Create a Profile for User According To His Past Choices Using Intelligent Algorithms

Habib Matin

Health Network malekshahi city, Ilam University of Medical Sciences

matin-h@medilam.ac.ir

Abstract

In this article, the researcher tries to classify the web user's navigation patterns automatically to achieve this, a new method is presented to classify user's navigation patterns and predict their upcoming requirements. Then to create users profile, a new method is introduced that by recording user's settings active and user's similarity measurement with neighboring users, the method is able to create the profile implicitly. Furthermore it updates the profile based on created changes. In fact, the researcher is trying to improve the function of recommender engine by using user's navigation patterns and clustering. This method that is based on user's navigation patterns is able to present the result of recommender engine based on user's requirement and interest. In addition, this method has the ability to help in customizing website and its more efficient organization.

Keyword: Web mining, clustering, recommender engine, web usage mining, neural network.

1. Introduction

Great explosion of World Wide Web during past 15 years has provided lots of growing data for the users. Finding relevant data and knowledge through too much data is not only difficult but also time-consuming. Then it seems necessary to have an intelligent system that is able to filter irrelevant interests or provide the required data in the time on the basis of user's interests. By suggesting items, recommender systems are trying to improve this problem for users.

In [2], a combined recommender system is presented that combines the results of two methods of recommender systems based on content and collaborative filtering in the form of two layer graph and based on partition graph. This system obtains the similarity between web pages and users and applies partition graph to classify users and web pages. This article puts into the issue the criteria of coverage and precision.

In [1], a combined algorithm is presented that uses user's browsing data and link among pages to suggest pages to users. To calculate the weight of pages visited by users, introduced criteria takes advantage from duration and frequency of page view. To classify pages, algorithmic based on learning automata and partition algorithm of graph are presented. In this article, the researcher uses the two meter coverage and precision.

In [9], by studying transaction of searching weblogs, Nicholas and et al explain that calculation devices discovered from these sources, are useful means in order to examine the rates of output and the rates of satisfaction and dissatisfaction of search engine. They applied both criteria of calculating passed time in searching session and the number of performed searches in each session to study the behavior of web users during data obtaining.

So in this article the researcher introduces a new method that creates the profiles implicitly and updates them based on created changes by recording user's setting and the criteria of similarity of active user and neighbor users. The mentioned way uses available data on log to recognize the user's requirements in an active way.

2. Research Background

In this part, the researcher explains the prerequisite subjects that seem to be necessary in our understanding. Web mining, customizing based on web usage mining, the method of clustering and neural network are concepts that appear sequentially in this article.

2-1-Web Mining

Web mining is using data mining technique that by means of it the user becomes able to find necessary data among documents and web services. Web structure mining: is a process that analyzes nodes and structural relations in a website by using models of graphs. Web content

mining: is a process that deals with discovering useful information from texts, images, voice and visual data in web. Web usage mining: this process concentrates on techniques that are able to

predict user's behavior interacting in web. The main functions in web usage mining are retrieving meaningful and useful data from profile storage and using web servers based on user's browse.

This process itself is divided into three parts: pre-processing, pattern discovery and pattern analysis.

2-2-Clustring

The collection of input models X= {x1, x2,…, xn} includes n objects. Each object from this collection equals with a vector with the length s of properties. These objects must be clustered in K groups n named C={C1 , C2 , … , Ck } that don't overlap with each other. In this article, k-means algorithm is used to cluster similar users. Despite of simplicity of this way, it is known as a basic way for lots of clustering way.

2-3-Neural Network

Artificial neural network is an idea for processing information inspired from biological neural network and this network processes information like the brain. This system consists of large numbers of processing elements called neuron that act harmoniously to solve the problem. The distinctive advantage of these networks is their excessive ability beside their simplicity while using.

