

Long/Short Term Memory (LSTM) Model

長短期記憶網路 介紹

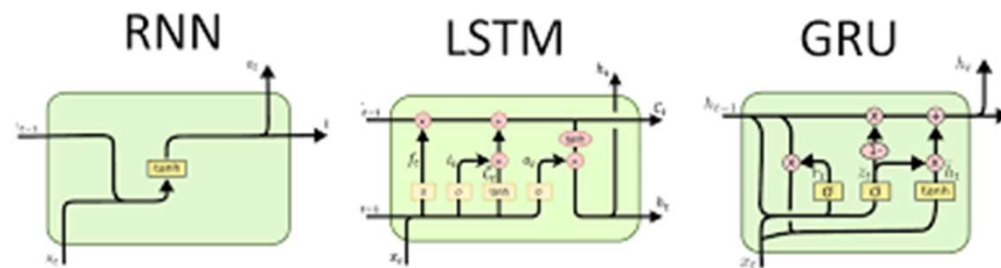


Sean, 王家祥

2020/11/20

CONTENTS

1. LSTM 簡介
2. LSTM 結構
3. LSTM圖解
4. LSTM Demo



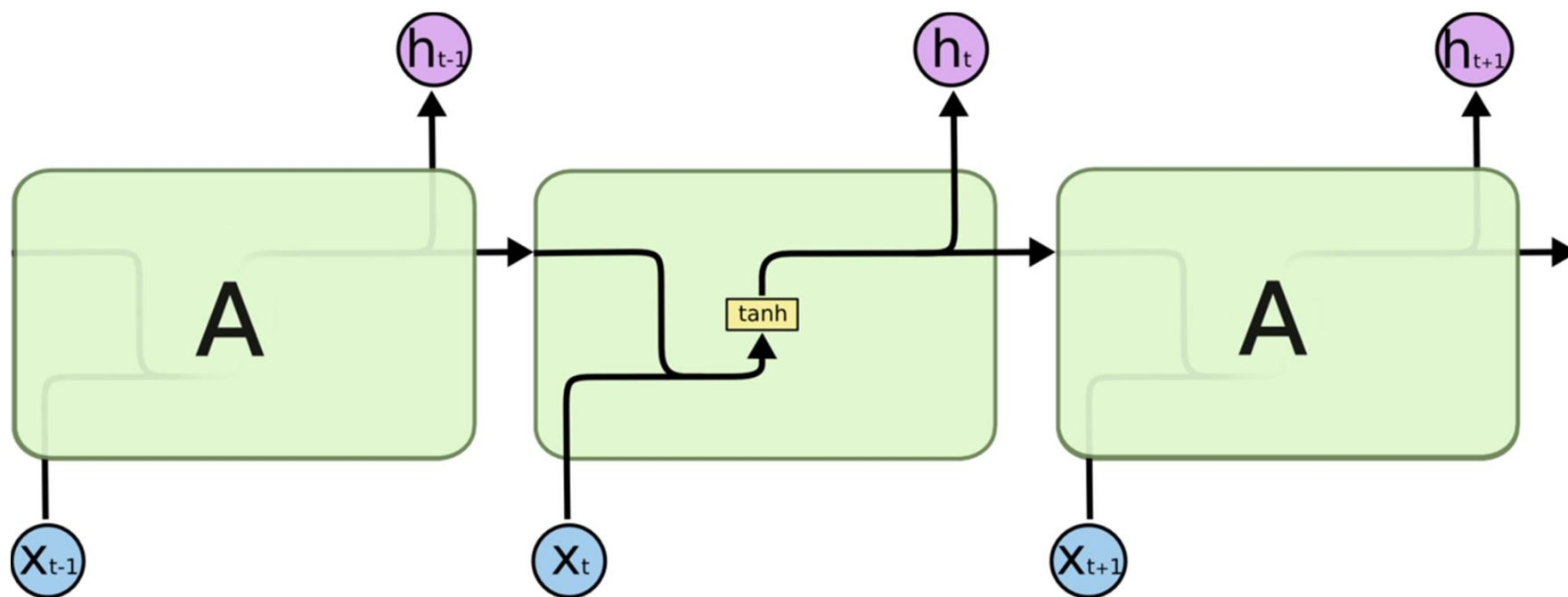


/01

LSTM 簡介

1.LSTM 簡介

- 所有 RNN 都具有一種重覆神經網絡模塊的鏈式的形式。在標準的 RNN 中，這個重覆的模塊只有一個非常簡單的結構，例如一個 \tanh 層，如下圖所示：



