

The Evaluation of a Hybrid Critiquing System with Preference-based Recommendations Organization

Li Chen and Pearl Pu
Human Computer Interaction Group
School of Computer and Communication Sciences
Swiss Federal Institute of Technology in Lausanne (EPFL)
CH-1015, Lausanne, Switzerland
{li.chen, pearl.pu}@epfl.ch

ABSTRACT

The critiquing-based recommender system mainly aims to guide users to make an accurate and confident decision, while requiring them to consume a low level of effort. We have previously found that the hybrid critiquing system of combining the strengths from both system-proposed critiques and user self-motivated critiquing facility can highly improve users' subjective perceptions such as their decision confidence and trusting intentions. In this paper, we continue to investigate how to further reduce users' objective decision effort (e.g. time consumption) in such system by increasing the critique prediction accuracy of the system-proposed critiques. By means of real user evaluation, we proved that a new hybrid critiquing system design that integrates the *preference-based recommendations organization* technique for critiques suggestion can effectively help to increase the proposed critiques' application frequency and significantly contribute to saving users' task time and interaction effort.

ACM Categories & Subject Descriptors

H5.2 [Information interfaces and presentation]: User Interfaces – Evaluation/methodology, Graphical user interfaces (GUI), User-centered design.

General Terms

Design, Experimentation, Human Factors.

Keywords

Hybrid critiquing-based recommender systems, preference-based recommendations organization, example critiquing, user study.

1. INTRODUCTION

Many highly interactive recommender systems engage users in a conversational dialog in order to learn their preferences and use their feedback to improve the system's recommendation accuracy. Such interaction models have been referred as conversational recommenders [9] or critiquing-based recommender systems [2,3]. The main component of the interaction is that of example-and-critique. The system simulates an artificial salesperson that

recommends example options based on a user's current preferences and then elicits her feedback in the form of critiques such as "I would like something cheaper" or "with faster processor speed". These critiques form the critical feedback mechanism to help the system improve its accuracy in predicting the user's needs in the next recommendation cycle.

The intelligent critiquing support was shown allowing users to more effectively refine the quality of their preferences and improve their decision accuracy up to a higher degree, compared to the non critiquing-based system such as a ranked list [7]. In recent years, two approaches have been mainly developed to realize such critiquing tools. One is the system-proposed critique generation method that aims at proposing a set of critiques for users to choose [1,6], and another is the user self-motivated critiquing support that stimulates users to freely compose and combine critiques on their own [2,7]. A comparative user evaluation indicates that the user self-motivated critiquing support enables users to achieve higher confidence in choice and decision accuracy, whereas the system-proposed critiques are more intuitive to use and would likely accelerate users' decision process if the critiques closely match to what users intend to make [2].

To converge the strengths from both the critiquing approaches, we have proposed to combine them into a hybrid critiquing-based recommender system [3], and found that this kind of hybrid system could positively affect users' subjective perceptions such as their decision confidence and trusting intentions, since they can not only obtain knowledge of the domain and easily perform critiquing via the proposed critiques, but also have the opportunity to freely create critiques by themselves if necessary with the aid of the user self-motivated critiquing support.

On the other hand, given the limitation of traditional system-proposed critique generation approaches in predicting users' intended critiquing criteria (due to their purely data-driven selection mechanism [6]), we have developed a preference-based recommendations organization technique to compute and organize critiques according to user stated and potential preferences. A simulation trial revealed that it could reach a higher level of critique prediction accuracy compared to other related methods, with the resulting potential benefit of saving users' decision effort to a certain extent [4].

Based on the previous findings, in this paper, we introduce a new hybrid critiquing system design by combining the *preference-based recommendations organization* (as system-proposed critiques) with a user self-motivated *example critiquing* agent [2,7]. In particular, we attempt to demonstrate from real users'

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

RecSys '07, October 19–20, 2007, Minneapolis, Minnesota, USA.
Copyright 2007 ACM 978-1-59593-730-8/07/0010...\$5.00.

evaluation that this hybrid critiquing system can not only enable users to promote high level of subjective perceptions, but also have the ability to improve their objective decision performance (e.g. time consumption and interaction effort) given the integration of the preference-based organization interface.

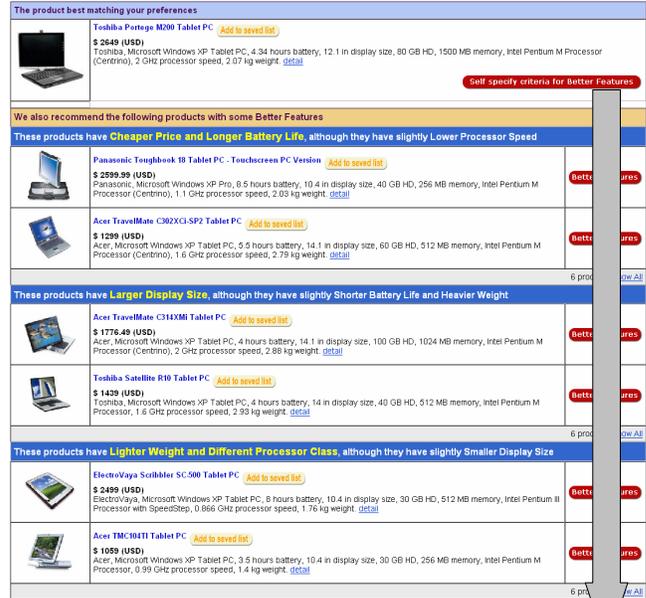
2. HYBRID CRITIQUING SYSTEM WITH PREFERENCE-BASED RECOMMENDATIONS ORGANIZATION

In the following, we first briefly describe the preference-based recommendations organization and the example critiquing support, and then introduce how we combine them into a hybrid critiquing system.

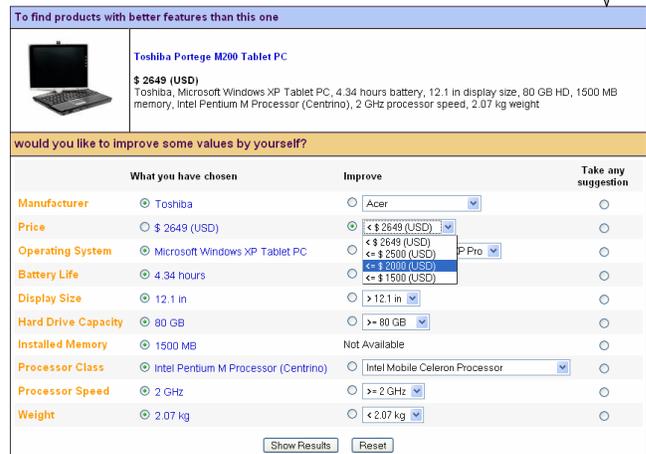
The key idea of the preference-based organization interface is to organize all recommendations (computed according to user preferences) into different categories based on their similar tradeoff properties (e.g. *improved* or *compromised* features) compared to the top candidate, and use the category titles (e.g. “these products have cheaper price and longer battery life, ...”) as proposed critiques (see Figure 1.a). To derive effective principles for this interface design, we tested more than 13 paper prototypes with real users and finally concluded five primary principles [4,8], including proposing improvements and compromises in the category title using conversational language, keeping the number of tradeoff attributes in the category title under five, including actual products in a category, and diversifying the proposed categories and their contained products.

We accordingly designed and implemented a computation algorithm to generate such organization interface. Briefly speaking, the algorithm contains three main steps (see details in [4]): 1) user preferences are first modeled based on the multi-attribute utility theory [5] since it can resolve conflicting preference values explicitly by considering tradeoffs; 2) the recommendations are computed based on the user’s preference model, and each of them (except the top candidate) is converted into a tradeoff vector, a set of (*attribute, tradeoff*) pairs where *tradeoff* indicates the *improved* or *compromised* property of the product’s *attribute* value compared to the same *attribute* of the top candidate. For the attributes without explicitly stated preferences, default properties are suggested (e.g. the cheaper, the better); 3) all of the tradeoff vectors are then organized into different categories by utilizing an association rule mining technique to discover the recurring subsets of (*attribute, tradeoff*) pairs used for category titles (i.e. critiques). Since a large amount of subsets would be produced by the mining algorithm, they are further ranked and diversified to select the ones with higher tradeoff utilities (i.e. gains vs. losses relative to user preferences) in terms of both the critiques themselves and their contained products, so as to be finally presented in the organization interface.