3-Research Methode

Data recovery often comes along with error since the available profiles in a server, saved sub sequentially, don't belong only to a user but they are available for different ones; furthermore different search information is kept for each user as well. Then these data should be pre-processed and prepared before using. Processing web logs includes data cleaning, user identification and user session identification. After preparing data and indentifying users and their session, session vector is made in the following way:

User's session can be explained in the form of a vector of weight of page views during a particular period. A session consists of all activities performed by the users from their arrival to site till their departure. A threshold is considered for the session duration .if this duration excesses from certain level, it is a sign of another access session of user. Based on this experiment, a thirty minute threshold is suggested for session duration [14-3].

Then we show the user session in the following way: suppose p as a collection of all accessible pages by site users like this p= {p1, p2, …, pM} provided that each pi be distinguished by a particular url.

The collection of S also shows a subset of access sessions of users provided that each Si be a subset of P.

$$S= \{s1, s2,\ldots,sn\}$$

Each session is a M dimensional vector like below:

$$Si=\{W(P1,Si),W(P2,Si),\ldots,W(Pm,Si)\}$$

Provided that the weight of each pages of pj be determined in i session and every page weight shows the amount of user's interest to that page.

In fact, to determine weight and amount of user's interest to the page, two factors of frequency and duration of page must be considered like this:

$$frequency(page) = \frac{Number\ of\ visits\,(page\,)}{\sum_{page\,\in visited\ pages} \left(Number\ of\ visits\,(page\,)\right)} \quad (1)$$

$$\text{Duration(page)} = \frac{Total\ Duration\ (page\,)/Leng\,ht(page\,)}{Max_{page\,\in visited\ pages}\left(Total\ Duration\ (page\,)/Leng\,ht(page\,)\right)} \quad (2)$$

The importance of whole page is obtained from compounding two mentioned criteria. In this system we use from the harmonious average of frequency and duration to explain the amount of user's interest to a web page in a session like below:

$$\text{Interest(page)} = \frac{2 * Ferequency(page) * Duration(page)}{Ferequency(page) + Duration(page)} \quad (3)$$

Finally, we will have a vector like this for every session provided that Wi determines the weight of the page i in a particular session. As the number of M dimension should not excess from defined number, then the pages that amount of their support is high or low should be cleared.

3-1-Creating User's Profile

Each user has k sessions in such a manner that S1, S2, …, Sk are collection of i user sessions. Average vector of Sui is considered as a criteria or ui user interest. Weight of each page in average vector obtains from the average weight of that page in all user sessions. To achieve more efficient results, in addition to history of user's behavior, his/her trivial session can be used as well.

3-2-Clustring Profiles

Now the vector of average sessions should be compared with each other and they should be clustered on the basis of their similarity. In this algorithm, the number of clusters should be

entered to an algorithm as an input parameter and cosines distance should be used to calculate the distance between two objects.

The collection of clusters is like this:

C={c1,c2,…,cK}

As a representative for each cluster, we obtain the average of each mc cluster that shows the user navigation pattern of each cluster in a particular collection of accessible web pages. At last, as the result of profile clustering, there will be a collection like this:

NP={np1,np2,…,npK}

Provided that each pi be a subset of web page collection p. after train neural network, by entrance of new user to site, we should prepare current user session in such a way that it is suitable for entrance to neural network. Now it should be determined that the profile of current session belongs to which navigation pattern, in other words, which clusters. Then, the profile of current session is given to the entrance of neural network and the network will determine suitable cluster for the session. when the number of cluster is determined , pages of cluster that haven't been visited in current session have a high potential to be next page which the user is interested in visiting them. So they will be put in suggested list.

4. Conclusion

In this article, to provide useful and required data for users, a way was introduced on the basis of user navigation patterns that obtained results by recommender engine are formed on the basis of user's requirement. In comparison to previous methods, suggested method makes possible more

efficient research by recording changes in each user's setting. This method separates the pages which are relevant to user's interest from irrelevant ones. To examine the effect of new method, the researcher did a survey on the structure of user's profile based on the history of their behavior. If in adjustable research for each user, the researcher concentrates on user's current

