

深度學習應用——作業一報告

B03902072 江廷睿

1 Model description

1.1 Preprocess

只有用到 MFCC 這個特徵。特徵中的每個維度都有各自進行縮放、平移，使得訓練資料中 39 維中的每一個維度的平均值都是 0，標準差都是 1。每句話作為一筆訓練資料。

1.2 參數與最佳化方式

- 批的大小：32
- 最佳化方式：Adam
- 學習率：0.001

1.3 RNN

- Dropout，丟失比例 0.2
- 隱藏狀態維度為 256 的 Bi-directional LSTM
- Dropout，丟失比例 0.2
- 隱藏狀態維度為 256 的 Bi-directional LSTM
- Dropout，丟失比例 0.2
- 隱藏狀態維度為 128 的 Bi-directional LSTM
- 對於每個訊框，各自線性轉換成 39 維的分數，並做 softmax

1.4 RNN + CNN

- 核大小 7，64 個過濾器的一維卷積層，補齊方式為 same
- 批標準化層
- 線性整流函數
- 以下重複 4 次：
 - 核大小 7，64 個過濾器的一維卷積層，補齊方式為 same

- 批標準化層
 - 線性整流函數
 - Dropout，丟失比例 0.1
 - 核大小 7，64 個過濾器的一維卷積層，補齊方式為 same
 - 與上上上上層的輸入相加
 - 批標準化層
 - 線性整流函數
 - Dropout，丟失比例 0.1
- 隱藏狀態維度為 128 的 Bi-directional LSTM
- 對於每個訊框，各自線性轉換成 39 維的分數，並做 softmax

1.5 Best

- 核大小 7，64 個過濾器的一維卷積層，補齊方式為 same
- 批標準化層
- 線性整流函數
- 以下重複 10 次：
 - 核大小 7，64 個過濾器的一維卷積層，補齊方式為 same
 - 批標準化層
 - 線性整流函數
 - Dropout，丟失比例 0.2
 - 核大小 7，64 個過濾器的一維卷積層，補齊方式為 same
 - 與上上上上層的輸入相加
 - 批標準化層
 - 線性整流函數
 - Dropout，丟失比例 0.2
- 對於每個訊框，各自線性轉換成 39 維的分數，並做 softmax

2 How to improve your performace

2.1 Dropout

Dropout 就是在訓練過程中，隨機將上一層的部分輸出設成 0 的技術。使用的主要理由是實驗上在驗證資料中的確能延遲過度擬合，並帶來比較高的準確度。

2.2 加深網路

實驗結果顯示，無論是 RNN 或 CNN，把網路加深都有助於模型的準確度。

	RNN	RNN + CNN	RNN
訊框準確度 (驗證集)	0.8008	0.8101	0.7989
Edit Distance (驗證集)	10.1299	10.0825	9.5440
Edit Distance (Public Leaderboard)	10.6384	10.9496	9.7853

表 1: RNN、RNN + CNN、CNN 三種架構的比較。

2.3 使用 LSTM

實驗結果顯示，從 GRU 換成 LSTM 也能帶來一些提昇，推測可能是因為 LSTM 比 GRU 有更強的表現能力。

2.4 殘差網路與批標準化

參考《Deep Residual Learning for Image Recognition》，這裡使用了殘差的架構，據稱可以增進訓練深度網路的效果。批標準化也是論文中建議使用的方法。

3 Experimental results and settings

3.1 RNN 與 RNN + CNN 與 CNN 的比較

表 1 三個模型的表現，他們的架構如第 1 節，其餘設定都相同。從表格中可以看到，雖然 RNN + CNN 的架構有最高的訊框準確度（所有訊框的預測結果中，正確的比例），但 edit distance 卻是三者中最差的。純粹 RNN 的架構表現在三個架構的中間，而純粹的 CNN 則有最好的 edit distance。

3.2 LSTM 與 GRU 的比較

也有實驗過將 LSTM 換成 GRU 的 RNN 模型。但 GRU 模型的各種表現都比不上 LSTM，訊框準確度為 0.7947，public leaderboard 上的 edit distance 也只有 11.8814。

3.3 不同深度的 CNN

實驗顯示不管有沒有用 RNN，使用較深 CNN 都可以有比較好的效果。以第 1 節提到的 RNN + CNN 架構為例，如果只使用 9 層的 CNN，訊框準確度只有 0.7949，edit distance 也只到 11.5593。

A 使用的套件

- bleach==1.5.0
- editdistance==0.3.1
- html5lib==0.9999999
- Markdown==2.6.9
- numpy==1.13.3
- pandas==0.20.3
- protobuf==3.4.0
- python-dateutil==2.6.1
- pytz==2017.2
- six==1.11.0
- tensorflow-gpu==1.3.0
- tensorflow-tensorboard==0.1.8
- tqdm==4.19.2
- Werkzeug==0.12.2