**机器学习纳米学位开题报告**

**猫狗大战**

**刘珅珅**

**2019年3月29号**

1 背景

计算机视觉是让计算机能够感知和理解图像的意义，在现实生活中有广泛的应用。图像分类是计算机视觉的一个重要的研究方向，通过对不同类别图像的进行特征提取并构建模型来预测未知图像的类别，在很多实际问题上收到良好的效果，例如udacity课程中讲解的皮肤癌检测的例子，通过对各种类别的皮肤病图片的分类识别来判断患者是否患有皮肤癌，其准确率甚至超过了一些经验丰富的医生。kaggle推出dogs-vs-cats比赛，就是一个典型的图像分类问题，通过机器学习的方式来识别出图片是猫还是狗。选择这个项目作为毕业项目的原因是本人对计算机图像处理感兴趣，并且项目本身会用到深度学习和卷积神经网络的知识，这也是目前人工智能的热点。

2 问题陈述

该项目是一个监督学习的二分类问题，训练集通过图片名称已经打好了标签，区分是猫还是狗，通过机器学习构建模型并训练，然后预测测试集中的图片是猫还是狗。

3 数据集与输入

本项目的数据集为kaggle提供的数据集，训练集为25000张图片，其中12500张为猫类图片，12500张为狗类图片。测试集为12500张猫狗图片。

观察数据集，并随机抽取猫和狗的图片，可以看出图片丰富多样，绝大部分图片的清晰度和分辨率都良好。另外根据<https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition/discussion/47074#latest-479615>描述的情况，图片中存在标识错误的情况，有部分图片与标识不匹配，鉴于这样的图片数目比较少，采用手动删除的方式来剔除掉这些错误的数据。猫类图片总共删除了cat.3216.jpg，cat.7377.jpg，cat.8456.jpg，cat.7564.jpg，cat.9171.jpg，cat.4688.jpg，cat.4085.jpg，cat.5351.jpg，cat.5418.jpg，cat.11184.jpg，cat.10029.jpg。狗类图片总共删除了dog.1043.jpg，dog.1773.jpg，dog.4367.jpg，dog.8736.jpg，dog.8898.jpg，dog.10237.jpg。

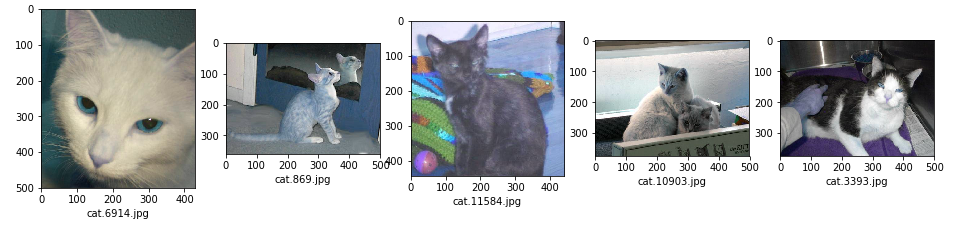


图1 猫的图片

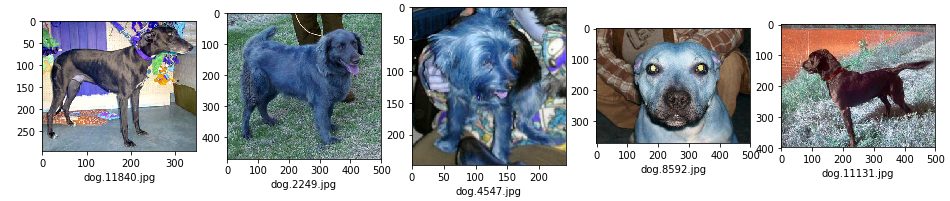


图2 狗的图片

删除异常数据后，训练集猫类图片有12489张，狗类图片有12494张，总共有24983张图片，考虑到交叉验证的情况，训练集中随机抽取20%共4996张图片作为验证集，最终的数据集是训练集19987张图片，验证集4996张图片，测试集12500张图片。

