第七讲
循环神经网络
Lecture 7
Recurrent Neural Networks

明玉瑞 Yurui Ming

yrming@gmail.com

声明
Disclaimer

- 本讲义在准备过程中由于时间所限，所用材料来源并未规范标示引用来源。所引材料仅用于教学所用，作者无意侵犯原著者之知识版权，所引材料之知识版权均归原著者所有；若原著者介意之，请联系作者更正及删除。

  The time limit during the preparation of these slides incurs the situation that not all the sources of the used materials (texts or images) are properly referenced or clearly manifested. However, all materials in these slides are solely for teaching and the author is with no intention to infringe the copyright bestowed on the original authors or manufacturers. All credits go to corresponding IP holders. Please address the author for any concern for remedy including deletion.

开环控制与闭环控制
Open Control and Closed-loop Control

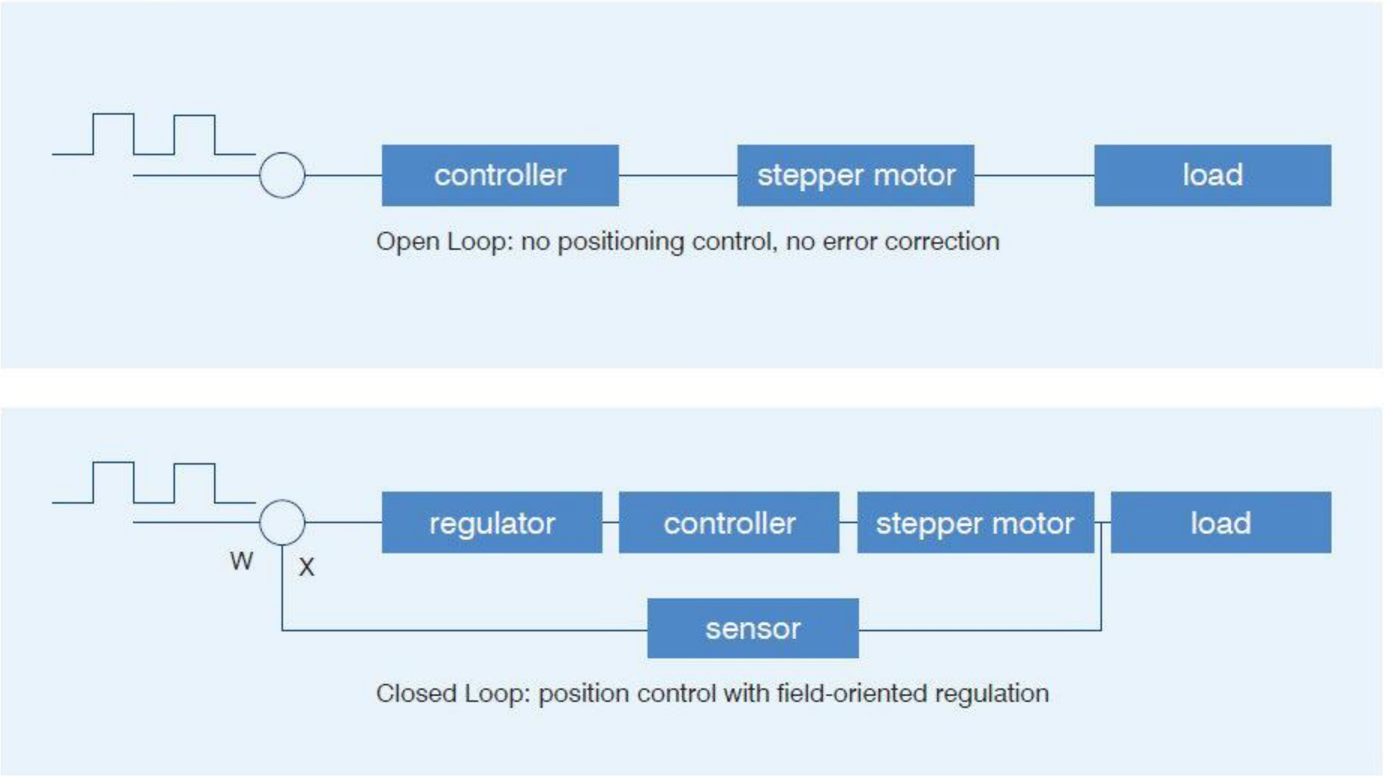
- 考虑一个例子，在运用步进电机的一些应用中，加有载荷与无载荷时，电机对电源的功率输出要求不一样。当载荷或电源功率变化时，步进电机的步进步数可能会有丢失。比如考虑传送过载的情况下，在开环控制情况下，可能载荷并不能达到指定位置。而在闭环控制下，当检测系统检测到载荷不能到位时，会继续指示控制器进行功率输出来驱动步进电机，以便载荷达到预定位置。

  Consider the example for controlling step motor in some applications. The load will incur different torques of the step motor which consequently requires the inconsistent output of power source. The alteration of power or load might lead to the loss of steps. For example, the overload impacting on motor especially under open control can incur the failure of proper delivery. However, under closed-loop control, when the sensor detects the improper delivery, it will continue to instruct the controller for power output to drive the motor, in order to fulfill the delivery.

开环控制与闭环控制
Open Control and Closed-loop Control

- 此种闭环控制，即是一种循环结构

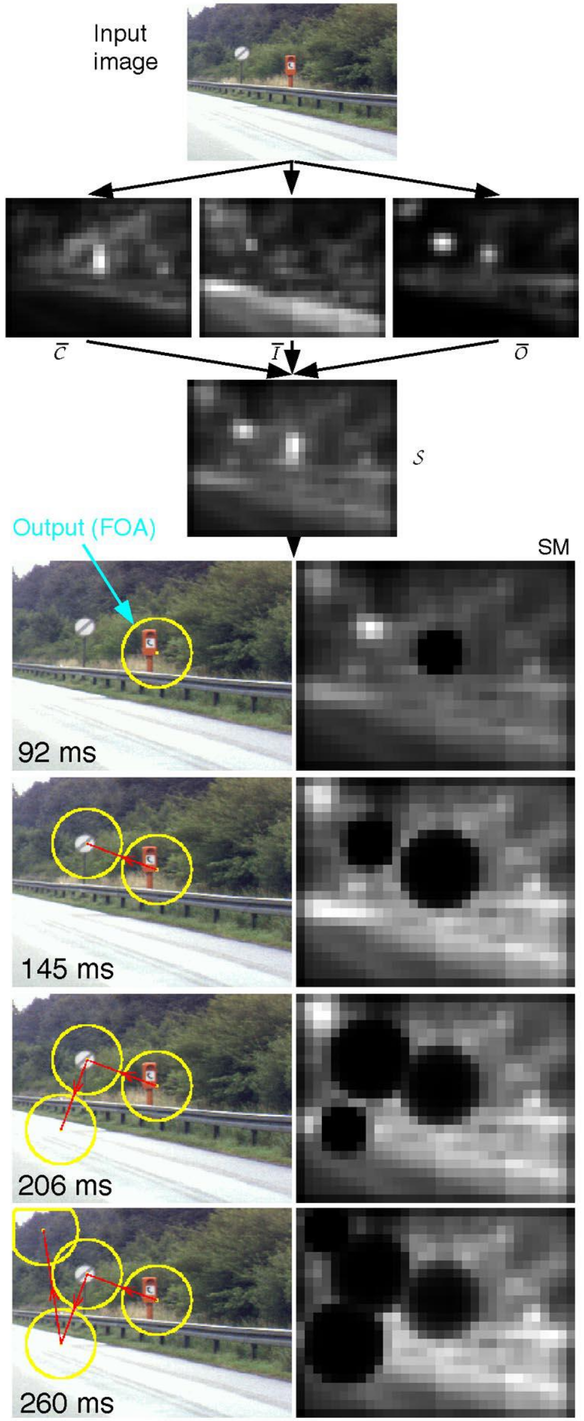
  The closed-loop control illustrates one recurrent structure.



