Volume 3, Issue 6, June 2013 ISSN: 2277 128X
International Journal of Advanced Research in
Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Web Usage Mining Using Self Organized Map

Amit Kumar Mishra, Mahendra Kumar Mishra, Vivek Chaturvedi, Santosh Kumar Gupta, Jaiveer Singh
Department of Computer Science & Engineering
Krishna Institute of Engineering & Technology,
Ghaziabad–201206, India

Abstract: - Web mining is used to discover relevant, useful and hidden information from the Web data. In the web usage mining, a category of web mining, we focused on knowledge discovery from the usage data of individual web site. It is an emerging field of research due to increase attraction of users towards World Wide Web. By understanding the behaviour of the user navigation pattern on a website, website personalization can be done thereby increasing the benefit of business, advertisement, and many more purposes. This paper attempts to personalize the website by the use of Self Organized Map and clustering technique user navigation behaviour based on past history.

Keywords: - Web mining, Web usage mining, Self Organized Map, Navigation pattern.

I. INTRODUCTION

There is an exponential increase of data available on the Web. The number of pages available on the Web is currently around 1 billion and is increasing at the rate of approximately 1.5 million per day. The Web-based business has been a key driving force for this rapid growth of the Web. Retailers on the Web need the ability to track users' browsing behavior history, which can increase the sale and build a strong customer relationship. This ability also can personalize the retailer's Web pages for different individual customers. Although Web log mining is a relatively new field, it has generated a lot of interest and research in the past ten years. As a sub research field of Web Usage Mining, Web log mining is the process of applying data mining technologies to discover usage patterns from the Web data. One important source to discover such patters is the Web log data that contains users Web browsing history. Most of Web log data is generated automatically by Web servers. From the last few decades, there has been witnessed an explosive growth in the information available on the World Wide Web (WWW). Today, web browsers provide easy access to myriad sources of text and multimedia data. More than millions of pages are indexed by search engines, and finding the desired information is not an easy task. The users want to have the effective search tools to find relevant information easily and precisely. The Web service providers want to find the way to predict the users' behaviors and personalize information to reduce the traffic load and design the Web-site suited for the different group of users. This profusion of resources has prompted the need for developing automatic mining techniques on the WWW, thereby giving rise to the term "web mining" [22]. Web mining is the application of data mining techniques to discover patterns from the Web. It can be classified into three categories[18][20]: Web Usage Mining, Web content Mining and Web Structure Mining, shown in the Fig. 1. Web usage mining [10] is a kind of web mining, which exploits data mining techniques to discover valuable information from navigation behavior of World Wide Web users. There are generally three tasks in Web Usage Mining: Preprocessing, Pattern analysis and Knowledge discovery [21]. Preprocessing cleans log file of server by removing log entries such as error or failure and repeated request for the same URL from the same host. The main task of Pattern analysis is to filter uninteresting information and to visualize and interpret the interesting pattern to users. The statistics collected from the log file can help to discover the knowledge. This knowledge collected can be used to take decision on various factors like Excellent, Medium, Weak users and Excellent, Medium and Weak web pages based on hit counts of the web page in the web site.

