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。また、この設定は先週の実験でoptunaで探索したものである

. 各割合に対し、50回ずつ実験を行った.

表 1: SVM の設定

kernel	poly
\mathbf{C}	3.654
γ	53.15
coef0	0.0
degree	3
dataset	cifar10
train	200
test	200

2.1 結果

表2に結果を示す. 結果としてはランダムなものが少ないほど精度は高く, かなり妥当なものであることが分かる.

表 2: 結果

random	accuracy
0	0.2538 ± 0.0344
1	0.2523 ± 0.0344
2	$0.2526 {\pm} 0.0364$
4	0.2514 ± 0.0314
8	0.2510 ± 0.0300
16	0.2390 ± 0.0323
32	0.2245 ± 0.0329
64	0.1989 ± 0.0286
128	0.1499 ± 0.0309

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.SVM のハイパーパラメータは optuna で最 適化した.今回はランダムなラベルの数と精度に対 する相関係数を目的関数とした.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.

前回との変更点として、データの前処理を正規化から標準化にした。また、前回 train と eval の分け方を世代ごとに変えるということをしていたが、それだけでは世代ごとの適応度のばらつきがあったので、今回はさらに5回の試行の平均を適応度とした。

表 3: GA の設定

個体数	30
世代数	100
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 実験 2.1

最適化関数の相関係数を求めるにあたり、250 枚 のラベル付き画像を train:test を 150 枚:100 枚に分け、train のうちランダムなラベル数を 0, 1, 4, 9, 16, 25, 36, 49, 64, 128 と変化させ、各 20 回ずつ試行した. こうして得られた精度とそれに対するランダムなラベル数の相関係数を最適化関数とした.また optuna の試行回数は 200 回とした.表 4 に探索したパラメータ及びその時の相関係数を示す.

3.2 実験 2.2

最適化関数の相関係数を求めるにあたり, 250 枚のラベル付き画像を train:test を 125 枚:125 枚に分

表 4: 実験 2.1 の SVM のパラメータ

	探索	best
kernel	rbf, sigmoid, poly, linear	poly
С	10^{-1} - 10^3	799.7
γ	10^{-4} - 10^2	$9.494*10^{-2}$
coef0	10^{-1} - 10	0.9060
固定		
degree		3
	相関係数	0.6833

け、train のうちランダムなラベル数を 100 から 119 と変化させ、各 20 回ずつ試行した. こうして得られた精度とそれに対するランダムなラベル数の相関係数を最適化関数とした. また optuna の試行回数は 200 回とした. 表 5 に探索したパラメータ及びその時の相関係数を示す.

表 5: 実験 2.2 の SVM のパラメータ

探索		best
kernel	rbf, sigmoid, poly, linear	linear
С	10^{-1} - 10^3	821.3
γ	10^{-4} - 10^2	$2.285*10^{-4}$
coef0	10^{-1} - 10	1.308
固定		
degree		3
	相関係数	0.3265

3.3 結果

図1,2に結果を示す.前回に比べ適応度のばらつきは押さえられていた.また,相関係数が示すように SVM で全域的には相関関係があり,ランダムなラベルが多い局所的な箇所では相関関係がないといえる.

4 来週の課題

• GA を FixMatch を用いて探索中

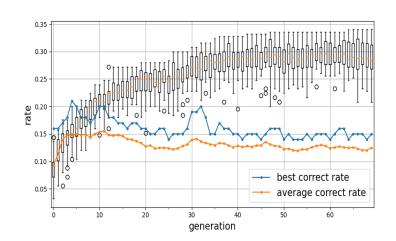


図 1: 実験 2.1 の結果

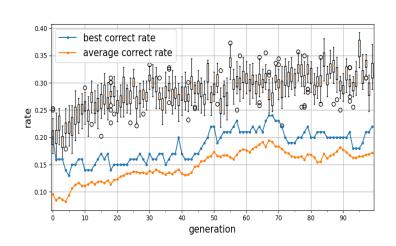


図 2: 実験 2.2 の結果