ISSN: 0976-3104
SPECIAL ISSUE: Emerging trends in Computer Engineering and Research (ECER)

ARTICLE

AN IMPROVED MECHANISM FOR USER PROFILING AND RECOMMENDATION USING CASE-BASED REASONING

Bhavithra Janakiraman[1*], Saradha Arumugam[2], Aiswarya Jayaprakash[1]

[1]Department of Computer Science and Engineering, Dr. Mahalingam College of Engineering and Technology, Pollachi, Tamilnadu, INDIA

[2]Department of Computer Science and Engineering, Institute of Road and Transport Technology, Erode, Tamilnadu, INDIA

ABSTRACT

Background: *Web Page recommendation system is the process of identifying suitable webpages that matches with user query and recommending them for further access. Today most of the research works have been focusing on applying machine learning techniques for improving the accuracy of recommendation process. Such algorithms predict suitable webpages by analyzing the past search experience of the currently active user. These past search history will be stored in the form of user profiles.* ***Methods:*** *This paper proposes an improved mechanism for developing user profiles using selective usage-based attributes. Attributes such as Page Rank, Page Weight, Bounce Rate, Exit Rate and Conversion Rate of each webpage visited by the corresponding user will be computed and stored in the profile of corresponding user. Case-Based Reasoning is further applied for clustering the user profiles based on similar search interest and generation of cluster summary. The subsequent webpages suitable for the currently active user is predicted based on cluster summary.* ***Results:*** *Experiments were conducted on eight categories of datasets comprising of benchmark and real-time web log data. The performance was analyzed in terms of accuracy and mean absolute error rate.* ***Conclusions:*** *Results expresses that, the proposed method of user profiling and clustering outperforms existing algorithms with improved accuracy and lower error rate.*

KEY WORDS
User Profile, Case-Based Reasoning, Bounce Rate, Exit Rate, Conversion Rate

INTRODUCTION

Current information era tries to engage plenty of online users by providing appropriate search results based on user query. Unfortunately, such results do not satisfy all the users of various categories. Information is to be retrieved from the web based on individual user needs. In other words, query search results must be customized accordingly to provide user satisfaction. Recommendation system is the thriving research area today, in which personalization is done to analyze user's search interest and provide better results even for those users who do not reveal their search interest explicitly [1]. Collaborative filtering approach provides recommendations based on other users who have similar interest and preferences. To support personalization in recommendation [2, 3, 4], it is necessary to record user's search history [5,6]. Web users may provide their search interest input either explicitly or implicitly [7]. Explicit input includes showing their interest by providing user-rating or subscription for articles and messages in the corresponding web sites. Implicit input are those that were given indirectly. Hence intelligent algorithms are to be executed to achieve those implicit inputs.

Web log mining is basically divided into three major categories [8] namely, Web Usage Mining, Web Structure Mining and Web Content Mining. Web Usage Mining focus on analyzing the URLs (web pages) visited by a normal user. In terms of analyzing the usage of a web page, each user's personal interest of that web page and overall search intention can be predicted. This type of mining plays a vital role in predicting the implicit interest of a user in web page recommendation system. Web structure mining is the process of analyzing the hyperlink structure within a web page. Web content mining is another predominant area of research where the actual search topic of the user is analyzed. Content based similarity should be considered with more weightage for any effective recommendation system.

Received: 12 May 2017
Accepted: 12 July 2017
Published: 21 Sept 2017

This research mainly targets on analysis of web usage log and the content log of frequent web surfers. User profile is constructed based on URLs visited by users from various IP addresses. This paper proposes an improved mechanism to generate user profile using such implicit inputs. The User profile will be created for individual users using attributes with respect to web usages such as Page Rank (PR), Page Weight (PW), Bounce Rate (BR), Exit Rate (ER) and Conversion Rate (CR). Also with these attributes, web content based attribute such as the frequency of keywords present in each of these corresponding pages will also be recorded in the user profile.

The following are some of the contributions of the proposed paper:

- An improved mechanism has been suggested for profiling user's search interest using Usage-Based attributes such as PR, PW, BR, ER and CR along with Content-Based attributes such as most frequent keywords.
- Clustering these user profiles initially using k-means and followed incrementally through Case-Based Reasoning (CBR). Incremental clustering approach is followed in order to dynamically cluster the users based on varying search interest and pattern. CBR is adapted suitably to generate a summary profile for each cluster [9]. This summary comprises of the mean values of Usage-based and content-based attributes.
- Finally, the recommendation system obtains the user query and recommends the web pages only based on matching cluster summary. Thus optimization is achieved with good accuracy and better response time.

