

Article

Not peer-reviewed version

Abstractive Summarization Model for Summarizing Scientific Article

[Mehtap Ülker](#)^{*} and Ahmet Bedri Özer

Posted Date: 16 May 2024

doi: 10.20944/preprints202405.1123.v1

Keywords: Text summarization; Abstractive method; SciBERT; ScilE; Graph transformer



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Abstractive Summarization Model for Summarizing Scientific Article

Mehtap ÜLKER * and A. Bedri ÖZER ¹

¹ Computer Engineering Department, Firat University, Elazig, Turkey; bedriozer@firat.edu.tr

* Correspondence: m.ulker@firat.edu.tr

Abstract: The rapid growth in scientific article publications allows us to access articles as soon as possible. Therefore, automatic summarization systems (ATSs) are widely preferred. In most studies, the entire source document is expected to be summarized, just as it would be summarized by a human. Summarizing long articles, such as scientific articles, is quite difficult due to token restraint and extraction of scientific words. To address this problem, a novel Graph-Based Abstractive Summarization (GBAS) model is proposed, which is a novel scientific text summarization model based on SciBERT and the graph transformer network (GTN). The document's integrity is maintained since the SciIE system uses the graph structure to create a terminology-based document structure. Therefore, long documents are also summarized. The proposed model is compared with baseline models and human evaluation. Human evaluation results show that the results of the proposed model are informative, fluent, and consistent with the ground-truth summary. The experimental results indicate that the proposed model outperforms baseline models with a 37.10 and 34.96 ROUGE-L score.

Keywords: Text summarization; Abstractive method; SciBERT; SciIE; Graph transformer

1. Introduction

Many researchers from different research areas publish scientific papers to contribute to the scientific community. The massive flow of scientific articles makes accessing the required information more challenging. Researchers need to get a handle on salient information without reading the entire document. Therefore, the tendency toward automatic text summarization has started to grow. Automatic text summarization (ATS) presents the content of the document as quickly and concisely as possible while preserving the original document's integrity. ATS can be performed with an extractive or abstractive approach depending on the selection, combination, or paraphrasing of salient sentences in the source document [1]. Extractive summarization generates a summary from sentences that best express the main idea in the source document [2], whereas abstractive summarization generates rewritten summaries with sentences or words that differ from the source document [3]. In this respect, the abstractive approach generates summaries that are similar to human-written summaries.

The rapid development of deep learning techniques has resulted in great advances in abstractive summarization methods [4,5]. However, it is still a challenge to intelligently summarize scientific articles via traditional neural summarization methods. The main reasons are: (I) These methods are trained for task-specific applications on public datasets consisting of news articles, blog posts, tweets, etc. (II) The data presented in scientific articles is quite a complex process because it requires knowledge to identify new technologies and their connections [6]. Therefore, information extraction (IE) systems may be preferred in practice to automatically determine these relations. However, the annotations of the relations extracted by these methods are not sufficient for a scientific summary of the text [7,8]. (III) In current studies [4], it has been mentioned that LSTM and attention-based models cannot handle hierarchical structures. In this paper, considering all the above problems, a novel scientific text summarization model based on SciBERT and the graph transformer network (GTN) has been proposed that generates abstracts from the introduction section of scientific articles. First, entity, co-reference, and relation annotations were extracted from the source document with the Scientific Information Extractor (SciIE) to handle the hierarchical structure of the document [7].

- Second, a knowledge graph has been constructed with this information. The output of the last hidden state of SciBERT was used to encode the introduction section. Finally, GTN [8] was performed to summarize text from the knowledge graph. The main contribution of this paper is:
- A novel model has been proposed for the summarization of scientific articles consisting of SciBERT trained on a large corpus of scientific text and a graph trans former that benefits from the relational structures of the knowledge graph without linearization or hierarchical constraints.

2. Related Work

Text summarization approaches are performed by directly extracting salient sentences from the source documents or rewriting these sentences with words that differ from the source documents [9,10]. Previous researchers have focused on extractive methods to summarize scientific articles because abstractive methods are more difficult and complex because they require advanced NLP techniques. These studies are summarized in Table 1. In this study [3], they proposed a generic summarizer model that is language-independent and based on a quantum-inspired approach for the extraction of important sentences. In this study [11], they proposed a graph-based framework that can also be applied to scientific articles without any domain or language constraints. This modelutilizes the advantages of graph-based, statistical based, semantic-based, and centrality-based methods. In this study [12], they proposed a regression-based model to high light salient sentences in scientific articles. They experimented on three different scientific datasets (CSPubSum, AIPubSum, and BioPubSum) to demonstrate the effective ness of their method. In this study [13], they constructed a large-scale manually annotated dataset (SciSummNet) for summarizing scientific articles. In addition, a hybrid summarization model was proposed. The effectiveness of their corpus on this model and the data-driven neural models was evaluated. In this study [14], they presented a new model for summarizing scientific articles by inspiring SummPipon [15]. In this study [16], they constructed a novel corpus (SciTLDR) consisting of scientific papers related to the computer science domain. In addition, a novel model (CATTS) was proposed to evaluate their corpus. The proposed model is appropriate for both extractive and abstractive methods.

Table 1. The summary of the scientific text summarization literature.

