

Easy-First Deep Biaffine Attention을 이용한

한국어 의존 파싱

홍승연⁰¹, 나승훈², 신종훈³, 김영길⁴

¹²전북대학교, ³⁴한국전자통신연구원

hongsy034@hanmail.net, nash@jbnu.ac.kr, jhshin82@etri.re.kr, kimyk@etri.re.kr

Easy-First Deep Biaffine Attention for Korean Dependency Parsing

Seung-Yean Hong⁰¹, Seung-Hoon Na², Jong-Hoon Shin³, Young-kil Kim⁴

¹²Jeonbuk National University, ³⁴ETRI

요약

기존의 그래프 기반 파서는 문장의 가능한 모든 트리를 찾고 가장 높은 점수를 갖는 트리를 취하는 방식이다. 하지만 점수를 계산하는데 있어서 노드 정보만을 사용하기 때문에 트리 구조의 특성을 반영하지 못하는 단점이 있다. 이를 위해 본 논문에서는 사전 학습된 모델에서 단어 간의 점수를 얻어 높은 점수를 가지는 단어를 미리 결정하고 결정된 의존성을 통해 부분 트리 만든다. 만들어진 부분 트리 정보를 사용하여 트리 구조의 특성을 반영할 수 있도록 하였다.

주제어: 의존 파싱, 딥러닝, 그래프 기반 파싱

1. 서론

최근 구문 분석은 자연어 처리에서 활발히 연구되고 있는 분야이다. 구문 분석은 문장을 그것을 이루고 있는 구성 성분으로 분해하고 구성 성분과의 관계를 분석하여 문장 구조를 파악하는 것을 말하며 크게 구 구조 파싱과 의존 파싱으로 나눌 수 있다. 그 중 의존 파싱은 단어 간의 의존 관계를 파악하는 연구로 자연어 처리에서 주로 연구되고 있으며 최근에는 딥러닝 모델들을 이용한 연구가 주로 이루어지고 있다. 어텐션 메커니즘(Attention Mechanism)[1]은 읽어 들인 입력에 대해 참조하고자 하는 위치에 가중치를 더 부과하는 메커니즘인데 이를 이용한 모델은 대표적으로 [2]가 있다. 이 모델은 Biaffine Attention을 사용하여 의존 파싱에서 좋은 성능을 보이고 있는 모델이다.

본 논문에서는 Deep Biaffine Attention 모델에 의존성이 있다고 판단되는 단어간의 관계를 미리 결정하여 트리 구조를 반영하고자 하였으며 이를 적용하여 기존 성능에 비해 향상된 성능을 얻었다.

2. 관련 연구

의존 파싱은 전이 기반 방식과 그래프 기반 방식으로 나눌 수 있다. 전이 기반 방식으로는 스택과 버퍼내의 모든 정보를 활용하기 위해 Stack LSTM을 이용한 모델인 [3]가 있고 [4]에서는 이를 한국어 의존 파싱에 적용하기 위해 형태소 표상을 합성하여 단어 표상을 얻어내는 방법을 제안 하였다. [5]는 어텐션을 사용하는 모델인 포인터 네트워크에 내부 스택을 만들어 포인터 네트워크가 출력하는 위치 정보를 이용하여 의존 트리를 하향식 방식으로 만들어가는 모델이다. 그래프 기반 방식으로는 [2]은 Biaffine Attention 사용하여 점수를 계산하여 의

존 트리를 구성하는 모델이다. [6]은 [2]모델을 한국어에 적용하기 위하여 형태소 표상으로부터 단어 표상을 얻어 한국어에 적용한 모델이다. [7]은 개선된 Biaffine Attention을 활용하여 한국어에 적용한 모델이다. [8]는 포인터 네트워크를 활용하여 점수를 계산하는 모델이다.

BERT[9], ELMo[10]는 문맥에 따라 다른 단어 벡터를 얻을 수 있어서 다양한 태스크에 적용되어 높은 성능을 얻고 있다. BERT, ELMo를 적용하여 [11], [12], [13]에서는 기존의 성능을 뛰어 넘는 성능을 보여주고 있다.

[14]은 본 논문의 아이디어의 기반이 되는 논문으로 결정하기 쉬운 의존성부터 결정하면서 트리를 구성해나가는 방식의 파싱을 제안하였다.

