

BERT 및 계층 그래프 컨볼루션 신경망 기반 감성분석 모델

장준준* · 신종호 · 안수빈 · 박태영 · 노기섭

청주대학교

BERT & Hierarchical Graph Convolution Neural Network based Emotion Analysis Model

Junjun Zhang* · Jongho Shin · Suvin An · Taeyoung Park · Giseop Noh

Cheongju University

E-mail : {zjj320, sjh0804_22 / ssubin0405 / mu07010 / kafa46}@cju.ac.kr

요 약

기존 텍스트 감성 분석 모델에서는 일반적으로 전체 텍스트를 직접 모델링하고, 텍스트 내용 간의 계층적 관계를 덜 고려한다. 그러나 감성분석의 구현에서는 많은 텍스트가 여러 감정으로 뒤섞여 있다. 전체의 의미론적 모델링을 직접 수행하면 감성분석 모델의 판단 난도가 높아져 혼합 감정 문장의 분류에 적용하기 어려울 수 있다. 따라서 본 논문에서는 텍스트 계층을 고려한 감성 분석 모델 BHGCN을 제안한다. 이 모델에서는 BERT의 각 레이어의 숨겨진 상태의 출력이 노드로 사용되며, 상위 레이어와 하위 레이어 사이에 직접 연결이 이루어져 의미 계층이 있는 그래프 네트워크를 구축한다. BHGCN 모델은 계층별 의미론에 주의를 기울일 뿐만 아니라 계층적 관계에도 주의를 기울이기 때문에 혼합 감성 분류 작업을 처리하는 데 적합하다. 본 논문에서는 비교 실험을 통해 제안하는 BHGCN 모델이 명백한 경쟁 우위를 보인다는 것을 입증하였다.

ABSTRACT

In the existing text sentiment analysis models, the entire text is usually directly modeled as a whole, and the hierarchical relationship between text contents is less considered. However, in the practice of sentiment analysis, many texts are mixed with multiple emotions. If the semantic modeling of the whole is directly performed, it may increase the difficulty of the sentiment analysis model to judge the sentiment, making the model difficult to apply to the classification of mixed-sentiment sentences. Therefore, this paper proposes a sentiment analysis model BHGCN that considers the text hierarchy. In this model, the output of hidden states of each layer of BERT is used as a node, and a directed connection is made between the upper and lower layers to construct a graph network with a semantic hierarchy. The model not only pays attention to layer-by-layer semantics, but also pays attention to hierarchical relationships. Suitable for handling mixed sentiment classification tasks. The comparative experimental results show that the BHGCN model exhibits obvious competitive advantages.

키워드

Sentiment Analysis; BHGCN Model; BERT; Hidden Layer; GCN

1. 서 론

NLP에서 가장 고전적인 과제 중 하나인 감성분석은 항상 학자들의 관심을 끌었다. 최근 몇 년 동

안 딥 러닝의 급속한 발전은 감성분석 기술의 연구를 성공적으로 촉진시켰다. 기존 텍스트 감성분석 모델은 일반적으로 전체 텍스트를 직접 모델링하는 반면, 문장 의미론 간의 상관관계에 대한 고려가 적다. 그러나 감성분석의 많은 텍스트는 긍정적인 감성, 부정적인 감성, 그리고 중립적인 감성

* speaker

이 혼합되어 있다. 혼합 감성 문장의 직접적인 의미론적 모델링은 감성 모델에서 판단의 어려움을 증가시킬 수 있다. 직관적으로 문장의 구조적 관계를 고려한다면 감성적 극성의 판단에 도움이 될 것이다. 최근, GNN[1]은 텍스트의 구조 정보를 모델링하는 데 매우 효과적인 것으로 나타났다. Wu, L. 등은 [2]에서 텍스트를 표현을 위해 그래프를 사용하는 것이 더 잘 얻어질 수 있음을 설명하기 위해 그래프 구성, 그래프 표현 학습 및 그래프 기반 인코더-디코더 모델의 세 가지 측면에서 NLP에 대한 GNN 연구를 종합적으로 진행하였다. 기존 연구는 대규모 사전 훈련된 모델이 그래프 구조 학습[3]의 잠재력을 활용하여 성능향상에 도움이 되는 것으로 알려져 있다.

본 논문에서는 관련 연구를 바탕으로 BERT-base 와 GCN 기반으로 감성 분석 모델을 제안하여 문장 수준 관계인 BERT and Hierachical Graph Convolution Network (BHGCN)을 제안한다. 이 논문의 기여는 한편으로는 계층적 정보에 대한 BERT 기반 인코딩 방법을 제안하는 것입니다. 다른 한편으로 각 문장의 의미 정보를 추출하기 위해 원래 BERT 계층을 추가하면서 BERT 레이어와 문장 사이의 계층적 관계를 기반으로 하는 계층적 그래프 구조가 도입하는 것이다. 마지막으로, 두 가지 방법에 따라 문장의 의미 정보와 문장 간의 구조적 관계를 모두 고려하는 새로운 통합 BHGCN 모델을 제안한 것이다.

