

A Recommender System for Web Mining using Neural Network and Fuzzy Algorithm

Maral Kolahkaj

Department of Computer
Engineering, Science and
Research khozestan Branch,
Islamic Azad University,
Ahwaz, Iran

Ali Harounabadi

Department of Computer
Engineering, Tehran Center
Branch, Islamic Azad
University, Tehran, Iran

Mehdi Sadegh Zade

Department of Computer
Engineering, Mahshahr Branch,
Islamic Azad University,
Mahshahr, Iran

ABSTRACT

Web mining is, using data mining tools to discover knowledge from various sources in the web and according to sources type that mined, to be classified in various fields of research.

One of the important tools in web mining is mining of web user's behavior that is considered as a way to discover the potential knowledge of web user's interaction. By identifying user's behavior in some cases like:

Targeted advertisement, e-commerce, and search engines, it would be possible to provide users with desired results.

By providing information that users are interested to view it, users can be converted into permanent customers.

In this article, by identifying user's behavior and use of neural and fuzzy techniques it would present a system that will predict user's interest and will propose them a list of pages based on their interests.

So that it would enjoy fuzzy clustering method. Due to the user's different interest and use of one or more interest in a time, their use may belong to several clusters. Fuzzy clusters provide a possible overlap. Then by resulting cluster it would extract fuzzy rules. After that, it will make user's movement pattern and with the help of neural network it will propose a list of suggested pages to the users.

The results show that the proposed algorithm is in a higher level of precision and recall compared with other algorithm.

Keywords

Web Personalization, Recommender System, Web Usage Mining, Fuzzy Clustering, Neural Network.

1. INTRODUCTION

Increasing development of available information in different web pages has been resulted in many problems for users and managers. When users try to find information about specific title among suggested pages, they face a difficult and time consuming process. When web designers and managers try to adapt themselves to the needs of users and when they try to make permanent consumers, these problems are emerged. This problem is more critical in electronic trade. Since these problems are mentioned about personalize web researches, there is an attempt to decrease such problems.

Personalized web is a process in which provided information or service of a web site is adapted to the needs of a user or a specific group of users by using received knowledge of user navigation behavior and his specific interests about structure of web site which can provide active suggestion according to users behavior pattern (Forsati et al [1]). User behavior

modeling is a fundamental factor in each personalized system which is done implicitly by user information or extracted users' samples (Qaderian [2]). Using web mining in order to extract knowledge from available web information is considered one of important approaches in web personalizing and it is classified to three active researches based on extracted web data area: web content mining, web usage mining, web structure mining (chen et al [3]). Purpose of this research is providing new method in order to model users' cooperation in web by using Fuzzy –neural algorithm. Modeling user behavior in web leads to personalize users' researches results without their direct activity. The innovative aspect of this research is: using neural network algorithm and using subtractive fuzzy clustering technique in order to model user's behavior for improving web personalizing process and providing a suggested list by suggested system in researcher motors. Suggested method tries to provide suggestions for users' future request carefully.

Included chapter in this article are as follows: the concept of web usage is provided in part two. Part three includes available approaches and methods in web personalizing based on web usage mining. Part four includes the suggested method in this research. Part five provides details about using recommended method, data collection, and instruments. Part six provides information about evaluating criteria and methods, results of experiments and compares them with other methods. Finally, part seven provides conclusion and advantages of suggested method.

2. BACKGROUND

this part, explains web usage mining which is used in this article.

2.1 WEB USAGE MINING

Web usage mining refers to process of discovering interested and suitable pattern of web user's application data. Application data refers to data which is stored in log file by users when they use web. Web usage mining [4-7] is the most applicable method for extracting users' behavior pattern in these files.

Web usage mining concentrates on the techniques which can predict user behavior while cooperating in web. The principle duty in web usage mining is recovery of web servers based on users' research. This item is divided in to three phases. Preprocessing, pattern discovery, and pattern analysis [4-8].

