

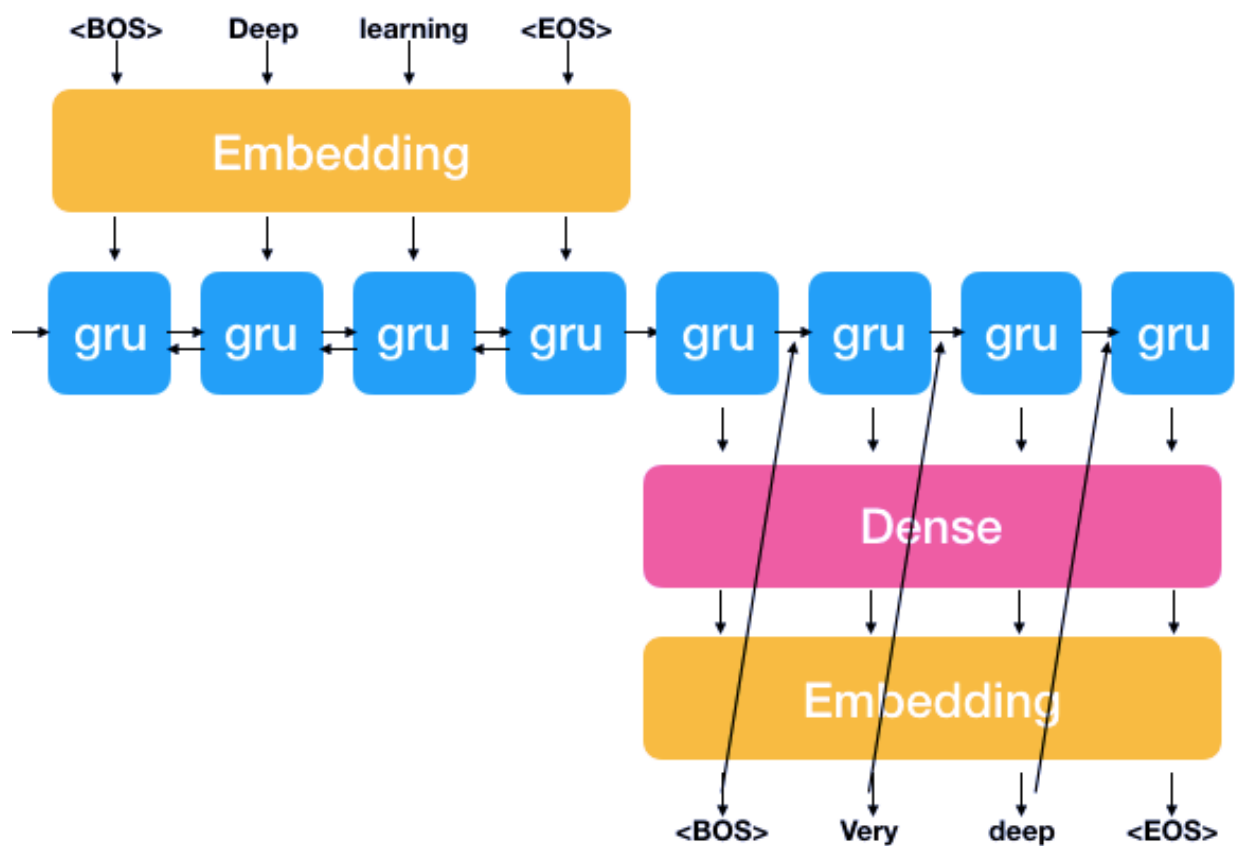
Chatbot

學號：b05902031, 系級：資工二, 姓名：謝議霆, 學號：b05902008, 系級：資工二, 姓名：王行健

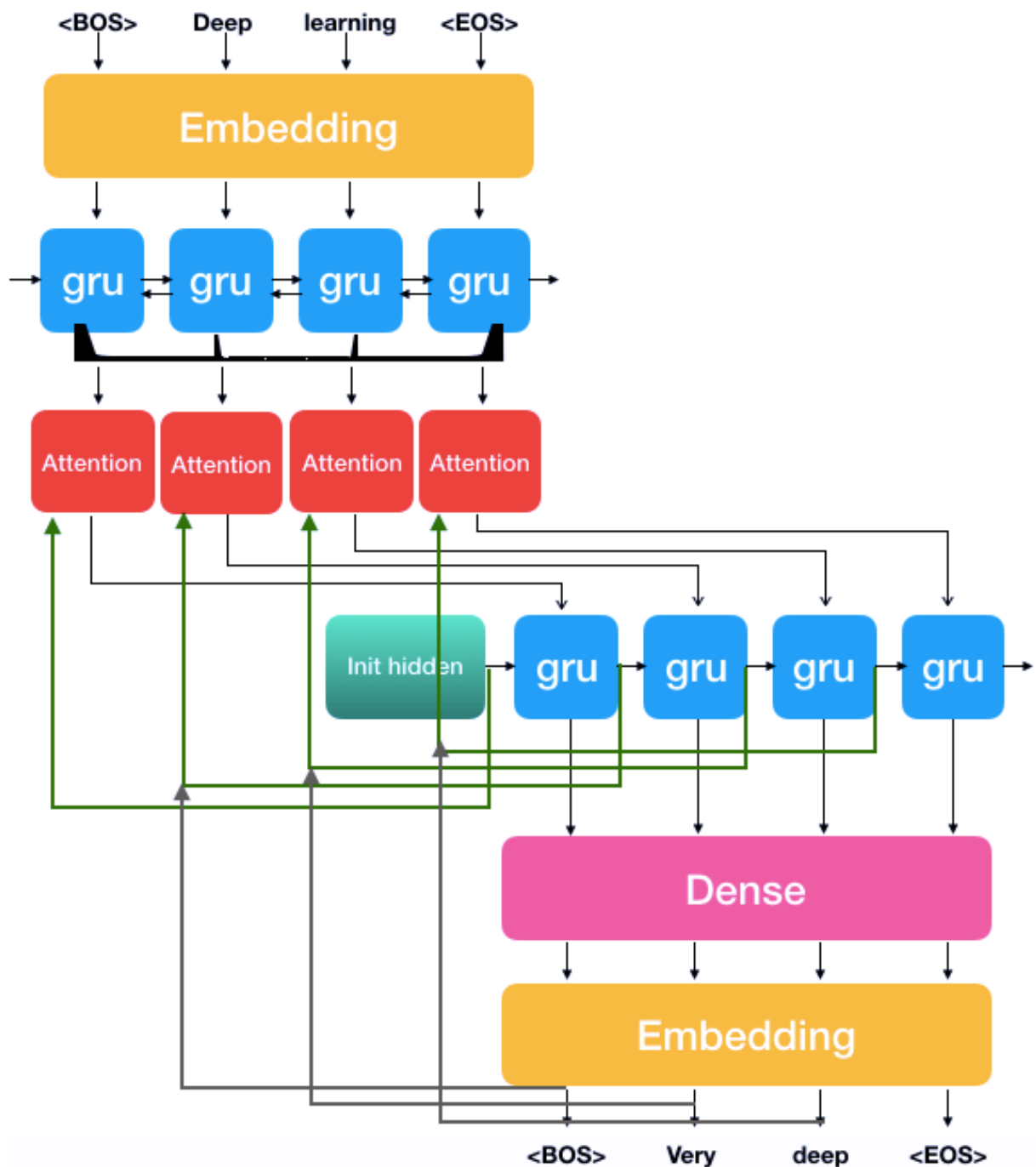
Model description

我們測試了兩種模型，相異點是有沒有使用attention

