

# 커리큘럼 학습의 우선순위 정책에 따른 이미지 분류 성능 비교

손건희 · 윤성현

국립공주대학교

## Performance Comparison according to the Priority Modes for Curriculum Learning in Image Classification

Gun-Hee Son · Sunghyun Yoon\*

Kongju National University

E-mail : songunhee5426@gmail.com / syoon@kongju.ac.kr

### 요 약

커리큘럼 학습은 인간의 학습 방법을 모방한 학습 방법론으로, 인간이 유년기 시절에는 쉬운 것을 배우고 나이가 들면서 점점 어려운 것을 배운다는 점에 착안해, 쉬운 데이터부터 시작해서 점차 어려운 데이터를 학습해 나가는 방법이다. 한편, 어려운 데이터 위주로 학습을 진행하는 커리큘럼 학습 방법을 통해 성능이 향상된 사례 또한 존재한다. 본 연구에서는 커리큘럼 학습 알고리즘 중 하나인 SuperLoss를 기반으로, 우선순위 정책 (e.g., 쉬운 데이터 우선, 어려운 데이터 우선) 에 따른 이미지 분류 성능을 평가하고 그 결과를 분석한다.

### ABSTRACT

Curriculum learning is a learning strategy that mimics human learning methods, training a model from easier data to harder data (i.e., easy-first). Meanwhile, there have been several studies applying the learning strategy of training from harder data to easier data (i.e., hard-first) and then achieving performance improvements. that achieves performance improvements with a variant of curriculum learning, where models are trained from harder data to easier data (i.e., hard-first). Based on SuperLoss, a simple yet powerful curriculum learning method, this study aims to evaluate and analyze the performances of image classification models according to the priority modes (e.g., easy-first and hard-first) for curriculum learning.

### 키워드

curriculum learning, deep learning, image classification, priority mode, superloss

### 1. 서 론

커리큘럼 학습은 쉬운 개념부터 학습한 뒤 점차 어려운 개념을 학습하는 인간의 학습 방식을 모방한 학습 방식이다 [1]. 모든 학습 표본의 중요도를 동등하게 하는 기존의 학습 방식과 달리, 커리큘럼 학습은 각 표본의 중요도를 다르게 설정하고, 중요도가 높은 표본일수록 모델 학습에 미치는 기여도를 높인다. 예를 들어, 손실함수의 값이 큰 표본 (i.e., 어려운 표본) 의 중요도는 낮추고 손실함수의 값이 작은 표본 (i.e., 쉬운 표본) 의 중요도는 높여

모델이 쉬운 표본을 먼저 학습하고 어려운 표본을 나중에 학습하도록 한다. 지금까지 다양한 종류의 커리큘럼 학습 방법이 개발되었으며, 여러 연구에서 학습 수렴 속도 및 성능의 향상이 확인된 바 있다 [2]. 본 연구에서는 표본의 중요도를 학습을 통해 추정하는 방식이 아닌, 표본의 중요도에 대한 해석적 해 (analytic solution) 가 존재하는 SuperLoss-s [3]를 기반으로 한다.

SuperLoss를 포함한 기존의 커리큘럼 학습 방법 대부분이 여러 연구에서 성능 개선이 확인된 것은 사실이다. 그렇지만, 개선되어야 할 부분 또한 존재한다. 본 연구에서는 다음의 두 가지 측면에 중점을 둔다. 첫째, SuperLoss를 포함한 대부분의 커

