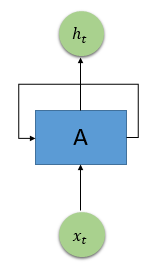
**循环神经网络**

人工神经网络和卷积神经网络的前提假设是元素之间是相互独立的，输入和输出也是相互独立的。但在现实世界中，很多元素都是相互连接的，比如，我出生在中国，从小在这里长大，所以我能说一口流利的。这里填空，都知道填“中文”，是因为能够根据上下文的内容进行推断，所以就出现了循环神经网络，循环神经网络的本质是像人一样拥有记忆能力，它的输出依赖于当前的输入和记忆。

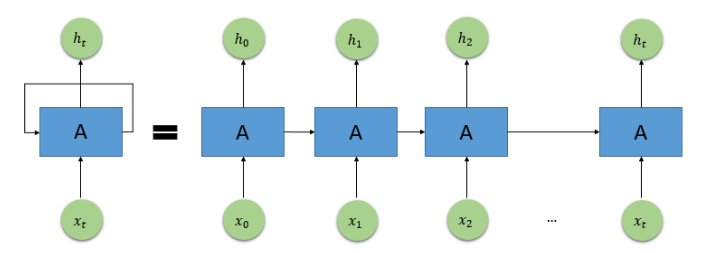
**RNN**

循环神经网络（Recurrent Neural Network，RNN）是一种序列模型，用来处理序列化数据（比如，语音、文本、视频等等），它的应用场景比较多，比如，语音识别、文本分类、机器翻译以及分词标注等等。RNN不能处理比较长的序列，会出现梯度消失问题，它本质上是一个BP网络，但与BP网络的区别是BP网络没有反馈回路，而RNN有反馈回路，它可以记住上一次的输出，两者都存在梯度消失现象。

如图下图所示，​是RNN的输入，A是RNN的一个节点，而​是输出。对这个RNN输入数据​​，然后通过网络计算并得到输出结果​​，再将某些信息传到网络的输出。

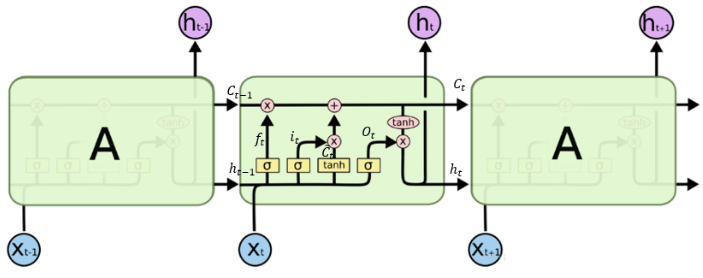


如果将RNN中的循环展开成一个个串联的结构，如下图所示，就可以更好地理解循环神经网络的结构了。RNN展开后，类似于有一系列输入x和一系列输出h的串联的普通神经网络，上一层的神经网络会传递信息给下一层。这种串联的结构非常适合时间序列数据的处理和分析。需要注意的是，展开后的每一个层级的神经网络，其参数都是相同的，并不需要训练成百上千层神经网络的参数，只需要训练一层RNN的参数，这里的共享参数和卷积神经网络的权值共享类似。RNN虽然被设计成可以处理整个时间序列信息，但其记忆最深的还是最后输入的一些信号，而更早之前的信号的强度则越来越低，最后只能起到一点辅助作用，甚至不起作用。



**LSTM**

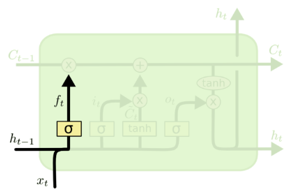
长短期记忆（Long Short Term Memory，LSTM）是一种特殊的RNN，主要是为了解决长序列训练过程中的梯度消失和梯度爆炸问题，简单来说，LSTM能够在更长的序列中有更好的表现。



LSTM的结构

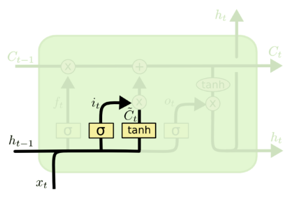
**LSTM内部三个阶段**

第一阶段：忘记门，决定哪些信息需要从细胞状态中被忘记，对上一个节点传进来的输入进行选择性忘记。忘记门是以上一层的输出​​和本层要输入的序列​​作为输入，通过激活函数，得到输出​​。

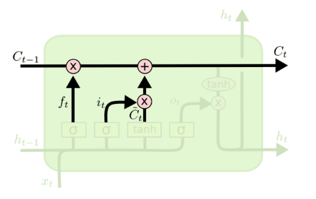


忘记门

第二阶段：输入门，输入门确定哪些信息能够被存放到细胞状态中，包括两个部分，第一部分为，第二部分为​。

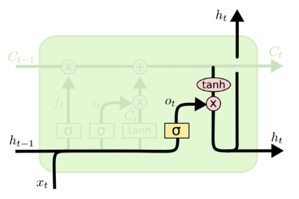


输出门



输出门

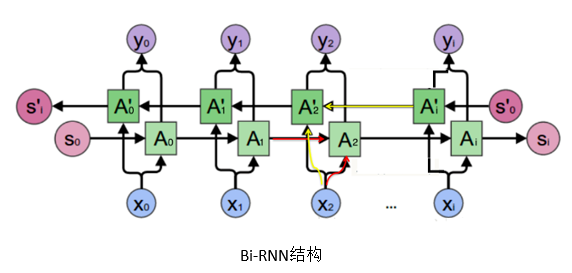
第三阶段：输出门，确定输出什么值，用来控制该层的细胞状态有多少被过滤。



输出门

**双向RNN**

双向RNN（Bidirectional Recurrent Neural Networks，Bi-RNN）的主要目的是增加RNN可利用的信息，它可以同时使用时序数据中某个输入的历史及未来数据，将时序方向相反的两个循环神经网络连接到同一个输出，通过这种结构，输出层就可以同时获取历史和未来的信息，也就是说，Bi-RNN由两个普通的RNN所组成，一个正向的RNN，利用过去的信息，一个逆向的RNN，利用未来的信息，这样在t时刻，既能够使用t-1时刻的信息，右可以使用t+1时刻的信息。一般来说，由于双向RNN能够同时利用过去时刻的信息和未来时刻的信息，会比单向RNN最终的预测更加准确。



**GRU**

门控循环单元（Gated Recurrent Unit，GRU）是循环神经网络的一种，和LSTM一样，都是为了解决长期记忆和反向传播中的梯度等问题而设计出来的。相比LSTM，GRU的结构更加简单，比LSTM少了一个门，因此计算效率更高，同时占用的内存也相对较少。在实际使用中，LSTM和GRU的差异不大，一般最后的准确率等指标都近似，但相对来说，GRU达到收敛状态时所需的迭代次数少，训练速度快。

GRU的结构包括两个门，一个重置门和更新门。候选隐含状态使用重置门来控制t-1时刻信息的输入，如果的结果为0，那么上一个隐含状态的输出信息将被丢弃。重置门决定过去有多少信息被忘记，有利于捕捉时序数据短期的依赖关系。候选状态使用更新门对上一时刻隐含状态和候选隐含状态进行更新。更新门控制过去的隐含状态在当前时刻的重要性，如果更新门一直趋近于1，t时刻之前的隐含状态将一直保存下去并全传递到t时刻，更新门有利于捕捉时序数据的中长期依赖关系。

