ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A WEB USAGE ASSOCIATION LEARNING METHODS BASED ON MINING TECHNIQUES FOR WEB PERSONALIZATION

Y.RAJU1, DR D. SURESH BABU2, DR.K.ANURADHA3
1Department of Computer Science and Engineering, ISL Engineering College, Hyderabad, India
2Department of Computer Science, Head, Kakatiya Government College, Kakatiya University, India
3Department of Computer Science and Engineering, GRIET, Hyderabad, India
E-mail: 1 raju.yeligeti@gmail.com , 2 sureshd123@gmail.com , 3 kodali.anuradha@yahoo.com

ABSTRACT

The increase in information resources on the World Wide Web allows users to find the information they need and navigate through multiple sites on the web. Because the web is huge and complex, users often unable to reach the lookout page when surfing the web. Web personalization is one of the potential conducts to solve this problem by leveraging the knowledge gained from the analysis of users accessing activities in the web usage logs to adapt the content and structure of the website to our needs. The existing approach focuses more on building user profiles that rely on web pages or documents that affect the effectiveness of web personalization. In this paper, we propose a web usage association (WUA) learning methods based on log usage association learning and personalized cluster mining technique for effective web personalization. The proposed method classifies the data using "frequent pattern mining (FPM)" and "Multi-Stage Association Rules (MAR) for the user's interest in navigation sites and personalization, and the chronic relationship of web usage using hierarchical methods and clustering. The Experimental evaluation has shown that the proposed approach has achieved effective personalization precision measurements for user interest and can be used in real-time personalization systems to minimize the storage cost and provide the provisioning for resources personalization in real time systems.

Keywords: *Web Usage, Association Learning, Log usage association learning, personalized cluster, Personalization.*

1. INTRODUCTION

The dependence of IR is growing with the growth of the Internet, and at the same time, the search engine has many difficulties in providing the most relevant and repeatable results for user queries. In common, search engines provide the identical search outcomes for different users while submitting the similar query without allowing for different information requirements and requirements. For the "information retrieval", the search engine retrieves information depends on client query input, however, the comeback result may ignore the intention page the user is looking for because of the extremely little range of query understanding. This means that results for different subtopics or semantics of the query are displayed mutually in the results record, which means that the user might need to explore a huge quantity of "irrelevant items" to find the item of importance. Nevertheless, because "search engines" are well-suited for search, but the search engine insufficient to recognize the user's intention after the search term, the search outcomes obtained are not all the time cooperative to the user browsing. A "common search engine" offers a comparable group of results exclusive of allowing for the user browsing intent [1] [2] [3]. This requires customized searches that can give accurate output to users with a higher relevancy.

The majority and significant concern that arises throughout the user personalization method is privacy infringement. Numerous users are unwilling to provide individual information on an implicit or explicit basis, do not hesitate to visit their website (if they know it exists), or disclose personal information on the registration form. In all cases, the user not unable to finds or uncertainty and knows that the entire setting is recorded and utilized exclusive of consent. Also, even if it agreed to provide "personal information" to the site with "cookie technology", this information may be substituted among sites and may be disclosed without our permission.

The W3C has recommended the changes recommended in data structure confidentiality as the "Platform for Privacy Preferences (P3P)" [26].

This average is "automatically retrieved" and "interpreted" with the utilized agent in a typical design that allocates Web sites to articulate their "privacy policies". The process of reading the privacy policy is automatically tailored to the data, because key information and personal exposure sites collected by the website that it automatically forward to users may differ from the method and consumer preferences [13].

Many traditional approaches have been extended to support web references such as "Web search", "collaborative filtering", and "hybrid context-based collaborative filtering approaches" [19] for web configuration [15], [20]. However, this approach suffers from the foremost disadvantage that mainly users browses the website anonymously through a proxy and find it difficult to hide and obtain their identity. Some adaptive personalization systems are developed based on user reaction or record interest. These systems are prolonged and users do not desire to utilize that approach. Current technologies are based on web usage [6], [7] aimed at finding motivating practice patterns resulting from data accumulated in "web server logs" [17].

Personalization in traditional approaches can be categorized into two main phases in the process of personalization as, *1) Data Collection Phase*, and *2) Learning Phase*. In the classification based on learning from data consists of "Memory Based Learning", and "Model Based Learning". The "Memory Based Learning" has a drawback of "requirement of enormous memory, Scalability issue, and flexible to changes", whereas, "Model Based Learning" has an advantage of due to "Limited memory requirement, Easily scalable, Learning phase offline and not adaptable to changes".

In the "Latent Variables Methods" variables that haven't been directly observed but have rather been inferred, hence have more become popular recently as a modelling approach in web usage mining. The two commonly used LVMs are, knows as "Finite Mixture Models", and "Factor Analysis". In the "Finite Mixture Models" the use a finite number of components to model such as page view, or user rating. whereas in the "Factor Analysis" it aims to summarize and find relationships within observed data and being utilized in pattern recognition, collaborative filtering and personalization based web usage mining.

The tradition web usage based models have a drawbacks of due to the process relies on user transactions or rating data. The new items or pages navigated are therefore never learned and also it does not utilize the knowledge from underlying domain and so cannot make more effective personalized result as required. To enhance the advantage of model based learning we present a novel "web personalization" is proposed through learning the web usage association (WUA) based on log usage association learning and personalized cluster mining methods in this paper.

