

Dogs vs. Cats Proposal

mnlife

mnlife@foxmail.com

Abstract

2013 年,kaggle 上举办了一个娱乐比赛: Dogs vs. Cats.而在去年,Machine learning 领域发生了很大变化,Deep learning 取得了巨大的突破,其中一个典型的例子就是在 IMAGENET 的挑战赛上, Classification error 由 2011 年的 26%, 降至 2012 年 16%,而其创新性的使用了 deep convolutional neural network. 之后, Classification error 逐年降低,说明了 Deep convolutional neural network 优越性.

而这个项目,使用 Residual Attention Network, 该网络由 Fei wang, Mengqing Jiang, Chen qian, shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, Xiaoou Tang 提出. 这个网络是一个高度模块化的图像分类网络架构.将这个网络迁移至该项目, 希望可以取得良好的效果.

1. Introduction

项目所要解决的为一个典型的二分类问题:给定一张图像,识别其是狗或猫.在图像中,它们有着各种不同的姿态, 这个识别带来了很大的困难, 尤其是在 2012 年 Deep learning 出现之前, 这几乎是一个不可能完成的任务. 不过, 随着近几年深度学习不断出现新的结构, 各种学习模型的不断优化, 使得该问题可以更好的被解决.

尤其是采用了 Residual Attention Network 之后, 对狗和猫进行识别, 只是用来验证该网络的优越性而已, 而以上的问题应该可以被很好地解决.

不过, 由于本次的数据集只有 25000, 所以很难使用 Residual Attention Network 模型进行恰当的拟合, 对模型的容量选择也需要进行不断的尝试.

2. Datasets and Inputs

在 kaggle 的 Dogs vs Cats 竞赛中, 公开了如下的数据集, 该数据集包含 train set 与 test set, 图像为猫或者狗, 任何一个图像都有与之对应的 labels. 其中, train set 包含 25000 张图像, test set 包含 12500 张. 对于其中的图像, 其中的 labels 都有特别的定义: dog = 1, cat = 0.

这样定义后, 可以使用 sigmoid 函数来表示网络的输出级, 并对相应的 feature 进行预测.

图像中包含各种各样姿态的猫与狗: 有站着的, 躺着的, 坐着的, 有正脸, 侧脸. 半身照, 全身照, 这给识别带来了相当大的麻烦. 不过, 现在的深度学习模型的容量也越来越大, 而且本次使用 ResNeXt 网络, 它可以构建很深的深度学习模型而不必担心 Gradient explosion 与 gradient vanishing. 在采用了该网络后, 理论上模型容量可以做得很大, 所以对于这些复杂姿势的拟合也不会有什么問題.

不过, 由于 train set 太小了, 当模型的容量太大时, 很容易导致 under fitting, 所以此时对于模型容量的选择也是一个需要 trade off 的方面.

此外, 如果用以上数据集对网络进行训练, 很大程度上可能会由于训练集太小而使网络 Underfitting, 所以对数据集进行扩充是很有必要的. 还有一个很好的数据集可供选择: The Oxford-IIIT Pet Dataset, 使用该数据集可以作为扩充数据集对网络进行训练.

对于数据预处理, 当输入数据的尺寸不一致时, 可以通过改变 kernel 的大小来变形为相同的向量; 之后对训练集进行划分时, 有必要将训练集拆分为 train set 与 validation set, 以便于更好进行模型的优化.

3. Solution Statement

采用了该网络后, 模型容量增大后导致的 underfitting 问题会变得非常显著, 而模型容量过小又不能很好的对这个分类问题进行表达, 这就成为了一个比较棘手的问题.

可以先从理论方面来分析一下: 本次模型的容量不是问题, 为了让模型泛化的更好的办法就是使用更多的数据进行训练. 但在本次任务中拥有的数据量很有限, 解决这一问题的一个方法是创建假数据并添加到训练集当中. 对于一些机器学习任务, 创建假的数据相当的简单.

数据集增强对一个具体的分类问题是特别有效的. 图像是高维的并包括各种巨大的变化因素, 其中有许多可以轻易的模拟. 即使模型已使用卷积和池化技术对部分平移保持不变, 但沿训练图像每个方向平移几个像素的操作通常可以大大改善泛化. 有许多操作已

被证明非常有效: 如平移, 旋转或缩放图像.

在神经网络的输入层注入噪声也是数据增强的一种方式. 神经网络被证明对噪声不是非常健壮(Tang and Eliasmith, 2000). 改善神经网络健壮性的一种方法是将随机噪声添加到输入再进行训练. 向隐藏单元施加噪声也可以,这可以看做在抽象层上对数据集进行增强. Poole et al.(2014)最近表明, 噪声的幅度被细心调整后该方法是非常有效的. 不过对于正则化策略 Dropout, 可以看做是通过与噪声相乘构建新输入的过程.

通过上面的理论分析, 人工设计的数据集增强方案可以大大减少机器学习技术的泛化误差, 可以通过使用上面的方法, 很好的解决训练数据不足引起的欠拟合问题, 大大改善模型的性能, 不过需要注意的是, 应该讲训练集, 扩充数据集, 测试集, 数据增强方式获得的数据集进行随机后, 再从中分割出 train set, test set, validation set.

进行数据增强后, 训练数据集还是稍显不足, 可以对 Residual Attention Network 采用迁移学习, 只训练后几层网络, 来达到更好的分类效果.

4. Benchmark Model

该问题所要参考的基准模型为 kaggle Dogs vs. Cats 的 leaderboard 中的 Public Leaderboard 的排行榜第一名, 他的分数为 0.03302,该分数为 log loss.

5. Evaluation Metrics

如下交叉熵损失公式为这次的评估指标:

$$\text{LogLoss} = -\frac{1}{n} \sum_{i=0}^n 1 [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

where:

n is the number of images in the test set

\hat{y}_i is the predicted probability of the image being a dog

y_i is 1 if the image is a dog, 0 if cat

$\text{Log}()$ is the nature (base e) logarithm

A small log loss is better.

6. Project Design

本次设计采用 Residual Attention Network 模型, 由于训练集太少, 采用迁移学习, 模型可以在较少的标注样本上进行训练. 对于该任务而言, 迁移学习只需要少量标注样本就可以显著的提升泛化性能.

本次计划采用 Keras, 可以用少量的代码来达到很好的性能. 该网络采用 softmax 输出 1000 分类, 所以对于本次项目, 应该将其更改为 sigmoid 二分类的输出, 进行迁移学习, 计划只对网络的深层进行训练, 而保留浅层的参数, 不对其进行改变

References

- [1] Dogs vs. Cats Redux: Kernels Edition: www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition
- [2] Stanford: Deep learning and computer vision Class
- [3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition. arxiv.org/abs/1512.03385, 2015. 12,10
- [4] S. Xie, R. Girshick, P. Dollar, Z. Tu, Kaiming He Aggregated Residual Transformations for Deep Neural Network. arxiv.org/abs/1611.05431. 2017. 04,11
- [5] Ian Goodfellow, Yoshua Bengio, Aaron Courville: Deeplearning Book. www.deeplearningbook.org
- [6] Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, Xiaoou Tang: Residual Attention Network for Image Classification. arxiv.org/abs/1704.06904. 2017. 04,23
- [7] Andrew Ng: Deeplearning.ai. mooc.study.163.com/smartSpec/detail/1001319001.htm