

## クラス間距離の最大化に基づく コサイン類似度ベース画像分類手法の検討

工藤 忍<sup>†</sup>  
Shinobu Kudo

谷田 隆一<sup>†</sup>  
Ryuichi Tanida

木全 英明<sup>†</sup>  
Hideaki Kimata

### 1. はじめに

近年、画像の分類タスクにおいてコサイン類似度を用いた手法が提案されている。この手法は入力画像から抽出した特徴ベクトルを超球面上に投影し、クラス代表ベクトルとのコサイン類似度が大きくなるように学習する手法であるが、クラス間距離(クラス代表ベクトル同士のなす角)が小さいため2つの問題を引き起こす。1つ目は隣接するクラスとの距離が小さいクラスでは隣接するクラスに誤って分類されやすい点である。2つ目は特徴量空間を最大限に活用できていないため、特徴量空間の表現能力が低下する点である。いずれも分類精度の低下を引き起こす。本稿ではクラス代表ベクトル間の最小距離を最大化することで上記2つの問題を解消し、従来手法よりも高精度な画像分類手法を提案する。

### 2. 従来技術

#### 2.1. コサイン類似度ベース画像分類手法

コサイン類似度ベースの画像分類手法は入力画像から抽出した特徴ベクトルを超球面上に投影し、クラス代表ベクトルとのコサイン類似度を利用した分類手法であり、従来の Softmax 関数を用いた手法と比較して、距離学習の性質を有している、手法が解釈しやすい、などの特徴を有する。超球面上でのペナルティの課し方で様々な手法が提案されており [1][2][3], その中でも Deng らの提案する ArcFace[3] は超球面上での角度  $\theta$  にペナルティを課す手法であり、高い分類精度が報告されている。

ArcFace は以下の目的関数を最小化する最適化問題として定式化される。

$$L_0 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i} + m))}}{e^{s(\cos(\theta_{y_i} + m))} + \sum_{j=1, j \neq y_i}^n e^{s(\cos \theta_j)}}. \quad (1)$$

ここで  $N$  はサンプル数,  $\theta$  は入力画像  $x$  の特徴ベクトル  $f(x)$  とクラス  $y$  の代表ベクトル  $W_y (0 \leq y < n)$  ( $n$  はクラス数) のなす角,  $s$  は超球面半径,  $m$  は角度マージンを表す。

#### 2.2. クラス間距離を考慮した手法

コサイン類似度ベースの手法ではクラス内分散が小さく、クラス間分散が大きくなるのが望ましい。そこで後者に関して明示的にクラス間距離をペナルティとして最適化に組み込む手法 [3][4] が提案されている。Deng ら [3] は以下の式でクラス間距離を定義し、正則化項と

表 1 従来技術を CIFAR100 に適用した場合のクラス間距離

| 手法                           | 最小値   | 最大値   | 平均値   |
|------------------------------|-------|-------|-------|
| ArcFace( $L_0$ )             | 0.207 | 0.414 | 0.346 |
| ArcFace( $L_0 + L_{inter}$ ) | 0.149 | 0.416 | 0.352 |

して最適化する検討を行なっている。

$$L_{inter} = -\frac{1}{\pi N(n-1)} \sum_{i=1}^N \sum_{j=1, j \neq y_i}^n \arccos(W_{y_i}^T W_j). \quad (2)$$

### 2.3. 従来技術の問題点

100 クラスの画像分類データセットである CIFAR100[5] に対して、ArcFace(目的関数を  $L_0$  あるいは  $L_0 + L_{inter}$  とした場合) を適用した時のクラス間距離(クラス代表ベクトル同士のなす角  $\theta$ ) の最小値, 最大値, 平均値を表 1 に示す。表 1 から ArcFace( $L_0$ ) はクラス間距離の平均が 0.346 であり、最大でも 0.414 と小さい値を示していることが分かる。ここから2つの問題点が挙げられる。

1つ目はクラス間距離が小さいと、隣接するクラス同士でお互いに誤ったクラスに分類されやすくなる点である。2つ目は特徴量空間の一部にクラス代表ベクトルが集中しているため、特徴量空間を最大限に活用できておらず、表現能力が低下する点である。

一方で、クラス間距離を考慮した ArcFace( $L_0 + L_{inter}$ ) ではクラス間距離を考慮していない ArcFace( $L_0$ ) と比較してクラス間距離の平均値を大きくすることができている。しかし、最小値が ArcFace( $L_0$ ) よりも小さくなっているため、上記の1つ目の問題がより顕著になり分類精度の低下を引き起こす。

