

DAKS: 도메인 적응 기반 효율적인 매개변수 학습이 가능한 한국어 문장 분류 프레임워크

김재민¹, 채동규²

¹한양대학교 인공지능학과 석사과정

²한양대학교 인공지능학과 교수

jaemink@hanyang.ac.kr, dongkyu@hanyang.ac.kr

DAKS: A Korean Sentence Classification Framework with Efficient Parameter Learning based on Domain Adaptation

Jaemin Kim¹, Dong-Kyu Chae²

¹Dept. of Artificial Intelligence, Hanyang University

²Dept. of Computer Science, Hanyang University

요 약

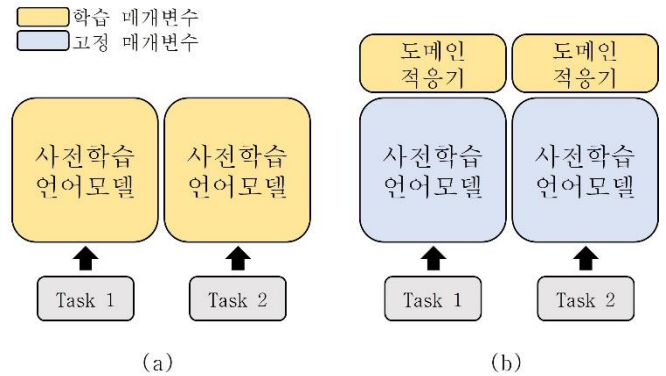
본 논문은 정확하면서도 효율적인 한국어 문장 분류 기법에 대해서 논의한다. 최근 자연어처리 분야에서 사전 학습된 언어 모델(Pre-trained Language Models, PLM)은 미세조정(fine-tuning)을 통해 문장 분류 하위 작업(downstream task)에서 성공적인 결과를 보여주고 있다. 하지만, 이러한 미세조정은 하위 작업이 바뀔 때마다 사전 학습된 언어 모델의 전체 매개변수(model parameters)를 학습해야 한다는 단점을 갖고 있다. 본 논문에서는 이러한 문제를 해결할 수 있도록 도메인 적응기(domain adapter)를 활용한 한국어 문장 분류 프레임워크인 DAKS(Domain Adaptation-based Korean Sentence classification framework)를 제안한다. 해당 프레임워크는 학습되는 매개변수의 규모를 크게 줄임으로써 효율적인 성능을 보였다. 또한 문장 분류를 위한 특징(feature)으로써 한국어 사전학습 모델(KLUE-RoBERTa)의 다양한 은닉 계층 별 은닉 상태(hidden states)를 활용하였을 때 결과를 비교 분석하고 가장 적합한 은닉 계층을 제시한다.

1. 서론

최근 자연어처리 분야에서는 사전 학습된 언어 모델을 통해 다양한 하위 작업에서 좋은 성능을 내고 있다. 특히, BERT[1]와 RoBERTa[2] 같은 트랜스포머 구조를 갖고 있는 사전 학습된 언어 모델은 큰 말뭉치를 이용한 비지도 학습을 통해 글의 맥락을 파악하고, 지도 학습을 통해 미세조정 학습을 한다. 이를 통해 인코더 구조를 갖은 사전학습 언어모델이 다양한 분류 작업에서 뛰어난 성과를 달성하고 있다. 하지만 이와 같은 미세조정 학습은 각 하위 작업에 대해 모델 전체 매개변수의 업데이트가 필요하다. 또한 각 하위 작업의 개수만큼 전체 모델의 매개변수를 학습시켜야 하는 물리적인 비용 문제가 발생한다.

이러한 비용적 문제를 해결하기 위해 프롬프트 학습[3], 도메인 적응[4, 5, 6]과 같이 사전 학습된 언어모델에 새로운 모듈을 추가하는 방법이 제안되었다. 즉, 사전 학습된 언어모델의 매개변수는 고정되고 새

로운 작업에 대해서만 프롬프트 학습 또는 도메인 적응기 매개변수만을 학습한다. 이를 통해 성능을 손상시키지 않으면서 반복적인 전체 매개변수 학습을 피할 수 있다는 장점이 있다. 아래의 [그림 1]은 도메인 적응기가 활용되었을 경우의 예시를 나타낸다.



(그림 1) 도메인 적응기 유무에 따른 차이점 비교. [그림 1]의 (a)는 일반적인 미세 조정이며 사전 학습

된 언어모델의 전체 매개변수가 학습된다. (b)는 도메인 적응기가 활용된 모델 구조이며 사전 학습된 언어모델의 매개변수는 고정되고 도메인 적응기의 매개변수만 학습된다.

본 논문은 정확하면서도 효율적인 한국어 분류 프레임워크 설계를 위해 앞서 설명한 도메인 적응 기반 효율적인 매개변수 학습 기법을 한국어 사전학습 모델인 KLUE-RoBERTa에 도입한다. 실험 결과 제안하는 DAKS(Domain Adaptation-based Korean Sentence classification framework)는 미세조정 학습의 효율성을 크게 높이면서 하위 작업 성능은 유지할 수 있었다. 또한 한국어 문장 분류 작업에 있어서 KLUE-RoBERTa의 어떤 계층의 은닉상태가 적합한지를 실험적으로 분석하고 최적의 은닉 계층을 제시한다.

2. 관련연구

2.1 도메인 적응기 (Domain adapter)

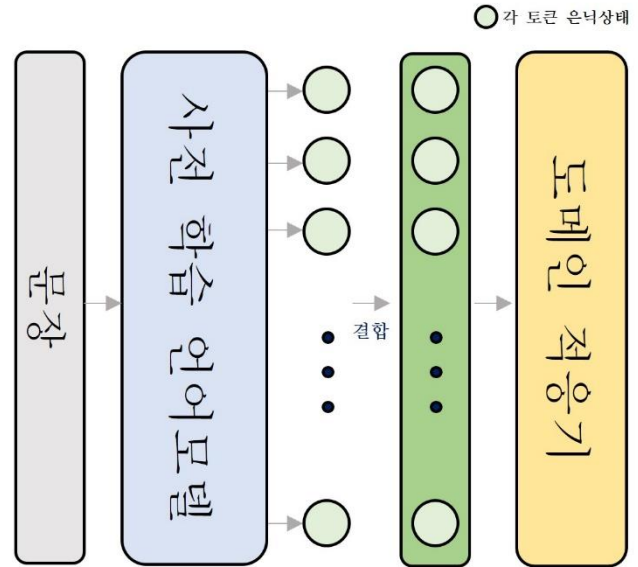
도메인 적응기는 사전 학습 모델을 학습하지 않아도 하위 작업에서 좋은 성능을 보여주면서, 학습해야 하는 매개변수의 양을 줄일 수 있는 방법으로 활용되고 있다. 특히 최근 사전 학습된 언어모델의 매개변수 크기가 커지면서 학습의 효율성을 높이기 위해 다양한 방법의 도메인 적응 연구가 이루어지고 있다. 자연어처리 분야의 다양한 하위 작업을 수행할 때 도메인 적응기를 통해 학습의 효율성을 더욱 향상시키는 방법을 탐색하거나 여러 개의 적응기를 결합하여 하위 작업의 성능을 개선하는 연구가 진행되고 있다 [4, 5, 6, 7].

2.2 문장 분류를 위한 합성곱 신경망 네트워크 (Convolutional Neural Networks; CNN)

참고문헌 [8]에서는 토큰 임베딩을 기반으로 합성곱 신경망 네트워크를 이용하여 문장 분류의 성능을 향상시켰다. 하지만 해당 논문에서는 사전 학습 모델의 은닉 상태가 아닌 Word2Vec을 문장의 특징으로 사용하였다. 본 논문에서는 Word2Vec 대신 사전 학습된 언어모델의 여러 계층의 은닉 상태를 기반으로 합성곱 신경망 네트워크를 활용하여 문장 분류를 수행하고자 한다.

3. 제안하는 방법

제안하는 프레임워크의 전체 구조는 [그림 2]에서 볼 수 있다. 사전 학습된 언어모델에서 추출한 문장의 토큰 별 은닉상태를 결합한 후 도메인 적응기에 입력으로 활용하여 학습한 후 한국어 문장을 분류한다. 해당 도메인 적응기는 합성곱 신경망 네트워크로 이루어져 있다.



(그림 2) 제안하는 DAKS 프레임워크의 전체적인 흐름. 도메인 적응기에 사전 학습 언어모델에서 나온 토큰들의 은닉상태를 활용하여 도메인 적응기로 문장을 분류한다.

도메인 적응기는 총 3개의 합성곱 계층으로 이루어져 있다. 사전 학습된 언어모델에서 출력된 토큰들의 은닉 상태(hidden states)를 결합한 후 합성곱 신경망 계층을 통과한다. 이 후 ReLU활성함수를 통과한 후 풀링 연산을 하여 완전 연결 계층을 통해 분류한다. 여러 개의 은닉 상태를 활용할 때, 은닉 상태를 결합한 후 풀링 계층을 거쳐 합성곱 신경망 네트워크로 이루어진 도메인 적응기에 입력되어 분류한다. 위 과정들을 수식으로 표현하면 아래와 같다.

$$x_{1:n} = x_1 \oplus x_2 \oplus \dots \oplus x_n \quad (1)$$

수식 (1)은 사전 학습된 언어모델의 토큰화된 문장의 은닉 상태를 이어 붙이는 과정을 나타낸다. 이 때 n은 문장의 최대 토큰 길이를 의미하며, \oplus 은 벡터 결합 연산을 의미한다.

$$y = \text{softmax}(W_2 \max(\text{ReLU}(W_1 * x_{1:n} + b_1))) \quad (2)$$

수식 (2)는 도메인 적응기의 과정을 나타낸 것이다. 이 때 W_1 는 합성곱 계층의 가중치 행렬, b_1 은 편향 벡터, ReLU는 활성화 함수, max는 풀링 계층, W_2 는 완전 연결 계층의 가중치 행렬이다. y는 각 레이블에 대한 확률 벡터이다. 사전 학습된 언어모델의 매개변수는 고정된 채로, 위의 작은 도메인 적응기의 매개변수만을 학습함으로써 획기적인 학습 시간 단축을 이룰 수 있다.

