FP-International Journal of Computer Science Research (IJCSR)
Volume 2, Issue 2, Pages 73-77, June 2015

Framework for Web Personalization Using Neural Network

Ramandeep Kaur
Research Scholar
Department of Computer Science
Chandigarh Group of Colleges, Gharuan
Punjab, India
88ramandeep@gmail.com

Iqbaldeep Kaur
Associate Professor
Department of Computer Science
Chandigarh University, Gharuan
Punjab, India
Iqbaldeepkaur.cu@gmail.com

ABSTRACT
In recent years there has been an increasing research interest in the problem of contextualizing search to the aim of overcoming the limitations of the "one size fits all" paradigm, which is generally applied by Search Engines and Information Retrieval Systems (IRSs). Most Information Retrieval Systems and Search Engines rely on the so called system-centered approach, where the IRS behaves as a black box, which produces the same answer to the same query, independently on the user context. This paper will propose a technique in which user context will take into regard using ANN method in which searching results can be made optimized using training and testing given by neural network.

Keywords
Neural Network, Web Personalization, Web Mining, Web Usage Mining

1. INTRODUCTION

Nowadays web users have been increased tremendously. Web users usually have very short life span in web portals while browsing and there is a big issue on getting the exact information what the web users actually wants. Here the website developers and publishers take the responsibilities of providing the efficient search results [1]. Thus they are stepping towards to optimize the web content. In order to provide the optimized search results to the users, who are in need of the proper results? In content optimization, personalization is an important feature of focusing the individual user interest rather than the common approach like "one size fits all". Personalized concepts involved in collecting, monitoring and storing present and past status of user interaction. This helps the client to get the right content at each iteration of searches [2]. During each time of searching there will be improvisation of search results by updating user interactions as specified in personalization concepts. Data mining is the powerful new technology with great potential to analyze important information in the data base. The application area of web usage mining is web personalization [3]. Web personalization is providing the user the relevant information from a large data set[4]. Here the proposed work will focus on the query system to find the most appropriate result using data mining and ANN technique.

Rest of the paper is organized as: Section II describes the concept of the web personalization, Section III gives rise to issues of data mining and Section IV the motivation of the work, Section V displays the proposed work information, Section VI displays the pseudo code, Section VII show shows the simulation result, and finally section VIII presents conclusion and future scope.

2. WEB PERSONALIZATION

World Wide Web (WWW) is the largest and most accessible source of information. Usually, web structures are large and sophisticated and users often miss the goal of their inquiry, or receive ambiguous results when they try to navigate through them [5]. Users seek a subject that they need information accordingly. Commonly, the search engines investigate the relevant web and pages according to the user query. Quite often the users find a lot of information for each subject through the web. However, one of the issues is to find the useful information from search result. Therefore, there are many studies on examining the importance of personalization in web search engine.

Personalization of search results is defined as any action to finding more relevant pages in search results list for particular user or a set of users [10]. The objective of a personalization system is to "provide information that users want or need exactly, without expecting from them to ask for it explicitly" [11].

The aim of Web usage mining is to discover patterns of user activities in order to better serve the needs of the users for example by dynamic link handling, by page recommendation etc [12]. The aim of a Web site or Web portal is to supply the user the information which is useful for him. There is a great competition between the different commercial portals and Web sites because every user means eventually money (through advertisements, etc.) [13]. Thus the goal of each owner of a portal is to make his site more attractive for the user. For this reason the response time of each single site have to be kept below 2s. Moreover some extras have to be provided such as supplying dynamic content

or links or recommending pages for the user that are possible of interest of the given user [14]. Clustering of the user activities stored in different types of log files is a key issue in the Web community. There are three types of log files that can be used for Web usage mining. Log files are stored on the server side, on the client side and on the proxy servers. By having more than one place for storing the information of navigation patterns of the users makes the mining process more difficult. Really reliable results could be obtained only if one has data from all three types of log file [15]. The reason for this is that the server side does not contain records of those Web page accesses that are cached on the proxy servers or on the client side. Besides the log file on the server, that on the proxy server provides additional information. However, the page requests stored in the client side are missing. Yet, it is problematic to collect all the information from the client side. Thus, most of the algorithms work based only the server Side data [16]. Web usage mining consists of four main steps:

A. Data collection

B. Preprocessing

C. Pattern discovery

D. Pattern analysis.

A. Data Collection: The first step in Web usage mining is the collection of data from different sources. Web usage mining applications can gather data mainly from 3 sources: (1) Web servers (2) Proxy servers (3) Web clients. The largest source of web data is web servers, the huge mass of data can be available there [7].

B. Preprocessing: The data collected in web log file is incomplete and not suitable for mining directly. Pre-processing is necessary to convert the data into suitable form for pattern discovery. Preprocessing can provide accurate, concise data for data mining [6]. Data pre-processing, includes data cleaning, user identification, user sessions identification, path completion and data integration [8].

