Improving Recommendations by Increasing Recommendation Set Quality Variety

Mark Graus (m.p.graus@student.tue.nl)
Human-Technology Interaction
Eindhoven, University of Technology

With the rise of high-speed internet, the amount of media available increases. Recommender systems have been designed in order to reduce the effort needed to find relevant media. These systems deploy collaborative filtering algorithms in order to provide their users with good recommendations. This paper investigates to what extent presenting a user with good recommendations causes a choice overload and if this overload can be reduced by having a recommendation set with more variety in terms of quality.

Introduction

With the rise of broadband internet the amount of media available to any person at any time has become overwhelmingly large. Video-On-Demand, Youtube, and the increasing number of channels provided by Digital TV are all examples of sources of entertainment. Where in the past people were restricted by what TV channels were airing, now they are offered an almost unlimited choice. This freedom seems like a positive development at first, but new problems arise. Instead of deciding whether or not to watch what is available, people have to decide what to watch from the endless possibilities, which can cause what is known as choice overload.

Recommenders

In order to help people make choices from this seemingly unlimited choice set of items, efforts have been made towards creating effective recommender systems. The goal of a recommender system in an entertainment context is recommending items that are relevant to a user. Recommendations are calculated algorithmically based on explicit and implicit information. An example of explicit information is when a user is asked how much she likes a certain item, an example of implicit information is the set of items a user watched in the past.

Two types of recommender systems can be distinguished: content-based filtering and collaborative filtering. The first type bases predictions on objective properties of the items like actors, directors, movie duration and language in the case of movies. If someone likes movies with one particular actor, chances are that she will like the newest movie starring this actor too.

Collaborative filtering techniques calculate their predictions by combining ratings given by a large amount of users. The rationale behind collaborative filtering is that users who have a similar way of rating will have similar preferences. The recommender system recommends items that a user has not rated yet, but users similar to her have rated high.

Recommender metrics. It is common practice in recommender systems to use accuracy as an indicator of performance. Accuracy is the extent to which a system can correctly predict ratings. Maximizing accuracy and thus algorithm performance is equal to minimizing the difference between the predicted rating of user i of movie j ($\hat{r}_{i,j}$) and the actual rating ($r_{i,j}$). The accuracy is often expressed as Mean Absolute Error (MAE, see Equation 1) or Root Mean Squared Error (RMSE, see Equation 2). Even though this metric is one of the few available performance measures, it has been argued that a high accuracy alone does not guarantee a useful recommender system (McNee, Riedl, & Konstan, 2006a).

$$MAE = \sum \frac{|\hat{r}_{i,j} - r_{i,j}|}{N} \quad (1)$$

$$RMSE = \frac{\sqrt{\sum \hat{r}_{i,j} - r_{i,j}^{2}}}{N} \quad (2)$$

Decision Making

A good prediction is not necessarily a good recommendation. Situational factors like mood or the audience a movie will be watched with can influence what predictions are good recommendations. In order to increase the probability that one of the recommendations is actually good, recommender systems typically present sets of between 5 and 20 items.

A large item set however can cause what Iyengar and Lepper (2000) coined the Choice Overload effect. They tested whether a large set was more attractive than a smaller set by setting up a booth in a supermarket, displaying 6 types of jam in the one condition and 24 types of jam in the other condition. The proportion of people that walked up to the booth after seeing it was significantly higher in the condition with the large set, implying that this larger set was more attractive. Customers were given the possibility to taste and compare the jams and were handed a reduction coupon for buying one of the jams on display. The proportion of people that actually bought a jam was found to be significantly higher in the condition with fewer jams, indicating that people were more motivated to buy. In a subsequent questionnaire the customers in the smaller set condition expressed a higher satisfaction with the purchased item.

In an attempt to further investigate the choice overload effect, Scheibehenne, Greifeneder, and Todd (2009) demonstrated that set size is correlated with perceived variety of an item set. In their research they presented people with either a long or a short list of restaurants. People were then asked to browse the list and after that choose a gamble with 1/40 chance of winning either 30 euros cash or a coupon worth 40 euros refundable in one of the restaurants on the list. Three variables were measured, the perceived variety of the list, the choice difficulty people experienced and what gamble people opted for. People browsing the larger list perceived it to be more varied and had more difficulty choosing a restaurant from the list. There was however no difference in the proportion of people choosing the coupon over conditions.

