

## Biaffine Average Attention 모델을 이용한 의미역 결정

남충현<sup>1</sup> · 장경식<sup>2\*</sup>

### Semantic Role Labeling using Biaffine Average Attention Model

Chung-Hyeon Nam<sup>1</sup> · Kyung-Sik Jang<sup>2\*</sup>

<sup>1</sup>Graduate Student, Department of Computer Engineering, Korea University of Technology and Education, Cheonan, 31253 Korea

<sup>2\*</sup>Professor, Department of Computer Engineering, Korea University of Technology and Education, Cheonan, 31253 Korea

#### 요 약

의미역 결정 작업은 서술어와 문장 내 행위자, 피행위자, 장소, 시간 등 서술어와 관련 있는 논항들을 추출하는 작업이다. 기존 의미역 결정 방법은 문장의 언어학적 특징 추출을 위한 파이프라인을 구축하는데, 파이프라인 내 각 추출 작업들의 오류가 의미역 결정 작업의 성능에 영향을 미치기 때문에 현재는 End-to-End 방법의 신경망 모델을 이용한 방법들이 제안되고 있다. 본 논문에서는 의미역 결정 작업을 위해 Biaffine Average Attention 구조를 이용한 신경망 모델을 제안한다. 제안하는 모델은 기존 연구에서 제안된 특정 시점에 대한 레이블 예측을 위해 주변 시점 정보를 이용하는 LSTM 모델 대신 문장 내 서술어와 논항의 거리에 상관없이 문장 전체 정보에 집중할 수 있는 Biaffine Average Attention 구조로 이루어져 있다. 제안하는 모델의 성능 평가를 위해 F1 점수를 이용하여 기존 연구에서 제안한 BERT 기반의 모델들과 비교하였으며, 76.21%의 성능으로 비교 모델보다 높은 성능을 보였음을 확인하였다.

#### ABSTRACT

Semantic role labeling task(SRL) is to extract predicate and arguments such as agent, patient, place, time. In the previously SRL task studies, a pipeline method extracting linguistic features of sentence has been proposed, but in this method, errors of each extraction work in the pipeline affect semantic role labeling performance. Therefore, methods using End-to-End neural network model have recently been proposed. In this paper, we propose a neural network model using the Biaffine Average Attention model for SRL task. The proposed model consists of a structure that can focus on the entire sentence information regardless of the distance between the predicate in the sentence and the arguments, instead of LSTM model that uses the surrounding information for prediction of a specific token proposed in the previous studies. For evaluation, we used F1 scores to compare two models based BERT model that proposed in existing studies using F1 scores, and found that 76.21% performance was higher than comparison models.

**키워드** : 딥러닝, 자연어 처리, 의미역 결정

**Keywords** : Deep Learning, Natural Language Processing, Semantic Role Labeling

Received 13 December 2021, Revised 18 March 2022, Accepted 5 April 2022

\* Corresponding Author Kyung-Sik Jang(E-mail:ksjang@koreatech.ac.kr, Tel:+82-41-560-1352)

Professor, Department of Computer Engineering, Korea University of Technology and Education, Cheonan, 31253 Korea

Open Access <http://doi.org/10.6109/jkiice.2022.26.5.662>

print ISSN: 2234-4772 online ISSN: 2288-4165

© 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.  
Copyright © The Korea Institute of Information and Communication Engineering.

## I. 서론

의미역 결정(Semantic Role Labeling) 작업은 문장 내 “누가 어디서 누구에게 무엇을 하였는가?”와 같이 문장에서 핵심이 되는 성분들 간의 관계성을 파악하고 구조화하는 작업이며, 의미역은 서술어(Predicate)와 행위자, 피행위자, 장소, 시간 등과 같이 서술어와 관련 있는 논항들(Arguments)로 구성되어진다. 이러한 의미역 결정 작업은 번역 시스템, 질의응답 시스템, 정보 추출 등 자연어 처리를 이용한 다양한 분야에서 활발히 적용되어지고 있다[1-2].

