



HAL
open science

Multi-dimensional Ratings for Research Paper Recommender Systems: A Qualitative Study

Shaikhah Alotaibi, Julita Vassileva

► **To cite this version:**

Shaikhah Alotaibi, Julita Vassileva. Multi-dimensional Ratings for Research Paper Recommender Systems: A Qualitative Study. International Symposium on Web Algorithms, Jun 2015, Deauville, France. hal-01171132

HAL Id: hal-01171132

<https://hal.science/hal-01171132v1>

Submitted on 2 Jul 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multi-dimensional Ratings for Research Paper Recommender Systems: A Qualitative Study

Shaikhah Alotaibi
University of Saskatchewan
181 Throvaldson Building
110 Science Place
Saskatoon, SK, Canada
Shaikhah.otaibi@usask.ca

Julita Vassileva
University of Saskatchewan
178 Throvaldson Building
110 Science Place
Saskatoon, SK, Canada
jiv@cs.usask.ca

ABSTRACT

Research paper recommender systems (RSs) aim to alleviate information overload for researchers. Existing approaches using collaborative filtering or hybrid approaches typically allow only one rating criterion (overall liking) for users to evaluate papers. We conducted a focus group qualitative study to explore the most important criteria for rating research papers that can be used to control the paper recommendation by enabling users to set the weight for each criterion. We investigate also the effect of using different rating criteria on the user interface design and how the user can control the weight of the criteria.

Keywords: Research paper recommender system; multi-dimensional rating; collaborative filtering; user-centric design

1 INTRODUCTION

Research papers allow researchers to communicate their most recent approaches and results and to receive continual updates about new research in their fields. Papers are also valuable learning resources for new researchers (e.g., graduate students). New researchers need help in choosing papers to read because they do not yet know much about the area in general, which papers contain important ideas or methods and who the pioneers and the active researchers in the field are. They need to read papers that satisfy their personal needs at the current time, but they can be overwhelmed with the huge amount of published papers available through digital libraries. Thus, paper recommender systems (RSs) have been developed to alleviate information overload.

RSs can be defined as “any system that produces individualized recommendations as output or has the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible options” [1]. Common RS approaches are content-based filtering (CBF), collaborative filtering (CF) and hybrid approaches. CBF is based on information retrieval techniques comparing a paper’s features (e.g., title, abstract, keywords, publication year) with the researchers’ features (e.g., interests, previous search queries) to find matches. CBF methods, which are based on text analysis, are widely used because each paper can be represented as a collection of words and CBF calculates the weight of each significant word in the paper to provide a measure of its importance. Examples of CBF for paper recommendations can be found in [2], [3]), and [4]). In contrast to CBF, the CF approach is generally helpful if few or no features about the items or users are available but user ratings are. CF for recommending papers could be based on readers’ explicit ratings (the ratings that are given by the user) [5][6], citation analysis (i.e., the relationship between papers by referencing or citing other papers) [7], or usage analysis (e.g., download, forward, print, annotate) [2]; [8]. In CF, similarities in the rating behavior of the users are recognized and used to provide each user with a personalized

list of papers based on her preferences of similar items (item-based CF) or ratings of similar users (user-based CF). The hybrid approach fuses both approaches to gain the advantages of using each approach alone and reduce their drawbacks.

Most CF algorithms require users to give just one overall (global) rating and then use the averages of all users’ ratings to correlate the items (or users) and compute “neighborhoods”. This approach is straightforward but not flexible enough to provide adequate details about the quality of the rated item/service. The inflexibility of global ratings produces biased recommendations because two users may give the same global rating from two different perspectives [9]. For example, two researchers may rate a paper the same, but the first researcher’s evaluation is based on the paper’s readability while the other’s is on the paper’s novelty. For this reason, some RSs are based on CF algorithms or hybrid approaches that use multi-criteria ratings based on two or more perspectives (dimensions). However, in all the existing multi-dimensional RS, the rating criteria are chosen by researchers and do not allow the users to change the importance weights of these criteria.

In this paper, we investigate researchers’ opinions of the most important criteria in rating the quality of a paper. The important decisions include how many and which criteria to include, whether to include the possibility of users’ assigning different weights to the criteria and how the user interface for ratings should be organized. There is a danger in giving more control to the user and increasing the complexity because that may lead to cognitive overload and reduce the user’s rating activity. Although all published research papers are peer-reviewed and evaluated for their significance and novelty, we are looking to the users reviews as consumers for these papers. The importance of that is to try to develop a recommender system that take into account the user’s preferences using the multiple ratings and the user’s weights for each rating criterion. Different weights will produce different a more personalized recommended paper list to each user.

