Prior Ratings: A New Information Source for Recommender Systems in E-Commerce

Guibing Guo, Jie Zhang
Nanyang Technological University, Singapore
gguo1,zhangj@ntu.edu.sg

Daniel Thalmann
Nanyang Technological University, Singapore
danielthalmann@ntu.edu.sg

Neil Yorke-Smith
American University of Beirut, and University of Cambridge
nysmith@aub.edu.lb

ABSTRACT

Lack of motivation to provide ratings and eligibility to rate generally only after purchase restrain the effectiveness of recommender systems and contribute to the well-known data sparsity and cold start problems. This paper 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.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Evaluation/methodology*

General Terms

Design, Experimentation

Keywords

Recommender systems, Prior ratings, Data sparsity, Cold start, User study, Virtual product.

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, 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* [4]. Although many approaches have been proposed to address these problems [11, 4, 5], few researchers [2] 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, contemporary websites, e.g., brides.com and Ray-ban.com are implementing novel interfaces and interactions through which users are allowed to virtually try on different wedding dresses and glasses, to better elicit user preferences. However, the available media and interactions are limited in comparison with the capabilities of virtual reality (VR), e.g., SecondLife.com which can provide users with immersive virtual product 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 [6].

While the emergence of 3D VR environments offers more adequate information which can be used to model user preference, research on recommender systems in VR is still in its infancy. Shah et al. [13] recommend to users locations of interest by analyzing users' login data to help them navigate in VR. Hu and Hang [7] design a system to recommend virtual furniture according to users' interest and requirements. Although a controlled prototype was implemented, the features of VR are not exploited to elicit user ratings.

This paper proposes a new information source, called *prior ratings*, built upon virtual product experiences. A conceptual model of prior ratings is proposed to provide a principled foundation, integrating the environmental factor *presence* whose effects on product evaluation have not been studied previously. For evaluation, we conduct user studies in two different modalities. The results demonstrate the validity of the conceptual model under our experimental settings. To our knowledge, this is the first work that defines and investigates the concept of prior ratings.

2. PRIOR RATINGS

We define the term *prior ratings* as users' assessment 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 [10]. 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). 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

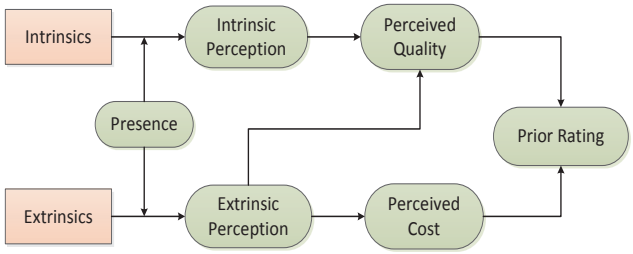

Figure 1: The conceptual model of prior ratings

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.

It is reported that VR real-time interactions enable users to possess a strong sense of being in a mediated environment and gain a lifelike shopping experience [10]. Hence, we draw a hypothesis as follows.

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.*

Jiang and Benbasat [9] 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.

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.*

Conceptual Model. A conceptual model of prior ratings is illustrated in Figure 1. Such a model allows a principled basis for the elicitation and analysis 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. 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.

Presence is defined as subjects' 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 [14]. In this paper, presence 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. Since 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.

[1]See question 2 for the tested environments in Figure 2.

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

Intrinsic attributes (e.g., workmanship, size) have a direct impact on perceived quality during the goal-directed process of pre-purchase product evaluation [3]. In this paper, 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.

Extrinsic attributes (e.g., price) 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 [1]. In this case, price is often used by users to infer the quality of products [1]. Brand name serves as a 'shorthand' for perceived quality by providing users with a bundle of product information [8]. In contrast, store name also has a positive but small impact on perceived quality [12]. Other extrinsic attributes (e.g., warranty) may also have effects on perceived quality.

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 (mainly price) and non-monetary (e.g., time, shipping) attributes.

Prior Ratings. 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.

HYPOTHESIS 5. *Perceived quality has significantly positive influence on prior ratings, and perceived cost will also positively influence prior ratings, if the price is acceptable.*

3. USER STUDY

Two kinds of mediated environments are developed: 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. Both user interfaces 'sell' our t-shirts whose source was the real-life commerce website 80stees.com. These 50 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. 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

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?*

Figure 2: Questions in the user study

can also adjust the avatar's shape as desired. 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. Four attributes, namely appearance, material, fit and price are identified as the major concerns of users via a pilot study. Together with attributes store and category, the six attributes are selected for the experiments.

The user study consisted of one session, structured as follows. All subjects started with a tutorial video so as to get familiar with functionalities of user interfaces. Once subjects were comfortable, they proceeded. Each subject experienced and evaluated eight different 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. In addition to t-shirts, the subjects were also 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. The questions are shown in Figure 2.