Ref.	Target	Methodology	Dataset	Metric	Performance
[3]	Extractive	Modified quantum-inspired genetic algorithm	DUC 2005 DUC 2007	Rouge	DUC 2005: R1:47.67, R2:12.87, RSU4: 18.85 DUC 2007: R1:41.06, R2:08.98, RSU4: 14.72
		Graph structure, statistical-based, semantic-based, and centrality-based methods	DUC 2001 DUC 2002		DUC 2001: R1:51.37, R2:27.16, RL:47.36, RSU4: 25.65 DUC 2002: R1:53.37, R2:28.58, RL:49.79, RSU4: 27.65
[12]	Extractive	Word2vec, Decision Tree, Random Forest, Multi-Layer	CSPubSum AIPubSum BioPubSumm	Rouge	CSPubSumm: RL:31.6 AIPubSum: RL:28.00 BioPubSumm: RL:28.90

		Perceptron, Gradient Boosting The graph- based multi- document			
[13]	Extractive	summarization, Reference paper, Graph Convolutional Network	CL- SciSumm SciSummNet	Rouge, Human Evaluation	SciSummNet: R2:29.30 R3:24.65 RSU4:18.56 CL-SciSumm: R2:18.46, R3 12.77 RSU4:12.21
[14]	Extractive Abstractive	SciBERT, PageRank	LongSumm	Rouge, Human Evaluation	R1:40.90, R2:9.52, RL:15.47
[16]	Extractive Abstractive	BART, Shuffled Data	SciTLDR	Rouge, Human Evaluation	R1:43.8, R2:21.3, RL:35.5 R1:31.7, R2:11.1 RL: 25.0
[5]	Abstractive	Graph network, auto- regressive generator, knowledge domain	CORD-19 PubMed	Rouge	CORD-19: R1:33.68, R2:22.56, RL:32.84 PubMed: R1:33.03 R2:13.51 RL:29.30
[17]	Abstractive	Seq2seq model, AraVec based word2vec	CNN/Daily Mail, Specific dataset	Rouge, Human Evaluation	R1:38.6, R1-Noorder:46.5, R1-stem: 52.6, R1- context:58.1
[4]	Abstractive	Graph Attention Networks, BioBERT, SciBERT Multiple timescales	CORD-19	Rouge, Human Evaluation	R1:44.56, R2:18.89, RL: 36.53
[18]	Abstractive	gated recurrent unit (MTGRU), The pointer- generator- coverage network	CNN/Daily Mail, Specific dataset	Rouge	CNN/Daily Mail: R1:40.94, R2:18.14, RL:38.57 Specific dataset: R1:56.91, R2:37.48, RL:54.02

Abstractive methods are closer to reality in terms of generating a summary by thinking human-like. Recently, re searchers have focused on it for generating scientific sum maries [9,10]. In this study

[5], they presented a graph network-based model based on a sentence-level denoiser and an auto-regressive generator. To demonstrate the effectiveness of their model, PubMed and CORD-19 datasets containing scientific articles in the biomedical domain were used. In this study [17], they proposed a sequence-to-sequence-based model with three encoders and one decoder. In addition, they proposed novel evaluation metrics, namely ROUGE1 NOORDER, ROUGE1-STEM, and ROUGE1-CONTEXT. In this study [4], they presented a SciBERT-based summarization model to summarize scientific articles related to COVID-19. This model consists of a graph attention network and a pre-training language model (SciBERT). To evaluate the proposed model, the CORD-19 (COVID-19 Open Research Dataset) consisting of scientific articles was used. In this study [18], they presented a novel model consisting of timescale adaptation over the pointer-generator-coverage network. It has been mentioned that this model is successful in summarizing long articles.

Pre-trained language models (PTLMs) have made significant progress in abstractive methods and many NLP tasks. The main aim of this study is to analyze the relations at the sentence/token level with large-scale corpora. Most researchers have proposed many language models to enhance specific NLP tasks. These are performed through two basic strategies: feature-based and fine-tuning. The feature-based approach requires task-specific architectures with pre-trained representations as additional features. In the fine-tuning approach, a classification layer must be added to the pre-trained model. Fine-tuning approaches are widely preferred to enhance the quality of the generated summaries in ATS [19,20].

PTLMs are preferred in general-domain text summarization tasks and have achieved successful results. The BERT [20] model was constructed to pre-train deep bidirectional representations of unlabeled text. This model has only an encoder. Therefore, it is stated that it cannot be suitable for abstractive approaches. To address this problem, many researchers have proposed novel models based on BERT. BERTSUMABS [21] is a model that consists of an encoder and a decoder. GSUM [22], based on a neural encoder decoder, is a model that takes several types of external guidance as input text. The Text-To-Text Transfer Transformer model(T5) [23], based on encoder-decoder architecture, aims to generate a novel text from the text it receives as input. Refactor [24] is a model consisting of a two-stage training process to identify candidate summaries from both document sentences and different base model outputs. To obtain semantic results from scientific articles, the researchers have focused on the SciBERT model, which has the same architecture and configuration as BERT. SciBERT is a model with a maximum sequence length of 512 tokens pre-trained with large-scale scientific papers (1.14M papers) collected from semantic scholars related to different disciplines [4,6].

GTN is a model designed to learn node representation and identify the relations between disconnected nodes in graph structures [25]. In most studies related to text summarization. [4,8,26–29], Graph Attention Transformer (GATs) are widely used among graph neural network-based approaches. GATs [30] are models with an attention-based architecture constructed to operate data in a graph structure. The main aim is to find the representations of each node in the graph by adopting the attention mechanism. With GATs, the hierarchical structure of the document can be handled as a whole. In addition, meaningful relationships can be revealed between sentences, tokens, or entities through the preservation of the global context [31].

According to [4], GATs are successful in representing word co-occurrence graphs because they utilize a masked self-attention mechanism to capture dependency between neighbors and prevent information flow between disconnected nodes. According to [26], GATs can extract the hierarchical structure of a document simultaneously as tokens, sentences, paragraphs, and documents. In addition, it provides consistency with the multi-head attention module in the BERT model. According to [27], the advantage of GAT is that it can enhance the impact of the most salient parts of the source document. According to [28], GATs can effectively capture the content in the source document through propagating contextual information [29] proposed mix-order graph attention networks for handling indirectly connected nodes inspired by the traditional model of GATs. According to [8], the use of self-attention in GATs constrains the vertex updates of information from adjacent nodes, despite eliminating the deficiencies of previous methods based on graph convolutions. Therefore, a graph transformer encoder built on the GATs architecture was proposed. It provides a more global contextualization of each vertex with a transformer-style architecture. However, GATs [29] have not achieved as many efficient results as GTN. In this paper, a graph-based abstractive summarization

(GBAS) approach consisting of three stages was proposed, inspired by the SciBERT and graph transformer, to generate a summary from the introduction sections of the source papers.

3. Proposed Model

The framework of the proposed model is illustrated in Figure 1. The proposed model consists of five stages: dataset preparation, introduction encoder, graph construction, graph encoder, and summary decoder. First, the dataset is prepared with the introduction section and abstract of the scientific articles and the features extracted through the SciIE system. Second, word embedding is obtained with the output in the last layer of SciBERT from the introduction section. Then, a knowledge graph is constructed with features. The knowledge graph is encoded with a graph transformer. In the last stage, a summary is obtained from the knowledge graph. These are explained in detail below.

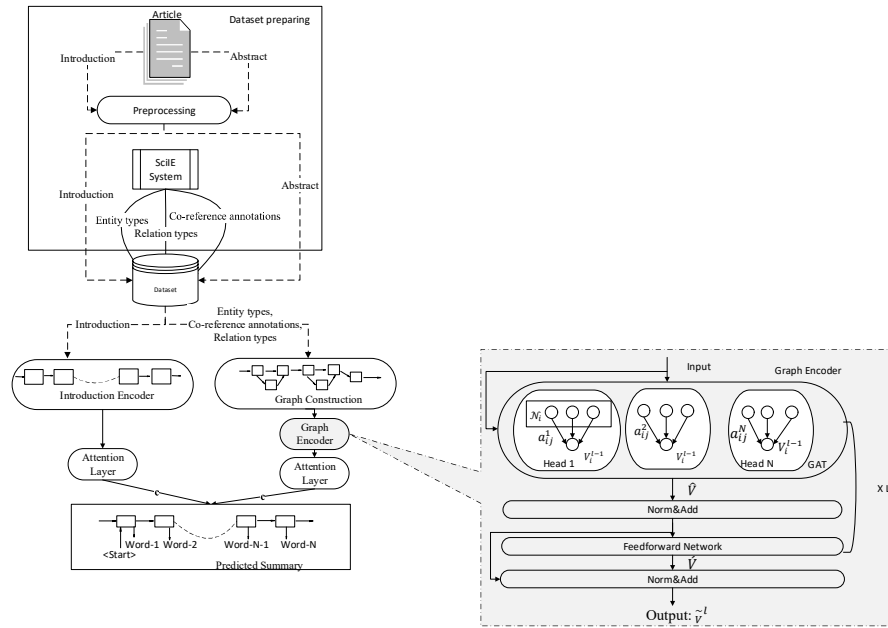


Figure 1. The overview of the proposed model.

Second, word embedding is obtained with the output in the last layer of SciBERT from the introduction section. Then, a knowledge graph is constructed with features. The knowledge graph is encoded with a graph transformer. In the last stage, a summary is obtained from the knowledge graph. These are explained in detail below.

3.1. Dataset

In the current research, abstractive summaries are generated from the title or full text of scientific articles [8,32]. Among the sections, the introduction is neither as short as the title nor as long as the full text. This section contains necessary and sufficiently salient information regarding the purpose and scope of the article. Therefore, it is foreseen within the scope of this study that the summaries generated from this section can improve the performance of scientific text summarization. The main aim of this paper is to generate a scientific summary from the Introduction section. First, scientific articles up to April 2022 from the arXiv website were crawled containing current topics related to computer science such as "fingerprint", "image processing", "natural language processing", "cyber security" and "machine learning". The dataset properties are given in Table 2. Considering the following factors, the SciIE system has been preferred for extracting salient information from scientific articles

Table 2. The properties of the datasets.

Dataset	Doc. Size	Target Avg. Word	Source Avg. Word	Avg. Entities	Avg. Relation
SciSummNet	710	132.90	338.18	13.35	6.36
SciTLDR	1548	166.63	3036.71	15.29	10.20
ArxivComp	16.807	158.24	549.06	16.14	5.92

- The most relation IE systems [33,34] in the scientific domain are designed to obtain these within sentences. However, SciIE makes it possible to extract information by taking it into account across sentences [7].
- The SciIE system is designed to identify six entity types (task, method, metric, material, other-scientific term, and generic), seven relationship types (compare, part-of, conjunction, evaluate-for, feature-of, used-for, and hyponym-of), and co-reference annotation is used to obtain entity types and relations annotations.

3.2. The Graph-Based Abstractive Summarization Model: (GBAS)

Introduction Encoder: The GBAS model generates a summary from the introduction section. The pre-trained language model (SciBERT) was used as the introduction encoder. SciBERT has a multi-transformer architecture [20].

In this study, the introduction section was tokenized. To obtain the corresponding sequences given in the samples as word embedding, we used the output of the last hidden state in SciBERT. Thus, word embedding was obtained for the introduction section of each article as follows:

$$S = \{\underbrace{w_{11}, w_{12}, \dots, w_{1k}}_{word_1}, \underbrace{w_{21}, w_{22}, \dots, w_{2l}}_{word_2}, \dots, \underbrace{w_{n1}, w_{n2}, \dots, w_{nm}}_{word_n}\} \quad (1)$$

where S represents source word sequences. The sub-word of each word is illustrated with w_{nm} , in which n and m indicate the order of words and sub-word order, respectively. In response to S, the target word sequence is obtained with $T = \{t_1, t_2, \dots, t_p\}$.

Graph Construction: The graph transformer operates the graph as an input. To construct the graph, the entities, their relations, and co-reference annotations for each introduction section were established as outlined in the dataset preparation. Then, it benefited from the graph preparation process of the GraphWriter model [8] based on [35]. Differently, the graph was constructed by considering the introduction and abstract sections together. According to the document structure given as an example in Table 3, the graph construction stage is as follows:

Table 3. The example of source document.

