Citation Sentence Generation Leveraging the Content of Cited Papers

Akito Arita
Osaka Institute of Technology
m21a02@oit.ac.jp

Hiroaki Sugiyama
NTT Communication Science Laboratories
h.sugi@ieee.org

Kohji Dohsaka,Rikuto Tanaka
Akita Prefectural University
{dohsaka,B20P025}@akita-pu.ac.jp

Hirotoshi Taira
Osaka Institute of Technology
hirotoshi.taira@oit.ac.jp

Abstract

We address automatic citation sentence generation, which reduces the burden on writing scientific papers. For highly accurate citation senetence generation, appropriate language must be learned using information such as the relationship between the cited source and the cited paper as well as the context in which the paper cited. Although the abstracts of papers have been used for the generation in the past, they often contain extra information in the citation sentence, which might negatively impact the generation of citation sentences. Therefore, this study attempts to learn a highly accurate citation sentence generation model using sentences from cited articles that resemble the previous sentence to the cited location, thereby utilizing information that is more useful for citation sentence generation.

1 Introduction

In recent years, the use of such preprint servers as arXiv (McKiernan, 2000) has increased the amount of scientific literature. With this, we need a lot of citations to write a new paper and writing the related work section has become time-consuming. The development of automatic citation sentence generation system can support the writing of papers and relieve scientific researcher's burden on tracking and editing citations (Wu et al., 2021; Narimatsu et al., 2021). There have been several studies on citation sentence generation. Hoang and Kan (2010) constructed a keyword-based tree from the cited papers and utilized to generate citation sentences. Xing et al. (2020) used a multi-source pointer-generator network with cross attention mechanism to generate a single citation sentence for a single citation. Wu et al. (2021) used the Fution-in-Decoder (FiD) model (Izacard and Grave, 2021) to generate citation sentences for citing multiple papers, which is commonplace in real papers. They also consider differences in citation intent (Cohan et al., 2019). There are many different relationships between citing paper and the cited papers. The expression of the citation depends on what the intent of the citation is.

Citation intent such as background information, methods, and comparison of results which is important to improve the quality of citation sentence generation.

Citation sentence generation methods, that have been proposed in recent years, often use deep learning, which has the limitation of word sequence size. For that reason, most previous works have used abstracts of the citing and cited papers (Xing et al., 2020; Ge et al., 2021; Wu et al., 2021), that are relatively short to the entire paper, to recognize the relationship between them and generate the citation sentence.

A single sentence in the abstract is compact in length and merely expresses an overview of the characteristics of the study. However, citation sentences are often sentences that describe in detail the differences in characteristics between the citing and cited papers. The information in the sentences in the abstracts tends to be rather coarse to generate a description of those relationships, and this is one of the reasons for the lower quality of citation sentence generation.

On the other hand, in the task of generating sentences describing the relationship between two papers, which is different from citation sentence generation, Luu et al. (2021) used sentences in the introduction, rather than in the abstract of the paper, to generate high-quality, sentences describing the relationship between the two papers.

Inspired by this work, we propose a method to use all the sentences in the cited and citing papers. In order to reduce the input size to the neural network, our method retrieves and uses useful sentences for generating citation sentences from all the sentences in the cited paper with reference to

Proceedings of the Third Workshop on Scholarly Document Processing, pages 170–174
October 12–17, 2022.

