



System Level Recommender System for Academic Venue Personalization: Multi vs. Linked Domain

Abir Zawali¹ and Imen Boukhris^{1,2}

¹ LARODEC, Institut Supérieur de Gestion de Tunis, Université de Tunis, Tunis,
Tunisia

abirzaouali1@gmail.com

² Ecole Nationale des Sciences de l'Informatique, Université de la Manouba,
La Manouba, Tunisia

Abstract. Standard academic venue recommender systems have been emerged aiming to help computer science researchers to find a suitable academic venue, in which they may publish their works. Therefore, using all the available user data provided from various domains is helpful to guide users, in their decision making process, to choose a new suitable academic venue that can be the right shape for their preferences in terms of academic venue type (conference/journal), publisher, ranking and/or location. In this context, to enhance the quality of delivered recommendation results, we propose a system level recommender system for academic venue personalization, based on multi domain vs. linked domain, comparative analysis. We investigate not only authors past publications, and authors, from the reference list, past publications from the DBLP dataset, but also from the IEEE dataset using system level multi and linked domain recommendation methods. Experimental results demonstrate the efficiency of our new system level multi domain recommender system.

Keywords: System level · Multi domain · Linked domain · Recommender system · Academic venues · Author's preferences

1 Introduction

As the number of academic venues increases, authors cannot choose the right one to submit their work to. In fact, even when the papers themselves are excellent, they may be rejected, because they are not relevant to the academic venue scope. In this context, standard academic venue recommender systems [2, 12, 13] upstanding as an effective solution to satisfy the authors preferences, interests, tastes, priorities and needs, and find the suitable academic venue that can fit their research scope [17, 18, 21].

However, works based on information extracted from single domain may hinder the standard academic venues recommender systems effectiveness. In fact,

users provide feedback in different ways and on different domains. Therefore, cross domain recommender systems burgeon as an expedient solution to incorporate information extracted from multiple domain and therefore to enhance the recommendation results.

Besides, a cross domain academic venue recommender system based on references is developed as addressed in [20] and refined in [19]. [20] is based on the hypothesis “for each paper, the target author may publish his work in one of the most appropriate venue in which one of the authors in the reference list has previously published”. For [19], it is based on the hypothesis “it is very important for an academic venue recommender system, that suggests personalized academic venue list, to take into consideration preferences provided by the authors (venue’s type, publisher, rank, location) and to filter out conferences or journals that do not match the author’s requirements”.

To ensure appropriate recommendations that correspond to the majority computer science researchers needs, recommender systems should leverage all available authors feedbacks implicitly and explicitly provided across maximum number of domains. In this context, we present system level recommender system based on multi domain, using DBLP and IEEE datasets as both source and target domain, compared to system level recommender system based on linked domain, using DBLP and IEEE datasets both as source domain and only DBLP dataset as target domain, to suggest personalized upcoming venues for computer science researchers, using information from references domain and authors domain for the both approaches. To do so, a personalized web interface is used on which, for each written paper, authors are able to specify their requirements, e.g., academic venue type (conference/journal), its publisher, its ranking and/or location, in order to get a list of recommended upcoming venues based on their interests without missing the submission deadline.

The rest of the paper is organized as follows: Sect. 2 gives necessary background on cross domain recommender systems. Section 3 exhibits data sources, explains the data extraction procedure, presents the proposed system level academic venue recommendation engine based on multi domain and shows the proposed system level academic venue recommendation engine based on linked domain. Section 4 discusses the experimental results by detailing comparative analysis. Finally, Sect. 5 concludes the paper and give an outlook over some future works.

2 Cross Domain Recommender Systems

2.1 Cross Domain Recommendation Tasks

The cross domain recommendation [9] is characterized by source domain (SD) and target domain (TD). There are three recommendation tasks [6], namely, cross domain, multi domain and linked domain.

Cross domain recommend items from the target domain using knowledge from the source domain.

Multi domain leverages knowledge from the source and the target domain to recommend items from both source and target domain.

Linked domain recommend items from the target domain using knowledge from the source and the target domain.

2.2 Notions of Domains

In the context of cross domain recommender systems, four levels have been defined [3] depending on the attributes and the type of recommended items [7], namely, system level, type level, item level, attribute level.

