An Efficient Web Recommendation System using Collaborative Filtering and Pattern Discovery Algorithms

R. Suguna
Assistant Professor
Department of Computer Science
and Engineering
Arunai College of Engineering
Thiruvannamalai – 606 603

D. Sharmila
Professor and Head
Department of Electronics and
Instrumentation Engineering
Bannari Amman Institute of Technology
Sathyamangalam- 638 401

ABSTRACT
Information is overloaded in the Internet due to the unstable growth of information and it makes information search as complicate process. Web recommendation systems assist the users to get the exact information and facilitate the information search easier. Web recommendation is one of the techniques of web personalization, which recommends web pages to the user based on the previous browsing history. It is done either content based approach or collaborative filtering approach. In this paper web usage mining is considered as the major source for web recommendation in association with Collaborative filtering approach, association rule mining and Markov model to recommend the web pages to the user.

Keywords: Web Recommendation, Apriori Algorithm, Markov model, Collaborative filtering, Web Usage Mining

1. INTRODUCTION
In today's world, people are very busy with engagements; the customary calendar day is not sufficient for them to comprehensive their needs. E-commerce is the most excellent solution for them to save their time and making their job comfort and ease. In order to facilitate the customers, companies are rushing to post their details worldwide. So, information is overloaded in the Internet and getting the pertinent information from the Internet is intricate task. In the framework of Web search, it is difficult to identify the user's anticipation and convince the user requirements. Web recommendation system will really help to satisfy the above addressable issues in the web search.

Web recommendation [1] is habitually related with web personalization to carry out the needs of the customers. Web personalization is defined as the task of building web based information systems suitable for the significance and outlook of the user. A personalized website recognizes its users and gathers the information based on their anticipation and attempt to modify or streamline the website contents. Web recommendation [2] is one of the techniques of web personalization, which refers to the recommendation of a set of hyperlinks that are coupled to the interests and preferences of the user. The management of the recommended links is done either in a separate frame of the Web page or in a pop-up window.

Web recommendation systems help the website visitors for easy navigation of web pages, quickly reaching their destination and to obtain relevant information. There are two types of approaches [1] to develop recommendation systems: (i) Content based filtering method, (ii) Collaborative filtering method, some cases combination of both the approaches [3] are preferred by the researchers.

Content-based filtering technique [3] is based on content learn from the target items. In content-based filtering technique, the web pages are recommended for a user exclusively on a profile built up by analyzing the content of items that the user has rated in the ancient times and/or user's personal information and preferences. The user's profile can be constructed by examining the responses to a questionnaire, item ratings, or the user's navigation information to infer the user's preferences and/or interests. By conclusion, in this method, recommendation can be done mainly from the past experience of the website visitor. The disadvantage of this method is, not all the times users give their ratings properly for a website or web pages even though it will be helped them for future.

In collaborative filtering approach, [4] web pages are recommended to a particular user when other similar kind of users also prefers those web pages. The definition of 'similarity' between users depends on applications which use the web recommendation system. For example, it may be defined as users having similar ratings of web pages or websites or users having related navigation behavior. A collaborative filtering system collects all information about users' activities on the web site from the web servers and calculates the similarity among the users. Users have similar characteristics will be categorized to the same group. This method had two disadvantages: 1. Sparsity [5] 2. Scalability [6]. In sparsity, the amount of ratings previously obtained is very miniature compared to the number of ratings that need to be predicted. Collaborative filtering requires explicit non-binary user ratings for similar products. With this problem, collaborative filtering based recommendations cannot accurately compute the future expectations (neighbourhood) of the product or web pages for recommendation. The second problem is related to scalability. Recommendation systems using collaborative filtering approach to find the neighborhood generally necessitate very long computation instance that grows linearly with both the number of customers and the number of products or web pages.

