특집논문 (Special Paper)
방송공학회논문지 제26권 제1호, 2021년 1월 (JBE Vol. 26, No. 1, January 2021)
https://doi.org/10.5909/JBE.2021.26.1.79
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

딥러닝 기반 분류 모델의 준 지도 학습 기법 분석

박 재 현[a)], 조 성 인[a)‡]

The Analysis of Semi-supervised Learning Technique of Deep Learning-based Classification Model

Jae Hyeon Park[a)] and Sung In Cho[a)‡]

요 약

본 논문에서는 소량의 레이블 데이터로 딥러닝 기반 분류 모델을 훈련할 때 적용되는 준 지도 학습 기법 (semi-supervised learning: SSL)에 대해서 분석한다. 기존의 준 지도 학습 기법은 크게 일관성 정규화 (consistency regularization), 엔트로피 기반 (entropybased), 의사 레이블링 (pseudo labeling)으로 구분할 수 있다. 우선, 각 준 지도 학습 기법의 알고리즘에 대해서 서술한다. 실험에서는 준 지도 학습 기법을 레이블 데이터의 수를 변화시키면서 훈련 후 분류 정확도를 평가한다. 최종적으로 실험 결과를 바탕으로 기존 준 지도 학습 기법의 한계에 대해서 서술하고, 분류 성능을 향상하기 위한 연구 방향을 제시한다.

Abstract

In this paper, we analysis the semi-supervised learning (SSL), which is adopted in order to train a deep learning-based classification model using the small number of labeled data. The conventional SSL techniques can be categorized into consistency regularization, entropy-based, and pseudo labeling. First, we describe the algorithm of each SSL technique. In the experimental results, we evaluate the classification accuracy of each SSL technique varying the number of labeled data. Finally, based on the experimental results, we describe the limitations of SSL technique, and suggest the research direction to improve the classification performance of SSL.

Keyword : convolutional neural network, image classification, semi-supervised learning

a) 동국대학교 멀티미디어공학과(Department of Multimedia Engineering, Dongguk University)
‡ Corresponding Author : 조성인 (Sung In Cho)
E-mail: csi2267@dongguk.edu
Tel: +82-2-2260-3339
ORCID: https://orcid.org/0000-0003-4251-7131
※ 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2020R1C1C1009662, NRF-2020X1A3A109).
※ This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP; Ministry of Science, ICT & Future Planning) (No. 2020R1C1C1009662, NRF-2020X1A3A109).
· Manuscript received December 3, 2020; Revised January 6, 2021; Accepted January 7, 2021.

Ⅰ. 서 론

딥러닝 기반 분류 모델은 K개의 클래스에 대응되는 입력 영상을 분류하는 것을 목적으로 다수의 컨볼루션 (convolution) 층과 완전 연결 층 (fully connected layer)을 쌓고 최상의 분류 예측 결과를 얻기 위해 훈련 파라미터를 최적화한다. 이러한 딥러닝 기반 분류는 다양한 학습 기법을 통해 전통적인 레이블이 있는 데이터 세트에서 높은 분류 정확도를 입증해왔다[1-4]. 하지만, 이러한 분류 모델은 훈련에 사용되지 않은 레이블이 없는 데이터에 대해서는 실제 애플리케이션에 적용할 만큼 충분히 만족스러운 분류 성능을 제공하지 않는다. 실제 산업에서는 방대한 수의 클래스 혹은 매우 유사한 클래스들에 대한 분류 성능을 요구하는 등 분류 난이도가 월등히 높기 때문이다. 또한, 사람에 의해 수집되는 레이블이 있는 데이터 세트는 상대적으로 레이블이 없는 데이터 세트에 비해 월등히 적은 수를 이용할 수밖에 없다. 수집되는 방대한 양의 각 데이터에 레이블을 부여하는 주석 (annotation)이 인프라를 구축하기 위한 재정적 비용과 더불어 많은 시간적, 인적 비용이 요구되기 때문이다. 초기의 딥러닝 기반 분류는 이러한 레이블이 없는 데이터의 유무를 고려하지 않은 조건에서 분류 정확도를 평가하는 것이 주를 이루었다[1-4]. 레이블이 없는 데이터를 이용하여 분류 모델을 훈련하기 이전에 레이블이 있는 데이터에 대한 분류 모델의 안정적인 성능이 우선적으로 확보되어야 하기 때문이다. 하지만, 소량의 레이블 훈련 데이터로는 과적합 (overfitting)과 같은 문제가 매우 높은 확률로 발생하였고, 대량의 레이블이 없는 데이터를 훈련에 함께 사용하여 소량의 레이블이 있는 데이터를 가지고 분류 모델을 훈련할 때보다 분류 성능을 향상시키기 위한 연구가 진행되었다. 가장 대표적인 기법이 레이블이 없는 데이터를 소량의 레이블이 있는 데이터와 함께 사용하여 모델의 분류 정확도를 향상시키는 준 지도 학습 기법 (semi-supervised learning)이다[5-11]. 그림 1에서 지도 학습 기법과 준 지도 학습 기법의 상이한 조건에 대한 대략적인 예시를 보여준다. 지도 학습에서는 가용 레이블 데이터를 전부 사용하여 분류 모델을 훈련하지만, 준 지도 학습에서는 제한된 레이블이 있는 데이터와 나머지 레이블이 없는 데이터를 훈련에 함께 사용하여 입력 영상을 분류한다. 준 지도 학습 기법에서는 레이블이 없는 데이터를 효율적으로 분류 모델의 훈련에 사용하여 레이블이 있는 데이터를 모두 이용하여 훈련한 (상대적으로 더 많이 활용한) 모델과 동일한 분류 성능을 이끌어 내는 것이 지향하는 연구 방향이다.

