Paper Recommendation Using Citation Contexts in Scholarly Documents

Tomoki Ikoma[1] and **Shigeki Matsubara**[1,2]
[1]Graduate School of Informatics, Nagoya University, Japan
[2]Information Technology Center, Nagoya University, Japan
ikoma.tomoki.d0@s.mail.nagoya-u.ac.jp

Abstract

This paper proposes a citation recommendation method that considers and utilizes citation contexts in scholarly papers. Citation contexts describe the purpose of the citation and how the cited paper is related to the citing paper. Citation contexts allow citation recommendation systems to select relevant papers based on the relationships between other existing papers. In the proposed method, the candidate selection model first projects the input paper and existing papers in the database to a vector space using the words appearing in the title and abstract. Then, a recommendation decision model estimates the probability that each candidate is a suitable citation for the input paper using a feed-forward network. Experiments conducted on scholarly papers in the medical domain demonstrate the benefits of using citation contexts for the citation recommendation task.

1 Introduction

Scholarly papers cite papers for various reasons, e.g., introducing the concepts applied in the research, referring to relevant tools and datasets used in experiments, and providing information about previous relevant studies.

However, finding papers to cite is becoming increasingly difficult because the number of papers is constantly increasing. To help authors identify appropriate papers for reference, several citation recommendation systems have been proposed (Bhagavatula et al., 2018; Dai et al., 2017; Mu et al., 2018; Guo et al., 2017; Yang et al., 2018; Han et al., 2018; Ebesu and Fang, 2017; Duma and Klein, 2014).

Most existing citation recommendation systems take the paper's title and abstract as the input, and they output a list of suitable papers to cite in the input paper (Färber and Jatowt, 2020).

This paper proposes a citation recommendation method that utilizes citation contexts in scholarly papers, which describe the purpose of the citation and how the cited paper is related to the citing paper. These citation contexts allow citation recommendation systems to select papers based on the relationships between other existing papers. Experiments conducted on scholarly papers in the medical domain have demonstrated the benefits of using citation contexts for citation recommendation systems.

2 Paper Recommendation

This section describes the target problem and the previously proposed citation recommendation systems.

2.1 Problem Settings

Studies on citation recommendation can be categorized as global citation recommendation or local citation recommendation. Global citation recommendation systems typically take the title and abstract of the user's paper as the input and output a list of relevant papers suitable to cite in the input paper (Bhagavatula et al., 2018; Dai et al., 2017; Mu et al., 2018; Guo et al., 2017). Local citation recommendation systems take an excerpt of the user's paper as the input and output suitable papers to cite in the input context (Yang et al., 2018; Han et al., 2018; Ebesu and Fang, 2017; Duma and Klein, 2014). The former category assumes situations where the user has completed writing the title and abstract and is currently writing the main text. The latter category assumes that the user has written a part of the text and wishes to cite papers to reinforce a given statement.

In this paper, we focus on the former category; thus, this paper proposes a method to identify papers that are relevant to the input title and abstract.

2.2 Related Work

Several previous studies have investigated citation recommendation systems (Färber and Jatowt, 2020; Beel et al., 2015; Ali et al., 2020, 2021; Ma et al.,

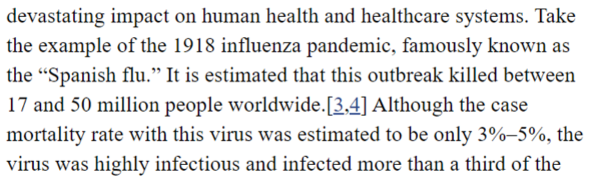
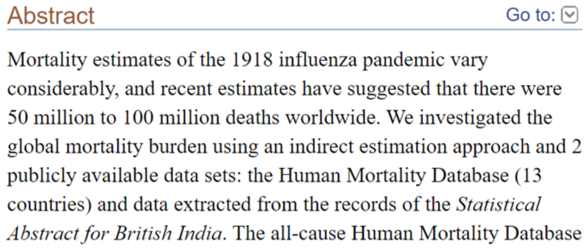

devastating impact on human health and healthcare systems. Take the example of the 1918 influenza pandemic, famously known as the "Spanish flu." It is estimated that this outbreak killed between 17 and 50 million people worldwide.[3,4] Although the case mortality rate with this virus was estimated to be only 3%–5%, the virus was highly infectious and infected more than a third of the

