

## 한글 음소 단위 딥러닝 모형을 이용한 감성분석

이재준\* · 권순범\*\* · 안성만\*\*\*

### Sentiment Analysis Using Deep Learning Model based on Phoneme-level Korean

Jae Jun Lee\* · Suhn Beom Kwon\*\* · Sung Mahn Ahn\*\*\*\*

#### ■ Abstract ■

Sentiment analysis is a technique of text mining that extracts feelings of the person who wrote the sentence like movie review. The preliminary researches of sentiment analysis identify sentiments by using the dictionary which contains negative and positive words collected in advance. As researches on deep learning are actively carried out, sentiment analysis using deep learning model with morpheme or word unit has been done. However, this model has disadvantages in that the word dictionary varies according to the domain and the number of morphemes or words gets relatively larger than that of phonemes. Therefore, the size of the dictionary becomes large and the complexity of the model increases accordingly.

We construct a sentiment analysis model using recurrent neural network by dividing input data into phoneme-level which is smaller than morpheme-level. To verify the performance, we use 30,000 movie reviews from the Korean biggest portal, Naver. Morpheme-level sentiment analysis model is also implemented and compared. As a result, the phoneme-level sentiment analysis model is superior to that of the morpheme-level, and in particular, the phoneme-level model using LSTM performs better than that of using GRU model. It is expected that Korean text processing based on a phoneme-level model can be applied to various text mining and language models.

Keyword : Sentiment Analysis, Deep Learning, Sequential Model, Phoneme Unit, LSTM, GRU

## 1. 서 론

다른 사람들의 생각은 의사결정에서 항상 중요한 역할을 한다. 인터넷이 없던 시절에는 주변 사람들의 의견을 참고하여 의사결정을 하였지만 인터넷 등장 이후 블로그나 구매후기 사이트에서 다양한 다른 사람의 의견을 손쉽게 접할 수 있게 되었다(Pang and Lillian, 2008). 특히 ‘후기’는 다양한 형태와 분야로 온라인상에 존재하며 미국의 한 조사에 따르면, 87% 성인들은 다른 사용자들의 후기가 구매에 영향을 미쳤다(Dave et al., 2003). 국내 시장조사전문기관 엠브레인 트렌드모니터에 의하면 광고보다 온라인 후기를 신뢰한다는 응답이 무려 73.8%로 나타났으며 실제로 상품을 구매하기 전 다른 사람들이 작성한 후기를 항상 챙겨보는 소비자가 79%에 이른다. 또한 이러한 후기는 소비자들의 주관적인 선호를 담고 있으며 실질적 구매결정에 도움을 주는 요소로 작용하고 있다(Zhu and Zhang, 2010).

이에 언어에 담겨져 있는 의견 및 태도와 같은 주관적 정보를 탐지하고 추출하는 방법인 감성분석에 대한 연구가 활발히 이루어지고 있다(Pang and Lillian, 2010). 다양한 분야에서 감성분석이 이루어졌으며, 온라인 후기의 감성을 판별하여 마케팅 의사결정에도 활용하고 있다(Yeon et al., 2011). 기존 감성분석은 부정 또는 긍정을 내포하는 단어들을 미리 사전 형태로 구축하여 후기의 감성을 판별 한다. 이후 머신러닝 기법들을 이용하여 후기의 극성(polarity)을 분류하는 연구가 진행되었으며(Pang et al., 2002) 최근 감성사전을 사용하지 않는 딥러닝(Deep Learning)을 이용한 연구가 이루어지고 있다(Gwon, 2017).

본 논문은 한글 영화 후기를 음소 단위로 분해하여 순환신경망의 두 가지 모형(LSTM, GRU)을 이용한 감성분석을 수행하고자 한다. 지금까지 한글문장에 대한 머신러닝 기법의 감성분석은 주로 형태소 단위로 분해한 연구가 대부분 이었는데, 제한한 음소 단위 분해와 감성분석의 정확도를 비교하고자 한다.

## 2. 관련 연구

### 2.1 텍스트 감성분석

감성분석은 학습 데이터에 기계학습을 적용한 지도방식(supervised method)과 미리 구축해 놓은 감성사전을 이용하여 텍스트의 감성을 판별하는 비지도 방식(unsupervised method)으로 구분된다(Pang and Lillian, 2010). 영어의 경우, 비지도 방식에 사용하는 SentWordNet(SWN)이 공개되어 있다. SWN은 전체 어휘에 대한 긍정 및 부정 감성 강도 값이 매겨진 어휘사전으로 전 세계 다양한 분야에서 활용되었다(Felman, 2013; Taboda et al., 2011). 한글의 경우에는 Openhangul에서 감성사전을 공개하였지만 현재는 서비스 이용이 불가능하다.

