Personalization and Privacy

John Riedl
University of Minnesota

Personalization has been a hot topic for nearly a decade now, and many new products and advanced algorithms have emerged in that time. Several companies now sell tools such as "recommender systems," which take input about users and products and generate recommendations about which products the users will like best. At their best, recommenders can be wonderful tools for users, helping them sort through myriad items they could read, buy, or watch to select those few that are most valuable to them. The algorithms that power these systems have evolved dramatically, and the best can produce rapid recommendations over data sets of millions of users and hundreds of thousands of products.

The other edge of the sword is that recommender systems provide perfect tools for marketers and others to invade users' privacy. After all, recommenders seek to learn everything about our preferences, including what we like to read, what we like to buy, how much money we spend, and what influences us to spend it. How a recommender deals with privacy decides whether its users view it as a boon or a bane. If the recommender only uses this information to help us find items to purchase on a Web site, we will probably value the feature – it might even bring us back to shop there again. On the other hand, if the Web site sells our information to other companies so they can more effectively bother us with phone calls at dinner time, we'll probably feel our privacy has been invaded. Privacy is a critical issue for recommender systems.

In the end, personalization is an important factor in developing effective Web sites because it creates a user experience that is both compelling and sticky. The experience is compelling because it helps users find exactly the information, products, and services they need. It is sticky because a personalized Web site trains itself over time to serve its users better, which makes those users less likely to go to a new site that they would have to train all over again.

Current Research

To realize the importance of personalization we need look no further than Amazon, the grandfather of e-commerce. To

1089-7801/01/$10.00 © 2001 IEEE

The Articles

The November/December 2001 issue of *IEEE Internet Computing* presents four articles exploring the ways in which personalization is changing the way we interact with the Internet. These articles examine the current practice, ongoing research, and social implications of personalization.

In "Inferring User Interest," (pp. 32-39) Claypool et al. demonstrate that implicit measuring of users' actions on the Web pages they visit can correlate closely with their interest in those items. Measuring user behavior has two advantages over asking explicit questions: You can get more data because every user interaction can be mined, and you can get more accurate data because it measures actual actions rather than just what users say.

Herlocker and Konstan present a first step toward combining personalization techniques with short-term user interests in "Task-Focused Recommendation" (pp. 40-47). They suggest that personalization should be tuned to suit the user's particular task or mood at the time of the interaction.

In "The Virtual World Gets Physical: Perspectives on Personalization," (pp. 48-52) McCarthy describes projects he has worked on that use personalization techniques to serve groups of people performing physical tasks in the real world. As computing devices become more embedded in everyday experience, personalization techniques are valuable not just in our virtual world interactions, but in our real-world work and play as well.

In "Privacy Risks in Recommender Systems," (pp. 53-61) Ramakrishnan et al. describe ways in which statistical attacks can be used to gain information about users in recommender systems. These users are often concerned about the data they share and are sometimes reassured by the fact that the information is not personally identifying. However, statistical methods might actually enable sites to match seemingly innocuous preference information to real-world data, such as addresses or phone numbers. During peer review, some reviewers commented that it would be difficult to carry out the proposed attacks, and that the paper does not describe how to stop such attacks if carried out. The authors clarified the risks, and explained that their intent was to identify this new privacy risk in recommender systems rather than to solve it. We thank the reviewers and authors for working together to make the article as clear and valuable as possible.

paraphrase the company's founder, Jeff Bezos: If I have 10 million visitors to my Web site, I should have 10 million Web sites for my visitors! Indeed, in a recent study of the site, we found no fewer than 23 independent applications of personalization. They ranged from simple "people who bought that item also bought" recommendations to deep personalization based on a user's entire history at the site and quick hits based on the last few mouse clicks. Clearly, personalization is a key factor in the company's strategy, and they are exploring how to expand the possibilities.

In my opinion, the top personalization algorithms in use today are already effective and fast enough. The most interesting research is not looking to squeeze out a few more milliseconds of performance or increase scalability by the final 2 percent. Instead, developers are focused on creating new ways to help users.

