

文章编号:1007-2780(2011)01-0116-07

双 X 特征下自组织竞争网络的手写字符识别

仲崇亮^{1,2}, 丁亚林¹, 付金宝^{1*}

(1. 中国科学院 长春光学精密机械与物理研究所, 吉林 长春 130033;

2. 中国科学院 研究生院, 北京 100039)

摘 要: 对图像预处理、字符特征提取和神经网络类型的选择、优化、训练及调整等进行研究, 提出运用双 X 特征作为输入的自组织竞争网络对手写字符进行识别的方法。首先进行图像预处理, 包括图像的灰度化、二值化、剪裁及缩放等。然后, 提出字符的双 X 特征提取方法, 并运用该方法提取各个字符的特征信息。最后, 将字符特征数据送入自组织网络中进行学习, 反复试验确定理想网络参数, 使其可以自组织地将各字符模式相区分。实验结果表明: 这种方法减少送入网络的数据量, 降低了冗余信息对网络的干扰, 使网络处理的复杂程度大大降低, 训练后的网络误差小, 网络的错判率约为 12%, 比传统方法有明显提高。

关 键 词: 模式识别; 图像处理; 特征提取; 神经网络; 自组织竞争网络

中图分类号: TP391.4 文献标识码: A DOI: 10.3788/YJYXS20112601.0116

Handwritten Character Recognition Based on Self-Organizing Network with Double-X Feature

ZHONG Chong-liang^{1,2}, DING Ya-lin¹, FU Jin-bao¹

(1. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China;

2. Graduate University of Chinese Academy of Sciences, Beijing 100039, China)

Abstract: By studying of image processing, character feature extraction and neural network including network type selection, network parameter optimization, training and adjustment, the handwritten character recognition method based on self-organizing network with double-X feature as input is proposed. First of all, the image is preprocessed including gray, binaryzation, tailoring and scaling. Second, the characters' double-X feature extraction method is utilized to extract feature information of various characters. Finally, the feature data of characters is input into self-organizing network for learning and to determine the ideal network parameters after many tests, so that it can be self-organized to distinguish the character-mode. Experimental results show that this approach reduces the amount of data inputed to the network, and the interference of redundant information to network, and the complexity of network processing was considerably reduced. The network training error is small and the network misjudged rate is just about 12%.

Key words: pattern recognition; image processing; feature extraction; neural network; self-organizing network

收稿日期: 2010-05-21; 修订日期: 2010-07-08

基金项目: 装备预研基金项目 (No. 9140A25060208ZK1003)

作者简介: 仲崇亮(1985-), 男, 黑龙江哈尔滨人, 硕士研究生, 研究领域为神经网络、图像识别、航空相机结构优化。

1 引 言

机器学习的一大目标是要从数据中学习其相关的规律性^[1]。认知学在计算机科学中最重要的产物就是如何使工具更为人性化,与之密切相关的领域包括了人机对话中手写字符的识别技术^[2]。手写字符的识别属于模式识别的范畴,在模式识别的研究中,盖然论是其最基本的方法,其中 Po^[3]和 Liou 及 Yang^[4]曾用贝叶斯决策理论去研究模式识别问题,并取得了一定成效。其他方法包括 Didaci 和 Giacinto^[5]在动态分类选择中用到的 k-最近邻分类方法等。决策论的方法实际上就是统计的方法。在决策理论方法中特征抽取尚无通用的理论指导,特征抽取后从特征空间映射到决策空间时,要引入鉴别函数。这样一来就使得该方法在形成通用算法之前要经过大量的统计计算,工作量大。

神经网络具有许多优良特性,如高度并行性、良好的容错能力和易于硬件实现等^[6]。更为重要的是神经网络具有自学习能力,对于具体问题而言,可以仅考虑输入输出内容,而不必对内部的具体算法进行干预,将神经网络用于模式识别领域显然较为合适。在字符识别中,最常用的就是人工神经网络及特征提取技术^[7]。在特征提取上,常规方法往往先对字符图像进行预处理,然后将其数字化,逐行存入向量作为训练集。对于手写字符而言这种方法使得送入网络的数据量大,处理过程复杂,另一方面,送入网络的冗余信息增加了网络训练难度,同时也使网络产生了更多的干扰。

基于对上述常规方法的考虑,本文提出了双 X 特征提取方法,该方法在保证实现模式分类的基础上,减少送入网络的数据量,降低了冗余信息对网络的干扰,使网络处理的复杂程度大大降低,训练后的网络误差小。误差负反馈网络需要训练较长时间才能记住所有反馈于网络的似然输入向量,同时又经常会由于泛化能力差而给出错误的判断^[8],可见解决网络的泛化能力是研究问题的关键。为了避免给出不良误导,本文选用自组织网络来实现模式分类,由于网络并不给出目标,而完全是在无教师的条件下学习的,一旦网络学习到各个特征并将其自动分类后,网络便天然具有了良好的泛化能力。

2 双 X 特征提取

2.1 图像预处理

受光线、纸张、笔及相机等因素影响,手写字符图像可能并不清晰。因此,在特征提取前需进行图像预处理,以减少图像冗余信息、去除噪声。通常字符图像为 RGB 多色彩形式,对图像进行灰度化处理,将其从三维矩阵转化到二维矩阵,经过灰度化处理后的手写字符图像如图 1 所示。



图 1 灰度化处理后的手写字符图像

Fig. 1 Handwritten character images after gray processing

为完成字符识别,采用基于直方图的方法^[9]对灰度化处理后的手写字符图像进行二值化处理,以使字符对象与背景分割开,二值化处理后的手写字符图像如图 2 所示。



图 2 二值化处理后的手写字符图像

Fig. 2 Handwritten character images after binarization

进一步对手写字符图像进行剪裁处理,去除周围空白,将字符满格置于图像的中心。以手写字符“A”为例,其预处理前后的图像数据矩阵分别如图 3(a)和图 3(b)所示。

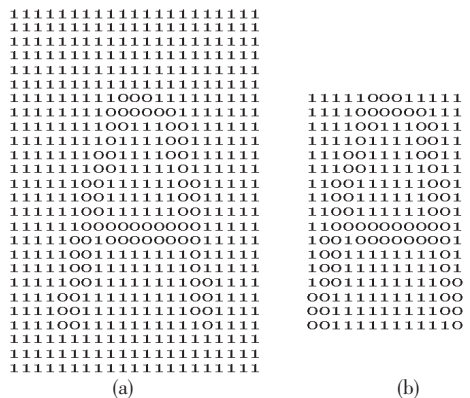


图 3 预处理前后手写字符“A”的图像数据矩阵对比
Fig. 3 Contrasts of the matrices corresponding to handwritten character “A” before and after image pre-processing

图像预处理滤除了图像的噪声、提高了图像的分辨性能、降低了字符图像数据矩阵的复杂性。

2.2 字符的双 X 法特征提取

在保证实现模式分类前提下,为简化神经网络的学习并提高网络识别能力,手写字符特征提取的数据应尽量精炼。字符特征提取前,应保证手写字符图像的数据矩阵维度相同,用缩放方式将各图像进行归一化处理,即:按特定高宽比,采用最邻近插值方法将图像缩放到相同大小。经反复验证,本文确定将图像矩阵缩放到 20×18 的数据矩阵较为适合。归一化后,各字符的笔画粗细不尽相同,为使笔画粗细相同,对各字符图像进行了骨骼化处理:在不打断字符笔画的前提下,移除字符边缘像素点,只保留字符笔画的骨架部分,使各字符的笔画宽度均只占一个像素点。归一化和骨骼化处理后的手写字符图像如图 4 所示。

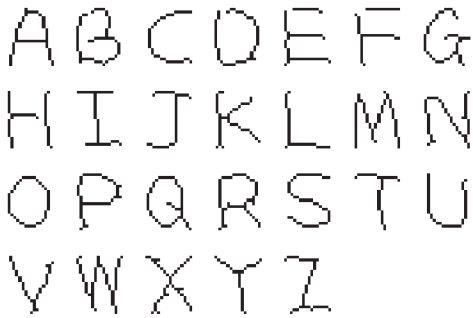


图 4 归一化和骨骼化处理后的手写字符图像

Fig. 4 Images of characters after normalization and skeleton processing

双 X 字符特征提取思想描述如下:首先对手写字符数据矩阵进行有代表性的特征区域划分和组合,然后统计各个特征区域中的特征值(即该特征区域中代表字符笔画的数字“0”的个数),各个特征区域中特征值的组合为特征向量,由特征向量表征字符特征。

