


캡스톤디자인 I 계획서

제 목	국문	시각 기반 질의응답에서 지식 베이스의 활용		
	영문	Knowledge Base for Visual Question Answering		
프로젝트 목표 (500자 내외)	<p>해당 프로젝트에서는 2022년도에 구축한 지식베이스 기반 시각 질의응답 데이터 (NIA 2-033-155)를 활용하여 시각 질의응답 시스템에서 지식베이스를 활용하는 딥러닝 기반 모델을 연구한다.</p> <p>우리의 데이터셋으로 최신 시각 질의응답 및 지식 기반 시각 질의응답 모델과 제안하는 모델의 성능을 비교, 분석한다.</p> <p>[논문 투고]</p> <p>SCI급 1편, BK21 IF 2 이상 학회 1편, KCC(컴퓨터 종합 학술대회) 수상을 목표로 한다.</p>			
프로젝트 내용	<p>실제 세계에서 사용하는 질문의 대부분은 외부 지식을 필요로 하는 질문이다. 이미지와 질의에 포함된 정보만 활용할 수 있는 것이 기존의 시각 질의응답 시스템의 한계점인데, 최근에는 이러한 점에 주목하여 지식베이스를 활용하는 시각 질의응답 시스템의 연구가 활발히 진행되고 있다. 유명한 벤치마크 데이터셋으로 KVQA(Knowledge-Aware Visual Question Answering), FVQA(Fact-based Visual Question Answering), OK-VQA(Outside-Knowledge based Visual Question Answering) 이 있다.</p> <p>KVQA와 FVQA는 트리플 {Head, Relation, Tail}로 이루어진 지식 베이스를 활용하고, OK-VQA는 위키피디아의 코퍼스를 지식 베이스로 활용하는 데이터셋이다. 우리가 구축한 데이터셋은 한 개의 이미지-질의 쌍에서 지식베이스가 필요한 질문 1개, 지식베이스가 필요하지 않은 질문 1개씩 구성되어 있다. 질의 종류에 따라 트리플과 위키피디아의 코퍼스를 지식 베이스로 포함한다.</p> <p>우리의 데이터셋은 새로운 벤치마크 데이터셋이 될 것이다. 프로젝트에서는 해당 데이터셋에서 지식베이스를 활용하는 딥러닝 기반 멀티모달 알고리즘을 개발한다.</p>			
중심어(국문)	시각 질의응답	지식베이스	자연어 처리	멀티모달 학습
Keywords (english)	Visual Question Answering	Knowledge Base	Natural Language Processing	Multimodal Learning
멘토	소속	서울과학기술대학교	이름	임경태
팀 구성원	학년/반	학 번	이 름	연락처(전화번호/이메일)
	4	20181796	김민준	010-8484-3956/mjkm20@gmail.com
	4	20181620	송승우	010-9164-6572/woo98220@gmail.com
	4	20202364	송지현	010-6889-3887/20202364@edu.hanbat.ac.kr
<p>컴퓨터공학과 캡스톤디자인 관리규정과 모든 지시사항을 준수하면서 본 캡스톤디자인을 성실히 수행하고자 아래와 같이 계획서를 제출합니다.</p> <p style="text-align: center;">2023 년 2 월 22 일</p> <p style="text-align: right;">책 임 자 : 김민준 </p> <p style="text-align: right;">희망 지도교수 : 장한얼</p>				

1. 캡스톤디자인의 배경 및 필요성

2015년에 질의응답에서 이미지를 활용하는 벤치마크 데이터셋인 VQA1.0이 공개되었다. 이 데이터셋을 필두로 시각 질의응답에 관한 연구가 국내외로 활발히 진행되었다. 하지만 일반적인 시각 질의응답 시스템은 이미지와 질의에 대한 얇은 이해만으로도 답변이 가능한 질문으로 구성되어 있다. 예를 들어, “이미지에서 고릴라는 몇 마리인가?”와 같은 질문은 답변이 가능하지만, “이미지 속 동물이 어떤 강에 속하는가?”와 같이 외부 지식을 필요로 하는 질문에는 답변이 불가능하다.

하지만 실제 환경에서는 대부분의 질의는 외부 지식을 필요로 하는 질의로 이루어진다. 이러한 점에 주목하여 국내외에서 지식베이스를 활용하는 질의응답 시스템에 관한 연구가 활발히 진행되고 있다. 유명한 벤치마크 데이터셋으로 KVQA, OK-VQA 등이 있다. 이 세가지 데이터셋은 외부 지식을 활용하는 VQA 데이터셋이기는 하지만, 이미지에 등장하는 객체가 무엇인지를 발견하는 것에 초점을 둔 데이터셋이다. 이러한 방식은 이미지에서 객체를 인식하는 연구에만 주목한다는 한계가 있다.

