration, correct recommendation, and privacy issues are new opportunities for researchers. The authors also throw light on importance of web personalization for its use in e-commerce sites [4].

M. Jalali et al. developed a recommendation system called WebPUM, an online prediction using longest common sequences algorithm (LCS) for classifying user navigation patterns to predict users' future intentions. To effectively provide online prediction, they proposed an approach for classifying user navigation patterns to predict users' future intentions. The approach is based on the new graph partitioning algorithm to model user navigation patterns for the navigation patterns mining phase. The Users and sessions are grouped using DFS Clustering. The system has been tested on CTI and MSNBC datasets. The results show an improvement in the quality of recommendations. The experiments on scalability prove that the size of dataset and the number of the users in dataset do not significantly contribute to the percentage of accuracy [5].

H. Liu and V. Keselj presented a system of automatic classification of web user navigation patterns for predicting user future moves. The approach is based on combined mining of web server logs and contents of retrieved web pages. Character N-gram is used to represent the content of web page. It is combined with web server log to represent user profiles. The approach is implemented as an experiment where they tested classification and prediction accuracy. They got around 70% classification accuracy and 65 % prediction accuracy [6].

Q. Yang et al. developed a technique for personalizing Web page recommendation via collaborative filtering and topic aware markov model .They tried to predict the next request of pages that Web users are potentially interested in when surfing the Web. They implemented a graph-based iteration algorithm to discover users' interested topics, based on which user similarities are measured. To recommend topically coherent pages, the authors derived a topic-aware Markov model to learn users' navigation patterns which capture both temporal and topical relevance of pages [7].

Y. AlMurtadha et al. proposed an Improved Web Page Recommendation system using profile aggregation based on clustering of transactions. The authors have built recommendation system for anonyms' users or visitors' .For this purpose they assigned the current user to the best navigation profile with similar navigation activities [8].

H. Ramadan et al. have focused on web usage mining techniques which are used in applications to model user behavior, finding access patterns as well as user interests. Clustering, Classification, server load balancing, content caching and data distribution insights are focused. Major Aspects and problems related to modeling user behavior are surveyed. Recent advancements related to automatic web navigation and specific page interest of users is discussed for implicit reorganization of web site [9].

S. Vijayalakshmi and V. Mohan developed strategy to discover frequent sequential patterns with minimum support provided. They adopted divide and conquer pattern growth principle approach to better serve the need of web based applications. The method combined tree projection and prefix growth features from pattern growth category and position coded feature from early pruning category [10].

J. Ren and X. Zhou worked on development of a new system to maintain sequential access patterns for an updated database. The authors made use of an efficient technique to reduce the cost of finding new patterns when some information is deleted from a database. They used the knowledge obtained from previous mining process for cost reduction. The MAD (Maintenance Algorithm for Deleted Information) is found to be better than GSP in all cases especially for low support [11].

C. Sumathi et al. made use of session based clustering for development of an application that analyzes Web pages of user interest from Web logs. Web log analysis is done based on automatic discovery of usage interest of web pages stored in web and application server access logs .The approach resulted in usage profiles and automatic identification of user interest in each profile [12].

C. Dimopoulos et al. modeled navigational history of users and weighted suffix trees of web page content were used in order to predict the usage of Web page. It is exploited in an online recommendation system of web site or web page cache system .The proposed method gave benefit in terms of constant efforts for every user as well as consumed relatively less memory space .The performance of the method is fair and better than similar systems [13].

R. Baraglia and F. Silvestri proposed a method SUGGEST for dynamic personalization of web sites without user intervention. They developed an all in one system for user profiling, model updating and recommendation generation. The system considered two level architecture (offline and online) for users' historical knowledge base creation and understanding of users' behavior. A Graph representation is incrementally updated as the request arrives at the system. The active user sessions are classified using graph partitioning algorithm [14].

R. Baraglia and F. Silvestri developed an online recommender system for large web sites. They considered the problem of scalability with large web sites. Some web sites are dynamic in nature and the contents increase on daily basis. The users visiting the web site may also increase regularly. It results in scalability issue of web site. The proposed system dealt with optimizing web server performance [15].

B. Mobasher et al. and M. Nakagawa presented a system for generation of hypertext links as a dynamic recommendation to active user. The system depends upon the combined mining of anonymous usage data and structure of web site. Aggregate usage profiles are generated by the application of data mining techniques clustering, association rule mining, and sequential pattern discovery. Matching pages of active user are used to generate a set of recommendations. These recommendations are assumed to be inserted as a hypertext link in the last requested page of a user [16] [17].

