

LIME 알고리즘을 이용한 한국어 감성 분류 모델 해석

남충현¹ · 장경식^{2*}

Korean Sentiment Model Interpretation using LIME Algorithm

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

한국어 감성 분류 작업은 챗봇, 사용자의 물건 구매 평 분석 등 실 서비스에서 사용되고 있으며, 현재 딥러닝 기술의 발달로 높은 성능을 가진 신경망 모델을 활발히 사용하여 감성 분류 작업을 수행하고 있다. 하지만 신경망 모델은 입력 문장이 어떤 단어들로 인해 결과가 예측되었는지 해석하는 것이 쉽지 않으며, 최근 신경망 모델의 해석을 위한 모델 해석 방법들이 활발히 제안되어지고 있다. 본 논문에서는 모델 해석 방법 중 LIME 알고리즘을 이용하여 한국어 감성 분류 데이터 셋으로 학습된 모델들의 입력 문장 내 단어들 중 어떤 단어가 결과에 영향을 미쳤는지 해석하고자 한다. 그 결과, 85.23%의 성능을 보인 양방향 순환 신경망 모델의 해석 결과, 총 25,283개의 긍정, 부정 단어를 포함했지만, 상대적으로 낮은 성능을 보인 84.20%의 Transformer 모델의 해석 결과, 총 26,447개의 긍정, 부정 단어가 포함되어 있어 양방향 순환 신경망 모델보다 Transformer 모델이 신뢰할 수 있는 모델임을 확인할 수 있었다.

ABSTRACT

Korean sentiment classification task is used in real-world services such as chatbots and analysis of user's purchase reviews. And due to the development of deep learning technology, neural network models with high performance are being applied. However, the neural network model is not easy to interpret what the input sentences are predicting due to which words, and recently, model interpretation methods for interpreting these neural network models have been popularly proposed. In this paper, we used the LIME algorithm among the model interpretation methods to interpret which of the words in the input sentences of the models learned with the Korean sentiment classification dataset. As a result, the interpretation of the Bi-LSTM model with 85.24% performance included 25,283 words, but 84.20% of the transformer model with relatively low performance showed that the transformer model was more reliable than the Bi-LSTM model because it contains 26,447 words.

키워드 : 딥러닝, 자연어 처리, 한국어 감성 분류, 모델 해석

Keywords : Deep learning, Natural language processing, Korean sentiment classification, Model interpretation

Received 7 September 2021, Revised 9 September 2021, Accepted 6 November 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.2021.25.12.1784>

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. 서 론

최근 한국어 감성 분류 작업은 딥러닝 기술의 발달로 챗봇, 사용자의 물건 구매 평 분석 등 실 서비스에서 활발히 사용되고 있다. 한국어 감성 분류 작업에 딥러닝 기술을 도입하는 경우, 정확도 평가를 통해 높은 성능을 가진 신경망 모델을 선정하여 사용하며, 정확도의 측정 은 신경망 모델의 예측 값과 사전 정의된 정답 값을 비교를 통해 이루어진다.

하지만 신경망 모델은 블랙박스 모델로 높은 정확도를 보이고 있더라도 정확도에 대한 신뢰하기 어렵다는 문제점이 있다. 즉, 입력 문장이 긍정 또는 부정인지 결과를 예측할 수 있지만, 어떤 단어들로 인해 그 결과가 예측되었는지 해석하는 것은 현재 불가능하다. 이는 결과 값이 중요한 영향을 미치는 실 서비스의 경우 신경망 모델을 신뢰하기 어려워진다.

앞서 언급한 문제점을 예시로 든다면, 부정 단어인 “우울”, “눈물”이라는 단어들 이 포함된 “나는 우울해서 눈물이 난다.”라는 입력 문장에 대해 신경망 모델이 부정 문장임을 예측하였을 때, 긍정, 부정과는 연관 없는 “나”라는 단어에 의해 부정 문장으로 예측되었을 수도 있다. 또한 학습 과정에서 “눈물” 단어가 “우울” 단어보다 학습 데이터에 상대적으로 많이 분포되어 있고, 대부분 “눈물” 단어는 긍정 문장, “우울” 단어는 부정 문장에 속해있다 가정할 때, 실제 부정 문장이라 할지라도 신경망 모델은 긍정 문장으로 예측할 수도 있다.

