

# 텍스트 요약 성능 향상을 위한 수정된 BART 모델

조찬영, 박채원, \*이선영  
순천향대학교

{ccy0531, fuyu1291, \*sunlee}@sch.ac.kr

## Modified BART model for improve text summarization performance

Chan Yeong Cho, Chae Won Park, Sun-Young Lee\*  
Dept. of Information Security Soonchunhyang Univ.

### 요약

텍스트 요약 기술은 뉴스나 기사의 요약, 학술 논문 요약 등 여러 분야에서 적용되고 있다. 본 논문에서는 어텐션 메커니즘을 문장과 단어의 두 단위로 나눠 문서처리에 높은 성능을 보이는 계층적 어텐션을 통한 BART 모델의 미세조정 방식을 제안한다. 제안하는 방식의 성능 평가를 위해 계층적 어텐션 적용 전과 후의 Rouge-score 를 비교한 결과 제안하는 방식이 효과적인 성능을 보임을 확인할 수 있었다.

### I. 서론

정보의 폭발적인 증가로 인해 방대한 양의 텍스트 데이터들이 생성되고 있으며, 이로 인해 시간을 절약하고 핵심 내용을 강조하기 위한 기술의 중요성이 부각되고 있다. 이에 따라 자동으로 핵심 내용만 요약하는 텍스트 요약 기술은 뉴스나 기사의 요약, 학술 논문 및 수업 내용의 요약 등 여러 분야에서 빠르게 적용되고 있다.

텍스트 요약은 원문에서 얻은 문장이나 단어를 추출하여 생성하는 추출적 요약(extractive summarization)과 원문에 없는 새로운 단어와 문장으로 생성하는 추상적 요약(abstractive summarization)으로 나눌 수 있다[1]. 추출적 요약의 예시로는 머신러닝 알고리즘인 TextRank, 추상적 요약의 예시로는 페이스북의 BART 가 있다. 요약문을 생성하기 위해서는 광범위한 자연어 처리가 필요하기 때문에 과거에는 추출 요약 연구가 주를 이루었지만, 딥러닝 기술이 발전함에 따라 현재 추상 요약에 관한 연구가 활발히 진행되고 있다. 언어 모델은 인간의 언어를 학습하는 과정에서 텍스트의 생성, 요약하는 목적을 가진 기술이다. 이러한 언어모델 중 BART 는 높은 수준의 정확도를 보인다.

기존 Seq2Seq 구조의 자연어 처리 모델은 텍스트 입력 단계에서 인코더 벡터의 정보손실 문제가 있다. 이러한 문제를 해결하기 위해 입력에 따른 출력 정확도의 보정이 가능한 어텐션 기법이 등장하였다. 어텐션 기법은 NLP, 음성인식 분야에서 다양하게 이용되고 있다[2][3][4].

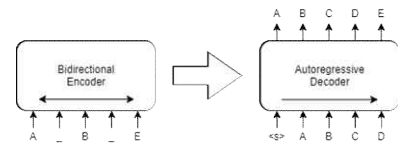
본 논문에서는 기존 BART 모델에 텍스트 처리에서 높은 성능을 보이는 계층적 어텐션 층을 쌓은 미세조정을 통해 요약문의 정확도와 모델의 학습 속도를 향상한다.

### II. 관련연구

#### 2.1 BART(Bidirectional Auto-Regressive Transformer) 모델

BART 모델은 페이스북에서 제안한 자연어 처리 모델로 Seq2Seq 의 트랜스포머 구조를 따른다. 인코더와 디코더를 통해 양방향 학습을 진행하는 추상적 요약 모델이다. 인코더에서 임의의 노이즈 제거 함수를 사용해

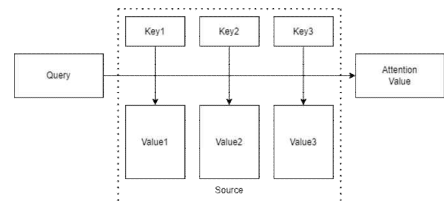
원문 텍스트를 손상시킨 후 디코더에서 원문 텍스트를 재구성하는 방식으로 학습을 진행한다[5]. <그림 1>은 BART 의 구조이다. BART 모델의 인코더는 텍스트 추출에 적합한 BERT 를 기반으로 구성되고, 디코더는 텍스트 생성에 적합한 GPT 모델을 기반으로 구성된다[5].



<그림 1> BART Architecture[5]

#### 2.2 계층적 어텐션(Hierarchical attention)

어텐션 메커니즘은 기존 RNN 에 기반한 Seq2Seq 모델의 인코더 벡터의 정보 손실 문제로 인해 고안된 기법이다. 어텐션 메커니즘은 <그림 2>와 같은 구조로 이루어진다.



<그림 2> Attention Mechanism[6]

계층적 어텐션은 문서처리에 대해 기존의 어텐션 메커니즘을 두 계층으로 나눈 구조를 띤다. 첫번째 계층은, 문서는 문장으로 이루어진 계층적 구조를 갖기 때문에, 문장 레벨의 어텐션 메커니즘을 가진다. 두번째 계층은, 단어의 의미에 따라 문장의 문맥이 달라지기 때문에, 단어 레벨의 어텐션 메커니즘을 가진다. 두 계층의 어텐션은 텍스트 데이터 분류를 결정할 때 단어와 문장의 기여도를 직관적으로 제공한다[7].

