



2024-06-11

☰ 태그	
👤 참석자	기흥 기흥 김

06/12 발표자료 구성

1. Whisper 구조 소개

- Whisper Github 소개 및 코드 구조 분석 설명
 - notion upload 내용 참조

2. 전체 연구 일정 (목표, 연구내용 등 포함)

- 기존 주제의 한계점 (초기에 잡았던 목표 등)
 - Baseline으로 두고 있던 whisper model의 학습 조건에 맞는 하드웨어 환경 구축이 불가.(서버 최소 20대 필요)
 - retraining→ 기존 model의 finetuning으로 방향성을 재설정
 - 기존에 수집한 영미권 News Contents data가 상용 STT model로 생성된 자막이 많다는 것을 확인.
→ data 신뢰성 하락.
 - 따라서 data domain을 방송 콘텐츠 중, 영화 부문에 초점을 두어 자막, 한국어 번역이 내포된 data를 활용하여 whisper model을 finetuning하는 것을 research 방향으로 설정.
- 변경된 목표 제안 및 방향성 소개
- 변경된 주제 관련 상세 설명
 - Domain 변경: 뉴스 콘텐츠 → 영화
 - En → En Transcribe, En → X Translate
 - 자막이 내포된 영화 콘텐츠 데이터를 수집하여 기존 whisper model 버전들을 fine tuning함

- metrics에서 기존 model보다 향상된 성능을 달성하는 것을 목표로 잡음.
- 초기 주요 일정 / 지금까지 진행된 사항 정리
- 3. 연구 결과물 시현 (현재까지의 연구결과물 / Whisper 시현?)
 - 필요 시 Whisper 시현
- 4. 하계 방학 일정 (하계방학 간 일정 및 목표, 연구내용)
 - 하계 방학 간 연구 내용 / 목표 및 일정
 - 하계방학: Inference Time 개선, 영화 콘텐츠 데이터 수집
 - 2학기: whisper fintuning
 - 현재 주제에 대한 주요 일정 (전체 일정)

-
- 5. 변경된 주제 관련 사항 소개 (2. 에서 소개)
 - 1. 영어 -> X
 - 2. Inference Time 개선
-

해야할 일

Github 업로드 및 정리 (회의록, 진행 코드 등)

회의록 정리

현재까지의 진행 코드 정리 (Whisper Inference Sh, AWS, GCP, Watson, etc)

Study 관련 자료 별도 정리 (Ubuntu, Docker, Transformer, Whisper Paper, Whisper Code)

발표자료 업로드

기말고사 발표자료 준비

기존 연구 내용의 한계점

- 이미 높은 Anchor 값 (이미 높은 정확도를 달성한 기 구현 모델)
- Model Re-Train에 필요한 서버 자원 부족 (Whisper의 경우, 서버 20대로 약 2주 소요)

- 데이터셋 수집 자체에서의 한계

변경된 연구 방향성

- 모델의 Inference 개선은 기존대로 진행 (방학 중 완료 예정 - 우선, GPU 측면의 Inference 개선)
- 기존 모델에 추가 학습 데이터를 통해, Model Fine-Tuning
- 데이터셋의 경우, 데이터셋 수집의 용이성을 고려하여 영화 데이터셋을 기반으로 추진
- 기존 En → En Transcription만 진행함을 넘어, 다국어 자막 파일 수집을 통해 En → X Translate 또한 수행(한국어 한정)
- 시간적 여건이 될 경우 && 방언 데이터셋을 받을 경우, 방언에 대한 추가 처리

방학 중 수행할 사항

- Whisper Model GPU Inference 개선 (e.g. Quantization, etc)
 - 수행 결과물: Inference 개선 코드 / 기존 Whisper 대비 WER 증가율 분석 결과
 - 해당 결과는 아래 진행 사항에 대한 BaseLine 코드로 구성
- 영화 Video Data (중 Audio Data) 및 영화 자막 데이터(En, Kr) 수집
 - 수행 결과물: Fine-Tuning에 있어 유의미한 결과를 도출할 수 있을 정도의 데이터셋 (Video, Audio, En/Kr 자막)
- Audio Data Noise Cancellation
 - 수행 결과물: Noise Cancellation을 수행한 위 데이터 (Audio)
- Parameter (e.g. Inference Chunk Size, etc)에 따른 WER 결과분석 (현, 30s) (방학 ~ 2학기)
 - 수행 결과물: Chunk Size에 따른 Inference Time, 수행 결과(WER 등) 분석

2학기에 수행할 사항

- Parameter (e.g. Inference Chunk Size, Padding Size, etc)에 따른 WER 결과분석 (현, 30s) (방학 ~ 2학기)
 - 수행 결과물: Chunk Size에 따른 Inference Time, 수행 결과(WER 등) 분석

- 방학 중 수집한 Dataset을 기반으로 Model Domain Adaptation (영화)
 - 영어에 대한 Transcription (Metric: WER)
 - 한국어에 대한 Translate (Metric: BLEU)
-