Chapter 3.11
Adaptive Web Presence and Evolution through Web Log Analysis

Xueping Li
University of Tennessee, USA

INTRODUCTION

The Internet has become a popular medium to disseminate information and a new platform to conduct electronic business (e-business) and electronic commerce (e-commerce). With the rapid growth of the WWW and the intensified competition among the businesses, effective web presence is critical to attract potential customers and retain current customer thus the success of the business. This poses a significant challenge because the web is inherently dynamic and web data is more sophisticated, diverse, and dynamic than traditional well-structured data. Web mining is one method to gain insights into how to evolve the web presence and to ultimately produce a predictive model such that the evolution of a given web site can be categorized under its particular context for strategic planning. In particular, web logs contain potentially useful information and the analysis of web log data have opened new avenues to assist the web administrators and designers to establish adaptive web presence and evolution to fit user requirements.

BACKGROUND

People have realized that web access logs are a valuable resource for discovering various characteristics of customer behaviors. Various data mining or machine learning techniques are applied to model and understand the web user activities (Borges and Levene, 1999; Cooley et al., 1999; Kosala et al., 2000; Srivastava et al., 2000; Nasraoui and Krishnapuram, 2002). The authors in (Kohavi, 2001; Mobasher et al., 2000) discuss the pros and cons of mining the e-commerce log data. Lee and Shiu (Lee and Shiu, 2004) propose an adaptive website system to automatically change the website architecture according to user browsing activities and to improve website usability from the viewpoint of efficiency. Recommendation systems are used by an ever-increasing number of e-commerce sites to help consumers find products to purchase (Schafer et al, 2001). Specifically, recommendation systems analyze the users' and communities' opinions and transaction history in order to help individuals identify products that are most likely to be relevant to their preferences (e.g. Amazon.com, eBay.com). Besides web mining technology, some researches investigate on Markov chain to model the web user access behavior (Xing et al., 2002; Dhyani et al., 2003; Wu et al., 2005). Web log analysis is used to extract terms to build web page index, which is further combined with text-based and anchor-based indices to improve the performance of the web site search (Ding and Zhou, 2007). A genetic algorithm is introduced in a model-driven decision-support system for web site optimization (Asllani and Lari, 2007). A web forensic framework as an alternative structure for clickstream data analysis is introduced for customer segmentation development and loyal customer identification; and some trends in web data analysis are discussed (Sen et al., 2006).

MAIN FOCUS

Broadly speaking, web log analysis falls into the range of web usage mining, one of the three categories of web mining (Kosala and Blockeel, 2000; Srivastava et al., 2002). There are several steps involved in web log analysis: web log acquisition, cleansing and preprocessing, and pattern discovery and analysis.

Web Log Data Acquisition

Web logs contain potentially useful information for the study of the effectiveness of web presence. Most websites enable logs to be created to collect the server and client activities such as access log, agent log, error log, and referrer log. Access logs contain the bulk of data including the date and time, users' IP addresses, requested URL, and so on. Agent logs provide the information of the users' browser type, browser version, and operating system. Error logs provide problematic and erroneous links on the server such as "file not found", "forbidden to access", et al. Referrer logs provide information about web pages that contain the links to documents on the server.

Because of the stateless characteristic of the Hyper Text Transfer Protocol (HTTP), the underlying protocol used by the WWW, each request in the web log seems independent of each other. The identification of user sessions, in which all pages that a user requests during a single visit, becomes very difficulty (Cooley et al., 1999). Pitkow (1995, 1997, 1998) pointed out that local caching and proxy servers are two main obstacles to get reliable web usage data. Most browsers will cache the recently pages to improve the response time. When a user clicks the "back" button in a browser, the cached document is displayed instead of retrieving the page from the web server. This process can not be recorded by the web log. The existence of proxy servers makes it even harder

6 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/adaptive-web-presence-evolution-through/54800

Related Content

The Effects of Culture on Planning and Technology in the Internationalisation of Medellín's SMEs
Luis Bustamante (2013). *Cultural and Technological Influences on Global Business (pp. 365-383).*
www.irma-international.org/chapter/effects-culture-planning-technology-internationalisation/76499

The Effect of Reciprocity on Mobile Wallet Intention: A Study of Filipino Consumers
Donald Amoroso, Ricardo Limand Francisco L. Roman (2021). *International Journal of Asian Business and Information Management (pp. 57-83).*
www.irma-international.org/article/the-effect-of-reciprocity-on-mobile-wallet-intention/274294

Women Entrepreneurship in Gulf Region: Challenges and Strategies In GCC
Viju Mathew (2019). *International Journal of Asian Business and Information Management (pp. 94-108).*
www.irma-international.org/article/women-entrepreneurship-in-gulf-region/216922

Cointegration and Causality Study Among Inward FDI, Economic Growth and Exports: An Indian Perspective
Jekka Chandrasekaran Sharmiladevi (2020). *International Journal of Asian Business and Information Management (pp. 63-77).*
www.irma-international.org/article/cointegration-and-causality-study-among-inward-fdi-economic-growth-and-exports/241927

National Intellectual Capital Stocks and Organizational Cultures: A Comparison of Lebanon and Iran
Jamal A. Nazari, Irene M. Herremans, Armond Manassianand Robert G. Isaac (2011). *Global Business: Concepts, Methodologies, Tools and Applications (pp. 2433-2456).*
www.irma-international.org/chapter/national-intellectual-capital-stocks-organizational/54910