

AdaLoRA의 스코어 함수 변형을 통한 파라미터 효율적 미세조정 성능 향상 연구

김명준, 박서운, 마예은, 안진현*

명지대학교

tjmjtjmj2237@mju.ac.kr, dhn04100@mju.ac.kr, k41031614@mju.ac.kr,

*wlsqus3396@mju.ac.kr

A Study on Performance Enhancement of Parameter-Efficient Fine-Tuning via Score Function Variation in AdaLoRA

Myoungjun Kim, Seoyun Park, Yeeun Ma, Jin-Hyun Ahn*

Myongji Univ.

요약

본 논문은 파라미터 효율적 미세조정(PEFT: Parameter-Efficient Fine-Tuning) 기법 중 하나인 적응형 저랭크 적응(AdaLoRA)의 스코어 함수 설계에 관한 연구를 다룬다. 기존 AdaLoRA는 학습 과정에서 파라미터 중요도를 평가하여 랭크를 동적으로 재분배한다. 하지만 스코어 함수의 정의가 제한적이기 때문에 다양한 데이터셋 및 모델 환경에서의 성능 최적화에 대한 한계가 존재한다. 이에 본 연구에서는 기존 스코어 함수 외에도 움직임 기반(Movement-based), 피셔 정보 기반(Fisher Information-based) 등 여러 형태의 스코어 함수를 제안하고, 이를 AdaLoRA에 적용하여 성능 변화를 분석하였다. GLUE 벤치마크 및 SQuAD 데이터셋을 활용한 실험 결과, 제안한 스코어 함수 중 일부는 기존 AdaLoRA 대비 최대 2.89%의 성능 향상을 달성하였으며, 특히 특정 태스크에서는 랭크 재분배 효율성이 크게 개선됨을 확인하였다. 본 연구는 스코어 함수 설계가 AdaLoRA 기반 PEFT 성능에 중요한 영향을 미친다는 점을 실험적으로 입증하였으며, 향후 대규모 언어모델 및 다양한 PEFT 기법 확장에 기여할 수 있을 것으로 기대된다.

I. 서론

최근 대규모 언어모델(LLM, Large Language Model)을 비롯한 거대 딥러닝 모델은 다양한 자연어 처리(NLP) 및 컴퓨터 비전(CV) 과제에서 우수한 성능을 보여주고 있다. 그러나 이러한 모델을 특정 태스크에 맞추어 전면 미세조정(full fine-tuning)하는 과정은 막대한 연산 자원과 메모리를 요구한다. 이에 따라, 전체 파라미터를 학습하는 대신 일부 모듈이나 저랭크(low-rank) 구조를 활용하여 효율적으로 모델을 적응시키는 파라미터 효율적 미세조정(PEFT, Parameter-Efficient Fine-Tuning) 기법이 활발히 연구되고 있다.

LoRA(Low-Rank Adaptation)는 대표적인 PEFT 기법으로, 사전 학습된 가중치에 저랭크 행렬을 삽입하여 효율적으로 파라미터를 학습할 수 있도록 한다.[1] LoRA는 전체 모델의 파라미터를 고정(freeze)한 상태에서 소수의 저랭크 행렬의 파라미터만 학습하므로 메모리와 연산 비용이 크게 절감된다. 그러나 LoRA는 학습 과정에서 저랭크 차원(rank)이 고정되어 있어, 태스크별 특성과 학습 단계에 따라 유연하게 자원을 배분하기 어렵다는 한계가 있다.

이를 개선하기 위해 제안된 AdaLoRA(Adaptive LoRA)는 학습 도중 파라미터 중요도를 평가하여 랭크를 동적으로 재분배함으로써, 자원 활용 효율성을 높이고 성능을 향상시킨다.[2] AdaLoRA에서 파라미터 중요도는 스코어 함수(score function)를 통해 산정되며, 이 값이 랭크 재분배의 핵심 기준이 된다. 그러나 기존 AdaLoRA의 스코어 함수는 학습 중 가중치의 방향이 아닌 움직임의 정도만 고려하기 때문에 다양한 데이터셋과 모델 환경에서의 성능 최적화에 한계가 있다. 따라서, 스코어 함수의 설계

방식을 확장하고 그 효과를 실험적으로 검증하는 연구가 필요하다.

