

# 다중 정보와 Self-Attention을 이용한 관계 추출

김정무<sup>10</sup>, 이승우<sup>2</sup>, 차정원<sup>1\*</sup>

창원대학교<sup>1</sup>, 한국과학기술정보연구원<sup>2</sup>  
{gersanga<sup>0</sup>,jcha\*}changwon.ac.kr, swlee<sup>2</sup>@kisti.re.kr

## Relation Extraction Using Self-attention with Multi Grained Information

Jeong-Moo Kim<sup>10</sup>, Seung-Woo Lee<sup>2</sup>, Jeong-Won Char<sup>1\*</sup>

Changwon National University<sup>1</sup>, Korea Institute of Science and Technology Information<sup>2</sup>

### 요약

관계 추출은 문서에서 존재하는 트리플(주어, 관계어, 목적어)형식에 해당하는 단어를 추출하는 작업을 뜻한다. 본 논문에서는 멀티헤드 셀프 어텐션을 이용하여 트리플 중 주어나 목적어를 찾는 구조를 제안한다. 한국어 위키피디아와 DBpedia의 관계어를 단어 임베딩을 통해 벡터를 생성하고 입력한다. 초록과 관계어의 어텐션 이후 멀티 헤드 셀프 어텐션 구조를 통해 초록 중 관계어와 관련 있는 단어들의 가중치가 높아진다. 멀티헤드 셀프 어텐션 과정을 반복하여 주요 단어들의 가중치가 계속해서 높아진다. 이를 입력으로 하여 정답이 될 단어의 시작과 끝을 선택한다. 제안 방법으로 직접 구축한 한국어 관계 추출 데이터셋을 대상으로 F1 0.7981의 성능을 보였다. 제안 방법은 관계어와 같이 단순한 정보만을 이용하고도 초록에서 적절한 정답 단어를 추출할 수 있음을 확인하였다. 관계어의 범위를 확장함으로써 나아가 육하원칙(5W1H)과 같은 이벤트의 추출에도 활용할 수 있을 것이다.

주제어: 관계 추출, 셀프 어텐션, 멀티헤드 어텐션, 기계학습, 딥러닝

### 1. 서론

수많은 문서들의 데이터화로 문서 내에서 핵심 내용을 파악하는 것이 중요시 되고 있다. 관계 추출(Relation Extraction)은 문서에서 존재하는 트리플(주어( $X$ ), 관계어( $R$ ), 목적어( $O$ ))형식에 해당하는 단어를 추출하는 작업을 뜻한다. 이러한 트리플을 문서 혹은 문장의 지식 베이스(Knowledge base)라고 불린다. 우리는 다음 예제와 같이 트리플형식 중 관계어를 입력으로 사용하여 주어나 목적어를 추출한다.

“윌 허드 Will Heard 1991년 4월 26일은 잉글랜드의 싱어송라이터이다”

( “윌 허드 “ , “직업” , “싱어송라이터” )

이 문제를 해결하려면 먼저 1) 주어와 목적어에 해당하는 구를 인지하고 2) 이들의 관계를 결정해야 한다. 일반적으로 주어와 목적어를 결정하는 작업은 개체명 인식기를 이용하여 진행한다. 또한 관계 결정은 정해진 관계에서 높은 연관성을 갖는 관계를 선택한다. 그렇지만 주어, 목적어 결정과 관계 결정을 하는 과정에서 오류가 발생하며 이 오류가 전체 성능 저하를 가져오게 된다.

본 연구에서는 관계 추출 문제를 ( $X, R, O$ ), ( $S, R, X$ )와 같이 주어와 관계를 주어졌을 때 목적어를 결정하거나 관계와 목적어가 주어졌을 때 주어를 결정하는 문제로 변형한다. 이 정의를 토대로 딥러닝을 이용하여  $X$  를 결정한다.

