

進捗報告

1 今週やったこと

GA を用いた DataAugmentaion

2 実験

2.1 実験データ

実験データは cifar10 を用いて、事前学習では epoch 数 150, train_data を各ラベル 4000 枚の計 40000 枚使用し、GA で学習する際は epoch 数 30, train_data は各ラベル 200 枚のオリジナルとそれらすべてを DataAugmentaion したものとを合わせ計 4000 枚とし、test_data 及び validation_data は共に 10000 枚とした。また事前学習での accuracy は 0.8281 である。

2.2 分散遺伝的アルゴリズム

多様性維持として分散遺伝的アルゴリズム (Distributed Genetic Algorithm:DGA)[1] を用いる。

2.2.1 DGA の概要

まず、GA の母集団を複数に分割をする。この分割されたものをサブ母集団といい、GA の遺伝子操作をサブ母集団ごとに施すことで、サブ母集団同士で多様性を維持することができる。

一方でサブ母集団は母集団よりも個体数が少なくなるため、初期収束が起こりやすくなる。そこで探索情報を共有するために次項で説明する移住 (Migration) という操作を用いる。

2.2.2 移住

移住はある一定の間隔でサブ母集団内の一部の個体を他のサブ母集団内の一部の個体と交換する操作である。移住の間隔を移住間隔 (Migration interval)、交換する個体数の割合を移住率 (Migration rate) という。移住は GA の選択、交叉、突然変異という操作のうち、選択と交叉の間で行われるものとした。また、今回移住先の制約として、移住元と移住先を

すべてつなげた時にリング状のグラフになるようにした。また移住個体はランダムに選択するものとした。

2.2.3 個体

これまでと同様 16 の操作に対し強度、確率、順序をもった個体とする。

2.2.4 選択

選択について、エリート選出によって最も適応度の高い 2 つの個体を選択する。なお、この二つは後述する交叉、突然変異は受けずに次の世代に追加する。残りの選出にはトーナメント選出を用いた。トーナメント選出は集団の中から任意の数 (トーナメントサイズ) の個体のうち最も適応度の高い個体を選出し次の世代に追加する。今回トーナメントサイズは 2 とした。

2.2.5 交叉

強度、確率を表す染色体については 2 点交叉、順序を表す染色体については部分写像交叉を用いた。2 点交叉は一对の親染色体をそれぞれ同じ場所で三分割し中央の染色体を入れ替えて交叉を行う

2.2.6 突然変異

強度、確率を表す染色体について、対象となる遺伝子の値を各 50% の確率で 1 増減させ、順序を表す染色体について、染色体の一部を逆順にする操作か、染色体を二つに分け前後を入れ替える操作のいずれかを行うものとした。

2.3 適応度

validation_data について target となる個体から得られた出力と、その個体の属するサブ母集団以外の個体全ての出力で比較し、target で正答したものの中で、他の個体で誤答した問題数をカウントして適応度とした。

2.4 アンサンブル学習

今回はサブ母集団ごとにもっとも適応度の高い個体を選び、それらをアンサンブルさせた。

2.5 実験

2.5.1 パラメータ

表 1 に学習パラメータを示す。表 2 に GA の設定

表 1: 学習パラメータ

optimizer	Adam
learning rate	0.001
loss function	categorical_crossentropy
batch size	128
epoch size	30

を示す。

表 2: 実験パラメータ

サブ母集団	個体数	10
	個数	10
総個体数		100
移住間隔		3 世代ごと
移住個体数		3
世代数		26
交叉率		0.9
突然変異率		
強度, 確率 (遺伝子ごと)		0.06
順序 (染色体ごと)		0.1

2.5.2 結果

図 1, 1 に 26 世代目の遺伝子座を tSNE で次元圧縮したものを示す。また, 図 2, 3, に val と test の推移を示す

また, train_data を全 4000 枚使用したものについて test の accuracy は 1 世代目が 0.8901, 26 世代目が 0.8947 となった。

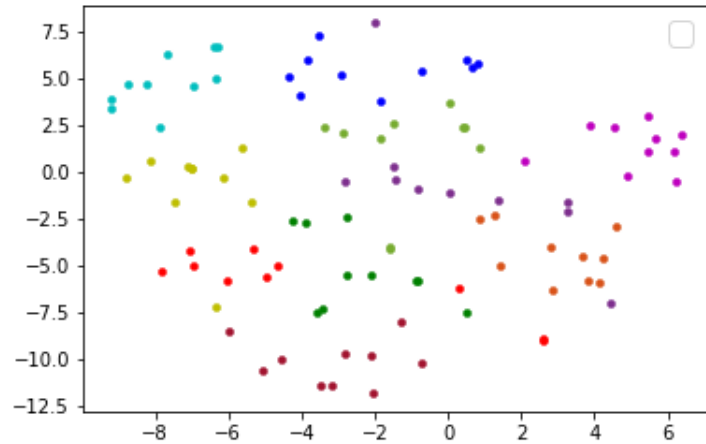


図 1: 26 世代目の遺伝子座に対する tSNE 図

2.6 まとめ

アンサンブル学習自体はうまくいっている気もするが, 初期世代と最終世代であまり大差がない。つまり, アンサンブルのなす多様性が DataAugmentation ではなく学習時に絞ったデータの分配に起因していると考えられる。DataAugmentation の違いでアンサンブルをすることの有用性を確認するためには学習のデータもすべて揃えなければならないのでは思った。もしくは本当に現在の操作だけではアンサンブルに寄与するほどのことではないかである。

