

基于组合特征的 Bp 神经网络数字识别方法^①

张翼成, 陈欣, 杨红军, 张玉建

(上海宝信软件股份有限公司, 上海 201900)

摘要: 提出一种组合特征作为 Bp 神经网络输入层向量实现数字字符识别算法. 该算法首先引入了数字字符结构特征中图段特征, 并结合数字字符的行列统计特征组合成为新的特征向量; 然后根据新的组合特征向量设计 Bp 神经网络分类器; 最后对已有的数字图像样本空间中的训练样本库按照 Bp 神经网络分类器训练方法进行训练, 并对测试样本库中的样本进行识别. 根据测试实验, 数字字符的识别准确率可达到 94% 以上.

关键词: 组合特征; Bp 神经网络; 分类器; 数字识别; 图段

Neural Network Digital Characters Recognition Based on the Combined Feature

ZHANG Yi-Cheng, CHEN Xin, YANG Hong-Jun, ZHANG Yu-Jian

(Shanghai Baosight Software Co., Ltd, Shanghai 201900, China)

Abstract: The digital characters recognition method based on the combined feature which is used as input layer feature vector in Bp neural network is proposed. First, segment feature, which is introduced in this paper, and the ranks of the statistical characteristics of the numeric characters are combined as a new feature. Then, Bp neural network classifier is designed according to the combined new feature. Finally, when the training sample set is trained according to the training method of the BP neural network, the recognition results of the test sample set can be acquired. According to the result, the recognition rate can reach at above 94%.

Key words: combined feature; Bp neural network; classifier; digital recognition; segment feature

1 前言

Bp 神经网络是基于模仿人类大脑的结构和功能而构成的一种信息处理系统, 是多层前馈型网络, 具有很强的自学习性、自组织性、容错性、高度非线性、高度鲁棒性、联想记忆功能和推理意识功能. Bp 神经网络算法通常用于模式识别相关领域, 是数字识别算法研究的热点之一. Bp 神经网络数字识别方法研究主要集中在两个方面, 第一是神经网络分类器设计中输入层结点特征向量提取, 常用的特征提取方法有逐像素特征提取法^[1]、骨架特征提取法^[2]和垂直方向数据统计特征提取法^[3], 还有一些新特征, 如文献[4]研究了一种 13 维特征提取方法与量子神经网络相结合来实现手写数字识别的方法; 第二是神经网络分类器的优化改进, 如文献[5]通过结合附加动量法和自适应学习

速率形成新的 Bp 改进算法实现数字识别. 本文针对 Bp 神经网络输入层结点特征向量, 提出了一种数字字符结构特征中图段特征与数字字符的行列统计特征组合而成的新的特征向量, 并根据此特征向量设计 Bp 神经网络分类器实现数字分类识别.

2 组合特征提取

组合特征主要包含两部分内容, 第一部分为图段的特征, 图段特征是组成数字字符整体结构特征的一部分, 能够准确反映数字字符的形状; 第二部分为数字字符的行列统计特征, 行列统计特征主要反映的数字字符的统计特性. 将图段特征和行列统计特征组成称为新的特征不仅能够反映数字字符的统计特性, 而且还能够反映数字字符的形状结构特性, 较全面的反

① 收稿时间:2012-08-13;收到修改稿时间:2012-09-18

映了数字字符的特征。

2.1 图段特征提取

对于二值图像，我们定义图段的概念^[6]，它是二值图像的每一行或每一列中被 0 或图像边界所分隔的值为 1 的单个像素或多个像素。图段的大小由其游程表示，最小是 1，最大不超过行或列的长度。

根据扫描方式的不同，图段可分为行图段和列图段。由逐行扫描方式产生的图段称为行图段，由逐列扫描方式产生的图段称为列图段。如图 1 所示，图 1(a)中有三个行图段，游程分别为 3、1、5，图 1(b)中有两个列图段，游程分别为 2、1。



图 1 行图段和列图段示意图

根据上述图段的定义，如下图 2 所示，数字字符 0 图像的大小为 8×16 像素，图段特征可以表示为：行图段特征(0,1,3,3,3,3,1,0)；列图段特征为(0,1,1,2,2,2,2,2,3,2,2,2,2,2,2,1)。

