See What You Want to See: Visual User-Driven Approach for Hybrid Recommendation

Denis Parra  
PUC Chile  
Santiago, Chile  
dparra@ing.puc.cl

Peter Brusilovsky  
University of Pittsburgh  
Pittsburgh, PA, USA  
peterb@pitt.edu

Christoph Trattner  
Know-Center  
Graz, Austria  
ctrattner@know-center.at

ABSTRACT
Research in recommender systems has traditionally focused on improving the predictive accuracy of recommendations by developing new algorithms or by incorporating new sources of data. However, several studies have shown that accuracy does not always correlate with a better user experience, leading to recent research that puts emphasis on Human-Computer Interaction in order to investigate aspects of the interface and user characteristics that influence the user experience on recommender systems. Following this new research this paper presents SetFusion, a visual user-controllable interface for hybrid recommender system. Our approach enables users to manually fuse and control the importance of recommender strategies and to inspect the fusion results using an interactive Venn diagram visualization. We analyze the results of two field studies in the context of a conference talk recommendation system, performed to investigate the effect of user controllability in a hybrid recommender. Behavioral analysis and subjective evaluation indicate that the proposed controllable interface had a positive effect on the user experience.

Author Keywords
Recommender systems; SetFusion; human factors; user interfaces; user studies

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation (e.g. HCI): User Interfaces

INTRODUCTION
Recommender systems have emerged as an important solution to help users in finding relevant items in a large item pool [14]. These systems have been in use for over 20 years recommending items in a wide range of domains such as news, movies, music, academic articles, jobs, or social network contacts. Over the years, several principal recommendation approaches have been developed and explored. These include collaborative filtering (user-based [18] and item-based [19]), content-based recommendation [1], as well as various hybrid-methods [5]. Nowadays, recommender systems are an essential component of many online services such as Amazon.com, Netflix, LinkedIn, Twitter, and Facebook.

Traditionally, research in this field has focused on improving the predictive accuracy of the recommendation algorithms. Recent studies have gone beyond the study of algorithms, exploring the importance of Human-Computer Interaction (HCI) on the user experience with recommender systems [3, 7, 10, 15, 20]. These studies have shown how visual features, enhanced interaction, and specific user characteristics affect the user engagement with the system and their decision to accept or dismiss recommendations beyond the off-line prediction accuracy paradigm. Some of the interface characteristics studied are transparency, explainability, and controllability.

In this paper we focus on one of the least explored HCI aspects of recommender systems - user controllability over the recommendation process. More specifically, we explore user controllability in the context of a hybrid recommender for conference talks. According to the experts in the area of hybrid recommenders [5], a considerable fraction of hybrid recommender systems deals with situations where the target system needs to fuse several recommendation sources to produce a single ranked list. Traditional approaches reviewed in [5] included weighted, mixed, and switching hybridization. In all these cases, the system decides how the sources should be integrated leaving the users nothing but browsing the integrated ranked list. We believe, however, that the user who understands the nature of the fused sources might be in a better position to choose the proper way to fuse them. We also believe that the final ranked list might not be the best source for the users to explore hybrid recommendation. A visual interface that offers the user a chance to control the fusion process and supports this controllability with an enhanced visualization of the fusion process might be more conductive user success and satisfaction than traditional “black box” system-driven hybridization.

To explore the value of user control and enhanced visualization in the context of hybrid recommendation we developed a novel interface that allows users to manually fuse different recommendation methods. In this paper we review the motivation behind this work, introduce our user-driven visual hybrid recommendation interface, and present the results of two studies that explored the value of this interface in the context of real research conferences.
Summarizing, the contributions of this work are: (a) presentation of a novel hybrid recommendation interface that combines Venn diagrams and sliders and allows users to fuse and inspect different recommendation methods; and (b) analysis of two field studies of the SetFusion interface in a conference talk recommendation context – a domain where user controllability has been rarely studied in the past.

BACKGROUND
This section reviews two lines of research that are related to our work: Visual Approaches for Recommendations and Recommender Systems for Research Talks or Articles.

