

深度学习笔记 | 第10讲：序列模型：RNN循环神经网络

原创：louwill 狗熊会 2018-10-22



新朋友点蓝色字免费订阅



大家好！又到了每周一狗熊会的深度学习时间了。在前面的分享中，小编基本都围绕着CNN卷积神经网络的主题在进行着讲解，咱们了解了各种CNN网络结构，对于图像分类、检测和分割都有了较为全面的了解。但这些都是针对图像任务的。深度学习的应用领域还包括语音识别和自然语言处理等重要领域，所以从本讲开始，小编将和大家介绍一种新的神经网络结构——循环神经网络（Recurrent Neural Network）。一般来说，RNN 有两种表现形态，即循环和递归，分别指时间递归和结构递归，本节小编将重点以时间递归网络进行介绍。



——1——

计算机视觉之外：从语音识别到自然语言处理

在之前的笔记中，小编花费了五到六讲的篇幅来介绍卷积神经网络，对深度学习在图像识别和计算机视觉领域的应用有了一个基本的了解。所以总的来说，卷积神经网络是致力于如何让机器能看懂图像这样一个任务。但是咱们天天宣传人工智能，机器仅仅会看可还远远不够。跟咱们人类相比，除了会看之外，机器

还必须得会听、会说，最好就是还能懂咱们人类的语言，这才是实现了人工智能的第一步。

咱们先来看看机器如何听的问题，也就是深度学习在语音识别方面的应用。语音识别应该是日常生活中比较常见的深度学习应用了，比如说苹果的siri，阿里的天猫精灵智能音箱等等，大家可以轻而易举的生成一段语音数据，siri收到你的语音信号后，通过内置的模型和算法将你的语音转化为文本，并根据你的语音指令给出反馈。那么语音识别这么高级的技术适用于深度学习方法吗？当然可以。相较于图像三维矩阵的存在形式，我们先来看看语音在计算机中是以何种形态呈现的。



语音通常是由音频信号构成的，而音频信号本身又是以声波的形式进行传递的，一段语音的波形通常是一种时序状态，也就是说音频是按照时间顺序播放的。通过一些预处理和转换技术，我们可以将声波转换为更小的声音单元，即音频块。所以在语音识别的深度学习模型中，我们的输入就是原始的语音片段经过预处理之后的一个个音频块，这样的音频块是以序列形式存在的，所以输入是一个序列。那么输出呢？也就是咱们语音识别的结果是啥。语音识别的结果通常以一段文字的形式呈现，比如siri会快速识别出你的语音指令并将识别结果以文字形式打印在手机屏幕上。这段文字也是一个按顺序排列的文本序列，所以我们的输出也是一个序列。那么如何建立由序列输入到序列输出之间的有监督机器学习模型呢？这便是本节的主题——RNN循环神经网络的主要任务。

能让机器会看懂图像、听懂语音还不够，最好还得能理解咱们人类的语言。所谓自然语言处理，就是让计算机具备处理、理解和运用人类语言的能力。咱们总不能天天说人工智能，可最后却只得到一个能看能听却不识字的文盲吧？那你可能又要问了，咱们人类语言那么多种，除了各国语言之外，即使是一个国家、

一个省份甚至一个地区大家说的话都不一样，我们自己都谈不上能充分理解人类语言，更何况去让机器理解？话虽如此，但在基于深度学习的自然语言处理上，我们确实能够做到一定程度的让机器理解我们的语言。



没有语言，我们的思维就无从谈起，那么对于机器来说，没有语言，人工智能永远都不够智能。所以从这个角度来说，自然语言处理代表了深度学习最高任务境界。这一下子到了最高境界，咱们的非线性回归分析框架还能用来分析这高大上的自然语言处理吗？必须能。我们以自然语言处理的一个应用案例——机器翻译来分析一下，看看基于深度学习自然语言处理问题是如何被规范为一个从输入到输出的有监督机器学习问题的。

