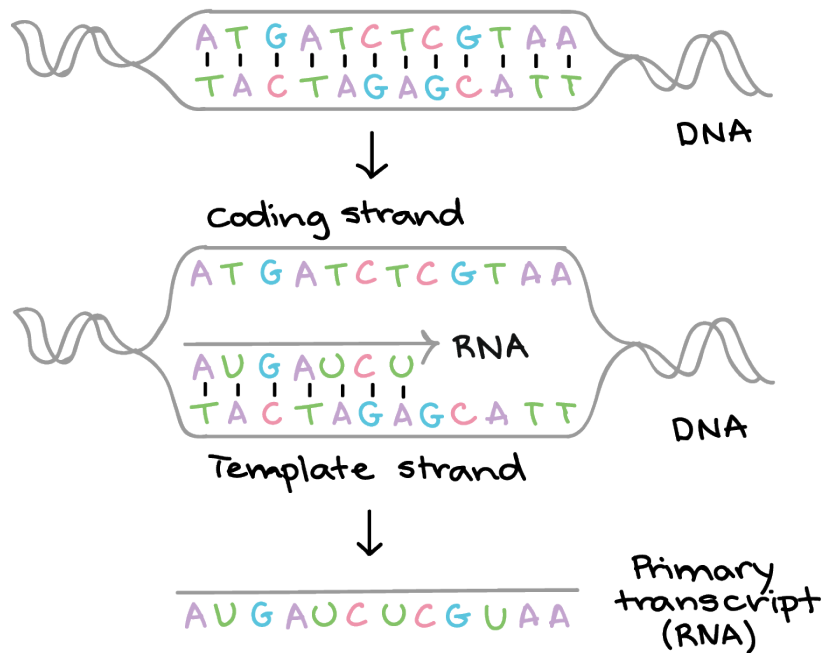




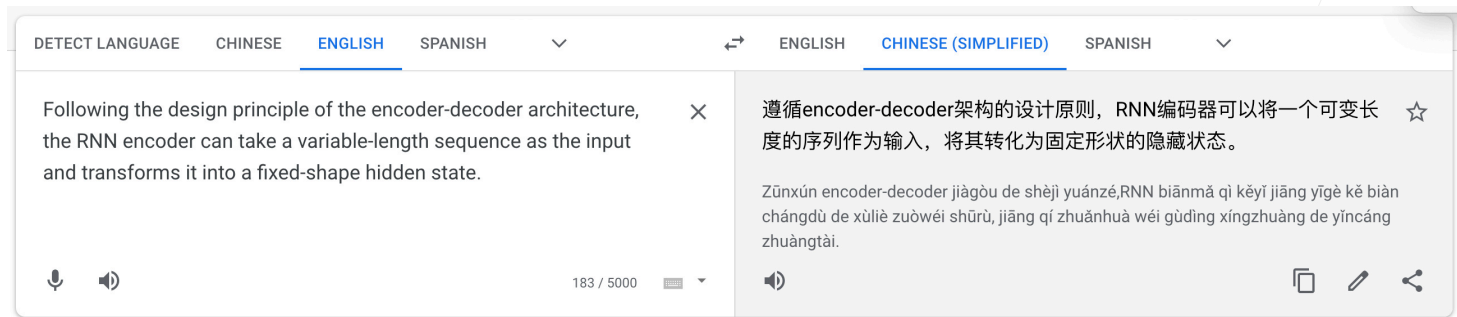
序列到序列学习 (seq2seq)



