H. Reiterer & O. Deussen (Hrsg.): Workshopband Mensch & Computer 2012
München: Oldenbourg Verlag, 2012, S. 95-102

Recommender Systems: Between Acceptance and Refusal

Thomas Neumayr, Mirjam Augstein

Communication and Knowledge Media, University of Applied Sciences Upper Austria

Abstract

Recommender Systems (RSs) are a prominent solution to the problem of information overload on the web. It is impossible for users to process or even understand all information presented to them. Also, it becomes more and more difficult for an individual to identify appropriate concrete pieces of information or information sources. RSs aim at adapting the presented content and the order in which it is presented to users' individual needs, based on their preferences and past behavior. Yet, a system can only provide accurate recommendations if it has been authentically used before, i.e., been able to collect information about a user. As authentic usage depends on a user's acceptance, the success of RSs in general is strongly dependent on acceptance also. If recommendations seem inappropriate, the trust in the system will fade. This paper presents a study analyzing how and to what extent different factors like transparency or controllability influence acceptance in the context of web-based recommendation.

1 Introduction

During the past decades, the amount of information offered on the web has been steadily growing, and, as (Gantz & Reinsel, 2011) suggest, it is likely to keep growing even exponentially. Obviously it is impossible for a web user to process and understand all of this information, which leads to a manifold information dilemma. In order to work with the web efficiently, users would have to be able to identify not only the pieces of information that are potentially fitting their requirements, but also information sources that are likely to provide the required information. Personalized systems are a successful approach to overcome the information dilemma and, with the advent of the Web 2.0, gained considerable popularity. Yet, the concept of personalization it not only relevant in the scope of the Web 2.0 and user-generated content but also made its way into more "traditional" fields like e-commerce. In general, a personalized web-based system aims at adapting its appearance, behavior and content to a user's individual characteristics and needs (Knutov, De Bra, & Pechenizkiy, 2009), based on a user model that represents the user's interests, current knowledge, goals, etc. (Brusilovsky & Millán, 2007). A user model is not only fed with information the user provided explicitly, but also with information a system gained by observing user interaction

(Brusilovsky & Maybury, 2002). The actual selection of data stored in a user model depends on the respective application area, e.g., in an e-learning system, a user's preferred learning approaches or previous knowledge might be most relevant whereas in an e-commerce system, a user's interests, profession or hobbies are more important.

In this paper, we concentrate on Recommender Systems (RSs) as a specific form of personalized systems. For an RS, content selection and personalized presentation are the most important aspects of personalization. An RS preselects elements based on user preferences, past behavior, and information about the elements (Frankowski et al., 2007), and presents them to the user in the order it considers to most likely comply with the user's interests. Yet, RSs do not only entail advantages compared to static systems, they also face several challenges. An RS is strongly dependent on its users – it might perform perfectly well if it is used extensively and authentically but might fail elsewise. The failure of an RS can manifest itself in, e.g., imprecise conclusions about users' interests, leading to inappropriate recommendations. This again could reduce users' trust and cause reduced usage authenticity, thus completing a vicious circle. Other challenges like security and robustness issues (regarded from a technical viewpoint) are not dealt with in this paper. The paper discusses factors that influence the acceptance of RSs in Section 2. Sections 3 and 4 describe a study performed in order to analyze how people perceive a particular RS and how the different factors introduced before affect acceptance, and Section 5 discusses related work, summarizes our conclusions and provides an outlook to future work.

2 Acceptance in Recommender Systems

As introduced before, the success of an RS is crucially dependent on its acceptance, as it can just perform well if authentically used. Users on the other hand interact with a system authentically and extensively only if they *i)* recognize the advantages for themselves and *ii)*, have the feeling they can trust the system (Hine, 1998), (Lam, Frankowski, & Riedl, 2006).