The system level, the same type of items, gathered from different datasets, is considered to belong to different domains, e.g., theater movies and TV movies.

The type level, similar items types that share certain attributes, are handled as belonging to different domains, such as movies and TV series.

The item level, different types of items that differ in most (if not all) attributes, are considered to be from different domains, e.g., books and film.

The attribute level, the same type of items with the same attribute have different values is regarded as belonging to different domains, e.g., a dramatic movie and a comedy movie.

2.3 User-Item Overlaps Scenarios

Across multiple domains, some relation needs to exist between users (U) and items (I). This relation is formed when users and items are found to be common in both source domain and target domain. Accordingly, we can identify four different cross domain scenarios [5], namely, no overlap, user - no item overlap, no user - item overlap, and user and item overlap.

- *No User - No Item overlap*: no users and no items are found to be common between source domain (SD) and target domain (TD), $U_{SD} \cap U_{TD} = \emptyset$ and $I_{SD} \cap I_{TD} = \emptyset$.
- *User - No Item overlap*: some users are found to be common between source domain (SD) and target domain (TD), but they have preferences for different items in both SD and TD, every item belongs to a single domain, $U_{SD} \cap U_{TD} \neq \emptyset$ and $I_{SD} \cap I_{TD} = \emptyset$.
- *No User - Item overlap*: some items are found to be common between source domain (SD) and target domain (TD), but they have been rated by different users from both SD and TD, every user belongs to a single domain, $U_{SD} \cap U_{TD} = \emptyset$ and $I_{SD} \cap I_{TD} \neq \emptyset$.
- *User - Item overlap*: users and items are found to be common between source domain (SD) and target domain (TD), $U_{SD} \cap U_{TD} \neq \emptyset$ and $I_{SD} \cap I_{TD} \neq \emptyset$.

3 System Level Academic Venue Recommendation: Multi vs. Linked Domain

The recommendations of academic venue are different from other items such as books or movies. In the academic venue recommendations [8, 10, 11], a researcher may attend the same conference several times. The user may submit his work to a conference that he already published in. Moreover, in the books or movies recommendation, the users are provided with a prediction of their future evaluation of items not yet rated.

Recently, cross domain recommender systems [19, 20] have been shown that recommendation based on cross domain can increase researchers satisfaction. In literature, [20] presents a cross domain recommender system which selects the most appropriate academic venue list based on the hypothesis “for each paper, the target authors may publish their work in one of the most appropriate venue in which one of the authors in the reference list has previously published”. It is covering the user interests across multiple domains, the references domain as source domain (SD) where the information is extracted from the papers in the reference list and the authors domain as target domain (TD) where the information is extracted from the authors papers. [19] recommends an academic venue list taking into consideration that researchers should easily interact with the recommender system and taking into account preferences provided by the authors (venue’s type, publisher, rank, location) to filter out irrelevant ones that do not matches the author’s requirements, without missing the submission deadline.

Despite, limiting cross domain recommendation engine to only two domains, i.e., authors and authors from the reference list domains may affect the cross domain recommendation forcefulness. In this context, we propose a comparative analysis between system level cross domain recommender system based on multi domain and also on linked domain to show that adding multi domains enhance the recommendation quality. Our proposed methods will notify researchers about academic venues that may match their preferences (venue’s type, publisher, rank, location). This is done using two steps, (see Fig. 1) for the multi domain system level and (see Fig. 2) for the linked domain system level.

3.1 Data Sources

Each recommender engine will require its own data sources that will provide the necessary information to generate recommendation. In what follows, we detail these data sources.

Past Publications DBLP Data Source. Information previously published authors papers is from “DBLP citation dataset” [14], it is publicly available on the Aminer website [1]. “DBLP citation dataset” is one of the most widely used datasets to provide academic venue recommendation. It contains information about 4,894,081 papers and 45,564,149 citation relationships published until April 2020.

Past Publications IEEE Data Source. Information about authors past publications are extracted from the “IEEE citation dataset”. It contains information about 11317 papers.

