Enhancing Recommendation Diversity by a Dual Recommendation Interface

Anonymous
For Blind Review

ABSTRACT

The beyond-relevance objectives of recommender system are getting more and more attention in the field. For example, a diversity-enhanced interface has been shown positively associate with the user satisfaction. However, little is known about how a diversity-enhanced interface can help users to fulfill the various real-world tasks. In this paper, we present a visual diversity-enhanced interface which presents recommendations in a two-dimensional scatter plot. Our goal was to design a recommender system interface to explore different relevance prospects of recommended items in parallel and stress their diversity. A within-subject user study with real-life tasks was conducted to compare our visual interface with a standard ranked list interface. Our user study results show that the visual interface significantly reduced the exploration efforts for the explored tasks. Also, the users' subjective evaluation shows significant improvement on many user-centric metrics. We show the user did explore a diverse set of recommended items while improving the user satisfaction.

KEYWORDS

Recommender System; Diversity; Beyond Relevance; User control

1 INTRODUCTION

Recommending *people* in a social system is a challenging task. The user may look for other people for a range of reasons, for example, to re-connect with an acquaintance or to find a new friend with similar interests [3]. This diverse used needs make it hard to generate a ranked list which fits all cases. A specific case where a single ranked list might not work well is a parallel hybrid recommendation system that fuses several recommendation sources. In this case different sources might be preferred for different needs (i.e., social similarity could work best for funding known friends while content-based similarity could be used to find people with similar interests). Several authors argued than the best approach in this situation is to offer user an ability to control the fusion by choosing algorithms [3, 4] or data sources [1]. However, it is not clear whether a casual user with no computer science background can fine-tune the provided interface to adjust the results to their exploration interests. Providing a visual interface that makes the process of fusion more transparent, for example, showing recommender sources and their overlaps as set diagrams [12, 19] could further address this problem. Yet the set-based approach has limited applicability since it ignores the strength of relevance (which is a continious variable). In this paper, we attempted to overcome the limitation of set-based visual fusion by exploring an visual fusion approach that represent a continuous nature of relevance aspects while keeping the fusion process transparent.

When selecting a visual metaphor for the transparent fusion of recommendation sources, we focused on better informing users about the diversity of recommender results. It has been demonstrated that a proper user interface could promote the diversity of information exploration. A diversity-enhancing interface evaluated in [6] lead to higher user satisfaction than the ranking list interface. Several attempts to design a diversity-focused interface using a dimensionality reduction technique to present the opinion similarity by latent distance are presented in [5, 15, 20]. However, the clustering distance was not easily interpretable for a user to make a personalized judgment. In this paper, we attempted to use a scatter plot two-dimensional visualization to present recommendations with several dimensions of relevance. Scatter plot is known as an intuitive way to present multidimensional data [8]. In our context, the scatted plot interface was used to help users combine different aspects of relevance for each recommended item. The user can further filter the recommendation result based on the extent of each dimension. We conducted a user study in an international conference, to compare the ranking list and scatter plot interface. Our user study results show that the new visual interface did reduce the exploration efforts on the proposed tasks. Also, the users' subjective evaluation shows significant improvement on many user-centric metrics. We provide empirical evidence that the user explore a diverse set of recommended item while improving the user stratification.

2 BEYOND THE RANKING LIST

We propose a recommender system to help conference attendees finding other relevant attendees to meet with a dual interface that included a ranking List and a scatter plot components. The ranking list is a classic way of presenting the recommended results in one dimension, from high to low relevance. The scatter plot was added as a diversity-promoting interface to show the recommended result in two dimensions with the second dimension used to reveal the diversity. Figure 1 illustrates the design of the dual interface. Section A is the proposed scatter plot. The interface presents each item (a conference attendee) on the canvas as a circle. The user can mouse over to highlight the selection. Section B is the control panel for the user to interact. The user can select the number of recommendation, the *major feature*, and *the extra feature* to visualize the recommendations on the scatter plot. The major feature is used to rank the results along the X axe and in the ranked list (section C) while the extra feature us used to show the diversity of results in the selected aspect along the Y axe. To further investigate diversity of displayed recommendations, the user can also use one data aspect as *category*

