

# 딥러닝과 정보검색을 결합한 질의응답 시스템

이현구<sup>○</sup>, 김학수

강원대학교 컴퓨터정보통신공학과

nlphglee@kangwon.ac.kr, nlpdrkim@kangwon.ac.kr

## Question Answering System that Combines Deep Learning and Information Retrieval

Hyeon-gu Lee<sup>○</sup>, Harksoo Kim

Kangwon National University Computer and Communication Engineering

### 요 약

정보의 양이 빠르게 증가함으로 인해 필요한 정보만을 효율적으로 얻기 위한 질의응답 시스템의 중요도가 늘어나고 있다. 그 중에서도 질의 문장에서 주어와 관계를 추출하여 정답을 찾는 지식베이스 기반 질의응답 시스템이 활발히 연구되고 있다. 그러나 기존 지식베이스 기반 질의응답 시스템은 하나의 질의 문장만을 사용하므로 정보가 부족한 단점이 있다. 본 논문에서는 이러한 단점을 해결하고자 정보검색을 통해 질의와 유사한 문장을 찾고 Recurrent Neural Encoder-Decoder에 검색된 문장과 질의를 함께 활용하여 주어와 관계를 찾는 모델을 제안한다. bAbI SimpleQuestions v2 데이터를 이용한 실험에서 제안 모델은 질의만 사용하여 주어와 관계를 찾는 모델보다 좋은 성능(정확도 주어:33.2%, 관계:56.4%)을 보였다.

주제어: 질의응답 시스템, Recurrent Neural Encoder-Decoder, 지식베이스, 정보검색

### 1. 서론

질의응답 시스템(Question ANswering System)은 자연어로 구성된 질의를 통해 사용자가 원하는 답변을 자동으로 생성해 주는 기술이다. 최근 다양한 정보가 빠른 속도로 증가하고 있지만 제한된 시간 안에 모든 정보를 습득하기 어려운 문제가 있어 질의응답 시스템의 중요성이 나날이 늘고 있다. 그런 이유로 IBM Watson[1], TREC(Text Retrieval Conference)[2], BioASQ[3] 등 다양한 분야에서 질의응답 시스템에 관한 연구가 활발히 진행되고 있다. 질의응답 시스템은 정보검색을 통해 문장을 찾고 검색된 문장에서 정답을 추출하는 정보검색 기반 질의응답 시스템[4]과, 질의에 해당하는 주어(Subject), 관계(Relation)를 찾고 관계 트리플(주어\_Subject, 관계\_Relation, 목적어\_Object)로 구성된 지식베이스에서 정답을 찾는 지식베이스 기반 질의응답 시스템[5]이 있다. 기존의 지식베이스 기반 연구는 규칙을 통해 주어 어휘와 관계 어휘를 추출하고 관계 트리플에 사상하였지만 유의어나 다른 형태로 나타난 문장의 경우 관계 트리플에 사상시키기 어려운 단점이 있었으며, 기계학습 방법으로 추출하기에는 질의문장 하나는 정보가 너무 적은 단점이 있었다. 본 논문에서는 지식베이스 기반 방법의 단점을 보완하고자 딥러닝을 통한 기계학습 방법에 질의와 유사한 문장을 정보검색 방법으로 찾고 질의와 함께 사용하여 성능을 향상시키는 모델을 제안한다.

### 2. 관련 연구

기존의 질의응답 시스템은 질의와 문장사이의 유사성

을 이용하기 위한 기계학습 방법[6], 질의에 해당하는 관계트리플을 찾고 질의와 관계 트리플을 신경망(Neural Network)을 통해 벡터화 시켜 유사도를 측정하는 방법[7] 등이 연구되었다. 최근 들어 순환 신경망(Recurrent Neural Network)을 적용하여 문장의 순서정보를 반영해 정답을 찾는 모델이 제안되었다. 대표적으로 BiGRU(Bi-directional Gated Recurrent Unit)와 BiGRU-CRF(BiGRU-Conditional Random Field)를 통해 주어와 관계를 찾는 모델[8], Sequence to Sequence 모델을 통해 위키피디아 문서를 읽어 정답을 찾는 모델[9]이 있다. 본 논문에서는 Recurrent Neural Encoder-Decoder[10]를 이용하여 문장을 인코딩 한 후 인코딩 벡터를 통해 주어와 관계를 추출하는 모델을 제안한다.

