

학습된 인공지능 모델

드라이브 url :

<https://drive.google.com/drive/folders/1XkHzcERgJRuzPy74w3VHCUYgIXDyaUJ1?usp=sharing>

1. 모델 개요

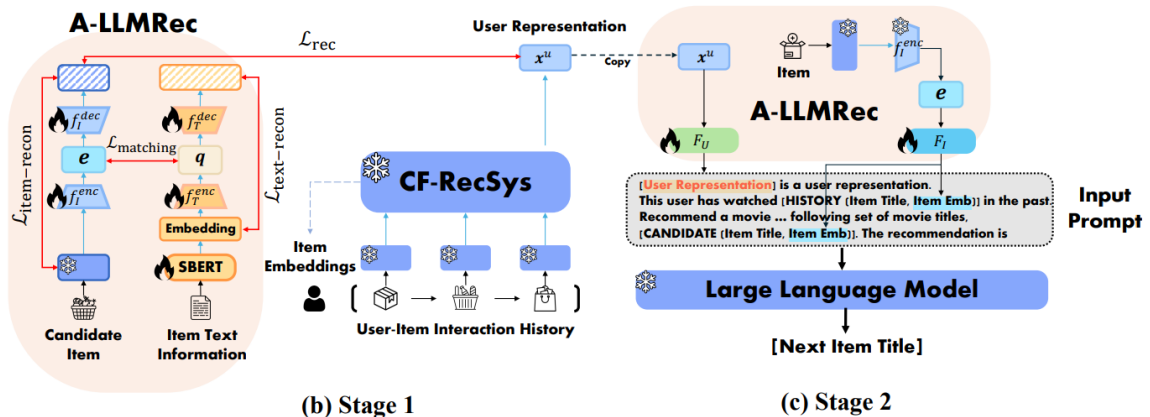
- **모델명** : A-LLMREC with GPT
- **목표** : CF의 문제점과 LLM의 문제점을 상호보완 하는 모델로 추천 품질이 향상된 모델을 대화형으로 사용하기 위함

2. 학습 데이터 및 전처리

- **데이터 출처** : Julian McAuley LAB(Amazon data)
- **데이터 구성**
 - **meta**
 - 상품의 제목 및 설명을 포함하는 정보
 - 제목과 설명을 추출
 - json
 - **review**
 - 사용자가 상품에 남긴 리뷰데이터
 - id와 관련 텍스트 추출
 - json
- **데이터 전처리** (전처리 문서를 참고해주시오)
 - **데이터 축소**
 - 카테고리화 연도, 리뷰 수를 제한하여 데이터 크기를 축소
 - **데이터 전처리**
 - 데이터 구성 내에서 필요하지 않은 것들을 제거, 너무 긴 정보들 요약 및 데이터 재구성
 - 376 GB → 540mb → 329KB

3. 모델 구조

- 기본 구조 :
 - CF 모델 : SASREC
 - 상품과 사용자의 상호작용을 학습
 - LLM 모델 : opt 6.7b / opt13b
 - CF의 후보들을 기반으로 상품을 추천
 - ALLMREC 모델
 - CF의 임베딩을 LLM의 임베딩 크기에 맞추고 아이템 설명을 이해하며 유사도 기반의 추천을 보완
- 구성 요소 및 흐름



4. 학습 단계

- CF의 학습
 - 사용자와 아이템간의 기본 상호작용을 학습함
 - 기본적인 추천 시스템의 성능을 확보
- phase1
 - 상호작용 이력과 콘텐츠 데이터를 모두 포함하는 결합된 협업-텍스트 임베딩을 생성
- phase2

- Phase 1에서 얻어진 협업-텍스트 임베딩을 LLM의 토큰 공간에 정렬, 이를 llm이 이해할 수 있게해줌

5. 추천 생성

- **ALLMREC모델**

- a_llmrec_model.py의 generate함수를 사용하면 모델을 사용하여 추천을 얻을 수 있음

- **with chatbot**

- 아직 미완성 단계이며 현재 진행 중에 있음
- chatbottesting.ipynb에서 진행 중
- 위 아래에 다른 llm을 붙이되 history는 공유하여 모델이 기억을 할 수 있게함
- 입력과 출력에 다른 프롬프트를 적용하여 진행예정
- 현재 구조는 변경 가능성이 있으며 완성된 것이 아님

6. 성능 평가 및 저장

- **ALLMREC모델**

- 모델 파일 내에 eval.py를 활용하여 평가가능
- 추천시스템에서 사용되는 hit@rate 와 NDCG가 존재
 - **hit** - 유저의 상품 기록 중 하나 제거 후 입력 모델이 제거된 답을 맞추면 hit로 판단 모델이 대답한 상품 후보 개수로 hit@1, hit@5가 됨
 - **모델에서 주로 사용하고 있는 평가지표**
 - **NDCG** - 추천 리스트에서 높은 순위에 있는 아이템이 더 많은 가중치를 가지도록 하여, 순위에 따른 추천 품질을 평가
- 현재 모델의 최고 점수 = 0.2