IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 17, Issue 1, Ver. V (Jan – Feb. 2015), PP 43-47
www.iosrjournals.org

Mathematical Programming Approach to Improve Website Structure for Effective User Navigation

Ms .Kshitija Potdar[1], Prof .RohiniBhosale[2]
[1](*Department of Computer Engineering, Pillai HOC college of Engineering & Technology, Rasayani / Mumbai University, India)*
[2](*Department of Computer Engineering, Pillai HOC college of Engineering & Technology, Rasayani / Mumbai University, India)*

Abstract:*Due to tremendous growth of web applications. It increases the complexity of web applications and web navigation. Designing well-structured website has been long challenge because while creating website web developers understanding of a how website should be structured can be different from that of user's preferences. Such differences result in cases users having difficulty to locate desired information on website. While various methods have been proposed to reorganize Webpages to improve users navigation using users navigation data but it still causes disorienting users and increases user's cognitive load. There is a need to improve the efficiency and the performance of a website for effective user navigation. Recommendations play an important role towards this direction. Our Recommendation is based on user Browsing patterns. Our approach presents a comprehensive overview of web mining methods and techniques used for the evaluation of reconciling systems to achieve better web navigation in order to improve the efficiency of web site. In this Paper we will propose Mathematical programming model to improve user navigation while minimizing alteration to current structure. Using our model it will possible to obtain optimal solution very quickly.*

Keywords: *clustering,mathematical programming model, mini session, Path threshold, user navigation,web mining*

I. Introduction

Now a day's internet has being used increasingly in everywhere As WWW grows tremendously it increases complexity of web application and web navigation. In order to satisfy needs of online customer business firms are investing in development and maintenance of their websites. Still it is find that finding desired information on website is not easy [1].The most of the people browsing the internet for retrieving information. But most of the time, they gets lots of insignificant and irrelevant document even after navigating several links. Website design has been examined from the lens of human-computer interaction which focuses on issues surrounding usability and interface design. Difficulty in navigation is reported as the problem that users leave a website even if its information of high quality and switch to competitor [2] [3].

Primary cause of poor website design is that web developers designing websites according to their own judgments so it is considerably different from the users. Thus, webpages should be organized in such way that it matches to users preferences .so designing well-structured website has long been challenge to facilitate effective user navigation. We will work on improving website navigation through use of user navigation data. The growth of internet has led to various studies on facilitating effective user navigation with knowledge mined from web server logs is classified into two categories to facilitate particular user by tailoring Webpages based on users profile and traversal path often referred as web personalization and changing site structure to ease the navigation for large set of users is referred as web transformation. While various methods have been proposed to improve user navigation through web site structure improvement by reorganization of website structure but there are drawbacks with website structure reorganization. Reorganized website structure is unpredictable because it changes organizational logic of websites which is designed by experts and costs of disorienting users remain unanalyzed. To overcome demerits of website reorganization we will propose mathematical programming model that facilitates effective user navigation on website with minimum changes to its current structure. Our model is appropriate for informational websites.

II. Related Work

Previous studies on website has focused on web mining approach for [4] improving user navigation as mining informative patterns, finding relevant pages of given pages and finding structure of informative website from web serverlogs. Web mining is classified into three approaches 1) web content mining 2) web structure mining 3) web usage mining. Web content mining is concerned with the retrieval of information from web into more structured forms and indexing the information to retrieve it quickly. Web usage mining is the process of identifying the browsing patterns by analyzing the user's navigational path. Web structure miningconsist of a

typical web graph consists of web pages as nodes, and hyperlinks as edges connecting related pages. It is the process of discovering structure information from the website to discover the model underlying the link structures of the Web pages, catalog them and generate information such as the similarity and relationship between them, taking advantage of their hyperlink topology.

