

BiLSTM 모델과 형태소 자질을 이용한 서술어 인식 방법

남충현¹ · 장경식^{2*}

Predicate Recognition Method using BiLSTM Model and Morpheme Features

Chung-Hyeon Nam¹ · Kyung-Sik Jang^{2*}

¹Graduate Student, Department of Computer Engineering, Korea University of Technology and Education, Cheonan, 31253 Korea

^{2*}Professor, Department of Computer Engineering, Korea University of Technology and Education, Cheonan, 31253 Korea

요 약

정보 추출 및 질의응답 시스템 등 다양한 자연어 처리 분야에서 사용되는 의미역 결정은 주어진 문장과 서술어에 대해 서술어와 연관성 있는 논항들의 관계를 파악하는 작업이다. 입력으로 사용되는 서술어는 형태소 분석과 같은 어휘적 분석 결과를 이용하여 추출하지만, 한국어 특성상 문장의 의미에 따라 다양한 패턴을 가질 수 있기 때문에 모든 언어학적 패턴을 만들 수 없다는 문제점이 있다. 본 논문에서는 사전에 언어학적 패턴을 정의하지 않고 신경망 모델과 사전 학습된 임베딩 모델 및 형태소 자질을 추가한 한국어 서술어를 인식하는 방법을 제안한다. 실험은 모델의 변경 가능한 파라미터에 대한 성능 비교, 임베딩 모델과 형태소 자질의 사용 유무에 따른 성능 비교를 하였으며, 그 결과 제안한 신경망 모델이 92.63%의 성능을 보였음을 확인하였다.

ABSTRACT

Semantic role labeling task used in various natural language processing fields, such as information extraction and question answering systems, is the task of identifying the arguments for a given sentence and predicate. Predicate used as semantic role labeling input are extracted using lexical analysis results such as POS-tagging, but the problem is that predicate can't extract all linguistic patterns because predicate in Korean language has various patterns, depending on the meaning of sentence. In this paper, we propose a Korean predicate recognition method using neural network model with pre-trained embedding models and lexical features. The experiments compare the performance on the hyper parameters of models and with or without the use of embedding models and lexical features. As a result, we confirm that the performance of the proposed neural network model was 92.63%.

키워드 : 딥러닝, 자연어 처리, 의미역 결정, 서술어 인식

Keywords : Deep learning, Natural language processing, Semantic role labeling, Predicate recognition

Received 13 October 2021, Revised 13 October 2021, Accepted 28 October 2021

* 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.2022.26.1.24>

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 문장과 서술어(Predicate)가 주어졌을 때, 서술어와 연관된 논항(Argument)들을 파악하는 작업으로 지식 그래프 구축, 사용자 질의에 대한 이해 및 답변을 위한 시스템 등 다양한 자연어 처리 분야에서 사용되고 있다.

여기서 서술어는 대부분 형태소 분석 결과, 동사 단어로 구성되어 있지만, 한국어 특성상 문장의 구조 및 의미에 따라 형용사, 명사 등의 품사가 사용될 수 있다. 이처럼 높은 성능을 가진 한국어 서술어 인식 시스템은 다양한 문장 패턴들에 모두 대응할 수 있어야 한다. 하지만 수많은 패턴들을 모두 대응하기 위해 모든 규칙을 정의하는 것은 많은 시간과 인력이 필요하다는 문제점이 따른다.

본 논문에서는 최근 자연어 처리 분야에서 활발히 연구되고 있는 신경망 모델을 이용하여 다양한 서술어 패턴에 대해 학습할 수 있는 한국어 서술어 인식 방법을 제안한다. 제안하는 방법에서 사용한 신경망 모델은 양방향 LSTM(Long Short-Term Memory) 모델과 CRF(Conditional Random Field) 모델을 사용한 모델이다. 또한 임베딩 층에는 각 단어들과 주변 단어들 간의 관계를 고려할 수 있도록 대량의 문장을 이용하여 사전 학습한 임베딩 모델을 사용할 수 있으며, 신경망 모델의 입력 자료로 기존 연구에서 사용된 각 입력 단어에 해당하는 형태소 자질[1, 2]을 함께 사용할 수 있는 구조로 이루어져 있다.

