

STT(Speech-To-Text)와 ChatGPT 를 활용한 강의 요약 애플리케이션

김진웅¹, 금보성¹, 김태국²

¹부경대학교 컴퓨터·인공지능공학부 학부생

²부경대학교 컴퓨터·인공지능공학부 교수

baseball0020@gmail.com, rmaqhtjd8364@naver.com, king@pknu.ac.kr

A Lecture Summarization Application Using STT (Speech-To-Text) and ChatGPT

Jin-Woong Kim¹, Bo-Sung Geum¹, Tae-Kook Kim²

^{1,2}School of Computer and Artificial Intelligence Engineering,
Pukyong National University

요 약

COVID-19 가 사실상 종식됨에 따라 대학 강의가 비대면 온라인 강의에서 대면 강의로 전환되었다. 온라인 강의에서는 다시 보기를 통한 복습이 가능했지만, 대면강의에서는 녹음을 통해서 이를 대체하고 있다. 하지만 다시 보기와 녹음본은 원하는 부분을 찾거나 내용을 요약하는데 있어서 시간이 오래 걸리고 불편하다. 본 논문에서는 강의 내용을 STT(Speech-to-Text) 기술을 활용하여 텍스트로 변환하고 ChatGPT(Chat-Generative Pre-trained Transformer)로 요약하는 애플리케이션을 제안한다.

1. 서론

COVID-19 이후의 강의들이 대부분 비대면에서 대면으로 진행됨에 따라, 강의 내용을 순간순간에 기록하는 것이 더욱 중요해졌다. 동영상 강의에서는 다시 보기가 가능했지만, 대면 강의에서는 강의를 녹음하는 방법으로 이를 대체하고 있다. 하지만 길이가 긴 녹음본에서 원하는 부분을 찾는 것은 최악의 경우 그 녹음본의 길이만큼 시간이 소요된다. 이는 학습효율의 저하로 이어진다.

이러한 문제를 해결하기 위해 본 논문에서는 강의 음성을 텍스트로 변환하고, 이를 요약본으로 만들어 주는 애플리케이션을 구현하였다. 이를 통하여 학습과 복습에 도움을 줄 것으로 기대된다.

2. 오픈 STT API 비교

제안한 애플리케이션은 음성을 텍스트로 변환하기 위한 기술인 STT(Speech-to-Text)와 텍스트 요약용 API(Application Programming Interface)를 사용하여 처리한다. 한국어의 음성을 텍스트로 변환하는 것은 강의 요약 애플리케이션에서 제일 중요한 부분 중의 하나이다. <표 1>은 상용으로 공개되어 있는 SST 오픈 API 서비스 목록과 특징을 나타낸다.

<표 1> SST 오픈 API 목록

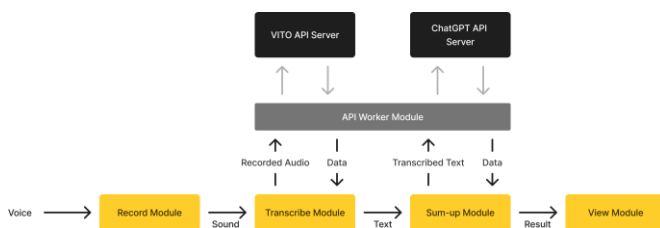
서비스	한국어 지원 여부	SDK 지원	가격 정책
Apple Speech[2]	Server, on-device 모두 지원	Apple 플랫폼 운영체제만 지원	- Server: 1분 단위 세션만 지원 - On-device: 무료
Google Cloud Speech-to-Text[3]	한국어 지원	- Rest API 제공 - Python 외 7개 언어 SDK 제공	- 월 60분 무료 - 초과 시, 분당 \$0.024

VITO Speech API[4]	한국어 지원	- Rest API 제공	- 월 100 시간 무료 - 100 시간 초과시 중단
AWS Transcribe[5]	한국어 지원	- Python 외 6개 언어 SDK 제공	- 월 60 분 무료

이외에도 한국어를 지원하는 STT 오픈 API 로는 Naver Clova, Microsoft Azure, 그리고 Kakao Speech 등이 있다[1]. 각 오픈 API 들을 비교하였을 때, VITO Speech API 는 월 100 시간이라는 사용량을 무료로 제공하고 한국어도 제공한다. 따라서 본 연구에서는 VITO Speech API 를 사용하여 강의 음성을 텍스트로 변환하는 SST 를 구현하였다.

3. 제안한 강의 요약 애플리케이션

제안한 강의 요약 애플리케이션은 각 기능을 모듈화하여 구현하였다. 그림 1 은 제안한 강의 요약 애플리케이션의 강의 요약 흐름도를 나타낸다. 모듈은 Record Module, Transcribe Module, Summarize Module, View Module 로 나눌 수 있다.



(그림 1) 제안한 강의 요약 애플리케이션 흐름도

3.1 Record Module

애플리케이션 서비스의 엔트리 포인트이자 음성을 녹음하여 음성파일로 만드는 과정이다. 입력으로 음성을 받아서 Apple 의 AVFoundation 프레임워크를 통하여 음성파일로 변환한다.

3.2 Transcribe Module

녹음된 음성파일로부터 음성을 추출하여 텍스트로 변환하는 역할을 담당한다. 음성 파일을 네트워크를 통해 전송하여 VITO API 서버에서 텍스트로 변환한다. 통신 모듈을 통해 입력 받은 음성 파일을 API 서버로부터 음성 텍스트 데이터를 얻어와 최종적으로

문자열 형식의 데이터로 변환하여 내보낸다.

3.3 Summarize Module

Transcribe Module 과 마찬가지로 네트워크 통신을 하는 모듈과 상호작용한다. 입력으로 문자열 형식의 텍스트 데이터를 받아 ChatGPT(Chat-Generative Pre-trained Transformer) API 를 통해 내용을 요약한다. 요약된 내용의 데이터를 받아 문자열 형태의 데이터로 변환하여 모바일 화면의 UI(User Interface)에 나타낼 수 있게 내보낸다.

3.5. View Module

각 모듈들을 통해 얻어온 텍스트 요약본을 UI 를 통해 보여준다.

4. 결론

본 연구에서 STT 와 ChatGPT 를 활용하여 강의 요약 애플리케이션을 구현하였다. 강의 음성 파일을 SST API 를 통해 텍스트로 변환하고, 변환된 텍스트는 ChatGPT API 를 활용하여 강의 내용을 요약하였다. 제안한 강의요약 애플리케이션은 모바일 디바이스를 활용하여 손쉽게 음성 녹음과 강의 요약본을 제공할 수 있다. 따라서 강의 요약본을 통해 학습효과 증진에 기여할 것으로 기대한다. 또한 제안한 애플리케이션은 회의록 요약 등 다양한 분야에도 활용될 수 있다.

참고문헌

- [1] 유현재, 김명화, 박상길, 김광용, “클라우드 기반의 음성인식 오픈 API 의 응용 분야별 한국어 연속음성인식 정확도 비교 분석,” 한국통신학회논문지, Vol.45, No.10, pp.1793-1803, 2020.
- [2] 애플, <https://developer.apple.com/documentation/speech/>
- [3] 구글, <https://cloud.google.com/speech-to-text/>
- [4] VITO, <https://developers.vito.ai/>
- [5] 아마존, <https://aws.amazon.com/ko/transcribe/>