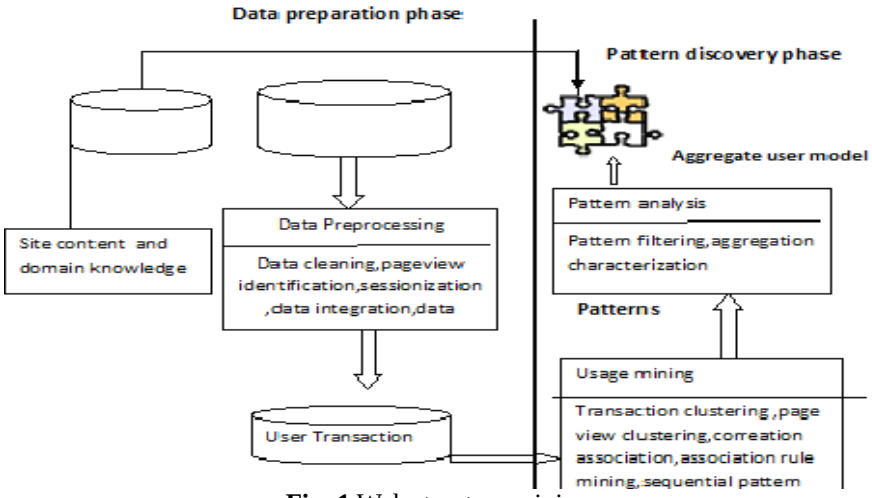

Fig. 1 Web structure mining

Our studies to Discover user navigation behavior is based on Web structure mining. Some algorithm have been proposed for web log mining such as page rank algorithm, HITS etc.As The Website is a rich source of the dynamic information, which is useful in various disciplines. There has also been much research related to improving the quality of information finding in the Websites. However, most of the work is still inadequate to satisfy demand from users [5]. Exploit the hyperlinks in the Web and propose a new approach called SFP in order to improve the quality of user navigation and research results obtain from search engines. The SFP algorithm evolves from the frequent pattern mining technique, which is a based on common data mining technique for conventional databases. The essential idea of this approach is to mine the frequent patterns of links from a given Web topology. By using the SFP algorithm, it is possible extract the authoritative pages and communities from the complex Web topology. Proposedapproach running on several experiments and show that the performance and functionalities of using the SFP in managing search results and user navigation are improved than other known methods such as HITS. Knowledge which is mined from webserver logs is classified into web personalization and web transformation. Web personalization is process of reconstituting web pages to the demand of particular users using information of the his profile and navigation behavior .the methods proposed by bamshadmobasher and honghudai [6], [7], [8] to evaluate cluster of users profile and page view from web server logs and generate links for users who are classified into different categories according to their access patterns through clustering methods. Bomshadmobasher[9] present framework for hybrid web personalization model that dynamically switch among different recommendation system based on degree of connectivity and current position of user within website.Perkowitz and Etzioni[10][11]proposed page gathering algorithm for discovering index pages which content links to particular topic based on users access logs to ease the user navigation.

On other hand web transformation approach involves changing structure of website for large number of users Instead of personalizing webpage for individual user.Gupta [12] proposed method of reorganizing website based on simulated annealing to relink Webpages for improving user navigation effectively. But this method doesn't yield optimal solution and take long computation time to run even small website. Further chang-chun-lin [13] develop integer programming model for reorganizing website which is based on cohesion among webpages for complex website structure. Proposed method reduces the information overload and search depth for users but computation time can be predicted to increase rapidly accordingly website scale. In addition a two stage heuristic model which contains two integers programming model is developed to reduce computation time. However this still requires very long computation time to get optimal result when website contains many links. To resolve this problem Tseng [14] propose ant colony method to reorganize website structure. Using this approach it is possible to find optimal solution relatively in short computation time. But it is scalable for small sized website and posing questions on large scale websites. For Website reorganizing approaches their drawbacks are obvious. First, the reorganized website structure is unpredictable and cost of disorienting users remains same. Second, complete reorganized website could change location of familiar items on website and

may disorient website users. This is because website structure is designed by expert web masters and it having organizational logic but this logic may no longer exist in new structure when website is reorganized.

III. Implemented System

After recognizing drawback of website reorganization it become challenge for webmasters to improve website structure rather than reorganizing it so user can locate target information on website in fewer clicks. Specifically The mathematical programming model provide effective user navigation on website with minimizing its current structure so it avoid disorienting users.[1] This model add minimum new links to website to reduce users cognitive load. It is appropriate for informative website whose content are static and stable over time. Examples of organization that have informational websites are universities, health care, education, sports and tourist attraction. This model allows web developers to specify goal for user navigation that improved structure should reach. This goal is deals with individual target pages and it is defined as maximum number of paths allowed to reach target pages in mini session. This goal is termed as path threshold. Thus the website structure must be altered in such way that number of paths required to locate desired page should be less than path threshold. For ease of navigation the graphical node Structure of web site is shown to users so it will help user to find desired Webpages very effectively. Proposed model is improved by graph portioning clustering algorithm. It is used to group the users with similar navigational pattern from weblog data. Clustering result include number of visits to single webpage.Fig.2 shows the process of mathematical model implementation and clustering of navigational data.