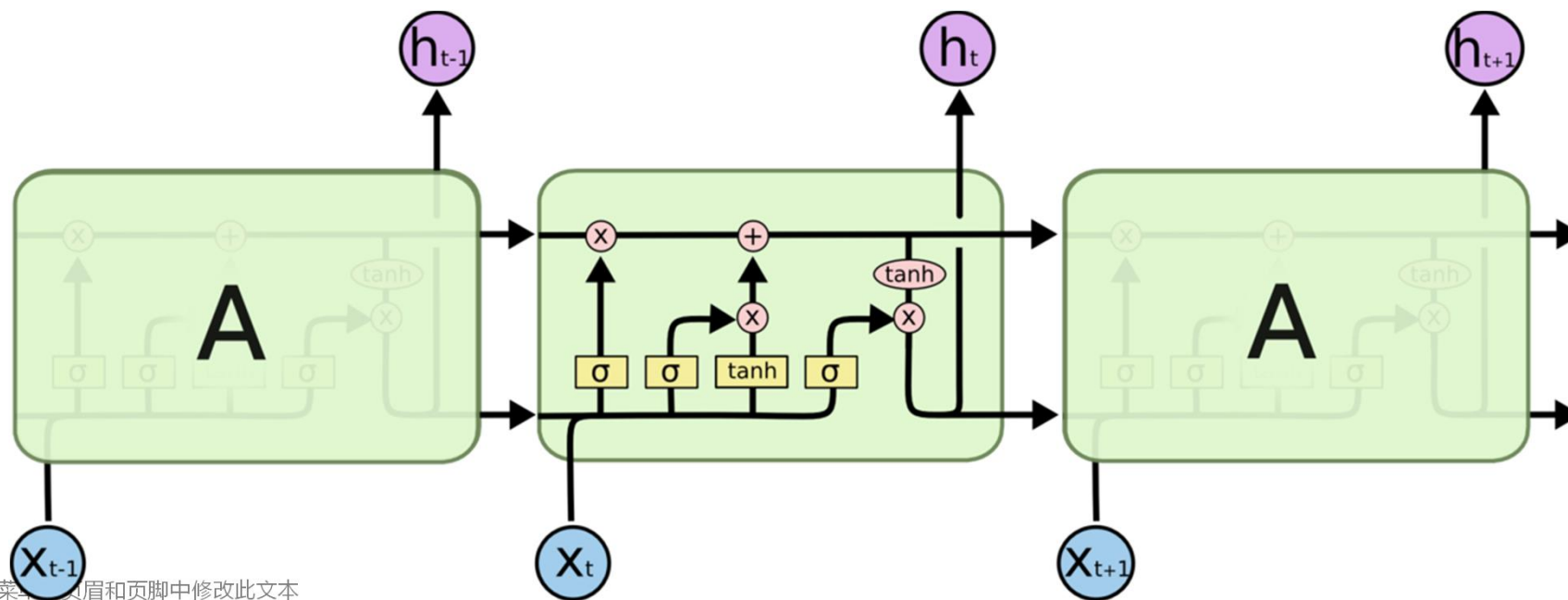
1.LSTM 簡介

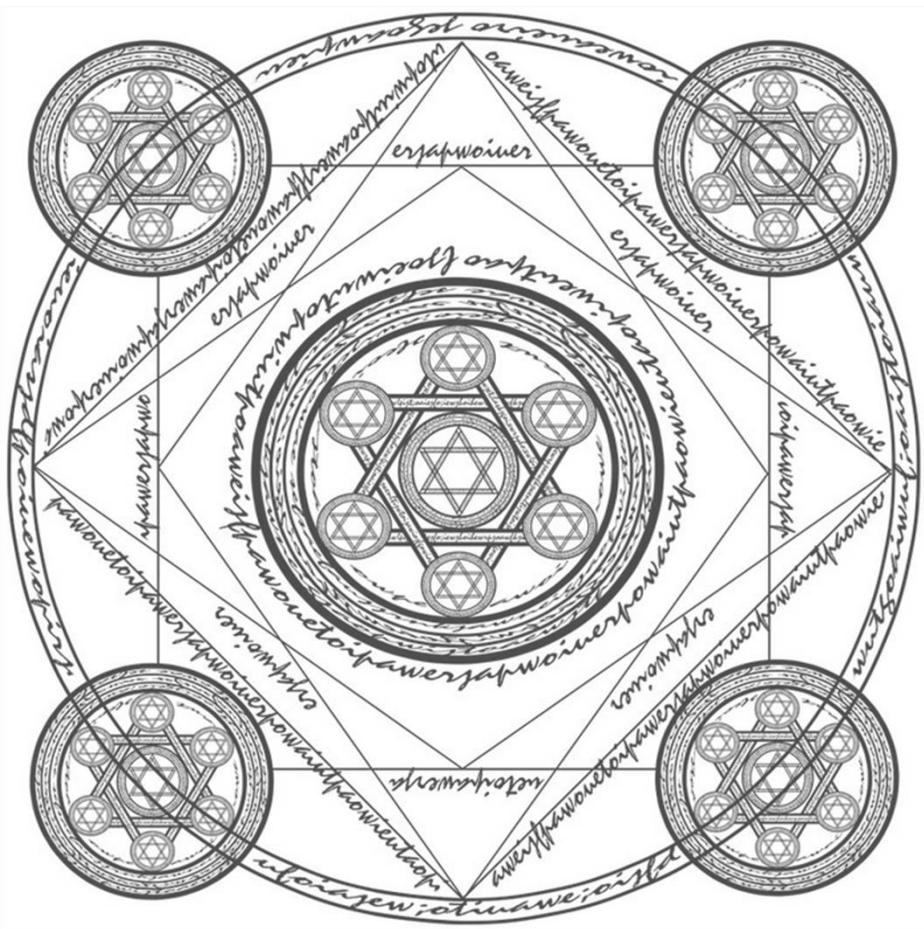
之前提過的RNN有個明確的缺點，無法捕捉長期時間（當序列的距離太大）之間的關聯。

簡單的 RNN 結構無法處理隨著遞歸權重指數級爆炸或消失的問題（Vanishing gradient problem）。

基於上述 RNN 的限制，可以透過 RNN 的變形，也就是 LSTM 來解決。

LSTM 的特色是能夠學習長距離的依賴關係（Long-Term Dependencies），它不同於 RNN 有個單一的神經網絡層（`tanh`），而是有四個層，以特別的方式進行溝通，如圖：





