Master's Thesis of Engineering

A Study on Aspect-oriented Summarization using Transformer

트랜스포머를 이용한 키워드 반영 추출 요약에 대한 연구

February 2023

Interdisciplinary Program in Bioengineering
The Graduate School
Seoul National University

Jiye Son

A Study on Aspect-oriented Summarization using Transformer

Academic adviser Jinwook Choi

Submitting a master's thesis of Engineering

February 2023

Interdisciplinary Program in Bioengineering
The Graduate School
Seoul National University

Jiye Son

Confirming the master's thesis written by
Jiye Son
February 2023

Chair	Hyung Jin Yoon	(Seal)
Vice Chair	Jinwook Choi	(Seal)
Examiner	Jae Sung Lee	(Seal)

Abstract

A Study on Aspect-oriented Summarization using Transformer

Jiye Son

Interdisciplinary Program in Bioengineering

The Graduate School

Seoul National University

Text summarization is well-known as a representative task in natural language processing. Text summarization methods generate brief written summaries of documents such as journal articles. In recent years, the performance of text summarization methods has improved significantly with the development of pretrained language models based on Transformer architectures such as BERT and GPT-3.

Recently, the development of language models designed to generate controllable output based on user preferences has attracted considerable attention as a topic of active research. Controllable

summarization methods such as query-focused or aspect-oriented summarization techniques have also emerged as promising approaches. In particular, aspect-oriented summarization generates a summary in terms of specific aspects provided as user input.

In this study, we propose a method to improve the performance of an aspect-oriented extractive summarization model presented in a previous work. The proposed method helps the model to generate aspect-oriented summaries by reflecting the relevance between sentence features and keyword features representing the aspect. To evaluate the performance of the proposed method, we constructed a new dataset consisting of articles on COVID-19 labeled in terms of two aspects: “Trend” and “Action.” The results showed that our proposed method outperformed a baseline model on the new dataset.

The proposed method exhibited higher performance than the baseline by roughly 3.6–4.3% in terms of “Trend,” and showed a relatively low impact with an improvement of less than 1% in terms of “Action.” However, in both aspects, we observed that even incorrect sentences included in a generated summary tended to be related to the defined aspect. Thus, we demonstrate that the proposed method generated more aspect-oriented summaries with content relevant to the defined aspect.

Keyword : Natural Language Processing, BERT, BERTSum, Transformer, Aspect−oriented Summarization, Text Summarization, Keyword Centered Summary

Student Number : 2021−22509

CONTENTS

List of Tables

List of Figures

Chapter 1. Introduction

1.1. Background

Text summarization is a representative task in natural language processing. It aims to generate brief written summaries of text documents such as articles. These methods can provide simple but useful content to readers in the forms of, for example, news spotlights on an article or an abstract in a paper. The development of deep learning−based language models has improved the quality of automatic summarization considerably. Recently, summarization models using deep learning have been applied in a wide variety of industries, which highlights the usefulness of these systems. Representing the important information in written documents concisely is critical to increase efficiency in various tasks, especially those involving lengthy documents. For example, this approach can be used to distill key content when performing other natural language retrieval tasks on the information in long documents, such as searching and question answering. By shortening the length of the text to be considered at a given time while maintaining the core information, the limitation of the input length of the system can be mitigated to some extent. Through this, summarization tasks can contribute to improving the performance and efficiency of other natural language processing methods.

Although text summarization has been extensively investigated in the literature, further research is necessary to develop methods to generate better summaries containing rich information, which would be applicable to a wide variety of fields. In particular, as deep learning models controlled by user preferences have become a focus of active work in several fields, considerable room for improvement remains in terms of the performance of existing methods.

1.2. Task Description

1.2.1. Text Summarization

Text summarization is the task of generating a summary of a written document. These techniques include both (1) extractive and (2) abstractive summarization methods [18].

Extractive summarization models generate a summary comprising sentences extracted from the original document. This requires the ability to understand text and select important sentences. These models typically comprise an encoder for understanding text and a classifier for selecting sentences.

By contrast, abstractive summarization models create summaries with newly generated sentences by paraphrasing sentences in the original document and generating new words. Because this requires not only understanding context but also generating text, these models consist of an encoder for understanding text and a decoder for generating new content.

In this study, we dealt with extractive summarization that generate a summary by extracting key sentences from a document, especially news articles about COVID-19.

1.2.2. Aspect oriented summarization

In general summarization, the output contains crucial information considered relevant to the entire text. By contrast, many studies have been conducted recently on generating different summaries for a given document depending on the purpose of the user. These methods have been referred to as “controllable,” “query-focused,” or “aspect-oriented” summarization according to the methods of control used or the specific design purposes considered.

Given that the goal of this task is to generate a summary oriented to a user-defined aspect indicated with keywords, we referred to such tasks as being aspect-oriented in this work, following [1].

In this study, we aimed to generate aspect-oriented summaries of articles. A given article may have diverse aspects. For example, we indicate two aspects of an article on COVID-19 as shown in Table 1.2. Some sentences highlight the trends of the novel coronavirus pandemic such as increasing numbers of infections and deaths, whereas others include information on active measures against COVID-19 such as restrictions or vaccines. As an example, the sentences that were considered most relevant to the action against COVID-19 and important would be extracted from the contents marked in blue in Table 1.2. to generate a summary about the aspect “Action” via extractive summarization.

Table 1.2. An example of article about COVID−19. The sentences about the trend of COVID−19 are marked in orange. The sentences about the action against COVID−19 are marked in blue.

Articles : South Sudan Becomes the 51st Country in Africa With COVID-19 as It Confirms Its First Case
1. JUBA, South Sudan — Officials in South Sudan say the country has its first case of COVID-19, making it the 51st of Africa's 54 countries to have the disease.
2. First Vice President Riek Machar and the U.N. mission in South Sudan confirmed the case of a U.N. worker who arrived in the country from Netherlands on Feb. 28.
3. The patient, a 29-year-old woman, first showed signs of the disease on April 2 and is recovering, said officials.
4. South Sudan, with 11 million people, currently has four ventilators and wants to increase that number, said Machar, who emphasized that people should stay three to six feet apart from others.
5. "The only vaccine is social distancing," said Machar.
6. The patient is under quarantine at U.N. premises and health workers are tracing the people who had been in contact with her, said David Shearer, head of the U.N. operations in South Sudan.
7. He said he hoped the measures would contain the case.
8. To prevent the spread of the virus in South Sudan, President Salva Kiir last week imposed a curfew from 8:00 p.m. to 6:00 a.m. for six weeks and closed borders, airports, schools, churches and mosques.
9. With the disease in South Sudan, now just three countries in Africa have not reported any cases of COVID-19: the tiny mountain kingdom of Lesotho in southern Africa, and the island nations of Comoros and Sao Tome and Principe.
10. Ethiopia on Sunday reported its first death from the virus and announced five more cases bringing its total to 43, most of them imported by travelers.
11. Ethiopia's Prime Minister Abiy Ahmed held discussions Sunday with opposition party leaders on measures to combat the virus.
12. A number of Ethiopia's regional states have implemented bans on movement of people and vehicles, but not yet in the capital Addis Ababa.

The goal of this study is to advance the existed model for aspect−oriented summarization which will be described in Section 2.3 for better aspect−oriented summary of COVID−19 articles. As delineated in Section 2.3., the existing method provides the information about user preference in a form of five keywords. Our key intuition is that reflecting the relevance to the keywords on each sentence would be helpful to generate more aspect−oriented summary. This study proposed some methods to explore this hypothesis.

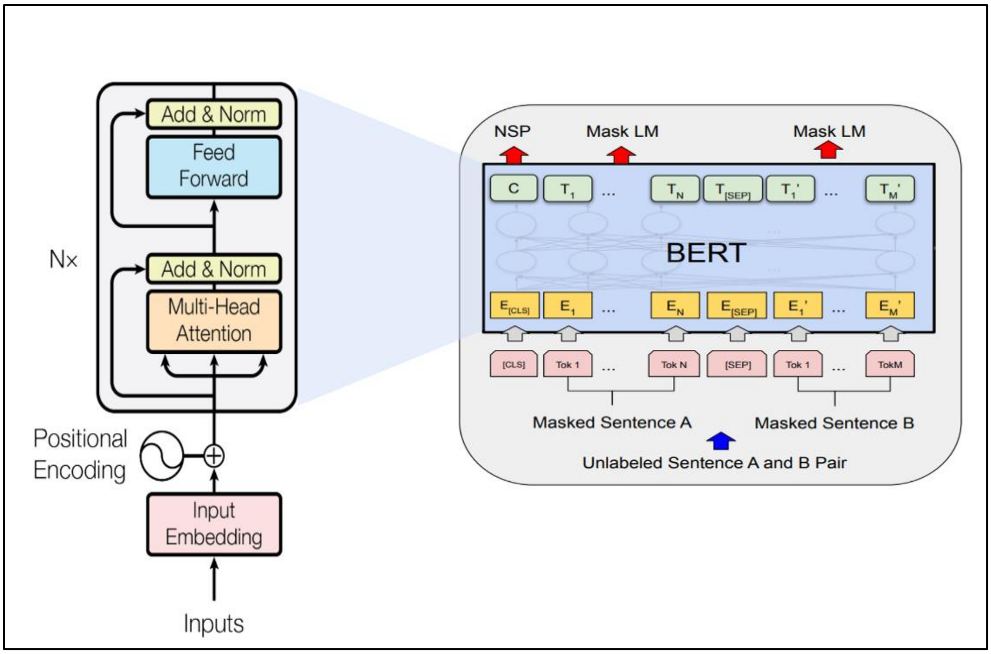
Chapter 2. Related Works

2.1. Extractive Summarization

Automatic summarization is the task of generating a brief summary from the lengthy document. The generated summary should contain the core contents of original text. Generally, summarization task is divided into two categories : extractive summarization and abstractive summarization. Extractive summarization is the task in this study. While the abstractive summarization model generates the summary by paraphrasing and generating, the extractive summarization model generates it by extracting important sentences from original document. As many related studies have conducted, the variety of models have appeared. TextRank is one of the representative algorithms of extractive summarization, which is the graph based ranking model[2]. As the development of deep learning, many researchers adopted the deep learning based model for extractive summarization. Nallapti et al.[3] suggested the model SummaRuNNer which is a Recurrent Neural Network(RNN) based sequence model for extractive summarization. With the advent of Transformer[4] using attention mechanism, transformer-based models pretrained on a massive dataset have begun to emerge and shown the significant performance in natural language processing[5-

7]. They were also developed to perform summarization task[8,9,16]. A representative model is BERTSum[9] whose architecture is BERT and finetuned to generate summaries. BERTSum includes BERTSumAbs, which performs the abstractive summarization, and BERTSumExt[10], which performs the extractive summarization. BERTSumExt consists of the encoder layer and classifier layer. The encoder understands and encodes the input to proper representations. The classifier serves to predict whether each sentence is included in a summary. Figure 2.1. shows the architecture of BERT and BERTSumExt. It has the modified segment embeddings to distinguish between two or more sentences. Furthermore, since the summarization task requires each representation of sentences, the [CLS] token was inserted into the front of all sentences. The embedding vector of [CLS] represents the following sentence, containing the feature of it. They are provided to the encoder layer Transformer. The output of encoder layer passes through the sigmoid classifier, which is the last layer. The obtained probability from the last layer indicates whether the corresponding sentence is included in the summary.