Preprocessing: the collected web information usually consists of large and heterogenous information. The information should be changed to homogeneous and adaptable information which is suitable for pattern discovery phase. Like many data mining application, preprocessing and

preparing data include completing omitted noise, changed and formatted data and adapting insufficiencies. In web usage mining, this phase includes information purging, recognizing user and their session, which are fundamental factors for discovery pattern.

Pattern discovery: includes processing and changing server session file to rules, patterns and statistic information.

Pattern analysis: seeks rules and statistic information acquired from pattern discovery phase which is very interesting for management site personnel.

2.2 RELATED WORKS

Purtorab [9] has provided a suggested motor based on usage mining which uses structure of web site for better suggestions, it means when user goes from one page to another one, a connection between those pages is created. In this method, a tree with the maximum depth of constant is used which makes research more simple. Also, by using classified lists, tree research time is improvised very much. The researcher has used three criteria for evaluating his system, accuracy, covering, and degree of effectiveness. The criteria improve degree of effectiveness, and accuracy of suggestion but the degree of covering is less than previous work.

Khademali [10] has provided a method for classifying user movement patterns and predicting user's next requests. After providing user profile, the user uses K-means clustering for making automatic user movement patterns. After making movement patterns, a list of suggested motor of user favorite pages is recommended. Suggested motor has found a useful movement pattern for user session by using perceptron neural network which predicts users' next requests. Then a list of favorite pages is provided.

Liu [11] has used a mixture of usage mining and content mining in order to predict users' next requests. In this article, user sessions are acquired by record file preprocessing and by using techniques of usage mining on the sessions; the representative user movement pattern is provided. Then, this pattern and content web are combined to provide movement pattern based on n-gram and it is used in experimental session collection for classification and prediction. Sujatha [12] has provided a mixed system for discovery and analysis of user movement. In this article, after preprocessing record file data, a clustering algorithm based on ants for pattern discovery is used which is out of system line phase.

Based on mixture of web content mining and web usage mining, Rashidi et al [13] has recommended an experimental system which clusters users movement algorithms by using Kmeans algorithms. In this article, first, users' profiles based on acquired information from server web record is made. Then, they are combined by web page contents and users movement algorithms for predicting users' next requests. Then, they are combining with related web pages content and users' movement algorithms for predicting user next requests. Keywords (n-grams) are used for acquiring page content. Then, by using neural networks of suggested neural network, a list of user favorite pages is produced.

Chen et al [3] and Yang et al [14] by using petri networks have analyzed web pages frame work in which places in petri networks, web pages and passages are considered as hyperlinks. After analyzing this frame work a contact matrix is provided which tries to observe pages and which uses features of this matrix for completing the passage.

3. PROPOSED METHOD

In this article a new method for personalizing is provided by using fuzzy –neural system. To achieve this result, after preprocessing log file data and extracting user's features from their session and providing profiles in offline phase, this method will classify them by using fuzzy clustering method. Then, by using fuzzy inference system, fuzzy rules based on related clusters and user's interests are extracted and after acquiring their user's movement patterns, they are entered in pattern of neural network. Since users usually mention their information about their interests and preferences in fuzzy and inexact form, fuzzy –neural systems can be suitable instrument for inexact and fuzzy information (Yeung [15]).

The output of fuzzy –neural network is the clusters which contain the interests of users. Therefore, after teaching on line network, the network has the ability to put the new users in related clusters. Users who were contained in clusters can change next repetition of their clusters. In suggested method, the answers which are more important and have high degree for the user have high degree of importance in fuzzy system.