Figure 1: Part of the citation context of (Spreeuwenberg et al., 2018) in (Khan et al., 2020)


Abstract Go to:

Mortality estimates of the 1918 influenza pandemic vary considerably, and recent estimates have suggested that there were 50 million to 100 million deaths worldwide. We investigated the global mortality burden using an indirect estimation approach and 2 publicly available data sets: the Human Mortality Database (13 countries) and data extracted from the records of the *Statistical Abstract for British India*. The all-cause Human Mortality Database

Figure 2: Part of the abstract of (Khan et al., 2020)

2020). Previously proposed methods are categorized as follows:

Content-based methods: Content-based methods utilize a set of features extracted from content of the input paper and existing papers. Such systems commonly use the words appearing in the text of the papers, and additional features include the paper topics and n-grams appearing in the texts. Note that most existing citation recommendation systems are based on this approach.

Collaborative filtering-based methods: Collaborative filtering-based methods recommend papers based on the target interest of the user. Such systems use reader evaluations of each paper to recommend papers appreciated by readers with similar interests to the user. The characteristics of this method are summarized as follows:

- There is no need to analyze the contexts of the papers.
- The method can consider the human-rated quality of the documents.

Graph-based methods: Graph-based methods search for papers based on the citation relationships among the papers. Such systems search for papers using the random walk algorithm starting from one or several papers, which are typically those cited in the paper the user is writing.

For the content-based methods, (Nogueira et al., 2020) retrieved recommendation candidates using a keyword-based approach and navigation-based expansion. Then, candidates were reranked using the candidates with pretrained transformer models, e.g., BERT (Devlin et al., 2019). In addition, (Yang et al., 2019) developed an encoder-decoder model that scores the suitability of each candidate for citation in the input paper. In terms of collaborative filtering-based methods, (Hu et al., 2020) proposed a method that searches for papers to recommend based on the reference relation between authors, the citation relationship between the author and the paper, and the authoritativeness of the authors and papers. In addition, (Bansal et al., 2016) proposed a method that leverages GRU to encode the text into a latent vector, which was utilized to select relevant papers to recommend. For the graph-based methods, (Tian and Jing, 2013) constructed a relational graph representing the similarity of the content and the interest of researchers, and then this graph was used to facilitate citation recommendation. (Kong et al., 2021) also proposed a method that utilizes vector representations of each paper extracted using a network embedding approach.

3 Usage of Citation Contexts

In the following, we describe citation contexts and their usage in this study.

3.1 Citation Contexts

The citation context of the cited paper is part of the citing paper's text that includes a description of the cited paper. Here, we define the citation context as the paragraph containing the reference tag.

Figure 1 shows part of the citation context of (Spreeuwenberg et al., 2018) found in (Khan et al., 2020), whose abstract appears in Figure 2. The corresponding citation tag is the underlined bracket in Figure 1.

Citation contexts contain various information, e.g., the reason the authors of the citing papers cited the paper and the relationship between the citing and cited papers. Therefore, the contents of citation contexts can be considered the most important points of the cited paper from the perspective of the citing paper's authors. For example, the citation context shown in Figure 1 states that the citing paper cited a paper on the Spanish flu to compare the COVID-19 pandemic and previous diseases.

