Eye-Tracking Study of User Behavior in Recommender Interfaces

Li Chen[1] and Pearl Pu[2]

[1] Department of Computer Science, Hong Kong Baptist University
Hong Kong, China
lichen@comp.hkbu.edu.hk
[2] Human Computer Interaction Group, Swiss Federal Institute of Technology in Lausanne (EPFL), Lausanne, Switzerland
pearl.pu@epfl.ch

Abstract. Recommender systems, as a type of Web personalized service to support users' online product searching, have been widely developed in recent years but with primary emphasis on algorithm accuracy. In this paper, we particularly investigate the efficacy of recommender interface designs in affecting users' decision making strategies through the observation of their eye movements and product selection behavior. One interface design is the standard list interface where all recommended items are listed one by one. Another two are layout variations of organization-based interface where recommendations are grouped into categories. The eye-tracking user evaluation shows that the organization interfaces, especially the one with a quadrant layout, can significantly attract users' attentions to more items, with the resulting benefit to enhance their objective decision quality.

Keywords: recommender systems, list interface, organization design, eye-tracking study, users' adaptive behavior.

1 Introduction

Although recommender systems have been widely developed in recent years to enable personalized decision supports (e.g., when users are searching for a movie, book, laptop, etc.), the focus has been mainly on the improvement of algorithm accuracy [1], less on studying the efficacy of interface designs from users' perspective. In fact, most of current recommender systems basically follow a ranked list structure, where all recommendations are listed one by one in the interface, according to the rank order of their predictive scores in matching users' interests as computed by the system.

Little is indeed known on the performance of such list interface and other possible display structures in influencing users' decision quality, although it has been claimed that users will likely adapt their decision strategies to the information presence on e-commerce sites, especially when they are confronted with a high-value product (e.g., computer, digital camera) for which the target product is not clear up front [7, 8]. For instance, Jedetski and Adelman have found that the amount of displayed alternatives will likely induce a significant effect [5]. When it is a small number (i.e., less than 30),

P. De Bra, A. Kobsa, and D. Chin (Eds.): UMAP 2010, LNCS 6075, pp. 375–380, 2010.

more compensatory strategies such as the weighted additive rule (WADD) [7] will be applied by users to produce more accurate decisions.

In this paper, we are particularly interested in exploring the effect of different recommendation displays on user behavior in the complex decision environment. Indeed, according to [3], users tend to focus on the top of a list, probably due to their cognitive limitations. Therefore, if such phenomenon also works when users face the list-based recommender interface, items that locate farther down in the list would attract little attention even though they may better match to the user's true interests. With this concern, we have attempted to understand: 1) how is users' actual visual searching pattern in the ranked list? And 2) are there more effective layout designs that can prompt users to consider more recommendations so as to potentially result in a more rigorous decision outcome?

We have accordingly designed a user study with the eye-tracker to answer the two questions. The experiment involved a comparison of the standard list view with category interfaces where recommendations are grouped into different categories and displayed in either vertical or quadrant layouts (see Fig. 1). The algorithm to generate the categories is called *preference-based organization method* that we have developed aiming to discover similar tradeoff properties among items (e.g., "these products are cheaper and lighter, but have slower processor speed") based on the association rule mining technique [2]. Prior simulation proved that this algorithm can obtain higher recommendation accuracy than related classification approaches due to its user-preferences focused clustering and selection strategies [2]. Thus, in this current work, we mainly aim at understanding whether and how the organization-based interface would in practice impact on end-users' cognitive searching process.

2 Experiment Setup

In our experiment, each user was asked to solve a decision problem (e.g., looking for a laptop to "buy"), for which the recommender was to assist them in locating interesting items and identifying the optimal choice. In the following, we will discuss in detail how the experiment was set up including materials used, participants recruited and experiment procedure, followed by results analysis.

