顏廷宇
資工四 B03902052
2017年10月28日

ADLxMLDS HW1

1. Model description

A, RNN :

一個 64 neurons 的 LSTM 接上 48 neurons 的 Densely-Connected Layer。Input 吃 (Batch_Size, Time_Step, Input_Dim) 格式的 mfcc 或 fbank 資料，Output 會為 48 dimensions 的 Vector。

(p.s. 故training前，須先將 training label 轉為 48-dim one-hot vector)

Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 777, 64)	44288
dense_1 (Dense)	(None, 777, 48)	3120

B, CNN+RNN :

一個 64 neurons, stride=(3,1) 的 Conv2D 接上 1 neuron, stride=(1,1) 的 Conv2D ，經過reshape後，接上 64 neurons 的 LSTM 與 48 neurons 的 Densely-Connected Layer。Input 吃 (Batch_Size, Time_Step, Input_Dim, 1) 格式的 mfcc 或 fbank 資料，Output 會為 48 dimensions 的 Vector。

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 777, 108, 64)	256
conv2d_2 (Conv2D)	(None, 777, 108, 1)	65
reshape_1 (Reshape)	(None, 777, 108)	0
lstm_1 (LSTM)	(None, 777, 64)	44288
time_distributed_1 (TimeDist	(None, 777, 48)	3120

2. How to improve your performance

A, Model Description :

三層 512 neurons 的 Bidirectional LSTM，加上四層 Densely-Connected Layer (neurons分別為 256, 128, 64, 48) 。Input 吃 (Batch_Size, Time_Step, Input_Dim) 格式的 mfcc 或 fbank 資料，Output 為 48 dimensions 的 Vector。

Layer (type)	Output Shape	Param #
bidirectional_1 (Bidirection	(None, 777, 1024)	2543616
dropout_1 (Dropout)	(None, 777, 1024)	0
bidirectional_2 (Bidirection	(None, 777, 1024)	6295552
dropout_2 (Dropout)	(None, 777, 1024)	0
bidirectional_3 (Bidirection	(None, 777, 1024)	6295552
dropout_3 (Dropout)	(None, 777, 1024)	0
dense_1 (Dense)	(None, 777, 256)	262400
dropout_4 (Dropout)	(None, 777, 256)	0
dense_2 (Dense)	(None, 777, 128)	32896
dropout_5 (Dropout)	(None, 777, 128)	0
dense_3 (Dense)	(None, 777, 64)	8256
dropout_6 (Dropout)	(None, 777, 64)	0
dense_4 (Dense)	(None, 777, 48)	3120

B, Methods :

在 input 方面，同時使用 mfcc 與 fbank 的資料， 這樣可以增加 feature 的數量。 LSTM 可以利用過去的資訊，但因為我們已經有未來的資料，所以用 Bidirectional LSTM 擁有 forward track 的特性，來增加準確度。而加深 Bidirectional LSTM 可以讓每層學到更好的資訊，以增加預測的準確性。增加 Densely-Connected Layer 是為了不讓每層 neurons 的下降太多，讓學到的資訊可以保留更多。增加 Dropout 是為了避免 overfitting 。最後，在 decode 的時候，會將連續出現次數少於二的聲音刪除，這是從 training label 中，發現到很少有連續出現次數少於二的聲音，故在 trimming 先將其刪除。

3. Experimental results and settings

A, RNN v.s. CNN :

	RNN	CNN+RNN
One Layer	15.34	14.18
Two Layers	13.49	12.65
Three Layers	13.57	12.91

從圖表中可以看到，加入 CNN 後，準確度都有提升，代表 CNN 可以將資訊擷取並放大其特徵，讓之後的 RNN 拿到更好分辨的資料。

B, RNN v.s. BLSTM :

	RNN	BLSTM
One Layer	15.34	8.21
Two Layers	13.49	7.92
Three Layers	13.27	7.86

BLSTM 的表現比 LSTM 好上一層，代表 BLSTM 可以利用未來的資訊做出更準確的判斷；增加層數也使表現越來越好，由此得知，深度的增加確實能讓每個 neurons 學到的資訊更精準。

C, BLSTM Deep or Wide?

	BLSTM
Three Layer - 512 neurons	7.86
Six Layer - 128 neurons	9.29
Ten Layer - 64 neurons	10.14

從圖中可以看到，neurons 數多，層數較淺的 model 表現較佳，推測是因過深的 model layer 之間沒有做 normalization，導致參數傳遞可能歪掉，使成效比淺層的差。