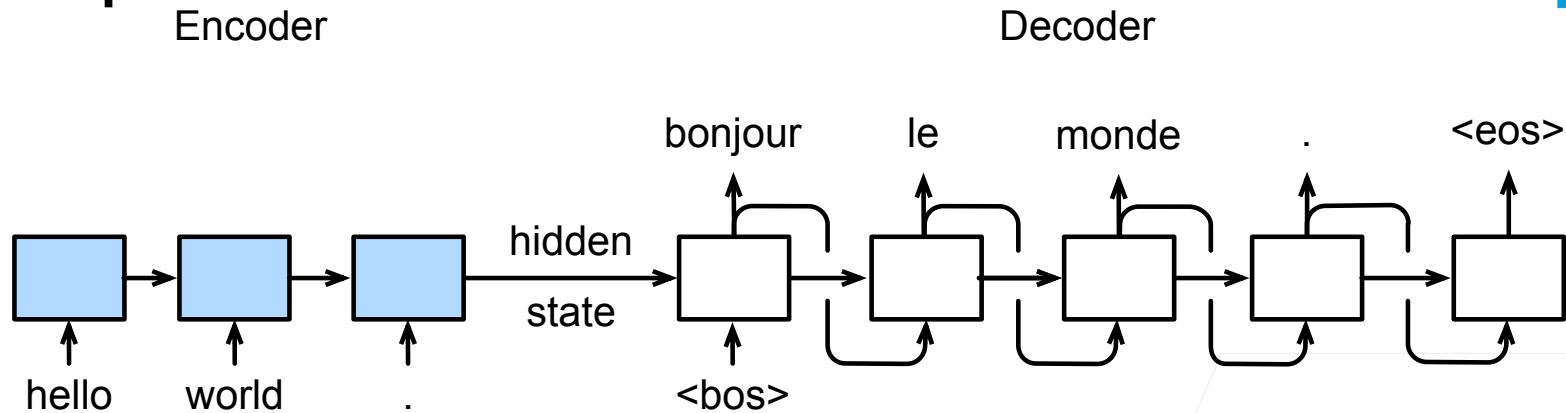
机器翻译



- 给定一个源语言的句子，自动翻译成目标语言
- 这两个句子可以有不同的长度

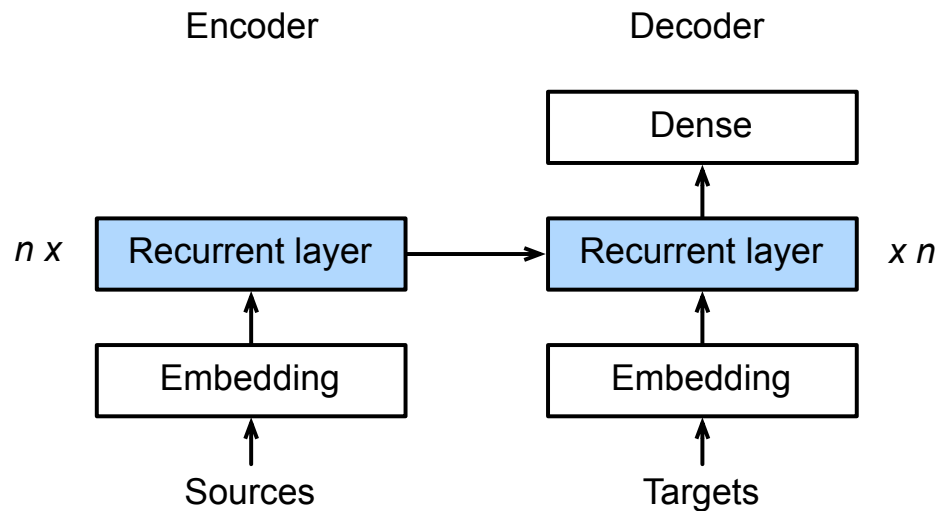


Seq2seq



- 编码器是一个RNN，读取输入句子
 - 可以是双向
- 解码器使用另外一个RNN来输出

编码器-解码器细节

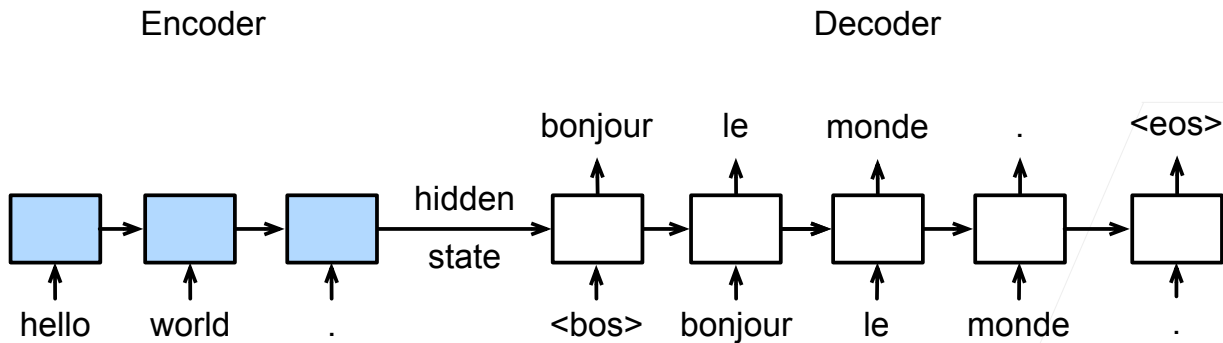


- 编码器是没有输出的RNN
- 编码器最后时间步的隐状态用作解码器的初始隐状态

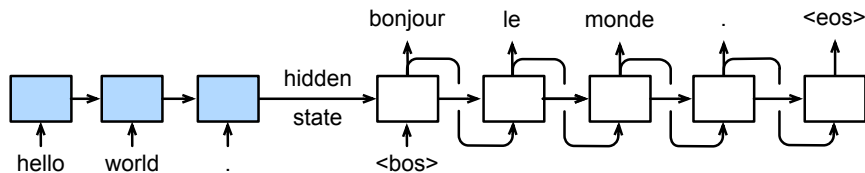
训练



- 训练时解码器使用目标句子作为输入



推理:





衡量生成序列的好坏的BLEU

- p_n 是预测中所有 n-gram 的精度

- 标签序列 $A B C D E F$ 和预测序列 $A B B C D$, 有

$$p_1 = 4/5, p_2 = 3/4, p_3 = 1/3, p_4 = 0$$

- BLEU定义

$$\exp \left(\min \left(0, 1 - \frac{\text{len}_{\text{label}}}{\text{len}_{\text{pred}}} \right) \right) \prod_{n=1}^k p_n^{1/2^n}$$

当 p_n 固定时, 会随着 n 的增长而增加

为什么?

预测的序列越短, 获得的 p_n 值越高

标签: ABCDEF
预测: AB
 $p_1 = p_2 = 1$

惩罚过短的预测

长匹配有高权重



总结

- Seq2seq从一个句子生成另一个句子
- 编码器和解码器都是RNN
- 将编码器最后时间隐状态来初始解码器隐状态来完成信息传递
- 常用BLEU来衡量生成序列的好坏