I. Cadez et al. discovered partitioning clustering method in Web-CANVAS to visualize user navigation paths of a cluster. To represent user sessions, categories of general topics for web pages are used. Predefined categories form the base for URLs from Web server log files. The URLs are assigned to one of the category to construct user sessions [18].

M. Perkowitz and O. Etzioni deigned a partitioning graph theoretic approach for Adaptive Web sites. These Web sites can improve the organization and presentation automatically for the end user. They used usage logs for knowledge mining. A new clustering method is proposed for PageGather algorithm. Clusters are group of cliques or coherent connected components. The larger clusters are computationally faster and easy. The PageGather algorithm results in creation of index page with hyperlinks to all pages in a cluster [19].

B. Zhou et al. used sequential access pattern mining for development of Sequential Web Access-based Recommender System (SWARS).They used CS-Mine as an efficient sequential pattern mining algorithm to find frequently occurring sequential web access patterns. A Pattern Tree is used to store access patterns. This Tree is used later in online phase to generate prediction list. The proposed system when tested found to be less efficient in terms of precision when number of recommended pages are more than five [20].

Summarizing the related work for personalization of a Web site as well as recommender systems, we found that all the authors tried to use and divide the reference architecture in two parts: offline and online. Offline part is used for preparing a knowledge base of web sites' usage and users interest as well as behavior. Online part is used for matching the active user session with previously created knowledge base and generating the recommendations in terms of web pages, links, items, products, advertisements etc.

4. METHODOLOGY

To provide effective suggestions, we have developed a novel web personalization technique. Our work aims at personalization of a particular web site. The proposed work is based on predicting user near future requests more accurately. The work is divided in two phases: Offline phase and Online Phase. The Offline phase preprocesses the data and creates clusters of web pages using navigation pattern mining. It creates knowledge base for online phase. The online phase uses this knowledge in order to generate list of recommended web pages. Fig. 1 shows the reference framework for proposed work.

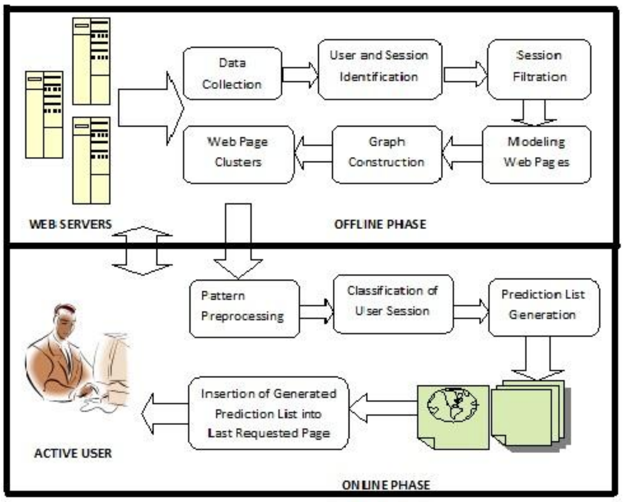

Fig.1 Framework of proposed work

Offline Phase: This phase aims at preprocessing of collected data from web servers. It includes collection of dataset, user and session identification, filtering unwanted and low size sessions. Next step is to create undirected weighted graph based on modeling of web pages. A graph based partitioning algorithm is used to divide the weighted graph. It results in generation of web page clusters. These clusters indicate web pages with coherent property of requests in similar sessions.ink.

Data Collection: In this step, Web server log files are collected to use in further steps.

User and session Identification: This step aims at finding users and sessions through raw web log files collected from server. We used reactive technique for user identification. It attempts to guess the users by IP address, operating system and browser used by anonymous users. Similarly session identification tries to approximate sessions by session duration method.

Session Filtration: It attempts at removal of size 1 sessions as well as sessions with low support page views. There are many sessions which are less useful from clustering concept. Hence we try to remove these sessions from dataset and create a consistent view of dataset obtained from web servers.

We did some more statistical analysis in terms of web log data to help us in formation of more appropriate clusters. It includes Total Duration of a Web Page across all sessions, Average Duration of a Web Page across all sessions, Number and list of sessions in which a web page occurs, Number and list of sessions in which a web page does not occur, Total number of pages visited in every session, Total number of unique pages visited in every session, Number of users and repeat users, Number of sessions before filtering and after filtering, Average Path length through a Web site, Most Referred Web Page of a Web site, Least Referred Web Page of a Web site, Total number of Web pages of website, Number of valid and useful Web pages of website.

