International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 3 (2015) pp. 6217-6232

http://www.ripublication.com

Markov Chain Monte Carlo Dynamic Clustering For Effective Search On Web Using Personalized User Behaviors

S.Sadesh[1] and Dr.R.C.Suganthe[2]

[1]Research Scholar and Assistant Professor,
Velalar College of Engineering and Technology, Thindal, Tamil Nadu, India
[2]Professor, Kongu Engineering College, Perundurai, Tamil Nadu, India

Abstract

Web data mining is used for several purposes including personalization, enhancement on web designing system based on each individual needs. Currently, personalized search is concentrated on addressing the challenge in web search community based on individual user preferences. However, similar preferences of users are not grouped effectively that adversely affects the results on online searching behavior. Therefore, retrieval of results based on the user behavior still remains a challenging task to be addressed. The recent studies on web based user search reveals that a clear input about the personal preferences are not provided in an automatic manner for effective clustering operation. To collect similar user behavior profiles and perform the automatic updation operation, a method called, Markov Chain Monte Carlo Dynamic Clustering based on the user behavior profiles (MCMC-DC) is designed. The main goal of the method, MCMC-DC is to group similar user behavior even on dynamic updation of user profiles. The method MCMC-DC is divided into three steps. Initially, Markov Chain in MCMC-DC is designed to develop an equilibrium distribution function to obtain different user behavioral profile with similar query search. Secondly, Dynamic Clustering is introduced with the equilibrium distribution function that reduces the updation time, and as a result, the running time gets minimized. The Dynamic Clustering in MCMC-DC method reviews the search histories and reversible-jump concept is introduced. Third step, briefly describes about the reversible-jump concept. The main goal of the reversible-jump concept is to review the history with updated profile and move to appropriate clusters by removing the replica information. Experiment is conducted on factors such as accuracy, similar user profile clustering rate, average running time of user queries.

Keywords: Dynamic Clustering, Web Data Mining, Reversible-jump Concept, User Profile Information, Markov Chain, Equilibrium Distribution

1. Introduction

With the increasing Web content, the most important and significant fact is the identification and retrieval of the users that matches accordingly to their needs. Therefore, the most essential requirement nowadays providing an effective search model based on personalized user behavior to enhance the applicability of Web site. Significant research works have been conducted on personalization of web. Amount of Invested Mental Effort (AIME) [1] also paved the way for information retrieval using pre-search and post-search questionnaire. However, user similar preferences were not grouped together that in a way highly affected online searching behavior results. Personalized Ontology Model (POM) [2] introduced representation of knowledge with the aid of multidimensional ontology mining. But, the results on web based user search provided a report that a precise user personal preferences were not provided in an automatic manner.

One of the pivotal roles played in increasing the Web intelligence and search model is the basis of knowledge and providing supporting information integration. The data sources on the Web were integrated with DBpedia [3] in order to extract information in a structured manner. This integrated model resulted in increased search results however, multilingual databases remained unaddressed. Wikipedia Miner Toolkit [4] with Java API not only provided parallelized processing but also improved the efficiency of search. But with the increasing Java APIs, human effort also increased. To reduce the human efforts, Bayesian approach was introduced in [5] extracted the information in an automatic manner.

In today's scenario dynamically generated databases are increasingly available on the web that produces dynamically generate Web pages for the queries provided by the users. A method called as Unsupervised Duplicate Detection (UDD) [6] was designed where the weights were assigned in a dynamic manner and was also not sensitive to false negative cases. But the time complexity increased due to the handling of both true and false negative cases. To address this problem Concept-based user profiles was designed in [7] that aimed on search engine personalization using query clustering model. However, relevance feedback remained unaddressed. To solve this issue, content-based image retrieval [8] was designed for effective content retrieval on web.

In order to perform effective web personalization, auto completion is one of the most significant services that the users communicate with as they form a part of queries. A new labeling strategy was introduced in [9] to improve the mean reciprocal rank. But the factor privacy was not introduced. To address the issues related to privacy in [10], user behavior was analyzed on Facebook and Gmail. Web Log Expert [11] was used to analyze the user behavior patterns on Internet with the aid of Web Access Logs and Log Analyzer. This greatly paved the way for analyzing the user behavior in relatively lesser amount of time.

Based on the aforementioned techniques and methods, we propose a Markov Chain Monte Carlo Dynamic Clustering based on the user behavior to collect similar user behavior profiles and perform the automatic updation operations. The framework collects similar user behavior profiles and performs automatic operation using Markov Chain Monte Carlo Dynamic Clustering. Using Markov Chain different user behavior profiles were developed with the aid of equilibrium distribution function. Finally, dynamic clustering and reversible jump is introduced to remove the replica information based on the review of search histories. Extensive experiments showed that our web personalization model increased the accuracy of automated updated profiles and increasing the similar user profile clustering rate with lesser average running time of user queries.

