Electronic Commerce Research and Applications 13 (2014) 440–455

Contents lists available at ScienceDirect

Electronic Commerce Research and Applications

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

Leveraging prior ratings for recommender systems in e-commerce

Guibing Guo [a,*], Jie Zhang [a], Daniel Thalmann [a], Neil Yorke-Smith [b,c]

[a] School of Computer Engineering, Nanyang Technological University, Singapore
[b] American University of Beirut, Beirut, Lebanon
[c] University of Cambridge, Cambridge, UK

ARTICLE INFO

Article history:
Received 16 August 2013
Received in revised form 26 June 2014
Accepted 14 October 2014
Available online 30 October 2014

Keywords:
Prior ratings
Recommender systems
Rating confidence
Similarity measure
Data sparsity
Cold start

ABSTRACT

User ratings are the essence of recommender systems in e-commerce. Lack of motivation to provide ratings and eligibility to rate generally only after purchase restrain the effectiveness of such systems and contribute to the well-known data sparsity and cold start problems. This article proposes a new information source for recommender systems, called *prior ratings*. Prior ratings are based on users' experiences of virtual products in a mediated environment, and they can be submitted prior to purchase. A conceptual model of prior ratings is proposed, integrating the environmental factor *presence* whose effects on product evaluation have not been studied previously. A user study conducted in website and virtual store modalities demonstrates the validity of the conceptual model, in that users are more willing and confident to provide prior ratings in virtual environments. A method is proposed to show how to leverage prior ratings in collaborative filtering. Experimental results indicate the effectiveness of prior ratings in improving predictive performance.

1. Introduction

User ratings are crucial for recommender systems in e-commerce in order to provide quality personalized product recommendations. However, users can lack motivation to provide ratings (why should I bother to report my experience of an item?), and ratings can generally be given only after purchase (how can I share my experience of an item I have not tried?). Without sufficient rating information for preference modelling, the effectiveness of recommender systems is hindered—as seen in well-known problems such as *data sparsity* and *cold start* (Su and Khoshgoftaar 2009).

The former problem, data sparsity, refers to the difficulty in finding a sufficient number of reliable users, since users in general only rate a small portion of items, while the latter problem, cold start, refers to the difficulty in providing accurate recommendations for those users who have rated a few items, e.g., less than five items. Cold start is an extreme case of the data sparsity problem. The key issue is that only limited rating information is available for preference modelling, whereby inherently and severely hindering the recommendation performance.

Although many approaches have been proposed to address these problems either by furthering the use of existent ratings (Ahn 2008; Guo et al. 2013b), or by including to additional information (Massa and Avesani 2007; Konstas et al. 2009; Guy et al. 2009; Jamali and Ester 2011; Guo et al. 2012, 2014a), few researchers have attempted to elicit more user ratings from the perspective of user interfaces, so as to inherently mitigate the severity of these problems. On the other hand, Virtual Reality (VR) environments (e.g., Second Life (Rymaszewski 2007)), have received considerable attention because of their ability to provide users with immersive virtual user experiences. Users can experience media more richly and can interact in real time with *virtual products*—the 'second existence' of real products in a mediated environment (Hemp 2006). Although these environments offer potentially useful information for preference modelling, research on e-commerce in VR is still in its infancy.

This article proposes a new information source, called *prior ratings*, built upon *virtual product experiences* (Li et al. 2003). Prior ratings can be issued prior to purchase by interacting with virtual products represented in a mediated environment. The aim of this article is to study (1) the concept and nature of prior ratings with respect to product attributes and environmental factors; and (2) the usefulness of prior ratings in coping with the data sparsity and cold start problems of recommender systems.

In particular, first, we propose a conceptual model of prior ratings to provide a principled foundation, integrating the environ-

* Corresponding author at: Institute for Media Innovation, 50 Nanyang Drive, Research Techno Plaza, XFrontiers Block, Level 03-01, Singapore 637553, Singapore. Tel.: +65 84005322.

E-mail addresses: gguo1@ntu.edu.sg (G. Guo), zhangj@ntu.edu.sg (J. Zhang), danielthalmann@ntu.edu.sg (D. Thalmann), nysmith@aub.edu.lb (N. Yorke-Smith).

http://dx.doi.org/10.1016/j.elerap.2014.10.003
1567-4223/

mental factor *presence* whose effects on product evaluation have not been studied previously. Five hypotheses and two research questions are proposed to verify the validity of the conceptual model. We recruited volunteers and performed user studies in both 2D (website) and 3D (virtual store) user interface modalities. The results demonstrate the validity of the conceptual model under our experimental settings, and indicate that users are more willing and confident to give prior ratings in a VR store (due to a stronger sense of presence) than in a website.

Then, second, by integrating the prior rating and confidence data collected from the user studies into a novel adapted collaborative filtering technique that we develop, we empirically demonstrate the usefulness of prior ratings in improving recommendation performance in terms of accuracy and coverage.

Our work sheds light on inherently alleviating the data sparsity and cold start problems by the design of user interfaces with rich media and interactions that elicit confident prior ratings from users.

Contribution. Summarized, the major contributions of this article are in three-fold: (1) we introduce a new information source (and its conceptual model) called *prior ratings*, which holds potential to benefit recommender systems in e-commerce; (2) we design a user study to validate the conceptual model of prior ratings; and (3) we propose and evaluate a collaborative filtering technique to demonstrate how to leverage prior ratings in predicting the ratings of products. A preliminary version of our work is published in (Guo et al. 2013a).

Outline. Section 2 gives an overview of related research in the literature. Section 3 details the proposed conceptual model of prior ratings, and proposes five related hypotheses and two research questions. Section 4 reports on a user study designed to validate the conceptual model. Then, Section 5 discusses the relationship between prior ratings and other information sources for recommender systems, and the limitations and implications of the user study and results. Based on the rating and confidence data collected from the user study, Section 6 introduces a variant of traditional collaborative filtering technique and demonstrates the usefulness of prior ratings in improving predictive performance. Finally, Section 7 concludes our work and outlines the future research.

2. Related work

Many approaches have been proposed to resolve the data sparsity and cold start problems. From the perspective of information source, we classify them into two categories. The first category adopts rating information only. There are two kinds of approaches, memory-based and model-based. For memory-based methods, various authors have proposed new similarity measures to better model user correlation to resolve the concerned issues, given the inefficiency of traditional similarity measures (Lathia et al. 2008). Specifically, Lathia et al. (2008) propose a concordance-based measure based on the amount of concordant, discordant and tied pairs of ratings between two users. It measures the extent to which the two users agree with each other. Ahn (2008) develops the PIP measure by studying the semantics of ratings in terms of *Proximity*, *Impact* and *Popularity*. The basic idea is that users with semantic agreements should be more similar than those with semantic disagreements. Bobadilla et al. (2012) design the *singularities measure* from the perspective of item singularity. The intuition behind is that ratings agreed on high-singular items should be counted more than those agreed on low-singular items in computing user similarity. Guo et al. (2013b) propose a novel *Bayesian similarity* by taking into account both the direction and length of rating vectors. The weights of evidences (i.e., rating pairs) are carefully computed and integrated into the Bayesian inference. Experimental results show that better performance can be achieved than the other similarity measures.

However, memory-based approaches do not scale well to large-scale data sets. In contrast, model-based methods possess better scalability and often perform better than memory-based ones (Koren et al. 2009). The reason is that not only ratings of two users but also ratings of other users are adopted to learn the features of users and items, and thus better handle the data sparsity and cold start issues. Gunawardana and Meek (2008) report to capture pairwise item interactions by using a Boltzmann machine, whose parameters are associated with item contents. They show that better performance is achieved in the case of cold-start situations. Gantner et al. (2010) attempt to learn a function mapping user/item attributes to latent features of a matrix factorization model. With such mappings, the latent factors learned by a matrix factorization can be applied to new users or new items. Liu et al. (2011) propose a representative-based matrix factorization method that aims to find out the most representative users and items in the system. Then, for the cold-start users, their preferences can be elicited by asking them to rate the most representative items; the same holds for the cold-start items. To combat the data sparsity problem, Ahmed et al. (2013) propose a method to learn user preferences over item attributes by applying a personalized Bayesian hierarchical model, which combines both globally and locally learned user preferences.

In summary, all these approaches, both memory-based and model-based, attempt to integrate user/item attributes into a certain recommendation model in order to handle the concerned issues. However, the attributes of users/items may not be available for a recommender system due to the concern of, e.g., privacy.

The second category adopts additional information other than ratings. For example, Konstas et al. (2009) take into consideration both the social annotation (tag) and friendships inherently established among users in a music track recommender system. By leveraging data from multiple channels including memberships in a project, Guy et al. (2009) build a system for recommending people of interest to active users. Ma et al. (2011) propose a matrix factorization model regularized by users' social friendships. The intuition is that a user-specific vector should be close to that of his friends. A stronger relationship than friendship is social trust, based on which Massa and Avesani (2007) develop a trust-aware recommender system. They show that data sparsity can be better handled without a significant decrement in predictive accuracy. Guo et al. (2014b) define trust in recommender systems as one's belief in the other's ability in providing accurate ratings. Guo et al. (2012, 2014a) merge the ratings of trusted neighbours to form a new rating profile for the active users by which the concerned problems are shown to be alleviated.

Other than these memory-based approaches, trust is also integrated into matrix factorization models for better recommendation performance. Ma et al. (2009) propose the social trust ensemble that forms a linear combination of a matrix factorization model and a trust-based neighbourhood model. Jamali and Ester (2010) propose a matrix factorization model where a user's user-specific vector is influenced by the average of her trusted neighbours. Tang et al. (2013) take into account both the global and local trust in the recommendation model, and show that predictive performance can be improved to some extent. Yang et al. (2013) report that the active user's ratings will be influenced by the ratings of users who trust her and those who she trusts. Experimental results show that their approach works the best among all other trust-based approaches. One of the problems of matrix factorization models lies in the difficulty of explaining how recommendations are generated, as these patterns are based on latent features. Another problem is that users' social information may not exist,

especially for the applications without built-in or linked social networks. Further, such information merely indirectly implies user preference, e.g., friends may or may not have similar preferences, and hence could be error-prone.

Therefore, although additional information sources such as friendship and trust have been widely applied in (social) recommender systems and although improvements have been demonstrated to some extent, the cold start problem remains a difficult issue to address. The reasons can be explained in three aspects. First, these kinds of information suffer from a number of inherent issues. As explained above, the semantics of friendship are ambiguous and error-prone. For example, friends may have different preferences because friendships can be built based on other relations (e.g., working affiliation) rather than common interest in items. It is usually low cost for a user to get connections with other users or even strangers (e.g., Facebook). Trust is only supported by few real systems (e.g., epinions.com and ciao.co.uk). Trust information is also very sparse (Guo et al. 2014b), i.e., the density of trust information is even much smaller than that of rating information in the same data set, since not all users who give ratings will be socially connected with other users. In many other systems, trust information is not available and thus it has to infer trust from users' behavioural patterns (Guo et al. 2014b,c). Another problem of social relationships is that they usually exist in the forms of social connections with no connection strength specified or available. For example, in trust networks, we only know the relationships about who trust whom, but it is unknown to what extent one will trust another. One explanation is due to the concern of, e.g., privacy. It is a commonplace that not all socially connected users should be equally weighted for recommendations. Fang et al. (2014) suggest to refine the trust values by training a support vector regression model based on four general decomposed trust factors, before taking as input to a matrix factorization model. They show that better performance can be achieved based on the refined trust values. The unweighted social relationships may further limit the utility of social recommender systems. Second, for 'extremely' cold users who have rated no items and linked to no one, it is difficult for a social recommender to provide accurate personalized recommendations. This is because user preferences cannot be inferred and modeled from their past behaviours. Third, even with the social information, the performance of cold users is still much worse than that of normal users, as demonstrated by the work of Yang et al. (2013). There is much room left for better recommendation performance. Hence, the cold start problem has not been well handled by the existing approaches and information sources. More efforts are required to further alleviate the cold start problem, including developing more advanced recommendation algorithms and designing new information sources.