4. 실험

4.1 실험 환경

한국어 문장 분류에 대한 실험은 토픽 분류 데이터 셋인 TC[9]와 영화리뷰 감성분석 데이터 셋인 NSMC¹에 대한 분류 정확도를 측정하였다. 한국어를 위한 사전 학습된 언어모델은 트랜스포머의 인코더만으로 구성된 KLUE-RoBERTa [9] 모델을 활용했다. 실험을 위한 딥러닝 프레임 워크로는 Pytorch 1.13 과 Huggingface Transformers 를 사용했다. 배치 사이즈는 64, 학습률은 5×10^{-5} , 최적화 알고리즘은 AdamW 를, epoch size 는 5 로 설정했다. 모든 실험들은 NVIDIA RTX 3090 GPU 가 탑재된 인텔 i9-12 세대 CPU, 128GB 메모리로 구성된 단일 컴퓨터에서 수행되었다.

4.2 실험결과

<표 1> 각 모델 별 정확도 결과 비교

RoBERTa	KLUE-TC	NAVER-NSMC
Fine-tuning	89.0	90.3
DAKS-last1	86.1	88.4
DAKS-last2	85.4	89.8
DAKS-last4	86.5	90.5

[표 1]은 KLUE-RoBERTa 모델에 대해 단순 미세조정 (fine-tuning)과 제안하는 DAKS의 미세조정 성능을 비교한다. 표의 각 행 DAKS-last1, last2, last4는 순서대로 은닉 계층의 마지막 계층의 은닉상태, 마지막 2개의 계층의 은닉 상태, 마지막 4개의 은닉 상태를 활용한 DAKS 프레임워크의 분류 정확도를 의미한다. 실험 결과 제안하는 DAKS는 전체 매개변수의 학습이 없음에도 불구하고 베이스라인인 fine-tuning과 거의 유사한 성능을 보였다. 또한 마지막 4개의 은닉 상태를 활용했을 때 가장 좋은 성능을 보였다.

<표 2> 하위 작업을 위해 학습한 매개변수 규모 비교. (단위: M(백만))

RoBERTa	#parameters to be fine-tuned
Fine-tuning	110M
DaKS	0.8M

[표 2]는 베이스라인인 전체 미세조정과 비교하여 DAKS의 학습 효율성을 보여주고 있다. 단순히 전체 미세조정을 수행할 경우 각 하위 작업에 약 110M개의 파라미터가 학습되어야 하는 반면, DAKS는 약 0.8M의 학습량만 필요하므로 매우 획기적으로 효율성이 좋아졌다고 볼 수 있다. 그럼에도 불구하고 전체

미세조정과 비슷한 성능을 내는 것을 확인할 수 있다.

5. 결론

본 논문에서는 도메인 적응기를 통한 효율적인 한국어 문장 분류 프레임워크인 DAKS를 제안한다. 해당 프레임워크를 통해 기존 미세조정과 비교하여 상당히 작은 매개변수 학습만으로도 하위 작업에서 비슷한 성능을 보여주었다. 더 나아가 한국어 문장 분류에서 문장 토큰들의 은닉 상태를 결합하여 활용할 때의 최적의 은닉 층을 제시했다.

향후 연구에서는 본 논문의 연구 성과를 기반으로 효율적인 문장 분류에 대한 설명성을 찾는 연구를 진행할 예정이다.

감사의 글

이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.2020-0-01373, 인공지능대학원지원(한양대학교))을 받아 수행되었음.

참고문헌

- [1] Devlin, Jacob, et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv preprint arXiv:1810.04805 (2018).
- [2] Liu, Yinhan, et al. "RoBERTa: A Robustly Optimized BERT Pretraining Approach." arXiv preprint arXiv:1907.11692 (2019).
- [3] Lester, Brian, Rami Al-Rfou, and Noah Constant. "The Power of Scale for Parameter-Efficient Prompt Tuning." arXiv preprint arXiv:2104.08691 (2021).
- [4] Houshy, Neil, et al. "Parameter-Efficient Transfer Learning for NLP." in ICML 2019.
- [5] Ankur Bapna and Orhan Firat. "Simple, Scalable Adaptation for Neural Machine Translation." in EMNLP 2019.
- [6] Asa Cooper Stickland and Iain Murray. "BERT and PALs: Projected Attention Layers for Efficient Adaptation in Multi-Task Learning." in ICML 2021.
- [7] He, Ruidan, et al. "On the Effectiveness of Adapter-based Tuning for Pretrained Language Model Adaptation." arXiv preprint arXiv:2106.03164 (2021).
- [8] Yoon Kim, et al. "Convolutional Neural Network for Sentence Classification." in EMNLP 2014.
- [9] Park, Sungjoon, et al. "KLUE: Korean Language Understanding Evaluation." arXiv preprint arXiv:2105.09680 (2021).

¹ <https://github.com/e9t/nsmc>