

## MLDS hw2-2 Report

### 0. Team work :

曾柏偉	程式撰寫、報告撰寫、比較結果
張嘉麟	無
劉宏國	程式撰寫，報告撰寫、比較結果

### 1. Model description :

#### A. Write down the method that makes you outstanding :

我使用 seq2seq with attention + beam search ，在使用助教 testing\_input.txt 得到的 prediction，可以得到我最好的 performance。

#### B. Why do you use it :

因為起初我直接使用 seq2seq 沒有接著 attention 時，loss 很難掉下去也 train 不起來，於是想起老師上課說目前 attention 幾乎都會加在 seq2seq model 裡面增加 model 的 performance。

#### C. Analysis and compare your model without the method :

沒有這個 attention 時，loss 幾乎都在 5.932 多附近卡住，但是加了 attention 我最好是可以下降至 2.8~v 3.2 附近，和助教最後釋出的數值感覺是相近的!!!!而且我 loss 的算法是會把 padding 的部分去除。

### 2. How to improve your performance :

最後那個 loss 是因為我使用雙向 LSTM(一層)encoder，decoder 則是基本的一層單向 LSTM，最後我上傳的 model 就是這個，使用助教的 testing\_input.txt 一萬句對話，預測產生的句子再送進去 language model 內得到 perplexity 竟然是 17.694，correlation 是 66.67!!不確定自己是否有使用錯誤，但是我自己看有時候我的 chat-bot 是可以回答到相關的問題!!!!

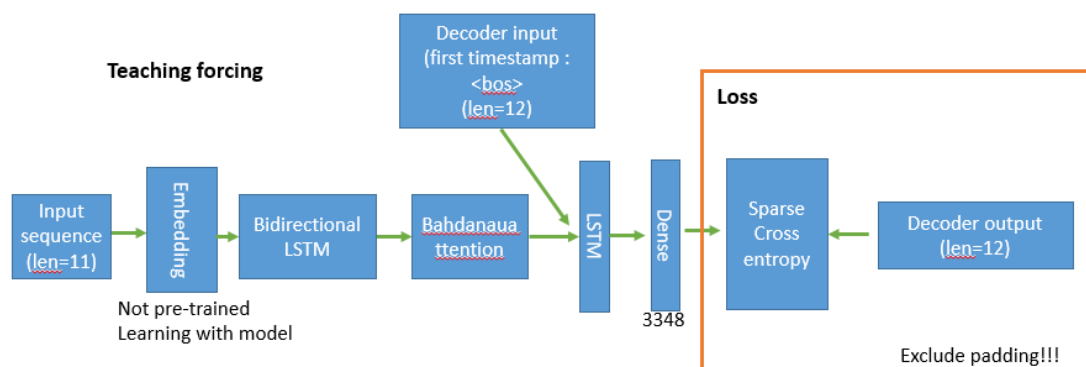
### 3. Experimental results and settings : parameter tuning, schedual sampling

先貼一個我覺得在 correlation 時所產生的句子，我是覺得對的滿好的，尤其是在你好帥那邊她竟然可以回答我的天，真的是滿合乎人性的!!! 所以先貼一個結果上來。

你好 你好 你好嗎 你是男生還是女生 我好帥 機器學習 哈哈 好無聊

你好  
你好  
很好  
生日快樂  
我的天  
是的  
哈哈  
是啊

然後在於 model 的參數設置，因為 tensorboard 實在是太複雜，看不出來 model 的架構，所以我採用別的方式：



Setting :

bidirectional units 都是 512 , attention

decoder LSTM 都是 1024 !!!

dropout : 0.5 , beam\_width = 5 !!!!

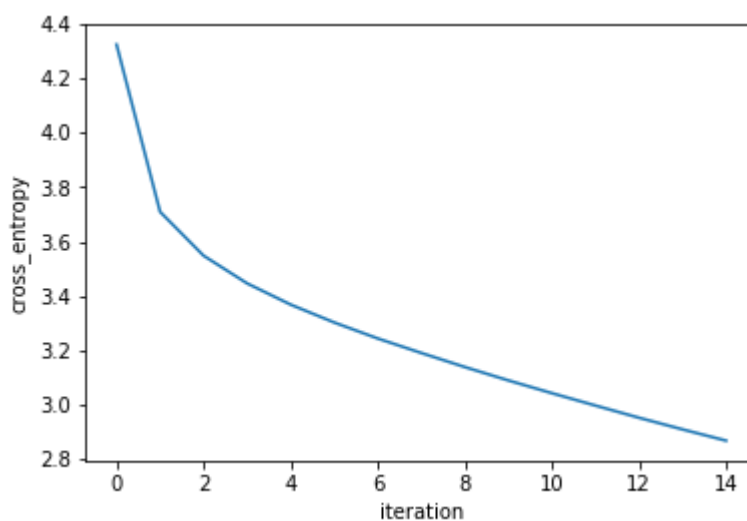
對於資料預處理上面，我這邊下了比較多的功夫去把 data 切成一個字一個字的，在中英文都是這樣，在數字上面就是選取 1996 ,20% ，就是切成我自己認知上為一個在生活上常用的字句、數字用法，在建立字典時因為記憶體及運算速度的問題，我的 min\_count = 50 ，才可以不 OOM 去做實驗。

至於 testing\_input.txt 的結果我附上幾句：

我知道	1	那 一點 也 不 難
?	2	如 你 所 知 ， " 癲 狂 " 這 東 西
這 是 什 麼 意 思 ?	3	就 好 比 一 股 引 力
機 器 人 嗎 ?	4	只 需 小 小 的 推 動 ， 人 就 飄 向 那 裏 了
你 們 好	5	登 特 ！
不 了 嗎 ?	6	戈 登 ， 你 的 人 接 走 瑞 秋 之 後
我 不 知 道	7	就 是 把 她 帶 到 了 這 兒
她 死 了	8	她 就 是 在 這 兒 死 的
我 不 知 道	9	我 知 道 ， 當 時 我 也 在 這 兒
你 救 了 我	10	想 要 救 她
我 沒 有	11	可 你 沒 救 成
我 不 能	12	我 無 能 為 力
我 明 白	13	不 對 ， 你 明 明 可 以
我 不 會	14	如 果 你 早 點 聽 我 的
我 什 麼 ?	15	如 果 你 能 站 出 來 對 抗 邪 惡
抱 歉	16	而 不 是 同 魔 鬼 簽 下 契 約
我 不 知 道	17	我 確 實 在 對 付 黑 幫 ！
就 是 這 樣	18	如 果 你 知 道 我 失 去 了 什 麼
曾 經 是 我 曾 經	19	你 就 決 不 敢 裝 得 這 麼 正 義
我 們 都 很 好	20	你 是 否 曾 經 和 你 最 愛 的 人 說
我 不 知 道	21	告 訴 他 們 一 切 都 很 好
	22	但 你 知 道 情 況 實 際 很 糟 ？
	23	如 果 沒 有 ， 那 你 馬 上 就 要 感 受 到 了 ， 戈 登
	24	然 後 你 就 可 以 盯 着 我 的 眼 睛
	25	告 訴 我 " 很 抱 歉 " 了

其實回答的句子都很短應該是我在 data 處理方面挑選最短的 decoder output，還有一個缺點就是他很常說"我不知道"，也是滿合乎常理的。他聽不懂一開始的問句是說甚麼，就會說:我不知道!!!!

以下是 training loss 的圖：



每次 iteration 訓練的句子總共有 1034508 句，共訓練 15 epochs。