<BERT>



<BERTSumExt>

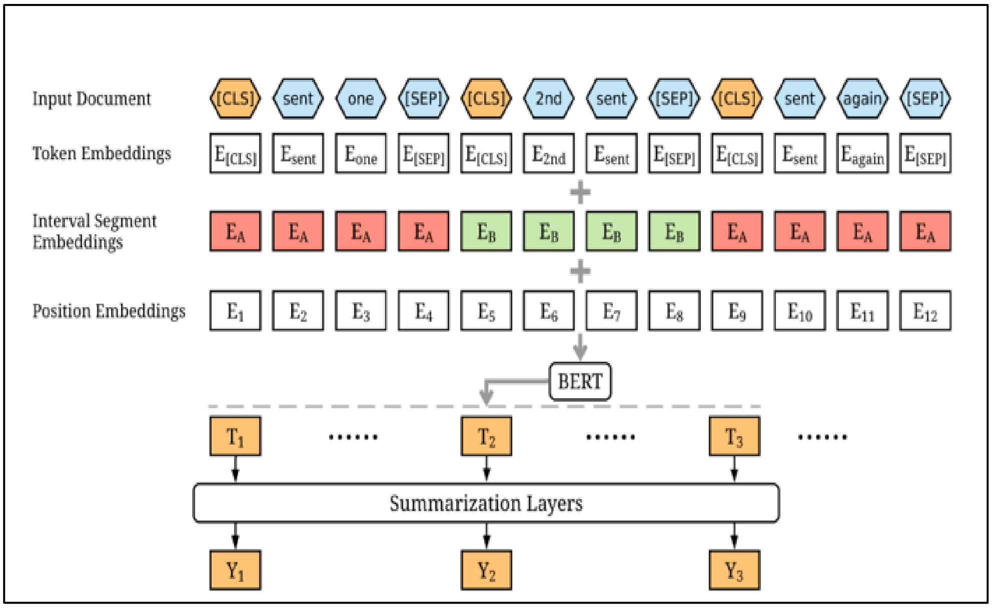


Figure 2.1. The architecture of BERT (top)[4][5] and BERTSumExt (bottom)[10].

2.2. Aspect Oriented Summarization

In recent years, the interest of many researches in deep learning has changed to produce appropriate output controlled by user's intent rather than generic output. This trend is also shown in the field of summarization. Fen et al.[11] proposed controllable summarization with the goal of controlling the shape of the summary such as the length, style or the entities. In some researches, the contents of summary can be adjusted to user preference. The research by Frermann and Alexandre[12] focused on generate a abstractive summary based on aspect using Pointer-generator networks. They incorporated different attention mechanism for reflecting the aspect into the encoder and decoder. He et al. proposed CTRSum[13], a framework for abstractive summarization where the users control the property of summary by providing the control tokens such as keywords or a description. There are not enough datasets for this task, they modified the existing dataset to train the model as in the previous works. Recently, Maddela et al.[14] introduced the dataset for entity centric summarization providing the summary centering around the entities extracted from the original document. For extractive aspect-oriented summarization, Ahuja et al.[1] proposed a dataset ASPECTNEWS and AOSUMM which is BERTSumExt model trained to generate aspect-oriented summary.

2.3. AOSUMM

In [1], they proposed the AOSUMM model with an architecture of BERTSumExt[10] which was finetuned on modified CNN/Daily Mail dataset for aspect-oriented summarization. Since CNN/Daily Mail dataset is a dataset for general summarization, they modified it for aspect-oriented summarization. Whereas the original CNN/Daily Mail dataset has the pair of document *D* and associated summary *S*, the modified dataset consists of triples *(D, K, S')* where *D* is a document, *K* is a set of keywords and *S'* is modified summary. The set of keywords are extracted from the text according to TF-IDF ranking system. For training, the extractive summary oriented to the keywords is required as the gold summary. To create the extractive oracle summary with respect to keywords, they developed the traditional method which finds the sentence which maximizes the ROUGE-2 between sentence and reference. The developed method is as follows:

$$final\ gold\ summary = \mathrm{argmax}_E\ BERTScore(E, S + nK) \qquad (1)$$

where E is a binary sequence of $\{\mathrm{E}_1, \mathrm{E}_2, .., \mathrm{E}_m\}$ indicating whether i^{th} sentence is included in summary, S is an original summary, K is keywords and n is hyperparameter.

The final training dataset comprise (1) the modified triples (D, K, S') and (2) the unmodified pairs (D, S) without keywords.

The model of which the architecture consists of BERT and Transformer encoder was finetuned in the same procedure of BERTSumExt model except that the sequence of keywords is prepended to the input text, which is referred AOSUMM. The architecture and input format of AOSUMM is described in Figure 2.3.

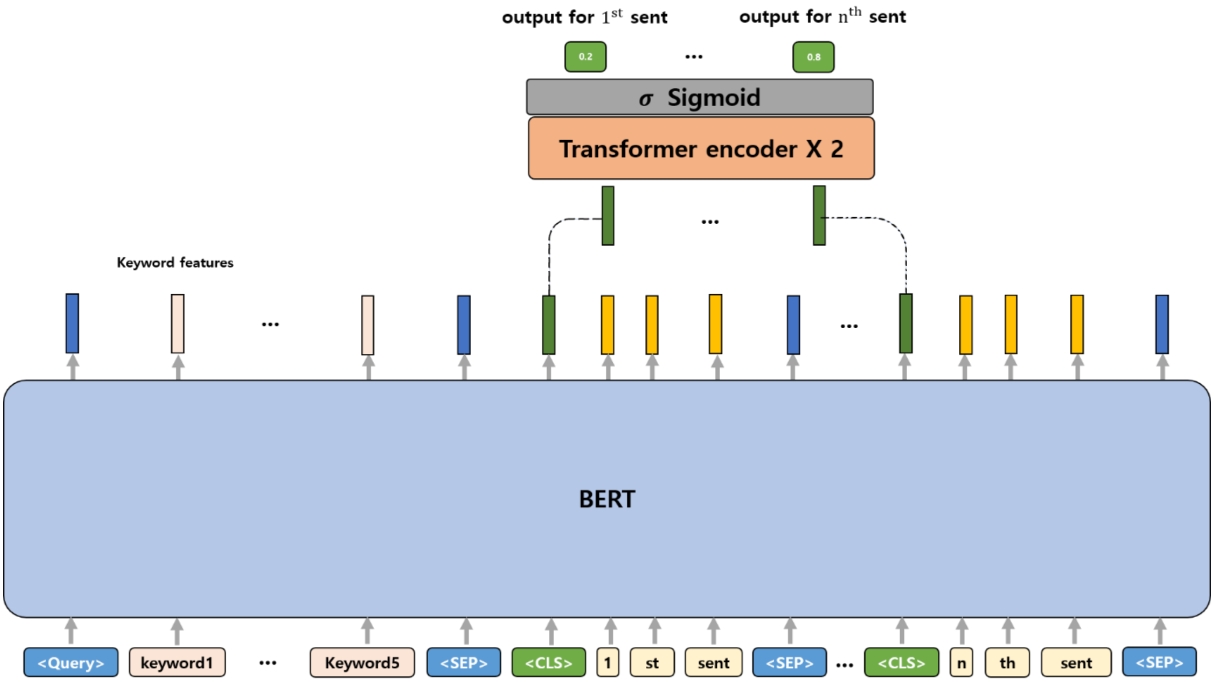


Figure 2.3. The architecture and input format of AOSUMM

Chapter 3. Materials and Method

3.1. Dataset for Training

The modified CNN/Daily Mail dataset was used to finetune the model with an architecture of BERTSumExt, following the previous research [1]. CNN/Daily Mail dataset[15] widely used for text summarization consists of news articles in CNN and Daily Mail websites and the corresponding summaries. Since CNN/Daily Mail dataset is not for aspect-oriented summarization, AOSUMM was finetuned on modified CNN/Daily Mail. The modified CNN/Daily Mail contains the triples including an article, keywords and an oracle extractive summary related to keywords. Following the procedure guided by [1] described in Section 2.3., we obtained the modified CNN/Daily Mail dataset. We split it into 287,227 articles for training and 26,736 articles for validation.

3.2. Dataset for Evaluation

For evaluation, we constructed a new dataset for aspect-oriented extractive summarization. The target domain is "COVID-19" of which articles include explicit several aspects in a single text and can be collected easily as it is a public concern. By considering the possible size of the dataset for human annotation and referring to the size of the dataset built in previous study[1] which was 100, 120 articles about COVID−19 was collected from free coronavirus news dataset published by AYLIEN[17].

3.2.1. Aspect Definition

In order to define aspects in this domain, all 120 articles were reviewed considering the useful information that readers would like to obtain from COVID−19 related articles. In each of the 120 articles of COVID−19, two aspects of "Trend" and "Action" were revealed in common. The aspect of "Trend" refers to contents related to the trend of the spread of the COVID−19, such as the infections or deaths. Meanwhile, the aspect of "Action" refers to contents related to the action against COVID−19, such as vaccine and the measures. The description of two aspects are shown in Table 3.2.1.

