

# 基于卷积神经网络的手写汉字识别方法

肖婷婷

(华东理工大学, 上海 201424)

**摘要:** 手写汉字识别在智能化处理手写文档时至关重要, 高准确率的预测结果对后续处理文本信息有着重要作用。为解决传统汉字识别特征提取繁琐等问题, 使用深度学习中的卷积神经网络, 利用 Keras 搭建网络离线识别手写汉字图片。在 HWDB1.1 的数据集上, 构建 6 层小网络, 预测图片中的汉字。分析实验结果得知, 深度学习可以有效提取图片特征,  $\tanh$ ,  $\text{sigmoid}$ ,  $\text{relu}$  三个激活函数中  $\text{relu}$  效果最好。

**关键词:** HCCR; 卷积神经网络; HWDB 数据集

中图分类号: TP391.4

文献标志码: A

文章编号: 2095-2945(2021)25-0114-03

**Abstract:** Handwritten Chinese character recognition is very important in the intelligent processing of handwritten documents, and the high accuracy prediction results play an important role in the subsequent processing of text information. In order to solve the problem of tedious feature extraction in traditional Chinese character recognition, convolutional neural network (CNN) in deep learning is used, and Keras is used to build a network for off-line recognition of handwritten Chinese characters. On the data set of HWDB1.1, a 6-layer small network is constructed to predict the Chinese characters in the image. Analysis of the experimental results show that deep learning can effectively extract image features, and  $\text{relu}$  is the best among the three activation functions of  $\tanh$ ,  $\text{sigmoid}$ ,  $\text{relu}$ .

**Keywords:** HCCR; convolutional neural network(CNN); HWDB data set

随着中国工业 4.0 的到来, 手写汉字识别 HCCR 应用变得广泛。1996 年 Nag 和 Casey 使用的模板匹配法识别出 1000 个印刷体, 引起了 HCCR 研究的热潮。高学等人在风险最小化准则上建立基于 SVM 的模型<sup>[1]</sup>, 并分析了识别手写汉字遇到的特殊问题。手写汉字具有随意的特点, 和印刷体的规范差距甚远, 采集合适的字体较困难, 且字形复杂, 有较多形似字。因此, 传统 HCCR 流程中的预处理、特征提取效果不好, 而 CNN 能够自动提取特征, 适合处理非线性关系, 对比而言是个好选择<sup>[2]</sup>。有其他研究者改进了 CNN, 如 Graham 等, 针对较少的数据集提出了解决方案, 进行了知识路径积分特征分析, 充分利用了联机时笔画的时序信息, 提高了准确率<sup>[3]</sup>。除 CNN 外, 有其他深度学习的方法也获得了好的效果, DBN 更适合处理一维的数据, 需要预训练, 这两方面均弱于 CNN, 在 HCCR 方面, CNN 效果更好<sup>[4]</sup>, 因而有人提出了 CNN 和 DBN 结合的方法<sup>[5]</sup>。手写单字的技术已较成熟, 但手写文本行依旧是难点, LSTM 和 RNN 适合提取序列信息, 解决难题可能性大。

## 1 数据集特点及预处理

本文采用的是 CASIA-HWDB 数据集, 其中脱机部分比如版本 1.0 和 1.1 总共有至少 7599 个汉字, 而  $\text{mnist}$  只要处理 10 个阿拉伯数字, 可见汉字识别的难度之大。

先将数据转成  $\text{tfrecord}$  格式, 同时记录标签。图像, 图像的长与宽。并且图片尺寸并不该作为变量输入到模型, 故将所有图像都转为  $64 \times 64$  像素。且图片的每一个像素点的范围是  $0 \sim 255$ , 统一将其转为以 0 为中心, 1 为半径的分布, 转为 zero-centered 数据, 加快收敛, 若输入全为正或负, 导致梯度只往一个方向更新, 阶梯状梯度会减慢收敛, 会大大影响深度神经网络。

