

멀티모달 검색을 위한 GraphRAG구현 및 성능 향상

문제점 개요서

9조

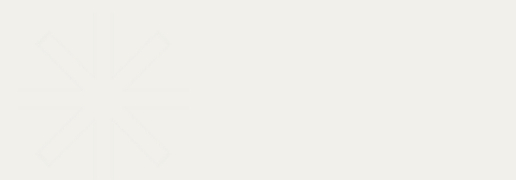
허정우, 김보배

목차

| | | | |
|--|---|--|--|
| | <p>＊ 01</p> <p>RAG기반 멀티모달 검색의 필요성</p> | <p>＊ 02</p> <p>멀티모달 RAG를 위해 설계해야 할 것</p> | <p>＊ 03</p> <p>RA-CM3 모델 소개</p> |
| | <p>＊ 04</p> <p>논문에서 제시한 한계</p> | <p>＊ 05</p> <p>연구 가능성 탐색</p> | <p>＊ 06</p> <p>GraphRAG 동작 방법과 기대 효과</p> |

■ 기존 멀티모달 모델의 한계

학습된 데이터를 바탕으로 생성하는 기존 모델들은
모든 지식을 매개변수에 저장하기 때문에 더 많은 정보를 위해서 학습 데이터와
매개변수 수가 계속 증가

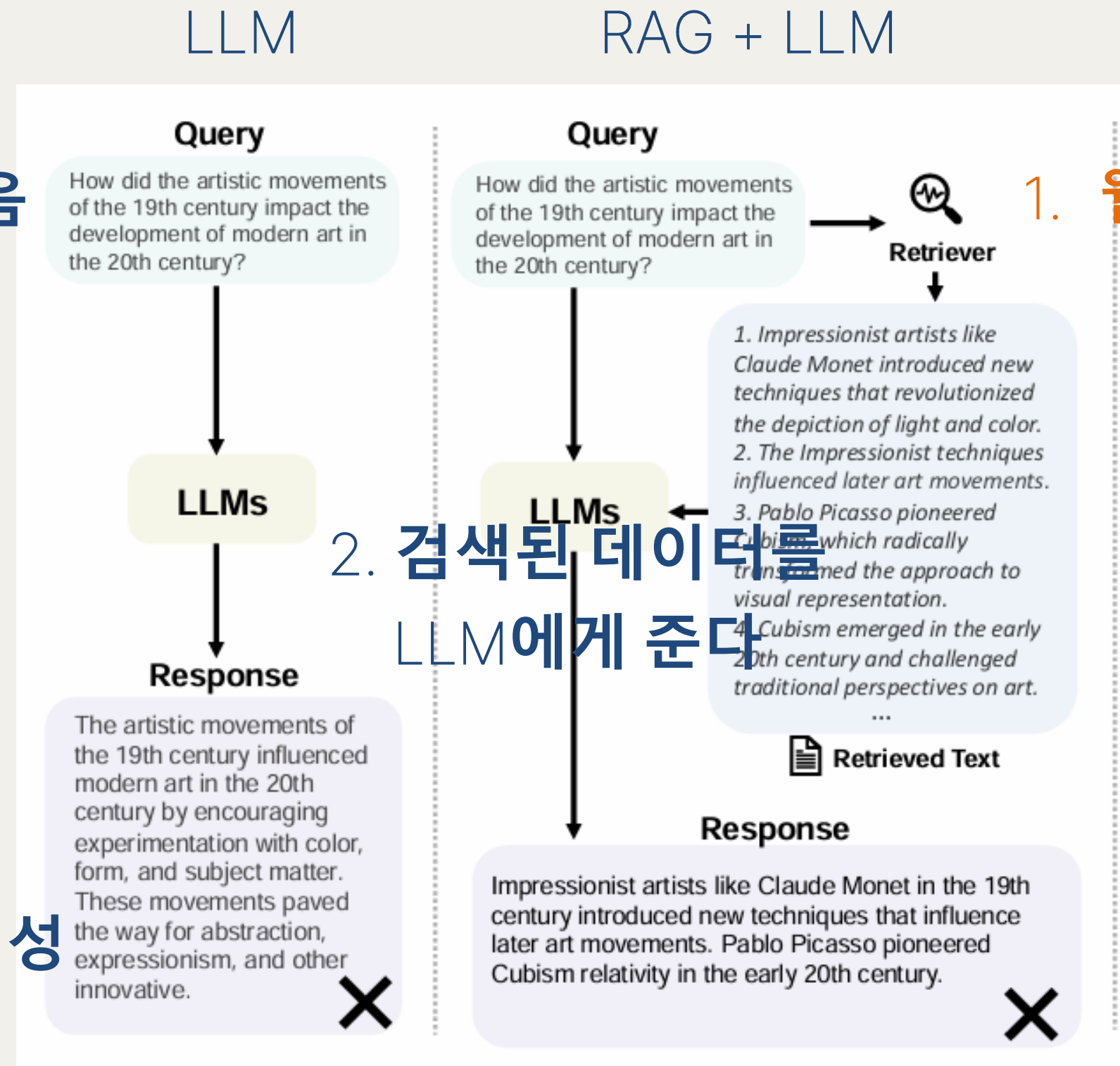


1 RAG기반 멀티모달 검색의 필요성

■ RAG란

1. 질문을 받음

2. LLM이 답변 생성



1. 웹 등에서 검색하는 과정(Retriever)