Table 3.2.1. Description and example topics of two aspects in the target domain "COVID−19".

Domain	Aspect	Description	Example of topics
Covid-19	Trend	It covers all of contents about the trend of the spread of COVID 19	infections or deaths, the speed of spread
	Action	It covers all of contents about the action against COVID-19 to prevent the spread of it and to treat	policy, vaccine, restrictions (e.g. lockdown or social distance)

3.2.2. Annotation

To generate the gold summary for 120 articles, 4 annotators were recruited. Four annotators were divided into two groups. In other words, each article was annotated by two annotators.

For each sentence, they gave a score considering how relative to the defined aspect and how essential in an article. The range of score is from 0 to 3. The score '0' indicates that the sentence is not related to the aspect. In the case of related sentence, they allocated score from 1 to 3, depending on the degree of importance and relation to aspect. After adding up the score given by two annotators, the top 3 sentences with the highest score in an article were selected as the final gold summary of that article. As shown in Table 3.2.2., for all 120 articles, the corresponding summary for each aspect was generated. Further details of the guidelines and example of annotation are attached to the Appendix A.

Table 3.2.2. Example of generated data. The number prepended to each sentence indicates the sentence number in an original article.

Articles : South Sudan Becomes the 51st Country in Africa With COVID-19 as It Confirms Its First Case
1. JUBA, South Sudan — Officials in South Sudan say the country has its first case of COVID-19, making it the 51st of Africa's 54 countries to have the disease. 2. First Vice President Riek Machar and the U.N. mission in South Sudan confirmed the case of a U.N. worker who arrived in the country from Netherlands on Feb. 28. 3. The patient, a 29-year-old woman, first showed signs of the disease on April 2 and is recovering, said officials. 4. South Sudan, with 11 million people, currently has four ventilators and wants to increase that number, said Machar, who emphasized that people should stay three to six feet apart from others. 5. "The only vaccine is social distancing," said Machar. 6. The patient is under quarantine at U.N. premises and health workers are tracing the people who had been in contact with her, said David Shearer, head of the U.N. operations in South Sudan. 7. He said he hoped the measures would contain the case. 8. To prevent the spread of the virus in South Sudan, President Salva Kiir last week imposed a curfew from 8:00 p.m. to 6:00 a.m. for six weeks and closed borders, airports, schools, churches and mosques. 9. With the disease in South Sudan, now just three countries in Africa have not reported any cases of COVID-19: the tiny mountain kingdom of Lesotho in southern Africa, and the island nations of Comoros and Sao Tome and Principe. 10. Ethiopia on Sunday reported its first death from the virus and announced five more cases bringing its total to 43, most of them imported by travelers. 11. Ethiopia's Prime Minister Abiy Ahmed held discussions Sunday with opposition party leaders on measures to combat the virus. 12. A number of Ethiopia's regional states have implemented bans on movement of people and vehicles, but not yet in the capital Addis Ababa.
Gold Summary for aspect "Trend"
1. JUBA, South Sudan — Officials in South Sudan say the country has its first case of COVID-19, making it the 51st of Africa's 54 countries to have the disease. 9. With the disease in South Sudan, now just three countries in Africa have not reported any cases of COVID-19: the tiny mountain kingdom of Lesotho in southern Africa, and the island nations of Comoros and Sao Tome and Principe. 10. Ethiopia on Sunday reported its first death from the virus and announced five more cases bringing its total to 43, most of them imported by travelers.
Gold Summary for aspect "Action"
4. South Sudan, with 11 million people, currently has four ventilators and wants to increase that number, said Machar, who emphasized that people should stay three to six feet apart from others. 8. To prevent the spread of the virus in South Sudan, President Salva Kiir last week imposed a curfew from 8:00 p.m. to 6:00 a.m. for six weeks and closed borders, airports, schools, churches and mosques. 12. A number of Ethiopia's regional states have implemented bans on movement of people and vehicles, but not yet in the capital Addis Ababa.

3.3. Evaluation Metric

For evaluation metrics, ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [19] is widely used in natural language task such as summarization and machine translation. ROUGE is a set of metrics. The following list is ROUGE score used in this study with corresponding description :

- Rouge-N : The ratio of overlapped words in N-gram between predicted and gold summary
- Rouge-L : The ratio of the longest overlapped sequence of words between predicted and gold summary

In ROUGE, precision and recall are calculated as follows :

$$Precision = \frac{Number\ of\ the\ overlapped\ words}{Total\ number\ of\ words\ in\ predicted\ summary} \tag{2}$$

$$Recall = \frac{Number\ of\ the\ overlapped\ words}{Total\ number\ of\ words\ in\ gold\ summary} \tag{3}$$

Using the calculated precision and recall, F1 is calculated as follows :

$$F1 = \frac{2*Precision*Recall}{Precision+Recall} \tag{4}$$

In extractive summarization, by adopting ROUGE as evaluation metric instead of simple precision or recall as classification task, the difference between wrong predictions can be calculated. Among incorrectly selected sentences, the sentence made up of similar words to gold summary can get higher score than those that do not.

Suppose that there is a problem to extract sentences about "sofa" as shown in Figure 3.3. Each candidate has one correct sentence and one wrong sentence. Using evaluation metrics for sentence−level, the two candidates have same Recall score of 0.5. However, while a wrong sentence in candidate 2 is only related to a dog, a wrong sentence in candidate 1 is a little closer to answer sentence as informing that a dog sat on the sofa. Using ROUGE−1 Recall score as described in equation (3), candidate 1 has 0.77 ROUGE−1 Recall score while candidate 2 has got 0.55 ROUGE−1 Recall score. As this example, ROUGE score can evaluate the difference among wrong predictions. For this reason, many extractive summarization researches adopted ROUGE score for evaluation[1][10][14].

Original Sentence
1. The sofa was old. 2. And the sofa was blue. 3. A dog sat on the sofa. 4. It looked sleepy.
Answer
1. The sofa was old. 2. And the sofa was blue.
Candidatae 1
2. And the sofa was blue. 3. A dog sat on the sofa.
Candidatae 1
2. And the sofa was blue. 4. It looked sleepy.

$$ROUGE - 1\ Recall\ for\ candidate\ 1 = \frac{7}{9} = 0.77$$

$$ROUGE - 1\ Recall\ for\ candidate\ 2 = \frac{5}{9} = 0.55$$

Figure 3.3. Example of evaluation using ROUGE.

3.4. Keyword Selection

Following the manner in the baseline AOSUMM model, the information of the aspect was fed into the model in the form of keywords. The keywords should be the representative words of the defined aspect. In general, keywords in a corpus are selected by considering the frequency and the uniqueness of word. TF-IDF is a well-known method based on this principle. Thus, we extracted keywords from the representative sentences of each aspect based on their frequency and relevance to aspect since this study required keywords related to defined aspect.

In the constructed dataset in section 3.2, the score of a sentence indicates the relation of the aspect and the sentence. Firstly, we collected the sentence whose score is 6, which is a maximum score, after summing up the scores of two annotators. It implies that those are the typical sentences having the defined aspect. The words in the set of representative sentences are the candidate for keywords. And then they were listed in order of frequency of appearance, and words unrelated to the defined aspect such as an article or a neutral word were excluded from the keyword candidate. The five most frequent word in the candidates were selected as a keyword. The list of five keywords for each aspect is shown in Table 3.4.

Table 3.4. The list of keywords for each aspect

Aspect	Keywords
Trend	outbreak, toll, death, infection, case
Action	vaccine, lockdown, restriction, measures, government

3.5. Method

The purpose of this study is to find the method to improve the performance for aspect-oriented summarization based on AOSUMM model. AOSUMM has the same architecture as BERTSumExt as shown in Figure 2.3. The keywords were provided into the model once and there is no information about keywords anymore providing to Transformer encoder which predicts the importance of each sentence.

In this study, we explored the effect of reflecting the relevance to the keywords on each sentence before entering the Transformer encoder, which is a step to predict the importance of each sentence. The Figure 3.5. describes the overview of the proposed method.

Prior to the experiments, following two steps were considered : (1) How to extract the keywords features, and (2) How to calculated the relevance between keywords features and sentence feature. In the following section, how to deal with these steps were explained.

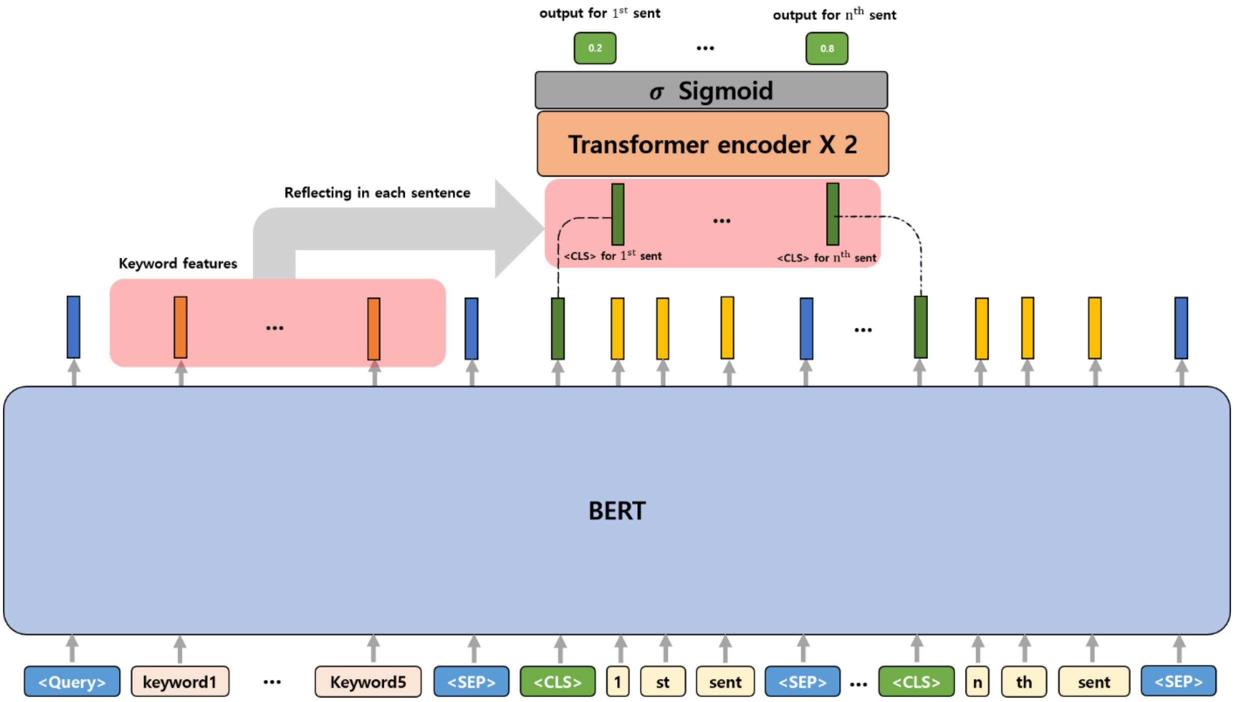


Figure 3.5. The overview of the proposed method

3.5.1. Extraction of keywords feature

The model is provided five keywords with respect to defined aspect. As shown in Figure 3.5., BERT outputs a representation of each token. For each sentence, each following [CLS] token before a sentence is used as the feature. Similarly, to obtain each feature of keywords, the special token [Q_SEP] was inserted before each keyword when providing to BERT. Like [CLS] token, the special token [Q_SEP] serve as a separator and representative. The output representation of [Q_SEP] is a feature of the following keyword.