Our work follows the second category, i.e., incorporating additional information in recommender systems, but differs in that we focus on introducing a new information source rather than the specific techniques to utilize such information. However, we do design a collaborative filtering technique to demonstrate the use of prior ratings.

Only a few works have attempted to study the concerned problems from the perspective of user interfaces. For example, Carenini et al. (2003) recognize that traditional recommender systems support only a limited model of interaction to elicit new users' ratings. They explore a set of elicitation techniques leading to a more conversational and collaborative interaction model. Offline experiments show that the effectiveness of recommender systems can be improved by applying these techniques. However, whether the new model of interaction is accepted by users and useful for online recommendations in practice is unknown. McNee et al. (2003) find that allowing systems to choose items for new users to rate works better than letting users choose the items, in order to bootstrap and build a recommendation model. Dong et al. (2012) develop a browser plugin to provide users with suggestions on writing better product reviews. Other users can hence better understand the performance of products before making a purchase decision. Most of these studies focus on interface design or assistance, so that users are more comfortable, enabled, or loyal in providing ratings. They are not particularly dedicated to resolving the two concerned recommender systems problems. By contrast, our motivation is to tackle the concerned problems through a new information source in a richer virtual environment.

Contemporary websites are implementing novel interfaces and interactions to better elicit user preferences. For example, brides.com allows users to virtually try on wedding dresses by uploading their own photos and adjusting the specific positions of dresses to fit. As another example, ray-ban.com offers users a virtual mirror through which users can calibrate their faces using a computer camera, and virtually try on different kinds of glasses. However, the available media and interactions are limited in comparison with the capabilities of virtual reality (VR) (e.g., Second Life (Rymaszewski 2007)). The emergence of 3D VR environments offers more adequate information which can be used to model user preference. Although the need to design new recommender agents for e-commerce in VR has been recognized (Xiao and Benbasat 2007), research on recommender systems in VR is still in its infancy. Eno et al. (2011) summarize several ways to model user preferences in VR. Shah et al. (2010) recommend to users locations of interest by analyzing users' login data to help them navigate in VR. Hu and Wang (2010) propose a system for virtual furniture recommendation according to users' interest and requirements. Although a controlled prototype implemented, the features of VR are not exploited to elicit more user ratings.

In this article we propose prior ratings as a means to make use of the information conveyed by the rich media and the real-time interactions in VR. Prior ratings represent a new information source distinct from the existing information sources noted earlier. First, prior ratings are issued by real users: hence they directly reflect users' preferences of products as well as standard type of user ratings. In this regard, they could be more reliable than other kinds of information, such as friendship and trust. Second, prior ratings differ from other extra information sources in that they do not depend on additional structures (e.g., social network) as required by the latter. Prior ratings only rely on the representations of virtual products in mediated environments, but these environments are the commonplace basis of e-commerce applications. Prior ratings are useful to deal with data sparsity and cold start because (1) more user ratings are incorporated to alleviate the sparsity of data; and (2) prior ratings can help model user preferences even if posterior ratings are few or none, and thus ensure the functionality of the recommenders. To our knowledge, there is no work that has defined the concept of and investigated the effectiveness of prior ratings for recommender systems.

3. Prior ratings

We define the term *prior ratings* as users' assessment or judgement of preference of products in the light of their *virtual product experiences*, referring to the psychological and emotional states that users undergo while interacting with virtual products in a mediated environment (Li et al. 2003). Hence, prior ratings are reported by users based on their interactions with virtual products in a mediated environment, and they can be issued prior to purchase or after purchase (if any). Therefore, although we focus on VR environments, prior ratings could be given in any other mediated environments, such as augmented reality, as long as they can provide reliable virtual product experiences.

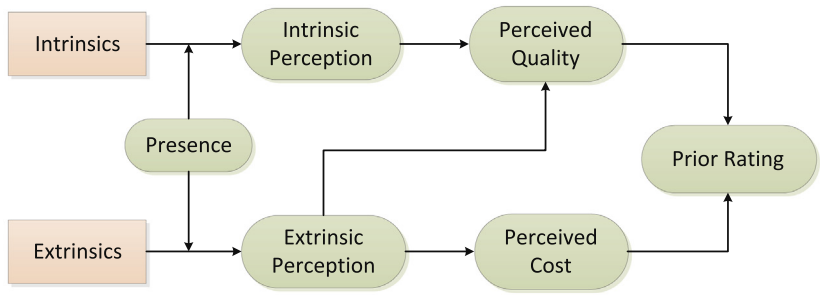


Fig. 1. The conceptual model of prior ratings.

We refer to the 'standard' type of ratings derived from 'posterior' product experiences as *posterior ratings*. By 'posterior', we mean experiences of a tangible product obtained via direct trials or use of the product in a physical environment. Since tangible products can be fully experienced usually only after purchase, posterior ratings are primarily post-purchase ratings. Prior ratings and posterior ratings are distinct and complementary in that they reflect different forms of user experiences. Note that for products without a tangible form, such as streamed movies, since users can only experience them virtually through some medium, users' ratings are necessarily prior ratings.

In this article, two kinds of mediated environments are investigated: traditional 2D websites (WS) and 3D VR environments. They differ in richness of both media and of interactions through which product information can be delivered. WS only supports limited media and user interactions; VR real-time interactions enable users to possess a strong sense of being in a mediated environment and gain a lifelike shopping experience (Li et al. 2001). Specifically, products are represented in 3D virtual models through which users can view, rotate, zoom, customize and even try them on. Considering the rich virtual product experiences that users obtain in VR, we posit that VR will motivate users to express their opinions by providing prior ratings to the products of interest, and hence make more informative purchase decisions while shopping in VR.

Hypothesis 1. Users are more willing to provide prior ratings to the items (e.g., products) that they have interacted with in VR than in WS.

Although prior ratings can be submitted in both WS and VR as long as the user interfaces enable the rating functionality, the confidence level may differ. Specifically, due to limited media and interactions available in WS, users may have less adequate information than in VR as a basis for their ratings. Jiang and Benbasat (2004) also contend that virtual products in VR help improve the perceived *diagnosticity* of products—the extent to which users believe a particular shopping experience is helpful to understand the quality and performance of a product. Therefore, users may feel more capable of forming direct, intuitive and concrete opinions about products in VR than in WS in terms of both rating confidence and rating values.

Hypothesis 2. (a) Users have more confidence in providing prior ratings in VR than in WS; (b) the average value of prior ratings in VR is closer to that of posterior ratings than that of prior ratings in WS.

3.1. Conceptual model of prior ratings

We now present a conceptual model of prior ratings as shown in Fig. 1. Such a model allows a principled basis for the elicitation and analysis of prior ratings. The objective of our conceptual model is to investigate a comprehensive understanding of the nature of prior ratings. Specifically, (1) how prior ratings are given by users, and (2) how other factors such as the presence of virtual reality and the attributes of virtual products impact on users' evaluation of prior ratings. Only after a proper understanding of prior ratings, we will be able to show how to leverage them in a newly-designed collaborative filtering technique so as to resolve the data sparsity and cold start problems—which are our main concern in this article—in Section 6. Note that the conceptual model is not used to justify the effectiveness of prior ratings in resolving the data sparsity and cold-start problems, but to give a better comprehension of prior ratings.

For a specific product, a number of intrinsic and extrinsic *attributes* are associated. In different environments, the perceptions of these attributes can differ according to the types of media and interactions that deliver information about them. For example, VR environments may have better perceptions of products than traditional websites as the former generally enables richer media and real-time interactions. The intrinsic and extrinsic *perceptions* indicate the quality of products as perceived directly and indirectly, respectively. In contrast, the *perceived cost* (e.g., time, price) refers to the cost that users have to bear in order to obtain the products. A prior rating is an overall evaluation of preference of products in terms of both perceived quality and cost, i.e., a combination of what we 'get' and what we 'give'.

We proceed to elaborate the details of the conceptual model in the following subsections.

3.2. Presence

Presence is defined as users' sense of "being there", the extent to which they experience the virtual environments as real or present and temporarily ignore where they are physically present (Slater et al. 2010). Two major determinants have been identified, namely *vividness* and *interactivity* (Steuer 1992). First, vividness reflects the representational richness of a mediated environment as defined by its formal media through which information can be presented. Two important elements of vividness are *sensory breadth* which refers to the number of sensory dimensions simultaneously presented, and *sensory depth* which refers to the resolution within each perceptual channel. Second, interactivity is defined as "the extent to which users can participate in modifying the form and content of a mediated environment in real time" (Steuer 1992). Three important elements, namely *speed*, *range*, and *mapping* describe the specification of a mediated environment in terms of response time, the amount of manipulable attributes, and the projections between human and environmental actions.

Hence, presence in this article is captured as the extent to which being in a mediated environment feels like being in a real environment,[1] given the richness in media and interactions. Picciano (2002) reports that the sense of social presence (i.e., the sense of belonging

[1] Compare question 2 for the tested environments in Fig. 3.

in a course and group) has a positive and statistically significant influence on the performance of students' written assignments in an online course. Phang and Kankanhalli (2009) study how the perceptions of virtual world can enhance online learning. They show that in 3D environments, presence can enhance students' concentration and enjoyment during the learning process, and thus improve students' learning outcomes. These two works show that (1) it is important for learners to perceive a realistic classroom experience; and (2) such sense of being there can help them concentrate more on the learning contents. In the case of e-commerce recommendations, the presence of virtual reality enables users a lifelike shopping experience, and thus users may concentrate more on the product experience and evaluation. In addition, Heeter (1992) stresses the importance of being able to change virtual environments, for instance, moving and painting a 3D object. A higher sense of presence can enable user interactions with 3D environments to be easier and more responsive. In our case, the 3D models of virtual products can respond to users' actions, e.g., rotating and zooming, and hence users may gain more direct comprehension about the properties (attributes) of products. Considering that the information concerning product attributes is conveyed by media channels and user interactions, presence can be an important environmental factor that will influence the perceptions of product attributes.

Hypothesis 3. Presence has positive influence on the perceptions of both intrinsic and extrinsic attributes.

Note that higher sense of presence does not necessarily mean better perceived quality. Perceived quality is based on the perceptions of product attributes; presence is a moderator of the perceptions of product attributes.

