

## Modeling user navigation behavior in web by colored Petri nets to determine the user's interest in recommending web pages

Maryam Bahadori<sup>a\*</sup>, Ali Harounabadi<sup>b</sup> and Mehdi Sadeghzadeh<sup>c</sup>

<sup>a</sup>Department of Computer, Dezful Branch, Islamic Azad University, Dezful, Iran

<sup>b</sup>Department of Computer, Central Tehran Branch, Islamic Azad University, Tehran, Iran

<sup>c</sup>Department of Computer, Mahshahr Branch, Islamic Azad University, Mahshahr, Iran

### CHRONICLE

#### Article history:

Received July 2, 2012

Received in revised format

28 October 2012

Accepted 11 November 2012

Available online

November 16 2012

#### Keywords:

Web mining

Personalization

Web recommender systems

Colored Petri nets

### ABSTRACT

One of existing challenges in personalization of the web is increasing the efficiency of a web in meeting the users' requirements for the contents they require in an optimal state. All the information associated with the current user behavior following in web and data obtained from pervious users' interaction in web can provide some necessary keys to recommend presentation of services, productions, and the required information of the users. This study aims at presenting a formal model based on colored Petri nets to identify the present user's interest, which is utilized to recommend the most appropriate pages ahead. In the proposed design, recommendation of the pages is considered with respect to information obtained from pervious users' profile as well as the current session of the present user. This model offers the updated proposed pages to the user by clicking on the web pages. Moreover, an example of web is modeled using CPN Tools. The results of the simulation show that this design improves the precision factor. We explain, through evaluation where the results of this method are more objective and the dynamic recommendations demonstrate that the results of the recommended method improve the precision criterion 15% more than the static method.

© 2013 Growing Science Ltd. All rights reserved.

## 1. Introduction

With extending the web, finding the useful information to meet the information required by users has become difficult. Search engines are responsible for addressing this undeniable need. In general, web search engines pursue one of the following procedures:

- Systems that use the relevant feedback of users,
- Systems in which users record their interests and statistical information,

\*Corresponding author.

E-mail addresses: M2010.bahadori@yahoo.com (M. Bahadori)

- Systems to recommend information to the user based on rank and advantage (Sugiyama et al., 2004).

The first and the second systems need the users' direct involvement. In these systems, users should record their personal information such as interests, age, etc. in advance and they need to enter their comments on the relatedness or unrelated information through numeric criteria or on a measure between bad and very good. These kinds of registrations, feedback gathering, and rankings take significant amount of time and most users do not tend to register and express their interests explicitly and prefer to use easier ways. Systems of the third type extract users' interests from their previous profile without users' direct involvement and recommendation of pages forms based on their best interest. Studying the user's behaviors in web is considered as an important tool in web mining domain for discovering knowledge related to users interaction with web. Research conducted in this real extracts knowledge mostly based on the necessary information obtained from users' interactions with web and it enjoys this knowledge in personalization, recommending the web pages, determining the relationship among documents, and web self-organizing. In order to provide recommendation to users, one of the following four various techniques can be chosen:

- 1) Prediction of a (or several) page(s) ahead,
- 2) Recommending the most probable next pages,
- 3) Prediction of the most popular paths ahead,
- 4) Providing recommendation created by semantic expansion of the page (Eirinaki et al., 2005).

There has been a tremendous growth in internet systems, which hold the capability of personalization of the content sent to the users. In recent years, many changes have occurred in sciences in the background of personalization. The primary objective of personalization systems is the same as the previous ones and includes providing what users need and would like without explicitly being asked from them. Personalization consists of supplying proper produces, services, information, and information specifically associated with productions or services.

Personalization is an extensive area, which covers web recommender systems, custom systems, and adaptive webs. In providing the desired services for users, three aspects of web are effective for the web utility. These aspects are the content presented in web, layout of the personal pages, and the general structure of the web.

Formal models such as Petri net and queuing help us analyze the user's behavior with mathematics. Recent developments in the area of web designing help increase the capability of concurrent exhibition of several web pages. Petri nets, on the other hand, are able to analyze and exhibit the actions and behaviors of the system and they have strong mathematical background to cover the concurrency discussion in systems. They can also be regarded as a formal model in simulation of web structure and modeling user's behavior in webs. Besides, their significant structural features can be employed to increase the quality of analysis.

The main drawback of the mentioned recommender systems is that they keep the users' models and change them slowly and they do not consider the fact that various sessions of a user or single sessions of users may express their different interests and purposes. To eliminate this defect, in the proposed method, the current session of user are followed step by step and due to the user's selection at each step, the recommended pages are suggested, dynamically. In each session, the user's favorites are defined separately in various content classes.

This research aims at presenting a Colored Perti Net (CPN) to model users' interactions in web to present recommendation list of pages to a user, personally. First, estimation of user's behavior in web proposes appropriate recommendations to users to continue in web and second, it leads to an accurate

understanding of users in abstract cases such as electronic business. In current study, the proposed pages are presented to users, personally and dynamically. After determining the value of web pages considering the number of previous visiting users, the web pages, classified based on content, are ordered from the most valuable pages to the least ones. With the user's entrance to the web, his (her) current session is followed and with respect to the amount of pause on the page and the page size, the user's interest in each page is pinpointed and on the basis of the obtained result, the most valuable pages of each category are suggested to user.

The layout of current research is as follow: in the second section, literature is explained, afterwards, a review of last related research is presented in the third section and the forth section includes a presentation of the proposed method and in fifth section, and evaluation, simulation, and results are discussed.

## **2. Literature**

### *2.1 Recommender systems*

Data mining consists of extracting important and valuable data from a multitude of data. Web mining is a subset of data mining technique to cover the used patterns of web and due to the extensiveness of the related domain; they are divided into three groups: web content mining, web structure mining, and web usage mining. One of the main applications of data mining techniques for the recorded files of large repositories of web data is web usage mining and the goal is to produce results in some aspects such as web designing, classifying users, designing adaptive websites, and web personalization. The web personalization systems are one of the outstanding applications in web usage mining (Thangaraj & Thangamani, 2011). Personalization is the process of selecting a web with respect to the need of a certain user. It refers to the type and manner of information exhibition in the web and it is feasible with regard to the history maintained of the use of the web, because users with various interests and tastes are in interaction with the web (Mulvenna et al., 2000).

Web recommender systems predict the users' need and supply them in recommendation format to personalize the users' navigation. The web-based interactive software agent system is a criterion attempting to forecast the user's priorities based on user data or access data. This task is executed to facilitate and to customize the user's online experience utilizing the recommendation lists of the proposed items. The suggested items can products such as books, films, music as well as online sources including web pages and online actions such as predicting a path. In general, a web recommender system encompasses two modules including online module and off-line module.

The offline module preprocesses the data to produce the user models, while the online module employs and updates these models during the work to recognize the user's objective and to predict a list of recommendations (Spiliopoulou, 2000).

### *2.2. Colored Petri Nets (CPN)*

Colored Petri nets and their analysis techniques were first created in Aarhus University in Denmark in Kurt Jensen's doctoral dissertation (Jensen, 1980). Then, the CP-net group in Computer Science Department of Aarhus University expanded its analysis methods and display tools and software package Design/CPN were designed by students and researchers in this university to implement Colored Petri nets (Jensen, 1997). The expansion of Colored Petri nets was performed with the aim of creating a modeling language.

The reason we use Colored Petri is because it provides the possibility of using beads, which carry various amount of data and they are distinct from each other. In fact, Petri nets are a graphic tool for formal description of dynamical systems, which preserves features such as concurrency, mutual exclusivity, and conflict and they maintain special characteristics of distributed environments.

Colored Petri nets are an expanded model of Petri nets where a Colored Petri nets model is displayed graphically by directed bipartite graph. In this model, places are demonstrated with circle and transitions with rectangle and the tokens are demonstrated in places in the form of black beads.