注意力
Attention

- 当人们观察一幅图片时，注意力并非平均分配，而是在会注重关注图片里的某个物体；从表面上看，注意力集中于某个物体是瞬时完成的，但实际是通过不同脑区的交互，特别是反馈过程渐进完成的。

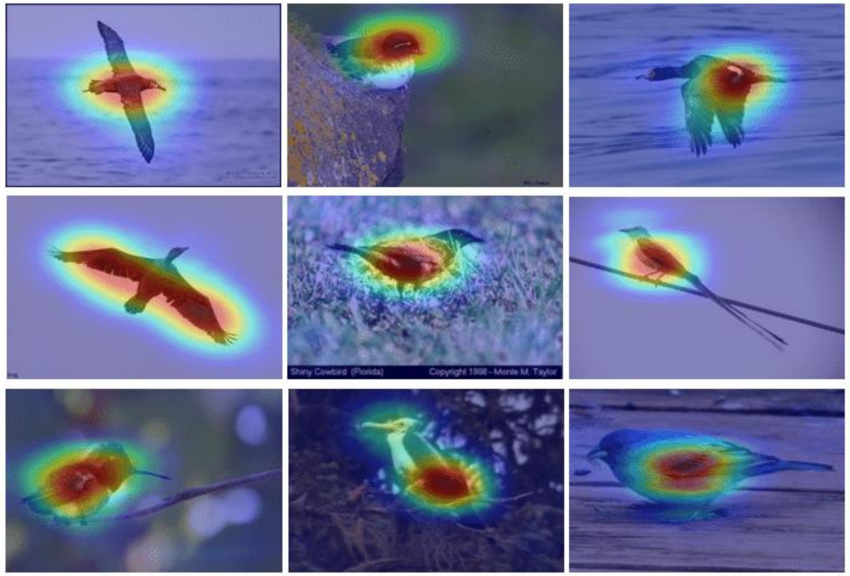
  When people observe a certain image, the attention is not scattering over the whole image, instead, just focusing on some objects in the image. Such a process seems to be superficially attained instantly; however, it is the consequence that different brain regions interacted with each other to achieve this in a procedural way.

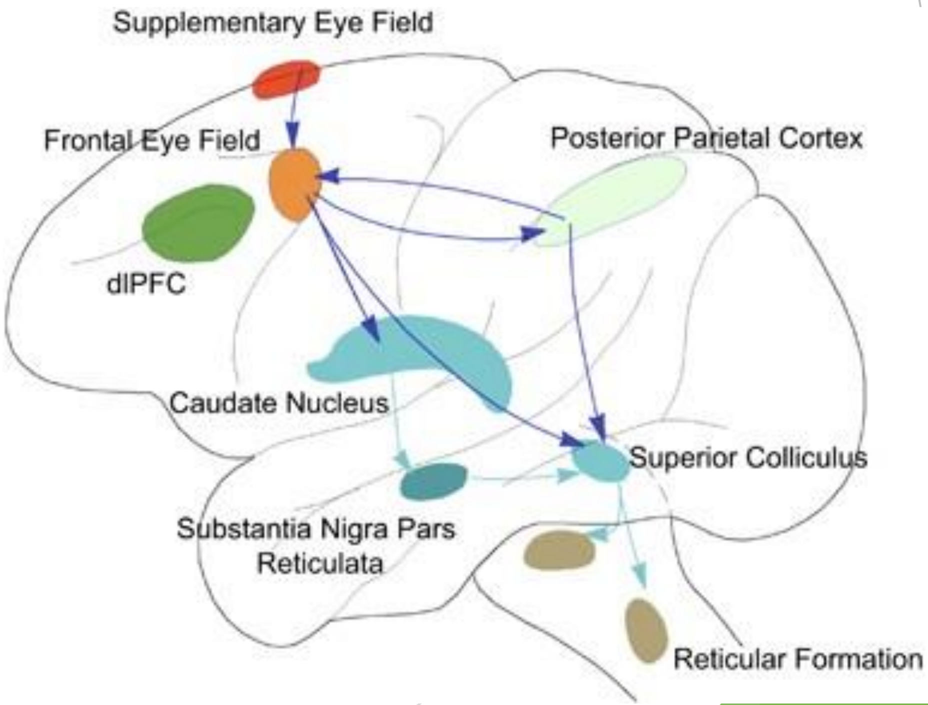


注意力
Attention

- 在定位显著物体的过程中，脑区连接的循环结构是显而易见的。

  The saccades for salient objects obviously involves recurrent structure of brain connectivity.

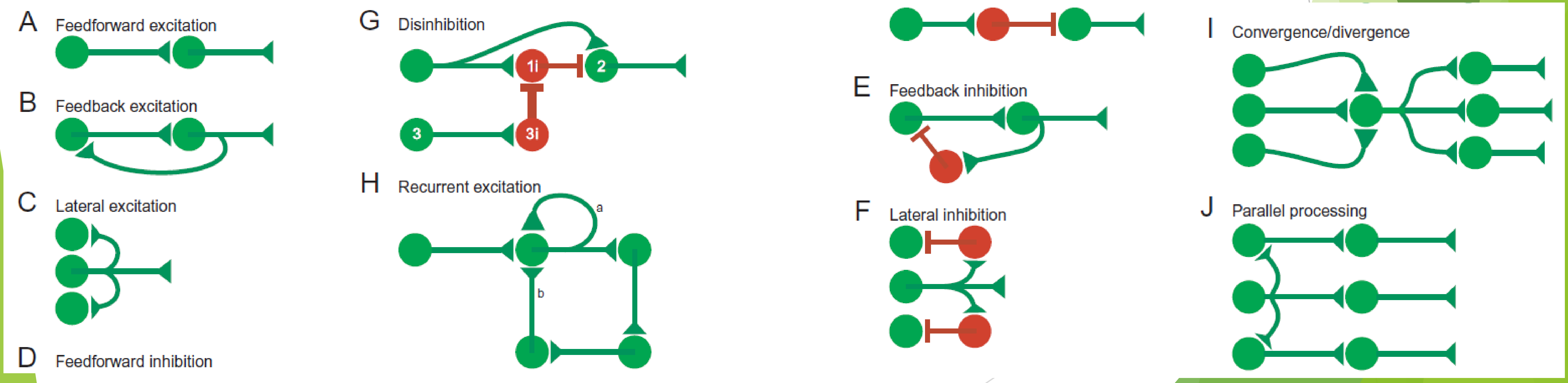




神经元的反馈结构
Feedback Circuits of Neurons

- 在神经元的种类丰富的连接中，反馈结构是非常常见的，特别对于抑制效应。

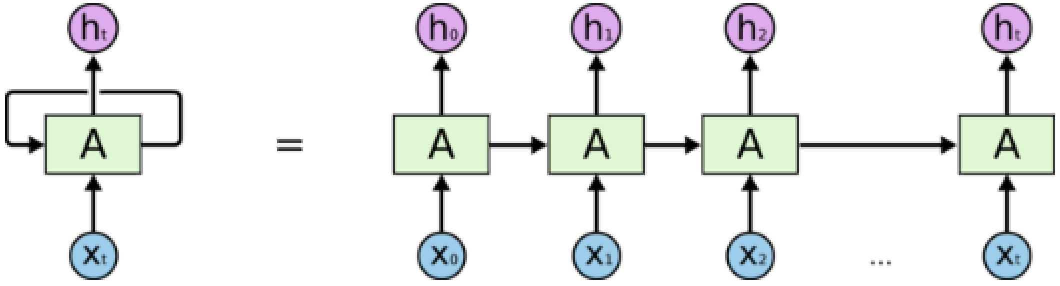
  For the variety of neuronal circuitries, the feedback is almost ubiquitous, especially for inhibition.



循环结构
Recurrent Structure

- 在前面的例子中，讲到闭环控制，循环交互，反馈回路，基本上讲的是同一个事物，即循环结构，只是在不同的学科里有不同的术语与使用套路。在人工神经网络学科，统称为有此类结构的网络称为循环神经网络。

  Among the previous examples, the referred closed-loop control, recurrent interaction, feedback circuits, they all talk about the same thing, aka, the recurrent structure. However, just different disciplines adopt different terminologies and tricks of application.



An unrolled recurrent neural network.