Following the framework proposed in [10]) we concentrate on two layers of the model: deciding upon adaptation (DA) and applying adaptation decisions (AA) to elicit the users’ opinions. We use a qualitative method by applying focus group discussion to involve the end users and their perspective throughout this study.

The paper is organized as follows: The next section presents the research methods. Section 3 presents the results obtained in the discussion and summarizes the ideas derived from the participants. We discuss the related work in section 4. Finally, in section 5, we summarize the conclusions of this work and outline future directions.

2 METHOD

In RS design and development, quantitative research methods are commonly used to evaluate the accuracy, efficiency and the effectiveness of proposed algorithms as well as user

satisfaction [9][11]. However, the qualitative method is well suited to exploring different options and user requirements. To our best knowledge, the question of which features (i.e., criteria) to consider in evaluating the quality of a paper has not been addressed in a systematic way, and all existing approaches that include multiple rating criteria seem to have been developed based only on the authors' intuition. We adopted a focus group approach, a qualitative research method suited for exploratory research that would allow us to examine 1) whether the participants thought including different ratings criteria is useful in finding higher quality research papers and 2) whether they support the idea of having the user control the weight of different criteria. We also want to extract and confirm some guidelines for user interface design that can make the RS more understandable and user-friendly.

A focus group, as defined by [12], is a moderated discussion on a predefined set of topics with 6–12 participants. The advantage of having small group discussion is allowing participants to develop shared understanding of the topics while voicing their opinions. The process uses open-ended questions to enrich the discussion with different ideas, perspectives and conflicts that reveal the similarities and differences in participants' thoughts. Some questions should be defined before conducting the focus group, whereas some may evolve as follow-up questions based on the participants' responses. One of the big advantages of a focus group as an exploratory method is its open-endedness, the fact that the direct interactions with and among the participants enable exploration and redefining the scope of ideas.

2.1 Participants

Eight people participated in this study: one postdoctoral fellow, two Ph.D. candidates in their fourth year of research, and five master's students in their second year of studies. The participants were recruited from the Computer Science Department of the University of Saskatchewan. We invited these people specifically because we aimed to engage active young researchers who still need to read research papers for their studies and for writing their theses. Graduate students usually need to read dozens of papers to be familiar with their research area, to find problems that they can contribute to solving and, finally, to write their thesis. Graduate students usually struggle to find good papers to read when they are starting their studies. Participants were invited personally through e-mail. Eleven of the invited people agreed to participate in the study, but three could not participate because of other commitments. The participants' demographic data are shown in Table 1. Most of the participants were international students who had studied previously in different countries, which meant their diverse educational backgrounds could enrich the discussion.

Table 1. Participants' Demographics Data

Participant	1	2	3	4	5	6	7	8
Degree	Master	Postdoc	Master	Ph.D.	Ph.D.	Master	Master	Master
Gender	Male	Female	Female	Female	Male	Male	Male	Male
Country of Origin	Iran	Iran	Canada	Nigeria	Nigeria	Taiwan	China	Canada

2.2 Focus Group Settings

The focus group was held in a quiet room in the Social Sciences Research Laboratories (SSRL), University of Saskatchewan, dedicated to conduct qualitative studies (e.g., focus groups or think-aloud sessions). Preparation of the

room included providing microphones for audio recording the session and a round table with comfortable chairs to enable the participants to communicate with each other easily.

When the participants arrived, they were welcomed and asked to read carefully and sign the consent forms. They were reminded that the focus group discussion was audio-recorded to make it easy to transcribe the session without missing any of the participants' opinions. They were also told to avoid interrupting any of the other participants so that all participants had the chance to complete their ideas. This also helps the transcription by preventing the audio file from being garbled.

Two types of questions were asked during the focus group session. First, we asked closed questions to examine how familiar the participants were with RSs:

- C1. Have you been recommended items, or are you familiar with any kinds of recommendations?
- C2. Are you familiar with research paper RSs?