위 예시처럼 신경망 모델은 긍정, 부정과 연관 없는 단어들로 인해 결과를 예측하거나, 토큰들의 상대적인 빈도 차이로 인해 학습이 편파적으로 이루어져 잘못 예측하는 경우 등의 오류를 범할 수 있다. 이처럼 모델의 예측 값을 이해하지 못하고 사용한다면 신경망 모델에 대한 신뢰성이 낮아지는 문제가 되어, 상용 감성 분류 서비스에서 결과를 오측하는 경우가 발생할 수 있다.

이러한 문제를 완화하고 신경망 모델의 신뢰성을 확보하기 위해 정확도 지표 기반의 학습된 신경망 모델의 맹목적인 사용보다는 입력 문장 내 단어들 중 어떤 단어에 근거하였는지 모델을 해석하는 것이 필요하며, 이는 모델 해석 방법을 통해 신뢰성을 향상시킬 수 있다.

모델 해석을 위해 제안된 LIME(Local Interpretable Model-Agnostic Explanations) 알고리즘[1]은 입력 문장 내 단어들을 각각 다르게 조합한 샘플 단어 벡터에 따라

예측 값이 달라짐을 이용하여 모델을 해석하며, 해석 가능한 모델을 이용하여 데이터의 지역적인 영역을 해석하는 것을 목표로 하는 알고리즘이다.

본 논문에서는 LIME 알고리즘을 이용하여 기존 연구에서 제안된 한국어 감성 분류 신경망 모델들이 긍정 또는 부정 단어들로 인해 결과 값을 예측하였는지 실험하여 모델을 해석하고자 한다. 자세히는 먼저 순환 신경망(Recurrent Neural Network)[2, 3], 합성곱 신경망(Convolutional Neural Network)[4-6], Transformer 모델 [7-8]등 대표적인 신경망 모델들을 한국어 감성 분류 데이터 셋을 이용하여 학습한 후 정확도를 측정한다. 다음으로 LIME 알고리즘을 이용하여 추출된 긍정 또는 부정 문장으로 예측하는 데 있어 영향을 미친 입력 문장 내 단어들과 사전 구축된 한국어 감성 사전을 비교하여 모델이 추론 과정에서 긍정, 부정 문장을 이루는 단어들을 예측 결과에 실제로 반영하였는지 실험하고자 한다.

그 결과 1층 양방향 순환 신경망 모델이 85.23%로 가장 높은 성능을 보였지만, 상대적으로 낮은 성능인 84.20%의 성능을 보인 6층 Transformer 모델이 샘플링된 56,849개의 긍정 단어와 49,322개의 부정 단어 중 12,143개의 긍정 단어와 14,304개의 부정 단어가 한국어 감성 사전에 속해 있었으며, 이는 샘플링된 57,191개의 긍정 단어와 53,481개의 부정 단어 중 11,273개, 14,010개의 긍정, 부정 단어를 보인 1층 양방향 순환 신경망 모델보다 상대적으로 더 많았음을 확인할 수 있었고, 문장의 긍정, 부정적인 성격에 영향을 미친 체언과 용언 빈도수 또한 순환 신경망 모델보다 Transformer 모델이 더 많았음을 확인할 수 있었다.