3.3. Intrinsic attributes

Intrinsic attributes (e.g., workmanship, size) have a direct impact on perceived quality during the goal-directed process of pre-purchase product evaluation (Gardial et al. 1994). Goering (1985) also considers that intrinsic quality of a product has an important influence on the perceived quality of a product. Specifically, the higher the intrinsic quality of a product is, the better it will perform. In addition, intrinsic attributes can also work as cues to infer product quality (Olson 1974). For example, the attribute 'nutrition content' can be used as a cue to assess the quality of a breakfast cereal.

The specific intrinsic attributes embedded can vary between different products. In this article, we classify intrinsic attributes into three types, namely *appearance*, *material*, and *functionality*. Appearance refers to the attributes related to the superficial representation of products, such as patterns, form, size, etc. Material refers to the attributes associated with what products are made of, such as fabric properties, weight, etc. Functionality refers to the attributes indicating the utility of products or the actions that products can perform or that can be performed on products. For example, an electronic watch contains the functionality of stopwatch and it may 'fit' someone well.

More generally, Nelson (1974) identified two different types of product attributes: *search attributes* and *experience attributes*. The former type refers to the attributes the information of which can be conveyed most effectively through secondhand sources, whereas the latter refers to the attributes the information of which can be evaluated most effectively by using products directly. Therefore, by definition, appearance and material attributes are more likely to be search attributes as their information can be easily obtained by searching. In contrast, functional attributes tend to be experience attributes since the effectiveness of functions requires the interactions with products, i.e., direct experiences.

Question 1. What are the major intrinsic attributes that influence the perceived quality of products in WS and VR?

3.4. Extrinsic attributes

Unlike intrinsic attributes, extrinsic attributes (e.g., price, product type) have no direct indications of perceived quality. Rather, they are often used as cues to infer the quality of products when the information of intrinsic attributes is incomplete (Dodds et al. 1991). For example, considerable theoretical and empirical evidence (Zeithaml 1988; Rao and Monroe 1989) shows that price is often used by users to infer the quality of products when it is the only available cue or when there is inadequate information about intrinsic attributes (Dodds et al. 1991). The rationale is that more cost is often required to produce high-quality products than low-quality products and the probability to charge high prices for low-quality products is low due to competitive pressures (Lichtenstein et al. 1993).

Other than price, brand and store are also well-studied in the literature. Brand name serves as a 'shorthand' for perceived quality by providing users with a bundle of information about the product (Jacoby et al. 1977). It helps reduce the perception of risk prior to purchase in terms of financial, time, performance and psychological risk (Ha 2002). In comparison with price and brand name, store name also has a positive but small (not significant) impact on perceived quality (Rao and Monroe 1989). Dodds et al. (1991) also show that favourable brand and store information positively influence the perceptions of quality and value.

In addition, products can be categorized into different types in the light of different kinds of product attributes. For example, according to the definitions of search and experience attributes, products can be classified as search products and experience products (Nelson 1974). Search products are those whose dominant attributes are search attributes and hence full information of them can be known prior to purchase without direct experience. In contrast, experience products are those whose dominant attributes are experience attributes and hence full information of them cannot be known until use of products (direct experience). Therefore, we have to highlight that not all products are suited in VR; specifically, experience products may perform better in VR whereas search products may perform better in WS in terms of user efforts in retrieving product information. In this article, we refer to product types as the *categories* of products. For example, there could be action and comedy movies. The categories are not deterministic and can be varied in different systems. Another reason for our definition of product type is that recommender systems usually focus on some specific product domains such as music and video rather than generic products.

Other extrinsic factors investigated in the literature may also have an effect on perceived quality, such as warranty (Bearden and Shimp 1982), packaging (Stokes 1974), advertising (Milgrom and Roberts 1986), etc. We do not consider every kind of extrinsic attributes, but examine the extrinsic factors that will most significantly influence the quality of products. As noted, WS is most efficient to deliver information of search attributes and VR to convey information of experience attributes. Further, intrinsic attributes rely more on direct experience whereas information of extrinsic attributes can be found without use of products. Therefore, for a product that can be represented in both WS and VR, we come to the following question and assumption.

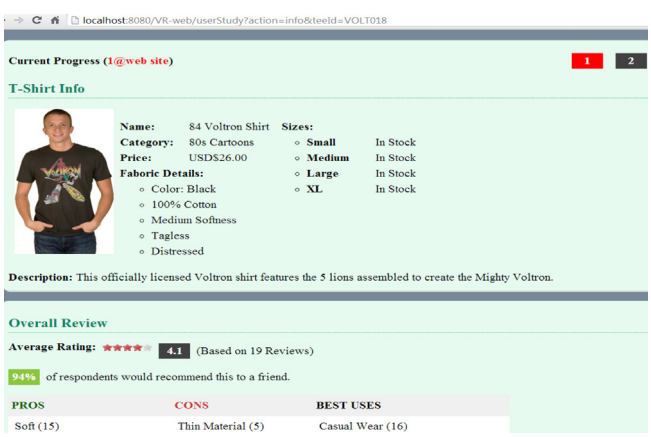
Question 2. What are the major extrinsic attributes that influence the perceived quality of products in WS and VR?

Hypothesis 4. Users depend more on extrinsic attributes than intrinsic attributes to evaluate the product quality in WS, whereas users depend more on intrinsic attributes than extrinsic attributes to evaluate the product quality in VR.

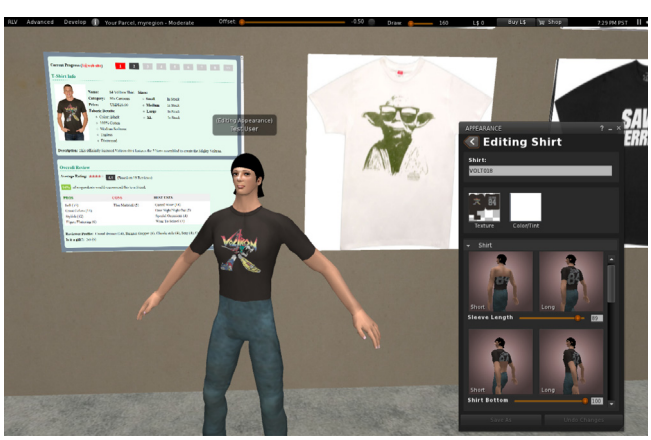
Besides quality, extrinsic attributes also contribute to *perceived cost*, a combination of monetary and non-monetary attributes (Zeithaml 1988). The former usually refers to price, and the latter includes energy, efforts and other costs (e.g., time, shipping). Other extrinsic attributes (e.g., brand name) may not have influence on perceived cost.

3.5. Prior ratings

For a given product in a pre-purchase phase, users go through a process (perhaps subconscious) of evaluating the benefits that they can get and the cost that they have to incur. The outcome of this process helps determine whether users will like the product in question. Other than perceived quality, we posit that prior ratings could also be positively enhanced if the perceived cost is acceptable. Intuitively, for a specific product interested in by a user in terms of quality, if the price of the product turns out to be acceptable, it is likely that the user will like the product as a whole. Recall that due to competitive pressures the price of a product is usually correlated with its quality, i.e., within a normal range (Lichtenstein et al. 1993). Therefore, we reach the following assumption.



(a) website



(b) virtual store

Fig. 2. Website and virtual store modalities.

Hypothesis 5. Perceived quality has significantly positive influence on prior ratings, and perceived cost will also positively influence prior ratings, given that price is within a normal range.

4. User study of prior ratings

In the previous section we introduced the concept of prior ratings and provided a conceptual model for them by drawing on various sources in the literature. This section reports a user study to validate the concept of prior ratings.

We developed two user interfaces with different levels of presence, namely *website* and *virtual store* (as seen in Fig. 2)—corresponding to the mediated environments of WS and VR, respectively. Both user interfaces 'sell' our t-shirts whose source was the real-life commerce website 80stees.com. From this website we derived 50 t-shirts in total as the products which will be evaluated by users in both interfaces. These t-shirts have average posterior ratings (on 80stees.com) in the range $[3.2, 4.9]$ (out of 5). The virtual store was built using OpenSimulator.org, an open source project for simulating 3D environments. T-shirts were displayed and arranged without a predefined order on the walls of virtual store. Users can interact with them by viewing, rotating, zooming, and even virtually trying on and customizing the t-shirts (on their avatar). They can also adjust the avatar's shape as desired to meet their personal specifications. In contrast, no interactions were available in WS: users can only imagine what the t-shirt would be like from text descriptions and static images. Six attributes were identified and studied in the experiments: three intrinsic attributes (appearance, material, fit) and three extrinsic attributes (price, category, store). Appearance included color, image pattern and size; material corresponded to fabric features; fit indicated how t-shirts can perform on avatars; category was the classification of t-shirts used by 80stees.com, such as '80s cartoon t-shirts'; store referred to the design of user interfaces.

4.1. Pilot study

In order to guide the above choices, and to understand whether our experimental settings are reasonable and useful, we conducted a questionnaire-based pilot study. Participants were asked to imagine online shopping for t-shirts and rate what product attributes (see Table 1) they were most concerned with. For each attribute, users rated its importance from 1 ("of very little or no importance") to 5 ("of utmost importance"). In August 2012, we recruited 23 volunteers to participate in the online questionnaire, by sending emails of participation to the students and staff on the campus of a technical university. Based upon subjects' ratings,

Table 1
Results of pilot study: importance of attributes.

	Attributes	Mean	*p*-value
Intrinsic	Appearance	4.348	1.29e−07
	Material	4.174	2.38e−06
	Fit	4.304	2.041e−08
Extrinsic	Price	4.130	7.62e−09
	Situation	3.044	0.433
	Customization	2.522	0.0227
	Rating	2.478	0.0152
	Brand	2.826	0.769
	Store	2.860	0.280
	Recommendation	2.826	0.253
	Category	2.522	0.0266
	Warranty	2.652	0.123
	Promotion	2.870	0.301
	Shipping	2.957	0.426

a one sample t-test for each attribute was conducted. To be specific, the *null* hypothesis was: *the mean importance of the attribute in question is moderate (mean = 3)*, and the alternative hypothesis was set: *the mean importance of the attribute in question is greater than moderate (mean > 3)*. A small $p < 0.05$ value will reject the null hypothesis and accept the alternative hypothesis.

The results from Table 1 show that four major attributes are mostly ($mean > 4$ and $p < 0.001$) concerned with by users when purchasing t-shirts online, namely appearance, material, fit and price. Other attributes are not significant. For example, whether the t-shirts can be customized is not important at all. One subject commented that "If it is a t-shirt I do not care 'experts' recommendation". In addition, subjects also suggested other expected attributes regarding the functionalities of t-shirts: "Matches my other clothes well", "fitness", "have enough text or image details" and have a "Recommender System" or to "have a dummy try-on".

In conclusion, considering that users usually have past experiences about t-shirts in real life, they are confident to evaluate the performance of t-shirts if sufficient information of the four major attributes is available online, especially if it is convenient to visualize or measure the wearing effects. We therefore selected the four significant attributes, together with the standard attributes store and category, for the experiments.

