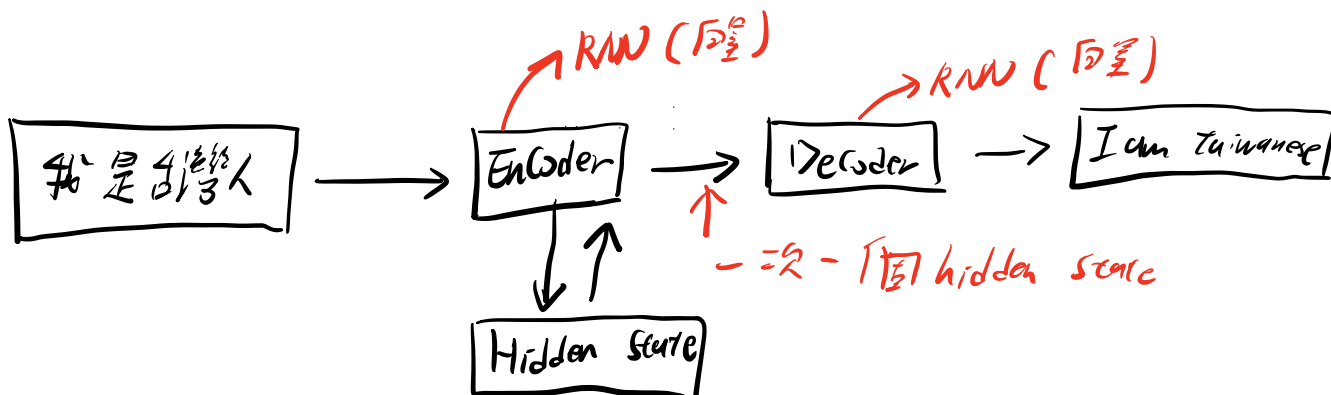
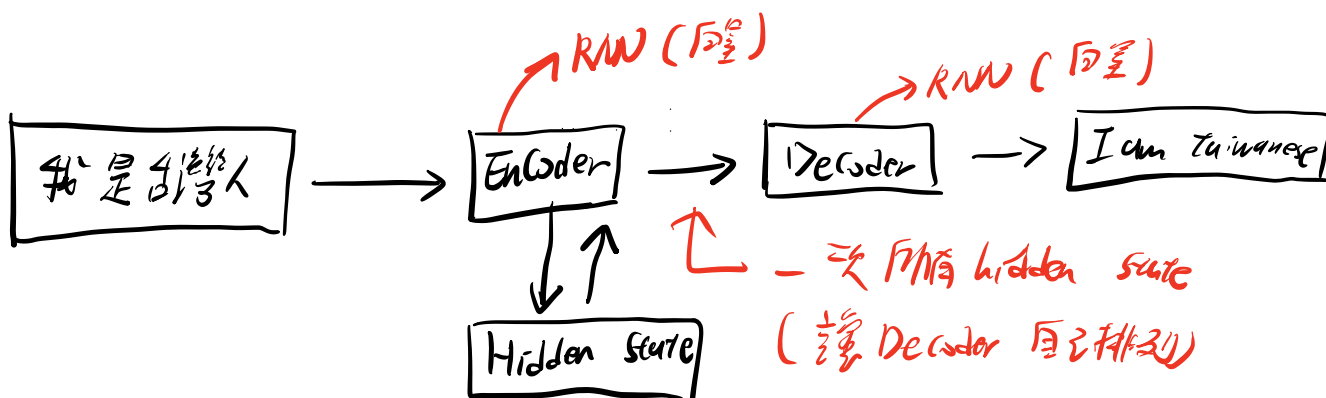


Encoder - Decoder Model (Seq2Seq Model)

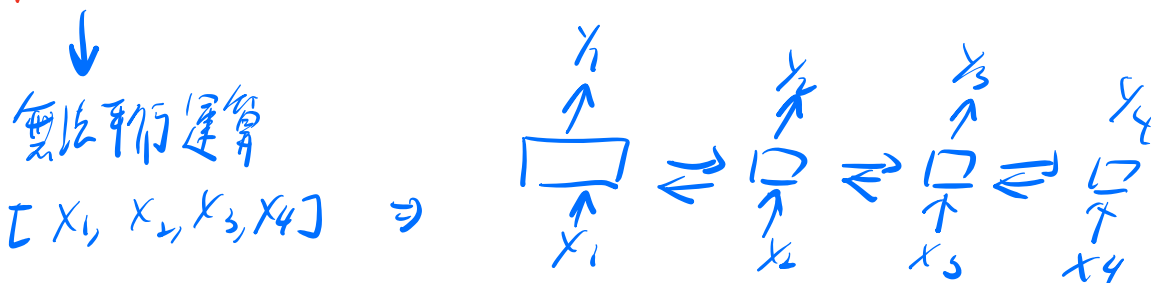


Encoder - Decoder Model + Attention Mechanism



Transformer (Seq2Seq Model + Self-Attention Mechanism)

RNN \rightarrow Self Attention Layer



Self Attention Layer:

利用矩阵运算. 一次执行完KVV!

Fig:

RNN:

$$h_t = W_1 h_{t-1} + W_2 \cdot x_t + b_t$$

$$y_t = A(h_t)$$

Self Attention Layer:

$$\begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_t \end{bmatrix}_{(Tx1)} = \begin{bmatrix} 0 & x_1 \\ h_1 & x_2 \\ \vdots & \vdots \\ h_t & x_t \end{bmatrix}_{(Tx2)} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}_{(2x1)} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_t \end{bmatrix}_{(Tx1)}$$