3-1-2018

Diversity-Enhanced Recommendation Interface and Evaluation

Chun-Hua Tsai

Diversity-Enhanced Recommendation Interface and Evaluation

Chun-Hua Tsai
University of Pittsburgh
Pittsburgh, PA 15260
cht77@pitt.edu

ABSTRACT

The beyond accuracy user experience of using recommender system is drawing more and more attention. For example, the system interface has been shown to associate positively with overall levels of user satisfaction. However, little is known about how the interfaces can constitute the user experience and the social interactions. In this paper, I plan to propose a visual diversity-enhanced interface that supports the user to inspect and control the multi-relevance recommendations. The goal is to let the users explore the different relevance prospects of recommended items in parallel and to stress their diversity. Two preliminary user studies with real-life tasks were conducted to compare the visual interface to a standard ranked list interface. The users' subjective evaluations show significant improvement in many metrics. I further show that the users explored a diverse set of recommended items while experiencing an increase in overall user satisfaction. A user-centered evaluation was used to reveal the mediating effects between the subjective and objective conceptual components. The future plans are discussed to extend the current findings.

KEYWORDS

Recommender System; Diversity; Beyond Relevance; User control

ACM Reference format:
Chun-Hua Tsai. 2018. Diversity-Enhanced Recommendation Interface and Evaluation. In *Proceedings of CHIIR'18, March 11-15, 2018, New Brunswick, NJ, USA.*, , 3 pages.
DOI: http://dx.doi.org/10.1145/3176349.3176357

1 MOTIVATION

The mainstream research on recommender system (RS) focuses on improving recommendation accuracy which determines the quality of a predictive model. The model is usually trained with user-generated data (e.g., bookmarking and rating) for optimizing a recommendation list in descending order. An accurate predictive model can filter the recommendations by relevance to the user. This approach has been proven useful for further eliciting users' interests or preference which reduces the effort of decision making and choice difficulties [20]. However, since the users interact with the complete RS interface instead of the relevance scores, the predictive model only partially constitutes the **user experience** [9]. The finding implies that high accuracy of the predictive model may not always be equivalent to user's perceived recommendation quality [9, 11]. There are more factors, e.g., control, transparency, trust, that need user-centric approaches to RS evaluation.

One challenge of the accuracy-oriented RS is to generate a "one fit for all" recommendation list for various user needs, which is not realistic in many real-world scenarios. For instance, a hybrid recommender system that fuses several recommendation sources can be diversified. Each source, which is considered as one-dimension relevance, may be preferred for different needs. In [4], the study showed the social-based similarity works best for finding known friends while content-based similarity could be used to find unknown people with similar interests. We can add some variation to the recommendation list, but it may be accompanied by the risk of lowering the user satisfaction due to the exposure of the beyond-expectation items [21]. To solve this issue, several authors argued to offer **user controllability** to fuse the multi-relevance features by choosing various algorithms [5] or data sources [1]. Providing a visual interface could make the fusing process more transparent; for example, showing recommender sources and their overlaps as set diagrams [12, 17] can further address this problem. However, it is not clear how the user interacts with the control functions and interfaces in the exploration tasks with multi-relevance. The effects on user experience, in this case, is also unknown.

The other challenge of a relevance-driven RS is to deliver a narrow set of recommendations to the user. All the recommended item are highly similar to the user's profile, which is a well-known over-specification (lack of diversity) problem that leads to a poor user experience [9]. Critics argued the personalized algorithm causes filter bubble effect, which shields the user from other viewpoints [2], facilitates an adverse effect in social fragmentation and creates an ideological polarization of discussions on social issues [10]. In response to those issues, the factor of **beyond relevance** are attracting more and more attention in the recent RS studies. The diversity-enhanced recommendations can be generated based on users' personality for broader exposure [19] and balance choice between novelty & similarity items [18]. Many studies further leveraged visual interfaces to enhance recommendation diversity [16], e.g., adopting a visual discovery interface [14] or organization interfaces [6] in a RS can increase the selection diversity. However, little research efforts have been focused on understanding how the user interactions can moderate the user experience.