4.2. Method and participants

The user study consisted of one session, structured as follows. All subjects started with a video introduction to the user study, including operations in two different environments. Specifically, for WS, subjects were guided to scan the overall and specific reviews of other customers along with the t-shirt specifications. For VR, subjects were introduced to try the VR hands-on, to become familiar with the functionalities of VR such as navigating, zooming in and out, virtual try-on, etc. Once subjects were comfortable, they proceeded. Each subject experienced and evaluated eight different, randomly-chosen t-shirts in each environment by giving ratings to the questions about product attributes.

Rating values were integers from 1 ("strongly disagree") to 5 ("strongly agree"). Subjects could also add textual comments for each t-shirt. To eliminate the influence of ordering, subjects were randomly determined into two groups. Specifically, of 30 volunteers recruited on a university campus, 16 subjects executed the user study first in WS and then VR, and 14 proceeded inversely. After subjects finished evaluating eight t-shirts in each environment, they were asked to rate the environment regarding the confidence (and state their reasons) and comfort in giving ratings, and the feelings of sense of presence. Finally, subjects could opt to state whether and in which environment they are willing and prefer to provide prior ratings. Subject demographics are reported in Table 2, and the questions shown in Fig. 3.

Table 2
Demographics of subjects in the user study.

Feature	Description
Gender	Male (24), Female (6)
Age	⩽20 (1), 20–29 (24), 30–39 (5)
Degree	Doctoral (16), Master (4), Bachelor (9), College (1)
Staff	Graphics (2), control system (2), engineering (1), telepresence (1)
Students	Computer Engineering (13), Electrical and Electronic Engineering (11)
Shopping[a]	1–2 times/week (3), 1–2/month (8), 1–2 months (16), Never (3)
VPEs[b]	<1 month (4), <3 months (5), <1 year (4), 1–2 years (2), Never (15)

[a] That is, the frequency of shopping online.
[b] That is, the frequency of prior virtual product experiences.

4.3. Results and analysis

Data cleaning is adopted to rule out noise of user data. In particular, the data from three users were discarded: one subject only completed the user study in virtual store, and two others stated that they were unfamiliar with the functionalities of VR even after an interactive introduction. A further user informed us that his evaluations on the first t-shirt in virtual store were not reflecting his real feelings due to misunderstanding of some terms in the first place. Thus his ratings on that t-shirt were also removed. After data cleaning, we had data of 27 users: 15 who tried WS then VR, and 12 in the inverse order. In total we collected 215 rating records from WS and 218 from VR. The statistics (in percentage) of collected ratings is illustrated in Table 4, including both prior and posterior ratings.

For Hypothesis 1, of 19 subjects who answered our questions regarding the willingness to rate t-shirts, 18 gave positive responses. More specifically, most subjects preferred to rate products in VR (14) rather than in WS (2). Two other subjects did not explicitly state their preference. Most subjects expressed that the reasons were "it can provide more detail information" and "this environment (VR) has really high engagement. I'd like to share my feeling". Only one subject did not want to provide prior ratings ("time consuming") but did indicate the willingness if "benefits or lucky draw" were offered. Thus, Hypothesis 1 is supported.

For Hypothesis 2(a), we conducted a number of paired two sample t-tests to investigate the mean differences of environmental factors, namely confidence, comfort, and presence. Table 3 reports the results. Since all $p < 0.01$, we find that users in VR have greater confidence and feel more comfortable in their prior ratings than in WS. The mean confidence in VR (3.778) is larger than that in WS (3.296). This may be partially explained by the fact that users have stronger sense of presence in VR than in WS.

Subjects also expressed their reasons of giving a specific rating to the confidence of tested environment, for instance:

- *I could not dress the T-shirt on my own body to check the looking effect, size and material. The image of T-shirt on model might not accurate and it's captured only from one side of view.* [in WS]
- *Everyone has a different figure and it is hard to image what it will look like when i wear this shirt.* [in WS]
- *Virtual environment gives a better understanding about how to t shirts looks on you.* [in VR]
- *It seems like that I was just staying in a real store, and then I can see the outfit directly. So I can make the judgement confidently.* [in VR]

These comments suggest that users in VR possess more confidence in their ratings because they can try t-shirts on their 'own' body rather than have to imagine the real wearing effect in WS. They also feel stronger sense of presence in VR as if being in a real store. Thus, Hypothesis 2(a) is supported.

For Hypothesis 2(b), as stated in Table 4, the final collected data consists of 215 prior ratings in WS ($R.ws$) and 218 records in VR ($R.vr$). Since all users only rated a handful of t-shirts, they are all regarded as cold-start users. The correlation between posterior ratings (R_p) and $R.ws$, denoted as $corr(R_p, R.ws)$ is -0.42 whereas $corr(R_p, R.vr) = 0.23$, signifying that the distribution of posterior ratings is distinct from prior ratings in WS, but marginally yet positively similar to prior ratings in VR. To have a better viewpoint, we classify the two rating values (i.e., 4, 5) which are larger than median scale value (i.e., 3) as positive, and the remainder (i.e., 1, 2, 3) as negative. Then we obtain clearer correlations: $corr(R_p, R.ws) = -1$ and $corr(R_p, R.vr) = 1$. In addition, the average posterior rating is 4.13 whereas the values for $R.ws$ and $R.vr$ are 2.94 and 3.56,

To what extent do you agree or disagree with the following statements?

For each t-shirt:

1. *The t-shirt has a good looking in terms of color, patterns, style, etc.*
2. *The t-shirt is made of good material.*
3. *The t-shirt fits you well.*
4. *The category of this t-shirt is of your favor.*
5. *The price of this t-shirt is acceptable, including price and shipping fees.*
6. *The website (virtual store) is well-designed.*
7. *In total, the quality of this t-shirt is good.*
8. *You need to spend a lot to obtain this t-shirt in price, time, effort, etc.*
9. *In total, this t-shirt is worthy purchasing.*
10. *Overall, you like this t-shirt.*

For each environment:

1. *You are confident about your ratings. When you gave ratings, you feel confident and no hesitations to make a judgement.*
2. *It feels the same that inspecting the t-shirt in the environment is just as if you were in a real store and had a real t-shirt in hand.*
3. *You are comfortable to give ratings in the tested environment.*
4. *You are (not) confident in your ratings because (state your reasons)*

For willingness (optional):

1. *Are you willing to rate the t-shirt of your interest or interacted with?*
2. *If yes, state your reason and indicate how confident in your ratings?*
3. *If no, state your reason. In what conditions, you will rate the t-shirts?*

Fig. 3. Questions in the user study.

Table 3
Evaluations of the environmental factors.

	Mean.ws	Mean.vr	Diff.	*p*-value
Confidence	3.296	3.778	0.482	3.300e−3
Comfort	3.444	3.963	0.519	6.653e−3
Presence	2.185	3.222	1.037	1.420e−4

respectively. In conclusion, prior ratings in VR are much closer to posterior ratings than those in WS. Thus, Hypothesis 2(b) is supported.

For Hypothesis 3, we conducted multiple linear regressions, each of which used 'presence' as independent variable and one of

Table 4
The distributions of collected ratings.

Scales	R_p (%)	$R.ws$ (%)	$R.vr$ (%)
1	3.82	11.63	3.67
2	4.08	18.60	10.55
3	7.15	35.81	27.52
4	27.77	25.12	42.66
5	57.18	8.84	15.60
1, 2, 3	15.05	66.04	41.74
4, 5	84.95	33.96	58.26
Total	1469	215	218

Table 5
The influences of presence on attributes.

Environment	Attributes	Estimate	*t* Value	Pr(> |*t*|)
WS	Appearance	0.142	2.131	0.0342
	Material	0.270	3.822	1.740e−4
	Fit	0.187	2.452	0.0150
	Category	0.130	1.880	0.0614
	Price	0.0921	1.294	0.197
	Store	0.269	3.216	1.500e−3
VR	Appearance	0.0860	1.259	<2e−16
	Material	0.244	3.388	8.370e−4
	Fit	0.216	3.349	9.580e−4
	Category	0.0698	1.092	0.276
	Price	0.209	3.295	1.150e−3
	Store	0.468	7.623	7.740e−13

Table 6
The evaluations of perceived quality.

Environment	Attributes	Estimate	*t* Value	Pr(> |*t*|)
WS	Appearance	−0.0665	−1.152	0.250
	Material	0.283	5.729	*3.52e−08****
	Fit	0.125	2.130	0.0343
	Category	0.311	5.115	*7.11e−07****
	Price	0.0462	0.975	0.331
	Store	0.212	3.748	*0.000231****
VR	Appearance	0.1958	3.217	0.00150
	Material	0.1413	2.941	0.00363
	Fit	0.2467	4.748	*3.79e−06****
	Category	0.1081	2.044	0.04222
	Price	0.1999	4.795	*3.07e−06****
	Store	−0.0059	−0.126	0.89976

* $p < 0.05$, ** $p < 0.01$.
*** $p < 0.001$

intrinsic or extrinsic attributes in WS and VR as dependent variable. The results are illustrated in Table 5. We see that presence in WS is most influential ($p < 0.01$) on material and store; in VR it is influential ($p < 0.001$) on all attributes except category. Hence, presence in WS has smaller effects on the perceptions of product attributes than that in VR. This can be attributed to the lower level of presence in WS as shown in Table 3. However, for attributes whose information can be adequately communicated by basic media (i.e., text descriptions, static images), such as category, presence may be of limited influence. One possible explanation for the different effects of price is that price in WS may be ignored as a cue (as in VR) to infer user preference as we will explain later for Hypothesis 5. Thus, Hypothesis 3 is partially supported.

For Questions 1 and 2 and Hypothesis 4, we conducted a multivariable linear regression with intrinsic and extrinsic attributes as independent variables and 'perceived quality' as dependent variable. The results, presented in Table 6, show that three attributes in WS are the major concerns for product quality, namely material, category and store. In addition, attribute 'fit' is also considered important but has smaller influence. A number of subjects commented that "It is difficult to judge the t-shirt", "I don't think t-shirt would fit me well. It does not even fit the model well." and "cannot see design". Note that the regression coefficients of category and store are greater than material and fit, which means that perceived quality relies more on extrinsic attributes than on intrinsic attributes. In contrast, the most important attributes in VR are appearance, material, fit and price. Most comments were focused on these four attributes, for example, "Nice and simple yet beautiful color", "I am a fan of ninja turtle, but it is not 100% cotton [sic].", "This shirt looks cute on the girl" and "The price is very cheap. good quality price ratio". Hence, subjects relied more on intrinsic attributes than extrinsic attributes to evaluate the quality of t-shirts in VR.

Table 7
The evaluations of prior ratings.

Environment	Attributes	Estimate	*t* Value	Pr(> |*t*|)
WS	Quality	0.619	8.129	3.58e−14
	Cost	0.0613	0.790	0.430
VR	Quality	0.670	10.521	< 2e−16
	Cost	0.141	2.206	0.028

Of the four major attributes identified from pilot study, we find that only one of them (material) is revealed in WS whereas all four attributes are correctly recognized in VR. One possible explanation is that when users have less or no direct experiences with products, they may tend to use extrinsic attributes (i.e., category, store, price) as cues to infer the product quality. On the other hand, if users have effective and direct interactions with products and thereby gain sufficient direct product experiences, they may rely more on intrinsic attributes to evaluate products. Thus, Hypothesis 4 is supported.

