ACSIJ Advances in Computer Science: an International Journal, Vol. 4, Issue 6, No.18 , November 2015
ISSN : 2322-5157
www.ACSIJ.org

ACSIJ
WWW.ACSIJ.ORG

Personalization Web Pages for Site Users, Utilizing Users' Interests and Sequential Patterns Discovery

Zeynab Fazelipour[1,2], Ali Harounabadi[3]

[1]MSc Student, Department of Computer, Khuzestan Science and Research Branch, Islamic Azad University Ahvaz, Iran

[2]MSc Student, Department of Computer, Ahvaz Branch, Islamic Azad University Ahvaz, Iran
Z.Fazelipour@gmail.com

[3]Department of Computer, Tehran Center Branch, Islamic Azad University Tehran, Iran
A.Harounabadi@gmail.com

Abstract

With the rapid growth of information on the Web and increase of users who are daily visiting the web sites, presenting information proportionate to requirements of users who are visiting a special website so that they could find their desired information would be essential. Therefore, analyzing browsing behavior of web users and modeling this behavior has particular importance. The aim of recommender systems is guiding users to find their favorite resources and meet their needs, using the information obtained from the previous users' interactions. In this paper, to predict the web pages with high precision, a hybrid algorithm of clustering technique, All-*K th*-Order Markov model, and neural network are presented. For this purpose, in order to model users' movement behavior, after clustering those with the same interests, the sequential patterns are extracted on users' sessions of each cluster using all-4th-order Markov model. Next, in the step of pages recommendation to a current user, which is performed in an online state, first, a current user session is assigned to a cluster using neural network. Then Markov model created on the cluster which has the nearest match to the current session, is applied and a sequence of pages, which the users are interested to view, is included in the list of recommendation. The implementation results demonstrate that the proposed algorithm has higher precision and recall comparing to other recommender systems.

Keywords: *Personalization Web Pages, Clustering, Neural Network, Markov Models*

1. Introduction

Web is an important resource for information retrieval where Increasing growth of information has caused finding the needed information to be more difficult. The main challenge users are facing is to effectively find relevant information with the minimum effort and time invested. To resolve these problems, personalization the web pages in order to customize the web environment has become a popular phenomenon. Web personalization is a process where information or services provided by a website are adapted to the needs of a user or a specific group of users by using received knowledge of user navigation behavior and his specific interests in the form of a combination with content and structure of the website can provide active suggestion according to users behavior pattern [1]. User behavior modeling is a fundamental factor in each personalized system which is done implicitly by user information or extracted users' samples [2]. Using web mining techniques in order to extract knowledge from available web information is considered one of important approaches in web personalizing and it is classified to three active researches field based on extracted web data area: web content mining, web usage mining, web structure mining [3]. So far, many personalized systems based on web-mining are created. In fact, personalization systems the web pages are an important component in a website to provide the desired or required information of the users without their explicit request. The aim of web personalization system is to recommend a series of items to the current user. Such recommendations include links, advertisements, texts, products, etc., in lines with the user's interests and preferences.

In this paper, a recommender system is presented using Markov model and based on users' interest, predicting the future requests of them. In the proposed method, after preprocessing log file and extracting of users' features from their sessions, the users profile is established in an offline state. Then, profile of users by k-means method is clustered and users' movement patterns are extracted. In

the next step, All-4 the-order Markov model is applied in order to model users' movement behaviors for each cluster. Finally, in an online state, the neural network assigns a new user to the cluster belonging to him/her and the Markov model corresponded to that cluster predicts future users' requests and gives a list of suitable pages to them. The proposed algorithm has been simulated on real-world data, and the results indicate that proposed algorithm has significantly enhanced the quality of the recommendations. The rest of this paper is organized as follows: In the second section the needed background material is provided. Section 3 includes several works done in this field Section 4 elaborates on the proposed method. The details of implementation, data set and evaluations are explicated in section 5 and finally, section 6 concludes the paper.

2. Background Material

In this section, web usage mining methods and All-K th-Order Markov model used in this paper are discussed.