Decision Making in Recommendation Sets

Size has been demonstrated to be an influence on the choice overload effect. However, neither study in the previous section considered the possible influence of item quality and variety. It is possible that adding items with similar quality to a list might cause choice overload, but adding lower quality items does not.

Ziegler, McNee, Konstan, and Lausen (2005) performed a study on the effect of variety of recommendation sets that demonstrates that set variety leads to higher satisfaction. In this study a taxonomy was used to calculate the intra-list similarity. This intra-list similarity is calculated following Equation 3. In this equation $c_o(b_k, b_e)$ is a measure of distance between item b_k and b_e, based on a separate taxonomy. $\Im P_{w_i}$ is the recommendation set for user i. Participants who got more diverse recommendations expressed having a higher satisfaction, showing that apart from size, subjective experience can also be altered by variety.

$$\mathrm{ILS}(P_{w_i}) = \sum_{b_k \in \Im P_{w_i}} \sum_{b_e \in \Im P_{w_i}, b_k \neq b_e} c_o(b_k, b_e) \qquad (3)$$

Recommender systems usually present the user the items with highest predicted ratings. A user will probably like all of these items to some extent, but some might be less useful because of the user's mood or because she has seen some movies before. In order to increase the probability of at least one good recommendation, this list is typically larger than 10, which can cause choice overload.

A way to reduce this choice overload might be to reduce the recommendation set size. However, reducing set size results in less transparency of the recommender system. People use the recommendations in a set to determine how reliable a recommender system is (Herlocker, Konstan, Terveen, & Riedl, 2004). A way to reduce choice overload, without reducing set size and thus transparency might be by taking items with a broader range of quality. This can be reached by presenting the top items, but replacing items lower in the list with lower quality items.

Because people should choose items with the highest quality, changing the recommendation set in the proposed fashion should not influence the choices people make, as both cases present the user with the best 5 items in the top of the list. In order to draw conclusions on subjective experience it is even a prerequisite that choice behavior is not influenced. Otherwise, any changes in subjective experience are caused by the ways in which the different lists influence the decision making process. This leads to the first hypothesis:

H1: Altering the recommendation set will not influence the choice people make, as long as the top items remain equal.

We also expect that choice difficulty is lower if a number of good items gets replaced with worse items. Bad items are less hard to discard from a consideration set than better items. This is in some sense also similar to what Scheibehenne et al. (2009) and Iyengar and Lepper (2000) demonstrate. Even though they do not explicitly mention item quality in their studies, we can regard the larger list as a list with more items with similar quality. Either replacing these items with options that are easier to discard or removing them altogether will reduce choice difficulty. The reduction of choice difficulty by removing these items is what has been demonstrated in the original studies.

H2: Choice difficulty decreases if a number of items are replaced with lower quality items.

Because the items are tuned towards the preferences of a user, similarity in the top items should be higher than similarity between top items and items lower in the list. Thus by increasing the range in quality, the perceived variety should increase.

H3: Perceived variety is higher when the range of item quality is broader.

Garbarino and Edell (1997) performed a study in which they demonstrated that choice difficulty can influence item satisfaction. An item set consisting of products of 4 different brands had to be evaluated. All items were described in terms of a number of attributes, but the effort required to compare the alternatives depended on the condition. In one condition attribute values were rounded numbers, in another condition attribute values were fractions. In the case of difficult item sets people expressed lower appraisal of all items. If **H2** holds, satisfaction should be higher in a item set with broad range. Similarly participants in the study by Iyengar and Lepper (2000) reported lower satisfaction in a larger item set, that has larger variety. Ziegler et al. (2005) demonstrated that variety increases satisfaction. If **H3** holds, satisfaction should thus be higher in a set where the proportion of suboptimal items is higher.

Combining these studies the last hypothesis is:

H4: A broader range of quality leads to higher satisfaction.

