Venue Recommendation Based on Paper's Title and Co-authors Network

Ramin Safa
Department of Engineering, University of Guilan, Guilan, Iran
r.safa@outlook.com
SeyedAbolghasem Mirroshandel*
Department of Engineering ,University of Guilan, Guilan, Iran
mirroshandel@gmail.com
Soroush Javadi
Department of Engineering , University of Guilan, Guilan, Iran
soroush.javadi@gmail.com
Mohammad Azizi
Department of Engineering , University of Guilan, Guilan, Iran
Moh.Azizi@gmail.com

Received: 22/Jul/2017 Revised: 16/Sep/2017 Accepted: 24/Dec/2017

Abstract

Information overload has always been a remarkable topic in scientific researches, and one of the available approaches in this field is employing recommender systems. With the spread of these systems in various fields, studies show the need for more attention to applying them in scientific applications. Applying recommender systems to scientific domain, such as paper recommendation, expert recommendation, citation recommendation and reviewer recommendation, are new and developing topics. With the significant growth of the number of scientific events and journals, one of the most important issues is choosing the most suitable venue for publishing papers, and the existence of a tool to accelerate this process is necessary for researchers. Despite the importance of these systems in accelerating the publication process and decreasing possible errors, this problem has been less studied in related works. So in this paper, an efficient approach will be suggested for recommending related conferences or journals for a researcher's specific paper. In other words, our system will be able to recommend the most suitable venues for publishing a written paper, by means of social network analysis and content-based filtering, according to the researcher's preferences and the co-authors' publication history. We used the minimum available free features and the minimum implementing facilities, which to the best of our knowledge have not seen up to now. In addition, it can be argued that the proposed system overcome the cold start problem which has always been a remarkable task in recommender systems. The results of evaluation using real-world data show acceptable accuracy in venue recommendations.

Keywords: Academic Recommender Systems; Social Network Analysis; Publication Venue Recommendation; DBLP.

1. Introduction

Recommender systems include software tools and techniques which recommend the most appropriate options by using different kinds of knowledge, user-related data, existing items and previous transactions. They will enable users to achieve their goal more quickly in a large amount of information [1]. Generally, there are three kinds of methods for recommendation: collaborative filtering, content-based filtering, and hybrid systems.

Briefly, in the collaborative filtering, the system identifies the items which may be interesting to a user by taking advantage of previous behaviors and finding similar rating patterns. In the other hand, in content-based filtering, a model of user preferences is created according to the features of the items. In this method, by identifying items similar to which the user liked contently before, and matching user's profile and items' features, the system presents recommendations [1–4]. However, both of these approaches have some issues. For example, data availability and data quality have always been significant subjects in collaborative filtering, and syntactic issues and compound nouns are important topics in content-based filtering. So, in hybrid systems, the quality of recommendations is improved by using the advantages of both aforementioned methods.

Recently, lots of attention have been paid to utilizing the information of users' social network and the existing relations for customization. There is also a growing trend in research on the use of recommender systems in social networks, especially in scientific environments [5]. Scientific social networks are resources that include relations among researcher, publications and bibliographic information which help knowledge development by sharing scientific publications. The large number of scientific bases, papers, and fields of research reveal the importance of using recommender systems and content personalization [6–8].

* Corresponding Author

In online scientific communities, applying recommender systems is observed in fields such as paper recommendation [9–14], expert recommendation [8,15], reviewer recommendation [16,17] and citation recommendation [18,19]. Due to the rapid growth of the number of the scientific events, especially in computer science [20], few attempts are done towards venue recommendation for publishing a new paper.

Searching for a conference or journal whose scope matches a new paper's topic can be difficult and time-consuming, and will not always lead to desirable results. This can motivate us to employ recommender systems, as the venues suitable for publishing a paper can be extracted by recognizing the researcher's preferences and applying the information filtering process. This could help the researchers, especially those who have no enough experience to choose the most appropriate venues from a lot of conferences and journals.

Despite the fact that some online publishers like Elsevier[1] and Springer[2] try to recommend their related journals by asking user to provide some information about written paper, the comprehensive system with the ability of recommending effective conferences and journals (taking diversity into account) has not been found in practice. In addition, small number of studies conducted in this area are mostly assuming that all of the information for implementing their approach are always available, which regardless of data gathering and matching cost issues, it cannot to be mapped to the real-world and seems to be not applicable. In this paper we present an approach for venues recommendation based on paper's title and co-authors network. As bibliographic information of papers and publication metadata such as title, author(s), year, and venue are freely available, can be used to compute similarity measures and find related documents. We also believe that people who are close to each other have the same taste. As a result, the publications of a co-author can be an informative clue for recommending scientific venues.