本文提出的手写字符区域划分和组合方法如图 5 所示。

在图 5 中,首先将整幅图像均分为 8 个矩形特征区域,然后将特征区域 1 和 4 的组合命名为特征区域 9,将特征区域 2 和 3 的组合命名为特征区域 10,将特征区域 6 和 7 的组合命名为 11 区,将特征区域 5 和 8 的组合命名为特征区域 12,将特征区域 1~8 的组合即整幅图像命名为特

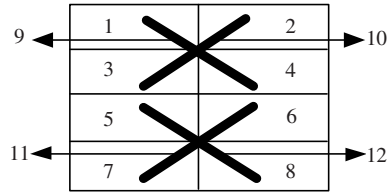


图 5 手写字符数据矩阵区域划分和组合

Fig. 5 Area partition and combination of handwritten Character

征区域 13。以手写字符“A”为例,对手写字符“A”的数据矩阵实施上述特征区域划分和组合方法如图 6 所示,依次统计特征区域 1~13 上的特征值(特征区域中“0”的个数),将特征值组合为特征向量 η_A ,由 η_A 表征手写字符“A”的特征。

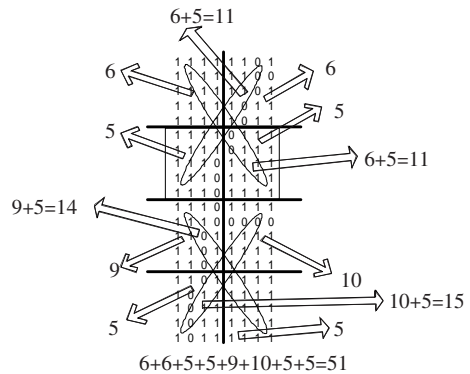


图 6 手写字符“A”的特征提取

Fig. 6 Characteristic extraction of handwritten character 'A'

表征手写字符“A”的特征向量 η_A 为: $\eta_A = [6 \ 6 \ 5 \ 5 \ 9 \ 10 \ 5 \ 5 \ 11 \ 11 \ 14 \ 15 \ 51]$,按同样的方法可对任意手写字符进行特征提取,得到 $\eta_A, \eta_B, \dots, \eta_Z$ 。本文将这种特征提取方法命名为双 X 特征提取方法。

2.3 基于双 X 特征的模式分类

本文选用自组织竞争网络,其模式分类是通过权矩阵分量学习输入模式来完成的,各输入向量间的距离越大越有助于网络学习。但是过大的距离又往往伴随更多的数据量,导致网络学习的复杂程度增加、干扰增多、学习时间增长、效率下降以及调整参数困难等问题。在权衡利弊的基础上,提出上述双 X 特征提取方法,其输入特征向量仅 13 维,网络完全可接受。

这里不失一般性地将某个字符特征向量表

示为:

$$\eta_i = [l_{i,1} \quad l_{i,2} \quad \dots \quad l_{i,13}] \quad (1)$$

式中: η_i 代表字符的特征向量, ($i = 1, 2, \dots, 26$); $l_{i,j}$ 为某个字符特征区域 j 中的特征值 ($j = 1, 2, \dots, 13$)。

对于由自然特征向量组 $\xi_1, \xi_2, \dots, \xi_{13}$:

$$\begin{cases} \xi_1 = [1 \quad 0 \quad 0 \quad \dots \quad 0 \quad 0]_{1 \times 13} \\ \xi_2 = [0 \quad 1 \quad 0 \quad \dots \quad 0 \quad 0]_{1 \times 13} \\ \vdots \\ \xi_{13} = [0 \quad 0 \quad 0 \quad \dots \quad 0 \quad 1]_{1 \times 13} \end{cases} \quad (2)$$

