

감정 분석을 위한 BERT 사전학습모델과 추가 자질 모델의 결합

(Combining Sentiment-Combined Model with Pre-Trained BERT Models for Sentiment Analysis)

이 상 아 *
(Sangah Lee)

신 효 필 **
(Hyopil Shin)

요약 대규모 코퍼스에 기반한 사전학습모델인 BERT 모델은 언어 모델링을 통해 텍스트 내의 다양한 언어 정보를 학습할 수 있다고 알려져 있다. 여기에는 별도의 언어 자질이 요구되지 않으나, 몇몇 연구에서 특정한 언어 지식을 추가 반영한 BERT 기반 모델이 해당 지식과 관련된 자연어처리 문제에서 더 높은 성능을 보고하였다. 본 연구에서는 감정 분석 성능을 높이기 위한 방법으로 한국어 감정 사전에 주석된 감정 극성과 강도 값을 이용해 감정 자질 임베딩을 구성하고 이를 보편적 목적의 BERT 모델과 결합하는 외적 결합과 지식 증류 방식을 제안한다. 감정 자질 모델은 작은 스케일의 BERT 모델을 적은 스텝 수로 학습하여 소요 시간과 비용을 줄이고자 했으며, 외적 결합된 모델들은 영화평 분류와 악플 탐지 문제에서 사전학습모델의 단독 성능보다 향상된 결과를 보였다. 또한 본 연구는 기존의 BERT 모델 구조에 추가된 감정 자질이 언어 모델링 및 감정 분석의 성능을 개선시킨다는 것을 관찰하였다.

키워드: BERT, 사전학습모델, 감정 분석, 외적 결합, 지식 증류, 감정 자질

Abstract It is known that BERT can capture various linguistic knowledge from raw text via language modeling without using any additional hand-crafted features. However, some studies have shown that BERT-based models with an additional use of specific language knowledge have higher performance for natural language processing problems associated with that knowledge. Based on such finding, we trained a sentiment-combined model by adding sentiment features to the BERT structure. We constructed sentiment feature embeddings using sentiment polarity and intensity values annotated in a Korean sentiment lexicon and proposed two methods (external fusing and knowledge distillation) to combine sentiment-combined model with a general-purpose BERT pre-trained model. The external fusing method resulted in higher performances in Korean sentiment analysis tasks with movie reviews and hate speech datasets than baselines from other pre-trained models not fused with sentiment-combined models. We also observed that adding sentiment features to the BERT structure improved the model's language modeling and sentiment analysis performance. Furthermore, when implementing sentiment-combined models, training time and cost could be decreased by using a small-scale BERT model with a small number of layers, dimensions, and steps.

Keywords: BERT, pre-trained model, sentiment analysis, external fusing, knowledge distillation, sentiment features

* 이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017S1A5A2A01023336)

** 이 논문은 2020 한국소프트웨어종합학술대회에서 '감정 분석을 위한 BERT 사전학습모델과 추가 자질 모델의 결합'의 제목으로 발표된 논문을 확장한 것임

† 학생회원 : 서울대학교 데이터사이언스대학원 박사후연구원
visualjan@snu.ac.kr

** 정 회 원 : 서울대학교 데이터사이언스대학원 교수
(Seoul Nat' Univ.)
hpshin@snu.ac.kr
(Corresponding author)

논문접수 : 2021년 3월 4일

(Received 4 March 2021)

논문수정 : 2021년 6월 18일

(Revised 18 June 2021)

심사완료 : 2021년 6월 22일

(Accepted 22 June 2021)

Copyright©2021 한국정보과학회 : 개인 목적이거나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.
정보과학회논문지 제48권 제7호(2021. 7)

1. 서론

트랜스포머 양방향 인코더 표현(Bidirectional Encoder Representations from Transformers, BERT) 모델[1]과 같은 대규모 언어 모델은 다른 문장 단위 임베딩을 위한 모델들과 함께 학습 과정에서 문장 내 어휘, 통사, 의미 정보를 포착하여 모델링한다고 알려져 있다[2]. 따라서 그 자체가 다양한 자연어처리 문제를 해결할 수 있는 보편적인 모델로서 기능한다. 그러나 최근 여러 연구에서 특정 언어 현상과 관련된 언어 지식을 추가적인 자질로 사용할 경우, 해당 자질과 관련된 자연어 처리 문제에서 모델의 성능이 향상된다고 보고하였다[3-6]. 특히 그 중에서도 단어나 문장의 감정 정보를 자질로 이용한 연구들[7,8]과 같이, 본 연구에서도 사전학습모델과 감정 자질을 결합하여 감정 분석 문제 해결에 도움이 되도록 하였다.

본 연구에서는 BERT 모델에 감정 자질을 추가하는 방법으로 외적 결합 방법(External Fusing Method)을 제시한다. 먼저 감정 자질을 포함하여 자질 기반 임베딩 모델을 학습시키는데, 이때 감정 자질은 문장 내의 각 토큰에 한국어 감정 분석 코퍼스(KOSAC)[9]에 대응하는 감정 극성(polarity)과 강도(intensity) 값을 부여한 것이다. 이 값들은 각각 임베딩으로 구성되어 기존의 BERT 임베딩에 더해지며, 이렇게 만들어진 임베딩을 학습한 것이 감정 자질 모델(sentiment-combined model)이 된다. 이는 각 자연어 처리 문제에 대한 모델의 미세 조정(fine-tuning) 단계에서 기존에 큰 규모로 사전학습되어 있는 임베딩 모델과 결합된다.

한편 본 연구에서는 추가적인 모델 결합 방식으로 감정 자질 모델에 기존의 대규모 사전학습모델이 학습하여 포함하고 있는 보편적인 언어 지식을 전이하는 지식 증류(Knowledge Distillation) 방법을 고려하였다. 지식 증류는 비교적 작은 규모로 학습된 감정 자질 모델이 대규모 사전학습모델의 문장 임베딩을 모사함과 동시에 감정 분석 문제에 대한 미세 조정을 수행하여 감정 분석 성능을 높이고자 한 것이다.

외적 결합 방식으로 구성된 모델은 기존에 공개된 사전학습모델들을 단독으로 사용한 경우와 비교하여 영화 평 분류, 악플 탐지 등 한국어 감정 분석 문제에서 향상된 성능을 보였다. 또한 감정 자질 모델은 기존 사전학습모델에 비하여 작은 규모와 더 적은 스텝 수로 학습되어 시간과 하드웨어 자원 등의 사용을 줄이면서도 우수한 성능을 유지한다는 장점을 갖는다. 그러나 지식 증류 방식으로 조정된 감정 자질 모델은 예상과 달리 성능의 향상을 보이지 못하여, 추가적인 분석 및 보강이 요구된다.

