Applied Mathematics and Computation 219 (2013) 11082–11095

Contents lists available at SciVerse ScienceDirect

Applied Mathematics and Computation

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

Optimization of multi-criteria website structure based on enhanced tabu search and web usage mining

Peng-Yeng Yin *, Yi-Ming Guo

Department of Information Management, National Chi Nan University, Nantou 545, Taiwan

ARTICLE INFO

Keywords:
Adaptive web site
Website structure optimization
Web usage mining
Tabu search
Genetic algorithm

ABSTRACT

With the rapid development in World Wide Web (WWW) technology, the number of webpages and the volume of information content have been overwhelming. It becomes increasingly important to help users find relevant webpage and information more easily and quickly. This situation causes widespread attention in constructing adaptive websites which automatically reorganize the structure or content by learning from the users' browsing behaviors, as such the usage of the websites is improved. In this study we propose a new formulation for the website structure optimization (WSO) problem based on a comprehensive survey of existing works and practice considerations. An enhanced tabu search (ETS) algorithm is proposed with advanced search features of multiple neighborhoods, adaptive tabu lists, dynamic tabu tenure, and multi-level aspiration criteria. The experimental result on 24 real-world problem instances shows that the proposed ETS algorithm can obtain a better value of web usage estimation than a genetic algorithm method. Moreover, ETS is computationally efficient due to the strategy that handles problem constraints on-the-fly when constructing the solution.

1. Introduction

The development of World Wide Web (WWW) in the recent decade has created successful applications such as search engines, e-commerce, e-government, weblogs, social network communications, etc. With the overwhelming websites created, WWW plays a central role not only in business communities but also in our daily lives. An empirical study [1] of user perceptions about websites discloses that the most important design features for different website domains include navigations, timeliness, readability, visualization, accuracy, and security. The easy-to-navigate feature is ranked among the top three for all domains. Web users look forward to more comfortable browsing experiences which require the WWW environment to be both *effective* and *efficient*. Effective browsing means that the users can easily search the most interesting website by specifying relevant keywords, while efficient browsing indicates the users can reach the target webpage in a website with just few clicks. Both requirements can be facilitated by using the *web mining* techniques in the design phase.

Cooley et al. [2] classified web mining into two classes, namely, web content mining and web usage mining. The web content mining analyzes various web objects (text, images, voices, videos, banners, etc.) to extract useful information regarding the websites. This conception is broadly applied in website search engine optimization (SEO) [3] and website content redesign [4]. The web usage mining aims to discover frequent browsing patterns based on click-stream data analysis. The significant patterns can be used to create adaptive websites. Perkowitz and Eizioni [5] defined adaptive website as sites which automatically improve their organization and presentations by learning from visitor access patterns. Two types of adaptive

* Corresponding author. Address: 470 University Rd., Puli, Nantou 545, Taiwan.
E-mail address: pyyin@ncnu.edu.tw (P.-Y. Yin).

0096-3003/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.amc.2013.05.033

websites have been developed. *Customization-based* websites display variable content for each user to satisfy particular need, while the *Transformation-based* websites modify the content or reorganize the link structure according to significant browsing consensus of most users [6]. To realize the design of adaptive web sites, data mining techniques such as classification, clustering, and association rules have been intensively used [7,8].

This paper focuses on transformation-based website structure optimization (WSO) by using an enhanced tabu search and web usage mining techniques. The contributions of this research are described as follows. *First*, we propose a new WSO problem formulation which is more realistic with complete considerations of real-world practice based on a comprehensive survey of existing works. *Second*, we observe that most existing methods for tackling WSO are based on heuristics or mathematical programming techniques which are either unable or requiring expensive time to obtain exact solutions for problems of large size. Website masters or owners prefer to obtain a feasible and quality, although not optimal, website structure within reasonable time. We thus design a metaheuristic approach for obtaining a near-optimal website structure that improves the usage. *Third*, the development of metaheuristics theory has been flourishing during the last decade. Among others, tabu search (TS) [9,10] is one of the most effective metaheuristics for solving the quadratic assignment problem (QAP) which has been previously adopted to formulate WSO. Therefore, in this paper we devise an enhanced tabu search (ETS) facilitated with sophisticated features including multiple neighborhoods, adaptive tabu lists, dynamic tabu tenure, and multi-level aspiration criteria, such that better solutions can be discovered. Experimental results on a real-world dataset containing 24 WSO problem instances show that the proposed ETS approach outperforms a competing method based on genetic algorithm.

The remainder of this paper is organized as follows. Section 2 reviews the related works for WSO. Section 3 presents the proposed method including user session identification, web usage evaluation, problem formulation, and an enhanced tabu search. Section 4 reports the experimental results. Finally, Section 5 concludes this work.

2. Related works

In this section we focus our discussion on existing works for adaptive websites which reorganize the link structure to improve the web usage. The discussion covers several aspects. First, the click-through data contained in the web log should be analyzed to extract prevailing path traversal patterns. Second, several web usage evaluation functions are introduced. Third, the web structure constraints considered in business practice are detailed. Fourth, existing optimization approaches to the WSO problem are discussed.

Web server logs contain tremendous amount of click stream incurred from daily user browsing activities. In order to extract meaningful and remarkable click-through traversal patterns, most existing works analyze the web server logs and disclose useful knowledge. For example, Fu et al. [11] identified the user session by reference to the IP address, browser software and operating system software facilitated at the client sites. The click stream is then divided into user sessions by using the time window technique. The click frequency of each direct link between webpages can be easily derived by scanning the user sessions. If necessary, the click frequency of traversal paths in a website can be calculated using the same technique.

One of the mostly used measures for evaluating the web usage is the number of clicks needed for accessing the target webpage [12,13]. With the frequency information associated with all of the traversal paths contained in the original website, the web usage can be improved by optimizing the website structure such that the most intensively visited traversal paths can be made shorter. Other criteria for evaluating the web usage include the download time of the visited webpages [14] and the probability of selling products [4] after the reorganization of the website structure.

In addition to the objective for improving the web usage, there are a number of web design constraints entailed by business practice. We summarize these constraints and the related references in Table 1 and give brief descriptions as follows.

- *Connectivity constraint.* The reorganization should not break the connectivity of the website structure, i.e., any page in the website is still accessible from the home page after the reorganization.
- *Out-degree constraint.* The number of outward hyperlinks from a webpage should be constrained by an upper bound. Containing too many outward hyperlinks would cause information overload and hinder the browsing fluency of the user.
- *Website depth constraint.* In order to prevent the user from losing interest in searching the target webpage, the length of the shortest path between the home page and the target webpage is constrained by a limit. In general, this constraint can be enforced by stipulating a maximal depth for the website hierarchy.
- *Basic link constraint. Some* basic links between specified webpages are essential to maintain the service logics deemed by the website design. The basic links are thus assumed permanent and are not removable from the website structure.
- *Page classification constraint.* Pages contained in a website can be classified into different categories, and each category of pages may require a different degree on a certain constraint. For instance, index pages usually have more outward links than content pages, and transactional pages may have more strict access control than advertisement pages.
- *Security constraint.* Some websites offer personal managerial pages which are only accessible by using the assigned user account. For security reason, these managerial pages are only accessed through the log-in pages with correct user account and password.