

# Seq2SPARQL: 신경망 기계 번역을 사용한 지식 베이스 질의 언어 자동 생성

홍동균, 심홍매, 김광민

(주)솔트룩스

e-mail: [dghong@saltlux.com](mailto:dghong@saltlux.com)

## Seq2SPARQL: Automatic Generation of Knowledge base Query Language using Neural Machine Translation

Dong-Gyun Hong, Hong-Mei Shen, Kwang-Min Kim  
Saltlux Inc.

## 요약

SPARQL(SPARQL Protocol and RDF Query Language)은 지식 베이스를 위한 표준 시맨틱 질의 언어이다. 최근 인공지능 분야에서 지식 베이스는 질의 응답 시스템, 시맨틱 검색 등 그 활용성이 커지고 있다. 그러나 SPARQL과 같은 질의 언어를 사용하기 위해서는 질의 언어의 문법을 이해하기 때문에, 일반 사용자의 경우에는 그 활용성이 제한될 수밖에 없다. 이에 본 논문은 신경망 기반 기계 번역 기술을 활용하여 자연어 질의로부터 SPARQL을 생성하는 방법을 제안한다. 우리는 제안하는 방법을 대규모 공개 지식 베이스인 Wikidata를 사용해 검증하였다. 우리는 실험에서 사용할 Wikidata에 존재하는 영화 지식을 묻는 자연어 질의-SPARQL 질의 쌍 20,000 건을 생성하였고, 여러 sequence-to-sequence 모델을 비교한 실험에서 합성곱 신경망 기반의 모델이 BLEU 96.8%의 가장 좋은 결과를 얻음을 보였다.

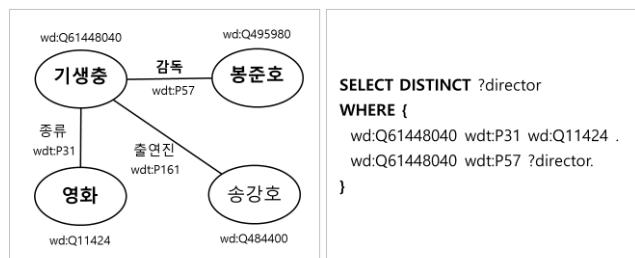
## 1. 서론

지식 베이스는 정형 데이터로써, 질의 응답이나 챗봇 같은 자연어 처리 분야에서 비정형 데이터와 함께 중요하게 사용된다. 대표적인 지식 베이스로는 Wikidata<sup>1</sup>, DBpedia<sup>2</sup>가 있으며, 주 활용 분야인 지식 베이스 기반 질의 응답 시스템에 대한 연구도 활발히 진행되고 있다[1,2].

지식 베이스는 지식의 집합이고, 지식은 개체 간의 관계(속성) 혹은 개체와 특정 값 사이의 관계를 하나의 사실로 표현한 것이다. 이때 사실은 <주어, 술어, 목적어> 트리플 형태를 가진다. 예를 들면 “대한민국 수도는 서울”이라는 하나의 사실에서 주어는 대한민국이고 술어는 수도이며 목적어는 서울이다.

지식 베이스는 다양한 형태로 구현할 수 있다. 그 중 대표적인 구현이 W3C에 의한 시맨틱 웹 표준을 사용한 구현이다. 이 경우 지식의 표현은 RDF, OWL이 쓰이며, 지식 베이스 질의 언어의 표준은 SPARQL이다<sup>3</sup>.

지식 베이스의 질의 언어 SPARQL은 관계형 데이터 베이스의 질의 언어인 SQL과 동일한 역할을 한다.

자연어 질의 영화 기생충의 감독은 누구야?

Knowledge Base (Wikidata)

SPARQL

(그림 1) Wikidata SPARQL 예시

SPARQL을 사용하면 지식 베이스의 지식을 조회하거나 생성, 삭제를 할 수 있다. 따라서 지식 베이스를 사용하기 위해서는 SPARQL의 기본적인 문법을 파악하는 것이 필수적이다.

