라벨링오류에 강인한 CNN 모델 학습기법

황태환, 백재순, 이준호, 박건율, 최준원
한양대학교

{thhwang, jsbaik, jhlee, konyulpark}@spa.hanyang.ac.kr, junwchoi@hanyang.ac.kr

Robust CNN model training scheme for labeling error

Hwang Tae Hwan, Baik Jae Soon, Lee Jun Ho, Park Kon Yul, Choi Jun Won
Hanyang Univ.

요 약

본 논문은 노이즈 레이블이 섞인 데이터세트에 대해 오직 한개의 네트워크만 사용하여 강인한 학습을 하는 기법을 제안한다. 제안된 방법은 학습단계에서 알맞게 라벨링된 샘플을 알아내기 위해 두 종류의 샘플 선택 기준을 사용한다. 첫 번째 기준은 이전 시간들의 모델 앙상블의 손실함수값이다. 시간적으로 분리된 모델 앙상블을 만들기 위해 주기적 학습률을 적용한다. 이전 시간적 앙상블에 대한 손실함수값은 깨끗한 샘플을 감지하는 기준을 제공한다. 두 번째 기준은, 데이터 변형을 적용하여 다시점으로 예측을 하게하여, 이 예측들의 일치가 학습과 샘플 선택을 하는데 둘다 사용되게 한다. 이러한 기준들을 같이 사용함으로 써 알맞게 라벨링된 샘플을 정확하게 감지할 수 있고, 여러 개의 네트워크를 사용한 방법보다 더 나은 성능을 도달할 수 있다. 본 연구는 CIFAR-10, CIFAR-100, MNIST 등 일반적으로 널리 사용되는 데이터세트를 이용하여 수행되었으며, 기존 방법들에 비해 높은 성능을 도출하는 것을 확인하였다.

Ⅰ. 서 론

본논문에서는 간단하지만 효과적으로 정확히 라벨링된 샘플을 확인할 자가 학습 방법을 제안한다. 제안된 방법은 학습된 하나의 네트워크로 얻어진 모델 앙상블을 이용한다. 먼저, 워밍업 단계 후에 주기적으로 학습률을 키웠다가 천천히 줄이는 주기적 학습률을 적용하여 각 주기마다 모델 앙상블을 생성한다. 학습률이 치솟을 때마다, 현재 주기에서 얻어진 앙상블 네트워크는 그 전 주기에서 얻어진 앙상블 네트워크와 분리된 새로운 네트워크가 된다. 그러므로 이전 주기에서 얻어진 시간적 앙상블에 기반하여 믿을만한 깨끗한 샘플을 감지할 수 있다. 다음으로, 제안된 방법은 데이터 샘플의 변형을 통해 다양한 시각으로부터 한 개 이상의 예측을 생산한다. 이 예측들 간의 일치는 일관성 정규화 손실 항목에 의해 받아들여지고 이 또한 샘플 선택 방법으로써 사용된다. 제안된 방법에 대한 전체 구조의 그림은 그림[1] 에서 확인할 수 있다.

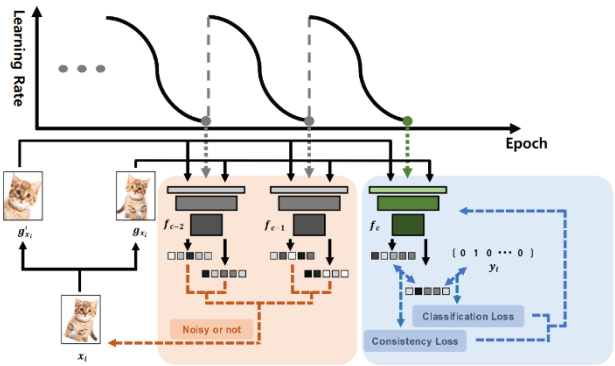

그림 1 제안하는 방법의 전체 구조

Ⅱ. 본론

2.1 문제 설정

본 논문에서는 학습 데이터 세트 $\mathcal{D} = \{x_i, y_i\}_{i=1}^{N}$ 에서 M개 클래스 이미지 분류 문제를 다룬다. x_i는 i번째 학습 데이터, $y_i \in \{0,1\}^M$ 은 i 번째 학습데이터의 라벨을 의미한다. 본 연구에서는 학습 데이터 세트 $\mathcal{D}$의 $\eta\%$가 라벨링 오류를 가지고 있다고 가정한다.

