

## BERT 기반 특징추출기를 이용한 감성 분석

Sentiment Analysis Using Feature Extractor Based on BERT

---

저자 (Authors)	신승엽, 최희원, 김한준 Seung-Yeop Shin, Hee-Won Choi, Han-Joon Kim
출처 (Source)	<a href="#">한국콘크리트학회 학술대회 논문집 31(1)</a> , 2019.5, 787-788(2 pages)
발행처 (Publisher)	<a href="#">한국콘크리트학회</a> Korea Concrete Institute
URL	<a href="http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE08602393">http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE08602393</a>
APA Style	신승엽, 최희원, 김한준 (2019). BERT 기반 특징추출기를 이용한 감성 분석. 한국콘크리트학회 학술대회 논문집, 31(1), 787-788
이용정보 (Accessed)	가천대학교 203.249.***.201 2019/09/02 17:09 (KST)

---

### 저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

### Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

# BERT 기반 특징추출기를 이용한 감성 분석

## Sentiment Analysis Using Feature Extractor Based on BERT

신 승 엽\*

최 희 원\*

김 한 준\*\*

Seung-Yeop Shin

Hee-Won Choi

Han-Joon Kim

### ABSTRACT

In this study, we proposed a deep learning model for sentiment analysis using BERT language model as a feature extractor. Unlike existing word embedding, feature extraction using BERT language model has the advantage of contextual word embedding. We derived the combination of BERT hidden layers best for sentiment analysis, and evaluated the accuracy of sentiment analysis.

### 요 약

이 연구에서는 BERT 언어모델을 특징추출기로 이용하여 텍스트 감성분석을 수행하는 딥러닝 모델을 제안하였다. 기존 단어 임베딩과는 달리 BERT 언어모델을 활용한 특징 추출은 문맥 기반 단어 임베딩이 가능하다는 장점이 있다. 감성분석에 적합한 BERT 은닉층의 조합을 도출하였으며, 이를 이용하여 감성분석 정확도를 평가하였다.

### 1. 서 론

소셜 미디어의 발전으로 실시간으로 생성되는 텍스트의 양이 증가했다. 감성 분석은 텍스트로부터 대중의 반응을 파악할 수 있는 좋은 방법이다. 이러한 감성 분석을 잘 수행하기 위해서는 텍스트로부터 좋은 특징을 추출하는 것이 중요하다. 이 연구에서는 BERT 언어모델을 특징 추출기로 이용하여 텍스트로부터 좋은 특징을 추출해 감성분석을 수행하는 딥러닝 모델을 제안한다. 이를 위해, 감성 분석에 적합한 BERT 언어모델의 은닉 층의 조합을 도출하였으며, 영화평 데이터를 이용하여 제안하는 감성분석 딥러닝 모델의 정확도를 평가하였다.

### 2. 감성분석 딥러닝 모델

본 논문에서 이용한 BERT 언어모델은 기본 버전으로 12개의 은닉 층으로 구성 된다[1]. 그림 1에 나타나있듯이 12개의 은닉층 중 마지막 4개 은닉 층의 합(Sum\_last\_4)을 감성 분석을 위한 문맥 기반 단어 임베딩으로 사용한다. 문맥 기반 단어 임베딩은 문장으로부터 100개까지의 단어에 대해 각 768차원의 벡터로 이루어진 특징을 추출한다. 그림 2는 단어 임베딩 이후 은닉층 3개 통해 텍스트의 긍정 및 부정을 분류하는 감성분석 딥러닝 모델의 구조를 나타낸다.