Retrieval + Augment & Generation

3. LLM은 생성된 데이터를 기반으로 응답(Generator)

1 RAG기반 멀티모달 검색의 필요성

■외부 데이터 기반 멀티모달 모델의 필요성

1. 외부 메모리 활용은 시간이 지나도 지식 갱신 가능
2. 엔티티 지식(고유 명사 같은 인물, 물건 등)이 필요한 경우 효과적
3. 설명 가능하고 신뢰할 수 있는 예측 제공

- 텍스트 + 이미지 검색이 가능한 검색기 (Retriever)
- 멀티모달 데이터를 활용해 텍스트 및 이미지를 생성하는 생성기(generator)

■ 검색기

- CLIP : 이미지, 텍스트를 벡터[0.1, -0.3, ...]의 형태로 변경해주는 모델
- 1. 데이터 검색을 통해 이미지-텍스트 쌍을 가져온다.
- 2. CLIP을 통해 각각 벡터로 변환시킨다.
- 3. 벡터를 평균, 정규화 한다.
- 4. MIPS(최대 내적을 구하는 방법)을 사용하여 사용자의 질의와 관련성 높은 문서를 찾는다.
- 5. 적합한 상위 K개의 문서를 활용한다.

■생성기

- CM3 Transformer : 이미지, 텍스트 생성이 가능한 모델
- 1. 검색된 K개의 문서를 학습된 데이터 앞쪽에 추가한다.
- 2. 검색된 문서와 미리 학습된 문서에 대한 손실 함수(예측 값과 실제 값의 차이)를 가중치를 두어서 계산한다.
- 3. 최적의 가중치를 찾는다.

■ RA-CM3의 기능

- 지식 집약적인 멀티모달 생성 가능
 - 희귀한 엔티티에 대해 검색을 활용해 정확하게 생성
 - 예) 명나라 도자기에 대해 이미지 검색을 통해 올바르게 생성



■ RA-CM3의 기능

- 희귀한 지식 조합 생성 가능
 - 예) 프랑스 국기 + 달에 대해 기존 모델은 주로 미국 국기 + 달을 보여줌

**RA-CM3
In-context**

French flag



RA-CM3 output



French flag waving on the moon's surface.

Baseline outputs

(Vanilla CM3)



(Stable Diffusion)



■ RA-CM3의 기능

- 제어된 이미지 생성 가능
 - 예) 삼각형의 집 모양 이미지를 사전에 주면 비슷한 결과를 출력

RA-CM3 In-context

(Specify the style to generate)



RA-CM3 output



Baseline outputs

(Vanilla CM3)



(Stable Diffusion)



Photo of a house taken on an autumn day.

- 웹 검색을 통해 이루어지기 때문에 대규모 노이즈가 포함된 이미지-텍스트 세트 존재하게 된다.
- 인종, 성별 등의 속성에 대한 편향이 포함될 수 있다.
- 성적 콘텐츠나 폭력과 같은 유해한 콘텐츠가 포함될 가능성이 있다.

