

ADLxMLDS 2017 Fall

HW1 - Sequence Labeling

B05901189 吳祥叡

October 25, 2017

1 Model Description

1.1 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.

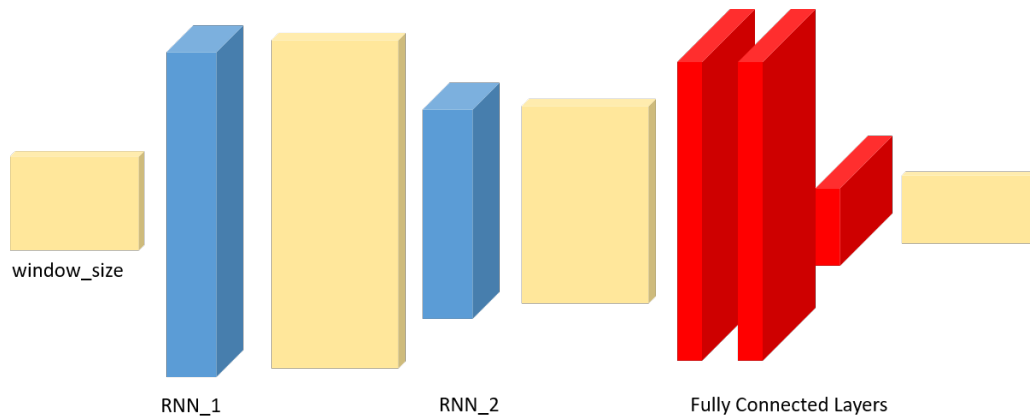


Figure 1: 兩層的 RNN Model

1.2 CNN Model

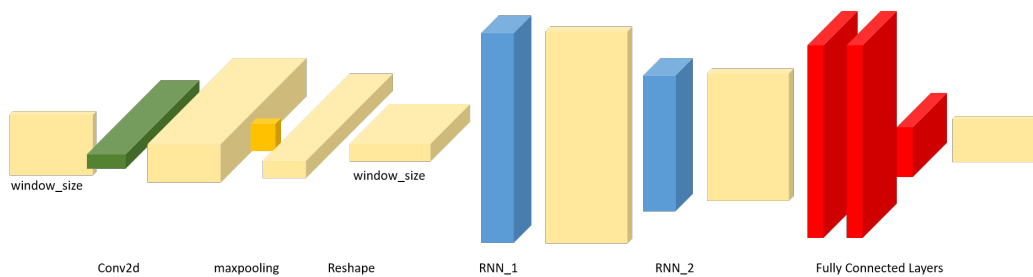


Figure 2: CNN Model

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:
2. RNN 層數:

3. Bidirection:
4. Window Size :
5. 刪掉不確定的 prediction :

3.3 Results

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