

# 한국 전통문화 말뭉치구축 및 Bi-LSTM-CNN-CRF를 활용한 전통문화 개체명 인식 모델 개발

김경민<sup>1</sup>, 김규경<sup>1</sup>, 조재춘<sup>2</sup>, 임희석<sup>3\*</sup>

<sup>1</sup>고려대학교 컴퓨터학과 석사과정, <sup>2</sup>고려대학교 컴퓨터학과 연구교수, <sup>3</sup>고려대학교 컴퓨터학과 교수

## Constructing for Korean Traditional culture Corpus and Development of Named Entity Recognition Model using Bi-LSTM-CNN-CRFs

GyeongMin Kim<sup>1</sup>, Kuekyeng Kim<sup>1</sup>, Jaechoon Jo<sup>2</sup>, HeuiSeok Lim<sup>3\*</sup>

<sup>1</sup>Student, Department of Computer Science and Engineering, Korea University

<sup>2</sup>Research Professor, Department of Computer Science and Engineering, Korea University

<sup>3</sup>Professor, Department of Computer Science and Engineering, Korea University

요 약 개체명 인식(Named Entity Recognition)시스템은 문서로부터 고유한 의미를 가질 수 있는 인명(PS), 지명(LC), 기관명(OG) 등의 개체명을 추출하고 추출된 개체명의 범주를 결정하는 시스템이다. 최근 딥러닝 방식을 이용한 개체명 인식 연구에서 입력 데이터의 앞, 뒤 방향을 고려한 LSTM 기반의 Bi-LSTM 모델로부터 출력 데이터 간의 전이 확률을 이용한 CRF를 결합한 방식의 Bi-LSTM-CRF가 우수한 성능을 보이고, 문자 및 단어 단위의 효율적인 임베딩 벡터생성에 관한 연구와 CNN, LSTM을 활용한 모델에서도 좋은 성능을 보여주고 있다. 본 연구에서는 한국어 개체명 인식시스템 성능 향상을 위해 자질을 보강한 Bi-LSTM-CNN-CRF 모델에 관해 기술하고 전통문화 말뭉치구축 방식에 대해 제안한다. 그리고 구축한 말뭉치를 한국어 개체명 인식 성능 향상을 위한 자질 보강 모델 Bi-LSTM-CNN-CRF로 학습한 결과에 대해 제안한다.

주제어 : 개체명 인식, 전통문화, 말뭉치, 딥러닝, 자질 보강

**Abstract** Named Entity Recognition is a system that extracts entity names such as Persons(PS), Locations(LC), and Organizations(OG) that can have a unique meaning from a document and determines the categories of extracted entity names. Recently, Bi-LSTM-CRF, which is a combination of CRF using the transition probability between output data from LSTM-based Bi-LSTM model considering forward and backward directions of input data, showed excellent performance in the study of object name recognition using deep-learning, and it has a good performance on the efficient embedding vector creation by character and word unit and the model using CNN and LSTM. In this research, we describe the Bi-LSTM-CNN-CRF model that enhances the features of the Korean named entity recognition system and propose a method for constructing the traditional culture corpus. We also present the results of learning the constructed corpus with the feature augmentation model for the recognition of Korean object names.

**Key Words** : Named Entity Recognition, Traditional culture, Corpus, Deep Learning, feature augmentation

\*This research is supported by Ministry of Culture, Sport and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Culture Technology(CT) Research&Development Program 2017. (No. R2017030045 ).

\*Corresponding Author : HeuiSeok Lim (Limhseok@korea.ac.kr)

Received September 30, 2018

Revised December 4, 2018

Accepted November 20, 2018

Published December 28, 2018

## 1. 서론

지도학습(Supervised Learning) 기반의 모델 생성을 위해서는 학습 과정에서 대량의 학습 말뭉치가 필요하다. 학습 말뭉치는 특정 도메인에 한정된 것이 아닌 지도학습 기반 모델에서는 필수 요소로 작용한다. 비정형 데이터로부터 연구자가 원하는 형태의 자질추출을 위해 추출 가능한 비정형 데이터로부터 수작업으로 직접 자질 데이터를 구축해야 한다. 그러나 이와 같은 방법은 모델을 개발함에 있어서 상당한 시간과 비용을 요구한다. 최근 준지도(Semi-supervised), 비지도(Unsupervised) 학습에서 이와 관련된 많은 연구가 진행되고 있다[5-7]. 그러나 지도학습 기반의 대부분 딥러닝 시스템 개발을 위해 새로운 도메인 학습에 필요한 말뭉치는 존재하지 않거나, 혹은 필요에 따라 새로 구축해야 한다.

