

심층 신경망을 활용한 진료 기록 문헌에서의 중단형 개체명 및 관계 추출 비교 연구*

- 파이프라인 모델과 결합 모델을 중심으로 -

A Comparative Research on End-to-End Clinical Entity and Relation Extraction using Deep Neural Networks: Pipeline vs. Joint Models

최 성 필 (Sung-Pil Choi)**

목 차

- | | |
|-----------------|------------------|
| 1. 서론 | 4. 성능 실험 및 결과 분석 |
| 2. 관련 연구 | 5. 결론 |
| 3. 방법론 (Models) | |

초 록

정보추출은 문헌 내에 존재하는 개체명을 인식함과 동시에 이들 간의 의미적 관계까지도 식별하여 최종적으로 문헌 내에 포함된 의미적 트리플을 자동으로 추출하여 활용할 수 있으므로 문헌에 대한 심층적인 분석과 이해에 많은 도움을 줄 수 있다. 그러나 지금까지 대부분의 정보추출에 대한 연구는 개체명 인식과 관계추출이 개별 연구로 각각 분리되어 진행되었으며, 그 결과 입력 문헌에 대한 정보추출의 최종 출력인 의미적 트리플 추출 성능에 대한 객관적이고 정확한 평가 제대로 이루어지지 않았다. 이에 본 논문에서는 진료 기록 문헌에 나타나는 개체명과 그들 간의 관계를 트리플 형태로 직접 추출할 수 있는 중단형 정보추출의 2가지 모델인 파이프라인 및 결합형 모델을 구축하는 구체적인 방법론을 제시하고 성능 비교 실험을 진행하였다. 우선 파이프라인 모델은 양방향 GRU-CRFs를 활용한 개체명 인식 모듈과 다중 인코딩 기반 관계추출 모듈로 구현되었고, 결합형 모델을 위해서는 다중 헤드 레이블링 기반의 양방향 GRU-CRFs이 적용되었다. 두 가지 시스템을 바탕으로 진료기록 문헌 내의 개체명과 관계를 모두 태깅하여 구축된 i2b2/VA 2010 데이터셋을 활용한 비교 실험에서 파이프라인 모델의 성능이 5.5%(F-measure) 더 높게 나타났다. 추가적으로, 대규모 신경망 언어모델과 수작업으로 구축된 자질 정보를 활용한 최고 수준의 기존 시스템과의 비교 실험을 통해, 본 논문에서 구현한 중단형 모델의 객관적인 성능 수준을 파악할 수 있었다.

ABSTRACT

Information extraction can facilitate the intensive analysis of documents by providing semantic triples which consist of named entities and their relations recognized in the texts. However, most of the research so far has been carried out separately for named entity recognition and relation extraction as individual studies, and as a result, the effective performance evaluation of the entire information extraction systems was not performed properly. This paper introduces two models of end-to-end information extraction that can extract various entity names in clinical records and their relationships in the form of semantic triples, namely pipeline and joint models and compares their performances in depth. The pipeline model consists of an entity recognition sub-system based on bidirectional GRU-CRFs and a relation extraction module using multiple encoding scheme, whereas the joint model was implemented with a single bidirectional GRU-CRFs equipped with multi-head labeling method. In the experiments using i2b2/VA 2010, the performance of the pipeline model was 5.5% (F-measure) higher. In addition, through a comparative experiment with existing state-of-the-art systems using large-scale neural language models and manually constructed features, the objective performance level of the end-to-end models implemented in this paper could be identified properly.

키워드: 개체명 인식, 관계추출, 중단형 정보추출, 심층 신경망, 텍스트마이닝

Named-entity recognition, Relation extraction, End-to-end information extraction, Deep neural network, Text mining

* 이 논문은 2020학년도 경기대학교 연구년 수혜로 연구되었음.

** 경기대학교 문헌정보학과 부교수(spchoi@kgu.ac.kr / ISNI 0000 0004 6772 9269)

논문접수일자: 2023년 1월 17일 최초심사일자: 2023년 2월 5일 게재확정일자: 2023년 2월 21일
한국문헌정보학회지, 57(1): 93-114, 2023. <http://dx.doi.org/10.4275/KSLIS.2023.57.1.093>

© Copyright © 2023 Korean Society for Library and Information Science

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

특정 분야의 기술 문헌을 심층 분석하여 구성하고 있는 핵심적인 개념 요소를 식별하고 추출하여 가공하는 작업은 해당 분야 전문가에게도 난해한 작업이다. 정보추출(Information Extraction)은 텍스트 내에 존재하는 개체명(Named Entity)을 식별함과 동시에 이들 간의 의미적 관계까지도 식별함으로써 이러한 기술 문헌 분석의 어려움을 경감시킬 수 있다(Miwa & Bansal, 2016). 정보추출 작업의 주요 목표는 비정형 텍스트 문헌에서 의미 트리플(semantic triple)과 같은 구조적 정보를 추출하는 것이며, 이를 통해서 인간과 컴퓨터가 전체 텍스트의 핵심적인 의미 요소들을 쉽게 파악하고 활용함으로써 기계번역(Machine Translation) (Wu et al., 2016)이나 질의응답(Question Answering) (Mollá et al., 2006) 등과 같은 지능적인 응용 시스템들의 성능을 개선할 수 있다.

전통적으로 대부분의 정보추출 방법론들은 개체명 인식(Named Entity Recognition, NER) (Lample et al., 2016)과 관계추출(Relation Extraction, RE) (Choi et al., 2014)을 순차적으로 적용하는 파이프라인(pipeline) 형태를 가정한다. 그러나, 많은 연구자들은 이 두 가지 하부 태스크 중에 하나의 태스크에만 집중하여 연구하고 있으며, 전체적인 정보추출 파이프라인을 실행하여 텍스트에서 완전한 형태의 개체명 간 의미적 관계들을 추출하는 시도가 거의 없었다(Miwa & Bansal, 2016; N. T. Nguyen et al., 2015). 결론적으로, 정보추출의 궁극적인 목표라고 볼 수 있는, 한 쌍의 개체명과 그들 사이의 관계로 구성된 의미 트리플들을 텍스트

에서 추출하는 연구는 지금까지 제대로 수행되지 않고 있었다고 볼 수 있다(Bhattacharyya et al., 2017; Gupta et al., 2016).

최근 들어, 심층 신경망 모델을 활용한 결합형 정보추출(Joint IE)에 대한 연구가 본격적으로 시작되고 있다(D. Q. Nguyen & Verspoor, 2019; Giorgi et al., 2019; Fang et al., 2021; Z. Ji et al., 2021). 이 기법은 어떠한 부가적인 정보가 포함되지 않은 원시 텍스트를 대상으로 개체명 인식과 관계추출이 단일 모델에 의해서 동시에 수행되어 최종적인 의미 트리플이 추출되는 것을 목표로 한다.

지금까지 대부분의 정보추출에 대한 연구는 개체명 인식과 관계추출이 개별 연구로 분리되어 진행되었으며, 그 결과 입력 문헌에 대한 정보추출의 최종 출력인 의미적 트리플 추출 성능에 대한 객관적이고 정확한 평가가 제대로 이루어지지 않았다. 앞서 언급한 바와 같이, 비록 초기 단계지만, 뉴스 기사 텍스트와 같은 범용 분야에서의 종단형 정보추출 연구에 대한 관심은 증대되고 있는 반면, 의생물학 분야 텍스트에서의 적용은 미비한 실정이다. 이에 본 논문에서는 진료 기록 문헌에 나타나는 개체명과 그들 간의 관계를 트리플 형태로 직접 추출할 수 있는 종단형 정보추출의 2가지 모델인 파이프라인 및 결합형 모델을 구축하는 구체적인 접근 방법을 제시하고 실제 구현된 두 시스템의 성능을 비교하는 연구를 수행한다. 이로써 문헌 내에서 의미적 트리플을 추출하기 위해서 독립적으로 개발된 두 시스템을 별도로 실행하는 기존의 방법에서 탈피하여 보다 효과적인 종단형 정보추출 모델을 확보하는데 필요한 전체 조건 및 성능 극대화를 위한 접근 방법 등에

대해서 논의한다.

본 논문은 다음과 같이 전개된다. 우선 2장에서는 정보추출과 관련된 지금까지의 연구사례를 분석하고, 3장에서는 논문에서 구현한 파이프라인 및 결합형 정보추출 모델에 대해서 세부적으로 설명한다. 4장에서는 구현된 시스템에 대한 실험 및 비교 분석을 하고 5장에서 결론을 맺는다.

2. 관련 연구

지금까지 개체명 인식 및 관계추출을 포함하는 정보추출 기술은 다양한 영역에 적용되어 많은 연구가 진행되어 왔다(Lample et al., 2016; D. Q. Nguyen & Verspoor, 2019; M. Zhang et al., 2017). 수많은 연구 중에서 여기서는 본 논문의 중점 접근 방향인 종단형 정보추출에 대한 연구 사례를 정리하였다.

