

정수 최적화 기반의 인공지능망 구조 탐색에서의 데이터 증강 기법 적용에 관한 연구

서지민, 이정우*
서울대학교, *서울대학교

jimin.seo@cml.snu.ac.kr, *junglee@snu.ac.kr

A Study on the application of data augmentation to integer optimization based neural architecture search

Jimin Seo, Jungwoo Lee*
Seoul National Univ., *Seoul National Univ.

요 약

본 논문은 정수 최적화 기반 인공지능망 구조 탐색 과정에서 학습 과정에서 나타나는 정확도 감소를 보완하기 위해 데이터 증강 기법을 적용하였으며, 이에 따른 성능의 향상 및 변화를 관찰하고 분석하고자 하였다.

I. 서 론

본 논문에서는 인공지능망 구조 탐색 (Neural Architecture Search, NAS)의 방법을 확장하기 위한 방법을 제안하고자 한다. 기존에 공개된 연구 중 LANA 기법을 확장하기 위해 CutMix 기반의 데이터 증강 알고리즘을 적용하였다. 이를 통하여 학습 전 단계에서의 성능 향상 및 변화를 관찰할 수 있었다.[1]

II. 본론

기존의 NAS 기법은 여러 방향의 연구들이 있었으며, 대표적으로 인공지능망 단위 구성 요소의 탐색을 통해 해당 단위 구성 요소가 어떤 연산으로 이루어져야 하는지를 선정하는 방법이 있다. 각 데이터에 이용 가능한 인공지능망 구조를 더 효율적으로 탐색할 수 있다는 점에서 강점을 가지고 있으나, 탐색에 필요한 시간이 많다는 약점을 가진다.

이러한 약점을 보완하기 위해 NAS에서는 탐색 자체를 더 빠르게 하기 위한 방법들이 많이 제안되었다. 이러한 방법 중 하나인 연산별 지식 증류 기법 (Blockwise Knowledge Distillation)은 기존에 주어진 인공지능망을 활용하여 목표로 하는 구조를 찾는 기법이다. 관련 기법 중 대표적인 예시인 DONNA에서는 간단한 선형 방정식을 응용한 성능 추정 함수를 학습한다. 표본 추정된 네트워크의 성능을 추정하여 네트워크 후보를 선택하는 기법을 이용한다.[2]

본 논문에서 이용한 LANA 알고리즘은 연산별 지식 증류 기법의 한 종류로서 탐색 단계에서 정수 최적화 기법을 이용한다. 이를 통해 설정한 목표에 부합하는 구조를 찾을 수 있다. 예를 들어, 경량화된 신경망

구성을 통해 일반적인 통신 환경에서의 응용이 가능한 형태로 발전시킬 수 있다.

$$\min_Z \min_W \sum_{(x,y) \in X_{tr}} \mathcal{L}(\mathcal{S}(x; Z, W), y)$$

where $\sum_{i=1}^N b_i^T z_i \leq \mathcal{B}$ and $1^T z_i = 1 \forall i \in [1, \dots, N]$

식 1. 인공지능망 구조 탐색의 목표 최적화 식. \mathcal{T} 는 l 개의 연산으로 구성된 목표 신경망, $\mathcal{S}(x; Z, W) = s_{l, z_l} \circ s_{l-1, z_{l-1}} \circ \dots \circ s_{1, z_1}(x; W)$ 은 후보 연산들의 l 개의 합성 함수로 이루어진 인공 신경망. $Z = [z_1, \dots, z_l]$ 은 후보 연산을 지정하는 변수.

학습 전 단계 (Pretrain)	탐색 단계 (Search)	최적화 단계 (Finetune)
$\min_{x \in X_{tr}} \sum_{i=1}^N \ s_i(x_{i-1}) - s_{i,l}(x_{i-1}; w_{ij})\ _2^2$	$\min_{z_i} \sum_{i=1}^N a_i^T z_i^{(k)}$ s.t. $\sum_{i=1}^N b_i^T z_i \leq \mathcal{B}$ and $\forall i \in [1, \dots, N]: 1^T z_i = 1$ and $\forall k' < k: \sum_{i=1}^N z_i^{(k')} \leq 0$	$\min_y \gamma_{CE} \mathcal{L}_{CE}(\mathcal{S}(x; Z, W), y) + \gamma_{KL} KL[\mathcal{T}(x) \mathcal{S}(x; Z, W)]$

도표 1. LANA 알고리즘의 단계별 구성

LANA 기법은 식 1의 최적화 과정을 두 과정으로 나누고, 이를 위해 크게 학습 전 단계 (Pretrain), 탐색 단계 (Search), 그리고 최적화 단계 (Finetune)으로 구성된다. 학습 전 단계에서는 대상 인공지능망의 각 위치에 대신 이용 가능한 연산들의 초기 가중치를 학습한다. 탐색 단계에서는 학습 전 단계를 마친 후의 성능 하락을 이용해 목표식에 부합하는 구조를 탐색한다. 마지막으로, 최적화 단계에서는 탐색한 구조를 대상 데이터에 대해 학습하는 단계이다. 해당 알고리즘은 각각의 단계를 통해 최적 구조를 찾아내게 된다.

