

Text summarization of dialogue based on BERT

Wongyung Nam*, Jisoo Lee*, Beakcheol Jang*

*Student, Graduate School of Information, Yonsei University, Seoul, Korea

*Student, Graduate School of Information, Yonsei University, Seoul, Korea

*Professor, Graduate School of Information, Yonsei University, Seoul, Korea

[Abstract]

In this paper, we propose how to implement text summaries for colloquial data that are not clearly organized. For this study, SAMSum data, which is colloquial data, was used, and the BERTSumExtAbs model proposed in the previous study of the automatic summary model was applied. More than 70% of the SAMSum dataset consists of conversations between two people, and the remaining 30% consists of conversations between three or more people. As a result, by applying the automatic text summarization model to colloquial data, a result of 42.43 or higher was derived in the ROUGE Score R-1. In addition, a high score of 45.81 was derived by fine-tuning the BERTSum model, which was previously proposed as a text summarization model. Through this study, the performance of colloquial generation summary has been proven, and it is hoped that the computer will understand human natural language as it is and be used as basic data to solve various tasks.

▶ **Key words:** Text Mining, BERT, Text Summarization, Abstractive Summarization, Dialogue Data

[요 약]

본 연구는 일목요연하게 정리되지 않은 구어체(대화)에 대한 텍스트 자동 요약 모델을 적용시키는 방법을 제안한다. 본 연구에서는 구어체 데이터인 SAMSum 데이터를 활용하였고, 선행연구에서 문어체 데이터 텍스트 자동 요약 모델 연구에서 제안한 BERTSumExtAbs 모델을 적용하였다. SAMSum 데이터셋은 70% 이상은 두 사람 간 대화, 나머지 약 30%는 세 사람 이상 간 대화로 구성되어 있다. 본 논문에서는 텍스트 자동 요약 모델을 구어체 데이터에 적용하여, ROUGE Score R-1 부문에서 42.43 이상의 결과를 도출해내었다. 또한, 텍스트 요약 모델로 기존에 제안된 모델인 BERTSum 모델을 fine-tuning하여, 45.81의 높은 점수를 도출했다. 본 연구를 통하여 구어체 데이터에 대한 텍스트 생성 요약의 성능을 입증하였으며, 앞으로 사람의 자연어를 있는 그대로 컴퓨터가 이해하여 다양한 task를 해결하는 데 기초 자료로 활용되길 바란다.

▶ **주제어:** 텍스트마이닝, BERT, 문서요약, 생성요약, 구어체데이터

-
- First Author: Wongyung Nam, Corresponding Author: Jisoo Lee, Beakcheol Jang
 - *Wongyung Nam (wonkn@yonsei.ac.kr), Graduate School of Information, Yonsei University
 - *Jisoo Lee (lee.js@yonsei.ac.kr), Graduate School of Information, Yonsei University
 - *Beakcheol Jang (bjang@yonsei.ac.kr), Graduate School of Information, Yonsei University
 - Received: 2022. 07. 13, Revised: 2022. 07. 26, Accepted: 2022. 08. 08.

I. Introduction

하루에도 수많은 매체로부터 텍스트, 이미지, 영상 등 다양한 종류의 정보를 접하며 살고 있다. 우리는 많은 정보를 쉽게 접하고 쉽게 저장할 수 있는 시대를 살아가고 있지만, 이러한 정보에 대해 순간적이고 집약적으로 의미를 파악하기란 어렵다. 특히 데이터 유형 중에서도, 직관적인 감각으로 느낄 수 있는 이미지와 영상, 음성 데이터보다, 텍스트 데이터는 인지 및 이해의 영역이 필수 불가결하다. 이처럼 사람이 직접 텍스트 데이터로부터 핵심적인 의미를 추출해내는 일은 단순하지 않고 매우 복잡하며, 이를 AI가 대신할 수 있도록 다양한 선행 연구들이 시도되어지고 있다. 가장 대표적으로는 BERTSum 모델을 제안한 Liu, Y., & Lapata, M. [1] 이 있다. 해당 논문에서는 뉴스기사 데이터를 대상으로 실험하여 Deep neural network 모델인 BERTSum의 높은 성능을 입증한 바 있다.

과연 일목요연하게 정리되어 있는 문어체인 뉴스기사 데이터에 대한 실험만 수행했던 해당 모델이 실제 인간 언어 생활에 얼마나 유용한가라는 의문에서 본 연구가 시작되었다. BERTSum에서 활용한 뉴스기사 데이터는 육하원칙하에 잘 정리되어 있는 문서 양식을 따르고 있다. 동양권 문화에서는 문어체에 대한 중요도가 높게 평가되고 있지만, 동서양을 막론하고 실제 일상생활에서의 SNS, 문자 메시지, 입으로 표현하는 현실적인 말 표현 등 구어체 사용 빈도는 문어체와 비교할 수 없을 정도로 매우 높다. 인간의 삶 속에서는 문어체보다는 구어체로 된 텍스트를 빈도 높게 사용하고 있으며, 사람만큼 구어체를 인지하고 이해하는 AI 기술이 필요한 시점이다. 실제로, 전 세계적으로는 물론, 국내에서도 인공지능과 사람이 자연어 형태로 대화하는 것을 목적으로 실생활과 밀접한 도메인에 대해 이해하여 답변 문장을 생성해주는 KoDialogPT2가 개발 [2]되는 등 다양한 시도가 이어지고 있다.

본 연구에서는 사람의 입으로 표현되는 구어체는 물론, 인간 세상에서 일어날 법한 이야기, 인간의 심리적인 묘사 등 다양한 표현이 존재하는 대화체 데이터를 분석하고, 텍스트 자동요약 모델을 적용하는 방법을 제안하고자 한다. 이로써 인공지능이 대화의 문맥을 얼마나 잘 이해하고, 이를 정확하게 요약하는지 성능을 평가해 보고자 한다. 즉, 문어체로만 서술되어 있는 뉴스기사와는 다른 특성을 가진 구어체 데이터 셋에 사전 학습된 언어 모델인 BERTSum을 적용해보고, 논리정연하게 정리되지 않은 구어체에서도 얼마나 좋은 성능을 보이는지 확인해보고자 한다.

II. Related Work

빠른 기술 발전으로 방대한 자료와 정보를 쉽게 접할 수 있는 만큼, 효율적인 정보 처리를 위한 요약 기술의 중요성이 점차 증대되고 있다. 텍스트 요약 연구는 컴퓨터 기술이 발전하기 이전인 1950년대 후반[3]부터 이어져 오고 있으며, 최근까지 새로운 기술과 접근이 시도되어 오고 있다.

기존 선행연구[4]에서는 ‘요약’에 대해 ‘중요한 정보를 전달하기 위하여 하나 이상의 텍스트로 구성된 문장 또는 글’이라고 규정하고 있다. 중요한 주요 정보를 놓치지 않고 유지해야한다고 강조하면서, 컴퓨터가 인간의 지식과 이해의 영역을 수행해야하는 작업이기 때문에 매우 어려운 작업이라 언급하였다. 이에 복잡하고 고난도 작업에 적합하고, 이미 11개의 Task 분야에서 성능이 입증된 BERT를 기반으로 한 BERTSum 모델을 본 연구의 Baseline 모델로 선정하였다.

자동 텍스트 요약 기술은, 본문 내에서 추출한 워드를 조합한 추출 요약(extractive summarization)과 본문에 쓰이지 않은 새로운 단어를 사용하여 문장을 만들어주는 생성 요약(abstractive summarization)으로 나누어 볼 수 있다. 추출 요약은 생성 요약 대비 상대적으로 간단한 방법이며, 문서에서 각 문장에 점수를 할당하고, 가장 높은 점수를 받은 문장이 요약되는 방법이며, 생성 요약은 원본을 재해석하여 본문과 다른 표현으로 요약글이 생성되는데, 추출 요약보다 복잡한 대신 현실에 더 적합하기 때문에 성능 면에서는 생성 요약이 더 뛰어나다[5].

