



中国石油大学
CHINA UNIVERSITY OF PETROLEUM

《油气人工智能基础及应用》

4.6 循环神经网络

董少群

dshaoqun@163.com

理学院数学系

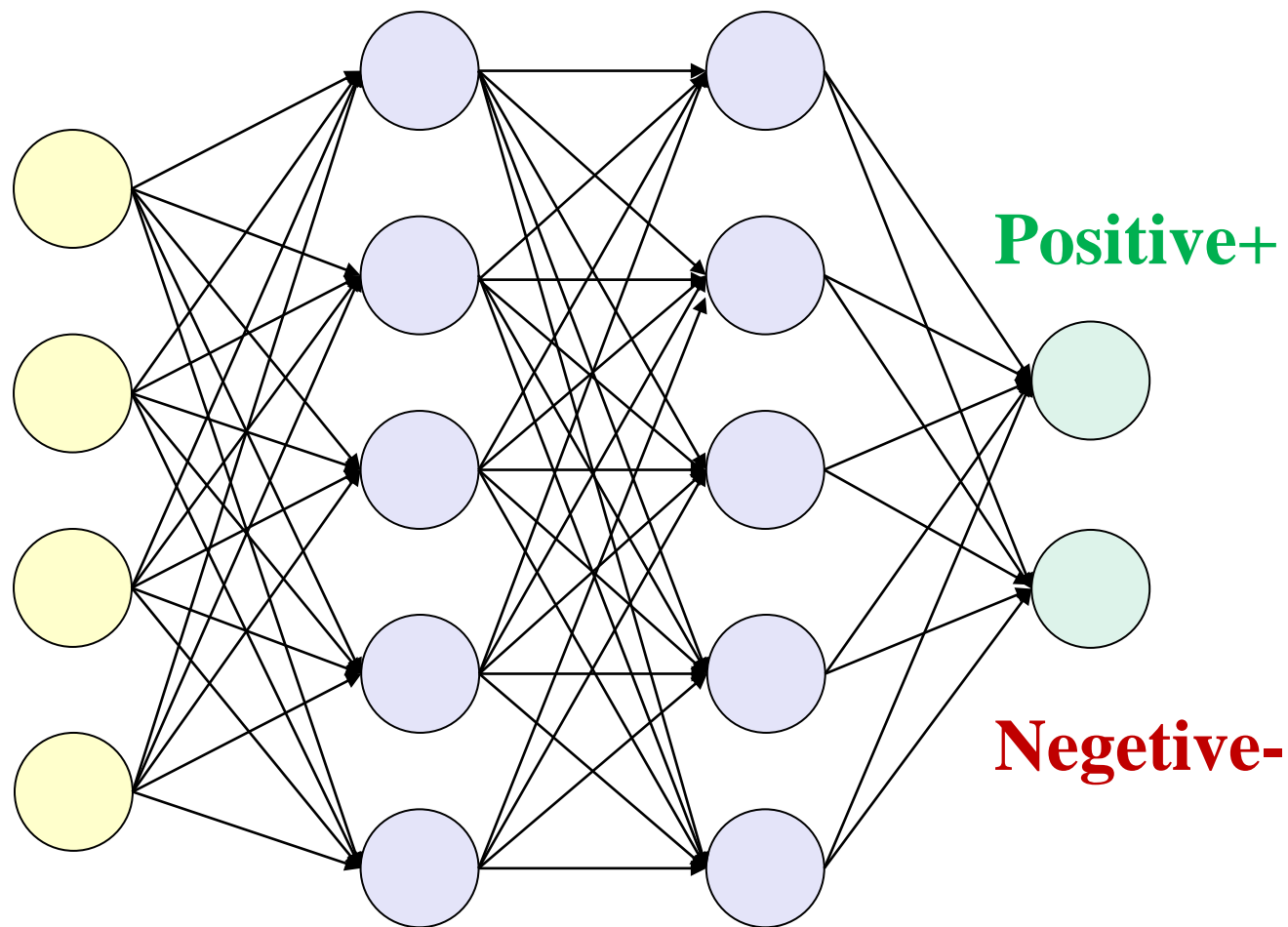
CUP

厚积薄发 开物成务

- 一、为什么需要循环神经网络
- 二、循环神经网络发展概述
- 三、循环神经网络改进算法

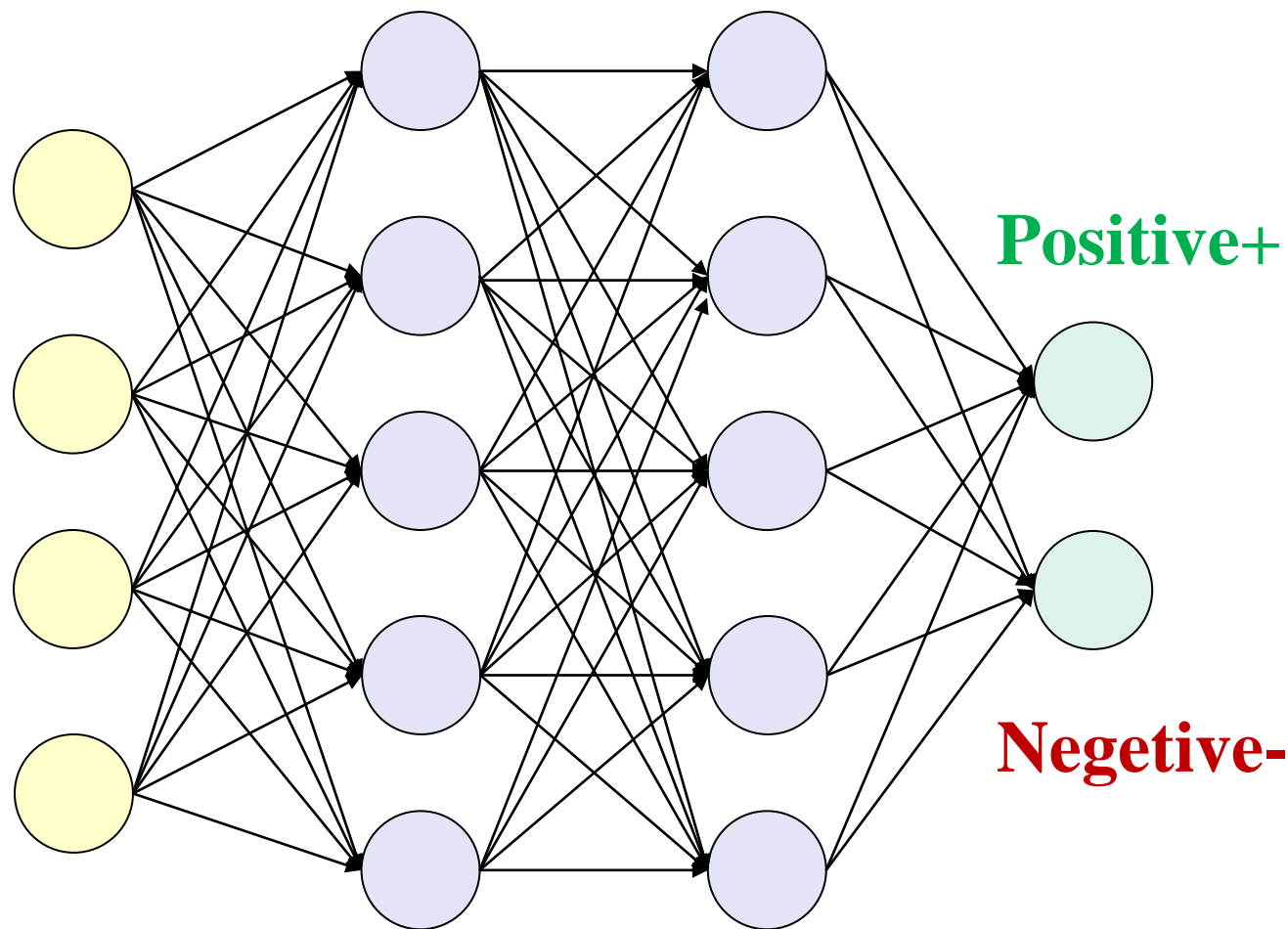
判断语句情绪

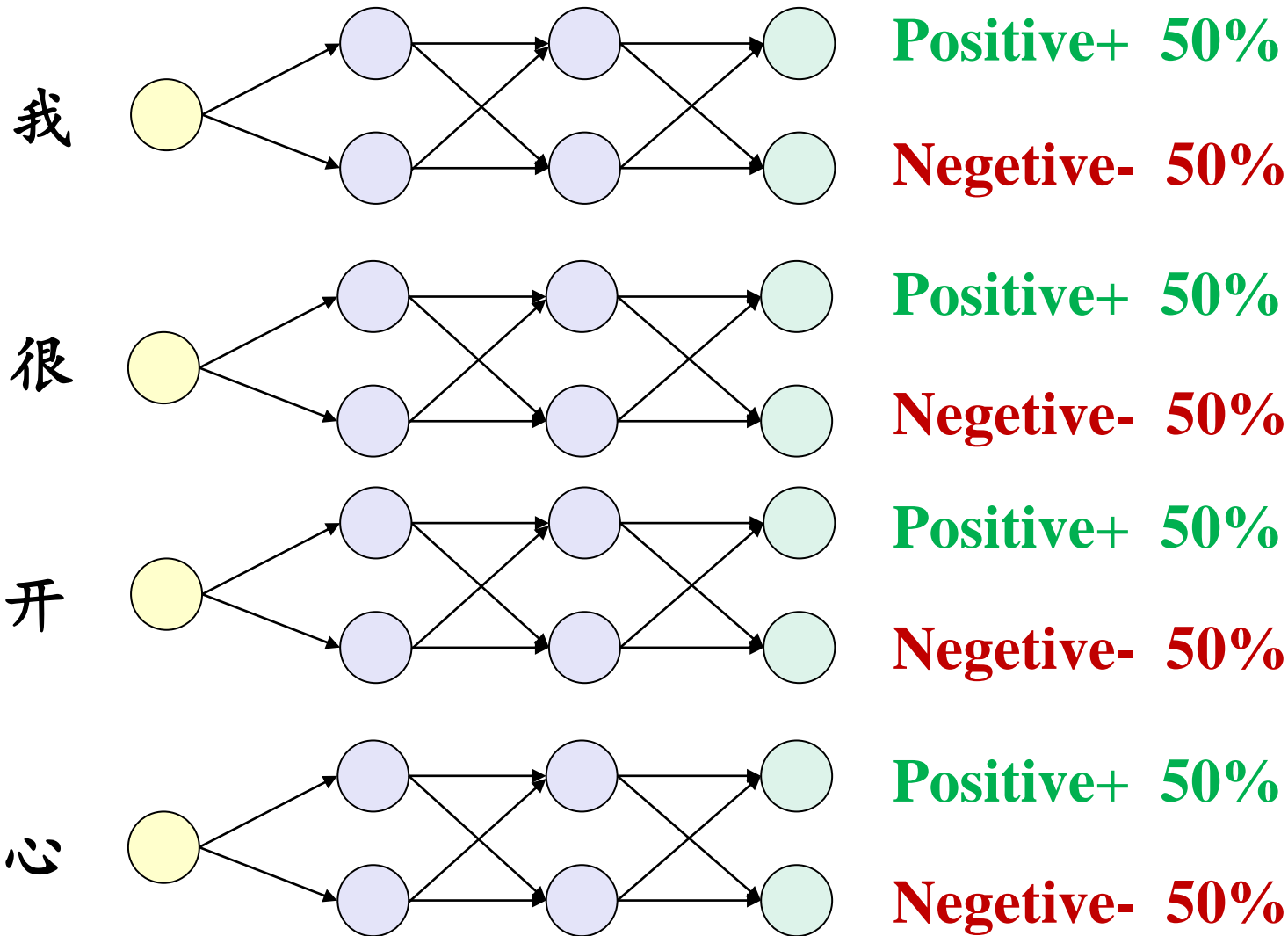
- 1.我很开心
- 2.我很难过
- 3.大家爱我
- 4.我爱大家
- 5.大家恨我
- 6.我恨大家
- 7.我中奖了
8.

