Journal of Engineering Science and Technology
Special Issue on ATITES2021, April (2021) 10 - 17

EXPLOITING SCIENTIFIC COLLABORATION NETWORK FOR ENHANCED ACADEMIC EVENT RECOMMENDATION

GHASSAN BILAL[1],
ABDULJALEEL AL-HASNAWI[1,*], GHAIDAA AL-SULTANY[3]

[1]Information Technology Center, University of Technology, Iraq
[2]College of Technical Al-Mussaib, Al-Furat Al-Awsat Technical University, Iraq
[3]Information Technology College, University of Babylon, Iraq
*Corresponding Author: adbuljaleel.alhasnawi@atu.edu.iq

Abstract

Academic events are widely attended by authors to publish their scientific findings, gain new knowledge, and expand their scientific collaboration network. With the massive scholarly data about diverse academic events (Incl. seminars, workshops, and conferences) held around the world, it is difficult for authors to find related information about the academic events in which they may be interested. Recommendation systems resolve the problem of data overload by providing a list of recommended events that are preferred and relevant to the target author. In this study, we present an enhancement to existing academic event recommendation approaches by exploiting the scientific collaboration network of the target author. Academic events attended by author's collaborators, who have a strength connection with the target author regarding research similarity, are nominated and filtered using collaboration filtering. Then, PageRank algorithm utilized to perform ranking of the filtered list of events to recommend a list of events that are most relevant to the target author. Experiments are conducted on data from the Computer Science Bibliography (DBLP) and Wiki Calls for Papers (WikiCFP). The experimental results show that the proposed approach enhances the recommendation precision with an average of 20% and enhances the recommendation recall rate with an average of 25%, as compared with others.

Keywords: Academic events, Collaborators, Recommendation, Social network.

1. Introduction

Nowadays, the academic community continues to grow and there are a large number of academic events held in diverse topics around the world. Finding the most suitable and relevant event to attend is a crucial task for researchers [1]. Furthermore, the numbers of researchers, publications, and academic venues have risen beyond the imagination for the rapid development of information technology. Therefore, obtaining informative information in big scholarly data is a challenge due to data overload [2]. The current search tools like Workshop Directory1 and Eventbrite2 that are used to retrieve information about academic events such as topics, organizers, host countries, submission and deadline, cannot effectively filter the massive information and wasted a lot of time due to information overload [3]. Recommender Systems (RS) are mainly designed to deal with the issues of information overload and help people make decisions by providing accessible and high-quality recommendations list that are preferred to the target users. The existing events recommendation tools face some limitations such as outdated event recommendations [4] unstructured event announcement data [5] and cold start challenges [6]. These aspects raise a challenge for the academic recommendation services that help authors navigate at the right and suitable venue as well as finding valuable collaboration opportunities.

This study presents an enhancement to the existing academic event recommendation approaches by exploiting the scientific collaboration network of a target author. The proposed approach is mainly relying on analysing the social interactions of an author in scientific collaboration network. The proposed approach utilizes Collaborative Filtering technique [7] to explore the author relationships in co-author networks and compute the similarity among the target author and his/her collaborators based on a set of academic social factors which are also used to predict the significant links between the target author and other candidate collaborators. Then, the PageRank algorithm [8] is used to rank the largest k-authors (a set of authors have the most significant interactions with a given target author) for the purposes of nominating new valuable partners to collaborate with target author.

2. Literature Review

In recent years, academic event recommendation systems have become emerging as information filtering systems to dealing with information overload and have employed for suggesting relevant events to authors. Hoang et al. [9] proposed event recommendation model based on research relatedness, collaboration and events attended at the same conference. These are important factors, but not enough to suggest the appropriate conference to the target user. Kong et al. [10] suggested a hybrid approach by utilizing content-based method to find author domain and employing text-clustering method to produce feature vectors. The random walk model has also utilized to compute the influence of each author in every domain after extracting titles from all the papers published by each this author. Finally, they used feature vectors to compute the similarity and produce the final suggested list.