对所张成的 R^{13} 空间而言, 显然对于 $\forall \eta \in R^{13}$, 在 R^{13} 空间中定义的 Euclid 范数为:

$$\|\eta_i\|_2 = \sqrt{\left(\sum_{j=1}^{13} |l_{i,j}|^2\right)} \quad (3)$$

则 $(R^{13}, \|\cdot\|_2)$ 成为赋范线性空间, 在此范数基础上, 可以诱导出距离:

$$d^{13}(\eta_i - \eta_{i+m}) = \|\eta_i - \eta_{i+m}\| = \sqrt{\sum_{j=1}^{13} |l_{i,j} - l_{i+m,j}|^2} \quad (4)$$

式中: d^{13} 表示 13 维特征向量间的距离, η_i, η_{i+m} 分别代表两个字符的特征向量, ($m = 0, 1, 2, \dots, 25$), $m=0$ 则表示相同字符的特征向量。

根据双 X 特征提取的方法, 如果仅对 1~8 矩形特征区域进行特征值统计提取, 即: 对任意字符向量 $\eta_i = [l_{i,1}, l_{i,2}, \dots, l_{i,13}]$ ($l_{i,9} = l_{i,10} = \dots = l_{i,13} = 0$)。此时计算任意两个字符间距离时, 相当于在 R^8 空间中进行计算。由公式(4)可以计算出任意两个字符间的距离。通过大量计算验证, 可以得到任意两个字符间的距离的最小值为:

$$\begin{aligned} \text{mind}^8(\eta_i - \eta_{i+m}) &= \min \|\eta_i - \eta_{i+m}\| = \\ \min \sqrt{\sum_{j=1}^8 |l_{i,j} - l_{i+m,j}|^2} &= 2.236 \quad (5) \end{aligned}$$

在原 8 维的特征向量的基础上, 增加双 X 特征变为 12 维后, 根据距离公式(4)及各字符特征向量可以计算任意两个字符间的距离, 这些距离中的最小值:

$$\begin{aligned} \text{mind}^{12}(\eta_i - \eta_{i+m}) &= \min \|\eta_i - \eta_{i+m}\| = \\ \min \sqrt{\sum_{j=1}^{12} |l_{i,j} - l_{i+m,j}|^2} &= 3.316 \quad (6) \end{aligned}$$

可见, $\text{mind}^{12}(\eta_i - \eta_{i+m}) > \text{mind}^8(\eta_i - \eta_{i+m})$, 且最小距离的位置与前次相同, 均为字符 G 和 Q 的距离。

用同样的方法来验证空间由 12 维扩展为 13

维的情况, 经计算可得:

$$\begin{aligned} \text{mind}^{13}(\eta_i - \eta_{i+m}) &= \min \|\eta_i - \eta_{i+m}\| = \\ \min \sqrt{\sum_{j=1}^{13} |l_{i,j} - l_{i+m,j}|^2} &= 4.358 \quad (7) \end{aligned}$$

显然, 如果继续将矩阵的维数扩大则可以进一步拉大各个字符特征向量之间的距离, 具体可表示为:

$$\text{mind}^{N_1}(\eta_i - \eta_{i+m}) > \text{mind}^{N_2}(\eta_i - \eta_{i+m}) \quad (8)$$

式中: N_1 和 N_2 均为正整数, 且 $N_1 > N_2$, 表示向量的维度。

根据前面论述, 在可识别前提下, 希望数据量较小, 分析及实验表明, 在 13 维特征向量下可以实现各手写字符的模式分类。

3 在自组织竞争网络下实现模式识别

3.1 基本思想

神经网络在众多领域均得到了广泛应用^[10-15]。在图像处理及模式识别等方面, 神经网络更是发挥了显著作用^[6,16]。如前所述: 字符特征向量之间实现模式分类的关键是距离, 所以在选择网络类型时应考虑用一种擅长分辨距离的网络。BP 网络和 RBF 网络常常使能量函数陷入局部最小而非全局最小^[17]。自组织竞争网络中的各权矩阵可以用距离接近的方式去学习各种不同类型的输入向量, 如果输入向量之间有良好的距离关系, 经过训练后的网络便可以自动将不同的模式进行区分。因此, 本文提出在双 X 法特征提取下, 用自组织竞争网络对字符进行模式分类, 达到手写字符识别目的。

