机器学习：深入理解LSTM网络 (二)

翻译 Matrix_11 2016-11-28 16:05:59 5574 收藏 1

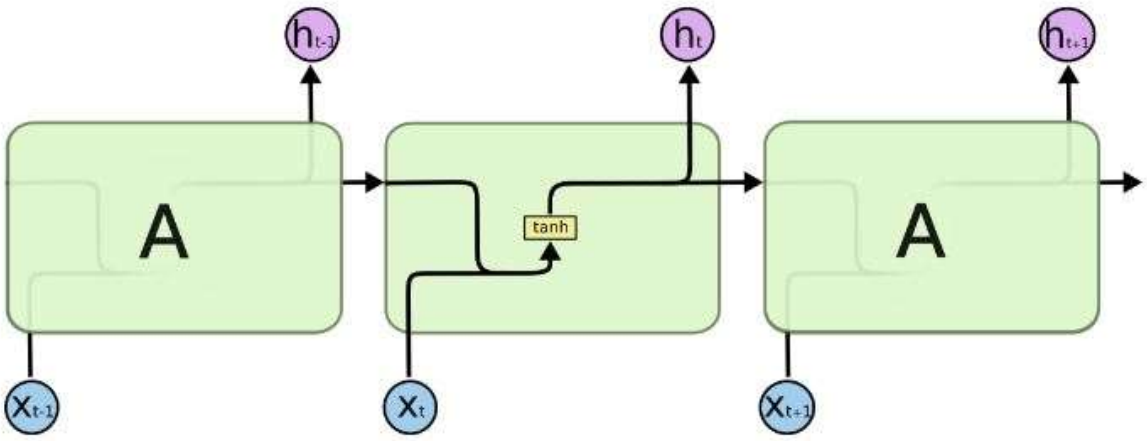
分类专栏： 机器学习 文章标签： 网络 机器学习

之前我们介绍了RNN 网络结构以及其所遇到的问题，RNN 结构对于关联度太长的时序问题可能无法处理，
简单来说，RNN对于太久远的信息不能有效地储存，为了解决这个问题，有人提出了LSTM的网络结构，LSTM 网络结构最早是由 Hochreiter & Schmidhuber 在1997 年提出的，随着后来研究者的不断改进，LSTM网络在很多问题上都有非常好的表现，并且得到广泛的关注与应用。

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。

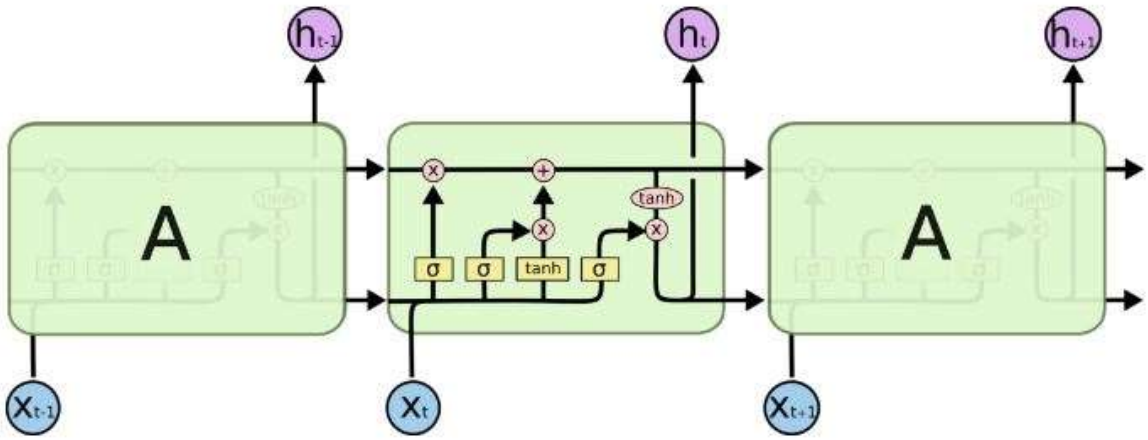
所有的RNN结构都是有一个不断重复的模块，在标准的RNN结构中，这个不断重复的模块是一个单层的tanh , 如下图所示：



The repeating module in a standard RNN contains a single layer.