초창기 연구에서는 체계적으로 조직화된 지식베이스와 함께 어휘 및 구문 자질들을 추가적으로 활용한 자질기반 지도학습(feature-based supervised learning) 기법이 주로 활용되었다. Kate와 Mooney(2010)는 ‘card-pyramid’ 구조와 이를 기반으로 한 구문분석 알고리즘을 고안하여 단일 문장 내의 가능한 모든 개체쌍에 대한 관계 조합을 고려한 결합형 개체-관계 추출을 시도하였다(Kate & Mooney, 2010). 이와 유사하게, Miwa와 Sasaki(2014)는 개체와 관계를 도표 구조로 도식화하고 여기에 이력 기반 구조적 학습(history-based structured learning)을 적용함으로써 전체 작업을 단순 슬롯 채우기(slot-filling) 문제로 변환하였다(Miwa

& Sasaki, 2014). 지금까지의 구문 구조적 방법에서 벗어나서, Q. Li와 Ji(2014)는 준 마르코프 체인(semi-Markov chain) 기반 세그먼트 디코더를 적용한 구조적 퍼셉트론을 활용하여 개체와 관계를 동시에 추출하는 새로운 접근 방법을 제시하고 ACE2025 말뭉치를 활용한 실험에서 우수한 성능의 파이프라인 모델보다 더 높은 성능을 보였다(Q. Li & Ji, 2014).

정보추출에 본격적으로 심층학습이 활용되면서 기존 결합형 모델의 성능은 획기적으로 향상되었다. 그 첫번째 시도로서, Gupta et al.(2016)는 위에서 언급한 Miwa와 Sasaki(2014)의 방법에 심층학습을 적용하여, 단어와 개체-관계 레이블 의존성의 의미적 구성을 활용하여 개체명 인식 및 관계분류 작업을 테이블 값 채우기 문제로 축소시키는 다중 태스크 순환 신경망을 제안하였고, CoNLL04 데이터셋을 이용한 실험에서 최고 성능을 달성했다(Gupta et al., 2016). 또한, Miwa와 Bansal(2016)은 문장의 어휘 나열 정보와 의존 트리 구조 정보를 한꺼번에 활용하기 위해서 양방향 트리구조 LSTM-RNNs와 양방향 순차 LSTM-RNNs 모델을 결합하여 종단형 관계추출 문제를 해결하고자 하였다. ACE2004, ACE2005는 물론 SemEval-2010 Task 8 데이터셋을 활용한 심도 깊은 실험을 통해서 자신들의 모델이 최고 성능을 달성했음을 입증하였다(Miwa & Bansal, 2016).

심층학습에 사용되는 모델의 규모가 커짐에 따라 다중 레이어 기반의 심도 깊은 신경망 구조를 활용하기 시작했다. 대표적으로 Zheng et al.(2017)은 결합형 모델을 구현하기 위해서 개체명 인식을 위한 양방향 LSTM 레이어와 관계추출을 위한 CNN 레이어를 겹겹이 쌓고 최

상단에 소프트맥스 레이어를 올리는 방식을 시도하였다(Zheng et al., 2017). 유사한 방법으로 Zhang et al.(2017)은 입력 표현을 강화하기 위해서 사전 학습된 의존 파서 기반의 은닉 양방향 LSTM 레이어에 의해 도출된 구문 정보를 활용하는 전역적으로 최적화된 신경망 모델을 고안하여 기존의 두 가지 방법(Miwa & Bansal, 2016; Miwa & Sasaki, 2014)보다 높은 성능을 보였다(Zhang et al., 2017).

한편, Li et al.(2017)은 학습된 개체명 인식 모델의 가중치를 구문 트리 정보 기반의 양방향 LSTM을 활용하는 관계추출 모델에 이식하는 방식으로 결합형 모델(joint model)을 구축하였다(Li et al., 2017). 이 연구는 생의학 분야에 결합형 모델을 적용한 최초의 사례로 볼 수 있다. Zheng et al.(2017)은 결합형 접근 방법을 시퀀스 태깅 문제로 환원하기 위한 통합 태깅 방법을 고안하고 이를 LSTM 기반의 인코더-디코더 모델에 적용하였고, Bekoulis et al.(2018b)는 Zheng et al.(2017)과 유사한 다중 헤드 레이블링 기법을 제안하고 이를 다중 문맥 기반 결합형 관계추출을 위한 적대적 모델(adversarial model)에 적용하였다(Bekoulis et al., 2018a). Nguyen와 Verspoor(2019) 또한 Zheng et al.(2017)와 유사하게 LSTM-CRF와 CNN 레이어들을 결합한 적층형 모델을 제안하였으나, 의존 구문분석에서 처음 사용된 심층 biaffine 주의집중 메커니즘(Dozat & Manning, 2016)을 새롭게 응용하는 방법을 고안했다는 점에서는 새로운 접근 방법이라고 평가받고 있다. 최근 들어, 수작업으로 세밀하게 구축된 추출 자질정보를 활용하거나(Ji et al., 2021) 대규모 신경망 언어모델을 적용한(Fang et al.,

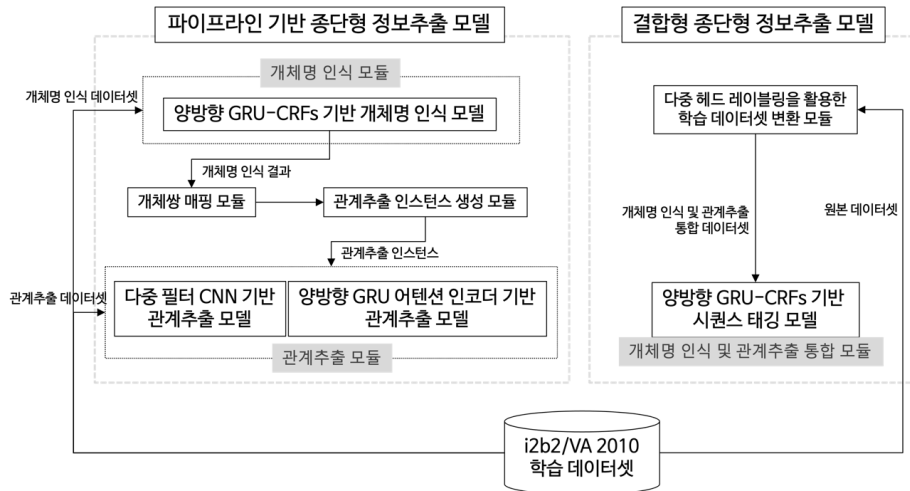
2021) 파이프라인 및 결합형 정보추출 모델 구축에 대한 연구도 진행되고 있으나 성능 평가 과정에서 정보추출의 최종 결과물, 즉 결과 트리플에 대한 정합성 검증 여부가 불투명하다.

위에서 설명한 바와 같이, 뉴스기사 등과 같은 범용 분야에서의 결합형 관계추출과 종단형 정보추출에 대한 많은 연구가 진행된 반면, 진료 기록 및 생의학 분야 자원을 대상으로 하는 특화된 연구는 많지 않다. 게다가 대부분의 분야 특화 연구가 정보추출의 특정 하부 모듈인 개체명 인식 혹은 관계추출에 국한되어 진행되고 있다. 이러한 상황에서 본 논문은 진료 기록 정보를 대상으로 하는 종단형 정보추출 방법론에 대한 심층적인 비교연구를 수행함으로써 각 방법론의 특성과 성능에 대한 구체적인 분석 결과를 제시하고자 한다.

3. 방법론(Methods)

본 논문에서는 종단형 정보추출을 위한 2가지 방법 즉, 파이프라인 모델과 결합형 모델에 대한 비교 실험을 위해서 두 시스템을 직접 구현하였다. 세부적인 내용에 대한 논의에 앞서, <그림 1>에서 비교 실험에 활용된 종단형 정보추출 시스템의 전체 구조를 보여주고 있다.

파이프라인 기반 종단형 정보추출 모델은 단일 개체명 인식 모델과 2가지 관계추출 모델로 구현되었고, 결합형 모델은 단일 시퀀스 태깅 모델을 활용한다. 개별 모델의 학습에 필요한 학습 데이터셋은 각 모델의 기능에 맞도록 변환되어 입력되며, 특히 결합형 모델을 위해서는 다중 헤드 레이블링(3.5) 방법을 활용하여



〈그림 1〉 논문에서 구현된 2가지 종단형 정보추출 모델 구조도

원본 학습 데이터셋을 변환하여 개체명 인식 및 관계추출 통합 데이터셋 형식으로 변환하게 된다. 또한 파이프라인 모델 내의 개체쌍 매핑 및 관계추출 인스턴스 생성 모듈은 개체명 인식 결과를 관계추출 모델의 입력으로 변환하기 위한 과정이다.

