

Class-Balanced Loss Based on Effective Number of Samples

202401819 김
언지

논문 개요

- 제목: *Class-Balanced Loss Based on Effective Number of Samples*
- 저자: Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, Serge Belongie
- 학회: CVPR 2019 (발표 2019, 인용 수 2,500+)
- 핵심 키워드:

#Class Imbalance #Long-Tailed Distribution

#Loss Re-weighting #Effective NumberCB Loss

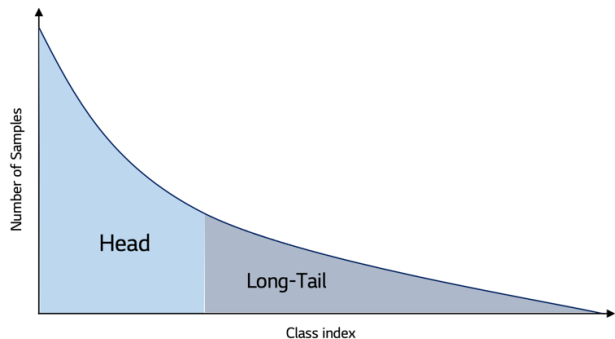
Introduction - problems / backgrounds

논문에서 제시하는 문제점

Long Tail data distribution

소수의 클래스가 대부분의 데이터를 차지하고,
대부분의 클래스는 데이터가 부족한 상황

(클래스가 총 1,000개 있는데, 상위 10개 클래스가 전체 데이터의 80%를 차지하고 나머지 990개는 샘플이 매우 적은 경우)



Class Imbalance

클래스별 샘플 수에 차이가 큰 경우

(고양이 클래스의 데이터 수가 70000개, 호랑이의 데이터 수가 70개 ... 와 같은 경우)

Introduction - problems / backgrounds

논문에서 제시하는 문제점

Long Tail data distribution / Class Imbalance 왜 문제일까? :

CNN은 ImageNet, CIFAR 같은 대규모 균형 잡힌 데이터셋에서 잘 작동하지만, 실제(real-world) 데이터는 위의 문제들을 가지고 있는 경우가 많음.

- 몇몇 클래스(major classes)는 데이터가 많고
- 대부분의 클래스(minor classes)는 데이터가 아주 적음.

이런 경우, 많은 데이터를 가진 쪽으로 가중치가 조정되기 때문에

적은 데이터를 가진 클래스에 대해 모델 성능이 매우 낮아지게 됨.

Introduction - problems / backgrounds

기존 해결 방법

리샘플링 (re-sampling)

- undersampling : 많은 데이터를 가지고 있는 class의 데이터를 버리는 방법
- oversampling : 적은 데이터를 가지고 있는 class의 데이터를 복원 추출하는 방법

리웨이팅 (re-weighting)

loss에 클래스마다 다른 가중치를 곱해주는 것.

손실 함수에서 적은 클래스에 더 큰 중요도(weight)를 부여

단순히 클래스 샘플 수가 적으면 가중치를 높이고, 많으면 줄이는 방식

$$\text{Weight}_c = \frac{1}{n_c} \quad \text{\#데이터수의역수}$$

$$\text{Weight}_c = \frac{1}{\sqrt{n_c}}$$

#데이터수의제곱근의역수

Introduction - problems / backgrounds

기존 해결 방법의 한계

리샘플링 (re-sampling)

- undersampling → 정보 손실
- oversampling → 과적합 위험

리웨이팅 (re-weighting)