The remainder of this paper is organized as follows. In Section 2, an overview of related work on web personalization. Section 3 analyzes the problem and provides solutions accordingly. Section 4 explains our proposed web personalization model based on user behavior profiles and Section 5 includes the experimental results. In Section 6, concluding remarks are included.

2. Related Works

The process of segmentation of queries in Web search engine involves the breaking down of user queries into single elements with the purpose of accompanying the process of query retrieval. With the application of Parts Of Speech (POS) [12] query segmentation process was significantly improved. Integrated techniques to support unsupervised representation of queries remained unaddressed. Mixed Script Space [13] for ad-hoc query retrieval was introduced to identify and assess huge amount of user generated content on the web.

Nowadays, search engines has become an essential constitute of our life with the capability of preponderance involvement and association by hundreds of millions of people around the world. Domain wise semantic personalized search and customized result using support vector machine algorithm was introduced in [14] to increase the fast retrieval process. This resulted in the accuracy of the user's interest profile, but was constrained on certain domains. User profile ontology was designed in [15] that in a way efficiently mapped the semantic user's natural language queries in an efficient manner. Though efficient in terms of retrieval rate, but user satisfaction feedback remained unaddressed. Conceptual weighting method was introduced in [16] to derive the concepts connected to phrases. With the aid of the threshold weight, concepts were derived resulting in the enhancement of the search process.

One of the most significant factors in the information retrieval system is user. Larger amount of marketing benefits can be derived from the information retrieved by the information provider (IP) according to the enhanced comprehension of the user interest profiles. Shanon entropy was introduced in [17] for effective query exchanges between users. As the queries submitted by the user interest profiles were short, a method based on Fuzzy based user profile [18] was introduced to address ambiguity issues and increased the rate of precision and recall for users formulating effective queries. A web search personalization approach was introduced in [19] based on Ontology for obtaining the information most desired to the user at a lesser span of time. Personalized Web Search system and Semantic Web Personalization [20] was introduced for optimizing the quality of output or query results obtained based on the user query profile.

In this paper, the importance of effective web search using personalized web user behavior has been recognized and hence proposed to incorporate the different user behavioral profiles in addition to review the search histories with updated profile. The web user behavior thus incorporates to capture both of the user behavior based on profile and personalized recommendation of web user request page for building a web personalized user model.

3. Markov Chain Monte Carlo Dynamic Clustering Based On The User Behavior Profiles

In this section, we introduce the Markov chain Monte Carlo dynamic clustering method to group similar user behavior files even on dynamic updation with the help of user behavior analyzing process and neat architecture diagram. The set of processes are worked together on the web to group similar user's behavior even on the dynamic updation of the user profiles.

The method, MCMC-DC is introduced to widely utilize the system and generate the result related to user query search. Web Knowledge Database as illustrated in Figure 1, collects the user's information and uses the taxonomical data for automatic updating of user profiles. One of the widely used methods in modeling the user behavioral based clustering is Markov chain. The Markov-chain is embedded with dynamic clustering to improve the accuracy on working with automatic updated user profiles. The user behavior analysis and search result to the recommended user query processing is illustrated in Figure 1.

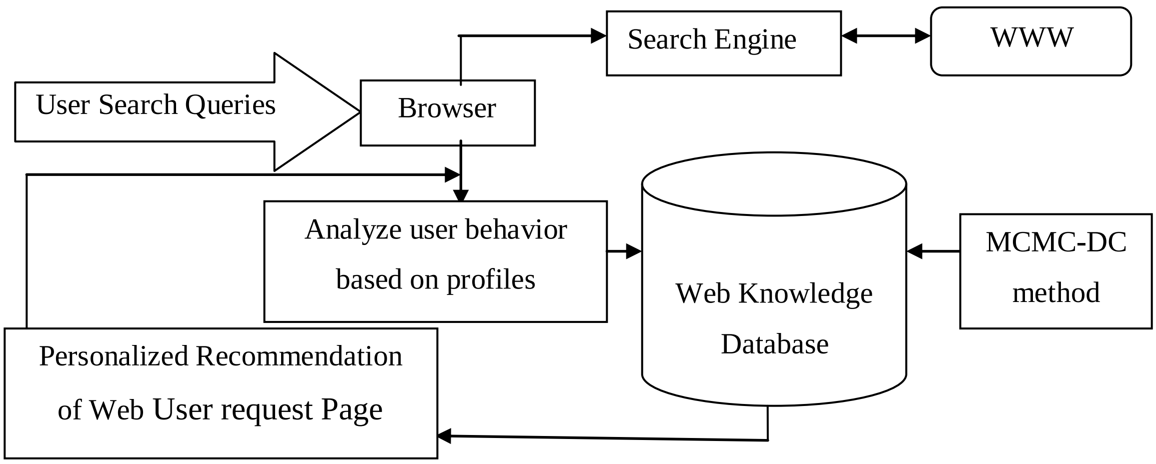

Figure 1 User Behavior Analyzing Process

As illustrated in Figure 1, the user queries are provided on the search engine that fetches the query result with the aid of web knowledge database. The World Wide Web includes all the web pages with titles name of each contents. The browser analyzes the user behavior based on each user profile and request query contents. Then the personalized web search results in as per the users need recommended through the browser. The browser keeps track of all the users' data using MCMC-DC method to collect automatically the newly updated data. The user behavior profiles include storing of user personal information along with the type of queries they request on web browser. The collection of user behavior profile in the web knowledge database produces the relevant search result suing the MCMC-DC method. The architecture diagram of MCMC-DC method is illustrated in Figure 2.