/02

LSTM結構

2.1 輸入門 遺忘門 輸出門

2.2 候選記憶細胞

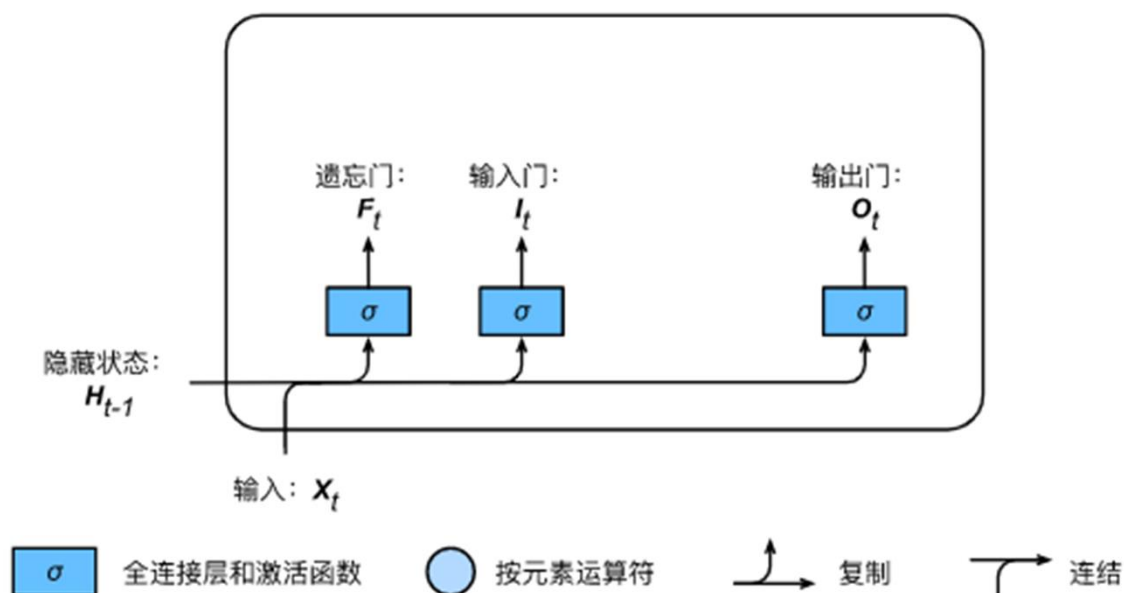
2.3 記憶細胞

2.4 隱藏狀態

2.1 輸入門 遺忘門 輸出門

LSTM 中引入了3個門，即輸入門（input gate）、遺忘門（forget gate）和輸出門（output gate），以及與隱藏狀態形狀相同的記憶細胞（某些文獻把記憶細胞當成一種特殊的隱藏狀態），從而記錄額外的信息。

長短期記憶的門的輸入均為當前時間步輸入 X^*_t 與上一時間步隱藏狀態 H^{*t-1} ，輸出由激活函數為sigmoid函數的全連接層計算得到。



2.1 輸入門 遺忘門 輸出門

具體來說，假設隱藏單元個數為 h ，給定時間步 t 的小批量輸入 $X_t \in \mathbb{R}^{n \times d}$ （樣本數為 n ，輸入個數為 d ）和上一時間步隱藏狀態 $H_{t-1} \in \mathbb{R}^{n \times h}$ 。時間步 t 的輸入門 $I_t \in \mathbb{R}^{n \times h}$ 、遺忘門 $F_t \in \mathbb{R}^{n \times h}$ 和輸出門 $O_t \in \mathbb{R}^{n \times h}$ 分別計算如下：

$$I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i),$$

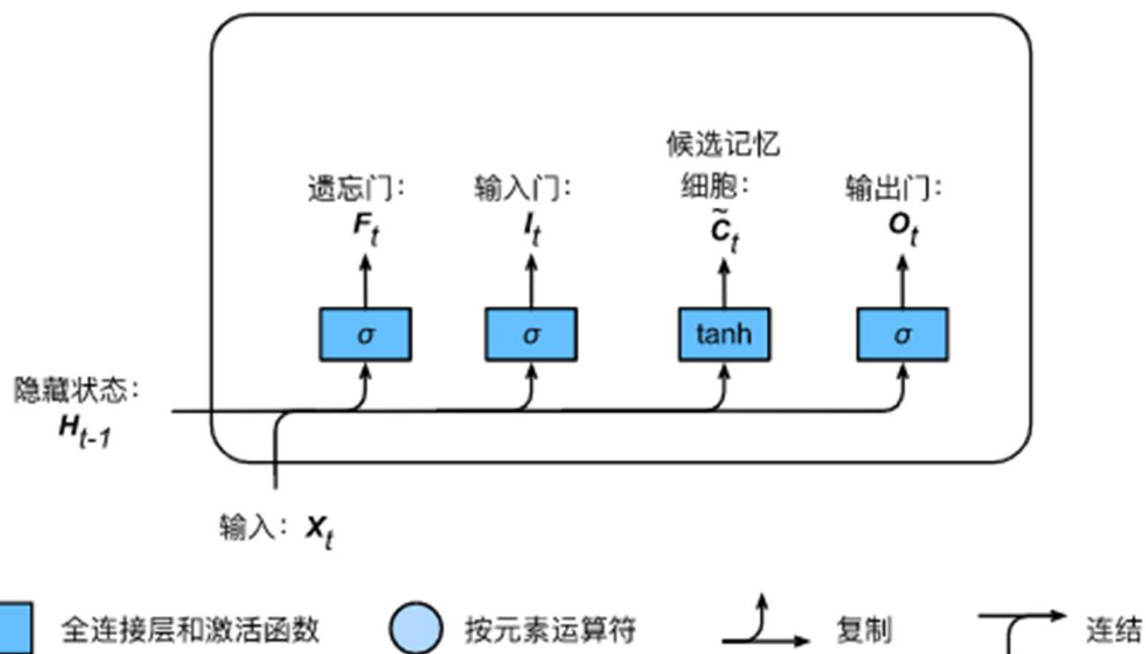
$$F_t = \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f),$$

$$O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o),$$

2.2 候選記憶細胞

- 接下來，長短期記憶需要計算候選記憶細胞* C_t 。它的計算與上面介紹的3個門類似，但使用了值域在 $[-1,1]$ 的tanh函數作為激活函數，時間步* t *的候選記憶細胞計算：

