

用于手写体数字不变性特征提取 及识别的自组织算法

林 列,常胜江

(南开大学现代光学研究所,天津 300071)

摘 要: 针对无约束手写体数字的特点,本文提出了一种用于手写体数字特征提取和识别的神经网络与视觉学习相结合的自组织学习算法,为了避免网络规模太大引起“过度训练”现象而导致网络的推广能力下降,提出了一种在学习过程中自适应删剪不重要权重的方法.计算机模拟结果显示该算法能有效地提取手写体数字的不变性特征,进而提高对无约束手写体的识别率.

关键词: 模式识别; 手写体; 手写体数字识别; 特征提取

中图分类号: TP183 **文献标识码:** A **文章编号:** 0372-2112 (2003) 10-1506-04

A Self-Organizing Algorithm for Invariance Extraction and Recognition of Handwritten Digits

LIN Lie, CHANG Sheng-jiang

(Optical Information Science Laboratory, Nankai University, SEC, Tianjin 300071, China)

Abstract: A self-organizing learning algorithm is proposed to perform invariance extraction and recognition of handwritten digits. In order to build adaptively an optimal network topology in the learning procedure for a given recognition problem so as to avoid overtraining, an adaptive pruning method is proposed. Computer simulation show that the proposed algorithm is effective on extracting invariance of handwritten digits and optimizing a network topological architecture.

Key words: pattern recognition; handwritten digit; handwritten digit recognition; feature extraction

1 引言

由于无约束手写体是由没有经过专门训练的人书写的,因此存在很大的畸变及尺度上的变化,使得它的分类变得非常困难.据邮电部门提供的数据表明,目前我们国家信件自动分检的准确率不到30%,主要原因是手写体数字的识别率太低.人工神经网络是一种仿脑信息处理模型,具有不依赖先验知识或规则为前提的自适应学习能力,这种学习能力决定了神经网络会为解决一些复杂的模式识别问题提供一种有效的方法.人们已经提出了多种神经网络模型和算法^[1,2]来自组织地提取模式的不变性表示,但对有较大平移、畸变的手写体样本的识别结果并不令人满意.为此,人们进一步提出采用稀疏迹网络结构模型^[3]来提取无约束手写体的不变性特征信息,取得了较好的效果,但由于网络训练采用有约束条件的背传(BP)算法,网络收敛速度慢,如果学习参数选择不当,会导致无法收敛或者使训练陷入局域最小,从而影响不变性特征的稳定性及识别结果.

用神经网络方法解决实际问题时,除了寻找适当的网络

权重,另一个重要的问题是确定适当的网络规模,如果网络规模过大,会引起过度训练(Overtraining)现象出现,导致网络的推广能力下降,即训练后的网络对训练样本的识别率很高,但对非训练样本却很低.解决此问题的有效方法是使用权重删剪方法^[4,5],但已有的方法需要额外的输入、输出样本对来计算Hessian矩阵,因此,这些方法不适用于在线学习情况.

为了有效地提取不变性特征以实现无约束手写体数字进行识别,我们把局域的广义卡尔曼滤波算法(Extended Kalman Filter)同视觉学习规则相结合用来自组织地提取模式的不变性特征.由于卡尔曼滤波算法自动估算出权重调整的最佳步长值,且需要调整的参数较少,同BP算法相比,网络的学习效率及收敛速度都会有较大提高.在上述算法的基础上,我们提出了一种权重的自适应删剪算法,通过删除不重要的权重,得到一个紧凑的网络结构.同时,为了提高不变性特征的稳定性,我们提出采用周期包函数作为激活函数.模拟结果表明该算法及周期包函数能够提高网络的收敛速度及不变性特征的稳定性.

2 激活函数

激活函数是影响网络收敛行为的一个重要的因素. 文[6]提出了一种 sigmoid 型周期函数, 由于该函数能把输出空间划分出无数个等间隔的子空间, 因此能显著地提高网络的分类能力和加快了网络的收敛速度. 同时, 由于该函数提供了较为平坦的收敛区域, 网络收敛的稳定性得到可靠的保证. 针对模式识别的特点, 我们提出使用周期包函数作为激活函数来提取模式的不变性特征. 周期包函数由多个不同振幅 $\{h_n\}$ 、斜率 $\{b_n\}$ 、周期 $\{r_n\}$ 、和位移 $\{c_n\}$ 的周期函数组成. 定义为:

$$f(\text{net}_j) = \sum_{n=1}^k \frac{h_n}{1 + e^{-b_n \sin(r_n \cdot \text{net}_j) + c_n}} \quad (1)$$

其中 net_j 为第 j 个神经元的净输入. 适当选取 K , h_n , b_n , r_n 和 c_n 的值, 周期包函数同 sigmoid 函数一样能够提供较平坦的收敛区域, 同时在中间区域提供一个慢变化的区域以降低模式变化对不变性特征的影响, 因此, 该激活函数既保证网络快速收敛又能保证不变性特征的稳定性.

3 自组织学习算法

迹学习规则^[1]首先被用于自组织的不变性替换, 又推广应用于神经网络的学习中. 迹学习规则可以归纳如下^[1,2]:

$$\begin{aligned} T_{mi}(t) &= \eta T_{mi}(t-1) + (1-\eta) y_i(t) \\ y_i(t) &= \psi \left[\sum_j w_{ij} x_j(t) \right] \end{aligned} \quad (2)$$

其中, $T_{mi}(t)$ 为输出层第 i 个神经元对第 m 类输入模式的迹, η 是一个控制参数, 用以控制迹和新的输入之间的相互影响, f 为激活函数, 函数 ψ 用来实现一个区域内的横向抑制, $y_i(t)$ 是第 i 个输出神经元的输出, 表 $x_j(t)$ 示第 t 个输入模式.

我们采用的网络结构为 3 层前馈型网络, 输入层、隐藏层和输出层的神经元数目分别为 N^i 和 N^h 和 N^o 个. 隐藏层来自组织地提取不变性特征信息, 而输出层利用隐藏层提取的不变特征信息对模式分类. 输入层到隐藏层为局域互连, 由式(2)中的函数 ψ 确定互连范围, 假设输入层有 L_i 个神经元与隐藏层的第 i 个神经元连接, 那么隐藏层的第 i 个神经元的输出为

$$x_i^h(t) = f(\text{net}_i^h) = f \left[\sum_{j=1}^{L_i} w_{ij}^h x_j(t) \right]$$

其中 f 为激活函数, $x(t)$ 表示第 t 个输入模式, w_{ij}^h 为第 j 个输入神经元与第 i 个隐藏神经元之间的互连. 隐藏层到输出层的连接为全互连. 当隐藏层的训练收敛后, 输出层利用提取的不变性特征对输入模式进行归类.

3.1 输入层到隐藏层互连的训练与修剪

在隐藏层对输入模式进行不变性特征提取时, 定义隐藏层第 i 个神经元的静态动力学方程为:

$$w_i^h(t) = w_i^h(t-1) + v_i^h(t) \quad (3)$$

$$T_i(t) = x_i^h(t) + \boldsymbol{\pi}_i^h(t) \quad (4)$$

其中, $\boldsymbol{\pi}_i^h(t)$ 和 $v_i^h(t)$ 为平均值为零、协方差为 $R_i^h(t)$ 和 $N_i^h(t)$ 的噪声.

由迹学习规则, 式(4)可写为:

$$T_i(t-1) = x_i^h(t) + \boldsymbol{\pi}_i^h(t) \quad (5)$$

$T_i(t-1)$ 是上一时刻的迹, 相对 t 时刻来说是个恒定常数. 网络权重 $w_i^h(t)$ 的估算可以利用标准 EKF 方法获得^[7]

$$\hat{w}_i^h(t) = \hat{w}_i^h(t-1) + K_i^h(t) [T_i(t-1) - \hat{x}_i^h(t)] \quad (6)$$

$$K_i^h(t) = [C_i^h(t-1) + N_i^h(t-1)] H_i^h(t)^T \{ \boldsymbol{\Gamma}_i^h(t) [C_i^h(t-1) + N_i^h(t-1)] H_i^h(t)^T + R_i^h(t) \}^{-1} \quad (7)$$

$$C_i^h(t) = C_i^h(t-1) + N_i^h(t-1) - K_i^h(t) H_i^h(t) [C_i^h(t-1) + N_i^h(t-1)] \quad (8)$$

