

# MAMLにおける計算効率の検討

日大生産工 ○五十嵐 堅一 日大生産工 山内 ゆかり

## 1. まえがき

近年、AIが目覚ましい成長を遂げている。その中でも強化学習は、ロボティクスや、医療など様々な領域で活躍している。そのため、様々なタスクに適応する汎化性能が必要とされている。Chelsea Finnらは、学習の仕方を学習する Model-Agnostic Meta-Learning(MAML)[1] を提案し、複数のタスクのサンプリングにより、新しいタスクへの適応能力が上がったことが報告がされている。しかし、複数のタスクにわたる経験を一つのモデルで学習するため、計算コストが非常に高くなるなどの問題がある。本研究では計算量の問題においてパラメータの再利用による計算量の削減手法を提案し、テスト環境における計算機実験により提案手法と比較し、その結果について報告する。

## 2. 従来研究

### 2.1 Model-Agnostic Meta-Learning

Model-Agnostic Meta-Learning(MAML)はメタ学習における代表的なアーキテクチャである。これは初期パラメータである $\theta$ を学習させることにより、別の似たタスクを解く際により早く学習させる手法である。これは学習により得られたパラメータがそのタスク一つを詳しく学習するのではなく、もっと大きくくりのものを学修するかもしれないという直感の下で提案されている。

以下の図1にフローチャートを示す。

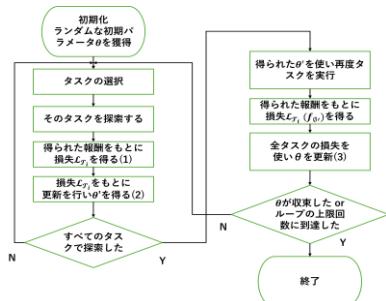


Figure 1 MAML のフローチャート

MAMLは主に二重のループによって形成される。一つ目は初期パラメータである $\theta$ を学習する外側のループである。これはすべてのタスク  $p(\mathcal{T})$  に適応するようなより良い初期パラメー

タを学習するためのループである。もう一つは個別のタスクに適応する初期パラメータ $\theta'$ を学習するための内側のループである。個別に学習された $\theta'$ 全てに適応するように $\theta$ を学習していく。

フローチャート上から、上から順に説明していく。学習される $\theta$ をランダムに初期化する。そして学習を行うタスクの分布を $\mathcal{T}$ として、その中から1つのタスク $T_i$ を選択し、そこで方策 $\pi_{\theta}$ に従い、Kエピソード分の探索を行う。そこで得られた損失 $L_{T_i}$ を計算する。

損失 $L_{T_i}$ の計算式を式(1)に示す。

$$\mathcal{L}_{T_i}(f_{\theta}) = -\mathbb{E}_{x_t, a_t \sim f_{\theta}, q_{\mathcal{T}}} \left[ \sum_{t=1}^H R_i(x_t, a_t) \right] \quad (1)$$

$f_{\theta}$ は現在のネットワークモデル全体を示す。また、 $H$ は1エピソードの長さ、 $R_i(x_t, a_t)$ が現在の状況 $x_t$ で行動 $a_t$ をとった時の報酬とする。ここで、 $\mathcal{L}_{T_i}(f_{\theta})$ は方策 $\pi_{\theta}$ に従った時の1エピソードの報酬の期待値の合計にマイナスをかけたものを表している。つまり報酬の期待値が高いほど損失は小さくなる。次にここでもとめた損失を使い、 $\theta'$ を求める。その式を(2)に示す。

$$\theta' = \theta - \alpha \nabla_{\theta} \mathcal{L}_{T_i}(f_{\theta}) \quad (2)$$

ここで $\nabla_{\theta} \mathcal{L}_{T_i}(f_{\theta})$ は $\mathcal{L}_{T_i}(f_{\theta})$ を $\theta$ について微分したものであり、 $\alpha$ は学習率である。この式は初期パラメータである $\theta$ に損失に学習率とマイナスをかけて引いたものである。このようにして、すべてのタスクについて $\theta'$ を求めてことで、内側のループは終了する。