In order to test these hypotheses a manipulation has to be formed that has similar top items, but a broader range. Having the same top items is crucial for testing the first hypothesis, while the three other hypotheses rely on the broader

Table 1
Predicted Rank of Items for both conditions

Top-N	**1**	**2**	**3**	4	**5**	6	**7**	8	9	**10**	11	12	13	**14**	15	16	17	18	19	20
exp	**1**	**2**	**3**	**5**	**7**	**10**	**14**	21	31	45	66	97	142	208	304	446	653	955	1399	2048

range. The dependent variables to test these hypotheses will be user satisfaction, perceived variety and choice difficulty. To see to what extent the findings by Iyengar and Lepper (2000) can be reproduced, recommendation set attractiveness will also be included in the independent variables.

Experiment

In order to test the hypotheses an online experiment was designed. The experiment was based on a movie recommender system, using the 1M MovieLens dataset (available at http://www.grouplens.org). This dataset consists of 1 million ratings of 6040 users of 3886 movies. The dataset was enriched with metadata, mined from the internet movie database (http://www.imdb.com). The metadata consisted of plot synopses, cover images, cast and directors.

The algorithm used as core for the recommender system is matrix factorization. Matrix factorization works like a principal component analysis in the sense that it describes movies and users in terms of K latent features. The predicted value for the rating of user i on item j is the dot product of their feature vectors. Schafer, Frankowski, Herlocker, and Sen (2007) give an extensive overview of various collaborative filtering algorithms.

The combination of the matrix factorization algorithm and this particular dataset resulted in a model with a RMSE of 0.875 and a MAE of 0.685, which is highly accurate. An overview of different metrics is given by Herlocker et al. (2004).

Method

In order to test the hypotheses, it is crucial to pick two samplings that have the same top items in terms of ranks, but different items further along the list. To achieve this, we decided to calculate recommendations for every participant the same way, but alter the items presented. Participants in one condition received the 20 items with highest predicted value (Top-N condition) as recommendations, the other condition received items exponentially sampled from ranks $2^{n*0.55}$ for $1 \leq n \leq 20$ (Exp condition).

We expect choice overload to occur in the Top-N condition more than in the Exp condition. This because participants are likely to like most items in the Top-N condition, whereas the proportion of items that they like in the Exp condition should be considerably lower.

The ranks of items in both conditions are displayed in Table 1, with the ranks of the items that occur in both sets in **bold face**.

Participants

Participants were students and employees of the university campus. For every 10 participants two cinema tickets worth 15 euros in total were raffled.

After removing participants with missing data caused by a number of software problems, a total of 39 participants (mean age: 24, sd: 3.8; 30 male and 9 female) remained. Because gender nor age influenced any of the other variables, they will be left out of the analysis.

Procedure

The experiment was designed as a web application and consisted of four parts. First participants were presented with an introduction and questionnaire to establish basic demographics.

The next phase was used to get enough information to provide the participants with personalized recommendations. In this training phase participants were presented with sets of 8 randomly picked movies out of which they were asked to rate the movies they knew. A button enabled the user to renew the list of movies if the current screen had no more movies that the participant could rate, until at least 10 ratings were gathered. The number of times participants pressed the button was logged. This method corresponds to how ratings in the original dataset are entered, ensuring valid recommendations.

In the next phase the user received her recommendations. All 20 items were presented on one screen. All items consisted of the movie title and the predicted rating in stars. If a user hovered over a movie title, more information appeared onscreen. The information consisted of a movie cover, the title of the movie, the cast, the director(s) and plot synopsis (Figure 1). People were asked to chose the movie they expected to enjoy watching most. The time participants took making their decision was measured as well as the duration and frequency participants looked at each item.

Predicted ratings were all 5/5 or 4/5 stars in the Top-N condition, and ranged from 5/5 to 2/5 in the Exp condition. In order to prevent making the recommendation set in the Exp condition too unattractive, items with predicted ratings than lower 2 stars were avoided. Apart from the stars, ratings rounded down to one decimal place were displayed. This was done to make the difference between ratings more salient.

After making the choice, people were asked to fill in a questionnaire consisting out of 31 items (See Appendix A). 23 items were designed to be entered into a factor analysis, resulting in 4 factors corresponding to the independent variables of recommendation set attractiveness, satisfaction with the chosen item, perceived variety and choice difficulty. The other items were used to measure possible covariates.