Type	Example
"relations types:"	["USED-FOR", "CONJUNCTION", "FEATURE OF", "PART OF", "COMPARE", "EVALUATE FOR", "HYPONYM-OF"]
"abstract entities":	"entities": ["brain tumor segmentation", "treatment outcome evaluation", "deep learning techniques", "brain tumor segmentation method", "neural networks FCNNs", "Conditional Random Fields CRFs", "unified frame work", "appearance and spatial consistency", ..., "T2 scans"] "relations": ["brain tumor segmentation- CONJUNCTION- treatment outcome evaluation", "deep learning techniques- USED-FOR- brain tumor segmenta tion method", "neural networks FCNNs- CONJUNC TION- Conditional Random Fields CRFs", "2D image patches- USED-FOR- deep learning based segmentation model", "2D image patches- CONJUNCTION image slices", ..., "T2 scans- USED-FOR- those"]
"document entities":	["brain tumor segmentation", "cancer diagnosis treat ment planning", "treatment outcome evaluation", "manual segmentation of brain tumors", "semiautomatic or automatic brain tumor segmentation methods", "brain

	tumor segmentation studies", "gliomas", "magnetic resonance imaging mri", "segmentation of gliomas", "appearance", "gliosis", "mri data gliomas", "fuzzy bound aries",..., "knee cartilage segmentation"]
"document relations":	["cancer diagnosis treatment planning- CONJUNCTION- treatment outcome evaluation", "manual segmentation- USED-FOR- manual segmentation of brain tumors", "mri data- USED-FOR- segmentation of gliomas", "discriminative model based methods CONJUNCTION- discriminative model", "probabilistic image atlases- USED-FOR- brain tumor segmentation problem", "outlier detection problem- USED-FOR- brain tumor segmentation problem", "discriminative model based methods- USED-FOR- tumor sementation problem",..., "back propagation algorithms- CONJUNCTION-mrfs crfs"]

- The "abstract-relations" and "document-relations" arrays are rebuilt according to the "relations types", "abstract-entities", and "document-entities" indexes.
- For instance, the "abstract-relations" array is converted to "0 1 1", according to the example of "brain tumor segmentation – CONJUNCTION – treatment outcome evaluation", so that the index of brain tumor segmentation is "0", the index of CONJUNCTION is "1" and the index of treatment outcome evaluation is "1".
- For instance, the "abstract-relations" array is converted to "2 0 3", according to the example of "Deep learning techniques – USED-FOR – brain tumor segmentation-method", so that the index of deep learning techniques is "2", the index of USED-FOR is "0" and the index of brain tumor segmentation method is "3".
- Entity names that could not be extracted because of spelling errors were excluded from the novel array. Accordingly, the "abstract-relations" array is [0 1 1; 2 0 3; 4 1 5; 9 0 8; 9 1 10; 10 0 8; 12 0 8; 10 0 8; 13 0 81; 9 0 17; 21 0 17; 23 0 22; 28 0 22; 28 1 29; 29 0 22].
- The same transformation was performed in the introduction section.
- The array of "document-relations" is [1 1 2; 19 0 18; 15 0 21; 25 0 26; 25 1 2; 28 0 26; 25 0 30; 31 0 5; 34 0 5; 34 0 36; 37 6 36; 39 6 36; 40 0 41; 49 1 50; 54 0 53; 55 6 54; 60 0 52; 63 0 66; 66 0 67; 34 0 75].
- As a result, a comprehensive graph was constructed by combining the new transformation array of "abstract-relations" and "document relations"

$$G = G_1 \cup G_2 \quad (2)$$

$$\text{All-entities} = \text{abstract-entities} \cup \text{document-entities} \quad (3)$$

$$\text{All-Relations} = \text{abstract-relations} \cup \text{document-relations} \quad (4)$$

Figure 2 shows the graph constructed for the introduction and abstract sections. G_1 and G_2 denote global nodes. G_1 and G_2 represent global nodes and contain entity name lists in Table 3. To maintain the flow of information in the graph, the global node is connected to all nodes. This node is used as the start of the decoder. Nodes consist of entity names. Each labeled edge is replaced by two nodes. One of them represents the forward direction of the relations (Rel.), and the other represents the reverse direction of the relations. A novel node is connected to nodes consisting of entities (Ent.), preserving the directions of previous edges.

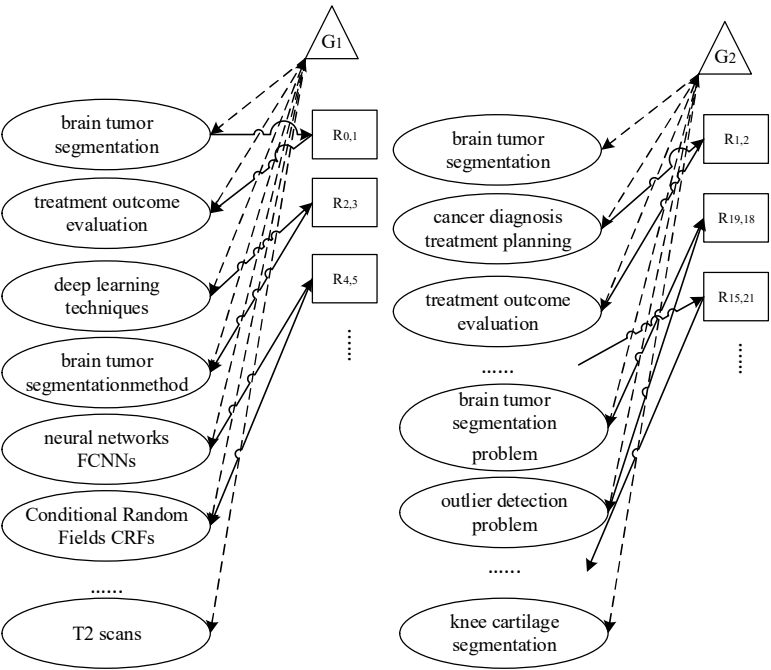


Figure 2. The example of relation graphs a) Graph of the abstract section b) Graph of the introduction section.

Within the scope of the study, entity names (abstract-entities, document entities), relations (abstract-relations, document relations) and global nodes (G_1 and G_2) in Equations 2, 3, and 4 were combined to create the comprehensive graph in Figure 3.

