Prefix-tuning에 기반한 한국어 자연언어 처리

민진우 [O1], 나승훈[2], 신동욱[3], 김선훈[4], 강인호[5]

전북대학교[12], 네이버[345]

Jinwoomin4488@gmail.com, nash@jbnu.ac.kr, {shin.dongwook, seonhoon.kim, once.ihkang}@navercorp.com

Prefix-tuning for Korean Natural language processing

Jinwoo Min[O1], Seung-Hoon Na[2], Dongwook Shin[3], Seon-Hoon Kim[4], Inho Kang[5]

Jeonbuk National University[12], Naver Corporation[345]

요 약

현재 BERT와 같은 대용량의 코퍼스로부터 학습된 사전 학습 언어 모델을 자연어 응용 태스크에 적용하기 위해 일반적으로 널리 사용되는 방법은 Fine-tuning으로 각 응용 태스크에 적용 시 모델의 모든 파라미터를 조정하기 때문에 모든 파라미터를 조정하는데 필요한 시간적 비용과 함께 업데이트된 파라미터를 저장하기 위한 별도의 저장공간이 요구된다. 언어 모델이 커지면 커질수록 저장 공간의 비용이 증대됨에 따라 이러한 언어모델을 효율적으로 튜닝 할 수 있는 방법들이 연구되었다. 본 연구에서는 문장의 입력 임베딩에 연속적 태스크 특화 벡터인 prefix를 추가하여 해당 prefix와 관련된 파라미터만 튜닝하는 prefix-tuning을 한국어 네이버 감성 분석 데이터 셋에 적용 후 실험결과를 보인다.

주제어: 언어 모델, Fine-tuning, Prefix-tuning, 프롬프트

1. 서론

Fine-tuning은 BERT[1]와 같은 사전 학습 언어 모델을 자연어 응용 태스크에 적용하기 위해 일반적으로 사용되는 방법론으로 fine-tuning 방법은 각 응용 태스크에 적용 시 모델의 모든 파라미터를 조정하기 때문에 업데이트된 파라미터를 저장하기 위한 별도의 저장공간이 요구된다. 이러한 문제를 해결하기 위한 방법으로 경량화(lightweight) Fine-tuning의 접근법이 존재하며 이 방법은 사전 학습 언어 모델의 대다수의 파라미터를 업데이트하지 않고 일부의 학습 가능한 모듈을 추가하여 모델을 증대하는 방법이다[2].

Prefix-Tuning[2]은 prompting 방법에서 영감을 얻어 일종의 설명을 나타내는 가상토큰으로 prefix라고 하는 하는 연속적 태스크 특화 벡터를 추가하여 해당 prefix와 관련된 파라미터만 튜닝하는 방법이다. 본 연구에서는 [2]와 달리 생성이 아닌 자연어 분류 한국어 분류 모델에 적용하여 Prefix-tuning이 성능을 어느 정도 보장하면서 모델을 경량화 할 수 있음을 보인다.

2. 관련 연구

경량화(lightweight) Fine-tuning은 대부분의 사전 훈련된 매개변수를 고정하고 사전 훈련된 모델을 작은 훈련 가능한 모듈로만 수정하는 Fine-tuning 방법으로 [3]에서는 일부 매개변수 제거하는 가장 간단한 방식을 제안하였다. 어댑터 조정[4]은 사전 훈련된 LM의 각 레이어 사이에 응용 태스크 별로 어댑터를 삽입하는 방법으로 성능을 큰 폭으로 떨어지지 않는 수준에서 전체 매개변수의 4%만을 조정하여 학습 속도 및 저장공간의 활용에서 효율을 보였다.

프롬프팅(Prompting)[5]은 태스크의 입력에 작업 지침과 몇 가지 예를 추가하여 언어모델에서 출력을 생성하는 것을 의미하고 초대형의 파라미터를 지닌 GPT-3은 다양한 작업에 대해 생성을 적용하기 위해 사람이 설계한 프롬프트를 사용하여 별도의 튜닝 없이 놀라운 자연어 생성 능력을 보였다. 수동으로 프롬프트를 추가하는 것은 비용이 많이 들어가는 작업으로 이를 해결하기 위해 자동으로 프롬프트를 생성하는 방법에 대한 연구를 진행되었고 그 중 AutoPrompt[6]방법은 사람이 프롬프트를 작성하는 것이 아닌 입력 문장의 단어들을 검색하고 얻은 결과를 프롬프트로 사용하는 자동화 프롬프트 방법이다.

