퓨샷 러닝 및 경량 딥러닝 응용을 위한 해석적 및 수치해석적 그레디언트 기반 미니배치학습 방법

정태순, 이재오, 김재현, 임재빈, 정찬호*

한밭대학교 전기공학과

20172566@edu.hanbat.ac.kr, 20181383@edu.hanbat.ac.kr, 20181422@edu.hanbat.ac.kr,
20167072@edu.hanbat.ac.kr, peterjung@hanbat.ac.kr*

Analytic and Numerical Gradient-based Mini Batch Training for Few-Shot Learning and Lightweight Deep Learning Applications

Taesun Chung, Jaeoh Lee, Jaehyeon Kim, Jaebin Lim, Chanho Jung*

Hanbat National University

요 약

본 논문에서는 퓨샷 러닝 및 경량 딥러닝 응용을 위한 해석적 및 수치해석적 그레디언트 기반 미니배치학습 방법을 제안한다. 제안하는 방법에서는 사전에 정의된 비율로 나눈 미니배치에 대해서 해석적 및 수치해석적 그레디언트 이용한 경사 하강법을 진행하였다. 본 논문에서는 퓨샷 러닝 응용을 위해 클래스당 학습 데이터를 제한하였다. 또한 경량 딥러닝 응용을 위해 간단한 구조를 가지는 신경망을 이용하였다. 제안하는 방법과 베이스라인에 대한 정량적 비교평가를 위해 MNIST 데이터셋을 이용하여 정확도를 측정하는 실험을 수행하였다. 실험결과는 전반적으로 베이스라인 및 제안하는 방법의 정확도가 비슷함을 보여주었다. 더불어 배치 사이즈와 뉴런의 개수가 각각 4, 50일 때 제안하는 방법이 베이스라인에 비해 비교적 우수한 성능을 제공함을 볼 수 있었다.

Ⅰ. 서 론

신경망을 학습시킬 때 손실함수가 최솟값이 되도록 매개변수를 갱신해 주는 경사 하강법을 이용한다. 경사 하강법은 매개변수를 갱신해 줄 때 손실함수의 그레디언트를 사용한다. 그레디언트는 해석적 및 수치해석적 미분 방법을 통해 구할 수 있다. 대부분의 신경망 학습에서는 해석적 그레디언트를 이용해왔다.

본 논문에서는 퓨샷 러닝 및 경량 딥러닝 응용을 위한 해석적 및 수치해석적 그레디언트 기반 미니배치학습 방법을 제안한다. 제안하는 방법에서는 사전에 정의된 비율로 나눈 미니배치에 대해서 해석적 및 수치해석적 그레디언트 이용한 경사 하강법을 진행하였다. 본 논문에서는 퓨샷 러닝 [1] 응용을 위해 클래스당 학습 데이터를 제한하였다. 또한 경량 딥러닝[2] 응용을 위해 매우 단순한 구조를 가지는 네트워크 아키텍처(two-layer network)를 이용하였으며, 은닉층의 뉴런 개수를 입력층의 뉴런 개수에 비해 매우 작은 값으로 설정하였다. 제안하는 방법과 베이스라인에 대한 정량적 비교평가를 위해 MNIST 데이터셋[3]을 이용하여 정확도를 측정하는 실험을 수행하였다. 실험결과는 제안하는 방법의 정확도 및 베이스라인(해석적 그레디언트 기반 경사 하강법)의 정확도가 비슷함을 보여준다. 더불어 배치 사이즈와 은닉층 뉴런의 개수가 각각 4, 50일 때 제안하는 방법이 베이스라인에 비해 비교적 우수한 성능을 제공함을 볼 수 있었다.

Ⅱ. 제안하는 방법

제안하는 방법은 경사 하강법 구현에서 다음과 같은 가정을 기반으로 설계되었다: “해석적 그레디언트에 비해 정확하지 않은 수치해석적 그레디언트를 통한 경사 하강법에서 사전에 정의된 비율로 나눈 미니배치에 대해서 각각의 그레디언트를 적용함으로써 성능 향상을 이끌어 낼 가능성이 있다.”. 즉, 미니배치에 대한 경사 하강법 적용 과정에서 그레디언트의 정확성과 딥러닝 모델의 정확성이 항상 비례하는 것은 아닐 수도 있음을 가정하였다. 본 논문에서는 미니배치에 적용되는 해석적 및 수치해석적 그레디언트의 비율을 각각 α, $1-\alpha$로 설정하였다. 그림 1은 미니배치에서 베이스라인과 제안하는 방법의 그레디언트 적용 방법 차이를 나타낸다.

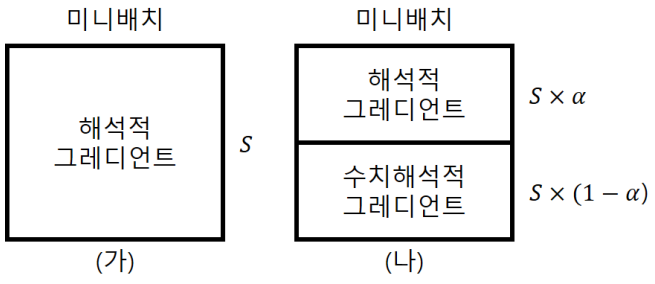

그림 1. 미니배치에서 베이스라인 및 제안하는 방법의 그레디언트 적용 방법 차이: (가) 베이스라인, (나) 제안하는 방법, S: 미니배치 사이즈, α: 해석적 그레디언트 적용 비율

Ⅲ. 실험결과

본 논문에서는 two-layer network를 이용하여 제안하는 방법의 성능을

표 1. 학습률, 배치 사이즈, 클래스당 학습 데이터 개수, 은닉층 내 뉴런 개수 변화에 따른 베이스라인 및 제안하는 방법 간의 성능 비교 (1행: 베이스라인(은닉층의 뉴런 개수: 20), 2행: 제안하는 방법(은닉층의 뉴런 개수: 20), 3행: 베이스라인(은닉층의 뉴런 개수: 50), 4행: 제안하는 방법(은닉층의 뉴런 개수: 50))

학습률	배치 사이즈	클래스당 학습 데이터 개수				
		1	2	3	4	5
0.1	2	0.5221	0.5505	**0.6138**	0.6165	**0.6494**
		0.5222	**0.5507**	0.6137	**0.6167**	0.6471
		0.5123	0.5546	**0.6106**	0.6430	0.6569
		0.5124	**0.5548**	0.6103	0.6430	0.6569
	4	0.5283	**0.5575**	**0.6139**	0.6145	**0.6459**
		0.5283	0.5573	0.6124	0.6145	0.6457
		0.5158	**0.5584**	0.6074	0.6387	0.6466
		0.5158	0.5583	**0.6076**	**0.6390**	0.6466
	8	**0.5236**	**0.5589**	0.6071	**0.6138**	**0.6460**
		0.5235	0.5588	0.6071	0.6137	0.6459
		0.5124	0.5567	**0.6081**	**0.6397**	0.6468
		0.5122	0.5567	0.6080	0.6395	**0.6470**
0.01	2	0.5298	**0.5604**	0.6062	0.6161	**0.6463**
		0.5300	0.5603	**0.6065**	0.6161	0.6461
		0.5153	**0.5582**	0.6086	**0.6378**	0.6474
		0.5154	0.5580	0.6086	0.6376	**0.6467**
	4	**0.5286**	0.5599	0.6072	**0.6169**	**0.6469**
		0.5284	**0.5601**	**0.6075**	0.6168	0.6468
		0.5165	**0.5586**	0.6084	0.6374	0.6475
		0.5166	0.5580	**0.6085**	**0.6375**	**0.6476**
	8	**0.5293**	0.5591	**0.6068**	0.6167	0.6464
		0.5291	**0.5592**	0.6066	**0.6169**	0.6464
		0.5169	**0.5564**	0.6085	0.6377	0.6477
		0.5171	0.5563	0.6085	0.6377	**0.6478**

평가하였다. 경량 딥러닝을 응용하며 은닉층 뉴런 개수에 따른 영향을 알아보기 위해 은닉층 뉴런의 개수를 각각 20, 50으로 설정하였다. 퓨샷 러닝을 위해 학습데이터를 추출하는 과정에서 클래스당 학습데이터 개수가 각각 1, 2, 3, 4, 5가 되도록 하였다. 제안하는 방법의 α를 0.5로 설정하였다. 또한 학습률 및 배치 사이즈의 변화에 따른 베이스라인과 제안하는 방법의 성능을 비교평가하였다. 학습률은 각각 0.1, 0.01이 되도록 설정하였고 배치 사이즈는 각각 2, 4, 8로 설정하여 미니배치학습을 수행하였다. 성능 평가를 위해 10,000개의 데이터로 구성된 MNIST 테스트데이터셋을 이용하였다. 표 1은 학습률, 배치 사이즈, 클래스당 학습데이터 개수 그리고 은닉층의 뉴런 개수 변화에 따른 베이스라인과 제안하는 방법의 성능 비교 결과를 보여준다. 표 1에서 보는 바와 같이 배치 사이즈와 뉴런의 개수가 각각 4, 50일 때 제안하는 방법이 베이스라인에 비해 비교적 우수한 성능을 제공함을 알 수 있었다. 또한 학습률, 배치 사이즈, 클래스당 학습데이터 개수, 히든 레이어의 뉴런 개수 변화에 따른 베이스라인 및 제안하는 방법 간의 성능 차이가 크지 않음을 볼 수 있었다.

Ⅳ. 결론

본 논문에서는 퓨샷 러닝 및 경량 딥러닝 응용을 위한 해석적 및 수치 해석적 그레디언트 기반 미니배치학습 방법을 제안하였다. 실험 결과를 통해 배치 사이즈와 뉴런의 개수가 각각 4, 50일 때 제안하는 방법이 베이스라인에 비해 비교적 우수한 성능을 제공함을 알 수 있었다.

참 고 문 헌

[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.