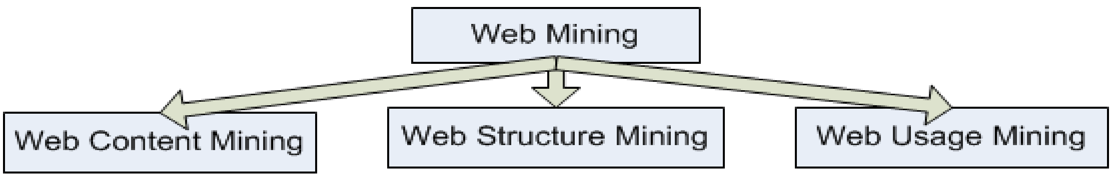

Fig. 1 Web mining categorization

II. Related Work

A lot of work in the field of Web Usage Mining has been done already. In [1], authors have proposed to mine Web usage data on client side or personal Web usage mining, as a compliment to the server side Web usage mining. By mining client side Web usage data, more complete knowledge about Web usage can be obtained. In this paper, author advocate personal Web usage mining as a complement to the server side Web usage mining. Our proposal shares similar goals with many those agents, our approach is automatic that it does not require user's explicit input. Moreover, we take a systematic approach to collect and comprehend user activities. We provide a general framework for collecting, mining, and search/query personal usage data, which may be employed by various agents. In addition, the applications of our framework are not limited to agents, but may include many others such as Web personalization, learning, and security. The authors have proposed the mining of client side Web usage data, which is termed personal Web usage mining. Based on our analysis, it is an interesting and important research area. In [2], author presented a Web mining strategy for Web personalization based on a novel pattern recognition strategy which analyzes and classifies both static and dynamic features. The results of experiments on the data from a large commercial web site are presented to show the effectiveness of the proposed system. The Approach used In this paper, they are interested in the web usage mining domain, which is usually described as the process of customizing the content and the structure of web sites in order to provide users with the information they are interested in, without asking for it explicitly. In this paper they describe their novel web usage mining strategy. It consists of two phases: in the first one a pattern analysis and classification is performed by means of an unsupervised clustering algorithm, using the registration information provided by the users and second one reclassification is iteratively repeated until a suitable convergence is reached. The authors have introduced an interesting solution based on pattern recognition techniques, and to classify a web user, based on its interaction with the web site. And off-line processing approach is used for classification tasks. In [3], authors describe a complete framework for web-usage mining to satisfy the challenging requirements of web-personalization applications. Web-usage mining should allow a compromise between scalability and accuracy to be applicable to real-life websites with numerous visitors. Within our web-usage-mining framework, they introduce a distributed user-tracking approach for accurate, scalable, and implicit collection of the usage data. They also propose a new model, the feature-matrices (FM) model, to discover and interpret users' access patterns. The Approach used in this paper, they defined a framework for web usage mining (WUM) that satisfies requirements of web-personalization applications. The framework is composed of an accurate tracking technique, and a new model (the FM model) to analyze users' access patterns. The FM Model, which is a generalization of the vector model, allows for a flexible, real time, and adaptive WUM. Dynamic clustering is possible since unlike the Markov model, incremental updating of the FM model has a low complexity. The authors defined a web usage mining (WUM) that satisfies requirements of web-personalization applications. And finally conducted several experiments, like high accuracy of accurate tracking technique, superiority of FM over the vector model, high precision of session classification when PPED is applied, tolerable accuracy of dynamic clustering as compared to K-Means, scalability of the dynamic clustering algorithm and capabilities of the FM model in capturing meaningful clusters in real data. In [4], the author presented frequent pattern mining in log data which is a heavily researched area in the field of data mining with wide range of applications. The aim of discovering frequent patterns in Web log data is to obtain information about the navigational behavior of the users. This can be used for advertising purposes, for creating dynamic user profiles etc. In this paper; there are three pattern used page sets, page sequences, page graphs. The Approach used In this paper, Web usage mining, from the data mining aspect, is the task of applying data mining techniques to discover usage patterns from Web data in order to understand and better serve the needs of users navigating on the Web. As every data mining task, the process of Web usage mining also consists of three main steps: (i) preprocessing, (ii) pattern discovery and (iii) pattern analysis. In this work pattern discovery means applying the introduced frequent pattern discovery methods to the log data. The authors' deals with the problem of discovering hidden information from large amount of Web log data collected by web servers. And they show that how frequent pattern discovery tasks can be applied on the web log data in order to obtain useful information about the user's navigation behavior. In [5], the authors present a survey of the recent developments in Web usage mining area that is receiving increasing attention from the Data Mining community. The Approach used in this paper, Web Usage Mining exploits consolidated statistical analysis techniques. In contrast, research in this area is mainly focused on the development of knowledge discovery techniques specifically designed for the analysis of web usage data. Most of this research effort focuses on three main paradigms: association rules, sequential patterns, and clustering. The authors present many commercial tools which perform analysis on log data collected from web servers. Most of these tools are based on statistical analysis techniques, With respect to Web Mining commercial tools, it is worth noting that since the review made in the number of existing products almost doubled. In most cases, Web Usage Mining tools are part of integrated Customer Relation Management (CRM) solutions for e-commerce. In [6], the authors presented in this paper, they determine which pages of the web site were accessed and how various web pages were reached, requires examining the raw data recorded in the log files created by the