2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA) | 978-1-7281-8577-4/20/$31.00 ©2020 IEEE | DOI: 10.1109/AICCSA50499.2020.9316516

Recommendation of Academic Papers Based on Heterogeneous Information Networks

Nana Du*, Jun Guo*, Chase Q. Wu†, Aiqin Hou*, Zimin Zhao*, and Daguang Gan‡
* School of Information Science and Technology, Northwest University, Xi'an, Shaanxi 710127, China
Email: {dunana,zhaozimin}@stumail.nwu.edu.cn, {guojun,houaiqin}@nwu.edu.cn
† Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102, USA
Email: chase.wu@njit.edu
‡ Wanfang Data, Beijing 100000, China
Email: gandg@wanfangdata.com.cn

Abstract—**The rapid advance in science and technology is made possible by research conduct and breakthroughs in a wide range of fields, which have resulted in a large number of academic papers. Searching through the enormous literature to find relevant information of one's research interest has become an increasingly important yet challenging problem for many researchers. Most existing methods for academic paper recommendation are based on the analysis of paper contents and only meet with limited success. We propose a novel method based on heterogeneous information networks for academic paper recommendation, referred to as HNPR. This method considers the citation relationship between papers, the collaboration relationship between authors, and the research area information of papers to construct two types of heterogeneous information networks. In such networks, a random walk-based strategy is used to simulate natural sentences for the discovery of relevance between two papers according to a mature natural language processing model. Extensive experimental results using real data in public digital libraries show that HNPR significantly improves the accuracy of academic paper recommendation in comparison with traditional content-based recommendation methods.**

Index Terms—**Heterogeneous information networks, academic paper recommendation, random walk, natural language model**

I. Introduction

The rapid advance in science and technology critically depends on the research progress being made in a wide range of fields on a daily basis. As a result, a large number of academic papers have been published at various venues and accumulated over many decades. Particularly, we have witnessed an explosive growth in recent years due to the increasing interest in open-access publications on the Internet. Therefore, searching through the enormous literature to find relevant information of one's research interest has become an increasingly important yet challenging problem for many researchers. In some sense, the speed and accuracy of finding relevant publications largely affect the progress of research conduct in general and the chance of making possible research breakthroughs.

Most existing methods for academic paper recommendation are based on the analysis of various paper contents and only meet with limited success. For example, nowadays, many researchers still heavily rely on general search engines on the Internet such as Google or Google Scholar to search for papers, which typically require the input of keywords from the user and impose certain limitations on the search results. Similarly, in most of the databases of academic literature, paper search is based on limited information such as year of publication, author name, and key technical terms. These existing search methods do not consider correlations between papers such as mutual citations, co-authors, and relevance of research areas. Such correlation information, if discovered, represented, and analyzed properly, could greatly improve the effectiveness in paper recommendation.

Recommendation algorithms in recommendation systems are mainly divided into three categories: collaborative filtering, content-based filtering, and graph-based filtering. Among them, collaborative filtering methods are established based on a rationale that researchers with similar research interests would publish and search for papers of certain similarity. Content-based filtering methods recommend a collection of papers based on the relevance of paper contents. Graph-based filtering methods combine paper contents and various pieces of derived information such as citation relationships, collaboration relationships, and research area relationships in a graphical form, oftentimes referred to as a heterogeneous information network, and use machine learning techniques for paper recommendation.

In this work, we propose to construct heterogeneous information networks and leverage natural language models to effectively identify words with identical meanings to recommend papers with high similarity. According to the natural language model, if the words appearing in the context of two words, respectively, are similar, it is concluded that these two words also have similar meanings [1]. We attempt to apply this model to the problem of paper recommendation: when two papers are of certain similarity in terms of citation, co-authorship, and relevance of areas, they are considered to be relevant. We conduct extensive experiments on real data in public digital libraries to validate the applicability of this model in paper recommendation. The experimental results show that this model can not only address the issue of one-sidedness in user-based collaborative filtering methods, but also overcome the limitation of content-based filtering methods that fail to take paper quality into consideration.

978-1-7281-8577-4/20/$31.00 ©2020 IEEE

The main contributions of our work are summarized as follows:

- We combine various pieces of information derived from papers such as author collaboration, paper citation, and research area to construct two types of heterogeneous information networks
- We employ a random walk-based strategy to conduct edge traversal in the constructed heterogeneous information networks and adopt natural language models to match word sequences for paper recommendation.
- The superiority of HNPR over existing methods is illustrated through extensive experiments on real data extracted from public digital libraries.

