

基于深度学习的猫狗识别项目开题报告

丁元虎

2018年03月05日

项目概述

Dogs vs. Cats 项目来源于2013年Kaggle举办的一场娱乐性质的图形分类识别比赛。时隔三、四年，基于深度学习的图像识别技术取得了突飞猛进的发展，先后出现了VGGNet (14.09), GoogLeNet (14.09), ResNet (15.12), Inception v3 (15.12), Inception v4 (16.02), Xception (16.10)等等很有名的模型，其中许多模型的识别准确率已经超过人类。两年前，Kaggle又重新把这个有名的项目作为 'playground competition' 带回来了。本项目使用的数据集即该竞赛使用的数据集，数据集包括训练数据集和测试数据集，其中训练数据集包括25000张标签好的猫或狗的图片，测试集包括12500张未标识的猫狗图片。

在图形识别领域，卷积神经网络 (Convolutional Neural Network)变得越来越流行。卷积神经网络是近年发展起来，并引起广泛重视的一种高效识别方法。20世纪60年代，Hubel和Wiesel在研究猫脑皮层中用于局部敏感和方向选择的神经元时发现其独特的网络结构可以有效地降低反馈神经网络的复杂性，继而提出了卷积神经网络 (Convolutional Neural Networks-简称CNN)。现在，CNN已经成为众多科学领域的研究热点之一，特别是在模式分类领域，由于该网络避免了对图像的复杂前期预处理，可以直接输入原始图像，因而得到了更为广泛的应用。参加该竞赛的选手很多即采用了卷积神经网络取得很好的排名，在本项目中本人也选择使用卷积神经网络技术。

问题描述

猫狗识别项目本质上是一个监督学习二分类项目。根据每张图片的特征，图片可以被分为猫和狗。该项目的目标即建立一个合理的模型，能根据图片的特征准确地识别出图片中的动物是属于猫还是狗。

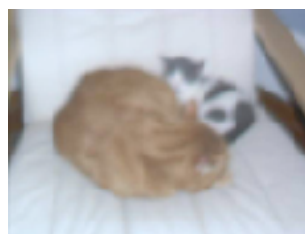
监督式学习 (英语: Supervised learning)，是一个机器学习中的方法，可以由训练资料中学到或建立一个模式 (函数 / learning model)，并依此模式推测新的实例。训练资料是由输入物件 (通常是向量) 和预期输出所组成。函数的输出可以是一个连续的值 (称为回归分析)，或是预测一个分类标签 (称作分类)。二分类问题即根据模型的输入特征把它分为两个类别，这个模型被称作分类器。

目前最广泛被使用的分类器有人工神经网络、支持向量机、最近邻居法、高斯混合模型、朴素贝叶斯方法、决策树和径向基函数分类。分类器的表现很大程度上地跟要被分类的资料特性有关。并没有某一单一分类器可以在所有给定的问题上都表现最好，这被称为'天下没有

白吃的午餐理论’。在机器视觉领域，近年来大热的深度学习网络即人工神经网络的一种，取得了很大的成功。本项目作为图像识别的项目，可以用深度学习技术来解决。

数据和输入

本项目使用的数据集即Kaggle竞赛的‘Dogs vs. Cats’数据集，数据集包括训练数据集和测试数据集，其中训练数据集包括25000张标签好的猫或狗的图片，测试集包括12500张未标识的猫狗图片。测试集25000张图片中，有12500张标注为cat，另外一半标注为dog，训练集的分布较为平衡。通过随机浏览方式查看了1000张标注为cat的照片和1000标注为dog的图片，未发现标注错误的情况，根据统计学原理可以大致估计训练集中标记错误的样本比例不会超过0.1%。图片质量参差不齐，文件大小大至60k小至3k的都有，有很多照片甚至是模糊的，图片尺寸也千差万别。猫狗的品种很多，在图片中的位置、数量及大小变化很大。



解决方案

在该项目中，我计划使用transfer learning技巧，利用pre_trained模型，比如已经在imageNet上训练过的ResNet50，通过开放其中的部分卷积层，在cats_vs_dogs数据集上进行fine_tuning。

参考基准

该项目在Kaggle上目前有1314组参赛者，其中成绩最好者的成绩是0.03302，最差的成绩是19.45806，Top10%的划分线是0.06127，Top5%的划分线是0.05038。在该项目上定的目标是保证进入Top10%，争取进入Top5%。

评价指标

该项目作为一个二分类项目，其评价指标有很多，比如可以用准确度（Accuracy）- 即正确分类的样本数除以总样本数得出；也可以利用混淆矩阵画出 'Precision-Recall curve' 来评判模型；其他还有F-score等等指标都可以对二分类模型进行评价。

在猫狗识别项目中，Kaggle官方指定的评价标准是Log Loss：

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

其中，

- n 代表数据集中图片的数量
- \hat{y}_i 代表当前图片被预测为dog的可能性
- y_i 为1当图片为dog的时候，当图片为cat的时候 y_i 为0
- $\log()$ 是底数为e的自然对数

Log Loss可以用来评估分类器的精度。当模型为每个类输出一个概率，而不是仅有可能的类时，就可以使用它。Log Loss是一种 "软" 的精度测量，包含了概率置信的概念。它与信息理论密切相关：Log Loss其实就是真实标签的分布与预测之间的交叉熵。简单来讲，熵即事物的不确定性。交叉熵融合了真实分布的熵，其用来衡量在给定的真实分布下，使用非真实分布所指定的策略消除系统的不确定性所需要付出的努力的大小。交叉熵越低，这个策略就越好，最低的交叉熵也就是使用了真实分布所计算出来的信息熵，因为此时，交叉熵 = 信息熵。这也是为什么在机器学习中的分类算法中，我们总是最小化交叉熵，因为交叉熵越低，就证明由算法所产生的策略最接近最优策略，也间接证明我们算法所算出的非真实分布越接

近真实分布。通过最小化交叉熵，可以最大化一个分类器的精度。该数值越小越好。那么在该项目中我们也使用LogLoss作为评价指标。

项目计划

在该项目中，可能需要用到的工具有开源的图像处理库opencv及开源的深度学习库TensorFlow、Keras。除此之外还会用到很多前辈的发现和提出的深度学习模型，比如ResNet、GoogLeNet等。

关于这个问题的解决方式，我认为可以分成5步来完成。

第一步，数据的准备和预处理。项目给定的数据都是图片文件，其中分为训练集和测试集，训练集是有标签的，标签即其文件名。首先要把图片和标签转化成模型训练和测试所需要的array和list。其次，把训练集随机分成训练集和验证集，验证集用来验证模型的分类效果；

第二步，利用TensorFlow、Keras 组建模型。参照当前已经在大数据集分类项目中取得较好成绩的模型，计划使用模型迁移和模型融合技术，把这些成功模型拿来为我所用。

第三步，利用提前准备好的训练数据训练模型，根据模型训练过程中输出的loss和acc值判断模型的分类效果。

第四步，调试模型，直至取得目标结果。

第五步，利用结果模型预测测试集的标签。验证测试结果的得分。

参考文献：

Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions[J]. 2016:1800-1807.

Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the Inception Architecture for Computer Vision[J]. Computer Science, 2015:2818-2826.

He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]// Computer Vision and Pattern Recognition. IEEE, 2016:770-778.

吴军. 数学之美.第2版[M]. 人民邮电出版社, 2014.

CyberRep. 知乎. <https://www.zhihu.com/question/41252833/answer/195901726>