

LSTM GRU

机器翻译的小技巧

RNN (从右往左)

序列生成模型

CNN 只能看到局部信息

BLEU

但可以做多个输出通道

Decoder:

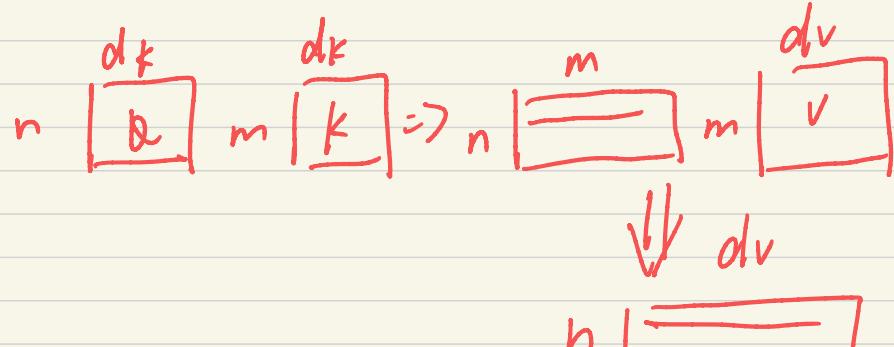
- Auto regressive
- residual connection

$d_{model} = 512$

$\underbrace{\text{R}, \text{P}, \text{P}, \text{P}}_{n}$ $\underbrace{\text{Q}, \text{K}, \text{V}}_m$ 加权

~(10.1). dk 太大时造成梯度消失

Attention($\text{Q}, \text{K}, \text{V}$): $\text{softmax}\left(\frac{\text{Q}^T \text{K}^T}{\sqrt{dk}}\right) \text{V}$

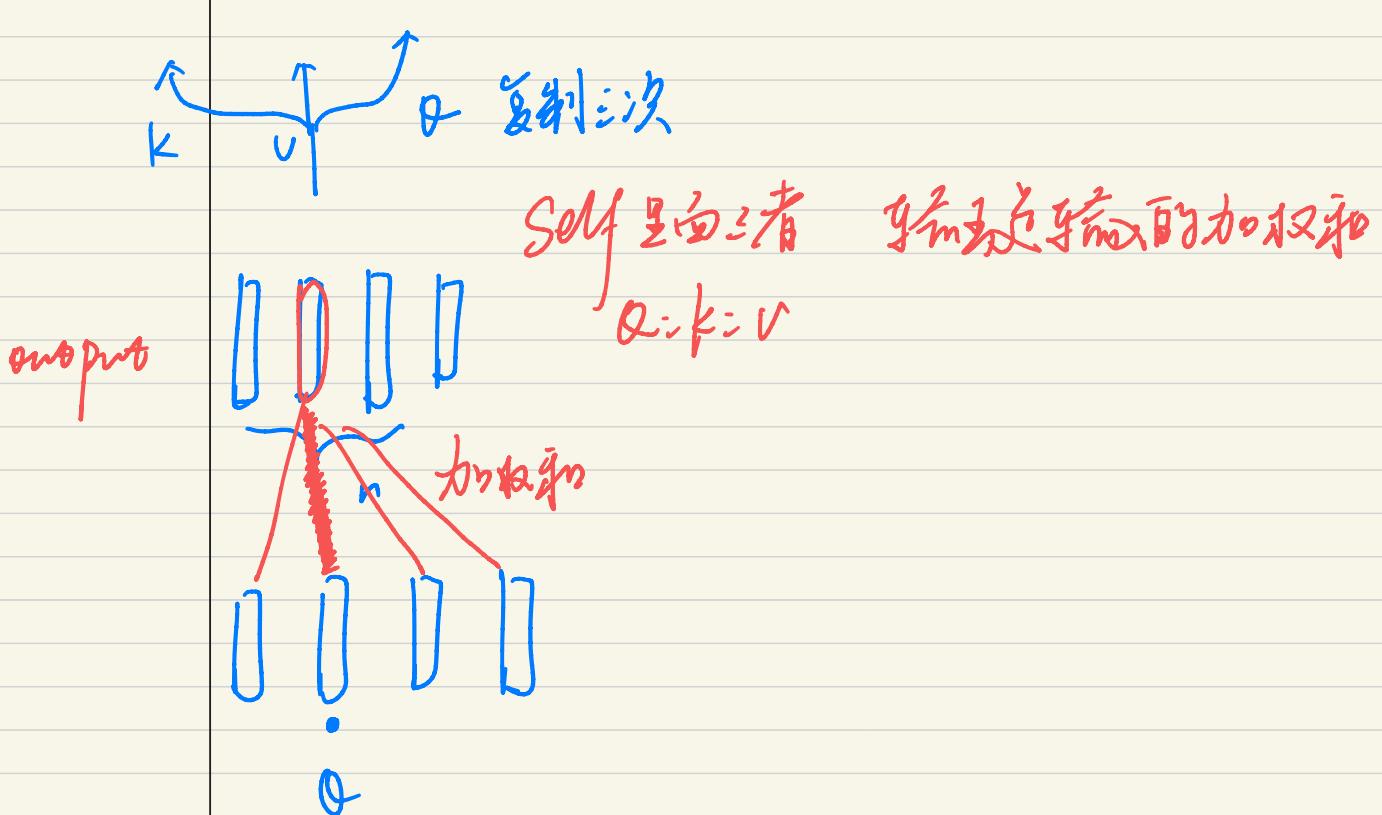


MASKED

$k_1, \dots, k_{t-1}, \underbrace{k_t, k_{t+1}, \dots, k_n}_{\text{masked}}$

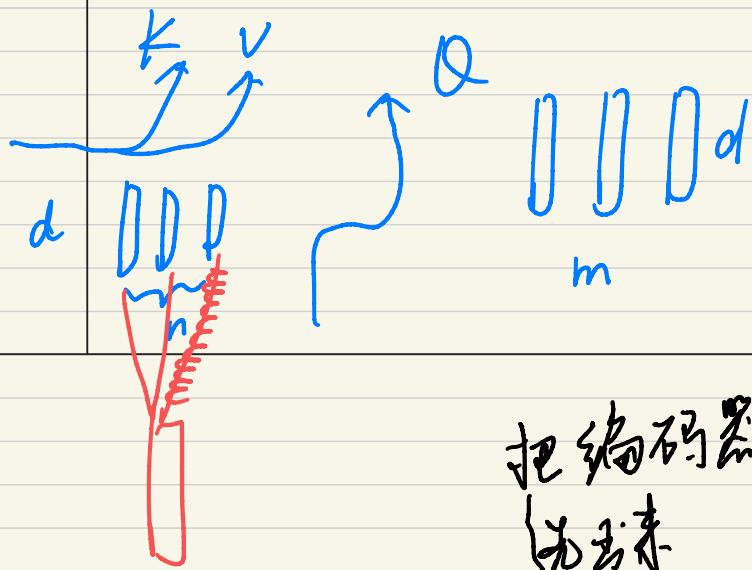
$\hookrightarrow -\infty \xrightarrow{\text{softmax}} 0$