II. 한국어 감성 분류 신경망 모델

2.1. 순환 신경망 모델

순환 신경망 모델은 입력 문장 $X = \{x_1, x_2, \dots, x_n\}$ 과 동일한 길이를 가진 출력 $Y = \{y_1, y_2, \dots, y_n\}$ 이 주어졌을 때, 수식1과 같이 이전 시점 $t-1$ 에 대한 입력 x_{t-1} 로부터 연산된 은닉 상태 h_{t-1} 와 현재 시점 t 에 대한 입력 x_t 을 이용하여 현재 은닉 상태 h_t 을 계산하는 재귀적 구조를 가지고 있다. 여기서 W_{hh} , W_{xh} 는 학습 가능한 파라미터, σ 는 활성화 함수를 의미한다.

$$h_t = \sigma(W_{hh}h_{t-1} + W_{xh}x_t) \quad (1)$$

기존 연구에서는 기울기 소실 문제를 완화하기 위해 현재 은닉 상태를 바로 다음 시점 이후의 은닉 상태의 입력으로 연결하여 하나 이상의 은닉 상태를 건너뛰는 Skip-Connected LSTM 모델을 적용한 연구[2]와 단방향 순환 신경망 모델은 학습 과정에서 현재 시점이 직전 과거 시점에 대해 수렴하는 경향을 보이며, 과거 시점에 대해서 현재 시점을 학습하기 때문에 역방향성을 가진 순환 신경망 모델을 추가적으로 적용하여 입력 문장에 대한 문맥을 이해할 수 있는 양방향 LSTM 모델[3]이 제안되었다.

2.2. 합성 곱 신경망 모델

영상 처리 분야에서 주로 사용되는 합성 곱 신경망 모델은 이미지 주변 픽셀 값을 고려한 지역적 학습을 위해 학습 가능한 필터를 이용하여 이미지 내 특징을 잘 반영할 수 있도록 학습되어진다. 합성 곱 신경망 모델은 지역적 학습에 용이하다는 장점을 갖고 있기 때문에 자연어 처리 분야에서도 입력 문장 내 주변 단어 간의 지역적인 정보를 학습할 때 활발히 사용된다[4-6].

합성 곱 신경망 모델의 학습을 위해 필요한 필터는 임베딩 크기인 e 로 이루어진 단어 표현 벡터 상에서 n 개의 단어를 동시에 학습하기 위한 $n \times e$ 크기의 필터를 사용한다.

기존 한국어 감성 분류 작업 연구에서는 합성 곱 신경망 모델의 입력으로 언어학적 자질을 이용한 다중 채널 구조가 제안되었으며[5], 자세히는 형태소, 음절, 자소 단위의 입력을 각각의 채널로써 합성 곱 신경망 모델의 입력으로 사용하여 입력 문장에 대한 표현을 확장시켰다. 또한 합성 곱 신경망 모델의 단어 임베딩 층 대신 Word2Vec와 같은 임베딩 모델[6]을 적용한 연구도 제안되었다.

2.3. Transformer 모델

Transformer 모델[7]은 인코더와 디코더 구조 모두 수식2와 같이 멀티 헤드 자가 주의집중 메커니즘(Multi Head Self-Attention Mechanism)을 이용하였으며, 입력 문장 내 하나의 단어에 대해 중요한 단어를 중점적으로 학습하여 직전 과거 시점뿐만 아니라 문장 전역 범위 내 중요한 단어들을 중점적으로 학습할 수 있어 직전 시점

에 대한 의존성을 제거할 수 있다는 장점을 갖고 있다.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

주어진 입력 문장 내에 포함된 단어들 중 Q 는 집중하고자하는 특정 단어에 대한 벡터이며, K 와 V 는 특정 단어 주위의 주변 단어에 대한 벡터이다. 각 단어 벡터들은 내적 연산으로 계산되어지며, $\sqrt{d_k}$ 는 내적 결과 값이 커진 경우 학습이 잘 되지 않을 수 있기 때문에 값을 낮추기 위해 사용된다.

Transformer 모델을 이용한 한국어 감성 분류 연구에는 워드 임베딩 방식과 합성 곱 신경망 모델과 순환 신경망 모델을 함께 사용한 연구가 제안되었다[8].

