ARTICLE

Enhanced Topic-Aware Summarization Using Statistical Graph Neural Networks

Ayesha Khaliq[1], Salman Afsar Awan[1], Fahad Ahmad[2,*], Muhammad Azam Zia[1] and Muhammad Zafar Iqbal[3]

[1]Department of Computer Science, University of Agriculture Faisalabad, Faisalabad, Punjab, 37300, Pakistan

[2]School of Computing, Faculty of Technology, University of Portsmouth, Southsea, Portsmouth, PO1 2UP, UK

[3]Department of Mathematics and Statistics, University of Agriculture Faisalabad, Faisalabad, Punjab, 37300, Pakistan

*Corresponding Author: Fahad Ahmad. Email: fahad.ahmad@port.ac.uk

ABSTRACT

The rapid expansion of online content and big data has precipitated an urgent need for efficient summarization techniques to swiftly comprehend vast textual documents without compromising their original integrity. Current approaches in Extractive Text Summarization (ETS) leverage the modeling of inter-sentence relationships, a task of paramount importance in producing coherent summaries. This study introduces an innovative model that integrates Graph Attention Networks (GATs) with Transformer-based Bidirectional Encoder Representations from Transformers (BERT) and Latent Dirichlet Allocation (LDA), further enhanced by Term Frequency-Inverse Document Frequency (TF-IDF) values, to improve sentence selection by capturing comprehensive topical information. Our approach constructs a graph with nodes representing sentences, words, and topics, thereby elevating the interconnectivity and enabling a more refined understanding of text structures. This model is stretched to Multi-Document Summarization (MDS) from Single-Document Summarization, offering significant improvements over existing models such as THGS-GMM and Topic-GraphSum, as demonstrated by empirical evaluations on benchmark news datasets like Cable News Network (CNN)/Daily Mail (DM) and Multi-News. The results consistently demonstrate superior performance, showcasing the model's robustness in handling complex summarization tasks across single and multi-document contexts. This research not only advances the integration of BERT and LDA within a GATs but also emphasizes our model's capacity to effectively manage global information and adapt to diverse summarization challenges.

KEYWORDS

Summarization; graph attention network; bidirectional encoder representations from transformers; Latent Dirichlet Allocation; term frequency-inverse document frequency

1 Introduction

Owing to the substantial and swift expansion of information and textual documents, propelled by the pervasive popularity of online content and the prevalence of big data, it necessitates a considerable amount of time to fully grasp the entire document. Hence, a considerable number of

academics are currently engaged in order to develop various technical solutions for the automated summarization of textual contents. The main objective is to provide a swift document review without compromising the documents originality [1]. Depending on the outcomes produced by a summary, their nature can be extractive or abstractive. The extractive summarization technique entails the systematic selection of sub-words, phrases, content, and sentences from the source document, while in the abstractive technique, the sentences, and key ideas generated undergo substantial alteration and involve paraphrasing in comparison to the sentences found in the original input document [2]. Conventional approaches to Extractive Text Summarization (ETS) encompass methods like TextRank and LexRank, both are centered on the computation of sentence similarity scores [3]. The acquisition and modeling of cross-sentence associations are pivotal stages in the process of deriving a summary containing profoundly significant sentences from the input content [4]. In a study referenced as [5], a framework for Single-Document Summarization (SDS) was constructed, featuring the inclusion of a hierarchical document encoder. The scholars in [6] proposed an innovative neural model designed for extractive summarization structured as a sentence ranking task. They augmented the ROUGE values through the integration of reinforcement learning techniques into their methodology. Further, the methodologies presented by [7,8] primarily employed the encoder-decoder, sequence-to-sequence architecture, wherein each sentence was encoded by using distinct neural components in various manners. However, the empirical observations have revealed that employing this paradigm to model inter-sentence relationships does not yield substantial performance enhancements in the context of summarization [9]. A few years ago, several studies exemplified by references [10] employed Recurrent Neural Networks (RNNs) to acquire and frame the inter-sentence relationship. Nonetheless, recurrent models face difficulties in capturing long-range dependencies among sentences and effective handling of computational factors, primarily because they only process input and output sequences in a strictly sequential manner [11].

In recent years, there has been a notable achievement in applying Graph Attention Networks (GATs) for the purpose to summarized documents [12]. This success stems from their proficiency in capturing complex inter-sentence relationships within documents [3,13–17]. To be precise, GNN models demonstrate the capability to effectively model intricate structural data that comprises semantic units (nodes) and the relationships (edges) connecting them. Authors of [18] employed the Rhetorical Structure Theory (RST) graph, which relies on joint extraction and syntactic compression, to model cross-sentence associations in the context of SDS. Reference [19] proposes an unsupervised graph-based model that leverages discourse information for hierarchical ranking, which is specifically tailored for extensive scientific documents, utilizing both intra-connections and inter-connections among sentences. The authors in [20] incorporate three types of nodes, namely Sentence, Entity Division Unit (EDU), and Entity Nodes. They utilize RST discourse parsing to capture the relationships between the EDUs. The researchers of [21] devised a Heterogeneous Transformer based model tailored for extractive summarization of lengthy text, incorporating multi-granularity sparse attention and treating tokens, entities, and sentences as distinct types of nodes in the process.

Another vital aspect of summarization pertains to modeling global information (topical information), which plays a pivotal role in the selection of sentences [22]. They concurrently trained with latent topics to model global information, employing the Neural Topic Model (NTM) for this purpose. The study conducted by the researchers cited in [23] offers an expanded model that employs NTM for abstractive text summarization. Within the graph structure, topics are extracted using clustering methods like K-Means and Gaussian Mixture Models (GMM). Even though the above methods have worked well, creating a powerful graph model extendable from single to Multi-Document Summarization (MDS) with sophisticated topical features extractor and robust encoder

that strengthens the process of better extractive summary generation is still an area of research that's not completely solved. Bidirectional Encoder Representations from Transformers (BERT)-Encoder has been pre-trained on extensive datasets, enabling it to acquire intricate features owing to its robust architecture. Building upon this, latent topic nodes are discovered using the widely recognized generative statistical topic modeling technique, Latent Dirichlet Allocation (LDA). In the subsequent stage, this research establishes a heterogeneous document graph, comprising sentence, word, and topic nodes, interconnected by edge features represented as Term Frequency-Inverse Document Frequency (TF-IDF) values. These node representations are updated through the utilization of GATs. In essence, this research offers several advantages: (i) Throughout the graph propagation process, sentence representations are enhanced by the infusion of topical information. This topical information can be seen as a feature at the document level, assisting the proposed model in extracting crucial content from the entire document. (ii) Topic nodes function as intermediaries, enhancing the proposed model's proficiency in capturing relationships between sentences. (iii) This research model could extend and adopted for both single and multi-document summarization. The key contributions of this research model are outlined as follows:

- In the context of prior related work, this study marks the first endeavour to introduce an innovative Enhanced Topic-Aware GATs for both single and multi-document extractive summarization. This approach combines LDA and pre-trained BERT, along with TF-IDF edge features with GATs.
- This research performs a quantitative examination to assess the impact of latent topics, and provide a comprehensive insight into how topical information facilitates the summarization process for both single and multi-document summarization.
- Based on the simulation results obtained from standard news dataset Cable News Network (CNN)/Daily Mail (DM) and Multi-News for single and multi-documents respectively, this research findings consistently attain satisfactory results across the board, compared with other topic aware and non-topic Heterogeneous Graphs (HG) models.

The structure of this study is systematically organized to facilitate a comprehensive understanding of the research. Section 2 outlines the Related Work, setting the context and foundation. Section 3 introduces the Methodologies for SDS and MDS. Section 4 provides an in-depth Performance Analysis, evaluating the proposed model against established benchmarks. Section 5 presents Discussion about the significance of the proposed techniques and their comparative advantages. Section 6 presents Conclusion of the study by summarizing the key findings and contributions, addresses the limitations and outlines potential directions for future research.

2 Related Work

This research literature review examines key methods in extractive summarization, starting with deep neural sequence-to-sequence models, and then exploring deep-learning transformers and BERT usage. This research also assesses techniques leveraging topical information to enhance inter-sentence relationships, and the application of neural graph-based networks in these tasks.

2.1 Sequence to Sequence Models

Various research in the past frequently employs sequence-to-sequences architecture-based models for various extractive summarization tasks. In the context of summarizing single legal documents, the authors implemented a deep learning method known as the Feedforward Neural Network (FFNN). This approach generates summaries by extracting key sentences, making it useful even without access

to specific features or expert knowledge [23]. In a scholarly investigation by [24] an encoder-decoder architecture was introduced as a comprehensive framework tailored for SDS. This system involves a many-to-many sequence problem and incorporates attention-based Long Short-Term Memory (LSTM) for generating the summaries [25]. The research introduced an innovative approach to automatic text summarization [26]. This strategy amalgamates adaptable fuzzy attributes with a neural sequence-to-sequence model integrating an attention mechanism.