We designed two ways how to utilize these keywords features.

a. Single feature covering all keywords

In the first way, a single feature which covers all of five keywords was used to calculate the relevance with each sentence feature. This representation of keyword feature was obtained by average pooling of five keyword features, that is, five [Q_SEP] tokens features. Average pooling is an operation down−sampled the feature map by using the average values of each feature. The operation of average pooling is shown in Figure 3.5.1. The relevance score in the first way implies how much each sentence relates to all keywords. In order to get a higher score, the sentence must be related to the entire five keywords.

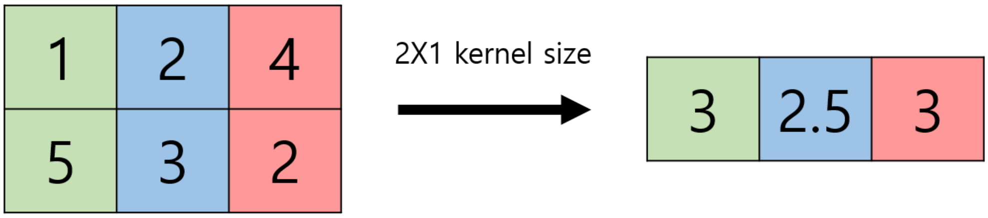


Figure 3.5.1. Example of Average Pooling operation with 2×1 kernel

b. Single feature of each keyword

However, the sentence that is strongly related to only one of keywords also can be included in the summary if it is a core statement in an article. Based on this idea, in the second way, the relevance score sc_i between keyword i and a sentence is calculated for all five keywords, respectively. Then, the maximum value of them becomes the final relevance score, **rs**. That is,

$$\mathbf{rs} = \max(\mathbf{rs_1}, \mathbf{rs_2}, \mathbf{rs_3}, \mathbf{rs_4}, \mathbf{rs_5}) \quad (5)$$

where $\mathbf{rs_j}$ is the relevance score between $\mathrm{j^{th}}$ keyword and a sentence.

3.5.2. Relevance Score

In order to reflect the relevance between keyword feature and sentence, the relevance score was multiplied to each sentence vector. After multiplying the relevance score, the transformed sentence vector has the feature of the relevance with keywords. To calculate the relevance score between keyword feature and sentence feature, we utilized cosine similarity.

a. Cosine similarity

In general, the similarity between two vectors is calculated by cosine similarity. The cosine vale is in the range of −1 and 1. The small angle between two vectors means that two vectors have similar directions, and the cosine similarity has a high value close to 1. Conversely, if the angle between two vectors is large, the similarity becomes low. The relevance score $\mathbf{rs}$ is as follows:

$$\mathbf{rs} = \text{Cosine similarity}(\boldsymbol{k}, \boldsymbol{s}) = cos(\theta) \tag{6}$$

where $\boldsymbol{k}$ is keyword feature, $\boldsymbol{s}$ is sentence feature, and θ is the angle between $\boldsymbol{k}$ and $\boldsymbol{s}$.

b. Modified Cosine similarity

Cosine similarity considers only the angle between two vectors. However, the length of vectors can be meaningful. To reflect the length ratio to the relevance score, we calculated the relevance score using the projected sentence vector. Firstly, we projected sentence vector $\boldsymbol{s}$ to keyword vector $\boldsymbol{k}$ as follows:

$$\boldsymbol{s}' = |\boldsymbol{s}| \times cos(\theta) \times \frac{\boldsymbol{k}}{|\boldsymbol{k}|} \tag{7}$$

where θ is the angle between sentence vector $\boldsymbol{s}$ and keyword vector $\boldsymbol{k}$, $|\boldsymbol{s}| \times cos(\theta)$ is the length of projected sentence feature, and $\frac{\boldsymbol{k}}{|\boldsymbol{k}|}$ is the unit vector of keyword vector.

Then, the projected sentence vector $\mathbf{s}'$ can be expressed as keyword vector $\boldsymbol{k}$ as follows :

$$\mathbf{s}' = \left(\frac{|s|}{|\boldsymbol{k}|} \times cos(\theta)\right) \times \boldsymbol{k} \tag{8}$$

$$\mathbf{rs} = \frac{|s|}{|\boldsymbol{k}|} \times cos(\theta) \tag{9}$$

The ratio of projected sentence vector $\boldsymbol{s}'$ and keyword vector $\boldsymbol{k}$ was used as the relevance score $\mathbf{rs}$. The higher $\mathbf{rs}$ indicates that for expressing the aspect of the sentence $\boldsymbol{s}$, the more keyword vectors are needed. In other words, the sentence $\boldsymbol{s}$ contains more contents related to the aspect.

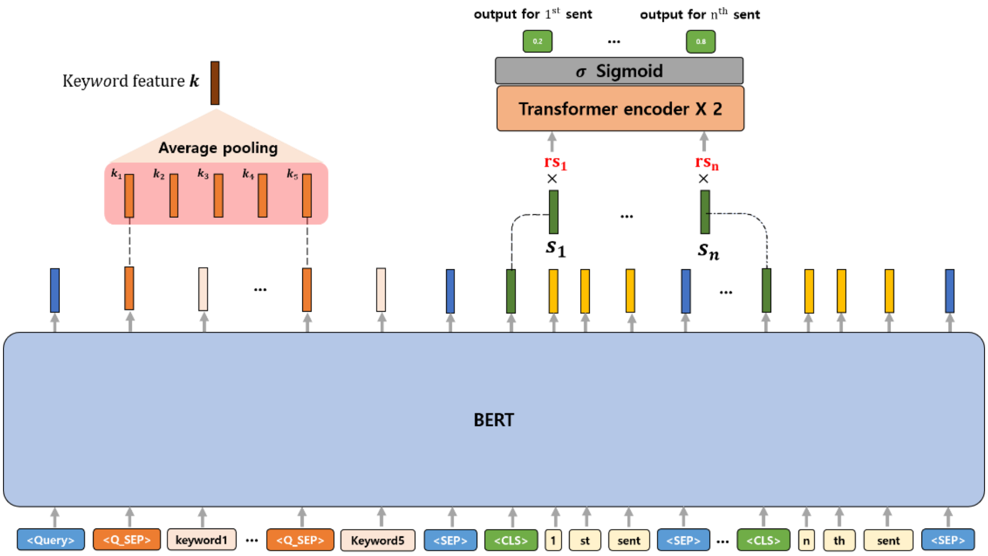
3.5.3. Proposed methods

According to the method to extract keyword features (Section 3.5.1.a and 3.5.1.b), we proposed two methods. Proposed method1 refers to the method utilizing a single keyword feature covering all five keywords. Proposed method2 refers to the method calculating relevance score for each keyword without any average pooling. We conducted additional experiments to find the proper relevance score (Section 3.5.2.a and 3.5.2.b) for each method. For both of two methods, the relevance score calculated by the ratio of projected sentence and keyword feature (Section 3.5.2.b) showed better performance on our COVID-19 dataset. The table of results is attached in Appendix B.

Hence, in this experiment, the relevance score calculated by Section 3.5.2.b. Final proposed methods are shown in Figure 3.5.3.

<Proposed method 1>

$$\mathbf{rs_i} = \frac{|\boldsymbol{s_i}|}{|\boldsymbol{k}|}\cos(\theta)$$



<Proposed method2>

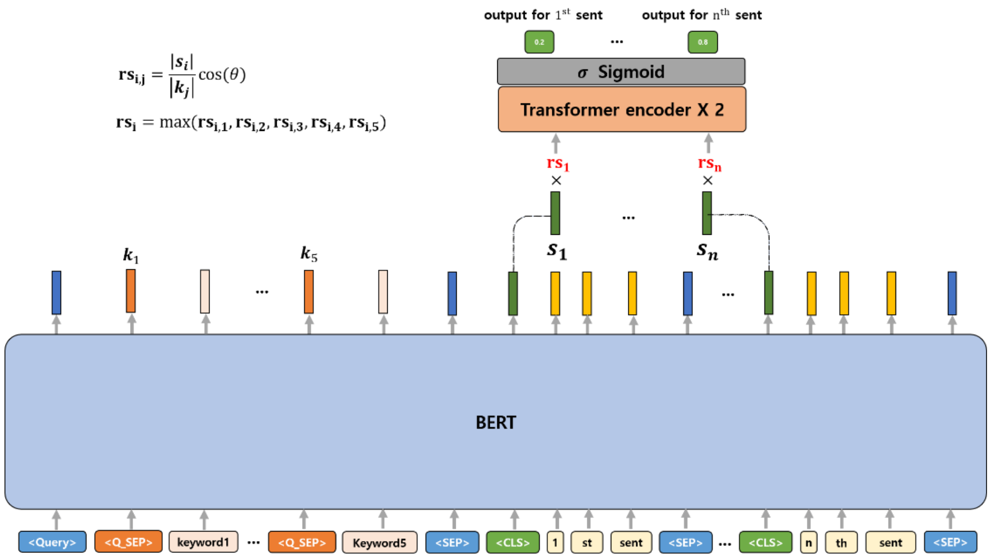


Figure 3.5.3. The architecture of the proposed method 1(top) and the proposed method 2(bottom). $\boldsymbol{s_i}$ refers to the representation vector of i$^{\text{th}}$ sentence. $\mathbf{rs_i}$ refers to the relevance score of i$^{\text{th}}$ sentence. $\mathbf{rs_{i,j}}$ refers to the relevance score between $\boldsymbol{s_i}$ and j$^{\text{th}}$ keyword feature.

