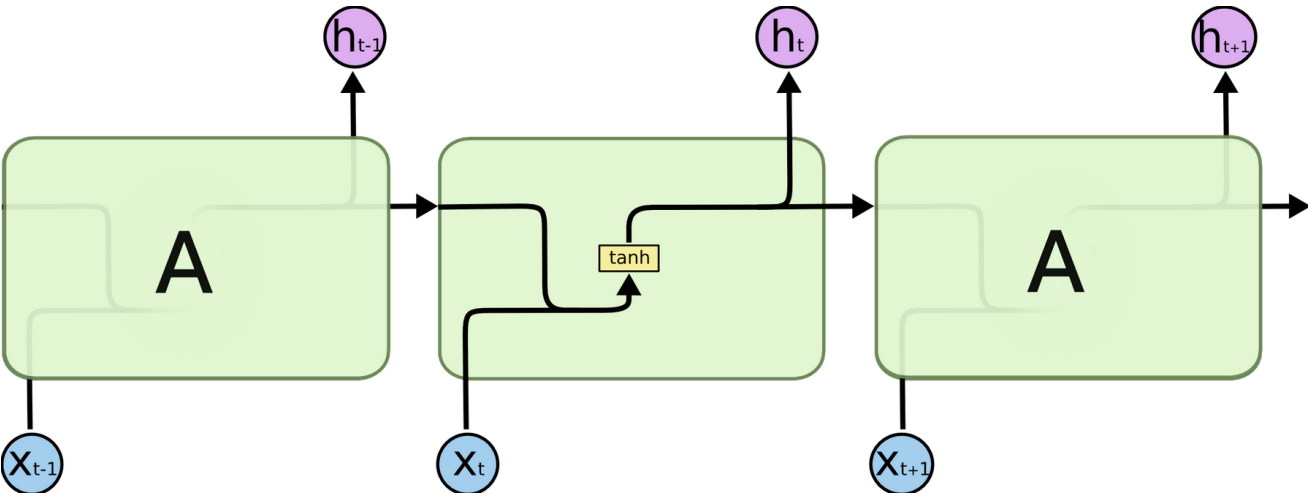


lstm网络介绍

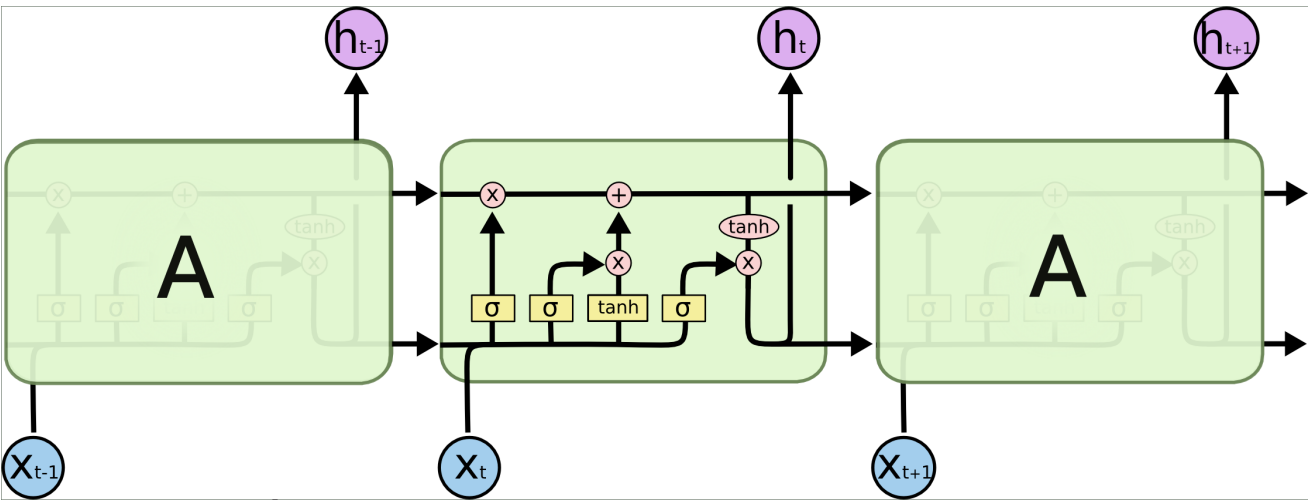
长短时记忆网络(Long Short Term Memory Network, LSTM)，一般就叫做 LSTM ——是一种 RNN 特殊的类型，可以学习长期依赖信息。它成功的解决了原始循环神经网络的缺陷，成为当前最流行的RNN，在语音识别、图片描述、自然语言处理等许多领域中成功应用。