### 3. 딥러닝과 정보검색을 결합한 질의응답 시스템

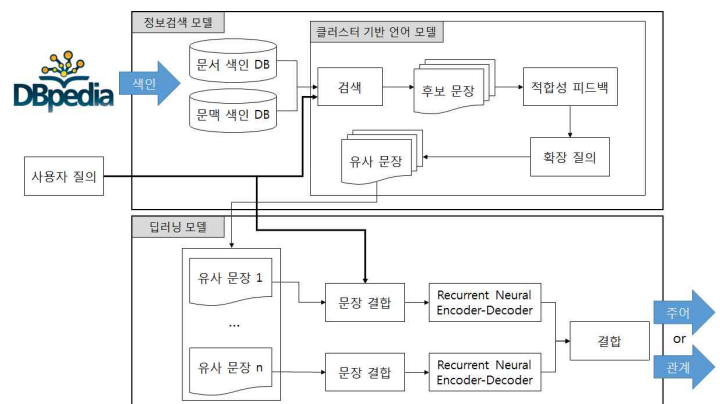


그림 1 제안 모델의 구조도

[그림 1]은 제안 모델의 전체 구조도를 보여준다. [그림 1]에서 보는 것과 같이 제안 모델은 정보검색을 통해 유사 문장을 검색하는 부분과 딥러닝을 통해 질의에 해당하는 주어와 관계를 찾는 부분으로 구성된다. 정보검색을 통해 얻어지는 유사 문장은 오류가 포함될 수 있어 클러스터 기반 언어 모델을 사용하여 검색 결과의 오류를 줄이고 질의와 검색된 유사 문장을 입력으로 모델을 학습시켜 주어와 관계를 찾는다.

### 3.1. 정보검색을 통한 유사 문장 검색

본 논문에서는 효과적으로 유사 문장을 찾기 위해 문서 단위의 색인과 문맥 단위의 색인을 한다. 문서 단위의 색인은 검색 대상이 되는 문서의 모든 내용을 색인하는 방법으로 질의와 유사 문장 사이의 어휘 불일치 문제를 완화하는 역할을 한다. 문맥 단위의 색인은 문서를 문장 단위로 분할한 후 각각의 문장과 문장 기준 앞, 뒤 문장을 같이 색인하는 방법으로 유사 문장 주위의 문맥 정보를 검색에 반영하기 위한 것이다.

질의에 해당하는 유사 문장을 찾기 위한 검색 모델은 적합성 피드백(Relevance feedback)[11]을 적용한 클러스터 기반 언어모델(Cluster based language model)[12,13]을 사용한다. 적합성 피드백은 질의 검색을 통해 예비 문서를 찾고 예비 문서 집합에서 나타나는 중요 단어들을 추가하여 검색결과의 성능을 향상시키는 방법이다. 클러스터 기반 언어모델은 식 (1)과 같다.

$$Sim_{IR}(Q, S) = \alpha Sim_{context}(Q, S) + (1 - \alpha) Sim_{doc}(Q, D) \quad \text{식 (1)}$$

식 (1)에서  $Q$ 는 질의,  $S$ 는 유사 문장,  $D$ 는 유사 문장  $S$ 를 포함하는 문서를 의미한다.  $Sim_{context}$ 와  $Sim_{doc}$ 은 언어 모델을 문맥 색인과 문서 색인에 적용하여 얻어진 질의와 문맥, 문서간의 유사도이다.  $\alpha$ 는 두 검색모델간의 가중치이며 본 논문에서는 0.8로 설정하였다. 식 (1)에서 보는 것과 같이 문맥 색인과 문서 색인에서 얻어진 유사도를 결합하여 질의와 유사 문장 사이의 최종 유사도를 계산한다. 클러스터 기반 언어 모델을 통해 질의와 유사 문장 사이의 어휘 불일치 문제를 해결할 뿐만 아니라 주변 문맥 정보 역시 반영이 되어 기존의 문장 검색 방법보다 높은 품질의 문장을 얻을 수 있다.