Chapter 4. Results and Discussion

4.1. Experiment Settings

The model is train to reduce the difference between the gold summary and predicted summary by optimizing the weights of the model to reduce the binary classification entropy loss as follows:

$$\text{BCE Loss}(\hat{y}, y) = -(y \times \log \hat{y} + (1 - y) \times \log(1 - \hat{y}) \qquad (10)$$

where $\mathbf{y}$ is a target and $\hat{y}$ is an output when given five keywords and an article to the model.

Since the training dataset consists of aspect-oriented summary with keywords and generic summary without keywords as followed by [1], the relevance score for the dataset without keywords was 1. In other words, the sentence vector whose input has no keyword was passed to Transformer encoder without any operation for reflecting the aspect.

We finetuned BERTSumExt model using two GPUs (Geforce RTX GPU, 190 3090). The hyperparameters for training was followed by AOSUMM. All models in this study were finetuned using the Adam optimizer with the learning rate scheduler. The learning rate scheduler was used followed by [4]. The initial learning rate was $\mathbf{2 \times 10^{-3}}$. The batch size was 3,000 in the unit of tokens and the total training iteration was 50,000. All results in this paper are obtained using the parameters of the best iteration where the loss on validation dataset was the lowest during training.

4.2. Results

To verify the effect of proposed methods, the results of test on our COVID19 dataset are compared to the baseline model. The baseline model is AOSUMM. The proposed methods are applied to AOSUMM model when both of finetuning and testing. There are three experiment setup for comparison : (1) baseline AOSUMM, (2) finetuning with the proposed method1 and (3) finetuning with the proposed method2.

4.2.1. Automatic Evaluation

The results are shown in Table 4.2.1. It shows the result on each aspect : “Trend” and “Action”. ROUGE score was used for the evaluation and expressed as a percentage.

For aspect “Trend”, both of proposed methods improved the performance compared to the baseline. The baseline model showed 64.1 of ROUGE-1 F1, 55.1 of ROUGE-1 F1, and 62.8 of ROUGE-L F1. The proposed method1 showed slight improvements with 65.1 of ROUGE-1 F1, 56.1 of ROUGE-2 F1, and 63.6 of ROUGE-L F1. The proposed method2 showed the best performance, achieving the improvements of ROUGE-1 F1 by 3.6%, ROUGE-2 F1 by 4.3% and ROUGE-L F1 by 3.7%.

Even in the aspect “Action”, the proposed method 2 showed better performance than others, but not remarkable as valid. The baseline model showed 60.8 of ROUGE-1 F1, 51.4 of ROUGE-2 F1, and 59.3 of ROUGE-L F1. The proposed mehtod1 had no effect, showing the similar performance to the baseline with 60.8 of ROUGE-1 F1, 51.3 of ROUGE-2 F1 and 59.1 of ROUGE-L F1. The proposed mehtod2 increased ROUGE-1 F1 by 0.5%, ROUGE-2 F1 by 0.8% and ROUGE-L F1 by 0.4%. Overall, the improvement in the aspect “Action” was less than that in the aspect “Trend”.

Table 4.2.1. The results of test on COVID-19 dataset. In each aspect, the significant high score is represented as **bold** and slight improvement is underlined. R refers to Recall. P refers to Precision. F refers to F1. Score is displayed as %. The score was rounded to the second decimal place.

Aspect	Model	Rouge-1			Rouge-2			Rouge-L		
		R	P	F	R	P	F	R	P	F
Trend	AOSUMM (baseline)	65.9	63.6	64.1	56.5	54.7	55.1	64.5	62.4	62.8
	Proposed1 (Ours)	<u>67.3</u>	<u>64.2</u>	<u>65.1</u>	<u>57.8</u>	<u>55.4</u>	<u>56.1</u>	<u>65.7</u>	<u>62.8</u>	<u>63.6</u>
	Proposed2 (Ours)	**69.4**	**67.5**	**67.7**	**60.9**	**59.1**	**59.4**	**68.1**	**66.3**	**66.5**
Action	AOSUMM (baseline)	60.6	62.1	60.8	51.2	52.3	51.4	59.1	60.4	59.3
	Proposed1 (Ours)	60.6	61.9	60.8	51.1	52.2	51.3	58.9	60.1	59.1
	Proposed2 (Ours)	60.6	<u>62.9</u>	<u>61.3</u>	<u>51.7</u>	<u>53.3</u>	<u>52.2</u>	59.2	<u>61.2</u>	<u>59.7</u>

To probe the reason, we investigated COVID-19 dataset, a test dataset, in particular the annotation results that annotators gave scores for each sentence depending on the importance and the relevance to each aspect. Figure 4.3.1. shows the number of sentences per an article with a score of 1 or more for each aspect, meaning that the annotators thought it was related to defined aspect. The sentence with a score of 3 indicates that it is related to defined aspect and very important. The sentence with a score of 1 indicates that it is related to the aspects, but not important. The sentences with a score of 1 or 2 can be interpreted as a supplementary sentence for a core sentence whose score is 3. As shown in Figure 4.3.1., there are more sentences with a score of 1 and 2 in the aspect "Action" than in the aspect "Trend". Conversely, more sentences with a score of 3 were found in the aspect "Trend" than in the aspect "Action". It implies that the contents about the aspect "Action" include more supplementary sentences than the contents about the aspect "Trend".

Figure 4.3.2. describes the ratio of complementary sentence with a score of 1 to important sentence with a score of 3. In terms of the aspect "Trend", the ratio is 0.74, while in terms of the aspect "Action", the ratio is 1.07.

Figure 4.3.3. describes the ratio when including the sentences with a score of 2 as the complementary sentence. Similar in Figure

4.3.2., the ratio of complementary sentences to core sentences in the aspect “Action” is 2.56 which is higher than the ratio in the aspect “Trend”.

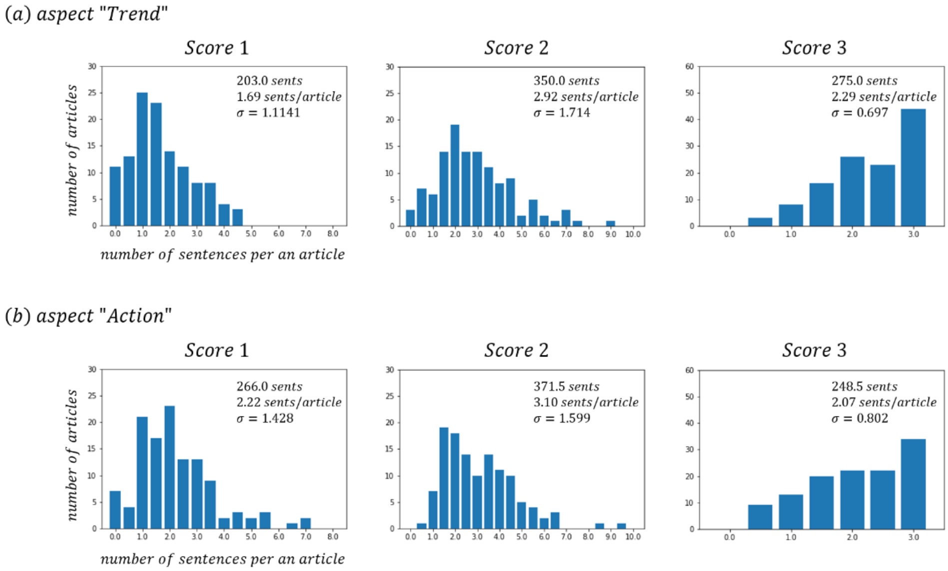


Figure 4.3.1. The number of sentences per each score. The x-axis represents the number of sentences per an article. The y-axis represents the number of articles. (a) Bar graph for the aspect “Trend”. (b) Bar graph for the aspect “Action”.

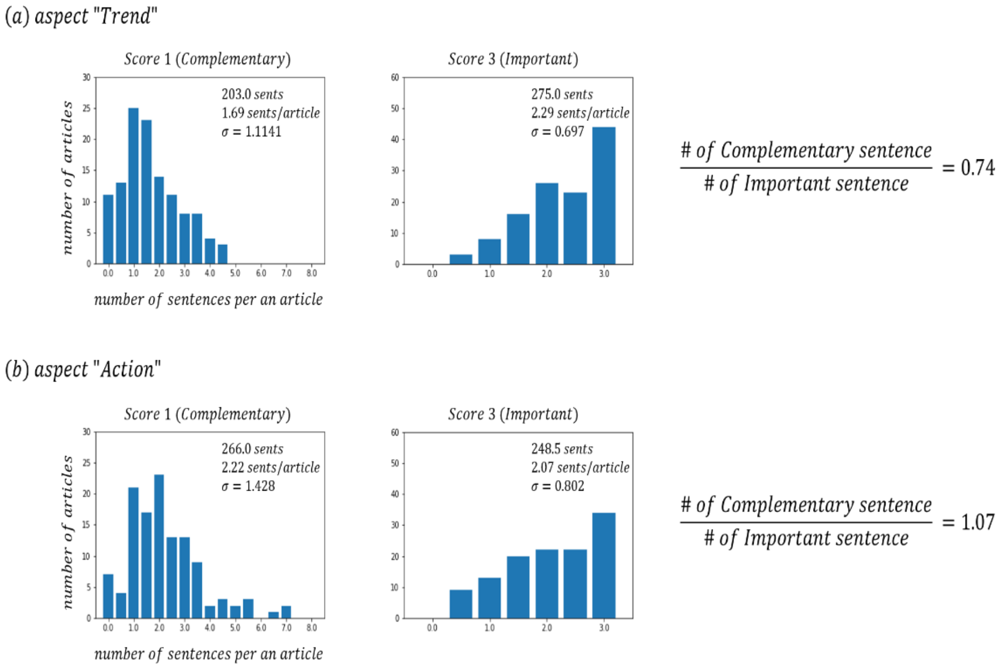


Figure 4.3.2. The ratio of complementary sentence with a score of 1 to important sentence with a score of 3.

