

基于非线性时序模型的神经网络 图形识别方法

熊沈蜀 周兆英

(清华大学精密仪器与机械学系 北京 100084)

摘要 提出了复数域非线性指数自回归(CNEAR)图形轮廓建模方法. 构造了用于 CNEAR 模型参数估计的神经网络模型, 利用该网络可解决非线性参数的估计问题. CNEAR 模型参数具有平移、旋转和比例不变性并与计算起始点的选择无关. 基于模型参数构造了特征向量, 并设计了复数神经网络分类器, 给出了复数神经网络学习算法. 实验结果表明 CNEAR 模型在较低阶次即能获得较高的识别率, CNEAR 模型对带噪声图形及形状差别较小图形的识别效果要好于复数域自回归模型方法.

关键词 图形识别, 复数域非线性指数自回归模型, 神经网络.

NONLINEAR TIME SERIES MODEL FOR SHAPE RECOGNITION USING NEURAL NETWORKS

XIONG Shenshu ZHOU Zhaoying

(*Department of Precision Instruments & Mechanology, Tsinghua University, Beijing 100084*)

Abstract A complex nonlinear exponential autoregressive (CNEAR) modeling approach for the closed shape boundary is suggested in this paper. A neural network structure is constructed to estimate the parameters of the CNEAR model. The nonlinear coefficient is easy to be determined by the network. The CNEAR parameters are invariant to translation, rotation around the origin, scale and choice of the starting point in tracing a boundary. The feature vector can be constituted by the CNEAR parameters. A complex neural network classifier is designed for learning and classification and its learning algorithm is also given. Experimental results show that high recognition rates can be achieved even for low-order models. It is also seen that the CNEAR model performs better than the complex autoregressive (CAR) model when shapes have random noise on the boundaries or have differentiating features at detailed levels.

Key words Shape recognition, complex nonlinear exponential autoregressive

model, neural network.

1 前言

图形识别技术是计算机视觉中的一个基本组成部分,已广泛地应用于工业自动化、图象分析与识别、文字识别等领域.图形识别系统一般包括以下四个部分:

1)获取图象信息——将图象用计算机能接受的形式表示.

2)预处理——去除噪声,增强有用信息,复原退化信息.

3)特征提取——对图形的某些物理性质进行数学描述,得到最能反映图形分类的本质特征.对不同的图形类别,特征的类内差别应尽可能的小,而类间差别应尽可能的大.特征应与图形的平移、旋转和比例变化无关.

4)学习和分类——学习是在建立识别系统时,用已选定的特征对系统进行训练以得到判决标准的过程;分类是在掌握分类规律后的实现阶段对连续输入的大量图形模式进行分类.

综上所述,图形识别系统的关键在于特征提取和模式分类两个阶段.如果能选取具有高度准确描述能力的特征,无疑对系统的建立具有重要意义,它可以较少地存贮表达较多的物理意义,而分类器的正确设计可使系统具有较高的稳定性和准确性.本文所做的研究也是集中于这两个阶段.

特征提取的方法很多,其中最常用的有不变矩、傅立叶描述符、时序模型等方法^[1].不变矩方法的缺点是有限的不变量特征对图形局部信息反映不够,在识别复杂图形或形状差别较小的图形时效果不好.傅立叶描述符方法虽然可以有足够的特征数量来反映图形轮廓局部的细微变化,但所需特征量大,使后续的模式识别工作负担较重,而且特征有冗余并不直接具备平移、旋转和比例不变性.另外,这两种方法都对噪声比较敏感.在图形建模的许多方法中,时间序列建模技术近年来被有效地应用在图形识别领域.这种方法的基本思路是:对封闭图形的轮廓进行采样可得到一个单周期内的数值序列,然后对此序列建立相应的时序模型,模型系数可以作为识别所需的特征.

