HCT NLP Week 3

问答摘要与推理
Seq2Seq（一）

Outline

- Encoder-Decoder结构
- Attention机制
- 模型Layer、Model构建
- Seq2Seq训练

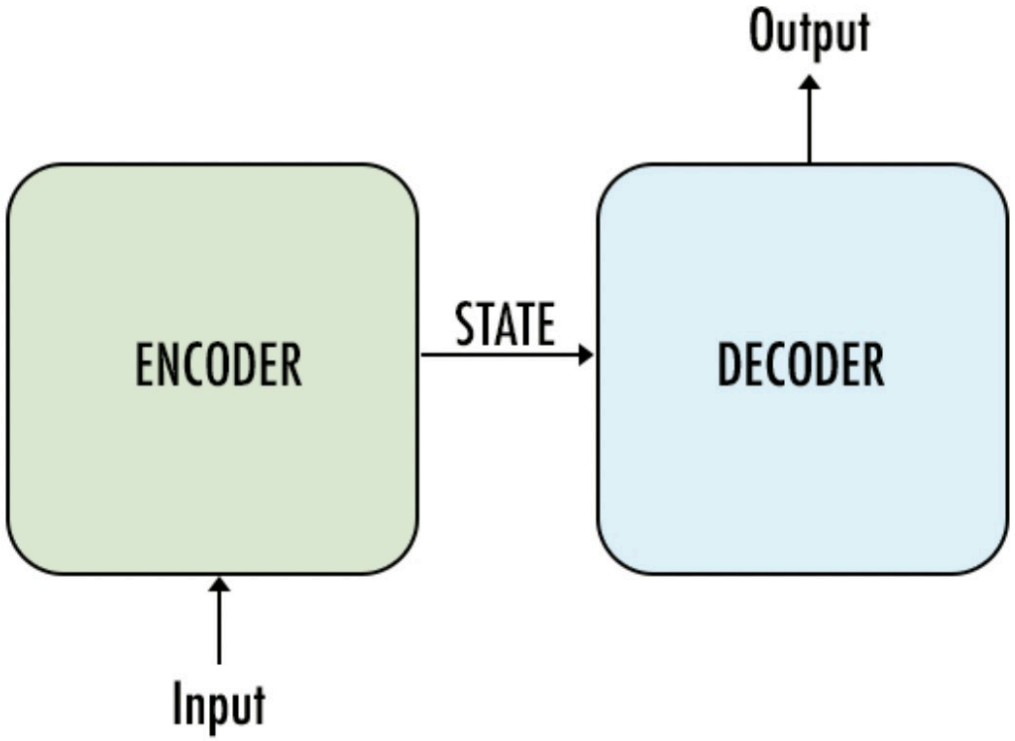
Outline

- Encoder-Decoder结构
- Attention机制
- 模型Layer、Model构建
- Seq2Seq训练

Seq2Seq

Encoder-Decoder结构

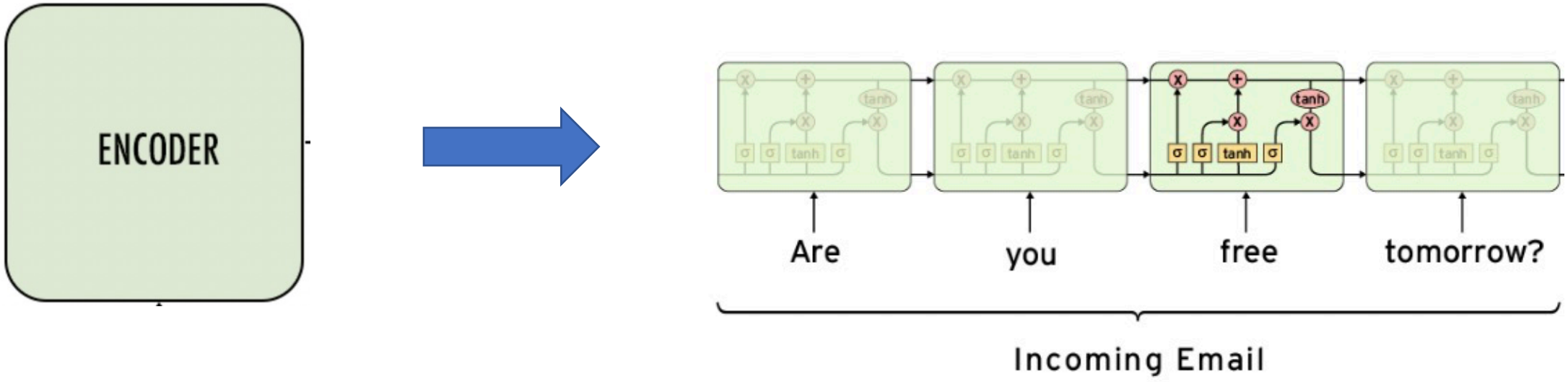
Sequence to sequence was first introduced by Google in 2014



1. Speech Recognition
2. Machine Language Translation
3. Name entity/Subject extraction
4. Relation Classification
5. Path Query Answering
6. Speech Generation
7. Chatbot
8. Text Summarization
9. Product Sales Forecasting