실험은 F1 점수를 이용하여 성능 평가를 하였으며, BiLSTM-CRF 모델의 변경 가능한 파라미터인 층 개수, 은닉 크기, 임베딩 크기 변화에 대한 성능 비교와 임베딩 모델 유무 및 형태소 자질 사용 유무에 따른 성능 비교를 실험으로 진행하였다. 또한, 미등록 단어 비율을 측정하는 실험을 진행하여, 사전 학습한 임베딩 모델의 미등록 단어 문제가 모델의 성능에 미치는 결과를 확인하였다.

실험 결과, 사전 학습된 임베딩 모델을 사용하지 않고 형태소 자질만을 사용한 경우 약 92.73%의 성능으로, 약 91.79%의 성능을 보인 형태소 자질을 사용하지 않은 경우에 비해 상대적으로 높은 성능을 보였음을 확인하였다.

또한 평가 데이터 셋에 대해 사전 학습된 임베딩 모델을 사용하지 않았을 경우 사전의 크기가 67,852이고 미

등록 단어 비율이 약 3.00%인 반면, 30,186의 사전 크기를 가진 fasttext 모델[3]은 약 16.00%의 비율을 가졌음을 확인하였다. 또한 사전 크기를 제한한 상황에 대해 실험을 진행하였으며, 그 결과, Word2Vec 모델[4]을 이용한 경우 92.63%로 가장 높은 성능을 보였으며, 임베딩 모델을 사용하지 않은 경우는 91.13%의 성능을 보였음을 확인하였다.

실험을 통해 형태소 자질이 신경망 모델의 성능에 영향을 미친다는 것을 확인할 수 있었으며, 또한, 사전 학습된 임베딩 모델의 미등록 단어 문제를 해결한다면 사용하지 않은 경우보다 성능을 개선할 수 있음을 확인할 수 있었다.

II. 관련 연구

머신 러닝 알고리즘 중 서포트 벡터 머신(Support Vector Machine) 알고리즘을 이용하여 서술어와 논항을 분류한 기존 연구[5]에서는 주어진 문장과 서술어에 대해, 논항의 경계를 파악하고 논항이 어떤 범주에 속하는지 서포트 벡터 머신 알고리즘을 이용하였다. 이 연구에서는 사전에 형태소 분석을 통해 품사 중 동사를 서술어라 설정하였으며, 논항 분류는 서포트 벡터 머신 입력 자료로 구문 분석을 이용해 생성한 구문 트리를 사용하였다. 비슷한 방법으로 구문 트리를 이용하여 서술어를 인식한 연구[6]에서는 구문 분석 결과인 구문 트리에 대해 특정 패턴에 적합한 단어를 서술어로 인식하는 방법을 사용하였다.

다음으로, 의미론적 구문 분석(Semantic Dependency Parsing)을 위해 통사론적 해석 방법과 의미론적 해석 방법을 단계적으로 사용한 연구[7]에서는 서술어를 인식하는 방법을 문장 내 단어들이 각각 서술어인지 아닌지 분류하는 이진 분류(Binary-Class Classification) 작업과 인식된 서술어가 어떤 범주에 속하는지 분류하는 다중 분류(Multi-Class Classification) 작업을 이용하였으며, 두 분류 작업을 순차적으로 수행하기 위해 파이프라인을 구성하였다. 먼저 서술어인지 아닌지 분류하는 이진 분류 작업은 “단어”, “형태소”, “어휘”, “단어의 접두사” 등을 자료로 고려하여 각 자질들을 조합한 다양한 패턴들을 생성하였으며, 생성된 자질들은 “단어”, “형태소”, “단어 + 형태소”, “단어 + 접두사” 등으로 이

루어졌다.

두 번째로 다중 분류 작업은 이진 분류 작업과 비슷하게 문장을 이루고 있는 단어, 형태소들의 집합, 서술어의 위치에 따른 순서 단어, 형태소 집합 등을 이용하여 자질 패턴을 만들었다. 이렇게 만들어진 자질 패턴들을 추론 과정에서 정수 선형 계획법(Integer Linear Programming)을 활용하여 최종적으로 서술어를 인식하였다.

마지막으로 의미역 결정 파이프라인을 통해 서술어와 논항을 인식하는 연구[8]에서는 주어진 문장에 대해 형태소 분석, 구문 분석, 어휘적 분석 등 전처리 과정을 통해 서술어 인식을 인식하였다.