본 논문은 1장 서론에 이어 2장에서는 관계 추출과 관

련된 연구를 설명한다. 3장에는 제안하는 구조에 대한 설명과 4장에서 구축한 데이터에 대한 설명을 진행한다. 5장에서 실험 설정 및 성능을 기술한다. 6장에서는 모델이 추출해낸 결과를 분석한다. 마지막 7장에서 결론을 도출한다.

### 2. 관련 연구

최근 딥러닝(deep learning) 모델을 이용한 관계 추출 연구가 많이 진행되었다. 첫 번째로 CNN(Convolution Neural Network) 방법론이다. [1]에서는 문장 내 유용한 자질을 학습하는 CNN의 특징을 이용했다. 구문 정보를 이용하기 위해 확장 최단 의존 경로를 자질로 사용하는 원거리 지도(Distant supervision) 학습을 연구했다 이를 통해 입력 데이터에서 부분적인 특징을 추출하여 관계를 표현한다. [2]에서는 단어 임베딩으로 표현된 문장 벡터와 개체 정보를 활용한다. 문장에서 개체와 각 단어들의 상대적 거리를 나타내기 위해 개체 위치 임베딩(Position Embedding)값을 입력으로 사용한다. 두 번째는 RNN(Recurrent Neural Network) 방법론을 이용한 연구들이 진행되었다. RNN은 문장의 문맥 정보를 효과적으로 활용할 수 있는 특징을 이용했다. 그 중 [3]은 개체가 문장 구조적으로 직접적으로 연결이 되어있다는 가정 하에 개체 간 구문 분석 정보를 자질로 사용한다. [4]는 개체 간에 발생하는 다중 관계 문제를 해결하기 위해 sequence-to-sequence를 활용하여 관계를

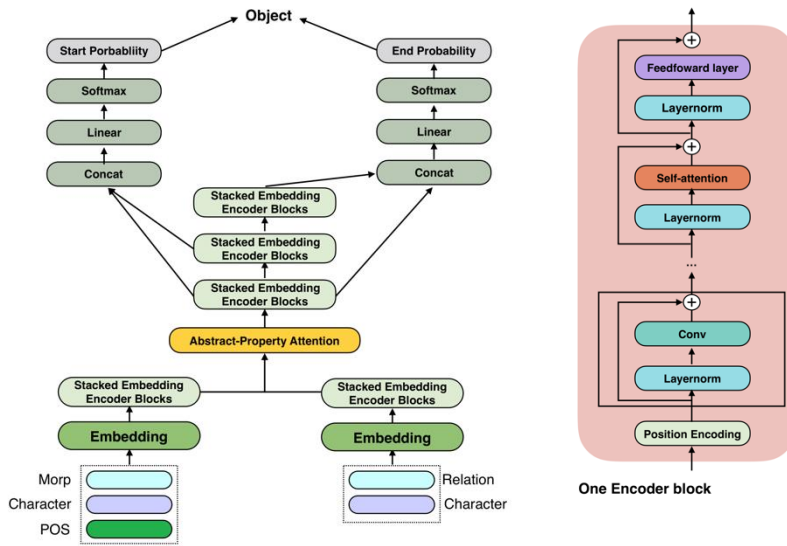


그림 1. 제안 모델 구조.

순차적으로 생성한다. 이를 통해 분석된 각 개체가 어떤 관계를 지니는지 학습하여 주어와 목적어 두 관계에 포함될 가능성이 있는 단어들에 대해 유연하게 대처하였다.

하지만 이러한 방법들은 단일문장에서 관계를 추출하거나 혹은 개체 분석, 문장 분석 등 다양한 오류를 야기할 수 있는 분석 방법을 사용하고 있다. 이에 우리는 입력 단위는 여러 문장으로 구성된 위키피디아를 사용하며, 분석 오류를 줄이기 위해 연관된 DBpedia의 관계어를 입력으로 사용하였다. 이는 [5]과 같은 기계 독해(Machine Reading Comprehension) 해결 방법을 활용한 것이다. 입력된 데이터들로 연관 있는 관계 단어를 추출하기 위해 멀티 헤드 어텐션 기법을 적용하여 정답을 추출했다.