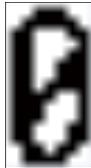


图 2 数字字符 0

2.2 垂直水平方向数据统计特征提取

对于二值化图像，根据扫描方式的不同，垂直水平方向数据统计特征也分为行方向数据统计特征和列方向数据统计特征。行方向数据统计特征是统计该行前景色像素的个数，列方向数据统计特征是统计该列前景色像素的个数。如图 1(a)所示该行的数据统计特征为 9，图 1(b)所示的列的数据统计特征为 3。而对于图 2 所示，数字字符 0 的行数据统计特征为(0,3,5,4,3,4,5,5,6,5,4,4,4,4,0)，列数据统计特征为(0,10,13,7,6,8,10,0)。

3 Bp神经网络数字识别原理

3.1 神经网络模型及原理

Bp 神经网络是基于误差反向传播算法的人工神经网络，也是目前应用最广泛的神经网络模型之一。

Bp 神经网络无需给出有关模式的经验知识和判别函数，只需要通过自身的学习机制揭示输入输出模型的映射关系，从而形成决策区域。由于 Bp 神经网络具有极强的非线性数据处理能力、容错能力和自主学习能力，因而在大多数的神经网络模型都是采用 Bp 网络和它的变化形式。Bp 神经网络为拓扑类型结构模型，通常包括输入层、隐层和输出层。如图 3 所示。

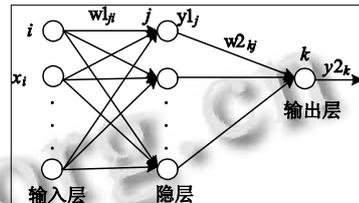


图 3 Bp 神经网络拓扑结构模型

如图 3 所示，假定输入为 X ，输入神经元数量为 n ，隐含层内神经元数量为 s ，输出为 $Y1$ ，激励函数为 f_1 ，输出层内神经元数量为 m ，对应的激励函数为 f_2 ，输出为 $Y2$ ，目标矢量为 T 。信息的正向传递为：

1. 隐含层中第 j 个神经元的输出为：

$$y1_j = f_1(\sum_{i=1}^n w1_{ji} x_i + b1_j); j = 1, 2, \dots, s$$

其中 $w1_{ji}$ 表示输入层第 i 个向量到隐层第 j 个向量的权值， $b1_j$ 表示隐层第 j 个向量的输出与第 j 个向量的输入之间的误差。

2. 输出层第 k 个神经元的输出为：

$$y2_k = f_2(\sum_{j=1}^s w2_{kj} y1_j + b2_k); k = 1, 2, \dots, m$$

其中 $w2_{kj}$ 表示隐层第 j 个向量到输出层第 k 个向量的权值， $b2_k$ 表示输出层第 k 个向量的输出与第 k 个向量的输入之间的误差。

3. 定义误差函数为：

$$E(W, B) = \frac{1}{2} \sum_{k=1}^m (t_k - y2_k)^2$$

t_k 表示输出层第 k 个向量的理想输出值， $y2_k$ 表示根据输入层节点计算获得的输出层节点值。

利用梯度下降法求权值变化及误差的反向传播。

1. 输出层的权值变化。从第 j 个输入到第 k 个输出的权值有：

$$\begin{aligned} \Delta w2_{kj} &= -\eta \frac{\partial E}{\partial w2_{kj}} = -\eta \frac{\partial E}{\partial y2_k} \cdot \frac{\partial y2_k}{\partial w2_{kj}} \\ &= \eta (t_k - y2_k) \cdot f_2' \cdot y1_j \\ &= \eta \cdot \delta_{kj} \cdot y1_j \end{aligned}$$

其中,

$$\delta_{kj} = (t_k - y_{2k}) \cdot f_2' = e_k f_2', \quad e_k = t_k - y_{2k}$$

f_2' 表示隐层到输出层激励函数的导数, η 表示步长, t_k, y_{2k}, e_k 同上.