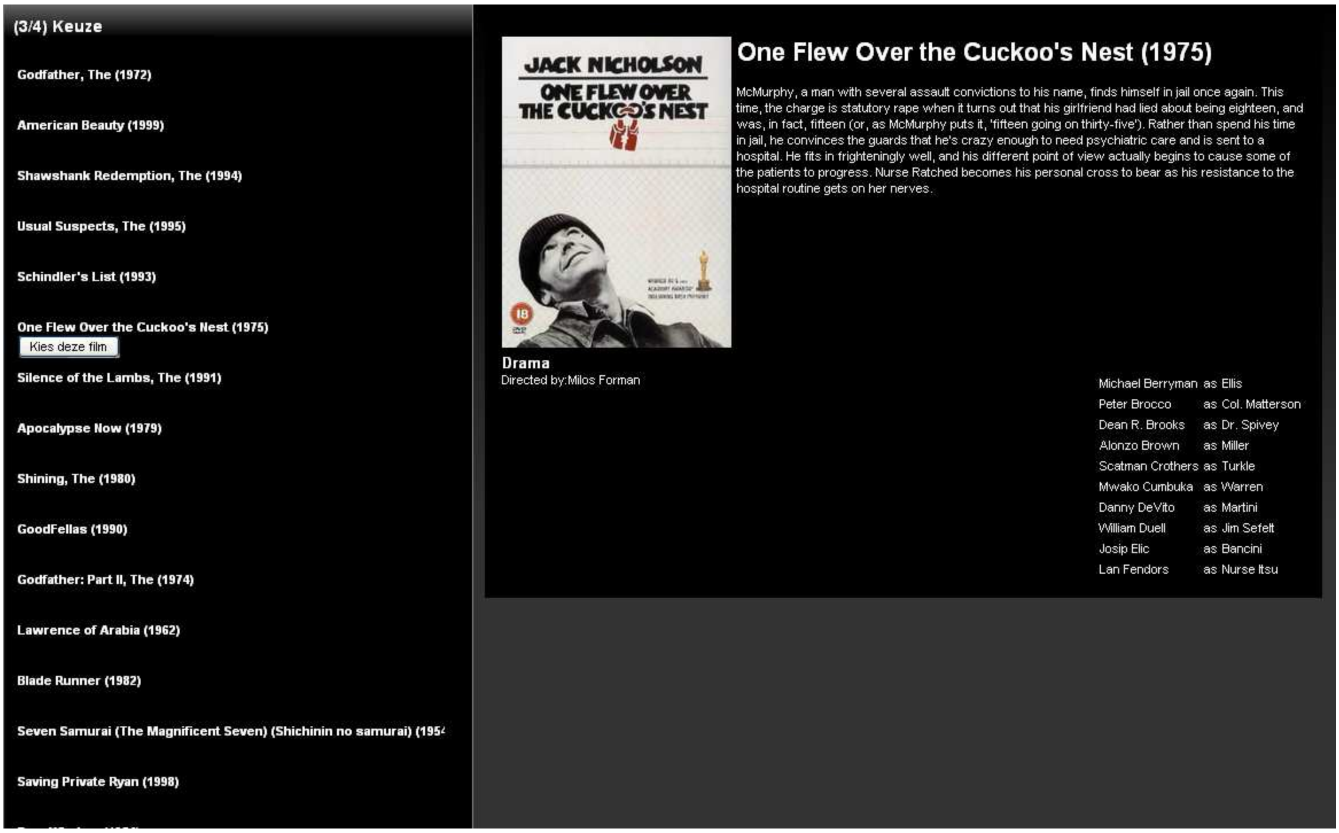

Figure 1. Recommendation Screen

Results

Chosen Item. The first variable to be investigated is the position of the chosen item. According to **H1** people should chose the item with similar rank in the entire set, regardless of position in the recommendation set. The data showed no significant difference ($t(38) = .370$, $p > .5$) between the average position people chose in the Exp condition ($M = 5.38$, $SD = 5.7$) and the Top-N condition ($M = 4.79$, $SD = 4.2$). Considering this p-value we cannot reject the first hypothesis.

