

## 본문-질의 비교를 활용한 오답 질의 분류

한상도<sup>○</sup>, 유환조, 이근배, 맹성현<sup>§</sup>

포항공과대학교, <sup>§</sup>한국과학기술원

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

### Incorrect query classification via context-query comparison

Sangdo Han<sup>○</sup>, Hwanjo Yu, Gary Geunbae Lee, Sung-Hyon Myaeng<sup>§</sup>

Pohang University of Science and Technology,

<sup>§</sup>Korea Advanced Institute of Science and Technology

#### 요약

본 논문은 딥 러닝 기반의 독해 기술이 풀지 못하는 문제를 분류해내는 기술에 관한 것이다. 해당 연구에서는 독해 데이터 및 시스템 결과 분석을 통해 시스템이 풀지 못하는 문제들의 특징을 도출해내고, 이에 알맞은 전략들을 시도해 보았다. 분석 결과에 따른 시도들은 각 목적에 부합하는 결과를 나타냈으며, 특히 독해 기술의 특징에 기반한 방법론이 효과적이었다. 본 논문에서 제안하는 방법은 본문과 질의 간 유사도 행렬을 활용하는 것으로, 기존의 독해 기술이 본문과 질의의 유사도를 활용하여 정답을 내는 것에 영감을 얻었다.

**주제어:** 딥 러닝, 독해 기술, 분류 기술

#### 1. 서론

독해 기술은 자연어 형태의 본문과 질의를 입력으로 받아 적절한 답을 출력하는 기술이다. 이 기술은 여러 가지 방법으로 발전되어 왔는데, 사용 가능한 지식 베이스에 알맞게 질의를 변형하는 방법[1] 그리고 본문과 질의의 임베딩을 활용한 딥 러닝 기반의 방법[2] 등이 있다. 특히 딥 러닝을 활용한 독해 기술은 최근 가장 높은 성능을 보이며 다양하게 활용되고 있다. 그러나 이 방법은 여러 단계의 추론이 필요한 질의나 상식 질문 등에 취약한 경향을 보인다. 본 연구에서는 다양한 독해 기술 간의 연계를 통한 성능 향상을 꾀하고자, 딥 러닝 기반의 독해 기술이 잘 풀지 못하는 문제들을 분류하는 시도를 해 보았다.

#### 2. 관련 연구

독해 문제는 최근에 각광받고 있는 연구 주제 중 하나로서, 위에서 언급하였다시피 자연어 본문과 질의를 입력으로 받아 이에 알맞은 정답을 출력하는 기술이다. 독해 문제는 SQuAD[4], newsQA[5], triviaQA[3] 등 다양한 공개된 데이터가 존재하며, 각 문제들은 풀고자 하는 문제의 특징이 있다. 예를 들어 newsQA의 경우 본문과 질의가 상이한 형태의 문제를 풀고자 했으며 triviaQA는 여러 문서를 아우르는 질문에 대한 답을 하고자 만들어진 데이터이다.

triviaQA는 워싱턴대학교에서 공개한 독해 데이터셋으로, 여러 문서를 아우르는 질의에 대답을 하고자 만들어진 것이 특징이다. 650,000쌍의 질의-정답-본문을 제공하여 딥 러닝 기반의 방법론을 적용할 수 있을 만큼의 규모를 제공한다. 이 데이터는 인터넷에 올라와 있는 질의

와 답을 기반으로, 질의를 가공한 검색어로 산출한 웹 검색 결과물을 본문으로 추가하였다.

documentQA[2] 시스템은 워싱턴대학교와 allenAI사에서 공개한 오픈 소스 시스템으로, 딥 러닝 기반의 독해 시스템이다. 본 시스템은 여러 문서를 본문으로 하는 문제에 적합한 방법론을 적용한 시스템으로, 본 연구에서 기준이 되는 독해 시스템으로 선정하였다. 이 논문은 해당 시스템이 주어진 질의를 풀어내지 못하는 질의들을 분류하는 분류기를 연구한 것의 기록이다.

