

캡스톤 디자인 I 최종결과 보고서

프로젝트 제목(국문): 시각 기반 질의응답에서 지식 그래프의 활용

프로젝트 제목(영문): Visual Question Answering Empowered by Knowledge Graph

프로젝트 팀(원): 학번: 20181796 이름: 김민준

프로젝트 팀(원): 학번: 20181620 이름: 송승우

프로젝트 팀(원): 학번: 20202364 이름: 송지현

1. 중간보고서의 검토결과 심사위원의 '수정 및 개선 의견'과 그러한 검토의견을 반영하여 개선한 부분을 명시하시오.

없음

2. 기능, 성능 및 품질 요구사항을 충족하기 위해 본 개발 프로젝트에서 적용한 주요 알고리즘, 설계방법 등을 기술하시오.

2-1) 지식 그래프를 VQA 모델에서 활용하기 위해 ConvKB(Knowledge Base Completion Based on Convolutional Neural Network)를 활용하여 지식 그래프의 임베딩을 얻는다.

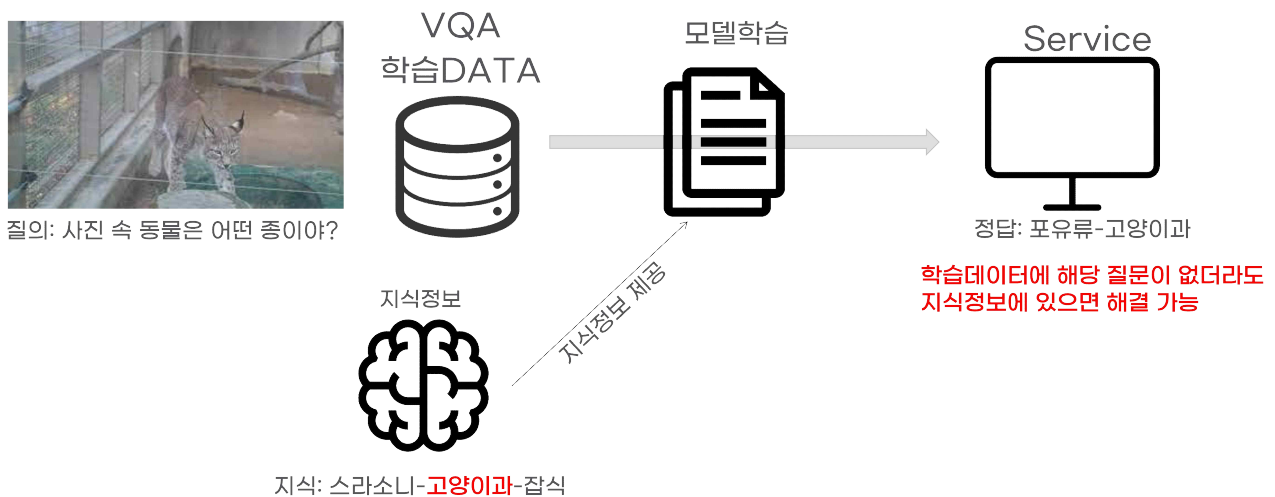
2-2) 실제 환경에서는 이미지-질의 쌍에 대응하는 외부 지식이 무엇인지 알 수 없기 때문에 Triple Classification 모듈을 설계하여 지식 그래프를 예측하며 VQA 학습을 진행하는 Real-Time Multi-Task Learning Model 설계.

2-3) Real-Time Multi-Task Learning Model은 학습 초기에 Triple classification의 성능이 낮기 때문에, 잘못된 방향으로 학습을 시작하게 되어 초기 학습에 불안정성이 존재한다. 이를 해결하기 위해 Triple classification에 Teacher-Forcing을 적용하여 학습 초기 불안정성을 완화.

