|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **[Euron] 22주차 프로젝트 회의록** | | | | |
| 팀 명  (작성자) | 유린이들  (김민지) | | 날짜 | 2024.02.05 |
| 진행 단계 | 도배 하자 질의 응답 처리 : 한솔데코 시즌2 AI 경진대회 Baseline 모델링 | | | |
| 회의 내용 | [논의 사항 1]: Baseline 코드 실행 | | | |
| 의견 | * Colab에서 돌리는데 Epoch 1회당 3시간씩 걸리는 등 과하게 시간이 오래 걸리는 문제 발생 * Colab에서 Baseline 코드를 돌리는 과정에서, utf-8 locale 관련 오류 발생 -> 회의를 통해 함께 트러블슈팅한 결과, 해결 완료! * Baseline 모델은 KoGPT2 모델 -> 다른 모델도 실험해보자 | | |
| 결론 | 실험환경: Google Colab - T4 GPU  KoELECTRA, KoBERT, KoGPT-3, Kollama 등 다양한 QnA 모델로 실험할 예정 | | |
| [논의 사항 2]: GPT API 활용 논의 | | | |
| 의견 | OpenAI사의 GPT-3.5 API를 통해 Vector를 추출한 결과 0.73xx대로 리더보드의 점수가 생각보다 높은 사례를 발견함.   * 대회 규칙에 다음과 같은 안내사항 발견: “API를 통한 외부데이터 수집, 데이터 전처리는 가능하나, API를 통한 추론은 불가능합니다. (Ex. ChatGPT API를 통한 추론 등 불가능) 반드시 언어 모델 학습의 과정이 존재해야하며, 학습된 언어 모델을 바탕으로 추론이 이루어져야합니다.” | | |
| 결론 | 결국 API 방식 대신, GPT 모델을 fine-tuning해서 새로운 모델을 구축하거나, 기존에 생각했던 방향처럼 다양한 모델 실험을 통해 점수를 내볼 예정. | | |
| 회의 인증 사진 | 스크린샷, 멀티미디어 소프트웨어, 소프트웨어, 멀티미디어이(가) 표시된 사진  자동 생성된 설명 | | | |
| 다음 회의 계획 | 날짜: 2/9(금) 저녁 10시  계획: NLP 모델에서의 데이터 전처리 방법(이상치 찾는법 등) 공부해오기, Huggingface 등에서 pre-trained된 모델로 실험한 결과 공유하기,  Retrieval-Augmented Generation(RAG)방식 공부하고 적용하기 | | | |

**선택한 공모전:** [도배 하자 질의 응답 처리 : 한솔데코 시즌2 AI 경진대회]

(<https://dacon.io/competitions/official/236216/overview/description>)

**지금까지 한 것:**

* 실험환경 통일 및 방향성 확정
  + Google Colab - T4 GPU
  + RAG(Retrieval Augmented Generation) 기술 적용
* Baseline 모델 코드 실행 및 탐구
  + Colab에서 돌리는데 Epoch 1회당 3시간씩 걸리는 등 과하게 시간이 오래 걸리는 문제 발생
  + Colab에서 Baseline 코드를 돌리는 과정에서, utf-8 locale 관련 오류 발생 -> 회의를 통해 함께 트러블슈팅한 결과, 해결 완료!
  + 제공된 Baseline 모델은 KoGPT2 모델 -> 다른 모델도 실험해보자
  + 실험사진:

스크린샷, 텍스트, 소프트웨어, 멀티미디어 소프트웨어이(가) 표시된 사진

자동 생성된 설명

**앞으로 할 것:**

* 데이터 전처리를 통한 정확도 향상 예정
  + NLP 모델에서 전처리하는 과정에 대한 공부 필요하다 판단 -> 이상치 찾기 등
* QA 모델을 통한 실험 예정
  + KoELECTRA, KoBERT, KoGPT-3, Kollama 등
  + Huggingface에서 pre-trained된 모델 적극 활용하여 실험할 예정
* Retrieval-Augmented Generation(RAG)방식 공부하고 적용하기
  + LLM에서 확장하여, 사용자의 질문이나 요청에 대해 더욱 신뢰 가능하고 정보가 풍부한 답변을 제공하는 데에 도움을 주는 기술 -> 대회 목적에 부합한다고 판단
* 다음 회의 일정: 2/9(금) 저녁 10시