

## 프롬프트 분해와 다중 과제 학습 기법을 통한 긴 문서 내 관계 추출 강화 방법

조성웅<sup>1</sup> · 오흥선<sup>2\*</sup><sup>1</sup>한국기술교육대학교 컴퓨터공학과 석사<sup>2\*</sup>한국기술교육대학교 컴퓨터공학과 교수

# Enhancing Relation Extraction for Long Documents with Prompt Decomposition and Multi-task Learning

Seongung Jo<sup>1</sup> · Heung-Seon Oh<sup>2\*</sup><sup>1</sup>Master, Department of Computer Science, KOREATECH, Cheonan 31253, Korea<sup>2\*</sup>Professor, Department of Computer Science, KOREATECH, Cheonan 31253, Korea

### [요약]

문서 단위 관계 추출 연구는 문서 내 중요한 정보와 해당 정보들의 관계를 찾는 것이 목적이다. 기존의 연구들은 문장 또는 단락 단위의 짧은 문서를 대상으로 문장 수준 관계 추출 모델을 확장하거나 긴 문서를 대상으로 사전 학습 모델 기반 방법을 연구하였다. 하지만, 입력 텍스트 길이가 제한된 사전 학습 모델의 사용은 긴 문서를 분할 하여 다루게 되고 정보 고립화 문제를 일으킨다. 이를 극복하기 위하여, 본 논문에서는 과제 분해, 프롬프트 설계 그리고 다중 과제 학습을 적용하여 관계 추출을 수행하는 MaV(Multi-aspect Verifier)를 제안한다. MaV는 긴 문서를 짧은 단위 텍스트로 분할 하여 결과를 얻고 두 단계 검증을 이용하여 최종 결과를 조합한다. 이를 통해 기존의 사전 학습 모델을 변경 없이 사용할 수 있으며 정보 고립화를 해소할 수 있다. 제안 방법을 벤치마크 데이터에 적용하고 최신 연구와 비교하여 우수성을 검증하였다.

### [Abstract]

Document-level Relation Extraction (RE) identifies relationships between important information within a document. Previous research has evaluated methods based on sentence-level relation extraction models for short documents, such as sentences or paragraphs, and a Pre-trained Language Model (PLM) for longer documents. However, using PLM with a limited input text length can result in the problem of information isolation because the final outputs are concatenated with the results of small unit texts. To address this issue, this study proposes the Multi-aspect Verifier (MaV), which performs relation extraction using task decomposition, prompting, and multi-task learning. MaV predicts the final relation by splitting long documents into shorter input texts, obtaining all possible answers, and combining the answers through a two-step verification process. This allows the existing small PLM to be utilized without modification and solves the problem of information isolation. The superiority of the proposed model is demonstrated through experimentation and comparison with existing studies on a benchmark dataset.

**색인어** : 딥 러닝, 관계 추출, 질의 응답, 프롬프팅, 문서 처리**Keyword** : Deep Learning, Relation Extraction, Question Answering, Prompting, Document Handling<http://dx.doi.org/10.9728/dcs.2023.24.12.3131>

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 31 August 2023; Revised 04 October 2023

Accepted 11 October 2023

**\*Corresponding Author; Heung-Seon Oh**

Tel: +82-41-560-1497

E-mail: ohhs@koreatech.ac.kr

### 1. 서론

문서의 분석과 이해를 위해서 개체(entity) 간의 관계 추출(Relation Extraction)을 포함한 정보추출(Information Extraction)은 오랫동안 연구되고 있다. 기존의 연구 [1]-[4]들은 문장 또는 단락 수준의 짧은 텍스트를 대상으로 정보추출 연구를 수행하였으나 최근 들어 문서 수준의 긴 텍스트를 대상으로 정보추출에 대한 중요성이 부각 되고 있다.

최근 딥러닝 기반의 관계추출 연구[5]-[8]들은 사전학습 모델(PLM; Pre-trained Language Model) 기반 질의응답(QA; Question Answering) 방법을 이용하여 짧은 텍스트를 대상으로 하는 여러 벤치마크 데이터셋[9],[10]에서 높은 성능을 보였다. 그러나 논문, 특허, 의학 자료 등과 같은 긴 텍스트를 대상으로 하는 벤치마크[11],[12] 등에서는 비교적 낮은 성능을 보였다. 이는 문서 수준의 긴 텍스트가 문장 또는 단락 수준의 짧은 텍스트와 비교해서 10~20배의 텍스트 양을 가지고 있어 기존의 방법을 이용하면 문서의 문맥 정보를 제대로 이용할 수 없기 때문이다.

기존의 방법을 적용하기 위해서는 긴 텍스트를 짧은 단위 텍스트로 분할하고 단위 텍스트를 대상으로 관계 추출을 수행해야 한다. 이때 전체 텍스트와 비교하여 단위 텍스트는 정

보 고립화 문제가 발생한다. 이는 PLM이 상대적으로 짧은 고정 길이의 텍스트만을 수용할 수 있기 때문이다. 대표적인 트랜스포머[13] 인코더 기반 PLM인 BERT[14]와 인코더-디코더 기반 PLM인 BART[15]의 최대 입력 텍스트 길이는 각각 512와 1024 토큰이다. 최근 템플릿 생성 기반 연구[16]에서 BART[15] 구조를 개선하여 입력 길이 문제를 완화하려고 시도하였다. 이 문제의 해결 방법은 입력 텍스트의 길이 크게 하여 PLM을 학습하는 것이지만 대용량 고성능 학습 환경이 필요하기 때문에 매우 비효율적이다.

본 논문은 위 문제를 완화하기 위해 긴 입력 텍스트를 대상으로 PLM 기반 관계 추출 모델에 적용가능한 다방면 검증 방법(MaV; Multi-aspect Verifier)을 제안한다. MaV는 기존의 관계 추출을 개체 추출(Entity extraction), 개체 형식 분류(Type classification), 지역 검증(LV; Local Verification) 그리고 전역 검증(GV; Global Verification)의 세부 과제로 분해하고 다중 과제학습(multi-task learning)을 수행한다. 이때 지역 검증과 전역 검증을 통해서 기존의 PLM을 이용하여 긴 입력 텍스트를 효과적으로 처리한다. 긴 텍스트를 짧은 단위 텍스트로 분할 하여 다루고 개체와 관계 추출 오류를 단계적으로 검증한다. 세부 과제를 통해 여러 측면을 학습하고 정보 고립화 문제를 완화한다. 각 세부 과제는 목적에 맞는 전용 프롬

