



LSTM几个关键“门”与操作

李倩 发表于 2018-10-09 16:17:07



Imagination Tech

+关注

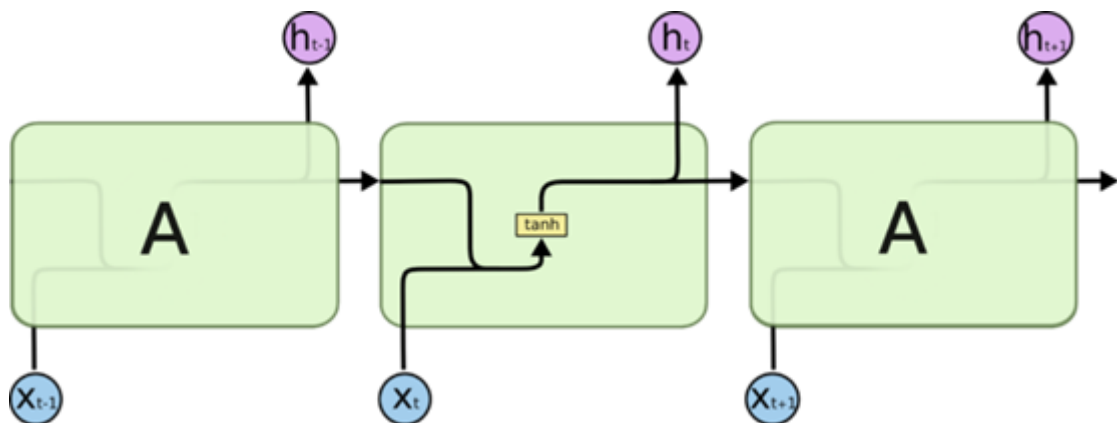
1. 摘要

对于RNN解决了之前信息保存的问题，例如，对于阅读一篇文章，RNN网络可以借助前面提到的信息对当前的词进行判断和理解，这是传统的网络是不能做到的。但是，对于RNN网络存在长期依赖问题，比如看电影的时候，某些细节需要依赖很久以前的一些信息，而RNN网络并不能很好的保存很久之前的信息，随着时间间隔不断增大，RNN网络会丧失学习到很远的信息能力，也就是说记忆容量是有限的。例如，对于阅读一本书，从头到尾一字不漏的阅读，肯定是越远的东西忘得越多。所以引入了LSTM网络，对于LSTM来解决梯度消失梯度爆炸问题。

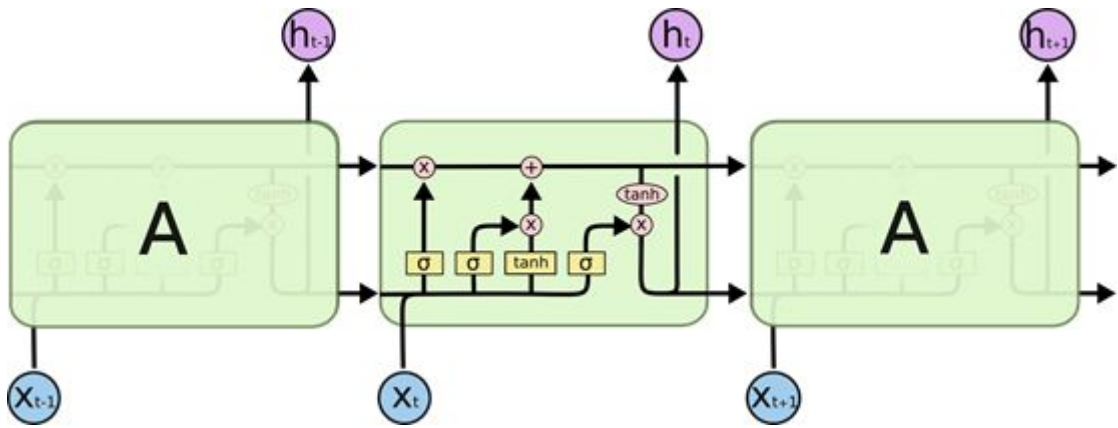
LSTM其实也是RNN的一种变体，大体结构几乎是一样的，但他们又有区别他的“记忆细胞”被改造过，该记忆的信息会一直传递下去，而不该记忆的东西会被“门”截断。

2. LSTM网络

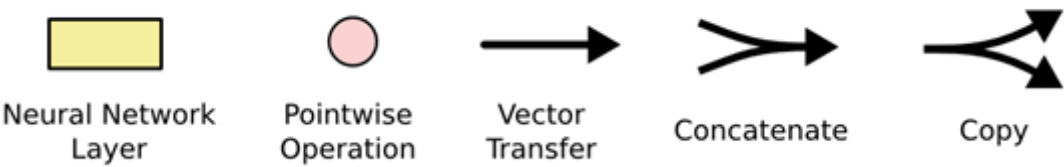
对于RNN网络都是具有重复神经网络模块的链式形式，在一个标准RNN中这个重复的模块只要一个非常简单的结构，例如一个tanh层。