\* Corresponding author

리큐럼 학습 방법은 쉬운 표본 우선 (easy-first) 정책을 기반으로 한다. 커리큐럼 학습의 동기가 되는 인간의 학습 방식이 쉬운 개념을 우선 학습하는 방식이기에 커리큐럼 학습 또한 기본적으로는 쉬운 표본을 우선시하기는 한다. 하지만, 쉬운 표본 우선 정책 이외에도 어려운 표본 우선 (hard-first), 중간 난도 표본 우선 (medium-first) 등의 정책들도 존재한다 [4]. 어떤 우선순위 정책이 가장 효과적 인지는 주어진 문제 상황에 따라 달라지므로, 이에 대한 검증이 필요하다. 둘째, 기존의 커리큐럼 학습 방법 대부분은 고정된 우선순위 정책을 기반으로 한다. 예를 들어, 쉬운 표본 우선 (easy-first) 기반 방식은 학습 시작부터 끝까지 쉬운 표본을 우선시한다. 본 연구에서는 고정된 우선순위 정책은 최적의 방법이 아닐 것이라 가정한다. 인간의 경우에도, 처음에는 쉬운 개념 위주로 학습해 기초를 다지는 데 중점을 두다가 (i.e., easy-first) 어느 정도 기본기가 잡힌 이후에는 심화 개념을 학습한다 (e.g., hard-first). 이러한 사실에 기초해, 기계학습에서의 커리큐럼 학습 또한 우선순위 정책을 전환하는 것이 효과적일 것이라 가정한다.

위에서 언급한 2가지 주제를 다루기 위해, 본 연구에서는 다음의 2가지 방법을 제안한다. 첫째, 쉬운 표본 우선 (easy-first) 방식인 기존의 SuperLoss (SL) 를 변형해, 어려운 표본 우선 (hard-first) SuperLoss (HFSL) 와 중간 난도 표본 우선 (medium-first) SuperLoss (MFSL) 를 제안한다. 둘째, 우선순위 정책의 변경이 가능한 SuperLoss (mode-changeable SuperLoss; MCSL) 를 제안한다.

본 연구에서는 이미지 분류 문제를 통해 제안한 방법의 효과를 확인하고자 한다. 이를 위해, 기존 방법 및 제안한 방법의 이미지 분류 정확도를 평가하고 그 결과를 비교·분석한다.

## II. SuperLoss 및 그 변형

[3]에서 제안한 SuperLoss는 식 (1)과 같이 정의된다.

$$SL(l_i) = (l_i - \tau)\sigma_i + \lambda(\log\sigma_i)^2 \quad (1)$$

여기서  $l_i$ 는  $i$ 번째 표본의 원 (original) 손실함수 (e.g., cross-entropy) 값,  $\tau$ 는 표본의 손실함수 값의 이동평균 (moving average),  $\sigma_i (> 0)$ 는  $i$ 번째 표본의 중요도,  $\lambda (> 0)$ 는  $l_i$ 가  $\sigma_i$ 에 미치는 영향을 조절하는 hyperparameter이다. 최적의  $\sigma_i$  값은 식 (1)의 해석적 해 (analytic solution) 에 해당하며,  $l_i$ 를 상수 취급하여 계산하면 얻을 수 있다.

$l_i$  값이 주어지면,  $\sigma_i$ 는 식 (1)의 해석적 해 (analytic solution) 를 구하는 방법을 사용해 직접 구할 수 있다.

어려운 표본 우선 SuperLoss (HF SuperLoss) 는 식 (2)와 같이 정의된다.

$$HFSL(l_i) = (l_i - \tau)\sigma_i - \lambda(\log\sigma_i)^2 \quad (2)$$

식 (1)의 SuperLoss와 마찬가지로, 식 (2)에서의 최

적의  $\sigma_i$  값은 해석적 해를 구함으로써 얻을 수 있다.

중간 난도 표본 우선 SuperLoss (MF SuperLoss) 는 쉬운 표본 우선 SuperLoss와 어려운 표본 우선 SuperLoss를 조합한 것으로 정의된다.  $l_i - \tau = 0$ 이 되는 지점을 기준으로, 왼쪽에는 HFSL의 그래프를  $l_i$  축으로  $-e^{-1}$  만큼 평행이동한 부분을, 오른쪽에는 EFSL의 그래프를  $l_i$  축으로  $e^{-1}$  만큼 평행이동한 부분을 연결함으로써 얻을 수 있다.

