Investigating the Effects of Two Types of Feedback in Recommendation Systems

Kris Jack

France Telecom R&D,
43 rue Pierre Marzin,
22 300 Lannion,
France
mrkrisjack@gmail.com

Liv Lefebvre

France Telecom R&D,
43 rue Pierre Marzin,
22 300 Lannion,
France
liv.lefebvre@gmail.com

RESUME

Cette recherche s'intéresse à la manière dont l'utilisation de feedback dans un système de recommandation augmente la satisfaction des utilisateurs. Cala a été testé au travers d'une étude dans laquelle les participants entraient leurs préférences cinématographiques au travers d'un système de recommandation. Il a été trouvé que le réordonnancment et la coloration de la liste d'items durant la phase d'expression des préférences peut augmenter la satisfaction des utilisateurs. Plus précisément, le feedback de reéordonnancement des items augmente le nombre d'items originaux qui ont été recommandés à l'utilisateur tandis que la coloration des items a un effet positif sur l'appréciation générale des recommandations.

MOTS CLES : feedback, interfaces, systèmes de recommandation.

ABSTRACT

This research investigates how improvements in a recommendation system's use of feedback can impact upon user satisfaction. A study is conducted in the field of cinematography in which participants enter their film-based preferences into a recommendation system. It is found that both reordering and colouring lists of items during preference elication can improve user satisfaction. In particular, giving feedback through reordering items increases the number of original items that are recommended to the user while colouring items has a positive effect on the user's general appreciation of such recommendations.

CATEGORIES AND SUBJECT DESCRIPTORS: H.5 INFORMATION INTERFACES AND PRESENTATION (e.g., HCI): H5.2 User Interfaces: *Ergonomics*; H.1 MODELS AND PRINCIPLES: H.1.2 User/Machine Systems: *Software psychology*.

GENERAL TERMS: Design; Experimentation.

KEYWORDS: Feedback; Interfaces; Recommendation Systems.

INTRODUCTION

Recommendation systems are typically designed to find items that would be liked by a given person. Such systems have been widely employed in the commercial sector, offering many types of items, such as books (e.g. Amazon) and films (e.g. Netflix), to customers (see [1] for a review). There is therefore considerable interest in producing usable interfaces that are both appreciated by users and allow them to find items of interest.

BACKGROUND

In order to make a good recommendation a system needs to have some information about the user. This information can then be exploited by the system with respect to the items that can be offered. For example, consider a user who would like a film recommendation. She informs the system that she loves Titanic. Using this information, the system applies its recommendation algorithm and suggests a number of films that this user should like, such as other romances and does not recommend action films. The user interface must therefore allow the user to enter their preferences and to receive recommendations.

Explicit preference entry interfaces tend to be used when gathering user data. They can take many forms from a flat list of items to 3D graphical visualisations [2]. The problem with these interfaces, however, is that the user tends not to receive any feedback during the preference entry process. As a result, they can often find themselves to be rather lost. Continuing the previous example, once the user has noted that she loves Titanic, she does not know if the system will be able to make good recommendations for her or not. As the system does not give her any feedback, she continues to score other films before explicitly asking for recommendations. Some systems are even designed to ask the user to rate several items, which is tedious, before receiving any form of system feedback is given (e.g. [4]). As a result, users can lose interest during the preference entry phase before they even receive their first set of recommendations. Keeping the user interested is thus of the utmost importance.

Deciding what type of feedback to provide the user is constrained by the workings of the system. Typical systems make use of a technique known as collaborative filtering [2]. Collaborative filtering essentially attempts to find correlations between users with respect to their stated appreciation of items, and recommends items to them that are liked by similar users. The information stored about the user is typically a set of scores for a number of items, where the number tends to be very small compared to the total number of known items. Recommendations cannot, therefore, be justified beyond typical "other customers have also liked these" explanations. The only type of output that the algorithm can produce is thus the predicted appreciation of items.

This research explores even such limited feedback can be more effectively exploited during the preference entry phrase. That is, the effect of the presence of different feedbacks will be measured in system usage.

SYSTEM EMPLOYED

A recommendation system has been designed and implemented that allows users to enter their preferences for films and then receive a list of recommendations. The system makes use of a collaborative filtering algorithm with weighted Pearson correlation similarities (see [7] for details). It contains a database of 17,770 films that have been rated by 480,189 users, using the MovieLens dataset.

An explicit preference entry interface was designed that shows a list of film titles (Figure 1). All films in the database are included in the list, in a random order. The user can express a monadic preference of like or dislike for a film by left clicking on its title. The first click expresses a like while the second click expresses a dislike. A third click cancels their declaration. Preferences are shown by colouring the film's title. Liked films appear in green while disliked films appear in red. In addition to showing user preferences, films are also coloured with respect to the user's predicted appreciation of the film (calculated by the recommendation engine). Four degrees of appreciation are shown in light pastel colours, signifying predicted love, predicted like, predicted dislike and predicted hate. When a user's explicit preference for a film has not be given, the predicted appreciation can be indicated. The user can also reorder the list, at any time, with respect their predicted appreciation of films, with the most liked films appearing higher in the list.

