키워드 추출을 활용한 향상된 추상적 대화문 요약 기법

유종재, 이환희, 정교민*
LG 전자, 서울대학교

chongjae.yoo@lge.com, wanted1007@snu.ac.kr, *kjung@snu.ac.kr

Improving Abstractive Dialogue Summarization Using Keyword Extraction

Yoo Chongjae, Lee Hwanhee, *Jung Kyomin
LG Electronics, Seoul National Univ.

요 약

본 논문은 회의 등 여러 화자가 다양한 주제로 대화하는 상황을 짧은 구절을 통해 요약하는 대화문 요약 태스크에 대해 다룬다. 본 논문은 대화문 요약 태스크에서 사전 훈련된 키워드 추출 모델을 통해 추출한 키워드를 활용하여 대화문에 추가적인 정보를 줌으로써, 대화문 요약 시스템의 성능이 향상되는 것을 DialogSum 데이터셋에 대해 확인하였다.

Ⅰ. 서 론

추상적 대화문 요약 태스크는 여러 화자가 말하는 특정 대화에 대한 중요한 내용을 포함하는 짧은 구절을 생성하는 것을 목표로 한다. 일반적으로 대화문을 요약할 때, 해당 대화의 주제를 포착하는 것은 구어체 텍스트의 특성으로 인해 어렵다. 본 논문에서는 키워드의 효율적인 사용을 통해 해당 문제를 완화하기 위해 대화의 주제를 쉽게 포착하는 키워드를 활용한 대화 요약 시스템을 제안한다. 본 방법은 먼저 사전 훈련된 키워드 추출기를 사용하여 입력 대화에서 키워드를 추출한다. 이어서 KADS 는 사전 훈련된 키워드 추출기를 사용하여 대화의 키워드 정보를 Transformer 기반 대화 시스템에 효율적으로 활용한다. DialogSum 벤치마크 데이터셋에 대해 수행한 실험은 제안된 방법이 기준 시스템을 능가한다는 것을 보여준다. 또한, 우리는 제안한 키워드를 활용한 대화 요약 시스템이 훈련 데이터의 수가 매우 제한적인 조건에서 상대적으로 더욱 고성능을 보이는 것을 확인하였다.

Ⅱ. 본론

2.1 키워드 활용 대화 요약 시스템의 구조

본논문에서는 주어진 대화에 대해 짧은 요약 문을 생성하는 것을 목표로 한다. 우리는 대화 요약 시스템을 개발하기 위해 BART[1]와 같은 사전 훈련된 언어 모델 기반의 seq2seq 모델을 훈련시킨다. 그러나, 대화문만 입력으로 사용하는 일반적인 요약 시스템과 다르게 각 발화에서 키워드를 추출하여 대화 요약 시스템의 입력에 추가하여 키워드 활용 대화 요약 시스템을 구축한다. 우리가 제안하는 키워드 활용 대화 요약 시스템 Keyword-aware Dialog Summarizer (KADS)의 전반적인 흐름은 그림 1 에 설명되어 있다. 우리 시스템은 키워드 추출기와 키워드를 활용한 요약문 생성 모델로 구성된다. 우리는 키워드 추출기를 활용하여 얻은 정보를 토대로 입력을 변경하고 사전 훈련된 언어 모델을 사용하여 개선된 요약을 제안한다.

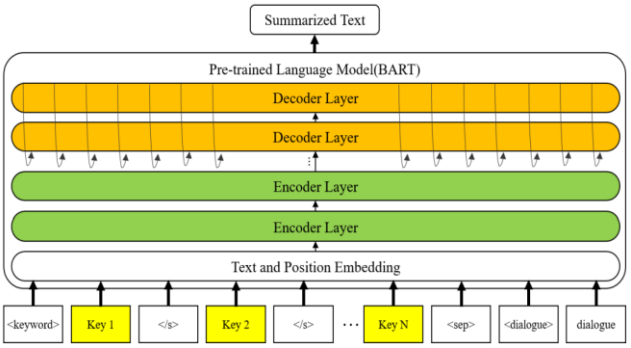

그림 1. 키워드 활용 요약 시스템 (KADS)의 전체 구조

키워드 활용 대화 요약 시스템을 개발하기 위해서는 적절한 키워드 추출기를 선택해야 한다. 따라서, 우리는 키워드 활용 요약 시스템의 주요 구성 요소인 키워드 추출기의 성능을 조사하여, 사용이 간편하고 성능이 우수한 KeyBERT [2]를 활용하였다. 그리고 가장 좋은 키워드 추출기를 통해 키워드를 추출하여 요약문 생성기의 입력에 키워드 정보를 추가하고 사전 훈련된 언어 모델을 사용하여 개선된 요약을 제안한다. KADS 는 그림 1 과 같이 KeyBERT 에서 추출된 키워드들을, 키워드 별로 스페셜 토큰 </s>를 통해 구분한다. 그리고 키워드들 뒤에 스페셜 토큰 <sep>을 넣어 키워드와 대화문을 구분할 수 있게 한 뒤 대화문을 입력에 추가하여 전체적인 모델의 입력을 만든다. 그리고 우리는 이 입력을 사전 훈련된 BART 에 통과시켜 키워드 정보를 활용해 대화문에 대한 요약문을 생성하도록 학습한다.