In contrast, the abstract of papers contains the most important points from the author's viewpoint,

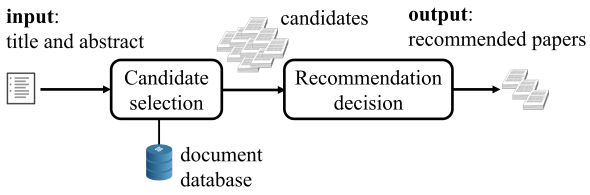


Figure 3: Configuration of the proposed method

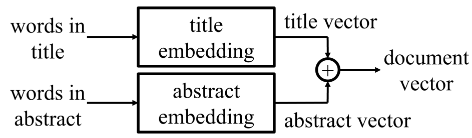


Figure 4: Embedding papers using candidate selection model

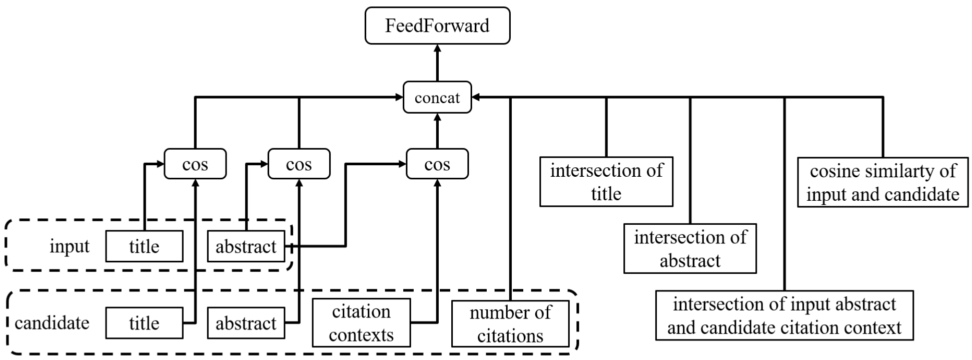


Figure 5: Configuration of recommendation decision model

such as the overview, the proposed ideas and concepts, and the experimental results. For example, the abstract shown in Figure 2 states that the paper revises the records of the 1918 Spanish flu.

3.2 Usage of Citation Contexts

In this study, the citation contexts are utilized in combination with the abstract of the input paper. Previously proposed methods (Bhagavatula et al., 2018; Dai et al., 2017; Mu et al., 2018; Guo et al., 2017) compared the abstracts of the input paper and recommendation candidates. However, these methods cannot select candidates based on how other existing papers cited each candidate.

In contrast, the proposed method compares the citation contexts of each candidate to the input abstract, which enables the citation recommendation system to select candidates based on information not included in the candidate abstract. In the example shown in Figure 1 and 2, with the help of citation contexts, the proposed method can select the candidate (Figure 2) considering that it references comparison with other diseases (Figure 1).

4 Method

The proposed method searches for candidate papers to recommend through two steps as shown in Figure 3. First, the proposed method extracts candidates for recommendation, and then it selects which candidates to recommend.

4.1 Candidate Selection

The candidate selection model projects the input paper and existing papers to a vector space using the words appearing in the title and abstract as show in Figure 4. Here, the model first learns the word embeddings to handle the title and abstract. The model then computes the title and abstract vectors as the weighted sum of each word. Finally, the model calculates the document vector as the weighted sum of the title and abstract vectors.

Note that we train the parameters, e.g., the embeddings and weights, such that papers with a citation relationship have a high cosine similarity. The candidate selection model calculates the cosine similarity between the input paper and each paper in the database. It then outputs a list of top N papers with the highest cosine similarity as candidates. In addition, papers directly cited in the N extracted documents are also selected as candidates.

4.2 Recommendation Decision

The recommendation decision model estimates the probability that each candidate is suitable for citation in the input paper via a feedforward network as shown in Figure 5. Then, it outputs documents with a higher probability than the given threshold t as the recommended documents.