추출 요약을 위해 활용되는 모델은 SummaRu-NNer[6], NeuSum[7], BERTSum[1] 등이 있으며, BERTSum 모델 이전의 추출 요약은 각 문장에 0(비선정) 또는 1(선정)로 점수를 매기고 선정된 문장을 추출하는 방식으로 이뤄졌으나, BERTSum은 연속값인 ROUGE Score를 산출하여 time step 마다 해당 문장의 추출여부를 ROUGE Score에 따라 판단하는 방법에 기인한다.

BERTSum 논문이 발표된 이후 생성 요약을 위해 PreSumm[8] 논문이 발표되었으며, 생성적 요약을 수행하는 BERTSumAbs와 BERTSumExtAbs 모델이 발표되었다. 이들은 BERTSum[1]과 구조적으로 동일하지만, transformer encoder가 아닌 decoder를 추가한 구조이다. 다만 두 모델의 가장 큰 차이점은, BERTSumExtAbs 모델은 fine-tuning을 두 번 진행한다는 점이다. 추출 요약을 학습한 인코더에 생성 요약에 대한 Task를 다시 한번 학습시킨다. 바로 생성요약을 진행하는 BERTSumAbs보다 2단계에 걸쳐 추출 요약을 먼저 학습한 후 생성 요약

을 진행하는 BERTSumExtAbs 성능이 더 뛰어나다[8].

텍스트 요약 기술의 다른 측면으로, 어떤 데이터를 활용하는지도 중요하다. BERTSum 논문에서는 뉴스 데이터를 활용하였고, 뉴스는 일목요연하게 육하원칙에 의해 정리된 데이터이다. 기존의 텍스트 요약 기술은 정돈된 문서 자료의 요약에서부터 시작되었으나, 컴퓨터가 일상 언어를 이해하고, 자연어를 그대로 처리하기 위해 대화형 데이터를 요약하는 기술들로 발전해가고 있다. 대표적으로는 트윗 데이터[9]가 있으며, 트윗 데이터 역시 글로 작성한 데이터로, 본 연구에서는 일상 대화체 데이터인 SAMSum을 활용하였다. 원문 그대로의 내용 전달이 중요한 경우에는 추출 요약이 유용하기도 하지만, 인간의 언어능력과 보다 가깝다고 평가되는 자동요약 기능은 생성 요약이라 할 수 있기 때문에, 본 연구의 데이터인 대화체 특성 상 쓰이지 않은 단어 외에도 새로운 단어 조합이 이뤄질 수 있도록 생성요약 방법을 활용하였다.

III. The Proposed Model

3.1 SAMSum Dataset

본 연구에서는 SAMSum Corpus[9] 데이터셋을 이용하여 구어체 데이터에 대한 추상적 요약(Abstractive summarization)을 수행하는 모델의 성능을 평가한다. 데이터는 14,732개의 Train 데이터, 818개의 Validation 데이터, 819개의 Test 데이터로 구성되어 있다.

SAMSum 데이터셋은 70% 이상의 대화가 두 사람간의 일상적인 채팅 내용이고, 나머지 약 30%이상의 대화는 세 사람 이상간의 대화로 구성되어 있다. 대화 내용에는 대화하는 사람의 이름과 속어, 이모티콘도 포함되어 있다. 각 대화 한 개에는 인간이 직접 작성한 요약문 한 개가 포함되어 있으며, 요약문은 한 문장에서 세, 네 문장까지 구성되어 있다. Table1은 데이터 중 한 개의 대화 내용을 보여주고 있다.