循环神经网络
Recurrent Neural Network

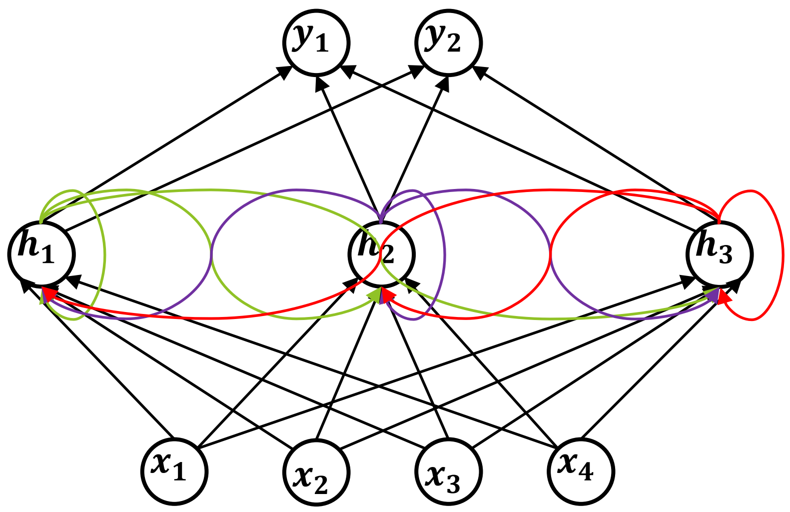
- 为什么需要循环神经网络？用机器学习方法所要解决的问题，比如语言翻译，视频理解，需要处理的都是序列数据（词的序列或图片的序列）。不引起歧义的情况下，我们称为时序数据。通常，组成这些时序的数据点是具有相关性的，引入循环神经网络，可为发掘这种关联性提供更为有效的模型。

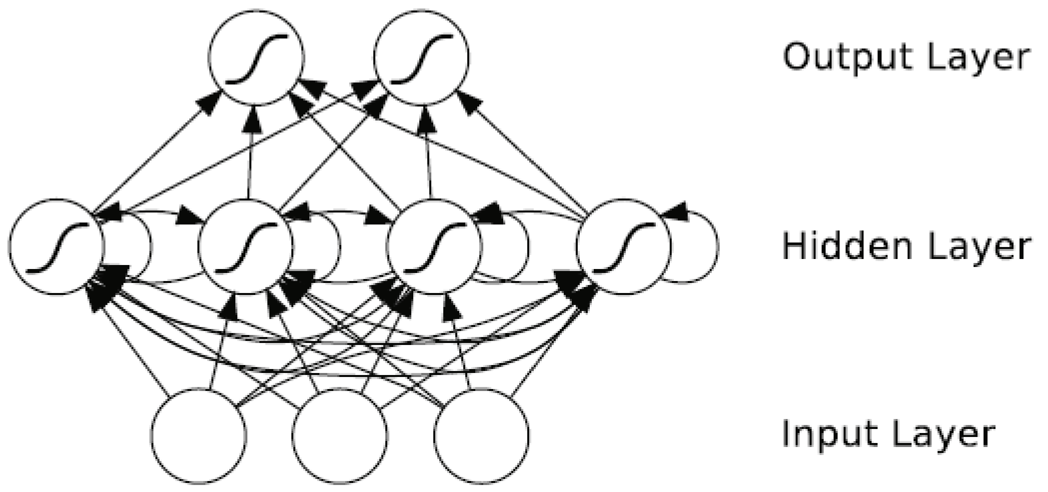
  The motivation for introducing recurrent neural network originates from the requirement to solve certain problems using machine learning method. Considering language translation, video understanding, etc., here the data to be processed are sequential data (data sequence or image sequence). We call these data time series data if no ambiguity arises. Generally, the data points which constituted the whole sample are of some correlations. By introducing recurrent structure, it empowers the model to harvest such correlations.

循环神经网络
Recurrent Neural Network

- 下图是神经网络的简单示例，对于隐含层而言，不仅接收来自输入层节点的输入，并且，每个隐含层节点的输出都有反馈至包括自身在内的其它隐含层节点的输入端。

  The following figure illustrates a simple recurrent neural network. For the hidden layer, each node not only receives the inputs from input layer, but also feed the output back to the input of each node of the same hidden layer, including itself.





前向传播
Feed-forward Propagation

- 符号假定（denotation）
  - I: 输入层节点数（number of input units）
  - H: 隐含层节点数（number of hidden units）
- 前向传播（Forward propagation）

$$z_h^t = \sum_{i=1}^{I} w_{ih} x_i^t + \sum_{j=1}^{H} v_{jh} y_j^{t-1},\ y_h^t = f(z_h^t)$$

z_h^t: the intermediate output of the h-th unit in hidden layer at time t

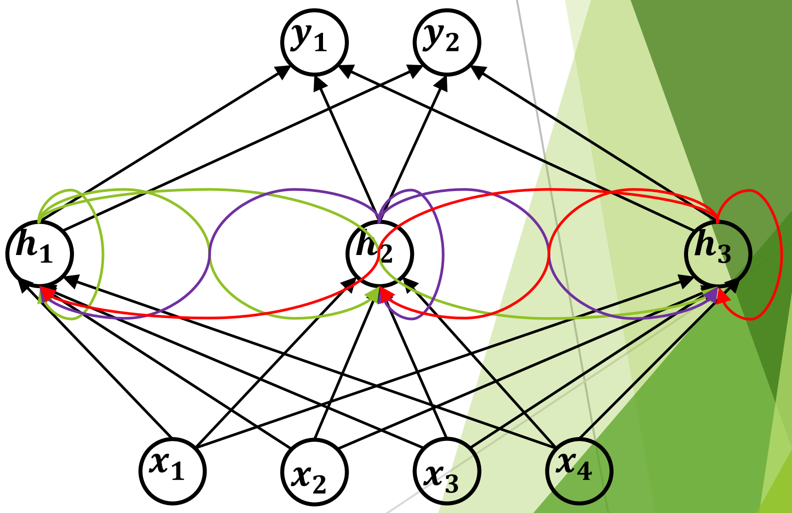
y_h^{t-1} 或 y_h^t: the activation of the h-th unit in hidden layer at time $t-1$ and t

w_{ih}: connection from input unit i to hidden unit h

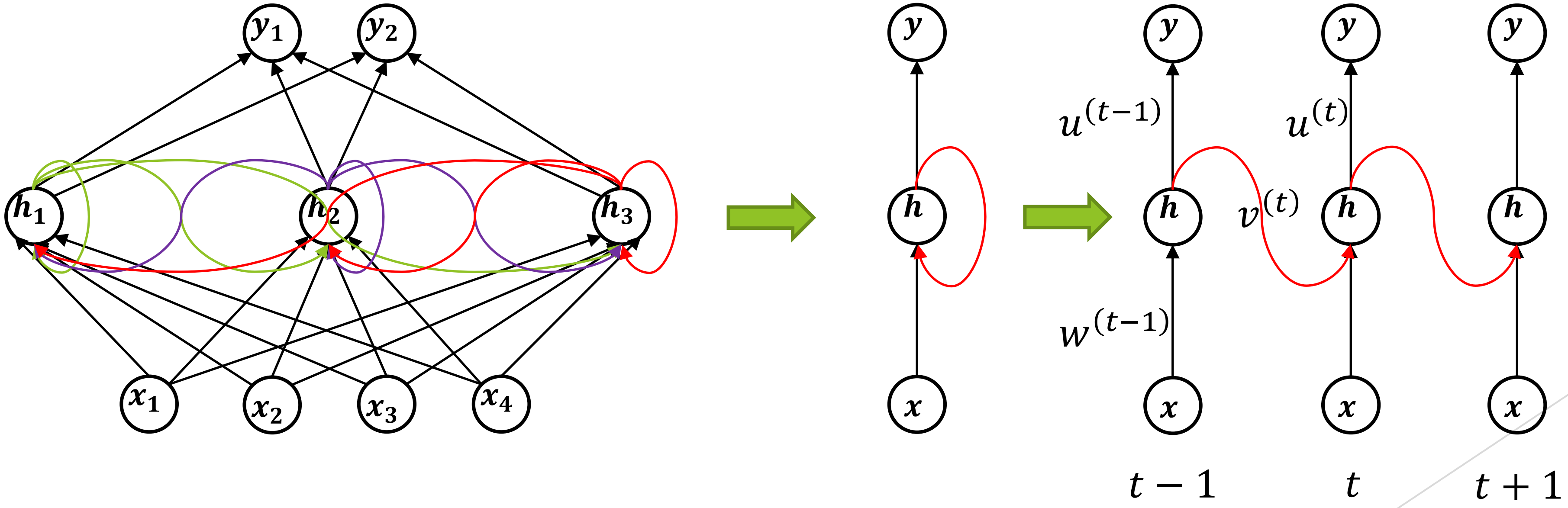
v_{jh}: connection from hidden unit j to hidden unit h