This following paper organized in four sections. The Section-2 presents the correlated works, Section-3 discuss the architecture and methods of the proposed work, Section-4 presents experiment result in analysis and Section-5 present the conclusion of the paper.

2. BACKGROUNG STUDY

A personalized web is a search for information that takes into account each user's interests [22]. As competition in the search market has accelerated, various search engines have launched customized search services. For instance, "Google's Custom Search" allows users to specify the categories of WebPages they're interested in. Some web personalization systems [23], require users to register demographic information early [25] to use relevance feedback or to provide better services to address related requirements. This approach is necessary because users must participate in additional activities to manually assign preferences rather than search. It need to develop an approach that is implicitly aware of the information needs. As the need for a personalized web grows, much research is needed to provide relevant information in consideration of the user's situation.

2.1 Web Usage Mining

Web-usage mining goes beyond keyword or single-user approaches and makes global questions about the behavior of user groups. Various methods and algorithms have been developed to provide reasonable recommendations to users, from simple keyword-based frequency counters to collaborative / content filtering techniques. Studies have shown that integrating the context and content of system and user data can improve the accuracy of recommendations. In recent years, there have been studies that use semantic knowledge systems to improve accuracy. Everyone is a unique person, but when many people access certain resources, they will see patterns that represent unique group behavior. If you can simply identify a group of behaviors, you are more likely to recommend webpages of interest with higher accuracy. Therefore, in this study, we aim to search for such behavior patterns and improve patterns using semantic information and order information and navigation logs existing on the page.

2.2 Web-Usage Learning Approaches

In the web-enabled mining sector, there are a variety of approaches that use a variety of technologies to choose how to cluster and how to cluster. The most common method is to cluster sessions. The study given in [24] suggests how to cluster user clickstream data by defining each user session as a path in the topology graph of the Web site. In [18], each user session is defined as a binary list of size N, where N is a list of all possible pages on your website. Similarity features are introduced according to the topology of the website. The combination of the variance of cosine similarity and this new similarity metric is used in the modified CARD algorithm to cluster sessions. When a new user session is introduced, a vector is also created for that session, and the closest cluster is selected based on the distance metric used for clustering. Then, a page subset of the cluster depending on the frequency cluster.

In [14] applies similar ideas to multivariate k-means clustering. In [26], the authors propose a potential basic approach while clustering data. The proposed technique generates a session pageview matrix by clustering URLs and directly assigning pageviews to the corpus. Unlike other approaches, pages are assigned a weight based on their visit time. The research presented in [15],[16] integrates content mining into web mining mining to ensure that the content of the page view is integrated with the usage profile.

Unsupervised algorithms have also been proposed in this area. The Self Organizing Map (SOM) algorithm [9] has been adopted by [3]. The author has experienced that SOM creates clusters that are stricter than k-means. An interesting approach is to use artificial ant colony clustering and linear genetic programming (LGP). First, the raw data are input into the ACLUSTER scheme proposed in [22], and the result is used as input to the LGP. This is a repetitive approach and must be broken at some point.

2.3 Related Works

Julia Hoxha et al. [9] solves the difficulty of suggesting resources from diverse domains by combining the semantic content of these resources with user "browsing behavior patterns". Overcome the be short of overlap among domains with developing associations supported on investigated the semantic content of web sources. If the user is at this time presentation a particular page, it will suggest an approach to applying a support vector machine to learn the relevance of the resource and to predict what is best for the user to recommend. We learning the impact of the organization on produce accurate suggestions from the actual data set of semantically rich logs of user browsing performance across "multiple Websites" and accomplish experiments to exhibit the usefulness of our approach.

Jie Yu et al. [15] described a "mining user context" based on "personalized interactive computing" for the Web. This explains how personalized searches are an achievement approach of doing the identical query and how to accomplish its "real-time information" needs as a main topic in "personalized search". Kim Han-joon et al. [14] described a conceptual network-based user profile for a personalized Web to build. Describes an "innovative approach" to construction utilize profiles for thought networks for "personalized searches". User profiles are characterized by a conceptual network, and each concept is roughly represented by the "Formal conceptual analysis (FCA) theory" [27].

Dong Zhou et al. al. [1] suggests a new model for organizing enhanced user profiles using outside corpus for "personalized query expansion". This representation put together the existing "state-of-the-art textual representation learning framework", such as "word insertion", through the "topic models" in the two pseudo-alignment document groups. Build two new query expansion technologies based on user profiles. These two techniques are based on topic relevance between topic-weighted word insertion and search terms within a user's profile. An in-depth experimental evaluation of two actual data sets using different outer corpuses showed that our approach is superior to existing technologies, including traditional "non-personalized" and "customized query expansion" methods.

F. Akhlaghian et al. [10] suggested a "personalized search engine using an ontology-based fuzzy concept network". The proposed proposal personalizes search engine output with the help of an "automatic fuzzy concept network". The most important purpose is to utilize the notion of ontology to advance the general "fuzzy concept network" [18] constructed according to the user's profile. C. Biancalana et al. [21] recommend an innovative method for "personalized Webs using social tagging in query expansion". The "Social networks" and "collaborative tagging systems" are rapidly recognized as the mainly significant means of data log usage association learning and sharing. Users can tag bookmarks to easily distribute information and visit later.

