后厂理工学院
houchangtech.com

# HCT NLP Week 3

## 问答摘要与推理
## Seq2Seq（一）

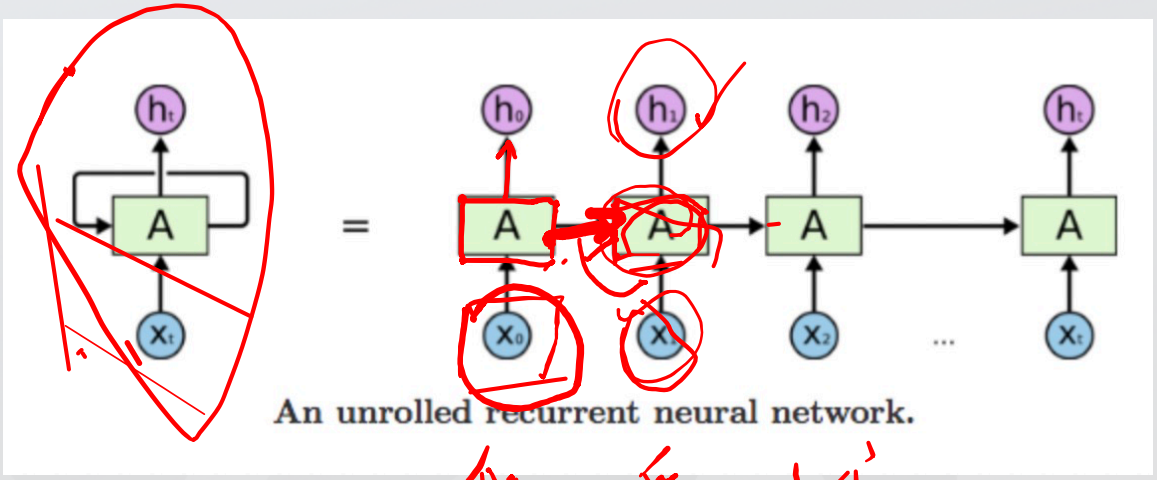
# Outline

1. RNN、LSTM、GRU
2. Encoder-Decoder结构
3. Attention机制
4. 作业

# 1. RNN、LSTM、GRU

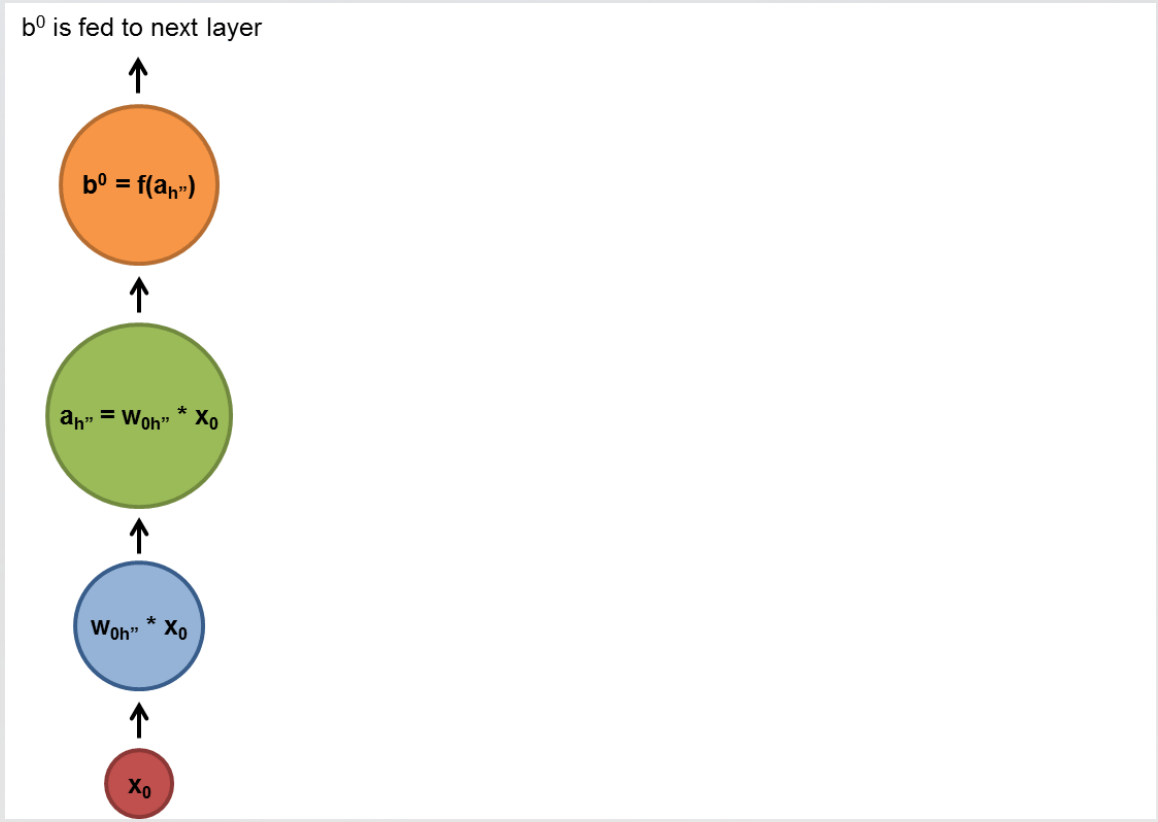
# RNN

## Recurrent Neural Network



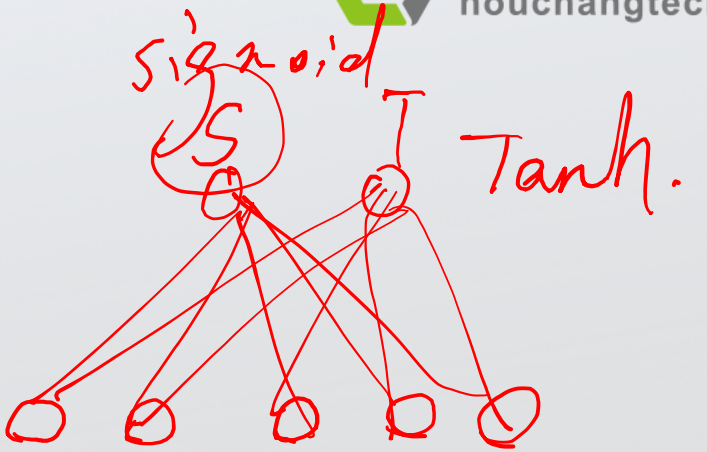

An unrolled recurrent neural network.

$$h_t = f(h_{t-1}, x_t)$$



# RNN

## Recurrent Neural Network




multiplication

addition

0 0 0 0
[1, 2] [4, 6]
0 0
[4, 3]

0 0
[5, 5]

concatenate

concat

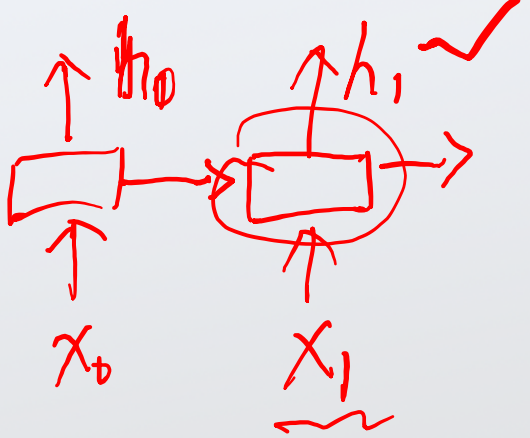
0 0 → 1~ 1~
[1, 2] [0, 1]

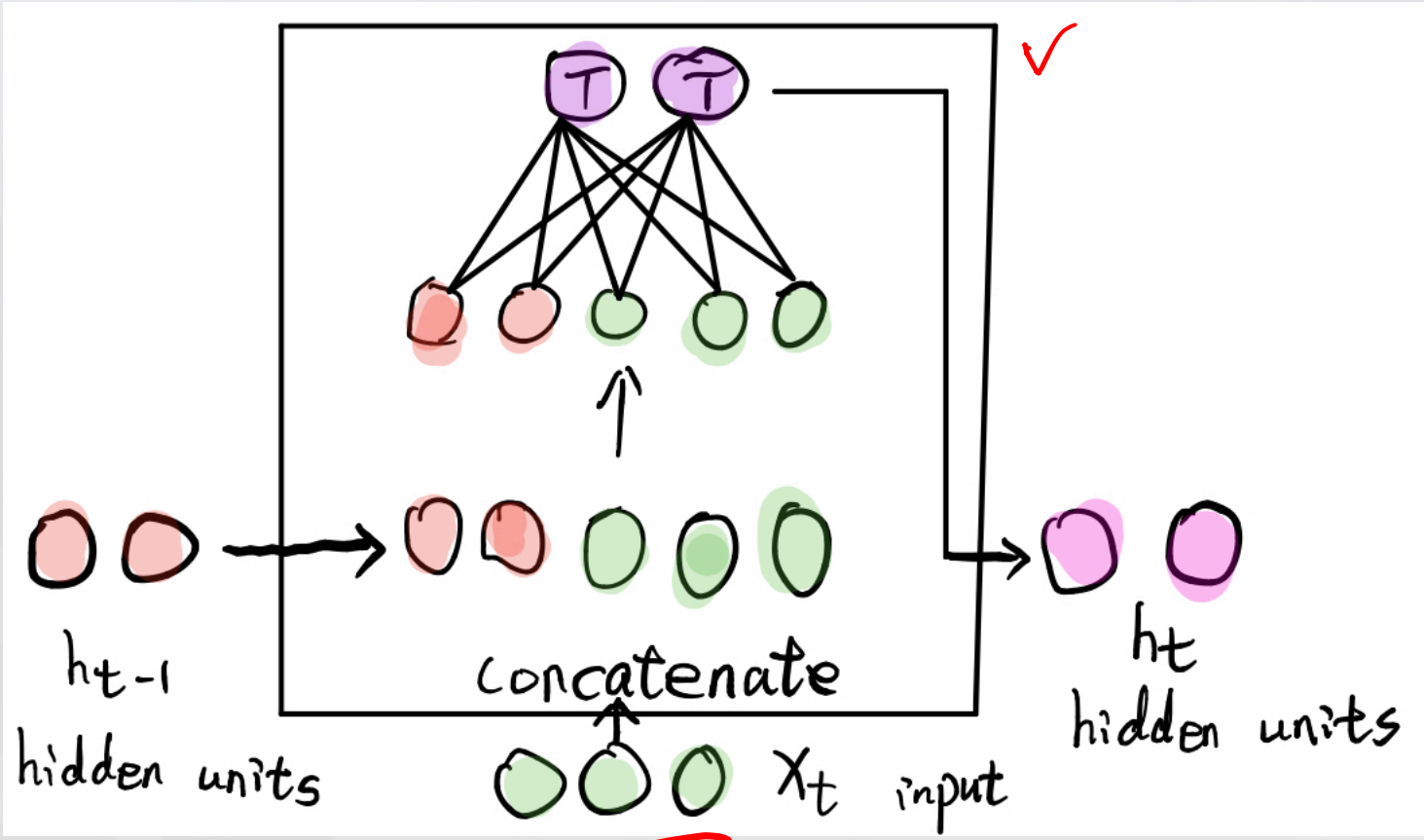
0 0
[1, 2]

