Available online at www.sciencedirect.com
ScienceDirect
Knowledge-Based Systems 20 (2007) 542–556
Knowledge-Based SYSTEMS
www.elsevier.com/locate/knosys

Trust-inspiring explanation interfaces for recommender systems

Pearl Pu, Li Chen *

Human Computer Interaction Group, School of Computer and Communication Sciences, Swiss Federal Institute of Technology in Lausanne (EPFL), CH-1015, Lausanne, Switzerland

Received 17 January 2007; accepted 17 April 2007
Available online 21 April 2007

Abstract

A recommender system's ability to establish trust with users and convince them of its recommendations, such as which camera or PC to purchase, is a crucial design factor especially for e-commerce environments. This observation led us to build a *trust model* for recommender agents with a focus on the agent's trustworthiness as derived from the user's perception of its competence and especially its ability to explain the recommended results. We present in this article new results of our work in developing design principles and algorithms for constructing explanation interfaces. We show the effectiveness of these principles via a significant-scale user study in which we compared an interface developed based on these principles with a traditional one. The new interface, called the organization interface where results are grouped according to their tradeoff properties, is shown to be significantly more effective in building user trust than the traditional approach. Users perceive it more capable and efficient in assisting them to make decisions, and they are more likely to return to the interface. We therefore recommend designers to build trust-inspiring interfaces due to their high likelihood to increase users' intention to save cognitive effort and the intention to return to the recommender system.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Recommender systems; Recommender agents; Interface design; Decision support; Explanation interfaces; Trust model; Competence perception; Trusting intentions; User evaluation

1. Introduction

The importance of explanation interfaces in providing system transparency and thus increasing user acceptance has been well recognized in a number of fields: expert systems [11], medical decision support systems [2], intelligent tutoring systems [29], and data exploration systems [4]. Being able to effectively explain results is also essential for product recommender systems. When users face the difficulty of choosing the right product to purchase, the ability to convince them to buy a proposed item is an important goal of any recommender system in e-commerce environments. A number of researchers have started exploring the potential benefits of explanation interfaces in a number of directions.

Case-based reasoning recommender systems that can explain their recommendations include ExpertClerk [27], Dynamic critiquing systems [12], FirstCase and TopCase [16,17]. ExpertClerk explained the selling point of each sample in terms of its difference from two other contrasting samples. In a similar way, FirstCase can explain why one case is more highly recommended than another by highlighting the benefits it offers and also the compromises it involves with respect to the user's preferences. In TopCase, the relevance of any question the user is asked can be explained in terms of its ability to discriminate between competing cases. McCarthy et al. [12] propose to educate users about product knowledge by explaining what products do exist instead of justifying why the system failed to produce a satisfactory outcome. This is similar to the goal of resolving users' preference conflict by providing them with partially satisfied solutions [25]. Some consumer decision support systems with explanation interfaces can be found on commercial websites such as Logical Decisions (www.logicaldecisions.com),

* Corresponding author.
E-mail addresses: pearl.pu@epfl.ch (P. Pu), li.chen@epfl.ch (L. Chen).

0950-7051/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.knosys.2007.04.004

Active Decisions (www.activedecisions.com), and SmartSort (shopping.yahoo.com/smartsort).

A number of researchers also reported results from evaluating explanation interfaces with real users. Herlocker et al. [10] addressed explanation interfaces for recommender systems using ACF (automated collaborative filtering) techniques, and demonstrated that a histogram with grouping of neighbor ratings was the most compelling explanation component among the studied users. They maintain that providing explanations can improve the acceptance of ACF systems and potentially improve users' filtering performance. Sinha and Swearingen [28] found that users like and feel more confident about recommendations that they perceive as transparent.

So far, previous work on explanation interfaces has not explored its potential for building users' trust in recommender agents. Trust is seen as a long term relationship between a user and the organization that the recommender system represents. Therefore, trust issues are critical to study especially for recommender systems used in e-commerce where the traditional salesperson, and subsequent relationship, is replaced by a product recommender agent. Studies show that customer trust is positively associated with customers' intention to transact, purchase a product, and return to the website [8]. These results have mainly been derived from online shops' ability to ensure security, privacy and reputation, i.e., the integrity and benevolence aspects of trust constructs, and less from a system's *competence* such as a recommender system's ability to explain its result. These open issues led us to develop a trust model for building user trust in recommender agents, especially focusing on the role of the competence construct. We pursue our research work in four main areas: (1) we investigate the inherent benefits of using explanation for trust building in recommender systems; (2) we examine whether competence-inspired trust provides the same trust-related benefits as other trust constructs, for example benevolence and integrity; (3) we seek promising areas to investigate interface design issues for building user trust, and (4) we develop sound principles and algorithms for building such interfaces. In the first stage of this work, we have developed a trust model for recommender systems[1] and evaluated its validity through a carefully constructed user survey [5]. We established that competence perception is an essential contribution to trust building and provides trust-induced benefits such as intention to return. As the second part of this work, it is therefore essential to concentrate on those design aspects of an interface that help the system increase its perceived competence. The work reported in this article emphasizes design principles and algorithms for generating competence-inspiring interfaces and testing these principles in empirical studies.

[1] We use recommender systems and agents interchangeably. However, the term "system" is used more often to refer to the entire computing environment, while "agent" is more frequently used to refer to the interface of a system and the perception it creates.

This article is organized as follows: Section 2 summarizes our previous work in developing a trust model for recommender systems and some results from a qualitative survey, which identified explanation interfaces as one of the most promising areas to address issues for building user trust; Section 3 describes a set of general principles derived from an in-depth examination of various design dimensions for constructing explanation interfaces, followed by an algorithm that we developed to optimize these principles; Section 4 presents a research model which explains more clearly how we developed the hypotheses on the main benefits of explanation interfaces, and discusses the design and implementation of a significant-scale empirical study to validate these hypotheses; Section 5 reports results from that study indicating that the organization-based explanation, where recommendations are organized into different categories according to their tradeoff properties relative to the top candidate, is more likely to inspire users' trust, given the fact users perceive it more capable and efficient in helping them interpret and process decision information (i.e., effort saving), and are more likely to return to it; Section 6 discusses the implication of this work to related work in this area, followed by the conclusion and future work.

The present article provides a number of follow-up results and more analytical detail to our earlier paper [24]. To better explain how the organization interface algorithm works in action, we use a step-by-step data flow diagram in Section 3.2 (organization algorithm) to illustrate the generation of such interfaces (see Fig. 1). Section 4.1 explains how we establish the hypotheses and their interrelationships to be tested in the empirical study. More discussions are given on the design of user tasks and their rationale (Section 4.3). Section 5.2 is added to include new results from path coefficient analyses to show the important causal relationships of trust constructs. Several important conclusions regarding user trust and its benefits such as users' intention to save cognitive effort are derived. To offer some explanations on why users prefer the organization based interfaces, we analyzed and have included users' actual comments in Section 5.3. Finally, we include more detailed discussion of the future work in Section 7 (Conclusion), particularly addressing the long-term trust issues and how trust relates to other issues such as user control and privacy.

2. Trust model and explanation interfaces

This section summarizes our earlier work and results on constructing a trust model for recommender systems [5]. It is intended to offer an overview of the overall research agenda and a roadmap identifying the most promising areas for investigating design issues for trust-inspiring interfaces.

2.1. Trust model for recommender systems

We have conceptualized a competence-focused trust model for recommender agents (see details in [5]). It consists

of three components: system features, trustworthiness of the agents, and trusting intentions. The system features mainly deal with those design aspects of a recommender agent that can contribute to the promotion of its trustworthiness. We classified them into three groups: the interface display techniques, the algorithms that are used to propose recommendations, and user-system interaction models, such as how an agent elicits users' preferences.

The agent trustworthiness is a trust formation process based on the users' perception of the agent's competence, reputation, integrity, and benevolence. It has been regarded as the main positive influence on the trusting intentions [8,15]. In this article, we primarily consider the competence perception and its essential contribution to trust-induced benefits.

The trusting intentions are the benefits expected from users once trust has been established by the recommender agents. The trusting intentions include the intention to purchase a recommended item, return to the store for more information on products or purchase more recommended products, and save effort. The intention to save effort is of particular interest to us because it examines whether upon establishing a certain trust level with the agent users will likely spend less cognitive effort or actual time in selecting the recommended items. As it turns out, trust formation is useful in making users feel that they expend less effort, even though they would expend a similar amount of actual task time, at least for the decision tasks that we have instructed our uses to perform . Please see Section 5.2 for detailed analyses of our results.

2.2. Trust building with explanation interfaces

As a first step, we primarily consider trust building by the different design dimensions of interface display techniques, especially those for the explanation interfaces, given their potential benefits to improve users' confidence about recommendations and their acceptance of the system [10,28]. We investigate the modality of explanation, e.g., the use of graphics vs. text, the amount of information used to explain (i.e., explanation richness), e.g., whether long or short text is more trust inspiring, and most importantly whether alternative explanation techniques exist that are more effective in trust building than the simple "why" construct currently used in most e-commerce websites.

