

## **Preface to the special issue on data mining for personalization**

**Bamshad Mobasher · Alexander Tuzhilin**

Received: 28 September 2008 / Accepted: 28 September 2008 / Published online: 11 November 2008  
© Springer Science+Business Media B.V. 2008

The continued growth and proliferation of e-commerce, online services, and Web-based information systems, has led to an explosion of available content and resources in digital form. As a result of user interactions with these resources, tremendous volumes of clickstream, transaction, and user data are collected by organizations in their daily operations. Mining and analyzing such data can help these organizations determine the life time value of clients, design cross marketing strategies across products and services, evaluate the effectiveness of targeted promotional campaigns, and thus provide more effective personalized content to visitors. Automatic personalization is a central technology used in user-adaptive systems to deliver dynamic content, such as textual information, links, advertisements, and product recommendations that are tailored to the needs and interests of a particular user or a segment of users.

The process of personalization comprises iterative invocations of the personalization cycle consisting of the stages of understanding the user and building user profiles and models; delivering personalized offerings to the users based on this understanding; measuring impacts of these offerings on the users; and adjusting all the previous stages based on the measured feedback. Data mining is one of the key technologies used in various stages of this personalization process, especially in building user models and delivering personalized offerings to the users. Therefore, it is crucial to develop appropriate data mining methods fitting the corresponding personalization tasks well.

This special issue of User Modeling and User-Adapted Interaction explores recent developments and applications of data mining and machine learning techniques in

---

B. Mobasher (✉)

School of Computing, DePaul University, 243 S. Wabash Ave., Chicago, IL 60604, USA  
e-mail: mobasher@cs.depaul.edu

A. Tuzhilin

Stern School of Business, New York University, 44 West 4th Street, Room 8-92, New York, NY 10012, USA

various aspects of personalization, user modeling, and user adaptive systems, including personalized or adaptive Web applications, e-commerce recommender systems, and intelligent personalized search. It contains five papers that have undergone an extensive reviewing process and cover various contributions of data mining to interface personalization.

The first paper by Sofia Stamou and Alexandros Ntoulas describes an approach to search personalization that utilizes both semantic domain information in the form of a concept hierarchy as well as user preference information by mining user queries and click-through data. The authors leverage a topical ontology for estimating users' topic preferences based on past queries and click-throughs on query results. A ranking function uses the discovered preference models in order to rank the search results to better match the interests of a given user. The type of technology described in this paper exemplifies an important trend in the new generation of intelligent search tools that attempt to personalize search results by taking into account automatically discovered preference models and user profiles. The combination of user preference models and semantic domain knowledge may be helpful in better understanding the underlying user search intent.

The paper by Shlomo Berkovsky, Tsvi Kuflik, and Francesco Ricci, focuses on the integration of personalization models across multiple types of recommender systems. The authors suggest that it would be beneficial in general to enrich the user models by integrating user models from other systems, a task they call cross-representation mediation. This conjecture is motivated by the observation that the quality of personalized recommendations often depends mostly on the size and accuracy of the underlying user models. In particular, this work evaluates cross-representation mediation of user models from collaborative filtering to content-based recommender systems: a content-based recommender system, having partial or no user profile data, can generate recommendations for users by mediating profile data of the same users, collected by a collaborative filtering system. The mediation process exploits the item descriptions that are typically not used by the collaborative filtering recommender systems.

Bhaskar Mehta and Wolfgang Nejdl, consider the vulnerabilities of collaborative recommender systems to shilling attacks. This problem has been brought to light in recent years and has significant implications for commercial systems on the Web. Like other social systems, collaborative recommender systems are quite vulnerable to manipulation by malicious users who may have an interest in promoting or demoting an item. Attack profiles can be injected into a collaborative recommender system with the intention of changing the behavior of the system with respect of one or more items, or for a group of users. The paper provides an in-depth analysis of such shilling attacks and describes new approaches to detect attack profiles. In particular, structural similarities in shilling user profiles are taken into account to separate them from normal user profiles utilizing an unsupervised dimensionality reduction approach. Based on these results, a novel collaborative filtering algorithm is proposed that is more robust in the face of shilling attacks.

In recent years, data mining has emerged as a powerful tool in personalized e-learning and web-based adaptive educational systems. Patterns discovered through data mining can be used directly by the teacher or author of the course in order to improve instruction or learning performance. This is the focus of the paper by Enrique

García, Cristóbal Romero, Sebastián Ventura, Carlos de Castro. The paper proposes a system to find and suggest to a teacher the most appropriate modifications to improve the effectiveness of the course. The authors describe an iterative methodology to maintain web-based courses. A critical component of the system involves association rule mining to discover interesting patterns from students' usage data in the form of IF-THEN recommendation rules. This is combined with a collaborative recommender system to share and score the rules obtained by teachers with similar profiles.

Finally, Markus Zanker and Markus Jessenitschnig explore the use of explicit customer requirements in e-commerce recommender systems. In such systems, the goal of personalization is often to suggest useful and interesting items in order to increase user satisfaction and online conversion rates. As in other adaptive systems, e-commerce recommenders utilize explicit or implicit user feedback such as ratings, purchase transactions, or clickstream data as the basis for model construction and recommendation generation. In this paper, however, the authors focus on explicitly formulated customer requirements as the sole type of user feedback. They provide a comparison of different recommendation approaches based on knowledge-based and utility-based methods, collaborative filtering, association rule mining, and some hybrid recommendation algorithms for eliciting explicit customer requirements. The work demonstrates how this type of user feedback can be used for personalization in e-commerce and suggests that explicit customer requirements can be a useful source of feedback for personalization, especially for new users.

Collectively, the five papers in this issue represent a number of new developments and emerging trends in automatic personalization and recommender system. Data mining is the underlying technology that enables the systems and approaches described in these papers to potentially serve as the basis for the next generation of personalized systems on the Web.

### Guest editors' vitae

**Bamshad Mobasher** is a professor of Computer Science and the director of the Center for Web Intelligence at the School of Computing of DePaul University in Chicago. His research areas include Web mining, Web personalization, predictive user modeling, agent-based systems, and information retrieval. He has authored over a hundred articles and edited several books in these areas. Dr. Mobasher received his PhD at Iowa State University in 1994. Prior to DePaul he was an assistant professor of Computer Science at the University of Minnesota, Twin Cities. As the director of the Center for Web Intelligence, Dr. Mobasher is directing research in Web mining, Web analytics, and personalization, as well as overseeing several related joint projects with the industry.

**Alexander Tuzhilin** is a professor of Information Systems and a NEC Faculty Fellow at the Stern School of Business in New York University. He joined Stern in 1989 after earning a Ph.D. degree in Computer Science from the Courant Institute. In addition to NYU, Dr. Tuzhilin has held visiting positions at The Wharton School of the University of Pennsylvania, Columbia University, and Ecole Nationale Supérieure des Télécommunications (Paris, France). He serves on the editorial boards of several journals, including the *IEEE Transactions on Knowledge and Data Engineering*, the *Data Mining and Knowledge Discovery Journal*, as well as on the organizing and program committees of numerous conferences. His research interests include knowledge discovery in databases (data mining), personalization and recommender systems. His industry experience includes two years as a software developer in the high-tech industry and various consulting engagements. His comments on technology-related topics and articles about his work have appeared in several major media publications, including The New York Times, The Wall Street Journal, Business Week, The Financial Times and The Los Angeles Times.