

# A Variation and An Optimization of Loss Weight in Hybrid Models for Semi-Supervised Learning

성균관대학교 데이터사이언스융합과 김한비 / 지도교수 김재광

## 연구목적 및 배경

- 기존 준지도 학습 (Semi-supervised learning) 하이브리드 모델 (Hybrid model)에서 모든 학습에 사용되는 배치(Batch) 별로 다른 데이터셋이 사용되지만, 고정된 하이퍼 파라미터 (Hyper-parameter)를 사용하는 것에 대한 문제 제기

## 연구 방법

- 점진증강법 (Ramp-Up method)을 활용하여, 해당 변수를 훈련이 지속됨에 따라 지속적으로 증강시키며 고정된 변수의 사용보다 유기적으로 변화는 변수의 사용이 성능이 더 좋다는 것을 증명  
- 강화학습(Reinforcement learning)의 Q-learning을 통해서 점진증강법에서 사용된 여러 비율을 최적화하여 훈련중 지속적으로 변화시키는 방법 제시

## 실험 데이터 및 환경

- 이미지 분류 벤치마크 데이터셋  
- 각 데이터셋 별 3개의 레이블 그룹 존재  
[ 적은 레이블 | 중간 레이블 | 많은 레이블 ]  
- 각 데이터셋 별 레이블 별 하이브리드 모델 적용 및 제안 방법 적용



Dataset Name	CIFAR-10	CIFAR-100	STL-10	SVHN
Images [examples]	Objects/Animals [cars, cats ...]	Objects/Animals [cars, cats ...]	Objects/Animals [cars, cats ...]	Street Numbers [0,1,2...]
Number of classes	10	100	10	10
Size of Images	32x32 Pixel	32x32 Pixel	96x96 Pixel	32x32 Pixel
Train Data Size [labeled condition]	50,000 [Fully labeled]	50,000 [Fully labeled]	5,000 x 100,000 [Partially labeled]	73,257 [Fully labeled]
Test Data Size	10,000	10,000	8,000	26,032

Table 4-1: Explanation of Dataset used in experiments

## 실험 결과

- [적은 레이블]: 점진증강법으로 지속적으로 정확도가 증가하는 것을 증명  
- [중간/많은 레이블]: 초반 구간에는 증가하지만 실험이 지속될 수록 정확도 감소  
- 점진증강법을 통해 레이블이 적을수록 하이퍼 파라미터 변화가 성능을 증가시킴  
- 강화학습을 통해 하이퍼 파라미터에 대한 최적화를 진행하여 성능 향상

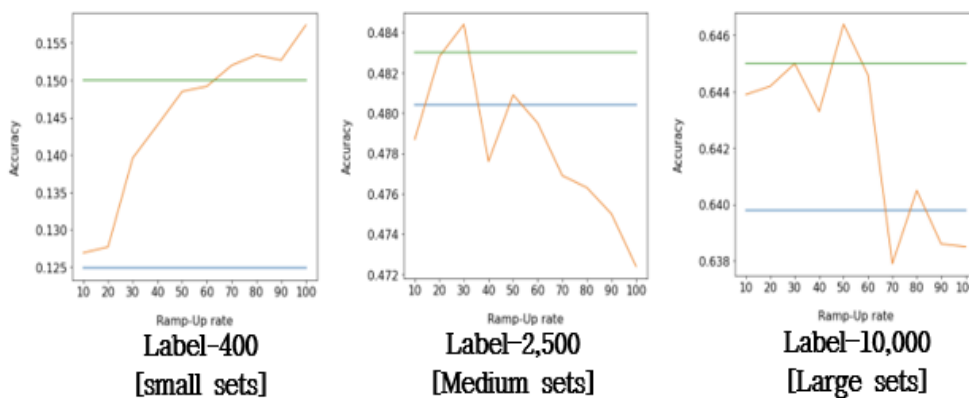


Figure 4-1: MixMatch CIFAR-100 results by number of labels compared with baseline, Ramp-Up, Q-learning method.

## 연구 결과 및 추후 과제 제안

- 연구 초기 가정을 증명하기 위해, 점진증강법을 통해 변하지 않는 하이퍼 파라미터를 비율별로 증가 시키며 해당 변수의 변화가 전체적인 모델의 정확도를 향상시키는 것을 증명하였다. 이후, 모든 비율을 시도하는 것 또한 많은 비용이 발생함으로 이를 강화학습을 통해 훈련 중에 학습하며 최적화를 진행시켜 최종적으로 더 좋은 성능을 이끌어냈다. 추후, 여러 강화학습 알고리즘을 적용해보고, 좀 더 복잡하고 사이즈가 큰 ImageNet 데이터셋 등에 적용해 보는 것을 통하여 실험 결과의 신뢰도를 높일 필요가 있다.

## 관련 홈페이지

Github : <https://github.com/Hanbi-Kim/Flexmatch-SSL>