

机器学习纳米学位

猫狗大战——开题报告

2018 年 9 月 2 日

项目背景

如果计算机能够像人一样识别出图片中的物体，那我们就能让计算机帮助我们做更多的事情，所以人们一直致力于让计算机拥有图像识别的能力。近年来，深度学习，尤其是神经网络在图像处理这一方面获得了巨大的发展，例如利用卷积神经网络进行手写数字识别[1]。

本项目希望参考近年来的研究成果，使用 kaggle 的训练数据（已经标记好猫/狗标签的图片）在计算机上对深度学习网络模型进行训练，让模型学习到判别猫狗的能力，对任意一张猫或狗的图片，识别出是猫还是狗。

问题描述

本项目的问题在于如何构建一个模型，用带标签的数据进行训练，然后对无标签的数据进行二分类预测（区分猫还是狗），这是一个监督学习的过程。我们可以对任意一张图片，预测该图为某个分类的概率，然后根据概率判定该图是否为某个分类。所以这个过程是可量化，可测量，可重复的。

数据或输入

数据全部来自 kaggle，包含 25000 张训练图片（猫狗各有 12500 张，文件名以 cat 或 dog 为前缀，放在同一个 test.zip 里面），12500 张测试图片（文件名为数字，无法区分猫狗）。

图片都是 3 通道的彩色图片，但是尺寸不尽相同，针对不同的模型，要 `resize` 成相应的尺寸才能使用。本项目打算在图片预处理阶段，使用 `keras.preprocessing.image.ImageDataGenerator` 对图片进行加载，并 `resize` 成相应尺寸。

原始训练数据集会被进一步划分为训练数据集和验证数据集，比例大概为 9:1。

方案描述

本项目拟使用深度学习中的卷积神经网络（Convolutional Neural Network，简称 CNN）来解决问题。卷积神经网络在图像处理方面取得了巨大的成就，它基于“权值共享”的策略，复合多个“卷积层”和“采样层”来加工输入信号，然后在连接层实现与输出目标的多个映射。[2]

基准模型

图像识别在业界已经做了非常丰富的研究，前人也给出了很多优秀的模型，本项目打算综合尝试使用论文指引中提到的 VGGNet、ResNet、Inception v3 等模型。

项目要求是最低要达到 kaggle Public Leaderboard 前 10%，即排在第 $1314/10 = 131$ 名选手之前，其得分为 0.06127，即本项目最终得分要小于该数值。

评估标准

评估标准采用 kaggle 官方指定损失函数：

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

参数说明

- n 为测试数据集的图片数量
- \hat{y}_i 为一张图片预测为狗的概率
- y_i 为类别标签，1 为狗，0 为猫
- $\log()$ 为自然对数

最终函数值越小，结果越好，代表模型性能越好。

项目设计

本项目将使用 Keras（以 tensorflow 为后端）来完成——基于“迁移学习”的理念，利用 Keras 已经预训练的模型，从一个较高的起点来训练我们的猫狗识别模型。[3]

项目大致设计为以下步骤：

- 1 数据预处理
 - 1.1 从 kaggle 下载图片
 - 1.2 将训练数据中的图片根据文件名称前缀分成 2 部分，分别是狗的图片、猫的图片。
 - 1.3 对图片进行 resize

- 1.4 初步浏览图片，发现异常图片很少，人工检查难度太大，考虑使用预训练模型检测出不是猫狗的图片（需引入猫狗的所有种类）
- 2 搭建模型
 - 2.1 从 Keras 的 Application 引入预训练的模型（考虑使用 VGGNet、ResNet、Inception v3 等模型）
 - 2.2 冻结预训练模型除了最后的全连接层和输出层以外的所有层（即使用已有的权重）
 - 2.3 添加自己的全连接层和输出层到模型
- 3 模型调参
 - 3.1 尝试使用 adam 优化器、kaggle 官方指定的损失函数对模型进行训练和评估，选出最佳模型。
- 4 模型评估
 - 4.1 使用训练完的模型对测试数据集进行预测，上传 kaggle 进行打分，判断是否达标
- 5 尝试对项目过程的一些数据进行可视化

参考文献

- [1]LeCun et al., 1998
- [2]周志华. 机器学习[M]. 北京:清华大学出版社, 2016:113-114.
- [3]量子位. 一文看懂迁移学习：怎样用预训练模型搞定深度学习？
- [OL]. https://mp.weixin.qq.com/s/1LoNvo2IE_Zhya8Vvqk6_w. 20170702