## 2 基于 Keras 的手写数字识别模型

Keras 由 Python 编写, 是 tensorflow 结合 CNTK 后端等的高层 API, 降低了 tensorflow 编写网络的难度, Keras 会自行根据是否有支持的显卡切换 CPU 和 GPU, 有模块化、简单化、扩展性好的显著优点。支持神经网络的常见方法, 比如数据的预处理, 神经网络训练, 评估和预测, 支持 Sequential 模型和函数化模型。本文运用了层的堆叠 Sequential 模型, 通过层的组合来搭建模型。Keras 最大优点和开发重点, 就是能快速搭建神经网络。

### 2.1 多层神经网络模型

多层神经网络模型(MLP)是包括输入层、隐藏层、输出层的前馈神经网络。MLP 可看作一个有向图, 每两个神经元之间连接的权重是边权, 输入层接受特征的输入, 整体是从输入层往输出层方向, 直到传输到输出层。误差反向传播 BP 算法用权重梯度更新权重, 而权重梯度根

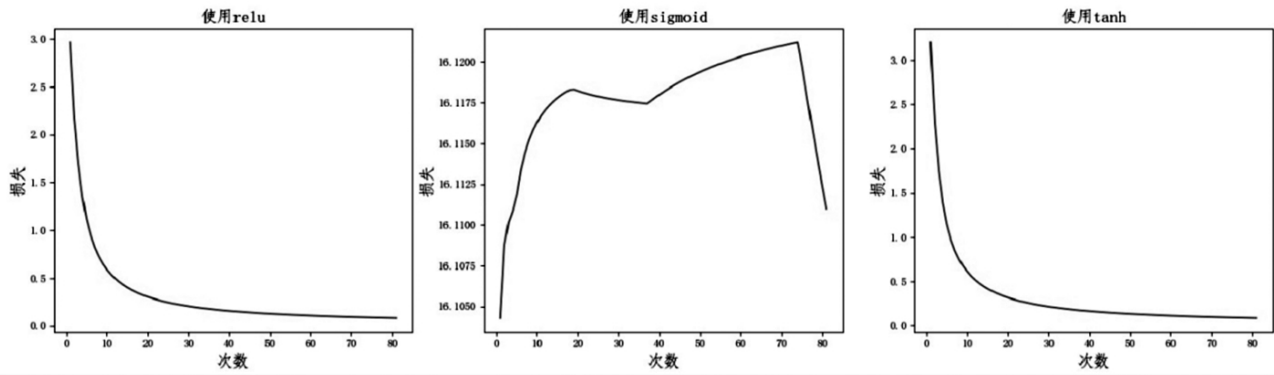


图 1 使用不同激活函数的 CNN 网络训练过程

据输出层预测值和实际标签的偏差,利用链式求导法则求偏差对权重的导数,中间变量为隐藏层的各输出变量。

以一个神经元为例,若  $x$  为输入列向量, $w$  为权重向量, $b$  为偏置, $y$  为输出,则:

$$y = g\left(\sum_{i=0}^M w_i x_i + b\right), \quad (1)$$

其中  $M$  为  $x$  的行数, $g(x)$  代表激活函数,有利于处理非线性的问题。激活函数的种类如表 1。sigmoid 导数大于 0,最大为 0.25,至少每一层会被缩小 1/4,特别是当 sigmoid 输入过大或过小,导数趋于 0,梯度减小快。导数涉及到幂的运算,深层网络耗时增加。输出数据非 0 中心,会导致后续梯度下降时呈现阶梯状。tanh 的函数图像是中心对称的,但仍存在 sigmoid 另两个问题。relu 正的输入数据的梯度为 1,缓解了梯度消失的问题,不涉及幂运算,负数的梯度为 0,降低了过拟合的可能性,但造成了一定可能性的梯度消失,因而提出了 leaky relu 等激活函数。