其中, $H_i^h(t) = \left[\frac{\partial x_i^h}{\partial w_{ij}^h} \right]_{w_{ij}^h = \hat{w}_{ij}^h(t-1)}$, K_i^h 为卡尔曼增益, C_i^h 为估算

的权重误差协方差矩阵, $\hat{x}_i^h$ 和 $\hat{w}_i^h$ 分别为网络的输出和权重的估算值. 假设训练在 t_0 次迭代后收敛, 相应的协方差矩阵 $C_i^h(t)$ 收敛到稳定值 C_i^h 上, 利用矩阵变换定理 $(A^{-1} + B R^{-1} B^T)^{-1} = A - A B (B^T A B + R)^{-1} B^T A$, 式(8)可以重写为

$$(C_i^h)^{-1} N_i^h(t-1) [C_i^h + N_i^h(t-1)]^{-1} = H_i^h(t)^T R_i^h(t)^{-1} H_i^h(t) \quad (9)$$

如果训练在 T 时刻停止, 且 $T > t_0$, 我们可以得到:

$$\begin{aligned} \frac{1}{T-t_0} \sum_{t=t_0}^T (C_i^h)^{-1} N_i^h(t-1) [C_i^h + N_i^h(t-1)]^{-1} \\ = \frac{1}{T-t_0} \sum_{t=t_0}^T H_i^h(t)^T R_i^h(t)^{-1} H_i^h(t) \end{aligned} \quad (10)$$

为简单起见, 假设 $N_i^h(t) = \delta I_{N^h \times N^h}$, $R_i^h(t) = I_{N^h \times N^h}$. 当 T 较大时, 式(10)右边的项将趋近 $H(t)^T H(t)$ 的期望值, 即:

$$\delta (C_i^h)^{-1} [C_i^h + \delta I]^{-1} = E[H_i^h(t)^T H_i^h(t)] \quad (11)$$

由于协方差矩阵是准对称正定的, 因而它可以分解为 $C_i^h = V G V^{-1}$ 形式, 其中 G 为对角矩阵, 它的对角矩阵元是 C_i^h 的本征值, 而 V 由相应的本征矢量组成. 对于网络能收敛的 δ 值一般远小于 C_i^h 的本征值, 因此, 式(11)可以被表示为:

$$\begin{aligned} E[H_i^h(t)^T H_i^h(t)] &= V \left[\frac{\delta}{G(G+\delta I)} \right] V^{-1} \\ &\approx V [\delta G^{-2}] V^{-1} \approx \delta (C_i^h)^{-2} \end{aligned} \quad (12)$$

将 $H_i^h(t)$ 的定义代入式(12), 得到方程

$$E \left[\left[\frac{\partial x_i^h}{\partial w_{ik}^h} \right]^2 \right] \approx \delta [(C_i^h)^{-2}]_{kk} \quad (13)$$

$[(C_i^h)^{-2}]_{kk}$ 表示它的第 k 个对角元素.

现在我们分析一下网络的误差, 训练后网络的方差期望值可以表示为:

$$E[(T_i - \hat{x}_i^h)^2] = E[(x_i^h + \boldsymbol{\pi}_i^h - \hat{x}_i^h)^2] \quad (14)$$

假定 $\hat{w}_{ik}^h$ 是当 $\hat{w}_i^h$ 的第 k 个元素 (即 $\hat{x}_{ik}^h$) 被置为 0 时相应的权重向量, $\hat{x}_i^h$ 为相应的网络输出, 假定网络输出的误差远小于噪声项, 则方差期望值可以写成

$$\begin{aligned} E[(T_i - \hat{x}_i^h)^2] &= E[(x_i^h + \boldsymbol{\pi}_i^h - \hat{x}_i^h)^2] \\ &= E[(x_i^h + \boldsymbol{\pi}_i^h - \hat{x}_i^h)^2] + E[(\hat{x}_i^h - \hat{x}_i^{h*})^2] \end{aligned} \quad (15)$$

