

모델링 테스트 계획 및 결과 보고서

**** 현 문서는 모델링의 테스트 계획 및 결과 보고서로 웹 화면과 그의 예외 처리는 후에 제출 할 배포 테스트 계획 및 결과 보고서에서 작성 예정**

테스트 계획

1. 테스트 목적

- 상품 추천 시스템이 사용자의 요청에 적절하게 응답하고, 관련성이 없는 요청을 정확하게 구별하는가
- 추천 트리거가 올바르게 작동하며, 추천 모델이 적절한 추천 결과를 생성하는가
- 프롬프트가 성능을 최적화하고 사용자 요청에 정확한 응답을 제공하는가
- 시간성능은 어느 정도되는가

2. 모델 개요

- 협업 필터링과 llm을 합쳐서 후보를 뽑고 추천을 해주는 시스템
- GPT-4o-mini와 A-LLMrec모델을 합쳐 놓은 대화형 추천 모델

테스트에 사용된 모델 정보

모델의 성능 지표

데이터 수 user 1588, item 4099

hit@1 : 0.12

사용된 모델 정보

입력 LLM : GPT-4o-mini

ALLMREC(내에서 사용되는 LLM) : OPT- 6.7B

출력 LLM : GPT-4o-mini

3. 테스트 범위

- 대상
 - GPT-4o-mini와 A-LLMrec(opt-6.7b)을 활용한 대화형 추천 모델

- 환경
 - colab환경
 - 시스템 RAM 83.5 GB
 - GPU(A100) RAM 40.0 GB
 - 디스크 235.7 GB

4. 테스트 항목

- **정확도 테스트** : 추천 요청에 대해 정확한 상품 추천이 이뤄지는지 확인
- **요청 분석 테스트** : 사용자의 요청에 따라 다른 상품을 추천하는지 확인
- **명확하지 않은 요청 테스트** : 불명확한 요청에 대한 처리가 올바른지 확인
- **응답 형식 테스트** : 추천이 필요한 경우와 그렇지 않은 경우 각각 올바른 언어 및 형식으로 응답하는지 확인
- **추천 제한 테스트** : 여러 추천이 가능한 경우 최대 3개의 추천을 제공하는지 확인
- **사용자에 따른 추천 테스트** : 다른 사용자(다른 구매 기록을 갖고 있는)에 따른 추천 내용 변화
- **성능 테스트** : 응답하는 데 걸리는 시간을 확인

4. 테스트 시나리오 및 결과

시나리오 1: 정확도 테스트

- **테스트 케이스** : 상품 추천과 관련된 구체적인 요청
- **사용자 입력** : "스킨케어 추천해줘"
- **기대 결과** : 추천 모델이 작동하고 적절한 추천을 해줌.
- **실제 결과**

사용자의 요청에 따라 다음과 같은 스킨케어 제품을 추천합니다:

1. Hyaluronic Acid Serum - Anti-Aging, Moisturizing, Timeless Glowing Skin.
- 피부에 수분을 공급하고 노화 방지 효과가 있어, 특히 건조한 피부에 적합합니다.
2. REBONCEL Aqua Rich Hydrating Face Foam Cleanser
- 수분을 공급하는 클렌저로, 피부를 부드럽게 세정하면서도 수분을 유지해줍니다.
3. Murad Hydration Intense Recovery Cream
- 강력한 보습 효과를 가진 크림으로, 피부 회복에 도움을 주며 건조한 피부에 적합합니다.

이 제품들은 사용자의 피부 타입과 요구에 맞춰 추천되었습니다. 필요에 따라 선택해 보세요!

소요시간 : 16초

시나리오 2 : 명확하지 않은 요청 테스트

- **테스트 케이스** : 추천과 무관한 요청 또는 일반적 질문을 입력
- **사용자 입력** : “배고파”
- **기대 결과** : 추천 모델이 작동하지 않고, “다시 자세하게 질문해주세요” 응답 출력
- **실제 결과** : “다시 자세하게 질문해주세요.”

소요시간 : 1초

시나리오 3 : 요청 분석 테스트

- **테스트 케이스** : 시나리오1과는 다른 구체적인 요청
- **사용자 입력** : “선크림도 추천해줘”
- **기대 결과** : 추천 모델이 작동하고 요구사항을 맞춘 추천을 해줌.
- **실제 결과** :

사용자의 요청에 따라 적합한 선크림을 추천해드리겠습니다:

1. Alteya Organics Sunscreen Rose Face Cream SPF 30 - 보습 성분이 포함되어 있어 건성 피부에 적합합니다.
2. OZNaturals Facial Sunscreen SPF 30 - 피부를 보호하면서도 수분을 공급해주는 제품입니다.
3. FASCY LAB GREEN+ Korean Sunscreen - SPF 50+ PA++++ - 높은 SPF로 강력한 자외선 차단 효과를 제공하며, 피부에 자극이 적습니다.

