



감성 사전을 활용한 트랜스포머 모델 기반의 감정분석

Text sentiment Analysis Based on Transformer Models using an emotional dictionary

저자 (Authors)	황소현 Sohyun Hwang
출처 (Source)	한국정보과학회 학술발표논문집 , 2021.6, 876-878 (3 pages)
발행처 (Publisher)	한국정보과학회 The Korean Institute of Information Scientists and Engineers
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE10583106
APA Style	황소현 (2021). 감성 사전을 활용한 트랜스포머 모델 기반의 감정분석. 한국정보과학회 학술발표논문집, 876-878.
이용정보 (Accessed)	한성대학교 220.66.103.*** 2021/08/16 05:00 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독 계약을 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.

감성 사전을 활용한 트랜스포머 모델 기반의 감성분석

황 소 현

고려대학교 컴퓨터정보통신대학원

sohyunhwang@korea.ac.kr

Text sentiment Analysis Based on Transformer Models using an emotional dictionary

Sohyun Hwang

Graduate School of Computer & Information Technology, Korea University

요 약

소셜미디어와 기술의 빠른 성장으로 인하여 비정형데이터인 문서, 댓글 등의 텍스트 데이터의 감성극성을 분류하는 감성분석 연구가 관심을 받아왔다. 초기의 감성분석 연구에서는 대부분 전통적인 단일 기계학습기법에 의존한 감성분류를 시도하였기에 분류 정확도 면에서 한계점이 있었다. 전통적인 기계학습기법 대신 대용량 데이터의 처리에 우수한 성능을 보이는 딥러닝 기법과 인공지능을 통한 감성분석(Sentiment Analysis) 분야에서 연구가 활발하게 진행되고 있다. 본 논문에서는 LSTM부터 한국어 BERT(Bidirectional Embedding Representations from Transformers)인 KoBERT, ELECTRA 등 다양한 트랜스포머 모델을 활용하여 감성분석을 해보고 감성사전을 활용하여 더 좋은 성능을 낼 수 있는 감성 분석 방법을 제안한다.

1. 서론

SNS와 인터넷이 활발해지면서, 사용자들은 온라인 상에서 댓글, 평점 등으로 의견을 자유롭게 표현할 수 있다. 이렇게 생성된 많은 데이터를 활용하여 감성 분석(Sentiment analysis)을 하면 인터넷 가짜 뉴스 판별, 고객 선호도 분석 등을 객관적으로 할 수 있어 서비스, 비즈니스, 브랜드 마케팅 등 다양한 분야에서 유용하게 활용되고 있다.

초기의 선행 연구들은 전통적인 기계학습 모델인 SVM, 나이브 베이즈 등을 이용한 감성 분석이 주를 이루었다. 하지만 분류의 정확도 측면에서 한계가 있고, 딥러닝 모델을 활용한 연구가 우수한 성과를 내고 있다.[1] 하지만 딥러닝을 수행하기 위해서는 많은 양의 데이터로 학습해야 좋은 모델을 만들 수 있다는 단점이 있다.

본 논문에서는 데이터에 따른 성능 비교와, 모델에 따른 성능을 비교하고자 한다. 데이터는 기존에 구축된 감성사전을 학습데이터로 사용하는 것과 사용하지 않는 것을 비교하고, 모델은 BERT와 한국어 기반 BERT모델인 KoBERT와 같이 사전 훈련된 모델들을 통해 트랜스포머 기반 모델이 기존 기계학습 모델에 비해 얼마나 성능의 향상이 있었는지, 학습데이터에 따른 성능 비교를 분석한다.

2. 관련 연구

2.1 감성사전

한국어 감성사전으로는 KNU 한국어 감성사전[2]이 있다. 표준국어대사전, SentiWordNet 및 SenticNet-5.0 등에서 주로 사용되는 긍/부정어를 추출하고 온라인에서

많이 사용되는 축약어, 긍/부정 이모티콘 목록을 통해 각 단어의 뜻을 분석하여 구축되었다. 1-gram, 2-gram, n-gram, 축약어, 이모티콘 등의 다양한 종류의 긍/부정어가 포함된 것이 특징이며, 표준국어대사전을 구성하는 모든 단어의 긍/부정 판별은 극성에 따라 매우 부정, 부정, 중립, 긍정, 매우 긍정으로 5단계로 나뉜다.