RecSys2017, Como, Italy

123-4567-24-567/08/06. . . $15.00
DOI: 10.475/123_4

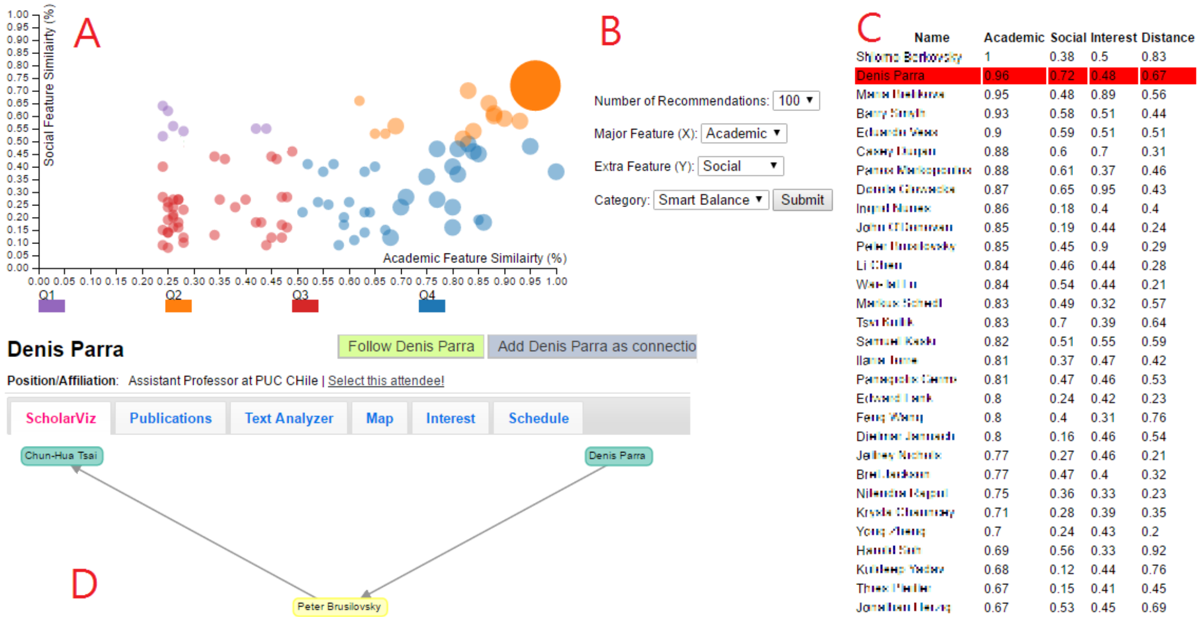

Figure 1: (A) Scatter Plot; (B) Control Panel; (C) Ranking List; (D) User Profile Page.

to color-code the results. The default category was *Smart Balance* which to color code by four quadrants with 0.5 ratio. Section C is the standard ranking list, more exactly is a combination of four ranked lists produced by four recommender engines explained below. To make four dimensions more clear, a normalized relevance of each user to the target user generated by each recommender engine is shown on the right side of the ranked list. Section D presents more detailed information about the person selected in the visualization or the ranked list. Among other aspects, four of six tabs explain visually how the relevance of the selected user to the target user is calculated by each recommender engine.

2.1 Personalized Relevance Model

To rank other attendees by their relevance to the target user, the system uses four separate recommender engines that rank other attendees along four dimensions that we call as *features*: text similarity of their *academic* publication, *social* similarity through the co-authorship network, current *interests* of CN3 activities and the *distance* of their affiliation place to the target user. Each of the feature was defined as below:

The Academic Feature is determined by publication similarity between two attendees using cosine similarity [9, 18]. The function is defined as: $Sim_{Academic}(x, y) = (t_x \cdot t_y)/\|t_x\|\|t_y\|$, where t is word vectors for user x and y.

The Social Feature feature approximates social similarity between the target and recommended user by combining co-authorship network distance and common neighbor similarity from publication data. We adopted the Depth-first search (DFS) method to calculate the shortest path p [14] and common neighborhood (CN) [11] for the number n of coauthor overlapping in two degrees. $Sim_{Social}(x, y) = p + n$ for user x and y.

The Interest Feature is determined by the the number of co-bookmarked papers and co-connected authors of the experimental social system.The function is defined as $Sim_{Interest}(x, y) = (b_x) \cap (b_y) + (c_x) \cap (c_y)$, where b_x, b_y represent the paper bookmarking of user x and y; c_x, c_y represents the friend connection of user x and y.

The Distance Feature is simply a geographic distance between attendees. We retrieve the longitude and latitude data based on attendees' affiliation information. We used the Haversine formula to compute the geographic distance between any two pair attendees [18]. $Sim_{Distance}(x, y) = Haversine(Geo_x, Geo_y)$, where Geo are pair of latitude and longitude for user x and y.

2.2 Diversity Navigation Model

The system determines the personalized relevance score for all conference attendees. Instead of rank the recommended people by ensemble value, the user can filter the items based on multiple aspects of relevance through our system. There are two kinds of diversification.