The explanation generation mainly comprises the steps of content selection and organization, media allocation, and media realization and coordination [4]. Content selection determines what information should be included in the explanations. For instance, the neighbors' ratings can be included to explain the recommended items computed by collaborative filtering technology [10]. Once the content is selected, we must know how to organize and display it. The simplest strategy is to display the recommendation content in a rank ordered list with a "why" component for each recommendation explaining the computational reasoning behind it. This strategy has been broadly embodied in the case-based reasoning systems and commercial websites [12,16,17,27].

As an alternative and potentially more effective technique, we have designed an organization-based explanation interface where the best matching item is displayed at the top of the interface along with several categories of tradeoff alternatives, each labeled with a title explaining the tradeoff characteristics of the items the respective category contains as to how they differ from the top candidate (see Fig. 4). This was inspired by the work of McCarthy et al. [13] which suggests that recommending products in groups of compound critiques enabled users to reach their decisions much faster.

2.3. Qualitative survey and results

We have conducted a survey with 53 users in order to understand the interaction among the three components of our trust model: the effect of an agent's competence in building users' trust, the influence of trust on users' problem solving efficiency and other trusting intentions, and the effective means to build trust using explanation-based interfaces. Nine hypotheses (see [5] for details) were established, each of which is a statement for which the participants indicated their level of agreement on a 5-point Likert scale (ranging from "strongly disagree" to "strongly agree").

Results indicate that the competence of recommender agents would not be the only contribution to users' trust formation process (mean = 3.15, $p = 0.121$), but it is positively correlated with the trusting intention to return (mean = 3.55, $p < 0.01$). In other words, if users possess a high perception of the recommender agent's competence, they would be more inclined to return to the agent for other product information and recommendations. However, they would not necessarily intend to buy the product from the website where the agent was found (mean = 4.23, $p < 0.01$), even though they established high competence perception. Post-survey discussion indicated that they would visit more websites to compare the product's price before making a purchase. The website's security, reputation, delivery service and privacy policy were also important considerations in buying a product.

With respect to the effect of explanation interfaces on trust building, users positively responded that explanation can be an effective means to achieve their trust (mean = 3.64, $p < 0.01$), and the organization interface is a more effective explanation technique than the simple "why" construct (mean = 3.91, $p < 0.01$). On the other hand, the modality and richness of an explanation interface did not seem to contribute much to the effectiveness of the interface (respectively, mean = 2.38, $p < 0.01$; mean = 2.85, $p < 0.01$). From the participants' viewpoints, these two aspects were mostly dependent on the concrete product domain. Users would prefer a short and concise conversational sentence for the so-called low-risk products such as

movies and books, but if they were selecting products which carry a high level of financial and emotional risks such as cars and houses, a more detailed and informative explanation would be favored. In addition, people with different professional outlooks (for example math vs. history majors) seemed to have different requirements for the media modality.

Based on results from the qualitative survey, we have decided to focus our attention on organization-based explanation interfaces and the related design issues for building users' trust.

3. Organization-based explanation interfaces

Traditional product search and recommender systems present a set of top-k alternatives to users. We call this style of display the k-best interface. Because these alternatives are calculated based on users' revealed preferences (directly or indirectly), these top-k items may not provide for diversity. Recently the need to include more diversified items in the result list has been recognized. Methods have been developed to address users' potentially unstated preferences [7,22], cover topic diversity [30], propose possible tradeoffs a user may be prepared to accept [16], and allow faster navigation to the target choice by critiquing the proposed items [3,13,26]. The organization-based explanation interface which we have developed can be regarded as a combination of the ideas of diversity, tradeoff reasoning, and explanation. Here we review a set of design principles that show promise for the design of such interfaces.

3.1. Design principles

We have implemented more than 13 paper prototypes of the organization-based interface, exploring all design dimensions such as how to generate categories, whether to use short or long text for category titles, how many tradeoff dimensions to include, whether to include example products in the categories or just the category titles, etc. We have derived 5 principles based on the results of testing these prototypes with real users in the form of pilot studies and interviews.

Principle 1. Consider categorizing the remaining recommendations according to their tradeoff properties relative to the top candidate.

We consider using explanation interfaces in the early stage of the interaction cycle between a user and a recommender agent. We assume that users are unlikely to have stated all of their preferences at this point. Consequently, they have not considered tradeoff alternatives of the product currently being considered. According to [23,26], integrating tradeoff support in a product search tool can improve users' decision accuracy by up to 57%. Thus, it suggests that displaying tradeoff alternatives in addition to the top candidate is likely to encourage users to consider uncertain or unstated preferences. Each category comprises a set of items having the same tradeoff properties. For example, one category contains the recommendations of notebooks that are cheaper but heavier than the top candidate, and another category's notebooks are lighter but more expensive. Each category indicates a potential tradeoff direction that users may consider in order to achieve their final decision goals.

Principle 2. Consider proposing improvements and compromises in the category title using conversational language, and keeping the number of tradeoff attributes no more than three to avoid information overload.

Here we consider designing a category's title in terms of its format and richness. After surveying some users, we found that most of them preferred category titles presented in natural and conversational language because that makes them feel at ease. For example, the title "these notebooks have a lower price and faster processor speed, but heavier weight" is preferred to the title "cheaper and faster processor speed and heavier." Moreover, the former title is also preferred to the title "they have a lower price and faster processor speed and bigger memory, but heavier weight and larger display size" which includes too many tradeoff properties. Many users indicate that handling tradeoff analysis beyond three attributes is rather difficult.

Principle 3. Consider eliminating dominated categories and diversifying the categories in terms of their titles and contained recommendations.

The third principle proposes to provide decision theoretic and diverse categories to users. Dominance relationship is an important concept in decision theory, originally proposed by Pareto [20]. A category is dominated by another one if the latter is superior to the former on all attributes. This principle suggests that we never propose dominated categories. For example a category containing heavier and slower portable PCs will never be shown next to a category containing lighter and faster products. This dominance relationship checking combined with diversity checking will likely ensure the decision quality and diversity of the suggested categories and the included items. In addition, the pilot study on category design showed that the total number of displayed categories is more effective when up to four since too many categories cause information overload and confusion.

Principle 4. Consider including actual products in a recommended category.

While comparing two interface designs, one displaying only category titles versus one displaying both category titles and a few actual products, users indicated a strong preference in favor of the latter design, mainly due to the fact that they were able to find their target choice much faster. Given the limitation of the display size and users' cognitive effort, a designer may consider choosing up to 6 items to include in each category.

Principle 5. Consider ranking recommendations within each category by exchange rate rather than similarity measure.

We have also performed a pilot study to compare the effects of two ranking strategies for the recommendations within the category. The *similarity* strategy is broadly used by early case-based and preference-based reasoning systems (CBR), which rank items according to their similarity degrees relative to a user's current query. We propose another strategy based on the *exchange rate* of an item relative to the top candidate, i.e., its potential gains versus losses compared to the top candidate (the detail formula for exchange rate calculation will be shown shortly). The study showed that users could more quickly find their target choice when the recommended items within each category were sorted by exchange rate rather than by similarity.

3.2. Organization algorithm

The organization algorithm was developed to optimize the overall objectives of the five principles. The top level of the algorithm can be described in four steps (see Fig. 1): generate all possible category titles by the Apriori algorithm [1]; exclude dominated categories; select a few prominent categories not only with longer tradeoff distance with the top candidate but also with higher diversity degree between each other; rank the recommended items within each category by their exchange rates relative to the top candidate. A resulting example based on the organization algorithm can be seen in Fig. 4.

Step 1. Generate all possible categories.

We generate the categories using the method presented in [13]. One modification is that we represent each recommendation as a tradeoff vector comprising a set of (*attribute*, *tradeoff*) pairs (the pair is also called an item in the Apriori algorithm [1]). Each *tradeoff* property (i.e., each (*attribute*, *tradeoff*) pair) indicates whether the *attribute* of the recommendation is *improved* (denoted as ↑)or *compromised* (denoted as ↓)compared to the same attribute of the top candidate. An example of a notebook recommendation is denoted by a tradeoff vector {(price, ↑), (processor speed, ↓), (memory, ↓), (hard drive size, ↑), (display size, ↑), (weight, ↓)}, indicating that this notebook has a lower price, more hard drive size, and larger display size, but heavier weight, slower processor speed, and less memory relative to the top recommended notebook. Thus a tradeoff vector describes how the current product compared to the top candidate in terms of its advantages and disadvantages, rather than the simple equality comparison used in dynamic critiquing (bigger, smaller, equal, different, etc.) [13]. After all tradeoff vectors are used as input to the Apriori algorithm [1], we obtain the frequent item sets in terms of their tradeoff potentials underlying all the recommendations.

