

Prototype Augmentation 과 Label Augmentation 을 이용한 MLP-Mixer 의 연속 학습 성능 개선

엄정민, 이재구*
국민대학교

*jaekoo@kookmin.ac.kr

Improving Performance in Continual Learning using MLP-Mixer with Prototype Augmentation and Label Augmentation

Jungmin Eom, Jaekoo Lee*
College of Computer Science, Kookmin Univ.

요 약

다중 퍼셉트론 레이어(Multi-Layer Perceptron, MLP)로만 구성된 MLP-Mixer 는 유도 편향(Inductive Bias)에 이점이 있지만 새로운 과업(Task)을 순차로 학습할 때 과거에 학습했던 정보를 잃어버리는 치명적 망각(Catastrophic Forgetting) 현상이 발생하는 문제가 있다. 이는 모델의 가중치가 새로운 과업에 편향되기 때문이다. 이를 해결하기 위해 본 논문은 MLP-Mixer 에 연속 학습 모델인 PASS 의 Prototype Augmentation(ProtoAug)과 Label Augmentation(LabelAug)을 적용할 것을 제안한다. 또한, 모델 규제를 위해 지식 증류(Knowledge Distillation, KD)도 추가로 적용했다. 그리하여 MLP-Mixer 기본 모델과 MLP-Mixer 모델에 ProtoAug, 지식 증류를 적용한 모델, 그리고 LabelAug 까지 모두 적용한 모델의 연속 학습 성능을 비교했다. 그 결과, ProtoAug 와 지식 증류를 사용한 모델이 기본 모델보다 평균 40.72%p 상승함을 확인할 수 있었다. LabelAug 까지 모두 적용한 모델은 앞서 2 가지를 적용한 모델보다 평균 3.49%p 상승함을 확인할 수 있었다.

I. 서 론

기존 컴퓨터 비전 분야는 합성곱 계층(Convolutional Layer)으로만 구성된 신경망 모델이 많다. 하지만 다중 퍼셉트론 (Multi-Layer Perceptron, MLP)로만 구성된 MLP-Mixer[1]는 유도 편향(Inductive Bias)을 완화하면서 합성곱 신경망(Convolutional Neural Network, CNN)에 준하는 성능을 낼 수 있다는 것을 보였다[1].

현재 많은 심층 신경망 모델은 새로운 과업(Task)을 순차적으로 학습하면 이전에 학습했던 정보를 잃어버리는 치명적 망각 현상(Catastrophic Forgetting)이 발생한다[2]. 이를 방지하고자 새로운 정보를 학습함과 동시에 이전에 학습했던 정보를 보존하도록 하는 연속 학습(Continual Learning) 연구가 진행되고 있다.

본 논문은 기존 MLP-Mixer 모델의 치명적 망각 현상을 확인하고, 연속 학습 모델 PASS[3]의 Prototype Augmentation(ProtoAug, PA)과 Label Augmentation(LabelAug, LA)[7]을 적용하여 MLP-Mixer의 치명적 망각 현상을 완화하고자 한다.

II. 본 론

MLP-Mixer는 다중 퍼셉트론만 구성된 모델이다. 작동 방식은 [그림 1]의 (a)와 같다. 입력 이미지를 패치(Patch) 단위로 잘라서 패치 임베딩(Patch Embedding)을 통과하여 토큰(Token) 단위로 임베딩 된다. 토큰은 믹서 층(Mixer Layer)을 통과하여 이미지의 위치별 특징(Per-Location feature)과 공간적 특징(Spatial Feature)을 학습하도록 한다. 믹서 층에서 나오는 특징(Feature)은 전역 평균 풀링(Global Average Pooling)을 통과하여 분류기(Classifier)에서 입력 이미지를 분류하도록 한다.

PASS의 주요 아이디어는 [그림 1]의 (b)와 같다. 현재 과업의 데이터는 특징 추출기(Feature Extractor)에 입력된다. 추출기에서 출력된 특징들은 종류(Class)별로 군집화된다. 군집화된 특징들은 평균 낸 후 저장하도록 하고, 이렇게 저장된 특징을 Prototype이라 한다.