After preprocessing of data and parameter analysis, we perform modeling of valid and useful Web pages to know the relationship among them. Valid and useful Web pages are the pages with a minimum support value greater than 0.001.For this we follow below mentioned enhanced modeling to overcome deficiencies of previous reference system.

Step 1) Find the Relationship Matrix for Web Pages of a Web site

Step 2) Creation of more appropriate weighted graph with respect to Relationship Matrix

Step 3) Partitioning of graph and formation of clusters

Step 1a: With the help of above statistical parameter analysis; we found that there are too many Web pages of a Web site which are not related to each other in many sessions. Hence we have proposed a new measure of finding the distance between requests of every two pages in a session.

$$RM_{x,y} = Avg.\left(\sum_{i=1}^{n} \frac{|d(xi) - d(yi)|}{Session\ size\ i\ of\ unique\ pages}\right) \quad (1)$$

Where d (xi) is the position of webpage x in i^{th} session, d (yi) is the position of webpage y in i^{th} session. We consider all the sessions where webpage x and webpage y both have occurred together. The More the value of $RM_{x,y}$,the less they are related whereas the less the value of $RM_{x,y}$ means the more they are related.

Step 1b: Occurrence Frequency measures the average occurrence of both pages in each session as shown below

$$FM_{x,y} = \frac{T_{xy}}{Avg.(T_x , T_y)} \quad (2)$$

Where T_{xy}= Total number of sessions where page x and page y both occur together, T_x = Number of sessions where only page x occurs, T_y=Number of sessions where only page y occurs, By considering average occurrence of both pages, we are trying to give equal importance to both pages

Step 1c: Weight Matrix represents the harmonic mean of relationship matrix and occurrence frequency matrix.

$$WM_{x,y} = \frac{(2 * RM_{x,y} * FM_{x,y})}{(RM_{x,y} + FM_{x,y})} \quad (3)$$

It tries to approximate the visits among web pages.

Step 2: We create an undirected weighted graph corresponding to the weights of weight matrix. The nodes of undirected weighted graph represent the Web pages of a Web site whereas the edges among the nodes represent the relationship found among the Web pages based on above two measures. To reduce number of edges, we use threshold value of such edges. We perform some iteration. In all iterations, we remove edges which are below threshold value. The threshold values range from 0.1 to 0.9.

Step 3: Partitioning of graph and formation of clusters

A Depth First Search algorithm is being used to partition the connected graph of Relationship Matrix into k disjoint sets known as clusters. A Depth First Search algorithm starts its process of splitting the connected graph into different clusters with one of the node x. It tries to reach all possible nodes from node x to result in connected component (cluster) of Web pages w. r. t. node x. It will start formation of another cluster beginning with a node which is not present in previously formed cluster. The process is repeated till all the nodes are assigned to one of the cluster. The resulting clusters represent the navigation patterns of users with similar browsing activities.

Two more parameters are considered while formation of clusters. The first one is the *edge threshold* and the other is *cluster threshold*. The *edge threshold* gets updated in all iterations from 0.1 to 0.9 whereas *cluster threshold* represents minimum number of pages in a cluster and is fixed to 3 pages for getting effective clusters. These clusters will be of optimum size and plays a vital role in online phase.

Online Phase:

This phase aims at generating list of recommended web pages to the active user. It utilizes the clusters of navigation patterns generated out of offline phase. Its objective is to predict user future requests before hand and produce short term view of potentially useful links. The links of recommended Web pages are inserted in the last requested Web page by the active user. For this we follow below mentioned steps.

Step 1) Preprocessing of user active session as well as navigation patterns

Step 2) Classification of active session by LCS algorithm

Step 3) Create and recommend a set of Web pages.

Step 1: Clusters of navigation patterns as well as user active session both are preprocessed in this step. Active user may visit the web pages in any random order. For example, Active User Session (AUS) = (P9, P2, P1, P6, P11, P5).Preprocessing of such AUS aims at Rearrangement of Web Pages in ascending order. i. e. AUS = (P1, P2, P5, P6, P9, P11).We use few pages of this active user session as an input to map it to one of the clusters. This is regarded as Input Window (IW). Size of IW is decided by average path length as found by our statistical parameter analysis. With our dataset it is 3 pages. Similarly Web pages inside a cluster of navigation patterns are also rearranged in ascending order. The preprocessing of both AUS as well as Web pages inside a cluster helps in better classification of active user session.