Turney는 후기의 성향을 나타내는 단어들을 형용사 및 부사로 설정하였다. 이 단어들에 대한 형태소 나열 방법을 사용하여 후기를 긍정 또는 부정으로 분류하였다(Turney et al., 2002).

Taboda는 상품 및 도서 등 다양한 분야의 후기를 감성사전(SWN)에 기반한 감성분석을 시도하였다. 긍정과 부정을 나타내는 어휘의 감성 값의 차이를 이용하였는데, 만약 한 어휘가 다양한 의미를 가지는 경우 각 어휘의 긍정 및 부정의 차이들의 평균을 이용하여 후기의 감성을 분석하였다(Taboda et al., 2011).

### 2.2 인공신경망을 이용한 감성분석

인공신경망(Artificial Neural Network)의 다양한 알고리즘을 이용한 감성분석 연구가 이루어졌다. Socher et al.(2013)는 재귀신경망(Recursive Neural Network)을 기반으로 한 재귀신경 텐서망(Recursive Neural Tensor Network) 모형을 구축하였고, 이를 이용하여 단어 단위 후기의 긍정 또는 부정을 분류하였다.

이미지 분류에 주로 사용하는 신경망 모형인 CNN(Convolutional Neural Network)도 텍스트 분류에 사용되었다. Kim(2014)은 단어 단위로 분할된

텍스트를 사진의 형태로 쌓아 입력 데이터로 사용하였다. 텍스트 수치화로는 원-핫 인코딩(One-hot encoding) 방식 대신 각 단어에 해당하는 벡터를 무작위 값으로 설정하여 임베딩(embedding) 하였다. 컨볼루션 레이어(Convolutional layer)와 맥스풀링(Max pooling)을 통과하여 문장들을 분류한 결과, 이전 연구 결과보다 뛰어난 성능을 보였다.

위 연구들은 모두 영어 텍스트를 사용한 반면, Gwon(2017)은 한글 영화 후기로 딥러닝 모델을 이용한 감성분석을 실험하였다. 텍스트 수치화에 사용하는 원-핫 인코딩 방식과 워드투벡(Word2Vec) 방식을 사용하였다. 수치화된 입력 데이터를 순환신경망(RNN/LSTM/GRU)과 k-NN 모델을 감성 분석에 사용하였다. K-교차검증(K-fold cross validation) 방법을 이용하여 정확도를 측정한 결과, 워드투벡과 k-NN 모델을 이용한 감성분석의 성능이 가장 뛰어났다. 딥러닝 모델의 경우에는 원-핫 인코딩 방식(One-hot encoding)과 LSTM을 이용한 모델의 성능이 가장 뛰어났다.

<Table 1>은 기존의 감성분석에 대한 연구를 두 그룹으로 정리한 것으로, 최근 감성사전을 사용하지 않고 인공신경망을 이용한 방법이 활용되고 있음을 알 수 있다.

영어의 경우 문자단위의 텍스트 연구가 활발히 진행되었지만 한글 문장의 경우 형태소 단위로 분해한 연구가 대부분이다. 이에 본 논문에서는 한글 문장을 형태소 보다 작은 단위인 음소 단위로 분해하여 딥러닝 기반 감성분석모형을 구축하고자 한다.

## 2.3 순환신경망

이번 장에서는 순환신경망 모델 각각의 장단점

과 차이점을 살펴본다.

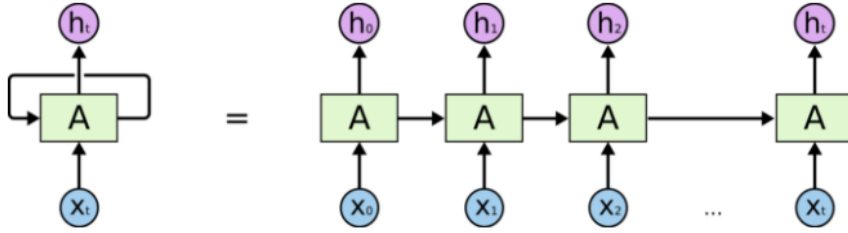
RNN은 문자 또는 음성과 같은 연속된 정보를 처리하는데 적합한 기본적인 순환신경망 모형이다 (Goodfellow et al., 2016). 기존 인공신경망(Artificial Neural Network)의 경우 각 입력 값은 독립적으로 사용한다. 그러나 문장이나 음성의 경우 t시점의 입력 값은 t+1시점의 입력 값에 영향을 준다. 이에 순환신경망의 은닉계층(hidden Layer)은 각 시점의 입력 데이터가 서로 영향을 주는 재귀적인 구조를 모형화 하였다. 은닉계층과 출력계층 노드 값은 아래와 같은 방법으로 계산한다.

$$\begin{aligned} h_t &= f(Ux_t + Wh_{t-1} + b) \\ o_t &= g(Vh_t + c) \end{aligned}$$

