開放平台軟體 期末報告

張友澤 李政憲 游登翔 張哲郡 劉彥麟

June 13, 2019

Outline

1. Introduction
   - Introduction to your team
   - Introduction to the problem you're trying to solve
2. Methodology
   - Input and output of your model
   - Each layer of your model
   - How you save and file size of your model?
   - What's your loss functions, and why?
   - What's your optimizer and the setting of hyperparameter?
3. Dataset
   - The size of our dataset should be larger than 1K
   - How you collect/build dataset?
   - How many paired training samples in dataset?
   - How many paired validating samples in dataset?
   - How many paired testing samples in dataset?
4. Experimental Evaluation
   - Experimental environment (CPU, GPU, memory,. . . ,etc.) and How many epochs you set for training?
   - Qualitative evaluation
   - Quantitative evaluation

Outline

1. Introduction
   - Introduction to your team
   - Introduction to the problem you're trying to solve
2. Methodology
   - Input and output of your model
   - Each layer of your model
   - How you save and file size of your model?
   - What's your loss functions, and why?
   - What's your optimizer and the setting of hyperparameter?
3. Dataset
   - The size of our dataset should be larger than 1K
   - How you collect/build dataset?
   - How many paired training samples in dataset?
   - How many paired validating samples in dataset?
   - How many paired testing samples in dataset?
4. Experimental Evaluation
   - Experimental environment (CPU, GPU, memory,. . . ,etc.) and How many epochs you set for training?
   - Qualitative evaluation
   - Quantitative evaluation

Introduction

Introduction to your team

1051438 張友澤
1051518 李政憲
1051540 游登翔
1051541 張哲郡
1051543 劉彥麟

Outline

1. Introduction
   - Introduction to your team
   - Introduction to the problem you're trying to solve
2. Methodology
   - Input and output of your model
   - Each layer of your model
   - How you save and file size of your model?
   - What's your loss functions, and why?
   - What's your optimizer and the setting of hyperparameter?
3. Dataset
   - The size of our dataset should be larger than 1K
   - How you collect/build dataset?
   - How many paired training samples in dataset?
   - How many paired validating samples in dataset?
   - How many paired testing samples in dataset?
4. Experimental Evaluation
   - Experimental environment (CPU, GPU, memory,. . . ,etc.) and How many epochs you set for training?
   - Qualitative evaluation
   - Quantitative evaluation

Introduction

Introduction to the problem you're trying to solve

有時候在廣播電台中、逛街途中會聽到自己覺得很好聽的音樂，但又不知道是誰唱得的時候，就可以錄一段音樂(要有人聲)，再丟進我們的程式，讓他告訴你是誰唱的。

Outline

1. Introduction
   - Introduction to your team
   - Introduction to the problem you're trying to solve
2. Methodology
   - Input and output of your model
   - Each layer of your model
   - How you save and file size of your model?
   - What's your loss functions, and why?
   - What's your optimizer and the setting of hyperparameter?
3. Dataset
   - The size of our dataset should be larger than 1K
   - How you collect/build dataset?
   - How many paired training samples in dataset?
   - How many paired validating samples in dataset?
   - How many paired testing samples in dataset?
4. Experimental Evaluation
   - Experimental environment (CPU, GPU, memory,. . . ,etc.) and How many epochs you set for training?
   - Qualitative evaluation
   - Quantitative evaluation

Methodology

Input and output of model

Input:
讀入MFCC向量特徵轉換後的.npy壓所檔，
將載入的train data與test data reshape為4個維度，
將train label 與 test label 類別變數轉為one-hot encoding，
即為欲輸入model的f所有資料

Output:
每個世代完成後，即輸出一HDF5檔案

Outline

1. Introduction
   - Introduction to your team
   - Introduction to the problem you're trying to solve
2. Methodology
   - Input and output of your model
   - Each layer of your model
   - How you save and file size of your model?
   - What's your loss functions, and why?
   - What's your optimizer and the setting of hyperparameter?
3. Dataset
   - The size of our dataset should be larger than 1K
   - How you collect/build dataset?
   - How many paired training samples in dataset?
   - How many paired validating samples in dataset?
   - How many paired testing samples in dataset?
4. Experimental Evaluation
   - Experimental environment (CPU, GPU, memory,. . . ,etc.) and How many epochs you set for training?
   - Qualitative evaluation
   - Quantitative evaluation

