関数の準凸性を利用したモデル修正の係数の最適化

松下　拓海†　　古川　翔大†
† 鹿児島工業高等専門学校 情報工学科

1. はじめに

近年，モデルを学習する際に事前学習済みモデルをファインチューニングして利用する手法が注目されてきている．本稿では，事前学習済みモデルの特定のタスクにおける精度を修正するモデル修正手法の PAINT [1]について，その係数の最適化手法を提案する．提案手法の目的は精度を保持するタスクと修正するタスクにおける損失関数の和を最小化することである．従って，目的関数を微分可能にするため精度の代わりに損失関数を用い，タスク間のスケール不変性を軽減するため正規化する．全探索と提案手法を比較する実験では損失が最大 2.1%向上した．また，モデル修正における損失関数に現れる性質を調査する．

2. モデル修正手法 PAINT

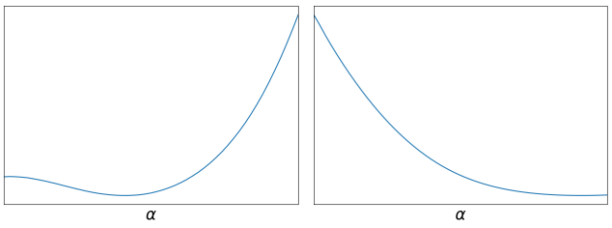
モデル修正の目的は，複数のタスクにおいて学習された事前学習済みモデルの特定のタスクの精度を他のタスクの精度を保ちながら向上させることである．特に，精度を保持するタスクを最適なタスク，精度を修正するタスクを最適ではないタスクと呼ぶ．PAINT [1]は，事前学習済みモデルと精度を修正したいタスクにおいてファインチューニングしたモデルの重みを平坦化し，そのベクトルの線形補間上から最適なモデルを探索する．

3. 提案手法

モデル修正の目的は精度を最大化することであったが，ここでは目的関数を微分可能にするため，精度の代わりに損失関数を用いる．提案手法は最適なタスクと最適ではないタスクにおける損失関数の和を最小化することを目的とする．このとき，損失関数の和は準凸関数であると仮定するが，これは実施した実験で一貫して現れた性質である．提案手法では，この損失関数の和の変曲点を挟み込む最小区間を求め，これを 5 次近似し，その微分から変曲点を求める．

4. 実験結果

既存手法と提案手法を比較するため，CLIP ViT-B/32 [2]をImageNet，CIFAR10,CIFAR100，STL10の精度を保持し，Cars，DTD，EuroSAT，GTSRB，MNIST，RESISC45，SUN397，SVHN，FashionMNISTの精度を向上させる実験を行った．その結果，CIFAR10とGTSRB，STL10とEuroSATを除く全ての組み合わせで全探索より低い損失の和が得られ，最大で2.1%向上した．



(b)　最適なタスク　　(a)　最適ではないタスク

図 1．損失関数の例

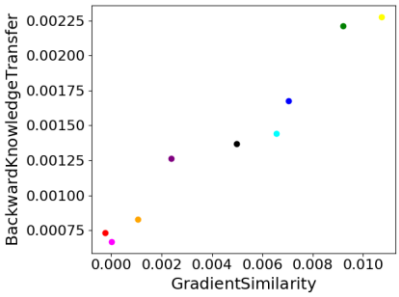


図 2．後方知識転移の定量的指標
最適なタスクをCIFAR10として，全ての最適ではないタスクの組み合わせでプロットした場合

5. 後方知識転移の定量的指標

後方知識転移とは，モデルを複数のタスクにおいて連続して学習させた場合に，新たなタスクにおける学習が既に学習されたタスクにおける精度の向上に寄与する現象である．実験結果より，最適なタスクにおける損失関数には後方知識転移が現れることが分かった．次に，この後方知識転移を予測するため，次のような勾配の類似度となる指標と後方知識転移の定量化を考えた．

$$\text{GradientSimilarity} = \frac{(\boldsymbol{w}_{\text{zs}} - \boldsymbol{w}_{\text{ft}})^{\top} \nabla_{\boldsymbol{w}} \mathcal{L}(\boldsymbol{w}_{\text{zs}}, \mathcal{D}_{\text{supp}})}{\|(\boldsymbol{w}_{\text{zs}} - \boldsymbol{w}_{\text{ft}})\|_2 \|\nabla_{\boldsymbol{w}} \mathcal{L}(\boldsymbol{w}_{\text{zs}}, \mathcal{D}_{\text{supp}})\|_2}$$

$$\text{BackwardKnowledgeTransfer} = \max_{\alpha \in [0,1]} \frac{\mathcal{L}(\boldsymbol{w}_{\text{zs}}, \mathcal{D}_{\text{supp}}) - \mathcal{L}(\alpha, \mathcal{D}_{\text{supp}})}{\alpha \|(\boldsymbol{w}_{\text{zs}} - \boldsymbol{w}_{\text{ft}})\|_2}$$

図 2 にこれらをプロットしたものを示す．図よりこれらの指標に正の相関が見られた．

6. 今後の課題

今後は破滅的忘却の定量的指標について調査を進める予定である．

参考文献

[1] Gabriel Ilharco, *et al.*, Patching open-vocabulary models by interpolating weights, 2022.
[2] Alec Radford, *et al.*, Learning transferable visual models from natural language supervision, 2021.