web server. And in these paper researchers discusses the first phase of web usage mining in order to identify the web usage patterns of the Computer Science & Information Systems web site at the University of Port Elizabeth. The Approach used in this paper is to discuss the development of a system to analyses the web usage of the CS&IS web site in order to identify firstly, any usage patterns that may exist and secondly, any potential problems with the information architecture. A web usage pattern or navigation pattern is the sequence of web pages or scripts accessed by a user during a user Session. However, only the portion of each user session relating to the CS&IS web site will be used for analysis, since access information is not publicly available from the vast majority of web servers. Web usage mining can be broken down into three main phases, namely preprocessing, pattern discovery and pattern analysis. The authors present a data warehouse is needed to store the relevant data in the log files created by the web server. This data represents the page accesses that take place on the CS&IS web site. This data warehouse will assist with the discovery of web usage patterns and determining problems with the information architecture of the web site. In [7], the authors presented in this paper, Web-based technology is often the technology of choice for distance education given the ease of use of the tools to browse the resources on the Web. Many sophisticated web-based learning environments have been developed and are in use around the world. The same technology is being used for electronic commerce and has become extremely popular. However, while there are clever tools developed to understand online customer's behaviors in order to increase sales and profit, there is very little done to automatically discover access patterns to understand learners' behavior on web based distance learning. Educators, using on-line learning environments and tools, have very little support to evaluate learners' activities. The Approach used In this paper, researchers discuss some data mining and machine learning techniques that could be used to enhance web-based learning environments for the educator to better evaluate the leaning process, as well as for the learners to help them in their learning endeavor. Web SIFT is a set of comprehensive web usage tools that is able to perform many data mining tasks and discover a variety of patterns from web logs. A versatile system, Web Log Miner, uses data warehousing technology or pattern discovery and trend summarization from web logs. The authors present the Web is an excellent tool to deliver on-line courses in the context of distance education. Web usage mining is a non-trivial process of extracting useful implicit and previously unknown patterns from the usage of the Web. Significant research is invested to discover these useful patterns to increase profitability of e-commerce sites. In addition, with the awareness of the potential advantages of integrated web usage mining and the insufficient data recorded by web servers, there is a need for more specialized logs from the application side to enrich the information already logged by the web server. In [8], the authors presented a web usage mining is the application of data mining techniques to discover usage patterns from Web data. Web usage mining consists of three phases, preprocessing, pattern discovery, and pattern analysis. This paper provides a detailed taxonomy of the work in this area, including research reports as well as commercial offerings. An up-to-date survey of the existing work is also provided. The Approach used **i**n this paper, researchers provides an up-to-date survey of Web Usage mining, including both academic and industrial research efforts, as well as commercial offerings. In Web Mining, data can be collected at the server side, client-side, proxy servers, or obtained from an organization's database. There are three main tasks for performing Web Usage Mining or Web Usage Analysis. Are preprocessing, pattern discovery and patterns analysis [14]. The authors proposed an attempted to provide an up-to-date survey of the rapidly growing area of Web Usage mining. With the growth of Web-based applications, specifically electronic commerce, there is significant interest in analyzing Web usage data to better understand Web usage, and apply the knowledge to better serve users. This article has aimed at describing such challenges, and the hope is that the research community will take up the challenge of addressing them. In [9], the authors present the Web usage mining techniques are applied to identify frequent item-sets, sequential patterns, clusters of related pages and association rules. Web usage mining can be used to support dynamic structural changes of a Web site in order to suit the active user, and to make recommendations to the active user that help him/her in further navigation through the site he/she is currently visiting. In the case of implementing Web usage mining system in the form of a proxy server, predictions about which pages are likely to be visited in near future can be made, based on the active users' behavior. The Approach used in this paper, the data from Web logs, in its raw form, is not suitable for the application of usage mining algorithms. The data needs to be cleaned and preprocessed. The overall data preparation process is following: Data Cleaning, Efficient User Identification, Session Identification and Path Completion, Transaction Identification and also presented web usage mining algorithm in this paper. After transactions are detected in the preprocessing phase, frequent item-sets are discovered using the A-priori algorithm. The authors proposed a Web usage mining differs from collaborative filtering in the fact that we are not interested in explicitly discovering user profiles but rather usage profiles when preprocessing a log file we do not concentrate on efficient identification of unique users but rather try to identify separate user sessions. These sessions are then used to form the so called Transactions. Web usage mining system in the form of a proxy server, predictions about which pages are likely to be visited in near future can be made, based on the active user's behavior. Such pages can be pre-fetched to reduce access times. In this paper researchers describe the algorithm for Web usage Mining.