2.2 Transformer and BERT-Based Models

The Transformer, a deep learning model, leverages a self-attention mechanism which is commonly used by many researchers for extractive summarization tasks. In the work of [27], the authors developed a model that enables stepwise summarization. This model integrates the previously generated summary as an additional sub-structure within the structured transformer. They focused on models like Hierarchical BERT (HiBERT) introduced by [28] and extended transformer [29] which offer an extractive, encoder-centric, stepwise approach to summarization. In the study conducted by [30], they introduce an extractive summarization model that utilizes layered trees. The researchers in [31] put forward a document-level discourse-based attention approach for extractive summarization by employing a smaller attention module size which is the necessary component of transformer architecture. The research in [32] introduces two separate transformer-based methods that perform sentiment analysis while extracting the most crucial words to produce a summary. The study in [20] presents a Heterogeneous Transformer (HETFORMER), an advanced pre-trained model based on heterogeneous transformers. This model is specifically tailored for the extractive summarization of long textual content and incorporates multi-granularity sparse attention. The researchers in [33] introduce a Star architecture-based model for extractive summarization. This model utilizes a Transformer with a self-attention mechanism and a star-shaped structure. Study [34] proposes a novel paradigm for extractive summarization known as DiffuSum. This approach creates summary sentence representations using transformer-based diffusion models and performs extraction based on sentence representation matching.

Furthermore, many researchers utilize the pre-trained model BERT for extractive summarization. In the study by [35], the authors introduce a "lecture summarization service". This service, implemented in Python as a RESTful API, leverages the BERT model for text embeddings and K-Means clustering. The work presented by [36] introduces a hybrid method that combines the advantages of both extractive and abstract approaches to create summaries for lengthy scientific texts using the BERT model. The fundamental research on utilizing BERT for extractive summarization, as conducted by [37], introduces an exceptionally efficient fine-tuning method that leverages the BERT model for extractive summarization. The authors of [19,28] also rely on the BERT model to provide contextualized representations for their summarization tasks. An unsupervised technique for multi-document extractive summarization based on transfer learning from the BERT sentence embedding model is proposed by the research described in [38]. In study [39], they delve into extractive summarization methods that leverage sentence embeddings derived from fine-tuned BERT models, coupled with the utilization of the K-Means clustering technique.

2.3 Topic Aware Text Summarization

Another vital aspect of summarization pertains to modeling global information (topical information), which plays a pivotal role in the selection of sentences [21]. In reference [40], the scholars utilized BERT to obtain contextual sentence representations. The study conducted by [41] explored an unsupervised approach for extractive summarization, integrating clustering with topic modeling to

mitigate topic bias. LDA was employed for topic modeling, and K-Medoid clustering was utilized for generating summaries. In work of [42], authors proposed an innovative approach by integrating BERT for contextual embeddings and an NTM for discovering topics. Moreover, a transformer network is employed to effectively capture long-term dependencies, facilitating the seamless exploration of topic inference and text summarization in an end-to-end fashion. Utilizing topic vectors and sequence networks to estimate latent information inside the document, reference [43] aimed to improve the precision and quality of the summary text. The researchers in [44] seamlessly combine the graph contrastive topic model with a pre-trained language model. The objective is to leverage both global and local contextual semantics effectively for the extractive summarization of lengthy documents. The study [45] proposes an innovative Topic-Aware GATs abstractive text summarization model. Initially, the document undergoes encoding with BERT to generate sentence representations, concurrently using NTM to discern the potential topic of the document.

2.4 Neural Graph-Based Summarization

Contemporary researchers often utilize a heterogeneous network, incorporating multiple nodes that undergo updates, instead of a homogeneous graph with uniform, unmodified nodes. Research [46] delves into the exploration of syntactic Graph Convolutional Networks (GCNs) to effectively model non-Euclidean document structures. Additionally, they incorporate an attention information gate to selectively highlight salient information for the purpose of generating text summarization. The findings by [47] introduce a HG network designed to effectively capture various levels of information, encompassing both words and sentences. The model emphasizes redundancy dependencies between sentences, iteratively refining sentence representations through a redundancy-aware graph. In the study presented by [48], a novel Multiplex Graph Convolutional Network (MGCN) is introduced to collectively model various types of relationships between sentences and words. Reference [49] primarily employs the Text Graph Multi-Headed Attention Network (TGA) to proficiently acquire sentence representations across various types of text graphs at different levels. In [50], they employ advanced techniques, such as hypergraph transformer layers, to iteratively update and learn robust sentence representations. This approach strategically fuses various types of sentence dependencies, including latent topics, keywords, coreference, and section structure. The detailed methodology of the novel proposed model is discussed in the following section.

The study situates itself uniquely among a spectrum of contemporary research in extractive summarization for text but video summarization is also a promising field that identifies key segments in untrimmed videos leverage dual streams and a context-semantics interaction layer to enhance content detection [51,52]. In the realm of sequence-to-sequence and transformer-based models, current methodologies for text summarization frequently utilize complex neural architectures like BERT and attention mechanisms to enhance the extractive process. However, proposed model diverges by not only employing these advanced techniques but also innovatively integrating GATs and LDA. This combination permits for a refined handling of inter-sentence relationships and topic coherence, which is less stressed in traditional models.

Moreover, while typical methods often focus on single-document contexts or connect simple neural networks for sentence extraction, this research model surpasses in both single and multi-document setups by effectively synthesizing information across various documents. This is apparent in the proposed model's ability to dynamically acclimate to the densities of the text, apprehending both the granular details and broader thematic elements more effectively than many prevailing models. Thus, this research approach not only aligns with the cutting-edge in technology but also pushes the boundaries by providing a more holistic and context-aware summarization solution.

3 Methodology

In this section, this research presents the methodology of proposed innovative model, developed to address both single and MDS tasks. This research builds a representation of an input document using an HG that includes three types of nodes: Words, sentences, and topics, and utilizes GATs to adeptly capture the information relationships among these nodes. The proposed network consists of five phases. The first phase focuses on representing the document as a HG. Following this initial phase, this research has four core trainable modules, which include a BERT encoder, LDA Topic Extractor, Graph Attention Layer, and Sentence Classification module. These stepwise processes are further explained in the upcoming subsections.

3.1 Graph Construction

Let's consider a graph representation, denoted as G = (V, E). In this representation, V corresponds to the set of nodes, and E signifies the edges connecting these nodes. In proposed approach, this research formally define the HG as follows: V = (V$_w$ U V$_s$ U V$_t$) and E = (Ew2s ∪ Es2w ∪ Ew2t ∪ Et2w). Here, V$_w$, V$_s$, and V$_t$ represent the three semantic components of a document, namely, words, sentences, and topics. Additionally, Ew2s, Es2w, Ew2t and Et2w correspond to four distinct types of edges: Word-to-sentence, sentence-to-word, and word-to-topic, respectively. Fig. 1 provides an overview of the proposed model, which primarily consists of four key components: the BERT Encoder, LDA Topic Extractor, HG Layer, and the Sentences Classification module.
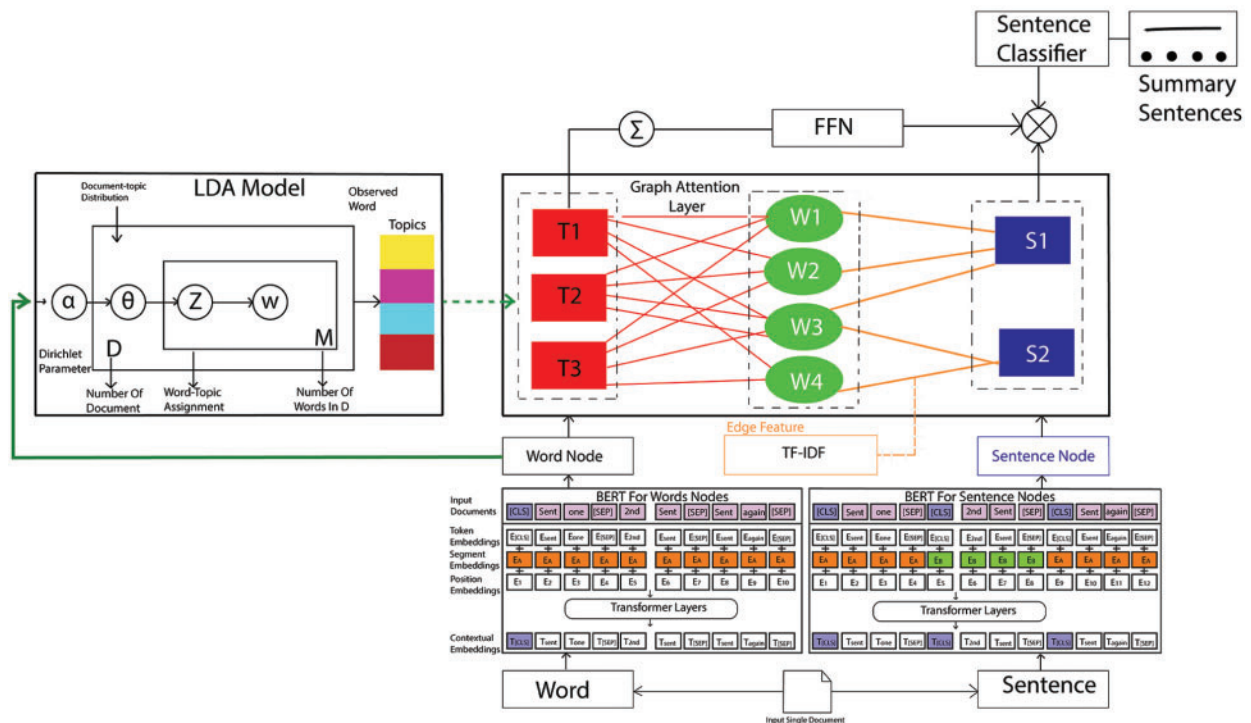