III. LIME 알고리즘

최근 제안된 신경망 모델은 각 층에 속한 노드들이 복잡하게 연결된 매우 깊은 다층 구조로 되어 있기 때문에 블랙박스 모델이며, 이러한 모델은 입력 데이터 중 어떤 요소가 결과에 큰 영향을 미쳤는지 직접적인 해석이 불가능하다.

복잡한 신경망 모델 해석을 위해 제안된 방법 중 하나인 대리 분석(Surrogate Analysis) 방법은 신경망 모델의 학습 과정이나 방법을 수정하지 않고 해석 가능한 입력과 해석 가능한 간단한 구조의 대리 모델(Surrogate Model)의 출력 값을 이용해 해석하는 방법이다.

LIME 알고리즘은 대리 분석 해석 방법 중 대표적인 알고리즘으로, 전체 데이터에 대한 해석이 아닌 각 데이터에 대해 사람이 해석 가능한 표현(Interpretable Representation)으로 재구성하여 해석하는 국소적 대리 분석(local Surrogate Analysis) 방법을 수행하는 알고리즘이다.

LIME 알고리즘은 입력 문장 x 이 주어졌을 때, 해석 가능한 모델들의 집합 G 중 신경망 모델 f 과 가장 근사한 결과를 예측하는 해석 가능한 모델 $g \in G$ 을 찾는 것을 목표로 한다.

이 알고리즘을 수행하기 위해 먼저 해석 가능한 표현 벡터 x' 과 이를 이용한 샘플 데이터 z' 을 만든다. 여기서 해석 가능한 표현은 입력 문장 x 가 있다고 가정할 때, 각 토큰들에 대한 이진 벡터 x' 을 의미한다. 예를 들어

“이번 영화 화이팅”이라는 입력 문장이 주어지고 띄어 쓰기 기준으로 문장을 나눈다면, “이번”, “영화”, “화이팅” 토큰들에 대해 해석 가능한 표현은 이진 벡터 $x' = \{1, 1, 1\}$ 가 되며, 샘플 데이터 z' 는 모두 0이 아닌 경우를 제외한 $\{0, 0, 1\}$, $\{0, 1, 1\}$ 등과 같은 벡터로 표현이 가능하다.

다음으로 샘플 데이터 z' 와 이 샘플 데이터를 해석하고자 하는 모델의 입력으로 사용할 수 있도록 기존 데이터 형식으로 변환한 데이터 z 을 각각 해석 가능한 모델 g 과 신경망 모델 f 의 입력으로 사용하여 각각의 예측 값을 얻은 후, 이 예측 값들의 차이를 이용하여 가장 모델 f 와 근사한 모델 g 을 찾는다.

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}} L(f, g, \pi_x) + \Omega(g) \quad (3)$$

수식3은 손실 함수 L 과 각 해석 가능한 모델 g 에 대한 복잡도 $\Omega(g)$ 을 이용하여 신경망 모델 f 의 결과와 가장 근사한 결과를 내는 해석 가능한 모델 g 을 찾는 수식이며, π_x 는 설명하고자 하는 데이터 x 와 샘플링 된 데이터 z 간의 유사도이다.

IV. 실험 및 결과

4.1. 실험 환경

모델 학습을 위해 본 논문에서는 PyTorch[9] 라이브러리를 사용하였다. 실험을 위한 신경망 모델은 순환 신경망 모델, 합성곱 신경망 모델, Transformer 모델을 사용하였으며, 순환 신경망 모델은 1, 2개의 층으로, Transformer 모델은 인코더와 디코더 블록을 2, 6개의 층으로 구분지어 실험하였다. 순환 신경망 모델의 은닉 상태 크기는 256으로 설정하였으며, 합성곱 신경망 모델은 윈도우 크기가 3, 4, 5인 3개의 필터를 사용하였다. 마지막으로 Transformer 모델은 기존 연구에서 사용된 파라미터와 동일하게 설정하여 학습을 진행하였다. 학습은 총 20회 진행되었으며, 실험에 사용된 신경망 모델 모두 Adam Optimizer 최적화 알고리즘을 이용하였다. 학습을 위해 형태소 분석 전처리를 수행하였으며 입력 문장 내 모든 단어들에 대해 LIME 알고리즘을 적용하고자 추가적인 불용어 제거나 정규화 전처리는 하지 않았다.

