The Journal of Systems and Software 86 (2013) 2995–3012

Contents lists available at ScienceDirect

The Journal of Systems and Software

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

Modeling users on the World Wide Web based on cognitive factors, navigation behavior and clustering techniques

Marios Belk*, Efi Papatheocharous, Panagiotis Germanakos, George Samaras

Department of Computer Science, University of Cyprus, Nicosia, Cyprus

ARTICLE INFO

Article history:
Received 14 April 2012
Received in revised form 8 March 2013
Accepted 15 April 2013
Available online 7 May 2013

Keywords:
User modeling
Cognitive styles
Web navigation behavior

ABSTRACT

This paper focuses on modeling users' cognitive styles based on a set of Web usage mining techniques on user navigation patterns and clickstream data. Main aim is to investigate whether specific clustering techniques can group users of particular cognitive style using measures obtained from psychometric tests and content navigation behavior. Three navigation metrics are proposed and utilized to find identifiable groups of users that have similar navigation patterns in relation to their cognitive style. The proposed work has been evaluated with two user studies which entail a psychometric-based survey for extracting the users' cognitive styles, combined with a real usage scenario of users navigating in a controlled Web 2.0 environment. A total of 106 participants of age between 17 and 25 participated in the study providing interesting insights with respect to cognitive styles and navigation behavior of users. Studies like the reported one can be useful for modeling users and assist adaptive Web 2.0 environments to organize and present information and functionalities in an adaptive format to diverse user groups.

1. Introduction

During the last decade, Web 2.0 has ingrained itself into everyday life and has contributed to the exponential increase of internet usage since the early 2000s. Its fundamental concept as a medium for collaboration and sharing of information has generated extensive enthusiasm driving many of the world's markets. Within this realm, adaptivity considerations in modern Web-based interactive systems are considered of paramount importance as it is known that traditional static interactive Web-based systems treat all users the same way, being unable to satisfy the heterogeneous needs and preferences of users (Brusilovsky, 2001). This often results in users experiencing orientation difficulties in the information space or being overwhelmed by redundant information. To elaborate on this by giving examples, a static electronic encyclopedia usually offers the same functionality and content to readers with different knowledge and interests. Similarly, a traditional e-commerce Web-site offers the same product catalog and tools to customers with different needs and preferences. Finally, a traditional educational hypermedia system presents the same static content to students with widely differing educational goals and educational background.

In this respect, adapting the functionality and content, of any Web-based interactive system, to satisfy the users' needs and increase their level of understandability and acceptability in an intuitive manner and empower them to complete specific tasks more efficiently and effectively is a challenging endeavor. It entails understanding and modeling human behavior for diverse user groups, with regard to structural and functional user requirements, which needs to be translated into usable computer–human interaction designs and workflows, whilst minimizing the overall users' cognitive, perceptual and learning load. Adaptive interactive systems (Brusilovsky et al., 2007; Brusilovsky, 2001; Bra et al., 2000; Schneider-Hufschmidt et al., 1993) provide an alternative to the "one size fits all" approach of static user interfaces by adapting the interactive system's structure, terminology, functionality and presentation of content to users' perceptions, needs and preferences, aiming to increase the usability of the interface and provide an improved user experience.

One of the distinctive features of an adaptive interactive system is its user model. The user model is a representation of static and dynamic information about an individual that is utilized by the adaptive interactive system aiming to provide adaptation effects (i.e., the same system can look different to users with different user models) (Brusilovsky and Millán, 2007; Frias-Martinez et al., 2005). To better explain this through some examples, an information retrieval system may adaptively select and prioritize the most relevant items to the user's goals and/or interests. An educational hypermedia system may provide adaptive navigation support by manipulating the links based on the user's knowledge and

* Corresponding author. Tel.: +357 22 892675.
E-mail addresses: belk@cs.ucy.ac.cy (M. Belk), efi.papatheocharous@cs.ucy.ac.cy (E. Papatheocharous), pgerman@cs.ucy.ac.cy (P. Germanakos), cssamara@cs.ucy.ac.cy (G. Samaras).

0164-1212/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jss.2013.04.029

learning goals. A security and privacy-preserving related mechanism in a commercial Web system may present the content adaptively to the user's level of knowledge toward security terms (e.g., provide novice users with personalized security information awareness by using simplified security terms and additional explanations).

The mechanism used for user modeling can be based on explicit or implicit information gathering approaches. Explicit information is provided directly by the user, usually through Web registration forms, questionnaires, or specially designed psychometric instruments. On the other hand, implicit information is extracted by the system automatically to infer characteristics about the user and is usually obtained by tracking the user's navigation behavior by interacting with the system. For example, such implicit information can be extracted based on the time spent on a particular Web-page by a user, which can be used to infer the interest of the user toward the main subject of that Web-page. Various research works have attempted to investigate the most effective source of information for user modeling (Gauch et al., 2007; Jawaheer et al., 2010; Wærn, 2004). Based on Gauch et al. (2007) it is yet not clear-cut whether implicitly created models are more or less accurate than explicitly created models. Nevertheless, since implicit information gathering does not affect the human computer interaction or the users' cognitive load (Gauch et al., 2007), it seems to be the preferable approach for collecting information about users. On the other hand, this approach is much more complex than explicit user feedback since in most cases the data obtained may be imprecise, incomplete and/or heterogeneous.