The rest of this paper is structured as follows: in the next section, we briefly survey the related work on venue recommendation. In Section 3, we present and detail our approach to recommendation using paper's title and co-authors network. Section 4 will be about another approach—singular value decomposition. In Section 5, we describe the experimental setup and discuss the results in detail. Finally, the paper will be concluded with a summary and future work in Section 6.

2. Related Work

Recommender systems help encounter problems resulted from the explosive growth of information and facilitate decision-making and selection based on user's interest through information filtering process [1]. It is proven that recommender systems are useful and valuable tools to encounter the problems resulted from information overload for online users. Nowadays, recommender systems are used in different fields—e.g., e-commerce, news, and entertainment—but researches show a high demand for utilizing these systems in scientific domains [6,21].

Social relations of people can influence their behavior and interests, hence we observe utilizing social network capabilities in different domains, and social network analysis in recommender systems is a new emerging topic [8,21]. By combining traditional methods of recommender systems and social network analysis, more effective recommendations can be achieved. Social interactions in scientific communities—such as co-authorship and participation in similar conferences—can influence recommendation quality.

Klamma et al [20] recommended conference venues to researchers by utilizing collaborative filtering concepts. Their system uses DBLP[3] and Eventseer.net information about venues. It extracts some useful information about individuals who participated in similar conferences to those the user participated in, to recommend related scientific events. The content-based approach is not employed in this research. Also, recommendations presented to a user are general, and not specific to a written paper.

It should be noted that in scientific communities, the semantic relations between papers and their publication venues are considered important, and collaborative filtering will not be able to extract these relations. This approach only takes interactions between users and items into account [12].

Martín et al [22] proposed a content-based filtering algorithm. Their algorithm uses textual information of call for papers and recent paper abstracts of each conference program committee member, and also the abstracts of the user's recent papers and their citation information.

Medvet et al [23] suggested a system which used paper title and abstract for recommending a publication venue. They extract conference papers in computer science from Microsoft Academic Search[4] engine, which most have the necessary features, and by matching the title and abstract of the user's new paper with conference selected papers try to recommend appropriate venue to the user.

Xu et al [24] studied a comprehensive system that covers all aspects of paper life cycle. In this work, conference and journal recommendation is mentioned as the most important part of the system. Their proposed system extracts the keywords from context, and then, recommendation process will be a subject-oriented query. It should be noted that their system does not utilize co-authors network and social network analysis.

To the best of our knowledge, utilizing authors social network for publication venue recommendation has been introduced by Luong et al [7] for the first time. Three methods are presented in their research to recommend

[1] Available from: http://journalfinder.elsevier.com
[2] Available from: https://journalsuggester.springer.com
[3] Available from: http://dblp.uni-trier.de/db
[4] Available from: http://academic.research.microsoft.com

relevant publication venues using social network analysis: (1) most frequent conference, (2) most frequent conference normalized by author, and (3) the second method combined with network topology. In the first method, using co-authors social network information, the conference in which these people had published the most number of papers will be recommended. The second method utilizes normalization in order to decrease the influence of the authors who wrote the most number of papers on the final results. In the third method, the weight of authors that have more previous collaborations with the main author will be considered more important. Results show the superiority of the network-based approach compared to the content-based one. It should be noted that paper content is not used in their work.

Beierle et al [25] identified six main ways for extracting relation of two authors in academic social graph, through common publications (co-authorship), affiliations, similar keywords (co-interests), commonly visited venues (co-activity), referencing or being referenced by the other. Based on their research, social relations can be used to derive author's preferences and exploited for conference recommendations. Furthermore, they found co-authorship, co-interests, and co-activity lead to the best recommendation accuracy. This is almost the same point mentioned by García et al [26].

We can also take a brief look at Peiris and Weerasinghe [27] work, who proposed an approach for ranking publication venues by considering publication history and citation network. They believe there are some aspects that contribute the importance of a publication, including citation it has received, the quality of the citing publications, the time metric and its authors.

In this paper, we present an approach to recommend related venues for a user's certain paper, by employing social network analysis concepts and content-based filtering. Experimental results using real-world data show that our approach can provide effective recommendations.

3. Proposed Method

Scientific recommendations are often done using content-based filtering [13] and based on paper content. It is necessary to mention that obtaining the full-text version of papers is usually not possible due to copyright issues, but bibliographic information of papers and publication metadata such as title, author(s), year, and venue, which can be a useful source of information, are freely available, and can be used to compute similarity measures and find related documents.

Generally, deriving a user's interests can be done in two ways: explicit and implicit. In the former, the user declares his/her preferences explicitly, while in the latter, the user's preferences are identified by monitoring and analyzing his/her activities [1,28]. In our studies, it is observed that some paper retrieval systems [13,29,30] and deriving users' profile algorithms [31] have used words appearing in a user's publication title as his/her interests.