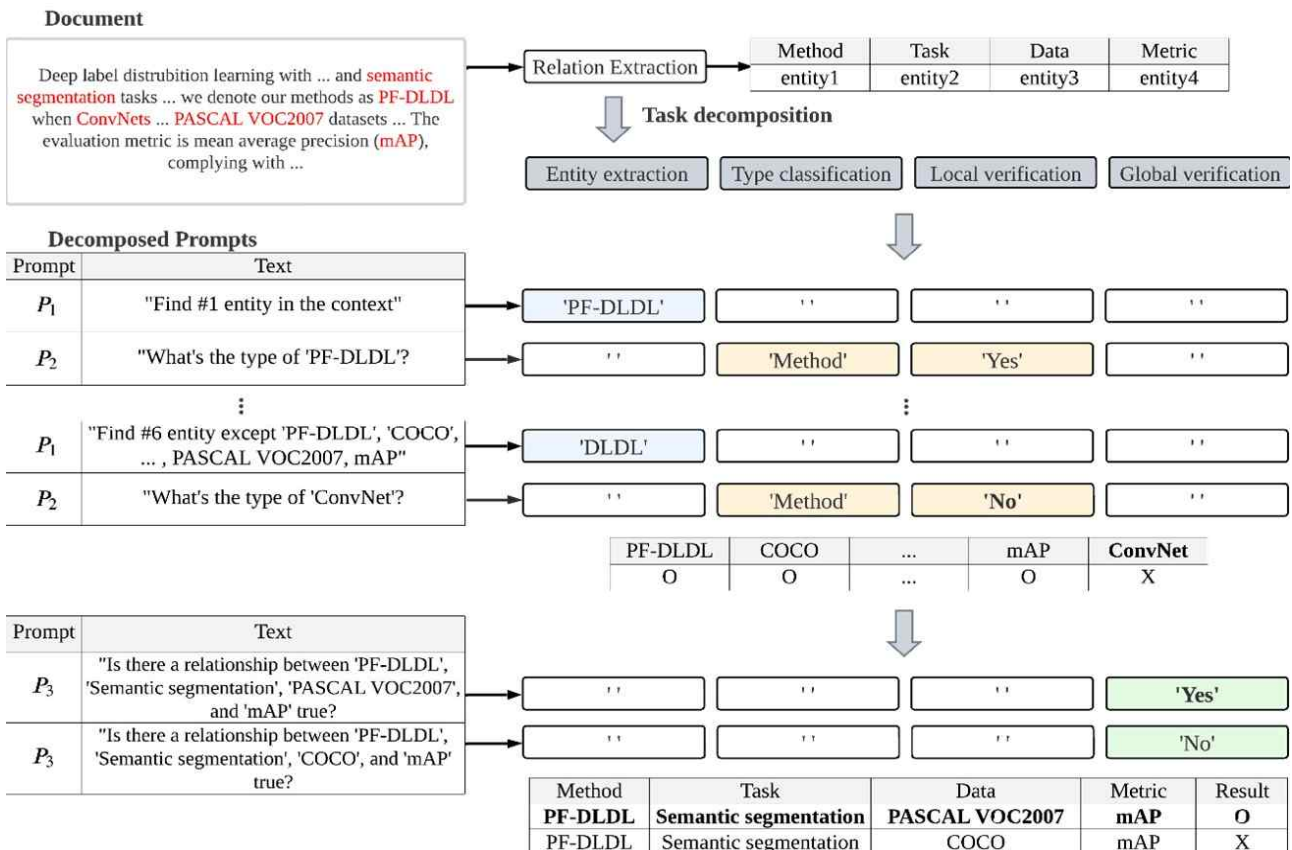


그림 1. MaV의 프롬프트 설계  
Fig. 1. Prompting in MaV

프트(prompt)를 설계하여 활용한다. 그림 1은 MaV의 프롬프트 설계를 통한 관계 추출 과정을 보여준다. 4개의 분해된 세부 과제를 3개의 프롬프트를 이용하여 추론하고 조합한다. 프롬프트  $P$ 는 3개 형식이 있으며 그림 1의 예시와 같이 동작한다.  $P_1$ 은 문서 내 개체를 추출하기 위한 프롬프트로 “Find 1 entity in the context”와 같은 질의를 만든다.  $P_2$ 는  $P_1$ 에서 추출한 개체의 형식을 분류하고 지역 검증에 대한 질의를 구성한다. 예시에서  $P_1$ 을 통해 개체 ‘PF-DL’을 찾았다면,  $P_2$ 를 통해 이 개체가 ‘Method’ 형식이고, 지역적으로 중요한 개체인지 추론한다. 마지막으로  $P_3$ 는  $P_1$ 과  $P_2$ 를 통해 추출한  $N$ 개의 개체 간의 관계를 전역 검증하는데 사용한다. 예시에서 추출한 4개 개체 간의 모든 관계를 고려하여 전역적으로 검증된 결과만을 확보한다. MaV는 전용 프롬프트를 통해 세부 과제들을 다중 과제 학습한다. 예시에서  $P_1$ 은 개체 추출 과제를 ‘PF-DL’로 나머지 3개 세부 과제를 공백 문자로 추론한다. 다중 과제를 통해 학습된 MaV는 프롬프트 형식에 따라 의도한 답변을 추론할 수 있다. 제안 방법은 최신 관계 추출 연구와 비교하여 성능 향상을 이루었다. 본 논문의 기여점은 다음과 같다.

1. 본 논문에서는 관계 추출을 4개의 세부 과제로 분해하고 전용 프롬프트 설계와 다중 과제학습 방법을 적용한 MaV를 제안한다. 이를 통해 긴 텍스트로부터 기존의 PLM을 수정하지 않고 관계 추출을 수행할 수 있다.

2. 본 논문에서는 긴 텍스트 대상 관계 추출 벤치마크 데이터셋인 SciREX [12]을 대상으로 최신 관계 추출 연구와 비교 분석하여 제안하는 모델의 우수성을 입증하였다.

논문의 남은 부분에서 다음과 같은 내용을 설명한다. 2장에서는 관계 추출 관련 연구, 3장에서는 제안하는 방법론, 4장에서는 실험과 결과 분석을 설명하고 5장에서 본 연구에 대한 결론을 전달한다.

## II. 관련 연구

관계 추출 연구는 초기 문장 수준의 연구에서 근래에 문서 수준의 연구로 확장되었다. 문장 수준 연구는 하나 혹은 몇 개의 문장에서 비교적 적은 개체와 관계를 다루었다. 짧은 길이의 텍스트를 CNN(Convolutional Neural Network), RNN(Recurrent Neural Network), Transformer 등 대표적인 딥러닝 모델 구조 기반으로 연구가 진행되었다. 이 후 BERT[6] 등 PLM을 사용한 구조에서 대부분의 문장 수준 연구가 인간과 비슷하거나 웃도는 성능을 보였다. 문장 및 문서 수준 관계 추출 연구는 크게 두 가지 방법으로 구분할 수 있다. 질의응답 및 분류 기반 방법(QA and classification based approach)과 생성 기반 방법(generation based approach)으로 두 방식 모두 PLM 구조 기반의 관계 추출 연구다.

질의응답 및 분류 기반 방법은 인코더와 분류 모델 기반 PLM, 질의 텍스트 입력을 통해 답변을 도출하는 초기 질의응답 연구의 확장 방법론이다. 주로 입력 텍스트인 문장이나 문서를 PLM 구조의 입력으로 하고 추가 분류 모델을 통과시켜 답변을 도출한다. [5]–[8], [17]–[24] 등의 대표 연구는 특정 개체에 대한 질의와 개체 추출을 답변으로 설계한다. 대부분 추출하고자 하는 대상에 대한 사전 정의된 고정적인 프롬프트를 만들어 사용한다. 문장 내 추출 대상이  $N$ 개라 하면 각  $N$ 개의 사전 정의된 질의가 필요한 방식이다. 본 방법론은 주로 인코더와 분류 모델 기반 PLM[14] 등을 사용한다. 최대 1~2개 정도의 단락(평균 512 토큰)을 한 번에 다룰 수 있는 것으로 여겨진다. 따라서, 입력 문서를 여러 번 쪼개어 매번 독립적인 문맥으로 다루므로, 각 추출 단계에 정보 고립화 [25] 문제가 발생한다. 순차적으로 각 단락의 연관 있는 여러 개체와 관계를 추출하고 조합하는 과정에 특정 단락에서 중요한 개체(혹은 정보)가 전체 문서에선 중요하지 않을 수 있기 때문이다. 해당 연구들은 성능이 좋은 PLM과 질의응답 방식의 확장으로 문장 및 단락 수준에서 좋은 성능을 보이며 관계 추출 연구의 새로운 장을 열었다. 하지만, 입력 길이가 긴 문서 수준 데이터셋이 점차 등장함에 따라 정보 고립화 및 추출 오류가 발생하는 한계가 있다.

