

# 深入浅出RNN——循环神经网络

原创 ARAI RoboticAI 2020-08-03



深入浅出RNN——循环神经网络

作者：机器学习组 聂小兰

## 简·介

本文主要是对循环神经网络RNN的简要介绍，关于什么是RNN，RNN的主要结构有哪些以及RNN的主要应用场景。

## 目·录

- ▶ CNN之后, 为什么还有RNN?
- ▶ 基本循环神经网络
- ▶ 双向循环神经网络和深度循环神经网络
- ▶ RNN的应用举例
- ▶ 小结

## CNN之后, 为什么还有RNN?

在介绍RNN之前, 我们先想一想, 为什么CNN (卷积神经网络) 之后, 还会有RNN?

了解过CNN的同学 (不了解也没关系) 应该知道, CNN是单独的去处理一个个的输入, 也就是说, 前一个输入和后一个输入是完全没有关系的。但是, 某些任务需要能够更好的处理序列的信息, 即前面的输入和后面的输入是有关系的。

比如下面这句话:

我昨天上学迟到了, 老师批评了\_\_\_\_\_。

当我们想让机器去理解这句话意思时, 孤立的理解这句话的每个词是不够的, 需要处理这些词连接起来的整个序列;

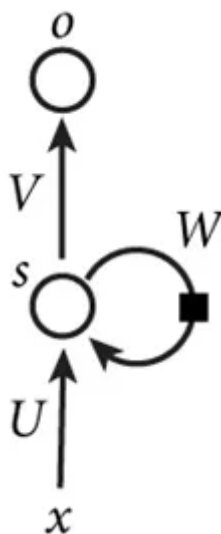
同样, 当我们处理视频的时候, 我们也不能只单独的去分析每一帧, 而要分析这些帧连接起来的整个序列。这时, 就需要用到深度学习领域中另一类非常重要神经网络: 循环神经网络(Recurrent Neural Network)。

概括来说: RNN是有“记忆”的

## 基本循环神经网络

为什么说RNN是有“记忆”的呢?

首先我们先来看一个简单的循环神经网络, 它由输入层、一个隐藏层和一个输出层组成:



对于刚入门的同学来说这张图可能有些难理解, 我们一点一点地看:

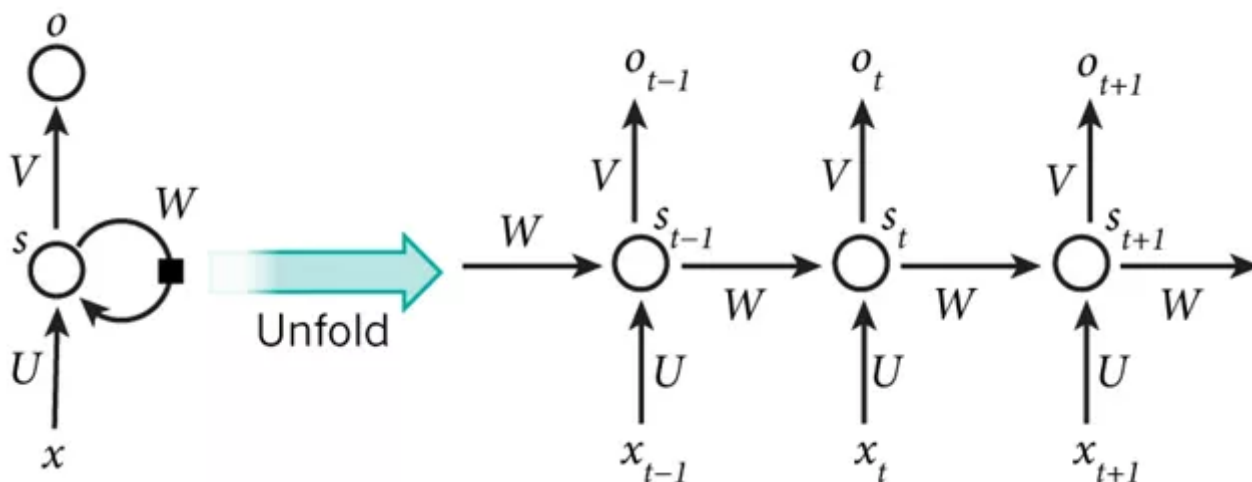
- 先把上面有W的那个带箭头的圈去掉, 它就变成了最普通的全连接神经网络。

- $x$  是一个向量，它表示输入层的值。（这里面没有画出来表示神经元节点的圆圈）
- $s$  是一个向量，它表示隐藏层的值。（这里隐藏层面画了一个节点，你也可以想象这一层其实是多个节点，节点数与向量  $s$  的维度相同）
- $o$  也是一个向量，它表示输出层的值
- $U$  是输入层到隐藏层的权重矩阵， $V$  是隐藏层到输出层的权重矩阵。

现在我们再来看看  $W$  是什么：

循环神经网络的隐藏层的值  $s$  不仅仅取决于当前这次的输入  $x$ ，还取决于上一次隐藏层的值  $s$ 。权重矩阵  $W$  就是隐藏层上一次的值作为这一次的输入的权重。

如果我们把这张图按时间线展开，循环神经网络也可以画成这样：



现在看上去就比较清楚了，这个网络在  $t$  时刻接收到输入  $x_t$  之后，隐藏层的值是  $s_t$ ，输出值是  $o_t$ 。关键一点是， **$s_t$  的值不仅仅取决于  $x_t$ ，还取决于  $s_{t-1}$** 。

我们可以用下面的公式来表示循环神经网络的计算方法：

用公式表示如下：

$$O_t = g(V \cdot S_t)$$

$$S_t = f(U \cdot X_t + W \cdot S_{t-1})$$

$s_t$  的值不仅仅取决于  $x_t$ ，还取决于  $s_{t-1}$

RNN公式

从上面可以看出，循环神经网络的输出值，是受前面历次输入值  $x_t$ 、 $x_{t-1}$ 、 $x_{t-2}$ 、 $x_{t-3}$ .....影响的，这就是为什么循环神经网络可以往前看任意多个输入值的原因。

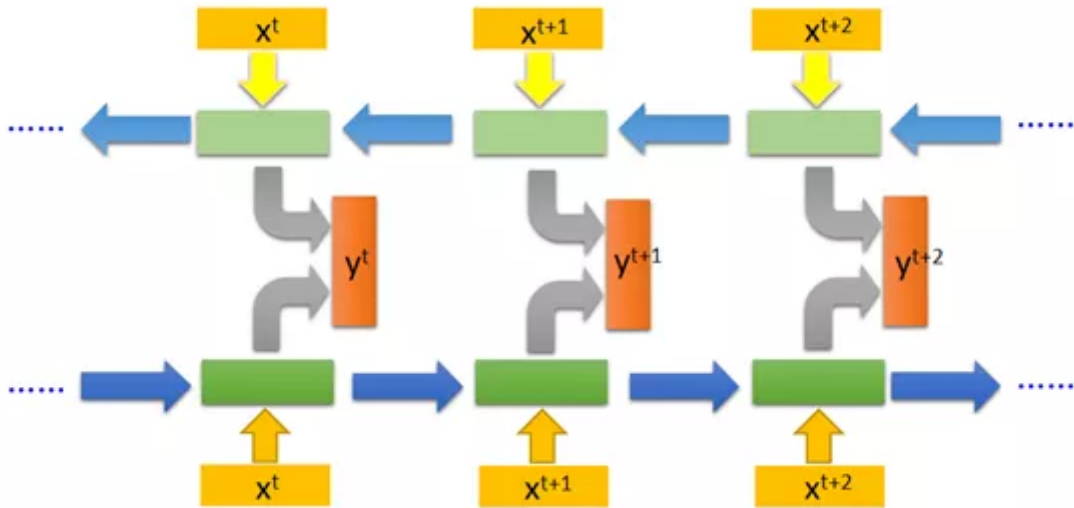
## 双向循环神经网络和深度循环神经网络

对于语言模型来说，很多时候光看前面的词是不够的，比如下面这句话：

我喜欢小明，可是小明\_\_\_\_小红

可以想象，如果我们只看横线前面的词，我喜欢小明，可是小明不喜欢我？可是小明有老婆了？这些都是无法确定的。但如果我们看到了横线之后的【小红】，那么横线上填【喜欢】这个词的几率就大得多了。