Figure 1: Proposed framework for single document summarization

- BERT encoder generates sentence and word nodes, encoding them for integration into the document graph.
- Subsequently, the LDA model is employed to extract latent topic features from each document. These features are then incorporated into the graph layer to extract semantic information.
- In the third phase, the HG updates these node depictions by iteratively transmitting messages among word, sentence, and topic nodes via GATs.
- In conclusion, the output sentence-topic representation is utilized for sentence classification, with the training objective focused on minimizing cross-entropy loss.

3.2 Bidirectional Encoder Representations from Transformers Encoder

Let's consider a graph representation, denoted as $G = (V, E)$. In this representation, V corresponds to the set of nodes, and E signifies the edges connecting these. The encoder employed in BERT utilizes an attention-based architecture, which offers distinct advantages over other standard language models. Notably, it employs deep bidirectional context training of the sequence. This means that during training, it considers both the left and right context, which enhances its ability to understand the sequence effectively. The output vectors generated by BERT are aligned with individual tokens rather than entire sentences, in accordance with the original design of the BERT model [53]. Simultaneously, when dealing with extractive summarization, the focus shifts to manipulating sentence-level representations. In the original BERT model, segmentation embeddings are exclusively used for sentence-pair inputs. Nevertheless, in the context of the extractive summarization task, this research needs to encode and handle inputs consisting of multiple sentences [37]. This research utilizes the modified version of BERT in which an external Classify token (CLS) token at the commencement of each sentence and a Separate token (SEP) token at the conclusion of each sentence is used. This systematic approach is employed to establish a consistent representation for individual sentences, in accordance with the method employed in [37]. The complete operational process of the BERT model for sentence encoding is visualized in Fig. 2 [54]. For the generation of word nodes, this research implements the foundational architecture of the BERT model, as illustrated in Fig. 3.
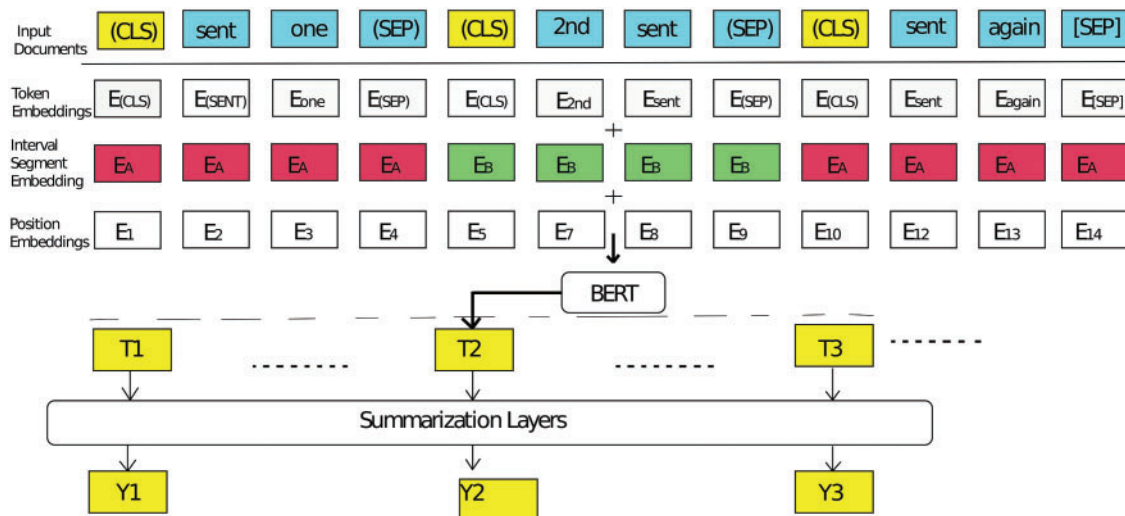

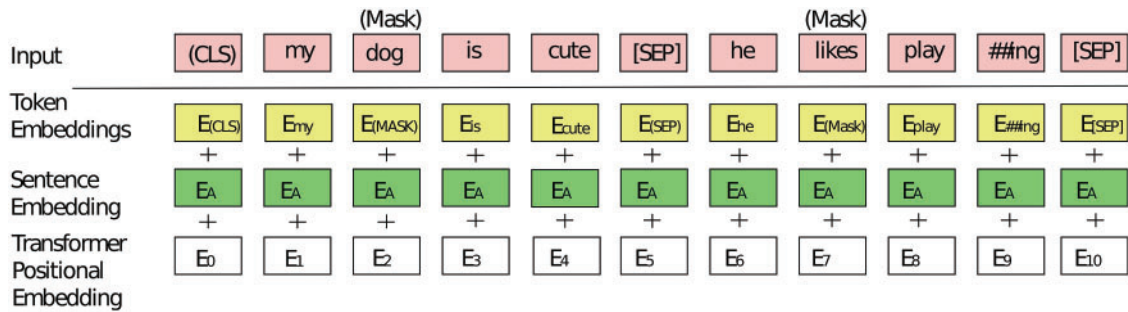**Figure 2:** Bidirectional encoder representations from transformers for sentence encoding

Figure 3: Bidirectional encoder representations from transformers for word encoding

3.3 Latent Dirichlet Allocation Topic Extractor

The core idea behind the proposed connection is to enhance global information by integrating latent topic nodes within the graph structure. Consequently, instead of resorting to neural models or K-Means and GMM [3,13] which are purely clustering models, this research opts for the utilization of the LDA model to extract the initial topic features from each document, leveraging pre-trained word embeddings. LDA is a commonly used unsupervised generative probabilistic method for corpus modeling. It is based on the premise that each document can be described by a probabilistic distribution of latent topics, with the topic distribution across all documents adhering to a shared Dirichlet prior. Each latent topic can be represented as a probability distribution over words, meaning some words are more likely to appear in specific topics. These individual word distributions share a common Dirichlet prior that influences the overall word combinations within each topic [55].

LDA is traditionally designed to operate with bag-of-words representations, not word embeddings, and this research needs to transform the data into a format that LDA can understand. To overcome this problem, the proposed work creates a Document-Term Matrix (DTM), which represents documents using word embeddings. In this matrix, rows correspond to documents, and columns correspond to the dimensions of the BERT embeddings. The complete generative process by LDA for latent topics discovery is as in below Algorithm 1:

Algorithm 1

For each topic t from 1 to T:

    Sample a word distribution βt from a Dirichlet distribution: $\beta t \sim$ Dirichlet (η).

    This step determines how words are distributed within each topic.

For each document d from 1 to D:

    Sample a topic distribution θd from a Dirichlet distribution: $\theta d \sim$ Dirichlet (α).

    This step defines the distribution of topics within each document.

For each word n in document d:

    Sample a topic assignment zd,n from the topic distribution θd: $zd,n \sim$ Multinomial (θd).

    Sample a word wd,n from the word distribution corresponding to the assigned topic zd,n : $wd,n \sim$ Multinomial $(\beta zd,n)$.

    This step represents how words are generated within documents based on their topic assignments.