생성 기반 방법은 템플릿 형태의 입, 출력과 디코더 기반 PLM을 통해 텍스트를 생성하는 연구의 확장 방법론이다. 문맥에서 특정 개체의 위치를 분류 모델을 찾아 답변하는 것이 아닌, 알맞은 답변 텍스트를 생성하는 구조다. 기존에서 변형된 템플릿은 보다 추상적이고 통합적인 형태를 보인다. 구체적으로 “위 문맥에서 도시 개체를 찾아라”라는 질의에서 “위 문맥에서 개체1을 찾아라”라는 형태의 질의를 사용하기도 한다. 이로 인해 개체 및 관계 수에 따라 사전 정의되는 프롬프트 수가 급격히 늘어나지 않는다. 본 방법론은 주로 디코더 구조에서 각 프롬프트에 의해 생성된 답변이 문서 하나에 할당된 전역 테이블의 값을 채워 넣는 방식이다. [25]–[29] 등의 대표 연구는 문서 단위 관계 추출 시 전역적 수준의 표를 통해 정보를 유지하는 방법을 제안 및 적용한다. 디코더 기반 PLM[15]은 2~3 단락(평균 1,024 토큰)을 입력으로 다룰 수 있으나, 이 또한 문서 전체를 한 번에 다룰 순 없다. 상대적으로 길어진 입력 문맥과 전역적 템플릿 생성 등의 보완 방법을 통해 비교적 높은 성능을 보인다. 하지만, 평균 문서 길이가 8,000 토큰에 육박하는 최근 벤치마크들을 다루게 되면 여전히 정보 고립화 및 추출 오류 발생하는 제한점이 있다.

본 논문에서는 문서 수준의 입력을 더 크게 확장하는 연구보다 관계 추출 시 발생하는 오류를 효과적으로 다루는 방법론을 연구하고자 한다.

### III. 제안 방법

그림 2는 제안하는 MaV 모델의 구조를 보여준다. MaV는 프롬프트 설계, MRC(Machine Reading Comprehension) model 그리고 다중 과제 학습 방법으로 구성된다. 세부 과제 별 제안 모델의 동작을 위해 목적 별로 전용 프롬프트를 설계한다. 모델 구조는 MRC model로 인코더 구조 PLM인 BERT[14]를 기반으로 하고 모델 하단에 세부 모듈들을 추가한다. 형식 별 프롬프트에 따른 모델의 답변을 통해 모든 세부 과제를 통합하여 학습한다.

#### 3-1 문제 정의

본 논문은 N차 관계(N-ary relation) 추출을 대상으로 한다. 관계 추출은 문서 내 K 개의 관계  $R = [r_1, r_2, \dots, r_k, \dots, r_K]$ 를 추출하는 과제다. N 차 관계 추출은 각 관계  $r_k$ 가 N 개의 서로 다른 형식 개체를 모두 포함한다.  $r_k = (e_1, e_2, \dots, e_N)$ 로 표현되고 N 개 형식의 개체들로 구성된다. 그림 1의 예에서 전체 개체 형식은 (Method, Task, Dataset, Metric)로  $|r|=N=4$ 인 경우이다. 이때 예측할 수 있는 관계는  $r_1 = (PF-DL DL, Semantic segmentation, VOC2007, mAP)$ 이 될 수 있다.

하나의 문서 D 를 여러 개의 짧은 단위 문맥 텍스트로 분할하여  $D = [C_1, \dots, C_i, \dots, C_{|D|}]$ 로 표현한다. |D| 는 단위 문맥 텍스트의 개수이고, 는 모델의 토큰 수용 크기에 맞춘 입력 단위인 단위 문맥 텍스트다. 예시로 길이가 2,000 토큰인 문서를  $|C_i|=500$  길이의 문맥으로 분할했다면,  $|D|=4$ ,  $D = [C_1, C_2, C_3, C_4]$ 가 된다.

모델의 입력은 문맥과 질의로 문맥 텍스트는  $C_i$ , 질의 텍스트는  $q_i^t$ 다. 하나의 문맥에 질의 셋  $Q_i = [q_i^1, \dots, q_i^T]$  내 질의들

을 사용한다. 문맥 별 질의는 최대 T 개로 모델의 하이퍼 파라미터로 설정된다. 모델의 출력은 각 질의에 대한 답변으로  $a_i^t = (a_i^{t,1}, a_i^{t,2}, a_i^{t,3}, a_i^{t,4})$ 가 된다. 순서대로 개체 텍스트, 개체의 형식, 개체의 핵심 유무(saliency) 그리고 관계의 전역적 검증 결과로 구성된 답변 튜플이다. 이어지는 두 개의 답변을 조합하여 개체 튜플  $E_j = (a_i^{t,1}, a_i^{t+1,2}, a_i^{t+1,3})$ 을 만든다. 그림 2의 예시에서  $E_1 = (PF-DL DL, Method, Yes)$ 와 같다. 모든 문맥  $C_1 \sim C_{|D|}$ 에 대해 반복하고 얻은 개체 튜플 셋을  $E = [E_1, \dots, E_j]$ 로 표현한다. 이는 문서 하나 D에 대해 추출한 모든 개체 튜플 셋이다.

최종 관계는 문서의 개체 튜플 셋을 통해 조합한다. 관계  $r_k = (e_1, \dots, e_N)$ 에서  $e$ 는  $r_k$ 를 구성하는 개체이다. 모든 N 개의 개체 형식이 포함되게 E의 조합을 사용한다. 그림 2의 예시와 같이  $|r|=N=4$  라면  $r_k = (e_1, e_2, e_3, e_4)$ 이고  $e_1 \sim e_4$ 는 각각 다른 형식의 개체이다. 해당 예시에서  $e_1$ 은 'Method'형식 개체로 조합한다. 구체적으로  $E_1^2 = 'PF-DL DL'$  혹은  $E_2^2 = 'ConvNet'$ 이 'Method'형식이라고 답변한 개체 셋이고  $e_1$ 이 될 수 있다. 동일하게  $e_2, e_3$ , 그리고  $e_4$ 도 'Task', 'Dataset', 'Metric' 형식으로 답변한 개체를 조합 한다.

#### 3-2 과제 분해 및 프롬프트 설계

MaV의 입,출력인 질의, 문맥 그리고 답변을 각 전용 프롬프트로 설계한다. 3개의 전용 프롬프트는 그림 1, 2의 예시와 같이 동작한다. 그림 1의 전용 프롬프트  $P_1 \sim P_3$ 는 각각 개체 추출( $P_{LV}$ ), 개체 형식 및 핵심 유무 판단( $P_{SP}$ ) 그리고 관계 판단( $P_{GV}$ ) 프롬프트이다. 분해된 과제들에 대해 도출한 답변들을 부분적으로 조합하여 최종 답변을 만든다.