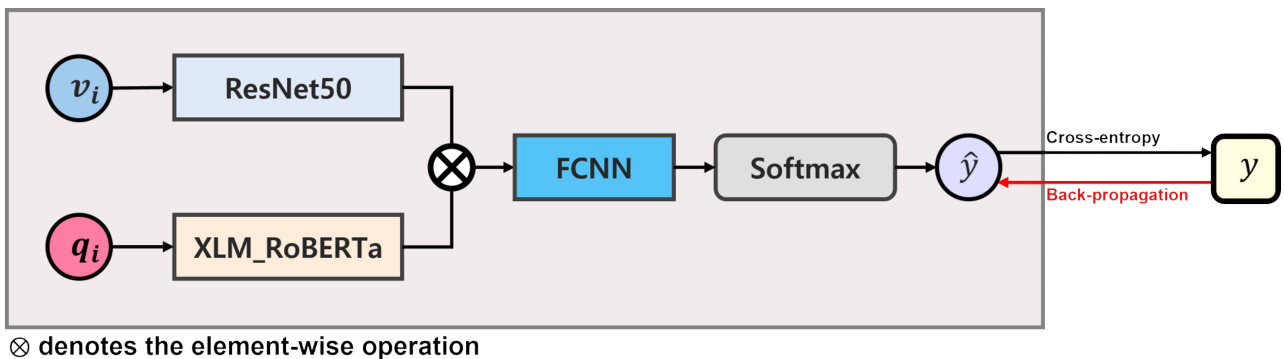
2-4) Triplet인 (Head, Relation, Tail)의 상대적 중요도를 학습하기 위하여 Triple Classification의 결과로 얻은 Embedding에 Self-Attention 적용.

3. 요구사항 정의서에 명시된 기능 및 품질 요구사항에 대하여 최종 완료된 결과를 기술하시오.