의미역 결정 작업에서 서술어는 대부분 동사 형태 단어로 이루어져 있기 때문에 문장 내 형태소 분석을 통한 결과 중 동사형 품사를 인식하여 사용하면 되지만, 논항 인식 작업은 문장 내 특정 서술어에 대해 사전 정의된 언어학적 패턴들과 형태소 분석, 의존 구문 분석과 같은 분석 방법을 이용하여 추출된 해당 문장의 언어학적 특징을 이용하여 인식한다.

하지만 앞서 언급한 전처리 작업들의 정확도는 논항 인식 작업 성능에 영향을 미치며, 전처리 작업에서 발생된 오류는 문장 내 논항의 범위를 잘못 인식하거나, 인식된 논항을 잘못 분류하게 되는 문제를 발생시킬 수 있다[2].

최근 이러한 문제점을 해결하기 위해 문장의 언어학적 특징 분석 및 추출 과정을 포함하는 전처리 작업 없이 오직 문장만을 입력으로 사용한 End-to-End 신경망 모델을 이용한 연구들이 제안되고 있다[3-6]. 특히, 문장을 이루는 각 단어 또는 토큰들에 대해 논항의 범위를 인식하고, 파악된 논항 범위가 어떤 논항에 속하는지 분류하는 작업으로 볼 수 있기 때문에 순차 레이블링(Sequence Labeling) 방법을 이용하여 학습된 신경망 모델을 의미역 결정 작업에 적용하는 방법들이 제안되고 있다.

의미역 결정 작업에서 사용된 대표적인 신경망 모델들은 LSTM(Long Short-Term Memory) 모델[3-4], BERT(Bidirectional Encoder Representations from Transformers) [5-6] 모델 등이 사용되었으며, 또한 성능 개선을 위해 CRF (Conditional Random Field) 층과 같은 신경망 구조를 함께 사용하였다.

본 논문에서는 의미역 결정 작업을 위해 Biaffine Average Attention 구조를 이용한 신경망 모델을 제안한

다. 기존 LSTM 모델은 특정 시점에 대해 주변 시점 정보를 이용하지만, 의미역 결정 작업은 문장 내 서술어와 논항이 가까운 위치뿐만 아니라 멀리 떨어져 있는 경우도 존재하기 때문에 LSTM 모델 대신 Biaffine Average Attention 구조를 사용하여 문장 전체 정보에 집중할 수 있도록 하였다.

본 논문에서 제안한 신경망 모델의 성능을 평가하기 위해 F1 점수를 이용하여 기존 연구에서 제안한 모델들과 비교하였다. 그 결과, 76.21%의 성능으로 BERT-CRF 모델과 BERT-BiLSTM-CRF 모델에 비해 각각 1.09%, 1.7%의 성능 개선이 있었음을 확인하였다.

## II. 관련 연구

신경망 모델을 사용하지 않은 의미역 결정 작업 연구는 형태소 분석, 의존 구문 분석 등 전처리 작업을 이용하여 품사, 구문 트리 등 언어학적 패턴을 사전에 구축한 후, 입력 문장의 언어학적 특징을 추출하기 위해 동일한 전처리 작업을 수행할 수 있는 파이프라인을 구축하여 작업을 수행하였다.

하지만 파이프라인 방법을 이용한 의미역 결정 작업의 문제는 사전에 언어의 다양한 패턴을 모두 정의할 수 없으며, 수많은 패턴을 정의하기 위해 대량의 말뭉치와 많은 시간, 인력이 필요하다는 점이다.

또한 구성된 파이프라인 내 각 작업들은 순차적인 작업들로 진행되기 때문에 작업 진행 과정에서 작업 결과들의 오류가 누적되는 경우, 최종 의미역 결정의 결과에 영향을 미치게 되어 성능을 저하시킬 수 있다는 문제점을 갖고 있다[2].

이러한 문제를 해결하기 위해 최근 의미역 결정 연구는 사전 언어학적 특징 정의 및 파이프라인을 구축하지 않고 신경망 모델을 접목한 End-to-End 추출 방식을 주로 제안하고 있으며, 특히 신경망 모델 중 대표적으로 LSTM 모델과 BERT 모델이 사용되어진다. 먼저 LSTM 모델은 시계열 데이터를 처리하기 위한 모델이며, 얼마나 과거 시점의 정보를 잊을지에 대한 망각 게이트(Forget Gate)와 현재 시점의 정보를 얼마나 기억할지에 대한 입력 게이트(Input Gate) 구조로 이루어져 있다.