Here D: Number of documents, W: Vocabulary size (number of unique words), T: Number of topics, Nd: Number of words in document d, θd: Topic distribution for document d, βt: Word distribution for topic t, zd, n: Topic assignment for the nth word in document d and wd, n: The specific word in document d at position n. In the above generative model, this research conducted parameter

estimation for α (document-topic distribution), η (topic-word distribution), and the assignment of topics to words within each document. The objective was to discover the most probable topics and their assignments to words based on the observed documents. The probability of the observed data D is computed for a corpus as follow in Eq. (1):

$$p\left(D|\alpha,\ \beta\right) = \prod_{d=1}^{M} \int p\left(\theta_d|\alpha\right) \left(\prod_{n=1}^{N_d} \sum_{z_{dn}} P\left(z_{dn}|\theta_d\right) p\left(w_{dn}|z_{dn},\ \beta\right)\right) d\theta_d \tag{1}$$

This understanding of topic distribution is then infused to Graph Structure to enrich the global information along with sentence and word nodes generated by BERT. To further emphasize the significance of the relationships between sentence, word, and topic nodes, this research incorporates TF-IDF values into the edge weights within this encoding phase.

3.4 Heterogenous Graph Layer

The vectors of nodes are initialized with embedding features, where H°s = Xs (sentence nodes) H°w = Xw (word nodes), and H°t = Xt (topic nodes), respectively. Sequentially, the node representations are updated with the Graph Attention Network (GAN).

3.4.1 Graph Attention Layer

To capture the complex interactions and individual features within the HG, this research employs the GATs to learn the hidden representations of each node. Specifically, supporting $h_i \in R^{dhi}$ and N_i express the input hidden representation and the surroundings of node *ith*, respectively, the computation of the GATs layer can be expressed as follows in Eqs. (2) and (3):

$$z_{ij} = LeakyRelu\left(\vec{a}^T \left(W_q \vec{h}_i || W_k \vec{h}^j\right)\right) \tag{2}$$

$$\alpha_{ij} = \frac{e^{z_{ij}}}{\sum_{k \in N_i} e^{z_{ik}}} = \vec{h}^{\prime i}\left(\sum_{j \in N_i} \alpha_{ij} W_u \vec{h}_j\right) \tag{3}$$

Here, W_q, W_k, W_v, and the vector αd represent trainable parameters that are optimized throughout the training process. The symbol ‖ is used to represent the concatenation operator. σ represents the non-linear transformation function, while h^i signifies the hidden state, which encapsulates information gathered from the neighboring nodes. Alternatively, to enhance performance, multi-head attention can be utilized, and it is computed in the following Eq. (4):

$$\vec{h}^{\prime i} ||_{K_{k=1}} \left(\sum_{j \in N_i} \alpha_{ij} W^k_u \vec{h}_j\right) \tag{4}$$

In addition, to tackle the challenge of gradient vanishing, a residual connection is incorporated. This connection is appended to the original representation, culminating in the derivation of the ultimate hidden state using Eq. (5):

$$h^{\overrightarrow{\prime\prime}\ \vec{\prime}\ i}_{ii} \tag{5}$$

3.4.2 Graph Updation

Following the initialization, the sentence nodes are updated by incorporating information from their neighboring word nodes as well as the topic nodes are also updated through the utilization of GATs and FFNN layers using Eqs. (6) and (7):

$$U_{T \leftarrow w}^{1} = GAT \left(H_{T}^{0}, \; H_{w}^{0}, \; H_{w}^{0} \right) \tag{6}$$

$$H_{T}^{1} = FFN \left(U_{T \leftarrow w}^{1} + H_{T}^{0} \right) \tag{7}$$

The text following an equation need not be a new paragraph. Please punctuate equations as regular text. During each iteration, the update process includes interactions between word-to-sentence, sentence-to-word, and word-to-topic connections. This can be formulated as shown in Eqs. (6)–(15):

$$U_{w \leftarrow s}^{t+1} = GAT \left(H_{w}^{t}, \; H_{s}^{t}, \; H_{s}^{t} \right) \tag{8}$$

$$U_{w \leftarrow T}^{t+1} = GAT \left(H_{w}^{t}, \; H_{T}^{t}, \; H_{T}^{t} \right) \tag{9}$$

$$U_{w \leftarrow s, T}^{t+1} = \sigma \left(U_{w \leftarrow s}^{t+1} + U_{w \leftarrow T}^{t+1} \right) \tag{10}$$

$$H_{w}^{t+1} = FFN \left(U_{w \leftarrow s, \; T}^{t+1} + H_{w}^{t} \right) \tag{11}$$

$$U_{s \leftarrow w}^{t+1} = GAT \left(H_{s}^{t}, \; H_{w}^{t+1}, \; H_{w}^{t+1} \right) \tag{12}$$

$$H_{s}^{t+1} = FFN \left(U_{s \leftarrow w}^{t+1} + H_{s}^{t} \right) \tag{13}$$

$$U_{T \leftarrow w}^{t+1}, \; A_{T \leftarrow w}^{t+1} = GAT \left(H_{T}^{t}, \; H_{w}^{T+1}, \; H_{w}^{T+1} \right) \tag{14}$$

$$H_{T}^{t+1} = FFN \left(U_{T \leftarrow w}^{t+1} + H_{T}^{t} \right) \tag{15}$$

Here, $A_{T \leftarrow w}$ signifies the attention matrix originating from word nodes towards the topic nodes. Next, this research combines the topic-specific features obtained from GATs to create a holistic representation of the document's thematic content. This combining process is explained as follows in Eq. (16):

$$\alpha_{i} = \frac{\sum_{n=1}^{N_d} c \left(w_n \right) * A_i, \; n}{\sum_{j=1}^{k} \sum_{n=1}^{N_d} c \left(w_n \right) * A_j, \; n} = \sum_{i=1}^{K} \alpha_i \; * \; H_{T_i} \tag{16}$$

In the above equation, $A_{i,j}$ represents the information that word j contributes to topic i. $C(wn)$ signifies the frequency of the word W_n in the document. Also, k stands for the number of topics, and αdi denotes the dominance level of topic i in relation to the overall document-topic. Furthermore, the hidden state of each sentence is intricately combined with the topic vector, culminating in the derivation of the sentence-topic representation. This process is elaborated as follows in Eq. (17):

$$H_{s_i, \; T_d} = FFN \left(H_{T_d} \right) \oplus H_{s_i} \tag{17}$$

3.4.3 Sentence Classification Module

Leveraging the learned sentence-topic representations, the final phase performs sentence classification. This involves optimizing the training objective, primarily through minimizing the cross-entropy loss function using Eq. (18):

$$L = \sum_{i=1}^{n} y_i log \left(\hat{y}_i \right) + (1 - y_i) \; log \left(1 - \hat{y}_i \right) \tag{18}$$

3.4.4 Single to Multi-Document

Presently, there is a scarcity of comprehensive research on MDS. The predominant challenge inherent in MDS lies in the considerable disparity between input documents in terms of their primary focus and perspective [3]. In essence, enhancing global information holds the potential to enhance the

performance of the MDS task. This enhancement is achieved by incorporating latent topics extracted from word nodes to represent the entirety of sentences across multiple documents. It should be highlighted that the proposed SDS model can be elegantly adapted for MDS, showcasing the versatility of this innovative research. To achieve this, the presented research introduced minor refinements to the proposed framework, resulting in a sophisticated MDS model depicted in Fig. 4. Fig. 1 illustrates the single-document summarization process, while Fig. 4 depicts the multi-document summarization process. The modifications applied to the proposed SDS model encompass the following: Initially, this research introduces latent topics to encompass themes across the entire corpus of relevant documents through an LDA model. In this context, rather than consolidating all the features related to the topic into a single representation, this research opts to maintain each individual topic feature representation separately. This approach was chosen to uphold information preservation, aligning with the methodology employed in the research. Secondly, the creation of word nodes and sentence nodes is facilitated by a collection of pertinent documents. This collection comprises a list of sentences and a set of distinct words derived from multiple documents, deviating from the conventional approach of relying on a singular document in the context of the SDS problem. Specifically, let $D = [d1, d2, \ldots, dn]$ represent the set of input multi-documents. This research re-calculates the output sentence-topic representation s_i as follows using Eqs. (19) and (20):