이 제품들은 건성 피부에 적합하며, SPF가 높아 자외선으로부터 피부를 잘 보호해줄 것입니다.

소요시간 : 7초

시나리오 4 : 사용자에게 따른 추천 테스트

- **테스트 케이스** : 다른 구매 기록을 가진 유저 비교
- **사용자 입력** : “스킨케어 추천해줘”
- **기대 결과** : 추천 모델이 작동하고 사용자별 맞춘 추천을 해줌.

- **실제 결과 :**

- **유저 0 : 그 전 사용 기록 없음**

사용자의 요청에 따라 다음과 같은 스킨케어 제품을 추천합니다:

Hyaluronic Acid Serum - Deep Moisture Anti Aging Face Serum for Dry Skin: 건조한 피부를 위한 수분 공급과 노화 방지 효과

SKEDERM EXHYDRO Ceramide Complex Moisturizer: 피부 장벽 강화 및 수분 유지

Murad Hydration Intense Recovery Cream: 강력한 수분 공급과 피부 진정 효과

이 제품들은 사용자의 피부 타입과 요구에 잘 맞을 것으로 보입니다.

소요시간 : 7초

- **유저 545 : 9 개 이상 상호작용함**

사용자의 요청에 따라 다음과 같은 스킨케어 제품을 추천합니다:

1. Hyaluronic Acid Serum - Anti-Aging, Moisturizing, Timeless Glowing Skin.

- 피부에 수분을 공급하고 노화 방지 효과가 있어, 특히 건조한 피부에 적합합니다.

2. REBONCEL Aqua Rich Hydrating Face Foam Cleanser

- 수분을 공급하는 클렌저로, 피부를 부드럽게 세정하면서도 수분을 유지해줍니다.

3. Murad Hydration Intense Recovery Cream

- 강력한 보습 효과를 가진 크림으로, 피부 회복에 도움을 주며 건조한 피부에 적합합니다.

이 제품들은 사용자의 피부 타입과 요구에 맞춰 추천되었습니다. 필요에 따라 선택해 보세요!

소요시간 : 7초

시나리오 5 : 사용자가 원하는 상품을 찾을 수 없는 경우

- **테스트 케이스 :** 다른 분야의 아이템 추천 요청

- 예시 : 장난감 추천해줘

- **기대 결과 :** 대상을 찾을 수 없을 알리고 대화를 중단 및 다시 시작할지 여부를 요청

- **ex)** 조건에 맞는 상품을 찾을 수 없습니다. 새 대화를 시작하시겠습니까?

- **실제 결과**

- 사용자의 요청에 따라 장난감 추천을 드리겠습니다. 다음은 추천할 만한 상품입니다:
 1. ****Bath Bombs for Kids with Surprise Toys Inside**** - 아이들이 좋아하는 목욕 폭탄으로, 안에 서프라이즈 장난감이 들어 있어 재미를 더합니다.
 2. ****Oukzon 40 Pcs Art and Crafts Supplies Kit for Kids**** - 다양한 미술 및 공예 용품이 포함된 키트로, 창의력을 키우는 데 도움을 줍니다.
 3. ****Disney Frozen Townley Girl Mega Nail Set**** - 디즈니의 인기 캐릭터가 포함된 네일 세트로, 아이들이 손톱 꾸미기를 즐길 수 있습니다.
 이 상품들이 도움이 되길 바랍니다!

소요시간 : 10초

시나리오 6 : 다른 상품에 대한 추천 속도 비교

- **테스트 케이스** : 다른 뷰티 분야의 아이템 추천 요청
 - 예시 : 풋 크림 추천해줘
- **기대 결과** : 상품을 추천해주되 다른 것들과 유사한 속도가 걸리는 것을 기대
- **ex)** 실행결과 5~ 10 초 내
- **실제 결과**

풋 크림 추천에 대한 요청에 따라 다음과 같은 상품을 추천합니다:

1. Stride Foot Cream
 - 깊은 보습과 진정 효과가 있는 풋 크림입니다.
2. Silver Miracles Colloidal Silver Foot Cream
 - 항균 효과가 있는 콜로이드 실버 성분이 포함된 풋 크림입니다.
3. Simply Dana Foot Restoration Cream
 - 발 피부 회복에 도움을 주는 크림입니다.

