Int. J. Human-Computer Studies 72 (2014) 367–382

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

How should I explain? A comparison of different explanation types for recommender systems ☆

Fatih Gedikli [a,*], Dietmar Jannach [a], Mouzhi Ge [b]

[a] TU Dortmund, 44221 Dortmund, Germany
[b] Bundeswehr University Munich, 85579 Neubiberg, Germany

ARTICLE INFO

Article history:
Received 14 August 2012
Received in revised form
15 December 2013
Accepted 23 December 2013
Communicated by M. Zanker
Available online 1 January 2014

Keywords:
Recommender systems
Decision support
Explanations
Interface design
Collaborative filtering
Tag clouds
User evaluation

ABSTRACT

Recommender systems help users locate possible items of interest more quickly by filtering and ranking them in a personalized way. Some of these systems provide the end user not only with such a personalized item list but also with an explanation which describes why a specific item is recommended and why the system supposes that the user will like it. Besides helping the user understand the output and rationale of the system, the provision of such explanations can also improve the general acceptance, perceived quality, or effectiveness of the system.

In recent years, the question of how to automatically generate and present system-side explanations has attracted increased interest in research. Today some basic explanation facilities are already incorporated in e-commerce Web sites such as Amazon.com. In this work, we continue this line of recent research and address the question of how explanations can be communicated to the user in a more effective way.

In particular, we present the results of a user study in which users of a recommender system were provided with different types of explanation. We experimented with 10 different explanation types and measured their effects in different dimensions. The explanation types used in the study include both known visualizations from the literature as well as two novel interfaces based on tag clouds. Our study reveals that the content-based tag cloud explanations are particularly helpful to increase the user-perceived level of transparency and to increase user satisfaction even though they demand higher cognitive effort from the user. Based on these insights and observations, we derive a set of possible guidelines for designing or selecting suitable explanations for recommender systems.

1. Introduction

Recommender systems point online users to possibly interesting or unexpected items, thereby increasing sales or customer satisfaction on modern e-commerce platforms (Linden et al., 2003; Senecal and Nantel, 2004; Zanker et al., 2006; Dias et al., 2008; Jannach and Hegelich, 2009). However, personalized recommendation lists alone might be of limited value for the end users when they have to decide between different alternatives or when they should assess the quality of the generated recommendations. In other words, only showing the recommendation lists can make it hard for the users to decide whether they can actually trust that the recommended items are actually useful and interesting without inspecting all of them in detail.

One possible approach to support the end user in the decision making process and to increase the trust in the system is to provide an explanation for why a specific item has been recommended (Herlocker et al., 2000; Bilgic and Mooney, 2005; Pu and Chen, 2006; Tintarev and Masthoff, 2007a, 2007b; Friedrich and Zanker, 2011). In general, there are many approaches of explaining recommendations, including non-personalized as well as personalized ones. An example of a non-personalized explanation would be Amazon.com's "*Customers who bought this item also bought...*" label for a recommendation list, which also carries explanatory information.

This work deals with questions of how explanations could be communicated to the user in a more effective way. This includes both questions of the *visual representation* as well as questions of the *content* to be displayed. In general, the type and depth of explanations a recommender system can actually provide depend on the types of knowledge and/or algorithms that are used to generate the recommendation lists. In knowledge-based recommendation or advisory approaches, explanations can be based on the rule base which encodes an expert's domain knowledge and the explicitly acquired user preferences (Felfernig et al., 2007; Jannach et al., 2009; Zanker, 2012).

☆This paper has been recommended for acceptance by M. Zanker.

* Corresponding author.
E-mail addresses: fatih.gedikli@tu-dortmund.de (F. Gedikli), dietmar.jannach@tu-dortmund.de (D. Jannach), mouzhi.ge@ebusiness-unibw.org (M. Ge).

1071-5819/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ijhcs.2013.12.007

For the most prominent type of recommender systems, collaborative filtering recommenders, Herlocker et al. (2000) and Bilgic and Mooney (2005) have proposed various ways of explaining recommendations to the user. Herlocker et al. have also shown that explanations can help to improve the overall acceptance of a recommender system.

In this paper, we continue the line of work of Herlocker et al. (2000), Bilgic and Mooney (2005), Vig et al. (2009), Tintarev and Masthoff (2007a, 2012), and our own work presented in Gedikli et al. (2011). In particular, we aim to contribute to the following research questions.