表 1 不同的激活函数

激活函数	特点	缺点
$f(z) = \frac{1}{1 + e^{-z}}$	把输入的实数范围转为 0 和 1 之间	易梯度消失,输出非 0 均值,幂运算耗时大
$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	输出 0 均值	梯度消失,幂运算耗时长
$\text{Relu} = \max(0, x)$	缓解了正区间梯度消失的问题,收敛速度快	输出非 0 均值

## 2.2 CNN 网络

本文构建网络主要思想是局部感受野,权重共享,池化三部分。局部感受野是使用一个卷积核和原图像部分(尺寸和卷积核的大小一致)进行卷积,此处涉及的局部区域,一个卷积核一次卷积只提取了部分区域的特征。而同一个卷积核以一定的步长值沿着  $x$  轴和  $y$  轴滑动,遍历了整个图片,代表图片的所有小区域共享一个卷积核,卷积核中的元素就是权重和偏移量。手写字识别只提取一种特征是不够的,提取不同特征就需要不同的卷积核,卷积核的数值不同,代表对某个区域的敏感度不同。而权

重共享最大的优点就是大大减少了模型参数,减小了计算量,对深层网络更有利。池化是某个区域取最大值或者均等值将区域的信息转为一个数值,对卷积层的输出进行了简化。池化保留了区域相对整体的信息,但丧失了更精确的位置数据。

## 3 超参数的选择

使用 tensorflow2.1 框架, cuda10.1 版本并行处理, cudnn7, 英伟达 GTX1050 显卡运算。

batch\_size 设为 512,训练 81 代,得出损失大小和训练集的准确率。使用 softmax 激活,将输出层的范围从  $(-\infty, +\infty)$  转为  $(0, 1)$ ,是每个样本属于各类的概率。将某个样本的特征量作为输入,得到  $T$  个类别分别对应的概率,概率最大的类别作为预测标签。实验使用了一个 batch 的数据平均损失来反向传播,输出了准确率。类别使用 onehot 编码,只有下标为真实标签的值 1,其他均为 0,交叉熵作为损失指标。

错误预测比正确损失大,较大错误程度的预测比小的损失更大。网络总体是为了训练出权重矩阵和偏置,使尽量多的样本概率最大的类别是真实类别。模型的优化器选取了 RMSProp,缓解了山谷震荡问题。山谷点的邻域内,即使横轴仍在往一个方向更新,但是纵轴却是来回震荡,甚至可能无法收敛,需要降低纵轴更新速度。 $w$  为权重, $\alpha$  为学习率, $\beta$  为平滑系数, $\varepsilon$  为极小值,公式如下:

$$s = \beta s + (1 - \beta) dw^2, \quad (1)$$

$$w = w - \alpha \frac{dw}{\sqrt{s + \varepsilon}}, \quad (2)$$

$\beta$  起平滑作用,震荡大的方向, $\beta$  值大,步长减小,震荡小的方向则增加了步长。 $\beta$  通常用 0.999, $\varepsilon$  典型值为  $10^{-8}$ 。

## 4 预测结果和分析

分析图 1 得到,第一层卷积使用了 relu 激活函数的 CNN 模型起初损失函数减少快,epochs 越大,损失函数降低得越平缓,最终收敛,训练集的准确率达到 98.42%。

(下转 118 页)

业将由粗放型转变为集约型,对于人力资源、财务和物资采取集约化的管理方式,构建大规划、大运行、大检修、大营销体系和大建设体系。通过“三集五大”模式进行管理,从而提高资源的利用率,使得电力企业发展更加规范化、集约化。

低压配网的安全运行也离不开高效的管理。电力企业的管理层应该加强对电力人员和维修人员的管理,使其对电力系统的检修工作尽职尽责,细心发现电力系统中出现的各种问题,并及时对问题进行解决。企业的发展离不开员工的努力,员工的发展也离不开企业,这两者相辅相成。电力系统安全输送电能离不开电力工作人员,只有加强对他们的管理才能促使他们及时地去解决电力系统中出现的问题,进而为用户提供更加安全、质量更好的电能。