준 지도 학습 기법은 크게 일관성 정규화 (consistency regularization)[7-9], 엔트로피 기반 (entropy-based)[10], 의사 레이블링 (pseudo labeling)[11] 기반 방법으로 분류할 수 있다. 일관성 정규화 기법은 고차원의 데이터를 축소하여 저차원에서 효과적으로 표현할 수 있는 매니폴드의 존재성에 대한 매니폴드 가설 (manifold hypothesis)을 기반으로 한다. 다시 말해, 동일한 클래스의 레이블이 있는 데이터와 레이블이 없는 데이터는 동일한 매니폴드 군집에 위치한다는 가정을 기반으로 분류 모델을 학습한다. 이는 동일한 클래스의 데이터는 매니폴드에서 같은 군집에 속해야 한다는 사전

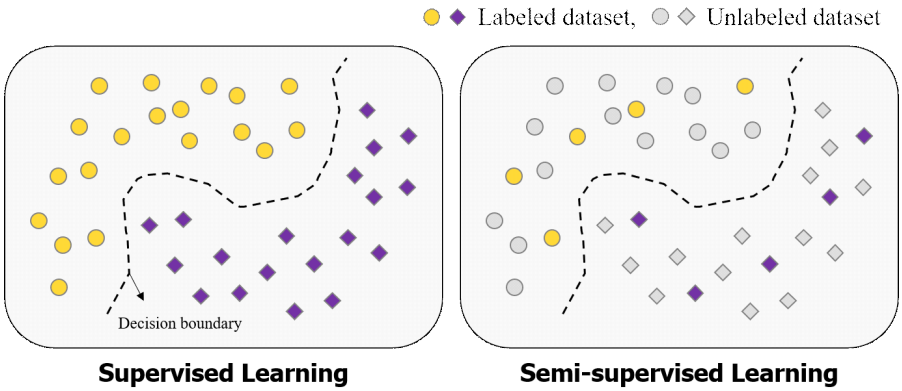


그림 1. 두 가지 클래스 분류 문제에서의 지도 학습과 준 지도 학습의 비교 (decision boundary)
Fig. 1. The comparison of supervised learning and semi-supervised learning on two classes

지식 (prior knowledge)이 적용된 것이다. 구체적으로, 동일 입력에 대해 드롭아웃 (dropout), 노이즈 등 출력에 영향을 주는 요소인 섭동 (perturbation)이 포함되어도 동일한 클래스로 분류 모델이 예측할 수 있도록 레이블이 없는 데이터를 이용하는 방법이다. 엔트로피 기반 준 지도 학습 기법[10]은 분류 모델이 자체적으로 높은 신뢰도 (high confidence)를 가지는 예측 확률을 제공하는 것을 목적으로 훈련하는 기법이다. 즉, 레이블이 없는 데이터에 대해 예측 확률의 엔트로피를 최소화함으로써 분류 성능을 증가시키는 기법이다. 의사 레이블링[11]은 분류 모델이 출력에 대한 높은 신뢰도로 예측한 결과를 정답 레이블로 여기고, 이에 대한 엔트로피 손실을 손실 함수에 추가하여 훈련하는 기법이다. 각 준 지도 학습 기법에 대해 이해하기 위해서 다음 장에서는 각 기법의 개념과 훈련 방법에 대해서 구체적으로 설명하도록 한다.

Ⅱ. 준 지도 학습 기법

본 장에서는 기존에 제안된 다양한 준 지도 학습 기법의 동작 원리에 대해서 서술한다. 우선 가장 성능이 우수한 것으로 알려진 일관성 정규화 기법에 대해서 설명한 후 엔트로피 최소화 및 의사 레이블링에 대해서 추가적으로 설명한다.

