RESEARCH ARTICLE OPEN ACCESS

Data Mining to Facilitate Effective User Navigation and Improve Structure of a Website

Shweta Mohod, Prof. Vishalgangawane
Department of Computer Technology, Veermata Jijabai Technological Institute

Abstract
Now a days for a company it is very important to have active presence on web to become successful in electronic market. This requirement is fulfilled by an interactive website of a company. Website can be used to sell goods, maintain customer relationships, promotion of goods. By looking at the customer's response towards website we can decide campaign for future products and services. So for this website should be interactive and must be used by user by the way designer wants to. So, it is very important to study navigational behavior of a customer. By analyzing the web logs which records the navigational activity of the customer we can get several sequential patterns which helps us to study whether user is pursuing site's goal or not. Sequences are analyzed with WUM (Web Utilization Miner) which gives g-sequence (Generalized Sequence) and aggregate tree as an output.
On the basis of the structure that is given by WUM, conclusions are studied and it is decided whether site structure needs improvement; The improvements suggested should be minimum but at the same time site should satisfy business goal. A mathematical programming model is used which suggest minimum improvements to site. Improvements which tends to change structure of site are avoided as it may be confusing for old site user.

I. Introduction

Internet has provided opportunity to explore large amount of information. It gives us chance to represent ourself on the web space. Web have presence of various companies and it is an international market place. Now a days, online business is gaining popularity and is providing huge amount of profit to the company. For a company to be successful in e – business should have remarkable presence on the web. This space can be used for promotion and selling of goods and also to study the behavior of the customers. For this properly designed websites are essential, as customers visualizes the whole business through the website only.[1]

Easily navigated websites can provide user with significant and relevant information within a fewer clicks. To achieve this websites should be designed from user's perspective. Website designer should understand user's point of view. Main cause of poor website design is having a little understanding of user's navigational behavior. What designer thinks of website structure is different from what users thinks about structure of a website. Web designer may just organize the pages based on their own understanding. So, it is very important to study the satisfaction of user towards website.

We could study this either by selecting appropriate user group and interviewing them directly or by studying their navigational behavior. First approach is very time consuming as users go on increasing the interviewing process becomes tedious. Second approach is to collect navigation log of the users while they browse the website. By studying the navigation log all visitors will be taken into account. It can be performed on regular basis as it just requires to collect the navigation log at different times. Users are not directly involved in the process, thisresults in saving time.It suggests the improvement points inside the site effectively. For this purpose we recommend to improve the success of site by studying the access log of users.

Previous studies focuses on finding relevant information on a page ,organize the structure of a website. Our study however focuses on improving the success of a site by studying user's navigational behavior. In this paper we focus on use of data mining with other appropriate technique and the model suggested in (ref no.)The mathematical model suggested in(ref no.) is applied on static websites which were having informative structure. We tried to develop and implement a model for dynamic website such as shopping website along with the data mining principles.[2]

Data mining principles are used to evaluate site's success and mathematical model is used to suggest improvements inside the site. Data mining principles are used in WUM(Web Utilization Miner) to evaluate parameters such as contact efficiency and conversion efficiency(ref no.).Mathematical model suggests the improvements in site with the help of parameters such as threshold specified in(ref .no).Improved website is again evaluated with contact efficiency and conversion efficiency.

In this paper, we suggest complete model to measure and improve success of a site by studying navigational behavior of users.Section2 contains related work. Section3 Model to improve Success of a site.Section 4 contains experiments and result. Section 5 contains the conclusion.

II. Related Work

Various attempts were made to evaluate success of a website since the web is invented. The mining of web server log can be used in two ways one is web personalization approach and other is transformation approach. Web personalization is the process of "tailoring" webpages to the needs of specific users using the information of the users' navigational behavior and profiledata [5]. Perkowitz and Etzioni [6] describe an approachthat automatically synthesizes index pages which containlinks to pages pertaining to particular topics based on theco-occurrence frequency of pages in user traversals, tofacilitate user navigation.

In [7], Eighmey presents an experimental setting for measuring the quality of commercial websites. Quality is modelled as a set of factors like the information utility of the presented contents, ease ofuse and attractiveness of the presentation metaphor. These factors do not constitutemeasures but are appropriatefor ranking. They were obtained from a sample of users that filled a questionnaire. Questionnairesare also used by Alpar et al. to evaluate the satisfaction of users with a number of web sites [8]. Intheir study, user satisfaction is again measured on the basis of various factors that can be measured (likeclicks performed or time spent in a site) or at least ranked. However, as noted in the introduction, suchmethods rely heavily on the selection of a representative sample of users and on the interaction with theseusers.

The objective of measuring the success of a site *with respect to the objective goals of its owner* is describedin [9]. Berthon et al. propose two measures of the site's success, the *contact efficiency* and the*conversion efficiency*. The first measure returns the fraction of users that spend at least a user-definedminimum amount of time exploring the site. The second measure returns the ratio of users that afterexploring the site becomes the customer of site. Hence, the success of the site is defined as its efficiency in"converting" visitors into customers and can be measured without the involvement of users.