Dubois^[2]等人最先提出将自回归(简称AR)模型应用于图形识别,他们在质心与轮廓之间以等角间距采样得到径向量的长度序列,然后对此序列应用AR模型.这种建模方法基本上只适用于凸状简单图形的识别.为了克服Dubois方法不能识别非凸图形的缺点,Pal^[3]等人采用了等分轮廓周长的采样方法,但这种模型对噪声非常敏感.Kartikyan^[4]等人认为线性AR模型只适用于识别那些形状明显不同的图形而对形状差别较小的图形识别能力较差,为此他们构造了非因果二次Volterra模型.但是这种模型的计算量很大,模型阶数不易选择,从而使特征集的形式很难统一并增大了模式分类的难度.在以上三种方法中,所用模型皆为一维模型.然而,平面图形轮廓本质上属于二维信号范畴,用一维模型描述图形轮廓显然既不充分也不精确.为此,Das^[5]等人采用了二维双变量AR模型,这种方法要估计的模型系数是相应同阶次一维模型的四倍,冗余性大,另外,此模型只是简单地将直角坐标 x 和 y 作为双变量来处理,忽视了它们之间的正交特性,导致模型系数并不直接具备旋转不变性.为了克服双变量AR模型的固有缺点,Sekita^[6]等人提出了复数域AR(CAR)模型,实验研究表明识别效果优于不变矩和傅立叶描述符方

法,尤其在有噪声的情况下.但是由于该模型是线性的,它不能反映图形轮廓的二维非线性封闭循环特性和局部特性^[1],所以在识别某些相似图形时效果不好.另外,CAR模型的特征具有镜象不变性,对镜象相似图形如“2”和“5”,“b”和“p”等很难进行识别.为解决这些问题,本文提出了复数域非线性指数自回归(CNEAR)图形轮廓建模方法,并研究基于CNEAR模型的神经网络图形识别方法.

2 CNEAR模型及特征提取

2.1 CNEAR模型

图形的封闭轮廓是最重要的图形表征形式,在物理意义上等效于两个一维等周期非线性振动分量的正交合成轨迹^[1].因此可以认为封闭图形轮廓与一维非线性振动系统的离散过程之间存在某些本质的联系.对于一个一维非线性振动系统,其 n 阶离散过程可以用指数自回归模型描述为

$$x_i = \sum_{i=1}^n [\varphi_i + \pi_i \exp(-\gamma x_{i-1}^2)] x_{i-1} + \alpha_i, \quad (1)$$

其中 $\varphi_i, \pi_i, \gamma, i=1, 2, \dots, n$ 均为模型系数, α_i 为白噪声.将上述模型推广至复数空间,本文构造了图形轮廓的CNEAR模型.

设 $\{(x_i, y_i) | i=0, 1, \dots, N-1\}$ 为对图形的轮廓曲线进行适当的等间距采样后得到的采样点相对轮廓质心的坐标序列,其中 N 为采样点的个数.封闭轮廓曲线具有周期特性,即有关系式 $x_j = x_{j+N}$ 及 $y_j = y_{j+N}, j=0, 1, \dots, N-1, N, \dots$.为描述直角坐标的二维正交关系,采用复数来表示轮廓的采样点坐标,即 $z_i = x_i + jy_i$,这样就可以得到图形轮廓采样点的复数坐标点序列 $\{z_i | i=0, 1, \dots, N-1\}$.对于沿轮廓顺时针方向的采样点序列 $\{z_i | i=0, 1, \dots, N-1\}$, m 阶CNEAR模型定义为

$$z_i = \sum_{k=1}^m (a_k + b_k e^{\frac{-\gamma |z_{i-1}|^2}{r_0}}) z_{i-k} + \varepsilon_i, \quad (2)$$

其中 $a_k, b_k, k=1, 2, \dots, m$ 为复数型的模型参数, γ 是实数型的非线性模型参数, $r_0 = \frac{1}{N} \sum_{i=0}^{N-1} |z_i|^2$, ε_i 为复数域内的白噪声.

2.2 基于神经网络的CNEAR模型参数估计

在CNEAR建模过程中,模型的非线性参数导致了一个非常复杂的非线性方程组,求解参数十分困难,而且求解时间长对图形识别应用是很不现实的.为解决这个问题,本文构造了神经网络估计器,利用图1给出的网络模型可以在使残差序列的均方差最小的准则下同时估计出CNEAR模型的全部参数.