Garcia et al. [11] proposed a technique for comparing and recommending conferences based on Co-authorship network. The proposed technique produces a

[1] International directory of workshops, conferences & festivals. www.workshop-directory.com
[2] Eventbrite. https://www.eventbrite.com.

new similarity measure based on the information of authors and the attended conferences, simultaneously. However, the speed and efficiency of their work went slower with the increase in network size. For reducing this problem, they used new two techniques, called *Cluster-WSCS-based* and *Cluster- MWSCS-based*. Li et al. [12] studied the accepted articles of a new conference for the participants and took both the textual information and associations among a researcher and authors papers as features for the experiment. Citation relationships and co-author relationships are elicited from the whole of the authors' papers and their citation. This approach relied on relations but did not employ filtering and ranking like our approach does. Personalized Academic Venue recommendation (PAVE) [13] is proposed based on combined academic factors: co-publication frequency, relationship weight, and research academic level. PAVE generates co-publication network by using the relationship between author and venue where the author published his work. PAVE walks to the next node on extracted co-publication network with modified probability, the walking stops when reaching convergence. Finally, recommended the top N venue after ranking a pool of venues. This approach utilized the relationship between author and venue, which is in our opinion not sufficient to provide an accurate recommendation for other venues (events) that do not have direct relation with a target author.

3. Research Method

3.1. Research design

Past collaborations, including past collaborators and their attended events, are considered as one of the major factors that affecting the recommendation process. The collaborative filtering technique are used to explore the author relationships in Co-Authorship Networks. Then, to find the value of interaction between the target author and his collaborators, we find the significant link value between them by extracting a set of collaboration factors, related to the authors' research history. By considering these factors, it is going to be feasible to build a co-author network for a given author. Whereas weighted links in co-author network determine the tie of interaction value between nodes, which will be used later to measure the similarity between the target author and other authors. The output of building co-author network stage is a list of k-authors who have high scores of interactions with the target author. After that, the PageRank algorithm is applied to supply a chain for ranking authority scores of the target authors. It ranks a number of k-author that has significant interaction with the target author and then seeks to nominate a new valuable partner to collaborate with target author and suggests a list of academic events that are attended by those valuable collaborators.

3.2. Research procedure

The similarity among authors has been computed based on aggregating a set of social factors. We consider a set $C = \{c_1, c_2 ..., c_k\}$ indicates all end-user clusters. If users x and y belong to the cluster (ci), that is means their interests are very close. Then, events similarity is added to exploited to discover new partners who have been never cooperated with the target author, as shown in Eq. (1)

$$F1 = \gamma \sum_{p_i \in p_{am} p_j \in p_{non}} sim(p_i, p_j) + (1 - \gamma) \sum_{C_i \in p C_{am} C_j \in C_{non}} sim(E_i, E_j) \qquad (1)$$

where $\gamma \in [0,1]$ is the weight of the similarity control, γ equal to 0.6. $sim(p_i, p_j)$ used to measure of similarity between target author and new collaborators' publications. While $sim(E_i, E_j)$ is used to return the similarity between events. For all attended events by target author a_m. Eq. (2) used to find similar events to the set of events attended by the target author a_m.

$$F2 = \frac{\sum_{e_i \in E_{am}, e_f \in E_f} sim(e_i, e_f)}{|E_{am}|} \tag{2}$$

where e_i is one event of the set of events has attended by target author (E_{am}), e_f is future events, and $sim(e_i, e_f)$ is a function of similarity. Eq. (3) used to calculate the similarity between event attended by the most interaction author A_k and future event e_f.

$$F3 = \frac{\sum_{a_i \in A_k} \left(\frac{\sum_{e_i \in C_k} sim(e_i, e_f)}{|E_k|} \times T(a_m, a_n) \right)}{\sum_{a_i \in A_k} T(a_m, a_n)} \tag{3}$$

Every author is an author in A_k who has interaction with target author with value (a_m, a_n) , the value of interaction $T(a_m, a_n)$ has an impact on the final decision on target author. $|E_k|$ refer to all numbers of events were attended by the most interaction authors. $sim(e_i, e_f)$ is returning the value of similarity between events of collaborator and future events. Combining Eqs. (2) to (4) to compute the recommender score of the future event.

$$F(a_m, e_f) = (1 - \beta) \times F2 + \beta \times F3 \tag{4}$$

The value β is a real number between [0, 1] which controls the weight of the similarity between $F2$ and $F3$. A Top N-recommended venue for target researcher a_m is produced based on the ranking of the recommended rate value of $F(a_m, e_f)$ which is a likelihood between [0, 1].

