

144 | 基于深度学习的计算机视觉技术（二）：基本的深度学习模型

2018-09-03 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 07:40 大小 3.52M



在上一期的介绍里，我们讨论了以深度学习为背景的计算机视觉技术，重点讲解了为什么需要深度学习，特别是从传统模型的眼光来看深度学习模型的特点。

今天，我们来聊一聊[应用到图像上的一些最基本的深度学习模型](#)。

前馈神经网络

前馈神经网络（Feedforward Networks）应该算是最基本的神经网络架构了。这种架构是理解其他神经网络结构的基础。

我们可以从最基本的线性模型（Linear Model）入手，来理解前馈神经网络。线性模型说的是有一组输入 x ，然后有一个输出 y ，我们学习到一组向量，有的时候也叫作系数 w ，来

通过 x 预测 y 。这种线性模型可以算是最简单的机器学习模型了。在图像的情况下，输入往往是一个向量，输出也是一个向量，这个时候，我们需要学习的系数就从一个向量变为一个**矩阵**了。

那么，试想一下，如果我们把多层的线性模型进行叠加，能否得到多层的神经网络结构呢？答案是否定的。即便是多层的线性模型，只要每一层的变换是线性的，那么最后的结果一定也是线性的。因此，要想构建多层的非线性模型，每一层的变换也一定要是非线性的。

那么，如何在线性模型的基础上，我们只进行一些最小的改动，就能引入非线性的因素呢？

在这里，我们引入一个叫“**激活函数**”（Activation Function）的概念。直观地理解，激活函数就是在线性模型输出的基础上进行非线性变换。一个最简单的激活函数就是**Sigmoid 函数**，也就是把负无穷到正无穷的实数给映射到 $0 \sim 1$ 这个范围内。我们经常提到的对数几率回归，其实就是对这种变换的另一种称呼。

利用了 Sigmoid 激活函数的线性模型，本质上就是在做二元分类。在神经网络发展的早期，Sigmoid 激活函数是一种普遍使用的非线性变换方式。遗憾的是，在之后的发展中，研究人员发现了 Sigmoid 函数在数值稳定性上存在严重的问题。

具体来说，在很多机器学习的优化算法中，我们都需要依赖“梯度下降”（Gradient Descent）的方法来优化目标函数。在前馈神经网络中，梯度下降也是一种简单的优化神经网络并且学习到系数矩阵的方法。但是因为有 Sigmoid 函数的存在，计算的梯度有可能会溢出或者归零。在这样的情况下，模型就无法得到正常的学习。

为了解决 Sigmoid 激活函数的问题，研究人员发明了“**线性整流函数**”（Rectified Linear Unit），或简称**ReLU 函数**。和 Sigmoid 函数相比，ReLU 函数直接留下大于 0 的数值，把小于 0 的数值统统归 0。在 ReLU 函数的帮助下，我们能够更容易地训练多层的前馈神经网络。

有了非线性的转换之后，前馈神经网络往往可以把多个非线性的转换给叠加在一起，形成多层的结构。有了这个多层的转换之后，最后一层往往是把已经有的信息再映射到最终需要的输出上。这可以是一个回归问题，也可以是一个分类问题。

从我们之前提到的特征提取的角度来讲，前馈神经网络的中间层次就是利用最原始的信息来提取数据的特征，而最后一层可以当作是我们之前讲过的线性模型层。只不过和手动构造复

杂特征有一个不同的地方，前馈神经网络是自动学习这样的特征。

卷积神经网络

了解了最基本的前馈神经网络之后，我们来看一看它是怎么应用到图像处理中的。我们在之前的计算机视觉基础知识中讲到过，一种最直观的表达图像数据的方法就是**把图像看成矩阵数据**。

比如，有一个长 32 像素、宽 32 像素并且有 3 个颜色通道 (Channel)，或者简称为“32 乘 32 乘 3”的图像，如果我们采用前馈神经网络来对这个图像进行建模，需要学习的系数或者说权重是多少呢？就是把这 3 个数乘起来，一共有 3072 个系数需要学习。如果说这还是一种可以接受的方案的话，那么一个长宽为 200 像素，也是 3 个颜色通道的图片就需要 12 万个系数。很显然，直接采用前馈神经网络来表达图像信息需要大量待学习的参数。那么，我们在这里就需要有一种方法能够更加简洁地表达数据。