前向传播
Feed-forward Propagation

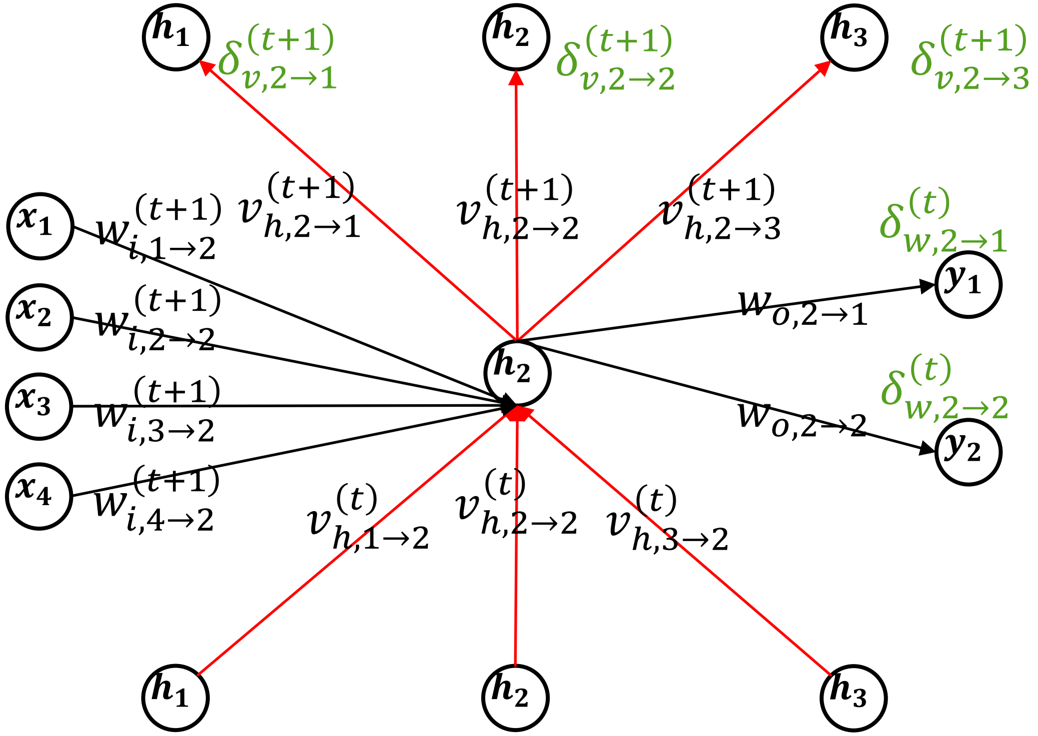
- $x^{(t)} = \begin{bmatrix} x_1^{(t)} & x_2^{(t)} & x_3^{(t)} & x_4^{(t)} \end{bmatrix} = \begin{bmatrix} 0.4 & 1.2 & -1.0 & 0.6 \end{bmatrix}$,
- $w_i^{(t)} = \begin{bmatrix} w_{1\to h}^{(t)} \\ w_{2\to h}^{(t)} \\ w_{3\to h}^{(t)} \\ w_{4\to h}^{(t)} \end{bmatrix} = \begin{bmatrix} 0.8 \\ 1.6 \\ -0.2 \\ -2.4 \end{bmatrix}$, $w_h^{(t)} = \begin{bmatrix} v_{1\to h}^{(t)} \\ v_{2\to h}^{(t)} \\ v_{3\to h}^{(t)} \end{bmatrix} = \begin{bmatrix} 1.8 \\ -1.2 \\ 0.4 \end{bmatrix}$,
- $y^{(t-1)} = \begin{bmatrix} y_1^{(t-1)} & y_2^{(t-1)} & y_3^{(t-1)} \end{bmatrix} = \begin{bmatrix} 2.0 & 0.2 & -1.2 \end{bmatrix}$
- $z^{(t)} = \begin{bmatrix} 0.4 & 1.2 & -1.0 & 0.6 \end{bmatrix} \begin{bmatrix} 0.8 \\ 1.6 \\ -0.2 \\ -2.4 \end{bmatrix} + \begin{bmatrix} 2.0 & 0.2 & -1.2 \end{bmatrix} \begin{bmatrix} 1.8 \\ -1.2 \\ 0.4 \end{bmatrix}$
- $y^{(t)} = \sigma\left(z^{(t)}\right)$

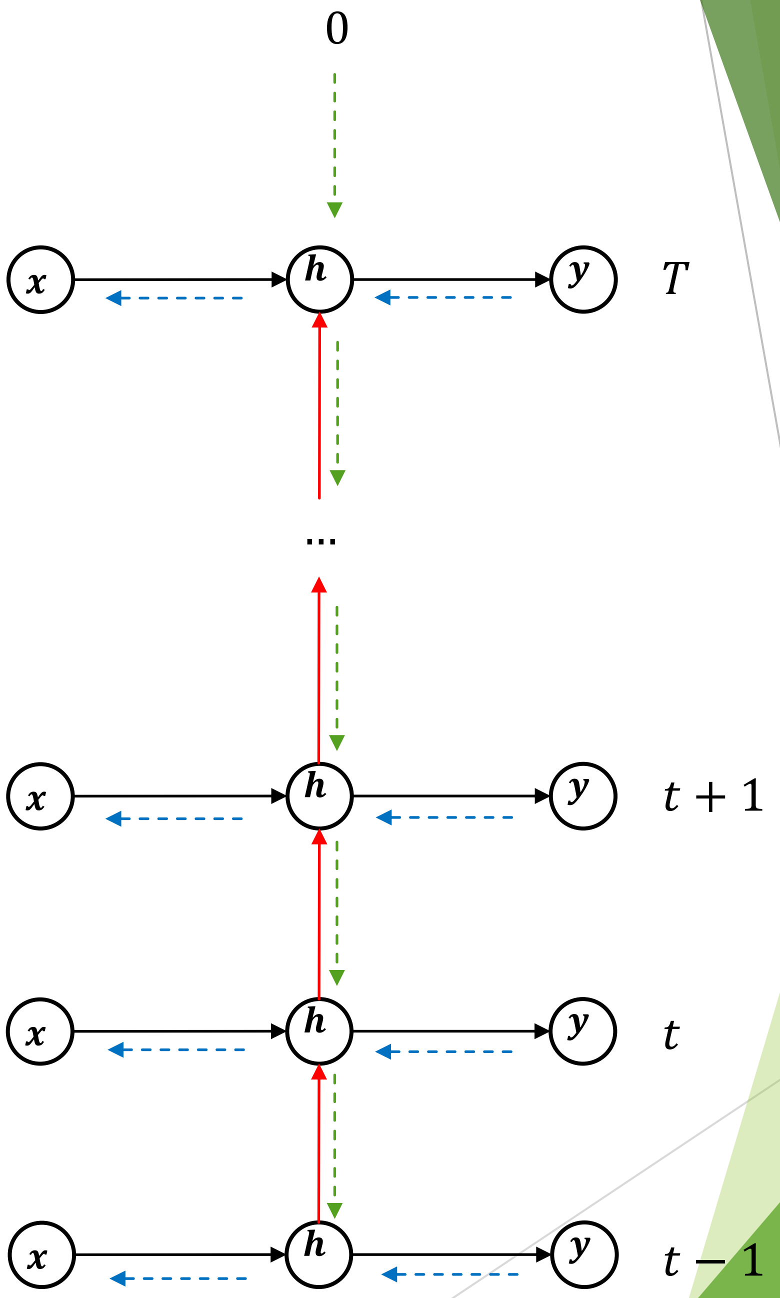


反向传播
Backward Propagation



反向传播
Backward Propagation





反向传播
Backward Propagation

- $\delta_{v,o}^{(t)} = \delta_{v,2\to1}^{(t+1)} v_{h,2\to1}^{(t+1)} + \delta_{v,2\to2}^{(t+1)} v_{h,2\to2}^{(t+1)} + \delta_{v,2\to3}^{(t+1)} v_{h,2\to3}^{(t+1)}$
- 下标中出现的o表示在h_2节点的输出端； i表示在h_2节点的输入端；