II. Related Work

Similar to the methods used in common recommendation systems, the methods for paper recommendation also fall in one of three categories: collaborative filtering [2, 3], content-based filtering [2, 4–6] and graph-based filtering methods [5–15]. Kazunari *et al.* [4] believed that a researcher's interest was disclosed in the past papers published by the researcher, and effective information could be extracted from the researcher's track record of publications to recommend potential papers of interest. Kazunari *et al.* [2] also believed that different sections of a paper contain information of different levels of importance for finding relevant papers. Therefore, in collaborative filtering methods, such section-aware importance is also considered to improve the accuracy of recommendation. In some online browsing support systems for paper recommendation, researchers may obtain relevant contents as part of a paper according to their needs for fast response time. Based on this idea, Ohta *et al.* [5] proposed a new paper recommendation method, which generates a bipartite graph from the chapter or section recommended in the online support system and the paper that contains the contents, and then uses the HITS algorithm to process the bipartite graph for paper recommendation. The recommended papers are further sorted to obtain the most relevant papers of interest.

In graph-based filtering methods, a combination of random walk strategy [16] and statistical model [17] can effectively extract hidden features in a network graph, such as the HINE algorithm proposed by Cai *et al.* [7], the HERec algorithm proposed by Shi *et al.* [9], and the PaperRank algorithm proposed by Gori *et al.* [12]. Cai *et al.* [7] established five graphs based on the mutual citation of papers, author cooperation, author-paper relationship, paper word-author relationship, and paper word-paper relationship, respectively, and generated paper representation vector using the proposed HINE algorithm. Shi *et al.* [9] proposed HERec, which is a heterogeneous information network embedding algorithm based on the random walk strategy of metapath. Zhao *et al.* [8] proposed another recommendation method to recommend users based on the author's research history and some of the knowledge that may be missing. They extracted research area information and built a concept map, which is used to compare the user's research background and current research objectives. To make up for the lack of knowledge, they employed the shortest conceptual path to explore papers that may be of interest to users.

In our work, we leverage natural language processing models [1, 17–19], which extract synonyms in natural languages, to extract relevant papers with similar research contents. Paper recommendation methods based on heterogeneous information networks need to embed the network in a vector representation. Traditional network embedding methods include Multi-Dimensional Scaling (MDS) [20], Laplacian Eigenmap, Local Linear Embedding (LLE) and Isometric Mapping (IsoMap) [21].

III. HNPR Algorithm

A. Heterogeneous Information Networks

In this section, we construct two heterogeneous information networks, i.e., paper-author network and paper-area network. When a paper-author network is used alone with a weak network structure, the accuracy of paper recommendation can be improved by adjusting the proportion of research area similarity.

1) Paper-Author Network: We denote a paper-author network as $G_1(V_1, E_1)$. There are two types of nodes, $V_1 = P \cup A$, where $P = \{p_i, i = 1, 2, \cdots, P_{num}\}$ represents all of the P_{num} papers in the database, and $A = \{a_j, j = 1, 2, \cdots, A_{num}\}$ represents all of the A_{num} authors in the database. There are three types of edges, i.e., $E_1 =< E_{pa}, E_{pp}, E_{aa} >$, where $E_{pa} = \{(p_i, a_j) | 1 \leq i \leq P_{num}, 1 \leq j \leq A_{num}\}$ denotes a set of undirected edges, indicating that a_j is an author of paper p_i, $E_{aa} = \{(a_i, a_j) | 1 \leq i, j \leq A_{num}\}$ denotes a set of undirected edges, indicating that authors a_i and a_j co-authored one or more papers, and $E_{pp} = \{(p_i, p_j) | 1 \leq i, j \leq P_{num}\}$ denotes a set of directed edges, indicating that paper p_j is a reference cited by paper p_i.

2) Paper-Area Network: Nowadays, most manuscript submission systems would require authors to select the most relevant research areas from a given list of areas during the submission of a paper. Typically, such a list of research areas are prepared and provided by the system in a systematic way, where a research area is represented by a carefully chosen word or phrase. Such words or phrases obviate the need for paper classification and can be used to provide significant auxiliary information for paper recommendation. Therefore, we also consider research area information in our work to facilitate paper recommendation.

