机器学习纳米学位

算式识别 金阳 Udacity

2018年12月16日

问题的定义

项目概述

使用深度学习识别一张图片中的算式。 从数据图片观察，图片中有噪点，且每个符号有一定角度的旋转，应该是验证码。这个数据集和mnist手写体识别挺像的，最大的区别是算是识别图片中，有多个字符。

问题陈述

这是一个图片识别问题，所以需要用到卷积神经网络（CNN），并且需要对图片数据做一些预处理。

算式图片中出现的长度是不定长的，需要用到递归神经网络（RNN）得到计算结果。 我决定使用卷积神经网络提取出特征之后，输入到递归神经网络中，识别出其中的算式。

评价指标

正确率=识别正确的算式数量/算式的总数

当算式图片识别出来的每个字符都正确时，该算式为识别正确。

基准模型

一个类似的项目，识别现实生活中音符照片的序列识别论文中，模型在测试集上的正确率有84%.参考了它的模型，我设计了基准模型如下：

其中：

- Convolution layer：kernel size：3*3，strides:1，padding:0
- MaxPooling layer：window size：2*2，strides:0
- 省略了激活函数：relu

使用该基准模型在数据上训练了20个批次，在验证集上准确率达到86%，目标准确率定位99%。

Type	Configurations
input	gray-image
Convolution	#maps:32
Convolution	#maps:32
Dropout	rate:0.2
BatchNormalization	-
MaxPooling	-
Convolution	#maps:64
Convolution	#maps:64
Dropout	rate:0.2
BatchNormalization	-
MaxPooling	-
Convolution	#maps:128
Convolution	#maps:128
Dropout	rate:0.2
BatchNormalization	-
MaxPooling	-
Convolution	#maps:256
Convolution	#maps:256
Dropout	rate:0.2
BatchNormalization	-
MaxPooling	-
Convolution	#maps:512
Convolution	#maps:1100
GlobalAveragePooling	-
Reshape	size:(11,100)
LSTM	#hidden units:128
Dropout	0.2
LSTM	#hidden units:128

Type	Configurations
Dropout	0.2
LSTM	#hidden units:256
Dropout	0.2
LSTM	#hidden units:256
Dropout	0.2
TimeDistributed	#hidden units:17
Activation	Softmax

数据以及观察结果

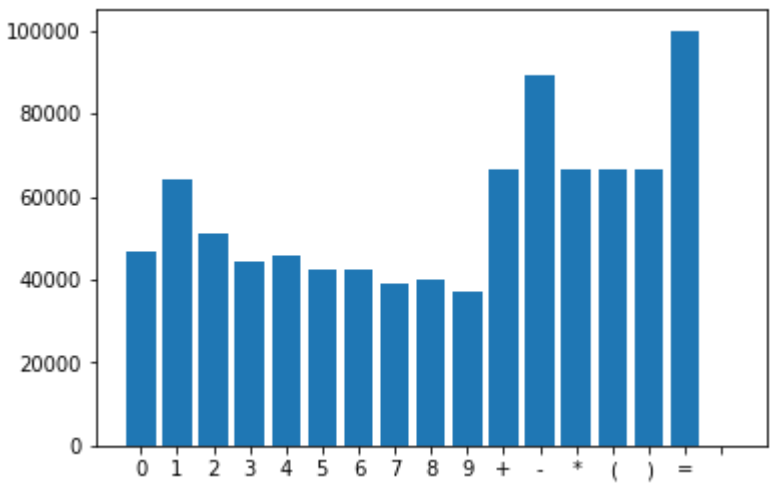
数据集可以通过这个链接下载：[数据下载链接]

此数据集包含10万张图片，每张图里面都有一个算式。

- 每个算式可能包含 `+-*` 三种运算符， 可能包含一对括号，可能包含0-9中的几个数字，以及每个算式包含一个等号。所以一共出现的字符总数是16种。
- 每个字符都可能旋转。
- 图片大小统一是300*64。
- 图片字体是各种颜色的，背景也是各种颜色的，但是背景都是浅色（接近白色）
- 图片中有一些噪点。

探索性可视化

统计标签中各个符号的数量，画柱形图，发现等号最多，每个算式都有，运算符号和括号平均比数字类型的符号多。



预期的解决方案

图片预处理后，直接将算式图输入模型，然后结合CNN和RNN，直接输出。

- 卷积神经网络（Convolutional Neural Network, CNN）是一种前馈神经网络，它的人工神经元可以响应一部分覆盖范围内的周围单元，对于大型图像处理有出色表现。

  卷积神经网络由一个或多个卷积层和顶端的全连通层（对应经典的神经网络）组成，同时也包括关联权重和池化层（pooling layer）。这一结构使得卷积神经网络能够利用输入数据的二维结构。与其他深度学习结构相比，卷积神经网络在图像和语音识别方面能够给出更好的结果。这一模型也可以使用反向传播算法进行训练。相比较其他深度、前馈神经网络，卷积神经网络需要考量的参数更少，使之成为一种颇具吸引力的深度学习结构。我使用它来提取图片特征。

- 长短期记忆（LSTM）是一种时间递归神经网络（RNN），论文首次发表于1997年。由于独特的设计结构，LSTM适合于处理和预测时间序列中间隔和延迟非常长的重要事件。

  LSTM的表现通常比时间递归神经网络及隐马尔科夫模型（HMM）更好，比如用在不分段连续手写识别上。本项目的算式识别，算是不分段连续手写识别的一种。所以在RNN部分使用LSTM。

具体设计：

- CNN部分：

  有5个模块组成，前四个模块的结构类似，以第一个为例，组成为：两个卷积核大小为3 * 3的卷积层，一个标准正则化层，一个relu激活层。然后每个模块卷积层的卷积核的数量。

  最后一个模块在两个卷积层后，通过全局平均池化层，把输出变成一维（1600），然后再把输出的形状调整成（11*100）后，为输入RNN部分做准备。

- RNN部分：

  每个模块为一个LSTM层，如此4个模块后，加上一个全连接层，一个*Softmax激活层，最终输出的形状是（11 * 17）

数据预处理

图片先转成灰度图，然后除以255方便模型计算。 把label转换成one-hot的形式。

项目设计

1. 数据预处理。
2. 设计一个结合CNN和RNN的模型。
3. 训练模型直到达到，大于等于99%正确率。

引用

- S. Hochreiter and J. Schmidhuber. Long short-term memory. Neural Computation, 9(8):1735–1780, 1997.
- A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke, J. Schmidhuber. A Novel Connectionist System for Improved Unconstrained Handwriting Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 5, 2009.
- Baoguang Shi, Xiang Bai，Cong Yao（2015）An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition
- Sebastian Ruder（2017）An overview of gradient descent optimization algorithms*