The formal definition of a colored Petri net includes nine tuples (Jensen, 1993):

$$\text{CPN} = (\Sigma, P, T, A, N, C, G, E, I)$$

where

- $\Sigma$  is a finite and non-empty set that is called color set.
- $P$  is a finite set of places.
- $T$  is a finite set of transitions.
- $A$  is a finite set of arcs that:  $p \cap T = p \cap A = T \cap A = \emptyset$
- $N$  is a Node function that is defined from  $A$  to  $P \times T \cup T \times P$ .
- $C$  is a color function that is described by  $P$  to  $\Sigma$ .
- $G$  is a guard function that is defined from  $T$  to relation like:

$$\forall t \in T: [\text{Type}(G(t)) = B \wedge \text{Type}(\text{Var}(G(t))) \subseteq \Sigma, \quad (1)$$

where,  $B$  is an indicator of Boolean type.

- $E$  is an arc expression function which is defined from  $A$  based on the following,

$$\forall a \in A: [\text{Type}(E(a)) = C(p)_{ms} \wedge \text{Type}(\text{Vary}(E(a))) \subseteq \Sigma] \quad (2)$$

where  $p$  stands for the place of  $N(n)$ .

- $I$  is an initialization function defined from  $P$  into closed expressions such that:

$$\forall p \in P: [\text{Type}(I(p)) = C(p)_{MS} \quad (3)$$

### 3. Review of the related studies

Different types of formal models have been used earlier in modeling of applications of web mining. Yang et al. (2007) and Chen et al. (2002) used simple and stochastic timed Petri nets (STPN) to model web structure with the purpose of facilitating web usage mining, analysis of web structure. They employed the reachability feature of the resultant adjacency matrix of Petri net where STPN helps preprocessing phases of data in path completion process and page view identification.

The other formal models widely utilized in modeling and user's navigation behavior are Markov models (Deshpande, M., & Karypis, 2001; Levene & Loizou, 2003; Borges & Levene, 2004; Sugiyama et al., 2004). These models have been used for various purposes such as extracting popular web paths, predicting the next steps of the current session, and selecting the most probable navigation path of user.

Mobasher et al. (2000) provided a framework for mining web log files, which tries to discover goal from the knowledge to create recommenders for the current users on the basis of search similarity to previous users. The required process to discover such the knowledge includes collecting and pre-

processing of the necessary data to discover the users' behaviors, applying the data mining techniques to discover usage patterns, and aggregation and filtering the data mining results to present the decision rules to customize the web content based on behaviors of individual users (Mobasher et al., 2000). Spiliopoulou (2000) explained the reasons for the log file data. The key element to keep the users is the efficiency of a web in terms of providing the user's requirements by the contents they need in the most optimal conditions. Traditional classification, based on techniques to measure the web structure effectively, is expensive and difficult to use, because there is a little knowledge on the basic population of web users and their demands. Through a type of mining in navigation patterns, Spiliopoulou (2000) presented a process, which could be used to obtain information within web usage behavior optimizing the current user population.

Etzioni and Perkowitz (Mobasher et al., 2000) introduced personalization as a process, which uses an internet web by automatic production of index pages for the web. In their study, adaptive websites were searched to improve their organizations and to display automatically by learning the visitors' access patterns. The adaptive websites mine the buried data in server log files for simpler production of websites with the navigability feature.

#### 4. The proposed method

Similar to other web recommender systems, this section includes an off-line phase, which ranks web pages, classifies them based on their content based on the number of visits of the pervious users. Afterwards, in the on-line phase, the amount of user's interest in any class of pages is pinpointed by following the current session of user and considering the time user devotes to each page as well as the size of the page. Then, on the basis of the amount of user's interest, some pages of the respective category are suggested to user. The reason to choose these criteria, including use behavior of the previous user, the page size, and the time allotted to visiting each page, are as follows:

- Following up the users' visits leads to some information utilized in page valuation. Among these cases is the repetition frequency of page (number of visits). The pages of web, which have been frequently visited by pervious higher value websites.
- When a user devotes prolonged time to visiting the page, most likely he/she is interested in that page and if a page is not attractive, the user quickly jumps to the other page. It should also be noted that the quick jump to another page is probably due to short length of the page.