An increasing interest, apart from research-oriented systems, has been observed during the last few years from major commercial Web 2.0 service providers such as Google (http://www.google.com), Bing (http://www.bing.com), and Amazon (http://www.amazon.com) that provide personalized results and recommendations, by employing various user modeling and adaptation techniques. In this respect, the notion of personalization has finally found its way in users' everyday interactions in Web interactive systems. Nevertheless, there is a need for more research on measuring the actual benefit for users, instead of merely developing elaborated user modeling and adaptation techniques. Even if personalization is the key to more efficient interactions and enhancing user experience, there is one undeniable issue to be resolved: how and why could users benefit even more? Individuals are certainly different from each other, but which would be the underlying theories that could guide research endeavors in producing measurable gains? A first approach would be to identify the levels in which individuals demonstrate considerable differences, such as demographics, social, cognitive and mental abilities, personality, goals, needs, and experiences, and to build a cohesive user model by including characteristics that could be important in adapting Web 2.0 interactive systems, with the aim to improve the user experience. Since content and functionality of Web 2.0 interactive systems is either presented visually or verbally, and users may have specific navigation behavior, e.g., holistic or analytic approach of navigation, this work suggests that cognitive styles of users, which describe the way individuals, perceive and organize information, might be applied effectively on designing adaptive Web 2.0 interactive systems. In this context, a Web environment's information presentation and structure adapted to users' cognitive style may increase the usability of a system in terms of efficiency and effectiveness, as well as provide a positive user experience (Germanakos et al., 2008).

To this end, the work presented is primarily driven by the need to apply User-Centered Design (UCD) methodologies related to the design and evaluation of adaptive mechanisms, and contributes toward this direction by proposing a user modeling approach for highlighting cognitive styles of users with explicit and implicit information gathering approaches by utilizing data analysis techniques. In particular, a specific psychometric measurement is used to highlight differences in cognitive styles of users, combined with data analysis techniques applied on the users' navigation patterns. Main objectives of the paper are to: (i) study the relation between users' cognitive styles and navigation behavior, (ii) investigate whether specific data analysis techniques can group users of particular cognitive style using measures obtained from psychometric tests, and (iii) propose navigation content metrics to find identifiable groups of users that have similar navigation patterns. The identification of users with specific cognitive and navigation style will ultimately help in defining various adaptation mechanisms which are required to be assembled to target a different user interface experience in Web-based environments for various cognitive typologies of users.

Overarching aim is to drive this research toward the design and development of a comprehensive adaptive interactive system which will be conceptually composed of two interconnected components; the user modeling and the adaptation component. The user modeling component will entail data about its users and their interactions. This information will be provided to the system either explicitly by the users (e.g., through registration forms or psychometric tests, etc.) and/or implicitly retrieved through user's interactions with the system aiming to enrich the user model and to infer information which is considered valuable in order to provide adaptive features (e.g., track the users' navigation behavior, and further infer their cognitive style). In this context, various data and cluster analysis techniques will be performed on the raw data acquired in order to generate the actual user models which, combined with various decision making and adaptation mechanisms, decide on the adaptation effects to be performed that will be further communicated to the adaptive user interface (e.g., in case users have a holistic approach in information organization and follow a linear approach in navigation, then provide personalized navigation tools that will assist the users process and organize information more efficiently and effectively).

The paper is organized as follows. In Section 2, we provide an overview of user modeling mechanisms giving emphasis on explicit and implicit gathering approaches, and cognitive-based user modeling systems. In Section 3 we present a user study based on the proposed approach. Consequently, we conclude the paper and describe our directions of future work in Section 4.

2. User modeling mechanisms

The ability of adaptation in interactive systems heavily depends on successful user modeling. A user model is created through a user modeling mechanism in which unobservable information about a user is inferred from observable information from that user (Frias-Martinez et al., 2005), for example, using the interactions with the system (i.e., time being active on a Web-page, buying history, ratings of products, bookmarked or saved content, etc.). User models can be created utilizing explicit information from the user, i.e., user guided modeling, and/or implicit information, i.e., dynamic user modeling (Gauch et al., 2007; Frasincar et al., 2004; Houben et al., 2004). These two categories are discussed in the next sections.

2.1. User guided modeling

User guided modeling methodologies rely on personal information provided by the users, typically via registration forms. The data collected usually contain demographic information (i.e., age, gender, and profession), interests and/or preferences. Common techniques for obtaining explicit information that allows specification of the users' model include the use of checkboxes, drop-down