Three recommender interfaces were prepared for this eye-tracking study. One is the standard list view where all recommendations are listed one by one, ranked by their satisfaction degrees according to user preferences (see Fig. 1.a "LIST"). More concretely, a set of 25 products (e.g., laptops), that have higher weighted utilities matched to the user's stated feature criteria as computed by the multi-attribute utility function [6], is returned as recommendations in this interface, and the highest ranked one is placed on the top, followed by others each with a "why" tool tip explaining the computational rational. The second one is the organization interface (see Fig. 1.b "ORG1"), where except for the ranked first item positioned as the top candidate, the remaining 24 products are organized into four categories. Each category is annotated with a title explaining how the attributes of products in that category provide benefits and compromises (i.e., tradeoff properties) in comparison with the top candidate (due to the space limit, please refer to [2] for detailed algorithm steps). In the third interface (see Fig. 1.c "ORG2"), instead of a vertical structure, the four categories are displayed in a quadrant arrangement with two categories laid out in parallel. The motivation for this new design actually came from [4], that indicates eye movements are likely to go to nearby objects. We were hence interested in knowing whether putting two categories at the same horizontal level would absorb attentions to more recommended products.

Each subject was randomly assigned one type of interface to evaluate. The main user task was to "*find a product that you would purchase if given the opportunity*" and the user was allowed to quit if s/he did not find any satisfactory product. A product catalog comprising 100 laptops extracted from a real e-commerce website was used for the generation of recommendations based on users' initially stated feature criteria (such as on the laptop's brand, price, processor speed, weight, etc.).

A Tobii 1750 eye-tracking monitor was used with a resolution setting of 1290x1024 pixels. It samples the position of the user's eyes by every 20ms. The monitor frame has a high resolution camera with near infra-read light-emitting diodes. This setting allows for more natural tracking of user behavior by not placing many restrictions on the participant. The ClearView software produces the log of eye-movements and user events, providing us with the screen coordinates for each area of interest (AOI).

Twenty-one participants (three females) volunteered to join in the study. They are students or employees in the university (ages ranging from 20 to 40) and from various countries (e.g., USA, China, Switzerland, Italy, Canada, India, etc.). More than half of them were interested in purchasing a laptop at the time of experiment, but no one was clearly certain of her/his targeted object before performing the experiment task.

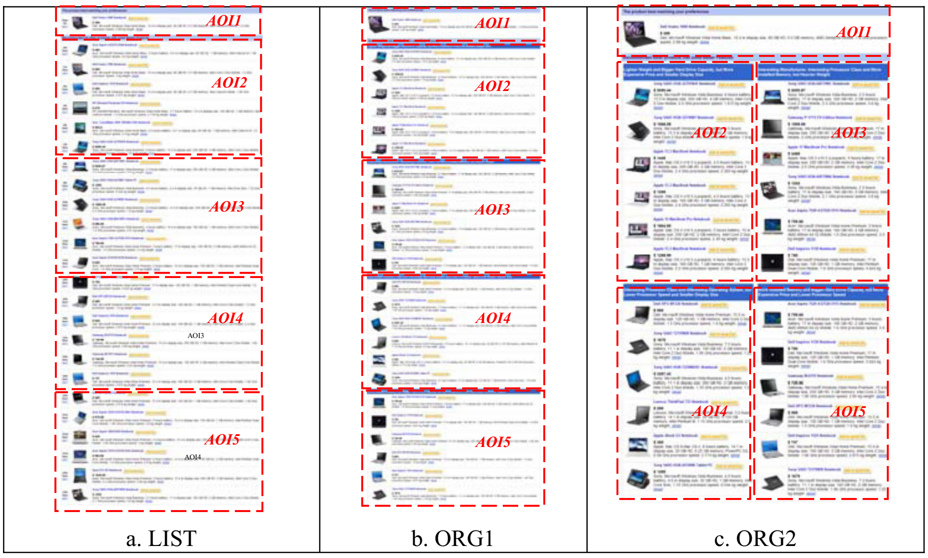

Fig. 1. Three recommender interface designs. Dashed boxes indicate Areas of Interest (AOIs).