Step 2: This step classify the current active user session in one of the cluster of navigation pattern. It takes Input Window (IW) as an input, Compare it to all the clusters by Longest Common Subsequence (LCS) algorithm and returns the cluster with highest degree of similarity. Similarity by

$$\text{LCS} = \frac{(2 * |\text{LCS}(S1, S2)|)}{(|S1| + |S2|)} \tag{4}$$

where S1 and S2 are the two sequences of Input Window (IW) and Cluster being matched respectively. We get different similarity values for different clusters. The cluster with highest similarity value is output of LCS algorithm. The working of LCS is as shown below.

Table 1 Cluster mapping example

User Active Session Input Window	**(P5, P9, P11)**
Clusters of Navigation Patterns	
C1	(P1, P2, P10, P15, P20)
C2	(P5, P25, P27, P29, P35, P38)
C3	(P9, P11, P17, P19, P23, P31, P47)
C4	(P45, P50, P52)
Output based on LCS => C3	

We get C3 as the matching cluster with highest degree of similarity in this example.

Step 3: This step create a set of Web pages as a recommendation to the active user. Initially all remaining pages of the matched cluster form the initial recommendation set. We do ranking of all these Web pages according to the relationship matrix generated in offline phase to generate Intermediate Recommendation Set. Finally, we set a threshold value for the Web pages. Only the Web pages with greater than threshold value will be recommended to the active user. Application of threshold to final recommendation set results in recommendation of optimum number of Web pages and thus greatly affects Accuracy of our system.

Table 2. Recommendation example

Initial Recommendation Set	Cluster Pages – Input Window (P9, P11, P17, P19, P23, P31, P47) **-** (P5, P9, P11) (P17, P19, P23, P31, P47)
Intermediate Recommendation Set	Ranking of Initial recommendation set based on Relationship Matrix (P17, P23, P47, P19, P31)
Applying Threshold to get Final Recommendation Set	(P17,P23,P47)
The final recommendation set of Web pages will be inserted into the last requested web page of the active user. In this example, it is (P17, P23 and P47).	

The Process of recommendation is repeated for next Input Window (IW). Many techniques are used to evaluate the web personalization and recommender systems. The proposed web personalization system is evaluated based on following parameters utilized both in off line phase and online phase.

5. SYSTEM EVALUATION AND EXPERIMENTAL RESULTS

Many techniques are used to evaluate the web personalization and recommender systems. The proposed web personalization system is evaluated based on following parameters utilized both in off line phase and online phase.

Visit Coherence: The quality of clusters produced during offline phase is evaluated by visit-coherence parameter. It measures the percentage of Web pages inside a user session, which belongs to the cluster that represents the session being considered [19].

We performed the evaluation of visit coherence by splitting the dataset into two halves: training and testing. The clustering process is applied on first half of the dataset whereas second half of dataset is used to test the quality of clusters produced from training part. Parameter β is defined to measure the number of Web pages in every session i that belongs to a navigation pattern (cluster) found for that session [19].

$$\beta_i = \frac{|\{p \in Si \mid p \in ci\}|}{N_i} \tag{5}$$

Where p is a page, Si is the ith session, Ci is the cluster representing session i, and Ni is the number of pages in the ith session. The average value for β over all N sessions in the evaluation part of dataset is shown as:

$$\alpha = \frac{\sum_{i=1}^{n} \beta_i}{N_S} \tag{6}$$

Where α is the percentage of the visit-coherence that should be considered for various range of Edge Threshold [19].

Outliers: The Outlier is a percentage of Web pages that does not belong to any navigation pattern (cluster), therefore does not contribute to the online phase. Outliers are calculated based on different values of Edge Threshold. We have tried to minimize the outliers. More the outliers means less efficient is the clustering process and vice versa.

