How to Identify Specialized Research Communities Related to a Researcher's Changing Interests

Hamed Alhoori
Department of Computer Science
Northern Illinois University
DeKalb, IL, USA
alhoori@niu.edu

ABSTRACT
Scholarly events and venues are increasing rapidly in number. This poses a challenge for researchers who seek to identify events and venues related to their work in order to draw more efficiently and comprehensively from published research and to share their own findings more effectively. Such efforts are hampered also by the fact that no rating system yet exists to assist researchers in culling the venues most relevant to their current readings and interests. This study describes a methodology we developed in response to this need, one that recommends scholarly venues related to researchers' specific interests according to personalized social web indicators. Our experiments applying our proposed rating and recommendation method show that it outperforms the baseline venue recommendations in terms of accuracy and ranking quality.

Keywords
Altmetrics, Personalized Recommendation, Scholarly Venues

1. INTRODUCTION
In addition to the challenges presented by the rising number of scholarly publications and venues, the task of identifying relevant research venues is further complicated because the research landscape is becoming less compartmentalized. There are, for example, increasingly complex academic sub-disciplines and emerging interdisciplinary research areas, events, and venues (e.g., journals, conferences, symposiums, workshops, and seminars). In this competitive and sophisticated research environment, researchers find it challenging to remain up to date on new findings, even within their own disciplines. Furthermore, "context-drift" in scholarly communities is becoming more prevalent as researchers expand, evolve, or adapt their interests in rapidly changing subject areas over time.

Generally, researchers learn of scholarly venues related to their research interests from limited sources: by word of mouth from lab members, departmental colleagues, and members of other scholarly communities; by conducting online searches and reviewing the research articles returned by these searches; from venue rankings; or from publishers' reputations. In the past, these approaches worked satisfactorily because relatively few related venues existed for any particular field. Today, however, given the more multifaceted scholarly environment, researchers become acquainted with newly available and specialized venues only by spending considerable time browsing and evaluating.

JCDL '16, June 19-23, 2016, Newark, NJ, USA
ACM 978-1-4503-4229-2/16/06.
http://dx.doi.org/10.1145/2910896.2925450

In this study, we report on the effectiveness of a personal measure for evaluating venues we built based on user-centric altmetrics [1] and readings rather than conventional citation-based metrics. When applied, the measure recommends semantically related scholarly venues based on the researcher's specific interests and thus augments their awareness of relevant communities. In creating this measure, we drew on data from CiteULike,[1] a well-known social reference management system.

Prior to our work, few studies focused on methods for recommending scholarly events and venues. Among these, Klamma et al. [2] developed an approach that recommended academic events based on a researcher's event participation history. Boukhris and Ayachi [3] proposed a hybrid recommender for upcoming conferences in computer science based on venues from co-authors, co-citers, and co-affiliated researchers. Pham et al. [4] clustered users on social networks and used the number of papers published in a venue by a researcher to derive the researcher's rating for that venue. Other venue recommendation approaches based ratings on topic and writing style [5], title and/or abstract [6], and personal bibliographies and citations [7].

Most research to date has used citation analysis and researcher's publication or participation history to recommend venues. This approach is not useful for new researchers or graduate students who have yet to establish a record of scholarly activity. Furthermore, using only the venues in which researchers have previously published work undermines the "discovery" aspect of the recommendation process, as researchers might be interested in new areas in which they have not yet published. With these deficiencies in mind, our study explored a way to draw on a researcher's current personal article collections and readings to recommend tailored venues.

2. METHODOLOGY
Research articles can be associated with several metadata fields to produce recommendations. However, no direct metadata or ratings exist for venues. Nevertheless, references in a researcher's library can provide indirect information pertaining to a researcher's interests. We used such references and the years in which each was added to a researcher's library as factors in the measurement, which we refer to as personal venue rating (*PVR*). PVR takes into consideration how a researcher's interest in a given venue has changed over time. In Equation 1, we define PVR as a weighted sum for researcher u and venue v, and we refer to it as $r_{u,v}$:

[1] http://www.citeulike.org/

$$r_{u,v} = \sum_{i=y}^{1} w \log(v_{u,i} + 1) \qquad (1)$$

$v_{u,i}$ denotes the number of references in a researcher's u library from a specific venue v, which the researcher added during a certain year of the total number of y years, during which the researcher followed venue v. The weight w increases the importance of newly added references and is equal to i. *PVR* favors researchers who have followed a venue for several years over researchers who have added numerous references from a venue over fewer years. The *log* minimizes the effect of adding numerous references and helps to reduce shilling attempts. The addition of one allows for the case of one reference to be added to a library in a year. We used the year that a reference was added to the researcher's library, as it is more personalized than the published year.