As illustrated in Figure 2, the Markov chain process on the web contains the collection of user web request routing sessions. Markov chain adopts equilibrium distribution to handle mixture of user profiles with similar information. The MCMC-DC equilibrium distribution combines different user profiles with similar search query. The Morkov chain converges with the equilibrium distribution to balance the user profiles with similar query search. MCMC-DC method uses higher order transitional probabilities to accurately identify the user replica information. The removal of replica information reduces the memory consumption rate and also the running time is minimized.

Next, Dynamic Clustering operation is carried out on the automatic updated profiles to balance the accuracy rate even on the dynamic web environment. The dynamic clustering carries the cluster head operation on each group to update automatically and maximize the use of the proposed model for personalized query search based on user behavior. The search history is reviewed periodically to improve the automatic updation of user profiles.

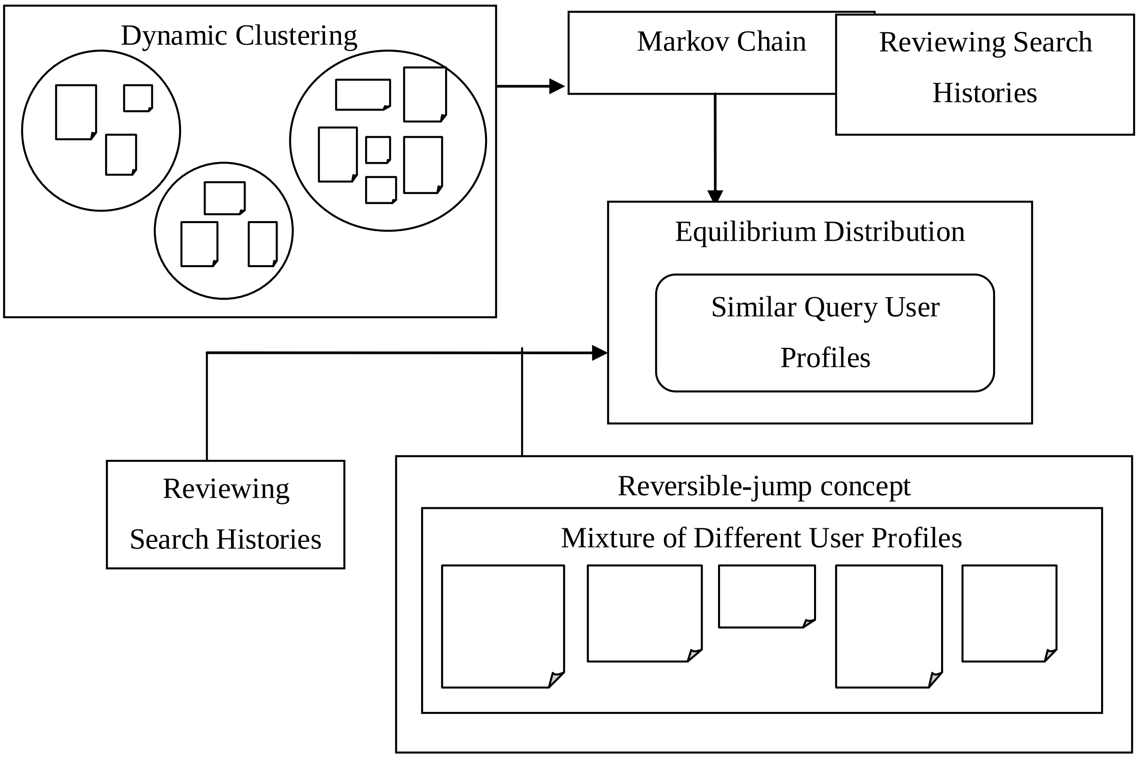

Figure 2 Architecture Diagram of MCMC-DC method

Finally, the main goal of the reversible-jump concept in MCMC-DC is to review the history with updated profile and move to the appropriate clusters. MCMC-DC with reversible-jump concept helps to effectively work on with mixture of different user profiles on web data mining. Moreover, the reversible-jump variant also includes higher order transitional probabilities to remove the replica on the updated user profiles. As a result, the MCMC-DC method is successful on clustering user behavior based on different set of test relationship from users.

In the forthcoming sections, we discuss in detail about the Markov chain equilibrium distribution process, followed by dynamic clustering operation with the help of an algorithm and conclude with the process of reversible-jump variant.