지식 베이스 기반의 질의 응답 (KBQA) 시스템은 자연어 질의에 대한 답을 지식베이스에서 찾아 반환하는 시스템이다. KBQA 시스템은 답을 얻기 위해 자연어 질의가 요구하는 답을 찾을 수 있는 SPARQL을 생성하고 지식베이스에 질의를 해야 한다. 이 과정에서 기존 KBQA 시스템들은 주로 시맨틱 파싱과 템플릿 매칭 방법을 사용한다[1]. 그리고 최근 연구들은 딥러닝을 활용하기도 한다[2]. 그러나 전자의 방법들은

<sup>1</sup> <https://www.wikidata.org/><sup>2</sup> <https://wiki.dbpedia.org/><sup>3</sup> <https://www.w3.org/standards/semanticweb/>

파싱이나 템플릿 생성을 위해 수작업이 많이 필요한 문제가 있고, 후자의 방법들은 비교형이나 관정형 질의에 대해서는 아직 어려움을 겪고 있다.

이처럼 SPARQL 을 일반적으로 활용하거나, 질의 응답 시스템에서 사용하는 과정에서 자연어 질의를 SPARQL 로 변환하는 작업은 문법 지식이 필요하고 수작업이 요구된다. 이에 본 논문에서는 최근의 번역 분야에서 좋은 성과를 보여주고 있는 신경망 기계 번역 모델[3,4,5]을 활용하여 자연어 질의로부터 SPARQL 을 자동 생성하는 방법을 제안한다.

우리는 한글 자연어 질의로부터 SPARQL 을 자동 생성하는 방법을 소개하고, 검증을 위해 Wikidata 지식 베이스를 활용한 실험을 수행하였다. 우리는 영화 도메인의 한글 질의와 SPARQL 의 쌍 20000 건의 데이터를 생성해 실험에 사용하였다. 우리는 실험을 통해 신경망 기계 번역 모델이 실제 질의 가능한 SPARQL 을 생성하는 것을 확인하였으며, 합성곱 신경망 기반의 신경망 기계 번역 모델이 BLEU 96.8%의 성능으로 가장 좋았음을 보였다.

## 2. 제안 방법

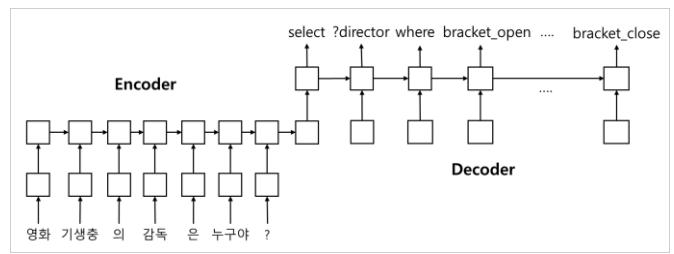
### 2.1 신경망 기계 번역

신경망 기계 번역 모델은 인공 신경망을 사용하여 소스 언어를 타겟 언어로 번역하는 모델을 의미한다. 특히 encoder-decoder 구조를 갖는 sequence-to-sequence 모델 [3]이 기계 번역을 포함하여, 음성 인식과 요약 분야에서도 좋은 성과를 보여주면서, 해당 구조가 신경망 기계 번역 모델에 일반적으로 사용되고 있다. 이러한 신경망 기계 번역은 기존 통계 기반 번역보다 더 좋은 성능을 보여주어 상용 번역 시스템에서 주요하게 활용되고 있다 [6].

과거 sequence-to-sequence 모델 [3]에서는 encoder 와 decoder 로 순환 신경망을 사용해왔다. 그러나 순환 신경망은 병렬로 학습이 어려워 속도의 한계를 보이는데, 최근에는 학습 속도가 빠르면서 성능 또한 더 우수한 모델들이 등장하였다 [4,5]. 먼저 [4]은 순환 신경망 대신 합성곱 신경망을 사용하여 sequence-to-sequence 모델을 구현한 것으로, 병렬 학습이 가능한 합성곱 신경망의 장점을 활용한 모델이다. 그리고 [5]는 순환 신경망은 사용하지 않고 attention 만을 이용하여 encoder-decoder 구조를 만들어 학습 속도를 개선한 모델이다. 이들은 학습 속도의 한계를 개선했을 뿐만 아니라 실제 번역의 품질도 우수하다는 것을 보였다.

### 2.2 SPARQL 자동 생성

본 논문에서는 신경망 기계 번역 모델을 사용해 한글 자연어 질의를 SPARQL 로 번역하는 것을 제안한다. 최근 신경망 기계 번역 모델은 언어 간의 번역만이 아니라, 관계형 데이터베이스로의 번역에 관한 연구가 활발해지고 있다[7]. 이와 더불어 SPARQL 로의 질의 파싱에 관한 연구도 시작 단계에 있으며 [8], 본 연구에서는 이 연구를 더 확장해보고자 한다.



