

문제점 개요서

Project Name	멀티모달 검색을 위한 GraphRAG구현 및 성능 향상
-----------------	--------------------------------

9조

201902770 허정우

202002470 김보배

지도교수: 임성수 교수님 (서명)

Document Revision History

REV#	DATE	AFFECTED SECTION	AUTHR
1	2025/03/21	1, 2	허정우, 김보배

Table of Contents

1.	SURVEY PAPER - LIMITATIONS FOCUS	5
2.	LIMITATIONS AND RESEARCH GAPS	9

List of Figure

그림 목차 항목을 찾을 수 없습니다.

1. Survey Paper - Limitations Focus

◇ 왜 RAG기반 멀티모달 검색인가

논문에선 최근 멀티모달 모델(DALL-E, CM3)등이 이미지 생성에서 놀라운 성과를 달성했지만 이 모델들은 모든 지식을 매개변수에 저장하기 때문에 더 많은 정보를 학습하기 위해 학습데이터와 매개변수 수가 계속 증가해야한다고 말함.

이에 따라서 웹 검색 등의 외부 메모리를 활용하여 학습하는 멀티모달 모델의 필요성을 강조함.

1. 외부 메모리 활용은 시간이 지나도 지식 갱신이 가능함
2. 엔티티 지식(고유 명사 같은 물건, 인물)이 필요한 경우 특히 효과적임
3. 설명 가능하고 신뢰할 수 있는 예측 제공

기존 RAG 모델 또한 자연어 처리 분야에서 높은 성능을 보여준다고 하지만 텍스트 데이터에 대해서만 너무 집중 되어있다고 지적함.

◇ 멀티모달 RAG를 구현하기 위해서 설계해야 할 것

1. 텍스트+이미지 검색이 가능한 검색기 설계
2. 멀티모달 데이터를 활용해 텍스트 및 이미지를 생성하는 생성기 설계

◇ 이 논문에서 제안한 모델인 RA-CM3

1. 입력 데이터: 텍스트, 이미지, 텍스트+이미지 문서 사용
2. 검색기 설계: DPR방식과 CLIP기반 모달 인코더 사용
 - DPR: 질의 q 와 외부 메모리 M 을 받아 관련 점수를 반환(고밀도 벡터 유사도 측정으로 쿼리와 검색 데이터 유사도를 측정)
 - CLIP: 이미지와 텍스트를 동일한 벡터 공간상에 표현하여 유사도 측정.
 - 입력 이미지, 텍스트에 대해 CLIP인코더를 통해 각각 벡터로 인코딩. 두 벡터를 평균하고 L2정규화하여 질의 인코더와 메모리 인코더에 사용(DPR에 사용)
 - 검색과정은 MIPS(최대 내적을 구하는 방법)을 사용하여 관련성 높은 문서를 찾아냄. 최종적으로 상위 K 개의 문서를 활용
 - 검색 전략: 이미지+텍스트, 단순 K 개 아닌 중복회피 전략
3. 생성기 설계: CM3 Transformer 기반, 검색된 문서를 앞에 추가하여 생성기 학습, token prediction loss기반 학습
 - 검색된 문서 k 개를 입력 시퀀스 앞쪽에 추가
 - 검색된 문서와 학습된 문서에 대한 손실 함수를 따로 가중치를 두어 생각함 ($L = L_{\text{학습}} + aL_{\text{검색}}$)
 - 최적의 a 값을 0.1이라 찾아냄

◇ 동작 방식

1. 쿼리에서 키워드를 찾아 웹 검색
2. 검색된 이미지, 텍스트, 텍스트+이미지 데이터를 CLIP 인코더로 벡터로 변환, DPR 방식으로 유사도 측정 후 상위 K 개의 데이터를 추출
3. K 개의 데이터를 시퀀스 앞쪽에 추가, 이미지 생성.

◇ 논문에서 사용한 데이터셋

1. LAION 데이터 셋으로 학습
2. MS-COCO 데이터 셋에서 평가

◇ RA-CM3의 기능

1. 지식 집약적인(knowledge-intensive) 멀티모달 생성 가능
 - 희귀한 엔티티를 검색을 활용해 정확하게 생성
 - 예) 명나라 도자기에 대해 웹 검색 이미지를 통해 검색 후 올바르게 생성
2. 희귀한 지식 조합(composition of knowledge) 생성 가능
 - 기존 모델은 주로 학습 기반으로 달+미국 국기를 보여주지만 달+ 프랑스 국기에 대해서도 잘 동작
3. 제어된 이미지(controlled image) 생성 가능
 - 예제 이미지를 주면 그 이미지의 시각적 특성을 반영해 새로운 이미지 생성
 - 예시) 삼각형 모양의 집 이미지와 호숫가 옆 집을 쿼리로 입력하면 삼각형 모양의 집 이미지가 나오게 된다.

