|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **[Euron] 23주차 프로젝트 회의록** | | | | |
| 팀 명  (작성자) | 유린이들  (조현지) | | 날짜 | 2024.02.12 |
| 진행 단계 | 도배 하자 질의 응답 처리 : 한솔데코 시즌2 AI 경진대회 Baseline 파인튜닝 | | | |
| 회의 내용 | [논의 사항 1]: Baseline 코드 파인튜닝 | | | |
| 의견 | * Baseline 코드에서 다른 모델로 변경 시도 -> 오류 발생 -> 모델은 그대로 사용하고 파인튜닝 적용하기로 함 * Epoch, lr, 데이터 포맷, loss 등 조절해보기로 함 * 데이터 추가하는 방법도 시도 | | |
| 결론 | 모델을 바꾸는데 오류가 지속적으로 발생해 모델을 그대로 사용하고 파인튜닝하는 방식으로 시도하기로 하였다 | | |
| [논의 사항 2]: 실험 공유 | | | |
| 의견 | * 데이터 포맷팅과 토크나이징 부분 수정하고 배치사이즈 조절 -> 평균 loss 값 감소, 하지만 제출 결과 그다지 높은 점수는 못받음 * Early stopping적용해 적절한 epoch수 확인 -> 코랩 무료 버전으로는 램용량에 한계가 있어 20으로 설정하기로 결정 * 데이터 전처리 함수 수정(특수 문자 및 숫자 제거, 중복 공백 제거, 양쪽 공백 제거) -> loss값 감소 | | |
| 결론 | - 평가 모델에 걸맞는 모델 찾기  - TTA 적용  - loss값 변경 (대회 평가지표인 cosine similarity이용)  - 데이터 증강 ( 단어 문장 일부 삭제, NLTK를 사용한 동의어 대체)  - test 데이터에 대한 pseudo labeling   * 실험 시 epoch = 5 & 10으로 진행해서 전처리 효과 있는지 loss값 확인하고 효과 있으면 20까지 돌려보기로 하였다 | | |
| 회의 인증 사진 |  | | | |
| 다음 회의 계획 | 날짜: 2/13(화) 저녁 10시  계획: 실험 진행 상황 공유하고 앞으로 할 것 공유하기 | | | |

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

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

**지금까지 한 것:**

* Baseline 코드 파인튜닝
* Baseline 코드에서 다른 모델로 변경 시도했으나 지속적인 오류가 발생해 모델은 그대로 사용하고 파인튜닝 적용하기로 하였음
* Epoch, lr, 데이터 포맷, loss 등을 조절하고 데이터를 추가하는 방법도 시도해보기로 함
* 실험 내용 공유
* 데이터 포맷팅과 토크나이징 부분 수정하고 배치사이즈 조절함 -> 평균 loss 값 감소, 하지만 제출 결과 그다지 높은 점수는 못받음
* Early stopping적용해 적절한 epoch수 확인 -> 코랩 무료 버전으로는 램용량에 한계가 있어 20으로 설정하기로 결정
* 데이터 전처리 함수 수정(특수 문자 및 숫자 제거, 중복 공백 제거, 양쪽 공백 제거) -> loss값 감소

**앞으로 할 것:**

- 평가 모델에 걸맞는 모델 찾기

- TTA 적용

- loss값 변경 (대회 평가지표인 cosine similarity이용)

- 데이터 증강 ( 단어 문장 일부 삭제, NLTK를 사용한 동의어 대체)

- test 데이터에 대한 pseudo labeling

-> 실험 시 epoch = 5 & 10으로 진행해서 전처리 효과 있는지 loss값 확인하고 효과 있으면 20까지 돌려보기로 하였음

- 다음 회의 일정: 2/13(화) 저녁 10시