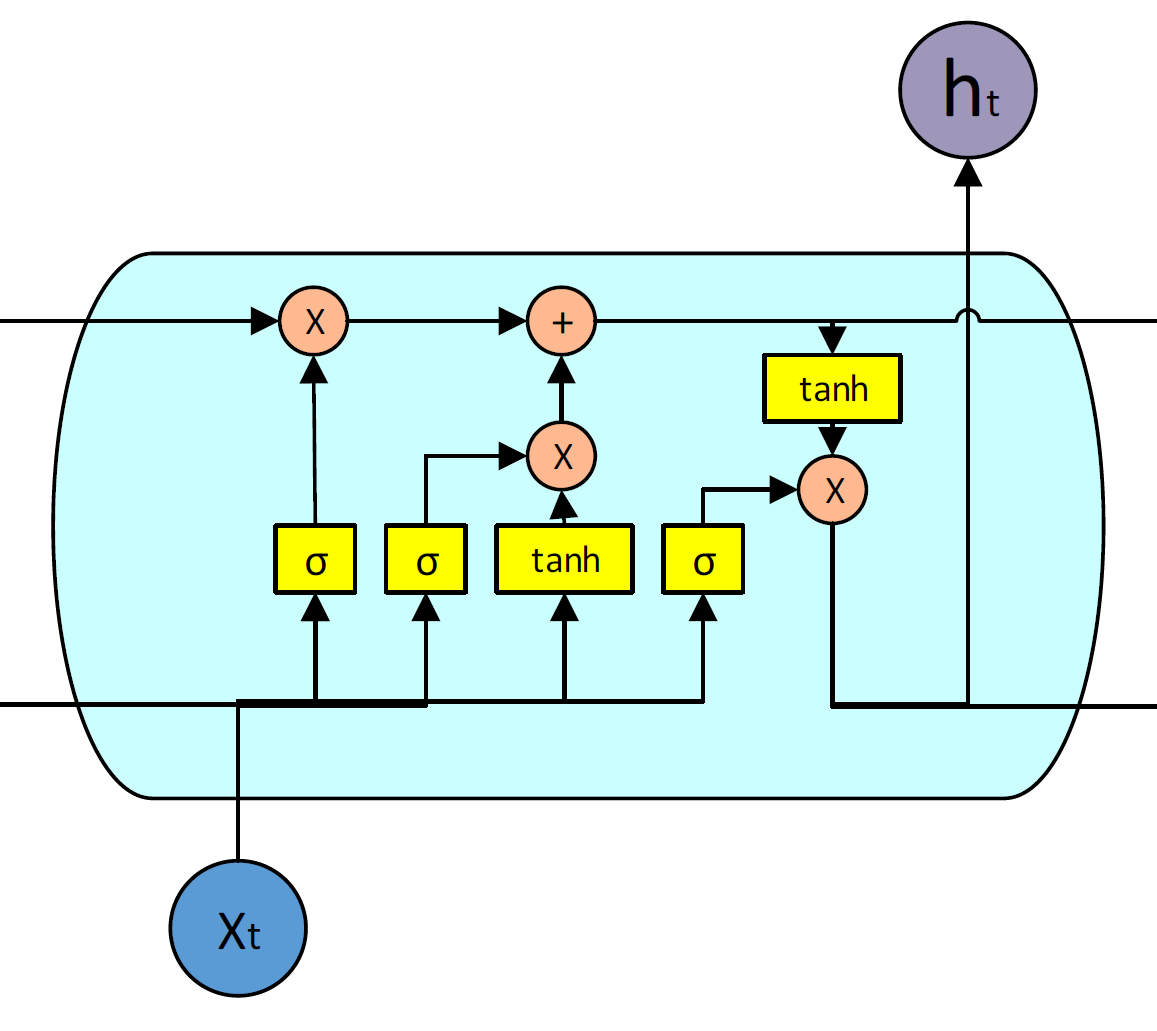
**LSTM模型概述**



Ct

Ct+1

*f*t

*i*t

t

ht

ht+1

*o*t

长短期记忆网络（Long Short-Term Memory, LSTM）是一种时间循环神经网络，与所有的RNN类似，都具有一种重复神经网络模块的链式形式。LSTM模型是为了解决传统RNN模型容易产生梯度消失，难以处理长序列数据的问题而设计的。作为非线性模型，LSTM可作为复杂的非线性单元用于构造更大型深度神经网络。[[1]](#footnote-1)

LSTM通过精心设计的隐藏层神经元来缓解传统RNN的梯度消失问题。[[2]](#footnote-2)LSTM模型中，每个序列索引位置t时刻除了向前传播隐藏状态*h*t，还有另一个隐藏状态*C*t，该状态被称为细胞状态(Cell State)。实际上，*C*t在LSTM中起到了RNN中*h*t的作用。除了细胞状态，LSTM神经元中还包括三种RNN所不具备的门控结构(Gate)，分别遗忘门，输入门和输出门。

**遗忘门**

遗忘门是图中标识为𝑓𝑡的一个sigmoid激活函数：

𝑓𝑡 = 𝜎(𝑊𝑓ℎ𝑡−1 + 𝑈𝑓𝑥𝑡 + 𝑏𝑓)，

函数输出值介于[0,1]之间，代表对上一层序列索引位置t-1时刻细胞状态进行遗忘处理的概率。当模型遇到新的主语𝑥𝑡并希望对其进行新预测时，可以通过遗忘门消除旧主语ℎt-1的一些无关特征。

**输入门**

输入门是图中标识为*i*𝑡的一个sigmoid激活函数和标识为t的一个双曲正切函数（tanh）：

𝑖𝑡 = 𝜎(𝑊𝑖ℎt-1 + 𝑈𝑖𝑥𝑡 + 𝑏𝑖)

𝑡 = tanh(𝑊𝐶ℎt-1+ 𝑈𝐶𝑥𝑡 + 𝑏𝐶)，

输入门负责控制是否将当前时刻的输入值𝑥𝑡融入细胞状态。𝑖𝑡的值介于[0,1]之间，代表记住这一层输入信息的概率。𝑖𝑡与𝐶𝑡的乘积表示当前细胞状态所需要添加的信息。

**输出门**

输出门是图中标识为*o*𝑡的一个sigmoid激活函数：

𝑜𝑡 = 𝜎(𝑊𝑜ℎ𝑡−1 + 𝑈𝑜𝑥𝑡 + 𝑏𝑜),

函数的值在[0,1]之间，决定了当前时刻细胞状态的输出部分，其目的在于过滤细胞状态，并且从𝐶𝑡中产生隐藏状态ℎ𝑡，输出给下一个序列索引位置t+1时刻。其数学表达式为：

ℎ𝑡 = 𝑜𝑡 ⨀ tanh (𝐶𝑡)

1. 陈亮, 王震, 王刚. 深度学习框架下LSTM网络在短期电力负荷预测中的应用[J]. 电力信息与通信技术, 2017(5):8-11. [↑](#footnote-ref-1)
2. 华泰9：p8 [↑](#footnote-ref-2)