

Machine Learning Engineer Nanodegree

Capstone Proposal

俞伟山

August 29th, 2018

Proposal

猫狗大战 ([Dogs vs. Cats Redux: Kernels Edition](#)) 。

项目背景

此项目最早源于 `kaggle` 2013年的[Dogs vs. Cats](#)比赛。那时，网站为了防止恶意攻击，一般会提供一些验证问题，用来区别人和机器，即 [CAPTCHA](#)(Completely Automated Public Turing test to tell Computers and Humans Apart)。这些问题，要设计得容易让人解决，而让计算机不好解决。在那时，[Asirra](#) (Animal Species Image Recognition for Restricting Access)，就是人容易解决，而计算机不好解决的问题。

猫狗大战属于图像识别的问题，那时已经有了一些机器学习算法应用于图像识别，[Machine Learning Attacks Against the Asirra CAPTCHA](#) 使用机器学习算法，可以在猫狗的图像识别中，达到 `80%` 的分类准确率。

随着这几年机器学习的发展，特别是深度学习和图像分析的发展，各种深度学习框架、CNN模型相继出现。`kaggle` 于2017年，再次举办了猫狗大战的比赛，排在第一名的，`LogLoss` 得分达到了 `0.03302` 。

图像识别问题，属于 `计算机视觉` 领域。`计算机视觉` 是一个跨学科领域，它处理计算机如何高度理解数字图像或视频的问题。它包含如何自动从图像和视频中提取、分析和理解一些有用的信息。

图像识别是 `计算机视觉` 的典型问题。目前，最好的解决图像识别问题的算法是基于CNN的算法。2012年，使用深度卷积网络在 `ImageNet` 的 `ImageNet Large Scale Visual Recognition Challenge` 中达到了 16%的错误率，被认为是深度学习的革命。

本人对 `计算机视觉` 比较感兴趣，而用深度CNN来处理图像识别问题又是目前比较常见的操作，所以，我选择这个毕业项目。

问题描述

使用深度学习方法识别一张图片是猫还是狗。

- 输入：一张彩色图片
- 输出：狗的概率

数据集和输入

数据集使用 [Dogs vs. Cats Redux: Kernels Edition](#)提供的数据集，

可以用[kaggle api](#)工具，下载整个数据集：

```
kaggle competitions download -c dogs-vs-cats-redux-kernels-edition
```

数据集包含训练集和测试集。

训练集包含25,000张猫和狗的图片，猫和狗的图片各占一半，图片以 `type.id.jpg` 形式命名。

测试集包含12,500张图片，图片根据序号命名。

解决办法

解决这个图像识别的问题，仍使用CNN模型。我们将基于 `ImageNet` 上比较成熟的CNN模型，来解决猫狗大战问题。

基准模型

根据 `Documentation for individual models` 的描述：

Model	Size	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth
Xception	88 MB	0.790	0.945	22,910,480	126
VGG16	528 MB	0.715	0.901	138,357,544	23
VGG19	549 MB	0.727	0.910	143,667,240	26
ResNet50	99 MB	0.759	0.929	25,636,712	168
InceptionV3	92 MB	0.788	0.944	23,851,784	159
InceptionResNetV2	215 MB	0.804	0.953	55,873,736	572
MobileNet	17 MB	0.665	0.871	4,253,864	88
DenseNet121	33 MB	0.745	0.918	8,062,504	121
DenseNet169	57 MB	0.759	0.928	14,307,880	169
DenseNet201	80 MB	0.770	0.933	20,242,984	201

我们将试验以下几种基准模型：

[Xception](#)

[InceptionV3](#)

[DenseNet201](#)

它们的表现相对较好，而且参数和深度相对不会太大。

评估指标

项目的评估指标参见 [Dogs vs. Cats Redux: Kernels Edition Evaluation](#)。它根据 `LogLoss` 来评估，`LogLoss` 越低越好。

`LogLoss` 定义如下：

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

在 `keras` 中, `loss函数` 使用 `binary_crossentropy` 即是 `LogLoss`。

根据 `kaggle` 上目前的 [Leaderboard](#) 排行榜, 要进入前10%, 即前130名, 则 `LogLoss` 至少要少于 `0.06114`。

设计大纲

总体思路, 借鉴[手把手教你如何在Kaggle猫狗大战冲到Top2%](#)和[Building powerful image classification models using very little data](#)的思路。

设计步骤如下:

1. 用 `opencv` 读取数据, 对图片数据进行图片大小统一处理。
2. 用 `keras` 提供的 `ImageDataGenerator` 对图片数据做数据增强。
3. 用 `ImageNet` 上的预训练模型Model对图片数据进行训练。
4. 提取Model中最后一个卷积层的 `feature map`, 并保存。
5. 综合多个Model的 `feature map` 作为输入层, 添加我们的 `Dense全连接层`, 并训练这个最终模型。

参考资料

1. [Dogs vs. Cats](#)
2. [Computer vision\[wiki\]](#)
3. [Machine Learning Attacks Against the Asirra CAPTCHA](#)
4. [手把手教你如何在Kaggle猫狗大战冲到Top2%](#)
5. [Building powerful image classification models using very little data](#)
6. [Documentation for individual models](#)