| 3팀 | | NLP를 활용한 회의록 요약 |
| --- | --- | --- |
| 2022.08.30 ~ 2022.10.17 |

o 팀원 : 강동현, 정영현, 방수진, 전아영, 성혜경, 지동주, 조혜리

| 프로젝트 개요(추후 변경 가능) |
| --- |
| 1. 목적 설정 2. 데이터, 모델 탐색 3. 데이터 전처리 4. 모델 성능확인 5. 모델 파인튜닝 6. DB 구축 7. Django 구현 8. 성능평가 |
| 프로젝트 목적(추후 변경 가능) |
| 1. 음성이나 텍스트로 입력이 가능한 일기장 프로그램 구현 2. 입력된 데이터를 전처리하여 감성분석, 요약, 키워드 추출 분석으로   사용자에게 편한 기록과 요약 시스템 제공과 |
| 프로젝트 방법(추후 변경 가능) |
| 1. 목적 설정    1. 음성, 텍스트 데이터의 일기장 어플 구현    2. 일기장 내용요약, 감정분석, 기타 등등 2. 데이터, 모델 탐색    1. 녹음, 동영상(음성) 데이터 저장 (DB 저장)       1. 녹음 >       2. 음성 데이터 >    2. 텍스트 데이터   i. 텍스트 가져오기>   1. 데이터 전처리    1. 전처리    2. 모델 생성    3. 테스트    4. 성능평가    5. 산출물 확인 2. Kobart 모델    1. 모델 생성    2. 테스트    3. 성능평가    4. 산출물 확인 3. Kobert 모델,    1. 모델 생성    2. 테스트    3. 성능평가    4. 산출물 확인 4. Django (어플 형태 구현)    1. 음성 데이터 db연결    2. 웹 디자인 5. 창출되는 서비스    1. 일기장 내용 요약    2. 일기장 내용 감성분석    3. 기타 등등 |
| 기대 효과(활용 또는 확장성) |
|  |

# 궁금한거

멘토링 오시면 지원 받을 수 있는 부분, 작업 중 에러, 충돌시 메모해놓기(에러내용, 컴퓨터 사양 등등)

유튜브 영상을 mp3로 다운받아야 하나?, wav로 가져올 방법은?

wav를 그대로 써야하나, 시계열데이터(변환)로 바꿔서 쓸 수 있나?

→ 실시간 음성 분석과 녹음 파일을 기반으로 분석하는 모델은 다르기 때문에 실시간이 아니라는 것을 강조할 필요가 있음

→ [논문] MFCC와 LPC 특징 추출 방법을 이용한 음성 인식 오류 보정

MFCC와 사용함수 공부

warp-rnnt : 코랩 사용 추천

음성데이터의 경우 발음이 같은 단어로 나타날 수 있는 위험 제거(ex : 탯줄->대출로 인식한경우)

> 유튜브 원본 음성데이터가 들어가서 전처리가 안된 상태임(음성, 잡음 전처리 후 대량 데이터 학습 후 확인 > 중복, 단어빈도 확인하고 필요없는거 제거

→ 오류는 전처리 과정에서 그대로 두고

→ nlp 과정에서 오류 보정 가능 ( 학습 데이터가 정확한 카테고리로 분류되어있을 경우 오류를 보완할 수 있다 – 도메인 특화 )

\*) 고려 사항

→ 말뭉치데이터를 학습데이터로 사용할 경우 일반적으로 모델을 형성 가능하나, 위 오류는 생길 수 있음

→ 도메인 특화로 진행 시 학습데이터를 구하기 어려울 수 있음

용량제한

> (regonizer\_google() 이게 구글꺼 빌려쓰는거라서 제한 걸려있음 > 모델을 직접 만들어서 raw data를 살리자 > KoSpeech 모델(카카오 개발자가 만듦) > 가상환경 실행 해보고 뜯어보기, 수행시간 체크

→ 손실된 데이터 부분을 교집합으로 겹쳐서 txt 데이터로 변환 후 문맥에 따라 교집합된 데이터 제거

영상을 토대로 설명하는 내용이므로 화면상으로만 설명되어 음성데이터에는 남아있지 않는경우

→:[영상시간] 요약된 내용

→ 영상 요약은 난이도가 높음

kospeech 모델 전처리 -> txt 언어 품질이 안좋다? > 모델 전처리 부분 수정

사투리, 축약어같은 경우 처리를

→축약어는 처리 할 수 없음

[질문사항]

* NVIDIA사용을 위한 내장 노트북을 구할수 있는지?

- 코랩으로 돌리고 안되면 모델을 변경

* 구글 -> youtube to mp3 대안툴 고민

-출처-

| 이름 | 내용 | URL |
| --- | --- | --- |
| Ko\_Speech | STT모델 | <https://github.com/sooftware/kospeech> |
| Ko\_Speech | ko\_speech 활용내용1 | <https://mingchin.tistory.com/230?category=986268> |
| Ko\_Speech | ko\_speech 활용내용2 | <https://velog.io/@letgodchan0/%EC%9D%8C%EC%84%B1%EC%9D%B8%EC%8B%9D-%ED%95%9C%EA%B5%AD%EC%96%B4-STT-1> |
|  |  | [파이썬으로 유튜브 영상 다운로드 하기 - 아무튼 워라밸 (hleecaster.com)](https://hleecaster.com/python-pytube/) |
|  |  | [파이썬으로 음악 파일 자르기 — Ainomix](https://ainomix.com/11) |
|  |  | <https://bebutae.tistory.com/17> |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

<감성분석>

영어 - vader (감성어 사전)

* 한국어 감성어 사전이 있는지?
* 임의로 분류한다면 분류기준은 어떻게할지?