Subjective Experience. In order to test the other hypotheses, 24 items in the questionnaire were entered into a factor analysis (see Appendix B). The factors to be extracted were satisfaction with the chose item (satisfaction), attractiveness of the recommendation set (attractiveness), choice difficulty (difficulty) and perceived variety (variety). The correlation between attractiveness and satisfaction caused these factors to collapse in a combined factor analysis. In order to prevent this, two separate factor analyses were performed.

Difficulty and variety were extracted in the first analysis. Because they appeared to be orthogonal, a Varimax rotation was used. The analysis resulted the 2 factors with 6 items loading on variety and 5 items on choice difficulty (see Table B1). Because of the correlation, the second analysis used Oblimin with $\delta = -.5$. It resulted in the two factors: satisfaction (5 items) and attractiveness (4 items) (see Table B2).

Of all of the questionnaire items, one item (9) originally formulated to load on variety ended up loading on attractiveness and 4 items (2, 4, 15 and 17) were removed from the analysis because they either loaded on two factors or had too low communalities. [1]

The main effect of the condition on the 4 factor means was tested in T-tests (see 2). Because the variables are extracted via a factor analysis, the scores over all participants have $M = 0$ and $SD = 1$. The marginally significant difference in choice difficulty supports the second hypothesis, as choice difficulty seems to decrease when the proportion of suboptimal items increases.

Surprisingly, the perceived variety in the Exp condition is not significantly higher than in the Top-N condition. This means there is no support for the third hypothesis. A possible reason is that the items in the questionnaire do not measure perceived variety correctly. Another possibility is the fact that the manipulation did not result in a difference of actual variety that the participants could perceive.

According to the fourth hypothesis, satisfaction should increase as variety increases. This however cannot be checked as the range of perceived variety is not large enough to perform a linear regression to test this hypothesis.

[1] The robustness of the factor analysis is affected negatively by the small sample size. However, the factor analysis was a confirmatory one and cross-validated in a twin research.

Table 2
Mean, SD and T-Test scores

Factor	Top-N Mean	Top-N SD	Exp Mean	Exp SD	T	significance
Variety	.094	1.06	-.110	.80	$t(37) = .668$	n.s.
Difficulty	.257	.87	-.220	.91	$t(37) = -1.669$	$p = 0.051$ (one-tailed)
Satisfaction	.384	.65	-.311	1.06	$t(36) = -2.366$	$p < 0.05$
Attractiveness	.343	.80	-.278	.97	$t(36) = -2.120$	$p < 0.05$

Process Data. Apart from what people report in the questionnaire, we can make inferences about their decision making by analyzing what and how long they look at each item. During the recommendation phase every time a person hovered over an item in the list, information appeared and this data was logged.

We consider it an acquisition when a user has information about a movie in view for at least 1 second. Shorter than that might indicate that the user did not really look at the information, but for example moved the mouse over the item in order to go from one item to another.

The total number of acquisitions does not significantly ($t(36) = .231$, $p > .5$) differ over means ($M_{exp} = 7.9$, $SD_{exp} = 4.95$ vs $M_{topn} = 7.5$, $SD_{topn} = 5.7$). The total time of acquisitions is marginally significantly higher ($t(36) = 1.973$, $p = .056$, $M_{exp} = 46.55$, $SD_{exp} = 29.68$ vs $M_{topn} = 30.02$, $SD_{topn} = 20.57$). However, standard deviations for this value are very high.

Another observation is the marginally significant difference of frequencies and total duration of acquisitions in the 3 top items over conditions. In the Exp condition people did look more often at the higher ranked items, whereas acquisitions were more uniformly distributed over all items in the Top-N condition (Figure 2).

However, because of the coarseness of the process data and the small sample size, interpreting conclusions from the process data has to be done with caution.

Domain Knowledge. In order to model domain knowledge, items 29 to 31 were summed. Domain knowledge did not differ significantly over conditions. Using a median split, the participants were divided over an experienced and a novice group.