删除 $\hat{w}_{ik}^h$ 引起的误差增量值为:

$$\begin{aligned} \Delta E_k &= E[(T_i - \hat{x}_i^h)^2] - E[(T_i - \hat{x}_i^h)^2] \\ &= E[(\hat{x}_i^h - \hat{x}_i^{h*})^2] \end{aligned} \quad (16)$$

对式中右边项进行 Taylor 展开, 忽略掉高次项和代入式

(13)的结果后,式(16)可表示为:

$$\Delta E_k^h \approx E \left[\left(\frac{\partial x_i^h}{\partial w_{ik}^h} \right)^2 \right] (\hat{w}_{ik}^h)^2 \approx \delta_i^h (C_i^h)^{-2} J_{kk} (\hat{w}_{ik}^h)^2 \quad (17)$$

式(17)描述了删掉权重 $\hat{w}_{ik}^h$ 所引起网络输出误差变化的大小,根据此变化的大小,可以判断权重 $\hat{w}_{ik}^h$ 的重要性.基于以上的分析,删剪过程如下进行:

(1)利用 EKF 方程及迹学习规则式(2)自组织地训练网络, i 以递增的顺序;

(2)用式(17)对所有权重 ($i = 1, 2, \dots, N^h$) 的重要性进行估算,根据删剪权重对 ΔE_i^h 值变化影响的大小,对所有权重的重要性进行排列,排列的方法为 $\{S_1, S_2, \dots, S_{N_w}\}$ (其中 N_w 为输入层到隐藏层总的权重数目), $\Delta E_{S_m}^h \leq \Delta E_{S_k}^h$ ($m < k$);

(3)令 $\Delta \hat{w}_{S_k} = \hat{w}_{S_k}$ (当 $k = 1, \dots, k_f$ 且 $\Delta \hat{w}_{S_k} = 0$ (当 $k > k_f$),用 $\Delta \hat{w}_{S_k}$ 重新构成 $\Delta \hat{w}^h$;

(4)估算由于权重(从 w_{S_i} 到 w_{S_k})被删除引起的平均期望误差的增量变化

$$\Delta E_{[S_i, S_k]} = \sum_{i=1}^k (\Delta \hat{w}_i^h)^T (C_i^h)^{-2} (\Delta \hat{w}_i^h) \quad (18)$$

如果这个估算的误差的变化量不超过某个确定的阈值,则定义 $k' = k' + 1$,转到步骤(3);否则停止,删掉从 w_{S_i} 到 $w_{S_{k'-1}}$ 的 $k' - 1$ 个权重.

3.2 隐藏层到输出层权重的 EKF 训练和删剪

训练和删剪隐藏层到输出层的权重采用全局 EKF 有导师学习算法,系统的静态动力学方程可以写为:

$$w^o(t) = w^o(t-1) + v^o(t) \quad (19)$$

$$d^o(t) = x^o(t) + \varepsilon^o(t) \quad (20)$$

d^o 和 x^o 分别为输出层的理想输出和实际输出,同上一节相同的推导方法,我们可以得到形式相同的 EKF 迭代公式及权重删剪过程,不同的是我们在输出层采用的激活函数为 $f(\text{net}_i)$