본 논문에서는 AdaLoRA에서 사용되는 스코어 함수를 다변화하고, 각각의 방식이 성능에 미치는 영향을 분석한다. 구체적으로 움직임 기반(Movement-based), 피셔 정보 기반(Fisher Information-based), 그라디언트-잡음 비율(GSNR, Gradient Signal-to-Noise Ratio)과 같은 스코어 함수를 설계하고 이를 GLUE, SQuAD 등 다양한 벤치마크 데이터셋에서 비교·평가한다. 이를 통해, 스코어 함수 설계가 AdaLoRA 기반 PEFT 기법의 성능 향상에 미치는 영향을 탐색한다.

II. 제안 방법

A. 연구 개요

일반적으로 k 번째 가중치 행렬에 대해 LoRA의 저랭크 행렬은 다음과 같이 두 행렬 $P_k \in R^{d_1 \times r}$ 와 $Q_k \in R^{r \times d_2}$ 로 구성되며, 이를 직교 기저로 표현하면 대각 성분 λ_k 와 양쪽 직교 행렬 P_k, Q_k 로 분해할 수 있다. AdaLoRA는 각 가중치 행렬의 i 번째 rank 성분을 다음과 같은 트리플렛 형태로 표현한다:

$$G_{k,i} = \{P_{k,i}, \lambda_{k,i}, Q_{k,i}\}$$

여기서 $P_{k,i}$ 와 $Q_{k,i}$ 는 좌/우 특이벡터이며, $\lambda_{k,i}$ 는 해당 성분의 특이값을 의미한다. 그리고 각 트리플렛의 중요도는 다음과 같이 계산한다:

$$S_{k,i} = s(\lambda_{k,i}) + \frac{1}{d_1} \sum_{j=1}^{d_1} s(P_{k,j,i}) + \frac{1}{d_2} \sum_{j=1}^{d_2} s(Q_{k,i,j})$$

여기서 $s(\cdot)$ 는 특정 요소의 중요도를 산출하는 스코어 함수이며, AdaLoRA에서는 다음과 같은 스코어가 사용된다.

$$s(w_{ij}) = |w_{ij} \nabla_{w_{ij}} \mathcal{L}|$$

즉, 가중치 w_{ij} 의 값과 해당 파라미터에 대한 손실 함수 \mathcal{L} 의 기울기(gradient)를 곱한 절대값을 스코어로 사용하여, 가중치의 크기와 학습 중 변화 민감도를 동시에 반영한다. 하지만 이 스코어는 해당 가중치가 모델의 학습 중, 0의 방향으로 이동하는 것을 제대로 반영하지 못한다.

B. 새로운 스코어 함수 설계

1) 움직임 기반 (Movement-based)

$$s(w_{ij}) = -w_{i,j} \nabla_{w_{i,j}} \mathcal{L}$$

해당 스코어는 [3]에서 제안된 아이디어를 기반으로 하며, 학습 과정에서 가중치가 0 으로부터 얼마나 멀리 이동하려 하는지를 고려한다.

2) 피셔 정보 기반 (Fisher Information-based)

$$s(w_{ij}) = E [\nabla_{w_{i,j}} \mathcal{L}^2]$$

피셔 정보(Fisher Information)는 통계학에서 관측 데이터가 모수(parameter)에 대해 제공하는 정보량을 정량화하는 척도로, 모수 추정의 불확실성을 이론적으로 설명하는 핵심 개념이다. 딥러닝에서는 파라미터의 기울기 분포로부터 피셔 정보를 근사하여 각 파라미터가 손실 함수에 미치는 민감도를 평가할 수 있다.

3) GSNR (Gradient Signal-to-Noise Ratio)

$$s(w_{ij}) = \frac{(E [\nabla_{w_{i,j}} \mathcal{L}])^2}{\text{Var} [\nabla_{w_{i,j}} \mathcal{L}]}$$

GSNR은 기울기 제공의 기댓값을 측정하여, 파라미터의 신호 세기를 잡음에 대비해 정량화하는 지표이다. 이 지표는 [4]에서 뉴런의 활성화 중요도를 평가하는 척도로도 활용되었다.

III. 실험 환경 및 설정

A. 데이터셋

제안한 스코어 함수를 적용한 AdaLoRA 기법의 성능을 다각도로 평가하기 위해 자연어 이해(NLU) 및 질의응답(QA) 테스트를 모두 포함하는 데이터셋을 사용한다.