기존 LSTM 모델을 사용한 연구[3]에서는 양방향 구조를 여러 층 쌓은 LSTM 모델과 CRF 층을 사용하였다.

또한 LSTM 모델과 어텐션 메커니즘(Attention Mechanism)을 사용한 연구[4]에서는 각 시점의 결과 값을 위해 주변 시점의 정보를 이용하는 LSTM 모델의 결과 값을 어텐션 메커니즘을 이용하여 종합적으로 봄으로써, 각 시점의 단어를 연관성 있는 다른 시점의 단어들에 집중할 수 있도록 하였다. 이 신경망 모델의 구조는 의미역 결정 작업 내에서 서술어와 논항들 간의 관계를 집중적으로 파악할 수 있도록 학습이 가능해지는 특징을 갖고 있다.

두 번째로, BERT 모델은 자기 집중 메커니즘(Self-Attention Mechanism)을 사용한 인코더(Encoder)와 디코더(Decoder) 구조로 이루어진 Transformers 모델에서 인코더 구조를 여러 층으로 구성된 모델이다[5]. 또한 BERT 모델은 사전에 대량의 말뭉치를 이용하여 언어 모델(Language Model) 역할을 할 수 있도록 학습되었으며, 가중치 미세 조절(Fine-Tuning) 방법을 이용하여 자연어 처리 내 다양한 분야에 맞게 학습하여 사용할 수 있으며, 이렇게 학습된 모델들은 기존 제안된 다른 모델들보다 높은 성능을 보였다.

BERT 모델을 이용한 의미역 결정 연구는 주로 BERT 모델을 문장 내 토큰들을 입력 받는 임베딩 층으로 사용하였으며, CRF 층을 추가한 BERT-CRF 모델 또는 LSTM 모델을 추가한 BERT-LSTM-CRF 모델로 사용하였다[6-7].

BERT 모델의 입력으로 사용된 문장은 Wordpiece 분할 방법을 이용하여 토큰들로 분할되며, 의미역 결정 작업에서 서술어 정보를 문장 토큰들과 분리된 정보로 인식할 수 있도록 [SEP] 토큰을 문장과 서술어 정보 사이에 삽입하여 모델에 입력하였다.

### III. 의미역 결정을 위한 신경망 모델

의미역 결정 작업은 각 입력 토큰들에 대해 특정 레이블을 할당하는 순차 레이블링 작업으로 볼 수 있다. 본 논문에서 제안하는 신경망 모델은 순차 레이블링 작업을 이용하여 의미역을 결정하며, 입력 문장  $S$ 이 주어졌을 때, 분할된  $N$ 개의 길이로 이루어진 입력 토큰  $T = \{t_1, t_2, \dots, t_N\}$ 들에 대해 모델이 각 토큰들에 대응하는 출력  $Y = \{y_1, y_2, \dots, y_N\}$ 을 예측하고 각 출력  $y_i$ 는  $L$ 개의 논항들 중 한 개의 논항을 할당하는 작업을 목표로 한다.

의미역 결정 작업은 서술어에 의존적이며, 서술어에 따라 문장 내 논항의 위치, 속성이 바뀔 수 있다. 하지만 순차 레이블링 작업을 이용하는 개체명 인식 작업 같은 경우 대부분 문장의 문맥보다는 사람, 지역과 같이 특정 단어의 특징 또는 속성에 따라 레이블이 결정되는 경우가 많다.

그렇기 때문에 개체명 인식 작업을 위한 신경망 모델은 주로 BERT 모델과 함께 양방향 다층 구조 LSTM 모델을 사용한다. 양방향 다층 구조 LSTM 모델은  $t$ 시점에 대한 출력은  $t-1, t+1$ 시점의 정보와  $t$ 시점의 입력으로 이루어지며, 이러한 주변 시점의 정보를 이용하는 LSTM 모델은 특정 시점의 예측에 있어 추가적인 특징을 제공하여 주변 시점에 의존적인 결과를 보일 수 있다는 특징을 갖고 있다.