***Corresponding Author**
Email: bavi.rr@gmail.com
Tel.: 9443170753

THE IIOAB JOURNAL
COMPUTER SCIENCE: Guest editor-Prof. S. Prabu & Prof. Swarnalatha P

The rest of this paper is organized as follows. In section II, related work to this paper has been discussed. Section III discusses about the User profiling mechanism. In Section IV, Case Based Reasoning is applied to effectively cluster the users based on profile attributes which produce cluster summary. Section V discusses about how web pages are recommended through matching cluster summaries. Section VI covers results and discussion. In section VII, final conclusion and future work is discussed

RELATED WORK

Predictions of web pages that could likely be visited by end users are recommended based on the user interest and previous navigation history. Various traditional methods such as collaborative filtering, association rules, clustering, sequential patterns, hybrid methods and semantic web [10, 11] are used for such predictions. Collaborative filtering [1] is one of the most common approaches used for providing recommendation by finding similar users. Pearson correlation coefficient and cosine based approach can be used to find similar users [12]. This traditional approach can still be improved by applying normal recovery Collaborative filtering [13]. But recommendation done using pure collaborative filtering approach may lead to problems such as popularity bias, cold start problem, handling dynamic pages etc. Case Based Reasoning generates the user profile and uses similarity knowledge to predict relevant profiles for the currently active user [14]. Such profile includes Page Rank [15] as a major feature which is computed using HITS and Page Rank algorithm [10].

Collaborative Filtering

Collaborative filtering is one of the most common approaches used for recommendation. Collaborative Filtering systems collect visitor opinions on a set of objects using ratings, explicitly provided by the users or implicitly computed. In explicit ratings, users assign ratings to items or web pages, or a positive (or negative) vote to some web pages or documents [10]. The implicit ratings are computed by considering the access to a Web page. A rating matrix is constructed where each row represents a user and each column represents an item or web page keywords [13]. Items could be any type of online information resources or products in an online community such as web pages, videos, music tracks, photos, academic papers, books etc. Collaborative filtering systems predict a particular user's interest in an item using the rating matrix. Alternatively, the item-item matrix, which contains the pair-wise similarities of items, can be used as the rating matrix. Rating matrix is the basis of CF methods. The ratings collected by the system may be of both implicit and explicit forms. Although CF techniques based on implicit rating are available for recommendation, most of the CF approaches are developed for recommending items where users can provide their preferences with explicit ratings to items

Content-based filtering

Content-based filtering is a type of information extraction system, where web pages are extracted based on the semantic similarity between the content in those web pages visited by users in past history [16, 17]. Web content mining applications mostly rely on content-based filtering approaches. Content-based filtering offers predominant support for web page recommendation system. In this technique, the keywords and its frequency of occurrence in those web pages that were previously visited are collected. Then, the semantic similarity between such keywords will be analyzed for further process. For example, consider two users "u1" and "u2" who frequently visit web pages based on their domain of interest [18]. Let u1 always focus on heath related web pages and u2 focus on gadget-related sites. Now, during the real time if any active academic user "ua" search for the query "apple", he will be mostly related to apple devices based sites, rather than apple fruit. So, he will be recommended the sites referred by u2. Similarly, when a dietician "ub" searched for "apple" he will be recommended the sites referred by u1. Recommendation engine classifies "ua" as an academic user and "ub" as a dietician based on the contents (keywords) of the web pages navigated in past history.

Case-based reasoning

Case Based Reasoning [14, 19, 20] is a process of finding solutions to new problems based on the solutions of similar past problems. This approach of refining new solutions can be considered as a clustering problem, where based on previous clustering structure and defined attributes, any new case can generally be grouped to any one of the previous structure. As the analysis of the new case is totally based on the previous cases and experiences, this approach generally provides good accuracy and cluster purity. Today many researchers work in the area of applying case-based reasoning in web mining concepts.

USER PROFILE GENERATION

This paper proposes an improved mechanism for gathering user's implicit search interest based on his previous search pattern and history [2, 21]. The following attributes are used for recording the implicit interest of any user by analyzing URLs visited by the corresponding user:

- Page Rank (PR)
- Page Weight (PW)
- Bounce Rate (BR)
- Exit Rate (ER)

- Conversion Rate (CR)
- Keywords (KEY)