对于LSTM同样是这样的结构，但是重复的模块变得复杂了。



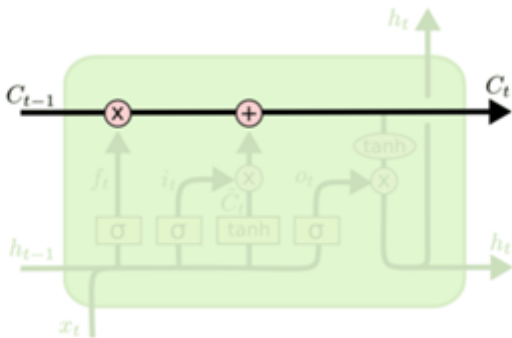
对于图中的一些细节，



在LSTM网络的图中，每一条黑线表示向量的传输，从一个节点的输出传入其它节点的输出。对于粉红色的圈代表着节点的操作，如向量的求和等。黄色的矩阵是学习到的网络层。Concatenate的线表示向量的连接，Copy的线表示内容复制，然后输出到不同的位置。

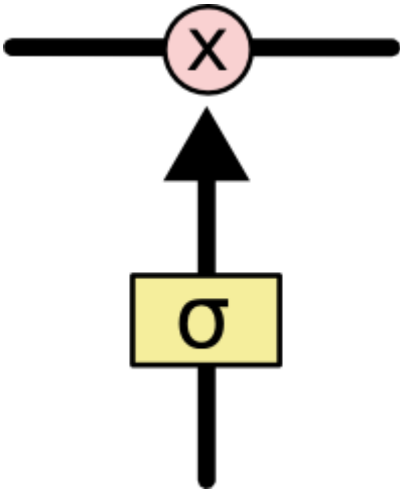
3. LSTM的关键点

对于LSTM的关键是细胞状态，细胞的状态类似于传送带，直接在整个链路上运行，只有一些少量的线性交互。



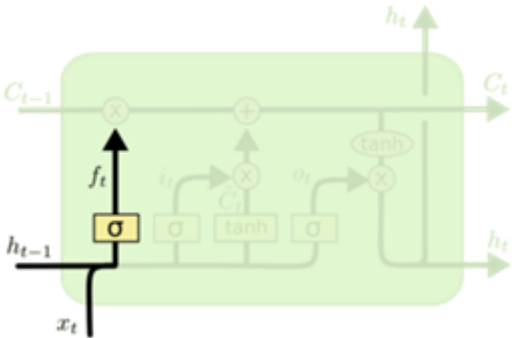
4. 细胞状态的控制

LSTM是通过“门”来控制细胞状态的，通过“门”来选择性通过，去除或者增加信息到细胞状态。它包含一个sigmoid网络层和一个pointwise乘法操作。Sigmoid层输入0到1之间的概率值，藐视每个部分有多少量可以通过，0代表不允许任何量通过，1代表允许任何量通过。LSTM是拥有三个门，来保护和控制细胞状态的。



5. LSTM几个关键 “门” 与操作

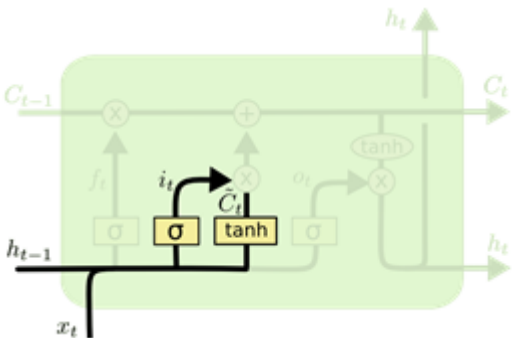
STEP 1: 决定从 “细胞状态” 中丢弃什么信息即 “忘记门” 。比如一篇小说的推导，小说中可能有几个人物，在读小说时候我们都会记住，但有些是不重要的这时候我就选择忘掉，而把重要的东西记住。总之，如果要决定之前的记忆以多大的程度被忘掉或者需要被保留，我们需要通过sigmoid函数去参考上一次的结果与当前考到的内容。