Methodology

Each layer of model

d0 = Input(shape=self.img_shape)
d1 = conv2d(d0, filters=32, f_size=2, stride=1, bn=True) 建立卷積層
d2 = maxpooling2d(d1, f_size=2, stride=2) 建立池化層
d3 = Dropout(0.25)(d2) Dropout層
d4 = flatten(d3) Flatten層
d5 = dense(d4, f_size=128, dr=True, lastLayer=False) 全連接層
d6 = dense(d5, f_size=5, dr=False, lastLayer=True) 全連接層

Outline

1. Introduction
   - Introduction to your team
   - Introduction to the problem you're trying to solve
2. Methodology
   - Input and output of your model
   - Each layer of your model
   - How you save and file size of your model?
   - What's your loss functions, and why?
   - What's your optimizer and the setting of hyperparameter?
3. Dataset
   - The size of our dataset should be larger than 1K
   - How you collect/build dataset?
   - How many paired training samples in dataset?
   - How many paired validating samples in dataset?
   - How many paired testing samples in dataset?
4. Experimental Evaluation
   - Experimental environment (CPU, GPU, memory,. . . ,etc.) and How many epochs you set for training?
   - Qualitative evaluation
   - Quantitative evaluation

使用save函式來儲存model至指定資料夾
每個Model size為2.15 MB
The Fig. 1

Figure: model相關資訊

Outline

1. Introduction
   - Introduction to your team
   - Introduction to the problem you're trying to solve
2. Methodology
   - Input and output of your model
   - Each layer of your model
   - How you save and file size of your model?
   - What's your loss functions, and why?
   - What's your optimizer and the setting of hyperparameter?
3. Dataset
   - The size of our dataset should be larger than 1K
   - How you collect/build dataset?
   - How many paired training samples in dataset?
   - How many paired validating samples in dataset?
   - How many paired testing samples in dataset?
4. Experimental Evaluation
   - Experimental environment (CPU, GPU, memory,. . . ,etc.) and How many epochs you set for training?
   - Qualitative evaluation
   - Quantitative evaluation

Methodology

loss functions and why

loss function使用'categorical_crossentropy'
因為用於多個分類，且目標值為分類格式(如:(1,0,0,0,0)、(0,1,0,0,0))，
所以選擇採用categorical_crossentropy作為損失函數

Outline

1. Introduction
   - Introduction to your team
   - Introduction to the problem you're trying to solve
2. Methodology
   - Input and output of your model
   - Each layer of your model
   - How you save and file size of your model?
   - What's your loss functions, and why?
   - What's your optimizer and the setting of hyperparameter?
3. Dataset
   - The size of our dataset should be larger than 1K
   - How you collect/build dataset?
   - How many paired training samples in dataset?
   - How many paired validating samples in dataset?
   - How many paired testing samples in dataset?
4. Experimental Evaluation
   - Experimental environment (CPU, GPU, memory,. . . ,etc.) and How many epochs you set for training?
   - Qualitative evaluation
   - Quantitative evaluation

optimizer採用'Adam'
metrics採用'accuracy'
The Fig. 2

```
self.CNN_Network.compile(loss='categorical_crossentropy', optimizer='Adam', metrics=['accuracy'])
```

Figure: optimizer and setting of hyperparameter

Outline

1. Introduction
   - Introduction to your team
   - Introduction to the problem you're trying to solve
2. Methodology
   - Input and output of your model
   - Each layer of your model
   - How you save and file size of your model?
   - What's your loss functions, and why?
   - What's your optimizer and the setting of hyperparameter?
3. Dataset
   - The size of our dataset should be larger than 1K
   - How you collect/build dataset?
   - How many paired training samples in dataset?
   - How many paired validating samples in dataset?
   - How many paired testing samples in dataset?
4. Experimental Evaluation
   - Experimental environment (CPU, GPU, memory,. . . ,etc.) and How many epochs you set for training?
   - Qualitative evaluation
   - Quantitative evaluation

