

A Multi-task Self-attention Model Using Pre-trained Language Models on Universal Dependency Annotations

Euhee Kim*

*Professor, Dept. of Computer Science & Engineering, Shinhan University, Gyeonggi-do, Korea

[Abstract]

In this paper, we propose a multi-task model that can simultaneously predict general-purpose tasks such as part-of-speech tagging, lemmatization, and dependency parsing using the UD Korean Kaist v2.3 corpus. The proposed model thus applies the self-attention technique of the BERT model and the graph-based Biaffine attention technique by fine-tuning the multilingual BERT and the two Korean-specific BERTs such as KR-BERT and KoBERT. The performances of the proposed model are compared and analyzed using the multilingual version of BERT and the two Korean-specific BERT language models.

▶ **Key words:** Universal Dependency, Multilingual BERT, KR-BERT, KoBERT, Fine-tuning, Pre-training

[요 약]

본 논문에서는 UD Korean Kaist v2.3 코퍼스를 이용하여 범용 품사 태깅, 표제어추출 그리고 의존 구문분석을 동시에 예측할 수 있는 보편적 다중 작업 모델을 제안하였다. 제안 모델은 사전학습 언어모델인 다국어 BERT (Multilingual BERT)와 한국어 BERT (KR-BERT와 KoBERT)을 대상으로 추가학습 (fine-tuning)을 수행하여 BERT 모델의 자가-집중 (self-attention) 기법과 그래프 기반 Biaffine attention 기법을 적용하여 제안 모델의 성능을 비교 분석하였다.

▶ **주제어:** 보편적 의존구문, 다국어 버트, KR-버트, Ko버트, 추가학습, 사전학습

• First Author: Euhee Kim, Corresponding Author: Euhee Kim
*Euhee Kim (euhkim@shinhan.ac.kr), Dept. of Computer Science & Engineering, Shinhan University
• Received: 2022. 10. 06, Revised: 2022. 10. 21, Accepted: 2022. 10. 27.

I. Introduction

자연어처리 응용에서 품사 태깅 (part-of-speech; POS tagging)은 문장을 이루는 각 단어의 품사를 찾는 문제이다. 표제어추출 (lemmatization)은 단어들이 다른 형태를 가지더라도, 그 뿌리 단어를 찾아가서 단어의 개수를 줄일 수 있는지 판단하는 문제이다. 의존 구문분석 (dependency parsing)은 문장 내 단어 사이의 의존관계 및 레이블을 인식하여 문장의 구조적, 의미적 중의성 문제를 해결하는 문제이다[1].

한국어 기반 자연어처리 응용은 한국어 품사 태깅, 표제어추출과 의존 구문분석에 대한 많은 국내 연구가 있지만 전 세계 언어를 하나의 보편적인 문법 기준으로 동일하게 분석하자는 목표로 한국어를 포함한 다국어 지원하는 자연어처리 응용 태스크에 대한 연구는 상대적으로 적다.

최근 한국어 의존 구문분석 태스크는 한국어 대용량 코퍼스 (corpus)로 사전학습 (pre-trained)한 한국어 기반 BERT 모델 위에 장단기메모리 (LSTM) 순환신경망을 추가하여 자가-집중 (self-attention) 기법을 적용한 모델들을 제안하여 높은 성능을 보여 주었다. 또한, 한국어 기반 BERT 모델을 기반으로 각 단어에 대해서 입력 문장 전체를 고려하여 지배소와 의존 관계를 결정하는 그래프 기반 Biaffine attention 모델도 많이 연구되고 있다. 최근에는 한국어 기반 BERT 모델의 자가-집중을 최대한 활용하기 위해 순환신경망을 대체하는 추가학습 (fine-tuning) 기법을 적용하여 한국어 의존 구문분석의 높은 성능을 보여 주었다[2-9].

본 논문에서는 한국어를 포함한 다국어 지원하는 자연어처리 응용 태스크를 위한 ‘다국어 범용 의존관계 주석체계’ (Universal Dependencies, 이하 UD) 지침을 따르는 프레임워크를 이용하여 범용 한국어 코퍼스를 대상으로 범용 품사 태깅, 표제어추출, 의존 구문분석을 동시에 예측할 수 있는 다중 작업 모델을 제안한다[10].

본 논문의 구성은 다음과 같다. 2장에서는 기존 한국어 의존 구문분석 연구들에 대해 설명하고, 3장에서는 제안 모델을 설명한다. 4장에서는 제안한 모델의 구현 방법에 대해 기술한다. 5장과 6장에서는 실험을 위한 데이터셋과 실험 결과를 기술하고, 7장에서는 결론을 맺는다.