개체명(Named Entity)은 문장에서 인명(PS), 지명(LC), 장소(LC) 등 특정 개체에 대해 의미 있는 정보를 갖는 단어 또는 어구를 말한다. 개체명 인식(Named Entity Recognition, NER)은 이러한 문서 내의 문장으로부터 고유한 의미를 갖는 개체명을 추출하고, 추출된 개체명에 적합한 범주(label)를 결정한다. 이러한 개체명 인식 시스템은 딥러닝을 활용한 개체명 인식 연구에서 우수한 성능을 보여주고 있는데 특히, 양방향 순환신경망(Bi-directional Recurrent Neural Networks, Bi-RNN)과 조건부 랜덤 필드(Conditional Random Fields, CRF)를 결합한 방식의 모델에서 우수한 성능을 보여주고 있다.

Long Short Term Memory(LSTM) 기반의 모델은 각 단어를 표현하는 단어 임베딩 벡터(Word Embedding Vector)를 입력으로 받기 때문에 해당 문장을 표현하는 자질에 의존적일 수밖에 없다. 대부분 연구에서는 사전에 학습된(pre-trained) 단어 임베딩 벡터를 활용하거나 추가적으로 단어의 자질을 추출하여 활용하였다. 최근 문자 단위(character-level)의 임베딩 벡터로부터 단어 단위(word-level)의 임베딩 벡터를 유도해내는 방법들도 연구되고 있으며 이러한 연구에서 LSTM과 Convolutional Neural Network(CNN)을 활용한 모델에서 좋은 성능을 보여주고 있다[1-4]. 최근 한국어 특성을 고려한 한국어 개체명 인식 모델에서 한국어 형태소 단위의 자질 표현을 위해 사용한 워드 임베딩 방식 중 Glove방식을 이용하여 우수한 성능을 보인 연구가 있다[15].

본 연구에서는 전통문화에 적합한 말뭉치구축을 위해 한국학중앙연구원 디지털 인문학 웹사이트<sup>1)</sup>에서 기사를

추출하고, 추출된 데이터로부터 구축한 말뭉치에 대해 기술한다. 또한, 구축한 말뭉치를 Bi-LSTM-CNN-CRF 모델에 활용하여 한국어의 특성에 맞게 다양한 실험을 적용한 결과에 대해 기술한다. 논문의 구성은 다음과 같다. 2장은 관련 연구를 소개하고 3장에서는 Bi-directional LSTM Conditional Random Fields (Bi-LSTM-CNN-CRFs) 모델에 대해 기술한다. 4장에서는 말뭉치구축 방식에 대한 내용과 5장에서는 전통문화 말뭉치 활용 결과 그리고 마지막 6장에서는 결론을 기술한다.

## 2. 관련 연구

개체명 인식을 위한 전통적인 연구방법으로는 Hidden Markov Model(HMM), Support Vector Machine(SVM), Conditional Random Fields(CRFs) 등과 같이 규칙 기반 모델이나 통계 기반의 기계학습을 이용한 연구가 진행되었다. 최근 연구 동향은 딥러닝 기반의 RNN(Recurrent neural networks), LSTM 모델을 활용한 연구가 활발히 진행 중이다. LSTM 기반 모델인 Bi-LSTM-CRFs 모델은 문자열을 양방향으로 입력받아 출력 계층의 개체명 태그 간의 전이 확률을 이용한 모델로, 개체명 인식과 품사 태깅(Part-of-speech Tagging)과 같은 순서 레이블링(Sequence Labeling) 문제 해결에 있어서 우수한 성능을 보인다[7-9].