2.1 Web usage mining

Web usage mining refers to process of discovering meaningful and suitable patterns of web user's application data. Application data refers to data which is stored in Web server log file by users when they use web. Web usage mining is the most applicable method for extracting users' behavior pattern in these files [4,5]. This approach concentrates on the techniques which can predict user behavior while cooperating in web. The principle duty in web usage mining is retrieve useful information from the web server records. This item is divided in to three phases. Data Preprocessing, Pattern Discovery, and Pattern Analysis [6,7].

Data Preprocessing: the collected web data usually consists of large and heterogonous information. The data should be changed to homogeneous and adaptable information which is suitable for pattern discovery phase. Like many data mining applications, preprocessing and preparing data include filling the missing values, deleting noise, changed and formatted data and eliminating incompatibility [8]. In web usage mining, this phase includes data purging, recognizing user and their session, which are fundamental factors for discovery pattern.
Pattern Discovery: In this step rules, patterns and statistic information, applying data mining techniques on user access sessions are discovered.
Pattern Analysis: seeks rules and statistic information acquired form pattern discovery phase which is very interesting for management site personnel.

2.2 All-Kth-order Markov model

Markov model is a model for studying stochastic processes which is efficient for modeling and predicting browsing behavior of website users. These models are a method for the discovery of sequential patterns in order to predict link. This modeling is based on the transition probability between web pages that are stored in the user session [9].

If $P = \{p_1, p_2, \ldots, p_m\}$ be a set of pages in a web site and W be a user session including a sequence of pages visited by the user in a visit. Assuming that the user has visited l pages, then prob($p_i|W$) is the probability that the user visits pages pi next. This probability, $prob(p_i|W)$, is estimated using sequences of page visited by all users in history data, denoted by W. This is done by creating a transition probability matrix. In principle, with this method, the probability of all the pages to be the next page is calculated and then the page with the highest probability is selected as a predication.
Of course, the Markov model starts calculating the highest probability of the last page visited because during a web session, the user can only link the page he is currently visiting to the next one. But, since a precise calculation of the whole conditional probabilities is not possible, the Markov process is used for predication of the next page. This process takes a limited number of previously viewed K pages. In other words, the probability of visiting a page pi does not depend on all the pages in the web session and on a small set of k previous pages, where $k << l$ [9].

So, the next page, p_{l+1}, that the user will visit, by equation (1) is expressed [9]:

$$P_{l+1} = arg\ max\ _{p \epsilon P} \{P(P_{l+1} = p | p_l, p_{l-1}, \ldots, p_{l-(k-1)})\} \quad (1)$$

Where, k denotes the number of the preceding pages and it determines the order of the Markov model .

The Markov model order is corresponded to the number of prior events in predicating a future event. Hence, the Markov model from k-th order predicts the probability of next event by viewing the previous K event. For calculate probabilities, if S_j^k be a state containing k pages, $S_j^k = \{p_{l-(k-1)}, p_{l-(k-2)}, \ldots, p_l\}$. The probability of $P(p_i|S_j^k)$ from a history (training) data set, using equation (2) is estimated [10]:

$$P(p_i|S_j^k) = \frac{Frequency(\langle S_j^k, p_i \rangle)}{Frequency(S_j^k)} \quad (2)$$

In most cases, the Markov models with lower order (first or second) cannot exactly predict the next page that is to be visited by user. This is because these models do not review profoundly the user's history. Thus, in order to achieve to a

better precision, the Markov models with higher order (third or fourth) must be used. But, higher-order Markov models have a number of limitations associated with high state-space complexity, low coverage, and sometimes even worse prediction accuracy. One method to overcome the problem of low coverage is to train varying order Markov models and then combine them for prediction. This scheme is called the All-K th-order Markov model.

Utilization of all-K th order Markov model usually contributes to produce separate models for each K-order. If the model is unable to predict by k-th order, it will try to do it by reducing the model order gradually. In this scheme, for each test instance, the highest-order Markov model that covers the instance is used for prediction. For example, if the first, second, and third-order Markov models is built, then, given a test instance, first a prediction is done using the third order model. If this model does not contain the corresponding state, then a prediction is done using the second-order model, and so on [11].

3. Review of the Previous Works

In [12] have proposed an experimental system which classifies users of browsing patterns using a combination of web usage and content mining. In this paper, firstly, the profile of users is created based on information achieved from web server logs. Subsequently, browsing pattern of each class of users is derived applying clustering algorithm to profiles. Afterwards, achieved result is combined with content of corresponding web page to generate users' browsing pattern for predicting future requests. At the final stage, recommender engine generates a list of user desired pages using neural network.

In [13] has proposed a hybrid model to improve web page prediction. The main goal of this paper is optimizing efficiency of Markov model by K-NN classification algorithm. This hybrid system consists of two parts. In the first one, a training dataset is selected and is classified into multiple classes that classes illustrate data pattern on data set. In the second part, Markov model is applied on classified data in order to predict next page.

In [14] has provided a mixed system for discovery and analysis of users' movement patterns. In this paper, after preprocessing log file data, a clustering algorithm based on ants for pattern discovery is used which is out of system line phase. In the online phase, movement patterns are classified using decision tree classification method and user's next request in a web site is predicted.

In [15] has suggested a method for predicting web access pattern. This model is a combination of Boosting and Bagging methods that improves the accuracy of the prediction model. It aims to create a prediction model for behavior of user's random browsing so that the next requested pages could be estimated according to previous visited pages.

Construction steps prediction model provided with Bagging, Boosting as follows:

- Data Preprocessing (Data cleaning, Session identification)
- Apply Bagging method on the training data set
- Apply Boosting method on the training Data set
- Combine bagging and boosting prediction results

The results show that the accuracy of the Bagging and Boosting prediction is improved as compared with Markov Model and Markov Model combined with ARM.

In [16], using users' behavior analysis, an approach, which makes use of preprocessing of log-file, is presented in order to pre-fetch, predict, and improve the performance of the web server. Clustering, Markov Model, and association rules are the three techniques used in recommending web pages. Therefore, after preprocessing and identifying the sessions, they are clustered using K-means clustering algorithm and measuring similarities. Every data-set is grouped in a different cluster. Then, Marco Model predicts the results. In case of ambiguity, association rules are applied to present accurate results. The knowledge base in this system is a reservoir of features which are mined using data-mining techniques. These features include the number of users, the visited web pages, and the time to access the pages.

4. Proposed Method

The proposed method in this paper is based on web usage mining approach. The proposed algorithm consists of two phases: offline phase and online phase. In the offline phase, at first web server registries are preprocessed, and then the users' sessions are extracted. Then, Users profile is created using a mean vector based on each user's interest and with K-means clustering method; the resulting profiles are clustered so that the users' movement patterns to be extracted. In the proposed system, the All-4th order Markov model is used to model users' movement behaviors in each cluster. Then, in online phase, the movement pattern (cluster) corresponding to an active user session are indentified using neural network and finally using predictive process of the corresponded Markov

model, the sequence of pages are predicted for users. The proposed algorithm includes the following steps:

1. Preprocessing of the server registries in order to extract the user's sessions. Preprocessing of the server registries include data cleaning, user identification, and session identification.

2. Dividing the obtained sessions into two categories of training and test sessions set.

3. Extraction of users' characteristics and weighting the viewed pages during sessions

4. Making users' profiles using training sessions

5. Apply k-means clustering method in order to cluster profiles and extract movement users' patterns

6. Modeling sessions of each cluster using all- 4th-order Markov model

7. Training the neural network using obtained movement patterns

8. Active user simulation using test sessions set

9. Utilization neural network in order to determine the relevant cluster using test sessions set. If the current session is placed in the relevant cluster, the algorithm continues, otherwise it goes back to step 5.

10. Recommend a list of predicted pages to the current user using Prediction Process by Markov model corresponding to detected cluster.

