

机器学习纳米学位

毕业项目：猫狗大战

2018年4月17日

开题报告

领域背景

“猫狗大战”项目的领域背景是深度学习中的计算机视觉。根据[维基百科](#)的定义，“计算机视觉”是一个交叉学科领域，该领域研究计算机如何对数字图像或者视频产生高层次的理解。它使得计算机模仿人类视觉系统并自动处理相关任务成为可能，例如对象识别，动作追踪，实景再现和图像复原等等。

计算机视觉的发展最早要追溯到20世纪60年代，那个时候世界上一些研究人工智能的大学就已经开展了相关领域的研究，目的是为了赋予智能机器人“看得见”的能力。后来在20世纪70年代进行的相关研究为这一领域奠定了基础，这些成果至今仍被沿用（包括边缘抽取，线段标注，多面体建模，动作估计和通过多个小结构的连接构成的对象表示等等）。20世纪80年代则开展了严格的数学分析和量化研究。到20世纪90年代末，计算机视觉与计算机图形学的交叉研究开始增多。最近几年计算机视觉研究的主流是与机器学习技术（特别是深度学习）和各种复杂优化方法相结合的，基于特征的方法。

本项目要解决的问题就是计算机视觉领域中的“图像识别”问题，具体来说就是训练计算机识别图片中是猫还是狗。这一问题已经有了很多成熟的解决方案。近几年的研究成果提出了很多以卷积神经网络为核心的图像识别解决方案，例如：

1. [VGGNet](#)
2. [ResNet](#)
3. [Inception v3](#)
4. [Xception](#)

我本人之所以选择这个项目，是因为自己对深度学习中计算机视觉这个细分领域很感兴趣，希望自己能够掌握这个领域的理论基础，培养工程实践能力，将来把图像识别技术应用于自动识别害虫，为农作物病虫害防治提供智能化解决方案。

问题描述

“猫狗大战”要解决的问题是训练计算机识别数字图片中是猫还是狗。这对于人类来说并不是什么大问题，而对于计算机来说有不小的难度，因为它所看到的图片是由数字构成的点阵。目前该问题已经有了很多的解决方案，其中基于卷积神经网络的深度学习解决方案效果最好。该问题属于二分类问题，可以被量化，因为我们最终的目标是为了通过调参和训练得到一个函数 $y=f(X)$ ，函数的输入是数字图像，输出是二分类的概率估计 p 和 q ，既图像有多大的概率是猫（ p ）或狗（ q ），并满足 $p + q = 1$ 。可以通过对预测得到的分类结果和实际的分类结果之间的差异来评估性能，因此是可测量的。最后，该问题显然是可复制的，定义良好的问题，所以它一定能够被很好地用各种方法解决。

数据集

本项目来自[Kaggle](#)，该项目提供了与之配套的数据集。数据集有800多M，分为训练数据和测试数据。训练数据包含25000张狗和猫的图片，每个图片文件都有相应标记；测试数据包含12500张没有标记的图片。该项目首先会使用训练数据对神经网络模型进行训练，最后在测试数据上评估模型的性能和泛化表现。

解决方案描述

拟采用“卷及神经网络”作为本项目的解决方案，这种特殊类型的神经网络在计算机视觉领域有着广泛的应用。具体地，将采取迁移学习的策略，使用在ImageNet上预训练过的四种卷积网络VGGNet，ResNet，Inception v3和Xception导出特征向量，然后在特征向量的基础上构建模型进行分类。这种方法有效的原因是卷积网络是一种分层架构，前几层能识别图像中的一些简单的图案，例如边缘等等，这些往往是每个图像识别问题所共有的特征，因此可以复用并节约时间，只需要训练和调整最后几层即可。

基准测试模型

In this section, provide the details for a benchmark model or result that relates to the domain, problem statement, and intended solution. Ideally, the benchmark model or result contextualizes existing methods or known information in the domain and problem given, which could then be objectively compared to the solution. Describe how the benchmark model or result is measurable (can be measured by some metric and clearly observed) with thorough detail.

评价指标

本项目要得到的最终结果是一个二分类问题，所以这里将用准确率结合二元交叉熵损失函数作为算法性能好坏的评估指标，将根据训练集和测试集的损失函数表现来评估算法性能。如果验证集损失还在下降，那么需要增加模型复杂度或者多训练几代；如果验证集损失上升，则出现过拟合，需要正则化或Dropout防止过拟合；如果验证集的损失出现震荡，则需要减小学习率；如果验证集的损失趋于稳定，则可以减少训练代数。

$$C = -\frac{1}{n} \sum_x [y \ln a + (1 - y) \ln(1 - a)]$$

项目设计

本项目的设计方案分为6个步骤。第一步首先是导入数据并做预处理，方便后面使用ImageDataGenerator。第二步，对数据进行检查，分别基于OpenCV和预训练的Resnet50开发一个人脸检测器和一个猫狗检测器，对所有的训练数据进行检查，看看是否存在既不是猫也不是狗的图片，然后将其删除，以免影响训练效果。第三步则是对已检查清洗好的数据进行基本的可视化分析。第四步，采用迁移学习的策略，使用预训练的VGGNet，ResNet，Inception v3和Xception导出特征向量并保存。第五步，载入已保存的特征向量，然后构建ensemble模型。第六步，训练模型，观察模型的表现并根据实际情况做优化，直到达到项目要求。

参考文献

1. [What Do We Understand About Convolutional Networks?](#)
2. [VGGNet](#)
3. [ResNet](#)
4. [Inception v3](#)
5. [Xception](#)