Dataset

The size of our dataset should be larger than 1K

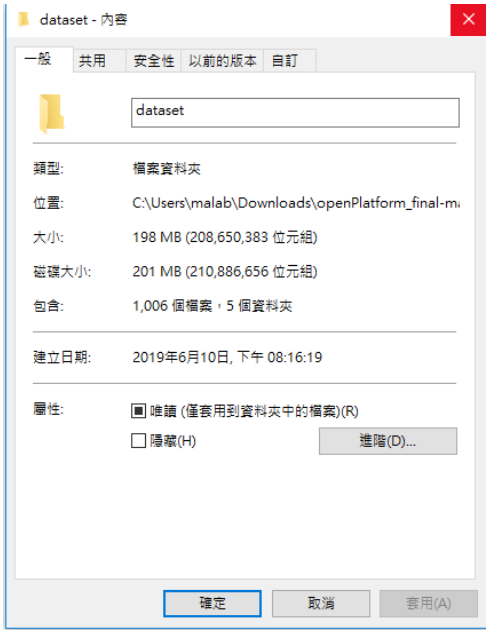

Figure: It's our Datasize

Outline

1. Introduction
   - Introduction to your team
   - Introduction to the problem you're trying to solve
2. Methodology
   - Input and output of your model
   - Each layer of your model
   - How you save and file size of your model?
   - What's your loss functions, and why?
   - What's your optimizer and the setting of hyperparameter?
3. Dataset
   - The size of our dataset should be larger than 1K
   - How you collect/build dataset?
   - How many paired training samples in dataset?
   - How many paired validating samples in dataset?
   - How many paired testing samples in dataset?
4. Experimental Evaluation
   - Experimental environment (CPU, GPU, memory,. . . ,etc.) and How many epochs you set for training?
   - Qualitative evaluation
   - Quantitative evaluation

Dataset

How you collect/build dataset?

1.把音樂下載成MP3的格式

2.用裁切軟體裁剪成每10秒一個人聲的音訊檔

3.把這些資料取mfcc特徵向量並製作成.npy壓縮檔

Outline

1. Introduction
   - Introduction to your team
   - Introduction to the problem you're trying to solve
2. Methodology
   - Input and output of your model
   - Each layer of your model
   - How you save and file size of your model?
   - What's your loss functions, and why?
   - What's your optimizer and the setting of hyperparameter?
3. Dataset
   - The size of our dataset should be larger than 1K
   - How you collect/build dataset?
   - How many paired training samples in dataset?
   - How many paired validating samples in dataset?
   - How many paired testing samples in dataset?
4. Experimental Evaluation
   - Experimental environment (CPU, GPU, memory,. . . ,etc.) and How many epochs you set for training?
   - Qualitative evaluation
   - Quantitative evaluation

Dataset

How many paired training samples in dataset?

使用此段code，從Dataset中每個類別取160筆資料(總共800筆資料)去訓練成模組

```
    self.train_data, self.test_data, self.train_labels, self.test_labels = self.get_train_test()

def get_train_test(self, split_ratio=0.8, random_state=42):
    # Get available labels
    labels, indices, _ = self.get_labels(DATA_PATH)

    # Getting first arrays
    X = np.load(labels[0] + '.npy')
    y = np.zeros(X.shape[0])

    # Append all of the dataset into one single array, same goes for y
    for i, label in enumerate(labels[1:]):
        x = np.load(label + '.npy')
        X = np.vstack((X, x))
        y = np.append(y, np.full(x.shape[0], fill_value=(i + 1)))

    assert X.shape[0] == len(y)

    return train_test_split(X, y, test_size=(1 - split_ratio), random_state=random_state, shuffle=True)
```

Figure: 利用function把1000筆資料分成800筆

Outline

1. Introduction
   - Introduction to your team
   - Introduction to the problem you're trying to solve
2. Methodology
   - Input and output of your model
   - Each layer of your model
   - How you save and file size of your model?
   - What's your loss functions, and why?
   - What's your optimizer and the setting of hyperparameter?
3. Dataset
   - The size of our dataset should be larger than 1K
   - How you collect/build dataset?
   - How many paired training samples in dataset?
   - How many paired validating samples in dataset?
   - How many paired testing samples in dataset?
4. Experimental Evaluation
   - Experimental environment (CPU, GPU, memory,. . . ,etc.) and How many epochs you set for training?
   - Qualitative evaluation
   - Quantitative evaluation

Dataset

How many paired validating samples in dataset?

