

저품질 샘플에 강인한 지식 증류를 위한 출력 분포 조정 방법

조유철*

국립한밭대학교

yccho@hanbat.ac.kr

Sample Quality-Based Output Adjustment in Knowledge Distillation

Yucheol Cho
Hanbat National University

Abstract

지식 증류(Knowledge Distillation, KD)는 대규모 교사(teacher) 모델의 지식을 소규모 학생(student) 모델로 전달하는 대표적인 모델 경량화 기법이다. 그러나 실제 이미지 데이터에는 종종 저품질(low-quality) 샘플이 포함되어 있어, 이들이 증류 과정의 성능을 저하시킬 가능성성이 존재한다. 본 연구에서는 이러한 저품질 샘플의 영향을 완화하기 위해 학습 과정에서 샘플 품질을 구분하고, 이에 따라 교사 모델의 soft label 분포를 보정하는 품질 인지형(quality-aware) 지식 증류 기법을 제안한다. 이를 위해 가중치가 0에 가까운 값을 더 많이 갖도록 유도한 희소(sparse) 모델을 구성하여 학습이 용이한 샘플과 어려운 샘플을 구분하고, 그 정보를 활용해 soft label의 분포를 조정하였다. 실험은 Dogs, CUB-200, MIT67, Stanford40 등 세밀한 구분이 필요한 fine-grained 인식 데이터셋에서 수행되었으며, 제안 기법은 기존 baseline 대비 우수한 성능을 보였다. 이를 통해 soft label 분포의 세밀한 조정이 학생 모델의 학습 효율과 일반화 성능 향상에 중요한 역할을 함을 확인하였다.

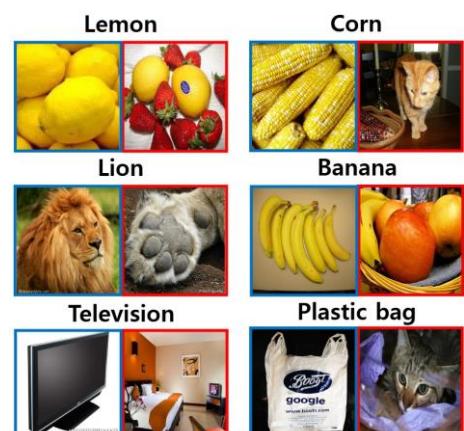
I. 서론

지식 증류(Knowledge Distillation, KD)는 교사 모델이 가진 풍부한 표현 능력을 학생 모델이 효율적으로 학습하도록 돋는 방법으로, 다양한 딥러닝 응용 분야에서 활용되고 있다 [1]. 하지만 실제 데이터셋에는 잡음이 많거나 저품질인 샘플이 포함되어 있으며, 이러한 데이터는 증류 과정에서 학생 모델의 학습을 방해하거나 과적합을 유발할 수 있다 [2]. 본 연구에서는 이러한 문제를 해결하기 위해, 샘플의 품질에 따라 교사 모델의 출력 분포를 동적으로 보정하는 방법을 제안한다. 이를 통해 학생 모델이 학습하기 용이한 데이터에 집중하도록 유도하고, 불필요한 잡음으로부터의 영향을 최소화하고자 한다.

II. 본론

먼저, 희소 모델을 생성하여 일부 가중치가 0으로 수렴하도록 유도한 보조 모델을 생성한다. 이 모델은 자연스럽게 샘플의 난이도나 품질을 구분할 수 있으며, 이를 통해 각 샘플의 학습 난이도를 평가한다. 이후, 교사 모델의 soft label 분포를 샘플 품질에 따라 조정하여, 품질이 높은 샘플에는 보다 뚜렷한(Sharper) 분포를, 품질이 낮은 샘플에는 완화된(Smoothen) 분포를

적용한다. 학생 모델은 이러한 품질 기반 soft label을 참조하여 학습함으로써, 효율적이고 안정적인 지식 증류가 가능하도록 설계하였다.



[그림 1] 품질이 높은 (파란색 박스) / 품질이 낮은 (빨간색 박스) 이미지들의 예시.

	CUB200	MIT67
KD	64.32	61.37
FitNet	65.03	62.34
AT	64.87	62.34
SP	64.09	61.96
VID	64.44	61.44
HKD	64.38	62.31
Ours	65.45	62.81

[표 1] 본 연구에서 제안하는 방법을 적용한 경우와 기존 KD 기법들에 대한 성능 비교.

III. 결론

본 연구에서는 지식 종류 과정에서 샘플 품질을 고려한 출력 분포 보정 방법을 제안하였다. 제안 기법은 저품질 샘플의 부정적 영향을 완화하고, 학생 모델이 보다 세밀한 차이를 인식할 수 있도록 학습을 유도한다. 실험 결과, fine-grained 인식 데이터셋 전반에서 기존 방법 대비 성능 향상을 확인하였으며, 이는 soft label 분포 조정이 지식 종류의 효율성과 강건성을 높이는 중요한 요인임을 시사한다.

REFERENCES

- [1] Cho, Yucheol, et al. "Ambiguity-aware robust teacher (ART): Enhanced self-knowledge distillation framework with pruned teacher network." Pattern Recognition 140 (2023): 109541.
- [2] Ham, Gyeongdo, et al. "Difficulty level-based knowledge distillation." Neurocomputing 606 (2024): 128375.