Then we asked open-ended questions as follows:

- O1. When you want to read a paper, how do you choose that paper? What criteria do you consider?
- O2. In using a RS for research papers, do you prefer to see overall ratings (one rating as general rating) for the paper or multiple criteria? Justify your answer.
- O3. If you are asked to evaluate and rate a research paper, in your opinion, what are the three most important rating criterion?
- O4. What do you think about giving the user the control to assign importance weights to those criteria?

3 ANALYSIS

The discussion began with questions C1 and C2. We noticed that the participants were very motivated to talk about the topic. All of them were familiar with RSs, specifically in recommending items to buy; seven of them were familiar with RSs used by scholarly websites, such as Academia.edu and Google Scholar Updates. Below are some direct quotes from participants as they expressed their hypotheses about the recommendation algorithms used:

[Participant 2 (P2): "I think they [Google Scholar Updates] may use some index [keywords] that you use in your papers. Try to match it and it will suggest to you some papers that have been recently published in your field."

[P8: "Google Scholar¹ seems to list them in order of how many citations they have. That's the only thing I've noticed on Google Scholar."

[P1: "I think Google Scholar itself is somehow [a] recommender system based on [the] number of citation[s] that the paper has? For example, if you search using keywords and you see a paper that has thousands of citations, so you prefer that one to the others." P5: "Yeah, but they [Academia.edu] just send me an e-mail with publications. So maybe they use the area of interest. I don't know."

[P4: "When you sign in to academia.edu it comments with recommendations so are you interested in this area? You either click [to] select it or not; once you selected it, it starts recommend papers in these areas. I was wondering one time that why it recommends some papers that I feel they are not in my research areas, I looked into my profile and I discovered that I've actually selected that before."

¹ <http://scholar.google.ca/>

[P4: “ResearchGate² send e-mail by questions and their answers using the keywords that I used previously to search.”

3.1 Brainstorming the Evaluation Criteria of Papers

Participants were asked to think about how they usually choose a paper to read and the reasons (i.e. criteria) that they choose to read a paper or not. They were asked to define each criterion so we would know exactly what they meant by each of them. They discussed the following 9 criteria.

1-Clarity: Participants converged toward defining clarity of the paper as how well the paper is written in English, whether it is using simple terms and how the style of writing makes the ideas presented by the paper obvious. They thought this is one of the criteria that make them decide to read the paper or not because the clearer the paper the more accessible and understandable it is. However, participants distinguished between two terms: clarity and understandability. They referred to clarity mostly as the quality and simplicity of the language used, while understandability was to mean how clear and logical the paper is in presenting the technical approach and the research methods used. For example, P5 said, “I think one is about grammar... and it was written poorly. So that might make it difficult to read. The second one is the paper is written well in English but it’s still difficult to get what that person is talking about, but I think that might be technical clarity.”

2-Technical clarity: participants felt that the technical content of the paper is one of the most important criteria in technical disciplines, such as engineering and computer science. However, the technical content may not be sound or may not be presented in an understandable way. For example, some authors use a lot of difficult technical terms that make the paper hard to understand. Similarly, complex mathematical formulas may be introduced without giving a hint or justification of why they are necessary. In some empirical or applied science papers the description makes it difficult to understand what the goals of the experiment are and why a specific evaluation or statistical evaluation method is chosen, assuming that the reader is very familiar with the subject.

3-Willingness to Cite This Paper: this criterion corresponds most closely to the general opinion of readers regarding liking or disliking the paper, if they intend to cite this paper in the future or not. This means they like it and find the information useful for their work. A high rating on this criterion would mean that, even though users may find the paper not useful in some aspects or may not like its clarity, it still brings value, can be useful and is worth citing.

4-Length of the paper: some participants thought that one of the criteria that encourage them to read the paper is its length. The shorter the paper, the higher the chance it will be read. “Sometimes a paper’s, like, 40 pages and you’re, like, I probably don’t want to read this,” P3 said.

However, they thought this could be a ranking criterion rather than a rating criterion, since the page length is an objective feature rather than a feature requiring a subjective evaluation (rating).

5-Closeness to purpose or task: participants defined this criterion as how much the paper is relevant to what the user is doing at the moment (e.g. writing a paper, finding another approach to solve the same problem). For example, if users read a paper, they might want to read another similar paper that uses a different approach or different method or frames the same problem from a different perspective. P2 commented that “Sometimes [the user] want[s] something newer. Sometimes [the user] want[s] something close.” This criterion was raised by one of the participants and did not receive support from the others.