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

ft是输出0到1的概率，表示已多大的程度忘记之前的信息。h表示上一次输出的结果，x表示当前输入的信息。

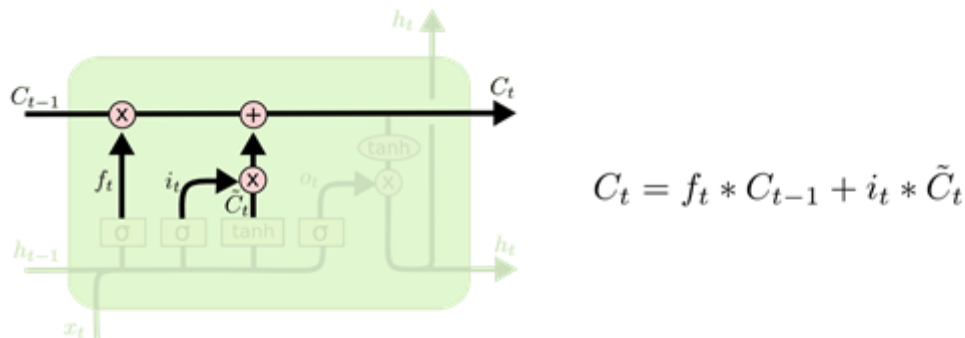
STEP 2: 决定放什么信息到 “细胞状态” 中。从step 1中我们在传送带上忘记了一些东西，那我们应该补充点东西，对细胞转态进行更新。也就是说哪些东西要记住，这一步就是君顶放什么心信息到 “细胞状态” 中。



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

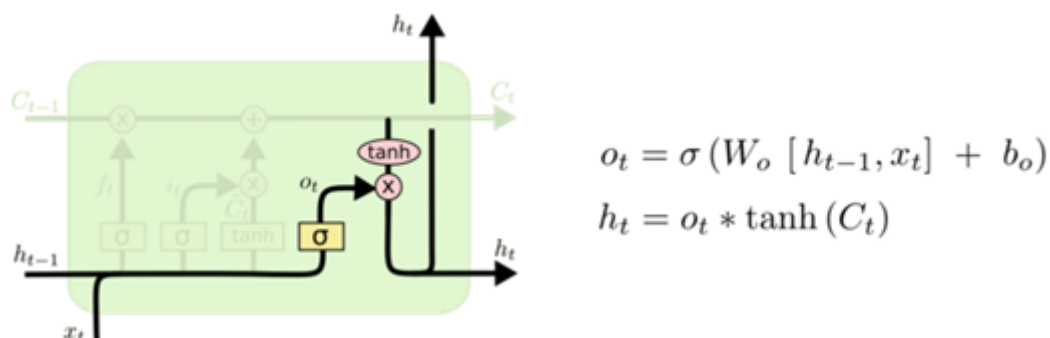
这里的Sigmoid层是决定什么值需要更新，也就是这一时刻我们新学到我们学到的东西哪些应该要记住，以多大的程度选择记住；Tanh层创建了一个新的候选值 C_t ，表示我们新学到的东西。

STEP 3: 更新细胞状态。前面两步就是为细胞状态的更新做准备的，更新 C_{t-1} 为 C_t 。也就是把旧状态与 f_t 相乘，丢掉我们确定需要丢弃的信息，加上 $i_t * C_t$ 。这就是新的候选值，根据我们决定更新每个状态的程度进行变化。



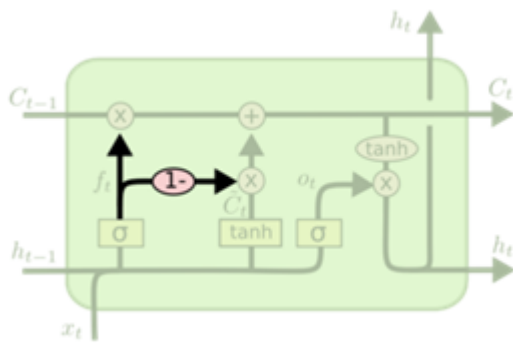
从数学的角度来看，与RNN相比，在求偏导的时候RNN是进行链式相乘的，有一个效果不好，那么就会影响整个系统，效果影响到整个网络。而在LSTM中，从以上的公式可看出，做完Tanh后没在进行求和，所以再求偏导的时候，就算有一个效果不好也不会影响到整个网络，因为这里是进行加法运算的，所以很大程度上解决了梯度消失梯度爆炸问题。

STEP 4: 基于“细胞状态”输出。首先运行一个Sigmoid层来确定细胞状态的哪个部分将输出，接着用Tanh来处理细胞状态，这是一个从-1到1的值。然后再讲它和Sigmoid门的输出相乘，输出我们确定输出的那个部分。比如我们可能需要单复数信息来确定输出“他”还是“他们”。最终输出我们想要输出的部分。