이 상품들은 발을 부드럽고 건강하게 유지하는 데 도움을 줄 것입니다.

소요시간 : 7초

결과 분석

시나리오 별 결과 분석

시나리오 1: 정확도 테스트

- **테스트 케이스** : 상품 추천과 관련된 구체적인 요청

- **정확성** : 추천 모델이 스킨케어 제품에 대한 요청을 정확히 인식하고 관련 제품을 추천
- **다양성** : 세럼, 클렌저, 크림 등 다양한 종류의 스킨케어 제품을 추천
- **상세성** : 각 제품의 주요 특징과 효과를 간단히 설명
- **반응 시간** : 16초의 소요 시간은 복잡한 추천 모델을 고려할 때 적절한 수준
- **개선 사항** : 현재 추천하는 상품의 개수는 3개 정도 이고 추후에는 더 많은 상품을 추가 하여 사용자에게 더 많은 선택지를 제공함으로써 정확도를 높여야함

시나리오 2 : 명확하지 않은 요청 테스트

- **테스트 케이스** : 추천과 무관한 요청 또는 일반적 질문을 입력
- **테스트 목적 달성** : 추천과 무관한 요청에 대해 시스템이 예상대로 반응
- **정확성** : 시스템이 명확하지 않은 요청을 적절히 처리
- **반응 시간** : 1초의 소요 시간은 매우 빠른 응답 속도이는 시스템이 명확하지 않은 요청을 신속하게 처리
- **개선 사항** : 현재 결과는 단순한 대답이 되지만 추후에는 프롬프트 수정을 통해 더 나은 질문 예시나 안내 사항을 추가 해야함

시나리오 3 : 요청 분석 테스트

- **테스트 케이스** : 시나리오1과는 다른 구체적인 요청
- **정확성** : 시스템이 사용자의 선크림 추천 요청을 정확히 이해하고 적절한 제품을 추천
- **다양성** : 시스템은 서로 다른 브랜드의 선크림 3가지를 추천하여 다양성을 제공
- **상세성** : 각 제품에 대해 SPF 지수와 주요 특징을 간단히 설명하여 사용자에게 유용한 정보를 제공
- **관련성** : 추천된 제품들이 건성 피부에 적합하고 높은 SPF를 가지고 있어 사용자의 요구 사항에 부
- **반응 시간** : 7초의 소요 시간은 이전 시나리오와 비교했을 때 적절한 수준
- **개선 사항** : 현재는 단순한 요청 분석이지만 추후에 다양한 요청 분석에 대한 샘플을 생성해 프롬프트에 추가 함으로써 고려해야 하는 상황을 늘려야함

시나리오 4 : 사용자에게 따른 추천 테스트

- **테스트 케이스:** 구매 기록이 다른 두 유저에 대한 추천 비교 (유저 0은 구매 기록이 전무함)
- **개인화:** 두 유저에게 일부 다른 제품이 추천되어 개인화가 적용되고 있음을 보여줌
- **일관성:** 두 유저 모두에게 Hyaluronic Acid Serum과 Murad Hydration Intense Recovery Cream이 추천되어 일부 일관성 유지
- **다양성:** 유저 545에게는 클렌저가 추가로 추천되어 제품 다양성 증가
- **상세성:** 두 경우 모두 제품의 주요 특징과 효과를 설명하여 유용한 정보 제공
- **반응 시간:** 두 유저 모두 7초로 동일한 응답 시간을 보여 일관된 성능 유지
- **개선 사항:**
 - 사용자의 구매 기록에 따른 추천의 차별화를 더욱 명확히 할 필요가 있음
 - 개인화된 추천 이유를 더 상세히 제공하여 사용자 경험 향상 가능
 - 사용자의 피부 타입이나 선호도를 고려한 추천 알고리즘 개선 필요