2.2 기계학습 및 딥러닝 기법을 이용한 감성분석

전통적인 기계학습에서는 분류에 대한 특징을 따로 추출해야 하지만, 딥러닝 모델들은 이미 주어진 단어특징 벡터를 활용해 모델을 학습하지 않고, 텍스트 정보를 입력해서 문장에 대한 특징 정보를 추출하기 때문에 분류의 정확도가 개선되는 장점이 있다.

국내에서는 한국어 문서에 적용하는 연구도 진행되어 왔다. 한국어 문서를 형태소 분석기를 통해 문장을 형태소 단어 단위로 토큰화하고 각 단어를 word2vec을 이용하여 임베딩하여 CNN 구조를 이용한 문서 분류 연구[3]와 RNN 구조를 이용한 문서 분류 연구[4]가 진행되었다.

또한 CNN-LSTM 조합을 이용하여 양자의 장점을 활용한 감성분석 연구[5]가 있다.

2.3 Transformer Model

대표적인 트랜스포머 모델은 2018년 구글에서 오픈소스로 공개된 BERT(Bidirectional Encoder Representation from Transformers)가 있다. 라벨링 되지 않은 데이터로 미리 학습시킨 후, 특정 자연어처리 태스크에 따라 별도의 아키텍처 없이 하나의 태스크 처리를 위한 레이어만 추가함으로써 지도학습을 수행하도록 하는 전이 학습 모델(Transfer Learning)이다. 사전학습(pre-training)과 미세 조정(fine-tuning)때의 아키텍처를 다르게 하여 전

이 학습을 쉽게 만드는 것이다.

SKT Brain의 KoBERT[6]는 기존 BERT 모델에 한국어 위키 데이터와 뉴스 2천만 문장을 추가 학습하여 한국어 데이터에 잘 대응할 수 있는 모델이다.

ELECTRA모델[7]은 구글에서 2020년 공개한 모델로, RTD(Real Token Detection) 방식의 새로운 기법을 적용하여 동일한 컴퓨터예산으로 성능을 향상시킨 새로운 사전 훈련 방식이다.

3. 실험 환경 및 실험 데이터

3.1 실험 환경

실험 환경으로는 GPU사용을 위하여 구글의 colab에서 수행하였다. 실험을 위한 모델은 구글의 오픈소스 라이브러리인 TensorFlow를 사용하였고, BERT와 SKT에서 만든 KoBERT, ELECTRA모델을 사용하였다. 최적화(Optimizer) 함수는 AdamW를 사용했으며, 모델 구현은 huggingface 라이브러리를 사용했다.

3.2 실험 데이터

실험데이터로는 NSMC(Naver sentiment movie corpus v1.0)데이터와 감정사전, 드라마 Friends 대본, IMDB 데이터를 사용하였다. 훈련 시 추가로 사용할 KNU 감성사전과 IMDB 데이터는 각 모델에 주어진 훈련데이터와 동일하게 라벨을 치환하여 개발을 수행한다.

3.2.1 한국어 데이터

네이버 영화리뷰 데이터 NSMC는 학습데이터 15만개, 테스트데이터 5만개로 구성된 총 20만개의 한글 리뷰 데이터로, 해당 영화리뷰가 중립의견은 제외하고, 긍정일 때 '1', 부정일 때 '0' 으로 표시된 레이블로 구성된다. 데이터 셋의 감성 레이블 통계는 다음 표 1과 같다.

표 1 네이버의 감성 레이블 수

감정분류		개수	비율
Train	Positive	74,827	49.9%
	Negative	75,173	51.1%
Test	Positive	25,173	50.3%
	Negative	24,827	49.7%

또한 모델을 학습시키기 위한 추가 데이터로 KNU한국어 감성 사전을 사용하였다. 0점 미만은 부정, 0점 초과는 긍정으로 치환하여 학습을 진행(0점은 배제)하였다. 공백은 제거하고, 정규 표현식에 어긋나는 데이터들을 제거한 결과 표 2와 같은 감성 레이블 분포를 보였다.

