

1. Attention model intuition

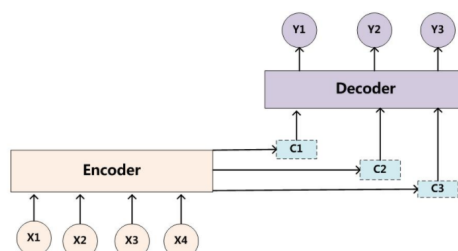
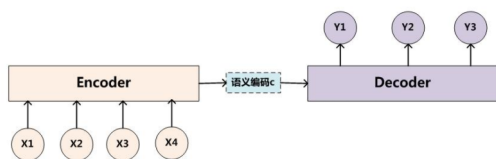
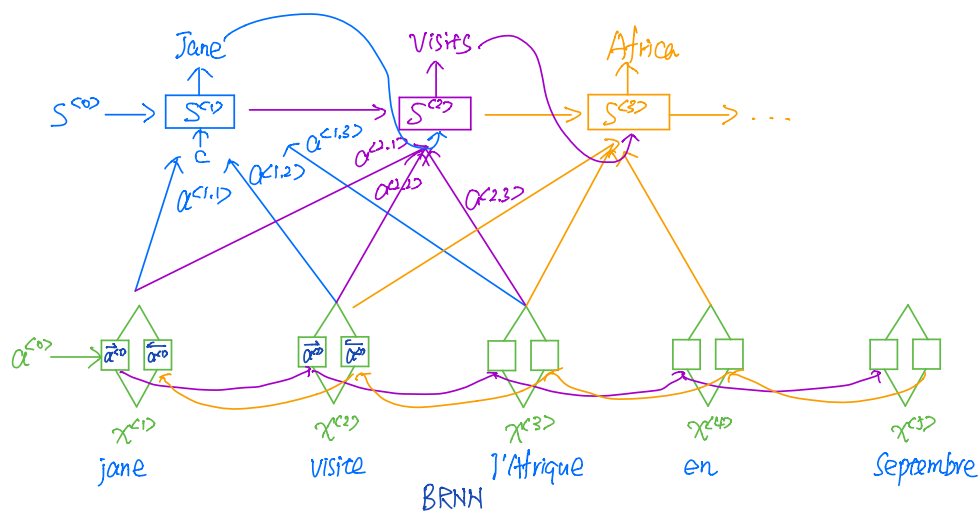


图2 引入AM模型的Encoder-Decoder框架

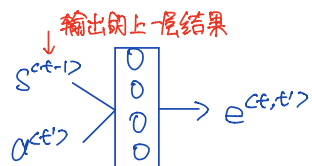
2. Detail

$$\alpha^{cs} = (\frac{1}{a} \cos, \frac{c}{a} \cos)$$

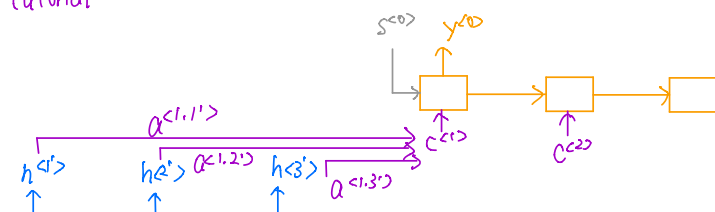
$\sum_t \alpha^{t,t} = 1$ $\alpha^{t,t} =$ amount of "attention" that y^t should pay to x^t

$$C^{(D)} = \sum_{t \in T} a^{(1,t)} a^{(t,3)}$$

$$a^{(t,t')} = \frac{\exp(e^{(t,t')})}{\sum_{t'=1}^T \exp(e^{(t,t')})}$$



Personal tutorial





Step1: 现在输入进Decoder中的不再是一个整体编码, 而是一个权重值, 我们先看 $c^{(1)}$,

$c^{(1)}$ 定义为与 $h^{(1,t)}$, $a^{(1,t)}$ 有关, t 代表 Encoder 的某一时刻, 注意有上标 t

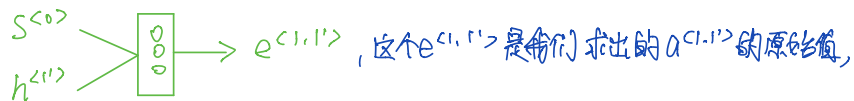
$h^{(1,t)}$ 定义为 Encoder 某一时刻的 hidden state, 为 more intuition. 设为 (1×30)

