IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 23 September 2024, accepted 16 November 2024, date of publication 25 November 2024, date of current version 4 December 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3505612

RESEARCH ARTICLE

PRISM: Personalizing Reporting With Intelligent Summarization Through Multiple Frames

KYOHOON JIN[1], SEJIN PAIK[2], JUNGMIN YUN[3], SOOJIN JANG[1], AND YOUNGBIN KIM[1], (Member, IEEE)
[1]Department of Image Science and Arts, Chung-Ang University, Dongjak, Seoul 06974, South Korea
[2]McCourt School of Public Policy, Massive Data Institute, Georgetown University, Washington, DC 20001, USA
[3]Department of Artificial Intelligence, Chung-Ang University, Dongjak, Seoul 06974, South Korea

Corresponding author: Youngbin Kim (ybkim85@cau.ac.kr)

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (Contribution Rate: 25%) under Grant NRF-2022R1C1C1008534; in part by the Institute for Information and Communications Technology Planning and Evaluation (IITP) Grant funded by the Korea Government [Ministry of Science and ICT (MSIT)] (Artificial Intelligence Graduate School Program, Chung-Ang University, Contribution Rate: 25%) under Grant 2021-0-01341; and in part by the Culture, Sports and Tourism Research and Development Program through the Korea Creative Content Agency Grant funded by the Ministry of Culture, Sports and Tourism in 2024 (Project Name: Developing Professionals for Research and Development in Contents Production Based on Generative AI and Cloud, Contribution Rate: 50%) under Project RS-2024-00352578.

ABSTRACT As the challenges of information overload and selection fatigue in the digital news environment persist, natural language processing techniques have become increasingly critical for personalizing information delivery. This paper introduces a novel, frame-specific summarization framework aimed at generating summaries that accurately reflect various frames within news content. Unlike conventional text summarization methodologies that predominantly focus on extracting key frames from a corpus of content, our approach utilizes media framing techniques, coupled with bag-of-words and deep learning model for framing analysis, to extract varied perspectives from a single content and infuse summaries with the relevant contextual semantics. Utilizing a conditional variational autoencoder structure, we created and trained data that embeds multiple frames within a single document, enabling the summary to reflect frame-specific information. Our framework's effectiveness is supported by its superior performance in terms of summarization accuracy on the benchmark dataset compared to baseline models. The implications of this work aim to support public interest by offering the potential to broaden perspectives on complex societal issues and enhance news personalization techniques.

INDEX TERMS Media framing, multi-frame dataset generation, controllable text summarization, multi-task learning.

I. INTRODUCTION

Media and communication studies emphasize the crucial role that framing plays in shaping how messages are conveyed through distinct semantic and visual constructs [1]. An examination of framing effects not only reveals the diverse perspectives embedded within a message but also demonstrates how these differences in framing influence an individual's perception and understanding of the content at hand. The utility of news and media framing analysis is twofold. First, it provides context for the portrayal of specific societal issues [2]. Second, it has the power to reveal the intentions of various actors - such as journalists or politicians - that shape the perspective of a news narrative. Considering the substantial influence that the dissemination of public affairs information wields on a society's political and societal dynamics, news framing analysis serves as a crucial instrument for illuminating potential biases embedded within specific narratives.

Historically, communication researchers have relied on manual methodologies, such as content analysis and surveys, to unearth patterns and characteristics of news framing [3],

The associate editor coordinating the review of this manuscript and approving it for publication was Massimo Cafaro.

[4], [5]. These traditional methods, however, have been constrained by the limited volume of accessible data. Classical framing scholars have identified various frame categories, such as episodic versus thematic framing [3], generic versus issue-specific [4], and gain versus loss [6]. Although these categories have proven to be pervasive, the increased availability and diversity of multinational news datasets necessitate an adaptive approach to frame identification that can effectively handle the complexity and breadth of the contemporary information environment.

Advancements in computational methodologies have driven the application of machine learning techniques to enrich news framing scholarship. Although previous work has leveraged Machine Learning (ML) and Natural Language Processing (NLP) techniques for automated tasks ranging from frame discovery and annotation [7], [8], [9] to detecting framing bias through summarization techniques [10], and creating tasks to reframe or reconstruct messages [11], [12], the utilization of framing to generate diversified news content remains largely unexplored. The primary goal of this paper is to employ summarization techniques to facilitate the automated generation of diversified, context-sensitive news narratives, providing a new direction for the application of deep learning in media and journalism research.

This study focuses on the detection of the multiplicity of frames within a single document, such as a news article. A given set of sentences or paragraphs within a single story contains diverse frames [13]. Thus, a deep learning model tasked with this challenge must simultaneously comprehend the content defining the frame within a document and identify the salient aspects for summarization. This dual requirement has not been extensively investigated in prior research, and available benchmark summary data do not provide frame-by-frame summarizations for a single document. This gap in existing studies underlines the need for new approaches to enable a more nuanced understanding of framing within documents and its impact on automated summarization tasks.

In this paper, we introduce a novel framework for generating context-specific news summaries infused with diverse frames extracted from a single document. The methodology unfolds as follows: First, topic modeling is employed for frame labeling. Subsequently, for each topic within a single document, extractive summarization is deployed to generate unique summaries, thereby addressing the constraints of earlier research by producing data that includes a summary for each frame in the same document. Furthermore, multi-task learning is utilized to train both frame classification and frame-dependent abstractive summarization using the conditional latent variable model. Empirically, our approach in both frame categorization and frame-triggered abstractive summarization outperforms existing baselines, yielding superior ROUGE scores compared to previously proposed frame-related deep learning models. This novel approach paves the way for an enhanced understanding and application of frame-based summarization in information processing tasks.

II. RELATED WORK

In this section, we provide a summary of the research on frame analysis methods and customizing summarization approaches.

A. FRAME ANAYLSIS

In the face of the overwhelming volume of accessible digital information in our modern era, the domain of communication and media studies promises considerable potential for advancing knowledge discovery through computational techniques. This potential is gradually being realized by scholars in communication technology and big data analysis, who are incorporating machine learning approaches to scrutinize extensive and diverse datasets. Of particular note is the integration of computational methods for media framing analysis. The essence of framing analysis lies in extracting predominant contexts and salient aspects of messages, facilitating a deeper understanding of the underlying intentions and motivations of specific texts [2], [14]. Framing enables the discovery of patterns and critical discourse in messages related to political, social, and cultural issues, providing insightful perspectives and enhancing our interpretation of these complex subjects [15].

In the sphere of communication scholarship, framing involves identifying specific attributes or elements of a topic that are emphasized within a given text [16]. The accumulation of these attributes or elements forms a frame. It is crucial to note that the mere presence or frequency of words does not directly translate to framing identification. Consequently, equating frames with topics would be inaccurate. A framework emerges when a semantic relationship is established among words and sentences, and these interconnected sentences yield a robust contextualization of meaning. The employment of computational methodology in framing analysis could lead to confusion, as numerous existing ML models are capable of generating topics from text. Thus, framing analysis is not merely the immediate output of ML-generated topics; it involves an additional step where the topics generated by ML models are synthesized into frames, requiring a nuanced and contextual approach to interpreting a message.

Historically, frame extraction was carried out manually, posing significant limitations in terms of scalability and objectivity. However, recent advancements have facilitated the emergence of automated or semi-automated methods for framing extraction. For example, Field et al. [17] embarked on a comprehensive study analyzing thirteen years of Russian news articles, aiming to dissect the portrayal of the United States within Russia's media system. Through the application of embedding-based methods, the study identified moral failings and threats as the most recurrently surfaced frames in describing the role of the U.S. Another study, FrameAxis, employed word embeddings to pinpoint the most salient semantic axes or biases of "microframes," providing a nuanced characterization of textual content. These endeavors underscore the potential of leveraging computational methods for intricate frame analysis.

Significant advancements have also been made in utilizing topic modeling techniques for frame identification. For instance, Walter et al. [18], [19] introduced the Analysis of Topic Model Networks (ANTMN), a novel approach tailored to decipher the frames within U.S. news articles. In a parallel development, Bhatia et al. [20] presented OpenFraming, an open-source tool engineered to facilitate framing analysis of multilingual data. This framework combines unsupervised machine learning through Latent Dirichlet Allocation (LDA) topic modeling and supervised learning via Support Vector Machines (SVM), thereby constructing a robust system capable of frame extraction. These efforts highlight the potential of integrating various machine learning techniques to reveal complex framing structures.

B. CUSTOMIZING SUMMARIES THROUGH SEMANTIC FRAMES

Existing methods that aim to control summarization from a specific perspective primarily focus on aspect- or keyword-specific data points [19], [21]. Studies have been conducted to generate query-based customized summaries based on aspect keywords in opinions [22]. Other researchers have explored entity-based text summarization models, which extract entities from the original texts and generate summaries using these guiding entities [23]. Efforts have also been made to create summaries that are focused on specific keywords and queries [24]. For instance, CTRLsum [25] allows users to manipulate various aspects of the generated summary using text inputs in the form of keywords. However, our proposed approach differs from the aforementioned methods. While keywords and aspects only provide partial fragments of information, frames constitute a higher-level construct that captures both semantics and structure, offering a more comprehensive understanding of entire sentences. Thus far, studies summarizing the integration of information, such as comprehensive text frames, have been somewhat limited.

Another area of research aims to improve the performance of document summaries by utilizing topic information [26], [27], [28]. These studies primarily focus on enhancing the quality of summary statements by leveraging topic distributions rather than generating summary statements for each specific topic. Our proposed method aligns with approaches such as Pointer-Generator [29], which aims to generate a summary for each topic. Past studies have largely relied on RNN-based seq2seq models and have commonly used a rule-based approach for generating training data. However, we propose a novel combination of deep learning and conditional latent variable models to generate frame-based summary sentences. Furthermore, in contrast to the conventional rule-based approach, we employ extractive summarization to generate training data for summary sentences based on topics.

TABLE 1. Keywords for seven example topics and their frames in CNN/DM dataset.

Keywords	Frame
League, Cup, Season, Club, Player, Games	Sports
Hospital, Love, Baby, Cancer, Mother	Healthcare
Murder, Arrested, Prison, Officers, Victim, Court	Law Enforcement
Obama, President, Military, Forces, Security	Defense Politics
Apple, Google, Company, Cent, Users	Information Technology
Water, Weather, Space, Earth, Snow, Storm	Weather
Labour, Minister, Tax, Pay, Party	Political Economy

Recent advances in Large Language Models (LLMs) may also be used to facilitate summarization. However, since LLMs are trained on large datasets, their summaries already have the potential for bias. LLMs have yet to completely overcome biases regarding underrepresented groups, races, and genders, as well as toxicity such as hate speech and profanity [30], [31]. In addition, LLMs, such as ChatGPT,[1] have limitations in representing various frames. ChatGPT responds by stating that it cannot provide answers to avoid political disputes, especially when asked to provide summaries from a specific political perspective. Instead, we propose a method to set a frame inspired by OpenFraming [20] and generate the news summary accordingly.

III. METHODOLOGY

Our model consists of three components. First, we identify and analyze the frame of the document. Then, we assign a frame to each document and generate multi-frame data using the data augmentation method. Finally, we introduce a conditional latent variable model for generating a frame-based summary using a deep learning model. Adding to the recent line of Transformer-based studies that utilize multi-task learning, we apply Transformer's encoder and decoder to perform frame classification and frame-based summarization. The overview of our method is illustrated in Figure 1.

A. FRAME ANAYLSIS

Our approach to frame analysis is inspired by the Openframing framework [20]. However, we replaced LDA topic modeling with a neural topic modeling method inspired by BERTopic [32]. Topic modeling methods, such as LDA and NMF, have primarily been used for computational frame analysis. However, the existing approach uses a Bag-of-Words (BoW) representation, which ignores the semantic relationships between words. As such, BoW representation

[1] https://openai.com/blog/chatgpt

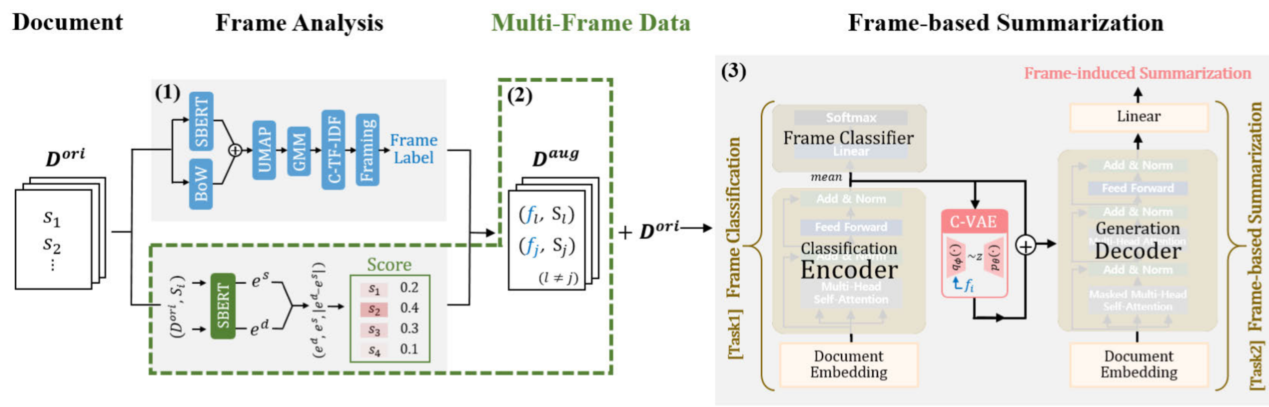

FIGURE 1. **Overall architecture of our proposed method. In the (1) frame analysis step, frames are assigned to each document. Then, in the (2) data augmentation step, we generate multi-frame data that contains multiple frames in a single document. Finally, in the (3) summarization with frame classification step, we generated frame-based summarization using the encoder and decoder of the Transformer with conditional latent variables.**

alone does not capture the contextual meaning of words within sentences. Therefore, there is a limitation to using BoW to represent long sentence documents, such as news articles. On the other hand, BERTopic clusters document embeddings to generate topic representations and extract keywords. However, our goal is not to end with topic modeling, but to identify the frames. Frame analysis aims to extract salient features while also understanding the context. To achieve this goal, we combine BoW and BERT to capture both the contextual meaning and semantic features of a word.

First, we extract contextualized embedding vectors using BERT and generate a BoW vector using the vocabulary of BERT. Then, we combine the two vectors by averaging the values of each token in the contextualized embedding vector. By combining the dimensions of the vector obtained through BERT with the dimensions of BoW, the resulting dimension becomes very large. For data in high-dimensional space, clustering performance may decrease [33]. To prevent this, we apply dimensionality reduction using UMAP [34] same as BERTopic.

Subsequently, we apply the Gaussian Mixture Model (GMM). A GMM is a probabilistic model that assumes all the data is generated from Gaussian distributions with unknown parameters. We use the combination of UMAP and GMM because it demonstrates good clustering performance for features extracted by UMAP [35]. Finally, topic modeling is performed by applying c-TF-IDF to each cluster, and important words are selected for each class. Inspired by BERTopic, we used c-TF-IDF, which generalizes the classic TF-IDF [36] to clusters of documents to determine the importance of words in clusters rather than focusing on the importance of words in individual documents. Furthermore, beyond simply selecting important words for each class, we generate a frame label by analyzing and interpreting framing based on its keywords. In this paper, we conducted a series of experiments to determine the optimal number of frames, ranging from 4 to 10, and identified 7 topics that provide the most coherent and straightforward explanations for the news. The results for the seven frames and their keywords are shown in Table 1.