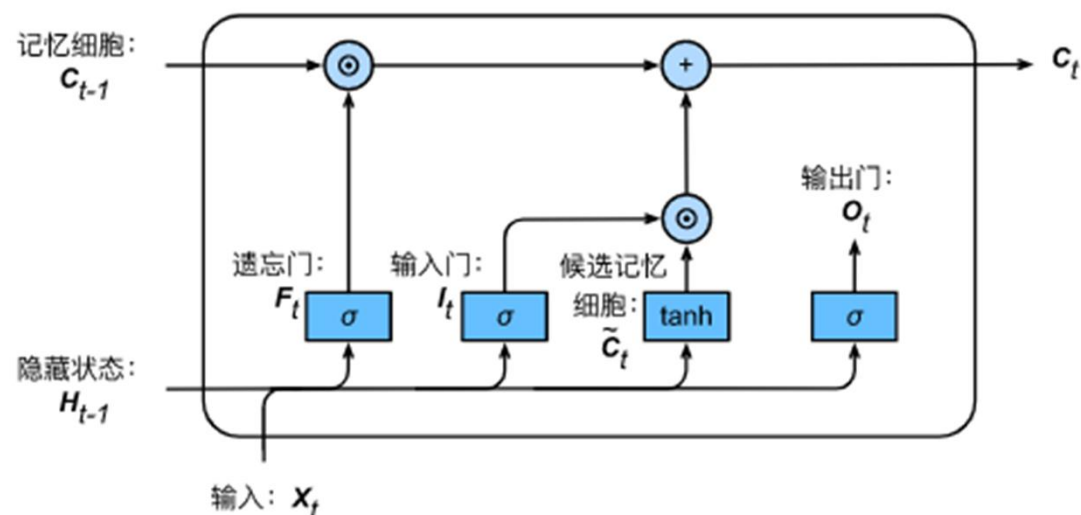
$$\tilde{C}_t = \tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c),$$



2.3 記憶細胞

- 我們可以通過元素值域在 $[0,1]$ 的輸入門、遺忘門和輸出門來控制隱藏狀態中信息的流動，
- 這一般也是通過使用按元素乘法（符號為 \odot ）來實現的。
- 當前時間步記憶細胞 $C_t \in \mathbb{R}^{n \times h}$ 的計算組合了上一時間步記憶細胞和當前時間步候選記憶細胞的信息，並通過遺忘門和輸入門來控制信息的流動：