Step: 1 collecting user navigation session from weblogs by assigning user id for each user.
Step: 2 Identifying user's session logs.
Step: 3 creating clusters of users having similar navigation pattern
Step: 4 Identifying Page connectivity information.
Step: 5 Performing MPM on the mini sessions to enhance existing links.

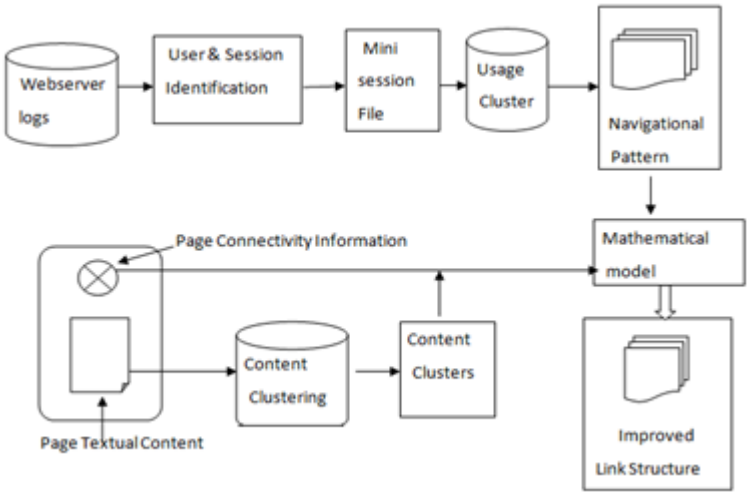

Fig.2: System architecture

IV. Mathematical Programming Model

This project model a website as a directed graph with nodes representing pages and edges representing links. Let N be set of all web pages and λ_{ij} where i, j □ N, denote page connectivity in website structure, with λ_{ij} indicating page i has a link to page j and $\lambda_{ij}=0$ otherwise. Current out-degree threshold for page i is denoted as $W_i=\sum j\epsilon N\, \lambda_{ij}$. For example suppose webpage 1 has links to page 2 and 4 then $W_1=2$. From webserver log, file of mini sessions T can be obtained. For a mini session S □ T, the target page of S is denoted as tgt(S). Let $L_m(S)$ be the length of S i.e., the number of paths in S. if the length of S is larger than out-degree threshold. Then there is need to alter site structure to improve user navigation to meet goal Otherwise, no improvement is needed for session S. Only relevant sessions and links are considered for mathematical model.

$$\sum_{(i,j)\epsilon\, E} x_{ij}\,[1-\lambda_{ij}\,(1-\varepsilon)]\,] + m \sum_{i\,\epsilon N_E} pi \ldots\ldots (1)$$

Subjected to

$$c^S_{kr} = \sum_{(i,j)\epsilon E} a^S_{ijkr}\, x_{ij}$$

Where $_{r = 1,2,\ldots,}L_{p(k,S)}$ is length of k^{th} path in mini session
k=1, 2,. . ., $L_m(S)$ is length of S .i.e. path in mini session S.

Notations:
x_{ij} = 1 if link from i to j is selected; else 0.
λ_{ij} = 1 if page i has a link to page j in current structure; 0 otherwise.
E =set of relevant links which can be selected for improving user navigation.
N_E = The set of source node of links in set of E.
a^S_{ijk} = 1 if i is the r^{th}page in k^{th}path and j is destination page in mini session S; else 0
c^S_{kr} =1 if a link from r_{th} page in the k_{th}path to target page is selected; else 0
m = penalty term set as 5.
p_i = Number of links exceeding the out-degree threshold.

The above equation (I) minimizes the cost needed to improve the website structure, where cost consist of two components 1) the number of new links to be added (the first summation) and 2) the penalties on pages containing excessive links i.e. more links than the out-degree threshold i.e. W_i (the second summation).some existing links often ignored by users because of poor design or ambiguous labels. MPM is improving such links before adding new links in existing web structure.

