

進捗報告

1 今週やったこと

- 前回の実験の正答率についての視覚化
- GA を用いた DataAugmentaion

2 前回の実験の正答率の視覚化

前回の実験では DGA を用いて実験した．サブ母集団 10 個に対し 10 個体とし，全体の個体として 100 個体を 25 世代回した．図 1, 2, 3, 4 に正答率を示す．このとき，時間短縮のため用いた学習済みモデルの正誤も共に示した．

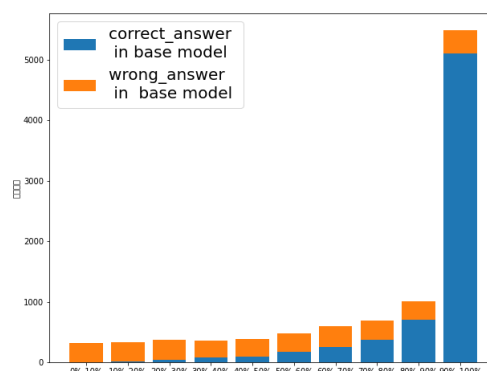


図 3: 最終世代における val 画像の正答率

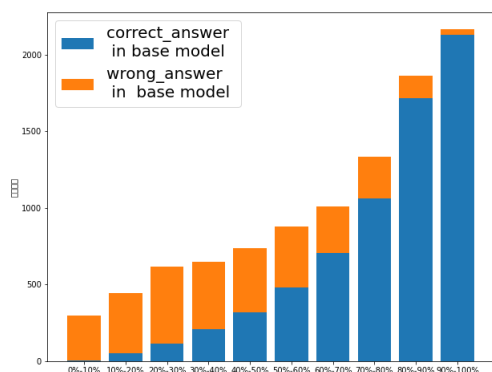


図 1: 初期個体における val 画像の正答率

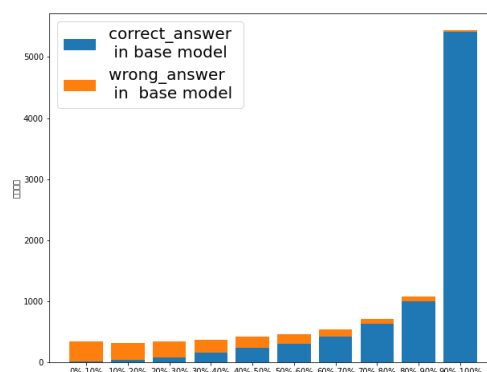


図 4: 最終世代における test 画像の正答率

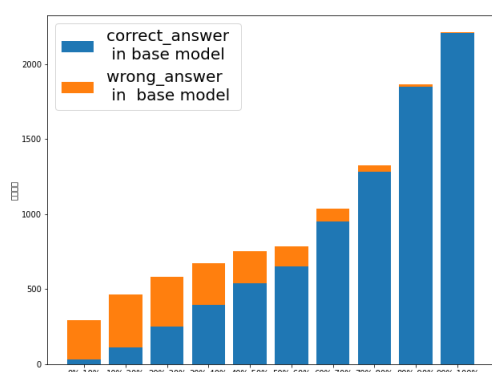


図 2: 初期個体における test 画像の正答率

3 実験

3.1 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，今回はオリジナルの 2000 枚は全ての個体で共通のものとした．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．

3.2 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.

3.2.1 移住

移住間隔を3世代ごとに, 移住個体数を各島2個体とした

3.2.2 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.

3.2.3 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サイズが2のトーナメント選出を用いた.

3.2.4 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

3.2.5 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%の確率に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.

3.3 適応度

前回言われていた通り, 他の島との比較を使わない, 正答にくい問題に重きを置くということを踏まえ, 事前学習における誤答した問題において個体の学習で新たに正答した数を適応度とした.

3.4 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.

3.5 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 | 6 |
| | 個数 | 6 |
| 総個体数 | | 36 |
| 移住間隔 | | 3世代ごと |
| 移住個体数 | | 2 |
| 世代数 | | 32 |
| 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 |

3.6 結果

図5, 6に結果の推移を示す. 前回よりもGA中のアンサンブル学習の精度が3%落ちていたが, これに関してまでtrain_dataが40000枚からランダムに2000枚選ばれていたのに対し, 今回は2000枚を固定したのでその差が現れたのだと思われる. といっても, base_lineと同じぐらいになってしまった理由として, base_lineは40000枚で学習していた分アンサンブルで差が埋まったのではないかと考えられる. 一方で, 得られた個体に対しtrain_data全40000枚を用いて学習したものについて, 0世代目は0.9009で, 32世代目は0.911であり, これは前回の実験での最終世代が0.8947であったため, 前回よりも適応

度の取り方が良いのではないかと考えられる。しかし，accuracy として1%しか上がっていないため，まだまだ適応度の決め方に気を配る必要がある。

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.

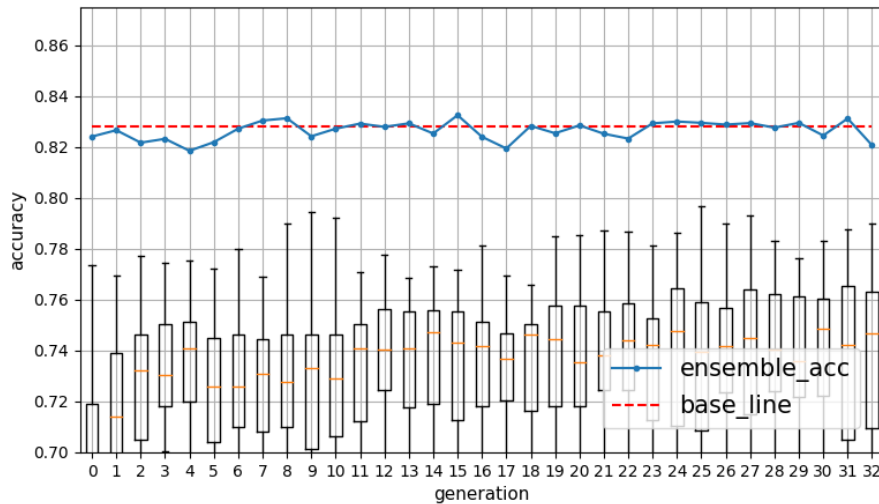


図 5: val の推移

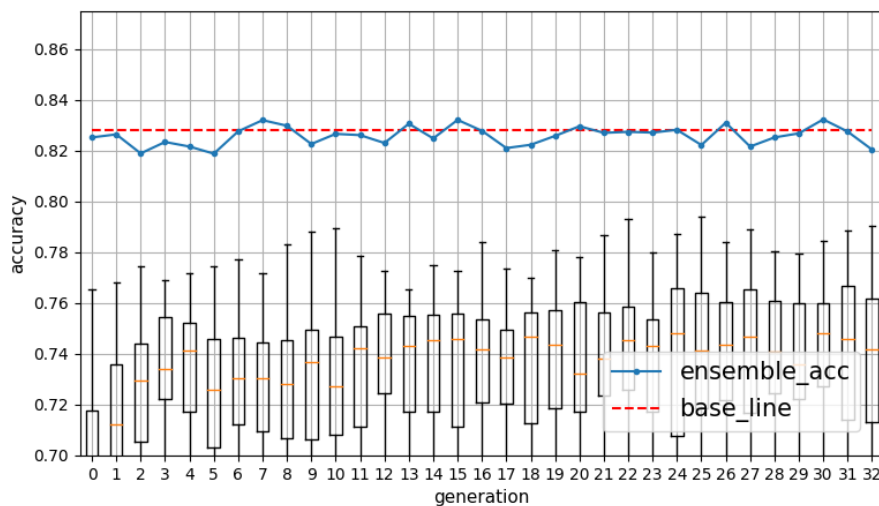


図 6: test の推移

4 来週の課題

- GA の適応度の取り方を工夫する