

進捗報告

水野泰旭

オーバーサンプリング

前回の手法

結果

冬休み期間中  
行ったこと

# 進捗報告

水野泰旭

弘前大学工学部電子情報工学科 4 年

January 16, 2023

# 目次

進捗報告

水野泰旭

オーバサンプ  
リング

前回の手法

結果

冬休み期間中  
行ったこと

- ① オーバサンプリング
  - 前回の手法
  - 結果
  
- ② 冬休み期間中行ったこと

# 前回の手法

進捗報告

水野泰旭

オーバサンプ  
リング

前回の手法

結果

冬休み期間中  
行ったこと

## 目的

使用したデータセットは不均衡データであり、一番データ数の多いクラスの画像の枚数だけ増やす。(オーバサンプリング)

前回のプログラムでは、一番多いクラスの画像の枚数になるように、画像を均一にコピーした。

# ランダムを用いたオーバサンプリング

進捗報告

水野泰旭

オーバサンプリング

前回の手法

結果

冬休み期間中  
行ったこと

一番多いクラスの画像枚数は 4,824 枚で、すべてのクラスの画像枚数が 4,824 になるようランダムに画像をコピー。訓練データは 24,120 枚となった。もとの訓練データは 7,106 枚

## Listing 1: rand\_augmentation.py

```
1 max_length = 0
2 for (key, val) in each_train_labels.items(): max_length = max(max_length, len(
    val))
3 augmentation_train_images = np.empty((0, 150, 150, 1))
4 augmentation_train_labels = np.empty(0)
5 for i in range(5):
6     repeated_perm = list()
7     for j in range(max_length): repeated_perm.append(random.randint(0, len(
        each_train_labels[i])-1))
8     augmentation_train_images = np.append(augmentation_train_images,
        each_train_images[i][repeated_perm], axis=0)
9     for _ in range(max_length): augmentation_train_labels = np.append(
        augmentation_train_labels, i)
10 print(augmentation_train_images.shape)
```

# 結果

進捗報告

水野泰旭

オーバーサンプリング

前回の手法

結果

冬休み期間中  
行ったこと

Table: 正解率と F1 スコアの比較

	Accuracy	$F1_{score}$
データ増やす前	0.9699	0.7598
データ増やした後	0.9495	0.8174

- 正解率下がったが、F1 スコアが上がった
- データを増やした後は、過学習が見られた

# 混同行列の比較

進捗報告

水野泰旭

オーバーサンプリング

前回の手法

結果

冬休み期間中  
行ったこと

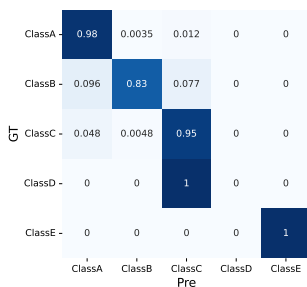


Figure: 増やす前



Figure: 増やした後

# 冬休み期間中行ったこと

進捗報告

水野泰旭

オーバーサンプ  
リング

前回の手法

結果

冬休み期間中  
行ったこと

- コードの修正
- 卒論
  - モデルに関する記述
  - $F1_{score}$  と計算結果の記述