My work is to propose a diversity-enhanced interface which supports the user perception of the multi-dimension social recommendation. An user-centric evaluation framework would be conducted using Structural Equation Modeling (SEM) method [9]. This framework aims to explain the user experience in multiple systems and user aspects. The primary goal of my proposal can be seen in three-fold: First, this proposal seeks to uncover the moderators which explaining the user experience while interacting the

DOI: http://dx.doi.org/10.1145/3176349.3176357

interface. It can help ground out the casualty between the interface design and user experience outcome. Second, this proposal aims to explain the mediation between the proposed interface and the recommendation diversity. It can help to explain the effects of adopting the interface on multi-relevance tasks. Third, this proposal aims to explain the social implications of applying an RS. The framework helps to investigate the issues of causing the filter bubble effects and how to react with a proper interface design.

2 RESEARCH QUESTIONS

My work focus on answering questions regarding user-centered evaluation of social recommender system. **RQ1:** How do objective system aspects (OSAs) affect the user perception (SSA) with a social recommender system? **RQ2:** How do OSAs affect the user experience (EXP) with a social recommender system? **RQ3:** How do OSAs affect the user interaction (INT) with a social recommender system? **RQ4:** How do personal and situational characteristics (PCs and SCs) affect the SSAs, EXPs, and INTs with a social recommender system? **RQ5:** How do the mediation effects help to explain the SSAs, EXPs, and INTs with a social recommender system?

3 PLANNED METHODOLOGY

To assess the value of the proposed interface, I plan to conduct user experiments to compare the proposed interface with differently controlled manipulation. The study is scheduled in a within-subject design; all participants were asked to use the interface for three designed tasks and to fill out a (pre) post-stage questionnaire at the (beginning) end of each manipulation. At the end of the study, participants were asked to compare interfaces regarding their explicit preference and situational awareness. The order of manipulation was randomized to control for the effect of ordering. The interface is embedded in the Conference Navigator System (CN3), a social support system for academic conferences [3]. To minimize the learning effect (becoming familiar with data), we used conference data from multiple years.

I plan to extend the User-Centric Evaluation Framework for Recommender systems from [8] on explaining the user experience. The framework represents as five interrelated conceptual components. **1) Objective System Aspects (OSAs):** the aspects of the proposed system that are currently being evaluated. In this proposal, the OSA represents the diversity-enhanced interface with different manipulations; **2) Subjective System Aspects (SSA):** the effects that mediating between EXPs, INTs and the OSAs. SSAs help to establish connections through user perception of certain system aspects; **3) User Experience (EXP):** the EXP is the subjective evaluation from the users. It helps to understand the user feedback on different system aspects. **4) Interaction (INT):** the INT is representing the logged data the user interaction with the system; **5) Personal and Situational Characteristics (PCs and SCs):** the PCs and SCs are used to test the influence of the user's characteristics and the situation on using the system. The factors are beyond system aspects but with a significant impact on EXPs.

4 PROGRESS MADE

Two attempts have been made to test the design of diversity-enhanced interface [15]. First, we proposed a recommender interface that explores the value of a two-dimensional scatter plot visualization to present recommendations with several dimensions of relevance (shown in Figure 1). In our context, the scatter plot interface was used to help users combine different aspects of relevance for recommended items as well as providing inspectability to the users. Second, we proposed a recommender interface that enhances the fusion control function within a ranked list with meaningful visual encoding for multiple dimensions of relevance (shown in Figure 2). The users can adjust the relevance weightings to customize the recommendation results, which provides the user with a higher level of control over their results.

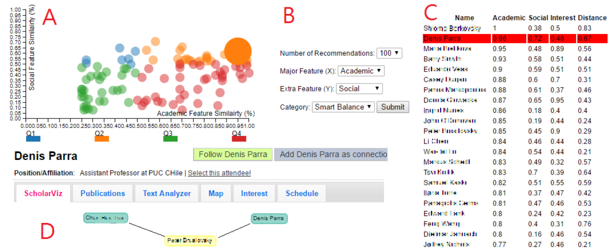

Figure 1: The design of Scatter Viz: (A) Scatter Plot; (B) Control Panel; (C) Ranked List; (D) User Profile Page. The user can select (or inspect) the recommendations with two relevance dimensions in the scatter plot.