3. Easy-First parsing

사람들은 일반적으로 문장을 읽을 때 왼쪽에서 오른쪽으로 순차적으로 읽지만 문장 구조를 파악하는 작업에서는 확실한 구조부터 결정하면서 문장 구조를 파악한다. [14]에서는 이러한 점을 이용하여 확실한 정보부터 문장 구조를 결정짓는 easy-first parsing을 제안하였다. 본 논문에서도 확실한 정보를 미리 결정하고 결정된 정보를 이용하여 부분 트리를 구성해 파싱하는 방식을 제안한다.

3.1. Model

그림 1은 제안한 파싱 모델을 나타내고 Biaffine Attention 모델[2]을 사용하고 있다. Biaffine Attention 모델에서 단어 표상을 얻기 위해 형태소 단위의 Bi-LSTM 기반 합성 방법을 사용하였으며 학습된 Glove 임베딩 벡터를 형태소 임베딩으로 초기화하여 사용하였다. 기존의 Biaffine Attention 모델은 단어 표상을 Bi-LSTM에 적용하여 인코딩 한 후 MLP 적용하여 지베소 표상과 의존소 표

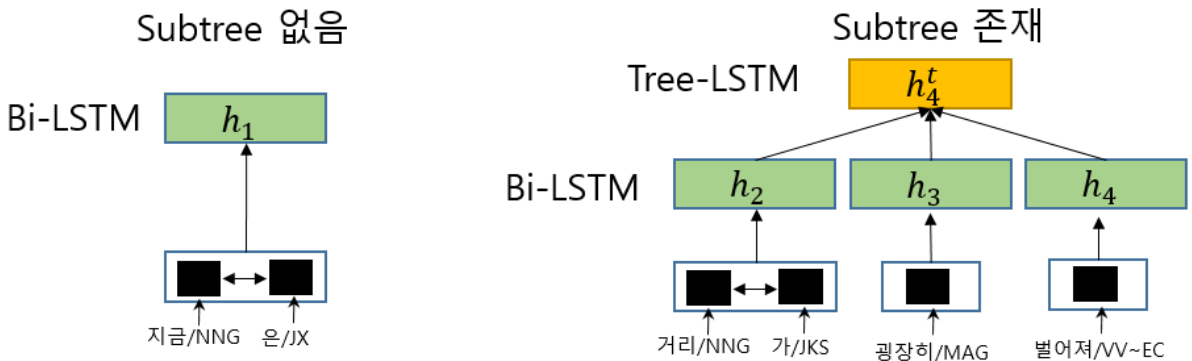
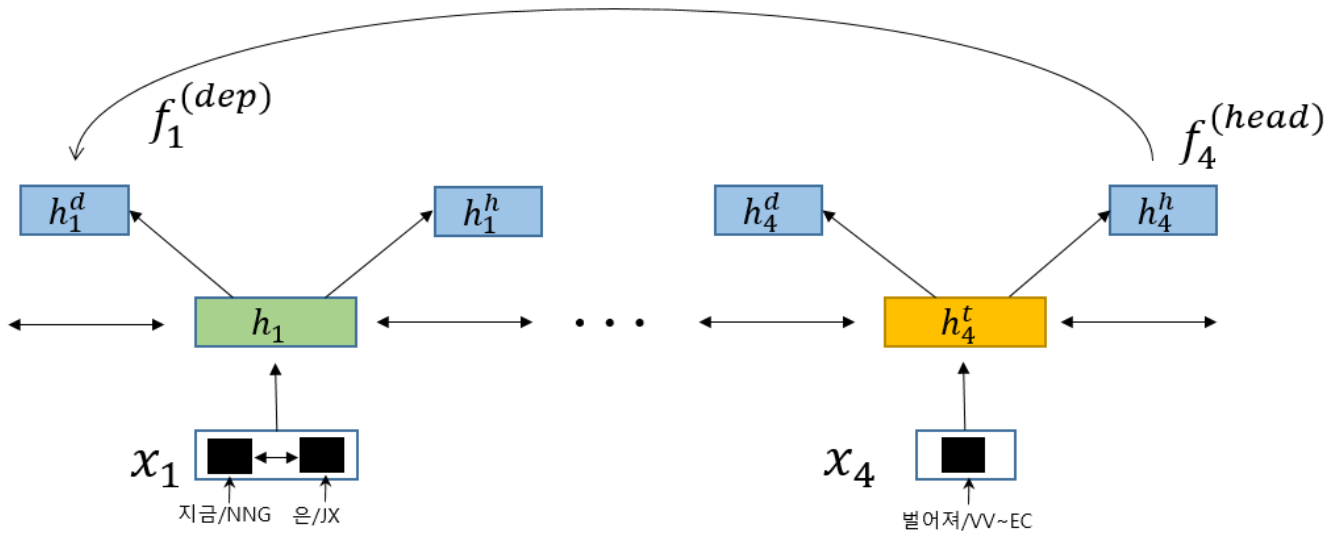