The proposed model in this paper utilizes the Transformer encoder to perform frame classification. Instead of using only the first token, like BERT, we use the average of the feature maps of each token that has been passed through the Transformer encoder. We let $D^{ori} = \{d^{ori(i)}\}_{i=1}^{N}$ be an original dataset of N documents, and $F = \{f^{(i)}\}_{i=1}^{N}$ be a documents D^{ori}'s frame. We have the following objective function:

$$\mathcal{L}_{cls} = \frac{1}{N}\sum_{i=1}^{N} log\, p_{enc}(f|d_i^{ori}) \tag{1}$$

B. GENERATING MULTI-FRAME DATASET

A single document can contain multiple frames. The longer the document is, the more likely its content will be spread across multiple frames [29]. However, readers' interest and attention may be limited to specific areas. To meet the information needs of these readers, it is helpful to generate summaries that focus on specific frames of interest. In order to train a model that can generate diverse summaries for the same document, we require a dataset that includes multiple sub-subjects and issue areas within the same document [37]. However, most existing datasets are not suitable for considering such a scenario when detecting multiple frames within text. To address this limitation, we propose a method for generating a dataset that enables learning to summarize each point of view within a single document that contains multiple frames.

The dataset consists of two different frames and summary pairs, denoted as (f_l, S_l) and (f_j, S_j), within one document d^{ori}. The procedure for creating the dataset is as follows: First, 1) frame labeling is performed through frame analysis, and secondly, 2) we use Sentence-BERT to calculate the relevancy score between the document and its individual sentences. Finally, 3) after extracting the summary sentence based on the score, we randomly select summary and frame pairs with different frame labels and combine them.

First, we apply our proposed frame analysis method to the CNN/Dailymail dataset to extract frame information for each document. Then, for each sentence in the document, we label it as 1 if the sentence is included in the summary and as 0 otherwise. After this preprocessing, we apply Sentence-BERT because it can extract the degree of semantic similarity between sentences rather than just the degree of overlap between individual words. Given a document d^{ori} that contains n sentences, we use pre-trained Sentence-BERT to generate embeddings for both the document d^{ori} and each individual sentence s. For document embedding, we apply mean pooling to the output of BERT to capture the overall meaning of all words without bias towards specific words. Given the embedding of document d^{ori} represented as e_d and the embedding of sentence s represented as e_s, we concatenate e_d, e_s, and $|e_d - e_s|$. Then, the combined values enter the input of the dense layer and output the final score. This final score represents the relevance between the sentence and the document. The higher the score, the more likely the sentence will be included in the document's summary.

Based on the sentence index idx corresponding to the highest score, we extract the summary sentence $S = \{s_{idx-2}, s_{idx-1}, s_{idx}, s_{idx+1}, s_{idx+2}\}$. By incorporating the sentence with the highest score along with the surrounding sentences, we can capture contextual information and avoid making S too similar to the gold summary.

Finally, we combine summary sentences S with different frames to create a combined document that encompasses multiple frames. For all summary sentences S, the following method is applied: For the summary sentence S_l and its corresponding frame f_l, a different frame $f_j (l \neq j)$ and its corresponding summary sentence S_j are randomly selected. Then, by combining S_l and S_j, one document d^{aug} consists of different frames f_l and f_j. Through this proposed method, the original document D^{ori} is augmented with a single document D^{aug} containing multiple frames. Finally, we train a frame-based summarization model using D which combines D^{ori} and D^{aug}.

C. FRAME-BASED SUMMARIZATION

We use the conditional latent variable model to incorporate information about a specific frame into the model. A latent variable refers to a variable that is inferred from another variable using mathematical methods. When a latent variable is identifiable within the data, it becomes possible to generate additional data that preserves the distinctive attributes of the original data. Models can utilize latent variables to extract frame information from the data, enabling the generation of summaries that are biased toward specific frames.

Among the latent variable models, the VAE stands out for its proficiency in generating data using latent variables. Nevertheless, VAE disregards the type of input data, as it directly models the latent variable z based on the input data. Consequently, it is limited to original documents and lacks any incorporation of frame-related information. However, the goal of our model is to generate summaries that reflect the frame information. To achieve this, we use a latent variable that effectively captures frame elements. This is accomplished through two methods: Conditional Variational Autoencoder (C-VAE) and multi-task learning.

We denote $Z = \{z^{(i)}\}_{i=1}^{N}$ as a subset of latent variables. Each $d^{(i)}$ is a document, and $z^{(i)}$ is the corresponding latent variable for $d^{(i)}$. In this paper, the latent variable z is generated conditionally based on f, which represents frame information. C-VAE encodes a prior distribution $p_\theta(z|d)$, which captures the specialized latent variable z for each frame. Subsequently, it learns the posterior distribution $q_\phi(z|d,f)$ using $p_\theta(d|z,f)$ parameterized by θ. C-VAE is trained to maximize the conditional log-likelihood by employing variational inference through the Evidence Lower Bound(ELBO):

$$\mathcal{L}_{C_{ELBO}} = \frac{1}{N}\sum_{i=1}^{N} log\ p_\theta(f|d, z^{(i)}) - KL(q_\phi(z|d,f)||p_\theta(z|d)) \tag{2}$$

The latent variable z is added to the feature vector that is passed through the encoder. The obtained value replaces the key value in the decoder. That is, it replaces the output value generated by the encoder. Instead of solely feeding the decoder with the original text information, the inclusion of frame information via C-VAE serves to balance the model, avoiding the risk of fixating excessively on frame details or diluting their impact.

The overall goal of this model is to create frame-based summarization. Achieving this objective depends on establishing a thorough understanding of the frame within the encoder. Therefore, we use the encoder that has already been learned and refined in frame classification through multi-task learning instead of retraining the entire transformer architecture from scratch to generate the summary. Finally, we have the following objective function $\mathcal{L}_{Total}$:

$$\mathcal{L}_{Total} = \alpha\mathcal{L}_{C_{ELBO}} + (1-\alpha)\mathcal{L}_{cls} \tag{3}$$

where α is a hyperparameter that $0 < \alpha \leq 1$. We set α to 0.5 in order to achieve a balance between training frame information and summary.