0 0 0 0
[1, 2, 4, 3]
0 0
[4, 3]

# RNN

## Recurrent Neural Network

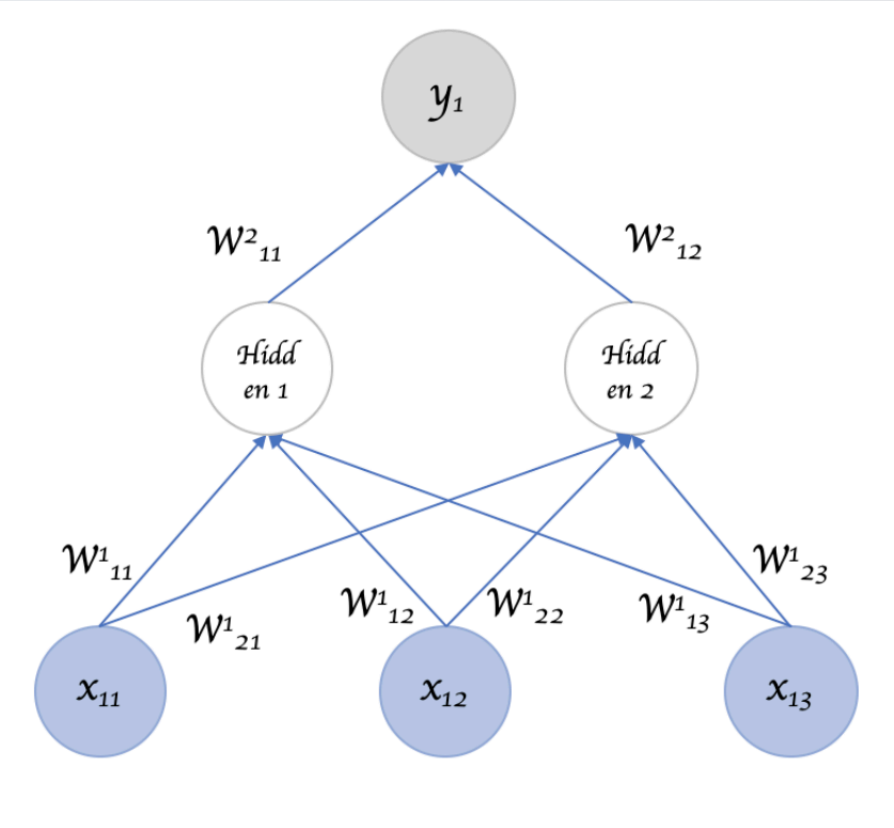


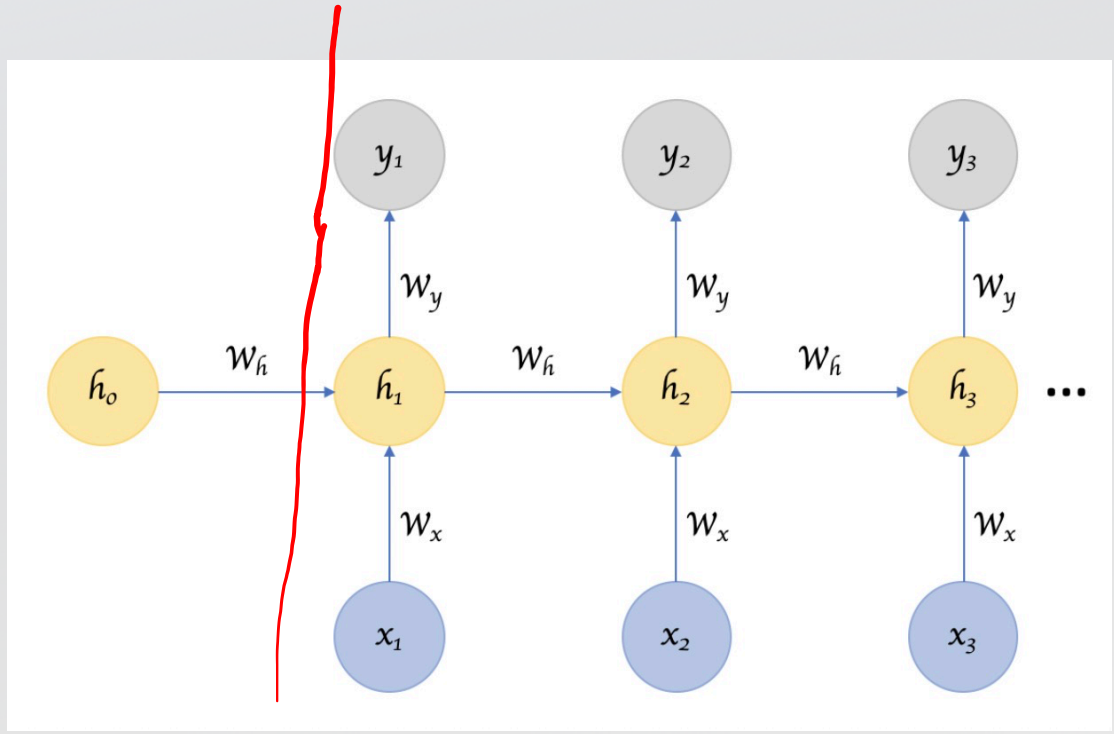




后厂理工学院
houchangtech.com

# RNN

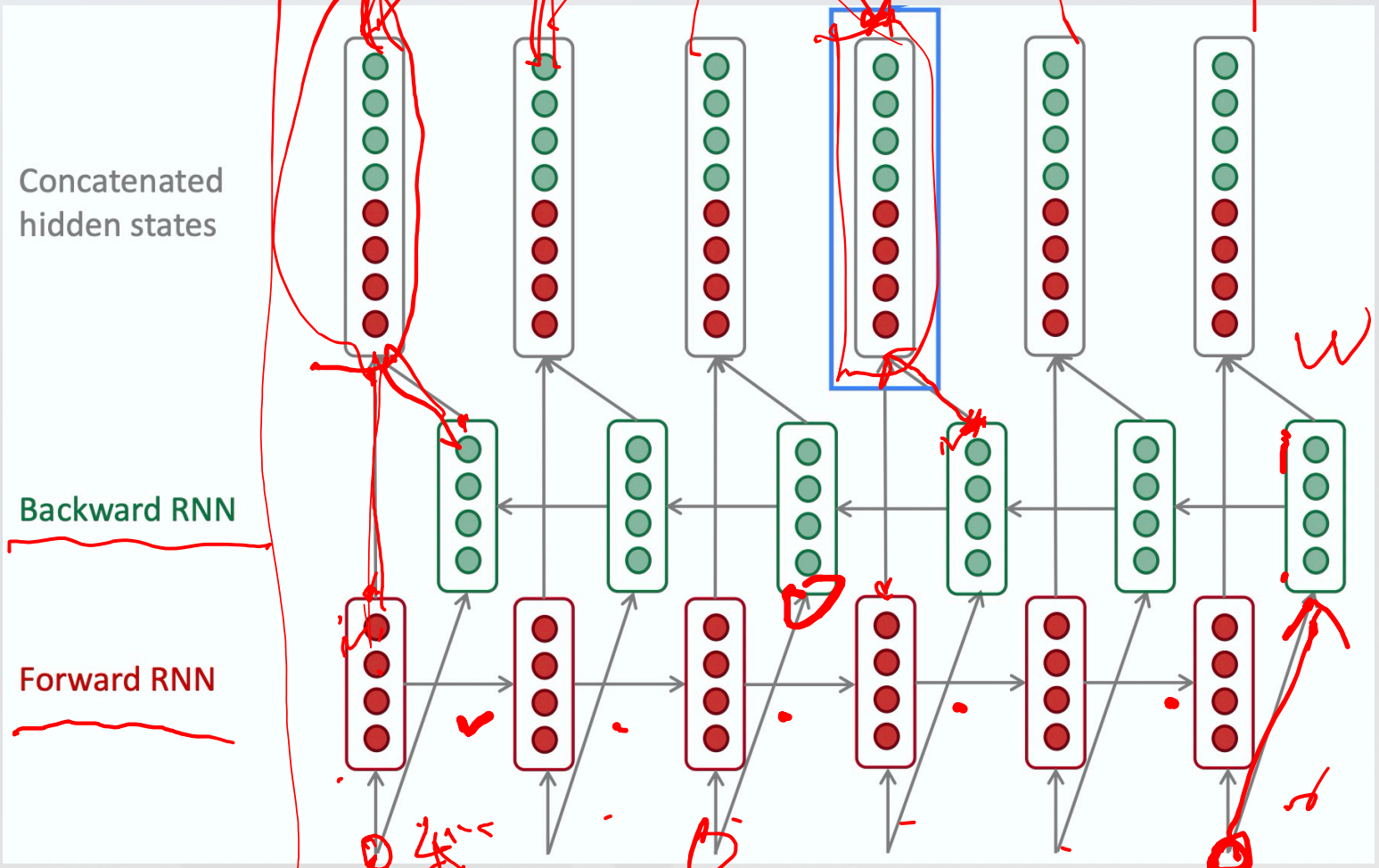
## parameter sharing



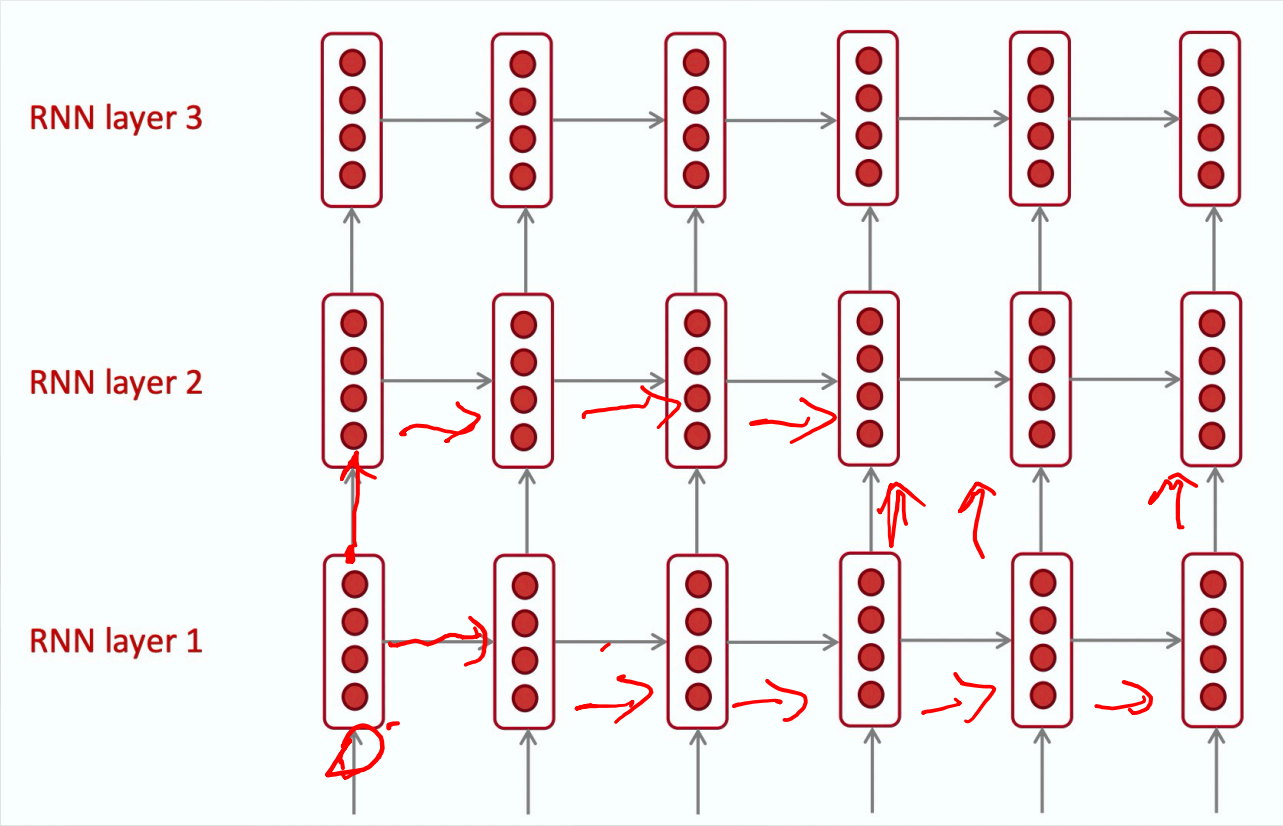