다음 과업을 학습할 때 저장된 Prototype에 가우스 잡음(Gaussian Noise)을 추가한다. 이를 ProtoAug라 하고 분류기에서 활용된다. 또한 PASS는 특징 추출기를 규제하기 위해 지식 증류(Knowledge Distillation, KD)[5, 6]를 사용한다. 그리고 각 과업을 학습할 때 종류마다 90°, 180°, 270°로 회전하여 레이블(Label) 개수를 추가하는 LabelAug를 사용한다.

ProtoAug는 심층 신경망이 학습한 특징 공간(Feature Space)이 과하게 왜곡되는 것을 방지하며, 분류기가 특정 과업에 편향되지 않도록 한다. 또한 LabelAug는 모델이 모든 과업에 대해 일반화된 특징을 학습하도록 한다[3].

본 논문은 기본 MLP-Mixer와 ProtoAug, 지식 증류를 적용한 MLP-Mixer(PA+ KD), 그리고 LabelAug까지 모두 적용한 MLP-Mixer(PA+ KD+ LA)의 성능을 비교하였다. 데이터 집합은 CIFAR-100[4]을 사용했다. 첫 과업에서는 50개의 종류를 학습한 후, 두 번째 과업부터 한 과업당 5개의 종류를 학습하여 총 11개의 점진적 종류 증가 과업(Incremental Class Task)을 학습하도록 했다. 각 과업의 테스트 결과는 첫 번째부터 해당 과업까지 모든 테스트 데이터를 통합한 정확도를 의미한다.

본 논문의 실험 결과는 [그림 2]와 같다. 기본 MLP-Mixer는 첫 번째 과업 이후의 성능이 매우 낮게 나왔다. 반면에 ProtoAug와 지식 증류를 적용한 MLP-Mixer(PA+ KD)는 기본 모델보다 평균 40.72%p 상승했