그러나, 학습 전 단계를 마친 가중치를 사용하여 성능 하락을 계산하였을 때 특정 깊이에서 성능의 하락이

크게 나타나는 경우가 나타난다. 특히 연산 중 공간적인 크기의 변화가 있는 연산에서 성능 하락이 나타나는 문제가 있는데, 이는 학습 전 단계에서 인공신경망의 가중치를 충분히 학습하지 못해 발생하는 문제로 판단된다. 이에 따라, 본 논문에서는 데이터 증강(Data Augmentation) 기법을 적용해 이 점을 개선하고자 한다. 특히, 본 논문에서는 CutMix 기법을 학습 전 단계 과정에 적용하여 성능을 확인하고자 한다. LANA 알고리즘 상에서는 CutMix 알고리즘을 적용 가능한 위치는 학습 전 단계와 최적화 단계이며, 이 중 학습 전 처리 과정에 적용하여 성능을 비교하고자 하였다.[3]

실험은 ImageNet 데이터 중 무작위 시드를 고정하고 추출한 5%의 데이터에 대하여 진행하였다. 일부 데이터에 대하여 진행하더라도 전체 데이터에서 얻은 결과와 유사한 수준의 성능을 얻을 수 있었으므로 유의미한 수준의 성능 비교가 가능하다. 실험에서는 δ 에 이용되는 후보 연산으로 3 개의 연산 후보군을 사용하였다. 실험에 사용한 매개변수들은 각각 학습 전 단계에서 학습률이 $8E-03$, $\gamma_{MSE} = 0.1$ 이다.

깊이	연산 별 정확도 [단위: $\Delta\%$]		
	(1)	(2)	(3)
0	-1.56	0	-5.92
1		-2.4	-4.28
2		0	0.68
3		-0.72	10.24
4		-0.92	1.88
5		0	-0.04
6		-1.92	-0.4
7		-1.52	-0.16
8		0	-0.08
9		-1.56	-0.08
10		-1.64	-0.04
11		-1.24	-0.2
12		0	-0.08
13		-1.8	-0.2
14		-1	-0.12
15		-1.4	0.16
16		0	0
17		-1.28	0.2
18		-1.44	0.2
19		-0.92	0.16
20		-2.36	0.12
21		0	0.12
22		-1.88	-0.04

표 1. 학습 전 단계 이후 성능 차이. 깊이는 연산 중 특정 깊이의 연산을 해당하는 열의 연산으로 교체한 경우를 나타냄. 정확도를 측정하였으며, (기준 기법) - (CutMix 적용 기법)을 계산. (1)은 대상 신경망의 연산, (2)는 Identity 연산, 그리고 (3)은 Mobile Inverted Residual Block 을 사용함.

표 1 에서 많은 수의 후보 연산들이 정확도 상승이 있음을 확인할 수 있다. 이에 따라, 데이터 증강 기법을 적용하면 학습 전 단계에서 더 높은 정확도를 얻을 수 있음을 확인할 수 있다.

위의 실험을 통해 CutMix 알고리즘이 학습 전 단계에 적용되는 경우 성능의 향상을 확인할 수 있다. 일반 함수 근사 정리(Universal function approximation theorem)에 의해 임의의 연산을 인공신경망이 근사할 수 있으므로 최소 정확도 감소는 0 이며, 학습 전 단계에서 추정된 연산의 성능은 각 연산이 잠재적으로 추정할 수 있는 성능을 예측하는데 도움이 될 것이다. 따라서, 이 과정에서 연산들이 더 높은 정확도를 얻을 수 있다면 인공신경망 후보의 최종 성능을 추정하는데 도움이 될 수 있다. 이 점에서 데이터 증강 알고리즘이 LANA 알고리즘의 탐색 과정을 보조함을 확인할 수 있다.

	일반 학습	CutMix 적용시
정확도 [단위: %]	57.76	43.80

표 2. 최적화 단계 후 성능 비교.

그러나 일부 연산에서는 큰 성능 향상을 얻지 못하는 경우가 있으며, 일부는 극단적인 성능 저하가 발생하는 경우가 나타난다. 또한 표 2 에서와 같이 이후 최적화 과정에 적용될 때 최종적인 성능에 악영향을 미치는 경우가 나타나는 경우를 확인하였다. 이는 CutMix 등의 기법을 직접적으로 적용하는 것이 학습 전 단계의 특정 연산 혹은 최적화 과정에 적합한 형태로 데이터 증강 기법을 적용하지 못해 발생한 것으로 추정된다. 추후 연구에서는 이러한 영향을 줄이기 위해 초매개변수 등의 조정과 관련된 연구가 필요하며, 더 나아가 정확도가 떨어지지 않는 것을 보장할 수 있도록 이론적인 접근이 필요하다.

III. 결론

본 논문에서는 NAS 계열의 연구 중 정수 최적화 기법을 활용한 LANA 기법을 보완하기 위해 CutMix 알고리즘을 적용하는 방법에 대하여 제안하였다. 이를 통해 더 좋은 구조를 선택할 수 있음을 확인하였다.

ACKNOWLEDGMENT

This work is in part supported by Institute of Information & communications Technology Planning & Evaluation (IITP, 2021-0-00106 (50%), 2021-0-02068 (50%)) grant funded by the Ministry of Science and ICT (MSIT), INMAC, and BK21-plus

참 고 문 헌

- [1] Molchanov, Pavlo, et al. "LANA: latency aware network acceleration." CoRR, abs/2107.10624 v2 8 (2021).
- [2] Moons, Bert, et al. "Distilling optimal neural networks: Rapid search in diverse spaces." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.
- [3] Yun, Sangdoo, et al. "Cutmix: Regularization strategy to train strong classifiers with localizable features." Proceedings of the IEEE/CVF international conference on computer vision. 2019.