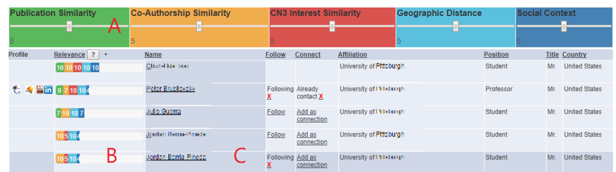

Figure 2: The design of the Relevance Tuner: (a) Relevance Slides; (B) Stackable Score Bar; (C) User Profiles. The user can inspect the recommendations with multi-relevance dimensions while controlling the weightings.

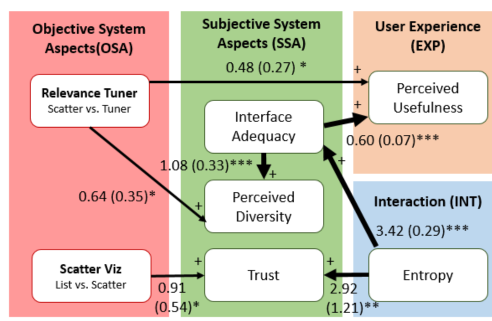

Figure 3: The structural equation model of the experiments. The number (thickness) on the arrows represents the β coefficients and standard error of the effect. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The model fit the statistics of $\chi^2(96) = 234.68$, $p < 0.01$, $RMSEA = 0.18$, $90\% CI : [0.152, 0.211]$, $CFI = 0.941$, $TLI = 0.922$.

The two interfaces were designed to explore the value of user-controllable and diversity-aware interfaces in a social recommender system. Each of the interfaces has been evaluated in a controlled field study in the target context (assigned exploration tasks in a conference), with 25 and 20 subjective respectively. The results show that the new visual interfaces reduce exploration efforts for a set of realistic tasks, and also make the users more aware of the diversity of recommended items. Also, the users' subjective evaluation shows a significant improvement in subjective metrics, i.e., perceiving useful and satisfaction. The experiments further showed the effects of the proposed interfaces for the users' interaction. We measure the user's selection diversity using information entropy [15]. The experiment results supported the two proposed interface can facilitate the user with higher selection diversity.

To better understand the mediation effects across the two interfaces, we conducted a structural equation model (SEM) analysis [7] to inspect the results of the two proposed interfaces on the user experiences. We used the logged data and questionnaire feedback from the two studies. There are two conditions and five summarized factors in the model (as shown in Figure 3). In OSAs, there are two manipulations based on the proposed interfaces. In SSAs and EXPs, we introduced four factors based on the classification by [13] and our post-experiment questions. In INT, we listed the entropy of the participant's selection diversity. The model shows that the two manipulations have different positive effects on the system, which helps to explain the EXP by the meditating impacts of SSA and INT.

The progress can be summarized as threefold: 1) we propose two interfaces that support the continuously controlled fusion of several relevance aspects with inspectability and controllability. 2) we provide evidence that the diversity-aware interface not only helps the user to perceive diversity but also helps the user to improve usability in the real world beyond simple relevance tasks. 3) finally, we discuss the user experience mediating effects on the proposed interfaces through a structural equation model analysis.

5 FUTURE PLANS

The preliminary study supports the two interfaces were useful in different aspects. The Scatter Viz is helpful on perceiving trust of the recommendations due to it reveals the relative relations of the multi-relevance, which may help to gain the transparency of the RS. The Relevance Tuner is let the user perceive usefulness and diversity due to the better inspectability and controllability. The next move is to consider the synergy of the two proposed interfaces designs. My plan can be summarized in three folds. **1) Full Design:** the two proposed interfaces were contributing to the different subjective system aspects (SSAs), i.e., perceived diversity and trust, which means they are useful in different contexts. I plan to combine the two designs as a Full Design, so a total four manipulations (Basic List, Scatter Plot, Relevance Tuner and Full Design) would be tested in a within-subject user study. The participants would be asked to find (explore) the scholars in academic conferences with specific criteria. **2) Generalizability:** one of the limitations of the two proposed studies is the small sample size, which decreased the robustness of the findings. A more extensive scale (size = 50) controlled study is scheduled to explain the conceptual components, which mentioned in the research questions section above. **3) Social Interactions:** I plan to extend the INTs with more diversity and usability metrics, e.g., the Gini index of the exploration diversity, the user's rating and the engagement on time spending. The goal is to correlate the logged user interactions to the user-centered evaluation framework, which aims to explain the causality through the conceptual components [9].