### 3. 제안 방법

#### 3.1 문제 정의

본 논문에서는 다음과 같이 문제를 정의한다. 본문  $D = \{S_1, S_2, \dots, S_n\}$  는  $n$  개의 문장으로 이루어져 있다. 문장  $S_j = \{w_1, w_2, \dots, w_m\}$  는  $m$  개의 단어(형태소)로 이루어져 있다. 또한 한 단어  $w_i = \{e_1, e_2, \dots, e_l\}$  는  $l$  개의 음절로 이루어져 있다. 본 논문에서는 본문  $D$ 와 주어 혹은 목적어  $w_i$ , 그리고 관계  $R = w_i$  가 주어졌을 때 본문에서  $A = \{e_i, e_{i+1}, \dots, e_{i+j}\}$  를 구하는 것이다.

#### 3.2 제안 방법 개요

그림 1은 본 논문에서 제안한 시스템의 모델 구성도이다. 모델은 크게 입력 인코더 층, 어텐션 층, 출력 층으로 나눌 수 있다. 입력 인코더 층에서 2 가지 임베딩(입력된 데이터 임베딩, 스택 임베딩 인코더 블록)을 수행한 후 어텐션 층에서 또한 2 가지의 어텐션(초록-관계어 어텐션, 스택 모델 인코더 블록)을 거친 후 출력 층에서 주어 혹은 목적어가 될 영역의 시작과 끝을 계산한다.

#### 3.3 입력 데이터 임베딩

데이터는 형태소 단위로 입력한다. 한글의 조사나 어미 등을 제외한 단어의 학습을 위한 것이다. 따라서 단어(형태소) 임베딩, 음절 임베딩, 형태분석 임베딩을 함께 사용한다. 본 논문의 실험에서는 사전 학습된 임베딩을 사용하지 않고 역전파를 통해 학습이 되도록 설정했다. 생성된 형태소 임베딩  $w_w$ , 킨볼루션을 거친 음절 임베딩  $x_c$ 와 형태분석 임베딩  $x_p$ 를 연결(concatenation)하여  $[x_w: x_c: x_p] \in R^{p_1+p_2}$  사용한다. 최종적으로 두 개 층의 하이웨이 네트워크 [6]를 통해 최종 문자 임베딩  $x$ 를 생성한다.

#### 3.4 임베딩 인코더 블럭

임베딩 인코더 블럭은 [합성곱 층 # + 자기 주의 층 + feed forward]를 쌓은 것이다. 먼저 단어 위치에 따라 가중치를 적용하기 위해 포지션 인코딩(Position Encoding)을 수행한다. 제안하는 구조에서는 단어 임베딩 과정에서 합성곱(Convolution)이나 리커런트(Recurrent)를 사용하지 않아 위치 정보가 포함되지 않는다. 따라서 위치 정보를 이용하기 위해 삼각 함수 중  $\sin(2n)$ ,  $\cos(2n-1)$  함수를 이용하여 위치 정보를 추가해준다 [7]. 이후 각 은닉 상태(hidden state)의 분포를 정규화해주고 기울기(gradient) 값이 안정적인 값을 가짐으로 빠른 속도로 학습이 가능하게 하는 레이어 정규화(Layer Normalization)을 진행한다. 분리 가능한 깊이별 킨볼루션(Depthwise separable convolution)으로 킨볼루션을 수행한다 [8]. 분리 가능한 깊이별 킨볼루션은 각 채널마다 독립적으로 킨볼루션을 수행한다. 그 후 각 채널의 결과들을 모아 1개의 채널로 압축하는 킨볼루션을 한번 더 수행한다. 이 방식을 통해 기존의 킨볼루션보다 연산량을 크게 줄일 수 있다. 커널의 크기는 7이고 필터의 수는 96, 그리고 한 블럭에서 합성곱의 수는 4이다. 이후 레이어

정규화를 재 진행하고 멀티-헤드 어텐션(multi-head attention) [7] 매커니즘의 셀프 어텐션을 거친다.