Therefore, the user's interest in a webpage can be estimated from the time spent to visit a page. Moreover, the user may give up the page during visiting the page and does the other tasks. In this case, such prolonged pause on the page is not an indication of the page value. So, in order to prevent such conditions, a threshold (20 minutes) is identified.

To identify the importance of the considered pages, higher rates are assigned to the most visited pages in ranking. The purpose of this research is to determine the user's interest to recommend the pages associated with his/her interest, dynamically. We assume that number of web pages associated with the database under study is constant during a day. In simulation tool of CPN, all pages are placed in token format in a place. Each token consists of page information including page size, class, and number of visits. The session of every user is also as a token, which includes information extracted from the sequence of the followed pages by user and time devoted to each page.

$$\text{PageValue}(pi) = \frac{\text{time}(pi)}{\text{size}(pi)} \cdot \frac{1}{\sum_{j=1}^n \frac{\text{time}(pj)}{\text{size}(pj)}} \quad (4)$$

$$\forall pi \in ck, ckValue = \sum \text{PageValue}(pi), \quad (5)$$

where  $N$  is the number of visited pages in sessions.

In this scenario, with each step the user take, more information is obtained about his/ her interests and more relevant pages are recommended to him (her).

Step1: Visiting the first the page by the user among the recommended pages or one of the defined category.

First state: if the user selects one of the recommended pages: the name of category, and size of the page of the token associated with page, and also the time spent to visit the page are considered of the token related to the user session and then on the basis of these factors, the user's interest in this class is calculated based on Eq. (4) and Eq. (5). Several new pages are replaced several final pages of the suggested list and the list pages are arranged based upon the number of visits.

Second state: If the user selects one of the given classes, the pages associated with other classes are filtered and the user can see only the pages relevant to that certain group. Then he/she chooses a page associated with that group and the process continues similar to the first state.

Step 2: In second step, the user selects the next page (second page). The name of the class and the page size of the token related to page, and the time spent to visit the page are extracted from the token related to the user session.

If the selected page is of same pervious class, the value of new page is added to the pervious one and sums of both of these shows the value of the relevant category to user. In this case, two new pages of the related group from the same higher rank class are again added to set of the recommended pages and two final pages of the suggested list are eliminated. This happens before adding new pages of the classes to the recommended list, pages of the list are checked. If the number of pages recommended from a category is more than a threshold, pages of that group are no longer added to that list. For example, if the length of the recommended list is considered as 25, if the number of the recommended pages of social class reaches to 25 and the user selects the social pages again, the number of social pages of the recommended list does not change. If the chosen page was from another class, the page value in session is calculated and on the basis of this value, pages of new class, which have higher rank in that class are added to the list and the same number of pages is deleted from the end of the recommended list. In order to determine the number of the recommended pages of each class added to the list at each step, the following process is followed:

Let us assume that the amount of interest of the user in each of the four recommended pages has been calculated as 0.11, 0.38, 0.51, and 0 respectively. We include six new pages for him/her in the recommended list. The calculated amounts are multiplied by these six new pages, respectively and the result is rounded to an integer number. Therefore, from the first to the fourth page, one, two, three, and zero page(s) are recommended, respectively.

## 5. Simulation and results

### Assumptions:

- The number of web pages (studied database) is steady during a day.
- The minimum time to visit a page is one minute and the maximum time is 20 seconds.
- The page number of the recommended list is 25.

All pages are placed in the format of tokens in a place. Each of these tokens encompasses information associated with page including size, name of class, and number of visits. The users' interests are shown as token including sequence information of the pages followed by the user and the time

devoted to each page. There is another place containing recommended pages and various classes where the user has access to them while entering the website.