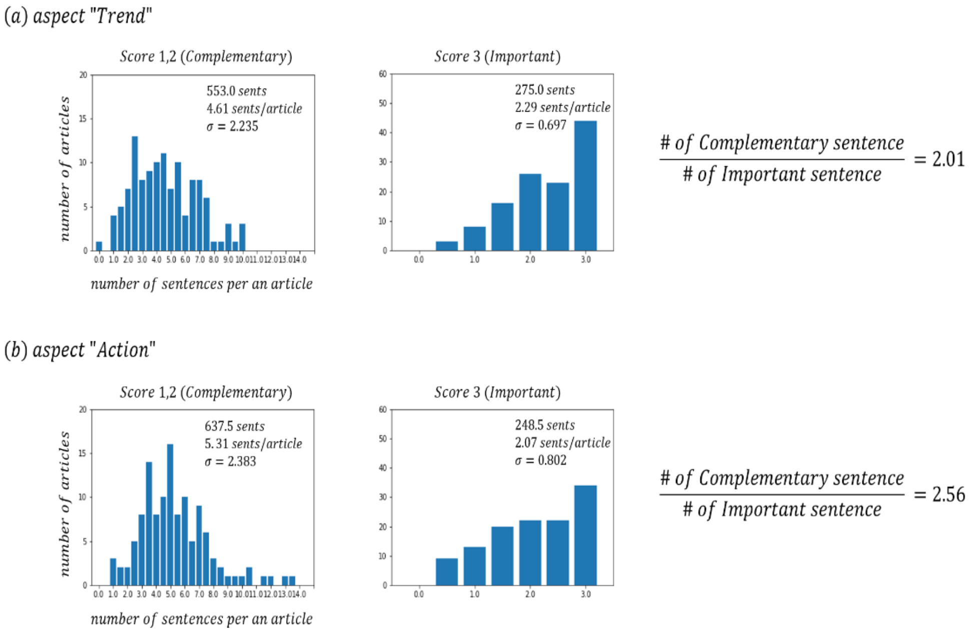


Figure 4.3.3. The ratio of complementary sentence with a score of 1 and 2 to important sentence with a score of 3.

4.2.2. Qualitative Evaluation

To probe the results in details, the predicted summaries of each model were shown in Table 4.2.2.1. and Table 4.2.2.2. The whole articles and more examples are appended to Appendix C.

In Table 4.2.2.1., the summaries generated by the proposed methods are closer to gold summary than a summary generated by the baseline. The baseline summary selected three wrong sentences. Even, two of them are unrelated to the aspect "Trend". As the ROUGE score showed, we observed that the overall quality of the summary generated by the proposed method 2 outperforms the output of the baseline.

Table 4.2.2.2. shows the output example for aspect "Action" test set. Compared to aspect "Trend", the overall quality of the generated summaries by all of models was lower. However, the summary of the proposed methods showed the better quality, closer to gold summary. Two of three sentences predicted by the proposed method were wrong, but one of wrong sentences is related to the aspect "Action". Given that the baseline model had two wrong sentences that are not related to the aspect "Action", it indicates that the proposed method helped the model to reflect the aspect of keywords to generate its summary.

Table 4.2.2.1. Example of generated summaries for aspect "Trend". Red font indicates that it is not included in gold summary. Green font indicates that it is not included in gold summary but related to aspect "Action".

Gold Summary
6. According to the Ministry of National Health Services, the nationwide tally of the coronavirus cases on Friday soared to 37,218 after 1,430 new cases were reported. 8. The total number of COVID-19 deaths in the country also reached 803 with 33 new fatalities reported during the last 24 hours. 9. So far, 344,450 tests have been conducted, including 13,700 during the last 24 hours.
Baseline Summary
1. Pakistan will resume domestic flights in a phased manner from Saturday after the government eased some of the restrictions imposed following the outbreak of the COVID-19 pandemic in the country that has claimed the lives of over 800 people. 2. The Pakistan government last week said that it would begin a phased lifting of the lockdown due to its effect on the economy and the workforce 3. Pakistan Civil Aviation Authority (PCAA) announced to resume the domestic flights operations from Saturday, The Express Tribune reported.
Proposed method1 Summary (Ours)
1. Pakistan will resume domestic flights in a phased manner from Saturday after the government eased some of the restrictions imposed following the outbreak of the COVID-19 pandemic in the country that has claimed the lives of over 800 people. 2. The Pakistan government last week said that it would begin a phased lifting of the lockdown due to its effect on the economy and the workforce. 6. According to the Ministry of National Health Services, the nationwide tally of the coronavirus cases on Friday soared to 37,218 after 1,430 new cases were reported.
Proposed method2 Summary (Ours)
5. It said that 68 flights will operate from Jinnah International Airport in Karachi, 32 from Allama Iqbal International Airport in Lahore, 32 from Islamabad International Airport, eight from Quetta International Airport and four from Bacha Khan International Airport in Peshawar. 6. According to the Ministry of National Health Services, the nationwide tally of the coronavirus cases on Friday soared to 37,218 after 1,430 new cases were reported. 8. The total number of COVID-19 deaths in the country also reached 803 with 33 new fatalities reported during the last 24 hours.

Table 4.2.2.2. Example of generated summaries for aspect “Action”. Red font indicates that it is not included in gold summary. Green font indicates that it is not included in gold summary but related to aspect “Action”. “Action”.

Gold Summary
6. Earlier wednesday , president trump said he was “in favor” of the states that had begun allowing businesses to re-open with restrictions and that social distancing guidelines would soon be “fading out .” 7. Experts have warned that abandoning restrictions too soon could lead to more surges in the virus ' spread and that the pandemic wo n't truly end until a vaccine is developed and widely distributed . 8. During that same oval office meeting , dr. anthony fauci said a recent drug trial showed remdesivir “can block” coronavirus .
Baseline Summary
1. The united states reached a new milestone in the coronavirus wednesday after it surpassed 60,000 deaths , according to data compiled by johns hopkins university . 2. The u.s. has recorded 60,207 deaths from coronavirus , and 1,030,487 cases , which account for about one third of the global total . 7. Experts have warned that abandoning restrictions too soon could lead to more surges in the virus ' spread and that the pandemic wo n't truly end until a vaccine is developed and widely distributed .
Proposed method1 Summary (Ours)
2. The u.s. has recorded 60,207 deaths from coronavirus , and 1,030,487 cases , which account for about one third of the global total . 5. Despite the rising numbers , more cities and states are easing restrictions . 6. Experts have warned that abandoning restrictions too soon could lead to more surges in the virus ' spread and that the pandemic wo n't truly end until a vaccine is developed and widely distributed .
Proposed method2 Summary (Ours)
2. The u.s. has recorded 60,207 deaths from coronavirus , and 1,030,487 cases , which account for about one third of the global total . 5. Despite the rising numbers , more cities and states are easing restrictions . 6. Experts have warned that abandoning restrictions too soon could lead to more surges in the virus ' spread and that the pandemic wo n't truly end until a vaccine is developed and widely distributed .

4.3. Discussion

We have three observations through the automatic evaluation and manual analysis.

Firstly, the proposed method 2 showed the better performance than the proposed method1, achieving remarkable improvement of performance, especially in the aspect “Trend”. The difference between the proposed method 1 and the proposed method 2 is how to extract keywords feature used when calculating the relevance score. While the proposed method 1 utilized a single representative feature covering all five keywords, the proposed method 2 used each keyword feature and calculated each relevance score. Since the sentences in a summary do not need to cover all of five keywords and it is enough to cover one of keywords, the strategy to reflect the highest relevance of one of keywords in the proposed method 2 was effective.

Secondly, the proposed methods were less effective to the aspect “Action” than the aspect “Trend”. To investigate the reason, we explored the annotation results of COVID-19 dataset as shown in Figure 4.3.1, Figure 4.3.2, and Figure 4.3.3. Figure 4.3.1. shows that the content of the aspect “Action” has more complementary sentences than the content of the aspect “Trend”, while it has less important sentences. We calculated the ratio of complementary sentences to important sentences in Figure 4.3.2. and Figure 4.3.3. In both of the results in Figure 4.3.2. and Figure 4.3.3., the ratio of complementary sentence to important sentence for the aspect Action is higher than that for the aspect “Trend”. Since complementary sentences are also related to the aspect, the relevance to the aspect is not enough factor to distinguish a key sentence from complementary sentences. For the aspect “Action” which has relatively more complementary sentences, it is likely that the proposed methods focused on finding aspect related sentences by reflecting the keywords features was not helpful to find key sentences. For a summary in this aspect, it is more necessary to find what the key sentence is than to pull out a keyword related sentence.

Lastly, while the effect of proposed methods was different depending on the aspect, the proposed methods encouraged the model to generate the better summary close to the defined aspects. Interestingly, in the Section 4.2.2, we observed that the selected sentences reflected the aspect of keywords, leading to real “aspect oriented” summarization for both aspects even though they are not exact answer.

However, the proposed methods still selected incorrect sentences that are unrelated to the aspects. There is still room for improvement of finding aspect related sentences. In addition, the result of this study is for our COVID-19 dataset consisting of 120 articles. To prove the effect of the proposed methods in depth, it is needed to collect more data for evaluation in the future work. Also, it will helpful for better aspect-oriented summarization to study on keyword selection and how to improve the ability to find the key sentences.

Chapter 5. Conclusion

This study proposed methods to improve the performance of the aspect-oriented summarization. For the experiment I created a new dataset consisting of COVID-19 articles and the corresponding summaries for defined aspects to evaluate the proposed models.