Table 1. SAMSum dialog and summary

Dialog
Aria: You won't believe who I've just met! Aria: Charlie Evans! Maverick: Oh God, I haven't seen him from ages! Maverick: How is he doing? Aria: He's doing great. :) Aria: He got married, he runs a small family business, which he is very passionate about and generally he seems to be a happy and fulfilled man. :) Aria: Oh, and he has two absolutely adorable daughters. :) Aria: It was so nice to meet him, he's such a sweet soul. Maverick: I'm glad to hear that. :) Maverick: Time flies so fast, doesn't it? Aria: It does. :) Recently I've met Cooper Roy, I'm sure you remember him, I constantly talked about him back then in high school, because I had a huge crush on him. He's changed a lot! Aria: I barely recognized him! Maverick: I remember, how you and Kylie used to gossip about him all the time. :) Aria: Well, we were madly in love with him. Aria: Or at least we thought so. :) Maverick: I sometimes miss those times. Maverick: Everything seemed to be easier, better and calmer back then. Aria: The world wasn't as crazy as it is now. Aria: I feel like politics got crazier, people - more radical and hostile and economics - less predictable... Maverick: Me too...:(
Summary
Aria has just run into Charlie Evans. He is now married, with two daughters, and a family business. She has also met Cooper Roy from high school. She used to have a crush on him, now she almost didn't recognise him. Maverick and Aria miss the old times and think the world has changed for the worse.

3.2 Model

우리는 구어체 텍스트의 추상적 요약에 대한 모델 성능을 측정하기 위한 실험을 수행했고, CNN, Daily Mail 뉴스 기사에 대한 실험 결과와 비교 분석하였다. 실험에서는 BERT(Pre-training of Deep Bidirectional Transformers)[10] 모델을 기반으로 하여 텍스트 요약에 대한 Fine-tuning하여 만든 BERTSum[1] 모델을 사용하였다. BERTSum 모델에서는 Bert와 동일하게 세 단계(Token, Segment, Position)의 embedding 구조로 이루어져 있다. 그러나, [CLS] 토큰을 입력 데이터의 처음에만 삽입하는 Bert와 달리, 문장과 문장 사이에 [SEP]와 [CLS] 토큰을 반복 삽입하여 문장을 구분하는 Embedding 방식을 사용하는 점에 차이가 있다. Table2에서는 SAMSum 데이터 중 하나의 대화 샘플의 Input document 형태를 확인할 수 있다.