### 3.2. 딥러닝 기반 질의응답 시스템

본 논문에서는 Recurrent Neural Encoder-Decoder 모델을 사용하여 주어와 관계를 찾는다. Recurrent Neural Encoder-Decoder 모델은 입력된 문장을 Neural Encoder를 통해 인코딩하여 문장의 의미가 담긴 중간벡터를 생성하고 인코딩된 중간벡터 값을 Neural Decoder에 전달하여 결과를 도출하는 모델이다. [그림 2]는 Recurrent Neural Encoder-Decoder 모델의 예시이다.

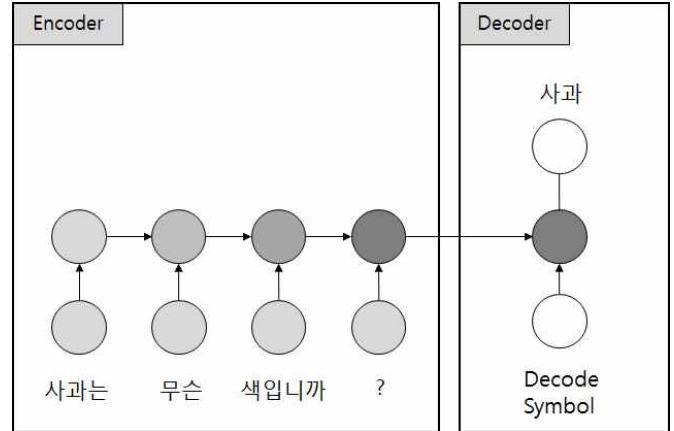


그림 2 Recurrent Neural Encoder-Decoder의 예

[그림 2]와 같이 예시문장 “사과는 무슨 색입니까?”를 입력하면 각 어절의 단어들이 순서 정보가 반영되어 인코딩된 중간벡터가 생성된다. 생성된 중간벡터를 디코딩하여 문장 정보가 모두 반영된 정답 “사과”를 찾게 된다. 본 논문에서는 Recurrent Neural Encoder-Decoder에 질의와 유사 문장 집합을 입력하여 주어와 관계를 찾는 모델을 제안한다.

[그림 3]은 본 논문에서 제안하는 질의응답 시스템의 구조도이다.

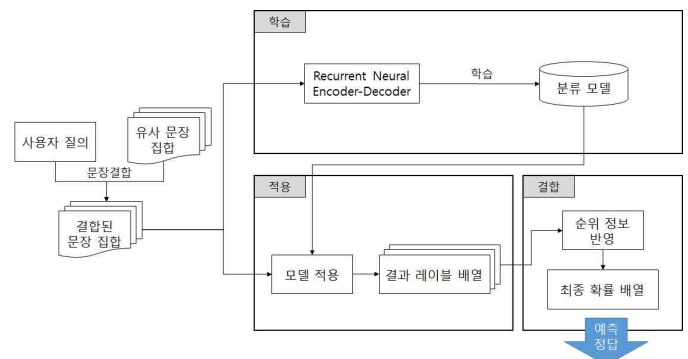


그림 3 딥러닝 기반 질의응답 시스템 구조도

[그림 3]과 같이 질의와 유사 문장 집합에 속하는 문장을 이용하여 학습하는 학습부분, 질의와 유사 문장 집합을 학습된 모델에 적용하여 결과 레이블 배열을 얻는 적용부분, 마지막으로 각 유사 문장별 결과 레이블 배열에 정보검색 순위정보를 반영하여 결합한 후 최종 정답을 분류하는 결합부분으로 구성된다.