同理可得:

$$\begin{aligned} \Delta b_{2kj} &= -\eta \frac{\partial E}{\partial b_{2kj}} = -\eta \frac{\partial E}{\partial y_{2k}} \cdot \frac{\partial y_{2k}}{\partial b_{2kj}} \\ &= \eta (t_k - y_{2k}) \cdot f_2' \\ &= \eta \cdot \delta_{kj} \end{aligned}$$

2. 隐含层的权值变化. 从第 i 个输入到第 j 个输出的权值有:

$$\Delta w_{1ji} = -\eta \frac{\partial E}{\partial w_{1ji}} = -\eta \frac{\partial E}{\partial y_{2k}} \cdot \frac{\partial y_{2k}}{\partial y_{1k}} \cdot \frac{\partial y_{1k}}{\partial w_{1ji}}$$

其中, $\delta_{ji} = e_j \cdot f_1', e_j = \sum_{k=1}^m \delta_{kj} w_{2kj}$,

$$\delta_{kj} = e_k f_2', e_k = t_k - y_{2k}$$

同理可得

$$\Delta b_{1ji} = \eta \cdot \delta_{ji}$$

误差反向传播过程实际上是通过计算输出的误差 e_k , 然后将其与输出层的激励函数的一阶导数 f_2' 相乘来求得 δ_{kj} , 由于隐含层中没有直接给出目标矢量, 所以利用输出层 δ_{kj} , 进行误差反向传递求出隐含层权值的变化量 Δw_{2kj} , 然后计算 $e_j = \sum_{k=1}^m \delta_{kj} w_{2kj}$, 并同时通过将 e_j 与该层激励函数的一阶导数 f_1' 相乘而求得, 一次求出前层的变化量 Δw_{1ji} .

3.2 神经网络设计

3.2.1 输出层结点设计

神经网络分类器主要根据输出层节点来进行分类和决策. 对于 0-9 数字识别系统, 所需分类数目为 10, 可以采用 10 结点型输出层和 4 结点型输出层. 对于 4 结点型输出层, 它采用 8421 码来识别. 如输出结点向量为 (0.1, 0.1, 0.1, 0.1) 表示识别数字 0, (0.9, 0.1, 0.1, 0.1) 表示识别数字 8, 等等. 对于 10 结点型输出层则比较直观, 结点向量元素的序号表示识别的数字, 如输出层结点向量 (0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.9, 0.1, 0.1, 0.1) 表示识别的数字为 3. 本文选择 10 结点型的输出层.

3.2.2 输入层结点设计

根据本文第二部分组合特征提取内容可知, 输入层结点向量包含两部分内容, 一部分为数字字符图像

的图段特征, 一部分为数字字符垂直水平方向数据统计特征, 假定图像的尺寸为 $w \times h$ (宽 \times 高), 那么输入层结点的数量为 $2 \times (w + h)$. 根据本文 Bp 神经网络的原理可知, 当输入层各结点值取值范围在 -1 到 1 之间的时候, 能够大幅减少误差反向传播的迭代次数. 而组合特征向量各元素的取值范围通常会大于 1, 因而应当将组合特征向量的各元素值转化为 -1 到 1 之间, 对于如图 2 所示的数字字符 0, 它的组合特征向量如下所示:

(0, 1/3, 1, 1, 1, 1, 1/3, 0, 0, 1/3, 1/3, 2/3, 2/3, 2/3, 2/3, 2/3, 1, 2/3, 2/3, 2/3, 2/3, 2/3, 2/3, 1/3, 0, 10/16, 13/16, 7/16, 6/16, 8/16, 10/16, 0, 0, 3/8, 5/8, 4/8, 3/8, 4/8, 5/8, 5/8, 6/8, 5/8, 4/8, 4/8, 4/8, 4/8, 4/8, 0).

3.2.3 隐含层结点设计及相关参数设计

3.2.3.1 确定隐含层结点数目

确定隐层节点数的最基本原则是: 满足精度要求的前提下取尽可能紧凑的结构, 即取尽可能少的隐层节点数目. 隐层节点数不能太少, 太少则网络可能根本不能训练或网络性能差; 同样隐层节点也不能太多, 太多虽然可使网络的系统误差减少, 但可能会使训练容易陷入局部极小点而得不到最优解. 本文根据上述的原则将隐层结点数目设定为 12.

3.2.3.2 确定初始权值

权值的初始值选择遵循两个原则: 第一, 所选权值不能相同; 第二, 在定义的值域空间均匀分布. 权值的初始值选得好则网络收敛快, 否则网络处于饱和区难于收敛或陷入局部极小, 导致输出节点误差难以满足, 训练失败. 本文选取 -1 到 +1 范围内的随机数作为神经网络的初始权值.