6-Relevance: the participants defined this criterion as the similarity between the content of the paper and the user’s research interests. The recommended papers must be relevant to the user’s interest somehow. The participants thought that this criterion is more important if the user has interests in multidisciplinary research fields. They wanted to be able to judge the relevance even though this can be automated (e.g. using CBF). They also suggested enabling the user to tag papers so that they can relate the paper to topics according to their own choice.

7-Comprehensiveness: participants described a paper as comprehensive if it reviews what has been done in the research area (i.e., literature review) without going deeply into detail. Some participants called this criterion “coverage.” They mentioned that comprehensive papers are more important for new researchers that need to read “something that at least gives a brief overview... a summary of what has been done like literature review”, P4 said.

8-Difficulty level: some participants insisted that the difficulty level of the paper is important to consider in deciding to read the paper, but they could not agree on what defines a paper as difficult to read. While there are linguistic metrics of text complexity that could be used, the participants felt that the problem with the difficulty of research papers is more nuanced and text complexity metrics are not sufficient to evaluate it. There was a strong conflict in the discussion of this criterion. Some participants argued that the paper could be difficult because it has many mathematical equations, theorems and proofs. Other participants said the mathematical issues were only in the applied sciences, but when we consider other fields such as social sciences, the papers do not usually contain any mathematics yet can still be difficult to understand. They recognize that some readers are good at mathematics and may feel that papers with lots of mathematics are easy. Papers that use many technical phrases, acronyms or jargon that are specifically used in that research field are difficult to understand to those who are unfamiliar with them. One participant commented, “Sometimes the paper is very hard to read. I don’t mean English but in the technical field. They use lots of English phrases that are specifically used in that field.” Depending on the reader’s preparation or experience, the paper may be easier or harder to read.

² <http://www.researchgate.net/>

Some other participants found that papers that discuss experiments or user studies are easier to read than the papers that describe theories. “But I think that is also subjective. It depends on the kind of research that the paper discusses. Like, if it is strictly theoretical research then it might be difficult for most people to research. But if it is, like, experimental... it is easier to understand,” said P4. Most participants stated that the distinction between theoretical and experimental papers is important but not available in any of the existing research paper-bookmarking websites, such as CiteULike³, Mendeley⁴ or Zotero⁵. P2 said, “If I want to trust or rely on any RSs, I want to see it ignore the paper that comes with lots of theoretical stuff.” This distinction can be done by enabling the user to categorize papers as experimental or theoretical. Then users can specify in their profile that they prefer one of them to the other (e.g., experimental), which can be another feature to be used by the content-based filtering or hybrid RS to find matches.

9-Value added by reading this paper. This criterion refers to how much new information the user gains by reading this paper. Some participants thought that it is useful to rate the usefulness of the content of the paper based on their knowledge, whereas other participants thought that it would be hard to rate old papers. P8 said, “I feel like if a paper’s really original and novel then it would be easier to give it a higher value added rating because you can see how many things you can do to build on this paper. But older papers you might say it seems like this old idea that you can’t build off of.” P5 commented that an older paper could not add value for them because they have already read many and newer papers in the same area.

3.2 Other criteria mentioned in the discussion

The participants mentioned some other criteria that are useful in evaluating a paper, such as how recent the paper is, the authors’ reputation (h-index) and the number of citations that the paper has received. They discussed the importance of the year of publication and the relationship between the publication date and the number of citations, acknowledging that it is not straightforward to base a conclusion about how important the paper is on its citations. P4 said: “The paper might be new. It has not had much citation, but it is very good.” The reputation of the author is also important; users may be interested in reading a paper written by author(s) that they know and follow. However, the participants realized that, although these criteria are important to consider, it is better to consider them as ranking criteria for the results rather than rating criteria for users, since all of these criteria have numeric values that are usually fixed at the time of considering them (e.g., the length) or change slowly (e.g., the number of citations, the year of publication, the h-index of the author).

After we combine the explicit and implicit users’ interests and the paper ratings and the content of the paper and using a hybrid approach, the objective criteria (e.g., paper length, text complexity, year of publication, authors’ reputation, and number of citations) can be used to narrow and order the list of results.

