Using latent features diversification to reduce choice difficulty in recommendation lists

Martijn C. Willemsen
Eindhoven University of Technology
Human-Technology Interaction group
IPO 0.17, P.O. Box 513
5600 MB Eindhoven, Netherlands
M.C.Willemsen@tue.nl

Bart P. Knijnenburg
University of California, Irvine
Department of Informatics
DBH 5091
Irvine, CA 92697
bart.k@uci.edu

Mark P. Graus
Eindhoven University of Technology
Human-Technology Interaction group
IPO 0.20, P.O. Box 513
5600 MB Eindhoven, Netherlands
M.P.Graus@tue.nl

Linda C.M. Velter-Bremmers
Eindhoven University of Technology
Human-Technology Interaction group
L.C.M.Bremmers@student.tue.nl

Kai Fu
Eindhoven University of Technology
Human-Technology Interaction group
K.Fu@student.tue.nl

ABSTRACT

An important side effect of using recommender systems is a phenomenon called "choice overload"; the negative feeling incurred by the increased difficulty to choose from large sets of high quality recommendations. Choice overload has traditionally been related to the size of the item set, but recent work suggests that the diversity of the item set is an important moderator. Using the latent features of a matrix factorization algorithm, we were able to manipulate the diversity of the items, while controlling the overall attractiveness of the list of recommendations. In a user study, participants evaluated personalized item lists (varying in level of diversity) on perceived diversity and attractiveness, and their experienced choice difficulty and tradeoff difficulty. The results suggest that diversifying the recommendations might be an effective way to reduce choice overload, as perceived diversity and attractiveness increase with item set diversity, subsequently resulting in participants experiencing less tradeoff difficulty and choice difficulty.

Categories and Subject Descriptors

H.1.2 [Models and principles]: User/Machine Systems software psychology; H.4.2 [Information Systems Applications]: Types of Systems-decision support; H.5.2 [Information Interfaces and Presentation]: User Interfaces-evaluation / methodology, interaction styles, user centered design

General Terms

Algorithms, Experimentation, Human Factors, Theory

Keywords

Choice Overload, Diversification, User-centric evaluation

1. INTRODUCTION

Recommender systems support users in content discovery and exploration by providing items that match their personal preferences. Based on preference information (e.g. the user's ratings of known items, or past purchases) a recommender system predicts which items the user will like. The output of a typical recommender system is a ranked list of items with the highest predicted ratings. Provided that these predictions are accurate, the personalized recommendations are highly attractive. There is however a downside to attractive recommendations: psychological research on choice overload suggests that choosing an item from such a set might be a difficult task. Previous research [2] has shown that longer lists of attractive personalized recommendations can result in choice difficulty and subsequently in choice overload. In this paper we will investigate to what extend the difficulty of choosing from the list is related to the diversity of the items, while keeping the overall attractiveness of the set constant. Our results will show that users like item sets that are diversified and experience less choice difficulty in these sets, which suggests that diversification might be effective in reducing choice overload.

2. RELATED WORK

2.1 Choice overload

Choice overload [8, 16] refers to the difficulty decision makers experience when choosing from a large set of good alternatives. Choice overload has originally been related to the *size of the item set*, which creates two opposing effects. On one hand, larger sets are more attractive as they potentially provide more benefits for the decision maker, but at the same time larger sets have increased opportunity costs such as comparison costs, potential regret of not choosing the best option, and increased expectations which might not be met by the large set [16, 19].

As these opportunity costs tend to increase faster than the benefits associated with larger sets, decision makers usually find it easier to choose from a smaller set, and are often more satisfied with their actual choice when choosing from a smaller set. For example, in the original study by Iyengar and Lepper [8] visitors of a supermarket where more attracted towards a tasting booth that displayed 24 types of jam, rather than 6 types. However, only 3% of people who visited the booth with the large set of items bought jam, whereas 30% of the visitors of the small set booth bought jam (and reported to be more satisfied with their purchase). Other researchers have shown similar effects for other consumer products such as gift boxes [15] and coffee [13].

Choice overload effects can also occur in item lists generated by recommender systems. Bollen et al. [2] performed a user study with a matrix factorization movie recommender. They used three conditions: a small top-5 list, a large high quality top-20 list and a large lower quality 20 item list, composed of the top-5 plus 15 items with lower-ranked movies. Users experienced significantly more choice difficulty when presented with the high quality top-20 item list, compared to the other two lists. This increased difficulty counteracted the increased attractiveness of the larger set, showing that in the end, choice satisfaction in all three list conditions was about the same. In other words, although users found the long high quality list more attractive, they engaged in increased effort when evaluating its items, compared to the other two lists. Behavioral data corroborated these findings.

2.2 Factors underlying choice overload

Within the psychological literature there is a strong debate as to how omnipresent the choice overload phenomenon is. A recent meta-analysis by Scheibehenne et al. [16] across 50 studies shows that the overall effect size is zero, showing that in some cases longer items lists are detrimental and in some cases beneficial. Several preconditions are identified as potential causes for choice overload. A larger choice set might only result in decreased satisfaction and increased choice deferral when there are no strong prior preferences or dominant options in the item set [3, 13]. Indeed, Scheibehenne et al. show in their meta-analysis that expertise and prior preferences reduce choice overload, showing that it is most likely to occur for sets in which there is little variability in the attractiveness of the individual items and in which no specific items stand out (i.e., the items all fit the preferences of the decision maker equally well). In the psychological and marketing literature, the choice overload effect is mostly studied using item sets that are not personalized and therefore contain a wide variety of different items that do not necessarily lie in a person's field of interest. Recommender systems, on the other hand, provide lists that are optimized for the decision-maker. For such sets the preconditions for choice overload are easily met, as recommendation lists contain many, highly attractive items, none of which are clearly dominating. This suggests (in line with Bollen et al. [2]) that choice overload might be an important (and undesirable) side effect of personalized recommender systems.

Scheibehenne et al. [17] suggest that there are several moderators not included in their meta-analysis that might influence the level of choice overload and that require more research. We will focus predominantly on moderators related to the composition and perception of the item set, as these are moderators that can be investigated effectively by manipulating the output of a recommender system.

Two important moderators related to the composition of the item set are recognized by Scheibehenne et al. [17]: the categorization of the list and the diversity of the list. The first moderator suggests that if items are categorized, cognitive effort is reduced, resulting in less choice overload and higher satisfaction [13]. Categorization of recommendations in a recommender system has been shown to increase user satisfaction, effectiveness and confidence [7]. Though it has not yet been established if categorization can also reduce choice overload in recommender systems, we will focus on the second moderator of choice overload that has been part of ongoing research in both recommender systems and psychology: item set diversity.

2.3 The role of item set diversity

Scheibehenne et al. [16] indicate that item set diversity is another important moderator of choice overload. They argue that until now diversity has not been precisely controlled in studies of choice overload, and that the lack of control over this variable might be one reason why studies on choice overload show such volatile results. The benefit of using recommender algorithms is that these do allow us to control item attractiveness and item set diversity at the same time, offering a precise control of the composition of the item set. But to understand better the relation between diversity and choice difficulty, we will first discuss some of the existing literature that has investigated the differences in choice difficulty between uniform and diverse item sets.

Reutskaja et al. [15] found that the more uniform an item set gets, the more difficult it becomes to make a choice. Fasolo et al. [4] studied real world assortments and showed that as the number of items in a set increase, the differences between the options decrease. Specifically, differences between the relevant attribute values become smaller and the density of the set increases. Density is the distance, measured one product attribute at a time, between one product and its closest neighbor (e.g., the inter product distance is larger for the attribute „duration" when the movies in the list are 60, 120 and 180 minutes long, than for a list of movies that are 60, 80 and 100 minutes long). As density increases the differences between products decrease (i.e. the items in a high density set are more uniform, whereas items in a low density set are more diverse).

Fasolo et al. [4] argue that in uniform sets, with attribute levels close to one another, it is hard to decide which option is better. Users of typical recommender systems, where item sets contain many items that are highly similar, might therefore experience a large amount of *choice difficulty*, as they may find it difficult to justify one decision over another. People prefer to make decisions they can easily justify [22], especially when item sets become larger [21]. Therefore, people may prefer diversified item sets over uniform item sets, as items from diversified lists provide clear reasons to be chosen.

However, there is another side of diversified sets, which has received little discussion in the literature concerning choice overload. As options become more diverse, they might encompass difficult *tradeoffs* that generate conflicts that require a lot of effort to resolve, as one always needs to sacrifice something when choosing one item over another. Scholten and Sherman [18] propose a double mediation model, which shows a U-shaped relation between the size of the tradeoffs in a set and the amount of conflict that is generated by the set. They suggest that both very uniform and very diverse sets might be difficult to choose from, compared to sets of average diversity. Choosing from a uniform choice set, compared to a diversified choice set, is harder because one lacks compelling reasons to pick any option. However, the tradeoffs one has to make are smaller, which makes it easier to choose because no great sacrifices need to be incurred. Conversely, making a choice from a diverse choice set, compared to a uniform choice set, is easier because better arguments can be made for a certain option. However, the tradeoffs are larger, which makes it more difficult to make the decision, because greater sacrifices need to be incurred.

