

문서 수준 관계 추출을 위한 개체 중심 문장 추출 및 Relation-Context Co-attention 방법

박성식^o, 김학수
강원대학교, 건국대학교 인공지능학과
a163912@gmail.com, nlpdrkim@konkuk.ac.kr

Entity-oriented Sentence Extraction and Relation-Context Co-attention for Document-level Relation Extraction

SeongSik Park^o, HarkSoo Kim
Kangwon University, Department of Artificial Intelligence, Konkuk University

요 약

관계 추출은 주어진 문장이나 문서에 존재하는 개체들 간의 의미적 관계를 찾아내는 작업을 말한다. 최근 문서 수준 관계 추출 말뭉치인 DocRED가 공개되면서 문서 수준 관계 추출에 대한 연구가 활발히 진행되고 있다. 또한 사전 학습된 Masked Language Model(MLM)이 자연어처리 분야 전체에 영향력을 보이면서 관계 추출에서도 MLM을 사용하는 연구가 진행되고 있다. 그러나 문서 수준의 관계 추출은 문서의 단위가 길기 때문에 Self-attention을 기반으로 하는 MLM을 사용하면 모델의 계산량이 증가하는 문제가 있다. 본 논문은 이 점을 보완하기 위해 관계 추출에 필요한 문장을 선별하는 간단한 전처리 방법을 제안한다. 또한 문서의 길이에 상관없이 관계 추출에 필요한 어휘 정보를 자동으로 습득 할 수 있는 Relation-Context Co-attention 방법을 제안한다. 제안 모델은 DocRED 말뭉치에서 Dev F1 62.01%, Test F1 59.90%로 높은 성능을 보였다.

주제어: 관계 추출, 문서 수준 관계 추출, 상호 주의집중 방법

1. 서론

관계 추출은 주어진 문장이나 문서에 존재하는 개체(Entity)들 중 서로 연관성을 갖는 개체를 찾아내고 그 관계명을 분별하는 작업을 말한다. 최근 연구는 문서 수준에서 관계를 추출하는 방법에 대한 연구가 중점적으로 진행되고 있다[1-6]. 표 1은 대표적인 문장 수준 관계 추출 말뭉치인 TACRED(TAC Relation Extraction Dataset)[7]와 문서 수준 관계 추출 말뭉치 DocRED(Document-level Relation Extraction Dataset)[1]를 비교한 것이다.

표 1. TACRED와 DocRED 말뭉치의 통계적 비교

	TACRED	DocRED
추출 대상 (max/avg)	문장 (1문장)	문서 (25/7.9 문장)
개체 수	2	41/19.5
트리플 수	1	70/12.5
관계 타입	42	96

표 1에서 보이는 바와 같이 문서 수준의 관계 추출은 문장 수준 관계 추출보다 더 많은 부분을 고려해야 한다. 문서에는 문장보다 월등히 많은 수의 개체들이 존재

하며 이에 따른 복잡한 형태의 관계들이 존재한다. 또한 문서 수준 관계 추출은 한 문장 내에서 표현된 관계뿐만 아니라 여러 문장에 걸쳐 표현된 관계나 멀리 떨어진 문장 간에 표현된 관계에 대한 추출 방법이 필요하다. 본 논문은 문서 수준 관계 추출을 수행하기 위해 개체를 중심으로 문장을 추출하는 방법을 제안한다. 문서 전체의 시퀀스(Sequence) 길이가 길기 때문에 간단한 방법으로 관계 추출에 필요한 문장만을 선별하여 데이터의 복잡성과 모델의 계산량을 줄인다. 또한 문서에서 각 관계 라벨(Label)을 설명하는 어휘 정보를 활용하기 위한 관계-문맥 간 상호 주의집중(Relation-Context Co-attention) 방법을 제안한다. 본 논문은 다음의 순서로 진행된다. 2장의 관련 연구에서 다양한 관계 추출 모델에 대한 소개를 하고 3장에서 제안 모델에 대한 상세한 설명을 진행한다. 4장에서 실험을 통해 제안 모델과 다른 모델들 간의 비교를 수행하고 5장에서 결론을 짓는다.

