



연구논문/작품 제안서

2022년도 제 2학기

논문/작품	○논문() ○작품(V) ※ 해당란에 체크
제목	온라인 회의 요약 시스템
GitHub URL	https://github.com/skku-skku
팀원명단	엄 계 현 엠티현 (학번: 2019315479) 이 예 진 이 (원)진 (학번: 2019315580)

2022 년 8 월 29 일
지도교수 : 이 진 규 서명

1. 과제의 필요성

시간을 돈으로 살 수 있다면 바쁜 현대인들은 기꺼이 지불하고자 할 것입니다. 또한, 코로나19의 영향으로 학교부터 회사까지 온라인 화상회의 사용빈도가 증가했습니다. 따라서, 비즈니스 세계에서 효율적인 시간 관리와 회의의 핵심 파악은 중요한 가치입니다. 또한, 긴 시간 회의를 진행했지만 끝나고나면 모든 회의 참가자들이 핵심을 이해하고 끝나는 경우가 많지 않습니다. 분명 다양한 사람들의 의견이 오고가며 생산적인 대화가 진행되었겠지만 지나간 회의를 정리하지 않으면 오래 기억할 수 없습니다.

따라서, 온라인 회의 요약 시스템은 녹화 혹은 녹음된 음성 데이터를 문서화 작업을 진행 후 요약/QA 과정을 진행해줍니다. 이를 통해 많고 긴 회의가 끝난 후에도 요약본을 통해 전체적인 회의 흐름을 빠르게 복기할 수 있습니다. 또한, QA를 통해 "8월 20일 회의 주제가 뭐였지?"라는 질문에 "졸업 작품 주제 선정"이라는 답변을 받을 수 있어 다음 회의 방향성을 잡는 길라잡이가 되어줄 수 있습니다.



그림 1 네이버웍스 영상회의 증가^[1]



그림 2 코로나19 재택근무 현황^[2]

2022년 상반기 한국 내 비즈니스 앱 상위 차트



그림 3 "줌" 누적 다운로드 10억 돌파^[3]

2. 선행연구 및 기술현황

2-1. 온라인 회의 요약 선행연구

1. 클로바 노트[4]

2022년에 네이버 클로바 노트에 AI를 기반으로한 요약 기능이 추가되었습니다. AI가 대화 주제별로 구간을 나누고 핵심을 자동으로 요약해 회의록이나 필기 내용을 한 눈에 빠르게 확인할 수 있다고 합니다.

2. AI기반 회의 녹취록 요약 경진대회[5]

작년에 LG AI Research에서 주최한 회의 녹취록 요약 경진대회입니다. 문서화된 회의 녹취록에서 핵심 내용을 요약하는 생성요약 AI모델 개발이 목적이었습니다. 저희는 이 경진대회에서 제공해준 녹취록 데이터를 이용해 개발을 진행할 예정입니다.

2-2. 음성인식 모델

1.. 구글 클라우드 음성 API (Speech-to-text)[6]

최고의 Google AI 연구 및 기술로 지원되는 API를 사용하여 음성을 텍스트로 정확하게 변환할 수 있습니다.

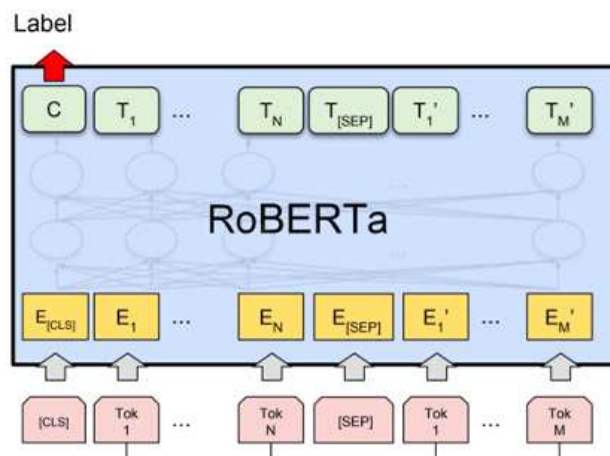
2.. 파이썬 STT 라이브러리 사용[7]

사람이 말하는 음성 언어를 컴퓨터가 해석해 그 내용을 문자 데이터로 전환하는 처리에 사용되는 라이브러리입니다.

2-3. 자연어 처리 모델

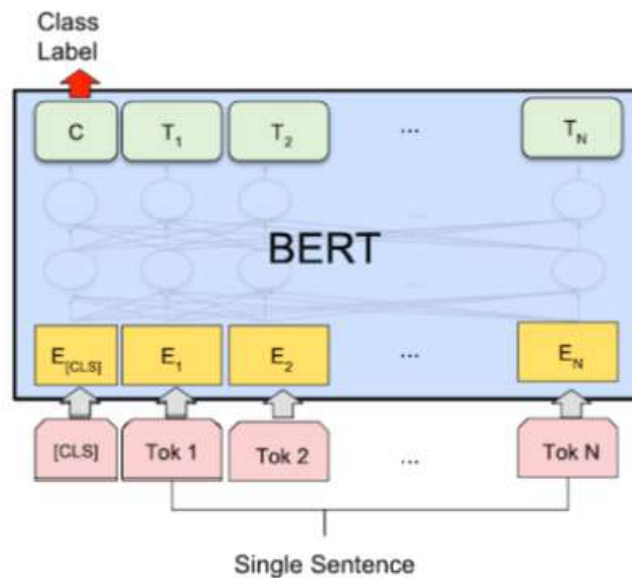
1. RoBERTa(CDQA)[8]