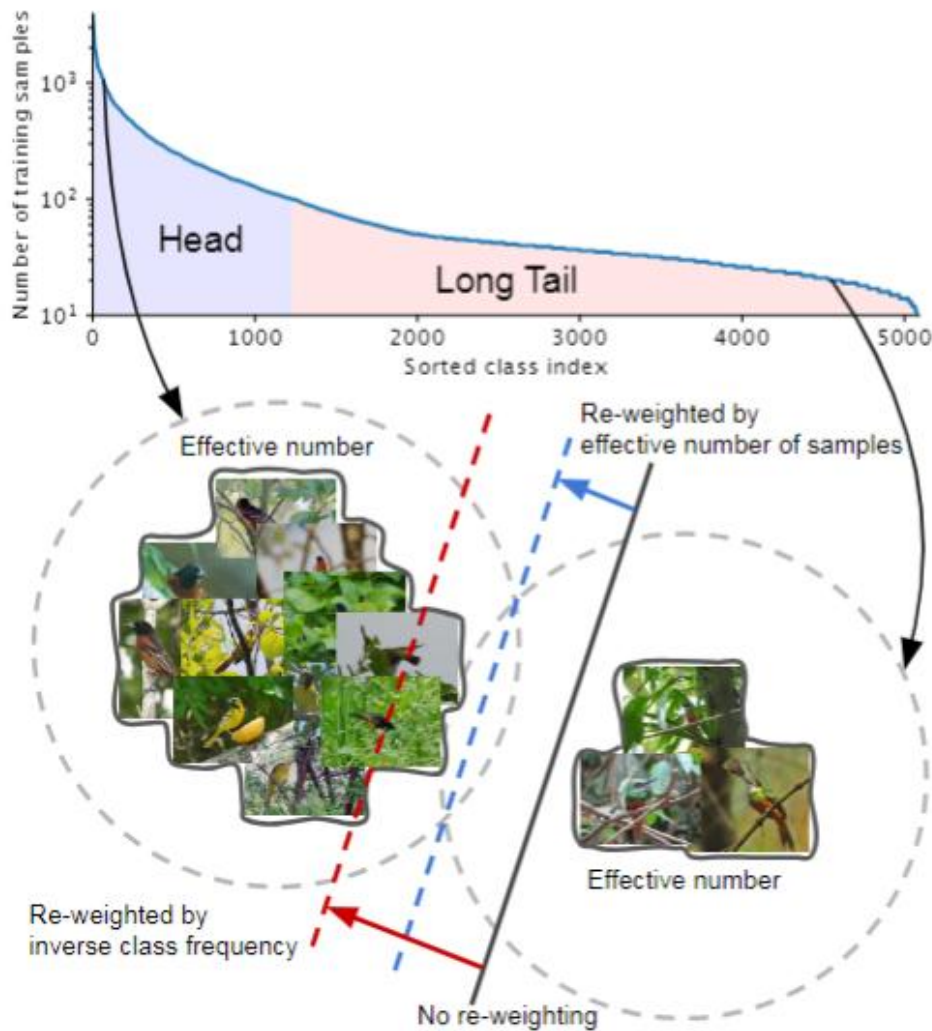
데이터 수가 많더라도 중복된 이미지가 많아 정보량은 적을 수 있음
→ 실질적인 정보량(effective information)을 고려하지 않음

$$\text{Weight}_c = \frac{1}{n_c} \quad \begin{array}{l} \# \text{ 너무 단순} \\ \# \text{ 극단적} \end{array}$$

Motivation & Key Idea

샘플 수가 증가할수록(클래스의 데이터 수가 증가할수록) 실질적인 정보량의 증가는 둔화함

- 데이터 중첩(data overlap) 고려
 - 데이터 하나하나를 점(point)이 아니라, 작은 영역으로 바라봄
 - 이 논문에서는 한 클래스의 샘플 수가 커질수록 중복이 커진다고 가정함
- 유효한 데이터 수(Effective Number of Samples) 고려
 - 한 클래스 내에서 데이터들끼리 서로서로 많이 겹친다는것은 실질적인 정보량이 작다는 것을 의미함.
 - 따라서 무조건 “클래스의 샘플 수가 많다 = 정보량이 많다” 가 아니라는 것.



NO re-weighting

: 클래스 불균형으로 인해 샘플 수가 많은 쪽으로 모델이 치우침

Effective number of samples re-weighting

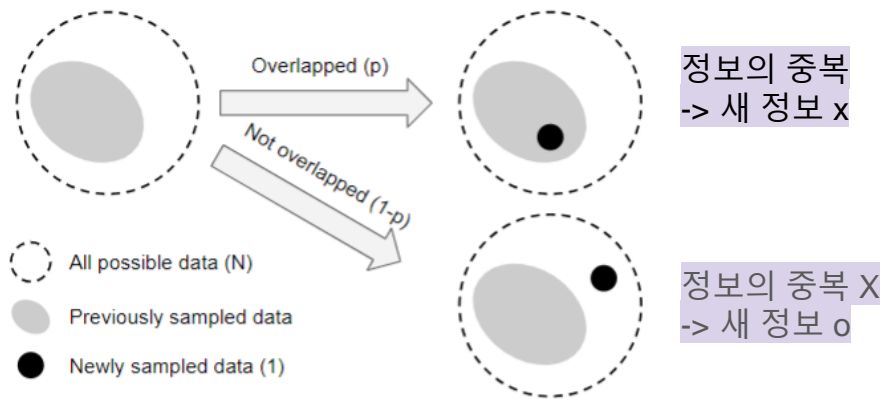
: (논문 제안) 효과적인 샘플 수 기반으로 두 개의 클래스를 가장 적절하게 나눔

Inverse class frequency re-weighting

: 너무 단순하며, 오히려 샘플 수가 적은 쪽으로 모델이 치우치는 등 성능이 나쁨

Theoretical Framework

데이터는 작은 영역 > 즉 전체 데이터(샘플)
는 그 공간들을 덩어리로 합친 부피



데이터들이 많이 겹친다면 부피 증가가 거의 없다.