A previously proposed model (Bhagavatula et al., 2018) uses the following features to calculate the probability:

Table 1: Tuning threshold in baseline model

Threshold	Precision	Recall	F1 score
.95	.0426	.1344	**.0647**
.90	.0323	.2205	.0563
.85	.0246	.2731	.0451
.80	.0194	.3102	.0365
.75	.0157	.3359	.0299
.70	.0131	.3550	.0252
.60	.0099	.3773	.0193
.50	.0083	.3869	.0163

- cosine similarity of the input and candidate title
- cosine similarity of the input and candidate abstract
- cosine similarity of the document vectors (calculated in the candidate selection step)
- number of times the candidate document has been cited
- the sum of scalar weights of the words appearing in both the input and candidate titles
- the sum of scalar weights of the words appearing in both the input and candidate abstracts

The proposed method includes the following additional features:

- cosine similarity of the input abstract and the citation contexts of the candidate document
- the sum of the scalar weights of the words appearing in both the input abstract and citation contexts of the candidate document

We train the parameters such as the word embeddings and the scalar weights for each word. Note that we newly train the word embeddings for the recommendation decision model rather than using the same embeddings as the candidate selection model.

5 Experiments

5.1 Dataset

The proposed method was evaluated experimentally on the PMC Open Access Subset[1], which is a corpus of open access papers in the medical domain. Here, we extracted approximately 320,000 papers randomly and split them into training, development, and testing sets according to the publication year. We used approximately 250,000 papers published prior to 2018 as the training data, 45,000 papers published in 2019 for development data, and 25,000 papers published in 2020 as the testing data.

[1]https://www.ncbi.nlm.nih.gov/pmc/tools/openftlist/

Table 2: Experimental results

	Precision	Recall	F1 score
Baseline	.0414	.1340	.0633
Proposed method	**.1130**	**.1472**	**.1279**

Table 3: Comparison of performance with different recommendation decision threshold values

Threshold	Precision	Recall	F1 score
.95	.1212	.1515	.1347
.90	.1051	.2447	**.1470**
.85	.0753	.2976	.1202
.80	.0529	.3310	.0913
.75	.0383	.3524	.0691
.70	.0284	.3677	.0527
.60	.0173	.3838	.0330
.50	.0119	.3931	.0232

5.2 Experimental Settings

We trained the candidate selection and recommendation decision models over 10 epochs using the training data. To evaluate the proposed method, we compared the list of recommended papers to the ground truth reference list of the input paper. For evaluation metrics, we considered precision, recall and their harmonic mean, i.e., the F1 score. Precision is defined as the fraction of ground-truth papers among the recommended documents, and recall is defined as the fraction of correctly recommended papers among the ground truth documents.

We compared the performance against a baseline method (Bhagavatula et al., 2018) that does not utilize citation contexts. Here, we set the recommendation decision threshold to 0.95 in accordance with the results shown in Table 1, and we set the number of candidates to be selected by the candidate selection model to 100.

5.3 Experimental Results

The experimental results are shown in Table 2. As can be seen, the proposed method outperformed the baseline in all evaluation metrics, which suggests the effectiveness of utilizing citation contexts in the citation recommendation task.

Table 4: Comparison of performance with different usage of citation contexts

Use citation contexts in	Precision	Recall	F1 score
None	.0426	.1344	.0647
Candidate selection	.1096	.1291	.1186
Recommendation decision	**.1212**	**.1515**	**.1347**
Both steps	.1046	.1191	.1114

6 Discussion

6.1 Recommendation Decision Threshold

We compared the performance of the proposed method on the development data with different recommendation decision threshold settings. Table 3 shows the results. Note that the number of recommended papers increases as the threshold decreases; thus, the recall value increases as the precision decreases. We found that the best F1 score was achieved with a threshold value of 0.9.

6.2 Utilizing Citation Contexts for Candidate Selection

The proposed method utilizes the citation contexts for only the recommendation decision. However, we can consider an alternative method that also utilizes the citation contexts for the candidate selection process. Thus, to analyze the effect of citation contexts on the candidate selection process, we trained a variant of the candidate selection model that also utilizes the cosine similarity between the input abstract and citation contexts of the candidates.