우리의 데이터셋은 KVQA와 OK-VQA 데이터셋과 비슷하게 지식 기반의 VQA 데이터셋이다. 이 데이터셋은 지식 그래프와 위키피디아의 코퍼스를 활용하여 VQA 시스템에 추가적인 정보를 제공한다. KVQA와 OK-VQA에 비해 더욱 포괄적이고 다양한 지식 기반을 제공하며 VQA 시스템 성능 향상에 기여할 것이다.

2. 캡스톤디자인 목표 및 비전

시각 질의응답 “이미지에 있는 동물은 무슨 과니?”라는 질문에 답하기 위해서는 질문, 이미지 외에 외부지식이 필요하다. 따라서 우리는 지식베이스 기반 데이터셋(NIA 20-330-155)을 사용한 VQA모델을 만들고자 한다. 데이터셋에 있는 지식기반 Triple(Head, Relation, Tail)을 활용하는 방안을 연구하고 기존의 모델의 성능과 비교한다.

3. 캡스톤디자인 내용

주요 기능

기능	내용
데이터 분석	- matplotlib, seaborn 으로 데이터 시각화 - 지식기반을 전처리하기 위하여 pandas 활용
질의 모델	- pytorch 기반의 bert계열 모델 사용 - 한국어, 영어가 있으므로 각 언어마다 모델에 맞게 Fine-tuning
이미지 모델	- pytorch 기반의 CNN계열 모델 사용
지식 활용	- abstract, triple 지식기반을 활용하는 방안 연구

비기능적 요구사항

	내용
성능	- 일반적인 VQA 모델의 성능과 지식기반을 활용한 VQA모델 성능 비교

	- 지식기반 활용 방안을 연구하여 성능 향상
보안	- 공공 데이터셋이므로 데이터 외부 사용 가능 - 논문을 통해 외부에 모델 공개
결과	- VQA모델을 사용하여 사용자에게 시각질의응답 서비스 제공 - 연구한 모델을 바탕으로 논문 작성

4. 캡스톤디자인 추진전략 및 방법

데이터셋(NIA 20-330-155)의 이미지는 60,360 개의 이루어져 있고, 질의는 120,720개로 대용량의 데이터셋이다. 또한 해당 데이터셋 학습을 위해서는 BERT 계열 모델과 CNN 계열 모델을 사용하기 위해서는 GPU가 필요하며 서버 메모리 64GB이상이 필요하다. 이를 해결하기 위해 서버를 활용하여 실험한다. 준비된 서버는 서버 메모리 128GB, GPU NVIDIA A100 80GB x 4 이다.

2022년 인공지능 학습용 데이터 구축 사업(NIA 20-330-155)에 참여하여 데이터셋에 대한 이해가 다소 있다. 또한 ‘인공지능’에서 다뤘던 CNN 기반의 모델에 대한 이해가 있으며, 연구실 활동을 통해 공부한 자연어처리 모델을 이해하고 사용할 수 있다.

서울과학기술대학교에서 MLP 연구실을 이끌고 있는 임경태 교수를 멘토로 섭외했다. 임경태 교수는 자연어 처리 기반의 멀티모달 러닝의 전문가이다.

프로젝트에 관련된 코드는 깃허브 <https://github.com/mjkmain> 에 업로드한다.

	팀 구성	성명	주요 역할
1	팀장	김민준	논문 및 자료조사, 모델 코드 작성
2	팀원	송승우	논문 및 자료조사, 모델 코드 작성
3	팀원	송지현	자료조사 및 UI 개발

사용 프레임워크



Hugging Face

5. 참고문헌

- [1] Aditya, S.; Yang, Y.; and Baral, C. Explicit reasoning over end-to-end neural architectures for visual question answering. AAAI 2018
- [2] Antol, S.; Agrawal, A.; Lu, J.; Mitchell, M.; Batra, D.; Zitnick, C. L.; and Parikh, D.

VQA: Visual Question Answering. ICCV 2015

[3] Fang, Y.; Kuan, K.; Lin, J.; Tan, C.; and Chandrasekhar, V. 2017. Object detection meets knowledge graphs. IJCAI 2017

[4] G. Narasimhan, M., and Schwing, A. Straight to the facts: Learning knowledge base retrieval for factual visual question answering. ECCV 2018

[5] Yuke Zhu, Oliver Groth, Michael S. Bernstein, and Li Fei-Fei. Visual7W: grounded question answering in images. CVPR, 2016

[6] Lee, C.-W.; Fang, W.; Yeh, C.-K.; and Wang, Y.-C. F. Multi-label zero-shot learning with structured knowledge graphs. CVPR 2018