III. 서술어 인식 신경망 모델

제안하는 서술어 인식 신경망 모델은 입력 문장 $S = \{s_1, s_2, \dots, s_N\}$ 가 주어졌을 때, 각 입력 토큰에 해당하는 레이블 $Y = \{y_1, y_2, \dots, y_N\}$ 이 서술어인지 아닌지 예측하는 순차 레이블링(Sequence Labeling) 작업으로 수행할 수 있다.

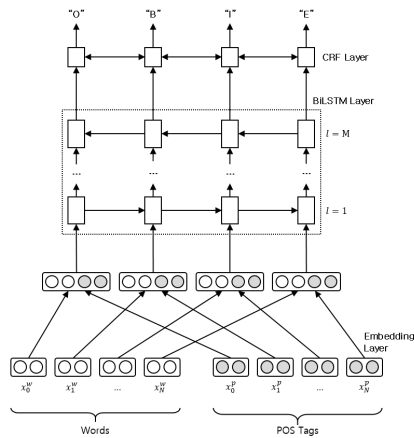


Fig. 1 Neural Network Model Architecture of Predicate Recognition

제안한 신경망 모델은 그림1과 같이 단어와 형태소 토큰을 입력하여 희소 벡터에서 밀집 벡터로 변환해주는 임베딩 층, 특정 시점에 대해 주변 시점에 대한 정보를 함께 고려할 수 있는 BiLSTM층과 CRF 층으로 구성

되어 있다.

3.1. Embedding Layer

신경망 모델은 사람의 개입 없이 언어학적 자질을 스스로 학습할 수 있다는 장점을 갖고 있기 때문에 입력 문장에 대해 사전 추출한 어휘 및 의미적 패턴을 추가로 입력하지 않아도 된다. 하지만 서술어의 특성상 동사, 형용사와 같이 특정 형태소에서 등장하는 경우가 많기 때문에 본 논문에서는 형태소 자질을 함께 입력할 수 있도록 신경망 구조를 설계하였다.

단어 자질 임베딩 층과 형태소 자질 임베딩 층은 각각 독립적으로 사용하였으며, BiLSTM층의 입력을 위해 두 개 임베딩 층의 각 출력 벡터들을 하나의 벡터로 합하였다.

$$\begin{aligned}
 e_i^w &= embed_{word}(x_i^w) \\
 e_i^p &= embed_{pos}(x_i^p) \\
 e_i &= e_i^w \oplus e_i^p
 \end{aligned} \tag{1}$$

수식1은 t 시점에서 단어 및 형태소 토큰인 x_i^w, x_i^p 가 임베딩 층을 통과한 후 결과 값이며, $embed_{word}(\cdot)$ 는 단어 임베딩 층을 의미한다. 또한, \oplus 는 결합 연산을 의미하고 사전 학습된 임베딩 모델로 변경하여 사용 가능하다.

3.2. BiLSTM-CRF Layer

BiLSTM 층은 과거, 미래 시점의 은닉 상태 벡터와 현재 시점의 단어, 형태소 자질이 결합된 벡터로부터 은닉 상태 벡터를 만드는 양방향 LSTM 층이 M 개로 이루어져 있다. 이러한 구조로 인해 BiLSTM 층은 현 시점의 단어에 대해 양방향으로 주변 단어들을 봄으로써 문장의 문맥을 파악할 수 있다는 장점을 갖는다. 다음으로 CRF 층은 순차 레이블링 작업에서 이전 레이블 정보를 고려하여 현재 시점의 레이블을 예측한다는 특징을 갖고 있어 각 시점에 대한 레이블 사이의 의존성을 고려할 수 있으며, 학습 과정에서 입력 토큰들에 대한 조건부 확률을 최대화 하는 방식의 학습 방법을 사용한다.

IV. 실험 결과 및 분석

4.1. 실험 환경

서술어 인식을 위한 신경망 모델 학습은 울산대학교 자연어 처리 연구실에서 공개한 의미역 결정 데이터 셋 [9]을 사용하였다. 공개된 의미역 결정 데이터 셋은 총 126,571개의 문장으로 이루어져 있으며, 학습, 검증, 평가 데이터는 8:1:1 비율로 무작위 분리하여 사용하였다.

