

# HW2-2 seq2seq

Environment :	Debian	CPU:	Intel Xeon E5-2650 v3
Lib :	Tensorflow 1.6	GPU:	Nvidia Tesla K80
Python:	2.7	CUDA:	8.0

## Model description

Model的參數設定:

使用 attention

layer\_number = 2

Layer size = 512

Bucket size. = [(5, 10), (10, 15), (20, 25), (40, 50)]

Batch\_size = 64

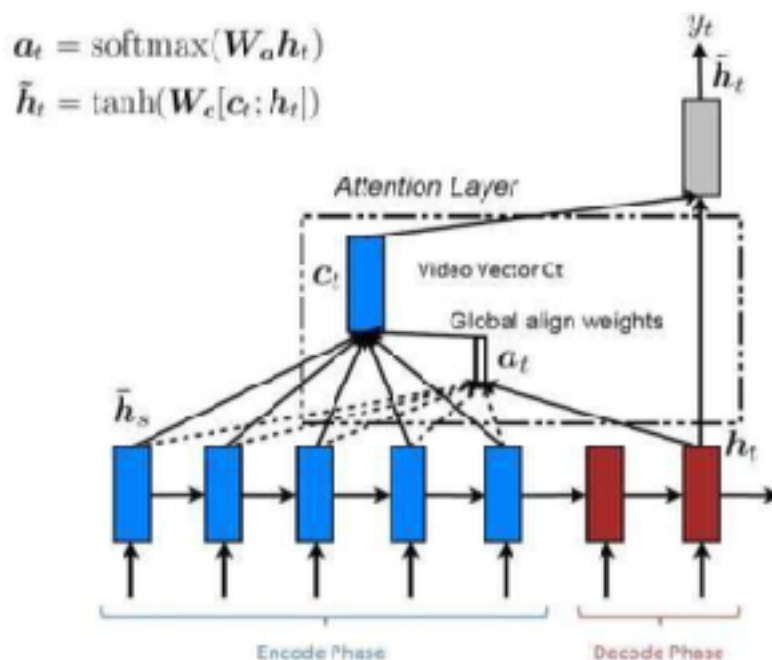
Learning rate = 0.001 往下調降

Dictionary size = 6781 以字為單位

只有使用助教提供的data中 clr\_conversation的部分，使用GRU cell  
seq2seq + attention

## Improvement

使用attention base model



這次的task遇到的第一個問題就是句子的長對不一，使得很短的句子也要做很多的padding使得memory用量很大而且，training速度會變慢。所以我們用四種不同大小的bucket分別代表不同的input與output長度，並且適當的分配對話進入bucket中，解決這個問題，我們只需要建立四種對應大小的training graph。

Model 部分我們用 s2s+ attention，大概train了半天，loss 落在 2.4左右

Inference 的時候，用的是schedule sampling，隨機挑選sample

## Experimental results and settings

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 EOS  
你 好 EOS  
你 好 EOS  
我 不 知 道 EOS  
我 們 要 去 哪 兒 ? EOS  
洲 洲 洲 洲 洲 洲 洲 洲  
我 不 知 道 EOS  
好 吧 EOS

最後我們跑evaluation的結果是

perplexity : 7.227757 (baseline: < 100)

correlation score : 0.99999 (baseline: > 0.45)

layer\_number = 2

Layer size = 512

Bucket size. = [(5, 10), (10, 15), (20, 25), (40, 50)]

Batch\_size = 64

Learning rate = 0.001 往下調降

Dictionary size = 6781 以字為單位

使用schedule sampling

## Team division

葛竑志 hw2-1

塗是澂 hw2-1

孫盟強 hw2-2