Academic Venues Data Source. The venue’s title, acronym, type (journal/conference), deadlines, locations and publisher are publicly available on the WikiCFP website [16]. We can distinguish five ranking categories, such as A* (flagship conference), A, B, C, and unranked (insufficient information is available to judge ranking), from the CORE Conference Portal [4].

Preferences Data Source. Each researcher has a customized interface to ensure their interaction with the recommender system. He created his own account by submitting his registration form for the first time in our proposed recommender system. After creating the account, he can log into the system by providing his email address and password. He then has the opportunity to enter or change information about his preferences, i.g., type (journal/conference), publisher, ranking and location. In addition, authors must provide a list of authors and a bibliography in order to select the appropriate venue for each written work. The interface facilitates interaction between authors and the system by allowing authors to indicate their preference for venue type, publisher, ranking, and location.

3.2 System Level Multi Domain

First Step: As explained in Sect. 2.1, multi domain uses information from both source and target domain to recommend items, for users, from both source and target domain. Although, for our multi domain academic venue recommender system, we used information, from IEEE as source domain and DBLP as target domain to recommend academic venues, for authors, from both of them (see Fig. 1). More in-depth, using information implicitly extracted from the authors past publications and the authors, from the reference list, past publications gathered from both IEEE as SD and DBLP as TD in the recommendation process.

Second Step: By reviewing the academic venue data, venues with submission deadlines out of date, are rejected. The list of recommended academic venue, obtained from Sect. 3.2, for the multi-domain, or Sect. 3.3, for linked domain approaches is refined in the same way according to the researcher’s requirements (e.g., venue’s type, publisher, rank, location) from the preferences data source.

To highlight academic venues that more closely match the author’s needs, the refinement process is performed by updating each upcoming academic venue score using a utility coefficient that describes the combination of preferences already specified in the author’s profile. Inappropriate venues based on author profiles will be rejected. By default, researchers prefer all types, publishers, rankings and locations. We assign 1 to the academic site if the type, publisher, rank

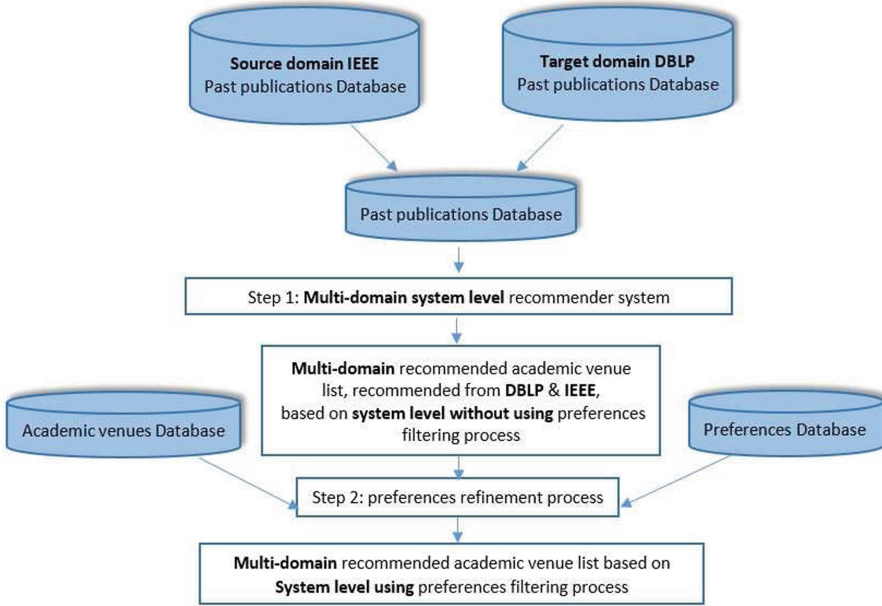


Fig. 1. System level multi domain recommender system.

and/or location of each academic site is the author’s preferred, otherwise we set it to 0. Finally, we multiply them together to get the updated final score. A set of venues is ordered by the final calculated score to be recommended.