# RNN

## Bidirectional RNNs



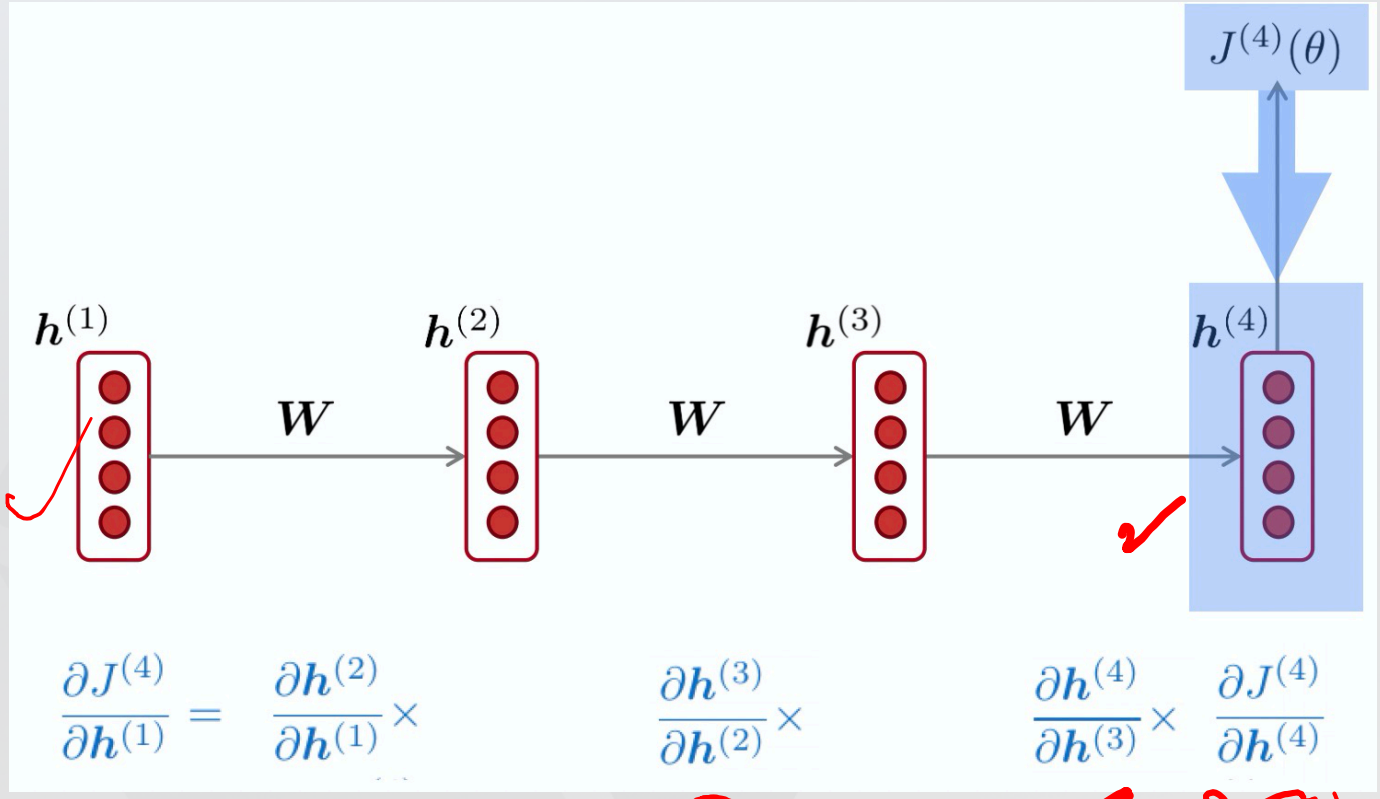
# RNN

## Deep RNNs



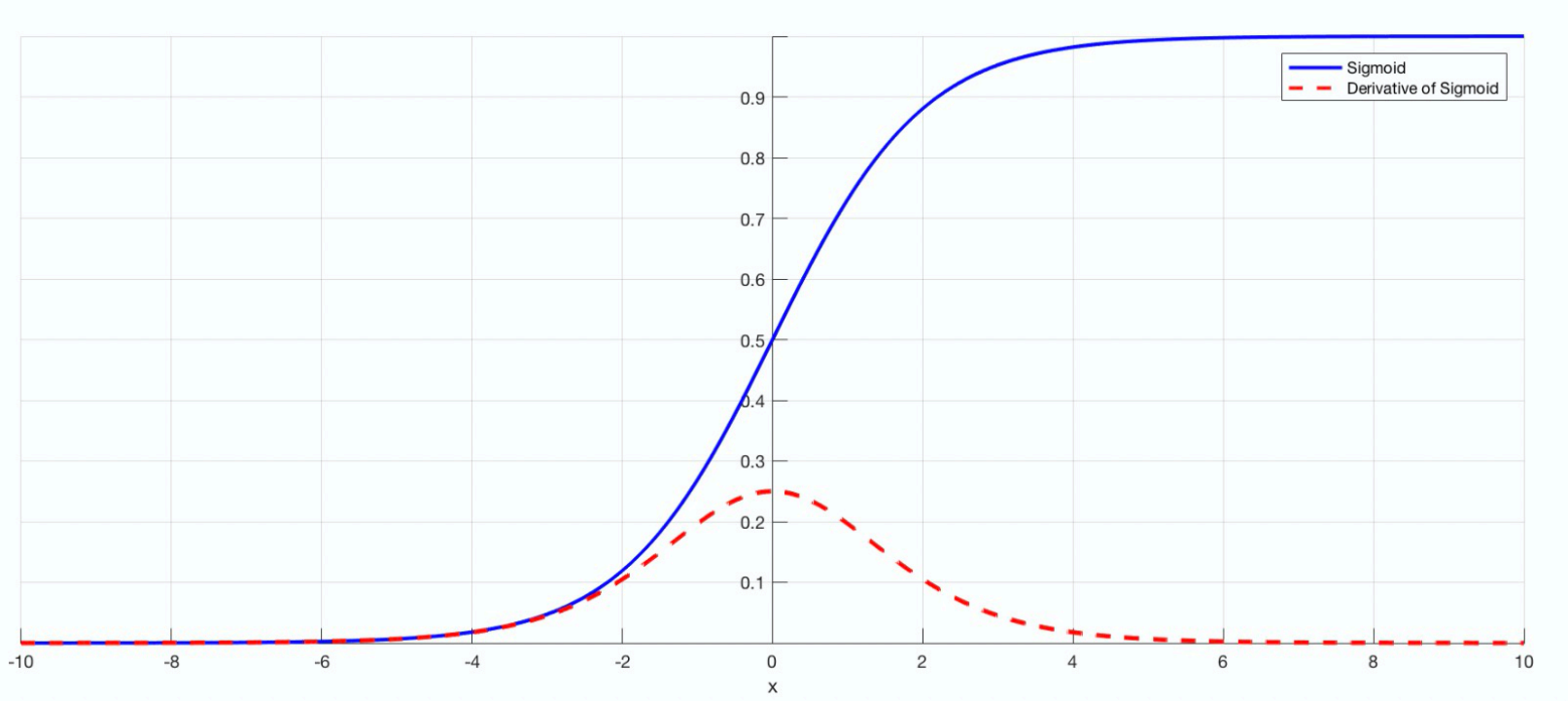
# Exploding and Vanishing Gradient Problem

后厂理工学院
houchangtech.com



# Exploding and Vanishing Gradient Problem

后厂理工学院
houchangtech.com