3 <http://www.citeulike.org/>

4 <http://www.mendeley.com/>

5 <https://www.zotero.org/>

3.3 Overall Rating versus Multiple Ratings

When we asked the participants Question O2 (whether they preferred to have the overall rating of the paper or multiple ratings based on different specific criteria), the participants expressed wishes for an overall rating and ratings based on three specific rating criteria for two reasons. First, the user sometimes strongly likes some aspects of the paper, but their general opinion of it may be low. All the participants agreed with the opinion of P5: “You still give your opinion on the different dimensions. But generally you still come up with the overall impression whether you accept or reject [the paper].” Second, the user might have different opinion from the available criteria for the ratings and that can change the overall rating of the paper.

The participants thought it is important to have different criteria that reflect different aspects of the paper’s quality, which would make the recommendations more personalized to their needs. All participants agreed that the obvious disadvantage of having multiple ratings for each paper is that “It’s too much work to click on all the rating criteria for each paper.” However, the solution is to “keep [the ratings] short and simple,” design the system carefully and display the benefits of having multiple ratings on recommendations that they receive.

3.4 Choosing the Most Important Three Rating Criteria

After the participants felt that there were no more criteria to add to the list, we asked them to choose the three most important criteria among those produced in the brainstorming discussion related to O1. We initiated the discussion with questions O3. One participant suggested starting by removing the least important criteria and focusing later on ranking the remaining ones. First, most (five or more) of participants suggested removing the “willingness to cite the paper” and “value added by reading the paper” because these criteria expressed a general impression about the paper that could be replaced by the overall rating. The participants thought that the “length of the paper” should be removed because it is an objective, static parameter that could be used by the RS as a ranking criterion after the set of possible recommendations was generated. Another criterion that participants felt that could be automatically inferred by the RS, instead of relying on user rating, is the “closeness to research area.” Each user has a user model that contains the user’s preferences, the research interests and any related topics of interest. The user model is built automatically using data that are entered during user registration or collected by the system based on the user’s history of searching, bookmarking, etc.

Table 3 shows the resulting ranking of the rating criteria. All participants agreed that the clarity of the paper is the most important rating criterion. It is important for the users to read well-written papers that use simple and well-defined terms and that are easy to read. One participant commented, “number one should remain clarity and that it is easy to read,” expressing the group consensus. However, there were some comments that suggested more nuanced views. Another participant pointed out that a paper’s clarity is related to its difficulty and that the difficulty criterion is subjective. Yet difficulty came only fourth in the ranking and was supported by only three of the participants.

Technical clarity, which can be “interpreted as the technical clarity” (P2), became the second most important criterion, as ranked by seven participants. The paper is technically clear when the technical details are presented so

that the work can be reproduced. Such details include consistent mathematical notation and formularization, detailed description of the research method and the experimental setup.

Table 2. Ranked List of the First Five Important Rating Criteria

Criterion	Ranking position	Number of users agreed on the ranking position
Clarity	1	8 (100%)
Technical clarity	2	7 (87.5%)
Relevance	3	5 (71.4%)
Difficulty level	4	3 (37.5%)
Comprehensiveness	5	2 (25%)

There was a long discussion to choose the third-ranked criterion among relevance, difficulty level and comprehensiveness. Finally, in a vote taken to resolve the question, five participants chose relevance.

3.5 Control Over Different Rating Criteria

When the participants have been asked O4, they showed their strong support of the idea of enabling the users to show their preferences for different rating criteria by assigning weight to the criteria. They felt that the ability to change their preferences could help users find better papers according to the criteria that they prefer most or to a combination of different criteria. They acknowledged that users' needs change with time, such as when senior researchers might give more weight to the technical clarity, while new researchers might give more weight to clarity and comprehensiveness. The participants discussed some ideas about how the user interface could enable user control over the weights of the criteria which are discussed in the next subsection.

3.6 Paper Recommender System and User Interface Design Implications

The focus group discussion proposed some interesting ideas for designing the paper RS (see Table 4). Participants showed a preference for having sliders to adjust the importance weight for each of the criteria with a default setting for each one (e.g., in the middle of the slider). However, they mentioned that it is important to make a minimum threshold value to not allow the user to ignore completely any of the criteria. For example, users would not be able to make the slider go less than 20% for the clarity criterion.

