Towards Three-Stage Recommender Support for Online Consumers: Implications from a User Study

Li Chen

Department of Computer Science, Hong Kong Baptist University
Hong Kong, China
lichen@comp.hkbu.edu.hk

Abstract. In this paper, a three-stage recommender support was implied from a user study. The purpose of the user study was to understand how to best utilize different types of social information (e.g., product popularity, user reviews) for facilitating online consumers' decision-making process in the e-commerce environment. Through both of in-depth tracking users' objective behavior and qualitative interviewing their reflective thoughts, we have not only refined a traditional two-stage decision process into a more precise three-stage process, but also identified at each stage what information users are inclined to seek for. Based on the study's results, suggestions were made to related recommender systems about their practical roles in the three-stage framework and how they can more effectively support users' information needs.

Keywords: user study, e-commerce, recommender supports, Flickr camera finder, complex decision making, high-value products, users' information needs.

1 Introduction

Social content has always been recognized to play important role in a consumer's hybrid decision process in which the decision maker seeks for advices for the purpose of reducing the uncertainty and the amount of information that must be processed to make a decision [9]. As a matter of fact, in recent years, recommender systems have been broadly developed in order to suggest items that users may be interested in [2]. For high-value, infrequently experienced products (e.g. a digital camera, a computer, and a house), the typically applied recommendation in existing e-stores is "people viewed this product also view others …" (or "people bought this product also bought …"). These recommended products are normally computed by the item-based method [2], which identifies a set of items that are most similar to the user's currently viewed one, in combination with other users' clicking or purchase histories.

However, most of related works have just focused on providing recommendations from the perspective of algorithm improvement. No much work has studied on when these recommendations will be most relevant to the user's actual needs within her/his whole decision process. Indeed, according to Adaptive Decision Theory [8], human's decision making is in nature a constructive and adaptive process, which basically follows two stages: 1) to screen down the number of available alternatives to a

L. Chen, P. Triantafillou, and T. Suel (Eds.): WISE 2010, LNCS 6488, pp. 365–375, 2010.

reduced consideration set, and 2) to in-depth examine the selected candidates and obtain the final choice. The questions are then that, at each stage, what kinds of social content (i.e., other users' generated product content) will be particularly interesting to the current user, and whether a recommender support that incorporates the required social resources could more effectively benefit the user.

With these questions, we have launched a user study with the goal of observing users' decision behavior when they are looking for a high-value product to buy. More concretely, two objectives of the study are: 1) tracing users' decision-making process to refine the basic two-stage model; and 2) understanding users' social information needs at each stage. Two kinds of social resources were especially investigated in the study. One is *product reviews*, which are what other users generate according to their post-purchase evaluation experiences with the product. The reviews are expressed in form of numerical ratings or natural languages such as in Yahoo shopping and Amazon. Another is *product popularity* info (e.g., "the top products"). In this regard, we have involved a newly released Flickr Camera Finder, since it primarily provides the popularity data based on the statistics of Flickr community members who have uploaded images or videos with a particular camera over a certain time. Its involvement is not only because the product domain is what we emphasize (i.e., the high-value product), but also its usage-driven popularity generation is different from traditional purchases or promotion driven method (as been used in most e-commerce sites). Through the empirical study, we can hence observe whether users could in reality perceive the difference between the two types of popularity data (i.e., one from social media site, and another from standard e-commerce site), and know whether the novel type could support users to make a more confident decision.

As a result of the user study, a three-stage recommender framework is established that suggests a set of directions regarding how to improve related decision systems and best assist users when they are at different decision stages. In the following, we will introduce our research questions and describe how our experiment was setup and the set of suggestions concluded from the user study.

2 Research Questions and Experiment Setup

2.2 Research Questions

The amount of cognitive effort applied to the decision process is in essence related to the level of importance that the consumers place on the specific product [5]. As for high-value products that are expensive and infrequently experienced (e.g., cameras, computers, cars), extensive decision-making effort is commonly spent by consumers in seeking information and deciding. Accordingly, as mentioned earlier, researchers from the psychology field describe a two-stage process in this condition, where the depth of cognitive load and information processing varies [8].