We conducted an offline experiment using our CiteULike dataset, collected as described in [8]. We used user-based collaborative filtering (CF), item-based CF, SGD, and SVD++ algorithms. We compared researchers with similar interests in terms of their PVRs. To identify similarities among the researchers, we used the cosine similarity, the Pearson correlation similarity, and the Euclidean distance similarity.

We used a Boolean recommendation as a baseline and compared it with recommendations for scholarly venues based on PVR implicit ratings. Boolean ratings assume that all venues added by researchers are good venues and receive the highest rating.

3. RESULTS

We compared similarities that used PVR ratings and the user-based CF algorithm with the Boolean recommendation. The results demonstrate that PVR implicit ratings achieved higher precision (Figure 1), recall, and NCDG at lower neighborhood sizes. Additionally, using the PVR we were able to provide recommendations for up to 98% of users. We measured NMAE and NRMSE at different neighborhood sizes, and found that the Euclidean-weighting achieved the lowest NMAE and the lowest NRMSE. We compared the performance of four algorithms that used PVR ratings at different percentages of the training set, and we found that SVD++ achieved the lowest NMAE and the lowest NRMSE (Figure 2).

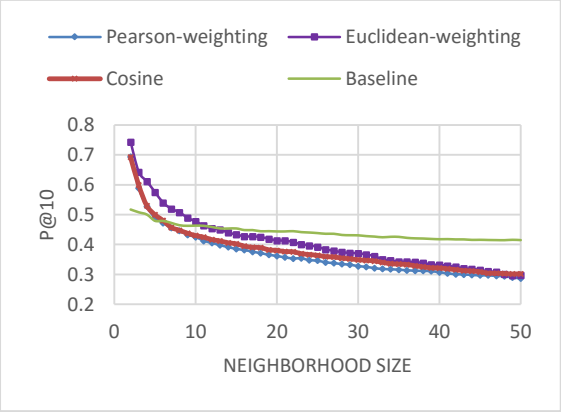

Figure 31. A comparison of user-based CF algorithm with similarities that use PVR ratings and the baseline at different neighborhood sizes

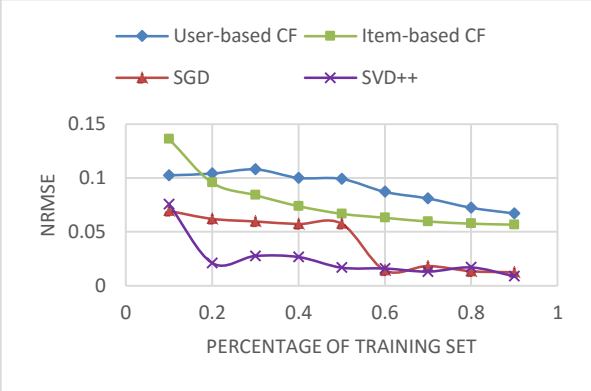

Figure 2. Comparison of algorithms at different training ratios

4. ACKNOWLEDGMENTS

This publication was made possible by NPRP grant # 4–029–1–007 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

5. REFERENCES

[1] H. Alhoori, R. Furuta, M. Tabet, M. Samaka, and E. A. Fox, "Altmetrics for country-level research assessment," in *Proceedings of the 16th International Conference on Asia-Pacific Digital Libraries*, 2014, vol. 8839, pp. 59–64.

[2] R. Klamma, P. M. Cuong, and Y. Cao, "You never walk alone : Recommending academic events based on social network analysis," in *Proceedings of the First International Conference on Complex Science*, 2009, vol. 4, no. 1, pp. 657–670.

[3] I. Boukhris and R. Ayachi, "A novel personalized academic venue hybrid recommender," in *IEEE 15th International Symposium on Computational Intelligence and Informatics*, 2014, pp. 465–470.

[4] M. C. Pham, Y. Cao, R. Klamma, and M. Jarke, "A clustering approach for collaborative filtering recommendation using social network analysis," *J. Univers. Comput. Sci.*, vol. 17, no. 4, pp. 583–604, Feb. 2011.

[5] Z. Yang and B. D. Davison, "Venue recommendation: Submitting your paper with style," in *Proceedings of the 11th International Conference on Machine Learning and Applications*, 2012, vol. 1, pp. 681–686.

[6] N. Kang, M. A. Doornenbal, and R. J. A. Schijvenaars, "Elsevier Journal Finder: Recommending Journals for your Paper," in *Proceedings of the 9th ACM Conference on Recommender Systems - RecSys '15*, 2015, pp. 261–264.

[7] O. Kucuktunc, E. Saule, K. Kaya, and U. V Catalyurek, "TheAdvisor: A webservice for ac ademic recommendation," in *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, 2013, pp. 433–434.

[8] H. Alhoori and R. Furuta, "Can social reference management systems predict a ranking of scholarly venues?," in *Proceedings of the 17th international conference on theory and practice of digital libraries*, 2013, vol. 8092, pp. 138–143.