Website Structure Improvement using Tailoring Method

Milind M Shinde, Vinod S Wadne, Kuldeep B Vayadande

Abstract— **The key problem to facilitate effective user navigation by designing well structured websites. A primary cause of poor website design is that the web developers' understanding of how a website should be structured can be considerably different from those of the users. This problem is difficult to avoid because when generating a website, web developers may not have a proper understanding of users' preferences and can only organize pages based on their own judgments. A mathematical programming model to improve the user navigation on a website while minimizing alterations to its current structure is proposed. Results from extensive tests conducted on a publicly available real data set indicate that our model not only significantly improves the user navigation with very few changes, but also can be effectively solved. So, to check the performance of the improved website using the real data set we define two evolution matrices.**

Index Terms — **User navigation, web mining technique, Website designing, mathematical programming model**

I. INTRODUCTION

Lot of investments in website design is currently going on. So, in this situation finding proper information in a website is difficult and designing effective websites is not a trivial task. Generally, understanding of web developer for designing a website should be structured in such a way that it can be different from those of the users. And it is the basic cause of the poor web designing. Due to this, users can not specify the proper information in the website, as it is designed according to developer's adjustment. So, when creating a website, web developers may not have a proper knowledge of users' preferences and can only organize pages based on their self judgments and ideas. However, the effectiveness of website should be calculated in the form of satisfaction of the users not in the form of developers. Thus, Web pages should be organized in a way that generally matches the user's model of how pages should be organized.

Manuscript received Nov, 2014.

***Milind Mahadeo Shinde**, Department of computer engineering, imperial college of engineering & research, Pune, India, 7709669202.*

***Vinod S. Wadne**, Department of computer engineering, imperial college of engineering & research, Pune, India, 9404996226,*

***Kuldeep B. Vayadande**, Department of computer science and engineering, bharati vidyapeeth's college of engineering , Kolhapur Kolhapur, India, 9762315473.*

Important question is how to improve website navigability through the use of user navigation data. Lot of work have done an effort to solve this question and they are classified into two categories:

A. Personalization:

The process of facilitating a particular user by dynamically reconstituting pages based on user profile and traversal paths is called as personalization.

B. Transformation:

The modification of website structure for proper navigation of users is known as transformation.

In these techniques, personalization relates with user behaviour and user profile, sessions and history of data also called as web logs. It is created by user's activity on web site, but transformations approaches mainly focuses on developing methods to completely reorganize the link structure of a website. So, we deal mainly with transformation approaches. Considering the limitations of website reorganization techniques, we focus on the question of how to improve the structure of a website rather than reorganize it and without making drastic change to its original structure. Especially, we develop a mathematical programming (MP) model that helps user navigation on web with few changes to its original structure. This model works very well with the informational websites whose contents do not change and remains comparatively stable over time.

Finding total no of outward links available in a page, i.e., the out degree is a crucial factor in modeling web structure. Previous studies mainly focus it as hard constraints so that pages in the new structure should not have links exceeding than a specified out-degree threshold, because if too too many links in a page are given, it can cause information overload to users and is not desirable.

The field data mining try to find necessary information that is present or hidden in conventional databases, and it is the emerging field of Web mining which aims at finding and extracting related information that is hidden in Web related data. Web mining is a multi-disciplinary and it finds techniques from fields like information or data retrieval, statistics, machine learning, natural language processing, and others. Web mining deals with three important facts: web content mining, web usage mining and structure mining in web.

Here we propose a mathematical programming model to improve the navigation of user on a website while minimizing changes to its current structure. Results from various tests conducted on a publicly available real data set which shows

ISSN: 2278 – 1323

that our model can improve the navigation of user with minimum changes, and it can be effectively solved. We define two evaluation metrics and use them to check the performance of the improved website using the real data set. Evaluation results show that the navigation of user on the improved structure is greatly increased. In this case, we get to know that heavily disoriented users can get benefit from the well developed structure than the less disoriented users.

II. RELATED WORK

For supporting decision making, web sites are intended to facilitate knowledge acquisition. Based on taxonomy of factors influencing Web site usability, hypotheses are developed for usability of alternative navigation structures. These hypotheses are checked via experiments that calculate performance of user in accomplishing knowledge gaining tasks and user perceptions of usability. Here two types of experimentation are conducted for simple and comparatively complex set of tasks. Results show that a usage-oriented hierarchy or a combined hierarchy is a navigation structure associated with significantly higher usability than subject-oriented hierarchies, for these simple as well as relatively complex knowledge acquisition tasks.