encoder

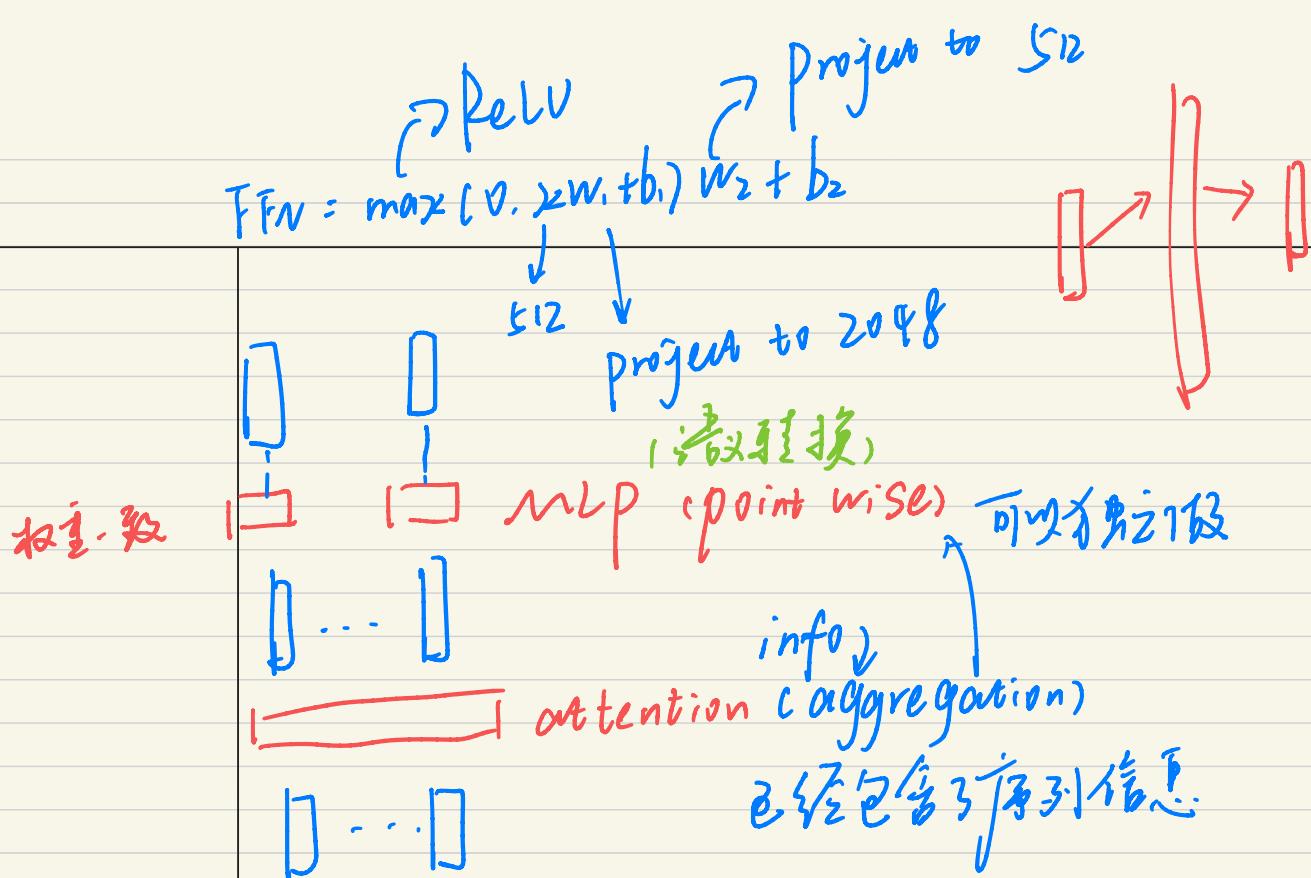


Decoder:
masked

encoder - decoder



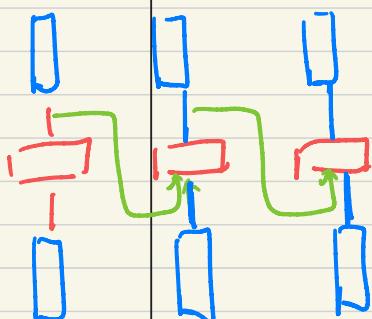
把编码器中的信息提供给 decoder
通过去



Rnn

embedding

$\times \sqrt{dmrde}$ 方便加口 Dos-camb

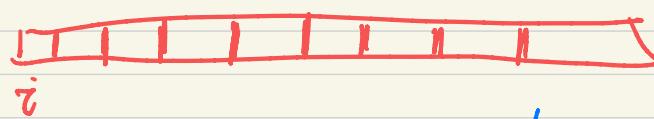


Pos-emb

Attention 不包含时序信息

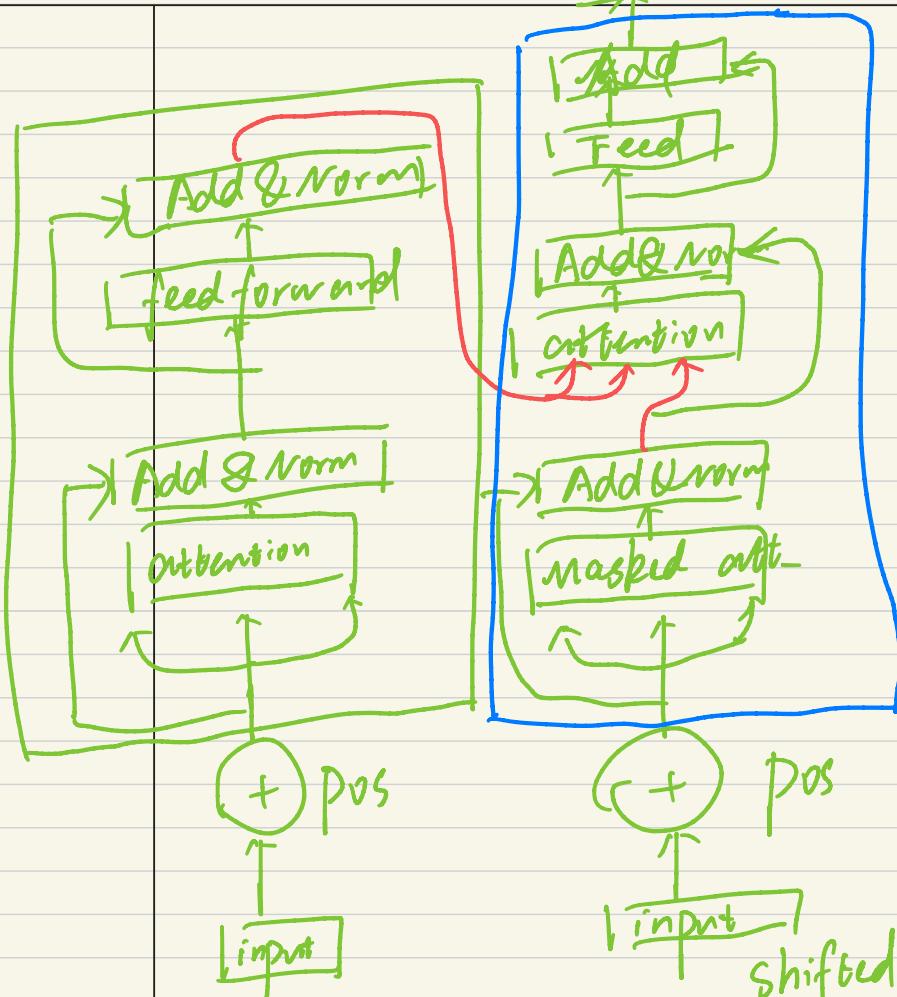