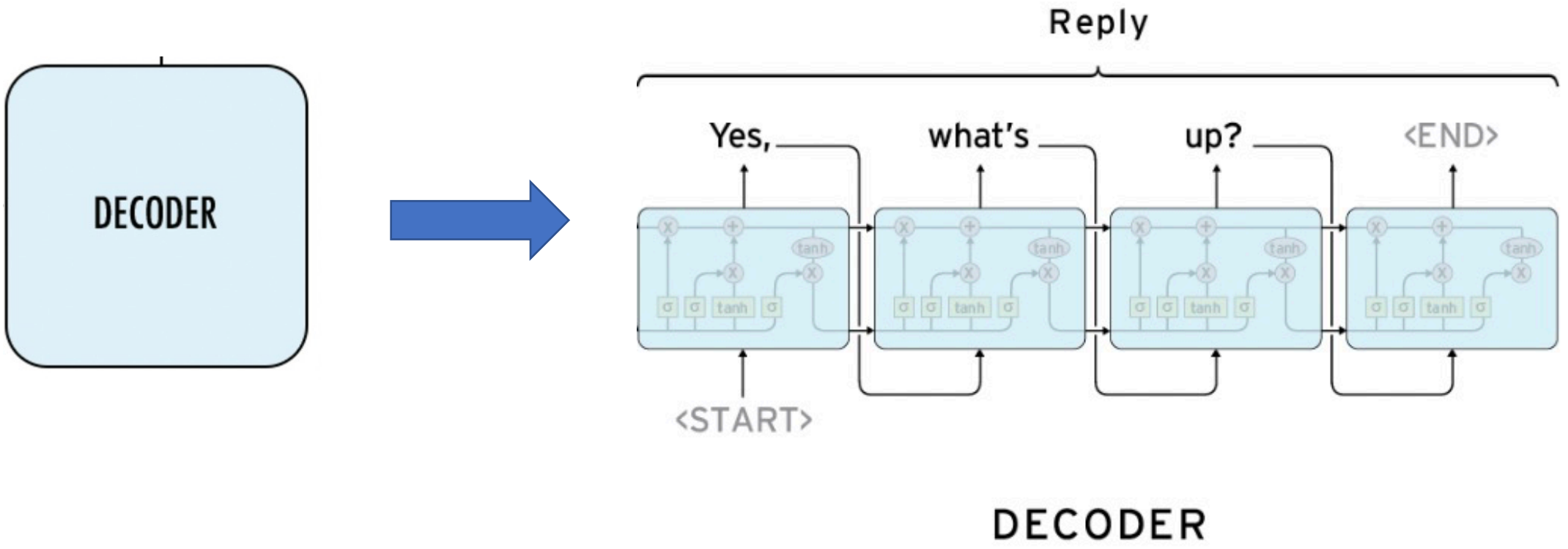
Seq2Seq

Encoder-Decoder结构



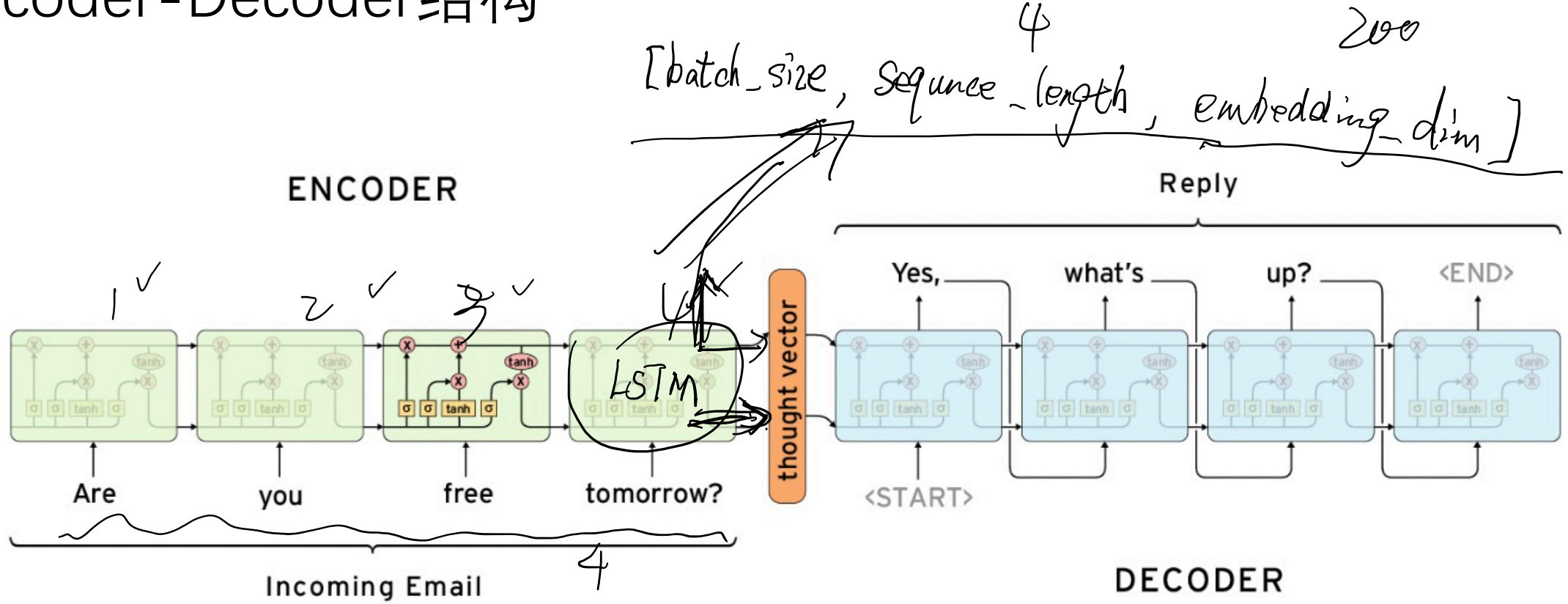
Seq2Seq

Encoder-Decoder结构



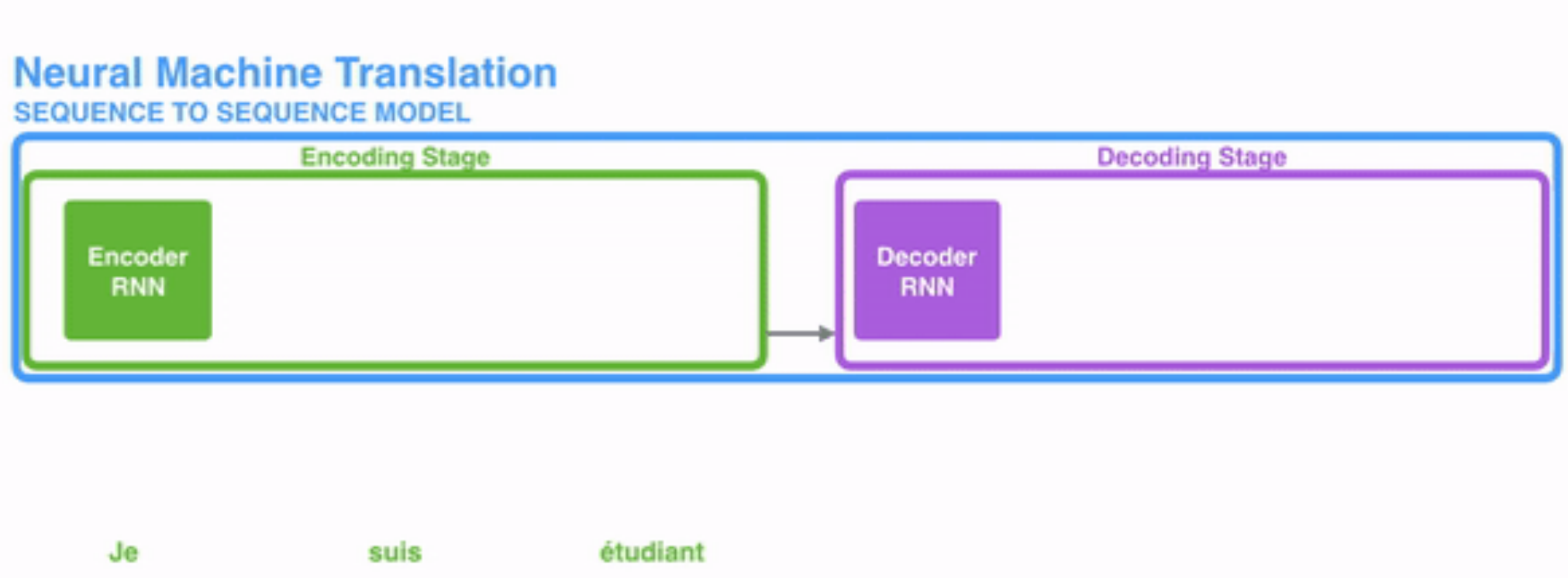
Seq2Seq

Encoder-Decoder结构



Seq2Seq

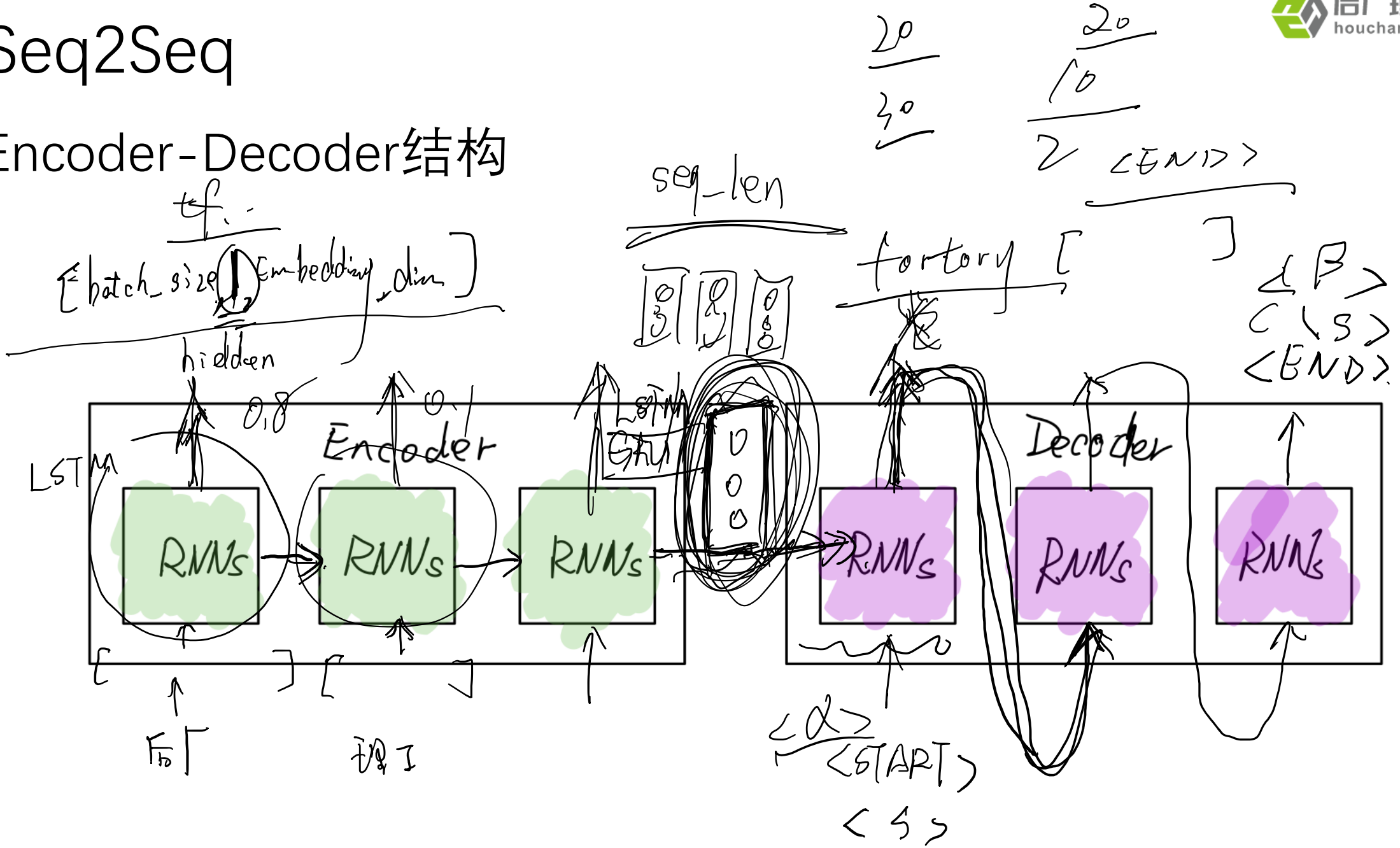
Encoder-Decoder结构



(source: Jay. Alammar, 2018)

