

机器学习纳米学位

图片中的数字识别

背景

在图片中富含大量文本信息可以提取，如果能便捷的提取这些信息将会极大的方便我们的生活，直接对银行卡拍照就能取出其中的卡号，上传发票的照片就能自动识别发票号这些都能减少我们的工作量。所以本毕业设计将专注于研究图像中的数字识别，抽取出图像中的数字。

问题描述

当输入一张图片时，程序能自动的识别出图片中的数字，并且输出图片所包含的所有数字(顺序无关)。如果找出的数字和图片匹配则认为正确，如果出现任何不匹配的数字，或者有未识别的数字则认为答错。当程序完成时图片的输入可以是直接命令行输入或者后期集成到微信上，支持用户上传图片并将结果返回。

注：[这个方案的灵感来源 \(https://github.com/ypwhs/wechat_digit_recognition\)](https://github.com/ypwhs/wechat_digit_recognition)

相关数据集

[MNIST \(http://yann.lecun.com/exdb/mnist/\)](http://yann.lecun.com/exdb/mnist/)数据集是一个包含6W个训练样本和1W个测试样本的手写数字数据集。在图片数字识别的问题中，很重要的一个功能就是能根据输入的像素块识别出数字，所以使用该数据集能很好训练一个合适的分类器。因为该数据集使用的是固定大小的单通道灰度图，所以在处理的时候可能需要将图片转灰度，然后在进行下一步的处理。

解决方案

主要分为两个部分：

1. 训练一个能够识别灰度图数字的模型，能根据输入的图片判断该图片是否为一个合法的数字(0~9)。使用MNIST的训练数据对全连接的神经网络模型进行训练，然后使用测试集验证。当模型的准确率达到标准时保存模型，后面在对图片进行Sliding Windows扫描时使用。
2. 使用Sliding Windows的方法遍历图片，找出图片中可能存在数字的方块，将可能存在数字的方块输入到模型中得到预测结果。

基准模型

数字识别已经有很多成熟的解决方案，包括kaggle上也有很多类似的[比赛 \(https://github.com/nd009/capstone/blob/master/capstone_proposal_template.md\)](https://github.com/nd009/capstone/blob/master/capstone_proposal_template.md)，使用CNN或者全连接的神经网络模型都能达到比较好的正确率(预测正确的结果/总预测次数)。

验证方法

验证总共分为两部分的评判标准：

1. 第一部分为单个数字的验证，使用MNIST的测试集来测试准确率，评估指标为预测准确率(预测正确的结果/总预测次数)。
2. 第二部分为图片中的数字识别准确率，数据自己随机生成一些带数字的图片，并且打上标签(图片中含有的数字)作为测试集。评估指标为预测准确率(预测正确的结果/总预测次数)。

方案设计

方案按照以下1~7的顺序依次执行：

1. 获取MNIST的数据，预览训练集和测试集，熟悉图片存储的数据格式，实现数据加载流程。
2. 采用DNN的模型训练神经网络。(初步考虑使用DNN，后期可以考虑使用CNN等更复杂的模型)
3. 使用测试集验证模型准确率，当准确率达到一定标准时完成第一阶段。
4. 随机构造100张含有5个数字的灰度图，作为测试集。
5. 使用Sliding Windows的方法依次滑动固定窗口大小，每次取出一小块图片输入到之前训练好的模型中进行预测，如果是数字的话则记录下来，不是数字的话继续滑动窗口直到遍历完成整个图片。最后每个图片都会被标识上一个数字序列。
6. 使用之前构造的100张灰度图进行测试，比对被标识的数字序列，完全一致才算正确，最后统计正确率。
7. 时间允许的话，将整个流程集成到微信中，支持识别上传的图片并返回结果。