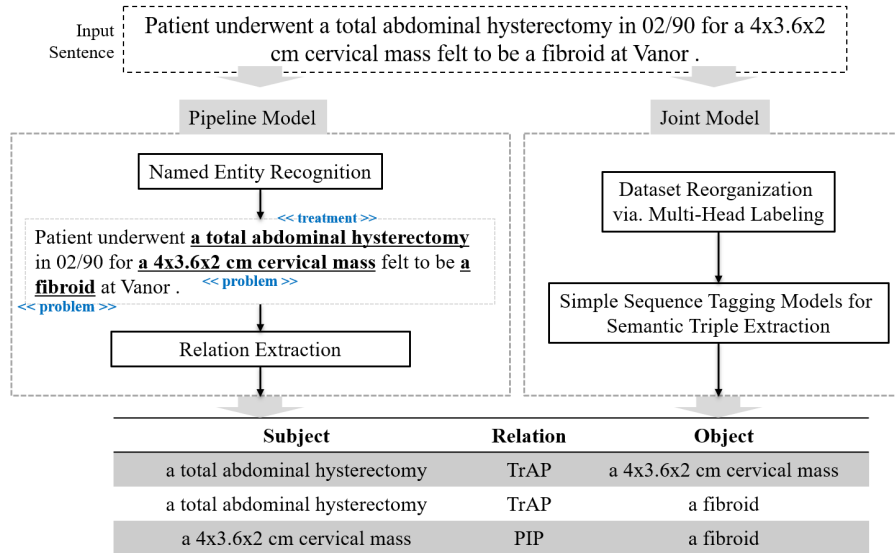
3.1 문제 정의(Task Definition)

본 논문에서는 ‘종단형 개체 및 관계추출(end-to-end entity and relation extraction)’을 단일 문장에 내포된 모든 의미적 트리플(semantic triple)을 추출하는 작업으로 정의한다. 이를 위해서, 해당 모델은 단계적으로 혹은 동시에 텍스트에서 모든 유효한 개체명과 그들 간의 의미적 관계쌍을 인식해야 한다. 따라서 전체적인 시스템은 개체명 인식(Named-Entity Recognition, NER)과 관계추출(Relation Extraction, RE)의 두 하부 모듈로 구성된다. 본 논문에서는 두 가지 종류의 종단형 정보추출 모델을 구현하였

다. 〈그림 2〉에서 보듯이, 우선 파이프라인 모델(Pipeline Model)은 두 개의 하부 모듈이 순차적으로 수행되는 모델이고, 결합형 방식(Joint Model)은 단일 모델이 문장에서 개체명 및 최종 의미 트리플을 추출한다.

3.2 양방향 GRU-CRFs 기반 개체명 인식 (Named-Entity Recognition with Bidirectional GRU-CRFs)

본 논문의 연구 목적이 진료 기록 문헌에서 서로 다른 2가지 형태의 종단형 정보추출 접근 방법을 집중적으로 비교 분석하는 것이므로, 개체명 인식을 위해서 비교적 단순하면서도 일반적으로 많이 활용되고 성능도 우수한 양방향 LSTM-CRFs(Huang et al., 2015) 형태의 구조를 채택하였다. 기존 양방향 LSTMs의 상단에 조건부 랜덤 필드(CRFs) 모델을 탑재한 형태로서 품사태깅(POS tagging), 기저구 인식(chunking) 및 개체명 인식 등을 위한 수많은



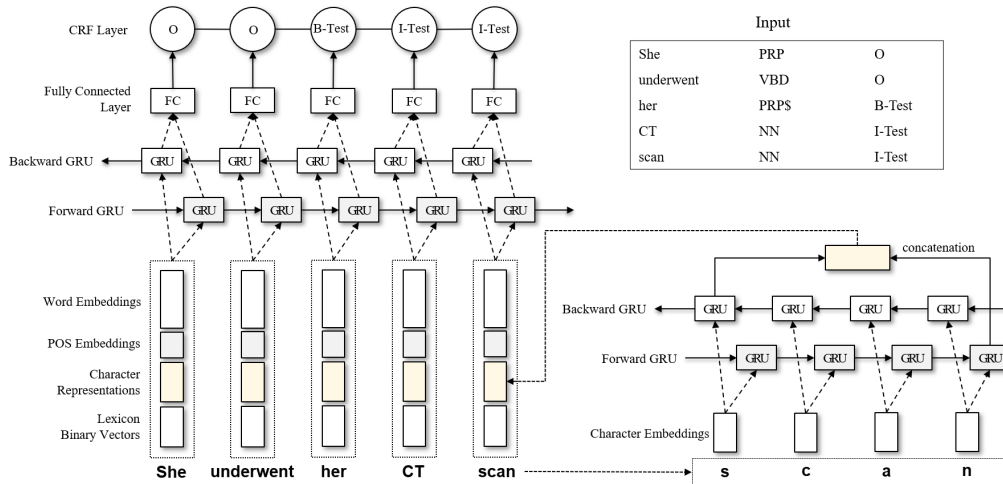
〈그림 2〉 종단형 정보추출을 위한 2 가지 모델인 파이프라인(pipeline) 모델과 결합형(joint) 모델 (파이프라인 모델은 NER과 RE를 순차적으로 수행하며, 결합형 모델은 문장에서 의미 트리플을 추출하기 위한 단일 신경망 모델이다.)

우수한 성능의 변형 모델들을 촉발시켰다(Chiu & Nichols, 2016; Dang et al., 2018; Misawa et al., 2017). 특히 본 논문에서는 위 골격은 그대로 유지하되 LSTM 셀을 Gated Recurrent Units (GRUs)(Cho et al., 2014) 셀로 교체하여, 성능은 그대로 유지하되 덜 복잡한 구조로 인한 계산 효율성 개선이 가능하였다(Chung et al., 2014).

〈그림 3〉은 논문에서 구현된 개체명 인식 모델의 전체 구조를 보여주고 있다. 먼저 원본 입력 문장에 대해서 Stanford CoreNLP(Manning et al., 2014)를 활용하여 토큰 분리 및 품사 태깅을 수행함으로써 그림의 우측 상단에 보이는 바와 같은 CoNLL 스타일의 입력 형태로 변환

한다(Sang & Meulder, 2003). 단어 임베딩을 위해서는 (Collobert et al., 2011)의 사전 학습 모델을 사용했으며, 품사 임베딩도 추가적으로 포함시켰다. 또한, 임베딩 어휘 목록에 포함되지 않은 미등록어나 희소 빈도 단어 등을 처리하기 위해 양방향 GRUs 기반 인코딩 레이어를 추가하여 문자 수준 자질까지도 추가했다(Lample et al., 2016). 마지막으로, 특수 입력 벡터인 “이진 레xicon 벡터(Lexicon Binary Vector)”를 구성하였는데, 이는 특정 단어가 전체 데이터셋에서 어떤 개체 유형에 포함되고 있는지를 벡터 형태로 표현한 것이다. 예를 들어, i2b2/VA 2010¹⁾ 데이터셋에서, “occlusion”이라는 단어는 두 종류의 개체 타입에 활용된다. 즉, 개체명

1) A National Center for Biomedical Computing, “I2B2 NLP Research Data Sets”, <https://www.i2b2.org/NLP/DataSets/>



〈그림 3〉 본 논문에서 구현된 진료 기록 대상 개체명 인식 모델(양방향 GRU-CRFs를 기반으로 단어, 품사정보, 문자정보 그리고 렉시콘 임베딩을 추가 자질로 활용하였다. 특히 문자정보 표현은 양방향 GRUs 인코더를 활용하여 생성하였다.)

인 “interventional neuroradiology occlusion”에서는 “I-Treatment”로 지정되어 있고, “distal left circumflex occlusion”에서는 “I-Problem”으로 지정되어 있다. 이렇게 개체명 관점에서의 특정 어휘에 대한 활용 패턴을 <표 1>과 같은 이진 벡터로 구성하고 이를 추가적인 입력 자질로 활용한다.

따라서, 입력 문장의 입력 자질 벡터는 4개 요소 벡터의 병합인 이다. 여기서 는 사전 학습된 단어 임베딩 벡터이고, 는 품사 임베딩이다. 그리고 는 해당 단어에 대한 문자 수준 자질 표현 벡터이며, 마지막으로 는 위에서 설명한 이진 렉시콘 벡터를 나타낸다. 이렇게 구성된 입력 자질 벡터는 표준 양방향 GRU 레이어로 입력되며 각 GRU 셀은 아래 수식을 통해서 d -차원의 벡터인 를 출력한다.

여기서 은 시그모이드(sigmoid) 활성화 함수이다. . . . , 와 는 전방 셀(forward cell)에 대한

가중치 행렬을 나타내고 . . . , 및 는 후방 셀(backward cell)에 대한 가중치 행렬이다. 따라서 모델 내에서 총 12개의 가중치 행렬과 6개의 바이어스가 학습되어야 한다. 는 완전 연결 노드()로 입력되어 각 토큰의 개체 타입에 대한 예측이 수월하도록 전체 레이블의 개수와 동일한 차원의 출력 벡터 를 도출한다. 각 토큰에 대한 개별적인 는 다시 CRF 레이어로 입력되어 레이블 전이 행렬과 추가적인 토큰 기반 자질들을 활용하여 현재 토큰에 대한 최적의 레이블을 예측하게 된다(Huang et al., 2015).

3.3 다중 인코딩 기반 관계추출

본 논문에서는 다양한 신경망 인코더들을 활용하여 특정 문장에 대해서 다른 방식으로 여러 번의 인코딩을 수행함으로써 다각도의 자질 정보를 추출하는 다중 인코딩(multi-encoding) 기

<표 1> Lexicon Binary Vector 예시(i2b2/VA 2010 데이터셋 내에서의 개체 타입 종류가 총 6개 (O는 제외, (B/I)-Problem (B/I-Pr), (B/I)-Treatment (B/I-Tr), (B/I)-Test (B/I-Te)) 이므로 Lexicon Binary Vector는 6차원의 이진 벡터임.)

Tokens	Labeling Patterns	Lexicon Binary Vectors					
		B-Pr	I-Pr	B-Tr	I-Tr	B-Te	I-Te
ETOH	B-Problem (<i>ETOH intoxication</i>) B-Test (<i>ETOH</i>)	1	0	0	0	1	0
occlusion	I-Problem (<i>distal left circumflex occlusion</i>) I-Treatment (<i>interventional neuroradiology occlusion</i>)	0	1	0	1	0	0
pressure	B-Problem (<i>pressure</i>) I-Problem (<i>increased blood pressure</i>) B-Treatment (<i>pressure stockings</i>) I-Treatment (<i>blood pressure control</i>) I-Test (<i>blood pressure</i>)	1	1	1	1	0	1
hematocrit	I-Problem (<i>decreased hematocrit</i>) I-Test (<i>discharge hematocrit</i>)	0	1	0	0	0	1

법에 기반한 문장 분류로 관계 추출을 시도하였다(Zhang et al., 2017). 자질 정보의 다양성 확보를 위해서 본 논문은 다중 필터 CNN 인코더(Multi-filter CNN Encoder)(Kim et al., 2017; Nguyen & Grishman, 2015)와 양방향 GRU 어텐션 인코더(Bidirectional GRU Attention Encoder)(Du & Huang, 2018; Ji et al., 2017; Yang et al., 2016)를 모두 사용하였다.