더 큰 배치사이즈로 좀 더 많은 데이터를 학습시킨모델입니다. 총 160GB의 데이터가 학습을 위해 쓰였으며, 최대한 많은 데이터를 수집하여 비교에 적합한 데이터 양을 찾아냈습니다. 기존 BERT보다 더 우월한 성능을 보였고, BERT 이후에 나온 후속 모델(XLNet 등) 보다도 더 좋은 성능을 보여주었습니다. 이전에 간과되었던 설계의 중요성 또한 대두되었습니다.



2. BERT^[9]

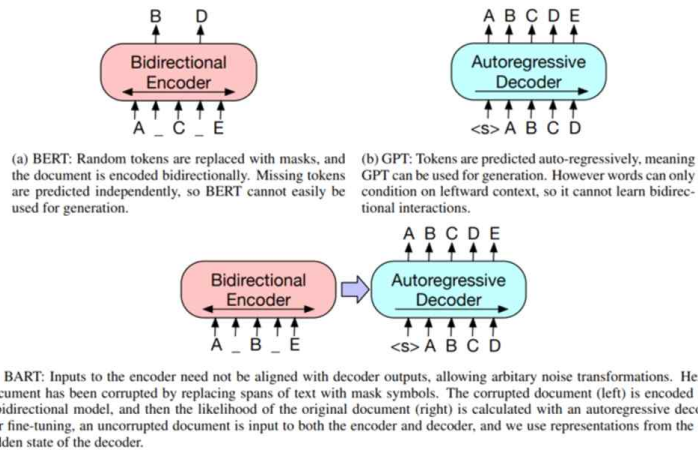
구글의 Devlin(2018)이 제안한 BERT는 사전 학습된 대용량의 레이블링 되지 않는(unlabeled) 데이터를 이용하여 언어 모델(Language Model)을 학습하고 이를 토대로 특정 작업(문서 분류, 질의응답, 번역 등)을 위한 신경망을 추가 하는 전이 학습 방법입니다.



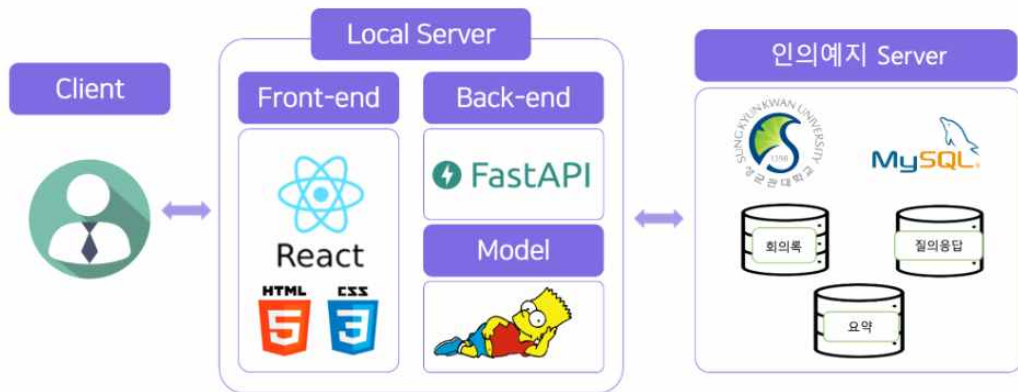
3. BART(요약)^[10]

BART는 corrupting documents를 reconstruction하는 것으로 pre-training 되었습니다. 기존의 pre-training 방식이 특정한 하나의 방식의 corrupting 을 사용했다면 BART는 다섯가지 방법으로 pre-training이 진행되었습니다.

- Token Masking : BERT에서 사용 한 것 처럼 랜덤한 토큰을 <MASK> 토큰으로 바꾸는 방식 입니다.
- Token Deletion : Random 한 토큰을 인풋에서 삭제하는 방법입니다. 이 방법은 Token masking 방식에서, 어디가 사라졌는지도 맞춰야 하는 난이도가 더 높은 방식이라고 할 수 있습니다.
- Text Infilling : Token masking 에서는 하나의 단어만 마스킹 하지만, text infilling 방식은 여러개의 토큰을 한번에 <MASK> 토큰으로 대체합니다.
- Sentence Permutation : 문장에 나온 단어의 순서 맞추기 방법입니다. 문장의 단어들이 Shuffle되어서 input이 되면 디코더는 원래 단어의 순서대로 문장을 출력해야 합니다.
- Document Rotation : 원래 문장의 단어 순서가 ABCDE 였다면, DEABC 로 순서가 회전된 문장이 주어집니다. 디코더는 다시 원래 형태인 ABCDE로 출력해야 하는 방식으로 학습이 진행됩니다.