한국어 STT

-기술 정의서-

| **rocognize\_google** | | | |
| --- | --- | --- | --- |
| 라이브러리 이름 | 사용방법 | 활용내용 | 출처 |
| pytube | pip install youtube  -from pytube  import YouTube | 유튜브 영상 추출 |  |
| os | pip install os  - import os | os 모듈을 이용하여 음성파일 디렉토리 설정 및 삭제 | <https://cyber0946.tistory.com/60> (os 모듈 속성 정리) |
| pydub | pip install pydub  -from pydub import AudioSegment | 음성 데이터 전처리 |  |
| speech\_recognition | pip install speech\_recognition  -import speech\_recognition as sr | 음성 데이터를 문자 데이터로 전환하는 처리 | <http://aispiration.com/nlp2/nlp-audio-transcribe.html> |
| time | pip install time  -import time | 영상 데이터를 시간별로 나누기 |  |
| math | pip install math  -import math | floor함수 사용 |  |

| **KoSpeech (Deep Speech 2)** | | |
| --- | --- | --- |
| **로드맵** | | |
| 이름 | 설명 | |
| *Deep Speech 2* | CTC 손실이 있는 ASR 작업에서 더 빠르고 정확한 성능  이전 End to End 모델에 비해 성능이 크게 향상 | |
| *Listen, Attend and Spell (LAS)* | Sequence-to-Sequence model 기반 음성 인식 모델  모든 컴포넌트를 End-to-End로 학습 | |
| *RNN-Transducer* | Attention 메커니즘을 사용하지 않는 sequence-to-sequence 모델의 한 형태  RNN-T는 입력 샘플을 지속적으로 처리하고 출력 기호를 스트리밍 | |
| *Speech Transformer* | Transformer는 NLP(자연어 처리) 분야의 강력한 아키텍처 | |
| *Joint CTC-Attention* | Joint CTC-Attention은 LAS 및 Speech Transformer와 함께 훈련 | |
| *Jasper* | 1D 컨볼루션 레이어의 블록으로 구성된 심층 시간 지연 신경망(TDNN)  각 모델이 서로 다른 수의 레이어를 갖는 모델 계열 | |
| *Conformer* | parameter-efficient 방식으로 audio sequence의 local 및 global dependency를 모두 모델링하기 위해 CNN과 Transformer를 결합하는 방법  음성 인식을 위한 Convolution-Augmented Transformer를 제안 | |
| **설치** | | |
| 라이브러리 이름 | 사용방법 | 활용내용 |
| Numpy | import numpy | 행렬/배열 처리 및 연산  난수 생성 |
| Pytorch | import torch | Scientific Computing 패키지  GPU를 이용하기 위한 Numpy 대체재  유연성과 스피드를 제공하는 딥러닝 연구 플랫폼 |
| Pandas | import pandas | 행과 열로 이루어진 데이터 객체 조작  대용량 데이터 처리하는데 편리 |
| Matplotlib | import matplotlib.pyplot | 자료를 차트나 플롯으로 시각화하는 패키지  정형화된 차트나 플롯 외에도 저수준 API를 사용한 다양한 시각화 기능 제공 |
| librosa | import librosa | 음원데이터 분석  파형가공, 변환기능, STFT, mel spectrogram 생성 |
| torchaudio | import torchaudio | 인공신경망을 만드는데 필요한 기본요소를 직관적이고 안정적인 API로 제공 |
| tqdm | from tqdm import tqdm  import time | 프로그램이 내가 의도한대로 돌아가는 중인지,  진행상황과 시간이 궁금할 때 사용 |
| sentencepiece | import sentencepiece | 사전 토큰화 작업없이 전처리하지 않은  데이터(raw data)에 바로 단어 분리 토크나이저를  사용할 수 있는 이점 구현 |
| warp-rnnt |  | CUDA가 warps라는 32개의 병렬 스레드 실행  32개의 스레드가 모두 실행경로에 동의할 때  완전한 효율성 실현  이것이 RNN Transducer 최적화하는데 사용 |
| hydra |  | configuration을 모듈로 세분화하여 관리 가능  미리 정의한 configuration을 hydra 환경에서 사용  실행된 hydra 환경에서의 실행 기록을 저장 |
| Configuration |  | 모듈별 폴더를 생성하여 구성  model에서 다양하게 구성해뒀다가 환경에 맞게 변경 |

| **KoNLP** | | | |
| --- | --- | --- | --- |
| 라이브러리 이름 | 사용방법 | 활용내용 | 출처 |
| Okt | from konlpy.tag import Okt | 한글 형태소 분석(정규화, 어구 및 명사 추출, 형태소 분석) |  |
| rhinoMorph | import rhinoMorph | 감성분석 예측 모델 |  |
| train\_test\_split | from sklearn.model\_selection import train\_test\_split | 훈련, 테스트 데이터 분류 |  |
| CountVectorizer | from sklearn.feature\_extraction.text import CountVectorizer | 피처 벡터화 |  |
| TfidfVectorizer | from sklearn.feature\_extraction.text import TfidfVectorizer | Tfid 벡터화 |  |
| cross\_val\_score | from sklearn.model\_selection import cross\_val\_score | 교차 검증 |  |
| LogisticRegression | from sklearn.linear\_model import LogisticRegression | 벡터화 데이터 회귀분석 통해 정확도 검증 |  |
| GridSearchCV | from sklearn.model\_selection import GridSearchCV | 모델 성능 향상 위한 파라미터 조절 |  |
| jpype | rhinoMorph 사용 위해 설치 필요 |  |  |