2. 관련 연구

[1]은 DocRED 말뭉치를 공개하면서 기존에 문장 단위의 관계 추출에서 높은 성능을 보이던 CNN[8], BiLSTM[9], Context-Aware[10] 모델을 문서 수준 관계 추출에 적용한 실험을 진행했다. 결과적으로 모든 모델이 큰 폭으로 성능이 하락하며 문서 수준의 관계 추출을

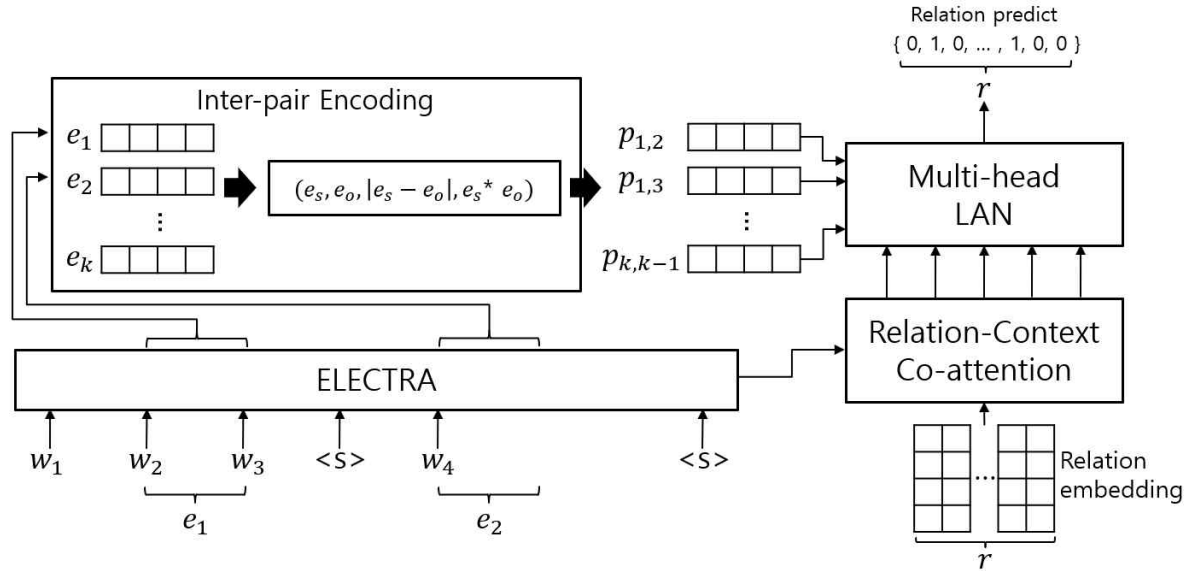


그림 1. 제안 모델 전체 구조도

위해서는 문장 수준과는 다른 접근 방법이 필요하다는 점을 보였다. 최근에는 BERT[11]처럼 사전 학습된 MLM(Masked Language Model)이 자연어처리 분야 전체적으로 큰 영향을 끼치면서 관계 추출에서도 MLM을 기반으로 한 문서 수준 관계 추출 연구가 주로 진행되고 있다. [3]은 관계 추출에 필요한 문장을 추출하는 문장 추출 모델을 관계 추출 모델과 함께 사용하는 BERT Two-step 모델을 제안했다. [4]는 문서 수준 관계 추출에서는 문서 내 개체들의 상호참조 정보가 중요하게 작용한다는 점을 강조하며 BERT 기반의 상호참조 모델을 제안하고 이를 활용한 관계 추출 연구를 진행했다. LSR(Latent Structure Refinement) 모델[5]은 정적인 그래프를 활용하던 GCN(Graph Convolutional Network)[12] 기반 관계 추출 모델을 개선하여 문서의 구조적 정보를 동적으로 학습하는 방법을 제안했다. [6]은 문장 수준 인코딩(Encoding)을 진행한 후 문서 수준 인코딩을 진행하는 계층적인 방식으로 지역적인 정보와 전역적인 정보를 관계 추출에 활용하는 HIN(Hierarchical Inference Network) 모델을 제안했다. 본 논문은 따로 딥러닝(Deep-learning) 모델을 구현할 필요 없이 간단한 전처리 알고리즘으로 관계 추출에 필요한 문장을 선별하는 방법과 문서에서 관계를 설명하는 어휘 정보를 자동 습득하고 활용하기 위한 Relation-Context Co-attention 방법을 제안한다.

