Heterogeneity in Customization of Recommender Systems By Users with Homogenous Preferences

Jacob Solomon
University of Michigan
Ann Arbor, MI, USA
jacobbs@umich.edu

ABSTRACT

Recommender systems must find items that match the heterogeneous preferences of its users. Customizable recommenders allow users to directly manipulate the system's algorithm in order to help it match those preferences. However, customizing may demand a certain degree of skill and new users particularly may struggle to effectively customize the system. In user studies of two different systems, I show that there is considerable heterogeneity in the way that new users will try to customize a recommender, even within groups of users with similar underlying preferences. Furthermore, I show that this heterogeneity persists beyond the first few interactions with the recommender. System designs should consider this heterogeneity so that new users can both receive good recommendations in their early interactions as well as learn how to effectively customize the system for their preferences.

Author Keywords

Recommender Systems, Customization

ACM Classification Keywords

H.1.2. User/Machine Systems: Human Factors

INTRODUCTION

Recommender systems have the challenge of matching items in their catalog, such as movies or consumer products, to users who have heterogeneous preferences for those items. An item suitable for one user will likely be unsuitable for another. One way recommender systems can deal with this heterogeneity in user preferences is to give users a high degree of control over the recommender's algorithm, allowing them to work collaboratively with the system to find items that match their preferences. This approach to designing interactive systems is known as *customization* [1], and considerable recent HCI research has explored ways to build customizable recommenders [10, 2, 11, 14].

CHI'16, May 07 - 12, 2016, San Jose, CA, USA

ACM 978-1-4503-3362-7/16/05?15.00
DOI : *http* : //*dx.doi.org*/10.1145/2858036.2858513

While the goal of customization is to help the recommender accurately find items that match preferences that vary from user to user, it may simultaneously create a new form of heterogeneity among users regarding the way that they choose to configure the system. Users can arrive at a system bringing wide variability in their experience using intelligent systems, their mental models of how they work, their expectations about how the system will be helpful, and knowledge about the decision or items they are seeking recommendations about. Systems that use complex logic but lack transparency about how that logic works [9] may further complicate things for users in trying to figure out how to make the system give them what they want.

Consider two users of a movie recommender for whom the movie *Sleepless in Seattle* would in actuality be a very good recommendation (i.e. users with homogenous preferences). One user may have a mental model of the system that says the system heavily relies on the cast of a movie, and another may have a mental model or expectation that the genre is the major determinant of recommendations. One user would likely then configure the system to focus on movies with Tom Hanks, while the other would try to filter for Romantic Comedies. *Sleepless in Seattle* would likely appear in lists for both configurations, but the set of recommendations could be very different for each user and this could impact whether the user eventually chooses to watch *Sleepless in Seattle* or some other movie that would be less preferred.

This variability in users' characteristics related to interaction with a recommender creates added heterogeneity that must be accounted for by a recommender system. I present data from user studies of two different customizable recommender systems to show that even when different users have similar preferences – meaning the system should in theory provide them with similar recommendations – they are likely to configure the system in widely different ways. The accuracy of customizable recommenders may suffer due to noise in its user profiles that comes from the process of interacting with the recommender, and not just in the overall heterogeneity in users' preferences.

BACKGROUND

In order to provide good recommendations to users, recommender systems must elicit information about users' preferences. A common approach to this is collaborative filtering in which users give explicit ratings for items (such as star ratings for movies), and then suggest items that are highly rated among other users who have given similar ratings [17]. Content-based recommenders suggest items that have similar content or attributes to items that users have highly rated [3].

Recent work in recommender systems has explored different ways to afford users control over how the system produces its recommendations beyond providing ratings for items, such as controlling the influence of some portions of a social network on the recommendations [7, 19], enabling users to sort or filter items based on attributes [8, 10], allowing users to give weight to specific attributes of items [2, 13], or providing an interface to critique recommendations [5]. A large body of research has demonstrated that giving users control over recommender systems improves the user experience [12, 10, 2, 11, 6, 7, 4], although it may also create some decision-making biases [16].