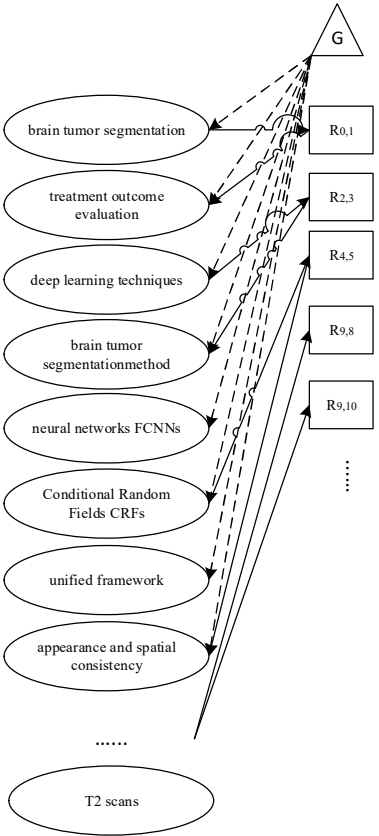


Figure 3. Abstract-introduction section graph model.

Graph Encoder: To encode the graph structure, a graph transformer architecture is based on GATs. This model uses the N-head self-attentional system, as shown in Figure 4. In this model, each vertex is contextualized by attending to another connected vertex in the graph. For the calculation of the N independent attentions, equation (5) is applied.

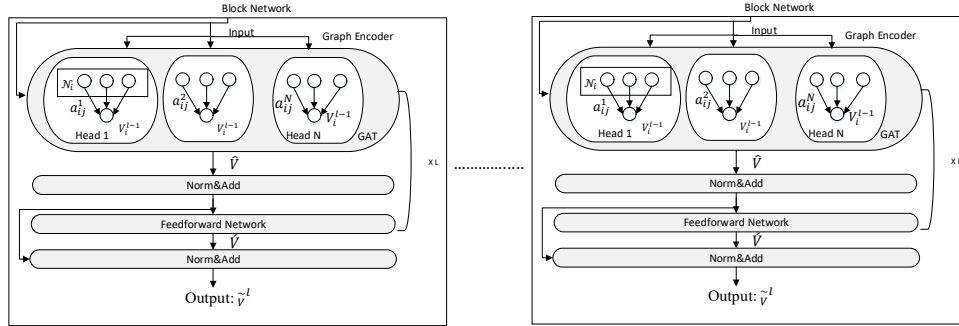


Figure 4. Graph encoder model [8].

$$\hat{v}_i = v_i + \parallel_{n=1}^N \sum_{j \in N_i} a_{ij}^n W_V^n v_j \quad (5)$$

$$a_{ij}^n = a^n(v_i, v_j) \quad (6)$$

where \parallel and $W_K W_Q \in \mathbb{R}^{d \times d}$ indicate the concatenation of the N head attention, the neighborhood of v_i in the graph structure, and attention mechanisms in equation (7). For each head, independent transformations are learned with α respectively.

$$a(q_i, k_j) = \frac{\exp((W_K k_j)^T W_Q q_i)}{\sum_{z \in N_i} \exp((W_K k_z)^T W_Q q_i)} \quad (7)$$

In this model, equations (8) and (9) are applied L times for each block: Where FNN(x) donates a two-layer feedforward network. As a result, each vertex encoding indicates with $V^L = [V_i^L]$ which consists of relation, entities, and global vertex.

$$\tilde{v}_i = \text{Layernorm}(v'_i + \text{Layernorm}(\hat{v}_i)) \quad (8)$$

$$v'_i = \text{FNN}(\text{Layernorm}(\tilde{v}_i)) \quad (9)$$

Summary Decoder: In this stage, the content vectors c obtained from the graph and introduction sequences are calculated by adding a decoder hidden state h_t at each t timestep. While vertex embedding V^L is used for graph sequences, and T is used for the introduction sequence. This is given in equation (10). In the last stage, it is given as input to RNN with h_t by concatenating the context vectors namely c_t from both graphs and the introduction.

$$c = \begin{cases} h_t + \parallel_{n=1}^N \sum_{j \in V} \alpha_j^n W_G^n V_j^L, & \text{graph sequence} \\ a_j = \alpha(h_t, V_j^L) \\ h_t + \parallel_{n=1}^N \sum_{j \in T} \alpha_j^n W_G^n T_j, & \text{introduction sequence} \\ a_j = \alpha(h_t, T_j) \end{cases} \quad (10)$$

To calculate the probability of copying from the input, it is applied in equation (11) [36]. Taking into account this equality, the probability of the final next token is given by equation (12).

$$c = \text{softmax}(W_{copy}[h_t \parallel c_t] + b_{copy}) \quad (11)$$

$$p * a^{copy} + (1 - p) * a^{vocab}, \quad (12)$$

where a^{copy} and a^{vocab} refer to the probability distribution over entities and input tokens, and remaining probability respectively. In a^{copy} , $a([h_t, \parallel c_t], x_i)$ is performed for $x_i \in V \parallel T$.

4. Experiments

To ensure a fair comparison of the proposed model performance, GBAS was performed using the default hyper-parameter settings for the three datasets. SciBERT was used [6] by following the default hyper-parameter settings. A graph transformer [8] based on the default parameter setting that activation function, dropout, attention head, graph layer (L), and feedforward network in block layer size were PReLU, 3, 6, 4, 2000, respectively, was used to encode graphs. To decode, a beam search was used with size 4. At the last stage, the words mentioned as < unk > were removed from the generated summaries.

4.1. Evaluation Metric

To evaluate the similarity of the references summary (human created) and generated summary by the GBAS model, the Rouge metric (Recall Oriented Understudy of Gisting Eval uation) was used. ROUGE calculates n gram-based recall whether overlaps the abstract written by the author and the summarization generated by the models [37]. The most commonly preferred metrics are ROUGE-1, ROUGE-2, and ROUGE-L, which overlap the uni-gram, bi-gram, and the longest common sequences (LCS) in the word level respectively. The calculation of the ROUGE-N is shown in equation (13):

$$ROUGE - N = \frac{\sum_{S \in H} \sum_{gram_n \in M} Count_{match}^{gram_n}}{\sum_{S \in H} \sum_{gram_n \in M} Count(gram_n)} \quad (13)$$

where N, $Count(gram_n)$, $Count_{match}^{gram_n}$ are represented as the length of the N-gram, the count of n-grams in the references, and the maximum number of matching words in the candidate summary, respectively.

