

论 文
Papers

基于多级神经网络结构的手写体汉字识别*

金连文 徐秉铮

(华南理工大学电子与通信工程系 广州 510641)

摘 要 本文提出了一种用于手写体汉字识别的多级神经网络结构(Multi-stage Neural Network Architecture, MNNA)模型。在该模型中,我们将多个神经网络和不同的特征提取方法有机地集成在一起而构成一个完整的模式识别系统。我们讨论了设计MNNA的一般原理,并提出了一个基于多层前馈神经网络的三级结构的手写体汉字识别实验系统。三种不同的特征提取方法被应用于各级子系统之中。对100个汉字15000个样本的实验我们得到了99.34%的识别率,0.36%的拒识率和0.3%的误识率,表明该模型是十分可行和有效的。

关键词 多级神经网络结构 手写体汉字识别 特征提取 多专家系统 多层前馈神经网络 K-L变换

A Multi-stage Neural Network Architecture for Handwritten Chinese Character Recognition

Jin Lianwen Xu Bingzheng

(Dept of Electronic & Comm Eng, South China Univ of Tech, Guangzhou 510641)

Abstract In this paper, we propose a Multi-stage Neural Network Architecture (MNNA) which integrates several neural networks and various feature extraction approaches into a unique pattern recognition system. General mechanism for designing the MNNA is presented. A three-stage fully connected feedforward neural networks system is designed for Handwritten Chinese Character Recognition (HCCR). Different feature extraction methods are employed at each stage. Experiments show that the three-stage neural network HCCR system has achieved impressive performance and the preliminary results are very encouraging.

Key words multi-stage neural network architecture, handwritten Chinese character recognition, feature extraction, multi-expert system, multilayer feedforward network, K-L transform

1 引言

人工神经网络由于其有效的学习、自组织、模仿人脑智能等方面的能力而越来越多地引起了国内外学者的重视,快速发展的神经网络技术已成功地应用于模式识别、图像处理、计算机视觉、信号处理及自动控制等许多领域中,理论和应用研究的新成果提供了大量的网络模型和

* 国家自然科学基金资助项目

学习算法。因此,神经网络也为字符识别(Optical Character Recognition, OCR)的研究提供了许多工具和方法。事实上,将神经网络方法应用于OCR一直是体现神经网络性能的一个重要应用领域^[1-4]。

近年来,手写体字符识别的研究取得了较大的进展,特别是对于手写体数字和英文字母的识别,已研制出许多识别率高、性能稳定的识别系统^[5-6]。然而,对于手写体汉字识别(Handwritten Chinese Character Recognition, HCCR),由于汉字词汇量大,字形结构复杂,相似汉字众多等因素,HCCR 仍然是一个需要解决的难题。尽管有许多识别算法已成功地用于手写体字符识别^[5-7],各种方法亦有其自身的优点,然而这些方法不可避免地也有一定的局限性和适用条件。很自然地我们会有这样的想法:能否有机地将各种方法结合在一起,各自取长补短,从而有效地提高整体系统的性能,这就是所谓的多专家系统(Multi-expert System)的基本思想^[8],每一种方法我们视为一个小专家系统。近年来,多专家系统的研究已越来越多地引起国内外学者的注意,并已成功的应用于手写字符识别的研究中^[6,8,9]。

基于神经网络技术和多专家系统的结构,在本文中,我们提出了一种多级神经网络结构模型(Multi-stage Neural Network Architecture, MNNA)并将之应用于手写体汉字识别。我们用了100种常用汉字共15000个手写体样本作为实验数据,其中10000字用于训练,另外5000字用于测试。对于测试数据,每级子系统的识别率分别为90.92%、89.58%及88.06%,而整个综合系统所得到的识别率为99.34%,表明这种多级结构非常有效地提高了整体系统的识别性能。

2 多级神经网络结构的设计原理

2.1 系统结构

设计多级神经网络系统的目的在于将各种基于神经网络识别的子系统有机地结合起来,一种简单的多级神经网络结构模型如图1所示:

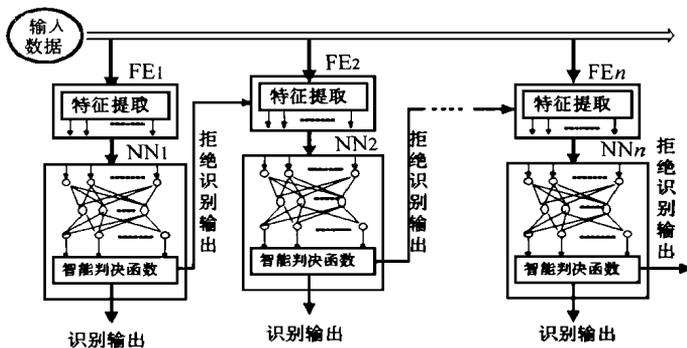


图1 多级神经网络结构示意图

在图1中,每一级子系统由一个特征提取模块 FE_i 和一个神经网络模块 NN_i 构成。特征提取模块对输入数据进行特征分析并提取出一组特征矢量,基于该特征矢量,各神经网络模块在学习阶段,将分别对所有训练数据进行学习和训练;进行识别时,神经网络模块则对输入到本级的数据进行智能的分类判决,判决包括两种情况:识别输出(包括正确识别和错误识别输出)和拒绝识别输出。我们不妨假定:在第 i 级的特征提取模块所提取的特征对于一部分输入样本是十分恰当的,能有效地代表它们的本质特征,因此这一部分样本在该级中将成功地进行

识别, 其余的样本由于特征不明显或含糊不清的将被拒绝识别。被拒识的样本则输入到下一级而进一步进行基于另外一种特征提取和神经网络结构的识别判决。每一级子系统将各自给出所得到的识别结果, 整个系统的拒绝识别输出由最后一级给出。

由上面处理过程我们看到, 设计性能良好的MNNA 关键在于如下两个方面:

- I. 有效的设计每级的特征提取器 FE_i , 尽可能使每级的特征提取方法相互独立, 互为补充。
- II. 合理设计各级神经网络结构 NN_i , 以使得该神经网络分类器能有效地拒绝识别那些在该级特征表征下含糊不清的样本, 从而使误识率尽可能的低。

2.2 MNNA 的性能分析

假设一个 n 级MNNA 系统第 i 级子系统的正确识别率为 r_i , 误识率为 s_i , 拒识率为 u_i , 如果各级的特征提取器 FE_i 相互独立, 则整体系统的识别率为:

$$\begin{aligned} R_n &= r_1 + r_2(1 - r_1 - s_1) + r_3(1 - r_1 - s_1)(1 - r_2 - s_2) + \dots + \\ &\quad r_n(1 - r_1 - s_1)(1 - r_2 - s_2)\dots(1 - r_{n-1} - s_{n-1}) \\ &= 1 - (1 - r_1 - s_1)(1 - r_2 - s_2)\dots(1 - r_{n-1} - s_{n-1}) - \\ &\quad s_1 - s_2u_1 - s_2u_1u_2 - \dots - (s_nu_1u_2\dots s_{n-1}) \end{aligned} \quad (1)$$

错误识别率为:

$$s_n = s_1 + s_2u_1 + s_2u_1u_2 + \dots + (s_nu_1u_2\dots s_{n-1}) \quad (2)$$

拒识率为:

$$U_n = (1 - r_1 - s_1)(1 - r_2 - s_2)\dots(1 - r_{n-1} - s_{n-1}) \quad (3)$$

从式(1)中可见, 为了获得较高的 R_n , 尽可能地降低各级误识率 s_i 是必要的, 但却没有必要使每级的识别率 r_i 都非常高。考虑这样一种理想情况: 假设每级的误识率 s_i 均为 0, 则:

$$R_n = 1 - (1 - r_1)(1 - r_2)\dots(1 - r_{n-1}) \quad (4)$$

此时不难证明:

$$\lim R_n = 1 \quad (5)$$

即在此情况下, 不管 r_i 取 $(0, 1]$ 中的任何值, 只要级数足够多, 则我们总可以达到趋于 100% 的正确识别率。

当然实际情况不可能完全作到使 s_i 均为 0, 但这给我们提供了一条设计多级结构的原则: 即尽可能地降低误识率而不必一定要追求非常高的正确识别率。这个性质是十分有用的, 因为在实际应用中, 设计一个低误识率的系统比设计一个高识别率的系统要容易得多。

作为一个例子, 考虑一个三级识别系统, 设每一级的识别率、误识率和拒识率分别为 80%、1% 和 19%, 则整体系统的识别率为:

$$\begin{aligned} R_3 &= 0.8 + 0.8(1 - 0.8 - 0.01) + \\ &\quad 0.8(1 - 0.8 - 0.01)(1 - 0.8 - 0.01) = 0.9808 = 98.08\% \end{aligned}$$

我们看到:MNNA 的整体性能在很大程度上依赖于每级子系统误识率的高低, 或者更准确地讲, 前几级的误识率尤其关键。

3 特征提取

在大多数模式识别系统中, 特征分析和提取起着十分重要的作用。在我们的实验系统中, 我们设计和应用了加权笔划密度特征、加权外围特征、网格像素概率分布特征等三种不同的特

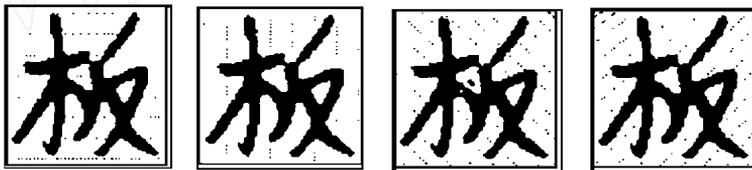
征提取方法。其中,笔划密度和外围特征是汉字结构上的特征,前者着重反映汉字整体上的特点,后者反映汉字外围特点;而网格像素概率分布特征是汉字的统计特征。我们预期通过MN-NA 将这三种不同的特征有机地结合在一起之后,能使它们互为补充,有效地进行识别。

在特征提取之前,我们预先将每个手写汉字样本扫描并归一化为 64×64 的二值图像。设 $f(i, j)$ 为汉字二维点阵图像中位置 (i, j) 处的像素点,有:

$$f(i, j) = \begin{cases} 1, & (i, j) \text{ 是黑像素点} \\ 0, & (i, j) \text{ 是白像素点} \end{cases} \quad 1 \leq i \leq 64, 1 \leq j \leq 64$$

3.1 加权笔划密度特征(Weighted Stroke Cross Counting Feature WSCCF)

笔划密度特征已被广泛地应用于汉字识别的研究中,一个汉字横、竖及左右对角线四个方向的笔划密度特征如图 2 所示,图中我们用一组假想的网线来扫描汉字图像,某条网线穿越笔划的次数即定义为该扫描线上的笔划密度特征。在图(a)和(b)中,8条均匀分布的扫描线被用来分别提取水平和垂直方向的特征;在图(c)和(d)中,16条均匀分布的扫描线被用来分别提取两个对角线方向上的特征。



(a) 水平 (b) 垂直 (c) 左右对角线 (d) 右左对角线

图 2 四方向的笔画密度特征

以往大多数的笔划密度特征都是在固定的均匀网线或非均匀网线上提取,图 2 中的 48 条均匀网线将决定 48 个笔划密度特征。然而对于手写体汉字,由于书写风格随意性大,因此仅仅考虑固定网线上的特征往往会因书写风格的不同而得到的特征不够稳定,为解决这一问题,我们设计了加权笔划密度特征的提取方法,具体作法是:在考察每一条扫描线上的笔划密度特征时,我们同时考虑了其领域特征的加权贡献。为此,我们使用了一权矢量 $W[9] = (0, 2, 0, 4, 0, 6, 0, 8, 1, 0, 8, 0, 6, 0, 4, 0, 2)$ 来对以一网线为中心的左右的邻域的笔划密度特征进行加权求和,作为该网线上的笔划密度特征,我们在实验中发现,加权密度特征要明显优于非加权笔划密度特征^[7]。

设 I_i 为水平方向第 i 条扫描线,则沿该扫描线所得到的加权笔划密度特征定义为:

$$WCCH_i = \frac{1}{2} \sum_{k=-4}^4 W[k+4] \sum_{j=1}^{N-1} f(I_i+k, j) f(I_i+k, j+1) \quad (6)$$

右对角线方向的加权笔画密度定义如下:

$$WCCL_i = \frac{1}{2} \sum_{k=-4}^4 W[k+4] \sum_{j=1}^{N-8i-5} f(I_i+k+j, j) f(I_i+k+j+1, j+1) \quad (7)$$

$$WCCL_{j+8} = \frac{1}{2} \sum_{k=-4}^4 W[k+4] \sum_{i=1}^{N-8j-5} f(i, J_i+k+i) f(i+1, J_i+k+i+1)$$