II. Preliminaries

1. Related works

1.1 Korean Universal Dependency Corpus

UD 프레임워크는 현재 114개 언어, 거의 200개 트리뱅크 (treebank) 코퍼스로 구성되어 여러 언어에 적용할 수 있는 형태/통사적 구문 특성을 찾는 것을 목표로 한다. 이 프레임워크는 개별 언어가 가지고 있는 언어 특성을 하나의 통일된 형식으로 변환하여 서로 다른 특성을 지닌 언어에도 공통적으로 적용할 수 있는 품사 주석 태그와 구문 주석 태그를 소개하고 있다[10].

최근에는 UD 가이드라인을 통해 전산 언어학 분야에 전 세계에서 통용될 수 있는 지침을 제공함으로써 실용 언어처리시스템에서 광범위하게 사용되고 있다. 또한 하나의 언어 분석을 위해 개발한 학습 모델을 다른 언어에 적용할 수 있는 가능성을 탐구하는 연구를 촉진하게 되었다.

UD 가이드라인에 맞춰 한국어에도 UD 주석을 적용시킨 한국어 트리뱅크 즉, the Penn Korean Treebank (이하 PKT-UD), Google UD Treebank (이하 GSD), KAIST Treebank (이하 Kaist-UD)와 같은 세 종류의 코퍼스가 공개되어 있다[11].

한국어의 경우 품사 주석 태그와 구문 주석 태그는 ‘21세기 세종계획’에서 설계한 주석체계가 일반적으로 활용되고 있다. 반면에 ‘21세기 세종계획’의 구문 주석 코퍼스는 여러 언어에 적용될 수 있는 범용 주석체계의 역할 하는데 한계가 있어서 두 번째 버전의 UD 지침에 맞춰 본래 구구조 구문 분석으로 주석되었던 세 개의 한국어 트리뱅크를 의존 관계 분석 트리뱅크로 변환하였다[12-14].

UD의 형태 주석 태그 (이하 UPOS)와 표제어추출 (이하 Lemmas), 구문 주석 태그 (이하 DEPREL)를 나타내는 형식인 CoNLL U-Format는 UD에 공개된 한국어 트리뱅크에 ‘21세기 세종계획’ 형태 주석 표지, 카이스트 형태 주석 표지 등이 대응되어 있다[15].

CoNLL U-Format는 모두 10개의 열로 구성된다. Table 1의 첫 번째 열은 문장을 구성하는 각 어절의 순서, 두 번째 열은 해당 어절의 형태, 세 번째 열은 해당 어절의 원형을 나타내고, 네 번째 열에는 해당 어절의 UPOS가 주어진다. 일곱 번째 열에서는 해당 어절의 지배소 (head)를, 여덟 번째 열에서는 해당 어절과 지배소 사이의 의존 관계를 표시하는 태그인 DEPREL를 할당하게 된다.

본 논문에서는 UD의 UPOS, Lemmas와 DEPREL에 논점을 맞추었기 때문에, 첫 번째부터 네 번째 열 그리고 일곱 번째부터 여덟 번째 열의 내용을 고려한다. 다음 Table

1은 예를 들어, ‘철수가 밥을 먹었다.’ 라는 문장을 CoNLL U-Format으로 분류한 표이다.

Table 1. CoNLL U-Format

# sent_id = s1									
# text = 철수가 밥을 먹었다.									
1	철 수 가	철수+가	noun	-	-	2	nsubj	-	-
2	밥 을	밥+을	noun	-	-	3	obj	-	-
3	먹 었 다	먹+었+다	verb	-	-	0	root	-	-
4	.		punct	-	-	-	punct	-	-

본 논문에서는 Chun et al.이 구축한 한국어 트리뱅크 Kaist-UD 코퍼스를 대상으로 CoNLL U-Format의 UPOS와 Lemmas, DEPREL을 동시에 분류 예측할 수 있는 다중 작업 모델을 구현하였다[12].

1.2 Korean based BERT

BERT는 구글에서 공개한 트랜스포머 (Transformer) 기반 사전학습 언어모델이다. BERT는 텍스트를 앞뒤로 확인하여 자연어를 처리하는 마스크 언어모델이다. BERT는 2가지 기법 즉, Masked Language Modelling (이하 MLM)과 Next Sentence prediction (이하 NSP)으로 모델을 사전 학습하는데 그 중 MLM 기법은 문장 내의 단어를 랜덤하게 마스킹하고 마스킹을 한 단어를 예측하는 방법이다. 그리고 NSP 기법은 두 문장이 주어졌을 때 뒤에 따라오는 문장이 앞에 문장 다음에 오기에 적합한지 판별하는 방법이다. 두 가지 학습 기법을 위해 BERT에는 토큰 임베딩, 포지션 임베딩, 세그먼트 임베딩 기법이 추가적으로 사용된다[2].