3.1 Markov Chain Equilibrium Distribution Process

The first step to be constructed in the design of MCMC-DC method is to obtain different user behavioral profile with similar query search using Markov Chain Monte Carlo process. Markov Chain Monte Carlo process on web search undergoes the transition process from one state (i.e., query search) to another state space. Each user profiles searches multiple queries using the MCMC-DC method using the transitional process. The underlying process of the Markov Chain Monte Carlo generates a chain of sample query search from users on the web to identify the equilibrium distribution. The Markov Chain Monte Carlo processing of chain user query is formularized as,

$$MCMC\,Chain\,Query = \left(U_{q+1} = u \middle| U_1 = u1, U_2 = u2 \ldots\ldots l\right) \tag{1}$$

The possible chain query ‘ ’ for different set of user profiles are acquired through the web browser. Here, the user ’ ’ request either single query or multiples of query repeatedly to fetch the result based on the personalized user profiles. The usage of equilibrium distribution function in MCMC-DC method obtains the result with similar queries for different user profiles. The Equilibrium distribution considers the web user query which is contiguous to one another query search. From the search result, the closest queries from the distributed user behavioral profiles are fetched using the MCMC-DC method as given below,

$$Equilibrium\ Distribution = \ lim_{T\to\infty}\frac{\Delta U}{T}(Sin \tag{2}$$

From the above (2) the query search at time interval ‘ ’ for different user profiles ‘ ’ is obtained. The limit of equilibrium distribution using MCMC-DC method is taken as infinity ‘ ’. The larger the infinity value, the maximum amount of time the equilibrium distribution carries out on the user query search based on personalized user behavioral profiles. In this way, similar query,‘ Si’ is fetched through Equilibrium distribution at time‘ ’.

With the aid of MCMC-DC method, the replica information is removed on the user profile updation using higher order transitional probabilities. The higher order transitional probabilities accurately identify the replica and reduce the running time. The higher order transitional probabilities is given as below,

$$Transitional\ Probabilities = \sum_{n \epsilon U} U_i^{(k} \tag{3}$$

The transitional probability with highest order number removes the replica information on personalized user profiles and therefore improves the running rate. In (2), the ‘ carries the ‘ ’ profile information whereas ‘ ’ represents the updated user information. The updated information is the duplicate information, the MCMC-DC method removes ‘ ’ duplicates from the ‘ ’ set of information.

3.2 Dynamic Clustering

The second step to be constructed in the design of MCMC-DC method is to review the search history and reduce the updation time using Dynamic Clustering. This is achieved by using the Dynamic Clustering that minimizes the sum of the squares distance of memory while maintaining the user behavior profile. The cluster obtains the ‘ ’ profile information and uses the equilibrium distribution function to group the similar profiles. With this, the similar profiles are clustered and each cluster maintains the cluster head. The cluster head in MCMC-DC method update the user profiles automatically, thereby reducing the running time. The dynamic clustering observed on the ‘ ’ user profiles with Markov Chain Monte Carlo process is given as below.

$$Dynamic\ Clustering = C\{Similar\ Q(U_1, U_2\ , U_3 \ldots . U_k)\} \tag{4}$$

The dynamic clustering of user web profiles makes use of the replica operation to remove the repeated behavior information on same profile. A dynamic clustering technique is an efficient way to group similar query on the user updated profiles. The MCMC-DC method aim is achieved by automatically updating the user behavioral profiles while controlling the replica factor. The algorithmic description of Markov Chain Dynamic Clustering is given below:

//Markov Chain Dynamic Clustering
Input: ‘k’ web user profile with ‘n’ behavioral information
repeat
for each ‘k’ user profiles
for each ‘Q’ queries fetched
Cluster similar query on personalized behavioral profile
end for

6: **end for**
7: **for** each Cluster ‘C’
8: form cluster head ‘H’
9: Update the cluster as per updated information
10: Review the history and update profile ‘k’
11: **end for**
12: **If** replica on update information then
13: Discard id
14: **else**
15:Add the information of user profile to appropriate Cluster ‘C
16: **end if**
16: **until** (all ‘k’ web user profile reached)
Output: Dynamic Cluster ‘C’ with automatically updated profiles