실험에 필요한 한국어 감성 분류 데이터 셋은 네이버

영화평 감성 분류 데이터 셋[10]을 이용하였으며, 학습 데이터와 평가 데이터는 각각 레이블링이 되어 있지 않은 데이터를 제외한 149995, 49997개의 데이터를 사용하였다.

입력 문장 내 예측에 영향을 미치는 단어들에 대해 실험하기 위해 LIME 알고리즘의 샘플 데이터 z' 의 개수를 학습 데이터와 평가 데이터의 평균 토큰 배수 곱을 취한 200개로 설정하였으며, 영향을 미친 주요 단어들을 각각 3개씩 추출하였다.

LIME 알고리즘을 통해 추출된 단어들이 긍정 또는 부정 단어로서 모델에 영향을 주었는지 확인하기 위해 n-gram으로 이루어진 어절과 축약어 등의 긍정, 부정 단어들이 포함되어있는 한국어 감성 사전[11]을 이용하였다. 한국어 감성 사전은 단어뿐만 아니라 어절로도 이루어져 있기 때문에 학습 과정에서 이루어진 전처리 과정과 동일하게 형태소 분석 작업을 사전에 처리하였으며, 형태소 분석된 긍정, 부정 단어들의 집합에 대해 각각 집합 내 공통적으로 출현한 단어들을 제거하여, 총 긍정 단어 1,416개, 부정 단어 3,189개를 실험에 사용하였다.

4.2. 신경망 모델 학습 결과

표1은 총 7개의 학습된 모델들의 정확도와 모델 내 매개 변수 개수를 나타낸 표이며, 실험을 통해 정확도는 순환 신경망 모델이 합성곱 신경망 모델이나 Transformer 모델보다 상대적으로 높은 성능을 보였으며, 1층을 가진 양방향 순환 신경망 모델과 합성곱 신경망 모델이 각각 85.23%, 83.71%로 가장 높은 성능과 가장 낮은 성능을 보였음을 확인하였다.

모델 내 매개 변수는 6층 Transformer 모델이 73,885K로 가장 많았으며, 합성곱 신경망 모델이 13,865K 개로 가장 적었다. 2층 구조를 가진 순환 신경망 모델과 1층 구조를 가진 양방향 순환 신경망 모델은 14,527K 개로 비슷한 매개 변수 개수를 갖고 있는데, 이는 2개의 순환 신경망 모델을 수평 또는 수직 배치를 했기 때문이다.

Table. 1 The Accuracy of Each Neural Network Models for Korean Sentiment Classification

Models	Accuracy(%)	# of Params(K)
CNN	83.71	13,865
RNN(layer=1)	84.96	14,132
RNN(layer=2)	85.01	14,527
Bi-RNN(Layer=1)	85.23	14,527

Models	Accuracy(%)	# of Params(K)
Bi-RNN(layer=2)	85.17	15,710
Transformer(layer=2)	83.87	61,261
Transformer(layer=6)	84.20	73,885

4.3. 한국어 감성 단어 적용 결과

표2는 LIME 알고리즘을 적용하여 각 모델들의 예측 값에 영향을 미친 단어들이 한국어 감성 사전에 속한 단어인지 확인한 결과이며, 각 모델들로부터 샘플링 된 긍정 단어(Sampled Positive Word)와 부정 단어(Sampled Negative Word)와 샘플링 된 단어들이 얼마나 한국어 감성 사전 내에 포함되어 있는 긍정 단어(Positive Word), 부정 단어(Negative Word)에 포함되어 있는지 개수를 나타낸다. 이를 통해 정확도는 순환 신경망 모델이 Transformer 모델보다 더 높지만 중립적인 단어보다 긍정 또는 부정 단어가 예측 결과에 상대적으로 크게 영향을 미치지 못하는 실험을 통해 확인하였다.

Table. 2 The Number of Words at Each Models using LIME Algorithm

Models	SPW	PW	SNW	NW
CNN	56,788	10,921	49,647	13,186
RNN(layer=1)	57,723	11,669	53,143	13,928
RNN(layer=2)	57,254	11,361	53,366	14,086
Bi-RNN(Layer=1)	57,191	11,273	53,481	14,010
Bi-RNN(layer=2)	57,182	11,424	53,971	13,956
Transformer(layer=2)	55,321	11,683	50,257	14,517
Transformer(layer=6)	56,849	12,143	49,322	14,304

다음으로 신경망 모델의 학습과 LIME 알고리즘을 수행하기 전, 전처리 과정으로 형태소 분석을 하였기 때문에 조사와 같은 중립 단어들과 형태학적 자질 중 문장의 긍정, 부정적인 성격에 영향을 미치는 명사, 대명사 등을 포함한 체언과 동사, 형용사 등을 포함한 용언, 그 외 품사들에 대한 빈도를 측정하였다.

Table. 3 The Number of Words at Each Models using Morpheme Analysis

Models	PN	PV	PE	NN	NV	NE
CNN	39,821	17,305	24,453	33,903	12,130	35,546
RNN(layer=1)	39,175	16,783	25,621	37,855	11,911	31,813
RNN(layer=2)	38,871	16,267	26,441	38,961	12,260	30,358
Bi-RNN(Layer=1)	39,617	16,842	25,120	38,276	12,224	31,079
Bi-RNN(layer=2)	39,464	16,546	25,569	28,411	13,002	30,166

Models	PN	PV	PE	NN	NV	NE
Transformer(layer=2)	41,559	17,078	20,419	36,218	12,620	19,751
Transformer(layer=6)	42,579	17,662	21,707	34,062	12,921	18,685

표3은 각 신경망 모델들에 LIME 알고리즘을 이용하여 추출된 단어들이 한국어 감성 사전 내 포함되어 있는 경우 각 단어들의 형태학적 자질에 대해 긍정 체언(PN), 부정 체언(NN), 긍정 용언(PV), 부정 용언(NV), 체언과 용언에 속하지 않는 긍정 단어(PE), 부정 단어(NE)로 나누어 분류하여 빈도를 측정 한 결과이다. 모델 중 문장의 긍정 또는 부정적인 성격에 영향을 주는 체언과 용언이 가장 많이 추출된 모델은 6층 Transformer 모델이며, 가장 낮은 모델은 1층 순환 신경망 모델을 실험을 통해 알 수 있었다.

실험을 통해 2층 양방향 순환 신경망이 비교 모델 중 가장 높은 성능을 보였으며, 상대적으로 6층 Transformer 모델의 성능이 낮았지만, LIME 알고리즘을 이용하여 문장에 영향을 미친 단어들을 추출하여 한국어 감성 사전 내 긍정 또는 부정 단어로 속해 있는지 확인한 결과 6층 Transformer 모델이 다른 비교 모델들에 비해 학습 과정에서 문장의 긍정, 부정적인 성격을 가지게 만드는 단어들을 학습하여 결과 값을 예측할 수 있었다.

V. 결론

한국어 감성 분류 작업은 챗봇, 사용자의 물건 구매 평 분석 등 다양한 곳에서 활용되고 있으며, 최근 이 작업에 신경망 모델을 적극적으로 적용하고 있다.

하지만, 기존 연구들은 높은 정확도를 가질 수 있게 순환 신경망 모델, 합성곱 신경망 모델, Transformer 모델들의 구조를 변형시키거나 층을 깊게 쌓는 등의 다양한 변형 방법이 제안되었지만, 입력 문장 내 어떤 단어를 근거로 하여 결과를 예측하였는지 해석할 수 없기 때문에 실 서비스에 적용 시 정확도에 대한 신뢰를 할 수 없다.

