

경사하강법에서 해석적 및 수치해석적 기울기 적용 방법

조혜정, 윤수용, 문지원, 임재빈, 정찬호*

한밭대학교

20201566@edu.hanbat.ac.kr, 20181435@edu.hanbat.ac.kr, 20221094@edu.hanbat.ac.kr,
20167072@edu.hanbat.ac.kr, peterjung@hanbat.ac.kr*

Deploying Analytic and Numerical Gradients in the Gradient Descent Method

Hyejeong Cho, Suyoung Yun, Jiwon Moon, Jaebin Lim, Chanhong Jung*

Hanbat National University

요약

본 논문에서는 퓨샷 러닝 및 경량 딥러닝 응용을 위한 경사하강법에서의 해석적 및 수치해석적 기울기 적용 방법을 제안한다. 제안하는 방법에서는 해석적 및 수치해석적 기울기를 반복(iteration)마다 번갈아 가며 적용하였다. 제안하는 방법의 효율성을 검증하기 위하여 MNIST 데이터셋을 이용하여 정확도를 측정하였다. 본 논문에서는 퓨샷 러닝을 위해 클래스 당 학습데이터 개수를 제한하였다. 또한 경량 딥러닝 응용을 위해 모델의 크기가 작고 연산량이 적은 네트워크 아키텍처를 이용하였다. 실험결과는 전반적으로 베이스라인 및 제안하는 방법의 정확도가 비슷함을 보여주었다. 더불어 클래스 당 학습데이터 개수가 2일 때 제안하는 방법이 베이스라인에 비해 비교적 우수함을 알 수 있었다.

I. 서론

딥러닝 모델을 학습시킬 때 일반적으로 최적화(optimization) 과정에서 가중치(weight) 매개변수에 대한 손실함수의 기울기를 사용한다. 기울기를 구하는 방법은 크게 두 가지로 나누어진다: 1) 해석적 기울기 및 2) 수치해석적 기울기. 대부분의 최적화 기법들에서는 해석적 기울기를 이용해 왔다.

본 논문에서는 퓨샷 러닝 및 경량 딥러닝 응용을 위한 해석적 및 수치해석적 기울기 적용 방법을 제안한다. 퓨샷 러닝[1] 및 경량 딥러닝[2]은 각각 그 중요성 때문에 최근 활발하게 연구되고 있는 주제들이다. 제안하는 방법에는 경사 하강법에서 해석적 및 수치해석적 기울기를 반복마다 번갈아 가며 적용하였다. 제안하는 방법의 효율성을 검증하기 위해 MNIST 데이터셋[3]을 이용하여 정확도를 측정하였다. 본 논문에서는 퓨샷 러닝을 위해 클래스 당 학습데이터 개수를 제한하였다. 또한 경량 딥러닝 응용을 위해 1) 모델의 크기가 작고 연산량이 적은 네트워크 아키텍처(two-layer network)를 이용하였으며, 2) 히든 레이어의 뉴런 개수를 입력 레이어의 뉴런 개수 대비 매우 작은 값으로 설정하였다. 실험결과는 베이스라인(해석적 기울기 기반 경사 하강법)의 정확도 및 제안하는 방법의 정확도가 비슷함을 보여준다. 또한 클래스 당 학습데이터 개수가 2일 때 제안하는 방법이 베이스라인에 비해 비교적 우수함을 알 수 있었다.

II. 제안하는 방법

제안하는 방법은 경사 하강법 구현에서 다음과 같은 가정을 기반으로 설계되었다: “해석적 기울기에 비해 정확하지 않은 수치해석적 기울기를 통한 가중치 업데이트를 2번의 반복마다 배치함으로써 기울기의 정확성과는 반대로 성능 향상을 이끌어 낼 수 있다.” 즉, 경사 하강법 적용 과정에서 기울기의 정확성과 딥러닝 모델의 정확성이 항상 비례하는 것은 아닐

수도 있음을 가정하였다. 그림 1은 경사 하강법에서 베이스라인 및 제안하는 방법의 기울기 적용 방법의 차이를 보여준다.

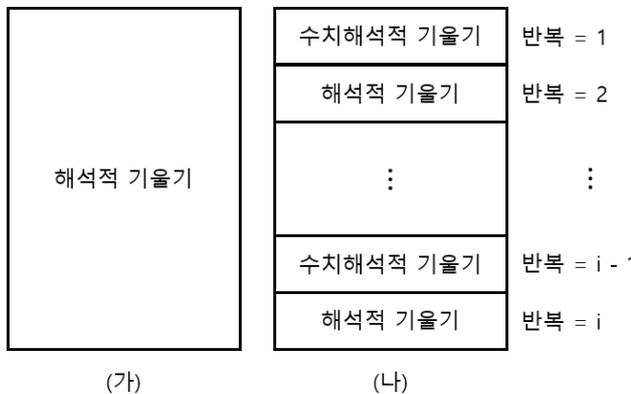


그림 1. 경사 하강법에서 베이스라인 및 제안하는 방법의 기울기 적용 방법 차이: (가) 베이스라인, (나) 제안하는 방법

III. 실험결과

본 논문에서는 제안하는 방법의 성능을 평가하기 위하여 two-layer network를 이용하였다. 히든 레이어의 뉴런 개수 영향을 평가하기 위하여 히든 레이어의 뉴런 개수를 각각 20, 50으로 설정하였다. MNIST 학습데이터셋으로부터 추출한 데이터를 two-layer network 학습을 위한 학습데이터셋으로 이용하였다. (퓨샷 러닝을 위한) 학습데이터 개수에 따른 성능 변화 정도를 알아보기 위하여 데이터를 추출하는 과정에서 클래스 당 학습데이터 개수가 각각 1, 2, 3, 4가 되도록 하였다. 성능 평가를 위해 10,000개의 샘플로 구성된 MNIST 테스트데이터셋을 이용하였다. 배치사이즈에 따른 성능 변화를 평가하기 위하여 배치사이즈를 각각 1, 2, 4, 8이