Focusing on the improvement of ability to find the related sentences with regard to the aspect defined as a form of keyword, the proposed methods shall reflect the information of the relevance between aspect keywords and each sentence to representation of each sentence. The results on COVID-19 dataset demonstrated that the proposed method is effective to help the model find the aspect-related sentences. Even though the effect of the proposed method was different depending on the aspect, the proposed method2 brought out the significant improvement by 3.6~4.3% for aspect of “Trend”. For aspect of “Action”, as showing the increase of less than 1%, the effect of the proposed methods was slight when compared to the results of aspect “Trend”. However, compared to baseline, the proposed methods showed a tendency to select the sentence whose attribute is closer to the defined aspect for both of aspects. It implied that the proposed methods helped the model to predict whether each sentence is related to the aspect keywords, however it was not enough to select the salient sentences.

It is expected that the additional study to improve the ability to find the core sentences and to select the proper keywords supplement the proposed methods in the future.

Recently, many researches have been conducted towards "user-controllable" learning in the field of natural language processing. The goal of those researches is the development of the advanced model to generate not plain output, but customized output considering the purpose of users. Also in the field of text summarization, the effort of researchers is ongoing to advance controllable summarization model. I hope that this study would be a helpful step forward to the advancement of aspect-oriented summarization.

Bibliography

[1] O. Ahuja, J. Xu, A. Gupta, K. Horecka, and G. Durrett, "ASPECTNEWS: Aspect−Oriented Summarization of News Documents", In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 2022, Volume 1: Long Papers, pages 6494–6506.

[2] R. Mihalcea and P. Tarau, "TextRank: Bringing Order into Text", In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, 2004, pages 404–411.

[3] R. Nallapati, F. Zhai, and B. Zhou, "Summarunner: A recurrent neural network based sequence model for extractive summarization of documents", Thirty−first AAAI conference on artificial intelligence, 2017.

[4] Vaswani, Ashish, et al., "Attention is all you need", Advances in neural information processing systems, 2017, 30.

[5] Devlin, Jacob, et al., "Bert: Pre−training of deep bidirectional transformers for language understanding", arXiv preprint arXiv:1810.04805, 2018.

[6] Brown, Tom, et al., "Language models are few−shot learners", Advances in neural information processing systems, 2020, 33: 1877−1901.

[7] Clark, Kevin, et al., "Electra: Pre−training text encoders as discriminators rather than generators", arXiv preprint arXiv:2003.10555, 2020.

[8] Kieuvongngam, V., Tan, B., and Niu, Y, "Automatic text summarization of covid−19 medical research articles using bert and gpt−2", arXiv preprint arXiv:2006.01997, 2020.

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

[10] Y. Liu, "Fine−tune BERT for extractive summarization", arXiv preprint arXiv:1903.10318, 2019.

[11] Fan, Angela, D. Grangier, and M. Auli, “Controllable abstractive summarization”, arXiv preprint arXiv:1711.05217, 2017.

[12] L. Frermann and A. Klementiev, “Inducing document structure for aspect-based summarization”, In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pages 6263–6273.

[13] He, Junxian, et al., “Ctrlsum: Towards generic controllable text summarization”, arXiv preprint arXiv:2012.04281, 2020.

[14] J. He, W. Krysci ´ nski, B. McCann, N. Rajani, and C. Xiong, “Ctrlsum: Towards generic controllable text summarization”, arXiv preprint arXiv:2012.04281, 2020.

[15] Hermann, K. M., Kocisky, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M., & Blunsom, P., “Teaching machines to read and comprehend”, Advances in neural information processing systems, 28, 2015.

[16] Zhong M, Liu P, Chen Y, et al., “Extractive summarization as text matching”, In Proceedings of the Annual Meeting of the Association for Computational Linguistics. Online, 6197–6208, 2020.

[17] https://aylien.com/blog/free-coronavirus-news-dataset

[18] Dong, Yue., "A survey on neural network-based summarization methods.", arXiv preprint arXiv:1804.04589, 2018.

[19] Lin, C.Y. “Rouge: A package for automatic evaluation of summaries”, In Text summarization branches out, 74-81, 2004.

Appendix

A. Guideline for annotation

COVID-19 관련 추출 요약문 생성을 위한 라벨링 작업

*정의 측면 2가지

Domain	**Aspect**		**예시 주제**
COVID-19	발생 경향	COVID-19의 발생 현황 및 경향에 대한 정보를 전달하는 문장	감염자, 감염 속도, 발생 국가 등
	코로나 대응	COVID-19에 대응하여 취하는 모든 행위에 대한 정보를 전달하는 문장	백신, 치료제, 정책, 규제 등

*각 측면에 대한 문장들의 관련성 및 중요도에 대하여 점수 부여
- 점수는 0점부터 3점까지 1점 단위로 부여 가능 (0점, 1점, 2점, 3점)
- 점수 0점은 해당 측면과 관련 없음을 의미
- 측면과의 관련도가 있을 경우 1점 이상의 점수를 주며, 중요도에 따라 점수를 차등 부여
- 다음 두가지 사항에 중점을 두어 점수 부여

i) **긍정/부정 상관없이** 정해진 측면에 해당하는지

ii) 기사의 내용에서 핵심적인 문장인지

*점수 부여 시, 주의 사항
- 한 기사 내 **최소 3개 문장 이상에 1점 이상**의 점수 부여
- 최고점 문장은 **최대 3개**

*작업 예시 : 엑셀 파일로 제공되며, 각 측면에 해당하는 란에 점수를 기입.

South Sudan Becomes the 51st Country in Africa With COVID-19 as It Confirms Its First Case	**발생 경향**	**관련 대책**
JUBA, South Sudan — Officials in South Sudan say the country has its first case of COVID-19, making it the 51st of Africa's 54 countries to have the disease.	3	0
First Vice President Riek Machar and the U.N. mission in South Sudan confirmed the case of a U.N. worker who arrived in the country from Netherlands on Feb. 28.	1	0
The patient, a 29-year-old woman, first showed signs of the disease on April 2 and is recovering, said officials.	2	0
South Sudan, with 11 million people, currently has four ventilators and wants to increase that number, said Machar, who emphasized that people should stay three to six feet apart from others.	1	3
"The only vaccine is social distancing," said Machar.	0	2
The patient is under quarantine at U.N. premises and health workers are tracing the people who had been in contact with her, said David Shearer, head of the U.N. operations in South Sudan.	0	1
He said he hoped the measures would contain the case.	0	0
To prevent the spread of the virus in South Sudan, President Salva Kiir last week imposed a curfew from 8:00 p.m. to 6:00 a.m. for six weeks and closed borders, airports, schools, churches and mosques.	0	3
With the disease in South Sudan, now just three countries in Africa have not reported any cases of COVID-19: the tiny mountain kingdom of Lesotho in southern Africa, and the island nations of Comoros and Sao Tome and Principe.	3	0
Ethiopia on Sunday reported its first death from the virus and announced five more cases bringing its total to 43, most of them imported by travelers.	3	0
Ethiopia's Prime Minister Abiy Ahmed held discussions Sunday with opposition party leaders on measures to combat the virus.	0	2
A number of Ethiopia's regional states have implemented bans on movement of people and vehicles, but not yet in the capital Addis Ababa.	0	2
AP journalist Elias Meseret in Addis Ababa, Ethiopia, contributed to this story.	0	0

< Example of annotation for each aspect >

B. Results of experiments for finding a proper relevance score

· Results of the proposed method1

Aspect	Relevance score	Rouge-1			Rouge-2			Rouge-L		
		R	P	F	R	P	F	R	P	F
Trend	$\cos(\theta)$	66.3	62.8	63.8	56.5	53.7	54.5	64.7	61.4	62.4
	$\frac{\lvert s\rvert}{\lvert k\rvert}\cos(\theta)$	67.3	64.2	65.1	57.8	55.4	56.1	65.7	62.8	63.6
Action	$\cos(\theta)$	60.7	62.5	61.1	51.3	52.7	51.6	59.1	60.1	59.5
	$\frac{\lvert s\rvert}{\lvert k\rvert}\cos(\theta)$	60.6	61.9	60.8	51.1	52.2	51.3	58.9	60.1	59.1

The ‘cosine’ method refers to “using cosine similarity” and the ‘projection’ method refers to “using the ratio of projected sentence feature and keyword feature”

Depending on the dataset, the effectiveness of the methods was different. However, since the increase rate by ‘the projection’ method in the “Trend” is higher than increase rate by the cosine method in the “Action”, the ‘projection’ method was adopted to calculate the relevance score in the proposed method 1.

· Results of the proposed method2

Aspect	Relevance score	Rouge-1			Rouge-2			Rouge-L		
		R	P	F	R	P	F	R	P	F
Trend	$\cos(\theta)$	66.5	63.2	64.1	57	54.2	55	64.9	61.8	62.7
	$\frac{\lvert s \rvert}{\lvert k \rvert}\cos(\theta)$	69.4	67.5	67.7	60.9	59.1	59.4	68.1	66.3	66.5
Action	$\cos(\theta)$	60.6	62.5	61.1	51.1	52.6	51.5	59.1	60.9	59.5
	$\frac{\lvert s \rvert}{\lvert k \rvert}\cos(\theta)$	60.6	62.9	61.3	51.7	53.3	52.2	59.2	61.2	59.7

The ‘projection’ method showed better performance than the ‘cosine’ method, the ‘projection’ method was used to calculate the relevance score in the proposed method2.

C. Examples of generated output

C.1 Article of Table 4.2.2.1.

(**Bold** : the gold summary for aspect “Trend”)

Title : COVID-19: Pak to resume domestic flights from Saturday
1. Pakistan will resume domestic flights in a phased manner from Saturday after the government eased some of the restrictions imposed following the outbreak of the COVID-19 pandemic in the country that has claimed the lives of over 800 people. 2. The Pakistan government last week said that it would begin a phased lifting of the lockdown due to its effect on the economy and the workforce. 3. Pakistan Civil Aviation Authority (PCAA) announced to resume the domestic flights operations from Saturday, The Express Tribune reported. 4. Pakistan International Airlines (PIA) and Serene Air will operate flights in the morning and afternoon. 5. It said that 68 flights will operate from Jinnah International Airport in Karachi, 32 from Allama Iqbal International Airport in Lahore, 32 from Islamabad International Airport, eight from Quetta International Airport and four from Bacha Khan International Airport in Peshawar. **6. According to the Ministry of National Health Services, the nationwide tally of the coronavirus cases on Friday soared to 37,218 after 1,430 new cases were reported.** 7. A total of 10,155 patients have so far recovered from the virus. **8. The total number of COVID-19 deaths in the country also reached 803 with 33 new fatalities reported during the last 24 hours.** **9. So far, 344,450 tests have been conducted, including 13,700 during the last 24 hours.** 10. After easing the restrictions for the last four days, the Punjab government on Friday imposed a three-day lockdown throughout the province. 11. All markets across the province will remain closed except the grocery stores which will remain open from 9am to 5pm. 12. Medical stores will operate even after 5pm. 13. Meanwhile, Prime Minister Imran Khan joined a call by world leaders for a "peoples' vaccine" to combat the coronavirus. 14. "We must work together to beat this virus," the Prime Minister's Office quoted Khan as saying.

