

進捗報告

1 今週やったこと

- 予測精度の実験
- GA の実験

2 精度とランダムラベルの割合

前回のランダムラベルでの実験について再度検証を行った．表 1 に設定を示す．

このラベル付き 500 枚のうち一部のラベルを別のものに変更して実験を行った．各割合に対し，5 回ずつ実験を行った．

表 1: FixMatch の設定

model	WideResNet28-2
data set	cifar10
labeled data	500
unlabeled data	49500
batch(labeled)	32
batch(unlabeled)	32 * 7
val data	10000
num_iterations	15000
optimizer	SGD(lr=0.1,momntum=0.9)
loss	cross_entropy_loss

2.1 結果

表 2 に結果を示す．また前回の実験ではコードの一部にミスがあり修正を行ったため，今回の結果が直観に沿ったものになったことが分かる．またこのことから，前々回 FixMatch で GA の結果が出ないと言っていたが，時間的な問題を除けば学習することはできそうではある．

3 GA の実験

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 3 に実験の設定を示す．目的関数は SVM 及び RandomForest の精度

表 2: 結果

random	accuracy
0	0.7764±0.0167
1	0.7651±0.0162
2	0.7663±0.0180
4	0.7553±0.0163
8	0.7554±0.0151
16	0.7410±0.0342
32	0.7375±0.0179
64	0.6731±0.0009
128	0.5710±0.0305
256	0.4458±0.0135

とした．また，表 4, 5 に optuna で最適化したパラメータを示す．また共に 5 交差検証で 200 回まわした．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 train と eval を世代ごとに分割している．

表 3: GA の設定

個体数	30
世代数	50
交叉率	1.0
突然変異率	0.02
全ラベル付き画像	250 枚
train:eval	100 枚:150 枚
search	100 枚

3.1 結果

図 3, 4 に結果を示す．また，指標として表 6 にラベル付き画像のみでの精度と探索した画像に正解ラベルを付与して学習したときの精度を示す．

前回のものと比べあまり GA による改善が見られない．また，図 1, 2 に RandomForest の時の散布図を示す．このことから各々の精度は良くなっている

表 4: SVM のパラメータ

探索		best
kernel	rbf, sigmoid, poly	poly
C	$1 - 10^2$	3.654
γ	$10^{-3} - 10^2$	53.15
固定		
coef0		0.0
degree		3

表 5: RandomForest のパラメータ

探索		best
n_estimators	$5 - 10^2$	64
max_features	$1 - 32$	22
max_depth	$2 - 16$	8
criterion	gini, entropy	entropy
固定		
mini_samples_split		2
mini_samples_leaf		1

が相関性があまり見られない。そのため、optuna で最適化する際のスコアについて、精度ではなく相関係数など別の指標をする必要があると考えられる。また、SVM や RandomForest ではやはり精度が頭打ちになってしまうので時間効率も考えたうえで実験する必要がある。

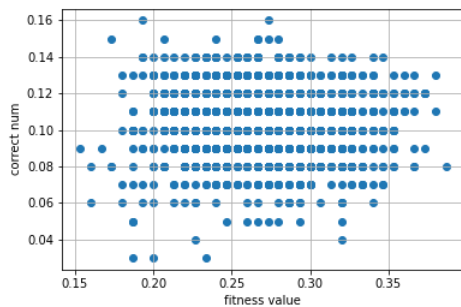


図 1: SVM における散布図

表 6: 精度

	SVM	RandomForest
ラベル付きのみ	0.200	0.273
ラベル付き+正解ラベル	0.267	0.287

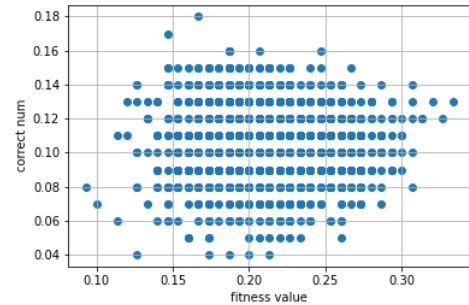


図 2: RandomForest における散布図

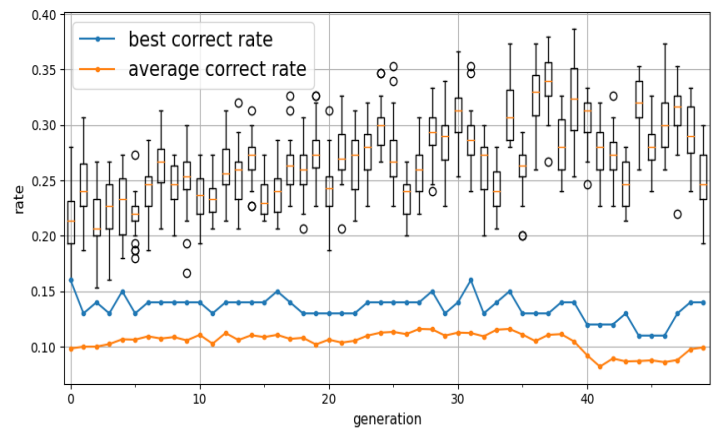


図 3: SVM 結果

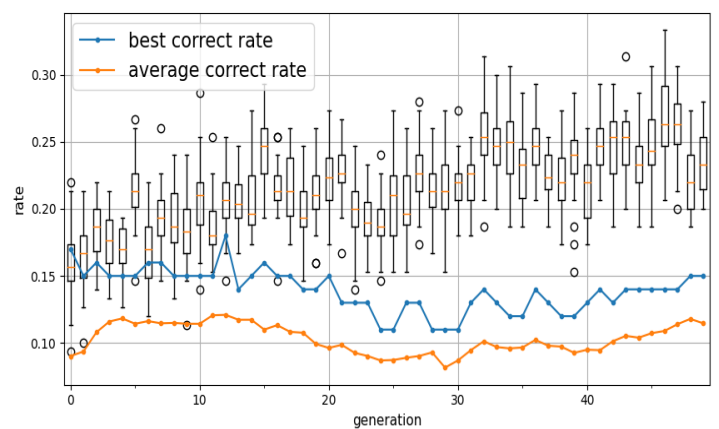


図 4: RandomForest 結果

4 来週の課題

- FixMatch の時間短縮 (model の縮小, 最終層のみの学習など)