1. 일관성 정규화 (consistency regularization)

일관성 정규화 기법은 매니폴드 가설에 입각하여 레이블이 없는 데이터에 대한 분류 모델의 예측 정보를 활용하여 매니폴드에서 서로 다른 클래스로 분류하는 결정 경계 (decision boundary)를 찾는 것을 목적으로 한다[12]. 이 때, 분류 모델을 훈련하기 위한 손실함수는 크게 두 가지로 구성되는데, 첫번째는 레이블이 있는 데이터를 이용한 교차 엔트로피 (cross entropy)이다. 즉 일반적인 분류에서 발생하는 손실을 반영하는 것이다. 두 번째는, 레이블이 없는 데이터에 랜덤하게 섭동을 주어 출력 데이터를 도출하고, 도출된 출력들의 분류 결과가 서로 다른 정도를 손실 값으로 활용한다. 예를 들어, 분류 모델의 초기화 (initialization), 입력의 변형 (transformation), 노이즈, 드롭아웃 등과 같은 요소들을 다르게 하여 동일 입력에 대해 서로 다른 출력 결과를 도출하도록 함으로써 모델의 분류 작업을 방해한다. 이 후, 분류 모델이 서로 다른 출력에 대한 예측 일관성을 유지하도록 함으로써 준 지도 학습 기법이 동작하도록 기여한다. 이러한 접근법을 일관성 정규화라고 한다. 일반적으로 일관성 정규화 기법에서 분류 모델을 훈련하기 위한 손실 함수의 형태는 다음과 같다.

$$\begin{aligned} Loss &= L_{CE}(y,p)+L_{SSL}(p,p') \\ L_{CE}(y,p) &= \sum_{k=1}^{K} -y_k \log(p_k), \quad k=1,2,...,K \end{aligned} \tag{1}$$

수식 (1)에서 K는 클래스의 수를 나타낸다. $L_{CE}(y,p)$는 정답 레이블인 y와 분류 모델의 예측 결과인 p에 대한 교차 엔트로피 손실 함수이다. p'은 입력에 섭동을 가하여 얻은 또 다른 출력을 의미한다. L_{SSL}은 레이블이 없는 데이터를 위한 정규화 (regularization) 항으로써 추가된다. 일반적으로 준 지도 학습 기법은 식 (1)의 형태의 손실 함수를 가지며, 레이블이 있는 데이터 뿐만 아니라 레이블이 없는 데이터도 훈련에 사용하여 분류 모델의 성능을 향상시키는 것을 목적으로 한다.

1.1 Π-model, Temporal Ensembling

[7]에서는 일관성 정규화 기법인 Π-model과 temporal ensembling(TE) 두 가지 학습 모델을 제안했다. Π-model[7]은 분류 모델의 확률적 증강 (stochastic augmentation)과 드롭아웃을 통해 서로 다른 두 출력의 예측 일관성을 오차 제곱의 평균 (mean squared error: MSE)을 감소시키며 유지하는 기법이다. 훈련을 위한 손실 함수는 다음과 같다.

$$\begin{aligned} L_{SSL} &= w(t)\frac{1}{|B|}\sum_{i\in B} \| z_i - \tilde{z}_i \|^2 \\ w(t) &= \exp\{-5(1-T)^2\} \\ T &= \min(t/W,1) \end{aligned} \tag{2}$$

수식 (2)에서 i는 레이블이 없는 데이터 샘플의 인덱스이고, B는 입력 데이터의 배치 (batch) 크기다. z_i는 분류 모델의 예측 벡터를 나타낸다. W는 $w(t)$를 조절하기 위한 상수 값이다. 훈련 스텝 (step) t가 W를 초과하는 순간부터 손실 함수의 계수는 1이 된다. $w(t)$는 지수적 램프-업 (expo-

nential ramp-up) 함수로 훈련 스텝 t가 증가함에 따라 정규화 항의 가중치를 증가시킨다. 훈련 초기 레이블이 없는 데이터에 대한 분류 모델의 예측 확률이 부정확하므로 레이블이 있는 데이터에 대한 분류 모델의 훈련이 먼저 적용되도록 레이블이 없는 데이터에 대한 훈련 손실 값을 서서히 증가시키기 위한 구간 [0,1)로 정의되는 함수이다.

TE[7]는 Π-model과 대부분 동일하나 이전 모델 출력과 현재 모델 출력의 앙상블 (ensemble)을 통해서 예측 벡터를 도출한다. 분류 모델의 이전 출력 결과를 취득함으로써 섭동의 변화에도 좀 더 안정적으로 클래스를 분류하도록 추가적으로 제안된 기법이다. 훈련을 위한 손실 함수는 다음과 같다.

$$\begin{aligned} L_{SSL} &= w(t)\frac{1}{|B|}\sum_{i\in B}\| z_i - \tilde{z}_i \|^2 \\ Z &= \alpha Z + (1-\alpha)z \\ \tilde{z}_i &= Z/(1-\alpha^t) \end{aligned} \tag{3}$$

수식 (3)에서 Z는 분류 모델의 이전 출력을 나타낸다. α는 (0,1)의 범위의 상수로 현재 분류 모델의 출력과 이전 출력을 컨벡스 (convex) 결합하기 위한 계수이다. 훈련 스텝 t가 증가함에 따라 신경망의 이전 출력인 Z의 섭동 출력에 대한 영향력이 감쇠하게 된다.