하지만 문장 내 서술어와 논항이 가까운 위치에 있지 않고 멀리 떨어져 있는 경우, 주변 시점 정보보다는 문장 전체에 대한 정보가 필요하며, 본 논문에서 제안하는 신경망 모델은 의미역 결정 작업을 위해 문장 전체에 대한 정보를 파악할 수 있는 Biaffine Average Attention 구조를 갖고 있는 모델을 제안한다.

제안하는 모델은 임베딩 층으로 사용하는 BERT 모델과 두 개의 전 완전 연결 층(Fully-Connected Layer)과 한 개의 Biaffine Attention 층 그리고 CRF 층을 이용한 구조로 구성되어 있다.

모델 설명에 앞서, 모델 입력은 BERT 모델을 이용한 기존 연구와 동일하게 문장과 서술어 정보 구분을 위해 구분 토큰 [SEP]을 추가하여, 문장, [SEP], 서술어 정보 순으로 구성하였다.

Biaffine Attention은 구문 분석 연구에서 제안된 구조 [8-9]로 임베딩 층에서 출력된 특징 벡터를 두 개의 전 완전 연결 층을 이용하여 각기 다른 추상화된 특징 벡터로 만들어, 쌍 선형 결합을 하는 구조로 이루어져 있으며, 이를 통해 입력 문장 내 각 토큰들이 어떤 토큰들과 관련이 있고, 집중해야 하는지 알 수 있게 된다.

본 논문에서 제안하는 모델을 구조화하는 데 있어 영감을 받은 Biaffine Attention을 이용한 개체명 인식 연구[9]에서는 입력 문장에서 개체에 해당하는 단어의 첫 토큰과 마지막 토큰에 집중할 수 있는 구조를 설계하였다. 본 논문에서도 동일하게 논항이 구성하는 어절들의 시작 토큰과 마지막 토큰을 집중할 수 있도록 학습하기 위해 Biaffine Attention 구조를 사용하였으며, 하나 또

는 비슷한 명사 품사로 결정되어지는 개체명과는 다르게 논항은 다양한 품사가 포함된 어절로 이루어진다는 차이점이 있기 때문에 마지막 토큰에 대한 특징 벡터를 사용하지 않고 집중된 모든 토큰 벡터들에 대해 평균을 취하는 구조로 구성하였다.

$$\begin{aligned} h_{start}(i) &= FFNN_{start}(t_i) \\ h_{end}(j) &= FFNN_{end}(t_j) \\ s_{i,j} &= h_{start}^\top(i) U h_{end}(j) \\ &\quad + W(h_{start}(i) \oplus h_{end}(j)) + b \end{aligned} \quad (1)$$

수식1은 Biaffine Attention 구조에 대한 수식이다. 먼저  $i, j$  시점에 대한 특징을 각각의 전 완전 연결 층을 이용하여 추상화된 특징 벡터  $h_{start}(i), h_{end}(j)$ 을 만든 후, 두 개의 특징 벡터들을 이용해 어절의 시작점  $i$ 와 끝점  $j$ 에 대해  $N \times N \times L$  크기로 이루어진 특징 벡터를 만든다. 여기서  $U, W, b$ 는 학습 과정에서 가중치가 변화될 수 있는 학습 가능한 파라미터이며,  $\oplus$ 는 결합(Concatenation) 연산이다.

다음으로 마지막 토큰에 대한 특징 벡터에 대한 평균 집중 값을 구하기 위해 수식2와 같이 이렇게 만들어진 특징 벡터의 끝 점을 뜻하는  $j$ 열에 대한 평균을 취하여  $N \times L$  크기로 이루어진 벡터  $\Phi_i$ 로 변환한다.

$$\Phi_i = \frac{1}{N} \sum_{j=1}^N s_{i,j} \quad (2)$$

이렇게 만들어진 벡터  $\Phi_i$ 는 CRF 층으로 입력되어진다. CRF 층은 현재 시점에 대한 논항을 예측하는 데 있어서 이전 시점에 대한 논항 상태도 고려하는 조건부 확률 최대화를 목표로 학습한다는 특징을 갖고 있다.