The ROUGE-L is based on the overlap of the longest common sequences between the candidate sentence and the reference sentence. The precision metric refers to the ratio of the number of common sentences in the candidate and reference summaries to the sentences in the candidate summaries. The recall metric refers to the ratio of the number of common sentences in the candidate and reference summaries to the sentences in the reference summaries.

4.2. Baseline Methods

To evaluate the performance of the proposed model GBAS, experiments were conducted in a baseline approach for both abstractive and extractive methods. These methods are summarized as follows:

- TextRank [38] is a graph-based method applied to ex tractive methods. The number of similar words between two sentences is calculated as similarity.
- LexRank [39] is a graph-based method that calculates based on cosine similarity.
- LSA [40] is an algebraic method for extractive methods. This model is performed in three stages: input matrix creation, singular value decomposition (SVD), and sentence selection.
- The performance of the proposed model was compared with a fine-tuned T5 model pretrained on 4515 English news articles¹.
- Billsum [41] is a T5-based fine-tuned model on the Billsum dataset which consists of scientific articles.
- BART [31], which uses the standard sequence-to-sequence Transformer architecture, is effective when the model is fine-tuned for the abstractive summarization task. The performance of the proposed model was compared to The BART fine-tuned model is trained to summarize scientific articles².
- GraphWriter [8] is a graph-based tool that generates a summary from the title for abstractive methods.

5. Discussion

The experiments were conducted with the above baseline methods on the SciTLDR and SciSummNet datasets and ArxivComp. The results of the experiment are shown in Tables 4–6.

Graph-based methods are highly successful among abstractive and extractive methods. Because Seq2Seq models are not very good at transferring information in long-term sequences, abstractive

¹ <https://huggingface.co/mrm8488/t5-base-finetuned-summarize-news>

² <https://huggingface.co/sana-ngu/bart-base-finetuned-summarize-scientific-articles>

methods concentrate on attention-based methods. Attention-based models fix this issue, but when taking the document’s hierarchical structure into account, they may result in a loss of semantic integrity. Therefore, graph-based methods are superior to other methods because they ensure the integrity of the document.

Table 4. The evaluation results on the ArxivComp dataset.

Model	Rouge-1	Rouge-2	Rouge-L
TextRank	28.52	9.20	25.67
LexRank	36.63	10.94	33.18
LSA	30.18	8.02	27.90
BART(fine-tuned)	24.39	9.52	23.17
T5 (fine-tuned)	20.38	2.28	12.62
Billsum	30.96	22.87	28.86
GraphWriter	43.63	18.63	36.31
GBAS(proposed)	45.05	19.35	37.1

Table 5. The evaluation results on the SciSummNet dataset.

Model	Rouge-1	Rouge-2	Rouge-L
TextRank	33.77	18.39	31.99
LexRank	34.22	18.18	32.39
LSA	34.25	18.13	32.30
BART(fine-tuned)	42.37	18.49	41.37
T5 (fine-tuned)	40.98	12.49	34.42
Billsum	43.27	35.79	39.65
GraphWriter	43.64	15.11	36.94
GBAS(proposed)	45.05	15.11	38.36

Table 6. The evaluation results on the SciTLDR dataset.

Model	Rouge-1	Rouge-2	Rouge-L
TextRank	16.11	4.34	13.67
LexRank	16.68	4.50	14.07
LSA	18.22	4.93	15.26
BART(fine-tuned)	25.24	4.41	25.25
T5 (fine-tuned)	14.85	8.48	7.51
Billsum	16.11	10.11	9.15
GraphWriter	42.19	15.46	34.62
GBAS(proposed)	42.93	14.90	34.96

According to the results shown in Table 4, the proposed model outperforms the baseline methods on the ArxivComp dataset. The GraphWriter model achieves the closest result to the proposed model among the abstractive methods. This model generates the summary from the title, whereas the proposed model generates the summary from the introduction section. This is how this model differs from the proposed model. Token embedding for the proposed model used SciBERT, which was trained using scientific articles. The introduction and abstract sections were combined to create the proposed model’s graph. The main difference between these approaches and others (BART,

T5-based, and Billsum) is the restriction on token sequence lengths. The introduction section is also of variable length for each article. Therefore, the summaries produced by models do not correspond with the author's summary for articles of different lengths. Graph-based techniques are better than other techniques because they guarantee document integrity.

In comparing the results of these models, the auto-regressive decoder in the BART model allows it to achieve the best summarization performance. As can be seen in Table 6, graph-based methods are more successful than baseline methods in summarizing long documents because they preserve the integrity of the document.

The performance of the model for both Billsum and T5-model approaches dramatically declines as the document length increases. However, with the SciSummnet dataset—whose average word length is shorter—better performance was obtained in both models. Based on these results, as the average word length, T5-based techniques are not sufficient for summarizing long scientific documents.

Examining the results of the extraction methods (TextRank, LexRank and LSA) reveal that these methods are not as successful as the abstraction methods. It is seen that the lowest results are obtained in the SciTLDR dataset. The main reason for this is that as the length of the document increases, the number of sentences containing more general information also increases. When sentence selections are performed with these algorithms, the overlapping rate of the summary is decreased.

As can be seen in Tables 4 and 6 the proposed method outperforms baseline methods on long documents. As can be seen in Table 5, it is a comparable method for documents with a shorter average word length. An advantage of the proposed method is that it handles the document hierarchically for long documents.

When the results of the extractive methods (TextRank, LexRank, and LSA) are examined, it is seen that these methods are not as successful as the abstractive methods. It is seen that the lowest results are obtained in the SciTLDR data set. The main reason for this is that as the length of the document increases, the number of sentences containing more general information also increases. When sentence selections are performed with these algorithms, the overlapping rate of the summary is decreased.

As can be seen in Tables 4 and 6 the proposed method outperforms baseline methods on long documents. As can be seen in Table 5, it is a comparable method for documents with a shorter average word length. An advantage of the proposed method is that it handles the document hierarchically for long documents.

