

Editorial - Measuring the impact of personalization and recommendation on user behaviour

Zanker, M., Ricci, F., Jannach, D., Terveen, L.

In recent years, several methods for generating and presenting personalized information have been developed. They are aimed at taming the information overload caused by the vast range of continuously growing content available on the World Wide Web. Examples include adaptive hypertext, which adapts the user interface to the users interests and needs (e.g., in mobile portals), and recommender systems that try to predict what items, e.g., news, movies or travels, are relevant for the user. Recommender systems use various AI and IR techniques to build their predictions, and they typically interact with the user for presenting and revising their recommendations. Many of these techniques and systems have made their way from research into commercial applications and are now widely used in major eCommerce portals. At the same time, new techniques are being proposed, for improving the recommendation effectiveness or for offering new ways for users to participate, as in social networks on Web 2.0 platforms. The assessment of the benefits, for the end users and the service providers, brought by the scientific and technological development of the personalization techniques is obviously an important issue. In fact, choosing the right evaluation method for personalized web applications, identifying the influential success factors behind different techniques or interpreting results coming from online experiments remain largely open research issues.

Off-line experiments, which estimate the recommendation prediction error using an existing dataset of recommendations (as described in the seminal article of Herlocker et al. [1]), is the traditional state-of-practice for evaluating recommender systems. A small survey on the evaluation designs adopted in the research articles on the topic of recommender systems, which appeared in the ACM Transactions on Information Systems over a period of 5 years (2004-2008), showed that three quarters of these studies instrumented offline experiments and half of them chose movie recommendations as their application domain [2]. In fact, many researchers have stressed the limitations of off-line evaluations, to the point that some have argued that the true accuracy of a recommender system could be never directly measured using these approaches.

However, the widespread use of recommender systems makes it crucial to develop evaluation approaches that can realistically and accurately assess the true effectiveness of the recommendations and their effects on the users. For

these reasons, this special issue contains five selected articles that focus on the understanding of how personalization and recommendation impact on the user expectations, beliefs and behavior during and after the interaction. These articles not only consider traditional websites, but also reach out into mobile and ubiquitous communication and interaction scenarios. Furthermore, the selected articles measure the impact of personalization and recommendation in diverse application domains, ranging from traditional e-commerce and news content personalization to mobile, in-vehicle and interactive TV interfaces.

Overview of the papers Selecting the recommendation approach that is best suited for a specific application or product domain is an open research issue, and only few works have discussed the topic of matching recommendation technologies with application domains [3]. In fact, only few empirical studies compare different recommendation strategies in different application settings to gain comparative evidence for deciding between different system designs [5].

The article by Ochi, Rao, Takayama and Nass (this issue) therefore constitutes one piece of additional evidence for building a general theory on the design parameters of recommender systems in specific application domains. They conducted an empirical study under lab conditions where a 2 (content-based vs. collaborative recommendation approaches) \times 2 (few vs. many feature dimensions used to generate recommendations) \times 2 (experience vs. search product type) mixed-design experiment was conducted. Users were initially asked to specify their profile by filling out a questionnaire with either social or content-related questions and were then presented a recommendation list. In order to control the effects of the recommendation list itself all participants were presented with exactly the same recommendations, no matter of the ostensible recommendation approach or the ostensible amount of used feature dimensions.

One key result of Ochi et al.'s study is that the type of recommendation approach (i.e. collaborative vs. content-based filtering) has a technical, but also a psychological impact. Even though the recommendations were identical in all respects, the participants believing that the recommendations were based on the preferences of similar people rather than on the preferred product characteristics, reacted very differently to the same suggestions. Furthermore, the authors identified additional two-way interaction effects between the recommendation approach, the product type and the feature dimensions that suggest that "one size fits all" approach for product recommendation systems is not optimal.

The problem of personalizing news content and measuring whether this personalization is perceived as a valuable service by the customers is at the core of the paper by Lavie, Sela, Oppenheim, Inbar and Meyer (this issue). In contrast to other works that solely rely on the users' past browsing and reading behavior, this study aims to analyze the value of the users' explicitly declared list of topics of interest for choosing the appropriate depth of personalization. In particular, the study was also designed to shed light on the question of whether a user that has declared interest in a specific subtopic might also be interested in articles that are related to more general ones.

