

BP Network Based on Handwritten Digital Image Feature Extraction

Xiaojuan Wang, Yanping Bai

North University of China, Taiyuan
Email: 920136560@qq.com

Received: Mar. 26th, 2014; revised: Apr. 23rd, 2014; accepted: May 3rd, 2014

Copyright © 2014 by authors and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

With the development of computer network, more and more electronic products use handwritten input. Therefore, the identification of these handwritten images becomes increasingly important. This paper uses the network model with one hidden layer, we extract 7 invariant moments (Hu moments), mean, variance and the characters of the image potential, with these characters to identify the images, and the recognition accuracy is above 80%.

Keywords

BP Network, Seven Invariant Moments, Characters Potential

基于BP网络的手写体数字图片特征提取

王晓娟, 白艳萍

中北大学, 太原
Email: 920136560@qq.com

收稿日期: 2014年3月26日; 修回日期: 2014年4月23日; 录用日期: 2014年5月3日

摘要

随着计算机网络的发展, 越来越多的电子产品会用到手写输入。因此, 对这些手写图片的识别变得越来

越重要。本文使用的是具有一层隐含层的BP网络模型，提取了图像的7个不变矩(Hu矩)特征、均值、方差以及图像的字符势能，用这些字符特征来对图像进行识别。识别的准确率在80%以上。

关键词

BP网络，7个不变矩，字符势能

1. 引言

BP 网络是一种反向传播网络，是对线性可微分函数进行权值训练的多层网络。BP 网络在模式识别方面有很重要的应用。像缺失二维码的识别、图像的识别、车牌识别等方面都广泛应用 BP 网络。这个网络具有高度的非线性运算性，不同的输入与输出的对应，使网络更加灵活。本文应用 BP 网络对图片的特征进行识别，得到很好的结论。识别效率在 80%。

矩函数[1]在图像分析[2]中有着广泛的应用，如模式识别[3]、目标分类、目标识别与方位估计、图像编码与重构等。本文主要应用的是其在模式识别方面的应用。一个从一幅数字图形中计算出来的矩集，通常描述了该图像形状的全局特征，并提供了大量的关于该图像不同类型的几何特性信息，比如大小、位置、方向及形状等。图像矩的这种特性描述能力被广泛的应用在各种图像处理、计算机视觉和机器人技术领域的目标识别与方位估计中。一阶矩与形状有关，二阶矩显示曲线围绕直线平均值的扩展程度，三阶矩则是关于平均值的对称性的测量。由二阶矩和三阶矩可以导出一组共七个不变矩。不变矩是图像的统计特性，满足平移、伸缩、旋转均不变的不变性，在图像识别领域得到了广泛的应用。

投影能表现图像的某种特征信息，指定方向上单条前景像素的个数。本文主要做了垂直方向的势能和水平方向的势能。

2. BP 网络原理[4]

2.1. BP 网络模型

图 1 中 P 表示输入，从图中可以看出这个网络有 r 个输入，共有 q 组这样的输入。 $W1$ 表示隐表示输出层的权值， $B2$ 表示输出层的误差。 k 输出层第 k 个神经元的实际输出， $F2$ 表示输出层的激活函数。从图中可以看出，输出层的输入是 $s1$ 维的，输出层也有 q 组这样的输入。