The Apriori algorithm has been widely used to resolve the market-basket analysis problem [1]. The main objective is to find regularities in the shopping behavior of customers by identifying sets of products that are frequently bought together. In dynamic critiquing systems, this algorithm was used to discover the available compound critiques from a given data set. More concretely, each critique pattern (reflecting the differences between one of the remaining products in the data set and the current recommendation) is equivalent to the shopping basket for a single customer, and the individual critiques correspond to the items in this basket. Through the Apriori algorithm, a set of compound critiques can be discovered as association rules of the form $A \rightarrow B$. In other words, from the presence of a certain set of critiques (A) one can infer the presence of highly related critiques (B) [13]. Each compound critique in set B is produced with a support value referring to the percentage of products that satisfy the critique. The dynamic critiquing approach select compound critiques with low support values to be included in B. That is, the approach only choose those products that are most likely to help users narrow their search.

In our approach, we use the same support value (i.e., 10%) when we apply the Apriori algorithm. Therefore, the number of products contained in each critiquing category is guaranteed to be at most 10% of all remaining products. We also set the Apriori's option "maximal number of items per set" to 3 in order to limit the number of attributes involved in each category title (according to principle 2). However, we do not select categories purely based on their support values, but largely consider their average tradeoff benefits according to the user's current preferences so as to improve the user's final decision accuracy.

Step 2. Exclude dominated categories.

If one category is strictly dominated by another category in terms of the item sets they contain in the titles, we will not show it to the user. Formally, each category title is denoted as a set of (*attribute*, *tradeoff*) pairs. A category title C_1 is dominated by another category title C_2 if they satisfy the following condition:

$|C_1| = |C_2|$ (meaning C_1 and C_2 contain the same number of items in the titles), and $\forall$ item $T_i \in C_1$, $\exists$ $T_j \in C_2$, where $T_i.attribute = T_j.attribute$ (with equal attribute name), $T_i.tradeoff \preceq T_j.tradeoff$ (with equal or less preferred tradeoff property, i.e., "↓" $\prec$ "↑"), and $\exists$ $T_p \in C_1$, $T_q \in C_2$, where $T_p.attribute = T_q.attribute$ and $T_p.tradeoff \prec T_q.tradeoff$ (i.e., at least one item is with less preferred tradeoff property). For example the title C_1 {(weight, ↓), (price, ↓), (processor speed, ↑)} is dominated by C_2 {(weight, ↓), (price, ↑), (processor speed, ↑)}, since its price is less preferred than C_2s (i.e., "↓" $\prec$ "↑") while the tradeoff properties of the other two attributes in C_1 and C_2 are equal.

Step 3. Select prominent categories with longer tradeoff distance and higher diversity degree.

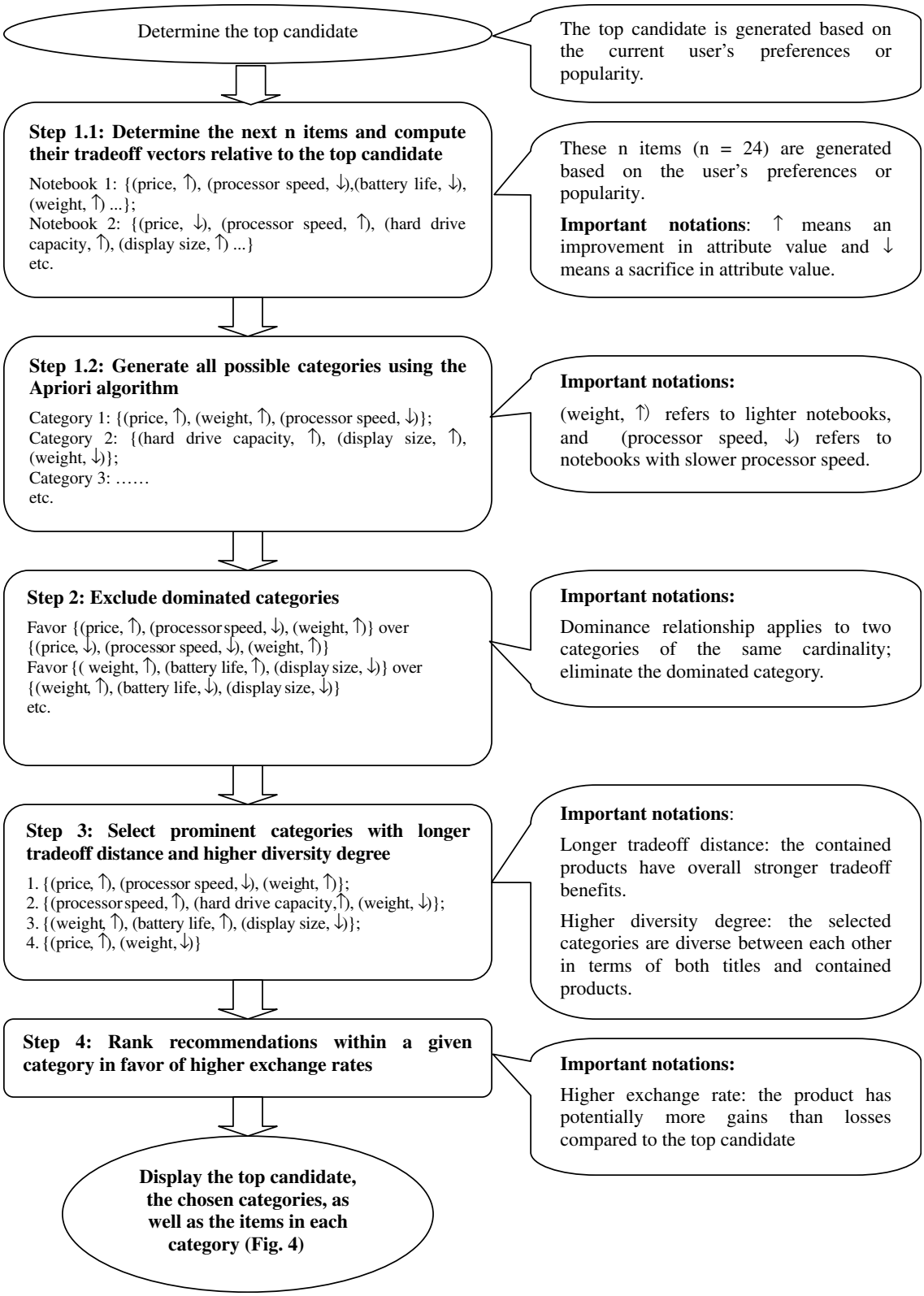


Fig. 1. Step-by-step data flow diagram of the organization algorithm.

This is another aspect where we depart from the dynamic critiquing method [13] which only uses the low support value to select categories. We use two criteria to select up to four categories: the maximal tradeoff distance with the top candidate and maximal diversity among each other in terms of their titles and contained recommendations (Principle 3). The tradeoff distance of each category is defined as the average sum of the

exchange rate of all recommendations which are contained in the category:

$$TradeoffDistance(C_i, TC) = \frac{1}{|SR(C_i)|}\sum_{R\in SR(C_i)}^{|SR(C_i)|} ExRate(R, TC)$$

where TC is the top candidate, $SR(C_i)$ is the set of recommendations contained in the category C_i, and $ExRate(R, TC)$ is the exchange rate of the recommendation R compared to the top candidate (see the $ExRate$ formula in Step 4). Intuitively, a category possessing higher values in tradeoff distance offers products with more gains than losses relative to the top candidate. Thus presenting this category is more likely to stimulate users to consider selecting them and thus improve their decision quality.

During the selection process, the category with the longest tradeoff distance will be initially selected as the first category. The second category will be selected if it has the biggest value of $F(C_i)$ in the remaining non-selected categories according to the following formula:

$$F(C_i) = TradeoffDistance(C_i, TC) \times Diversity(C_i, SC)$$

where C_i is the current considered category in the remaining set, TC is the top candidate, and SC denotes the set of categories so far selected. $F(C_i)$ is the combination of the category's tradeoff distance and diversity degree with respect to the categories selected so far. The subsequent categories are selected according to the same rule. The selection process will end when the desired k categories have been selected.

The global diversity of C_i with SC is the average sum of its local diversity with each category in the SC set. The local diversity of two categories is further determined by two factors: the title diversity and recommendation diversity, according to Principle 3.

$$Diversity(C_i, SC) = \frac{1}{|SC|}\sum_{C_j\in SC}^{|SC|} TitleDiv(C_i, C_j) \times RecomDiv(C_i, C_j)$$

The title diversity determines the degree of difference between the two category titles (C_i and C_j), respectively, represented as a set of (*attribute*, *tradeoff*) pairs:

$$TitleDiv(C_i, C_j) = 1 - \frac{|C_i \cap C_j|}{|C_i|}$$

The recommendation diversity measures the different recommendations contained in the two compared categories:

$$RecomDiv(C_i, C_j) = 1 - \frac{|SR(C_i) \cap SR(C_j)|}{|SR(C_i)|}$$

where $SR(C_i)$ represents the set of recommendations included in category C_i.

