基于 TensorFlow 卷积神经网络的猫类图像识别

李昂

2017 年 8 月 24 日

摘要

这里是摘要。

关键词： 卷积神经网络; 深度学习; 图像识别

1 题目重述

给定一个已经标记有“猫”和“非猫”的训练数据集（Training Dataset）以及一个结构类似的测试数据集（Testing Dataset），建立一个简单的图像识别模型来准确地将猫的图片与其他的图片进行分类。每个图像都具有固定宽度与长度、以 RGB 值来表示每个像素。

2 符号说明

3 基本假设

4 问题分析

题目中 Training Set 给出了三组数据，除 list_classes 为介绍说明之外，其余两组分别为训练图像与标签。其中 training_set_x 是一个 $209 \times 64 \times 64 \times 3$ 的四维向量空间，即 209 个图像中，每个图像均为 64×64 像素，每个像素由一组 RGB 值（8-bit, 0-255）表示。以下为单个图像降维之后的向量表示样例:

而且 Training Set 之中每个图像（第一维度的 209 个）都有 label（training_set_y）对应。根据如上的 Training Set 与 Test Set 的数据特征，我们引入卷积神经网络来提取并学习图像的特征。

5 模型建立

考虑到过于复杂的神经网络模型会导致时间开销过大，我们建立了两层卷积层（Convolution Layer，每层包含池化过程）、两层普通前馈层（Feed-Forward Layer）的神经网络，并定义卷积核为 5×5 大小，并使用 Python 语言与 TensorFlow 框架编写程序。

其中第一层卷积层导出 32 层卷积核，作为提取的特征的一部分输入到池化 / 下一层卷积过程中。第二层卷积层导出 64 层卷积核，并通过 reshape 过程将两层卷积 / 池化导出的 $5 \times 5 \times 64$ 卷积核展开，导入之后的 4096 个前馈神经元，并通过 2 层前馈神经网络得到最终结果。

5.1 卷积与池化

对于两个参与卷积运算的矩阵

$$A = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1m} \\ a_{21} & a_{22} & ... & a_{2m} \\ ... & ... & ... & ... \\ a_{m1} & a_{m2} & ... & a_{mm} \end{bmatrix}, B = \begin{bmatrix} b_{11} & b_{12} & ... & b_{1n} \\ b_{21} & b_{22} & ... & b_{2n} \\ ... & ... & ... & ... \\ b_{n1} & b_{n2} & ... & b_{nn} \end{bmatrix}$$

（满足 $m > n$），将卷积核 B 在二维平面上平移，并且卷积核的每个元素与被卷积图像对应位置相乘，再求和。

$$A * B = \begin{bmatrix} \sum_{i=n}^{m}\sum_{j=n}^{m} b_{ij}a_{i-n+1,j-n+1} & \sum_{i=n}^{m}\sum_{j=n}^{m} b_{ij}a_{i-n+1,j-n+2} & \cdots & \sum_{i=n}^{m}\sum_{j=n}^{m} b_{ij}a_{i-n+1,j} \\ \sum_{i=n}^{m}\sum_{j=n}^{m} b_{ij}a_{i-n+2,j-n+1} & \sum_{i=n}^{m}\sum_{j=n}^{m} b_{ij}a_{i-n+2,j-n+2} & \cdots & \sum_{i=n}^{m}\sum_{j=n}^{m} b_{ij}a_{i-n+2,j} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{i=n}^{m}\sum_{j=n}^{m} b_{ij}a_{i,j-n+1} & \sum_{i=n}^{m}\sum_{j=n}^{m} b_{ij}a_{i,j-n+2} & \cdots & \sum_{i=n}^{m}\sum_{j=n}^{m} b_{ij}a_{ij} \end{bmatrix}$$

卷积运算常见于计算机视觉、图像处理等方面，由于具备将图像信息压缩的能力，在深度学习、图像分类中常用于提取图像特征。但是，对于题目中 64×64 像素的 RGB 图像，若通过学习得到了 400 个压缩成 32×32 大小的图像特征，那么每一个图像特征与原图像卷积都会得到 $(64 - 32 + 1)^2$ 的图像特征，而且只是单个特征的计算量；并且庞大的运算容易失去控制而导致过拟合（overfitting）。所以我们引入池化（pooling）过程，对于不同位置的特征进行统计，降低特征维度，同时不影响卷积过程中产生的新的特征信息。