4 解决方案

过去几年中，卷积神经网络在图像分类中有很好的效果，考虑到数据集较大并和ImageNet的图片集有很大的相似性以及Udacity课程中深度神经网络有关迁移学习的介绍，可以采用迁移学习[1]的方式来解决问题。将问题分为两部分：特征提取和模型分类。特征提取是采用预训练模型提取实际图片的边缘特征、轮廓特征以及更高级的能区分猫狗的特征等。模型分类是利用之前提取的特征，输入到一个模型中进行分类训练和预测，并根据训练过程中的参数如准确率、损失大小等来判断是否满足要求来调整模型参数，最终得到符合要求的分类结果。

5 基准模型

毕业项目要求模型的LogLoss必须在kaggle pubic learnboard中排入前10%，即LogLoss<0.06127。考虑到可以对模型进行优化调整，基准模型的LogLoss可以稍大于这个值，所以设置基准模型的LogLoss为0.08.

6 评价指标

在kaggle竞赛中，该项目采用的评价指标为对数损失LogLoss，即对数损失。本项目采用与之相同的评价指标。

对于一个二分类问题，LogLoss的计算公式如下所示：



其中为测试集样本数，代表图片的真实标签，如果是狗，否则。表示模型预测图片为狗的概率。

从LogLoss的计算公式可以看出，模型不仅需要预测正确，而且预测的概率要尽可能的大，即对预测结果非常肯定，否则LogLoss的值就会较大。

7 项目设计

1) 开通阿里云服务器，并在Linux系统安装Anaconda，Python3，Numpy，Pandas，Matplotlib，Tensorflow-gpu，Keras等。

2) 下载数据集并将训练集数据安装不同分类拆分为cat和dog两部分，手动删除其中的异常数据，然后随机抽取20%的训练集数据生成一个验证集。

3) 本项目拟分别采用VGG16，VGG19[2]，ResNet50[3]，InceptionV3[4]，Xception[5]这5种预训练模型通过迁移学习来完成特征提取，这5种模型在ImageNet的图像分类比赛中取得了很好的效果，而猫狗分类隶属于ImageNet分类的一部分。在Keras中5种模型都有对应的预处理函数，VGG16，VGG19，ResNet50的预处理方式是相同的，需要图片每个通道都减去ImageNet的mean值[103.939, 116.779, 123.68]，并将图片从RGB转换为BGR；InceptionV3，Xception的预处理方式是相同的，将图片数据缩放到[-1, 1]之间。

4) 使用keras中的ImageDataGenerator类通过实时数据增强批量将图片转换为张量图像数据，其中VGG16，VGG19，ResNet50默认的图像尺寸为(224, 224)，InceptionV3，Xception默认的图像尺寸为(299, 299)。

5) 采用上述5种预训练模型进行图片特征提取。

6) 构建分类模型，设置相应优化器，评估指标和损失函数。由于是二分类问题，损失函数采用二分类的交叉熵，评估指标一般为准确率。图像分类常用的优化器[6]有SGD,RMSprop,Adam，可以分别尝试然后选择最优结果。模型构建完成后，使用特征数据进行训练，在训练过程中保存下训练过程信息，主要是每个epoch的训练准确率、验证准确率、训练LogLoss和验证LogLoss，通过这些信息对模型进行参数调优，包括调整Dropout，优化器和优化器的学习率等。

7) 对测试集进行预测，并将结果上传kaggle并查看结果是否满足要求。如果效果不满足要求，可以考虑模型微调[6]和模型融合等。

参考文献

[1] 庄福振，罗平，何清，等 迁移学习研究进展[J]. 软件学报，2015,26(1):26-39

[2] Karen Simonyan, Andrew Zisserman Very Deep Convolutional Networks for Large-Scale Image Recognition[EB/OL] <https://arxiv.org/abs/1409.1556>, April 10, 2015

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun Deep Residual Learning for Image Recognition[EB/OL] <https://arxiv.org/abs/1512.03385>, December 10, 2015

[4] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna Rethinking the Inception Architecture for Computer Vision[EB/OL] <https://arxiv.org/abs/1512.00567>, December 2, 2015

[5] François Chollet Xception: Deep Learning with Depthwise Separable Convolutions[EB/OL] <https://arxiv.org/abs/1610.02357>, October 7, 2016

[6] Keras Documentation[Z] <https://keras.io/>