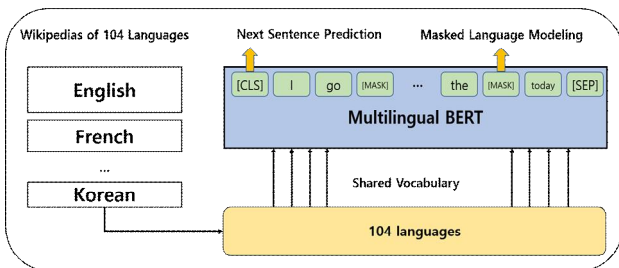


Fig. 1. Multilingual BERT

위의 Fig. 1의 Multilingual BERT (이하 M-BERT)는 구글에서 오픈소스로 공개한 BERT의 다국어 버전으로 BERT 모델을 학습시킨 것과 동일한 방식으로 MLM과

NSP 기법을 사용하면서 104개 언어의 위키피디아 텍스트를 모두 사용하여 학습시킨 모델이다.

M-BERT는 104개 언어에 걸쳐 모두 11만 개의 워드피스 (wordpiece)로 구성되며 다운스트림 태스크에 맞춰 자연어처리 멀티-태스크에서 높은 일반화 성능을 보여주고 있으며, 특히 충분한 양의 레이블을 갖는 학습 데이터의 부족으로 학습을 할 수 없었던 언어에 대해 추론을 수행하는데 높은 성능을 기록하여 주목받고 있다.

또한 한국어로 특화된 BERT의 한국어 버전으로서 SKT-brain에서 공개한 KoBERT가 있으며 서울대학교에서 공개한 KR-BERT 모델 등이 있다[16-17].

다음 Table 2는 BERT 계열의 사전학습 언어 모델을 비교 정리한 표이다.

Table 2. A summary of Korean-specific models

Size	M-BERT	KoBERT	KR-BERT (character)
vocab	119,547	8,002	16,424
data	Wikipedia data(104 languages)	25M(sents), 324M(words)	2.47GB, 20M(sents), 233M(words)

본 논문에서는 사전학습을 진행한 한국어 기반 BERT 모델과 다국어 기반 M-BERT 모델을 한국어 트리뱅크 Kaist-UD 코퍼스를 대상으로 각각 추가학습을 진행하여 CoNLL U-Format의 UPOS, Lemmas와 DEPREL를 동시에 예측할 수 있는 처리 성능이 어느 정도 인지지에 대한 비교 분석하였다.

1.3 Dependency Parsing Model

최근 한국어 의존 구문분석 모델로는 사전학습 한국어 BERT 언어모델 위에 추가적인 순환신경망을 적용하거나 BERT 언어모델의 자가-집중을 활용하여 추가학습만을 이용하는 모델 등이 있다[4,8].

Park et al. 연구는 한국어 BERT 언어모델을 이용하여 입력 문장을 토큰 열로 분할한 이후, 언어모델의 문맥 반영 벡터를 계산하였고, 어절 범위 자질 임베딩, 형태소 범위 자질 임베딩 벡터를 추가하여 양방향 LSTM 순환신경망을 적용한 후, 각 어절의 첫 토큰에 대한 지배소 및 의존 관계 레이블 인식 모델을 적용하였다[4].

Lim et al. 연구는 한국어 BERT 언어모델 위에 추가적인 순환신경망을 사용하지 않고, 입력 문장을 형태소 분석과 언어모델의 토큰 분리를 통하여 토큰 단위로 자가-집중을 이용하여 문맥 표현을 계산한다. 그 후 문맥 반영 벡터를 어절로 표현한 후 추가학습만을 이용한 각 어절 별로

지배소와 의존소 그리고 의존관계 레이블을 인식하는 모델을 적용하였다[8].

최근 특정 언어에 특화되지 않고 다국어에 공통적으로 적용할 수 있도록 UD 트리뱅크를 사전학습 다국어 기반 BERT 언어모델로 추가학습만을 적용한 의존 구문분석 모델 등이 있다[18-19].

Kondratyuk et al. 연구는 사전학습 다국어 M-BERT를 기반으로 75개 언어를 사용하여 124개의 UD 트리뱅크에 대해 UPOS, Lemmas, DEPREL 등을 동시에 자동으로 예측할 수 있는 모델을 제안하였다. Biaffine attention 모델과 각 UD 태스크에 대한 간단한 softmax 분류기만을 적용하여 다국어 M-BERT 모델을 추가학습을 통해 최고 성능을 보였다[19].

본 논문에서는 Kondratyuk et al. 연구에서 제안한 모델을 수정하여 범용 한국어 코퍼스를 대상으로 범용 품사 태깅, 표제어추출, 의존 구문분석을 동시에 예측할 수 있는 다중 작업 모델을 제안하였다.