먼저, 개체 추출과 개체 형식 및 핵심 유무를 판단하는 2개의 프롬프트는 3개의 세부 과제인 개체 추출, 개체 형식 분류

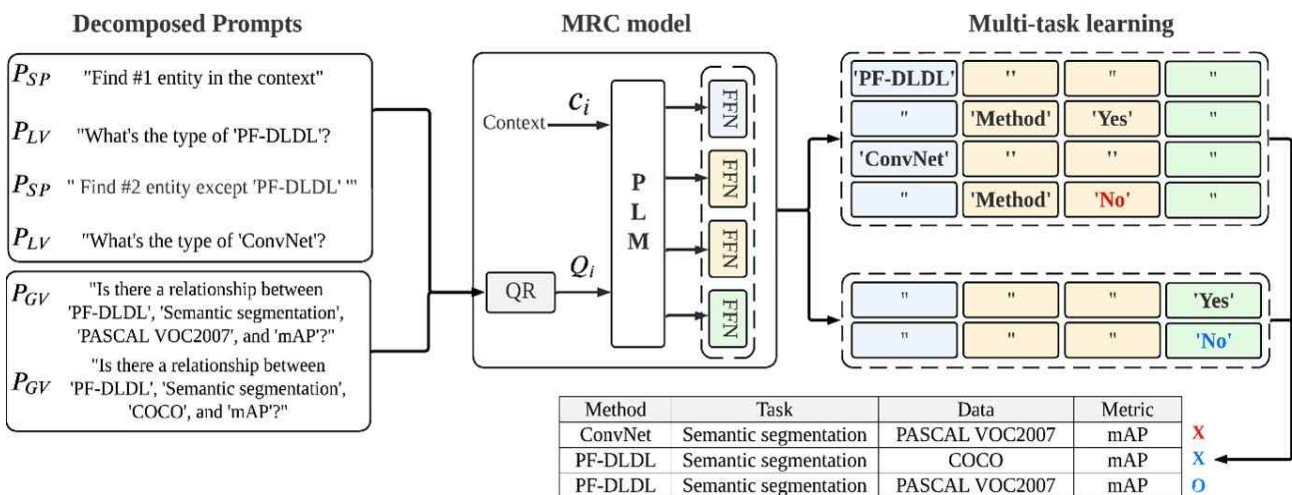


그림 2. 전체 모델 구조  
Fig. 2. Overall architecture



그리고 지역 검증에 사용된다. 수식 1은 질의 재구성(QR; Query Reformulation) 함수로 경우에 따라 다른 프롬프트를 사용하여 질의 텍스트를 구성한다.

$$QR(H,t,k) = \begin{cases} P_{LV}(H,t) & \text{if } t \% 2 = 0 \\ P_{SP}(H,t,k) & \text{otherwise} \end{cases} \quad (1)$$

이때,  $P_{LV}$ 는 문맥에서 개체를 추출하기 위한 질의,  $P_{SP}$ 는 추출한 개체를 지역적 검증하기 위한 질의를 만든다.  $P_{SP}$ 는 이전에 추출한 개체를 히스토리 셋  $H$ 에 순차적으로 저장하여 사용한다. 개체 히스토리  $H$ 의 일부인  $H_{1:s}$ , 현재 새로 추출할 개체의 순서  $k$ , 현재 질의 순서  $t$ 를 사용한다. 그림 2의 예시에서 ‘ConvNet’을 추출하기 위해  $q^3$ : “Find #2 entity except ‘PF-DLDDL?’”를 만드는데 이때  $k=2$ ,  $t=3$ ,  $s=1$ 이 된다.  $P_{LV}$ 는  $P_{SP}$ 를 통해 추출한 개체로 “What’s the type of ‘ConvNet?’”의 질의를 만들며  $H_{t-1}$  개체의 형식과 핵심 유무를 추론한다. 입력 문맥은 문서를 PLM의 수용 가능 토큰 길이로 분할한 단위 텍스트  $C_i$  시퀀스를 사용한다.

프롬프트  $P_{GV}$ 는 수식 2로 최종 관계를 조합하고 전역 검증하기 위한 질의를 만든다.  $P_{GV}$ 는  $P_{SP}$ 와  $P_{LV}$ 에 의한 문서 단위 결과인  $E$ 를 사용한다. 3.1과 같이 개체 셋  $E = [E_1, \dots, E_j]$ 를 개체 형식을 기준으로 분해 및 조합한다. 하나의 관계  $r_k$ 에 각 형식 별 개체가 모두 포함되게 조합하며, 그림 2에서  $P_{GV}$ 로 만든 관계  $r_1 = (\text{PF-DLDDL}, \text{Semantic segmentation}, \text{VOC2007}, \text{mAP})$ 과 같다.

$$QR(E) = P_{GV}(E) \quad (2)$$

질의 텍스트는  $r_1$ 의 개체를 토큰 시퀀스로 나열한 것으로 그림 2에서  $N=4$ 라면 “Is there a relationship between PF-DLDDL, Semantic segmentation, VOC2007 and mAP?”와 같다.  $P_{GV}$  프롬프트를 사용할 때 입력 문맥은 기존과 상이하다.  $r_k$ 을 구성하는 각 개체는 서로 다른 문맥으로부터 추출했으므로 각 개체를 추출했던 문맥의 토큰을 나열하여 새로운 문맥을 구성한다.

### 3-3 MRC model과 다중 과제 학습

MRC model은 그림 2와 같이 인코더 기반 PLM과 FFN(Feed Forward Network) 4개(FFN1~FFN4)로 구성된다. MRC model의 입력으로 문맥과 질의를 사용하고, 출력으로 답변을 추출한다. 모델의 입,출력은 3.2에서 설계한 프롬프트에 맞춰 동작한다.

FFN1은 단위 텍스트인 문맥 내 개체를 추출한다. 문맥 내 특정 토큰 시퀀스의 범위를 찾아내는 것으로 해당 범위의 텍스트를 답변으로 한다. FFN2~FFN4은 각 세부 과제를 위한 분류 모델이다. 입력 프롬프트에 따라 각각 개체 형식 판단,

개체 핵심 유무 판단 그리고 관계의 전역적 검증을 수행한다. 각 모델은 공백 문자 답변을 위한 클래스를 추가한 분류 모델이다. 예를 들어 그림 2의 예시에서 개체 형식  $N=4$ 이므로 FFN은 개체형식 4개와 공백 답변을 포함한 5개의 클래스에 대한 분류 모델이 된다.

MaV는 위 과정을 단계별로 학습하지 않고 다중 과제 학습 방법으로 통합 학습한다. 모든 학습 인스턴스는 4개로 분해한 과제를 동시에 학습한다. 각 전용 프롬프트는 모든 과제를 위한 답변을 하지만, 현재 프롬프트의 목적이 아닌 과제는 공백 문자를 답변하게 학습한다. 그림 2의 다중 과제 학습 예시에서  $P_{LV}$ 에 대한 답변으로 개체 형식 ‘Method’와 핵심 유무 ‘Yes’ 그리고 나머지 2개 과제에 대해 공백 문자를 답변하는 것과 같다. 추론 시에는 개체 추출부터 전역 검증까지 이전 답변을 기반으로 순차적으로 진행한다.