3.3 System Level Linked Domain

First Step: As explained in Sect. 2.1, linked domain uses information from both source and target domain to recommend items, for users, from the target domain. Although, for our linked domain academic venue recommender system, we used information, from IEEE as source domain and DBLP as target domain to recommend academic venues, for authors, from the target domain (see Fig. 2). More in-depth, using information implicitly extracted from the authors past publications and the authors, from the reference list, past publications gathered from both IEEE as SD and DBLP as TD in the recommendation process.

Second Step: For the linked domain academic venue recommender system, the filtering process follows the same instructions as the multi domain academic venue recommender system, to obtain a refined academic venue list using the system level linked domain method.

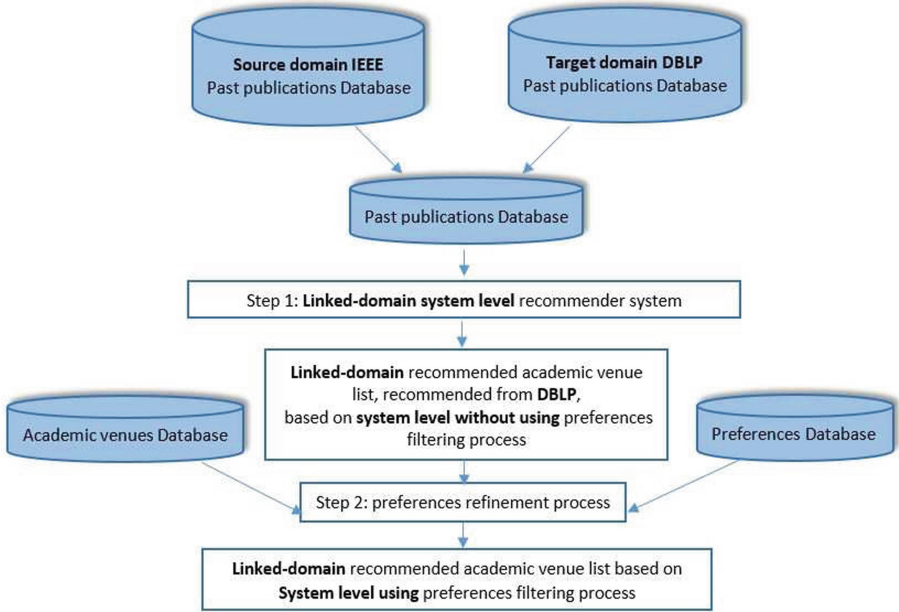


Fig. 2. System level linked domain recommender System.

4 Experimental Results and Discussions

4.1 Evaluation Datasets

To compare the proposed system level multi domain recommender system based on information extracted from IEEE and DBLP datasets versus the refined cross domain based on references recommender system [19] and the proposed system level linked domain based on references and IEEE and DBLP datasets recommendation method, experiments were conducted on the “DBLP citation dataset” and “IEEE citation dataset”, already explained in Sect. 3.

The DBLP and the IEEE citation datasets are noisy and should be cleaned up. For IEEE each paper’s information is in a separate .JSON file. We cleaned DBLP and IEEE data based on the publication with missing necessary information (authors, venue).

For the system level multi domain approach and after the dataset cleaning, a set of 13000 papers published in 2017 has been chosen randomly, from IEEE and DBLP datasets, as target papers. For each target one, a venue list will be recommended based on different subsets extracted progressively in the recommendation process, from IEEE and DBLP datasets, to count the authors past publications in 2015 and 2016 as explained in Sect. 3 to achieve approximately 550,000 papers.

For the single domain approach and after the dataset cleaning, a set of 6500 papers published in 2017 has been chosen randomly, from DBLP dataset, as target papers. For the each target one, a venue list will be recommended based on

different subsets extracted progressively in the recommendation process, from DBLP dataset, to count the authors past publications in 2015 and 2016 as explained in Sect. 3 to achieve approximately 500,000 papers.

For the system level linked domain approach and after the dataset cleaning, a set of 6500 papers published in 2017 has been chosen randomly, from DBLP dataset, as target papers. For each of them, a venue recommendation list will be recommended based on different subsets extracted progressively in the recommendation process, from IEEE and DBLP datasets, to count the authors past publications in 2015 and 2016 as explained in Sect. 3 to achieve approximately 550,000 papers.