  For the subscript, o indicates the output, whilst i indicates the input.
- $\Delta w_{o,2\to1} = \sum_{t=1}^{T} \delta_{w,2\to1}^{(t)} y_2^{(t)}$， $\Delta w_{o,2\to2} = \sum_{t=1}^{T} \delta_{w,2\to2}^{(t)} y_2^{(t)}$

- $\delta_{w,o}^{(t)} = \delta_{w,2\to1}^{(t)} w_{h,2\to1}^{(t)} + \delta_{w,2\to2}^{(t)} w_{h,2\to2}^{(t)}$
- $\delta_{h,o}^{(t)} = \delta_{v,o}^{(t)} + \delta_{w,o}^{(t)}$
- $\delta_{h,i}^{(t)} = \delta_{h,o}^{(t)} \sigma'\left(y^{(t)}\right) z^{(t)}$

反向传播
Backward Propagation

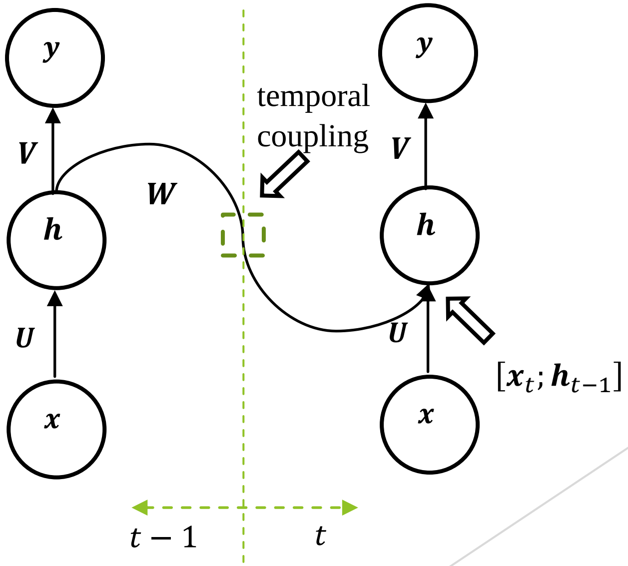
- $\Delta w_{i,1\to 2} = \sum_{t=1}^{T} \delta_{h,i}^{(t)} x_1^{(t)}$， $\Delta w_{i,2\to 2} = \sum_{t=1}^{T} \delta_{h,i}^{(t)} x_2^{(t)}$
- $\Delta w_{i,3\to 2} = \sum_{t=1}^{T} \delta_{h,i}^{(t)} x_3^{(t)}$， $\Delta w_{i,4\to 2} = \sum_{t=1}^{T} \delta_{h,i}^{(t)} x_4^{(t)}$

- $\Delta v_{h,1\to 2} = \sum_{t=1}^{T+1} \delta_{h,i}^{(t)} h_1^{(t-1)}$， $\delta_{h,i}^{(T+1)} = 0$， $h_1^{(0)} = 0$
- $\Delta v_{h,2\to 2} = \sum_{t=1}^{T+1} \delta_{h,i}^{(t)} h_2^{(t-1)}$， $\delta_{h,i}^{(T+1)} = 0$， $h_2^{(0)} = 0$
- $\Delta v_{h,3\to 2} = \sum_{t=1}^{T+1} \delta_{h,i}^{(t)} h_3^{(t-1)}$， $\delta_{h,i}^{(T+1)} = 0$， $h_3^{(0)} = 0$

长短期记忆网络
Long Short-Term Memory Network

- 普通的神经网络在处理时序数据的数据点时，会假设相邻数据点之间的强相关，这是由于同一隐含层节点之间的直接反馈连接造成的，此点在循环网络于时间轴上展开时，会更加明显：

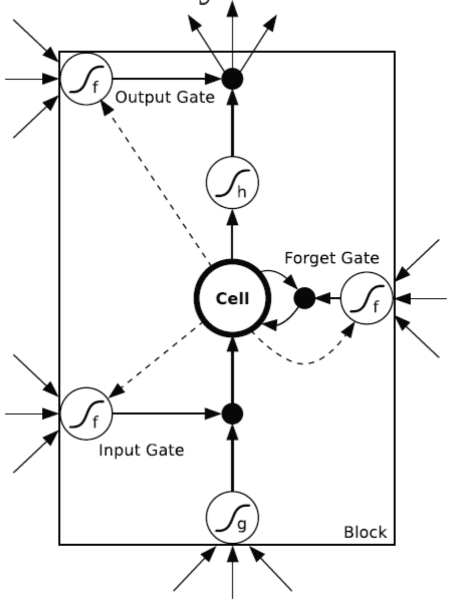
  The vanilla recurrent neural network is assuming the strong correlation of adjacent data points which constitute the time series data sample. This is caused by the direct link among hidden nodes in the same layer, and it is more manifest if the recurrent neural network is unfolded along the time axis.



长短期记忆网络
Long Short-Term Memory Network

- 为了克服这个缺点，有学者引入了长短期记忆网络。其想法是根据当前数据与历史所学，通过各种门的调制，来缓和这种强耦合。

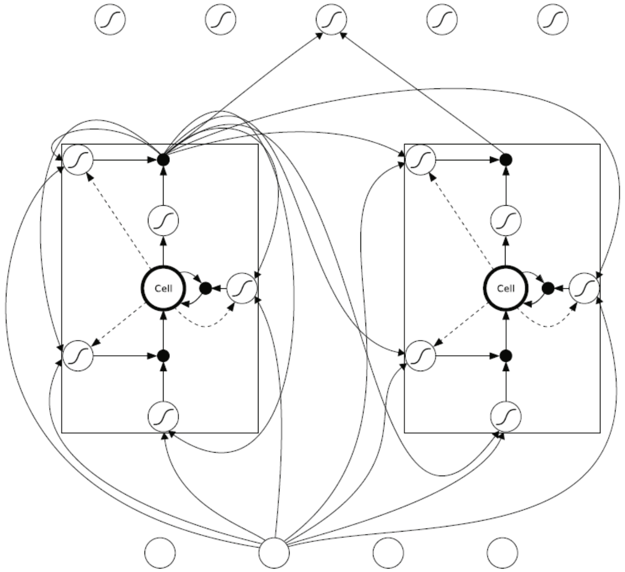
  To overcome such a shortcoming, some scholars introduced Long Short-Term Memory Network. The idea is to introduce various gates to mitigate such strong coupling based on the current input data and information from history data.



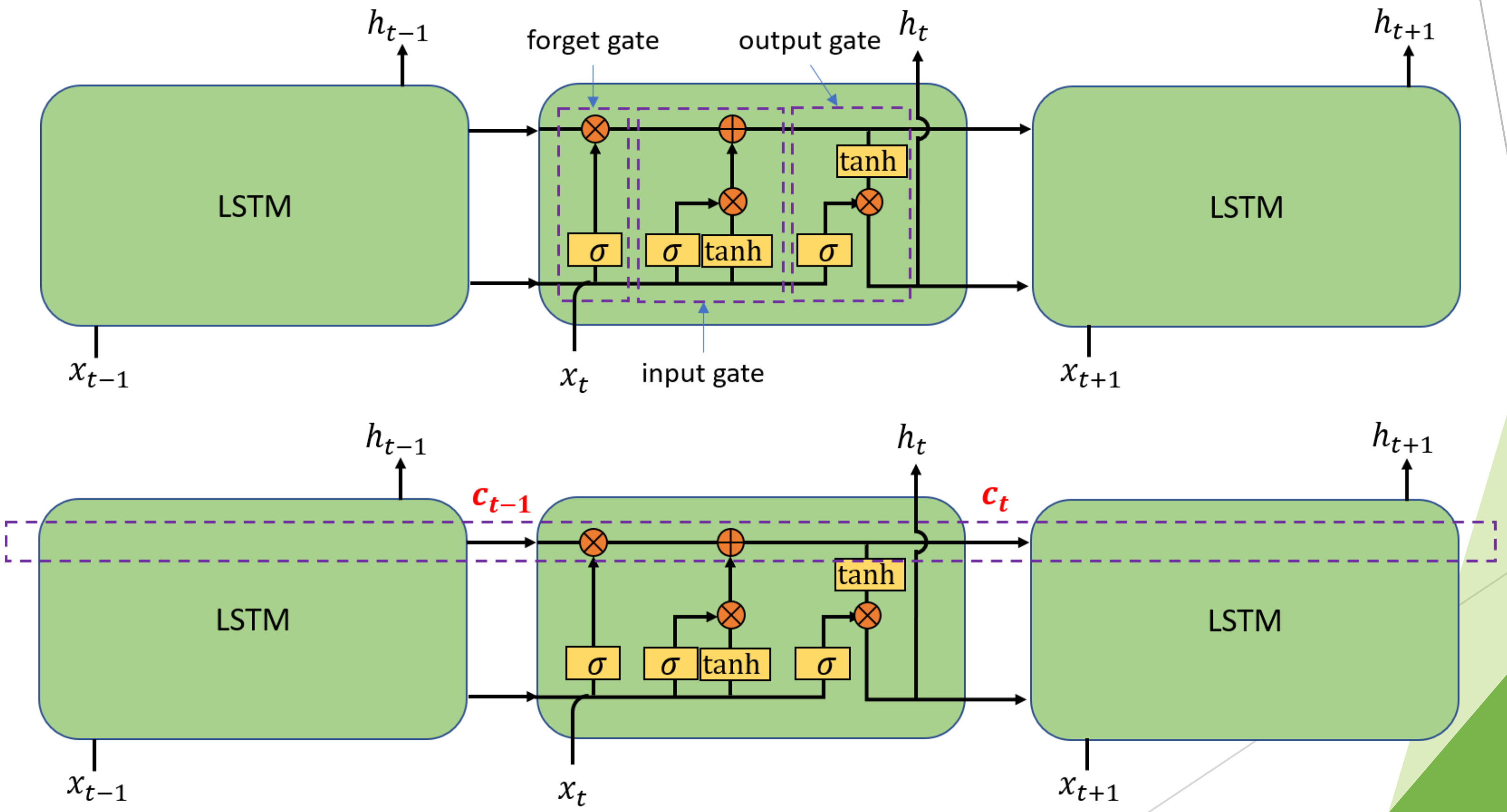
长短期记忆网络
Long Short-Term Memory Network

- 下图展示了一个简单的LSTM网络，其有四个输入端，由单个单元组成的两个LSTM模块，及5个输出节点组成。注意有的连线没有画出，每个LSTM模块有四个输入，一个输出。

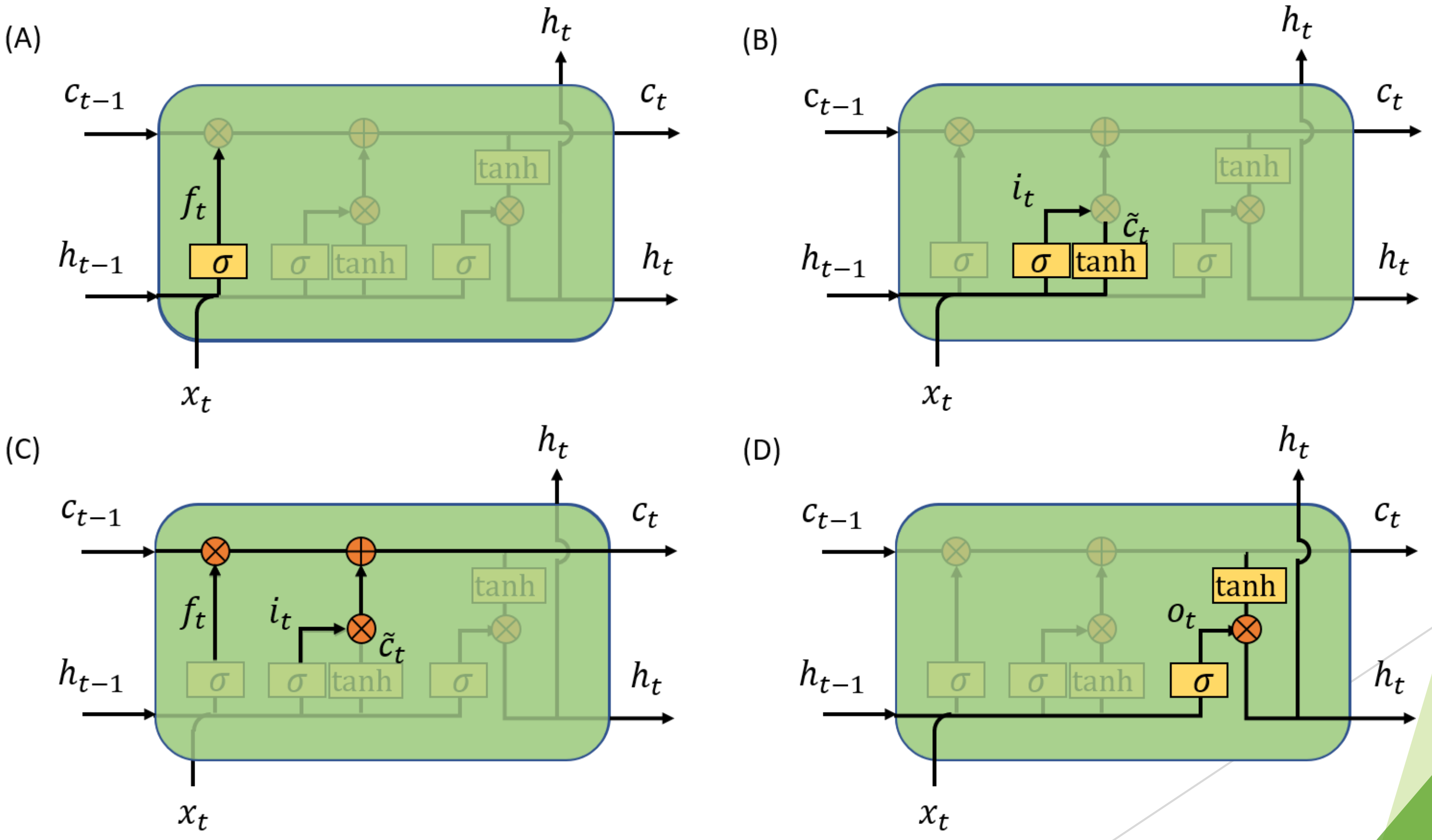
  The following diagram illustrate an LSTM network. The network consists of four input units, a hidden layer of two single-cell LSTM memory blocks and five output units. Not all connections are shown. Note that each block has four inputs but only one output.