The example critiquing support [2,7], as another important component of the hybrid critiquing system, is a purely user self-motivated critiquing agent since it emphasizes on facilitating users to freely build and combine critiques (i.e. identifying a single or a set of features to improve or compromise) on their own (see Figure 1.b). It is also able to support different critiquing types such as similarity-based (e.g. “Find some products similar to this one”), quality-based (e.g. “Cheaper”, “Bigger”), and more concrete critiquing based on quantity (e.g. “\$500 cheaper”, “1 inch bigger”) (see details in [2]).



a. The preference-based recommendations organization (Pref-ORG).



b. The user self-motivated example critiquing interface (EC).

Figure 1. The hybrid critiquing-based recommender system of combining Pref-ORG and EC.

Recent series of experiments revealed the respective potential advantages of the preference-based recommendations organization technique (henceforth Pref-ORG) and the example critiquing support (henceforth EC) [2,4]. In Pref-ORG, since critiques are generated and proposed according to user preferences, they could be more accurate in matching to the user’s intended critiquing so as to likely accelerate her decision process. They could also expose to users the remaining recommendation opportunities and stimulate them to make critiques for a better choice. As for EC, the most obvious advantage is its ability to give users total control over the process of their critiques making, so that the users’ decision accuracy could be highly enhanced.

The goal of developing the hybrid critiquing system is therefore to combine the two techniques’ advantages and make them compensate for each other to better serve real users. More concretely, Figure 1 shows the two main interfaces in the hybrid system. After a user specifies her initial preferences, the best

matching product will be first returned at the top followed by four preference-based critique suggestions, each containing 2 sample products that satisfy the critique (see Figure 1.a). If the user is interested in one of the proposed critiques, she could click to see more products under the critique. Among all of the products recommended with those critiques, the user can either choose one as her final choice, or select one near-target and click the “Better Features” (along with it) to see a new set of suggested critiques relative to the selected product.

Even more, if no critique and product interests the user in the organization interface, she could switch to make the self-motivated critiquing by clicking the button “Self specify criteria for Better Features” along with the top candidate (see Figure 1.a). At this point, the example critiquing interface will be shown (see Figure 1.b) to facilitate the user to freely create critique on one single feature or combine critiques on any set of multiple features simultaneously. After this critiquing process, a set of tradeoff alternatives that best match the user’s self-specified critiquing criteria will be returned to her to compare.

Either the action of selecting the system-proposed critique or making the user self-motivated critiquing is regarded as one critiquing cycle during which the user’s preference model will be automatically refined to respect her current needs.

3. USER EVALUATION

In order to understand whether the integration of the preference-based organization interface in the hybrid critiquing system could particularly contribute to save users’ objective decision effort as well as improving their subjective perceptions, we conducted an experimental evaluation involving real users to compare the hybrid system (henceforth Pref-ORG+EC) with an original design (as described in [3]) that combines EC with a traditional system-proposed critiquing interface, the dynamic critiquing agent [6] (henceforth this hybrid system noted as DC+EC). In the dynamic critiquing interface, although critiques are also automatically computed by mining the recurring sets of unit differences between the current recommendation and remaining products, they are filtered and selected purely according to the critiques’ support values (i.e. favoring the critiques with less percentage of products that satisfy them), rather than user preferences that are emphasized in Pref-ORG. We therefore called the traditional dynamic critiquing approach the data-driven critique generation.

The user evaluation was conducted in a between-group design. All participants were randomly and evenly divided into two groups, and each group was assigned one system (Pref-ORG+EC or DC+EC) to evaluate. In addition, both the hybrid critiquing systems were developed for two product catalogs: tablet PC (55 products, each described by 10 main features) and digital camera (64 products characterized by 8 main features), and every participant was randomly assigned one product domain to search.

A total of 44 (8 females) volunteers participated in the experiment. Most of them are students in the university, but they are from a variety of different countries (France, Italy, Switzerland, China, etc.), studying varied subjects (computer science, mechanics, manufacturing, etc.) and pursuing different levels of educational degrees (bachelor, master, or Ph.D.). The main user task was to “find a product you would purchase if given the opportunity” with the assigned system. After the choice was made, the participant was asked to fill in a post-study

questionnaire asking about his/her subjective perceptions (e.g. perceived cognitive effort, decision confidence, and trusting intentions) with the interfaces s/he just used.

3.1 Time and Interaction Effort

Experimental results show that in terms of the time consumption, Pref-ORG+EC demands significantly less time than DC+EC ($t = 2.32, p < 0.05$). More concretely, the participants who used Pref-ORG+EC spent average 4.07 minutes in locating their choice, while the other group with DC+EC consumed more time (5.98 minutes; see Figure 2 of the time distribution).