서술어는 한 개 또는 그 이상의 토큰으로 이루어져 있기 때문에 BIOES 태깅 방식을 사용하였다. 입력 토큰의 개수는 한국어 특성상 동사는 문장의 마지막에 빈번하게 나타나기 때문에 긴 문장에 대응할 수 있는 256으로 설정하였으며, 설정한 크기보다 적은 경우 문장 뒤에 [PAD] 토큰을 부착하였다.

제안하는 서술어 인식 신경망 모델 성능 평가를 위해 변경 가능한 모델 파라미터에 따른 성능 평가, 사전 학습된 임베딩 모델 사용 유무와 형태소 자질 사용 유무에 따른 성능 평가로 총 2가지의 실험을 진행하였다.

4.2. 모델 파라미터에 따른 성능 평가

첫 번째 실험은 다양한 모델 파라미터 중 가장 높은 성능을 보이는 파라미터를 찾기 위해 다른 파라미터 변경 없이 임베딩 크기, 은닉 상태 크기, BiLSTM 층의 개수를 변화시키며 성능을 평가하였다.

Table. 1 F1 Score of each model parameters

Layers	Hidden	Embedding	Valid Scores(%)	Test Scores(%)
2	128	64	89.40	91.22
		200	89.74	91.46
		256	89.89	91.65
	256	64	89.41	91.34
		200	89.48	91.25
		256	89.44	91.29
3	128	64	89.13	91.03
		200	89.45	90.89
		256	89.20	91.13
	256	64	89.70	90.74
		200	90.04	91.84
		256	90.43	91.79

임베딩 크기는 64, 200, 256으로 하였으며, 128이 아닌 200을 실험 대상 파라미터로 선정한 이유는 사전 학습된 모델을 사용할 때 임베딩 크기가 200이기 때문에 동일한 크기로 설정하기 위해서이다. 또한 BiLSTM 층의 은닉 상태 크기는 128, 256으로 설정하였으며, 층 개

수는 2, 3개로 총 12개 경우의 수에 대한 실험을 진행하였다. 그 외, 파라미터 크기는 성능 변화가 크게 변하지 않기 때문에 실험 결과에 추가하지 않았다.

실험 결과는 표1과 같이 학습 과정에서 사용된 학습 데이터를 제외한 검증 데이터와 테스트 데이터로 평가를 하였다. 그 결과, 검증 데이터 셋에 대해 대부분 89%의 성능을 보였지만 은닉 상태 크기가 256이고, 임베딩 크기가 256일 때 90.43%로 가장 높은 성능을 보였다. 테스트 데이터 셋은 대부분 91%의 성능을 보였음을 확인하였으며, 파라미터는 3층, 은닉 상태 크기 256, 임베딩 크기 200으로 설정하였을 때 91.84%로 가장 높은 성능을 보였다.

4.3. 추가 모델 및 자질 사용 유무에 따른 성능 평가

다음 실험은 서술어 인식 모델의 파라미터를 직전 실험에서 가장 높은 성능을 보인 파라미터로 고정한 후, 임베딩 모델을 사용하지 않은 경우(None)와 사전 학습된 임베딩 모델인 Word2Vec, fasttext 모델을 각각 사용하는 경우와 형태소 자질 사용 유무에 대해 모델 성능 평가를 진행하였다. 여기서 Word2Vec 모델과 fasttext 모델은 공개된 사전 학습된 모델[10]을 사용하였으며, 형태소 자질 임베딩 크기는 32로 설정하였다.

Table. 2 F1 Score of each word embedding models

W. E. Models	Valid Scores(%)	Test Scores(%)
None	90.43	91.79
Word2Vec	82.01	85.76
fasttext	81.02	84.60

먼저 표2는 사전 학습된 임베딩 모델 사용 유무에 따른 성능 결과이다. 사전 학습된 임베딩 모델을 사용하지 않은 서술어 인식 모델의 경우 테스트 데이터 셋에 대해 약 91.79%의 성능을 보인 반면, Word2Vec, fasttext 모델을 사용한 경우 약 85.76%, 84.60%의 성능으로 약 6% 정도의 성능 저하가 있었음을 확인하였다.

표3은 형태소 자질을 함께 사용한 경우에 대한 성능 결과를 나타내는 표이다. 표2에서 보인 성능보다 사전 학습된 임베딩 모델과 형태소 자질을 함께 사용한 경우 약 7% 이상 성능이 향상되었다. 또한 사전 학습된 임베딩 모델을 사용하지 않은 경우에도 약 1%의 성능 향상을 보였음을 확인할 수 있었다.