Step 4. Rank recommendations within a given category by exchange rate.

The global exchange rate for each recommendation R is formulated as:

$$ExRate(R, TC) = \sum_{i=1}^{p} w_i exrate(v_{r,i}, v_{tc,i})$$

where p is the number of attributes, w_i is the weight of attribute i, and $exrate$ is the local exchange rate computed for each attribute ($v_{r,i}$ and $v_{tc,i}$ are the values of the ith attribute of R and TC, respectively). For numeric attributes, $exrate(v_i, v_j) = q \times \frac{v_i - v_j}{range}$. The parameter $q = 1$ if the attribute i is in increasing order (i.e., the more, the better), and $q = -1$ if i in decreasing order (i.e., the less, the better). For nominal attributes, $exrate(v_i, v_j) = 1$ if $v_i \neq v_j$ and v_i is preferred to v_j, or -1 if contrarily, or 0 if $v_i = v_j$.

Therefore, the exchange rate motivates a user to consider alternative choices. A positive and higher exchange rate means that there are potentially more gains than losses (i.e., higher decision quality) of an alternative product compared to the top candidate.

4. User evaluation

In order to understand whether the organization interface based on the design principles and algorithm is a more effective way to explain recommendations, we conducted a significant-scale empirical study that compared our organization interface with the traditional "why" interface in a within-subjects design. The main objective was to measure the difference in users' trust level in terms of the perceived competence and trusting intentions (the intention to save effort and to return) in the two interfaces. To compare the subjective attitudes with actual behaviors, we also measured users' actual task time while selecting the product that they would purchase.

4.1. Research model and hypotheses

Based on the qualitative survey results [5], we have developed a research model (see Fig. 2) representing the various parameters to be measured in our user experiment regarding the effects of explanation interfaces on building users' trust. The trust is mainly assessed by three constructs: perceived competence, the intention to save effort, and the intention to return. The intention to save effort is further measured by the perceived cognitive effort and actual completion time consumed. The survey indicated that users' intention to purchase was not necessarily associated with a recommender system's perceived competence. Therefore, we have excluded the intention to purchase from the research model.

According to this model, our main hypothesis was that users would build more trust in the organization-based explanation interface than the simple "why" interface. That is, users would perceive the organization interface more competent and more helpful in saving their cognitive

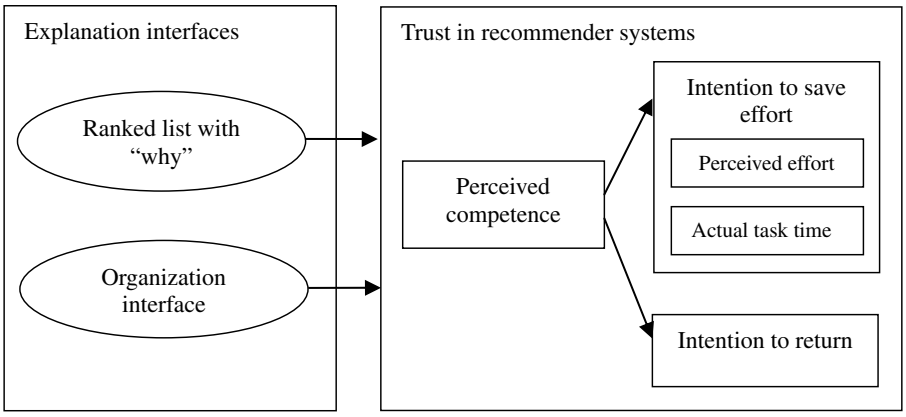


Fig. 2. Research model for the hypotheses evaluated in the user study.

effort for making decisions, and would be more likely to return to it.

Hypothesis 1. The organization-based explanation interface enables users to build more trust in recommender systems compared to systems employing a list view with the simple "why" component.

Using the relationships established in our trust model, we will be able to prove this hypothesis if we can prove the following hypotheses.

Hypothesis 1.1: The organization-based explanation interface enables users to perceive the recommender system more competent in terms of its ease of use and usefulness in comparing products.

Hypothesis 1.2: The organization-based explanation interface enables users to consume less cognitive effort in decision making.

Hypothesis 1.3: The organization-based explanation interface enables users to increase their intention of returning to the interface for future use.

In addition, we also hypothesized that a positive perception of the agent's competence could necessarily enable users to experience less cognitive effort (both subjectively and objectively measured) and increase their intention to return to the agent. Although it was widely agreed in the survey that higher competence perception can increase the users' intention to return, we decided to further prove this point by a quantitative evaluation. As for the benefit of competence perception to effort saving, we were even more motivated to clarify it through the quantitative empirical study since the relevant qualitative survey result was rather inconclusive.

Hypothesis 2. An increased level of perceived competence in a recommender agent leads to more intention to save effort for making decisions.

Hypothesis 3. An increased level of perceived competence in a recommender agent leads to users' increased intention to return to the agent for future use.

4.2. Participants

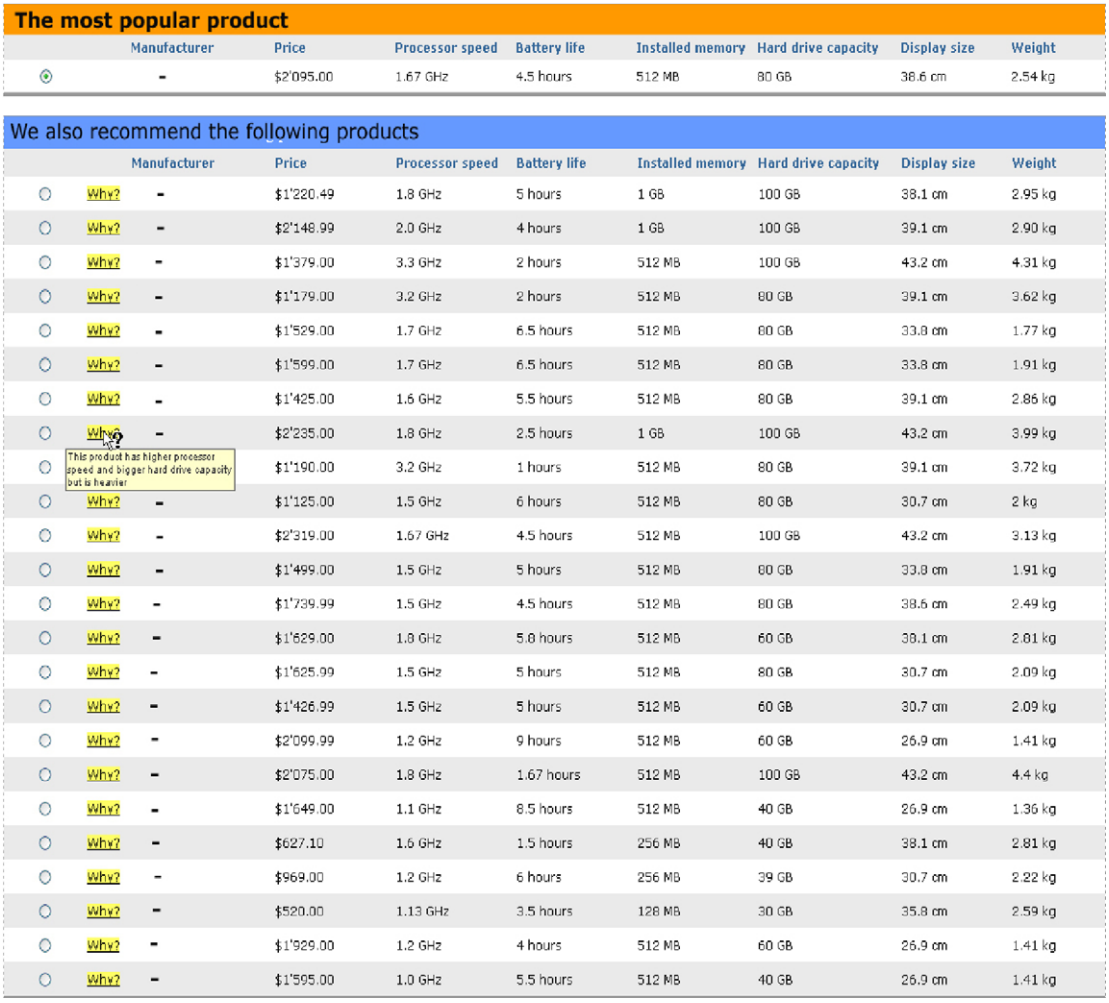
A total of 72 volunteers (19 females) were recruited as participants in the user study. They come from 16 different countries (Spain, Canada, China, etc.), and have different professions (student, professor, research assistant, engineer, secretary, sales clerk and manager) and educational backgrounds (high school, bachelor, master and doctorate degrees). Most of the participants (62 users) had some online shopping experiences. In addition, 54 had bought a notebook in the past two years and 59 users had bought a digital camera. Furthermore, most participants intended to purchase a new notebook (57 users) and digital camera (60 users) in the near future. Table 1 shows some of their demographic characteristics.