4.2 Evaluation Metrics

Two evaluation metrics are dedicated to assessing the prediction accuracy in the context of academic venue recommendations, namely, 0/1 subset accuracy and Hamming loss.

0/1 Subset Accuracy. The 0/1 subset accuracy [7, 15] calculates the percentage of instances where the predicted set exactly matches its real set. There is no difference between a fully correct and partially correct prediction. In other words, if any of the outputs of recommender system matches the conference of the publication, it can be considered a successful recommendation. Note that the highest 0/1 subset accuracy value indicates better recommendation quality.

$$0/1 \text{ subset accuracy} = \frac{1}{N} \sum_{i=1}^N [Z_i = Y_i] \quad (1)$$

For instance, return 1 if each paper's recommended academic venues contain the true one, otherwise return 0. Note that N is the number of instances, Y_i is the true set, and Z_i is the predicted items set.

Hamming Loss. Hamming loss [7] measures the percentage of RS that makes incorrect recommendations. This metric takes into account cases where the correct academic venue is not predicted and when an incorrect academic venue is predicted. Note that the lower the Hamming loss value, the more accurate the prediction.

$$\text{Hamming loss} = \frac{1}{N} \sum_{i=1}^N [Z_i \neq Y_i] \quad (2)$$

For instance, return 0 if each paper's recommended academic venues contain the true one, otherwise return 1. Note that N is the number of instances, Y_i is the true set, and Z_i is the predicted items set.

4.3 Results and Discussions

We perform a comparative evaluation over our system level multi domain academic venue recommender system and respectively the refined cross domain academic venue recommender system [19] and the proposed system level linked domain academic venue recommender system, in order to highlight the extent to which system level Multi domain can improve the suggested academic venues results.

In fact, recommending only one academic venue to assess recommendation scores is very strict. However, we experimented on selected subsets by varying the number of top N recommended academic venues between 1 and 6 each time. For each top N , we compute the 0/1 subset accuracy and the Hamming loss for each target paper. Finally, average each of the top N results used in the experiment.

Note that the lower the Hamming loss values, the more accurate the predictions while the highest values of the 0/1 subset accuracy indicate a better recommendation quality.

The performance of the system level multi domain academic venue recommendation is compared to that of the academic venue recommender system based on refined cross domain [19] as well as the system level linked domain academic venue recommendation process. It can be seen that multi domain approach performs better than the refined cross domain and the linked domain one.

Single vs. Multi Domain Recommendation Results. It can be seen that the system level multi domain method performs better than the [19] method. For instance, for Top 5, we notice an improve of 12.34% in terms of 0/1 subset accuracy (see Fig. 3). It acquires the highest value 91.84% compared to 79.5% against [19]. Indeed, the system level multi domain approach attends the lowest average values in terms of Hamming loss, as shown in Table 1, for Top 5 (8.16% compared to 20.5%) against [20].

Table 1. Single vs. multi domain in terms of hamming loss.

Top N	Single domain recommendation based on DBLP	Multi domain recommendation based on DBLP & IEEE
Top 1	67.3	23.23
Top 2	47.8	14.27
Top 3	34.2	11.01
Top 4	26.9	9.29
Top 5	20.5	8.16
Top 6	20.5	8.16

