

土木与建筑工程工程 CAE 大作业

基于 BP 神经网络的手写数字识别

蔡诗瑶 2016310048

1 选题背景

字符识别是模式识别中一个重要的领域，从上世纪 50 年代开始，就有许多的研究者对字符识别进行了探索。字符识别处理可以分为两大类，包括文字信息和数据信息。手写数字识别就是其中的一个分支，即利用计算机自动辨认人手写纸上的阿拉伯数字。随着社会信息化程度的提高，国内有大量数据需要输入计算机中，如邮件分拣系统、银行支票处理系统等。但是，现有的手写数字识别方法的可靠性和识别率尚未达到完美的程度。

神经网络是一种模仿生物神经网络的结构和功能的数学模型或计算模型，仿照生物大脑方式进行信息处理。对神经网络的研究起源于上世纪 40 年代，中间经历过低潮期，到了 80 年代开始又呈现出热门的趋势。1986 年，BP（Backpropagation）神经网络算法被提出。Backpropagation 即反向传播，是“误差反向传播”的简称，目前已成为一种广泛使用的网络。神经网络理论在人工智能、自动控制、机器人、模式识别等方面都具有重要的应用价值。神经网络为手写数字识别提供了新的手段。

2 选题目的

本研究希望通过学习 BP 人工神经网络的原理和图像预处理技术，利用 BP 神经网络进行样本训练和数字识别测试。同时也希望在此研究中，熟悉 Matlab 的基本操作以及相应的工具包。

3 研究内容

利用 BP 神经网络进行手写数字的识别，需要以下几个步骤。

（1）训练样本与测试样本采集

神经网络需要通过大量的样本进行学习，从而建立可靠的识别规则。

样本采集的方法很多，本研究采用的方法是将纸上书写的数字进行扫描并转换为计算机中的图形文件。

手写数字识别的一大难点在于：不同人书写的习惯不同。为了获取更多类型的手写数字样本，本研究邀请了 11 名同学，连同本人共 12 人，进行数字书写。最终得到每个数字 400 个（每个数字 4 幅图片，每幅图片 100 个），0-9 共 4000 个手写数字样本，其中随机选取 3600 个作为训练样本，剩余 400 个作为测试样本。另有手写电话号码 20 个作为测试样本。为了简化问题、降低识别难度，本研究要求用于训练的单个数字之间没有连笔、电话号码为没有倾斜的一行，并剔除了一些过于潦草的样本（上述样本中已进行了剔除）。

