Prioritizing the Links on the Homepage:
Evidence from a University Website

Lian-lian SONG\textsuperscript{1,a}\textsuperscript{*} and Geoffrey TSO\textsuperscript{2,b}

\textsuperscript{1}Nanjing University of Aeronautics and Astronautics, Nanjing, China
\textsuperscript{2}City University of Hong Kong, Hong Kong, China
\textsuperscript{a}songll@nuaa.edu.cn; \textsuperscript{b}msgtso@cityu.edu.hk

\textsuperscript{*}Corresponding author

Keywords: Homepage; Links; Random Utility Theory; Clickstream Data

Abstract. A common problem encountered in Web design is how to arrange the homepage links of a Website. Previous empirical studies have not complete systematic guidelines to resolve this problem. This paper uses a random-utility theory for studying visitors’ choice behaviors to prioritize the links on a homepage. We propose a total of three multiple-choice demand models and apply them to clickstream data collected from an educational institute over a seven-and-a-half month period. The model with the best performance is selected as the tool for constructing a metric, utility loss. Empirical results show that the proposed metric is highly efficient for prioritizing the links on a homepage and the algorithms can also be used to study the feasibility of introducing a new function in a Website.

Introduction

With the increasing importance of Websites and related issues, studies that investigate Web design are also emerging rapidly (Simon, 2001; Chen et al., 2009; Masseglia et al., 2009; Caballero-Luque et al., 2010; Wang and Lee, 2011). Surprisingly, however, studies on link-prioritization on the homepage based on statistical models built from clickstream data are scarce. There is no complete empirical work investigating specific guidelines for link arrangement. We present a first attempt to produce the instrument of link-prioritization guidelines based on clickstream data in this article.

Imagine a Website that contains \( n \) Web pages. Each visitor to this Website spends different amounts of time (some could be zero) to view each of these pages. In order to study the amount time spent by a visitor on different pages, one could use the concept of “utility” in economics to measure the reward of viewing a page as utility is generated whenever a page is viewed. Here the utility is defined as the utility of information; when visitors (whether they are repeat visitors or not) read a Web page for one minute, they get information and understand more about the products shown at the Website. The more time they spend reading the page, the more information they obtain and thus more is the utility they get, but at a decreasing rate due to satiation. The time allocation model is under the constraint of limited time spent online and this is a reasonable assumption since each visitor has limited time. This is just like a consumer going to a supermarket having limited money and people having limited entertainment time in weekends. Based on this utility theory, three new models are proposed in this paper to study how visitors allocate time on different links on a Website.

We illustrate the use of the developed models by answering the questions of prioritizing the links on the homepage: 1) which links should be included on the homepage? 2) Which links should be removed from the homepage and added as secondary links (We define the links on the homepage as the primary links. Clicking a primary link to open a Web page, the links on this Web page are secondary links)? 3) how to arrange useful links? Which one should be put in the first place and which should be put at the end of the page? Related to this content placement strategy, practitioners agree that packing more information and links onto the homepage in general increases click-throughs (Murphy, 1999). However,
given the limited space of a homepage, links of only a subset of Web pages can be displayed on the homepage for direct clicks. Use of small fonts for links on the homepage of the Website results in two problems: One is that the homepage does not have a pleasant look and becomes unattractive because of non-artistic design with so many words clamping together on it; the other problem is that it is also a waste of time for a visitor to find the right information among many links. Consequently, understanding visitors’ preference for links can help identify links that should be shown on the homepage. Using the developed models, we are able to create a metric to measure the relative attractiveness of links. The metric is the utility loss caused by removing each link from the homepage. We test our models using clickstream data of an educational institute. Empirical results confirm that the popular model in economics can achieve good results in studying Web browsing behavior. The proposed metric is highly efficient for prioritizing

**Model Construction**

We define the total reading time (TRT) as the sum of all time that a viewer spends on all chosen Web pages, which is of course limited for each visitor. The reading time is used to calculate the utility gained from this visit. Based on the random-utility framework principle, a Website visitor allocates the available time efficiently so as to maximize the utility. Similar concepts have been studied in transportation and entertainment also, for problems such as time allocation to different combinations of types of vehicles and multiple types of activities.