LSTM 모델은 각 단어를 표현할 수 있는 단어 임베딩 벡터값을 입력으로 받는데, 여기서 word2vec, glove, fastText와 같은 워드 임베딩 방법들을 이용하여 단어의 임베딩 공간(Embedding space)을 구성하고 문장의 자질을 표현한 연구가 있다[11-13]. LSTM 모델은 입력 데이터인 단어 임베딩 벡터에 의존적이기 때문에 효율적인 임베딩 벡터를 구상하기 위한 많은 연구가 진행되고 있다. 대량의 말뭉치를 이용하여 사전에 학습된(pre-trained) 단어 임베딩 벡터를 활용하고 이에 추가적으로 문장에서 문자 단위를 사용하여 CNN에 활용한 연구가 있다[1-4]. 최근 영어 개체명 인식에서 문자 단위 임베딩 벡터를 활용하여 단어 단위 임베딩 벡터로 유도한 방법을 한국어 개체명 인식에 적용하여 실험한 연구들이 한국어에서 좋은 성능을 보여주었다[14]. 그리고 LSTM과 CNN, 그리고 LSTM과 CNN을 결합하여 우수한 성능을 보인 연구들이 있다[2-4].

1) <http://dh.aks.ac.kr/>

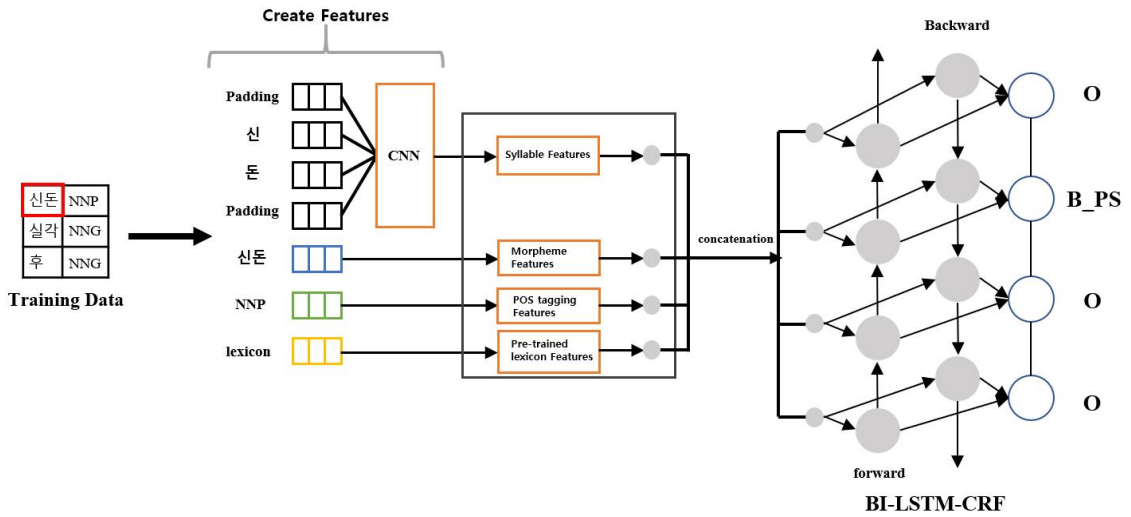


Fig. 1. System Architecture of Named Entity Recognition Using Bi-LSTM-CNN-CRF

### 3. 전통문화 개체명 인식 모델

본 논문에서 사용한 Bi-LSTM-CNN-CRF 모델은 Fig. 1에서 전통문화 개체명 인식 데이터를 예시로 활용하여 전체 구조도를 나타낸다. 해당 데이터는 한국학중앙연구원 디지털 인문학 웹사이트에서 크롤링 및 전처리 과정을 거친 데이터를 사용했으며 결과적으로 전체 문장 수 4702로부터 형태소 단위의 15만 개의 말뭉치 데이터를 구축했다.

모델은 “신돈 실각 후”라는 3개의 단어와 그에 해당하는 품사 태깅 결과를 입력 데이터로 받는다. 입력받은 데이터는 4가지 방법으로 문서를 표현하는데 첫 번째, 음절 단위의 자질 표현 구성을 위해 단어를 음절 단위 자질로 구성한 뒤 CNN을 통하여 음절 자질을 추출한 후 이를 음절 단위 자질 표현으로 활용한다. 두 번째, 형태소 단위의 데이터는 Glove 워드 임베딩 모델을 이용하여 워드 임베딩을 구성하고 이를 형태소 단위의 자질로 활용한다. 세 번째, 품사 태깅 결과를 품사 태깅에 대한 임베딩으로 구성하여 이를 자질로 활용한다. 네 번째, 구축된 사진을 활용하여 사진 자질을 표현한다. 학습 데이터로부터 표현한 각각의 자질(음절, 형태소, 품사, 사진)들을 연결(concatenation)하고 Bi-LSTM의 입력 데이터로 사용한다. LSTM 모델은 은닉 계층(hidden layer)을 통해 데이터를 계산하여 출력하고, 해당 데이터를 CRF의 입력으로 사용하여 전이 확률값을 계산한 후 최종적으로 입력값(형태소)에 대응하는 개체명을 예측한다.