外側ループでは、内側で求めた $\theta'$ を使い、再度同じタスクを実行する。これにより同じタスクの新しいデータに対して $\theta'$ がどれだけ適応できるかをテストできる。そして内側ループと同様に損失 $\mathcal{L}_{T_i}(f_{\theta'})$ を求める。これを他のタスクの $\theta'$ に対して同様に求め、最後に目的である $\theta$ を更新する。その式は(3)に示す。

$$\theta = \theta - \beta \nabla_{\theta} \sum_{T_i \sim p(\mathcal{T})} \mathcal{L}_{T_i}(f_{\theta'}) \quad (3)$$

ここで $\nabla_{\theta} \mathcal{L}_{T_i}(f_{\theta'})$ は $\mathcal{L}_{T_i}(f_{\theta'})$ を $\theta$ について微分したものであり、 $\beta$ は学習率を示す。このようにして、

$\theta$ を学習させ、上限回数に到達するか $\theta$ が収束すると学習は終了する。しかし、MAMLには $\nabla_{\theta} \mathcal{L}_{\mathcal{T}_i}(f_{\theta'})$ の計算に二階微分が必要であり、計算量が非常に多くなってしまうという欠点がある。

## 2.2 First-Order Meta-Learning Algorithms

First-Order Meta-Learning Algorithms (FOMAML)[2]はMAMLに存在した二階微分による計算量の増大を近似により解決するアーキテクチャである。具体的には、 $\nabla_{\theta} \mathcal{L}_{\mathcal{T}_i}(f_{\theta'})$ を計算する際に $\mathcal{L}_{\mathcal{T}_i}(f_{\theta'})$ を $\theta$ ではなく $\theta'$ について微分するというものである。これによりほとんど性能を落とすことなく計算量の大幅な削減に成功している。一方で、近似をしている関係で確実に最適解に収束するという保証は失われた。

## 3. 提案手法

この膨大な計算量という問題に対して、パラメータの再利用による計算量の更なる削減手法を提案する。通常のMAMLでは、初期パラメタ $\theta$ からタスクの数だけ $\theta'$ を生成し、さらにこれらについての評価を行って最終的な $\theta$ の更新値を決める。しかし、これには冗長な部分が含まれていると考えることが出来る。タスクをまたぐ汎用性を持つ初期パラメータが学習したいので、1つの個別のタスクに適応するパラメータを作成する必要はない。そこで、本研究では1つの $\theta'$ を作成した後、それを最初から複数のタスクに適用し、そのパラメータの評価を行う。今後これをタスクスキップ数と称する。本研究では、タスク数12に対しタスクスキップ数が4の場合、ランダムにタスク3つを選びそれにたいして3つの $\theta'$ を生成する。そしてそれらを選ばれたそれぞれのタスクを含む4つのタスクに適応する。というようにする。これにより計算量が1/4になる。

## 4. 実験および検討

このタスクスキップ数に関して実験を行い、学習効率を下げずに計算量を削減する手法に関して検討する。環境にはArcade Learning Environment (ALE)を採用した。これはAtari 2600のゲームを強化学習エージェントのベンチマークとして利用するためのプラットフォームである。このAtari2600のゲームの中から12個のゲームを事前学習した従来のFOMAMLと、事前学習していないネットワーク、タスクスキップ数が2,3,4,6のFOMAMLの6つを使い、事前学習に使用しなかった他のゲームの学習速度を比較する。

## 5. まとめ

本研究では、タスクスキップによる計算量の削減を提案した。これにより、学習精度の低下が起こることが予測されるが、一方で計算量の大幅な削減により学習の効率化が起こると予測される。

## 参考文献

- 1) Chelsea Finn, Sergey Levine, Pieter Abbeel and Model-Agnostic, “Meta-Learning for Fast Adaptation of Deep Networks”, Proceedings of the 34th International Conference on Machine Learning (2017) p.1126-1135.
- 2) Alex Nichol, Joshua Achiam and John Schulman, “On First-Order Meta-Learning Algorithms”, arXiv:1803.02999 (2018)