3. 문서 수준 관계 추출 모델

그림 1은 제안 모델의 구조도를 보여준다. 중점적으로 다뤄질 부분은 크게 세 부분이다. 첫 번째는 문서 인코딩으로부터 개체 인코딩을 추출하고 그를 이용해 개체 쌍 인코딩을 수행하는 Inter-pair 인코딩 계층이다. 두 번째는 문맥에서 각 관계 라벨을 설명하는 부분을 자동적으로 강조하는 역할을 하는 Relation-Context Co-attention 계층이다. 세 번째는 Inter-pair 인코딩과 Co-attention의 결과 값을 입력하여 해당 개체 쌍의 관

계를 판별하는 Multi-head LAN(Label Attention Network)[13] 계층이다.

3.1. Inter-pair 인코딩 계층

문서는 ELECTRA[14]를 통해 인코딩이 진행된다. ELECTRA는 사전 학습 시 Masking된 단어를 판별하는 방법으로 자가 지도 학습(Self-supervised Learning)을 진행한 MLM이다. 문서 내 개체의 인코딩 벡터는 각 개체 언급(Mention)에 해당하는 인코딩 벡터들의 평균값을 사용한다. 수식 (1)은 개체 인코딩 과정을 나타낸다.

$$\begin{aligned} enc(e_k) &= \frac{1}{M} \sum_{m_i \in e_k} enc(m_i) \\ enc(m_i) &= \frac{1}{H} \sum_{h_j \in m_i} h_j \end{aligned} \quad (1)$$

(1)에서 $enc(e)$, $enc(m)$ 은 각각 개체 인코딩 벡터와 해당 개체의 언급 인코딩 벡터를 의미한다. M, H 는 각각 개체의 언급 횟수와 언급에 해당하는 단어 수를 의미한다. h_j 는 문서에서 j 번째 단어의 인코딩 벡터를 의미한다. 개체 인코딩 벡터는 모든 개체 쌍 조합에 대한 인코딩에 사용된다. 개체 쌍 인코딩에 사용된 수식은 (2)와 같다.

$$enc(p_{s,o}) = [e_s, e_o, |e_s - e_o|, e_s * e_o] \quad (2)$$

(2)에서 e_s, e_o 는 각각 Subject, Object에 해당하는 개체의 인코딩 벡터이다. 수식은 두 문장 사이의 연관 정보를 구하는데 효과를 증명한 Infer-Sent[15] 모델의 수식을 참고했다. 모든 개체의 조합 쌍 인코딩 벡터 $enc(p)$ 는 위 과정을 통해 계산된다.