In this evaluation, we compared performance of the following models on the development data:

- citation contexts are not used
- citation contexts are used for the candidate selection
- citation contexts are used for the recommendation decision
- citation contexts are used for both candidate selection and recommendation decision

The results are shown in Table 4. As can be seen, the best performance was obtained when the citation contexts were only used for the recommendation decision, which implies that considering the citation contexts in both candidate selection and recommendation decision processes does not improve performance.

7 Conclusion

This paper has proposed a citation recommendation method that utilizes citation contexts in scholarly documents. The proposed method was evaluated on paper data from the medical domain, and the results showed that the performance of citation recommendation is improved by using citation contexts.

In future work, we plan to consider a method to extract the citation contexts more appropriately. In this paper, we defined the citation contexts uniformly as the paragraph containing the citation tag. However, a previous study (Xing et al., 2020) proposed a method to extract descriptions of the contents of cited papers from the paragraph containing the citation. We expect that such methods can be applied to extract the most appropriate part of the text as the citation context for use in the citation recommendation task.

In addition, we will apply the proposed method to the local citation task. In the current study, we utilized the citation contexts to search for suitable papers to cite in an input paper. However, we can also consider using citation contexts to identify papers for citation in the input text. We expect that using the citation contexts in combination with the input text will be highly beneficial for this task.

Limitations

Unavailability of Citation Contexts

The proposed method is heavily dependent on citation contexts to determine which candidate papers to recommend. However, collecting a large number of citation contexts is not a straightforward task. This is mostly due to the fact that, generally, only a limited proportion of existing papers are available as open access[2]. In addition, in some cases, the candidate paper has not received any citations, particularly for recently published papers. In such cases, citation context information is unavailable; thus,

[2]For example, PubMed stores 35,991,116 papers as of July 26th, 2023; however, open access to the full text is provided for only 30.8% (11,094,672 papers) of those papers.

the proposed method cannot be applied for such candidate papers. Therefore, traditional features, e.g., the similarity of abstracts, are still required for candidate papers without available citation contexts.

Difficulties of Performance Evaluation

The experimental results demonstrated that the proposed model outperformed the baseline method; however, the specific precision, recall and F1 score values were still very low. This is mainly due to the difficult task setting, in which the models must retrieve the few ground truth papers from thousands of candidates stored in the database. In addition, the ground truth papers used for this evaluation were taken from the reference list of each input paper. However, the papers that the users truly wish to find and those that are suitable for reference do not always match. The goal of this paper was to propose a method to retrieve the former; however, the actual evaluation was based on how well the model can retrieve the latter. Thus, additional evaluations from other perspectives, e.g., the user experiences, are required to gain a better understanding of the model's performance.

Acknowledgments

This work was partially supported by the Grant-in-Aid for Challenging Research (Exploratory) (No. 23K18506) of JSPS and by JST SPRING, Grant Number JPMJSP2125.

References

Zafar Ali, Pavlos Kefalas, Khan Muhammad, Bahadar Ali, and Muhammad Imran. 2020. Deep learning in citation recommendation models survey. *Expert Systems with Applications*, 162:113790.

Zafar Ali, Irfan Ullah, Amin Khan, Asim Ullah Jan, and Khan Muhammad. 2021. An overview and evaluation of citation recommendation models. *Scientometrics*, 126(5):4083–4119.

Trapit Bansal, David Belanger, and Andrew McCallum. 2016. Ask the GRU: Multi-task learning for deep text recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 107–114. Association for Computing Machinery.

Joeran Beel, Bela Gipp, Stefan Langer, and Corinna Breitinger. 2015. Research-paper recommender systems: a literature survey. *International Journal on Digital Libraries*, 17(4):305–338.

Chandra Bhagavatula, Sergey Feldman, Russell Power, and Waleed Ammar. 2018. Content-based citation recommendation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 238–251. Association for Computational Linguistics.

