スーパークラスを考慮した mixup による Deep Mutual Learning の高精度化

EP17099 藤井駿伍　　指導教授：山下隆義

1. はじめに

Deep Mutual Learning（DML）[1] は，複数のモデル間で知識を伝達し，相互に学習を行う手法である．DML の基となる蒸留（Distillation）[2] では，学習済みモデルによるクラスの予測確率を教師として用いることで，モデル性能を向上させている．予測確率は Soft Target と呼び，正解クラスと他のクラスとの相対的な類似性を表現している．しかしながら，予測確率はモデル依存であり，人間が目標とした予測確率にならない．

入力データを混合して人間が目標とした予測確率となるような正解ラベルを作成できる方法として，mixup[3] がある．mixup は，Data Augmentation 手法として用いられ，認識精度の向上を図ることができる．一方で，mixup は混合するデータのクラスによらず同じベータ分布に従い混合するため，類似性の低いクラス同士の中間的な混合データが生成されることもある．そこで，本研究では，クラスの上位集合であるスーパークラスを利用した mixup を提案する．

2. Deep Mutual Learning

DML は，複数のモデル間で予測確率を知識として伝達し，単体学習を上回る精度を実現する手法である．DML の概略図を図 1 に示す．学習用データを用いた通常の学習に加え，入力データに対する予測確率がモデル間で互いに近づくように学習を行う．

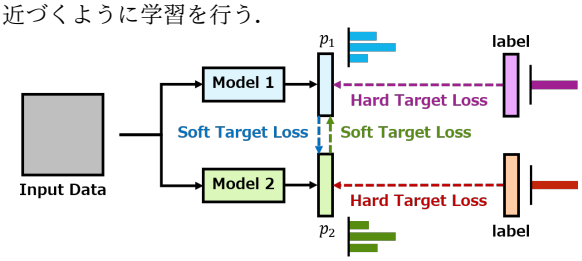

図 1：Deep Mutual Learning

3. mixup

mixup は，2 枚の画像及びラベルを混合し，新たなデータを生成する手法である．画像及びラベルの混合処理を式 (1)，式 (2) に示す．

$$\tilde{x} = \lambda x_i + (1-\lambda)x_j \quad (1)$$

$$\tilde{y} = \lambda y_i + (1-\lambda)y_j \quad (2)$$

ここで，$\tilde{x}$ は混合画像，λ は混合比率，x は処理前画像，$\tilde{y}$ は混合ラベル，y は x に対応する処理前ラベルである．x_i，x_j はランダムに選択され，λ はハイパーパラメータを α とするベータ分布 $Beta(\alpha, \alpha)$ に従う確率で決められる．

4. 提案手法

提案手法では，混合対象のスーパークラスを考慮して mixup の混合比率を選択し，生成する混合データの傾向を調整する．スーパークラスは，各クラスをより大きなカテゴリで分類する上位集合であり，概念辞書である WordNet[4] に従って定義する．提案手法の概念図を図 2 に示す．スーパークラスが同じである orchids クラス，tulips クラスの場合，中間の混合比率になりやすくする．一方，スーパークラスが異なる tulips クラス，trout クラスの場合，いずれかのクラスの比率が高くなるように混合する．これにより，DML において類似するクラス間の予測確率を重点的に学習可能となる．

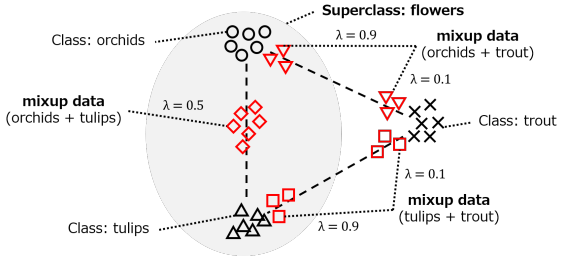

図 2：スーパークラスによる混合データの調整

提案手法における混合比率の選択は，予め設定した選択肢を用いる方法（選択制）と，ベータ分布を用いる方法が考えられる．前者は，混合対象のスーパークラスが同じ場合は中間の混合比率で混合し，異なる場合はいずれかのクラスの比率が高くなるように混合する．後者は，スーパークラスが同じ場合は α を 8.0，異なる場合は 0.2 とする．ベータ分布を用いることで，スーパークラスの関係に反した比率での混合を許容する．提案手法を適用した画像の例を図 3 に示す．図 3 では，orchids クラスと tulips クラスはスーパークラスが同じであるため，中間の混合比率として 0.5，tulips クラスと trout クラスはスーパークラスが異なるため，tulips クラスの比率が高くなる 0.9 や trout クラスの比率が高くなる 0.1 を選択している．

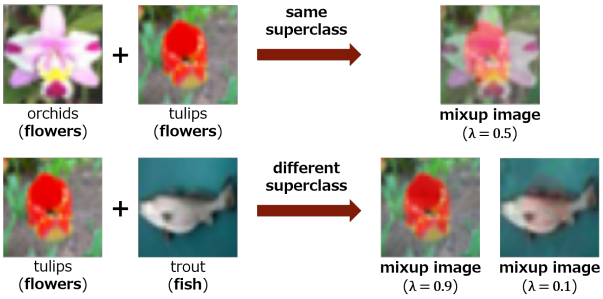

図 3：提案手法を適用した画像

5. 評価実験

評価実験により，提案手法の有効性を評価する．

5.1. 実験概要

モデルは ResNet32 とし，学習及び評価データには CIFAR100 を用いる．CIFAR100 では，100 種類のクラスが 20 種類のスーパークラスに 5 種類ずつ分類されている．損失関数には交差エントロピー，予測確率を近づけ合うモデル間の損失関数には KL-Divergence を使用する．通常の mixup における α の値は 1.0 とする．

5.2. 実験結果

各手法の精度を表 1 に示す．表 1 より，提案手法が最も高精度であり，ベータ分布を使用することによって更に精度が向上していることが分かる．DML において，スーパークラスを利用した mixup は有効であり，スーパークラスの関係に反した例外的な混合データを僅かに含めることも重要であることが分かる．

表 1：定量的評価 [%]

手法		モデル 1	モデル 2
mixup 無し		71.51	71.46
mixup ($\alpha = 1.0$)		71.37	71.08
提案手法	選択制	72.23	71.93
	ベータ分布	**73.28**	**73.01**

6. おわりに

本研究では，蒸留においてクラスの予測確率の知識が重要であることから，DML で混合データを用いて予測確率を効果的に学習する手法を提案した．提案手法では，データを混合する際にスーパークラスを考慮させ，mixup を行う．評価実験では，提案手法の有効性及び例外的な混合データの重要性を確認した．今後は，混合画像の注視領域やモデルの潜在表現による定性的な分析，異なるモデルやデータセットにおける有効性の調査を行う．

参考文献

[1] Y. Zhang, *et al.*, “Deep mutual learning”, CVPR, 2018.

[2] G. Hinton, *et al.*, “Distilling the knowledge in a neural network”, NeurIPS, 2015.

[3] H. Zhang, *et al.*, “mixup: Beyond Empirical Risk Minimization”, ICLR, 2018.

[4] G. A. Miller, “WordNet: A Lexical Database for English”, Commun. ACM, Vol. 38, No. 11, pp. 39–41, 1995.