3.3.1 다중 필터 CNN 인코더

다중 필터 합성곱 신경망(Convolutional Neural Network, CNN)은 특정 문장에 서로 다른 크기를 가진 여러 개의 필터를 적용함으로써 풍부한 자질을 출력할 수 있다. 일반적으로 CNN은 합성곱(convolution) 레이어와 풀링(pooling) 레이어로 구성되는데, 다중 필터가 사용되면 각 필터는 독립적인 고유의 합성곱 및 풀링 연산(convolution and pooling operation)을 수행함으로써 다중 자질 벡터를 추출하게 된다.

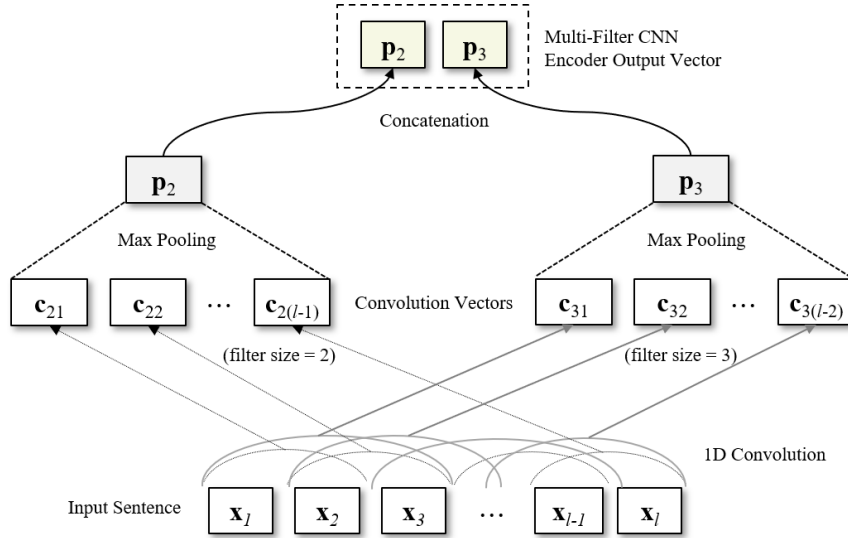
<그림 4>는 본 논문에서 사용된 다중 필터

CNN 모델을 나타낸다. 입력 문장에 대한 합성곱 벡터를 얻기 위해서 아래의 수식을 이용하여 개별 길이가 s인 단어 범위(word span)에 걸쳐서 합성곱 연산이 수행되었다.

여기서 는 합성곱 벡터의 인덱스를 나타내고, 는 필터의 길이를, 그리고 는 길이가 인 필터 가중치이며 는 바이어스이다. 또한 하이퍼볼릭 탄젠트 함수(\tanh)를 활성화 함수로 활용하였다(T. H. Nguyen & Grishman, 2015). 이렇게 도출된 합성곱 벡터 는 맥스 풀링(max pooling) 레이어로 입력되어 축약된 형태의 효과적인 자질 벡터 가 만들어진다. 이를 기반으로 길이가 에서 인 다중 필터를 적용하면, 모든 자질 벡터를 병합함으로써 최종적인 인코더 출력 벡터 를 얻게 된다.

3.3.2 양방향 GRU 어텐션 인코더 (Bidirectional GRU Attention Encoder)

또 하나의 인코딩 기법은 최근 들어 관계추



〈그림 4〉 다중 필터 CNN 인코더 기반 관계추출 모델(길이가 (2, 3)인 1차원 합성곱 필터를 사용하였으며 인코더의 출력 벡터를 도출하기 위해서 2개 이상의 필터를 적용할 수 있다.)

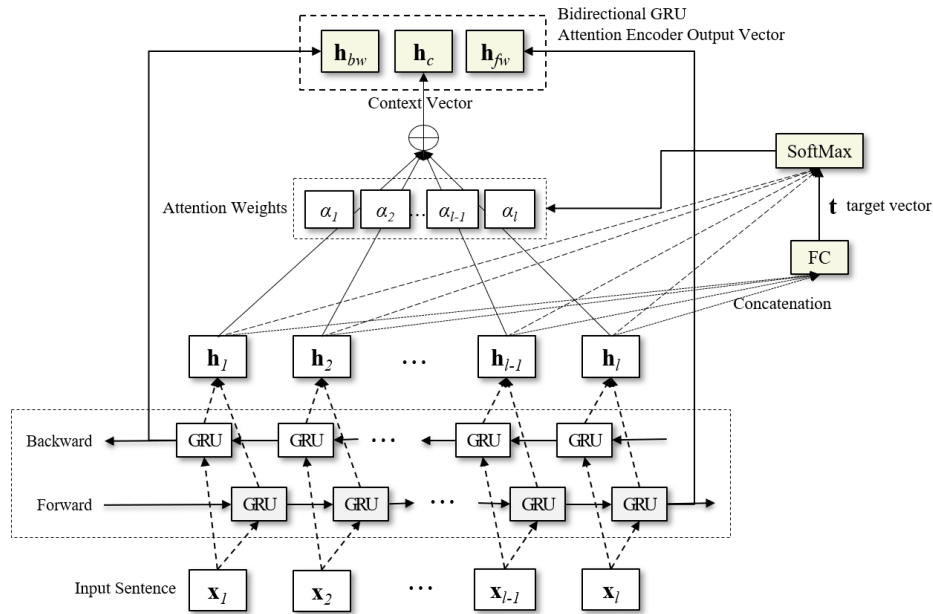
출 및 문서분류에 자주 활용되는 어텐션 기법을 동반한 양방향 RNN 네트워크 모델이다(Du & Huang, 2018). 본 논문에서는 전체 모델의 학습 속도와 매개변수 크기를 고려하여 GRU (Gated Recurrent Units)(Cho et al., 2014)를 사용하였다.

〈그림 5〉에서 보는 바와 같이, 순차적 자질 분석을 위해 입력 임베딩 벡터가 양방향 GRU 네트워크로 주입되어 순방향 상태 벡터 h_{fw} 와 역방향 상태 벡터 h_{bw} 가 도출된다. 그 과정에서 네트워크 내의 모든 상태 벡터들이 병합되어 완전 연결 레이어로 주입되고, 최종적으로 전체 토큰 시퀀스를 표현할 수 있는 타깃 벡터가 출력된다. 어텐션 가중치는 다음의 softmax 함수로 계산된다.

여기서 h_i, h_j 는 양방향 GRU 네트워크의 상태 벡터이고, n 는 생성된 타깃 벡터, 그리고 e 는 토큰 시퀀스의 전체 길이를 나타낸다. 이러한 셀프 어텐션 메커니즘은 최종적으로 컨텍스트 벡터를 출력하는데 이 벡터는 어텐션 가중치에 의한 모든 상태 벡터의 가중 평균을 수행함으로써 계산된다. 결론적으로, 인코더의 출력 벡터는 다음과 같은 병합 벡터 형태로 표현될 수 있다.

이 벡터는 다시 전술한 다중 필터 CNN 인코더²⁾의 출력과 병합되어 최종적으로 출력 차원이 관계 레이블의 수와 동일한 완전 연결 레이어로 입력된다. 해당 레이어의 출력은 다음 수식과 같이 softmax 함수를 거쳐서 레이블에 대한 예측 확률로 변환된다.

2) 본 논문에서는 길이가 각각 2, 3인 2개의 필터를 사용하였다.



〈그림 5〉 양방향 GRU 어텐션 인코더 기반의 관계추출 모델(출력 인코딩 벡터는 셀프 어텐션 메커니즘에 의해 생성된 컨텍스트 벡터와 최종 역방향 및 순방향 상태 벡터로 구성된다.)

여기서 은 완전 연결 레이어의 가중치 행렬 이고 은 바이어스를 나타내며, 그 외에 다른 매개변수들은 앞에서 설명한 인코딩 벡터들이다. 내에서 확률 최고값에 대한 관계 레이블을 선택함으로써 예측이 완료된다.

3.4 파이프라인 모델(Pipeline Model)

파이프라인 모델은 특정 처리 단계의 출력이 다음 단계의 입력으로 직접적으로 연결되어 전체 데이터 처리가 이루어지는 모델이다. 이 모델은 최종 결과물 도출을 위해서 2개 이상의 작업이 필요할 때 자주 활용되며, 단계별 병렬처리 및 입출력 버퍼링 등을 통해서 전체적인 시스템의 효율성 확보라는 장점이 있다. 그러나

처리 과정에서 이전 단계에서 발생한 다양한 오류들이 다음 단계로 그대로 전파되어 모델의 전체 성능이 현저하게 악화될 수 있는 위험성이 있다.

