Electronic Commerce Research and Applications 21 (2017) 65–78

Contents lists available at ScienceDirect

Electronic Commerce Research and Applications

journal homepage: www.elsevier.com/locate/ecra

Context preserving navigation redesign under Markovian assumption for responsive websites

Vinay Kumar *, Mamata Jenamani

Department of Industrial and Systems Engineering, Indian Institute of Technology, Kharagpur 721302, India

ARTICLE INFO

Article history:
Received 15 June 2016
Received in revised form 23 December 2016
Accepted 23 December 2016
Available online 26 December 2016

Keywords:
Click-stream data
Contextual relevance
Implied deviation
Markov process
Meta-heuristics
Minimum derived out-degree
Mobile Internet
Responsive web design
Website effectiveness

ABSTRACT

Identification and sorting of contextually relevant links are important for navigation design of responsive websites. While *average click ratio* is often used as a metric to evaluate contextual relevance of navigation structure from users' view, we propose a metric called *implied deviation* to quantify the same from designer point of view. Average click ratio minimization problem typically solved using meta-heuristics has issues such as loss of designer-defined contextual relevance and loss of connectivity among webpages. To solve this problem, we propose a *deviation minimization framework* to suggest context-preserving navigation structure. The proposed framework consists of three stages. Stage 1 models user navigation behavior as a Markov process and generates a transition probability matrix. Then we use the transition probabilities as weights to relax the original average click ratio minimization problem, and bring it to a form similar to a transportation model. The corresponding solution is considered as the initial basic feasible solution of the original problem. In Stage 2, transition probability guided meta-heuristics improve upon the initial basic feasible solution. Specifically, we use modified simulated annealing and ant colony optimization algorithms. This resulting solution is further weighted with transition probabilities to make a tradeoff in Stage 3. We obtained experimental results based on two education and university-related datasets to show that the proposed framework can achieve a good tradeoff for contextual relevance in terms of both the average click ratio and implied deviation.

1. Introduction

Today, millions of people have fast, pervasive access to the Internet through desktop machines, smart phones, tablets, and other portable devices (Mobile Marketing Statistics, 2015). This diversity creates a challenge for the web designers to build sites that are responsive enough to adjust the content to different screen sizes. *Responsive web design* offers a solution to the content redesign problem, but its navigation design is still a challenge. Such sites when viewed on mobile devices suppress many links which otherwise appear on desktops (Lestari et al., 2014). In this context, the web designer faces two challenges: (1) selecting the right links to be displayed on a page; and (2) displaying the important links in the order of relevance with the screen size in mind. The problem of redesigning a website's structure based on user navigation pattern can be formulated as an optimization problem. The objective is to minimize the number of clicks made by a user per webpage visited during a session (Lin and Tseng, 2010; Singh and Kaur, 2014b). This is called the *average click ratio* (Gupta et al., 2007).

As a solution to this problem, the original website structure is viewed as a graph and replaced with another graph with a different link structure (Lin and Tseng, 2010). In this graph, the nodes are the webpages connected by the links which appear on each webpage. The initial matrix, including the original site navigation structure, represents the contextual relevance of the webpage connectivity from designer's point of view. We refer to this as the *designer-defined context*. The solution matrix represents the new navigation structure that balances the *designer-defined context* and contextual relevance from the users' viewpoint based on click-stream data.

The success of the proposed system hinges on improving user experience in a website. User centric design aims to make a software system progressively more usable. The system can be a website, a web application, desktop software or some form of human-computer interaction interface. Improving the navigability of websites through the addition or modification of structural elements has been shown to improve user performance and perception about the site, and adds to usability improvement (Spyridakis et al., 2007). Navigation design can affect user perception in two

* Corresponding author.
E-mail addresses: vinay2k2@gmail.com (V. Kumar), mj@iem.iitkgp.ernet.in (M. Jenamani).

http://dx.doi.org/10.1016/j.elerap.2016.12.003
1567-4223/

ways (Sundar, 2007). First, there is the sheer presence effect: well-structured navigation on a site gives it credibility. Second, the characteristics of the hyperlinks or of the linked content can cause the user to perceive the site as more or less credible (Flavián et al., 2006). A measure for the effect of navigability on attitudes may be related to the increased self-efficacy associated with helping users complete specific tasks quickly (Hsieh et al., 2015). For existing customers, the website is a place to become more familiar with the organization and to obtain information on its products and services (Zhou, 2016). An effective website also must feature well-designed navigation and ease of operation to create a positive *user experience* (Siegel, 2007; Tan and Wei, 2007).

The behavior and information needs of website user change continuously. Hence, the link structure should be modified accordingly, or else users may leave the website if it takes too long to find the required webpage quickly (Singh and Kaur, 2016). Frequent changes, however, create confusion for returning users (Chen and Ryu, 2013). User browsing consensus encourages changing the site structure, whereas designer-defined context prompts keeping the existing linkage structure intact.