2. 관련 연구

많은 임베딩 모델들이 영어 텍스트를 이용하여 학습되고 따라서 영어로 구성된 자연어처리 문제를 해결하였으나, 점차 다양한 언어로 확장 연구되고 있다. 그 초기 접근으로 Google Research가 공개한 multilingual BERT 모델(M-BERT)은 104개 언어에 대하여 적용할 수 있도록 고안되었다. 그러나 학습 데이터가 104개 언어의 Wikipedia 데이터만으로 구성되어 그 장르가 한정적이라는 것과, 그로부터 구성된 사전에서 각 개별 언어가 차지하는 비율이 매우 적다는 단점이 있다. 따라서 그 규모에 비하여 각 언어에 적용했을 때 성능이 개별 언어 BERT 모델들보다 저조하게 나타난다[10-13]. 이는 공개된 몇 가지 한국어 기반 BERT 모델을 통해서도 확인할 수 있는 경향이다(KorBERT¹⁾, KoBERT²⁾, KR-BERT[13], HanBERT³⁾).

2.1 감정 자질을 이용한 전통적인 방법론

초기 감정 분석 연구에서는 나이브 베이즈(Naive Bayes), 서포트 벡터 머신(Support Vector Machine), 최대 엔트로피(Maximum Entropy) 등을 이용한 분류 방법들이 널리 쓰였고, 이후 CNN, RNN, LSTM 등의 신경망 모델과 ELMo, GPT, BERT와 같은 임베딩 모델을 활용하는 방법들도 나타났다. 이러한 모델들은 텍스트에 포함되어 있는 감정 정보를 포착하기 위해 다양한 종류의 자질들을 고안하고 사용해 왔다. 가장 기본적인 n-gram의 존재 여부 및 빈도수 즉 bag-of-words 자질과 단어의 품사 정보, 그리고 감정 어휘 사전 등이 그 예이다.

특히 감정 어휘 사전은 단어의 감정적 방향성 즉 극성과 강도 등에 따라 일정한 값을 할당하여 목록화한 것이다. Teng et al. (2016)[14]은 단어들의 극성 정보에 따른 점수 형태의 값을 weight로 하여 LSTM 모델에 포함시키는 어휘 기반 가중합 모델을 구축하고, 이를 통해 문장 단위의 감정 분류를 수행하였다. Sun et al. (2018)[15]에서는 감정 어휘와 논증, 의존 구조를 LSTM을 통해 인코딩한 뒤, 문장 내 각 단어들과의 attention을 계산하여 각 단어에 대한 가중치가 되게 하였다. 이에 따라 각 단어의 은닉 상태값(hidden states)의 가중합을 구하여 의미 벡터를 얻었다. Bao et al. (2019)[16]는 어휘 정보로부터 얻은 각 단어의 감정 극성값을 LSTM 모델과 결합하여 속성(aspect) 기반 감정 분석 문제에 이용하였다. Li and Caragea(2019)[17]는 감정 어휘 목록과의 매칭을 통해 시퀀스 안의 단어 각각이 해당 목록에 포함되어 있는지의 여부를 벡터화하여 사용하였다.

1) https://aiopen.etri.re.kr/service_dataset.php

2) <https://github.com/SKTBrain/KoBERT>

3) <https://github.com/tbai2019/HanBert-54k-N>

2.2 BERT 사전학습모델을 이용한 감정 분석 연구

임베딩 모델들은 다양한 언어 지식을 내부적으로 자동 모델링할 수 있어서, 추가적인 자질을 만들어 넣지 않더라도 그 스스로 감정 극성값과 같은 언어 자질을 인코딩할 수 있다고 알려져 있다[2]. 그러나 몇몇 모델은 자질을 더하여 관련 자연어 처리 문제에서 더 개선된 성능을 보고하였다. 특히 감정 분석에 도움이 되는 추가 정보를 사용한 BERT 모델 기반 연구들은 다음과 같다.

SentiLARE[7]은 단어와 품사 정보를 이용하여 Senti-WordNet 사전에서 각 단어에 대한 감정 극성값을 얻고, 이를 품사 태그와 문장 단위 감정 극성 정보와 함께 기존의 BERT 임베딩에 더하여 학습한 새로운 BERT 모델이다. 이는 문장 단위의 감정 정보를 포함한 리뷰 텍스트(Yelp Dataset Challenge 2019)를 학습 데이터로 사용하고, BERT 모델의 내부 학습 태스크 중 하나인 Masked Language Modeling(MLM)을 개선하여 사용하였다. 모델은 단어뿐 아니라 해당 단어의 품사와 감정 극성 정보 역시 예측하여 각각의 loss를 합산하는 방식으로 학습을 수행하고, 그에 따라 단어와 문장, 감정 정보 사이의 내포적인 관계를 학습하고자 하였다.

SKEP[8]은 감정 정보를 갖는 단어들과 측면-감정 쌍(aspect-sentiment pairs) 정보를 사전학습 과정에 포함시킨 모델이다. 적은 수의 seed words를 이용하여 최대한 비지도적인 방법(Pointwise Mutual Information)으로 시퀀스 내 단어의 극성값을 자동 주석하고, 감정 극성을 갖는 단어의 근처 단어들에 간단한 매칭을 적용하여 속성으로 간주하였다. 이렇게 얻은 측면-감정 쌍과 감정 단어에 대해 마스킹 및 예측을 적용하고 각 서브태스크의 loss를 모두 합하여 최종적인 성능을 계산한다.

본 연구에서도 위 연구들과 같이 감정 정보를 BERT 모델에 추가하는 방식을 사용하였다.

2.3 지식 증류

전이 학습(transfer learning)은 한 가지 도메인의 데이터나 문제에 대해 모델이 학습한 지식을 관련된 다른 데이터와 문제에 이동시켜 적용하는 방법론을 뜻한다. 그 하위 개념인 지식 증류(knowledge distillation) 기법은 큰 모델(teacher model)과 작은 모델(student model)을 설정하여, 큰 모델이 포함하는 지식을 전이시켜 작은 모델의 성능을 향상시키고자 한 것이다. 지식 증류는 모델의 압축 및 경량화에도 많이 사용되나, 학습된 지식을 전달한다는 점을 이용하여 한 언어에서 다른 언어로, 한 도메인에서 다른 도메인으로의 모델 전이 학습에도 적용되고 있다.

Chen et al. (2020)[18]은 하나의 도메인에 대해 미세 조정된 BERT 모델로부터 지식을 전이받은 전통적인 시퀀스 투 시퀀스 모델을 이용해 문장 생성 문제를 해

결하였다. Tan et al. (2019)[19]는 개별 언어에 대해 학습한 기계 번역 모델로부터 다중 언어에 대해 학습한 모델에 지식을 전이하여 다양한 언어쌍의 기계번역 문제에서 개선된 성능을 얻도록 하였다. Reimers and Gurevych (2020)[20] 역시 지식 증류 방법론을 이용하여 큰 모델에 해당하는 영어 문장 임베딩 모델을 작은 번역 모델의 개선에 적용하고자 하였다. 이는 번역된 언어의 문장 임베딩이 기존 영어 문장 임베딩과 벡터 공간상에서 같은 위치에 표현될 수 있도록 매핑하고자 한 것으로, 임베딩 간의 벡터 공간상의 거리는 평균 제곱 오차를 이용해 계산하였다.