定义网络的误差函数为

$$E = \sum_{i=0}^{N-1} |z_i - \hat{z}_i|^2 = \sum_{i=0}^{N-1} (z_i^* z_i - \mathbf{W}^* \bar{X}_i z_i - z_i^* X_i^T \mathbf{W} + \mathbf{W}^* \bar{X}_i X_i^T \mathbf{W}), \quad (3)$$

其中 $\hat{z}_i = X_i^T \mathbf{W}, X_i = [z_{i-1}, e^{\frac{-\gamma |z_{i-1}|^2}{r_0}} z_{i-1}, \dots, z_{i-m}, e^{\frac{-\gamma |z_{i-1}|^2}{r_0}} z_{i-m}]^T$,

$$\mathbf{W} = [a_1, b_1, \dots, a_m, b_m]^T.$$

E 关于 γ 的偏导数及向量 \mathbf{W}^* 的偏导向量分别为

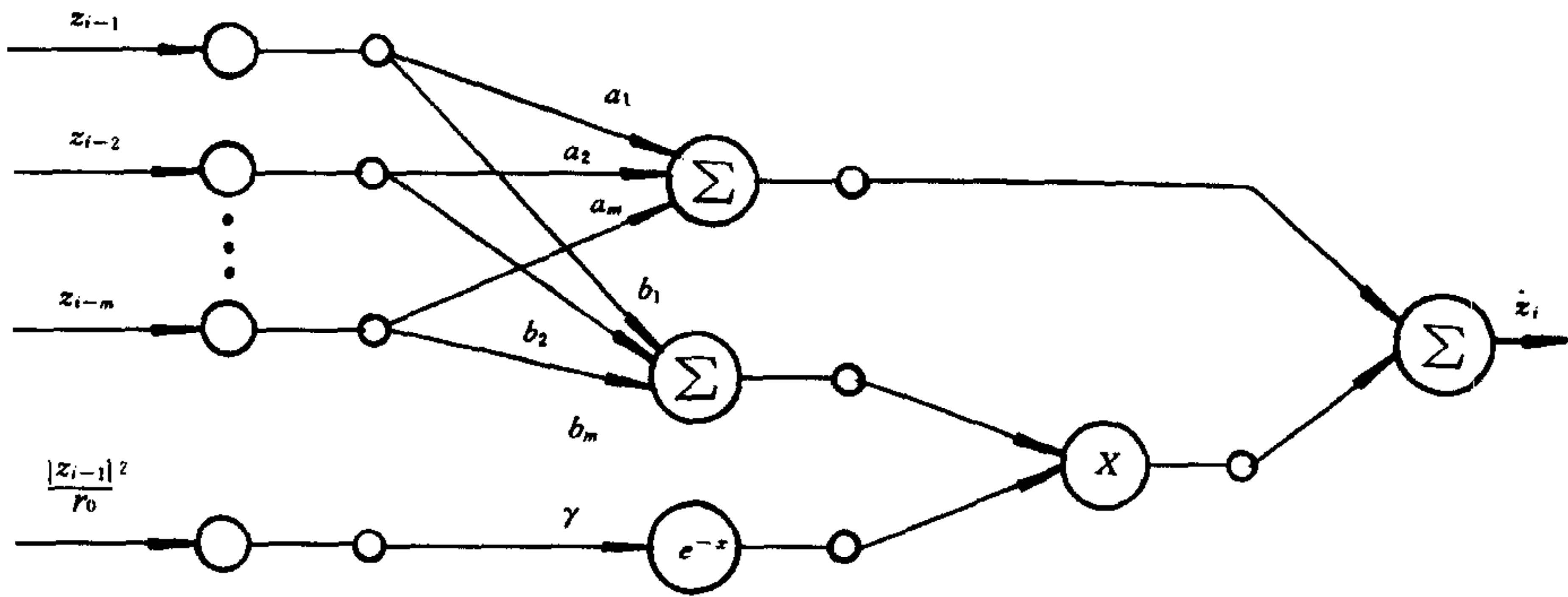


图1 用于 CNEAR 模型参数估计的神经网络结构

$$\frac{\partial E}{\partial \gamma} = \sum_{i=0}^{N-1} \left(\frac{|z_{i-1}|^2}{r_0} e^{-\frac{\gamma |z_{i-1}|^2}{r_0}} \left(\left[\left(z_i - \sum_{k=1}^m a_k z_{i-k} \right) \sum_{k=1}^m \bar{b}_k \bar{z}_{i-k} \right]_{\text{Re}} - 2e^{-\frac{\gamma |z_{i-1}|^2}{r_0}} \left| \sum_{k=1}^m b_k z_{i-k} \right|^2 \right) \right), \quad (4)$$

$$\frac{\partial E}{\partial W^*} = \sum_{i=0}^{N-1} (-\bar{X}_i z_i + \bar{X}_i X_i^T W) = -\sum_{i=0}^{N-1} \bar{X}_i e_i. \quad (5)$$

这里,下标 Re 表示复数的实部, $e_i = z_i - \hat{z}_i = z_i - X_i^T W$.