The definition of Clustering or cluster analysis is, technique of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those objects remains in other groups (clusters). Clustering is a main task of data mining, and it is a commonly used technique for statistical data analysis, used in various fields, which includes machine learning, pattern identification, analysis of image, information and data retrieval, and bioinformatics etc. Various clustering techniques are proposed. In this paper comparison of three clustering technique is given. Here we start with most basic k-means clustering algorithm, The term "k-means" was first used by James MacQueen in 1967. The k-means algorithm creates k-cluster for n no. of objects according to closest mean. By randomly generating k points in the data space, K-means initialize the cluster means. This is specifically done by creating a value uniformly at random within the range for each dimension.

Every iteration of K means contains two steps: i) cluster assignment, and ii) centroid update. One of the most important aspects of k-mean is the sum of squared errors scoring function. Another algorithm we selected is Connectivity based clustering, and it is also known as hierarchical clustering, is based on the main idea of objects being more related to nearby objects than to objects farther away. These algorithms forms "clusters" by connecting "objects" based on their distance. Here the set of objects represent as hierarchical tree. This technique includes a set of N items to be clustered according to similarity measures of single linkage (shortest distance), complete linkage (larger distance) and average linkage using median and mode, we compared only single linkage.

To analyze web resource details web mining techniques are used [1]. Some important types of web mining are: Content mining, structure mining and usage mining. The contents in web page are analyzed in the content mining process. Structure mining technique is used to analyze the web site and page layouts. User access particulars are analyzed using usage mining methods. The structures web sites are changed to help and improve the user navigations. Web personalization technique rebuilds the linking of page with reference to the traversal path and profile of a particular user, while the transformation technique is used to modify the site structure for all users. Navigability of web pages can be improved by linking user navigation data to web pages. The main concept in navigability is 'Out degree threshold'. The total number of outward links in a page is known as out degree. Out degree threshold is used to control the number of links in a page to decrease information overload in a page. Page-stay time information is used to identify the targeted pages. Mini sessions are identified with processed logs and path threshold information. Mathematical programming model is used to improve the user navigation on a website with minimum changes to its current structure. To calculate backtracking pages from mini sessions Backtracking algorithm is preferred. To evaluate the navigation performance two metrics are used. These are Average user navigation and benefited user count. The web site restructuring scheme is enhanced with frequent pattern mining mechanism. Dynamic out degree threshold estimation model is preferred for the system. For the navigation pattern analysis relative link information is preferred. Sequential patterns are used to complete target page identification process.

Our study mainly focuses on improvement of web navigation efficiency. It is achieved by reorganizing Web structure [2]. Navigation efficiency is defined mathematically for both navigation with or without target destination pages, e.g. for experienced and new users. Structure stability is taken into account to help some experienced users who are not going to lose their orientation. Stability constraint can also help for website designers control the maintaining effort of Web. To reorganize Web structure for achieving better navigation results, our study proposes a mathematical programming method. The web designer can notify user requirements. Designers can also specify up to which extent the website structure should be stable. A good example for this is an e-banking. It is given to explain in which way a method performs in scenarios where user surfs with target destination. This study has the benefit of estimating and improving navigation efficiency and of relieving the designer of tedious chore to develop the structure in transformation.

To enhance the linkage format in websites for both wired and wireless devices, Gupta et al [3] found a heuristic approach and applied it to two different sites. For improving navigability it re-linked the nodes in that. The proposed technique could be used to help a designer in using initial structure, given final content pages and a set of anticipated content bundles with subjective estimates of their frequencies. Preferences of user, both at the individual and at the aggregate level, can then be measured from user access log files. Reorganized site can't take to longer sessions which include more content pages. In the Cost per benefit analysis, costs include the direct cost of changing the site, and the cost of disorientation of users after the change is a problem which needs proper investigation.

A programming model [4] proposed for reorganizing websites based on cohesion between web pages. They uses two techniques like grouping of same session and grouping of pages with co-occurrence frequency on which they performed clustering and association rule mining for pages involved in

ISSN: 2278 – 1323

the session. In this model constraint is given as length of shortest path from home page to each page.

