

Personalization on the Net using Web Mining

Maurice D. Mulvenna, Sarabjot S. Anand and Alex G. Büchner

Guest Editors

On the Internet, we have seen a massive growth in systems that can personalize content delivered to Internet users. The science behind personalization has undergone tremendous changes in recent years, yet the basic goal of personalization systems remains the same, which is to provide users with what they want or need without them having to ask for it explicitly.

Personalization is the provision to the individual of tailored products, services, information or information relating to products or service. It is a broad area, also covering recommender systems, customization, and adaptive web sites.

Three aspects of a web site affect its utility in providing the intended service to its users. These are the content provided on the web site, the layout of the individual pages and the structure of the entire web site itself. The relevance of each of the objects comprising a web page to the users' needs will clearly affect their level of satisfaction. The structure of the web site, defined by the existence of links between the various pages restricts the navigation performed by the user to predefined paths and therefore defines the ability of a user to access "relevant" pages with relative ease. The definition of "relevance" is however subjective. It is here that there is a potential mismatch between the perception of what the user needs, on the part of the web site designer, and the true needs of users. This may have a major impact on the effectiveness of a web site.

Personalization technology is software that learns patterns, habits and preferences. On the Internet, its use is primarily in systems that support e-business. Personalization works in this context because it helps users to find solutions, but perhaps more importantly, it also empowers e-business providers with the ability to measure the quality of that solution. In terms of the fast emerging area of Customer Relationship Management (CRM), personalization enables e-business providers to implement strategies to lock-in existing customers, and to win new customers.

Daniel E. O'Leary from the University of Southern California coined the phrase 'AI Renaissance' in 1997 [1], to describe how Artificial Intelligence (AI) can make the Internet more usable. Personalization technology is part of that renaissance.

In parallel with the academic progress covered in this special section, the commercial world is witness to unprecedented growth in personalization technology companies. It is indeed sometimes difficult to find a commonality in technology foundation that spans the breadth of commercial product offerings and global academic efforts in personalization, as well as the broad swathe of emerging 'touchpoints' in digital markets.

Initial attempts at achieving personalization on the Internet have been limited to *check-box* personalization, where portals allow the user to select the links that they would like on their "personal" page. However, this has limited use as it depends on the users knowing *beforehand* the content that is of interest to them. Arguably, collaborative filtering was the first attempt at using AI for achieving personalization in a more "intelligent" manner. This allows users to take advantage of other users behavioural activities based on a measure of similarity between them. These techniques require the user to divulge some personal information on their interests, likes and dislikes, information that many web users would not necessarily wish to divulge.

An alternative, to check-box personalization, is *observational* personalization that attempts to circumvent the need for users to divulge any personal information. The underlying assumption in this approach is that

hidden within logs of a user's previous navigation behaviour, is knowledge that provides clues of how services, products and information need to be personalized for enhanced web interaction.

In our view, there are three principal components to observational personalization: analytics, representation, and deployment. Web Mining provides the tools to analyse web log data in a user centric manner such as segmentation, profiling and clickstream discovery. The knowledge mined by using these tools are increasingly being represented using W3C standards such as XML and the deployment of the knowledge on web servers may be carried out through personalization or recommender systems.

Perhaps for the first time since the beginning of O'Leary's AI renaissance, research and commercial goals are identical. These goals are high performance, high quality analytics, flexible and novel representational schema, and real-time application or deployment of the represented knowledge on web servers, WAP servers, etc.

This special issue contains four articles chosen to reflect various aspects of personalization on the Net using web mining.

Spiliopoulou provides a rationale on why web log data should be mined. The effectiveness of a web site in providing users with the content they need in the most optimized manner is the key to retaining them. Traditional focus group based methods for evaluating the effectiveness of a web sites structure are expensive and difficult to implement as little is known about the underlying population of users of the web site and their needs. Spiliopoulou suggests a process by which mining for navigational patterns may be used to gain insights into a web site's usage and optimality with respect to its current user population.

Cingil, Dogac and Azgin describe the need for interoperability when mining the web and how the various W3C standards can be used to achieve personalization applications. In particular the article looks at how the recent data exchange, meta data and privacy standards from the World Wide Web Consortium (W3C), namely, Extensible Markup Language (XML), Resource Description Framework (RDF) and Platform for Privacy Preferences (P3P) may be used to support personalization activities.

Mobasher, Cooley and Srivastava provide a framework for mining web log files to discover knowledge for the provision of recommendations to current users based on their browsing similarity with previous users. The process for discovering such knowledge includes the gathering and preprocessing of the data necessary for discovering user behaviors, the application of data mining techniques to discover usage patterns, and aggregation and filtering of the data mining results in order to create decision rules for customizing Web site content based on an individual users behavior.

Perkowitz and Etzioni address personalization as a process that adapts an Internet site through the automated generation of index pages for the web site. The article explores adaptive Internet sites: sites that automatically improve their organization and presentation by learning from visitor access patterns. Adaptive web sites mine the data buried in web server logs to produce more easily navigable web sites.

We hope that you find these articles interesting and useful, and that they show the potential and diverse uses available for personalization systems. In February 1999, Gartner Analytics stated that, 'matching direct or inferred reader requests through content personalization will be the most dramatic development in the Internet (...) through 2002, and will help differentiate the Web as a new medium'. Clearly this is an important technology, and is being applied in many Internet systems in use today. As digital markets converge, personalization will be ubiquitous in digital interactive devices, from Personal Digital Assistants (PDAs) through mobile telephony devices to digital TV.

Little attention to date has been paid on evaluating the success of web sites in terms of their utility to the users. Server statistics provided by web log analysis tools provide metrics for evaluating the success of the server in serving pages to users, however, no insights on how useful the content provided was to the user. It is agreed that evaluation of the effectiveness of a web site must be based on the business goal of the e-business and these metrics would provide the basis of evaluating the success of personalization activities. Web Intelligence tools based on web mining have an important role to play in the development of these e-

Metrics. August 2000's ACM KDD conference in Boston topically discusses e-Metrics and the role of web mining in personalization.

The success of personalization on the web depends on the ability of the personalization community in promoting "responsible" use of the technology by e-businesses. A recent survey of web users suggests that over 70%¹ of web users find online solicitation from a web site as a hindrance rather than being helpful and over 50% of web users are worried about their privacy². How the providers of personalized services can manage to tread the fine line between personalization and personal intrusion and control the hype around it will determine the future of this field. Technologically the scene is set to enable businesses to develop, for the first time, equitable relationships with their customer.

References

[1] O'Leary, D., "The Internet, Intranets, and the AI Renaissance", IEEE Computer, January 1997

¹ <http://www.personalization.org>

² These are users that always read the privacy statement on a web site before divulging any personal information