采用梯度算法来计算网络权值

$$\Delta \gamma(n+1) = -\frac{\lambda_\gamma}{r_0} \frac{\partial E(n)}{\partial \gamma} + \eta \Delta \gamma(n), \quad (6)$$

$$\Delta W(n+1) = -\frac{\lambda_W}{r_0} \frac{\partial E(n)}{\partial W^*} + \eta \Delta W(n). \quad (7)$$

上式中 n 表示时序变量; $\lambda_\gamma, \lambda_W$ 分别是权值 γ 及权值向量 W 的学习步长; η 为惯性因子; $\Delta \gamma(n+1) = \gamma(n+1) - \gamma(n)$; $\Delta W(n+1) = W(n+1) - W(n)$. 网络收敛的条件取为 $(\|\Delta W(n)\|_{C^{2m}} + |\Delta \gamma(n)|) < \epsilon$, ϵ 为给定的精度, $\|\cdot\|_{C^{2m}}$ 表示 $2m$ 维酉空间的 2-范数.

2.3 特征向量的构造

可以证明^[1],当坐标原点设在轮廓质心时,CNEAR 模型参数具备平移、旋转和比例不变性并与轮廓跟踪的起始点选择无关.另外,由于非线性的不可逆性,CNEAR 模型不具备镜象不变性,这条性质对于分辨镜象相似图形很有用处,如“b”和“p”,“2”和“5”等.还须指出,CNEAR 模型同时包含线性和非线性项,不但能提取图形轮廓的基本形状,还能将轮廓的局部信息提取出来.因而可以识别形状比较接近但类别不同的图形.

这些性质符合特征向量的构造要求,可以使相应模式分类器的设计不必考虑对图形正常几何变换不变性的要求.在本文中,构造特征向量为 $v = [a_1, b_1, a_2, b_2, \dots, a_m, b_m, \gamma]^T$.

3 神经网络分类器的设计

相应于第2节构造的特征向量,常用的模式分类方法是统计决策方法.在统计模式识别方法中,当参考总体内样本的模式向量分布不服从预先假定的正态分布,或输入模式受到噪声、畸变干扰,或模式形状比较复杂时,基于统计决策分类器的性能就会有所下降.神

神经网络所具有的信息分布式存贮、自适应学习能力、高度的鲁棒性等特点为上述问题提供了解决途径.