Seq2Seq

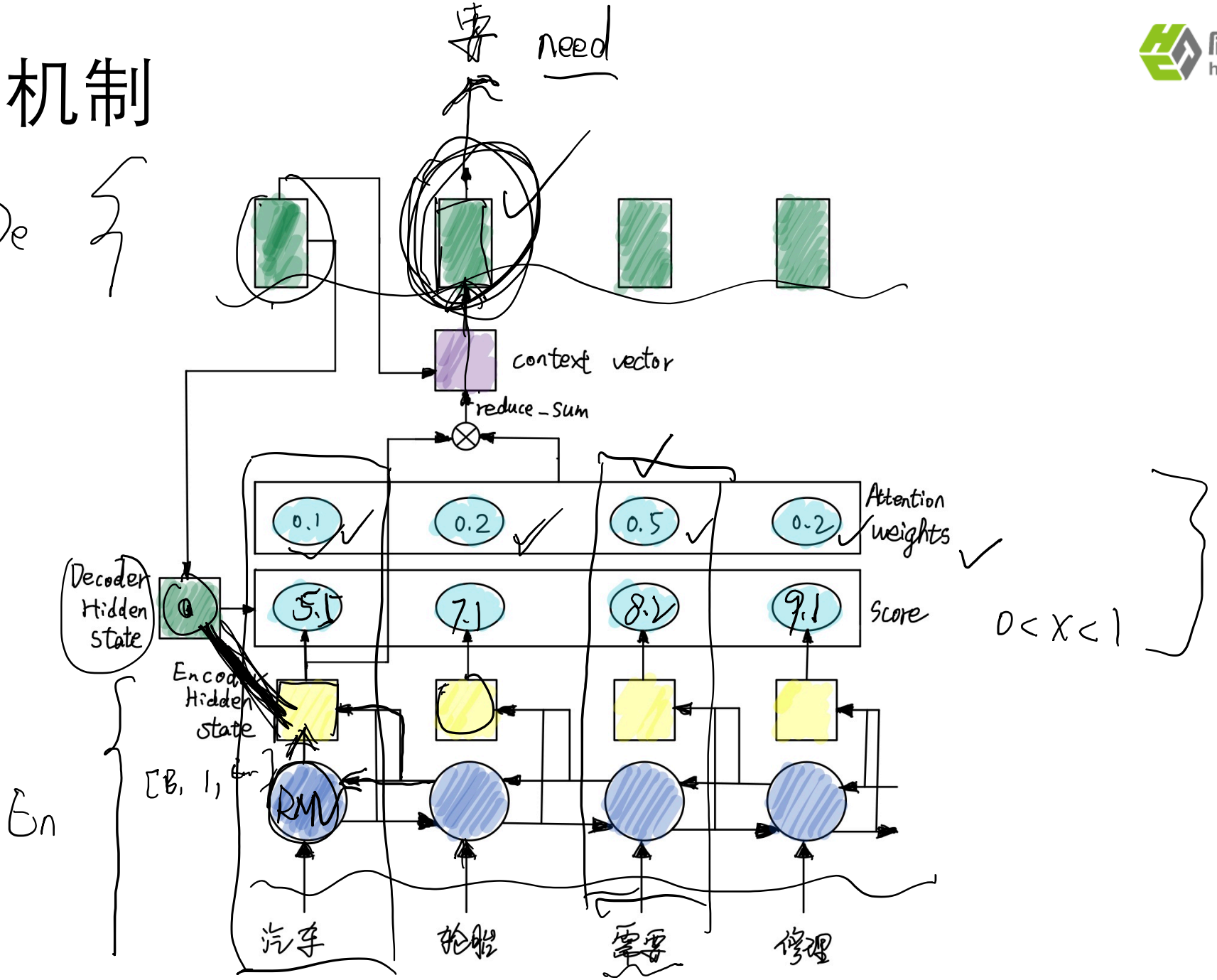
Encoder-Decoder结构



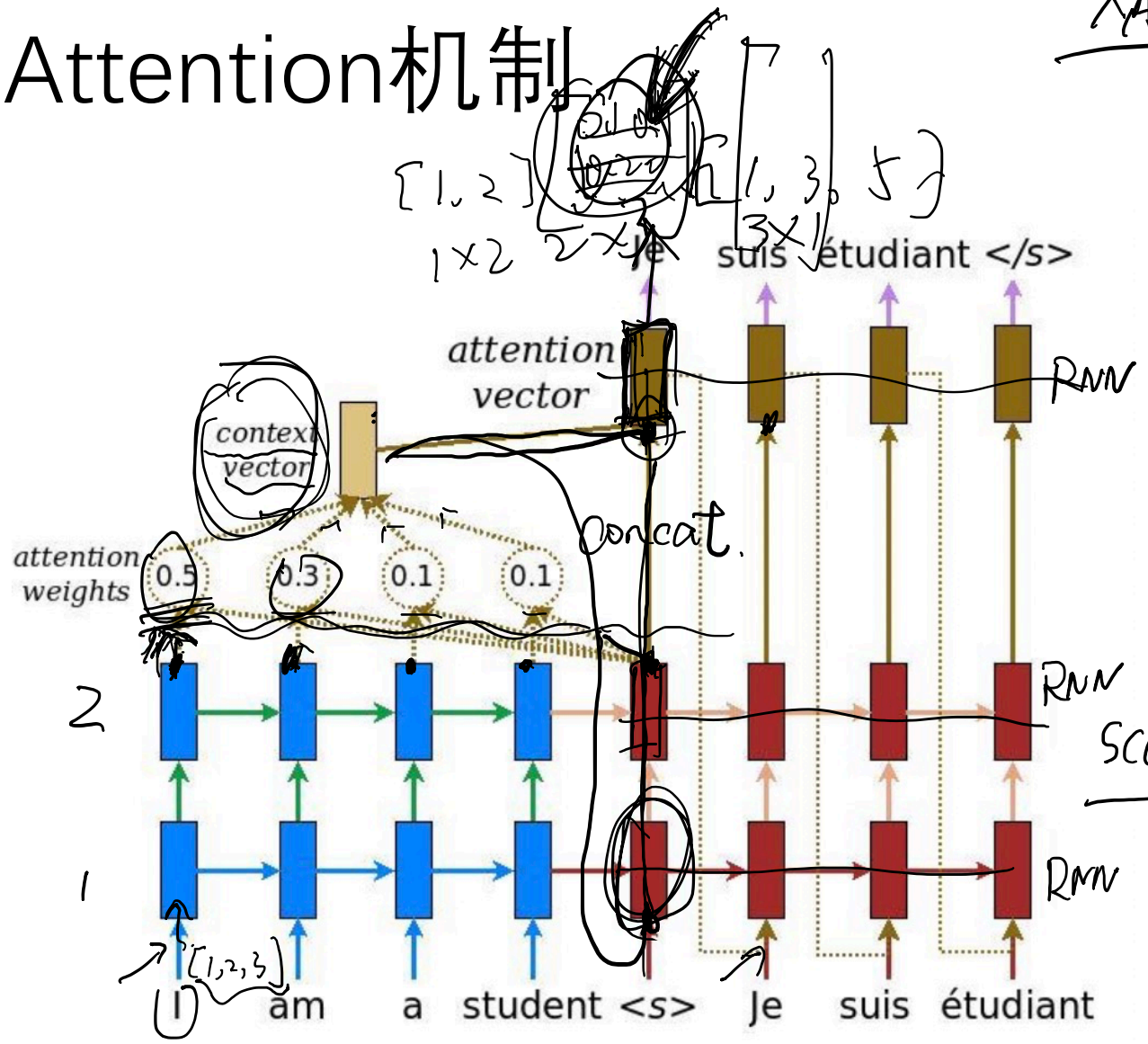
Outline

- Encoder-Decoder结构
- Attention机制
- 模型Layer、Model构建
- Seq2Seq训练

Attention机制



Attention机制



XAI

~~dedo~~

decoder hidden : h_t

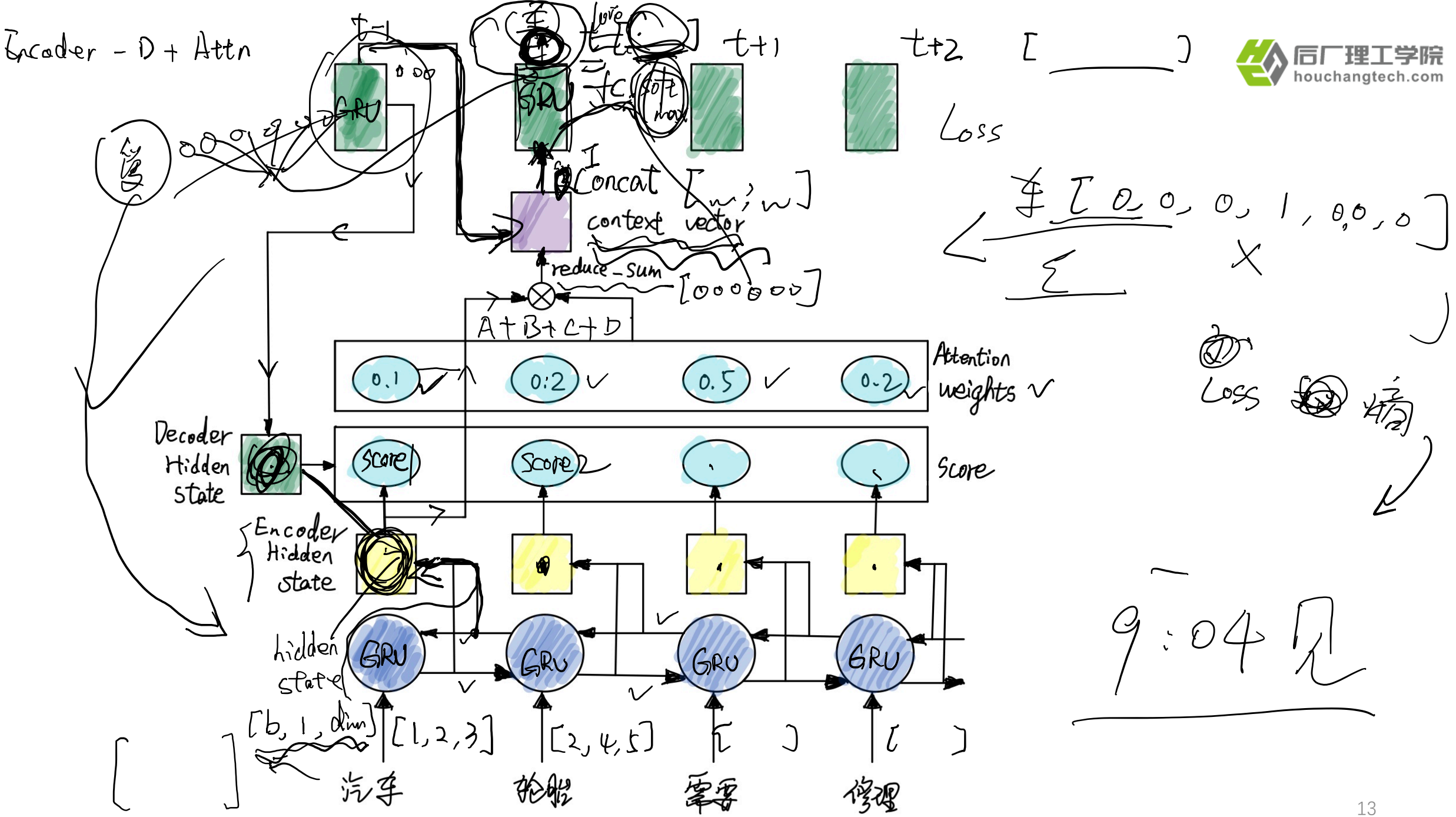
encoder hidden : $\bar{h}_s$

score

$$\alpha_{ts} = \frac{\exp(\text{score})}{\sum_{s'=1}^{S} \exp(\text{score})}$$

$$\text{score}(h_t, \bar{h}_s) = \begin{cases} h_t^T W \bar{h}_s & \text{[Luong's multiplicative style]} \\ v_a^T \tanh(W_1 h_t + W_2 \bar{h}_s) & \text{[Bahdanau's additive style]} \end{cases}$$

$$\text{Context} = \sum \alpha_{ts} \bar{h}_s = 0.5 \bar{h}_{s_1} + 0.3 \bar{h}_{s_2} + \dots$$


Encoder - D + Attn
t-1
t+1
t+2
GRU
Loss
Concat
context vector
reduce_sum
A+B+C+D
0.1
0.2
0.5
0.2
Attention weights
Decoder Hidden state
score
Encoder Hidden state
hidden state
[b, 1, dim]
[1,2,3]
[2,4,5]
汽车
轮船
需要
修理
Loss
9:04 见