Satisfaction and difficulty both differed marginally significant over groups. (Satisfaction: $t(37) = 1.968$, $p = .058$, $M_{expert} = .29$, $SD_{expert} = 1.01$, $M_{novice} = -.28$, $SD_{novice} = 0.81$, Difficulty: $t(38) = 1.888$, $p = 0.067$, $M_{expert} = .27$, $SD_{expert} = 1.06$, $M_{novice} = -.27$, $SD_{novice} = 0.67$). This implies that novice users are less satisfied than expert users. They also find it less difficult. Because of the low number of participants investigating the possible interaction between domain knowledge and the manipulations is not worthwhile to investigate.

Conclusion

Conclusion

There is support for the first and second hypothesis. Actual choice behavior is not influenced when a number of items is replaced with lower quality items, keeping the top equal. However, altering the choice set in this way does reduce choice difficulty.

Perceived variety did not appear to be influenced by the manipulation, which is unexpected. When considering the other hypotheses this should be taken into account. Also, the fourth hypothesis is not supported.

Another finding is that even though the choice difficulty is lower in a set with broader range, the set is found less attractive and satisfaction with an item from this set is lower.

Discussion

The main conclusion is that the manipulation did successfully reduce choice difficulty, but also reduced item set attractiveness and the satisfaction with the chosen item. However, a number of discrepancies between the original assumptions and the data exist, with the major discrepancy being the lack of significant difference in perceived variety over conditions.

One of the reasons might be domain knowledge having too much of an influence in this small sample. Having seen or not having seen a movie before influences the usefulness of recommending it and thus the satisfaction if this item is chosen. On the other hand, people with domain knowledge can more easily make judgments about the recommendation sets and chosen option.

Regarding the study of Iyengar and Lepper (2000), some similarities were found. In the less varied set, people reported more choice difficulty. Choice difficulty could explain the lower proportion of people purchasing a jam in the original experiment. The effect of a higher satisfaction was not established, but this might be due to the different way in which our study measures satisfaction. Furthermore, satisfaction could be influenced by the fact that users in our study were forced to make a choice.

The effect demonstrated by Scheibehenne et al. (2009) was not found. Variety did not lead to a higher satisfaction. Our study is however different in the sense that perceived variety was not influenced enough, nor did our study manipulate recommendation set size.

Satisfaction was also not positively influenced by reducing choice difficulty, like found in Garbarino and Edell

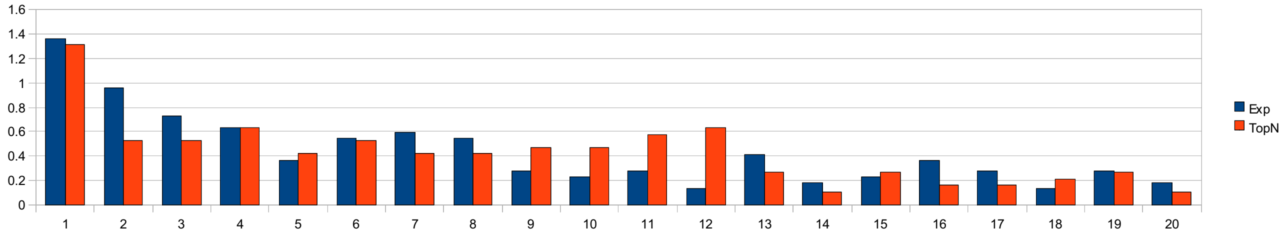

Figure 2. Average Number of Data Acquisitions per Position

(1997). A crucial difference between our study and Garbarino and Edell (1997) is that in our study people had to make a choice, instead of give a judgment about the alternatives. It is known that whether people choose or judge influences the appraisal of the items in a choice set (Billings & Scherer, 1988).

The relationship between variety and satisfaction as established by Ziegler et al. (2005) was not reproduced either. However, Ziegler et al. (2005) measures actual variety, whereas we measured variety perceived by the subjects. This leads to two possible causes. Either difference in variety was not perceived in Ziegler et al. or our manipulation did not cause variety, as introducing variety in quality might not necessarily guarantee a variety in items. In terms of dimensionality, quality is only one attribute of movies and movies with the same quality can have low similarity (for example good quality movies in different genres) and movies with different quality can have a high similarity (a good and a bad movie in the same genre).

