

Capstone_proposal

2019 年 5 月 21 日

1 AI 算法工程师

1.1 毕业项目开题报告

沈捷 2019 年 3 月 28 日

1.1.1 背景

2013 年, kaggle 举办了一项趣味性赛事“猫狗大战”, 即猫狗图像分类。许多学者及行业人士参与其中, 贡献了大量的解决方案, 机器学习领域的局势由之改变。[?] 现今图像分类在许多行业已展开深入研究, 如医学上热门的癌组织与良性组织的鉴别, 甚至有学者尝试利用深度学习方法鉴别出男女大脑 MRI 图像的差异 [?], 都是类似的问题。但医学领域常由于患者隐私保护政策, 难以获得大量的训练数据, 所以为了更好地学习、研究新的方法, 也有学者会利用猫狗大战数据进行探索尝试和验证 [?], 同理可以延伸至许多应用领域。

1.1.2 问题描述

区分图片是猫还是狗的照片, 显然是一个二分类问题, 通过输入图像特征, 获得一个概率, 通过概率来判断属于哪一个类。

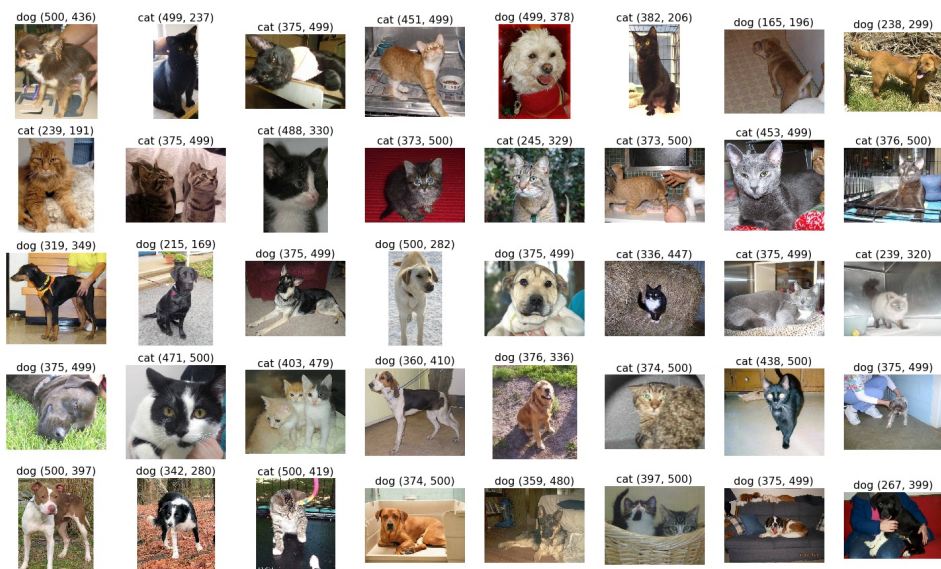
1.1.3 数据输入

该数据集已切分为训练集和测试集。训练集包含 25000 张图片, 猫和狗图片比例为 1:1, 可以从文件名 (标签. 编号.jpg) 中获得标签; 测试集有 12500 张图片, 文件名只有编号。

随机抽取样本, 简单探查这些图片:

可见图片长或宽多在 300~500 像素之间, 大小不一。背景也各种各样, 也许会造成一定程度的干扰, 不过目标主体基本清晰居中。只不过有的主体占图片的比例较小, 有的图片上有两个或以上的目标, 也可能会影响模型训练, 且尚不知是否有分类错误的训练样本。

对于错误分类样本的检测, 决定采用预训练的 ResNet50 模型进行初始预测, 根据 ImageNet1000 的标签, 其中有 118 个狗的品种和 7 个猫的品种, 可以做为参考, 找到那些预测与标签不符的图片, 再做人工确认。如下为 ResNet50 模型中找到的部分错误标签的图:



samples



miss_labeled

可见，这些图中有的目标主体太小，有些有遮挡，有些模糊，有些曝光度太高以至特征消失，甚至一张图片上同时有猫和狗或两者都不是，这都是对模型训练的干扰，这部分图片将剔除。

其余训练集图片会经过随机旋转、缩放、裁剪等预处理，使模型能有更好的泛化能力。这些图像将统一为 224×224 大小，并分解为 RGB 三个颜色通道的色值，作为模型的输入。

1.1.4 解决方案

处理好输入数据后，将采用一个预训练的神经网络进行迁移学习，预计将采用 ResNet50 模型。该模型要求输入为 (224, 224, 3) 的 tensor，所以图片也会随之调整，同时神经网络的输入节点也应为 224×224 大小，输出节点数量为 1，采用 sigmoid 激活函数将图像对某一类的线性预测值映射到(0,1)之间，即为概率（狗=1，猫=0）。sigmoid 函数公式：

$$S(x) = \frac{1}{1 + e^{-x}}$$

训练模型的过程中，将调整 epoch、隐藏层节点、学习速率等超参数，以尽量高准确率，降低对数损失（log loss）。

1.1.5 基准模型

本例没有基准模型。直接将迁移学习训练好的模型在测试集上进行预测，提交到 kaggle 上进行比较判断。模型的训练目标即是达到 kaggle 的前 10%，即 LogLoss < 0.05603。

1.1.6 评估指标

模型将采用对数损失函数（log loss）进行评价。[?] 对数损失函数需要输入每个分类的预测概率与标签，对错误的分类进行惩罚，从而对准确率（Accuracy）进行量化 [?]。损失越少，准确率越高。对数损失函数公式 [?]:

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

其中：* n 是样本数量 * \hat{y}_i 是图像*i* 为狗的预测概率值 * y_i 是图像*i* 的标签， $y_i = 1$ 是狗， $y_i = 0$ 是猫 * $\log()$ 是以自然数*e* 为底的对数函数

此外，预测结果将上传到 _Kaggle_ 进行排名对比，目标是要达到 _Kaggle_ 排名的前 10%，即 LogLoss 达到 0.05603。

1.1.7 项目设计

综上，本项目是一个图像的二分类问题。数据集已经被划分为训练集 25000 张图片和测试集 12500 张图片。对于训练集图像，首先要进行随机翻转、缩放、裁剪等预处理，并根据预训练模型统一成 224×224 大小，再转化为 RGB 三个通道的 Tensor，与预训练模型的标准化参数进行匹配，即均值为 [0.485, 0.456, 0.406]、标准差为 [0.229, 0.224, 0.225]，作为模型的输入。

将采用 ResNet50 为预训练模型框架进行迁移学习，输入节点根据图像大小设为 224×224 即 50176，输出节点为 1，用 Sigmoid 函数输出图像的预测概率（狗 =1，猫 =0），即预测为狗的概率。当概率小于 0.5 时，可以认为预测倾向于猫。

在训练过程中，将调整 epoch、隐藏层数、隐藏节点数、学习速率、dropout 等超参数，尽量提高准确率，降低对数损失。可对 logloss 的变化绘制折线图，观察 logloss 收敛情况。

在测试集上进行预测之后，提交到 kaggle 上进行评价对比。kaggle 则采用对数损失函数 (log loss) 进行模型准确率的评价。本模型的目标是达到 Kaggle 排名的前 10%，即 LogLoss < 0.05603。

1.1.8 参考文献

In []: