

2024년 전기 졸업과제 착수 보고서

LLM을 이용한 추천시스템 성능 개선 방안 연구



지도교수

조준수 교수님

팀명

Llecommend

팀원

201824482 박지환

201924604 하현진

201924562 장재혁

- 1 프로젝트 개요
 - 1.1 배경 및 필요성
 - 1.2 문제상황 및 기대효과
- 2 요구조건 분석 및 요소 선정 기준
 - 2.1 요구조건 분석
 - 2.2 시스템 요소 선정 기준
 - 2.2.1 평가지표 선정
 - 2.2.2 추천시스템 선정
 - 2.2.3 LLM 모델 선정
 - 2.2.4 데이터세트 선정
- 3 구체적 시스템 선택
 - 3.1 모델 선택(LLM, RecSys)
 - 3.2 데이터 전처리, 품질 관리 방안
- 4 현실적 제약사항 및 대책
- 5 개발 일정 및 역할 분담
- 6 참고 자료

1. 프로젝트 개요

1.1 배경 및 필요성

기존 추천 시스템은 사용자의 과거 행동 데이터를 기반으로 아이템을 추천하는 협업 필터링 기법을 주로 사용해왔다. 이 방식은 사용자가 이전에 구매하거나 평가한 아이템 정보를 바탕으로 유사한 사용자나 아이템을 찾아내어 추천을 제공한다. 그러나 이러한 기법들은 데이터의 불완전성과 제한된 분석 능력으로 인해 사용자의 선호도를 정확히 파악하기 어려웠다. 특히 새로운 사용자나 아이템에 대한 데이터가 부족할 때 시스템의 성능이 현저히 떨어지는 cold-start problem도 발생한다. 또한, 기존 시스템은 주로 협업 필터링이나 콘텐츠 기반 필터링에 의존하여 다양한 추천을 제공하지 못하며, 이는 사용자 경험을 단조롭게 만든다.

최근 등장한 LLM(Large Language Model)은 이러한 문제들을 해결할 수 있는 새로운 가능성을 제공한다. LLM은 방대한 양의 텍스트 데이터를 학습하여 언어 이해 및 생성 능력을 갖추고 있으며, 이를 통해 사용자의 의도와 맥락을 보다 정확히 반영할 수 있다. 예를 들어, 사용자의 리뷰나 대화 데이터를 분석하여 현재 사용자가 어떤 종류의 영화를 보고 싶어하는지, 어떤 주제나 스타일을 선호하는지 파악할 수 있다. 또한, 아이템의 메타데이터와 설명을 활용하여 설명 가능한 추천 시스템을 구축할 수 있으며, 이는 사용자가 왜 특정 아이템이 추천되었는지 이해하는 데 도움을 준다.

본 프로젝트의 목표는 LLM을 기반으로 기존 추천 시스템의 한계를 극복하고 개선된 추천 시스템을 개발하는 것이다.

1.2 문제 상황 및 기대 효과

1.2.1 문제상황

기존 추천 시스템은 사용자-아이템 간 상호작용 데이터만을 활용하여 추천을 수행하기 때문에 다음과 같은 한계점이 있다.

1) 데이터 부족으로 인한 cold start problem 발생

협업 필터링 기법은 사용자나 아이템에 대한 데이터가 충분히 누적되지 않은 경우, 즉 새로운 사용자나 새로운 아이템에 대해 추천할 때, 적절한 추천을 제공하는 데 어려움을 겪는다. 이로 인해 새로운 사용자나 아이템이 추천 시스템에 효과적으로 반영되지 못해 초기 추천의 정확도가 저하된다.

2) 사용자 맥락과 의도 반영 미흡

기존 추천 시스템은 사용자의 과거 행동 데이터를 기반으로 하기 때문에 현재의 맥락이나 의도를 반영하지 못하는 한계를 갖는다. 사용자의 관심사나 선호도가 변할 수 있는데, 과거 데이터만을 이용한 추천은 이러한 변화를 반영하지 못해 현재의 사용자 요구를 만족시키지 못한다.

3) 편향된 추천

협업 필터링 기법은 특정 그룹의 사용자나 인기 있는 아이템에 편향될 가능성이 높다. 이는 다양한 사용자에게 균형 잡힌 추천을 제공하지 못하고, 소수의 인기 아이템에 대한 추천이 반복적으로 이루어지는 문제를 야기한다. 이러한 편향은 사용자 경험의 다양성을 저해하고, 시스템의 공정성에도 부정적인 영향을 미친다.