Future Research

Recommender Efficiency. This experiment shows that subjective experience can be manipulated, without manipulating the algorithm. Measuring subjective experience when interacting with a recommender system is not common practice. This is argued by McNee, Riedl, and Konstan (2006b), who propose a framework to incorporate subjective user experience into recommender research. Their main point is that there are no metrics that express subjective perception of the result of recommender systems, like for example variety.

Where efforts in improving recommender systems have mainly been aimed at increasing algorithm accuracy, the data from this experiment shows that there is also room for improvement on a more psychological level. The emphasis on trying to predict item quality will have to shift towards trying to predict item relevance. This experiment showed that even though the recommender can successfully predict quality, people saw no change in variety when we sampled differently. This implies that items in the top of the list are not more similar than top and bottom items, something we would expect in a recommender system.

User Perception. The result of a recommender system is a recommendation set that consists out of a number of movies. However, we do not know how people perceive movies. The dimensionality is unknown, which makes it hard to generalize findings in decision making like the choice overload effect into a recommender system context.

In decision making experiments, choice sets have values on clear attributes. Being able to describe the items in a recommender system in terms of a set of interpretable attributes would enable new ways to navigate the recommendations. Users could directly express what attributes they find important, which could lead to an improved user experience. Aside from this, the people who implement recommender systems could diversify recommendation sets more objectively, which could lead to a reduction in choice difficulty and an increase in user satisfaction.

References

Billings, R. S., & Scherer, L. L. (1988). The effects of response mode and importance on decision-making strategies: Judgment versus choice. *Organizational Behavior and Human Decision Processes*, *41*(1), 1 - 19.

Garbarino, E. C., & Edell, J. A. (1997). Cognitive effort, affect, and choice. *Journal of Consumer Research*, *24*, 147-158.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, *22*(1), 5–53.

Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, *79*(6), 995–1006. Available from `http://dx.doi.org/10.1037/0022-3514.79.6.995`

McNee, S. M., Riedl, J., & Konstan, J. A. (2006a). Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *Chi '06: Chi '06 extended abstracts on human factors in computing systems* (pp. 1097–1101). New York, NY, USA: ACM.

McNee, S. M., Riedl, J., & Konstan, J. A. (2006b). Making recommendations better: an analytic model for human-recommender interaction. In G. M. Olson & R. Jeffries (Eds.), *Chi extended abstracts* (p. 1103-1108). ACM.

Schafer, J., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. In (pp. 291–324).

Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2009). What moderates the too-much-choice effect? *Psychology and Marketing*, *26*(3), 229–253. Available from `http://dx.doi.org/10.1002/mar.20271`

Ziegler, C.-N., McNee, S. M., Konstan, J. A., & Lausen, G. (2005). Improving recommendation lists through topic diversification. In *Www '05: Proceedings of the 14th international conference on world wide web* (pp. 22–32). New York, NY, USA: ACM.