2.1 General Facilitating and Inhibiting Factors

In order to derive measures to foster acceptance, facilitating and inhibiting factors have to be identified. (Paramythis, Weibelzahl, & Masthoff, 2010) describe a user-centered evaluation approach for adaptive systems and identify evaluation criteria that can be regarded factors influencing acceptance: *Transparency/comprehensibility*, *predictability*, *privacy*, *controllability/scrutability*, *breadth of experience/serendipity*, *unobtrusiveness*, *timeliness*, *aesthetics* and *appropriateness/necessity*. They suggest this compilation of criteria as a basis for heuristic evaluation. Yet, in the context of this paper we use it in a more general sense as a basis for identifying factors facilitating and inhibiting acceptance. (Jameson, 2009) names *diminished predictability and comprehensibility*, *diminished control* and *obtrusiveness* as decisive factors for usage. He defines *predictability* as "the extent to which a user can predict the effects of her actions", *comprehensibility* as "the extent to which she can understand system actions and/or has a clear picture of how the system works", and *obtrusiveness* as "the extent to which the system places demands on the user's attention which reduce the user's ability to concentrate on her primary tasks". Summing up, we deduce that *predictability* and *comprehensibility* are strongly related to *transparency* and identify a relation between *diminished control* and *controllability/scrutability*. The latter can potentially also lead to increased *precision*.

2.2 Definitions and Hypotheses

We define four chosen factors representing a common denominator in literature, especially reported by (Paramythis et al., 2010), (Jameson, 2009), (Pu, Chen, & Hu, 2012), (Cramer et al., 2008) and (Knijnenburg, Willemsen, Gantner, Soncu, & Newell, 2012): A system is *transparent* if a user has an understanding of its logic (Sinha & Swearingen, 2002), i.e., a user knows why an element has been suggested. A system is *unobtrusive* if its actions do not disturb the user unnecessarily and the user's approval of system actions is not sought too often (Paramythis et al., 2010). A system is *scrutable* if it "allows users to correct reasoning and system assumptions where needed" (Pu et al., 2012), and if a user can influence the system's behavior (Kay, Kummerfeld, & Lauder, 2003). *Precision* can be understood as the *perceived precision* of a user, i.e., "the percentage of items in a recommendation list that the user would rate as useful" (Schafer, Frankowski, Herlocker, & Sen, 2007).

Based on the synopsis of the findings of literature review and these definitions, we construct hypotheses to be used in the study described below. We suppose that *transparency* and *acceptance* are positively correlated, thus an increase in *transparency* leads to an increase in *acceptance [H1]*, *obscurity* and *scrutability* are negatively correlated, thus an increase in *obscurity* leads to a decrease in *scrutability [H2]*, *obtrusiveness* and *acceptance* are negatively correlated, thus an increase in *obtrusiveness* leads to a decrease in *acceptance [H3]*, and a system's *speculation* and *precision* are negatively correlated, thus an increase in *speculation* leads to a decrease in *precision [H4]*. *H1* is restricted to cases where users are in consent with the basic functionality of a system. Otherwise, an increase in transparency might reveal unsolicited details and thus decrease acceptance. Regarding *H4*, speculation can be understood as the opposite of scrutability. In addition to the hypotheses, our further derivations are based on two invariants: An increase in *speculation* leads to a decrease in *obtrusiveness [I1]*, and an increase in *transparency* leads to a decrease in *obscurity [I2].*

3 Study

This section describes a study conducted to find out if and how different factors like transparency or obtrusiveness are related to acceptance.

3.1 Relevant Variables

Based on *I1* and *I2*, the four dimensions of transparency vs. obscurity and speculation vs. obtrusiveness can be reduced to two: transparency and obtrusiveness. Precision and scrutability, both found to lead to increased acceptance, are additionally considered. In our study, a system's transparency and obtrusiveness are varied while on the other hand acceptance (as measured by precision and scrutability) is observed. The relevant constructs and derived variables investigated in the study are shown in Table 1.

Construct	**Items**	**Source**
Transparency	*TR1*: I understand why especially this product is recommended to me. *TR2*: I have the feeling that the order of the recommended products has a specific meaning.	(Sinha & Swearingen, 2002) (Herlocker, Konstan, & Riedl, 2000)
Scrutability	*SC1*: I think/know that I can change the system goal-oriented/purposeful. *SC2*: I know which actions I have to perform in order to let the system know that it has made a failure so that I do not see the undesired product here in the future.	(Kay et al., 2003) (Pu et al., 2012)
Obtrusiveness	*OB1*: Are explanations of system's actions not disturbing unnecessarily and too often? *OB2*: Is the user's approval of system actions not sought too often, when it is not really needed?	(Paramythis et al., 2010)
Precision	*PR1*: Overall, I think that the recommended products are interesting to me and assembled well. *PR2*: The system provides me with products I am not interested in. [negative wording reversed]	(Schafer et al., 2007)

Table 1: Relevant constructs composed of items and their respective source of relevant literature.

