

文章编号: 2095-2163(2020)08-0024-03

中图分类号: TP391.7

文献标志码: A

基于深度神经网络的手写数字识别方法研究

徐英卓, 梁学斌

(西安石油大学 计算机学院, 西安 710065)

摘要: 手写数字识别技术的发展随着人工智能的进步也得到了体现, 并已渗入到人们的生活中。文章运用深度神经网络模型(DNN), 来完成手写数字识别。通过改进代价函数, 使用 MNIST 数据集对模型进行训练, 测试评估模型的准确识别率, 手动书写阿拉伯数字输入模型进行测试。实验结果表明, 数字的正确识别率平均约为 98.24%, 改进模型在手写数字识别上有较高的准确性, 具有一定的使用价值。

关键词: 神经网络; 手写数字识别; DNN; MNIST

Research on handwritten digit recognition method based on deep neural network

XU Yingzhuo, LIANG Xuebin

(School of Computer, Xi'anShiyou University, Xi'an 710065, China)

[Abstract] The development of handwritten digit recognition technology has been reflected with the advancement of artificial intelligence, and it has increasingly penetrated into people's lives. The article uses a deep neural network model (DNN) to complete the handwritten digit recognition task. By improving the cost function, the model is trained using the MNIST data set, then the accuracy of the recognition rate of the model is tested, and finally the Arabic numerals are manually written into the model for testing. Experimental results show that the correct recognition rate of digits is about 98.24% on average, and it can be considered that the improved model has good accuracy in handwritten digit recognition.

[Key words] neural networks; handwritten numeral recognition; DNN; MNIST

0 引言

在财政、金融、银行、税务等领域中, 每天都有大量的手写数字产生, 而运用计算机技术能够识别出手写数字, 将会极大的减少工作量。机器学习的快速发展进步使得手写数字识别在这些领域获得了普遍的应用。虽然仅有 10 个阿拉伯数字, 但每个人的书写方式和习惯往往不同, 导致同一个数字会有多种不同的风格, 有时计算机很难识别。因而, 提高识别的准确率^[1]和准确精度是识别技术中的重中之重。

手写数字识别技术方法中常见的有 3 种: 隐马尔科夫模型(HMM)^[2], 支持向量机(SVM), 人工神经网络(ANN)。“自学习”和“记忆”作为神经网络模型的特点, 使其识别技术方法有着较高的正确识别率。深度神经网络因其性能较佳, 在手写数字识别技术中应用的较为广泛。本文将手动书写的 10 个阿拉伯数字数据集作为研究对象, 使用多层感知器训练(DNN)模型对其进行学习和训练, 改进模型代价函数, 对神经网络模型的基本结构和交叉熵代

价函数做出了相应的阐述。此外, 通过训练和测试数据集中的图片来验证其识别准确度, 最终实现了对手写数字识别的测试。

1 神经网络模型设计

1.1 深度学习神经网络的基本结构

深度学习神经网络^[3-4]是一种多层的前馈神经网络模型, 其特点是自学习、自组织及自适应能力较强。深度神经网络的实质是误差反向传播算法, 该算法采用训练神经网络误差的平方作为目标函数, 通过梯度下降方法获取误差平方的最小值。因此, 深度学习神经网络模型的训练过程也就是将样本输入模型后, 多次迭代网络的权重和阈值, 使得输出结果逐渐逼近预期值。简单的 3 层深度学习神经网络结构如图 1 所示。

由图 1 可知, DNN 网络结构分为 3 层。即: 输入层, 隐藏层和输出层, 一般第一层是输入层(Layer1), 最后一层是输出层(layer3), 而中间的层数都是隐藏层(layer2)。隐藏层的输出为:

$$H_j = f\left(\sum_{i=1}^n w_{ij}x_i - a_j\right). \quad (1)$$

基金项目: 陕西省自然科学基金基础研究计划项目(2019JM-383)。

作者简介: 徐英卓(1964-), 女, 硕士, 教授, 主要研究方向: 智能信息处理、决策支持系统; 梁学斌(1995-), 男, 硕士研究生, 主要研究方向: 智能信息处理与可视化。