For Hypothesis 5, we investigated the correlations among prior ratings, perceived quality and cost in WS and VR by applying a linear regression analysis. The results are shown in Table 7. It is observed that the coefficient of perceived quality is positive and large (> 0.6), and that it has a significant influence ($p < 0.001$) on prior ratings. In addition, the cost is demonstrated as relatively small yet positively important in VR ($0.14, p < 0.05$) rather than in WS ($0.06, p > 0.1$). The price of collected t-shirts ranges from US \$3.99 to \$32.00, and so the price is within a normal range in general in the context of the study. Besides, the mean, median and mode of product price ratings are 3.25, 3, and 3, respectively (where 3 (out of 5) here means "slightly disagree or slightly agree" that the price is within a normal range). The results indicate that most subjects do not think the price is unacceptable. In other words, price tends to be indifferent when evaluating the preferences, and works as a confirmation that the perceived quality is matchable with the price and thus its quality is trustworthy. As a consequence, users are more likely to like products as a whole given good quality estimated. However, as users may not correctly judge the quality of products in WS, the price may fail to be or less considered when assessing their preferences. In addition, if the price seems exceptional—too high or too low—users may suspect that the real quality that they will received cannot be competitive with the quality that the product is claimed. In this case, the perceived cost may have negative influence on the prior ratings. Thus, Hypothesis 5 is supported.

5. Discussion

In this section, we discuss the motivation for users to give prior ratings, the usage of prior ratings, the similarities and differences between prior ratings and other information sources, the limitations to our current experiments, and the implication of our work in real applications.

5.1. Motivation to provide prior ratings

As a new information source intended for recommender systems, it is important for users to be properly motivated to provide their feelings, understandings and evaluation towards the experience of virtual products, i.e., prior ratings. In this article, we consider a number of possible ways to motivate users to provide prior ratings.

First, it is easy, comfortable and convenient to rate virtual products. For posterior ratings, users have to wait for product delivery, try out products (offline) and then go back to rate them (online), if they decide to rate. The time and effort in such a procedure can cause users to forget or not wish to rate products after use. In addition, users may not have a fresh memory regarding their offline experience and thus bring in noise in their posterior ratings (Nguyen et al. 2013). In contrast, the presence of virtual reality and responsive virtual products provide and enable users life-like shopping experience. More importantly, users can immediately go through the virtual products of interest, and preserve a fresh memory leading to the prior ratings closer to their real feelings. Hence, users may feel it low effort, comfortable, and convenient to give prior ratings.

Second, new and novel incentive mechanisms (Zhang et al. 2012) can be developed to pro-actively encourage users to rate products. Virtual product experiences provides an opportunity for retailers to design novel ways to motivate users to try out more products and speak out their experiences, i.e., providing prior ratings. How to design incentive mechanisms is beyond the scope of the present work. Nevertheless, the media-rich new environments do have the potential to boost new opportunities for incentive mechanisms.

Third, an additional possibility is not to request users to rate explicitly, but to implicitly transform users' experience into prior ratings automatically. In this way, users can focus more on their virtual products and do not worry about giving ratings. As an initial attempt, Guo and Elgendi (2013) study how to capture users' emotional signals during their experience with virtual products, and then convert these signals (in the form of electroencephalography EEG) into ratings prior to purchase. Arapakis et al. (2009) use web cameras mounted on top of computer screens to capture users' real-time facial expressions during watching videos, and show that better recommendation performance can be achieved by integrating with other implicit feedback. Literature works such as Yeasin et al. (2006) have shown that facial expressions can be automatically transformed into numerical level of interest, i.e., user preference. Such techniques could be adapted in virtual reality to derive prior ratings.

Lastly, our evaluation in Section 6.2.3 will show that even with a small amount of prior ratings, the performance of recommendations can still be improved to some extent. In other words, we do not expect users to provide prior ratings of every products that they experience; a small amount of ratings is sufficient.

5.2. Usage of prior ratings

Since prior ratings are issued by users who have experienced the products of interest in virtual reality, the ratings are not used to recommend these products back to those who have already formed their evaluation and opinions (virtually or physically). Rather, prior ratings are used to help model user preference for a recommender system to recommend users with items that they are not aware of. Specifically, prior ratings can be used in at least two cases: (1) the prior ratings of an active user can be used to help identify other users who share similar preferences, and hence the system can recommend to the active user products that she has not experienced or purchased; and (2) the prior ratings of similar users can be used to form a proper prediction on the unknown items for an active user. Further, in Section 6 we will propose a specific collaborative filtering technique that exploits prior ratings to improve the performance of recommendations, as a feasible solution to demonstrate the use of prior ratings in a real case.

5.3. Prior ratings vs. posterior ratings

The most commonly used information source in recommender systems is posterior ratings given by users after they purchased and experienced the products of interest offline. The values of prior and posterior ratings will be expected to differ (but have some correlation, see Hypothesis 2), because they are made by users based on different information. In general, both types of ratings indicate user preferences towards a certain product based on some form of user experience, but they differ in the confidence of user ratings. Specifically, prior ratings reflect users' virtual product experience according to their interactions with virtual products that are represented in a mediated environment. We find that environments with higher sense of presence motivate users to rate products[2] (see Hypothesis 1), and lead to more confident prior ratings and rating values closer to posterior ratings (see Hypothesis 2). In contrast, posterior ratings reflect more tangible forms of user experience based on 'physical' interactions with real products in real world; but the issue is that users often lack incentive to provide their ratings—one cause of the data sparsity problem.

Users' confidence of prior ratings may be lower than that of posterior ratings due to the less tangible form of product experience. We use users' stated rating confidence, if any, to distinguish prior ratings from posterior ratings. In Section 6 we evaluate the impact of confidence on recommendation performance.

As pointed out by Nguyen et al. (2013), posterior ratings could also be noisy *per se* as users do not have fresh memory when they get back to rate the products after the experience. Prior ratings, instead, can be given in a shorter time after the experience with virtual products, and easier to 'get back' to rate them. In other words, prior ratings could alleviate the motivation issue of posterior ratings to some extent.

Lastly, it is commonly understood that users usually browse or experience more products than they actually purchase, especially in the virtual environments where product information is easy to reach. In this regard, we suggest that prior ratings can complement traditional posterior ratings and help inherently alleviate the data sparsity and cold start problems.

5.4. Prior ratings vs. social information sources

We discuss two kinds of social relationships of users, namely friendship and social trust which are widely adopted in social recommender systems. It has been demonstrated that social networks provide an additional and useful source of information to improve the quality of recommendations (Crandall et al. 2008).

Specifically, friendship is readily available from social networks and has been demonstrated to be helpful for recommender systems (Ma et al. 2008). However, it has also been reported that online friendship sometimes cannot work well for recommenders due to its inherent ambiguity as a relational descriptor (Boyd 2006). It becomes easy, simple and low cost to connect with other users in social networks, and it is not surprising to find that a number of strangers appear within a user's circle of friends. An even weaker connections could be only one-sided, for example, users in Twitter can easily 'follow' other users whereas the others are free not to link back. In contrast, trust relationships are much stronger than friendships as the former are often built upon positive evaluation towards the others in conducting some expected actions (Mayer et al. 1995), e.g., providing reliable ratings. It has been shown that trust-based recommender systems are able to provide better performance (Yang et al. 2013). However, a critical problem of trust information is its sparsity and the difficulty of building it. Existing publicly-available datasets that include trust information show that only a small portion of users has specified others as trustworthy (Guo et al. 2014b). Only a few online systems

[2] That is, we mean that users are more willing to give prior ratings in VR than in WS, rather than that users are more likely to give prior ratings than posterior ratings.

(e.g., epinions.com, ciao.co.uk) support the concept of trust, whereas most other systems do not build user connections based on trust evaluation, or even do not have an inherent social network. In conclusion, friendships are more common and easy to collect, but more ambiguous towards user preferences; trust is much stronger than friendships but itself suffers from the data sparsity problem. In addition, both information sources require a social network structure inherently supported by a recommender system. Besides, these kinds of information are usually represented in the form of 'who connects whom' without a numerical value indicating the strength of social ties, whereas not all the friends/trusted users should be equally weighted. This problem may further limit the usefulness in recommender systems. Moreover, although social relationships-based recommender systems can help mitigate the cold start problem, recent work shows that the performance of cold users is still much worse than that of normal users Yang et al. (2013). Therefore, it is necessary to identify other possible information sources that are more reliable and less constrained, and can be effectively used to model user preferences for recommender systems.

Prior ratings are just such a kind of information sources that has the potential to overcome the drawbacks of existing information sources, and help reveal and model user preference to improve the performance of recommendations. First, prior ratings are issued by users based on their evaluation of virtual products of interest prior to purchase, which are similar to the posterior ratings with respect to real products. Prior ratings directly indicate user's likeness towards the products that they experienced. As a comparison, it makes sense that even friends or trusted users may have different preferences, especially when these relationships are built upon offline relations (e.g., classmates, colleagues, families). Second, prior ratings have no requirement and constraint of a social network that should be supported by a social recommender system. Third, prior ratings may less suffer from the data sparsity problem than social trust, considering the following viewpoints: (1) since users usually experience more products than they purchase (and could even more in the case of e-commerce environments), it has a potential to attract more prior ratings than posterior ratings which are richer than trust; and (2) as elaborated in Section 5.1, a number of ways can be used to motivate users to provide prior ratings. As a result of more available information, the data sparsity and cold start problems could be better resolved than other kinds of information sources. Lastly, prior ratings can coexist with social information sources. Prior ratings refer to only individuals' personal preferences, and thus they do not prevent users from building connections with other users in virtual reality. In fact, it is possible to infer implicit trust from prior ratings in the same way as from posterior ratings, to resolve the data sparsity problem of explicit trust (Guo et al. 2014b,c). Ultimately it would be more powerful to make use of all possibly available information sources (e.g., prior ratings, posterior ratings, social relationships) in order to better resolve the data sparsity and cold start problems. Such future work is beyond the scope of this article.

5.5. Prior ratings: An alternative information source

To summarize the discussions in the last two subsections, we propose prior ratings as a good alternative information source to social relationships, and as a complementary to posterior ratings for modelling user preferences. The speciality of prior ratings lies in the differences from other information sources: they provide a unique source of information that is similar to (yet less confident than) posterior ratings, and is more reliable than social relationships as real users' preferences on products. Prior ratings are easy to issue, require no purchase payment, and have the potential to be denser than posterior ratings due to the fact that users often experience more products than they purchase. They are also distinguished in formulating the product experience prior to purchase in the form of usable information source. This kind of product experience exists for a long time in e-commerce, but has not been studied and investigated for recommender systems till now.