멀티-헤드 어텐션은 전체 차원에 대해 설정한 헤드(head) 수 만큼 나누어 어텐션을 적용하는 것이며, 각각 헤드의 관점에서 보고 추후에 합치는 과정을 의미한다. 어텐션 층에서는 먼저 본문과 관계어 각각에 대한 멀티-헤드 어텐션이 수행된다. 본 논문에서는 헤드 수를 2로 설정하였다. 이후 레이어 정규화 [9]를 수행한 후 순방향 층(Feed-Forward layer)를 거친다. 이는 각 헤드가 동일한 입력으로 셀프 어텐션을 하게 되면 각 헤드에 따라 어텐션이 일정한 값으로 치우치는 것을 방지하기 균등하게 섞는 역할을 한다.

### 3.5 본문-관계 어텐션

$D$ 의 단어들 중 관계어와 연관된 단어 즉  $A$ 를 구하기 위해 본문-관계 어텐션을 수행한다. 이 어텐션은 총 4개의 벡터를 구하여 연결하여 사용한다. 인코딩 된 본문 벡터( $V$ )-관계 벡터( $G$ )를 이용하여 유사도 매트릭스  $S \in R^{n \times m}$ 를 구한다. 유사도 매트릭스를 이용하여 각 행과 열에 대해 softmax를 취해 정규화 한 유사도 매트릭스( $\bar{S}, \bar{S}^T$ )를 구한다. 정규화 한 유사도 매트릭스와 관계 벡터를 이용하여  $A = \bar{S} \cdot g^T \in R^{n \times d}$ 과 같이 어텐션을 계산한다. 여기서 사용하는 유사도 함수는 [10]에서 제안한 함수  $f(g, c) = W_0[G, c, G \odot c]$ 를 사용한다. 여기서  $\odot$ 는 요소별 곱이고  $W_0$ 는 학습 가능한 변수이다. 또한 [11]이 제안한 DCN을 이용하여 관계 벡터( $G$ )-본문 벡터( $V$ ) 어텐션을 구한다. 그리고 난 후에 관계 벡터( $G$ )-본문 벡터( $V$ )를 함수  $B = \bar{S} \cdot \bar{S}^T \cdot C^T$ 를 사용하여 구한다.

### 3.6 모델 인코더

[11]에서 제안한 방법과 유사하게 각 위치에서 이 층의 입력은  $[c, a, c \odot a, c \odot b]$ 이다. 여기서  $a, b$ 는 어텐션 행렬  $A, B$ 의 열에 해당한다. 이 층의 파라미터는 블록 안의 합성곱의 수가 2과 블록의 수가 7인 것을 제외하면 3.4의 임베딩 층과 유사하다. 우리는 인코더의 3번 반복 사이의 가중치는 공유한다.

### 3.7 출력 층

이 층에서는 주어와 목적어 해당하는 어구의 처음과 끝을 결정한다. 여기서  $M_0, M_1, M_2$ 는 모델 인코더의 출력 결과이다.  $W_i$ 는 가중치이고 마지막으로 목적 함수는 정답이 시작과 끝에 대한 출력 결과의 차이로 정의된다.

$$p^1 = \text{softmax}(W_1[M_0; M_1]),$$

$$p^2 = \text{softmax}(W_2[M_0; M_2])$$

여기서  $y_i^1, y_i^2$ 는  $i$ 에 대한 시작과 끝의 정답위치이다.  $\theta$ 는 학습된 파라미터를 담고 있다.

$$L(\theta) = -\frac{1}{N} \sum_i^N [\log(p_{y_i^1}^1) + \log(p_{y_i^2}^2)]$$

## 4. 실험

### 4.1 데이터 구축

실험에는 한국어 위키피디아, DBpedia 트리플 데이터를 이용하였다. 위키피디아의 문서 제목과 DBpedia의 주어가 완전히 일치하는 경우에 두 데이터를 합쳐 학습 데이터를 생성하였다.