Xuan Wu et. al. [8] explore "multiple semantic relationships in social tagging systems", including between tags, between words, and between tags and words. Three similarity graphs are created based on the tags and functions derived from the word. It also standardizes the smoothness of multiple associations through three similar graphs, incorporating physician-related feedback information from top-level documents. The objective of this work is to improve the customized search results by considering the three similarity graphs above as a new query expansion model. Experiments performed on real data sets validate the proposed approach.

Web personalization can improve the speed and accuracy of web information retrieval instantly. Because users can have diverse interests and expectations about pose queries, search engine personalization can yield better results based on a user's profile or usage, and improved counterpart the general importance of individual users [4], [5], [16].

3. PROPOSED WEB USAGE ASSOCIATION LEARNING METHODS

We propose a "Web Usage Association (WUA) learning method based on log usage learning and personalized cluster mining techniques for effective web personalization [12]. The framework of "WUA learning" is shown in Fig. 1. It describes two mechanisms for personalization through processing the "Log Usage Association Learning" and "Personalized Results Creation" to implement user web usage processing and result personalization.

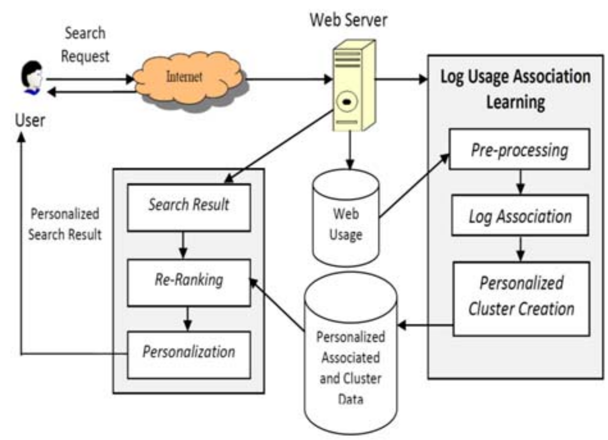

Figure -1: Web Personalized Framework

The mechanism of the personalization builds through the activity when the user performs a search demand query to the Web server, which carries outs a search process and sends the search results for the personalization. The "web server" records all user activity in the "web usage log". User access log records include user "IP address", "request time", "requested URL", and "status code". The use of logged logs configures the cluster data model by means of "hierarchical cluster methods" and "frequent pattern mining log" usage association learning. The "log usage processing" is a "background process" that periodically keep informed cluster data. This approach decreases processing costs in real time. The collected data is used for "ranking" and "personalization results".

3.1 Log Usage Association Learning

The processing of web usage logs is made to discover the information stored in web server log files. Through utilizing "statistical" and "data mining methods" to web usage data, we can recognize the relationship patterns of a user's interest behavior with the web pages, and correlate them with website, page, and user. Weblog usage process carries out in three different phases as: "Pre-processing", "Log Association Learning" and "Personalized cluster" to do the data preparation, the log pattern creation and relationship identification and construction of user-personalized cluster creation.

A. Pre-processing

The Web server logs provide the information of the "Web page access" in the server log file. The "Web log data" is pre-processed to recognize "users", "conversations", "page views", and so on. The server log is a simple text file that records events on the server. Each row in the log file represents a request. If a "visitor requests" a page that includes images, then along with link info these image info also be added to the log file.

Preprocessing helps to improve the quality of data by eliminating the noise and inconsistency in the log file. It clears Web log data from web-based access to irrelevant entries (such as errors, graphics, script files, and so on). The process parses the log file and converts the data to normalize so that the log uses associative learning and personalized clustering.

The Pre-processing of web logs is generally complex and time-consuming. It performs four major tasks to create the data required for association learning as, : *(i) "data cleaning", (ii) "Session identification", (iii) "retrieving of information about page content and structure"* and *(iv) "data formatting".*

- The data cleanup step consists of removing all items / data from the web logs. For example, it is not relevant to the mining process and is not useful. Requests for graphical page content (such as jpg and gif images) or robots and web spiders are considered irrelevant and useless. Robots and related items that are not related to the web spider can be found by looking at the user agent or checking the text file. robots.txt. A "heuristic-based approach" can be used when a robot sends incorrect information, such as an invalid user agent, to an HTTP request. In this approach, user sessions and robot sessions are separate.

- The session identification step includes identification of different user sessions. This session is identified using an incomplete information form weblog. Using a proxy server introduces caching issues that affect session identification. This allows you to reorganize sessions using search-oriented heuristics, time-driven heuristics, or using cookies. In some situations, cookies do not solve the problem. In this case, the URL is rewritten by including the session-id in the original URL. Therefore, the web log will contain the modified URL instead of the original URL. The main problem with this solution is to insert a software agent on the server side to perform these tasks.

- Web browser caching affects the creation of consistent paths. Web users visit the previous page by clicking the Back button multiple times. However, web logs cannot contain this information. You can use a heuristic approach to reconfigure a consistent navigation path. In many web-based mining applications, the visit URL is used as a major source of information for mining purposes. In addition to URLs, you can perform web page categorization based on the content type of the web page. This classification is then used during mining operations. If sufficient classification is not possible, you can build using Web structure mining.