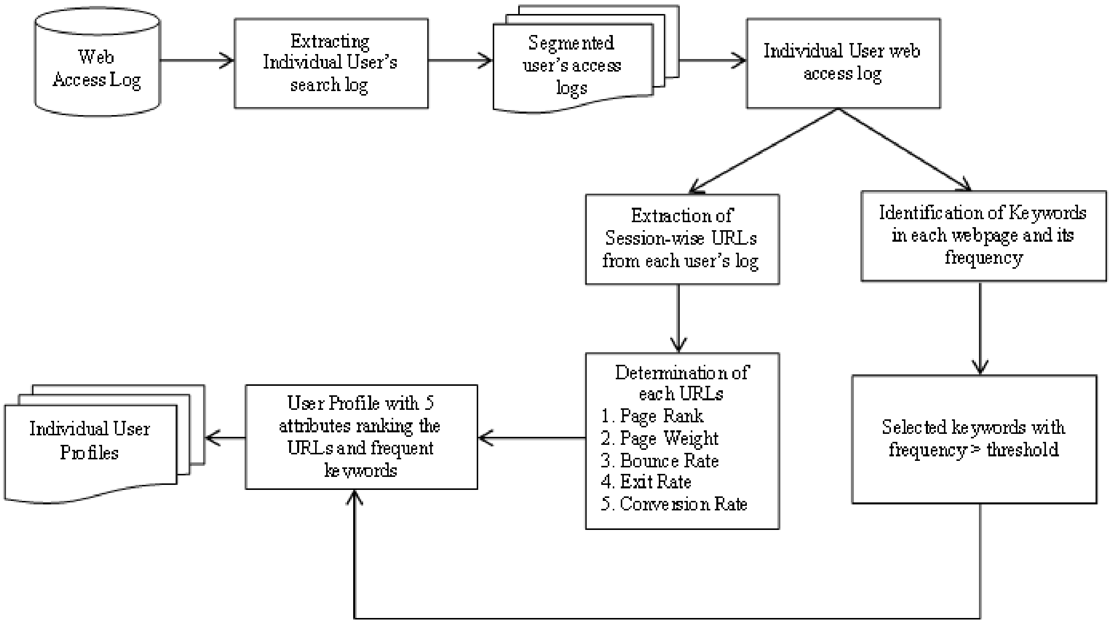

Fig. 1: Overall architecture of user profile generation from web access log dataset.

The overall architecture of User-Profile generation system is shown in [Fig. 1]. The following paragraph discusses about the procedures used for user profile generation.

Page rank (PR)

Page Rank is a numerical value that measure's a webpage importance among the group of similar web pages [22]. Such page rank is computed based on Random Surfer model [23]. This algorithm computes the page rank based on the link structure of the web page [23]. A page gets hold of high rank if the addition of the ranks of its backlinks is high. The rank of the given page is thus computed using the following equation (1)

$$PR = \frac{1}{N}\left[(1-d) + d \sum_{v \in B(u)} Page_{Wt} \times \frac{PR(v)}{N_v}\right] \quad (1)$$

Where, u represents a web page. B(u) is the set of pages that point to u. PR(v) is the page rank of page v that points to page u. Nv is the number of outgoing links of page and d is the damping factor that is set between 0 and 1. The damping factor is the decay factor that represents the chance of a user stop clicking links within a current page and then requesting another random page [22]. $Page_{Wt}$ is termed as Page weight which is calculated based on frequency and duration as in equation (2).

Page weight (PW)

Page weight is calculated based on frequency and duration as in Equation (2).

$$Page_{Wt}(PW) = NV(pi) \times \tau pi \quad (2)$$

Where τpi is total time spent by the user on particular web page. A quick jump might also occur due to the short length of a web page so the size of page may affect the actual visiting time. Hence, duration is normalized by the length of the web page, i.e. the total bytes of the page. NV(pi) is the number of times that a page is accessed by different users; computed by equation (3).

$$NV(pi) = \sum_{j=1}^{N} \sum_{pi \in sj} n \quad (3)$$

Where,

IIOAB JOURNAL
ISSN: 0976-3104

$$n = \begin{cases} 0, & \text{if pi is not present atleast once in Session Sj} \\ 1, & \text{if pi is present atleast once in Session Sj} \end{cases}$$

Bounce rate (BR)

The web page access percentage with respect to session wise grouping of access pattern is called as bounce rate. Today BR plays a vital role in web analytics. Web pages access pattern is grouped into sessions based on date and time difference between two consecutive page requests. If the date and time difference is exceeding the certain time limit of 10 minutes, we have grouped the access patterns as clusters called as sessions. The page pi's access rate between all such sessions is computed as BR using the equation (4)

$$BR(pi) = \frac{\sum_{s \in S} \sum_{(pi \in s) \cap (i=1)} \tau pi}{NS(pi)} \times TS \qquad (4)$$

Where 's' represents each session from the complete set of sessions 'S' and TS represents the total number of sessions active by a web user. NS(pi) denotes the total number of sessions where page pi has been accessed.