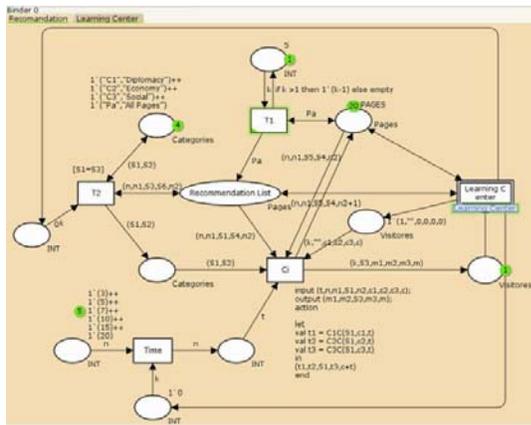


Fig. (1-a). modeling of the recommender system

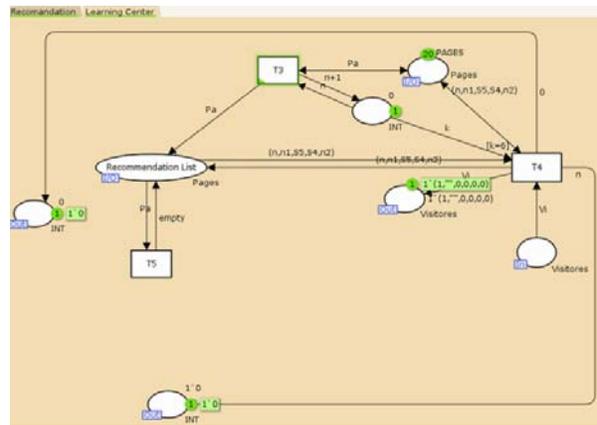


Fig. (1-b). Learning center subnet

With respect to modeling of the recommender system presented in Fig. (1-a), while entering the site, the user can choose one of the recommended pages or the available classes (transition T1). If he/she selects a particular class, the pages within the place including the page are filtered and only the pages of that particular class are available to the user (transition T2). If one of the recommended pages is selected, the features of that page are included in user's interests and its calculations are performed based on Eq. (4) and Eq. (5) in learning center subnet (Fig. (1-b)). Afterwards, the recommended list is updated and transferred to recommendation list place through the subnet output and it is given to user for the next selection. In user's next steps, the same cycle is repeated.

**6. Results and conclusion**

Given the fact that, a recommender system estimates and recovers the documents associated with user's interest among a set of documents, to evaluate a recommender system, evaluation criteria related to information retrieval can be used. In information retrieval concepts, the precision is calculated based on two sets of documents. These two sets are "set of retrieved documents" and "set of documents related to the given topic". The result of division of "the number of related retrieved documents" by "the total number of retrieved documents" is called precision. In current study, "the number of related retrieved documents" is equivalent to the number of pages in recommended list, which are the same as the content with the user's selected page. "The total number of retrieved documents" is the number of the recommended elements.

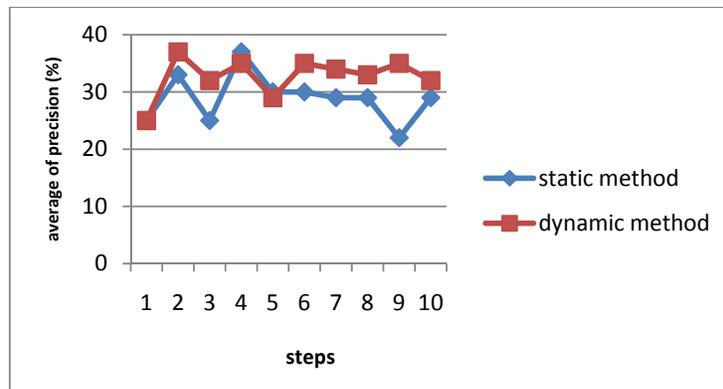


Fig. 2. the comparison of precision average of the dynamic recommended algorithm and the static algorithm

