

机器学习算法工程师纳米学位

毕业项目

陈正和 Sam Chen April 20st, 2019

猫狗大战项目

背景知识 Domain Background

本项目属于计算机视觉领域的范畴。早在1960年代，计算机视觉正式成为一门学科领域，当时的目标是自动化图像分析的过程：用计算机模拟人类的视觉系统，让计算机告诉我们它看到了什么。(Amir Jirbandey, A brief history of Computer Vision and AI Image Recognition, 2018)

到了2010年，得益于深度学习领域的发展与突破，我们图像识别被推到了全新的高度。其中不得不提到的是大型开源项目：[ImageNet](#) 数据库。ImageNet数据库，包含超过1千万张手动标记的图片，这些图片来自于1000个不同的图片类别。(Udacity,算法工程师纳米学位课程)2010年开始，ImageNet每年都会举办大规模图像识别大赛(ImageNet Large Scale Visual Recognition Competition)。在这个比赛中，不同的团队会协力创造准确率最高的图像识别CNN模型。(Udacity,算法工程师纳米学位课程)

其中的重大突破，包含2012年多伦多大学提出的AlexNet架构；2014年牛津大学的VGG Net 架构；2015年微软研究部门的ResNet架构等。(Udacity,算法工程师纳米学位课程)从每年胜出者的架构当中，我们可以学习到很多设计CNN架构的技巧，例如：使用了ReLU激活函数、Dropout等技巧，避免过拟合的问题；增加横跨多层的线路，解决过深的架构梯度消失的问题。(Udacity,算法工程师纳米学位课程)

问题描述 Problem Statement

本项目源自于Kaggle数据竞赛平台上，[猫狗大战图像识别项目](#)。项目包含了37,500张猫或者狗的图片，欲通过训练一个深度学习图像识别的模型，识别图片中的主体是狗的概率。(若模型识别出图片中的主体是猫，则概率为0；若模型识别出图片中的主体是狗，则概率为1。)一个潜在的解决方案，是以ImageNet的获胜者模型(VGG, ResNet等)为基底，通过迁移学习的方式，搭建模型。最后搭建一个函数，若识别出的前 n 个结果有 x 个为狗的品种之一，则这个图片为狗的概率为 x/n 。

数据集 Datasets and Inputs

本项目的数据集由Kaggle提供，其中包含25,000张训练集数据；12,500张测试集数据。数据集可以在[这个地址](#)获取。

图片可能存在几个问题：

1. 图片中未包含猫或者狗
2. 图片的尺寸不固定
3. 图片中的猫或者狗可能在图片比较边缘的位置

在实际执行项目的过程中，将会跟据上述几个问题进行处理。

解决方案 Solution Statement

这个问题，可以通过迁移学习解决。以ImageNet的获胜者模型(VGG, ResNet等)为基底，通过迁移学习的方式，去掉最后一层，并在其后加上池化层、Dense层等。迁移学习搭建的模型，可以识别图片中的主题为某个狗/猫品种的概率。在迁移学习的模型之后，加上一个函数，若识别出的前 n 个结果有 x 个为狗的品种之一，则这个图片为狗的概率为 x/n 。

模型标杆 Benchmark Model

这个项目中，在原始有效的比赛时间里，共有1314个提交。以Leaderboard前10%做为标杆，Log Loss值需要达到 0.06320以下。

评估指标 Evaluation Metrics

Kaggle的原始比赛中，提供了这个项目的评估指标 -- log loss值:

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

where n is the number of images in the test set \hat{y}_i is the predicted probability of the image being a dog y_i is 1 if the image is a dog, 0 if cat $\log()$ is the natural (base e) logarithm

项目设计 Project Design

1. 数据预处理

- 数据探索：
 - 基础信息探索：探索数据集中图片的数量等，数据的基本信息。
 - 可视化探索：随机检索某些图片，了解图片数据的分布情况。
- 数据异常处理：发现图片潜在的问题，针对异常的图片进行处理。例：删除图片中未包含猫/狗的数据，避免干扰模型。
- 数据增强：
 - 对图片进行不同的操作(例：旋转、翻转、平移)等，增加模型的泛化能力。

2. 模型搭建

- 迁移学习：使用Keras，导入欲进行迁移学习的模型(VGG, ResNet等)。
- 聚合函数：撰写一个函数，计算迁移学习的模型，预测结果前 n 个属于猫/狗的品种之一的概率为多少。

文章引用

- [Amir Jirbandey, A brief history of Computer Vision and AI Image Recognition, 2018](#)
- [Dogs vs. Cats Redux: Kernels Edition](#)