V. Results

Table1: Testing result on website that needs link structure to be improved

Time Threshold	Multiplier	Path Threshold	Session Timeout	Number Of Links To Be Improved
1 Min	0	1	0.20	2
		2	0.40	1
		3	0.60	0
	1	1	0.20	0
		2	0.40	1
		3	0.60	0
2 Min	0	1	0.25	2
		2	0.32	1
		3	0.60	0
	1	1	0.22	0
		2	0.40	1
		3	0.66	0

- Table 1 shows testing result on website that needs link structure to be improved.
- For column name 'number of links to be improved' indicates number of links to be improved so navigation on website can be enhanced for path threshold values 1, 2 and 3 respectively.

Table 2: Results of evolution on improved website using percentage of mini session enhanced

Multiplier for penalty term	Evolution path threshold(C=6) C=6	Mini session Enhanced
0	1	70.56%
	2	77.56%
	3	80.78%
1	1	60.35%
	2	67.78%
	3	70.16%

- After applying MPM on website it enhances mini session and reduces path required to reach target page.
- Using MPM user can locate target page within fewer clicks and also reduces bandwidth required for browsing data on website.

VI. Conclusion

This Paper addresses how to improve a website structure without introducing substantial changes. Specifically, new method a mathematical programming model (MPM) has been proposed to improve user navigation on a website with minimizing alterations to its current structure. Further model is enhanced using graph clustering mechanism. Using MPM it is possible to obtain optimal solution effectively but MPM is applicable for only static websites.The model could be further improved by incorporating additional constraints that can be identified using data mining methods.

References

[1]. Min Chen and young U.RYU "Facilitating effective user navigation through website structure improvement "IEEE,Volume 25,pp 571-588,2013.
[2]. Pingdom, "Internet 2009 in Numbers," http://royal.pingdom com/2010/01/22/internet-2009-in-numbers/, 2010.
[3]. T. Nakayama, H. Kato, and Y. Yamane, "Discovering the Gap between Web Site Designers' Expectations and Users' Behavior," Computer Networks, vol. 33, pp. 811-822, 2000.
[4]. R.Shrikant and Y. Yang"Mining web logs to improve web site organization", ," Proc. 10th Int'l Conf. World Wide Web, pp. 430-437,2001.
[5]. R. Cooley, B. Mobasher, and J. Srivastava, "Data Preparation for Mining World Wide Web Browsing Patterns," Knowledge and Information Systems, vol. 1, pp. 1-27, 1999.
[6]. B. Mobasher, R. Cooley, and J. Srivastava, "Automatic Personalization Based on Web Usage Mining," Comm. ACM, vol. 43, no. 8, pp. 142-151, 2000.
[7]. B. Mobasher, R. Cooley, and J. Srivastava, "Creating Adaptive Web Sites through Usage-Based Clustering of URLs," Proc Workshop Knowledge and Data Eng. Exchange, 1999.
[8]. W. Yan, M. Jacobsen, H. Garcia-Molina, and U. Dayal, "From User Access Patterns to Dynamic Hypertext Linking," Computer Networks and ISDN Systems, vol. 28, nos. 7-11, pp. 1007-1014, May 1996.
[9]. M. Nakagawa and B. Mobasher, "A Hybrid Web Personalization Model Based on Site Connectivity," Proc. Web Knowledge Discovery Data Mining Workshop, pp. 59-70, 2003.
[10]. B. Mobasher, "Data Mining for Personalization," The Adaptive Web: Methods and Strategies of Web Personalization, A. Kobsa, W. Nejdl, P. Brusilovsky, eds., vol. 4321, pp. 90-135, Springer-Verlag, 2007.
[11]. M. Perkowitz and O. Etzioni, "Towards Adaptive Web Sites: Conceptual Framework and Case Study," Artificial Intelligence, vol. 118, pp. 245-275, 2000.
[12]. R. Gupta, A. Bagchi, and S. Sarkar, "Improving Linkage of Web Pages," INFORMS J. Computing, vol. 19, no. 1, pp. 127-136, 2007.
[13]. C.C. Lin, "Optimal Web Site Reorganization Considering Information Overload and Search Depth," European J. Operational Research, vol. 173, no. 3, pp. 839-848, 2006.
[14]. C.C. Lin and L. Tseng, "Website Reorganization Using an Ant Colony System," Expert Systems with Applications, vol. 37, no. 12, pp. 7598-7605, 2010.