3.2.3.3 确定激励函数、步长和误差阈值

激励函数通常是非线性函数, 常见的阈值函数 (Heaviside)、分段线性函数、Sigmoid 函数、对数 S 型函数 (logistic 函数) 四种类型. Bp 算法要用到各层激励函数的一阶导数, 所以要求其激励函数处处可微^[7]. 本文选取的激励函数为:

$$f(x) = \frac{1}{1 - e^{-x}}$$

误差的反向传递调整权值的过程根据 Bp 神经网络的原理可知, 与激励函数、步长和误差阈值有密切的关系, 本文将步长设定为 $\eta = 0.015$, 误差阈值 $e = 0.0010$.

根据上述 Bp 神经网络的设计, 将输入层节点向量, 输出层节点向量, 初始权值, 误差阈值和步长设定阈值代入本文 3.1 节中相关原理计算公式, 即可最终求得输入层节点到隐层节点和隐层节点到输出层节点的权值矩阵.

3.3 Bp 神经网络数字识别原理

Bp 神经网络数字识别可以分为训练和识别两大部分. 训练通常是对一定数量的数字字符训练样本空间的每一个样本采集组合特征作为输入层结点向量, 然后根据 Bp 神经网络分类器的设计原理, 生成输入层结点向量到输出层结点向量的对应关系. 这个过程可以获得输入层结点到隐含层结点的权值矩阵和隐含层结点到输出层结点的权值矩阵. 训练中采用的样本空间可以采用外来的数字字符样本空间, 也可以采用实际采集的样本中作为训练用的样本空间. 但是不管训练的样本空间来源何处, 都必须具有相同的结构和形式, 且这种结构和形式要和识别部分数字图像预处理的结果保持一致. Bp 神经网络训练流程如下图 4 所示.

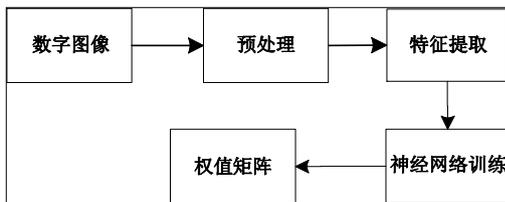


图 4 Bp 神经网络训练流程

如图 4 所示, 当采集获得一定数量的数字图像, 首先需要经过图像预处理, 图像预处理过程通常包含灰度化、二值化、梯度锐化、噪声滤除、图像分割和归一化处理几个环节. 经过图像预处理后的数字图像被分割成为大小相同的单个数字字符图像, 如图 5 所示. 然后对每一个数字字符图像(8×16)提取组合特征, 送入训练网络进行训练, 获得输入层结点到隐含层结点权值矩阵和隐含层结点到输出层结点权值矩阵, 从而确定输入层到输出层的映射关系.



图 5 数字字符图像处理过程生成图

Bp 神经网络数字识别的识别部分步骤是: 1) 将样本空间中的每一个测试样本进行图像预处理后, 分割并归一化为 N 个单个字符图像; 2) 对每一个单个字符图像提取组合特征作为输入层结点向量, 根据训练获得权值矩阵, 计算输出层结点向量; 3) 根据输出层节点向量的值及所在的节点序号确定分类识别结果. 如下图 6 所示:

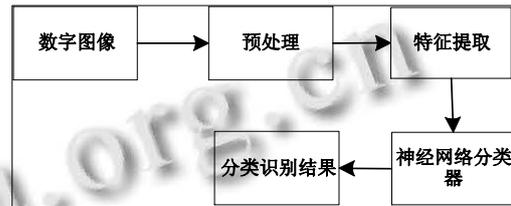


图 6 Bp 神经网络数字识别流程

4 实验及结论

按照上面的原理, 首先对训练样本库进行训练, 样本库的图像名存储在一个 txt 文件中, 训练时调用这样一个 txt 文件就可以对所有的训练样本图像进行训练, 可以获得输入层到隐含层和隐含层到输出层的权值矩阵, 程序界面如下图 7 所示.



图 7 识别程序界面

训练完成之后, 将测试识别的图片调入, 进行数字识别 (下转第 54 页)

获取大量用户访问历史信息. 见图 8.

