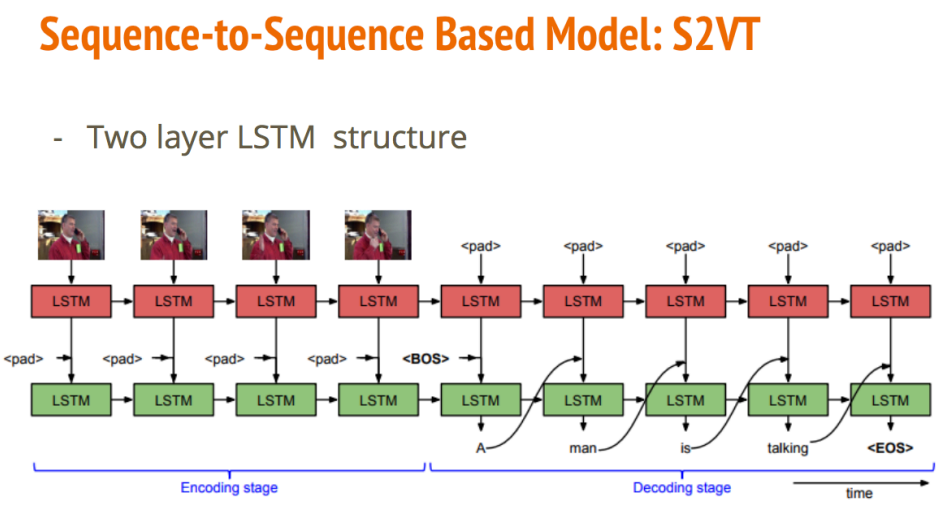
ADLxMLDS2017 HW2 report

姓名：徐有慶
學號：R05922162

1. Model description

- 本次作業中需實做一個 sequence to sequence based model，輸入為一段影像，輸出為描述這個影像的一句話，輸入的影像為助教前處理過後，80 個 4096 維的 frames，而輸出的每個字，使用 one-hot vector encoding 表示
- 參考助教投影片所提供的 S2VT model，建出兩層的 LSTM model，其中，前半部份為 Encoder，後半部份為 Decoder，如下圖所示



另外在第一層 LSTM 前加上 embedding layer，將 4096 維轉換成和第一層 LSTM 的 hidden layer size 一樣大，也在第二層 LSTM 後加上 projection layer，將第二層 LSTM 的輸出轉換成和 vocabulary size 一樣大

2. Attention mechanism

- 在 encoder 部分，將第二層 LSTM 中每個 timestamp 的 output 堆疊起來，產生一矩陣 Hs，大小為(80, hidden layer size)，在 decoder 部分，將第一層 LSTM 每個 timestamp 的 output 記錄成 ht，大小為(hidden layer size, 1)，算出 alpha = softmax(Hs * W * ht)，大小為(80, 1)，可以得出每個 frame 各自的權重，便可以得到 context = Hs^T * alpha，將 context 當作輸入餵給 decoder 的第二層 LSTM

- **Evaluations**

  S2VT：

  Originally, average bleu score is 0.2614

  By another method, average bleu score is 0.574

  S2VT + attention：

  Originally, average bleu score is 0.2726

  By another method, average bleu score is 0.5977

  如果單看 BLEU 分數的話，加上 attention 後效果的確有比較好，不過出來的句子沒有很顯著的差異。其實也不知道自己的作法是不是正確的，如果看上課的投影片會覺得要採用 LSTM 出來的 hidden state 來做，但看大部分的做法都是使用 LSTM 的 output 來做，而要在哪個時機點加上 LSTM 也是個搞不太清楚的部分，可能的原因應該跟不清楚為什麼原先的 S2VT 要那樣建置有關係，因為也只是參考助教提供的 model 就建出來了，沒有實際理解他為什麼要這樣建

3. How to improve my performance

- **Schedule sampling**

  在 training 時，原先 decoder 的部分，考慮第二層 LSTM 的第 i 個 timestamp Ti，會先將 Ti-1 的 ground truth 和第一層 LSTM Ti 的 output 串接在一起，再一起餵進 Ti，以此類推。而 testing 的時候，由於沒有 ground truth，所以會將 training 時 Ti-1 的 ground truth 改為第二層 LSTM Ti-1 預測出來的 word。但這樣的做法在 testing 的時候，因為是使用預測出來的 word，如果預測錯的話可能導致後面的結果也出錯，因為在 training 的時候沒有看過這種情況。

  所以使用 schedule sampling 的方式，將原來 training 餵 ground true 的部分更改為，先隨機產生一個 0-1 的數，如果大於 0.5 餵則 ground true，反之則餵預測出來的 word，這樣的做法可以讓 training 時也能夠訓練到如果是餵預測出來的 word 要怎麼處理，那在 testing 時結果也會較佳。

4. Experimental results and settings

- **Experimental settings**

  Training Epoch = 200

  Batch size = 64

  Optimizer = AdamOptimizer

  Loss function = softmaxt cross entropy with logits

  LSTM dimension = 256

  Start learning rate = 0.001

  vocab size = 6143

- **Experimental results**

  S2VT：

  Originally, average bleu score is 0.2614

  By another method, average bleu score is 0.574

  S2VT + attention：

  Originally, average bleu score is 0.2726

  By another method, average bleu score is 0.5977

  S2VT + schedule sampling：

  Originally, average bleu score is 0.2852

  By another method, average bleu score is 0.644