Table. 3 F1 Score of POS-tag features

W. E. Models	Valid Scores(%)	Test Scores(%)
None	91.88	92.73
Word2Vec	91.51	92.63
fasttext	91.24	92.43

이를 통해 두 가지 실험을 통해 서술어 인식 모델을 학습하는 경우 형태소 자질이 성능 향상에 영향을 미친다는 것을 알 수 있었다. 하지만 사전 학습된 임베딩 모델만을 이용한 경우 성능이 더욱 떨어지는 문제점이 발생하였으며, 본 논문에서는 해당 실험에서 사용한 데이터 셋에 임베딩 모델의 단어 자질이 잘 반영되고 있는지 확인하기 위해 미등록 단어 비율을 확인하는 추가 실험을 실시하였다.

Table. 4 Vocabulary size and the out of vocabulary of each validation dataset and test dataset

W. E. Models	Vocab Size	V-O Rate(%)	T-O Rate(%)
None	67,852	1.85	3.00
Word2Vec	30,185	18.20	19.59
fasttext	30,186	15.28	16.00

표4는 검증 데이터 셋과 테스트 데이터 셋에 대해 사전 학습 모델의 유무에 따른 단어 사전 크기와 검증 데이터 셋에 대한 미등록 단어 빈도와 테스트 데이터 셋 미등록 단어 빈도를 측정된 결과이다. 형태소 분석 기준, 검증 데이터 셋과 테스트 데이터 셋은 총 215,933개, 371,668개의 단어를 포함하고 있었다.

실험 결과, 사전 임베딩 모델을 사용하지 않은 경우 학습 데이터 내 모든 단어를 사용하기 때문에 미등록 단어 비율이 약 1.85%, 3.00%이었다. 하지만 Word2Vec 모델의 경우 약 18.20%, 19.59%의 비율을 보였으며 fasttext 모델의 경우 약 15.28%, 16.00%의 비율을 보였다.

이를 통해, 사전 학습된 임베딩 모델을 사용한 경우에 비해 사용하지 않은 경우가 더 높은 성능을 보이는 현상은 미등록 단어 문제였음을 확인할 수 있었다.

다음으로 사전 임베딩 모델을 사용하지 않은 경우 사전의 크기를 사전 임베딩 모델과 동일하게 하였을 때 성능 차이에 대한 실험을 진행하였다.

실험을 위해 Word2Vec 모델과 fasttext 모델 중 미등록 단어 비율이 상대적으로 낮은 fasttext 모델과 동일한 단어 집합 크기인 30,186개에 대하여, 기존 학습 데이터로 만들어진 단어 집합의 크기인 67,852개에서 랜덤으

로 추출하여 동일한 크기를 제한하여 생성하였다.

Table. 5 F1 Score about vocabulary size and a reduced vocabulary size

Vocab Size	Valid Scores(%)	Test Scores(%)
67,852	91.88	92.73
30,186	89.66	91.13

그 결과, 67,852개 단어 집합 크기를 가진 모델의 성능이 각 데이터 셋에 대해 약 91.88%, 92.73%에 비해 89.66%, 91.13%로 성능이 낮아졌음을 확인할 수 있었고, 이 성능은 사전 학습된 임베딩 모델을 사용한 서술어 인식 모델의 성능에 비해서도 약 1% 이상의 성능 차이가 남을 확인할 수 있었다.

결론적으로 사전 학습된 임베딩 모델을 사용한 경우 미등록 단어 문제를 해결한다면 사용하지 않은 경우에 비해 성능이 높음을 알 수 있었으며, 추후 서술어 인식 모델의 성능 향상을 위해 사전 학습된 임베딩 모델의 미등록 단어 문제를 완화할 수 있는 연구를 진행하고자 한다.

V. 결 론

정보 추출, 질의응답 시스템 등 다양한 자연어 처리 분야에서 의미 분석 및 처리 과정으로 활용되어지는 의미역 결정 작업은 주어진 문장과 서술어를 입력으로 하여, 서술어와 연관성 있는 논항들의 관계를 파악하는 작업이다.