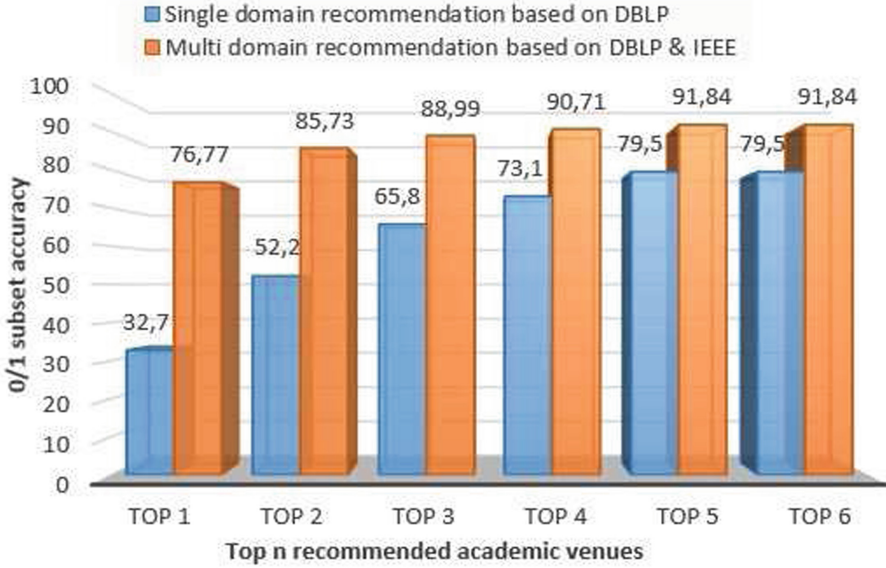


Fig. 3. Single vs. multi domain in terms of 0/1 subset accuracy.

Linked vs. Multi Domain Recommendation Results. In terms of 0/1 subset accuracy, for Top 5, we notice an improve of 13.28% (see Fig. 4), for the system level mutli domain method, to attend the highest value 91.84% compared to 78.56% for the system level linked domain approach. For Top 5. Indeed, the system level multi domain approach attends the lowest average values in terms of Hamming loss, as shown in Table 2, for Top 5 (8.16% compared to 21.44%) against the linked domain one.

Table 2. Linked vs multi domain in terms of hamming loss.

Top N	Linked domain recommendation from DBLP & IEEE to DBLP	Multi domain recommendation from DBLP & IEEE to DBLP & IEEE
Top 1	56.07	23.23
Top 2	48.38	14.27
Top 3	36.88	11.01
Top 4	28.27	9.29
Top 5	21.44	8.16
Top 6	21.44	8.16

To summarize, using information extracted from both DBLP and IEEE, to incorporate system level in the multi domain recommendation approach that joins both authors domain and references domain leads to a good predictive quality versus linked domain recommendation and single domain results.

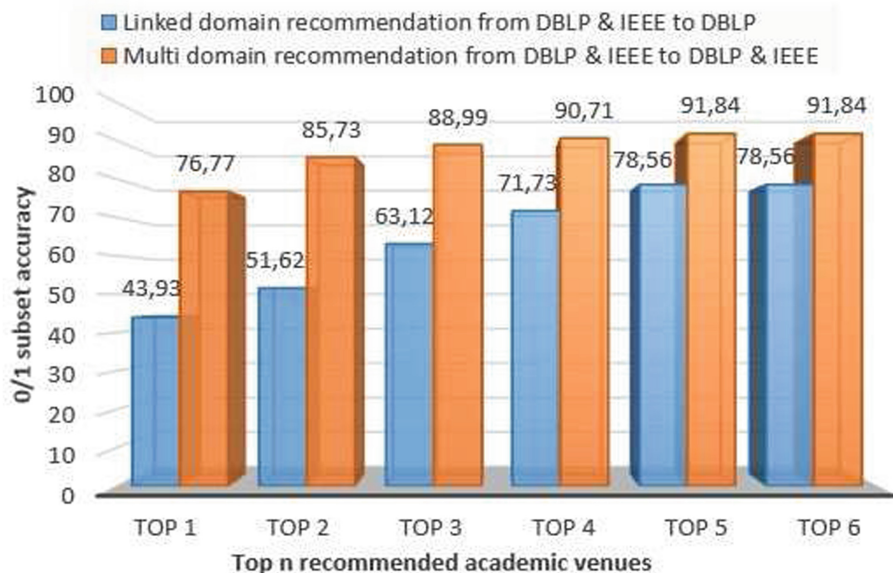


Fig. 4. Linked vs multi domain in terms of 0/1 subset accuracy.

5 Conclusion

In this paper, we compared system level multi domain recommender system, that suggests an appropriate academic venue list for the computer science researchers to publish their work, to the linked domain one. They are based on the authors from the reference list past publications aside from the target paper's authors past publications, using information extracted from IEEE and BDLP citation datasets. Our recommender engine filters out inappropriate academic venue that does not satisfy the authors preferences. It is able to deal with authors changing interests and also useful for young researchers who have not publications yet. With regard to this work, the experimental results showed that using multi domain recommendation approach can improve the author's satisfaction. As future work, we plan to integrate other author's interests, as features in the author's profile, to get more personalized recommendations (e.g., affiliation). We plan also to use other cross domain levels and integrate other domains aiming to ameliorate the recommendation results.