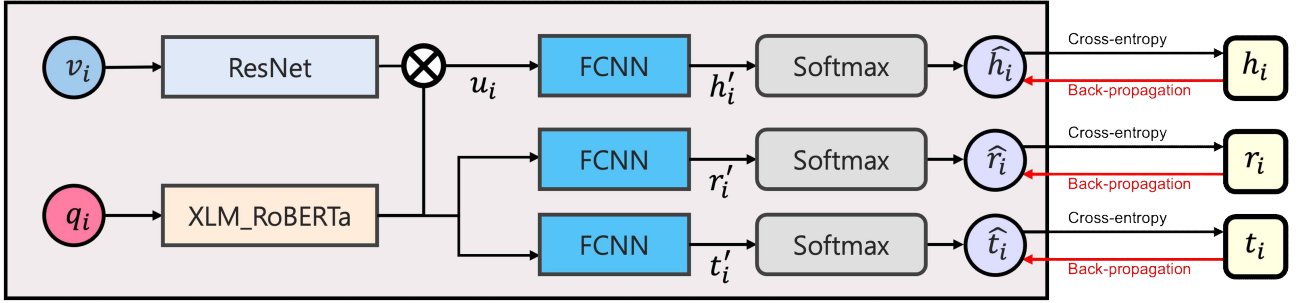
3-1) 전체 프로세스



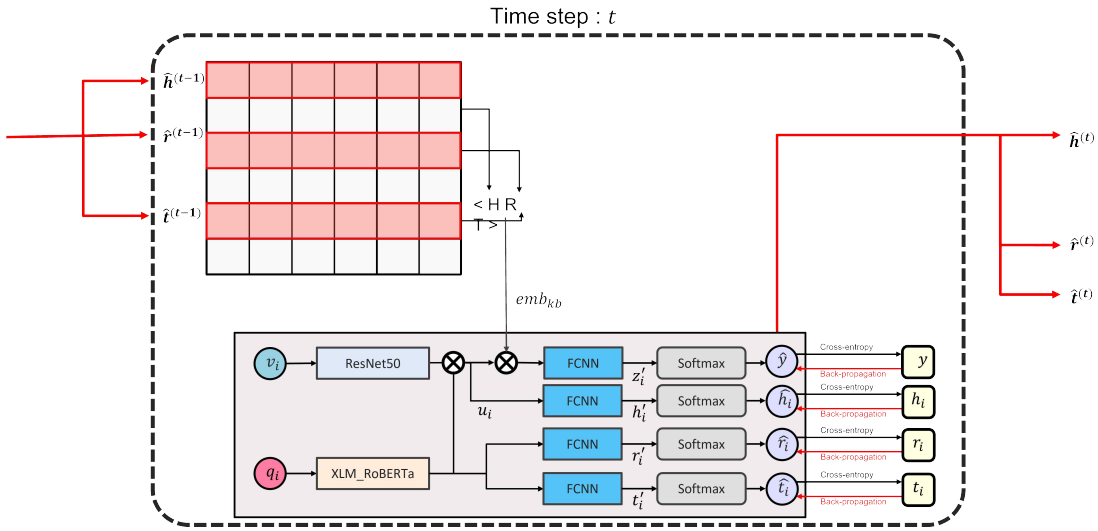
3-2) VQA 베이스라인 모델 구조



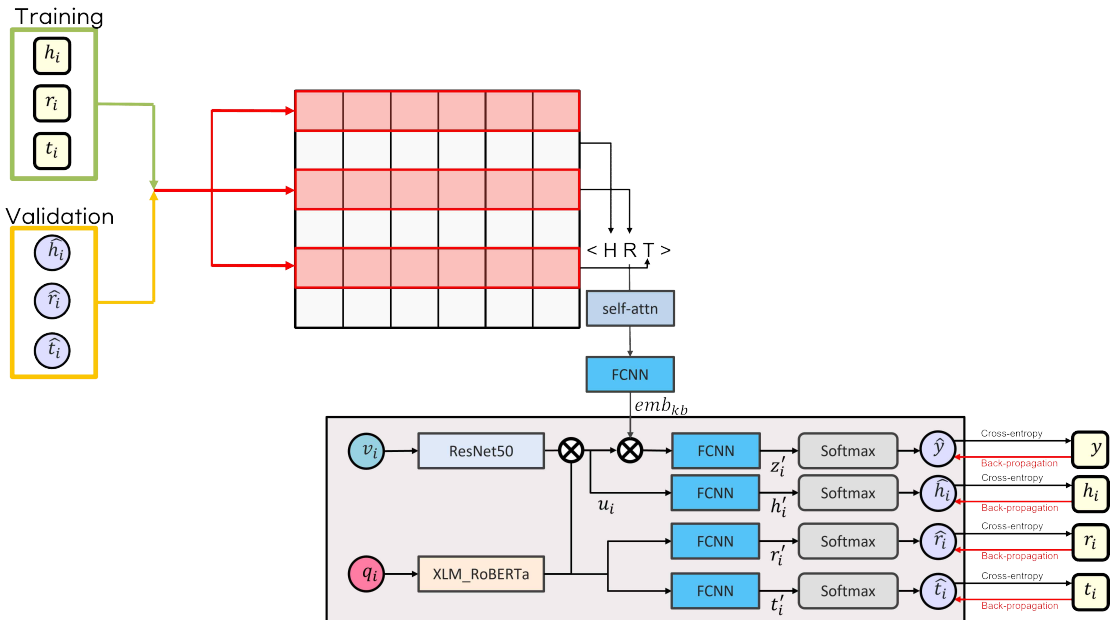
3-3) Triple Classification Module 구조



3-4) Real-Time Multi-Task Learning Model 구조



3-5) Teacher-Forcing Multi-Task Learning Model 구조



4. 구현하지 못한 기능 요구사항이 있다면 그 이유와 해결방안을 기술하시오.

| | | |
|---------|----------------------|-------------------------------------|
| 최초 요구사항 | 구현 여부(미구현, 수정, 삭제 등) | 이유(일정부족, 프로젝트 관리미비, 팀원변동, 기술적 문제 등) |
| | | 해당사항 없음 |

5. 요구사항을 충족시키지 못한 성능, 품질 요구사항이 있다면 그 이유와 해결방안을 기술하시오.

| | | |
|------------------------|-------------------|-------------------------------------|
| 분류(성능, 속도 등) 및 최초 요구사항 | 충족 여부(현재 측정결과 제시) | 이유(일정부족, 프로젝트 관리미비, 팀원변동, 기술적 문제 등) |
| | | 해당사항 없음 |

6. 최종 완성된 프로젝트 결과물(소프트웨어, 하드웨어 등)을 설치하여 사용하기 위한 사용자 매뉴얼을 작성하시오.