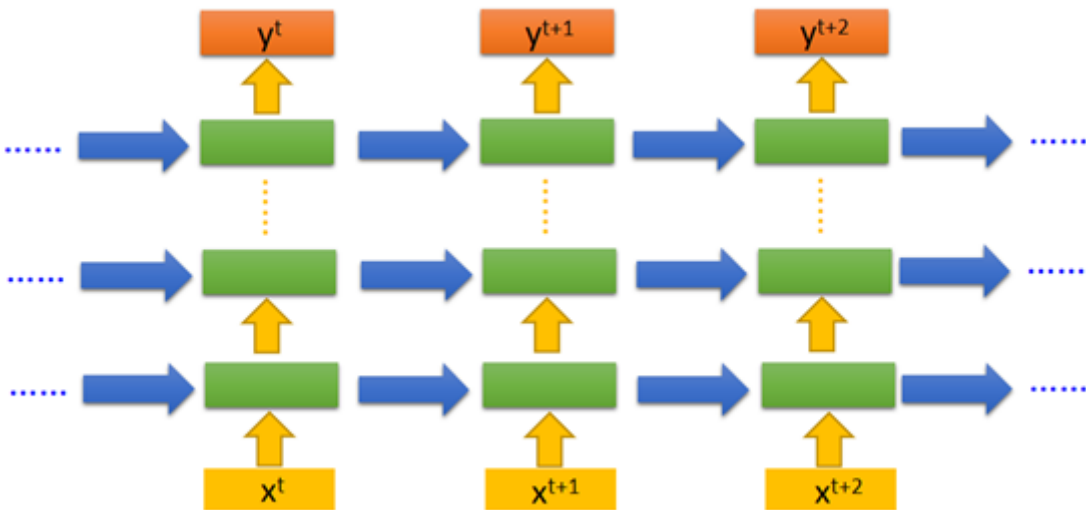
用上一小节中的基本循环神经网络是无法对此进行建模的，因此，我们需要双向循环神经网络，如下图所示：



【黄色】：输入层  
【绿色】：隐藏层  
【橘色】：输出层

从上图可以看出，双向循环神经网络的隐藏层要保存两个值，一个参与正向计算，另一个值参与反向计算。

除此之外还有深度循环神经网络，就是堆叠两个以上的隐藏层，如下图所示：



## RNN的实际应用非常广泛，主要有以下几个方面：

- 机器翻译：输入为源语言中的一个序列数据（比如：英语），输出是目标语言序列数据（比如：汉语），比如谷歌翻译、有道翻译等。
- 情感分类：输入为一段文本或者一段评论数据，输出相应的打分数据，来判别这段话的情感。
- 文本摘要：输入为一段文本序列，输出是这段文本序列的摘要序列
- 图像描述：输入为一张图片，输出是对这张图片的描述信息。
- 语音识别：输入为语音新号序列数据，输出为文本序列，比如 ASR 功能。
- 阅读理解：输入为一段文章和问题，对其分别编码，在对其进行解码，输出得到问题的答案。

## 小结

至此，我们讲完了基本的循环神经网络以及双向循环神经网络和深度循环神经网络。RNN比较烧脑，想要继续探索学习的读者可以观看李宏毅老师的《机器学习》系列课程。实际上，循环神经网络这个话题并没有完结。基本的循环神经网络存在梯度爆炸和梯度消失问题，并不能真正的处理好长距离的依赖（虽然有一些技巧可以减轻这些问题）。所以真正得到广泛的应用的是循环神经网络的一个变体：长短时记忆网络（LSTM）。它内部有一些特殊的结构，可以很好的处理长距离的依赖，感兴趣的读者欢迎关注我们的公众号继续学习~

## 往期回顾

- 01 | 哪个男孩不想拥有一个属于自己的机械狗呢？
- 02 | 我发现了海胆的秘密
- 03 | 看！电缆驱动4-DOF上肢康复机器人来了！
- 04 | 数据挖掘技术在智能水务中的应用探讨