One issue with customization is that it is generally more preferred among expert users than novices or users with little domain knowledge relating to the decision being made [18, 8]. This creates a difficult paradox for the system in that 1) new users are the ones that the system needs the most input from in order to be effective (i.e. the "cold start" problem [15]) and 2) new users may never become effective at customizing the system to accurately provide recommendations if they do not get practice at using the system. However, if a system does afford customization to new users, those users may be ill-equipped to effectively express their preferences in their early interactions, which could lead the system to create inaccurate profiles of these new users that hurt the quality of recommendations it provides. If users with similar preferences (i.e. users who should be receiving similar recommendations) provide widely varying input to the system, it would add considerable noise that could inhibit its overall effectiveness.

To study this issue of heterogeneity in configuration, I sidestep the nuisance factor of heterogeneity in preferences for items by *assigning* those preferences to users, so that the sample of users in these studies form a homogenous group of users who all have identical preferences for the items being recommended. This feature of the study design allows us to see how much variability in the behavior of new users can be attributed to their inexperience with the system and not to the natural underlying heterogeneity in their preferences. In other words, this design examines how much variation there is among groups of users that would form "neighborhoods" of users if preference information could be perfectly elicited and measured. I show that even within a group of homogenous users, there is considerable variation in how they will customize a recommender, and that this variation does not quickly reduce as these users become more experienced.

TRAVEL AGENT STUDY

I created an interface to a prototype recommender system called Travel Agent for recommending travel destinations that distinguished important attributes of a destination and allowed users to specify their preference along a continuous scale for these attributes by adjusting a slider (see Figure 1).

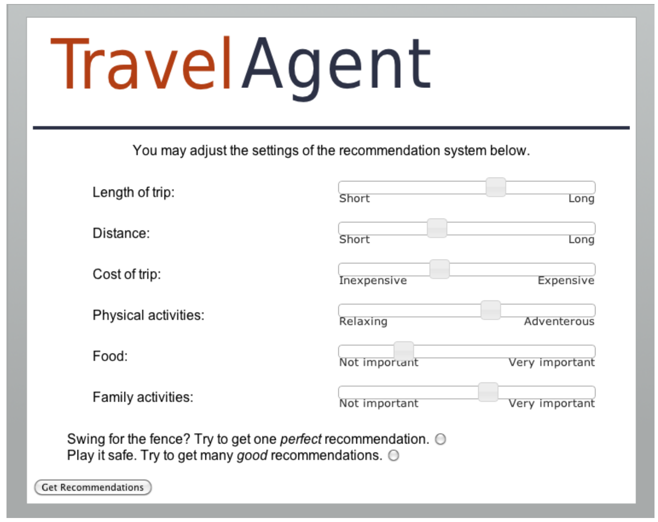

Figure 1. Travel Agent interface.

Rather than seeking a recommendation for themselves, subjects were trying to get recommendations for a fictional user persona. The persona was a 51 year old financial analyst from Chicago with 3 children looking to plan a vacation in October. The persona described details of his personality, hobbies, travel experience, and budget.

I recruited 375 subjects from Amazon Mechanical Turk to participate in this evaluation of Travel Agent. These subjects read instructions for the study and the details of the persona, and were required to pass a quiz on this information prior to using Travel Agent. After passing this quiz, they proceeded to use Travel Agent and generate recommendations, after which they answered survey questions about the recommendations.

Results

Despite the fact that all users were trying to produce recommendations for the same person, there was remarkable variability in the way users configured Travel Agent. Figure 2 shows the distribution of configuration for each option across the different users. A value of 100 corresponds with moving the slider for that option all the way to the right. For all options, the distribution spans the majority of the scale with relatively normal distributions, although the large spikes at the far-left setting are noteworthy. It should be noted that by default the sliders were pre-set to the mid-point of the scale. For each feature, users set the sliders at levels across the entire scale, although there were relatively normal distributions around a mean level for each feature.

To analyze this variance further, I conducted a k-means cluster analysis on the matrix of users by configuration options. Subjects were clustered together based on the similarity of all seven of their configuration choices. I determined that the within-cluster variance continually dropped until 20 clusters were formed, suggesting that there are about 20 distinct patterns of configuration among the 375 subjects. Figure 3 shows how each of these clusters configured the system. These clusters represent unique combinations of settings of the seven options.

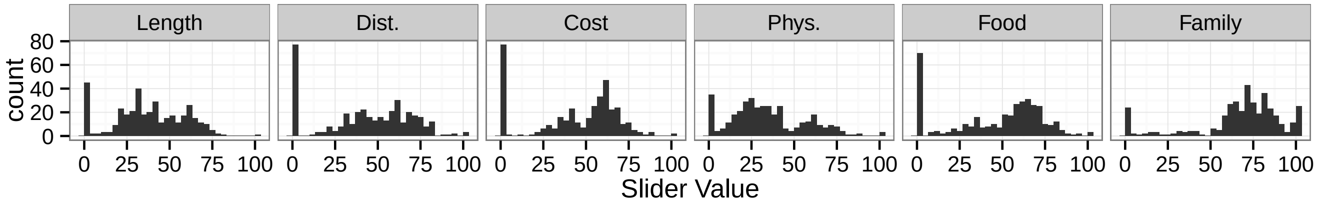

Figure 2. Distribution of configurations for each option in Travel Agent. A value of 100 corresponds to moving the slider all the way to the right.

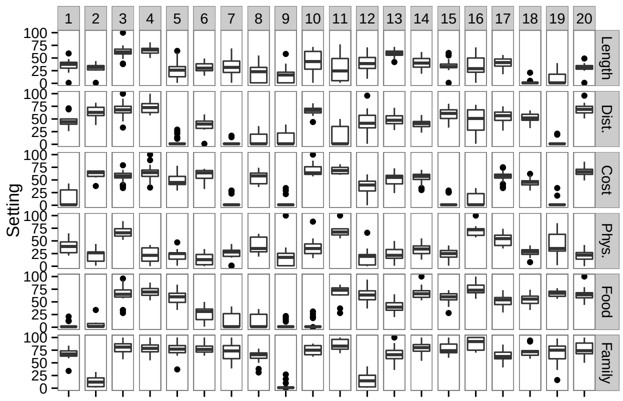

Figure 3. Variation in the way different clusters of subjects configured Travel Agent, even though all subjects were configuring for the same persona. Each column is a a single cluster.

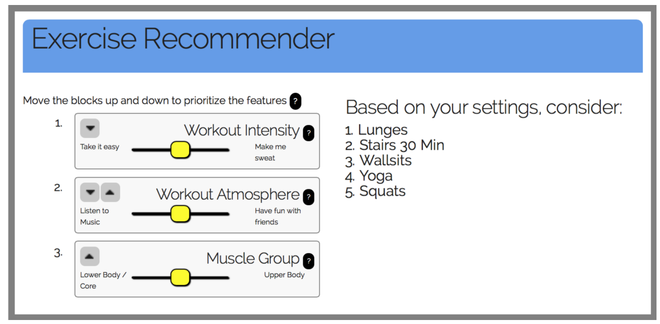

Figure 4. Exercise Recommender interface.

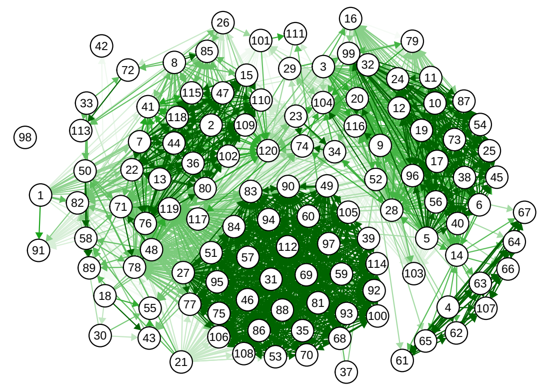

Figure 5. Similarity network of configurations for a single profile. Clusters represent groups who used similar configurations of the system.