#### 3.1 CNN(Convolutional Neural Network)

기존에 활용한 CNN은 주로 이미지 처리에 활용되었으나 최근에는 자연어처리와 관련된 문제 해결을 위한 방법으로 많은 연구가 진행되고 있다[1-4]. 문장을 구성하는 단어 단위의 자질을 추출하기 위해 양 끝을 padding 처리하여 처음부터 끝까지 K개의 convolution filter를 적용하고 max pooling을 취하여 단어를 표현하는 벡터 자질로 표현하는 방법이다. 또 다른 연구에서는 LSTM과 CNN으로 추출한 단어 임베딩 벡터를 결합하여 추가함으로써 좋은 성능을 보여주었다[2-4]. 본 논문의 모델에서는 자소 단위의 특징을 표현하였으며 Fig. 2는 자소 단위의 CNN을 활용하여 자질을 추출하는 과정을 나타낸다.

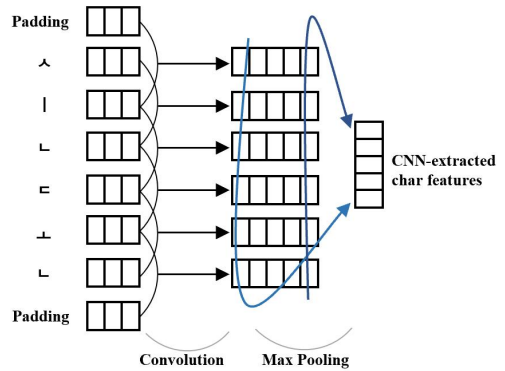


Fig. 2. CNN-extracted char features

### 3.2 Bi-LSTM-CRF

#### 3.2.1 Long Short Term Memory(LSTM)

순환 신경망 구조를 갖는 RNN은 그래디언트 소멸 문제(Vanishing Gradient Problem)가 발생한다는 단점이 있다. 이를 보완하고자 LSTM 등이 제안되었으며 자연어처리를 비롯한 음성 인식 외의 다양한 분야에서 사용되고 있다. LSTM은 아래의 식(1)과 같이 정의된다.

$$\begin{aligned}
 f_t &= \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \\
 i_t &= \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \\
 o_t &= \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \\
 g_t &= \tanh(W_{gx}x_t + W_{gh}h_{t-1} + b_g) \\
 c_t &= f_t \odot c_{t-1} + i_t \odot g_t \\
 h_t &= o_t \odot \text{hanh}(c_t)
 \end{aligned}
 \tag{1}$$

LSTM의 memory cell은 해당 데이터의 전체 상태를 기억하여 다음 단계(state)로 전달하는 역할을 한다. forget gate( $f_t$ )는 과거 cell의 정보 제거 여부를 나타내고 input gate( $i_t$ )는 입력 게이트로 cell에 입력된 현재 정보를 나타낸다.  $i_t \odot g_t$ 를 통해 현재 정보를 기억할지 계산하고 현재 cell 값과 출력값과의 계산을 통해 다음 hidden state( $h_t$ )에 전달한다.

#### 3.2.2 Bidirectional LSTM CRF(Bi-LSTM-CRF)

순서 레이블링(Sequence Labeling)에 LSTM을 적용할 경우 RNN의 그래디언트 소멸 문제를 어느 정도 해결할 수 있지만, 은닉 상태(Hidden state)를 고려해보았을 때 다음 단어가 이전 단어에 미치는 관계정보를 알 수 없으므로 관계정보를 추가로 고려하기 위해 Bi-LSTM을 사용한다. Bi-LSTM을 사용할 경우 전(forward), 후

