

Algorithm Design for Robust Web Personalization Using Neural Network

Ramandeep Kaur ^[1], Er. Iqbaldeep Kaur ^[2]

Research Scholar ^[1], Associate Professor ^[2]

Department of Computer Science and Engineering

CGC Gharuan

Chandigarh University, Gharuan

India

ABSTRACT

In the recent few spans the World Wide Web has become the leading and most general way of communication and information sharing. World Wide Web is a huge source of web pages and links. The size of information available on the internet is increasing exponentially, as around one million pages are added every day. So there is a problem of information overload. The dynamic and heterogeneous nature of the web, makes web site search process very difficult for the end users. The users are provided with more information and service options. It becomes very difficult for end users to access the useful and relevant information from the web. Usually, every user has different information requirements for their query. But typical search engines return the same result for the same query submitted by different users. To solve the information overload problem and provide the relevant information to users Web Personalization is used. Web Personalization increase the accuracy of search engine, simplifies the searching process, save the time and provide relevant information to users. In this paper we present an approach based on Feed forward Back propagation Neural Network for web personalization of web content.

Keywords:- Web Personalization, Data Mining, Neural Network, Feed forward Backpropagation Neural Network

I. INTRODUCTION

In these days Data Mining has an important place in World Wide Web. As World Wide Web contains huge amount of data and information. So processing of data available on World Wide Web is very important to extract the useful information and knowledge for the large amount of data. Web mining is the application of data mining techniques to automatically determine and mine information from Web documents and facilities [1]. Web Usage Mining is a process of applying data mining techniques to discover interesting patterns from Web usage data. Web usage mining provides better understanding for serving the needs of Web-based applications [2].

1.1 Web Personalization:

The World Wide Web has become the most popular way of communication and the huge repository of information that can come either from the Web pages publicly available, or from the web usage logs daily collected by all the servers around the world to record the user's accesses [3]. The World Wide Web has been adopted by the mass market very quickly. Every day the billions of pages are added and accessed by the users from the different part of earth therefore the volume of

information available on the internet is growing exponentially and web logs files are also growing at a faster rate and the size of information available on internet is becoming huge. With the explosive growth in the size; the users are provided with more information and service options but they have to spend more time on the web to find the relevant and interesting information. Web Personalization is used for provide the more relevant and interesting information to users. It is difficult to personalize World Wide Web because web is a place for human to human communication whereas personalization requires software system to take part in interaction [4]. The aim of a personalization system is to deliver information which customers want or need correctly, without expecting from them to request for it openly [5].

1.2 Neural Network:

Neural Network is also known as Artificial Neural Network. An Artificial Neural Network (ANN) is an information-processing standard that is inspired by the biological nervous systems, such as the brain. A neural network is also referred as Neuron computers, connectionist networks, parallel distributed processors etc. [6]. Artificial neural

network is a knowledge processing paradigm that inspired from biological nervous system [7].

Artificial Neural Networks (ANNs) have the capability to made difficult non-linear connections and are skilled of approximating any computable function [8]. We 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. The major studies reporting the use of neural networks in web searching have been carried out in web content mining, personalization, clustering and web content classification and result relevancy [9]. The feedforward neural network was the first and simplest type of artificial neural network. In this network, the data or information moves only in one direction, i.e. in forward direction. The information or data is move from the input nodes to the hidden nodes and then to the yield nodes. There are no sequences or cycles in the network. Neural networks are basically works in three layers as input, hidden and output layer.

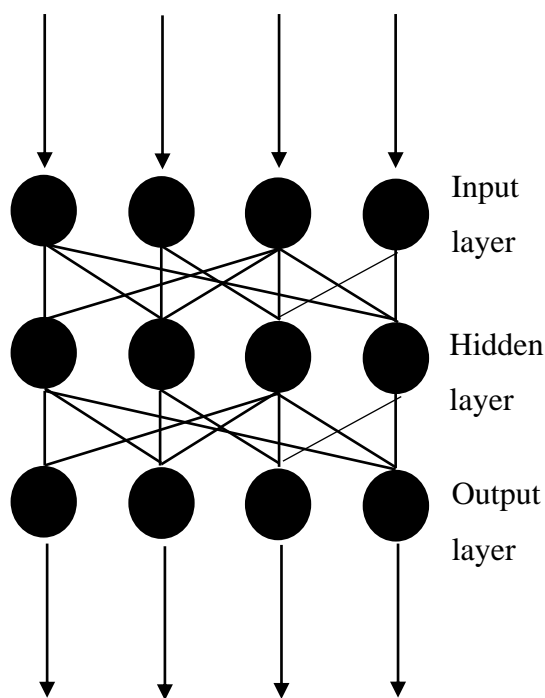


FIG. 1. Layers of Neural Network

Advantages of Neural Network:

- 1) **Adaptive Learning:** It mean the capability to learn in what way to do jobs based on the data specified for training or personalise experience.
- 2) **Real Time Operation:** ANN calculations might be accepted in equivalent and distinct hardware devices are being planned and

manufactured that take benefit of this ability [9].