모델 학습은 수식 3의 목적 함수(objective function)를 사용한다. 각 FFN 4개의 오차  $L_1 \sim L_4$ 를 더하여 총합 오차(total loss)  $L$ 로 계산 하는 공동 학습(joint learning) 방식이다. 각 목적 함수는 기존 분류 및 질의응답에 사용하는 목적 함수 교차 엔트로피(CE; Cross-Entropy)이다.

$$L = \alpha L_1 + \beta L_2 + \gamma L_3 + \delta L_4 \quad (3)$$

각 오차 값과 곱하는 인자  $\alpha, \beta, \gamma, \delta$ 는 교정 수치(scaling factor)로 세부 과제 간의 오차를 비슷하게 하여 안정적인 학습을 유도한다.

본 논문에서는 노이즈 기반으로 학습 인스턴스를 추가한다. 추가한 학습 인스턴스는 관계를 구성하는 개의 개체 중  $m$ 개 개체를 노이즈인 오답 인스턴스(negative instance)로 변경한다. 오답은 ‘비핵심’ 개체를 통해 빈도수 기반 분포(frequency distribution)를 만들고 샘플링 한다. 그림 3의 예시는 ‘COCO, F1, Age estimation’등의 오답 개체를 포함한 학습 인스턴스를 추가한 것을 보여준다. 위 과정을 통해  $P_{GV}$ 는 다양한 인스턴스를 학습하고 적절한 관계 추론을 할 수 있다.

### 3-4 지역 및 전역적 검증

LV는 MRC model에서 추출한 개체를 검증한다. 개체에 대한 답변 중 개체 핵심 유무를 기반으로 비 핵심으로 분류한 개체들을 제외한다. 각 개체는  $P_{LV}$ 와  $P_{SP}$ 로 얻은 순차적인 답변을 통해 조합한다. 그림 4에서 개체(‘ConvNet’, ‘Method’, ‘No’) 예시는 ‘ConvNet’은 ‘Method’형식 개체이며 지역적으로 핵심 개체가 아니라고 답변했기 때문에 최종 개체 셋에 포함하지 않는다. (‘PF-DLDDL, Method, Yes’), (‘COCO, Data, Yes’)은 최종 개체 셋  $E$ 에 순차적으로 저장한다.

GV는 MRC model에서 조합한 관계를 검증한다. 관계는

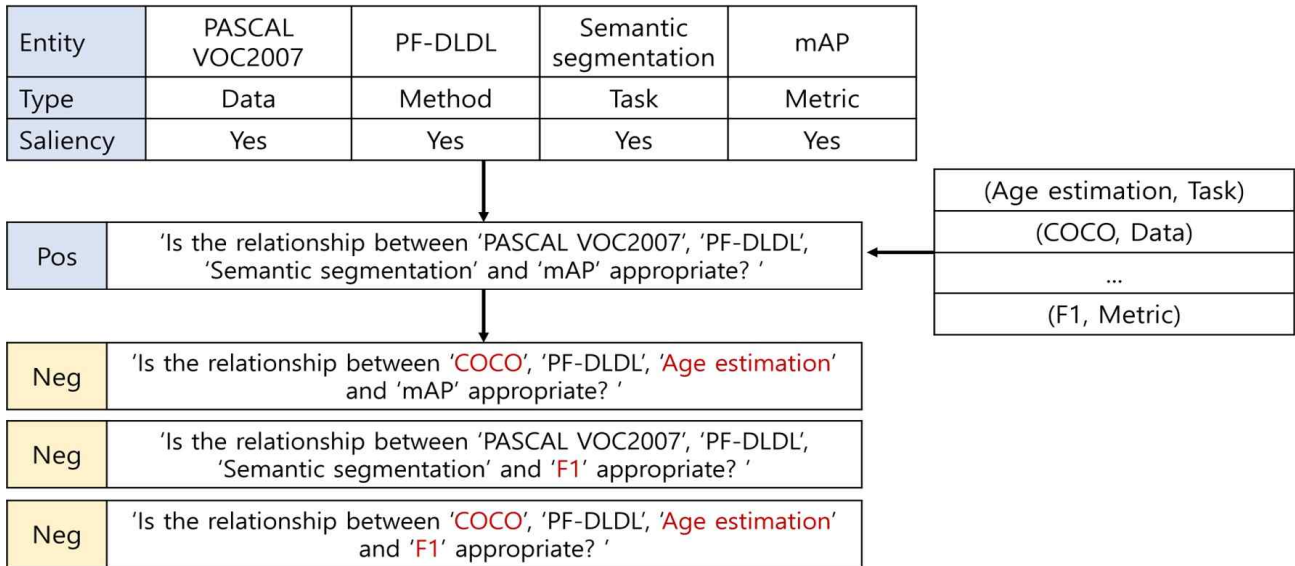


그림 3. 학습 인스턴스 추가 예시  
 Fig. 3. Example of making negative instances

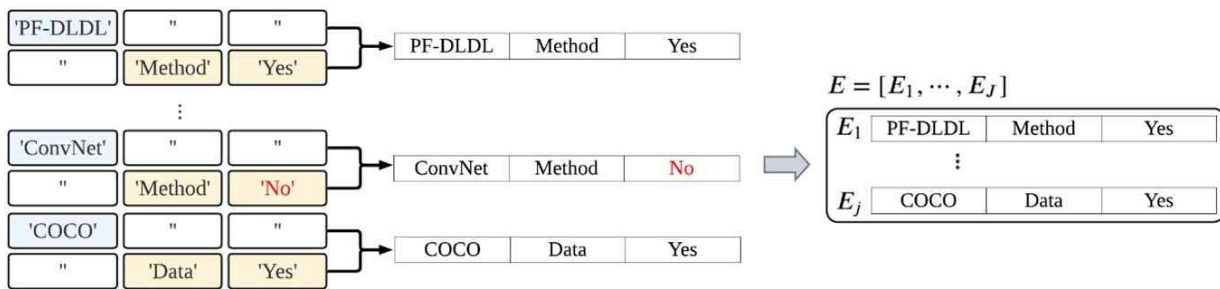


그림 4. 지역 검증 과제 예시  
 Fig. 4. Example of LV

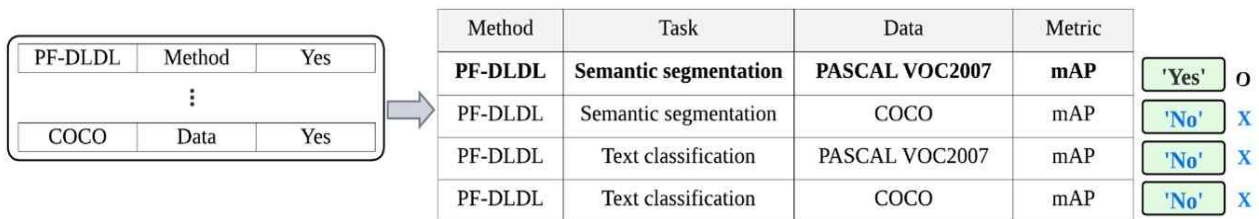


그림 5. 전역 검증 과제 예시  
 Fig. 5. Example of GV

문서  $D$ 에서 지역 검증된 개체 셋인  $E = [E_1, E_2, \dots, E_j, \dots, E_N]$ 를 통해 조합된다. 관계  $r_k = (e_1, e_2, \dots, e_N)$  조합은 각 개체 답변 중 개체 형식을 이용한다.  $N$  개의 개체 형식 클래스가 모두 포함되게 조합한  $r_k$ 는  $P_{GV}$  답변을 통해 전역 검증한다. 그림 5에서  $N=4$ 이고  $r_1 = ('PF-DL DL', 'Semantic segmentation', 'PASCAL VOC2007', 'mAP')$  와 같이 각 개체 형식이 모두 포함되게 관계를 조합할 수 있다. 예시의  $r_1 \sim r_4$  관계를  $P_{GV}$  프롬프트를 통해 질의 텍스트로 만들고 답변을 추론한다. 추

론한 답변에 따라 전역 검증 답변이 'Yes'인 관계  $r_1$ 을 최종 관계로 선정한다.