먼저 질의로 검색된 유사 문장 집합의 각 문장을 질의와 연결한 질의-유사 문장 쌍을 입력으로 사용하고 정답을 출력으로 하여 학습한다. 예를 들어 질의에서 주어를 찾은 문제일 경우 질의가 “사과는 무슨 색입니까?”이고 검색된 유사 문장 집합이 “사과는 빨간색 과일로 과즙이 풍부하다.”, “사과는 세계적으로 가장 널리 재배되는 과일이다.”로 이루어져 있을 때 학습의 입력으로 사용하기 위한 질의-유사 문장 쌍은 “사과는 무슨 색입니까?<EOS>사과는 빨간색 과일로 과즙이 풍부하다.”, “사과는 무슨 색입니까?<EOS>사과는 세계적으로 가장

널리 재배되는 과일이다.”가 되고 출력은 “사과”가 된다. 위와 같은 방법으로 입력 데이터를 생성 후 학습된 모델에 적용하여 각 정답 레이블의 확률 값을 가지는 결과 레이블 배열을 얻는다. 마지막으로 질의와 유사 문장들로 얻은 결과 레이블 배열들에 정보검색 순위결과 가중치를 반영하여 각 요소(각 레이블의 확률 값)를 더한 최종 확률 배열을 생성한 후 확률이 가장 높은 레이블 하나를 예측 값으로 출력한다.

[그림 4]는 본 논문에서 제안하는 모델의 적용부분과 결합부분의 예시이다.

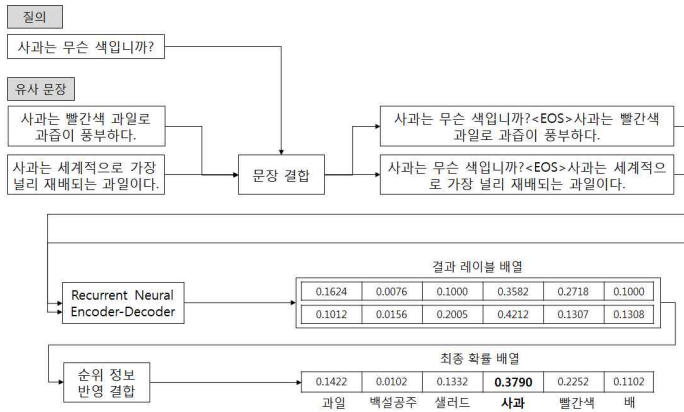


그림 4 적용 및 결합의 예

[그림 4]와 같이 질의와 유사 문장들을 학습된 모델에 적용시켜 결과 레이블 배열을 출력하고 각 요소에 정보검색 순위정보 가중치를 반영하여 합한 최종 확률 배열을 계산한다. 마지막으로 최종 확률 배열에서 가장 높은 확률을 가지는 요소를 가져오며 [그림 4]의 예시는 네 번째 요소의 단어인 “사과”를 출력하게 된다.

## 4. 실험 및 평가

### 4.1. 실험 준비

본 논문에서는 정보검색 모델의 색인 데이터로 DBpedia 2015-10[14]의 영어 위키피디아 abstracts만으로 이루어진 long\_abstracts.ttl을 사용한다. 질의응답 시스템 실험용 데이터로는 질의와 FreeBase\_5MB[15]에 존재하는 관계 트리플의 쌍으로 이루어진 bAbI SimpleQuestions v2[16]에서 학습 데이터 10,000개, 평가 데이터 1,000개를 무작위 추출하여 사용한다. 정보검색 모델을 위한 색인에 앞서 문단으로 이루어진 abstracts를 문맥 단위로 색인하기 위해 LingPipe[17]를 통해 문장 분할을 시행하였다.

학습을 위한 Recurrent Neural Encoder-Decoder의 파라미터는 <표 1>과 같다.