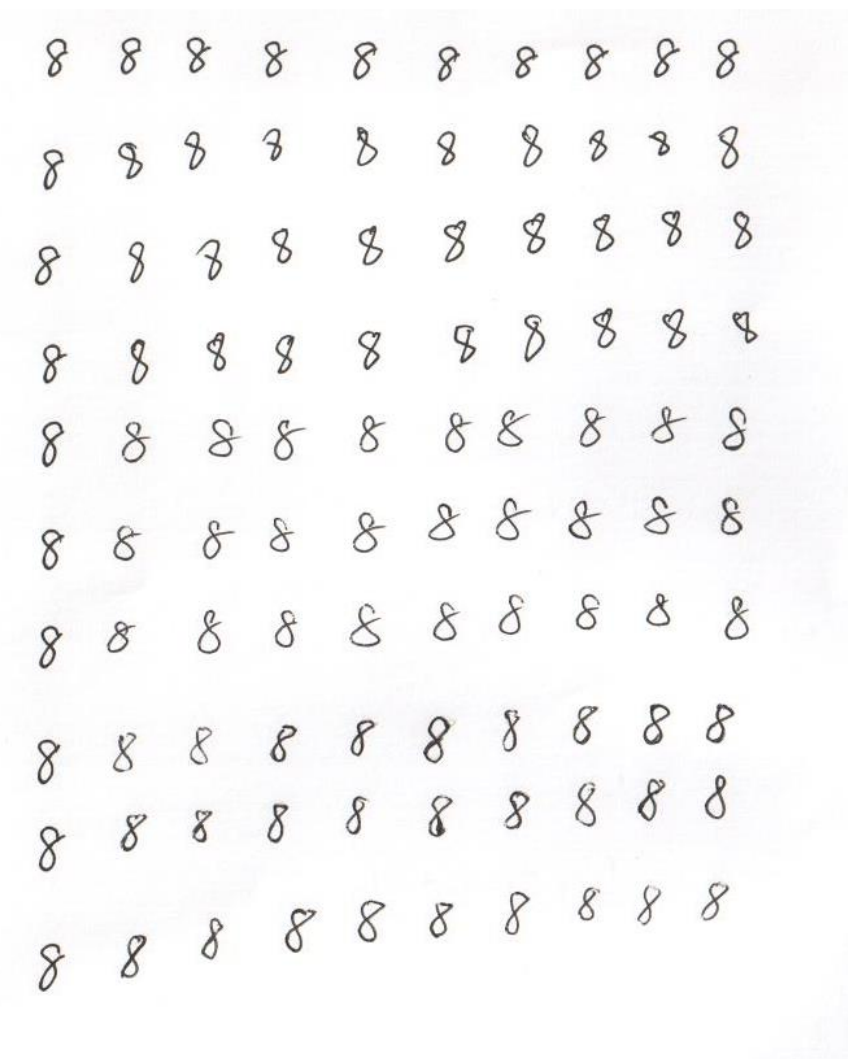


图 100 个数字 8 的训练样本

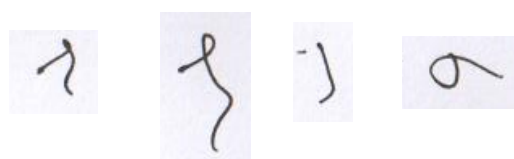


图 被剔除的部分数字样本（左起依次为 9、9、3、5）

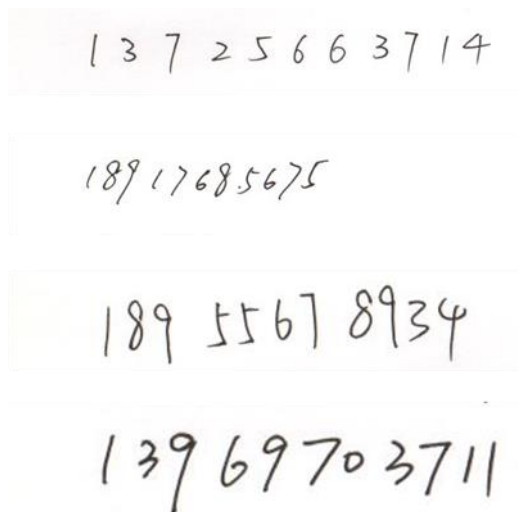


图 部分手写电话号码样本

(2) 图像预处理

图像预处理的流程如图所示：

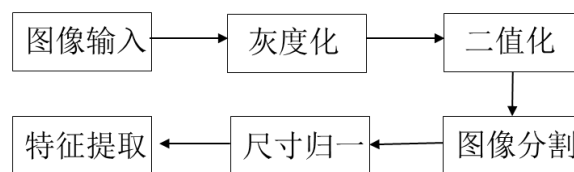


图 图像预处理流程

在图像预处理的过程当中，采用了一些图像处理的技术，如：图像数据读取、图像的二值化、字符切分、缩放和特征提取等。对于倾斜度调整、去噪等问题，本研究暂未考虑。

在手写数字的识别中，我们只需要知道数字的笔划经过哪些像素，而对数字的颜色、各像素点的灰度值信息则没有要求。因此，为了减小后续处理的计算负担，在图像读入后，需要进行灰度化和二值化，即将彩色图像转化为灰度图像，然后转化为由 0, 1 表示的二值像素矩阵形式（例如，没有笔划经过的像素点为 0，有笔划经过的为 1）。二值化需要设定灰度阈值，当某个像素点的灰度值大于阈值时，认为该点为背景；小于阈值时，则认为有笔划通过。本研究采用了 Matlab 工具箱提供的 `graythresh` 函数，该函数使用最大类间方差法得到阈值，即是的前景和背景像素灰度值的类间方差最大。该方法能有效地获取图像二值化所需的阈值。

获得二值化图像后，需要进行数字切分，获得独立的数字样本。字符切分有

多种方法，本研究选择了连通域分析法、直方图投影法这两种方法进行试验。其中，连通域分析法首先对图像开运算，去除细小的噪声，然后对图像中八邻域方式连接的连通域进行标记，计算连通区域的框线特征，得到左上顶点坐标、长宽，并依照这些数据对图像进行分割。在本研究的 4000 个样本中，每个数字有 4 张图片，每张图片含 10 行 10 列，共 100 个数字，且行、列的排列不整齐，此时用连通域分析法能进行快速划分。但是，对于有连笔的数字（如一行电话号码），用连通域分析法就比较难处理。而对于单行的电话号码，只需考虑水平方向的切分，直方图投影法比较适用。直方图投影法的原理是：对于每个相同的 x 坐标，统计所有 y 坐标上有效像素的个数。考虑噪声和连笔情况时，可以设置较低的阈值，判断有效像素较少的 x 坐标处是否为字符间隔。但是，对于字符距离过近、连笔过于明显，以及部分数字本身中部像素密度低（如数字 4）的情况，直方图投影法也无法很好地解决。

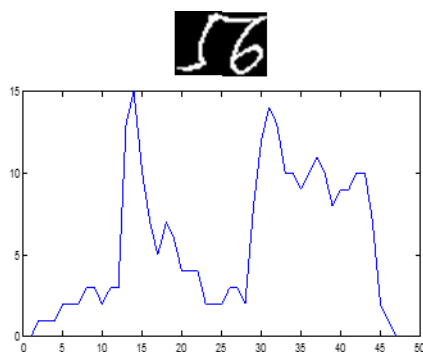


图 带连笔的 56 的直方图投影



图 被中间截断的数字 4

为统一后续的处理和计算，需要将分割后的数字图像调整为相同的大小，并将每个数字图像统一后的像素形成向量，作为神经网络的输入。本研究尝试了两种方法。第一种方法是统一为 20x20 像素，将 400 个像素构成一个向量作为神经网络输入；第二种方法是统一为 70x50 像素，然后进行粗网格划分，以 10x10 像素为单位，再将 70x50 的图像分割为 7x5 个粗网格，统计每个小区域中图像像素所占百分比作为特征数据。为了避免对图像进行过分拉伸，对于宽度过小的图像（如数字 1），本研究采用了在两边补全空白像素的方法。根据后续的试验，发现粗网格划分后能大大加快训练速度，且对识别效果没有明显影响，因此最终采用了粗网格划分的方法。

(3) 神经网络训练与测试

人工神经网络的结构包括输入层、隐含和输出层，每层节点仅和下一层节点相连。在本研究中，隐含层的激励函数采用 `logsig` 函数，而输出层则采用线性函数。

在本研究中，输入层包含 35 个神经元，输出层包含 10 个神经元，分别表示数字 0~9。根据样本数量和模型的复杂程度，选择 1 个隐含层，25 个神经元。

4 识别效果

在人工神经网络训练和测试的过程中，本研究使用了 `Matlab` 提供的神经网络工具包中的函数，也尝试了根据神经网络迭代的原理自己实现算法。下表摘录了两种算法在某次运行时的识别准确率。

表 1 数字识别准确率 (%)

算法	识别内容	1	2	3	4	5	6	7	8	9	0
Matlab 函数	单个数字	100	94.4	97.2	93.3	87.0	100	100	90.2	89.1	97.8
	电话号码	90.9	80.0	81.5	66.7	90.0	81.0	90.3	71.4	83.3	85.7
自己实现	单个数字	100	97.1	84.6	87.5	68.8	79.2	87.5	25.6	87.8	100
	电话号码	97.7	86.7	85.2	33.3	70.0	81.0	87.1	28.6	55.6	71.4

利用 `Matlab` 函数时，训练时间约 23s，对单个数字和电话号码的识别准确率相对较高，且准确率较为稳定。自己实现的算法测试时，训练时间约 12s，准确率相对低一些，且不是很稳定，每次运行后，准确率特别低的数字不一定相同。这可能与汇报时林佳瑞师兄提出的问题有关，即自己实现算法时，是将数字一个一个放入神经网络训练的，后面的数字会对前面的训练结果产生影响。

5 写在最后

本次小课题的研究，使我对图像的基本处理、神经网络的核心原理，以及 `Matlab` 的使用都有了一定的了解。以前觉得“人工智能”、“模式识别”、“神经网络”这些词汇非常高端以至于遥不可及，现在虽然依然很遥远，但是却感觉它们变得亲切了，甚至感觉自己与这个时代的最前沿之间的鸿沟变得浅了一些、窄了一些。最初选题的时候，其实想达到一个更好的效果，但是后来由于时间所限和自己的拖延，只好止步于目前的水平。无论是“土木与建筑工程 CAE”这门课程，还是手写数字识别这个课题，都让我非常开心，希望以后会有机会接触到更多这样有趣的技术和知识，也希望自己以后更加积极地去尝试，少一点拖延。