We denote a paper-area network as $G_2(V_2, E_2)$. Similar to G_1, the node set contains two types of nodes, $V_2 = P \cup D$, where P is the same as in G_1, and $D = \{d_k, k = 1, 2, \cdots, D_{num}\}$ denotes the set of all D_{num} research areas in the database. The research area information is comprised of the names of the areas as specified and arranged by the system. The edge set $E_2 = \{(p_i, d_k) | 1 \leq i \leq P_{num}, 1 \leq k \leq D_{num}\}$ is a set of undirected edges, where edge (p_i, d_k) indicates that research area d_k is one of the research areas specified for paper p_i.

B. Random Walk

The random walk algorithm [16] has been widely used to generate random walk (RW) sequences, for example, to simulate natural language sequences. In the Natural Language Processing (NLP) model [17], two similar words can be identified without losing the ability to encode each word into a different representation. In the language model, if the contexts of two words are similar, it is always concluded that these two words have similar meanings. Similarly, in the random walk-based sequences of papers generated from the above heterogeneous information networks, if the front and back RW sequences of two papers are similar, we conclude that these two papers have similar research contents.

In practice, researchers would be interested in two papers with similar contents in terms of one or multiple layers of references, cited papers, and authors. We introduce in detail the algorithm used for generating random walk sequences as follows.

We use w_{vi} to denote a RW starting from node $v_i \in V_1$, which is represented by a sequence of $w_{vi}^1, w_{vi}^2, ..., w_{vi}^k, ..., w_{vi}^{L_{num}}$. The next hop from node w_{vi}^k in sequence w_{vi} is node w_{vi}^{k+1}, which is randomly selected among the adjacent nodes of node w_{vi}^k. Note that a directed edge can be traversed only along the direction of the edge, while an undirected edge can be traversed bidirectionally. From any node in G_1 network, random walk repeats RA_{num} times and the length of each random walk sequence is denoted as LA_{num}, forming a set R_1 of random walk sequences. Similarly, starting from any node in G_2 network, random walk repeats RD_{num} times and the length of each walk sequence is denoted as LD_{num}, forming a set R_2 of random walk sequences.

Fig. 1 shows an example paper-author network G_1, where $V_1 = \{p_1, p_2, p_3, p_4, a_1, a_2\}$, $P = \{p_1, p_2, p_3, p_4\}$, $A = \{a_1, a_2\}$, and $E_1 = \{(p_1, p_2), \cdots, (p_4, p_2), (p_1, a_1), \cdots, (p_2, a_2)\}$. Let RA_{num}=2 and LA_{num}=3. According to the random walk algorithm, for $w_{p1}^1 = p_1$, w_{p1}^2 can be randomly selected from the set $\{p_2, a_1, a_2\}$, which is the set of p_1's adjacent nodes, and we set $w_{p1}^2 = a_1$. By repeating this process, the w_{p1} sequence can be obtained as $\{p_1, a_2, p_2\}$. Table I shows several examples of the random walk sequence set R_1 generated using this example network. Since R_1 is randomly generated, it is not unique. The random walk sequence set R_2 generated from paper-area network G_2 is obtained in a similar manner.

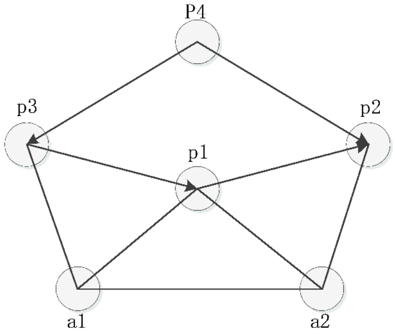

Fig. 1. An example paper-author network.

TABLE I
RANDOM WALK SEQUENCE SET EXAMPLES.

V_i	W_{vi}	R_1
p_1	(p_1,a_2,p_2)	{(p_1,a_2,p_2), (p_1,a_1,p_3), (p_2,a_2,p_1), (p_2,a_2,p_1), (p_3,p_1,p_2), (p_3,a_1,a_2), (p_4,p_3,p_1), (p_4,p_2,a_2), (a_1,p_1,a_1), (a_1,a_2,p_2), (a_2,p_2,a_2), (a_2,a_1,p_3)}
	(p_1,a_1,p_3)	
p_2	(p_2,a_2,p_1)	
	(p_2,a_2,p_1)	
p_3	(p_3,p_1,p_2)	
	(p_3,a_1,a_2)	
p_4	(p_4,p_3,p_1)	
	(p_4,p_2,a_2)	
a_1	(a_1,p_1,a_1)	
	(a_1,a_2,p_2)	
a_2	(a_2,p_2,a_2)	
	(a_2,a_1,p_3)	