$$\overline{H}_{T_D} = FFN \sum_{k=1}^{K} H_{T_k} \tag{19}$$

$$H_{si}, T_D = \sigma \left(FFN \left(\overline{H}_{T_D} \oplus H_{si} \right) \right) \tag{20}$$
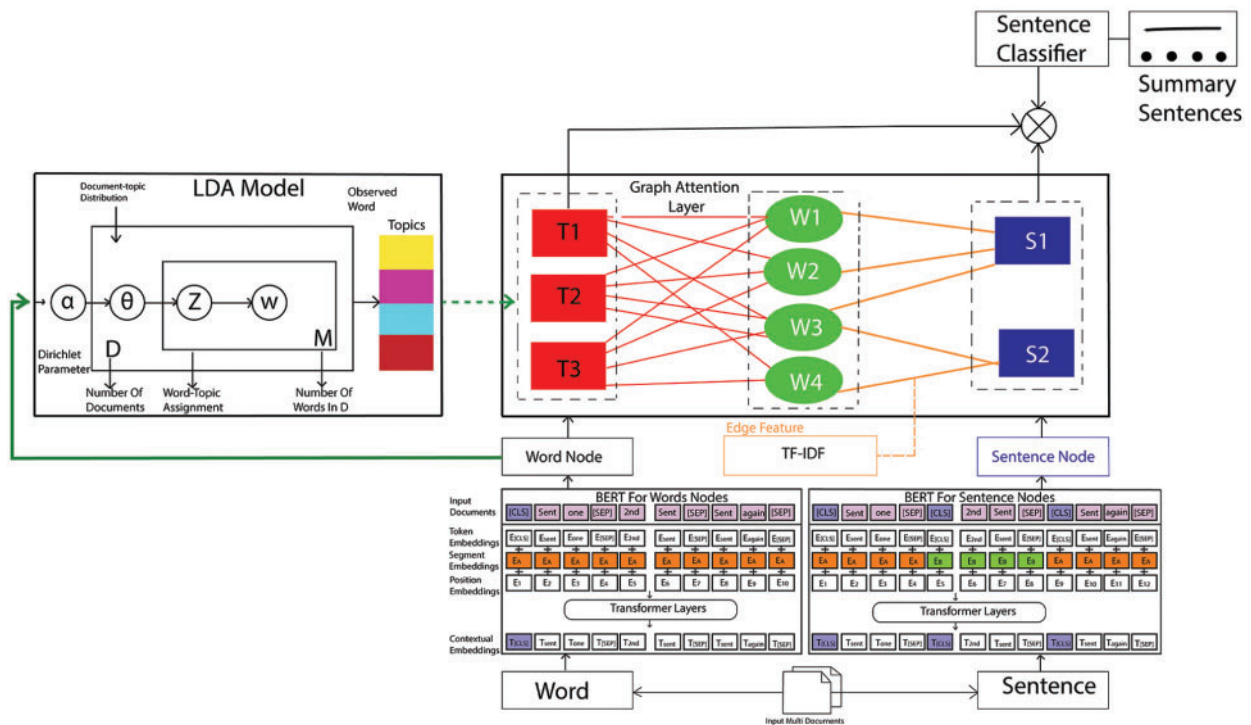

Figure 4: Proposed framework for multi-document summarization

In this equation, *k* represents the quantity of topics associated with the multiple documents, and ‖ signifies the concatenation operation. Subsequently, the resulting matrix undergoes a transformation into a vector via a flattened layer, a step integral to the final classification.

4 Performance Evaluation

This research evaluation strategy is focused on rigorously testing the proposed model's effectiveness in extractive summarization against contemporary state-of-the-art models. The performance assessment utilizes the CNN/DM and Multi-News datasets, which are standards in the field for SDS and MDS tasks, respectively. This deliberate choice confirms a broad and challenging test environment, reflecting real-world applications and the complexity of integrating multiple document contexts.

Through systematic comparisons, proposed model is tested against well-known benchmarks and recent advances in neural graph-based models. These assessments are not just limited to performance metrics like ROUGE scores but also extend to an examination of how well each model assimilates and processes linguistic and thematic information across different document types. This research examines the performance in relation to various node configurations and graph structures used in competing models, providing a comprehensive view of the proposed model's abilities.

Furthermore, the assessment comprises a detailed review of the implementation settings, such as vocabulary size, tokenization standards, and hyperparameters like learning rates and dropout settings. These particulars underscore the tailored approach of proposed model in handling the refinements of extractive summarization, highlighting its innovative integration of advanced neural network techniques and its adaptability to both SDS and MDS summarization challenges. This laborious evaluation framework ensures conclusions are well-founded and validate clear advancements over existing methodologies.

4.1 Dataset Description and Selection Rationale

For the evaluation of proposed summarization model, this research employed two widely recognized benchmark datasets: CNN/DM for SDS and Multi-News for MDS. The CNN/DM dataset is chosen due to its substantial volume and diversity, comprising 287,227 training samples, 13,368 validation, and 11,490 testing samples, which signify a broad spectrum of news stories [17,20]. This dataset is instrumental in training models to handle a variety of topics and writing styles, making it a standard in SDS research.

The Multi-News dataset, consisting of 44,972 training, 5622 validation, and 5622 testing samples, was selected for its unique composition of news articles paired with human-written summaries [56]. Each instance comprises 2–10 source documents, providing a complex challenge of synthesizing information across multiple texts, thereby testing the model's capability in MDS scenarios [3,17,20]. The relevant statistics for the two benchmark datasets are outlined in Table 1.

These datasets were specifically chosen not only for their relevance and challenge to the task of extractive summarization but also for their established use in prior research, allowing for direct comparison with state-of-the-art models in the field. This rigorous evaluation framework ensures that proposed model's performance is benchmarked against the highest standards in ETS, highlighting its innovation and effectiveness in handling both single and multiple document summarization tasks.

Table 1: Identified datasets statistics

	CNN/DM dataset	Multi-news dataset
Training	28,227	44,972
Validation	13,368	5622
Testing	11,490	5622
Vocabulary size	717,951	666,515

4.2 System Configuration and Implementation Settings

Proposed innovative system configuration utilized a Lenovo Mobile Workstation, featuring a 12th Generation Intel Core i9 processor, 128 GB DDR4 memory, a 4 TB SSD, and an NVIDIA RTX A4090 graphics card. Operating on Windows 11 with Python 3.12.0, this setup facilitated the robust testing of proposed model across various datasets.

The pre-processing phase in this study involves meticulous text cleaning, tokenization, and normalization to prepare the datasets for processing. This comprises the removal of stop words, lemmatization, and the reduction of vocabulary by cleaning out low TF-IDF terms. Feature selection is innovatively handled through the integration of advanced neural network techniques that dynamically acclimatise to the details of the text. Proposed model distinctively leverages a combination of GATs, BERT embeddings, and LDA to extract and exploit both semantic and structural features effectively. These steps ensure that proposed model not only comprehends the fundamental elements of the text but also grasps deeper thematic and contextual nuances.

In terms of execution, neural network architecture was precisely designed for both SDS and MDS tasks. The network utilizes a 50,000-token vocabulary and integrates an uncased BERT-based model to create 300-dimensional embeddings. These embeddings are influential in generating word and sentence nodes, with a transformation of the 769-dimensional output from BERT to 300 dimensions tailored specifically for the Graph layer. The pre-processing routine enhances model efficiency by removing stop words, applying lemmatization, and filtering out low TF-IDF terms, which reduces the vocabulary by 10%. The graph layer is powered by a GATs with a hidden size of 64, a feedforward layer with a 512-inner size, and an 8-head multi-head attention, all integrated with a 0.1 dropout rate to prevent overfitting. This research optimized model training using the Adam optimizer, with early stopping triggered after one epoch if no loss improvement is noted, setting the learning rate at 5e-4 and the batch size at 64. To tailor the summaries, lengths are capped at 3 tokens for single-document and 11 for multi-document tasks, ensuring concise and relevant output. This comprehensive setup underlines our commitment to leveraging advanced technologies to push the boundaries of text summarization.

4.3 Results and Discussion

This section presents the detailed simulation results of the proposed model, focusing on its performance relative to other contemporary models. The evaluation begins with the proposed model's testing on the widely used CNN/DM dataset for SDS and extends to the Multi-News dataset for MDS.

The study organizes the evaluation into four distinct parts, the first part lists the performance of baseline models like Lead-3 and Oracle. The second part reviews models that utilize unique embedding techniques and graph networks. The third focuses on topic-aware models tailored for extractive summarization. The final part discusses the performance of proposed model, which incorporates an

enriched topical information-based BERT graph structure, enabling it to effectively discern cross-sentence relationships.

4.3.1 Performance Insights on CNN/DM Dataset

Table 2 categorizes models into single document models and topic-aware models, comparing them against baseline models like Lead-3 and Oracle. Oracle, expectedly, shows the highest scores across all metrics due to its idealized methodology of sentence selection based on maximum overlap with reference summaries. Among single document models, DIFFU-SUM shows a significant strength in R-L scores, which indicates better retention of the sentence structure from the original text. The topic-aware models, particularly the proposed model, perform well in all three metrics, with a notable R-2 score, suggesting efficient handling of phrase-level information that contributes to the coherence and fluency of the summaries.