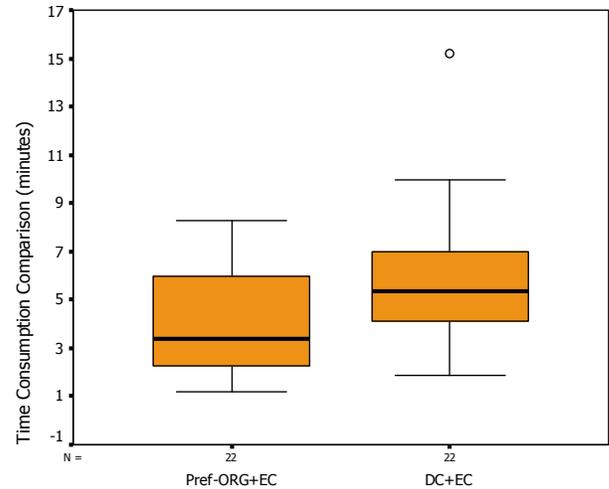


Figure 2. The task time distribution (minimum, median, maximum, etc.) in the two systems.

We were further interested in exploring the actual interaction effort users expended within their task time. Specifically, the interaction effort was measured as the whole interaction session (i.e. the total number of visited pages) the user took while using the system. The pages include the initial preferences entering page, the search results page, the critiquing page, the product’s detailed specification page and the saved list page. All of these pages were implemented in both systems. The result showed that in Pref-ORG+EC the average interaction effort is 6.23, which is significantly less than the average effort consumed in DC+EC (mean = 10.59; $t = 2.85, p < 0.01$). Furthermore, with respect to the number of products users viewed in both systems, we found that, likely due to the organization interface design, in Pref-ORG+EC 53.5 products (including repeated ones) were averagely displayed for each user, versus 22.3 products in DC+EC ($t = -3.73, p < 0.01$).

We additionally in depth measured the critiquing cycles that were contained in the interaction effort to indicate the frequency users consulted with the critiquing agent (either the system-proposed or user self-motivated) to refine their preferences. It was revealed that the participant was on average involved in 3.68 critiquing cycles while using Pref-ORG+EC, compared to 4.86 cycles with DC+EC ($t = 1.42, p = 0.16$). Moreover, 54.3% of the critiquing cycles in Pref-ORG+EC were used to pick the preference-based critique suggestions, against 23.4% critiquing cycles in DC+EC to select the proposed critiques based on traditional dynamic critiquing technique (see Figure 3). It can be hence implied that the preference-based critique generation may have a better

prediction on users' intended critiquing criteria (since users chose to use it more actively in reality), than the purely date-driven dynamic critiquing method. This result again verifies our previous simulation finding [4].

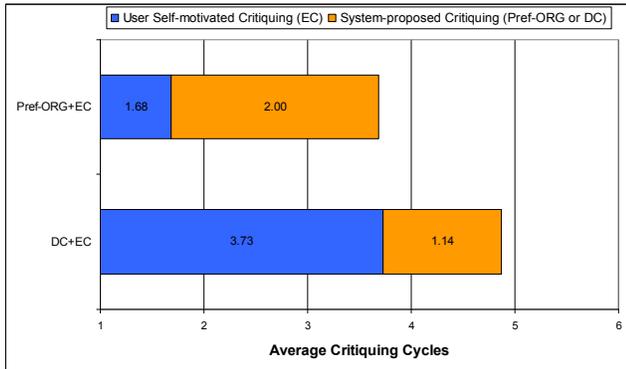


Figure 3. The critiquing applications in the two systems.

3.2 Subjective Perceptions

In addition to the two main objective standards (i.e. the time and interaction effort participants consumed), it is also important to measure their subjective perceptions with the interfaces.

As discussed in [2,3], we primarily consider four subjective aspects: the *perceived cognitive effort* indicating the amount of subjective effort users perceived to exert, the *decision confidence* questioning about whether users were confident that they made the best choice with the system, and two trusting intentions inferring whether the system can convince its users to purchase a product (i.e. *intention to purchase*) and stimulate them to return to the system for future use (i.e. *intention to return*). The actual questions used can be found in [3]. Each question elicited a response on a 5-point Likert scale.

