

作業一 Sequence Labeling

劉彥廷 B03902036

1. 模型敘述

1.1. RNN

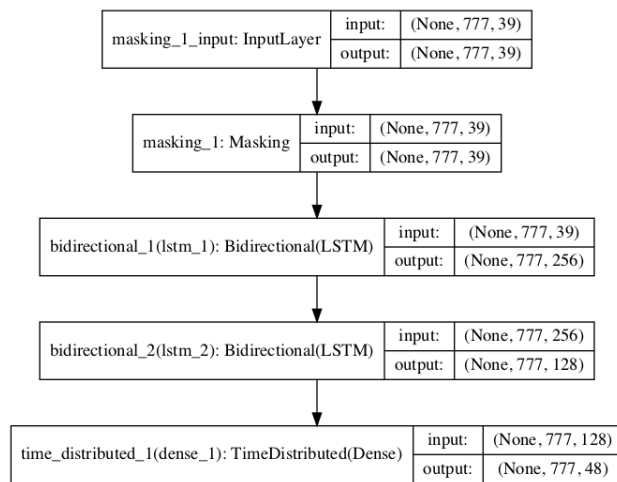


Figure 1: 繳交的 RNN 模型

Input Layer 使用了以 0 作為遮罩的 Masking Layer 讓模型忽略所有使用 0 在特徵維度進行 padding 的 frames，亦即，如果一個句子長度不足 777 個 frames，會提供各個維度特徵皆為 0 的 frames 來補足。

整體設計上依照作業規格的投影片所指示的 LSTM，總計兩層 Bidirectional 的 LSTM。兩層各自有一組的 forward 與 backward 層，各組的單元數分別為 128×777 與 64×777 。其中 777 這個數字乃為了滿足訓練的資料組最長的句子有 777 個 frames。

Output Layer 由一層 Dense Layer 構成，裡頭的單元數恰好吻合 phones 的數量 (48)，activation 的方式為 softmax。

1.2. RNN + CNN

TBA

2. 優化方式

2.1. 策略

本次作業採取的策略為

1. 使用 MFCC 與預設的 batch size (32) 以及論文 [2] 所採用的 10 個 epochs，但選用不同的模型。
2. 在參考過 RNN 與 LSTM (在 Keras 當中的分類) 的 framw-wise 準確率以後，選用 LSTM (爾後增加了 Bidirectional Wrapper)。

3. 調整 optimizer 與初始化的 kernel (最後沒有依照論文採用的 RandomUniform)，並調整 batch size 與 epochs 大小，觀察 loss 變化與 GPU 使用率和訓練時間。
4. 反覆上述步驟。

在這過程當中並沒有使用系統性的在 hyperparameter 的空間 (從策略當中，可以調整的主要為 batch size、epochs 與初始化的 kernel) 當中搜尋最佳值，僅人工的固定間隔取樣 (例：epochs 以 10 為單位從 10 調整到 100 觀察結果)。

2.2. 嘗試過的方法

本次作業裡頭嘗試過了

- 1 到 3 層的 GRU
- 1 到 3 層的 LSTM
- 2 層 LSTM 與 Bidirectional LSTM
- Bidirectional LSTM 使用 SGD 與 Adam 兩種不同的 optimizer
- Bidirectional LSTM 使用亂數 (範圍從 -0.1 到 0.1) 與全為 0

最後決定了維持使用 Fig. 1 這個模型基於跨過 baseline 且 50 個 epochs 的訓練時間約莫為 1 小時，允許我嘗試多種 hyperparameter 的設置 (batch size、epochs)。

3. 結果

3.1. fbank 與 mfcc 比較

filter bank 的計算緣起於聲音信號的自然特徵 (由不同頻率所組成)，合併上人類耳蝸的樣式；MFCC 源起於某些演算法的先天限制，而需要使用 DCT 對 filter bank 的係數做 decorrelate。根據 [1] 的建議，如果演算法不會受到信號當中彼此高度耦合的現象影響的話，可以選用 filter bank，反之則會建議使用 MFCC。從實驗結果當中，在固定模型與 hyperparameters 的情況下，fbank 的結果總會比 MFCC 差了將近 50% 的 frame-wise 準確率 (79% 與 50% 的準確率差異)。考慮到我們使用了 LSTM 且使用了 Time Distributed Wrapper 讓模型可以針對時序信號自適應，信號本身隨時間的相依性可能會造成模型誤判結果 (因為時間上的連續已經在模型當中被考慮了)，故 MFCC 是較佳的選擇。

3.2. epochs 數

3.3. 模型

4. 參考文獻

- [1] Fayek, H. (2016). Speech processing for machine learning: Filter banks, mel-frequency cepstral coefficients (mfccs) and what's in-between. <http://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html>. (Accessed on 10/28/2017).
- [2] Graves, A. and Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural Networks*, 18(5-6):602-610.