EXERCISE RECOMMENDER STUDY

A limitation of the Travel Agent study is that the heterogeneity in the way that users customized the system may simply represent variability in the way people interpreted the persona's preferences. To address that limitation, I conducted a second study using a system called Exercise Recommender for recommending exercise activities (Figure 4) that specified preferences more specifically and gave greater incentive for users to customize the system for those preferences. I drew on experimental economics research to develop a decision task in which users are assigned preferences for attributes of an item by giving them a "payoff" for choosing an item (an exercise activity in this case) that has a given attribute. Subjects were shown five attributes of exercises they prefer (e.g. a cardio exercise, a group activity, a convenient activity etc.) and if their selected activity matched those attributes (as determined by an external panel of judges) they received an additional payment beyond their baseline compensation for participation in the study. This method has been shown to effectively *induce* preferences [20] in experiments by giving them an incentive to make decisions that fit their assigned preferences rather than their own personal preferences. By assigning concise preferences and incentivizing subjects to match those preferences, I was able to again replicate a group of homogenous users within a sample of recruited subjects.

113 subjects recruited from Amazon Mechanical Turk participated in the study. Subjects customized the system in two ways. First, they indicated through a 5-point slider their preference on three dimensions of an exercise (Workout Intensity, Social Recreation, and Muscle Group). Then, they could prioritize these dimensions by moving their input block up and down, such that the system would place greater emphasis on matching the dimension at the top of the list. The Exercise Recommender returned 5 recommended activities using a recommender formula suggested in [14]. After choosing an activity, they were shown their payoff for that choice and then redirected back to the Exercise Recommender to complete the task again using a new preference profile that gave different payoffs for different attributes. Each subject completed this task ten times so that their learning over repeated use of the system could be assessed.

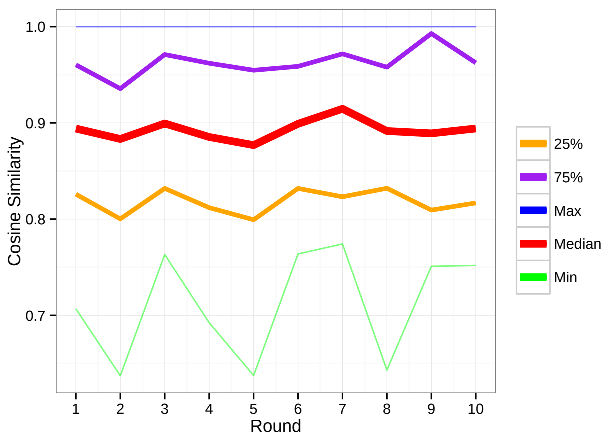

Figure 6. Similarity within cohorts using the same profile in the same round number. There is no trend for any aspect of the distributions, suggesting that subjects did not converge over time to more similar choices in customizing the Exercise Recommender.

Results

Figure 5 illustrates the variety of different ways that people tried to configure the Exercise Recommender for just one of the profiles. I calculated the similarity between each subject's configuration when using the system for a particular profile and every other subject when using that same profile as the cosine similarity between the six configuration values (Muscle Group setting, Muscle Group priority, Social Recreation setting, etc. ...) of each subject.

As in the Travel Agent study, there was again significant variability in the ways that people customized the Exercise Recommender even when trying to get recommendations to match an identical set of preferences. In the profile represnted in Figure 5, there is a single dominant cluster and three smaller clusters, as well as a non-trivial number of unclustered configurations. The dominant cluster for this profile in fact accounts for only 28% of the pool, leaving 72% of customization choices spread out among many different approaches. Across the 10 profiles, there were between two and four clear profiles for each cluster, with the largest cluster never accounting for more than 50% of users. This variability is further evidence that a customizable algorithm presents a difficult usability challenge to users who must figure out how to express their preferences and control the recommender, but may have widely varying mental models or intuitions about how to do that successfully.