表达式简单来说就是： $h_t = \tanh(W_h \cdot [h_{t-1}, X_t] + b_h)$ ht=tanh(Wh·[ht−1,Xt]+bh)

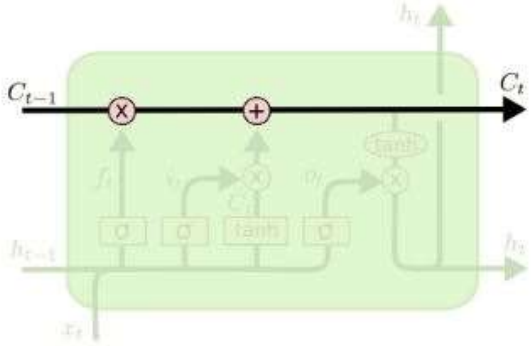
LSTM 网络也是有一个不断重复的模块，但是这个模块不是一个简单的tanh层，而是有复杂的四个网络层，用一种特殊的方式连接在一起，如下图所示：



The repeating module in an LSTM contains four interacting layers.

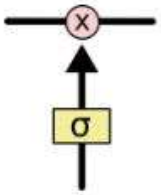
LSTM 的核心思想

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：

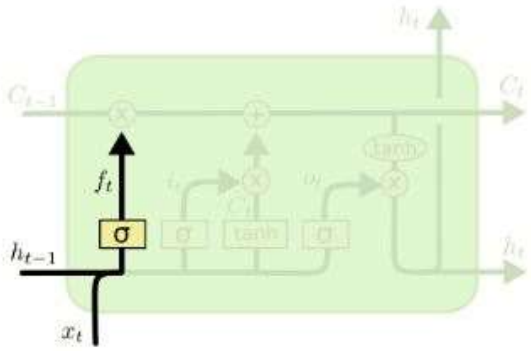


其中，sigmoid 层输出 0-1 之间的数，控制信息传递的概率，1表示信息完全通过，0表示信息完全不能通过，一个典型的LSTM 网络有三个这样的传送门用来控制 cell state.

逐步深入LSTM

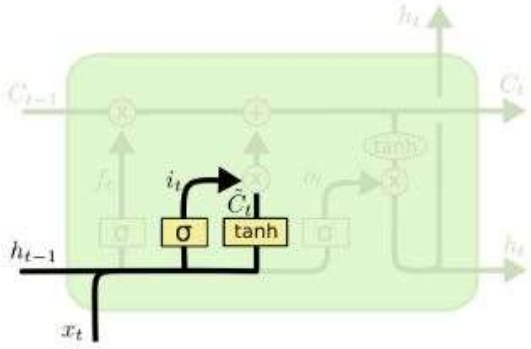
LSTM 网络的第一步就是决定哪些信息将从 cell state 中剔除掉，这一步是由一个sigmoid 层来负责的，sigmoid 层会根据输入的h_{t-1}ht-1 和 $x_t$xt输出一系列 0-1 之间的数，这些数表示了状态 C_{t-1}Ct-1 中信息保

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：



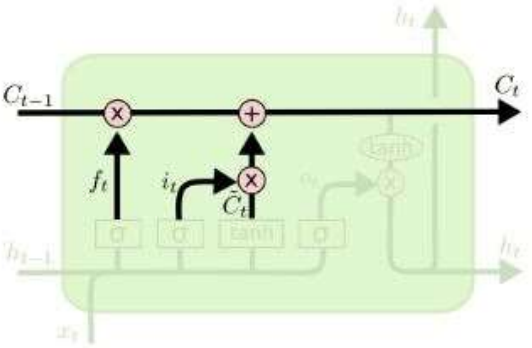
$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

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 $\tilde{C}_t$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：



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

现在，就是对网络的旧状态C_{t-1}Ct−1 进行更新到新的状态 C_tCt , 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：

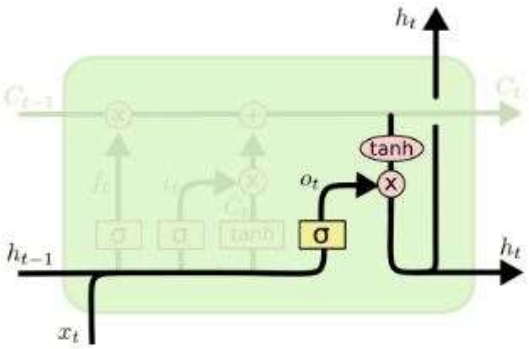


$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

$f_t$ft 和 $i_t$it 是两个控制门，C_{t-1}Ct−1 是网络的旧状态，$\tilde{C}_t$C~t 是网络更新的信息，$f_t$ft 表示有多少旧信息会被剔除，而 $i_t$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C_tCt 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：



$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

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。

这里介绍的只是最常见的一种LSTM结构，实际上还有很多LSTM的变种，更加详细的介绍，可以参考colah 的博客。

http://colah.github.io/posts/2015-08-Understanding-LSTMs/