

# ADLxMLDS HW1 報告

---

b03902083 陳力

## Model Description

---

- Feature:  
在兩個模型中，我所採用的 feature 皆為 39 維的 MFCC，且皆經過 z score 做過正則化 ( $\vec{x}_t \leftarrow \sigma^{-1}(\vec{x}_t - \mu)$ ，其中  $\mu$  與  $\sigma$  分別為向量  $\vec{x}_t$  的平均值與標準差)。
- Activation Function:  
我所使用的 activation function 皆為 ReLU unit ( $f(x) = \max(0, x)$ )。
- Loss Function:  
我所使用的 Loss Function 皆為 frame-wise 的 Cross Entropy Loss ( $L(y, c) = -\log y_c$ )
- Validation set:  
隨機得從 training set 中選取 100 個 sequence 作為 validation set。
- Optimizer:  
使用 pytorch 套件內建的 ADAM 演算法，learning rate 固定在 0.0001。

## RNN

我所使用的 RNN 模型是多層雙向 LSTM (DBLSTM)。在 DBLSTM 裡，每個 frame 在兩個方向的輸出  $\vec{h}, \tilde{h}$  會被線性組合成單一向量表示各種 label 的機率分佈 ( $y = W_1 \vec{h} + W_2 \tilde{h} + b$ )。

## CNN + RNN

在我的 CNN + RNN 模型裡，每個 frame 會先被 CNN 做過轉換，然後這些轉換過的 frame 所成的 sequence 再送進上述的 RNN model 裡做預測。

CNN 的部分則是每個 frame 會配上前後各一個 frame，成為一個  $3 \times 39$  的二維 tensor，然後送進 convolution layer。convolution layer 則是一個輸入 channel 對到一個輸出 channel，kernel 的大小是  $3 \times 2$ ，pool 的大小是 2，所以輸出會是一個  $1 \times 19$  的 tensor；之後每個值再送進 ReLU 裡。

## Performance Improvements

---

### Best model

我表現最好的模型參數如下：

6層 DBLSTM，hidden cell 與 batch 大小皆為 100，訓練時 LSTM 裡的 dropout 機率为 0.5，訓練的回合數為 300。

在 public testing set 達到 8.02259 的分數。

## Technique

在 trimming 的過程裡，主動捨棄連續出現不超過 2 次的 label。

## Why?

1. 層數越多，model 的描述能力就越強，但也越容易 overfit，所以需要引進 dropout 的技巧來控制。
2. 在 training set 裡，因為 frame 所代表的聲音長度很短，全部的 label 在 sequence 都會連續出現好幾次。這個觀察讓我們在做預測時，可以主動捨棄零星出現的 label，增加準確率。

## Experimental Results

RNN + Naive Trim:

層數	hidden cell 大小	dropout	epoch 數	Public score
5	100	0.3	56	15.69491
5	100	0.3	172	10.97175
5	100	0.3	231	10.53672
5	100	0.3	300	10.27683

RNN + 移除零星 label (連續出現  $\leq 2$  次):

層數	hidden cell 大小	dropout	epoch 數	Public score
5	100	0.3	300	8.85875
6	100	0.5	300	8.02259
6	100	0.5	400	8.02824
7	100	0.5	500	8.02259
6	1024	0.5	5	8.55932
6	1024	0.5	9	7.81920
6	1024	0.5	20	7.71186

RNN + 移除零星 label (連續出現  $\leq 3$  次):

層數	hidden cell 大小	dropout	epoch 數	Public score
6	100	0.5	300	9.55932
6	100	0.5	400	9.50282

RNN + 移除零星 label (連續出現  $\leq 2$  次) + MFCC + FBANK:

層數	hidden cell 大小	dropout	epoch 數	Public score
6	1024	0.5	5	8.55932
6	1024	0.5	9	7.81920
6	1024	0.5	15	7.74576
6	1024	0.5	18	7.75141
6	1024	0.5	20	7.71186

CNN + RNN + 移除零星 label (連續出現  $\leq 2$  次):

kernel 大小	pooling 大小	層數	hidden cell 大小	dropout	epoch 數	Public score
3 x 2	2	5	100	0.3	300	9.84745

從以上的結果可以猜測:

1. Train 越多回合對結果在大部分時候有好的影響，即使 model 在 validation set 的結果早已收斂。
2. Trim 的方法對分數有顯著的影響。
3. 我的 CNN model 對結果有負面的影響，壓縮掉太多資訊。
4. MFCC + FBANK 兩種 feature 同時用會有好的效果。

## Residual Network

Deep residual network 是現在很流行的模型，我將他接在 RNN 之後，做了一些實驗。

模型長相:

有四組 residual block，分別為

1. 1 channel to 16 channels, kernel size 3, 6 layers
2. 16 channels to 8 channels, kernel size 3, 2 layers
3. 8 channels to 4 channels, kernel size 3, 2 layers
4. 4 channels to 2 channels, kernel size 3, 2 layers

DBLSTM + Residual network:

層數	hidden cell 大小	dropout	epoch 數	Public score
2	128	0.2	100	10.00000

Deep residual network

## Reference

<http://cs231n.stanford.edu/reports/2017/pdfs/804.pdf>

(<http://cs231n.stanford.edu/reports/2017/pdfs/804.pdf>)

<https://arxiv.org/pdf/1611.07174.pdf> (<https://arxiv.org/pdf/1611.07174.pdf>)

[https://www.cs.toronto.edu/~graves/asru\\_2013.pdf](https://www.cs.toronto.edu/~graves/asru_2013.pdf) ([https://www.cs.toronto.edu/~graves/asru\\_2013.pdf](https://www.cs.toronto.edu/~graves/asru_2013.pdf))

<https://arxiv.org/pdf/1303.5778.pdf> (<https://arxiv.org/pdf/1303.5778.pdf>)