相信不少朋友都用过机器翻译，谷歌翻译、百度翻译、有道翻译，可供选择的工具就有很多，将大段的英文复制粘贴到谷歌翻译中直接机翻的经历应该很多人都有过。在这样的一个问题里，我们的输入毫无疑问就是一段待翻译的中文、英文或者是任意国家的文字，总的来说输入是由一个个单词或者文字组成的序列文本。那么作为翻译的结果，输出也是一个个单词或者文字组成的序列文本，只不过换了一种语言，所以在机器翻译这样一个自然语言处理问题中，研究的关键在于如何构建一个深度学习模型来将输入语言转化为输出语言。可以看到，这个问题跟前面语音识别的例子很像，它们的输入输出形式都是序列化的。针对这样的序列建模问题，深度学习给出的网络方案和语音识别一样，都是循环神经网络。稍后小编回和大家学习 RNN 的基本结构和细节知识。

当然，对于博大精深的自然语言处理来说，机器翻译还仅仅是一个小的方向，除此之外，自然语言处理还包括很多有趣的研究与应用方向：句法语义分析、文本挖掘、信息检索、问答系统等等。但是不管是哪个方向的应用，只要它是属于监督机器学习性质的深度学习问题，我们都可以将其归纳为一个从输入到输出的有监督机器学习问题。

2

RNN：网络架构与技术

相较于 DNN 和 CNN，RNN 网络结构有什么特别之处？它与前两者又有哪些不一样的结构设计？在对 RNN 的结构进行深入了解之前，我们必须对使用 RNN 面临的问题场景进行梳理。假设我们在进行语音识别时，给定了一个输入音频片段 X，要求我们输出一个文本片段 Y，其中输入 X 是一个按照时间播放的音频片段，Y 是一个按照顺序排列的单词组成的一句话，所以在 RNN 中我们的输入输出都是序列性质的。针对这样的输入输出 (X,Y) 的有监督学习，最适合的神经网络结构就是循环神经网络。为什么循环神经网络就最适用这种场景？在正式介绍 RNN 前，我们先来看下对于序列问题使用常规的神经网络看看会有什么问题。

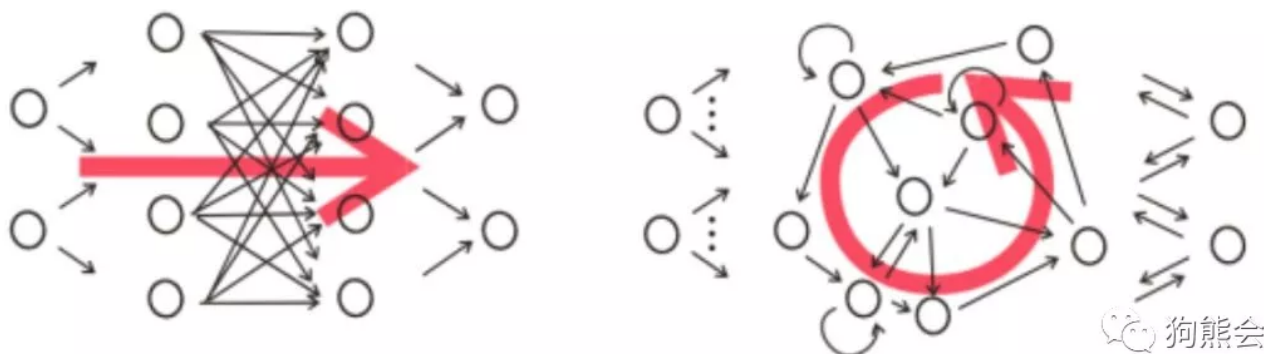
假设我们现在需要对输入的一段话识别其中每个单词是否是人名，即输入是一段文本序列，输出是一个每个单词是否是人名的序列。假设这段话有9个单词，我们将其转化为 9 个 one-hot 向量输入到标准神经网络中去，经过一些隐藏层和激活函数得到最终 9 个值为 0/1 的输出。但这样做的问题有两个：

一是输入输出的长度是否相等以及输入大小不固定的问题。在语音识别问题中，输入音频序列和输出文本序列很少情况下是长度相等的，普通网络难以处理这种问题。

二是普通神经网络结构不能共享从文本不同位置学到的特征，简单来说就是如果神经网络已经从位置 1 学到了 louwill 是一个人名，那么如果 louwill 出现在其他位置，神经网络就可以自动识别到它就是已

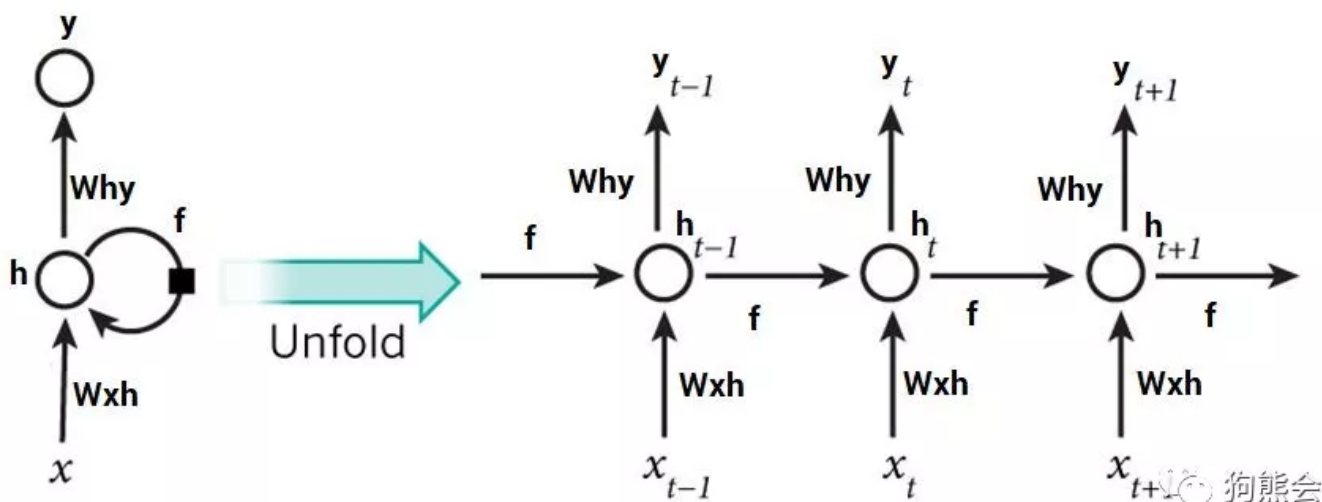
经学习过的人名，这种共享可以减少训练参数和提高网络效率，普通网络不能达到这样的目的。

所以直观上看，普通神经网络和循环神经网络的区别如下图所示：



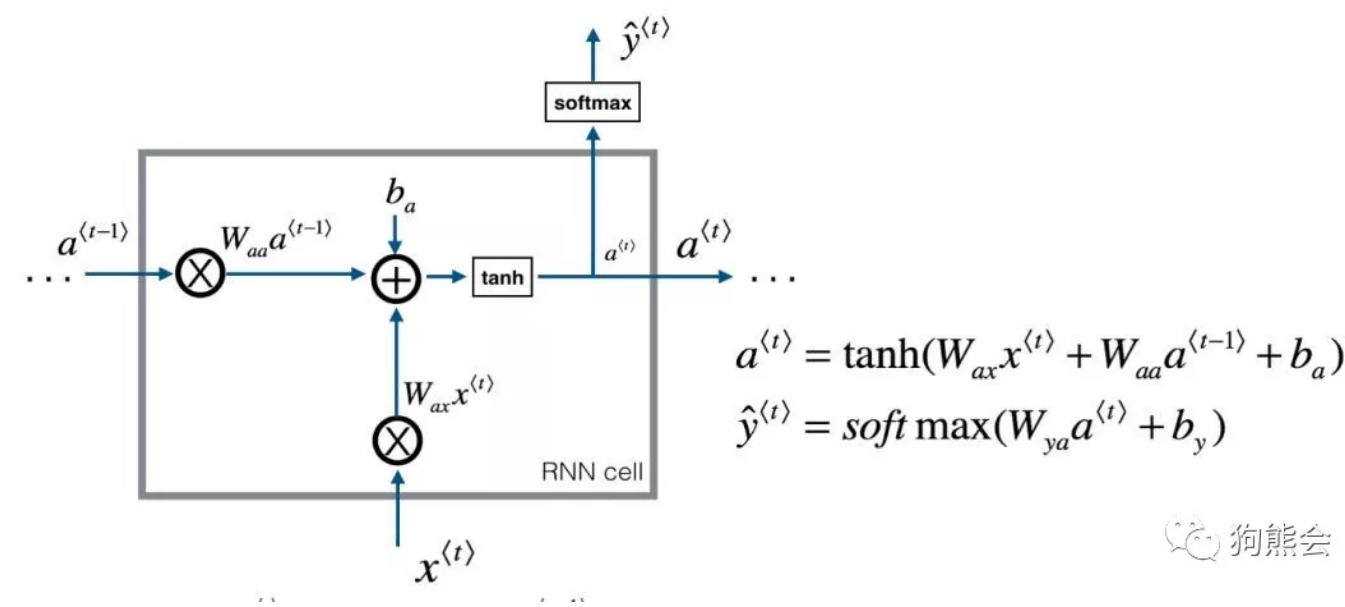
那么 RNN 到底长什么样子呢？

假设我们将一个句子输入 RNN，第一个输入的单词就是 x_1 ，我们将 x_1 输入到神经网络，经过隐状态得到输出判断其是否为人名，即输出为 y_1 。同时网络初始化隐状态激活值，并在隐状态中结合输入 x_1 进行激活计算传入到下一个时间步。当输入第二个单词 x_2 的时候，除了使用 x_2 预测输出 y_2 之外，当前时间步的激活函数会基于上一个时间步的进行激活计算，即第二个时间步利用了第一个时间步的信息。这便是循环（Recurrent）的含义。如此下去，一直到网络在最后一个时间步输出 y_n 和激活值 a_n 。所以在每一个时间步中，RNN 传递一个激活值到下一个时间步中用于计算。



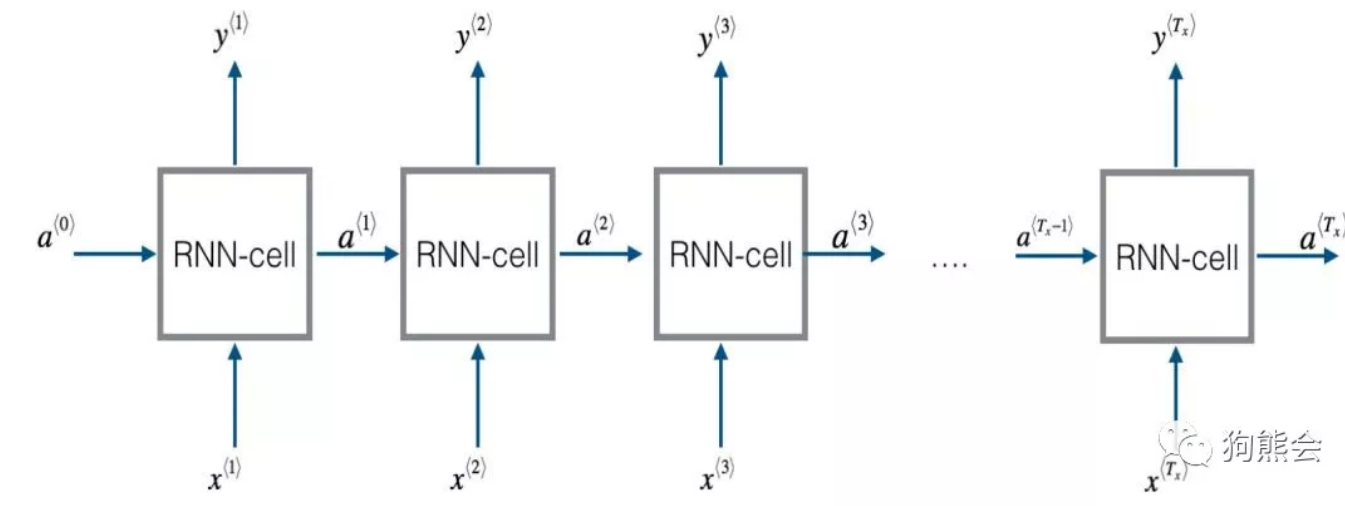
上图便是循环神经网络的基本结构。左边是一个统一的表现形式，右边则是左边的展开图解。在这样的循环神经网络中，当我们在预测 y_t 时，不仅要使用 x_t 的信息，还要使用 x_{t-1} 的信息，因为在横轴路径上的隐状态激活信息得以帮助我们预测 y_t 。