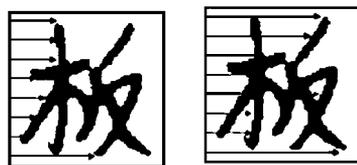
式中, $I_i = 8i + 4, 0 \leq i \leq 7, J_i = 8j + 4, 0 \leq j \leq 7, f(i, j)$ 是 $f(i, j)$ 的非值。垂直方向及左对角线方向的笔画密度特征同理可定义。

3.2 加权外围特征(Weighted Peripheral Feature WPF)

外围特征是汉字的一种轮廓特征^[10],如图 3 所示:一阶外围特征定义为从汉字边缘到垂

直于该边缘方向第一次碰到黑像素点的距离, 二阶外围特征定义为从边缘点到第二次碰到的黑色像素点的距离。

我们分别从汉字上下左右四边 8 个等距分布的边缘点处检测该字符的一阶及二阶外围特征。同样, 我们采用上节所述的加权方法对每一边缘点八邻域的特征进行加权和。这样, 最终将得到 32 个一阶加权外围特征和 32 个二阶加权特征, 构成一个 64 维的特征矢量。实验证明, 加权外围特征要明显优于非加权外围特征^[7]。



(a) 一阶 (b) 二阶

图 3 外围特征示意图

3.3 网格像素分布特征(Grid Pixel Probabilistr Distribution Feature GPPDF)

该特征提取方法类似于文献[11]中的方法, 所不同的是: 我们选用了 8×8 的网格, 而非 10×10 的网格来提取特征, 这样得到了 256 维的特征矢量, 同时使用 K-L 变换对特征空间进行降维, 在保持了 92.6% 的精度度的情况下, 将原来的 256 维特征空间变换到 128 维空间, 比起文献[11]中所用的 400 维特征矢量, 我们所选用的特征维数要小得多, 因而运算速度也就要快得多。

4 实验

4.1 实验系统结构

如图 4 所示, 我们设计了一个具有三级结构的多级神经网络汉字识别系统, 使用 100 个常用字来进行实验, 每一个汉字含有 150 个不同的书写样本, 其中随机选出 100 个字作为训练样本, 余下的 50 个字用于测试, 这样共有 10000 个训练样本和 5000 个测试样本。

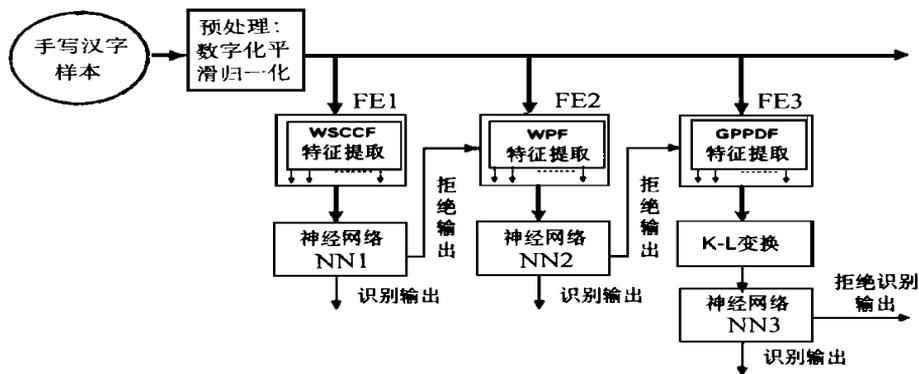


图 4 三级神经网络结构框图

4.2 拒绝识别策略

前面我们提到, 要想使MNN 系统得到良好的性能, 除了合理地提取特征之外, 采用恰当的拒绝策略和设计适当的判决函数也十分关键。一般而言, 在进行模式分类时, 容易被错误识别的样本主要有两种: 一种是该样本在特征空间比较接近两类或多类的分界处, 因而容易被错误地识别, 我们称这类样本为含糊样本(Ambiguous samples), 另一种是该样本在特征空间离任何一类都太远, 以致于不能被所学习出来的分类曲面所覆盖, 我们称这类样本为离异样本

(Outliers)。

设 X 为输入到神经网络的特征矢量, 经过神经网络的运算后, i_1 和 i_2 为两个最激活的输出神经元, f_1, f_2 为它们相应的输出值, 我们使用如下的拒绝策略来判断是否应该拒绝识别 X 。