2.2. 节点输出

节点的输出主要是包括隐含层和输出层节点的输出，其实两者的输出是一致的，都是前层节点的输

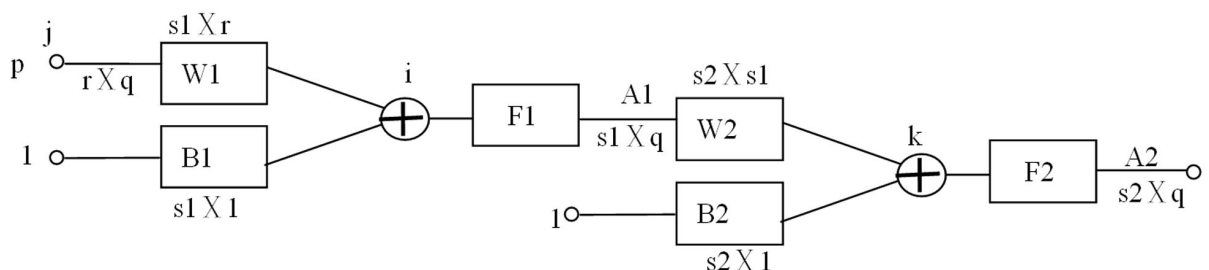


Figure 1. BP network model
图 1. BP 网络模型

出加和再加上误差的影响，如下：

隐节点输出模型：

$$a_i = f1\left(\sum_{j=1}^r w1_{ij} p_j + b1_i\right) \quad (1)$$

其中 a_i 表示隐含层的输出。

输出节点输出模型：

$$a_k = f2\left(\sum_{i=1}^r w2_{ki} a_i + b2_k\right) \quad (2)$$

其中 a_k 表示输出层的输出。

2.3. BP 网络的激活函数

BP 网络经常有的是 S 型的对数或正切激活函数和线性函数。因为 S 型函数具有非线性放大系数功能，它可以把输入从负无穷大到正无穷大的信号，转换成 -1 到 1 之间的输出。对较大的输入信号，放大系数较小；而对较小的输入信号，放大系数则较大。所以采用 S 型激活函数可以处理和逼近非线性的输入/输出关系。不过，如果在输出层采用 S 型函数，输入则被限制到一个很小的范围；若采用线性激活函数，则可采用网络输入任何值。所以只有当希望对网络的输入进行限制，如限制在 0 和 1 之间，那么在输入层应当包含 S 型激活函数，在一般情况下，均是在隐函数采用 S 型激活函数，而输入层采用线性激活函数。

2.4. 误差的反向传播

误差计算模型是反映神经网络期望输出与计算输出之间误差大小的函数：

$$E = (W, B) = \frac{1}{2} \sum_{k=1}^{s_2} (t_k - a2_k)^2 \quad (3)$$

2.5. BP 网络的学习规则

BP 算法是一种监督式的算法。主要是通过输入样本，通过网络得到的实际输出样本和已知与输入样本对应的输出样本之间的误差来修正权值的。使得实际输出样本和已知的输出样本尽可能的接近，也就是使得网络的输出层的误差的平方和最小。它是通过连续不断地在相对于误差函数斜率下降的方向上计算网络权值和偏差的变化而逐渐逼近目标的。每一次的全职和偏差的变化都与网络误差的影响成正比，并以反向传播的方式传播到前面的每一层。

3. 7 个不变矩和字符势能

3.1. 二值字符的 7 个不变矩(Hu 矩)特征

在连续情况下，图像函数为 $f(x, y)$ ，那么图像的 $p + q$ 阶几何矩(标准矩)定义为：

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy, \quad p, q = 0, 1, 2, \dots \quad (4)$$

$p + q$ 阶中心距定义为：

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy, \quad p, q = 0, 1, 2, \dots \quad (5)$$

其中 \bar{x} 和 \bar{y} 代表图像的重心,

$$\bar{x} = m_{10}/m_{00} \quad (6)$$

$$\bar{y} = m_{01}/m_{00} \quad (7)$$

对于离散的数字图像, 采用求和号代替积分:

$$m_{pq} = \sum_{y=1}^N \sum_{x=1}^M x^p y^q f(x, y), \quad p, q = 0, 1, 2, \dots \quad (8)$$

$$\mu_{pq} = \sum_{y=1}^N (x - \bar{x})^p (y - \bar{y})^q, \quad p, q = 0, 1, 2, \dots \quad (9)$$

N 和 M 分别是图像的高度和宽度;

归一化的中心距定义为:

$$\eta_{pq} = \mu_{pq} / (\mu_{00}^\rho); \quad \text{其中 } \rho = (p+q)/2+1 \quad (10)$$

利用二阶和三阶归一化中心矩构造了 7 个不变矩 $M1 \sim M7$:

$$M1 = \eta_{20} + \eta_{02} \quad (11)$$

$$M2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \quad (12)$$

$$M3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \quad (13)$$

$$M4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (14)$$

$$M5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) \left((\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2 \right) \\ + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \left(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right) \quad (15)$$

$$M6 = (\eta_{20} - \eta_{02}) \left[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (16)$$

$$M7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) \left((\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2 \right) \\ - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03}) \left(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right) \quad (17)$$

