ADLxMLDS HW1 Report

R06922055 資工所 碩一 吳均庭

Model description

- RNN

  對於每個batch，透過FC input layer接進1層LSTM(64 hidden node)，Reshape後再由FC output layer 將hidden node 轉為48維output，最後經過softmax使用tensorflow實作的masked loss function將padding貢獻之loss權重設為0，計算cross entropy loss。

- CNN + RNN

  在load data時會先在頭尾各padding一個frame，輸入shape為 (batch, time_step + 2, feature, 1) 通過一層kernel 為 1*3 的 conv2d layer，輸出會形成與前述RNN input shape相同的vector，接著透過FC input layer接進1層LSTM(64 hidden node)，Reshape後再由FC output layer 將hidden node 轉為48維output，最後經過softmax使用tensorflow實作的masked loss function將padding貢獻之權重設為0，計算cross entropy loss。

How to improve your performance

1. 增加Epoch數量

   讓model loss能夠收斂，至最低點，但須注意validation accuracy 的變化，不宜過多，需要注意是否發生overfitting之現象。

2. 手動降低Leaning Rate

   模型訓練時，一般使用Adam Optimizer learning rate=0.001，當train到正確率無法繼續上升時，便手動降低learning rate 至 0.0005或 0.0001，可以讓正確率再上升一些。

3. 層數加深

   在訓練模型，直覺上便是把模型層數增加，來達到更好的效果，正確率會隨著層數增加逐漸增加至收斂，我嘗試過最多至10層的bidirection rnn，但因為層數過多，訓練十分緩慢，且繼續增加層數，可能會造成梯度無法傳遞造成模型訓練失敗。

4. Bidirectional RNN

   單向LSTM只有考慮過去的資訊，透過有Bidirectional同時跑順向和逆向的LSTM，能夠

使用未來的資訊來做單前frame的判斷，做法也符合直覺，結果也確實有所提升。

5. 消除雜訊

   考量到人不太可能有20ms的發音，所以使用sliding window來消除 aaabaaa 以及 aaaabccccc這兩類雜訊，可以在測試結果中發現正確率有所提升。

6. Dropout layer

   Dropout能夠一定程度上避免模型發生overfit，在使用較複雜的model時加入dropout 0.5 或 0.75，來避免model太早發生overfitting，訓練上確實有顯著的效果。

7. 調整Neuron數量

   增加LSTM內的neurons數量，作業中發現，增加neuron數量，會增加訓練所需時間，實驗中也試著將層數降低，增加neurons，發現loss能夠更快收斂。

8. Optimizer

   試過SGD、RMSprop、Adam三種optimizer，發現使用adam效果最好。

9. Normalization

   避免每句 Sequence當中有不同的分佈行情，所以處理data時先透過normalization讓 mean = 0, standard deviation = 1。

Experimental results and settings

本次作業，我使用Tensorflow進行實作，使用mfcc、fbank 以及 fbank+mfcc三種data，發現使用mfcc + fbank效果最佳。另外CNN+RNN之目的為考慮前後Frame的資訊，來幫助提高正確率，但作業中發現，與LSTM RNN model 相比，並沒有讓正確率提升，於是我嘗試使用Bidirectional RNN來考慮前後文的資訊，並且不會受到CNN window szie的限制，比起RNN有顯著的提升。另外自己實作masked loss 和 accuracy函數，來避尾部padding讓sil的frame太多，而影響到正確率以及loss之計算，也能認正確率繼續提升。

Best_model使用 10層 bidirectional RNN，LSTM大小為64，training batch=16，dropout 0.5，training acc為 98%、valid acc 為 81%，在Public Set 上 Edit distance 為 8.03 , private set 降低到 7.79。contest結束之後，發現在所有submission，最好的 private score 為7.621使用的model 為5層的BiRNN，hidden node=512，表示淺層較寬的model效果甚至超越深層model，且訓練時間較短，比起深層model能更快收斂，可能是因為model太深會

造成gradient難以傳遞，另外輸出都使用前部分提到消除雜訊之方法，能夠將結果再提升1左右。

model	validation accuracy	kaggle public	kaggle private
1 LSTM		20.5	19.7
3 LSTM	0.72	16.20	16.21
2 CNN + 4 biLSTM64	0.74	13.21	13.52
2 CNN + 6 biLSTM64	0.76	10.44	10.6
2 CNN + 8 biLSTM64	0.79	8.6	8.7
8 biLSTM64	80.3	9.06	8.82
5 biLSTM 512	0.82	8.2	7.62
6 biLSTM 400	0.81	8.20	8.25
10 biLSTM 64 dropout 0.5	0.81	8.03	7.79