이를 통해 학습 과정에서 논항 속성을 가진 어절의 시작 토큰이 중간 토큰에 또는 중간 토큰이 마지막 토큰에 영향을 주게 되며, CRF 층을 사용하였을 때, 사용하지 않았을 경우보다 더 높은 성능이 나왔음을 실험을 통해 확인하였다.

## IV. 실험 및 결과

### 4.1. 실험 환경

실험을 위한 의미역 결정 학습 데이터 셋은 Ai hub에

서 제공하는 일반 상식 데이터 셋 중 상호참조 복원 태깅 데이터 셋[10]을 사용하였다. 데이터 셋에는 총 14,461개의 데이터가 포함되어 있으며, 본 논문에서는 실험을 위해 데이터 셋을 각각 8:1:1로 나누어 학습, 검증, 평가에 각각 사용하였다. 사용한 데이터 셋에는 필수 격, 보조 격 의미역이 포함된 총 19개의 의미역으로 이루어져 있었으며, 순차 레이블링 방식을 적용하기 위해 Begin, Inside, Outside, End, Single로 이루어진 BIOES 태깅 방식을 이용하여 각각 의미역을 레이블링 하였다.

성능 평가 기준은 F1 점수 방법을 사용하였고, 성능 비교를 위한 모델 중 임베딩 층으로 BERT 모델을 이용한 연구들에서 LSTM 모델과 비교하여 더 높은 성능을 보였기 때문에 LSTM 모델을 임베딩 층으로 사용한 실험은 제외하고, BERT 모델을 이용한 모델들에 대해서만 실험을 진행하였다.

### 4.2. 각 모델에 대한 성능 평가

제안한 모델의 성능 평가를 위해 기존 연구에서 제안한 모델인 BERT-CRF 모델과 BERT-BiLSTM-CRF 모델을 비교 모델로 선정하였다.

**Table. 1** F1 score(%) of each models using validation dataset and test dataset

Models	Valid Dataset	Test Dataset
BERT-CRF	75.09	75.12
BERT-BiLSTM-CRF	75.05	74.51
Our model (without CRF)	75.07	75.09
Our model (with CRF)	75.32	76.21

모델들의 공통된 임베딩 층인 BERT 모델의 파라미터는 앞서 언급한 것처럼 동일한 파라미터로 선정하였지만, BERT-BiLSTM-CRF 모델 내 사용된 LSTM 모델의 하이퍼 파라미터는 본 논문에서 사용한 학습 데이터를 이용하여 학습하는 과정에서 가장 높은 성능을 보인 모델을 따로 저장하여 비교 모델로 사용하였다. 그 결과, LSTM 모델은 2층 양방향 구조를 갖고, 은닉 상태 크기는 256으로 설정하였을 때 가장 높은 성능을 보였다.

표1은 비교 모델 2개와 CRF 출력 층의 사용 유무에 따른 제안한 모델의 성능 실험 결과를 나타낸다. 성능 결과 비교에 앞서, 표에는 언급하지 않았지만 CRF 출력 층을 사용하지 않은 경우에는 손실 함수에 대한 추가적인 성능 평가를 위해 크로스 엔트로피 손실 함수와

Focal 손실 함수[11]를 이용하여 각각 학습시켰다. 크로스 엔트로피 손실 함수는 다중 분류 문제에서 주로 사용되는 손실 함수이며, Focal 손실 함수는 순차 레이블링 작업은 데이터 불균형 문제가 빈번하게 발생하는데, 이러한 문제를 완화하기 위해 제안된 손실 함수이다.

Focal 손실 함수를 사용한 경우 평가 데이터 셋을 이용하여 평가한 경우 약 74.95%의 성능을 보였으며, 이는 비교 모델 중 BERT-BiLSTM-CRF 모델보다 높은 성능을 보였지만, 크로스 엔트로피 손실 함수에 비해 상대적으로 적은 성능을 보였기 때문에 표에 명시하지 않았다.