은닉계층의 노드( $h_t$ )는 t시점의 입력 값( $x_t$ )과 이전 시점의 은닉계층 노드값( $h_{t-1}$ )의 선형결합을 활성화 함수(f)를 이용하여 표현한다. 활성화함수는 시그모이드(sigmoid), 탄에이치(tanh)와 같은 비선형함수로 최근 ReLU(Rectified Linear Unit)를 신경망에 적용하고 있다. 출력계층의 노드 값( $o_t$ )은 활성화 함수(g)를 이용하여 0과 1사이의 확률 값으로 표현한다. 이는 문장 생성 실험에서 다음에 출현할 단어나 문자를 예측하고, 감성 분석의 경우 긍정(1) 또는 부정(0)을 예측한다. 식에서 사용한 U, W, V는 가중치를 나타내는 행렬이며 b, c는 선형결합에 사용한 편향을 의미한다. 가중치와 편향으로 이루어진 목적함수(Cost function)를 최소화시키는 최적의 파라미터를 추정하기 위해 역전파 알고리즘(Back propagation)을 이용한다. 역전파알고리즘은 가중치와 편향으로 이루어진 목적함수의 편미분계수를 사용하여 반복적으로 파라미터를 갱신한다.

<Table 1> Previous Research for Sentiment Analysis

Text sentiment analysis	Classified reviews by setting adjectives and adverbs(Turney et al., 2002)
	SWN-based sentiment analysis, using difference of sentimental value of positive and negative vocabulary(Taboda et al., 2011)
Sentiment analysis using artificial neural network	Recursive Neural Tensor Network model to classify positive or negative text(Socher et al., 2013)
	Text classification using Convolutional Neural Network(Kim, 2014)
	Sentiment analysis using Word2vec and neural network(Gwon, 2017)



<Figure 1> Recurrent Neural Network(Olah, 2015)

<Figure 1>에서 셀(Cell)이 확장하면 RNN 모델의 출력 값은 무한한 과거 시점의 출력 값에 의존(long-term dependency)하는 것을 확인할 수 있다. 이는 목적함수의 편미분 계수 값이 0에 가까워지거나 매우 커지는 문제가 발생한다(Hochreiter, 1991; Bengio et al., 1993). 편미분 계수 값이 0에 가까워지면 목적함수를 최적화시키기 위한 파라미터 학습이 제대로 이루어지지 않으며, 매우 커지게 되면 학습이 불안정해지는 문제가 발생한다.

위에서 언급한 RNN의 문제점을 보완하기 위한 방법 중 하나인 LSTM(Long Short-Term Memory) 모형은 RNN과 비슷한 형태이나 은닉계층이 메모리 셀(memory cell)이라 불리는 구조를 지닌다. 이러한 구조는 오래된 시점에서의 출력 값과 현 시점의 입력 값을 어느 정도 반영할 지 게이트(gate)를 이용해 학습하도록 고안되었다. RNN에 비해 상대적으로 복잡한 형태의 식을 가지며 다음과 같다.

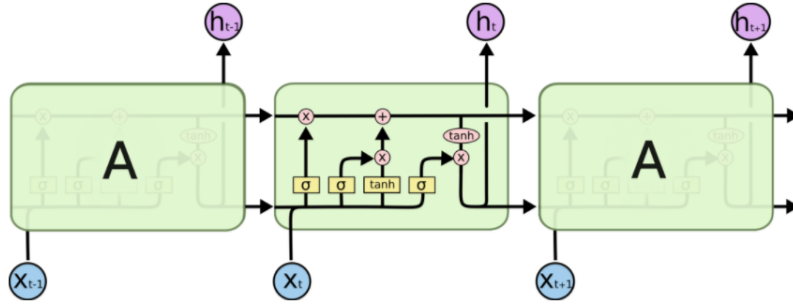
$$\begin{aligned} i_t &= \sigma(x_t U^i + h_{t-1} W^i) \\ f_t &= \sigma(x_t U^f + h_{t-1} W^f) \\ o_t &= \sigma(x_t U^o + h_{t-1} W^o) \\ \bar{c}_t &= \tanh(x_t U^c + h_{t-1} W^c) \\ c_t &= \bar{c}_t \circ i_t + c_{t-1} \circ f_t \\ h_t &= \tanh(c_t) \circ o_t \end{aligned}$$

현 시점의 입력 값( $x_t$ )과 이전 시점의 은닉노드 값( $h_{t-1}$ )의 선형결합을 활성화 함수를 통해 내부은닉노드 값( $\bar{c}_t$ )을 계산한다. 이후 입력 게이트(input gate), 망각 게이트(forget gate), 출력 게이트(output gate) 출력 값  $i_t$ ,  $f_t$ ,  $o_t$ 은 이전 시점의 은닉노드