REFERENCES

[1] 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.

[2] Engin Bozdag and Jeroen van den Hoven. 2015. Breaking the filter bubble: democracy and design. *Ethics and Information Technology* 17, 4 (2015), 249–265.

[3] Peter Brusilovsky, Jung Sun Oh, Claudia López, Denis Parra, and Wei Jeng. 2016. Linking information and people in a social system for academic conferences. *New Review of Hypermedia and Multimedia* (2016), 1–31.

[4] Jilin Chen, Werner Geyer, Casey Dugan, Michael Muller, and Ido Guy. 2009. Make new friends, but keep the old: recommending people on social networking sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 201–210.

[5] Michael D Ekstrand, Daniel Kluver, F Maxwell Harper, and Joseph A Konstan. 2015. Letting users choose recommender algorithms: An experimental study. In *Proceedings of the 9th ACM Conference on Recommender Systems*. ACM, 11–18.

[6] Rong Hu and Pearl Pu. 2011. Enhancing recommendation diversity with organization interfaces. In *Proceedings of the 16th international conference on Intelligent user interfaces*. ACM, 347–350.

[7] Bart P. Knijnenburg, Svetlin Bostandjiev, John O'Donovan, and Alfred Kobsa. 2012. Inspectability and Control in Social Recommenders. In *6th ACM Conference on Recommender System*. 43–50. http://dl.acm.org/citation.cfm?id=2365966

[8] Bart P Knijnenburg and Martijn C Willemsen. 2015. Evaluating recommender systems with user experiments. In *Recommender Systems Handbook*. Springer, 309–352.

[9] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction* 22, 4-5 (2012), 441–504.

[10] Q Vera Liao and Wai-Tat Fu. 2013. Beyond the filter bubble: interactive effects of perceived threat and topic involvement on selective exposure to information. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2359–2368.

[11] Sean M McNee, John Riedl, and Joseph A Konstan. 2006. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI'06 extended abstracts on Human factors in computing systems*. ACM, 1097–1101.

[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.

[13] Pearl Pu, Li Chen, and Rong Hu. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*. ACM, 157–164.

[14] Choon Hui Teo, Houssam Nassif, Daniel Hill, Sriram Srinivasan, Mitchell Goodman, Vijai Mohan, and SVN Vishwanathan. 2016. Adaptive, Personalized Diversity for Visual Discovery. In *Proceedings of the 10th ACM Conference on Recommender Systems*. ACM, 35–38.

[15] Chun-Hua Tsai and Peter Brusilovsky. 2017. Enhancing Recommendation Diversity Through a Dual Recommendation Interface. In *Workshop on Interfaces and Human Decision Making for Recommender Systems*.

[16] Chun-Hua Tsai and Peter Brusilovsky. 2017. Leveraging Interfaces to Improve Recommendation Diversity. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*. ACM, 65–70.

[17] 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.

[18] Jacek Wasilewski and Neil Hurley. 2016. Intent-Aware Diversification Using a Constrained PLSA. In *Proceedings of the 10th ACM Conference on Recommender Systems*. ACM, 39–42.

[19] Wen Wu, Li Chen, and Liang He. 2013. Using personality to adjust diversity in recommender systems. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media*. ACM, 225–229.

[20] Bo Xiao and Izak Benbasat. 2007. E-commerce product recommendation agents: use, characteristics, and impact. *MIS quarterly* 31, 1 (2007), 137–209.

[21] Yuan Cao Zhang, Diarmuid Ó Séaghdha, Daniele Quercia, and Tamas Jambor. 2012. Auralist: introducing serendipity into music recommendation. In *Proceedings of the fifth ACM international conference on WSDM*. ACM, 13–22.