$a^{(1,t)}$ 代表 Decoder 第 1 时刻的输入应内别对 Encoder 的 t 注意多少.

最后 $c^{(1)} = \sum_{t=1}^T h^{(1,t)} \cdot a^{(1,t)} = h^{(1,1)} \cdot a^{(1,1)} + h^{(1,2)} \cdot a^{(1,2)} + h^{(1,3)} \cdot a^{(1,3)}$ (后面有解释怎么算) ①

问题来了, 怎么求 $a^{(1,t)}$ 呢? 先别一上来就 softmax, 先说 $a^{(1,1)}$ 怎么求?

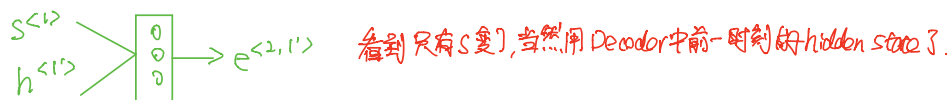
$a^{(1,1)}$ 与 $h^{(1)}$ 和 $s^{(0)}$ 有关, 记住 $a^{(1,t)}$ 与 $h^{(1,t)}$ 和 $s^{(t-1)}$ 有关, 先记住我们把 $h^{(1)}$, $s^{(0)}$ 扔进 NN 中, $s^{(0)}$ 也是 (1×30)



同理我们求出 $e^{(1,2)}$, $e^{(1,3)}$, 属于实数 R , 为了进 softmax, 将 $e^{(1,t)}$ 转为概率, $a^{(1,t)}$ 当我们拥有 $a^{(1,t)}$ 后, 就算出 $c^{(1)}$, 好了结合 $s^{(0)}$ 扔进第一个 RNN cell 中, 等待 $y^{(0)}$

接着, 我们要求 $c^{(2)}$

$c^{(2)} = \sum_{t=1}^T h^{(2,t)} \cdot a^{(2,t)}$, 求 $a^{(2,t)}$ 要求 $e^{(2,t)}$

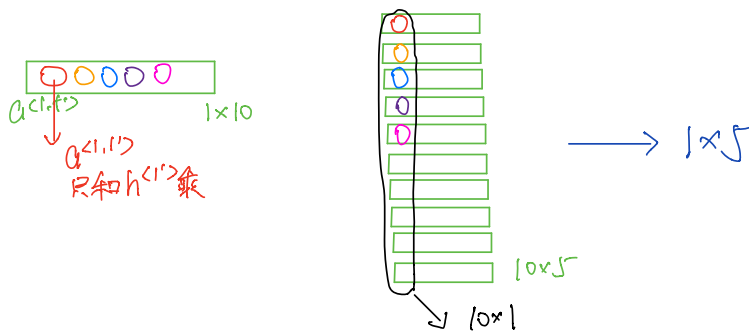


① 解释

$a^{(1,t)}$ 是一个数字, $h^{(1,t)}$ 为 $(1 \times \text{hidden-size})$

把所有 $a^{(1,t)} \cdot h^{(1,t)}$ 加起来怎么求?

假设句子长度为 10, i.e. $T_x = 10$, 那我们有 10 个 $a^{(1,t)}$, 假设 $\text{hidden-size} = 5$



第一次矩阵乘 $(1 \times 10) \cdot (10 \times 5)$

各自 $a^{(1,t)}$ 和各自 $h^{(1,t)}$ 乘, 乘完还加起来, 完全做完就实现了 \sum