2:15

后厂理工学院
houchangtech.com

Exploding Gradient

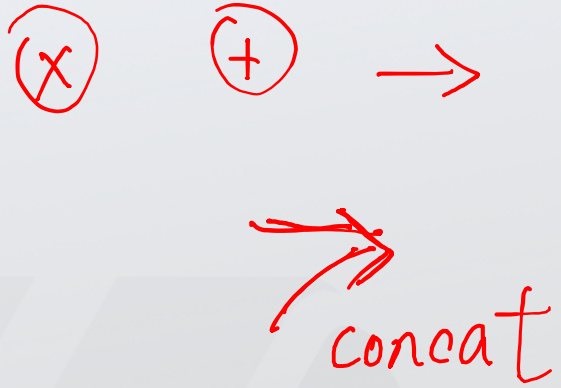
- gradient clipping

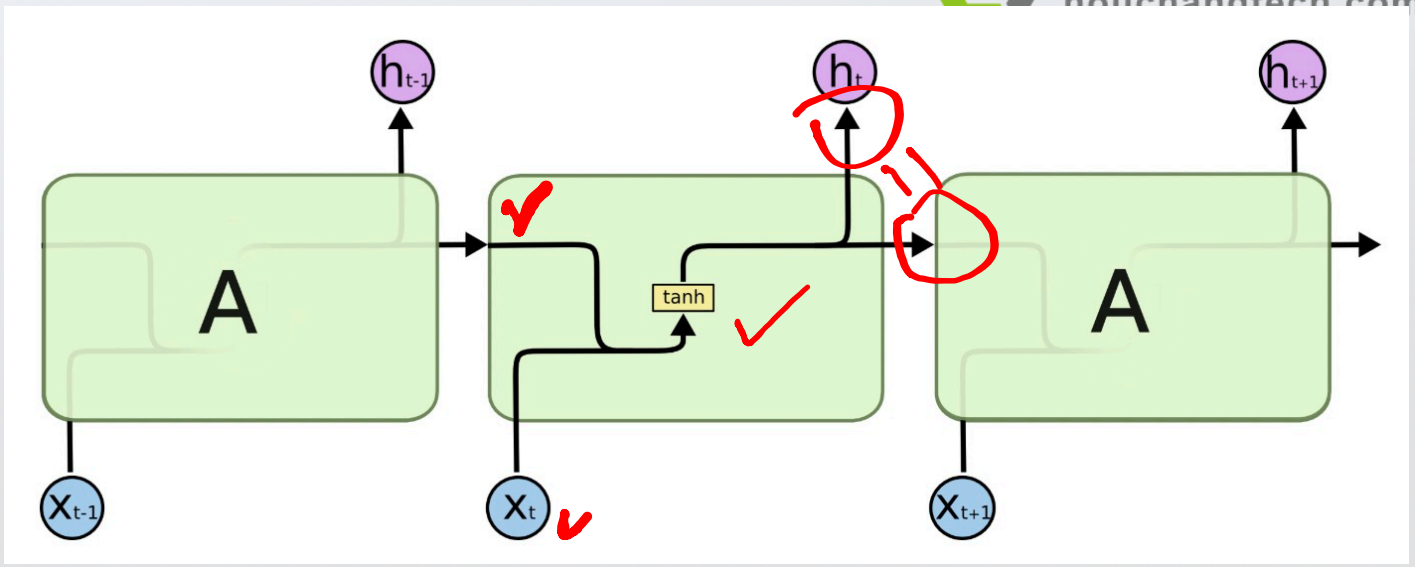
Vanishing Gradient

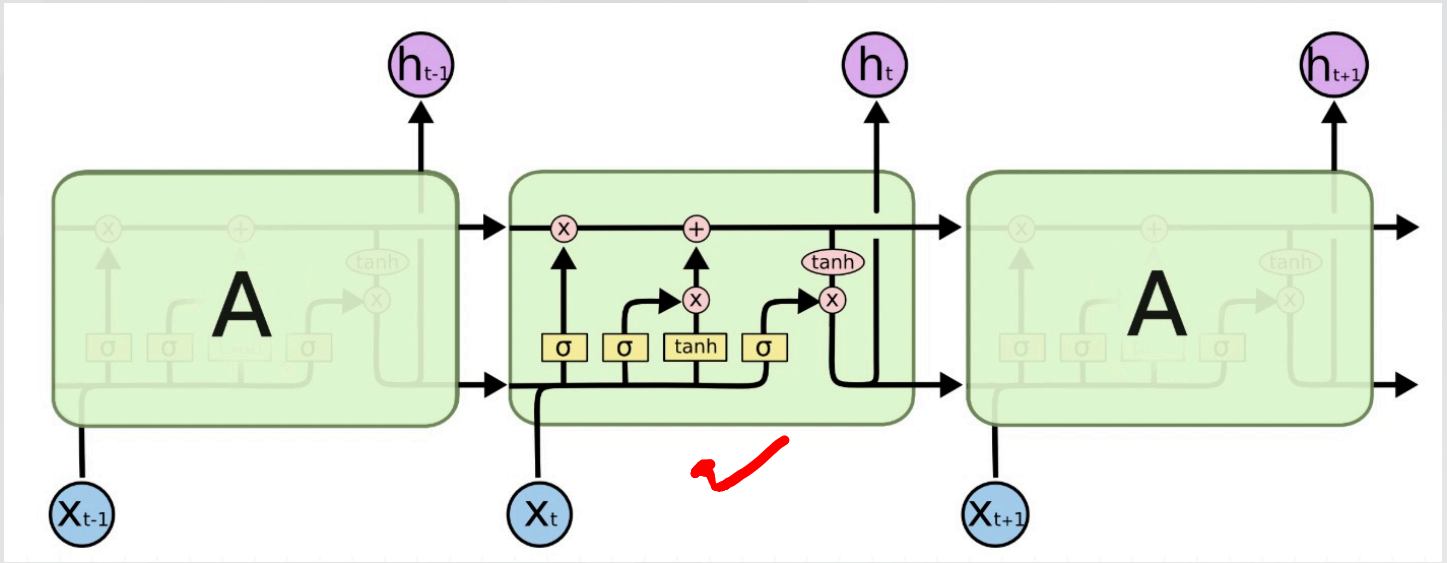
- Identity Initialization
- LSTM
- Residual Networks
- Batch Normalization