IV. EXPERIMENTS

In this section, we describe the experimental setup and present qualitative and quantitative results of frame analysis and frame-based summarization. The results of the proposed method's hyperparameter experiments and ablation study are also provided.

A. EXPERIMENTAL SETTING

We conducted a summarization experiment on the CNN/Dailymail dataset[2] [38] to evaluate the capability of our

[2]https://huggingface.co/datasets/cnn_dailymail

proposed model to generate frame-induced summaries. To address the limitations of existing studies and datasets, we created a dataset that includes a summary for each frame in a multi-frame document. This newly generated data consists of the sentences that are most relevant to the summary, along with the sentences that precede and follow it. While this configuration imbues the data with frame-related information, it decreases the amount of content that the model needs to summarize within the document as a whole compared to the existing CNN/Dailymail dataset. To address the limitation of insufficient information, we first combine the existing CNN/Dailymail dataset D^{ori} with a newly generated dataset D^{aug}. Therefore, the model is trained to detect frame-specific contextual information and generate summaries based on this awareness. Secondly, we use the combined dataset for the initial 5 epochs and subsequently train the model using the existing CNN/Dailymail dataset. This strategic approach allows us to train a model that can generate summaries specifically tailored to the desired frame while still maintaining proficiency in generating comprehensive summaries.

Since there are not many existing studies on models that generate summaries frame by frame, we compared them with the Pointer-Generator [29] using topic modeling and LSTM and PPLM using GPT-2. The topic-aware pointer-generator network proposes an attention-based RNN framework that generates multiple summaries for a single document, tailored to different topics of interest. This network consists of an LSTM encoder and decoder and utilizes an attention mechanism and topic vector to generate summary sentences. The Plug and Play Language Model (PPLM) is based on GPT-2 [39]. This model guides GPT-2 to create attributes by incorporating user-defined bag-of-words or lightweight classifiers. In PPLM, sentences are generated and compared using the BoW method. The results exclude non-overlapping topics, such as sports, and focus solely on overlapping topics, like politics.

For data preprocessing, special characters are first removed, and words from the stopwords list provided by NLTK [40] are excluded. The headings are extracted using the WordNetLemmatizer provided by NLTK. The maximum length of the input text is 768, and the output text is limited to a maximum length of 300. We use a pre-trained BART-Base model with 139M parameters. Ralamb plays a role as an optimizer for learning. The learning rate is set to 5e-5, and we also utilize a warm-up learning rate scheduler. We manually select hyperparameters based on the ROUGE score. For all tasks, we use four RTX 3090 GPUs. During inference, summaries are generated using beam search, and the beam size is set to 3.

B. FRAME CLASSIFICATION RESULTS

To evaluate the performance of our model in frame classification, we conducted classification training using multiple models with our frame-labeled data. For comparison, we use models based on LSTM-RNN liu2016recurrent and CNN kim2014convolutional, which are primarily used in text classification. Also, as used in Openframing, BERT [41], one of the large-scale pre-training language models, is trained. Finally, we compare the performance of our model when trained using multi-task learning against its performance without such training. Following recent studies, we use Macro and Micro F1 scores for evaluation.

TABLE 2. The performance comparison of different models for frame classification of 7 frames by macro and micro F1-score.

	Macro-FI	Micro-F1
CNN	45.80	21.99
RNN	47.14	23.23
BERT	62.87	43.21
BART	63.46	39.80
Ours	68.23	53.71

TABLE 3. The performance comparison of different models for frame-based summarization by ROUGE score. The experiment was conducted using all 7 data frames.

	ROUGE-1	ROUGE-2	ROUGE-L
Pointer-generator	34.1	13.6	31.2
PPLM	38.7	16.6	35.2
Ours	**43.4**	**21.9**	**37.5**
LEAD-3	38.3	16.7	34.7
BART-Base	43.5	20.6	37.4
T5-Small	41.1	19.7	38.3

TABLE 4. Ablation study. Ours is the result of both augmentation and multi-task learning.

	ROUGE-1	ROUGE-2	ROUGE-L
Ours	43.4	21.9	37.5
w/o augmented data	29.1	23.1	34.5
w/o multi-task learning	38.9	15.1	36.5

Table 2 shows the results of frame classification. Previous studies on frame detection have primarily utilized deep learning models, such as CNN and LSTM. These previous studies showed good performance even without a complex process, owing to the fact that frame detection was performed within a given title. However, it's worth noting that extracting a frame from a title is inevitably restricted. The primary objective of this study is to consider diverse frames within the same article. So, we searched for the frame within the body of the article rather than solely relying on the title. Compared to other deep learning models such as CNN and LSTM, the Transformer architecture demonstrates superior performance. As the large-scale, pre-trained language model based on Transformer showed better performance, other recent frame-related studies also used the BERT model. However, more recent models, such as BART, have demonstrated superior capabilities compared to these BERT-based models. In particular, the model trained through our proposed multi-task learning approach shows the best performance. It is important to emphasize that an accurate and thorough comprehension of the entire document is crucial for both summarization and frame classification.

TABLE 5. Frame-based summaries generated by our proposed model for articles from CNN/Dailymail dataset.