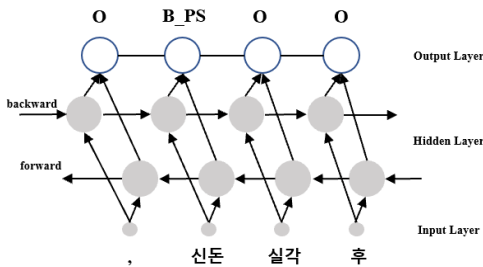


Fig. 3. Bi-LSTM-CRF

(backward) 관계에 있는 모든 정보의 관계성에 대해 파악할 수 있다. 또한, 문장 간의 의존성을 양방향에서 고려하여 그 인접한 태그 간의 의존성을 고려할 수 있다. Fig. 3 본 논문은 Bi-LSTM-CRF 모델을 이용하여 한국어 개체명 인식을 수행한다.

## 4. 말뭉치구축

한국학중앙연구원 디지털 인문학 웹사이트는 우리나라의 역사와 문화에 대한 정보를 데이터 자원으로 구축한 사이트이다. 말뭉치구축을 위해 해당 사이트의 백과사전 기사에서 전통문화와 관련된 기획기사 및 중심기사 5개로부터 각 기사의 개요와 내용에 대한 문장들을 크롤링하여 형태소 단위의 데이터를 생성하였다. Fig. 4

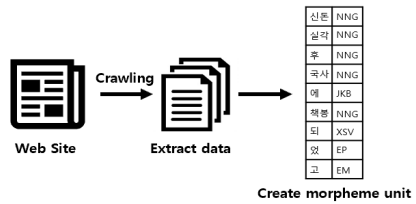


Fig. 4. Extract data and Create morpheme unit

데이터 구축을 위해 5명의 인원이 하루 평균 5시간씩 한 달 동안 진행했다. 문장 내의 태그 정보 추출을 수작업으로 진행하였으며 전체 2351개 기사로부터 4702개의 문장과 15만 형태소 단위의 데이터를 추출하였다.

Table 1. Category and Tag ratio among total corpus

Category	Count	Frequency
B_PS(Person)	4231	2.92%
B_DT(Date)	2399	2%
B_LC(Location)	2217	1.53%
B_OG(Organization)	740	0.51%
B_TI(Time)	53	0.04%
I (Tag I)	3765	2.6%

말뭉치구축을 위한 태깅 방식은 BIO (Begin, Inside, Outside)를 활용한다. 개체명 말뭉치의 각 카테고리에는 총 5개로, 인물(PS), 장소(LC), 기관(OG), 날짜(DT), 시간(TI)이다. 각 태그명 앞에는 'B'를 붙여 태그의 시작을 표기하고 연결된 어미는 'I'로 앞 단어와의 연결성을 나타

낸다. 그리고 문장 경계는 각 문장의 마지막 ‘과 문장 간 공백을 활용한다.

각 태그 중 인물(PS)이 가장 많이 태그되었으며 날짜(DT), 장소(LC) 순서로 태그되었다. I는 앞 단어가 B\_ 태그이면서 의미가 연결되는 단어를 태그한다.

### 5. 전통문화 말뭉치 활용 결과

전체 4702문장으로부터 구축한 15만 형태소 말뭉치를 Bi-LSTM-CNN-CRF 모델에 학습시킨 결과, 일반적인 문장뿐만 아니라 전통문화에 적합한 문장을 입력으로 넣어도 개체명을 정확하게 인식하는 것을 확인할 수 있다.

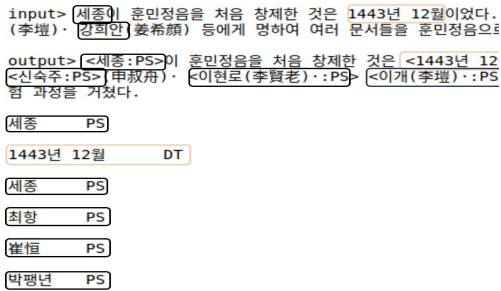


Fig. 5. Results using Traditional culture Corpus

Fig. 5에서 사용한 문장은 “세종이 훈민정음을 처음 창제한 것은... (이후 생략)” 이다. 모델 결과 ‘세종’, ‘최항’, ‘박팽년’ 등을 사람(PS) 개체명으로 인식하고 ‘1443년 12월’을 날짜(DT) 개체명으로 인식한다. 단어 간 의존도가 고려되어 최항(최항) 또한 사람 개체명으로 인식된 것을 확인할 수 있다.