4.1 Determining Weight of Pages

When the users' sessions are indentified, the redundant pages are deleted. The reason is that if the number of pages is beyond the normal limit, the clustering process needs too much time. Therefore, in the proposed method, we use the effective dimension reduction method in order to improve the clustering outcomes. For this purpose, inspired by [17], we omit the pages whose degree of support (a ratio of the number of sessions containing that page to the total pages) is too low or too high. The pages with poor support are those that are rarely visited and not viewed over total sessions. These pages have little informational value and not suitable to include in the movement patterns. In contrast, the pages with strong support are those that are almost always viewed during sessions. They are homepages of a site that are displayed in most of sessions. Thus, the pages are deleted when access to them is below 10% over the maximum accesses. Those pages viewed above 80% are also omitted, like the home pages. Then, all the user sessions with length less than 5 pages are deleted because they cannot be predicted by 4-th order Markov model.

We represent each session s_i as an m dimensional vector over the space of web pages, $s_i = \{w(p_1, s_i), w(p_2, s_i), \ldots, w(p_m, s_i)\}$, where $w(p_j, s_i)$ is a weight assigned to the jth web page $(1 \leq j \leq m)$ visited in the session si. In this paper, the weight $w(p_i, s_j)$ is defined as the interest degree of a particular user to the page, which is the harmonic mean of page observation time and page observation frequency to represent this interest [12].

4.2 Creating Users' Profile

This system module is used to create users' profiles. For this purpose, session vectors related to different users separated. Assuming $\{s_1, s_2, \ldots, s_k\}$ be a set of session vectors of ith user (u_i). We compute a mean vector sui for the user u_i as its representation. This mean vector represents web pages, which are interesting in by the users. The weight of each web page in the mean vector is computed by the average weight of the web pages across total access sessions of the user $\{s_1, s_2, \ldots, s_k\}$.

4.3 Clustering Profiles

In order to cluster the obtained profiles, the K-means clustering algorithm is used. The following algorithm is a basic one for this method.

1- First, K points are selected as center points of clusters.

2- Each data case is assigned to the cluster whose center has the shortest distance to that data.

3- Calculation of new centers of each cluster. Each new center will have its own mean value of clustered data points.

4- The above steps 2 and 3 are reiterated until no variation occurs in the cluster centers any more.

The relation (3) is expressed as an objective function [18]:

$$SSQ(N, C) = \sum_{i=1}^{k} \sum_{X \in N_i} d(X, C_i) \qquad (3)$$

Where, $d(X,C_i)$ is a measure of distance between the points and C_i is ith cluster center.

After apply the k-means clustering method to the obtained profiles, several cluster centers are resulted as $C = \{c_1, c_2, \ldots, c_m\}$ in which every c_i $(1 \leq i \leq k)$ indicates a subset of the users' sessions, and k indicates the number of clusters. We

compute a mean vector (m_c) for each cluster $c \in C$ as its representation. The mean value for each web page in the mean vector is equal to the mean of the weight of that page to total sessions that cluster. Every average vector shows the browsing pattern of users in a cluster in a special class of accessed web pages. As the results of profiles clustering, *NP = {np1, np2, ...,npk}* is used to represent the set of users navigation patterns, in which each *npi* is a subset of *P*, the set of web pages.

4.4 Creation Markov Model on Sessions of Each Cluster

After clustering the profiles, all-4th order Markov model is applied to model the users' movement sessions for each cluster. In this way, on each cluster, the transition probability matrix is established between web pages for 1-4th order Markov model. To create a transition probability matrix in the Markov model, the value of transition probability is estimated between web pages in the sessions using the relation (2) that is introduced in the section (2-2).

4.5 Finding Corresponding Cluster Utilizing the Neural Network

In online mode, we use neural network to find the most similar cluster to the user's current access session. Therefore, we train neural network using users' navigation patterns. The navigation patterns have been considered as the inputs of the network and the relevant cluster's number as the output of the one.
Neural network input is a vector of weights of web pages visited in the session. So, a profile for the users' current session based on the weight of the pages is created. Now, should be determined that current session profile belongs to which cluster (navigation pattern). For this purpose, current session profile is made, to the neural network input are given and the network determines relevant cluster's number for the session.

4.6 Prediction of Web Pages Using All-4th-order Markov Model

In this step, having identified the current session cluster, the all-4th-order Markov model related to that cluster predicts the preferred page for users. For each test session, the transition probability matrix of 4th order Markov model related to that cluster is employed to predicate the current page corresponding to the test session. The pages with the highest transition probability are included in a list of recommended pages. If each state is not covered by 4th order matrix, the 3, 2 and 1th order matrices are matched so that the sequence of pages that the user is interested to view is included in the recommendation list.