> 실질적 정보량이 적다는 뜻

데이터들이 적게 겹친다면 부피 증가가 매우 크다

> 실질적 정보량이 많다는 뜻

Theoretical Framework

각 클래스에 대해 실제로 유효한 샘플 수를 계산하는 이론적인 프레임워크를 제시

E_n

= Effective Number of Samples

= “효과적으로 "의미"를 가지고 있는 샘플의 수.

= $(1 - \beta^n) / (1 - \beta)$

$$E_n = \frac{1 - \beta^n}{1 - \beta}$$

n : 샘플 개수

β : 데이터 간의 중복 정도를 조절하는 파라미터 (0~1 사이), 새로운 샘플이 기존 정보와 겹치지 않을 확률

$\beta \rightarrow 1$ 은 중복이 거의 없다는 가정, $\beta \rightarrow 0$ 은 중복이 매우 많다는 가정

단순히 샘플 개수 n 만 가지고는 실질적인 정보량을 알 수 없음.

겹치는 정도를 반영해서 실질적으로 유효한 샘플 수(E_n)를 정의

Effective Number of Samples

$$\frac{(1-\beta^n)}{1-\beta}$$

* 각각의 샘플은 β 만큼의 확률로 기존 샘플과 중복이 된다 *
 ↓
 (= 각 샘플은 $(1-\beta)$ 만큼의 확률만큼 중복이 안 된다 (= 새로운 정보이다) *

P 기댓값의 P
 0 총 합 0

$$E_n = \sum_{i=0}^{n-1} \beta^i = \frac{1-\beta^n}{1-\beta}$$

↓
 E_n 정보 샘플 수.

첫 번째 샘플 → 무조건 새로운 정보 → 1만큼의 실질적 정보를 가짐

두 번째 샘플 → 즉 새로운 정보를 줄 확률은 $1-\beta$
 → 새 정보에 대한 기여량은 β^1 로 1보다 작아짐..

세 번째 샘플 → 새 정보에 대한 기여량은 β^2 로 β^1 보다 작아짐..

네 번째 샘플

⋮

n 번째 샘플

(정보 기여량)

1

β

β^2

β^3

⋮

β^{n-1}

n개의 샘플을 가진
 데이터의 정보
 기여량 총 합

$E_n =$

$$\sum_{i=0}^{n-1} \beta^i = \frac{1-\beta^n}{1-\beta}$$

↗
 강 등비수열의
 합이잖아..

β 는 실험적으로 설계
 된 값으로 지정

보통은 0.9나 0.99처럼,
 1에 가까운 값으로 설
 정

CB Loss Function

Effective Number of Samples > class-balanced weight > class-balanced loss

$$E_{n_i} = \frac{1 - \beta^{n_i}}{1 - \beta}$$

$$\alpha_i = \frac{1 - \beta}{1 - \beta^{n_i}}$$

$$\text{CB}(p, y) = \frac{1 - \beta}{1 - \beta^{n_y}} \cdot L(p, y)$$

(효과적인 샘플 수의 역수) (로스 값에 클래스-밸런스 가중치 곱해주기)

1. 기존의 inverse frequency re-weighting보다 더 부드러운 조절이 가능 (빈도가 낮을수록 가중치를 더 크게 주는 방식)
2. 클래스 빈도가 너무 작거나 크더라도 안정적 손실 조절 가능
3. 모델 및 손실 함수의 종류에 관계 없이 적용 가능하므로 재사용성 높음.

모델 및 손실 함수의 종류에 관계 없이 적용 가능하므로 재사용성 높음.

$$\text{CB}_{\text{softmax}}(\mathbf{z}, y) = -\frac{1 - \beta}{1 - \beta^{n_y}} \log \left(\frac{\exp(z_y)}{\sum_{j=1}^C \exp(z_j)} \right) \quad \begin{array}{l} \text{SOFTMAX} \\ + \text{CB} \end{array}$$