본 연구는 Reimers and Gurevych (2020)[20]의 지식 전이 구조를 참고하되 기계 번역이 아닌 도메인 간의 지식 전이를 목적으로 하여, 보편적인 언어 모델링을 수행한 사전학습모델로부터 감정 분석 문제를 해결하는 감정 자질 모델로의 지식 전이를 수행하는 방법론을 고려했다.

3. 감정 자질을 이용한 모델 학습

3.1 감정 자질과 임베딩

본 연구에서는 한국어 감정 어휘 사전(sentiment lexicon) 중 하나인 한국어 감정 분석 코퍼스(KOSAC)[9]에 주석된 값을 이용해 감정 자질을 구성하였다. KOSAC은 세종코퍼스와 뉴스 기사 데이터로부터 감정 표현 17,582개를 추출하여 주석한 것이다. 주석된 감정 정보로는 표현의 주관성, 감정 극성, 강도, 표현 유형 등이 있다. 본 연구는 그 중에서 각 단어에 할당된 감정 극성과 강도 값을 자질로 이용하고자 하였다. 주석된 극성 값은 POS(긍정), NEUT(중립), NEG(부정), COMP(복합), None(없음)의 다섯 가지로, 해당 단어가 갖는 감정의 정도가 얼마나 강한지를 나타내는 강도 값은 High(높음), Medium(중간), Low(낮음), None(없음)의 네 가지로 구성된다. 감정 극성과 강도 값은 사전에 주석된 것들 중에서 단일 형태소로 구성된 3,276개 항목에 대한 것을 사용하였다.

이러한 자질 값들은 BERT의 WordPiece 토큰나이저에 의해 분석된 데이터의 각 토큰에 대응, 할당된다. 예를 들어 '너무'라는 단어는 KOSAC에 정의되어 있는 값에 따라 극성 'NEG'와 강도 'Medium'을 부여받는다. 또한 BERT 모델의 사전에는 들어 있지만 감정 사전에는 포함되어 있지 않아서 부여받은 감정 자질 값이 없는 토큰의 경우, 기본값으로 극성과 강도 모두 'None'을 부여받도록 하였다. BERT의 학습 및 연산에 필요한 특별 토큰인 [CLS], [SEP], [PAD] 등에도 'None'을 부여하였다.

이렇게 각 토큰에 부여된 자질은 정수 인덱스 벡터 형태로 극성 임베딩과 강도 임베딩을 구성하고, BERT 기본 임베딩에 더해진다(element-wise sum). 그림 1은

| | | | | | | | | | | | | |
|----------------------|-------------|--------------|--------------|-------------------|--------------------|-------------------------|------------|-------------|-----------------|----------------------------|------------|-------------|
| Input | [CLS] | 이 "this" | [MASK] | 너무 nemwu "too" | 재미 caymi "amus" | ##있따 ##issta "##ing" | . | [SEP] | 참 cam "very" | 좋았다 cohassta "was good" | . | [SEP] |
| Token Embeddings | $E_{[CLS]}$ | E_i | $E_{[MASK]}$ | E_{nemwu} | E_{caymi} | $E_{##issta}$ | $E_{.}$ | $E_{[SEP]}$ | E_{cam} | $E_{cohassta}$ | $E_{.}$ | $E_{[SEP]}$ |
| Segment Embeddings | + | + | + | + | + | + | + | + | + | + | + | + |
| | E_A | E_A | E_A | E_A | E_A | E_A | E_A | E_A | E_B | E_B | E_B | E_B |
| Position Embeddings | + | + | + | + | + | + | + | + | + | + | + | + |
| | E_0 | E_1 | E_2 | E_3 | E_4 | E_5 | E_6 | E_7 | E_8 | E_9 | E_{10} | E_{11} |
| Polarity Embeddings | + | + | + | + | + | + | + | + | + | + | + | + |
| | E_{None} | E_{POS} | E_{None} | E_{NEG} | E_{POS} | E_{None} | E_{None} | E_{None} | E_{POS} | E_{None} | E_{None} | E_{None} |
| Intensity Embeddings | + | + | + | + | + | + | + | + | + | + | + | + |
| | E_{None} | E_{Medium} | E_{None} | E_{Medium} | E_{Medium} | E_{None} | E_{None} | E_{None} | E_{High} | E_{None} | E_{None} | E_{None} |

그림 1 감정 자질을 포함한 임베딩 구성
Fig. 1 Visualization of the embedding composition

‘이 [MASK] 너무 재미있다. 참 좋았다.’라는 예시 시퀀스를 가지고 임베딩이 결합되는 모습을 시각화한 것이다. 이때 [MASK]는 원래 ‘영화’였던 토큰이 BERT의 학습 서브태스크 중 하나인 Masked Language Modeling (MLM)에 따라 가려졌음을 나타낸 것이다.

3.2 감정 자질 모델 학습

위 3.1에서 구성한 감정 자질을 포함하여 새로 학습한 BERT 기반 모델이 감정 자질 모델(sentiment-combined model)이다. 본 연구에서는 학습 데이터와 사전 구성에 따라 두 가지 서로 다른 감정 자질 모델을 학습하였다.

- KR-BERT-KOSAC-small: 기존의 KR-BERT 모델과 같이 한국어 위키피디아 데이터와 뉴스 기사로 구성된 2.47GB의 텍스트로 학습한 모델. 사전은 해당 학습 데이터를 통해 Byte-Pair Encoding (BPE) 방식으로 구성된 10,000개의 단어로 한국어 텍스트에 자주 사용되는 라틴 알파벳, 한자, 일본어 등의 문자들을 추가한 16,424개의 항목으로 구성되어 있다.
- CH-BERT-KOSAC-small: 한국어 뉴스 기사 댓글 데이터⁴⁾와 한국어 악플 탐지 데이터셋[21]의 미주석된 텍스트를 합한 12.7GB의 학습 데이터에 기반해 구축한 모델. 사전은 위와 같이 데이터로부터 BPE 방식으로 정렬한 20,000개의 단어로 구성하였다.

두 감정 자질 모델 모두 8개의 인코더 은닉층, 8개의 어텐션 헤드, 512차원 임베딩의 비교적 작은 규모로 구성하였다. 또한 데이터 시퀀스의 최대 길이를 128, 학습 배치사이즈를 64, 학습률을 1e-4로 설정하였으며 Adam optimizer를 사용하였고, 각각 학습 성능이 안정적으로 수렴하기 시작하는 50만 스텝까지 학습하였다. 그림 2는 CH-BERT-KOSAC-small 모델을 100만 스텝까지 학습하면서 관찰한 학습 loss 값의 추이로, 50만 스텝 지점

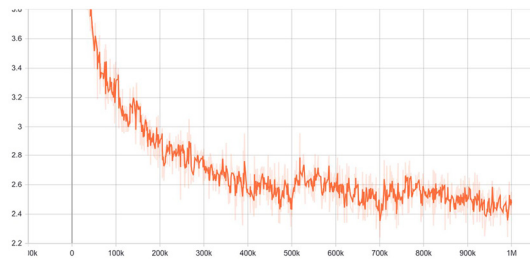


그림 2 CH-BERT-KOSAC-small의 학습 loss
Fig. 2 Training loss of the CH-BERT-KOSAC-small model

전후에서부터 그 값이 크게 변화하지 않는 것을 볼 수 있다. 이에 따라 본 연구는 두 가지 감정 자질 모델의 학습 스텝수를 50만으로 설정하였다.