3.2 Method and Study Setup

Several challenges are connected to the evaluation of RSs. As personalized systems need time until they can provide the user with valuable recommendations, it would be necessary to set up a longitudinal experiment, which can be costly. Thus, alternatives are often demanded, e.g., (Paramythis et al., 2010) examined (indirect) user tests or simulated users. Yet, these methods do not provide a solution to the problem of the delay until the evaluation can start. Also, timesaving options often lack reliability, e.g., indirect user test and simulated users run the risk of less authentic interaction. The simulated users method is not seen as capable of assessing qualitative aspects like the feeling of trust. Thus, we decided to use real user models for our study. Several factors suggested using an existing system as the only way to circumvent both the reliability drawbacks of indirect or simulated user tests and the cold-start problem (Schein, Popescul, Ungar, & Pennock, 2002). For two reasons, the commercial RS Amazon.com[1] was selected. Firstly, Amazon has a high number of users, thus it seemed likely to recruit many different types of probands. Secondly, several measures to facilitate an increased perception of transparency have already been deployed there.

An online survey, aiming at covering all constructs listed in Table 1, was conducted. Participants were asked to log in to Amazon, perform several operations there and switch back and forth between their Amazon account and the survey to tell their observations. They were asked the same set of questions for three different system states (comparable to scenarios) *S1*, *S2* and *S3*. These states could be reached over the same navigation paths and showed the same wireframes, like "here all probands can see a few articles that are recommended", while concrete content, e.g. the products, may and will have differed. *S1* contained a rather undetailed recommendation list, *S2* a more detailed list, and *S3* details for a specific recommendation. Measures to facilitate transparency (like explanations) were increased from *S1* to *S2* and from *S2* to *S3,* and (objectively) about the same level of obtrusiveness was deployed to all three states (by simulated system explanations for *S1* and *S2* and "real" ones for *S3*).

4 Results

The survey was active for ~20 days; it was completed by 224 (out of ~900) participants. N=207 (101 M, 106 F) participants remained after a few potentially corrupt datasets were removed. The survey was distributed to students and staff of the University of Applied Sciences Upper Austria and via related mailing lists and Facebook. Demographic and usage data were gathered in the form of gender, age, educational level, the date Amazon was joined as well as the frequency of buying, selling, rating and writing reviews.

The survey revealed the following general trends: The perceived level of transparency, as measured by *TR1* and *TR2* (see Table 1), remained relatively constant from *S1* through *S3*, the perceived level of scrutability rose between *S1* and *S3*, and a significant decrease in the perceived level of obtrusiveness could be observed between *S2* and *S3* (the decrease between *S1* and *S2* was not significant). Furthermore, as expected, the decrease in perceived obtrusiveness led to an increase in perceived precision.

A regression analysis highlighting multidirectional influences of the factors on each of the system states helped to interpret the results more in-depth (see Figure 1). Scrutability and precision were used as dependent variables whereas transparency and obtrusiveness served as independent variables. Independents with a T value ≥ 2 and sign. ≤ 0.05 were considered meaningful. According to these criteria, transparency had a positive impact on precision on *S1* and revealed that people who find the additional effort to handle an explanation more okay (related to obtrusiveness) also perceived the system as more scrutable. These findings

[1] http://amazon.com, last access 27 June 2012

could generally be verified on *S2*. On *S3*, a positive impact of transparency on scrutability was revealed.