4.3. Materials and user task

In order to avoid any carryover effect due to the within-subjects design, we developed a four (2×2) experiment condition and each condition was assigned to a distinct material. The manipulated factors were explanation interfaces' order (organized view first vs. list view with "why" first) and product catalogs' order (digital camera first vs. notebook first). The 72 participants were evenly assigned to one of the four experiment conditions, resulting in a sample size of 18 subjects for each condition cell. For example, the 18 users in one experiment

Table 1
Demographic characteristics of all participants (total 72)

Gender	Female 19 (26.4%)		Male 53 (73.6%)
Education	High school, Bachelor, Master, Doctor		
Nationality	16 countries (Spain, Canada, China, etc.)		
Age	20–30 64 (88.9%)	30–40 4 (5.56%)	>40 4 (5.56%)
Online shopping experience	Yes 62 (86.1%)		No 10 (13.9%)



The most popular product

	Manufacturer	Price	Processor speed	Battery life	Installed memory	Hard drive capacity	Display size	Weight
◉	-	$2'095.00	1.67 GHz	4.5 hours	512 MB	80 GB	38.6 cm	2.54 kg

We also recommend the following products

		Manufacturer	Price	Processor speed	Battery life	Installed memory	Hard drive capacity	Display size	Weight
○	Why?	-	$1'220.49	1.8 GHz	5 hours	1 GB	100 GB	38.1 cm	2.95 kg
○	Why?	-	$2'148.99	2.0 GHz	4 hours	1 GB	100 GB	39.1 cm	2.90 kg
○	Why?	-	$1'379.00	3.3 GHz	2 hours	512 MB	100 GB	43.2 cm	4.31 kg
○	Why?	-	$1'179.00	3.2 GHz	2 hours	512 MB	80 GB	39.1 cm	3.62 kg
○	Why?	-	$1'529.00	1.7 GHz	6.5 hours	512 MB	80 GB	33.8 cm	1.77 kg
○	Why?	-	$1'599.00	1.7 GHz	6.5 hours	512 MB	80 GB	33.8 cm	1.91 kg
○	Why?	-	$1'425.00	1.6 GHz	5.5 hours	512 MB	80 GB	39.1 cm	2.86 kg
○	Why?	-	$2'235.00	1.8 GHz	2.5 hours	1 GB	100 GB	43.2 cm	3.99 kg
○			$1'190.00	3.2 GHz	1 hours	512 MB	80 GB	39.1 cm	3.72 kg
○	Why?	-	$1'125.00	1.5 GHz	6 hours	512 MB	80 GB	30.7 cm	2 kg
○	Why?	-	$2'319.00	1.67 GHz	4.5 hours	512 MB	100 GB	43.2 cm	3.13 kg
○	Why?	-	$1'499.00	1.5 GHz	5 hours	512 MB	80 GB	33.8 cm	1.91 kg
○	Why?	-	$1'739.99	1.5 GHz	4.5 hours	512 MB	80 GB	38.6 cm	2.49 kg
○	Why?	-	$1'629.00	1.8 GHz	5.8 hours	512 MB	60 GB	38.1 cm	2.81 kg
○	Why?	-	$1'625.99	1.5 GHz	5 hours	512 MB	80 GB	30.7 cm	2.09 kg
○	Why?	-	$1'426.99	1.5 GHz	5 hours	512 MB	60 GB	30.7 cm	2.09 kg
○	Why?	-	$2'099.99	1.2 GHz	9 hours	512 MB	60 GB	26.9 cm	1.41 kg
○	Why?	-	$2'075.00	1.8 GHz	1.67 hours	512 MB	100 GB	43.2 cm	4.4 kg
○	Why?	-	$1'649.00	1.1 GHz	8.5 hours	512 MB	40 GB	26.9 cm	1.36 kg
○	Why?	-	$627.10	1.6 GHz	1.5 hours	256 MB	40 GB	38.1 cm	2.81 kg
○	Why?	-	$969.00	1.2 GHz	6 hours	256 MB	39 GB	30.7 cm	2.22 kg
○	Why?	-	$520.00	1.13 GHz	3.5 hours	128 MB	30 GB	35.8 cm	2.59 kg
○	Why?	-	$1'929.00	1.2 GHz	4 hours	512 MB	60 GB	26.9 cm	1.41 kg
○	Why?	-	$1'595.00	1.0 GHz	5.5 hours	512 MB	40 GB	26.9 cm	1.41 kg

Fig. 3. The "why" interface used in the user evaluation.

condition evaluated the ranked list view with "why" explanations for finding a digital camera (similar to Fig. 3 but with digital cameras as the product domain), and then the organization interface for finding a notebook (Fig. 4).

Both product domains comprise 25 up-to-date items from a commercial website (www.pricegrabber.com). Each notebook has 8 attributes (manufacturer, price, processor speed, battery life, etc.) and each digital camera contains 9 attributes (manufacturer, price, megapixels, optical zooms, etc.). To prevent the brand of products from influencing users' choice, we replaced them by fictitious manufacturers which do not exist (they are illustrated with dashes in Figs. 3 and 4).

The top candidate in both interfaces was the most popular item (Figs. 3 and 4). In the "why" interface the remaining 24 products were sorted by their exchange rates relative to the top candidate, where the "why" tool-tip explains how one product compares to the most popular item (Fig. 3). In the organization interface, the remaining items were grouped in four ($k = 4$) categories generated based on our organization selection and ranking algorithms (Fig. 4). The radio button alongside with each item was used by participants to select the product that they were prepared to purchase.

In designing the actual user task, we considered a number of aspects. If users were to find an ideal product in the traditional way, i.e., by first stating their initial preferences, revising them based on available options shown to them and then choosing the final product, we would likely observe different behavioral patterns in terms of how users consult and evaluate the explanation offered. For those who are very certain about their preferences, their likelihood in consulting the explanation and considering recommendations other than the top candidate is rather low. For those who feel less certain about their initial preferences, they will behave in an opposite way because they are more likely to be influenced by the recommendations and explanations (for more details on the adaptive nature of human decision behavior and how they process information, please see [21]). In order to encourage users to consult the explanations as often and as equally likely as possible despite their difference in decision behavior, we decided to recommend the top 25 most popular products from www.pricegrabber.com and ask users to "*select a product that you would purchase if given the opportunity*." Since