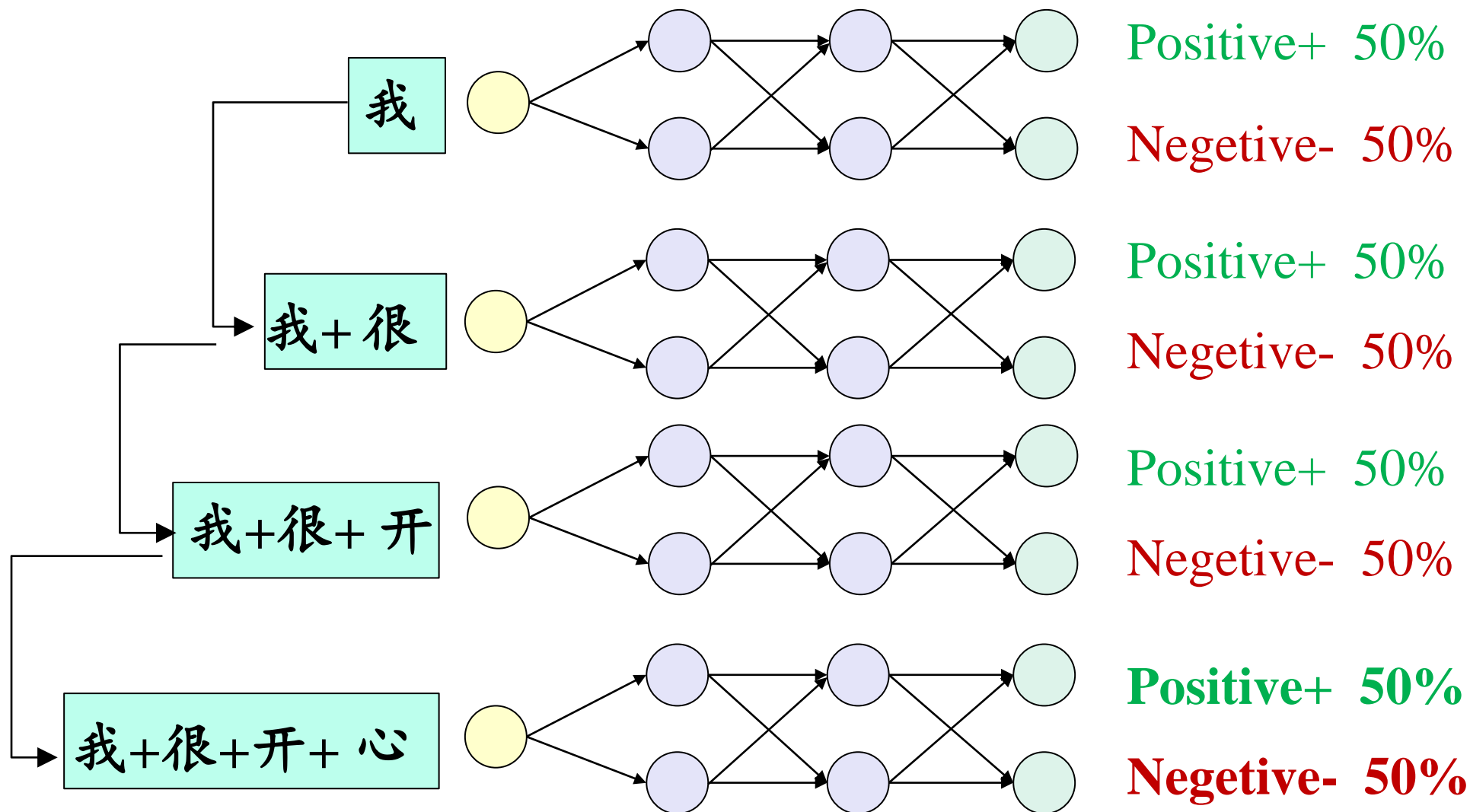


判断语句情绪

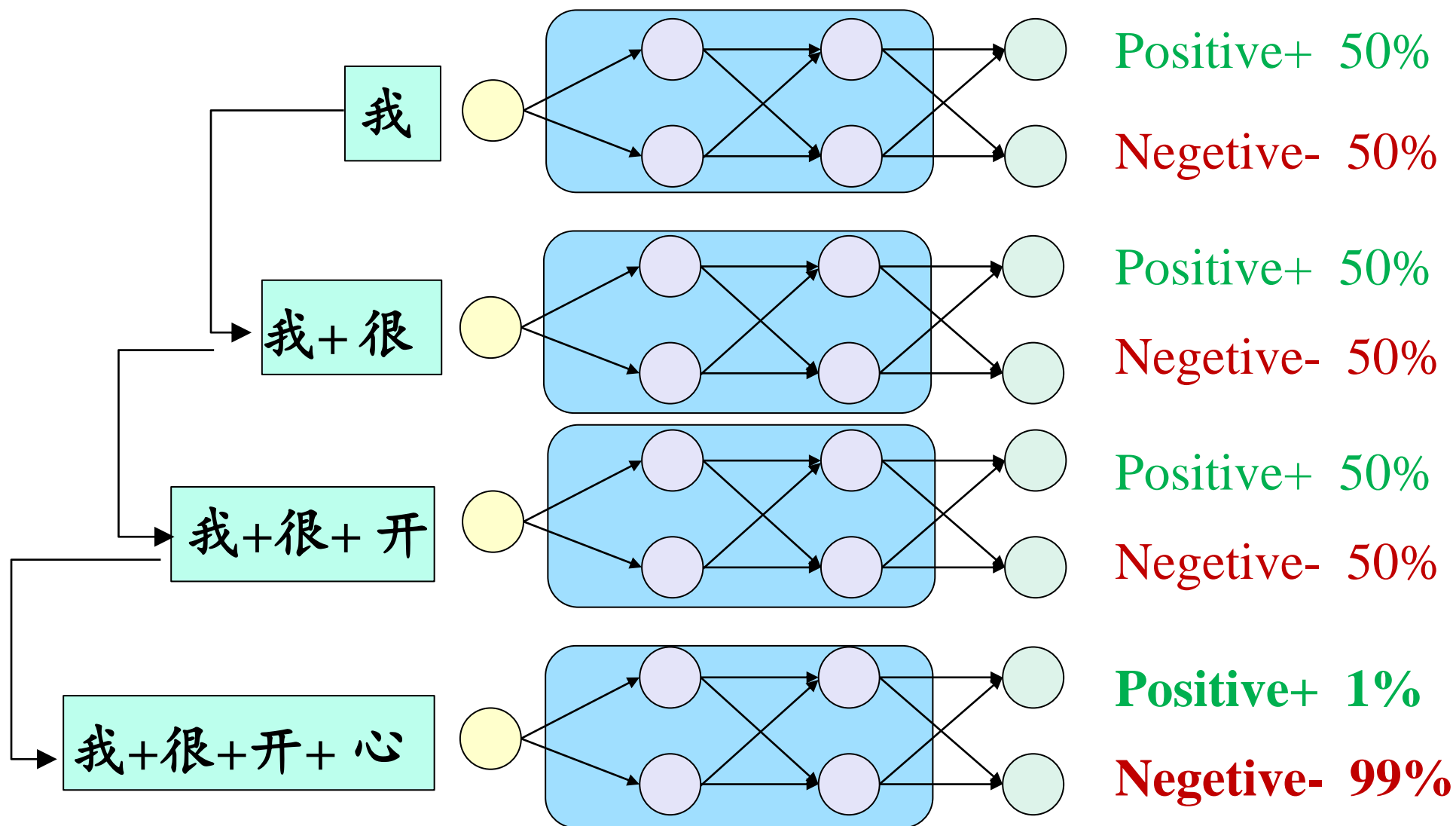
- 1.我很喜欢我手机的颜色
- 2.我很开心
- 3.我很讨厌我手机的相机功能
- 4.我emo了
- 5.今天天气很好
- 6.我得到了大家的认可
- 7.今天有人批评我了
8.







