

## 進捗報告

## 1 今週やったこと

- Fixmatch と GA を組み合わせた実験

## 2 FixMatch + GA

ラベルなし画像 100 枚取り出し、それらに対するラベルを GA によって探索する。表 1, 2 に実験の設定を示す。表 1 での精度を GA の目的関数とした。

遺伝子は 0 から 9 の整数値をとる整数値コーディングとした。

選択はサイズ 2 のトーナメント選択、交叉には二点交叉、突然変異は別の数値にランダムに移るように設定した。

また、事前学習を表 3 の設定で行い、精度 0.55(エラー率:0.45) であり、初期個体について、事前学習から予測したラベルに対し、各遺伝子座にエラー率の割合で突然変異させたものとした。

表 1: FixMatch の設定

model	WideResNet16-2	
data set	cifar10	
train data	labeled	100
	unlabeled	49650
	search	100
batch size	labeled+search	64
	unlabeled	64 * 7
val data	150	
num_iterations	10000	
optimizer	SGD(lr=0.1,momntum=0.9)	
loss	cross_entropy_loss	

表 2: GA の設定

個体数	5
世代数	8
交叉率	1.0
突然変異率	0.06

表 3: 事前学習の設定

model	WideResNet16-2	
data set	cifar10	
train data	labeled	50
	unlabeled	49750
	batch size	64
	unlabeled	64 * 7
val data	200	
num_iterations	10000	
optimizer	SGD(lr=0.1,momntum=0.9)	
loss	cross_entropy_loss	

## 2.1 結果

表 4, 図 1 に結果を示す。

表 4: 結果

	適応度		正答数	
	最大	平均	最大	平均
1 世代	0.167	0.128	18	15.4
2 世代	0.167	0.120	17	15.8
3 世代	0.167	0.121	17	14.2
4 世代	0.167	0.124	17	13.8
5 世代	0.167	0.115	16	14.4
6 世代	0.180	0.131	16	13.6
7 世代	0.180	0.131	15	13.0
8 世代	0.180	0.136	14	12.6

時間の都合上個体数、世代数共に低かったためあまり探索できていない。また図 1 より、相関関係があまり成り立っていないことが分かる。また、1 個体の学習について 1 時間半ほど要した。

## 2.2 考察

出鱈目なラベルの割合が多いと精度との相関関係が薄くなる可能性がある。元々のデータ数に対して search するデータ数の割合を小さくする必要がある。

一世代あたりの個体数が小さいために結果は悪い方

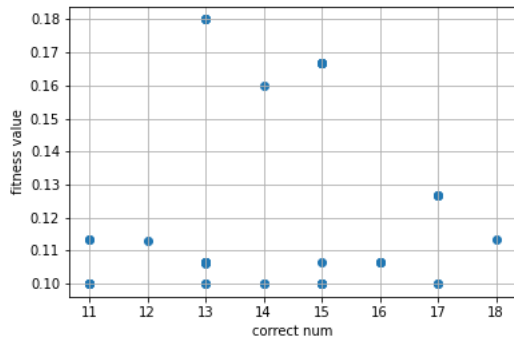


図 1: 散布図

向に収束していったと考えられる。あるいは，トーナメント選択で適応度の高いものが選ばれなかったことがあった．そのため適応度の低いものに制限をかけることも必要なのではないかと考えられる．また選ばれたデータが悪いことを考えると，train と val のデータを世代ごとに入れ替えて，交叉検証のような評価をする必要があるのではないかと思う．

### 3 来週の課題

- 間違っラベルの割合と精度の関係について調べパラメータの調整をする．
- cnn や randomforest といった時間短縮につながる手法を組み込み実験を行う．