3.2. Relation-Context Co-attention 계층

Relation-Context Co-attention 계층에서는 문서의 모

든 단어 인코딩 벡터와 관계 라벨 임베딩 벡터 간의 Attention을 통해 각 단어가 어떤 관계와 연관이 있는지 계산한다. 관계 임베딩 벡터는 임의 초기화하며 모델 학습이 진행되며 그 값이 미세 조정(Fine tune)된다. Attention은 Bilinear attention을 기반으로 수행되며 수식 (3)과 같다.

$$\begin{aligned} score_l &= r_l W_a H^\top + b_l \\ a_l &= softmax(score_l) \\ c_l &= a_l H \end{aligned} \quad (3)$$

(3)에서 r_l, H 는 각각 l 번째 관계 라벨의 임베딩 벡터와 문서의 모든 단어 인코딩 벡터들을 의미하며 W_a 는 Bilinear attention의 가중치 행렬이다. $score$ 는 두 벡터 간의 연관도 점수이며 이에 $softmax$ 를 취해 연관도 확률 분포인 a 를 만든다. 이후 a 는 단어 인코딩 벡터 H 와의 행렬 곱을 통해 r_l 에 관련된 단어의 의미가 강조된 벡터 c_l 계산에 사용된다. Co-attention 계층은 문서의 길이에 상관없이 문서에서 관계를 설명하는 부분의 정보를 습득하는데 도움을 줄 수 있다.

3.3. Multi-head LAN 계층

Multi-head LAN 계층에선 개체 쌍 인코딩 벡터와 Relation-Context Co-attention 벡터 간의 Attention을 수행하여 각 개체 쌍이 갖는 관계를 추출한다. Multi-head LAN은 Multi-head Attention[16]의 수식을 기반으로 하며 수식은 (4)와 같다.

$$\begin{aligned} P &= FNN([p_{1,2}, \dots, p_{k,k-1}]) \\ C &= FNN([c_1, c_2, \dots, c_l]) \\ score_i &= softmax\left(\frac{P_i C_i}{\sqrt{d}}\right) (0 \leq i \leq n) \\ O &= \frac{1}{n} \sum_i score_i \end{aligned} \quad (4)$$

(4)에서 P, C 는 개체 쌍 인코딩 벡터들과 Co-attention 벡터들에 FNN(Feed-forward Neural Network)를 적용해 추상화한 값이다. n 은 사용할 head의 개수를 의미하며 i 는 head의 인덱스(index)를 의미한다. 개체 쌍의 관계 점수는 두 벡터의 내적을 기반으로 계산되며 최종 관계 예측 점수 O 는 각 head의 점수를 평균 낸 값이다. 모델의 학습을 위한 손실 함수는 다중 분류(Multi Labeling)를 위해 Binary Cross-Entropy를 사용하며 관계 예측 점수 O 와 실제 정답 사이의 손실 값이 계산된다.

3.4. 개체 중심 문장 추출

문서 수준의 관계 추출에선 문서 자체의 길이가 매우 길기 때문에 대부분의 MLM에서 사용하는 Self-Attention의 계산량이 폭발적으로 증가한다. 또한 관계 추출을 할 때 꼭 필요한 문장을 제외한 나머지 문장은 모델에 잡음