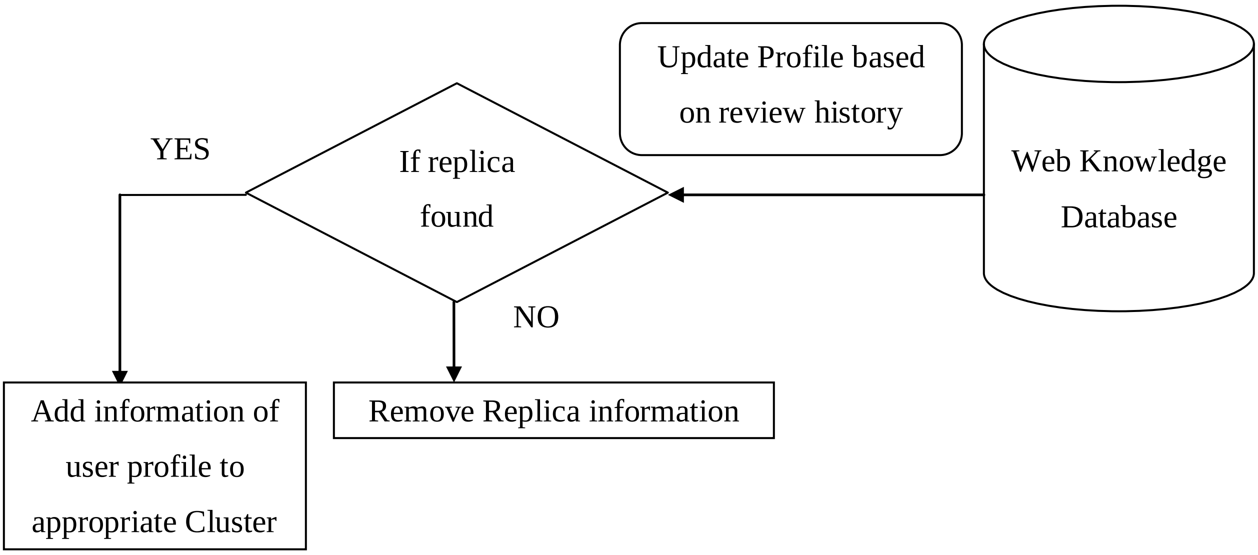

Figure 3 Flow diagram of Reversible-Jump Process

The algorithmic step in the MCMC-DC method for automatic updation of profile is the reversible jump approach. The main goal of the reversible jump approach is to remove the replica information and update the profile with minimal running time. In step 7, Cluster ‘C’ carries the cluster head operation ‘H’ (in step 8) for easier updation of the personalized behavior of the user profiles. With this, similar preferences of users are grouped effectively to improve the online searching behavior result.

3.2.1 Process of Reversible-Jump

Finally, to review the history with updated profiles, the process of reversible-jump is introduced in MCMC-DC method. The MCMC-DC method with reversible-jump concept helps to effectively work on with mixture of different user profile information and to cluster the web data by avoiding replicas. The flow diagram of reversible-jump using MCMC-DC is clearly illustrated in Figure 3.

The reversible-jump concept main goal is to review the history with updated profile and move to the appropriate clusters by removing the replica information. Figure 3 as illustrated performs the reversible-jump processing for obtaining the user result without any replica information. The user query results are again altered as per the automatic updation of user profiles using MCMC-DC method. The Markov chain Monte Carlo process on using the reversible jump concept establishes the web search effective after the user profile updation. The automatic user profile updated is as given below,

$$\iiint_{1t,2t}^{n} P(U, du') = \iiint_{2t,1t}^{n} P \quad (5)$$

With the help of the reversible jump, one transition ‘ ’, two transitions ‘ ’, upto ‘ ’ order transitions are noted on updating web user profile automatically through reversible-jump. In (5), ‘ ’ denotes the web users where the jump is carried out in reverse order to perform the query result search as per the updated user profile.

4. Experimental Evaluation

Markov Chain Monte Carlo Dynamic Clustering based on the User behavior profiles (MCMC-DC) method is used on retrieving the accurate retrieving result. In order to evaluate the proposed system performance, Amazon EC2 instances clustered database is used. Freebase Data Dump is the Public Amazon Web Services which is used for experimenting MCMC-DC method. A data dump of all the present facts and declaration in the Freebase system is provided with an open database covering millions of theme.

A data dump is the essential information provided on identifying the facts concerning each subject in Freebase. Freebase is an open database of the world's information, cover millions of theme in hundreds of group. Drawing from large open datasets like Wikipedia, MusicBrainz, and the SEC archives simultaneously contain prearranged information on several popular topics, including movies, music, people and locations. The information includes the historic events, European railway stations and chemical properties of common food ingredients.

The MCMC-DC method compares with the existing work such as Amount of invested mental effort (AIME-OS) [1] in online searching and Personalized Ontology Model (POM) [2]. The method, MCMC-DC method is experimented on factors such as accuracy, relative error rate, similar user profile clustering rate, and average running time of user queries.

The accuracy ‘ on working with automatic updated user profiles using MCMC-DC method is the proportion ratio of true results $Number\ of\ true$

from web knowledge database $Number\ of\ all\ assessme$. It is measured in terms of percentage (%).

$$A = \frac{Number\ of\ true\ resu}{Number\ of\ all\ assessmen} \quad (6)$$

The similar user profile clustering rate using MCMC-DC method is measured using the equilibrium distribution $Equilibrium\ Distri$ obtained from (2) for ‘ ’ different user profiles as given below. Higher, the similar user profile clustering rate, higher is the efficiency of the method to be proven. It is measured in terms of percentage (%).

$$UPCR_{similar} = \sum_{i=1}^{n} Equilibrium\ Distribution_{i-} \quad (7)$$

The average running time of user queries ART_{use} using MCMC-DC method is the measure of running time of ‘ ’ user queries by applying higher order transitional probabilities $U_i^{(k)} *$. The average running time is measured in terms of milliseconds or ms.

