Stanford Alpaca 보고서

2023/04/05 박성완

도표이(가) 표시된 사진

자동 생성된 설명

**Generate\_instruction.py**

1. 175개의 seed를 가지고 task pool에 넣어두기
2. Task Pool에서 언어모델(GPT)를 가지고 Instruction data를 생성
3. 생성된 data에 Classification/Identification 작업 수행
4. Filtering 과정 진행 후 최종 Instruction을 Task Pool에 넣고 다시 2번 과정 실행
5. 총 52000개의 Instruct data 생성

도표이(가) 표시된 사진

자동 생성된 설명

**Train.py**

Pretrain된 LLaMA에 alpaca\_data.json을 fine-tuning시킴.

Hugging face의 transformers를 이용해서 fine-tuning 진행.

Alpaca\_data.json의 format만 지키면서 원하는 데이터를 넣으면 알아서 바로 fine-tuning 가능.