우선순위 정책 변경이 가능한 SuperLoss (MCSL) 는 위의 3가지 SuperLoss 방법들을 조합해 구현할 수 있다. 예를 들어, 변경 전/후 우선순위 정책이 각각 easy-first, hard-first 라면, MCSL은 식 (3)과 같이 정의된다.

$$MCSL_{EF \rightarrow HF}(l_i) = \begin{cases} SL(l_i) & (e_c \leq e_{thr}) \\ HFSL(l_i) & (e_c > e_{thr}) \end{cases} \quad (3)$$

여기서  $e_c$ 는 현재 epoch,  $e_{thr}$ 는 우선순위 정책을 변경하는 시점에서의 epoch를 의미한다.

## III. 실험 설계

본 연구에서는 CIFAR-10 데이터셋을 사용해 이미지 분류 모델을 학습하고 평가했다. 정답 레이블 (label) 에 노이즈 (noise) 가 존재하는 상황도 추가로 고려하기 위해, 전체 학습 데이터의 10%, 30%, 50%에 해당하는 표본을 무작위로 추출하고 이들의 레이블을 무작위로 할당했다.

이미지 분류 모델은 compact convolutional transformer (CCT) [5]를 사용해 구현했다. RAdam optimizer, 학습률 0.0005, 배치 크기 128 조건에서 100 epochs 만큼 반복 학습했다.

실험은 크게 2가지로 진행했다. 먼저, baseline (i.e., 커리큐럼 학습 미적용) 과 3가지 SuperLoss의 성능을 비교했다. 다음, 3가지 SuperLoss 중 분류 정확도가 가장 높은 방법을 변경 전 정책으로 하고, 남은 2가지 SuperLoss 방법을 변경 후 정책으로 하는 MCSL 방법의 분류 정확도를 평가했다.  $\lambda = 1$ ,  $e_{thr} = 20$ 으로 설정했다.

## IV. 결과 및 분석

그림 1 및 표 1은 CIFAR-10 평가 세트에서의 각 조건 별 분류 정확도를 나타낸다. 레이블 노이즈가 없는 경우 (i.e., 0% corruption ratio), 방법에 관계없이 비슷한 성능을 보였다. 노이즈가 약한 (e.g., 10%, 30%) 경우에는 우선순위 정책을 쉬운 표본 우선에서 중간 난도 우선으로 변경하는 (easy-to-medium) 방법이 가장 높은 정확도를 보였고, 중간 난도 표본 우선 방법 (medium-first) 과 쉬운 표본 우선 (easy-first) 방법이 그 뒤를 이었다. 노이즈가 강한 (e.g., 50%) 경우에는 쉬운 표본 우선 (easy-first

st) 방법이 가장 높은 정확도를 보였다. 어려운 표본 우선 (hard-first) 방법은 노이즈가 있는 조건에서는 일관되게 낮은 성능을 보였다.

일반적으로, 노이즈가 없는 표본은 쉬운 표본에, 노이즈에 의한 왜곡이 적은 (i.e., 변경 전/후 레이블의 특성이 비슷한 경우) 표본은 중간 난도의 표본에, 노이즈에 의한 왜곡이 심한 (i.e., 변경 전/후 레이블의 특성 차이가 큰 경우) 표본은 어려운 표본에 해당한다고 볼 수 있다. 이러한 점을 고려하면, 약한 노이즈 조건에서 easy-to-medium의 성능이 높게 나타난 것은 학습 초기에는 노이즈가 없는 깨끗한 표본 위주로 학습해 학습 세트의 대략적인 분포를 모델링한 뒤, 이후에는 왜곡이 적은 표본 위주로 학습해 보다 복잡한 분포를 모델링함으로써 일반화 성능을 높였기 때문으로 설명할 수 있다. 깨끗한 표본을 사용한 학습 이후에 곧바로 왜곡이 많은 표본 위주로 학습하는 경우, 급격하게 변화하는 학습 세트의 분포로 인해 오히려 왜곡된 결과를 초래할 수 있음을 easy-to-hard의 결과를 통해 확인할 수 있었다.

