

문장 의미 비교를 활용한 독해 시스템 개선 방안 모색

한상도[○], 유환조, 이근배, 맹성현[§]
포항공과대학교, [§]한국과학기술원

{hansd, hwanjoyu, gblee}@postech.ac.kr, myaeng@kaist.ac.kr

Reading comparison study through sentence-level comparison

Sangdo Han[○], Hwanjo Yu, Gary Geunbae Lee, Sung-Hyon Myaeng[§]
Pohang University of Science and Technology,
[§]Korea Advanced Institute of Science and Technology

요 약

본 연구에서는 질의-본문 간 문장 비교 정보가 reading comprehension task 의 성능 향상에 도움이 되는지를 확인해 보았다. 기존의 reading comprehension 방법론이 질의-본문 간 의미 비교정보를 활용하지만, 본문 전체를 대상으로 한 비교이기 때문에 문장 단위의 정보가 활용되지 못하는 단점이 있었다. 실험에 사용한 데이터는 대표적인 RC 데이터 중 하나인 NewsQA[5] 를 이용하였으며, 질의-본문 문장 간 비교를 통한 성능 향상의 잠재력을 확인하였다.

주제어: 질의응답, 문장 임베딩, 질의응답, 독해, 문장 의미 비교, question answering, reading comprehension

1. 서론

독해(reading comprehension)는 질의응답(question answering)의 하위 분야 중에서도 가장 활발히 연구되고 있는 분야이다. 독해는 일반적인 질의응답과 같이 자연어 질의에 대한 답을 찾는 연구이지만 방대한 규모의 web 에서 정답을 찾는 것이 아닌, 작은 규모의 주어진 본문 내에서 답을 찾는 연구이다.

이러한 문제를 해결하기 위해서 다양한 방법론들이 제시되어 왔다[4,6]. 이러한 방법론들은 질의와 본문의 의미 유사도를 활용하여 정답을 추출하는 공통점을 가진다. 그러나 기존의 방법론들은 질의를 본문 전체와 비교하기 때문에 질의-본문 문장 간 의미 비교 정보를 활용하지 못하는 단점이 있다.

본 연구에서는 질의-본문 문장 비교 정보를 활용하기 위하여 본 연구에서는 기존의 방법론을 기본적으로 유지하되, 문장 비교 정보를 자질로서 추가하거나 문장 비교 네트워크를 추가하여 보았다. 실험 결과, 질의-문장 간 의미 비교를 통한 성능 향상 가능성을 확인할 수 있었으며, 차후 적용할 새로운 방안도 모색할 수 있었다.

2. 관련 연구

독해는 주어진 본문을 기반으로 질의의 정답을 출력하는 연구이다. 정답을 출력하는 방법은 본문으로부터 정답의 색인을 탐지하는 방법[3,5]과 본문의 내용을 기반으로 생성하는 방법[2]이 있는데, 본 논문에서는 정답에 해당하는 본문의 색인을 탐지하는 방법에 대해 연구하였

다. 정답 탐지의 유명한 데이터는 스탠포드 대학에서 공개한 SQuAD[3] 와 마이크로소프트 말루바(Maluuba) 에서 공개한 NewsQA[5] 가 있으며, 이들은 위키피디아 문서 혹은 뉴스 기사를 본문으로 만든 데이터이다.

본 연구에서는 NewsQA 데이터를 사용하였다. 이 데이터는 본문 내용을 기반으로 질의를 생성하는 다른 데이터와는 달리 뉴스 헤드라인을 기반으로 질의가 만들어졌다. 이로써 질의를 구성하는 단어가 본문 내 단어의 유의어로 구성되는 경우가 더 많기 때문에 질의와 본문의 형태가 더 상이한 특징을 가지고 있다.

NewsQA 를 비롯한 독해 문제를 해결하고자 많은 방법론들이 제시되어왔다. 이들의 공통점은 순차적 네트워크(sequential network)를 이용하여 질의-본문 간 비교 정보를 활용한다는 것이다. 대표적인 네트워크가 BiDAF[4] 시스템으로, 질의와 본문을 비교하여 정답의 앞 뒤 색인을 찾아낸다. 본 논문에서는 BiDAF 시스템과 유사한 fastQA[6] 네트워크를 기반으로 아이디어를 적용해 보았다. fastQA 는 주목기술(attention) 을 활용하여 질의를 embed 하고, 질의 embedding을 본문과 비교하여 정답을 찾는 구조를 가진다.