Reference
Over 30 Premiership and Championship games scheduled this weekend. More than half the games are set to be affected by delays and disruption. Engineering works affect routes to London Euston, hitting west coastline. Labour says ministers have failed to learn from the Boxing Day chaos.
Sports Frame
More than 30 Premiership and Championship games scheduled for the long weekend. But large parts of the rail network are set to be closed for repair works. More than half the games will be affected by delays and disruption. This weekend's crunch Premiership fixture, Arsenal versus Liverpool in London, will be hit by the rail chaos.
Political Economy Frame
More than 30 Premiership and Championship games are scheduled for the long weekend. But large parts of the rail network are set to be closed for repair works. Labour has accused the Government of failing to learn from the Boxing Day chaos when nearly a million football fans faced nightmare journeys.

C. QUANTITATIVE RESULTS OF SUMMARIZATION

We compare our proposed model against the baselines using the ROUGE metrics[3] [42]. The results are shown in Table 3. Our model demonstrated the best performance compared to models that generate summaries by frame. Transformer-based models outperform LSTM-based models, such as the Topic-aware pointer-generator. This is because Transformer-based models perform better with long sentences compared to RNN-based models. Since there are not many existing studies on models that generate summaries frame by frame, we compared them with PPLM using GPT-2, and our model achieved a higher ROUGE score. Our model outperformed PPLM in terms of comprehensively understanding the frame. This is because PPLM did not learn the contents of the frame and instead treated the frame information as an attribute.

In addition, we conducted an experiment to verify how the proposed model performed compared to the existing state-of-the-art models. As a baseline, we used BART, which has only been pre-trained without using any of the techniques proposed in this paper; T5, which shows good performance among pre-trained language models; and LEAD-3, which is commonly employed for comparing summary performance. To ensure a fair comparison based on parameter count, we used T5-small and BART-Base, which align with the number of parameters in our model. As can be seen from Table 3, our proposed model outperforms the LEAD-3 and BART models, excelling not only in frame-based summarization but also in comparison with state-of-the-art models. In particular, it demonstrates strong performance in general summarization compared to T5, which is pre-trained using large-scale data.

Additionally, we conducted an ablation study to evaluate the effectiveness of the proposed method. The results of the ablation study are shown in Table 4. First, without augmented data, we achieved a decent ROUGE-L score of 34.5. This indicates a partial inability to fully discern where the model should focus more on extracting frame-related information. Similar observations arise from multi-task learning. Although the frame classifier is not trained together, some frame information can be obtained through augmented data. However, the best performance can be seen when all the methods proposed in this paper are combined. This underscores the efficiency of obtaining frame information by combining augmented data and multi-task learning.

D. QUALITATIVE RESULTS OF SUMMARIZATION

We assess summary generation by applying distinct frames to the same article for qualitative evaluation. The original story revolves around soccer fans being unable to attend a game due to a railway closure. Two frames are discernible in this story. One is on the sports frame regarding a soccer game. The other frame is political, as indicated in the last sentence, where Labor addresses the government's failure to improve the transportation system. This included discussions about political affairs and government responsibilities. The summary is presented in Table 4.

When we applied our frame-based summarization task, we discovered several valuable insights. One of them occurred between the results for sports- and political-specific frames. While most of the core summarization content of the output for each frame remained the same, there were some noticeable differences. For example, the last sentence

[3]We use TorchMetrics to calculate ROUGE metrics; https://torchmetrics.readthedocs.io/en/latest/

emphasizes the content related to the specific frame assigned between the sports frame and the political frame. In the sports frame, the last sentence highlights the influence of railroads on soccer teams' performance and schedules. For the political frame, the final sentence offers more context on the consequences of the government's failure to support public transportation infrastructure. As another example of the result, the name of a specific soccer team is excluded from the political frame but is mentioned more frequently in the sports frame. This is supported by the higher weight that our model assigns to team names when learning about the characteristics of a sports frame. These examples demonstrate that our model has successfully comprehended each frame parameter and has been able to incorporate appropriate contextual phrasing and/or words for those frames.

V. CONCLUSION

In this paper, we introduce a novel deep learning framework that analyzes frames and generates summaries that reflect the frames. Our approach leverages topic modeling and framing to identify frames within an entire document. To enhance the model's comprehension of diverse frames, we generate additional data, enabling the model to gain insights into each frame. Frame-based summarization is performed using conditional latent variables. Experimental results show that our proposed model is superior to baseline models based on the ROUGE score. The summary sentences generated for different frames effectively capture the assigned frames, demonstrating the effectiveness of our model.

Our method excels in single-frame scenarios, but users might be interested in multiple frames. Recent studies have led to the exploration of multi-label frameworks, which involve identifying two or more frames, including the characteristics of multi-lingual frames. Thus, our future work intends to investigate the capability and potential of our model for generating multi-label and multi-lingual frame-based summarization.

While our focus is on frame-based summarization, not general abstractive summarization, we conducted our experiments using the CNN/Dailymail dataset due to the absence of datasets specifically designed for frame-specific summaries. Therefore, our evaluation primarily involves quantitative performance assessment, with limited qualitative evaluation focused on detecting frame-related biases. For future work, a comprehensive frame-based summarization dataset should be used for a more robust performance evaluation.

ACKNOWLEDGMENT

(Kyohoon Jin, Sejin Paik, and Jungmin Yun contributed equally to this work.)

REFERENCES

[1] S. D. Reese and S. C. Lewis, "Framing the war on terror: The internalization of policy in the U.S. press," *Journalism*, vol. 10, no. 6, pp. 777–797, Dec. 2009.

