

1. はじめに

クラス分類タスクを対象とした深層学習モデルの学習は、正解ラベルと予測確率間の損失を最小化することを目的として、パラメータを更新する。この目的を達成するために、様々な損失関数が提案されている。一方で各クラスのデータが不均衡な Long-Tailed 物体認識においては、各損失関数の効果について深く分析されていない。そこで本研究では、Long-Tailed 物体認識での各損失関数の認識精度および再現率を評価し、損失関数の組み合わせ傾向を調査する。

2. 損失関数

画像認識を対象とした深層学習モデルの損失関数は、「出力空間に対するもの」と「特徴空間に対するもの」の2つに大別できる。

2.1. 出力空間に対する損失関数

Cross Entropy Loss Cross Entropy Loss (CE Loss) は、正解クラスの予測確率を 1 に近づける損失関数である。CE Loss (H) は、 N をサンプルの総数、 y を正解ラベルが 1 の確率分布、 $\hat{y}$ を予測確率分布をすると、式 (1) となる。

$$H(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^N y_i^T \cdot \log(\hat{y}_i) \quad (1)$$

Complement Entropy Complement Entropy (Comp. Loss)[1] は、不正解クラスのエントロピーを考慮する損失関数である。 g を正解クラスのインデックスとすると Comp. Loss は式 (2) となり、 $C(\hat{y})$ が最大となるように学習する。Comp. Loss の最大化により、高い確率の不正解クラスが正解クラスと競合する可能性を低下させる働きを持つ。

$$C(\hat{y}) = -\frac{1}{N} \sum_{i=j}^N \sum_{j=1, j \neq i}^k \frac{\hat{y}_{ij}}{1 - \hat{y}_{ig}} \log \left(\frac{\hat{y}_{ij}}{1 - \hat{y}_{ig}} \right) \quad (2)$$

2.2. 特徴空間に対する損失関数

Prototype Conformity Loss Prototype Conformity Loss (PC Loss) [2] は、特徴空間において、同じクラスの特徴量を近づけ、異なるクラスの特徴量を離す損失関数である。異なるクラスの特徴量を離すことで、識別境界線付近にあるサンプルの誤分類を抑制する。学習可能なクラス重心 w^c と特徴量 f_i の関係を $\alpha_i = \|f_i - w_{y_i}^c\|_2$, $\beta_i = \|f_i - w_j^c\|_2$, $\gamma_i = \|w_{y_i}^c - w_j^c\|_2$ とすると PC Loss は式 (3) となる。これより、 α_i を最小化することで特徴量をクラス重心に集め、 $(\beta_i + \gamma_i)$ を最大化することでクラス間を離すように働く。

$$L_{pc} = \sum_i \left\{ \alpha_i - \frac{1}{k-1} \sum_{j \neq y_i} (\beta_i + \gamma_i) \right\} \quad (3)$$

3. 評価実験

本実験では、実世界のデータ不均衡を表した Long-Tailed データセットを用いて各損失関数で学習した際の最適な組み合わせを調査する。

3.1. 実験概要

本実験では、PC Loss, Comp. Loss をクラス全体および Head クラス, Tail クラスに適用して学習を行う。ここで、データ数が 100 枚以上のクラスを Head クラス、100 枚未満のクラスを Tail クラスと定義する。

3.2. 実験条件

ネットワークモデルは ResNet-20、活性化関数は ReLU、バッチサイズは 128、学習の更新回数は 300 エポック、最適化手法は CE Loss, Comp. Loss の場合に Momentum SGD, PC Loss の場合に SGD を用いる。学習用データセットは CIFAR-100-LT を用い、不均衡率は 100 とする。また、CE Loss で 10 エポック事前学習したモデルを用いる。

3.3. 実験結果

各損失関数による精度比較を表 1 に示す。表 1 より、Comp. Loss を適用する事で Tail クラスの精度向上を確認した。PC Loss は、全クラスに適用すると精度が低下する一方、Head クラスに適用すると精度が大幅に向上することを確認した。これは、Head クラスの特徴空間を優先的に形成する事で、Tail クラスを容易に学習できるからであると考えられる。

表 1: 各損失関数による精度比較 [%]

Comp. Loss		PC Loss		All	Head クラス	Tail クラス
Head クラス	Tail クラス	Head クラス	Tail クラス			
				41.14	69.34	25.95
✓	✓			41.97	69.62	27.07
✓				42.05	68.65	27.72
	✓			41.17	69.40	25.96
		✓	✓	40.35	68.94	24.95
		✓		43.14	70.45	28.43
			✓	41.55	68.74	26.90
✓			✓	41.47	68.45	26.93
	✓	✓		43.61	69.95	29.47

Head クラスに PC Loss, Tail クラスに Comp. Loss を適用した場合に最も精度が向上することが確認できる。特に Tail クラスの精度が大幅に向上している。これは、PC Loss により Head クラスのクラスタ分布を小さくする事で、Tail クラスのクラスタ分布との重なりが減り、精度が向上したと考える。

3.4. 考察

学習初期と学習後期のクラス毎の認識精度を可視化し、CE Loss のみの場合と比較する。Test データにおける学習初期と学習後期のクラス毎の認識精度を図 1 に示す。図 1 より、学習初期では Head クラスの正解率が高く、学習後期になると Tail クラスの正解率が高くなる事が確認できる。特に、学習初期に PC Loss は、Head クラスに焦点を当てていることがわかる。これにより、Head クラスの学習を優先的にしながら、Tail クラスおよび全体が学習されていると言える。以上より、PC Loss と Comp. Loss の組み合わせは実世界のデータ分布に有効であると考えられる。

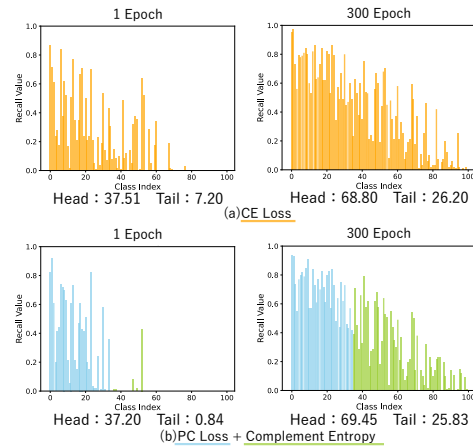


図 1: 学習初期と学習後期の認識精度

4. おわりに

本研究では、実世界のデータ分布における各損失関数の組み合わせ傾向を調査した。PC Loss を Head クラス、Comp. Loss を Tail クラスに適用する事で、Long-Tailed データセットの学習を効率的にできることを確認した。今後は、データセットを変更して実験を行う予定である。

参考文献

- [1] H. Chen, *et al.*, "COMPLEMENT OBJECTIVE TRAINING", ICLR, 2019.
- [2] A. Mustafa, *et al.*, "Adversarial Defense by Restricting the Hidden Space of Deep Neural Networks", ICCV, 2019.