
Fact 검증 기반 정답 재순위화를 이용한 다중추론 질의응답

이정두¹ · 나승훈¹ · 김태형² · 최윤수² · 장두성²
¹전북대학교, ²KT 융합기술원

Content

1. Introduction
2. Related Work
3. Proposal
4. Data
5. Experiment
6. Conclusion
7. Q & A



Introduction

- **질의 응답(Question Answering)**
 - 기계가 문맥을 이해하고, 주어진 질문에 대한 답을 찾는 태스크

Q: 피카소는 어느 나라 사람이야?



피카소는 ... 파리의 발전상을
고향 **스페인**에 알리는 등의 노력을 했다....



Introduction

- 다중 추론 질의응답(Multi-hop Question Answering)

Single hop QA	
Question	스티브 잡스가 CEO로 있던 회사는?
Answer	애플

Multi hop QA			
Bridge question		Comparison question	
Question	스티브 잡스가 CEO로 있던 회사의 현재 CEO는?	Question	스티브 잡스와 팀 쿡 중에 키가 더 큰 사람은?
Answer	팀 쿡	Answer	스티브 잡스

➤ Multi hop QA 는 2개 이상의 정보를 필요로 하는 QA



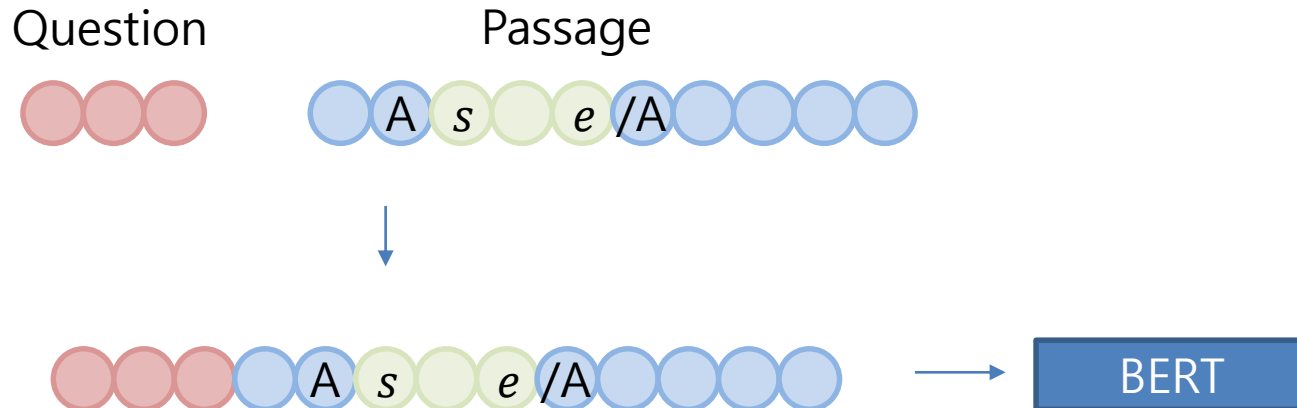
Related Work

- Open domain 다중 추론 질의응답(Multi-hop Question Answering)
최근 이슈
 1. 관련 문서를 찾기 어렵다.
 - Answering Complex Open-Domain Questions with Multi-Hop Dense Retrieval [Xiong et al 2021]
 - Learning to Retrieve Reasoning Paths over Wikipedia Graph for Question Answering[Asai et al 2019]
 - Simple yet Effective Bridge Reasoning for Open-Domain Multi-Hop Question Answering[Xiong et al 2019]
 - HopRetriever: Retrieve Hops over Wikipedia to Answer Complex Questions[Li et al 2020]



Related Work

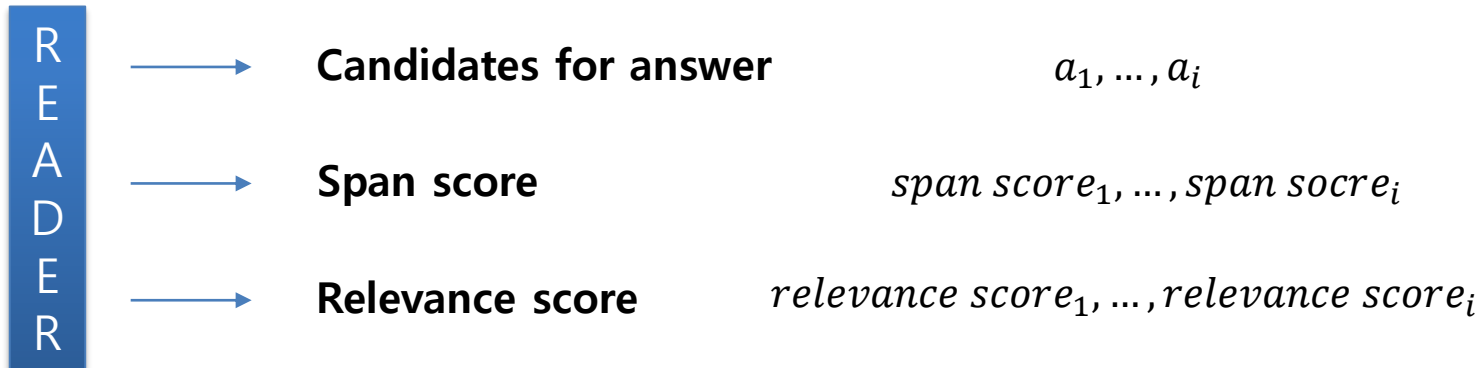
- Open domain 다중 추론 질의응답(Multi-hop Question Answering)
최근 이슈
 2. 동일 type의 정답 후보들 중 실제 정답을 선택하기 어렵다.
 - RECONSIDER: Re-Ranking using Span-Focused Cross-Attention for Open Domain Question Answering [Iyer et al 2020]



Proposal

- **IDEA**

- 질문과 정답 후보에 대해 Fact checking 을 하여 Reranking



Proposal

- **Novelty**

- 정답 후보를 사용하여 새롭게 검색을 진행
- QA 모델의 구조와 상관 없이 적용 가능
- Fact checking 분야가 연구될수록 성능 향상의 가능성



Proposal

- **Retriever**

- 기존 multi hop QA 에서 사용한 model 과 동일
- 정답 후보만 추가적으로 쿼리로 사용

- **Fact verification**

$$\text{fact score}_i = \mathbf{W}_0 \mathbf{u}_i$$

$$\mathbf{u}_i = \text{BERT}_{\text{mean}}(q; a_i; P_i)$$

where, $\mathbf{W}_0 \in \mathbb{R}^h$: learnable parameter

$P_i = \{p_0, p_1\}$: $q; a_i$ 로 얻은 passage pair

answer candidate set $A = \{a_0, \dots, a_i\}$

$$\text{Answer score}_i = \gamma * (\lambda * \text{relevance score}_i + (1 - \lambda) * \text{span score}_i) + (1 - \gamma) * \text{fact score}_i$$



데이터

- Multi hop QA 한국어 데이터

학습 데이터 (KorQuad 1.0)

다중추론 질문

Q: 바그너는 교향곡 작곡을 어디까지 쓴 뒤에 중단을 부활시켜 연주한 사람은? A: 리스트

첫 번째 단일추론 질문

Q: 바그너는 교향곡 작곡을 어디까지 쓴 뒤에 중단했는가?
A: 1악장

두 번째 단일추론 질문

Q: 1악장을 부활시켜 연주한 사람은? A: 리스트



데이터

- Multi hop QA 한국어 데이터

평가 데이터

다중추론 질문

Q: 애니메이션 '크리스탈요정 지스쿼드'를 만든 회사의 설립연도는? A: 2001년

사용된 위키피디아 문서

크리스탈 요정 지스쿼드: <<크리스탈요정 지스쿼드>>는 대한민국의 에네메스가 만들어 SBS에서 방송한 3D 애니메이션이다. 평범한 초등학생 소녀인 ...

에네메스: 주식회사 에네메스 (Enemes Co, Ltd.)는 대한민국의 애니메이션 회사다. 2001년에 설립했으며, 본사는 서울특별시 금천구 가산동에 있다.

...

데이터 통계

Train	Dev	Test
50,708	500	1,000



실험 결과

- Oracle performance of retriever for fact verification on test set

- 검색된 문서 집합 중 정답이 포함된 문서가 있을 경우에 대한 성능

입력	질문, 정답	질문
Hit@5(*2)	84.7	74.6
Hit@10(*2)	88.5	80.6
Hit@20(*2)	90.8	84.7

- 검색된 문서가 실제 추론에 필요한 문서일 경우에 대한 성능

입력	질문, 정답			질문		
	EM	First	Second	EM	First	Second
Hit@1(*2)	40.0	56.1	44.8	33.3	50.0	39.0
Hit@5(*2)	55.3	67.2	62.8	50.9	60.8	61.4
Hit@10(*2)	58.7	70.4	67.5	56.1	65.0	68.0
Hit@20(*2)	61.7	73.0	72.7	60.7	68.1	72.9



실험 결과

- Performance of Reranking using fact verification module on test set

Multi hop QA		Multi hop QA + Fact verification	
EM	F1	EM	F1
48.6	61.59	49.2	62.09

- 결론

- Fact checking task 를 Multi hop QA task 에 접목하여 QA 성능 향상
- Fact checking 분야가 더 연구되어 Model 이 정교해 지면 추가적인 성능 향상을 기대해 볼 수 있다.
- 향후 Multi hop 검색 모델을 개선하는 연구를 진행할 예정



Q & A



Thank You!