Web Usage Mining plays a vital role to address the above mentioned issues. Many researchers [6][7] used web usage mining as a tool for their recommendation systems. The task of web usage mining [9] has two broad phases: (i) Data Preparation, (ii) Pattern Discovery. During data preparation, data are collected from web a server which records the user wise session details and make them suitable for applying

pattern discovery algorithms. In pattern discovery phase, [1][10] data mining techniques such as clustering, classification, association rule mining etc., are applied to discover the patterns. The discovered patterns [11] are useful for making suitable decision in website restructuring, web recommendation, site modification etc.

The most commonly used web recommendation system with web usage mining applying association rules to find the frequently visited web pages by the users or website visitors. Most of the website visitors have a confusion that, Where can get relevant information? How can navigate the web pages? How can reach target place? An excellent website will assist its users to resolve the above confusions by make use of an efficient recommendation system.

The author [7] found that, the worth of the recommendation system has a significant outcome on the customer's future shopping behavior. Poor recommendations can cause two types of characteristic errors: false negatives, which are products that are not recommended to the customers, though the customer would like them, and false positives, which are products that are recommended to the customers, though the customer does not like them. In an e-commerce environment, the false positives are avoided strictly, otherwise these types of mistakes will lead to irritated customers and thus they will be unlike to revisit the site.

In this paper, a recommendation is done effectively using the web usage mining as follows:

i. Web logs are preprocessed to eliminate the inconsistency.
ii. Users are grouped based on similar browsing behavior.
iii. Association rule mining algorithm is applied to find the frequently used web pages.
iv. Markov Model is applied to recommend the web pages.

This paper is organized as follows. Chapter 2 deals with Literature Survey and chapter 3 describes the Proposed Methodology, Chapter 4 explains the Result and Experimentation and finally Chapter 5 discusses the Conclusion and Future Direction.

2. LITERATURE SURVEY

The authors [12] have proposed a new algorithm called Profile Aggregations based on Clustering Transactions, which is used to group the similar kind of transactions and in the second level page view clustering is applied to identify the similar pages in each transaction.

Authors [7] have improved the efficiency of the collaborative filtering approach based recommendation which overcomes the problem of sparsity and scalability by using web usage mining, decision tree induction method, association rule mining algorithms and data warehousing technologies. They have used web logs as a source to find the frequent patterns using Apriori algorithm and built product taxonomy. Decision tree induction method is used to classify the customers, finally a recommendation system with five levels have introduced to recommend the items to the customer.

Research done by [13] and they have proposed the recommendation techniques which collectively used web usage and web content information for recommendation. They have used Concept Logs for web recommendation. The research by [14] proposed an intelligent web recommendation system based on Fuzzy approximation reasoning. They have used web usage mining to extract the user profile and grouping the user sessions using hierarchical unsupervised niche clustering method. Fuzzy approximation reasoning techniques are used for recommendation.

This research [8] has taken effort to expose the role of web usage mining in the area of web personalization. They have classified the personalization process into four major areas: (i) Memorization, which the system record user's browsing details, and will returns the past browsing history when the user log on to the system, (ii) Guidance, which returns the couple of web pages based on user's preferences, also termed as recommendation systems, (iii) Customization, which modify the content, structure etc., based on the user's preference and (iv) Task performance support, which executes the action on behalf of the user.

A novel recommendation method [15] which combines web content semantics with users' navigational behavior is proposed by the researchers. Semantically coherent clusters are formed by extracting the keywords from the web contents based on user navigation. Domain ontology is formed based on the keywords extracted from the web contents. The resultant clusters and ontology mapping are then used to produce recommendations to the end user that are semantically relevant to his current navigational behavior. The authors [16] done a recommendation using web usage mining with two major data mining algorithms such as clustering and association rule mining. They have used Hierarchical Bisecting Mediods for clustering the users with respect to time framed session. Association rules are applied to above formed groups to find the similar kind of students in future. [17] proposed an intelligent web recommender system namely SWARS (Sequential Web Access-based Recommender System) that uses a sequential pattern mining technique for predicting the next web pages. The paper also proposes a compact data model, called Pattern-tree, which stores the sequential web access patterns, and an efficient approach for user pattern matching and recommendation rules generation.