3.2 网络结构和学习算法

自组织竞争网络具有简单的网络结构和学习算法, 其连接形式如图 7 所示。

其中 $i = 1, 2, \dots, N, j = 1, 2, \dots, M$, 输入层权

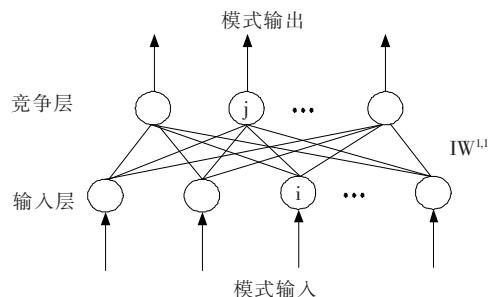


图 7 自组织竞争网络连接形式

Fig. 7 Connection form of organization competition network

矩阵 $IW^{1,1}$ 的分量 $w_{i,j}$ 满足:

$$\sum_{i=1}^N w_{i,j} = 1 \tag{11}$$

在竞争层中,只有惟一的获胜神经元,其输出为 1,其余神经元均输出 0,获胜神经元表明该模式所对应的分类。图 8 为自组织竞争网络的网络结构图。

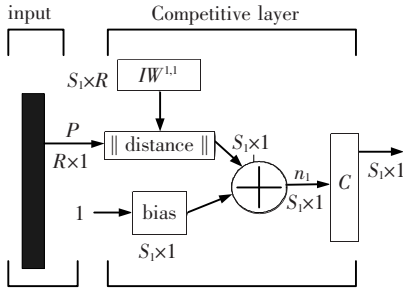


图 8 自组织竞争网络结构图

Fig. 8 Structure drawing of organization competition network

图 8 中: S_1 为竞争层神经元的个数; R 为输入向量的维数; P 为输入向量的个数。框图中 distance 接收 $IW^{1,1}$ 和 P , 并计算 $IW^{1,1}$ 中各行与 P 中各分量的距离:

$$\|dis\| = [IW_1^{1,1} - p, IW_2^{1,1} - p, \dots, IW_{S_1}^{1,1} - p]^T \tag{12}$$

在这些距离中,数值最小的神经元将成为竞争网络中的胜利者,从而以 1 的形式作为网络的最终输出,而其他神经元均输出 0。胜利神经元的权值调整,即学习算法为 Kohonen 学习法则:

$$iW_1^{1,1}(q) = iW_1^{1,1}(q-1) + \alpha[p(q) - iW_1^{1,1}(q-1)] \tag{13}$$

式(13)表示的是第 i 个神经元获胜,其权值的调整方式,式中的 α 为学习参数: $0 < \alpha \ll 1$ 。 n_1 接收 $\|dis\|$ 及 bias, 并将二者作和, 即: Kohonen 学习法则让权矩阵中的分量去学习输入向量的形式, 该网络可以实现模式识别。从其学习算法中不难看出, 权向量最接近输入向量的神经元将取胜, 同时经过调整后当遇到类似输入时, 具有更大取胜的可能性。

由于可能存在一些权向量的初始值远离任何一种输入模式的“死元”, 最终导致网络无论经过多长时间的训练, 这些神经元均无法取胜。所以在给出神经元余量的同时, 应该引进 bias, 即偏

置。这些偏置将给予“死元”获得胜利的机会, 从而达到优化网络的目的。因此, 送入转移函数 f 中的并非 $\|dis\|$, 而是 n_1 , 且:

$$n_1 = \|dis\| + bias = [IW_1^{1,1} - p, IW_2^{1,1} - p, \dots, IW_{S_1}^{1,1} - p]^T + bias \tag{14}$$

将 n_1 送入转移函数进行自组织学习, 学习后的网络中的权矩阵中的各分量将学习到输入向量, 即各个字符的特征, 当输入类似时, 网络便可以自动判断, 并给出结果。