# LSTM

## Long Short Term Memory



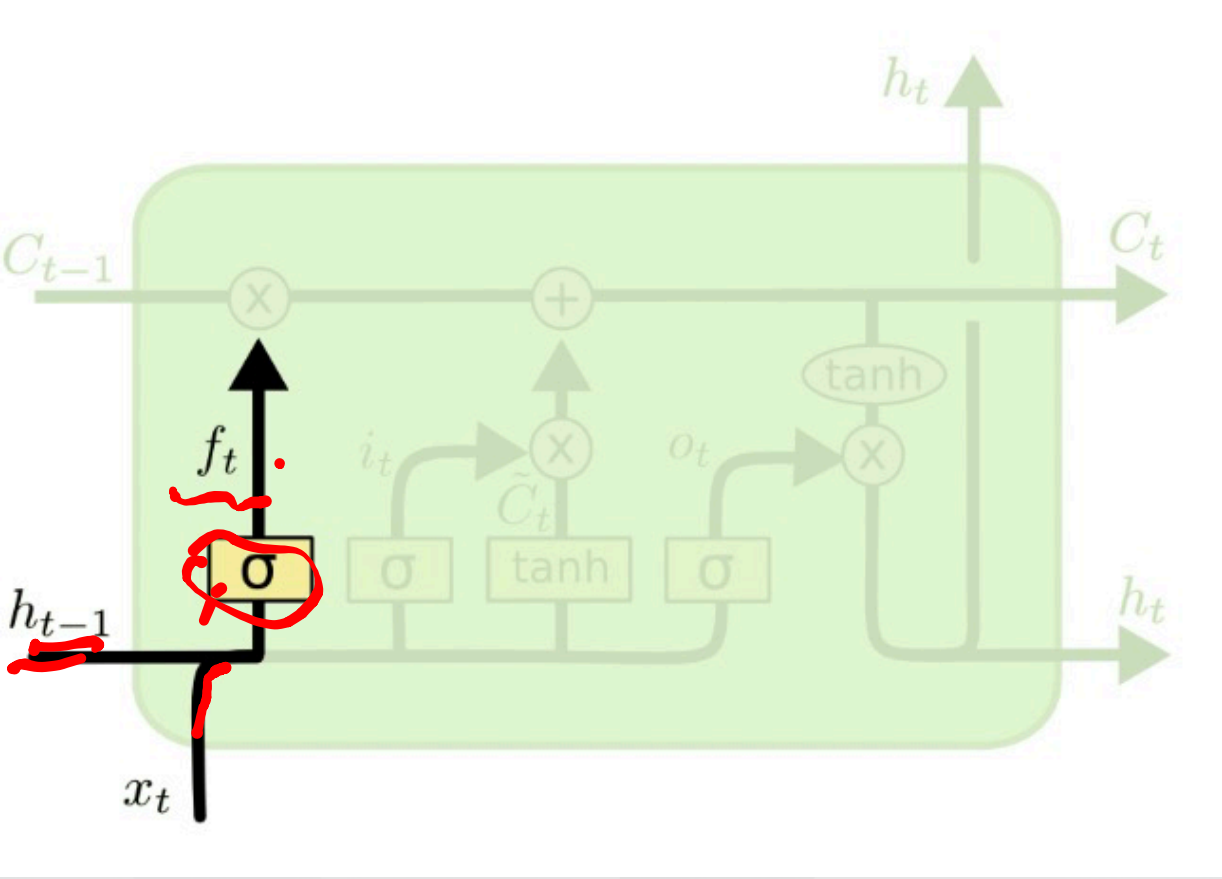
# LSTM

3个门 gates

input

forget

output




~~input gate~~

forget gate

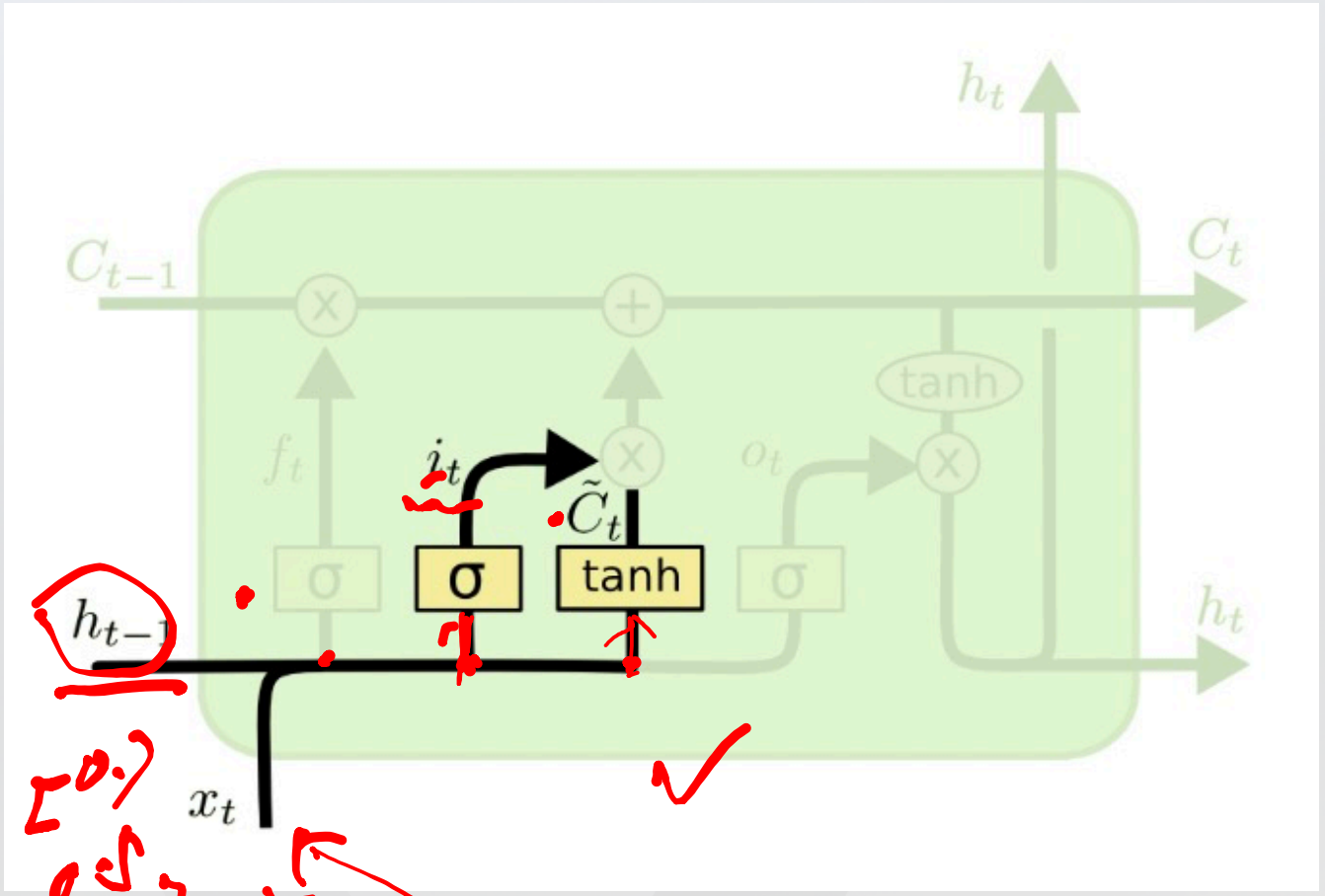
$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right)$$

$W_f$

# LSTM

input gate

后厂理工学院
houchangtech.com




$$i_t = \sigma(W_i \cdot [h_{t-1}; x_t] + b_i)$$

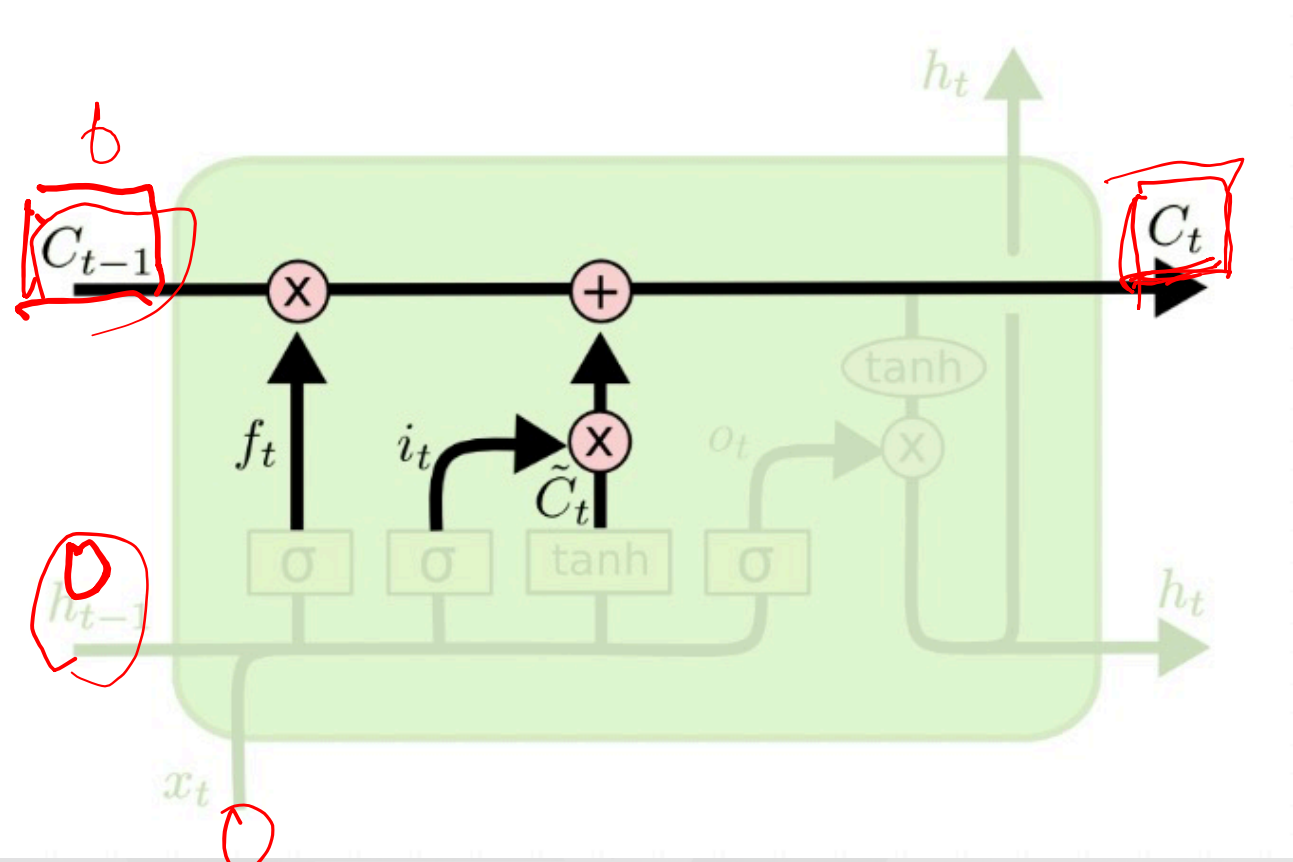
$$\tilde{C}_t = \tanh(W_C [h_{t-1}; x_t] + b_C)$$