References

1. Aminer homepage. <https://www.aminer.org/citation/>. Accessed 14 Apr 2022
2. Boukhris, I., Ayachi, R.: A novel personalized academic venue hybrid recommender. In: 2014 IEEE 15th International Symposium on Computational Intelligence and Informatics, pp. 465–470 (2014)

3. Cantador, I., Fernández-Tobías, I., Berkovsky, S., Cremonesi, P.: Cross-domain recommender systems. In: *Recommender Systems Handbook*, pp. 919–959 (2015)
4. Core homepage. <http://portal.core.edu.au/conf-ranks/>. Accessed 14 Apr 2022
5. Cremonesi, P., Tripodi, A., Turrin, R.: Cross-domain recommender systems. In: *2011 IEEE 11th International Conference on Data Mining Workshops*, pp. 496–503 (2011)
6. Fernández-Tobías, I., Cantador, I., Kaminskas, M., Ricci, F.: Cross-domain recommender systems: a survey of the state of the art. In: *Spanish Conference on Information Retrieval*, vol. 24 (2012)
7. Gibaja, E., Ventura, S.: A tutorial on multilabel learning. *ACM Comput. Surv.* **47**(3), 1–38 (2015)
8. Kang, N., Doornenbal, M.A., Schijvenaars, R.J.: Elsevier journal finder: recommending journals for your paper. In: *Proceedings of the 9th ACM Conference on Recommender Systems*, pp. 261–264 (2015)
9. Khan, M.M., Ibrahim, R., Ghani, I.: Cross domain recommender systems: a systematic literature review. *ACM Comput. Surv.* **50**(3), 1–34 (2017)
10. Klamma, R., Cuong, P.M., Cao, Y.: You never walk alone: Recommending academic events based on social network analysis. In: *International Conference on Complex Sciences*, pp. 657–670 (2009)
11. Küçüktunç, O., Saule, E., Kaya, K., Çatalyürek, Ü.V.: TheAdvisor: a webservice for academic recommendation. In: *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 433–434 (2013)
12. Mhirsi, N., Boukhris, I.: Exploring location and ranking for academic venue recommendation. In: *International Conference on Intelligent Systems Design and Applications*, pp. 83–91 (2017)
13. Sato, R., Yamada, M., Kashima, H.: Poincare: recommending publication venues via treatment effect estimation. *J. Informetrics* **16**(2), 101–283 (2022)
14. Sinha, A., et al.: An overview of Microsoft academic service (MAS) and applications. In: *Proceedings of the 24th International Conference on World Wide Web*, pp. 243–246 (2015)
15. Wang, D., Liang, Y., Xu, D., Feng, X., Guan, R.: A content-based recommender system for computer science publications. *Knowl. Based Syst.* **157**, 1–9 (2018)
16. Wikicfp homepage. <http://www.wikicfp.com/cfp/>. Accessed 14 Apr 2022
17. Yang, Z., Davison, B.D.: Venue recommendation: submitting your paper with style. In: *2012 11th International Conference on Machine Learning and Applications*, vol. 1, pp. 681–686 (2012)
18. Zawali, A., Boukhris, I.: A group recommender system for academic venue personalization. In: *International Conference on Intelligent Systems Design and Applications*, pp. 597–606 (2018)
19. Zawali, A., Boukhris, I.: Academic venue recommendation based on refined cross domain. In: *International Conference on Intelligent Systems Design and Applications*, pp. 1188–1197 (2022)
20. Zawali, A., Boukhris, I.: Cross domain collaborative filtering recommender system for academic venue personalization based on references. In: *2020 IEEE Symposium Series on Computational Intelligence*, pp. 2829–2835 (2020)
21. ZhengWei, H., JinTao, M., YanNi, Y., Jin, H., Ye, T.: Recommendation method for academic journal submission based on doc2vec and XGBoost. *Scientometrics* **127**, 1–14 (2022)