收稿日期: 2020-06-22

式(1)是激励函数,它主要处理模型的输入并得到一定范围内的输出值。一般激励函数包括 Sigmoid 函数、双曲正切函数等。

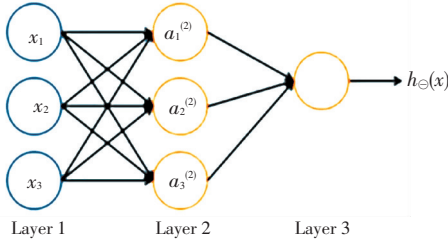


图 1 DNN 基本结构

Fig. 1 DNN basic structure

DNN 相邻层的神经元是全部连接的。由于这个特性,模型容易出现过拟合的现象。目前能够在一定程度上解决过拟合的方法是 Dropout 技术,其本质就是在神经网络的训练过程中,随机的将一部分神经元的节点权重值归为零,使得神经网络有很好的鲁棒性。

1.2 深度神经网络模型设计

神经网络来源于感知机,是感知机的延伸。所以有时也把 DNN 叫多层(是指具有很多的隐藏层)感知机(MLP)或深度神经网络^[3]。

感知机模型是一种浅层的机器学习模型,由两层神经元组成(若干输入和一个输出),激活函数仅仅处理最后一层的神经元。也就是说,只拥有一层的功能神经元,学习能力非常有限,是比较简单的线性模型,因此无法应用在工业上。

本文设计的神经网络模型共有四层:一个输入层,一个输出层和两个隐藏层。每个隐藏层的大小为 100,输出层的大小是随着数据集的类别判断的,因共有 0-9 数字,所以输出层大小为 10。选取 Softmax 函数作为输出层激活函数,使用 Adam 方法进行优化。使用交叉熵代价函数代替均方差代价函数与 sigmoid 激活函数组合。定义 sigmoid 函数表达式为:

$$\sigma(z) = \frac{1}{1 + e^{-z}}. \quad (2)$$

定义均方差代价函数为:

$$L_1(\omega, b) = -\frac{1}{2}(y - a)^2. \quad (3)$$

其中, $a = \sigma(z)$, $z = \omega x + b$,本文采用梯度下降法,分别对 ω 和 b 求偏导数:

$$\frac{\partial L_1}{\partial \omega} = -|y - \sigma(z)| \sigma'(z)x, \quad (4)$$

$$\frac{\partial L_1}{\partial b} = -|y - \sigma(z)| \sigma'(z). \quad (5)$$

其式(2)对应的 sigmoid 函数图像,如图 2 所示。

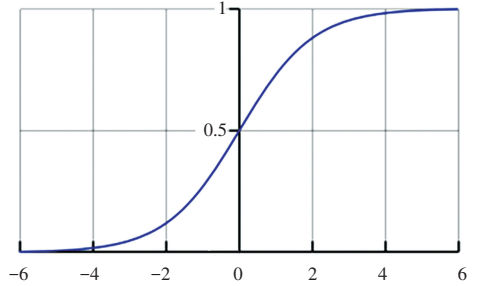


图 2 sigmoid 函数图像

Fig. 2 Sigmoid function image

从图 2 可以看出,当神经元输出接近 0 或 1 时,曲线变得非常平缓,使其导数 $\sigma'(z)$ 越接近于 0。这使得均方差代价函数的 ω 和 b 的偏导数接近于 0,网络更新速度很慢。

假设一个二分类问题,交叉熵代价函数如下:

$$\text{loss} = -\frac{1}{n} [y \ln a + (1 - y) \ln(1 - a)]. \quad (6)$$

交叉熵函数偏导为

$$\frac{\partial \text{loss}}{\partial \omega} = \frac{1}{n} x_j \sum_x [\sigma(z) - y]. \quad (7)$$

由交叉熵偏导可知,权重的学习主要取决于输出值与期望值的误差,好处是当偏差大的时权重更新较快,偏差小时权重更新的较慢,不存在为 0 的情况。所以选取交叉熵函数为代价函数。

2 实验过程与分析

2.1 数据集

本文所使用的是美国国家技术与标准研究所收集并且制作的阿拉伯手写数字数据集(MNIST 数据集),共含有 70 000 张图片,其中 60 000 张为训练数据集,10 000 张为测试数据集^[4]。它是由 250 位不同的人手写而得来。

2.2 实验过程

训练阶段设置初始学习率为 0.001, batch 设置为 100。每 100 个 batch 打印一次信息(准确率与误差)。每训练一轮,计算每轮的平均误差和准确率。并打印测试结果的平均准确率和平均误差。训练模型结果如图 3、图 4 所示。

由训练损失数据变化率和准确变化率可知,模型训练到 80 000 步后,损失率已基本趋于稳定,准确率趋近于 1,可以认为该网络模型训练基本完成。接下来制作一个 600 张手写模型预测数据集,每个数字为 60 张,其识别结果见表 1。(下转第 32 页)