1. The main question we seek to answer in this paper is which effects different explanation types for recommendations have on users. In the existing literature on recommender system explanations, authors often limit their analysis to some specific explanation goals (Herlocker et al., 2000; Bilgic and Mooney, 2005; Pu and Chen, 2006, 2007) or explanation types (Vig et al., 2009; Gedikli et al., 2011). In our work, however, we aim at evaluating different explanation types in a comprehensive manner and consider the desired effects and quality dimensions *efficiency*, *effectiveness*, *persuasiveness*, *perceived transparency*, and *satisfaction* (Tintarev and Masthoff, 2011) in parallel. To that purpose, we conducted a laboratory study involving 105 subjects in which we compare several existing explanation types from the literature (Herlocker et al., 2000) with a tag-based explanation approach.
2. Going beyond existing research which focuses only on one explanation goal or analyze trade-offs between two quality dimensions,[1] we aim at detecting interdependencies between more than two quality dimensions. In particular, our goal is to analyze the influence of efficiency, effectiveness, and perceived transparency on user satisfaction. Based on the dependencies between the different effects of explanation types, we aim to derive a first set of possible guidelines for the design of effective and transparent explanations for recommender systems and sketch potential implications of choosing one over the other. These guidelines were validated through a qualitative interview-based study involving 20 participants.
3. We finally aim to obtain a deeper understanding of the value of the recently proposed tag- and preference-based explanation types proposed in Gedikli et al. (2011). We included two variants of this explanation method in our experimental study and compare their performance with the other explanation types in the different quality dimensions. Since acquiring explicit tag preferences is costly and can be cumbersome for the user, one of the two tag-based explanations incorporates a new method to automatically estimate the user's detailed preferences from the item's overall ratings.

The paper is organized as follows. Section 2 summarizes the quality factors for recommender system explanations and discusses related and previous work. Section 3 introduces the different explanation types compared in our study. Section 4 describes the experimental setup. Section 5 provides a discussion of the obtained results and our first set of possible design guidelines. Section 6 finally summarizes the main findings of this work and gives an outlook on future work.

2. Explanations in recommender systems

In recent years, the concept of explanations has been widely discussed in the area of recommender systems (Pu and Chen, 2007; Tintarev and Masthoff, 2008; Vig et al., 2009; Friedrich and Zanker, 2011; Tintarev and Masthoff, 2012). An explanation can be considered as a piece of information that is presented in a communication process to serve different goals, such as exposing the reasoning behind a recommendation (Herlocker et al., 2000) or enabling more advanced communication patterns between a selling agent and a buying agent (Jannach et al., 2010). Up to now, however, there exists no standard definition of the term "explanation" in the context of recommender systems. According to Tintarev and Masthoff (2012), a popular interpretation of the term explanation in recommender systems is that explanations "*justify*" the recommendations. Since this definition might be too narrow, we propose to characterize explanations through the possible aims which one might want to achieve with them in a recommendation system. Tintarev and Masthoff identify seven possible aims of explanations for recommender systems as shown in Table 1.

This paper investigates the impact of different explanation types on the first five factors in this list. Next, we will characterize these factors in more detail and sketch how to measure each of them.

2.1. Efficiency

An explanation is usually considered to be efficient when it helps the user to decide more quickly or when it helps to reduce the cognitive effort required in the decision process. In the context of conversational recommender systems, Thompson et al. (2004), for example, measure efficiency by computing the total interaction time between the user and the recommender system until the user has found a suitable item. McSherry (2005), in contrast, measures efficiency through the number of required dialogue steps before a user accepts one of the system's recommendations. In other papers, efficiency is sometimes calculated by measuring the time used to complete the same task with and without an explanation facility or with different types of explanations, see, e.g., the study of Pu and Chen (2006).

In our work, we adopt an efficiency measure that is based on the decision time required by a user. We distinguish between "item-based" and "list-based" efficiency. Typically, recommender systems can produce two types of output: (a) a rating prediction showing to what degree the user will like or dislike an item and (b) a list of n recommended items. Therefore, efficiency can be measured either for each individual item or for a given list of recommendations. Item-based efficiency thus considers the decision time required by a user to evaluate a single candidate item at a time (see, e.g., Gedikli et al., 2011). An appropriate protocol for list-based efficiency would be to measure the overall time required by a user to decide on one single best item given a larger candidate set with explanations (see, e.g., Thompson et al., 2004; McCarthy et al., 2005). In order to make the results comparable to our prior work on explanations (Gedikli et al., 2011), we decided to measure the item-based efficiency in this work.

Table 1
Possible goals of using explanations in recommender systems.

(1) Efficiency	Reducing the time used to complete a task
(2) Effectiveness	Helping the users make better decisions
(3) Persuasiveness	Changing the user's buying behavior
(4) Transparency	Explaining why a particular recommendation is made
(5) Satisfaction	Increasing usability and enjoyment
(6) Scrutability	Making the system's user model correctable
(7) Trust	Increasing the user's confidence in the system

[1] See Table 2 in Tintarev and Masthoff (2012).