Table 2: ROUGE scores on CNN/DM dataset

Models	Models	R-1	R-2	R-L
Lead-3	Lead-3	40.42	17.62	36.67
Oracle	Oracle	55.61	32.84	51.88
Single document models				
HSG	HSG	42.31	19.51	38.74
HSG + Tri-blocking	HSG + Tri-blocking	42.95	19.76	39.23
JECS	JECS	41.70	18.50	37.90
HETFORMER	HETFORMER	44.55	20.82	40.37
MULTI-GCN	MULTI-GCN	43.16	20.14	39.49
DISCO-CORELATION-GRAPH	DISCO-CORELATION-GRAPH	43.61	20.81	41.12
DIFFU-SUM	DIFFU-SUM	44.83	22.56	40.56
Topic aware models				
Topic-GraphSum		44.02	20.81	40.55
GRETEL		**44.62**	20.96	**40.69**
THGS-GMM		43.28	20.31	39.67
DEEP-SUM		43.30	19.0	38.90
Proposed model		44.46	**20.98**	40.51

4.3.2 Performance Insights on Multi-News Dataset

Table 3 evaluates the models' performance on the Multi-News dataset, which involves summarizing content from multiple documents. This is a more multifaceted task as it needs not only summarizing information from individual articles but also mixing information across numerous sources. SG-SUM and the proposed model stand out, with SG-SUM achieving the highest R-L score, indicating its superior capability in maintaining structural coherence across multiple documents. The proposed model displays a balanced performance with competitive scores in R-1 and R-2, reflecting its capability to accurately capture essential content and present it comprehensibly.

Table 3: ROUGE scores on multi-news dataset

Models	R-1	R-2	R-L
HDSG	40.05	16.35	42.08
Match-sum	46.20	16.51	41.89
HETFORMER	46.21	17.49	42.43
THGS-GMM	46.66	16.90	42.73
SG-SUM	**47.36**	18.61	43.31
Proposed model	47.02	17.68	**43.40**

4.3.3 Highlighting the Innovation and Contribution of the Proposed Model

The proposed model's innovative method is demonstrated by its integration of advanced neural network technologies—combining GATs, Transformer-based BERT, and LDA with TF-IDF enhancements. This incorporation enables the model to analyse and understand not only the linguistic and syntactic aspects of the text but also its thematic structures, thereby producing summaries that are both contextually relevant and content-rich.

The innovation lies in how these technologies are synergized. GATs help in understanding the relationships and dependencies between different parts of the text, BERT brings powerful context-aware text encoding capabilities, and LDA assists in identifying the latent topics within the texts. The addition of TF-IDF metrics further improves the model's ability to weigh and prioritize text based on the significance of terms as they appear across the corpus.

4.3.4 Comprehensive Evaluation and Implications

Both tables collectively validate that while traditional models like Lead-3 and even complex ones like HETFORMER and THGS-GMM show strong performances, the proposed model's architecture offers a nuanced understanding of textual data. It excels not just in summarizing texts from single documents but also shows a profound competence in handling the complexities of multi-document datasets.

This comprehensive performance underscores the proposed model's potential as a handy tool in real-world applications where the quick summarization of extensive and diverse data sets is required. Whether for news aggregation, research, or legal document analysis, the model's ability to quickly cleanse essential information and present it in a coherent, succinct manner could significantly enhance efficiency and decision-making processes.

The presented study not only advances the field of text summarization through innovative technological incorporation but also sets a new benchmark in the performance of automated summarization systems, paving the way for further research and development in this area. Overall, proposed model demonstrates consistently satisfactory results in both single and multi-document extractive summarization tasks, combining advanced architecture with resource efficiency and ease of implementation across diverse settings.

5 Discussion

Table 4 provides an overview of Single Document models, Table 5 in list the Topic-Aware models for extractive summarization while Table 6 outlines the models designed for Multi-Document models.

Table 4: Comparison with single document models

Author(s)	Model name	Description and comparative performance
Wang et al. [17]	HSG	Utilizes a hypergraph-based model for extractive summarization. Incorporates a GNN that encodes word occurrences and sequences using CNN and BiLSTM. Reported to achieve an R-1 score of 42.31 on the CNN/DM dataset, highlighting its effectiveness in capturing word-level interactions.
Xu et al. [57]	JECS	Focuses on reducing redundancy through sentence compression and dependency tree pruning. Achieves notable efficiency in content reduction, with an R-1 score of 41.70 on CNN/DM, emphasizing its utility in summarizing concise content.
Liu et al. [20]	HETFORMER	A model based on heterogeneous transformers tailored for long texts, integrating multi-granularity sparse attentions. Demonstrates superior handling of long content with an R-1 score of 44.55 on CNN/DM, outperforming many traditional models.
Jing et al. [48]	MULTI-GCN	This model introduces a novel MGCG to model relationships among words across sentences, focusing on both inter-sentential and intra-sentential relationships. Its performance, with an R-1 score of 43.16 on CNN/DM, underscores its capability in detailed relational analysis.
Huang et al. [19]	DiscoCorelation-GraphSum	Innovatively uses a hypergraph structure that includes entity, EDU (Elementary Discourse Units), and sentence nodes. By leveraging RST (Rhetorical Structure Theory) discourse parsing, it captures intricate relationships with an R-1 of 43.61 on CNN/DM.
Zhang et al. [34]	DiffuSum	Employs transformer-based diffusion models to create and match sentence representations, enhancing extraction accuracy. Its novel approach yields an R-1 of 44.83 on CNN/DM, demonstrating its effectiveness in matching sentence-level semantic representations.

Overview of Summarization Model Performance

The provided Tables 5 and 6 detail the performance of various text summarization models on two distinct datasets: CNN/DM and Multi-News. These tables assess model effectiveness using the ROUGE metrics: ROUGE-1 (R-1) for unigram overlap, ROUGE-2 (R-2) for bigram overlap, and ROUGE-L (R-L) for the longest common subsequence, which are standard measures to evaluate the quality of text summarization by comparing generated summaries against reference summaries.

Table 5: Comparison with topic aware models

Author(s)	Model name	Description and comparative performance
Cui et al. [13]	Topic-graph-sum	Integrates BERT with latent topics using the Neural Topic Model (NTM) to improve sentence representation. Achieved an R-1 of 44.02 on the CNN/DM dataset, reflecting its robust ability to integrate global thematic elements successfully.
Xie et al. [44]	GRETEL	Employs a graph contrastive topic model alongside a pre-trained language model, focusing on both global and local semantic features. It accomplishes robustly with an R-1 of 44.62 on CNN/DM, showcasing its skill in handling detailed semantic structures.
Phan et al. [3]	THGS-GMM	Uses a graph structure augmented with K-Means and GMM for topic extraction. This model highlights its efficacy in thematic content summarization with an R-1 of 43.28 on CNN/DM, indicating a strong grasp of topic-related nuances.
Joshi et al. [43]	DeepSumm	Focuses on finding hidden themes within documents through topic vectors and sequential learning techniques. It offers a new dimension in summarization with an R-1 of 43.30 on CNN/DM, underscoring its innovative approach to discovering latent thematic elements.

Table 6: Comparison with single to multi-documents models

Author(s)	Model name	Description and comparative performance
Wang et al. [17]	HDSG	Extends document-level associations through super-nodes. This method models complex document structures, demonstrating its efficacy with an R-1 of 40.05 and an R-L of 42.08 on the multi-news dataset, illustrating its robustness in multi-document contexts.
Phan et al. [3]	THGS-GMM (Multi-Docs)	Adjusts their single document model to address multi-document contexts by segregating latent topics from diverse documents. Attained an R-1 of 46.66 and an R-L of 42.73 on multi-news, highlighting its precision in thematic separation and integration.
Chen et al. [58]	SgSum	Rethinks MDS as a sub-graph selection issue, using graphs to link sentences effectively. This model's innovative approach results in an R-1 of 47.36 and an R-L of 43.31 on multi-news, demonstrating superior information synthesis and summary coherence.

This table encapsulates a diverse range of SDS models, highlighting each model's unique approach and effectiveness in handling text summarization challenges. The incorporation of performance scores (ROUGE-1) alongside model descriptions provides a clear comparative perspective, illustrating how

each model stands in relation to benchmarks in the field and particularly against the proposed model. This detailed comparative analysis ensures a comprehensive understanding of the current state of extractive summarization technologies and sets a clear benchmark for the innovation introduced in this study.

This table summarizes the distinct approaches and performance of various topic-aware models in text summarization, comparing their methods and effectiveness as measured by ROUGE-1 scores on the CNN/DM dataset. Each model establishes unique capabilities in handling the complexity of topic integration and thematic accuracy, providing a solid framework for proposed model's evaluation. Through this comparative analysis, this research aims to position proposed model's innovative contributions within the context of these established methodologies, highlighting its advancements and diverse features in handling topic-aware extractive summarization.