5.1. Human Evaluation

The Rouge metric compares the generated summary based on the overlapping of the n-grams with the ground-truth summary. However, it is not sufficient to prove the quality of the generated summaries. To overcome this problem, the generated summaries were also evaluated with human judgment. The evaluation criteria are as follows: 1) Concise ness(Con.) is whether you avoid redundant information; 2) Informativeness (I) is whether it contains salient information; 3) Coherence(Coh.) is whether the content of the generated summary is appropriate for the ground-truth summary; 4) Readability(R) means that the generated summary is easy to understand and fluent; and 5) Grammatically(G), the question is whether the sentences are appropriate to the grammar rules.

Five expert volunteers rated the summaries from 1 (worst) to 5 (best) for each criterion. Fleiss's Kappa analysis was performed to determine whether the evaluator scores were compatible with each other.

From the results in Table 7, it is seen that the evaluations of the volunteers mostly agree with each other in that the result of each criterion is greater than 0.5. According to the results, the generated summary is generally informative, fluent, and overlaps with a ground-truth summary. However, it has been observed that grammatical problems remain. For instance, some words repeat more than once in the generated summary. In addition, it adversely affects conciseness and readability because < unk > and some entities extracted as "generic" by the SciIE system are not learning. It caused inconsistencies in the meaning of sentences because these words were removed from the generated summaries at the last stage. Because the proposed model is pretrained with topics related to computer

science, it will produce successful summaries on these topics. In Table 8, sample abstracts generated for scientific articles.

Table 7. Human evaluation results.

Dataset	Con	In.	Coh.	R.	G.
Mean/Var.	3.11(0.57)	4.06(0.61)	4.03(0.56)	3.92(0.51)	3.15(0.56)
Fleiss kappa	0.671	0.611	0.534	0.603	0.613

Table 8. The samples of gold and gen summaries.

Summary Type	Summary
Gold	"This work presents a self-supervised method to learn dense semantically rich visual concept embeddings for images inspired by methods for learning word embed dings in NLP. Our method improves on prior work by generating more expressive embeddings and by being applicable for high-resolution images. Viewing the generation of natural images as a stochastic process where a set of latent visual concepts give rise to observable pixel appearances, our method is ..."
Gen	"Masked language models are an efficient tool used across NLP. It is considered task to implement many self supervised learning in the complexity-based view sampling. In this paper, we explore effective stochastic process and apply it to derive regional contextual masking such as dense semantically rich visual concept embed dings. We demonstrate how to train better complexity based view sampling when training data can achieve state-of-the-art performance on multiple NLP."
Gold	"Sentence embedding methods offer a powerful approach for working with short textual constructs or sequences of words. By representing sentences as dense numerical vectors, many natural language processing (NLP) ap plications have improved their performance. However, relatively little is understood about the latent structure of sentence embeddings. Specifically, research has not addressed whether the length and structure of sentences impact the sentence embedding space and topology. This paper reports research on a set of comprehensive clus tering and network analyses targeting sentence and sub sentence embedding spaces..."
Gen	"In this paper, we demonstrate how sentence embedding models can be used to improve the performance of natural language processing NLP applications. We perform topology by means that sentence embedding models such as network analyses require more relevant to sentence and sub-sentence embedding spaces and topology can achieve the same sentence embedding models performance is compared to short textual constructs".

6. Conclusions

Summarizing scientific articles while preserving the integrity of the document remains a challenging problem in terms of extracting salient words or sentences. In this study, a novel model that generates a summary from the introduction to correspond to the SciBERT token length has been proposed. The proposed model is comparable to the baseline methods on the SciSummnet dataset, whose document length is shorter than the ArxivComp and the SciTLDR dataset. However, when the proposed model is evaluated on the ArxivComp and SciTLDR datasets, it outperforms the baseline methods. The results of the human evaluation show that the proposed model generally generated an informative, fluent, and overlapping ground-truth summary. The proposed model is trained with current topics related to computer science. The proposed model provides superior summaries of topics in this field. The limitation of the model is that it is insufficient to summarize documents containing mathematical expressions, figures, and tables. These sections were eliminated from the

study during the preprocessing phase. In light of these, it may be of great value to condense the articles for future study.