의미역 결정의 입력으로 사용되는 서술어는 동사 품사를 가진 단어가 지배적이지만, 한국어 특성상 상황에 따라 형용사, 명사 등 문장의 의미에 따라 다양한 패턴을 가질 수 있기 때문에 서술어에 대한 모든 언어학적 패턴을 정의하는 것은 많은 시간과 인력이 필요하다는 문제점이 있다.

본 논문에서는 한국어 서술어 인식을 위한 사전 임베딩 모델과 형태소 자질을 추가한 신경망 모델을 제안하였다. 실험은 F1 점수를 이용하여 성능 평가를 하였으며, 모델의 변경 가능한 파라미터에 대한 성능 비교, 임베딩 모델과 형태소 자질의 사용 유무에 따른 성능 비교로 총 2가지 실험과 미등록 단어 문제와 모델의 성능에 대한 실험을 추가적으로 진행하였다.

실험 결과, 테스트 데이터 셋에 대해 형태소 자질을

추가로 사용한 경우 92.73%의 성능으로 사용하지 않은 경우에 비해 상대적으로 높은 성능을 보였다. 또한 사전 학습된 임베딩 모델을 사용하지 않은 경우 단어 집합의 크기를 임베딩 모델과 동일하게 제한하였을 때, Word2Vec 모델을 이용한 신경망 모델이 92.63%로 가장 높은 성능을 보였으며, 이는 실제 시스템에서 특정 단어와 주변 단어 간의 관계를 학습한 사전 학습된 임베딩 모델을 사용한 서술어 인식 모델이 사용하지 않은 모델보다 상대적으로 성능이 높다고 할 수 있다.

현재는 사전 임베딩 모델을 사용하지 않은 경우가 가장 높은 성능을 보였지만 앞서 실험한 결과를 통해 제한된 사전 크기 내에서는 임베딩 모델을 사용하는 것이 성능을 향상시키는 데 영향을 미쳤음을 확인할 수 있었으며 추후 한국어 서술어 인식 모델의 성능을 향상시킬 수 있는 임베딩 모델에 대한 연구를 진행하고자 한다.

ACKNOWLEDGEMENT

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

REFERENCES

- [1] B. J. Kim, C. M. Park, Y. Y. Choi, M. J. Kwon, and J. Y. Seo, "Korean Named Entity Recognition using Joint Learning with Language Model," in *Proceeding of the 29st Conference on Human and Cognitive Language Technology*, Daeju, pp. 333-337, 2017.
- [2] M. S. Choi and B. W. On, "A Comparative Study on the Accuracy of Sentiment Analysis of Bi-LSTM Model by Morpheme Feature," in *Proceeding of the 2019 KIIT Conference*, Daejeon, pp. 307-309, 2019.
- [3] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135-146, Aug. 2017.
- [4] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," *arXiv:1301.3781v3*, 2013.
- [5] A. Moschitti and C. A. Bejan, "A Semantic Kernel for Predicate Argument Classification," in *Proceeding of the Eighth Conference on Computational Natural Language*

Learning, Boston, pp. 17-24, 2004.

- [6] Y. Miyao and J. Tsujii, "Deep linguistic analysis for the accurate identification of predicate-argument relations," in *Proceeding of the 20th international conference on Computational Linguistics*, Switzerland, pp. 1392-1398, 2004.
- [7] W. Che, Z. Li, Y. Hu, Y. Li, B. Qin, T. Liu, and S. Li, "A Cascaded Syntactic and Semantic Dependency Parsing System," in *Proceeding of the 12th Conference on Computational Natural Language Learning*, Menchester, pp. 238-242, 2008.
- [8] D. Larionov A. Shelmanov E. Chistova, and I. Smirnov, "Semantic Role Labeling with Pretrained Language Models for Known and Unknown Predicates," in *Proceeding of the Recent Advances in Natural Language Processing*, Varna, pp. 619-628, 2019.
- [9] Ulsan University Korean Language Processing Laboratory UCorpus-DP/SR [Internet]. Available: <http://nlplab.ulsan.ac.kr/>.
- [10] K. B. Park. wordvectors [Internet]. Available: <http://www.github.com/Kyubyong/wordvectors/>.



남충현(Chung-Hyeon Nam)

2019년 한국기술교육대학교 컴퓨터공학부 공학사
 2019년 ~ 현재 한국기술교육대학교 컴퓨터공학과 공학석사과정
 ※ 관심분야 : Natural Language Processing, Image Processing, Incremental Learning, Data Mining



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