### IV. 실험 및 분석

#### 4-1 데이터 셋

본 논문에서 제안한 방법론과 기존 방법론을 비교하기 위

하여 SciREX[12] 벤치마크 데이터셋을 사용했다. SciREX는 과학 문서(scientific article), 특히 최신 기술 논문의 핵심 내용을 다룬다. 논문의 전문과 전문 내에 존재하는 개체들 간의 관계를 포함하며, 각 개체들은 논문의 방법론과 평가 방법 등의 핵심 내용에 해당한다. 논문은 하나의 문서로 평균 22개의 단락, 5,737개의 단어로 구성되며 총 438개의 문서가 존재한다.

각 문서의 최소 1개에서 5개의 핵심 기술 관계가 표기되어 있다. 핵심 기술 관계는 주요 과제(task), 데이터 셋(dataset), 평가 지표(metric) 그리고 제안 방법론(method) 네 가지 개체로 구성된다. 네 가지 개체가 모두 포함된 관계는 4차 관계(4-ary relation)이라 하고, 네 가지 개체 중 두 가지 개체가 유효한 관계를 2차 관계(binary relation)이라고 표현한다. 각 개체는 언급(mention), 핵심 유무(saliency) 그리고 개체 형식(type) 정보가 있다. 언급은 개체가 문맥에 어떤 텍스트로 언급 되어있는지, 핵심 유무는 해당 개체가 현 문서에서 주요한 개체인지 그리고 개체 형식은 위 네 가지 중 한 개로 문맥, 데이터 셋, 평가 지표, 제안 방법론 중 하나로 표현된다. 문서 당 평균 언급, 핵심 개체 수, 각 관계 수는 표 1과 같다.

표 1. SciREX 데이터 셋 내 개체 통계

Table 1. SciREX dataset statistics

Properties (per document)	
# of average mentions	360
# of average salient entities	8
# of average binary relations	16
# of average 4-ary relations	5

#### 4-2 실험 환경

실험은 파이토치(PyTorch) 프레임워크를 이용하여 구현하였고 NVIDIA RTX A6000 \* 8 환경에서 동일하게 학습 및 추론하였다.

SciREX는 공식적으로 전체 데이터(438문서)가 학습, 검증 및 추론 데이터가 7:3:3의 비율로 분할되어 있다. 본 연구에서 제안하는 모델의 성능을 실험하기 위해 MRC model의 기본 입, 출력 형태인 질의, 문맥 쌍으로 학습 및 평가 데이터를 구성하였다. 기존 연구와의 동일한 비교를 위하여 모델은 인코더 기반 PLM BERT[14] 구조에 과학 기술 분야 말뭉치 (scientific corpus) 로 사전학습한 SciBERT-base 를 사용하였다. 또한, 데이터 전, 후 처리 및 포함된 메타 데이터 등의 사용은 TEMPGEN[29]과 동일한 환경을 구성했다. SciBERT[30]는 기존 BERT-base와 동일하게 12개의 transformer 인코더 계층을 쌓아 구성하였다. 각 과제를 위한 세부 모델 FFN1~FFN4는 모두 동일한 세팅으로 같은 파라미터 사이즈와 학습 파라미터로 진행하며 총합 오차를 구성하는 각 과제 별 교정 수치는 0.7, 0.1, 0.1, 0.1로 통합 1.0

이 되게 했다. 표 2는 실험에 사용된 모델의 주요 하이퍼 파라미터와 구성이다.

표 2. 하이퍼 파라미터 설정

Table 2. Hyper parameter setup

# hidden layers	12
# attention head	12
Embedding size	768
FFN hidden size	3072
Dropout rate	0.1
Epoch	30
Optimizer	Adam
# prompts	8
# perturb entities	3
# history set	5
Scaling factor	0.7, 0.1, 0.1, 0.1
Max sequence length	512

#### 4-3 비교 모델 및 성능 평가 방법

제안 모델은 SciREX에 대한 기존 방법론과의 성능 비교를 통해 평가를 진행한다. SciREX-P[12], TANL[22], DyGIE[23] 그리고 GRIT[24]은 질의응답 및 분류 기반 방법으로 인코더 구조 PLM BERT[14]를 사용한다. TEMPGEN[29]은 생성 기반 방법으로 디코더 구조 기반 PLM BART[15] 를 사용한 관계 추출 모델이다. 제안 방법 및 비교 방법론들은 모두 개체 추출부터 관계 추출 단계까지 하나의 파이프라인으로 이어진 end-to-end 방식으로 실험한 성능이다. [23]의 경우와 같이 원본 연구에서 문장 단위로 실험한 방법론은 문서 단위로 재실험한 성능을 기재했다.

평가 지표는 벤치마크 데이터 셋을 제안한 기존 연구[12]에서 제안한 평가 방식으로 binary F1 과 4-ary F1 지표를 사용했다. 4-ary F1은 관계를 구성한 4개 개체 모두가 정답과 일치한 지 판단하고, binary F1 은 관계 내 2개 개체 쌍이 정답과 일치하지 모든 개체 쌍인 6가지 경우의 수를 평균 내어 계산한다.

#### 4-4 결과 및 분석

MaV와 기존 방법론들의 SciREX에 대한 주 성능 평가는 표 3과 같다. Binary F1 지표에서 MaV와 같은 방법론인 질의응답 및 분류 방법 SciREX-P[12], TANL[22], DyGIE[23], GRIT[24] 대비 최대 + 4.93 높은 성능을 기록했으며, 생성 기반 방법 TEMPGEN[29] 대비 비슷한 성능을 달성했다. 4-ary F1 지표의 경우 2배 이상 긴 입력 길이를 사용하는 기존 연구 대비 최대 + 2.12 높은 성능을 보였다. 이를 통해 단순한 대량의 문맥 정보 사용 및 최신 생성 기반 방법 적용보다, 긴 입력 문서를 효과적으로 분해하고 중요 정보를 재조합하여 다루는 방법이 중요함을 알 수 있다.