Another proposed model is structure mining based on number of links traversed in a session [5], and here instead of directly hanging structure they added some extra links between web pages which are more frequently used and browsed.

The order in which rank assigned on page which is most popular according to user behaviour, a weighted page rank algorithm [6] is used in this model. In that k-means clustering algorithm is used. This technique also focus on parameters like in-links, out-links on every webpage.

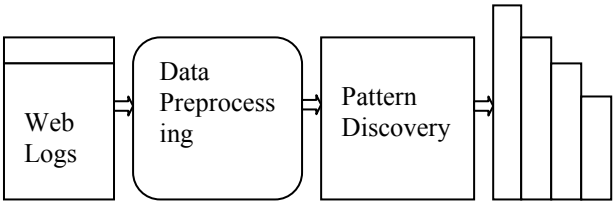
Clustering on frequent item-sets and their frequent item-sets [7] are treated as user session and access patterns. The techniques proposed for clustering and association rule mining [8] are different. Apriori algorithm is used to enhance association rule mining. It is based on pre-processed logs and for clustering co-occurrence of pages is used.

III. ANALYSIS

A. Knowledge discovery from web logs:

Calculating the Web access logs would help in predicting the behaviour of user and preparing the web structure. By considering the applications point of view, information extracted from web usage patterns can be applied to work out activities related to e-business, e-services, e-education, and on-line communities and so on. But On the other side, there is tremendous increase in the density of data and its size. So the information supplied by current web log file analysis tools may facilitate insufficient information and hence more intelligent mining techniques are needed.

For the purpose of web usage mining, web log files play an important role. The information and knowledge that can be collected from web log files is the navigation pattern of user. Different users have associated different navigational patterns with them. It is not easy to obtain such knowledge, because the users are continuously changing their focus. User can use the navigation pattern knowledge for two purposes: To help users by predicting their future request and for the personalization of websites.



B. Web Personalization and data mining for personalization:

An action which makes web experience of user personalized for taste of user is called "web personalization". It is the process of "tailoring" web pages according to requirements of users using the information of the users' navigational behaviour and profile data [9]. Web Personalization helps a particular user by rearranging pages residing on the user profile and traversal path. In this technique information related with the users is summarized. The concept of Personalization implemented in clustering results and tailoring web pages to satisfy requirements of users using the information of the users' navigational pattern and profile information.

Web Personalization can be treated as an application of data mining [10] which needs the support for all steps of data mining cycle. These steps contain data collection and pre-processing, pattern discovery and evaluation.

C. Mathematical programming model:

To recognize the active links and to insert new links mathematical programming model is performed on the mini sessions. Some parameters like out-degree threshold, path threshold and multiplier for the penalty word would be calculated. To authorize the working of the mathematical programming model two metrics are used. The first one calculates whether the enhanced structure can ease users to reach their goals. While second metric calculates the trouble of users in navigation and can get advantage from the enhancements done to the site.

A mathematical programming model is introduced to reduce alteration to the existing structure of site by enhancing the user navigation as below [11]:

$$\text{Minimize} \sum_{(i,j)\in E} x_{ij}[1-\lambda_{ij}(1-\varepsilon)] + m\sum_{i\in N_E} p_i$$

subject to

$$c_{kr}^S = \sum_{(i,j)\in E} a_{ijkr}^S x_{ij}; r = 1,2,\ldots,L_p(k,S), \quad k = 1,2,\ldots,L_m(S), \forall S \in T^R \tag{1}$$

$$\sum_{k=1}^{b_j}\sum_{r=1}^{L_p(k,S)} c_{kr}^S \geq 1; \forall S \in T^R, j = tgt(S) \tag{2}$$

$$\sum_{j:(i,j)\in E} x_{ij}(1-\lambda_{ij}) + W_i - p_i \leq C_i; \forall i \in N_E \tag{3}$$

$$x_{ij} \in \{0,1\}, p_i \in \{0\} \cup Z^+, \forall (i,j) \in E, i \in N_E. \tag{4}$$

To enhance the website organization, the objective function reduces the cost. Here cost contains two parts: 1) the number of new links to be created i.e. first summation, and 2) the penalties on pages including additional links, i.e., more links than the out-degree threshold (Ci), in the enhanced structure.

