

How Much Knowledge Can You Pack Into the Parameters of a Language Model?

EMNLP 2020

Tag: Closed-Book Question-Answering, T5

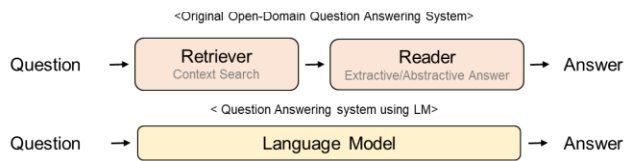
0. Summary

- Language model(T5)의 능력을 QA system을 통해 평가함

1. Introduction & Background

- Language model이 pre-trained 되고 나면, 일종의 "knowledge base"를 가짐

> Question-Answering System



- 기존 ODQA: Retriever-Reader 구조로 되어있음
- 본 논문: 언어 모델 안에 정답이 있다고 간주하고, Retriever 없이 정답을 도출 (**Closed-Book**)

2. Experiments

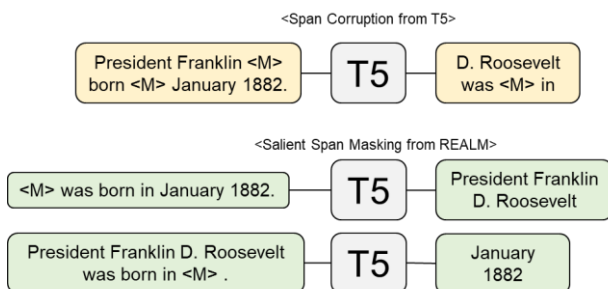
> Datasets

- Natural Questions (NQ), WebQuestions (WQ), TriviaQA (TQA) dataset사용
- 3개 Data 모두 Questions + Context로 이루어져 있음
- 본 논문에서는 Context를 제외한 Questions만 사용함

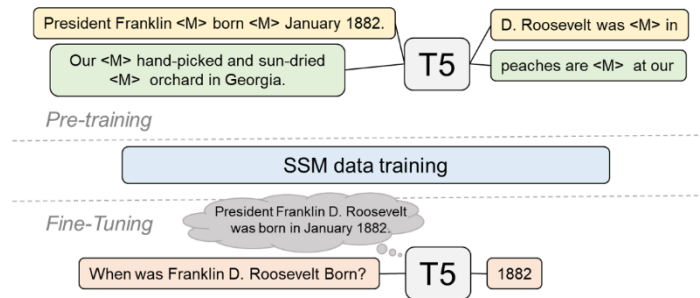
> Training, Model

- Size별: T5-Base(220M), Large(770M), 3B, 11B 사용
- Task별: T5, T5.1.1(pre-trained on unlabeled data only)
- 위 모델들에 QA dataset에 대한 Fine-Tuning 진행

> Salient Span Masking (SSM) (Guu et al., 2020)



- BERT로 Saliency Span(named entities and dates)을 가진 Sentence 탐색
- 15% masking을 진행할 때 "world knowledge"를 가진 대상을 학습 가능 → 위키피디아에서 이름, 날짜 들어간 데이터가 세계적인 지식을 갖고 있을 것이라는 가정



- T5 Fine-Tuning 단계 이전에 SSM data를 이용하여 Pre-training 추가

3. Result

- T5 모델 사이즈가 커질수록 성능 향상
- SSM을 활용 시 성능 향상 상당함
- T5.1.1-XXL + SSM에서 WQ dataset에 대해 SOTA

4. Human Evaluation

- T5는 Generation 모델이지만 Extraction Base Model과 비교하여 좋은 성능을 냄
- 의미는 같은데 형식이 다른 FN이 발생하지 않았을까?
- TN: 62%, Phrasing Mismatch: 13.3%, Incomplete Annotation: 13.3%, Unanswerable: 11.3%
- 성능 개선 여지 있는 것으로 보여짐

5. Conclusions

- T5 모델에는 많은 양의 세계 지식이 존재함을 확인

> Limitations

- 11B 모델에서만 SOTA를 달성함
- Open-book 모델은 질문에 답을 할 때 어떤 정보에 접근했는지 알 수 있지만 (Retrieved 된 문서 알 수 있음) 본 모델은 그렇지 않음 (해석 불가능)
- Maximum-likelihood objective가 사용되었기 때문에 사실을 학습했는지 알 수 없음
- Trivia style 지식을 학습했기 때문에, 추론 능력 필요한 것에 대한 성과 측정에 대한 실험이 필요함



질문 있으시면 정민지에게 DM 주세요!!