Prefix-tuning[2]은 각 응용 태스크에 대해 연속형의 prefixes를 최적화 하는 방법으로 언어모델을 효율적으로 조정하기 위해 사용되어 자연어 생성에서 저용량의 파라미터만 조정하고 저장하면서도 기존의 Fine-tuning 방법과 유사한 성능을 유지함을 보였다. 본 연구에서는 Prefix-tuning을 자연어 생성이 아닌 한국어 감성 분석 태스크에 적용하여 기존 파라미터 대비 소폭의 용량을 사용하면서 일정 수준의 성능을 보장함을 보인다.

3. 모델

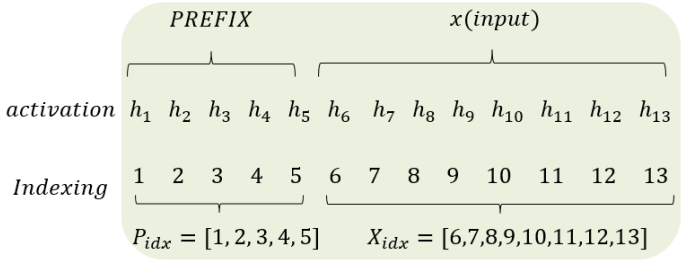

그림 1. 그림 2. Prefix-tuning의 입력 예

Prefix-tuning은 [2]에서와 같이 자연어 생성에서 인코더와 인코더 모두에 대한 Prefix를 추가하여 z =

[prefix; x; $prefix'$; y]로 입력을 주나 본 연구에서는 자연어 분류를 수행하기 위해 인코더에 대해 그림 1과 같이 prefix를 추가한 의 형태 z = [prefix; x]의 형태이다. P_{idx}는 접두사 인덱스의 시퀀스를 나타내며 $|P_{idx}|$는 접두사의 길이를 나타낸다.

목적함수에서 fine-tuning과 달리 훈련 가능한 매개변수 세트가 변경되어 언어 모델 파라미터 ϕ는 고정되어 prefix 파라미터 P_θ만이 오직 훈련 가능한 매개변수로 활용된다. [2]논문에서 prefix 파라미터를 직접 최적화하는 것은 효율이 떨어짐을 실험결과로 보였다. 그래서 본 논문에서도 동일하게 추가적인 MLP 층을 이용하여 $P_\theta = MLP_\theta(P'_\theta)$형태로 재모수화하고 여기서 MLP는 높은 차원의 feedforward Neural Networks로 즉, P_θ와 P'_θ는 동일한 행 차원(즉, 접두사 길이)을 같지만 열 차원은 다르게 설정된다.

4. 실험

4.1 실험 세팅

실험 및 평가를 위해 네이버 영화리뷰 데이터 셋[7]을 사용하였으며 위 데이터 셋은 문장, 문장의 총 20만 문장으로 구성되어 있으며 각 리뷰 문장 마다 긍정/부정의 이진레이블을 포함하고 있다. 학습데이터의 인 만 문장을 개발 셋으로 사용하였다.

사전학습 언어 모델로 ETRI 어절 BERT[8]와 한국어 RoBERTa[9]를 사용하였으며 어절 BERT는 어절에 대해 별도의 형태소 분석기 없이 byte-pair encoding(bpe)[10] 토큰화를 수행하는 토크나이저를 사용한 BERT 모델이고 한국어 RoBERTa는 3만개의 형태소 단어장에 있으면 형태소 단위로 토큰 단어장에 없으면 bpe 토큰화를 수행하는 하이브리드 토크나이저를 적용한 모델이다.

4.2 실험 결과

아래는 표1, 2는 네이버 감성 분석 데이터 셋에서의 ETRI BERT와 언어모델의 prefix tuning 실험 결과를 보여준다. 각 prefix 길이와 MLP 층에 따라 전체 파라미터 규모가 달라지고 이를 실제 언어모델의 규모 대비로 비율을 괄호로 표기한다.