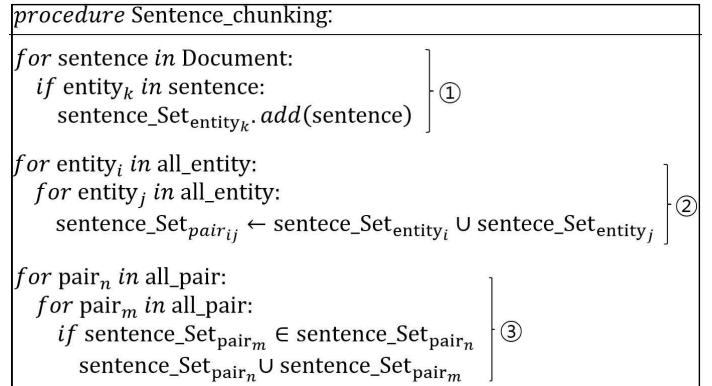


그림 2 개체 중심 문장 추출 의사 코드

(Noise)로 작용 할 수 있는 가능성이 있다. 본 논문에서는 이를 보완하기 위해 관계 추출에 필요한 문장을 간단하게 추출하는 방법을 제안한다. 추출 전처리 과정은 그림 2의 의사 코드에 나타나 있다. 전처리 과정에서 우선 개체가 존재하는 문장을 찾아내어 각 개체마다의 문장 집합을 만들어낸다(①). 그 다음 모든 개체들로 조합이 가능한 개체 쌍에 대한 문장 집합을 만든다. 개체 쌍 문장 집합은 두 개체의 문장 집합을 합하여 만들어진(②). 그 후 각각의 개체 쌍 문장 집합을 비교해 보면서 포함 관계에 해당하는 문장 집합이 존재하면 두 문장 집합 중 더 큰 문장 집합에 다른 하나를 합한다(③). 위 과정을 진행하고 나면 기존 문서보다 짧고 관계 추출에 필요한 문장을 충분히 포함하고 있는 새로운 문장 집합이 만들어진다.

4. 실험

4.1. 실험 환경

표 2. 개체 중심 문장 추출 전후 비교

이름	전처리 전		전처리 후	
	값		값	
	최대	평균	최대	평균
단어 수	511	197.7	463	85.7
문장 수	25	7.9	15	3.5
개체 수	41	19.5	32	8.5
트리플 수	70	12.5	58	3.0

표 3. 학습에 사용된 Hyper-parameter 정보

이름	값
label embedding size	256
bilinear weight size	512
multi-head LAN heads	8
multi-head LAN head size	128
drop-out	0.1
learning rate	0.0005

본 논문에서는 제안한 모델의 학습 및 평가를 위해 DocRED 말뭉치를 사용한다. 학습용 문서는 3,053개가 존재하며 개발용, 평가용 문서는 둘 다 1,000개이다. 말뭉치는 3.4절에서 설명한 전처리를 수행했으며, 그 결과 학습용 문장 집합은 27,739개, 개발용 문장 집합은 9,159개, 평가용 문장 집합은 8,649개를 습득했다. 전처리 전후 말뭉치의 상세 정보는 표 2와 같다. 제안 모델에서 사용한 사전 학습된 ELECTRA는 base 모델을 사용했다. 학습 시 ELECTRA의 파라미터 미세 조정도 함께 진행했다. 제안 모델에서 사용한 Hyper-parameter 정보는 표 3과 같다.

4.2. 실험 결과

표 4. 뒷받침 문장 추출 성능

Recall	Precision	F1
93.0	55.9	69.8

표 4는 문장 추출 전처리를 진행했을 때 관계가 존재하는 개체 쌍에 대하여 실제 관계 추출에 뒷받침(Evidence)이 되는 문장이 추출됐는지에 대한 성능표이다. 결과적으로 Precision이 상대적으로 낮지만 Recall이 93%로 높은 성능을 보임을 알 수 있다. 이는 전처리 후 얻어낸 문장 집합들이 관계 추출에 필요한 문장들을 충분히 포함하고 있다는 것을 의미한다. 표 2의 전처리 결과와 연관 지으면 문장 추출 전처리 과정이 문서의 길이를 크게 줄이면서 관계 추출에 적합한 문장을 충분히 선별 할 수 있다는 점을 알 수 있다.

표 5. 제안 모델 자체 성능 비교표

Model	Dev F1
LSTM-Bilinear	52.76
+ Chunking	54.80
ELECTRA-Bilinear	60.12
ELECTRA-Inter-pair + LAN	61.66
+ Co-attention	62.01