두 모델의 학습에는 2개의 Tesla V100 GPU를 이용하여 70시간 정도 소요되었다. 이는 일반적인 사전학습 모델이 12개 이상의 은닉층과 어텐션 헤드, 768차원 이상의 임베딩 등을 사용하여 TPU와 같은 고비용 장비를 통해 학습되는 것에 비해 매우 경제적인 규모이다.

또한 MLM을 위해 각 시퀀스의 최대 15%까지의 토큰을 마스킹하도록 하였다. 이때 CH-BERT-KOSAC-small 모델의 경우 학습 데이터의 대부분이 한두 문장 정도의 댓글로 이루어져 있어, 문장 간의 관계를 모델링하는 BERT의 서브태스크인 Next Sentence Prediction (NSP)을 올바르게 수행하지 못할 것이라 생각하였다. 따라서 CH-BERT-KOSAC-small 모델은 BERT의 두 서브태스크 중 MLM만을 사용하여 학습하였다.

한편 위와 같은 소규모 모델과 성능을 비교하기 위해 기본 BERT 모델("BERT-Base")과 같은 규모인 감정 자질 모델인 KR-BERT-KOSAC을 추가로 학습하였다. 이 모델은 기존 공개된 KR-BERT 모델 및 KR-BERT-KOSAC-small과 같은 학습 데이터와 사전을 이용하고,

4) <https://github.com/Beomi/KcBERT>

마찬가지로 12개의 인코더 은닉층, 12개의 어텐션 헤드, 768차원 임베딩으로 구성하였다. 데이터 시퀀스의 최대 길이를 512, 학습 배치사이즈를 64, 학습률을 1e-4로 설정하고 Adam optimizer를 사용하였으며, KR-BERT와 같은 200만 스텝까지 학습을 수행하였다.

3.3 외적 결합

본 연구에서는 서로 다르게 학습된 사전학습모델 두 개를 각각 불러와서 하나의 자연어처리에 대해 함께 미세 조정하고 결과를 내는 외적 결합 방식을 제안한다. 그림 3에서 볼 수 있듯, 한쪽 모델로는 공개되어 있는 한국어 사전학습모델을, 다른 한쪽 모델로는 3.2에서 학습한 감정 자질 모델을 이용하였다. 특히 한국어 텍스트를 다룰 수 있는 사전학습모델 중에서 TensorFlow 기반으로 공개된 음절 단위 모델인 KorBERT와 KR-BERT, KR-BERT-MEDIUM⁵⁾, 그리고 M-BERT 모델을 선택하였다.

결합 방법은 각 모델의 체크포인트 파일로부터 미리 학습되어 있는 weights를 불러오고, 이를 이용해 로드한 모델 각각이 데이터 시퀀스에 대해 개별 loss를 계산하는 것이다. 각 모델이 구성한 문장 임베딩의 마지막 은닉 상태 중 맨 앞의 토큰인 [CLS]에 해당하는 부분이 문장 전체의 의미를 연산하여 포함하고 있으므로 이를 이용한다(CLS-pooling). 두 임베딩으로부터 각각 계산

한 loss는 선형 결합한 뒤 Adam optimizer를 통해 한번에 최적화된 뒤 각 문장의 정답 예측에 사용된다.

그에 따라 두 개의 서로 다른 성격을 가진 모델들이 상호작용을 하여 성능 향상에 도움을 줄 것이다. 특히 현재 공개되어 있는 사전학습 모델들은 대규모 코퍼스를 이용하여 충분한 스텝수로 언어 모델링을 수행한 것이므로, 그 자체가 보편적인 언어 모델로서 텍스트에 포함된 다양한 언어 정보를 잘 포착해낼 것으로 기대된다. 또한 감정 자질 모델은 추가된 감정 자질에 따라 텍스트에 녹아 있는 관련 정보를 인식하고 감정 분석 문제에서 유리하게 기능할 것이다. 이렇게 두 모델이 서로 다른 측면에서 장점을 가지고 있으므로 성능 향상에 도움이 될 것이라고 생각된다.

4. 실험 및 결과

4.1 실험 설정

본 연구는 감정 분석 문제로 영화평 분류와 악플 탐지 문제를 수행하였다.

- 영화평 분류: 네이버 영화평 분류 코퍼스(NSMC)⁶⁾를 이용하여, 각 리뷰가 영화에 대해 내리는 평가를 ‘긍정’, ‘부정’의 두 가지로 분류하는 문제이다. 학습 데이터 15만 개, 테스트 데이터 5만 개 리뷰로 구성되어 있으며 본 연구에서는 학습 데이터 중 2만 개 리뷰를 분리하여 검증 데이터로 이용하였다. 모델 성능은 예측 정확도(accuracy)를 이용하여 평가한다.
- 악플 탐지: 공개된 악플 탐지(Hate Speech Detection) 데이터셋²¹⁾ 중 주석된 부분을 이용하며, 악플 여부는 표현의 정도에 따라 ‘hate(개인이나 단체에 대한 공격적인 표현 포함)’, ‘offensive(명백한 공격은 아니나 불쾌할 수 있는 표현 포함)’, ‘none(공격적인 표현을 포함하지 않음)’의 세 가지로 나뉜다. 학습 데이터 7,896개, 검증 데이터 471개, 테스트 데이터 974개 텍스트로 구성되어 있으며, 중간 검증 성과와 테스트 데이터 성능은 공개된 리더보드⁷⁾에 따라 F1 점수를 이용하여 평가한다.

두 문제 모두에 대하여 3.3에서 결합한 모델을 학습 배치사이즈 128, 검증 배치사이즈 8, 데이터 시퀀스의 최대 길이 128, 학습률 5e-5로 설정하여 5 에폭 동안 파인튜닝하였다. 감정 자질 모델을 추가하지 않은 단독 사전학습 모델들의 경우, 공개된 토큰나이저와 사전, 모델 파일 등을 이용하여 본 연구가 재구성한 실험 결과를 보고한다. 이들 사전학습모델은 감정 자질 및 외적 결합 방법의 효과를 확인하기 위한 베이스라인이 되며,

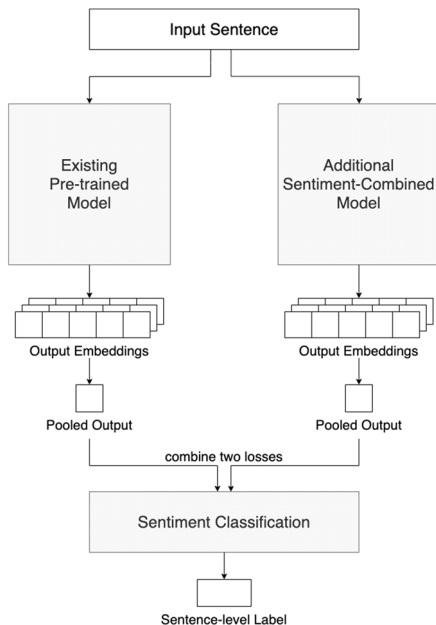


그림 3 외적 결합 방식

Fig. 3 External Fusing Method

5) <https://github.com/snunlp/KR-BERT-MEDIUM>

6) <https://github.com/e9t/nsmc>

7) <https://www.kaggle.com/c/korean-hate-speech-detection/leaderboard>

KorBERT와 KR-BERT, KR-BERT-MEDIUM, M-BERT의 네 가지 모델을 포함한다.