In the last decade, recommender systems have been widely developed, but most works were oriented to low-value, public taste products (e.g., music, movie, webpage, book) [6,7]. Little has exploited the effect of social information (i.e., user-generated content) on producing recommendations for high-value products. In this area, most systems have been still based on products' static attributes (e.g., the camera's optical zoom, megapixels) to model users' information needs. For instance, the critiquing-based recommenders such as Dynamic-Critiquing, FindMe and Example-Critiquing [4] can support the incremental refinement of user preferences via providing the critiquing facility (e.g., "I would like something cheaper", "with faster processor speed"), but the user's preference model is purely built on static attribute values. [10] has conducted a tentative experiment that discovered merits of product ratings and reviews as part of recommendations to influence users' searching strategies. However, it is still not clear how the social content can be integrated with static product attributes to construct more accurate user model.

Given the existing limitations, our objective was to identify the relative importance of social content (e.g., product popularity, product reviews, etc.) in users' complex decision process. Specifically, we have targeted to answer the following research questions by means of a user study:

1) How would consumers in practice follow the basic two-stage process when interacting with the e-stores? Would the two-stage be further refined, and at each stage, how would different kinds of social content act to assist the user? More deeply, we have examined the following three aspects:

   a. How would users perceive the so called personalized recommendations that suggest items based on the user's current click (e.g., "*people viewed this product also viewed others*")?
   b. Would there be any differences of users' perception and usage of the *product popularity info* as from social media sites (e.g., Flickr), relative to from the standard e-commerce sites? And if there are differences, which type of popularity would be more effective?
   c. As for *product reviews*, what would be their primary role in users' decision process and which kind of reviews (i.e., positive vs. negative reviews) would be more interesting to the buyer?

2) How would we develop more useful recommender supports applicable at each stage so as to maximize the benefits of social content that users will intentionally rely on?

2.2 Experiment Setup

In order to obtain answers to above questions, we have conducted an experiment that recorded users' decision behavior when they were assigned the task of looking for an item to buy. As mentioned before, we used Flickr Camera Finder as one experiment material because it provides the unique source of usage-based popularity info. Besides, a standard e-commerce site was also offered, which provides the complete set of static attribute values, the traditional popularity data (so as to be compared with Flickr's), product reviews and personalized recommendations. For this site, we have considered a number of options, including Amazon, Yahoo Shopping, shopping.com, etc, and finally selected Yahoo Shopping because it is not only based on the same product database that Flickr Camera Finder (CF) uses, but also can be representative of other e-commerce sites regarding information amount and information diversity.

The experiment was concretely designed in a free-choice scenario where users were allowed to freely select and examine any product info that can be obtained from the two sites: Flickr Camera Finder (www.flickr.com/cameras/, Flickr CF for short) and Yahoo Shopping (shopping.yahoo.com) for Cameras (Yahoo CF for short). Note that this is not a comparative study and we were not to identify which site is better. Our goal was to reveal what kind of social info that the two sites provide can be in reality adopted by users.

We have finally recruited twelve motivated volunteers (three females) because they were interested in buying a digital camera at the time of our experiment. They are Master or PhD students in our department with ages between 20 and 40. They have often visited e-stores (at least once every three months), which indicates that they can be representative of online consumers to some extent. It is also worth noting that all subjects were first-time encounters to the two websites (i.e., Flickr CF and Yahoo CF), so that their behavior would not be biased by any of previous usage experience.

In the experiment, an initial warm-up period (10 minutes) was first given to each participant for her/him to be familiar with the two sites' facilities as much as possible. Then s/he was assigned the task: "*Imagine you are prepared to buy a digital camera, please use the provided two sites to examine product information and find a product that best meets your needs*." During the formal trial, all of the user's interaction events, including on-screen mouse moves, clicks and keyboard inputs, were automatically captured by a screen observer software (Morae). Then, after the participant finished the task, a semi-structured interview was conducted by the administrator in order to obtain the participant's reflective thoughts and opinions on various examined aspects.