Visual Approaches for Recommendations. We can name just a handful of interfaces that present recommended items in a visual form rather than as a traditional ranked list. Examples include SFViz [8], a sunburst visualization to allow users finding interest-based content in Last.fm, and Pharos [21] a social map visualization of latent communities. Other examples that also include a richer user interaction are PeerChooser [15], and SmallWorlds [9] which focus on representing collaborative filtering, and TasteWeights [3], an interactive visual interface for a hybrid music recommender. What differentiates SetFusion from TasteWeights [3] is the broader depth of field provided by the interactive Venn diagram widget, a characteristic that allows users to keep their attention on the recommended talks (the details) but also on the intersection among the recommender approaches (the high-level view) [11]. Verbert et al. [20] introduced TalkExplorer, focusing on both rich interaction and transparency of recommendation. It allowed users to explore and to find relevant conference talks by analyzing the connections of talks to different entities such as user bookmarks, recommender algorithms and user tags. A study of TalkExplorer found that the effectiveness and probability of item selection both increase when users are able to explore and interrelate multiple entities. Although TalkExplorer had good results, it had limitations: its visualization was unnecessary complex and some users had difficulty understanding the “intersections” of entities. In SetFusion, we applied a more straightforward Venn diagram rather than the clustermaps used in TalkExplorer to show set intersections, while adding fusing sliders to increase user control over source integration.

Recommender Systems for Research Talks or Articles. Recommending scientific and technical articles has been approached with a diverse range of methods and information sources. Basu et al. [2] used content-based (CB) filtering and collaborative filtering (CF) for recommending papers to reviewing committee members. McNee et al. [13] used the citation network to recommend citations of papers. They tested 4 CF methods (co-citation, user-item, item-item, bayesian) with two non-CF that also used articles content (graph search, google search). By performing offline evaluations and a user study, they suggested combining the algorithms or using different algorithms depending on the task: CF methods are more appropriate for recommending novel papers, while CB filtering might be more accurate when recommending familiar related work. Ekstrand et al. [6] focused on building introductory research lists by using "augmented" versions of CB, CF, and hybrid methods including the influence of papers within the web of citations. By conducting off-line experiments and a user study, they found that, for the task researched, CF outperformed CB and the CB-CF hybrid methods.

THE SETFUSION VISUAL RECOMMENDER
Conference Navigator
Conference Navigator 3 (CN3) [17] is the third version of a web system aimed at supporting conference attendees. CN3 offers users information about talks (conference program, proceedings, paper details, most popular papers), people (list of authors, list of attendees, groups) while also collecting and representing the user’s personal information (bookmarked talks, tags, connections, recommendations, and profile information). Among user-personalized features, the system offers several kinds of talk recommendations, which in the past were presented as a set of traditional ranked lists.

The Recommender Interface
CN3 leverages several sources of knowledge to generate recommendations including talk content (title and abstract), user tags, and user social connections [4]. Given their particular strengths and weaknesses, we believed that users should be aware of which sources were used to recommend a specific talk and have some level of control over the source selection in a recommender. However, the ranked lists of recommendations produced by traditional hybrid recommenders (i.e., recommender systems that fuse several sources of recommendation) do not allow the users to control and to combine them on demand. It typically does not even show which source produced which result. As mentioned above, SetFusion follows the set-based approach of TalkExplorer to visualize the sources that produced each relevant item, while using an easier-to-understand visual paradigm for set presentation – Venn diagrams. A screenshot of the SetFusion controllable recommender is presented in Figure 1. While a fused list of recommended talks is the central part of the interface (Figure 1(a)), the fusion process in SetFusion is both controllable and transparent. SetFusion allows the user to control the importance of three recommender methods by using sliders (Figure 1(b)), and provides a Venn diagram to examine and filter items recommended by one or more methods (Figure 1(c)). It also offers clear indications of the source of the recommendations in both visualization and ranked list views.
SetFusion Interface Interactions

The recommender interface offered the user several interaction areas: the list of recommendations, the sliders widget and the Venn diagram:

- **List of talks recommended** (Figure 2)
  1. **Open and close abstract**: by clicking on the link provided by each paper title, the users could see the abstract of the article.
  2. **Hover over color bar**: users could hover over the color bar to obtain an explanation of the method used to recommend the paper.
  3. **Bookmark a paper**: at the very end of each paper’s title, an icon indicates if the paper is bookmarked or not. This same icon allows the user to bookmark or remove the paper from the list of relevant items.
  4. **See 10 more**: By default, the system shows the top 30 recommended items. If the user wants to see more items beyond that point, she can click on the button “See 10 more”.

- **Sliders widget** (Figure 3)
  1. **Hover over explanation icon**: this action allows the user to obtain a more detailed explanation of the method by displaying a black floating dialog.
  2. **Move sliders**: by moving the sliders (or typing a number in the text box), the users change the relative importance of each method used to generate the list.
  3. **Update recommendation list**: after moving the sliders to adjust the importance of each method, the user must click on the button “Update Recommendation List” in order to sort the list of recommendations on the right-side panel.