On the other hand, it is proven in different researches that people tend to accept recommendations that are presented by people around them, rather than those who are far from them but have similar tastes. This opens many research opportunities subsumed under social recommender systems [1,28,32]. In scientific domain, studying models and methods presented in various researches, it can be concluded that utilizing the information about people around a user increases recommendation accuracy [7,14,15,33].

The main idea is that we can use similarity measures with co-authors' published papers and the target paper, to recommend related conferences and journals for publishing the paper.

As shown in figure 1, our recommender system takes author(s) identity and the new paper's title as input. It finds similar papers of co-authors by matching their titles with the input title. The venues of extracted papers will be ranked and recommended. This procedure can be done recursively for co-authors of co-authors and so on.

Figure 1 shows the architecture of our system which mainly consists of three components: linguistic information extractor, matcher, and ranker.

3.1 Linguistic Information Extractor

Documents should be changed into a structured form to be interpretable for the system in order to apply our similarity measure. In natural language processing, there is a procedure called stop word removal, which is done to remove the most frequent used words. In this component, a definite list of stop words taken from MySQL website[1] is removed from each paper's title. It is necessary to mention that to enhance the results, words "a", "i" and "based" are added to this list.

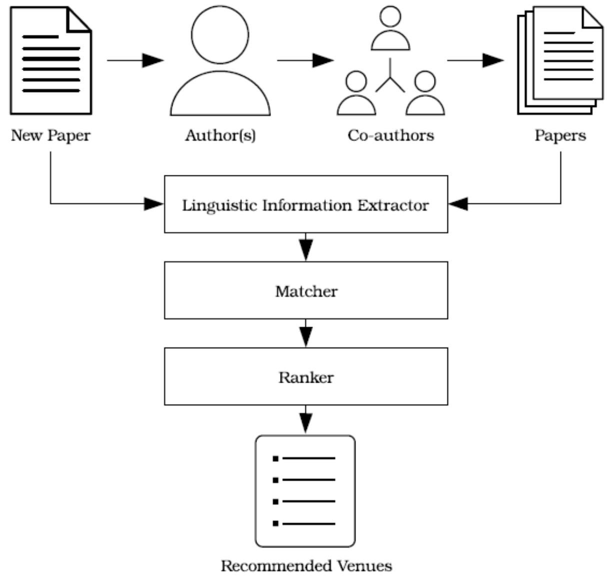

Fig. 1. Proposed method overview

In language morphology and information retrieval, there is another process named stemming, that aims towards reducing a word to its root, by removing some

[1] Available from: http://bit.ly/mysqlstopwords

parts of the word (the affixes). Applying stemming is very important in matching process [34].

After removing stop words, stemming is done by the Porter2 stemmer[1], which is the improved version of the Porter stemmer with minor modifications. Porter is one of the most well-known tools for stemming and it is commonly used for the English language due to its high performance. Some of Porter's rules are as follows [35]:

- sses → ss
- ies → i
- s → *(null)*
- izer, ization → ize
- ator, ation, ational, → ate

In other words, after taking papers' titles, stop word removal is applied to remove uninformative words, and then stemming is done by the Porter2 stemmer in order to increase matching accuracy by reducing the diversity of words.

3.2 Matcher

In order to apply matching and similarity measures, each document is represented by a vector of term weights (i.e., term frequencies). Next, cosine similarity is used, which is one of the most appropriate similarity measures. This measure is defined as follows, where $A \cdot B$ denotes the dot product of the vectors A and B [1]:

$$\mathrm{similarity}(\mathrm{A},\mathrm{B}) = \cos\theta = \frac{\mathrm{A}\cdot\mathrm{B}}{\|\mathrm{A}\|\|\mathrm{B}\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2} \times \sqrt{\sum_{i=1}^{n}(B_i)^2}} \quad (1)$$

3.3 Ranker

After identifying the venues of papers similar to the input paper, weighting is done in order to identify the most relevant venues at ranker component as follows, where P_i is the i^{th} paper published in the venue V, and R is the new paper:

$$\mathrm{weight}(\mathrm{V}) = \sum_{i=1}^{n} \mathrm{similarity}(\mathrm{P_i},\mathrm{R}) \quad (2)$$

Finally, venues are sorted based on their weights and presented to the user.

4. SVD Approach

Singular value decomposition (SVD) is one of the dimensionality reduction techniques which factors an *item × features* matrix *A* into three different matrices: an *item × concepts*, a *concept strength*, and a *concept × features* as the following:

$$\mathrm{A} = \mathrm{U}\lambda\mathrm{V}^{\mathrm{T}} \quad (3)$$

The most well-known application of SVD in natural language processing is latent semantic analysis (LSA), which is a theory and method for extracting and representing the meaning of words by statistical computations applied to a large corpus of text [36]. LSA can work with a term-document matrix which describes the occurrence of terms in documents. Since it can be used as a prediction tool [37], we considered applying it to our problem as an alternative method.