3.3. Data acquisition

Data were provided by two datasets: DBLP[3] as a source of collaborators, and Wiki for Calls for Papers (WikiCFP[4]) as a source for academic events. DBLP released in 2018, it contains information about a set of 2.1 million authors with 4.1 million publications published in more than 1500 journal articles, with about 5400 conferences. Data were provided by DBLP in the form of XML files. The system is trained using the papers' information that published within the time interval [2008-2010] as an input training dataset, which has 312,486 nodes (authors), with 1,055,435 edges (co-author relations), 258,313 papers, and 2,595 conferences. The proposed model is tested using the papers published in the time interval [2011-2012] as testing dataset. WikiCFP is an online resource that provides bibliographic information on major computer science conference proceedings and journals. Our model has been performed with a set of randomly selected target authors.

4. Results and Discussion

To prove the effectiveness of the proposed model, extensive experiments are conducted. All experiments were performed on a 64-bit Windows-based operation

[3] DBLP computer science bibliography: https://dblp.org/
[4] Wiki for Calls for Papers: http://www.wikicfp.com/cfp/allcat

system, Intel Core i7-4710 HQ 250 GHz processor with 8 GB RAM. All the programs are implemented with Python on the PyCharm IDE community edition.

4.1. Scientific collaborative network

Two experiments are repeated stimulatingly for 13 target authors. First, a set of past collaborators is recognized and only nodes with strong connections are nominated, as shown in Fig. 1(a). Then, a set of non-collaborated authors are recognized and only nodes with highest values of links with a target author are recommended, as shown in Fig. 1(b). The experimental results show that, 38% of past collaborators are recommended as valuable collaborators; 50% of non-collaborated authors are recommended as new collaborators.

The precision and recall values are compared to the recommended past collaborators, as shown in Fig. 2(a). The results show that increasing the number of recommended past collaborators from 1 to 13 slightly decreases the precision, while the recall rate is increasing accordingly. The highest precision value (58%) where the number of recommended items is low (2), while the highest recall rate (71%) where the recommended items is high (13). Further, the precision and recall values are also compared with respect to the number of recommended new collaborators, as shown in Fig. 2(b). The results show that the highest precision value (89%) where the number of recommended items is very low (1) and the highest recall rate (55%) where the number of recommended items is very law (1).

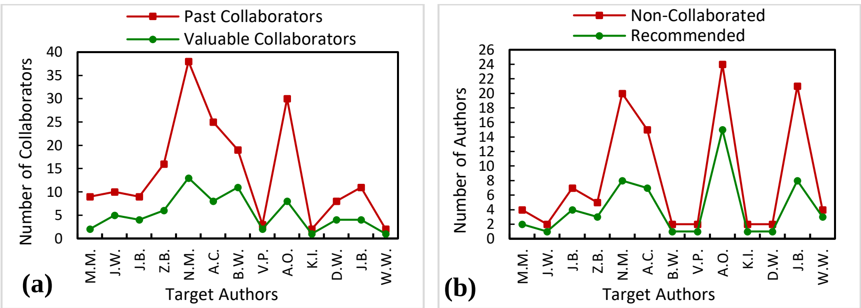


Fig. 1. Experimental results (a) valuable and (b) new collaborators.

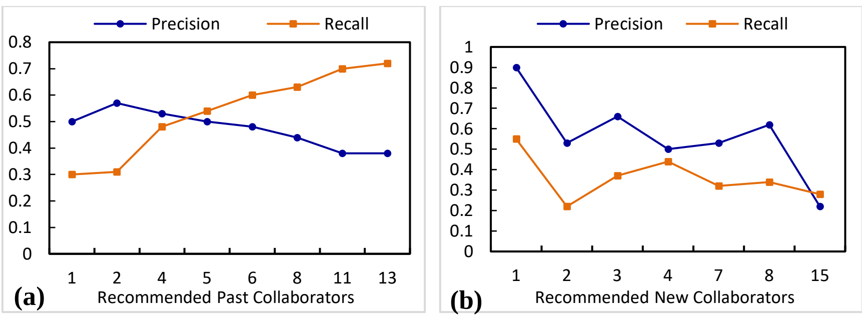
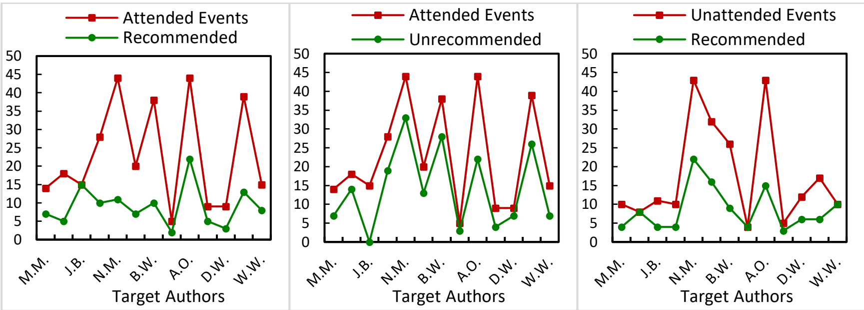


Fig. 2. Performance of the model (a) valuable, and (b) new collaborators.