D. Web metrics:

A broad range of set of Web metrics [12] are considered in this paper. For quantification of web graph properties, web page significance, web page resemblance, search and retrieval, usage characterization and information theoretic properties, we introduce origins, measurement functions, formulations and comparison of familiar Web metrics. So for enhancing web data access and its use these metrics can be used.

ISSN: 2278 – 1323

E. Path Threshold:

Path threshold represents the target for user navigation in such a way that enhanced structure can be achieved in various ways. First it is feasible to recognize when visitors exit a website before reaching the goal from analysis of weblog files. So it allows in making proper evaluation for the path thresholds. In second case by surveying website structures, visitors can assist proper understanding of user's hopes and make sensible selections on the path threshold values. In third case a lot of client-side web usage data is gathered from firms over a broad range of websites. So analyzing every data set can also provide better features into choice of path threshold values for various types of website structures.

F. Out-Degree Threshold:

Web pages are classified into two types [13]: index pages and content pages. A page generated to assist users for proper navigation and can contain lot of links called as "index page". While a page containing information in which users are interested in and should not have too many links called as "content page". In this way, the out-degree threshold for a page is greatly reliant on the purpose of the page and the website. The out-degree threshold for content pages should be lesser than index pages. Generally out-degree threshold for a page could be set at a little bit values when most web pages have comparatively small links and new links added the threshold can be continuously improved.

IV. CONCLUSION

In this paper, a mathematical programming model is proposed to enhance the navigation efficiency of a website while decreasing alteration to its current structure, an important factor that has not been examined in the literature. Our model is suitable for informational websites whose contents are comparatively steady over time. It enhances a website instead of reorganizing it and hence it is appropriate for website maintenance on a progressive basis. Various tests on a real website showed that our model could provide major improvements to user navigation by increasing few new links. Best possible solutions were suddenly obtained, which suggest that our model is very efficient to real world websites. To scale up the performance, the MP model was observed. It can solve large-sized problems in a few seconds and generally on a desktop PC. To check the performance of our model, we have used two metrics and used them to calculate the enhanced website using simulations.

REFERENCES

[1] Ms. Jissin Mary Kunjukutty, Ms. A.Priya, "Optimizing User Navigation with Pattern based Web Site Restructuring Scheme", IJIRCC and ICGICT, Vol.2, Special Issue 1, March 2014

[2] May Wang and Benjamin Yen, "Web Structure Reorganization to Improve Web Navigation Efficiency", 11th Pacific-Asia Conference on Information System

[3] Gupta R., Bagchi A. and Sarkar S. (2007) 'Improving Linkage of Web Pages', INFORMS J.Computing Vol. 19 No. 1 pp.127-136

[4] C.C. Lin, "Optimal Web Site Reorganization Considering Information Overload and Search Depth," European J. Operational Research.

[5] R. Gupta, A. Bagchi, and S. Sarkar, "Improving Linkage of Web Pages," INFORMS J. Computing, vol. 19, no. 1, pp. 127-136, 2007.

[6] Amar Singh,navjot Kaur,"To Improve the Convergence Rate of KMeans Clustering Over K-Means with Weighted Page Rank Algorithm" IJARCSE August 2013.

[7] Bamshad Mobasher,Robert Cooley, Jaideep Srivastava "Creating Adaptive Web Sites Through Usage-Based Clustering of URLs".

[8] Bamshad Mobasher,Robert Cooley, Jaideep Srivastava"Automatic Personalization Based on Web Usage Mining".

[9] Eirinaki M. and Vazirgiannis M. (2003) 'Web Mining for Web Personalization', ACM Trans. Internet Technology Vol. 3 No. 1 pp.1-27

[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] Min Chen and Young U. Ryu, "Facilitating Effective User Navigation through Website Structure Improvement", VOL. 25, NO. 3, MARCH 2013

[12] Devenish Dyane, NG WeeKeong and Sourav S Bhowmick, "A Survey of Web Metrics", ACM Journal Name

[13] Y. Fu, M.Y. Shih, M. Creado, and C. Ju, "Reorganizing Web Sites Based on User Access Patterns," Intelligent Systems in Accounting, Finance and Management, vol. 11, no. 1, pp. 39-53, 2002.

ISSN: 2278 – 1323