5.6. Limitations of current experiments

There are several potential limitations in our current experiments. First, certain attribute information (e.g., warranty, shipping) was not available for our user study. Although these are less relevant for the product type studied, they may be more important for other kinds of products. Second, due to lack of devices, our prototype implementation uses only visual information in VR: users cannot touch the t-shirts and feel the material. Tactile feedback may be important for user evaluation of preferences. Nevertheless, as analyzed in Section 4.1, this limitation may not greatly influence the general conclusion since we exploited abstract attributes rather than some specific attributes. Further, as shown in Section 6, prior ratings can improve recommendation performance even if based solely on visual information, reducing the requirements of expensive VR devices. In this regard, t-shirts represent a kind of products with simple representations in VR. Third, most subjects in our study were computer or electrical engineering students on a university campus, and the sample size was modest. A larger and more heterogeneous sample may allow for more confident generalization of our research findings.

5.7. Implications for real systems

Our work has practical implications for real systems. First, as stated by Hypotheses 1 and 2, users usually are more willing and feel more comfortable and confident in sharing their opinions in VR than in WS. This indicates that for the product sellers, it would be value to market products through e-commerce systems in VR, which can provide better online product experiences than traditional websites. Besides, as we will demonstrate in Section 6, prior ratings can benefit recommender systems by solving the data sparsity and cold start problems. In other words, the VR e-commerce systems can help expose more products to users, and recommend them more accurate products of interest.

Second, as pointed out by Hypothesis 3, a greater sense of presence can enhance users' perceptions of products, and thus form better opinions regarding the product qualities. Therefore, for the designers of VR, it is necessary to enhance the environmental presence by enabling richer types of interactions and media to better convey product information.

Third, our foundational work in introducing the concept of prior ratings opens up future discussions around customers' pre- and post-purchase product evaluations and their purchase intent decisions.

6. Leveraging prior ratings

Having defined prior ratings and observed their value, we now leverage prior ratings in product recommendations. Since no existing commonly used datasets contains information of prior ratings, we rely on the prior ratings collected from the user studies for quantified evaluation.

To facilitate discussion, we first introduce a number of notations. Let the sets of all users, all items, all ratings and all rating confidences be $\mathbb{U}, \mathbb{I}, \mathbb{R}$ and $\mathbb{C}$, respectively. We keep the symbols $u, v \in \mathbb{U}$ for users and $i, j \in \mathbb{I}$ for items. Let $r_{u,i} \in \mathbb{R}, c_{u,i} \in \mathbb{C}$ represent the rating and confidence given by user u on item i, respectively. The confidence $c_{u,i}$ indicates the degree to which users are 'certain'

about their evaluation $r_{u,i}$ on item i. Since the collected data does not include the confidence of posterior ratings, we set it as 1.0 to ensure that users are more certain in their posterior ratings. Then the task of a recommender can be modeled as: given a set of user-item-rating-confidence $(u, i, r_{u,i}, c_{u,i})$ quaternions, provide a prediction $(u, j, ?, ?)$ for user u on an unknown item j. The prediction pair is denoted as $(\hat{r}_{u,j}, \hat{c}_{u,j})$.

6.1. Prior ratings-based CF (PRCF)

As a new information source, prior ratings have not been used in any previous research, and no recommendation algorithms have been developed based on prior ratings so far. We integrate prior ratings with the conventional collaborative filtering (CF), based on a *confidence-aware* distance similarity. Note that we aim to evaluate the usefulness of prior ratings by providing a feasible solution to make use of prior ratings rather than to provide a perfect algorithm resulting in the best performance. The algorithms based on other popular techniques such as matrix factorization may be proposed and work better than CF, but that is beyond the discussion of this article. We leave the exploration for a better recommendation algorithm based on prior ratings as a line of future research.

The first step of CF is to identify the like-minded users who have similar preferences with the active user. Since all the users only rated a few items, traditional similarity measures such as Pearson correlation coefficient and cosine similarity often fail to work effectively in this condition (Guo et al. 2013b). More importantly, existing similarity measures cannot accommodate the new information source in terms of prior ratings and their confidences. In this article, we therefore propose a confidence-aware distance similarity by taking into consideration three factors: the distance of ratings, the distance of rating confidences and the semantics of rating values. Intuitively, the greater the rating distance is, the lower the similarity will be. This intuition also holds for the distance of rating confidences. Further, if two ratings have the same positive or negative opinions towards the same item, they are regarded as semantically indifferent. We regard a rating as positive if its value is greater than the median rating scale; otherwise it is negative. In addition, ratings generally have more influence on similarity than rating confidences. Hence, we compute user similarity as follows:

$$s_{u,v} = 1 - \frac{1}{3N}\sum_{i \in I_{u,v}}\left(\frac{|r_{u,i} - r_{v,i}|}{R_{max} - R_{min}} + \frac{|c_{u,i} - c_{v,i}|}{|c_{u,i} - c_{v,i}| + 1} + sw_i\right), \tag{1}$$

where $s_{u,v} \in [0, 1]$ is the similarity between users u and v based on their ratings $(r_{u,i}, r_{v,i})$ on commonly rated items $I_{u,v}$ with cardinal N, and R_{max} and R_{min} are respectively the maximum and minimum rating scales defined by a recommender system. The semantic weight sw_i is defined by:

$$sw_i = \begin{cases} \frac{1}{d_i+1} & \text{if } d_i \geqslant 0; \\ \frac{|d_i|}{|d_i|+1} & \text{otherwise}, \end{cases} \tag{2}$$

where $d_i = (r_{u,i} - R_{med})(r_{v,i} - R_{med})$ denotes the extent to which two users have the same opinions towards item i relative to the median scale R_{med}. The settings of sw_i capture the intuition that if two users have closer opinions, the computed similarity $s_{u,v}$ will be greater and vice versa. The semantic weight is used to distinguish the case where two pairs of ratings have the same rating distance but possess distinct semantic meaning essentially. For example, assume that there are two pairs of ratings (5, 4) and (4, 3) from two users and that the rating scales are integers from 1 to 5 predefined by a certain system. Although the rating distance is the same (1), the semantic meaning is different. Specifically, both ratings 5 and 4 are greater than the median rating scale (3) and hence positive whereas the ratings 4 and 3 are different opinions in real life. In other words, the rating distance 1 in the first pair reflects the difference in liking whereas the same distance in the second pair reflects the differences between liking and disliking.

The second step of CF is to select a set of similar users in order to predict the rating and confidence of an unknown item for an active user. Specifically, we adopt the users who have rated item j and whose similarity is greater than a predefined threshold, i.e., $U_{u,j} = \{v \mid s_{u,v} > \theta, \exists r_{v,j}, v \in U\}$, where θ is a similarity threshold. In this article, all users with positive correlations will be adopted, i.e., $\theta = 0$.

The third step of CF is to generate the predictions by using either simple weighted average (WA) or Resnick's formula (RF) (Su and Khoshgoftaar 2009):

$$\hat{p}_{u,j} = \frac{\sum_{v \in U_{u,j}} s_{u,v} p_{v,j}}{\sum_{v \in U_{u,j}} |s_{u,v}|} \tag{WA}$$

$$\hat{p}_{u,j} = \bar{p}_u + \frac{\sum_{v \in U_{u,j}} s_{u,v}(p_{v,j} - \bar{p}_v)}{\sum_{v \in U_{u,j}} |s_{u,v}|} \tag{RF}$$

where $p_{u,j}$ corresponds to $r_{u,j}$ (or $c_{u,j}$), and $\hat{p}_{u,j}$ to $\hat{r}_{u,j}$ (or $\hat{c}_{u,j}$), respectively, and $\bar{p}_u$ represents the average rating or confidence reported by user u. (WA) and (RF) may produce different rating predictions and hence both are used to predict item ratings. Since the confidence of posterior ratings is unavailable, we set it to 1.0 by default; thus the confidence difference in (RF) will be always 0 if only posterior ratings are available. Hence, we adopt (WA) only to predict rating confidences.

PRCF Variants. One potential drawback of the PRCF approach is the reliance on rating confidences. In practice, users may not be motivated to provide the rating confidences, which could become for them an additional cognitive burden. Hence, we adapt PRCF to two scenarios. First, when no confidence data is available we set 0.5 as the default confidence for all prior ratings (considering that we set 1.0 as the default confidence for all posterior ratings). This algorithm variant is denoted as *PRCF-1*. Second, when only limited confidence data is available, i.e., only few users reported confidence while others did not, we use the average of all available confidence data as the default confidence for all prior ratings. This algorithm variant is denoted as *PRCF-2*.

Discussion. This section presented a new user-based collaborative filtering technique (i.e., PRCF) that we developed by exploiting the prior ratings. The new method is not trivial considering the following aspects. First, we proposed a new similarity measure based on both the values and confidence of prior ratings (see Eq. (1)), due to the inability of traditional similarity measures to accommodate the prior ratings. As pointed out by Guo et al. (2013b), the definition of the similarity measure is important since it plays two critical roles in collaborative filtering: (1) identifying a number of similar users; and (2) weighting the ratings of similar users to generate predictions. Second, our approach uses a new information source that is not supported by traditional approaches. Third, a new predictive metric will be proposed in next subsection (see Eq. (5)) to account for the rating confidence. Therefore, we regard PRCF as a new user-based collaborative filtering method rather than applying others' existing work.

6.2. Results and analysis

Using the collected real data, we conduct a number of experiments to investigate the effectiveness of prior ratings in predicting item ratings. We further verify the usability of our collected modest data in revealing performance patterns. The leave-one-out approach is applied to the rating data collected from user study (see Table 4). In particular, each rating is hidden iteratively and

then predicted by adopting the proposed PRCF approach or its variants.

Predictive performance is usually estimated in terms of mean absolute errors (MAE) between predictions ($\hat{r}_{u,j}$) and the ground truth, and the rating coverage (RC) of predictable ratings over all testing ratings. In particular, they are computed as follows:

$$\mathrm{MAE} = \frac{\sum_u \sum_j |\hat{r}_{u,j} - r_{u,j}|}{M}; \tag{4a}$$

$$\mathrm{RC} = \frac{P}{M} \times 100\%, \tag{4b}$$

where M is the total number of testing ratings, and P is the number of predicable ratings. Since MAE does not consider the influence of rating confidences ($\hat{c}_{u,j}$), we propose a confidence-aware metric, termed mean absolute confidence errors (MACE):

$$\mathrm{MACE} = \frac{\sum_u \sum_j \hat{c}_{u,j} |\hat{r}_{u,j} - r_{u,j}|}{\sum_u \sum_j \hat{c}_{u,j}}. \tag{5}$$

Note that if all rating confidences such as of posterior ratings are the same, MACE will be the same as MAE. Generally, smaller MAE and MACE mean better predictive accuracy and higher value of RC indicates better coverage.

6.2.1. Performance of PRCF

We aim to investigate how posterior ratings can be better predicted by involving prior ratings. We denote R_p as the set of posterior ratings, R_{pw} the union set of R_p and $R.ws$, and R_{pv} the union set of R_p and $R.vr$. Two subsets will be used as testing data: *All* is the subset including all the posterior ratings, and *Pred* is the subset only including the posterior ratings that can be predicted by PRCF when no prior ratings are used. The performance of PRCF is reported in Table 8.

