I J C T A, 9(41) 2016, pp. 799-803 ISSN: 0974-5572

E-COMMERCE WEBSITE DESIGN IMPROVEMENT BASED ON WEB USAGE MINING

Divya Lal[1], Adiba Abidin[1], Vikas Deep[2], Naveen Garg[2] and Dr. Deepti Mehrotra[3]

Abstract:Designing an effective website is challenging task, but its need is increased with more usage of ecommerce trading. To design an effective website the user's usage pattern can be considered as an input and Web usage mining for this objective. The web usage mining use the web log data. The Web Data contains the huge amount of information which needs to be preprocessed for extracting the relevant information. A large number of attributes are present in the web log and dimension reduction approach can help to identify most critical parameter that may affect to purchase of the product. In this paper web log of the ecommerce website is analyzed and recommendations are given that may improve the design of the website.

Key Words: *Web mining, server logs, Rapid Miner, URL;*

I. INTRODUCTION

Data mining is the extraction of information from expansive measure of observational data sets, to find unsuspected relationship and example covered up in data outline the data in novel approaches to make it justifiable and valuable to the data clients.On of wide usage of data mining approach is done in field of mining web dataset referred as web mining. Web mining is the utilization of data mining techniques to extract the useful information from the web pages. Web mining can be broadly classified into three classes: a) web usage mining which is utilization of data mining strategies to find fascinating usage designs from Web information keeping in mind the end goal to comprehend and better serve the requirements of Web-based applications. Usage data captures the identity or origin of Web users along their searching conduct at a Web website. Web usage mining can be further classified based on usage of data i.e. Web Server Data, Application Server Data, Application Level Data [2]. b) Web structure mining is the way towards utilizing graph theory to analyze the structure of website. The structure contains nodes and its connections. It identifies the relationship using different hyperlinks. Hyperlinks are the integral part of the web that connects different web page by information. Web structure mining extracts different patterns from these hyperlinks. It contains web graph, node, edge, in degree, out degree.[2]c)web content mining which use text mining approach to mine the web content mining. It is the extraction, mining and integration of web information. Web content mining focuses on knowledge discovery, in which the primary articles are the traditional collections of interactive media reports, for example, pictures, video, and sound, which are inserted in or connected to the website pages. It is

[1] *Student, Department of Information Technology, Amity University Uttar Pradesh, Noida, India,*
Email: lal.divya92@gmail.com, adibaabidin@gmail.com
2. Asst Professor, Department of Information Technology, Amity University Uttar Pradesh, Noida, India,
Email: vikasdeep8@gmail.com, er.gargnaveen@gmail.com
3.Professor, Department of Information Technology, Amity University Uttar Pradesh, Noida, India
Email: dmehrotra@amity.edu

likewise not same as Data mining since Web information are for the most part semi-organized and/or unstructured, while Data mining bargains essentially with organized information.

The web usage mining considers web activities as input. Web activity contains everything from server logs to web browser activity. Log File contains all the information of the customers who visited the website. It contains information like IP address, request, response; bytes sent, browser, date and time, referrer. Data is fetched from the log file and stored in the database which further will be analyzed. Authors have used the this technique to classify whether the customer will procure the given product or not for a e-commerce website Along with this input it can also we used to provide information that can help to redesign the website to make it more effective.

There are various tools available in the market for mining data such as Rapid miner, Weka, R-Programming, Orange, KNIME, NLTK and many more. In this work Rapid Miner tool for analyzing the data [10]. It is utilized for business applications, exploration, education, prototyping, and application advancement and supports all steps of data mining i.e. data preprocessing, data cleaning, data visualization and data optimization. In this paper analysis of weblog data is done with the objective to increase the effectiveness of website. While redesigning the website administrator should identify which parameter has highest information gain. Decision tree and rule inductionapproach is used for achieve the desired objective.

II. LITERATURE SURVEY

As we know evaluating the effectiveness of e-commerce Web site design is an important, yet complex problem for e-commerce retailers. Their success hinges largely on their ability to provide a high quality Web site. So e-commerce retailers need to constantly monitor the effectiveness of their web-based storefronts to identify those Web site areas that are problematic. So, in regard with this issue we reviewed and compared several works, drawn from various areas of research.

There are many usability problems related to e-commerce websites and there are factors such as navigability, content, design, ease of use which plays important role so Sandeep Kumar Panda, Santosh Kumar Swain, Rajib Mall [4] suggested the sequence of their importance in usability i.e navigation, content, design were highest priority for evaluating the usability of e-commerce websites whereas ease of use and structure were the less important as compared to the above mentioned from the overall usability value calculation. The tool used by in the mentioned paper is Camtasia which is an automated open source tool.

Yanlong Zhang, Hong Zhu, Sue Greenwood [5], highlighted a system that investigate for measurement of website navigation using matrix approach. A large number of matrices were suggested on the basis of structural complexity of website measurement. The matrices were validated using Weyuker's software. The tool used by the researchers consists of three parts. First one is site download agent which downloads the complete website and the aim is to generate the html site map for the url. Second part is filter which discards multimedia including animation, audio but keeping the text file contributing the structural complexity. Third part is matrix calculation and report agent. This paper concluded that the structural complexity has a vital role for web navigation.