표 2 감성사전의 감성 레이블 수

감정분류		개수	비율
Positive	매우 긍정(2)	2,603	17.8%
	긍정(1)	2,246	15.3%
Negative	부정(-1)	5,016	34.2%
	매우 부정(-2)	4,799	32.7%

3.2.2 영어 데이터

미국드라마 Friends의 대본 데이터는 등장인물의 발화에 대한 감정 레이블이 담겨 있다. 학습용 10,561개, 검증용 1,178개, 테스트용 2,764개, 총 14,503개의 데이터로 구성되며 neutral, non-neutral, joy, sadness, fear, anger, surprise, disgust의 8개 감정이 존재하는 다중분류 문제이다. 학습 데이터로 함께 사용한 IMDB 리뷰 데이터는 스탠포드대학교에서 발표한 2011년에 논문에서 사용된 데이터로 25,000개의 텍스트와 해당 리뷰가 긍정일 때 '1', 부정일 때 '0' 으로 표시한 레이블로 구성된 데이터이다. 부정은 'angry' 로 긍정은 'joy' 로 치환하여 학습데이터로 실험하였다.

4. 실험결과

4.1 한국어 감정분석 모델

본 모델인 LSTM 모델과 bi-LSTM모델, 한국어 전용 모델인 KoBERT 모델을 비교하였다. BERT기반 모델의 학습 진행은 NSMC데이터와 추가데이터인 감정사전 데이터에 미세 조정(fine-tuning)을 통해 진행하였다. 모든 모델들은 형평성을 위해 입력 토큰의 길이를 각각 32, 64, 128로 나누어 진행하였다. LSTM모델과 bi-LSTM모델 학습 진행시 해당 문장을 형태소 단위로 자르는 모듈로 코모란(Komoran)을 사용하였다.

NSMC데이터만을 훈련데이터로 사용하였을 경우, LSTM과 bi-LSTM은 큰 차이는 보이지 않았다. 하지만 트랜스포머기반의 KoBERT는 더 좋은 성능을 보여주었다. 그 이유는 방대한 코퍼스로 이미 사전학습을 많이 하여, 해당 언어를 잘 이해하고 처리할 수 있기 때문이라고 판단한다. 반면 NSMC데이터와 KNU감성사전을 함께 사용하였을 경우, 모든 모델의 정확도와 F1-score가 높아짐을 확인할 수 있었다.

표 3 한국어 모델 성능 평가

DATA	MODEL	Accuracy	F1-score
NSMC	LSTM	83.47	83.09
	Bi-LSTM	83.93	83.60
	KoBERT	89.90	90.04
NSMC + KNU	LSTM	83.87	83.83
	Bi-LSTM	84.45	84.84
	KoBERT	91.64	90.76

NSMC만 사용하여 훈련한 경우, KoELECTRA가 조금 높은 정확도를 보였지만, KNU감성사전을 추가하였을 때 KoBERT가 가장 좋은 성능을 보여주었다. 이는 감정사전을 긍정과 부정을 분류가 정확도를 높이는데 영향을 주는 것으로 판단될 수 있다.

4.2 영어 감정분석 모델

영어모델은 BERT와 ELECTRA(small, base, large), DistilBERT 모델의 미세 조정을 통하여 실험을 수행하였다. 또한 학습량 증가를 위하여 IMDB 데이터를 함께 넣고 돌린 모델과 비교해 보았다. 배치 사이즈는 전체 데

이터를 쪼개어 여러 번 학습하는 것이기 때문에 학습 시 필요한 메모리와 학습속도에 영향을 끼치며, loss의 감소도가 차이가 난다.

표 4는 Friends데이터로만 실험하였을 경우와 라벨링을 치환한 IMDB데이터를 함께 학습을 진행하고, Friends 테스트 데이터셋으로 평가한 결과이다. 실험 시 epoch은 모두 동일하게 20으로 진행하였다.

표 4 영어 모델 Micro-f1 성능 평가

DATA	MODEL	Batch size	
		32	64
Friends	BERT	55.50	55.14
	ELECTRA-small	55.50	56.87
	ELECTRA-base	53.65	55.53
	ELECTRA-large	54.05	55.13
	DistilBERT	53.98	53.94
Friends + IMDB	BERT	55.50	55.57
	ELECTRA-small	56.51	56.98
	ELECTRA-base	55.39	56.65
	ELECTRA-large	54.34	55.86
	DistilBERT	54.59	53.58

Friends 데이터만 훈련 데이터로 넣었을 경우 BERT와 DistilBERT는 Batch size가 32일 때, 좀 더 좋은 성능을 보였다. 반면 ELECTRA모델은 Batch size가 64일 때 높은 정확도를 보여주었다. Friends와 IMDB데이터를 함께 학습 데이터로 사용하였을 경우, DistilBERT를 제외한 모든 모델은 Batch size 64일 때 다소 높은 경향을 보였다.