III. The Proposed Model

Fig. 2는 Kondratyuk et al.가 제안한 모델 구조를 수정하여 설계했으며[19], CoNLL-U format에 따라 사전학습 BERT 언어모델의 토큰 단위 문맥을 이용하여 한국어 UD 주석에 대한 다중 예측 작업을 수행하는 모델의 구성도이다.

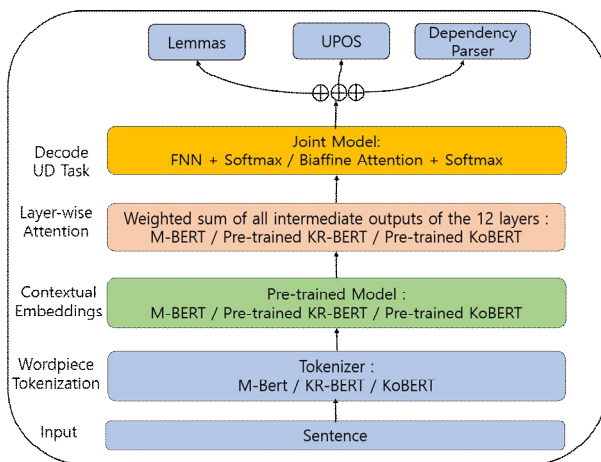


Fig. 2. A multi-task self-attention model

본 논문에서 제안한 Fig. 2의 다중 작업 모델 네트워크는 첫째, 입력 문장 (Input)을 사전학습 언어모델의 입력 단위로 토큰화 (Wordpiece Tokenization), 둘째, 언어모

델에 의해 각 토큰을 문맥 임베딩 벡터로 표현 (Contextual Embeddings), 셋째, 언어모델의 각 레이어 별로 해당 토큰에 대한 자가-집중을 통한 가중치를 추출하여 가중 합 계산 (Layer-wise Attention), 그리고 마지막 단계에서 한국어 UD의 각 품사 태깅, 표제어추출 및 의존 구문 태그별 주석을 예측 수행하는 예측 분류모델 (Decode UD Task)로 구성된다.

IV. The Model Implementation

본 장에서는 제안한 모델의 각 구성 요소에 대한 구현 방법을 기술한다. 본 실험에서는 어휘가 늘어나는 문제를 피하기 위해 모든 입력에 대해 BERT의 wordpiece 토큰 나이저를 직접 사용하였다. 문맥 임베딩 표현 방법으로서 사전학습 언어모델인 M-BERT, KoBERT, 그리고 KR-BERT 모델을 각각 사용하였다.

1. Multi-Task Learning Joint Model

한국어를 대상으로 보편적인 형태 및 구문분석 관계를 동시에 처리하는 다중 작업 학습 (multi-task learning) 방법은 한국어 UD 데이터 셋을 이용해 품사 태깅, 표제어 추출 그리고 의존 구문분석을 동시에 수행하는 공동 모델 (joint model)을 구현하였다.

이 공동 모델은 사전학습 BERT 모델의 Wordpiece Tokenization, Contextual Embeddings, Layer-wise Attention 기법을 공유한다.

1.1 UPOS Tagger

보편적 품사 태깅 태스크 (UPOS)는 입력 문장의 각 단어마다 특정한 UD 주석 레이블을 다는 문제이다.

품사 태깅 모델은 각 입력 단어를 사전학습 모델 BERT의 문맥 임베딩 벡터로 표현하여 Layer-wise Attention 기법을 적용하여 FNN 레이어와 softmax 함수를 추가하였다.

1.2 Lemmatization

보편적 표제어추출 태스크는 입력 단어로부터 표제어 (Lemmas)를 찾아가는 과정이다.

Kondratyuk et al. 연구에서 제안한 표제어추출 규칙을 사용하였고, 특정한 UD 표제어추출 규칙을 분류하기 위해 UPOS 태깅 모델 구조와 같은 모델을 구현하였다.

1.3 Dependency Parser with Biaffine Attention

의존 구문분석 (Dependency parser) 태스크는 문장 단어들의 지배소와 의존소 관계를 찾는 문제이다.

Fig. 3의 보편 의존 구문분석 모델은 Dozat와 Manning 연구에서 제안한 그래프 기반 Biaffine attention 모델을 수정하여 구현하였다[21].

제안 모델은 양방향 LSTM 심층신경망 대신 사전학습 언어모델 BERT로 대체하여 추가 학습을 수행하였다. 각 입력 단어의 BERT 임베딩 레이어, Layer-wise Attention 레이어, 다층 퍼셉트론 (Multi-Layer Perceptron; MLP) 신경망을 추가해서 단어 간의 지배소와 의존소 관계를 분류한다.