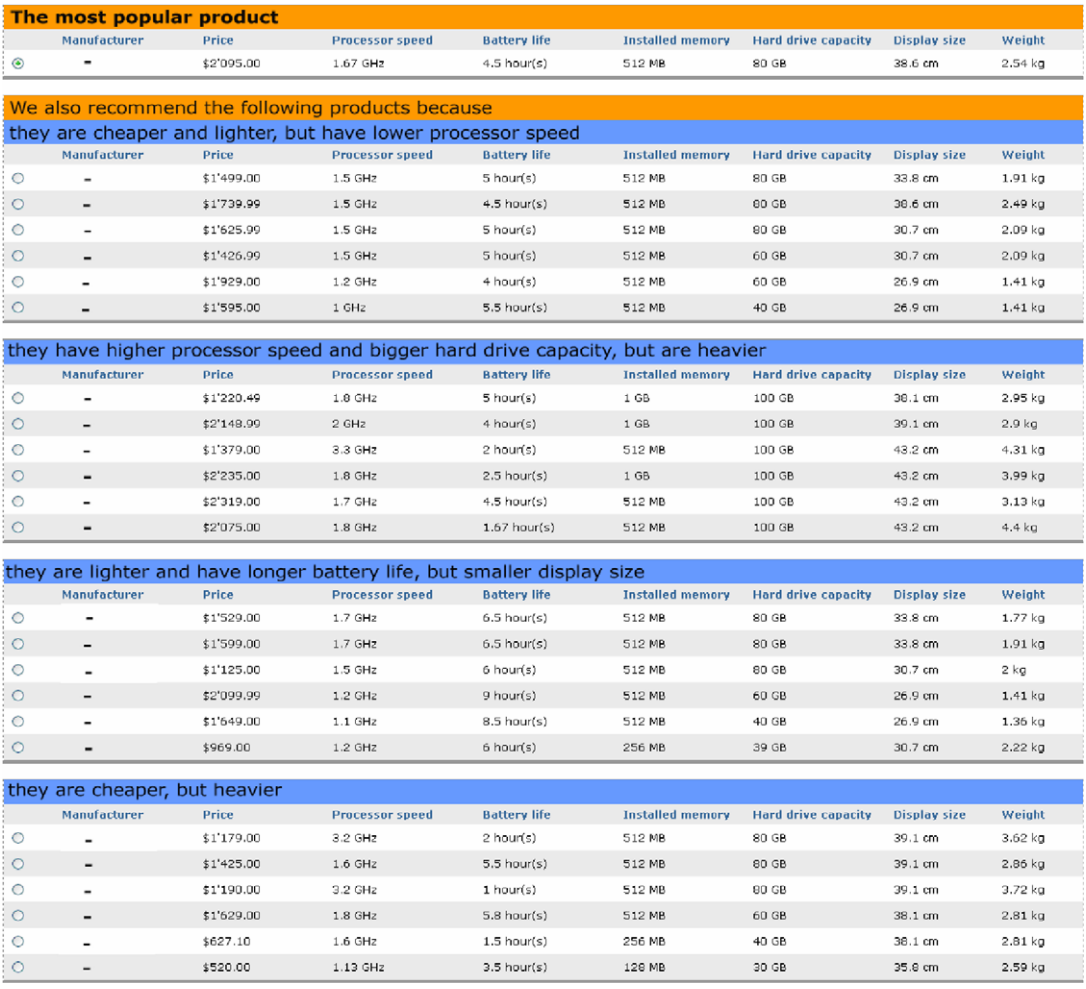


Fig. 4. The organization interface used in the user evaluation.

popularity reflects opinions from other consumers, we judged that users would more likely examine the other 24 products and consult the explanations in order to be "convinced". As the results show, our initial judgment was correct since less than 11.3% of users selected the top candidate in the "why" interface, and only 8.3% in the case of the organization interface.

4.4. Procedure

The user study was conducted at places convenient for the participants (office, home, cafeteria, etc.) with the help of a provided notebook or desktop computer. An online procedure containing the instructions, evaluated interfaces and questionnaires was implemented so that users could easily follow, and also for us to record automatically all of their actions in a log file. The same administrator presided in each user study to address questions from the participants as well as taking notes. The online experiment was prepared in two versions, English and French, since these were the main languages spoken by our participants.

At the beginning of each session, the participants were first asked to choose the language that they preferred, and then they were debriefed on the objective of the experiment and the upcoming tasks. In particular, they were asked to evaluate two graphical recommendation interfaces and to determine which interface was more helpful in recommending products to users. Thereafter, a short questionnaire was to be filled out about their demographics, e-commerce experience and product knowledge. Participants would then start evaluating the two interfaces one by one corresponding to the order defined in the assigned experiment condition. For each interface, the main user task was to "*select a product that you would purchase if given the opportunity*", followed by a total of 6 questions about his/her overall opinions (i.e., trust assessment) regarding the interface. Users were also encouraged to provide any comment on the interface.

5. Results analysis

Results were analyzed for each measured variable using the paired samples *t*-test.

Table 2
Construct composition, items' mean value, construct validity and reliability

Construct	Items of the construct	Mean (St.d.)		Construct validity (factor loading)	Construct reliability (Cronbach's α)
		Organized view	List view with "why"		
Perceived competence	I felt comfortable using the interface	3.24 (1.12)	2.78 (1.31)	0.85	0.84
	This interface enabled me to compare different products very efficiently	3.38 (1.19)	2.72 (1.24)	0.85	
Perceived cognitive effort	I easily found the information I was looking for (*reverse scale*)	2.47 (1.09)	3.07 (1.25)	0.77	0.73
	Selecting a product using this interface required too much effort	2.61 (1.15)	3.14 (1.26)	0.75	
Intention to return	If I had to buy a product online in the future and an interface such as this was available, I would be very likely to use it	3.11 (1.09)	2.56 (1.24)	0.93	0.91
	I don't like this interface, so I would not use it again (*reverse scale*)	3.40 (1.22)	2.79 (1.35)	0.91	

5.1. Trust assessment

5.1.1. Perceived competence

Users' subjective perception of the competence in the interface was mainly measured by their perception of the interface's ease of use and efficiency in comparing products. Each assessment was asked by one item (i.e., a question) in the post-questionnaire marked on a 5-point Likert scale ranging from 1 ("strongly disagree") to 5 ("strongly agree"). Table 2 indicates participants' mean responses and standard deviation to each item for the two interfaces. The construct validity and reliability, respectively, represent how well the two items are related to the construct "perceived competence" and how consistently they are unified for the same construct (the significant benchmark of factor loading for construct validity is 0.5 [9], and of Cronbach's α for construct reliability is 0.7 [19]).

Both items were responded to be on average higher for the organization interface, which shows that most users regarded the organization-based explanation interface more comfortable to use and perceived it to be more efficient in making product comparisons. The overall level of perceived competence of the organization interface is thus higher than that provided by the "why" interface (see Table 3 and Fig. 6; mean = 3.31 for the organization vs. mean = 2.75 for the "why" interface, $t = 3.74$, $p < 0.001$). The constructs' overall mean values in Table 3 were calculated as the average of the mean values for each item contained in the respective constructs in Table 2. For example, the overall mean value for the organized view is 3.31 which is the average of 3.24 and 3.38, respectively, from the two contained items. In addition to mean values, Table 3 also shows the overall median and mode values for the two interfaces.

5.1.2. Intention to save effort

5.1.2.1. Perceived cognitive effort. Cognitive effort refers to the effort associated with gathering and processing of information in order for a person to reach her final decision. The perceived cognitive effort is a subjective evaluation from the user on the overall information processing effort required by a tool and its interface. Like the perceived competence, it was also made up of two items (or questions), respectively, responded on a 5-point Likert scale (see Table 2 for the items and their mean responses). One of the questions was asked on a reverse scale, meaning that the scale ranges from 1 ("strongly agree") to 5 ("strongly disagree").

The lower mean rate therefore represents a smaller amount of cognitive effort an average user perceived during his/her interaction with the corresponding interface. As a result, the overall cognitive effort was perceived significantly lower ($t = -3.89$, $p < 0.001$) on the organization-based explanation interface (see Table 3 and Fig. 6; mean = 2.54 for the organization vs. mean = 3.10 for the "why" interface).

5.1.2.2. Actual completion time. Contrarily to the perceived cognitive effort, the actual completion time is an objective

Table 3
Overall descriptive statistics for trust constructs

Constructs	Mean (St.d.)		Median		Mode	
	Organized view	List view with "why"	Organized view	List view with "why"	Organized view	List view with "why"
Perceived competence	3.31 (1.05)	2.75 (1.20)	3.5	3	4	3.5
Perceived cognitive effort	2.54 (0.96)	3.1 (1.13)	2.5	3	2	3.5
Intention to return	3.27 (1.11)	2.67 (1.24)	3.5	2.5	4	1

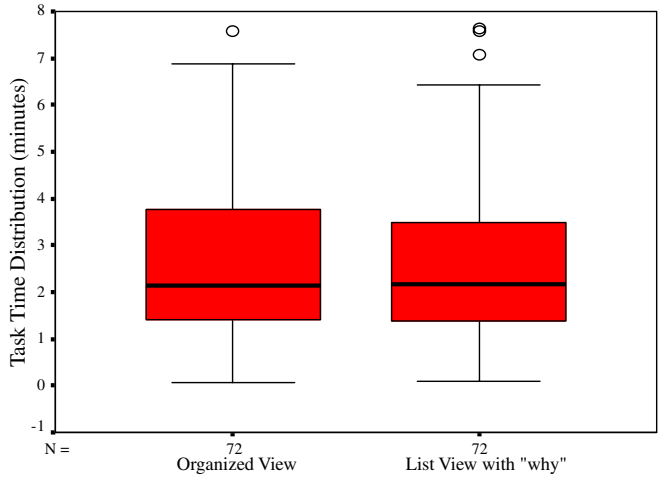


Fig. 5. Actual task time distribution in the two explanation interfaces.

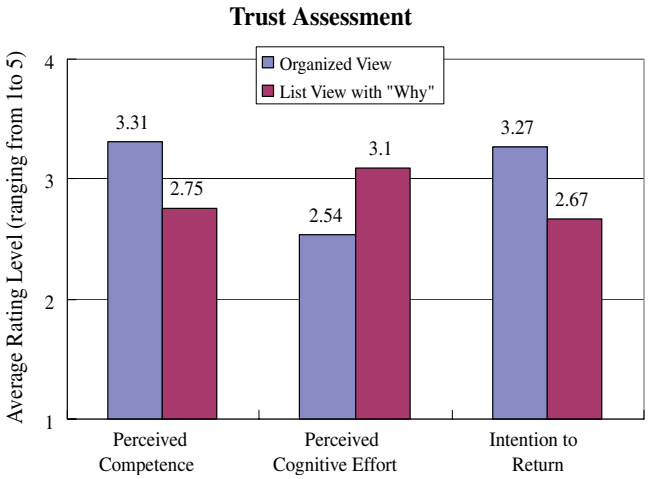


Fig. 6. Mean difference of participants' trust formation in the two explanation interfaces.

measure, defined as the amount of time a participant takes in obtaining and processing information while accomplishing the task of locating a desired product in the interface. No significant difference was found between the two interfaces in terms of task completion time (mean = 2.62 min, SD = 1.67 for the organization vs. mean = 2.60 min, SD = 1.74 for the "why" interface, $t = 0.13$, $p = 0.45$; see Fig. 5 of the actual time distribution in the two explanation interfaces). Users took slightly less time to complete the tasks using the organization interface when compared by the median time (median = 2.13 for the organization vs. 2.18 min for the "why" interface). Combined with the results from measuring the perceived cognitive effort, it indicates that even though users expended a similar amount of time in processing information during their decision making process, they perceived the decision task executed in the organization interface as less demanding.