4.2 실험 결과

두 감정 분석 문제에 대한 실험 결과는 표 1과 같다. 네 가지 기본 사전학습모델 M-BERT, KorBERT, KR-BERT, KR-BERT-MEDIUM 아래 +KR-small, +CH-small로 표시된 행들은 각각 사전학습모델에 감정 자질 모델인 KR-BERT-KOSAC-small과 CH-BERT-KOSAC-small을 결합한 실험 결과를 나타낸다. 맨 아래 행은 KR-BERT-KOSAC 모델의 단독 실험 결과를 보인다.

네 가지 사전학습모델 모두가, 두 가지 감정분석 문제 모두에서 감정 자질 모델과 결합했을 때 명확한 성능 향상을 보였다. 영화평 분류 문제에서는 M-BERT에 감정 자질 모델을 결합했을 때 성능이 2~3%p 상승하였고, KorBERT와 KR-BERT 역시 단독으로 실험했을 때보다 감정 자질 모델을 함께 사용했을 때 성능이 1%p 가까이 향상되었다. KR-BERT-MEDIUM의 경우 단독 모델에 비하여 정확도가 소폭 상승하였으나 본 연구의 실험 결과 중 가장 우수한 성능을 보였다. 특히 KorBERT와 KR-BERT, KR-BERT-MEDIUM에서 영화평의 검증 데이터와 테스트 데이터로부터 얻은 90.50 이상의 정확도는 현재까지 공개된 네이버 영화평의 실험 성능 중 가장 높은 수준이다.

악플 탐지 문제에서는 감정 자질 모델을 결합했을 때 성능의 상승폭이 영화평 분류 문제에서보다 더 크게 나타났다. M-BERT와 KR-BERT, KR-BERT-MEDIUM을 단독으로 사용했을 때보다 검증 데이터와 테스트 데이터

의 F1 점수가 상승하였고, KorBERT 역시 감정 자질 모델을 결합했을 때 테스트 데이터 성능이 향상되었다. 특히 M-BERT에 CH-BERT-KOSAC-small 모델을 결합했을 때 검증 데이터의 예측 정확도는 8%p 향상되었다.

악플 탐지 문제의 경우 두 가지 감정 자질 모델에 의한 성능 상승폭에 차이가 나타난다. CH-BERT-KOSAC-small 모델을 결합했을 때 성능이 KR-BERT-KOSAC-small 모델을 결합했을 때보다 크게 높아지는 것이다. 반면 영화평 분류 문제에서는 이러한 차이가 뚜렷하게 나타나지 않는다. 이는 KR-BERT-KOSAC-small 모델의 학습 데이터가 위키피디아 데이터와 뉴스 기사 본문으로부터 추출되어 비교적 정형적인 문장들로 이루어져 있는 반면, CH-BERT-KOSAC-small 모델의 학습 데이터가 뉴스 기사에 대한 댓글로 구성되어 악플 탐지 데이터셋과 유사한 장르적 특성을 띠고 있기 때문이라고 판단된다. 댓글은 다양한 사람들에 의해 온라인에서 작성된 비정형적인 텍스트로서 줄임말, 신조어, 이모티콘, 오타자 등을 포함할 가능성이 높으며, 이런 형태적 특성 중 일부는 작성자의 주관이나 감정을 드러내는 장치로서 기능하기도 한다.

한편 일반 규모의 감정 자질 모델인 KR-BERT-KOSAC은 별도의 사전학습모델과의 결합 없이 단독으로 실험했을 때 감정 분석 문제에서 전반적으로 우수한 성능을 보였다. 그러나 외적 결합된 모델들의 성능이 대부분의 경우 단독 사용된 KR-BERT-KOSAC보다 높게 나타났다. 그에 따라 사전학습모델과 감정 자질 모델을 결합한 경우 감정 분석 문제의 수행 성능이 단독 모델을 사용했을 때보다 개선된다는 것을 알 수 있었다.

4.3 감정 자질 모델 분석

KR-BERT-KOSAC은 위에서 학습한 KR-BERT-KOSAC-small과 비교해서는 모델의 규모만이 확장된 것이고, 기존의 KR-BERT과 비교해서는 감정 자질이 추가되었다는 점이 차이점이 된다. 따라서 이 두 가지 모델의 비교를 통해 모델 규모 축소와 감정 자질 추가의 효과를 확인해볼 수 있을 것이다.

KR-BERT-KOSAC은 다른 조건은 모두 같은 상태에서 감정 자질만을 넣지 않은 모델인 KR-BERT에 비해 전반적으로 감정 분석 문제들에서의 성능이 향상되는 경향을 보인다.

표 2는 KR-BERT와 KR-BERT-KOSAC의 학습 성능을 나타낸 것이다. 표에서 MLM과 NSP는 각각 BERT 모델의 학습에 사용되는 서브태스크를 나타내고, acc는 정확도를 뜻한다. 두 모델의 학습 성능은 MLM에서 특히 차이를 보이는데, 감정 자질을 추가한 경우인 KR-BERT-KOSAC 모델에서 큰 폭으로 향상됨을 알 수 있다. 이에 따라 각 토큰에 추가된 감정 극성과 강도

표 1 외적 결합 모델의 감정 분석 문제 성능

Table 1 Task performances of externally-fused models

| | Movie Review | | Hate Speech | |
|----------------|--------------|--------------|--------------|--------------|
| | eval | test | eval | test |
| M-BERT | 87.08 | 86.82 | 73.39 | 52.03 |
| +KR-small | 89.34 | 89.03 | 75.00 | 54.23 |
| +CH-small | 89.39 | 89.18 | 80.22 | 55.63 |
| KorBERT | 90.48 | 89.81 | 78.74 | 54.33 |
| +KR-small | 90.75 | 90.49 | 77.70 | 55.64 |
| +CH-small | 91.00 | 90.68 | 79.34 | 57.49 |
| KR-BERT | 89.86 | 89.74 | 78.18 | 54.53 |
| +KR-small | 90.51 | 90.12 | 78.97 | 55.52 |
| +CH-small | 90.60 | 90.53 | 78.90 | 56.28 |
| KR-BERT-MEDIUM | 90.90 | 90.29 | 81.93 | 57.91 |
| +KR-small | 91.21 | 90.89 | 82.25 | 57.74 |
| +CH-small | 90.93 | 90.81 | 83.63 | 58.90 |
| KR-BERT-KOSAC | 90.30 | 89.82 | 80.00 | 53.81 |

