

進捗報告

1 今週やったこと

- GA の実装と実験

2 実験

2.1 問題

DARTS の初期値依存性やグラフの収束不安定のため GA を使用する。個体をアーキテクチャ α とし, w を共有することで, 学習速度を維持しつつ, 学習の安定を図る。

評価の問題

- 通常: 与えられたパラメータ (個体) でモデルを学習し性能を評価する
- 今回: w を共有するので呼ぶごとに評価が変化する

GA の流れ α, w の学習はアーキテクチャ探索として分離して, 評価では α, w に変化を与えずテストデータのロスを用いる。

- 初期化
- (アーキテクチャ探索)
- 選択
- 交叉・突然変異
- 評価・世代交代

α の個体表現 α は行列であるため, 交叉をどうするかが難しい。初期段階の方法として, 行列を 2 次元配列にして 2 点交叉をした。何が適しているかを調査する必要がある。

2.2 実験設定

表 1, 2 にモデルと GA の設定を示した。前回までの設定ではショートカットを持たない状態で学習を始めるが, GA の場合多様性がなくなるため各層に 1 本ずつ持たせる設定で学習した。

表 1: モデルの設定

base model	VGG19
Optim(w)	SGD(lr=0.001, momentum=0.9)
Optim(α)	Adam(lr=0.003, $\beta=(0.5, 0.999)$)
Loss	Cross Entropy Loss
dataset	cifar10
pretrain	true
batch size	64
train size	2500
valid size	2500

表 2: GA の設定

個体数	10
世代数	10
選択	トーナメント
サイズ	3
交叉	2 点交叉
交叉率	0.5
変異	ガウス分布
変異率 (遺伝子座ごと)	0.2 0.1

2.3 結果

図 1, 2 に GA の結果の精度とロスを示した。ただしこの結果は α から作成したモデルの性能ではないため, 本来の値とは異なる。図では精度は世代ごとに向上するが, 2 世代から損失は悪化した。モデルが出力する確信度が平均的な α に最適化された結果, 逆にすべての α との差が損失に大きな誤差を与えたと思われる。

今回の設定の場合, 1 世代にかかる時間は 3 分であった。データをすべて使う場合, 1 個体あたり 3 分となり, 20 個体の場合 1 世代に 1 時間かかる。

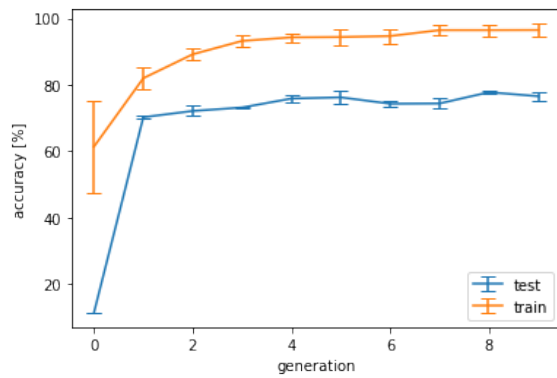


図 1: 世代ごとの精度 (平均と標準偏差)

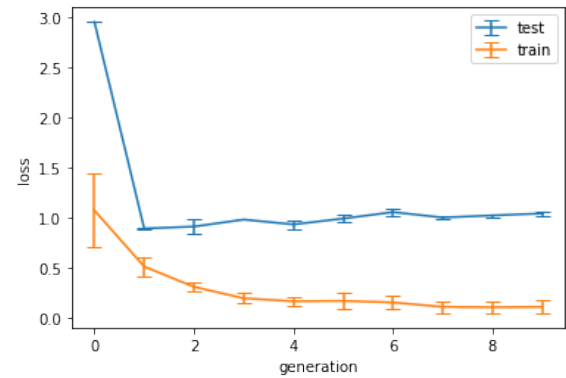


図 2: 世代ごとのロス (平均と標準偏差)

3 今後の予定

まず GA はメモリの問題もなく動いた. しかし交叉や突然変異などは適した手法が分からないため, GA を調査して設定を見直したい. うまくいけば来週サーバーで長時間動かす.

4 ソースコード

github の notebook リポジトリ参照.