II. 관련 연구

기존 감성분석 방법은 일반적으로 감성사전 기반 방법, 전통적인 기계학습 기반 방법 및 딥러닝 기반 방법으로 구분한다. 처음 두 가지 방법은 시간이 많이 걸리고 노동 집약적이며 비용이 많이 드는 기능 엔지니어링에 크게 의존한다. 최근 몇 년 동안 인공지능 기술의 발달로 감성분석 연구는 점차 딥러닝으로 전환되고 있다. 특히 BERT [4] 사전훈련된 언어모델의 등장은 사전훈련과 다운스트림 작업으로 생성된 단어 벡터 간의 관계를 완전히 변화시켰고, 다수 NLP 작업에서 최상의 결과를 달성했다. 그러나 BERT는 텍스트의 내부 구조 정보를 마이닝 하는데 일정한 제한이 있다. 최근 GNN[1]은 텍스트의 구조 정보를 명시적으로 모델링하여 기계번역, 감성분류 및 문법 검출과 같은 일부 NLP 작업에서 최첨단 결과를 달성하는 것으로 나타났다. 우리는 [5]에서 영감을 받아 사전 훈련된 모델과 그래프 네트워크 모델을 최대한 활용하여 감성분석 작업을 위한 BHGCN 모델을 제안한다. BHGCN 모델은 문장의 의미 정보를 완전히 특성화할 수 있을 뿐만 아니라 문장 간의 구조적 관계를 포착할 수 있다. 본 논문에서 제시한 모델링 구조는 혼합 감성분류 작업을 더 잘 처리할 수 있다.

III. 모델 아키텍처

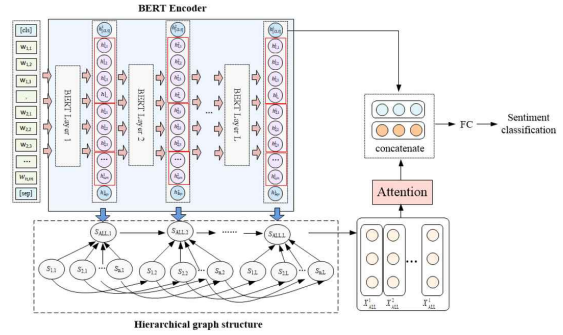


그림 1. BHGCN 모델 구성도

가. BERT 정보기반 인코딩

입력한 문장이 $S=\{s_1, s_2, \dots, s_n\}$ 일 때 s_i 는 i 번째 절(일정 수 단어가 결합된 문장의 구성단위)이다. 우리는 문장의 처음과 끝에 각각 [CLS]와 [SEP] 토큰을 추가하여 BERT의 입력 포맷으로 맞춘 다음 BERT의 L 번째 레이어를 이용해서 인코딩한다. BERT의 j 번째 레이어의 경우, 각 레이어의 은닉층의 representation H^j 를 얻을 수 있다.

$$H^j = BERT^j(x) \tag{1}$$

$$= [x_{[CLS]}^j, x_{1,1}^j, x_{1,2}^j, \dots, x_{n,1}^j, x_{[SEP]}^j],$$

$$\text{where } H^j \in R^{(n+2) \times d}$$

수식 (1)에서 d 는 은닉 레이어의 벡터 크기이며, l 은 절의 최대 단어 수이며, n 은 문장에서 절의 수이다. 문장의 수이다. 평균 풀링 연산을 통해 j 번째 레이어에서 i 번째 절의 의미론적 표현은 수식 (2)와 같다(그림 1의 빨간색 박스 참고).

$$h_i^j = \text{MeanPooling}([x_{i,1}^j, x_{i,2}^j, \dots, x_{i,l}^j]),$$

$$h_i^j \in R^{1 \times d}$$

(2)

나. GCN 기반 계층형 그래프 네트워크

이 논문은 그림 1의 하반부에 표시된 것처럼 계층형 그래프 네트워크를 구성한다. 여기서 V 는 노드 집합이고 E 는 간선 집합이다. 네트워크 계층의 수는 L 이며, 각 계층은 BERT의 각 계층에 해당한다. 그래프의 노드는 하위 문장 노드와 전체 문장 노드이며 상위 수준 노드와 하위 수준 노드는 그림에 제시된 방식으로 연결한다. 하위 문장 노드의 초기 특징은 BERT기반 인코딩하여 얻고, 전체 문장 노드의 특징은 무작위로 초기화한다. 노드 집계 방법은 수식 (3)과 같다.

$$H_g^l = (\tilde{A}\tilde{D}^{-1}XW^l) \quad (3)$$

\tilde{A} 는 자체 연결을 포함하는 그래프 인접 행렬, \tilde{D} 는 그래프의 정도(node degree) 행렬, X 는 특징(feature) 행렬, W^l 은 레이어별 훈련가능한 가중치(weight params)이다.