This table provides a comparative overview of models that bridge SDS and MDS tasks, featuring innovations that enhance document-level understanding and thematic interconnections. These models are evaluated based on their performance on the Multi-News dataset, with each showing distinct advantages in handling the complexities associated with multi-document contexts. This analysis not only benchmarks their effectiveness but also situates the proposed model amidst these advancements, underscoring contributions to the evolution of summarization technology. This detailed comparison aids in illustrating the innovative aspects of the proposed model, particularly in how it manages and synthesizes information across multiple documents more effectively.

The analysis, synthesizing results from Tables 4–6, reveals the competitive edge and innovative aspects of the proposed model across both SDS and MDS tasks. In the realm of SDS models as in Table 2, the proposed model shows robust performance with an R-1 score of 44.46, closely approaching the benchmark set by Oracle and surpassing other advanced models such as HETFORMER and DIFFU-SUM. Notably, it exhibits a fine balance between content fidelity R-1 and sentence structure preservation R-L with scores of 20.98 and 40.51, respectively.

In MDS shown in Table 3, the proposed model further distinguishes itself by achieving an R-1 of 47.02 and an R-L of 43.40, outperforming sophisticated models like Match-Sum and THGS-GMM, highlighting its superior ability to synthesize information from multiple sources. This performance is particularly impressive given the complexity of integrating thematic and contextual nuances across different documents.

The comparisons drawn with both single and multi-document models, including HDSG, GRE-TEL, and SgSum, underscore proposed model's enhanced capabilities in both thematic depth and structural coherence. The proposed model's architecture leverages advanced neural network techniques to dynamically adapt to the intricacies of the text, setting new standards in extractive summarization. This places proposed research at the forefront, pioneering next-generation summarization solutions that effectively address both the granularity of content and the coherence of summaries.

6 Conclusion, Limitations and Future Work

This study presents a novel integration of GATs with Transformer-based BERT and LDA, further augmented by TF-IDF improvements. This approach significantly refines extractive summarization by effectively capturing inter-sentence relationships and integrating thematic information to ensure global context relevance. Empirical evaluations on the CNN/DM and Multi-News datasets validate that the proposed model not only excels in single-document scenarios, achieving a ROUGE-1 score of 44.46, but also handles the complexities of multi-document datasets with high competence, reflected

by a ROUGE-L score of 43.40. This showcases the proposed model's capability to outperform other advanced summarization models, including THGS-GMM and SG-SUM, in both settings.

However, the model's current formulation does not broadly address resource limitations, which could hamper its deployment in constrained environments. Moreover, the integration of richer semantic units such as entire paragraphs and named entities remains unexplored, which could potentially raise the summarization quality.

Future research will focus on optimizing the model for resource-limited settings and experimenting with the incorporation of richer semantic structures. By expanding the depth of thematic analysis and enhancing the contextual understanding, this research aims to push the boundaries of automated summarization technology further.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Ayesha Khaliq, Salman Afsar, Fahad Ahmad; data collection: Ayesha Khaliq, Muhammad Azam Zia, Muhammad Zafar Iqbal, Fahad Ahmad; analysis and interpretation of results: Ayesha Khaliq, Salman Afsar, Fahad Ahmad; draft manuscript preparation: Ayesha Khaliq, Salman Afsar, Muhammad Azam Zia, Muhammad Zafar Iqbal, Fahad Ahmad. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The dataset used in this study is publicly available.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] A. P. Widyassari *et al.*, "Review of automatic text summarization techniques & methods," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 4, pp. 1029–1046, 2022. doi: 10.1016/j.jksuci.2020.05.006.

[2] K. Sandeep and A. Solanki, "An abstractive text summarization technique using transformer model with self-attention mechanism," *Neural Comput. Appl.*, vol. 35, no. 25, pp. 18603–18622, 2023. doi: 10.1007/s00521-023-08687-7.

[3] T.-A. Phan, N. D. Nguyen, and K.-H. N. Bui, "Extractive text summarization with latent topics using heterogeneous graph neural network," in *Proc. 36th Pacific Asia Conf. Lang., Inform. Comput.*, 2022, pp. 749–756.

[4] J. Chen, "An entity-guided text summarization framework with relational heterogeneous graph neural network," *Neural Comput. Appl.*, vol. 36, no. 7, pp. 3613–3630, 2024. doi: 10.1007/s00521-023-09247-9.

[5] P. P. S. Bedi, M. Bala, and K. Sharma, "Extractive text summarization for biomedical transcripts using deep dense LSTM-CNN framework," *Expert. Syst.*, vol. 41, no. 7, pp. e13490, 2024. doi: 10.1111/exsy.13490.

[6] M. Bidoki, M. R. Moosavi, and M. Fakhrahmad, "A semantic approach to extractive multi-document summarization: Applying sentence expansion for tuning of conceptual densities," *Inform. Process. Manage.*, vol. 57, no. 6, pp. 102341, 2020. doi: 10.1016/j.ipm.2020.102341.

[7] S. Bano, S. Khalid, N. M. Tairan, H. Shah, and H. A. Khattak, "Summarization of scholarly articles using BERT and BiGRU: Deep learning-based extractive approach," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 35, no. 9, pp. 101739, 2023. doi: 10.1016/j.jksuci.2023.101739.

[8] J. Dan, W. Hu, and Y. Wang, "Enhancing legal judgment summarization with integrated semantic and structural information," *Artif. Intell. Law*, vol. 34, pp. 1–22, 2023. doi: 10.1007/s10506-023-09381-8.

[9] Q. Xie, P. Tiwari, and S. Ananiadou, "Knowledge-enhanced graph topic transformer for explainable biomedical text summarization," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 4, pp. 1836–1847, Apr. 2024. doi: 10.1109/JBHI.2023.3308064.

[10] R. Nallapati, F. Zhai, and B. Zhou, "SummaRuNNer: A recurrent neural network based sequence model for extractive summarization of documents," in *Proc. AAAI Conf. Artif. Intell.*, San Francisco, CA, USA, 2017.

[11] A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inform. Process. Syst.*, vol. 30, pp. 1–17, 2017.

[12] L. He, L. Bai, X. Yang, H. Du, and J. Liang, "High-order graph attention network," *Inf. Sci.*, vol. 630, pp. 222–234, 2023. doi: 10.1016/j.ins.2023.02.054.

[13] P. Cui, L. Hu, and Y. Liu, "Enhancing extractive text summarization with topic-aware graph neural networks," arXiv preprint arXiv:2010.06253, 2020.

[14] H. Jin, T. Wang, and X. Wan, "Multi-granularity interaction network for extractive and abstractive multi-document summarization," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguist.*, 2020, pp. 6244–6254.

[15] T.-A. Phan, N.-D. N. Nguyen, and H. N. Bui, "HeterGraphLongSum: Heterogeneous graph neural network with passage aggregation for extractive long document summarization," in *Proc. 29th Int. Conf. Comput. Linguist.*, Gyeongju, Republic of Korea, 2022, pp. 6248–6258.

[16] Z. Song and I. King, "Hierarchical heterogeneous graph attention network for syntax-aware summarization," in *Proc. AAAI Conf. Artif. Intell.*, 2022, vol. 36, no. 10, pp. 11340–11348. doi: 10.1609/aaai.v36i10.21385.

[17] D. Wang, P. Liu, Y. Zheng, X. Qiu, and X. Huang, "Heterogeneous graph neural networks for extractive document summarization," arXiv preprint arXiv:2004.12393, 2020.

[18] S. E. Nourbakhsh and H. B. Kashani, "ConHGNN-SUM: A contextualized heterogeneous graph neural network for extractive text summarization," in *2024 20th CSI Int. Symp. Artif. Intell. Signal Process. (AISP)*, Babol, The Islamic Republic of Iran, 2024, pp. 1–6. doi: 10.1109/AISP61396.2024.10475307.

[19] Y. J. Huang and S. Kurohashi, "Extractive summarization considering discourse and coreference relations based on heterogeneous graph," in *Proc. 16th Conf. Eur. Chapter Assoc. Comput. Linguist.*, 2021, pp. 3046–3052.

[20] Y. Liu *et al.*, "HETFORMER: Heterogeneous transformer with sparse attention for long-text extractive summarization," arXiv preprint arXiv:2110.06388, 2021.

[21] W. Ting *et al.*, "A study of extractive summarization of long documents incorporating local topic and hierarchical information," *Sci. Rep.*, vol. 14, no. 1, pp. 10140, 2024. doi: 10.1038/s41598-024-60779-z.

[22] Y. Huang, Z. Li, Z. Chen, C. Zhang, and H. Ma, "Sentence salience contrastive learning for abstractive text summarization," *Neurocomputing*, vol. 593, pp. 127808, 2024. doi: 10.1016/j.neucom.2024.127808.

[23] D. Anand and R. Wagh, "Effective deep learning approaches for summarization of legal texts," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 5, pp. 2141–2150, 2022. doi: 10.1016/j.jksuci.2019.11.015.