표 2 KR-BERT와 KR-BERT-KOSAC의 학습 성능
Table 2 Training performances of KR-BERT and KR-BERT-KOSAC

| | KR-BERT | KR-BERT-KOSAC |
|----------|--------------------|-------------------|
| loss | 0.9154538 | 0.54173344 |
| MLM acc | 77.34 | 85.15 |
| MLM loss | 0.9511529 | 0.56670845 |
| NSP acc | 99.50 | 99.38 |
| NSP loss | 0.013327285 | 0.02148062 |

표 3 단일 자질 모델의 감정 분석 문제 성능
Table 3 Task performances of ablated models

| | Movie Review | | Hate Speech | |
|---------------|--------------|--------------|--------------|--------------|
| | eval | test | eval | test |
| KR-BERT | 89.86 | 89.74 | 78.18 | 54.53 |
| +polarity | 89.61 | 89.34 | 77.60 | 54.46 |
| +intensity | 89.85 | 89.37 | 78.00 | 55.84 |
| KR-BERT-KOSAC | 90.30 | 89.82 | 80.00 | 53.81 |

값이 추가적인 정보로서 문맥에 기반한 언어 모델링에 도움을 준다고 예상할 수 있다.

또한 본 연구에서 사용한 감정 자질인 극성과 강도의 효과를 파악하기 위하여, 두 자질 중 극성과 강도 각각 하나씩만을 포함한 감정 자질 모델을 학습하였다. 표 3은 이들 모델의 감정 분석 성능을 나타낸 것이며, 두 번째 행은 BERT 기본 임베딩에 극성 자질만을 추가한 모델의, 세 번째 행은 강도 자질만을 추가한 모델의 실험 결과를 보인다.

극성과 강도 단일 자질만을 포함시켜서 학습한 경우 두 가지 감정 분석 문제에서의 모델 성능은 향상되지 않고, 오히려 소폭 감소하였다. 그러나 두 가지 자질을 모두 이용한 KR-BERT-KOSAC 모델은 감정 자질을 포함하지 않은 KR-BERT 모델에 비해 개선된 성능을 보였다. 이러한 성능의 비교는 간접적인 방법이므로 감정 자질 각각이 BERT 임베딩에 정확히 어떤 영향을 미치는지는 알 수 없으나, 감정 극성과 강도 두 가지 자질이 함께 사용되었을 때 상호작용하여 모델의 감정 분석 성능을 향상시키는 데 도움을 주었으리라고 생각해 볼 수 있다.

표 4는 사전학습모델에 결합되는 감정 자질 모델의 규모에 따른 성능을 비교한 것이다. 기반이 되는 사전학습 모델은 표의 첫 번째 행에 있듯 KR-BERT로 설정하고, 표의 두 번째, 세 번째 행은 그에 각각 KR-BERT-KOSAC-small과 CH-BERT-KOSAC-small 모델을 결합한 결과를, 마지막 행은 KR-BERT-KOSAC을 결합한 결과를 나타낸다. KR-BERT-KOSAC은 소규모로 학습된 두 자질 모델과 달리 많은 사전학습모델과 같은 규모

표 4 감정 자질 모델의 규모에 따른 성능 비교
Table 4 Comparison of training performances of externally-fused models by scale

| | Movie Review | | Hate Speech | |
|----------------|--------------|--------------|--------------|--------------|
| | eval | test | eval | test |
| KR-BERT | 89.86 | 89.74 | 78.18 | 54.53 |
| +KR-small | 90.51 | 90.12 | 78.97 | 55.52 |
| +CH-small | 90.60 | 90.53 | 78.90 | 56.28 |
| +KR-BERT-KOSAC | 90.73 | 90.28 | 79.48 | 53.74 |

(“BERT-Base”)로 구성되었는데, 소규모 모델을 결합한 경우와 감정 분석 문제들의 성능에서 큰 차이를 보이지 않았다. 영화평 분류와 악플 탐지 문제의 테스트 데이터에서는 소규모 모델인 CH-BERT-KOSAC-small을 결합한 경우가 좀 더 높은 성능을 기록하기도 하였다. 이에 따라 감정 자질 모델을 소규모로 구성하더라도 결합 성능에 큰 영향을 끼치지 않는다는 것을 확인할 수 있다.

이러한 실험 결과를 토대로, 본 연구는 감정 자질의 부착이 모델의 감정 분석 성능에 도움이 된다는 것을 확인하였다. 또한 감정 자질 모델을 작은 규모와 적은 스텝수로 학습하여 사용했음에도 불구하고 확실한 성능 개선을 관찰하였다는 점에서, 비교적 짧은 학습 시간과 적은 자원을 사용하는 경제적인 방법을 사용하면서도 성능 향상의 효과를 얻을 수 있었다고 판단된다.

5. 지식 종류

앞서 4장까지의 실험과 분석을 통해, 본 연구가 제시한 사전 기반의 감정 정보와 이를 이용한 감정 자질 모델이 기존의 사전학습모델과 결합하여 감정 분석 문제에서 향상된 성능을 보인다는 것을 확인하였다. 특히 보편적 목적의 대규모 사전학습모델이 학습한 언어 지식과 감정 자질 모델이 포함하는 감정 정보가 결합 과정에서 상호 작용한다는 특성에 주목하여, 본 연구에서는 이러한 정보의 결합을 위해 지식 종류 기법을 이용한 방법론을 시도해 보았다. 그림 4에서 볼 수 있듯 보편적 목적의 사전학습모델을 큰 모델(teacher model)로, 본 연구가 학습한 감정 자질 모델을 작은 모델(student model)로 하여, 후자가 감정 분석 문제 및 데이터셋에 대한 모델의 미세 조정을 진행함과 동시에 큰 모델이 학습한 언어 지식을 전달받아서 감정 분석의 성능을 개선시킬 수 있도록 하였다.

미세 조정 과정에서 작은 모델은 데이터셋으로부터 구성한 문장 임베딩을 큰 모델이 구성한 임베딩과 의미 공간 상의 거리가 가까워지도록 학습하며, 여기에는 코사인 유사도에 기반한 loss가 사용된다. 또한 작은 모델은 일반적인 BERT 모델의 미세 조정 과정에서처럼 데

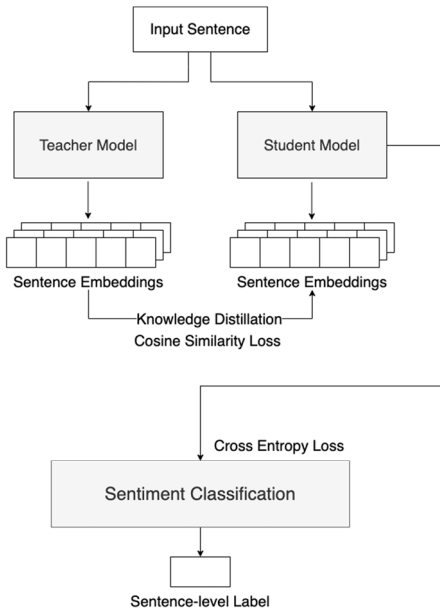


그림 4 지식 증류 방식

Fig. 4 Knowledge Distillation Method

이터셋의 정답을 예측해 나가며, 이때 교차 엔트로피에 기반한 loss를 사용한다.