값( $h_{t-1}$ )과 현 시점 입력 값( $x_t$ )의 선형결합을 활성화 함수( $\sigma$ )를 거쳐 0과 1사이의 값으로 계산한다. 입력 게이트 출력 값  $i_t$ 은 내부 은닉노드 값( $\bar{c}_t$ )과 결합하여 이를 얼마큼 통과시킬 지 결정한다. 망각 게이트 출력 값  $f_t$ 은 이전 시점의 은닉노드 출력 값( $c_{t-1}$ )과 결합하여 얼마큼 반영할지 결정한 후  $\bar{c}_t \circ i_t$ 와 선형결합 하여 현 시점 은닉노드 출력 값( $c_t$ )을 생성한다. 탄에이치(tanh)를 통과한  $c_t$ 은 출력 게이트 출력 값  $o_t$ 과 결합하여 다음 네트워크로 얼마큼 반영할지 계산한 후 최종 은닉노드 출력 값( $h_t$ )을 생성한다. 식에서 사용한 U, W는 가중치를 나타내는 행렬이며 위첨자 g, i, f, o는 현 시점 은닉노드( $\bar{c}_t$ ), 입력 게이트, 망각 게이트, 출력 게이트에서의 계산을 의미한다.  $\circ$ 은 벡터 간 각 요소의 곱이며  $\sigma$ 는 활성화 함수로 LSTM 모형에서는 주로 시그모이드(sigmoid)를 사용한다.

RNN에 비해 상대적으로 복잡한 계산과정을 거치지만 무한한 과거 시점의 출력 값에 의존함을 줄일 수 있다. 하지만 각 게이트마다 사용하는 가중치의 양이 많아 모형 학습속도가 RNN에 비해 떨어진다.

학습속도가 RNN에 비해 상대적으로 느린 LSTM을 보완하기 위한 다양한 변형모델 중 하나인 GRU는 리셋 게이트(reset gate)와 업데이트 게이트(update gate)로 이루어졌다. 리셋 게이트는 새로운 입력을 이전 메모리와 얼마나 합칠지 조정하며 업데이트 게이트는 이전 메모리를 얼마큼 기억할 지 조절한다. LSTM에 비해 게이트수가 적어 다음과 같은 간단한 형태의 식을 가진다.



〈Figure 2〉 Long Short-Term Memory(Olah, 2015)

$$\begin{aligned}
 r_t &= \sigma(x_t U^r + h_{t-1} W^r) \\
 u_t &= \sigma(x_t U^u + h_{t-1} W^u) \\
 \bar{h}_t &= \tanh(x_t U^h + (r_t \circ h_{t-1}) W^h) \\
 h_t &= \bar{h}_t \circ (1 - u_t) + h_{t-1} \circ u_t
 \end{aligned}$$

현 시점의 입력 값( $x_t$ )과 이전 시점의 은닉노드 출력 값( $h_{t-1}$ )의 선형결합을 활성화 함수( $\sigma$ )를 통해 0과 1사이의 값을 갖도록 리셋 게이트, 업데이트 게이트 출력 값  $r_t, u_t$ 을 계산한다. 리셋 게이트 출력 값  $r_t$ 과 이전 시점의 은닉노드 출력 값( $h_{t-1}$ )을 결합한 값을 현 시점의 입력 값( $x_t$ )과 선형결합하여 탄에이치(tanh)를 통해 내부 은닉노드 값( $\bar{h}_t$ )을 계산한다. 이후 내부 은닉노드 값( $\bar{h}_t$ )과 1에서 업데이트 게이트 출력 값을 뺀 값( $1 - u_t$ )의 결합은 이전 시점의 은닉노드 출력 값( $h_{t-1}$ )과 업데이트 게이트 출력 값  $u_t$ 의 결합과 더해져 최종 은닉노드 출력 값( $h_t$ )을 생성한다. 식에서 사용한 U, W는 가중치를 나타내는 행렬이며 위첨자 r, u, h는 내부 은닉노드 값( $\bar{h}_t$ ), 업데이트 게이트, 리셋 게이트에서의 계산을 의미한다.  $\circ$ 는 벡터 간 각 요소의

곱이며  $\sigma$ 는 활성화 함수로 GRU 모형에서도 마찬가지로 시그모이드(sigmoid)를 사용한다.

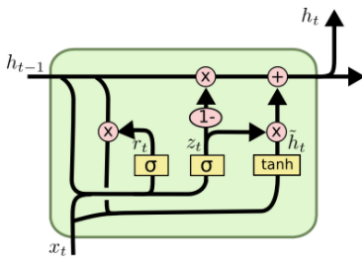
LSTM 모형에서의 입력 게이트와 망각 게이트가 업데이트 게이트로 합쳐진 것으로 확인할 수 있다. 또한 GRU 내에서 최종 은닉노드 출력 값( $h_t$ )은 활성화 함수를 거치지 않는다.

LSTM과 GRU의 성능차이는 Chung et al.(2014)과 Jozefowicz et al.(2015)에 의하면 큰 차이가 없다고 한다. GRU의 경우 가중치와 편향의 수가 LSTM에 비해 적어 소규모 데이터로도 학습이 가능하지만, 충분한 데이터의 경우 LSTM의 결과가 뛰어날 수 있다.

본 논문은 LSTM, GRU 순환신경망을 이용하여 한글 음소 단위 감성분석모형과 형태소 단위 감성분석모형을 구축하여 비교분석 한다.

### 3. 순환신경망을 이용한 음소 단위 감성분석모형

본 논문에서는 한글문장을 음소 단위로 분할한 뒤 입력하여 문장의 감성을 부정 또는 긍정으로 예측하는 모형을 구축한다. 문장들을 음소 단위로 분해하여 유일한 음소로 이루어진 사전을 구축한 뒤, k개의 음소로 이루어진 문장벡터  $x_n = (c_1, c_2, \dots, c_k)^T$ 가 주어지면 긍정과 부정이 될 확률을 계산하여 출력한다. 위 사전은 감성사전이 아닌 학습에 사용한 데이터의 유일한 음소를 구축한 것으로 문장벡터를 생성할 때 사용한다.



〈Figure 3〉 Gated Recurrent Unit(Olah, 2015)



〈Table 3〉 Examples of Reviews from NAVER

	examples of reviews
negative	<ul style="list-style-type: none"> <li>- 인상 깊지 않은 영화...(Unimpressive movie...)</li> <li>- 몰입하기 힘들었다...(It was hard to be immersed...)</li> <li>- 아이는 좋아하는데 엄마아빠는 지루해요(The child liked it, but parents are boring)</li> <li>- 그놈의 피라미드 포홀스맨 올 싱아 이 지긋지긋 하다 새로움도 없고 화려함도 없었다(The Pyramid of Forthus Mann Olsinia is boring. There is no newness and glamor)</li> <li>- 감독이 바뀌어서 그런가 애니랑 느낌이 다른 어중간한데서 끊어짐(The director changed, so the feeling was broken)</li> </ul>
positive	<ul style="list-style-type: none"> <li>- 아이들과 같이 봤는데~~어른인 저도 재밌었네요(I saw them with the kids~~ and was funny as well)</li> <li>- 슬플 줄 알았는데 재밌었어요ㅎ(I thought it was sad, but fun ha)</li> <li>- 왜 강동원 강동원하는지 알겠다!! 강동원의 매력이 터지는 영화!!! 재밌게 봤어요~(I know why it is Kang Dong Won!! Gang Dong Won's charm pops!!! I have enjoyed it~)</li> <li>- 멋진 제주풍경들~~바다~~윤여정님과 김고은의 연기도 최고였죠(Nice Jeju scenery~~Sea~~ Yoon Ji-jung and Kim Go-eun were the best performers)</li> <li>- 매직브러시 소재가 썩나서 보여줬는데, 조카가 진짜 신나하네요 ㅎㅎ(The magic brush material showed me a lot of fun, but my nephew is really excited)</li> </ul>

〈Table 4〉 Summary of Experiment Data

	Phoneme	Morpheme
Number of data	30,000(positive 50%, negative 50%)	
How to collect	crawling Naver movie reviews	
Review length	333	143
dictionary size	78	2006

음소 단위 감성분석과 형태소 단위 감성분석모형을 비교하기 위해 10-교차 검증을 이용하였다. 두 모형은 50 배치(batch) 크기와 100 에포크(epoch)로 학습되었으며 이후 실험데이터를 이용하여 정확도와 로스(loss)를 계산하였다. 또한 학습 데이터에 의한 과적합(overfitting)을 방지하기 위해(Srivastava et al., 2014), 매 계층에서 40%를 dropout 하였다. 실험은 Intel Xeon CPU 1개(16코어)와 NVidia GeForce GTX 1080 GPU 1개가 장착된 PC에서 진행하였으며 텐서플로(TensorFlow)를 백엔드(Backend)로 설정한 케라스(Keras) 패키지를 사용하였다(Chollet, 2015).

#### 4.1 실험 결과

세 모형의 정확도와 학습 데이터의 로스는 각각 〈Table 5〉, 〈Table 6〉에 나타나 있다. 음소 단위

LSTM 모형과 형태소 단위 LSTM 모형은 각 집단의 정확도가 큰 차이를 보이지 않으며 최종 정확도는 10개 집단의 정확도 평균으로 음소 단위모형이 약 1% 높았다. 로스(loss)의 경우, 각 학습 데이터 집단의 정확도와 반비례하여 최종 로스(loss)도 음소 단위 모형이 0.01 낮게 측정되었다.