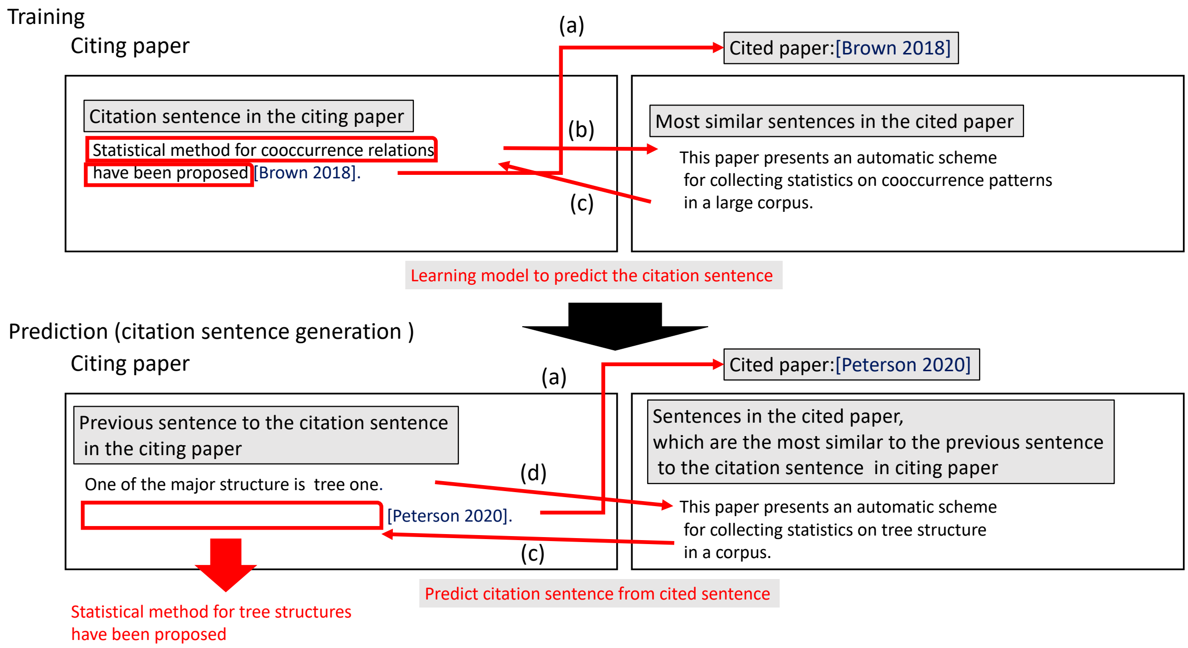

Figure 1: Overview of our method

the contents of the citing paper. The mehod finds a sentence from the cited paper, semantically similar to the previous sentence of the citation sentence to be generated, and uses it as input for generating it. Experiments with an evaluation dataset show that the method improves accuracy by about 2 points in ROUGE evaluation, compared to the method that uses only abstracts as input to generate citations.

2 Proposed Method

Citation generation is the task of generating the citation sentence to describe a cited paper under the context in a citing paper.

Figure 1 illustrates an overview of our proposed method. In the training phase, the mehod consists of three steps: a) preparing the full text of the cited papers contained in the citation sentences in the training data; b) extracting semantically similar sentences to each citation sentence from the cited papers using cosine similarity; c) learning to generate citation sentence from the semantically similar sentences.

In the prediction phase, the mehod consists of three steps: a) preparing the full text of the cited papers contained in the quoted sentences in the test data; d) extracting sentences from the cited papers that are semantically similar to the previous sentence in the target citation using cosine similarity; c) learning to generate citations from semantically similar sentences.

The major difference between training and prediction is in steps b) and d). In step b) of training phase, the system extracts sentences from the cited papers, that are similar to the citation sentence and useful for generating the citation sentence.

On the other hand, in step d) of prediction phase, the system extracts the two sentences immediately before the citation sentence, because we cannot use the citation sentence, which is the sentence itself to generate and does not exist in the phase.

To utilize the all sentences of a cited paper, excluding its abstract, the text is divided into sentences using NLTK (Loper and Bird, 2002), and we calculated the embedded representation of each sentence using SentenceBERT (Reimers and Gurevych, 2019).

In the step c), we performe fine-tuning a pre-trained model for generating citation sentences. We used T5 (Raffel et al., 2020) as a pre-trained model.