C. Natural Language Model

We use the natural language model in the paper-author network and the paper-area network, respectively. The similarity *sim*$_1$ of two papers, which extracts the citation relationship between papers and the collaboration relationship between authors is measured in the paper-author network. The similarity *sim*$_2$ of two papers, which extracts the research area information is measured in the paper-area network. We use an adjustment factor α to balance the weights of *sim*$_1$ and *sim*$_2$, and then compute a combined similarity *sim* of two papers. The first *Re_Num* papers with the largest *sim* values to a specific paper are the final recommended papers. We describe the application of the natural language model [22, 23] in detail as follows.

We first introduce the vector representation method for each node in paper-author network G_1 using the natural language model. Each random walk sequence generated from G_1 is regarded as a sentence in the natural language model, and each node is a word. According to the natural language model [22], we establish the following objective function:

$$\sum_{r \in R_1} \sum_{i}^{LA_{num}} \log P(\{v_{i-w} : v_{i+w}\} \setminus v_i | v_i)$$
$$= \sum_{r \in R_1} \sum_{i}^{LA_{num}} \sum_{j=i-w, j \neq i}^{i+w} \log P(v_j | v_i), \qquad (1)$$

where w denotes the window size, which is the maximum number of hops before and after node v_i in the random walk sequence $r \in R_1$, and $P(\{v_{i-w} : v_{i+w}\} \backslash v_i | v_i)$ is a conditional probability that the set $\{v_{i-w} : v_{i+w}\} \backslash v_i$ appears within the w hops before and after node v_i on the premise of the appearance of node v_i.

In particular, we employ the Neural Probabilistic Language Models (NPLM) [1] to obtain the vector representation set VEC of each paper and author. SkipGram model and CBOW are among the most important models to maximize the objec-

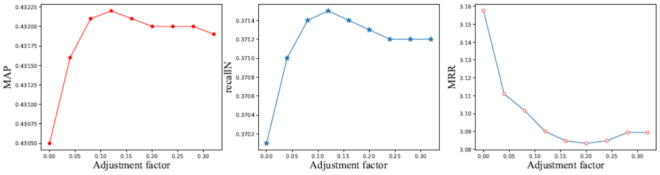

Fig. 2. Measurements of recall@N, MRR, and MAP indicators in response to the adjustment factor α in the N2Y paper set.

tive function defined in Eq. (1). NPLM is a fast hierarchical language model along with a simple feature-based algorithm for automatic construction of word trees from the data. It replaces the unstructured vocabulary of NPLM by a binary tree that represents a hierarchical clustering of words in the vocabulary. Each word corresponds to a leaf in the tree and can be uniquely specified by the path from the root to that leaf. The word tree can be automatically created by using expert knowledge, data-driven methods, or a combination of both.

Once the vector representation is obtained through NPLM, we compute the cosine similarity [24] to measure the similarity sim_1 of two papers, as follows:

$$sim_1 = \cos(\theta) = \frac{V_i \bullet V_j}{\|V_i\| \, \|V_j\|}, V_i, V_j \in VEC, \tag{2}$$

where V_i and V_j denote the vector representations of paper p_i and p_j, respectively. In the heterogeneous information network G_1, the structural characteristics of authors and citations of two papers are expressed in sim_1, but there is insufficient consideration of the contents of two papers. In order to address this deficiency, many researchers extract feature words of papers based on word segmentation and clustering, and then extract the similarity in the contents of two papers. This method shows a significant advantage, but the list of carefully designated research areas could effectively replace word segmentation and clustering. Note that the research area information is prepared and provided by the system and hence is of higher credibility.

In order to utilize research area information, we also carry out the above procedures based on random walk and natural language processing models in the paper-area network G_2. Similar to G_1, we obtain a vector representation of each paper, but this time the vector representation extracts the structural information of G_2. The similarity sim_2 of two papers in G_2 is also computed using the cosine similarity as shown in Eq. (2).