(그림 2) 신경망 기계 번역 기반 SPARQL 생성

신경망 기계 번역 모델 기반의 SPARQL 자동 생성은 그림 2 와 같이 진행된다. 한글 자연어 질의가 입력되면, 모델은 영문으로 인코딩 된 SPARQL 문을 출력한다. 이후 출력 영문을 다시 디코딩하면 실제 지식 베이스에 질의할 수 있는 SPARQL 이 완성된다. 본 연구에서는 기존 연구 [8] 가보인 순환 신경망 기반의 모델을 한글에 적용해보는 작업을 진행했다. 그리고 더 나아가서 최근의 더 좋은 번역 성능을 보여준 [4,5]의 모델을 활용하여 SPARQL 생성에 적용하고자 한다.

## 3. 실험 및 결과

### 3.1 실험 환경

신경망 기계 번역 모델 기반의 SPARQL 생성 성능을 평가하기 위하여 지식 베이스 기반의 한글 자연어 질의-SPARQL 쌍 20000 건을 생성하여 사용하였다. 지식 베이스로는 Wikidata 를 사용하였고 SPARQL 은 Wikidata 기준으로 작성하였다. 자연어 질의는 영화, 영화 감독, 배우와 관련된 지식을 묻는 질문으로 구성하였다. 질문의 형태는 단일 사실을 묻는 질문과 함께 두 개체를 비교하는 질문과 개체의 개수 등을 묻는 질문 등 20 개 유형이 존재한다. 각 유형 별로 1000 개의 질문을 생성했는데, 각 질문은 서로 다른 개체에 관한 질의이다. 해당 자연어 질의는 모두 실제 Wikidata 에 지식이 존재하여 질의가 가능하도록 생성하였다. 자연어 질의와 대응하는 SPARQL 의 예는 그림 2 와 같다.

기계 번역 학습을 용이하게 하기 위하여 기존 연구[8]에서처럼 SPARQL 에 존재하는 특수 기호(대괄호, 중괄호)와 변수 등을 그대로 사용하지 않고 영문으로 치환하여 학습 데이터를 구성하였다. 추후 모델의 출력 영문을 다시 원래대로 치환하면 실제 Wikidata 에 질의할 수 있도록 하였다. SPARQL 과 영문의 치환은 [8]의 것을 기반으로 수행했고, 한글에서 개체의 마지막 글자의 받침 유무에 따라 질문의 조사가 달라져야 하는 부분을 고려할 수 있도록 개선했다.

실험에서 사용한 신경망 기계 번역 모델은 Tensorflow 기반의 NMT<sup>4</sup>와 Pytorch 기반의 fairseq<sup>5</sup>의 것을 사용하였다. 우리는 NMT 를 사용하여 순환

<sup>4</sup> <https://github.com/tensorflow/nmt>

<sup>5</sup> <https://github.com/pytorch/fairseq>

신경망 기반의 모델을 실험하였다. nmt-rnn 은 [8]에서 사용한 기준 모델로써 단방향 순환 신경망을 2 개의 레이어로 구성하였고, hidden unit 의 수는 128 개로 하였다. 그리고 nmt-rnn\_att 은 어텐션 메커니즘을 추가한 모델이다. fairseq 에서는 합성곱 신경망 기반의 모델[4]과 Transformer[5]를 사용했다. 두 모델은 fairseq 에서 다양한 구조와 파라미터를 갖도록 구현되어 있는데, 실험에서는 번역 분야에서 좋은 성능을 보여준 기본 구조인 fconv 와 transformer 를 각각 사용하였다.<sup>6</sup>

### 3.2 실험 결과

<표 1> 한글 질의-SPARQL 번역 정확도 및 실제 동작률

	BLEU (%)	Accuracy (%)
nmt-rnn	83.8	66.8
nmt-rnn_att	85.9	66.9
fairseq-fconv	96.8	90.3
fairseq-transformer	58.6	70.35

실험은 20000 건 데이터 중 80%를 학습 데이터로 활용하고, 10%를 검증 데이터, 나머지 10%를 테스트 데이터로 사용하였다. 실험 결과는 <표 1>과 같다. <표 1>의 BLEU 점수는 번역 모델이 생성한 문장이 정답 문장과 얼마나 유사한지를 계산하는 방법이고, 정답 문장과 동일한 문장을 번역한 경우의 BLEU 점수는 100(%)이다.