- The final step in preprocessing is to properly format the data and then provide the formatted data for mining purposes. Data can be formatted in a variety of ways. Use a relational database to store data extracted from web logs, a signature tree to index logs, or a tree to store access sequences. Even structures such as cubes can be used to store session information.

B. Log-Association Learning

The log usage association learning is applied to the data mining method using the detection mode. It helps to extract relevant data according to some rules.

In [3] a personalized diversification of search results is ranked using methods that are not implicitly or explicitly supervised. It has adopted a different perspective on this issue based on structured learning. In [9] the learning of semantic content of the resources is used together with patterns of user browsing behavior to recommend resources from various domains. If the user is currently viewing a particular page, it suggest an approach to apply support vector machines to learn the relevance of resources and to predict what is most relevant to recommending to users.

The general idea in [5] is to extract domain-level objects from user sessions and aggregate them according to the weight and merge functions to

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

create a user profile for each user. Suppose that a domain level ontology for a website already exists and the merge function is easily defined for all attributes of the object. Another study, given in [12], combines the Apriori algorithm with item and attribute insertions to extract frequent patterns. There are studies that use sequence information in clustering [21],[25]. In [21], sequence-based clustering of sessions is used in conjunction with the Markov model. Similar to [25], sequence information is used to cluster similar click streams to determine the user's navigation behavior.

We aim to propose multi-phase association learning strategies for personalized diversification of to construct the pattern. The multi-phase association rule mining is used to construct the pattern of preprocessing data on the proposed framework. The multi-stage association rule can effectively exploit the log data using the conceptual hierarchy calculation based on the "support" and "confidence" of each log data.

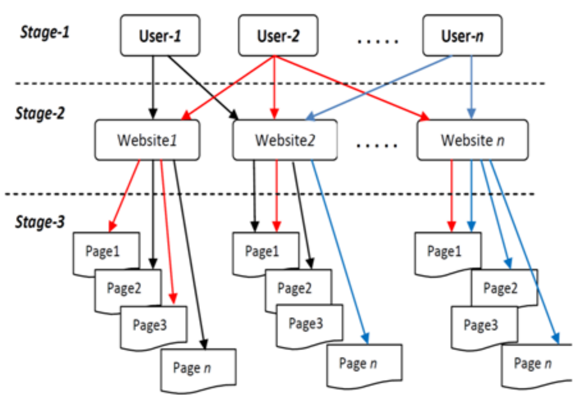

Figure -2: Multi-Stage Association Through Association Learning S Of Users Log Data

In general, the "top-down strategy" as shown in Fig. 2, is used to calculate the count of associated itemsets at every "concept level", initiating at concept stages "1, 2, and 3", and at the stage down to further precise concept levels during no additional correlated item set are to be found.

C. Personalized cluster Creation

The process of personalized clustering is divides data into groups based on the "Inter-cluster similarities", and "Intra-cluster similarities". In relate to users-based clustering it represent a user segments based on common navigational behaviour through associating a groups of pages or items that are commonly accessed or purchased together. The degree of association rules involves in frequent itemsets which satisfying a minimum support threshold and a minimum confidence threshold.

A Personalized clusters treat the categorized data as similar to each other within the same group and in a different group than the other data in the group. Data clusters can be considered as a form of data compression because they can be treated as a group collectively. Using Logs Association learning is an efficient means of grouping, except it necessitate the appropriate gathering and classification of many sets of patterns that a classifier utilizes to model every group. An illustration is being shown in Fig. 3, where *P* represents as web pages.

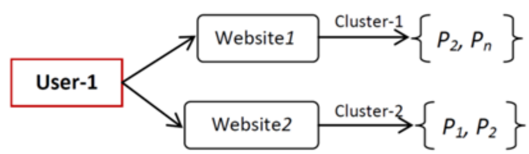

Figure-3: Personalized Cluster Of User1 Log Data Based On The Above Log Usage Association Learning

The framework implements "hierarchical clustering methods" to group data top-down into a cluster tree. Starting with the top-level item in the cluster, this "top-down strategy" breaks the cluster into smaller, smaller pieces, until each item forms its own cluster or meets certain exit conditions. The acquired cluster patterns are stored for further analysis in the process of re-ranking the web personalization.

3.2 Personalized Result Creation

The framework implements a "re-ranking algorithm" to personalize search results. A "search query, *Q*", demanded by a "user as, *U*", to retrieve search results. We suppose that the "cluster pattern data set as, C_d", such as the user *U*, is accumulated in the framework's "cluster database" [11].

To calculate the "re-ranking for personalization", we calculate how often the user's "website is visited frequency as, W_{freq}", "link the access frequency as, A_{freq}", and the "average personalization value as, P_{avg}" using "W_{freq}" and "A_{freq}".

Let's consider that a collection of results as "S_R" consist of top 10 results has retrieved representing as, "R_1 . . . . . R_{10}" and each outcome has relationship reference to a site as "L_1 . . . . . L_{10}" as on pretending as a "query, Q" to the search engine.