$$C_t = F_t \odot C_{t-1} + I_t \odot \tilde{C}_t$$



请在插入菜单—页眉和页脚中修改



全连接层和激活函数



按元素运算符



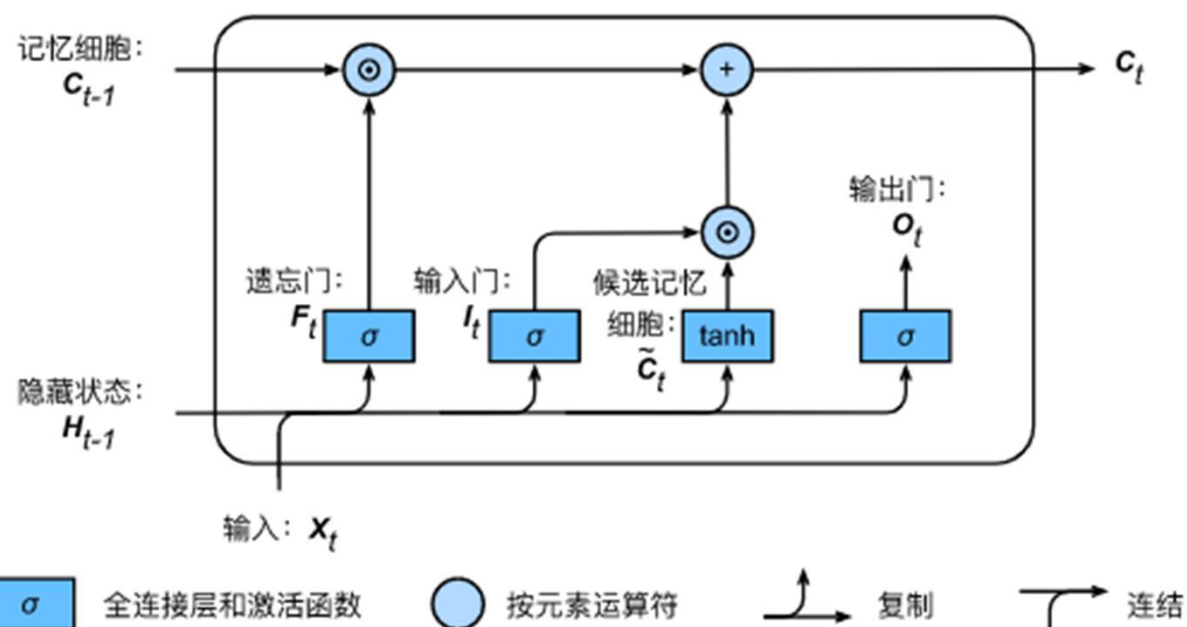
复制



连结

2.3 記憶細胞

- 遺忘門控制上一時間步的記憶細胞 c_{t-1} 中的信息是否傳遞到當前時間步，
- 而輸入門則控制當前時間步的輸入 x_t 通過候選記憶細胞 \tilde{c}_t 如何流入當前時間步的記憶細胞。
- 如果遺忘門一直近似1且輸入門一直近似0，過去的記憶細胞將一直通過時間保存並傳遞至當前時間步。這個設計可以更好地捕捉時間序列中時間步距離較大的依賴關係。

