

뉴로-심볼릭 구조 기반 온톨로지 생성기 제안

박형철^{1,+}, 윤은수^{1,+}, 김민정², 배희재¹, 신유진³, 이지항^{4,*}

¹상명대학교 융합전자공학과 학사과정

²상명대학교 전기공학과 학사과정

³상명대학교 휴먼지능정보공학과 석사과정

⁴상명대학교 휴먼지능정보공학 교수

(+: 공동 1 저자, *: 교신저자)

cchptr204@gmail.com, hkhk0331@gmail.com, minjeong7656@gmail.com, baehj100@naver.com,
lynn2867@naver.com, jeehang@smu.ac.kr

Developing the Deep Text-to-Ontology Generator based on Neuro-Symbolic Architecture

Hyeoung-Cheol Park¹, Eun-Su Yun¹, Min-Jeong Kim², Hui-Jae Bae¹, Yu-Jin Shin³, Jee-Hang Lee^{4,*}

¹Dept. of Converged Electronics Engineering, Sangmyung University

²Dept. of Electrical Engineering, Sangmyung University

³Dept. of AI & Informatics, Sangmyung University

⁴Dept. of Human-Centered Artificial Intelligence, Sangmyung University

요 약

본 논문은 뉴로-심볼릭 구조를 바탕으로 일반 텍스트로부터 온톨로지 생성이 가능한 심층 신경망 기반 온톨로지 추출기를 제안한다. 온톨로지 추출 단계를 (i) 온톨로지 학습 및 (ii) 온톨로지 생성의 2 단계로 상정, (i) 일반 텍스트로부터 문장 구조 및 논리적 관계를 학습하는 트랜스포머 기반 심층 생성 신경망 출력을 이용하여 (ii) 계층적으로 결합한 심볼릭 추론기로 온톨로지를 생성하는 뉴로-심볼릭 구조 온톨로지 추출기를 구현하였다. 1800 개 훈련 집합으로 학습 후 200 개 테스트 집합으로 평가한 결과, 정확도 91.9%, Precision 100%, Recall 99.1%로 비교 모델 OpenIE의 성능에 비해서 각각 83.8%, 1.8%, 3.5% 개선된 것을 확인하였다. 정성적 품질에 있어서, 복잡한 문장 (예: 관계대명사, 접속사, 중첩 구조)에서도 비교 모델에 비해 더 정밀한 온톨로지 생성 결과를 보였다.

1. 서론

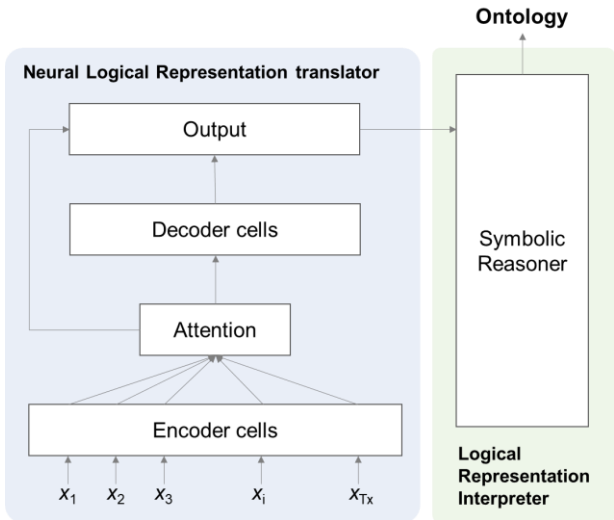
온톨로지 (Ontology)는 지식의 의미론적 표현으로 객체 사이의 관계를 컴퓨터가 처리 및 이해할 수 있도록 표현해 놓은 합의된 지식들의 연결을 의미한다. 일반적으로 새로운 지식에 대한 온톨로지 생성은 각 영역 전문가들의 수작업에 의존하는 바, 시간과 비용이 많이 소요되었다 [3]. 최근 심층 신경망 발전에 따른 신경 기계 번역 성능이 높아짐에 따라, 온톨로지 추출 기술을 일반 텍스트로부터 온톨로지를 번역하는 문제로 치환하거나, 단대단 방식 온톨로지 생성 기술이 제안되고 있다 [1].