Based on both of the above similarities sim_1 and sim_2, we compute a combined similarity, which uses an adjustment factor $0 \leq \alpha \leq 1$ to balance the influences of G_1 and G_2, as follows

$$sim = \alpha \cdot sim_1 + (1 - \alpha) \cdot sim_2. \tag{3}$$

According to the measurements of *sim*, our proposed HNPR algorithm recommends the first *Re_Num* papers with the highest similarity measurements to a given paper, as the final recommendation result.

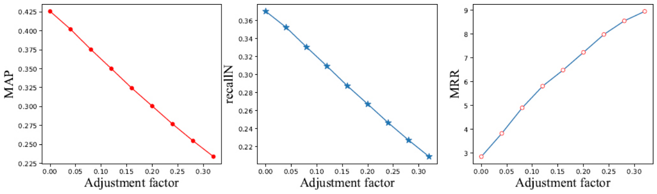

Fig. 3. Measurements of recall@N, MRR, and MAP indicators in response to adjustment factor α in the MA paper set.

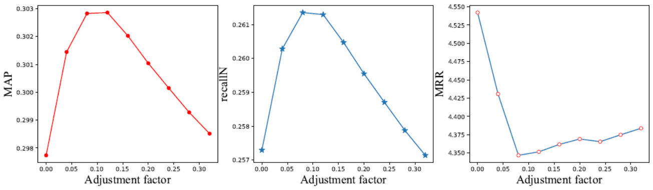

Fig. 4. Measurements of recall@N, MRR, and MAP indicators in response to the adjustment factor α in the IP paper set.

IV. Experimental Results and Performance Analysis

A. Experiment Dataset

We conduct experiments on Aminer Dataset[1] and Wanfang Dataset[2]. Aminer Dataset contains more than 4 million papers, from which we extract papers in the past 2 years (2018-2019) (N2Y), papers in the Mathematical Analysis (MA) area, and papers in Image Processing (IP) area as three independent experimental paper sets. The detailed information of the dataset is shown in Table II.

- Paper: the number of papers.
- Author: the number of authors.
- Research Area: the number of research areas.
- Citing: the number of citation relationships.
- Cooperation: the number of cooperation relationships between authors.
- Paper-Author: the number of paper-author relationships.
- Paper-Area: the number of paper-area relationships.

TABLE II
Statistics of the test dataset.

	N2Y	MA	IP
Paper	221,076	98,702	49,098
Author	503,945	117,183	107,290
Research Area	45,072	19,573	17,875
Citing	2,686,359	578,242	737,423
Cooperation	1,746,008	225,838	305,662
Paper-Author	885,295	245,099	203,421
Paper-Area	2,090,960	1,067,671	712,730

TABLE III
Parameter settings.

Intri_Num	*Re_Num*	*RAnum*	*LAnum*	*RDnum*	*LDnum*
30	50	30	20	30	20

[1] https://www.aminer.cn/data
[2] http://www.wanfangdata.com.cn/index.html

TABLE IV
COMPARISON OF RECOMMENDATION QUALITY.

	TF-IDF	HINE	HNPR/α=0.12	HNPR/α=0.08	
Recall@N	0.1378	0.3701	**0.3715**	0.3714	2018-2019 paper set (N2Y)
MRR	8.2061	3.1575	**3.0901**	3.1017	
MAP	0.1598	0.4305	**0.4322**	0.4321	
	TF-IDF	HINE	HNPR/$\alpha = 0$	HNPR/$\alpha = 0.04$	Mathematical analysis paper set (MA)
Recall@N	0.1772	0.3512	**0.3701**	0.3404	
MRR	5.3159	3.8316	**2.8576**	4.0059	
MAP	0.2100	0.4020	**0.4256**	0.4052	
	TF-IDF	HINE	HNPR/$\alpha = 0.08$	HNPR/$\alpha = 0.12$	Image processing paper set (IP)
Recall@N	0.1284	0.2572	**0.2613**	0.2612	
MRR	7.8658	4.5419	**4.3468**	4.3516	
MAP	0.1508	0.2977	**0.3028**	0.3028	

We use at most *Intri_Num* papers with the highest number of identical references as paper p as the relevant paper set to paper p. This method of extracting the relevant paper set is only useful for a small number of papers in the database, because we assume that two papers with many identical references have content relevance. However, most of the related papers in the database may not have any identical citation. Hence, this method is only suitable for extracting a test paper set.