III. Web Usage Mining

Analyzing the behavior of a Web Site user, also known as Web Usage Mining [11][13]. It is a research field which consists in adapting the data mining methods to access log files records. These files collect data such as the IP address of the connected machine, the requested URL, the date and other information regarding the navigation of the user. Web Usage Mining [16] techniques provide knowledge about the behavior of the users in order to extract relationships in the recorded data. Among available techniques, the sequential patterns are particularly well adapted to the log study. Web usage mining also enables Web based businesses to provide the best access routes to services or other advertisements. When a company advertises for services provided by other companies, the usage mining data allows for the most effective access paths to these portals. Usage mining is also valuable to e-businesses whose business is based solely on the traffic provided through search engines. The use of this type of web mining helps to gather important information from customers visiting the site. This enables an in-depth log to complete analysis of a company's productivity flow. E-businesses depend on this information to direct the company to the most effective Web server for promotion of their product or service. Therefore, usage mining has definite valuable utility to the marketing of businesses and a direct impact on the success of promotional strategies. Web Usage Mining analyses the usage patterns of web sites in order to get an improved understanding of the user's interests and requirements. This information is especially valuable for E-Business sites in order to achieve improved customer satisfaction.

3.1 Approaches of Web Usage Mining

Various approaches have been discussed in literature for web usage mining [5].

- Self organized map.
- Association rules.
- Apriori algorithm.
- Sequential rules.
- Genetic algorithm.

3.2 Self Organized Map

The Self-Organizing Map (SOM) [17] was developed by professor Kohonen. It is one of the most popular neural network models. It belongs to the category of competitive learning networks Based on unsupervised learning, which means that no human intervention is needed during the learning and that little need to be known about the characteristics of the input data. SOM is used for clustering the data without knowing the class. The SOM can be used to detect features inherent to the problem and thus has also been called SOFM, the Self-Organizing Feature Map, Provides a topology preserving mapping from the high dimensional space to map units. Map units, or neurons, usually form a two-dimensional lattice and thus the mapping is a mapping from high dimensional space onto a plane. The property of topology preserving means that the mapping preserves the relative distance between the points. Points that are near each other in the input space are mapped to nearby map units in the SOM. The SOM can thus serve as a cluster analyzing tool of high-dimensional data.

3.3 SOM Algorithm

1. Select output layer network topology.
   Initialize current neighborhood distance, D (0), to a positive value.
2. Initialize weights from inputs to outputs to small random values.
3. Let t=0
4. While computational bounds are not exceeded do (t<=1).
   i) Select an input sample $t_{i,k}$.
   ii) Compute the square of the Euclidean distance of $t_{i,k}$. From weight vectors (wj) associated with each output node.
   $$\sum_{k=1}^{n} \left(t_{i,k} - w_{j,k}(t) \right)^2$$
   iii) Select output node j^* that has weight vector with minimum value from step 2.
   iv) Update weights to all nodes within a topological distance given by D(t) from j^*, using the weight update rule:
   $w_j(t+1) = w_j(t) + n(t)(t_l - w_j(t))$
   v) Increment t.
5. End while.

IV. Proposed Methodology

First Euclidian distance between the user transaction and cluster centers is calculated and transaction is assigned to nearest cluster. After each assignment weights are updated. This process continues until there are no more changes in the weights. This is an unsupervised learning mechanism because no teacher is needed to train the system. The procedure for clustering the transactions is shown in Fig. 2.

