경사하강법에서 해석적 및 수치해석적 미분 기반 최적 파라미터 선택 방법

문지원, 윤수용, 조혜정, 정태순, 임재빈, 정찬호*

한밭대학교

20221094@edu.hanbat.ac.kr, 20181435@edu.hanbat.ac.kr, 20201566@edu.hanbat.ac.kr,
20172566@edu.hanbat.ac.kr, 20167072@edu.hanbat.ac.kr, peterjung@hanbat.ac.kr*

Selecting Optimal Parameters based on Analytic and Numerical Gradients in the Gradient Descent Method

Jiwon Moon, Suyoung Yun, Hyejeong Jo, Taesun Chung, Jaebin Lim, Chanho Jung*

Hanbat National University

요 약

본 논문에서는 퓨샷 러닝 및 경량 딥러닝 응용을 위한 해석적 및 수치해석적 미분 기반 최적 파라미터 선택 방법을 제안한다. 제안하는 방법에서는 경사 하강법에서 매 반복(iteration)마다 해석적 및 수치해석적 기울기를 기반으로 한 갱신을 통해 서로 다른 2개의 가중치 파라미터(weight parameter)들을 얻어낸다. 이어서 손실함수 값을 최소로 하는 가중치 파라미터를 최적의 파라미터로 선택한다. 본 논문에서는 퓨샷 러닝을 위해 클래스당 학습데이터 개수를 제한하였으며, 경량 딥러닝 응용을 위해 매우 단순한 구조를 가지는 딥러닝 모델을 사용하였다. MNIST 데이터셋을 이용한 실험을 통해 기존 방법 및 제안하는 방법의 성능을 비교하였다. 실험결과에 대한 분석을 통해 1) 제안하는 방법의 정확도와 베이스라인의 정확도가 전반적으로 비슷하며, 2) 배치사이즈가 1일 때 제안하는 방법이 베이스라인보다 비교적 향상된 성능을 제공하는 것을 확인할 수 있었다.

Ⅰ. 서 론

경사 하강법에서 기울기를 구하는 방법은 크게 1) 해석적 미분, 2) 수치해석적 미분으로 나눌 수 있다. 대부분의 경사하강법에서는 해석적 기울기를 사용해왔다.

본 논문에서는 퓨샷 러닝 및 경량 딥러닝 응용을 위한 해석적 및 수치해석적 미분 기반 최적 파라미터 선택 방법을 제안한다. 희귀 클래스 분류 및 경량 디바이스에서 딥러닝 모델을 응용하고자 최근 퓨샷 러닝[1, 2] 및 경량 딥러닝[3] 연구가 활발하게 이뤄지고 있다. 제안하는 방법에서는 경사 하강법에서 해석적 및 수치해석적 기울기를 기반으로 갱신한 가중치 파라미터들 중 최적의 파라미터를 선택하여 적용하였다. 제안하는 방법의 성능을 검증하기 위해 MNIST 데이터셋을 이용하여 정확도를 측정하였다. 본 논문에서는 퓨샷 러닝을 위해 클래스당 학습데이터 개수를 제한하였다. 또한 경량 딥러닝 응용을 위해 1) 매우 단순한 구조를 가지는 네트워크 아키텍처(two-layer network)를 이용하였으며, 2) 은닉층의 뉴런 개수를 입력층의 뉴런 개수에 비해 매우 작은 값으로 설정하였다. 실험결과는 제안하는 방법의 정확도 및 베이스라인(해석적 및 수치해석적 미분 기반 경사 하강법)의 정확도가 비슷함을 보여준다. 또한 배치사이즈가 1일 때 제안하는 방법이 베이스라인에 비해 비교적 우수함을 알 수 있었다.