1.2. Mean Teacher (MT)

TE에서는 지수적 감쇠된 분류 모델의 이전 출력을 누적하여 출력 벡터를 생성했다면, MT[8]에서는 학생 (student) 모델의 훈련 파라미터를 지수 이동 평균 (exponential moving average: EMA)하여 누적하는 교사 (teacher) 모델로 출력 벡터를 생성한다. TE 모델에서는 이전에 훈련된 분류 모델의 결과를 고려한 앙상블 모델이 교사 모델로 학생 모델에 지식 증류 (knowledge distillation)가 되는 것과 같고, MT에서는 이전 분류 모델의 훈련 파라미터가 전달되는 교사 모델로부터 학생 모델에 지식 증류가 이루어지는 것과 같다. 손실 함수는 다음과 같이 정의된다.

$$\begin{aligned} L_{SSL} &= w(t)\frac{1}{|B|}\sum_{i\in B}\|f(x_i,\theta',\eta') - f(x_i,\theta,\eta)\|^2 \\ \theta' &= \alpha\theta_{t-1} + (1-\alpha)\theta_t \end{aligned} \tag{4}$$

수식 (4)에서 θ와 η는 분류 모델의 파라미터와 노이즈를 나타내며, $f(x_i,\theta,\eta)$와 $f(x_i,\theta',\eta')$은 서로 다른 섭동이 적용된 분류 모델의 출력을 나타낸다. $f(x_i,\theta',\eta')$은 교사 모델로 EMA를 통해 이전 훈련에서 도출된 훈련 파라미터를 전달받기 때문에 섭동이 주어져도 학생 모델 $f(x_i,\theta,\eta)$보다 좀 더 안정적인 분류 결과를 도출할 수 있고, 이러한 교사 모델로부터 학생 모델에 지식 증류가 이루어지는 학습 기법이다. 다시 말해, 섭동이 적용된 분류 문제를 보다 안정적으로 풀도록 학습된 교사 모델과 학생 모델의 출력의 차이를 감소시켜 레이블이 없는 데이터에 대한 학생 모델의 분류 정확도를 향상시키는 기법이다.

1.3. Virtual Adversarial Training (VAT)

VAT[9]는 Π-model과 MT와 같은 모델 내부의 확률적 (stochasticity) 기반으로 섭동 출력을 생성하는 것이 아닌 입력에 노이즈를 생성하여 분류 모델의 출력 차이를 만들어낸다. 이러한 노이즈는 모델의 손실 함수에 가장 적대적 (adversarial)인 노이즈로 생성되며, 분류 모델의 큰 출력 차이를 생성하게 된다. 손실 함수는 다음과 같이 정의된다.

$$\begin{aligned} L_{SSL} &= w(t)\frac{1}{|B|}\sum_{i\in B} D[f(x_i,\theta), f(x_i + r_{adv},\theta')] \\ r &= \xi \times rand(h,w) \\ g &= \nabla_r D[f(x_i,\theta), f(x_i + r_{adv},\theta')] \\ r_{adv} &= \epsilon\frac{g}{\| g \|_2} \end{aligned} \tag{5}$$

수식 (5)에서 ξ는 입력 영상의 크기와 동일하게 생성되는 노이즈의 강도를 조절하는 상수이고, $D[\bullet]$는 분류 모델의 서로 다른 입력에 대한 출력의 거리 함수를 나타낸다. VAT에서는 KL 확산 (Kullback-Leibler divergence)을 적용한다. KL 확산은 두 입력 확률 분포의 유사도 (similarity)를 도출한다. 그 다음, 노이즈 성분 r에 대한 도출된 유사도의 변화도 (gradient)가 도출된다. 또한, 정규화 (normalize)된 변화도와 상수 파라미터 ϵ을 통해서 적대적 노이즈 (adversarial noise) 성분인 r_{adv}를 생성한다. 최종적으로, VAT에서는 도출된 적대적 노이즈 성분이 추가된 $f(x_i + r_{adv},\theta')$와 $f(x_i,\theta)$의 KL-divergence를 준 지도 학습

의 손실 함수로써 최소화한다. 따라서, 서로 다른 두 출력의 유사도를 감소시킴으로써, 신경망은 레이블이 없는 데이터에 대한 분류를 가장 방해하는 노이즈 성분인 r_{adv}를 발생시키게 된다.