session more than his/her. Search history, it will lead to more efficient results. This system uses neural network to determine classification of user similar and common interests. In comparison to systems based on associative rules, the below results are obtained for precision and coverage:
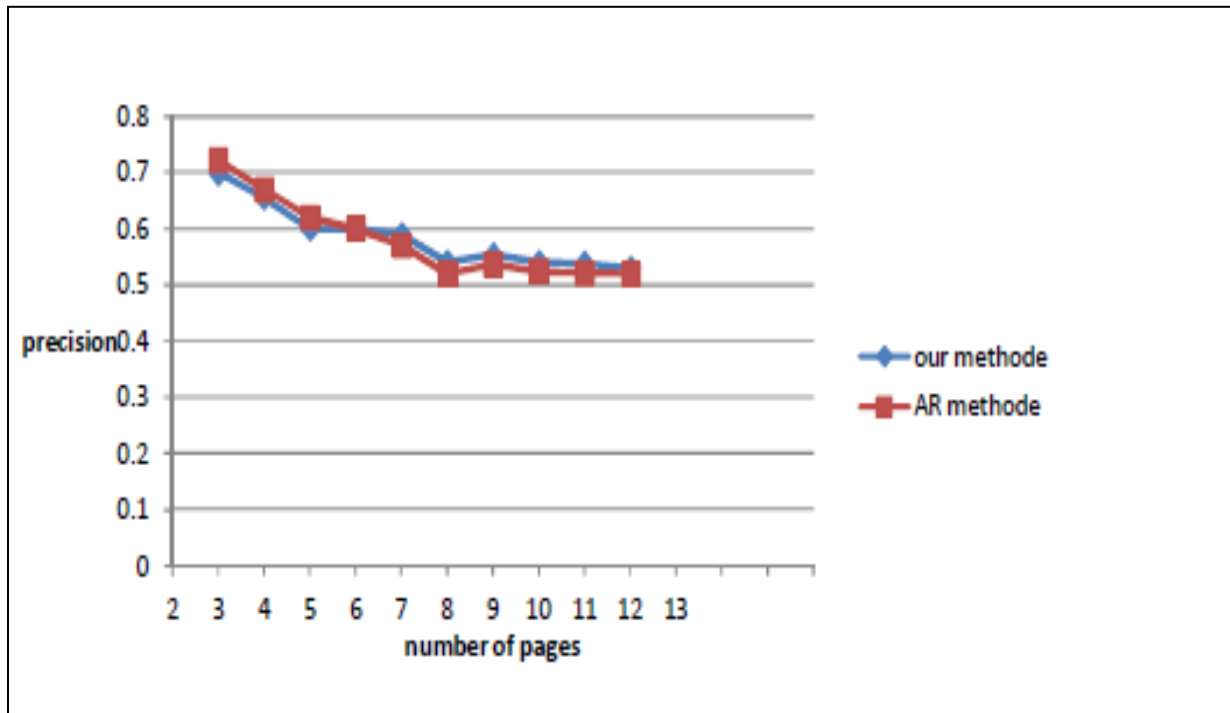
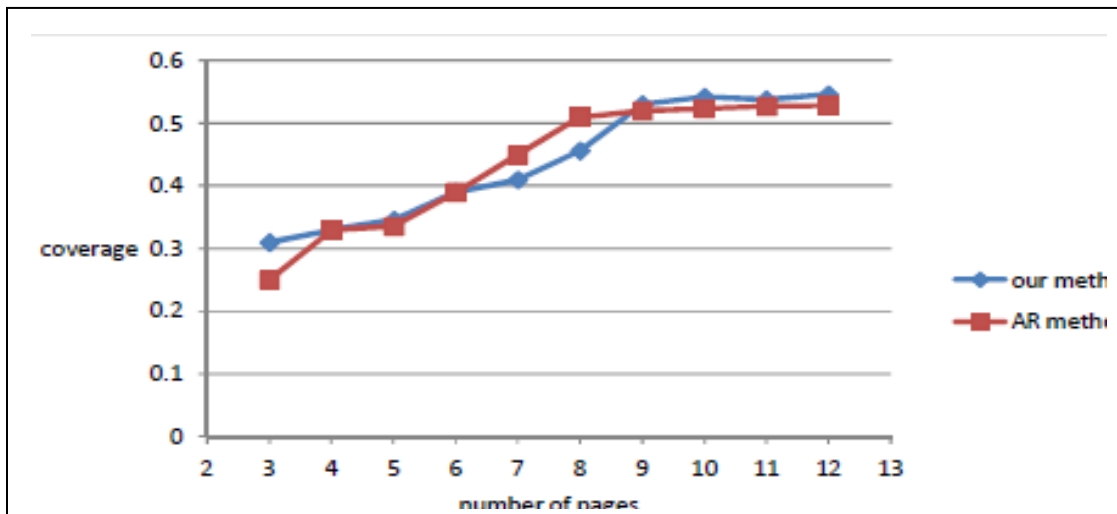Figure1: precision of the recommendations

