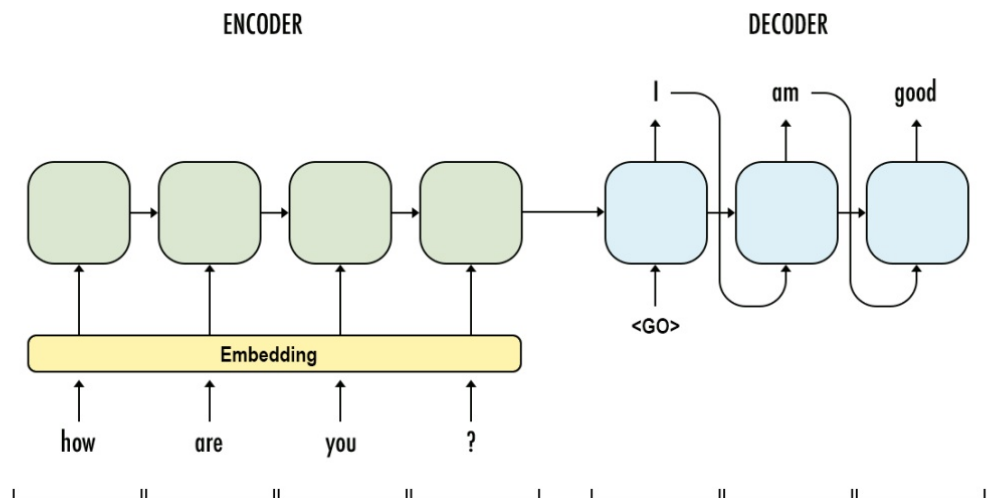


## 2-2

- Model description (2%)
  - Describe your seq2seq model



我使用的 model 是最基本的 seq2seq。encoder 先將 input 句子通過 embedding，轉成以字為單位的 one-hot embedding，再輸入 GRU（不輸入最後的 <PAD>），得到最後的 hidden state，作為 decoder 的 initial hidden state。decoder 再以 <BOS> 為 input，一個一個字 output，直到 <EOS>。

- How to improve your performance (3%)  
(e.g. Attention, Schedule Sampling, Beamsearch...)
  - Write down the method that makes you outstanding (1%)
  - Why do you use it (1%)
  - Analysis and compare your model without the method. (1%)
    1. Teacher Forcing  
使用 teacher forcing 可以使 model 較好訓練，loss 掉的比較快，比較快收斂。除此之外，teacher forcing 也可以讓輸出的句子有較好的文法。若沒有用 teacher forcing，model 較難訓練且輸出的句子通常沒有什麼文法。
    2. Beam Search  
用 beam search 的目的是希望能找到最大機率的句子，但用 beam search 輸出的句子，文法變得比較差，correlation 也沒有變好，整體效果反而比原本還差。
    3. pad\_packed\_sequence  
training data 中會有句子長短不同，結尾會加上 <PAD> 的情況，而我們會預期機器會學到此情況，使得 encoder 不會受到最後那些 <PAD> 的影響。但我實作中卻發現，將所有句子 pad 成同樣長度下去訓練，loss 無法降下來，一直卡在 20 左右。pytorch 中的 pad\_packed\_sequence 這個函式可以讓 rnn 不要 input 最後的那些 pad，使用這個函式之後，loss 就可以降到 3 左右了，是我可以成功訓練的最大原因。
- Experimental results and settings (1%)

- parameter tuning, scheduled sampling ... etc

Parameter	
GRU hidden size	512
batch size	200
Adam learning rate	0.001
dropout	0.1
vocabulary size	3000 (單位：字)

### **Experiment results:**

1.

	use teacher forcing	no teacher forcing
perplexity	41.77	267.09
correlation	0.319	0.296

有用 teacher forcing 的話整體都比較好，沒有用的話比較難訓練，結果也比較差，尤其是文法。

2.

	not using beam search	beam search
perplexity	41.77	212.67
correlation	0.319	0.293

用 beam search 的文法變得很差，但 correlation 卻沒有變好。

分工表：

2-2: 朱柏澄