The results show that if only posterior ratings (R_p) are used, a small ratio (4.56%) of testing ratings is predicted. The rating coverage can be greatly increased by involving prior ratings (R_{pw}, R_{pv}) since more ratings are thus available. The difference in coverage between R_{pw} and R_{pv} is due to the fact that subjects in the user study rated different t-shirts in different environments. Another observation is that the accuracy based on (WA) and (RF) may be different in terms of MAE and MACE. MACE generates relatively smaller values than MAE as it considers rating confidences. For *All*, R_{pw} and R_{pv} may produce worse (WA) or better (RF) results than R_p, and R_{pw} has larger variations between (WA) and (RF) than R_{pv}. This may indicate that prior ratings in WS produce less reliable predictions than those in VR. Considering that the newly predicted item ratings also contribute to predictive errors, the performance on *Pred* is more comparable among different approaches and better demonstrates the effectiveness of prior ratings. Although the performance of R_{pw} may increase (RF) or decrease (WA), R_{pv} consistently and significantly obtains better accuracy. Specifically in *Pred*, 7.86% improvements in MAE and 9.50% in MACE can be achieved using (WA) while 8.0% (MAE) and 10.34% (MACE) are obtained using (RF). In conclusion, prior ratings in VR can significantly improve both the coverage and accuracy whereas those in WS only show consistent improvements in the coverage rather than accuracy.

6.2.2. Performance of PRCF variants

We further study the effectiveness of prior ratings when rating confidences are missing or incomplete. The performance of PRCF variants is shown in Tables 9 and 10. Comparing with PRCF, PRCF-1 obtains similar but superior results in Table 9: prior ratings in VR can consistently improve the predictive accuracy, and the improvement is more significant especially in terms of MACE. Specifically for R_{pv} in *Pred*, 7.97% improvements in MAE and 11.35% in MACE can be achieved using (WA) while 8.46% (MAE) and 13.48% (MACE) are obtained using (RF) relative to the performance of R_p. One possible explanation is that the rating confidences reported by users are unreliable and noisy due to the fact that only visual information is available to evaluate product quality and performance. This indicates that a richer environment with higher sense of presence can help improve the quality of user evaluation, and hence enhance the utility of prior ratings in predicting the ratings of unknown items.

For PRCF-2, the default confidences adopted are 0.659 and 0.756, corresponding to the average confidences in WS and in VR after normalization, respectively. Analogously, the results presented in Table 10 are similar to those in Table 8: R_{pw} works worse in (WA) but better in (RF) whereas R_{pv} achieves much better accuracy in both cases. We note that MAE in Tables 9 and 10 is the same. This may be due to two reasons. First, confidence in Eq. (1) has smaller influence than ratings. Second, MAE does not take into account rating confidence. However, in terms of MACE, PRCF-1 performs better. Hence, MACE is important in evaluating overall predictive performance.

Since PRCF performs closely with PRCF-2, which in turn is inferior to PRCF-1, we conclude that prior ratings can effectively improve the recommendation performance when the rating confidences are robust and correctly based on sufficient product information. In case that the confidence data is missing or incomplete, prior ratings with a default smaller rating confidence offer a performance improvement. We can also recommend higher-fidelity mediated environments, such as VR to WS, because the former achieves consistent improvements in both coverage and

Table 8
The predictive performance of PRCF.

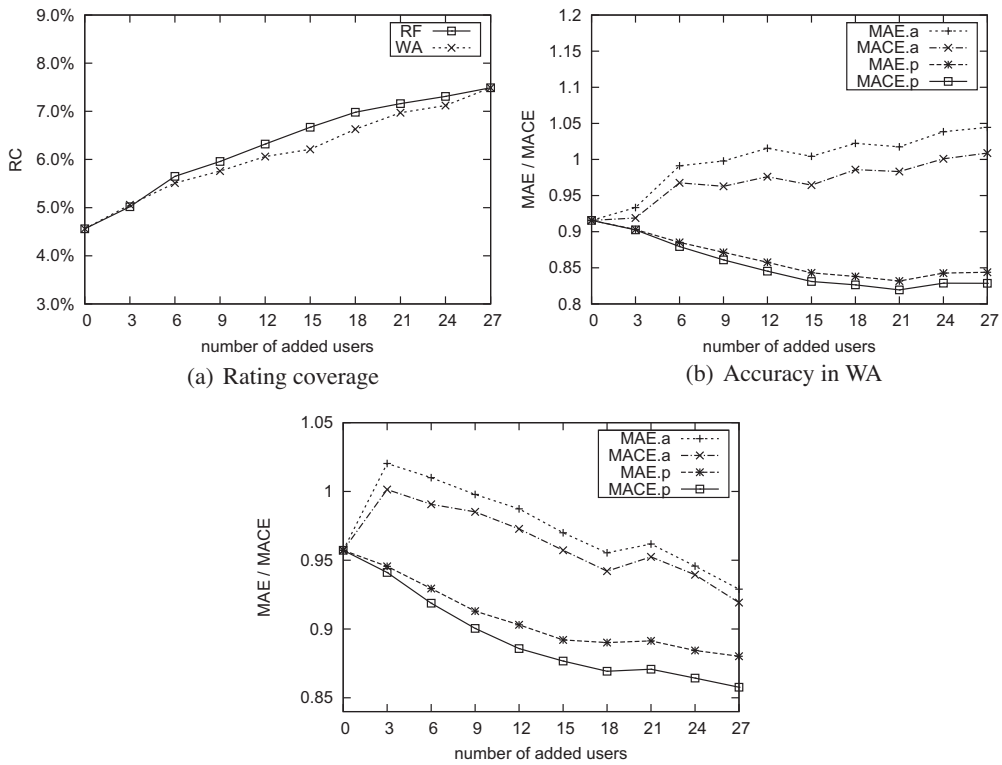
		All			Pred	
		MAE	MACE	RC (%)	MAE	MACE
(WA)	R_p	0.916	0.916	4.56	0.916	0.916
	R_{pw}	1.394	1.346	9.26	1.070	1.048
	R_{pv}	1.044	1.009	7.49	0.844	0.829
(RF)	R_p	0.957	0.957	4.56	0.957	0.957
	R_{pw}	0.798	0.815	9.26	0.926	0.910
	R_{pv}	0.929	0.919	7.49	0.880	0.858

Table 9
The predictive performance of PRCF-1.

		All			Pred	
		MAE	MACE	RC (%)	MAE	MACE
(WA)	R_p	0.916	0.916	4.56	0.916	0.916
	R_{pw}	1.407	1.311	9.26	1.095	1.057
	R_{pv}	1.046	0.964	7.49	0.843	0.812
(RF)	R_p	0.957	0.957	4.56	0.957	0.957
	R_{pw}	0.798	0.810	9.26	0.925	0.896
	R_{pv}	0.926	0.877	7.49	0.876	0.828

Table 10
The predictive performance of PRCF-2.

		All			Pred	
		MAE	MACE	RC (%)	MAE	MACE
(WA)	R_p	0.916	0.916	4.56	0.916	0.916
	R_{pw}	1.407	1.351	9.26	1.095	1.071
	R_{pv}	1.046	1.013	7.49	0.843	0.830
(RF)	R_p	0.957	0.957	4.56	0.957	0.957
	R_{pw}	0.798	0.805	9.26	0.925	0.907
	R_{pv}	0.926	0.906	7.49	0.876	0.855



(a) Rating coverage (b) Accuracy in WA

(c) Accuracy in RF

Fig. 4. The predictive accuracy and coverage with incremental prior ratings.

accuracy. Although prior ratings in WS may also have a certain positive influence on recommendations, its performance varies by the different prediction approaches. Lack of high confidences in prior ratings could be one of the critical reasons. The rationale is that prior ratings with low confidences *per se* are noisy and hence less useful.

6.2.3. Performance of incremental prior ratings

Finally, we examine whether the modest sample size in our experiments has validity. We randomly and incrementally (step $k = 3$) incorporate the prior ratings of new users. We apply PRCF to generate predictions, and repeat the whole process five times in order to generate different sequences of prior ratings. The mean results are shown in Fig. 4, where only posterior ratings are used if $k = 0$.

Fig. 4(a) shows that, as more prior ratings are available, more ratings of unknown items can be predicted. The trends of rating coverage are similar in (WA) and (RF). Fig. 4(b, c) show the changes of predictive accuracy when the number of prior ratings is adjusted in (WA) and (RF), respectively. For prediction set *All*, the predictive errors (MAE.a, MACE.a) in (WA) increase whereas in (RF) they decrease. As explained earlier, newly-predicted ratings may contribute to the overall errors (with view *All*), and hence the better view of performance is based on the results of prediction set *Pred*. The predictive errors (MAE.p, MACE.p) gradually decrease as the amount of prior ratings increase. This indicates that even a small number of prior ratings can improve recommendation performance.

7. Conclusion and future work

Towards the design of recommender systems for e-commerce in virtual reality (VR), this article proposed a new information source, called *prior ratings*. By leveraging the effective interactions between users and virtual products represented in a mediated environment, prior ratings capture users' opinions about products that result from virtual product experiences, usually prior to purchase. We presented a conceptual model of prior ratings that provided their principled foundation. We conducted a user study in two different environments, namely website and virtual store, in which users virtually interacted with t-shirt products. Particularly, unlike in the traditional environments (WS), users felt more comfortable and were motivated to rate products in VR, and provided more confident prior ratings that were closer to posterior ratings due to the higher sense of presence. We found that the presence had positive influence on the perceptions of some experience-related product attributes, both intrinsic and extrinsic. To estimate product quality, users relied more on extrinsic attributes in WS while users relied more on intrinsic attributes in VR since direct experiences can be obtained. Besides, both perceived quality and cost positively influence prior ratings. We stress that prior ratings can complement posterior ratings and help ameliorate the data sparsity and cold start problems due to the fact that users usually browse or experience more product than they actually purchase. In conclusion, the results validated the conceptual model of prior ratings under our experimental settings. In addition, since higher presence may result in more confident prior ratings, it follows that the design of virtual stores should emphasize the sense of presence by increasing the media richness or the effectiveness of user interactions.

We furthered our contribution by demonstrating how to leverage prior ratings in predicting items' ratings using a collaborative filtering technique. Specifically, we introduced a new similarity measure by taking into account three important factors: the distances between ratings, the distances between confidences and the semantics of rating values, each of which captures the different considerations of user similarity. Using this measure, we proposed a confidence-aware performance measurement. Using the data col-

lected from the user study, we conducted a number of experiments to evaluate the usefulness of prior ratings in improving the recommendation performance. The results indicated that prior ratings are effective and valuable, and hold potential as a new information source to bootstrap recommender systems. We also showed that even a small number of prior ratings may benefit the recommender systems.

Our current research focuses on the validation of conceptual model of prior ratings and on a feasible solution to leverage both prior and posterior ratings to improve recommendation performance. For future work, we intend to investigate additional benefits such as diversity of prior ratings. In addition, an interesting study that builds on our work would be to compare user purchase behaviours in a system with and without prior ratings.

Acknowledgements

We thank the reviewers for their constructive comments. Guibing Guo acknowledges the Ph.D.Grant from the Institute for Media Innovation, Nanyang Technological University, Singapore. Neil Yorke-Smith thanks the Operations group at the Judge Business School and the fellowship at St Edmunds College, Cambridge.