본 논문에서는 각각 3.2와 3.3에서 설명한 개체명 인식 및 관계추출 모델을 단계별로 실행하여 문장에 존재하는 의미 트리플을 추출하는 파이프라인 모델을 구성하였다. 전체 과정에 대한 개념도는 〈그림 1〉의 좌측에 제시하였다. 두 개의 독립적이고 순차적인 작업은 그 사이에 일련의 부가적인 작업을 수행함으로써 자연스럽게 연결될 수 있다. 다시 말해서, 개체명 인식을 수행한 후에 관계추출을 수행하기 위해서는 개체쌍 매핑(entity pair mapping)과 관계 인스턴스 생성(relation instance generation)

작업이 필요하다.

관계추출 작업은 문장 내 한 쌍의 개체명 사이의 의미적 관계를 예측하는 것을 목표로 하므로, 이전 개체명 인식 결과 중에서 개체명이 존재하지 않거나 1개만 존재하는 문장을 제거해야 한다. 개체쌍 매핑을 통해서 2개 이상의 개체명이 포함된 문장에 대해서 관계 인스턴스를 생성하기 위해서는 모든 가능한 조합의 개체쌍을 연결시켜야 하며 개별 개체쌍 각각에 대한 방향성도 고려해야 한다. 예를 들어, “Went to OSH c furter ICD shocks, **VT** on monitor but stable VS.”라는 문장에서 개체명은 2개 존재하므로 2개의 개체쌍, 즉 “**VT** → **monitor**”, and “**monitor** → **VT**”가 관계 인스턴스로 도출될 수 있다.

한편, 길이가 긴 문장에 대해서 개체명 인식을 수행한 결과 인식된 개체명이 많을 경우, 과도하게 많은 개체쌍이 관계 인스턴스로 생성될 수 있다. 이럴 경우를 대비해서 문장 내에서 15 단어 이상 떨어진 두 개체명에 대해서는 서로 간의 의미적 관계가 성립하기 어렵다고 판단하고 관계 인스턴스로 선택하지 않고 제외시킨다 (T. H. Nguyen & Grishman, 2015).

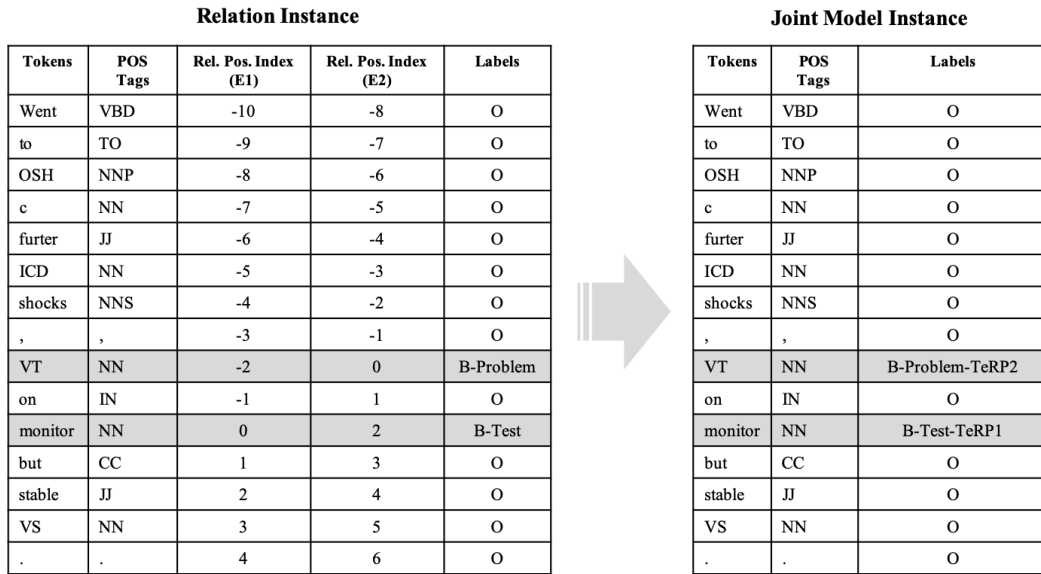
앞에서 설명했듯이, 2개 이상의 개체명을 포함하는 단일 문장 내에서 모든 가능한 조합의 개체쌍을 도출하고 이들 각각에 대한 방향성까지 고려한다면 여러 개의 관계 인스턴스가 생성될 수 있다. 하나의 문장에서 생성되는 다중 인스턴스 각각에 대해서 관계추출 대상 개체명을 표시하기 위해 위치 인덱싱 기법(position indexing)을 사용한다. 다시 말해서, 각각의 대상 개체명에 대한 문장 내에서의 상대적 위치 정보를 모델에 추가적으로 입력하여 관계추출을 수행하

게 된다. 예를 들어, 현재 대상 개체명의 위치정보를 0으로 지정하고 나머지 단어들은 이 위치를 기준으로 한 상대적 인덱스를 지정하는 것이다(Nguyen & Verspoor, 2019).

3.5 결합형 모델(Joint Model)

본 논문에서는 개별 토큰에 부착되는 각각의 레이블에 개체 유형 뿐만 아니라 (만일 특정 개체와의 관계가 존재한다면) 관계 유형도 함께 포함하는 다중 헤드 레이블링(multi-head labeling) (Bekoulis et al., 2018b)을 활용한 시퀀스 태깅(sequence tagging) 방법으로 결합형 모델을 구현하였다. <그림 5>는 파이프라인 모델과 조인트 모델의 데이터 레이블링 기법을 비교해서 보여주고 있다.

<그림 6>의 좌측에서 보듯이 일반적인 관계추출을 위한 입력인 관계 인스턴스는 단어, 품사, 개체 유형 및 개별 개체에 대한 상대적 위치 정보가 포함되어 있다. 그리고 관계추출 과정에서 해당 입력에 대한 자동분류를 수행함으로써 문장 내 두 개체명의 의미적 관계를 예측한다. 반면에, 다중 헤드 레이블링 기법을 적용한 <그림 6>의 우측에서는 개체명 및 관계 유형이 레이블에 포함되어 있다. 예를 들어, <그림 6>에서 “B-Problem-TeRP2”는 해당 단어의 개체 유형이 “Problem”이며, 문장 내에서 또 다른 개체명과 “TeRP”라는 관계로 연관되어 있음을 나타낸다. 이 때, 레이블의 맨 마지막 숫자 2는 해당 개체가 트리플 상에서 두 번째 개체임을 표시한다. 이를 기반으로 다양한 종류의 시퀀스 레이블링 방법을 통해서 개체명 인식과 관계추출을 동시에 수행할 수 있다. 본 논문에서



monitor (Test) :: TeRP :: VT (Problem) * TeRP : Test reveals medical problem

〈그림 6〉 파이프라인 및 결합형 모델 각각을 위한 레이블링 기법 비교 예시(파이프라인 모델을 위한 관계 인스턴스에는 앞에서 언급한 개체 위치 인덱스가 포함되어 있으나, 결합형 모델에서는 레이블 자체에 모든 정보가 포함되어 있으므로 이 정보가 불필요하다.)

서는 3.2절에서 설명한 양방향 GRU-CRFs에 렉시콘 이진 벡터를 적용한 방법을 결합형 모델에 적용하였고 앞에서 설명한 파이프라인 모델과의 성능 비교 실험을 수행하였다.

4. 성능 실험 및 결과 분석

4.1 데이터셋과 실험환경

본 논문에서는 앞서 언급한 바와 같이 i2b2/VA 2010 데이터셋을 이용하여 파이프라인 모델과 결합형 모델의 비교 실험을 진행하였다. 이 데이터셋은 병원에서 작성되는 다양한 진료

기록(clinical records) 문헌을 수작업으로 분석하여 기록 내부에 존재하는 개체명 뿐만 아니라 개체명 간의 의미적 관계까지도 태깅한 데이터셋으로서 의료분야 종단형 정보추출 연구에 활발하게 활용되고 있다. 〈그림 7〉은 데이터셋의 내용 일부를 보여주고 있다.

〈그림 7〉의 윗부분은 특정 환자에 대한 특정 일자의 병원 진료 기록 문헌의 일부이다. 그리고 아랫부분에는 해당 문헌 내에 존재하는 개체명(의학개념문구, clinical concepts)과 그 관계를 기술하고 있다. 이해를 돕기 위해서, 첫번째 항목을 세부적으로 살펴보면, 진료 기록 문헌의 20번째 라인에 존재하는 “chest x-ray”라는 개체명은 “mild pulmonary edema”라는 개

진료 기록 문헌(일부)

1: Admission Date :
 2: 2011-03-10
 3: Discharge Date :
 4: 2011-03-15
 ... (중략)
 7: Service :
 8: Medford
 9: HISTORY OF PRESENT ILLNESS :
 10: This is an 83-year-old woman with a history of hypertension who presents with slurred speech when dehydrated .
 11: She has paraplegia secondary to HTLV exposure while on vacation in the Bahamas six years ago .
 ... (중략)
 19: The patient had coarse rhonchi on examination .
 20: Chest x-ray revealed mild pulmonary edema .
 21: Chest CT scan was negative for pulmonary embolism but positive for consolidation .
 22: She was started on Levofloxacin but the patient became hypotensive at that point with blood pressure of 70/45 and received a normal saline bolus to boost her blood pressure to 99/60 ; however the patient was admitted to the Medical Intensive Care Unit for overnight observation because of her somnolence and hypotension .
 ... (중략)

문헌에 포함된 개체명 및 관계정보

```

...
c="chest x-ray" 20:0 20:1||r="TeRP"||c="mild pulmonary edema" 20:3 20:5
c="chest ct scan" 21:0 21:2||r="TeRP"||c="pulmonary embolism" 21:6 21:7
c="chest ct scan" 21:0 21:2||r="TeRP"||c="consolidation" 21:11 21:11
c="blood pressure" 22:14 22:15||r="TeRP"||c="hypotensive" 22:9 22:9
c="a normal saline bolus" 22:20 22:23||r="TriP"||c="hypotensive" 22:9 22:9
c="levofloxacin" 22:4 22:4||r="TrCP"||c="hypotensive" 22:9 22:9
c="overnight observation" 22:44 22:45||r="TeCP"||c="her somnolence" 22:48 22:49
c="overnight observation" 22:44 22:45||r="TeCP"||c="hypotension" 22:51 22:51
...
        
```

〈그림 7〉 i2b2/VA 2010 데이터셋의 내용 일부

체명과 “TeRP”라는 관계로 연결되어 있음을 나타내고 있다. 이 데이터셋에 존재하는 개체명 유형의 종류는 총 3가지이며, “Problem(의학적 문제)”, “Treatment(의학적 치료법/대응)”, “Test(의학적 검사)”으로 구성되어 있다. 또한 관계 유형의 종류는 총 8개이며 각 유형에 대한 세부적인 설명은 <표 2>와 같다.