Praphul Chandra, GeethaManjunath[1], portrays a way for evaluating navigation complexity quantitatively for user to interact on the web. The method uses the graph and another theoretical complexity explanation. The approach was completely structured using bits that can be used for both traditional browsing and emerging web interconnection.The researcher work was applied on an number of web interaction where it was assume that the staring web page is known to achieve the goal. The proposed system helps in reducing the navigational complexity using widget. The authors concluded a way to reduce complexity for web navigation.

Min Chen, Young U. Ryu[6], displayed a mathematical programming system in order to increase efficiency for website navigation for a user while decreasing other transformation to the current system. The test was taken for the given data where two metrics were calculated which was used to improve the proficiency of the website. The main aim of the researchers is to improve the navigation efficiency for a website with minimum changes. The problem was taken as a special graph optimization problem using directed graphs with nodes that represent pages and arcs that represent links. The result showed the high improvement in the efficiency and it seemed disoriented users have more advantages over the system

III. PROPOSED SYSTEM

This paper proposesmining the Web Log data of an ecommerce website. Weblog contain a large amount of information in raw form. Very transaction is stamped in weblog and corresponding to each step a series of entries can be observed in the web log. Thus to fetch the relevant information, preprocessing of weblog is most important step it involves the both removing the not required data and converting the data in the desired form. The processing of log file includes few steps as shown in figure 1.

3.1 Data Collection:

There are two types of log; client web log and server web log. Client web log is responsible for analyzing which pages client has visited and can be analyzed by the cache. Whereas a server web log automatically created and maintained by a server consisting of a list of activities it performed [9]. The data of web server log fileof the E-commerce website is taken.It includes various attributes such as IP address, Browser, Date and Time, URL, Request, Response, Byte sent, Page requested, user agent. IP address of the client is from where the request has been initiated. Browser shows from which browser the client has been logged on. Date and Time signifies the specific duration of the client accessing the website. URL includes the complete path of the client browsing the website. Request is the type of web service used by the client i.e. GET, POST. Response is the web status code i.e. HTTP code send by the server to the client. Byte sent is the number of bytes used by the client. Page requested is the page that is request by the client. User agent informs the website about the client's browser and operating system.

3.2 Data Preprocessing:

The steps in the data preprocessing is the data cleaning, attribute selection and data processing. Cleaning involves removing undesirable data and incomplete data which can't be used in analyzing process, validating values corresponding to its range we require, irrelevant data is removed from the log file. The log file now contains the attributes required for given work. Attribute selection involves picking up the attributes which are of use during the analysis. The attribute selected for the analysis for the given paper are IP address, referrer and browser. The last step of data preprocessing involves data processing wherein some attributes were calculated from the given one including depth of page on website, cart and time of stay.

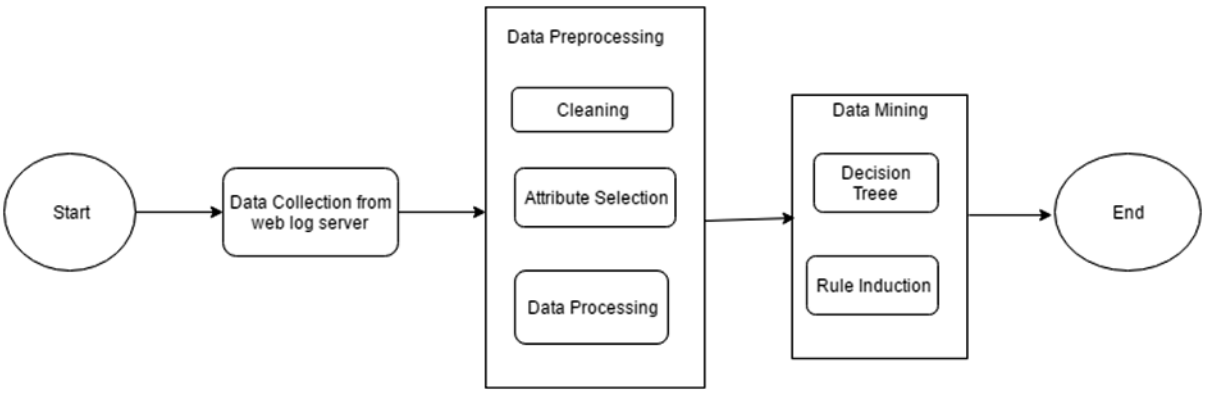

Figure1. Steps describing the data flow

3.3 Data mining:

Data mining is the final step of processing of the data. Using Rapid Miner, decision tree was created for the data we have processed.Decision tree helps in determining what all parameters are important for making the website efficient in terms of customer navigation and product cart. Along with this rule induction, an operator was also created which learns pruned sets of rules with respect to the information gain from the given data set.

IV. EXPERIMENT AND RESULT

On using Rapid Miner, Decision tree was created using the attributes IP address, Referrer, Depth, Cart, Time of Stay, and Browser. For the given data, prediction was made and Decision tree created wasto analyze which factors playimportant role to determine client has reached to cart or not.