음소 단위 모형의 비교로는, GRU 모형은 각 집단의 정확도의 차이가 크며(+/- 6.81%) 일정한 학습이 이루어지지 않음을 확인할 수 있다.

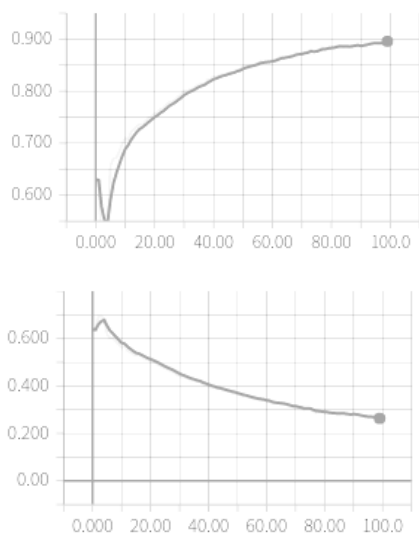
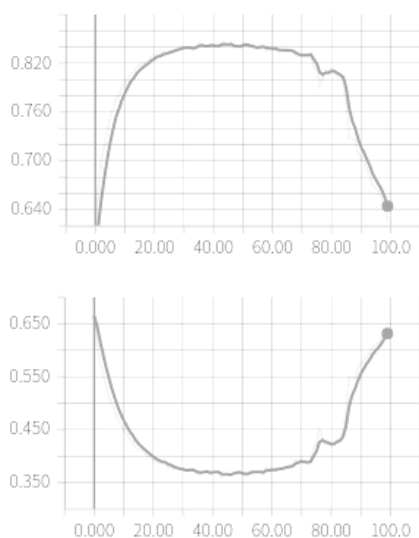
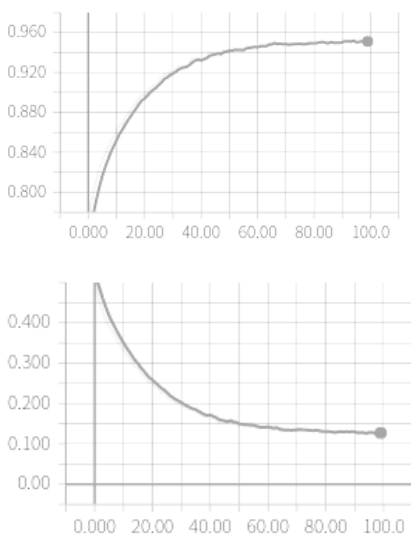
〈Table 5〉 Accuracy for 10-Groups of Three Models

	Phoneme-unit model LSTM 4 layer	Phoneme-unit model GRU 3 layer	Morpheme-unit model LSTM 1 layer
1-fold	78.7333(%)	77.9667(%)	76.8333(%)
2-fold	79.9(%)	76.9333(%)	78.3333(%)
3-fold	78.8(%)	78.7(%)	78.9(%)
4-fold	77.4667(%)	61.6667(%)	78.3667(%)
5-fold	77.4667(%)	77.4(%)	77.9333(%)
6-fold	77.4333(%)	62.5(%)	78.2667(%)
7-fold	77.8667(%)	68.3333(%)	77.2667(%)
8-fold	78.2333(%)	74.0333(%)	78.2(%)
9-fold	78.3667(%)	63.2(%)	77.3667(%)
10-fold	78.7333(%)	64.2333(%)	77.1(%)
<b>final</b>	<b>78.73%</b> <b>(+/- 0.79%)</b>	<b>70.50%</b> <b>(+/- 6.81%)</b>	<b>77.86(%)</b> <b>(+/- 0.64%)</b>

〈Table 6〉 Loss for 10-Groups of Three Models

	Phoneme-unit model LSTM 4 layer	Phoneme-unit model GRU 3 layer	Morpheme-unit model LSTM 1 layer
1-fold	0.4602	0.4725	0.4884
2-fold	0.4479	0.4848	0.4667
3-fold	0.4561	0.4570	0.4549
4-fold	0.4750	0.6171	0.4713
5-fold	0.4444	0.4773	0.4558
6-fold	0.4714	0.6427	0.4587
7-fold	0.4725	0.6112	0.4766
8-fold	0.4617	0.5615	0.4609
9-fold	0.4624	0.6355	0.4588
10-fold	0.4588	0.6353	0.4612
<b>final</b>	<b>0.4644</b>	<b>0.5703</b>	<b>0.4701</b>

〈Figure 5〉, 〈Figure 6〉, 〈Figure 7〉는 세 모형의 학습 데이터를 임의로 10개 집단으로 나눠 학습할 때, 마지막 10번째 집단의 학습과정을 텐서보드(TensorBoard)를 이용하여 그래프로 나타낸 것이다. 가로축과 세로축은 각 epoch, 정확도를 의미한다.