Appendix A
Questionnaire Items

No.	Question Text	Range	Lower Anchor	Upper Anchor
1.	Hoe tevreden bent u met de film die u gekozen heeft?	-3/+3	Erg ontevreden	Erg tevreden
2.	Ik denk dat ik de beste film heb gekozen uit de mogelijkheden	-3/+3	Helemaal niet mee eens	Helemaal mee eens
3.	Het lijkt me leuk de door mij gekozen film te kijken	-3/+3	”	”
4.	Ik ken meerdere films die beter zijn dan degene die ik heb gekozen	-3/+3	”	”
5.	De film die ik gekozen heb zou een van mijn favorieten kunnen zijn	-3/+3	”	”
6.	Ik zou de film die ik gekozen aan anderen aanbevelen	-3/+3	”	”
7.	Ik zou liever een andere film huren dan degene die ik gekozen heb	-3/+3	”	”
8.	De lijst van aanbevelingen was gevarieerd	-3/+3	Helemaal niet mee eens	Helemaal mee eens
9.	De lijst van aanbevelingen had tenminste n film die ik leuk vind	-3/+3	”	”
10.	Er waren veel films in de lijst die verschillen van andere films in de lijst	-3/+3	”	”
11.	Alle aanbevelingen leken op elkaar	-3/+3	”	”
12.	De lijst van aanbevelingen bevatte films uit veel verschillende genres	-3/+3	”	”
13.	Geen twee films in de lijst van aanbevelingen leken op elkaar	-3/+3	”	”
14.	De lijst van aanbevelingen was	-3/+3	Erg beperkt	Erg breed
15.	De lijst van aanbevelingen was aantrekkelijk	-3/+3	Helemaal niet mee eens	Helemaal mee eens
16.	Ik vond geen enkele aanbeveling uit de lijst leuk	-3/+3	”	”
17.	De film die ik heb gekozen was de beste van de slechtsten	-3/+3	”	”
18.	De lijst van aanbevelingen kwam overeen met mijn voorkeuren	-3/+3	”	”
19.	Hoeveel aanbevelingen vond u leuk?	-3/+3	Geen	Allemaal
20.	Hoe makkelijk/moeilijk was het een keuze te maken?	-3/+3	Heel gemakkelijk	Erg moeilijk
21.	Hoe frustrerend vond u het keuzeproces?	-2/+2	Totaal niet	Heel erg
22.	Ik ben meerdere keren van gedachten veranderd voordat ik tot mijn keuze kwam	-3/+3	Helemaal niet mee eens	Helemaal mee eens
23.	De taak van het kiezen was overweldigend	-3/+3	”	”
24.	Uiteindelijk twijfelde ik tussen	4 items	1-2 films	Meer dan 10 films
25.	Ik heb het gevoel dat de beoordelingen invloed hebben gehad op de aanbevelingen	-3/+3	Helemaal niet mee eens	Helemaal mee eens
26.	Ik beoordeel liever films die ik goed vind dan films die ik slecht vind	-3/+3	Helemaal niet mee eens	Helemaal mee eens
27.	Van de lijst van aanbevelingen heb ik ... films al gezien	4 items	1-2 films	Meer dan 10 films
28.	Ik heb de film die ik gekozen heb al gezien	2 items	Ja	Nee
29.	Ik ben een filmliefhebber	-3/+3	Helemaal niet mee eens	Helemaal mee eens
No.	Question Text	Range	Lower Anchor	Upper Anchor
30.	Ten opzichte van anderen in mijn kennissenkring ben ik een expert op het gebied van films	-3/+3	Helemaal niet mee eens	Helemaal mee eens
31.	Ten opzichte van anderen in mijn kennissenkring kijk ik veel films	-3/+3	Helemaal niet mee eens	Helemaal mee eens

Appendix B
Factor Analyses

Table B1
Factor Analysis: Factors Variety and Choice Difficulty

	Factor		
	Variety	Difficulty	
8	,708		De lijst van aanbevelingen was gevarieerd
10	,777		Er waren veel films in de lijst die verschillen van andere films in de lijst
11	-,585		Alle aanbevelingen leken op elkaar
12	,793		De lijst van aanbevelingen bevatte films uit veel verschillende genres
13	,668	-,258	Geen twee films in de lijst van aanbevelingen leken op elkaar
14	,772		De lijst van aanbevelingen was
20	-,281	,660	Hoe makkelijk/moeilijk was het een keuze te maken?
21		,560	Ik heb het keuzeproces ervaren als
22		,777	Ik ben meerdere keren van gedachten veranderd voordat ik tot mijn keuze kwam
23		,534	De taak van het kiezen was overweldigend
24		,767	Uiteindelijk twijfelde ik tussen

Table B2
Factor Analysis: Satisfaction and Attractiveness

	Factor		
	Satisfaction	Attractiveness	
1	,8285		Hoe tevreden bent u met de film die u gekozen heeft?
3	,503	,335	Het lijkt me leuk de door mij gekozen film te kijken
5	,819		De film die ik gekozen heb zou een van mijn favorieten kunnen zijn
6	,776		Ik zou de film film die ik gekozen aan anderen aanbevelen
7	-,580		Ik zou liever een andere film huren dan degene die ik gekozen heb
9		,733	De lijst van aanbevelingen had tenminste n film die ik leuk vind
16		-,680	Ik vond geen enkele aanbeveling uit de lijst leuk
18		,795	De lijst van aanbevelingen kwam overeen met mijn voorkeuren
19		,784	Hoeveel aanbevelingen vond u leuk?