A mathematical model is proposed in [3] which works best on website containing static pages. This model allows the web master to specify the threshold. On the basis of threshold Sessions get created. Out of created sessions the sessions whose length is greater than threshold value are chosen for improvement. Mathematical Model is applied on these sessions and the modifications are suggested inside the site.

III. Model to improve Success of a site

Success of site should be measured in terms of ease of navigational behavior. Website should be designed such that it allows user to find desired information without backtracking. Hence user should perceive sites goal within few steps. Goal of site can be the exposure of a particular page to the user.To make the site's goal explicit for the analysis of user behavior, we characterize the site's pages in termsof their function in pursuing this goal.

An "action page" is a page whose invocation indicates that the user is pursuing the site'sgoal. A "target page" is a page whose invocation indicates that the user has achieved the site's goal.

In an e-commerce site providing facility for shopping of an antique items from museum along with description of museum; pages providing description of items which are available for sell can be an action pages. An page providing facility for generation of bill can be an action page. Here we are assuming that target page could not be reached without reaching to the action page. In the framework of web-based marketing, it is anticipate that the success of a site is measured bythe percentage of its visitors that get engaged in exploring it ("contact efficiency") and in the percentage ofthe visitors that finally become customers ("conversion efficiency").

The efficiency of a site can be measured and these parameters are further used for improvement inside the site. Improvements should be done such that site structure should not be altered and user can reach to the target pages within specified path threshold length.Here we can set a goal for usernavigation for each target page, which is denoted by $b_j$andis termed the path threshold for page j.For a mini session S having target page j,we can determine whether user reach with minimum steps or not by comparing sessions length denoted by L with path threshold b_j..If it's length is less than path threshold then there is no improvement needed for that session but if it is greater than the path threshold then improvement is needed for S.

Table1 shows the mini sessions from E-Bombay Museum Website. We have considered E-Bombay Museum site for our experiment. We have collected the data from this site locally with the help of Local Area Network(LAN). In this website target page is the page where the bill is generated. But for our experiment we use the page which contains items for shopping as target page because user can not directly jump to the page where the bill is generated. User has to go to the page where items are displayed for shopping. Here we have to take the threshold more than 5;as the website is not just the mere collection of static pages but it is the collection of pages which needs to visit in particular order.

	Mini Session
1	Index.jsp,loginaction.jsp,Museumshop.jsp,others.jsp,ViewCart.jsp, GenerateBill.jsp
2	Index.jsp, loginaction.jsp,,Museumshop.jspart.jsp,ViewCart.jsp, GenerateBill.jsp
3	Index.jsp, loginaction.jsp,events.jsp,collectives.jsp

Table1

IV. Experiments and Results

First we have collected the access log of a E-Bombay Museum site and evaluate the parameters as contact efficiency and conversion efficiency. The data is gathered by making 50 users to browse the website. Dataset is collected locally with the help of LAN connectivity. Users browse at http://localhost:8080/mms. Fig.1 represents the value of these parameters in graphical format. Then we apply the model given [3].

We have taken path threshold as b_j=5.We have to take threshold greater than 5 because to reach to target page after visiting four pages. As target page is page which generate bill. Hence to generate bill user have to traverse the pages which provides the functionality for shopping. These will make user to visit minimum four pages.

Our model suggested the sessions needs to be improved and additional links which needs to be introduced in the site to improve effective user navigation. The required changes are made in the site and again access log is collected and the parameters contact efficiency and conversion efficiency are evaluated.

Fig.1 shows the increase in contact efficiency of all the action pages.Fig.2 shows comparative increase in relative contact efficiency.Fig.3 shows comparative increase in conversion efficiency