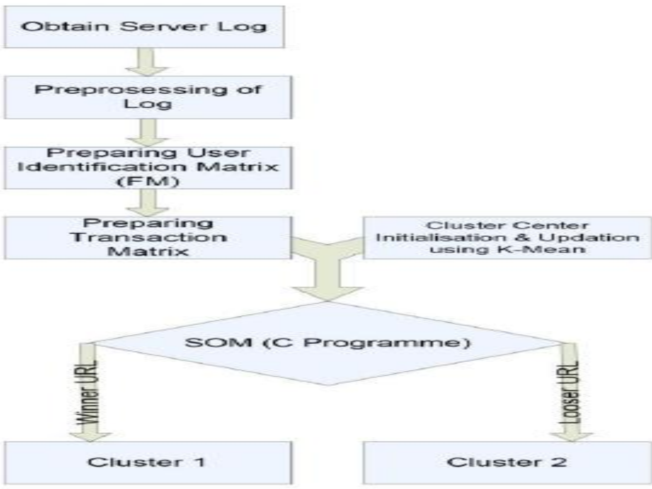

Fig. 2 Processing Prior to the execution of SOM

4.1 Data Preprocessing

Main purpose of this mining approach is to facilitate more effective and efficient navigation hence only those keep log entries need to keep which are relevant to the purpose of organizing the Web pages. Some irrelevant log entries are deleted from the log file. Sometimes a user requests a page that does not exist. This will create an error entry in the log. Here data cleaning process is applied on the web server log obtained from KIET central lab and its size is reduced from thousands of entries to hundreds of entries.

4.1.1 User Identification [19]: Identifying unique users is a complex work because different users can access web pages from the same machine. Whole session can be divided in different user sessions or transaction, where each session represents a different user.

4.1.2 User Session Identification [12]: same machine can be used to access different server Web sites by different users. Therefore, log entries are divided into different user-sessions through a session timeout. If two time slots between page requests exceed a defined limit, it is assumed that the pages are requested by two different user-sessions, no matter IP address of machine is same.

4.1.3 Transaction Identification: Transactions are the group of URL accessed by a user within a defined span of 5 minute time.

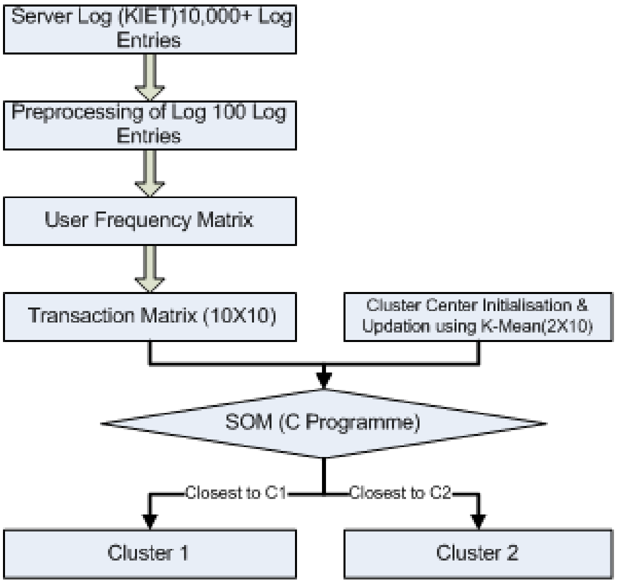

Fig. 3 Clustering of KIET Server Log

4.2 Server Log KIET (DATE-2013-04-06)

This summary is based on the server log [14] collected from KIET central Lab Server. Purpose of this summary is to select that machine which made maximum number of URL request. Here Machine IP 192.168.30.121 shown in the Table 1 have maximum no of URL request so it is selected for further analysis.

Table 1 Access Frequency Table

Machine IP Addre	Access Frequency
10.21.40.40	33
10.21.40.224	1
192.168.30.44	22
192.168.30.100	22
192.168.30.13	2
192.168.40.194	18
192.168.30.121	97
10.21.40.165	1
192.168.30.57	14
192.168.40.15	6
10.21.60.128	9
192.168.30.16	7
10.21.40.49	11
10.21.40.32	7
192.168.30.128	24

4.2.1 Access log by machine 192.168.30.12

Machine IP 192.168.30.121 made a maximum of 97 URLs requests as shown in above Table 1. These 97 URLs are categorized in transactions based on a specific time span as shown in Table 2.