5. Implementation and Evaluation of the Proposed Method

To implement the components of the proposed system, Microsoft SQL Server and MATLAB software were utilized. The log file data of this research is collected based on Nasa log file web server. After preprocessing and identifying sessions, in order to train and test the proposed system, the 70% of sessions were selected for training and the rest were used for test and evaluation. Then, the proposed system is taught using the educational data. The optimal number of cluster compactness and cluster separation proposed in [19]. The clustering method applied and produced 7 clusters. A perceptron network was used to learn from the data. After training the system, the experimental data are used which were not responsible in making the movement patterns, and simulated the active user. The aim of personalization is to calculate a proposed set (rs) for the user's current session which has the highest correspondence with the user's interests. This part is the only online component of the system and must be highly efficient and precise.

5.1 Evaluation Metrics

Two criteria of precision and recall are effective parameters in system performance that using formulas (4) and (5) are obtained. Precision refers to the capability of recommender system to generate precise recommendations. In other words, the precision of recommendation equals to the ratio of correct recommendations to total number of recommendations. Recall refers to the ability of the recommender system to generate suggestions which could be seen by the user. In fact, recall is the ratio of diagnosed correct recommendations to remained pages in the continuation same session [20].

$$Precision(rs, rp) = \frac{|rs \cap rp|}{|rs|} \quad (4)$$

$$Recall(rs, rp) = \frac{|rs \cap rp|}{|rp|} \quad (5)$$

Where, rs is the output of the proposed system (suggested set) and rp is pages viewed by the user in the continuation same session.

5.2 Comparing the Proposed Algorithm with other Methods

In this section, In order to evaluate suggested method as in Figures 1, 2 and 3 are shown, first, we compared the precision of our method with association rule Markov model [21], the integrated Markov model (second order) with association rule and clustering [10] and the hybrid Markov model (third order) with clustering [22] for predication of next web page. The precision of next page prediction is equal to percentage of the number of correctly predicted sessions to total test sessions.

As well as, the precision and recall of the proposed algorithm are compared respectively, with the performance of the recommender systems NEWER [17] and IPACT [23] than the number of suggested different pages with the window length of 4.
The experimental results demonstrate that the proposed algorithm has higher precision and recall than algorithms that were compared.
As it can be seen in Figure (1), the proposed method outperformed the other methods.

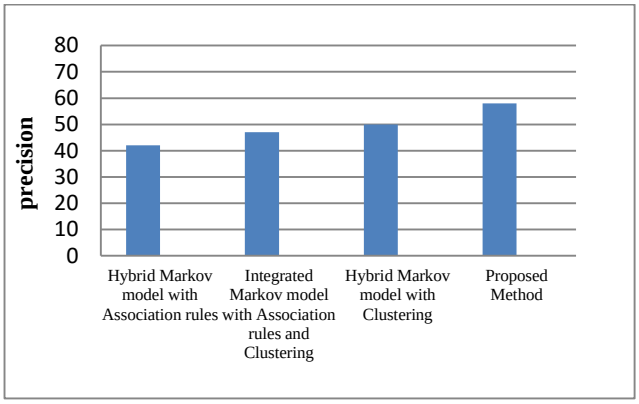

Fig. 1 Comparison of the precision algorithms for predicting the next web page

As it can be seen in Figure (2), except in the number of proposed page of 3, in which NEWER method has higher accuracy than the proposed algorithm, in other parts the proposed algorithm is more accurate compared to the other two methods.

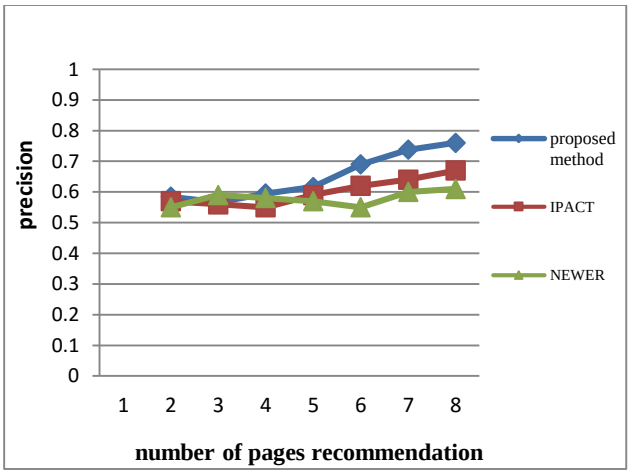

Fig. 2 Comparison of the precision algorithms for the number of suggested different pages

Figure (3) indicates the recall of the algorithms being compared with one another. By increasing the number of proposed pages, the proposed algorithm, as the figure (3) shows, has a greater recall in comparison with the NEWER method except about size of suggested pages 3, 5 and 6. Such improvement is also true compared with the IPACT method except in the range of 2 to 4 pages.

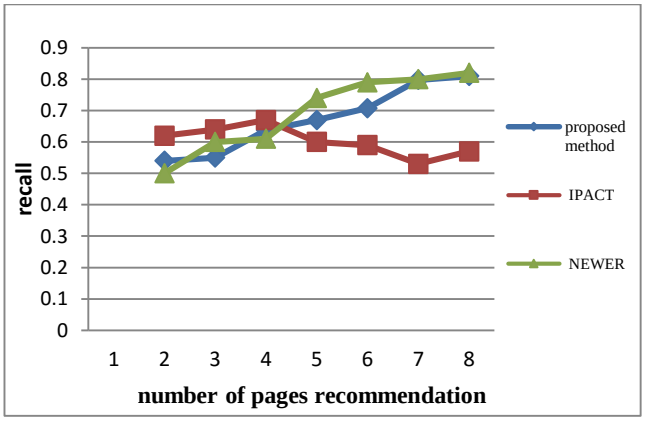

Fig. 3 Comparison of the recall algorithms for the number of suggested different pages

6. Conclusions

In this paper, a new hybrid algorithm of clustering technique, All-*Kth*-Order Markov model and neural network was presented. In this algorithm, the profiles of users who have similar movement behaviors are clustered

using K-means clustering method. Then, to model users' movement behavior, sequential patterns were extracted from the users' sessions of each cluster using all-4th-order Markov model. Next, in the step of pages recommendation that was done in the state of online, first, a current user session is assigned to a cluster using neural network. Then, the Markov model established on that cluster, which has the nearest match to the current session, is applied, and the sequence of pages which the user desires to view is included in the recommendation list. It can be concluded that a better efficiency of the system results from modeling and prediction of higher order Markov Model with all orders. Also, the analysis of the model was done on users' cluster which could model the users who had the same interests and produce more precise prediction using the more homogenous and limited sessions. Evaluation results revealed that the proposed method has high precision and recall while recommending pages to users.

References

[1] Eirinaki, M., Vazirgiannis, M. ," Web mining for web personalization," ACM Transactions on Internet Technologies (TOIT), NY, USA, Vol. 3, No. 1, 2003, pp. 1-27.

[2] Forsati, R., Meybodi, M., "An Algorithm Based on Structure of Connected Pages and Information of Users for Suggesting Web Pages", The Second Iran Data Mining Conference, industrial Amir Kabir university, 2008.

[3] Chen, P.Z. and Sun, C.H. and Yang, S.Y., " Modeling and Analysis the Web Structure Using Stochastic Timed Petri Nets", Journal of Software, Vol. 3, No. 8, 2008, pp. 19-26.

[4] Tyagi, N.K. and Solanki, A.K. and Wadhwa, M., "Analysis of server log by web usage mining for website improvement," International Journal of Computer Science Issues (IJCSI), Vol. 7, No. 8, 2010, pp. 17-21.

[5] Verma, V., Verma, A.K. and Bhatia, S.S., "Comprehensive analysis of web log files for mining," International Journal of Computer Science Issues (IJCSI), Vol. 8, No. 6, 2011, pp. 199-202.

