

## 進捗報告

### 1 今週やったこと

GA を用いた DataAugmentaion

### 2 実験 1

前回に引き続き GA を用いたアンサンブル学習のための DataAugmentation の実験を行った。

#### 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 300, train\_data を各ラベル 5000 枚の計 50000 枚使用し、GA で学習する際は epoch 数 100, 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 は共に 10000 枚とした。また事前学習での accuracy は 0.8475 である。

#### 2.2 遺伝的アルゴリズム

##### 2.2.1 探索空間

探索する水増し操作として画素値操作 (Sharpness, Posterize, Brightness, Autocontrast, Equalize, Solarize, Invert, Contrast, ColorBalance), 変形操作 (Mirror, Flip, Translate X/Y, Shear X/Y, Rotate) の 16 種類の操作であり、今回はそれらすべてを個別にどの程度強くかけるかおよびどの順序でかけるかということを探索する。各操作についての強度の最大最小を設定し、それを -100% から 100% まで 25% ずつ分 11 段階の度合いとする。ただし、Autocontrast, Equalize, Invert, Mirror については適用するか否かであるためパラメータが 0 以上で適用するとした。強度は 0 から 5 の整数値を持つ 15 個の遺伝子を実数値コーディングによって表現する。また、適用順序に関しては同様に 15 個の遺伝子を持つ順列コーディングによって表現する。確率は 10% ごと 11 段階の実数地コーディングによって表現する。つまり、探索空間は  $2^5 * 11^{11} * 15! * 11^{16}$  となる。

##### 2.2.2 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。

##### 2.2.3 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。部分写像交叉は親遺伝子を二分割し入れ替える際重複をなくす交叉法で、重複のあった遺伝子について、それに該当した重複する遺伝子座を見つけ、それに対してとなっているもう一方の親の遺伝子を参照する。

##### 2.2.4 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。

##### 2.2.5 多様性維持

多様性を維持するために、上記 3 つの操作 (選択、交叉、突然変異) を行った集団に対し、適用順序を表す染色体について一致するものが 3 つ以上あれば、それが 2 つになるように一部の個体を突然変異させたうえで次の世代の集団とした。

##### 2.2.6 適応度

前回までは適応度を各個体の accuracy としたために、各個体同氏が似通ったものとなりアンサンブルしたものはあまり改善が見られなかった。そこで、個体

$i$  の予測値の集合を  $pred(i)$ , accuracy を  $acc(pred(i))$  とし, 予測値の集合の集合  $A$  に対するアンサンブルによる accuracy を  $ens\_acc(A)$  とすると,

$$fitness_i = \alpha * \left( \frac{\sum_{A \subseteq U | A \text{ have } pred(i)} ens\_acc(A)}{2^{n-1}} - acc(pred(i)) \right) + acc(pred(i))$$

とした. 今回は  $\alpha = 0.9, 1, 1.1, 1.5$  について行った.

## 2.3 実験

### 2.3.1 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

個体数	15
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

### 2.3.2 結果

, 図 1, 図 2, 図 3, 図 4 に accuracy の最良値及び平均値の推移を示す.

表 3, 表 4 にアンサンブル学習の最良値を示す.

図より  $\alpha > 1$  のとき個体自体のばらつきは広がっているが accuracy が全体的に下がっている.  $\alpha = 0.9, 1$  のときは前回の最良値が 0.8588 であったことに比べると精度は上がった.

表 3:  $\alpha = 1$ , 13 世代目

1st	0.8605	2nd	0.7899
3rd	0.8512	4th	0.8560
5th	0.8547	6th	0.8513
7th	0.8574	8th	0.8587
9th	0.8462	10th	0.8574
11st	0.8440	12nd	0.8543
13rd	0.8487	14th	0.8487
15th	0.8556		
ensemble		0.8828	

表 4:  $\alpha = 0.9$ , 15 世代目

1st	0.8553	2nd	0.8618
3rd	0.8553	4th	0.8598
5th	0.8527	6th	0.8599
7th	0.8526	8th	0.8557
9th	0.8569	10th	0.8604
11st	0.8555	12nd	0.8526
13rd	0.8513	14th	0.8570
15th	0.8560		
ensemble		0.8846	

$\alpha = 1$  について得られた個体から元のデータ数 50000 枚を倍に水増ししたものについて同様にアンサンブル学習を行った. 表 5 に結果を示す. 前回の実験である適応度が各個体の accuracy だけだったものの結果が 0.8732 であったので改善はされているが, 前々回のアンサンブル学習を目的としなかった実験ではアンサンブルの結果は 0.9205 であり, それと比べるとあまり良い結果とはいえない.

## 2.4 まとめ

今回と前回の結果から, アンサンブル学習を目的とし多様性を考慮した場合, 各個体だけの精度を考えるだけでなく, アンサンブル学習にどの程度寄与するかも考えて適応度関数を決定したほうが良い. 今回と前々回から, 当然ではあるがアンサンブル学習に使う個体のうち精度の最大値は高いほうが良い

表 5: 最終結果

1st	0.8723	2nd	0.8769
3rd	0.8787	4th	0.8778
5th	0.8787	6th	0.8727
7th	0.8782	8th	0.8719
9th	0.8521	10th	0.8684
11st	0.8738	12nd	0.8817
13rd	0.8809	14th	0.8755
15th	0.8755		
ensemble		0.9048	

が、それに付随する他の個体の精度は少し低くてもよい。

したがって、多様性を考えずに個々の精度を上げる集団とそれに平行して、多様性を考えてかつ前者の集団を含めたアンサンブル学習の精度を上げる集団との二つの集団で同時に学習を行えば多少うまくいくのではないかと思った。