C. Pattern Discovery: After identifying user sessions, the various techniques of web usage pattern discovery are applied in order to detect interesting and useful patterns [6]. Some of pattern discovery techniques are Path Analysis, Association Rules, Support Vector Machines Sequential Patterns, Clustering and Classification. By using this approach, web marketers can predict future visit patterns which will be helpful in placing advertisements aimed at certain user groups.

D. Pattern Analysis: The last and final step in Web Usage Mining process is Pattern analysis. Mined patterns are not suitable for interpretations and judgments. So it is important to filter out uninteresting rules or patterns from the set found in the pattern discovery phase [9]. In this stage tools are provided to facilitate the transformation of information into knowledge. Knowledge query mechanism such as SQL is the most common method of pattern analysis. Another method is to load usage data into a data cube in order to perform OLAP operations.

3. VARIOUS ISSUES IN DATA MINING

Caching: The data is stored on the server in cache hierarchy. It is possible to mismatch in the local cache data access patterns and web server log records [17]. E.g. user has visited page hierarchy as page 1, page2, page1, page3 but due to data in caching server has recorded log as page1, page2, page3 as second time access of page 1 would directly been from cache. So the second entry of page 1 is missed from log [18]. So, we cannot say that log that every time 100% correct data. Thus, caching is a very big issue for accessing web data.

CGI Data: CGI is referred as Common Gateway Interface and it is used to pass variables and user entered data to respective server. CGI has a functionality to hide the username and value pairs from URI. So, the data is accessed by whom cannot be tracked by usage mining methods.

Session Identification: Tracking and finding the session creation and usage duration and especially when parallel login with same account through different machines makes identification complex.

Dynamicity of Pages: Dynamic pages may change their content according to User request or fixed time interval. So, even minor change in content makes the log data huge as result.

Transaction Uniqueness: Issues in identifying unique users and their unique transactions as same account multiplicity is available.

4. MOTIVATION

Personalized system helps every user to overcome the mentioned problems. First of all, the system has to extract interesting keywords for each user. There are two ways to finding interest keywords: explicitly and implicitly. In the explicit approach, user should fill up the registration forms or rate on the visited pages, while implicitly approach finds interesting keywords by examining the historical search and analyzing the user behavior in the web browsing. Second process for personalization system is to apply these keywords for exploring more relevant pages in search results list by filtering and re-ranking techniques [19].

5. PROPOSED ALGORITHM

Table 1. The Algorithm for Web Personalization

Algo_Web

Start

Initialize globaltesting_datauser_found

{

extracting the words from the paragraph

}

result=generator(current_word)

load architecture

Select A Training File To Upload

collecting data from the excel sheet

for setting up the target;

for (i=1,i<=rows,i++)

{

for k=1,k<=coloums,k++)

{

If

Target(j)=p;

}

Else

p=p+1;

}

initiating the neural network

- loadsvm_group
- loadtraining_data;

for (fortempind (tp_{ind})=1, tp_{ind}<=itrind, tp_{ind++})

{

tst=test(tempind,:);

C=Cb;

T=Tb;

u=unique(C);

N=length(u);

c4=[ ];

c3=[ ];

j=1;

k=1;

if

{

(N>2)

itr=1;

classes=0;

cond=max(C)-min(C);

while((classes~=1)&&(itr<=length(u))&& size(C,2)>1 &&cond>0)

if you increase the data you will have to adjust the groups also

}

Else

Database Updated

}

Draw Plots

End

6. RESULTS AND IMPLEMENTATION

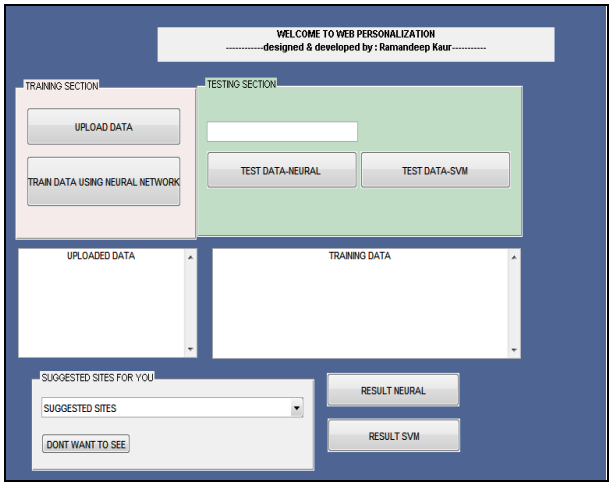

Figure 1. Main Working Window

Theabove figure shows the main working window of the personalization. It has all the training and testing window components in which the personalized data can be trained through the Neural Network and Support Vector Machine. Training Model for SVM as well as Neural Network.

