ADL HW1-TIMIT Report

b03705012 資管四 張晉華

1.Model description

- RNN

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
bidirectional_1 (Bidirection (None, 776, 1000)         1620000
_________________________________________________________________
bidirectional_2 (Bidirection (None, 776, 1000)         4503000
_________________________________________________________________
time_distributed_1 (TimeDist (None, 776, 1000)         1001000
_________________________________________________________________
dropout_1 (Dropout)          (None, 776, 1000)         0
_________________________________________________________________
time_distributed_2 (TimeDist (None, 776, 1000)         1001000
_________________________________________________________________
dropout_2 (Dropout)          (None, 776, 1000)         0
_________________________________________________________________
time_distributed_3 (TimeDist (None, 776, 39)           39039
=================================================================
Total params: 8,164,039
Trainable params: 8,164,039
Non-trainable params: 0
```

(Model Summary)

  - 兩層的 Bidirectional GRU:Units = 500, dropout = 0.5, activation = ReLU
  - 兩層 TimeDistributed Dense: Units = 1000, activation = ReLU, 加上 TimeDistributed 讓 Output Shape 維持 776(frames)*1000
  - 最後一層 TimeDistributed Dense 作為 Output Layer: Units = 39(事先用 48_39.map,48phone_char.map 將 label mapping 成 char 了), activation = Softmax, 加上 TimeDistributed 一樣為了讓 Output Shape 維持 776(frames)*39(probability of per label)

- RNN+CNN

```
Layer (type)                 Output Shape              Param #
=================================================================
conv1d_1 (Conv1D)            (None, 776, 800)          1217600
_________________________________________________________________
time_distributed_1 (TimeDist (None, 776, 39)           31239
_________________________________________________________________
bidirectional_1 (Bidirection (None, 776, 1000)         1620000
_________________________________________________________________
bidirectional_2 (Bidirection (None, 776, 1000)         4503000
_________________________________________________________________
bidirectional_3 (Bidirection (None, 776, 500)          1876500
_________________________________________________________________
bidirectional_4 (Bidirection (None, 776, 500)          1126500
_________________________________________________________________
time_distributed_2 (TimeDist (None, 776, 39)           19539
=================================================================
Total params: 10,394,378
Trainable params: 10,394,378
Non-trainable params: 0
```

(Model Summary)

  - 一層的 Conv1d:Filters = 800, Filters Shape = (39)
  - 兩層的 Bidirectional GRU:Units = 500, dropout = 0.3, activation = ReLU
  - 兩層的 Bidirectional GRU:Units = 250, dropout = 0.3, activation = ReLU
  - 最後一層 TimeDistributed Dense 作為 Output Layer: Units = 39(事先用 48_39.map,48phone_char.map 將 label mapping 成 char 了), activation = Softmax, 加上 TimeDistributed 一樣為了讓 Output Shape 維持 776(frames)*39(probability of per label)

2.How to improve your performance

- Standardize

先將 mfcc 的 data 做標準化，將他整理成每個 column 都是標準差為 1，平均為 0 的資料。（由於 mfcc 是 1 個對數能量和 12 個倒頻譜參數再加上差量運算和差差量運算所形成的 39 維資料，因此以 column 做標準化單位而不用 row)

- Output Filter

  由於最後 performance 的評估方式是以 edit distance 來做評價，所以我決定忽略 frame 的連續數量不足的 label，只輸出連續數量含或超過 3 個以上的 label。

  - Compare(以同一模型做比較)
    - before:10.14124(Kaggle Public Score)
    - after:9.55367(Kaggle Public Score)

- Ensemble

  將 3 個不同的模型做 ensemble，以避免個別模型的 bias，有實作過 voting 跟 average 兩種作法。

  - Compare(Before 是以單一模型的最高分為準)
    - Voting
      - before:9.55367(Kaggle Public Score)
      - after:9.53672(Kaggle Public Score)
    - Average
      - before:9.55367(Kaggle Public Score)
      - after:8.81920(Kaggle Public Score)

3.Experimental results and settings

- Compare and analyze the results between RNN and CNN
  - Compare
    - RNN:11.41807(Kaggle Public Score)
    - CNN+RNN:13.18644(Kaggle Public Score)
  - Analyze
    - CNN+RNN 在 training 時收斂速度比 RNN 還快但容易 overfit

        - CNN+RNN 由於 frame 的 feature 在 CNN 時先通過了 Filters，因此到 RNN 時對前後相關性的判斷資訊稍微不足，因此綜合的結果表現比起純 RNN 似忽略差。

- Compare and analyze the results with other models
    - SimpleRNN 和 GRU 的比較

      在同樣的架構下 GRU 的表現比 RNN 好，收斂速度也較快

    - 4 層 GRU 和 2 層 GRU+2 層 Dense 的比較

      4 層 GRU 對於前後相關性的判斷能力比起 2 層 GRU+2 層 Dense 來的好一些，雖然計算時間較久、收斂速度稍慢一點，但最後的 Performance 較好。

        - 4 層 GRU:10.14124(Kaggle Public Score)
        - 2 層 GRU+2 層 Dense:11.02824(Kaggle Public Score)