1.2.2 기대효과

LLM 을 도입하면 텍스트 형태의 다양한 정보를 추천 시스템에 사용할 수 있게되어 사용자 리뷰, 대화 등을 통해 맥락과 의도 반영에 용이하고, 다양한 정보를 추천에 이용할 수 있으며, 추천시 설명을 제공해 투명성 및 신뢰성을 향상할 수 있다.

1) 아이템 메타데이터와 설명을 활용하여 cold start problem 완화

LLM 은 아이템 메타데이터와 설명을 활용하여 새로운 사용자나 아이템에 대한 데이터를 보완할 수 있다. 이를 통해 새로운 사용자나 아이템이 추천 시스템에 빠르게 반영되며, 초기의 낮은 추천 성능 문제를 해결할 수 있다.

2) 사용자 리뷰, 대화 등을 통해 맥락과 의도 반영

사용자 리뷰, 대화, 검색 기록 등을 통해 현재의 맥락과 의도를 반영한 개인화된 추천을 수행할 수 있다. LLM 은 자연어 처리 능력을 통해 사용자의 최신 요구를 파악하고, 이에 맞는 추천을 제공할 수 있다.

3) 자연어 기반 설명 제공으로 투명성 및 신뢰성 향상

자연어 기반 설명을 제공함으로써 추천의 투명성과 신뢰성을 높일 수 있다. LLM 은 추천된 아이템에 대한 설명을 자연어로 제공할 수 있어, 사용자가 추천의 이유를 쉽게 이해하고 수용할 수 있다. 이를 통해 사용자는 추천 시스템에 대한 신뢰를 높이고, 시스템과의 상호작용이 향상된다.

4) 편향 감소

LLM 은 다양한 데이터를 학습하여 특정 그룹이나 인기 있는 아이템에 편향되지 않은 공정한 추천을 제공할 수 있다. 다양한 사용자 그룹에 대해 균형 잡힌 추천을 제공함으로써 사용자 경험의 다양성을 높일 수 있다

2. 요구 조건 분석 및 요소 선정 기준

2.1 요구 조건 분석

기존 추천시스템에 대규모 언어 모델(LLM)을 통합하여 추천 성능을 향상시키는 것을 목표로 합니다. 이를 통해 사용자 만족도를 증대시키고, 시스템의 정확도와 효율성을 개선합니다. 이를 위해 적절한 평가지표, 추천시스템, LLM 모델을 선정해야 합니다.

2.2. 시스템 요소 선정 기준

2.2.1 평가지표 선정

추천시스템의 성능 변화를 객관적으로 평가하기 위해 아래와 같은 평가지표를 고려할 수 있습니다.

리스트 상위 k 개의 아이템 중 올바르게 추천된 아이템의 비율을 측정하는 Precision@ k , 상위 k 개의 추천 리스트에서 사용자가 실제로 선호한 아이템이 차지하는 비율을 측정하는 Recall@ k , 추천 리스트의 순서에 따른 추천의 질을 평가하는 Normalized Discounted Cumulative Gain (NDCG)을 사용합니다. PALR에서는 추천된 리스트 중 적어도 하나의 아이템이 사용자의 실제 선호도에 포함되는지를 측정하는 Hit Rate (HR), 첫 번째로 올바르게 추천된 아이템의 순위에 기반한 평균 역수 순위를 계산하는 Mean Reciprocal Rank (MRR), 그리고 시스템이 추천할 수 있는 고유한 아이템의 비율을 측정하는 Coverage를 사용합니다. 이러한 지표들은 추천 시스템의 정확성, 재현율, 순위 기반 평가, 성공적 추천 여부, 다양성을 종합적으로 평가합니다.

2.2.2 추천시스템 선정

LLM과 통합하기에 적합한 모델을 장단점을 파악하여 선정한다. 협업 필터링(Collaborative Filtering), 콘텐츠 기반 필터링(Content-Based Filtering), 하이브리드

추천시스템(Hybrid Recommender Systems)등 현재 가장 흔하게 사용되는 알고리즘을 선택해 LLM을 이용한 개선 방안을 제시한다.

2.2.3 LLM 모델 선정

추천 시스템의 추천 정확도와 성능 향상을 위해 text classification, feature extraction 성능을 중점적으로 하여 각 모델의 성능, 훈련 데이터 요구사항, 연산 비용 등을 비교하여 적절한 LLM 모델을 선정한다.