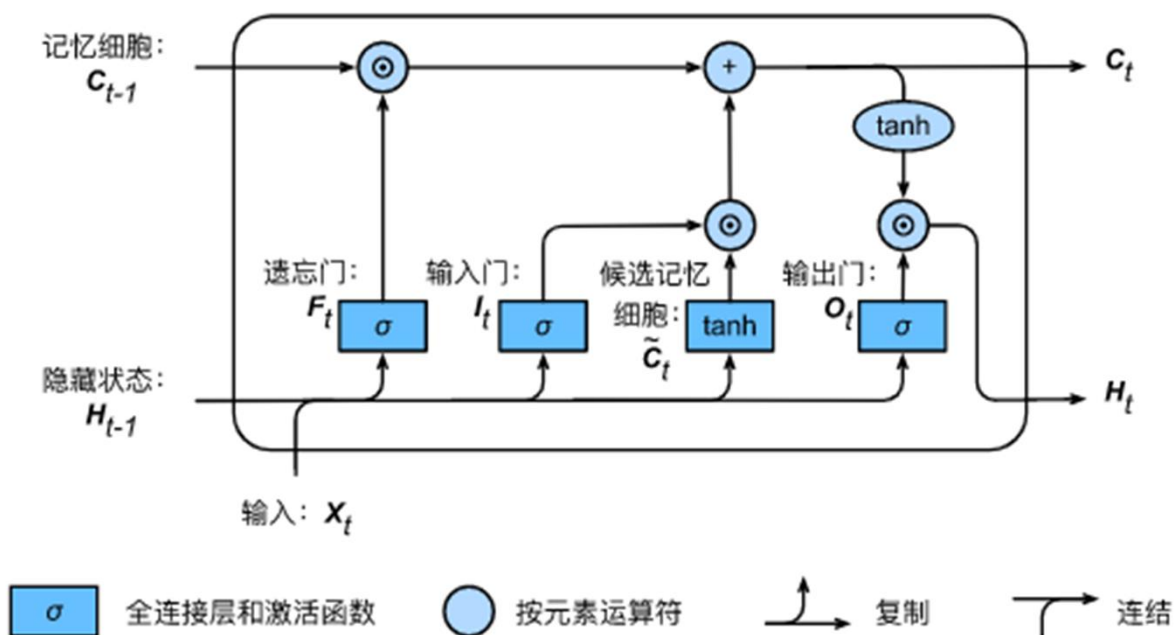


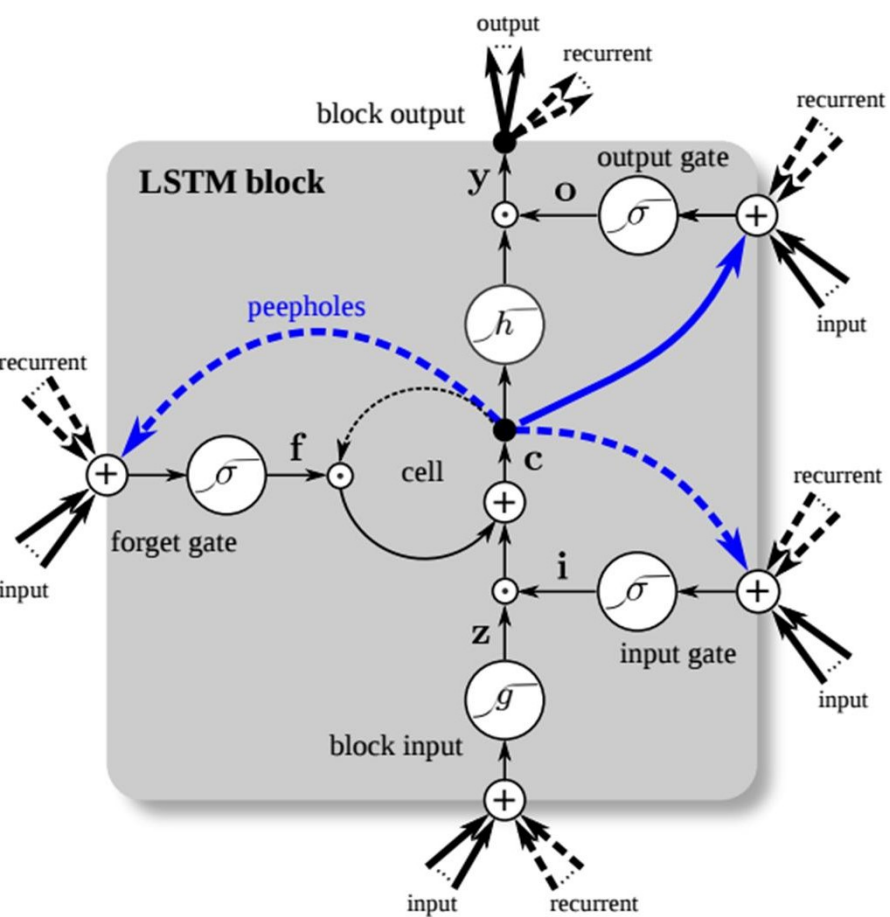
请在插入菜单一页眉和页脚中修

2.4 隱藏狀態

- 有了記憶細胞以後，接下來我們還可以通過輸出門來控制從記憶細胞到隱藏狀態 $H^t \in \mathbb{R}^{n \times h}$ 的信息的流動：

$$H_t = O_t \odot \tanh(C_t).$$





/03

LSTM圖解

核心思想

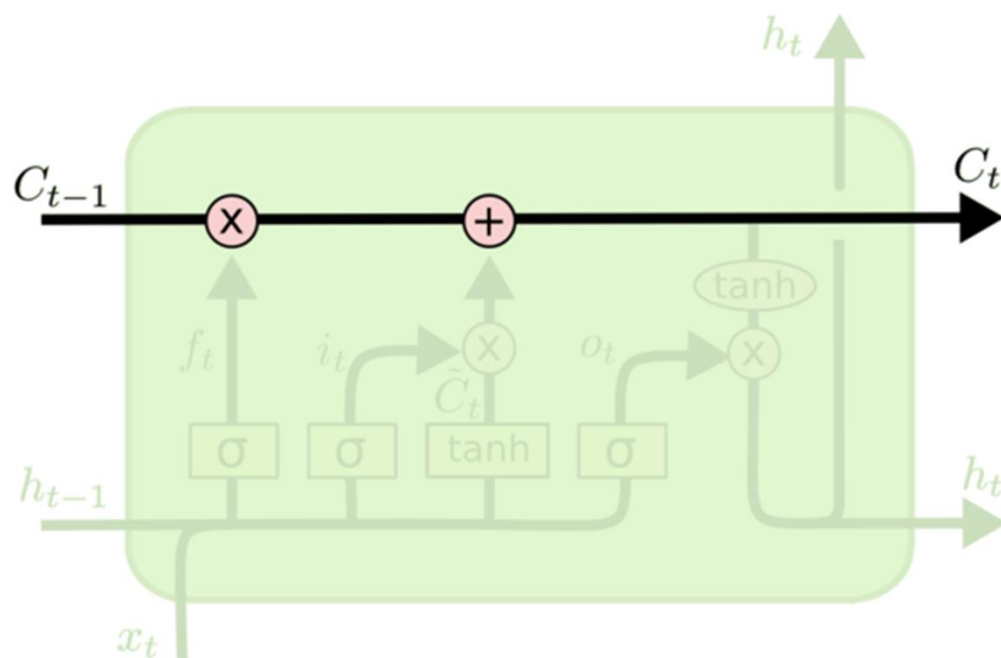
Forget-gate 忘記門

Input-gate 輸入門

Output-gate 輸出門

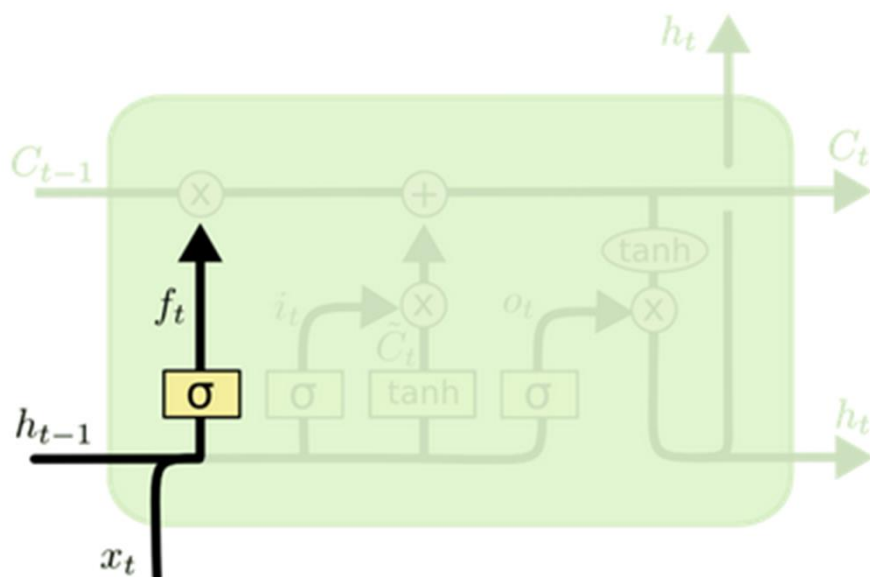
3.LSTM圖解 - 核心思想

- LSTM 的關鍵就是細胞狀態，水平線在圖上方貫穿運行。
- 細胞狀態類似於傳送帶。直接在整個鏈上運行，只有一些少量的線性交互。信息在上面流傳保持不變會很容易。示意圖如下所示：