2. Entropy Minimization (EM)

EM[10]은 레이블이 없는 데이터에 대한 분류 모델의 예측 확률의 엔트로피를 감소하는 방향으로 훈련하는 준 지도 학습 기법이다. 훈련을 위한 손실 함수는 다음과 같다.

$$L_{SSL} = w(t)\frac{1}{|B|}\sum_{i\in B} E(z_i)$$
$$E(p) = \sum_{k=1}^{K} -p_k \log(p_k), \quad k = 1,2,...,K \tag{6}$$

수식 (6)에서 E는 예측 확률 벡터 z의 엔트로피를 나타내며, 엔트로피는 특정 클래스에 대한 확률이 높을 경우 낮아지고, 모든 클래스에 대한 예측 확률이 모두 동일할 때 가장 높은 값을 나타낸다. 따라서, EM은 레이블이 없는 데이터에 대한 분류 모델의 예측 확률 벡터 z가 어느 한 클래스에 대해서 높은 확률을 나타내도록 훈련하는 기법이다.

3. Pseudo Labeling (PL)

일반적으로 의사 레이블링은 레이블이 없는 데이터에 대한 분류 모델의 출력 중 예측에 대한 신뢰도 (confidence)가 높은 샘플을 정답 레이블로써 가정하고 새로운 훈련 데이터로 추가하는 기법이다. 하지만, 이러한 기법에서는 오답 레이블로 예측된 결과에 대해서 분류 모델이 높은 신뢰도를 나타낼 경우 분류 모델의 극심한 성능 저하를 나타낼 수 있다. 또한, 높은 신뢰도를 보이는 샘플에 대해서만 치중하기 때문에 분류 모델의 과적합 문제가 발생할 수 있다. 이와는 다르게 PL[11]은 분류 모델의 임계치 (threshold value) 이상 높은 신뢰도를 나타내는 레이블이 없는 데이터 샘플에 대해 출력의 엔트로피를 도출하고, 이를 전체 훈련 손실 함수에 추가하여 최소화하는 훈련 기법이다.

Ⅲ. 실험 결과

실험에 사용된 데이터 세트는 CIFAR-10[13]과 SVHN[14]이다. CIFAR-10은 총 10개의 클래스를 가진 분류 데이터

(a) (b)

그림 2. 실험에 사용된 데이터 세트 예시 영상 (a) CIFAR-10, (b) SVHN
Fig. 2. The example images of dataset used for experiment (a) CIFAR-10, (b) SVHN

세트로 훈련 영상 50,000장, 테스트 영상 10,000장으로 32×32 크기의 컬러 영상으로 구성되어 있다. 실험에서는 훈련 영상 50,000장 중 45,000장은 훈련에 사용하고, 5,000장은 평가 (validation) 영상으로 사용했다. SVHN은 1부터 10까지 숫자 영상으로 구성된 32×32 크기의 컬러 영상 데이터 세트로 총 훈련 영상 73,257 장, 테스트 영상 26,032장으로 구성되어 있다. 총 훈련 영상 중 65,931장은 훈련에 사용하고, 나머지 7,326은 평가에 사용했다. 두 데이터 세트에서 레이블이 있는 데이터는 훈련 영상에서 클래스 별로 동일한 분포로 추출된 1,000장과 5,000장에 대해서 실험하고, 추출된 영상을 제외한 나머지 훈련 영상은 전부 레이블이 없는 데이터로 사용했다. 추론 (inference)에는 테스트 영상 전부를 사용했다. 그림 2는 실험에 사용된 데이터 세트의 샘플 영상을 나타낸다. CIFAR-10은 동물, 선박, 기계 등의 다양한 클래스로 구성되어 있고, SVHN은 숫자를 나타내는 표지판 형식의 영상으로 구성되어 있다.

[12]에서는 준 지도 학습 기법은 동일한 백본 신경망 및 조건에서 평가할 것을 언급했다. 이는 분류 모델의 복잡도 (complexity)가 다른 백본 신경망에서 준 지도 학습 기법을 비교할 경우 정확한 분류 성능 비교를 하기 어렵기 때문이다. 이에 따라서 [12]에서 제안한 백본 신경망인 Wide-ResNet-28-2[15]를 사용하고, 하이퍼 파라미터도 동일하게 적용하여 실험을 진행했다. Wide-ResNet은 딥러닝 기반의 분류 모델로 신경망 깊이 (depth)가 깊어짐에 따라 학습의 난이도가 어려워지자 이를 잔차 블록 (residual block)[16]의 채널을 늘리면서 분류 성능을 향상시킨 신경망이다. 실험에 적용된 Wide-ResNet-28-2 신경망의 깊이는 28이고, 빌딩 블록 내의 컨볼루션 층의 수는 2이다. 본 실험에는 드롭아웃은 적용되지 않았다. 레이블이 있는 데이터의 수는 1,000개, 5,000개를 사용했을 때를 비교하며, 나머지 훈련 영상은 레이블이 없는 데이터로 사용한다. 게다가 [12]에서는 엔트로피 최소화 기법의 분류 정확도를 다른 준 지도 학습 기법들과 비교하지 않았으나, 본 논문에서는 다양한 준 지도 학습 기법의 분류 성능을 확인하기 위해서 실험 결과를 추가했다. 표 1에서 All은 분류 모델을 훈련할 때 훈련 영상 전부를 사용한 것이고, labeled only는 레이블이 있는 데이터만 훈련에 사용한 것이다. 백본 신경망과 하이퍼 파라미터가 기존 논문과 다르기 때문에 실제 각 논문에서 제공된 분류 성능과 다소 차이가 있을 수 있다. 각 준 지도 학습 기법의 실험에 적용된 하이퍼 파라미터는 표 1에서 확인할 수 있다.