표 3. SciREX 벤치마크에 대한 기존 연구와의 성능 비교

Table 3. Performance comparison of different RE model

Model	Binary			4-ary		
	P	R	F1	P	R	F1
TANL [22]	0.74	0.67	0.60	0.00	0.00	0.00
GRIT [24]	-	-	-	0.00	0.00	0.00
DyGLE++ [23]	2.90	12.80	3.80	-	-	-
SciREX-P [12]	6.50	<b>41.10</b>	9.60	0.79	<b>17.30</b>	0.80
TEMPGEN [29]	<b>17.11</b>	13.56	14.47	3.19	4.26	3.55
MaV (ours)	15.02	16.41	<b>14.53</b>	<b>6.38</b>	5.31	<b>5.67</b>

표 4, 5 그리고 6은 지역 및 전역적 검증 방법에 대한 분석 평가이다. 표 4는 MaV에서 각각 지역, 전역적 검증 과제(해당 프롬프트와 모델)를 제외하여 어떤 영향을 주었는지 분석한다. 전역적 검증 방법인 GV는 3장에서 설명한 모든 부분을 제거하였다. 개체 추출과 직접적인 연관이 있는 LV는 추출 관련 부분은 유지하고 핵심 개체를 판단하는 프롬프트와 모델 등을 제거하여 비교하였다. 표 4를 통해 LV와 GV는 모두 MaV의 최종 관계 추출 성능에 영향을 미치는 것을 알 수 있다. 각각 제거한 경우와 두 방법을 같이 제거한 경우 모두 성능이 대폭 하락하였기 때문이다. 또한, LV보다는 GV가 전체 성능 폭 하락에 큰 영향을 미침을 보인다. LV만 제거하고 GV를 유지한 방법은 4-ary 지표에서[29]를 뛰어 넘는 성능을 보인다. 이는 지역 수준에서 개체 추출에 대한 오류가 있더라도 전역 수준의 검증 단계에서 최종적으로 관계 오류를 다시 검증하기 때문이다.

표 4. 지역 및 전역 검증 방법 유무에 따른 성능 비교

Table 4. Performance comparison according to LV and GV

Model	LV	GV	Binary			4-ary		
			P	R	F1	P	R	F1
MaV	O	O	15.02	16.41	<b>14.53</b>	<b>6.38</b>	5.31	<b>5.67</b>
-LV	X	O	<b>16.07</b>	11.04	12.32	2.77	6.94	3.88
-GV	O	X	8.14	<b>46.09</b>	11.36	0.77	<b>32.40</b>	1.38
-LV, GV	X	X	5.10	38.58	7.89	0.13	24.07	0.27

표 5는 LV의 각 세부 옵션에 따른 비교 및 분석이다. 표 5를 통해 개체 히스토리 셋  $H$ 에서 가져오는 개체 수  $s$ 가 늘어남에 따라 성능이 하락하는 것을 볼 수 있다. SciREX[12] 전체 문서의 단락 별 평균 개체 수는 6개로  $s$ 가 커질수록 반복된 개체를 추출할 수 있기 때문이다. 추가로 단락 별  $H$ 가 아닌 문서 단위  $H$ 를 유지하면 성능이 하락함을 볼 수 있다. 현재 다루는 단위 텍스트와 문맥적으로 다루지 않는 개체를 사용할 수 있어 발생한 문제로 분석된다. 마지막으로  $H$ 를 사용하지 않으면 동일 개체를 반복 추출하여 성능이 매우 저조한 것을 볼 수 있다.

표 5. 지역 검증 단계에 대한 세부 성능 비교

Table 5. Performance comparison according to methods in LV

Model	Binary			4-ary		
	P	R	F1	P	R	F1
MaV	15.02	16.41	14.53	6.38	5.31	5.67
Using history set						
$s = 5$	15.02	16.41	14.53	6.38	5.31	5.67
$s = 7$	-	-	-	6.38	3.72	4.39
- history set	-	-	-	0.62	7.44	1.02
Using $D$ history set	7.04	46.45	10.66	0.71	27.31	1.11

표 6은 GV의 각 세부 옵션에 따른 비교 및 분석이다. 주로 문맥 재구성 함수, 오답 인스턴스 구성, 최종 조합 및 검증한 관계의 top-k 선정에 따른 비교를 진행했다. 단위 텍스트인 문맥을 재구성할 때 현재 방법이 아닌 임의의 단락 하나를 선택하는 방법은 성능 하락의 원인이 된다. 이는 4개의 개체를 모두 이해할 수 있는 문맥 정보가 결여되었기 때문이다. 오답 인스턴스 추가 시 빈도수 분포 기반 샘플링 제거와 노이즈 개체 수인  $m$ 을 다양하게 섞어 쓰는 것은 성능 측면에서 좋지 않다. 마지막으로 관계의 최종 유추에 사용하는  $k$ 가 커질수록 잘못된 유추한 관계가 더 많아져 성능이 하락하는 것을 분석할 수 있다.

표 6. 전역 검증 단계에 대한 세부 성능 비교

Table 6. Performance comparison according to methods in GV

Model	Binary			4-ary		
	P	R	F1	P	R	F1
MaV	15.02	16.41	14.53	6.38	5.31	5.67
Context reformulation						
$c_j$ with top-1 entity	16.96	11.63	12.96	4.25	3.19	3.54
Negative instance augmentation						
- negative sampling	16.07	11.12	12.46	6.38	2.39	3.30
Mixture $m(2\sim4)$	-	-	-	6.72	3.19	3.74
Top-k retrieval						
$k = 3$	12.99	16.36	13.33	2.83	2.92	2.69
$k = 5$	-	-	-	0.80	4.25	1.41

## V. 결론

본 논문은 효과적인 문서 내 관계 추출을 위해 MaV를 제안한다. MaV는 질의응답 및 분류 기반 방법의 확장 모델로 긴 문서를 분할 처리하면서 발생한 지역 및 전역적 오류를 검증하게 설계한 방법이다. 제안모델은 과제 분해, 프롬프트 설계 그리고 다중 과제 학습 방법을 통해 작은 PLM으로 긴 문서를 효과적으로 다룰 수 있게 한다. 생성 기반 방법론에 비



해 약 2배 작은 길이의 입력 텍스트를 사용함에도 문서 내 개체와 관계를 적절히 추론한다. MaV는 긴 문서 벤치마크인 SciREX에서 준수한 성능을 달성하였으나, 여전히 인코더 구조의 구조적 문제 및 효율적인 학습 방법의 부재 등 한계점이 존재한다. 따라서, 추후 연구에서는 인스턴스 추가 방법 등을 통한 새로운 학습 방법 고안과 인코더-디코더 구조 모델 확장을 통한 심화 연구가 가능할 것으로 예상할 수 있다.

## 감사의 글

본 연구는 2023년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업(2021RIS-004)과 2023년도 한국기술교육대학교 교수 교육 연구진흥과제의 지원을 받아 수행되었습니다.