◇ 논문에서 지적인 한계

1. 웹 검색을 통해 이루어지기 때문에 대규모 노이즈가 포함된 이미지-텍스트 셋이 존재하게됨
2. 인종, 성별 등의 속성에 대한 편향이 포함될 수 있다.
3. 포르노그래피나 폭력과 같은 유해한 콘텐츠가 포함될 가능성이 있다.

작성 요령(제출 시 삭제할 것)

◦ 최신 연구 중심으로

- ✓ 연구 개요서에서는 기존 연구의 배경을 설명하는 것이 주요 목표였지만, 이 문서는 연구의 한계점을 구체적으로 분석하는 것이 핵심임
- ✓ 최근 연구 중심으로 정리해야 함(최소 3~5년 이내)
- ✓ 해결되지 않은 문제와 한계를 구체적으로 분석한 뒤 한계를 도출하는 것이 중요함.
(예: 연구의 실용성(실제 적용 가능성), 데이터셋, 실험 방법의 제약 등)

◦ 저명한 학회 및 저널 인용 필수

- ✓ 저명한 학회 및 저널에서 발표된 논문을 인용하는 것이 중요함
(CVPR, ICML, ACL, EMNLP, ICLR, IEEE Transactions, ACM CCS)
- ✓ 될 수 있으면 국내보다는 국제 학술지, 컨퍼런스 발표된 영어 논문을 인용하는 것을 권장함
- ✓ 목표로 하고 있는 컨퍼런스나 저널이 있다면, 해당 분야의 논문을 중심으로 읽고 연구 방향을 설정해보는 것을 추천함

번호	연구 제목(저자)	저널/컨퍼런스 (연도)	주요 내용 요약	한계점
1	Retrieval-Augmented Multimodal Language Modeling	ICML 2023	멀티모달 데이터 검색 성능을 높인 RAG기반 LLM 검색기 설계: DPR방식과 CLIP기반 모달 인코더 사용 생성기 설계: CM3 Transformer 기반, 검색된 문서를 앞에 추가하여 생성기 학습 기존 CM3보다 Text-to-Image, Image-to-Text 성능이 높음	웹 검색을 통해 이루어지기 때문에 대규모 노이즈가 포함된 이미지-텍스트 셋이 존재하게 되고 편향이 생길 수 있다. 또한 GraphRAG를 사용했을 때 개선될 수 있을 점이 보인다.

2. Limitations and Research Gaps

◇ 연구 가능성 탐색

1. 검색기, 생성기를 더욱 발전시킬 가능성 탐색
2. 검색 속도 최적화
 - ANNS(approximate nearest neighbor search) 중 그래프 기반 검색인 (hierarchical navigable small world) 이용을 할 수 있을 것 같다.
3. 검색된 데이터(이미지+텍스트)에 대한 가중치 조정
4. 검색 품질 향상(편향 조절과 노이즈 제거)
5. 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

◇ GraphRAG를 사용했을 때 방식

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 양식’ 을 검색할 때
 - 기존방식: 고대 이집트 문명, 건축 양식 키워드가 포함된 문서를 전부 검색
 - 고대 이집트 -> 건축 양식 -> 피라미드/오벨리스크 등과 같이 연관된 정보를 찾아서 바로 검색.
3. 검색된 데이터 기반으로 응답 생성

◇ GraphRAG를 사용했을 때 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.

종합설계 1

번호	기존 연구	한계점	연구 필요성	본 연구의 기여
1	Retrieval-Augmented Multimodal Language Modeling (Michihiro Yasunaga, ArmenAghajanyan, WeijiaShi, RichJames, JureLeskovec, PercyLiang, MikeLewis, LukeZettlemoyer, Wen-tauYih)	웹 검색을 통해 이루어지기 때문에 대규모 노이즈가 포함된 이미지-텍스트 셋이 존재하게 되고 편향이 생길 수 있다. 또한 GraphRAG를 사용했을 때 성능, 속도 부분에서 개선될 수 있을 점이 보인다.	<ul style="list-style-type: none"> ● 유사도 측정 방식을 바꾸어 속도 향상 가능성 ● 검색기, 생성기 단계에서 발전시킬 가능성 ● 편향 조절과 노이즈 제거 	<p>기존 검색 과정에서 GraphRAG를 사용하였을 때 기대효과</p> <ol style="list-style-type: none"> 1. 속도 향상과 정확성 향상 2. 엔티티 지식에 대한 검색 결과 성능 향상