

Multi Perspective Scientific Document Summarization With Graph Attention Networks (GATS)

Abbas Akkasi

Computer Engineering Department,
Istanbul Gelisim University,
aakkasi@gelisim.edu.tr

Abstract

It is well recognized that creating summaries of scientific texts can be difficult. For each given document, the majority of summarizing research believes there is only one best gold summary. Having just one gold summary limits our capacity to assess the effectiveness of summarizing algorithms because creating summaries is an art. Likewise, because it takes subject-matter experts a lot of time to read and comprehend lengthy scientific publications, annotating several gold summaries for scientific documents can be very expensive. The shared task known as the Multi perspective Scientific Document Summarization (Mup) is an exploration of various methods to produce multi perspective scientific summaries. Utilizing Graph Attention Networks (GATs), we take an extractive text summarization approach to the issue as a kind of sentence ranking task. Although the results produced by the suggested model are not particularly impressive, comparing them with the state-of-the-arts demonstrates the model's potential for improvement.

1 Introduction

A summary is a clear and accurate representation of the input text that distills the main ideas from the source. It is important to maintain the text's interword and inter-sentence reliance. A novel method for ascertaining an article's main objective is text summarization. The article summary assists users in rapidly determining whether a paper is pertinent to their study areas and focusing on them. Regardless of the type of documents that need to be summarized, there are two methods for automatic text summarization: extractive and abstractive. While abstractive summarization attempts to recreate the key content in a fresh way after interpreting and analyzing the text with more sophisticated techniques, extractive summarization is based on identifying important sections of the text and producing a subset of the sentences from the original text (Kadriu

and Obradovic, 2021; El-Kassas et al., 2021; Syed et al., 2021; Magdum and Rathi, 2021).

For news articles, automatic summary has recently produced impressive results; nevertheless, summarizing scholarly publications has gotten less attention (Yasunaga et al., 2019; Cohan and Goharian, 2018; Patil et al., 2022; Huang et al., 2021). Published papers differ from other sorts of material, including news, in a few key respects. They are typically longer and feature more complex subjects and technical jargon. Scientific publications are also citeable and contain citations. Additionally, these documents usually contain tables, charts, and figures, which complicates the summary process. Last but not least, another characteristic of scientific publications is that they may have unintended effects after being published.

The majority of the current research on scientific document summarization assumes only one optimal gold summary. Because creating summaries is a subjective process, having only one perfect summary makes it difficult to assess how well summarization systems are working. On the other hand, annotating several gold summaries for scientific publications can be quite expensive because it calls for specific topic experts to read and comprehend lengthy scientific documents.

As the first collaborative activity, Multi Perspective Scientific Document Summarization aims to investigate techniques for producing multi-perspective summaries. In this attempt, we proposed a model using Graph Attention Networks (GATs) with data preparation based on transformers to deal with the issue.

The remainder of this paper is organized as follows. Recent related work is presented in the next section. Sections 3 and 4 are dedicated to the model explanation and the experiments' results, respectively, and finally, the paper is ended by Section 5 as a conclusion.

2 Related Work

Although scientific document summarization has been studied for a long time, there are still many outstanding questions about how to do it effectively Paice (1980); Elkiss et al. (2008); Lloret et al. (2013). Liu and Lapata (2019) has reported on the state-of-the-arts' results in abstractive and extractive summarization in the general text domain (news). The authors used pretrained encoders to build their summarizers and provided a two-stage technique in which the encoder is fine-tuned twice, once for extractive summarizing and once for abstractive summarizing. No official model that can provide a reasonable level of data independence has been mentioned (Kadriu and Obradovic, 2021).

By choosing important passages from a text and replicating them word for word, extractive summarization creates a subset of the original text's phrases. On the other hand, an abstractive summarizer recreates crucial content in a new way after reading and analyzing the text using sophisticated natural language algorithms to create a new shorter text that offers the most important information from the original one El-Kassas et al. (2021); Patil et al. (2022); Syed et al. (2021).