$$ART_{user\ queries} = Time\ (\frac{U_i^{(k)}}{} \quad (8)$$

The relative error rate for MCMC-DC method is the ratio of difference between the true query result $True_{que}$ and the measured query result $Measured_{que}$ with query result as per the updated user profile $Query\ Result_{updated\ use}$.

$$RER = 100 - \frac{True_{query\ result} - Measured_{query\ result}}{Query\ Result_{updated\ user\ profile}} * \quad (9)$$

5. Results On MCMC-DC

Markov Chain Monte Carlo Dynamic Clustering based on the user behavior profiles (MCMC-DC) is compared against the existing Amount of Invested Mental Effort in Online Searching (AIME-OS) [1] and Personalized Ontology Model (POM) [2]. The evaluation value given below with the aid of table and graph outlines the MCMC-DC on user behavior profiles improves the accuracy rate with automatic updated user profiles. Table 1 evaluates the accuracy rate measured in terms of percentage (%) achieved using different number of queries ranging from 50 to 350. Comparison is made with the two existing schemes namely, Amount of Invested Mental Effort in Online Searching (AIME-OS) [1] and Personalized Ontology Model (POM) [2].

Table 1 Comparison of different accuracy rate using MCMC-DC, AIME and POM

Number of queries (Q)	Accuracy with automatic updated user profiles (%)		
	MCMC-DC	AIME	POM
50	63.45	58.43	46.40
100	67.85	62.81	50.78
150	71.35	66.33	54.30
200	73.45	68.43	56.40
250	77.35	72.32	60.29
300	70.25	65.21	53.18
350	81.45	76.40	64.37

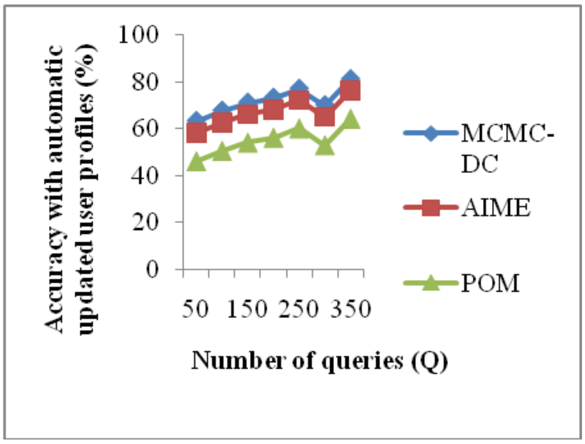

Figure 4 Measure of Accuracy with automatic updated user profiles

Figure 4 illustrate the accuracy based on the automatic updated user. Our proposed MCMC-DC method performs relatively well when compared to two other methods AIME [1] and POM [2]. The rate of accuracy using the proposed method, MCMC-DC is increased with the application of dynamic clustering operation that efficiently carries out the cluster head operation on each group resulting in the increase of accuracy by 6 – 7 % compared to AIME. Moreover, the dynamic clustering minimizes the sum of squares distance of memory while maintaining the user behavior profile resulting in the increase of accuracy by 20 – 26 % compared to POM.

Table 2 Significance results of Similar User Profile Clustering Rate

Number of Clusters (C)	Similar User Profile Clustering Rate (%)		
	MCMC-DC	AIME	POM
5	0.214	0.194	0.164
10	0.235	0.230	0.185
15	0.256	0.226	0.206
20	0.267	0.237	0.217
25	0.245	0.215	0.195
30	0.275	0.245	0.250
35	0.305	0.275	0.255

The Similar User Profile Clustering Rate efficiency of our work MCMC-DC with the existing two schemes namely AIME [1] and POM [2] is provided in table 2. As listed in table 2, the MCMC-DC method measures the Similar User Profile Clustering Rate which is measured in terms of percentage (%). The resultant Similar User Profile Clustering Rate using our method MCMC-DC offers comparable values than the state-of-the-art methods.

The targeting results of Similar User Profile Clustering Rate using MCMC-DC method with two state-of-the-art methods [1], [2] in figure 5 is presented for visual comparison based on the varied number of clusters. Our method differs from the AIME [1] and POM [2] in that we have incorporated Markov chain that applies equilibrium distribution to handle mixture of user profiles with similar information resulting in the increase of similar user profile clustering rate by 2 – 11 % compared to AIME. Besides, the Markov Chain Monte Carlo processing of chain user query result with similar queries for different user profiles results in the increase of similar user profile clustering rate by 9 – 23 % compared to POM.