Researchers [18] proposed a recommendation system mainly based on applying association rule mining concept in the web log files for better recommendation. They have introduced one algorithm for finding the association rule from the web logs called Formal Concept Analysis (FCA) using lattice theory. They have proved that, the new algorithm is better than the Apriori algorithm.

Authors [19] done a recommendation based on collaborative Web recommendation scheme based on Latent Dirichlet Allocation (LDA) model. [20] developed recommendation model which combines web usage data, content data, and structure data in a web site to generate user navigational models.

Researchers [21] have developed a novel recommendation system for the students who are all themselves learning a technologies through e-learning environment using web usage mining. Their recommendation system will automatically

suggest the educational resources for the students based on their browsing history. Learner and Content module is built in offline then recommendations are suggested based on the above built model. [22] used distributed learning automata to learn the behavior of previous users' and cluster pages based on learned pattern. One of the challenging problems in recommendation systems is dealing with unvisited or newly added pages. They have introduced a novel Weighted Association Rule mining algorithm, to address the above problem and HITS algorithm is used to extend the recommendation set.

Web usage mining techniques [23] is used by the researchers for determining the interest of "similar" Users. The complete process for recommendation broadly consists of two components: offline component and online component. The offline component involves Data Preprocessing, Pattern Discovery and Pattern Analysis. The outcome of the offline component is the derivation of aggregate usage profiles using web usage mining techniques. The online component is responsible for matching the current user's profile to the aggregate usage profiles. The scope of this paper is to match an online user's navigational activity with the aggregate usage profiles obtained through mining tasks and provide suitable page recommendations which may be of interest to the user.

Recommendation [3] is done the authors by combining collaborative and content based method. They have used web content mining as the source and Fuzzy C-Means and Ant colony clustering techniques are applied to the web contents as the offline process. A hybrid recommendation systems checked the page matching with the previous similar users and done the suitable recommendation. [24] have done the recommendation based on collaborative filtering technique only for the trustworthy customers. Entropy based similarity measure is used to identify the similarity between the users.

Authors [25] had done research on semantic web personalization. In this paper, web contents are modified based on the user's searching and navigational behavior. [26] have presented a web page recommendation algorithm using weighted sequential patterns and markov model. [27] proposed a technique that incorporates web recommendation and personalization of websites based on the user interest. They have taken web logs for their source. They have used the data structures such as Web-Interest Matrix, User-Interest Matrix, Class-Interest Matrix and Frequent-Path Matrix to keep track of user interest and change the website based on the impact of the users.

The researchers have [28] constructed a new model to understand the user personal behavior about webpage navigation using Latent Dirichlet Allocation (LDA). They have also proposed three types of recommendation models namely pure-LDA, LDA-knn, and LDA-tran. [29] introduced recommendation systems for news and story readers based on context trees. The recommender system provides high-quality news recommendation to new visitors based on present browsing behavior. They proved that context-tree recommender systems provide good prediction accuracy and recommendation novelty, and they are sufficiently flexible to capture the unique properties of news articles.

Table 1 shows the tabular form of techniques used by the various authors in the field of web recommendation.

3. PROPOSED METHODOLOGY

This research work has proposed a new system for efficient web page recommendation using Web usage mining and Markov model which are coupled with the pattern discovery algorithms such as clustering and association rule mining. The new recommendation system is developed with the following four processes: (i) Data preparation (also termed as data preprocessing in web usage mining field), (ii) Clustering the web log files based on user wise, (iii) Determining the associative patterns, and (iv) Web page recommendation.

3.1 Data Preparation

Whenever the user interact the website, the interaction details are recorded in the web server in the form of web log files. Web log files [1] are maintained in the web servers in the form of plain text files. It is too difficult to use the web log files directly.

