

Machine Learning Engineer Nanodegree

Capstone Proposal – Distracted Driver Detection

万立佳

August 26, 2018

Proposal

Domain Background

本项目源自美国的 State Farm 公司，State Farm 希望根据画面捕捉准确检测驾驶员的分心行为，以提醒用户专心开车，来更好的保证用户的安全。根据美国 CDC 的统计，在 2015 年，有 3,477 人死于分心驾驶，并且有 391,000 人因为分心驾驶而受伤。根据 CDC 的分析，分心的类型分为三大类：Visual（眼睛离开路面）、Manual（手离开方向盘）及 Cognitive（心思不在开车上）[1]。

此问题属于计算机视觉领域，图像识别问题。由于深度学习在自然语言处理及计算机视觉领域的出色表现，有力的促进了深度学习成为人工智能领域的主流方向之一，在 Google 学术中搜“图像识别 深度学习”关键词有超过 15k 条结果。LeNet 是最早用于深度学习的卷积神经网络之一，ImageNet 竞赛极大的促进了卷积神经网络在图像识别领域的发展，陆续推出了 AlexNet、ZF NET、GoogleNet、VGGNet、ResNet 等网络模型，并且在 2016 年 ILSVRC 的错误识别率已经达到约 2.9%，远低于人类的识别错误率 5.1%[2]。综上所述，通过卷积神经网络应该能解决驾驶员分心识别的问题。

Problem Statement

此问题是图像分类问题，State Farm 公司提供了已经标注的图像样本（训练集、测试集）。在图像识别领域卷积神经网络表现出色，我们可以利用卷积神经网络来提取图像的特征，识别驾驶员的分心状态，例如：ResNet、VGG 等。分类问题是可以被度量的，例如：准确率（Accuracy）。此问题亦是复现的，即在测试集上验证效果，模型应具有良好的泛化能力，以保证在不同的测试集上都有稳定的准确率。

Datasets and Inputs

数据集有三份文件，具体如下：

imgs.zip – 训练、测试集的压缩文件

sample_submission.csv – 一份正确格式的样本提交文件

driver_imgs_list.csv – 截图与驾驶员 id、分心状态 id 的对应关系

输入的数据（imgs.zip）是来自对驾驶员驾驶视频中的不同状态截图，并且已经被标注，标注类型有十类，具体如下：

c0: 安全驾驶
c1: 右手打字
c2: 右手打电话
c3: 左手打字
c4: 左手打电话
c5: 调收音机
c6: 喝饮料
c7: 拿后面的东西
c8: 整理头发和化妆
c9: 和其他乘客说话

图像数据中已经被去除 `metadata`，例如：创建时间等。训练集数据和测试集数据是根据司机区分的，即如果一个司机在训练集那就不会出现在测试集。[3]

Solution Statement

根据以上内容，此问题合理的方案应该是采用卷积神经网络来实现。卷积神经网络通过配置不同的卷积层、池化、全连接层、GAP、Dropout、BN、激活层等，来提取图像中的特征，CNN 的网络模型可以非常灵活，可以顺序型的网络模型，也可以构建复杂的网络模型，并且为了降低参数的量还可以参数共享，这样不仅可以降低参数的量，还可以实现模型的平移不变性。经典的 CNN 模型有 LeNet、AlexNet、VGG、ResNet、Inception 等。他们在网络结构、模型的泛化能力、运算效率等方面做了很多的尝试和突破，这些经典的模型有大量值得我们学习和利用的优点，用以解决图像分类的问题。

Benchmark Model

度量指标用 kaggle 对该竞赛的度量指标 `logloss`。

kaggle 排行榜第 144 名 (Top 10%，共有 1440 团队参加比赛) 的 `logloss` 得分: **0.25634**，即我们的 Benchmark[4]。

Evaluation Metrics

度量指标是：multi-class logarithmic loss，公式如下：

$$\log loss = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij})$$

其中 N 代表测试集样本的数量， M 代表分类的个数，本项目中即为 10。 \log 是自然对数， y_{ij} 有两种取值，当样本 i 属于 j 类时取值为 1，否则取值为 0。 p_{ij} 是样本 i 属于 j 类的概率值。`log loss` 指标比 `accuracy` 考虑的更加细致，因为不仅考虑是否预测该类别的概率最大，而且考虑到了和其他类别的区分度。在结果中并不要求对每张照片的预测结果的和必需等于 1，因为每个概率会除以所有图片所有概率的和，同时为了防止出现 $\log(0)$ ，会取 $\max(\min(p, 1 - 10^{-15}), 10^{-15})$ 来替代归一化以后的概率直[5]。

Project Design

环境：用 aws 的 p3.2xlarge 服务器。Tensorflow 作为 backend 的 keras。

阶段 1: 数据处理：

- a. 数据加载，通过 `wget` 下载数据到服务器，读取数据到内存。
- b. 利用 `opencv` 处理数据格式。
- c. 数据预处理，例如：除以 255.

阶段 2: 模型基础搭建:

- a. 搭一个三层卷积层，GAP 层，Dropout，激活层的网络结构，不做任何调优得到一版初步的结果。
- b. 通过调整 `batch_size`, Dropout, active method, optimizer, epochs 等参数，来寻找参数的最优组合。
- c. 输出 kaggle 约定的结果格式，将结果上传到 kaggle 得到 log loss 结果。

阶段 3: 优化网络结构，使 log loss 降低到 0.2534

- a. 利用迁移学习，借鉴经典的 CNN 的网络结构，来优化结果，例如：VGG, ResNet, Inception, Dense 等。
- b. 输出 kaggle 约定的结果格式，将结果上传到 kaggle 得到 log loss 结果。
- c. 如结果不满足，则通过降低学习率，防止过拟合，提升模型泛化能力等不通的手段优化模型，直至符合结果要求。

Reference

- [1] https://www.cdc.gov/motorvehiclesafety/distracted_driving/
- [2] <http://hongchaozhang.github.io/blog/2018/05/02/cnn-and-image-classification/>
- [3] <https://www.kaggle.com/c/state-farm-distracted-driver-detection/data>
- [4] <https://www.kaggle.com/c/state-farm-distracted-driver-detection/leaderboard>
- [5] <https://www.kaggle.com/c/state-farm-distracted-driver-detection#evaluation>