LSTM 通过刻意的设计来避免长期依赖问题，记住长期的信息在实践中是 LSTM 的默认行为。

所有 RNN 都具有一种重复神经网络模块的链式的形式。在标准的 RNN 中，这个重复的模块只有一个非常简单的结构，例如一个 tanh 层。

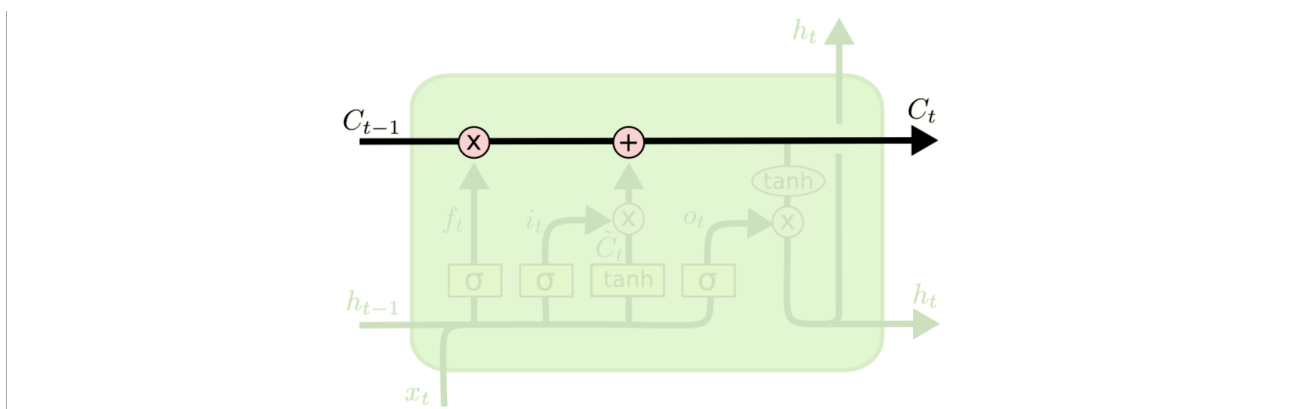


LSTM 同样是这样的结构，但是重复的模块拥有一个不同的结构。不同于 单一神经网络层，这里是有四个，以一种非常特殊的方式进行交互。



LSTM的模型结构解剖

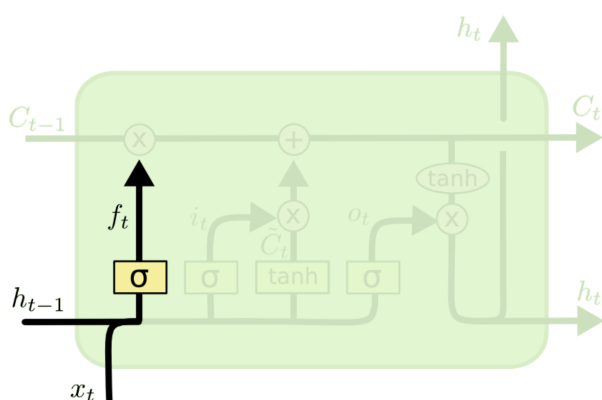
LSTM 的关键就是细胞状态，水平线在图上方贯穿运行。细胞状态类似于传送带。直接在整个链上运行，只有一些少量的线性交互。信息在上面流传保持不变会很容易。每个序列索引位置 t 时刻向前传播的除了和RNN一样的隐藏状态 $h(t)$ ，还多了另一个隐藏状态，如图中上面的长横线。这个隐藏状态我们一般称为细胞状态(Cell State)，记为 $C(t)$ 。



除了细胞状态，LSTM还有一些门控结构。LSTM在每个序列索引位置t的门一般包括遗忘门，输入门和输出门。

LSTM--遗忘门

遗忘门（forget gate）顾名思义，是控制是否遗忘的，在LSTM中即以一定的概率控制是否遗忘上一层的隐藏细胞状态。



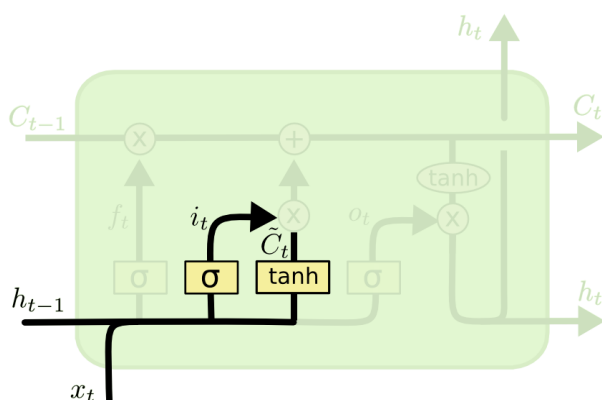
$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

图中输入的有上一序列的隐藏状态 $h(t-1)$ 和本序列数据 $x(t)$ ，通过一个激活函数，一般是sigmoid，得到遗忘门的输出 $f(t)$ 。由于sigmoid的输出 $f(t)$ 在 $[0,1]$ 之间，因此这里的输出 $f(t)$ 代表了遗忘上一层隐藏细胞状态的概率。

其中 W_f, U_f, b_f 为线性关系的系数和偏移。 σ 为sigmoid激活函数。

LSTM--输入门

输入门（input gate）负责处理当前序列位置的输入，它的子结构如下图：



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

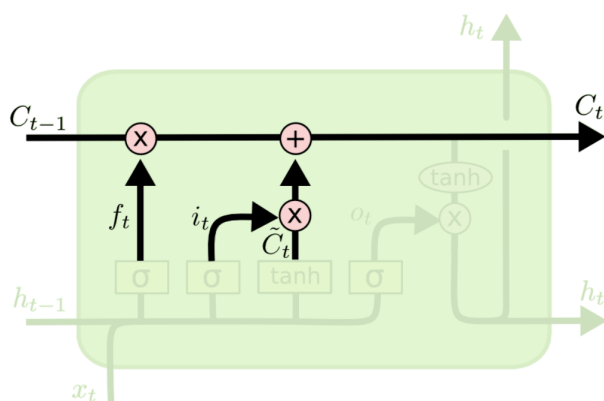
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

输入门由两部分组成，第一部分使用了sigmoid激活函数，输出为 i_t ，第二部分使用了tanh激活函数，输出为 $a(t)$ ，两者的结果后面会相乘再去更新细胞状态。

其中 W_i, b_i, W_c, b_c 为线性关系的系数和偏移。 σ 为sigmoid激活函数。

LSTM--细胞状态更新

现在是更新旧细胞状态的步骤，前面的遗忘门和输入门的结果都会作用于细胞状态 $C(t)$ 。

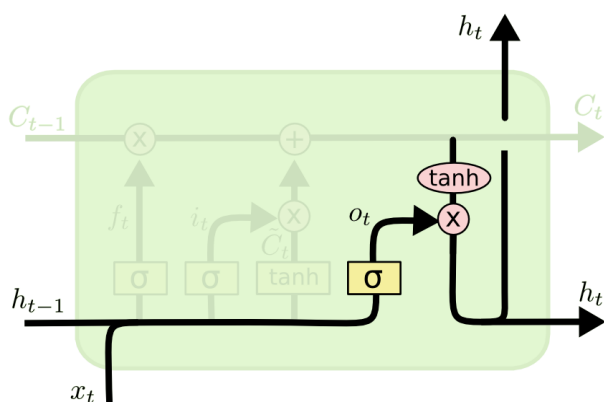


$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

细胞状态 $C(t)$ 由两部分组成，第一部分是 $C(t-1)$ 和遗忘门输出 $f(t)$ 的乘积，第二部分是输入门的 $i(t)$ 和 $a(t)$ 的乘积。

LSTM--输出门

子结构如下：



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

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

隐藏状态 $h(t)$ 的更新由两部分组成，第一部分是 $o(t)$ ，它由上一序列的隐藏状态 $h(t-1)$ 和本序列数据 $x(t)$ ，以及激活函数sigmoid得到，第二部分由隐藏状态 $C(t)$ 和tanh激活函数组成。

LSTM虽然结构复杂，但是只要理顺了里面的各个部分和之间的关系，进而可以深入理解LSTM的原理，以及前向传播和反向传播原理。