

机器学习纳米学位毕业项目

猫狗大战开题报告

Udacity 丁小刚

2018 年 03 月 18 日

一、项目背景

猫狗大战[1]最开始是 Kaggle 在 2013 年举办的一个关于图像识别的比赛，目的是尽可能准确地识别出图片中的内容是猫还是狗。在此之前有计算机视觉专家表示，人们所用方法的准确率一般都很难超过 60%。近年来以神经网络为代表的一系列机器学习方法使得该类问题的准确率有了大幅提升。尤其是深度卷积神经网络在图片识别方面的应用[2, 3]。得益于此，几年前很难解决的图片识别问题，现在可以得到比较好的结果了，在 Kaggle 的猫狗大战公开算法排名中，第一名的准确率已经达到了 98.533%。

二、问题描述

该问题的目的就是设计一个机器学习算法，使得电脑能够对图片中的猫狗进行准确的识别，属于一个非猫即狗的二分类问题。算法的准确率应该达到 Kaggle 猫狗大战公开排行榜的前 10%，也就是对于测试集中的猫狗图片，算法的 $\text{score}(\log \text{ loss})$ 应该低于 0.06127，并且能够推广到现实应用场景中。

三、输入数据

Kaggle 猫狗大战中已经提供了用于比赛的数据集，该数据集来自微软研究院，可以在 Kaggle 网站上注册后直接下载获得。

数据集分为两部分，一种是用来学习使用的训练数据集，另一种则是用来测试最终算法准确率的测试数据集。

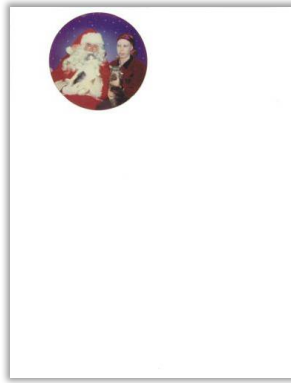


图 1. cat.10029. jpg

训练数据集中包含有 25000 张图片，全部为彩色 jpg 格式，图片尺寸不一致，大多为 300x500 不等。其中 12500 张图片名含有 cat，并且绝大部分确实是猫，发现其中一张图片 cat.10029. jpg 并不是猫，这应该属于误标的图片。剩余 12500 张图片名都包含有 dog，绝大部分都是狗。整体来看，这是一个已经标注了的猫狗图片集，并且包含有异常数据，即非猫非狗的图片。

测试数据集中包含有 12500 张图片，同样为彩色 jpg 格式，尺寸不统一，全部以数字命名。算法最终的准确率就要通过该数据集测试获得，也就是对于该数据集中的图片，算法的 $\text{score}(\log \text{ loss})$ 应该低于 0.06127。

由于训练数据集中包含有异常数据，所以这里准备采用基于 ImageNet 的预训练模型 Inception v4[4]、Xception[5]初步筛选出训练集中的非猫非狗图片，这些预训练模型已经能够分类 118 种狗和 7 种猫，最后将这些非猫非狗图片从训练集当中剔除，这样有利于算法最终准确率的提升。

四、 解决方法

首先，对数据进行预处理。先将所有图片缩放至统一尺寸，这里应该根据原始图片尺寸以及计算平台性能的考虑来选择尺寸，在此选择预训练模型的默认尺寸 299x299。然后利用预训练模型剔除非猫非狗的干扰图片。再对训练数据集进行划分，将训练数据集中的猫和狗分别随机排序，按照 10:1 的比例划分为训练集和验证集，生成训练猫集、验证猫集以及训练狗集、验证狗集。

其次，构建深度神经网络算法，利用猫狗训练集和验证集对算法进行训练，使得在验证集的表现达到预期。

最后，利用已经训练好的神经网络算法，在测试集上进行测试，得出最终的测试成绩，如果测试成绩未达到预期，则表明算法的泛化能力太差，整个神经网络需要重新设计和训练。

五、 评估标准

评估标准采用 Kaggle 上提供的交叉熵计算方法：

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

其中，n 表示测试集中所有图片的数量， \hat{y}_i 表示测试结果为狗的概率， y_i 的值在图片是狗时取 1，是猫时则取 0， $\log()$ 表示自然对数。

LogLoss 的值越小越好，最终的结果应该小于目标 0.06127。

六、 基准模型

这里选择 Kaggle 猫狗大战公开排行榜的前 10% 成绩为比较基准，也就是算法最终的 LogLoss 要小于 0.06127。

七、 项目设计

计划采用如下步骤实施项目：

1. 获取数据集，确保使用的数据集为 Kaggle 提供，未经过修改。
2. 数据预处理，包括图片缩放，异常图片筛选，数据混洗，训练集验证集的划分。
3. 构建深度卷积神经网络，基于预训练模型，准备使用迁移学习。以预训练模型的参数为初始值，将其中一部分参数设置为可训练状态，其余参数保持不变，去掉预训练模型最后的全连接层，用分类器将多个预训练模型组合在一起，将多个模型融合。
4. 在训练集和验证集上进行学习训练，并适当调整网络结构。
5. 在测试集上测试算法，在 Kaggle 上提交测试结果，获得 Kaggle 验证后的最终准确率。

参考文献

- [1] Kaggle. (2017, Mar. 3). *Dogs vs. Cats Redux: Kernels Edition*. Available: <https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition>
- [2] M. A. Nielsen. (2015). *Neural Networks and Deep Learning*. Available: <http://neuralnetworksanddeeplearning.com/index.html>
- [3] S. Mallat, "Understanding deep convolutional networks," *Philos Trans A Math Phys Eng Sci*, vol. 374, no. 2065, p. 20150203, Apr 13 2016.
- [4] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet

and the Impact of Residual Connections on Learning," *ArXiv e-prints*, vol. 1602, Accessed on: February 1, 2016 Available: <http://adsabs.harvard.edu/abs/2016arXiv160207261S>

- [5] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *ArXiv e-prints*, vol. 1610, Accessed on: October 1, 2016 Available: <http://adsabs.harvard.edu/abs/2016arXiv161002357C>