3.LSTM圖解-Forget-gate 忘記門

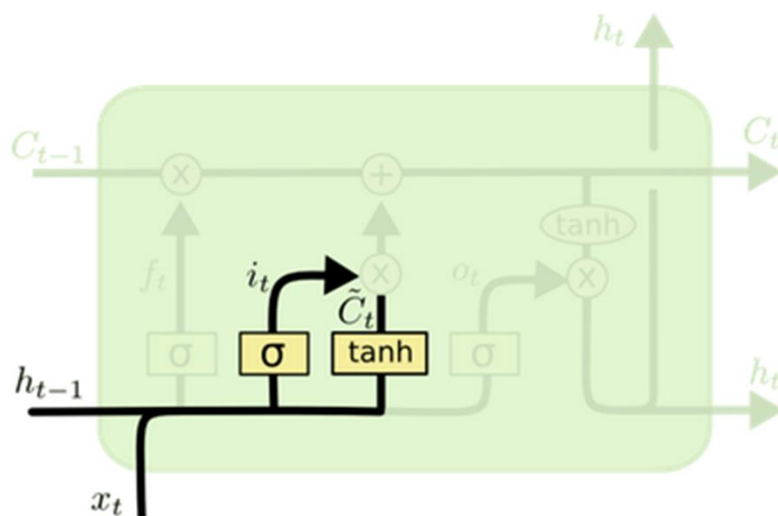
- 作用對象：細胞狀態。
- 作用：將細胞狀態中的信息選擇性的遺忘。
- 操作步驟：該門會讀取 h_{t-1} 和 x_t ，輸出一個在 0 到 1 之間的數值給每個在細胞狀態 c_{t-1} 中的數字。1 表示“完全保留”，0 表示“完全舍棄”



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

3.LSTM圖解-Input-gate 輸入門

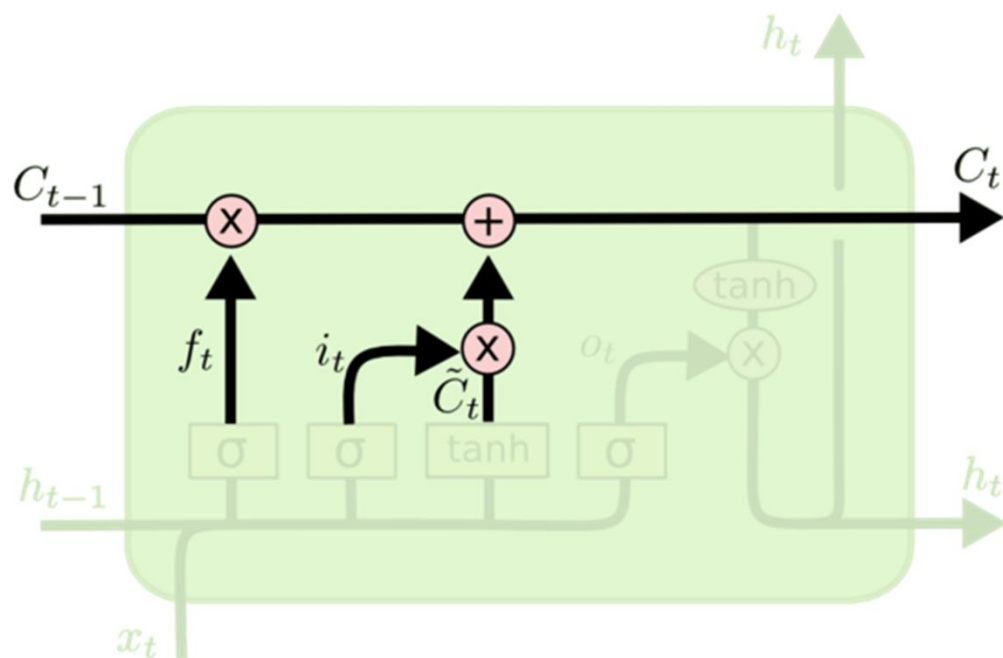
- 作用對象：細胞狀態
- 作用：將新的信息選擇性的記錄到細胞狀態中。
- 操作步驟：
 - 步驟1.sigmoid 層稱 “輸入門層” 決定什麼值我們將要更新。
 - 步驟2.tanh 層創建一個新的候選值向量 \tilde{C}_t 加入到狀態中。：



$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

3.LSTM圖解-Input-gate 輸入門

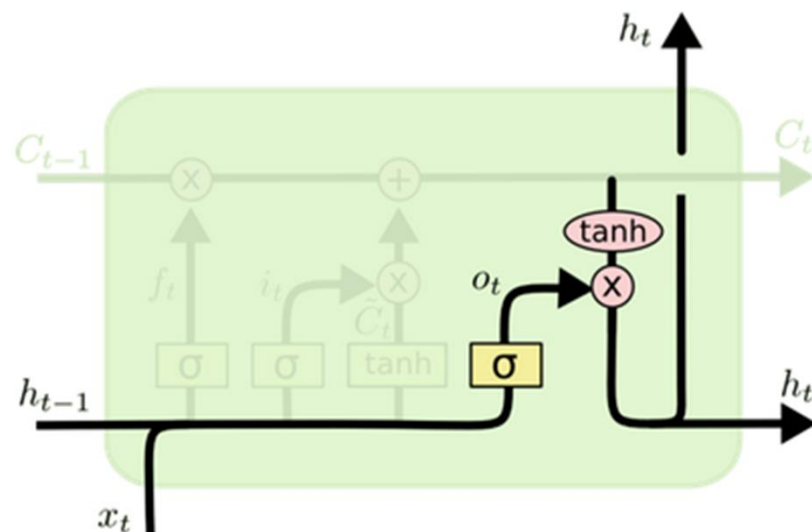
- 步驟3：將 c_{t-1} 更新為 c_t 。將舊狀態與 f_t 相乘，丟棄掉我們確定需要丟棄的信息。接著加上 $i_t * \tilde{c}_t$ 得到新的候選值，根據我們決定更新每個狀態的程度進行變化。