Table 4. Implication of Raised Issues on the Paper RS design

Issue or Aspect	Effect on the RS Design
Topic relevancy to the user's interests	Use hybrid recommendation approach based on model of user interests and ratings.
User control over the weights of the rating criteria	Use sliders for each criteria; use checkboxes to choose which criterion the user wants to change the weight; or enable ranking the criteria
Overall rating or multiple ratings	Use both overall rating and multiple ratings; overall rating is used to reflect the general impression of the user about the paper
Experimental or theoretical papers	Enable users to classify the papers according to predefined categories (e.g., ACM classifications) and specify preferred categories in their profile

Users do not like to fill in much information	Keep criteria few, well defined and show tips that give clear descriptions
Users cannot do the multiple ratings before reading the paper	Enable multiple ratings only for users who have read the paper (e.g., by using checkboxes)
Author reputation, publication date, number of citations, paper length, text complexity (measured with text analysis tools as one component of the difficulty level of the paper)	Recommendation algorithm to use such "static" parameters as ranking criteria for ordering the recommendations
Relate papers to topics and evaluate the paper relevancy	Enable tagging

One of the participants suggested ranking the three criteria instead of giving weight value by using sliders, "and just put this is the first one, this is the second one, and the third one... that would be way easier." Another participant suggested enabling the user to choose three criteria out of five, for example, by selecting them in a checkbox, and then the user could change the weight for the selected criterion (or criteria). He commented, "Maybe it should be dynamic."

4 BACKGROUND AND RELATED WORK

Although recommendation approaches using multi-criteria rating are very promising, little work has been done in recommending research papers. The previous multi-criteria paper RSs are summarized in Table 5. Tang and McCalla [13] use multi-criteria ratings to recommend papers for learners. Their algorithm clusters learners based on their interests and categorizes papers based on their pedagogical attributes. The learner's knowledge level (novice or advanced) is matched with the paper's technical level to recommend a personalized list of papers to the learner. Tang and McCalla also proposed a multi-dimensional RS that takes into consideration not only the learner's interest, but also the pedagogical value of the recommended papers [14]. This pedagogical value includes factors such as the potential learning value from reading those papers, the difficulty level, the probability of recommending the paper to other learners and the relatedness to the learner's work. The evaluation via a field study showed that interest is not the key factor in the recommendation, that incorporating multi-dimensional ratings improves the recommender performance and that the users were highly satisfied with the recommendations received. The study was then extended to compare different learning groups of undergraduate and graduate students [15]. The two experiments showed that including features other than overall ratings for the papers and the learner's interest can improve recommendations.

Another work relevant to the study of paper recommenders is a study of recommending learning objects (LOs) by [16]. The LOs' rating criteria include whether the LO is easy to use, the degree of the facilitation of learning and the relevance of the LO to the learning topics used. However, the evaluation is only of the proposed algorithm and is done by simulation, not by involving real users. Another simulation study has been conducted to test the Papyrus system [9], which is a paper recommendation system based on 10 different evaluation criteria (e.g., originality, readability, organization, literature review, etc.). All the above studies were used explicit ratings entered by the users.

Zarrinkalam and Kahani [17] described a multi-criteria paper RS that uses three implicit ratings. These criteria relate papers to each other using references, citations and co-author

relationships. The paper’s neighborhood is first identified, CF is used, and a separate matrix that relates paper to paper based on each of the three criteria is built and combined with the results of a CF algorithm to produce an ordered list of recommendations. This work differs from our work because the objective is to find paper citations and it does not look at the user evaluation of the paper’s quality.

Table 5. Different Paper Recommender Systems that Apply Multi-criteria Ratings

Previous Work	Criteria	Evaluation focus
Tang and McCalla [13]	Three <i>ad hoc</i> criteria: technical level, readability, usefulness	System functionality
Matsatsinis et al. [18]	Seven user-defined criteria: publication year, keyword relevance, impact factor of publication’s journal, citation number, author’s average citation number, acknowledgements and affiliation	Not evaluated; an example is described
Tang and McCalla [14]	Three <i>ad hoc</i> criteria: value added, difficulty level and probability to recommend to others	Examine the user satisfaction and the effect of considering pedagogical elements in making recommendation and
Winoto, Tang, and McCalla [15]	Four <i>ad hoc</i> criteria: facilitate learning, topical relevance, popularity and ease of understanding	Learners’ acceptance
Manouselis et al. [16]	Three <i>ad hoc</i> criteria: ease of use, facilitate learning and topical relevance	Algorithm performance
Papyres RS [5]	10 <i>ad hoc</i> criteria: contribution, originality, literature review, readability, organization, technical quality, testing, procedure and quality of references	Tests and compares five different approaches of defining user’s neighborhood in terms of accuracy
Zarrinkalam and Kahani [17]	Implicit ratings, three relationships between papers: references, citations and co-author	Assess the benefits of using Linked Data ⁶ , hybrid and multi-criteria RS for improving the performance of citation RS

