

手写数字识别实验报告

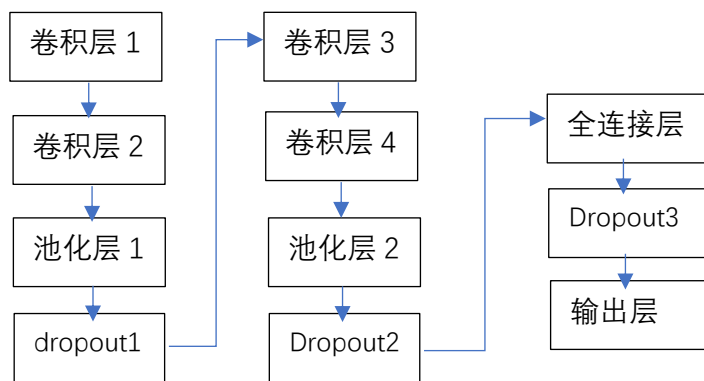
一、模型结构图：

采用 CNN 的架构来训练模型。

我先在模型中定义四个嵌套定义函数，用来构建四个层。在输入数据后，我将一维的数据重定型为 28×28 的矩形，输入进两层相连的， 5×5 卷积，32 通道的卷积层。再连接一层 2×2 池化层，并增加一层 0.25dropout。之后再连接两层 5×5 , 64 通道的卷积层，加上 2×2 池化和 0.25dropout。最后得到的是 $7 \times 7 \times 64$ 的图。将结果再输入进一层 256 个单元的全连接层，取 0.5dropout。最后将全连接层输出到 10 个单元的输出层。卷积层和全连接层都使用 relu 激活函数。

使用配合 softmax 的交叉熵损失函数计算其损失。最后使用 Adam 优化器，并采用指数衰减学习率。

在平台数据集的基础上，我将图片分别向上下左右平移三个像素，得到 5 倍的数据进行训练。将数据以 batch size 为 100 的形式输入进网络。



二、实验结果

最好结果：

Submission and Description	Public Score
results.csv 2 hours ago by Helium using CNN	0.99642

总提交次数： 7

Submission and Description	Public Score	Use
results6.csv just now by Helium cnn with more data	0.99100	
results5.csv a few seconds ago by Helium cnn with dynamic lr	0.98471	
results4.csv 12 hours ago by Helium cnn with regu	0.98128	
results3.csv 12 hours ago by Helium nativ_cnn	0.98871	
results2.csv 13 hours ago by Helium cnn	0.99614	
results_lstm.csv 13 hours ago by Helium using lstm	0.97642	
results.csv 13 hours ago by Helium using CNN	0.99642	
No more submissions to show		

三、参数效果

在试验中我比较了三种参数：L2 正则化，学习率衰减，数据增加
我先构建了一个基本的 CNN 架构，不采用以上参数，将双层卷积层换成单层。
得到的结果正确率为 0.98871。
在全连接层和计算损失上采用 L2 正则化后，得到结果为 0.98128，可以发现效果不显著。我认为原因是 L2 正则化参数不够理想；数据较为丰富，使用 L2 正则化对过拟合的影响较小等。
采用指数学习率衰减后，正确率为 0.98471，效果不显著，但是我猜测是由于训练时间不够，模型未完全收敛。使用衰减的学习率可以得到更准确的最优点。
采用数据增加后，正确率为 0.99100，提升显著。因为更多的数据使模型有更多的机会学习到更有用的信息，而且平移图片防止了过拟合。

四、模型比较

在 CNN 之外，我还实现了单层 256 单元 lstm 模型，将行与行的关系作为时序输入，正确率为 0.97642，离 cnn 距离较大。
分析模型，lstm 并不能完全解读图片中相邻像素的关系，而且以行为单位理解图片不过透彻。CNN 可以多通道（方式）完整解读。但是 lstm 结构较为简单，时间空间表现都更加优秀。

五、问题思考

1、实验训练什么时候停止是最合适的？简要陈述你的实现方式，并试分析固

定迭代次数与通过验证集调整等方法的优缺点。

我认为在迭代到一定数量的 epoch，或者验证集正确率不再上升后即停止。因为采用数据扩增的模型一个 epoch 较大，我采用监控验证集的方式节约时间。

固定迭代次数可以使数据充分参与训练，但是一般耗时过长。

通过验证集调整可以节约时间，并且可以根据验证集的正确率定向调节参数。

- 2、实验参数的初始化是怎么做的？不同的方法适合哪些地方？（现有的初始化方法为零均值初始化，高斯分布初始化，正交初始化等）

在卷积层、全连接层和池化层中采用默认的初始化，在输出层采用高斯分布初始化 w ，常数初始化 b 。用小随机数初始化权重保证网络不会因权值过大而进入饱和状态，同时不同值保证网络可以正常地学习。同时也不能过小，使模型学习过慢。

- 3、过拟合是深度学习常见的问题，有什么方法可以方式训练过程陷入过拟合。采用 dropout，使用卷积层的共享权重，增加正则项，扩大数据等方式。

- 4、试分析 CNN（卷积神经网络）相对于全连接神经网络的优点。

CNN 可以自行抽取图像的特征信息，面对位移、缩放、扭曲等有较强的鲁棒性，自适应能力强。空间连接紧密，能较好地反映图片的空间信息。

与全连接网络相比，CNN 的稀疏连接、共享权重和池化操作使参数更少，学习更快。

六、心得体会

这次实验时我第一次完整使用 TensorFlow 的 CNN 模块，让我对多个神经网络模型有了更深的体会。同时也培养了我分析问题，寻找优秀解决问题的能力。我认为为了培养学生处理实际科研、工作问题的能力，此类大作业应该在课程学习中占更大的比重。