We utilized SVD to capture latent relationships between terms and venues which allow us to compute our proposed matching algorithm in a different space. To achieve this goal, we started with a term-venue matrix wherein each column represents one of the venues to be ranked, and terms are the words of the preprocessed titles of all published papers in those venues in recent years. Each matrix entry indicates the frequency of the corresponding term in the corresponding venue. Table 1 is an example of term-document matrix for ten venues and a limited number of their terms.

Table 1. An example of term-document matrix

Venue / Keyword	CAMAD	EUNICE	HAISA	HPCC-ICESS	IJESMA	ISCA	KMIS	NMR	SPRINGL	SSV
algorithm	2	8	0	24	0	5	0	2	1	1
cellular	2	1	0	1	0	0	0	0	0	0
game	1	1	0	1	0	0	1	1	0	0
hardwar	1	0	1	4	0	18	0	0	1	0
internet	2	6	2	0	2	0	0	0	0	0
mobil	10	8	0	6	17	5	2	0	2	0
network	58	60	4	38	2	25	12	0	3	0
search	0	1	0	1	2	4	1	0	0	0
secur	4	4	29	5	1	12	3	0	4	0
web	0	2	0	3	3	1	13	0	2	0

We computed the singular value decomposition of the term-venue matrix and extracted the singular values, left singular vectors and right singular values from the SVD matrix. Next, we considered the search component of latent semantic indexing. Queries are computed by taking the centroid of the term vectors corresponding to the terms in the query, where the query is the preprocessed title of the new paper. The centroid is computed pointwise, by adding the values in each dimension. This is then matched against the venue vectors using the scaling provided by the singular values, resulting in a score (weight) for each venue. We used dot product and cosine similarity methods for matching.

5. Evaluation and Discussion

The common way to measure the performance of the proposed method is to compare the results of applying it to a standardized dataset with the results of other researchers. But as noted earlier, the number of tasks performed in this field is not high and the datasets of the related works are not available. In the other hand, we received some dataset from the related works, but they do not have all the features and metadata we intended to use and were customized for specific work. Due to different datasets, the comparison of existing methods with our method is not a valid and meaningful comparison.

[1] Available from: http://bit.ly/porter2stemmer

Therefore, we decided to use real-world data and a strict approach to measure the method performance.

We used DBLP data to evaluate the recommendation method. DBLP is the computer science bibliographic website that provides bibliographic information about papers, events and computer science journals. The dataset of this website is saved periodically in an XML file[1] and the last update of this file at the time of evaluation (May 29, 2014) is used to measure the accuracy of the system. This dataset contains eight types of entities which are shown in table 2. As it is described in the table, conference and journal papers are the two major types in this dataset.

Table 2. Types of DBLP entities

Type	Description	Count
Article	Journal Article	1,131,735
Book	Book	10,932
In Collection	Publication Cited in a Collection	26,739
In Proceedings	Publication Published in Conference Proceedings	1,431,399
Master's Thesis	Master's Thesis	9
Ph.D. Thesis	Ph.D. Thesis	6,937
Proceedings	Conference Proceedings	23,146
WWW	Author Links	1,412,090

Each entity contains some of these metadata, among which *title* is the only one that has to exist [38]: *author*, *editor*, *title*, *booktitle*, *pages*, *year*, *address*, *journal*, *volume*, *number*, *month*, *url*, *ee* (electronic edition), *cdrom*, *cite*, *publisher*, *note*, *crossref*, *isbn*, *doi*, *series*, *school* and *chapter*.

In order to evaluate suggested method, we selected a random sample of 20,000 papers among a total of 205,880 papers published in 2013. We used the papers published from 2008 to 2012 as recent papers. Also, to obtain the co-authors of an author, we used the metadata of papers published from 2003 to 2012.

It is necessary to mention that there were some challenges during the preparation of the dataset. In many cases, we observed that the same venue has different names. This is sometimes due to typographical errors, sometimes part of the name is removed or added, and sometimes it is displaced.

Some examples of typographical errors:

- Internet Measurement Conference,
- Internet Measurement Comference, and
- Internet Measurment Conference;
- Computer Supported Activity Coordination, and
- Computer Supported Acitivity Coordination;
- Adaptive Agents and Multi-Agent Systems, and
- Adaptive Agents and Multi-Agents Systems.

Some examples of removal or addition of part of conference name:

- IEEE VAST, and
- VAST;
- GI Jahrestagung, and
- GI-Jahrestagung;
- ICWS, and
- ICWS-Europe.

An example of displacement:

- KR4HC/ProHealth, and
- ProHealth/KR4HC.