I wanted to see whether this variability was stronger in the earlier rounds of the study than in later rounds to see whether users would begin to homogenize in their configurations after gaining some experience with the system and feedback about decisions. I divided the dataset into cohorts of subjects who used the same preference profile in the same round number. Since the preference profiles were randomly ordered for each subject, this resulted in 100 cohorts of 9 to 13 subjects. I calculated the full cosine similarity matrix within each cohort, and extracted the quartiles of each matrix. I found that over time, there was no trend of configurations becoming more or less similar to each other among people with the same preference profile. Figure 6 illustrates the pattern of change over time in the quartiles of cohorts' distributions, and regression analyses suggested that there was no meaningful trend towards either greater or lesser similarity over time.

DISCUSSION

Customizable recommender systems provide an alternative to more traditional ratings-based approaches such as collaborative filtering or content-based recommenders in eliciting preference information from users. Customizable recommenders provide an interface for users to interact more directly with the system's algorithm or recommender logic. While this can have an advantage of providing a positive user experience [4], it may not be an effective way to elicit reliable information about preferences, particularly from novice users of a system. Customizing an algorithm directly may be somewhat of a skill that requires training, experience, and knowledge to perfect. This heterogeneity may also be a significant source of noise for customizable systems, since users who want the same thing won't necessarily do the same thing within the system.

These findings suggest that customizable algorithms actually require less flexibility than what is apparent in user interfaces to those algorithms. Users should be given multiple paths to reach the same destination, meaning that there should be several different ways to configure a system within the UI that effectively result in the same recommendations. In the Exercise Recommender, an ideal design would have given about three different paths that would have led to similar recommendations, since there were typically about 3 distinct clusters of strategies for configuring the system. Additionally, these results suggest that customizable recommender systems need to explore feedback mechanisms to help users perceive what effect they have on a recommender algorithm. This will help users adapt their mental models and their configuration choices to better fit the algorithm.

A limitation of this study is that it merely quantifies the degree of heterogeneity that designers might expect, but does not provide specific information about the different mental models users have of an algorithm. As these mental models may be highly specific to particular systems or decision contexts, a critical part of a good user-centered design process will involve user research to determine all the specific mental models or customization strategies that users will take, and building affordances into the customization process that fit the varying mental models.

ACKNOWLEDGMENTS

Rick Wash, Emilee Rader, Joshua Introne, and Wei Peng, and members of the BITLab at Michigan State University provided valuable feedback on various stages of this work. This work was funded by the College of Communication Arts and Sciences at MSU.

REFERENCES

1. Jan Blom. 2000. Personalization: a taxonomy. In *CHI'00 extended abstracts on Human factors in computing systems*. ACM, 313–314. DOI: http://dx.doi.org/10.1145/633292.633483

2. Svetlin Bostandjiev, John O'Donovan, and Tobias Höllerer. 2012. Tasteweights: a visual interactive hybrid recommender system. In *Proceedings of the sixth ACM conference on Recommender systems*. ACM, 35–42. DOI: http://dx.doi.org/10.1145/2365952.2365964

3. Robin Burke. 2002. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction* 12, 4 (2002), 331–370. DOI: http://dx.doi.org/10.1023/A:1021240730564

4. Dina Burkolter, Benjamin Weyers, Annette Kluge, and Wolfram Luther. 2014. Customization of user interfaces to reduce errors and enhance user acceptance. *Applied ergonomics* 45, 2 (2014), 346–353. DOI: http://dx.doi.org/10.1016/j.apergo.2013.04.017