图形分类器可以看作是一个映特征空间的图形空间的算子 T . 在本文中, T 将紧集 $U = \{v_i | v_i \in C^{2m+1}\}$ 映射到输出空间 $Y = \{y_i | y_i \in C^N\}, i = 1, 2, \dots$, 其中 v_i 表示图形类别 y_i 的特征向量. 这里要说明一点, 为统一起见, 将 γ 也视为复数, 其虚部为 0. 本文构造了图 2 所示的三层前馈网络可以用来辨识算子 T , 该网络模型就是图形分类器. 特征向量直接作为网络的输入, 输入节点数目 N_I 与特征向量的维数相同, 即 $N_I = 2m + 1$. 输出节点数目 N_O 与待识别的图形类别的数目相同. 隐节点个数取为 $N_H = 4m + 3$. 网络节点的激励函数取如下复变函数

$$\sigma(z) = \frac{1}{1 + \exp(-z_{\text{Re}})} + j \frac{1}{1 + \exp(-z_{\text{Im}})}, \quad (8)$$

其中 $z = z_{\text{Re}} + jz_{\text{Im}}$, 下标 Re 和 Im 分别表示复数的实部和虚部(以下同). 该网络的输入向量、输出向量、权值向量以及激励函数均为复值, 为区别于传统神经网络本文定义该种网络为复数神经网络.