Ⅱ. 제안하는 방법

본 논문에서는 매 반복마다 손실함수의 값을 최소로 하는 최적의 파라미터를 선택한다. 이를 위해 동일한 가중치 파라미터를 기준으로 해석적 및 수치해석적 기울기를 (매 반복마다) 구한다. 제안하는 방법은 다음과 같이 동작한다.

$$\mathbf{W}_i = \underset{\mathbf{W}\in\{\mathbf{W}_{i,a},\mathbf{W}_{i,n}\}}{\arg\min} L(\mathbf{W}) \qquad (1)$$

식 (1)에서 $\mathbf{W}_{i,a}$ 및 $\mathbf{W}_{i,n}$은 각각 i번째 반복에서 1) 해석적 기울기로 갱신된 가중치 파라미터 및 2) 수치해석적 기울기로 갱신된 가중치 파라미터를 나타낸다. $\mathbf{W}_i$는 i번째 반복에서 선택된 최적의 파라미터를 나타낸다. $L(\mathbf{W})$는 two-layer network에 대한 모든 가중치 $\mathbf{W}$에 대한 손실 L을 나타낸다.

Ⅲ. 실험결과

본 논문에서는 (경량 딥러닝 응용을 위한) 제안하는 방법의 성능을 평가하기 위하여 은닉층의 뉴런 개수가 20개인 two-layer network를 이용하였다. MNIST 학습데이터셋으로부터 각 클래스당 학습데이터의 개수를 제한하여 추출해 two-layer network 학습을 위한 학습데이터셋으로 이용하였다. (퓨샷 러닝을 위한) 학습데이터 개수에 따른 성능 변화를 알아보기 위하여 클래스당 학습데이터의 개수가 각각 1, 2, 3이 되도록 하였다. 성능 평가를 위해 10,000개의 샘플로 구성된 MNIST 테스트데이터셋을 이용하였다. 배치사이즈에 따른 성능 변화를 평가하기 위하여 배치사이즈를 각각 1, 2, 4가 되도록 설정하였다. 즉, 클래스당 학습데이터 개수를 일정하게 유지시키면서 배치사이즈를 각각 1, 2, 4로 하여 미니배치학습을 수행하였다. 표 1은 배치사이즈, 클래스당 학습데이터 개수의 변화에 따른 베이스라인 및 제안하는 방법 간의 성능 비교 결과를 보여준다. 표 1에서 보는 바와 같이 배치사이즈, 클래스당 학습데이터 개수 변화에 따른 제안하는 방법 및 베이스라인 간의 성능 차이가 크지 않음을 확인할 수 있었다. 또한 배치사이즈가 1일 때 제안하는 방법이 베이스라인에 비해 비교적 높은 평균 성능을 제공함을 알 수 있다.

Ⅳ. 결론

본 논문에서는 경사 하강법에서 해석적 및 수치해석적 미분 기반 최적 파라미터 선택 방법을 제안하였다. 실험결과를 통해 배치사이즈가 1일 때

표 1. 배치사이즈, 클래스당 학습데이터 개수에 따른 베이스라인 및 제안하는 방법 간의 성능 비교 : (a) 수치해석적 미분 기반 경사 하강법, (b) 해석적 미분 기반 경사 하강법, (c) 해석적 및 수치해석적 미분 기반 최적 파라미터 선택 방법. AVG는 클래스당 학습데이터 개수 1, 2, 3에 대한 각 방법의 평균 정확도.

Batch Size	Method	Number of Training Data per Class			AVG
		1	2	3	
1	(a)	0.5085	**0.5478**	0.5247	0.5270
	(b)	0.5085	0.5453	0.5559	0.5366
	(c)	0.5085	0.5453	**0.5597**	**0.5382**
2	(a)	0.5222	0.5461	**0.6139**	0.5607
	(b)	0.5221	**0.5505**	0.6138	**0.5621**
	(c)	**0.5223**	0.5462	0.6138	0.5608
4	(a)	0.5282	0.5571	0.6124	0.5659
	(b)	0.5283	**0.5575**	0.6139	**0.5666**
	(c)	**0.5284**	0.5571	**0.6140**	0.5665

제안하는 방법이 베이스라인에 비해 비교적 우수한 평균 성능을 제공함을 알 수 있었다.

참 고 문 헌

[1] Sung, Flood, et al. “Learning to compare: Relation network for few-shot learning,” *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 1199-1208, 2018.

[2] Wang, Yaqing, et al. “Generalizing from a few examples: A survey on few-shot learning,” *ACM Computing Surveys*, pp. 1-34, 2020.

[3] Lee, Y. J, et al. “Recent R&D Trends for Lightweight deep learning,” *Electronics and Telecommunications Trends*, pp. 40-50, 2019.