$W$

$[0.1, 0.2, 0.3]$

tensor

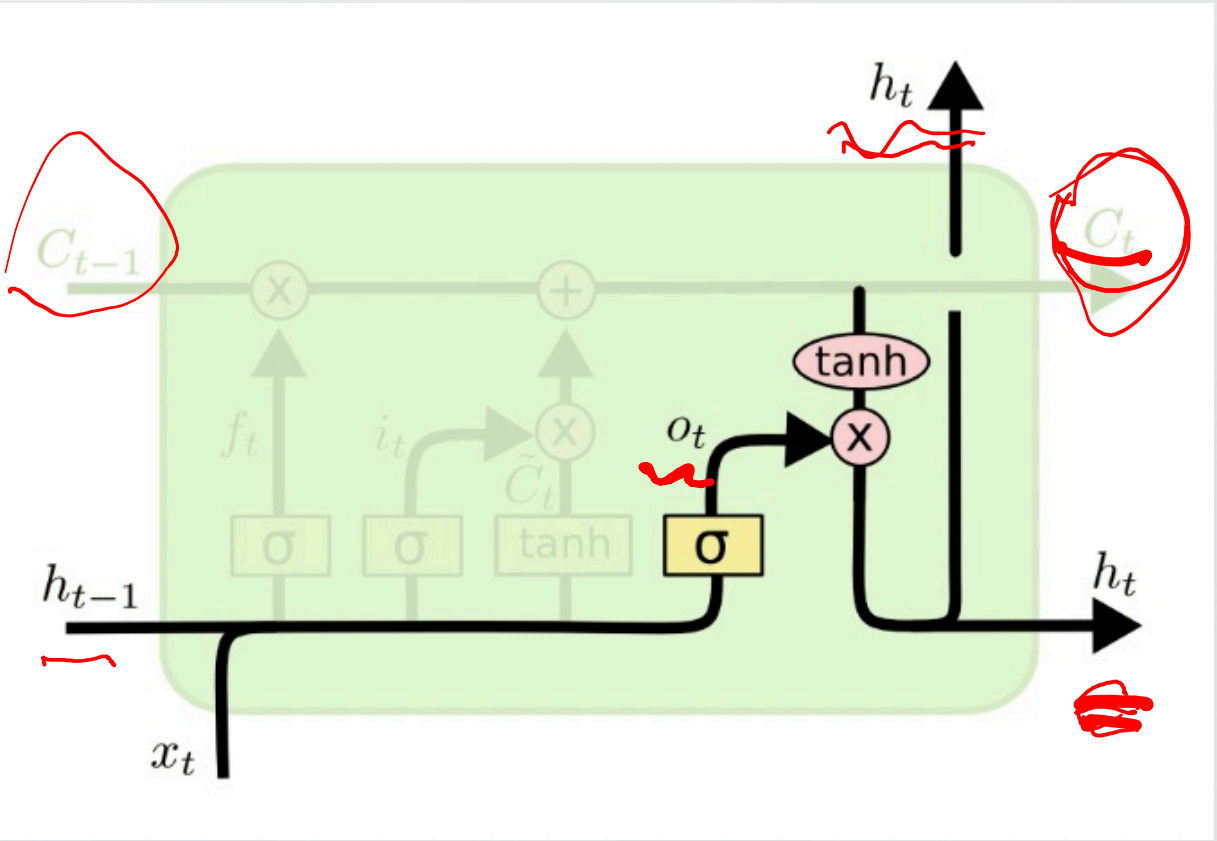
# LSTM




$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

# LSTM

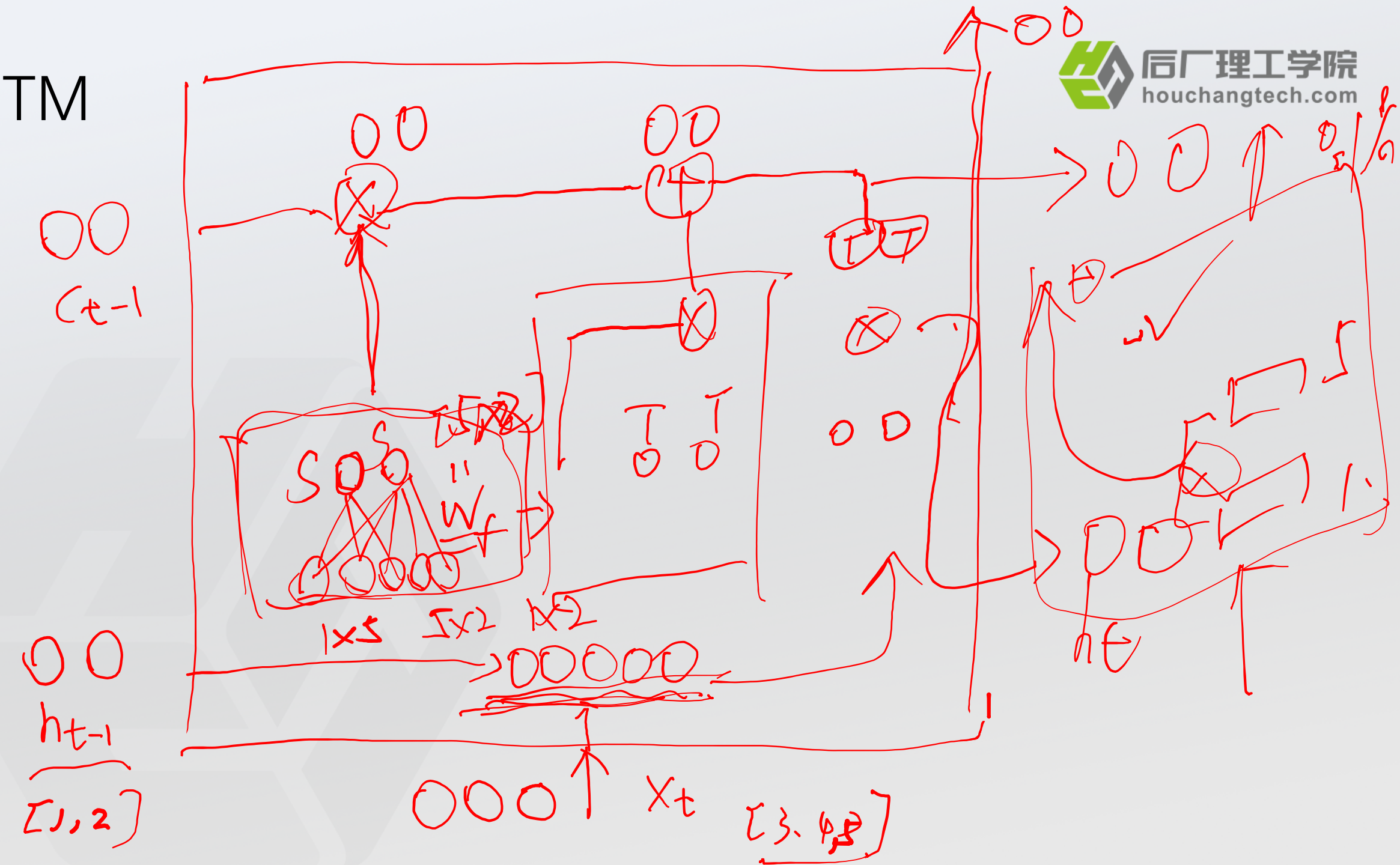
Output gate




$$o_t = \sigma(W_o[h_{t-1}; x_t] + b_o)$$

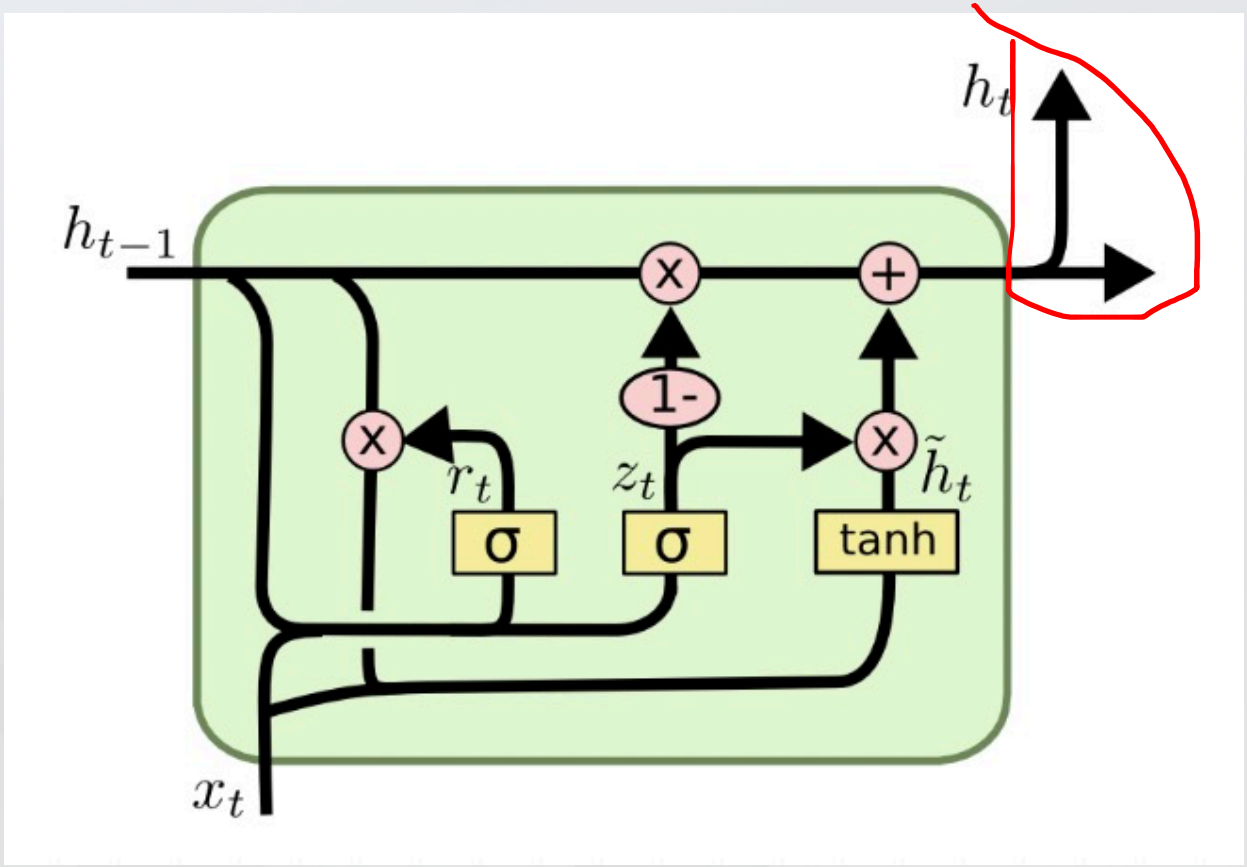
$$h_t = o_t * \tanh(C_t)$$

# LSTM



# GRU

后厂理工学院
houchangtech.com




