세종대학교

팀번호	21

2025-하계 세종창의학기제(집중이수제) 주간학습보고서 (1주차)

창의과제	스마트노트				
이름	임홍철	학습기간	7월 7일 ~ 7월 11일		
학번	20011815	학습주차	2주차	학습시간	12
학과(전공)	데이터사이언스학과	과목명	자기주도창의전공 I	수강학점	3
※ 수강학점에 따른 회차별 학습시간 및 10회차 이상 학습 준수					
금주 학습목표	whisper large 모델을 활용하여 지난주에 있었던 번역 오류 해결 임베딩 한 것을 실제 데이터랑 코사인 유사도를 통해 임베딩이 제대로 되었는지 확인				
학습내용	whisper 모델 구조에 대해 파악 - 우선 음성 신호를 25ms 길이의 프레임 단위로 분할하고 10ms 단위로 서로 겹치게 한다. 그리고 각 프레임에 대해 기계학습시간에 배웠던 멜 스펙트로그램으로 변환한다. 이후에 멜 스펙트로그램을 시간 차원에 따라 2개의 1d convolution layer와 gelu 활성화 함수를 통해 임베딩 한다. 여기서 1d를 쓰는 이유는 음성 데이터가 시계열 데이터이기 때문이다. 임베딩 된 값을 트랜스포머 encoder에 넣은 후에 cross attention을 통해 트랜스포머 decoder 부분에 넣는다. 그리고 디코더 입력으로는 토크나이저를 거친 텍스트 토큰 시퀀스인데 여기에는 이전 토큰값, 유저 프롬프트, 언어, 전사인지 번역인지, 시간라인을 원하는 지 등이 들어간다. 생성된 임베딩 값들이 제대로 들어갔는지 내가 원하는 값과 코사인 유사도를 통해 인접한 값들이 무엇이 있는지 확인 가능하게 구현. 예시 임베딩을 하기 위해 beomi님의 github에서 2019년 기사 댓글들을 스크랩 한 것을 전처리하여(특수문자제거,중복제거) 5400만개의 단어 데이터를 얻었고 이를 바탕으로, 추후 rag를 진행할 때 의미 파악 증강에 활용할 것이다.				
학습방법	인터넷 서칭(github 코드 탐구)				
학습성과 및 목표달성도	아직 stt 모델에서 영어의 한국 발음으로 생기는 오류를 잡지는 못하였다. 그러나 이전보다 대폭의 성능 상승이 있긴했다.				
참고자료 및 문헌	https://github.com/openai/whisper/tree/main https://github.com/Beomi/KcBERT				
내주 계획	위에서 만든 단어세트를 통해 rag를 구현 하기전 비슷 단어들을 잘 뽑는지 검증해볼것이고, 추가 데이터 3개년을 구해서 이것도 변환하는 작업을 해볼 것이다.				

2025 년 7 월 15 일

지도교수 전창재 (인)