3 Experiments

We observed changes in the accuracy of the generated citation sentences by combining the citation intent, the citing paper's abstract, the citation context, the cited paper's abstract, and the cited paper's content. Then we investigated which in-

Table 1: Experimental results for each combination of inputs

Model	Citing abstract	**Citing context**	Cited abstract	**Cited content**	Citation Intent	ROUGE-1	ROUGE-2	ROUGE-L
A	✓		✓		✓	20.87	2.60	15.40
B	✓			✓	✓	21.02	2.54	14.30
C		✓	✓		✓	19.44	2.14	14.11
D		✓		✓	✓	**22.08**	**3.43**	**16.52**

formation contributes to the generation of citation sentences.

3.1 Experimental Data

We used the citation sentence generation dataset created by Xing et al. (2020) for the evaluation data. It is based on the ACL Anthology Network (AAN) corpus (Radev et al., 2013), which consists of 21,121 papers in computational linguistics and contains citation relationship information for them. The dataset is based on the assignment of pseudo-labels for all of the citations in the AAN corpus, using a model trained by 1,000 manually labeled sentences. The training data consisted of 85,652 sentences, and the test data consisted of 400 sentences. However, since we found that some of the test data were also included in the training data, we removed 103 duplicated sentences from the training data.

3.2 Experimental Settings

The input available size for the deep neural network was limited, and we could not use all sentences in the cited paper for learning to generate the citation setence. Therefore, we used the top six sentences in the cited paper, with a cosine similarity of 0.6 or more. If the number of sentences more than the threshold was less than three, we used the top three sentences. These extracted similar sentences, which were to be used as the cited paper's content,were concatenated for both training and prediction.

We used the following citation intent categories defined by Cohan et al. (2019): "Background information," "Method" and "Result comparison." Since "Result comparison" is divided into two labels, "supportive" and "not supportive," we have a total of four labels. These four citation intent categories were automatically assigned to the citation sentence by the Cohan et al. (2019) model.

We assigned a prefix token to the beginning of the text so that the citation generation model could recognize the type of data given during training. The citation intent was assigned a prefix token such as "cit_intent:".

In our experiments, we used T5-base (Raffel et al., 2020) as a pre-trained model for generating citation sentences and performed fine-tuning. We used ROUGE-1, ROUGE-2, and ROUGE-L (Lin, 2004), to calculate the abstract evaluation score.

3.3 Experimental Results

We combined the input data and show the resulting accuracy of our experiments in Table 1. We compared the two types of methods to test whether the abstracts of cited papers or their content contributed to accuracy and confirmed that the cited content only improved accuracy when combined with the citation content.

First, we compared A and B in Figure 1. A and B use the abstract as the information on the cited paper side, and A uses abstract as the information on the cited paper side, while B uses content. Compared to A , B is 0.15 points higher in the ROUGE-1 evaluation, and 0.06 points and 1.1 points lower in the ROUGE-2 and ROUGE-L evaluations. Second, we compared C and D in Figure 1. C and D use the citing context as the information on the citing paper side, and C uses abstract as the information on the cited paper side, while D uses content. Compared to C, D showed that ROUGE-1, ROUGE-2, and ROUGE-L improved by 2.64, 1.29, and 2.41 points,when the cited content was used. These results confirm that cited content alone is not particularly meaningful, and that accuracy can only be improved by using the citing and cited content.

Next, examples of the citation sentence generation results using the proposed method and a baseline method using abstracts as input, are shown in Table 2.

Our proposed method is expected to extract sen-

Table 2: Example of citation sentence generation using the proposed method

Citation intent : Background
Previous sentences to citation sentence (citation context) :
For example, whereas the first sentence of a news paper might be an effective abstract of its contents.
Of course ... identify what genre or genres a text belongs to.
Sentences in cited paper (three of sentences most similar to the citation context):
(1st) The genre of a text can also be very important
(2nd) Genres in terms of author/speaker purpose, while text types classify texts
(3rd) Which form the basis for assigning a given text to a certain genre are reflected ···
Target (ground truth):
Fortunately, there is a growing body of work on genre based text classification, including.
Baseline method's output (input both abstracts):
The resulting results are based on the results of #REFR, which is a German equivalent of the Brown corpus.
Proposed method's output (using cited paper content):
This is a problem that has been explored in previous work on genre of text categorisation.

tences that are semantically similar to the citation context in the cited paper's content. In the actual example, some similar words appear: "text," "genre," "belongings," and "assigning," indicating that keywords that are basically common to a topic.