## 참고문헌

- [1] L. Yao, H. Liu, Y. Liu, X. Li, and M. W. Anwar, "Biomedical Named Entity Recognition Based on Deep Neutral Network," *International Journal of Hybrid Information Technology*, Vol. 8, No. 8, pp. 279-288, August 2015. <https://doi.org/10.14257/ijhit.2015.8.8.29>
- [2] T. H. Nguyen, A. Sil, G. Dinu, and R. Florian, "Toward Mention Detection Robustness with Recurrent Neural Networks," arXiv:1602.07749, February 2016. <https://doi.org/10.48550/arXiv.1602.07749>
- [3] S. Zheng, F. Wang, H. Bao, Y. Hao, P. Zhou, and B. Xu, "Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Canada, pp. 1227-1236, July-August 2017. <https://doi.org/10.18653/v1/p17-1113>
- [4] F. Zhai, S. Potdar, B. Xiang, and B. Zhou, "Neural Models for Sequence Chunking," in *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI-17)*, San Francisco: CA, pp. 3365-3371, February 2017. <https://doi.org/10.1609/aaai.v31i1.10995>
- [5] X. Li, J. Feng, Y. Meng, Q. Han, F. Wu, and J. Li, "A Unified MRC Framework for Named Entity Recognition," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, pp. 5849-5859, July 2020. <https://doi.org/10.18653/v1/2020.acl-main.519>
- [6] X. Li, F. Yin, Z. Sun, X. Li, A. Yuan, D. Chai, ... and J. Li, "Entity-Relation Extraction as Multi-Turn Question Answering," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, pp. 1340-1350, July-August 2019. <https://doi.org/10.18653/v1/p19-1129>
- [7] D. Wadden, U. Wennberg, Y. Luan, and H. Hajishirzi, "Entity, Relation, and Event Extraction with Contextualized Span Representations," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China, pp. 5784-5789, November 2019. <https://doi.org/10.18653/v1/d19-1585>
- [8] T. Zhao, Z. Yan, Y. Cao, and Z. Li, "Asking Effective and Diverse Questions: A Machine Reading Comprehension Based Framework for Joint Entity-Relation Extraction," in *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI-20)*, Yokohama, Japan, pp. 3948-3954, January 2021. <https://doi.org/10.24963/ijcai.2020/546>
- [9] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "SQuAD: 100,000+ Questions for Machine Comprehension of Text," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP 2016)*, Austin: TX, pp. 2383-2392, November 2016. <https://doi.org/10.18653/v1/d16-1264>
- [10] M. Joshi, E. Choi, D. S. Weld, and L. Zettlemoyer, "TriviaQA: A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Canada, pp. 1601-1611, July-August 2017. <https://doi.org/10.18653/v1/p17-1147>
- [11] Y. Yao, D. Ye, P. Li, X. Han, Y. Lin, Z. Liu, ... and M. Sun, "DocRED: A Large-Scale Document-Level Relation Extraction Dataset," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, pp. 764-777, July 2019. <https://doi.org/10.18653/v1/p19-1074>
- [12] S. Jain, M. van Zuylen, H. Hajishirzi, and I. Beltagy, "SciREX: A Challenge Dataset for Document-Level Information Extraction," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, pp. 7506-7516, July 2020. <https://doi.org/10.18653/v1/2020.acl-main.670>
- [13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, ... and I. Polosukhin, "Attention is All You Need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS '17)*, Long Beach: CA, pp. 6000-6010, December

2017. <https://doi.org/10.48550/arXiv.1706.03762>
- [14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019)*, Minneapolis: MN, pp. 4171-4186, June 2019. <https://doi.org/10.48550/arXiv.1810.04805>
- [15] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, ... and L. Zettlemoyer, "BART: Denoising Sequence-to-Sequence Pre-Training for Natural Language Generation, Translation, and Comprehension," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, pp. 7871-7880, July 2020. <https://doi.org/10.18653/v1/2020.ac1-main.703>
- [16] X. Du, A. Rush, and C. Cardie, "Template Filling with Generative Transformers," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Online, pp. 909-914, June 2021. <https://doi.org/10.18653/v1/2021.naacl-main.70>
- [17] X. Han, W. Zhao, N. Ding, Z. Liu, and M. Sun, "PTR: Prompt Tuning with Rules for Text Classification," *AI Open*, Vol. 3, pp. 182-192, November 2022. <https://doi.org/10.1016/j.aiopen.2022.11.003>
- [18] J. Liu, Y. Chen, K. Liu, W. Bi, and X. Liu, "Event Extraction as Machine Reading Comprehension," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020)*, Online, pp. 1641-1651, November 2020. <https://doi.org/10.18653/v1/2020.emnlp-main.128>
- [19] Z. Chen and C. Guo, "A Pattern-First Pipeline Approach for Entity and Relation Extraction," *Neurocomputing*, Vol. 494, pp. 182-191, July 2022. <https://doi.org/10.1016/j.neucom.2022.04.059>
- [20] L. Qiu, H. Zhou, Y. Qu, W. Zhang, S. Li, S. Rong, ... and Y. Ru, "QA4IE: A Question Answering Based Framework for Information Extraction," in *Proceedings of the 17th International Semantic Web Conference (ISWC 2018)*, Monterey: CA, pp. 198-216, October 2018. [https://doi.org/10.1007/978-3-030-00671-6\\_12](https://doi.org/10.1007/978-3-030-00671-6_12)
- [21] O. Levy, M. Seo, E. Choi, and L. Zettlemoyer, "Zero-Shot Relation Extraction via Reading Comprehension," in *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, Vancouver, Canada, pp. 333-342, August 2017. <https://doi.org/10.18653/v1/k17-1034>
- [22] X. Du and C. Cardie, "Document-Level Event Role Filler Extraction Using Multi-Granularity Contextualized Encoding," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, pp. 8010-8020, July 2020. <https://doi.org/10.18653/v1/2020.ac1-main.714>
- [23] G. Paolini, B. Athiwaratkun, J. Krone, J. Ma, A. Achille, R. Anubhai, ... and S. Soatto, "Structured Prediction as Translation between Augmented Natural Languages," arXiv:2101.05779, December 2021. <https://doi.org/10.48550/arXiv.2101.05779>
- [24] X. Du, A. Rush, and C. Cardie, "Grit: Generative Role-Filler Transformers for Document-Level Event Entity Extraction," in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*, Online, pp. 634-644, April 2021. <https://doi.org/10.18653/v1/2021.eacl-main.52>
- [25] X. Wu, J. Zhang, and H. Li, "Text-to-Table: A New Way of Information Extraction," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, Dublin, Ireland, pp. 2518-2533, May 2022. <https://doi.org/10.18653/v1/2022.acl-long.180>
- [26] S. Li, H. Ji, and J. Han, "Document-Level Event Argument Extraction by Conditional Generation," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Online, pp. 894-908, June 2021. <https://doi.org/10.18653/v1/2021.naacl-main.69>
- [27] L. Xu and J. Choi, "Modeling Task Interactions in Document-Level Joint Entity and Relation Extraction," in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Seattle: WA, pp. 5409-5416, July 2022. <https://doi.org/10.18653/v1/2022.naacl-main.395>
- [28] X. Du, S. Li, and H. Ji, "Dynamic Global Memory for Document-level Argument Extraction," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, Dublin, Ireland, pp. 5264-5275, May 2022. <https://doi.org/10.18653/v1/2022.acl-long.361>
- [29] K.-H. Huang, S. Tang, and N. Peng, "Document-Level Entity-Based Extraction as Template Generation," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Punta Cana, Dominican Republic, pp. 5257-5269, November 2021. <https://doi.org/10.18653/v1/2021.emnlp-main.426>

- [30] I. Beltagy, K. Lo, and A. Cohan, "SciBERT: A Pretrained Language Model for Scientific Text," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China, pp. 3615-3620, November 2019. <https://doi.org/10.18653/v1/d19-1371>



**조성웅(Seongung Jo)**

2020년 : 한국기술교육대학교  
컴퓨터공학 (공학사)  
2023년 : 한국기술교육대학교  
컴퓨터공학 (공학석사)

※ 관심분야 : 딥러닝(Deep Learning), 자연어처리(Natural Language Process) 등



**오흥선(Heung-Seon Oh)**

2006년 : 한국항공대학교 컴퓨터공학  
(공학사)  
2009년 : 한국과학기술원 전산학  
(공학석사)  
2014년 : 한국과학기술원 전산학  
(공학박사)

2013년~2018년: 한국과학기술정보연구원(KISTI)

2018년~현재: 한국기술교육대학교 컴퓨터공학부 교수

※ 관심분야 : 딥러닝(Deep Learning), 자연어처리(Natural Language Process) 등