Hu. M. K 在 1962 年证明了由这 7 个不变矩[5]构成一组特征量具有旋转, 缩放和平移不变性。实际上, 在对图片的识别中, 并不是所有的 Hu 矩都保持三个特性, 其实只有 $M1$ 和 $M2$ 不变性保持的比较好, 剩下的几个不变矩对不变性带来的误差就比较大了, 有学者认为只有基于二阶矩的不变矩对二维物体的描述才是真正的具有旋转、缩放和平移不变性($M1$ 和 $M2$ 刚好都是由二阶矩组成的)。

由 Hu 矩组成的特征量对图片进行识别时, 避免了用图片直接进行识别所引起的信息量大, 从而速度很快, 但是也存在一些缺点, 比如识别率比较低, 因为图片的特征量不能很好的反应图片的本质。对于纹理丰富的图片, 像手势识别[6], 识别就很低。由于 Hu 不变矩只用到低阶矩(现在最多三阶矩), 所以并不能对图像的细节描述清楚, 图像的特征量不明显。

Hu 不变矩一般用来识别图像中大的物体, 对于物体的形状描述得比较好, 图像的纹理特征不能太复杂, 像识别水果的形状, 或者对于车牌中的简单字符的识别效果会相对好一些。

3.2. 二值字符的字符势能投影特征

针对以往字符投影(见图 2)的密度统计特征的不足, 我们进行字符势能特征提取[7], 下面简要介绍一

下字符势能。

最后我们得到的特征数据是：

2 2 3 8 1 9 5 4 2 2 5 1 6 6

4. 识别过程

图片的势能有垂直方向势能(见图 3)和水平方向势能(见图 4)，本文提取了手写体数字图片的 7 个不变矩以及每张图片的字符势能，然后对提取的这两方面的特征向量再求均值、方差等，下面我们看一下几幅图的特征和字符势能：

我们选取两副图片，来看看他们的不变矩和字符势能。

我们选取的图片(见图 5)都是从画图工具插入的 28×28 的图像。

数字 1 的不变矩特征： $bubianju1 = [1.74809618654275; 13.0086683512297; 15.4218332622757;$

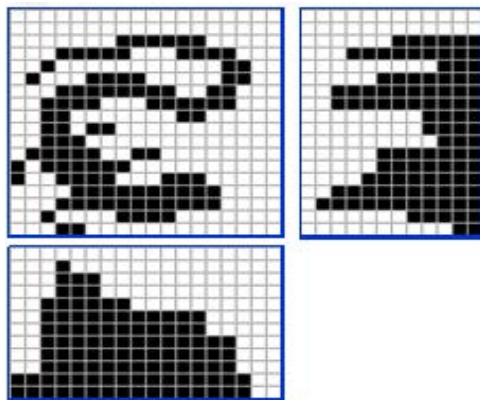


Figure 2. Horizontal and vertical projection
图 2. 水平和垂直方向投影

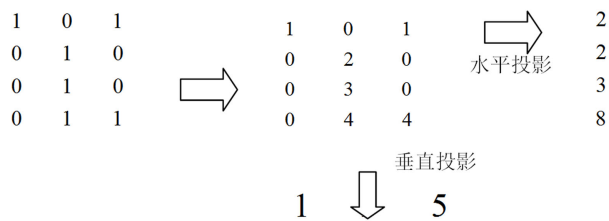


Figure 3. Potential in the vertical direction
图 3. 垂直方向势能

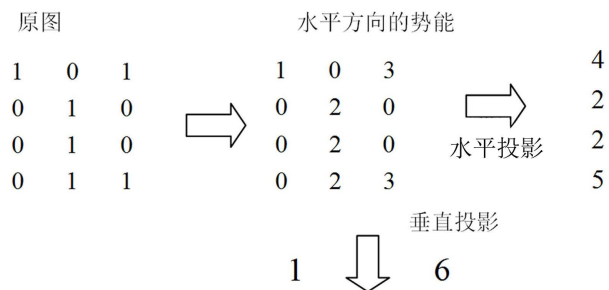


Figure 4. Potential in the horizontal direction
图 4. 水平方向势能

14.6951883042133; 30.6959439883039; 22.6999061621170; 29.7542843350497; 14.4311926605505;
14.5412844036697; 8.12073733243560; 8.18963636280428; 2.05369581043094; 3.46709986839808]。

图 6 显示的是数字 1 的不变距的 13 个特征的大小。

数字 2 的不变距特征: $bubianju2 = [1.69054859157154; 11.0446320281519; 15.3965741740709; 15.7757991761264; 32.1016844108935; 21.6370923441087; 31.5191547719385; 14.4979423868313; 14.4622770919067; 8.11966197545053; 8.28834848639378; 2.10797164696553; 3.53488626570687]$ 。

图 7 显示的是数字 2 的不变距的 13 个特征的大小。

可以看出，两幅图片的不变距特征都是 13 维的向量，为了能让我来识别，我们要对这些数字进行归一化处理，当然，这很好实现，只需找到数组中的最大值和最小值，然后数组中的各个数字减去最小值，比商最大值就能归一化到[0,1]这个区间。

下面，我们再来看一下图片的字符时能。



Figure 5. Handwritten pictures
图 5. 手写体图片

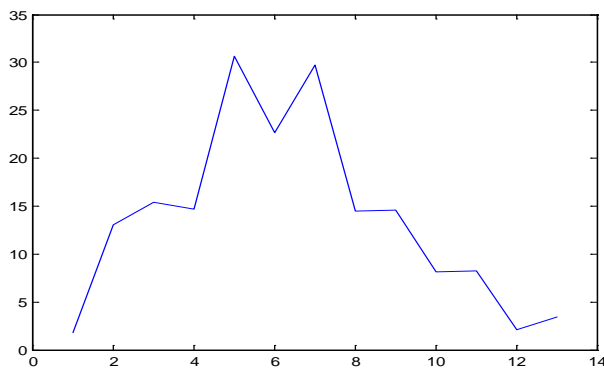


Figure 6. The distribution of moment invariants of number 1
图 6. 数字 1 不变距的分布

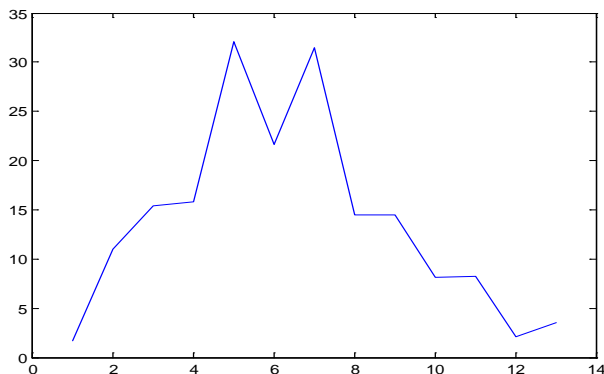


Figure 7. The distribution of moment invariants of number 2
图 7. 数字 2 不变距的分布

Table 1. The recognition correct rate of each number

表 1. 各个数字的识别正确率

数字	1	2	3	4	5	6	7	8	9	0
准确率	100%	81.82%	81.82%	90.91%	54.54%	90.91%	100%	81.82%	100%	81.82%

高，数字 1、7、9 的识别正确率都是 100%，数字 2、3、4、5、6、0 的识别正确率也在 80% 以上，不过数字 5 的识别正确率很低，不到 60%。主要是因为输入的时候，数字 5 的输入不是很规范，而且数字 5 本身和数字 3 和 8 也有一定得相似性，导致你误识。但是，经过特征提取，数字总体的识别正确还是增加了很多，总体的识别正确率达到了 86.36。

项目基金

2013 年国家自然科学基金(编号：61275120)。

参考文献 (References)

- [1] 王耀明 (2002) 图像的矩函数. 华东理工大学出版社, 上海.
- [2] 王惠明, 史萍 (2006) 图像纹理特征的提取方法. *中国传媒大学学报: 自然科学版*, **13**, 49-52.
- [3] Al-Nuaimy, W., Huang, Y., Nakhkash, M., *et al.* (2000) Automatic detection of buried utilities and solid objects with GPR using neural networks and pattern recognition. *Journal of Applied Geophysics*, **43**, 157-165.
- [4] 从爽 (2003) 面向 MATLAB 工具箱的神经网络理论与应用. 中国科学技术大学出版社, 合肥.
- [5] 黄勇, 王崇骏, 王亮等 (2004) 基于形状不变矩的图像检索算法的研究. *计算机应用研究*, **7**, 256-260.
- [6] Freeman, W.T. and Roth, M. (1995) Orientation histograms for hand gesture recognition. *International Workshop on Automatic Face and Gesture Recognition*, **12**, 296-301.
- [7] 陈雪松 (2011) 基于图像势能理论的特征提取方法研究. 华中科技大学博士学位论文.