- 3) **Inexpensive:** ANN’s are relatively low-cost to construct and train
- 4) **Self-Organization:** An ANN can generate its personal association or representation of the data it accepts in learning time [10].
- 5) **Fault tolerance via multiple information copies:** partial destroys or failure of network cannot affect the performance of the network [11].

II. RELATED WORK

The web personalization has become an important tool for both Web-based organizations and for the end users. The focus of the researchers is on automatic, dynamic or a combination of the two approaches over customized personalization [12]. **L. Page** proposed the personalized web search by modifying the global Page Rank procedure with the input of bookmarks or addresses of a user [13].

Letizia is considered to be the first system that records the user’s navigation behaviour and gives interesting recommendations to the user [14]. WebWatcher is web content based system that provides navigation hints to the user. This system uses personal profiles of users and recommends other items or pages based on their content similarity to the items or pages that are in the user’s profile [15]. **Goecks** proposed a different technique to develop an intelligent web browser for web personalisation that found user’s concern without the requirement for openly score pages. They considered mouse movement activity in addition to user surfing activity [16].

Haveliwala used personalized PageRank scores to enable topic sensitive web exploration. They decided that the usage of personalized PageRank scores can improve web exploration, however the number of hub vectors (e.g., number of remarkable web pages used in a bookmark) used was restricted to 16 due to the computational necessities [17]. **Gao et al.** suggested a recommendation technique for personalized service in digital recourse that combined partition-based collaborative filtering and meta-data filtering. In partition-based collaborative filtering the user-item ranking matrix can be divided into short dimensional compact matrices using a matrix clustering algorithm [18]. **L. Bentley et al.** investigates multidimensional binary search trees from the viewpoint of the database designer. Various types of search in KD have been discussed by the author as exact match query, partial match query, range query, best match query and other query [19]. In 2000 **Mobasher** proposed the web usage-based Web personalization system called Web Personalizer for recommending Web pages on Server-Side to users. The Web

Personalizer provides a personalization framework based on web log mining and using data mining techniques for extraction of knowledge for generating the recommendations to current users based on their browsing navigational history [20]. **Castellano et al.** applied neuro-fuzzy strategy to develop a Web personalization system that dynamically suggests interesting URLs for the current user [21].

III. ALGORITHMS USED IN WEB PERSONALIZATION

3.1 Brute Force Algorithm

Brute force Algorithm is a very general problem-solving technique. Brute-force search is easy to implement, and will every time discover a solution if it exists. Brute-force search is usually used when the problem size is restricted and when the ease of implementation is more significant than speed. Brute-force search or exhaustive search is a very common problem-solving method which consists of systematically computing all possible applicants for the solution and inspection whether each contestant satisfies the problem's statement [22].

TABLE 1. Brute Force Algorithm

Algo_Brute-Force
Start
Algorithm (A[0..n-1])
for i ← 0 to n-2 do
min ← i
for j ← i + 1 to n-1 do
if A[j] < A[min] min ← j
swap A[i] and A[min]
Stop

3.2 K-D Tree Algorithm

It is the type of structure which stores data. It is used for arranging some number of points in a space. It has k dimensions. So it is a BST with some constraints applied on it. KD trees are very efficient for nearest neighbour searches and range searching. The aim of Algo is to divide space. It divides such that it remains with small number of cells. So bigger input objects are not taken by cell. This makes Algo fast. Algorithms make KD trees by dividing point sets. Algo works in different dimensions. It divides data set points in all directions. In parent node, the children nodes are divided to equal sides. The process goes on. Separation stops at level n [23].

TABLE 2. K-D Tree Algorithm

Algo_K-D Tree
This algorithm is passed at root P.
Check for NULL tree, If ROOT= v, then set ROOT= P.
Compare If K (P) = K (Q) for 0 < I < k-1
Move Down: Set Q ← SON (Q)
Insert New Node Set SON (Q) ← P
Where SON is field of tree.
Stop

3.3 K-Nearest Neighbor Algorithm:

Supervised learning is done in KNN. It is used in various fields like statistical pattern recognition, data mining, etc. It uses the Euclidean distance. Although there are numbers of distance measures but one of them is Manhattan distance [22].