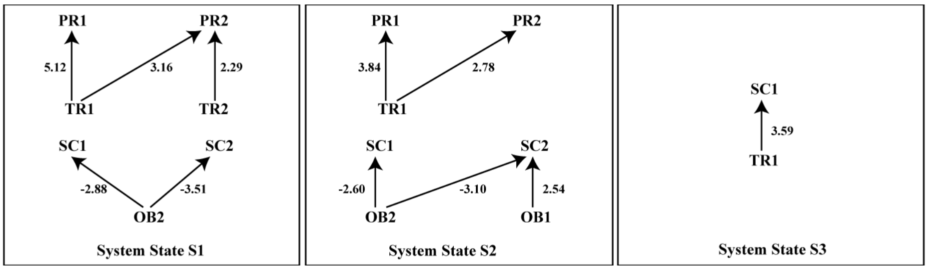

Figure 1: Relevant links between dependent and independent variables. T values are depicted. See Table 1 for abbreviations.

Consequently, analyses suggested that users who have a certain understanding of a system's logic (transparency) tend to perceive a system as more precise as opposed to users who don't. As there is common agreement on the positive correlation between perceived precision and acceptance (see, e.g., (Pu et al., 2012)), *H1* could be verified in this context. Further, the analyses showed that users with a good understanding of a system's logic find it easier to manipulate the system according to their needs, which can be regarded a verification of *H2*. Explanations provided by the system can facilitate both transparency and scrutability and have been seen as mostly welcome by the probands. It could even be the case that additional actions for handling explanations are not perceived as obtrusive/disturbing "enough" to outweigh the advantages of increased scrutability. Therefore, *H3* could not be verified in its original form in this context but could be refined as follows: *An increase in obtrusiveness does not automatically lead to a decrease in acceptance. If other acceptance-enhancing factors are promoted through the increase in obtrusiveness, overall acceptance may well rise. H4* could not be verified in the context of this study. None of the regression analyses suggested a significant relation between obtrusiveness and precision.

5 Conclusions and Related Work

In this paper we discussed the issue of acceptance in RSs, identified facilitating and inhibiting factors like transparency or obtrusiveness, and presented the results of a related study. The hypotheses derived in Section 2.2 were tested by extensive regression analyses and basic hints for researchers interested in increasing the acceptance of RSs were provided. For instance, the importance of explanations in RSs has been verified insofar as they have been shown to lead to both increased transparency and scrutability, which, in turn lead to an increase in the overall acceptance of a system. Yet, increasing transparency and scrutability are only two of the possible motives behind offering explanations, the others being *trust*, *effectiveness*, *persuasiveness*, *efficiency*, and *satisfaction* (Tintarev & Masthoff, 2012). Further, (Tintarev & Masthoff, 2012) presented studies on the trade-off between *effectiveness* and *satisfaction*, showing that "Contrary to expectation, personalization was detrimental to effectiveness, though it may improve user satisfaction". Some findings like the heterogeneity of the effect of explanations and their comprehensive usefulness could generally be verified in our research. An approach similar to ours was described before by (Van Velsen, 2011), who reported a large-scale web survey aiming at understanding user acceptance of online content personalization. He measured different aspects of trust, perceived controllability and their impact on a user's decision to (not) use a technology and defined related hypotheses. Next, scenarios, comparable to the different system states discussed in this paper, were created. The results of Van Velsen's survey showed that perceived controllability has strong influence on the decision to use a technology. The established usability principle *user control* is clear common ground between these findings and those presented in this paper, as here, scrutability and transparency, which can potentially be facilitated by explanations, have also been identified as important.

The study discussed in this paper placed a strong focus on the individual. (Pu, Chen, & Hu, 2011) discussed a similar approach, letting probands use items that are of special interest to them as a basis for their survey, which seems to be a legitimate way to introduce a user-centered stance. Yet, our data also contains several group-specific variables like gender, age and patterns of use (see Section 4). It seems likely that users belonging to different "groups" react differently to acceptance-enhancing measures like explanations, etc. Thus, future analyses will concentrate on possible differences and similarities between the groups to derive further measures to tailor the appearance of RSs to users' needs and therefore contribute to increased acceptance.

Acknowledgement

Please note that this paper contains a condensed summary of the findings reported in the first author's master thesis (Neumayr, 2012).

References

Brusilovsky, P., & Maybury, M. T. (2002). From Adaptive Hypermedia to the Adaptive Web. *Commun. ACM*, *45*(5), 30–33.

