

# マルチタスク学習のためのオンライン Fine-Tuning における動的パラメータ調整手法の検討

佐藤 陽斗† 佐藤 裕二‡

法政大学情報科学部コンピュータ科学科†

## 1. まえがき

Deep Learning において、データ不足に対処するための手法として、転移学習や Fine-Tuning がある。これらの手法は、学習済みモデルの層を再利用することで学習を効率化する。Fine-Tuning を用いた研究として、マルチタスク学習のためのオンライン Fine-Tuning が提案されている[1]。[1]は、2つのタスクを並列して学習し、その学習中に Fine-Tuning を用いて CNN のフィルタを転移する手法である。転移を伴わない学習(スクラッチ学習)と比較して、最大 2%精度が向上した。一方、[1]では転移頻度や転移フィルタの割合を人手によって決定する必要があり、適切なパラメータを設定するのは困難である。そこで本研究では、転移タイミングにおける誤答率(=100-精度)を計算し、前回転移時の誤答率との差によってパラメータを動的に調整する手法を提案する。

## 2. 先行研究とその問題点

マルチタスク学習のためのオンライン Fine-Tuning は、ニューラルネットワーク(NN)の圧縮手法であるプルーニングの指標を用いてフィルタの重要度をランク付けし、重要度が低いフィルタを転移によって上書きする。この手順により、学習時に精度への影響が少ないと思われるフィルタのみを転移対象のフィルタとして上書きすることで有用な知識の忘却を防ぎつつ、転移性能を向上させることができる。一方、転移を行う間隔(転移頻度)や転移するフィルタの割合(転移割合)は人手によってあらかじめ決定する必要がある。これらのパラメータは試行錯誤して決定する必要があり、タスクや学習の進行状況によって最適な設定が異なる可能性がある。この問題を解決するためには、学習状況によってパラメータを動的に調整する必要があると考えられる。

## 3. 誤答率の差に注目した動的パラメータ調整手法の提案

本研究では、オンライン Fine-Tuning において学習状況に応じて転移頻度や転移割合を調整する手法を提案する。提案手法は、転移時における誤答率と前回転移時における誤答率の差によってパラメータを調整する。最初の転移には前回の転移が存在しないため調整は行わない。図 1 に、誤答率の差によるパラメータ調整のイメージ図を示す。横軸は学習の進行を表し、縦線は転移タイミングを表している。e は転移頻度(epoch)、P は転移割合(%)であり、これらを本研究における調整対象のパラメータとする。誤答率の差が減少した場合、前回転移時より誤答率が小さくなっているため、共通の特徴を獲得するための転移が十分に行われていると判断して次の転移を遅らせるよう e を調整し、転移するフィルタの枚数を減らすよう P を調整する。誤答率の差が増加した場合、逆に e を小さく、P を大きくなるよう調整する。次に、誤答率の差を用いたパラメータ調整方法を以下に示す。

- 1) 誤答率の差(%)を変化量とする  
例: 変化前. 30, diff. -5% → 変化後.  $30+5=35$
- 2) 誤答率の差を変化割合とする(小数点以下偶数丸め)  
例: 変化前. 30, diff. -5% → 変化後.  $30+0.05*30=32$
- 3) 転移回数を係数として誤答率の差に乗算したものを変化量とする  
例: 変化前. 30, diff. -5%, 転移回数. 3  
→ 変化後.  $30+3*5=45$
- 4) 転移回数を係数として誤答率の変化割合にかけたものを変化量とする  
例: 変化前. 30, diff. -5%, 転移回数. 3  
→ 変化後.  $30+3*0.05*30=34$
- 5) 4)の係数部分を学習進度の割合として適用  
例: 変化前. 30, diff. -5%, 現在の epoch 数. 100, 学習 epoch 数. 600  
→ 変化後.  $30+600/100*0.05*30=39$