3 Results

We mainly measured fixation frequency and duration. The fixation is gaze point with a minimum threshold of 100ms. Three participants were screened out from our analysis due to calibration difficulties or incomplete data with the eye-tracker, leaving us with eighteen participants for the analysis.

3.1 Areas of Interest

In total, 8389 gaze data points were recorded. Five major AOIs were defined on each recommender interface. In ORG1 and ORG2, each category (containing 6 products) represents an area of interest, in addition to the "top candidate" region (see Fig. 1). Accordingly, the list interface was also divided into five AOIs: the top candidate, 2nd to 7th recommended products, 8th to 13th ones, 14th to 19th ones, and 20th to 25th recommendations. The aim was hence to achieve the maximum comparability between the list interface and the organization interfaces.

The chance of looking at each AOI was first calculated. It shows that, although almost all studied users scanned over all AOIs on each interface, the focus was quite different as revealed by their fixation frequency and duration time (see Fig. 2.a). Specifically, in LIST, most of its average user's attentions were placed on AOI1 (the top candidate) and AOI2 (respectively 24.9s and 21.3s accumulated, covering 80.4% of the user's total duration time). It therefore indicates that users are likely to fixate on the top results in the rank list, though they were with the task goal of making a product choice among all displayed alternatives.

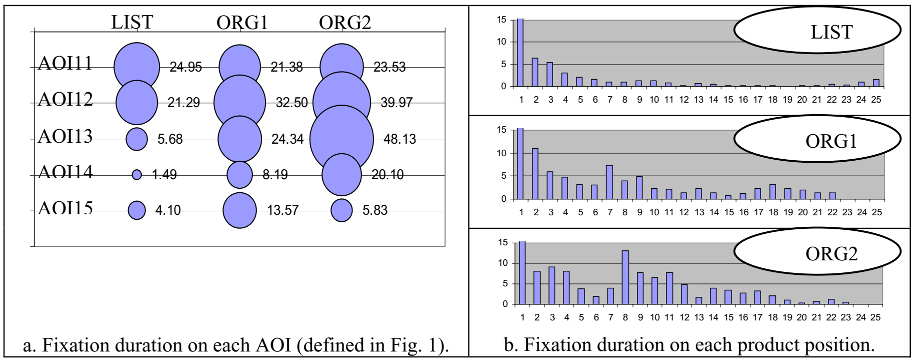

Fig. 2. Mean fixation duration (in *seconds*) on each AOI (left figure) and on individual product position (right figure) in the three interfaces

However, the fixation was dramatically changed in ORG1 and ORG2. In fact, more fixations were observed in these two interfaces in terms of both frequency and duration time (average 484.6 fixations with 108.4s in ORG1, and 595.6 with 149.4s in ORG2, vs. 336.5 with 71.7s in LIST). Comparing respective gazes on AOIs reveals that all of the four areas (from AOI2 to AOI5) received more attentions, relative to those in the LIST interface. It is especially of significant differences w.r.t. AOI3 and AOI4 (respectively $F = 5.14$, $p = 0.02$; $F = 3.84$, $p = 0.045$, by ANOVA test regarding duration). Post-hoc multiple comparisons further tell that the durations through AO2 to AOI4 were all averagely higher and more equally distributed in ORG2. Another interesting phenomenon is that users paid more attentions to AOI5 than AOI4 in ORG1 and LIST (both with the vertical layout), whereas AOI4 got more gazes than AOI5 in ORG2 (with the quadrant arrangement).