输入输出序列信息

512

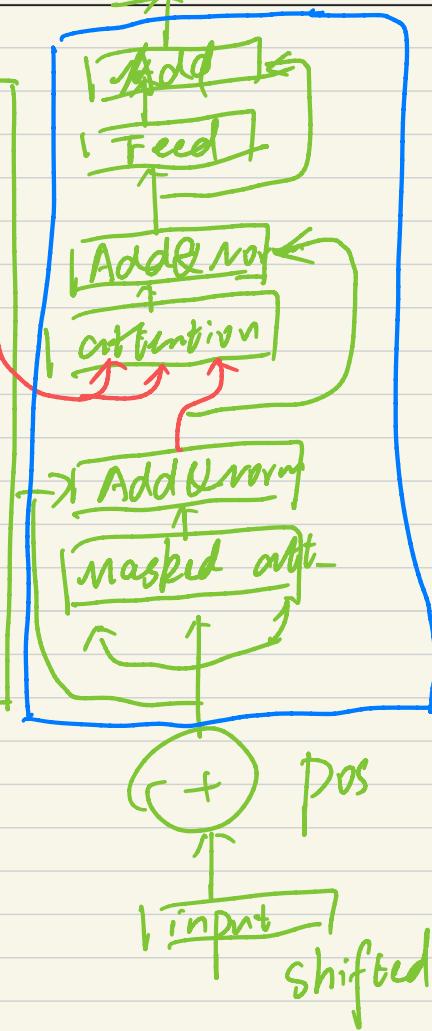


$$P_6(pos, z) = \sin\left(\frac{pos}{10000}z\right)/d_{model}$$

$$PE(pos, z_i+1) = eosC(pos) / \sqrt{d_{model}}$$



encoder



decoder

$$\begin{aligned} &\text{Softmax} \\ &\text{MaskedMultiAtt} \\ &+ \\ &\text{Scale} \\ &\text{Maximum} \end{aligned}$$

θ K