Preprocessing techniques are necessary for the web logs to discover the knowledge from them. Commonly the web log files are maintained in Web Servers (Web Logs for all the users), Proxy Servers (Maintained Somewhere), or/and Browser (Web Logs for the particular client). Traditionally there are three types of web log formats. They are

- W3C Extended Log file Format.
- Microsoft IIS Log File.
- NCSA Common Log file Format.

The common format for the web log files are "Common Log Format" (CLF) or "Extended Log Format" (ELF) which consists of the following fields [10]

<ip_addr><base_url><date><method><file><protocol><code><bytes><referrer><user-agent>

The data preprocessing is considered as the important activity in web usage mining technique and treated as a key to success. It consists of the following steps:

Data Collection
Data collection is the first step in web log preprocessing. The web logs may be collected from web server, proxy server or client machine.

Data Cleaning
Data cleaning is the important step in preprocessing. In this step, irrelevant and noisy data are removed from the log files. The result of data cleaning has the fields like date, time, client ip, URL access, Referrer and Access log files.

User Identification
User Identification is the difficult job of web log preprocessing. Here, ip address and user id (name) is considered as the unique user.

Session Identification
The following is the rules we use to identify user session in our experiment:

1) If there is a new user, there is a new session
2) In one user session, if the refer page is null, there is a new session
3) If the time between page requests exceeds a certain limit(30 or 25.5mintes), it is assumed that the user is starting a new session.

Path Completion

There are chances of missing pages after constructing transactions due to proxy servers and caching problems. In such a condition it becomes necessitate identifying the user's access path, and adding the missing paths. Because of local buffers existence, some requested pages will not be recorded in access log. The goal of path completion is to fill in all the missing references that are not recorded. The solution for path completion is that if a requested page can be reachable by a hyperlink from any of the visited pages by the user, we assume that it should be added in the session. When there are two or more pages which have a super link to it in one session, then it should be placed before the latest visited page.

The preprocessed web logs are taken into consideration for the next step (i.e) Web log clustering. In this proposed method, [30] Modified bird flocking algorithm is used for grouping the web logs. In order to support the algorithm, the web logs are preprocessed and the following fields are extracted from the web logs and each field as considered as boid as per the bird flocking algorithm.

***b* = < *ip, user, url, session, frequency* >**

Where, b is the representation of the boid, which has values, "ip" the ip address, "user" user name, "url" web address, "session" session duration of the user, "frequency" the number of visits by the user.

3.2 Web Log Clustering

In this paper, Clustering has been used for grouping the users with common browsing behavior. User interest level information is extracted by clustering the web logs based on modified bird flocking algorithm.

3.2.1.Bird Flocking Algorithm

The bird flocking algorithm is swarm intelligence algorithm based on the behavior of bird flock. The bird flocking was introduced by C. Reynolds as boids model. The Boid algorithm is used to group the websites. The boids are considered as moving data points and hence the clustering should be considered in diverse manner. The parameters used to cluster the boids are affinity and centroid based calculation. If there are n number of objects, then there would be n number of boids considered in a dataset.

$$D = [x_1, x_2, ..., x_n]$$

Where, D denotes the dataset and x denotes the boids in the dataset and n is the number of boids.

Affinity Calculation

The affinity value defines relationship of two boids which has the similarities amid same objects. The similarity amid the related objects is indicated quantitatively by the affinity amid the two boids. The affinity amid the related boids is high, if it has more similar objects. To group the data the boids algorithm uses the affinities of boids as an important parameter. The affinities amid two boids are calculated using the following formula:

$$A_{ij} = \left(\sqrt{\sum_{k=1}^{L} (x_{ik} - y_{jk})^2} \right)$$

Where,

X_i, Y_i → boids considered for similarity calculation

L → Total length between X_i and Y_i

Centroid rule and Merging rule

Every boid travelling in the environment is considered as centroid and each centroid has many objects as per their similarities. For instance, if a database has hundred objects, the algorithm will have hundred boids at first and each cluster contains single object. A probabilistic rule is then applied to merge centroid. Another probabilistic rule is also applied to create new centroid, to split the previously merged groups and to form two discrete groups.