3 Results Analysis and System Implications

3.1 Result 1: Three-Stage Decision Process

The analysis of all users' decision-making behavior surprisingly shows that they all exhibited a more precise, three-stage process: 1) to screen all alternatives and select one for in-depth evaluation; 2) to view the selected product's details and save it in wish list if near-satisfactory; 3) to compare candidates in the wish list and make the final choice. Moreover, the transition between these stages did not follow a sequential order, but was iterative in nature and the size of consideration set gradually decreased. Concretely, at the start, all users were with some initial preferences in mind, e.g., looking for a camera that is "*easy to use*", "*easy to carry*", *"with colorful images", "of high cost performance"*, "*better for night scenes*", or "*better for long distance picture-taking"*. Some users (6) had specific criteria on product's static attributes (e.g., on type, megapixels, screen size, battery, focal length). As for price, 5 users expressed the constraint (e.g., "cheap", "not expensive", "under $500").

Their interaction logs further show that they maximally considered four brands within the whole decision session. Their information-seeking process was typically brand-based. That is, they first looked into one preferred brand. If anyone product(s) interested them when they browsed the brand's product list (Stage 1, see Figure 1 left), they went to examine the product's details and saved it into the wish list if the details were near-satisfying (Stage 2). After reviewing one brand's products, they switched to another preferred brand and performed the similar browsing process. The iterative cycle between stages 1 and 2 continued until a set of candidates was determined (the average number of candidates is 4. See the analysis later). At this point, they entered into Stage 3 to compare candidates in their wish list and confirmed the final choice. Due to the common behavior exhibited by all participants, we came up with the three-stage decision process model (see Figure 1 left) with the processed input and output at each stage.

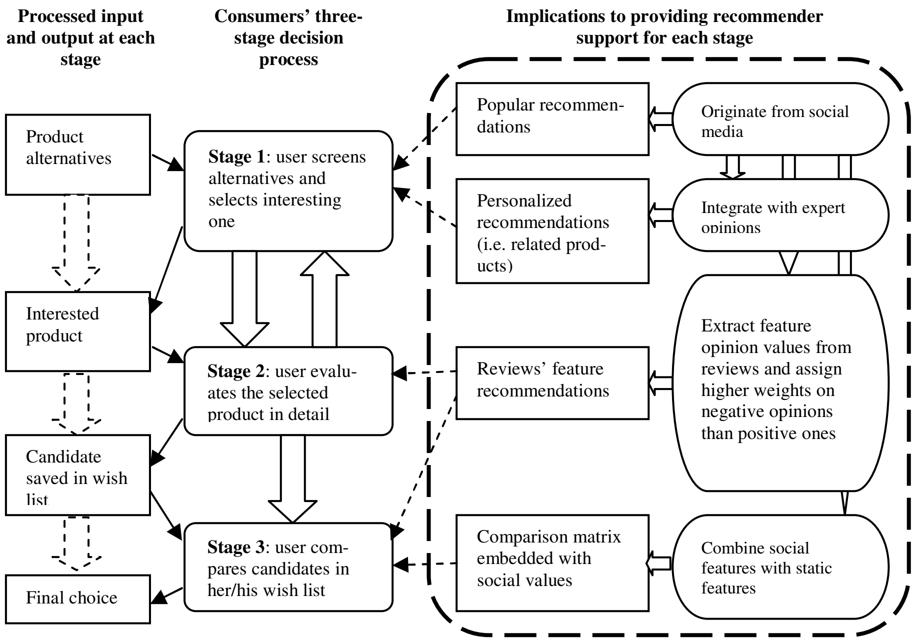

Fig. 1. Three-stage recommender framework, with the goal of facilitating an online buyer's information seeking need at each decision stage

3.2 Result 2: Implications to Providing Recommender Supports at Each Stage

It is then interesting to know what product information (i.e., social/static content) that users have processed at each stage and what kind of recommender supports could be relevant to their information need. Figure 1 (right) summarizes system implications. In the following, we will in detail explain how the implications were derived from the user study's results.