#### 3. 방법 및 결과

이 연구에서는 documentQA 시스템이 triviaQA 데이터, 특히 triviaQA 중 wikipedia 를 본문으로 하는 질의를 대상으로, 주어진 질의를 대상으로 documentQA 시스템이 오답을 낼 지를 판별하는 분류기를 연구하였다.

분류기를 제작하기에 앞서, documentQA가 풀지 못하는 질의들을 정성적으로 분석하였고, 세 가지 특징을 관찰하였다. 첫째, 본문과 질의에서 사용하는 어휘가 상이하였다. 딥 러닝 기반의 독해 기술은 본문과 질의의 단어의 유사도를 활용하기 때문에 이를 활용하지 못하는 경우에 취약점을 보였다. 둘째, 특정한 질의 형식을 가지고 있었다. 딥 러닝 기반의 독해 기술은 보통 단문의 형태로 특정 사실을 묻는 데이터를 대상으로 한다. 그러나 여러 후보군을 주고 그 안에서 답을 찾게 하는 질의 혹은 주어진 정보에서 누락된 정보를 찾게 하는 등의 질의에는 취약하다는 것을 확인할 수 있었다. 셋째, 본문 내에 정답이 구조화된 정보 내에 존재하는 경우이다. 예를 들어 정답이 표에 존재하거나 표를 해석해야 하는 경우에는 훈련 데이터가 많지 않기 때문에, 그리고 기존의 방법론이 그러한 문제에 적합한 형태가 아니기 때문에

취약한 모습을 나타내었다.

본 연구에서는 이러한 문제를 해결하기 위해 다양한 방법론을 시도해 보았다. 최종 방법론을 소개하기에 앞서, 먼저 시도한 세 가지 방법에 대해 소개하겠다. 가장 먼저 시도한 방법은 기초 성능을 위한 방법으로, 무작위 추출이다. 두 번째 방법은 기존의 독해 기술이 본문과 질의 간 의미 비교를 활용하기 때문에 시도해 본 방법으로, 로지스틱 회귀 방법을 사용했다. 데이터에서 직접 자질을 추출한 후 그 자질로 분류하는 방법이었다. 자질은 본문에 질의의 단어가 얼마나 나오는지, 본문의 문장 중 질의와 단어가 겹치는 최대 비율은 무엇인지 등을 0에서 1사이의 스케일로 계산하여 사용하였다. 세 번째 방법은 독해 기술이 특정한 질의 형태에 취약한 모습을 보였기 때문에 시도한 방법으로, 질의 임베딩을 활용하였다. pre-trained 된 GloVe 임베딩을 LSTM 층에 거쳐 pooling 하는 방법이다. 위의 정성적 분석에서 언급했듯 독해 기술이 취약한 형태의 질의 중 사용하는 어휘가 정해져 있는 경우가 있었기 때문에 (which is 등) 어휘 정보를 가장 잘 활용할 수 있도록 max-pooling으로 최종 벡터를 생성하였다.