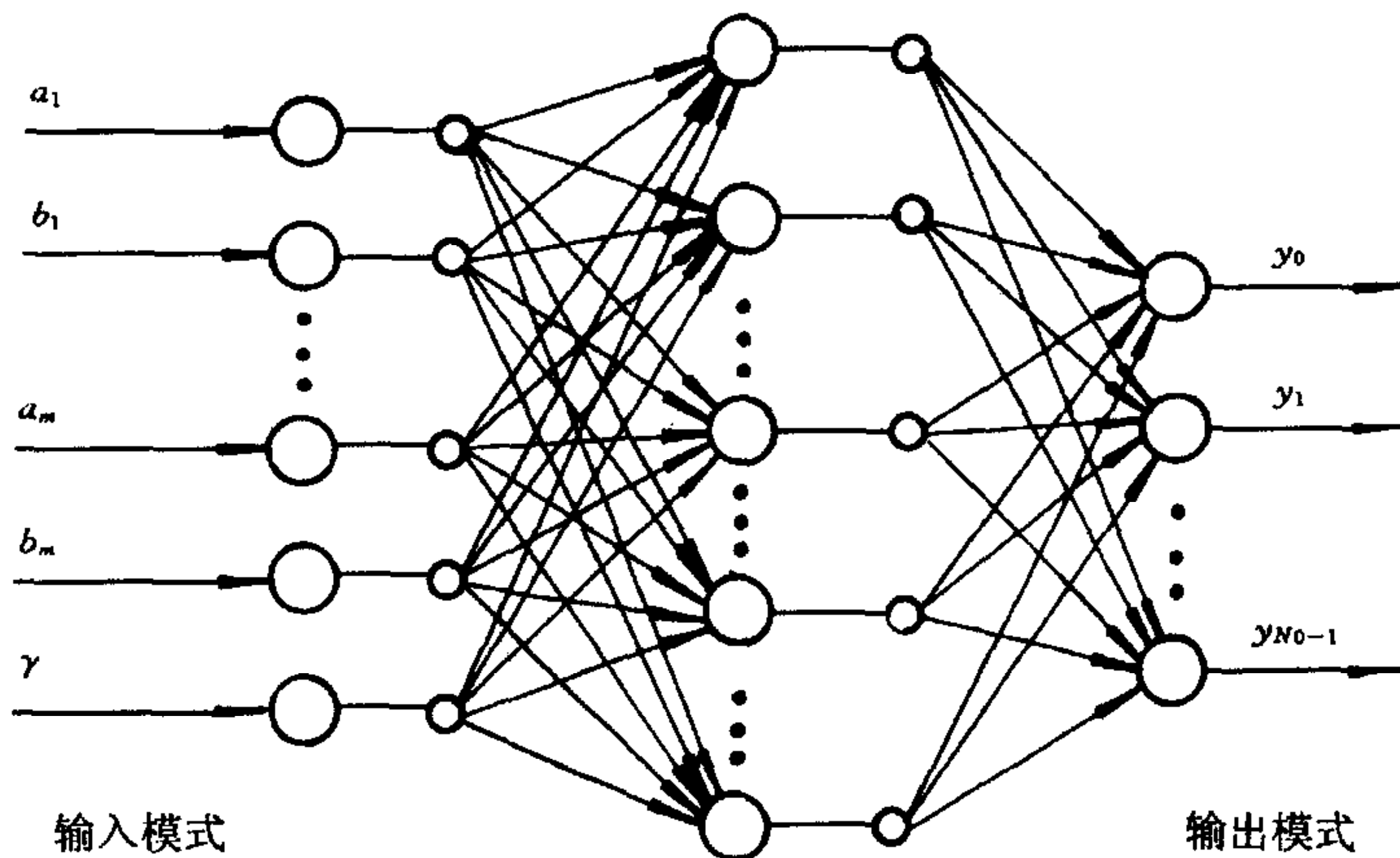


图2 复数神经网络分类器

下面给出复数神经网络学习算法. 为了便于分析, 对网络各层的输入、输出及权值用统一的符号来描述. $O_{ki}^{(l)}$ 表示第 l 层中第 i 个神经元的输出, 下标 k 表示学习过程中所用的训练样本编号; $w_{ir}^{(l+1)}$ 表示第 l 层的第 r 个神经元至第 $l+1$ 层的第 i 个神经元的权重系数; d_{ki} 表示网络总输出向量的各个分量的理想值, 在本文定义为

$$d_{ki} = \begin{cases} 0.9 + j0.9, & \text{如果输入为第 } i \text{ 类图形,} \\ 0.1 + j0.1, & \text{如果输入不为第 } i \text{ 类图形,} \end{cases} \quad i = 0, 1, \dots, N_o - 1. \quad (9)$$

对于训练样本 k , 网络的各层输出分别为

$$O_{k,2i-2}^{(1)} = a_i, \quad O_{k,2i-1}^{(1)} = b_i, \quad O_{k,2m}^{(1)} = \gamma, \quad i = 1, 2, \dots, m; \quad (10)$$

$$O_{ki}^{(2)} = \frac{1}{1 + \exp(-\sum_{r=0}^{N_I-1} w_{ir}^{(2)} O_{ki}^{(1)})_{\text{Re}}} + j \frac{1}{1 + \exp(-\sum_{r=0}^{N_I-1} w_{ir}^{(2)} O_{ki}^{(1)})_{\text{Im}}}, \quad (11)$$

$$i = 0, 1, \dots, N_H - 1;$$

$$O_{ki}^{(3)} = \sum_{r=0}^{N_H-1} w_{ir}^{(3)} O_{ki}^{(2)}, \quad i = 0, 1, \dots, N_o - 1. \quad (12)$$

网络的总体误差函数定义为

$$E = \sum_k E_k = \sum_k \sum_{i=0}^{N_o-1} |d_{ki} - O_{ki}^{(3)}|^2. \quad (13)$$

采用共轭梯度法来计算网络各个权重系数的调整量 $\Delta w_{ir}^{(l)}$,

$$\Delta w_{ir}^{(l)} = -\lambda \frac{\partial E}{\partial (w_{ir}^{(l)})^*} = -\lambda \left(\frac{\partial E}{\partial w_{ir,Re}^{(l)}} + j \frac{\partial E}{\partial w_{ir,Im}^{(l)}} \right). \quad (14)$$

当 $l = 3$ 时, $\Delta w_{ir}^{(3)} = \lambda \sum_k (d_{ki} - O_{ki}^{(3)}) \bar{O}_{kr}^{(2)}, i = 0, 1, \dots, N_o - 1, r = 0, 1, \dots, N_H - 1;$

$$(15)$$

当 $l = 2$ 时, $\Delta w_{ir}^{(2)} = \lambda \sum_k \delta_{ki} \bar{O}_{kr}^{(1)}, i = 0, 1, \dots, N_H - 1, r = 0, 1, \dots, N_I - 1,$