In this scenario, the research problem is how to make changes in the linkage structure while balancing both factors. *Transformation-based websites* reorganize their link structure according to the major browsing consensus of the users (Perkowitz and Etzioni, 2000). The input to such websites is user navigation behavior captured in log files based on the existing website structure, and the output expected is improved website structure. In this research, we extend this idea to propose transformation-based changes that captures browsing behavior and preserve the designer-defined context to make the right tradeoff.

Experts use techniques such as card sorting and iterative tree testing to evaluate information architecture to get ideas for restructuring sites and to increase searchability and discovery. Analysis of website content, information hierarchy, and classification schemes help them to spot problem areas and identify opportunities for improvement (Nielsen Norman Group, 2016). Current academic research ignores the designer-defined context, especially the assumptions made while creating page linkages. The literature has been too user centric, and this has led to loss of contextual relevance in the page linkages. This may be happening due to the nature of the HTTP protocol, which fails to capture the user access sequence at the server level. At the same time, disregarding user browsing behavior can hinder progressive site design. This study makes a tradeoff between the *designer-defined context* and the *user-defined context* to help restructure websites with minimal changes and contextual loss.

A typical commercial website contains thousands of pages, resulting in a colossal integer programming problem for *average click ratio* minimization. This cannot be solved online nor can an exact solution be obtained offline. Meta-heuristics based approaches that are typically used to solve such problems decrease the click ratio (Gupta et al., 2007; Lin and Tseng, 2010). Our experience of working with a large data set shows that the suggested link structure has two major problems. First, contextual relevance is often lost; and second, some of the pages become disconnected. We suspect this may be happening because solutions obtained from meta-heuristics are sub-optimal, and the major source for collecting navigation data is access log files. But these do not contain all the requests made to the server through intermediate caches. So the identification of exact client sessions is difficult due to the stateless HTTP protocol. As a result, a number of assumptions must be made for data cleaning and session identification (Gupta et al., 2007). The suggested navigation structure based on such data may be out of context though. And deviation from the original navigation structure creates confusion in the minds of regular visitors (Chen and Ryu, 2013).

We propose a navigation redesign solution which offers important links which can be selected according to different screen sizes. Fig. 1 shows the basic ideas and the research steps to realize them.

We formulate this problem as a web structure optimization problem. We pre-process and obtain an initial basic feasible solution in Stage 1. Next, we optimize and contrast the results that are obtained by applying different meta-heuristics, including *simulated annealing* and *ant colony optimization* in Stage 2. We improve on both of the algorithms by using transition probabilities. In Stage 3, we accommodate link changes by further weighting them with transition probabilities. We store this data and the modified link structure in a database. When they are needed, the top *N* links may be retrieved in real time and displayed in various environments. The average click ratio minimization problem has been addressed by Gupta et al. (2007) and Chen and Ryu (2013). Our solution approach is based on Markovian analysis of users' navigational behavior. We represent their patterns using a transition probability matrix where the rows and columns represent webpages.

This assumption addresses the stochasticity involved in user navigation, limits the changes in the site structure, and preserves the contextual relevance of the original site. The initial website

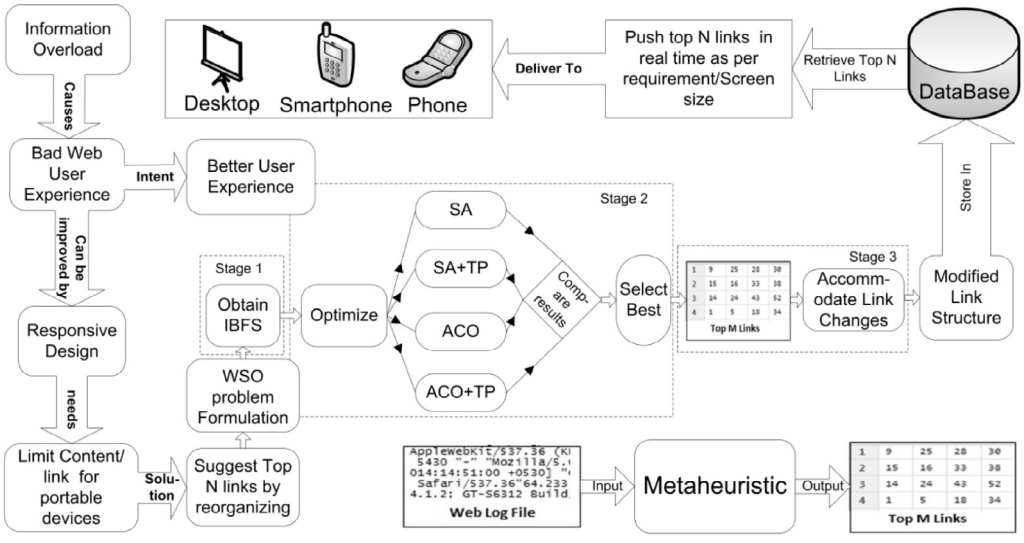

Fig. 1. Steps in the research.