C.2 Article of Table 4.2.2.2.

(**Bold** : the gold summary for aspect “Action”)

Articles : U.S. passes 60,000 coronavirus deaths
1. The United States reached a new milestone in the coronavirus Wednesday after it surpassed 60,000 deaths, according to data compiled by Johns Hopkins University. 2. The U.S. has recorded 60,207 deaths from coronavirus, and 1,030,487 cases, which account for about one third of the global total. 3. The death toll represents a larger death toll than the Vietnam War. 4. However, according to the Centers for Disease Control and Prevention, the number of deaths may be underreported. 5. Despite the rising numbers, more cities and states are easing restrictions. **6. Earlier Wednesday, President Trump said he was “in favor” of the states that had begun allowing businesses to re-open with restrictions and that social distancing guidelines would soon be “fading out.”** **7. Experts have warned that abandoning restrictions too soon could lead to more surges in the virus’ spread and that the pandemic won’t truly end until a vaccine is developed and widely distributed.** **8. During that same Oval Office meeting, Dr. Anthony Fauci said a recent drug trial showed remdesivir “can block” coronavirus.**

C3. Additional Example

Bold font refers to gold summary. Red font refers to wrong sentence and unrelated to target aspect. And green font refers to wrong sentence, but related to target aspect.

· Example for the aspect “Trend”

Title : SXSW cancelled due to coronavirus concerns
1. Democrats call for a more competitive market in the tech sector at South by Southwest in Austin over the weekend. 2. City officials from Austin, Texas, announced the cancellation of the South by Southwest (SXSW) media and music festival on Friday, for fear that the close contact of so many might hasten the spread of the novel coronavirus. 3. Mayor Steve Adler said, after consulting various health officials, he's decided that the best course of action would be to scrap the event entirely. 4. "Based on the recommendation of our public health officer and director of public health ... I've gone ahead and ... issued an order that effectively cancels South by Southwest for this year," he said. 5. Travis County Judge Sarah Eckhardt also spoke at the presser and said she would be signing a disaster order in an effort to limit the transmission of the virus. 6. "I am signing a companion declaration of disaster, which will apply countywide to festival gatherings that are attracting individuals from areas that have documented cases of person-to-person transmission of COVID-19," she added. 7. "This is a medical and data-driven decision," Eckhardt continued. **8. Austin Public Health Interim Director Dr. Mark Escott called the decision a "proactive" measure and noted that there are no confirmed cases of COVID-19 in Travis County as of yet.** 9. He also called for calm and said the declaration should be viewed as a positive. 10. "Now is not the time to panic," Escott added. 11. Now is the time to prepare and to provide a measured response to that threat. 12. Not all mass gatherings need to be canceled, he continued. 13. However, with no vaccine or treatment available, local leaders thought it best to err on the side of caution. **14. Globally, the virus has spread to 85 countries, infecting 99,624 people and claiming the lives of more than 3,400 others.** 15. The bulk of the cases are in mainland China, where the outbreak originated in December 2019. **16. The U.S. has 244 confirmed cases in 18 states and 12 deaths, the majority in Washington state.**
Baseline Summary
2. City officials from Austin, Texas, announced the cancellation of the South by Southwest (SXSW) media and music festival on Friday, for fear that the close contact of so many might hasten the spread of the novel coronavirus. 15. The bulk of the cases are in mainland China, where the outbreak originated in December 2019. 16. The U.S. has 244 confirmed cases in 18 states and 12 deaths, the majority in Washington state.
Proposed method1 Summary (Ours)
8. Austin Public Health Interim Director Dr. Mark Escott called the decision a "proactive" measure and noted that there are no confirmed cases of COVID-19 in Travis County as of yet. 15. The bulk of the cases are in mainland China, where the outbreak originated in December 2019. 16. The U.S. has 244 confirmed cases in 18 states and 12 deaths, the majority in Washington state.
Proposed method2 Summary (Ours)
8. Austin Public Health Interim Director Dr. Mark Escott called the decision a "proactive" measure and noted that there are no confirmed cases of COVID-19 in Travis County as of yet. 15. The bulk of the cases are in mainland China, where the outbreak originated in December 2019. 16. The U.S. has 244 confirmed cases in 18 states and 12 deaths, the majority in Washington state.

· Example for the aspect "Action"

Title : 'Very fierce' COVID-19 second wave to come in colder months, warns Taiwanese health expert

1. MANILA - A "very fierce" second wave of the COVID-19 pandemic will come during colder weather when people are more susceptible to respiratory infections, a Taiwanese infectious disease specialist said Tuesday.
2. The southern hemisphere of the globe is currently experiencing the height of the pandemic while it is slowing down in the northern hemisphere due to warmer temperature, said Dr. Mei-Shang Ho, president of the Taiwan Health Corporation.
3. "But the second wave will come.
4. How soon it's hard for me to say.
5. Definitely in the fall or winter, one would experience a very fierce second wave," she told ANC.
6. She noted that while there are not as many cases of COVID-19 in the summer, "the virus probably doesn't go away."
7. "It might silently transmit among mild, asymptomatic cases in such a way that where it might seed in the community is unknown," she added.
8. **Taiwan, which has been deemed as a "success story" in battling the pandemic, implements isolation, identification of infection source, physical distancing and practice of good personal hygiene to combat the virus, according to Ho.**
9. "These are very basic (measures) that we continue to follow," she said.
10. **Ho, meantime, said a vaccine for the new strain of virus might become available next year.**
11. **"For popular use, probably next year, with the massive effort and resources pouring in globally.**
12. Probably next year with reasonable amount for use.
13. That's a very optimistic view, assuming everything goes well," she said.
14. It would be better for Taiwan, which has so far reported 426 cases of coronavirus and 6 deaths, to join the World Health Organization, Ho said.
15. "Personally, I sort of look at it in a humble way.
16. We just won the first battle.
17. We still have a long way to go.
18. It's better that we are with the whole pack, we are freely communicating with others," she said.

Baseline Summary

1. MANILA - A "very fierce" second wave of the COVID-19 pandemic will come during colder weather when people are more susceptible to respiratory infections, a Taiwanese infectious disease specialist said Tuesday.
2. The southern hemisphere of the globe is currently experiencing the height of the pandemic while it is slowing down in the northern hemisphere due to warmer temperature, said Dr. Mei-Shang Ho, president of the Taiwan Health Corporation.

10. Ho, meantime, said a vaccine for the new strain of virus might become available next year.

Proposed method1 Summary (Ours)

2. The southern hemisphere of the globe is currently experiencing the height of the pandemic while it is slowing down in the northern hemisphere due to warmer temperature, said Dr. Mei-Shang Ho, president of the Taiwan Health Corporation.

9. "These are very basic (measures) that we continue to follow," she said.
10. Ho, meantime, said a vaccine for the new strain of virus might become available next year.

Proposed method2 Summary (Ours)

2. The southern hemisphere of the globe is currently experiencing the height of the pandemic while it is slowing down in the northern hemisphere due to warmer temperature, said Dr. Mei-Shang Ho, president of the Taiwan Health Corporation.

9. "These are very basic (measures) that we continue to follow," she said.
10. Ho, meantime, said a vaccine for the new strain of virus might become available next year.

초록

텍스트 요약(Text Summarization)은 자연어 처리 분야의 대표적인 작업 중 하나이다. 텍스트 요약의 목적은 신문 기사와 같은 문서를 간결하지만, 핵심적인 내용을 중심으로 요약하는 것이다. BERT, GPT-3와 같은 트랜스포머 기반의 사전학습 모델들이 개발됨에 따라, 요약 모델의 성능이 크게 향상되었다.

최근에는 사용자의 목적 혹은 선호도를 반영하여 출력을 생성하는 언어 모델을 개발하기 위해 많은 연구들이 진행되고 있다. 텍스트 요약 분야에서도 이러한 흐름에 따라 쿼리 중심(Query focused) 혹은 측면 중심(Aspect oriented) 요약과 같이 제어 가능한 요약문 생성에 대한 연구들이 등장하고 있다. 측면 중심 요약(Aspect oriented)은 사용자가 알고 싶은 특정 측면에 대해서 요약문을 생성하는 것을 목표로 한다.

본 논문에서는 선행 연구에서 제안한 측면 중심 요약 모델의 성능 향상을 위한 방법을 제안한다. 제안된 방법은 문장의 표현 벡터와 측면을 대표하는 키워드 표현 벡터들 사이의 연관성을 기존의 문장 표현 벡터에 반영함으로써 모델이 측면과 관련된 요약문을 생성하도록 했다. 평가를 위해서, "발생 현황"과 "관련 대응"이라는 두 가지 측면을 가지는 COVID-19 관련 기사로 구성된 새로운 데이터셋을 구축하였다. 제안된 방법들은 새로운 데이터셋에 대하여 기존 모델보다 더 좋은 성능을 보여주었다.

제안된 방법은 '발생 현황' 측면에서는 3.6~4.3%로 높은 성능

향상을 가져왔으며, ‘관련 대응’ 측면에서는 1%미만의 향상으로, 비교적 낮은 효과를 보여주었다. 하지만 두 측면 모두에서 오답이라 하더라도 측면과 관련된 문장을 선택하는 것을 관찰했다. 이를 통해, 제안된 방법이 모델의 측면 지향 요약에 도움을 주었음을 확인할 수 있었다.

Keyword : 자연어 처리, BERT, BERTSum, 트랜스포머, 측면 중심 요약, 텍스트 요약, 키워드 중심 요약

Student Number : 2021-22509