

進捗報告

1 今週やったこと

- GA を用いた DataAugmentaion
- SSL の論文読み

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, 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て二つのデータセットを用意した. 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.

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 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

2.2.2 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.

2.2.3 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.

2.2.4 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

2.2.5 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.

2.3 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, 正答し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での学習で新たに正答した数を適応度とした.

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 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		17
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

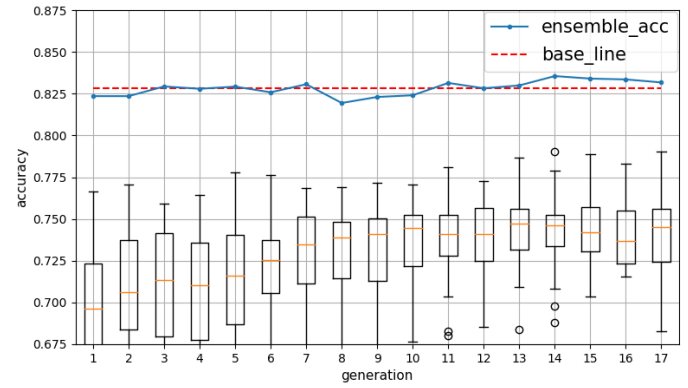


図 2: data_set2 の val の推移

2.6 結果

図 1, 2 に結果の推移を示す. 前回とは別の固定されたデータを用いても結果としてあまり変わらないものとなり, 固定したデータが test や val のデータと相性が悪いということにはなかったと考えられる. また前回と同じ条件のため初期収束ぎみである.

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.

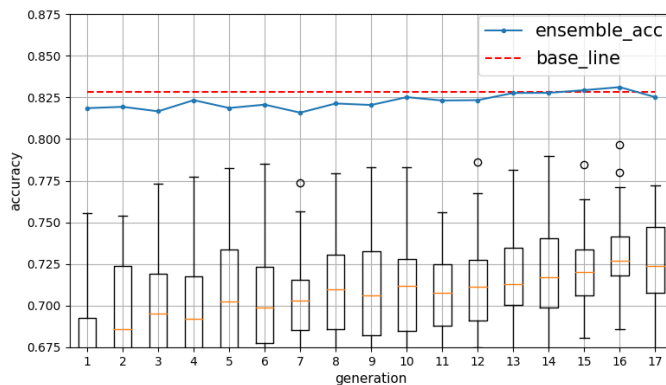


図 1: data_set1 の val の推移

3 来週の課題

- GA を用いた SSL のラベル付けの実験

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