For illustration, as shown in Fig. 1, p_2 and p_3 are two references of p_4, and p_2 is also cited by p_1, so p_2 is one identical reference of p_4 and p_1. Paper p_1 has 0, 0, and 1 identical references as p_2, p_3, and p_4, respectively. We ranked all papers in the database by the total number of identical references. For example, p_1 has 1 total identical reference as $0+0+1=1$. We select the top 5000 articles with the largest total number of identical references as the test paper set T_p, and extract the first *Intri_Num* papers with the most number of identical references with the test paper p_i ($p_i \in T_p$) as the relevant paper set of the test paper p_i ($p_i \in T_p$). The parameter settings are provided in Table III.

The experimental results show that a proper adjustment of the above parameters does not have a significant impact on the recommendation results.

B. Performance Metrics for Evaluation

We consider the following performance metrics:

- *Recall@N*: This is the proportion of correctly recommended papers in the relevant paper set [7].
- *MRR*: This is used to measure the average value of the order, in which relevant papers appear for the first time in the recommended set [7], defined as

$$MRR = \frac{1}{|T_p|} \sum_{p_i \in T_p} rank_{\text{first}}(p_i), \tag{4}$$

  where T_p is the test paper set, and $rank_{\text{first}}(p_i)$ is the order of the paper successfully recommended for the first time in the recommended paper set.
- *MAP*: In order to take into account the number of successfully recommended papers and the order of all successfully recommended papers in the recommended set [6], *MAP* is computed as follows:

$$MAP = \frac{1}{|T_p|} \sum_{p_i \in T_p} \frac{1}{|T_s|} \sum_{r_j \in \mathrm{T_s}, rank(r_j) \neq 0} \frac{rank(r_j)+1}{rank(r_j)}, \tag{5}$$

  where T_s denotes the relevant paper set corresponding to the test paper p_i. If a relevant paper $r_j \in T_s$ is not in the recommended paper set, $rank(r_j) = 0$.

C. Experimental Results and Analysis

In order to evaluate the recommendation performance of HNPR, we select two classic recommendation algorithms for comparative analysis: HINE [7] and TF-IDF [4]. Figs. 2 to 4 plots the performance measurements in terms of *recall@N*, *MRR*, and *MAP* in the N2Y, MA, and IP paper sets, respectively, where the horizontal axis represents the adjustment factor α. According to the trend of the three performance measurements in the corresponding paper sets, we observe an optimal adjustment factor in each paper set, i.e., $\alpha_{\text{N2Y}} = 0.12$, $\alpha_{\text{MA}} = 0$, and $\alpha_{\text{IP}} = 0.08$, respectively. We select two optimal adjustment factors in each paper set, and compare the recommendation results with the results of HINE and TF-IDF, as shown in Table IV.

From Table IV, we observe that the recommendation accuracy of TF-IDF, which is a traditional recommendation algorithm based on paper contents, is significantly lower than that of the graph-based recommendation algorithms HINE and HNPR, and only recommends 12.84%-13.78% of relevant papers correctly. In the N2Y paper collection, 37.15% of relevant papers are correctly recommended by HNPR, which is higher than 37.01% achieved by HNIE. Qualitatively similar recommendation results are produced in the MA and IP paper sets. We would also like to point out that HNPR uses a much smaller number of variables than HINE because HNPR uses area information. Hence, HNPR has a lower time complexity and runs faster than HINE.

According to the measurements of *MRR*, the recommendation order of the first correctly recommended relevant papers by HINE and HNPR is about the third on average. In the MA paper set, when $\alpha = 0.12$, the *MAP* measurement of HNPR is 5.87% higher than HINE. In the paper sets of N2Y and IP, as the value of α increases, we observe better recommendation results, indicating that research area information can significantly improve recommendation performance.

V. Conclusion

In this paper, we proposed a method, HNPR, for paper recommendation based on heterogeneous information networks, by taking into account the structural and content information of papers. We leveraged the natural language models to identify papers that appear in similar contexts. The experimental results showed that the proposed HNPR algorithm significantly outperforms the traditional content-based recommendation algorithm TF-IDF, and correctly recommends more than 30% of relevant papers, with a higher accuracy than the classical graph-based algorithm HINE.

Acknowledgement

This work was supported by the National Key Research and Development Program of China under Grant No. 2017YFB1400301.

References

[1] A. Mnih and G.E. Hinton. A scalable hierarchical distributed language model. In *Proc. of Advances in Neural Information Processing Systems*, pages 1081–1088, Vancouver, British Columbia, Canada, Dec. 2009.

[2] K. Sugiyama and M.-Y. Kan. Exploiting potential citation papers in scholarly paper recommendation. In *Proc. of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 153–162, Jul. 2013.

[3] N. Sakib, R. B. Ahmad, and K. Haruna. A collaborative approach toward scientific paper recommendation using citation context. *IEEE Access*, 8:51246–51255, 2020.

[4] K. Sugiyama and M.-Y. Kan. Scholarly paper recommendation via user's recent research interests. In *Proc. of the 10th Annual Joint Conference on Digital Libraries*, pages 29–38, Aug. 2010.

[5] M. Ohta, T. Hachiki, and A. Takasu. Related paper recommendation to support online-browsing of research papers. In *Proc. of the 4th Int. Conf. on the Applications of Digital Information and Web Technologies*, pages 130–136. IEEE, Aug. 2011.

[6] T. Strohman, W.B. Croft, and D. Jensen. Recommending citations for academic papers. In *Proc. of the 30th Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pages 705–706, Amsterdam, The Netherlands, Jul. 2007.

[7] X. Cai, J. Han, S. Pan, and L. Yang. Heterogeneous information network embedding based personalized query-focused astronomy reference paper recommendation. *Int. J. of Computational Intelligence Systems*, 52(1):591–599, Mar. 2018.

[8] W. Zhao, R. Wu, and H. Liu. Paper recommendation based on the knowledge gap between a researcher's background knowledge and research target. *Information Processing & Management*, 52(5):976–988, Sep. 2016.

[9] C. Shi, B. Hu, W.X. Zhao, and S. Y. Philip. Heterogeneous information network embedding for recommendation. *IEEE Trans. on Knowledge and Data Engineering*, 31(2):357–370, 2018.

[10] F. Meng, D. Gao, W. Li, X. Sun, and Y. Hou. A unified graph model for personalized query-oriented reference paper recommendation. In *Proc. of the 22nd ACM Int. Conf. on Information & Knowledge Management*, pages 1509–1512, Oct. 2013.

[11] D. Zhou, S. Zhu, K. Yu, X. Song, B. L. Tseng, H. Zha, and C. Giles. Learning multiple graphs for document recommendations. In *Proc. of the 17th Int. Conf. on World Wide Web*, pages 141–150, Jan. 2008.

[12] M. Gori and A. Pucci. Research paper recommender systems: A random-walk based approach. In *Prof. of IEEE/WIC/ACM Inf. Conf. on Web Intelligence*, pages 778–781, Hong Kong, China, Dec. 2006.

[13] L. Pan, X. Dai, S. Huang, and J. Chen. Academic paper recommendation based on heterogeneous graph. In *Chinese computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*, pages 381–392. Springer, 2015.

[14] W. Tanner, E. Akbas, and M. Hasan. Paper recommendation based on citation relation. In *Proc. of IEEE Int. Conf. on Big Data*, pages 3053–3059, 2019.

[15] X. Ma and R. Wang. Personalized scientific paper recommendation based on heterogeneous graph representation. *IEEE Access*, 7:79887–79894, 2019.

[16] B. Perozzi, R. Al-Rfou, and S. Skiena. Deepwalk: Online learning of social representations. In *Proc. of the 20th ACM Int. Conf. on Knowledge Discovery and Data Mining*, pages 701–710, Mar. 2014.

[17] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin. A neural probabilistic language model. *Journal of Machine Learning Research*, 3:1137–1155, Mar. 2003.

[18] F. Morin and Y. Bengio. Hierarchical probabilistic neural network language model. In *Proc. of the 10th Int. Workshop on Artificial Intelligence and Statistics*, volume 5, pages 246–252, Jan. 2005.

[19] G. A Miller. *WordNet: An electronic lexical database*. MIT press, Jan. 1998.

[20] M. A. Cox and T. F. Cox. Multidimensional scaling. In *Handbook of Data Visualization*, pages 315–347. Springer, 2008.

[21] J. B. Tenenbaum, V. De Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[22] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[23] M. Allamanis, D. Tarlow, A. Gordon, and Y. Wei. Bimodal modelling of source code and natural language. In *Proc. of Int. Conf. on Machine Learning*, pages 2123–2132, 2015.

[24] A. Huang. Similarity measures for text document clustering. In *Proc. of the 6th New Zealand Computer Science Research Student Conference*, volume 4, pages 9–56, Christchurch, New Zealand, 2008.