3.3 网络训练

选择 5 组由不同人书写的共 130 个手写字符作为网络的训练集, 训练过程中随着实验的进行, 对结果进行分析, 并且适当调整训练集中元素, 同时修正网络的各个训练参数。最终使网络可以将训练集分为 26 类, 且同一字母归为一类。仿真结果: 26 个英文字符对应的神经元编号如表 1 所示, 该编号也代表了该字母所对应的类别, 即模式。

表 1 字母对应的神经元编号

Table 1 Neuron serial numbers corresponding to characters

字符	模式	字符	模式
A	9	N	18
B	5	O	19
C	26	P	25
D	24	Q	3
E	7	R	23
F	8	S	17
G	16	T	22
H	21	U	20
I	12	V	14
J	13	W	15
K	10	X	2
L	4	Y	1
M	11	Z	6

从表 1 中可以看到, 在不给定目标的前提下, 由于该网络中权向量分量分别学习输入模式, 而输入模式又是距离可分的, 这样一来, 网络便可以自组织地将输入向量进行自动分类, 并最终给出每个输入向量所对应的模式代号。由于网络可以将不同的字符自组织地分成 26 组, 可见网络的确学习到了各个字符的特征, 对 26 个英文手写字符模式可分。

4 实验验证

4.1 学习能力测试

另取训练集以外的 26 个字母送入网络,检验网络的输出,实验结果如表 2 所示。

表 2 模式分类结果

Table 2 Pattern classification result

字符	初始模式	网络判断	字符	初始模式	网络判断
A	9	9	N	18	18
B	5	5	O	19	19
C	26	26	P	25	25
D	24	24	Q	3	3
E	7	7	R	23	3
F	8	8	S	17	17
G	16	16	T	22	22
H	21	21	U	20	14
I	12	12	V	14	14
J	13	13	W	15	15
K	10	10	X	2	2
L	4	4	Y	1	1
M	11	11	Z	6	6

从表 2 中可以看到,网络出现了两个误判:将 U 判断为 V,将 R 判断为 Q。试验中重新给网络 10 个新的手写字母 U,网络给出了 8 次正确的判断。由此可以看出,虽然网络容易将这两个字符误判,但对于已训练的网络而言,这两个字符的确是可分的。

4.2 误判率测试

随机抽取训练集外的 500 个样本送入网络,

检验网络的输出,并与实际模式进行对比,以检测网络的识别能力。误判率如图 9 所示。

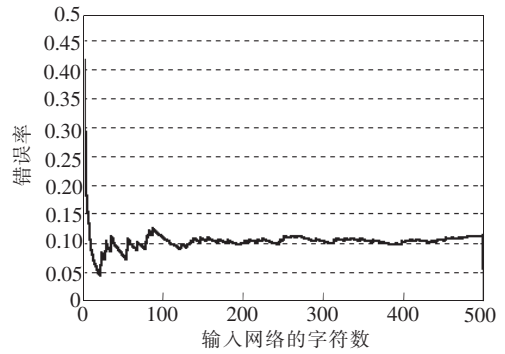


图 9 误判率曲线

Fig. 9 Misjudgment ratio curve

图 9 中误判率为:

$$K_{err} = \frac{n}{N} \times 100\% \quad (15)$$

式中: n 表示误判的字符数; N 表示当前输入网络的字符总数。从图中可以看到,网络的误判率约为 12%。

5 结 论

双 X 特征法对手写字符进行特征提取可以大大减少送入网络的数据量,降低了冗余信息对网络的干扰,使网络处理的复杂程度大大降低,训练后的网络误差小;自组织神经网络可以在无教师监督的条件下进行学习,经过反复试验及参数调整,训练后的网络一旦对于手写字符可分,便天然具有了良好的泛化能力。实验结果表明:网络的误判率约为 12%,比以往方法有了很大提高(以往方法错误率约为 23%~13%^[8])。

参 考 文 献:

[1] 杨晓敏,吴炜,何小海,等. 应用 SLLE 实现手写体数字识别[J]. 光学精密工程,2009,17(3):641-647.

[2] Faaborg A J. Using neural networks to create an adaptive character recognition system[R]. Ithaca NY: Cornell University, 2002.