To estimate the result of the "W_{freq}" for the retrieved "S_R", we need to find the summation of the frequency of the URL as "t" being visited by the user utilizing the personalized cluster created with the total distinct pages visited as,

$$W_{freq} = \frac{\sum_{i=0}^{n}((t_i \in C_d) \rightarrow U)}{distinct\left(((t \rightarrow C_d) \neq t_i) \rightarrow U\right)} \quad (1)$$

and, in order to calculate "A_{freq}", we need to find the summation of the frequency of URL as "t" and links as "L", from the personalized cluster against the total distinct link visited by the user,

$$A_{freq} = \frac{\sum_{i=0}^{n}((t_i,\ L_i \in C_d) \rightarrow U)}{distinct\left(((t,L \rightarrow C_d) \neq t_i,\ L_i) \rightarrow U\right)} \quad (2)$$

In order to personalize the ranking value, we use "W_{freq}" and "A_{freq}" as to calculate the personalize average "A_{freq}", as,

$$P_{avg} = \frac{(W_{freq} + A_{freq})}{2} \times 100 \quad (3)$$

The personalize average "P_{avg}" based on getting each result will be sorted individually. The personalized results are sent to the user as a search response. This reordering method will be valid because the user uses the log as a collection of all browsing activities. Users may visit the site directly instead of a web search. The proposed method uses search and direct access logs to build personalized results, allowing users to search for web pages and meet the desired interests.

4. EXPERIMENT EVALUATION AND RESULTS

In order to determine the usefulness of the proposed method, we measured the personalize "precision (PP_R), recall (PR_R), accuracy (PA_R) and error rate (PE_R)" of the result as the based on the equations given below.

$$PP_R = \frac{|No.of\ \ True\ result\ |}{|Total\ No.of\ Associated\ Result|} \quad (4)$$

$$PR_R = \frac{|No.of\ \ True\ result\ |}{|Total\ No.of\ Related{+}Associated\ Result|} \quad (5)$$

$$PA_R = \frac{|Total\ No.of\ Associated\ Result\ |}{|Total\ No.of\ Search\ Result|} \quad (6)$$

$$PE_R = \frac{|No.of\ \ True\ result\ |}{|Total\ No.of\ Search\ Result|} \quad (7)$$

Based on the quantitative results to determine the personalized results measurement towards re-ranking.

We assume that the "search result as, S_R", the "personalized result as, P_R", given a set of "personalized cluster data as, C_D". If "P_R (r) ≠ SR (r)" and "$P_{avg} > 0$", where "r" is the outcome record.

The experiment was performed through installing the "Web proxy" in the server to log the user log. We have carried out cyclic searches, through directly access various websites to download links related music files. In order to build a cluster data model, we run a background java program which does processing to execute the "Usage Association Learning" to construct the required the "Personalized cluster using the hierarchical clustering method to achieve the Personalized Result Creation.

In order to evaluate the proposed method, we selected three users and each user-submitted multiple search query related to "*download music*" to the popular Google search engine. We bring together the "top 10 results" from the search engine and execute a "re-ranking algorithm" for every result to calculate "W_{freq}", "A_{freq}" and "P_{avg}".

We create a "personalized result" depend on the "P_{avg}" value by rank and sends to the user as a search result. This practice is repetitive for every user when submitting a search query. Since "cluster pattern generation" is a continuous program running at periods, all searches and searches will increase the user's web usage log to generate a more relevant pattern for the user. The enlarge in user cluster patterns instantly improves the accuracy of personalization and user concern.

4.1 Personalized Result

Fig. 4 and 5 shows the user's "Personalization Precision" and "Personalized

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Recall" rate with the increasing number of clustered pattern. Fig. 4 depicts the Precision improvements with increasing Personalized cluster data utilizing the web usage logs as
input for user personalization of web and suggested access effectiveness. It shows the as the number of correct positive predictions divided by the total number of related results predictions. The best precision is 1.0, whereas the worst is 0.0. The measure shows an good quality of improvisation from an avg. of 0.8 and increases to 0.98 with increasing number of clustered patterns.

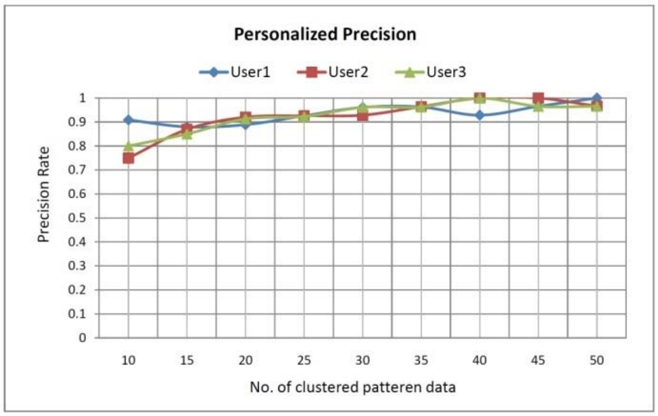

Figure-4: Personalized Precision percentage

Whereas, Fig. 5 depict the recall rate calculated as the number of correct positive predictions divided by the total number of related and associated results. It also shows the improvisation with increasing number of clusters patterns.