Outline

- Encoder-Decoder结构
- Attention机制
- 模型Layer、Model构建
- Seq2Seq训练

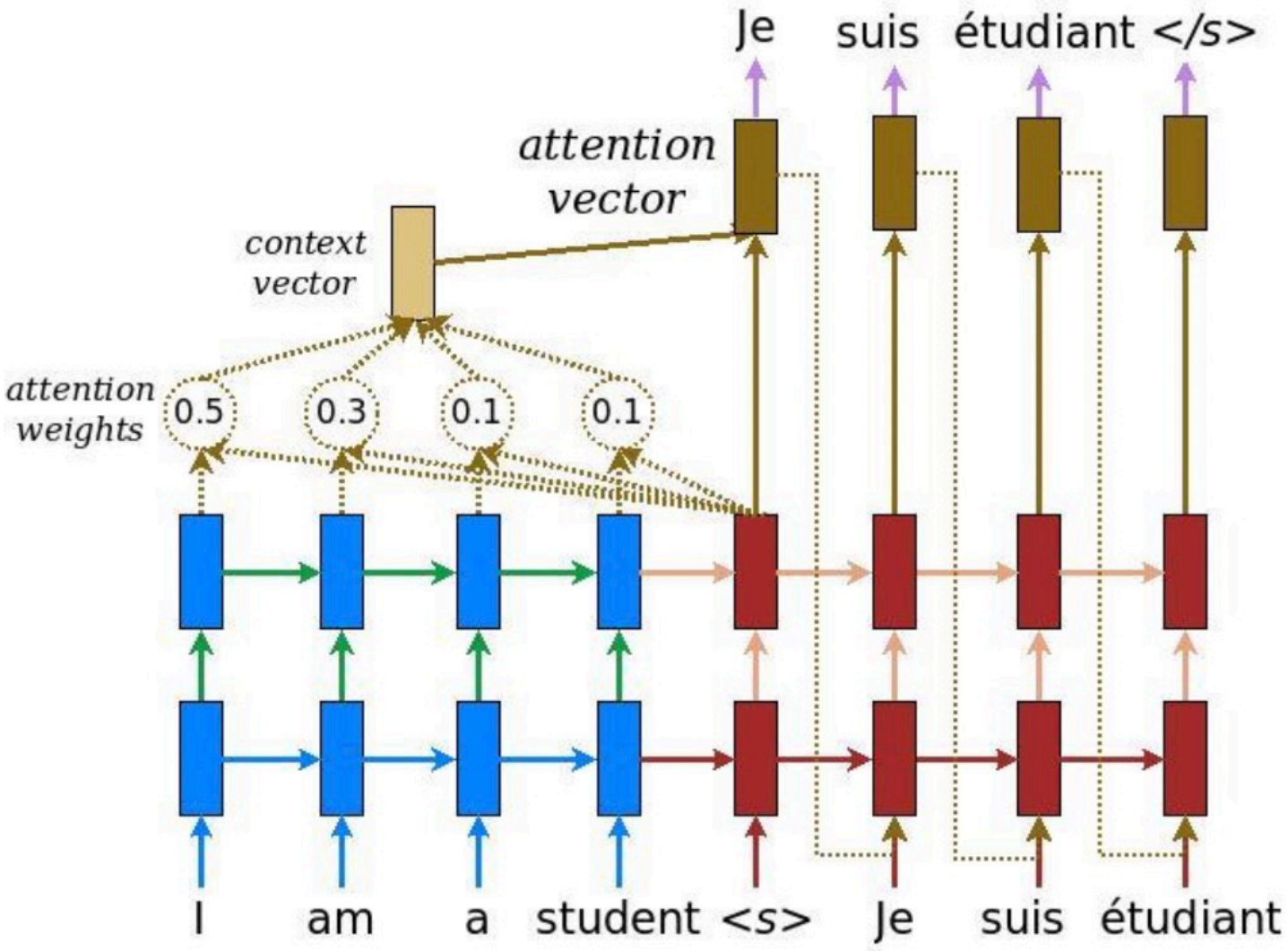
模型Layer、Model构建

代码见jupyter

Outline

- Encoder-Decoder结构
- Attention机制
- 模型Layer、Model构建
- Seq2Seq训练

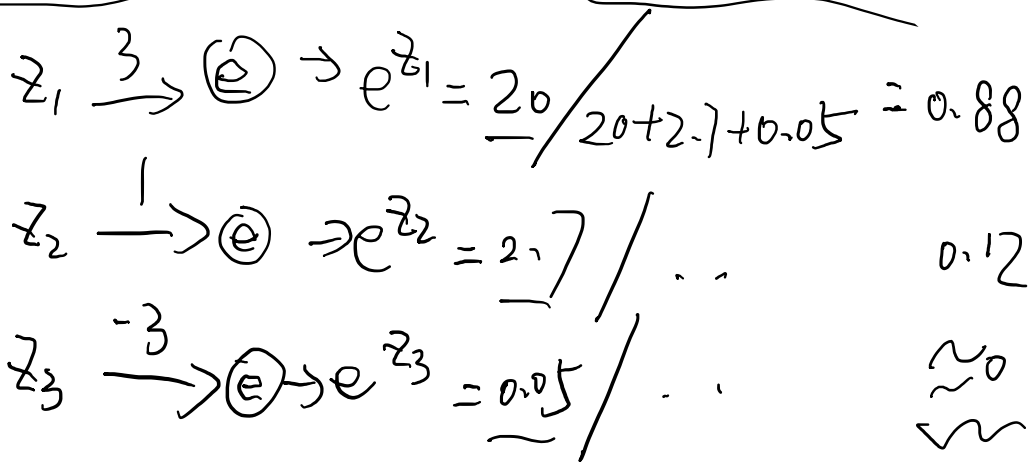
Seq2Seq训练

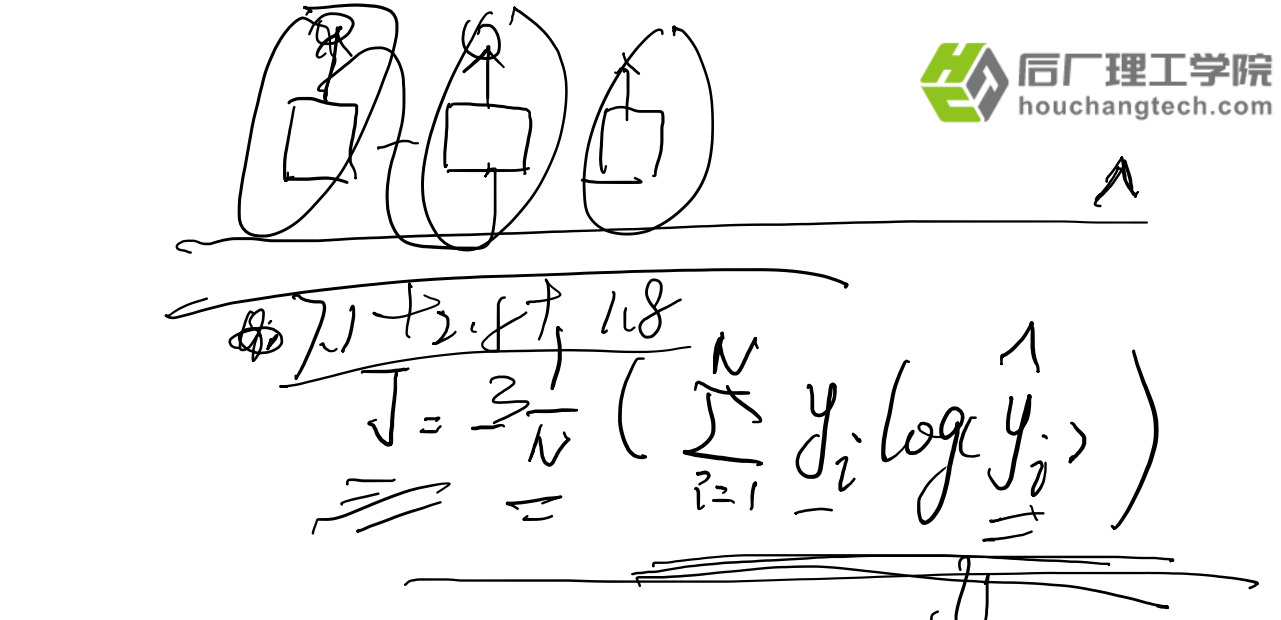


How is it trained?