[3] Wu P H. Handwritten character recognition [D]. Queensland, Australia: School of Information Technology and Electrical Engineering, the University of Queensland, 2003.

[4] Liou C Y, Yang H C. Hand printed character recognition based on spatial topology distance measurement[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18:941-945.

[5] Didaci L, Giacinto G. Dynamic classifier selection by adaptive k-nearest-neighbourhood rule[J]. Lecture Notes in Computer Science, 2004,3077:174-183.

[6] 姜帅,丁振良,袁峰,等. 应用 Hopfield 神经网络和小波域隐 Markov 树模型的图像复原[J]. 光学精密工程,2009,

- 17(11):2828-2834.
- [7] Brown E W. Applying neural networks to character recognition[EB/OL]. [2010-05-15]. <http://www.ccs.neu.edu/home/feneric/charrecnn.html>.
- [8] Ganapathy V, Liew K L. Handwritten character recognition using multiscale neural network training technique[J]. *World Academy of Science, Engineering and Technology*, 2008, 39:32-37.
- [9] 张建伟,张启衡. 基于块遍历的直线边缘特征提取[J]. *光学精密工程*, 2009, 17(3):662-668.
- [10] 潘国兵,颜国正,张明卿,等. 概率神经网络与差异演化在胶囊内窥图像出血识别中的应用[J]. *光学精密工程*, 2010, 18(6):1435.
- [11] 朱福珍,李金宗,朱兵,等. 基于径向基函数神经网络的超分辨率图像重建[J]. *光学精密工程*, 2010, 18(6):1451.
- [12] 赵章荣,邬义杰,顾新建,等. 用神经网络结构实现超磁致伸缩智能构件滑模控制[J]. *光学精密工程*, 2009, 17(4):778-786.
- [13] 赵章荣. 基于神经网络的超磁致伸缩智能构件滑模控制[J]. *光学精密工程*, 2009, 17(4):778-786.
- [14] 耿洁,刘向东,陈振,等. Preisach 迟滞逆模型的神经网络分类排序[J]. *光学精密工程*, 2010, 18(4):855-862.
- [15] 沈铨武,王志乾,刘畅,等. BP 神经网络在多位置捷联寻北系统中的应用[J]. *光学精密工程*, 2009, 17(8):1890-1895.
- [16] 邓超,张涛,姚清华. 应用小波神经网络处理 CCD 图像噪声[J]. *光学精密工程*, 2008, 16(2):345-351.
- [17] 张煜东,吴乐南. 用改进的 Paik 型 Boltzmann 机实现图像复原[J]. *光学精密工程*, 2009, 17(7):1737-1744.

学术论文英文摘要写作质量标准(试行)

为了让中文学术期刊学术论文以更高的质量走向国际,以利于国际科技界对中国科技事业的了解和交流,有必要规范英文摘要的质量标准:

一、英文摘要是应用符合英文语法的文字语言,以提供文献内窗口梗概为目的,不加评论和补充解释确切地论述文献重要内容的短文。

二、英文摘要必须符合“拥有与论文同等量的主要信息”的原则。为此,英文摘要应重点包括 4 个要素,即研究目的、方法、结果和结论。在这 4 个要素中,后 2 个是最重要的。在执行上述原则时,在有些情况下,英文摘要可包括研究工作的主要对象和范围,以及具有情报价值的其他重要的信息。当前学术期刊上英文摘要的主要问题是要素不全,繁简失当。

三、英文摘要的句型力求简单,通常应有 10 个左右意义完整、语句顺畅的句子。

四、英文摘要不应有引言中出现的内容,也不要对论文内容作诠释和评论,不得简单重复题名中已有的信息;不用非公知公用的符号和术语,不用引文,除非该论文证实或否定了他人已发表的论文,缩略语、略称、代号,除了相邻专业的读者也能清楚理解的以外,在首次出现时必须加以说明;科技论文写作时应注意的其他事项,如采用法定计量单位,正确使用语言文字和标点符号等,也同样适用于英文摘要的编写。

摘自:中国科协学会学术部 2002.9 通知