1) Feature Diversification: the user can select any two pair of proposed features and spot the recommended items through the relevance intersection. All of the proposed features were calculating in a different scale. For example, the distance feature is physical distance by miles versus academic feature by percentage. To let all the features are comparable. We adopted standard Z-Score to normalize all the features to the same scale from 0 to 1. The function was defined as: $ZScore = \frac{x_i - u_j}{\sigma_j}$, where x_i is ith recommended item and j represents the corresponding features from 1 to 4. Then we use the standard Z-table to convert the $ZScore$ to the corresponding percentile p_ij. Hence, we can list all the features on the same scale for presenting in a ranking list or scatter plot diagram.

2) Coverage Diversification: a diversification model to help the user select the recommended item from different category. [7]. In SCATTER interface, we color-code the item from different categories, e.g. title, position and country. In RANK interface, we listed the category as one column for a user to utilize.

We can then measure the user selection diversity through the two diversification model. We observe the user interaction with from different "Quadrants" (feature intersection) [17], e.g. high academic and high social feature or high academic and low social features. Both of the diversity is measured by *Entropy*: $d_u = -\sum_{i=1}^{4} p_i log_4 p_i$, where p_i is the probability for a particular quadrant (feature or category) and the proportion of all the user's selection [10].

3 EXPERIMENT

3.1 Data and Participants

The recommendations produced by all four engines are mostly based on data collected by the Conference Navigator 3 (CN3) system [2]. The system has been used to support 38 conferences at the time of writing this paper and has data about 6,398 articles presented at these conferences, 11,939 authors, 6,500 users (attendees of these conferences), 28,590 bookmarks, and 1,336 social connections. To mediate the cold start issue for academic and social engines that occurs when users have no publications or co-authorship within CN3 [16], we used the Aminer dataset [13]. This dataset includes 2,092,356 papers, 1,712,433 authors and 4,258,615 co-authorship. By combining the CN3 data and Aminer database.

A total of 25 participants (13 female) were recruited in the user study. All of the participants were attendees at the 2017 Intelligent User Interfaces Conference (IUI 2017). Since the main goal of our system was to help junior scholars connecting to other people in the field, we specifically selected junior scholars such as graduate students or research assistants. The participants came from 15 different countries; their age ranged from 20 to 50. All of them can be considered as knowledgeable in the area of the intelligent interface for at least one academic publication from IUI 2017. To control for their experience in the field of the recommender system, we included a question about in the background questionnaire. The average answer score was 3.28 in a five-point scale, means most of them are familiar with the recommender systems.

3.2 Experiment Design and Procedure

To assess the value of the diversity visualization, we compared the dual interface with the scatter plot and the ranked list (SCATTER) with a baseline interface using only ranked list (RANK) with part A removed. The study used a within-subjects design. All participants were asked to use each interface consecutively for three tasks and fill a post-stage questionnaire at the end of work with each interface. At the end of the study, they were asked to compare interfaces along six aspects explicitly. The order of using interfaces was randomized to control for the effect of ordering. In other words, half of the participants started the study with the SCATTER interface. To minimize the learning effect (getting familiar with data), we used data from two years of the same conference: the SCATTER used papers and attendees from IUI 2017 while the RANK used the same data from IUI 2016.

	Control Panel Usage		Explanation Tab Usage	
	RANK	SCATTER	RANK	SCATTER
Task 1	3.88	4.12	8.56	8.56
Task 2	2.88	2.88	6.56	4.8
Task 3	2.56	2.84	8.12	6.76
Overall	9.32	9.84	23.23	20.12

Table 1: Usage Analysis: control panel usage, explanation tab usage. Column 2 & 3 shows the comparison of user clicks between RANK / SCATTER interfaces.

	Hover	Click	Time	Engage
Task 1	-37.16%	-69.71%(*)	+9.21%	+161.7%(*)
Task 2	-59.53%(*)	-63.67%(*)	-11.91%	+115.2%(*)
Task 3	-55.51%(*)	-66.45%(*)	+50.14%	+179.6%(*)
Overall	-48.35%(*)	-67.07%(*)	+9.47%	+134.8%(*)

Table 2: Efficiency Analysis: the frequency of hover, click, task time (seconds for finish each task) and engage time (seconds between each click). All columns show incremental changes between RANK and SCATTER interfaces. (*) indicates statistical significance at the 0.05 level.