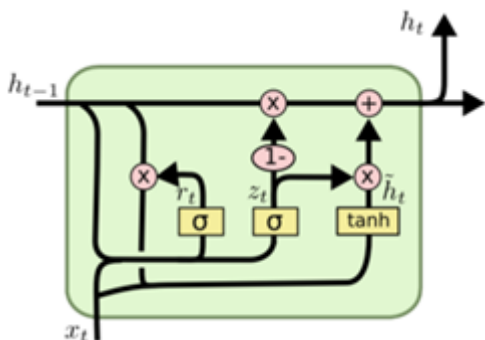
6. LSTM主要变体

变种1：通过使用coupled忘记和输出门，之前是分开确定需要忘记和添加的信息，这里是一同作出决定。



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

变种二：Gated Recurrent Unit (GRU)，2014年提出的，他将忘记门和输出门合成一个单一的更新门，同样还混合了细胞状态和隐藏转态，和其他一起改动，与标准的LSTM简单。



$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

相关推荐：

[深度学习和神经网络](#)

[1.1神经元工作原理](#)

[1.2神经网络介绍](#)



赞



收藏



相关话题

神经网络

+关注

RNN

+关注



电子发烧友网

收听电子行业动态，抢先知晓半导体行业

查看

评论(0)

[登录](#) 后参与评论

相关文章



参与英飞畅知识小测验，
赢精美礼品

[L-com](#) 2019-07-23



深度学习和神经网络

[发烧友学院](#) 1970-01-01



华强芯城|限时领现金红包，
下单赢华为P30pro、
西门子冰箱等豪礼

2019-07-18



人工智能算法能否复制人类的大脑功能

38次阅读 2019-07-30



如何自制18W全兼容快充
充电器？你只需这份清单

2019-05-13



微软Xiaowuhu根据自己亲身经历的学习，归纳出来以下教程

340次阅读 2019-07-30



【有干货】RTOS视频、
文档教程免费下载学习

[RT-Thread](#) 2019-07-18



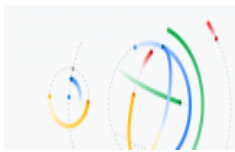
戴尔易安信关注FPGA的原因是什么

51次阅读 2019-07-29



人工智能在工业云时代的关键云让人更智能

51次阅读 2019-07-29



预测深度神经网络泛化性能差异

288次阅读 2019-07-28



【点击报名】穿戴市场应用技术和方案

在线研讨会 2018-07-29



神经网络的优缺点讨论

50次阅读 2019-07-27



单v100 GPU，4小时搜索到一个鲁棒的网络结构

202次阅读 2019-07-27

姓名	毕业院校	职位	入职时间	薪资范围
张一	清华大学	研发	2015-2017	1.2-1.5M, 年薪150W-200W
李二	北京大学	研发	2016-2018	年薪180W, 年薪200W
王三	复旦大学	研发	2017-2019	年薪200W, 年薪250W
赵四	上海交通大学	研发	2018-2020	年薪250W, 年薪300W
陈五	浙江大学	研发	2019-2021	年薪300W, 年薪350W
周六	中国科技大学	研发	2020-2022	年薪350W, 年薪400W

华为首次公开2019顶尖学生的年薪，最高201万！

707次阅读 2019-07-27



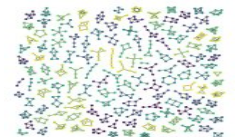
【免费参会】直击国产FPGA最新系列产品，抢占市场先机

在线研讨会 2019-07-24



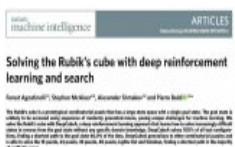
某些嵌入式AI处理器比其它更智能？

319次阅读 2019-07-23



介绍你需要的图建模论文，图中的各种隐藏知识、强大且复杂的工具

650次阅读 2019-07-22

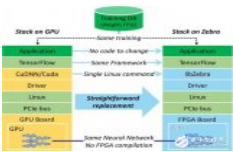


继围棋、德扑之后，AI现在又在魔方领域碾压人类

187次阅读 2019-07-22

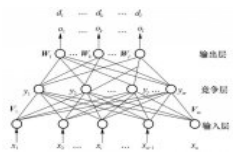


【有干货】RTOS视频、文档教程免费下载学习
RT-Thread 2019-07-18



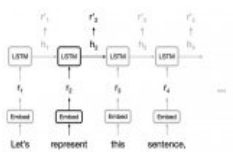
Mipsology Zebra在深度学习推断的应用

52次阅读 2019-07-21



对偶传播神经网络的详细资料说明

303次阅读 2019-07-20

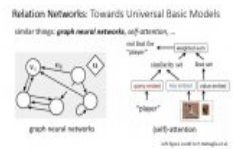


循环神经网络卷积神经网络注意力文本生成变换器编码器序列表征

236次阅读 2019-07-19



【点击报名】穿戴市场应用技术和方案
在线研讨会 2018-07-29



Microsoft最新研究提基于关系网络的视觉建模

385次阅读 2019-07-19