본 연구는 이러한 지식 증류 과정을 Reimers and Gurevych(2020)[20]가 공개한 코드⁸⁾를 참고하여 구현하였다. 해당 코드의 구조에 따라, 큰 모델과 작은 모델 모두 불러온 뒤 풀링 모델(pooling model)이 추가된 형태로 이용된다. 이때 풀링 모델로는 텍스트 의미 처리에 문장 내 모든 토큰 임베딩의 평균을 이용하는 평균 풀링(mean pooling) 모델을 이용하였다.

앞서 4장의 실험에서 수행한 영화평 분류와 악플 탐지 문제에 대하여 이러한 지식 증류 방식의 감정 분석 능력을 확인하고자 한다. 먼저 큰 모델로는 KR-BERT-MEDIUM 모델에 평균 풀링을 추가한 모델인 KR-SBERT⁹⁾를, 작은 모델로는 본 연구에서 학습한 감정 자질 모델인 CH-BERT-KOSAC-small을 선택하였다. 이들 모델은 4장의 실험에서 전반적인 성능이 가장 높게 나타난 조합에 따른 것이다. 또한 다른 모델 조합을 이용해서도 실험해 보았는데, 한국어 댓글 데이터셋에 기반하여 학습된 KcBERT[22]를 큰 모델로 하고, 본 연구에서 학습한 KR-BERT-KOSAC 모델을 작은 모델로 하였다. 이때 악플 탐지 문제는 테스트 데이터의 정답이 공개되어 있지 않은 리더보드 형태이므로 검증

데이터에 대한 F1 점수만을 도출하였다.

모든 경우 AdamW optimizer를 사용하였고, 학습 배치사이즈를 64, 학습률을 2e-5로 설정하였으며 10 에폭 동안 학습을 수행하였다. 데이터 시퀀스의 최대 길이는 영화평 분류의 경우 50, 악플 탐지의 경우 60으로 설정하였는데, 이는 이전의 휴리스틱한 방법으로 설정한 128과 달리 패딩되는 토큰의 수를 줄이기 위한 것이다. 영화평 분류 데이터 시퀀스 길이가 중간값 27.0과 75분위 값 43.0, 악플 탐지 데이터의 경우 중간값 31.0과 75분위 값 53.0으로 나타났기 때문이며, 특히 공개된 PyTorch 기반 한국어 BERT 모델들이 설정한 영화평 길이의 최대 데이터 시퀀스 길이가 대부분 50으로 설정된 점 역시 고려한 것이다. 이에 따른 감정 자질 모델의 지식 증류 학습은 2개의 Tesla V100 GPU를 이용하여 영화평 분류 데이터의 경우 약 80시간, 악플 탐지 데이터의 경우 약 4시간 소요되었다.

본 연구에서 시도한 지식 증류 방법론의 감정 분석 실험 결과는 다음 표 5와 같이 나타난다.

먼저 영화평 분류 문제에서, KR-SBERT가 학습한 언어 지식을 전이받은 CH-BERT-KOSAC-small 모델(표 5에서 KR-SBERT→CH-small로 표시)은 검증 데이터와 테스트 데이터의 예측 정확도가 모두 50 이하로 나타났다. 또한 KcBERT를 큰 모델로 한 KR-BERT-KOSAC 모델(표 5에서 KcBERT→KR-KOSAC-BERT로 표시) 역시 50 정도의 낮은 예측 정확도를 보였다. 이는 앞서 표 1에서 보고한 사전학습모델 및 외적 결합 방법론의 성능이 90 전후인 것을 고려하면 매우 저조한 수준이다. 한편 악플 탐지 문제에서는 두 가지 감정 자질 모델 모두 검증 데이터에서 70.70과 76.50의, 기존 성능에 비교적 가까운 F1 점수를 기록하였으나 여전히 개선되지 못한 성능을 보였다.

특히 모든 경우에서 10 에폭 동안의 예측 정확도와 F1 점수의 변화폭이 크지 않았으므로, 지식 증류 과정에서 큰 모델의 언어 지식이 작은 모델로 충분히 전달되지 못했을 가능성이 있다. 현재는 두 가지 모델 임베딩 사이의 거리를 좁히는 것이 코사인 유사도에만 의존하고 있는데, 추후 작은 모델의 문장 임베딩을 큰 모델

표 5 지식 증류 모델의 감정 분석 문제 성능

Table 5 Task performances of knowledge-distilled models

| | Movie Review | | Hate Speech | |
|----------------------|--------------|-------|-------------|------|
| | eval | test | eval | test |
| KR-SBERT→CH-small | 49.74 | 43.11 | 70.70 | - |
| KcBERT→KR-KOSAC-BERT | 50.24 | 50.47 | 76.50 | - |

8) <https://github.com/UKPLab/sentence-transformers>

9) <https://github.com/snunlp/KR-SBERT>

의 문장 임베딩에 가까워지도록 학습하는 과정을 좀 더 정제된 방법으로 구성해 보고자 한다. 한편 본 연구가 고려한 두 가지 모델은 기존 Reimers and Gurevych (2020)에서와 달리 두 모델이 서로 다른 목적으로 학습 되어 각각 포함하는 언어 지식이 다를 것으로 기대되므로, 임베딩 간의 거리를 감소시키는 것 외의 지식 전이 방법 역시 고려해 보아야 할 것이다.

6. 결론 및 향후 연구

본 연구에서는 BERT 모델에 감정 자질을 추가하는 새로운 방법을 제안하였다. 한국어 감정 사전으로부터 구성된 감정 자질을 더한 모델은 기존의 사전학습모델과 외적 결합되었을 때 감정 분석 문제에서 각 사전학습모델을 단독으로 사용하였을 때보다 높은 성능을 보였다. M-BERT, KorBERT, KR-BERT, KR-BERT-MEDIUM 네 가지의 공개된 모델들을 이용하여 감정 자질 모델의 사용이 성능 향상에 도움이 된다는 것을 관찰하였다. 본 연구에서 구현한 감정 자질 모델은 비교적 적은 스태프로 학습한 소규모 모델로, 감정 자질을 추가하면서도 계산상의 비용과 학습에 드는 시간을 줄이고자 한 것이다. 한편 본 연구는 추가적으로 두 가지 모델을 지식 종류 방식으로 결합하는 방법 역시 시도하였으나, 앞서 외적 결합 방식을 적용했을 때와 달리 감정 분석 문제에서의 성능 개선을 관찰할 수 없었다. 현재는 본 연구가 제안하는 두 가지 BERT 기반 모델의 외적 결합이 간단하게 구성된 방법하면서도 가장 높은 성능을 보고하고 있으며, 지식 종류 방법론의 효과에 대해서는 아직 확실히 결론지을 수 없으나 추가 연구를 통해 더 발전되어야 할 것이다.