그림 1 Easy-First Biaffine Attention을 이용한 한국어 의존 파싱 모델

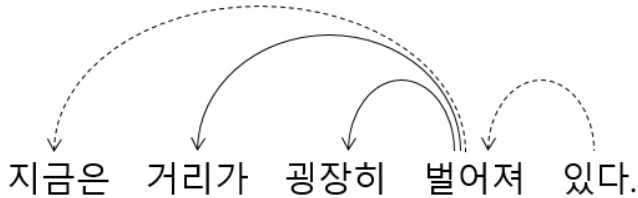


그림 2 의존성이 결정된 의존 트리 구조

상을 언어 점수를 계산하는 형태를 가지고 있는데 이때 지배소 표상과 의존소 표상은 노드 정보만 가지고 있고 트리 특성의 정보를 가지고 있지 않다. 본 연구에서는 트리 정보를 제공하기 위하여 확실한 의존성을 지니는 단어간의 정보를 미리 결정하고 결정된 정보를 이용하여 부분 트리 만들었다. 확실한 정보를 미리 결정하기 위해 선형 학습한 Biaffine Attention 모델의 점수 정보를 사용하여 확실한 의존성을 결정하고 부분적으로 결정된 의존성을 이용하여 각 단어마다 부분 트리를 만들어 트리 정보를 반영하고자 하였다. 그림 2에서와 같이 “지금은 거리가 굉장히 벌어져 있다.” 문장이 주어졌을 때 점수 정보를 이용하여 확실하다고 판단되는 의존성을 결정하였다. 실선은 의존성을 결정 지은 것이고 점선은 예측해야 하는 의존성을 나타낸다. 그림 1은 의존성이 결정되었을 때 단어의 부분 트리가 존재하는 경우와 존재하지 않는 경우 표상을 얻는 과정을 보여준다. “지금은”의 단어 표상은 의존성 결정되었을 때 자식이 없기 때문에

부분 트리를 갖지 않는다. 부분 트리를 갖지 않는 경우 Bi-LSTM을 통해 표상을 얻는다. “벌어져”의 단어 표상은 확실한 의존성이 결정되었을 때 “거리가”, “굉장히” 두 단어의 표상을 자식으로 갖는다. 이 정보를 이용하여 부분 트리를 구성하여 Tree-LSTM[15]를 통해 표상을 얻는다. 본 논문에서는 Tree구조를 반영하기 위하여 Tree-LSTM을 채택하였다. 부분 트리가 존재하지 않은 경우는 수식 (1)과 같이 BiLSTM을 통해 얻은 표상을 사용하였고 부분 트리가 존재하는 경우는 수식 (2)와 같이 수식(1)에서 얻은 표상 정보와 자식 표상 정보를 이용하여 Tree-LSTM 통해 얻은 표상 정보를 사용하였다.

$$h_1, \dots, h_n = BiLSTM(x_1, \dots, x_n) \quad (1)$$

$$h_i^t = TreeLSTM(h_i, child(h_i)) \quad (2)$$

이후 얻어진 표상을 MLP 적용하여 지배소 표상과 의존소 표상을 통해 어텐션을 수행한다. 어텐션 단계에서 Biaffine Attention을 적용하여 점수를 얻고 이를 이용하여 가장 높은 점수를 갖는 의존 트리를 결정한다.

