* 神经网络的发展过程
* 感知机

拥有输入层、输出层和一个隐含层。输入的特征向量通过隐含层变换达到输出层，在输出层得到分类结果。

* 多层感知机（NN）

多层感知机，顾名思义，就是有多个隐含层的感知机。

多层感知机可以摆脱早期离散传输函数的束缚，使用sigmoid或tanh等连续函数模拟神经元对激励的响应，在训练算法上则使用Werbos发明的反向传播BP算法。

多层感知机解决了之前无法模拟异或逻辑的缺陷，同时更多的层数也让网络更能够刻画现实世界中的复杂情形。

多层感知机给我们带来的启示是，神经网络的层数直接决定了它对现实的刻画能力——利用每层更少的神经元拟合更加复杂的函数

* DNN（神经网络“具有深度”）

为了克服梯度消失，ReLU、maxout等传输函数代替了 sigmoid，形成了如今 DNN 的基本形式。

单从结构上来说，全连接的DNN和上图的多层感知机是没有任何区别的。

* CNN（卷积神经网络）

对于CNN来说，并不是所有上下层神经元都能直接相连，而是通过“卷积核”作为中介。同一个卷积核在所有图像内是共享的，图像通过卷积操作后仍然保留原先的位置关系。

在普通的全连接网络或CNN中，每层神经元的信号只能向上一层传播，样本的处理在各个时刻独立，因此又被成为前向神经网络(Feed-forward Neural Networks)。

* RNN（循环神经网络）

全连接的DNN还存在着另一个问题——无法对时间序列上的变化进行建模。对了适应这种需求，就出现了另一种神经网络结构——循环神经网络RNN。

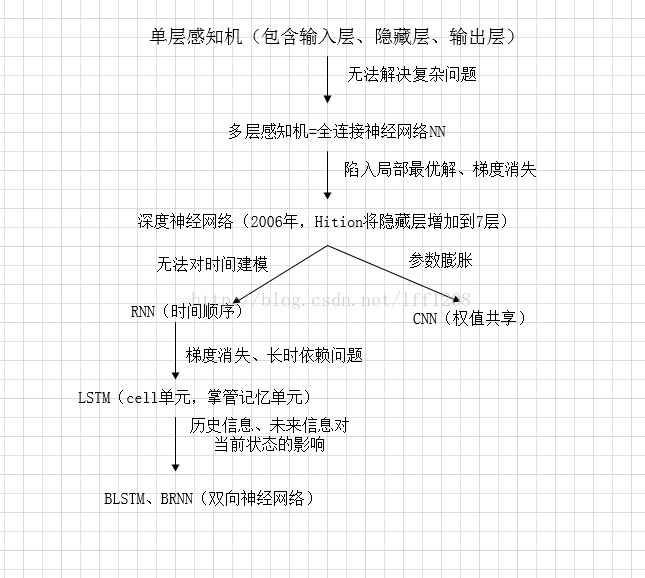
RNN可以看成一个在时间上传递的神经网络，它的深度是时间的长度。

* LSTM（长短期记忆网络）

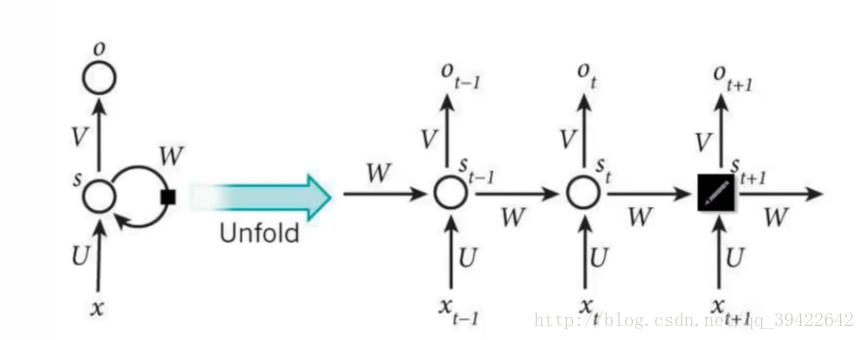
为了解决时间上的梯度消失，机器学习领域发展出了长短时记忆单元 LSTM，通过门的开关实现时间上记忆功能，并防止梯度消失。

* GRU

GRU是LSTM网络的一种效果很好的变体，它较LSTM网络的结构更加简单，而且效果也很好，因此也是当前非常流形的一种网络。GRU既然是LSTM的变体，因此也是可以解决RNN网络中的长依赖问题。



* RNN原理
* 神经网络包含输入层、隐层、输出层，通过激活函数控制输出，层与层之间通过权值连接。激活函数是事先确定好的，那么神经网络模型通过训练“学“到的东西就蕴含在“权值“中。
* 网络结构：