References

Ahmed, A., Kanagal, B., Pandey, S., Josifovski, V., Pueyo, L.G., and Yuan, J. Latent factor models with additive and hierarchically-smoothed user preferences. In Proceedings of the 6th ACM International Conference on Web Search and Data Mining (WSDM), 2013, 385–394.

Ahn, H., 2008. A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem. Information Sciences 178, 37–51.

Arapakis, I., Moshfeghi, Y., Joho, H., Ren, R., Hannah, D., and Jose, J.M. Enriching user profiling with affective features for the improvement of a multimodal recommender system. In Proceedings of the ACM International Conference on Image and Video Retrieval (CIVR), 2009, 29:1–29:8.

Bearden, W., Shimp, T., 1982. The use of extrinsic cues to facilitate product adoption. Journal of marketing research 19, 229–239.

Bobadilla, J., Ortega, F., Hernando, A., 2012. A collaborative filtering similarity measure based on singularities. Information Processing and Management 48, 204–217.

Boyd, D. Friends, friendsters, and myspace top 8: Writing community into being on social network sites. First Monday 11, 2006. URL: <http://journals.uic.edu/ojs/index.php/fm/article/view/1418>.

Carenini, G., Smith, J., and Poole, D. Towards more conversational and collaborative recommender systems. In Proceedings of the 8th International Conference on Intelligent User Interfaces (IUI), 2003, 12–18.

Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., and Suri, S. Feedback effects between similarity and social influence in online communities In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2008, 160–168.

Dodds, W., Monroe, K., Grewal, D., 1991. Effects of price, brand, and store information on buyers' product evaluations. Journal of marketing research 28, 307–319.

Dong, R., McCarthy, K., O'Mahony, M., Schaal, M., and Smyth, B. Towards an intelligent reviewer's assistant: recommending topics to help users to write better product reviews. In Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces (IUI), 2012, 159–168.

Eno, J., Stafford, G., Gauch, S., and Thompson, C. Hybrid user preference models for second life and opensimulator virtual worlds. In Proceedings of the 19th International Conference on User Modeling, Adaption and Personalization (UMAP), 2011, 87–98.

Fang, H., Bao, Y., and Zhang, J. Leveraging Decomposed Trust in Probabilistic Matrix Factorization for Effective Recommendation, 2014.

Gantner, Z., Drumond, L., Freudenthaler, C., Rendle, S., and Schmidt-Thieme, L. Learning attribute-to-feature mappings for cold-start recommendations. In Proceedings of the 2010 IEEE 10th International Conference on Data Mining (ICDM), 2010, 176–185.

Gardial, S., Clemons, D., Woodruff, R., Schumann, D., Burns, M., 1994. Comparing consumers' recall of prepurchase and postpurchase product evaluation experiences. Journal of Consumer Research 20, 548–560.

Goering, P., 1985. Effects of product trial on consumer expectations, demand, and prices. Journal of Consumer Research 12, 74–82.

Gunawardana, A., and Meek, C. Tied boltzmann machines for cold start recommendations. In Proceedings of the 2nd ACM Conference on Recommender Systems (RecSys), 2008, 19–26.

Guo, G., Elgendi, M., 2013. A new recommender system for 3D e-commerce: an EEG based approach. Journal of Advanced Management Science 1, 61–65.

Guo, G., Zhang, J., and Thalmann, D. A simple but effective method to incorporate trusted neighbors in recommender systems. In Proceedings of the 20th International Conference on User Modeling, Adaptation and Personalization (UMAP), 2012, 114–125.

Guo, G., Zhang, J., Thalmann, D., 2014a. Merging trust in collaborative filtering to alleviate data sparsity and cold start. Knowledge-Based Systems (KBS) 57, 57–68.

Guo, G., Zhang, J., Thalmann, D., Basu, A., and Yorke-Smith, N. From ratings to trust: an empirical study of implicit trust in recommender systems. In Proceedings of the 29th Annual ACM Symposium on Applied Computing (SAC), 2014b, 248–253.

Guo, G., Zhang, J., Thalmann, D., and Yorke-Smith, N. Prior ratings: a new information source for recommender systems in e-commerce. In Proceedings of the 7th ACM Conference on Recommender Systems (RecSys), 2013a, 383–386.

Guo, G., Zhang, J., Thalmann, D., and Yorke-Smith, N. ETAF: an extended trust antecedents framework for trust prediction. In Proceedings of the 2014 International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2014c.

Guo, G., Zhang, J., and Yorke-Smith, N. A novel Bayesian similarity measure for recommender systems. In Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI), 2013b, 2619–2625.

Guy, I., Ronen, I., Wilcox, E. Do you know?: recommending people to invite into your social network. In Proceedings of the 14th International Conference on Intelligent User Interfaces (IUI), 2009, 77–86.

Ha, H., 2002. The effects of consumer risk perception on pre-purchase information in online auctions: brand, word-of-mouth, and customized information. Journal of Computer-Mediated Communication 8.

Heeter, C., 1992. Being there: the subjective experience of presence. Presence: Teleoperators and Virtual Environments 1, 262–271.

Hemp, P., 2006. Avatar-based marketing. Harvard Business Review 84, 48–57.

Hu, X., and Wang, K. Personalized recommendation for virtual reality. In Proceedings of the 2010 International Conference on Multimedia Technology (ICMT), 2010, 1–5.

Jacoby, J., Szybillo, G., Busato-Schach, J., 1977. Information acquisition behavior in brand choice situations. Journal of Consumer Research 3, 209–216.

Jamali, M., and Ester, M. A matrix factorization technique with trust propagation for recommendation in social networks. In Proceedings of the 4th ACM Conference on Recommender Systems (RecSys), 2010, 135–142.

Jamali, M., and Ester, M. A transitivity aware matrix factorization model for recommendation in social networks. In: Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI), 2011, 2644–2649.

Jiang, Z., Benbasat, I., 2004. Virtual product experience: effects of visual and functional control of products on perceived diagnosticity and flow in electronic shopping. Journal of Management Information Systems 21, 111–147.

Konstas, I., Stathopoulos, V., and Jose, J. On social networks and collaborative recommendation. In Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2009, 195–202.

Koren, Y., Bell, R., Volinsky, C., 2009. Matrix factorization techniques for recommender systems. Computer 42, 30–37.

Lathia, N., Hailes, S., and Capra, L. The effect of correlation coefficients on communities of recommenders. In Proceedings of the 23rd Annual ACM Symposium on Applied Computing (SAC), 2008, 2000–2005.

Li, H., Daugherty, T., Biocca, F., 2001. Characteristics of virtual experience in electronic commerce: a protocol analysis. Journal of Interactive Marketing 15, 13–30.

Li, H., Daugherty, T., Biocca, F., 2003. The role of virtual experience in consumer learning. Journal of Consumer Psychology 13, 395–407.

Lichtenstein, D., Ridgway, N., Netemeyer, R., 1993. Price perceptions and consumer shopping behavior: a field study. Journal of Marketing Research 30, 234–245.

Liu, N.N., Meng, X., Liu, C., and Yang, Q. Wisdom of the better few: cold start recommendation via representative based rating elicitation. In Proceedings of the 5th ACM Conference on Recommender Systems (RecSys), 2011, 37–44.

Ma, H., King, I., and Lyu, M. Learning to recommend with social trust ensemble. In Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2009, 203–210.

Ma, H., Yang, H., Lyu, M., and King, I. SoRec: social recommendation using probabilistic matrix factorization. In Proceedings of the 31st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2008, 931–940.

Ma, H., Zhou, D., Liu, C., Lyu, M.R., and King, I. Recommender systems with social regularization. In Proceedings of the 4th ACM International Conference on Web Search and Data Mining (WSDM), 2011, 287–296.

Massa, P., and Avesani, P. Trust-aware recommender systems. In Proceedings of the 2007 ACM Conference on Recommender Systems (RecSys), 2007, 17–24.

Mayer, R., Davis, J., Schoorman, F., 1995. An integrative model of organizational trust. Academy of management review 20, 709–734.

McNee, S., Lam, S., Konstan, J., and Riedl, J. Interfaces for eliciting new user preferences in recommender systems. In Proceedings of the 9th International Conference on User Modeling (UMAP), 2003, 148–148.

Milgrom, P., Roberts, J., 1986. Price and advertising signals of product quality. The Journal of Political Economy 94, 796–821.

Nelson, P., 1974. Advertising as information. The Journal of Political Economy 82, 729–754.

Nguyen, T.T., Kluver, D., Wang, T.Y., Hui, P.M., Ekstrand, M.D., Willemsen, M.C., and Riedl, J. Rating support interfaces to improve user experience and recommender

accuracy. In Proceedings of the 7th ACM Conference on Recommender Systems (RecSys), 2013, 149–156.
Olson, J., 1974. Cue properties of price: literature review and theoretical considerations. 20, Pennsylvania State University.
Phang, C.W., and Kankanhalli, A. How do perceptions of virtual worlds lead to enhanced learning? an empirical investigation. In Proceedings of the 30th International Conference on Information Systems (ICIS), 2009, 167:0–167:17.
Picciano, A.G., 2002. Beyond student perceptions: issues of interaction, presence, and performance in an online course. Journal of Asynchronous learning networks 6, 21–40.
Rao, A., Monroe, K., 1989. The effect of price, brand name, and store name on buyers' perceptions of product quality: an integrative review. Journal of marketing Research 26, 351–357.
Rymaszewski, M. Second Life: The Official Guide. Sybex, 2007.
Shah, F., Bell, P., and Sukthankar, G. A destination recommendation system for virtual worlds. In Proceedings of the 21st International Florida Artificial Intelligence Research Society Conference (FLAIRS), 2010, 475–476.
Slater, M., Spanlang, B., Corominas, D., 2010. Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. ACM Transactions on Graphics (TOG) 29, 92:1–92:9.
Steuer, J., 1992. Defining virtual reality: dimensions determining telepresence. Journal of communication 42, 73–93.
Stokes, R. The effects of price, package design, and brand familiarity on perceived quality. Ph.D. thesis. ProQuest Information & Learning, 1974.
Su, X., Khoshgoftaar, T., 2009. A survey of collaborative filtering techniques. Advances in Artificial Intelligence, 2009, 4:2–4:2.
Tang, J., Gao, H., Hu, X., and Liu, H. Exploiting homophily effect for trust prediction. In Proceedings of the 6th ACM International Conference on Web Search and Data Mining (WSDM), 2013, 53–62.
Xiao, B., Benbasat, I., 2007. E-commerce product recommendation agents: use, characteristics, and impact. MIS Quarterly 31, 137–209.
Yang, B., Lei, Y., Liu, D., and Liu, J. Social collaborative filtering by trust. In Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI), 2013, 2747–2753.
Yeasin, M., Bullot, B., Sharma, R., 2006. Recognition of facial expressions and measurement of levels of interest from video. IEEE Transactions on Multimedia 8, 500–508.
Zeithaml, V., 1988. Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. The Journal of Marketing 52, 2–22.
Zhang, J., Cohen, R., Larson, K., 2012. Combining trust modeling and mechanism design for promoting honesty in e-marketplaces. Computational Intelligence 28, 549–578.