From another perspective, scientific paper summarization may be classified into two types: Summarization based on *Content* or *Citation* Sefid and Giles (2022); Khurana and Bhatnagar (2022). The summarizer just accepts the content of a document as input in content-based summarizing. In citation-based summarization, along with the original paper's content, external knowledge in the form of citations is also leveraged. The community has given those citations for the paper at hand. The majority of current studies in this domain are of the second category. Nevertheless, being cited by other research works is required here, which means that newly published papers may not be accurately summarized in their initial days of publication. Qazvinian and Radev (2008) proposed one of the first models for the scientific text summarization task. They suggest a clustering method where communities are generated in the lexical network of the citation summary and sentences are retrieved from various clusters. They claimed that, for this particular issue, their method outperforms LexRank, one of the most widely used multi-document summarizing algorithms. ScisummNet, (Yasunaga et al., 2019), is a large annotated corpus for scientific paper summarization considering the papers' ci-

tations. This dataset is suitable for data-driven approaches due to its size. Besides the corpus, the author presented a graph convolutional network for the paper summarization. An et al. (2021) used the citation graph to improve the work of summarizing scientific papers. In order to produce the final abstract, summarization algorithms specifically can find relevant data from the relevant research community from the citation graph, in addition to using the document information from the original publication. Additionally, they created a novel citation graph-based model that takes into account both the features of an article and its references. SciSummpip (Ju et al., 2020), is another unsupervised paper summarization pipeline which uses a transformer-based language model for contextual text representation and PageRank for sentence selection. Cachola et al. (2020), proposed a model to generate summaries for long papers. They used high source compression in their system for creating summaries, which entails complicated domain-specific language that needs to be understood by experts. SciBERTSUM (Sefid and Giles, 2022), is another model designed for summarizing lengthy texts, such as scientific publications with more than 500 sentences. By 1) incorporating a section embedding layer to include section information in the sentence vector and 2) employing a sparse attention mechanism, which allows each sentence to pay attention to nearby sentences locally while only a small number of sentences pay attention to all other sentences globally, SciBERTSUM extends BERTSUM to long documents. Another scientific document summarization system is proposed by Mishra et al. (2022). They introduced a new approach for summarizing scientific documents that makes use of multi-objective differential evolution. Making use of citation contextualization, different important sentences are first retrieved. The idea of multi-objective clustering is used to further group these sentences. Ibrahim Altmami and El Bachir Menai (2022), summarized almost all the recent works in the domain of scientific article summarization. They categorized the work based on different factors and reported the achieved results in terms of popular evaluation metrics.

3 Proposed Model

We adopted an extractive summary technique for the MuP shared task. For this, we generate a graph for each article after selecting the three sentences

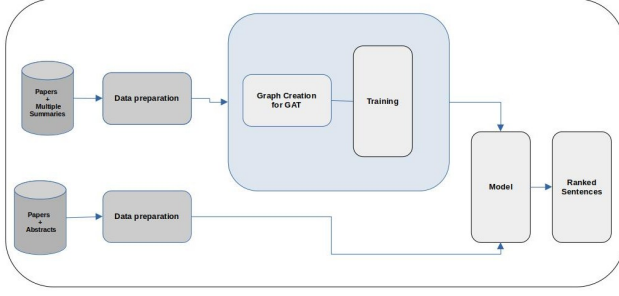


Figure 1: Suggested model for the Mup.

that are closest to the each summary sentences based on the cosine similarity between the their sentence embeddings. The oracle rank is defined as the normalized average cosine similarity score between each sentence and the provided summaries' embeddings. Figure 1 demonstrates the suggested model.

3.1 Graph Attention Networks

A particular kind of convolutional neural network called a "graph convolutional network" (GCN) may operate directly on graphs and benefit from their structural data. The fundamental tenet of GCN is that we gather feature information about each node from all of its neighbors as well as the feature itself. It resolves the issue of categorizing nodes in graphs (like citation networks) where labels are only accessible for a small portion of nodes (such as documents) (semi-supervised learning) (Zhang et al., 2019). The normalized sum of the node features of neighbors is what is produced for GCN by a graph convolution process as Formula 1.

$$h_i^{(l+1)} = \sigma \left(\sum_{j \in N(i)} \frac{1}{c_{ij}} W^{(l)} h_j^{(l)} \right) \quad (1)$$

Where $N(i)$ is the set of its one-hop neighbors, $c_{ij} = \sqrt{|N(i)|} \sqrt{|N(j)|}$ is a normalization constant based on graph structure, σ is an activation function (e.g. ReLU), and $W^{(l)}$ is a shared weight matrix for node-wise feature transformation.

The attention mechanism is a replacement for the statically normalized convolution operation in Graph Attention Networks (GATs) (Figure 2).