수집된 위키피디아 데이터는 총 125,858개, 트리플 데이터는 총 980,155개이며 관계어의 종류는 총 643개이다. 트리플의 목적어나 주어의 단어가 초록에 존재하지 않는 경우 해당 트리플은 삭제하여 학습에 사용하지 않았다. 관계어의 경우 강화 학습을 기반으로 위키피디아 관계 데이터의 오류를 수정한 연구 [12]의 항목과 일치하는 관계어만 사용하였다. [12]의 기준은 각 관계어 별로 수집된 문장의 수가 4000개 이상인 30개의 관계어를 사용하였다. [12]와 수집된 데이터 중 일치하는 관계어는 총 12개이다. 선택된 관계어를 기준으로 학습 데이터를 구성하였다. 학습, 평가, 테스트 데이터의 통계는 표 1과 같다. 관계어 종류와 각각의 데이터의 수는 표 2와 같다.

표 1 학습 데이터셋 통계

	초록	트리플
학습	27,498	42,315
개발	3,099	5,261
테스트	3,125	5,269

표 2 관계어 종류 및 데이터 수

관계어	데이터 수	관계어	데이터 수
직업	23,173	국가	2,170
출생지	12,072	사망지	1,338
포지션	3,397	행정구역	1,121
위치	2,814	지휘관	959
국적	2,722	배우자	399
장소	2,373	언어	307

### 4.2 성능 및 실험 설정

표 3은 구축한 한국어 관계 추출 데이터셋에 대한 성능 결과이다. 성능 측정은 F1 스코어와 Exact Match(EM)로 측정한다. 또한 관계어 각각의 성능 및 micro, macro 성능을 포함했다. 우리는 표 3과 같이 음절 단위와 형태소 단위의 입력으로 실험을 진행했다. 형태소 단위의 입력의 경우 3.3에서 언급한 바와 같이 조사나 어미에 영향을 받지 않기 위함이다. 형태소 분석은 창원대학교의 형태소 분석기 [13]으로

표 3 관계 추출 실험 성능 표

모델	성능	관계어											Micro F1	Macro F1	
		직업	출생지	포지션	국적	장소	위치	국가	사망지	배우자	행정 구역	지휘관			언어
음절 단위	F1	75.27	92.16	89.07	94.12	39.54	56.50	78.04	56.43	77.44	74.08	36.48	55.70	79.01	68.77
	EM	69.98	89.56	80.79	92.49	22.65	30.63	75.45	53.84	57.14	57.14	4.76	38.46	73.88	56.07
형태소 단위	F1	73.50	90.44	85.55	91.92	26.60	50.43	73.58	51.28	68.40	50.34	28.23	23.07	77.21	59.44
	EM	70.83	88.85	84.24	91.73	22.72	37.08	72.64	47.69	60.86	42.86	23.52	23.07	74.80	55.51
형태소 단위 (+ POS 임베딩)	F1	75.09	91.62	91.32	91.73	40.86	58.20	77.36	63.07	86.95	81.74	38.82	43.58	<b>79.81</b>	<b>70.04</b>
	EM	71.42	90.86	89.04	91.73	37.01	42.69	77.36	63.07	78.26	71.42	29.41	38.46	<b>77.00</b>	<b>65.05</b>
형태소 단위 (오류 제외)	F1	78.28	92.60	88.38	93.52	31.20	56.92	76.76	52.45	0	84.61	50.00	45.45	81.79	62.51
	EM	78.28	92.60	88.38	93.52	31.20	56.92	76.76	52.45	0	84.61	50.00	45.45	81.79	62.51

실행하였다. 형태소-오류 제외한 실험은 형태소 분석이 완벽하지 않은 데이터를 모두 삭제한 실험이다. 삭제 대상은 표 4와 같이 조사에 붙은 음절이 하나의 단어가 될 수도 있는 경우 등이다.

경우이다. 두 번째는 관계어와 관련 있는 단어의 후보들이 가까운 위치에 없으며 그 수가 적은 경우이다. 표 6-8의 첫 번째 열의 정답 단어는 파란색으로, 모델이 출력한 오답 단어는 빨간색으로 표시하였다.