표 1. ETRI BERT에서의 실험 결과

	정확도
RoBERTa 감성분석 [9]	89.88%
LM-parameter Fix (0%)	78.13%
Prefix-length10 (9.22%)	87.26%
Prefix-length100 (9.28%)	87.71%
Prefix-length200 (9.35%)	87.81%
Prefix-length200-MLPdim1024 (19.35%)	87.84%
Fine-Tune	**90.16%**

표 2. 한국어 RoBERTa에서의 실험 결과

	정확도
RoBERTa 감성분석 [9]	89.88%
LM-parameter Fix (0%)	68.16%
Prefix-length10 (9.22%)	85.19%
Prefix-length100 (9.28%)	86.52%
Prefix-length200 (9.35%)	86.71%
Prefix-length200-MLPdim1024 (19.35%)	86.71%
Fine-Tune	**89.94%**

먼저 ETRI BERT의 실험 결과에서 Prefix-tuning 중 가장 높은 성능을 보이고 있는 길이 200에 MLP 1024이나 메모리를 10%가량 많이 사용하고 있으나 성능은 0.03% 향상으로 Prefix-length200이 메모리 대비 효율적인 세팅으로 보이고 사전학습 언어모델의 파라미터를 fix한 실험에 비해서는9.68% 높고 fine-tuning에 비해서는 2.35%가량 낮은 성능을 보인다.

다음으로는 RoBERTa모델에서도 Prefix-length200가장 높은 성능을 보이고 있고 ETRI BERT와 달리 MLP 차원 변화에 따른 성능 향상은 보이지 않았다. 위 모델에서는 사전학습 언어모델의 파라미터를 fix한 실험에 비해서는18.55% 높고 fine-tuning에 비해서는 3.23%가량 낮은 성능이다.

두 모델에서 모두 prefix 길이를 10에서 100으로 올렸을 때의 성능 변화의 폭이 높았으며 100 이후 prefix길이의 변화 및 차원 증가에 따른 성능 향상의 폭이 미미하여 성능과 계산 복잡도 면에서 prefix 길이 100일때 가장 효율적인 것으로 보인다.

5. 결론

prefix라고 하는 하는 연속적 태스크 특화 벡터를 프롬프트로 추가하여 기존의 파라미터의 변경 없이 자연어 분류 태스크인 한국어 영화 리뷰 감성분석에 적용하여 실험결과를 얻었다. 이는 파인튜닝 방식 대비 소폭의 성능 하락을 보이지만 메모리에 강점을 보여 향후 초대형의 언어모델을 응용태스크에 적용 시 활용가치가 높다.

향후 연구로는 다양한 한국어 자연어 이해 태스크와 생성 태스크에도 prefix-tuning과 다양한 프롬프트 방법론을 적용하여 기존 fine-tuning과의 성능을 비교할 예정이다.

참고문헌

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In North American Association for Computational Linguistics (NAACL).

[2] Xiang Lisa Li and Percy Liang. 2021. Prefixtuning: Optimizing continuous prompts for generation. arXiv preprint.

[3] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. arXiv preprint arXiv:2005.14165, 2020
[4] Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. 2019. Parameter efficient transfer learning for NLP. In Proceedings of the 36th International Conference on Machine Learning, volume 97 of Proceedings of Machine Learning Research, pages 2790-2799, Long Beach, California, USA. PMLR.
[5] Mengjie Zhao, Tao Lin, Martin Jaggi, and Hinrich Schutze. Masking as an Efficient Alternative to Finetuning for Pretrained Language Models. arXiv:2004.12406, 2020.
[6] Taylor Shin, Yasaman Razeghi, Robert L. Logan IV, Eric Wallace, and Sameer Singh. 2020. Autoprompt: Eliciting knowledge from language models with automatically generated prompts.
[7] https://github.com/e9t/nsmc
[8] https://aiopen.etri.re.kr/service_dataset.php
[9] 민진우, 나승훈, 신종훈, 김영길, “RoBERTa를 이용한 한국어 자연어처리: 개체명 인식, 감성분석, 의존파싱”, KSC 2019
[10] Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. arXiv preprint arXiv:1508.07909, 2015.