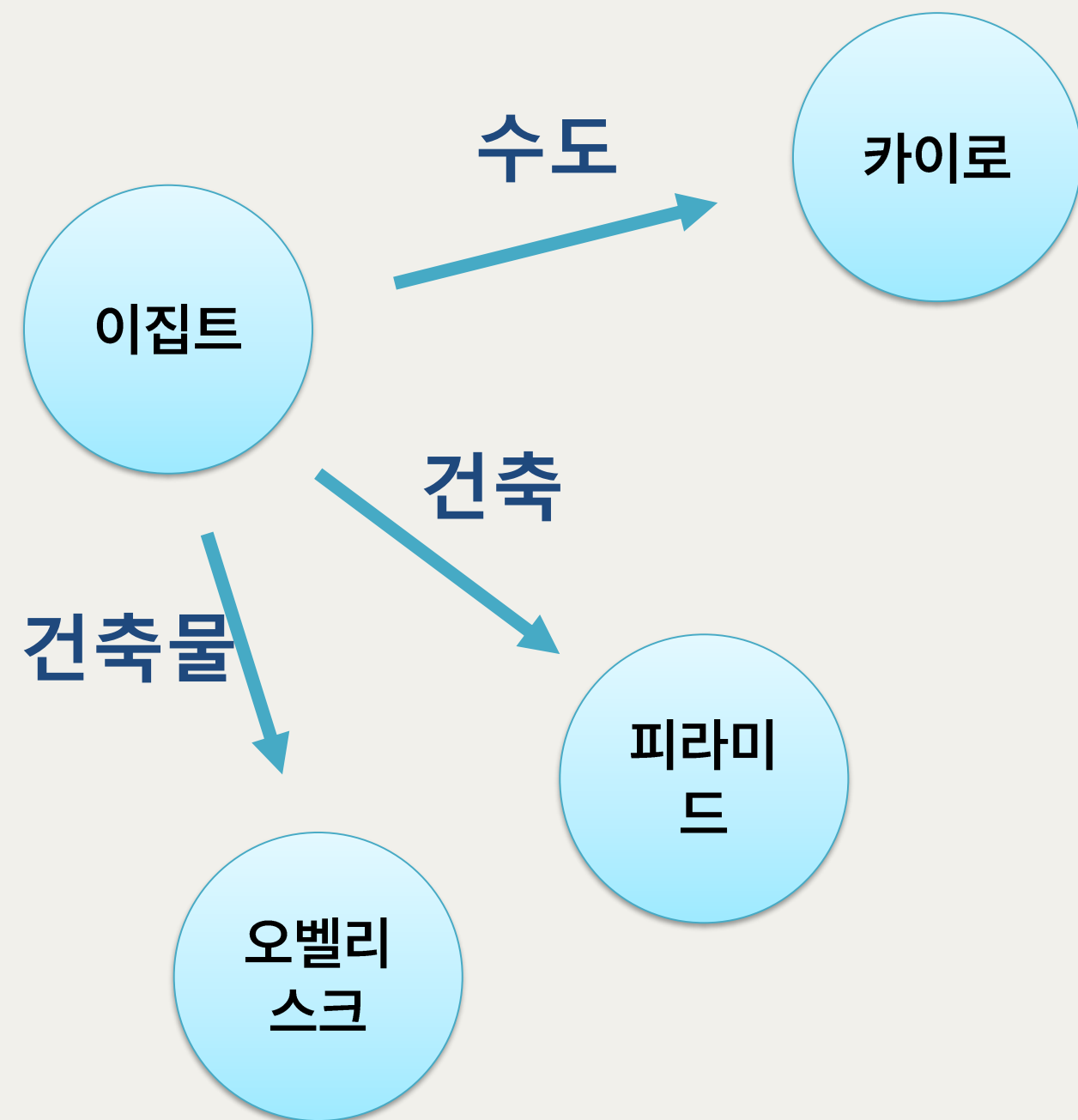
- 검색기, 생성기를 더욱 발전시킬 가능성 탐색
- 검색 속도 최적화

MIPS(내적)이 아닌 다른 알고리즘으로 유사도 측정 가능

- 검색된 데이터(이미지+텍스트)에 대한 가중치 조정
- 검색 품질 향상(편향 조절과 노이즈 제거)
- 검색을 굳이 하지 않아도 되는 데이터에 대한 경우에 대한 처리

예: 일반적인 푸들 사진의 경우 굳이 검색 안하고 사전 학습 데이터만을 이용하는 것이 훨씬 빠르다

■ Graph 데이터 베이스란



■ GraphRAG를 추가한다면

1. 데이터와 관계를 먼저 그래프 기반 데이터 베이스에 저장해둔다.
2. 쿼리에서 키워드를 뽑아 그래프 기반 데이터 베이스에서 유사한 개념을 같이 웹에서 검색
예) '고대 이집트 문명의 건축 양식을 검색할 때
기존 방식: 고대 이집트 문명, 건축 양식 키워드가 포함된 문서를 전부 검색
그래프: 고대 이집트 -> 건축 양식 -> 피라미드/오벨리스크 등과 같이
연관된 정보를 찾아서 바로 검색.
3. 검색된 데이터 기반으로 응답 생성

■ GraphRAG 사용 기대 효과

1. 쿼리에서 찾아낸 키워드만 검색하는 것이 아닌, 그래프 데이터베이스에 연관성 있는 개념들까지 같이 검색이 가능하다. (속도 향상과 정확성 향상)
2. 엔티티 지식에 대한 조금 더 나은 검색 결과를 기대해 볼 수 있다.

예) 스페이스 x가 화성을 탐사하는 모습

그래프 데이터베이스에서 스페이스x -> 화성 탐사선처럼 정보를 찾아 화성 탐사선에 대한 정보를 포함시킬 수 있다.

- Graph Retrieval-Augmented Generation: A Survey(2024)
- Retrieval-Augmented Multimodal Language Modeling(2023)