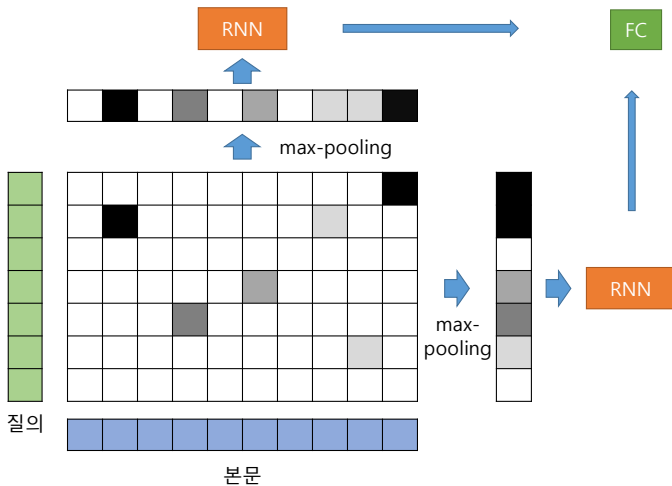


그림 1 유사도 행렬을 활용한 질의 분류 방법

본 연구에서 제안하는 방법론은 유사도 행렬을 활용한 방법(그림 1)으로, 딥 러닝 기반의 독해 기술에서 얻은 영감으로 적용한 방법이다. 딥 러닝 기반 독해 기술이 본문과 질의의 단어 의미 비교를 활용하여 정답을 찾았던 것과 같이, 본 방법에서는 본문과 질의의 유사도 행렬을 만들고, 이를 인코딩 하여 분류하였다. 유사도 행렬에 max-pooling 을 적용하여 유사도 벡터를 만들고, 유사도 벡터들은 recursive neural network(RNN)을 활용하여 인코딩하였다. 인코딩 된 두 벡터는 결합한 후 fully-connected network(FC)를 활용하여 분류하였다. 각 시도의 성능은 표 1에 나와 있으며, 성능은 negative prediction value를 사용하였다. 이는 딥 러닝 기반의 독해 시스템이 일반적으로 다른 독해 방법론에 비해 성능이 높기 때문에, 풀지 못하는 문제만을 다른 시스템에 전달하도록 하기 위함이었다.

유사도 행렬 방법론에 사용한 단어 임베딩은 GloVe 300차원 데이터를 사용했으며, RNN의 hidden은 400차원, 학습률은 0.0001이었다. 각 방법론의 성능은 표 1에 기입되어 있으며, 유사도 행렬 방법론이 가장 높은 성능을 기록하였다.

	Negative prediction value
무작위	0.41
로지스틱 회귀	0.54
질의 임베딩	0.59
유사도 행렬	0.65

표 1 방법론에 따른 질의 분류 성능

#### 4. 고찰

표 1과 같이 유사도 행렬을 활용한 방법론의 성능이 가장 높았던 이유는 두 가지로 정리된다. 첫째, 딥 러닝 기법을 활용하여 최적화된 자질을 활용할 수 있었다는 점이다. 로지스틱 회귀와 같이 수작업으로 생성한 자질은 최적화에 방해가 되기 때문에 유사도 행렬을 생성한 후 자동으로 자질을 추출해내도록 하는 방식이 효과적이었다. 둘째, 본문과 질의를 함께 활용하였다는 점이다. 질의 임베딩을 활용한 방법은 로지스틱 회귀 방법론보다는 성능이 개선되었지만, 독해 기술이 기본적으로 본문과 질의 간의 의미 비교를 주로 하기 때문에 이 자질을 활용하였을 때 성능이 더 좋아지는 것을 확인할 수 있었다. 추후 연구에서는 anomaly detection과 같이 풀리지 않는 질의의 데이터가 많지 않을 때에도 적용 가능한 방법론을 모색할 계획이다.

#### Acknowledge

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2013-0-00179, (엑소브레인-3세부) 컨텍스트 인지형 Deep-Symbolic 하이브리드 지능 원천 기술 개발 및 언어 지식 자원 구축)

#### 참고문헌

[1] Berant, Jonathan, et al. "Semantic parsing on freebase from question-answer pairs." Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. 2013.  
 [2] Clark, Christopher, and Matt Gardner. "Simple and Effective Multi-Paragraph Reading Comprehension." Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018.  
 [3] Joshi, Mandar, et al. "TriviaQA: A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension." Proceedings of the 55th Annual Meeting

of the Association for Computational Linguistics (Volume 1: Long Papers). 2017.

[4] Rajpurkar, Pranav, et al. "SQuAD: 100,000+ Questions for Machine Comprehension of Text." Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016.

[5] Trischler, Adam, et al. "NewsQA: A Machine Comprehension Dataset." Proceedings of the 2nd Workshop on Representation Learning for NLP. 2017.