

Chapter 2.23

Integrating Semantic Knowledge with Web Usage Mining for Personalization

Honghua Dai

DePaul University, USA

Bamshad Mobasher

DePaul University, USA

ABSTRACT

Web usage mining has been used effectively as an approach to automatic personalization and as a way to overcome deficiencies of traditional approaches such as collaborative filtering. Despite their success, such systems, as in more traditional ones, do not take into account the semantic knowledge about the underlying domain. Without such semantic knowledge, personalization systems cannot recommend different types of complex objects based on their underlying properties and attributes. Nor can these systems possess the ability to automatically explain or reason about the user models or user recommendations. The integration of semantic knowledge is, in fact, the primary challenge for the next generation of personalization systems. In this chapter we provide an overview of approaches for incorporating semantic knowledge into Web usage mining and personalization processes. In particular, we

discuss the issues and requirements for successful integration of semantic knowledge from different sources, such as the content and the structure of Web sites for personalization. Finally, we present a general framework for fully integrating domain ontologies with Web usage mining and personalization processes at different stages, including the preprocessing and pattern discovery phases, as well as in the final stage where the discovered patterns are used for personalization.

INTRODUCTION

With the continued growth and proliferation of e-commerce, Web services, and Web-based information systems, personalization has emerged as a critical application that is essential to the success of a Website. It is now common for Web users to encounter sites that provide dynamic recommendations for products and services, tar-

geted banner advertising, and individualized link selections. Indeed, nowhere is this phenomenon more apparent as in the business-to-consumer e-commerce arena. The reason is that, in today's highly competitive e-commerce environment, the success of a site often depends on the site's ability to retain visitors and turn casual browsers into potential customers. Automatic personalization and recommender system technologies have become critical tools, precisely because they help engage visitors at a deeper and more intimate level by tailoring the site's interaction with a visitor to her needs and interests.

Web personalization can be defined as any action that tailors the Web experience to a particular user, or a set of users (Mobasher, Cooley & Srivastava, 2000a). The experience can be something as casual as browsing a Website or as (economically) significant as trading stocks or purchasing a car. Principal elements of Web personalization include modeling of Web objects (pages, etc.) and subjects (users), categorization of objects and subjects, matching between and across objects and/or subjects, and determination of the set of actions to be recommended for personalization. The actions can range from simply making the presentation more pleasing to anticipating the needs of a user and providing customized information.

Traditional approaches to personalization have included both content-based and user-based techniques. Content-based techniques use personal profiles of users and recommend other items or pages based on their content similarity to the items or pages that are in the user's profile. The underlying mechanism in these systems is usually the comparison of sets of keywords representing pages or item descriptions. Examples of such systems include Letizia (Lieberman, 1995) and WebWatcher (Joachims, Freitag & Mitchell, 1997). While these systems perform well from the perspective of the end user who is searching the Web for information, they are less useful in e-commerce applications, partly due to the lack

of server-side control by site owners, and partly because techniques based on content similarity alone may miss other types of semantic relationships among objects (for example, the associations among products or services that are semantically different, but are often used together).

User-based techniques for personalization, on the other hand, primarily focus on the similarities among users rather than item-based similarities. The most widely used technology user-based personalization is collaborative filtering (CF) (Herlocker, Konstan, Borchers & Riedl, 1999). Given a target user's record of activity or preferences, CF-based techniques compare that record with the historical records of other users in order to find the users with similar interests. This is the so-called *neighborhood* of the current user. The mapping of a visitor record to its neighborhood could be based on similarity in ratings of items, access to similar content or pages, or purchase of similar items. The identified neighborhood is then used to recommend items not already accessed or purchased by the active user. The advantage of this approach over purely content-based approaches that rely on content similarity in item-to-item comparisons is that it can capture "pragmatic" relationships among items based on their intended use or based on similar tastes of the users.

The CF-based techniques, however, suffer from some well-known limitations (Sarwar, Karypis, Konstan & Riedl, 2000). For the most part these limitations are related to the scalability and efficiency of the underlying algorithms, which requires real-time computation in both the neighborhood formation and the recommendation phases. The effectiveness and scalability of collaborative filtering can be dramatically enhanced by the application of Web usage mining techniques.

In general, *Web mining* can be characterized as the application of data mining to the content, structure, and usage of Web resources (Cooley, Mobasher & Srivastava, 1997; Srivastava, Cooley, Deshpande & Tan, 2000). The goal of Web min-

26 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/integrating-semantic-knowledge-web-usage/22283

Related Content

Digital Inclusion Programs in South America

María Gladys Ceretta and Javier Canzani (2016). *Handbook of Research on Comparative Approaches to the Digital Age Revolution in Europe and the Americas* (pp. 444-458).

www.irma-international.org/chapter/digital-inclusion-programs-in-south-america/138049

The Moderating Effect of Organizational Safety Climate on Text Message Use and Work-Related Accidents: An Organizational-Level Investigation

Brian E. Kufner and Laura E. Plybon (2012). *International Journal of Social and Organizational Dynamics in IT* (pp. 52-67).

www.irma-international.org/article/moderating-effect-organizational-safety-climate/76385

Digital Divide: Introduction to the Problem

Liudmila Burtseva, Svetlana Cojocar and Constantin Gaidric (2007). *Information Communication Technologies and Human Development: Opportunities and Challenges* (pp. 57-76).

www.irma-international.org/chapter/digital-divide-introduction-problem/22619

Exploring College Students' Deeper Learning Perceptions in the Blended Learning Environment: Scale Development, Validation, and Experimental Comparison

Dan-Dan Shen and Chiung-Sui Chang (2022). *International Journal of Technology and Human Interaction* (pp. 1-21).

www.irma-international.org/article/exploring-college-students-deeper-learning-perceptions-in-the-blended-learning-environment/313184

The Contribution of Different Body Channels to the Expression of Emotion in Animated Pedagogical Agents

Saikiran Anasingaraju, Nicoletta Adamo-Villani and Hazar Nicholas Dib (2020). *International Journal of Technology and Human Interaction* (pp. 70-88).

www.irma-international.org/article/the-contribution-of-different-body-channels-to-the-expression-of-emotion-in-animated-pedagogical-agents/261216