5.1.3. Intention to return

As demonstrated in our previous work [5], the most remarkable benefit of the competence-inspired trust is its positive influence on users' intention to return. Accordingly, we regard the "intention to return" as an important criterion to judge the trust achievement of explanation-based recommendation interfaces. In our user study, it was assessed by two interrelated post-questions (still using the 5-point Likert scale), which asked participants, positively then negatively, about their genuine intention to use the interface again for future shopping (see Table 2). Note that the negative question was asked on a reverse scale so that the higher the rate the better it is.

The results show that most of participants had a stronger intention of returning to the organization-based explanation interface in the future, than the simple "why" list view. The difference in overall mean value proved to be highly significant (see Table 3 and Fig. 6; mean = 3.27 for the organization vs. mean = 2.67 for the "why" interface, $t = 4.58$, $p < 0.001$).

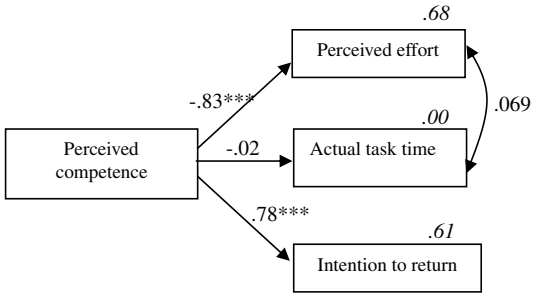


Fig. 7. Standardized path coefficients and explained variance for the measured variables (***indicating the coefficient is at the $p < 0.001$ significant level; explained variance R^2 appearing in italics over the box).

5.2. Path analysis between constructs

Using the path coefficient analysis, we aimed at investigating the causal relationships (see Fig. 7) between the trust constructs. Such an analysis can help us validate Hypotheses 2 and 3: "could an increased level of perceived competence in a recommender agent lead to more intention to save effort in making decisions?" and "could an increased level of perceived competence in a recommender agent lead to users' increased intention to return to the agent for future use?"

The model for the path analysis contains one independent variable, the perceived competence, and three dependent variables: perceived effort, actual completion time, and intention to return to the recommender agent. The path coefficients are partial regression coefficients which measure the extent of effect of one variable on another in the path model using a correlation matrix as the input. The results indicate that an increased level of perceived competence can significantly lead to users' experiencing a lesser amount of perceived cognitive effort in decision making ($b = -0.83$, $p < 0.001$) and an increased intention to return to the recommender system ($b = 0.78$, $p < 0.001$). These findings were further corroborated by the phenomenon that

approximately 68% of variance in cognitive effort ($R^2 = 0.68$) and 61% of variance in intention to return ($R^2 = 0.61$) can be accounted for by the perceived competence (both exceeding the 10% benchmark recommended by Falk and Miller [6]).

However, the path coefficient from perceived competence to actual completion time does not indicate a significant level ($b = -0.02$, $p = 0.829$). The in-depth examination of the correlation between perceived cognitive effort and actual time (i.e., objective effort) reveals that these two aspects of decision effort are actually not significantly associated with each other (correlation = 0.069, $p = 0.414$). This means that even though less task time is spent on the interface, it does not predict that users would perceive the interface to be less demanding, and vice versa.

Thus, the results indicate that an increased level of perceived competence in a recommender agent can necessarily lead to users' experiencing less perceived cognitive effort for making decisions, but it does not infer that they will actually spend less time to locate the choice. In addition, an increased level of perceived competence can also definitely lead to users' increased intention to return to the recommender agent for future use.

5.3. User comments

Further analysis of users' comments made the reasons more explicit as to why the organization interface was subjectively preferred to the simple "why" list by the majority of participants. Many users considered it well structured and easier to compare products from different categories or in one category. Some users found it a little surprising at the beginning, but they soon got used to it and found it useful. It was also accepted as a good idea to label each category to distinguish it from others. In other words, the grouping allowed most of them perceiving the location of a product matching their needs more quickly than the ungrouped display. Although some users also liked the "why" component in the ranked list since it provided a quick overview of advantages and disadvantages of the product compared to the top candidate, they felt too much information was provided in the list view that required more concentration and effort for the decision making than the organized view. Table 4 summarizes users' comments in more detail respecting the two explanation interfaces.

5.4. Discussion

Most of our hypotheses are well supported by the empirical user study. Participants on average built more trust in the organization-based explanation interface, given the fact that they perceived the organization interface more competent and more helpful in processing decision information. They also indicated a higher level of intention to return to it for future use. In addition to the straightforward comparison, most of the participants were actually positively responding to assessment questions concerning the organization interface given the fact that the mean values of these variables are all above the midpoint (i.e., 3) of the Likert scale, whereas the mean values for assessing the "why" interface were all below the midpoint.

The study also shows that a higher level of competence perception does not necessarily lead to reduction in actual time spent on the corresponding interface, which means that users are likely to take nearly the same amount of time to make decisions as in the interface with lower perceived competence. However it is worth pointing out that a more favorable perception of an agent's competence is positively correlated with a reduction in perceived effort. That is, even though users may expend the same amount of actual time in finishing their decision tasks, they are likely to feel as though they did not put in as much effort.

6. Implication to related work

Results from our empirical study strongly support a current trend in displaying a diverse set of recommendations rather than the k-best matching ones. McGinty and Smyth [14] maintain that showing diverse items can reduce the recommendation cycles. McSherry [16] advocates that the displayed items should cover all possible tradeoffs that the user may be prepared to accept. Faltings et al. [7] propose to show products that can be potentially acceptable to users had they stated all of their preferences. In the same spirit, Price and Messinger [22] propose to generate the displayed set taking into account users' preference uncertainty. Our work demonstrates that displaying a diverse set of results in an organization-based interface more effectively enables users' trust formation compared to the simple k-best interface even after the "why" enhancement. We believe that similar trust-related benefits can be obtained for the diversity-driven interfaces proposed by other researchers in this field.

7. Conclusion and future work

We have developed a trust model for recommender agents, and we have shown that explanation interfaces have a great potential in building competence-inspired trust relationships with users. A carefully designed survey indicates that a recommender agent's competence is positively correlated with users' intention to return, but not necessarily with their intention to purchase. It also shows that an organization-based explanation interface is likely to be more effective than the simple "why" interface, since most participants felt that it would be easier for them to compare different products and make a quicker decision.

Based on these results, we have pursued a set of five principles for the design of organization interfaces and an algorithm for generating the content of such interfaces. We reported here a significant-scale comparative study to further quantify users' trust formation and trusting intentions. Results show that while both interfaces enable