To merge two boids into single centroid, the merging centroid rule is used. Initially, each group has single object in the database. When two boids are within the sight area of each other, the probability Pm_{xy} of two groups x and y are merge as a single group. The probability is proportional to the affinity amid the two boids.

$$Pm_{xy} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

When two boids merge, one boid get disappear from the environment. But sometimes it is possible that the merged boid may not belong to the same group and it may get divide from that group. To exclude that problem, a new rule is introduced to create a new centroid. When two or more objectes are grouped, there is a probability that one of them leaves the group and return to the environment to look for a new group.

The probability of an individual leaving its cluster is proportional to the difference between its affinity with the current group and its greater affinity with the other groups. Initially, the current affinity with its group (cag) is estimated by finding the affinity between the boid and the centroid of the group which it belongs. Then it calculates the affinity of this boid with the centers of all the other groups and the highest one is named as gag (greatest affinity group). The probability P_i of the boid leaving the group is directly proportional to the difference between cag and gag.

$$cag = \sqrt{\sum_{i=1}^{n} (b_i - c_i)^2}$$

$$gag = \sqrt{\sum_{i=1}^{n} (b_i - g_i)^2}$$

$$P_i = gag - cag$$

From the above formulas, b is the object that is represented by the boid, c is the centroid (average values) of the current group and g is the centroid of the group with greatest affinity with b. If the value of the object changes, this rule reflects this change and it may remove the object from its current group.

New Separation, Alignment and Cohesion Rules

According to the new rules, the three behavioral rules are redefined. The rules are redefined based on their affinity value. The rules are as follows:

- Separation rule is defined as the strength of the separation between two boids is inversely proportional to the affinity between the boids. If the value of affinity is less, the separation force between them is high.
- Alignment rule is defined as the degree of alignment varies with respect to the affinity between the boids. If the affinity value between the boids is high, the alignment between them is harder.
- Cohesion rule is defined as the strength of cohesion varies with respect to the boids affinity value. If the affinity value is high, the cohesion is stronger.

3.3 Finding Associative Patterns

An improved Apriori Algorithm is used to find the frequently visited web pages by the user. The Apriori algorithm is modified in assigning the minimum support value.

Apriori Algorithm

Apriori algorithm is a sequential pattern mining algorithm of association rule mining. It is initially used for Market Basket Analysis for finding the customer's purchasing behavior. The steps for finding association rules can be stated as follows:

As example of an association rule is: Contains (T,"baby food") → Contains (T, "diapers") [Support= 4%, Confidence=40%].

The interpretation of such rule is as follows:

• 40% of transactions that contains baby food also contain diapers.

• 4% of all transactions contain both of these items.

The calculations of the Support(S) and Confidence(C) are very simple:

$$\text{confidence}(A \rightarrow B) = \frac{\text{no of tuples containing both A and B}}{\text{No. tuples containing A}}$$

$$\text{support}(A \rightarrow B) = \frac{\text{no of tuples containing both A and B}}{\text{No. total no. of tuples}}$$

The main problem is to find all association rules that satisfy minimum support (min_sup) and minimum confidence (min_conf) thresholds, which are provided by user and/or domain experts.

Pseudo-code for Apriori:

Lk: Set of frequent itemsets of size k (with min support)
Ck: Set of candidate itemset of size k (potentially frequent itemsets)
L1 = {frequent items};
for (k = 1; Lk !=∅; k++) do
Ck+1 = candidates generated from Lk;
for each transaction t in database do
increment the count of all candidates in Ck+1 that are contained
in t
Lk+1 = candidates in Ck+1 with min_support
return ∪k Lk;