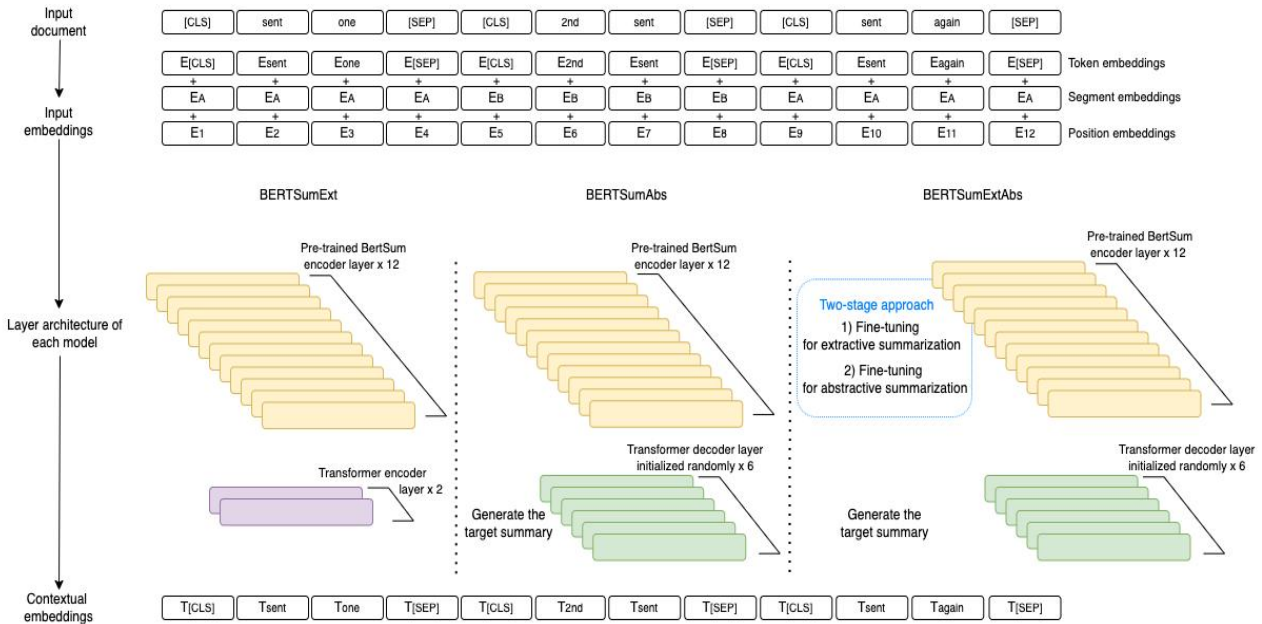


Fig. 1. Workflow of BERTSumExtAbs

본 실험에서는 베이스라인 모델로 TransformerAbs 모델과 BERTSum 기반 모델로 BertExtAbs 모델을 비교 모델로 선정하였다. 베이스라인 모델 Transformer -Abs는 텍스트 추상 요약에 위해 제안된 인코더와 디코더 구조의 Transformer 모델[11][12]로 구성되어 있다. Transformer 모델은 OpenNMP library[13]를 사용하여 훈련되었다.

Table 2. SAMSum dialogue and Input document sample

Dialogue
Hannah: hey, do you have Betty's number?
Amanda: let me check
Amanda: sorry, can't find it
Amanda: ask Larry
Amanda: he called her last time we were at the park together
Hannah: i don't know him well
...
Amanda: bye bye
Input document
[CLS] 0 , " hannah : hey , do you have betty's number ? [SEP] [CLS] amanda : le ### me check [SEP] [CLS] amanda : sorry , ca n ##' ##t find it . [SEP] [CLS] amanda : ask larry [SEP] [CLS] amanda : he called her last time we were at the park together [SEP] [CLS] hannah : i do n ##' ##t know him well [SEP] [CLS]
...
[SEP] [CLS] amanda : bye bye [SEP]

텍스트 추상적 요약 Task에 특화되어 설계된 BertExtAbs 모델은 12개의 사전 학습된 BERTSum 인코더

와 6개의 Transformer layer로 구성된 구조이다. 추출 요약에 최적화된 BERTSumExt 모델은 사전 학습 (Pre-trained)된 BERTSum 인코더(Encoder)와 Transformer layer의 Encoder로 구성되어. 마지막 인코더 위에 Transformer layer가 쌓여 있는 구조이다. Transformer layer(L) 2개가 쌓인 모델이 가장 좋은 성능으로 뽑혀 BERTSumExt 모델이라 이름 붙여져 있다. 본 논문에서 사용한 모델 BERTSumExtAbs 모델은 이러한 추출적 요약 모델인 BERTSumExt 모델과 달리, Token embedding, Segment embedding, Position embedding 단계를 거친 데이터가 인코더에서 추출적 요약을 위해 먼저 fine-tuning 된 후, 추상적 요약을 위한 fine-tuning 과정을 한 번 더 거친다는 점에서 차이가 있다. Transformer decoder에서는 예측된 추출 요약을 생성한다. 텍스트 추상적 요약에 특화되도록 두 단계의 fine-tuning되어 배포된 BERTSumExtAbs 모델을 비교 모델로 선정하여 모델 성능을 비교하였다. 추출적 요약(Extractive summarization)을 수행하도록 fine-tune된 인코더를 처음 사용한 뒤, 추출적 요약을 수행하도록 설계된 모델이다.

IV. Experiments

4.1 Experimental Set-up

실험은 Pytorch와 OpenNMT 코드를 기반으로 하여 수행하였다. 텍스트 토큰화(Tokenization)에는 BERT의 서브

워드 토큰라이저(Subwords tokenizer)를 사용하였다. 본 논문에서 실험의 훈련은 사전학습된 모델을 이용하여 수행되었다. 사전학습된 모델의 세부 설계 내용은 Table3와 같다. 모든 모델에 GTX 1080 Ti 4개의 GPU를 이용하여 트레이닝 되었으며, 5스텝마다 Gradient Accumulation되도록 설정되었다. 모델의 훈련은 매 2,000 스텝마다 Checkpoint가 저장되도록 설계되었고, 인코더와 디코더에 Dropout 0.1옵션을 적용하여 훈련하였다.

Table 3. Experiments Set-up

Models	Train steps	Learning rate	Batch
TransformerAbs	50,000	0.05	500
BERTSumExtAbs	164,000	0.05	512

전체 데이터 셋 16,369개의 데이터 중 약 90%는 Train 데이터로, 약 5%는 Valid 데이터와 Test 데이터로 실험에 사용되었다. 베이스라인 모델의 경우 훈련데이터의 정확도가 73.05%, Perplexity(PPL)점수 2.57를 기록하였으나, Valid 데이터의 경우 각각 53.24%와 12.03의 점수를 기록하였다. PPL 점수는 문장 길이를 반영하여 확률값을 정규화한 값으로, 언어 모델 성능을 평가하는 또 다른 척도이다. PPL은 Cross entropy loss 값에 Exponential을 취한 값으로도 이해할 수 있는데, 이는 낮을수록 좋은 점수를 의미한다. BERTSumExtAbs모델의 경우 훈련데이터의 정확도가 71.43%, PPL이 3.19를 기록하였으나, Valid 데이터의 경우 각각 58.23%, 7.34로, 베이스라인 모델이 Train 데이터와 Valid 데이터간에 보인 격차보다는 다소 낮은 격차를 보이고있음을 확인할 수 있다. 즉, 아래 실험에서 더 좋은 성능을 보여준 BERTSumExtAbs모델이 Train데이터와 Valid 데이터간의 정확도 격차가 다른 모델에 비해, 미비하다고 할 수 있다.

4.1 Results

실험에 대한 평가는 ROUGE[14]의 F1 Score를 기준으로 평가하였다. ROUGE 점수는 자연어의 생성 모델 성능을 평가에 보편적으로 사용되는 지표 중 하나이다. 본 논문에서는 모델의 추론으로 생성된 자연어를 사람이 미리 작성해놓은 정답 요약문(Ground truth)와 비교, 대조하여 성능을 계산한다. R-1은 추론된 요약문과 정답 요약문간의 겹치는 Unigram의 수를 계산하는 지표이며, R-2는 두 요약문간 겹치는 Bigram의 수를 계산해주는 지표이다. R-L은 가장 긴 공통 부분의 Sequence를 찾는 문제에 기

반하여 점수를 측정하는 지표로, 가장 길이의 매칭되는 문자열을 계산해주는 지표이다.

본 연구에서는 베이스라인 모델과 BERT 기반의 BERTSumExtAbs 모델 모두 자동 문서 요약 모델의 성능이, 뉴스 데이터에서뿐만 아니라 일상 대화 데이터의 텍스트 요약에 대해 높은 성능을 보여주고 있음을 입증하였다. 베이스라인 모델과 BERTSumExtAbs 모델의 ROUGE Score를 Table4에서 확인할 수 있다. 두 모델의 성능 차이는 Transformer layer decoder의 유무와 두 번의 fine-tuning과정이 더 이루어졌다는 점에 있다. 두 모델 모두 R-L 점수측면에서는 다소 떨어지는 성능을 보여주지만, R-1 점수부분에서는 충분히 유의미한 성능을 보여주는 점수를 달성하였다. 특히, 추출적 요약을 기반으로 다시 Fine-tuning하여 훈련된 BERTSumExtAbs모델의 경우, 일상 대화에 대해서 텍스트 자동 요약 모델이 훌륭하게 기능하고 있다는 것을 확인할 수 있다.