3.2.1 Stage 1: Screening Out Interesting Products

3.2.1.1 Users' Objective Behavior

At this stage, we measured how many products were selected (i.e. clicked by the user for details) and from where these products were located. It indicates that on average 9.67 (*St.d.* = 4.78) products were chosen to view details, among which 5.42 were located in Yahoo, and 4.25 were in Flickr CF. Figure 2 (left) concretely shows the distribution of these products' locations. In fact, basic static features provided by

Yahoo for the browsing/filtering got the highest chance that enabled the average user to obtain 39.79% interesting products. The second and third winners came to Flickr CF's popularity-based sorting list (27.51%) and brand popular products (12.18%) respectively. In comparison, Yahoo's popularity list got much less hits (5.28%). There were in fact only 2 participants who accessed "Top Digital Cameras" in Yahoo, against 9 in Flickr CF. As for the remaining selected products, they were either found through keyword-based search (for example, the user input a model's name) (6.53%), or through Flickr's photo-related products (4.83%), or Yahoo's recommendations (i.e., "*shopper who viewed this also viewed …*", 3.89%).

Thus, in total, above half of interested products (53.69%) were stemming from social sources (see Figure 2, right). Product popularity was shown particularly more active relative to the other social contents at this stage. As one user said, "*popularity is a suitable proxy to measure the product's quality when I am not familiar with a brand or uncertain about what I want*".

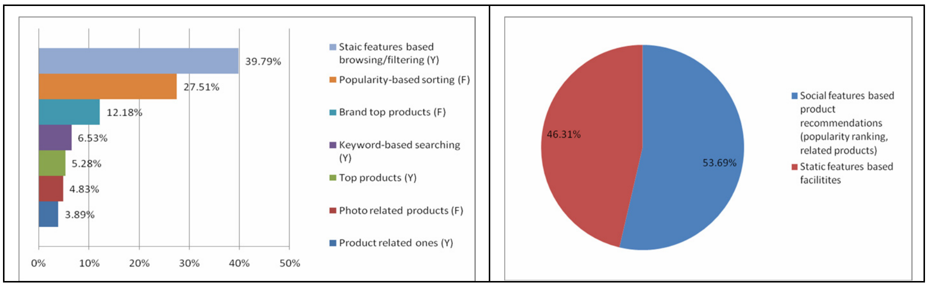

Fig. 2. (Stage 1) Distribution of product locations in Flickr CF (*F*) and Yahoo CF (*Y*), and their overall distribution as from social sources vs. from static sources

3.2.1.2 Users' Qualitative Comments

Users' qualitative comments further exposed their reflective thoughts on two aspects:

Credibility of product popularity. When being asked why they went to Flickr CF for accessing "product popularity", they replied that because it was perceived more trustworthy: "*I trust the information on the social forum.*" "*I trust Flickr's popularity information because of its large amount of users.*" "*Flickr is more neutral and credible.*" "*Although this is my first-time using this website, the information sounds credible since it should be based on actual usages.*" As for the product popularity on Yahoo CF, they commented "*the 'top products' in Yahoo may be only dependent on users' clicks or for companies' promotion purpose.*" "*The popularity information in Yahoo may be faked. It looks more trustworthy and real in Flickr.*" "*Flickr is more neutral because it is a consumer-operated website. The information on Yahoo may be not so real because it is more commercial-oriented.*" It can be hence seen that users propend to trust the data from the social media site like Flickr, because it is more dependent on a large community's real usages and less of commercial interests.