3 来週の課題

- 本当に DataAugmentation がアンサンブルに有効か確認する
- 絵本についてのリサーチを進める

参考文献

- [1] 廣安知之, 三木光範, 上浦二郎, et al. 実験計画法を用いた分散遺伝的アルゴリズムのパラメータ推定. 情報処理学会論文誌数理モデル化と応用 (TOM), 43(SIG10 (TOM7)):199–217, 2002.

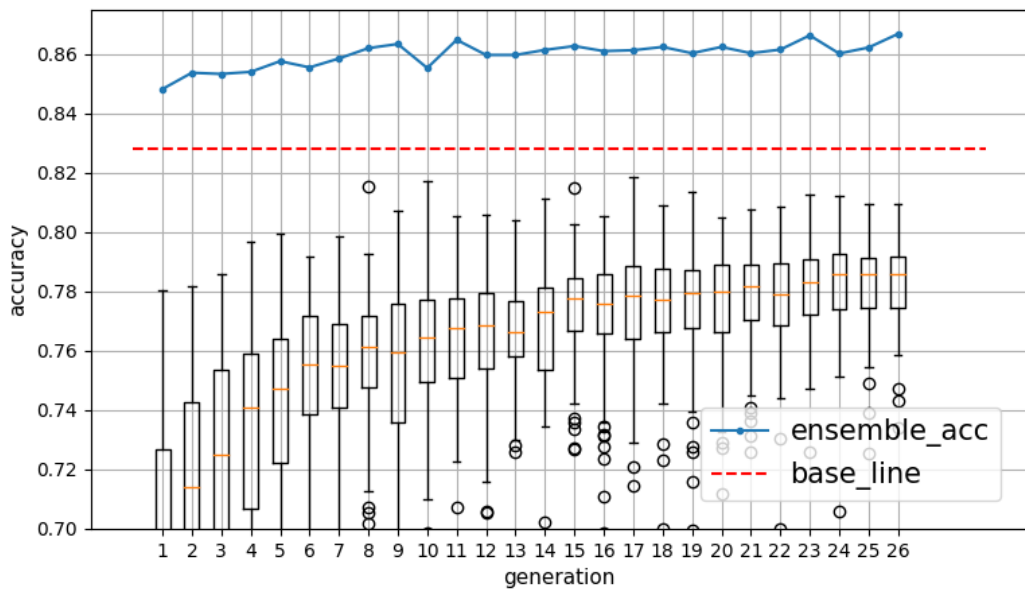


図 2: val の推移

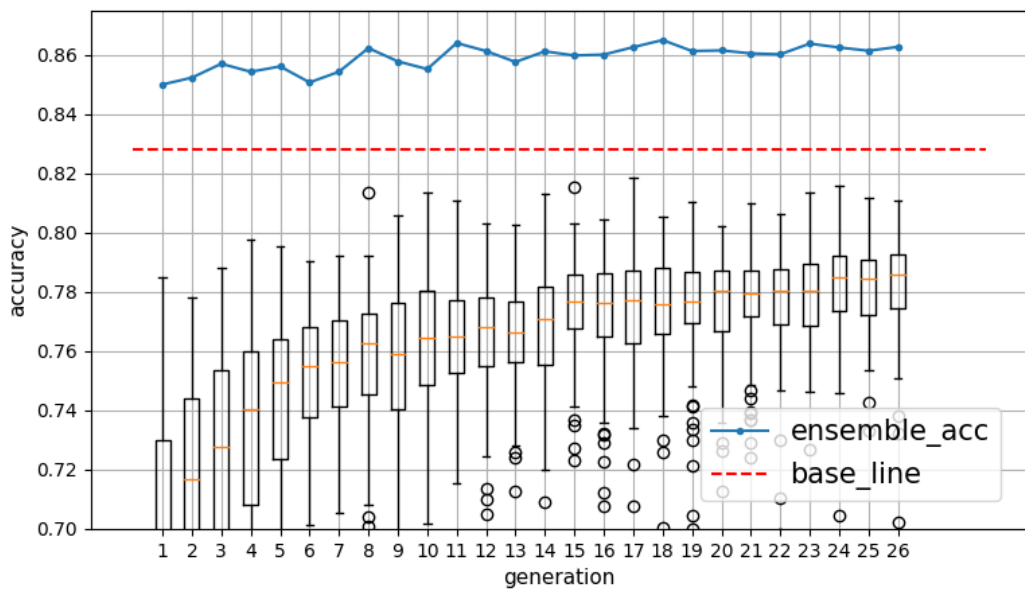


図 3: test の推移