표 5는 제안 방법의 적용 유무에 따른 모델 비교 성능을 보여준다. RNN-Bilinear는 [1]에서 실험한 LSTM 기반 모델을 재구현한 모델이다. Chunking은 문장 추출 전처리 후 데이터를 학습에 사용했을 때의 성능이다. 전처리만 수행했음에도 성능이 약 2%p 가량 상승함을 알 수 있다. 이는 관계 추출에 필요한 문장이 충분히 포함된 문장 집합을 학습에 이용하면서 모델의 계산량, 복잡도가 감소했기 때문이라고 판단된다. 이후 실험은 추출된 문장 집합으로만 학습을 진행했다. ELECTRA-Bilinear는 문서 인코딩을 ELECTRA로 대체했을 경우의 성능이다. 사전 학습된 ELECTRA에서 풍부한 언어 자질을 습득 할 수 있기 때문에 5.2%p의 큰 성능 향상이 있다고 판단된다. ELECTRA-Inter-pair + LAN은 개체 쌍 인코딩에 Bilinear 대신 Inter-pair를 사용하고 multi-head LAN을 적용한 성능이다. 이전과 비교 했을 때, 약 1.5%p 정도의 성능

향상을 보인다. 이는 Bilinear보다 Inter-pair 인코딩이 개체 쌍 간 관계 정보를 추출하는데 적합하기 때문이라고 생각된다. 마지막으로 Relation-Context Co-attention을 적용 했을 때, F1 62.01%로 가장 높은 성능을 보임을 알 수 있다. 이는 Co-attention 계층에서 관계 추출에 필요한 어휘 정보를 가져오면서 Multi-head LAN에 더 풍부한 정보가 반영됐기 때문이라고 판단된다.

	a	daughter	of	the	industrial	charles
child	5.27E-06	8.84E-01	1.43E-04	3.33E-06	1.02E-05	1.56E-04
composer	0.00012	0.00024	1.43E-05	1.03E-04	4.93E-03	0.00071
conflict	0.00073	1.43E-03	1.57E-05	1.31E-04	1.15E-03	2.06E-03

그림 3. Co-attention 분포 예시

Co-attention 계층에서의 Attention 분포도는 그림 3에 나타나 있다. 결과적으로 관계 추출 학습이 진행되며 “child”에 해당하는 관계를 추출 할 때 문서에서 “daughter” 어휘에 자동으로 가중되는 것을 확인 할 수 있다.

표 6. DocRED 실험 모델 간 비교 성능

Model	Dev F1	Test F1
BiLSTM[1]	51.72	49.82
GREG[2]	52.88	-
BERT-Two_Step[3]	54.42	53.92
HIN-BERT[6]	56.31	55.60
BERT-LSR[5]	59.00	59.05
Coref-BERT_Base[4]	57.51	56.96
Coref-RoBERTa_Large[4]	59.93	59.91
Proposed model	62.01	59.90

표 6은 DocRED를 실험한 기존 연구들과의 성능 비교를 보여준다. 실험 결과 개발용 말뭉치에 대해서 제안 모델이 F1 62.01%로 가장 높은 성능을 보임을 알 수 있다. 그러나 평가용 말뭉치에 대해서 Coref-RoBERTa_Large과 비슷하지만 약간 낮은 성능을 보이는데 이는 MLM의 크기 차이 때문인 것으로 생각된다. 제안 모델과 같은 크기의 MLM을 사용한 Coref-BERT_Base와 비교하면 제안 모델이 더 높은 성능을 보인다. 제안 모델도 MLM Large를 사용하면 현재보다 높은 성능을 보일 것으로 예상된다.

5. 결론

본 논문은 문서 수준의 관계 추출을 위해 개체 중심으로 문장을 추출하는 전처리 방법을 제안했으며 기존 방법들과 다른 방식으로 개체 쌍 인코딩을 제안했다. 또한 문서로부터 관계 추출에 필요한 어휘 정보를 습득하기 위한 Relation-Context Co-attention 방법을 제안했다. 결과적으로 DocRED를 사용한 실험에서 Dev F1 62.01%, Test F1 59.90%로 높은 성능을 보였다. 향후 연구로 Curriculum Learning[17]을 사용한 문서 수준 관계 추출 연구를 진행할 예정이다.