We briefly introduce the additive utility function, constructed as:

\[
U = \sum_{j=1}^{J} \psi_j (t_j + 1)^{\alpha_j} e^{\varepsilon_j}, \varepsilon_j \sim \text{GEV} \tag{1}
\]

Where \(\psi_j\) is the baseline utility parameter, \(t_j\) represents the time allocation to variety \(j\). \(\alpha_j\) is the satiation parameter. \(\psi\) and \(\alpha\) together decide the rate of diminishing marginal utility. The high values of \(\psi\) and \(\alpha\) indicate high base preference and low satiation, which results in most part of total time being allocated to this variety. However, the two parameters are interactive with each other.

Equation (1) is considered as a valid function under the restrictions that \(\psi_j > 0\) and \(0 < \alpha_j < 1\). The constraint \(\sum t_j = T\) is specified to guarantee a finite utility. Under the assumption that individual varieties are imperfect substitution to each other, we are in search of equilibrium of time allocation over the varieties to maximize the total utility. \(\varepsilon\) is the error term, which captures the unobserved elements that impact the utility. We use exponential form to ensure the positivity of the utility. GEV is short for generalized extreme value. Here, we follow the step of Bhat (2005) to assume \(\varepsilon_j\) a standard extreme value distribution, since such distribution leads to a closed form of probability expression.

In light of the theory of additive utility function, we extend Kim et al.’s model by developing two other versions of the utility function, which would be more useful in application. The first version is similar to the one used in Kim et al. but all \(\alpha\) are assumed equal across varieties to reduce the interaction with baseline utility. The second version is the simplest one, containing only one parameter for each variety. We compare the performance of the three models and choose the best to help prioritize the links on a homepage.
Empirical Study

The data used in empirical work are the clickstream data of the admission Website of City University of Hong Kong collected from Oct. 1st 2009 to May 15th 2010. During this period, more than 100,000 individuals visited the Website. The web-log file records second-by-second visitors’ clicking behavior including type and duration of Web pages visited. This information can be used to develop our models. The TRT spent on the Website is used to select samples. We select 500 visitors whose TRTs are between 117 to 134 minutes (7000 to 8000 seconds) on the Website.

The maximum likelihood estimation (MLE) method was adopted to estimate the parameters in the models. All estimation procedures in the following sections employed the programming language R 3.0.2. First, we should handle the identification problem: the nonuniqueness of the parameters. Values of the baseline utility/satiation parameters for different varieties are relatively defined and hence one of them must be fixed. We fix the baseline utility of variety 1, $\psi_1=1$, and the satiation of variety 2, $\alpha_2=0.1$.

There are a total of three versions: Full Model with different satiation parameters, the Reduced Model with the same satiation across all varieties and the Simplest Model without satiation parameter. The objective function (Obj) is the negative log likelihood function, which we aim to reduce to a minimum. The comparisons of objective function values shown in Table 1 suggest that the Full Model outperforms the other two models. The BIC value has reduced in Full Model. However, please pay attention to BIC values of Full Model and Reduced Model in the Browsers group; the difference is very small, indicating the performance of the two models is almost the same. When confronting a large number of varieties, Reduced Model may be a better choice since it would achieve an approving performance but needs much less effort to estimate parameters.

<table>
<thead>
<tr>
<th>Models</th>
<th>Obj</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Model</td>
<td>5290.50</td>
<td>5301.28</td>
</tr>
<tr>
<td>Reduced Model</td>
<td>5316.39</td>
<td>5323.12</td>
</tr>
<tr>
<td>Simplest Model</td>
<td>5337.78</td>
<td>5343.17</td>
</tr>
</tbody>
</table>

Practical Implications

It is infeasible to include links for all pages on the homepage. An excessively large number of links on the homepage may make it look overcrowded and confusing. Moreover, it is also a waste of time to find the needed page among many links. As a result, managers have to decide which links to include on the homepage. Which one should be put in the first place and which should be put at the end (or corner) of the page? Which links should be removed from the homepage and added as secondary links? These problems directly influence the attraction of a Website.