Additionally, users suggested several ways to improve the generation of product popularity. One user suggested involving the geographical distribution of community members, because "*one camera model is suitable for European, but probably not for*

Chinese." "*People from the same cultural background may have similar preferences.*" Another user proposed to take into account of the time dimension, given that "*it should be easier to compare the popularity values of different products if they were released at the same time.*"

Expert opinions to be integrated into "Related Products". As for "Related Products" (e.g., "shoppers who viewed this product also viewed …"), users commented that if they could integrate with experts' professional opinions on the relevance of currently viewed product with others, they can be more meaningful than being purely dependent on other consumers' clicking behavior. As one user said, "*imagine the friends around you all use XX, you would be not familiar with YY. But if there is a comparison table from an expert explaining what their differences are, I will go to see YY's products.*" Another user also noted "*because people sometimes just randomly clicked, the information from 'shopper viewed this product also viewed others' cannot be so credible. Experts' suggestions can be more useful to be regarded as important references.*" Moreover, users suggested that the recommendation can be even better if it incorporates their hard constraints, like on price range and product type, because "*the 'related products' are useful, but I will not be interested in them if they are beyond my price expectation.*" "*If the products are of the type that I prefer, I will more likely consider them.*" Users' comments can hence explain why they did not select many items from "Related Products" (as shown above). It also infers if this type of recommendation could be well integrated with experts' opinions and be matched to the user's constraints if any, the adoption degree will be potentially highly increased.

3.2.1.3 System Implications for this Stage

Thus, the findings indicate two system implications for the first stage (see Figure 1).

1) It is suggested that "product popularity" should better originate from the social media site, because in such platform, it can reflect real usages from a community of like-minded users (like from Flickr), and can be hence perceived more credible than in standard e-commerce sites. Moreover, a popularity-based recommender tool should be particularly referential and helpful when users have not formed their clear targets at the beginning. To provide such support, the popularity can be customized to involve contextual factors like contributors' regional properties and products' releasing time, so as to dynamically match to the current user's context.
2) "Related Products" can be likely enhanced by means of involving expert opinions and users' stated attribute constraints, so as to augment users' adoption degree of such recommendation. Furthermore, we believe that the provision of a hybrid mechanism that well unifies benefits from popularity-based recommendation and "Related Products" (i.e., the personalized recommendation) should be capable of facilitating users to locate more interesting products at the first stage.

3.2.2 Stage 2: Evaluating Product in Detail

3.2.2.1 Users' Objective Behavior

At the 2nd stage, we were interested in knowing what detailed info that users have evaluated after selecting a product from the 1st stage. The analysis of their page visits shows that on average 42.86% of selected products were in detail evaluated on Yahoo (that provides the product's full specifications and consumer ratings/reviews), 30.44% on Flickr CF (that provides the product's usage trend and resulting photos), and 26.70% on both sites' product pages. Among the evaluated products, 45.82% were finally saved into the average user's wish list (i.e., mean = 4 products, *St.d.* = 1.95). The page evaluations respectively contributed 39.09% (1.50 products), 6.25% (0.25) and 91.67% (2.17), to establishing the wish list (the % means the percent of products saved as candidates after the corresponding page evaluations. See Figure 3). It hence infers that the evaluation of product details from both Flickr CF and Yahoo CF can most likely convince the user to take the product as a candidate. The correlation is indeed highly significant ($p < 0.001$) by Pearson Coefficient. Another fact is that 91.7% users' final choice was a product that underwent this combined evaluation.

