

Hw2

b03902101 楊力權 b03902093 張庭維 b03902102 廖廷浩

分工

楊力權 hw2-1, hw2-2 40%

廖廷浩 hw2-2 30%

張庭維 hw2-2 30%

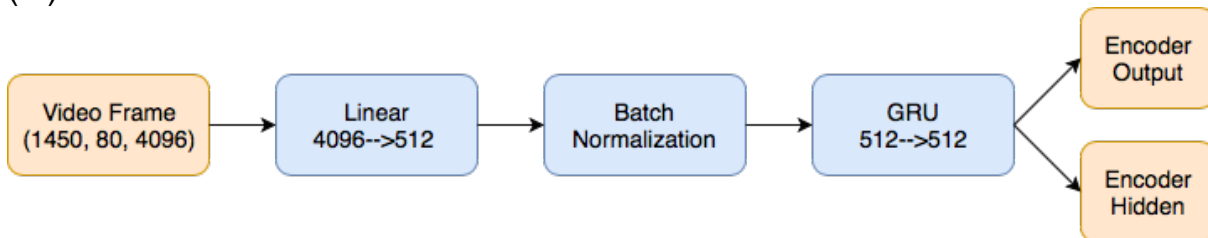
2-1 Video Caption Generation

一、Model Description

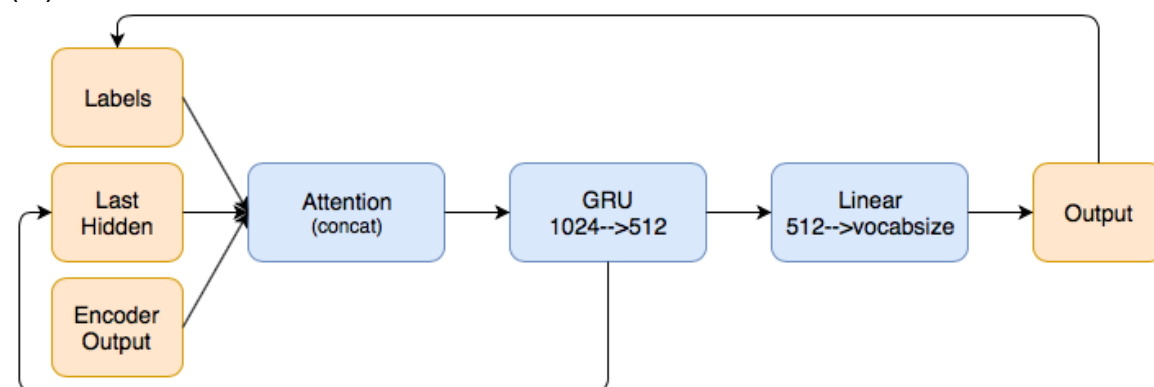
(a.)Preprocess

把每個單字都轉成小寫，只留下出現次數大於3的單字，小於等於3的全當作<unk>處理，並將每個單字都轉成一個onehot的index。

(b.)Encoder



(c.)Decoder



(d.)Training Detail

- Optimizer : Adam
- Learning Rate : 0.001
- Loss Function : CategoricalCrossEntropy

二、How to improve performance and Some Experiments

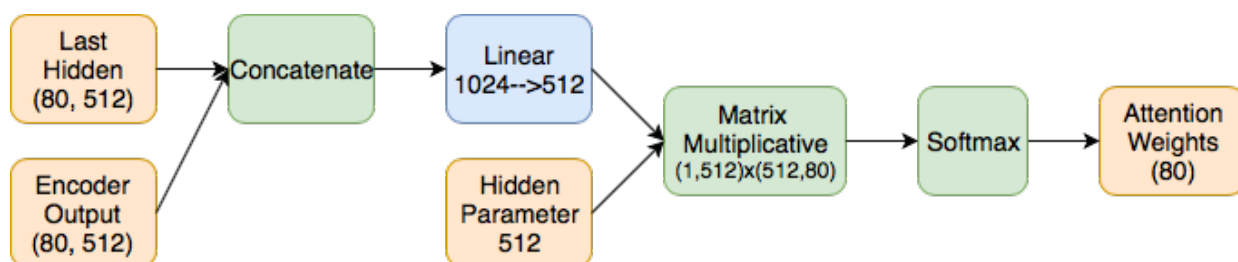
我們試了Attention與Scheduling兩個方式。

(a.)Attention

因為output往往與input sequence的output state有關係，而並非只有最後一個hidden state，因此在每個input word進入RNN前，加入Attention的encoder output。