Moreover, for venues in whose names there was an at-sign character, we used the part after the at-sign character as the venue name. For example, DUBMOD@CIKM papers were considered as CIKM papers. Also, in a few cases, the paper title contained LaTeX codes, which we ignored.

Among the total 5,865 distinct venues, 611 cases had the aforementioned problems. Venue names containing an at-sign where trimmed automatically. For other cases, we used regular expression search to find them, and corrected them manually. In addition, 164 documents were removed after linguistic preprocessing phase, because no word remained in their title.

After applying the proposed method with depth 1 traversal for co-authors, we observed that in 70.69% of cases, the accurate venue—the venue in which the input paper was actually published—does exist in the system output, and for the top-20 recommendations, accuracy reaches to 48.53%. This seems interesting, because the evaluation was done on a completely random set of DBLP data, and we did not select specific papers with predetermined domains. On the other hand, the average number of total venues reaching the ranker component of our system was 350 venues for each input paper, and the top-20 list of recommendations contains only less than 6% of these possible recommendations. We used the term "Oracle score" in our results to show accountability of the system, which refers to existence of exact venue in total recommendation list of each depth.

For measuring the quality of our recommendations, we could ask human experts to evaluate the output of our system. However, evaluation by human experts is very time-consuming, expensive, and human-dependent. For resolving these issues, we have utilized accuracy measure for evaluation. In our recommender system, accuracy is even stricter measure than evaluation by human experts. In accuracy measure, if the recommended venue is exactly equal to the real published venue, system achieve a score, otherwise the answer will be regarded as unrelated. However, in our observation, lots of recommendation were totally related to the paper, but accuracy did not score these recommendations. We should again emphasis that the satisfaction of researchers (human experts) is the best way for evaluation of the system. In this situation that it is not feasible to evaluate using the mentioned measure, we used accuracy measure.

Co-authors network can be represented as a graph containing authors as nodes and co-author relationship as edges. In the process of evaluation, we learned that the first recommendation in depth 1 of the graph, with an accuracy of 15.18%, has the best accuracy in a recommendation list. The results of the evaluation for recommendation number 1 to 20 can be seen in table 3.

Table 3. Evaluation results

Recommendation no.	No. of accurate recommendations	Accuracy (%)
Only 1st	3,036	15.18

[1] Available from: http://dblp.uni-trier.de/xml

Only 5th	470	2.35
Only 10th	256	1.28
Only 15th	132	0.66
Only 20th	119	0.59

Table 4 shows system oracle score for depth 1 to 4, that depth 4 has a significant accountability of 88.14%.

Table 4. The system oracle score for depths 1 to 4

Depth	Oracle count	Oracle score (%)
1	14,139	70.69
2	16,361	81.81
3	17,323	86.61
4	17,627	88.14

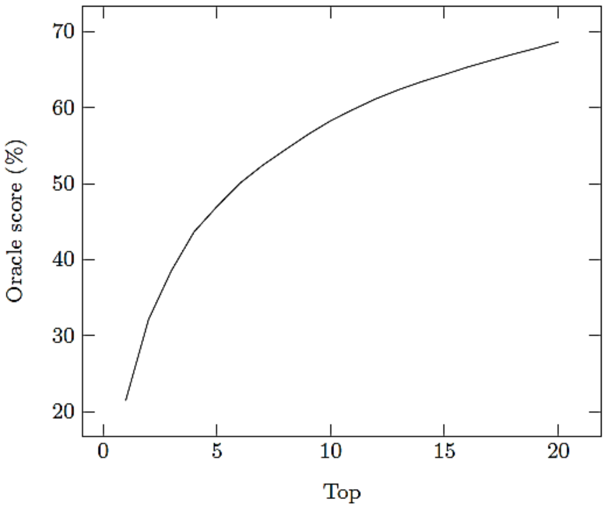

Fig. 2. The oracle score in different tops for depth 1

Tables 5 and 6 show the number of accurate recommendations, the accuracy and oracle score for depth 1 and 2 traversals in different tops.

Table 5. Details of depth 1 evaluation

Recommendations	No. of accurate recommendations	Accuracy (%)	Oracle score (%)
Top-1	3,036	15.18	21.47
Top-2	4,540	22.70	32.11
Top-3	5,447	27.23	38.52
Top-4	6,171	30.86	43.65
Top-5	6,641	33.20	46.97
Top-6	7,073	35.37	50.02
Top-7	7,408	37.04	52.39
Top-8	7,699	38.49	54.45
Top-9	7,984	39.92	56.47
Top-10	8,240	41.20	58.28
Top-11	8,452	42.26	59.78
Top-12	8,652	43.26	61.19
Top-13	8,820	44.10	62.38
Top-14	8,968	44.84	63.43
Top-15	9,100	45.50	64.36
Top-16	9,237	46.19	65.33
Top-17	9,358	46.79	66.19
Top-18	9,478	47.39	67.03
Top-19	9,587	47.94	67.81
Top-20	9,706	48.53	68.65

Evaluation results show that the recommendation accuracy decreases as the traversal depth for co-authors increases. The reason is that, as the depth increases, co-authors grow exponentially in number, but become less relevant. So, even though co-authors of co-authors are obviously less relevant than the co-authors, they are much greater in quantity, hence affect the system accuracy negatively. On the other hand, this decrease in accuracy is accompanied by an increase in the oracle score of the system. It seems that the recommendation accuracy can be enhanced by assigning a proper weight to each depth in order to utilize this accountability in the future.

Table 6. Details of depth 2 evaluation

Recommendations	No. of accurate recommendations	Accuracy (%)	Oracle score (%)
Top-1	2,611	13.05	15.96
Top-2	3,878	19.39	23.70
Top-3	4,696	23.48	28.70
Top-4	5,313	26.57	32.47
Top-5	5,836	29.18	35.67
Top-6	6,296	31.48	38.48
Top-7	6,670	33.35	40.77
Top-8	6,967	34.84	42.58
Top-9	7,235	36.17	44.22
Top-10	7,526	37.63	46.00
Top-11	7,770	38.85	47.49
Top-12	8,002	40.01	48.91
Top-13	8,203	41.02	50.14
Top-14	8,396	41.98	51.32
Top-15	8,558	42.79	52.31
Top-16	8,734	43.67	53.38
Top-17	8,899	44.49	54.39
Top-18	9,040	45.20	55.25
Top-19	9,175	45.88	56.08
Top-20	9,287	46.44	56.76

However, we use the information of author's own network, one may ask about the percentage of new venues in our proposed search space and finally in our recommendations. The answer of this question is shown in table 7. As it is visible in this table, the average number of extracted venues from depth one and two of co-authors network are 349.83 and 1,619.89, respectively. These large numbers show the comprehensiveness of search space. Another interesting column in table 6 is the average percentage of new venues, for the target author, in our top-20 recommendation. Again, this percentage is quite high (i.e., about 80%) and it shows our algorithm is able to discover new venues for recommendation to the user.

Table 7. average number of extracted venues and the average percentage of unfamiliar venues in top-20

Depth	Average no. of extracted venues	Average percentage of new venues in top-20
1	349.83	78.96%
2	1,619.89	83.25%

5.1 SVD Results

Tests of this approach are done on the same dataset. In all tests, cooperation and recent papers are extracted respectively from ten and five recent years. For calculating the SVD, the number of factors (latent semantic dimensions) is restricted in each test.

First, a test in depth 1 is applied to the dataset, with 20 factors and both dot product and cosine and methods. The accuracy of top-20 recommendations for the dot product method was 19.75%, and for the cosine method, was 20.45%. So the cosine similarity method was slightly more accurate.

Increasing the number of factors to 60 in the cosine method improves the accuracy of top-20 recommendations to 22.34%, which is 1.89% better than the same test with 20 factors. The result of this test is shown in table 8.

Table 8. SVD test results in depth 1, with 60 factors and the cosine similarity method

Recommendations	Accuracy (%)
Top-1	4.28
Top-2	6.76
Top-3	8.66
Top-4	10.28
Top-5	11.60
Top-6	12.86
Top-7	13.84
Top-8	14.76
Top-9	15.58
Top-10	16.38
Top-11	17.05
Top-12	17.80
Top-13	18.53
Top-14	19.15
Top-15	19.79
Top-16	20.34
Top-17	20.84
Top-18	21.39
Top-19	21.91
Top-20	22.34

In depth 2, with 20 factors and the cosine method, the accuracy of top-20 recommendations was 10.58%, which is 9.87% less than that of the same test in depth 1.

6. Conclusion and Future Work

The goal of this paper is to design and apply an algorithm to recommend appropriate venues to researchers for publishing scientific papers, by utilizing the title of papers and the co-authors network. We evaluated the proposed method with real-world data from the DBLP computer science bibliography website, and some explanations are presented about the preparation of the dataset and the challenges we encountered. The results of the evaluation show that our method is able to present effective recommendations only by publication metadata and minimum implementing facilities: with depth 1 traversal of the co-authors network, the oracle score of the system was 70.69% and the accuracy was 48.53% for the top-20 recommendations. In this evaluation, it was cleared that system responsibility increases as the traversal depth increases, but with weight tuning for each depth, the accuracy reduction should be prevented.

Our system depends on previous publications of researchers to provide them with venue recommendations, but since novice researchers usually get help from experts in the field of research, it can be argued that the proposed system resolves the cold start problem. Moreover, the combination of the suggested approach with data mining techniques and machine learning algorithms to create a model to find proper venues can be an interesting topic.