[6] Chitraa, V. and Davamani, A.S., " A survey on preprocessing methods for web usage data," International Journal of Computer Science and Information Security, Vol.7, No. 3, 2010, pp.78-83.

[7] Santra, A.K. and Jayasudha, S., "Classification of web log data to identify interested users using naïve Bayesian classification," International Journal of Computer Science Issues (IJCSI), Vol. 9, No. 2, 2012, pp. 381-387.

[8] Pamutha, T., Chimphlee, S., Kimpan, C., Sanguansat, P., "Data preprocessing on web server log files for mining users access patterns," International Journal of Research and Reviews in Wireless Communications (IJRRWC), Vol. 2, No. 2, 2012, pp. 92-98.

[9] Thwe, p., "proposed approach for web page access prediction using popularity and similarity based pagerank algorithm", international journal of scientific & technology research (ijstr). Vol. 2, No. 3, 2013, pp. 240-246.

[10] Khalil, F., H., Li, Wang, H., "An Integrated Model for Next Page Access Prediction", International Journal knowledge and Web Intelligence (IJKWI), Vol.1, No.1/2, 2009, pp.1-18.

[11] Deshpande, M. and Karypis, G., "Selective Markov models for predicting Web Page accesses," ACM Transactions on Internet Technology (TOIT), Vol. 4, No. 2, 2004, pp 163-184.

[12] Rashidi, S.F., Harounabadi, A., Abasi Dezfouli, M., "Prediction of Users Future Requests Using Neural Network", Management Science Letters. Vol. 2, No. 6, 2012, pp. 2119-2124.

[13] Kaushal, P. ,"Hybrid markov model for better prediction of web page," International Journal of Scientific and Research Publications(IJSRP), Vol. 2, No. 8, 2012, pp. 1-4.

[14] Sujatha, V., Punithavalli , "An approach to user navigation pattern based on ant based clustering and classification using decision tress," International Journal of Advanced Engineering Sciences And Technologies, Vol. 1, No. 2, 2010, pp. 112 – 117.

[15] Girija P., Kavitha V., "An Approach for Predicting User's Web Access Pattern", International Journal of Computer Science and Management Research, Vol.2, No.5, 2013, pp. 2585-2589.

[16] Makker, S., Rathy. R.K, "Web Server Performance ptimization using prediction pre fetching Engine", International Journal of Computer Applications, Vol. 23, No. 9, 2011, pp. 19-24.

[17] Castellano, G., Fanelli, A.M., Torsello, M.A., "NEWER: A System for Neuro-fuzzy Web Recommendation", Applied Soft Computing, Vol. 11, No.1, 2011, pp. 793-806.

[18] Jain, A. K., Murty, M. N. and Flynn, P. J., "Data Clustering: A Review," in: ACM Computer, Vol. 31, No. 3, 2000, pp. 264-323.

[19] Liu, H., Keselj, V., "Combined mining of web server logs and web contents for classifying user navigation patterns and predicting users' future requests," Data and Knowledge Engineering, Elsevier, Vol. 61, No.2, 2007, pp. 304-330.

[20] Nakagawa, M., Mobasher, B. ,"A hybrid web personalization model based on site connectivity," In Web KDD Workshop at the ACM SIGKKDD International Conference on Knowledge Discovery and Data Mining, 2003, pp. 59-70.

[21] Khalil, F. and Li, H. and Wang, H.,"A framework of combining Markov model with association rules for predicting web page accesses," Proc. of the 5th Australasian Conf. on Data Mining (AusDM'06), Australian Computer Society, 2006, Vol. 61, pp. 177-184.

[22] Pandey, T., Kumari, R., Tripathy, A., and Sahu, B., "Merging data mining techniques for web page access prediction: Integrating Markov model with clustering," International Journal of Computer Science (IJCSI), Vol. 9, No.1, 2012, pp. 188-193.

[23] AlMurtadha, Y. and Sulaiman, M., Bin, N., Mustapha, N. and Udzir, N.I., "IPACT: Improved web page recommendation system using profile aggregation based on clustering of transactions," American Journal of Applied Sciences, Vol. 8, No. 3, 2011, pp. 277-283.