Table 4. Experiments Results

Models	R-1	R-2	R-L
TransformerAbs	42.43	17.58	34.45
BERTSumExtAbs	45.81	22.36	38.38

ROUGE Score는 많은 텍스트 요약 모델의 성능을 측정하는 지표로 사용되지만, 같은 의미이지만 다른 단어로 표현되는 동음이의어 등에 대해서는 정확도를 잡아내지 못한다는 단점이 있다. 따라서, 우리는 Table5에 모델이 추론하여 생성한 텍스트 요약문의 정확도를 확인할 수 있도록 추론된 요약문의 샘플을 첨부하였다. Table5에서 Ground truth는 Dialogue 대화에 대해 인간이 직접 요약문을 작성한 정답 요약문이다. 본 연구에서 베이스라인 모델이 된 TransformerAbs 모델과 이보다 좋은 성능을 보인 BERTSumExtAbs 모델이 생성한 요약문을 Table5 Inferred summary에 첨부하였다. TransformerAbs 모델도 문장의 흐름을 대략적으로 파악하여 주어-서술어-목적어가 들어간 온전한 문장으로 요약 표현하는 것을 확인할 수 있다. 또한, 대화 흐름에서 가장 중요한 내용을 신뢰할만한 수준으로 요약하고 있다는 것도 확인할 수 있다. 그러나 베이스라인 모델에서는 학교폭력을 보고하고 있는 사람이 Corbin임을 포착하였지만, 학교폭력의 희생자와 Corbin이 어떤 관계인지 파악하는 문맥은 제외되어있고 상담사가 Corbin의 친구를 도와주려고 한다는 문장으로 요약되어 있다. 반면에, BERTSumExtAbs모델의 경우에

는 학교폭력을 보고하고 있는 사람이 Corbin이며 그 폭력 대상자는 Corbin의 친구라는 중요한 문맥을 잘 포착하여, 베이스라인 모델과 비교했을 때, 대화에서 중요한 문맥을 보다 정확하게 포착하여 요약문을 추론하고 있다는 점을 확인할 수 있다.

Table 5. Dialogue and inferred summary with TransformerAbs and BERTSumExtAbs

Dialogue
Corbin: Is this the department in charge of school violence? Dimitri: Yes, it is. Corbin: I want to report school violence in our school. Dimitri: Okay. What school are you in? Corbin: Jungang high school. The student who is victim of the violence is my friend. They are not hitting him any more. But you should help my friend. If they notice i was the one who reported, they will hit me as well. Dimitri: First, calm down. Give us your phone number. You will be safe. Corbin: 486-984-324. It is. Dimitri: Don't worry. I will call you now, ok?
Ground truth
corbin reported to the department in charge of school violence that his friend has been beaten
Inferenced summary with TransformerAbs
corbin wants to report school violence in his school violence. dimitri wants to help his friend.
Inferenced summary with BERTSumExtAbs
corbin wants to report school violence in jungang high school . the student who is victim of the violence is corbin 's friend .

V. Conclusions

텍스트 요약 중에서도 ‘생성 요약’이란, Input DATA의 text를 활용하지 않고, 새로운 단어로 문장을 생성해주는 기술이다. 원문 그대로의 내용 전달이 중요한 경우에는 ‘추출 요약’이 활용되기도 하지만, 사람의 대화를 자연어 그대로 이해하고 대화체 특성 상 사용되지 않은 새로운 단어를 조합하여 요약문을 생성하는 ‘생성 요약’ 기술의 중요성이 점차 증대되고 있다. 최근 AI 기술 발달로 문어체(문서, 뉴스 등)의 요약 기술의 성능은 이미 많은 모델과 다양한 데이터로 꾸준히 증명되고 있으나, 구어체의 요약 기술 성능은 아직 부족한 실정이다. 이에 본 연구에서는 구어체인 일상 대화문 데이터로 생성 요약의 성능을 검증하였으며, 그 결과 텍스트자동요약 모델의 성능이 높은 수치를 보이는 것을 확인하였다.

향후 연구는 영어로 된 대화체 데이터를 활용한 점을 꼽을 수 있으며, 향후 한글로 된 대화체 데이터를 활용한 생성 요약 성능 연구가 필요하겠다. 이를 위해서는 BERTSum 모델을 한국어 버전에 맞춘 KoBERTSum 모델이 필요하겠으며, 기존 선행연구에서도 KoBERTSum 모델을 기반으로 한 법률문서 요약 연구는 있었으나 대화체 연구는 없었기에, 한국어 대화체 연구가 필요하다.