표 2에서 확인할 수 있듯이, 적대적 잡음을 생성함으로써

표 1. 하이퍼 파라미터
Table 1. Hyperparameters

Method	Key	Value	
		WRNet	DNN
Common	Learning rate decay	0.2	0.2
	Iteration	500,000	50,000
	Batch	100	100
	W	200,000	20,000
All, Labeled only	Initial learning rate	0.003	0.001
	Batch	200	200
Π-model[7]	Initial learning rate	0.0003	0.001
	Max coefficient of ramp-up	20	1
MT[8]	Initial learning rate	0.0004	0.001
	Max coefficient of ramp-up	8	1
	EMA decay	0.95	0.99
VAT[9]	Initial learning rate	0.003	0.001
	Max coefficient of ramp-up	0.3	1
	Epsilon	6	{8,3}
	Xi	10^{-6}	10^{-6}
	Coefficient of entropy	0.06	1
EM[10]	Initial learning rate	0.003	0.001
	Max coefficient of ramp-up	1	1
PL[11]	Initial learning rate	0.003	0.001
	Max coefficient of ramp-up	1	1
	Confidence threshold	0.95	0.95

*WRNet: Wide-ResNet-28-2

표 2. Wide-ResNet-28-2에서 비교 방법 분류 정확도
Table 2. The classification accuracy of benchmark method (Backbone network: Wide-ResNet-28-2)

Method	CIFAR-10		SVHN	
# Labels	1000	5000	1000	5000
All	0.9252		0.9678	
Labeled only	0.6251	0.8161	0.8729	0.9291
Π-model[7]	0.6407	0.8544	0.9172	0.9477
MT[8]	0.6672	0.8505	0.9371	0.9447
VAT[9]	0.7552	**0.8735**	**0.9411**	0.9519
VAT + EM[9]	**0.7700**	0.8688	0.9394	**0.9542**
EM[10]	0.6453	0.8008	0.8765	0.9372
PL[11]	0.7114	0.8570	0.9287	0.9500

분류 모델의 예측을 크게 변화시킨 VAT의 성능이 전반적으로 가장 우수한 것을 확인할 수 있다. 따라서, 일관성 정규화 기법에서는 효과적인 섭동을 생성하는 것이 가장 주요하게 작용하게 된다. 실험 결과에서 알 수 있듯이 EM의 경우 레이블이 있는 데이터만 사용했을 때보다 낮은 분류 정확도를 보여준다. 이는 EM에서는 출력에 대한 신경망의 예측 신뢰도를 높이는 방향으로 훈련하기 때문에 과적합 문제를 초래할 수 있다고 [12]에서도 밝힌 바 있다. 따라서, 엔트로피 기반 기법은 이러한 과적합 문제를 해결할 수 있는 방안이 필수적이다. 전반적으로 준 지도 학습 기법은 레이블이 있는 데이터만 사용했을 때 보다 눈에 띄는 성능 향상을 보여주는 기법이 있는 반면, 그렇지 않은 기법들도 있다. 추가적으로, 기존의 준 지도 학습 기법[7-11]은 13개의 층으로 구성된 깊은 신경망 (deep neural network: DNN)[7]을 백본 신경망으로써 적용하여 CFIAR-10 및 SVHN 데이터 세트에 대한 분류 성능을 비교한다. 표 3은 13개의 층으로 이루어진 신경망인 DNN 모델의 구조를 보여준다. 실험에 적용된 하이퍼 파라미터는 표 1과 같다. VAT 기법에서 각 데이터 세트 CIFAR-10과 SVHN에 적용되는 적대적 노이즈의 강도를 조절하는 상수 파라미터 ϵ는 각각 8, 3으로 설정했다. 각 준 지도 학습 기법[7-11]에서 사용한 하이퍼 파라미터를 동일하게 적용했으나, 훈련의 총 반복 횟수, 배치 크기 및 러닝 레이트가 다소 다르므로 논문에서 제공하는