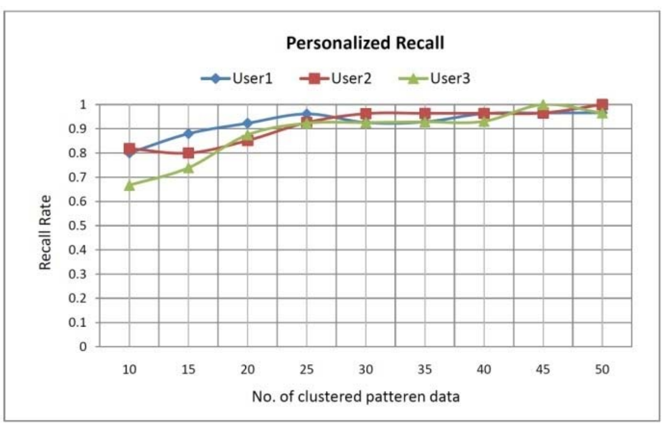

Figure-5: Personalized Recall percentage

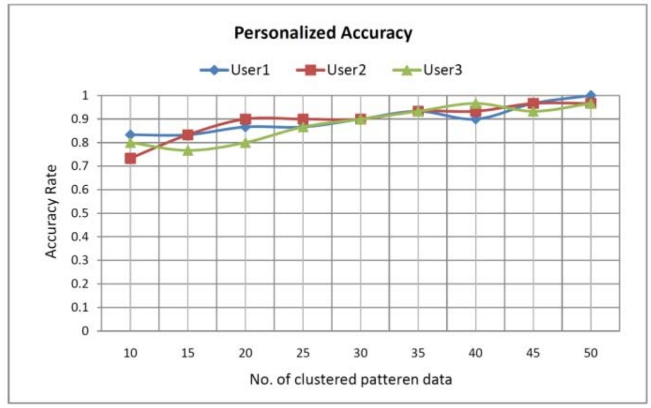

Figure-6: Average Personalized Accuracy Rate

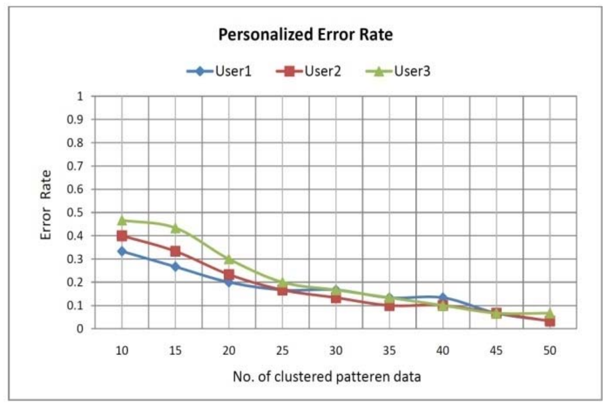

Figure-7: Average Personalized Error Rate

To measure the impact of the clustered patterns over the personalization we measure the personalized "accurate and error rate" as shown in Fig.6 and Fig. 7 based on the obtained result against the query.

Fig.6, shows the personalization accuracy rate, it measured based on the number of all correct results predictions divided by the total number of the data records. It illustrate that with increasing number of personalized clusters improve the results accuracy. Fig.7, shows the average personalization error rate, it illustrate that with increasing number of personalized clusters can minimize the error rate and improvise the result accuracy.

Based on the evaluation result of Fig.4 , 5, 6 and 7 describes the effectiveness of the creation of the personalized cluster for the accurate information extraction. To further, evaluate the performance of the proposal we perform a comparison with the existing personalization approach and measures the precision, recall, accuracy and error rate results as presented in the following section.

4.2 Comparison Results

In order to evaluate the proposed method is effective or better than other methods, we compare with the "cross-domain recommendation (CDR)" proposed in [9] and a fuzzy conceptual based documents personalize proposed in [10]. The "CDR"[9] the evaluation of the semantic recommendations that take advantage of a combination of new ways of semantics of web pages and usage patterns inherent in navigation logs. It first analyzing the behavior of multiple Web sites using structured markup data and later the impact of a structure that predicts related personalized recommendations for various domains for Web users.

In [10] the constructs user personalization based on fuzzy concept network based on the user's profile and in [9] implement "Support Vector Machines" for learning the significance of sources and forecast which ones are the nearly appropriate to suggest to a user.

We define a User-1 profile that includes "30 relevance concepts" related to music. For the proposed approach, we generated "30 personalized clusters" using generated web usage logs. The retrieval of "precision rate" and "recall rate" for both the approaches are being shown in Fig. 8 and 9.