$$= \frac{1}{1 + e^{-b \sin(r, \text{net}_i)}}$$

4 计算机模拟结果

4.1 网络的学习能力

本文使用 Lovell^[8]和 Wallis^[2]在其文献中使用的一套无约束手写体数字库来研究网络对学习样本的分类能力.此样本的每一数字由 20×20 个像素组成.同 Lovell 和 Wallis 的方法一样,400 个手写体数字随机进入网络学习 800 次.其中的网络参数为 $\eta = 0.9$, $N^i = 20 \times 20$, $N^h = 14 \times 14$, $N^o = 1 \times 10$, $L_i = 5 \times 5$, $C_i^h(0) = C_i^o(0) = I$,迹的初始值 $T(0)$ 定为零,权重的初始值在区间 $[-0.1, 0.1]$ 中随机选取,隐藏层神经元激活函数的参数为 $K=2$, $b_1 = b_2 = 8.5$, $r_1 = r_2 = 0.4$, $h_1 = h_2 = 0.5$, $c_1 = 4$, $c_2 = -4$,输出层神经元激活函数参数为 $b = 7.5$, $r = 0.4$.网络收敛以后,根据删剪算法删除不重要的权重.表 1 给出了 3627 个权重(隐藏层 3069 个,输出层 558 个)被删除后网络性能的一些数据,同时还给出了 Lovell 和 wallis 报道的结果.可以看到,我们的网络用的神经元及权重数目最少,但网络的分类能力最佳,说明上述算法及周期函数确能提高网络的分类能力.

表 1 三种模型或算法的分类能力比较

网络模型或算法	训练样本识别率	神经元个数	权重个数
EKF	100 %	206	3233
Wallis ^[2]	74.5 %	410	10400
Supervised Neocognitron ^[8]	72.5 %	11320	1200000

4.2 小规模数据库的识别能力

我们采用一个小规模的手写体数字库^[2]作为考察网络性能的样本.图 1 给出了所有的 200 个样本.参数 $\eta = 0.7$.图 2(a)、(b)显示了平均方差的估算值和实际值与隐藏层和输出层中删减权重数目的关系,可以看出,估算的方差与实际方差非常逼近,这说明上述权重重要性的估算是非常精确的.在模拟中我们发现权重的删剪处理能够提高网络的推广能力,删剪后的网络对测试样本的识别率提高了 2%.表 2 给出了当 4700 个权重(隐藏层 3780 个,输出层 920 个)被删除后网络性能的一些数据.需要说明的是, Peng^[3]的模拟是基于另一个规模相同的数据库.把周期函数同 Sigmoid 函数的结果比较,可以看到采用周期型激活函数提高了网络的收敛速度和识别率.图 3(a)、(b)给出了稳定的迹和隐藏层神经元输出的模式.它们高度的相似性表明我们提出的算法能够有效地提取输入模式的不变性特征.

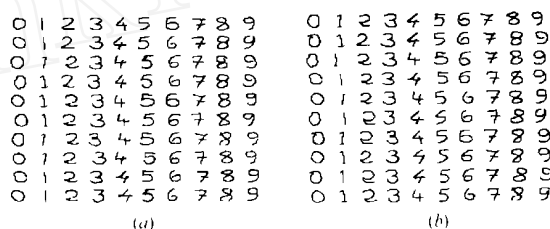


图 1 20 套手写体数字 (a) 训练样本; (b) 测试样本

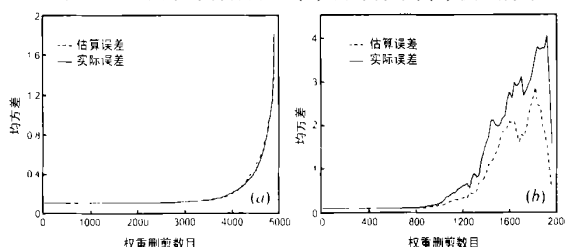


图 2 估算误差、实际误差与权重删剪数目的关系曲线 (a) 隐藏层; (b) 输出层

表 2 五种模型或算法对小规模样本集的性能比较

网络模型或算法	迭代次数	训练样本识别率	测试样本识别率
EKF (Periodic)	~150	100 %	76 %
EKF (Sigmoid)	~270	100 %	70 %
Wallis ^[2]	~70	95 %	55 %
Peng ^[3]	~500	95 %	64.83 %
BP net ^[3]	~500	95 %	57.92 %