실험에서 사용된 4가지 모델 중 검증 데이터 셋과 평가 데이터 셋에 대한 성능이 각각 BERT-CRF 모델에서 75.09%, 75.12%, BERT-BiLSTM-CRF 모델이 75.05%, 74.51%의 결과를 보였지만 본 논문에서 제안한 CRF를 사용한 모델의 F1 점수는 각각 75.32% 76.21%로 기존 제안된 모델보다 검증 데이터 셋과 평가 데이터 셋 모두 상대적으로 더 높은 성능을 보였다. 또한 제안한 모델의 CRF 출력 층의 사용 유무에 따라 평가 데이터 셋에 대해 약 1.1% 정도의 성능 차이가 있었음을 확인하였다.

실험 결과를 통해 Biaffine Average Attention 구조를 이용한 의미역 결정 작업을 위한 신경망 모델이 기존 제안된 모델보다 높은 성능을 보였음을 실험적으로 확인하였다.

그림1은 본 논문에서 제안한 모델이 예측한 논항을 의미하는 가로축과 정답 논항을 의미하는 세로축에 대한 Confusion Matrix이며, 논항이 아닌 O 태그를 포함한 총 20개의 레이블에 대해 얼마나 예측한 논항이 정답 논항과 맞았는지, 틀렸다면 어떤 논항으로 예측하였는지 나타내고 있다.

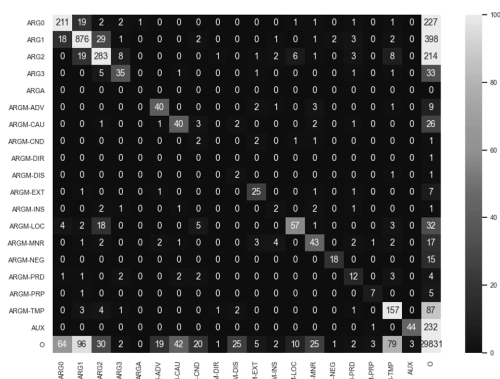


Fig. 1 Confusion matrix of our proposed model

필수 격을 의미하는 ARG0, ARG1, ARG2, ARG3 중 ARG1, ARG2, ARG3은 각각 876, 283, 35개로 예측 논항이 정답 논항에 가장 많이 맞았음을 확인하였으며, ARG0 논항 같은 경우 211개의 예측 논항이 정답 논항과 일치하였지만, O 태그임에도 불구하고 227개의 예측 논항이 ARG0으로 오인한 경우가 더욱 많았다.

평가 데이터 셋 내 부가 격 태그들은 필수 격 태그들에 비해 상대적으로 적은 빈도를 갖고 있음에도 불구하고 대부분 정확한 태그를 예측하였지만, O 태그로 잘못 예측하거나, O 태그를 논항으로 인식하는 경우가 빈번하게 존재하였다. 이러한 문제는 논항이 서술어와 의존적인 관계를 이루고 있고, 서술어에 따라 논항이 바뀌기 되어 대부분의 문장 내 토큰에는 O 태그가 할당되기 때문에 O 태그에 치중하여 학습되는 문제이다. 이러한 문제를 데이터 불균형(Data Imbalance) 문제라고 하며, 추후 데이터 불균형 문제 완화를 위한 학습 데이터 증폭 및 신경망 모델 개선 연구를 통해 이러한 문제를 완화하여 성능을 개선하고자 한다.

## V. 결론

의미역 결정 작업은 문장 내 핵심이 되는 성분들 간의 관계성을 파악하고 구조화하는 작업으로, 의미역에는 서술어와 행위자, 피행위자, 장소, 시간 등과 같이 서술어와 관련 있는 논항들로 구성되어진다.

최근 문장의 언어학적 특징 분석 및 추출을 위한 파이프라인 없이 주어진 문장만을 입력으로 사용한 End-to-End 신경망 모델을 이용한 연구들이 제안되고 있다. 기존 연구에서 제안한 LSTM 모델은 특정 시점에 대해 주변 시점 정보를 이용하지만 의미역 결정 작업은 문장 내 서술어와 논항이 가까운 위치뿐만 아니라 멀리 떨어져 있는 경우도 존재하기 때문에 본 논문에서는 문장 전체 정보에 집중할 수 있는 Biaffine Average Attention 구조를 갖고 있는 신경망 모델을 제안하였다.