위의 방법론들은 질의-본문 간 비교 정보를 활용하지만 본문 전체를 대상으로 하기 때문에, 본 연구에서는 질의-본문 문장 간 비교 정보를 사용되지 않는 단점을 개선하고자 문장 비교 정보를 추가한 방법을 통해 독해 기술 향상 방안을 모색하였다. 기존의 문장 비교 활용 방법[1]의 경우 문장을 선택하는 네트워크를 따로 두어 two-step approach 으로 시도하였는데, 이는 문장 선택 단계에서의 성능에 최종 성능이 영향을 받기 때문에 향상

폭이 크지 않았다. 본 연구에서는 문장 비교 정보를 추가하는 end-to-end 의 간단한 방법으로 성능 향상이 되었음을 확인하였다.

3. 방법론

본 연구에서는 독해 기술에 질의-본문 간 문장 의미 비교를 적용하여 개선 가능성을 모색하였다. 새로운 방법론을 적용하기에 앞서 질의와 본문의 의미 비교 정보가 정답을 찾는 데 도움이 될 수 있는지 확인해보고자 의미 비교를 통한 정답 문장 랭킹을 시도해 보았다. Long short term memory model(LSTM)과 max pooling 을 활용하여 질의와 본문의 문장을 encoding 하고, 질의와 정답 문장 간의 유사도(similarity)가 높아지도록 훈련하여 정답 문장 랭킹을 해 보았다. 그 결과, 문장 embedding 을 활용한 정답 문장 랭킹 결과가 유의미하였고(그림 1), 이는 질의-본문 간 문장 비교 정보가 정답을 찾는 데 도움이 될 것이라는 가설을 뒷받침해 주었다.

정답 문장 랭킹 결과

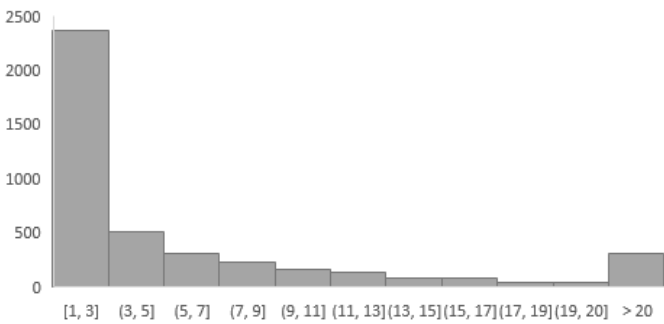


그림 1. 질의-본문 문장 간 유사도를 이용한 정답 문장 랭킹 결과. x 축 : 정답 문장의 랭킹 결과, y 축 : 정답 문장이 해당 랭킹에 해당하는 본문의 수. MRR=0.424

fastQA 는 위에서 언급한 대로 질의와 본문의 비교를 통해 정답의 앞 뒤 색인을 찾아내는 네트워크이다. 그림 2의 구조는 fastQA 의 구조를 기반으로 가공한 구조로서, 질의-본문 문장 간 의미 비교를 추가하였다. 해당 구조는 기존의 fastQA 구조에 정답 문장 선택 모듈을 추가한 것으로서, 추가된 구조는 질의 문장 embedding 과 본문 문장 embedding 간 의미 유사도를 기반으로 정답 문장을 분류하는 네트워크이다. 해당 네트워크는 두 가지 방법으로 훈련하여 보았는데, 정답 문장 분류 단을 pre-train 한 후 기존의 정답 탐지 네트워크를 훈련하는 방법과 정답 문장 분류 loss 를 정답 탐지 loss 에 더하여 훈련하는 방법이었다. 해당 네트워크의 성능을 보기에 앞서 기존의 구조에 질의-본문 문장 비교 자질을 추가한 구조도 함께 훈련해 보았다. 이 때 비교 자질은 질의 embedding 과 본문 embedding 의 hadamard product 벡터를 사용하였다.