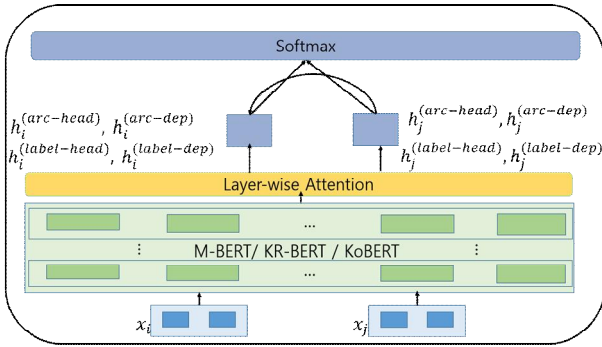


Fig. 3. Biaffine Attention classifier for Dependency Parsing

각 단어의 의존 정보와 의존 레이블정보를 얻기 위해 각 단어로부터의 BERT 모델 위의 Layer-wise Attention 레이어로부터 입력 표상인 x_i 에 대해 각 단어의 표상 r_i 을 추출한다. 그리고 MLP 신경망을 통해 각 r_i 에 대한 비선형 연산과 활성화 함수 ReLU를 적용하여 각 의존소에 대한 지배소의 $h_i^{(arc-head)}$ 와 $h_i^{(arc-dep)}$ 그리고 관련된 레이블 $h_i^{(label-dep)}$ 와 $h_i^{(label-head)}$ 를 결정한다. 그 다음 Biaffine attention 함수를 적용하여 $s_{i,j}^{arc}$ 와 $s_{i,j}^{label}$ 을 계산한다. 계산된 어텐션 스코어는 softmax 함수에 의해 의존 파싱 결과를 출력한다.

$$h_i^{(arc-dep)} = ReLU(MLP^{(arc-dep)}(r_i)),$$

$$h_i^{(arc-head)} = ReLU(MLP^{(arc-head)}(r_i)),$$

$$h_i^{(label-dep)} = ReLU(MLP^{(label-dep)}(r_i)),$$

$$h_i^{(label-head)} = ReLU(MLP^{(label-head)}(r_i)),$$

$$s_{i,j}^{(arc)} = h_i^{(arc-head)} U h_j^{(arc-dep)} + w_i h_j^{(arc-head)}$$

$$s_{i,j}^{(label)} = h_i^{(label-dep)} U h_j^{(label-head)} + w_i h_j^{(label-dep)}$$

2. Fine-tuning BERT on UD Annotations

Fig. 2의 사전 훈련된 BERT 언어 모델위에 공동 모델 (joint model)을 추가하여 한국어 UD 태스크에 대해 추가 학습할 때 Kondratyuk et al. 연구에서 제안한 세 가지 최적화 기법을 적용하였다[19].

2.1. Discriminative Fine-tuning

자연어처리 응용의 모든 다운스트림 태스크에 적용할 수 있는 보편적인 전이 학습 (transfer learning) 방법으로서 Kondratyuk et al.는 사전 학습된 다국어 언어모델 M-BERT을 Discriminative fine-tuning (이하 Discr) 알고리즘을 적용하여 특정 UD 태스크에 맞게 미세 조정한다. Discr 알고리즘은 신경망 모델의 모든 계층에 대해 동일한 학습률을 사용하는 대신 각 계층을 서로 다른 학습률을 분별하여 미세 조정하는 것을 말한다[19].

본 실험에서는 다중 작업 학습 모델의 UD 주석 예측을 위해 적용한 첫 번째 최적화 기법으로서 Kondratyuk et al.이 수정한 Discr 알고리즘을 적용하여 특정 태스크에 맞게 미세 조정하였다. 제안한 다중 작업 모델을 추가 학습할 때 BERT 네트워크의 기본 학습율은 $5e^{-5}$ 로 설정하고 분류기 네트워크에서는 e^{-3} 로 다르게 설정하였다. 또한 vanishing gradient 문제와 underfitting 문제를 피하기 위해 inverse square root learning rate decay와 linear warmup을 적용하여 모델의 가중치를 조정하였다.

2.2. Layer-wise Attention

UD 주석 예측을 위한 추가학습 단계에서 적용한 또 다른 최적화 기법인 Kondratyuk et al.이 제안한 Layer Attention 기법을 적용하였다.

Layer Attention 기법은 사전 학습 M-BERT의 12개 블록을 이용하여 레이어별로 임베딩 벡터 표상을 추출하여 dot-product attention 기법을 적용하여 출력 층에서는 입력 문장의 각 토큰의 위치에 FNN 레이어와 softmax 함수를 사용하여 UD 주석에 대한 예측을 하게 된다.