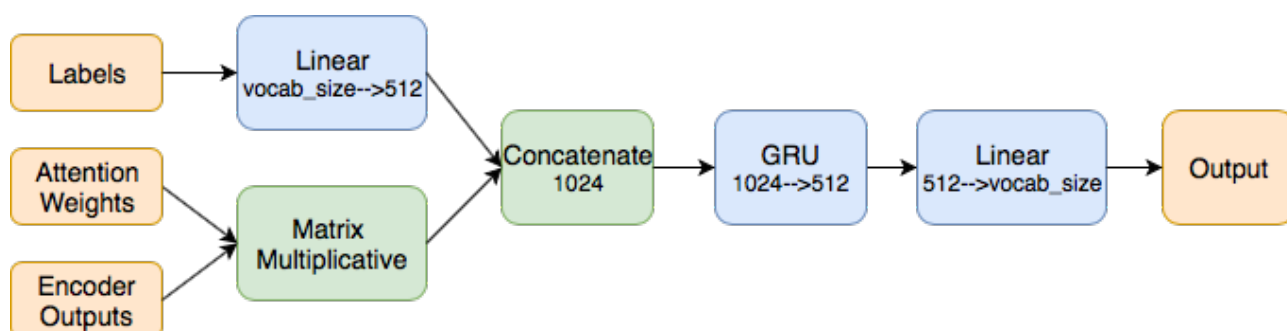
- Attention Weights

把Last Hidden與Encoder Output接在一起，對一個embedding作乘績並取softmax得到Attention Weights，也就是每個Encoder Output的權重。



- Attention Model

將Attention Weights與Encoder Output內積，並與Label相接，當作RNN中GRU的輸入。



- 結果

在同樣的條件下，dim=512，num_layer=1，使用Schedule Sampling比例0.7

	沒有Attention	有Attention
Bleu Score	0.6382	0.6945

有使用Attention的model，Bleu Score明顯比沒有使用Attention model來得好，但句意其實仍然是模糊不清。

(b.)Schedule Sampling

- 方法

因為若只有使用自己predict出來的字，會有一步錯步步錯的情況，然而全部使用 training data的label卻非常容易overfitting。因此採居中的方式在餵給RNN時的input word，有機率p會用training label的字，而有1-p用神經網路上一個預測出來的字。

- 結果與實驗

在同樣的條件下，dim=512，num_layer=1，使用Attention。

Schedule Sampling ratio (上述的p)	Bleu Score
0(全部用predict結果)	0.6102
0.25	0.6712
0.5	0.6720
0.7	0.6945

1(全部用training label)	0.6273
----------------------	--------

(c.)全部比較

	無Attention & Schedule Sampling	加上 Attention	加上 Schedule Sampling	加上Attention & Schedule Sampling
Bleu Score	0.5882	0.6102	0.6382	0.6945

可以發現，在同時使用Schedule Sampling與Attention時，可以達到最好的結果，雖然句意仍然奇奇怪怪，但是有些主詞與動詞已經比原本更正確。

三、Experimental results and settings

除了上面對Attention與Schedule Sampling的參數實驗，還有下列一些參數實驗。同樣條件下，使用Attention，使用Schedule Sampling ratio=0.7

	Bleu Score
dim=128 num_layer=1	0.6359
dim=256 num_layer=1	0.5927
dim=256 num_layer=2	0.6310
dim=512 num_layer=1	0.6945
dim=512 num_layer=2	0.6659
dim=512 num_layer=4	0.6990

• 結論

這份作業的大概在training 10 epoch內可以得到最高的Bleu Score，但是雖然在Bleu Score上的成績不差，但是其實生出的句子都是奇奇怪怪的，常常有A man is就結束的句子，只有一些會出現playing, riding, eating等動作以及potato, table, ball等受詞，不過這些也常常不是影片正確內容。