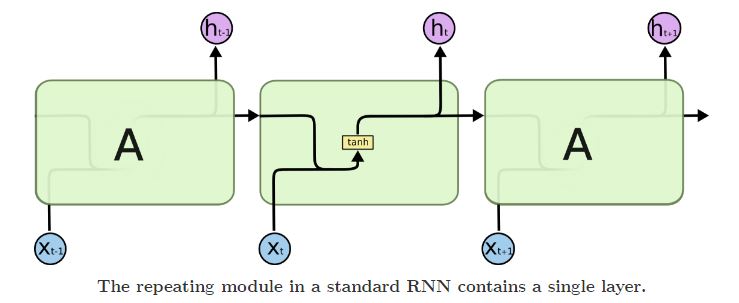
本周学习：尝试理解LSTM（其二）

接着上次的尝试理解LSTM（其一）。

LSTM 网络

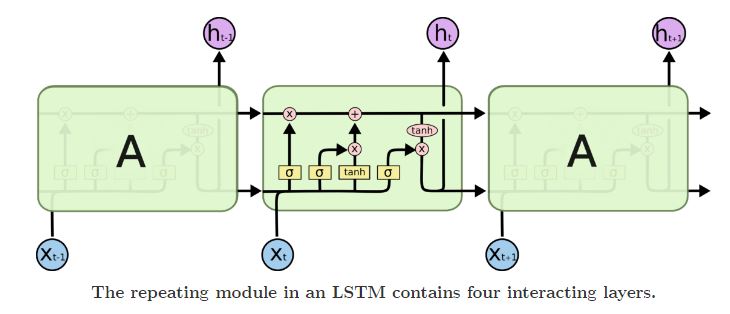
LSTM 结构的一个优势在于可以很好的解决 “long-term dependency” 的问题，”长期记忆”是LSTM结构与生俱来的特性，而不需要刻意地去学习。

所有的RNN结构都是有一个不断重复的模块，在标准的RNN结构中，这个不断重复的模块是一个单层的tanh , 如下图所示：



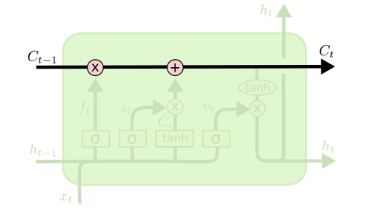
表达式简单来说就是： ht=tanh(Wh⋅[ht−1,Xt]+bh)

LSTM 网络也是有一个不断重复的模块，但是这个模块不是一个简单的tanh层，而是有复杂的四个网络层，用一种特殊的方式连接在一起，如下图所示：



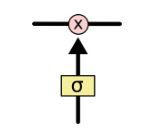
LSTM 的核心思想

LSTM 网络的关键是 cell state，就是网络结构中最上面的那条水平线，如下图所示：



这条水平线贯穿整个网络，与一些线性组合相结合，可以将信息无改变的传递。

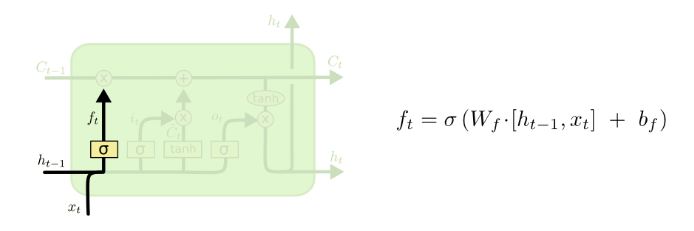
LSTM 网络具备的另外一种能力就是移除或者增加一些信息，这个过程是由一些称为传送门的结构来控制的，传送门可以让信息有选择的通过，这种门结构由sigmoid 层 与 点乘运算符组成。如下图所示：



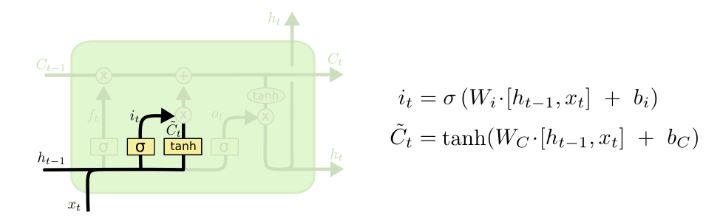
其中，sigmoid 层输出 0-1 之间的数，控制信息传递的概率，1表示信息完全通过，0表示信息完全不能通过，一个典型的LSTM 网络有三个这样的传送门用来控制 cell state.

逐步深入LSTM

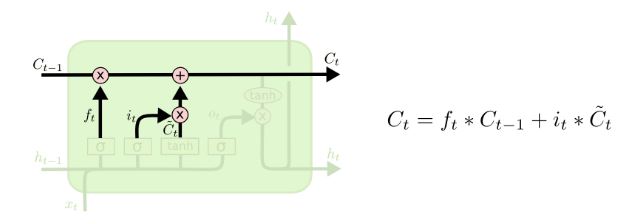
LSTM 网络的第一步就是决定哪些信息将从 cell state 中剔除掉，这一步是由一个sigmoid 层来负责的，sigmoid 层会根据输入的ht−1 和 xt输出一系列 0-1 之间的数，这些数表示了状态 Ct−1 中信息保存下来的概率，1 表示完全保存，而 0 表示完全剔除。结构及表达式如下图所示：



接下来的一步是要决定哪些新的信息需要存储在cell state 中，这一步有两部分，首先，一个称为 “input gate layer”的sigmoid 层会决定哪些信息要被更新，然后 一个 tanh 层会创建一个新的向量 C~t , 这个新的向量有可能被加入到 cell state 中，接下来的一步，我们会结合这两部分对cell state 创建一个更新，结构及表达式如下图所示：

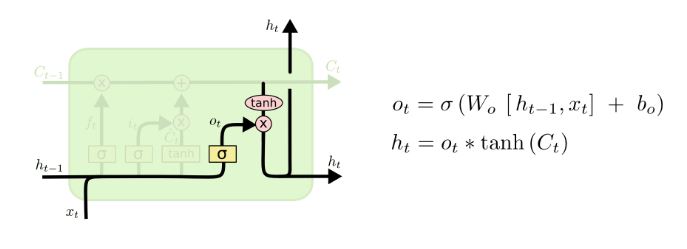


现在，就是对网络的旧状态Ct−1 进行更新到新的状态 Ct , 前面两部已经做好了所有的准备的工作，我们只需要进行简单的线性组合运算即可，结构及表达式如下图所示：



ft 和 it 是两个控制门，Ct−1 是网络的旧状态，C~t 是网络更新的信息，ft 表示有多少旧信息会被剔除，而 it 表示会有多少新的信息加入进来。

最后，我们需要给出输出，我们同样需要一个sigmoid层来决定Ct 中哪些是需要被输出的，然后我们让 cell state 通过一个 tanh 层 将值映射到到 [-1, 1]之间，然后乘以sigmoid层的输出，这样最终输出的就是我们决定输出的。结构与表达式如下图所示：



总得来说，LSTM结构，利用了几个传送门来控制信息的删除与更新，通过一些设计好的连接方式，可以拥有“长期记忆”的能力。与标准的RNN结构最大的区别就在于，LSTM是利用模块层里的神经网络来控制信息，而RNN是利用模块本身的连接方式来处理信息。所以与RNN相比，LSTM处理时序信息的能力要更强。性能也更稳定。

参考资料

1. <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
2. <http://blog.csdn.net/matrix_space/article/details/53376870>