본 실험에서는 M-BERT와 사전학습 한국어 BERT를 사용하여 각 12개 BERT 블록 출력 값이 가진 정보는 전부 서로 다른 종류의 정보를 갖고 있을 것이므로, 이들을 모두 활용한다는 점에서 12개 블록의 출력 값을 연결하고 동

일한 가중치로 가중치 합을 계산하는 블록 i 에 대한 j 번째 단어(w_j)에 대한 표상을 다음과 같이 정의한다[19].

$$e_j^{task} = \gamma^{task} \sum_i^{12} BERT_{ij} \cdot s(w_j)^{task} \quad (1)$$

위의 수식 (1) 에서 $BERT_{ij}$ 는 i 번째 블록에서 j 번째 토큰에 대한 임베딩 벡터로 표상한다. 특정한 한국어 UD 태스크의 문장에 대한 표상 $e^{(task)}$ 는 각 토큰별 BERT 블록의 출력 벡터 $BERT_{ij}$ 을 모두 더하는데 이 때 각각에 softmax 함수를 통한 동일한 가중치인 $s(w_j)^{task}$ 을 스칼라 곱 연산을 진행한 뒤 더하게 된다. 마지막에 최종적으로 γ^{task} 을 곱하게 되는데 스칼라 파라미터이다. $s(w_j)^{task}$ 와 γ^{task} 는 모두 학습 파라미터가 된다.

2.3. Layer Dropout

다중 작업 학습 모델의 한국어 UD 주석 예측을 위해 적용한 나머지 최적화 기법으로서 Kondratyuk et al.이 제안한 Layer Dropout 기법을 적용한다[19].

Layer Dropout 기법은 한국어 UD 주석 예측 모델의 추가학습 단계에서 임의의 블록의 정보에 과적합 되는 것을 방지하기 위해 각 학습 단계에서 각 매개변수 w_j 가 $-\infty$ 일 때 확률 0.1로 설정하였다. UPOS, Lemmas, 그리고 Dependency Parser 각각에 대해 하나의 $s(w_j)^{task}$ 와 γ^{task} 매개변수 세트를 사용하여 수식 (1)의 e_j^{task} 를 계산한다[19].

V. Experiments

본 논문에서 제안한 한국어 UD 주석에 대한 다중 작업 모델 실험은 파이토치(PyTorch)를 이용하여 구현하였으며, 제안한 모델의 성능을 분석하기 위해 사용한 사전학습 언어모델은 M-BERT, KoBERT, 그리고 KR-BERT 모델을 적용하였다. 실험에서 사용한 하드웨어는 Table 3과 같다.

Table 3. Hardware configuration

Name	Version
GPU	RTX A5000
CPU	i9-10980XE
RAM	32 GB
HDD	2 TB
SSD	1 TB

1. Datasets

다중 작업 모델은 Chun et al.이 구축한 UD v2.3의 UD Korean Kaist 트리뱅크를 기반으로 학습하였다[12].

UD의 주석체계를 따르는 UD Korean Kaist 트리뱅크는 CoNLL 형식의 데이터로 구성되며 UPOS, Lemmas, DEPREL 주석을 제공한다.

다중 작업 모델을 훈련하기 위해 학습 데이터, 개발 데이터 그리고 평가 데이터로 ko_kaist-ud-train.conllu, ko_kaist-ud-dev.conllu, ko_kaist-ud-test.conllu 데이터셋을 사용하였다[12]. 실험 데이터의 크기는 Table 4와 같다.

Table 4. UD Korean Kaist Treebank v2.3

Corpus	Size
Sentences	27,363
Tokens	350,090
Tokens (11%) that are not followed by a space	38,923

2. Metrics

제안 모델의 성능 평가 항목으로 F1 score, UAS 그리고 LAS을 이용하여 평가를 수행하였다.

F1 score는 정밀도와 재현율을 결합하여 만든 지표로서 Lemmas와 UPOS 태그 분류에 대한 평가를 수행하였다.

의존 구문분석 모델에 대한 평가는 UAS(Unlabeled attachment score)와 LAS(Labeled attachment score) 지표를 사용하였다. UAS는 전체 단어 중 지배소를 올바르게 인식한 단어 정확도이고, LAS는 지배소와 의존관계 레이블을 모두 올바르게 인식한 단어 정확도를 의미한다.

3. Hyperparameters

제안한 한국어 UD 주석에 대한 다중 작업 모델 실험에 사용했던 주요 하이퍼파라미터는 Table 5와 같다.

모델 학습은 epoch 10, learning rate $1e^{-3}$, 배치크기 32를 모든 태스크에 동일하게 적용하였다. 지배소와 의존소 인식 및 레이블 인식 모델 학습의 dropout 비율은 0.5를 적용하였다.