표 4 형태소 분석 오류 예

기존 단어	정답	형태소 분석 결과
시네그로스 주의	시네그로스 주	시네그로스 주의
모델	모델	모데 르
영화연출가	영화연출가	영화연출 가

실험에 사용한 파라미터는 표 5와 같다. 초록의 경우 학습 시 최대 길이가 제한되며 정답의 길이는 학습, 테스트 모두 길이에 제한을 둔다. 초록의 길이 제한은 설정 값 이상의 정보는 정답을 찾을 수 있는 충분한 정보량으로 제한을 둔 것이다. 정답의 길이 제한은 기계 독해에서도 동일하게 발생하는 문제로 정답의 길이가 길어질수록 모델이 적절한 정답을 찾기 어려워지는 문제로 적정 길이에 대한 정확도를 높이기 위해 제한한다.

표 5 실험 파라미터

파라미터	설정 값	파라미터	설정 값
초록 최대 길이	400	헤드 수	2
정답 최대 길이	10	Early Stop	10
Dropout	0.1	음절 임베딩 차원	64
형태소 임베딩 차원	300	POS 임베딩 차원	300

### 5. 결과 분석

제안 모델에서 정확한 정답을 추출할 수 있는 경우는 표 6의 예와 같이 크게 2가지로 볼 수 있다. 첫 번째는 초록 내에 관계어와 관련된 적절한 정보가 있는

표 6 정답 추출 예

초록	관계어	정답	추출
황태영 黃泰榮 1961년 7월 6일 음력 5월 24일 2017년 11월 22일 은 대한민국의 전 금융인자 수필가 칼럼리스트 기업인 문화기관단체인이다 ...	직업	금융인	금융인
로열캐리비안 인터내셔널 Royal Caribbean International 은 미국의 크루즈 기업이다 ... 본사는 미국 플로리다주 마이애미에 위치해 있다	출생지	대한민국	대한민국
	장소	마이애미	마이애미

추출한 결과가 오답으로 되는 공통적인 조건은 경우 또한 표 7의 예와 같이 크게 2가지로 볼 수 있다. 첫 번째는 하나의 초록에 동일한 관계어가 두 개 이상으로 정답 단어가 다양한 경우이다. 두 번째는 초록 내에 관계어와 관련 있는 단어의 후보들이 두 개 이상인 경우이다.

표 7 오답 추출 예

초록	관계어	정답	추출
실비아 헤르트루디스 마르티나 후스 Sylvia Gertrudis Martyna Hoeks	직업	모델	배우

1983년 6월 1일 는 네덜란드의 배우 모델이다 ...	직업	배우	배우
마르코 리시 Marco Risi 1951년 6월 4일 는 이탈리아의 영화 감독 각본가 영화 제작자이다 영화 감독 디노 리시의 아들이다 ...	직업	영화 제작자	영화 감독

그 외 관계어 마다 오답의 경향이 다름을 분석 할 수 있었다. 사망지와 배우자, 지휘관의 경우 장소나 인물 주변에 관계어와 연관 있는 정보가 부족할 경우 오답률이 높았다. 장소와 위치같이 번지수나 도로명까지 정답 단어로 구성되어있는 경우 제안 모델에서는 전체 영역에 대해 제대로 된 학습을 하지 못한 결과를 볼 수 있었다. 또한 DBpedia의 트리플 작성자에 따라 정답 단어를 구성하는 방식이 다름을 확인하였다. 이를 정답 범주의 다양성이라 하고 그 예는 표 8과 같다. 정답에 해당하는 모든 단어들로 정답을 정의하거나 각각의 단어들로 정답을 정의하는 경우이다. 이러한 경우 실제 정답과 다르지만 실제 의미는 일치하는 경우도 발생하였다.