- 沒使用attention

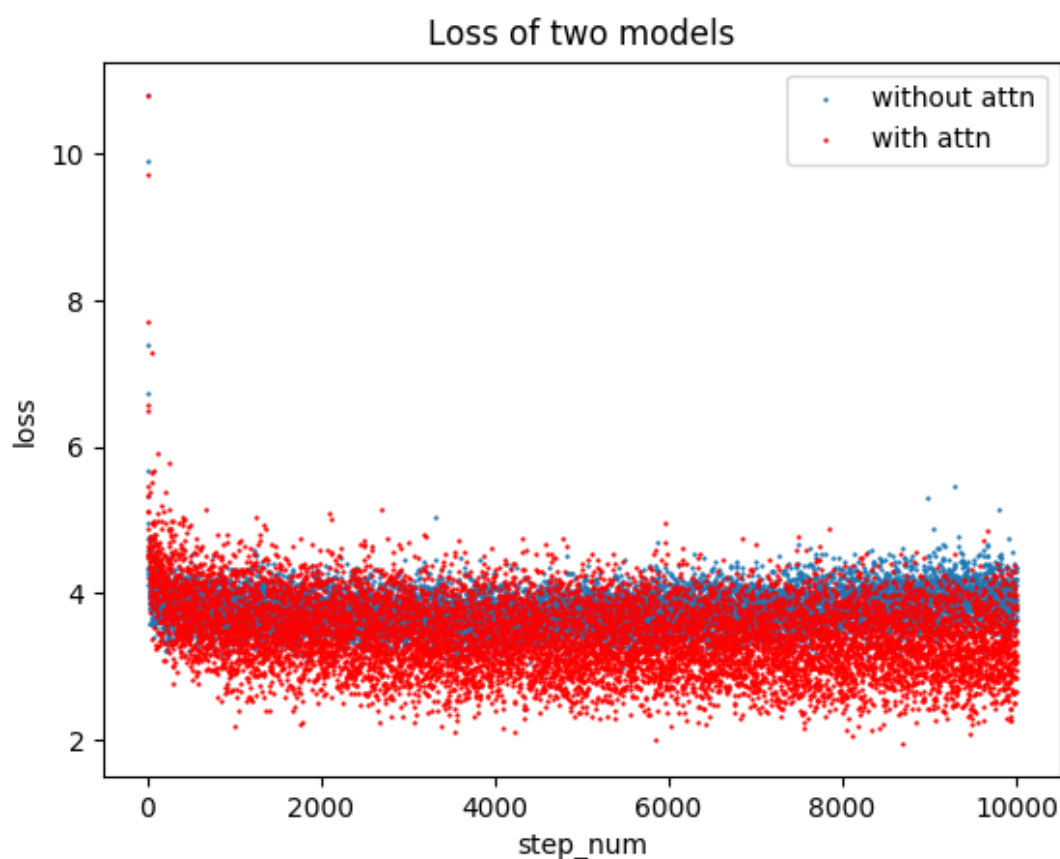


- 有使用attention



How to improve your performance

- 我們在這次的model中使用attention，從encoder中的所有output，數量有輸入至encoder的長度(最長句子長度)，拿前一次的decoder hidden和前一個字的embedding當資訊，從encoder的output中，挑選一組詞的資訊來做下一次的decode。
- 在使用第一種最簡單的model時，輸出的句子有很多是"我 是"、"你 的"，原因是訓練資料中就有許多這樣的句子，所以model理當會傾向這些輸出，而忽略了原句子裡面重要的部分。
- 以下是有使用attention和沒使用attention中loss的比較