Improved Apriori Algorithm

The improved Apriori algorithm is modified with respect to support value. The candidate item sets are generated with respect to minimum support value and the time duration spent on each web page by the user. The time duration spent on each web page is an important consideration for identifying the user interest level while visiting the web pages. So, the support value is calculated as follows

$$\text{support}(A \rightarrow B) = \frac{\text{no of tuples containing both A and B and time duration spent on A and B greater than or equal to the specified time}}{\text{No. total no. of tuples}}$$

3.4 Markov Model for Web Page Recommendation

We have used Markov model to recommend the web pages. This model is used to identify the next pages based on the sequence of previously visited pages by the users. When a new user enters to obtain the suggestion of web page, the sequence path of that user is compared with the associative pattern and it would recommend the web page using the probability definition.

Table:1 Techniques used by Authors for Web Recommendation

Author Details	Techniques used for Web Recommendation		
Bamshad Mobasher,et.al., 2002	Collaborative Filtering approach	Web Usage Mining	Profile aggregation clustering
Yoon Ho Cho, et. al, 2002	Collaborative filtering approach	Web Usage mining	Decision tree induction and Association rule mining
Eirinaki, M, et. al, 2003	Web Usage Mining	Web Content Mining	Concept Logs
Olfa Nasraoui and Chris Petenes, 2003	Web Usage Mining	Hierarchical Unsupervised Niche Clustering	Fuzzy Approximation Reasoning
Dimitrios Pierrakos, et. al.,2003	Web Usage Mining	Machine Learning	User Modeling
Magdalini Eirinaki, et.al., 2004	Web Content Mining	Semantic Web	Domain Ontology
Feng-Hsu Wanga et. al.,2004	Collaborative filtering method	Web usage mining	Hierarchical bisecting clustering and Association rule mining
Bayao Zhou, et. al., 2004	Sequential Pattern Mining	SWARS Recommender System	Pattern Tree
Bayao Zhou, et. al., 2005	Web Usage Mining	Association Rule Mining	Formal Concept Analysis using Lattice Theory
Guandong Xu, et. al., 2008	Collaborative Filtering Approach	Latent Dirichlet Allocation model	Web Content, Structure and Usage Mining
Mohamed Koutheaïr Khribi, et. al., 2009	Content and Collaborative filtering approach	Web usage mining	Browsing History for e-learning
Rana Forsati, et.al., 2009	Distributed Learning Automata	Weighted Association Rule Mining	Hits algorithm
Sumathi, C., P., et.al., 2009.	Web Usage Mining	Aggregate Usage Profile	Recommendation for Online
Haritha Mehta, et. al.,2011	Collaborative based approach	Entropy based similarity measure	Web Content Mining
T.Venkata Ramana and K.Venugopala Rao, 2010	Web Content Mining	Semantic Web	User searching an navigational behavior
K. Suneetha and Dr.M.Usha Rani, 2012	Weighted sequential Pattern mining	Web Usage Mining	Markov model
Haibo LIU, et. al., 2012	Web logs	data structures such as, Class-Interest Matrix and Frequent-Path Matrix	Web-Interest Matrix, User-Interest Matrix
Qinjiao Mao, et. al., 2013	Web Usage Mining	Latent Dirichlet allocation	Recommendation Model
Florent Garcin, et. al., 2013	Web Usage Mining	Present browsing behavior	Context Tree Recommender system

4. RESULT AND DISCUSSION

The data set are taken from msnbc web server to implement our work. The performance of the result is evaluated based on the following three measures: (i) precision, (ii) applicability and (iii) hit ratio. We have used Java (jdk 1.6) which the system has i5 processor with 4GB RAM.

$$Precision = \frac{C^+}{C^+ + I^-}$$

$$Applicabil\,ity = \frac{C^+ + I^-}{|N|}$$

$$Hit\ ratio = Precision \times Applicabil\,ity = \frac{C^+}{|N|}$$

Where,

C^+ - number of correct recommendations
I^- - denotes the number of incorrect recommendations
N - denotes the total number of given requests.