5. Li Chen and Pearl Pu. 2012. Critiquing-based recommenders: survey and emerging trends. *User Modeling and User-Adapted Interaction* 22, 1-2 (2012), 125–150. DOI: http://dx.doi.org/10.1007/s11257-011-9108-6

6. Yoshinori Hijikata, Yuki Kai, and Shogo Nishida. 2012. The relation between user intervention and user satisfaction for information recommendation. In *SAC '12 Proceedings of the 27th Annual ACM Symposium on Applied Computing*. 2002–2007. DOI: http://dx.doi.org/10.1145/2231936.2232109

7. Bart P Knijnenburg, Svetlin Bostandjiev, John O'Donovan, and Alfred Kobsa. 2012. Inspectability and control in social recommenders. In *Proceedings of the sixth ACM conference on Recommender systems*. ACM, 43–50. DOI: http://dx.doi.org/10.1145/2365952.2365966

8. Bart P Knijnenburg, Niels JM Reijmer, and Martijn C Willemsen. 2011. Each to his own: how different users call for different interaction methods in recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*. ACM, 141–148. DOI: http://dx.doi.org/10.1145/2043932.2043960

9. Todd Kulesza, Simone Stumpf, Margaret Burnett, and Irwin Kwan. 2012. Tell me more?: The effects of mental model soundness on personalizing an intelligent agent. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1–10. DOI: http://dx.doi.org/10.1145/2207676.2207678

10. Benedikt Loepp, Katja Herrmanny, and Jürgen Ziegler. 2015. Blended Recommending: Integrating Interactive Information Filtering and Algorithmic Recommender Techniques. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 975–984. DOI: http://dx.doi.org/10.1145/2702123.2702496

11. Denis Parra. 2013. *User controllability in a hybrid recommender system*. Ph.D. Dissertation. University of Pittsburgh.

12. Denis Parra and Peter Brusilovsky. 2015. User-controllable personalization: A case study with SetFusion. *International Journal of Human-Computer Studies* 78 (2015), 43–67. DOI: http://dx.doi.org/10.1016/j.ijhcs.2015.01.007

13. J Ben Schafer, Joseph A Konstan, and John Riedl. 2002. Meta-recommendation systems: user-controlled integration of diverse recommendations. In *Proceedings of the eleventh international conference on Information and knowledge management*. ACM, 43–51. DOI: http://dx.doi.org/10.1145/584792.584803

14. J Ben Schafer, Joseph A Konstan, and John Riedl. 2004. View through MetaLens: usage patterns for a meta-recommendation system. *IEE Proceedings-Software* 151, 6 (2004), 267–279. DOI: http://dx.doi.org/10.1049/ip-sen:20041166

15. Andrew I Schein, Alexandrin Popescul, Lyle H Ungar, and David M Pennock. 2002. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 253–260. DOI: http://dx.doi.org/10.1145/564376.564421

16. Jacob Solomon. 2014. Customization bias in decision support systems. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 3065–3074. DOI: http://dx.doi.org/10.1145/2556288.2557211

17. Xiaoyuan Su and Taghi M Khoshgoftaar. 2009. A survey of collaborative filtering techniques. *Advances in artificial intelligence* 2009 (2009), 4. DOI: http://dx.doi.org/10.1155/2009/421425

18. S Shyam Sundar and Sampada S Marathe. 2010. Personalization versus customization: The importance of agency, privacy, and power usage. *Human Communication Research* 36, 3 (2010), 298–322. DOI: http://dx.doi.org/10.1111/j.1468-2958.2010.01377.x

19. Katrien Verbert, Denis Parra, Peter Brusilovsky, and Erik Duval. 2013. Visualizing recommendations to support exploration, transparency and controllability. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. ACM, 351–362. DOI: http://dx.doi.org/10.1145/2449396.2449442

20. Vernon L. Smith. 1976. Experimental Economics: Induced Value Theory. *The American Economic Review* 66, 2 (May 1976), 274–279. http://www.jstor.org/stable/1817233