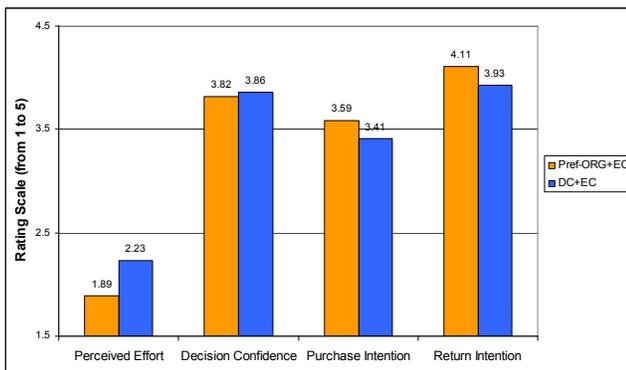


Figure 4. Users' subjective perceptions with the two systems.

Analysis of users' answers shows that both groups of participants indicated positive perceptions regarding the four subjective aspects for both systems, and Pref-ORG+EC gained slightly better scores on most of them (see Figure 4). More concretely, both types of interfaces allow a low level of cognitive effort, which was additionally perceived marginally significantly lower ($t = 1.71$, $p = 0.09$) in Pref-ORG+EC (mean = 1.89 vs. 2.23 in DC+EC). As for decision confidence, the average rate was around equally high for both systems (mean = 3.82 in Pref-ORG+EC vs. 3.86 in DC+EC, $t = 0.31$, $p = 0.78$), which again verifies our

previous statement that since in hybrid critiquing systems users can have the total freedom of identifying their truly-intended critiquing criteria with different supports, they would be much confident of the accuracy of their final choice. Furthermore, both groups of participants also expressed high trusting intentions respectively in both systems, and the group with Pref-ORG+EC had slightly higher intention to purchase the chosen product (mean = 3.59 vs. 3.41 with DC+EC, $t = -0.75$, $p = 0.45$) and higher intention to return to the system for future use (mean = 4.11 vs. 3.93 in DC+EC, $t = -0.83$, $p = 0.41$).

4. CONCLUSION

As a conclusion, in this paper, we introduced and evaluated a new hybrid critiquing system design that combines the *preference-based recommendations organization* technique with the user self-motivated *example critiquing* support so as to assist users in making critiques and decisions in an informative way. Since in the hybrid critiquing system, users can not only feely refine their preferences with the aid of the example critiquing support, but also have the opportunity to learn the remaining recommendation opportunities and accelerate their critiquing process in the preference-based organization interface, they promoted highly positive level of subjective perceptions (i.e. perceived cognitive effort, decision confidence and trusting intentions) with the system as shown in the experiment. More notably, it is demonstrated from the user evaluation that given the integration of the preference-based recommendations organization for critiques suggestion, the system enables users to consume significantly less time and interaction effort in decision making, relative to user performance in another hybrid critiquing system with traditional data-driven critique generation approach.

5. REFERENCES

- [1] Burke, R., Hammond, K., and Young, B. The FindMe approach to assisted browsing. *IEEE Expert: Intelligent Systems and Their Applications* 12, 4 (1997), 32-40.
- [2] Chen, L. and Pu, P. Evaluating critiquing-based recommender agents. In *Proc. AAAI'06*, 2006, 157-162.
- [3] Chen, L. and Pu, P. Hybrid critiquing-based recommender systems. In *Proc. IUI'07*, 2007, 22-31.
- [4] Chen, L. and Pu, P. Preference-based organization interfaces: aiding user critiques in recommender systems. In *Proc. UM'07*, 2007, 77-86.
- [5] Keeney, R. and Raiffa, H. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge University Press, 1976.
- [6] McCarthy, K., Reilly, J., McGinty, L., and Smyth, B. Experiments in dynamic critiquing. In *Proc. IUI'05*, 2005, 175-182.
- [7] Pu, P. and Chen, L. Integrating tradeoff support in product search tools for e-commerce sites. In *Proc. ACM EC'05*, 2005, 269-278.
- [8] Pu, P. and Chen, L. Trust building with explanation interfaces. In *Proc. IUI'06*, 2006, 93-100.
- [9] Thompson, C.A., Goker, M.H., and Langley, P. A personalized system for conversational recommendations. *Journal of Artificial Intelligence Research* 21 (2004), 393-428.