Table 2 Access log of a machine

Transaction1	2013-04-06 10:58:58 192.168.30.121(50862)->206.190.36.45(80) in.yahoo.com/jsal 2013-04-06 10:58:57 192.168.30.121(50860)->206.190.36.45(80) in.yahoo.com/p.gif; 2013-04-06 10:58:57 192.168.30.121(50851)->216.115.100.103(80) l1.yimg.com/dh/ap/default/130326/yuvi-392.jpg 2013-04-06 10:58:55 192.168.30.121(50860)->206.190.36.45(80) in.yahoo.com/jsal 2013-04-06 10:58:55 192.168.30.121(50855)->206.190.36.45(80) in.yahoo.com/p.gif;_ylc=x3odmtvlbzjsymfsbf9tazk3njg0mtqybgedtfnfrmvuzybtahvpih 2013-04-06 10:58:55 192.168.30.121(50851)->216.115.100.103(80) l1.yimg.com/nn/fp/rsz/040413/images/smush/bedfp_1365057363.jpg
Transaction2	2013-04-06 10:58d:50 192.168.30.121(50855)->206.190.36.45(80) in.yahoo.com/jsal 2013-04-06 10:58:50 192.168.30.121(50853)->206.190.36.45(80) in.yahoo.com/p.gif;_ylc=x3odmtvjazhvdnq5bf9tazk3njg0mtqybgedrklox0hvdybtdwnoig 2013-04-06 10:58d:49 192.168.30.121(50855)->206.190.36.45(80) in.yahoo.com/jsal/ygang
. . .	. . .
Transaction 10	2013-04-06 10:58:19 192.168.30.121(50766)->206.190.36.45(80) in.yahoo.com/p.gif;_ylc=x3odmtu3chfzbjaybf9tazk3njg0mtqybgedq1jjx1dhdwdoihr3aw 2013-04-06 10:58:19 192.168.30.121(50764)->216.115.100.103(80) l1.yimg.com/nn/fp/rsz/040413/images/smush/waughs_1365077898.jpg 2013-04-06 10:58:19 192.168.30.121(50767)->206.190.36.45(80) in.yahoo.com/jsal 2013-04-06 10:58:18 192.168.30.121(50761)->67.195.141.200(80) in.omg.yahoo.com/photos/the-ugly-side-style-stunted-stars-slideshow/

4.2.2 Unique URL Frequency Matrix

Following Table 3 assign unique IDs to URLs accessed by machine 192.168.30.121 by different transactions:

Table 3 URL Frequency Table

ID	URL	FRQ
URL1	206.190.36.45(80)	22
URL2	216.115.100.103(80)	52
URL3	67.195.141.200(80)	7
URL4	98.138.47.199(80)	2
URL5	98.138.4.126(80)	2
URL6	66.94.245.1(80)	2
URL7	125.56.200.41(80)	2
URL8	69.192.223.139(80)	1
URL9	216.115.100.102(80)	7
URL10	192.221.76.125(80)	1

4.2.3 Transaction Matrix ($t_{i,k}$)

Transaction Matrix given in Table 4 shows URLs accessed in a particular transaction. Transaction T1 had made request to URL1 and URL2 so ‘1’ is placed in the corresponding URLs and placed ‘0’ in the rest of URLs. This process is repeated for all Transactions from T1 to T10 as shown in Table 4.