- 1) GLUE Benchmark: 일반 NLU 능력 측정[5]
- 2) SQuAD v1.1: 정답 추출 기반 QA[6]
- 3) SQuAD v2.0: QA + 불확실성 판별[7]

B. 학습 환경

실험은 NVIDIA A100 GPU 40GB 환경에서 수행하였다. 베이스 모델로는 DeBERTa-v3 base를 사용하였으며, 모든 데이터셋에 대해 동일한 학습 설정을 적용하였다. 학습 에폭(epoch)은 3 으로 설정하였고, AdaLoRA 초기 랭크(r_{init})는 16, 최종 랭크(r_{final})는 8 로 두어 랭크 축소 과정을 거치도록 하였다. 모든 모델은 Adam 옵티마이저와 학습률 스케줄러를 사용하였으며 배치 크기와 학습률은 데이터셋 특성에 따라 사전 실험을 통해 최적화하였다.

IV. 실험 결과 및 분석

Table 1: GLUE Benchmark

Method	MRPC Acc	SST-2 Acc	CoLA Acc	QQP Acc
AdaLoRA(Original)	89.95	95.41	67.58	89.81
AdaLoRA(Movement)	90.44	94.49	68.23	90.51
AdaLoRA(Fisher)	89.46	95.64	69.69	91.64
AdaLoRA(GSNR)	88.73	94.38	67.28	91.60
QNLI	RTE	MNLI	STB-B	All
Acc	Acc	m/mm	Corr	Ave.
93.86	83.39	89.38/89.37	91.77	87.84
93.70	86.28	89.78/89.61	91.27	88.26
93.89	85.56	90.41 /90.39	91.33	88.67
93.76	85.56	90.36/ 90.50	90.35	88.06

Table 2: SQuAD v1.1 and SQuAD v2.0

Method	SQuAD v1.1 EM/F1	SQuAD v2.0 EM/F1
AdaLoRA(Original)	87.38/93.52	84.53/87.58
AdaLoRA(Movement)	87.54/93.68	85.98/88.88
AdaLoRA(Fisher)	88.12/93.93	85.67/88.50
AdaLoRA(GSNR)	88.48/94.23	86.09/88.97

실험 결과, 제안한 스코어 함수들은 기존 AdaLoRA 대비 대부분의 테스트에서 의미 있는 성능 향상을 보인다. 이는 스코어 함수 설계가 랭크 재분배 효율성 및 모델의 최종 성능에 중요한 영향을 미친다는 점을 입증한다.

V. 결론

본 연구에서는 파라미터 효율적 미세조정(PEFT) 기법인 AdaLoRA에서 사용되는 스코어 함수의 설계가 성능에 미치는 영향을 분석하였다. 기존 AdaLoRA의 한계를 파악하고 이를 해결하기 위해 다양한 스코어 함수를 설계하고 평가하였다.

다양한 NLU와 QA 데이터셋을 대상으로 실험을 진행한 결과, 제안한 스코어 함수는 대부분 기존 AdaLoRA 대비 성능이 향상되었음을 확인하였다. 향후 연구에서는 본 연구에서 제안한 스코어 함수를 대규모 언어모델과 다양한 도메인에 적용하여 분석할 계획이다. 이를 통해 AdaLoRA뿐 아니라 파라미터의 중요도를 판단하는 지표로써, 보다 범용적인 파라미터 효율적 미세조정 전략 개발에 기여할 수 있을 것으로 기대한다.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00212836)

참고 문헌

- [1] Hu, Edward J., et al. "Lora: Low-rank adaptation of large language models." *ICLR* 1.2 (2022): 3.
- [2] Zhang, Qingru, et al. "Adalora: Adaptive budget allocation for parameter-efficient fine-tuning." *arXiv preprint arXiv:2023.10512* (2023).
- [3] Sanh, Victor, Thomas Wolf, and Alexander Rush. "Movement pruning: Adaptive sparsity by fine-tuning." *Advances in neural information processing systems* 33 (2020): 20378-20389.
- [4] Liu, Jinlong, et al. "Understanding why neural networks generalize well through gsnr of parameters." *arXiv preprint arXiv:2001.07384* (2020).
- [5] Wang, Alex, et al. "GLUE: A multi-task benchmark and analysis platform for natural language understanding." *arXiv preprint arXiv: 1804.07461* (2018).
- [6] Rajpurkar, Pranav, et al. "Squad: 100,000+ questions for machine comprehension of text." *arXiv preprint arXiv: 1606.05250* (2016).
- [7] Rajpurkar, Pranav, Robin Jia, and Percy Liang. "Know what you don't know: Unanswerable questions for SQuAD." *arXiv preprint arXiv: 1806.03822* (2018).