\* 서울시립대학교 전자전기컴퓨터공학부, 석사과정

\*\* 서울시립대학교 전자전기컴퓨터공학부, 교수

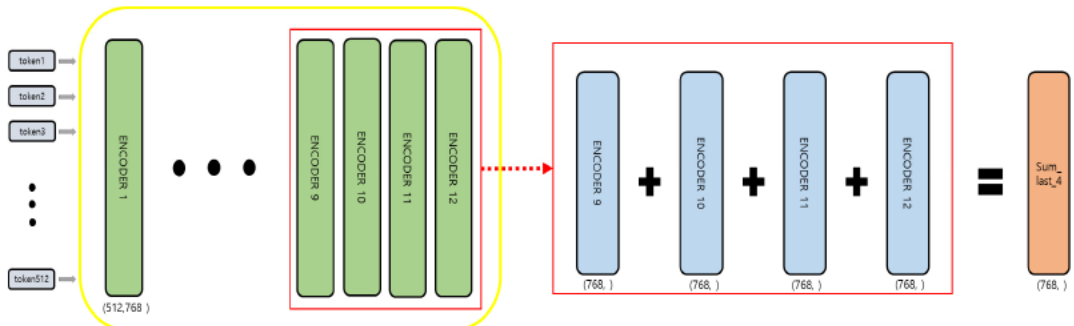


그림1. 감성분석에 적합한 BERT 은닉층의 조합

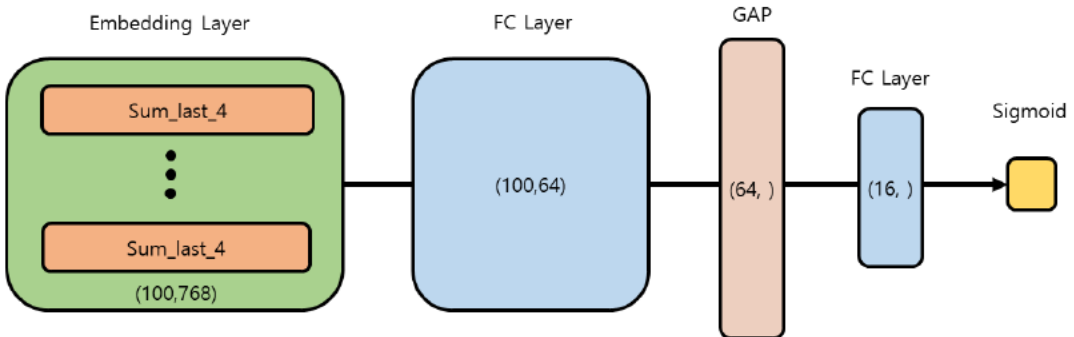


그림2. 감성분석 딥러닝 구조

### 3. 실험 및 결과

실험에 사용한 데이터는 Stanford IMDB movie review이다. IMDB 데이터는 영화평 리뷰 사이트로부터 영화평을 수집하여 긍정, 부정으로 분류한 데이터 셋이다. 실험은 25,000개의 IMDB 훈련 데이터를 서플한 후 학습데이터 10,000개, 검증 데이터 5,000개, 테스트 데이터 10,000개로 나누어 사용하였다. BERT 은닉층의 4가지 조합별 감성분석 정확도 및 GloVe[2]를 임베딩으로 사용한 감성분석 정확도가 표 1에 나타나있다. 4가지 은닉 층의 조합은 표 1의 순서대로 마지막 은닉 층(Last), 마지막에서 두 번째 은닉 층(Second-to-Last), 마지막 4개 은닉 층의 합(Sum Last Four), 모든 12개 은닉 층의 합(Sum All 12)이다. 표 1에 따르면 마지막 4개 은닉 층의 합을 단어 임베딩으로 사용한 경우 83.65%의 정확도로 다른 은닉층 조합보다 높으며 GloVe를 임베딩으로 사용한 경우보다 높은 정확도를 보였다.

표1. 은닉층조합과 GloVe 감성분석 정확도비교

Embedding	Last	Second-to-Last	Sum Last Four	Sum All 12	GloVe
정확도(%)	82.82	83.41	83.65	81.66	81.57

#### 감사의 글

이 연구는 국토교통부 도시건축연구사업의 연구비지원(19AUDP-B100356-05)에 의해 수행되었습니다.

#### 참고문헌

- Devlin, Jacob, et al. "BERT: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).
- Pennington, Jeffrey, Richard Socher, and Christopher Manning. "Glove: Global vectors for word representation." Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014.