Given the fact that recommender systems are prone to induce choice overload and that researchers in psychology do not yet agree whether increased diversity leads to more or less difficulty, there is a need to study the effect of item set diversity on choice difficulty and tradeoff difficulty in greater detail. In the present study, we employ a diversification algorithm that allows us to tightly control the diversity of the recommended item set, thereby allowing us to investigate how different levels of diversification influence the user"s perception and experiences (in terms of tradeoff difficulty and choice difficulty) of the item set. For this purpose, we chose a recommender using a matrix factorization algorithm, as the latent features used by these algorithms provide ideal means to control diversity and tradeoffs.

2.4 Matrix factorization and tradeoffs

Matrix factorization algorithms try to express movies and users in terms of vectors on a set of latent features. Features are extracted in such a way that the relative positions of an item-user pair can be used to predict the rating for the user on that item as accurately as possible. The recommendations provided are those items with highest predicted rating for the user.

Koren, Bell and Volinsky [11] claim that the vectors in matrix factorization model "measure the extent to which the item possesses those [features], positive or negative", and that these features relate to real-world concepts, for example „Geared towards males/females‟ and „Escapist/Serious‟. This is similar to how choice sets are described in Multi-attribute utility theory (MAUT) used in the psychology of judgment and decision making [1]. MAUT describes choice options on a set of common dimensions (attributes; e.g., hard disk space, processor speed or battery life for a notebook), and assumes that one‟s preferences can be described by a series of weights that denote the relative importance of these attributes to a decision maker.

Given the similarity between attributes and the latent features in matrix factorization, the latent feature space could be used to vary the diversity of recommendation lists. The latent feature space allows us to select subsets of recommendations with a specific density (i.e. a specific distribution across the latent features), while keeping its overall attractiveness (in terms of predicted ratings) constant. Arguably, this diversity manipulation directly affects underlying psychological concepts responsible for choice difficulty and tradeoff difficulty and therefore is more effective in helping us understand choice overload than other diversity manipulations that are often based on external information. For example, Ziegler et al. [23] investigated that diversification by using a subset of the Top-50 recommendations and diversifying this subset in terms of Intra-List Similarity based on a separate, external ontology.

3. EXPERIMENT

3.1 Goal and hypotheses

The present study aims to investigate how diversification of a list of recommended items affects the perceived diversity and perceived attractiveness of the list, and how these factors subsequently affect tradeoff difficulty and choice difficulty. Using a diversification algorithm on the latent features of a matrix factorization algorithm, we vary the density of the list of recommendations while keeping the overall attractiveness of the list (in terms of the predicted ratings) constant. By using three levels of diversification we investigate whether the relation between diversity and difficulty is linear (higher diversity always simplifies choosing as it is easier to find reasons to choose one over another, the predominant view in the literature on choice overload) or U shaped (diversity only helps to a certain level, but a high diversity might result in large tradeoffs between items that are effortful to resolve and that might for some people result in too high sacrifices, as suggested by Scholten and Sherman [18]). To accurately measure the role of diversity, we employ a within-subject design in which each participant is presented (sequentially) with a low, medium and high diversity list. To prevent possible order effects, the order of these lists is randomized over participants.

Between subjects we also vary the number of items in the list on 5 levels (5, 10, 15, 20 or 25), as the literature suggests that diversification might have a stronger impact for larger lists. However, given that recommenders output personalized and highly attractive sets of items, we might find that even for short lists, diversification has a strong impact on experienced difficulty.

To measure the subjective perceptions and experiences of these recommendations after the presentation of each list, we employ the user-centric framework for user experience of recommender systems as described in Knijnenburg et al [10]. Based on this framework we expect that the effect of diversification of the recommender output on subjective experience with the list (how difficult is it to make tradeoffs and choose from the list) is mediated by subjective perceptions of the diversity and attractiveness of the list. In particular, we expect that item lists that are more diverse (i.e., have a lower density on the attributes) are perceived as more varied and potentially also as more attractive, and that these two factors affect the experience of tradeoff difficulty and choice difficulty.

3.2 Manipulating diversity in a matrix factorization model

Our study uses a 10-dimensional prediction model. The goal of our diversification algorithm is to generate three lists of movies that are about equally attractive, but that differ in how much they vary on the latent features of the algorithm. Our diversification is performed on the personalized top-200 of the ranked output of the recommender algorithm. In a prior study this number was established to allow for a large enough range in potential diversification while at the same time not differing too much in attractiveness: the difference between the highest and lowest predicted ratings in a typical top-200 list for this prediction model is 0.76 on a five-point scale, which is not much higher than the mean absolute error in the predictions of the prediction model we used.

For every participant three different sets of N movies were extracted from these top 200 recommendations. The low diversification set consisted of the N movies closest to the centroid of the top 200 recommendations (see Figure 1). For the high diversification, a greedy algorithm was used to select the set of N movies with the highest inter-item distances (using city block distance). The algorithm started with the movie closest to the centroid. The distance to each of the remaining movies was calculated and the one with the highest distance was added to the recommendation list. For the next steps, the distances between all items in the recommendation list and all remaining items were

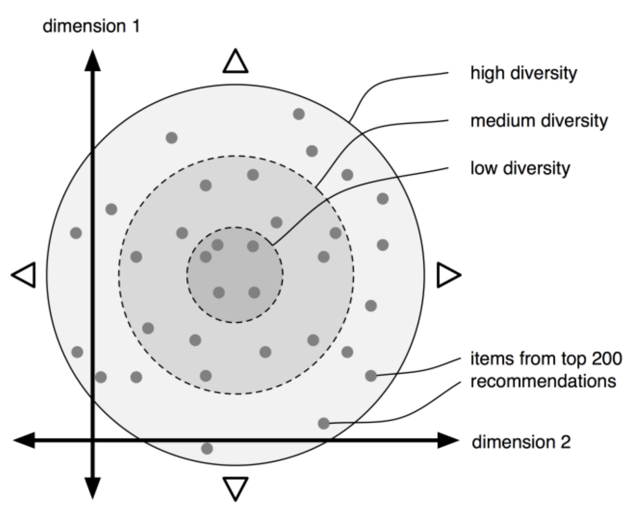


Figure 1: Schematic representation (simplified in two dimensions) of our item diversification in the top 200 set

calculated and the item with the maximum distance was added to the recommendation list, until the required number of recommendations was reached. Initial analyses on the used data set suggested that a boundary around 50% would result in movies with a medium density, so to derive a set with medium diversification level, the same greedy algorithm was used, but it was restricted to the 100 movies closest to the centroid instead of using the entire set.

3.3 System

For the study a movie recommender was developed based on a web-interface used previously in the MyMedia project[1], using a Matrix Factorization algorithm for the calculation of the recommendations. The dataset used for the experiment was the 10M MovieLens dataset[2]. In order to maximize the probability that users knew some of the movies presented in the initial rating task, movies from before 1994 and their corresponding ratings were removed from the dataset, resulting in a set of 5.6 million ratings by 69820 users on 5402 movies. We further enriched the MovieLens dataset with a short synopsis, cast, director and a thumbnail image of the movie cover taken from the Internet Movie Database. The Matrix Factorization algorithm used 10 latent features, a maximum iteration count of 100, a regularization constant of 0.0001 and a learning rate of 0.01. Using a 5-fold cross validation on the used dataset, this specific combination of data and algorithm resulted in an RMSE of 0.854 and an MAE of 0.656, which is up to standards. An overview of metrics is given by [5].

3.4 Design and procedure

The study consisted of three parts. In the first part, participants answered a set of questions to measure a number of individual characteristics. In their meta-analysis, Scheibehenne et al. [16] show that the characteristics expertise and prior preference are important moderators of choice overload. Therefore, we constructed a set of items to measure movie expertise and strength of preferences. We also measured maximizing tendency of our participants, using the short 6-item version [14] of the maximization questionnaire by Schwarz [20]. Schwarz defines people who always try to make the best possible choice as maximizers, and people who aim for "good enough" as satisficers. Maximizers consider more options whereas satisficers stop looking when they have found an item that meets their standards. Therefore the search costs of maximizers are higher and consequently it is suggested that they are more prone to choice overload.

After these questions, the second part of the study was used to gather rating information from the participant to be able to calculate and provide personalized recommendations. In this phase the participants were asked to rate a total of ten movies. They were presented with ten randomly selected movies at a time, with the instruction to rate only the movies they were familiar with (ratings were entered on a scale from 1 to 5 stars). After inspecting and (possibly) rating some of the ten movies shown, users could get a new list of movies by pressing a button. When the participant had entered ten or more ratings in total, they would be guided to the third part.

In the third part the participant sequentially received three times a list with recommendations, each time with a different level of diversification (the order of diversification levels was randomized). Depending on the condition, the participant was shown a rank-ordered list of between 5 and 25 movies (list length was manipulated between subjects) represented by a movie title. The predicted rating (in stars and one point decimal value) was shown next to the title. If the participant hovered over one of the titles, additional information appeared in a separate preview panel. This additional information consisted of the movie cover, the title of the movie, a synopsis, the name of the director(s) and part of the cast. Before moving to the next list, participants were presented with a short questionnaire of 16 items, measuring choice difficulty, tradeoff difficulty, perceived diversity and perceived attractiveness of the presented list. Participants thus answered these questions about each of the three lists.

3.5 Participants

Participants for this study were gathered using an online participant database. Participants were compensated with 3 euro (about 4 US dollars) for participating. 97 participants completed the study (mean age: 29.2 years, sd=10.3, 52 females and 45 males).

4. RESULTS

4.1 Measures

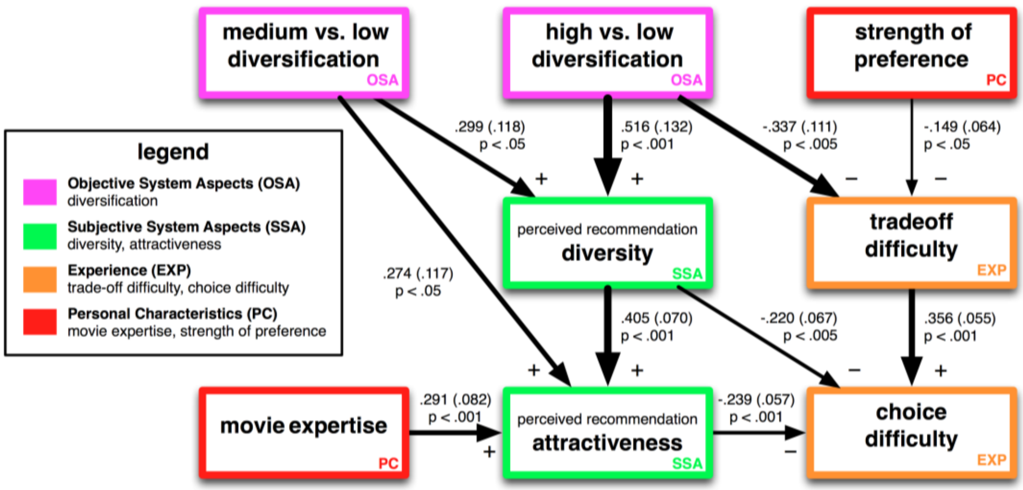
The items in the questionnaires were first submitted to an exploratory factor analysis (EFA) to determine whether their covariances naturally reproduced the predicted constructs. The EFA used repeated ordinal dependent variables, a weighted least squares estimator and Geomin rotation. After deleting items with low communalities or high cross-loadings, the analysis produced five correlated factors with a good model fit ($\chi^2(86) = 176.4$, $p < .001$, CFI = .988, TLI = .977, RMSEA = .060)[3], a good factor definition (lowest relevant loading: .576, highest irrelevant loading: .225) and a good discriminant validity (highest factor correlation: .396). These are the resulting five factors and their underlying items:

- Perceived recommendation diversity (5 items):
  - The list of movies was varied.
  - All the movies were similar to each other.
  - Most movies were from the same genre.
  - Many of the movies in the list differed from other movies in the list.
  - The movies differed a lot from each other on different aspects.
- Perceived recommendation attractiveness (5 items):
  - I would give the recommended movies a high rating.
  - The list of movies showed too many bad items.
  - The list of movies was attractive.
  - I didn't like any of the recommended items.
  - The list of recommendations matched my preferences.
- Strength of preference (3 items):
  - I have clearly defined preferences concerning movies.
  - I know what kind of movies I like.
  - Most of the time I let someone else pick a movie for me.

[1] See http://www.mymediaproject.org/

[2] The MovieLens dataset is available at http://grouplens.org.

[3] Based on extensive simulations, Hu and Bentler [6] propose cut-off values for these fit indices to be: CFI > .96, TLI > .95, and RMSEA < .05.



- Movie expertise (4 items):
  - I am a movie lover.
  - Compared to my peers I watch a lot of movies.
  - Compared to my peers I am an expert on movies.
  - I only know a few movies.
- Maximizing tendency (3 items):
  - No matter what I do, I have the highest standards for myself.
  - I never settle for the second best.

Two additional items were selected as indicators:

- Choice difficulty ("I would find it difficult to choose a movie from this list")
- Tradeoff difficulty ("I had to put a lot of effort into comparing the different aspects of the movies")

Nine additional items failed to contribute to a single factor, and were therefore deleted from the analysis.

4.2 Manipulation checks

We compared the resulting diversity, predicted ratings and variance of the predicted ratings in our data to check our diversification algorithm. As can be seen in Table 1, our diversification algorithm indeed increases the average range of the scores on the 10 matrix factorization features (calculated through Equation 1), a proxy of the level of attribute diversity.

$$AFSR = \sum_{1}^{10} \frac{\max(x_i) - \min(x_i)}{10} \quad (1)$$

Table 1: Diversity and predicted ratings of the presented items in our study

diversity	**Average feature score range (AFSR)** mean (SE)	**Predicted rating** mean (SE)	**SD predicted rating** mean (SE)
Low	0.959 (0.015)	4.486 (0.042)	0.163 (0.010)
Medium	1.273 (0.016)	4.486 (0.041)	0.184 (0.011)
High	1.744 (0.024)	4.527 (0.039)	0.206 (0.013)

At the same time the predicted ratings do not differ between the three levels of diversification showing that we manipulated diversity independent of (predicted) attractiveness. The standard deviation of the predicted ratings for the three set does increase slightly with increasing diversity.

4.3 SEM Model

The subjective constructs from the EFA were organized into a path model using a confirmatory structural equation modeling (SEM) approach with repeated ordinal dependent variables and a weighted least squares estimator. In the resulting model, the subjective constructs are structurally related to each other and to the conditions (list length and diversification level). The model was constructed based on the user-centric framework for user experience of recommender systems described in Knijnenburg et al. [10]. In the final model, the maximizer scale did not relate to any other variable, and was therefore removed from the analysis. The manipulation "list length" (whether participants were shown 5, 10, 15, 20 or 25 recommendations) also did not have a significant influence on the other variables. The results are therefore collapsed over these conditions. We also did not observe any effect of the order in which the three lists were presented.

The final model had a good model fit ($\chi^2(179) = 256.5$, $p < .001$, CFI = .989, TLI = .987, RMSEA = .041). Figure **2** displays the effects found in this model. Factor scores in the final model are standardized; the numbers on the arrows (A → B) denote the estimated mean difference in B, measured in standard deviations, between participants that differ one standard deviation in A. The number in parentheses denotes the standard error of this estimate, and the p-value below these two numbers denotes the statistical significance of the effect. As per convention, only effects with $p < .05$ are included in the model. The medium and high diversification conditions are compared to the low diversification baseline condition; numbers on the arrows originating in the conditions denote the mean differences between participants in medium or high diversification condition and participants in the low diversification condition.

To better understand the effects, we plotted the marginal means of the subjective constructs in the mid and high diversification condition relative to low diversification condition in Figure 3. Our diversification algorithm significantly affects the perceived diversity in linear fashion, with medium and high diversification

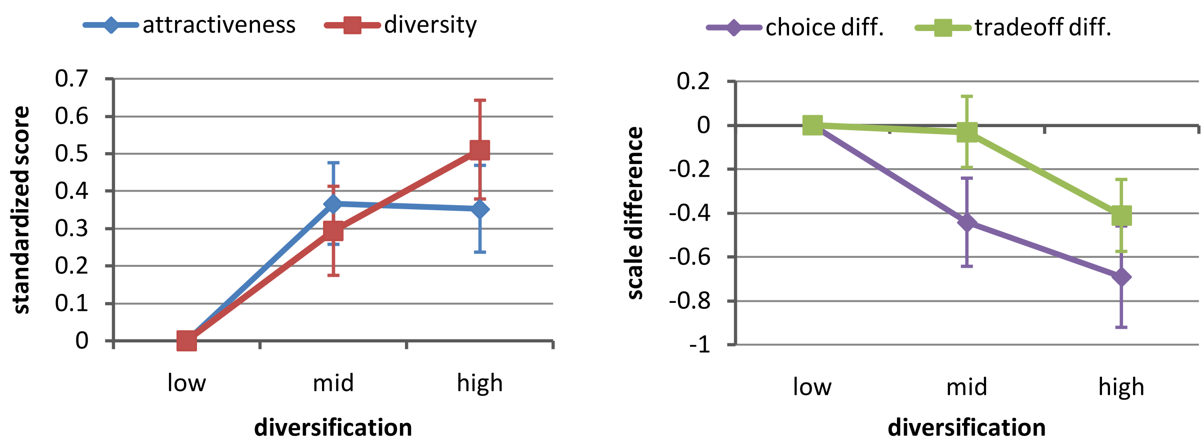


Figure 3: Left side: marginal means of perceived attractiveness diversity scores for mid and high diversification relative to low diversification: Right side: relative scale differences of mid and high diversification scores relative to the low diversification scores. Error bars are one std. err of the mean.

resulting in significantly higher perceived diversity than the low diversification condition. Higher perceived diversity subsequently increases the perceived attractiveness of the recommendations. The medium diversification condition also has a direct positive effect on attractiveness, making medium diversification as attractive as the high diversification (and both are significantly more attractive than low diversification, see Figure 3). There is also a direct effect of expertise, a personal characteristic, on attractiveness, showing that expert participants report higher attractiveness ratings.

In terms of tradeoff difficulty, we observe that this is significantly (and negatively) influenced by the high diversification condition, as well as a main effect of strength of preferences. So people experience less tradeoffs with very high diversification and if their self-reported strength of preference is higher, they also experience less tradeoff difficulty. The negative effect of diversification goes against the expectation that higher diversification leads to options that encompass larger tradeoffs between the attributes. Potentially the high diversification does not generate items that encompass apparent tradeoffs in the specific domain of movies.

All these constructs together influence the choice difficulty experienced by the user, which goes up with increased tradeoff difficulty, but goes down with increased diversity and attractiveness. The net result of our diversification manipulation on choice difficulty is negative: the higher the diversity of the set, the more attractive and diverse, and the less difficult it is to choose from the set (the marginal effects in Figure 3 suggest that choice difficulty decreases almost linearly with diversification level).

Our two experience constructs, tradeoff and choice difficulty, suggest that difficulty within this domain and for this recommender system is caused predominantly by lack of diversity in the item set, probably because the items are too similar to the user be able to make up her mind and find an option that is easy to justify. We do not find any evidence for the tradeoff difficulty related to conflicting preferences, which according to Scholten and Sherman [18] may occur when items become too diverse and tradeoffs become difficult to resolve.

5. Conclusion and Discussion

This paper shows how choice difficulty and tradeoff difficulty associated with a list of recommendations are influenced by the diversity of the list on the underlying latent features. By increasing the diversity of the items on these underlying dimensions, we increase the perceived diversity and attractiveness of the set, and subsequently reduce choice difficulty. Our net result thus is not a U-shaped relation between diversity and choice difficulty, but rather a simple linear downward trend.

We also do not observe an effect of item size on the perceived diversity or the experienced difficulty. Though intuitively, one would expect such an effect, we did not observe large variations across the different lengths. Given that our diversification algorithm finds items that are maximally spaced out from each other on the latent features within the set of equally attractive options, this might be not very surprising: when all items are good, diversification helps for both small and large sets. We would thus expect the effect of diversification to be roughly equal for different list lengths. Moreover, as the 5 different list lengths were manipulated between subjects, we have limited statistically power to detect such small differences.

We may have found a lack of an effect of item size on choice difficulty (which deviates from previous studies of choice overload) because we did not ask our participants to actually make a choice from the item sets. In the current study we explicitly did not investigate choice overload in the classical sense (showing that people defer choice or are less satisfied with their chosen option afterwards) but tried to investigate first whether diversifying on the underlying latent features of the recommender algorithm could influence the perceived diversity, attractiveness and difficulty. In a follow-up study, we plan to use this diversification to further investigate if manipulating the diversity (combined with item set size) will influence choice overload. In that study we will explicitly ask people to choose an item from the list of recommendations and also measure their satisfaction with the chosen item (as in Bollen et al. [2]).

Our results seem to corroborate the idea that latent features in matrix factorization algorithms have psychological meaning, as was suggested before [11]. Others [9, 12] already suggested that many of the findings from decision-making research could be very relevant for the development of recommender systems. Our data suggests that manipulating the underlying latent feature spaces could be one key to further understand and explore the role of important effects such as context and reference points in recommender systems.

6. ACKNOWLEDGMENTS

We would like to thank Niels Reijmer and Dirk Bollen for helpful discussions on the study.

7. REFERENCES

[1] Bettman, J.R., Luce, M.F. and Payne, J.W. 1998. Constructive Consumer Choice Processes. *Journal of Consumer Research*. 25, 3 (Dec. 1998), 187-217.