[24] M. Zhang *et al.*, "From coarse to fine: Enhancing multi-document summarization with multi-granularity relationship-based extractor," *Inform. Process. Manage.*, vol. 61, no. 3, pp. 103696, 2024. doi: 10.1016/j.ipm.2024.103696.

[25] D. Debnath, R. Das, and S. Rafi, "Sentiment-based abstractive text summarization using attention oriented lstm model," in *Intell. Data Eng. Anal.: Proc. 9th Int. Conf. Front. Intell. Comput.: Theory Appl. (FICTA 2021)*, Aizwal, India, Springer, 2022, pp. 199–208.

[26] Z. Fatima *et al.*, "A novel approach for semantic extractive text summarization," *Appl. Sci.*, vol. 12, no. 9, pp. 4479, 2022. doi: 10.3390/app12094479.

[27] D. Varab and Y. Xu, "Abstractive summarizers are excellent extractive summarizers," in *Proc. 61st Annu. Meeting Assoc. Comput. Linguist.*, Toronto, ON, Canada, 2023, pp. 330–339.

[28] X. Zhang, F. Wei, and M. Zhou, "HIBERT: Document level pre-training of hierarchical bidirectional transformers for document summarization," arXiv preprint arXiv:1905.06566, 2019.

[29] J. Ainslie *et al.*, "ETC: Encoding long and structured inputs in transformers," arXiv preprint arXiv:2004.08483, 2020.

[30] J. Kwon, N. Kobayashi, H. Kamigaito, and M. Okumura, "Considering nested tree structure in sentence extractive summarization with pre-trained transformer," in *Proc. 2021 Conf. Empir. Methods Nat. Lang. Process.*, Punta Cana, Dominican Republic, 2021, pp. 4039–4044.

[31] J. J. Navjord and J.-M. R. Korsvik, "Beyond extractive: Advancing abstractive automatic text summarization in norwegian with transformers," M.S. thesis, Norwegian Univ. of Life Sciences, Norway, 2023.

[32] L. Bacco, A. Cimino, F. Dell'Orletta, and M. Merone, "Explainable sentiment analysis: A hierarchical transformer-based extractive summarization approach," *Electronics*, vol. 10, no. 18, pp. 2195, 2021. doi: 10.3390/electronics10182195.

[33] K. Shi, X. Cai, L. Yang, J. Zhao, and S. Pan, "StarSum: A star architecture based model for extractive summarization," in *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, 2022, vol. 30, pp. 3020–3031. doi: 10.1109/TASLP.2022.3207688.

[34] H. Zhang, X. Liu, and J. Zhang, "DiffuSum: Generation enhanced extractive summarization with diffusion," arXiv preprint arXiv:2305.01735, 2023.

[35] S.-N. Vo, T.-T. Vo, and B. Le, "Interpretable extractive text summarization with meta-learning and BI-LSTM: A study of meta learning and explainability techniques," *Expert. Syst. Appl.*, vol. 245, pp. 123045, 2024. doi: 10.1016/j.eswa.2023.123045.

[36] D. Jain, M. D. Borah, and A. Biswas, "Summarization of lengthy legal documents via abstractive dataset building: An extract-then-assign approach," *Expert. Syst. Appl.*, vol. 237, pp. 121571, 2024. doi: 10.1016/j.eswa.2023.121571.

[37] Y. Liu and M. Lapata, "Text summarization with pretrained encoders," arXiv preprint arXiv:1908.08345, 2019.

[38] S. Lamsiyah, A. E. Mahdaouy, S. E. A. Ouatik, and B. Espinasse, "Unsupervised extractive multi-document summarization method based on transfer learning from BERT multi-task fine-tuning," *J. Inf. Sci.*, vol. 49, no. 1, pp. 164–182, 2023. doi: 10.1177/0165551521990616.

[39] Y. Qiu and Y. Jin, "Engineering document summarization: A bidirectional language model-based approach," *J. Comput. Inf. Sci. Eng.*, vol. 22, no. 6, pp. 061004, 2022. doi: 10.1115/1.4054203.

[40] Y. Gu, Y. Wang, H.-R. Zhang, J. Wu, and X. Gu, "Enhancing text classification by graph neural networks with multi-granular topic-aware graph," *IEEE Access*, vol. 11, pp. 20169–20183, 2023. doi: 10.1109/ACCESS.2023.3250109.

[41] R. Srivastava, P. Singh, K. Rana, and V. Kumar, "A topic modeled unsupervised approach to single document extractive text summarization," *Knowl.-Based Syst.*, vol. 246, pp. 108636, 2022. doi: 10.1016/j.knosys.2022.108636.

[42] T. Ma *et al.*, "T-BERTSum: Topic-aware text summarization based on BERT," *IEEE Trans. Comput. Soc. Syst.*, vol. 9, no. 3, pp. 879–890, 2021. doi: 10.1109/TCSS.2021.3088506.

[43] A. Joshi, E. Fidalgo, E. Alegre, and L. Fernández-Robles, "DeepSumm: Exploiting topic models and sequence to sequence networks for extractive text summarization," *Expert. Syst. Appl.*, vol. 211, pp. 118442, 2023. doi: 10.1016/j.eswa.2022.118442.

[44] Q. Xie, J. Huang, T. Saha, and S. Ananiadou, "GRETEL: Graph contrastive topic enhanced language model for long document extractive summarization," arXiv preprint arXiv:2208.09982, 2022.

[45] M. Jiang, Y. Zou, J. Xu, and M. Zhang, "GATSum: Graph-based topic-aware abstract text summarization," *Inf. Technol. Control.*, vol. 51, no. 2, pp. 345–355, 2022. doi: 10.5755/j01.itc.51.2.30796.

[46] H. Xu, Y. Wang, K. Han, B. Ma, J. Chen and X. Li, "Selective attention encoders by syntactic graph convolutional networks for document summarization," in *ICASSP 2020–2020 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Barcelona, Spain, 2020, pp. 8219–8223. doi: 10.1109/ICASSP40776.2020.9054187.

[47] Z. Henghui, W. Zhang, M. Huang, S. Feng, and Y. Wu, "A multi-granularity heterogeneous graph for extractive text summarization," *Electronics*, vol. 12, no. 10, pp. 2184, 2023. doi: 10.3390/electronics12102184.

[48] B. Jing, Z. You, T. Yang, W. Fan, and H. Tong, "Multiplex graph neural network for extractive text summarization," arXiv preprint arXiv:2108.12870, 2021.

[49] T. Vo, "An approach of syntactical text graph representation learning for extractive summarization," *Int. J. Intell. Robot. Appl.*, vol. 7, no. 1, pp. 190–204, 2023. doi: 10.1007/s41315-022-00228-0.

[50] Z. Zhang et al., "AHCL-TC: Adaptive hypergraph contrastive learning networks for text classification," *Neurocomputing*, pp. 127989–128000, 2024.

[51] Y. Zhang, Y. Liu, and C. Wu, "Attention-guided multi-granularity fusion model for video summarization," *Expert. Syst. Appl.*, vol. 249, pp. 123568, 2024. doi: 10.1016/j.eswa.2024.123568.

[52] Y. Zhang, Y. Liu, W. Kang, and R. Tao, "VSS-Net: Visual semantic self-mining network for video summarization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 4, pp. 2775–2788, Apr. 2024. doi: 10.1109/TCSVT.2023.3312325.

[53] Y. Wang, Y. Sun, Y. Fu, D. Zhu, and Z. Tian, "Spectrum-BERT: Pretraining of deep bidirectional transformers for spectral classification of Chinese liquors," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–13, 2024. doi: 10.1109/TIM.2024.3427843.

[54] H. Bao et al., "Fine-tuning pretrained transformer encoders for sequence-to-sequence learning," *Int. J. Mach. Learn. Cybern.*, vol. 15, no. 5, pp. 1711–1728, 2024. doi: 10.1007/s13042-023-01992-6.

[55] H. Jelodar et al., "Latent dirichlet allocation (LDA) and topic modeling: Models, applications, a survey," *Multimed. Tools Appl.*, vol. 78, pp. 15169–15211, 2019. doi: 10.1007/s11042-018-6894-4.

[56] B. Ma, "Mining both commonality and specificity from multiple documents for multi-document summarization," *IEEE Access*, vol. 12, pp. 54371–54381, 2024. doi: 10.1109/ACCESS.2024.3388493.

[57] J. Xu and G. Durrett, "Neural extractive text summarization with syntactic compression," arXiv preprint arXiv:1902.00863, 2019.

[58] M. Chen et al., "SgSum: Transforming multi-document summarization into sub-graph selection," arXiv preprint arXiv:2110.12645, 2021.