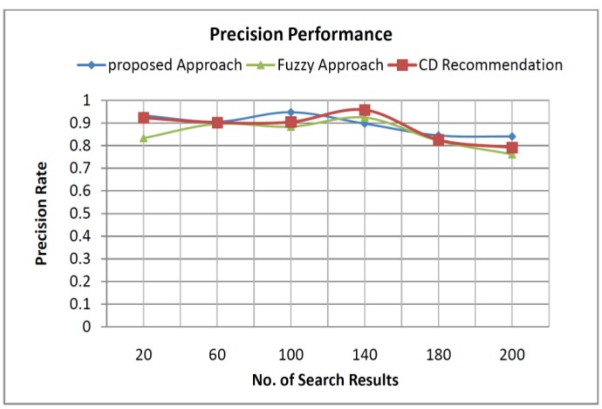

Figure-8: Precision Comparison

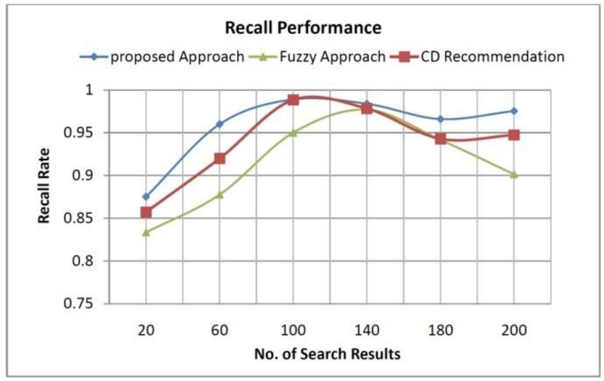

Figure-9: Recall Comparison

Fig. 8 and 9 shows the comparison results between proposed and the "CDR" and "Fuzzy approach". It shows an improvisation of the precision and recall rate due to the dynamically generated web usage cluster as in the profile-based configuration of user personalization using the "fuzzy concept network". And increasing the number of search results by defining the limits of the user interest concept.

However, in the case of "SVM-based CDRs", learning similarities suggests similar improvisation compared to the proposed approach and continues to be effective and effective. As the web data grows multidimensional, it can be concluded that the proposed approach is more suitable for feature personalization based on user-used log data clusters, since the personalization-based profile definition is less accurate compared to the proposed approach.

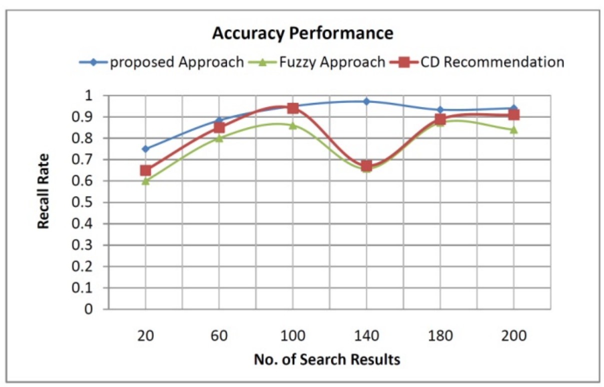

Figure-10: Accuracy Comparison

Fig. 10 and 11 shows the accuracy and error rate comparison results between the proposed, "CDR" and "Fuzzy approach". The results depicts that with increasing number of search results accuracy is improved and error rate decreases. It is because of the support of the designed clustered personalized being proposed in the approach shows low error rate in comparison to others. The overall interpretation of the comparison results in Fig.8, 9, 10 and 11 suggested the improvisation of the proposal.

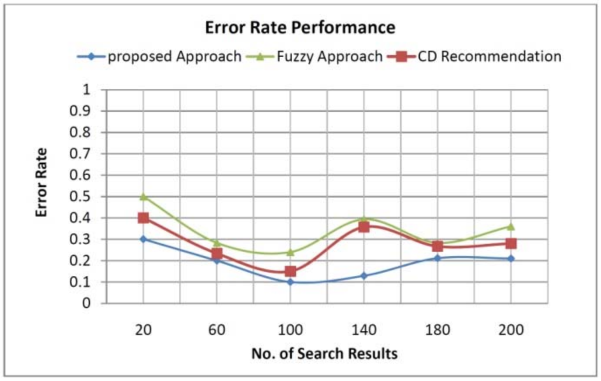

Figure-11: Error Rate Comparison

5. CONCLUSIONS

Web personalization can only use a short fuzzy query to meet user information requirements. The basic difficulty lies in information retrieval and personalized search, for each user to provide customized search results. In this paper, we present a novel Web usage log using associative learning and personalized clustering methods for effective Web personalization, which is used to study how the search can be personalized to precisely recognize the user's "Web usage log data". Personalization requires the pre-provision of support resources to reduce the cost of individuality. The experiment shows an impromptu representation of the percentage of personalized precision, indicating that the Web usage log can be used for user site personalization and user interest for user input.

REFERENCES

[1]. D. Zhou, Xuan Wu, Wenyu Zhao, Séamus Lawless, Jianxun Liu, "Query Expansion with Enriched User Profiles for Personalized Search Utilizing Folksonomy Data" IEEE Transactions on Knowledge and Data Engineering Volume: 29, Issue: 7, Pages: 1536 - 1548, 2017.

[2]. C. Chen, Xiangwu Meng, Zhenghua Xu, Thomas Lukasiewicz, "Location-Aware Personalized News Recommendation With Deep Semantic Analysis", IEEE Access, Volume: 5 Pages: 1624 - 1638, 2017.

[3]. S. Liang, Fei Cai, Zhaochun Ren, Maarten de Rijke, "Efficient Structured Learning for Personalized Diversification", IEEE Transactions on Knowledge and Data Engineering Volume: 28, Issue: 11 Pages: 2958 - 2973, 2016.

[4]. L. Wang, Bin Wu, Juan Yang, Shuang Peng, "Personalized recommendation for new questions in community question answering", IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) Pages: 901 - 908, 2016.