표 1. 모델 파라미터

파라미터	값
학습률(Learning rate)	0.1
Decay factor	0.99
Mini-batch size	64
Hidden Layer 개수	3
Layer 차원 수	1024

### 4.2. 실험 결과

표 2. 주어와 관계 모델에서의 성능 비교

모델	정확도(Accuracy)	
주어	Baseline	27.4%
	제안 모델	<b>33.2%</b>
관계	Baseline	51.3%
	제안 모델	<b>56.4%</b>

<표 2>은 제안 모델의 성능을 보여준다. <표 1>에서 Baseline은 제안 모델의 성능을 확인하기 위해 질의만 입력으로 했을 때 주어나 관계가 나오도록 학습시킨 모델이다. <표 2>에서 보는 것과 같이 본 논문에서 제안한 정보검색을 통해 검색된 유사 문장을 같이 사용한 주어, 관계 모델의 성능이 Baseline보다 각각 5.8%, 5.1% 향상되는 것을 볼 수 있다. 그러나 전체 성능이 낮은 것을 확인 할 수 있는데 주어는 80,990개, 관계는 1,777개중 하나를 찾는 문제라 분류할 내용이 많아 성능이 낮다고 판단된다.

### 4.3. 오류 분석

본 논문에서 제안하는 모델을 통해 발생하는 오류는 다음과 같다.

#### 1) 의미는 비슷하나 다른 단어

- 주어 Seoul이 주어 Tokyo로 분류되는 경우 : 수도라는 의미는 비슷하나 결과적으로 다른 단어

#### 2) 의미는 같으나 범주가 다른 단어

- 관계 film/genre가 관계 music/album/genre로 분류되는 경우 : 장르라는 의미를 지니고 있으나 범주가 달라 의미가 달라진다. 문장에서 film과 music을 구별해 줄 힌트가 나타나지 않는 경우 잘못된 분류를 하게 된다.

#### 3) 잘못된 분류

- 의미적, 형태적으로 전혀 다른 단어로 분류가 되는 경우

이러한 오류는 두 가지 이유로 인해 발생하는 것으로 추측된다. 먼저 신경망 모델을 사용하였기에 단어 표현

이 의미적 특성을 따라가 의미는 비슷하나 다른 단어로 인식돼 (오류 1), (오류 2)와 같은 문제가 발생한다. 또 다른 이유로 정보검색 모델을 통해 얻어진 유사 문장 집합에 오류가 많을 경우 (오류 3)과 같이 완전히 잘못 분류되는 문제가 발생하게 된다.

## 5. 결론 및 향후 연구

본 논문에서는 정보검색을 통해 검색된 유사 문장과 Recurrent Neural Encoder-Decoder를 사용하여 질의에서 묻고자하는 주어와 관계를 찾는 방법을 제안하였다. 본 논문에서 제안한 모델은 질의를 검색하여 유사 문장 집합을 찾고 검색된 결과와 질의를 통해 주어와 관계를 찾았다. 실험 결과 질의만을 통해 주어와 관계를 찾는 모델보다 주어, 관계 분류 정확도가 각각 5.8%, 5.1% 향상되는 결과를 얻었다. 또한 실험 데이터는 영어지만 자질이 들어가지 않고 어휘만을 이용한 언어 독립적인 방법을 제안하였기에 한국어로 구성된 학습 데이터만 구축되면 한국어에도 적용이 가능 할 것으로 보인다. 향후 연구로 주어와 관계의 타입 정보를 통해 단어 임베딩을 추가하고, Sequence Labeling 방법을 통해 정답 가능성이 있는 단어를 추론하는 가지치기 모델을 추가할 계획이다.

## 감사의 글

본 연구는 LG전자 산학연구용역 과제의 지원을 받아 수행되었음. 또한 우수기술연구센터 사업 중 “링크드데이터 기반 대화형 질의응답 검색 프레임워크 개발 (과제 번호 : 10048448)” 과제의 지원을 받아 수행되었습니다.