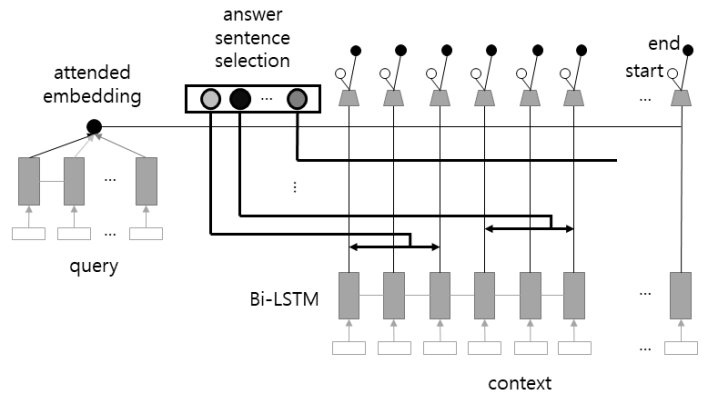


그림 2. FastQA 에 answer sentence selection network 를 추가한 network architecture. 굵은 선이 answer sentence selection network를 의미한다.

4. 실험 및 결과

실험 결과는 표 1 과 같았다. 기존 네트워크 구조에 문장 비교 자질을 추가한 경우, f1과 em 점수가 함께 증가하는 것을 확인하였다. 정답 탐지 네트워크를 추가한 방법의 경우 em 점수가 상대적으로 많이 향상했지만 f1 점수가 하락하는 현상이 관찰되었다.

System	F1	EM
Baseline	48.04	29.93
Baseline + 자질 추가	48.26	30.62
정답 문장 선택 구조 (pretrain)	46.45	30.95
정답 문장 선택 구조 (loss 합)	46.88	32.09

표 1. NewsQA 데이터에 정답 문장 선택 정보를 활용한 결과

훈련 시 batch size 는 16으로 하였고, 그 외의 기본 훈련 환경은 fastQA 의 조건을 따랐다. 다만 정답 문장 선택의 경우 stochastic gradient descent algorithm 으로 pre-train 하였으며, 이 때 learning rate 는 1로 설정하였다. 단, 본 논문에서는 character embedding 이 적용되지 않은 FastQA를 baseline 시스템으로 사용하였다.

5. 결론

기존의 독해 방법론은 순차적 네트워크를 활용하여 본문 전체에서 정답을 검색하였다. 우리는 질의와 비슷한 문장에 답이 있을 확률이 높다는 점에 착안하여 문장 단위의 의미 비교를 추가한 방법론을 시도해 보았다. 문장 비교 벡터를 기존의 방법론에 자질 벡터로써 사용한 방법론과, 기존의 네트워크에 문장 비교 네트워크 단을 추가하는 간단한 방법론을 시도하였다.

실험 결과, 질의-본문 간 문장 비교 정보를 사용하는 간단한 방법으로 성능 향상을 확인하였으며, 추후 자질 개선과 네트워크 구조 개선을 통한 발전을 기대할 수 있었다. 자질 벡터를 추가한 경우 전반적인 성능 향상이 있었으나 향상 폭이 크지 않아 자질 벡터의 개선이 필요해 보였으며, 정답 탐지 네트워크를 추가한 방법의 경우 em 점수가 더 많이 상승하는 것을 관찰 할 수 있었다. 후자의 경우 f1 점수가 감소하는 것으로 보아 개선된 오류의 특성을 확인할 수 있었는데, 이는 기존 방법론이 정답을 매우 크게 탐지하는 오류가 줄어든 것으로 보인다.

후속 연구에서는 위의 장점을 결합한 개선안을 마련하는 것을 포함하여 기존의 embedding 기반의 문장 비교가 아닌, 다른 방식의 비교 방식을 활용하거나 순차적 네트워크 비교 기반이 아닌 새로운 구조의 새로운 네트워크를 설계해 볼 계획이다.

Acknowledge

"본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 Grand ICT연구센터지원사업의 연구결과로 수행되었음" (IITP-2018-2015-0-00742)

참고문헌

- [1] Min, S., Zhong, V., Socher, R., & Xiong, C. "Efficient and Robust Question Answering from Minimal Context over Documents." ACL, 2018.
- [2] Nguyen, Tri, et al. "MS MARCO: A human generated machine reading comprehension dataset." NIPS, 2016.
- [3] Rajpurkar, Pranav, et al. "Squad: 100,000+ questions for machine comprehension of text." EMNLP, 2383-2392p, 2016.
- [4] Seo, Minjoon, et al. "Bidirectional attention flow for machine comprehension." ICLR, 2017.
- [5] Trischler, Adam, et al. "Newsqa: A machine comprehension dataset." ReplANLP, 191-200p, 2017.
- [6] Weissenborn, Dirk, Georg Wiese, and Laura Seiffe. "Making neural QA as simple as possible but not simpler." CoNLL, 271-280p, 2017.