The Algorithm of the KNN is as preceding:

1. Define K. K should be equivalent to nearest neighbor.
2. Compute distance among the training samples and query examples.
3. All the training samples distance should be calculated. Calculate nearest neighbour.
4. Obtain all the classes and sort according to K.
5. Nearest neighbor is used predict values of the query instance.

IV. PROPOSED WORK

This work proposed an approach for web Personalization using neural network. Neural Network is computational model that is developed based on biological nervous system. Artificial Neural Network is based on artificial neurons that are connecting with each other. In Proposed work feedforward backpropagation neural network and SVM are used. SVM stands for Support Vector Machine.

TABLE 3. New Algorithm

Algo_New
Start
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 (tpind)=1, tpind <=itrind,
tpind++)
{
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
    
```

V. CONCLUSION

It is very difficult task to satisfying individual user’s needs. Due to the development of internet, people are getting more and more dependent on the search engines for their information needs. Web search engines are trying to satisfying the user’s information needs. But still there are some challenges for web search engines. Especially when the same query is submitted by the different users for their different needs but the same result is

returned by the search engine to different users. To solve this problem Web Personalization approach has been proposed. This paper provides the overview of Web Usage mining, Web Personalization, Neural Network and the different algorithms used for Web personalization.

REFERENCES

- [1] Raymond Kosala and Hendrik Blockeel (2000) “Web Mining Research: A Survey” SIGKDD Explorations, Volume 2, Issue 1.
- [2] Bayir Murat Ali (2006) “A New Reactive Method for Processing Web Usage Data”
- [3] Rajesh K Shukla (2012) “Existing Trends and Techniques for Web Personalization” International Journal of Computer Science Issues, Vol. 9, Issue 4.
- [4] Rutuja S .Lachake (2014) “A Survey on Personalize Search: An Web Information Retrieval System”
- [5] S.A. Sarabjot and Maurice D. Mulvenna (2000) “k on the Net using Web mining: introduction” Communications of the ACM, Vol. 43, No. 8.
- [6] Er. Romil V.Patel (2012) “Introduction to Integrating Web Mining with Neural Network” International Journal of Computer Science and Information Technology & Security (IJCSITS), Vol. 2, No.6.
- [7] Sonali Muddalwar, Shashank Kawan, (2012) “Applying Artificial Neural Networks in web Usage Mining”, international journal of computer science and management research, Vol. 1, Issue 4.
- [8] Sunita Yadav (2011) “Neural Network based Approach for Predicting User Satisfaction with Search Engine” International Journal of Computer Applications, Volume 18, No. 5.
- [9] P. Arumugam “Advanced Web Usage Mining Algorithm using Neural Network and Principal Component Analysis” International Journal of Computer Science & Communication Networks, Vol 3, Issue 3.
- [10] Anshuman Sharma (2012) “Web Usage Mining using Neural Network” International Journal of Reviews in Computing, Vol. 9.
- [11] Mr. M. Jagani Jaykumar “Survey on Web Usage Mining with Neural Network and Proposed Solution on several issues” Journal of information, knowledge and research in computer engineering, Volume 2, Issue 2.

- [12] D. S. Weld, et al., (2003) “Automatically Personalizing User Interfaces” International Joint Conference on Artificial Intelligence, pp. 34-39.
- [13] L. Page, et al., (1999) “The PageRank Citation Ranking: Bringing Order to the Web” Technical Report.
- [14] H. Lieberman, (1995) “Letizia: An agent that assists web browsing” Fourteenth International Joint Conference on Artificial Intelligence, pp. 924–929.
- [15] T. Joachims, D. Freitag, and T. Mitchell (1995) “Webwatcher: A tour guide for the World Wide Web” 15th International Conference on Artificial Intelligence, Nagoya, Japan.
- [16] 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, pp.129-132.
- [17] 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.
- [18] F. Gao, et al., (2007) "Personalized Service System Based on Hybrid Filtering for Digital Library". Tsinghua Science & Technology, Vol 12 issue1, pp. 1-8.
- [19] L. Bentley, (1979) “Multidimensional Binary Search Trees in Database Applications”, IEEE Transactions on Software Engineering, Vol. SE-5, No. 4.
- [20] B. Mobasher, R. Cooley, and J. Srivastava (2000) “Automatic personalization based on web usage mining” Communications of the ACM, Vol.43, pp 142-151.
- [21] A. M. Castellano, Fanelli and M. A. Torsello (2011) “NEWER: A system for NEuro-fuzzy Web Recommendation G” Proceedings of the 7th Asia Pacific Industrial Engineering and Management Systems, pp 793–806.
- [22] Stoimen., (2012) “Stoimen’s web log: Computer Algorithms: Brute Force String Matching”.