### 3. 提案法

#### 3.1. 着眼点

従来のクラス間距離を考慮した正則化は全クラス間距離の平均を最大化しているため、特定のクラス間距離が小さくなってしまいう可能性があった。そこで本稿では全クラス間距離の中から最小のクラス間距離を最大化するように最適化を行う。これによりクラス間距離の最小値が引き上げられることでクラス代表ベクトルが特徴量空間でより広く分布し、特定のクラス同士が近くなるという問題が解消されることが期待される。

<sup>†</sup>日本電信電話株式会社, NTT メディアインテリジェンス研究所

表2 各手法の分類精度及びクラス間距離

| 手法                            | Top1 精度 [%]  | CIFAR10 |       |       | Top1 精度 [%]  | CIFAR100 |       |       |
|-------------------------------|--------------|---------|-------|-------|--------------|----------|-------|-------|
|                               |              | クラス間距離  |       |       |              | クラス間距離   |       |       |
|                               |              | 最小値     | 最大値   | 平均値   |              | 最小値      | 最大値   | 平均値   |
| ArcFace( $L_0$ )              | 90.95        | 0.370   | 0.371 | 0.371 | 66.68        | 0.207    | 0.414 | 0.346 |
| ArcFace( $L_0 + L'_{inter}$ ) | 90.70        | 0.376   | 0.376 | 0.376 | 66.99        | 0.149    | 0.416 | 0.352 |
| 提案法 ( $L_0 + L'_{inter}$ )    | <b>90.99</b> | 0.381   | 0.381 | 0.381 | <b>68.88</b> | 0.337    | 0.428 | 0.355 |

### 3.2. クラス間距離の最大化に基づくコサイン類似度ベース画像分類手法

提案法の目的関数を以下に示す.

$$L = L_0 + L'_{inter}. \quad (3)$$

ここで  $L'_{inter}$  は本稿で提案するクラス間距離のペナルティであり, 以下の式で定義する.

$$L'_{inter} = -\frac{1}{\pi} \min(\arccos(W^T W)). \quad (4)$$

ここで  $\min(\cdot)$  は行列の中から最小の要素を取り出す操作を表す.

## 4. 評価実験

提案法の有効性を確認するためシミュレーション実験を行った.

### 4.1. 実験条件

比較手法として ArcFace(目的関数を  $L_0$  及び  $L_0 + L'_{inter}$  としたもの) との比較を行った. 本実験では目的関数の影響のみを比較するため, 提案法及び比較手法のネットワーク構成を Resnet18[6] で統一し, Resnet18 との変更点として最終層の直前にドロップアウト ( $p = 0.2$ )[7] を挿入し, 最終層の特徴ベクトルは 512 次元とした. データセットは CIFAR10 及び CIFAR100[5] を使用し,  $224 \times 224$  画素にリサイズして用いた. 最適化は Adam[8] を用い, バッチサイズは 128, 反復回数は 100epoch に設定し, 学習率は 0.001 から始め 30epoch 毎に 1/10 倍に減衰させた. また, その他のパラメータは  $s = 64, m = 0.5$  とした.

### 4.2. 結果及び考察

表2に CIFAR10 と CIFAR100 について各手法の分類精度 (Top1) 及びクラス間距離の最小値, 最大値, 平均値を示す. CIFAR10 では提案法が最も高い分類精度を示したものの他の比較手法と顕著な差は無く, クラス間距離の差もほとんど認められなかった. 考えられる要因として CIFAR10 の分類クラス数が 10 クラスと少なかったため, 従来の ArcFace( $L_0$ ) でも十分にクラス間距離を大きくすることができたためであると考えられる. 一方で分類クラス数が 100 クラスである CIFAR100 の結果では提案法の分類精度が比較手法よりも約 2.0% 向上していることが確認できる. クラス間距離の結果が示す通り, 提案法は狙い通り最小値が比較手法よりも大きくなっていることが示され, これが分類精度の向上に寄与したと考えられる.

## 5. まとめ

本稿では全クラス間距離の中から最小のクラス間距離を最大化するコサイン類似度ベースの分類手法を提案した. シミュレーション実験により提案手法がクラス数が多い場合において, 従来手法を顕著に上回る分類精度を達成することを示した.

## 参考文献

- [1] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition, 2018.
- [2] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphreface: Deep hypersphere embedding for face recognition, 2017.
- [3] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition, 2018.
- [4] Weiyang Liu, Rongmei Lin, Zhen Liu, Lixin Liu, Zhiding Yu, Bo Dai, and Le Song. Learning towards minimum hyperspherical energy, 2018.
- [5] Alex Krizhevsky. Learning multiple layers of features from tiny images, 2009.
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.
- [7] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, Vol. 15, No. 56, pp. 1929–1958, 2014.
- [8] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.