One especially hot area is in developing systems that use multiple data types. A Web site might track which pages each user visits and how long they remain on them, for instance, and use that data in analyzing the relationship between user interest profiles and which products they actually purchase. Each of these individual pieces of information is in some way flawed as a predictor, but taken together they can give real insight into users' interests and behaviors.

We also need personalization techniques that consider a user's current goals or state of mind. A task-oriented user who needs to find a book, buy it, and get back to work would probably prefer a focused list of recommendations. On the other hand, a personalized library shelf interface might be more appropriate for a user who is simply in the mood to browse. The best sites provide different types of recommenders for the different moods.

The task-oriented recommender might be integrated with a sort feature that could whittle the 1,555 Java books Amazon's search returns down to the top 5 books selected just for you. The personalized library shelf interface might group books based on their similarity to each other, where similarity is computed by combining the Library of Congress number with an understanding of how the current visitor thinks about books.

Most systems have been designed to learn about long-term interests, but until recently they have paid less attention to current activities for cues. If a user is looking for a book to take along on a beach trip, for example, recommending Java texts related to those purchased for work is worse than useless!

From Virtual to Physical Space

Personalization was one of the first Internet technologies to migrate from the Internet to interactions in the physical world. Using nearly the same technology as on the Web, catalog sales centers can also personalize their responses to customers. The

sales center software tells the operator which items to suggest based on each caller's interests. The best of these recommenders can even suggest items that fit with the items the customer is buying right now as well as with long-term interest profiles.

Personalization is reaching into other areas of the physical world through things like in-store kiosks, cash register coupon printers, message boards in public spaces, and wireless PDAs. While these technologies promise benefits like greater convenience, they also highlight important privacy issues. It is impossible to personalize a site without information about users, and the deeper the information, the more personalized the experience can be. On the other hand, there are legitimate concerns about how sites use the detailed data they collect about users' habits and interests. Some systems balance personalization and privacy concerns by tracking only information about preferences. Most users are comfortable giving this information as long as it remains disconnected from their physical selves. E-commerce sites also collect addresses, credit card numbers, and other personal identifiers, however, which can create very detailed portraits of customers.

The Balancing Act

How can we balance users' desire for personalized service with their concerns about privacy? One proposal for solving this problem is the Platform for Privacy Preferences project (http://www.w3.org/P3P/). P3P is designed to be integrated into browsers to let users easily manage relationships with multiple Web sites and identify what types of information they are willing to share with each. In its current proposed form, however, P3P has some problems.

There are questions, for example, about initial settings: How private should P3P be by default? Various parties have differing opinions on whether P3P clients should automatically share information with vendors, ask users whether to share information, or to share no information unless users explicitly approve it.

P3P has a policing problem because it was designed to specify privacy policies, not to enforce them. Enforcement must be carried out by third parties that audit businesses to ensure that they are following their stated policies.

Some sites are volunteering to store P3P data for everyone. (This service is a centerpiece of Microsoft's .Net strategy.) This type of centralized data storage would be a valuable service to users because it would let them access a single repository of all their preference data from anywhere. However, it would also create a convenient target for hackers to access information about millions of people by cracking a single site. Overall, P3P is a step forward for user management of privacy, but future versions must do a better job of balancing personalization and privacy. Other mechanisms might also be necessary as new technologies continue to emerge.

Today, nearly all enterprise Web sites either provide personalization capabilities or plan to soon. This special issue of *IEEE Internet Computing* gives a glimpse into a future in which personalization will be widespread across many domains of human interaction, virtual and physical.

Acknowledgments

The guest editor would like to thank the diligent reviewers for their insightful reviews and the authors for responding so thoughtfully to the feedback they received. To avoid potential conflicts of interest, EIC Munindar Singh made all decisions on the articles by Karypis, Konstan, and Herlocker, with whom Riedl has previously published.

John Riedl is an associate professor in the computer science department at the University of Minnesota and director and chief scientist at Net Perceptions. His research focuses on collaborative systems. Riedl has been codirector of the GroupLens research project on collaborative information filtering since 1992, and he cofounded Net Perceptions to commercialize GroupLens in 1996. He received a BS in mathematics from the University of Notre Dame and MS and PhD degrees in computer science from Purdue.

Readers can contact the author at riedl@cs.umn.edu.