ADL x MLDS Fall 2017

Homework 1 Report

Student: CSIE R06922068 林裕景

# Introduction

　　這次作業的目的是藉由TIMIT數據集的語音資料，利用深度學習建立能判斷其語音(phone)的模型。助教已經先對原始音檔進行feature extracting by frame，透過moving window方法擷取小段時間內的音訊特徵，提供了mfcc和fbank兩種不同方法取出的features，分別是39維和69維，而每個frame也都有標上標籤表示屬於哪種語音(共48種)。

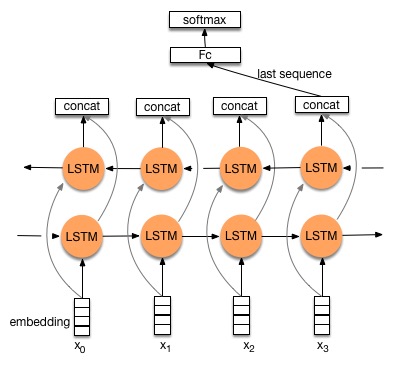
　　由於語音資料具有時間上的關聯性，因此使用RNN類型的神經網路應該會有不錯的效果，其中我選擇的普遍被廣泛使用的LSTM，許多研究指出LSTM的gate機制能使它有效的記住或遺忘不定時間長度的數值，在預測時間序列資料有顯著的效果。而在進入RNN前使用CNN，從原始資料的特徵間關聯彼此，雖然是助教要求的一種模型，但我也覺得很值得嘗試。在經過LSTM後，接39維度的Dense層加上softmax activiation function，以輸出分類的結果。在實驗的期間，我也有嘗試在LSTM和最後的Dense間增加幾層Dense，提高從時間序列學出的特徵之非線性關係，期望能找出隱藏的關聯。

　　最後，由於需要大量測試不同模型的組合，以及調整參數已找到最佳的配置，我選擇使用TensorFlow+Keras作為主要的工具進行實驗。

# C:\Users\ZeROa\AppData\Local\Microsoft\Windows\INetCache\Content.Word\greff_lstm_diagram.pngModel Description

　　作業中共設計了三個模型，第一個是單純的RNN，第二是先CNN+RNN，最後是我試過各種排列組合後找出最好的模型。

　　在我設計的三個模型中，我所使用的RNN都是LSTM(如右圖)，有著較好的能力去記憶不定時間間隔的數值，我認為對於語音資料的訓練應該會有幫助，比方說會有某個音出現後，過兩三個音節後通常會出另一個音的情況。

　　而在每個LSTM外，我還包了Bidirectional Layer (如右圖)在它的外面，藉由在時間序列上雙向的propagation，能使neurons接收到時間前和時間後的訊息，比起一般只往前傳遞的LSTM，我認為雙向能有更好的效果。

# Model Implementation

## Preprocessing

## Training

## Predicting

# How did I improve the performance of my models?

# Experiment Settings and Results