Artificial Intelligence(AI)의 궁극적 목표는 인간의 세계를 컴퓨터의 언어로 이해하는 것이다. 본 연구에서는 보다 인간의 언어 능력과 밀접한, 생성적 요약에 관한 텍스트 자동 요약의 성능을 입증하였다. 해당 연구가 향후 인간의 언어생활에 기반한 AI 기술의 발전 초석이 되리라 믿는다.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea Fund of NRF-2022R1F1A1063961.

REFERENCES

- [1] Liu, Y. (2019). Fine-tune BERT for extractive summarization. arXiv preprint arXiv:1903.10318.
- [2] Oh, Dongsuk, et al. "KoDialoGPT2: Modeling Chit-Chat Dialog in Korean." Annual Conference on Human and Language Technology. Human and Language Technology, 2021.
- [3] Luhn H. "The automatic creation of literature abstracts." IBM J Res Dev 2, pp. 159-165, 1958.
- [4] Dragomir R Radev, Eduard Hovy, and Kathleen McKeown. "Introduction to the special issue on summarization." Computational linguistics, Vol. 28, No. 4, pp. 399-408, 12, 2002.
- [5] Gambhir, Mahak, and Vishal Gupta. "Recent automatic text summarization techniques: a survey." Artificial Intelligence Review, Vol. 47, No. 1, pp. 1-66, 1, 2017.
- [6] Nallapati, Ramesh, Feifei Zhai, and Bowen Zhou. "Summarunner: A recurrent neural network based sequence model for extractive summarization of documents." Thirty-first AAAI conference on artificial intelligence. 2017.
- [7] ZHOU, Qingyu, et al. "Neural document summarization by jointly learning to score and select sentences." arXiv preprint arXiv:1807.02305, 2018.
- [8] Liu, Y., & Lapata, M. "Text summarization with pretrained

encoders.”. EMNLP, 2019.

- [9] Feigenblat, G., Gunasekara, C., Sznajder, B., Joshi, S., Konopnicki, D., & Aharonov, R. “TWEETSUMM--A Dialog Summarization Dataset for Customer Service.” EMNLP, 2021.
- [10] Bogdan Gliwa, Iwona Mochol, Maciej Biesek, Aleksander Wawer, "SAMSum Corpus: A Human-annotated Dialogue Dataset for Abstractive Summarization," Proceedings of the 2nd Workshop on New Frontiers in Summarization, Association for Computational Linguistics, November 2019.
- [11] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. “Bert: Pre-training of deep bidirectional transformers for language understanding.”. NIPS, 2019.
- [12] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30, 2017.
- [13] Abigail See, Peter J. Liu, and Christopher D. Manning. “Get to the point: Summarization with pointer generator networks,” In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1073–1083, Vancouver, Canada, 2017.
- [14] <https://github.com/OpenNMT/OpenNMT-py>
- [15] Chin-Yew Lin, “ROUGE: A package for automatic evaluation of summaries. In Text Summarization Branches Out,” pages 74–81, Barcelona, Spain, 2004.

Authors



Wongyung Nam received the B.S. in Business Administration from Sangmyung University, Korea. Wongyung Nam is currently a student of Business Analytics at Graduate School of Information, Yonsei

University, Seoul, Korea, for M.S. degree. She is interested in Natural Language Processing, Artificial Intelligence, and Bigdata Analytics.



Jisoo Lee received the M.A. in Educational Psychology from Sookmyung University, Korea. Jisoo Lee is currently a student of Business Analytics at Graduate School of Information, Yonsei University, Seoul, Korea,

for Ph.D. degree. She is interested in Timeseries Forecasting, Artificial Intelligence, and Bigdata Analytics.



Baekcheol Jang received the B.S. in Computer Science from Yonsei University, Seoul, Korea, M.S. in Computer Science from KAIST, Daejeon, Korea and Ph.D. in Computer Science from North Carolina State

University, NC, USA, respectively. Baekcheol Jang is currently a Professor in the Graduate School of Information at Yonsei University, Korea. He is interested in Wireless Networking and Artificial Intelligence.