한편 본 연구는 사전학습모델에 감정 자질을 추가 사용할 경우 모델 학습 중 언어 모델링 과정의 성능과 관련 자연어 처리 문제의 성능에 긍정적인 역할을 할 것이라고 예상하고 이를 확인하였다. 그러나 이것은 정확한 loss, F1 점수 등의 척도를 통해 간접적으로 판단한 결과이다. 따라서 BERT 모델이 내부적으로 다양한 언어 현상을 어떻게 포착하고 있는지를 검증하는 연구들을 참고하여, 감정 자질과 그에 기반한 모델을 결합하는 방법의 영향을 더욱 구체적으로 밝혀 나가려고 한다. 또한 모델 학습 데이터의 크기와 장르 다양성을 확장하고, 감정 자질과 그를 이용한 모델을 기존의 사전학습모델과 결합하는 더 효과적인 방법에 대해 연구를 계속하고자 한다.

References

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers

for language understanding," *Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 4171–4186, 2018.

[2] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," *Proc. of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014.

[3] J. Zhou, Z. Zhang, H. Zhao, and S. Zhang, "LIMIT-BERT: Linguistic informed multi-task bert," arXiv preprint arXiv:1910.14296, 2019.

[4] Z. Zhang, X. Han, Z. Liu, X. Jiang, M. Sun, and Q. Liu, "ERNIE: Enhanced language representation with informative entities," arXiv preprint arXiv:1905.07129, 2019.

[5] L. Huang, C. Sun, X. Qiu, and X. Huang, "GlossBERT: BERT for word sense disambiguation with gloss knowledge," arXiv preprint arXiv:1908.07245, 2019.

[6] Y. Levine, B. Lenz, O. Dagan, D. Padnos, O. Sharir, S. Shalev-Shwartz, A. Shashua, and Y. Shoham, "Sensebert: Driving some sense into bert," arXiv preprint arXiv:1908.05646, 2019.

[7] P. Ke, H. Ji, S. Liu, X. Zhu, and M. Huang, "Sentilare: Linguistic knowledge enhanced language representation for sentiment analysis," *Proc. of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 6975–6988, 2020.

[8] H. Tian, C. Gao, X. Xiao, H. Liu, B. He, H. Wu, H. Wang, and F. Wu, "Skep: Sentiment knowledge enhanced pre-training for sentiment analysis," arXiv preprint arXiv:2005.05635, 2020.

[9] H. Shin, M. Kim, H. Jang, and A. Cattle, "Annotation scheme for constructing sentiment corpus in korean," *Proc. of the 26th Pacific Asia Conference on Language, Information, and Computation*, 2012.

[10] L. Martin, B. Muller, P. J. O. Suárez, Y. Dupont, L. Romary, É. V. de la Clergerie, D. Seddah, and B. Sagot, "Camembert: a tasty french language model," arXiv preprint arXiv: 1911.03894, 2019.

[11] W. de Vries, A. van Cranenburgh, A. Bisazza, T. Caselli, G. van Noord, and M. Nissim, "BERTje: A Dutch BERT Model," arXiv preprint arXiv:1912.09582, 2019.

[12] A. Virtanen, J. Kanerva, R. Ilo, J. Luoma, J. Luotolahti, T. Salakoski, F. Ginter, and S. Pyysalo, "Multilingual is not enough: Bert for finnish," arXiv preprint arXiv: 1912.07076, 2019.

[13] S. Lee, H. Jang, Y. Baik, S. Park, and H. Shin, "A Small-Scale Korean-Specific BERT Language Model," *Journal of KIISE*, Vol. 47, No. 7, pp. 682–692, 2020. (in Korean)

[14] Z. Teng, D. T. Vo, and Y. Zhang, "Context-sensitive lexicon features for neural sentiment analysis," *Proc. of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1629–1638,

- 2016.
- [15] Q. Sun, Z. Wang, Q. Zhu, and G. Zhou, "Stance detection with hierarchical attention network," *Proc. of the 27th International Conference on Computational Linguistics*, pp. 2399-2409, 2018.
- [16] L. Bao, P. Lambert, and T. Badia, "Attention and lexicon regularized LSTM for aspect-based sentiment analysis," *Proc. of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pp. 253-259, 2019.
- [17] Y. Li and C. Caragea, "Multi-task stance detection with sentiment and stance lexicons," *Proc. of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 6300-6306, 2019.
- [18] Y-C. Chen, Z. Gan, Y. Cheng, J. Liu, and J. Liu, "Distilling Knowledge Learned in BERT for Text Generation," *Proc. of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 7893-7905, 2020.
- [19] X. Tan, Y. Ren, D. He, T. Qin, Z. Zhao, and T-Y. Liu, "Multilingual Neural Machine Translation with Knowledge Distillation," *Proceedings of the 7th International Conference on Learning Representations*, 2020.
- [20] N. Reimers and I. Gurevych, "Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation," *Proc. of the 2020 Conference on Empirical Methods in Natural Language Processing*, pp. 4512-4525, 2020.
- [21] J. Moon, W.-I. Cho, J. Lee, "BEEP! Korean corpus of online news comments for toxic speech detection," *Proc. of the Eighth International Workshop on Natural Language Processing for Social Media*, 2020.
- [22] J. Lee, "KcBERT: Korean Comments BERT," *Proc. of the 32nd Annual Conference on Human and Cognitive Language Technology*, pp. 437-440, 2020. (in Korean)



이 상 아

2013년 서울대학교 언어학과, 컴퓨터공학부 졸업(학사). 2016년 서울대학교 언어학과 졸업(석사). 2021년 서울대학교 언어학과 졸업(박사). 2021년 3월~현재 서울대학교 데이터사이언스 대학원 박사후연구원. 관심분야는 자연어처리, 문장 임베딩 모델, 감정분석, 논증 마이닝



신 효 필

1988년 서울대학교 언어학과 졸업(학사)
1990년 서울대학교 언어학과 졸업(석사)
1994년 서울대학교 언어학과 졸업(박사)
1997년 12월 University of Missouri, Computer Science 졸업(석사). 1998년 1월~2001년 1월 Computing Research Lab, New Mexico State University. 2001년 1월~2001년 12월 YY Technology in Silicon Valley. 2001년 9월~2003년 2월 서울대학교 전자공학부 BK교수. 2003년~2020년 1월 서울대학교 인문대학 언어학과 교수. 2020년 1월~현재 서울대학교 데이터사이언스 대학원 교수. 관심분야는 사전학습모델 구축과 이를 활용한 자연어처리.