[2] Bollen, D., Knijnenburg, B.P., Willemsen, M.C. and Graus, M. 2010. Understanding choice overload in recommender systems. *Proceedings of the fourth ACM conference on Recommender systems - RecSys '10* (Barcelona, Spain, 2010), 63-70 DOI=http://doi.acm.org/10.1145/1864708.1864724

[3] Chernev, A. 2003. Product assortment and individual decision processes. *Journal of Personality and Social Psychology*. 85, 1 (2003), 151-162.

[4] Fasolo, B., Hertwig, R., Huber, M. and Ludwig, M. 2009. Size, entropy, and density: What is the difference that makes the difference between small and large real-world assortments? *Psychology and Marketing*. 26, 3 (Mar. 2009), 254-279.

[5] Herlocker, J.L., Konstan, J.A., Terveen, L.G. and Riedl, J.T. 2004. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.* 22, 1 (Jan. 2004), 5–53.

[6] Hu, L.-tze and Bentler, P. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*. 6, 1 (1999), 1-55.

[7] Hu, R. and Pu, P. 2011. Enhancing recommendation diversity with organization interfaces. *Proceedings of the 16th international conference on Intelligent user interfaces* (New York, NY, USA, 2011), 347–350.

[8] Iyengar, S.S. and 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 (2000), 995-1006.

[9] Jameson, A., Gabrielli, S., Kristensson, P.O., Reinecke, K., Cena, F., Gena, C. and Vernero, F. 2011. How can we support users‟ preferential choice? *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems* (Vancouver, BC, Canada, 2011), 409–418.

[10] Knijnenburg, B.P., Willemsen, M.C., Gartner, Z., Soncu, H. and Newell, C. Explaining the User Experience of Recommender Systems. *accepted to User Modeling and User-Adapted Interaction. http://t.co/cC5qPr9*.

[11] Koren, Y., Bell, R. and Volinsky, C. 2009. Matrix Factorization Techniques for Recommender Systems. *IEEE Computer*. 42, 8 (2009), 30-37.

[12] Mandl, M., Felfernig, A., Teppan, E. and Schubert, M. 2011. Consumer decision making in knowledge-based recommendation. *Journal of Intelligent Information Systems*. 37, 1 (Aug. 2011), 1-22.

[13] Mogilner, C., Rudnick, T. and Iyengar, S.S. 2008. The Mere Categorization Effect: How the Presence of Categories Increases Choosers‟ Perceptions of Assortment Variety and Outcome Satisfaction. *Journal of Consumer Research*. 35, 2 (Aug. 2008), 202-215.

[14] Nenkov, G.Y., Morrin, M., Ward, A., Schwartz, B. and Hulland, J. 2008. A short form of the Maximization Scale: Factor structure, reliability and validity studies. *Judgment and Decision Making Journal*. 3, 5 (Jun. 2008), 371-388.

[15] Reutskaja, E. and Hogarth, R.M. 2009. Satisfaction in choice as a function of the number of alternatives: When “goods satiate.” *Psychology and Marketing*. 26, 3 (Mar. 2009), 197-203.

[16] Scheibehenne, B., Greifeneder, R. and Todd, P.M. 2010. Can There Ever Be Too Many Options? A Meta- Analytic Review of Choice Overload. *Journal of Consumer Research*. 37, 3 (Oct. 2010), 409-425.

[17] Scheibehenne, B., Greifeneder, R. and Todd, P.M. 2009. What moderates the too-much-choice effect? *Psychology and Marketing*. 26, 3 (Mar. 2009), 229-253.

[18] Scholten, M. and Sherman, S. 2006. Tradeoffs and theory: The double-mediation model. *Journal of Experimental Psychology: General*. 135, 2 (May. 2006), 237-261.

[19] Schwartz, B. 2004. The tyranny of choice. *Scientific American*. 290, 4 (Apr. 2004), 70-75.

[20] Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K. and Lehman, D.R. 2002. Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality and Social Psychology*. 83, 5 (2002), 1178-1197.

[21] Sela, A., Berger, J. and Liu, W. 2009. Variety, Vice, and Virtue: How Assortment Size Influences Option Choice. *Journal of Consumer Research*. 35, 6 (Apr. 2009), 941-951.

[22] Shafir, E., Simonson, I. and Tversky, A. 1993. Reason-based choice. *Cognition*. 49, 1-2 (Nov. 1993), 11-36.

[23] Ziegler, C.-N., McNee, S.M., Konstan, J.A. and Lausen, G. 2005. Improving recommendation lists through topic diversification. *Proceedings of the 14th international conference on World Wide Web* (Chiba, Japan, 2005), 22–32.