We apply the developed models to build the metric utility loss. The calculation process is under the assumption that visitors are always optimally allocating their time, in accordance with the marginal utility theory. It is easy to understand that no one would like to spend time reading useless information. After removing a link from homepage and adding it as a secondary link (included in other variety), we assume visitors’ preference for the remaining varieties remains unchanged. This is a reasonable assumption since visitors do not know the exact place (URL) of the removed variety. Their attitudes towards the remaining varieties during the information search cannot change. Since the total time spent
on the Website by different visitors varies, we obtain the utility difference for each visitor in the sample. The final utility loss caused by removing one link is the average of the utility loss over the whole sample. Table 2 reports the utility loss for removing each of the five links for the browsers and the applicants. We can see from Table 2 that removing link 1 from the homepage leads to the biggest utility loss, deducting 86.9% of the original total utility gained from five links. Removing either link 3 or 4 brings the least effect on reduction of visitors’ utility. The standard errors of the utility loss are very small, indicating the effects are quite homogeneous over the population.

Table 2. Average utility loss of each link for browsers.

<table>
<thead>
<tr>
<th>Link</th>
<th>Utility loss</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>link 1 (variety 1)</td>
<td>1.078 (7.98e-03)</td>
<td>86.9</td>
</tr>
<tr>
<td>link 2 (variety 2)</td>
<td>0.093 (1.09e-03)</td>
<td>7.5</td>
</tr>
<tr>
<td>link 3 (variety 3)</td>
<td>0.007 (2.24e-05)</td>
<td>0.6</td>
</tr>
<tr>
<td>link 4 (variety 4)</td>
<td>0.003 (2.23e-05)</td>
<td>0.2</td>
</tr>
<tr>
<td>link 5 (variety 5)</td>
<td>0.023 (7.11e-05)</td>
<td>1.9</td>
</tr>
</tbody>
</table>

(Figures in the parentheses are standard errors and percentages do not necessarily sum to 100)

Conclusion

Managers always encounter the problem of how to organize the links of the homepage of a Website. They need to decide which link should be put in the most important place, which one at the end of the homepage. When the number of links is very large, the limited space of homepage cannot accommodate all links; some need to be added as secondary links. What is the evidence to support the decision idea? With the help of the proposed models, we construct a metric to study the effects of removing a link from the homepage. That is “utility loss”, which is defined as the difference in maximum utility obtained by a visitor before and after the removal of a link. When a link is removed from homepage and added as a secondary link, visitors need to spend more time on finding the needed information or may not even find it, which causes the loss of total utility. The metric, utility loss, is used to measure the negative effect of transferring a primary link to a secondary link. It is very useful to resolve the link-prioritization problem.

Moreover, utility loss can also be applied to prioritize links of search engines, for example, Google, Yahoo, etc. The search Web page is generally full of links collected by keyword matching. However, not all good-matching links are shown on the primary place are attractive. Utility loss can detect the importance of each link and decide which should be deleted. In the similar theory, utility loss can be used for examining feasibility of introducing a new feature to a Website. Another extended application of utility loss is to split a popular primary link into two primary links. When utility loss confirms a very important link which is favorite by nearly all visitors, managers can consider dividing the contents of the link into two parts. In this case, visitors are able to more conveniently and quickly find the useful information.

Acknowledgement

This research is financial supported by NSFC (No.71602089), Natural and Social Science Foundation of Jiangsu Province, (No.BK20160785, 16GLC001), and the Fundamental Research Funds for the Central Universities (NJ20160084).
References