To test another approach, we used SVD, and by applying this method to the same dataset with 60 factors, we found out that our suggested system is by far more efficient. However, increasing the number of factors may enhance the results, but is not economic.

We will focus on using more advanced methods for extracting keywords, in order to improve the matching process. Also, we aim to utilize some additional information related to venues, e.g., deadlines and locations. Bibliographic information also contains valuable data. For example, in some researches, citation information is considered a useful clue. We will also expect to launch this method as a web-based application to help researchers.

References

[1] F. Ricci, L. Rokach and B. Shapira (2011). Introduction to recommender systems handbook. Springer.

[2] D. Jannach, M. Zanker, A. Felfernig and G. Friedrich (2010). Recommender systems: An introduction. Cambridge University Press.

[3] J. Bobadilla, F. Ortega, A. Hernando and A. Gutiérrez (2013). Recommender systems survey. Knowledge-Based Systems 46:109–132.

[4] X. Su and T. M. Khoshgoftaar (2009). A survey of collaborative filtering techniques. Advances in Artificial Intelligence 2009:4.

[5] J. Beel, B. Gipp, S. Langer and C. Breitinger (2016). Paper recommender systems: A literature survey. International Journal on Digital Libraries 17(4):305–338.

[6] C. Pan and W. Li (2010). Research paper recommendation with topic analysis. In: Computer Design and Applications (ICCDA), 2010 International Conference on, IEEE, vol 4, pp V4–264.

[7] H. Luong, T. Huynh, S. Gauch, L. Do and K. Hoang (2012). Publication venue recommendation using author network's publication history. In: Asian Conference on Intelligent Information and Database Systems, Springer, pp 426–435.

[8] T. Huynh, K. Hoang and D. Lam (2013). Trend based vertex similarity for academic collaboration recommendation. In: International Conference on Computational Collective Intelligence, Springer, pp 11–20.

[9] C. Bancu, M. Dagadita, M. Dascalu, C. Dobre, S. Trausan-Matu and A. M. Florea (2012). ARSYS–article recommender system. In: Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), 2012 14th International Symposium on, IEEE, pp 349–355.

[10] M. Gori and A. Pucci (2006). Research paper recommender systems: A random-walk based approach. In: 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings) (WI'06), IEEE, pp 778–781.

[11] Y. Jiang, A. Jia, Y. Feng and D. Zhao (2012). Recommending academic papers via users' reading

purposes. In: Proceedings of the sixth ACM conference on Recommender systems, ACM, pp 241–244.

[12] J. Sun, J. Ma, X. Liu, Z. Liu, G. Wang, H. Jiang and T. Silva (2013). A novel approach for personalized article recommendation in online scientific communities. In: System Sciences (HICSS), 2013 46th Hawaii International Conference on, IEEE, pp 1543–1552.

[13] G. Cabanac (2011). Accuracy of inter-researcher similarity measures based on topical and social clues. Scientometrics 87(3):597–620.

[14] M. S. Pera and Y. K. Ng (2014). Exploiting the wisdom of social connections to make personalized recommendations on scholarly articles. Journal of Intelligent Information Systems 42(3):371–391.

[15] Z. Zhan, L. Yang, S. Bao, D. Han, Z. Su and Y. Yu (2011). Finding appropriate experts for collaboration. In: International Conference on Web-Age Information Management, Springer, pp 327–339.

[16] C. Basu, H. Hirsh, W. W. Cohen and C. Nevill-Manning (2001). Technical paper recommendation: A study in combining multiple information sources. Journal of Artificial Intelligence Research 14:241–262.

[17] D. Conry, Y. Koren and N. Ramakrishnan (2009). Recommender systems for the conference paper assignment problem. In: Proceedings of the third ACM conference on Recommender systems, ACM, pp 357–360.

[18] B. Gipp and J. Beel (2009). Citation proximity analysis (CPA)–a new approach for identifying related work based on co-citation analysis. In: Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09), Rio de Janeiro (Brazil): International Society for Scientometrics and Informetrics, vol 2, pp 571–575.

[19] C. Caragea, A. Silvescu, P. Mitra and C. L. Giles (2013). Can't see the forest for the trees?: A citation recommendation system. In: Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries, ACM, pp 111–114.

[20] R. Klamma, P. M. Cuong and Y. Cao (2009). You never walk alone: Recommending academic events based on social network analysis. In: International Conference on Complex Sciences, Springer, pp 657–670.

[21] D. H. Park, H. K. Kim, I. Y. Choi and J. K. Kim (2012). A literature review and classification of recommender systems research. Expert Systems with Applications 39(11):10,059–10,072.

[22] G. H. Martín, S. Schockaert, C. Cornelis and H. Naessens (2013). An exploratory study on content-based filtering of call for papers. In: Information Retrieval Facility Conference, Springer, pp 58–69.

