IMPROVE USER NAVIGATION ON A WEBSITE BY USING WEB STRUCTURE MINING

Mohini B. Deshmukh , Sonali S. Harjit , Seema T. Jadhav , Dhanashri R. Kadam

*1 (B.E Student SVPM’s College Of Engineering, Malegaon)
*2 (B.E Student SVPM’s College Of Engineering, Malegaon)
*3 (B.E Student SVPM’s College Of Engineering, Malegaon)
*4 (B.E Student SVPM’s College Of Engineering, Malegaon)

Mohinideshmukh29@gmail.com, sonaliharjit29@gmail.com, seemajadhav271994@gmail.com, dhanashri.r.kadam01@gmail.com

Abstract
Website designing is easy but to effective user navigation is difficult task since user behaviour repeatatively change. Webmaster develop website with their own judgment that are different from user expectation. So, users suffer from different problem of searching in the network. Completely restructured website is highly unpredictable that increases the number of strange users. We propose mathematical programming to facilitate the user navigation on website with minimal changes to its current structure. We provide dynamically links to the users according to their access pattern which ensures that decreases the number of disoriented users. Our goal is to improve user navigation with minimizing alteration to its current structure by using web structure mining technique.

Keywords— Website designing, user navigation, mathematical programming, web structure mining, searching.

1. Introduction
In 21st century, Internet is become big need of peoples in world. Website having number of web pages and on these pages number of hyperlinks are present for navigation on website so it is very difficult to find information in which user is interested. Now a day’s evolution in internet on large scale, still people give preferences to ordinary brick & mortal store Instead of online store. When users are unsatisfied to find needed information so they leave website immediately. To resolve this problem web developer must think according to user point of view. The web pages should be organized in a way that generally matches the user’s model of how pages should be organized. The problem of improving user navigation on a website with minimal changes to the current structure. In last few years’ research peoples focus on various issues like understanding of web structures, finding relevant pages of a given page, mining informative structure of news website and extracting template from web pages.

There are two approaches of facilitating effective user navigation. First, personalization approach [6] in which according to individual user provide dynamically Reconstituting pages depends on user profile and traversal path. Second, In Transformation approach, to change the structure of website to easy navigation for all users. Due to completely change in the structure of website users are unfamiliar with particular website and number of disoriented users become large. Our project model is applicable to information website whose contents are statically stable over time like university, college websites, hospital, sports, travel websites etc. We use web log to improve structure of website.

2. Literature Survey
Many web users use browser to search particular information in which they are interested but, still it is very hard to retrieve information quickly.

Nakagawa and Mobasher[6] use clustering technique to generate dynamic links according to user profile and traversal path. But, this system consider only local structure instead of complete website structure.

Perkowitz and Etzioni[7] describe an approach that automatically find out index pages which contain links to pages having particular topics depends upon the frequently occurrence of pages in user navigations, to improve user navigation.
Gupta[4] propose a model in which uses aggregate user preference data to relink web pages which improve user navigation but time require for computation is very large.

Lin develop integer programming model according to relation between different web pages which reduce searching time and searching depth. This model reduce computation time but applicable to only small website.

Lin and Tseng[4] develops Ant colony system which reorganize website but problem of scalability still remains unanalyzed. So, our approach is to complete reorganization of website change location of familiar items and increases number of disoriented users. So, we facilitate user navigation with minimal changes in website structure[1].

3. System Design

In this section we give overview of the system design. Referring to Figure 1, the system consist of two phases Online phase and Offline phase. In Online phase web server is responsible for maintain user session information and it is also responsible for update website structure. Offline phase responsible for web log analysis and create dynamic links.

Offline phase contain modules such as Maintain weblog, Find target pages and Create dynamic links. The main task of offline phase is to maintain weblog. The page is said to be a target page when its stay time is greater than time threshold. After that we established dynamic links that actually improve user navigation on website.