Inputs: examples, a set of examples, each with input x = x1; x2; : : : ; xn and output y

Inputs: network, a perceptron with weights Wj ; j = 0; : : : ; n and activation function g

Repeat for each e in examples do inPnj = 0Wjxj [e] Err y[e] - g(in)

WjWj + _ _ Err _ g0(in) _ xj [e]

End

Until all examples correctly predicted or stopping criterion is reached Return network.

Figure 2 represents the architecture of the Neural Network. Neural network contains of input and hidden layers. Each and every layer has weight and bandwidth of the data. Hidden Layer contains epochs that means iteration. The maximum iteration provided over here is 50 but it is not necessary that the neural will run till 50. It would cross check the validations and would provide the results required. The results can also be checked by the following graphs.

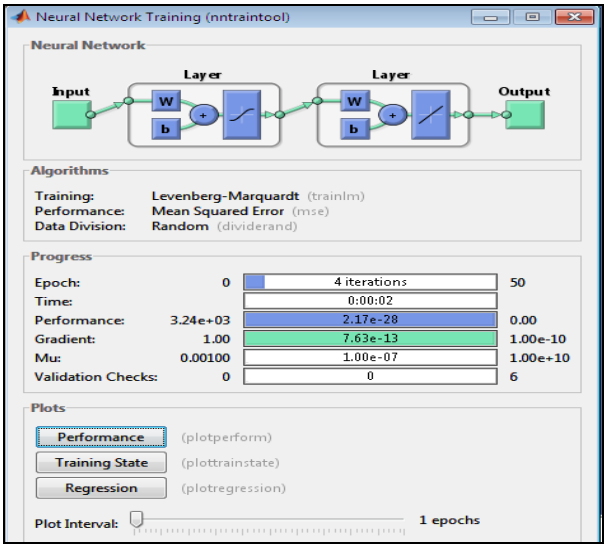

Figure 2. Neural Network Architecture

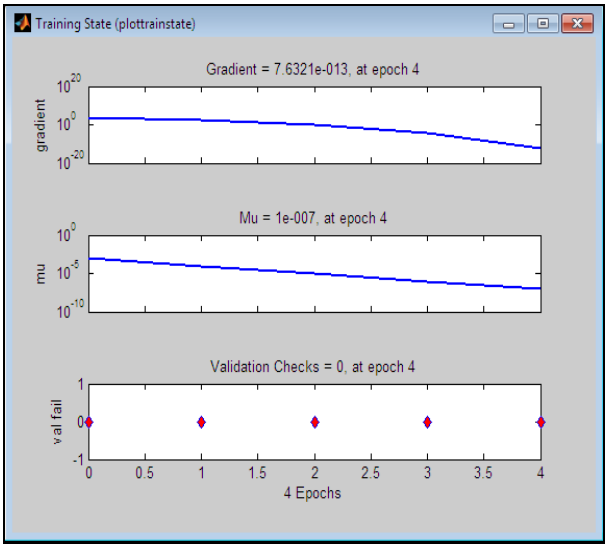

Figure 3. Detailed Neural Architecture

The figure 3represents detailed neural architecture. This figure represents the architecture over which the neural has been tested and trained. There is one validation denoted by the pink line and has been achieved on the 4th Iteration.

Figure 4. Personalizing Option

The above figure provides the option to personalize the system according to the choice of the user. Here the user can banned those website link which he or she does not want to see in the future.

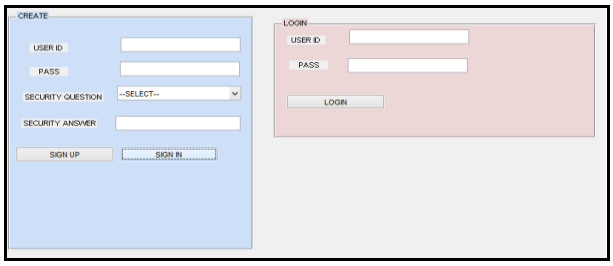

Figure 5. Login Window

This figure shows the login window. If the user is new or not registered then he can use sign up option for registration. After filling details, user is registered. User is already registered he can log in using the User ID and Password.

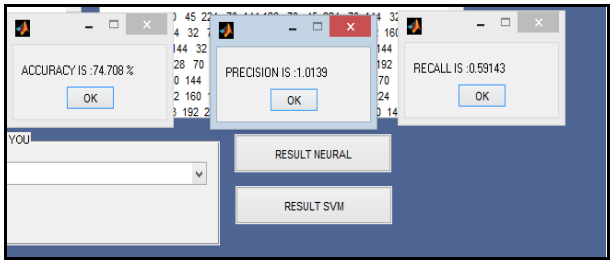

Figure 6. Performance Parameters

The above figure depicts the different parameters after clicking on result neural button the Accuracy, Precision, and recall parameters can be calculated. The same parameter can be calculated by SVM also.