In a preliminary survey that involved over 100 participants, the correlation of the participants' interest in a general topic and specific subtopics was analyzed for different topic categories such as "Science and Technology" or "Health". The survey revealed that these correlations are strongly dependent on the breadth and number of the subtopics in a category. In the following study the participants first declared their topics of interest and then were asked to rate the interestingness of several articles in order to determine their real interests. The analysis showed that the value of explicit interest statements for personalization purposes is strongly influenced by the news topic. For some topics, for example, the users' declared interests correlated well with the actual interests while in other cases this correlation was relatively low. From these and other observations, Lavie et al. then derive a small set of guidelines for personalizing news content based on the users' expressed interest in topics and general features of a news article.

Overall, the study indicates that personalization systems operating on the basis of explicit interest statements provided by the users have to be carefully designed. This holds in particular for news personalization, an area in which the user interests are often not well-defined and aspects such as up-to-dateness or the general importance of a news article have to be considered when determining the personalized list of news recommendations.

Also the paper by Chiu, Kao and Loet (this issue) is concerned with the consumption of information and entertainment services in non-stationary situations. In this case, the application scenario is the personalized distribution and consumption of user generated web logs (blogs) on mobile devices. Users and providers are confronted with the limitations of the usage environment; the screen size of the mobile devices and the bandwidth limitations. Therefore, the amount of information and the actual content that can be "pushed" to mobile customers must be carefully chosen. In their paper, Chiu et al. report the findings of a study concerned with the effects of personalization of content suggested with a WAP push service of a Taiwanese telecommunications provider. The experiment involved over 20.000 customers, who used the blog delivery service; the nearly 3.000 blogs themselves covered a wide variety of topics from technology over finance to news and beauty.

The developed system rely on a combination of various information retrieval techniques for document categorization and recommender systems, but it also exploits human experts for the selection of content that is appropriate for mobile consumption.

In order to determine the effects of personalization of the distributed content, both the click rates and the time spent on individual documents have been measured and compared with what observed in the interaction with non-personalized content. The experiment showed that click rates (item selection) could be significantly increased in different runs. Moreover, there is evidence that users spend more time on personalized pages even though the personalization itself is only based on the user's browsing history and not on more sophisticated preference elicitation methods. In summary, the paper represents additional evidence that demonstrates the value of personalization and recommendation in commercial

settings.

Today, fuelled by the increased availability of low-cost wireless network connectivity, information, communication and entertainment services are increasingly consumed “on-the-move”. Beside the rapid developments of new smartphones, more and more car manufactures start to equip their vehicles with (internet-enabled) information and communication appliances. In contrast to mobile phones or stationary devices, the design of such systems must be safety-critical as such systems can distract their users from their main driving task. In order to minimize such distractions, is necessary to automate individual tasks as much as possible, e.g., by allowing the user to perform the most common actions without much effort. Such personalized automation can be implemented in an adaptive system whose behavior is based on the user’s past actions.

The paper by Lavie and Meyer (this issue) presents a study, in which the effects of different levels of such in-vehicle adaptation system behavior on the performance of drivers are analyzed under lab conditions. The experiment design included test drives on a driving simulator, during which the subjects performed various routine- and non-routine tasks such as reading and answering text messages, reading news messages or simply changing from radio to CD. Other variables included the adaptivity level and the difficulty of the test drive. The effect of the adaptation was mainly measured in the driving performance, i.e., to which extent the drivers could keep the optimal lane position.

The experiments reported by Lavie and Meyer indicate that adaptive user interfaces can be a helpful to support the driver in such situations, but also that the adaptation level of these systems has to be very carefully designed in order to avoid undesired system behavior in particular in case of non-routine tasks and unfamiliar driving situations. Overall, the paper offers a novel method and experiment design for measuring the effect of system-side adaptation in in-vehicle information-consumption scenarios which will most probably receive increased attention in the next years.

The last article in this special issue also addresses the evaluation of recommender systems in ubiquitous environments. The work by Funk, Rozinat, Karapanos, Alves de Medeiros and Koca (this issue) proposes an evaluation framework that captures situated judgements in the field. Capturing the situational context of user feedback is crucial for correctly interpreting what the user actually meant. Therefore, they employ an event-triggered experience sampling approach where user feedback is collected at any moment after the test products have been placed in the field.