위 데이터셋에 대해서 먼저 단어 토큰화와

품사 태깅 작업을 거친 이후 개체명 매핑 작업을 수행하였다. 학습 집합(training set)과 실험 집합(test set)이 분리되어 있으므로 이를 그대로 활용하였고, 검증 집합(validation set)을 마련하기 위하여 학습 집합을 무작위로 섞어 9:1의 비율로 나누었다. i2b2/VA 2010 데이터셋 내에 존재하는 개체명에 대한 분포는 다음의 <표 3>과 같다.

〈표 2〉 i2b2/VA 2010 데이터셋의 관계 유형

관계유형	설명
TrIP	치료법(treatment)이 의학적 문제(problem)를 개선하다.
TrWP	치료법(treatment)이 의학적 문제(problem)를 악화하다.
TrCP	치료법(treatment)이 의학적 문제(problem)를 야기하다.
TrAP	치료법(treatment)이 의학적 문제(problem)에 대응하다.
TrNAP	치료법(treatment)이 의학적 문제(problem)때문에 실행되지 않다.
TeRP	검사(test) 결과 의학적 문제(problem)가 발견되다.
TeCP	검사(test)가 의학적 문제(problem)을 조사하기 위해서 수행되다.
PIP	의학적 문제(problem)가 다른 의학적 문제(problem)를 나타내다.

〈표 3〉 i2b2/VA 2010 데이터셋의 개체명 분포 정보

개체 유형	데이터셋 규모			개체명 일치 비율(%)	
	학습 집합	검증 집합	실험 집합	검증 집합	실험 집합
Problem	6,347	725	12,592	48.13	40.39
Treatment	4,343	498	9,344	62.65	47.64
Test	4,177	427	9,225	60.83	62.64
전체 (평균)	14,867	1,652	31,161	57.36	50.22

〈표 3〉에서 개체명 일치 비율을 계산한 이유는 학습 집합 내에 태깅된 모든 개체명 집합을 수집하여 이를 검증 집합과 실험 집합에 존재하는 개체명과의 일치율을 계산함으로써 3.2에서 소개한 개체명 인식용 추가 자질인 이진 렉시콘 벡터(Binary Lexicon Vector)의 효용성을 확인하고자 함이다. 확인 결과 평균적으로 50%가 넘는 일치율을 보이고 있으므로 해당 자질의 효용성은 어느 정도 입증된 것이라 볼 수 있다.

추가적으로 관계추출을 위한 i2b2/VA 2010 원본 데이터 가공 작업을 진행하였다. 3.4.1절과 3.4.2에서 설명한 것과 같이 한 쌍의 개체명을 포함한 관계 인스턴스들을 생성하고, 각 관계

인스턴스를 구성하는 토큰에 상대적 위치 정보를 추가하였다. 이 과정에서 개체명 간 거리가 15 토큰 미만이고 관계로 맺어지지 않은 한 쌍의 개체명을 포함하는 별도의 네거티브 인스턴스들을 임의로 생성하여 “None” 관계 레이블로 지정하였다. 이렇게 구성된 관계추출 데이터셋의 통계 정보는 〈표 4〉와 같다.

〈표 3〉와 〈표 4〉에서 설명한 데이터셋은 각각 별도의 개체명 인식 모델(3.4)과 관계추출 모델(3.5)의 학습에 활용되어 각 모델이 단계별로 수행되는 파이프라인 모델을 도출한다.

이와는 별도로, 결합형 모델을 위한 데이터셋도 별도로 구성해야 한다. 앞서 3.5절에서 설

〈표 4〉 i2b2/VA 2010에서 도출한 관계추출 데이터셋 통계 정보

관계 유형	전체 관계추출 데이터셋		
	학습 집합	검증 집합	실험 집합
TrAP	619	66	1,379
TrCP	142	13	281
TrIP	39	1	114
TrWP	19	1	80
TrNAP	40	4	92
TeRP	697	74	1,579
TeCP	117	13	268
PIP	568	71	1,234
None (관계없음)	24,701	2,753	52,862
전체 (None 제외)	2,241	243	5,027
전체	26,942	2,996	57,889

명한 바와 같이 본 논문에서는 개체 유형과 관계 유형이 동시에 표현된 멀티 헤드 레이블링 기법을 활용한다. 즉, <그림 6>의 우측과 같이 단일 문장에서 관계로 맺어진 한 쌍의 개체명 각각에 대해서 개체 유형은 물론 관계 유형 및 트리플 내에서의 위치정보까지도 포함한 복합 레이블("B-Problem-TeRP2")을 부착하게 된다. 그런데 개체 및 관계 유형이 결합된 형태의 복합 레이블을 구성하는 과정에서 레이블의 종류가 과도하게 증가되어 전체적인 성능이 악화되는 상황이 발생한다³⁾. 따라서 이러한 과도한 레이블 생성을 일부 제한하기 위해서 본 논문에서는 단일 문장에 동일한 관계 유형이 여러 번 발생할 경우 이들 중 오직 1개의 관계 유형만을 유지하고 나머지는 제외하는 방법을 적용하였다. 단일 문장 내에서 동일한 관계 유형들은 유사한 표현 맥락을 공유하기 때문에 1개만 유지하더라도 해당 관계 유형만의 고유한 문맥적 자질을 학습 집합이 충분히 포함할 것이라는 가정을 적용하였다. 이렇게 구성된 결합형

모델에 활용될 통합 데이터셋의 관계 유형별 분포는 <표 5>와 같다.

결합형 모델에서는 "None" 관계 유형을 가진 인스턴스가 불필요하기 때문에 제외시켰다. 또한, 이전의 <표 4>에 비해서 유형별 관계 인스턴스 개수가 감소했음을 알 수 있는데, 이는 앞에서 설명했듯이 과도한 복합 레이블 생성을 방지하기 위한 필터링 과정에서 일부 관계 정보가 배제되었기 때문이다. 따라서 파이프라인 모델과의 객관적인 성능 비교를 위해서 이 데이터셋에서 배제된 관계 인스턴스에 대해서 <표 4>의 데이터셋에서도 동일하게 제외시켰다.

4.2 모델 구현 및 매개변수 최적화

본 논문에서는 3개의 세부 모델이 실제로 구현되어 학습되었다. 우선, 단어, 품사, 문자 및 개체명 사전 기반의 양방향 GRU-CRFs를 활용한 개체명 인식 모델(3.2)과 2개의 인코더를 결합한 다중 인코딩 기반 관계추출 모델(3.3)은

<표 5> 결합형 모델을 위한 통합 데이터셋 통계 정보

관계 유형	통합 데이터셋		
	학습 집합	검증 집합	실험 집합
TrAP	432	63	873
TrCP	87	13	186
TrIP	29	1	84
TrWP	10	1	56
TrNAP	27	4	67
TeRP	510	70	1,068
TeCP	79	13	176
PIP	366	66	775
전체	1,587	231	3,285

3) BIO 기반의 개체 레이블 종류가 총 7가지(B-Problem, I-Problem, ...B-Test, I-Test)이고, 이에 관계 유형(8가지)까지 결합한다면 총 57가지(None 포함) 종류의 레이블이 발생할 수 있다.

파이프라인 모델을 구성한다. 또한, 파이프라인 모델에서의 양방향 GRU-CRFs를 활용한 개체명 인식 모델은 데이터셋만 변경하여 별도의 결합형 모델을 구성한다. 이 3가지 모델에 대하여 검증 집합을 활용한 매개변수 최적화 과정을 수행했고 그 결과는 <표 6>과 같다.

3가지 모델 모두 최적화 함수로 Adam을 사용하였고, 학습률(learning rate)도 0.01로 동일하게 설정되었으며, 모델 내부의 가중치 차원과 입력 차원들도 유사하게 지정되었다. 관계추출 모델을 위한 CNN 필터는 길이가 각각 2, 3, 5인

3가지를 지정하여 적용하였다.

4.3 실험 결과 및 분석

우선, 4.2에서 설명한 파이프라인 모델을 구성하는 개체명 인식(Bi-GRU-CRFs) 및 관계추출(Multi-Encoding) 모델 각각의 성능 수준을 객관적으로 평가하기 위해서, <표 3>과 <표 4>에서 설명한 데이터셋을 이용하여 각 모델에 대한 개별적인 실험을 진행하였고 그 결과를 <표 7>에서 보여주고 있다.