[2] R. M. Entman, "Framing: Towards clarification of a fractured paradigm," *McQuail's Reader Mass Commun. Theory*, vol. 390, p. 397, Jun. 1993.

[3] T. E. Patterson, "Is anyone responsible? How television frames political issues. By Shanto Iyengar. Chicago: University of Chicago press," *Amer. Political Sci. Rev.*, vol. 86, no. 4, pp. 1060–1061, Dec. 1992.

[4] M. C. Nisbet, "Communicating climate change: Why frames matter for public engagement," *Environ., Sci. Policy Sustain. Develop.*, vol. 51, no. 2, pp. 12–23, Mar. 2009.

[5] P. S. Hart and L. Feldman, "Threat without efficacy? Climate change on U.S. network news," *Sci. Commun.*, vol. 36, no. 3, pp. 325–351, Jun. 2014.

[6] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," in *Handbook of the Fundamentals of Financial Decision Making: Part I*. Singapore: World Scientific, 2013, pp. 99–127.

[7] V.-A. Nguyen, J. Boyd-Graber, P. Resnik, and K. Miler, "Tea party in the house: A hierarchical ideal point topic model and its application to republican legislators in the 112th congress," in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process.*, 2015, pp. 1438–1448.

[8] E. Baumer, E. Elovic, Y. Qin, F. Polletta, and G. Gay, "Testing and comparing computational approaches for identifying the language of framing in political news," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Human Lang. Technol.*, 2015, pp. 1472–1482.

[9] D. Card, A. E. Boydstun, J. H. Gross, P. Resnik, and N. A. Smith, "The media frames corpus: Annotations of frames across issues," in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process.*, 2015, pp. 438–444.

[10] N. Lee, Y. Bang, T. Yu, A. Madotto, and P. Fung, "NeuS: Neutral multi-news summarization for mitigating framing bias," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Human Lang. Technol.*, 2022, pp. 3131–3148.

[11] W.-F. Chen, K. A. Khatib, B. Stein, and H. Wachsmuth, "Controlled neural sentence-level reframing of news articles," in *Proc. Findings Assoc. Comput. Linguistics*, 2021, pp. 2683–2693.

[12] T. Chakrabarty, C. Hidey, and S. Muresan, "ENTRUST: Argument reframing with language models and entailment," 2021, *arXiv:2103.06758*.

[13] A. F. Akyürek, L. Guo, R. Elanwar, P. Ishwar, M. Betke, and D. T. Wijaya, "Multi-label and multilingual news framing analysis," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 8614–8624.

[14] D. Chong and J. N. Druckman, "Framing theory," *Annu. Rev. Political Sci.*, vol. 10, no. 1, pp. 103–126, 2007.

[15] D. A. Scheufele and D. Tewksbury, "Framing, agenda setting, and priming: The evolution of three media effects models," *J. Commun.*, vol. 57, no. 1, pp. 9–20, Mar. 2007.

[16] J. Matthes and M. Kohring, "The content analysis of media frames: Toward improving reliability and validity," *J. Commun.*, vol. 58, no. 2, pp. 258–279, Jun. 2008.

[17] A. Field, D. Kliger, S. Wintner, J. Pan, D. Jurafsky, and Y. Tsvetkov, "Framing and agenda-setting in Russian news: A computational analysis of intricate political strategies," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2018, pp. 3570–3580.

[18] D. Walter and Y. Ophir, "News frame analysis: An inductive mixed-method computational approach," *Commun. Methods Measures*, vol. 13, no. 4, pp. 248–266, Oct. 2019.

[19] X. Liang, J. Li, S. Wu, J. Zeng, Y. Jiang, M. Li, and Z. Li, "An efficient coarse-to-fine facet-aware unsupervised summarization framework based on semantic blocks," in *Proc. 29th Int. Conf. Comput. Linguistics*, 2022, pp. 6415–6425.

[20] V. Bhatia, V. P. Akavoor, S. Paik, L. Guo, M. Jalal, A. Smith, D. A. Tofu, E. E. Halim, Y. Sun, M. Betke, P. Ishwar, and D. T. Wijaya, "OpenFraming: Open-sourced tool for computational framing analysis of multilingual data," in *Proc. Conf. Empirical Methods Natural Lang. Process., Syst. Demonstrations*, 2021, pp. 242–250.

[21] F. Kunneman, S. Wubben, A. van den Bosch, and E. Krahmer, "Aspect-based summarization of pros and cons in unstructured product reviews," in *Proc. Int. Conf. Comput. Linguistics*, 2018, pp. 2219–2229.

[22] A. Bražinskas, R. Nallapati, M. Bansal, and M. Dreyer, "Efficient few-shot fine-tuning for opinion summarization," 2022, *arXiv:2205.02170*.

[23] C. Zheng, Y. Cai, G. Zhang, and Q. Li, "Controllable abstractive sentence summarization with guiding entities," in *Proc. 28th Int. Conf. Comput. Linguistics*, 2020, pp. 5668–5678.

[24] M. T. R. Laskar, E. Hoque, and J. X. Huang, "WSL-DS: Weakly supervised learning with distant supervision for query focused multi-document abstractive summarization," in *Proc. 28th Int. Conf. Comput. Linguistics*, 2020, pp. 5647–5654.