Table.2. Performance of Improved Apriori Algorithm compared with Apriori Algorithm

Algorithm Name	No. of Candidate generation	Precision	Applicability	Hit ratio
Apriori	18	70.52	83.24	82.2
Improved Apriori	12	84.23	90.2	94.7

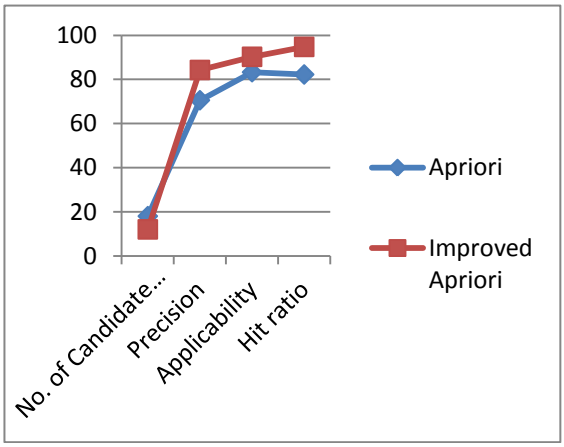

Fig. Performance Comparision of Apriori Algorithm with Improved Apriori Algorithm

5. CONCLUSION AND FUTURE WORK

Web usage mining as a tool for recommendation system. The traditional Apriori algorithm is improved by adding the time duration spent on each web page. Markov model is used for recommending the web pages based on user's past history. In future, FP-Tree algorithm with more improvements in using the minimum support value will be applied for finding the associative patterns for more accuracy.

6. REFERENCES

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

[2] Dimitrios Pierrakos, Ge Orgios Paliouras, Christos Papatheodorou and Constantine D., Spyropoulos. 2003. Web Usage Mining as a Tool for Personalization: A Survey. User Modeling and User-Adapted Interaction 13: 311-372, Kluwer Academic Publishers. Printed in the Netherlands.

[3] Shiva Nadi, Mohammad Hossein Saraee, Ayoub Bagheri. 2011. A Hybrid Recommender System for Dynamic Web Users, International Journal Multimedia and Image Processing (IJMIP), 1(1).

[4] Riecken, D. 2000. Personalized Views of Personalization. Communications of the ACM 43, (8) 27–28.

[5] Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D., and Sartin, M. 1999. Combining content-based and collaborative filters in an online newspaper. ACM SIGIR '99 Workshop on Recommender Systems, Berkely.

[6] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. 2000. Analysis of recommendation algorithms for e-commerce. Proceedings of ACM Ecommerce Conference. 158–167.

[7] Yoon Ho Cho, Jae Kyeong Kim, Soung Hie Kim. 2002. A personalized recommender system based on web usage mining and decision tree induction. Expert Systems with Applications 23, 329–342.

[8] Dimitrios Pierrakos, Ge Orgios Paliouras, Christos Papatheodorou and Constantine D. Spyropoulos. 2003. Web Usage Mining as a Tool for Personalization: A Survey. User Modeling and User-Adapted Interaction. 13: 311-372. Kluwer Academic Publishers. Printed in the Netherlands.

[9] Cooley, R., Mobasher, B., and Srivastava, J. 1999. Data preparation for mining world wide web browsing patterns. Journal of Knowledge and Information Systems.

[10] Mobasher, B., Dai, H., Luo, T., Sun, Y., and Zhu, J. 2000b. Integrating web usage and content mining for more effective personalization. Proceedings of the EC-Web 2000, 165–176.

[11] Srivastava, J., Cooley, R., Deshpande, M., and Tan, P. 2000. Web usage mining: Discovery and applications of usage patterns from web data. SIGKDD Explorations, 1(2), 1–12.

[12] Bamshad Mobasher, Honghua Dai, Tao Luo, Miki Nakagawa. 2002. Improving the Effectiveness of Collaborative Filtering on Anonymous Web Usage Data.