3.2 Viewed and Selected Products

We further analyzed the average fixation duration on individual product position in the three interfaces. It indicates that users in fact carefully looked at more products (that exceed 1s duration) in both organization interfaces (average 12.3 and 15.2 products in ORG1 and ORG2, against 7.7 in LIST). The difference between ORG2 and LIST is even significant ($p = 0.046$). More specifically, in ORG1, except for the top candidate (at the 1st position), the first two categories were carefully examined in terms of both titles and contained products (product positions from 2 to 9, see Fig. 2.b "ORG1"). The fixations of products in the other two categories were relatively less, but still exhibited a certain amount of interests. In ORG2, products in the first four AOIs were given more in-depth examinations in respect of their details (i.e., product positions from 1 to 17 through AOI1 to AOI4, see Fig. 2.b "ORG2"). On the contrary, in LIST, only the top five products were fixated but in a linear reduction manner, and users rarely placed attention to the other products below (see Fig. 2.b "LIST").

To further measure the objective decision quality achieved in each interface, we counted number of users who have finally made the product choice. It shows that respective 71.43% and 100% users have chosen products in the two organization interfaces, relative to 50% users in LIST (meaning that the other 50% quitted without selecting any item in LIST). Because participants were allowed to save several products into their shopping cart, we found that more products were on average selected for the basket in ORG1 and ORG2 (1.86 and 3.2 respectively, vs. 1.33 in LIST). Moreover, in ORG1 and ORG2, the selected products came from almost all AOIs, whereas it was either the top candidate or from AOI2 in LIST (see Table 1). It thus infers that when users are motivated to review more options (i.e., by the category interfaces), they will likely carefully consider what they see and choose more items if satisfactory, which unfortunately will be ignored in the list view by chance.

Table 1. Actual product selections and their sources (i.e. AOIs) in the three interfaces

	% of users who made product choices	Average selections	AOI1 (% of selected products from this AOI)	AOI2	AOI3	AOI4	AOI5
LIST	50%	1.33	25%	75%			
ORG1	71%	1.86	23%	31%	15%	8%	23%
ORG2	100%	3.2	12.5%	37.5%	37.5%	12.5%	

4 Conclusion and Future Work

In conclusion, the eye-tracking results interestingly show that users did practically adapt their searching behavior to different recommendation displays. In the ranked list, most of attentions were paid to the top, whereas in the organization-based interfaces users were attracted to view more recommended items. As a result, above 70% users have made product choices in ORG1 and ORG2, against 50% in LIST. It hence suggests that the category structure can more likely lead to a rigorous consideration process, enabling users to make informed decisions at the end. More notably, the quadrant category layout was demonstrated more competent in prompting users' fixations and augmenting their decision quality in the experiment.

The findings therefore point to a promising direction, motivating us to conduct more studies in the future. One objective will be to recruiting more users from diverse origins (e.g., females, professions) to consolidate the results. Another area is to in-depth investigating users' perceptual processes and discovering the reason that causes their behavior difference between the quadrant category layout and the vertical one. We will also target to build predictive models of users' cognitive architecture through continuous collection of their eye gaze patterns.

References

1. 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)
2. Chen, L., Pu, P.: Preference-based Organization Interfaces: Aiding User Critiques in Recommender Systems. In: Conati, C., McCoy, K., Paliouras, G. (eds.) UM 2007. LNCS (LNAI), vol. 4511, pp. 77–86. Springer, Heidelberg (2007)
3. Guan, Z., Cutrell, E.: An Eye Tracking Study of the Effect of Target Rank on Web Search. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 417–420. ACM Press, New York (2007)
4. Halverson, T., Hornof, A.J.: A Minimal Model for Predicting Visual Search in Human-Computer Interaction. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 431–434. ACM Press, New York (2007)
5. Jedetski, J., Adelman, L., Yeo, C.: How Web Site Decision Technology Affects Consumers. IEEE Internet Computing, 72–79 (2002)
6. Keeney, R.L., Raiffa, H.: Decisions with Multiple Objectives: Preferences and Value Trade-offs. Wiley, New York (1976)
7. Payne, J.W., Bettman, J.R., Johnson, E.J.: The Adaptive Decision Maker. Cambridge University Press, Cambridge (1993)
8. Pu, P., Chen, L., Kumar, P.: Evaluating Product Search and Recommender Systems for E-Commerce Environments. Electronic Commerce Research 8(1-2), 1–27 (2008)