The equations to calculate the node embedding of layer $l+1$, $h^{(l+1)}$, from its layer l embeddings

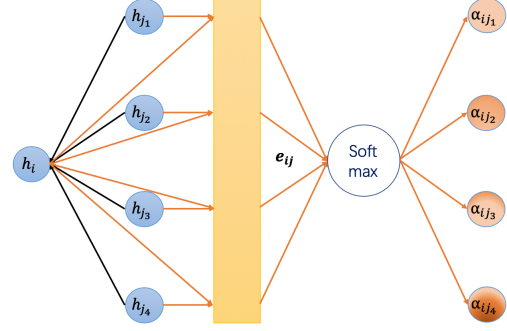


Figure 2: Graph Attention Networks, after (Veličković et al., 2018)

are listed below (Veličković et al., 2018).

$$\begin{aligned} z_i^{(l)} &= W^{(l)} h_i^{(l)}, \\ e_{ij}^{(l)} &= \text{LeakyReLU}(\alpha^{(l)T} (z_i^{(l)} || z_j^{(l)})), \\ \alpha_{ij}^{(l)} &= \frac{\exp(e_{ij}^{(l)})}{\sum_{k \in N(i)} \exp(e_{ik}^{(l)})}, \\ h_i^{(l+1)} &= \sigma(\sum_{j \in N(i)} \alpha_{ij}^{(l)} z_j^{(l)}) \end{aligned}$$

The main concept behind using GAT was to choose the most essential phrases based on inter-sentence relationships within the articles, using the attention mechanism to concentrate on more effective sentences.

3.2 Evaluation

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) was presented as an automated evaluation approach in 2003. It is a series of metrics based on the similarity of n-grams¹. Other ROUGE scores include ROUGE-L, which is a longest common sub-sequence measure, and ROUGE-SU4, which is a bigram measure that allows at most four unigrams inside bigram components to be skipped (ROUGE, 2004). In this task, the intrinsic evaluation using the ROUGE-1, -2, and -L metrics is applied. The final ranking also takes into account the average of the ROUGE-F scores achieved against the various summaries.

4 Experiments

The datasets made available by the task organizers were used for all of the experiments. We made use of two-layers GATs implemented with the Pytorch-Geometrics library with 10 epochs and a learning

¹A sub-sequence of n words from a particular text is referred to as an n-gram.

rate of 0.0001 to train the suggested model. Applying the trained model on new data, the highly ranked sentences are selected as

4.1 Data preparation

In the first step, we prepared data to be used for graph generation. For this purpose, for each sentence of available summaries, the most similar 3 sentences to each summary sentence from the input article are taken as the graph nodes. The cosine similarity between the embeddings of body sentences and provided summary sentences is used as similarity metrics. Two pretrained transformer-based language models, SPECTER (Cohan et al., 2020) and all_mpnnet_v2², are utilized to generate the sentence embeddings. The duplicate sentences are also ignored. The sentence embedding, which has a length of 768, is regarded as a node feature for each node. In addition, the dot product between the relevant pairs of nodes is used as a feature for the edges.

4.2 Results

Tables 1 and 2 demonstrate the obtained results by applying the trained model on development and test datasets respectively.

The tables show that the results on development data are marginally superior to the test data. Since for test data, there was no available summaries, we made use of the abstract's sentences as provided summaries in graph sentence selection and graph creation processes. Lower test data findings could be attributed to this. We experimented with SAGE, GCN, and other forms of graph neural networks in addition to GATs, but the results were not any better than those that had already been reported.

5 Conclusion

In this study, we used Graph Attention Networks to perform the Multi-Perspective Scientific Documents summary problem while adhering to the extractive summarization methodology. We first chose the three sentences from the input article that most closely resembled each summary sentence to produce the graphs for our node ranking task. Then, using each selected sentence as a node, the sentence embedding produced by the pretrained transformer-based language model is taken as the node features, and the dot product between the pairs of nodes is taken as the corresponding edge feature. Because

of the discrepancy between the results published by other teams and the ones obtained by the suggested model, it can be inferred that preprocessing techniques and the use of external knowledge may improve the results.