표 3. DNN 모델 구조[7]
Table 3. The architecture of DNN

Name	Description	Output size
Input	32×32 RGB images	-
Transform	Gaussian noise σ = 0.15, Random flip, Random crop	32×32
Conv__1	128 filters, 3×3, pad='same', BN (128), LReLU (α = 0.1)	
	128 filters, 3×3, pad='same', BN (128), LReLU (α = 0.1)	
	128 filters, 3×3, pad='same', BN (128), LReLU (α = 0.1)	
pool	Max pooling (2×2)	16×16
drop	Dropout (p=0.5)	
Conv__2	256 filters, 3×3, pad='same', BN (256), LReLU (α = 0.1)	
	256 filters, 3×3, pad='same', BN (256), LReLU (α = 0.1)	
	256 filters, 3×3, pad='same', BN (256), LReLU (α = 0.1)	
pool	Max pooling (2×2)	8×8
drop	Dropout (p=0.5)	
Conv__3	512 filters, 3×3, pad='valid', BN (512), LReLU (α = 0.1)	6×6
	256 filters, 1×1, BN (256), LReLU (α = 0.1)	
	128 filters, 1×1, BN (128), LReLU (α = 0.1)	
pool	Global average pooling (6×6 → 1×1)	1×1
Output	Fully connected (128 → 10), softmax	1×10

*Conv: convolution layer, BN: batch normalization[17], LReLU: leaky ReLU[18]

표 4. DNN에서 비교 방법 분류 정확도
Table 4. The classification accuracy of benchmark method (Backbone network: DNN)

Dataset	# Labels	Labeled only	Π-model[7]	MT[8]	VAT[9]	EM[10]	PL[11]
CIFAR-10	1000	0.5978	0.6088	0.6256	**0.7209**	0.6067	0.6157
SVHN	1000	0.8576	0.8786	0.8836	0.8476	0.8867	**0.9210**

실험 결과와 차이가 있을 수 있다. 표 4에서 볼 수 있듯이, CIFAR-10 데이터 세트에서는 VAT 기법이 가장 우수한 분류 정확도를 제공하고, SVHN 데이터 세트 에서는 PL 기법이 가장 우수한 분류 정확도를 제공한다. 또한, 준 지도 학습 기법의 분류 정확도는 이전 Wide-ResNet-28-2를 백본 신경망으로 실험했을 때와 전반적으로 동일한 양상을 보여주지만 분류 정확도는 감소된 것을 확인할 수 있다.

Ⅳ. 결 론

본 논문에서는 제한된 수의 레이블이 있는 데이터와 많은 수의 레이블이 없는 데이터를 훈련에 사용하는 준 지도 학습 기법에 대해서 서술하고, 대표적인 기법에 대해서 전통적인 데이터 세트로 훈련하여 분류 정확도를 비교해보았다. 준 지도 학습 기법에서는 레이블이 없는 데이터에 대해 서로 다른 섭동을 적용하여 분류 모델의 출력에 차이를 발생시키고, 이러한 차이를 최소화하면 분류 정확도를 향상시킬 수 있다는 사실을 활용하여 분류 정확도 향상을 유도하였다. 준 지도 학습 기법의 효과는 실험적으로 분류 모델의 성능 향상에 유의미하다는 것을 확인할 수 있었다. 결과적으로, 준 지도 학습 기법은 레이블이 있는 데이터와 레이블이 없는 데이터의 출력의 차이를 감소시키는 방향으로 훈련하므로, 레이블이 있는 데이터에 대한 분류 모델의 경향성이 가장 큰 비중을 차지하게 된다. 따라서, 레이블이 있는 데이터에 대한 다양한 변형 및 증대 기법을 적용하여 가용 정보를 늘려주는 것이 효과적일 것으로 보인다. 또한, 백본 신경망에 따라서 준 지도 학습 기법의 분류 정확도는 달라질 수 있으며, 데이터 세트에 따라서 분류 정확도의 변화를 보였으며, 분류 모델의 성능이 하이퍼 파라미터에 따라 민감하게 변화하는 것을 확인할 수 있었다. 따라서, 준 지도 학습 기법은 하이퍼 파라미터의 최적화 작업이 필수적으로 이루어져야 한다.