Exit rate (ER)

The rate at which, the web page (pi) will be at the end of the session is computed as ER. Here, the occurrence of pi being the last entry within the session is calculated to identify the exit rate using the equation (5)

$$ER(pi) = \frac{\sum_{s \in S} \sum_{(pi \in s) \cap (i=N)} \tau pi}{NS(pi)} \times TS \qquad (5)$$

Conversion Rate (CR)

The conversion rate for each web page is computed as the ratio between total sessions accessed by a user to the total number of sessions that contains the page pi. Equation (6) computes the conversion rate of page pi.

$$CR(pi) = \frac{TS}{NS(pi)} \times 100 \qquad (6)$$

Keywords (KEY)

The set of keywords present in all the visited web pages is collected together as a local dictionary for that corresponding user. The set of identical words after performing tokenization, removal of stop words and stemming is generally called as Keywords. A matrix as shown in [Table 1] is developed after keywords identification. The matrix is populated with term frequency which corresponds to the number of times a keyword (term) occurs within a particular web page (denoted by its URL). Pruning operation is also performed through the elimination of those keywords that are below the certain threshold level. The threshold level is computed using equation (7). Finally the top "n" keywords from "k" web pages are selected for further clustering process.

$$Thershold = \frac{1}{k} \sum_{i=1}^{k} Term\_Freq(KEYi) \qquad (7)$$

Table 1: Keyword matrix comprising of frequency of keywords in local dictionary of top k webpages

URLs	KEY1	KEY2	KEY3	KEY4	…	KEYN
URL 1	12	5	13	16	…	7
URL 2	4	17	16	7		14
URL 3	17	14	12	2		1
URL 4	11	15	8	18		12
………………………………						
URL K	10	13	5	1	…	12

CLUSTERING USING CASE-BASED REASONING

Case-based reasoning is then applied for effective clustering of user profiles based on the similarity in their search and access pattern. For example, users those who are always searching for product based web pages will be clustered together. Similarly, users those who search for "Web Mining" based research articles will be united into a single cluster. Such similarities among the search histories were identified using keyword match. The following algorithm (1) applies k-means method of clustering for initially segmenting the users. The initial set of clustering will result in a matrix as shown in [Table 2]. In order to incrementally cluster/ re-cluster the active members, case-based reasoning is used. The final cluster summary generated using CBR approach is shown in [Table 3].

IIOAB JOURNAL ISSN: 0976-3104

Algorithm 1: Applying Case-Based Reasoning for clustering user profiles

Input: User profiles with Usage-based features such as PR,PW,BR,ER,CR and Keywords(1..N)
Output: Cluster summary comprising of keywords(1..N) and Top URLs for further recommendation
Initialize: N← Max_KEY; K←No_of_UserProfiles; M← No_of_Clusters; c[1..M]← No_of_users in cluster

Begin
1. Retrieve all user profiles and store as Hash Table
   1.1. UP_KEY[i] ← [K1i,K2i,K3i, … ,KNi]
   1.2. UP_URL[i] ← [PRi,PWi,BRi,ERi,CRi]
2. Initially cluster user profiles based on keywords using k-means algorithm
   2.1. Randomly assign 'K' users into 'M' clusters; Such that each cluster contains 'c' users (where c ← K/M)
   2.2. For all clusters c = 1 to M, identify the mean of term_frequency for each keyword KEY = 1 to N
   2.3. Compare individual user profile's KEY[1..N] term_frequency with KEY_MEAN[1..N] of that corresponding cluster
   2.4. Move the user profile which has the nearest match between its KEY[1..N] term_frequency with KEY_MEAN[1..N]
   2.5. Repeat from step 2.2 until there are no more moves
   2.6. Return initial level of clusters.
3. Apply Case_Based Reasoning for further clustering which includes updated/new user profiles
   3.1. Identify the mean value for all usage-based features resulting in Avg(PR), Avg(PW), Avg(BR), Avg(ER), Avg(CR)
   3.2. Compare the Keyword similarity of updated/new user profile with cluster's mean.
   3.3. Identify top 'T' clusters that nearly matches with updated/new user profile
   3.4. Compare the updated/new user profile's usage-features with those 'T' clusters
   3.5. Move the updated/new user profile to the corresponding cluster with the best match
4. Identification of Cluster Summary
   4.1. For each cluster c from 1 to M repeat the following
   4.2. Sort the cluster members based on usage-based features such as PR,PW,BR,ER,CR
   4.3. List the top 'S' User profile's Keywords (1..N) and their URL's as the corresponding cluster's summary.
   4.4. Return the summary of all clusters c[1..M].