사전학습 언어모델의 토큰에 대한 임베딩 벡터 표상을 얻기 위해 M-BERT는 bert-base-multilingual-cased, KR-BERT는 KR-BERT-char16424, KoBERT는 kobert-base-v1 모델을 사용하였다.

Table 5. A summary of model hyperparameter

Hyperparameter	Value	
model: token embedders	pretrained model	M-BERT
		KR-BERT
		KoBERT
	requires_grad	True
	drouput	0.15
	layer dropout	0.1
combine layers	all	
model : classifiers	tasks	UPOS
		Lemmas
		Dependency Parser
model : encoder	input dim	768
model : decoder: UPOS, Lemmas	input dim	768
model: Dependency parser	tag dim	256
	arc dim	768
	input dim	768
trainer: optimizer	type	Adam
	β_1, β_2	0.9, 0.99
	weight decay	0.01
	learning rate	1e-3
trainer	epochs	10
trainer: learning rate scheduler	type	ulmfit_sqrt
	warmup steps	392
	discriminative fine tuning	True
iterator	batch size	32
	samples per batch	100

VI. Results

본 장에서는 UD 가이드라인에 따라 CoNLL U-Format 으로 변환한 한국어 UD Korean Kaist 트리뱅크 코퍼스를 사용하여 다국어 M-BERT, 한국어 특화된 KR-BERT와 KoBERT 기반 다중 작업 모델들을 대상으로 추가학습을 수행한 실험 결과를 비교 기술한다. 제안한 모델의 실험 결과는 Table 6과 같다.

Table 6. A summary of the model performance

Metric	M-BERT	KR-BERT	KoBERT
UPOS	95.15	95.23	56.49
Lemmas	88.69	88.29	36.89
UAS	86.23	86.87	55.24
LAS	83.32	83.99	32.56

제안한 다중 작업 모델은 평가 세트를 대상으로 M-BERT 기반 모델은 95.15% UPOS, 88.69% Lemmas, 86.23% UAS, 83.32% LAS 성능을 보였다. KR-BERT 기반 모델은 95.23% UPOS, 88.29% Lemmas, 86.87%

UAS, 83.99% LAS 성능을 보였다. KoBERT 기반 모델은 56.49% UPOS, 36.89% Lemmas, 55.24% UAS, 32.56% LAS 성능을 보였다.

자가-집중 기법을 이용한 Layer Attention과 Biaffine Attention 기법을 사용하여 추가학습을 수행한 KR-BERT 기반 모델은 M-BERT 기반 모델보다 95.23% UPOS, 86.87% UAS, 83.99% LAS 로 약간 높은 성능을 보였다. 이는 CoNLL U-Format 한국어 UD Korean Kaist 트리뱅크 코퍼스 대상으로 실험한 결과를 비교하였을 때 다국어 M-BERT 기반 모델보다 한국어 KR-BERT가 보편적인 UD 문법체계 인식 성능 개선이 가능함을 보인 결과이다. 그러나 KoBERT 기반 모델을 사용하였을 경우, 다른 M-BERT와 KR-BERT 기반 모델들에 비하여 매우 낮은 성능을 보였다. 이는 Table 2를 참고로 하면 사전학습 언어모델의 학습 데이터와 전처리 방식의 차이에 기인한다고 볼 수 있다.

VII. Conclusions

본 논문에서는 사전학습 BERT 계열 다국어 버전 (M-BERT)과 한국어 버전 (KR-BERT, KoBERT) 모델을 기반을 둔 UD 태스크 다중 작업 모델을 제안하였다.

Kondratyuk et al.의 모델 구조를 수정하여 설계한 제안 모델을 통해 UD Korean Kaist 트리뱅크 코퍼스를 대상으로 사전학습 언어모델 BERT가 가지는 토큰 단위 문맥 표상을 이용하여 자가-집중 기법과 그래프 기반 Biaffine attention 기법을 적용하여 보편적 품사 태깅, 표제어 추출, 의존 구문 파싱을 동시에 처리하였다.

제안 모델은 순환신경망 계층을 사용하지 않고, BERT 모델의 추가학습만을 이용한 UD 한국어 태스크를 동시에 수행하였다. 또한 모델의 성능 개선을 위하여 BERT 모델의 12개 블록 계층을 이용한 Kondratyuk et al.가 제안한 Layer Attention 기법을 적용하였다.

제안 기법은 UD Korean Kaist 평가 세트를 대상으로 M-BERT 기반 모델은 95.15% UPOS, 88.69% Lemmas, 86.23% UAS, 83.32% LAS 성능을 보였고, KR-BERT 기반 모델은 95.23% UPOS, 88.29% Lemmas, 86.87% UAS, 83.99% LAS 성능을 보였고, KoBERT 기반 모델은 56.49% UPOS, 36.89% Lemmas, 55.24% UAS, 32.56% LAS 성능을 보였다.