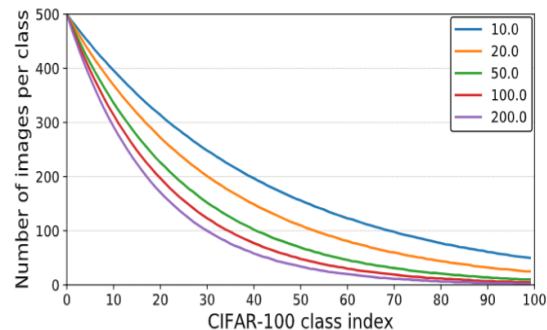
$$\text{CB}_{\text{sigmoid}}(\mathbf{z}, y) = -\frac{1 - \beta}{1 - \beta^{n_y}} \sum_{i=1}^C \log \left(\frac{1}{1 + \exp(-z_i^t)} \right) \quad \begin{array}{l} \text{SIGMOID} \\ + \text{CB} \end{array}$$

$$\text{CB}_{\text{focal}}(\mathbf{z}, y) = -\frac{1 - \beta}{1 - \beta^{n_y}} \sum_{i=1}^C (1 - p_i^t)^\gamma \log(p_i^t) \quad \begin{array}{l} \text{FOCAL} \\ + \text{CB} \end{array}$$

Experiments / Datasets

다양한 모델 크기에서도 CB Loss가 일관되게 효과적인지 검증

Dataset Name	# Classes	Imbalance
Long-Tailed CIFAR-10	10	10.00 - 200.00
Long-Tailed CIFAR-100	100	10.00 - 200.00
iNaturalist 2017	5,089	435.44
iNaturalist 2018	8,142	500.00
ILSVRC 2012	1,000	1.78



CIFAR-10/100 (long-tailed) : 클래스 간 샘플 수 비율을 10~200배로 인위적으로 조정

iNaturalist 2017/2018 : 현실 세계의 자연스러운 long-tail 분포

ImageNet (ILSVRC 2012) : 비교적 균형 잡힌 분포 (imbalance ratio: 1.78)

Results 1

CB Loss vs 기존 Loss 성능 비교 (CIFAR 10 과 CIFAR 100에서)

Dataset Name	Long-Tailed CIFAR-10						Long-Tailed CIFAR-100					
Imbalance	200	100	50	20	10	1	200	100	50	20	10	1
Softmax	34.32	29.64	25.19	17.77	13.61	6.61	65.16	61.68	56.15	48.86	44.29	29.07
Sigmoid	34.51	29.55	23.84	16.40	12.97	6.36	64.39	61.22	55.85	48.57	44.73	28.39
Focal ($\gamma = 0.5$)	36.00	29.77	23.28	17.11	13.19	6.75	65.00	61.31	55.88	48.90	44.30	28.55
Focal ($\gamma = 1.0$)	34.71	29.62	23.29	17.24	13.34	6.60	64.38	61.59	55.68	48.05	44.22	28.85
Focal ($\gamma = 2.0$)	35.12	30.41	23.48	16.77	13.68	6.61	65.25	61.61	56.30	48.98	45.00	28.52
Class-Balanced	31.11	25.43	20.73	15.64	12.51	6.36*	63.77	60.40	54.68	47.41	42.01	28.39*
Loss type	SM	Focal	Focal	SM	SGM	SGM	Focal	Focal	SGM	Focal	Focal	SGM
β	0.9999	0.9999	0.9999	0.9999	0.9999	-	0.9	0.9	0.99	0.99	0.999	-
γ	-	1.0	2.0	-	-	-	1.0	1.0	-	0.5	0.5	-

#오차율(Error Rate)

Results 2

iNaturalist, ImageNet 성능 분석

#정확도(Accuracy)

					iNaturalist 2017		iNaturalist 2018		ILSVRC 2012	
Network	Loss	β	γ	Input Size	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
ResNet-50	Softmax	-	-	224×224	45.38	22.67	42.86	21.31	23.92	7.03
ResNet-101	Softmax	-	-	224×224	42.57	20.42	39.47	18.86	22.65	6.47
ResNet-152	Softmax	-	-	224×224	41.42	19.47	38.61	18.07	21.68	5.92
ResNet-50	CB Focal	0.999	0.5	224×224	41.92	20.92	38.88	18.97	22.71	6.72
ResNet-101	CB Focal	0.999	0.5	224×224	39.06	18.96	36.12	17.18	21.57	5.91
ResNet-152	CB Focal	0.999	0.5	224×224	38.06	18.42	35.21	16.34	20.87	5.61
ResNet-50	CB Focal	0.999	0.5	320×320	38.16	18.28	35.84	16.85	21.99	6.27
ResNet-101	CB Focal	0.999	0.5	320×320	34.96	15.90	32.02	14.27	20.25	5.34
ResNet-152	CB Focal	0.999	0.5	320×320	33.73	14.96	30.95	13.54	19.72	4.97

CB Focal Loss 사용 시 ResNet-50이 ResNet-152급 성능