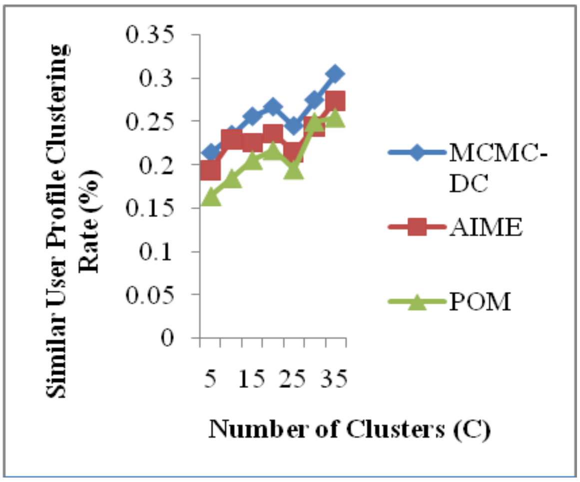

Figure 5 Measure of Similar User Profile Clustering Rate

Table 3 Experimental results of Average Running Time of User Queries

Number of Queries (Q)	Average Running Time of User Queries (ms)		
	MCMC-DC	AIME	POM
50	33.5	45.6	57.8
100	38.4	50.5	62.7
150	41.3	53.4	65.6
200	44.5	56.6	68.8
250	40.2	52.3	64.5
300	48.3	60.4	72.6
350	52.5	64.6	76.8

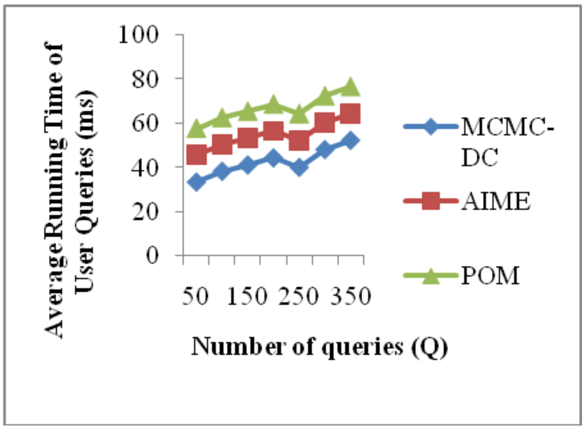

Figure 6 Measure of Average Running Time of User Queries

More accurately the influence of average running time of user queries with respect to the number of queries provided as input is listed in table 3 and comparison is made with two other existing schemes. It can also be seen that the average running time of user queries increases with the increase in the number of queries though not observed to be linear.

Table 4 Tabulation for Relative Error Rate

Number of Clusters (C)	Relative Error Rate (%)		
	MCMC-DC	AIME	POM
5	43.5	48.6	53.8
10	47.8	52.9	57.0
15	53.3	58.4	63.6
20	48.5	53.6	58.8
25	55.8	60.9	66.0
30	60.2	65.3	70.4
35	63.4	68.5	73.6

Figure 6 shows the Average Running Time of User Queries for MCMC-DC method, AIME [1] and POM [2] versus increasing number of queries of sizes from Q = 50 to Q = 350. The average running time of user queries gradually increased as the number of queries gets increased, though comparatively lesser using the MCMC-DC method. For example when Q = 200, the percentage improvement of MCMC-DC method compared to AIME is 27.19 percent and compared to POM is 54.60 percent, whereas for Q = 250 the improvements are around 30.09 and 60.44 percent compared to AIME and POM respectively.

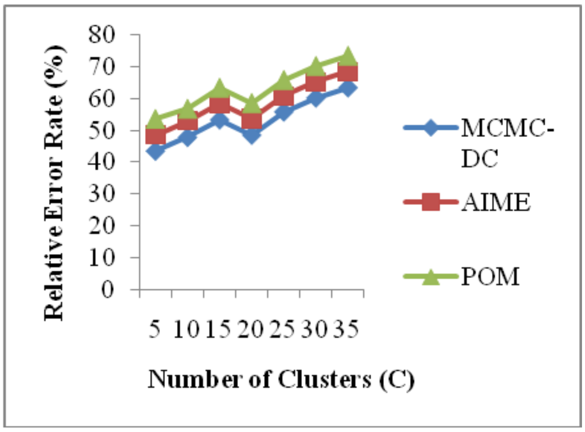

Figure 7 Measure of Relative Error Rate

The reason is that the average running time of user queries is reduced by adopting higher order transitional probabilities that accurately identify the user replica information. The removal of the respective user replica information results in minimizing the average running time of user queries by 23 – 36 % compared to AIME. In addition, the higher order transitional probabilities reduce the average running time by 46 – 72 % compared to POM not only carrying the 'n' profile information but also removing the 'k' duplicated information.

Table 4 and Figure 7 illustrate the Relative Error Rate for varied cluster sizes of range 5 to 35. As illustrated in the figure, the Relative Error Rate increases linearly with the increasing cluster sizes. Though comparatively, the relative error rate outperformed the other two methods by providing lesser rate. The relative error rate is reduced using the proposed MCMC-DC method by introducing reversible-jump variant that even removes the replica on updated user profiles. With the application of reversible-jump operation, jump is carried out in reverse order to perform the query result search as per the updated user profile minimizing the relative error rate by 8 – 11 % and 16 – 23 % compared to AIME and POM respectively.

