

Transformer기반의 언어모델 Bert와 GPT-2 성능 비교 연구

유연준¹, 홍석민¹, 이협건¹, 김영운¹

¹한국폴리텍대학 서울강서캠퍼스 데이터분석과

Yeanjun0222@gmail.com, ghdtjrals3@gmail.com, hglee67@kopo.ac.kr,
luckkim@kopo.ac.kr

Transformer-based Language model Bert And GPT-2 Performance Comparison Study

Yean-Jun Yoo¹, Seok-Min Hong¹, Hyeop-Geon Lee¹, Young-Woone Kim¹

¹Dept. of Data Analysis, Seoul kangseo Campus of Korea Polytechnics

요 약

최근 자연어처리 분야에서는 Bert, GPT 등 Transformer기반의 언어모델 연구가 활발히 이뤄지고 있다. 이러한 언어모델은 대용량의 말뭉치 데이터와 많은 파라미터를 이용하여 사전학습을 진행하여 다양한 자연어처리 테스트에서 높은 성능을 보여주고 있다. 이에 본 논문에서는 Transformer기반의 언어모델인 Bert와 GPT-2의 성능평가를 진행한다. 성능평가는 ‘네이버 영화 리뷰’ 데이터 셋을 통해 긍정 부정의 정확도와 학습시간을 측정한다. 측정결과 정확도에서는 GPT-2가 Bert보다 최소 4.16%에서 최대 5.32% 높은 정확도를 나타내었지만 학습시간에서는 Bert가 GPT-2보다 최소 104초에서 116초 빠르게 나타났다. 향후 성능 비교는 더 많은 데이터와 다양한 조건을 통해 구체적인 성능 비교가 필요하다.

1. 서론

자연어 처리는 기계가 자연어를 이해하고 해석하여 처리할 수 있도록 하는 일을 칭한다. 자연어 처리는 주로 텍스트 분류, 감성 분석, 텍스트 요약, 텍스트 군집화와 유사도 측정, 기계번역, 대화 시스템 및 자동 질의 응답 시스템에 사용되고 있다. 최근 자연어처리 분야에서는 Bert, GPT 등 Transformer 기반의 언어모델 연구가 활발히 이뤄지고 있다. 이러한 언어모델은 대용량의 말뭉치 데이터와 많은 파라미터를 이용하여 사전학습을 진행하고 fine-tuning을 적용하여 다양한 자연어처리 태스크에서 높은 성능을 보여주고 있다.

이에 본 연구에서 Transformer 기반의 언어모델인 Bert와 GPT-2의 성능을 비교 분석한다.

본 논문의 구성은 다음과 같다. 2장에서는 언어 모델에서 사용되는 Attention Mechanism, Transformer, Bert, GPT를 살펴본다. 3장에서는 Bert와 GPT를 감성분류의 정확도를 측정하고 분석한다. 4장에서는 결론 및 향후 연구 과제를 제시한다.

2. 관련 연구

본 장에서는 Attention Mechanism, Transformer, Bert, GPT를 살펴본다.

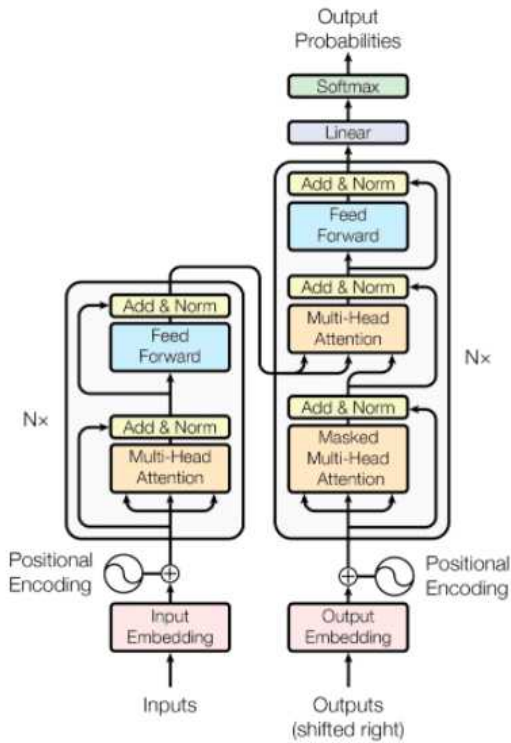
2-1. Attention Mechanism

Attention이란 기계번역을 하는 경우, 새로 만들어 내는 단어들 때문에 문장이 길어지고, 언어 특유의 어순에 따라서 멀리 있는 단어를 참조해야하는 경우가 많아져서 성능을 저하시키는 문제를 해결하고자 나온 구조이다. Attention 구조는 각 입력 단어가 출력 상태에 연결되는 신경망을 추가로 두어서, 바로 직전 단어뿐만이 아니라 모든 단어들에 얼마나 현재 결과에 기여하는지 가중치의 형태로 나타내는 것이다.

2-2. Transformer

Transformer는 2017년 구글이 발표한 논문인 ‘Attention is all you need’에서 나온 모델로 기존의 seq2seq의 구조인 인코더 디코더를 따르면서도, 논문의 이름처럼 Attention으로만 구현한 모델이다. Transformer는 크게 Encoder Block과 Decoder Block으로 나누어진다. Encoder Block은 문장을 잠재 표

현으로 변환하는 부분으로 Multi-head Attention층과 Fully-connected층으로 구성된다. Decoder Block은 실제 단어를 출력한다.