자연어처리 응용에 있어서 문장의 구조적 중의성 해소 관련한 한국어 의존 구문분석에 대한 많은 국내 연구가 있

지만 전 세계 언어를 하나의 보편적인 형태/통사적 구문 문법 기준으로 동일하게 분석하자는 목표로 구축 중인 UD 트리뱅크 코퍼스를 이용하여 한국어를 포함한 다국어 지원 자연어처리 응용 태스크 분석에 대한 연구는 상대적으로 적다. 따라서 향후 연구계획은 한국어 특성을 잘 반영한 개선된 한국어 UD 체계를 활용한 트랜스포머 언어 모델에 적용하여 다국어 기반 자연어 처리 태스크 연구 결과와 비교하고자 한다.

ACKNOWLEDGEMENT

This work was supported by Shinhan Univ. Sabbatical Leave Program for Tenured Faculty in 2021.

REFERENCES

- [1] L. Joon-Ho et al., "Korean Dependency Guidelines for Dependency Parsing and Exo-Brain Language Analysis Corpus," *proc. of the 27th Annual Conference on Human & Cognitive Language Technology*, pp. 224-239, Oct. 2015.
- [2] J. Devlin et al., "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," <https://arxiv.org/abs/1810.04805v2>
- [3] P. Cheoneum et al., "Korean Dependency Parsing with BERT," 2019 Proceedings of the Korean Information Science Society Conference, pp. 530-532, 2019
- [4] L. Joon-Ho et al., "Korean Dependency Parsing using the Self-Attention Head Recognition Model," *Journal of Korean Institute of Information Scientists and Engineers*, Vol. 46, No. 1, pp. 22-30, Jan. 2019.
- [5] P. Seongsik et al., "Korean Dependency Parsing Using ELMo and Multi-head Attention," *Proceedings of the 30th Annual Conference on Human and Cognitive Language Technology*, pp. 8-12, 2018
- [6] A. Hwiyeon et al., "Korean Dependency Parsing Using Deep Bi-affine Network and Stack Pointer Network," *Proceedings of the 30th Annual Conference on Human and Cognitive Language Technology*, pp. 689-691, 2018
- [7] L. Jinu et al., "Empirical Research on Segmentation Method for Korean Dependency parsing," *Proceedings of the 33th Annual Conference on Human and Cognitive Language Technology*, pp. 427-432, 2021
- [8] L. Joon-Ho et al., "Korean Dependency Parsing using Token-Level Contextual Representation in Pre-trained Language Model," *Journal of Korean Institute of Information Scientists and Engineers*, Vol. 48, No. 1, pp. 27-34, Jan. 2021.
- [9] N. Seung-Hoon et al., "Deep Biaffine Attention for Korean Dependency Parsing," *Proceedings of the Korean Information Science Society Conference*, pp. 584-586, 2017
- [10] J. Nivre et al., "Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection," *Proceedings of the 12th Conference on Language Resources and Evaluation*, pp. 4034-4043, 2020.
- [11] Universal Dependencies, <https://universaldependencies.org/>
- [12] Chun et al., https://universaldependencies.org/treebanks/ko_kaist/
- [13] C. Jayeol et al., "Building Universal Dependency Treebanks in Korean," *Proceedings of the 11th International Conference on Language Resources and Evaluation*, pp. 2194-2202, 2018
- [14] P. Hye-Jin et al., "Universal POS Tagset for Korean," *Proceedings of the 30th Annual Conference on Human and Language Technology*, pp. 417-421, 2018.
- [15] O. Taehwan et al., "Manual Revision of Penn Korean Universal Dependency Treebank," *Proceedings of the 33th Annual Conference on Human and Cognitive Language Technology*, pp. 61-65, 2021.
- [16] KR-BERT, <https://github.com/snunlp/KR-BERT>
- [17] KoBERT, <https://github.com/SKTBrain/KoBERT>
- [18] M. Straka et al., "Evaluating Contextualized Embeddings on 54 Languages in POS Tagging, Lemmatization and Dependency Parsing," <https://arxiv.org/abs/1908.07448>
- [19] D. Kondratyuk et al., "75 Languages, 1 Model: Parsing Universal Dependencies Universally," *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 2779-2795, 2019
- [20] T. Dozat and C. D. Manning, "Deep Biaffine Attention for Neural Dependency Parsing," <https://arxiv.org/abs/1611.01734>

Authors



Euhee Kim received the M.S. degrees in Computer Engineering from Dongguk University, Korea, in 2002 and Ph.D. degrees in Mathematics from The University of Connecticut, U.S.A in 1995.

Euhee Kim is currently a Professor in the Department of Computer Science & Engineering at Shinhan University. She is interested in AI, NLP and Big Data computing.