표 1. 배치사이즈, 클래스 당 학습데이터 개수, 히든 레이어의 뉴런 개수 변화에 따른 베이스라인 및 제안하는 방법 간의 성능 비교 (1행: 베이스라인(히든 레이어의 뉴런 개수: 20), 2행: 제안하는 방법(히든 레이어의 뉴런 개수: 20), 3행: 베이스라인(히든 레이어의 뉴런 개수: 50), 4행: 제안하는 방법(히든 레이어의 뉴런 개수: 50))

배치사이즈	클래스 당 학습데이터 개수			
	1	2	3	4
1	0.5094	0.5456	0.5559	0.5986
	0.5093	0.5456	0.5549	0.5986
	0.5062	0.5593	0.6072	0.6423
	0.5062	0.5593	0.5930	0.6262
2	0.5271	0.5621	0.6145	0.6168
	0.5270	0.5621	0.6144	0.6169
	0.5129	0.5608	0.6110	0.6439
	0.5130	0.5613	0.6110	0.6437
4	0.5325	0.5643	0.6163	0.6147
	0.5325	0.5643	0.6143	0.6146
	0.5160	0.5628	0.6110	0.6423
	0.5160	0.5630	0.6109	0.6426
8	0.5250	0.5638	0.6076	0.6138
	0.5252	0.5629	0.6077	0.6137
	0.5134	0.5608	0.6091	0.6421
	0.5134	0.5615	0.6091	0.6421

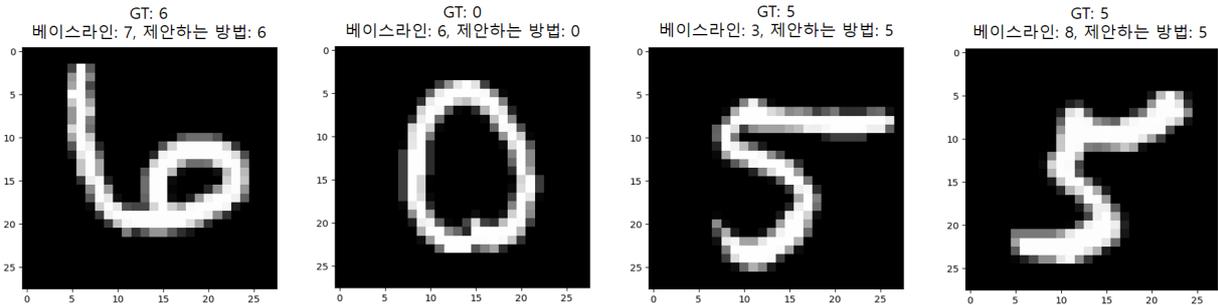


그림 2. 동일한 테스트샘플에 대한 베이스라인 및 제안하는 방법의 예측결과

되도록 설정하였다. 즉, 히든 레이어의 뉴런 개수 및 클래스 당 학습데이터 개수를 일정하게 유지시키면서 배치사이즈를 각각 1, 2, 4, 8로 하여 미니배치학습을 수행하였다. 표 1은 배치사이즈, 클래스 당 학습데이터 개수, 히든 레이어의 뉴런 개수 변화에 따른 베이스라인 및 제안하는 방법 간의 성능 비교 결과를 보여준다. 표 1에서 보는 바와 같이 클래스 당 학습데이터 개수가 2일 때 제안하는 방법이 베이스라인에 비해 비교적 우수함을 알 수 있었다. 또한 전반적으로 제안하는 방법이 베이스라인과 비슷한 성능을 제공함을 볼 수 있었으며, 배치사이즈, 클래스 당 학습데이터 개수, 히든 레이어의 뉴런 개수 변화에 따른 베이스라인 및 제안하는 방법 간의 성능 차이도 크지 않음을 볼 수 있었다. 그림 2는 1) 배치사이즈가 8, 2) 히든 레이어의 뉴런 개수가 50, 3) 클래스 당 학습데이터 개수가 2일 때 동일한 테스트샘플에 대한 베이스라인 및 제안하는 방법의 예측결과를 보여준다.

IV. 결론

본 논문에서는 퓨샷 러닝 및 경량 딥러닝 응용을 위한 경사하강법에서의

해석적 및 수치해석적 기울기 적용 방법을 제안하였다. 실험결과를 통해 클래스 당 학습데이터 개수가 2일 때 제안하는 방법이 베이스라인에 비해 비교적 우수함을 알 수 있었다. 즉, 퓨샷 러닝을 위한 제안하는 방법의 적용 가능성을 볼 수 있었다.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2021R1I1A3041191).

참고 문헌

- [1] J. Snell, S. Kevin, and Z. Richard, "Prototypical networks for few-shot learning," *Advances in Neural Information Processing systems*, 2017.
- [2] C. H. Wang, K. Y. Huang, Y. Yao, J. C. Chen, H. H. Shuai, W. H. Cheng, "Lightweight Deep Learning: An Overview," *IEEE Consumer Electronics Magazine*, 2022.
- [3] L. Deng, "The mnist database of handwritten digit images for machine learning research," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141-142, 2012.