문장 융합 레이어는 각 레이어의 문장 노드의 특징을 추출하고 Attention 메커니즘을 통해 융합하는 레이어로 구현 방법은 수식 (4)와 같다.

$$H_{structure} = \sum_{i=1}^l W_i X_{ALL}^i \quad (4)$$

여기서 W_i 는 훈련가능한 가중치 행렬이며, $X_{ALL} = [X_{ALL}^1, X_{ALL}^2, \dots, X_{ALL}^l]$ 이다. 이후 BERT 마지막 레이어의 [CLS] 토큰과 수식 (4)의 $H_{structure}$ 를 연결하고, 선형 레이어를 통해 차원 축소 후 Softmax를 통해 최종 예측을 수행한다.

IV. 실험 및 분석

BHGCN 모델의 성능을 검증하기 위해 공개 데이터 세트 MR을 사용하였다. MR 데이터 세트에 대한 자세한 내용은 Github 해당 페이지에서 확인할 수 있다. 성능비교를 위한 베이스라인 아키텍처는 TextCNN, BERT, TextGCN을 선정하였다. 실험 설정은 배치 크기는 16, 옵티마이저는 Adam, 학습률은 $1e-5$, 구의 수는 10, 최대 단어수는 100으로, Epoch은 100으로 설정하였다. 과적합 방지 및 훈련 시간 단축을 위해 Early Stop을 적용하였다. Early Stop은 F1-score가 3회 이상 향상되지 않으면 학습을 종료하는 것으로 설정하였다. 실험 결과는 표 1에 제시하였다.

표 1. BHGCN과 베이스라인 모델 비교

Model	Accuracy	Precision	Recall	Micro-F1
BERT-TextCNN	77.69	77.7	77.71	77.69
BERT	86.22	86.21	86.42	86.2
TextGCN	76.74	76.77	76.75	76.74
BHGCN (ours)	86.32	86.31	86.32	86.32

표 1에서, BHGCN은 비교대상 모델보다 높은 성능을 보여주었다. BHGCN 성능은 BERT가 약간 높으며, 다른 모델보다 월등하다. TextCNN 모델은 BERT에서 좋은 초기 기능을 얻지만 문장 구조 파악에서 성능이 여전히 낮습니다. TextGCN은 다른 기준 모델보다 더 나은 결과를 보여주지 않는데, 가능한 이유는 GCN이 텍스트 작업에서 매우 중요한 단어 순서 정보를 무시하기 때문이다.

V. 결론

본 논문에서는 현재 감성 분석에 존재하는 문제 해결을 목표로 대규모 사전 학습된 모델과 그래프 신경망의 장점을 결합하여 새로운 모델 BHGCN을 제안하였다. 제안된 BHGCN 모델은 BERT에서 의미론적 특징을 추출하고 GCN에서 구조적 특징을 추출한다. 이전 연구 방법론과 달리, 본 논문에서 구성하는 그래프 네트워크는 BERT의 각 레이어의 hidden states 출력을 기반으로 구축한다. 한편, BERT의 각 계층의 정보를 충분히 활용하여 정보 손실을 방지할 수 있다. 반면에, 확립된 단단계 계층 관계 다이어그램은 각 계층의 의미론뿐만 아니라 계층 관계도 고려할 수 있다. 결과적으로 BHGCN 모델의 혼합 감성분류에서 더 좋은 성능을 달성할 수 있다. 실제 성능은 공개 데이터를 통한 실험을 통해 입증하였다.

References

- [1] P. W. Battaglia, Hamrick, J. B. Bapst, V. Sanchez-Gonzalez, A. Zambaldi, V. Malinowski, M. ... & R. Pascanu, "Relational inductive biases, deep learning, and graph networks," arXiv preprint arXiv:1806.01261, 2018.
- [2] L. Wu, Y. Chen, K. Shen, X. Guo, H. Gao, S. Li, ... & B. Long, "Graph neural networks for natural language processing: A survey," arXiv preprint arXiv:2106.06090, 2021.
- [3] Y. Lin, Y. Meng, X. Sun, Q. Han, K. Kuang, J. Li, F. Wu, "BertGCN: Transductive Text Classification by Combining GCN and BERT," ACL-IJCNLP, August 1-6, online meeting, 2021.
- [4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018.
- [5] Z. Xiao, J. Wu, Q. Chen, & C. Deng, "BERT4GCN: Using BERT Intermediate Layers to Augment GCN for Aspect-based Sentiment Classification." arXiv preprint arXiv:2110.00171, 2021.