연구과제명: 화자분리와 대본을 이용한 뉴스 음성 요약

|  |  |  |  |
| --- | --- | --- | --- |
| 화자 별 내용 요약을 생성할 방법에 대한 조사를 진행하였다. 전체 요약의 경우 기존 방송 대본 데이터에서 요약문을 산출하는 데이터를 모델에 학습시켜 전체 요약문을 제공할 예정이지만 화자 별 요약의 경우 해당하는 유형의 데이터가 없기 때문에 같은 방법으로 결과를 산출하기 어렵다. 이 경우 gpt모델의 few-shot learning을 이용할 예정이다. Few-shot learning은 기존 학습된 모델에 적은 수의 데이터를 학습시켜 결과를 산출하는 방식의 meta-learning인데, 데이터를 직접 학습하는 것이 아닌 데이터에서 결과를 산출하는 방법을 학습한다는 특징이 있다. 해당 참고자료에서는 gpt-3의 few-shot learning 성능이 기존의 sota모델보다 뛰어난 성능을 가지고 있다는 점을 확인하고 있다. 고로 우리 조는 기존 대본에서 화자 별 요약을 만드는 요약문 데이터를 직접 만든 뒤 해당 데이터를 기반으로 모델을 학습할 예정이다. 화자 별 요약문 중 전체 데이터의 일부를 test set으로 구분한 뒤 training set으로 학습한 모델을 통해 화자 별 요약을 산출하고, 이를 test set과 비교하여 성능을 평가할 예정이다. 평가 방식은 전체 요약문과 같은 방법인 rouge, meteor를 사용할 예정이다. 전체 요약문의 경우 생성 모델에 데이터를 학습하여 fine-tuning을 해야 하는데 gpt3이후의 gpt모델은 모델을 공개하지 않기에 gpt-2를 사용하여 데이터를 학습시켜 결과를 산출해야 할 것이다.  다음은 지난 주차 대본 생성시스템의 단점을 보완한 시스템의 흐름도이다. 지난 시스템과의 차이점은 분리된 segment가 연속되어 stt가 수행되는 경우 해당 두 음성을 이어 붙이는 기능이 추가된 점과, stt수행 이전 전체 음성에서 무음 부분을 제거하는 부분이 추가된 점이다. 첫번째 기능의 경우 stt모델은 음성이 길수록 모델의 정확도가 정확하다는 기존 연구 결과에 기반한 흐름도 수정이고, 두번째 기능의 경우 실험적으로 무음 구간에서 정확하지 않은 문장이 산출되기에 추가한 기능이다. 이를 반영한 흐름도는 다음과 같다.  해당 흐름도를 요약하면 segment 내에 마지막 음성은 바로 stt를 실행하지 않고 queue에 저장하고, segment내에 첫번쨰 음성의 경우 queue에 저장된 음성과 화자 인식을 수행하여 stt를 할지, queue에 삽입할 지를 결정한다. 이를 통해 위 기능을 구현하였다. 현재 90%가량 구현하였는데, 음성에서 무음 구간을 제거하는 기능을 추가해야 한다. 이번 주 까지 위 시스템을 완성하고, 다음 주부터는 계획도에 따라 화자 별 요약 데이터를 만들고, 대본을 통해 전체 요약을 내는 모델을 학습해야 한다. | | | |
| **작성자** | **일자 2023-05-12** | **확인자** | **일자 2023-05-12** |
| **서명 신원철** | **서명 김유성** |