End

Table 2: User profile matrix used for clustering using case-based reasoning (an example)

User_ID	PR	PW	BR	ER	CR	KEY1	…	KEYN	Cluster_ID
User 1	3	32	58	45	4.67	12		9	3
User 2	2	21	41	38	2.45	44		23	1
User 3	5	14	68	23	7.54	17		6	2
User 4	17	12	32	18	3.56	11		3	3
User 5	1	37	67	34	9.87	16		13	2
User 6	8	45	58	24	1.34	11	…	13	3
User 7	1	32	32	12	3.67	34		15	1
User 8	4	11	18	37	5.24	45		32	1
User 9	2	21	44	43	6.13	23		12	2
User 10	3	37	24	23	8.44	22		9	4
…………………………………									
User K	16	22	54	32	3.59	19	…	7	4

Table 3: Cluster summary generated using CBR algorithm (cluster comprising of book searchers profiles)

KEY1	…	KEYN	Top URLs visited by users in the corresponding cluster
44		23	www.goodreads.com,www.amazon.com, www.infibeam.com
34		15	www.openingthebook.com
17		13	www.quora.com, www.bookadventure.com
16	…	13	www.educatorstechnology.com
12		9	www.readingrockets.org
12		6	www.magicbox.com
11		2	www.kidsreads.com

RECOMMENDATION PROCESS

This is the only online activity done by web servers as soon as end user provides search query. The cluster summary generated using CBR approach will be used for further recommendation. As the user provides the query through online, information retrieval and further recommendation must be obtained with quick response time. Hence recommendation algorithms should concentrate both on accuracy and as well as good response time. The recommendation process of the proposed system uses cluster summary in order to efficiently predict the next web page that will be possibility expected by the currently active user. The overall recommendation process is given in [Fig. 2].

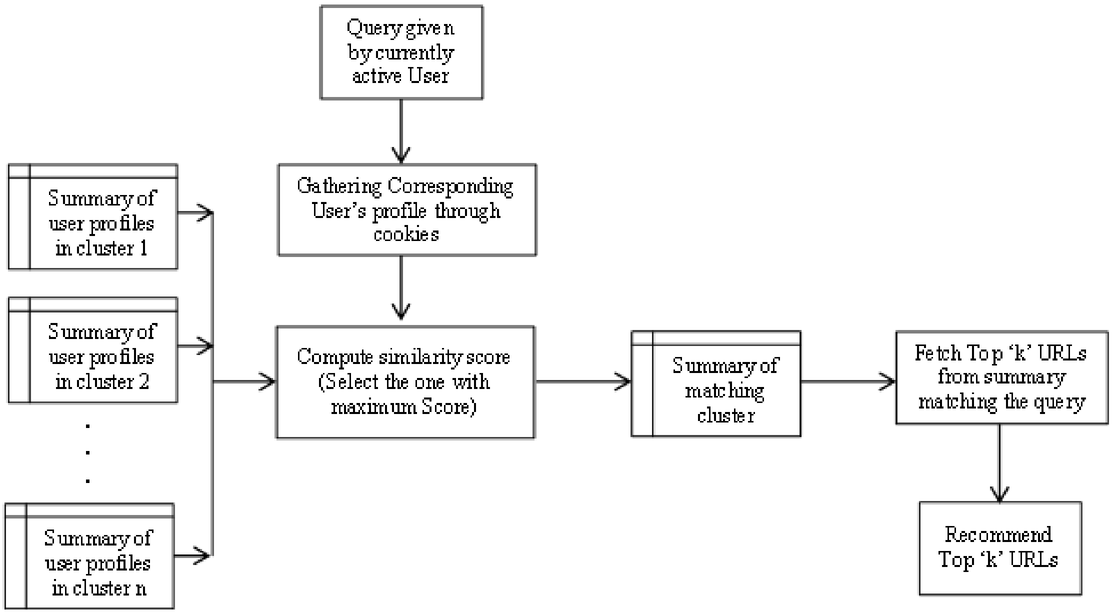

Fig. 2: Recommendation of web pages based on similarity between active user and cluster summary

Recommendation system fetches the profile of the corresponding user who provided the search query. These user profiles can be either saved within web servers, search engines or even at the client side itself. If the profile is stored within the client end, cookies are used for fetching the profiles. One of the issues faced here would be that if end user deletes the cookie files, recommendation accuracy may be reduced. After successfully retrieving the profile, the cluster summary of the corresponding cluster where the user is intended to be grouped is identified. From the list of URLs stored in the summary, top 'k' URLs will be recommended for the currently active user. If the recommended web pages are selected by the end users, the page's PR, PW and BR will automatically raise favoring the corresponding web page to be given with high weightage to be considered for the next time. Also, if the user does not visit a page, its view, weight and bounce rate will automatically diminish again favoring the recommendation process by lowering the selection probability of low-rated web pages to be recommended.