$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh(W[r_t * h_{t-1}, x_t])$$

$$(1-p)h + pk$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

后厂理工学院
houchangtech.com

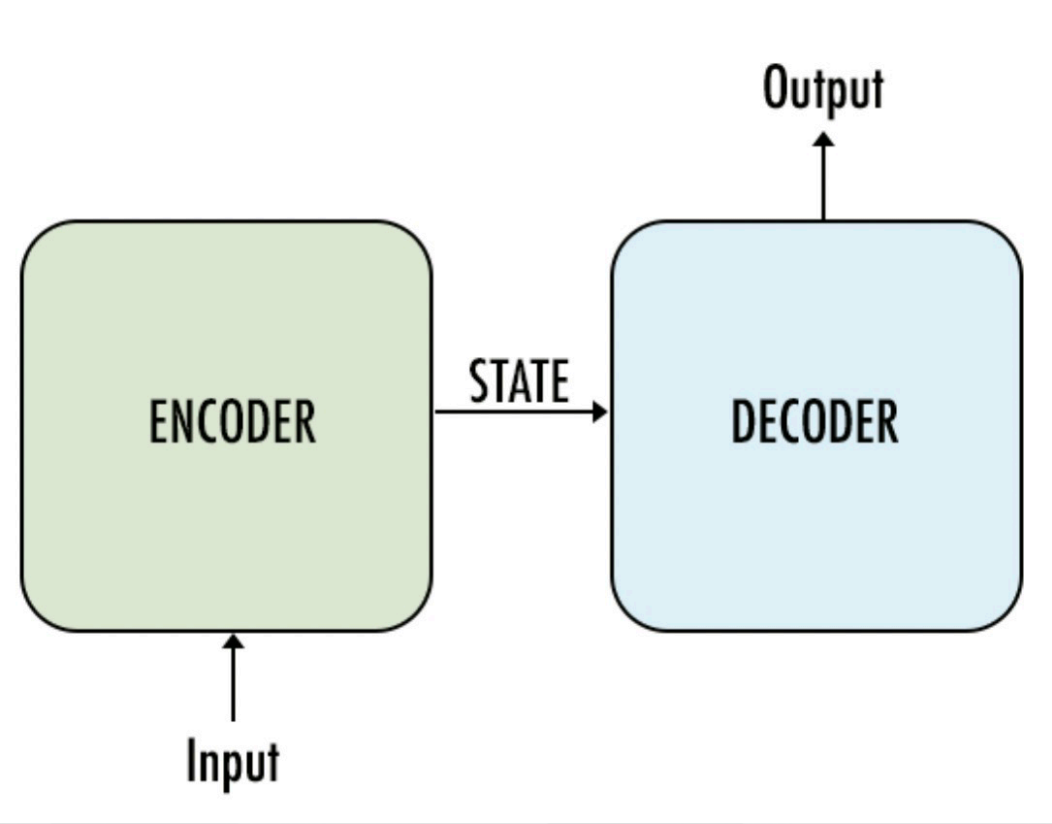
# 2. Encoder-Decoder结构

22:00

# Seq2Seq

## Encoder-Decoder结构

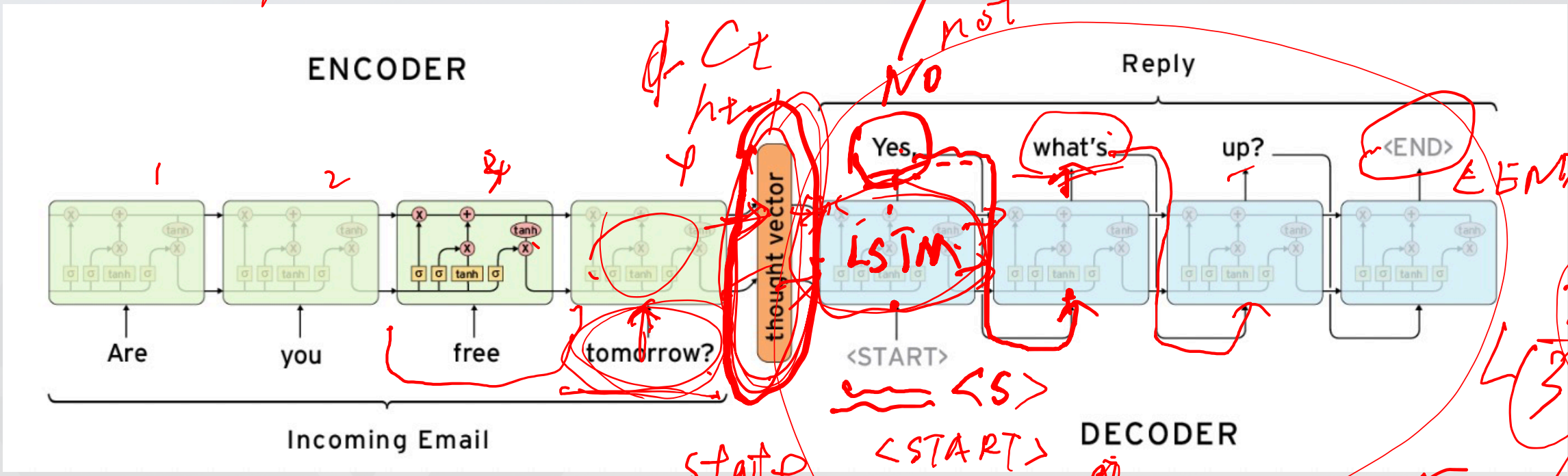
Sequence to sequence was first introduced by Google in 2014




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

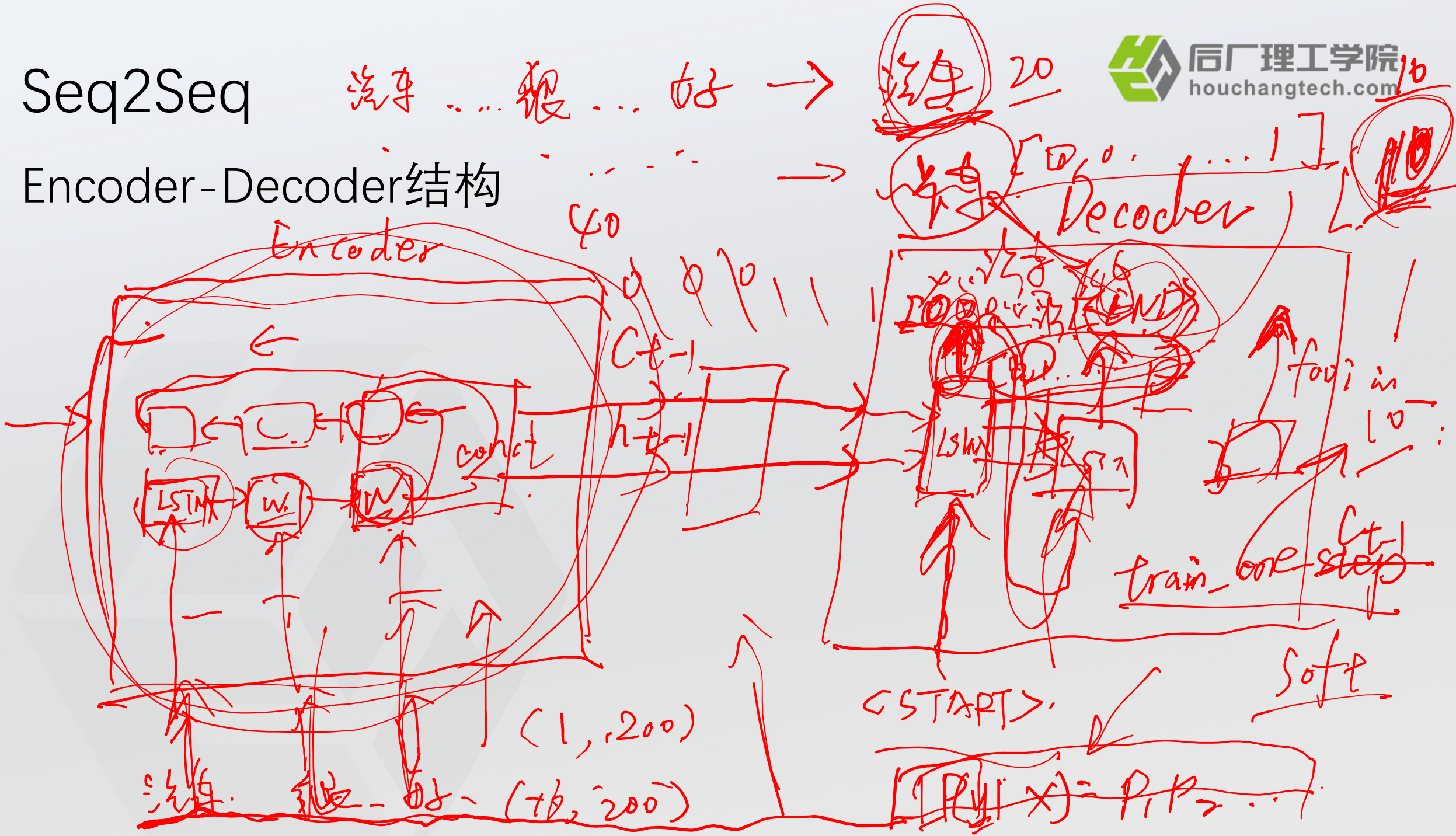
# Seq2Seq

## Encoder-Decoder结构



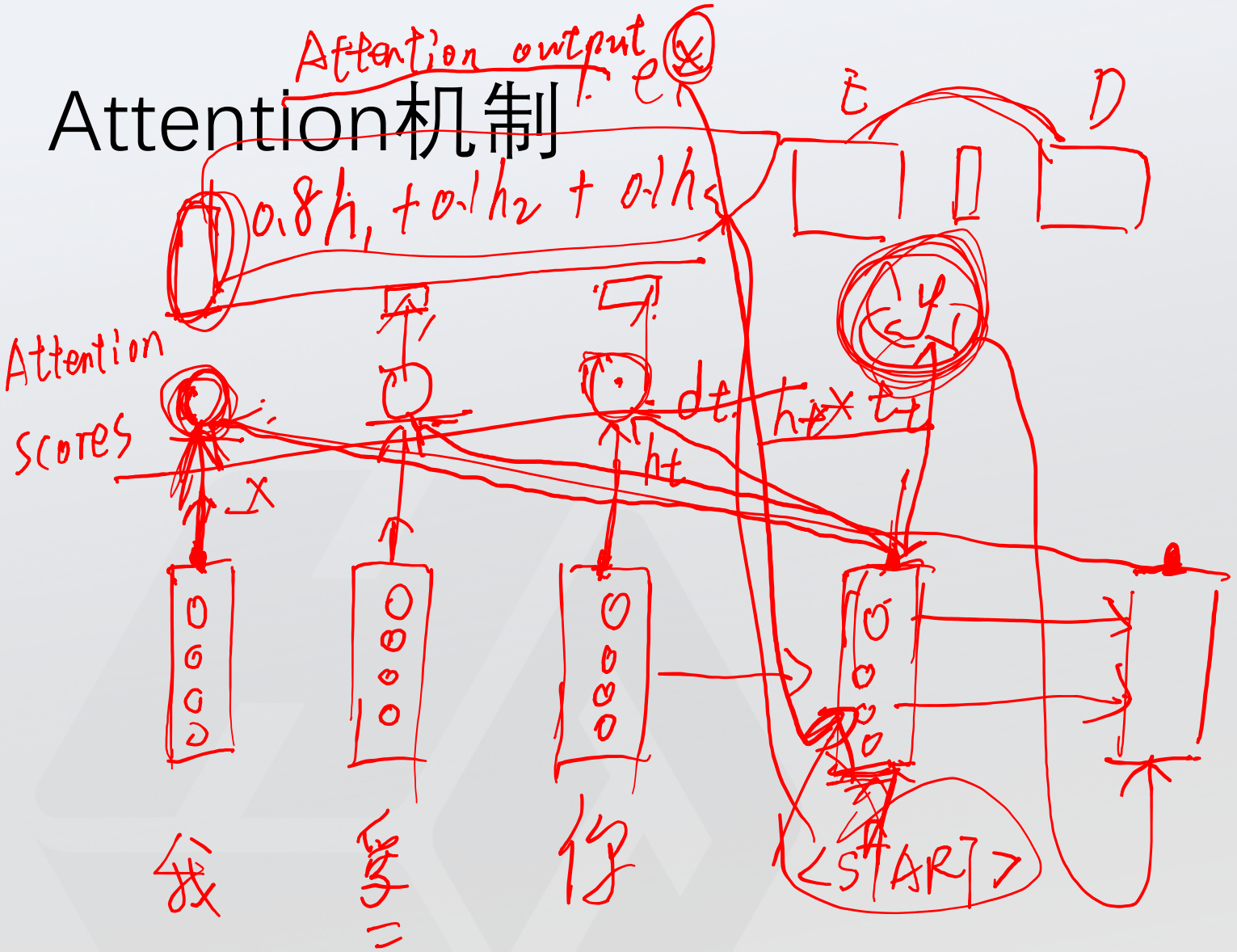
# Seq2Seq

## Encoder-Decoder结构



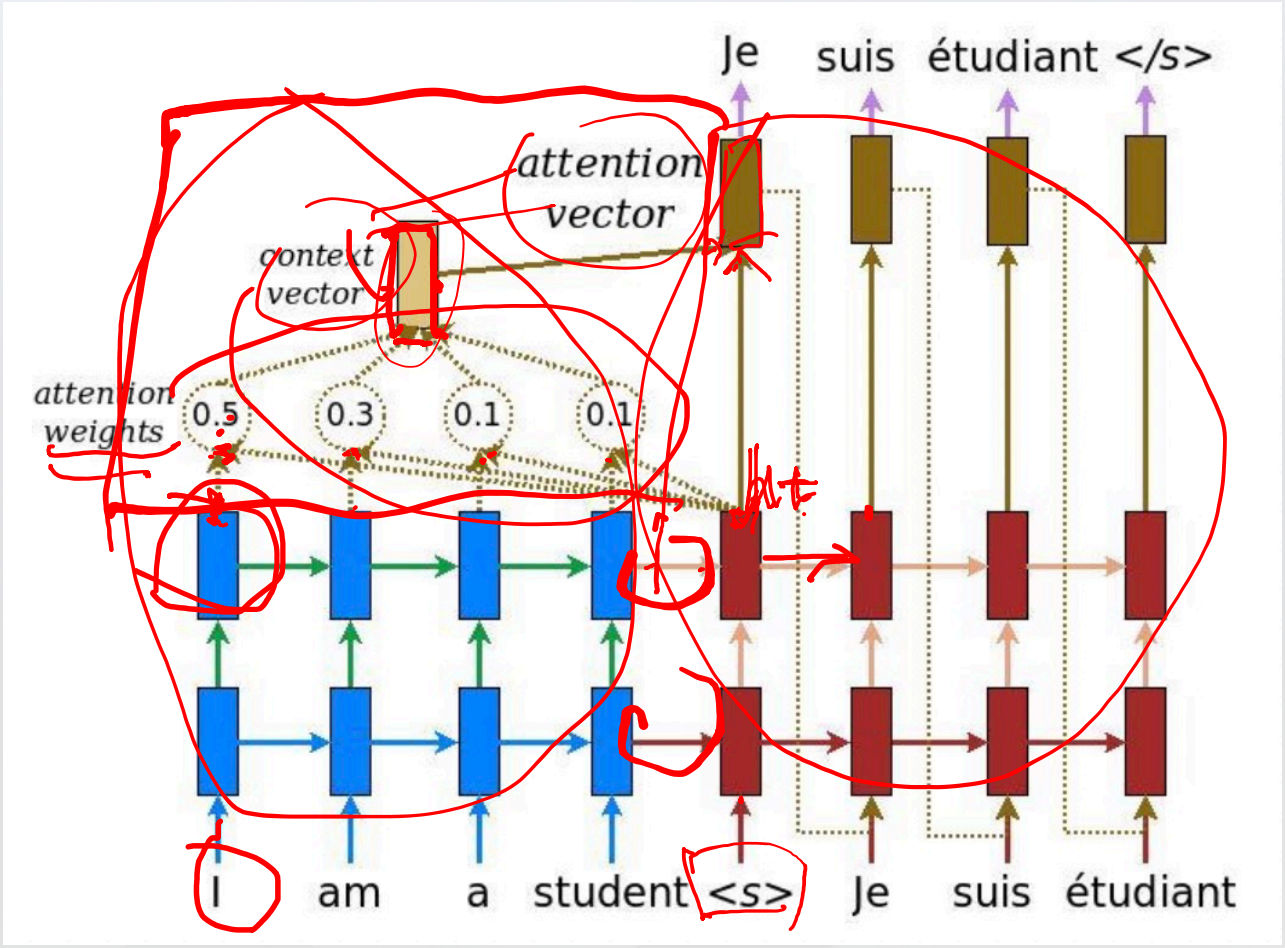
# 3. Attention机制

# Attention机制




后厂理工学院
houchangtech.com

# Attention机制



后厂理工学院
houchangtech.com

# 4. 作业

# Bye ！