3.1 DETERMINING WEIGHT OF PAGES

According to Chan [10], the suggested algorithm is used for determining degree of user for each page based on page observation time and pages observation frequency of each user criteria. These criteria are acquired according to functions 1 and 2.

$$\begin{aligned} & \text{Frequency}(\text{page}) \\ &= \frac{\text{number of visits}(\text{page})}{\sum_{\text{page} \in \text{visited pages}} (\text{number of visits}(\text{page}))} \end{aligned} \tag{1}$$

$$\begin{aligned} & \text{Duration}(\text{page}) \\ &= \frac{\text{Total Duration}(\text{page})/\text{Lenght}(\text{page})}{\max_{\text{page} \in \text{visited pages}} (\text{Total Duration}(\text{page})/\text{Lenght}(\text{page}))} \end{aligned} \tag{2}$$

After determining time and frequency observation of pages based on function 3, the degree of importance of each page is specified.

$$\begin{aligned} & \text{Interest}(\text{page}) \\ &= \frac{2 * \text{Ferequency}(\text{page}) * \text{Duration}(\text{page})}{\text{Ferequency}(\text{page}) + \text{Duration}(\text{page})} \end{aligned} \tag{3}$$

3.2 USERS CLUSTERING

After session patterning phase, it is necessary that the available numbers in each pattern be changed to standard criteria (between zero and one). Therefore, each number is divided to collection of available numbers. The reduction clustering method is provided by Chiu [17], the potentialities of the data points are considered instead of network points in order to find center of clusters. The reduction clustering method advantage is avoidance of network separation which compromise between computations accuracy and complication.

Reduction clustering method, potentiality of each data point x_i is acquired by using function 4.

$$P_i = \sum_{i=0 \text{ to } n} e^{-\alpha |x_i - x_j|^2} \quad (4)$$

In this function, $\alpha = 4/(r_a^2)$. r_a is a constant and positive number which determines the neighboring radius for each data point. When the potentiality of all data points is computed, the data point with the high potentiality is selected as the first center of cluster. Then, the potentiality of each data point x_i is reformed by function 5:

$$P_i = P_i - P_1^* e^{-\beta |x_i - x_1|^2} \quad (5)$$

In this function, $\beta = 4/(r_b^2)$, x_1 is the first place of center of cluster and p_1 is the potential amount of first center cluster.

r_b is a constant and positive number which determines the neighboring radius r_b .

It is considered bigger than r_a in order to avoid accuracy center of clusters in neighboring distances ($r_b = 1.5r_a$). when the potentiality of all data centers is reformed based on function 5, the data point with the highest remained potentiality is selected as the second cluster centers.

Acquiring the new cluster center and reforming potentiality is continued until the function 6 is provided.

$$P_k^* < \varepsilon P_l^* \quad (6)$$

Selection of ε and r_a has an important effect on the results. If very big factors are selected, fewer data points are accepted as cluster centers and if they are very small, the acquired number of cluster centers is very large. After using reduction clustering method on acquired profiles, the number of acquired cluster are determined by using function 7 for degree membership of each cluster center.

$$\mu_i = e^{-\alpha |y - x_i|^2} \quad (7)$$

Some of features of this algorithms are (Bataineh [18]):

- Providing adaptable results in different repetition compared to FCM.
- Higher accuracy in relation to FCM.
- No need for determining the number of primary clusters in relation to k-means.
- Reducing the number of calculation in relation to other clustering algorithms.

3.3 FUZZY RULES EXTRACTION

After acquiring the center clusters and degree of membership of each cluster, this method will acquire membership functions by Gaussian distribution. Providing membership functions is done in order to improve processing in real world by using Gaussian distribution. Then, it can has fuzzy rules by using fuzzy deduction system based on mamdani method.

mamdani method is used because it is comprehensible for human entrances.

3.4 Recommending system for neural network

In previous phases, after acquiring users' Profiles, clustering profiles and determining degree of membership for users for each cluster, the fuzzy rules are extracted by using users' profiles and degree of membership of each cluster. In this phase, the system is taught by using neural network. After teaching neural network, the recommended system can predict users' interests and the related pages in this regard.

4. IMPLEMENTATION OF PROPOSED METHOD

We use Matlab software, in order to implement suggested method. The log file data of this research is collected based on nasa log file web server. After preprocessing phase and session identification, the clustering operation based on users' session is done.

Having determined cluster centers, to produce membership function and to extract rules, this method used three linguistic variables which include small, medium, and high in order to express the degree of users interest to the pages and degree of membership of each user to each cluster.