In the Figure 2 by applying decision tree analysis it has been found that the most importantparameter i.e. the parameter on which basis the decision will make is “depth” which means how many pages the visitor underwent of a website. If the depth is greater than 3.5 then not a single visitor go to the cart, if it is less than 3.5 then again we considered depth, if depth is greater than 1.5 then the next important parameter is browser. If depth is greater than 1.5 and browser is Mozilla then the visitor goes to the cart but if the browser is opera then 75% of visitors didn’t go to the cart but 25% did. If the depth is less than or equal to 1.5 then 75% of the visitors didn’t go to the cart but 25% did.

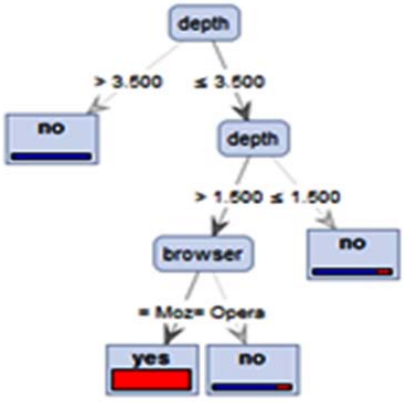

Figure2. Decision Tree

In the Figure 3 by applying Rule Induction it has been found that if the time of stay is greater than 6.5 then out of 73 outcomes none of them is greater than 6.5. If browser is opera then out of 11 outcomes only 1 didn’t go to the cart else if browser is not opera out of 9 outcomes 2 visited the cart

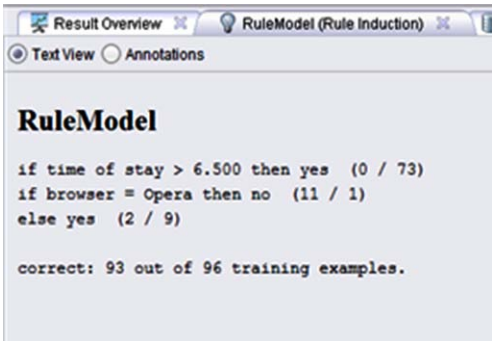

Figure3. Rule Induction

V. CONCLUSION AND FUTURE WORK

The data mining technique was used for analysis of weblog of an e-commerce website. Pre-processing the web log data is a significant and prerequisite phase in Web mining. It removes irrelevant items and identifies users and sessions along with the browsing information. From the above analysis, we can conclude that the attributes which affect most the web navigation pattern are the Time of stay, Cart, Depth and Browser. It assists the web site designers to improve the performance of the web by giving preference to the patterns navigated by the regular interested users. In future we can work on making the website more efficient and attractive by analyzing the web log to bring the frequently used products and the products which are of interest to the customer on top level hierarchy of the website in order to increase the purchase of the product.

REFERENCES

[1] Praphul Chandra and Geetha Manjunath.” Navigational Complexity in Web Interactions.” in WWW 2010, April 26–30, 2010, Raleigh, North Carolina, USA, 2010, pp. 1075-1076

[2] “What is Web Mining?” Internet: http://www.scaleunlimited.com/about/web-mining/ [Sept 16, 2016]

[3] “Web mining.” Internet: https://en.wikipedia.org/wiki/Web_mining, Sept 2, 2016 [Sept 24, 2016]

[4] Sandeep Kumar Panda, Santosh Kumar Swain and Rajib Mall. ”An Investigation into Usability Aspects of E-Commerce Websites Using Users’ Preferences.” In ACSIJ, 2015, PP.65-73

[5] Yanlong Zhang, Hong Zhu and Sue Greenwood. “Website Complexity Metrics for Measuring Navigability.” In IEEE Proceeding, 2004.

[6] Min Chen and Young U. Ryu. “Facilitating Effective User Navigation through Website Structure Improvement.” In IEEE Transaction, 2013 pp. 571-588

[7] RapidMiner, Internet: https://en.wikipedia.org/wiki/RapidMiner, Aug 29, 2016 [Sept 24, 2016]

[8] Server Log, Internet: https://en.wikipedia.org/wiki/Server_log, Aug 1, 2016 [Sept 10, 2016]

[9] Rapidminer, Internet: https://my.rapidminer.com/nexus/account/index.html#login [Sept 10, 2016]

[10] L.K. Joshila Grace, V.Maheswari and Dhinaharan Nagamalai. “ANALYSIS OF WEB LOGS AND WEB USER IN WEB MINING” in IJNSA, 2011 pp. 99-110

[11] Navin Kumar Tyagi, A. K. Solanki and Manoj Wadhwa. “Analysis of Server Log by Web Usage Mining for Website Improvement.” In IJCSI, 2010 pp. 17- 21

[12] Sana Siddiqui and Imran Qadri. “Mining Web Log Files for Web Analytics and Usage Patterns to Improve Web Organization.” In IJARCSSE, 2014 pp. 794-802

[14] Mr. Ankit Rathi and Prof. Abhijeet Raipurkar. “Web Usage Mining - A Review.” In IJARCCE, 2016 pp. 496-498