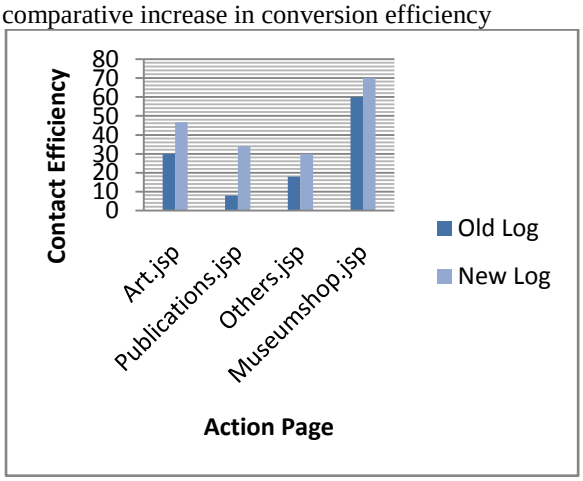

Fig.1Comparision of contact Efficiency of old log and new customer log

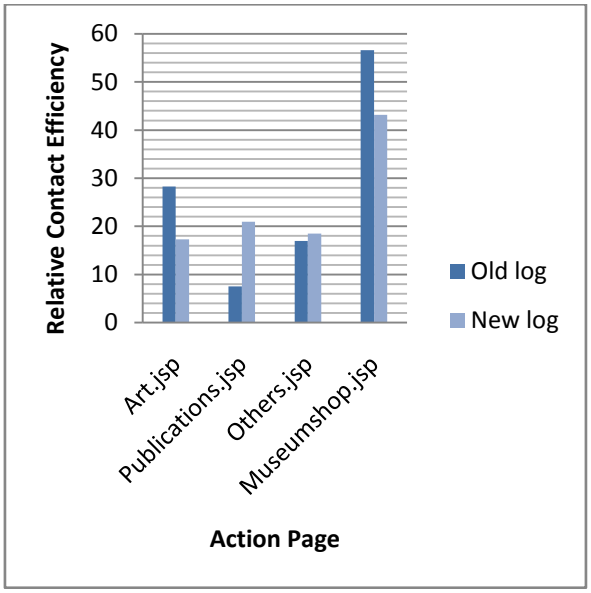

Fig.2 Comparison of Relative Contact Efficiency of old log and new customer log

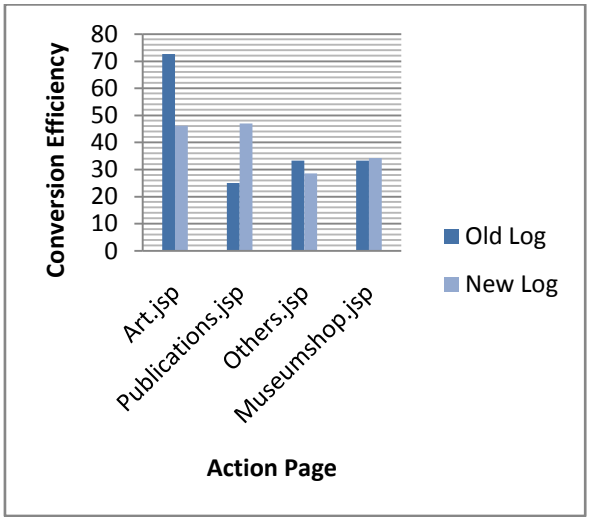

Fig.3Comparison of Conversion Efficiency of old log and new customer log

V. Conclusion and Future Work

We used contact efficiency and conversion efficiency for measuring the success of a site. We calculate these parameters from access log collected. Then apply the mathematical model on the same access log to suggest the modification inside the site. The modifications were in terms of the additional links needs to be introduced in site in order to ease user's navigational behavior while browsing the site.

When site was again evaluated it is found that contact efficiency and conversion efficiency of most of the pages were increased. For one page it decreased but it was above the threshold value. The mathematical model was originally designed for website containing static web pages but we tried to implement it on site containing static pages as well as

pages providing shopping facility and we got quite promising results.

As this model and its theory were completely based on transformation approach we can extend this to the personalization approach along with the other data mining techniques as like association rule where we can find out the results as user which were interested in movies were also interested in sports. We can try to apply this method on the websites whose contents are not relatively stable over time.

References

[1] Chintan R. Varnagar, Nirali N. Madhak, Trupti M. Kodinariya, Jayesh N. Rathod, *Web Usage Mining: A Review on Process, Methods and Techniques*,

[2] Myra Spiliopoulou, CarstenPohle *Data Mining to Measure and Improve the Success of web*,arXiv:cs.LG/0008009 v1 15 Aug 2000

[3] Min Chen and Young U. Ryu, *Facilitating Effective User Navigation through Website Structure Improvement*, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 3, MARCH 2013

[4] Myra Spiliopoulou, Lukas C. Faulstich, *WUM: A Web Utilization Miner*

[5] M. Eirinaki and M. Vazirgiannis, "*Web Mining for WebPersonalization*," ACM Trans. Internet Technology, vol. 3, no. 1,pp. 1-27, 2003

[6] M. Perkowitz and O. Etzioni, "*Towards Adaptive Web Sites: Conceptual Framework and Case Study*," Artificial Intelligence, vol. 118, pp. 245-275, 2000.

[7] John Eighmey. *Profiling user responses tocommercial web sites*. Journal of Advertising Research, 37(2):59–66,May-June 1997.

[8] Paul Alpar. *Satisfaction with a web site. In August-Wilhelm Scheer and Markus Nüttgens*, editors,*4. Internationale Tagung Wirtschafts informatik 1999*. PhysicaVerlag, Heidelberg, 1999.

[9] Pierre Berthon, Leyland F. Pitt, and Richard T.Watson. *The world wide web as an advertisingmedium*. *Journal of Advertising Research*, 36(1):43–54, 1996.