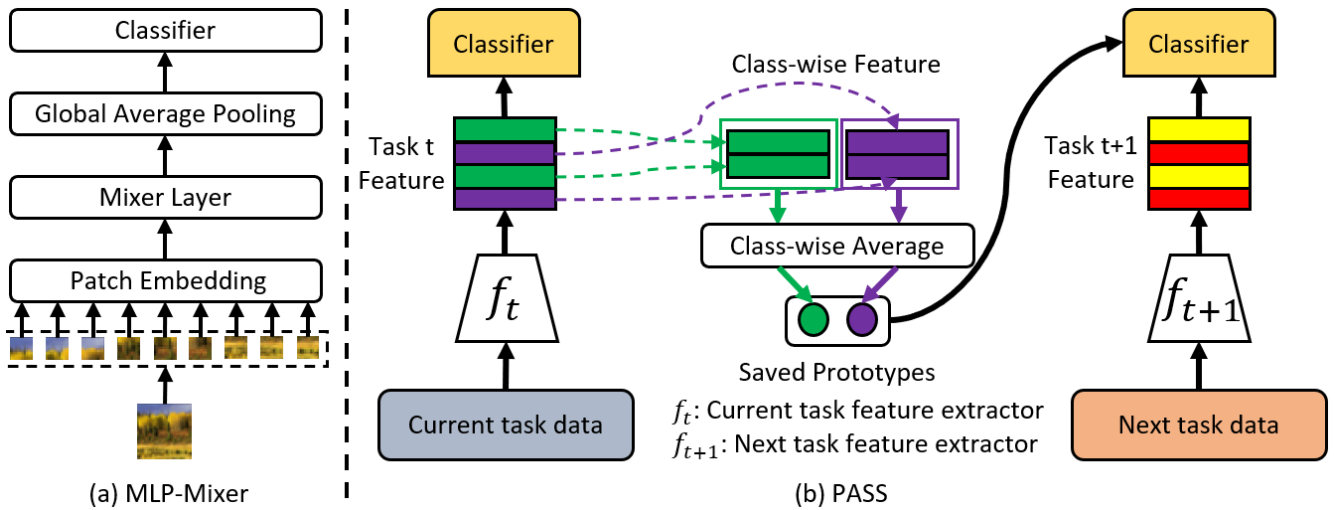


그림 1. MLP-Mixer 와 PASS 의 구조

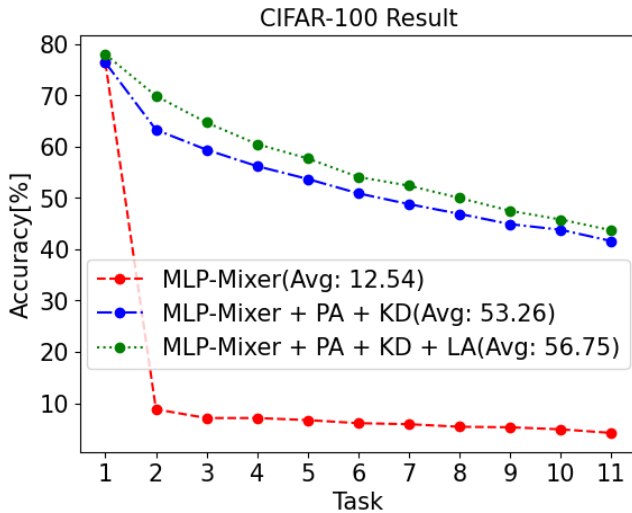


그림 2. MLP-Mixer 의 연속 학습 실험 결과

다. 그리고 LabelAug까지 모두 적용한 MLP-Mixer(PA+ KD+ LA)는 MLP-Mixer(PA+ KD)보다 평균 3.49%p 상승했다.

MLP-Mixer는 모델의 특징 추출기와 분류기 모두 새로운 과업에 편향되어 이전 과업들의 종류를 분류하지 못했다. 반면에, MLP-Mixer(PA+ KD)는 새로운 과업 데이터와 과거 과업의 특징을 같이 학습하므로 모델이 새로운 과업에 편향되지 않는다. MLP-Mixer(PA+ KD+ LA)는 LabelAug를 통해 모든 과업에서 일반화가 잘 된 특징을 학습하였다.

따라서 본 논문은 유도 편향 이점이 있는 MLP-Mixer가 순차적인 데이터를 효과적으로 연속 학습하기 위해 PASS의 3가지 핵심 요소를 적용한 MLP-Mixer(PA+ KD+ LA)를 제안한다.

III. 결 론

본 논문에서는 순차로 데이터를 학습할 때 치명적 망각 현상이 발생하는 MLP-Mixer를 개선하기 위해 ProtoAug와 LabelAug를 적용했다. MLP-Mixer(PA+ KD)는 기본 모델보다 평균 성능이 40.72%p 상승한 것을 확인했다. 이는 지식 종류가 모델을 적절히 규제하면서 ProtoAug가 모델의 특징 추출기와 분류기의 새로운 과업 편향 현상을 완화한 것이다. 모두 적용한 MLP-Mixer(PA+ KD+ LA)는 MLP-Mixer(PA+ KD)보다 평균 3.49%p 상승했다. 이는 LabelAug를 통해 모든 과업에 일반화가 잘 된 특징을 학습한 것이다.

이를 통해 본 논문은 순차적 데이터를 학습할 수 있는 MLP-Mixer(PA+ KD+ LA)를 제안하며, 현재 방대한 이미지가 생성되는 상황에서 데이터를 순차로 학습할 때 연속 학습 모델의 중요성을 확인할 수 있었다.

ACKNOWLEDGMENT

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.RS-2022-00167194, 미션 크리티컬 시스템을 위한 신뢰 가능한 인공지능).

참 고 문 헌

- [1] Tolstikhin, Ilya O., et al. "Mlp-mixer: An all-mlp architecture for vision." *Advances in Neural Information Processing Systems* 34 (2021): 24261–24272.
- [2] Van de Ven, Guido M., and Andreas S. Tolias. "Three scenarios for continual learning." *arXiv preprint arXiv:1904.07734* (2019).
- [3] Zhu, Fei, et al. "Prototype augmentation and self-supervision for incremental learning." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.
- [4] Krizhevsky, Alex, and Geoffrey Hinton. "Learning multiple layers of features from tiny images." (2009): 7.
- [5] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. "Distilling the knowledge in a neural network." *arXiv preprint arXiv:1503.02531*, 2015.
- [6] Saihui Hou, Xinyu Pan, Chen Change Loy, Zilei Wang, and D. Lin. "Learning a unified classifier incrementally via rebalancing." In *CVPR*, pages 831–839, 2019.
- [7] Lee, Hankook, Sung Ju Hwang, and Jinwoo Shin. "Self-supervised label augmentation via input transformations." *International Conference on Machine Learning*. PMLR, 2020.