RESULTS

Experiments were conducted using eight categories of datasets including benchmark web server repository called AOL web access log dataset and real time web access log from Dr.Mahalingam College of Engineering and Technology (MCET), Pollachi, Tamilnadu, India. The web access log dataset is divided into eight samples of equal size with 50 records as mentioned in [Table 4]. The descriptions of these datasets have been discussed below:

AOL Web Access Log (AOL)

AOL web access log dataset [24] was used. The log file contains web query log data from ~650k users. In order to have privacy preservation, IP addresses of individual users were anonymized. Hence each user is represented by unique ID. The schema of this log dataset is: {AnonID, Query, Query Time, Item Rank, ClickURL}. Where, AnonID denotes an anonymous user ID number; Query denotes to the query issued by the user; Query Time refers the time at which the query was submitted for search; Item Rank states that if the user clicked on a search result, the rank of the item on which they clicked is listed; Click URL states that if the user clicked on a search result, the domain portion of the URL in the clicked result is listed.

MCET Web Access Log (MCET)

The log file was collected from students browsing history, given by Dr.Mahalingam College of Engineering & Technology, Pollachi, Tamilnadu, India. The size of the dataset used was about 23,709 KB with 1 lakh entries, consists of 77 different IP addresses, date & time of visiting the web pages, method URL/protocol, status, received byte and connection type. Each webpage may include advertisements, pictures, videos, textual content etc. The banned and invalid URLs are ignored during web content extraction. The preprocessed log will contain IP Address, date & time and URL.

IIOAB JOURNAL ISSN: 0976-3104

Table 4: Various sample datasets used for experimentation process

Data Set Title	Description
AOL_SET 1	The top 50 users who access web frequently were selected based on the maximum length (no. of URLs) within each session.
AOL_SET 2	The top 50 users who do not access web frequently were selected based on the minimum length (no. of URLs) within each session.
AOL_SET 3	The top 50 users having profile with maximum number of identical search keywords were selected.
AOL_SET 4	The top 50 users having profile with minimum number of identical search keywords were selected.
MCET_SET 1	Without any conditions, access log of 50 users were selected randomly
MCET_SET 2	Uniform sampling was performed to select one user after each 50 records.
MCET_SET 3	The query was analysed and categorized into various domains. 50 users accessed under academic category were selected.
MCET_SET 4	The query was analysed and categorized into various domains. 50 users accessed under entertainment category were selected.

The model is trained using both AOL dataset and real time MCET dataset covering the samples excluding the eight sets mentioned in [Table 4]. Testing experiments were conducted with traditional Collaborative Filtering (CF) algorithm and the proposed User Profiling and Recommendation using CBR (UPR-CBR) approach using these eight samples.

Evaluation metrics

The performance of existing (CF) and proposed (UPR-CBR) systems are evaluated using Precision, Recall, F-Measure [16] and Mean Absolute Error (MAE) [13]. Let 'R' denote the total number of web pages in the collection, 'A' denotes set of web pages that are answered and 'Ra' denotes Set of relevant web pages that are retrieved. Precision is calculated as shown in equation (8) which is defined as the ratio of a number of relevant web pages retrieved to the total number of answered web pages. Recall is defined as the ratio of a number of relevant web pages retrieved to the total number of web pages. It is expressed as a percentage and calculated as shown in Equation (9). F-Measure is the harmonic mean of precision and recall, the F-measure or balanced F-score is calculated as shown in equation (10). Mean Absolute error (MAE) is the average absolute deviations of predictions to the ground truth values. It measures the deviation of actual value and predicted value using the equation (11)

$$\text{precision} = \frac{|R_a|}{|A|} \quad (8)$$

$$\text{Recall} = \frac{|R_a|}{|R|} \quad (9)$$

$$\text{Harmonic Mean} = \frac{2}{\frac{1}{r} + \frac{1}{p}} \quad (10)$$

$$\text{MAE} = \frac{\sum_{u,i} |r_{u,i} - \hat{r}_{u,i}|}{N} \quad (11)$$

Where,

- r denotes recall
- p denotes precision
- $r_{u,i}$ denotes actual accuracy of webpage i observed by the user u
- $\hat{r}_{u,i}$ denotes the predicted value of webpage I for user u
- N denotes the total number of predicted value

The following [Table 5] compares the performance of Collaborative Filtering (CF) algorithm and User-Profiling based Recommendation using Case-Based Reasoning (UPR-CBR) approaches in terms of accuracy and error rate as defined in equations (11) and (12) respectively. It has been found that in all the samples of datasets both in benchmark access logs and real time MCET access logs, the proposed system outperforms existing collaborative recommendation. The error rate is also found to be low for UPR-CBR in all datasets. Multiple experiments were conducted with varying cluster values (M). An average F1-measure of all experiments under the same sample set is summarized in [Table 5].