Table 4 Transaction Matrix

Transaction	Url1	Url2	Url3	Url4	Url5	Url6	Url7	Url8	Url9	Url10
T1	1	1	0	0	0	0	0	0	0	0
T2	1	0	0	0	0	0	0	0	0	0
T3	1	1	1	0	0	0	0	0	0	0
T4	1	1	1	1	1	1	0	0	0	0
T5	1	1	1	0	0	0	1	0	0	0
T6	0	1	1	0	0	0	0	1	0	0
T7	0	1	0	1	1	1	0	0	0	0
T8	1	1	0	0	0	0	0	0	1	1
T9	0	1	0	0	0	0	0	0	1	0
T10	1	1	1	0	0	0	0	0	0	0

4.2.4 Weight Matrix ($w_{j,k}$)

Although any number of cluster centers can be initialized according to the clustering need but here Two Custer centers are taken and randomly initialized with values between 0 and 1. After assignment of each transaction to a cluster center, weights of that particular cluster center are updated using classical K-Mean algorithm [15] which uses the following weight updating formula:

$$\mathbf{w_j(t+1) = w_j(t) + n(t)(i_l - w_j(t))}$$

After every update of weight, some changes occur in cluster center weight values. Update processes terminated when there is no change in weight values in consecutive 2 or 3 update and these weight values are assumed as final weight values. Following weight matrix is taken with random values between 0 and 1 and assumed that no further change occurs after update. Final weights values are shown in Table 5.

Table 5 Weight Matrix

Cluster Center 1	0.5	0.4	0.3	0.2	0.1	0.6	0.8	0.9	0.0	0.1
Cluster Center 2	0.4	0.5	0.3	0.3	0.6	0.1	0.1	0.8	0.8	0.1

4.2.5 Euclidian Distance Calculation

Distance of each cluster is calculated from both of cluster centers by using following formula-

$$\sum_{k=1}^{n} \left(\mathbf{t_{l,k} - w_{j,k}(t)} \right)^2$$

After calculation of distance, a transaction is assigned to cluster center closest to it. Assume transaction T1, Which accessed the following URLs as shown in row 1 of above Table 4

Transaction T1 - 1, 1, 0, 0, 0, 0, 0, 0, 0, 0,

Now distance of T1 is calculated from both of cluster centers as follows-

Distance from cluster Center 1-

$(1-0.5)^2 + (1-0.4)^2 + (0-0.3)^2 + (0-0.2)^2 + (0-0.1)^2 + (0-0.6)^2 + (0-0.8)^2 + (0-0.9)^2 + (0-0.0)^2 + (0-0.1)^2 = 2.57$

Distance from cluster Center 2-

$(1-0.4)^2 + (1-0.5)^2 + (0-0.3)^2 + (0-0.3)^2 + (0-0.6)^2 + (0-0.1)^2 + (0-0.1)^2 + (0-0.8)^2 + (0-0.8)^2 + (0-0.1)^2 = 2.46$

Above result shows that distance of T1from cluster center2 is less than from cluster center 1. Hence T1 is closest to cluster center 2 therefore assigned to cluster 2. Same process in done for all the transaction and result is shown in the following table 6.

Table 6 Transaction assignment to Clusters

Transaction	Distance from Cluster Center 1	Distance from Cluster Center 2	Winning Cluster
T1	2.57	2.46	Cluster2
T2	2.37	4.60	Cluster1
T3	2.97	2.86	Cluster2
T4	4.17	3.86	Cluster2
T5	2.37	3.66	Cluster1
T6	2.17	2.06	Cluster2
T7	3.77	3.26	Cluster2
T8	4.37	2.66	Cluster2
T9	2.57	2.46	Cluster2
T10	2.57	2.46	Cluster2

The above results clearly show that if user is navigating in a particular web page which is belonging to some cluster, then the probability of navigation by that particular user is other web pages belonging to the same cluster.

V. Conclusion And Future Work

Above result generated by the SOM network shows that our approach can effectively discover usage pattern. User's browsing behavior can also predict based on the past history. Here transaction-5 & transaction-2 belongs to cluster-1. And rest of the transactions belongs to cluster-2. So navigation of a web user from any URL related to transaction 2 is more likely to the URL related to transaction 5, (and vice – versa) than URL related to any other transactions. And same is in the case of cluster 2. For huge server logs automatic software can be used. With these software data cleaning becomes a challenging task which is needed to be separately handled. Varying time slot can be used for identification of users which may provide better results.

References

[1] F. Yongjian and M.Yi Shih , A Framework for Personal Web Usage Mining, Department of Computer Science University of Missouri-Rolla, 2005.