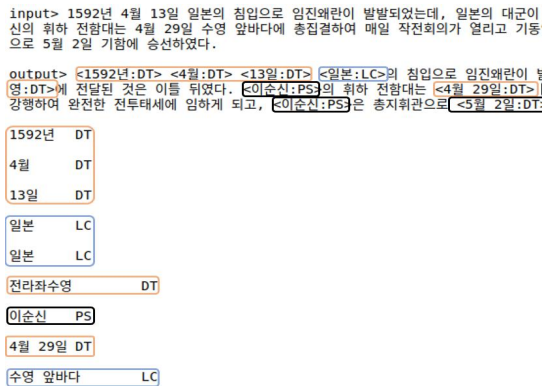


Fig. 6. Results using Traditional culture Corpus2

Fig. 6에서 사용한 문장은 “1592년 4월 13일 일본의 침입으로 임진왜란이 발발되었는데.. (이후 생략)” 이다. 모델 결과 ‘이순신’을 사람 개체명으로 인식하고 ‘1592년 4월 13일’, ‘전라좌수영’, ‘4월 29일’을 날짜 개체명으로, ‘일본’과 ‘수영 앞바다’를 장소 개체명으로 인식한다.

“개체명 인식 결과” “<세종:PS>이 훈민정음을 처음 창제한 것은 <1443년 12월:DT>이었다. 이후 <세종:PS>은 <최항:PS>·<박팽년:PS>·<이현로:PS>·<이개:PS>·<강희안:PS> 등에게 명하여 여러 문서들을 훈민정음으로 작성하는 등의 시정 과정을 거쳤다. <1446년 9월 29일:DT>, 마침내 훈민정음의 완성을 공식 발표하였다. <세종:PS>은 ‘용비어천가(龍飛御天歌)’를 훈민정음으로 번역하거나 ‘삼강행실도(三綱行實圖)’에 훈민정음으로 해설을 달아 발표하는 등 훈민정음 보급을 위해 여러 가지 노력을 기울였다. 또, <동국정운:OG>·<東國正韻:LC>·‘홍무정음역판’ 등을 편찬하여 훈민정음으로 한자의 정확한 발음을 정리하는 작업도 추진하였다.”

Fig. 7. Incorrect tagged NER system

Fig. 7은 기존의 다른 개체명 인식기 모델 시스템에 Fig. 5에서 사용한 문장과 동일한 문장을 입력으로 넣은 결과로 인물에 대해 잘못 예측하거나 예측하지 못하는 경우를 확인할 수 있다. 전통문화 말뭉치를 활용하여 학습한 모델은 새로운 전통문화 관련 단어를 예측하지만 대부분 기존 모델에서의 전통문화 키워드는 잘못 태그되거나 태그되지 않았다.

### 6. 결론

본 논문은 한국어 개체명 인식 시스템 성능 향상을 위해 제안한 Bi-LSTM-CNN-CRF 모델을 활용하여 전통문화 영역에서 활용 가능한 말뭉치를 구축하고 해당 모델에 적용하였다. 해당 모델에서는 음절 자질, 형태소 자질, 품사 태깅 자질, 구축된 사전 자질 보강을 통해 한국어 특성을 고려한 개체명 인식 성능을 향상시켰다. 전통문화라는 기존 개체명 인식 영역과 다른 영역의 시스템은 기존 개체명 인식 시스템에서 인식하지 못하는 부분이 발생했다. 따라서 한국학중앙연구원 디지털 인문학 웹사이트를 통해 전통문화 말뭉치를 구축하였고 구축한 4702문장의 15만 어절 형태소 말뭉치 데이터로 학습시킨 결과 해당 모델은 전통문화와 관련된 문장에서 대부분의 해당 개체명을 인식할 수 있었다.

### REFERENCES

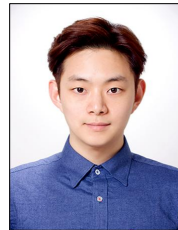
[1] Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). *Neural architectures for named entity recognition*. arXiv preprint arXiv:1603.01360.