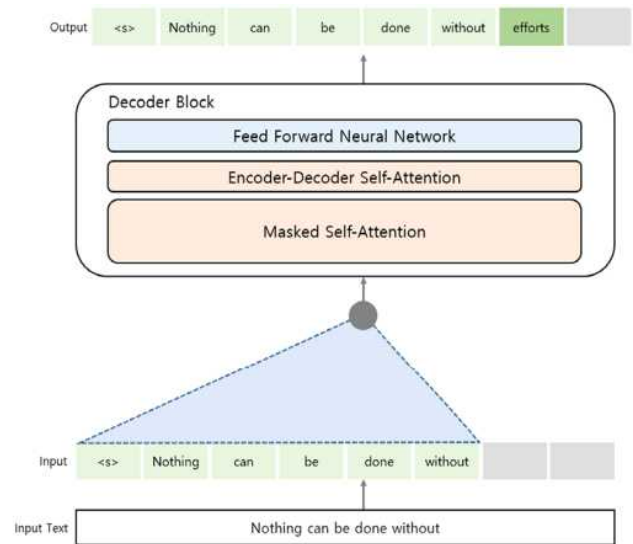
(그림 1) Transformer 모델 구조

2-3. Bert Model

Bert는 Transformer구조의 인코더 블록만을 활용한 모델이다. 특정 작업을 수행하기 전에 사전 훈련이 필요한 모델이며, Bert는 학습 말뭉치 문장의 토큰들을 랜덤으로 마스킹하고 마스킹 된 위치의 토큰을 예측하는 Masked MLM과 연속하는 두 문장의 순서가 적절한지를 예측하는 NSP 을 이용하여 사전 학습을 진행한다. 문장의 순서를 고려해야 하므로 문단 단위로 분리하였고 많은 데이터 정제 작업을 수행하였다.

2-4. GPT-2 Model

GPT-2는 Transformer구조에서 디코더 블록만 활용한 모델이다. 그림2는 GPT-2모델의 구조이다. GPT-2의 Masked Self-Attention구조는 진행 중인 시스템의 오른쪽 단어들은 고려하지 않는다. 또한 자기회기 모델로 이전의 출력이 다음의 입력이 된다. 즉 입력 받은 텍스트의 다음 토큰을 예측하게 된다.



(그림 2) GPT-2 모델 구조

3. 성능평가

본 장에서는 Transformer기반의 언어모델인 Bert와 GPT-2의 성능평가를 수행한다

3-1 성능 평가 실험

성능 평가 실험은 감성분석을 위해 '네이버 영화 리뷰'를 활용하였고 훈련용 리뷰 개수 150000개와 테스트용 리뷰 개수 50000개를 사용한다.

성능 평가 실험은 훈련용 데이터를 Google Colab 에 TPU를 사용하여 Bert와 GPT-2 언어모델에 학습 시킨 후 테스트용 데이터를 사용하여 감성분석의 정확도 및 학습시간을 측정하였다.

3-2 성능 평가

<표 1> 감성분석 정확도

Learning Count	Bert(%)	GPT-2(%)	Rating(%)
1	0.8062	0.8593	0.0531
2	0.8611	0.9035	0.0424
3	0.8884	0.9301	0.0417

<표 2> 감성분석 학습시간

Learning Count	Bert(s)	GPT-2(s)	Rating(s)
1	1262	1366	104
2	1254	1366	112
3	1254	1370	116

Bert와 GPT-2의 학습모델의 성능 비교 결과, GPT-2는 Bert에 비해 최소 4.17%에서 최대 5.31%까지 정확도에서 차이가 나는 것으로 측정되었지만 학습시간에서는 최소 104초에서 최대 116초까지 Bert가 GPT-2보다 속도가 빠른 것으로 측정되었다.

4. 결론

최근 자연어 처리 분야에서는 Bert, GPT 등 Transformer 기반의 언어모델 연구가 활발히 이뤄지고 있다. 이러한 언어모델은 대용량의 말뭉치 데이터와 많은 파라미터를 이용하여 사전학습을 진행하고 fine-tuning을 적용하여 다양한 자연어처리 태스크에서 높은 성능을 보여주고 있다.

이에 본 논문에서는 Bert와 GPT-2의 언어모델에 ‘네이버 영화 리뷰’ 데이터를 학습시켜 성능을 비교 분석 실시하였다. GPT-2는 BERT에 비해 최소 4.17%에서 최대 5.31%까지 정확도에서 차이가 나는 것으로 측정되었고 학습시간에서는 Bert가 GPT-2보다 최소 104초에서 최대 116초까지 빠른 것으로 측정되었다.

향후 성능 비교는 더 많은 데이터와 다양한 조건을 통해 구체적인 성능 비교가 필요하다.

참고문헌

- [1] 이승훈, 나승훈, 최윤수, 이해우, 장두성 “ALBERT를 이용한 한국어 자연어처리: 감성분석, 개체명 인식, 기계독해”, 한국컴퓨터종합학술대회 논문집, 2020
- [2] 김상운, 신원철 “GPT-2를 이용한 지식 그래프 완성”, 정보과학회논문지, 제48권, 제 12호, 8
- [3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, Illia Polosukhin “Attention Is All You Need”, NIPS ,2017