### III. 계층적 어텐션을 활용한 BART 미세조정

#### 3.1 미세조정 방식

본 논문에서는 BART 모델의 성능 향상을 위해 기존의 사전학습 모델을 그대로 이용하는 방식으로 미세조정을 진행한다. 텍스트 요약 성능 향상을 위해 텍스트 처리에 적합한 계층적 어텐션을 BART 모델의 연결 계층에 추가하는 방식으로 미세조정을 진행한다.

### 3.2 실험 및 결과 분석

실험 데이터셋으로는 Xsum 데이터셋을 사용하였다. 안하는 미세조정 방식의 성능평가를 위해 요약 과제 평가에 널리 사용되는 Rouge-score 를 사용하였다. 모델 최적화 방법은 ADAM 을 사용하였다.

<표 1>은 미세조정 전 후의 Rouge-score 를 보여준다. 계층적 어텐션을 사용한 경우 모든 지표에서 약간 높은 성능을 보였다. 이는 계층적 어텐션이 문서의 구조를 잘 이해하고, 이에 따라 더 정확하고 포괄적인 요약을 생성할 수 있음을 의미한다.

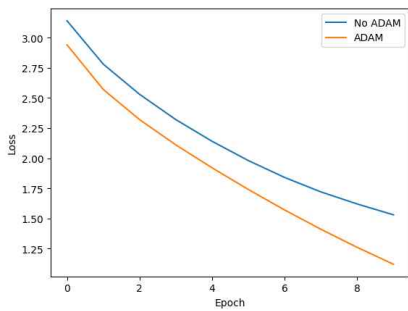
Metric	BART	Finetuned-BART
sssssss	0.394	0.377
Rouge-2	0.314	0.302
Rouge-L	0.364	0.351

<표 1> Comparison before and after applying hierarchical attention

<표 2>는 미세조정된 모델에 최적화 알고리즘인 ADAM 을 적용했을 때의 Rouge-score 를 보여준다. 최적화를 진행했을 때 최적화 전 보다 더 높은 Rouge-score 를 보였고, 학습 속도 또한 약 12% 향상됐음을 확인할 수 있었다. <표 3>은 ADAM 사용 여부에 따른 속도를 비교하기 위한 그래프이다.

Metric	ADAM	No ADAM
Rouge-1	0.403	0.377
Rouge-2	0.323	0.302
Rouge-L	0.373	0.351

<표 2> Comparison before and after optimization with ADAM



<표 3> Comparison of speed before and after using Adam

## IV. 결론

본 논문에서는 텍스트 요약 모델의 학습 속도 향상과 기존 BART 모델의 요약 정확도 향상을 위해 텍스트 처리에 뛰어난 계층적 어텐션 층을 이용한 미세조정 방식을 제안한다.

제안하는 미세조정 방식의 성능 평가를 위해 BART 모델과 계층적 어텐션의 결합과 최적화 알고리즘인 ADAM 을 사용해 모델을 최적화하여 실험을 진행하였다. 계층적 어텐션을 추가하기 전과 추가한 후의 Rouge-score 를 비교한 결과 계층적 어텐션을 적용한 경우 더 높은 정확도의 텍스트 요약 성능을 확인할 수 있었다. 학습속도 향상을 위해 최적화 알고리즘을 적용한 결과 적용 전보다 약 12% 빠른 학습속도를 확인할 수 있었다.

향후 제안하는 미세조정 방식을 BART 모델 외의 다른 모델에도 적용한 실험을 통해 계층적 어텐션을 활용한 미세조정 방식의 일반성 검증이 필요하다.

## ACKNOWLEDGMENT

본 연구는 정부(과학기술통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF. 2021R1A4A2001810)

## 참 고 문 헌

- [1]Widyassari, Adhika Pramita, Supriadi Rustad, Guruh Fajar Shidik, Edi Noersasongko, Abdul Syukur, Affandy Affandy, et al. "Review of Automatic Text Summarization Techniques & Methods." Journal of King Saud University-Computer and Information Sciences 34, no. 4 (2022): 1029- 1046.
- [2]Bahdanau, D., Cho, K. H., and Bengio, Y. "Neural Machine Translation by Jointly Learning to Align and Translate." arXiv preprint arXiv:1409.0473 (2015).
- [3]Galassi, A., Lippi, M., and Torroni, P. "Attention in Natural Language Processing." IEEE Transactions on Neural Networks and Learning Systems 32, no. 10 (2020): 4291-2308.
- [4]Cho, K., Courville, A., and Bengio, Y. "Describing Multimedia Content Using Attention-Based Encoder-Decoder Networks." IEEE Transactions on Multimedia 17, no. 11 (2015): 1875-1886.
- [5]Lewis, Mike, Liu, Yinhan, Goyal, Naman, Ghazvininejad, Marjan, Mohamed, Abdelrahman, Levy, Omer, Stoyanov, Veselin, and Zettlemoyer, Luke. "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension." In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 7871- 7880. Online: Association for Computational Linguistics, 2020.
- [6]유원준. "딥 러닝을 이용한 자연어 처리 입문."
- [7]Yang, Zichao, Yang, Diyi, Dyer, Chris, He, Xiaodong, Smola, Alex, and Hovy, Eduard. "Hierarchical Attention Networks for Document Classification." In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1480- 1489. San Diego, California: Association for Computational Linguistics, 2016.
- [8]Young-Pill Ahn, and Hyun-Jun Park. "Fine-tuning of Attention-based BART Model for Text Summarization." Journal of the Korea Institute of Information and Communication Engineering 26, no. 12 (2022): 1769-1776.