このうち、転移頻度の調整には 1)-5)を用い、転移割合の調整には 2),4),5)を用いる。

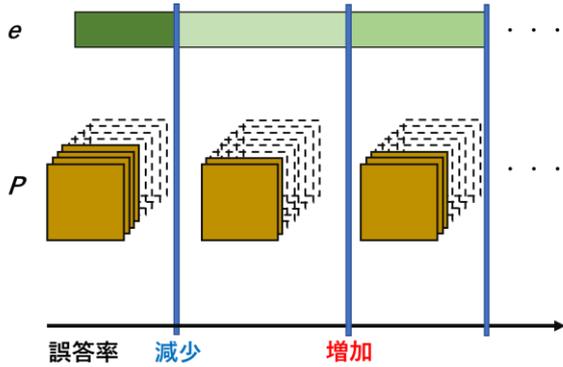


図1: 誤答率の差によるパラメータ調整のイメージ図

#### 4. 評価実験

##### 4.1 実験方法

パラメータ調整なし, 転移頻度のみ調整, 転移割合のみ調整, 両方調整のそれぞれの場合で実験を行う. 実験には CIFAR-100 のデータセットを用いる. CIFAR-100は画像分類タスクで, データセットは 100 クラスに分類される. また, 各クラスはさらに抽象的な 20 個のスーパークラスに属している. 実験では, スーパークラス 1-10 に分類されるクラスのうち, それぞれ 1 クラスずつ抽出した計 10 クラスをタスク 1, 同様にスーパークラス 11-20 から 1 クラスずつ抽出した計 10 クラスをタスク 2 とする. これらのサブセットを 2 つのモデルに割り当てて学習を行う. また, 学習 epoch 数 600epoch, 初期転移頻度 25epoch, 初期転移割合 10%とする. 各実験において 10 回の試行を行い, 平均誤答率を算出する.

##### 4.2 実験結果と考察

図 2 に転移頻度のみ調整した場合, 図 3 に転移割合のみ調整した場合, 図 4 に転移頻度と転移割合の両方を調整した場合の各調整法における平均誤答率を示す. 図 4 において, 調整法( $m, n$ )は転移頻度を調整法  $m$ により調整し, 転移割合を調整法  $n$ により調整することを表す. また, 表 1 に転移頻度のみ調整した場合の各調整法における平均転移回数とプログラムの平均実行時間を示す. 表 1 の結果から, 転移回数が少ないほど実行時間が短いことがわかる. 転移頻度の調整を行った場合, 調整なしと比べ少ない転移回数で学習が行えているため, 提案手法を用いることで学習にかかる時間を短縮できると考える. また, 図 2, 3, 4 より, いくつかの調整法において調整なしより平均誤答率が低い. これは, より適切なタイミングでより適切な割合だけフィルタを転移しているためであると考えられる.

表 1: 転移頻度を調整した場合の平均転移回数と実行時間

	転移回数		実行時間
	タスク 1	タスク 2	
調整なし	23	23	3359.4
調整法 1)	14.2	13.4	3125.7
調整法 2)	19.6	19.3	3255.672
調整法 3)	11.0	10.1	3063.6
調整法 4)	22.5	22.3	3335.6
調整法 5)	9.2	8.6	2977.2

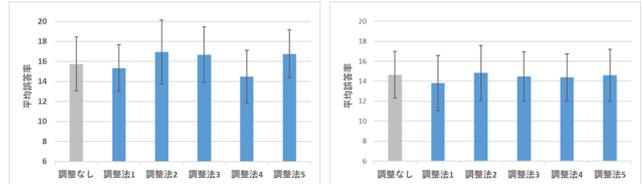


図 2: 転移頻度のみ調整した場合の各調整法における平均誤答率(左タスク 1, 右タスク 2)

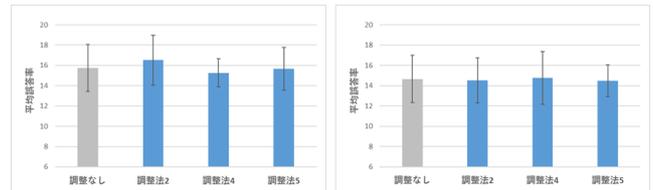


図 3: 転移割合のみ調整した場合の各調整法における平均誤答率(左タスク 1, 右タスク 2)

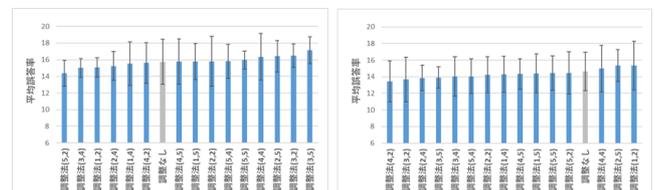


図 4: 転移頻度と転移割合の両方を調整した場合の各調整法における平均誤答率(左タスク 1, 右タスク 2)

#### 5. むすび

本研究では, オンライン Fine-Tuning において効率的な学習のため動的にパラメータ調整を行う手法を提案した. 提案手法は誤答率の差によって学習中にパラメータ調整を行い, 精度および実行時間への影響を検証することを目的とした. 結果として, 調整方法によっては精度を悪化させずに学習時間を短縮することができることを示した.

#### 文献

[1] S. Ikawa, Y. Sato: Multi-task Learning Using Online Fine-Tuning Considering the Importance of Each Filter, Proceedings of the 23rd Asia Pacific Symposium on Intelligent and Evolutionary Systems, Springer, pp. 106-117, 2019.