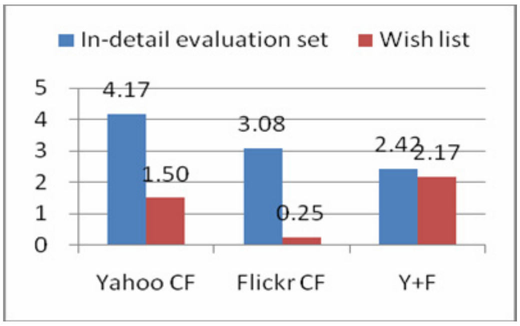

Fig. 3. (Stage 2) Correlation between evaluated products and saved ones to the wish list

3.2.2.2 Users' Qualitative Comments

Product reviews: negative versus positive reviews. Most of users noted that product ratings/reviews were very important when they examined a product, since "*they help me judge the product's true quality.*" As for ratings, the extreme ratings are more helpful than neutral ones: "*the middle rate cannot say anything.*" "*Human normally has vague consistency on the interpretation of middle scores, so extreme scales should be more useful.*" They said that when the product rating was low, they read user reviews (i.e., the textual commens). They liked the separation of reviews into *pros* and *cons* categories because it facilitates the comparison: "*The motivation of buying a product is not because it is very perfect, but is whether you can stand its drawbacks.*" "*Every product should have flaws, and what I want to get from user reviews is whether they can disclose these negative aspects.*" All participants agreed that "*I will not buy a product only because it has positive ratings and reviews, but will certainly not buy it if it has negative reviews, especially on features that I am concerned about.*" Moreover, the quantity of user reviews also takes a certain effect: "*less number of reviews will have lower credibility*", but it is still better than zero because "*in the case that two products both have few numbers of user reviews, I will still read the reviews to get the feeling of which product should be better.*"

3.2.2.3 System Implications for this Stage

Users' favor on negative comments than positive ones can be supported by a previous claim that "when information about an object or firm comes through the opinions of another person, negative information can be more credible and generalizable than positive information" [11]. Thus, we believe that the mining and exposure of negative opinion values and the indication of their relevance to users' interests can be likely to support more informed and accurate product evaluation at the stage 2 (see Figure 1). To achieve this goal, the sentimental analysis and opinion mining techniques [12] can be first applied to extract features (e.g., ease of use, image quality) from reviews and then associate opinion scores (i.e., positive, neutral and negative) to the features. Then, the recommendation can be returned in form of *{feature, opinion score}* sets, placing higher weights on negative opinions and further personalizing the sets with the user's specific feature concerns.

3.2.3 Stage 3: Comparing Candidates and Confirming the Final Choice

At the last stage when users nearly made the "purchase" decision, they compared candidates in their wish list and then identified the best one. In order to know which factors they mainly considered during the comparison, we recorded items they have viewed after their wish list was established. It shows that 66.7% (8 out of 12) users went to Flickr CF to compare candidate products' usage trends or photos (as taken by Flickr community), and 33.3% emphasized product specifications or reviews provided in Yahoo CF (see Figure 4 left).

From Figure 4 (right), we can see that totally 75% participants replied on social features (i.e., usage trends, product reviews, and community photos) to accomplish the choice confirmation. The social features are hence demonstrated more influential than static features (i.e., specifications) in convincing users about their "purchase" decision. Users' comments reflected that "*I would like to rely on the social content to identify which product should be better than others.*" "*The product's usage trend can help me form a correct judgment and decrease the uncertainty from purely evaluating its static specifications.*"

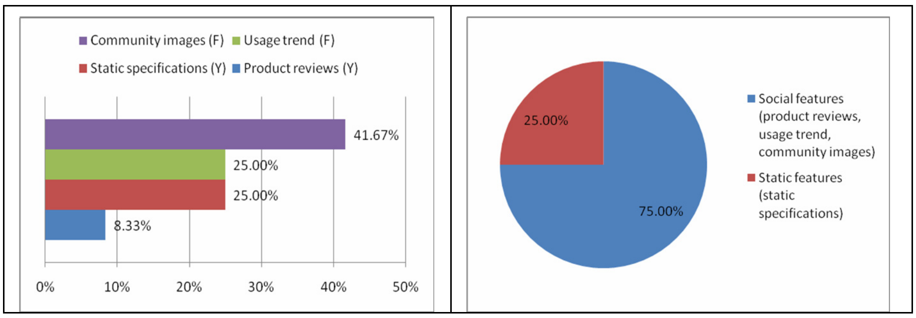

Fig. 4. (Stage 3) Major factors that users compared among candidates before they made the final choice, and these factors' overall distribution as from social sources vs. from static sources