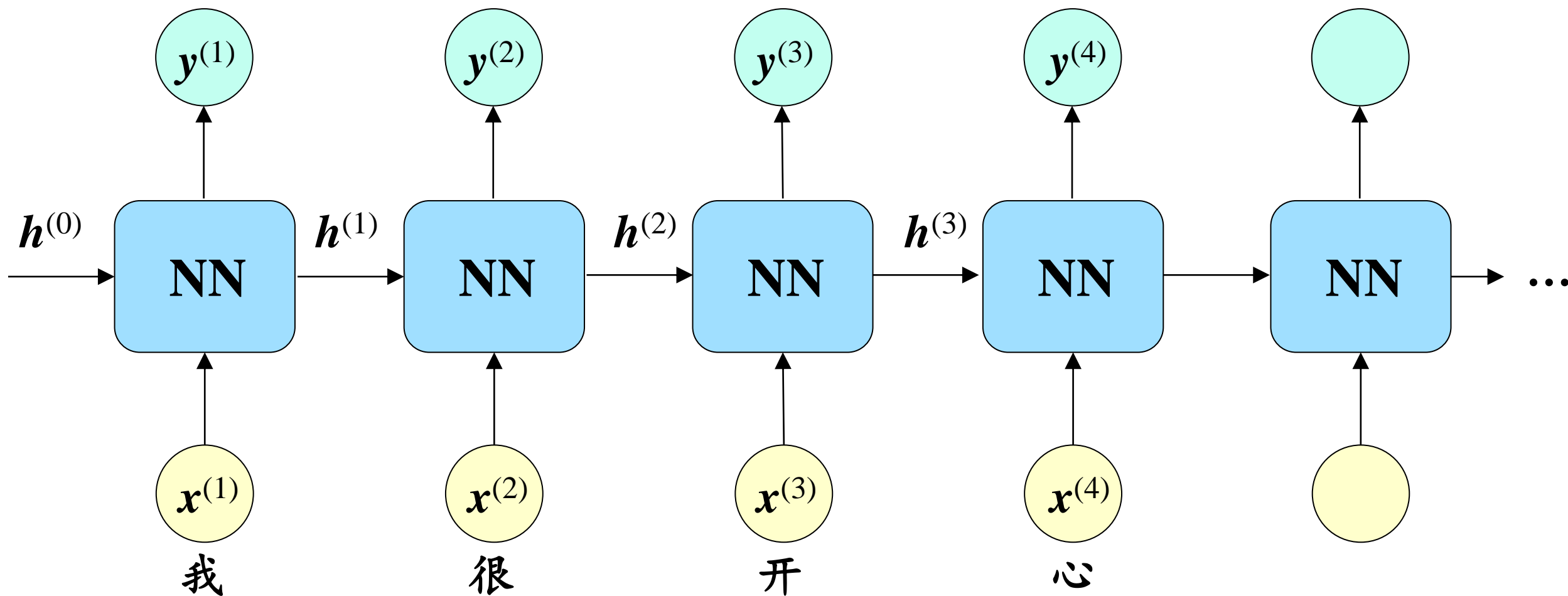
NN

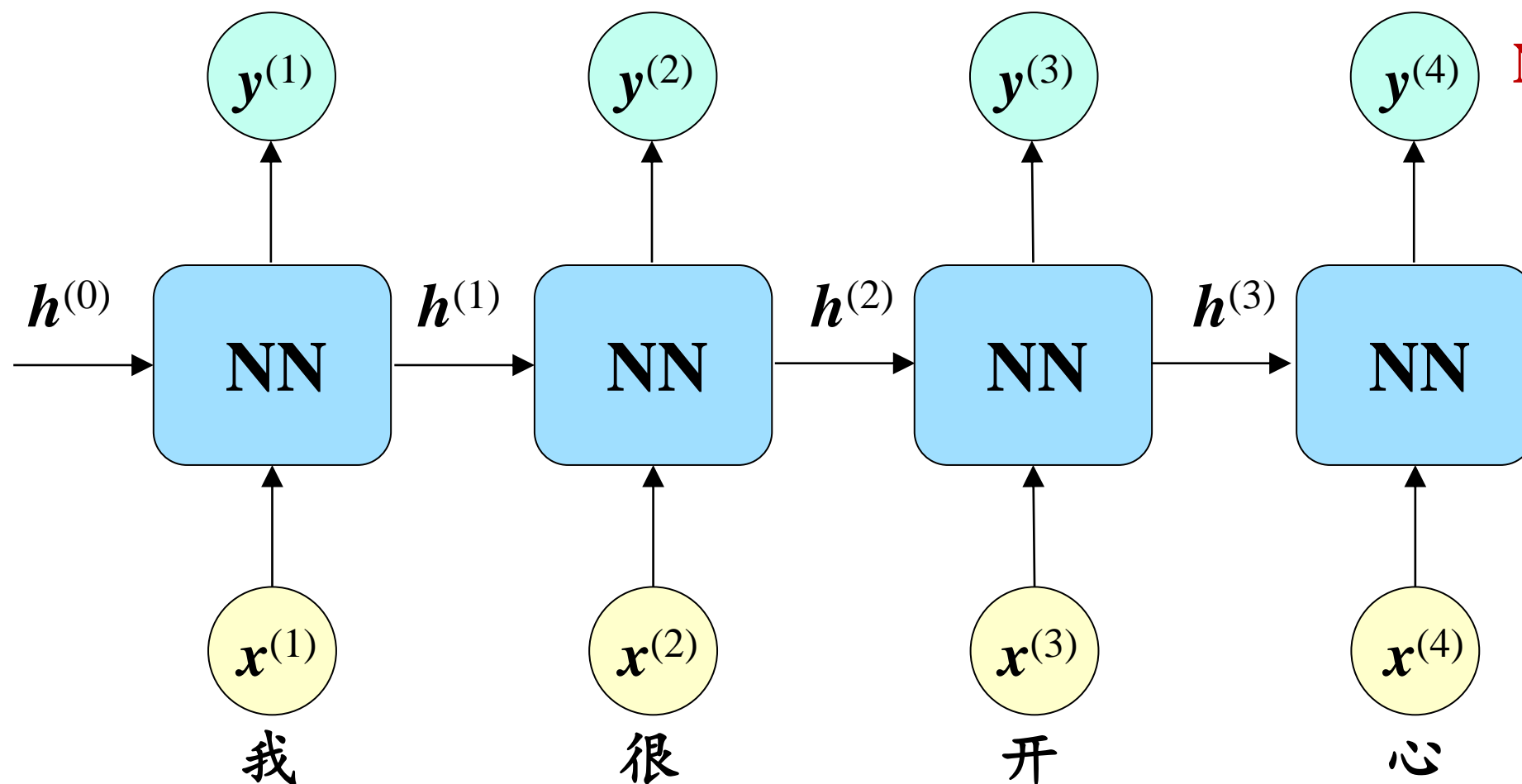


循环神经网络 (Recurrent neural network, RNN)



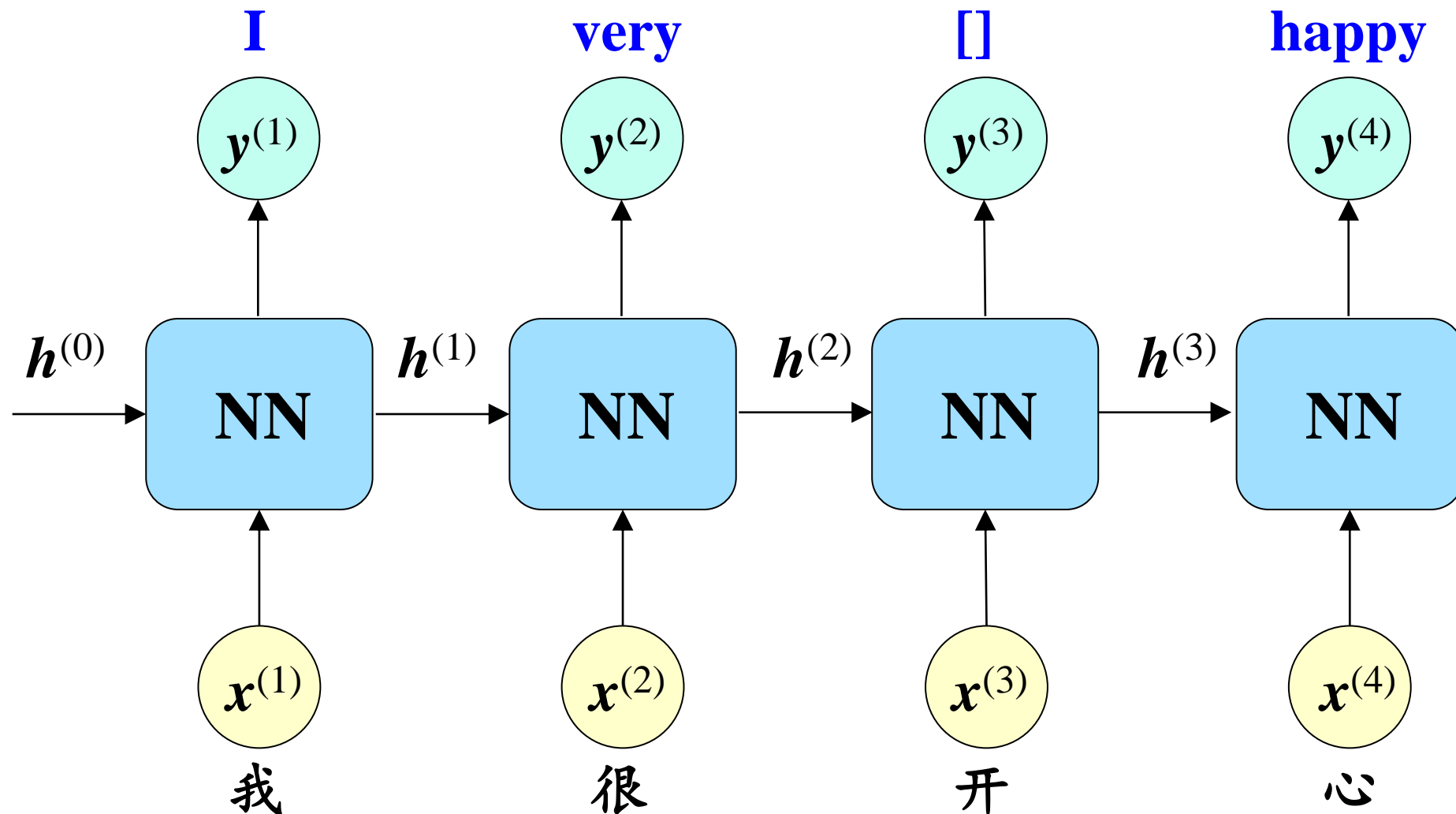
中国石油大学
CHINA UNIVERSITY OF PETROLEUM





Positive+ 1%

Negetive- 99%



翻译:

小米小爱鼠标



- 短按语音键，语音变文字
聊天、码字更轻松
- 长按语音键，唤醒小爱同学
查天气、开空净，样样精通



翻译键

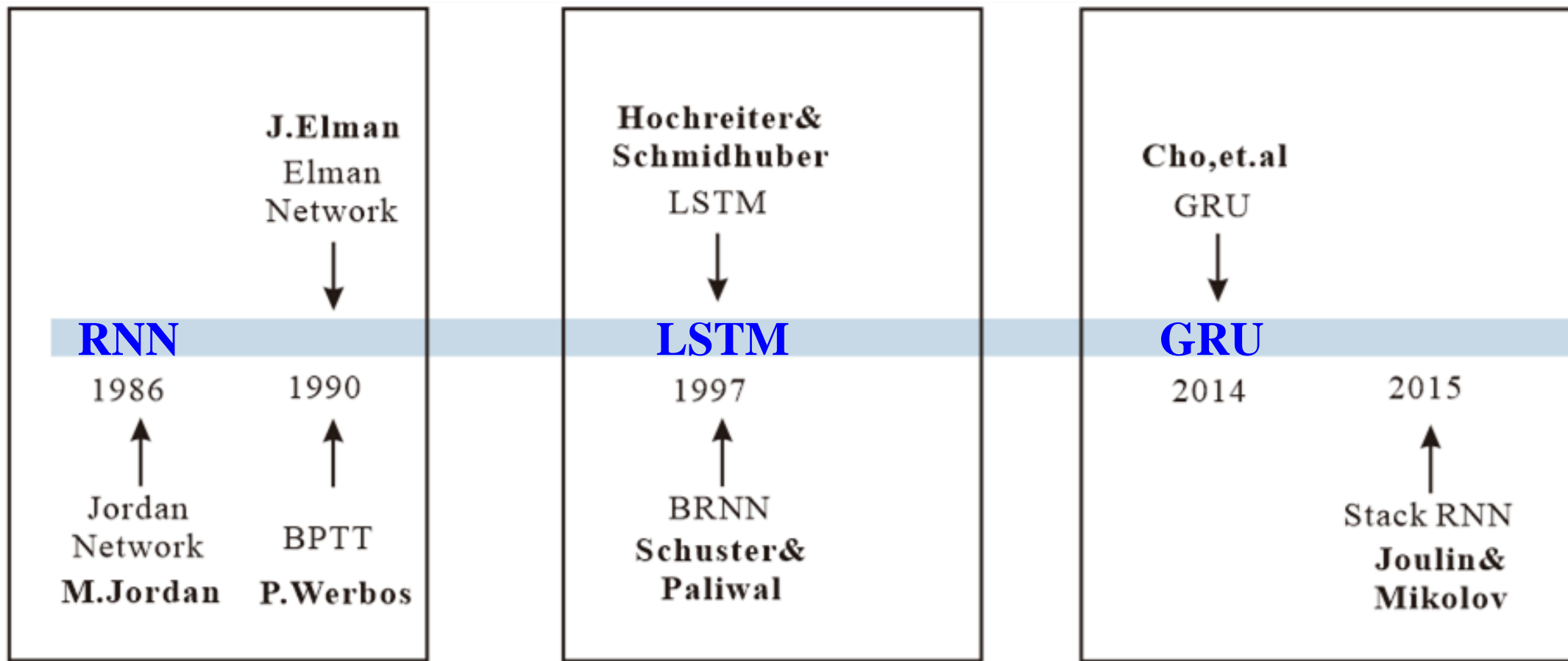


- 一、为什么需要循环神经网络
- 二、循环神经网络发展概述
- 三、循环神经网络改进算法

二、循环神经网络发展概述



RNN (M.Jordan,1986) 是一类用于处理序列数据的神经网络，应用于输入数据具有依赖性且是序列模式时的场景，即前一个输入和后一个输入是有关系的。



二、循环神经网络发展概述



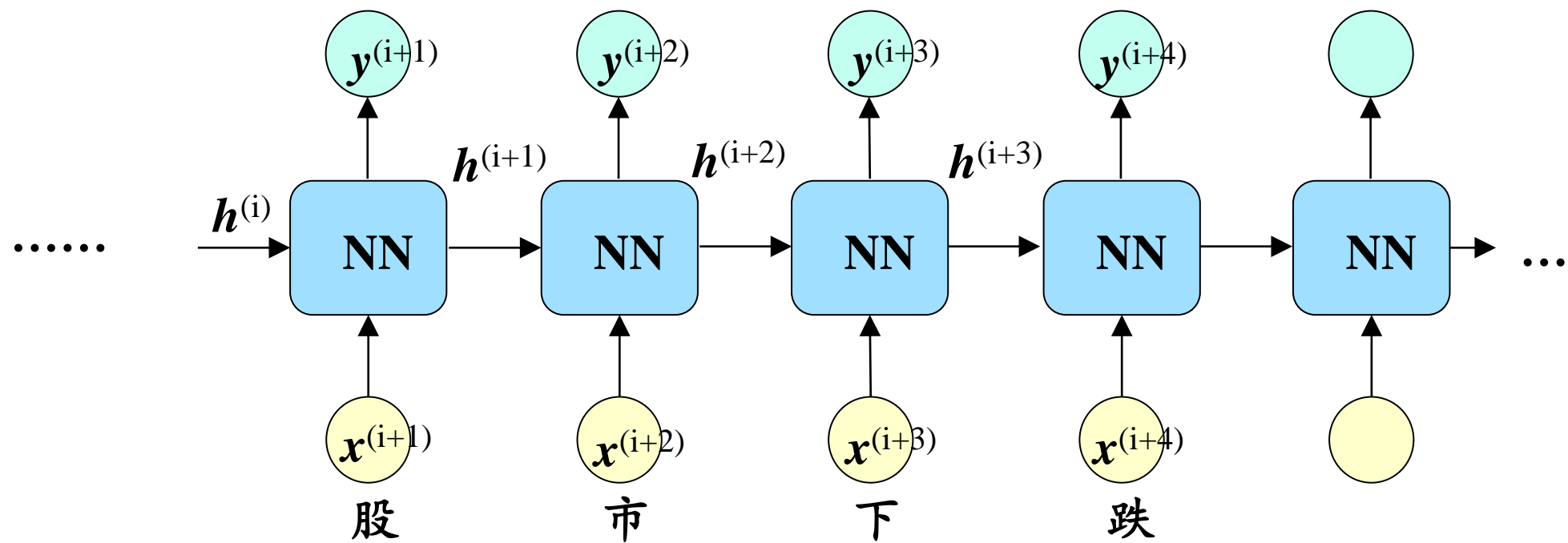
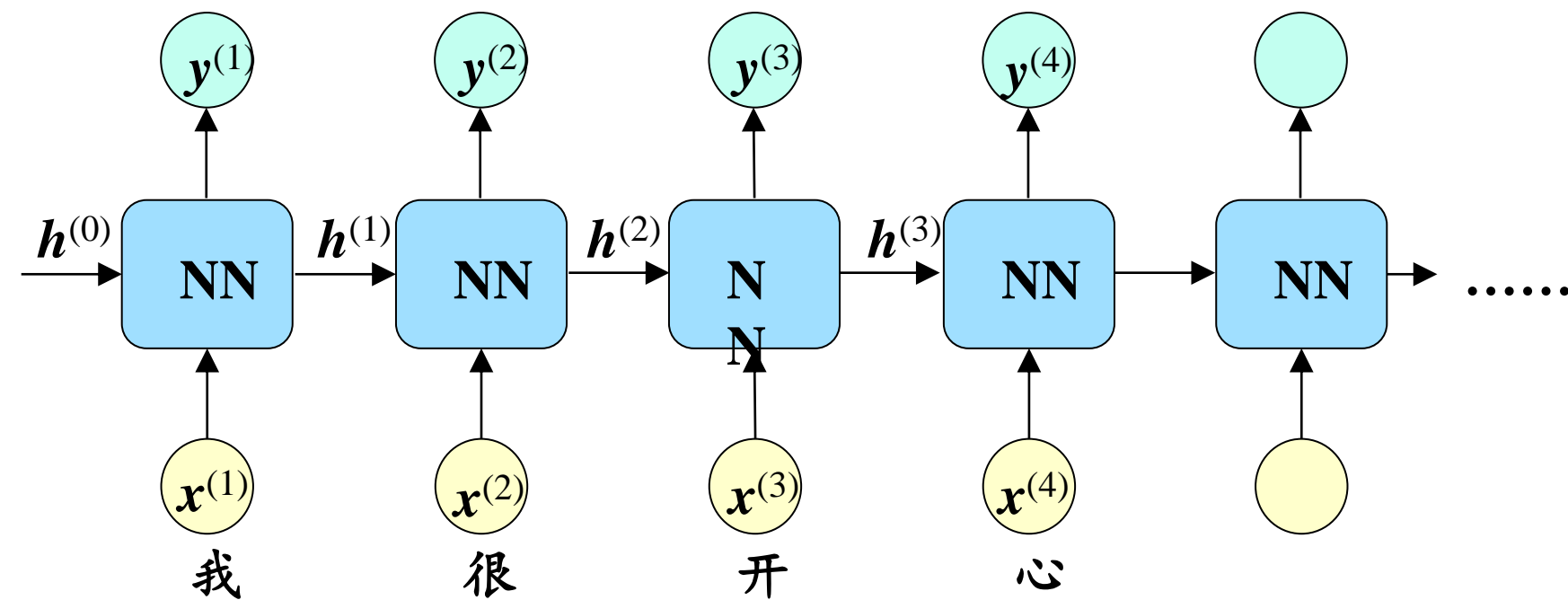
年份	人物	网络	简要说明
1986	Jordan	Jordan Network	Jordan网络的每个隐含层节点都与一个状态单元相连以实现延时输入，并使用logistic函数作为激活函数。
1990	Elman	Elman Network	提出了第一个全连接的RNN，即Elman网络。
1997	Hochreiter&Schmidhuber	LSTM	为解决长期依赖问题，提出长短期记忆网络。
	Schuster&Paliwal	BRNN	提出了具有深度结构的双向循环神经网络，并对其进行了语音识别试验。
2014	Cho	GRU	Cho提出了门控循环单元网络，该方法是LSTM之后另一个受到关注的RNN门控构架。

二、循环神经网络发展概述



年份	方法	人物	论文	引用
1986	Jordan Network	Jordan	Jordan, M I. Serial order: a parallel distributed processing approach. Technical report, June 1985-March 1986. United States: N. p., 1986. Web.	
1990	Elman Network	Elman	Elman J L. Finding structure in time[J]. Cognitive science, 1990, 14(2): 179-211.	14166
1997	LSTM	Hochreiter & Schmidhuber	Hochreiter, Sepp & Schmidhuber, Jürgen. (1997). Long Short-term Memory. Neural Computation. 9. 1735-80.	72946
	BRNN	Schuster & Paliwal	Schuster M, Paliwal K K. Bidirectional recurrent neural networks[J]. IEEE transactions on Signal Processing, 1997, 45(11): 2673-2681.	7544
2014	GRU	Cho	Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv preprint arXiv:1406.1078, 2014.	21023
2015	Stack RNN	Joulin & Mikolov	Joulin, A., & Mikolov, T. (2015). Inferring algorithmic patterns with stack-augmented recurrent nets. In Advances in Neural Information Processing Systems (pp. 190-198).	406

- 一、为什么需要循环神经网络
- 二、循环神经网络发展概述
- 三、循环神经网络改进算法



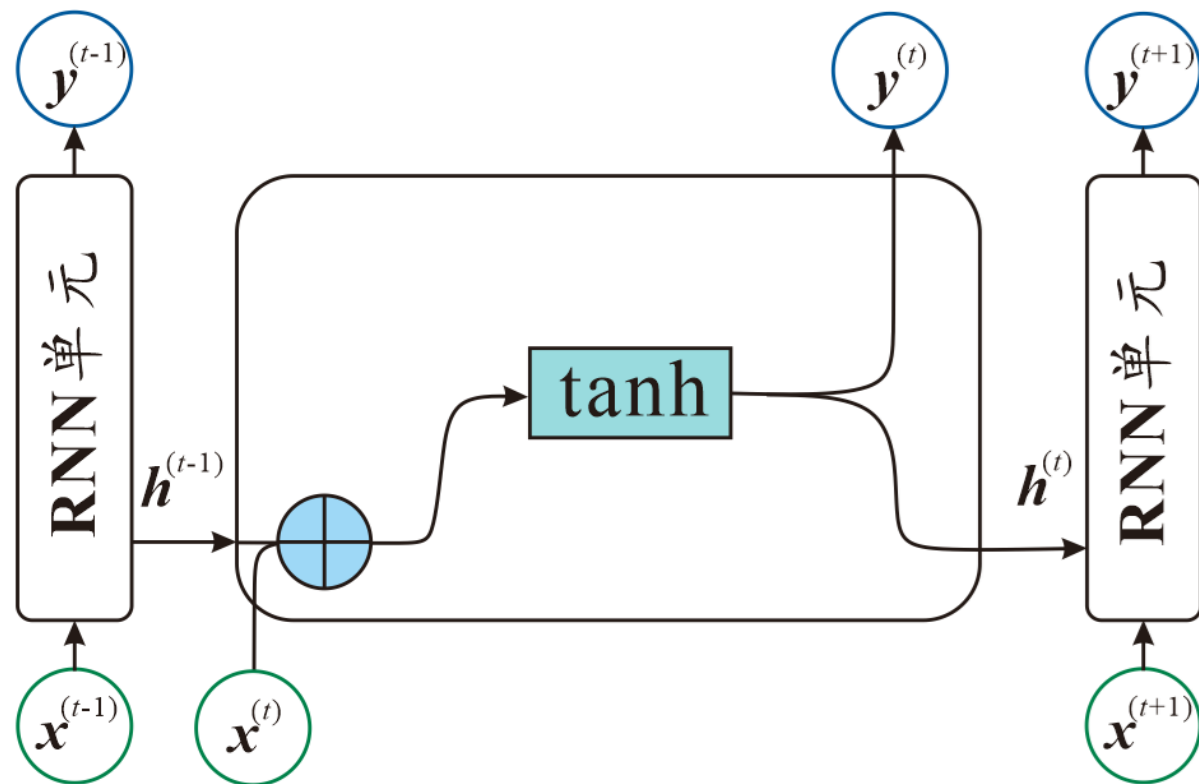
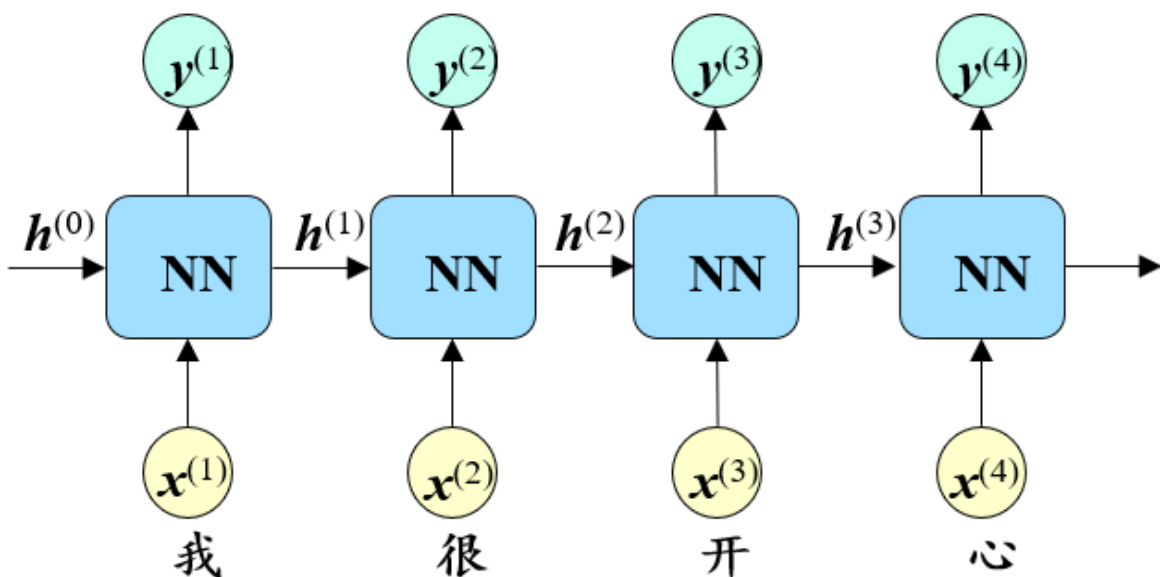
三、循环神经网络改进算法



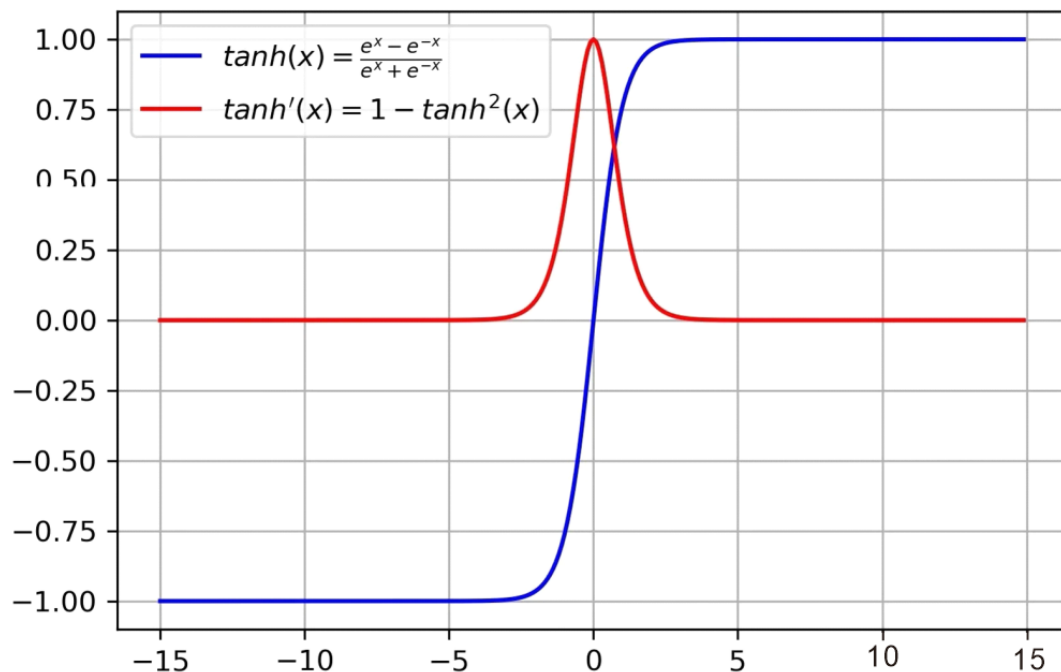
➤ 循环神经网络 (RNN)

$$\mathbf{h}^{(i)} = \alpha_h (\mathbf{W}_{hh} \mathbf{h}^{(i-1)} + \mathbf{W}_{hx} \mathbf{x}^{(i-1)} + \mathbf{b}_h)$$

$$\mathbf{y}^{(i)} = \alpha_y (\mathbf{W}_{yh} \mathbf{h}^{(i)} + \mathbf{b}_y)$$



三、循环神经网络改进算法



tanh 函数的基本形式: $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

tanh 函数的导数形式: $\tanh'(x) = (1 - \tanh^2(x))$

➤ 梯度消失

当 V 处于0~1之间时，时间序列足够长（ t 足够大）：

$$\prod_{j=k+1}^t \tanh' V \longrightarrow 0$$

➤ 梯度爆炸

当 V 的值很大， t 足够大：

$$\prod_{j=k+1}^t \tanh' V \longrightarrow \infty$$

三、循环神经网络改进算法



➤ 长程依赖问题

简单循环网络理论上可以建立长时间间隔的状态之间的依赖关系，但由于梯度爆炸或消失问题，实际上只能学习到短期的依赖关系，很难建模这种长距离的依赖关系。

➤ 改进方案



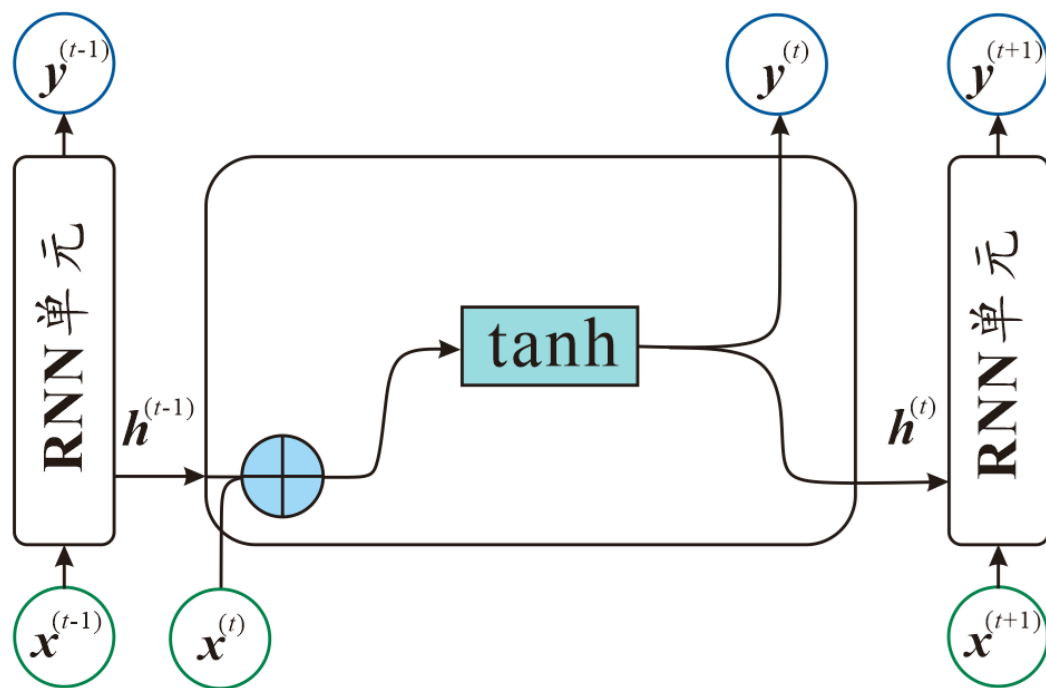
三、循环神经网络改进算法



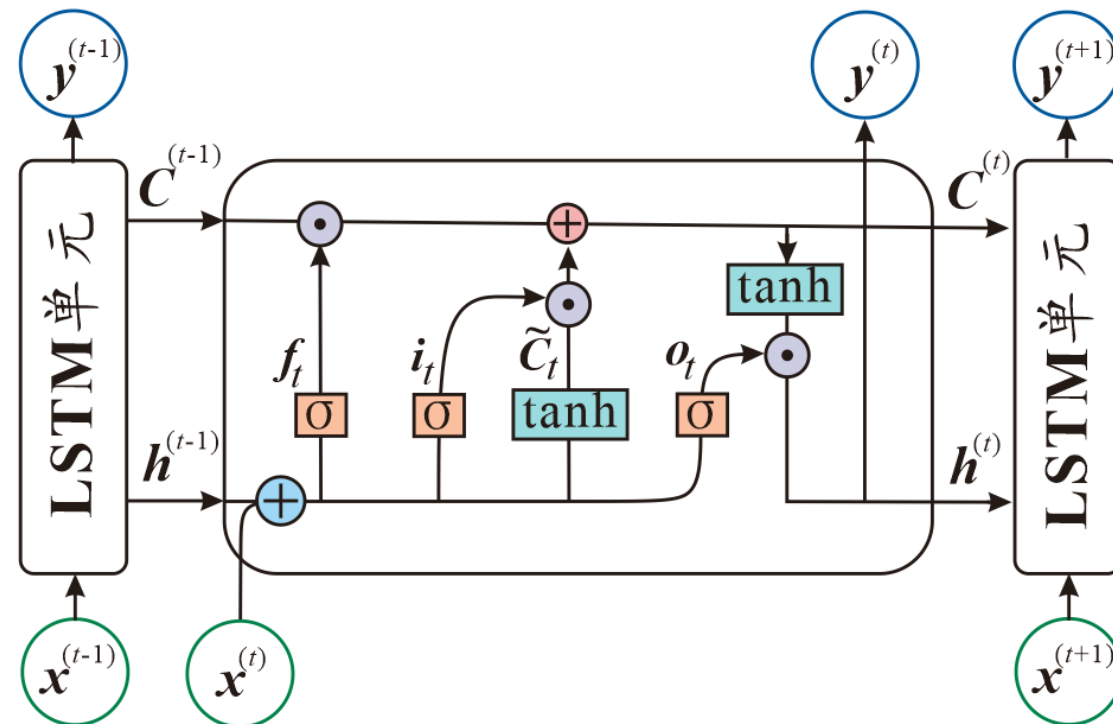
➤ 长短期记忆神经网络 (Long Short-Term Memory Network, LSTM)

加入3个门结构来控制信息在序列中的传递，使信息能够长期传递下去。

遗忘门、输入门、输出门



RNN



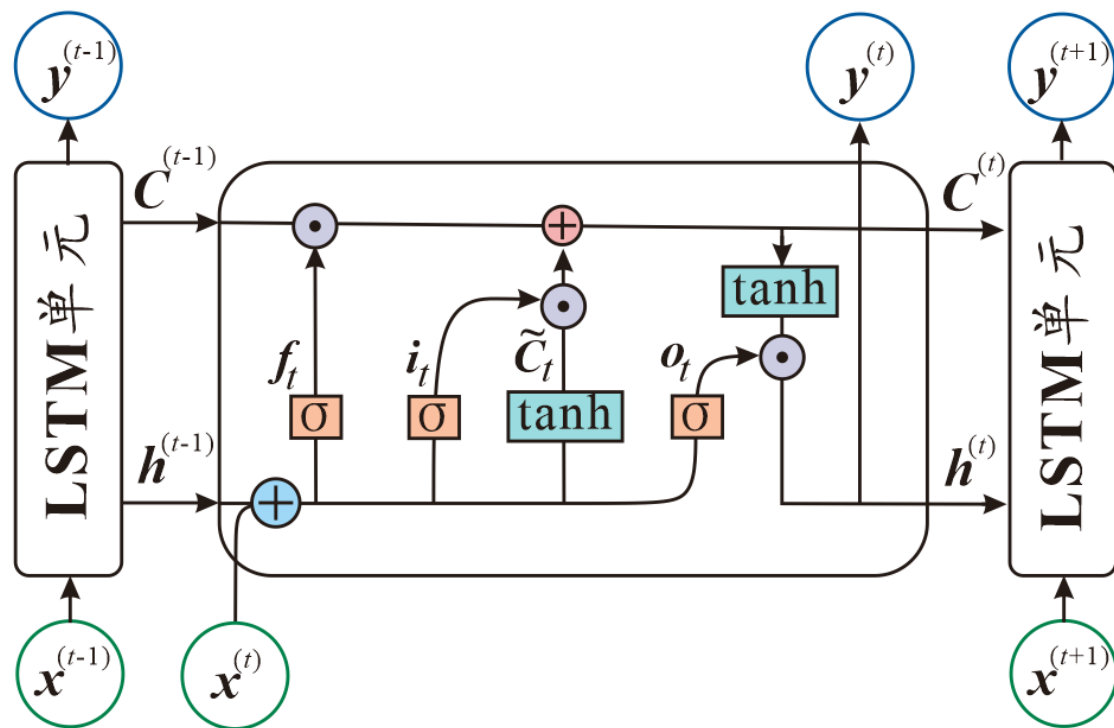
LSTM

三、循环神经网络改进算法

LSTM



中国石油大学
CHINA UNIVERSITY OF PETROLEUM



$$h^{(t)} = o_t \odot \tanh(C^{(t)})$$

$$o_t = \sigma(W_o x^{(t)} + U_o h^{(t-1)} + b_o)$$

$$C^{(t)} = f_t \odot C^{(t-1)} + i_t \odot \tilde{C}^{(t)}$$

$$f_t = \sigma(W_f \cdot x^{(t)} + U_f \cdot h^{(t-1)} + b_f)$$

$$i_t = \sigma(W_i \cdot x^{(t)} + U_i \cdot h^{(t-1)} + b_i)$$

$$\tilde{C}^{(t)} = \tanh(W_c \cdot x^{(t)} + U_c \cdot h^{(t-1)} + b_c)$$

$$C^{(t)}, h^{(t)} = \text{LSTM}(x^{(t)}, C^{(t-1)}, h^{(t-1)})$$

tanh \tanh 激活函数

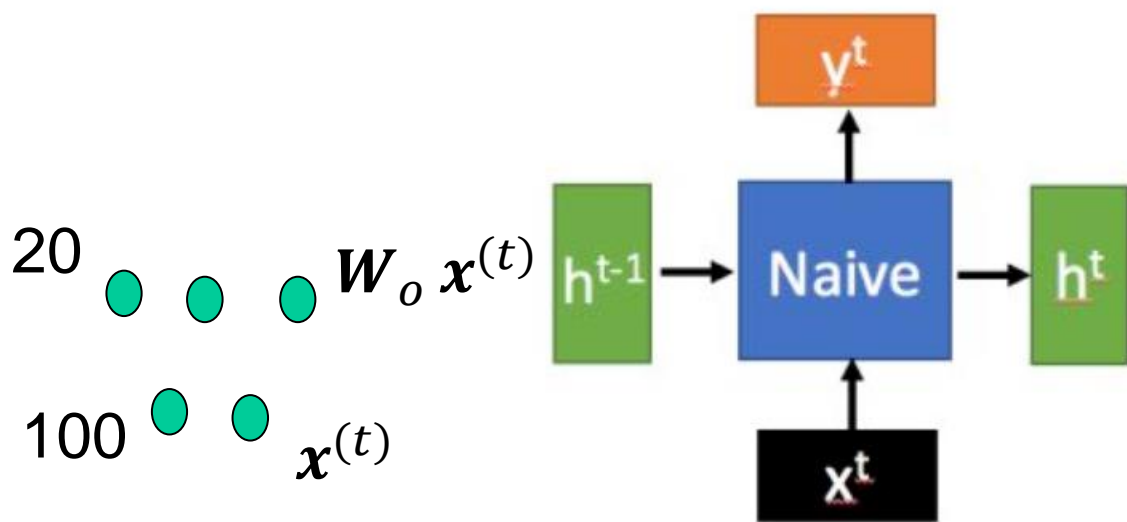
σ σ 为 sigmoid 激活函数

+ 向量拼接 **⊕** 向量和

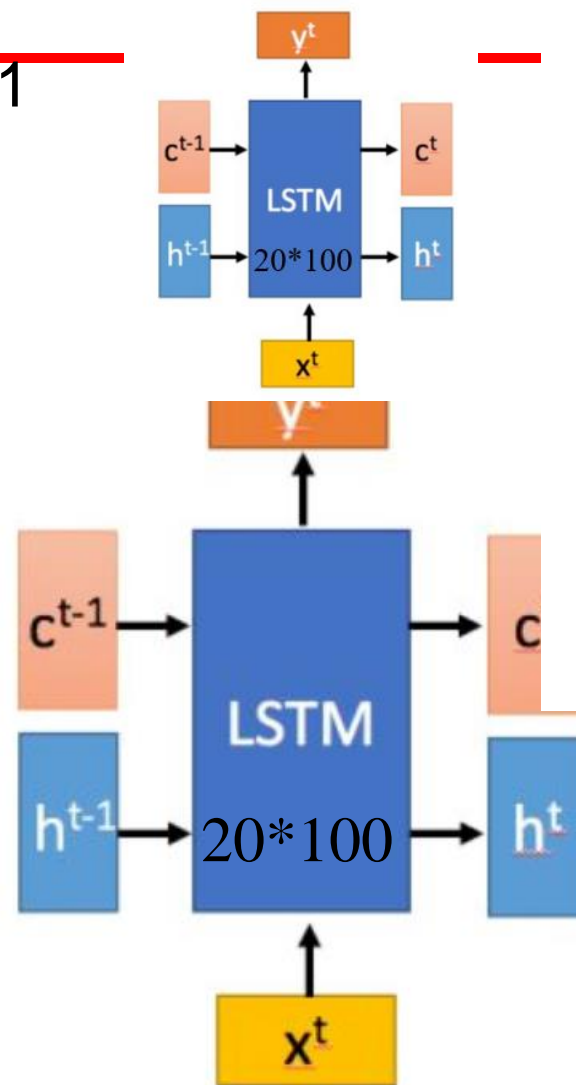
⊙ \odot 表示 Hadamard 积，即逐元素对应相乘

- `input_size`: 输入的维度=`embedding_size`
- `hidden_size`: `h`的维度
- `num_layers`: 堆叠LSTM的层数, 默认值为1

LSTM



字典200000



[1,0,...,1] 100

我

`input_size=100`

- **input**: tensor of shape (L, H_{in}) for unbatched input, (L, N, H_{in}) or (N, L, H_{in}) when `batch_first=True` containing the features. `input` can also be a packed variable length sequence. See [torch.nn.utils.rnn.pack_sequence\(\)](#) for details.
- **h_0**: tensor of shape $(D * \text{num_layers}, H_{out})$ for unbatched input, containing the initial hidden state for each element in the input sequence provided.
- **c_0**: tensor of shape $(D * \text{num_layers}, H_{cell})$ for unbatched input, containing the initial cell state for each element in the input sequence provided.

where:

N = batch size

L = sequence length

$D = 2$ if `bidirectional=True` otherwise 1

H_{in} = `input_size`

H_{cell} = `hidden_size`

H_{out} = `proj_size` if `proj_size > 0` otherwise `hidden_size`

60 W

128

$[1,0,\dots,1]^T$

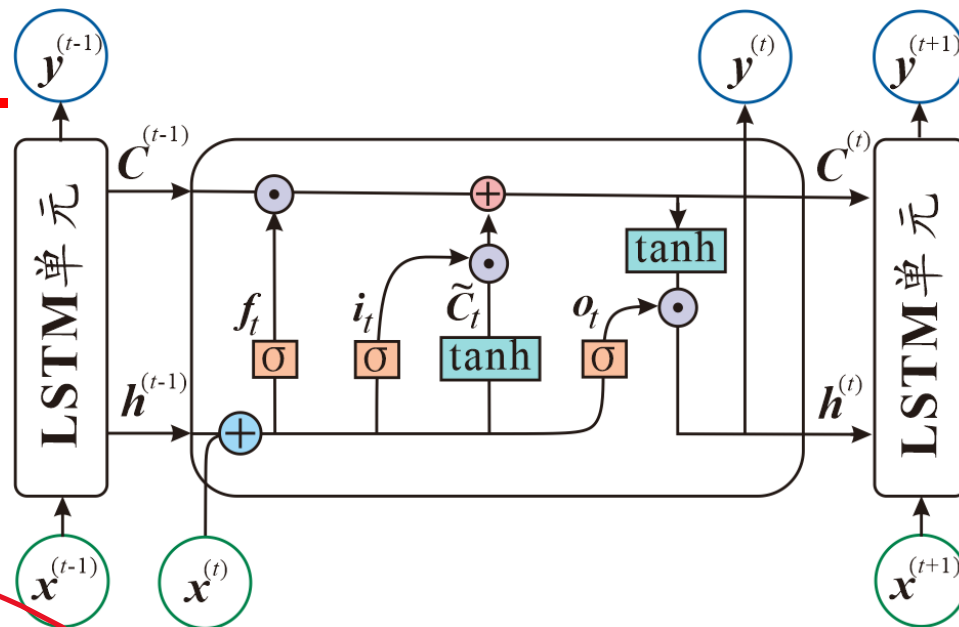
$[1,0,\dots,1]^T$

$[0,0,\dots,1]^T$

$[1,1,\dots,1]^T$

我 和 你

$[1,2,3]^T$

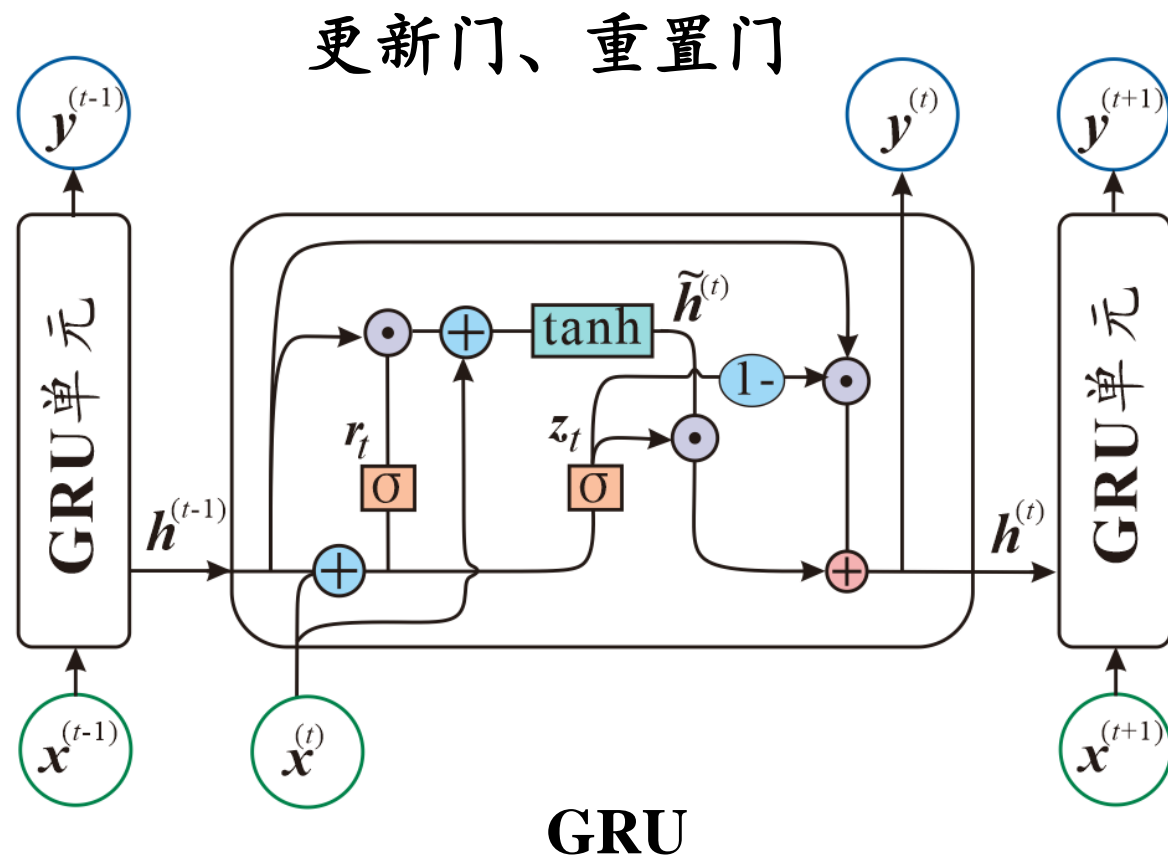
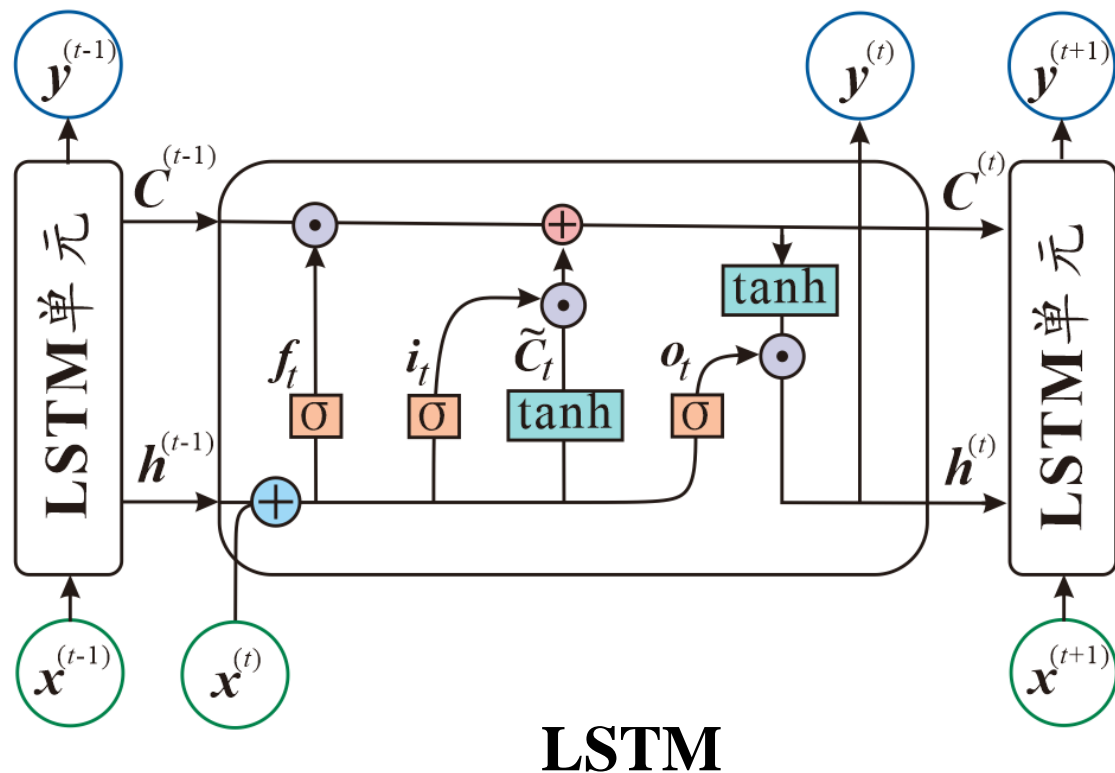


三、循环神经网络改进算法



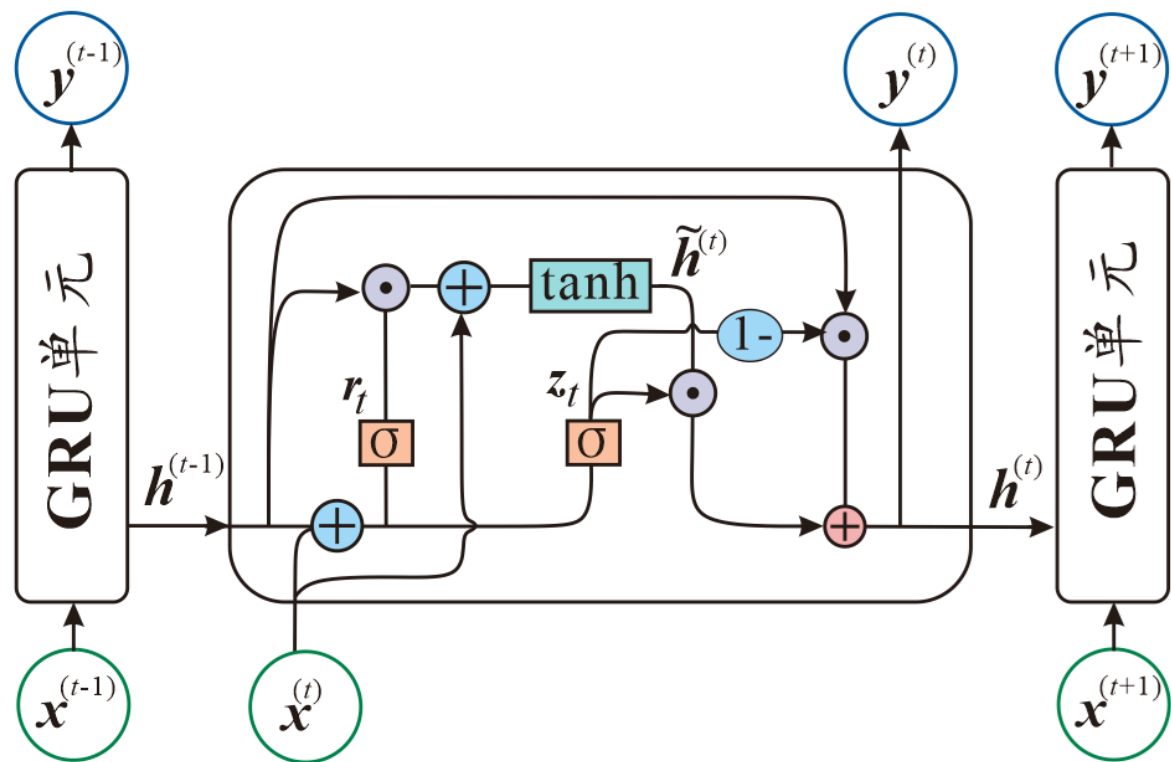
➤ 门控循环单元网络 (Gated recurrent unit, GRU)

GRU网络 [Choi et al., 2014; Chung et al., 2014] 是LSTM的变体, 和LSTM不同, GRU不引入额外的记忆单元, 是一种比LSTM网络更加简单的循环神经网络。



三、循环神经网络改进算法

GRU



tanh \tanh 激活函数

σ σ 为 sigmoid 激活函数

\oplus 向量拼接 **\oplus** 向量和

\odot \odot 表示 Hadamard 积，即逐元素对应相乘

$$h^{(t)} = (1 - z_t) \odot h^{(t-1)} + z_t \odot \tilde{h}^{(t)}$$

更新门

$$z_t = \sigma(W_z \cdot x^{(t)} + U_z \cdot h^{(t-1)} + b_z)$$

$$\tilde{h}^{(t)} = \tanh(W \cdot x^{(t)} + U(r_t \odot h^{(t-1)}) + b)$$

重置门

$$r_t = \sigma(W_r \cdot x^{(t)} + U_r \cdot h^{(t-1)} + b_r)$$

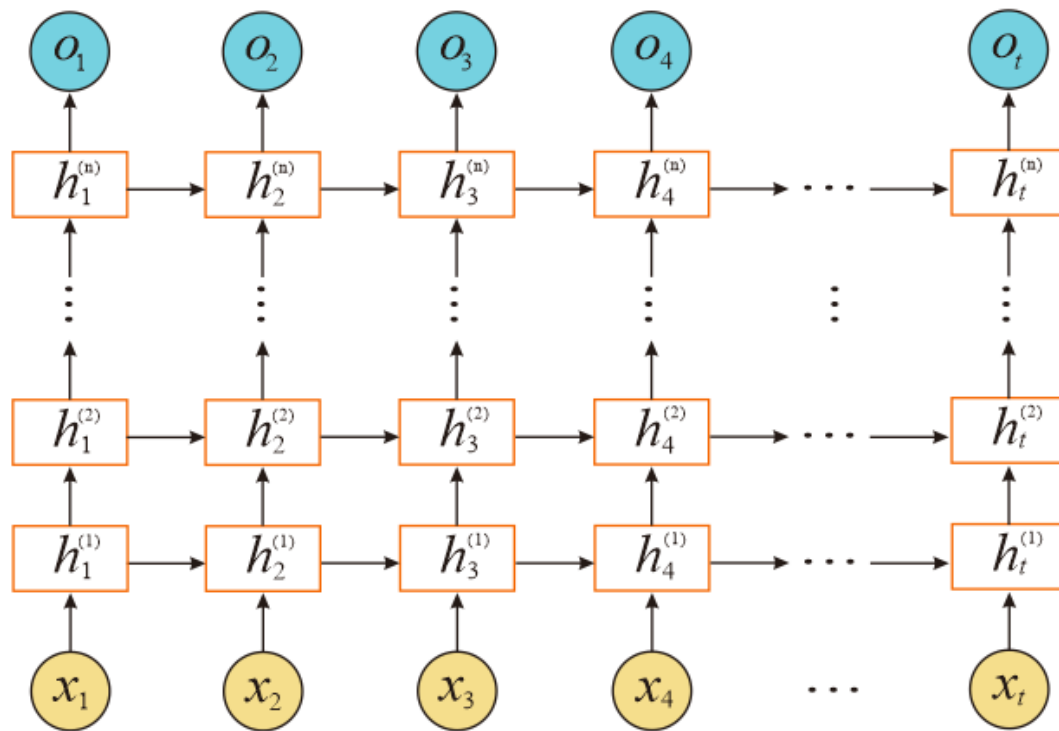
$$h^{(t)} = \text{GRU}(x^{(t)}, h^{(t-1)})$$

三、循环神经网络改进算法



➤ 堆叠循环神经网络 (Stacked Recurrent Neural Network, SRNN)

堆叠循环神经网络 (2015, Joulin&Mikolov) 可以增强循环神经网络的预测能力。一种常见的增加循环神经网络深度的做法是将多个循环网络堆叠起来。



$h_t^{(n)}$ 为在时刻 t 时第 n 层的隐状态:

$$h_t^{(n)} = f \left(V^{(n)} h_{t-1}^{(n)} + W^{(n)} h_t^{(n-1)} + b^{(n)} \right)$$

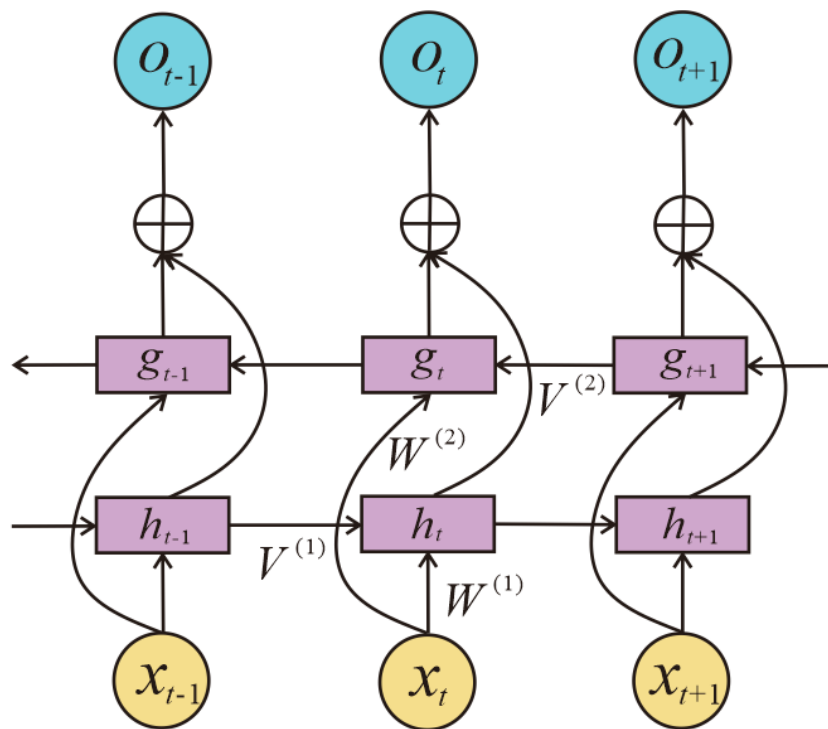
按时间展开的SRNN网络

三、循环神经网络改进算法



➤ 双向循环神经网络 (Bidirectional Recurrent Neural Network, Bi-RNN)

Bi-RNN(1997, M. Schuster&K. Paliwal)由两层循环神经网络组成，它们的输入相同，只是信息传递的方向不同。



双向循环神经网络

顺序:

$$h_t = f(V^{(1)}h_{t-1}^{(1)} + W^{(1)}x_t + b^{(1)})$$

逆序:

$$g_t = f(V^{(2)}g_{t+1}^{(2)} + W^{(2)}x_t + b^{(2)})$$

隐藏层输出:

$$h_t \oplus g_t \quad \oplus \text{ 为向量拼接操作}$$

二、改进循环神经网络



➤ 循环神经网络优缺点

优点：

面对对时间序列敏感的问题赫和任务，RNN(如LSTM)通常会比较合适。
RNN用于序列数据，并且有了一定的记忆效应；

缺点：

- (1) 并行处理上存在劣势。与一些最新的网络相对效果一般；
- (2) RNN的梯度问题在LSTM及其变种里面得到了一定程度的解决，但还是不够。它可以处理100个量级的序列，而对于1000个量级，或者更长的序列则依然会显得很棘手；
- (3) 计算费时。



本节课结束！
谢谢！