7. CONCLUSION AND FUTURE SCOPE

Web usage mining is the process of identifying representative trends and browsing patterns describing the activity in the web site, by analyzing the users' behavior. Web site administrators can then use this information to redesign or customize the web site according to the interests and behavior of its visitors, or

improve the performance of their systems [21, 22]. Moreover, the managers of e-commerce sites can acquire valuable business intelligence, creating consumer profiles and achieving market segmentation. There exists number of methods but none has been achieved good quantity. This work presented an approach based on neural network for web personalization of web content.

Future lies to investigate the hybrid usage-structure ranking that can be applied to a unified web/navigational graph which expands out of the limits of a single web site. Such approach would enable a "global" importance ranking over the web, enhancing both web search results and the recommendation process.

8. REFERENCES

[1] Sugiyama, K., K. Hatano, and M. Yoshikawa, (2004) "Adaptive web search based on user profile constructed without any effort from users", Proceedings of the 13th international conference on World Wide Web, ACM: New York, NY, USA, pp. 675-684

[2] Jeh, G. and J. Widom, (2003) "Scaling personalized web search", Proceedings of the 12th international conference on World Wide Web, ACM: Budapest, Hungary, pp. 271-279

[3] Tanudjaja, F. and L. Mui, (2002) "Persona: A Contextualized and Personalized Web Search", IEEE Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS'02), pp. 1232-1240

[4] Teevan, J., S.T. Dumais, and E. Horvitz, (2005) "Personalizing search via automated analysis of interests and activities", Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, ACM: Salvador, Brazil, pp. 449-456

[5] S.Gauch, J. Chaffee, and A. Pretschner, (2003) "Ontology-based personalized search and browsing", Web Intelligence and Agent Systems, Vol.1, Issue 3, pp. 219-234

[6] B.Ketul Patel, (2012) "Process of Web Usage Mining to find Interesting Patterns from Web Usage Data" International Journal of Computers & Technology, Vol 3, No. 1

[7] Mr. Jaykumar, (2013) "A Survey on Web Usage Mining with Neural Network and proposed solutions on several issues", journal of Information, Knowledge and Research in Computer Engineering, Vol. 02, Issue 02, Page 330

[8] Li Mei, Feng Cheng, (2010) "Overview of Web Mining Technology and Its Application in E-commerce", 2nd International Conference on Computer Engineering and Technology, Vol. 7, pp. 277-280

[9] V.Chitraa(2010) "A Survey on Preprocessing Methods for Web Usage Data" International Journal of Computer Science and Information Security

[10] L. Bentley, (1979) "Multidimensional Binary Search Trees in DatabaseApplications", IEEE Transactions On Software Engineering, VOL. SE-5, issue 4

[11] http://pages.cs.wisc.edu/~bolo/shipyard/neural/local.html

[12] http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html

[13] K. Jung, (2001) "Modeling web user interest with implicit indicators", MSc Thesis, Florida Institute of Technology

[14] J.Goecks, and J. Shavlik, (2000) "Learning users' interests by unobtrusively observing their normal behavior", Proceedings of the 5th international conference on intelligent user interfaces, ACM: New Orleans, Louisiana, United States, pp.129-132

[15] D. Goldberg, et al., (1992) "Using collaborative filtering to weave an information tapestry". Communication ACM, Vol.35 Issue 12, pp. 61-70

[16] F.Gao, et al., (2007) "Personalized Service System Based on Hybrid Filtering for Digital Library". Tsinghua Science & Technology, Vol.12, Issue1, pp. 1-8

[17] M.Balabanovi, and Y. Shoham, (1997) "Fab: content-based, collaborative recommendation". Communication ACM, vol.40 issue 3

[18] L. Page, et al., (1999) "The PageRank Citation Ranking: Bringing Order to the Web", Technical Report

[19] S.Brin, R. Motwani, and T. Winograd, (S.) What can you do with a web in your pocket, Data Engineering Buletin, Vol.21,Issue2, pp. 37-47

[20] K. Bharat and G.A. Mihaila, (2002) "when experts agree: using non-affiliated experts to rank popular topics" ACM Trans. Inf. Syst., Vol.20, Issue1, pp. 47-58

[21] T.H.Haveliwala, et al., (2002) "Evaluating strategies for similarity search on the web", in Proceedings of the 11th international conference on World Wide Web, ACM: Honolulu, Hawaii, USA, pp. 432 - 442

[22] Z.Dou, et al., (2009) "Evaluating the Effectiveness of Personalized Web Search". IEEE Transactions on Knowledge and Data Engineering, Vol.21, Issue 8, pp. 1178-1190