[2] M. Albanse, A. Antonio, Carlo and L. Sansone “Web Personalization System based on Web Usage Mining Techniques” Department of the Information Sistemistica University di Napoli Federico II via Claudio, 2180125 Napoli, Italy 2007.

[3] C. Shahabi , F. Banaei-Kashani , Efficient and Anonymous Web-Usage Mining for Web Personalization Department of Computer Science, Integrated Media Systems Center, University of Southern California, Los Angeles ,2004.

[4] R. Iváncsy , I. Vajk , Frequent Pattern Mining in Web Log Data Department of Automation and Applied Informatics, and HAS-BUTE Control Research Group Budapest University of Technology and Economics , 2006.

[5] F. M. Facca and P. L. Lanzi, Recent Developments in Web Usage Mining Research, Artificial Intelligence and Robotics Laboratory Department of the Electronics and Information Politecnico di Milano.

[6] P. Craig Oosthuizen , J. Wesson & C. Cilliers , Processing Web Logs in order to Mine Web Usage Patterns , Department of Computer Science and Information Systems University of Port Elizabeth , 2005.

[7] R.O. Zaýane , Web Usage Mining for a Better Web-Based Learning Environment, Department of Computing Science University of Alberta Edmonton, Alberta, Canada.

[8] J. Srivastava_ y , R. Cooleyz , M. Deshpande, Pang-N. Tan , Web Usage Mining on Discovery and Application of Usage Pattern from Web Data Department of Computer Science and Engineering University of Minnesota, 2000.

[9] M. Grèar, User Profiling: Web Usage Mining Department of Knowledge Technologies Jozef Stefan Institute Jamova 39, 1000 Ljubljana, Slovenia.

[10] A. Sharma,Web Usage Mining Using Neural Network Associate Prof (Comp.Sc), Peoples University, Bhopal 2012.

[11] P. S Raghavendra, Shreya Roy Chowdhury, Srilekha Vedula Kameswari, Web Usage Mining using Statistical Classifiers and Fuzzy Artificial Neural Networks, Department of Information Technology, National Institute of Technology Karnataka, 2011.

[12] R. Baeza-Yates and C. Castillo. Relating web structure and user search behavior, In 10th World Wide Web Conference, Hong Kong, China, May 2001.

[13] Vaishali A.Zilpe, Dr. Mohd Atique, Neural Network Approach for Web Usage Mining, Government College of Engineering, Amravati Maharashtra, India 2011.

[14] Kerkhofs J., Vanhoof K., Pannemans D., Web Usage Mining on Proxy Servers: A Case Study, Limburg University Centre, 2001.

[15] D. Qi and C. Chih Li, Self-Organizing Map based Web Pages Clustering using Web Logs, School of Information Technology, Illinois State University, USA 2005.

[16] S. Bai ,Q. Han ,Q Liu ,X. Gao “Research of an Algorithm Based on Web Usage Mining” China, IEEE, 2009.

[17] P. Britos, D. Martinelli, H.. Merlino, R. García-Martínez, Web Usage Mining Using Self Organized Map, PhD Computer Science Program, National University of La Plata. Software & Knowledge Engineering Center, Buenos Aires Institute of Technology, Intelligent Systems Laboratory, University of Buenos Aires , Argentina, 2007.

[18] R. Cooley, B. Mobasher, J. Srivastava, Web Mining: Information and Pattern Discovery on the World Wide Web, Proceedings of the ninth IEEE international conference on Tools with AI, 1997.

[19] A. Kate Smith and Alan Ng , Web page clustering using a self-organizing map of user navigation patterns, School of Business Systems, Monash University, Australia 2003.

[20] J. Srivastava, WEBKDD –Web Mining for Usage Patterns & Profiles, University of Minnesota 4-192 EECS Bldg, 200 Union Street , 2002.

[21] Mrs Geeta , Prof. S. kumar , Dr. P. Reddy , Amalgamation of Web Usage Mining and Web Structure Mining , GMR Institute of Technology, Departments of IT & CSE , JNT University, Andhra Pradesh, India 2009.

[22] F. Johnson and S. K. Gupta, Web Content Mining Techniques: A Survey, International journal of computer applications (0975-888), Volume 47– No.11, June 2012.