Brusilovsky, P., & Millán, E. (2007). User Models for Adaptive Hypermedia and Adaptive Educational Systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The Adaptive Web*. Springer-Verlag Berlin Heidelberg. 3-53

Cramer, H., Evers, V., Ramlal, S., van Someren, M., Rutledge, L., Stash, N., Aroyo, L., et al. (2008). The Effects of Transparency on Trust in and Acceptance of a Content-based Art Recommender. *User Modeling and User-Adapted Interaction*, *18*(5), 455–496.

Frankowski, D., Lam, S. K., Sen, S., Harper, F. M., Yilek, S., Cassano, M., & Riedl, J. (2007). Recommenders Everywhere: the WikiLens Community-maintained Recommender System. *Proceedings of the 2007 International Symposium on Wikis*, WikiSym '07. New York, NY, USA: ACM. 47-60.

Gantz, J., & Reinsel, D. (2011). Extracting Value from Chaos. *IDC Research Report IDC Research Report, Framingham, MA, June. Retrieved September*, *19*, 2011.

Herlocker, J. L., Konstan, J. A., & Riedl, J. (2000). Explaining Collaborative Filtering Recommendations. *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, CSCW '00. New York, NY, USA: ACM. 241-250.

Hine, C. (1998). Privacy in the Marketplace. *The Information Society*, *14*(4), 253–262.

Jameson, A. (2009). Adaptive Interfaces and Agents. *Human-Computer Interaction: Design Issues, Solutions, and Applications*, 105.

Kay, J., Kummerfeld, B., & Lauder, P. (2003). Managing Private User Models and Shared Personas. *UM03 Workshop on User Modeling for Ubiquitous Computing*. 1–11.

Knijnenburg, B., Willemsen, M., Gantner, Z., Soncu, H., & Newell, C. (2012). Explaining the User Experience of Recommender Systems. *User Modeling and User-Adapted Interaction*, *22*(4), 441–504.

Knutov, E., De Bra, P., & Pechenizkiy, M. (2009). AH 12 Years Later: A Comprehensive Survey of Adaptive Hypermedia Methods and Techniques. *New Review of Hypermedia and Multimedia*, *15*(1), 5–38.

Lam, S., Frankowski, D., & Riedl, J. (2006). Do You Trust Your Recommendations? An Exploration of Security and Privacy Issues in Recommender Systems. In G. Müller (Ed.), *Emerging Trends in Information and Communication Security*, Lecture Notes in Computer Science (Vol. 3995). Springer Berlin / Heidelberg. 14-29.

Neumayr, T. (2012). *Recommender Systeme: Im Spannungsfeld zwischen Vertrauen und Ablehnung*. Unpublished Master Thesis, Submitted 25 June 2012.

Paramythis, A., Weibelzahl, S., & Masthoff, J. (2010). Layered Evaluation of Interactive Adaptive Systems: Framework and Formative Methods. *User Modeling and User-Adapted Interaction*, *20*(5), 383–453.

Pu, P., Chen, L., & Hu, R. (2011). A user-centric evaluation framework for recommender systems. *Proceedings of the fifth ACM conference on Recommender systems*, *RecSys '11*. New York, NY, USA: ACM. 157-164.

Pu, P., Chen, L., & Hu, R. (2012). Evaluating Recommender Systems from the User's Perspective: Survey of the State of the Art. *User Modeling and User-Adapted Interaction*, *22*(4), 317–355.

Schafer, J., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative Filtering Recommender Systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The Adaptive Web*, Lecture Notes in Computer Science (Vol. 4321). Springer Berlin / Heidelberg. 291-324.

Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2002). Methods and Metrics for Cold-start Recommendations. *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, *SIGIR '02*. New York, NY, USA: ACM. 253-260

Sinha, R., & Swearingen, K. (2002). The Role of Transparency in Recommender Systems. *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, *CHI EA '02*. New York, NY, USA: ACM. 830-831.

Tintarev, N., & Masthoff, J. (2012). Evaluating the Effectiveness of Explanations for Recommender Systems. *User Modeling and User-Adapted Interaction*, *22*(4), 399–439.

Van Velsen, L. S. (2011). *User-centered Design for Personalization*. University of Twente