### 3 来週の課題

- 前期発表の資料の仕上げ
- 絵本についてのリサーチを進める

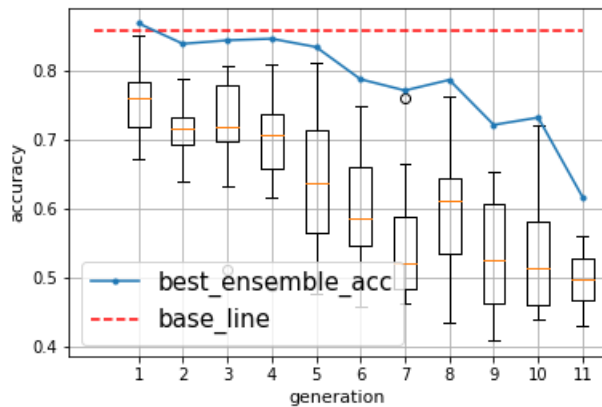


図 1:  $\alpha = 1.5$  の推移

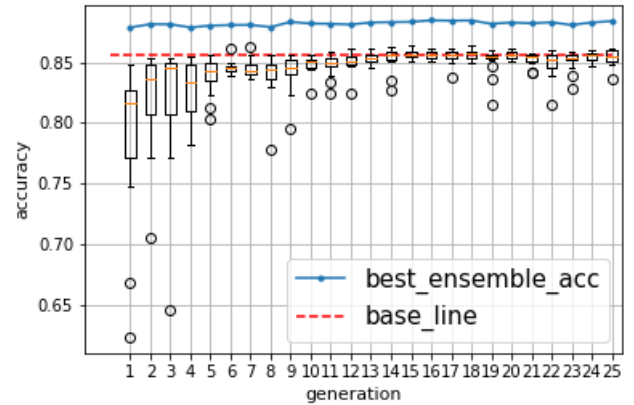


図 2:  $\alpha = 0.9$  の推移

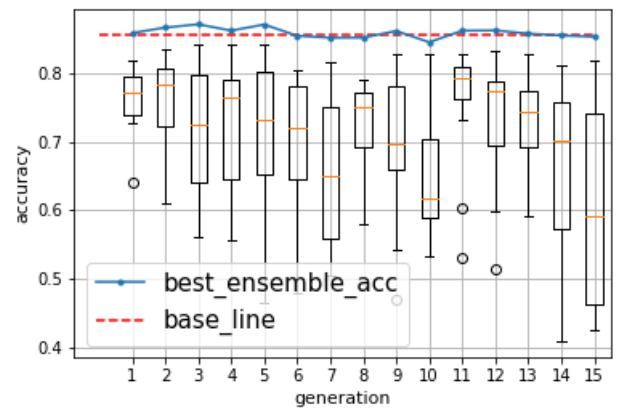


図 3:  $\alpha = 1.1$  の推移

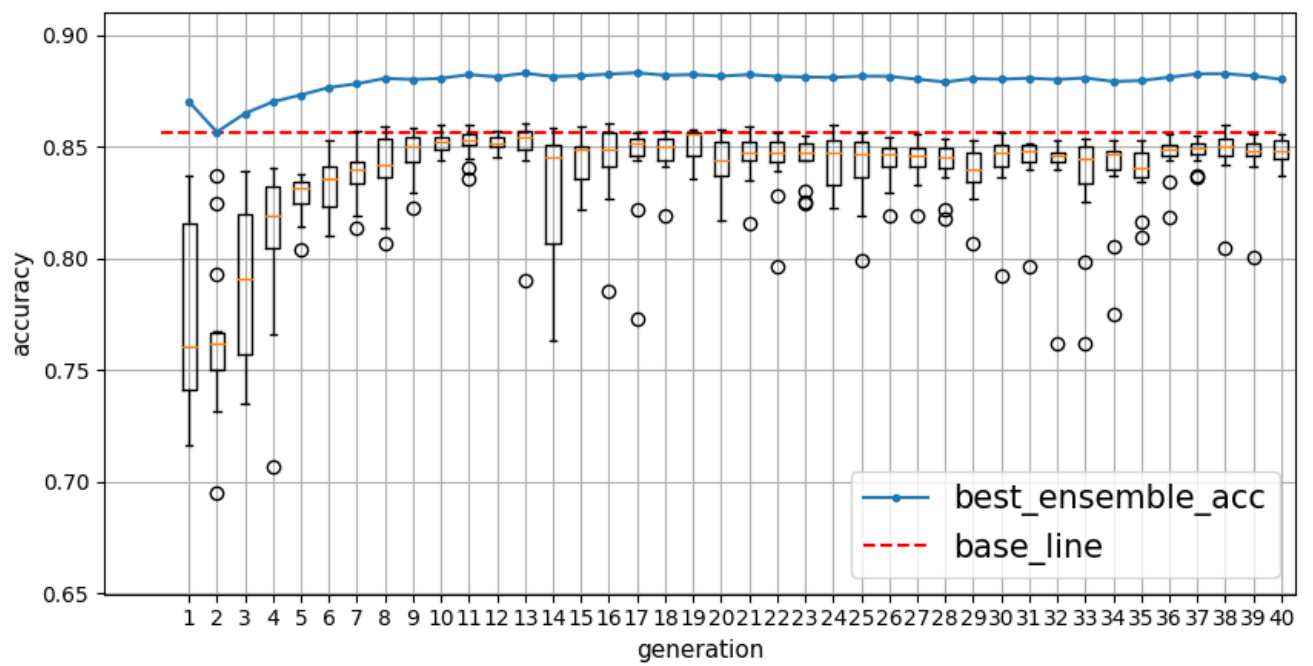


図 4:  $\alpha = 1$  の推移