## 참고문헌

- [1] A. Gliozzo, O. Biran, S. Patwardhan and K. McKeown, “Semantic Technologies in IBM Watson,” *Proceedings of the Fourth Workshop on Teaching NLP and CL*, pp. 85-92, 2013.
- [2] E. Agichtein, D. Carmel, D. Harman, D. Pelleg and Y. Pinter, “Overview of the TREC 2015 LiveQA Track,” *Proceedings of The Twenty-Fourth Text REtrieval Conference(TREC 2015). National Institute of Standards and Technology(NIST)*, 2015.
- [3] G. Tastsarionis, G. Balikas, P. Malakasiotis, I. Partalas, M. Zschunke, M. R. Alvers, D. Weissenborn, A. Krithara, S. Petridis, D. Polychronopoulos, Y. Almirantis, J. Pavlopoulos, N. Baskiotis, P. Gallinari, T. Artières, Axel-Cyrille N. Ngomo, N. Heino, E. Gaussier, L. Barrio-Alvers, M. Schroeder, I. Androutsopoulos, and G. Paliouras, “An Overview of the BIOASQ Large-Scale Biomedical Semantic Indexing and Question Answering Competition,” *BioMed Central Bioinformatics*, vol. 16, no.1, Apr. 2015.
- [4] H. Lee, M. Kim, H. Kim, J. Kim, S. Kwon, J. Seo, J. Choi and Y. Kim, “KSAnswer: Question-answering System of Kangwon National University and Sogang University in the 2016 BioASQ Challenge,” *Proceedings of BioASQ workshop at the BioNLP workshop of Association for Computational Linguistics 2016*, 2016.
- [5] C. Unger, L. Bührmann, J. Lehmann, A. N. Ngomo, D. Gerber and P. Cimiano, “Template-based Question Answering over RDF Data,” *Proceedings of the 21<sup>st</sup> international conference on World Wide Web*, ACM, pp. 639-648, 2012.
- [6] L. Yu, K. M. Hermann, P. Blunsom and S. Pulman, “Deep learning for answer sentence selection,” *arXiv preprint arXiv:1412.1632*, 2014.
- [7] A. Bordes, S. Chopra and J. Weston, “Question Answering with Subgraph Embeddings,” *arXiv preprint arXiv:1406.3676*, 2014.
- [8] Z. Dai, L. Li and W. Xu, “CFO: Conditional Focused Neural Question Answering with Large-scale Knowledge Bases,” *arXiv preprint arXiv:1606.01994*, 2016.
- [9] D. Hewlett, A. Lacoste, L. Jones, I. Polosukhin, A. Fandrianto, J. Han, M. Kelcey and D. Berthelot, “WikiReading A Novel Large-scale Language Understanding Task over Wikipedia,” *arXiv preprint arXiv:1608.03542*, 2016.
- [10] K. Cho, B. V. Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk and Y. Bengio, “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” *arXiv preprint arXiv:1406.1078*, 2014.
- [11] J. J. Rocchio, “Relevance feedback in information retrieval,” 1971.
- [12] X. Liu and W. B. Croft, “Cluster-Based Retrieval Using Language Models,” *Proceedings of The 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, pp. 186-193, 2004.
- [13] A. Merkel and D. Klakow, “Comparing Improved Language Models for Sentence Retrieval in Question Answering,” *LOT Occasional Series*, vol. 7, pp. 35-50, 2007.
- [14] DBPedia version 2015-10, Available from World Wide Web: <http://wiki.dbpedia.org/dbpedia-dataset-version-2015-10>.
- [15] K. Bollacker, C. Evans, P. Paritosh, T. Sturge and J. Taylor, “Freebase: a collaboratively created graph database for structuring human knowledge,” *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*,

ACM, 2008.

- [16] A. Bordes, N. Usunier, S. Chopra and J. Weston,  
“Large-scale simple question answering with  
memory networks,” *arXiv preprint*  
*arXiv:1506.02075*, 2015.
- [17] B. Baldwin and B. Carpenter, “LingPipe,”  
Available from World Wide Web: <http://alias-i.com/lingpipe>.