使用此段code，從Dataset中每個類別取40筆資料(總共200筆資料，不會與train dataset的資料重複)來驗證模組的準確度

```
        self.train_data, self.test_data, self.train_labels, self.test_labels = self.get_train_test()

def get_train_test(self, split_ratio=0.8, random_state=42):
    # Get available labels
    labels, indices, _ = self.get_labels(DATA_PATH)

    # Getting first arrays
    X = np.load(labels[0] + '.npy')
    y = np.zeros(X.shape[0])

    # Append all of the dataset into one single array, same goes for y
    for i, label in enumerate(labels[1:]):
        x = np.load(label + '.npy')
        X = np.vstack((X, x))
        y = np.append(y, np.full(x.shape[0], fill_value=(i + 1)))

    assert X.shape[0] == len(y)

    return train_test_split(X, y, test_size=(1 - split_ratio), random_state=random_state, shuffle=True)
```

Figure: 利用function把1000筆資料分成200筆

Outline

1. Introduction
   - Introduction to your team
   - Introduction to the problem you're trying to solve
2. Methodology
   - Input and output of your model
   - Each layer of your model
   - How you save and file size of your model?
   - What's your loss functions, and why?
   - What's your optimizer and the setting of hyperparameter?
3. Dataset
   - The size of our dataset should be larger than 1K
   - How you collect/build dataset?
   - How many paired training samples in dataset?
   - How many paired validating samples in dataset?
   - How many paired testing samples in dataset?
4. Experimental Evaluation
   - Experimental environment (CPU, GPU, memory,. . . ,etc.) and How many epochs you set for training?
   - Qualitative evaluation
   - Quantitative evaluation

Dataset

How many paired testing samples in dataset?

總共50筆資料來測試模組

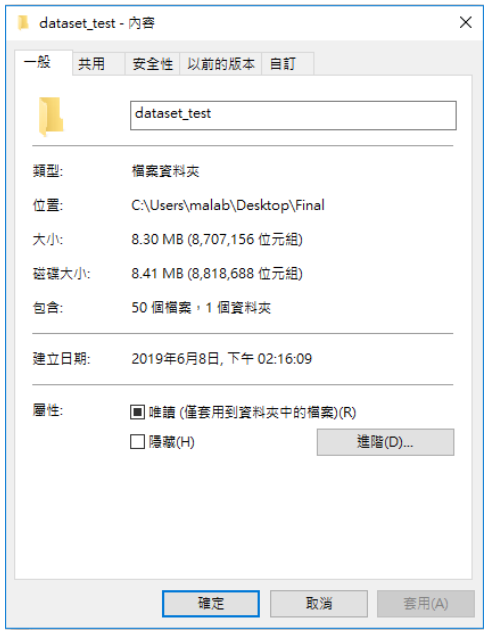

Figure: 50筆隨機歌手人聲的Test Data

Outline

1. Introduction
   - Introduction to your team
   - Introduction to the problem you're trying to solve
2. Methodology
   - Input and output of your model
   - Each layer of your model
   - How you save and file size of your model?
   - What's your loss functions, and why?
   - What's your optimizer and the setting of hyperparameter?
3. Dataset
   - The size of our dataset should be larger than 1K
   - How you collect/build dataset?
   - How many paired training samples in dataset?
   - How many paired validating samples in dataset?
   - How many paired testing samples in dataset?
4. Experimental Evaluation
   - Experimental environment (CPU, GPU, memory,. . . ,etc.) and How many epochs you set for training?
   - Qualitative evaluation
   - Quantitative evaluation

Experimental environment and how many epochs set for training?