3.2.3.1 System Implications for this Stage

For this stage, the implication is mainly about the comparison support (see Figure 1). Many existing e-commerce sites provide a comparison matrix, by which users can perform pair-wise comparison between multiple products in an *alternatives* (rows) x *attributes* (columns) matrix. Although this method has been demonstrated to enable higher decision quality than the condition without it [8], most of its applications are still limited to the simple display of products' static feature values. As implied from above findings, it will be likely more effective if the traditional comparison matrix can be improved to embed products' social values. For instance, the extracted *{feature, opinion}* pairs (implied in Section 3.2.2.3) can be also adopted here to complement standard *{feature, static value}* pairs. Besides, other types of social content (e.g., usage trend) can be also included in the matrix to facilitate users' comparison actions.

4 Conclusion

Thus, our research questions were well answered through the user study. All subjects' common behavior reflects how an online consumer normally interacts with e-stores and what social info s/he does require during a complex decision process (i.e., looking for a high-value product to buy). Based on the study's observations, we suggest a three-stage recommender framework (as illustrated in Figure 1): 1) at the 1st stage when users browse alternatives, show popular recommendations that are derived from usage-driven social media. Personalized recommendation is also fit for this stage, which can be potentially more likely adopted by users if being integrated with expert opinions and users' personal constraints; 2) at the 2nd stage when users evaluate a product in detail, provide *{feature, opinion}* recommendations that are extracted from product reviews and place higher weights on negative opinions; 3) at the 3rd stage when users make comparison among candidates that were saved in their wish list, include social feature values in a comparison matrix to facilitate them confirming the final choice.

We believe that these implications can be very useful to related works, including ones that have attempted to fuse product popularity and reviews into recommenders [1,3]. In the future, we will be engaged in building the three-stage recommender system so as to be adaptive to users' various information needs. We will also conduct more user studies to test the system and consolidate the current study's findings.

References

1. Aciar, S., Zhang, D., Simoff, S., Debenham, J.: Recommender system based on consumer product reviews. In: Proc. WI-IAT 2006, pp. 719–723 (2006)
2. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering 17(6), 734–749 (2005)
3. Amatriain, X., Lathia, N., Pujol, J.M., Kwak, H., Oliver, N.: The wisdom of the few: a collaborative filtering approach based on expert opinions from the web. In: Proc. SIGIR 2009, pp. 532–539 (2009)
4. Chen, L., Pu, P.: Evaluating critiquing-based recommender agents. In: Proc. AAAI 2006, pp. 157–162 (2006)

5. Engel, J.F., Blackwell, R.D., Miniard, P.W.: Consumer Behavior. Dryden Press, Orlando (1990)
6. Groh, G., Ehmig, C.: Recommendations in taste related domains: collaborative filtering vs. social filtering. In: Proc. GROUP 2007, pp. 127–136 (2007)
7. Guy, I., Chen, L., Zhou, M.X.: Workshop on social recommender systems. In: Proc. IUI 2010, pp. 433–434 (2010)
8. Häubl, G., Trifts, V.: Consumer decision making in online shopping environments: the effects of interactive decision aids. Marketing Science 19(1), 4–21 (2000)
9. Kim, Y.A., Srivastava, J.: Impact of social influence in e-commerce decision making. In: Proc. ICEC 2007, 293-302 (2007)
10. Leino, J., Räihä, K.: Case Amazon: ratings and reviews as part of recommendations. In: Proc. RecSys 2007, pp. 137–140. ACM Press, New York (2007)
11. Mizerski, R.: An attribution explanation of the disproportionate influence of unfavorable information. Journal of Consumer Research 9, 301–310 (1982)
12. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval l 2(1-2), 1–135 (2008)