RM 문제 원인 및 해결 방안

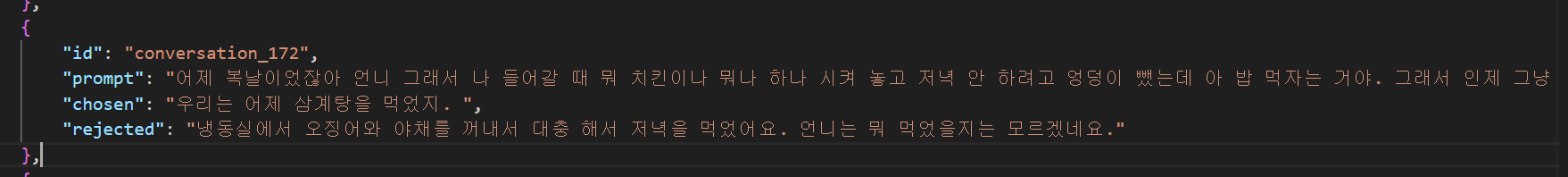
2023.08.23

1. **Test set이 이상하다.**
   1. 다른 팀은 똑같은 TestSet으로 성능이 오름
   2. 하지만 현재 데이터셋은 일상대화가 없음
      * **각 카테고리마다(일상대화, 번역, 혐오, Evol) 다시 테스트해볼 예정**
      * **사람이 직접 모든 카테고리에서 40개씩 양질의 TestSet 만들기**
2. **Chosen, Rejected 사이에 큰 차이가 없다.**
   1. Chosen, Rejected 사이의 큰 차이가 있어야 RM 학습에 효과가 있음
   2. 우리의 SFT+는 이미 너무 잘하므로 Chatgpt와 큰 차이가 없음
      * **다변의 차이가 적게 나는 Ko-Alpaca 12.8B 모델로 다시 response 생성하고 Rejected로 설정해서 데이터셋 수정**
3. **데이터 검수 부족**
   1. Chosen, Rejected가 바뀐 데이터 존재 가능

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

* 1. 이상한 Prompt 존재 가능



* 1. 이상한 Chosen, Rejected 존재 가능

텍스트, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

스크린샷, 텍스트, 폰트이(가) 표시된 사진

자동 생성된 설명

* **15 음절 이하인 답변이 있으면 다 삭제하기**
* **일상 대화 데이터셋 1500개는 Prompt는 퀄리티에 문제가 있는 상황이므로 사람이 직접 모든 데이터셋을 검수하기로 함**

**향후 일정:**

공통: 500개씩 일상대화 Instruction Set 검수

1. 각 RM Train 데이터셋 카테고리마다 Test 다시해보기 - 성완

2. Ko-Alpaca Model로 5000개 Response 생성 - 현빈

3. 모든 카테고리에 대해 50개씩 Test Set 만들기 – 유민

월요일까지 새로운 Reward 모델 제작 및 테스트 예정