机器学习纳米学位毕业项目开题报告

张戬昊 heeroz@gmail.com

项目背景

深度学习最近正在成为机器学习领域中极其重要的一部分，这主要得益于最近 GPU 的发展[1]，作为游戏玩家的自己应加入这股浪潮。在 Web 应用中，为了防止程序滥用服务，会使用一种叫验证码的人机识别功能，一般显示一张人类可以分辨，但计算机很难识别的图像，并要求输入识别结果。其中有种验证码是让用户区别一张图像是猫还是狗[2]，这种图像的好处在于人类很好分辨，且会觉得很有趣。这个项目的目的就是使用深度学习，让计算机高正确率的识别猫狗图像。让计算机识别这个图像，不但可以展示深度学习的能力，也能让人们重新思考验证码和安全的问题。

问题描述

为了尽可能正确的识别猫狗图像，我需要建立一种监督学习的机制，使用标记过的猫狗图像，让深度学习神经网络学习，其中要用到卷积和池化。目标是在测试集上达到 90%以上的正确率。最后我将把训练好的模型保存下来，在自家的宠物照片上实验，验证是否能正确进行识别。

数据或输入

本项目将使用 kaggle 上公开的猫狗识别项目数据集[4]，以及 The Oxford-IIIT Pet Dataset[5]，这些数据及包含 25000 张以上的猫狗图像以及标识图像的标签，其中 Oxford 的数据集甚至包含宠物品种的标签，这 2 个数据集可以符合项目的要求。数据输入时会进行一些预处理，比如灰度或 0 均值，来更适用于机器学习。

解决方法描述

我准备使用多层卷积网络（CNN）和池化（Pool），最后输出 0 或 1 值，标识猫或狗，的方式来对图像进行识别。这些使用 Tensoflow 库的 Keras 框架来实现。

卷积和池化能够提取图片中的重要信息，并通过反向传播算法不断提高输出结果。输出的结果可以适用精确度来衡量效果，以及用 logloss 来测量损失。

基准模型

由于输出只有二值，因此随机输出的结果可以达到 50%的精确度（假设测试集中猫狗的数量相等），可以认为 50%是基准。因为猫狗的颜色有部分不重合，估计通过简单的分辨颜色也能进一步提高精确度。根据这篇论文 [3]，60%是以前认为比较难超过的基准。我决定使用 SVM 或线性分类器中的一种作为基准模型，由于 SVM 可能执行太慢且没有 GPU 支持，线性分类器可能更好。

评估标准

我将把数据集区分成 3 份，分为训练、验证、和测试集，测试集只在最后评估环节使用，作为对模型的测量。测量方法是把图片的标识，猫和狗转化为 0 和 1 数值，然后和模型识别的输出结果（0 或 1）比较，使用 logloss 来判断模型的好坏，辅助以正确结果数除以测试集数量作为正确率来评估模型。

项目设计

整个项目将写在 notebook 中并按以下流程完成：

1. 通过代码下载解压图像数据集
2. 预处理这些图形，统一尺寸和 0 均值化。我也可能尝试用灰度化调优模型。
3. 打乱处理后的图形，并分出训练、测试和验证集，最后保存到 pickle 文件中方便读取。由于图片过多，可能会尝试用边训练边读取的方法来防止内存不足。
4. 应用多层 CNN+MaxPool 最后输出二值并最小化 logloss 的神经网络在训练集上训练。
5. 通过验证集评估精确度，反复试验优化神经网络模型，达到一个满意的结果，用测试集测试并保存模型。
6. 自己尝试拍取宠物照片并用该模型进行识别。

引用

[1] J Schmidhuber 2014 - Deep learning in neural networks: An overview

[2] Asirra

[3] Philippe Golle - Machine Learning Attacks Against the Asirra CAPTCHA

[4] https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition

[5] http://www.robots.ox.ac.uk/~vgg/data/pets/