X\_t:表示t时刻的输入，o\_t:表示t时刻的输出，S\_t:表示t时刻的记忆

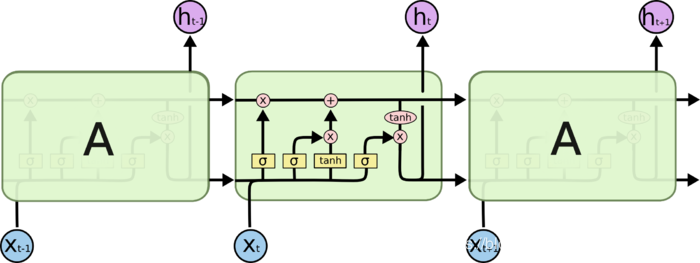
* 标准RNN的还有以下特点：

权值共享，图中的W全是相同的，U和V也一样。

每一个输入值都只与它本身的那条路线建立权连接，不会和别的神经元连接。

* RNN的标准结构，属于多输入多输出，并且输入与输出的个数是相同的，即每次输入都会对应一个输出。（一一对应关系）
* LSTM原理
* Long Short Term 网络—— 一般就叫做 LSTM ——是一种 RNN 特殊的类型，可以学习长期依赖信息。

LSTM 同样是这样的结构，但是重复的模块拥有一个不同的结构。不同于单一神经网络层，这里是有四个，以一种非常特殊的方式进行交互。整体上除了h在随时间流动，细胞状态c也在随时间流动，细胞状态c就代表着长期记忆。



* 遗忘门

LSTM 中的第一步是决定我们会从细胞状态中丢弃什么信息。这个决定通过一个称为遗忘门完成。

* 输入门

下一步是确定什么样的新信息被存放在细胞状态中。

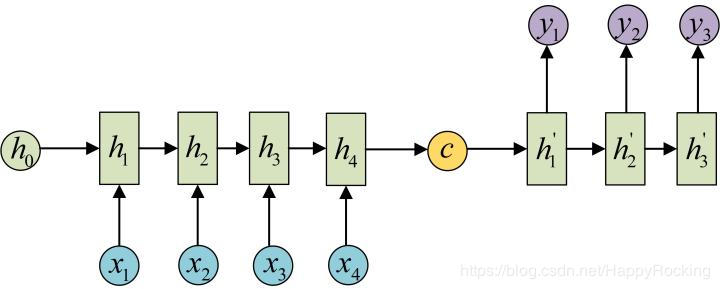
1. sigmoid 层称 “输入门层” 决定什么值我们将要更新。
2. 一个 tanh 层创建一个新的候选值向量，C\_t会被加入到状态中。

* 输出门

1. 运行一个 sigmoid 层来确定细胞状态的哪个部分将输出出去。
2. 细胞状态通过 tanh 进行处理（得到一个在 -1 到 1 之间的值）并将它和 sigmoid 门的输出相乘

* Seq2Seq原理
* 多输入多输出的结构，其输入与输出并不是一一对应的。

这种结构又叫Encoder-Decoder模型，也可以称之为Seq2Seq模型。



1. RNN为什么会有梯度消失或爆炸，分别如何解决？

* 答：循环神经网络的设计初衷之一就是能够捕获长距离输入之间的依赖。
  + 由于预测的误差是沿着神经网络的每一层反向传播的，因此当雅克比矩阵的最大特征值大于1时，随着离输出越来越远，每层的梯度大小会呈指数增长，导致梯度爆炸；
  + 反之，若雅克比矩阵的最大特征值小于1，梯度的大小会呈指数缩小，产生梯度消失。
  + 梯度爆炸的问题可以通过梯度裁剪来缓解，即当梯度的范式大于某个给定值时，对梯度进行等比收缩。
  + 深度残差网络是对前馈神经网络的改进，通过残差学习的方式缓解了梯度消失的现象，从而使得我们能够学习到更深层的网络表示；而对于循环神经网络来说，长短时记忆模型（LSTM）及其变种门控循环单元（Gated recurrent unit，GRU） 等模型通过加入门控机制，很大程度上弥补了梯度消失所带来的损失。

1. LSTM有哪些门，作用是什么？

答：输入门i\_t 、遗忘门f\_t 以及输出门o\_t 三个门和一个内部记忆单元c\_t 。

* 输入门

控制当前计算的新状态以多大程度更新到记忆单元中；

* 遗忘门

控制前一步记忆单元中的信息有多大程度被遗忘掉；

* 输出门

控制当前的输出有多大程度上取决于当前的记忆单元。

1. Seq2Seq模型引入注意力机制是为了解决什么问题？

答：Seq2Seq模型的核心思想是，通过深度神经网络将一个作为输入的序列映射为一个作为输出的序列，这一过程由编码输入与解码输出两个环节构成。在经典的实现中，编码器和解码器各由一个循环神经网络构成，既可以选择传统循环神经网络结构，也可以使用长短期记忆模型、门控循环单元等。在Seq2Seq模型中，两个循环神经网络是共同训练的。

* Seq2Seq模型中引入注意力机制解决的问题：

Seq2Seq模型的输出序列中，常常会损失部分输入序列的信息，这是因为在解码时，当前词及对应的源语言词的上下文信息和位置信息在编解码过程中丢失了。

* 直观的理解

在生成一个输出词时，会考虑每一个输入词和当前输出词的对齐关系，对齐越好的词，会有越大的权重，对生成当前输出词的影响也就越大。