3.1 An Example:

We use an example to describe concepts and extraction of the metric from weblog files. To analyze the interaction between web users and a website, the weblog files must be split into web user sessions. A session is a collection of activities performed by a web user during his navigation on website. In this definition, a session may contain one or more target pages, as a user may visit several goals during a individual session. Since the metric used in our analysis is the number of paths traversed to reach one target, we use a another term mini session to define as a set of paths visited by a web user for only single target. Thus, a session may include one or more mini sessions, each of which comprises a group of pages traversed to find the target.
Specifically, we identify whether a page is the target page by calculating if the time spent on that page is greater than a timeout threshold. We depict in Fig. 1 a hypothetical website that has 10 webpages. Fig. 2 describes a mini session, where a user starts from 0, browses 3 and 6, and backtracks to 3, from where he visits 2, 1, 4, 8 and backtracks to 1. Then, this user goes from 1 to 5 and finally reaches the target 9. We denote the mini session by $S = \{0, 3, 6\}, \{2, 4, 8\}, \{5, 9\}$ where an element in $S$ shows a path traversed by the web user. In this example, mini session $S$ has three different paths as the user backtracks at 6 and 8 before reaching the target 9. Note that 3 and 1 only appear once in $S$ because of caching.

To achieve the web user efficient and effective navigation goal, the website structure must be modified in a way such that the group of pages required to find the targets in the new improved structure is not greater than the path threshold. From given example in Fig. 2, the user has traversed three different paths before finding the target. A solution to help this web user locate the target easier by introducing more links. There are number of ways to insert extra links. If a link is inserting from 3 to 9, the user can directly reach 9 via 3, and hence locate the target in the only single path. Thus, inserting this link “saves” the web user two paths. Similarly, establishing a link from 1 to 9 enables the web user to locate the target in the second path. Hence, this saves him single path. We could also add a link from 4 to 9, and this is considered the same as linking 1 to 9. This is because both 1 and 4 are web pages visited in the second path, so linking either one to 9 saves only single path. Yet, another possibility is to link 2 to 5, a non target page. In this case, we assume that the web user does not insert the new link, because it does not directly connect given page to the target.

4. Flow of Project

Our goal is to reconcile the user on website. The problem is concern with graph optimization problem[1]. We model website as directed graph with nodes representing pages and edges representing links between two pages. When The system consist of following components:

4.1 Registration:

The main requirement of our system is registration. After registration we are able to identify particular user on website that useful to maintain the user sessions.

4.2 Maintain Weblog:

For maintain weblog user navigate different pages on the website to maintain user session. Weblog contains number of sessions, it is nothing but the time period between login to logout. One session contain number of minisession. We use a term minisession to refer to a number of pages visited by a user to achieve single target.

4.3 Find Target Pages:

From weblog we find target pages, we can say that the page is target page if stay time of that page is greater than time threshold. Stay time is nothing but time spend on that page. Time threshold value is set by webmaster.

4.4 Create Links Dynamically:

By using weblog we find target pages and also find the optimal path to achieve the goal. After finding the target pages create links dynamically. We set penalty term for each page to prevent the structure from adding the new links to the pages which already having more links and update website structure which improve user navigation.

The problem of facilitating the user navigation on a website with minimal changes to its current structure can be formulated using mathematical programming model as follows.
The objective function minimizes the cost required to improve the website structure, where the cost including two components: 1) the number of new links to be established (the first summation), and 2) the penalties on pages containing excessive links, i.e., more links than the out-degree threshold $C_i$, in the improved structure (the second summation). Constraint (2) requires that the goal for user navigation be achieved for all relevant mini sessions, where the goal is defined as path threshold $p_i$. Particularly, for a mini session $S \in T_R$ in which the user navigation is below the specified goal, i.e., $L_m(S) > b_j$ for $j \leftarrow \text{tgt}(S)$, at least one link from pages visited on or before the $b_j$th path to the target page $j$ is either established or improved so that the user can reach the page $j$ within the path threshold set by the Webmaster. Constraint (3) uses $p_i$ to capture the number of links exceeding the out-degree threshold $C_i$ for page $i \in NE$. This value ($p_i$) is then used to compute penalties in the objective function. The degree of penalty can be controlled by the multiplier for the penalty term ($m$). Constraint (4) imposes that decision variables are binary and $p_i$ are non negative integers.

