

猫狗大战

机器学习工程师 纳米学位 开题报告

张洪阳 2018年4月15日

项目背景

猫狗大战源自 Kaggle 于 2013 年举办的一个娱乐竞赛项目，要求分辨给定测试图片是猫还是狗，属于计算机视觉领域图像分类问题，适合使用以卷积神经网络（Convolutional Neural Network, CNN）构建模型、最小化对数损失函数作为策略、梯度下降与反向传播作为算法的深度学习方法。

卷积神经网络根据生物视觉神经中感受野（Receptive Field）概念的启发，使用卷积核与上层的稀疏连接取代传统 DNN（Deep Neural Network）层与层之间的全连接权重，极大降低了网络参数数量，而理论上相比全连接网络表示能力只是略微降低，并且更容易训练。由于计算机视觉相关问题数据结构的特殊性（图像、视频等），往往每个像素只与附近区域的像素高度相关，非常适合使用 CNN 模型。

问题陈述

猫狗大战本质上属于监督学习，但是图像数据的特征不易提取，甚至不易理解，人类本身也无法准确描述分辨猫狗的核心特征。但若使用整个像素向量作为特征向量，以像素向量的高维度性，使用传统机器学习方法将带来维度灾难（Curse of dimensionality），需求像素维度指数级别的均匀样本，以及指数增长的模型规模与计算量。

数据描述

训练数据集包括 25000 张图片，并在文件名中标注了图片为猫还是狗，猫狗数量各占一半，测试数据集包括 12500 张图片，没有标注类别，文件以数值 ID 命名。

解决方案

针对维度灾难，深度网络的层次结构可以将模型容量（Capacity）增长为样本维度的指数级别，而没有增加太多的模型参数，并且，高维度的图像数据向量并不是每种组合都有意义，模型需要学习的模式通常为数据向量的某种低维流形表示，随机像素组合大都没有意义，能表示猫狗甚至有意义的图像的像素向量只占极少数，并不需要样本维度指数级别的训练样本，使用深度网络模型满足解决此问题的理论基础。然而，对于图像数据，DNN 模型引入的参数太多，随着网络规模增加，几乎无法训练，由此，根据像素值往往只与附近区域的像素高度相关这一先验特征，CNN 模型成为了当前最适合进行图像分类的先验模型。

基准模型

使用 ResNet50、Inception v3、Xception 组合模型输出特征向量进行逻辑回归 (Logistic Regression)。

评价指标

模型输出单元使用 Sigmoid 激活函数，损失函数采用对数损失函数，即伯努利分布下的交叉熵。

项目设计

项目使用以 Tensorflow 为框架封装的 Keras API 进行模型实现：

数据预处理

剔除异常样本，异常值判定使用四分位数法，判定的统计指标可以是图片尺寸、像素均值等，也可以使用预训练模型输出图像特征向量，根据特征向量到均值向量的欧氏距离作为判定值。异常样本删除需慎重，轻微异常的样本有益于模型的泛化性能，最好对判定为异常的样本图片进行人为判断是否保留。图像尺寸统一，一般统一为所使用模型需要的输入尺寸，对于放大的图像需要进行插值处理。

基本模型

由于使用迁移学习方法，可以选择一个在 ImageNet 数据集中预训练的模型，其卷积层已经学习到常见图像数据中的关键特征。一般做法是通过预训练模型的卷积层前馈输出训练数据的特征向量，并使用自适应学习方法其进行逻辑回归，快速得到一个基本模型，并记录模型在训练集和验证集的损失函数取值 Loss 以及预测准确率，绘制学习曲线。