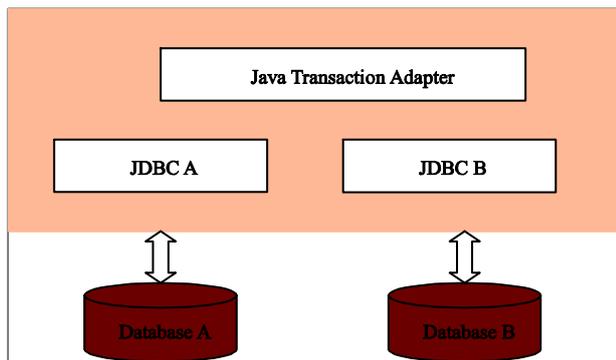


图 7 直接访问的原理框图

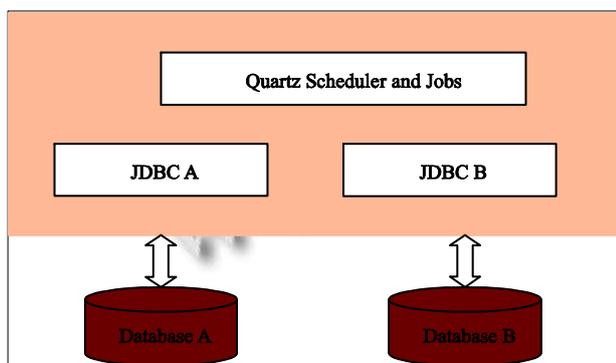


图 8 批量获取的原理框图

5 结语

通过两个系统的整合, 对实验教学管理有所创新. 一方面提高了实验室管理水平, 各种实验室相关信息能集中在该平台上, 便于服务和管理; 另一方面通过学生自主预约开放性实验室, 提高了学生的创新能力, 也提高了实验室的使用效率, 可谓一举多得.

致谢 本文所涉及之系统在前期开发和后期维护阶段得到杭州甲子科技有限公司大力支持, 特此表示感谢.

参考文献

- 1 白云, 柴钰. 加强开放式实验教学培养学生的创新能力. 实验室研究与探索, 2010, (8): 109-112.
- 2 谢辉, 魏金岭, 马楠. 通用标准化高校教学管理新系统分析与设计. 计算机系统应用, 2009, 18(10): 24-26, 43.
- 3 徐萍萍, 张贺文, 张国英, 纪宝伦. 基于 Internet 开放式实验教学管理系统的开发. 实验技术与管理, 2004, 21(6): 126-128.
- 4 袁晓辉. 基于 Web 的开放实验室管理系统. 实验室研究与探索, 2005, 24(7): 55-58.
- 5 徐彬彬, 叶柏龙, 陆兰. 一种基于 J2EE 的异构数据库的同步方法. 微计算机信息, 2011, 27(7): 205-206, 40.

(上接第 116 页)

字识别, 如上图 7 所示, 从上图中可以看出能够准确的识别出图像中的数字.

在具体测试过程中, 分别采用本文的组合特征和逐像素特征提取方法来对数字字符进行特征提取, 并将这些特征作为 Bp 神经网络的输入层节点进行训练和识别, 并分别记录它们识别数字的准确率. 训练样本的单个数字字符图像数量为 500 幅, 测试样本的数字字符数量为 300 个 0~9 数字字符图像, 测试实验结果如下表 1 所示.

表 1 本文方法与逐像素特征提取法数字识别比较

	误识率	拒识率	准确率
逐像素法	5.14%	3.27%	91.59%
本文组合特征提取法	3.68%	2.06%	94.26%

从表 1 可以看出, 本文提出的组合特征 Bp 神经网络数字识别方法误识率、拒识率明显低于逐像素 Bp 神经网络数字识别方法, 正确率较之于逐像素法得到显著提高.

参考文献

- 1 陈艳平, 赵磊. 基于 8 邻域编码逐像素跟踪法的指纹细节特征提取算法. 电脑知识与技术: 技术交流, 2008, 3(11): 932-934.
- 2 李政广, 吴黎明, 等. 骨架提取在 IC 晶片缺陷机器视觉识别中的研究. 半导体技术, 2007, 32(4): 53-56, 86.
- 3 刘春丽, 吕淑静. 基于混合特征的孟加拉手写体数字识别. 计算机工程与应用, 2007, 43(20): 214-215.
- 4 肖婧, 谭阳红. 基于新特征提取法和量子神经网络的手写数字识别. 电子测量技术, 2009, 32(6): 84-87.
- 5 王婷, 江文辉, 肖南峰. 基于改进 BP 神经网络的数字识别. 电子设计工程, 2011, 19(3): 108-111.
- 6 林小竹, 沙芸, 籍俊伟, 万建邦. 计算二维图像欧拉数的新公式. 微电子学与计算机, 2005, 22(11): 158-161.
- 7 柯大观, 李少平. 神经网络的发展及应用. 福建电脑, 2002, (3): 5-6.