

徐早立

11/27/2017

开题报告

项目背景

人工智能的研究最早兴起于 1960 年代，而其后的发展可谓一波三折 [1]。1980 年代兴起了人工智能的第二波热潮，一批专家系统登上历史舞台，却又很快消失了。从那以后人工智能跌入了近 30 年的寒冬，直到最近几年才又“火”了起来。那么，人工智能为什么又“火”了呢？一个重要的节点是围棋人工智能 AlphaGo 的出现 [2]。2016 年 3 月，AlphaGo 击败了前世界冠军李世石，惊艳世界。更让人觉得不可思议的是，AlphaGo 甚至可以通过自我博弈就在短时间内进化，并在 2017 年又击败了世界第一的柯洁，全面超越人类最高水平，让人惊呼人工智能这一次是真的来了。通过 AlphaGo，人们开始去了解其背后的原理，于是深度学习和神经网络等词汇开始频繁地出现在大众的视野。事实上，深度学习与神经网络不仅可以用来训练计算机下棋，其作为一种通用的机器学习算法，还被广泛应用到了各种领域，这其中就包括了计算机视觉（Computer Vision）[3]。在 2012 年的 ImageNet 竞赛上，深度学习和神经网络开始展现出其算法的优势，基于卷积神经网络的 AlexNet 算法取得了 15.4% 的错误率，远低于第二名的 26.2%。从那之后，卷积神经网络进一步发展，新的网络结构和优化算法使得其逐渐成为处理计算机视觉任务的主流方法。目前具备看图识物能力的人工智能已经被训练出来，用来替代人类来完成某些简单枯燥的看图识物任务。

问题描述

本项目希望能够训练出可以区分猫和狗的人工智能。这一人工智能需要能够识别出一张彩色照片中的动物是猫还是狗。

解决方法描述

区分猫和狗这样的识别只需要知道识别的是是什么，而不需要知道为什么识别的是是什么。进一步来说，人是怎么区分猫和狗的呢？事实上，我们并不能准确地说出猫和狗的外观区别在哪里，我们只是模糊地知道，这样的动物是猫，那样的动物是狗。这样的任务，可以通过归纳学习的机器学习算法来训练，而卷积神经网络又是表现最好的算法。因此，本文会采用卷积神经网络来建立具有猫狗识别能力的人工智能，希冀其识别能力能够与人类相当。

数据或输入

本项目的数据来自数据网站 Kaggle 的一个猫狗识别项目（<https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition>）。数据分为训练集和测试集，其中训练集包含 25000 张猫或狗的照片，测试集包含 12500 张猫或狗的照片。训练集的 25000 张照片会被用来训练具有猫狗辨识能力的人工智能，而测试集的照片会用来评判这一人工智能的猫狗辨识能力。在训练集中，包含猫或者狗的照片各有 12500 张，如图 1 所示。训练集中猫和狗的照片示例如图 2 所示。初步来看，训练集的数据特征有：

1. 猫狗照片数量均等；
2. 照片的尺寸不一致。

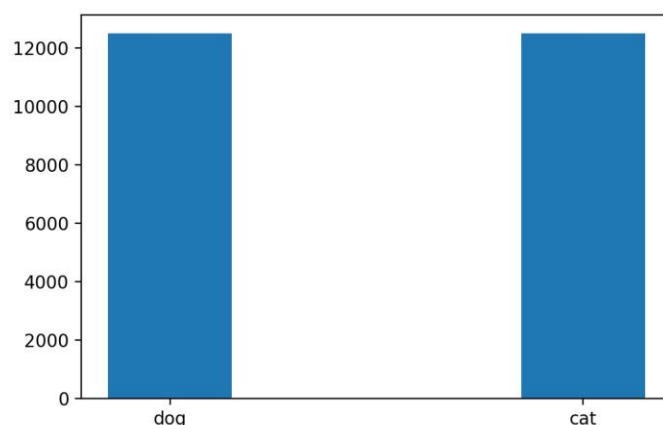


图 1 训练集中猫/狗照片数量柱状图

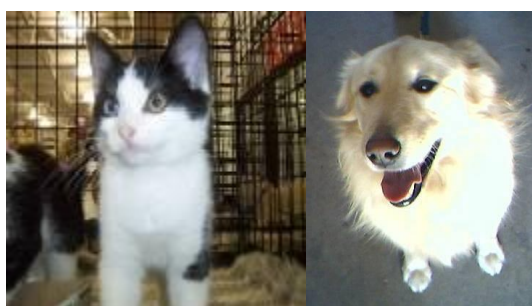


图 2 训练集中猫（左图）和狗（右图）照片示例

基准模型

本项目的基准模型是 AlexNet 的模型。首先会基本参照 AlexNet 来创建神经网络，并根据本项目的具体情况对模型进行相应的修改。如果建模结果已经达到预期效果（训练集的 Loss 值小于或等于 0.05），则不做进一步改进；如果建模结果不如预期，则会参考其他更多的优化模型（比如 VGGNet，ResNet，Inception 等等）对模型进行修改，以期达到预想的预测效果。

评估标准

本项目的损失函数（cost function）为平均交叉熵（average cross-entropy 或叫做 log loss 值），定义为：

$$J = \frac{1}{m} \sum_{i=1}^m [y^i \ln \hat{y}^i + (1 - y^i) \ln (1 - \hat{y}^i)],$$

其中 y 表示实际图片中是猫（ $y=0$ ）还是狗（ $y=1$ ），而 \hat{y} 表示神经网络的预测值。 m 为训练图片总数。所建立的猫狗识别人工智能需达到以下标准：模型在全部测试集上的平均交叉熵小于或等于 0.05。

项目设计

本项目的过程基于数据可分为以下四个阶段：

1. 数据搜集（因为 **kaggle** 已提供数据，因此这一步可以简化）；
2. 数据预处理。针对已有数据集，首先做统计分析，统计已有数据中有多少猫的图片，有多少狗的图片，以及有没有其他类型的图片。再做异常分析，数据集中是否有异常的图片需要被剔除；最后对数据做必要的归一化处理为下一步建模打好基础；
3. 数据建模与优化。如前所述，所用的基准模型为 **AlexNet** 卷积神经网络，以及可能用到的其他经典卷积神经网络模型（比如 **VGGNet**，**ResNet**，**Inception** 等等）；
4. 数据可视化。本项目最好的可视化方法是生成一种应用，通过实际拍一张猫或者狗的照片，看看这一人工智能应用是否在实际应用中有效。计划是在 **iOS** 平台上开发一个猫狗识别的应用。

参考文献

[1] http://mp.weixin.qq.com/s/-wSYLu-XvOrsST8_KEUa-Q, 浅谈人工智能：现状、任务、构架与统一——正本清源，2017，朱松纯，视觉求索

[2] <https://en.wikipedia.org/wiki/AlphaGo>

[3] <https://adeshpande3.github.io/adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html>, 2016, Adit Deshpande, The 9 Deep Learning Papers You Need To Know About (Understanding CNNs Part 3)