2.2 제안하는 방법

주기적 학습률 스케줄링은 주기적으로 학습률을 증가 및 감소시켜 시간적으로 분리 된 네트워크를 생성한다. 스케줄링은 최대 학습률 r_1 , 최소 학습률 r_2 및 epoch 수의 주기 c로 지정된다. t번째 epoch 의 학습률 $r(t)$는 다음과 같이 주어진다.

$$r(t) = \big(1 - s(t)\big)r_1 + s(t)r_2 \quad (1)$$

$$s(t) = \frac{1 + \big((t-1)\ mod\ c\big)}{c} \quad (2)$$

주기적 학습률이 계속 진행되면 네트워크의 시간적 앙상블 시퀀스 $f_1(x), f_2(x), \dots, f_k(x)$ 가 생성된다. 여기서 k는 주기를 나타낸다.

제안하는 방법은 네트워크가 k 주기의 학습률로 학습되었다고 가정할 때, 노이즈가 있는 라벨의 영향을 피하기 위해 올바르게 라벨이 지정된 깨끗한 샘플로만 모델 학습에 사용하는 것을 목표로 한다. 이를 위해 배치에서 학습에 사용되는 샘플 수는 손실이 가장 작은 샘플의 $(100-\eta)\%$를 유지한다. 현재 시점의 모델만으로 손실 함수를 구할 때 모델의 바이어스로 인해 실수로 노이즈 라벨을 깨끗한 샘플이라고 감지한 경우 다시 이 샘플을 노이즈 샘플로 지정하기 어려울 수 있다. 따라서 학습중인 현재 모델에서 분리 된 앙상블 네트워크를 통해 배치 샘플에서 평가할 필요가 있다.

Noise rate	Co-teaching	Co-teaching+	JoCoR	Ours
Symmetric-20%	95.12±0.23	97.81±0.03	98.06±0.04	**98.51**±0.11
Symmetric-50%	89.97±0.17	95.80±0.09	96.68±0.10	**97.37**±0.06
Symmetric-80%	79.74±0.36	58.92±14.73	84.89±4.55	**87.89**±4.17
Asymmetric-40%	92.97±3.69	93.28±0.43	95.57±0.23	**97.51**±0.11

표 1 MNIST 의 평균 테스트 정확도 결과

Noise rate	Co-teaching	Co-teaching+	JoCoR	Ours
Symmetric-20%	78.36±0.62	78.81±0.26	85.73±0.19	**89.78**±0.23
Symmetric-50%	71.30±0.13	57.12±0.59	79.41±0.25	**85.63**±0.27
Symmetric-80%	27.59±3.76	24.90±2.13	27.78±3.06	**29.95**±3.33
Asymmetric-40%	76.00±1.32	69.88±0.33	76.36±0.49	**84.66**±0.30

표 2 CIFAR-10 의 평균 테스트 정확도 결과

주기적인 학습률을 이용하면 이러한 앙상블 네트워크를 쉽게 얻을 수 있다. 샘플 선택이 k번째 주기에서 수행 될 때 손실 함수는 최근 T 주기에서 얻은 앙상블 네트워크 $f_{k-1}(x), \dots, f_{k-T}(x)$ 를 사용하여 구하게 된다. 특히 각 샘플 x_i 에 대해, 손실 함수의 합계 $\sum_{t-1}^{T}\mathcal{L}(x_i, f_{k-t})$를 구하는데, $\mathcal{L}(x_i, f_{k-t})$는 입력 샘플 x_i와 최근 t 주기의 모델 f_{k-t} 에 대한 손실 함수이다. 손실 함수 계산 후에 값이 큰 상위 $\eta\%$ 의 샘플을 배치에서 제거한다.

일관성 정규화는 입력 또는 모델의 작은 변형에 대한 일관성 있는 예측을 수행하도록 모델을 강제한다. 널리 사용되는 방법은 근소하게 변형된 데이터 또는 모델에서 예측의 차이를 최소화 하는 정규화 손실 항을 추가하는 것이다. 일관성 정규화 손실은 학습 과정에서 깨끗한 샘플에 대해 일관된 예측을 수행하도록 모델을 안내하지만 레이블이 잘못된 샘플에 대해서는 일관된 예측을 하기가 어렵다. 이러한 이유로 일관성 정규화를 사용하면 깨끗한 샘플과 라벨이 잘못 지정된 샘플을 더 잘 구별 할 수 있다. 제안된 방법은 일관성을 측정하기 위해 이미지 뒤집기 및 자르기와 같은 데이터 변환을 사용한다. 다르게 변환된 입력들에 대한 네트워크의 출력들은 다시점 모델 예측으로 간주 할 수 있다. i번째 학습 샘플 (x_i, y_i)와 관련된 전체 손실 항 $\mathcal{L}(x_i, f)$은 교차 엔트로피 손실 항 $\mathcal{L}_{CE}(x_i, y_i, f)$ 과 정규화 손실 항 $\mathcal{L}_{JS}(x_i, f)$로 구성된다.

$$\mathcal{L}(x_i, f) = \lambda\mathcal{L}_{CE}(x_i, y_i, f) + (1-\lambda)\mathcal{L}_{JS}(x_i, f) \quad (3)$$