卷积神经网络 (Convolutional Neural Networks)，简称为 CNN，就是来解决这类问题的一种神经网络架构。其实，CNN 最初就是专门为视觉问题而提出的。那么，如何利用 CNN 来解决这个问题呢？

首先，**卷积神经网络试图用向量来描述一个矩阵的信息**。从工具的角度来讲，卷积神经网络利用两个特殊的架构来对图像数据进行总结，一个叫“卷积层” (Convolutional Layer)，一个叫“池化层” (Pooling Layer)。

在简单理解卷积层和池化层之前，我们先来看一下卷积神经网络是采用怎样的架构来处理图像的。一般来说，图像的输入要先经过一个卷积层，再经过一个线性整流函数，然后经过一个池化层，再经过一个全联通层（也就是前馈神经网络），最后得到一个向量的表达。需要注意的是**卷积神经网络的优势就是直接处理 3 维数据，进行 3 维的变换**。

现在我们来看卷积层是如何工作的。卷积层直接作用在 3 维的输入上，利用另外一个 3 维的“过滤器” (Filter) 来对原始的数据进行卷积操作。例如对于一个“32 乘 32 乘 3”的图像来说，我们可以使用一个“5 乘 5 乘 3”的过滤器来对其进行卷积。

在这里，我不展开来讲卷积操作，如果你有兴趣可以在网上找到卷积操作的数学定义。在这里，我们只需要理解**卷积操作是利用一个小的过滤器来对原始的更大的图像进行函数变换，从而希望能够提取图像的局部特征**。

在刚才这个例子中可以看到，因为有了卷积层，我们就不需要针对原始大小的图像进行表达了，而仅仅需要学习过滤器上的参数，这种方法就大大减少了参数的数目。

池化层的目的是对数据进行进一步的高度总结和概括。**对于一个矩阵来说，池化层针对某一个矩阵的局部，采用平均值、最大值来总结这个区域的矩阵数值。**例如，我们有一个“4 乘 4”，一共 16 个单元的矩阵，如果我们针对每个“2 乘 2”的区域加以最大值（Max Pooling）池化，那么我们就可以把 16 个单元的数据总结为“2 乘 2”，也就是一共 4 个单元的数据。每个单元是原来矩阵中“2 乘 2”区域中的最大值。

当一个图像经过了卷积和池化等一系列的操作以后，我们就说已经提取了这个图像的关键特征。这个时候，我们往往会把数据经过基本的前馈神经网络来进一步融合，最后能够完整地总结数据信息。在前馈神经网络之后，这就又是一个线性的决策层，可以是回归，也可以是分类。

小结

今天我为你讲了基于深度学习的计算机视觉技术的第二篇分享，帮助你对深度学习的基本模型，包括前馈神经网络和卷积神经网络有一个基本的认识。

一起来回顾下要点：第一，我们从线性模型入手讨论了前馈神经网络架构，重点介绍了激活函数和线性整流函数；第二，我们讲了在前馈神经网络基础之上的卷积神经网络，核心是用向量来描述一个矩阵的信息。

最后，给你留一个思考题，卷积神经网络中池化层的必要性是什么？为什么需要总结周围的单元到一个单元？这难道不是丢失了很多信息吗？

欢迎你给我留言，和我一起讨论。


AI 技术内参

你的360度人工智能信息助理

洪亮劼

Etsy 数据科学主管
前雅虎研究院资深科学家



新版升级：点击「 请朋友读」，10位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 143 | 基于深度学习的计算机视觉技术（一）：神经网络入门

下一篇 145 | 基于深度学习的计算机视觉技术（三）：深度学习模型的优化

精选留言 (3)

 写留言



郁郁达

2018-09-09

 2

池化从一定程度上讲，有防止过拟合的作用

展开 ∨



医用地狗精...

2018-09-05

 2

池化是为了让学习不过于注重不必要讯息，比如说辨认一张脸，知道有眼睛就可以了，不需要知道眼睛是在图的哪一个地方。通过池化还可以降低计算成本，提高效率。

展开 ∨



郁郁达

2018-09-09



池化从一定程度上说，起到防止过拟合的作用

展开 ∨