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

3.LSTM圖解-Output-gate 輸出門

- 作用對象：隱層* h_t *
- 作用：確定輸出什麼值。
- 操作步驟：
 - 步驟1：通過sigmoid 層來確定細胞狀態的哪個部分將輸出。
 - 步驟2：把細胞狀態通過 tanh 進行處理，並將它和 sigmoid 門的輸出相乘，最終我們僅僅會輸出我們確定輸出的那部分。



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

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

RNN特點

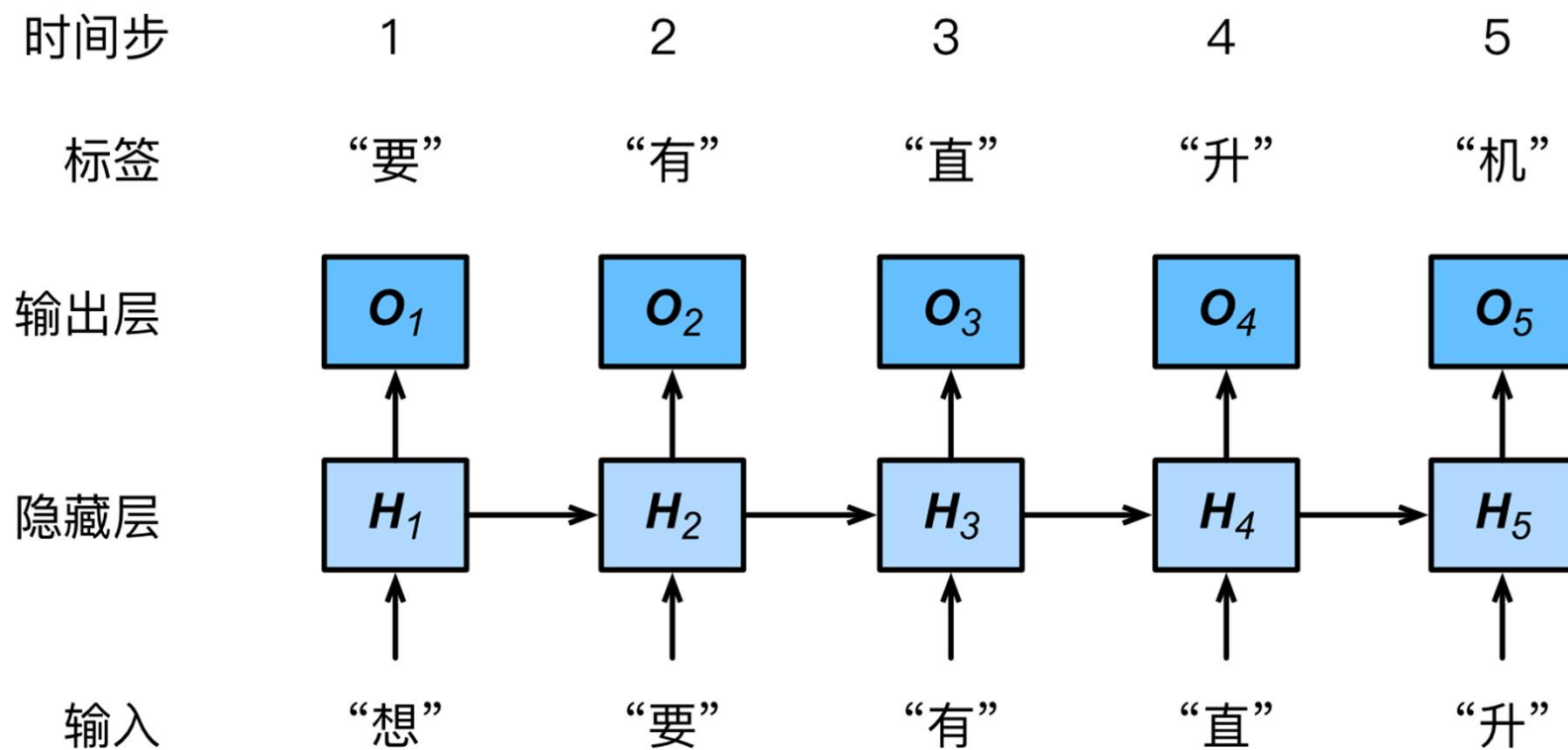
- 1.RNNs主要用于处理序列数据。对于传统神经网络模型，从输入层到隐含层再到输出层，层与层之间一般为全连接，每层之间神经元是无连接的。但是传统神经网络无法处理数据间的前后关联问题。例如，为了预测句子的下一个单词，一般需要该词之前的语义信息。这是因为一个句子中前后单词是存在语义联系的。
- 2.RNNs中当前单元的输出与之前步骤输出也有关，因此称之为循环神经网络。具体的表现形式为当前单元会对之前步骤信息进行储存并应用于当前输出的计算中。隐藏层之间的节点连接起来，隐藏层当前输出由当前时刻输入向量和之前时刻隐藏层状态共同决定。
- 3.标准的RNNs结构图，图中每个箭头代表做一次变换，也就是说箭头连接带有权值。
- 4.在标准的RNN结构中，隐层的神经元之间也是带有权值的，且权值共享。
- 5.理论上，RNNs能够对任何长度序列数据进行处理。但是在实践中，为了降低复杂度往往假设当前的状态只与之前某几个时刻状态相关



/04

LSTM Demo

LSTM Demo



LSTM Demo

- (代碼部分 Demo 有些部分尚待整理解修正)
- Present 時會再修正

thanks!

Complicated to be Simple

Sean