	Diversity	Coverage - Country	Coverage - Position
Task 1	-20.4%(*)	-6.42%	-15.10%
Task 2	+24.29%(*)	+46.59%(*)	-17.16%
Task 3	+35.8%(*)	+45.45%(*)	-23.07%

Table 3: Diversity Analysis: the test of diversity and coverage with two category variable. All columns show incremental changes between RANK and SCATTER interfaces.

Participants were given the same three tasks for each interface. $Task1$: Your Ph.D. adviser asked you to find four Committee Member candidates for the dissertation defense. You need to find candidates with the expertise close to your research field while trying to lower the travel cost to the defense. $Task2$: Your adviser asked you to meet four attending scholars, preferably from different regions across the world with a close connection to your research group. $Task3$: You want to find four junior scholars (not yet faculty members) with reasonably similar interests among the conference attendees to establish your networking. The participants were asking to pick up the suitable candidates among conference attendees based on best judgment in each task. When designing the tasks, we attempted to make them realistic, yet focusing on multiple aspects of relevance as many real tasks are. We consider task 2 & 3 are diversity-oriented.

4 RESULT ANALYSIS

4.1 User's Objective Evaluation

Table 1 shows system usage for two interfaces. The data indicates that participants extensively used the control panel and explanation tabs to complete the tasks. There is no significant difference

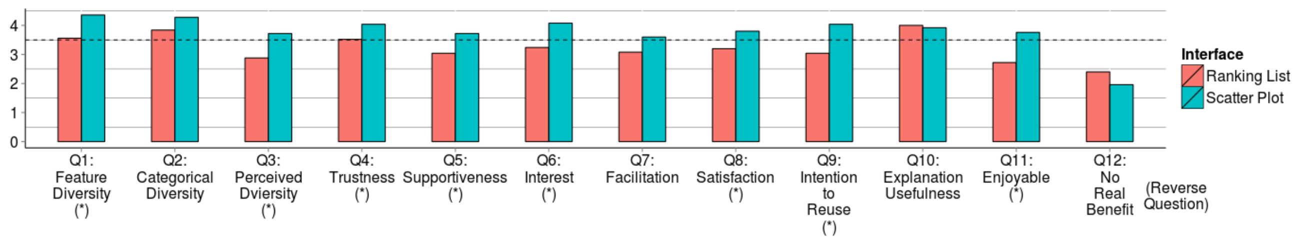

Figure 2: Usability and user satisfaction assessment results. A cut off value at 3.5 on the 5 point scale. (*) means significant differences at the 5% level (p-value < 0.05)

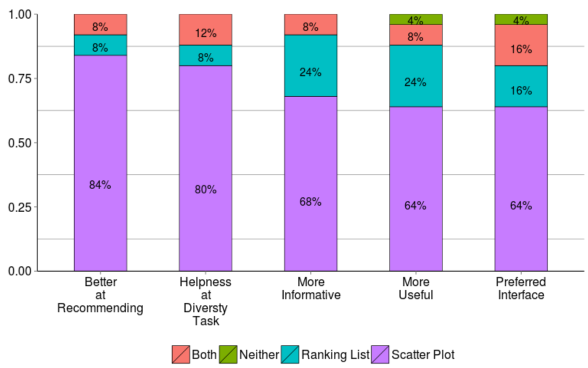

Figure 3: Preference Results: the final preference test after user experienced the two interfaces.

between the interfaces, although, in SCATTER interface, the users tend to use the explanations less.

Table 2 shows the work efficiency comparison between two interfaces. We counted how many mouseovers (hover) and clicks the users made to complete each task and expressed the number of actions done in SCATTER as a percentage increase or decrease from RANK. The data shows, with SCATTER, user completed the same tasks with 40-60% fewer mouseovers and about 66% fewer clicks. At the same time, we found no significant difference in the time spent on the tasks. The data hints that each action in SCATTER delivered more interesting information to explore, Indeed, we found that with SCATTER, the users spent significantly more time between clicks engaged in analyzing results.

Table 3 shows the diversity analysis for each task and interface. We found the diversity and coverage measurement is showing task difference. All three tasks are with a significant feature diversity difference between two interfaces but in the different aspect of features. In task 1 (relevance-oriented) is with less diversity on Academic/Distance features and less coverage on country and position variable. The SCATTER interface helps to explore the attendees with multi-relevance more accurately. The task 2 & 3 (diversity-oriented) are with more diversity of Interest/Distance and Social/Distance features, respectively as well as higher coverage in the country category. The result shows the users response to the same task with a different pattern of exploration on diversity and coverage.