추가적으로 라벨링이 된 IMDB데이터를 넣었을 때가 좋지 않았을 때 보다 높은 Micro-f1 점수를 기록하였다. 이는 라벨링이 잘 된 학습데이터가 성능을 끌어올리는데 도움을 준다고 볼 수 있다.

표 5 테스트 데이터의 각 감정 별 F1-score

Emotion	개수 (비율)	MODEL		
		BERT	ELECTRA	DistilBERT
anger	759 (5.2%)	33.33	28.92	31.08
disgust	331 (2.3%)	13.33	0	4.17
fear	246 (1.7%)	12.9	0	3.77
joy	1,710 (11.8%)	53.31	55.01	52.91
neutral	6,530 (45.0%)	74.16	76.4	73.34
non-neutral	2,772 (19.1%)	30.44	29.35	29.78
sadness	498 (3.4%)	33.12	30.95	31.58
surprise	1,657 (11.4%)	47.04	46.58	44.08

위에서 나온 결과를 바탕으로 각 감정 별 F1-score를 비교하였다. ELECTRA 모델은 ELECTRA-small을 사용하였고, DistilBERT는 Batch32로 수행하였다. 실험 결과는 위의 표 5에서 확인할 수 있다.

세 모델 중 Micro-f1이 가장 높았던 모델은 ELECTRA였으나, 가장 sparse한 감정이었던 fear와 disgust는 하나도 맞지 못하는 결과를 보였다. 학습데이터가 제일 많았던 감정 joy의 결과는 높지만 훈련데이터의 개수가 적었던 감정에는 상대적으로 정확도가 낮은 양상을 보였다.

반면 BERT모델의 경우, 상대적으로 disgust와 fear의 감정의 F1-score가 높음을 확인할 수 있다.

5. 결론 및 향후 연구

본 연구에서 감정사전과 같은 정확한 학습데이터가 성능에 영향을 주는 것을 확인할 수 있었다. 또한 LSTM, bi-LSTM과 같은 딥러닝 모델을 시작으로 BERT, KoBERT, ELECTRA, DistilBERT까지 다양한 모델을 활용하여 학습해보았을 때, 기존 딥러닝 모델보다 트랜스포머 기반 딥러닝 모델이 성능이 더 좋음을 확인할 수 있었다. 또한 감정사전을 긍정과 부정을 분류가 정확도를 높이는데 영향을 주는 것을 확인할 수 있었다.

하지만 단순 정확도가 높다고 해서 모든 감정을 잘 분류하는 것이 아니라 감정의 분포 별로 편차가 존재할 수 있다는 것을 확인하였다. 이는 영어 데이터셋의 데이터 레이블 불균형으로 인해 정확도가 낮아졌다고 판단한다. 따라서 향후에는 레이블의 불균형이 존재할 경우 레이블 별로 가중치를 주는 방식의 WCE를 최적화를 통하여 더 높은 성능을 내는 모델에 대한 연구가 필요하다.

참고 문헌

- [1] Devlin, J., Chang, M. W., Lee, K., and Toutanova, K., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. of NAACL, 2019. <http://dx.doi.org/10.18653/v1/N19-1423>
- [2] 군산대학교 소프트웨어융합공학과 "KNU 한국어감성사전", <https://github.com/park1200656/KnuSentilex>
- [3] 김도우, 구명완, "Doc2Vec과 Word2Vec을 활용한 Convolutional Neural Network 기반 한국어 신문 기사 분류", 정보과학회논문지, 44(7), pp.742-747, 2017.
- [4] 김정미, 이주홍, "Word2vec을 활용한 RNN기반의 문서 분류에 관한 연구", 한국지능시스템학회 논문지, 27(6), pp.560-565, 2018.
- [5] 박호연(Ho-yeon Park), & 김경재(Kyoung-jae Kim). "Cnn-Lstm 조합모델을 이용한 영화리뷰 감성분석", 지능정보연구, 25(4), pp.141-154, 2019.
- [6] SKTBrain, "Korean BERT pre-trained cased (KoBERT)", <https://github.com/SKTBrain/KoBERT>
- [7] Clark, Kevin, et al. "Electra: Pre-training text encoders as discriminators rather than generators." arXiv preprint arXiv:2003.10555, 2020.