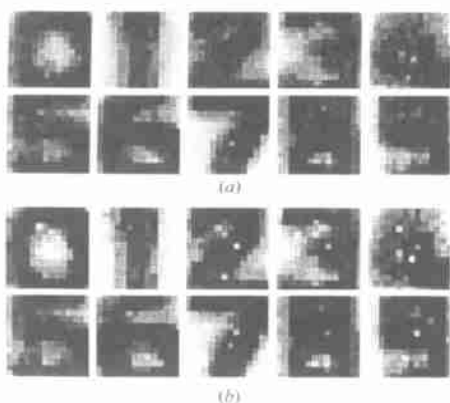


图3 (a)隐藏层训练收敛后的迹;
(b)一组模式送入网络后得到的输出

4.3 大规模数据库的识别能力

为了进一步测试网络的性能,我们在一个大规模数据库^[3]上进行了模拟工作.这个数字库共有 6000 手写体数字,其中随机选取 4800 个作为学习样本,另外 1200 个作为测试样本.表 3 给出了当 2738 个权重(w^h 中 2358 个, w^o 中 380)被删掉后的识别结果,同时表中还列出了其他几种方法对该数据库的模拟结果,同这些方法比较,我们的算法同样获得了较高的识别率.

表 3 四种模型或算法对大规模样本集的性能比较

网络模型或算法	训练样本识别率	测试样本识别率
EKF(Periodic)	99.83 %	92.8 %
EKF(Sigmoid)	99.4 %	91.5 %
Peng ^[4]	91 %	90.5 %
BP net ^[4]	99 %	90.5 %

5 结论

本文把局域 EKF 算法同迹学习规则相结合用于无约束手写体数字的自组织不变性特征提取和识别中,在此算法的基础上,提出了一种在线的权重自适应删剪学习算法.为了提高不变性特征的鲁棒性,我们提出采用周期包函数作为激活函数.模拟结果表明该算法及周期包函数不仅能够有效地提取手写体数字的不变性特征,同时也能自适应地找到网络的最佳规模,从而获得较高的识别率.

参考文献:

- [1] P Foldiak. Learning invariance from transformation sequence [J]. Neural Computation. 1991, 3:194 - 200.
- [2] G Wallis. Using spatio-temporal correlations to learn invariant object recognition [J]. Neural Network. 1996, 9(9):1513 - 1519.
- [3] H Peng, L Sha, Q Gan, Y Wei. Energy function for learning invariant in multilayer perceptron [J]. Electronics Letters, 1998, 34:292 - 294.
- [4] Y LeCun et al. "Optimal brain damage", Advances in Neural Information Processing System. 2, D. S. Touretsky, Ed., 1990, pp. 396 - 404.
- [5] B Hassibi, D G Stork. Second-order derivatives for network pruning: Optimal brain surgeon [M]. Advances in Neural Information Processing system, Hanson et al., Eds., 1993, pp. 164 - 171.
- [6] S J Chang, K W Wong, C S Leung. Periodic activation function for fast on-line EKF training and pruning [J]. Electronics Letters, 1998, 34: 2255 - 2256.
- [7] Y Liguni, H Sakai, H Tokumaru. Real-time learning algorithm for a multilayered neural network based on the extended Kalman filter [J]. IEEE Transaction on signal Procession, 1992, 40:959 - 967.
- [8] D R Lovell, A C Tsoi. The performance of the neocognitron with various S-cell and G-cell transfer functions. Paper posted to the "Connectionists" Internet mail group. Produced in the Dept. of Electrical Engineering, University of Queensland, Queensland 4072, Australia, 1992.

作者简介:



林 列 男, 1998 年毕业于南开大学现代光学研究所, 获博士学位, 现任信息技术学院副院长、副研究员, 主要研究方向有神经网络及其应用、光学/数字图像处理等.



常胜江 男, 1996 年毕业于南开大学现代光学研究所, 获博士学位, 1998 年至 2000 年在香港城市大学作博士后研究工作, 现任南开大学现代光学所教授, 主要研究方向有神经网络及其应用、飞秒激光及与物质相互作用、数字图像处理等.