- [2] Ma, X., & Hovy, E. (2016). *End-to-end sequence labeling via bi-directional lstm-cnns-crf*. arXiv preprint arXiv:1603.01354.
- [3] Ling, W., Trancoso, I., Dyer, C., & Black, A. W. (2015). *Character-based neural machine translation*. arXiv preprint arXiv:1511.04586.
- [4] Chiu, J. P., & Nichols, E. (2015). *Named entity recognition with bidirectional LSTM-CNNs*. arXiv preprint arXiv:1511.08308.
- [5] Nadeau D., Turney, P. D., & Matwin, S. (2006). Unsupervised named-entity recognition: Generating gazetteers and resolving ambiguity. *In Conference of the Canadian Society for Computational Studies of Intelligence* (pp. 266-277). Springer, Berlin, Heidelberg. DOI : 10.12811/JKCS.201.11.2.129
- [6] Zhu, X. (2006). Semi-supervised learning literature survey. *Computer Science, University of Wisconsin-Madison*, 2(3), 4. DOI : 10.22156/JKCS.2018.7.1.001
- [7] Derczynski, L., Maynard, D., Rizzo, G., van Erp, M., Gorrell, G., Troncy, R., ... & Bontcheva, K. (2015). *Analysis of named entity recognition and linking for tweets*. *Information Processing & Management*, 51(2), 32-49.
- [8] Graves, A., Mohamed, A. R., & Hinton, G. (2013, May). Speech recognition with deep recurrent neural networks. *In Acoustics, speech and signal processing (icassp), 2013 IEEE international conference on* (pp. 6645-6649). IEEE.
- [9] Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). *On the properties of neural machine translation: Encoder-decoder approaches*. arXiv preprint arXiv:1409.1259.
- [10] Santos, C. D., & Zadrozny, B. (2014). Learning character-level representations for part-of-speech tagging. *In Proceedings of the 31st International Conference on Machine Learning (ICML-14)* (pp. 1818-1826).
- [11] Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. *In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).
- [12] Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2016). *Enriching word vectors with subword information*. arXiv preprint arXiv:1607.04606.
- [13] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. arXiv preprint arXiv:1301.3781.
- [14] S. H. Na & M. W. Min. (2016). Character Based LSTM CRFs for Named Entity Recognition, *Korea Computer Congress (KCC)*.
- [15] D. Y. Lee, W. H. Yu, & H. S. Lim. (2017). Bi-directional

LSTM-CNN-CRF for Korean Named Entity Recognition System with Feature Augmentation. *Journal of the Korea Convergence Society[KCI]*, 8(12).

김 경 민(Kim, Gyeongmin)

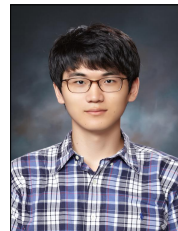
[학생회원]



- 2017년 8월 : 백석대학교 정보통신학부(이학사)
- 2018년 3월 ~ 현재 : 고려대학교 컴퓨터학과 석사과정
- 관심분야 : 딥 러닝, 자연어처리
- E-Mail : totoro4007@korea.ac.kr

김 규 경(Kim, Kuekyeng)

[학생회원]



- 2016년 2월 : 고려대학교 정보통신학부(이학사)
- 2018년 3월 ~ 현재 : 고려대학교 컴퓨터학과 석사과정
- 관심분야 : 딥 러닝, 이미지 처리 및 분석, 자연어 처리

· E-Mail : overmind22@korea.ac.kr

조 재 춘(Jo, Jaechoon)

[정회원]



- 2010년 2월 : 제주대학교 컴퓨터교육과(이학사)
- 2012년 2월 : 고려대학교 컴퓨터교육과(이학석사)
- 2018년 2월 : 고려대학교 컴퓨터과(공학박사)

· 2018년 3월 ~ 현재 : 고려대학교 정보대학 연구교수  
 · 관심분야 : AI in Education, NLP, Deep Learning, ITS  
 · E-Mail : jaechoon@korea.ac.kr

임 희 석(Lim, Heuseok)

[중신회원]



- 1992년 : 고려대학교 컴퓨터학과(이학학사)
- 1994년 : 고려대학교 컴퓨터학과(이학석사)
- 1997년 : 고려대학교 컴퓨터학과(이학박사)

· 2008 ~ 현재 : 고려대학교 컴퓨터학과 교수  
 · 관심분야 : 자연어처리, 뇌신경 언어 정보 처리  
 · E-Mail : limhseok@korea.ac.kr