All the previous works show that using multi-criteria ratings to recommend research papers has a positive effect on the recommendation accuracy as measured by the standard precision-recall metrics for recommender algorithms performance and evaluated using simulations with datasets. However, all these approaches define their own criteria in an ad hoc fashion as decided by the researchers. None of these studies has based the choice of different criteria on a study of the user’s opinions or needs. We believe that choosing the rating criteria must be done by users and that the

recommendation accuracy is not the only factor that determines whether the user accepts and consumes the recommended items. Integrating the user in the recommendation process allows the user to give feedback about the recommendations to validate or tell the RS that it is wrong. The only study that recommended considering a user-centric design is the work done by [18]. They exploited a methodology based on decision theory and calculated the weight of each of seven papers’ attributes such as keywords relevance and the citation index of a paper (see Table 5 for complete list) based on the user’s input of preferences.

However, we focus on the content quality of the paper by enabling the user to rate the paper by three quality criteria selected from this user study. We try to increase the user–RS interaction to maximize the personalization and the user satisfaction and increase the user’s trust in the RS [19]. The system can enable users to adjust their preferences in all three of the criteria they have chosen to find relevant papers based on those criteria by enabling an explanation-based user interface for filtering out irrelevant recommendations and producing recommendations interactively and in real time.

5 CONCLUSIONS, LIMITATIONS AND FUTURE WORK

We presented a focus group study aimed to elicit the users’ views on rating the papers using different rating criteria evaluating different paper quality aspects. In addition, we discussed with the participants issues related to user control of the rating criteria and the user interface design for research paper RSs. The results can be summarized as follows:

- [Participants thought that using different rating criteria is very useful to find more personalized recommendations.
- [Participants showed the strongest support for the following three criteria: clarity, technical clarity and relevance.
- [Participants liked the idea of controlling the weight of the different rating criteria and observing the resulting change in the list of recommendations. User control seems to contribute in the level of trust in the recommendations and in the RS.
- [Trusting the RS can be increased if the users can provide the RS with feedback to tell the RS whether it satisfies their needs (i.e., making the system scrutable [20]).

Design considerations for recommender algorithms:

- [It is important to find a way to judge fairly the recently published papers rather than relying on the number of citations as a criterion for paper quality.
- [It is important to make the RS more dynamic by giving some control to the user such as the ability to change the weight of each rating criterion and choosing the ranking order of the recommended papers.
- [Paying more attention to the presentation style of the recommendation as well as the user interface design can increase user satisfaction and trust in the RS.
- [Explaining to users how the recommendation list is chosen is important in their accepting the recommendation. The participants of this study only guessed about the recommendations that they saw in some of the paper recommendations, such as Academia.edu and ResearchGate (see the list of their hypotheses in section 3).

⁶ It is to link related data that are not previously linked. <http://linkeddata.org/>

However, the study has limitations. Using a focus group as a research method implies collecting the participants' subjective opinions, and we are not sure if their justifications can be helpful for other users. To address with this limitation, we selected users that represented different levels of experience, various cultural and educational backgrounds and both genders. However, they were all computer scientists. A focus group comprising social or life scientists may come to different findings. Therefore, our future work will repeat the study with graduate students from other science, applied science and humanity disciplines to discover if their opinions are similar or different from the participants from the area of computer science. We would also like to conduct a field study with experienced researchers (10 years or more in doing research) to discover if their criteria for finding good papers are different from those employed by graduate students. Another study that will be helpful, using a developed paper RS as a tool for the focus group, would be to investigate in more detail the significance of the objective criteria (e.g., publication date, number of citations) in ordering the recommendation results and to discuss with the users the idea of enabling the users to choose (and change upon their request) which criteria to use in ranking. We want to question if these changes would make the recommendations more satisfying and the RS more trusted by users.