이에 발맞추어, 본 논문에서는 일반 텍스트로부터 온톨로지 생성이 가능한 뉴로-심볼릭 구조 기반 온톨로지 추출기를 제안한다. 온톨로지 추출 단계를 (i) 온톨로지 학습 및 (ii) 온톨로지 생성의 2 단계로 구분하여, 먼저 일반 텍스트로부터 문장 구조 및 논리적 관계를 학습하여 구조화된 문장을 재생성하는 트랜스포머 기반 심층 생성 신경망에, 계층적으로 결합한 심볼릭 추론기로 온톨로지를 생성하도록 하였다.

2. 뉴로-심볼릭 구조 기반 온톨로지 생성기

본 논문이 제안하는 생성기는 Seq2Seq 기반 문장 논리 표현 생성 신경망[2]과 심볼릭 추론기의 결합으로 이루어져 있으며, 구조는 그림 1 과 같다. 임의의 문장에서 Head, Relation, Tail 로 구성된 온톨로지 트리플렛(Triplet)과 지식 그래프를 출력하며, 한 문장에 접속사, 관계 대명사 등 복잡한 문장 구조를 지닌 텍스트로부터, 한 문장에서 추출 가능한 모든 객체를 온톨로지 표현이 가능하다.

문장 논리 표현 생성 신경망에 문장이 입력되면, 입력 문장을 논리 표현 (Logical representation) 형태로 변환한다. 변환된 논리문장은 단어들 간의 관계가 논리 기호로 표현되어 단어 간 관계를 쉽게 파악할 수 있다. 이를 활용하여 심볼릭 추론기는 논리 기호 간의 정의된 규칙을 통해 문장 내의 단어들 간의 관계를 head - relation - tail 형태의 트리플렛으로 해석한다. 심볼릭 추론기는 논리 표현에 있는 논리 기호들을 기준으로 기호 전과 후에 있는 단어들을 추출하여 Head,-Relation-Tail 관계로 정렬한 후 출력한다.



(그림 1) 뉴로-심볼릭 구조 기반 온톨로지 생성기 구조도

알고리즘 1 은 심볼릭 추론기의 동작을 보여준다. 논리 기호 중 ‘:=’ 를 기준으로 좌항을 head, 이후 우항의 술부를 relation, tail 로 정하였다. Relation 의 경우, 술부 내 기호 ‘E’ 혹은 ‘A’ 와 ‘:=’ 사이에 나오는 단어들로 정하였고, tail 의 경우 Relation 이후 단어들로 정하였다. 마지막으로 포함 관계를 나타내는 논리 기호들을 고려하여, 한 문장에서의 여러 relation 에 대해 모든 tail 들을 결합할 수 있도록 하였다.

ALGORITHM 1: SYMBOLIC REASONER

```

Input: sentence, Seq2Seq Model
Output: Ontology triplets
1 logic sentence <- Seq2Seq(sentence)
2 logic symbols index <- find_logic_symbol(logic_sentence)
3 for word in logic sentence do //find head, relation, tail in logic sentence
4   if (word.index < logic symbols index [“:=”]) then
5     head ← word
6   else if ((word.index < logic symbols index [“E” or “A”]) and
7           (word.index > logic symbols index [“:=”]) then
8     if word = be verb then
9       append ‘be’ to relation set
10    else then
11      append word to relation set
12    End
13  else then
14    append word to tail set
15 End
16 combine head with relations and tails in each set
    
```

3. 실험 결과

제안된 모델을 이용하여 온톨로지를 생성하고, 온톨로지 생성 분야 대표 모델 OpenIE [4] 결과와 비교하여 성능을 측정하였다. 성능 측정용 데이터셋은 <문장, 논리 formula>쌍으로 구성된 2k-closed 데이터셋 [2]을 사용하였으며, 2000 쌍의 데이터 중 1800 쌍을 학습, 200 쌍을 테스트에 사용하였다. 표 1 에서 보는 바와 같이, 제안 기술이 Accuracy, Precision, Recall 에서 우월한 성능을 보였다.

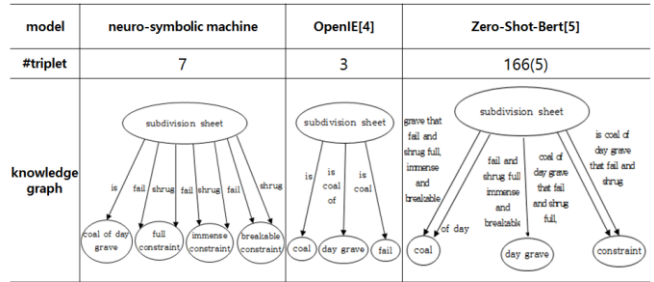
그림 2 는 관계 대명사와 접속사로 이루어진 복잡한 구조의 예시 문장을 입력하였을 때 생성된 온톨로지 트리플렛을 보여준다. 예시문장에 대하여 OpenIE 는 that 이후의 문장 정보는 추출하지 못하였으며,

Zero-Shot-Bert[5]는 각 단어의 관계를 서술하는 relation 의 정보를 올바르게 추출하지 못하였다. 반면, 본 논문에서 제안한 기술은 복잡한 문장 구조 내의 모든 정보를 반영하여 단어 간의 관계를 올바르게 추출한 것을 확인할 수 있다.

<표 1> 온톨로지 생성 성능 비교

	ACCURACY	PRECISION	RECALL
제안기술	0.919	1.0	0.991
OPENIE	0.081	0.982	0.956

Sentence: subdivision sheet are coal of day grave that fail and shrug full, immense and breakable constraint.



(그림2) 제안 기술의 온톨로지 생성 사례

4. 결론

본 논문에서는 일반 텍스트로부터 문장 구조를 학습하여 논리적 표현으로 생성하는 심층 생성 신경망과 심볼릭 추론기를 계층적으로 결합한 뉴로-심볼릭 구조 기반 온톨로지 생성기를 제안하였다. 제안한 모델을 통해 입력된 일반 텍스트로부터 다양한 정보를 추출하여 온톨로지 지식 구조로 변환한 결과를 확인하였고, 이를 비교모델에서 추출한 온톨로지 결과와 비교하여 제안한 모델의 우수성을 확인하였다. 그러나, 문장의 의미론적 정보 추출 가능 범위가 [2]에서 제공한 데이터에 제한되는 바, 추후 벤치마크용 데이터셋을 적용하여 제안 모델을 학습한 후 그 결과를 확인해 볼 예정이며, 최신 연구들에서 제시하는 지표를 통해 제안 모델의 성능을 확인해 볼 예정이다.

Acknowledgement

이 논문은 2023 년도 정부재원(과학기술정보통신부 여대학원생공학연구팀제 지원사업; 협약번호 WISSET-2023-125 호)으로 과학기술정보통신부와 한국여성과학기술인육성재단의 지원을 받아 수행되었으며, 과학치안진흥센터 (인공지능과 클라우드를 활용한 아동목격자 맞춤형 비대면 진술조서 지원 시스템 개발; 과제번호 RS-2023-00281194)의 지원을 받아 수행되었음.

참고문헌

[1] 신유진 외, “기계번역 개념을 활용한 Text-to-Ontology 변환 방법 제안”, ACK 2021(추계) 학술 발표대회, 여수, 2021, 891p.
 [2] Giulio Petrucci *et al.*, “Expressive ontology

- learning as neural machine translation” Journal of Web Semantics, Vol. 52-53, pp. 66-82, 2018.
- [3] Judith Reitman Olson *et al.*, “Extracting expertise from experts: Methods for knowledge acquisition, Expert Syst. (ISSN: 1468-0394) 4 (3) pp.152-168 1987.
- [4] Stanford NLP Group. “Stanford NLP Group Official Website.” [The Stanford Natural Language Processing Group](#) (access:2023.09.23)
- [5] Chenguang Wang *et al.*, “Zero-Shot-Information Extraction as a Unified Text-to-Triple Translation”, Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2021, 14p