

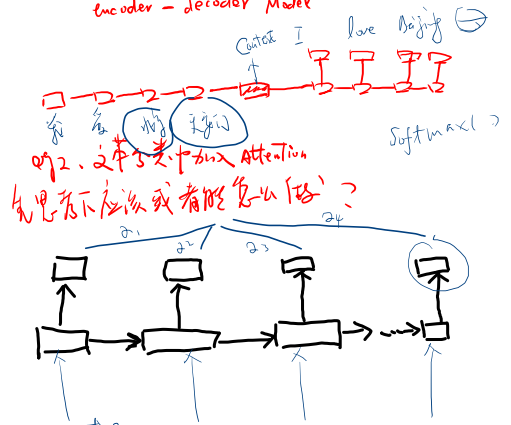
Attention 是什么:

1. 注意力机制: 在文本/图像/语音/广告等领域, 可能涉及神经网络中, 可能涉及到的信息有很多, 但是哪些应该聚焦, 哪些是不需要记住的信息.

2. 注意力机制的几种方式: ——

3. 神经网络中最经典的例子.

encoder-decoder Model



例2. 文本分类中加入 Attention
先看下应该或者能怎么做?

$$\text{Softmax}(h_{\text{total}}) = \sum_i \alpha_i h_i$$

Attention 的几种不同的形式:

1. 从权重分布2的计算方式角度.
2. attention q, v, k 的维度 \Rightarrow Self-Attention
3. 范围角度. 4. hard/soft.

$$\alpha = W^T \tanh(H)$$

$$[h_1, \dots, h_n]$$

1. 权重分布2的计算方式角度:

- 1.1. 内积: $\alpha_i = v^T u$
- 1.2. 全连接: $\alpha = V^T \tanh(Wu + b)$
- 1.3. 双线性形式: $\alpha = V^T W U$

注意: 这里的22维计算的是 attention 形式的具体. 这里问: 如何计算又线性?

$$= V^T (U^T W) U$$

$$= (VU^T) (WU)$$

2. q, k, v 的维度. 只需要一种

3. ? 4. ? self-attention

Answer: ——