Tao Dai, Li Zhu, Xiaoyan Cai, Shirui Pan, and Sheng Yuan. 2017. Explore semantic topics and author communities for citation recommendation in bipartite bibliographic network. *Journal of Ambient Intelligence and Humanized Computing*, 9(4):957–975.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Daniel Duma and Ewan Klein. 2014. Citation resolution: A method for evaluating context-based citation recommendation systems. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 358–363. Association for Computational Linguistics.

Travis Ebesu and Yi Fang. 2017. Neural citation network for context-aware citation recommendation. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1093–1096. Association for Computing Machinery.

Michael Färber and Adam Jatowt. 2020. Citation recommendation: approaches and datasets. *International Journal on Digital Libraries*, 21(4):375–405.

Lantian Guo, Xiaoyan Cai, Fei Hao, Dejun Mu, Changjian Fang, and Libin Yang. 2017. Exploiting fine-grained co-authorship for personalized citation recommendation. *IEEE Access*, 5:12714–12725.

Jialong Han, Yan Song, Wayne Xin Zhao, Shuming Shi, and Haisong Zhang. 2018. hyperdoc2vec: Distributed representations of hypertext documents. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2384–2394. Association for Computational Linguistics.

Donglin Hu, Huifang Ma, Yuhang Liu, and Xiangchun He. 2020. Scientific paper recommendation using author's dual role citation relationship. In *IFIP Advances in Information and Communication Technology*, pages 121–132. Springer International Publishing.

Gulfaraz Khan, Mohamud Sheek-Hussein, Ahmed R. Al Suwaidi, Kamal Idris, and Fikri M. Abu-Zidan. 2020. Novel coronavirus pandemic: A global health threat. *Turkish Journal of Emergency Medicine*, 20(2):55.

Xiangjie Kong, Mengyi Mao, Wei Wang, Jiaying Liu, and Bo Xu. 2021. VOPRec: Vector representation learning of papers with text information and structural identity for recommendation. *IEEE Transactions on Emerging Topics in Computing*, 9(1):226–237.

Shutian Ma, Chengzhi Zhang, and Xiaozhong Liu. 2020. A review of citation recommendation: from textual content to enriched context. *Scientometrics*, 122(3):1445–1472.

Dejun Mu, Lantian Guo, Xiaoyan Cai, and Fei Hao. 2018. Query-focused personalized citation recommendation with mutually reinforced ranking. *IEEE Access*, 6:3107–3119.

Rodrigo Nogueira, Zhiying Jiang, Kyunghyun Cho, and Jimmy Lin. 2020. Navigation-based candidate expansion and pretrained language models for citation recommendation. *Scientometrics*, 125(3):3001–3016.

Peter Spreeuwenberg, Madelon Kroneman, and John Paget. 2018. Reassessing the global mortality burden of the 1918 influenza pandemic. *American Journal of Epidemiology*, 187(12):2561–2567.

Geng Tian and Liping Jing. 2013. Recommending scientific articles using bi-relational graph-based iterative RWR. In *Proceedings of the 7th ACM Conference on Recommender Systems*, pages 399–402. Association for Computing Machinery.

Xinyu Xing, Xiaosheng Fan, and Xiaojun Wan. 2020. Automatic generation of citation texts in scholarly papers: A pilot study. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6181–6190. Association for Computational Linguistics.

Libin Yang, Zeqing Zhang, Xiaoyan Cai, and Tao Dai. 2019. Attention-based personalized encoder-decoder model for local citation recommendation. *Computational Intelligence and Neuroscience*, 2019:1–7.

Libin Yang, Yu Zheng, Xiaoyan Cai, Hang Dai, Dejun Mu, Lantian Guo, and Tao Dai. 2018. A LSTM based model for personalized context-aware citation recommendation. *IEEE Access*, 6:59618–59627.