The initial static algorithm used in the studied website, offers a fixed recommended list to all users, elements of this list have been selected just due to the number of visits of previous users, and it does not make into consideration the user's current session and interests. Fig. 2 demonstrates the results associated with comparison of precision average of the dynamic recommended algorithm and the aforementioned static algorithm. In this figure, the horizontal axis shows the order of steps of users' consequent visits and vertical axis depicts the average of calculated precision of the studied users. As seen in Fig. 2, the obtained result does not follow a steady trend and it experiences a fluctuating trend. This is because of the fact that the manner of users' web surfing varies much. When the user follows pages of the same class consequently, the recommended algorithm increases the value of that class from the users' viewpoints and the number of the recommended pages of that class increments in the suggested list. When the user goes to a page of another class, the performance undergoes considerable drop. One explanation is that the pages have been classified in a mere way, i.e. a page is a member of a class or not.

## References

- Borges, J., & Levene, M. (2004). *A Dynamic Clustering-Based Markov Model for Web Usage Mining*. CS.IR/0406032.
- Chen, P., Sun, C.H., & Yang, S.Y. (2008). Modeling and analysis the web structure using stochastic timed Petri nets. *Journal of Software*, 3(8), 19-26.
- Deshpande, M., & Karypis, G. (2001). Selective Markov models for predicting web-wage accesses. *ACM transactions on internet technology (TOIT)*, 4(2), 163-184.
- Eirinaki, M., Vazirgiannis, M., & Kapogiannis, D. (2005). Web path recommendations based on page ranking and Markov models. *7th annual ACM international workshop on web information and data management (WIDM)*, ISBN: 1-59593-194-5, 2-9.
- Jensen, K. (1997). A brief introduction to colored Petri nets. Springer-Verlag London, the Third International Workshop on Tools and Algorithms for Construction and Analysis of Systems, ISBN. 3-540-62790-1, 203-208.
- Jensen, K. (1980). *A method to computer the descriptive power of the different types of Petri nets*. Springer-Verlag, 88, 348-361.
- Jensen, K. (1993). *An Introduction to the Theoretical Aspects of Coloured Petri Nets*. Springer-Verlag London, ISBN. 3-540-58043-3, 230-272.
- Levene, M., & Loizou, G. (2003). Computing the entropy of user navigation in the Web. *International Journal of Information Technology and Decision Making*, 2, 459-476.
- Li, J., & Zaïane, O. R. (2004). Combining usage, content, and structure data to improve website recommendation. *5<sup>th</sup> international conference on electronic commerce and web technologies (EC-Web)*, 305-315.
- Mobasher, B., Cooley, R., & Srivastava, J. (2000). Automatic Personalization Based on Web-Usage Mining. *Communication of ACM*, 43(8), 142-151.
- Mulvenna, M. D., Anand, S. S., & Büchner, A. G. (2000). Personalization on the Net using Web Mining: introduction. *Communications of the ACM*, 43(8), 122-125.
- Spiliopoulou, M. (2000). Web Usage Mining For Web Site Evaluation Making A Site Better Fit Its Users. *Communications of the ACM*, 43(8), 127-134.
- Sugiyama, K., Hatano, K., & Yoshikawa, M. (2004). Adaptive web search based on user profile constructed without any effort from users. *World Wide Web Conference Series - WWW*, 675-684.
- Thangaraj, P., & Thangamani, C. (2011). Survey on web usage mining: Pattern discovery and applications. *International Journal of Computer Science and Information Security*, 9(10), 78-83.
- Yang, S.Y., Chen, P.Z., & Sun, C.H. (2007). Using Petri Nets to enhance web usage mining. *Acta polytechnica Hungarica*, 4(3), 113-125.
- Zhu, J., Hong, J., & Hughes, J.G. (2002). *Using Markov Chains for Link Prediction in Adaptive Web sites*. Springer, 2311, 60-73.