Table 4
Users' qualitative comments respecting the two explanation interfaces

The organized view Comments in favor of this interface	**The list view with "why"** Comments in favor of this interface
"Better than the list view because it is organized."	"At first sight, the interface seemed less useful than the one for the digital cameras. However, it turned out to be much more convenient for comparing the features of the products. I had quickly made my choice"
"Much better! Easy to compare features!"	"The yellow "Why?" is a good idea; it provides a quick overview of advantages and disadvantages of the product compared to the most popular that comes first"
"The grouping allows finding a camera according to one's needs quickly, better than an ungrouped display."	"Cependant l'aide donnée par le "Why?" semblait objective et par conséquent digne d'intérêt." (However, the "why" explanation seems objective and therefore is worthwhile)
"It was easier to find the desired product with this interface, because it was not necessary to look at the "why" pop-up. The groups helped me go directly to the most interesting features (optical zoom in that case)."	"Le bouton "why?" m'a permit de valider mon choix de me montrer les détail que je n'avais pas remarqué sur le produit." (The "why" button allowed me to validate my choice by giving me details which I had not noticed from the product)
"The way the interface provides the information is a little surprising in the beginning, but once you got used to it, it comes in quite handy."	"Le lien du WHY? permettant de faire des comparatifs rapides est très bien. L'explication est courte, mais cible bien les caractéristiques." (The WHY? link, allowing a quick comparison (of the products), is very good. The explanation is short but to the point (in terms of characteristics being compared))
"Classification of products makes it a little bit easier to compare the products"	
"In this particular case, I found what I was looking for. What if the category that corresponds best to my needs does not exist?"	
"The interface is easy to use and it is comfortable to compare each item"	
"Better than the list view + good idea to put a key like "cheaper but heavier" etc. + intuitive ways to group items"	
"Le tableau pour comparer est une bonne chôse." (The table for comparing items is a good thing)	
"Par contre, la classification est bonne." (Compared to the list view, the classification is good)	
" La classification des offres permet de trouver le portable idéal tout de suite." (The classification of offers allows one to find the ideal portable computer right away)	
"J'ai apprécier la façon de séparer les produits." (I appreciate the separation of products)	
"La disposition d'un tableau convient bien." (The table view is convenient)	
Comments suggesting improvements	*Comments suggesting improvements*
"The idea is interesting, but there are way too many choices in each category."	"Although products are well presented, it is not easy to compare them because you cannot order them by any of their features."
"Grouping of items by weight and screen size is a good idea, but still listing of elements is a bit annoying…"	"Too messy! Too much time lost in comparing products for (the) same feature!"
"Still (the display is) too static, it would be better to be able to reorder or filter by my own expectations. But (it is) better than (the list view)."	"I liked the explanation tooltip, but found it annoying that I had to try to memorize the advantages and disadvantages of each as only one tooltip could be shown at a time"
	"No structure! The features are mixed up, no order."
	"less categories"
	"The "Why" was annoying and practically useless unless you already had some products in mind. However you get so frustrated trying to find some products of interest and remembering their positions, you never get to use the "Why"."
	"Listing of items is too intensive; too much information at once; requires a lot of concentration."
	"Too much information without a clear classification. All static. (It) would be better to be able to reorder the list by any characteristic."
	"Total absence of any kind of order for grouping the same kind of products."
	"Je préfères a ce moment la de classer les ordinateurs dans les différentes catégories pour mieux choisir!!" (I prefer at this point to classify the computers in the different categories in order to choose (the most preferred item)).
	"Le Why ne m'a pas amené grand-chose." (The Why did not help me much)

trust-building, the organization-based explanation interface is significantly more effective given the fact that users perceived it more competent, were more likely to experience less perceived effort while making decisions, and were more likely to use the agent again. As for the interrelationship of the various trust constructs, we found that a higher level of perceived competence in a recommender agent can significantly lead to users' increased intention to return to the agent for more products' information and also a decrease in their perceived cognitive effort consumed in decision making. In addition, the actual time spent looking for a product does not have a significant impact on users' subjective perceptions. This indicates that less time spent on the interface, while very important in reducing decision effort, cannot be used alone in predicting what users may subjectively experience.

We are exploring several directions to carry out the future work of this research. In the short-term future, we plan to design a user study that measures users' trust intentions when they revisit the explanation interface. Such results will shed light on the long term benefits of trust such as competence perception, users' intention to save effort, and a recommender's perceived accuracy. In addition to subjective attitudes, we will also analyze users' actual behavior such as decision accuracy and task time. In the long-term future, we would like to investigate the user control issue in relation to the user's trust and measure their causal or correlation relationships. We will design some experiments to examine whether explanation interfaces (or scrutable interfaces [18]) play an important role in helping users feel that they have control over their personal data and preference models, and whether such feeling-of-control can help users overcome fears associated with the loss of privacy. On the technical side, we intend to address preference conflict resolution using the tradeoff-based organization interfaces presented here. We would also like to investigate the effect of other system design features, such as recommender algorithms and user-system interaction models, on trust promotion.

References

[1] R. Agrawal, T. Imielinski, A. Swami, Mining association rules between sets of items in large databases, in: International ACM SIGMOD Conference, 1993, pp. 207–216.

[2] E. Armengol, A. Palaudàries, E. Plaza, Individual prognosis of diabetes long-term risks: a CBR approach, Methods of Information in Medicine 40 (2001) 46–51.

[3] R. Burke, K. Hammond, B. Young, The FindMe approach to assisted browsing, Journal of IEEE Expert 12 (4) (1997) 32–40.

[4] G. Carenini, J. Moore, Multimedia explanations in IDEA decision support system, Working Notes of the AAAI Spring Symposium on Interactive and Mixed-Initiative Decision Theoretic Systems, 1998.

[5] L. Chen, P. Pu, Trust building in recommender agents, Workshop on Web Personalization, Recommender Systems and Intelligent User Interfaces at the 2nd International Conference on E-Business and Telecommunication Networks, 2005, pp. 135–145.

[6] R.F. Falk, N.B. Miller, A Primer for Soft Modeling, first ed., The University of Akron Press, Akron, Ohio, 1992.

[7] B. Faltings, P. Pu, M. Torrens, P. Viappiani, Designing example-critiquing interaction, in: International Conference on Intelligent User Interfaces, 2004, pp. 22–29.

[8] S. Grabner-Kräuter, E.A. Kaluscha, Empirical research in on-line trust: a review and critical assessment, International Journal of Human-Computer Studies 58 (2003) 783–812.

[9] J.F. Hair, R.E. Anderson, R.L. Tatham, W.C. Black, Multivariate Data Analysis with Readings, fourth ed., Prentice-Hall, Englewood Cliffs, NJ, 1995.

[10] J.L. Herlocker, J.A. Konstan, J. Riedl, Explaining collaborative filtering recommendations, in: ACM Conference on Computer Supported Cooperative Work, 2000, pp. 241–250.

[11] D.A. Klein, E.H. Shortliffe, A framework for explaining decision-theoretic advice, Artificial Intelligence 67 (1994) 201–243.

[12] K. McCarthy, J. Reilly, L. McGinty, B. Smyth, Thinking positively – explanatory feedback for conversational recommender systems, in: Workshop on Explanation in CBR at the Seventh European Conference on Case-Based Reasoning, 2004, pp. 115–124.

[13] K. McCarthy, J. Reilly, L. McGinty, B. Smyth, Experiments in dynamic critiquing, in: International Conference on Intelligent User Interfaces, 2005, pp. 175–182.

[14] L. McGinty, B. Smyth, On the role of diversity in conversational recommender systems, in: Fifth International Conference on Case-Based Reasoning, 2003, pp. 276–290.

[15] D.H. McKnight, N.L. Chervany, What Trust Means in e-commerce Customer Relationships: Conceptual Typology, International Journal of Electronic Commerce (2002) 35–59.

[16] D. McSherry, Similarity and compromise, in: International Conference on Case-Based Reasoning Research and Development, 2003, pp. 291–305.

[17] D. McSherry, Explanation in recommender systems, in: Workshop Proceedings of the 7th European Conference on Case-Based Reasoning, 2004, pp. 125–134.

[18] J. Kay, B. Kummerfeld, Scrutability, user control and privacy for distributed personalization, in: Proceedings of the CHI 2006 Workshop on Privacy-Enhanced Personalization, 2006, pp. 21–22.

[19] J. Nunnaly, Psychometric Theory, McGraw-Hill, New York, 1978.

[20] V. Pareto, Cours d' Economie Politique, Technical report, Rouge, Lausanne, Switzerland, 1896.

[21] J.W. Payne, J.R. Bettman, E.J. Johnson, The Adaptive Decision Maker, Cambridge University Press, 1993.

[22] B. Price, P.R. Messinger, Optimal recommendation sets: covering uncertainty over user preferences, in: National Conference on Artificial Intelligence, 2005, pp. 541–548.

[23] P. Pu, L. Chen, Integrating tradeoff support in product search tools for e-commerce sites, in: ACM Conference on Electronic Commerce, 2005, pp. 269–278.

[24] P. Pu, L. Chen, Trust building with explanation interfaces, in: International Conference on Intelligent User Interface, 2006, pp. 93–100.

[25] P. Pu, B. Faltings, M. Torrens, Effective interaction principles for online product search environments, in: IEEE/WIC/ACM International Joint Conference on Intelligent Agent Technology and Web Intelligence, 2004, pp. 724–727.

[26] P. Pu, P. Kumar, Evaluating example-based search tools, in: ACM Conference on Electronic Commerce, 2004, pp. 208–217.

[27] H. Shimazu, ExpertClerk: a conversational case-based reasoning tool for developing salesclerk agents in e-commerce webshops, Artificial Intelligence Review 18 (2002) 223–244.

[28] R. Sinha, K. Swearingen, The role of transparency in recommender Systems, in: Extended Abstracts of Conference on Human Factors in Computing Systems, 2002, pp. 830–831.

[29] F. Sørmo, A. Aamodt, Knowledge communication and CBR, in: ECCBR-02 Workshop on Case-Based Reasoning for Education and Training, 2002, pp. 47–59.

[30] C.N. Ziegler, S.M. McNee, J.A. Konstan, G. Lausen, Improving recommendation lists through topic diversification, in: 14th International World Wide Web Conference, 2005, pp. 22–32.