λ는 두 손실 항의 균형을 맞추는 매개 변수이다. 교차 엔트로피 손실 항은 다음과 같이 주어지다.

$$\mathcal{L}_{CE}(x_i, y_i, f) = -\sum_{m=1}^{M} y_i log f^m(x_i) \quad (4)$$

$f^k(x_i)$ 는 k 번째 클래스 확률을 나타낸다. 정규화 손실은 두개의 변환 함수 $g(\cdot)$ 와 $g'(\cdot)$ 을 샘플 x_i 에 적용시킨 후 다시점 예측 간의 Jensen-Shannon (JS) divergence 에 의해 주어진다.

$$\mathcal{L}_{JS}(x_i, f) = D_{KL}\left(f(g(x_i))\middle|f(g'(x_i))\right) + D_{KL}\left(f(g'(x_i))\middle|f(g(x_i))\right) \quad (5)$$

$D_{KL}\left(f(g(x_i))\middle|f(g'(x_i))\right)$ 는 Kullback-Leibler (KL) divergence 로 다음과 같이 나타낸다.

$$D_{KL}\left(f(g(x_i))\middle|f(g'(x_i))\right) = \sum_{m=1}^{M} f(g(x_i)) log \frac{f(g(x_i))}{f(g'(x_i))} \quad (6)$$

2.3 실험 결과

본 연구에서 3 개의 공개 데이터 세트인 MNIST, CIFAR-10 및 CIFAR-100 을 사용하여 실험 진행하였다.

Noise rate	Co-teaching	Co-teaching+	JoCoR	Ours
Symmetric-20%	44.08±0.54	49.27±0.03	53.21±0.34	**62.42**±0.27
Symmetric-50%	34.96±0.50	40.04±0.70	43.49±0.46	**55.79**±0.23
Symmetric-80%	15.15±0.46	14.40±0.58	15.49±0.98	**26.91**±0.75
Asymmetric-40%	31.75±0.28	34.75±0.86	32.80±0.21	**40.22**±0.45

표 3 CIFAR-100 의 평균 테스트 정확도 결과

노이즈 레이블 생성을 위해, 이전 연구에서 합성 라벨 노이즈를 생성하는 방식[1-3]을 따라 대칭 및 비대칭 레이블 노이즈 모두 생성하였다.

모든 데이터 세트에 대한 깨끗한 테스트 데이터 세트의 테스트 정확도의 평균 및 표준편차는 표[1-3] 에서 확인할 수 있다. 성능을 평가하기 위해 다른 무작위 시드로 실험을 5 회 반복하여 결과를 측정했다. 제안된 방법은 Co-teaching[1], Co-teaching+ [2] 및 JoCoR[3] 등의 기존 방법 과 비교하였다. 모든 데이터 세트와 모든 noise rate 에 대해서 기존 방법들에 비해 성능이 월등히 높은 것을 확일 할 수 있다.

Ⅲ. 결론

본 논문에서는 학습 중에 깨끗한 샘플을 선택하는 새로운 기준을 제시하여 노이즈 라벨에 대한 효율적이고 강력한 학습 방법을 제안하였다. 제안된 방법은 주기적 학습률을 통해 단일 네트워크를 학습시켜 시간적으로 분리된 앙상블을 생성한다. 깨끗한 샘플은 이전 사이클에서 얻은 시간적 앙상블 네트워크를 기반으로 검출되었다. 주기적 학습률은 여러 네트워크를 사용하지 않고도 딥 뉴럴 네트워크가 노이즈 레이블을 기억하여 발생하는 영향을 완화할 수 있음을 보였다. 입력 샘플을 여러 개로 변환하고 다시점 예측 간의 불일치를 측정하는 정규화 손실을 사용하여 네트워크를 학습하였다. 전체 손실 함수는 배치에서 올바른 라벨을 선택하는데 사용되었다. 실험 결과는 자기 앙상블 방법이 적은 계산 오버 헤드로도 기준 성능에 비해 상당한 성능 향상을 제공했으며 제안된 방법이 대부분의 범주에서 기존 방법을 능가한다는 것을 보여주었다.

ACKNOWLEDGMENT

이 연구는 2021 년도 한국연구재단 부설 정보통신기획평가원 연구비지원에 의한 연구임(과제번호 : 202100000001840)

참 고 문 헌

[1] HAN, Bo, et al. Co-teaching: robust training of deep neural networks with extremely noisy labels. In: *Proceedings of the 32nd international Conference on Neural Information Processing Systems*. 2018. P. 8536-8546.

[2] YU, Xingrui, et al. How does disagreement help generalization against label corruption?. In: *International Conference on Machine Learning*. PMLR, 2019. P. 7164-7173.

[3] WEI, Hongxin, et al. Combating noisy labels by agreement: A joint training method with co-regularization. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020. P. 13726-13735.