

机器学习工程师纳米学位 – 猫狗大战

- 开题报告

- 张海鹰

- 2018.07.02

项目背景

猫狗大战是 kaggle 平台上的一个比赛项目，最终的要求是提供一个模型来识别图片中的对象是猫还是狗，所以抽象来看这是一个图像识别的问题。

图像识别是人工智能的一个重要领域，目前发展十分迅速，应用范围相当广泛，手写数字识别、邮政编码识别、汽车牌号识别、汉字识别、条形码识别，以及如人脸、指纹、虹膜识别等已经在人类日常生活中广泛应用，对经济、军事、文化及人们的日常生活产生重大影响。

支撑其应用实现的重要技术就是深度学习，深度学习被誉为通往人工智能的必经之路，在 2016 年 3 月 Google DeepMind 研发的 AlphaGo 4:1 战胜了世界冠军李世石后，深度学习已经成为现今最为火热的人工智能技术。

问题描述

项目要求的最终输出是辨别图片是猫还是狗，因此属于机器学习领域的分类问题，由于没有要求对狗和猫的品种再次细分，因此是二分类。具体的量化方法就是让模型输出该图片是猫和狗的概率。模型如何可以输出概率了？就是用深度学习来进行训练，让模型通过给定的训练数据，不断学习如何识别猫狗的特征。

数据输入

全部数据都由 kaggle 平台提供，分为了训练集和测试集；

训练集共有图片 25000 张，其中猫 12500 张，狗 12500 张；

测试集共有图片 12500 张；

我会在训练集中使用 k-fold 再次划分训练和验证数据集，来训练和验证模型，最后在测试集上测试模型，其中要注意的是不能对模型泄露测试集数据。

解决方案

深度学习在图像识别中目前最有代表性的就是卷积神经网络。

普通神经网络处理图像识别问题中，输入层的每一个神经元可能代表一个像素的灰度值。但这种神经网络用于图像识别有几个问题，一是没有考虑图像的空间结构，识别性能会受到限制；二是每相邻两层的神经元都是全连接，参数太多，训练速度受到限制。

而卷积神经网络就可以解决这些问题。卷积神经网络使用了针对图像识别的特殊结构，可以快速训练。因为速度快，使得采用多层神经网络变得容易，而多层结构在识别准确率上又很大优势。

因此本项目会采用卷积神经网络

评估标准

采用 log loss 作为模型评估标准;
其定义如下:

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

基准模型

使用 ResNet50, InceptionV3, Xception 作为基准模型, 期望能在项目中综合利用三个成熟模型的优势, 站在它们的肩膀上, 利用他们已经训练好的特征, 在此基础之上自己再训练出一个能够在猫狗这个细分领域更加优秀的模型。

项目设计

Step1: 数据预处理

- 简单查看图片, 思考是否要处理异常值, 以及如何处理
- 为了配合 keras 的 ImageDataGenerator, 将猫狗图片分别存储到 cat 和 dog 文件夹

Step2: 模型搭建

- 尝试一下 ResNet50, InceptionV3, Xception 这些网络的单独辨识情况
- 尝试在三种预训练模型后面训练自己的全连接层, 查看其辨识情况
- 尝试综合利用三种模型的特征向量, 然后训练出自己的模型

Step3: 模型训练

- 在训练集中划分出 20%保留为验证集, 然后进行训练和验证

Step4: 模型调参

- 尝试调整 learn rate, dropout 等参数

Step5: 模型评估

- 采用 log loss 作为模型评估标准

Step6: 可视化

- 把模型训练过程中每一个 epoch 的 acc, loss 用曲线图像展现, 这样可以直观的看见模型训练中的进展情况。