Figure2: coverage of the recommendations

As it was explained the suggested method of this survey emphasizes on the structure of user's profile. As upcoming activities, we try to consider similar criteria in obtained clusters to be able

to calculate the quality of suggestions provided by other users. In addition, we want to attribute the users to several clusters (overlapped clusters) and use the clusters for suggestion since people usually have different interest in real occasions.

Appreciation

References

[1]forsati, R., meybodi, M.R., 2009, algorithmic based on structure pages and user's usageinformation for recommending web pages, the 2th Iran data mining conference.

[2]mohamadi dostdar, H., forsati, R., meybodi, M.R., 2011, recommender system of combined web based on 2-layers graph and partition of graph, the 5th Iran data mining conference.

[3] Berendt, B., Mobasher, B., Spiliopoulou, m., Wiltshire, j., 2001, Measuring the Accuracy of Sessionizers for Web Usage Analysis, Proceedings of the Web Mining Workshop at the 1st SIAM International Conference on Data Mining.

[4]breeding, m., 2005, analyzing web server logs to improve a site's usage, Computers in Libraries, Volume 25, Number 9, Page(s) 26-29.

[5]Cooley, R., Mobasher , B., Srivastava, J, 1999, Data preparation for mining world wide web browsing patterns, Department of Computer Science and Engineering University of Minnesota.

[6]kosala, r., blockeel, h., 2000, web mining research:a survey, ACM SIGKDD Explorations Newsletter, Volume 2, Issue 1, Pages 1-15.

[7]lin, w., alvarez, s.a., ruiz, c., 2000, Collaborative recommendation via adaptive association rule mining, Data Mining and Knowledge Discovery, In Proceedings of the Web Mining for E-Commerce Workshop (WebKDD'2000).

[8] mobasher, b., 1999, A web personalization engine based on user transaction clustering, Proc. of the 9th Workshop on Information Technologies and Systems (WITS'99).

[9] Mobasher, B., Dai, H., Kuo, T., Nakagawa, M., 2001, Effective Personalization Based on Association Rule Discover from Web Usage Data, In Proceedings of WIDM 2001, Atlanta, GA, USA, pp. 9–15.

[10] Nicholas, d., Huntington, p., jamali, h.r., Watkinson, a., 2006, the information seeking behaviour of the users of digital scholarly journals, Information Processing & Management, Volume 42, Issue 5, Pages 1345-1365.

[11] Nicholas, d., Huntington, p., Watkinson, a., 2005, scholarly journal usage: the results of deep log analysis, Journal of Documentation, Vol 61, issue 2, pp.248 – 280.

[12] Nicholas, d., Huntington, p., jamali, h.r., tenopir, c., 2006, Finding Information in (Very Large) Digital Libraries: A Deep Log Approach to Determining Differences in Use

According to Method of Access,journal of Academic Librarianship, v32, n2, p119-126. [13] Phatak, D.S., Mulvaney, R., 2002, Clustering for personalized mobile webusage, Proceedings of the IEEE International Conference on Fuzzy Systems, pp. 705-710.

[14] Spiliopoulou, m., Mobasher, B., Berendt, b., Nakagawa, m., 2003, A Framework for the Evaluation of Session Reconstruction Heuristics in Web-Usage Analysis, INFORMS Journal on Computing, Volume 15 Issue 2, Pages 171 - 190.

[15] Tuzhilin, a., Adomavicius, G., 2005, Towards the Next Generation of Recommender