2.2. 키워드 활용 대화 요약 시스템의 성능 평가

우리는 제안한 요약 시스템의 성능을 평가하고자 대표적인 공개된 대화 요약 벤치마크 데이터셋 중 하나인 DialogSum[3]을 사용했다. 또한 방법의 관점에서 DialogSum 데이터셋은 긴 대화 데이터를

포함하여 심층 신경망 기반 방법을 사용한 대화 요약 연구를 용이하게 할 수 있다. 특히나 대부분의 대화 데이터셋은 채팅 대화를 기반으로 한 대화 또는 일상 대화가 부족한 경우가 많지만, DialogSum 은 실제 회의 등의 대화를 포함하고 있다.

우리는 제안한 요약 시스템의 대화 요약 성능을 표 1 에 첨부하였다. 우리는 대화 요약 시스템의 성능 평가를 위해 ROUGE[4]와 BERTScore[5]를 활용하였다. 우리는 BART 를 Baseline 시스템으로 사용했고 널리 쓰이는 또 다른 모델 중 하나인 T5 [6]를 활용한 실험도 수행했다. 우리는 우리가 제안한 KADS 가 모든 경우에서 기준 시스템보다 개선된 성능을 보이는 것을 관찰했다. 표 1 에서 보는 바와 같이, BART-large 에서는 베이스라인과 KADS 에 비해 약 2.7%, T5-large 에서는 7.6%, T5-large 에서는 9.4%의 성능이 각각 향상되었다. 그리고 그 결과는 Baseline 모델의 성능이 낮을수록 제안된 키워드 활용 방법을 통해 개선 효과가 크다는 것을 보여준다.

Model	ROUGE-1	ROUGE-2	ROUGE-L
Baseline			
BART-base	44.8874	19.6440	37.0678
BART-large	46.1996	21.0814	38.8086
T5-base	41.5242	16.4631	33.6254
T5-large	42.1325	17.3326	34.4822
KADS			
BART-base	45.9874	20.9440	38.1678
BART-large	**47.2237**	**22.1353**	**39.8665**
T5-base	44.2605	18.8368	36.2043
T5-large	45.2232	18.9618	37.7235

표 1. DialogSum 데이터셋에 대한 성능 비교

또한, 일반적으로 대화 요약을 위한 데이터의 수는 문서 요약 데이터셋에 비해 상대적으로 적다. 따라서 대화 요약 시스템을 학습하는 것은 종종 어렵다. 따라서 더 적은 수의 데이터만 존재하는 상황에서 키워드를 활용하는 방식이 효율적인지 확인할 필요가 있다. 우리는 데이터가 적은 시나리오에서 제안된 시스템의 성능을 검증하기 위해, 우리는 학습 데이터셋의 절반만 사용하여 시스템을 훈련하고 그 결과를 표 2 에 제시하였. 그리고 우리는 제안된 키워드 활용 요약 시스템이 이러한 적은 데이터셋의 조건에서 Baseline 시스템과 비교하여 특히나 더 효과적이라는 것을 발견했다.

Model	ROUGE-1	ROUGE-2	ROUGE-L
Baseline			
Full	46.1996	21.0814	38.8086
Half	44.0717	20.1916	36.6705
KADS			
Full	**47.2237**	**22.1353**	**39.8665**
Half	45.2720	20.5406	38.1776

표 2. 데이터가 적은 상황에서의 시스템의 성능 비교

Ⅲ. 결론

본논문에서는 대화 요약 시스템의 성능을 향상시키기 위해 키워드 정보를 효율적으로 활용하는 대화 요약 시스템인 KADS 를 제안했다. 우리는 키워드 추출기 성능이 대화 요약 결과에 상당한 영향을 미칠 수 있음을 보여주었다. 널리 사용되는 대화 요약 데이터셋에 대한 실험 결과는 제안된 키워드 활용 대화 요약이 기준 시스템보다 향상된 것으로 나타났다. 향후 우수한 키워드 추출기가 제안될 경우 KADS 의 성능이 추가로 향상될 수 있다고 본다. 또한, 우리는 제안된 시스템이 자원이 적은 조건에서 특히 효율적이라는 것을 보여주었다.

.

ACKNOWLEDGMENT

이 연구는 서울대학교 자동화연구소와 2022 년도 BK21 FOUR 정보기술 미래인재 교육연구단 지원을 받아 수행된 연구결과임.

참 고 문 헌

[1] M. Lewis et al.,, 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, 2020, pp. 7871- 7880.

[2] M. Grootendorst, Keybert: Minimal keyword extraction with bert. (2020). doi:10.5281/zenodo.4461265. URL https://doi.org/10.5281/zenodo.4461265

[3] Y. Chen et al., Dialogsumm: A real-life scenario dialogue summarization dataset, CoRR abs/2105.06762. arXiv:2105.06762.

[4] C.-Y. Lin, ROUGE: A package for automatic evaluation of sum- maries, in: Text Summarization Branches Out, Association for Computational Linguistics, Barcelona, Spain, 2004, pp. 74- 81.

[5] T. Zhang et al., Bertscore: Evaluating text generation with BERT, CoRR abs/1904.09675. arXiv:1904.09675.

[6] C. Raffel et al.,, Exploring the limits of transfer learning with a unified text-to-text transformer (2019). doi:10.48550/ARXIV.1910.10683.