시나리오 5 : 사용자가 원하는 상품을 찾을 수 없는 경우

- **테스트 케이스 :** 다른 분야의 아이템 추천 요청
- **테스트 목적 달성 실패 :** 시스템이 기대한 대로 작동하지 않았음 장난감이 데이터베이스에 없음에도 불구하고 추천을 시도했습니다.
- **부적절한 응답 :** 시스템은 뷰티 제품 데이터베이스에서 장난감을 추천하려함
- **응답 시간 :** 10초의 소요 시간은 이전 시나리오들과 비교했을 때 약간 길지만, 복잡한 처리 과정을 고려하면 수용 가능한 수준
- **개선 사항**
 - 시스템이 데이터베이스에 없는 항목에 대한 요청을 인식하고 적절히 처리할 수 있도록 개선
 - 사용자에게 데이터베이스의 한계를 인지, 가능한 대안을 제시하거나 새로운 대화를 시작할 수 있는 옵션을 제공

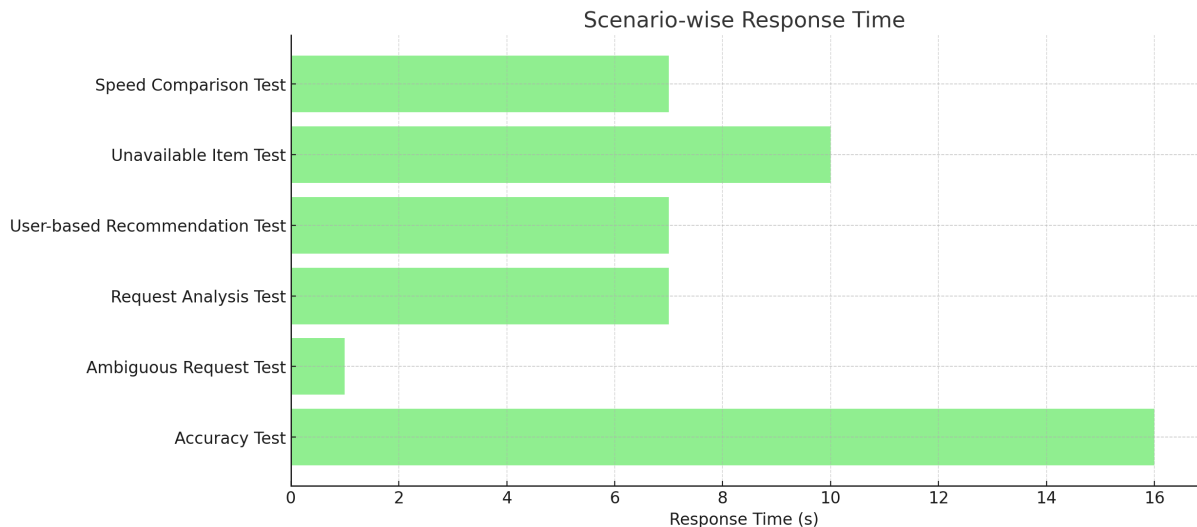
시나리오 6 : 다른 상품에 대한 추천 속도 비교

- **테스트 케이스 :** 다른 뷰티 분야의 아이템 추천 요청

- **테스트 목적 달성:** 다른 뷰티 분야(풋 크림)에 대한 추천 요청에 시스템이 성공적으로 응답
- **추천의 적절성:** 시스템은 3가지 다른 풋 크림을 추천하여 다양성을 제공
- **상세성:** 각 제품에 대해 간단한 설명을 제공하여 사용자에게 유용한 정보를 전달
- **반응 시간:** 7초의 소요 시간은 기대 결과인 5~10초 내, 다른 시나리오들과 비교했을 때 일관된 성능을 보여줌
- **개선 사항 :** 일관된 시간이 나오지만 특정 경우에 대답이 오래 걸리는 경우도 있음 (시나리오 1의 경우) 원인을 파악하고 개선해야함

시각적 자료

시나리오 별 응답 속도 비교



결론 요약

각 시나리오를 통해 시스템의 다음과 같은 특성과 한계가 있음을 확인

1. **정확도와 반응성 :** 대부분의 시나리오에서 추천 모델은 사용자 요청을 잘 이해하고 적절한 추천을 제공, 평균 7~10초 내외의 반응 시간을 유지하여 성능이 적절한 것으로 평가
2. **명확하지 않은 요청 처리 :** 비추천성 요청에 대해 모델이 신속하게 응답하며 적절히 거부하는 능력을 보였지만, 안내 메시지를 개선해야함

3. **사용자별 개인화** : 구매 이력에 따른 차별화된 추천이 이루어짐 하지만, 더 심화된 개인화와 설명 기능의 향상이 필요
4. **카테고리 한계 처리** : 시스템은 카테고리에 맞지않는 항목에 대해 비정상적인 추천을 시도, 적절한 예외 처리가 필요함을 확인