Chapter 14

Web Usage Mining Approaches for Web Page Recommendation: A Survey

H. Hannah Inbarani
Periyar University, India

K. Thangavel
Periyar University, India

ABSTRACT

The technology behind personalization or Web page recommendation has undergone tremendous changes, and several Web-based personalization systems have been proposed in recent years. The main goal of Web personalization is to dynamically recommend Web pages based on online behavior of users. Although personalization can be accomplished in numerous ways, most Web personalization techniques fall into four major categories: decision rule-based filtering, content-based filtering, and collaborative filtering and Web usage mining. Decision rule-based filtering reviews users to obtain user demographics or static profiles, and then lets Web sites manually specify rules based on them. It delivers the appropriate content to a particular user based on the rules. However, it is not particularly useful because it depends on users knowing in advance the content that interests them. Content-based filtering relies on items being similar to what a user has liked previously. Collaborative filtering, also called social or group filtering, is the most successful personalization technology to date. Most successful recommender systems on the Web typically use explicit user ratings of products or preferences to sort user profile information into peer groups. It then tells users what products they might want to buy by combining their personal preferences with those of like-minded individuals. However, collaborative filtering has limited use for a new product that no one has seen or rated, and content-based filtering to obtain user profiles might miss novel or surprising information. Additionally, traditional Web personalization techniques, including collaborative or content-based filtering, have other problems, such as reliance on subject user ratings and static profiles or the inability to capture richer semantic relationships among Web objects. To overcome these shortcomings, the new Web personalization tool, nonintrusive personalization, attempts to increasingly incorporate Web usage mining techniques. Web usage mining can help improve the scalability, accuracy, and flexibility of recommender systems. Thus, Web usage mining can reduce the need for obtaining subjective user ratings or registration-based personal preferences. This chapter provides a survey of Web usage mining approaches.

DOI: 10.4018/978-1-4666-2542-6.ch014
Web Usage Mining Approaches for Web Page Recommendation

INTRODUCTION

Web mining is the use of data mining techniques to automatically discover and extract information from World Wide Web documents and services. Web mining is a technique to discover and analyze the useful information from the Web data. Web mining can be defined roughly as data mining using data generated by the Web and includes the following sub areas: Web content mining, Web usage mining, and Web structure mining (Srivastava, et al., 2000). In Web Content Mining (WCM) useful information is extracted from the content of Web pages (Pal, et al., 2002), as e.g. free text inside a Web page, semi-structured data such as HTML code, pictures, and downloadable files. Web Structure Mining (WSM) aims at generating a structural summary about the Web site and Web pages. While Web content mining mainly focuses on the structure of inner document, Web structure mining tries to discover the link structure of the hyperlinks at the inter document level. Web Usage Mining (WUM) is applied to the data generated by visits to a Web site, especially those contained in Web log files. Other sources can be browser logs, user profiles, user sessions, bookmarks, folders and scrolls (Pal, et al., 2002; Arayaa, et al., 2004). Personalization of Web sites is a very challenging field of both, current research as well as applications that have as goals e.g. individualized marketing for E-Commerce or dynamic recommendations to a Web visitor based on his/her profile and usage behavior. Analyzing Web data can also be used for system improvements providing the key to understanding Web traffic behavior. Advanced load balancing, data distribution or policies for Web caching as well as higher security standards are potential benefits of such improvements. Similar analyses could be used for modification of Web sites. Understanding visitors’ behavior in a Web site provides hints for adequate design and update decisions. Business intelligence covers the application of intelligent techniques in order to improve certain businesses, mainly in marketing.

The rest of this chapter is organized as follows: section 2 discusses the applications of Web usage mining techniques, section 3 presents a detailed review of various association rule mining techniques, section 4 provides a review of sequential pattern mining techniques for Web personalization, section 5 provides a detailed review of various clustering techniques for Web page recommendation, section 6 presents a review of graph theoretic modeling techniques for Web personalization, section 7 presents a review of Latent variable model for Web personalization, and section 8 provides a review of bi-clustering techniques for Web personalization in the literature.

APPLICATIONS OF WEB USAGE MINING

Personalization

Web Personalization simply means to understand the needs and interests of the visitors of the site and respond accordingly. Such a Web site recognizes each visitor and customizes itself by various ways such as determining the information that should be shown to the visitor or automatically changing the site structure in a way that will be useful and attractive for the current user. Personalization is an attractive research topic, because it is critically important for the success of e-commerce companies.

Making dynamic recommendations to a Web user, based on her/his profile in addition to usage behavior is very attractive to many applications, e.g. cross sales and up-sales in e-commerce. Web usage mining is an excellent approach for achieving this goal, as illustrated in Cooley et al. (1999). The WebWatcher (Joachims, et al., 1997), SiteHelper (Ngu, et al., 1997), Letizia (1995) and Yan et al. (1996) have all concentrated on providing Web site personalization based on usage information. The system proposed in Yan et
16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:

www.igi-global.com/chapter/web-usage-mining-approaches-web/71916?camid=4v1


www.igi-global.com/e-resources/library-recommendation/?id=1

Related Content

Walking the Information Overload Tightrope
www.igi-global.com/chapter/walking-information-overload-tightrope/56223?camid=4v1a

Cooperation Protocol Design Method for Repository-Based Multi-Agent Applications
Wenpeng Wei, Hideyuki Takahashi, Takahiro Uchiya and Tetsuo Kinoshita (2013). International Journal of Software Science and Computational Intelligence (pp. 1-14).
www.igi-global.com/article/cooperation-protocol-design-method-for-repository-based-multi-agent-applications/101315?camid=4v1a

Computational Intelligence for Information Technology Project Management
www.igi-global.com/chapter/computational-intelligence-information-technology-project/56216?camid=4v1a

Modeling and Analysis for Production Performance: Analysis of U.S. Manufacturing Companies
Purnendu Mandal and Enrique (“Henry”) Venta (2008). Handbook of Computational Intelligence in Manufacturing and Production Management (pp. 454-474).
www.igi-global.com/chapter/modeling-analysis-production-performance/19372?camid=4v1a