

少ないデータによる識別モデルの訓練

MAML (メタ学習)

MAMLが解決したい課題

深層学習モデルの開発に必要なデータ量を削減したい

深層学習モデル開発に必要なデータ量

データセット名	画像の枚数
MNIST	7万枚
ImageNet (ILSVRC2012)	約120万枚
Open Image Dataset V6	約900万枚
MegaFace	約570万枚

- 訓練に必要なデータ量が多い
 - ・ 人手のアノテーションコスト
 - ・ データ自体を準備できるかどうか
 - 少ないデータの問題点
 - ・ 過学習が発生しやすい
- ➔ 少ないデータで学習させたい

MAMLのコンセプト

タスクに共通する重みを学習し、新しいモデルの学習に活用

事前学習

- オートエンコーダー (教師なし学習)
- ImageNetを使った事前学習モデル (教師あり学習)

タスクAのためのモデル

モデル全体の重みを学習

転移学習

タスクBのためのモデル

タスクAの重みを使用 この部分のみ学習

ファインチューニング

タスクBのためのモデル

モデルAの重みを初期値として学習

MAML (メタ学習)

タスク共通の重みを学習

タスクAのためのモデル

共通重みからファインチューニング

タスクBのためのモデル

共通重みからファインチューニング

タスクCのためのモデル

共通重みからファインチューニング

タスクDのためのモデル

共通重みからファインチューニング

MAMLの学習手順

タスクごとの学習を行った結果を共通重みに反映させ学習

Outer Loop

θ が収束するまで繰り返し

共通重み θ をランダムに初期化

タスクの個数分繰り返し

Inner Loop

タスク集合 T からタスク T_i を取り出し

タスク T_i に重み θ を最適化し θ'_i を得る

$(\theta'_1, \theta'_2 \dots)$ を集める

集めた重みで共通重み θ を更新 (SGD)

MAMLの効果

Few-Shot learningで既存手法を上回る精度を実現

Omniplotデータセット(50種類の文字)を使ったクラス分類

C. Finn+, arXiv:1703.03400

	5クラス		20クラス	
	5-way Accuracy		20-way Accuracy	
	1-shot	5-shot	1-shot	5-shot
Omniplot (Lake et al., 2011)				
MANN, no conv (Santoro et al., 2016)	82.8%	94.9%	—	—
MAML, no conv (ours)	89.7 ± 1.1%	97.5 ± 0.6%	—	—
Siamese nets (Koch, 2015)	97.3%	98.4%	88.2%	97.0%
matching nets (Vinyals et al., 2016)	98.1%	98.9%	93.8%	98.5%
neural statistician (Edwards & Storkey, 2017)	98.1%	99.5%	93.2%	98.1%
memory mod. (Kaiser et al., 2017)	98.4%	99.6%	95.0%	98.6%
MAML (ours)	98.7 ± 0.4%	99.9 ± 0.1%	95.8 ± 0.3%	98.9 ± 0.2%

↑ ↑ ↑ ↑
サンプル1枚 サンプル5枚 サンプル1枚 サンプル5枚

この他にも、回帰問題、強化学習などでも効果が確認された

MAMLの課題と対処

タスクごとの学習と共通パラメータの学習で計算量が多い

- MAMLは計算量が多い

- タスクごとの勾配計算と共通パラメータの勾配計算の2回が必要
- 実用的にはInner loopのステップ数を大きくできない

- 計算コストを削減する改良案 (近似方法)

- First-order MAML: 2次以上の勾配を無視し計算コストを大幅低減
- Reptile: Inner loopの逆伝搬を行わず、学習前後のパラメータの差を利用