Accuracy: We measure the performance of recommendations by Accuracy parameter. It means number of relevant Web pages actually retrieved by user divided by total number of Web pages in recommendation set. Initially we divide the dataset in two parts: training set and testing set. Each navigation pattern npi (a session in the dataset) in the testing set is divided in two parts. The first n page views in npi are used as input to produce recommendation set of Web pages. The remaining part of npi is used to evaluate generated recommendation set.

$$\text{Accuracy}\{\text{Rec. Set}_{(AS,ET)})\} = \frac{|\,\text{Rec. Set}_{(AS,ET)} \cap \text{Test Set}_{(np-n)}\,|}{\text{Rec. Set}_{(AS,ET)}} \tag{7}$$

Where Rec. Set (AS, ET) is the generated Recommendation set with respect to Active Session window(AS) and Edge Threshold(ET) . Test Setnp-n is the part of test set session by discarding first n page views. |Rec. Set (AS, ET) ∩ Test Setnp-n| is the common Web pages in both recommendation set and test set.

Coverage: Another evaluation parameter for recommendation set is Coverage. It means number of relevant Web pages actually retrieved by user divided by the total number of pages present in the user session. It measures ability to produce all the page views which can be visited by user.

$$\text{Coverage}\{\text{Rec. Set}_{(AS,ET)})\} = \frac{|\,\text{Rec. Set}_{(AS,ET)} \cap \text{Test Set}_{(np-n)}\,|}{\text{Test Set}_{(np-n)}} \tag{8}$$

F1: F1 is a Harmonic Mean of Accuracy and Coverage. It achieves maximum value when both Accuracy and Coverage achieve maximum values. It is given by

$$\text{F1} = \frac{2 * \text{Accuracy}\{\text{Rec. Set}_{(AS,ET)}\} * \text{Coverage}\{\text{Rec. Set}_{(AS,ET)}\}}{\text{Accuracy}\{\text{Rec. Set}_{(AS,ET)}\} + \text{Coverage}\{\text{Rec. Set}_{(AS,ET)}\}} \tag{9}$$

Experimental Results and Evaluation:

To evaluate the proposed system, we used following specifications of operating environment and dataset. **Dataset:**

We have implemented our work on DePaul University CTI log file dataset (www.cs.depaul.edu). This data set contains the data for the main DePaul CTI Web server (http://www.cs.depaul.edu). The data is based on a random sample of users visiting this site for a 2 week period during April of 2002. The original (unfiltered) data contained a total of 20950 sessions from 5446 users. The filtered data files are produced by filtering low support page views, and eliminating sessions of size 1. The filtered data contains 13745 sessions and 683 page views.

Two experiments are conducted using above mentioned implementation specifications. In the first experiment, after modeling of Web pages clustering of Web pages is done using DFS algorithm. In the second experiment, user active session is classified in one of the cluster using LCS algorithm.

Table 3. Experimental Evaluation

Edge Threshold Value	No. of Clusters	No. of Outlier Web Pages	Percentage of Outliers	Visit Coherence (Average Value of Alpha)	Visit Coherence (Percentage Value of Alpha)	Accuracy Value	Coverage (Average Value)	Coverage (Percentage Value)	F1
0	01	00	00	01	100	16.33	0.80	80.40	27.14
0.1	03	07	1.02	0.99	99.28	18.42	0.77	77.01	29.72
0.2	22	90	13.17	0.91	91.49	24.78	0.72	72.48	36.93
0.3	65	184	26.93	0.45	45.27	34.12	0.35	35.19	34.64
0.4	81	303	44.36	0.42	42.32	44.61	0.30	30.37	36.13
0.5	80	390	57.10	0.41	41.27	55.70	0.28	28.47	37.68
0.6	77	462	67.64	0.27	26.99	61.42	0.34	34.56	44.23
0.7	65	528	77.30	0.23	23.16	52.86	0.30	30.10	38.35
0.8	54	567	83.01	0.15	15.36	39.47	0.28	28.22	32.91
0.9	36	606	88.72	0.10	9.90	20.32	0.29	29.16	23.95
1	00	683	100	00	00	00	00	00	00

6. CONCLUSION AND FUTURE WORK

We have developed a new web personalization technique. The proposed technique is based on finding more appropriate weights among the web pages of a web site. The modeling among the Web pages is done by a novel formula measuring the distance relationship as well as occurrence frequency. The enhanced clustering done on this relationship matrix helped us to form more appropriate clusters. We classified the active users using LCS. The Threshold used by us in the last phase of recommendations improves the accuracy of our system. However it affects in lowered coverage. Semantic knowledge about the underlying domain may improve quality of recommendations further.

ACKNOWLEDGEMENT

We are thankful to North Maharashtra University Jalgaon for providing assistance and support for successful completion of project under Vice Chancellor Research Motivation Scheme (VCRMS).