[5]. T M Veeragangadhara swamy, G T Raju, "A Novel Prefetching Technique through Frequent Sequential Patterns from Web Usage Data", COMPUSOFT An international journal of advanced computer technology, vol. 4, no. 6, June 2015.

[6]. M. Dhandi, Rajesh Kumar Chakrawarti "A comprehensive study of web usage mining", Symposium on Colossal Data Analysis and Networking (CDAN) Pages: 1 - 5, 2016.

[7]. C. Ramesh, K. V. Chalapati Rao, A. Govardhan "Ontology-based web usage mining model", International Conference on Inventive Communication and Computational Technologies (ICICCT) Pages: 356 - 362, 2017.

[8]. X. Wu, Dong Zhou, Yu Xu, Séamus Lawless, "Personalized query expansion utilizing multi-relational social data" 12th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP) Pages: 65 - 70, 2017.

[9]. J. Hoxha, Peter Mika, Roi Blanco, "Learning Relevance of Web resources across Domains to make recommendations", 12th international conference on Machine Learning and Applications, vol. 2, pp. 325-330, 2013.

[10]. F. Akhlaghian, B. Arzanian and P. Moradi, "A Personalized Search Engine Using Ontology-Based Fuzzy Concept Networks", International Conference on Data Storage and Data Engineering (DSDE), Pages. 137 - 141, 2010.

[11]. M. Sarwat, Raha Moraffah, Mohamed F. Mokbel, James L. Avery, "Database System Support for Personalized Recommendation Applications" IEEE 33rd International Conference on Data Engineering (ICDE) Pages: 1320 - 1331, 2017.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[12]. H. Yilmaz, P. Senkul, "Using ontology and sequence information for extracting behavior patterns from web navigation logs", Data Mining Workshops (ICDMW) 2010 IEEE International Conference on, pp. 549-556, Dec. 2010.

[13]. K.W.-T. Leung, D.L. Lee and Wang-Chien Lee, "Personalized Web search with location preferences", IEEE 26th International Conference on Data Engineering (ICDE), Pages. 701 - 712, 2010

[14]. H.-joon Kim, Sungjick Lee, Byungjeong Lee and Sooyong Kang, "Building Concept Network-Based User Profile for Personalized Web Search", 9th International Conference on Computer and Information Science (ICIS), Pages. 567 - 572, 2010.

[15]. J. Yu and Fangfang Liu, "Mining user context based on interactive computing for personalized Web search", 2nd International Conference on Computer Engineering and Technology, Vol.-2, Pages. 209-214, 2010.

[16]. J. Lai and B. Soh, "Personalized Web search results with profile comparisons," 3rd International Conf. on Information Technology and Applications-2005, Vol. -1, Pages. 573 - 576, 2005

[17]. H. Liu and V. Keselj, "Combined mining of Web server logs and web contents for classifying user navigation patterns and predicting users future requests", Data and Knowledge Engineering, Elsevier, 2007.

[18]. K.-Joong Kim and Sung-Bae Cho, "A personalized Web search engine using fuzzy concept network with link structure", Joint 9th IFSA World Congress and 20th NAFIPS International Conference, Vol. 1, Pages. 81 - 86, 2005.

[19]. X. Pan, Zhengcheng Wang and Xinjian Gu, "Context-Based Adaptive Personalized Web Search for Improving Information Retrieval Effectiveness", International Conference on Wireless Communications, Networking and Mobile Computing, Pages. 5427 - 5430, 2007.

[20]. D. Pierrakos, G. Paliouras, C. Papatheodorou, and C. D. Spyropoulos,"Web Usage Mining as a Tool for Personalization: A Survey", User Modeling and User-Adapted Interaction, 13(4):311-372, 2003.

[21]. C. Biancalana and A. Micarelli, "Social Tagging in Query Expansion: A New Way for Personalized Web Search", International Conference on Computational Science and Engineering (CSE '09), Vol. 4, Pages. 1060 - 1065, 2009.

[22]. F. Liu, C. Yu and Weiyi Meng, "Personalized Web search for improving retrieval effectiveness", IEEE Transactions on Knowledge and Data Engineering, Vol.-16, No.-1, Pages. 28 - 40, 2004.

[23]. M. Eirinaki and M. Vazirgiannis,"Web Mining for Web Personalization", ACM TOIT, 3(1):1-27, 2003.

[24]. M. Kutub, R. Prachetaa and M. Bedekar, "User Web Search Behaviour," 3rd International Conference on Emerging Trends in Engineering and Technology, Pages. 549 - 554, 2010.

[25]. N. M. Khan and Shahbaz, "Using Personalized Web Search for Enhancing Common Sense and Folksonomy Based Intelligent Search Systems", International Conference on Web Intelligence (IEEE/WIC/ACM), Pages. 423 - 426, 2007.

[26]. J. A. Konstan, B. N. Miller, D. Maltz, J. Herlocker, L. Gordon, and J. Riedl. Grouplens," Applying Collaborative Filtering to Usenet News", Communication of ACM, 40(3):77-87, 1997.

[27]. B. Ganter and R. Wille. Formal Concept Analysis: Mathematical Foundations. Springer-Verlag, 1997.