[13] Eirinaki, M., Vazirgiannis, M., and Varlamis, I. 2003. SEWeP: Using Site Semantics and a Taxonomy to Enhance the Web Personalization Process. SIGKDD '03, August 24-27. Washington, DC, USA. Copyright 2003 ACM 1-58113-737-0/03/0008.

[14] Olfa Nasraoui and Chris Petenes. 2003. An Intelligent Web Recommendation Engine Based on Fuzzy Approximate Reasoning. Proceedings of the IEEE International Conference on Fuzzy Systems - Special Track on Fuzzy Logic and the Internet.

[15] Magdalini Eirinaki, Charalampos Lampos, Stratos Paulakis, Michalis Vazirgiannis. 2004. Web Personalization Integrating Content Semantics and Navigational Patterns. WIDM'04, November 12-13. Washington, DC, USA. Copyright 2004 ACM 1-58113-978-0/04/0011

[16] Feng Hsu Wanga, Hsiu-Mei Shao. 2004. Effective personalized recommendation based on time-framed navigation clustering and association mining. Expert Systems with Applications. 27, 365–377.

[17] Baoyao Zhou, Siu Cheung Hui and Kuiyu Chang. 2004. An Intelligent Recommender System using Sequential Web Access Patterns. Cybernetics and Intelligent Systems. IEEE Conference on Cybernetics and Intelligent Systems.

[18] Baoyao Zhou, Siu Cheung Hui, Kuiyu Chang. 2005. A Formal Concept Analysis Approach for Web Usage Mining. Intelligent Information Processing II IFIP International Federation for Information Processing. 163, 437-441.

[19] Guandong Xu, Yanchun Zhang, and Xun Yi. 2008. Modelling User Behaviour for Web Recommendation Using LDA Model. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.

[20] Jia L , Osmar R., Zaïane. 2004. Combining Usage, Content, and Structure Data to Improve Web Site Recommendation. 5th International Conference on Electronic Commerce and Web Technologies (EC-Web).

[21] Mohamed Koutheaïr Khribi, Mohamed Jemni1 and Olfa Nasraoui. 2009. Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval. Educational Technology and Society, 12 (4), 30–42.

[22] Rana Forsati, Mohammad Reza Meybodi, Afsaneh Rahbar. 2009. An Efficient Algorithm for Web Recommendation Systems. 978-1-4244-3806-8/09.

[23] Sumathi, C., P., Padmaja Valli, R., and Santhanam, T. Automatic Recommendation of Web Pages in Web Usage Mining. (IJCSE) International Journal on Computer Science and Engineering 02(09),3046-3052.

[24] Harita Mehta, Shveta Kundra Bhatia, Punam Bedi and Dixit, V., S. 2011. Collaborative Personalized Web Recommender System using Entropy based Similarity Measure. IJCSI International Journal of Computer Science. 8(6),3, 1694-0814.

[25] Venkata Ramana, T., and Venugopala Rao, K. 2010. User Search Personalization in Semantic Web Mining, International Journal of Advanced Research in Computer Engineering and Technology. 1(3).

[26] Suneetha, K., and Usha Rani, M. 2012. Web Page Recommendation Approach Using Weighted Sequential Patterns and Markov Model. Global Journal of Computer Science and Technology. 12(9).

[27] Haibo Liu, Hongjie Xing, Fang Zhang. 2012. Web Personalized Recommendation Algorithm Incorporated with User Interest Change. Journal of Computational Information Systems 8(4), 1383-1390.

[28] Qinjiao Mao, Boqin Feng, Shanliang Pan,. 2013. Modeling User Interests Using Topic Model. Journal of Theoretical and Applied Information Technology. 48(1).

[29] Florent Garcin, Christos Dimitrakakis, Boi Faltings. 2013. Personalized News Recommendation with Context Trees. ACM Journal.

[30] Suguna, R., Sharmila, D. 2012. User Interest Based Web Usage Mining using a Modified Bird Flocking Algorithm, European Journal of Scientific Research (EJSR), Vol 86(2), pp. 218-231.