6. Conclusion

In this paper, a Markov Chain Monte Carlo Dynamic Clustering (MCMC-DC) based on the user behavior profiles is presented to collect similar user behavior profiles and perform the automatic updation operation. The MCMC-DC method utilizes Markov Chain Monte Carlo function to efficiently derive different user behavioral profile with similar query search. It also applied Dynamic Clustering to minimize the updation and running time with the aid of equilibrium distribution function. Finally, reversible-jump process was applied to remove the replica information based on the mixture of different user profiles and add the information of user profile to the appropriate cluster. Experimental results demonstrate that the proposed MCMC-DC method not only leads to noticeable improvement over the parameters precision and similar user profile clustering rate, but also outperforms average running time of user queries over state-of-the-art methods.

References

[1] Soo Young Rieh, Yong-Mi Kim, Karen Markey, "Amount of invested mental effort (AIME) in online searching," Information Processing and Management, *Elsevier journal*, 2012

[2] Xiaohui Tao, Yuefeng Li and Ning Zhong,"A Personalized Ontology Model for Web Information Gathering,"*IEEE Transactions On Knowledge And Data Engineering*, VOL. 23, NO. 4, APRIL 2011

[3] Christian Bizer, Jens Lehmann, Georgi Kobilarov, Soren Auer, Christian Becker, Richard Cyganiak Sebastian Hellmann," DBpedia - *A Crystallization Point for the Web of Data*", Elsevier, May 2009

[4] David Milne, Ian H. Witten," An open-source toolkit for mining Wikipedia", *Artificial Intelligence*, June 2012

[5] Tak-Lam Wong and Wai Lam," Learning to Adapt Web Information Extraction Knowledge and Discovering New Attributes via a Bayesian Approach", *IEEE Transactions On Knowledge And Data Engineering*, VOL. 22, NO. 4, APRIL 2010

[6] Weifeng Su, Jiying Wang, and Frederick H. Lochovsky," Record Matching over Query Results from Multiple Web Databases", *IEEE Transactions On Knowledge And Data Engineering*, VOL. 22, NO. 4, APRIL 2010

[7] Kenneth Wai-Ting Leung and Dik Lun Lee," Deriving Concept-Based User Profiles from Search Engine Logs *IEEE Transactions On Knowledge And Data Engineering*, VOL. 22, NO. 7, JULY 2010

[8] Ja-Hwung Su, Wei-Jyun Huang, Philip S. Yu and Vincent S. Tseng," Efficient Relevance Feedback for Content-Based Image Retrieval by Mining User Navigation Patterns", *IEEE Transactions On Knowledge And Data Engineering*, VOL. 23, NO. 3, MARCH 2011

[9] Milad Shokouhi," Learning to Personalize Query Auto-Completion", *ACM*, Aug 2013

[10] Jianqiang Shen, Oliver Brdiczka, Yiye Ruan," A comparison study of user behavior on Facebook and Gmail", Computers in Human Behavior, *Elsevier*, June 2013

[11] Neha Goel, C.K. Jha," Analyzing Users Behavior from Web Access Logs using Automated Log Analyzer Tool", *International Journal of Computer Applications (0975 – 8887)* Volume 62– No.2, January 2013
[12] Rishiraj Saha Roy, Yogarshi Vyas and Niloy Ganguly, Monojit Choudhury," Improving Unsupervised Query Segmentation using Parts-of-Speech Sequence Information", *ACM*, July 2014
[13] Parth Gupta, Kalika Bali, Rafael E. Banchs," Query Expansion for Mixed-Script Information Retrieval", *ACM*, July 2014
[14] Archana Y. Panpatil," Domain Personalized Multimedia Encyclopedia Search by Web Mining", *International Journal of Advance Foundation and Research in Computer (IJAFRC)* Volume 1, Issue 3, March 2014
[15] Ahmed Al-Nazer, Tarek Helmy," Semantic Query-Manipulation and Personalized Retrieval of Health, Food and Nutrition Information", 4th International Conference on Ambient Systems, Networks and Technologies (ANT-2013), *Elsevier,*
[16] Amita Jain, Kanika Mittal, Smita Sabharwal," Conceptual Weighing Query Expansion based on User Profiles", *International Journal of Computer Applications*, Feb 2012
[17] David Rebollo-Monedero, Jordi Forne, and Josep Domingo-Ferrer," Query Profile Obfuscation by means of Optimal Query Exchange between Users*", IEEE Transactions On Dependable And Secure Computing*, Dec 2011
[18] Rahul Verma, Prof. Kshitij Pathak," Deriving Personalized Concept and Fuzzy Based User Profile from Search Engine Queries", *International Journal of Scientific & Engineering Research* Volume 4, Issue3, March-2013
[19] Shamkala Waghmare, Prof. R. K. Krishna," Implementation of Personalized User Model Based on Ontology: A Review Study", *International Journal of Computer Science and Information Technologies*, Vol. 4 (6) , 2013, 787-790
[20] T.Venkata Ramana, Dr.K.Venugopala Rao," User Search Personalization in Semantic Web Mining*", International Journal of Advanced Research in Computer Engineering & Technology* Volume 1, Issue 3, May 2012