5.2 前馈神经网络

人工神经网络（Artificial Neural Network）发展自人类对于动物神经系统工作方式、学习与记忆方式的观察。科学家最早提出的神经网络模型被称为感知机（Perceptron），由多个输入数据 $a_1, a_2, ..., a_n$、对应多个权重值 $\omega_1, \omega_2, ..., \omega_n$、一个输出阈值 t 与一个输出数据 b 组成。

$$b = \begin{cases} 1, & \omega \cdot a \geqslant t, \\ 0, & \omega \cdot a < t \end{cases}$$

ω（权重）通过人工给定初始值，在遍历训练集的过程中，通过指定增减学习率（learning rate）、不断修正错误（b 导出值与实际不符）来调整 ω。更复杂的前馈神经网络具有 1 到 2 层的隐藏层（hidden layer），用于处理高维输入值、提高模型准确率，同时允许多个输出值。

同时，为了规避单一感知机的线性不可分问题（TODO 此处应有论文引用），隐藏层、输出层的神经元在接受上一层输入后，会通过特定的激活函数（Activate Function）将线性的向量内积转化为非线性的空间划分。我们采用了 Sigmoid 函数

$$y = \frac{1}{1 + e^{\omega \cdot a}}$$

作为该神经元的输出。

在解决这一问题的过程中，我们采用了 2 层前馈网络。第一层为输入层（在整个网络中是隐藏层），接收上一层卷积 / 池化层输入的特征；第二层则为整个网络中最终的输出层，通过一个神经元输出最终的 0 和 1。

6 模型求解

随机初始化路径权值 $\omega = (\omega_1, ..., \omega_n)$，并且服从 $\mu = 0.2, \sigma = 0.1$ 的正态分布。遍历训练数据集（train_catvnoncat.h5 文件中的 train_set_x、train_set_y）部分。对于单个图像 A_m，首先将其与预设置的 5×5 卷积核 B 进行步长为 1 的卷积，得到该层输出

$$A * B = C_{62\times62}^{(1,1)}$$

，其中 C 为该卷积层输出 32 层的其中一层；进入第一层池化进行最大池化方案（max-pooling）：

$$\mathcal{A}^{(1)}(C_{62\times62}^{(1,1)}) = C_{32\times32}^{(1,2)}$$

第二层卷积 / 池化类似：

$$C_{32\times32}^{(1,2)} * B = C_{30\times30}^{(2,1)}$$

$$\mathcal{A}^{(2)}(C_{30\times30}^{(2,1)}) = C_{16\times16}^{(2,2)}$$

，其中 C 为该卷积层输出 64 层中的其中一层。至此完成卷积、数据提取过程。

将 64 层 $C_{16\times16}^{(2,2)}$ 展开为一维向量

$$\mathcal{B}(64C_{16\times16}^{(2,2)}) = c_{16384\times1}^{(3)}$$

，其中 $16384 = 64 \times 16 \times 16$；并且使 $c^{(3)}$ 各维度与 4096 个前馈神经元全连接。对于单个该层神经元 $K^{(1)}$ 则会产生输出

$$K^{(1)} = \frac{1}{1 + e^{\omega^{(3)} \cdot c^{(3)}}}$$

，最终输出层则为

$$K^{(2)} = \frac{1}{1 + e^{\omega^{(4)} \cdot K^{(1)}}}$$

。实际运行中，以上过程对于单一数据集进行了 15 次遍历（TODO 引用），以在防止过拟合的情况下获得最佳结果。

7 模型检验

我们在以上计算机求解过程中，每次遍历完成后都会通过 test_catvnoncat.h5 文件中的数据集检验一次运行情况。15 次的全部运行情况如下：

8 模型评价与推广

附录

A 程序源码

带有运行结果的源码与部分互动界面可以通过 Jupyter Notebook 打开根文件夹中的 /code 文件夹内的 ipynb 文件来查看。

以下代码均为 Python 语言（Python 3）。

B 关于训练模型的说明

1. 受制于计算机运行能力（CPU Intel i5, 2.70GHz，超频 3.10GHz，无 GPU）带来的运行时间，模型最终确定为两层卷积与两层前馈的神经网络。