They implement this framework in an interactive TV set-top box prototype that recommends video content and offers search and browsing functionality. Therefore, Funk et al.’s main contribution is the specification of the evaluation framework for assessing systems in ubiquitous environments and its application in a field study that serves as a proof of concept. By using the concept of event-based experience sampling they could show that the different types of user activities that can be performed with the system score very differently with respect to the constructs valence and goal orientedness. Therefore, this article

lays the foundation for a better understanding of the subjective and situated nature of the perceived quality of a system's personalization and recommendation services in ubiquitous environments.

Finally, we would like to express our gratitude to the more than 30 reviewers for their commitment and conscientious efforts in developing this special issue. We hope that this collection of papers and their results, will be useful for practitioners and researchers, and will contribute to further develop knowledge in this exciting and useful research area.

Markus Zanker
University Klagenfurt
Institute of Applied Informatics
Universitaetsstrae 65-67
9020 Klagenfurt, Austria

Francesco Ricci
Free University of Bozen-Bolzano
Faculty of Computer Science
Piazza Domenicani 3
39100 Bozen-Bolzano, Italy

Dietmar Jannach
Dortmund University of Technology
Faculty of Computer Science
Baroper Str. 301
44227 Dortmund, Germany

Loren Terveen
University of Minnesota
Department of Computer Science and Engineering
200 Union Street SE
Minneapolis, MN 55455, USA

References

- [1] J. L. Herlocker, J. A. Konstan, L. G. Terveen, J. T. Riedl, Evaluating collaborative filtering recommender systems, *ACM Trans. Inf. Syst.* 22 (1) (2004) 5–53. doi:<http://doi.acm.org/10.1145/963770.963772>.
- [2] D. Jannach, M. Zanker, A. Felfernig, G. Friedrich, *Recommender Systems - An Introduction*, Cambridge University Press (CUP), 2010, in print.
- [3] R. Burke, M. Ramezani, *Recommender Systems Handbook*, Springer, 2010, in print, Ch. Matching Recommendation Technologies and Domains.

- [4] F. Ricci, L. Rokach, B. Shapira, P. Kantor (Eds.), *Recommender Systems Handbook*, Springer, 2010, in print.
- [5] B. Xiao, I. Benbasat, E-commerce product recommendation agents: Use, characteristics, and impact, *MIS Quarterly* 31 (1) (2007) 137–209.
- [6] P. Ochi, S. Rao, L. Takayama, C. Nass, Predictors of user perceptions of web recommender systems: How the basis for generating experience and search product recommendations affects user responses, *International Journal of Human-Computer Studies* this issue (2010) –. doi:DOI: 10.1016/j.ijhcs.2009.10.005.
URL <http://www.sciencedirect.com/science/article/B6WGR-4XP381S-1/2/ae4141f382e58d974c34>
- [7] T. Lavie, M. Sela, I. Oppenheim, O. Inbar, J. Meyer, User attitudes towards news content personalization, *International Journal of Human-Computer Studies* this issue (2010) –. doi:DOI: 10.1016/j.ijhcs.2009.09.011.
URL <http://www.sciencedirect.com/science/article/B6WGR-4XKBYWV-1/2/426044a2b90103813962>
- [8] P.-H. Chiu, G. Y.-M. Kao, C.-C. Lo, Personalized blog content recommender system for mobile phone users, *International Journal of Human-Computer Studies* this issue (2010) –. doi:DOI: 10.1016/j.ijhcs.2010.03.005.
URL <http://www.sciencedirect.com/science/article/B6WGR-4YT05CW-1/2/408442158ceb8f371ef3>
- [9] T. Lavie, J. Meyer, Benefits and costs of adaptive user interfaces, *International Journal of Human-Computer Studies* this issue (2010) –. doi:DOI: 10.1016/j.ijhcs.2010.01.004.
URL <http://www.sciencedirect.com/science/article/B6WGR-4YBMFB4-1/2/02486e954879e0aa86d3>
- [10] M. Funk, A. Rozinat, E. Karapanos, A. A. de Medeiros, A. Koca, In situ evaluation of recommender systems: Framework and instrumentation, *International Journal of Human-Computer Studies* this issue (2010) –. doi:DOI: 10.1016/j.ijhcs.2010.01.002.
URL <http://www.sciencedirect.com/science/article/B6WGR-4Y8G605-1/2/9a2bc25588b09e7ef56e>