3. 작품/논문 전체 진행계획 및 구성



3-1. 음성인식

① 데이터셋

데이콘 회의 녹취록 요약 경진대회의 의회 데이터를 사용해 구축합니다. 구축된 데이터를 기반으로 요약모델의 학습이 진행됩니다.

② 학습

한국어 음성 인식 라이브러리인 파이썬 STT를 사용해 음성 데이터를 문자로 바꾸어 문서화를 진행합니다. 문서화된 데이터가 문서 요약과 QA의 데이터로 활용됩니다.

3-2. KoBART 모델 문서 요약 / RoBERTa 모델 QA

① 학습

구축된 데이터셋을 기반으로 pre-trained KoBART모델을 fine-tuning 진행합니다. Pytorch 프레임워크를 사용할 예정입니다.

② QA

구축된 데이터셋을 기반으로 pre-trained RoBERTa모델을 fine-tuning 진행합니다. Pytorch 프레임워크를 사용할 예정입니다.

3-3. 웹 서비스

① Frontend

React를 이용하여 모바일 웹페이지를 개발할 예정입니다.

웹 서비스는 음성 녹음 업로드, 요약, QA 기능에 대한 인터페이스를 제공합니다.

② Backend

FastAPI를 이용하여 백엔드 API를 개발할 예정입니다.

MySQL 데이터베이스와 연동하여 녹취록과, 회의록 파일을 저장합니다. 회의록을 바탕으로 요약된 회의록도 저장되며, 회의록 질문과 답변목록에 데이터베이스에 저장됩니다. 서버는 인의예지 서버를 이용하여 개발할 예정입니다.

③ 기능

사용자가 파일에 저장된 음성 회의를 웹에 추가합니다.

추가된 음성 데이터는 일정 시간동안 인공지능 모델을 통해 학습하고 문서화시켜 사용자에게 정보를 전달할 준비를 마칩니다.

사용자가 특정 회의의 궁금한 점을 입력합니다. (예시. 8월 29일 회의 주제가 뭐였지?) 요약된 데이터로부터 사용자의 질문에 해당하는 답변을 제공해줍니다. (예시: 졸업작품 제안서)

4. 기대효과 및 개선방향

1. 기대효과

- 1.1. 수많은 회의를 진행하는 직장인들에게 빠르게 필요한 회의의 내용을 복기할 수 있습니다.
- 1.2. 음성 회의 요약물 통해 다음 회의의 방향성을 보다 확실하고 빠르게 잡을 수 있습니다.
- 1.3. 회의 내용을 바탕으로 검색기능이 가능해집니다.

2. 확장성

- 2.1. 회의에서의 사용 뿐만아니라 학생들의 공부에서도 사용이 가능할 것입니다. CDQA를 사용하여 강의자료를 토대로 검색이 가능할 것입니다.
- 2.2. 웹 서비스 뿐만 아니라 애플리케이션으로도 개발을 하여 접근성을 높일 수 있습니다.
- 2.3. 자체 어플 내에서 회의가 가능하게 개발할 수 있습니다.
- 2.4. 실시간으로 요약가능한 기능을 추가하여 효율성을 더 높일 수 있습니다.

이 사항 중 실현 가능한 기능(데이터셋 확보가 가능하거나, 활용 가능한 정도의 정확도가 나오는 기능)들을 최대한 구현할 예정입니다.

3. 차별성

기존의 회의요약 시스템은 회의를 요약만 해주는 서비스인 반면에, 우리의 회의요약 서비스는 요약뿐만 아니라 사용자가 회의 내용에 대하여 검색을 할 수 있는 Qustion and Answering 서비스를 제공합니다. 이를 통해 빠른 시간내에 이전 회의 내용을 보다 더 효율적으로 복기할 수 있습니다.

5. 기타

5-1. 팀원간의 역할분담

엄계현

백엔드/음성 인식

이예진

프론트엔드/요약 모델

공동 작업

1. QA 모델 학습

확보한 음성 데이터를 문서화 및 요약을 진행한 후 QA 모델로 학습시킵니다.

2. 데이터셋 전처리

원활한 학습을 위하여 구축한 데이터의 전처리를 진행합니다.

6. 참고문헌

- [1] <https://naver.worksmobile.com/blog/220325/>
- [2] <https://www.epnc.co.kr/news/articleView.html?idxno=201393>
- [3] <https://www.news1.kr/articles/?4756107>
- [4] <https://clovanote.naver.com/>
- [5] <https://dacon.io/competitions/official/235813/overview/description>
- [6] <https://cloud.google.com/speech-to-text?hl=ko>
- [7] <https://api.ncloud-docs.com/beta/docs/ai-naver-clovaspeechrecognition-stt>
- [8] <https://arxiv.org/abs/1907.11692>
- [9] <https://arxiv.org/abs/1910.13461>
- [10] <https://arxiv.org/abs/1910.13461>