所以，RNN 单元结构通常需要两次计算，一次是结合上一个时间步的隐状态值和输入的计算，另一个是基于当前隐状态值的输出计算。一个 RNN 单元和两次计算如下图所示：



其中隐藏层的激活函数一般采用 `tanh`，而输入输出的激活函数一半使用 `sigmoid` 或者 `softmax` 函数。

当多个 RNN 单元组合到一起便是 RNN 结构：



这样的带有时间和记忆属性的神经网络模型使得深度学习可以胜任语音识别和自然语言处理等序列建模问题。当然，本节介绍的 RNN 结构还是最基础也是最经典的网络结构，在这个结构的基础上，RNN 针对输入输出序列的长度异同和记忆强度有各种各样的变种模型，基本如下：

- 一对多 (1 VS N)：根据类别生成对应音乐
- 多对一 (N VS 1)：文本情感分类

- 多对多（等长）（N VS N）：命名实体识别
- 多对多（不等长）（N VS M）：seq2seq，机器翻译、语言识别...

至于 RNN 更多的变体网络结构和应用，小编将在下一讲继续分享。



在本节内容中，小编和大家介绍了语音识别和自然语言处理这两类序列模型的基本特点和采用 RNN 作为网络结构的主要形式。对 RNN 的基本结构和原理进行了相对深入的探讨。在下一讲中，小编将继续分享关于 RNN 更深入的内容。



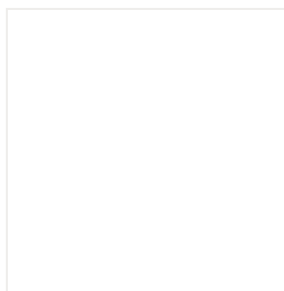
【参考资料】

deeplearningai.com

<https://zhuanlan.zhihu.com/p/22930328>

作者简介

鲁伟，狗熊会人才计划一期学员。目前在杭州某软件公司从事数据分析和深度学习相关的研究工作，研究方向为贝叶斯统计、计算机视觉和迁移学习。



识别二维码，查看作者更多精彩文章

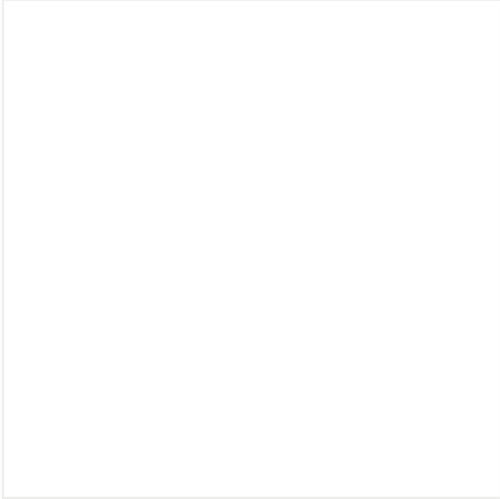
识别下方二维码成为狗熊会会员！

友情提示：

个人会员**不提供数据、代码，**

视频**only！**

个人会员网址：<http://teach.xiong99.com.cn>



点击“[阅读原文](#)”，成为狗熊会会员！

[阅读原文](#)