METHOD

Participants

16 volunteers (10 men and 6 women) participated in this study. They were 29 years old on average and they all had a good level of experience in computer science.
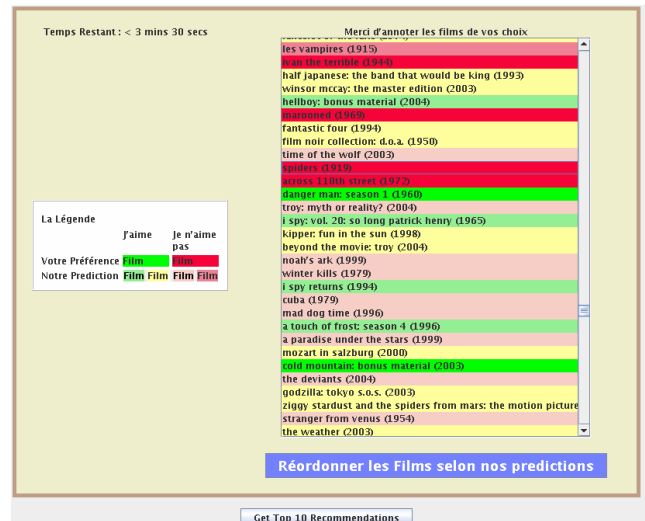

Figure 1: Preference entry interface with reordering option available and coloured list items

Design

They were asked to choose films that they liked and disliked using the system. Four conditions were proposed, which correspond to the four experimental conditions. We manipulated two within-subjects variables: the colouring of recommendations (with and without) and the possibility to reorder the list of films (with and without). These two independent variables were crossed and participants were randomly assigned to one of four counterbalanced groups (Table 1).

		Colouring	
		Without	**With**
Reordering	**Without**	R-C-	R-C+
	With	R+C-	R+C+

Table 1: The four experimental conditions

Procedure

We informed participants that they had 4 minutes to complete each session, but they can demand a list of recommendations when they want (with a button in the interface).

When the session's time was over, the system proposed a list of recommendations composed of 10 films. For each film recommended, participants were asked if they knew it or not and were asked to rate it between "I like" (5) and "I don't like" (1). We also provided the rating "I don't know". After that, they completed a short questionnaire on there impressions of the system. Finally, participants completed a final questionnaire where after an explanation of the conditions, they were asked "Which system did you most prefer?" and "Which system did you least prefer?"

Measurements

There were measurements taken of different variables in the two phases of the sessions: the phase of profile completion and phase of system recommendation. In profile completion phase, we measured the time spend, the number of preferences formulated, the percentage of liked preferences, and the number of preferences confirmed and contradicted with respect to the system's prediction.

In the system recommendation phase, we measured: the number of recommendations known and not known, the number of these that were scored, and the average score of recommendations. Concerning the subjective evaluation of recommendations, we measured:

- the average score for the list of recommendations
- percentage of agreement with question: "I am satisfied by the list of recommendations"
- percentage of agreement with question: "I want to reuse this system"
- percentage of agreement with question: "I easily found films in the list of recommendations."
- percentage of agreement with question: "It was fun to fill my profile"
- among the conditions, those which were most or least preferred.

Hypothesis

We expect that each type of feedback (reordering and colouring) proposed will have positive effects on the use of the system. We hypothesize that feedback influences profile completion. Simultaneously we think participants will be more satisfied by the recommendations and with the system in general when reordering and colouring are present.

RESULTS

Preferences entered

The type and number of preferences entered have been compared under the four conditions. In every condition, participants enter more likes than dislikes. There were no significant differences in their proportion under the different conditions. 14% more preferences were given with reordering. 19% less preferences were given with colouring. Participants also confirmed the system's predicted recommendations while entering their preferences more when reordering was available. Colouring did not have an effect upon the confirmations or contradictions made. In the questionnaire, participants noted that it was easier to find films that they knew when reordering was available. In every case, participants took the full 4 minutes to enter their preferences.

In the questionnaire, participants reported to prefer the condition when both colouring and reordering were present (Figure 2). There was a clear preference for the systems with colouring in general, whereas reordering alone was no more preferred than the condition that made use of neither reordering nor colouring.
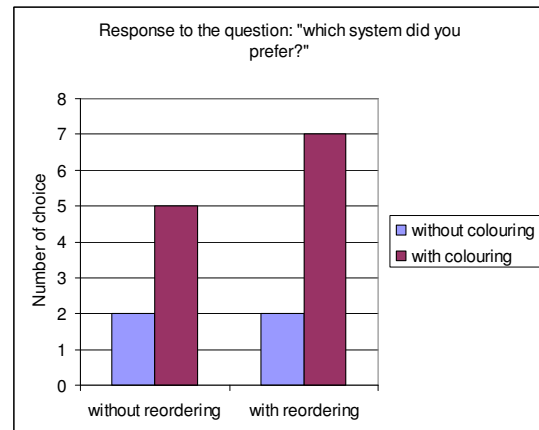

Figure 2: Participant's choice concerning the most preferred system

Recommendations

Participants were given 10 recommendations to score after entering their preferences at the end of each condition. The more films that the participant was familiar with in the list the higher that she tended to score them; correlations are as follows: R-C- (r(16)=.31; NS); R+C- (r(16)=.67; p < 0,05); R-C+ (r(16)=.56; p < 0,05); R+C+ (r(16)=.63; p < 0,05). Scores given for known films recommended were R-C- (M=4.1); R+C- (M=3.7); R-C+ (M=3.8); R+C+ (M=3.5). Scores given for unknown films recommended were R-C- (M=1.8); R+C- (M=1.7); R-C+ (M=2.1); R+C+ (M=2.1). When there was reordering, the participant received 15% more original recommendations. Colouring did not affect the originality of recommendations.