표 8 정답 범주의 다양성

초록	관계어	정답	추출
전북고속 全北高速 은 전라북도의 시외버스 업체이다 주 사무소는 전라북도 전주시 덕진구 가리내로 30 금암동 70510 전주시외버스터미널 내에 있다 계열사로는 전주고속이 있다	장소	전라북도	덕진구 가리내로 30
		덕진구	
		가리내로 30	
마리오 레미나 Mario Lemina 1993년 9월 1일 는 사우샘프턴 FC에서 중앙 미드필더로 활약하고 있는 가봉의 축구 선수이다 FC ...	포지션	미드필더	중앙 미드필더

## 6. 결론 및 향후 연구

제안 모델 예측을 분석한 결과 제안하는 구조로 적절한 정답을 추출할 수 있음을 확인하였다. 하나의 임베딩을 입력하는 것에 비해 여러 개의 임베딩을 입력하는 것이 성능 향상에 도움이 됨을 확인했다. 형태소 단위와 형태 분석 정보 또한 정확한 정답을 찾는 것에 도움이 되었다. 향후 연구로는 출력 층의 기능을 개선하여 동일한 관계어의 입력에도 다양한 정답을 추출할 수 있도록 연구를 진행할 예정이다. 또한 관계어의 영역을 넓힘으

로써 나아가 육하원칙(5W1H)에 해당하는 단어와 같이 초록의 전체적인 이벤트를 추출할 수 있도록 연구를 진행할 예정이다.

## 감사의 글

본 연구는 2019년도 한국과학기술정보연구원(KISTI) 주요사업 과제로 수행되었음.

## 참고문헌

- [1] S.Kmar, "A survey of Deep Learning Methods for Relation Extraction", arXiv:1705.03645v1, 2017
- [2] Rupesh Kumar Srivastava, Klaus Greff, and Jurgen Schmidhuber, "Highway network", CoRR, abs/1505.00387, <https://arxiv.org/abs/1505.00387>, 2015.
- [3] D.Zeng, K.Liu, S.Lai, G.Zhou, and J.Zhao, "relation Classification via Convolutional Deep Neural Network", In Colling, pp.2335-23344, 2014
- [4] M.Miwa and M.Bansal, "End-to\_end Relation Extraction using LSTMs on Sequences and Tree Structures", arXiv:1601.00770v3, 2016
- [5] X.Zeng, D.Zeng, S.He, K.Liu and J.Zhao, "Extracting Relational Facts by an End-to-End Neural Model with Copy Mechanism", Proceedings of the 56<sup>th</sup> Annual Meeting of the Association for Computational Linguistics(ACL), pp.506-514, 2018
- [6] Pranav Rajpukar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. "Squad: 100,000+ questions for machine comprehension of text", In Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2016
- [7] A.Vaswani, N.Shazeer, N.Parmar, J.Uszkoreit, L.Jones, A.N.Gomez, L.Kaiser, I.Polosukhin, "Attention Is All You Need", Nernal Information Processing Systems(NIPS), pp.5998-6008, 2017.
- [8] Francois Chollet, "Xception: Deep learning with depthwise separable convolutions", CoRR abs/1610.02357, <http://arxiv.org/abs/1610.02357>, 2016
- [9] Lei Jimmy Ba, Ryan Kiros, and Geoffrey E.Hinton, "Layer Normalization", CoRR, abs/1607.06450, <https://arxiv.org/abs/1607.06450>, 2016
- [10] Caiming Xiong, Victor Zhong, and Richard Socher, "Dynamic coattention networks for question answering", CoRR, abs/1611.01604, <https://arxiv.org/abs/1611.01604>, 2016
- [11] Min Joon Seo, Aniruddha Kembavi, Ali Farhadi, and Hannaneh Hajishirzi, "Bidirectional attention flow for machine comprehension", CoRR abs/1611.01603, 2016
- [12] Sooji Yoon, Sangha Nam, Eun-kuyung Kim, Key-Sun Choi, "Relation Extraction Model for Noisy Data Handling on Distant Supervision Data based on

Reinforcement Learning” , The 30<sup>th</sup> Annual  
Conference on Human & Cognitive Language  
Technology, pp. 55-60, 2018

[13] Espresso:

<http://air.changwon.ac.kr/~airdemo/Espresso/>

(downloaded 2019, Aug. 16)