Table 5: Performance analysis of algorithms in terms of accuracy and error rate

THE IIOAB JOURNAL

COMPUTER SCIENCE: Guest editor-Prof. S. Prabu & Prof. Swarnalatha P

IIOAB JOURNAL ISSN: 0976-3104

Training Dataset	Accuracy (Harmonic Mean)		Mean Absolute Error (MAE)	
	UPR-CBR	CF	UPR-CBR	CF
AOL_SET1	70	56	0.41	0.54
AOL_SET2	66	61	0.47	0.52
AOL_SET3	68	57	0.47	0.58
AOL_SET4	68	56	0.46	0.58
MCET_SET1	64	60	0.48	0.53
MCET_SET2	71	63	0.46	0.55
MCET_SET3	71	66	0.43	0.47
MCET_SET4	76	73	0.36	0.40

CONCLUSION

This paper proposes an improved mechanism of profiling user's web pages of interest through their search history and navigation by using usage-based features such as Page Rank, Page Weight, Bounce Rate, Exit Rate and Conversion Rate. The frequency of keywords in the web pages visited by corresponding users was recorded to identify the similarity among user community for effective recommendation. Hence proposed system's user profiling mechanism includes both collaborative approach and content-based filtering systems. Case-based reasoning was then applied for clustering the user profiles among vast set of user community with similar web navigation history. CBR concludes the clustering by providing a cluster summary comprising of top level web pages with high usage-based features and content-based features. During online information retrieval as soon as the user provides the search query, the corresponding user profile is just compared with the matching cluster summary and web pages are recommended with quick response time. The usage-based and content-based attributes will be dynamically updated favoring relevant pages to be recommended with high probability and the probability of considering irrelevant pages will be reduced gradually based on user's click rate. Experiments were conducted using eight categories of datasets from benchmark and real time log entries. Results infer that the performance of proposed system is found to be improved in terms of accuracy and error rate.

CONFLICT OF INTEREST
There is no conflict of interest with any author regarding publication of this article

ACKNOWLEDGEMENTS
The authors of this paper express their gratitude to the respected management team and network administration team of Dr.Mahalingam College of Engineering and Technology, Pollachi, Tamilnadu, India for providing the web access log files of the college for this research in order to perform real time testing.

FINANCIAL DISCLOSURE
This research work was carried out without financial support from any organizations.

REFERENCES

[1] Zibin Zheng, Hao Ma, Michael Lyu R, Irwin King. [2009] Wsrec: A Collaborative Filtering Based Web Service Recommender System, IEEE International Conference on Web Services, 437-444.

[2] Sarabjot Singh Anand, Bamshad Mobasher. [2005] Intelligent Techniques for Web Personalization, Springer-Verlag Berlin Heidelberg, 1-36.

[3] Abdul Manan Ahmad, Mohd. Hanafi Ahmad Hijazi. [2004], Web Page Recommendation Model for Web Personalization, Springer-Verlag Berlin Heidelberg, 587–593.

[4] Magdalini Eirinaki, Charalampos Lampos, Stratos Paulakis, Michalis Vazirgiannis. [2004] Web Personalization Integrating Content Semantics and Navigational Patterns, WIDM'04, ACM.

[5] Zheng Lu, Hongyuan Zha, Xiaokang Yang, Weiyao Lin, Zhaohui Zheng, [2013] A New Algorithm for Inferring User Search Goals with Feedback Sessions, IEEE Transactions on Knowledge and Data Engineering, 25(3): 502-513.

[6] Dimitrios Pierrakos, Georgios Paliouras. [2010] Personalizing Web Directories with the AIDOF Web Usage Data, IEEE Transactions on Knowledge and Data Engineering, 22(9): 1331-1344.

[7] Şule Gunduz-Oguducu, Tamer Ozsu, M. [2006] Incremental Click-Stream Tree Model: Learning from New Users for Web Page Prediction, Distributed Parallel Databases, Springer Science, 19: 5–27.

[8] Şule Gunduz - Oguducu. [2010] Web Page Recommendation Models: Theory and Algorithms, Synthesis Lectures on Data Management, 2: 1-85.

[9] Yahya Al Murtadha, Md. Nasir Bin Sulaiman, Norwati Mustapha, Nurlzura Udzir. [2010] Improved Web Page Recommendation System using Profile Aggregation Based on Clustering of Transactions, American Journal of Applied Sciences, 8(3):277-283.

[10] Pooja Devi, Ashlesha Gupta, Ashutosh Dixit. [2014] Comparative Study of HITS and PageRank Link Based Ranking Algorithms, International Journal of Advanced Research in Computer and Communication Engineering, 3: 5749-5754.

[11] Murat Goksedef, Şule Gunduz-Oguducu. [2010], Combination of Web Page Recommender Systems, Elsevier Journal on Expert Systems with Applications, 2911–2922.

[12] Zibin Zheng, Hao Ma, Michael Lyu, R, Irwin Kin. [2011] QoS-aware Web Service Recommendation by Collaborative

Filtering, IEEE Transactions on Services Computing, 4:140-152

[13] Huifeng Sun, Zibin Zheng, Junliang Chen, Michael Lyu R. [2012] Personalized Web Service Recommendation via Normal Recovery Collaborative Filtering, IEEE Transactions on Services Computing, 6: 573-579.

[14] Yong-Bin Kang, Shonali Krishnaswamy, Arkady Zaslavsky. [2014], A Retrieval Strategy for Case-Based Reasoning Using Similarity and Association Knowledge, IEEE Transactions on Cybernetics, 44: 473- 487.

[15] Huifeng Sun, Yong Peng, Junliang Chen, Chuanchang Liu, Yuzhuo Sun. [2011] A New Similarity Measure Based on Adjusted Euclidean Distance for Memory-Based Collaborative Filtering, Journal of Software, 6:993-1000.

[16] Niki, R, Kapadia, Kinjal Patel. [2012] Web Content Mining Techniques – A Comprehensive Survey, International Journal of Research in Engineering & Applied Sciences, 2(2)1869, 1877.

[17] Lina Yao, Quan Z Sheng, Anne HH Ngu, Jian Yu, Aviv Segev. [2015] Unified Collaborative and Content-Based Web Service Recommendation, IEEE Transactions on Services Computing, 8(3): 453-466.

[18] Thi Thanh Sang Nguyen, Hai Yan Lu, Jie Lu. [2014] Web-Page Recommendation Based on Web Usage and Domain Knowledge, IEEE Transactions on Knowledge and Data Engineering, 26(10): 2574-2587.

[19] Daxa K. Patel. [2014], A Retrieval Strategy for Case-Based Reasoning using USIMSCAR for Hierarchical Case, International Journal of Advanced Engineering Research and Technology, 2(2): 65-69.

[20] Shweta Tyagi, Kamal, K Bharadwaj. [2012] A Hybrid Recommender System Using Rule-Based and Case-Based Reasoning, International Journal of Information and Electronics Engineering, 2(4): 586-590.

[21] Aditi Shrivastava, Nitin Shukla. [2012] Extracting Knowledge from User Access Logs, International Journal of Scientific and Research Publications, 2(4)

[22] Pooja Devi, Ashlesha Gupta and Ashutosh Dixit [2014] Comparative Study of HITS and PageRank Link Based Ranking Algorithms, International Journal of Advanced Research in Computer and Communication Engineering, 3: 5749-5754.

[23] Ciobanu D, Dinuca CE. [2012] Predicting the Next Page that will be visited by a Web Surfer using Page Rank Algorithm, International Journal of Computers and Communications, 6(1): 60-67.

[24] Michael G Noll. [2006] AOL Research Publishes 650,000 User Queries, 2006, http://www.michael-noll.com/blog/2006/08/07/aol-research-publishes-500k-user-queries/

[25] Yahya Al Murtadha, MD Nasir Bin Sulaiman, Yahya Norwati Mustapha, Nur Izura Udzir, Zaiton Muda. [2011] Web Page Recommendation System for Anonymous Users Based on Web Usage Mining, Advances in Communications, Computers, Systems, Circuits and Devices, ISBN: 978-960-474-250-9.

[26] Tranos Zuva, Sunday, O. Ojo, Seleman M. Ngwira, Keneilwe Zuva. [2012] A Survey of Recommender Systems Techniques: Challenges and Evaluation Metrics, International Journal of Emerging Technology and Advanced Engineering, 2(11): 382-386.

[27] Wesley Chu, Tsau Young Lin. [2005] Foundations and Advances in Data Mining - Studies in Fuzziness and Soft Computing, Springer Verlag, vol. 180.

[28] Pearl Pu, Li Chen, Rong Hu. [2012] Evaluating Recommender Systems from the User's Perspective: Survey of the State of the Art, Springer Science 22: 317–355.

[29] Schröder G, Thiele M, Lehner W. [2011] Setting Goals and Choosing Metrics for Recommender System Evaluations, Proceedings of the Second Workshop on User-Centric Evaluation of Recommender Systems and Their Interfaces (UCERSTI 2).