[25] J. He, W. Kryscinski, B. McCann, N. Rajani, and C. Xiong, "CTRLsum: Towards generic controllable text summarization," in *Proc. Conf. Empirical Methods Natural Lang. Process.* Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, 2022, pp. 5879–5915. [Online]. Available: https://aclanthology.org/2022.emnlp-main.396
[26] Y. Zou, L. Zhao, Y. Kang, J. Lin, M. Peng, Z. Jiang, C. Sun, Q. Zhang, X. Huang, and X. Liu, "Topic-oriented spoken dialogue summarization for customer service with saliency-aware topic modeling," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 16, pp. 14665–14673.
[27] X. Liang, S. Wu, C. Cui, J. Bai, C. Bian, and Z. Li, "Enhancing dialogue summarization with topic-aware global- and local- level centrality," 2023, *arXiv:2301.12376*.
[28] P. Cui, L. Hu, and Y. Liu, "Enhancing extractive text summarization with topic-aware graph neural networks," in *Proc. 28th Int. Conf. Comput. Linguistics*, 2020, pp. 5360–5371.
[29] K. Krishna and B. V. Srinivasan, "Generating topic-oriented summaries using neural attention," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2018, pp. 1697–1705.
[30] S. Gururangan, D. Card, S. Dreier, E. Gade, L. Wang, Z. Wang, L. Zettlemoyer, and N. A. Smith, "Whose language counts as high quality? Measuring language ideologies in text data selection," in *Proc. Conf. Empirical Methods Natural Lang. Process.*. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, 2022, pp. 2562–2580.
[31] S. Gehman, S. Gururangan, M. Sap, Y. Choi, and N. A. Smith, "RealToxicityPrompts: Evaluating neural toxic degeneration in language models," in *Findings Assoc. Comput. Linguistics, EMNLP*. Stroudsburg, PA, USA: Association for Computational Linguistics, Nov. 2020, pp. 3356–3369.
[32] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," 2022, *arXiv:2203.05794*.
[33] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, "On the surprising behavior of distance metrics in high dimensional space," in *Proc. Int. Conf. Database Theory*. Cham, Switzerland: Springer, 2001, pp. 420–434.
[34] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform manifold approximation and projection for dimension reduction," 2018, *arXiv:1802.03426*.
[35] C. Malzer and M. Baum, "A hybrid approach to hierarchical density-based cluster selection," in *Proc. IEEE Int. Conf. Multisensor Fusion Integr. Intell. Syst. (MFI)*, Sep. 2020, pp. 223–228.
[36] J. Ramos, "Using TF-IDF to determine word relevance in document queries," in *Proc. 1st Instructional Conf. Mach. Learn.*, 2003, vol. 242, no. 1, pp. 29–48.
[37] H. Hayashi, P. Budania, P. Wang, C. Ackerson, R. Neervannan, and G. Neubig, "WikiAsp: A dataset for multi-domain aspect-based summarization," *Trans. Assoc. Comput. Linguistics*, vol. 9, pp. 211–225, Mar. 2021.
[38] K. M. Hermann, T. Kocisky, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, and P. Blunsom, "Teaching machines to read and comprehend," in *Proc. 28th Adv. Neural Inf. Process. Syst.*, vol. 28, Sep. 2015, pp. 1–9.
[39] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," *OpenAI Blog*, vol. 1, no. 8, p. 9, 2019.
[40] S. Bird and E. Loper, "NLTK: The natural language toolkit," in *Proc. ACL Interact. Poster Demonstration Sessions*, 2004, pp. 214–217.
[41] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, vol. 1, Jun. 2019, pp. 4171–4186.
[42] C.-Y. Lin, "Rouge: A package for automatic evaluation of summaries," in *Proc. Workshop Text Summarization Branches Out*, 2004, pp. 74–81.

KYOHOON JIN received the B.S. degree in applied statistics from Gachon University, Seongnam, South Korea, in 2019, and the M.S. degree from the Graduate School of Advanced Imaging Science, Multimedia and Film, Chung-Ang University, where he is currently pursuing the Ph.D. degree. His research interests include deep learning and natural language processing.

SEJIN PAIK received the B.A. degree in international studies from Emory University, the M.A. degree in computational journalism from Stanford University, and the Ph.D. degree in communication, specializing in AI in journalism, political psychology, human-centered AI, and intelligent social systems from Boston University. She is a Postdoctoral Fellow within the Georgetown's McCourt School of Public Policy, Massive Data Institute (MDI). Her current research examines how individuals interact with AI-generated content and develops various AI models for a safe and trustworthy information ecosystem.

JUNGMIN YUN received the B.S. degree in library and information science and the M.S. degree in artificial intelligence from Chung-Ang University, where she is currently pursuing the Ph.D. degree with the Graduate School of Artificial Intelligence. Her research interests include deep learning and natural language processing.

SOOJIN JANG received the B.S. degree in information and communication engineering from Sun Moon University and the M.S. degree from the Graduate School of Advanced Imaging Science, Multimedia and Film, Chung-Ang University, where she is currently pursuing the Ph.D. degree. Her research interests include deep learning and computer vision.

YOUNGBIN KIM (Member, IEEE) received the B.S. and M.S. degrees in computer science and the Ph.D. degree in visual information processing from Korea University, in 2010, 2012, and 2017, respectively. From August 2017 to February 2018, he was a Principal Research Engineer with Linewalks. He is currently an Assistant Professor with the Graduate School of Advanced Imaging Science, Multimedia and Film, Chung-Ang University. His current research interests include data science and deep learning.