3.2 학습 및 평가

학습을 위해 학습 데이터는 무작위로 의존성을 결정할 데이터와 의존성을 결정하지 않은 두 데이터를 학습 데이터로 하여 실험하였고 무작위로 의존성을 결정할 때에는 문장에서 50%를 무작위로 의존성을 결정하였다. 두

학습 데이터는 학습 시 동일 파라미터를 사용하여 학습이 진행된다.

평가 시에는 사전 학습된 모델에서 얻은 평가 데이터의 점수 정보를 이용하여 확실한 의존성을 결정 후 평가를 진행하였다. 선행 학습된 모델에 점수 이외에 확실한 의존성 여부 판단을 위해 선행 학습된 파싱 모델에 MC Dropout [16]을 적용하여 모델이 불확실성을 가지게 되어 샘플링을 통해 확신도를 측정하였다. 이렇게 얻은 확신도는 확실한 의존성 결정 여부를 판단하는 지표로 사용하였다. 성능 산출 시에는 결정되지 않은 의존성의 모델 예측 결과와 결정된 의존성과 결정되지 않은 의존성의 모델 예측 결과를 모두 합한 방법으로 성능을 산출하였다. 결정된 의존성은 모델에 예측 결과를 사용하지 않고 사전 학습된 모델의 결과를 이용하였다.

4. 실험 및 결과

4.1 실험

실험 데이터 셋으로 SPMLR' 2014의 공개 말뭉치인 [17]의 데이터를 사용하였다. 확실한 의존성 결정은 문장 주어였을 때 점수, 확신도를 보고 높은 순서대로 결정하여 50%만을 결정하였다. 사전 학습한 모델에서 얻은 점수와 확신도가 확실한 의존성을 결정하기에 적절한지 파악하기 위해 점수, 확신도를 높은 순으로 나열하여 높은 순서로 결정된 모델과 낮은 순서로 결정된 모델의 두 결과를 비교하였다. 50%의 점수를 활용한 모델은 Score50%, 50%의 확신도를 활용한 모델은 Confidence50%로 표현하였다.

4.2 실험 결과

표 1 점수, 확신도에 따른 모델 성능

의존 파싱 모델	UAS	LAS
Easy First-Score50%	86.38%	85.86%
Easy First-Score50%(역)	92.63%	92.10%
Easy First-Confidence50%	86.37%	85.88%
Easy First-Confidence50%(역)	97.91%	97.28%

표 1은 점수, 확신도를 높은 순으로 나열한 후 높은 순으로 50%, 낮은 순(역)으로 50%를 선택하여 의존성을 결정 후 얻은 모델의 결과이다. 모델의 성능 측정은 의존성이 결정되지 않은 경우만을 측정하였다. 모델의 점수나 확신도가 낮은 순으로 선택한 경우 모델이 확실한 의존성들을 찾는 것이기 때문에 높은 성능이 예상되고 높은 순으로 선택한 경우에는 그 반대로 낮은 성능이 예상된다. 모델의 성능은 예상과 같이 낮은 순으로 선택한 결과의 성능이 높게 나오고 있고 높은 순으로 선택한 결과가 높게 나오고 있다. 이를 통해 점수, 확신도가 높은 경우에는 성능이 높은 의존성이 선택되기 때문에 의존성 결정을 위해 사용하기 적절하다는 판단을 할 수 있다.

표 2 결정되지 않은 의존성 모델 성능(의존성 수 14755)

의존 파싱 모델	UAS	LAS
ML Biaffine-Score50%	85.87%	85.29%
Easy First-Score50%	86.38%	85.86%
ML Biaffine-Confidence50%	85.90%	85.39%
Easy First-Confidence50%	86.37%	85.88%

결정되지 않은 의존성만을 성능 측정한 결과이다. 제안 모델이 기존 모델에 비해 0.5%가량 높은 성능을 보이고 있다.

표 3 모델 성능(의존성 전체 28366)

의존 파싱 모델	UAS	LAS
ML Biaffine-Score50%	91.68%	91.08%
Easy First-Score50%	91.94%	91.36%
ML Biaffine-Confidence50%	91.68%	91.07%
Easy First-Confidence50%	91.92%	91.35%

모든 의존성에 대하여 성능 측정한 결과이다. 제안 모델이 기존 모델에 비해 0.3%가량 높은 성능을 보이고 있어 제안 내용이 효과가 있음을 증명하였다.