References

1. Alomari, A., Idris, Sabri A.Q.M., Alsmadi, I., Deep reinforcement and transfer learning for abstractive text summarization A review. *Computer Speech & Language*, **2022**; 71, 101276.
2. Zhong, M., Liu, P., Wang, D., Qiu, X., Huang, X., Searching for effective neural extractive summarization: What works and what's next, arXiv preprint arXiv:1903.1907.03491, **2019**.
3. Mojriani, M., Mirroshandel, S.A., A novel extractive multi-document text summarization system using quantum-inspired genetic algorithm: Mtsqiga, *Expert systems with applications*, **2021**; Volume 171, 114555.
4. Cai, X., Liu, S., Yang, L., Lu, Y., Zhao, J., Shen, D., Liu, T., Covidsum: A linguistically enriched scibert-based summarization model for covid19 scientific papers., *Journal of Biomedical Informatics*, **2022**; Volume 127, 103999.
5. Du, Y., Zhao, Y., Yan, J., Li, Q.T., Ugdas: Unsupervised graph-network based denoiser for abstractive summarization in biomedical domain, *Methods*, **2022**; Volume 203, pp. 160–166.
6. Beltagy, I., Lo, K., Cohan, A., Scibert: A pretrained language model for scientific text, Available: arXiv preprint arXiv:1903.10676, **2019**.
7. Luan, Y., He, L., Ostendorf, M., Hajishirzi, H. Multi-task identification of entities, relations, and co-reference for scientific knowledge graph construction, arXiv preprint arXiv:1808.09602, **2018**.
8. Koncel-Kedziorski, R., Bekal, D., Luan, Y., Lapata, M., Hajishirzi, H., Text generation from knowledge graphs with graph transformers, arXiv preprint arXiv:1904.02342, **2019**.
9. Fan, A., Grangier, D., Auli, M., Controllable abstractive summarization, arXiv preprint arXiv:1711.05217, **2017**.
10. Liang, Z., Du, J., Li, C., Abstractive social media text summarization using selective reinforced seq2seq attention model, *Neurocomputing*, **2020**; Volume 410, pp. 432–440.
11. El-Kassas, W.S., Salama, C.R., Rafea, A.A., Mohamed, H.K., Edgesumm: Graphbased framework for automatic text summarization, *Information Processing & Management*, **2020**; Volume 57, 102264.
12. Cagliero, L., La Quatra, M., Extracting highlights of scientific articles: A supervised summarization approach, *Expert Systems with Applications*, **2020**; Volume 160, 102264.
13. Yasunag, M., Kasai, R. J. Zhang, Fabbri, A.R., Li, I., Friedman, D., Radev, D.R. Scisummnet: A large annotated corpus and content-impact models for scientific paper summarization with citation networks, arXiv preprint arXiv:1909.01716, **2019**.
14. Ju, J., Liu, M., Gao, L., Pan, S., Scisummpip: An unsupervised scientific paper summarization pipeline, arXiv preprint arXiv:2010.09190, **2020**.
15. Zhou, Q., Wei, F., Zhou, M., At which level should we extract? an empirical analysis on extractive document summarization, arXiv preprint arXiv:2004.02664, **2020**.
16. Cachola, I., Lo, K., Cohan, A., Weld, D.S., Tldr: Extreme summarization of scientific documents, arXiv preprint arXiv:2004.15011, **2020**.
17. Suleiman, D., Awajan, A., Multilayer encoder and single-layer decoder for abstractive arabic text summarization, *Knowledge-Based Systems*, **2022**; Volume 237, 107791.
18. Moirangthem, D.S., Lee, M., Abstractive summarization of long texts by representing multiple compositionality with temporal hierarchical pointer generator network, *Neural Networks*, **2020**, Volume 124, pp.1-11.
19. Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., Improving language understanding by generative pre-training, 2018,. Available online: <https://www.mikecaptain.com/resources/pdf/GPT-1.pdf> (accessed on 16.05.2024).
20. Devlin, J., Chang, M.W., Lee, K., Toutanova, K., Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint arXiv:1810.04805*, **2018**.
21. Liu, Y., Lapata, M., Text summarization with pretrained encoders, *arXiv preprint arXiv:1908.08345*, **2019**.
22. Dou, Z.Y., Liu, P., Hayashi, H., Jiang, Z., Neubig, G., Gsum: A general framework for guided neural abstractive summarization, *arXiv preprint arXiv:2010.08014*, **2020**.
23. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ..., Liu, P.J., Exploring the limits of transfer learning with a unified text-to-text transformer, *arXiv preprint arXiv:1910.10683*, **2020**.
24. Liu, Y., Dou, Z.Y., Liu, P., Refsum: Refactoring neural summarization, *arXiv preprint arXiv:2104.07210*, **2021**.
25. Yun, S., Jeong, M., Kim, R., Kang, J., Kim, H.J., Graph transformer networks. advances in neural information processing systems, *arXiv preprint arXiv:1911.06455*, **2019**.
26. Zheng, B., Wen, H., Liang, Y., Duan, N., Che, W., Jiang, ... D., Liu, T., Document modeling with graph attention networks for multi-grained machine reading comprehension, *arXiv preprint arXiv:2005.05806*, **2020**.
27. Zhou, Y., Shen, J., Zhang, X., Yang, W., Han, T., Chen, T., Automatic source code summarization with graph attention networks, *Journal of Systems and Software*, **2022**; Volume 188, 111257.

28. Chen, H., Hong, P., Han, W., Majumder, N., Poria, S., Dialogue relation extraction with document-level heterogeneous graph attention networks, *arXiv preprint arXiv:2009.05092*, **2020**.
29. Zhao, Y., Chen, L., Chen, Z., Cao, R., Zhu, S., Yu, K., Line graph enhanced AMR-to-text generation with mix-order graph attention networks, *In Proceedings of the 58th Annual meeting of the association for computational linguistics*, **2020**, pp. 732–741.
30. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., Graph attention networks, *arXiv preprint arXiv:1710.10903*, **2017**.
31. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., ..., Zettlemoyer, L., Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, *arXiv preprint arXiv:1910.13461*, **2019**.
32. Xie, F., Chen, J., Chen, K., Extractive text-image summarization with relation-enhanced graph attention network, *Journal of Intelligent Information Systems*, **2022**; Volume 61, pp.1-17.
33. Kumar, Y., Kaur, K., Kaur, S., Study of automatic text summarization approaches in different languages *Artificial Intelligence Review*, **2021**; Volume 54, pp. 5897–5929.
34. Augenstein, I., Das, M., Riedel, S., Vikraman, L., McCallum, A., Semeval2017 task 10: Scienceie-extracting keyphrases and relations from scientific publications, *arXiv preprint arXiv:1704.02853*, **2017**.
35. Beck, D., Haffari, G., Cohn, T., Graph-to-sequence learning using gated graph neural networks. *arXiv preprint arXiv:1806.09835*, **2018**.
36. See, A., Liu, P.J., and Manning, C.D. Get to the point: Summarization with pointer-generator networks, *arXiv preprint arXiv:1704.04368*, **2017**.
37. Lin, C.Y., Rouge: A package for automatic evaluation of summaries, *In Text Summarization Branches Out*, **2014**; pp.74-81.
38. Mihalcea, T. R., Textrank: Bringing order into text., *Conference on empirical methods in natural language processing*, **2004**; pp. 404-411.
39. Erkan, G., Radev, D.R., Lexrank: Graph-based lexical centrality as salience in text summarization, *Journal of Artificial Intelligence Research*, **2004**, Volume 22, 457–479.
40. Gong, Y., Liu, X., Generic text summarization using relevance measure and latent semantic analysis., *In Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, **2001**; pp. 679-688.
41. A. Kornilova and V. Eidelman, Billsum: A corpus for automatic summarization of us legislation, *arXiv preprint arXiv:1910.00523*, **2019**.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.