$$(16)$$

$$\delta_{ki} = \delta_{ki,Re} + j\delta_{ki,Im}, \quad (17)$$

$$\begin{aligned} \delta_{ki,Re} = & O_{ki,Re}^{(2)} (1 - O_{ki,Re}^{(2)}) \sum_{p=0}^{n-1} ((d_{kp} - O_{kp}^{(3)})_{Re} \cdot w_{pi,Re}^{(3)} \\ & + (d_{kp} - O_{kp}^{(3)})_{Im} \cdot w_{pi,Im}^{(3)}), \end{aligned} \quad (18)$$

$$\begin{aligned} \delta_{ki,Im} = & O_{ki,Im}^{(2)} (1 - O_{ki,Im}^{(2)}) \sum_{p=0}^{n-1} (- (d_{kp} - O_{kp}^{(3)})_{Re} \cdot w_{pi,Im}^{(3)} \\ & + (d_{kp} - O_{kp}^{(3)})_{Im} \cdot w_{pi,Re}^{(3)}). \end{aligned} \quad (19)$$

为了提高网络的学习速度,可以采用惯性校正学习算法

$$\Delta w_{ir}^{(l)}(n+1) = -\lambda \frac{\partial E(n)}{\partial (w_{ir}^{(l)})^*} + \eta \Delta w_{ir}^{(l)}(n), \quad (20)$$

其中 n, λ, η 分别表示学习的时序编号、步长和惯性系数.

4 实验研究

4.1 实验准备

图3(a)和图3(b)分别给出了八种不同类别的工件和四种形状相似的飞机模型. 对于每种图形,通过将原始样本随机地以不同位置、角度放置并按任意比例进行摄取,分别收集了200个训练样本和40个待识样本. 为了测试在噪声干扰下 CNEAR 模型的识别性能,采用文[6]中介绍的轮廓噪声产生方法对图形集 A 中的待识样本加入随机噪声,图3(c)给出了一些噪声样本.

4.2 采样点数选取及模型定价问题讨论

从信号采集的角度看,采样点数 N 相当于采样频率. 对于大多数图形对象而言,图形轮廓的形状一般不存在十分剧烈的变化,都属于低频缓变信号,只要选择的 N 值不是太小,正常情况下对轮廓的采样均可满足采样定理. 考虑到空间分辨率,采样相邻点间距不应小于2~3个像素点.

在图形识别应用中,模型阶数的确定以能满足准确识别为准,它与轮廓的形状、采样点数都有关. 一般来讲,随着模型阶数升高,模型残差会逐步下降,表明模型逼近真实系统的准确性提高,然而,模型阶数升高意味着模型参数增多,计算误差也随之增大,权衡两方

面因素,应能找到一个较为合适的阶次.

为了研究采样点数与模型阶数对识别率的影响情况,本文分别取 $N=30,60,100$ 及 $m=1,2,\dots,10$ 对图形集 A 中样本进行识别实验. 对于每一个 N 和 m , 利用第2节介绍的神经网络模型构造出1600个学习样本的特征向量,并用这些特征向量训练第3节给出的神经网络分类器. 分类器训练好后,对320个待识样本进行识别测试,测试结果见表1.