CPU: Intel i5-4570 3.40GHz
RAM: 16GB
作業系統: Windows 10企業版
系統類型: 64位元作業系統，x64型處理器
Pycharm 2019.1.1 (Professional Edition)
(沒有使用GPU跑model)

本專題訓練了99個epochs

Outline

1. Introduction
   - Introduction to your team
   - Introduction to the problem you're trying to solve
2. Methodology
   - Input and output of your model
   - Each layer of your model
   - How you save and file size of your model?
   - What's your loss functions, and why?
   - What's your optimizer and the setting of hyperparameter?
3. Dataset
   - The size of our dataset should be larger than 1K
   - How you collect/build dataset?
   - How many paired training samples in dataset?
   - How many paired validating samples in dataset?
   - How many paired testing samples in dataset?
4. Experimental Evaluation
   - Experimental environment (CPU, GPU, memory,. . . ,etc.) and How many epochs you set for training?
   - Qualitative evaluation
   - Quantitative evaluation

Experimental environment

Qualitative evaluation

Outline

1. Introduction
   - Introduction to your team
   - Introduction to the problem you're trying to solve
2. Methodology
   - Input and output of your model
   - Each layer of your model
   - How you save and file size of your model?
   - What's your loss functions, and why?
   - What's your optimizer and the setting of hyperparameter?
3. Dataset
   - The size of our dataset should be larger than 1K
   - How you collect/build dataset?
   - How many paired training samples in dataset?
   - How many paired validating samples in dataset?
   - How many paired testing samples in dataset?
4. Experimental Evaluation
   - Experimental environment (CPU, GPU, memory,. . . ,etc.) and How many epochs you set for training?
   - Qualitative evaluation
   - Quantitative evaluation

The Fig. 7

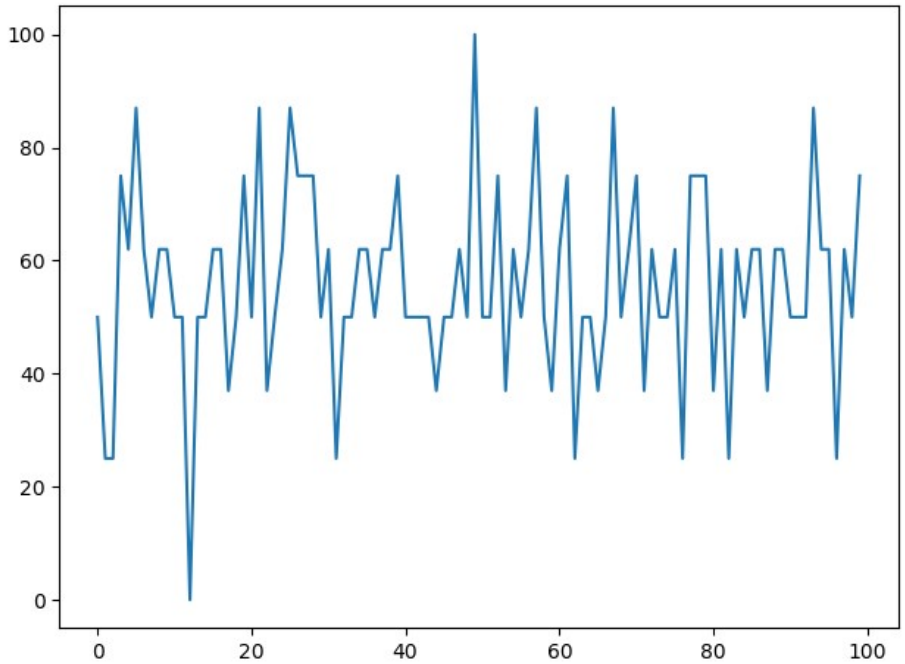

Figure: 每個世代model訓練正確率

Authorship

Job scheduling of your team

05/31-06/03 每個人上傳自己選擇的歌手的dataset(200個檔案)
06/03-06/09 寫codetrain完model到99世代
06/09-06/13 Latex建置presentation，SRS

Authorship

Contribution of each team member with evidence

張友澤：dataset，SRS
李政憲：dataset，Presentation
游登翔：dataset，Presentation
張哲郡：dataset，大部分code
劉彥麟：dataset，SRS