拒绝策略一: 如果 $f_1 = 0$ 则拒绝 X 。

拒绝策略二: 如果 $|f_1 - f_2| > \lambda$ 则拒绝 X 。

拒绝策略一主要用来处理离异样本, 因为此时 f_1 将可能变得非常小。拒绝策略二主要用来处理含糊样本, 因为此时 f_1 与 f_2 将可能变得十分接近。这里, θ, λ 为预先确定的常量参数。

4.3 实验结果

表 1 各级子系统及整体系统的识别结果

我们用 10000 个手写汉字样本对每一级的神经网络分别进行训练, 学习算法采用标准的 BP 算法, 然后用另外 5000 个样本分别对每一级子系统及整个系统进行识别测试, 实验结果如表 1 所示。

级数	误识率	拒绝率	识别率
1	0.16%	9.12%	90.72%
2	0.22%	10.2%	89.58%
3	0.38%	11.56%	88.06%
整体	0.3%	0.36%	99.34%

实验结果表明:

a、每一级的误识率都较低, 说明拒绝策略是较成功的;

b、尽管各级的识别率都不是很高(最高的仅为 90.72%), 然而整体系统的识别率却十分高(99.34%), 可见多级结构有机地将三种子系统结合在一起之后非常有效地提高了系统的识别率, 同时这也说明我们所选用的三种特征提取方法能有效地互为补充, 是比较恰当的。

实验在 Sun Sparc-20 工作站上进行, 训练每一级神经网络的时间约为 3~ 4 小时, 整体系统的识别速度约为 25 字/秒。

5 结束语

本文提出了一种多级神经网络结构并将之应用于手写体汉字识别, 初步的实验表明这种结构能非常有效地提高系统的识别率, 取得了令人满意的结果。

本文的实验只是在较小的词汇量上进行的, 对于大词汇量的手写汉字识别(如识别国标一级字库 3755 个汉字), 采用这种基于神经网络的多级结构将可能由于目前硬件容量、速度等方面因素的限制而变得不太切实可行。在此情况下, 一方面可以采用一个较好的预分类算法来将汉字粗分为一系列的小词量集之后再用此方法进行识别; 另一方面可选用或设计快速学习的神经网络模型(如 Radial Basis Function 等), 将值得我们更进一步的研究。

致 谢

香港大学陈国平博士提供了实验数据并给予了許多帮助, 谨此表示感谢。

参 考 文 献

- 1 Skrzypek J, Hoffman J. Visual Recognition of Script Characters and Neural Network Architectures Neural Networks: Advance and applications Gelenbe E ed USA: Elsevier Science Pub, 1991: 109~ 114
- 2 Guyon I Application of neural networks to character recognition International Journal of Pattern Recognition and Artificial Intelligence, 1991 5(3): 353~ 382
- 3 Chang H D, Wang J F, Kuo S C A Bayesian neural network for separating similar complex handwritten Chinese characters Pattern Recognition letters, 1994 15(4): 403~ 408

- 4 Francoise F S, Viennet E, Lamy B. Multimodular neural network architectures: applications in optical character and human face recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 1993 7(4): 721~ 755
- 5 Mori S. Historical Review of Theory and Practice of Handwritten Character Recognition. In: *Fundamentals in Handwriting Recognition*. Edited by Sebastiano and Impedovo. Berlin: Springer-Verlag, 1994, 43~ 69
- 6 Suen C Y, Legault R, Nadal C, Cheriet M, Lam L. Building a new generation of handwriting recognition system. *Pattern Recognition Letters*, 1993, 14(4): 303~ 315
- 7 金连文. 手写体汉字识别的研究 [博士学位论文]. 华南理工大学, 1996
- 8 Marek W K. On the multistage bayes classifier. *Pattern Recognition* 1988 21(4): 355~ 357
- 9 Huang Y S, Suen C Y. Combination of multiple classifiers with measurement values. In: *Proceedings of the Third International Conference of Document Analysis and Recognition*, Japan, 1993, 598~ 601
- 10 张忻中. 汉字识别技术 清华大学出版社, 1992
- 11 Li T F, Yu S S. Handprinted Chinese character recognition using the probability distribution feature. *International Journal of Pattern Recognition and Artificial Intelligence*, 1994, 8(5): 1241~ 1258
(1995-11-21 收到, 1997-01-24 改定)

www.cnki.net