제안한 신경망 모델의 성능을 평가하기 위해 F1 점수를 이용하여 기존 연구에서 제안한 모델들과 비교하였으며, 그 결과 제안한 모델의 성능이 76.21%로 BERT-CRF 모델과 BERT-BiLSTM-CRF 모델에 보다 각각 1.09%, 1.7%의 성능 개선이 있었음을 확인하였다.

**ACKNOWLEDGEMENT**

This paper was supported by Education and Research Promotion Program of KoreaTech.

**REFERENCES**

- [ 1 ] L. He, K. Lee, M. Lewis, and L. Zettlemoyer, "Deep Semantic Role Labeling: What Works and What's Next," in *Proceeding of the 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Canada, pp. 473-483, 2017.
- [ 2 ] L. He, K. Lee, O. Levy, and L. Zettlemoyer, "Jointly Predicting Predicates and Arguments in Neural Semantic Role Labeling," in *Proceeding of the 56th Annual Meeting of the Association for Computational Linguistics*, Melbourne, Australia, pp. 364-369, 2018.
- [ 3 ] J. S. Bae and C. K. Lee, "Korean Semantic Role Labeling using Stacked Bidirectional LSTM-CRFs," *Journal of KIISE*, vol. 44, no. 1, pp. 36-43, Jan. 2017.
- [ 4 ] K. H. Park and S. H. Na, "A Neural Attention model for Korean Semantic Role Labeling," in *Proceeding of the 2017 Korea Software Congress*, Busan, South Korea, pp. 512-514, 2019.
- [ 5 ] J. Devlin, M. W. Chang, K. Lee, K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceeding of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Minneapolis, US, pp. 4171-4186, 2019.
- [ 6 ] J. S. Bae, C. K. Lee, S. J. Lim and H. K. Kim, "Korean Semantic Role Labeling with BERT," in *Proceeding of the 2019 Korea Computer Congress*, Jeju, South Korea, pp. 512-514, 2019.
- [ 7 ] T. Dozat, C. D. Manning, "A Neural Attention model for Korean Semantic Role Labeling," in *Proceeding of the 5th International Conference on Learning Representations*, Busan, South Korea, pp. 512-514, 2019.
- [ 8 ] S. H. Na, J. R. Li, J. H. Shin and K. I. Kim, "Deep Biaffine Attention for Korean Dependency Parsing," in *Proceeding of the 2017 Korea Computer Congress*, Jeju, South Korea, pp. 584-586, 2017.
- [ 9 ] J. Yu, B. Bohnet and M. Poesio, "Named Entity Recognition as Dependency Parsing," in *Proceeding of the 58th Annual Meeting of the Association for Computational Linguistics*, Kuala Lumpur, Malaysia, pp. 6470-6476, 2020.
- [10] AI Hub Common sense AI dataset [Internet]. Available: <https://www.aihub.or.kr/>.
- [11] T. Y. Lin, P. Goyal, K. He and P. Dollar, "Focal Loss for Dense Object Detection," in *Proceeding of the 2017 IEEE International Conference on Computer Vision(ICCV)*, Venezia, Italiana, pp. 2980-2988, 2017.



**남충현(Chung-Hyeon Nam)**

2019년 한국기술교육대학교 컴퓨터공학부 공학사  
 2019년 ~ 현재 한국기술교육대학교 컴퓨터공학과 공학석사과정  
 ※ 관심분야 : Natural Language Processing, Natural Language Understanding, Information Extraction



**장경식(Kyung-Sik Jang)**

1987년 고려대학교 전자공학과 공학사  
 1989년 한국과학기술연구원 전기전자공학과 공학석사  
 1998년 동경공업대학 전기전자공학 공학박사  
 1999년 ~ 현재 한국기술교육대학교 컴퓨터공학부 교수  
 ※ 관심분야 : Embedded System, Natural Language Processing, Incremental Learning