4.2 Subjective Evaluation

To compare user subjective feedback, responses to the post-stage questions were analyzed using paired sample t-tests. The result is shown in Figure 2. We compared the eight aspects of subjective feedback from the participants. Among them, SCATTER interface received a significantly higher rating for six aspects: Trust (Q4), Supportiveness (Q5), Interest (Q6), Satisfaction (Q8), Intention to Reuse (Q9) and Enjoyable (Q11). In two questions, facilitation (Q7) and the control reversed Benefit Question (Q12), SCATTER scored higher, but not significantly. It is Interesting to see RANK interface scored a bit higher (not significantly) on explanation usefulness hinting that the lack visualization made explanations more important in RANK. In the final preference test, the SCATTER interface received much stronger support than RANK in all six aspects (Figure 3). Most importantly, the dominating majority of users considered SCATTER as a better system for recommending attendees and a better help in diversity-oriented tasks as well as better in recommending.

5 CONCLUSION

In this paper, we present a dual visual interface for recommending attendees at a research conference. A research conference context introduces several dimensions of attendee relevance such as social, academic, interest and distance similarity. In this regard, the traditional ranked list makes it hard to express the diversity of recommended items (attendees). By spreading ranking in two dimensions, the suggested interface helps users in exploring recommendations and recognizing their diversity in several aspects. Our approach can be applied to any recommender system with multiple relevance features and item categories. To assess the visual approach, we conducted a user study in a real conference environment comparing our interface (SCATTER) with a traditional ranked list (RANK) in three practical tasks.

Our experimental result shows the tangible incremental impact the metrics of system usage, efficiency, and diversity. We found the SCATTER interface is benefited more on the aspect of perceived and helps on the diversity tasks. The final preference survey is shown a strong preference on the SCATTER interface. Interestingly, we also found the SCATTER interface was benefited more on the feature diversity tasks. The user feedback suggests they would easier to find and category variable through the RANK interface. However, even user feedback indicates an ease of use for selecting and inspecting an item by category through the RANK interface. The user who used the SCATTER interface still shows significantly higher coverage measurement between tasks.

The main contribution of this paper is to prove the enhanced diversity interface not only help the user to perceive the diversity [6] but also help the user to improve the usability on the real world beyond relevance tasks. We provide empirical evidence on how to design a recommender system interface for the user to explore a diverse set of recommended item while improving the user stratification.

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] 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.

[3] 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.

[4] 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.

[5] Siamak Faridani, Ephrat Bitton, Kimiko Ryokai, and Ken Goldberg. 2010. Opinion space: a scalable tool for browsing online comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1175–1184.

[6] Rong Hu and Pearl Pu. 2011. Helping Users Perceive Recommendation Diversity.. In *DiveRS@ RecSys*. 43–50.

[7] Marius Kaminskas and Derek Bridge. 2016. Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 7, 1 (2016), 2.

[8] Hannah Kim, Jaegul Choo, Haesun Park, and Alex Endert. 2016. Interaxis: Steering scatterplot axes via observation-level interaction. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 131–140.

[9] Christopher D Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to information retrieval*. Vol. 1. Cambridge university press Cambridge.

[10] Jennifer Moody and David H Glass. 2016. A Novel Classification Framework for Evaluating Individual and Aggregate Diversity in Top-N Recommendations. *ACM Transactions on Intelligent Systems and Technology (TIST)* 7, 3 (2016), 42.

[11] Mark EJ Newman. 2001. Clustering and preferential attachment in growing networks. *Physical Review E* 64, 2 (2001), 025102.

[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] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. 2008. Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 990–998.

[14] Robert Tarjan. 1972. Depth-first search and linear graph algorithms. *SIAM journal on computing* 1, 2 (1972), 146–160.

[15] 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.

[16] Chunhua Tsai and Peter Brusilovsky. 2016. A personalized people recommender system using global search approach. *IConference 2016 Proceedings* (2016).

[17] Chun-Hua Tsai. 2017. An Interactive and Interpretable Interface for Diversity in Recommender Systems. In *Proceedings of the 22Nd International Conference on Intelligent User Interfaces Companion (IUI '17 Companion)*. ACM, New York, NY, USA, 225–228. DOI: http://dx.doi.org/10.1145/3030024.3038292

[18] Chun-Hua Tsai and Yu-Ru Lin. 2016. Tracing and Predicting Collaboration for Junior Scholars. In *Proceedings of the 25th International Conference Companion on World Wide Web*. International World Wide Web Conferences Steering Committee, 375–380.

[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.

[20] David Wong, Siamak Faridani, Ephrat Bitton, Björn Hartmann, and Ken Goldberg. 2011. The diversity donut: enabling participant control over the diversity of recommended responses. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1471–1476.