2.2.4 데이터세트 선정

MovieLens, Amazon Product Review, Netflix Prize, Goodreads등 아이템, 아이템에 대한 리뷰, 평점이 제공되는, 사용자에게 대한 정보를 추출할 수 있는 데이터셋을 선정해 모델을 테스트해볼 수 있도록 한다.

3. 구체적 시스템 선택

3.1. 모델 선택(LLM, RecSys)

- LLM : Meta-Llama-3-8B
 - 모델 선정 이유는 아래와 같다.
 - 높은 성능 : 4월에 공개된 최신 모델 중 하나로 GPT3.5 보다 높은 성능을 보여주며, 추론 성능 개선이 되었다. 따라서 후보군과 유저 히스토리를 통한 추천 아이템 선정에 적절하다고 판단했다.
 - 저렴한 비용 : 오픈 소스 모델로 보다 저렴하게 사용할 수 있다.
 - 대규모 데이터셋으로 학습 : 다양한 데이터를 학습해 다양한 지식을 바탕으로 추천을 진행해 높은 성능이 기대된다.
- RecSys : BERT4Rec
 - 모델 선정 이유는 아래와 같다.
 - LLM 과의 결합 용이성 : BERT 는 Transformer 를 기반으로 한 양방향 인코더 추천 시스템으로, LLM 과 결합이 용이하다. 이에 따라 본 연구의 주제인 LLM 와 RecSys 의 결합 면에서 보았을 때, 모델로 적합하다고 판단하였다.
 - 양방향성 : 사용자의 과거 행동에 대한 효과적인 이해 및 sequence representations learning 을 위하여 양방향에서 문맥을 통합하는 것이 필요한데 이를 반영하여 보았을 경우 여러 RecSys 후보군 중 해당 모델이 가장 적합할 것이라고 판단하였다.

3.2. 데이터 전처리, 품질 관리 방안

- 중복 데이터 제거: 데이터셋에서 중복된 영화 정보나 사용자 리뷰 등을 제거하여 데이터의 중복성을 최소화한다.
- 결측치 처리: 데이터셋에서 결측치가 있는 경우, 해당 부분을 적절히 처리(제거, 대체)하여 데이터의 완전성을 유지한다.
- 데이터 정규화: 데이터의 스케일을 조정하여 모델의 학습을 안정화한다.

- 데이터 형식 변환: 영화 정보나 사용자 리뷰 등의 텍스트 데이터를 토큰화하고, 필요에 따라 단어 임베딩 등의 기법을 활용하여 숫자 형식으로 변환한다.

4. 현실적 제약사항 및 대책

4.1. LLM에서의 문제점

- 계산 자원 요구량 과다: LLM은 거대한 모델 크기로 인해 많은 계산 자원을 필요로 한다. 이로 인해 비용이 증가하고, 실시간 응답성을 제공하는 데 어려움이 있을 수 있다.
 - 대책: 클라우드 GPU 자원을 적극 활용하고, 필요한 경우 분산 학습을 도입하여 계산 자원을 효율적으로 사용할 수 있다. 또한, 추론 시에는 생성 토큰 수를 최소화하는 전략을 통해 계산 부하를 줄일 수 있다. 예를 들어, 토큰 수를 제한하거나, 중요한 정보만 추출하여 모델에 입력하는 방식을 고려해야 한다.
- 추천 시스템 적용을 위한 입출력 형식 변환 필요: LM을 추천 시스템에 적용하려면 사용자의 입력 데이터를 모델이 이해할 수 있는 형식으로 변환하고, 모델의 출력을 사용자에게 이해하기 쉽게 변환하는 과정이 필요하다.
 - 대책: 프롬프트 엔지니어링을 통해 목적에 맞게 입력을 구성하고, 모델의 출력을 적절히 조정하는 방법을 개발한다. 이를 위해 사용자의 요청을 명확히 정의하고, 원하는 출력 형식을 지정하여 모델이 일관된 결과를 제공할 수 있도록 한다. 다양한 템플릿과 프롬프트 조정을 통해 적절한 입력 형식을 찾아내는 과정이 필요하다.
- 프롬프트 길이 조정: LLaMA3 7B 모델은 8k 길이의 컨텍스트 창을 가지고 있다. 과도하게 긴 입력이 주어지면 모델이 제대로 동작하지 않을 수 있으며, 중요한 정보가 손실될 가능성이 있다.
 - 대책: 사용자의 입력 데이터를 적절히 요약하고, 중요하지 않은 오래된 history를 제거하는 기준을 설정한다. 또한, 핵심 정보만을 추출하여 프롬프트를 구성함으로써 모델이 중요한 맥락을 잃지 않도록 한다. 이를 통해 모델의 효율성을 높이고, 응답 시간을 단축할 수 있다.