〈Figure 5〉 Phoneme-unit Model(LSTM) 10<sup>th</sup> Group Learning Graph〈Figure 6〉 Phoneme-unit Model(GRU) 10<sup>th</sup> Group Learning Graph〈Figure 7〉 Morpheme-unit Model(LSTM) 10<sup>th</sup> Group Learning Graph

음소 단위 LSTM 모형의 경우, 10번째 집단의 학습 데이터로 90%의 정확도를 가진다. 하지만 실험데이터로 모형을 평가하면 78.73%의 정확도를 가져 과적합이 발생하였으며 초기에 학습이 이루어지지 않은 것을 확인할 수 있다. 음소 단위 GRU



모형은 정확도가 82%가 된 이후에 낮아지면서 음소 단위 LSTM 모형에 비해 학습이 원활하게 이루어지지 않았다. 형태소 단위 LSTM 모형은 학습 데이터로 약 96%의 정확도를 가지지만 실험데이터로 평가하면 77%의 정확도를 가진다. 학습과정 그래프와 최종 정확도를 살펴본 결과, 세 모형은 모두 과적합 된 것을 확인할 수 있다.

## 5. 결 론

본 논문에서는 한글 음소 단위 감성분석모형을 구축하여 형태소 단위 감성분석모형과 비교하였다. 모형 평가의 척도로 정확도와 로스(loss)는 음소 단위 모형이 형태소 단위 모형보다 미세한 차이로 성능이 좋았으며 음소 단위 모형은 GRU 보다 LSTM을 적용하였을 때 더 뛰어났다.

본 논문에서 제시한 음소 단위 감성분석모형은 기존 감성사전을 구축하여 문장의 감성을 분류하는 모형과 달리 감성사전을 구축하지 않아도 되는 장점이 있다. 또한 음소 단위 모형은 형태소 단위 모형보다 사전의 크기가 감소되어 학습하는 가중치 행렬의 크기도 감소한다. 이는 모형 학습의 복잡성을 줄여 학습시간을 단축시키며 형태소 단위 모형에서 사용하는 임베딩 레이어(embedding layer)를 사용하지 않아 모형 학습이 효율적이다. 형태소 보다 작은 음소 단위로 문장들을 분할하기 때문에 긍정 또는 부정을 나타내는 후기의 문맥을 상세히 파악하여 감성을 분류할 수 있다.

하지만 학습 데이터로 후기를 긍정 또는 부정으로 분류할 때, 긍정 또는 부정 기준의 모호성이 존재하는데 예로 부정으로 분류한 별 5개 후기가 긍정을 내포하는 경우가 있다. 또한 학습할 때의 정확도와 실험데이터로 측정한 정확도의 차이가 커지는 과적합이 발생하는 한계점이 존재 한다.

이러한 감성분석모형은 영어에 비해 연구가 부족한 한국어 기반 언어처리 또는 음성인식 분야에 활용될 것으로 기대한다. 향후 연구 계획으로 본 논문에서 제시한 모형과 감성사전을 활용한다면

보다 성능이 뛰어난 감성분석모형을 구축할 수 있을 것으로 보인다.

## References

- Ahn, S.M., Y.C. Chung, J.J. Lee, and J.H. Yang, "Korean Sentence Generation Using Phoneme-Level LSTM Language Model", *Journal of Intelligence and Information*, Vol.23, No.2, 2017, 71-88.
- (안성만, 정여진, 이재준, 양지현, "한국어 음소 단위 LSTM 언어모델을 이용한 문장 생성", *지능정보연구*, 제23권, 제2호, 2017, 71-88.)
- Bengio, Y., P. Frasconi, and P. Simard, "The Problem of Learning Long-term Dependencies in Recurrent Networks", *IEEE International Conference on Neural Networks*, San Francisco, US, 1993, 1183-1188.
- Chollet, F., "Keras", 2015, Available at <https://github.com/fchollet/keras>(Downloaded December 1. 2017).
- Chung, J.Y., C. Gulchre, K.H. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling", *arXiv preprint arXiv : 1412.3555*, 2014.
- Dave, K., S. Lawrence, and D.M. Pennok, "Mining-thepeanutgallery : Opinion Extraction and Semantic Classification of Product Reviews", *Proceedings of the 12<sup>th</sup> International Conference on World Wide Web*, 2003, 519-528.
- Feldman, R., "Techniques and Applications for Sentiment Analysis", *Communications of the ACM*, Vol.56, No.4, 2013, 82-89.
- Goodfellow, I., Y. Bengio, and A. Courville, *Deep Learning*, The MIT Press, 2016.
- Gwon, S.J., "Sentiment Analysis of Movie Reviews using the Word2vec and RNN", University of Dongkuk, 2017.