감사의 글

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2020-0-00368, 뉴럴-심볼릭(neural-symbolic) 모델의 지식 학습 및 추론 기술 개발). 또한 이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2013-0-00131, (엑소브레인-총괄/1세부) 휴먼 지식증강 서비스를 위한 지능진화형 WiseQA 플랫폼 기술 개발)

참고문헌

- [1] Y. Yao, D. Ye, P. Li, X. Han, Y. Lin, Z. Liu, Z. Liu, L. Huang, J. Zhou and M. Sun, DocRED: A Large-Scale Document-Level Relation Extraction Dataset, Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp.764-777, 2019.07.
- [2] K. Kim, Y. Hur, G. Kim and H. Lim, GREG: A Global Level Relation Extraction with Knowledge Graph Embedding, Applied Sciences, Vol.10(1181), pp.1-12, 2020.02. (doi:10.3390/app10031181)
- [3] H. Wang, C. Focke, R. Sylvester, N. Mishra and W. Wang, Fine-tune Bert for DocRED with Two-step Process, arXiv:1909.11898v1, 2019.09.
- [4] D. Ye, Y. Lin, J. Du, Z. Liu, M. Sun and Z. Liu, Coreferential Reasoning Learning for Language Representation, arXiv:2004.06870v1, 2020.04.
- [5] G. Nan, Z. Guo, I. Sekulic and W. Lu, Reasoning with Latent Structure Refinement for Document-Level Relation Extraction, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp.1546-1557, 2020.06.
- [6] H. Tang, Y. Cao, Z. Zhang, J. Cao, F. Fang, S. Wang and P. Yin, HIN: Hierarchical Inference Network for Document-Level Relation Extraction, Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2020), pp.197-209, 2020.05.
- [7] Y. Zhang, V. Zhong, D. Chen, G. Angeli and C. D. Manning, Position-aware Attention and Supervised Data Improve Slot Filling, Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp.35-45, 2017.09.
- [8] D. Zeng, K. Liu, S. Lai, G. Zhou and J. Zhao, Relation classification via convolutional deep neural network, Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers (COLING 2014), pp.2335-2344, 2014.08.
- [9] R. Cai, X. Zhang and H. Wang, Bidirectional Recurrent Convolutional Neural Network for Relation Classification, Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pp.756-765, 2016.08.
- [10] D. Sorokin and I. Gurevych, Context-Aware Representations for Knowledge Base Relation Extraction, Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp.1784-1789, 2017.09.
- [11] J. Devlin, M.W. Chang, K. Lee and K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Proceedings of North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2019), pp.4171-4186, 2019.06.
- [12] T.N. Kipf and M. Welling, Semisupervised classification with graph convolutional networks. Proceedings of 5th International Conference on Learning Representations, pp.1-14, 2017.04.
- [13] H. Kim and H. Kim, Integrated Model for Morphological Analysis and Named Entity Recognition Based on Label Attention Networks in Korean, Applied Sciences (SI: Machine Learning and Natural Language Processing), Vol.10(3740), pp.1-10, 2020.05. (DOI: 10.3390/app10113740)
- [14] K. Clark, M.T. Luong, Q.V. Le and C.D. Manning, ELECTRA: Pre-training text encoders as discriminators rather than generators, Proceedings of 8th International Conference on Learning Representations, pp.1-14, 2020.04.
- [15] A. Conneau, D. Kiela, H. Schwenk, L. Barrault, A. Bordes, Supervised Learning of Universal Sentence Representations from Natural Language Inference Data, Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp.670-680, 2017.09.
- [16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser and I. Polosukhin, Attention Is All You Need, Proceedings of Neural Information Processing Systems (NIPS), pp.5998-6008, 2017.12.
- [17] B. Xu, L. Zhang, Z. Mao, Q. Wang, H. Xie and Y. Zhang, Curriculum Learning for Natural Language Understanding, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp.6095-6104, 2020.07.