We have classified pages based on content and structure of web sites into six groups which include computing, economical, historical, amusing, and structural subjects.

Based on three linguistic variables and six groups of different pages, the number of extracted fuzzy rules will result in 36 rules. The low validity rules according to (Ziyaratban et al [19]) will be omitted from the rules collection.

The number of extracted rules is 11 and two of these extracted rules are shown in table 1.

Table 1. Extracted rules

Rule.1	If (s1 is small) and (s2 is small) and (s3 is small) and (s4 is small) and (s5 is high) and (s6 is small) then (μ_1 is high)(μ_2 is small)(μ_3 is small)(μ_4 is small)(μ_5 is small)(μ_6 is small)(μ_7 is small)(μ_8 is small)(μ_9 is small)
Rule.2	If (s1 is small) and (s2 is small) and (s3 is small) and (s4 is small) and (s5 is high) and (s6 is medium) then (μ_1 is medium)(μ_2 is medium)(μ_3 is small)(μ_4 is small)(μ_5 is small)(μ_6 is small)(μ_7 is small)(μ_8 is small)(μ_9 is small)

In this table, s1 refers to users interest about computing pages, s2 refers to economical, s3 refers to historical, s4 refers to amusing, s5 refers to structural and s6 refers to the other subjects. c1-c9 expresses the title of cluster or related cluster of each pattern as well.

5. EVALUATING PROPOSED METHOD

5.1 SUGGESTED MODEL

In order to teach neural net, this method has used 70 percent of session collections for learning and 30 percent of them for testing.

5.2 EVALUATION METRICS

Precision refers to the ability of suggested system to produce accurate suggestions. In other words, suggestion precision means the ratio of true suggestion to the total number of suggestions.

Recall refers to ability of suggested method to produce suggestions which are obvious for users. In fact; recall refers to ratio of true recognized suggestions in relation to remained pages in the same session.

5.3 SIMULATION RESULTS

In order to evaluate suggested method, it compared with the provided method in (Castellano [20]).The newer method is used for evaluation method because it is more accurate than KNN, FAR, and NP method.

The comparison between, precision and recall of recommended method with newer method is shown in Figure 1 and Figure 2.



Figure 1. Copmarison of Precision Algorithms

As can see in Figure 1, the precision of suggested method is more than newer method about size of suggested pages of 2, 5, 6, and 8.



Figure 2. Comparison of Recall Algorithms

Also can see in Figure 2, the recall of this method is more than newer method about size of suggested pages of 2, 3, 5 and 6.

6. CONCLUSION

As mentioned earlier, web is developing in a decentralized and distracted process. This process results in a large mass of connected results which does not have any logical organization.

Since presenting knowledge in real world is not trustworthy to some degree and interests of users is different from person to person, it is difficult to acquire user's movement pattern. In this research, the fuzzy and modeling clustering models are presented by the fuzzy-neural network. This article has presented a structure which helps to deal with insufficiencies of previous researches. In addition, fuzzy theory which is available in fuzzy-neural networks is related to the quality of human thought and language and it is different from probability theory. The purpose of suggested method is production of user's profiles and finding their common movement pattern implicitly. Finding user's movement pattern is done by using fuzzy clustering technique and web usage mining. The presented system finds the useful movement pattern of users and it presents user's future requests by using fuzzy -neural network.

Result of experiment shows that, suggested algorithm has very high degree of precision and recall in recommending pages for users.

7. REFERENCES

- [1] Forsati r., Meybodi.,1387, an algorithm based on structure of connected pages and information of users for suggesting web pages, the second Iran data mining conference, industrial Amir Kabir university.
- [2] Qaderian m.,1387, improving user model in website automatically by using semantics of specific degree of concepts. M.A. thesis, Amir Kabir University, information technology and computer engineering faculty.
- [3] Chen P.Z., Sun C.H. and Yang S.Y., 2008, Modeling and Analysis the Web Structure Using Stochastic Timed Petri Nets, Journal Of Software, Vol. 3, No. 8, pp. 19-26.
- [4] Verma V., Verma A. K. & Bhatia S. S., 2011, Comprehensive Analysis of Web Log Files for Mining, International Journal of Computer Science Issues, Vol. 8, Issue 6, No 3, pp. 199-202.
- [5] Thakare S. B., Gawali S. Z., 2010, An Effective and Complete Preprocessing for Web Usage Mining, International Journal on Computer Science and Engineering, Vol. 02, No. 03, pp. 848-851.
- [6] Tyagi N. K., Solanki A. K. & Wadhwa M., 2010, Analysis of Server Log by Web Usage Mining for Website Improvement, International Journal of Computer Science Issues, Vol. 7, Issue 4, No 8,pp. 17-21.
- [7] Santra A. K., Jayasudha S., 2012, Classification of Web Log Data to Identify Interested Users Using Naïve Bayesian Classification, International Journal of Computer Science Issues, Vol. 9, Issue 1, No 2, pp. 381-387.
- [8] Chitraa V., Davamani A. S., 2010, A Survey on Preprocessing Methods for Web Usage Data, International Journal of Computer Science and Information Security, Vol. 7, No. 3, pp. 78-83.
- [9] Purtorab a.,1390, a suggested web motor based on user activity mining pattern, M.A. thesis, khozestan research and since university ,computer engineering faculty.

- [10] khademali z.,1391, providing slides for users based on his cooperation in web by using intelligent algorithms, M.A. thesis, Dezful Islamic Azad university, computer engineering faculty.
- [11] Liu H., Keselj V., 2007, Combined mining of Web server logs and web contents for classifying user navigation patterns and predicting users 'future requests, *Data & Knowledge Engineering*, Vol. 61, Issue 2, pp. 304–330.
- [12] Sujatha V., Punithavalli, 2010, An Approach To User Navigation Pattern Based On Ant Based Clustering And Classification Using Decision Tress, *International Journal Of Advanced Engineering Sciences And Technologies*, Vol No. 1, Issue No. 2, pp. 112 – 117.
- [13] Rashidi S. F., Harounabadi A. and Abasi Dezfouli M., 2012, Prediction of users' future requests using neural network, *Management Science Letters*, Vol. 2, Issue 6, pp. 2119–2124.
- [14] Yang S.Y., Chen, P.Z. and Sun, C.H., 2007, Using Petri Nets to Enhance Web Usage Mining, *Acta Polytechnica a Hung Arica*, Vol. 4, No. 3, pp. 113-125.
- [15] Yeung, D. S., Liu, J., Shiu, S. C., Fung, G. S., 1996, Fuzzy Coloured Petri Nets In Modeling Flexible Manufacturing Systems, *ISAI, IFIS, ITESM*, pp. 100-107.
- [16] Chan P.K., 1999, A non-invasive learning approach to building web user profiles, in: *Workshop on Web usage analysis and user profiling, Fifth International Conference on Knowledge Discovery and Data Mining, San Diego*.
- [17] Chiu S.L., 1994, Fuzzy model identification based on cluster estimation, *journal of intelligent and fuzzy systems*, Vol. 2, pp. 267-278.
- [18] Bataineh K.M., Najia M., Saqera M., 2011, A Comparison Study between Various Fuzzy Clustering Algorithms, *Jordan Journal of Mechanical and Industrial Engineering*, Vol. 5, No. 4, pp. 335 - 343.
- [19] Ziyaratban M., Moradi M.h., Azuji M., 1385, Improving practicality of fuzzy classification by teaching membership functions and selecting rules in order to recognizing Persian numbers, *The fourth conference of visual machine and Iran processing pictures, Mashad Ferdosi University*.
- [20] Castellano G., Fanelli A.M. and Torsello M.A., 2011, NEWER: A system for Neuro-fuzzy Web Recommendation, *Applied Soft Computing*, Vol. 11, Issue 1, pp. 793-806.