DISCUSSION

Preferences entered

The number of preferences given was higher under reordering lower with colouring. It is reasonable to assume that the reordering algorithm did a good job of ordering films by the participant's real preference, thus allowing them to navigate throughout the list with a real sense of order. The elevated number of prediction confirmations supports this. The questionnaire also revealed a subjective sense that it was easier to find films when reordering was available. This confirms results found in previous studies that organisation helps in making recommendations [6]. The colouring strategy, however, may take more cognitive effort to interpret, demanding more of the participant as they search for films. It is recommended that designers include a reordering function if they wish to maximise the number of preferences entered. In each case, however, participants used their full 4 minutes and even remarked that it was too short and that they would be happy to spend more time.

Subjectively, participants favour the use of at least one of the forms of proposed system feedback over none. They were amused by the presence of colouring and found it easy to understand. It is important to show the feedback of the system's activity [5]. The user makes an internal mental representation of the system's activity, which is, from the outset, based on superficial features such as the way in which the interface presents information. Colouring is a way of showing the system's predictions, and thus workings to an extent, that can be easily understood by users. Participants were also happy with the reordering feature. When the two features are combined in the same interface, however, their individual effects are not cumulative. Reordering alone was the preferred condition. Perhaps the participants' expectations rise when more functions are available but are not met by the results, giving a sense of dissatisfaction.

Recommendations

In general, participants score recommended films that they already know rather highly. That is, the system does a good job of finding films that the participant likes. While unknown films are not scored so highly, this does not mean that they are not appreciated, but instead reflects the different semantics behind giving scores for a known or unknown film. That is, when a film is known, the score is a mark of preference by experience while, when the film is unknown, the score is a desire to see the film. In comparing the conditions, known films receive the best scores when neither reordering nor colouring are present. This suggests that the presence of additional actions has a negative influence on the quality of known films that are recommended. The aim of recommendation systems, however, is not to recommend items that are already known by the user. On the contrary, such systems should introduce the user to new items that they are not yet familiar with. In testing, participants scored unknown film recommendations higher when colouring was present, suggesting that colouring guides participants into giving better preferences for use in recommendation systems. The presence of reordering also had a positive effect upon the quantity of original recommendations made. Condition R+C+ thus provide users with more and better original recommendations compared to the other conditions. Given that these are typical aims in recommendation system design, it is recommendable to include such feedback to improve the quality of recommendations.

CONCLUSION

It is the responsibility of the system designer to decide what kind of recommendations their system should make from the outset. Different problems can have different requirements. For example, one recommendation system may be designed to offer novel items to it's users that are very different from what they know while another may aim to recommend items that are similar to what the user is known to likes. With these requirements, the designer can then consider what kind of interface is most suited to their needs. These results are directly relevant for system designers, showing how system feedback can significantly effect the output of a system. In this case, the re-ordering and colouring of items in a list can have a real effect upon the quality and type of recommendations offered to the user. In particular, the inclusion of the option to reorder items leads to a significant increase in the number of original items that are recommended and the inclusion of colouring leads to a better overall liking of original recommendations.

REFERENCES

1. Adomavicius, G., & Tuzhilin, A. (2005). 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.

2. Candillier, L., Meyer, F., & Boullé, M. (2007). Comparing state-of-the-art collaborative filtering systems. In Perner, P. (Ed.), 5th International Conference on Machine Learning and Data Mining in Pattern Recognition (pp. 548–562), Leipzig, Germany. Springer Verlag.

3. Jack, K., & Duclayee, F. (2008). Improving Explicit Preference Entry by Visualising Data Similarities. In Intelligent User Interfaces, International Workshop on Recommendation and Collaboration (ReColl). Spain.

4. Miller, B., Albert, I., Lam, S., Konstan, J. and Riedl, J., MovieLens unplugged: experiences with an occasionally connected recommender system. in IUI '03: Proceedings of the 8th international conference on Intelligent user interfaces, (2003), ACM, 263-266.

5. Norman, D. (1986). User Centered System Design, Lawrence Erlbaum Associates.

6. Pu, P. and Chen, L., Trust building with explanation interfaces. in IUI '06: Proceedings of the 11th international conference on Intelligent user interfaces, (2006), ACM, 100.

7. Resnick et al., 1994) Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). Grouplens: An open architecture for collaborative filtering of netnews. In Conference on Computer Supported Cooperative Work (pp. 175–186). ACM.