雖然不明顯，但可以看出有使用attention的loss可以降的比較低，不過從perplexity和correlation的分數來看

	有attention	無attention
Perplexity	34.488438	34.814837
Correlation	0.49576	0.44365

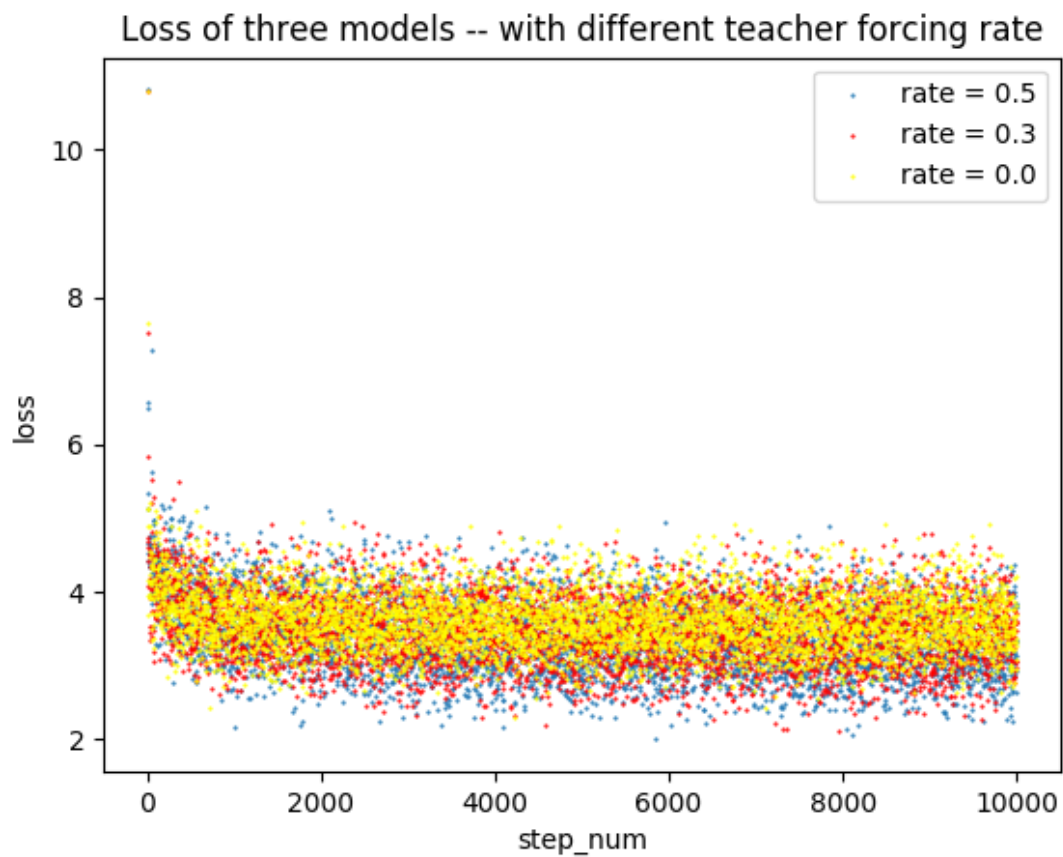
- 會有這樣的差別還有一個可能的原因是兩個模型的參數量不同
 - 有使用attention的參數量： $81012000 + 131847516 = 212859516$
 - 沒使用attention的參數量： $32514000 + 116071004 = 148585004$
- 除了attention，我們還使用了teacher forcing，這樣就model比較少會說出重複的詞彙，比較的部分在第三部分(experimental results and setting)
- 回答(predict)的方面，會發現有些句子在model輸出"我"之後，很難說得下去，雖然這方面的情況在使用attention之後已經有改善了，但是還是有一些，所以我們會適當得在回答不好的情況下，用輸入的第一個詞來回答接下來的句子，讓model比較能準確的回答問題。

Experimental results and setting

- Embedding size = 1500
 - 相較於2_1的英文句字，中文的詞明顯多很多，一開始我們將embedding size設成跟2_1一樣，訓練了很久都很難在訓練資料中講出難一點的詞和複雜一點的句子，將embedding size調高之後這方面的缺點就改善了許多。

- Encoder GRU hidden size = 500, encoder GRU with bidirectional
- Decoder GRU hidden size = 1000
- Batch size = 16
- Optimizer : Adam, learning rate = 0.001
- Iteration = 300000
- 最長句子長度：12
- Teacher forcing rate = 0.5

Different teacher forcing rate : 從loss可以稍稍看出，在rate = 0.0的情況下，loss是普遍偏高的，在回答training data方面，也幾乎都是"我 我 我"類似的回答。



分工表

B05902008 : 資料處理，report中測試的圖片，參數調整

B05902031 : 模型架構，程式編寫