- (권수정, “Word2vec과 RNN을 이용한 영화 리뷰의 감성분석”, 동국대학교, 2017.)
- Hochreiter, S. and J. Schmidhuber, “Long Short-Term Memory”, *Neural Computation*, Vol. 9, No.8, 1997, 1735-1780.
- Jozefowicz, R., W. Zaremba, and I. Sutskever, “An Empirical Exploration of Recurrent Network Architectures”, *ICML'15 Proceedings of the 32nd International Conference on International Conference on Machine Learning, Lille, France*, 2015, 2342-2350.
- Kim, Y., “Convolutional Neural Networks for Sentence Classification”, *arXiv Preprint arXiv : 1408.5882*, 2014.
- Lee, J.J. and S.M. Ahn, “Sentiment Analysis of Deep Learning Model Using Hangul Phoneme Unit : Focused on Translated Movie Review(IMDB)”, *Proceedings of the Korea Intelligent Information System Society Conference(SCW-2017)*, 2017, 113-114.
- (이재준, 안성만, “한글 음소 단위를 적용한 딥러닝 모형의 감성분석 : 번역된 영화리뷰(IMDB)를 중심으로”, 한국지능정보시스템학회, 춘계학술대회(SCW-2017), 2017, 113-114.)
- Olah, C., “Understanding LSTM Networks”, 2015. Available at <http://colah.github.io/posts/2015-08-Understanding-LSTMs>(Accessed December 1. 2017).
- Pang, B. and L. Lillian, “Opinion Mining and Sentiment Analysis”, *Foundations and Trends® in Information Retrieval*, Vol.2, No.1-2, 2008, 1-135.
- Pang, B. and L. Lillian, “Sentiment Analysis and Subjectivity”, *Handbook of Natural Language Processing*, 2010.
- Pang, B., L. Lillian, and V. Shivakumar, “Thumbs up? : Sentiment Classification Using Machine Learning Techniques”, *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, Vol. 10, 2002, 79-86. doi : 10.3115/1118693.1118704 (Download February 27, 2017).
- Socher, R., A. Perelygin, J.Y. Wu, J. Chuang, C.D. Manning, A. Ng, and C. Potts, “Recursive Deep Models for Semantic Compositionality over A Sentiment Treebank”, *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, US*, 2013, 1631-1642.
- Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout : A simple Way to Prevent Neural Networks from Overfitting”, *Journal of Machine Learning Research*, Vol.15, No.1, 2014, 1929-1958.
- Taboada, M., J. Brooke, M. Tofiloski, K. Voll, and M. Stede, “Lexicon-based Methods for Sentiment Analysis”, *Computational Linguistics*, Vol.37, No.2, 2011, 267-307.
- Turney, P.D., “Thumbs up or Thumbs Down? : Semantic Orientation Applied to Unsupervised Classification of Reviews”, *Proceedings of the 40<sup>th</sup> Annual Meeting on Association for Computational Linguistics, Philadelphia, Pennsylvania, US*, 2002, 417-424.
- Yeon, J.H., D.J. Lee, J.H. Shim, and S.G. Lee, “Product Review Data and Sentiment Analytical Processing Modeling”, *The Journal of Society for e-Business Studies*, Vol.16, No. 4, 2011, 125-137.
- (연종흠, 이동주, 심준호, 이상구, “상품 리뷰 데이터와 감성 분석 처리 모델링”, 한국전자거래학회지, 제16권, 제4호, 2011, 125-137.)
- Zhu, F. and X. Zhang, “Impact of Online Consumer Reviews on Sales : The Moderating Role of Product and Consumer Characteristics”, *Journal of Marketing*, Vol.74, No.2, 2010, 133-148.

## ◆ About the Authors ◆



**Jae Jun Lee (wowns4787@kookmin.ac.kr)**

Jae Jun Lee is currently in a master's course in Data Science from the Graduate School, Kookmin University. He received a B.S. degree in Mathematics from Kookmin University in 2016. His research interests include artificial intelligence, neural network and deep learning.



**Suhm Beom Kwon (sbkwon@kookmin.ac.kr)**

Suhm Beom Kwon is currently a professor in the School of Business Administration (Bigdata Business and Statistics Major), Kookmin University, Seoul, Korea. He has a MS and Ph.D. in Management Information Systems from KAIST and BA from Seoul National University. His research interests are on-line platform business model, business ecosystem, servitization and intelligent information systems.



**Sung Mahn Ahn (sahn@kookmin.ac.kr)**

Sung Mahn Ahn is a Professor with School of Business Administration, Kookmin University, Seoul, Korea. He holds a bachelor's degree in business from Seoul National University and M.S. from KAIST, and a Ph.D. in information technology from George Mason University, Fairfax, Virginia, USA. His research areas include statistical learning, data mining, neural networks, and deep learning.