## 5 结束语

时代的发展带动经济的发展,经济的发展带动电力行业的发展。电力行业的稳步发展离不开电能的安全高效输送。现今人们的日常生活需求离不开电力的供应,人们的用电质量需求也在逐步升高,这使得电力公司不仅仅只是输送电力给用户,还应该输送高质量、高效率的电能。因此,电力公司要稳抓电力系统的各个方面。电力系统中三相不平衡问题的出现,导致电力系统设施设备、用户电器、用电情况还有一些动力企业的电动机等都受到

了一定程度的影响,只有对低压配网中的三相电流平衡进行稳固,才能减少这些方面的影响。为此,需要电力人员积极进行检测,电力公司积极引进先进的技术等等。通过这些举措进行节能降耗,提高用户用电质量。

## 参考文献:

- [1]曾楚云.浅谈低压配电网三相不平衡运行的影响及治理对策研究[J].现代国企研究,2018,4(8):162.
- [2]许家益,李小省,凌晓斌,等.低压配电网中三相不平衡装置设计与研究[J].科技创新与应用,2018(8):1-4.
- [3]欧承谦.低压配电网三相不平衡运行的影响及治理措施[J].科学大众,2018(9):32+54.
- [4]白静.低压三相负荷不平衡治理控制策略研究[J].普洱学院学报,2018,34(3):59-60.
- [5]邹水强.低压配电网三相不平衡问题及处理对策[J].自动化应用,2018,4(11):111-112.
- [6]刘永.基于电力载波技术的低压配电网三相负荷不平衡自动调节系统的研究[J].电力系统装备,2019,(14):51-52.
- [7]潘岩.含扶贫光伏配电网电能质量问题及对策研究与应用[D].合肥:安徽大学,2018.
- [8]杨帆.低压配电网三相不平衡治理相关问题探讨访安徽大学教育部电能质量工程研究中心测试评估研究所所长朱明星[J].电气应用,2018,37(8):6-9.
- [9]李沃林.桥头低压配电网三相不平衡问题及对策的研究[D].广州:华南理工大学,2017.

(上接 115 页)

激活函数换为 sigmoid,损失函数呈现略微上升的趋势,迭代了 81 代后未收敛,准确率始终小于 1%,出现了 gradient vanishing 现象。tanh 激活函数下,收敛但在经过相同代数,比 relu 的损失大。一方面,sigmoid 导数最大值为 0.25,而 tanh 导数最大值为 1,故使用 sigmoid 激活函数更易梯度消失。另一方面,relu 正数输入的导数为 1,比 tanh 多数情况大,故更快收敛。实验得到损失误差和准确率如表 2,sigmoid 损失最大,relu 最小。

表 2 激活函数效果对比

激活函数	训练集损失	训练集准确率%
relu	0.083609	98.42
sigmoid	16.110989	0.02651
tanh	0.086384	98.35

## 5 结束语

本文将 HWDB1.1 数据集预处理后,构建了一个简单而有效的卷积神经网络,使用 softmax loss 计算损失,模

型优化使用 RMSProp,根据第一层卷积层的不同激活函数,观察训练集的准确率。实验表明,relu 激活函数在训练同等 epochs 下,分类效果最好。sigmoid 易出现梯度消失的问题,可使用 batch normalization,改用 leaky relu 等激活函数。

## 参考文献:

- [1]高学,金连文,尹俊勋,等.一种基于支持向量机的手写汉字识别方法[J].电子学报,2002(5):651-654.
- [2]Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge [J]. International journal of computer vision, 2015, 115(3):211-252.
- [3]Graham B. Spatially-sparse convolutional neural networks [J]. arXiv preprint arXiv:2014.1409.6070.
- [4]Du J, Hu J S, Zhu B, et al. A study of designing compact classifiers using deep neural networks for online handwritten Chinese character recognition [C]//2014 22nd International Conference on Pattern Recognition. IEEE, 2014:2950-2955.
- [5]李兰英,周志刚,陈德运.DBN 和 CNN 融合的脱机手写汉字识别[J].哈尔滨理工大学学报,2020,25(3):137-143.