LSTM前向传播
LSTM Feed-forward Propagation



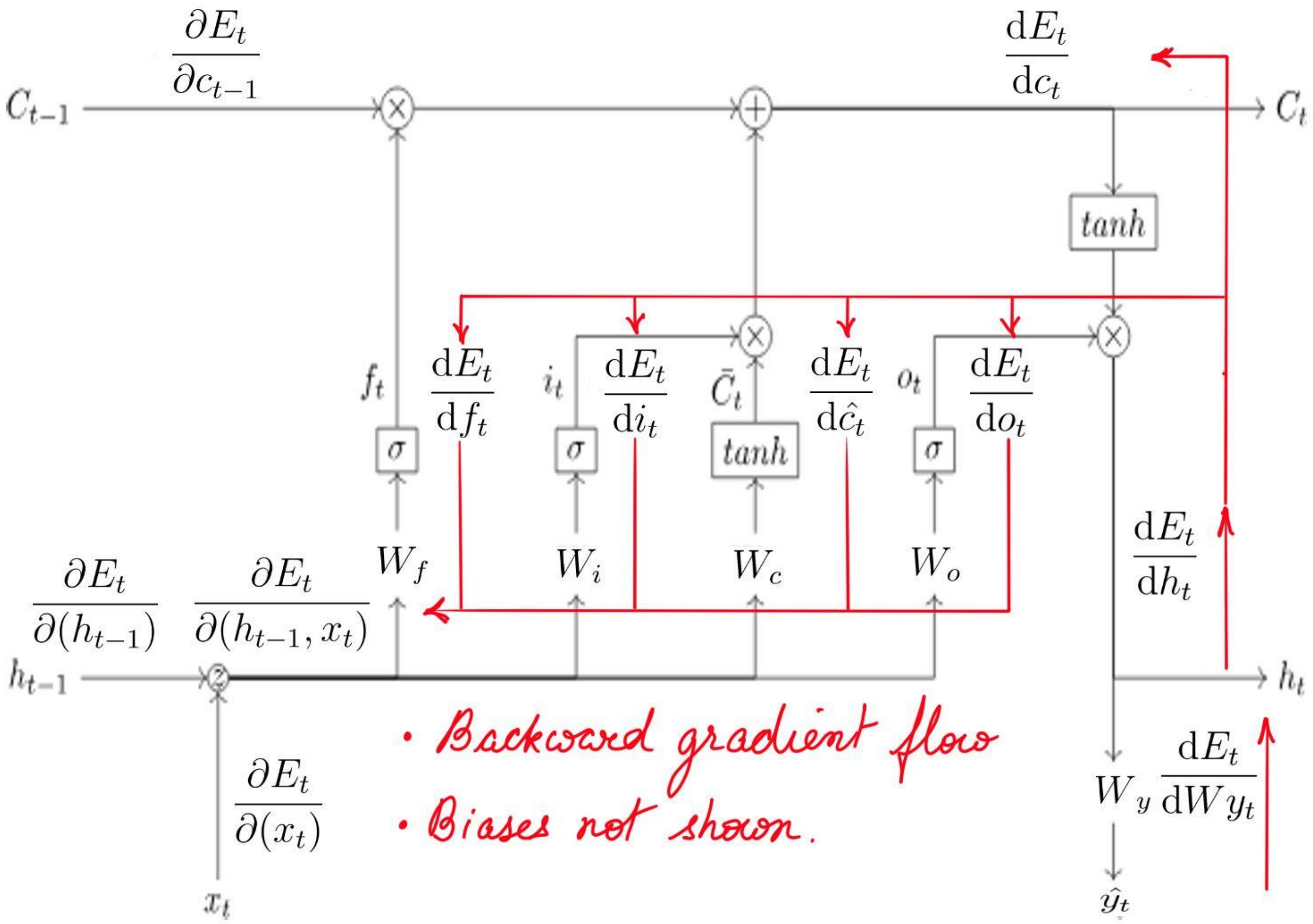
LSTM前向传播
LSTM Feed-forward Propagation



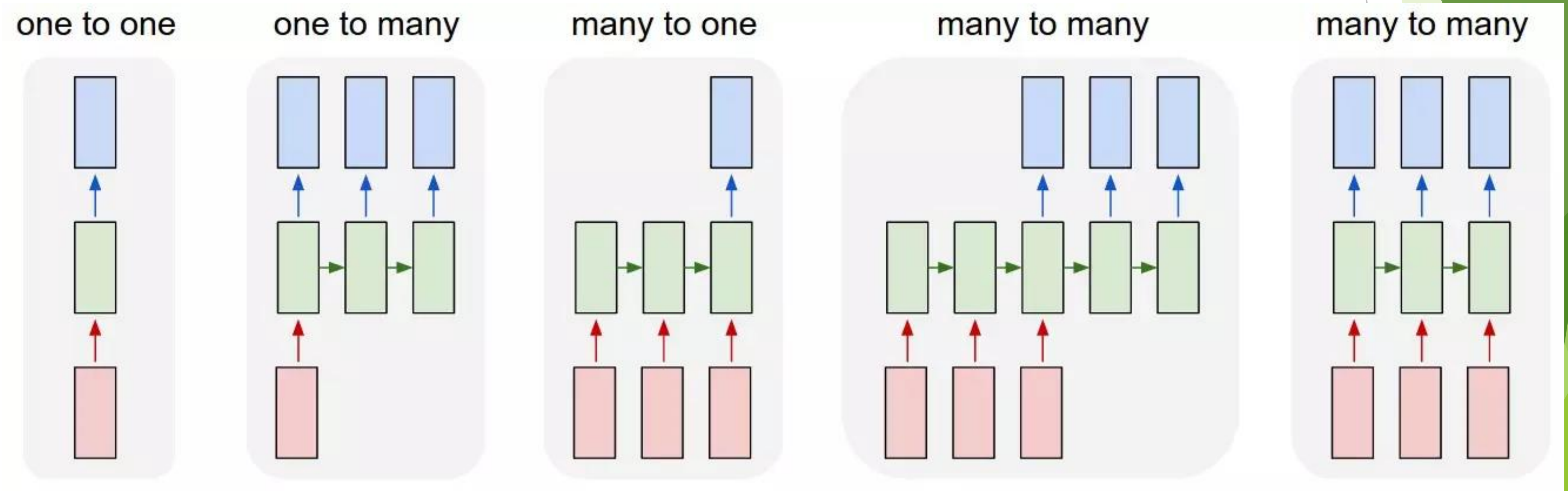
LSTM前向传播
LSTM Feed-forward Propagation

- $f_t = \sigma\left(w_f[h_{t-1}, x_t] + b_f\right)$
- $i_t = \sigma(w_i[h_{t-1}, x_t] + b_i)$
- $\tilde{c}_t = \tanh(w_c[h_{t-1}, x_t] + b_c)$
- $c_t = f_t * c_{t-1} + i_t * \tilde{c}_t$
- $o_t = \sigma(w_o[h_{t-1}, x_t] + b_o)$
- $h_t = o_t * \tanh(c_t)$

LSTM反向传播
LSTM Backward Propagation



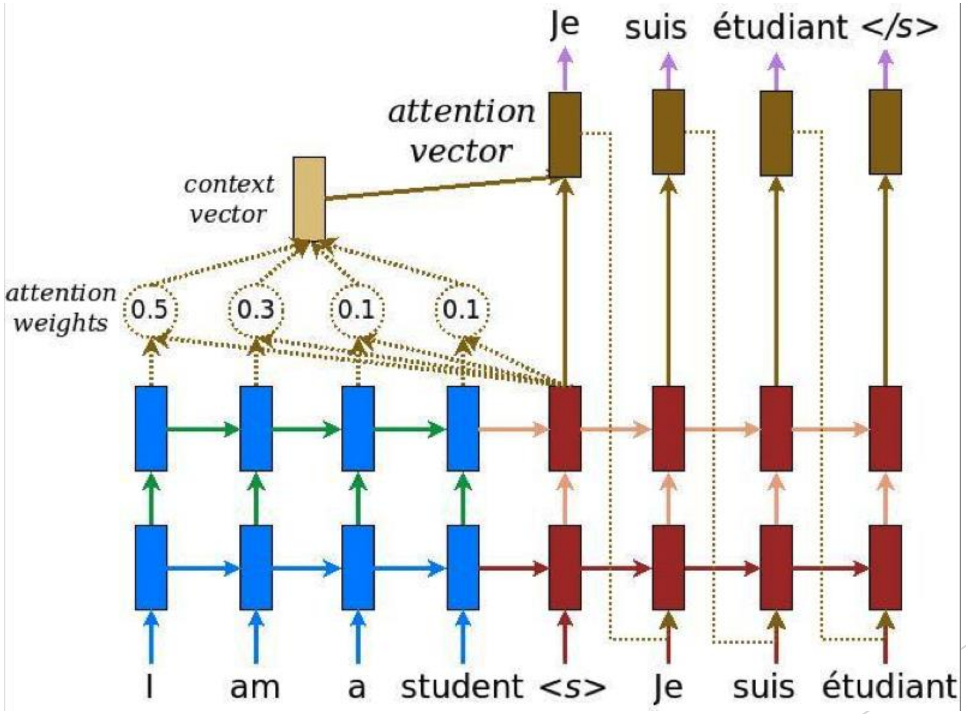
循环网络输入输出关系模式
Paradigm Of Input vs Output for RNN



应用
Application

- 下图展示了可用于不同种类语言翻译的一种神经网络架构.