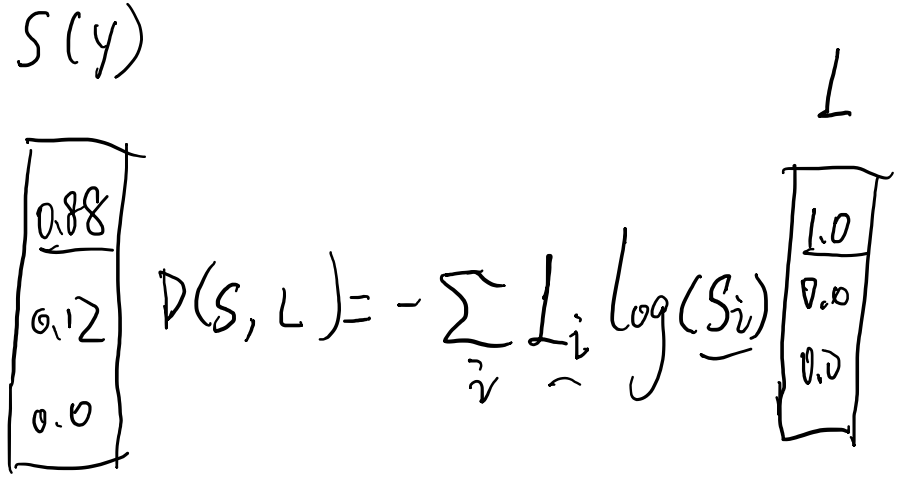
Seq2Seq训练

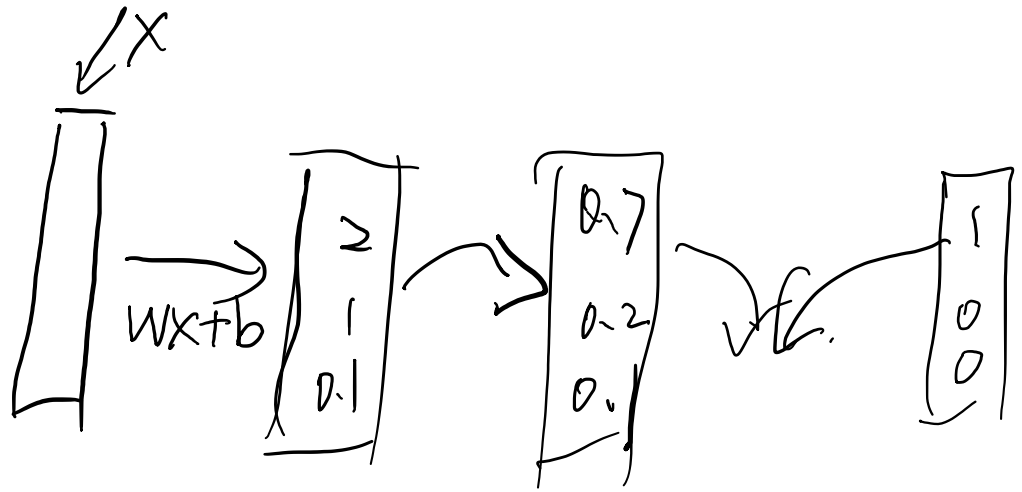
cross-entropy loss 交叉熵





$$J(w) = -\frac{1}{N}\sum_{n=1}^{N}\left[y_n \log \hat{y}_n + (1-y_n)\log(1-\hat{y}_n) \right]$$



