

机器学习纳米学位——猫狗大战

开题报告

张嘉

2018年3月28日

项目背景

猫狗大战 (Dogs Vs. Cats) 项本次项目是kaggle上的一个竞赛题目，目标是训练一个模型从给定的图片中分辨出是猫还是狗，这个是计算机视觉领域的一个问题。

深度学习是近十年来人工智能领域取得的最重要的突破之一。它在语音识别、自然语言处理、计算机视觉、图像与视频分析、多媒体等诸多领域都取得了巨大成功。现有的深度学习模型属于神经网络。神经网络的起源可追溯到20世纪40年代，曾经在八九十年代流行。神经网络试图通过模拟大脑认知的机理解决各种机器学习问题。1986年，鲁梅尔哈特(Rumelhart)、欣顿(Hinton)和威廉姆斯(Williams)在《自然》杂志发表了著名的反向传播算法用于训练神经网络，该算法直到今天仍被广泛应用。

深度学习在计算机视觉领域最具影响力的突破发生在2012年，欣顿的研究小组采用深度学习赢得了ImageNet图像分类比赛的冠军。排名第2到第4位的小组采用的都是传统的计算机视觉方法、手工设计的特征，他们之间准确率的差别不超过1%。欣顿研究小组的准确率超出第二名10%以上。这个结果在计算机视觉领域产生了极大的震动，引发了深度学习的热潮。

自此以后每年的ImageNet图像分类比赛都是神经网络夺得冠军

- 2012年冠军 AlexNet, top-5错误率16.4%，使用额外数据可达到15.3%，8层神经网络
- 2014年亚军 VGGNet, top-5错误率7.3%，19层神经网络)
- 2014年冠军 InceptionNet, top-5错误率6.7%，22层神经网络
- 2015年的冠军 ResNet, top-5错误率3.57%，152层神经网络

问题描述

项目需要识别出猫狗，本质上是二分类问题。对应于监督学习就是使用现有的标签的图片训练模型，完成训练后对没有标签的图片进行分类。因此也可以使用监督学习方法如SVM解决此问题。项目要求使用深度学习方法识别一张图片是猫还是狗，通过训练模型，任意一张测试的图片，模型总能将输入数据映射为是猫或者狗的概率。因此该问题是可量化的、可衡量、可复制的。

输入数据

输入数据来自 kaggle猫狗大战，一个包含两个zip文件，分别是train.zip和test.zip。其中train.zip用来训练模型，test.zip用来对训练出来的模型进行预测。

训练数据共有 25000 张图片，猫和狗各占一半，每张图片都带有类别标签。因为两个分类的数据量相同，所以不用担心。测试数据共有 12500 张图片。按照训练验证测试4:1的比例对数据进行划分。

在上述所有图片中，都是彩色图片都包含 RGB 三通道的信息，但是图片质量差异很大，因此在进行数据输入的时候需要进行数据预处理，将图片裁剪大小一致的图片。

数据大小不一致没有办法直接输入到神经网络中使用，需要进行resize，设置一个图片的size,偏大的图片需要裁剪，偏小的使用0填充。本次实验使用keras的ImageDataGenerator.flow_from_directory进行图片裁剪,其中参数target_size: 整数tuple,默认为(256, 256).图像将被resize成该尺寸。

解决方法

项目要求使用深度学习的方法解决问题，这里拟使用卷积神经网络（CNN）。卷积神经网络(Convolutional Neural Network, CNN)是深度学习技术中极具代表的网络结构之一，在图像处理领域取得了很大的成功，在国际标准的ImageNet数据集上，许多成功的模型都是基于CNN的。CNN相较于传统的图像处理算法的优点之一在于，避免了对图像复杂的前期预处理过程（提取人工特征等），可以直接输入原始图像。CNN网络对图片进行多次卷积层和池化层处理，在输出层给出两个节点并进行softmax计算得到两个类别各自的概率。

基准模型

使用基于keras的resnet，Xception，VGG16等网络模型去完成项目。在kaggle上，总共有1314只队伍参加了比赛，本项目的最低要求是 kaggle Public Leaderboard 前10%。所以需要最终的结果排在131位之前，131位的得分是0.06127，我们的结果小于这个就好。

评估指标

采用对数损失来衡量：

$$LogLoss = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

其中：

- n 是图片数量
- \hat{y}_i 是模型预测为狗的概率
- y_i 是类别标签，1 对应狗，0 对应猫
- $\log()$ 表示自然对数

对数损失越小，代表模型的性能越好。上述评估指标可用于评估该项目的解决方案以及基准模型。

设计大纲

数据预处理

- 从kaggle下载好图片
- 为keras.ImageDataGenerator准备数据，要求猫和狗在不同的文件夹以示分类
- 对图片进行裁剪，保持输入图片信息大小一致（不需要人工裁剪，设置图片大小使用keras.ImageDataGeneratorn能）
- 对训练数据进行随机偏移、转动等变换图像处理，这样可以尽可能让训练数据多样化

模型搭建

Kera的应用模块Application提供了带有预训练权重的Keras模型，这些模型可以用来进行预测、特征提取和微调整和。

- 使用ResNet50等现有的去掉了全连接层预训练模型
- 添加自己的全连接层到ResNet50网络

模型训练&模型调参

- 导入预训练的网络权重
- 冻结除了全连接成的所有层，获得bottleneck特征
- 尝试使用不同的优化器 adam,adadelat等对模型进行训练，选择最佳模型

模型评估

- 使用 $Logloss$ 进行模型评估,上传Kaggle判断是否符合标准

可视化

- 进行数据探索并且可视化原始数据
- 可视化模型训练过程的准确率曲线，损失函数曲线等

参考文献

【1】 《中国计算机学会通讯》 第8期 《专题》

【2】 [LaTeX 各种命令， 符号](#)

【3】 Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks

【4】 <http://keras-cn.readthedocs.io/en/latest/>

【5】 Oren Rippel, Jasper Snoek, Ryan P. Adams. Spectral Representations for Convolutional Neural Networks