한편, 강한 노이즈 조건에서 easy-first의 성능이 높게 나타난 것은, 라벨의 왜곡이 심한 표본은 어느정도 학습된 모델 입장에서도 설명하기 어려운 표본임을 의미한다. 이 경우, easy-first는 설명하기 어려운 어려운 표본을 배제하고, 노이즈가 없는 쉬운 표본 및 쉬운 표본의 분포를 크게 왜곡하지는 않아 일반화 성능 향상을 기대해볼 수 있는 중간 난도의 표본을 위주로 학습하기 때문에 성능이 향상된 것이라 설명할 수 있다.

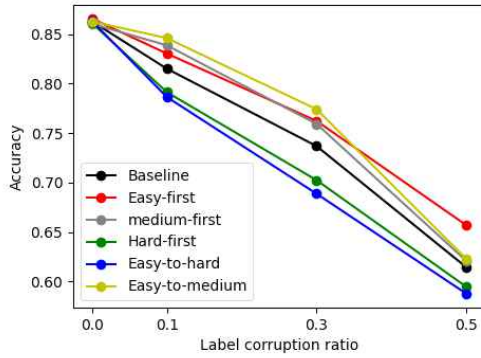


그림 1. Label corruption ratio에 따른 각 방법의 CIFAR-10 평가 세트에서의 분류 정확도

표 1. Label corruption ratio에 따른 각 방법의 CIFAR-10 평가 세트에서의 분류 정확도

Corruption \ Method	0%	10%	30%	50%
Baseline	0.86	0.81	0.74	0.61
Easy-First	<b>0.87</b>	0.83	0.76	<b>0.66</b>
Hard-First	0.86	0.79	0.70	0.59
Medium-First	0.86	0.84	0.76	0.62

Easy-to-Hard	0.86	0.79	0.69	0.59
Easy-to-Medium	0.86	<b>0.85</b>	<b>0.77</b>	0.62

## V. 결 론

본 논문에서는 커리큘럼 학습 방법을 이미지 분류 문제에 적용하는 실험을 진행했다. CIFAR-10 데이터세트 및 CCT를 사용해 진행한 실험 결과, 노이즈가 없는 상황에서는 모든 방법의 성능이 비슷하게 나타났다. 노이즈가 심하지 않은 경우에는 easy-to-medium 정책이, 노이즈가 심한 경우에는 easy-first 정책의 성능이 가장 높게 나타났다.

향후 연구로는 실환경에 보다 가까운 조건에서의 경향을 확인하기 위해 ImageNet과 같은 대규모 데이터세트에서 실험을 진행하고자 한다. 또한, CCT 외의 다른 구조의 모델에 대해서도 동일한 경향이 관찰되는지 확인할 계획이다.

## Acknowledgement

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. RS-2022-00165827).

## References

- [1] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th annual International Conference on Machine Learning (ICML)*, pp. 41-48, 2009.
- [2] X. Wang, Y. Chen, and W. Zhu, "A survey on curriculum learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 9, pp. 4555-4576, 2021.
- [3] T. Castells, P. Weinzaepfel, and J. Revaud, "SuperLoss: A generic loss for robust curriculum learning," *Advances in Neural Information Processing Systems (NIPS)*, pp. 4308-4319, 2020.
- [4] X. Zhou and O. Wu, "Which samples should be learned first: Easy or hard?," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1-15, 2023.
- [5] A. Hassani, S. Walton, N. Shah, A. Abuduweili, J. Li, and H. Shi, "Escaping the big data paradigm with compact transformers," 2021, *arXiv:2104.05704*.