6 REFERENCES

- [1] R. Burke, "Hybrid Recommender Systems: Survey and Experiments," *User Model. User-Adapt. Interact.*, vol. 12, no. 4, pp. 331–370, Nov. 2002.
- [2] B. Gipp, J. Beel, and C. Hentschel, "Scienstein: A Research Paper Recommender System," presented at the International Conference on Emerging Trends in Computing (, 2009, pp. 309–315.
- [3] C. Basu, H. Hirsh, W. W. Cohen, and C. Nevill-Manning, "Technical paper recommendation: a study in combining multiple information sources," *J Artif Int Res*, vol. 14, no. 1, pp. 231–252, May 2001.
- [4] Q. He, J. Pei, D. Kifer, P. Mitra, and L. Giles, "Context-aware citation recommendation," in *Proceedings of the 19th international conference on World wide web*, New York, NY, USA, 2010, pp. 421–430.
- [5] A. Naak, H. Hage, and E. Aïmeur, "A Multi-criteria Collaborative Filtering Approach for Research Paper Recommendation in Papyrus," vol. 26, G. Babin, P. Kropf, and M. Weiss, Eds. Springer Berlin Heidelberg, 2009, pp. 25–39.
- [6] T. Y. Tang and G. McCalla, "A Multidimensional Paper Recommender: Experiments and Evaluations," *IEEE Internet Comput.*, vol. 13, no. 4, pp. 34–41, Aug. 2009.
- [7] S. M. McNee, I. Albert, D. Cosley, P. Gopalkrishnan, S. K. Lam, A. M. Rashid, J. A. Konstan, and J. Riedl, "On the recommending of citations for research papers," in *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, New York, NY, USA, 2002, pp. 116–125.
- [8] R. Torres, S. M. McNee, M. Abel, J. A. Konstan, and J. Riedl, "Enhancing digital libraries with TechLens+," in *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, New York, NY, USA, 2004, pp. 228–236.
- [9] A. Naak, H. Hage, and E. Aïmeur, "A Multi-criteria Collaborative Filtering Approach for Research Paper Recommendation in Papyrus," vol. 26, G. Babin, P. Kropf, and M. Weiss, Eds. Springer Berlin Heidelberg, 2009, pp. 25–39.
- [10] A. Paramythis, S. Weibelzahl, and J. Masthoff, "Layered evaluation of interactive adaptive systems: framework and formative methods," *User Model. User-Adapt. Interact.*, vol. 20, no. 5, pp. 383–453, Dec. 2010.
- [11] S. Watanabe, T. Ito, T. Ozono, and T. Shintani, "A paper recommendation mechanism for the research support system Papis," in *International Workshop on Data Engineering Issues in E-Commerce, 2005. Proceedings, 2005*, pp. 71–80.
- [12] D. W. Stewart and P. N. Shamdasani, *Focus groups: theory and practice*. Sage Publications, 1990.
- [13] T. Tang and G. McCalla, "Smart Recommendation for an Evolving E-Learning System: Architecture and Experiment," *Int. J. E-Learn.*, vol. 4, no. 1, pp. 105–129, 2005.
- [14] T. Y. Tang and G. McCalla, "A Multidimensional Paper Recommender: Experiments and Evaluations," *IEEE Internet Comput.*, vol. 13, no. 4, pp. 34–41, Aug. 2009.
- [15] P. Winoto, T. Y. Tang, and G. I. McCalla, "Contexts in a paper recommendation system with collaborative filtering," *Int. Rev. Res. Open Distance Learn.*, vol. 13, no. 5, pp. 56–75, Nov. 2012.
- [16] Manouselis, R. Vuorikari, and F. Van Assche, "Collaborative recommendation of e-learning resources: an experimental investigation," *J. Comput. Assist. Learn.*, vol. 26, no. 4, pp. 227–242, 2010.
- [17] F. Zarrinkalam and M. Kahani, "A multi-criteria hybrid citation recommendation system based on linked data," in *2012 2nd International eConference on Computer and Knowledge Engineering (ICCKE)*, 2012, pp. 283–288.
- [18] N. F. Matsatsinis, K. Lakiotaki, and P. Delias, "A System based on Multiple Criteria Analysis for Scientific Paper Recommendation," in *11th Panhellenic Conference in Informatics*, Patras, Greece, 2007, pp. 135–149.
- [19] B. P. Knijnenburg, M. C. Willemsen, Z. Gantner, H. Soncu, and C. Newell, "Explaining the user experience of recommender systems," *User Model. User-Adapt. Interact.*, vol. 22, no. 4–5, pp. 441–504, Oct. 2012.
- [20] N. Tintarev, "Explanations of recommendations," in *Proceedings of the 2007 ACM conference on Recommender systems*, New York, NY, USA, 2007, pp. 203–206.