



Chapter 7

A Survey of Web-Usage Mining: Techniques for Building Web-Based Adaptive Hypermedia Systems

Martha Koutri, University of Patras, Greece

Nikolaos Avouris, University of Patras, Greece

Sophia Daskalaki, University of Patras, Greece

Abstract

This chapter discusses Web usage mining techniques that can be applied for building adaptive hypermedia systems. These techniques are used for uncovering hidden patterns within Web access data and then for building the user model that lies in the heart of each adaptive system. Web access data, traditionally stored in the server log files, constitute a rich source of data collected in a non-intrusive way that guards the privacy of users. Several Web usage mining approaches have been proposed for exposing usage patterns, with the most prominent ones being cluster mining, association rule mining, and sequential pattern mining. This chapter provides an overview of the state of the art in research of Web usage mining, and discusses the most relevant criteria for deciding on the suitability of these techniques for building an adaptive Web site. Moreover, the different types of patterns revealed from Web usage mining are correlated with different adaptation aspects.

Introduction

The Web constitutes a huge, widely distributed and dynamically growing hypermedium, supporting access to information and services. Given the diversity of Web users, a desired characteristic of many Web systems is to personalise user interaction for tasks like navigation and information retrieval. Web-based hypermedia systems that take into consideration certain user characteristics in order to guide user interaction are referred as *adaptive Web systems*. The study and development of adaptive Web systems constitute an area of active research that stimulates the interest of different research communities, such as user modelling, machine learning, data mining, human-computer interaction, and so forth. Given the availability of a large volume of user data collected over the Web and the difficulty to extract user characteristics directly from the users, Web usage mining techniques play a vital role in adaptive Web systems research and practice (Webb, Pazzani, & Billsus, 2001).

The main objective of this chapter is to provide a comprehensive overview of Web usage mining techniques used for building adaptive Web-based hypermedia systems. Therefore, the focus is on the main approaches used and their effectiveness in various adaptive Web systems. These approaches bear different characteristics and can be used in a variety of Web systems as discussed in the following sections.

Web User Modelling

The *user model (UM)* constitutes a basic component of every adaptive Web system. The UM represents user characteristics that enable the system to distinguish among different users and is built using data that are requested either directly by the users or obtained by logging user interactions. Different types of user data can be used for building the UM. For example:

- user characteristics (age, gender, location, etc.);
- user preferences and interests;
- user knowledge and skills; and
- user behavioural patterns.

While certain information may be derived directly from data, other times it is inferred after processing and interpreting usage data. User models may refer to either individual users or communities of users. In the latter case, each user is considered to be a member of a generic user community, and an *aggregate user model* is used for describing the different communities of users.

A number of adaptive systems have been proposed for recommending Web documents relevant to users' interests (Armstrong, Freitag, Joachims, & Mitchell, 1995; Lieberman,

23 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/survey-web-usage-mining/4182

Related Content

Computer Simulations and Scientific Knowledge Construction

Athanassios Jimoyiannis (2011). *Gaming and Simulations: Concepts, Methodologies, Tools and Applications* (pp. 57-74).

www.irma-international.org/chapter/computer-simulations-scientific-knowledge-construction/49374

Our Future: With the Good, the Bad, or the Ugly eServices? Case Finland

Maija R. Korhonen (2018). *Digital Multimedia: Concepts, Methodologies, Tools, and Applications* (pp. 1541-1559).

www.irma-international.org/chapter/our-future/189541

Leadership Competencies for Managing Global Virtual Teams

Diana J. Wong-Mingji (2008). *Multimedia Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 1303-1310).

www.irma-international.org/chapter/leadership-competencies-managing-global-virtual/27157

OTT Platforms From the Perspective of Critical Algorithm Studies: The Algorithmic Paradox of the Audience

Elif Karakoç Keskin (2024). *The Rise of Over-the-Top (OTT) Media and Implications for Media Consumption and Production* (pp. 95-113).

www.irma-international.org/chapter/ott-platforms-from-the-perspective-of-critical-algorithm-studies/337668

An H.264/AVC Error Detection Algorithm Based on Syntax Analysis

Luca Superiori, Olivia Nemethovaand Markus Rupp (2009). *Multimedia Transcoding in Mobile and Wireless Networks* (pp. 215-234).

www.irma-international.org/chapter/264-avc-error-detection-algorithm/27203