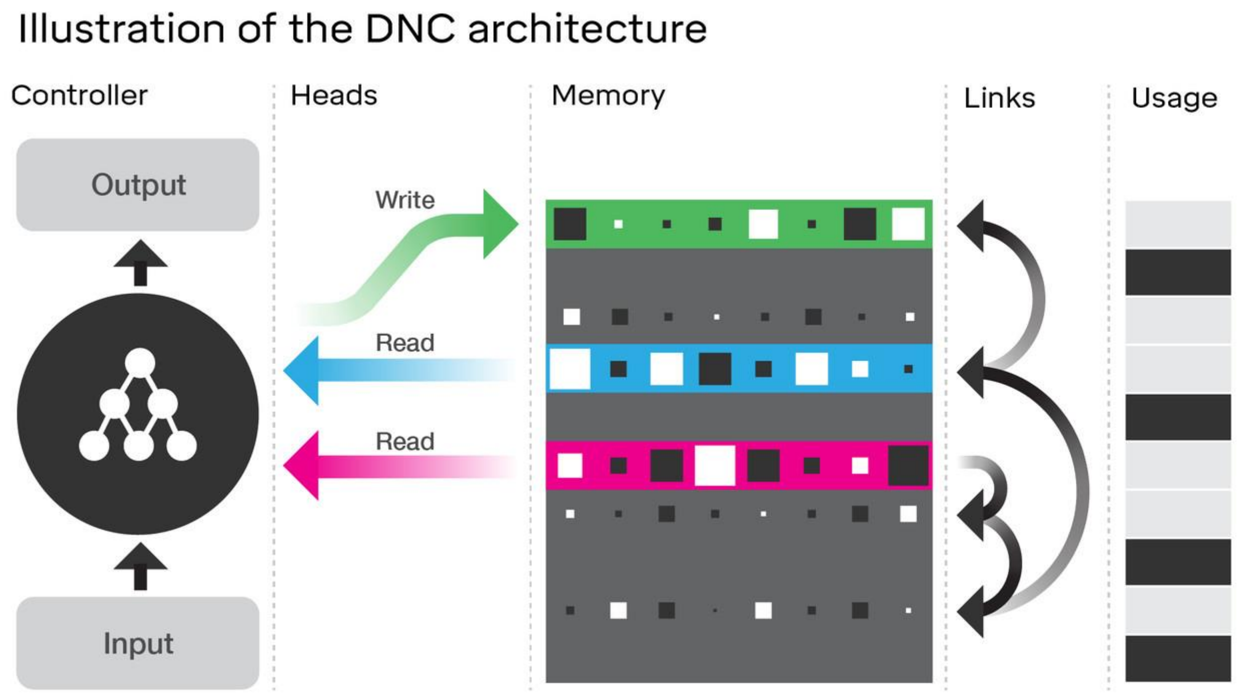
  The following figure illustrates one RNN architecture for language translation.



应用
Application

- 可微神经计算机

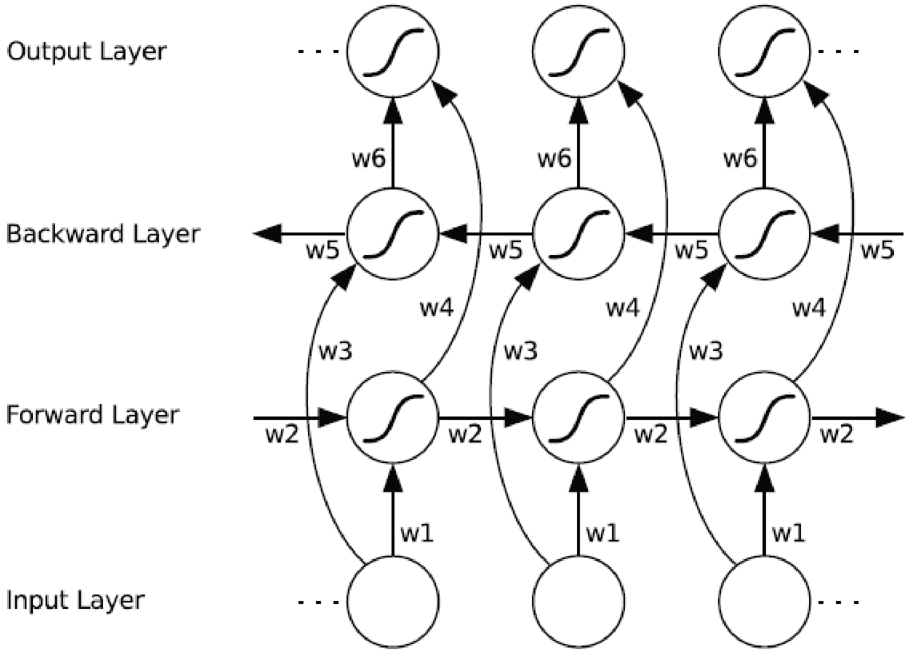
  Differentiable neural computer



习题
Problems

1. 查阅文献，了解什么是双向循环网络，描述其前向传播过程。

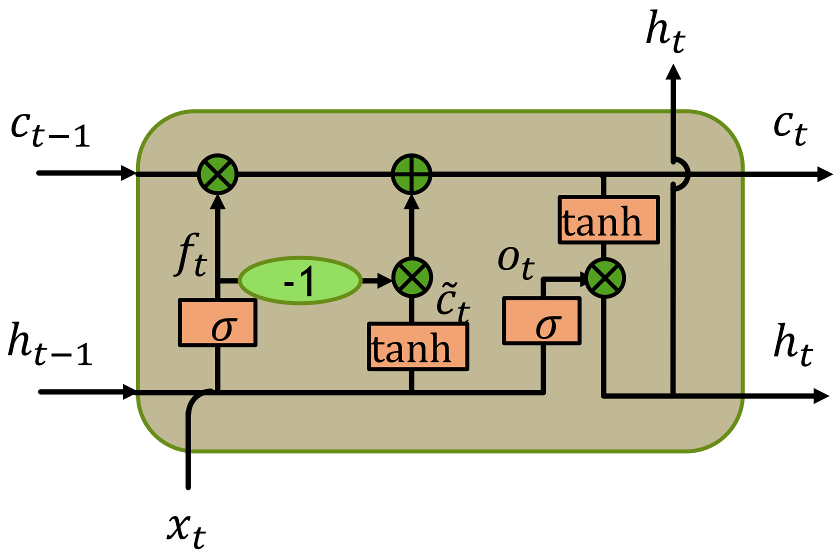
   Explain bi-directional recurrent neural network via literature survey. Describe the process of feedforward propagation.



习题
Problems

2. LSTM的一个变型是通过互补关系将遗忘门与输入门耦合在一起，如下图所示，请推导其前向传播公式。

   One variant of LSTM is to coupling forget gate and input gate together via the complementary relationship, as illustrated below. Deduce the formula of feedforward propagation.



习题
Problems

3. 人类的记忆可分为工作记忆，短期记忆与长期记忆。一般来说，工作记忆看成和日常认知活动有关的一种机制，而不是侧重需要记忆的内容本身。请说明LSTM网络更和工作记忆接近，而非短期或长期记忆。

   The memory of human being can be categorized as working memory, short-term memory and long-term memory. Generally speaking, working memory emphasizes on the mechanism instead of the content to be memorized. Indicate that LSTM network is more analogous to the working memory instead of short- or long-term memory.