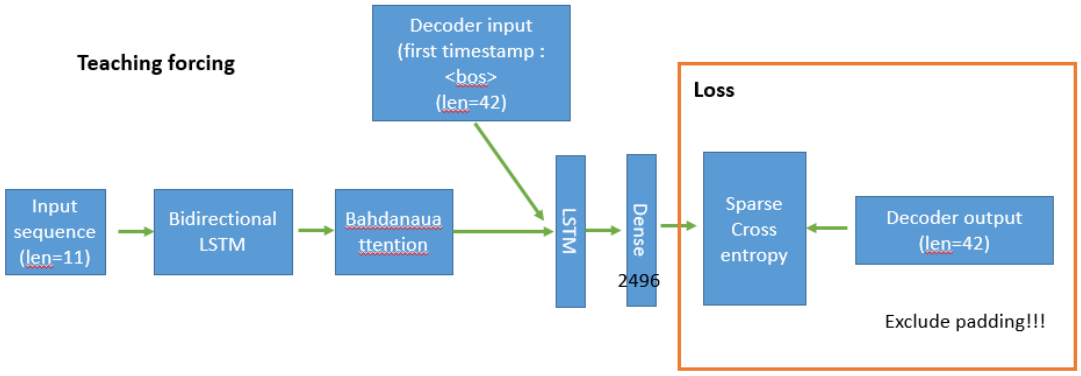
MLDS hw2-1 Report

1. 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

2. Model description：

A. Write down the method that makes you outstanding：



使用 BahdanuAttention 跟 一層雙向 LSTM，訓練至 250epochs ，得到剛好過 basline 的分數。之後加上使用 beam search， 還有一些 label 隨機配對的資料方式，才達到我最高的分數，而且產生的句子卻是較具有文法以及彈性。(後面會細談)

B. Why do you use it：

因為在使用雙向 LSTM 時，我的分數頂多超過 baseline 一點點(不知道哪邊出了問題)，所以最後才使用這個方法。但是還有一個重點地方是說: 因為 training label 每一筆都有這麼多選擇，那我到底該每個都去學還是針對一個去學呢？

C. Analysis and compare your model without the method：

- 以沒加 attention 及只使用 2 層單向的 LSTM 來說：

  model 說明：

  video feature 經過一層 GRU、

  text feature 經過 masking 跟一層 GRU、

  最後再把兩個 output concatenate。

  得分：

  bleu score is 0.379。 (keras 組員一完成)

- 以下為使用 tensorflow 並且得到較高 bleu 分數的做法：
  加了 attention 及 encoder 為雙向 LSTM，
  decoder 為一層 LSTM 的 model)
  Model 模式即為第二題圖例。
  得分：
  bleu score is 0.6037 (此時 beam width 為 1)

- 採用 beamsearch :
  提高 beam width 至 3~5 時，以及多訓練 50epochs，得到最高分為 0.63379(但是最高的總是產生答案的第一個)，之後就採用這個方向繼續去加強我們的模型。下一部分將會討我前面說的隨機配對 label 的方式。

3. How to improve your performance：

就像上述說的，我加入 beam search 時，分數也就是只能到達 0.63，我覺得說為什麼要讓機器單一學習該影片就只能配對一組回答方式呢，於是在每次 epochs 我訓練時我都會隨機為這 1450 筆 embedding vectors，隨機選擇一個 label 下去學，雖然在 loss 上面並不可觀(多了針對一個 label 去學多了 0.5 以上)，但是在實際測 bleu 時卻得到較高的成績!

4. Experimental results and settings：parameter tuning, schedual sampling

A. Experimental results

- 沒加 attention 及只使用 2 層單向的 LSTM 來說：
  得分：
  bleu score is 0.379

  句子範例：
  ScdUht-pM6s_53_63.avi
  A man is cutting a cucumber into half with a knife

  wkgGxsuNVSg_34_41.avi
  A man standing in a kettle is being face by a his hand

  BtQtRGI0F2Q_15_20.avi
  A man is riding a motorcycle with a motorcycle back as it behind him

- 以加了 attention 及使用 8 層雙向 LSTM 的 model 來說：

  得分：

  bleu score is 0.7

  句子範例：

  ScdUht-pM6s_53_63.avi
  A woman is playing a
  wkgGxsuNVSg_34_41.avi
  A woman is a a
  BtQtRGI0F2Q_15_20.avi
  A woman is a a

  針對上面的結果，我抱持懷疑，但是分數卻是我最高的….

- 以下是 使用 beach seach 以及隨機 label 配對產生句子的範例 ：

```
q7pOFn8s4zc_263_273.avi,A man is talking
mtrCf667KDk_134_176.avi,A woman is slicing a vegetable
Olh_UWF9ZP4_62_69.avi,A woman is stirring some thing
JntMAcTlOFO_50_70.avi,A man is walking in a forest
7NNgO_n-bS8_21_30.avi,A band is performing
IhwPQL9dFYc_124_129.avi,A woman is putting a block of tofu
BAf3LXFUaGs_28_38.avi,A man is singing
6q1dX6thX3E_286_295.avi,A man is talking
```

  竟然會有 tofu 這個詞在上面，可見這真的是不錯的結果。

B. Settings training detail

此處參數設定以較佳的 model 為例：

- 字典大小：training caption 全部 43000 多句子，min_count =3 (參考助教)
- 一層雙向 lstm(encoder)、單向 LSTM(decoder)、attention unit = 256
- batch_size = 128
- dropout 0.5
- training epoch 250
- optimizer：Adam (learning_rate : 0.001)

下圖為訓練過程的 Loss: