

机器学习工程师纳米学位

● 项目背景

选择计算机视觉中的猫狗大战作为毕业项目。完成课程五的学习和图像分类项目后，对深度学习产生了较为浓厚的兴趣。计划通过猫狗大战项目对卷积网络有更加深刻地理解，加强对深度学习框架的配置和使用。

此项目属于图像分类问题，需要为输入图像打上猫或者狗的标签。图像分类是计算机视觉领域的核心问题之一，许多研究者对此已有了大量研究并且诞生了许多优秀的算法。目前图像分类已在不同的实际场景得到了应用^[2]。

● 问题描述

给定一组猫与狗标注以后的彩色图像，通过使用传统机器学习方法或者 CNN 卷积神经网络对数据集进行建模与学习，之后将模型用于一组新图像的分类，判断图像中是猫还是狗^[1]。

一般对于结构数据的分析，使用传统机器学习方法有比较好的结果。但对于非结构化数据，比如图像，图像特征的提取较为关键，目前卷积神经网络是处理这类问题的很有效的方法。

● 数据集和输入

数据集来源于 Kaggle，这里必须提到 Asirra。它是一种通过询问用户识别猫和狗照片的人类交互证明 (HIP)，这个任务对于计算机是比较困难的。人却可以快速准确地完成任务。Asirra 与世界最大的为流浪宠物寻找家网站 Petfinder.com 合作，他们向微软研究院提供了超过三百万张猫和狗的图像，并由美国各地动物收容所的人员进行了手动分类。Kaggle 提供了这些数据的一个子集，用于娱乐和研究^[1]。

数据集分为训练集和测试集。训练集中包含 25000 张狗和猫的图像。文件夹中每个图片的标签都作为文件名的一部分。测试集中包含 12500 张图像，以数字 ID 命名。对于测试集中每张图片，将预测它是狗的概率，其中 1 代表狗，0 代表猫^[3]。

● 解决方案描述

此项目处理的数据是图像，属于图像分类问题。与处理结构化数据相似。首先，需要对图像数据进行一定的预处理。其次，需要对图片进行特征提取，目前卷积神经网络对于图像特征提取有着很好的效果。最后是有全连接网络进行分类。网络训练过程中也需要进行交叉验证，根据评估指标选取较好的网络模型和超参数。

● 基准模型

选择 CNN 中比较有名的 AlexNet 或者 VGG 网络为基础的模型作为基准模型。

AlexNet 在 2012 年的 ImageNet 图像分类竞赛中远远超过当年的第二名。AlexNet 在 CNN 基础上使用了：1) 非线性激活函数，ReLU；2) 防止过拟合的方法：Dropout，Data augmentation；3) 大数据训练，百万级 ImageNet 图像数据；4) GPU 实现，LRN 归一化层的使用^[5]。

VGG 也是比较稳定和经典的模型，它在 AlexNet 基础上引入了更多的网络层，由 AlexNet 的 8 层变为了 19 层。也使用了更小的滤波器尺寸，对于参数初始化进行了调整，取得了更好的效果^[4]。

● 评估指标

使用 Log 损失来评估模型性能^[3]：

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

其中，

n 表示测试集的图片数量；

\hat{y}_i 表示预测图像是狗的概率；

y_i 是 1 表示图像是狗，0 则为猫；

$\log()$ 表示自然对数（底为 e）；

log 损失越低模型性能越好。

● 设计大纲

环境配置：模型构建和训练需要计算量巨大，考虑使用 GPU。借助现有配备有 GPU 的云服务器作为运行环境，首先安装和配置好所需的深度学习等第三方库。

实现模型：

1. 图片数据集探索和预分析；
2. 将数据集预处理，分为测试集和训练集，并拆分训练集分为训练集和验证集；
3. 根据已有文献，基于现有深度学习框架实现 CNN 模型；
4. 训练模型，观察训练时损失值和正确率变化，调整超参数，并在验证集上评估模型性能。选出最优模型和参数，使用测试集测试模型性能。
5. 实现和部署基准模型，与已实现的模型进行对比，分析不同模型特点和优劣。

参考文献：

- [1] Dogs vs. Cats: <https://www.kaggle.com/c/dogs-vs-cats>
- [2] 对比图像分类五大方法: <https://www.jiqizhixin.com/articles/2017-05-15-7>
- [3] Dogs vs. Cats Redux Kernels Edition: <https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition/data>
- [4] 常用 CNN 模型: <https://www.cnblogs.com/52machinelearning/p/5821591.html>
- [5] AlexNet 介绍: http://blog.csdn.net/cyh_24/article/details/51440344