Recommendation Algorithms

Three recommendation algorithms were used to produce a fused list of talk recommendations. The methods were:

*Content-based Recommendation*. In this method the user profile, composed of keywords, is matched to the title and abstract of non-bookmarked talks in the conference. The user profile is represented as a vector of tokens extracted from the title and abstracts of papers that the user bookmarked from the current and previous conferences hosted in CN3.

The user model can then be described as a vector of tokens \( \vec{u} = \{ w_1, w_2, \ldots, w_n \} \) using tf-idf weighting, and a document in the collection of conference talks is represented analogously as \( \vec{d} = \{ w_1, w_2, \ldots, w_n \} \). The recommended documents are ranked based on their similarity to the user profile. To this end, we computed the cosine-based similarity between the user profile \( \vec{u} \) and each document \( \vec{d} \) in the collection \[12\]. Apache Solr\(^1\) was used to index the conference talks and to implement the content-based recommender. The content-filtering functions in Apache Solr receive several parameters, the following four controlled by us: (a) \( \text{min.tf} \): The minimum frequency that a term must have to be considered in the user profile, (b) \( \text{min.df} \): The minimum number of documents that a word must appear in the collection, (c) \( \text{min.wl} \): The length

\(^1\)http://lucene.apache.org/solr/
The study logs recorded an extensive use of various SetFusion features. User questionnaires reported positive user attitudes to the SetFusion approach. Both sets of data provided good evidence that SetFusion was valuable for CN3 users. While the remaining subsections of this paper focus on the analysis of user behavior and perception in the UMAP 2013 field study, the bottom line CSCW 2013 data are presented in all cases to compare the impact of SetFusion in a study-adjusted and natural settings.

**User Participation at UMAP 2013**

The analysis of user participation and engagement data shown in Table 1, in comparison with CSCW 2013 pilot study, shows a remarkable increase in most of the participation parameters following the move from less natural to more natural preference specification. While the total number of users who had a chance to notice and use the SetFusion interface was comparable (84 vs. 95), the fraction of users who used the interface more than doubled (22 vs. 50) and twice as many users made an extensive use of it by bookmarking papers (6 vs. 14). The same proportion can be observed when comparing the number of users returning to the recommender interface. In brief, we observed that the engagement impact of the visual interface is about twice as large than registered in the less natural context of the pilot CSCW study.

A comparison of user bookmarking activities also provides some evidence that the UMAP 2013 interface was highly more productive. While the number of users more than doubled, the number of bookmarks made with the visual interface increased almost 4 times. While differences between the conferences don’t allow us to attribute the growth to the recommender, we could argue that in the presence of a more realistic recommender interface that properly takes into account users past bookmarked talks, the visual interface can provide better help to their users in locating relevant talks. Another interesting productivity observation is that the users were able to bookmark more talks while making fewer supportive actions such as talk ratings or talk details openings. The first kind of actions was perceived as important to improve content-based recommendation and get relevant talks closer to the top of the rating list, and the second kind was critical to choosing talks when title and authors provided insufficient relevance evi-

<table>
<thead>
<tr>
<th>Measure</th>
<th>SF CSCW13</th>
<th>SF UMAP13</th>
</tr>
</thead>
<tbody>
<tr>
<td># Users exposed to recommendations</td>
<td>84</td>
<td>95</td>
</tr>
<tr>
<td># Users who used recommender page</td>
<td>22</td>
<td>50</td>
</tr>
<tr>
<td># Users who bookmarked</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td># Papers rated / avg per user</td>
<td>130 / 13</td>
<td>86 / 10.75</td>
</tr>
<tr>
<td># Users who rated</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td># Users who answered survey</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>Average user rating</td>
<td>3.73</td>
<td>3.62</td>
</tr>
</tbody>
</table>

**Usage at Recommender Page**

<table>
<thead>
<tr>
<th>Measure</th>
<th>CSCW13</th>
<th>UMAP13</th>
</tr>
</thead>
<tbody>
<tr>
<td># Talks explored (user avg.)</td>
<td>16.84</td>
<td>14.9</td>
</tr>
<tr>
<td># Talks bookmarked / user avg.</td>
<td>28 / 4.67</td>
<td>103 / 7.36</td>
</tr>
<tr>
<td># People returning to recommender page</td>
<td>7 (31.8%)</td>
<td>14 (28%)</td>
</tr>
<tr>
<td>Average time spent in page (seconds)</td>
<td>261.72</td>
<td>353.8</td>
</tr>
</tbody>
</table>

**Bookmarking Popularity.** This is a non-personalized community-based recommender approach. It simply ranks papers based on their popularity in the conference community, i.e., the number of people who bookmarked the talk.

**USER STUDIES**

SetFusion was evaluated in the context of CN3 talk recommendation through a sequence of two field studies. A between-subjects field study was performed during the CSCW 2013 conference. The purpose of the study was to pilot-test SetFusion in a conference context and to compare the user response to our interactive interface with a non-controllable baseline (ranked list) recommender interface. To keep a reasonable level of control, which is important for a between-subjects study, the pilot study used a rather unnatural setup stage in its recommender interface: users had to complete two steps (choosing favorite authors from previous versions of the conference and picking authors’ most relevant papers) in order to see the recommendation list. Following the pilot study, we fixed some issues with the pilot SetFusion version and performed another field study of SetFusion at the UMAP 2013 conference. The availability of comparative data collected at CSCW 2013 allowed us to avoid baseline balancing and run the UMAP 2013 study in a more natural setting (i.e., we let the users to freely interact with the interface), which was important in assessing the true impact of the visual interface. To engage users in both studies, we promoted our recommender interface by e-mail among conference attendees and also by presenting a promotion image on CN3 home page. To measure the impact of SetFusion, we logged user activity with the systems. At the end of each conference we e-mailed all SetFusion users an invitation to answer a survey about the system.

**A summary of CSCW 2013 study.** Due to the lack of space, we are not able to present in detail the results of the pilot study conducted at CSCW 2013; however, these results can be found in [16]. The importance of the CSCW pilot study is that we found preliminary evidence that the controllable interface is more engaging for the conference attendees than the non-controllable one. There was also some evidence that the visual version was able to offer better ranking performance. The study logs recorded an extensive use of various SetFusion features. User questionnaires reported positive user attitudes to the SetFusion approach. Both sets of data provided good evidence that SetFusion was valuable for CN3 users. While the remaining subsections of this paper focus on the analysis of user behavior and perception in the UMAP 2013 field study, the bottom line CSCW 2013 data are presented in all cases to compare the impact of SetFusion in a study-adjusted and natural settings.

**Table 1. Participation and engagement metrics in the SetFusion interface of CSCW13 and UMAP13.**

- The study logs recorded an extensive use of various SetFusion features. User questionnaires reported positive user attitudes to the SetFusion approach. Both sets of data provided good evidence that SetFusion was valuable for CN3 users. While the remaining subsections of this paper focus on the analysis of user behavior and perception in the UMAP 2013 field study, the bottom line CSCW 2013 data are presented in all cases to compare the impact of SetFusion in a study-adjusted and natural settings.

**Author-Based Popularity Recommendation.** In this method, we ranked the papers considering the popularity of their authors based on the numbers of citations they had received in the ACM Digital Library. We collected a dataset from the ACM DL, and then we ranked the recommended papers following this procedure:

(a) List the papers of the conference hosted in CN3.
(b) Obtain the names of the authors from papers found in (a).
(c) Match the names of the authors with those in our ACM DB.
(d) For each author found in our ACM DB, obtain the number of references.
(e) Calculate the popularity of each paper found in (a) by aggregating (adding up or choosing the maximum of) the logarithm of the number of references of each of its authors.

**Bookmarking Popularity.** This is a non-personalized community-based recommender approach. It simply ranks papers based on their popularity in the conference community, i.e., the number of people who bookmarked the talk.
dence. While increases in these parameters provide evidence of user engagement and determination to get to relevant talks, a lower yield, or ratio of bookmarked talks to the total number of supportive actions indicates that the users had to work relatively harder to end up with the same number of bookmarks (which are the true output of the process). Finally, the analysis of time spent in the system provides evidence about both engagement and productivity. While the total time spent working with the the system has further increased, along with an increase in other engagement parameters, its increase was slightly lower than the increase of the number of bookmarks (1.35 vs 1.57), i.e., the users were able to work a bit more productively spending less time per bookmark.

Action Analysis for UMAP 2013

On average, users updated the list of recommendation after manipulating the sliders 4.36 times (over 11 users), which is clearly greater than the 2.25 times average usage of the CSCW 2013 SetFusion (over 8 users). To compare the usage of specific sliders, we need to consider that the recommender methods in CSCW and UMAP were not the same. In CSCW 2013, methods A, B, and C corresponded to: (A) frequently-cited authors, (B) content-based matching, and (C) co-authors of favorite authors. On the other hand, in UMAP the mapping is: (A) talk popularity, (B) content-based matching, and (C) frequently-cited authors.

If we observe the usage distribution over sliders (change SliderX in Figure 5), the participants of UMAP 2013 showed a more uniform behavior, with 5.62, 6.62 and 5 changes of the sliders C (frequently-cited authors), B (content-based recommendation) and A (talk popularity), respectively. This distribution differs from CSCW 2013, where participants performed only 2.5 changes to the slider B (content-based recommendation), whereas they made greater use of the sliders A (6.8 times, frequently cited authors) and C (5.5 times, articles written by co-authors).

Users increased their usage of the Venn diagram in UMAP compared with CSCW, particularly considering the hover CircleX actions. In UMAP, users were more likely to hover over the circles of the Venn diagram to inspect the talks recommended by a single method or by a fusion of methods. As seen in Figure 5, five areas of the Venn diagram were explored more than four times in average: A (4.75), B (7.83), C (6.9), BC (10), and ABC (5.43), compared to only two that received the same average number of user actions during the CSCW: B (6.9), and C (4.71). This behavior provides more evidence of the Venn diagram’s role in identifying the papers recommended by one or more methods.

User Feedback Analysis for UMAP 2013

After UMAP 2013 had ended we e-mailed a link to the post-study survey to all CN3 users that tried SetFusion. They rated their agreement to several statements from 1 to 5. We highlight that SetFusion users, in general, perceived that they understood why the talks were recommended (M=4.13, S.E.=.25), that they felt in control using the sliders (M=4.25, S.E.=.4), and that they intended to recommend the system to colleagues (M=4.25, S.E.=.33). The same survey was administered at the end of the CSCW pilot study and we found two significant differences. Using t-test, we found that SetFusion users in UMAP significantly disagreed more with the statement that a talk recommender system is not necessary in CN3 (M=1.5, S.E.=.21) compared to CSCW users (M=2.36, S.E.=.2), p<0.05; and they also gave a stronger indication of recommending this system to their colleagues (M=4.25, S.E.=.33) than CSCW 2013 users (M=3.36, S.E.=.28), p<0.05. These results might indicate that the natural setting of the UMAP study increased users’ appreciation for the visual and controllable features.

DISCUSSION, CONCLUSIONS AND FUTURE WORK

This paper explores the issues of controllability and transparency in recommender system interfaces. The focus of the paper is SetFusion, a visual hybrid recommender that fuses several recommendation sources. In contrast to the traditional hybrid recommender approaches known as weighted, mixed, and switching hybridization where the recommender system decides how the sources should be integrated in the single ranked list, SetFusion uses an interactive Venn Diagram and a set of sliders to make the fusion process transparent, controllable, and exploratory. To assess the value of the visual controlled hybridization, we implemented SetFusion in the context of a conference support system (CN3) and performed two field studies of the system at academic conferences CSCW 2013 and UMAP 2013. The CSCW study was designed to compare the impact of SetFusion with a traditional ranked list. To balance two conditions, it used a rather unnatural label, and explorable. To assess the value of the visual controlled hybridization, we implemented SetFusion in the context of a conference support system (CN3) and performed two field studies of the system at academic conferences CSCW 2013 and UMAP 2013. The CSCW study was designed to compare the impact of SetFusion with a traditional ranked list. To balance two conditions, it used a rather unnatural 2-stage preference elicitation at the start. However, it also allowed to reliably demonstrate the benefits of SetFusion over the ranked list approach [16]. UMAP study was designed as a one-condition study to assess the impact of SetFusion in natural settings. A comparison of study results provided in the paper demonstrated that SetFusion in a natural recommendation mode has an even greater impact on user motivation, performance, and attitude. First, several parameters indicated that SetFusion is more engaging to the users in its natural form – with a comparable number of users exposed to the systems, twice as many users used SetFusion, bookmarked talks with it and used it repeatedly. The users also bookmarked almost 4 times more talks and spent considerably more time with the
In our future work we hope to invest more time in improving SetFusion and exploring it in other recommendation contexts. In a broader scope, SetFusion is an important addition in a sequence of visual recommendation approaches explored by our team. We hope that this experience with SetFusion will allow us to distill more features that are critical for visual recommender interfaces and help us to advance our research in the area of visual user-controlled recommendation.

ACKNOWLEDGMENTS
The last author of this paper is supported by the Know-Center.

REFERENCES