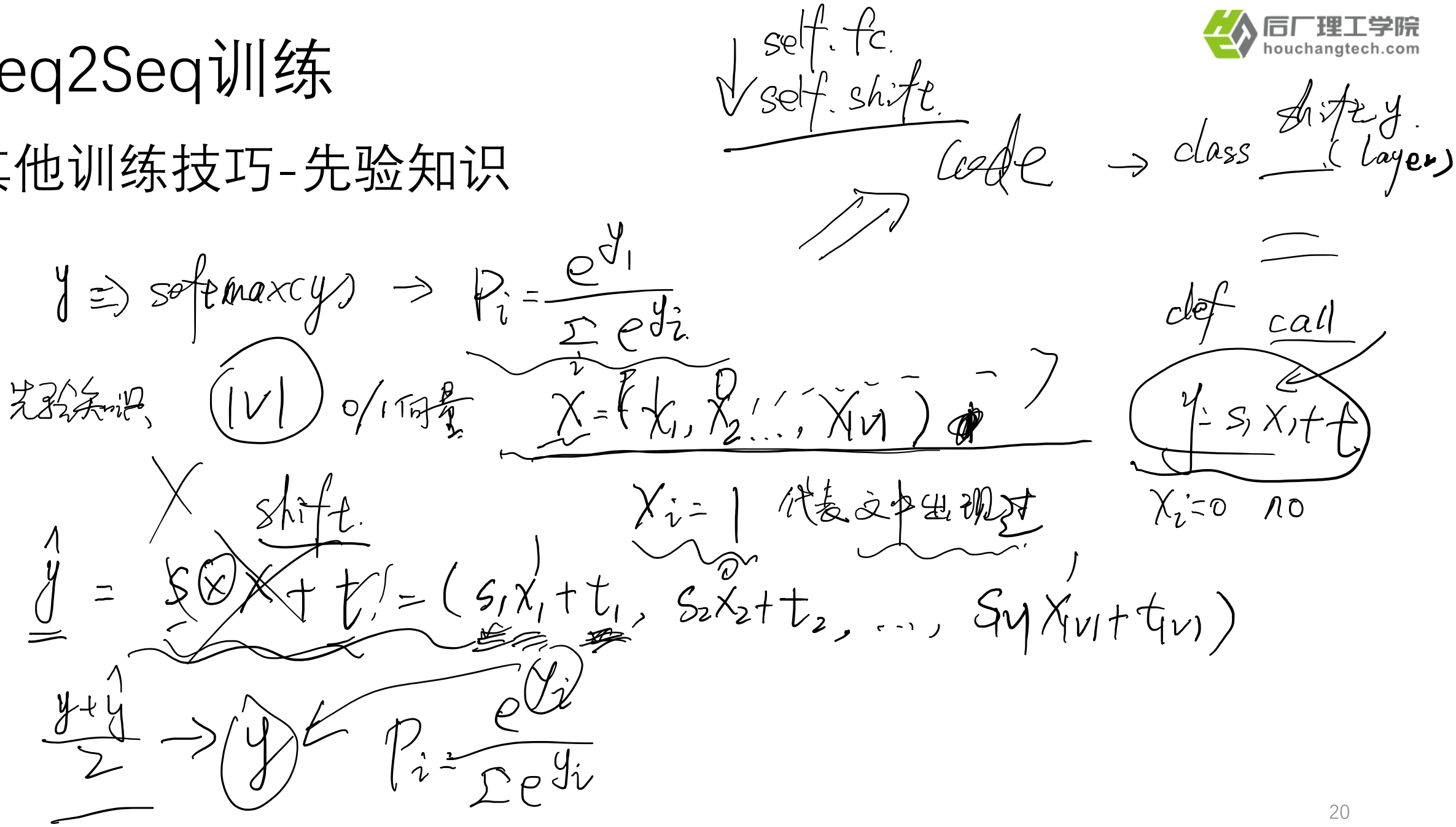


Seq2Seq训练

cross-entropy loss 交叉熵

Seq2Seq训练

其他训练技巧-先验知识

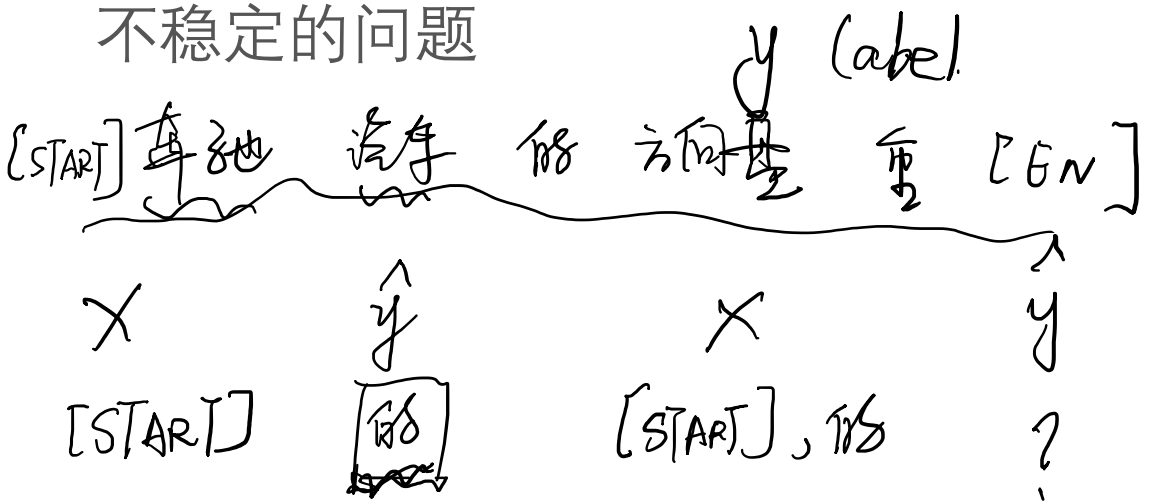


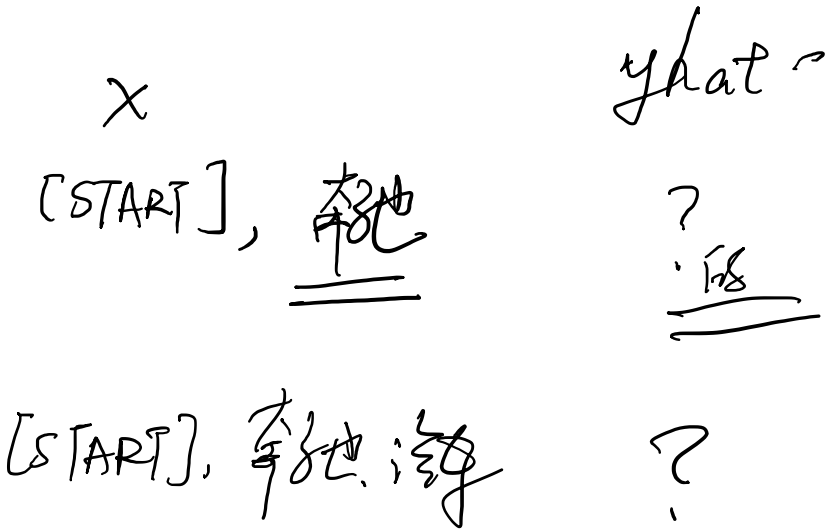
Seq2Seq训练

Teacher Forcing

RNN模型是用前一步的输出作为输入

Teacher Forcing方法能够解决收敛速度慢和不稳定的问题





Seq2Seq训练

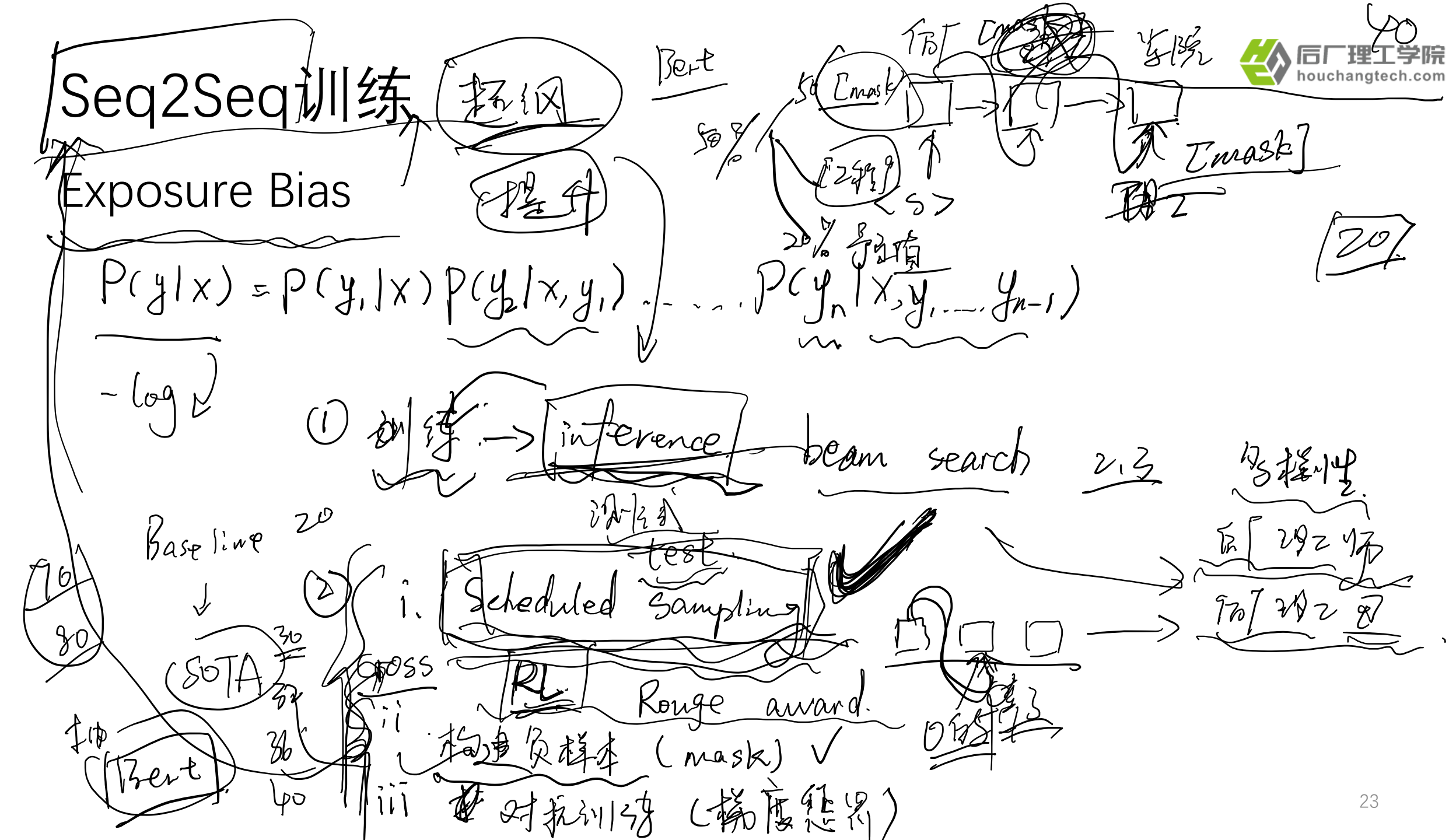
Teacher Forcing

Seq2Seq训练

Exposure Bias

$$P(y|x) = P(y_1|x)\,P(y_2|x,y_1)\cdots P(y_n|x,y_1,\ldots,y_{n-1})$$

$-\log$



作业

Bye ！