실험 결과 합성곱 신경망 기반의 fairseq-fconv 모델이 96.8%로 가장 높은 BLEU 점수를 보여주었다. 이는 번역 분야의 결과와 동일하게 합성곱 신경망 기반의 모델의 번역 품질이 더 좋은 경향으로 설명할 수 있다. 그러나 기대와 다르게 fair-transformer 모델이 58%로 가장 낮은 BLEU 점수를 보여주었다. 이는 일부의 결과를 살펴봤을 때, 다른 모델들과 비교하여 정답 SPARQL 의 길이보다 짧은 출력을 내는 경향을 보였기 때문으로 보인다. 이 결과는 추가적인 실험을 통해 더 상세한 분석을 할 예정이다.

SPARQL 은 기존 언어 간의 번역과는 다르게 BLEU 가 아무리 높아도 문법적인 오류가 있다면 생성한 SPARQL 은 실제로 동작하지 않는다. 예를 들어 중괄호 하나만 빠지더라도 문법 오류로 인해 생성한 SPARQL 은 동작하지 않는다.

따라서 실제 모델이 생성한 SPARQL 을 실제 Wikidata SPARQL endpoint 에 질의하여 정상적인 SPARQL 을 얼마나 생성할 수 있는지 평가하였다. <표 1>의 Accuracy 는 각 모델이 생성한 2000 건의 SPARQL 중 실제 동작하는 SPARQL 의 비율이다. 동작률 역시 fairseq-fconv 모델이 가장 높은 결과를 보였다. 전반적으로는 번역 품질인 BLEU 가 높은 것과 비교해서 실제 동작하는 SPARQL 의 비율은 더 낮은 것을 확인 할 수 있었다.

<sup>6</sup> <https://github.com/pytorch/fairseq/tree/master/fairseq/models>

### 4. 결론 및 향후 연구 방향

본 논문에서는 신경망 기계 번역 모델을 활용한 SPARQL 자동 생성 방법을 제안하였다. 우리는 합성곱 신경망 기반의 모델을 사용하여 높은 BLEU 를 얻을 수 있었다. 그러나 생성한 SPARQL 中 실제 동작하는 비율은 높지 않았다.

우리는 생성한 SPARQL 의 실제 동작률이 낮은 경향을 확인하였고, 이는 SPARQL 문법을 고려하여 개선할 수 있을 것으로 기대한다. 따라서 향후 연구에서는 신경망 기계 번역 모델을 연구들처럼[9] SPARQL 의 문법을 decoder 가 고려할 수 있는 모델을 활용해볼 것이다. SPARQL 의 문법은 복잡하지 않기 때문에, 기본적인 문법 제약을 고려하는 것만으로도 실제 동작하는 SPARQL 생성 비율은 더 높아질 것이라고 예상한다. 나아가서, 본 연구를 지식 기반 질의응답 시스템에 실제로 적용하여, 기존 시맨틱 파싱 기반 혹은 딥러닝 기반의 질의 응답 시스템과의 성능을 비교하는 연구를 진행할 계획이다.

### Acknowledgement

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2013-0-00109, WiseKB: 빅데이터 이해 기반 자가학습형 지식베이스 및 추론 기술 개발)

### 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] Yin, Wenpeng, et al. "Simple question answering by attentive convolutional neural network." arXiv preprint arXiv:1606.03391(2016).
- [3] Cho, Kyunghyun, et al. "Learning phrase representations using RNN encoder-decoder for statistical machine translation." arXiv preprint arXiv:1406.1078 (2014).
- [4] Gehring, Jonas, et al. "Convolutional sequence to sequence learning." Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017.
- [5]. Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems. 2017.
- [6] Wu, Yonghui, et al. "Google's neural machine translation system: Bridging the gap between human and machine translation." arXiv preprint arXiv:1609.08144 (2016).
- [7] Zhong, Victor, Caiming Xiong, and Richard Socher. "Seq2sql: Generating structured queries from natural language using reinforcement learning." arXiv preprint arXiv:1709.00103 (2017).
- [8] Soru, Tommaso, et al. "Neural Machine Translation for Query Construction and Composition." arXiv preprint arXiv:1806.10478 (2018).
- [9] Lin, Kevin, et al. "Grammar-based Neural Text-to-SQL Generation." arXiv preprint arXiv:1905.13326 (2019).