실험에 사용된 CIFAR-10과 SVHN과 같은 데이터 세트는 실제 산업에서 분류하고자 하는 데이터 세트의 크기에 비해서 상대적으로 적다고 볼 수 있다. 따라서, 준 지도 학습 기법의 단계적 향상을 위해서 광범위한 클래스 데이터를 가지고 있는 데이터 세트에 대하여 적용하는 것이 필수적인 것으로 보인다. 하지만, 대량의 데이터가 포함된 데이터 세트의 경우 훈련에 방해가 되는 잡음 (noise) 데이터와 오분류 클래스 데이터를 포함할 수 있기 때문에 이러한 데이터를 검출 및 제거한다면 준 지도 학습 기법의 분류 성능은 좀 더 우수한 결과를 도출할 수 있을 것으로 보인다. 또한, 데이터 세트의 클래스 수가 증가하거나 해상도가 큰 경우에는 분류를 위해 요구되는 파라미터가 증가하기 때문에 분류 작업의 복잡도 (complexity)에 따라서 적절한 백본 신경망을 사용하는 것이 필요하다.

참 고 문 헌 (References)

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Proc. Adv. Neural Inf. Process. Syst., pp. 1097 – 1105, 2012.

[2] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Proc. Int. Conf. Learn. Represent., pp. 1 – 14, 2015.

[3] C. Szegedy et al., “Going deeper with convolutions,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 1 – 9, 2015.

[4] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 770-778, 2016.

[5] P. Bachman, O. Alsharif, and D. Precup, “Learning with pseudo ensembles,” in Proc. Advances Neural Inf. Process. Syst., pp. 3365 – 3373, 2014.

[6] M. Sajjadi, M. Javanmardi, and T. Tasdizen, “Regularization with stochastic transformations and perturbations for deep semi-supervised learning,” in Proc. 30th Int. Conf. Neural Inf. Process. Syst., pp. 1171 – 1179, 2016.

[7] S. Laine and T. Aila, “Temporal ensembling for semi-supervised learning,” in Proc. Int. Conf. Learn. Represent., pp. 1-13, 2017.

[8] A. Tarvainen and H. Valpola, “Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results,” in Proc. Adv. Neural Inf. Process. Syst., pp. 1195 – 1204, 2017.

[9] T. Miyato, S.-I. Maeda, M. Koyama, and S. Ishii, ‘‘Virtual adversarial training: A regularization method for supervised and semi-supervised learning,’’ IEEE Trans. Pattern Anal. Mach. Intell., Vol. 41, No. 8, pp. 1979 – 1993, Aug 2019.

[10] Y. Grandvalet and Y. Bengio, “Semi-supervised learning by entropy minimization,” in Proc. Adv. Neural Inf. Process. Syst., pp. 529 – 536, 2004.

[11] D.-H. Lee, “Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks,” in Proc. Workshop Challenges Represent. Learn. (ICML), pp. 2-7, 2013.

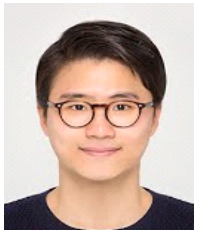
[12] A. Oliver, A. Odena, C. Raffel, E. Cubuk, and I. Goodfellow, “Realistic Evaluation of Deep Semi-Supervised Learning Algorithms,” in Adv. in

Neural Inf. Process. Syst., pp. 3235-3246, 2018.
[13] A. Krizhevsky and G. Hinton, “Learning Multiple Layers of Features from Tiny Images,” technical report, Univ. of Toronto, 2009.
[14] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, “Reading digits in natural images with unsupervised feature learning,” In NIPS Workshop on Deep Learning and Unsupervised Feature Learning, 2011.
[15] S. Zagoruyko and N. Komodakis, “Wide residual networks,” in Proc. Brit. Mach. Vis. Conf., pp. 87.1 – 87.12, 2016.
[16] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition.” In Proc. IEEE Conf. Comput. Vis. Pattern Rcognit., pp. 770-778, 2016.
[17] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in Proc. Int. Conf. Mach. Learn., pp. 448 – 456, 2015.
[18] A. L. Maas, A. Y. Hannun, and A. Y. Ng, “Rectifier nonlinearities improve neural network acoustic models,” in Proc. ICML, Vol. 30, No. 1, p. 3, Jun. 2013.

저 자 소 개

박 재 현

- 2019년 2월 : 대구대학교 전자전기공학부 공학사
- 2019년 9월 ~ 현재 : 동국대학교 멀티미디어공학과 공학석사
- ORCID : https://orcid.org/0000-0002-6233-4394
- 주관심분야 : 영상 분석 및 화질 개선, 톤매핑 알고리즘, 딥러닝 분류 시스템



조 성 인

- 2010년 : 서강대학교 전자공학부 공학사
- 2015년 : 포항공과대학교 전기컴퓨터공학 공학박사
- 2015년 ~ 2017년 : LG디스플레이 선임연구원
- 2017년 ~ 2019년 : 대구대학교 전자전기공학부 조교수
- 2019년 ~ 현재 : 동국대학교 멀티미디어공학과 조교수
- ORCID : https://orcid.org/0000-0003-4251-7131
- 주관심분야 : 영상 화질 개선, 비디오 처리, 멀티미디어 신호 처리, LCD 및 OLED 시스템 회로 설계