[23] E. Medvet, A. Bartoli and G. Piccinin (2014). Publication venue recommendation based on paper abstract. In: Tools with Artificial Intelligence (ICTAI), 2014 IEEE 26th International Conference on, IEEE, pp 1004–1010.

[24] Z. Xu, Y. Yang, F. Wang, J. Xu, Z. Li, F. Mu and L. Li (2014). A recommendation system for paper submission based on vertical search engine. In: Computer Engineering and Networking, Springer, pp 201–208.

[25] F. Beierle, J. Tan and K. Grunert (2016). Analyzing social relations for recommending academic conferences. In: Proceedings of the 8th ACM International Workshop on Hot Topics in Planet-scale Mobile Computing and Online Social Networking, ACM, pp 37–42.

[26] G. M. García, B. P. Nunes, G. R. Lopes, M. A. Casanova and L. A. P. P. Leme (2016). Comparing and recommending conferences. In: Proceedings of the 5th Bra, SNAM–Brazilian Workshop on Social Network Analysis and Mining.

[27] D. Peiris and R. Weerasinghe (2015). Citation network based framework for ranking academic publications and venues. In: Advances in ICT for Emerging Regions (ICTer), 2015 Fifteenth International Conference on, IEEE, pp 146–151.

[28] S. Perugini, M. A. Gonçalves and E. A. Fox (2004). Recommender systems research: A connection-centric survey. Journal of Intelligent Information Systems 23(2):107–143.

[29] X. Ren, Y. Zeng, Y. Qin, N. Zhong, Z. Huang, Y. Wang and C. Wang (2010). Social relation based search refinement: Let your friends help you! In: International Conference on Active Media Technology, Springer, pp 475–485.

[30] Y. Zeng, Y. Yao and N. Zhong (2009). DBLP-SSE: A DBLP search support engine. In: Web Intelligence and Intelligent Agent Technologies (WI-IAT'09), IEEE/WIC/ACM International Joint Conferences on, IET, vol 1, pp 626–630.

[31] D. Tchuente, M. F. Canut, N. Jessel, A. Peninou and F. Sèdes (2013). A community-based algorithm for deriving users' profiles from egocentrics networks: Experiment on Facebook and DBLP. Social Network Analysis and Mining 3(3):667–683.

[32] M. Eirinaki, J. Gao, I. Varlamis and K. Tserpes (2018). Recommender systems for large-scale social networks: A review of challenges and solutions.

[33] H. C. Wang and Y. L. Chang (2007). PKR: A personalized knowledge recommendation system for virtual research communities. Journal of Computer Information Systems 48(1):31–41.

[34] J. B. Lovins (1968). Development of a stemming algorithm. MIT Information Processing Group, Electronic Systems Laboratory Cambridge.

[35] C. D. Manning, P. Raghavan and H. Schtze (2008). Introduction to information retrieval. Cambridge University Press.

[36] T. K. Landauer and S. T. Dumais (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological Review 104(2):211.

[37] B. Sarwar, G. Karypis, J. Konstan and J. Riedl (2000). Application of dimensionality reduction in recommender system–a case study. Tech. rep., DTIC Document.

[38] M. Ley (2009). DBLP: some lessons learned. Proceedings of the VLDB Endowment 2(2):1493–1500.

Ramin Safa studied Computer Engineering at the Islamic Azad University of Lahijan, Guilan, Iran. In 2012 he got his B.Sc. degree and was accepted to the University of Guilan to earn a M.Sc. degree in Information Technology which he received in 2014. In 2015 he was accepted to pursue his studies as a full-time Ph.D. student in Software Systems at Islamic Azad University of Rasht. Today he is a Ph.D. candidate with research interests including Data Mining Techniques and Recommender Systems as well as their applications.

Seyed Abolghasem Mirroshandel received his B.Sc. degree from University of Tehran in 2005 and the M.Sc. and Ph.D. degree from Sharif University of Technology, Tehran, Iran in 2007 and 2012, respectively. Since 2012, he has been with Faculty of Engineering at University of Guilan in Rasht, Iran, where he is an Assistant Professor of Computer Engineering. Dr. Mirroshandel has published more than 50 technical papers in peer-reviewed journals and conference proceedings. His current research interests focus on Data Mining, Machine Learning, and Natural Language Processing.

Soroush Javadi received his B.Sc. degree in Software Engineering from University of Guilan, Rasht, Iran in 2015. His main research interests include Machine Learning and Natural Language Processing.

Mohammad Azizi received the B.Sc. and M.Sc. degree in Software Engineering from Islamic Azad University of Lahijan in 2012 and University of Guilan in 2016, respectively. His main research interests include Data Mining Techniques and Recommender Systems.