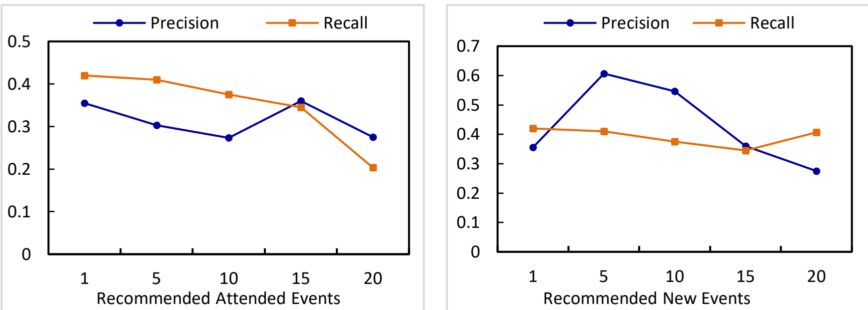
4.2. Enhanced event recommendation

Three experiments are repeated stimulatingly for 13 target authors. First, attended events are recognized and only events with strong associations with target author are recommended, as shown in Fig. 3(a). Second, attended events are recognized and only events with weak associations are unrecommended, as shown in Fig. 3(b). Third, unattended events are recognized and only events with highest association values are recommended, as shown in Fig. 3(c). The experimental results show that 40% of attended events are recommended, 60% of attended events are unrecommended, and 48% of unattended events are recommended. The precision and recall values of the proposed model are compared with respect to the number of recommended attended events, as shown in Fig. 4(a). The results show that increasing the number of recommended events from 1 to 20 slightly decreases the precision, the recall rate is also slightly decreasing from 1 to 10 then it increases for 15 then decreases for 20 recommended events. The highest precision value (42%) where the number of recommended events is low (1) and while the highest recall rate (36%) where the number of recommended items is high (15). Further, the precision and recall values of the proposed model are compared with respect to the number of recommended new events, as shown in Fig. 4(b). The results show that the highest precision value (60%) where the number of recommended items is very low (5) and the highest recall rate (42%) where the number of recommended items is very law too (1).



(a) (b) (c)

Fig. 3. Experimental results for (a) attended and recommended events, (b) attended and unrecommended events, and (c) unattended and recommended events.

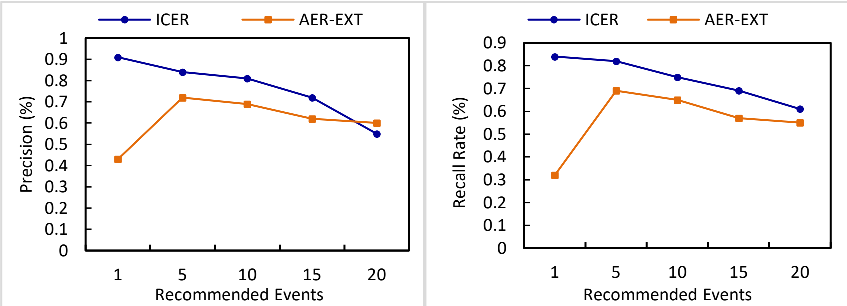


(a) (b)

Fig. 4. Performance of the proposed model in recommending (a) attended events, (b) new events.

4.3. Comparison with existing model

To measure the performance of the proposed model as compared with AER-EXT model [14], the same target authors has been chosen to test the two models. Experimental results, illustrated in Figure 5(a), show that precision is slightly decreasing with increasing the number of recommended events for both models except the case for AER-EXT model in which the precision is sharply increasing where the number of recommended events increases from 1 to 5. The proposed model has higher precision than AER-EXT for k = 1, 5, 10, and 15, and it has lower precision for k = 20. In average, our proposed model is 20% overperforms AER-EXT with respect to precision measure. Regarding recall measure, Fig. 5(b), the experimental results show that recall rate is slightly decreasing with increasing the number of recommended events for both models except the case for AER-EXT model in which the recall is sharply increasing where the number of recommended events increases from 1 to 5. The proposed model has higher recall rate than AER-EXT for all k values. In average, the proposed model is 25% overperforms AER-EXT model with respect to recall measure. Therefore, the proposed model has shown a better performance than the AER-EXT model regarding precision and recall rate measures.



