

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, G A 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 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査する. 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 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.

2.2.4 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 3 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。

2.2.6 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 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

サブ母集団	個体数	7
	個数	5
総個体数		35
世代数		29
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, にアンサンブル学習の accuracy の推移を示す

図 2, に各個体の accuracy の推移を示す

.

2.4 まとめ

今回の実験は前回に比べ train_data が少ないがために全体的には精度が下がっているのは明らかではある。サブ母集団の個体が 7 つに対し、DGA や突然変異率の設定が甘かったせいか、初期収束を起こし、個々の val_accuracy が base_line を超えていない。また、アンサンブル学習で val が上がっているのに対し、test が全く上がっていないのが気になる。val が全探索なのに対し、test は val の最良時に使われた個体という 1 つの組み合わせしか見ていない。また、val に合わせて fitness が決定しているので上昇しているのはわかるが、どちらも弱学習器の学習には使われていないはずであるので、val と test のデータが相性の悪いものとなっている可能性がある。

3 来週の課題

- DGA の設定を変えて実験を行う
- 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.

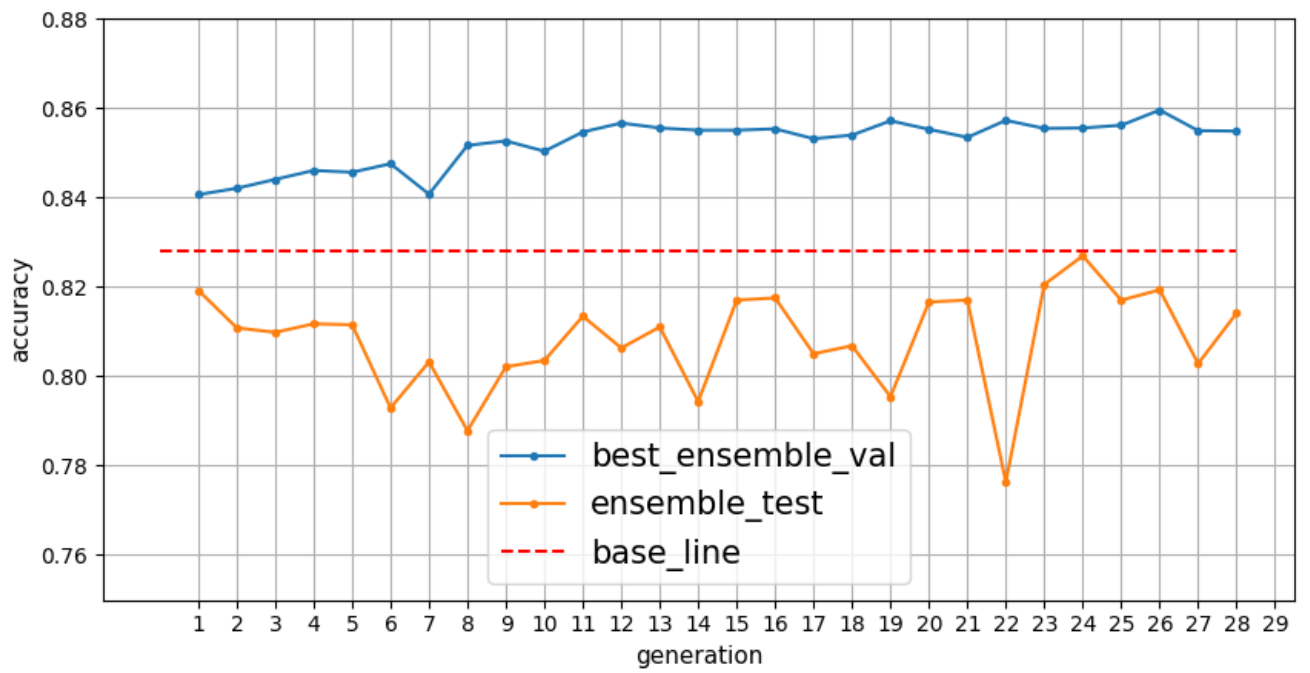


図 1: ensemble_accuracy の推移

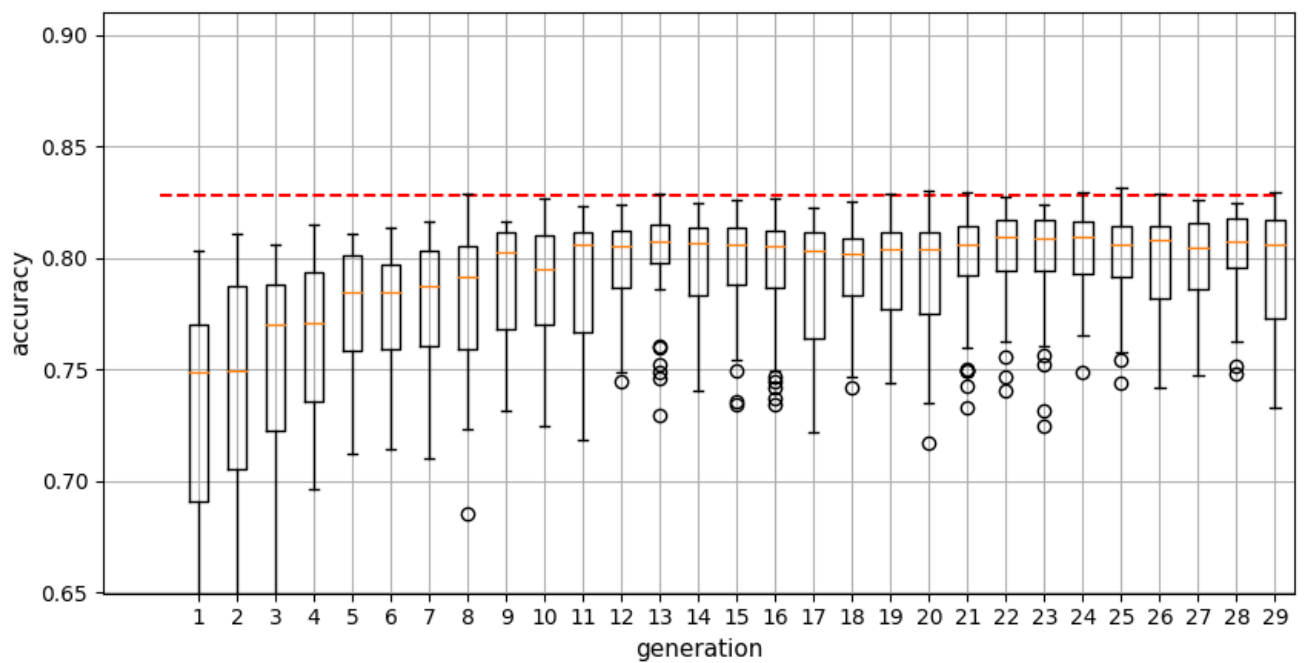


図 2: 各個体の val_accuracy の推移