본 논문에서는 한국어 감성 분류 모델들을 학습하여 정확도를 측정하고, LIME 알고리즘을 이용하여 예측에 영향을 미친 입력 문장 내 단어들을 추출한 후, 한국어 감성 사전 내 속해 있는 단어인지를 확인하여 각 모델들을 해석하는 실험을 진행하였다.

실험을 통해 1층 양방향 순환 신경망 모델이 85.23%로 가장 높은 정확도를 보였으며, 합성곱 신경망 모델이 가장 낮은 성능인 83.71%를 보였다. 하지만 1층 양방향 순환 신경망 모델보다 상대적으로 낮은 성능인 84.20%를 보인 6층 Transformer 모델에 LIME 알고리즘을 적용한 결과, 긍정 또는 부정 단어들인 한국어 감성 사전에 포함된 빈도수는 샘플링 된 56,849개의 긍정 단어와 49,322개의 부정 단어 중 12,143개의 긍정 단어와 14,304개의 부정 단어가 한국어 감성 사전에 속해 있었다. 이는 샘플링 된 57,191개의 긍정 단어와 53,481개의 부정 단어 중 11,273개, 14,010개의 긍정, 부정 단어를 보인 1층 양방향 순환 신경망 모델보다 상대적으로 더 많았음을 확인할 수 있었다.

ACKNOWLEDGEMENT

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

REFERENCES

- [1] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You?: Explaining the Predictions of Any Classifier," in *Proceeding of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, pp. 1135-1144, 2016.
- [2] J. S. Bae and C. K. Lee, "Sentiment Analysis with Skip-Connected LSTM," in *Proceeding of the 2017 Conference on Korea Software Congress*, PyeongChang, pp. 633-635, 2017.
- [3] E. J. You, J. H. Lee, and S. Y. Park, "The Sentiment Classification of News Articles using LSTM," in *Proceeding of the 2018 Conference on Korea Software Congress*, PyeongChang, pp. 1949-1951, 2018.
- [4] Y. Kim, "Convolutional Neural Networks for Sentence Classification," in *Proceeding of the 2014 Conference on Empirical Methods in Natural Language Processing*, Doha, pp. 1746-1751, 2014.
- [5] W. W. Kim and K. H. Park, "Design of Korean Text Emotion Classifier Using Convolution Neural Network," in *Proceeding of the 2017 Conference on Korea Computer Congress*, Jeju, pp. 642-644, 2017.
- [6] M. Kim, J. H. Byun, C. H. Lee, and Y. S. Lee, "Multi-channel CNN for Korean Sentiment Analysis," in *Proceeding of the 2018 Conference on Human and Language Technology*, Seoul, pp. 79-83, 2018.
- [7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," in *Proceeding of the 31st Conference on Neural Information Processing Systems*, Long Beach, pp. 5998-6008, 2017.
- [8] C. E. Park and C. K. Lee, "Sentimental Analysis of Korean Movie Review using Variational Inference and RNN based on BERT," *The KIISE Transactions on Computing Practices*, vol. 25, no. 11, pp. 552-558, Aug. 2019.
- [9] A. Paszke, PyTorch [Internet]. Available: <https://pytorch.org>.
- [10] E. J. Park, Naver Korean Sentiment Classification Dataset [Internet]. Available: <https://github.com/e9t/nsmc>.
- [11] B. W. On, S. M. Park, and C. W. Na. KnuSentLex [Internet]. Available: <https://github.com/park1200656/KnuSentiLex>.



남충현(Chung-Hyeon Nam)

2019년 한국기술교육대학교 컴퓨터공학부 공학사
2019년 ~ 현재 한국기술교육대학교 컴퓨터공학과
공학석사과정

※관심분야: Natural Language Processing,
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