(a) **(b)**

Fig. 5. Comparison of ICER and AER-EXT regarding (a) Precision, and (b) Recall.

5. Conclusions

This paper recognized the importance of the interrelationship between collaborators and events in academic social networks toward for improving the academic recommendation process. Based on that, we have proposed a model that assure the enhancement of the academic events recommendation by analysing the authors' historical interactions in his/her scientific collaboration network to recognize past collaborators and attended events. The proposed model utilized a set of academic social factors to find the most associated collaborators from the set of past collaborators. Moreover, these factors enabled the proposed model to produce a list of new collaborators (who never collaborated with the target author before) based on research relatedness and attended events. Then, PageRank algorithm is employed to rank all collaborators and recommended those who have the strongest link with the target author. Based on the outcomes of this process, a list of academic events attended by recommended collaborators are recognized and recommended only the most relevant events to the target author.

Extensive experiments on DBLP and WikiCFP datasets are conducted. The experimental results have shown that aggregating six academic social factors and employing PageRank algorithm for collaborators and their attended events ranking lead to enhancement in event recommendation of 20% regarding precision measure and 25% regarding recall measure as compared with other works.

References

1. Othman, W. S. N. A. W.; Mohd Shuib, N. L.; and Normadhi, N. B. A. (2017). AcadEvent: Recommender system for academic events. *1st International Conference on Industrial, Enterprise, and System Engineering (ICoIESE 2017)*, 171-176, Bandung, Indonesia.
2. Khusro, S.; Ali, Z.; and Ullah, I. (2016). Recommender systems: issues, challenges, and research opportunities. *Information Science and Applications (ICISA)*, 1179-1189, Springer, Singapore.
3. Mhirsi, N.; and Boukhris, I. (2017). Exploring location and ranking for academic venue recommendation. *International Conference on Intelligent Systems Design and Applications*, 83-91, Springer, Cham.
4. Yin, H.; Zou, L.; Nguyen, Q.V.H.; Huang, Z.; and Zhou, X. (2018). Joint event-partner recommendation in event-based social networks. *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, 929-940.
5. Korde, V. (2016). Information extraction for personalised services based on conference alerts. *International Journal of Data Mining, Modelling and Management*, 8(1), 93-105.
6. Zhang, W.; and Wang, J. (2015). A collective bayesian Poisson factorization model for cold-start local event recommendation. *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1455-1464.
7. Koren, Y.; and Bell, R. (2015). Advances in collaborative filtering. *Recommender Systems Handbook*, 77-118, Springer, Boston, MA.
8. Musto, C.; Semeraro, G.; de Gemmis, M.; and Lops, P. (2017). Tuning personalized pagerank for semantics-aware recommendations based on linked open data. *European Semantic Web Conference*, 169-183, Springer, Cham.
9. Hoang, D. T.; Hwang, D.; Tran, V.C.; and Nguyen, N.T. (2016). Academic event recommendation based on research similarity and exploring interaction between authors. *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 004411-004416.
10. Kong, X.; Jiang, H.; Yang, Z.; Xu, Z.; Xia, F.; and Tolba, A. (2016). Exploiting publication contents and collaboration networks for collaborator recommendation. *PloS one*, 11(2), e0148492.
11. Garcia, G.M.; Nunes, B.P.; Lopes, G.R.; Casanova, M.A.; and Paes Leme, L.P. (2017). Techniques for comparing and recommending conferences. *Journal of the Brazilian Computer Society*, 23(1), 1-14.
12. Li, S.; Brusilovsky, P.; Su, S.; and Cheng, X. (2018). Conference paper recommendation for academic conferences. *IEEE Access*, 6, 17153-17164.
13. Yu, S.; Liu, J.; Yang, Z.; Chen, Z.; Jiang, H.; Tolba, A.; and Xia, F. (2018). PAVE: Personalized academic venue recommendation exploiting co-publication networks. *Journal of Network and Computer Applications*, 104, 38-47.
14. Hoang, D.T.; Tran, V.C.; Nguyen, V.D.; Nguyen, N.T.; and Hwang, D. (2017). Improving academic event recommendation using research similarity and interaction strength between authors. *Cybernetics and Systems*, 48(3), 210-230.