(작성요령: 여기에서 작성하는 사용자 매뉴얼은 개발한 시스템(환경)을 설치하여 사용할 수 있을 정도로 상세히 기술합니다)

Ubuntu 20.04.4 LTS

Python 3.8.10

CUDA 11.6

GPU : NVIDIA A100 80GB x 4

CPU : AMD EPYC 7352 24-Core Processor x 48

Memory : 126G

1. git clone https://github.com/mjkmmain/VQA_NIA.git

2. pip install -r requirements.txt

- 사용자 매뉴얼과 실행파일

모델 학습 방법

python main.py --[옵션]

- train_data : str, choice ['A', 'B', 'all'] (default : 'all')

- A : A형 질의 데이터

- B : B형 질의 데이터

- all : 모든 데이터 (A형 + B형 질의)

- n_epoch : int (default : 50)

- 에폭 수

- lr : float (default : 3e-5)

- 학습률

- batch_size : int (default : 512)

- 배치 크기

- max_token : int (default : 50)

- token max length

- use_transformer_layer : store_true

- 해당 옵션을 사용하면 멀티모달 표현 fusion 후 Transformer Encoder Layer를 추가합니다

예시 python main.py --n_epoch 30 --batch_size 128 --use_transformer_layer

추론 방법

python inference.py --[옵션]

model_path : str (default : "./results/all_1228_1734/infer_model_57.15.pt")

- 저장된 모델 경로

infer_data : str, choice ['all', 'abstract', 'triple', 'vqa'] (default : 'all')

- 추론하고자 하는 데이터 종류

예시

```
python inference.py --model_path ./results/all_1228_1734/infer_model_57.15.pt --infer_data vqa
```

7. 캡스톤디자인 결과의 활용방안

- 캡스톤 디자인을 통하여 완성된 프로젝트가 미치는 사회적/기술적/경제적 파급효과, 기대 효과 등을 자유롭게 기술함

외부지식을 활용하는 VQA 시스템은 사람들이 일상 생활에서 접하는 다양한 시각적 정보를 이해하고 분석하는 데 도움을 줄 수 있다. 특히, 시각적 장애인들에게는 이러한 시스템이 환경 인식 및 이해의 도구로서 중요한 역할을 한다. 또한, 어린 아이들의 학습에 제안하는 시스템을 활용하면, 아이들이 주변 환경을 이해하고 학습하는 것에 큰 도움이 될 것이다. 외부 지식을 활용하는 VQA시스템은 아이들이 책의 그림, 사진 등을 보며 질문을 제기할 때, 외부 지식을 통해 그 질문에 대한 답을 제공하여 아이들의 지식 확장에 도움이 될 것이다.

본 연구에서 제안하는 외부 지식을 활용하는 VQA 시스템의 발전은 기계학습, 컴퓨터비전, 자연어처리 등 다양한 인공지능의 분야에서 진보를 의미한다. 이는 궁극적으로 인공지능 기술의 전반적인 수준을 높이는 역할을 한다.

또한, 외부 지식을 활용하는 VQA시스템은 소비자와 기업 모두에게 가치를 제공할 수 있다. 소비자는 제품이나 서비스에 대한 질문에 대해 답변을 시각적으로 제공받고 정보를 더 잘 이해하고, 결정을 내릴 수 있을 것이다. 기업들은 이 시스템을 활용하여 고객 서비스를 개선하거나 새로운 서비스나 제품 개발에 기여할 수 있다. 이러한 VQA시스템의 보다 많은 발전과 활용은 우리 사회의 정보 접근성을 높이고, 정보의 이해를 돕는 것부터 산업 분야에서의 새로운 기회 창출에 이르기까지 다양한 혜택을 제공할 것으로 보인다.