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 1 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. This may be partially explained by the fact that users have stronger sense of presence in VR than in WS. Besides, subjects also commented that the main reasons are due to the fact that they can try t-shirts on their 'own' body rather than have to image 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.

Table 1: The 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**

Table 2: 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

For Hypothesis 2(b), the collected data consists of 215 prior ratings in WS ($R.ws$) and 218 records in VR ($R.vr$) (see Table 2). 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 rating values (4, 5) larger than median scale (3) as positive, and the remainder (1, 2, 3) as negative. Then we can 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, respectively. 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 intrinsic or extrinsic attributes in WS and VR as dependent variable. The results are illustrated in Table 3. 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 1. 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 in two environments is that price in WS may be ignored as a cue to infer user preference. Thus, Hypothesis 3 is partially supported.

For Hypothesis 4, we conducted a multi-variable linear regression: intrinsic and extrinsic attributes as independent variables and 'perceived quality' as dependent variable. The results, presented in Table 4, 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. Note that the regression coefficients of category and store are greater than material and fit, which means that perceived quality re-

Table 3: The influences of presence on attributes

Env.	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 4: The evaluations of perceived quality

Env.	Attributes	Estimate	T Value	Pr(> |t|)
WS	appearance	-0.06649	-1.152	0.250485
	material	0.28331	5.729	**3.52e-08**
	fit	0.12482	2.130	0.034338
	category	0.31082	5.115	**7.11e-07**
	price	0.04622	0.975	0.330593
	store	0.21357	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

lies more on extrinsic attributes than on intrinsic attributes. In contrast, the most important attributes in VR are appearance, material, fit and price, and most comments are focused on these four attributes. Hence, subjects relied more on intrinsic attributes than extrinsic attributes to evaluate the quality of t-shirts in VR. Of the four major attributes identified from pilot study, we find that only one of them (material) is recognized 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 by applying a linear regression analysis. The results show 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, indicating that the price is acceptable in general. 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. Thus, Hypothesis 5 is supported.

4. DISCUSSION AND CONCLUSION

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 class 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 aforementioned, this limitation may not greatly influence the general conclusion since we exploited abstract attributes rather than some specific attributes. Third, most subjects 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. 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. The current research focused on the validation of conceptual model of prior ratings. For future work, we intend to investigate the benefits of prior ratings in improving recommendation performance.

Acknowledgement. Supported by the Institute for Media Innovation at Nanyang Technological University, Singapore.

5. REFERENCES

[1] W. Dodds, K. Monroe, and D. Grewal. Effects of price, brand, and store information on buyers' product evaluations. *J. of Marketing Research*, 28(3):307–319, 1991.

[2] R. Dong, K. McCarthy, M. O'Mahony, M. Schaal, and B. Smyth. Towards an intelligent reviewer's assistant: recommending topics to help users to write better product reviews. In *Proc. IUI'12*, pages 159–168, 2012.

[3] S. Gardial, D. Clemons, R. Woodruff, D. Schumann, and M. Burns. Comparing consumers' recall of prepurchase and postpurchase product evaluation experiences. *Journal of Consumer Research*, 20(4):548–560, 1994.

[4] G. Guo, J. Zhang, and D. Thalmann. A simple but effective method to incorporate trusted neighbors in recommender systems. In *Proc. UMAP'12*, pages 114–125, 2012.

[5] G. Guo, J. Zhang, and N. Yorke-Smith. A novel bayesian similarity measure for recommender systems. In *Proc. IJCAI'13*, 2013.

[6] P. Hemp. Avatar-based marketing. *Harvard Business Review*, 84(6):48–57, 2006.

[7] X. Hu and K. Wang. Personalized recommendation for virtual reality. In *Proc. ICMT'10*, pages 1–5, 2010.

[8] J. Jacoby, G. Szybillo, and J. Busato-Schach. Information acquisition behavior in brand choice situations. *Journal of Consumer Research*, 3(4):209–216, 1977.

[9] Z. Jiang and I. Benbasat. 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(3):111–147, 2004.

[10] H. Li, T. Daugherty, and F. Biocca. The role of virtual experience in consumer learning. *Journal of Consumer Psychology*, 13(4):395–407, 2003.

[11] P. Massa and P. Avesani. Trust-aware recommender systems. In *Proc. RecSys'07*, pages 17–24, 2007.

[12] A. Rao and K. Monroe. The effect of price, brand name, and store name on buyers' perceptions of product quality. *J. of Marketing Research*, 26(3):351–357, 1989.

[13] F. Shah, P. Bell, and G. Sukthankar. A destination recommendation system for virtual worlds. In *Proc. FLAIRS'10*, pages 475–476, 2010.

[14] M. Slater, B. Spanlang, and D. Corominas. Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. *ACM Transactions on Graphics (TOG)*, 29(4):92, 2010.