

# ADLxMLDS 2017 Fall

## HW1 - Sequence Labeling

B05901189 吳祥叡

October 26, 2017

### 1 Model Description

#### 1.1 RNN Model

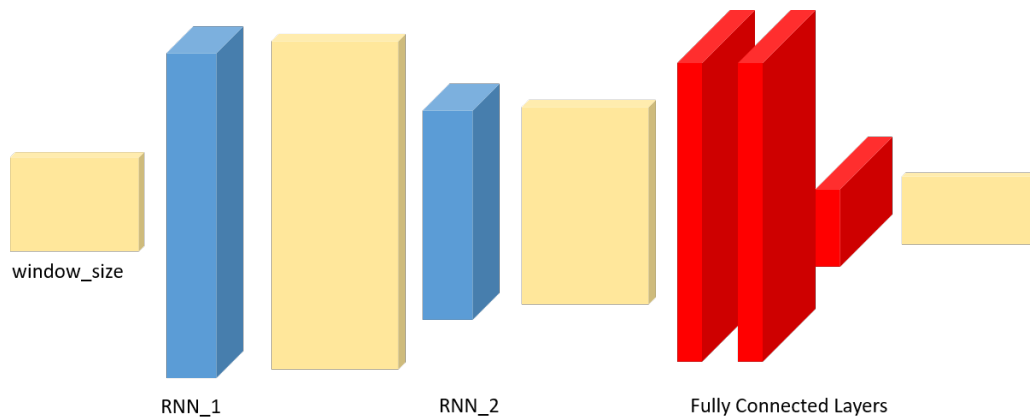


Figure 1: 兩層的 RNN Model

圖中縱方向為 feature 的長度. 經過兩層 RNN layer 之後再通過 Fully Connected Layer 產生 frame-wise phone 判斷.

#### 1.2 CNN Model

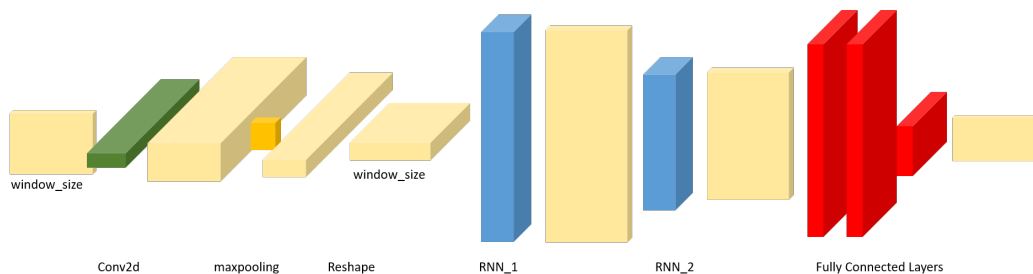


Figure 2: CNN Model

CNN Model 的後半部和 RNN Model 一樣. 前段使用兩個 convolution layer 和 maxpooling layer, 為了把其中一個 dimension 變成 window size 要再做 reshape. 我覺得對時間軸 maxpooling 可能丟失太多資訊所以也試過沒有 maxpooling, 但效果比較不好.

## 2 Boosting Performance

### 2.1 後處理

1. 用 sliding window 做投票: 因為這些 model 只能根據一段固定長度的 context window 對每個 frame 做預測, 因此我讓 window 一次只往前移動一個 frame 再做預測, 這樣每個 frame 就可以有 window size 次的投票.
2. 忽略前段和後段的投票:  
因為前段和後段可能因為沒有看到比較多前後文所以我選擇忽略掉他們的投票. 我做了三次實驗:  
如果不忽略: kaggle 分數 23  
前後各忽略 10 個: kaggle 分數 20  
前後各忽略 20 個: kaggle 分數和忽略 10 個一樣
3. 去除 glitch :  
我對 training data 做過統計發現每個 phone 最少都會由 3 個 frame 組成, 所以決定在把 frame-wise prediction 變成 phone sequence 的時候刪掉小於 3 個 frame 的 phone. 如果不刪掉的話 kaggle 分數 20, 用了這個 trick 可以直接進步到 12 分. 或許用 seq2seq 方法做 trimming 可以有更好的結果.

## 3 Experiment Settings and Results

### 3.1 Experiment Settings

1. 運算資源: 有兩個 K80 GPU 的 Azure NC6
2. 使用套件: Tensorflow
3. 前處理: 將全部 timit dataset 串起來以便之後產生 batch
4. Validation: 把串好的 training data 切出後 10%.
5. Kaggle Metric : 每段聲音的平均 Phone Error (Edit Distance), 非平常的 Phone Error Rate.

## 3.2 參數試驗

1. 不同 RNN cell: BasicGRU 和 BasicLSTM Cell 表現都差不多
2. RNN 層數: 發現 2 層的收斂性比較好
3. Bidirection: 為了比較固定參數量兩種 model 的表現, 在使用 Bidirection 的時候的 RNN latent dimension 都是正常的一半. 出乎意料的就算在 RNN model, 用 Bidirectional RNN 完全沒有提昇表現.
4. Window Size : 實驗發現 32 以下的 window size 在 validation acc 和 loss 表現比較不好. 超過 50 之後表現都差不多. 原因應該就和要刪除前段和後段類似.
5. 刪掉不確定的 prediction : 因為原本的 prediction 是沒有正規化的, 所以我取通過 softmax 機率加總為一的 prediction, 設定不同的 threshold, 刪掉不確定的 prediction. 實驗發現表現都沒有刪除前段和後段的方法好.

## 3.3 Results

	RNN model	CNN model
Validation Loss	0.98	0.83
Validation Accuracy	69 %	72 %
Kaggle Public Score	12.2	9.6

兩種 model 大約都花一個小時就訓練完, 但也代表這兩個 model 都很快就 overfit, 是一個待解決的問題.