References

- Chenxin An, Ming Zhong, Yiran Chen, Danqing Wang, Xipeng Qiu, and Xuanjing Huang. 2021. Enhancing scientific papers summarization with citation graph. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 12498–12506.
- Isabel Cachola, Kyle Lo, Arman Cohan, and Daniel S Weld. 2020. Tldr: Extreme summarization of scientific documents. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4766–4777.
- Arman Cohan, Sergey Feldman, Iz Beltagy, Doug Downey, and Daniel S Weld. 2020. Specter: Document-level representation learning using citation-informed transformers. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2270–2282.
- Arman Cohan and Nazli Goharian. 2018. Scientific document summarization via citation contextualization and scientific discourse. *International Journal on Digital Libraries*, 19(2):287–303.
- Wafaa S El-Kassas, Cherif R Salama, Ahmed A Rafea, and Hoda K Mohamed. 2021. Automatic text summarization: A comprehensive survey. *Expert Systems with Applications*, 165:113679.
- Aaron Elkiss, Siwei Shen, Anthony Fader, Güneş Erkan, David States, and Dragomir Radev. 2008. Blind men and elephants: What do citation summaries tell us about a research article? *Journal of the American Society for Information Science and Technology*, 59(1):51–62.
- Luyang Huang, Shuyang Cao, Nikolaus Parulian, Heng Ji, and Lu Wang. 2021. Efficient attentions for long document summarization. *arXiv preprint arXiv:2104.02112*.
- Nouf Ibrahim Altmami and Mohamed El Bachir Menai. 2022. Automatic summarization of scientific articles: A survey.
- Jiaxin Ju, Ming Liu, Longxiang Gao, and Shirui Pan. 2020. Monash-summ@ longsumm 20 scisummpip: An unsupervised scientific paper summarization pipeline. In *Proceedings of the First Workshop on Scholarly Document Processing*, pages 318–327.
- Kastriot Kadriu and Milenko Obradovic. 2021. Extractive approach for text summarisation using graphs.
- Alka Khurana and Vasudha Bhatnagar. 2022. Investigating entropy for extractive document summarization. *Expert Systems with Applications*, 187:115820.

²https://www.sbert.net/docs/pretrained_models.html

Table 1: Results on Development data

	r1_f	r1_r	r2_f	r2_r	rL_f	rL_r	Average
	35.46	35.08	9.53	10.42	19.63	22.27	21.96

Table 2: Results on Test data

	r1_f	r1_r	r2_f	r2_r	rL_f	rL_r	Average
	33.85	38.05	7.40	8.33	17.74	20.13	19.66

- Yang Liu and Mirella Lapata. 2019. Text summarization with pretrained encoders. *arXiv preprint arXiv:1908.08345*.
- Elena Lloret, María Teresa Romá-Ferri, and Manuel Palomar. 2013. Compendium: A text summarization system for generating abstracts of research papers. *Data & Knowledge Engineering*, 88:164–175.
- PG Magdum and Sheetal Rathi. 2021. A survey on deep learning-based automatic text summarization models. In *Advances in Artificial Intelligence and Data Engineering*, pages 377–392. Springer.
- Santosh Kumar Mishra, Naveen Saini, Sriparna Saha, and Pushpak Bhattacharyya. 2022. Scientific document summarization in multi-objective clustering framework. *Applied Intelligence*, 52(2):1520–1543.
- Chris D Paice. 1980. The automatic generation of literature abstracts: an approach based on the identification of self-indicating phrases. In *Proceedings of the 3rd annual ACM conference on Research and development in information retrieval*, pages 172–191.
- Priyadarshini Patil, Chandan Rao, Gokul Reddy, Ritteesh Ram, and SM Meena. 2022. Extractive text summarization using bert. In *Proceedings of the 2nd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications*, pages 741–747. Springer.
- Vahed Qazvinian and Dragomir Radev. 2008. Scientific paper summarization using citation summary networks. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 689–696.
- Lin CY ROUGE. 2004. A package for automatic evaluation of summaries. In *Proceedings of Workshop on Text Summarization of ACL, Spain*.
- Athar Sefid and C Lee Giles. 2022. Scibertsum: Extractive summarization for scientific documents. *arXiv preprint arXiv:2201.08495*.
- Ayesha Ayub Syed, Ford Lumban Gaol, and Tokuro Matsuo. 2021. A survey of the state-of-the-art models in neural abstractive text summarization. *IEEE Access*, 9:13248–13265.
- Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph attention networks. In *International Conference on Learning Representations*.
- Michihiro Yasunaga, Jungo Kasai, Rui Zhang, Alexander R Fabbri, Irene Li, Dan Friedman, and Dragomir R Radev. 2019. Scisummnet: A large annotated corpus and content-impact models for scientific paper summarization with citation networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 7386–7393.
- Si Zhang, Hanghang Tong, Jiejun Xu, and Ross Maciejewski. 2019. Graph convolutional networks: a comprehensive review. *Computational Social Networks*, 6(1):1–23.