표 4 기존 모델과 비교

의존 파싱 모델	UAS	LAS
Easy First-Score50%	91.94%	91.36%
Deep Biaffine 나승훈[6]	90.85%	89.31%
Deep Biaffine + dual-민진우[18]	91.07%	N/A
Multi-level Biaffine-홍승연[19]	91.84%	91.29%

위 모델들은 BERT, ELMo를 사용하지 않는 기존의 모델들의 성능에 비해 개선된 성능인 UAS에서 91.94%, LAS에서 91.36%의 성능 향상을 가져옴을 확인할 수 있다.

5. 결론

기존의 그래프 기반이 가지고 있던 트리 정보를 활용하지 못하는 단점을 해결하였고 기존 모델보다 개선된 성능을 얻었다. 앞으로 세종 데이터에서 결과를 측정하고 BERT를 적용하여 더욱 향상된 성능을 얻을 수 있을 것을 기대하고 있다.

감사의 글

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (R7119-16-1001, 지식증강형 실시간 동시통역

원천기술 개발)

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (R7119-16-1001, Core technology development of the real-time simultaneous speech translation based on knowledge enhancement)

참고문헌

- [1] T. Luong, H. Pham, D. Maning. “Effective approaches to attention based neural machine translation”, EMNLP, 2015.
- [2] T. Dozat, C. Manning. “Deep Biaffine Attention for Neural Dependency Parsing”, ICLR, 2017.
- [3] C. Dyer, M. Ballesteros, W. Ling, A. Matthews, N. A. Smith. “Transition-Based Dependency Parsing with stack Long Short Term Memory” ACL, 2015.
- [4] 나승훈, 김강일, 김영길. “Stack LSTM을 이용한 전이 기반 한국어 의존 파싱” KCC, 2016.
- [5] X. MA, Z. Hu, J. Liu. “Stack-Pointer Networks for dependency Parsing” ACL, 2018.
- [6] 나승훈, 이건일, 신중훈, 김강일. “Deep Biaffine Attention을 이용한 한국어 의존 파싱” KCC, 2017.
- [7] 오동석, 우종성, 이병우, 김경선. “한국어 의존 구문 분석을 위한 개선된 Deep Biaffine Attention”, HCLT, 2018.
- [8] 박천음, 이창기. “멀티 태스크 학습 기반 포인터 네트워크를 이용한 한국어 의존 구문 분석” KCC, 2017.
- [9] J. Devlin, M. Chang, K. Lee, K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, arXiv:1810.04805, 2018.
- [10] M. Peters, M. Neumann, M. Iyyer, M. Gardner, “Deep contextualized word representations”, NAACL, 2018.
- [11] 박천음, 김기훈, 임준호, 류지희, 김현기, 이창기. “BERT기반 Deep Biaffine을 이용한 한국어 상호참조해결”, KCC, 2019.
- [12] 홍승연, 신중훈, 김영길, 나승훈. “BERT와 ELMo 문맥화 단어 임베딩을 이용한 한국어 의존파싱”, KCC, 2019.
- [13] 박성식, 오신혁, 김홍진, 김시형, 김학수. “ELMo와 멀티헤드 어텐션을 이용한 한국어 의존 구문 분석”, HCLT, 2018.
- [14] Y. Goldbert, M. Elhadad. “An Efficient Algorithm for Easy-First Non-Directional Dependency Parsing”, ACM, 2010.
- [15] S. Tai, R. Socher, D. Manning. “Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks”, ACL, 2015.
- [16] Y. Gal, Z. Ghahramani. “Dropout as a bayesian approximation: Representing model uncertainty in deep learning”, JMLR, 2016.
- [17] D. Seddah, S. Kubler, R. Tsarfaty. “Introducing the SPMRL 2014 Shared Task on Parsing Morphologically-Rich Languages” SPMRL-SANCL, 2014.
- [18] 민진우, 나승훈, 신중훈, 김영길. “Dual Decomposition을 이용한 뉴럴 그래프 기반 한국어 의존 파싱” KCC, 2018.
- [19] 홍승연, 나승훈, 신중훈, 김영길. “Multi-level Biaffine Attention을 이용한 한국어 의존파싱”, KSC, 2018.