4.2. RecSys에서의 문제점

- 데이터 부족: BERT4Rec 의 특성 상 시퀀스 데이터에서 학습을 진행하므로 대규모의 데이터가 필요할 가능성이 있다.
 - 대책 : 시퀀스 변형 등 데이터 증강 기법을 적용하거나, 외부 소스에서 추가 데이터를 확보하여 모델 학습에 적용한다.
- 초기 사용자 문제 : Cold start problem 의 가능성이 존재한다.
 - 협업 필터링 및 콘텐츠 기반 필터링을 활용하여 적용한다.
 - 유사 프로필 사용자의 데이터를 활용하여 해당 문제를 완화한다.
- 계산 자원 요구량 과다 : 대규모 데이터를 처리하므로 요구하는 컴퓨팅 자원이 많다.
 - 모델 경량화 기법(Quantization, Knowledge Distillation 등)을 적용하여 전반적인 모델의 크기를 축소한다.
 - 하드웨어 자원(GPU 등)을 활용한 분산 학습 기법을 적용, 병렬로 학습을 진행한다.

4.3. 데이터에서의 문제점

- 데이터 부족: 새로운 사용자 및 아이템에 대한 데이터 부족
 - 대책 : 다양한 데이터 소스를 활용하여 새로운 사용자와 아이템에 대한 데이터를 수집하고, 데이터 증강 기법을 도입하여 부족한 데이터를 보완한다.
- 데이터 편향: 특정 그룹의 사용자나 인기 있는 아이템에 편향
 - 대책 : 데이터 수집 단계에서 다양한 사용자와 아이템을 포함하도록 하고, 편향을 보정하는 알고리즘을 도입한다. 예를 들어, 데이터 샘플링 기법을 통해 다양한 사용자 그룹을 대표하는 데이터를 균형 있게 수집하고, 추천 알고리즘에 편향 보정 기술을 적용하여 특정 아이템이나 사용자 그룹에 대한 과도한 추천을 방지한다. 또한, 추천 결과를 주기적으로 평가하여 편향 여부를 모니터링하고, 필요한 경우 알고리즘을 조정하는 방법 등을 고려해볼 수 있다.

5. 개발 일정 및 역할 분담

5.1. 개발 일정

6월				7월				8월				9월			
1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
데이터셋 선정 및 데이터 전처리															
		가선택한 모델 테스트 및 모델 파인튜닝													
				LLM, 추천 시스템 통 합 및 오류 수정											
						중간보고 서 작성									
								평가지표 점수 개선 및 추가 파인 튜닝							
									기능 점검 및 개선 방안 탐색						
												오류 수정 및 최종 보고서 작성 및 발표 준비			

5.2. 역할 분담

공통	<ul style="list-style-type: none"> ● 담당 분야 관련 논문 탐독 및 개인 연구 결과 공유
박지환	<ul style="list-style-type: none"> ● 추천 시스템 관련 작업 수행 : 시스템 구현 및 성능 평가 ● 강화 학습 관련 작업 수행 : LLM 모델 담당자와 연계한 학습 과정 및 파라미터 업데이트 과정 구현 ● 경량화 작업 수행
하현진	<ul style="list-style-type: none"> ● LLM 모델 관련 작업 수행 : LLaMA3-8b 모델의 구현 및 테스트 ● 강화 학습 관련 작업 수행 : 추천 시스템 담당자와 연계한 학습 과정 및 파라미터 반영 과정 구현
최재혁	<ul style="list-style-type: none"> ● 프롬프트 엔지니어링 ● 데이터셋 개발 : 필요한 데이터셋 구현 및 전처리 수행 ● 전반적 작업 내용 기록 및 지원

6. 참고 자료

- PALR: Personalization Aware LLMs for Recommendation
- LlamaRec: Two-Stage Recommendation using Large Language Models for Ranking
- A Large Language Model Enhanced Conversational Recommender System