REFERENCES

[1] M. Eirinaki and M. Vazirgiannis, “Web Mining for Web Personalization”, in ACM Transactions on Internet Technology (TOIT), Vol.3, No.1, pp.1-38, 2003.

[2] G. Adomavicius and A. Tuzhilin, “Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions”, in IEEE Transactions on Knowledge and Data Engineering, Vol.17, No.6, pp.734–749, 2005..

[3] R. B. Wagh and J. B. Patil, “Web Personalization and Recommender Systems: An Overview”, in proceedings of the 18th International Conference on Management of Data (COMAD), pp.114, 2012.

[4] Z. Malik and C. Fyfe, “Review of Web Personalization”, in Journal of Engineering Technologies in Web Intelligence, Vol. 4, No. 3, 2012.

[5] M. Jalali, N. Mustapha, N. Sulaiman and A. Mamat, “WebPUM: A Web-based Recommendation System to Predict User Future Movements”, in Expert Systems Applications, Vol.37, pp. 6201-6212, 2010.

[6] H. Liu and V. Keselj, “Combined Mining of Web Server Logs and Web Contents for Classifying User Navigation Patterns and Predicting Users’ Future Request”, in Data and Knowledge Engineering, Vol.61(2), pp.304-330,2007.

[7] Q. Yang, J. Fan, J. Wang, and L. Zhou, “Personalizing Web Page Recommendation via Collaborative Filtering and Topic-Aware Markov Model”, in IEEE International Conference on Data Mining, Vol. 1, pp. 1145-1150, 2010.

[8] Y. AlMurtadha, N. Sulaiman, N. Mustapha and N. Udzir, "IPACT: Improved Web Page Recommendation System Using Profile Aggregation Based On Clustering of Transactions", in American Journal of Applied Sciences, Vol. 8, No.3, pp. 277-283, 2011.

[9] H. Ramadhan, M. Hatem, Z. Al-Khanjri, "A Classification of Techniques for Web Usage Analysis", in Journal of Computer Science, Vol.1, pp. 413-418, 2005.

[10] S. Vijayalakshmi and V. Mohan, "Mining Sequential Access Pattern with Low Support from Large Pre-processed Web Logs", in Journal of Computer Science, Vol. 6, pp. 1293-1300, 2010.

[11] J. Ren and X. Zhou, "A New Incremental Updating Algorithm for Mining Sequential Patterns", in Journal of Computer Science, Vol.2, pp.318-321, 2006.

[12] C. Sumathi, R. Valli and T. Santahnam, "An Application of Session Based Clustering to Analyze Web Pages of User Interest from Web Log Files", in Journal of Computer Science, Vol.6, pp.785-793, 2010.

[13] C. Dimopoulos, C. Makris, Y. Panagis, E. Theodoridis and A. Tsakalidis, "A Web Page Usage Prediction Scheme using Sequence Indexing and Clustering Techniques", in Data Knowledge Engineering, Vol.69, pp.371-382, 2010.

[14] R. Baraglia and F. Silvestri, "Dynamic Personalization of Web Sites without User Intervention",in Communications of the ACM,Vol.50,No.2 ,pp.63-67,2007.

[15] R. Baraglia and F. Silverstri, "An Online Recommender System for Large Web Sites", in IEEE/ACM/WIC International Conference on Web, Beijing, China, 2004.

[16] B. Mobasher, R. Colley, J. Shrivastav, "Automatic Personalization based on Web Usage Mining", in Communications of the ACM, Vol. 43, No. 8, pp. 142-151,2000.

[17] M. Nakagawa and B. Mobasher, "A Hybrid Web Personalization Model based on Site Connectivity", in the fifth WEBKDDD workshop, pp. 59-70, at Washington DC, USA, 2003.

[18] I. Cadez, D. Hackerman, C. Meek, P. Smyth, S. White, " Visualization of Navigation Patterns on a Web Site using Model-based Clustering" , in sixth ACM SIGKDD International Conference on Data Mining and Knowledge Discovery, Boston, Massachusetts, United States,2000.

[19] M. Perkowitz and O.Etzioni, "Towards Adaptive Web Sites: Conceptual Framework and Case Study", in Artificial Intelligence, Vol. 118(1-2), pp. 245-275, 2000.

[20] B. Zhou, S. Hui, K. Chang, "An Intelligent Recommender System using Sequentail Web Access Patterns", in IEEE International Conference on Cybernetics and Intelligent Systems, at Singapore, pp. 393-398, 2004.