**TABLE I**

Summary of Notation

<table>
<thead>
<tr>
<th>Notation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S$</td>
<td>It is mini session that contains group of pages visited by User to reach target page.</td>
</tr>
<tr>
<td>$T$</td>
<td>Set of all mini sessions.</td>
</tr>
<tr>
<td>$T_R$</td>
<td>The set of all relevant mini sessions i.e. that need to be facilitated for given path threshold.</td>
</tr>
<tr>
<td>$N$</td>
<td>The set of all web pages.</td>
</tr>
<tr>
<td>$\lambda_{ij}$</td>
<td>1 if page $i$ has link to page $j$; otherwise 0.</td>
</tr>
<tr>
<td>$E$</td>
<td>Set of candidate links which can be selected for improving user navigation.</td>
</tr>
<tr>
<td>$E_R$</td>
<td>Set of all relevant candidate links that can help to meet the navigation goal.</td>
</tr>
<tr>
<td>$NE$</td>
<td>The set of source nodes of links in set $E_r$.</td>
</tr>
<tr>
<td>$W_i$</td>
<td>The current out-degree of page $i$.</td>
</tr>
<tr>
<td>$C_i$</td>
<td>The out-degree threshold for page $i$.</td>
</tr>
<tr>
<td>$p_i$</td>
<td>The no. of links exceed out-degree threshold $C_i$ in page $i$.</td>
</tr>
<tr>
<td>$m$</td>
<td>The multiplier for penalty term in objective function.</td>
</tr>
<tr>
<td>$b_j$</td>
<td>The path threshold for mini session $S$ in which a $S_{ijkr}$ is the rth page in kth path and j is target page in mini session $S$; otherwise 0.</td>
</tr>
<tr>
<td>$a_{ijkr}$</td>
<td>1 if $i$ is the rth page in kth path and $j$ is target page in mini session $S$; otherwise 0.</td>
</tr>
<tr>
<td>$x_{ij}$</td>
<td>1 if the link from page $i$ to page $j$ selected; otherwise 0.</td>
</tr>
<tr>
<td>$C_{kr}$</td>
<td>1 if in mini session, a link from rth page in kth path to the target is selected; otherwise 0.</td>
</tr>
<tr>
<td>$\text{tgt}(S)$</td>
<td>Target page of mini session $S$.</td>
</tr>
</tbody>
</table>
5. Our Contribution

We propose a mathematical model in our system in that we find target pages based on their stay time also confirm that selected page is actually target page or not. For example, when user surfing on a website if any obstacle occur such as phone call so unfortunately user stay time on current page increases but current page is not its target page .The page is actually target page if it’s stay time is greater than time threshold and page is frequently visited by user. By using optimization technique, we find optimal path to achieve target.

6. Result Analysis

Our system not changes the current structure of website only give expected pages links to the user. We maintain weblog for individual user to understand their behavior. From the weblog we provide optimal path to visitor and create dynamic links to reach target fastly .Our system reconcile the user navigation on website.

7. Conclusion

In this paper, We proposed the mathematical programming model to improve user navigation on website with few changes to its current structure. Our model is particularly suitable for informational websites whose contents are relatively stable over time .It improve website structure instead of completely restructured it. It provided dynamically links to the user for efficient navigation with their behavior. We test on the real website showed that our model could provide Significant improvements to user navigation by adding only few new links. This gives the very important information to the webmaster for providing efficient and easier navigation to the website.

References
