

進捗報告

水野泰旭

F1 スコア

前回の問題点

検証データ 20%の混同行列

Macro-F1 スコアの再計算

データ数を揃えて学習

ソースコード

訓練データ数の推移

学習結果の比較

進捗報告

水野泰旭

弘前大学理工学部電子情報工学科 4 年

December 19, 2022

目次

進捗報告

水野泰旭

F1 スコア

前回の問題点

検証データ 20%の混同行列

Macro-F1 スコアの再計算

データ数を揃えて学習

ソースコード

訓練データ数の推移

学習結果の比較

- ① F1 スコア
 - 前回の問題点
 - 検証データ 20%の混同行列
 - Macro-F1 スコアの再計算
- ② データ数を揃えて学習
 - ソースコード
 - 訓練データ数の推移
 - 学習結果の比較

前回の問題点

進捗報告

水野泰旭

F1 スコア

前回の問題点

検証データ 20%の混同行列

Macro-F1 スコアの再計算

データ数を揃えて学習

ソースコード

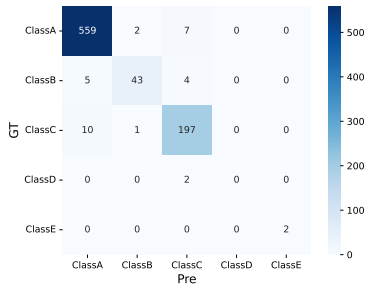
訓練データ数の推移

学習結果の比較

クラス D の TP と FP が共にゼロとなり、

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

の値を取る適合率がゼロ除算となってしまう。現在は訓練データと検証データは 9:1 なので、8:2 として学習してみる。



検証データ 20%の混同行列

進捗報告

水野泰旭

F1 スコア

前回の問題点

検証データ 20%の混同行列

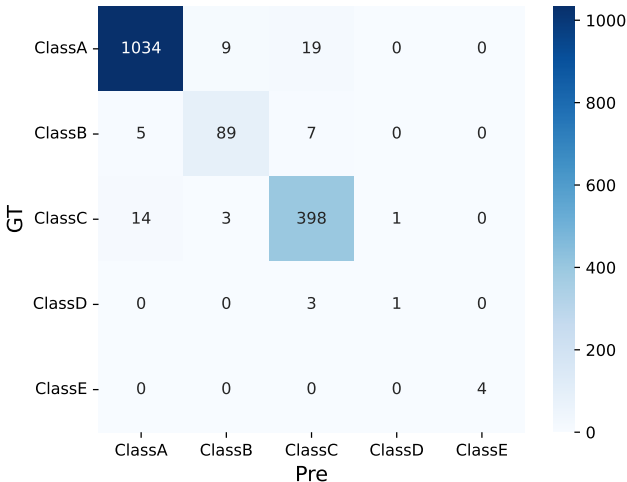
Macro-F1 スコアの再計算

データ数を揃えて学習

ソースコード

訓練データ数の推移

学習結果の比較



Macro-F1 スコアの再計算

進捗報告

水野泰旭

F1 スコア

前回の問題点

検証データ 20%の混同行列

Macro-F1 スコアの再計算

データ数を揃えて学習

ソースコード

訓練データ数の推移

学習結果の比較

Table: クラスごとの Macro-F1 の計算

Class	Accuracy	Precision	Recall	Macro-F1
A	0.97038	0.98195	0.97363	0.97777
B	0.98487	0.88118	0.88118	0.88118
C	0.97038	0.93208	0.95673	0.94424
D	0.99810	1.00000	0.25000	0.40000
E	1.00000	1.00000	1.00000	1.00000
Average	0.98475	0.95904	0.81231	0.84064

sklearn.f1_score で macro-F1 スコアの計算をすると、0.82730 が出力され、手計算と異なる値となった。

ソースコード

進捗報告

水野泰旭

F1 スコア

前回の問題点

検証データ 20%の混同行列

Macro-F1 スコアの再計算

データ数を揃えて学習

ソースコード

訓練データ数の推移

学習結果の比較

画像の枚数が少ないクラスに対して、画像を適当な回数コピーしてだいたい同じ枚数にする。

Listing 1: augmentation.py

```
1  # ラベルごとに抽出
2  index = list()
3  for i in range(5): index.append(train_labels == i)
4  each_train_images = dict()
5  each_train_labels = dict()
6  for i in range(5):
7      each_train_images[i] = train_img[index[i]]
8      each_train_labels[i] = train_labels[index[i]]
9
10 max_l = 0
11 for (key, val) in each_train_labels.items(): max_l = max(max_l, len(val))
12 augmentation_train_images = np.empty((0, 150, 150, 1))
13 augmentation_train_labels = np.empty(0)
14 for (key, val) in each_train_images.items():
15     for i in range(max_l // len(val)):
16         aug_train_images = np.append(aug_train_images, val, axis=0)
17         aug_train_labels = np.append(aug_train_labels, each_train_labels[key])
```

訓練データ数の推移

進捗報告

水野泰旭

F1 スコア

前回の問題点

検証データ 20%の混同行列

Macro-F1 スコアの再計算

データ数を揃えて学習

ソースコード

訓練データ数の推移

学習結果の比較

Table: もとの訓練データ

クラス	データ数
A	4824
B	411
C	1844
D	11
E	16
SUM	7106

Table: 増やした訓練データ

クラス	データ数
A	4824
B	4521
C	3688
D	4818
E	4816
SUM	22667

学習結果

進捗報告

水野泰旭

F1 スコア

前回の問題点

検証データ 20%の混同行列

Macro-F1 スコアの再計算

データ数を揃えて学習

ソースコード

訓練データ数の推移

学習結果の比較

Table: もとのデータと増やしたデータの学習結果

	Accuracy	Macro-F1
もとのデータによる学習結果	0.9699	0.7598 ¹
水増しデータによる学習結果	0.9519	0.7656

¹クラス D の TP と FP が共にゼロとなりあまり正確ではない

混同行列の比較

進捗報告

水野泰旭

F1 スコア

前回の問題点

検証データ 20%の混同行列

Macro-F1 スコアの再計算

データ数を揃えて学習

ソースコード

訓練データ数の推移

学習結果の比較

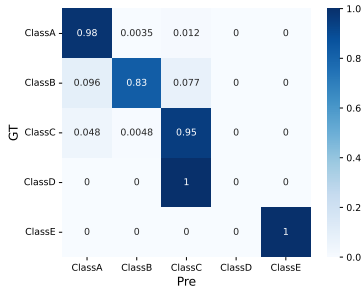


Figure: もとのデータ

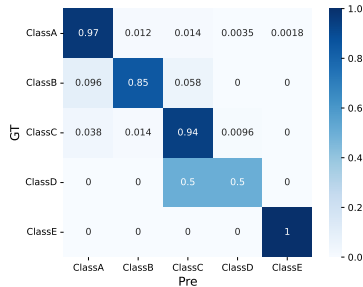


Figure: 水増しデータ