<표 6> 세부 모델별 매개변수 최적화 결과

모델 매개변수	파이프라인 모델(pipeline model) ([표 3], [표 4]의 데이터셋 학습)		결합형 모델(joint model) ([표 5]의 데이터셋 학습)
	개체명 인식 모델 (Bi-GRU-CRFs)	관계추출 모델 (Multi-Encoding)	통합 추출 모델 (Bi-GRU-CRFs)
품사 임베딩 차원 수	30	-	50
문자 임베딩 차원 수	50	-	50
단어 임베딩 차원 수	100	100	100
문자 레벨 GRU 차원 수	100	-	100
단어 레벨 GRU 차원 수	100	-	50
CNN 필터 크기	-	[2, 3, 5]	-
CNN 차원 수	-	100	-
Learning rate	0.01	0.01	0.01
Dropout ratio	0.9	0.5	0.7
Optimizer	Adam	Adam	Adam

<표 7> 파이프라인 모델에서의 개체명 인식 및 관계추출 성능 비교(i2b2/VA 2010)
(P: Precision, R: Recall, F1: F-score)

성능 비교 모델	개체명 인식(NER)			관계추출(RE)			누적 성능	평균 성능
	P	R	F1	P	R	F1		
Unanue et al. (2017)	-	-	83.35	-	-	-	-	-
Li et al. (2019)	-	-	-	75.69	73.03	74.34	-	-
Giorgi et al. (2019)	-	-	89.26	-	-	63.02	56.25	76.14
Ji et al. (2021)	85.33	82.55	83.92	51.74	37.31	43.36	36.39	63.64
Fang et al. (2021)	91.77	91.74	91.76	57.15	53.08	55.04	50.50	73.40
본 논문의 파이프라인 모델	82.52	84.32	83.41	74.38	75.91	75.13	62.67	79.27

〈표 7〉에서 Fang et al.(2021)을 비롯하여 Giorgi et al.(2019) 및 Ji et al.(2021)은 의학 분야에 특화된 대규모 신경망 언어모델을 사용하고 있다. 이 시스템들의 개체명 인식 성능은 본 논문의 시스템보다 높게 나타나고 있으나 관계추출 성능이나 오류 전파를 고려한 누적 성능⁴⁾(62.67%) 혹은 평균 성능(79.27%) 측면에서는 제안 시스템이 대등한 성능을 보임을 알 수 있다. 앞에서 설명했듯이 종단형 정보추출의 최종 목표 측면에서는 이 전체 성능 혹은 평균 성능이 더욱 중요하며 본 논문에서 제안된 파이프라인 및 결합형 모델의 성능 수준을 간접적으로 확인할 수 있었다. 그런데 〈표 6〉에서 제시된 누적 성능과 평균 성능은 실제 종단형 정보추출의 성능이 아니며, 단지 개체명 인식과 관계추출 각각의 자체 성능을 기준으로 계산된 예측 성능일 뿐이다. 이후부터는 3.4, 3.5에서 설명한 파이프라인 및 결합형 모델에 대한 실증 실험을 통해서 종단형 정보추출 성능을 비교한다.

우선 파이프라인 모델 구성을 위해서는 〈표 7〉의 실험에서 학습된 두 모델을 그대로 사용했으며, 개체명 인식 결과 중에서 문장 내 개체명 간 거리 제약(15 단어 이하)을 적용한 총 26,718개의 개체 쌍을 도출하고 이를 관계추출의 입력으로 사용하였다. 결합형 모델은 3.5에서 설명한 멀티 헤드 레이블링 기반 시퀀스 태깅 모델을 활용하여 〈표 5〉의 학습집합으로 학습을 진행하였다. 파이프라인 및 결합형 모델이 각각 학습한 학습집합의 규모와 내용이 서로 다르기 때문에 객관적인 성능 비교가 불가능하다는 의견이 있을 수도 있지만, 제한적인 학습집

합 활용이라는 단점도 어떤 측면에서는 결합형 모델 고유의 특성으로 볼 수 있으므로 〈표 5〉의 실험집합을 활용하여 비교 실험을 진행하였다.

〈표 8〉은 본 논문에서 구현한 파이프라인 및 결합형 모델에 대해서 관계가 존재하지 않는 인스턴스를 제외한 실험 집합에 대해서 수행한 성능 평가 결과를 보여주고 있다. 또한 〈표 7〉에서 제시된 비교 모델 중에서 성능이 가장 높은 모델들도 함께 제시하였다. 앞에서 언급하였듯이 비교 모델들은 평가의 세부 기준이 명확하게 제시되지 않아서 확실하지는 않으나, 본 논문에서 구현한 아래 두 모델들은 실험집합을 대상으로 두 모델이 출력하는 의미 트리플 집합의 정합성을 실제로 직접 평가한 성능을 기재하였다. 따라서 표에서 파이프라인 모델의 최종 성능은 개체명 인식의 결과를 그대로 관계추출에 적용했으므로 관계추출의 성능(45.82%)과 동일하다. 결합형 모델에서는 두 단계가 단일 프로세스로 수행되므로 개체명 인식의 성능은 크게 의미가 없으나 별도로 측정하여 추가하였다.

〈표 8〉을 살펴보면, 우선 파이프라인 기반의 종단형 정보추출 모델의 성능이 〈표 7〉에서의 계산된 누적 성능 혹은 평균 성능에 비해 매우 낮다는 사실을 알 수 있다. 이는 전체 학습 데이터셋의 규모가 이전에 비해 작은 것과 개체명 인식 후의 결과 중에서 조건에 맞지 않아서 제거되는 데이터가 존재한다는 이유도 있지만, 개체명 인식과 관계추출이 순차적으로 수행되는 과정에서 발생한 오류의 전파가 더 심각하다는 것을 의미한다.

4) 누적 성능은 개체명 인식 및 관계추출 과정에서 발생하는 오류를 모두 적용한 성능으로, 파이프라인 시스템을 구성하는 두 모델 각각의 개별 성능을 곱하여 계산된다.

〈표 8〉 종단형 정보추출(End-to-End IE)의 성능 비교(i2b2/VA 2010)

모델	개체명 인식(NER)			관계추출(종단형 정보추출 성능)		
	Precision	Recall	F1	Precision	Recall	F1
Ji et al. (2021)	85.33	82.55	83.92	51.74	37.31	43.36
Fang et al. (2021)	91.77	91.74	91.76	57.15	53.08	55.04
본 논문의 파이프라인 모델	82.20	82.37	82.28	68.34	34.46	45.82
본 논문의 결합형 모델	53.84	59.68	56.60	69.20	27.49	39.35

비교 모델 중의 하나인 Ji et al.(2021)은 수작업으로 구축한 세밀한 자질 정보 기반의 모델을 활용하였으며, Fang et al.(2021)은 의생물학 분야에 특화된 대규모 언어모델을 파인튜닝하여 구성한 모델을 사용하였다. 본 논문의 최종 목표가 성능의 극대화가 아니라 종단형 정보추출을 위한 2가지 즉, 파이프라인 및 결합형 모델의 실험적 비교라는 점에서 타 시스템과의 성능 비교가 중요한 사항은 아니지만, 〈표 8〉에서 보는 바와 같이 현재 자연어처리 분야에서 가장 높은 성능을 보이고 있다는 신경망 언어모델을 사용한 모델과의 성능 차이가 0.9% 정도로 나타나고 있어, 외부 자원을 전혀 활용하지 않은 모델임을 감안했을 경우 비교적 준수한 수준으로 볼 수 있다.

본 논문에서 제안한 파이프라인 및 결합형 모델을 비교하면, 정확도는 비교적 대등하게 나타났지만 재현율에서 7% 정도의 성능 차이가 발생했고, 결과적으로 최종 F1 점수는 5.5% 정도로 파이프라인 모델의 성능이 높았다. 이러한 결과는 〈표 4〉와 〈표 5〉에서 제시된 학습 데이터셋의 차이에서도 기인하지만 결정적으로 파이프라인 모델에서 총 26,718개의 개체명 인식 결과쌍을 바탕으로 모든 경우에 대해서 관계추출을 시도한 결과가 재현율 향상을 이끈 것으로 보인다. 부가적으로 두 모델 간의 처리

속도에 대한 실험도 진행하였는데 결합형 모델이 파이프라인 모델에 비해서 약 100배 정도 빠른 속도로 최종 트리플을 추출할 수 있었다. 이는 결합형 모델이 하나의 인스턴스에 대해서 별도의 추가적인 처리 없이 트리플을 식별하여 추출하는 반면, 파이프라인 모델은 개체명 인식 후에 개체쌍을 구성하고 이들 개체쌍 발생 개수만큼 관계추출을 여러 번 시도하기 때문이다. 비록 결합형 모델이 학습집합의 부족과 과도하게 많은 레이블의 수 등으로 인해 낮은 성능을 보이고 있으나, 이를 극복한다면 장기적으로는 종단형 정보추출의 효율성 측면에서 파이프라인 모델에 비해 활용 가능성이 더 높을 것으로 판단된다.

〈표 9〉는 결합형 모델을 통해서 도출된 정보추출 결과를 보여주고 있다. 표에서 제시된 3개의 문장 중에서 두 문장에서는 올바른 트리플이 추출되고 있으나, 마지막 문장에서는 트리플이 아예 식별되지 않고 있다. 이는 앞서도 언급하였듯이 근본적으로는 학습집합의 부족 문제에서 기인하지만 단일 문장에 2개 이상의 개체명이 존재할 경우에 트리플 추출이 더 어렵다는 것을 부분적으로 보여주고 있다. 특이한 점은 마지막 문장에서 정답 트리플은 “gentamincin-TrAP-sepsis”로 지정되어 있으나, 추가적으로 “ampicillin-TrAP-sepsis”도 정

<표 9> 결합형 모델에 의해서 입력 문장으로부터 추출된 트리플 예시

입력 문장	출력 트리플		비고
Selective coronary angiography showed an occluded right coronary artery in its midportion.	Subject	selective coronary angiography	True
	Relation	TeRP	
	Object	an occluded right coronary artery in its midportion	
A colonoscopy showed anteriordysplasia, no bleeding was seen.	Subject	A colonoscopy	True
	Relation	TeRP	
	Object	anteriordysplasia	
The infant received 48 hours of ampicillin and gentamicin for rule out sepsis.	Subject	- (gentamicin)	False
	Relation	- (TrAP)	
	Object	- (sepsis)	

답 트리플로 간주될 수 있다는 점이다. 이러한 점은 향후 학습집합과 해당 모델의 수정 및 보완을 통해서 해결해야 하는 한계점이다.

5. 결 론

본 논문은 진료 기록 문헌에 나타나는 개체명과 그들 간의 관계를 트리플 형태로 직접 추출할 수 있는 종단형 정보추출의 2가지 모델인 파이프라인 및 결합형 모델을 구축하는 구체적인 방법론을 제시하였다. 우선 파이프라인 모델은 양방향 GRU-CRFs를 활용한 개체명 인식 모듈과 다중 인코딩 기반 관계추출 모듈로 구현되었고, 결합형 모델을 위해서는 다중 헤드 레이

블링 기반의 양방향 GRU-CRFs이 적용되었다. 두 가지 시스템을 바탕으로 진료기록 문헌 내의 개체명과 관계를 모두 태깅하여 구축된 i2b2/VA 2010 데이터셋을 활용한 비교 실험에서 파이프라인 모델의 성능이 5.5%(F-measure) 더 높은 반면, 추출 속도는 결합형 모델이 약 100배 정도 빠른 것으로 나타났다. 또한 수작업으로 세밀하게 구축한 자질 정보를 활용하거나 대규모 언어모델을 기반으로 한 모델과의 성능 비교에서 본 논문의 파이프라인 모델이 비교적 대등한 성능을 나타내었다. 향후 연구 과제로 진료기록 문헌 대상 종단형 정보추출 모델의 성능 개선은 물론, 정확하고 객관적인 성능 평가를 수행할 수 있는 다양한 데이터셋을 기반으로 한 광범위한 실험적 연구가 필요하다.

참 고 문 헌

Bekoulis, G., Deleu, J., Demeester, T., & Develder, C. (2018a). Adversarial training for multi-context joint entity and relation extraction. ArXiv:1808.06876 [Cs].
<http://arxiv.org/abs/1808.06876>

- Bekoulis, G., Deleu, J., Demeester, T., & Develder, C. (2018b). Joint entity recognition and relation extraction as a multi-head selection problem. *Expert Systems with Applications*, 114, 34-45.
- Bhattacharyya, P., Pawar, S., & Palshikar, G. K. (2017). End-to-end Relation Extraction using Neural Networks and Markov Logic Networks. *EACL*.
- Chiu, J. P. & Nichols, E. (2016). Named entity recognition with bidirectional LSTM-CNNs. *Transactions of the Association for Computational Linguistics*, 4, 357-370.
- Cho, K., van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: encoder-decoder approaches. *ArXiv:1409.1259 [Cs, Stat]*.
- Choi, S.-P., Lee, S., Jung, H., & Song, S. (2014). An intensive case study on kernel-based relation extraction. *Multimedia Tools and Applications*, 71(2), 741-767.
- Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *ArXiv:1412.3555 [Cs]*.
- Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12, 2493-2537.
- Dang, T. H., Le, H.-Q., Nguyen, T. M., & Vu, S. T. (2018). D3NER: Biomedical named entity recognition using CRF-biLSTM improved with fine-tuned embeddings of various linguistic information. *Bioinformatics (Oxford, England)*, 34(20), 3539-3546.
- Dozat, T. & Manning, C. D. (2016). Deep biaffine attention for neural dependency parsing. *ArXiv:1611.01734 [Cs]*.
- Du, C. & Huang, L. (2018). Text classification research with attention-based recurrent neural networks. *International Journal of Computers Communications & Control*, 13(1), 50-61.
- Fang, X., Song, Y., & Maeda, A. (2021). Joint extraction of clinical entities and relations using multi-head selection method. *2021 International Conference on Asian Language Processing (IALP)*, 99-104.
- Giorgi, J., Wang, X., Sahar, N., Shin, W. Y., Bader, G. D., & Wang, B. (2019). End-to-end named entity recognition and relation extraction using pre-trained language models. *ArXiv:1912.13415 [Cs]*.
- Gupta, P., Schütze, H., & Andrassy, B. (2016). Table filling multi-task recurrent neural network for joint entity and relation extraction. *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2537-2547.
<https://aclweb.org/anthology/papers/C/C16/C16-1239/>
- Huang, Z., Xu, W., & Yu, K. (2015). Bidirectional LSTM-CRF models for sequence tagging.

ArXiv:1508.01991 [Cs].

- Ji, G., Liu, K., He, S., & Zhao, J. (2017). Distant supervision for relation extraction with sentence-level attention and entity descriptions. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 3060-3066.
- Ji, Z., Ghiasvand, O., Wu, S., & Xu, H. (2021). A discrete joint model for entity and relation extraction from clinical notes. *AMIA Summits on Translational Science Proceedings, 2021*, 315-324.
- Kate, R. J. & Mooney, R. (2010). Joint entity and relation extraction using card-pyramid parsing. *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, 203-212.
- Kim, S.-W., Yu, S. J., Lee, M.-H., & Choi, S.-P. (2017). A comparative study on deep learning topology for event extraction from biomedical literature. *Journal of the Korean Society for Library and Information Science*, 51(4), 77-97.
- Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). Neural architectures for named entity recognition. *ArXiv:1603.01360 [Cs]*.
- Li, F., Zhang, M., Fu, G., & Ji, D. (2017). A neural joint model for entity and relation extraction from biomedical text. *BMC Bioinformatics*, 18(1), 198.
- Li, Q. & Ji, H. (2014). Incremental joint extraction of entity mentions and relations. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 402-412.
- Li, Z., Yang, Z., Shen, C., Xu, J., Zhang, Y., & Xu, H. (2019). Integrating shortest dependency path and sentence sequence into a deep learning framework for relation extraction in clinical text. *BMC Medical Informatics and Decision Making*, 19(1), 22.
- Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. *Association for Computational Linguistics (ACL) System Demonstrations*, 55-60.
- Misawa, S., Taniguchi, M., Miura, Y., & Ohkuma, T. (2017). Character-based Bidirectional LSTM-CRF with words and characters for Japanese Named Entity Recognition. *Proceedings of the First Workshop on Subword and Character Level Models in NLP*, 97-102.
- Miwa, M. & Bansal, M. (2016). End-to-end relation extraction using LSTMs on sequences and tree structures. *ArXiv Preprint ArXiv:1601.00770*.
- Miwa, M. & Sasaki, Y. (2014). Modeling joint entity and relation extraction with table representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*

- (EMNLP), 1858-1869.
- Mollá, D., Zaanen, M. V., & Smith, D. (2006). Named entity recognition for question answering. In Lawrence Cavdon and Ingrid Zukerman, Editors, Proceedings of the 2006 Australasian Language Technology Workshop, 51-58.
- Nguyen, D. Q. & Verspoor, K. (2019). End-to-end neural relation extraction using deep biaffine attention. ArXiv:1812.11275 [Cs], 11437, 729-738.
- Nguyen, N. T., Miwa, M., Tsuruoka, Y., Chikayama, T., & Tojo, S. (2015). Wide-coverage relation extraction from MEDLINE using deep syntax. BMC Bioinformatics, 16(1), 107.
- Nguyen, T. H., & Grishman, R. (2015, June). Relation Extraction: Perspective from Convolutional Neural Networks. Proceedings of NAACL Workshop on Vector Space Modeling for NLP.
- Unanue, I. J., Borzeshi, E. Z., & Piccardi, M. (2017). Recurrent neural networks with specialized word embeddings for health-domain named-entity recognition. Journal of Biomedical Informatics, 76, 102-109.
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, L., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., & Dean, J. (2016). Google's neural machine translation system: bridging the gap between human and machine translation. ArXiv:1609.08144 [Cs].
- Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016). Hierarchical attention networks for document classification. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1480-1489.
- Zhang, J., Liu, Q., & Zhou, J. (2017). ME-MD: An effective framework for neural machine translation with multiple encoders and decoders. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, 3392-3398.
- Zhang, M., Zhang, Y., & Fu, G. (2017). End-to-end neural relation extraction with global optimization. 1730-1740. <https://doi.org/10.18653/v1/D17-1182>
- Zheng, S., Hao, Y., Lu, D., Bao, H., Xu, J., Hao, H., & Xu, B. (2017). Joint entity and relation extraction based on a hybrid neural network. Neurocomputing, 257, 59-66.
- Zheng, S., Wang, F., Bao, H., Hao, Y., Zhou, P., & Xu, B. (2017). Joint extraction of entities and relations based on a novel tagging scheme. ArXiv:1706.05075 [Cs].