Next we discuss a case where the most accurate citation context and the cited paper's content are used as input, based on the generation results. The proposed method's generation results show that words are generated that are synonymous with the common words discussed earlier: "genre," "text," and "categorisation." Words that are synonymous with "genre," "text," and "classification" were also generated in the actual citation sentence. The above results confirm that the characteristic keywords overlap. This suggests that the reason for the large increase in accuracy when the citation context and the cited paper's content are input as a set is that the keywords appear multiple times in both the citation context and the cited paper's content.

Next we analysed the training data by examining the proportion of words that overlap with the citations in each set of paper abstracts, citation contexts, and the cited paper's content. The results showed that the proportion of words overlapping with citations is 24% in the abstracts and 30% for the citation contexts and the cited papers's content. This is 6 points increase indicates that unnecessary information is more likely to be included in the generation of citations than in abstracts.

Finally, we discuss the generation results of our proposed method when the citation context and the cited paper's content are entered as a set, and when the baseline paper abstracts are entered. The baseline generation results are quite different compared to the actual citations that we used, because a paper's abstract summarizes an entire paper. Hence it is unclear which sentences of a given text should be focused on to generate citations. This situation resembles the results analysed above, which show that citations are more likely to contain unnecessary information.

4 Conclusion

We performed the task of generating an appropriate citation sentence from a citing paper, cited papers, and the citation context. While citation sentence generation in previous studies has been based on sentences in abstracts, we proposed citation sentence generation based on sentences in the citing paper and the cited papers. Experimental results show that our proposed method is more accurate in generating citation sentences than the conventional method of using sentences in abstracts. In the future, we will evaluate using people or other methods than ROUGE and larger citation datasets.

References

Arman Cohan, Waleed Ammar, Madeleine van Zuylen, and Field Cady. 2019. Structural scaffolds for citation intent classification in scientific publications. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 3586–3596.

Yubin Ge, Ly Dinh, Xiaofeng Liu, Jinsong Su, Ziyao Lu, Ante Wang, and Jana Diesner. 2021. BACO: A background knowledge-and content-based framework for citing sentence generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*, pages 1466–1478.

Cong Duy Vu Hoang and Min-Yen Kan. 2010. Towards automated related work summarization. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING)*.

Gautier Izacard and Edouard Grave. 2021. Leveraging passage retrieval with generative models for open domain question answering. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 874–880.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *proceedings of the ACL-04 Workshop: Text summarization branches out*, pages 74–81.

Edward Loper and Steven Bird. 2002. NLTK: The natural language toolkit. In *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*, pages 63–70.

Kelvin Luu, Xinyi Wu, Rik Koncel-Kedziorski, Kyle Lo, Isabel Cachola, and Noah A Smith. 2021. Explaining relationships between scientific documents. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*, pages 2130–2144.

Gerry McKiernan. 2000. arXiv.org: the Los Alamos National Laboratory e-print server. *International Journal on Grey Literature*, 1:127–138.

Hiromi Narimatsu, Kohei Koyama, Kohji Dohsaka, Ryuichiro Higashinaka, Yasuhiro Minami, and Hirotoshi Taira. 2021. Task definition and integration for scientific-document writing support. In *Proceedings of the Second Workshop on Scholarly Document Processing (SDP)*, pages 18–26. Association for Computational Linguistics.

Dragomir R Radev, Pradeep Muthukrishnan, Vahed Qazvinian, and Amjad Abu-Jbara. 2013. The ACL anthology network corpus. *Language Resources and Evaluation*, 47(4):919–944.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992.

Jia-Yan Wu, Alexander Te-Wei Shieh, Shih-Ju Hsu, and Yun-Nung Chen. 2021. Towards generating citation sentences for multiple references with intent control. *arXiv preprint arXiv:2112.01332*.

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 (ACL)*, pages 6181–6190.