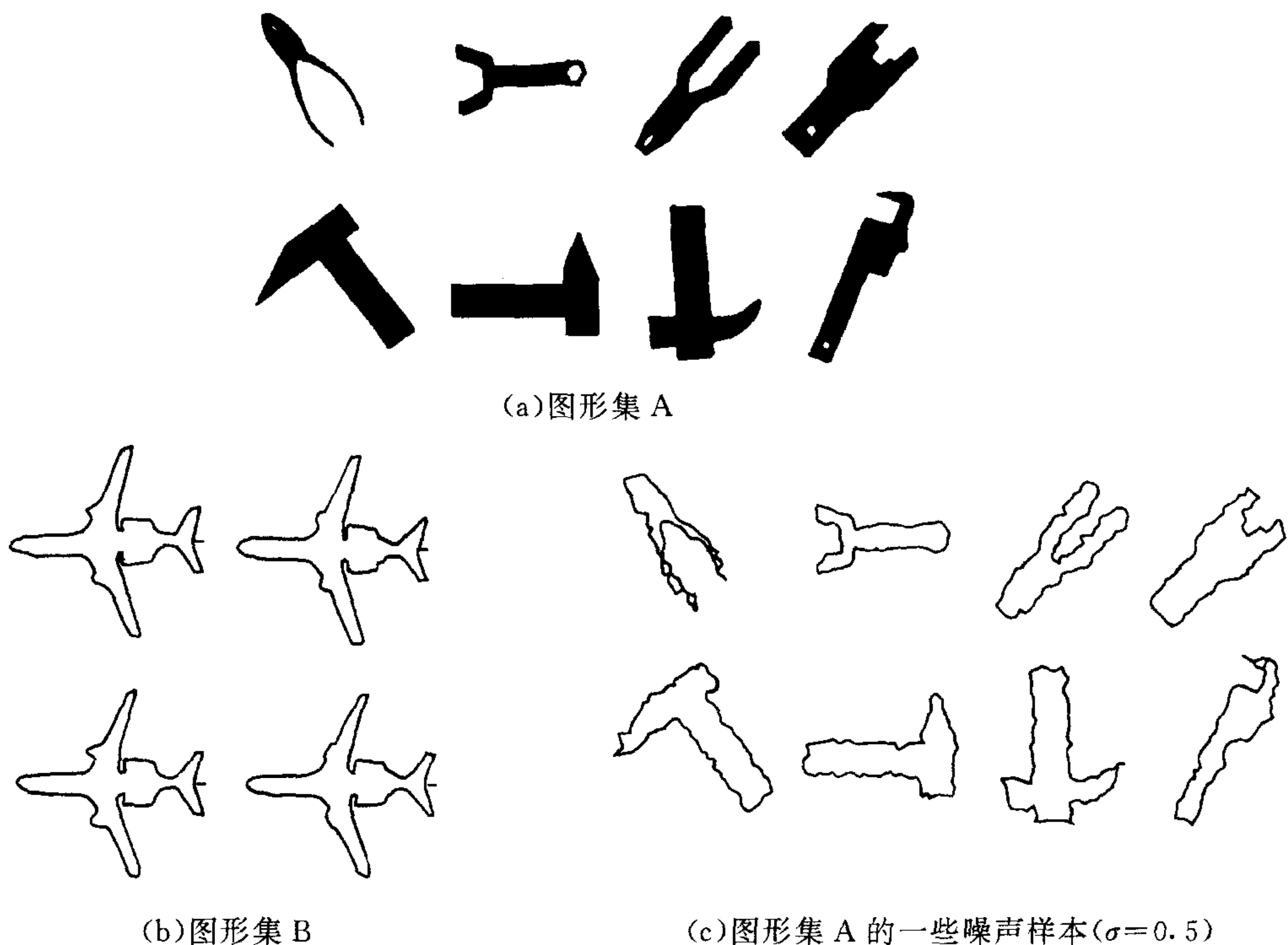


图3 实验样本集

表1 不同采样点、模型阶数下的识别结果(%)

$N \backslash m$	1	2	3	4	5	6	7	8	9	10
30	89.06	98.75	100	100	100	100	100	96.38	96.81	98.56
60	89.69	99.06	100	100	100	100	100	100	97.18	99.69
100	91.56	100	95.93	100	99.38	99.38	97.81	98.44	97.31	93.5

从表1可以看出 CNEAR 模型在较低阶次即能较好地拟合一个封闭轮廓,因此可以认为低阶次的模型已包含了图形的足够信息. 从表中还可以看出采样点过多或模型阶次过高会使识别性能下降.

4.3 与 CAR 模型方法的比较

为了比较 CNEAR 模型和 CAR 模型在噪声条件下及对相似图形的识别效果,本文对两种方法在相同条件下的识别性能进行了测试. 图4为两种模型对图形集 A 的噪声样本的测试结果.

在采样点数 $N=30$ 、模型阶次 $m=3$ 时, CNEAR 模型对图形集 B 待识样本的识别率为100%,而 CAR 的识别率仅为73.13%.

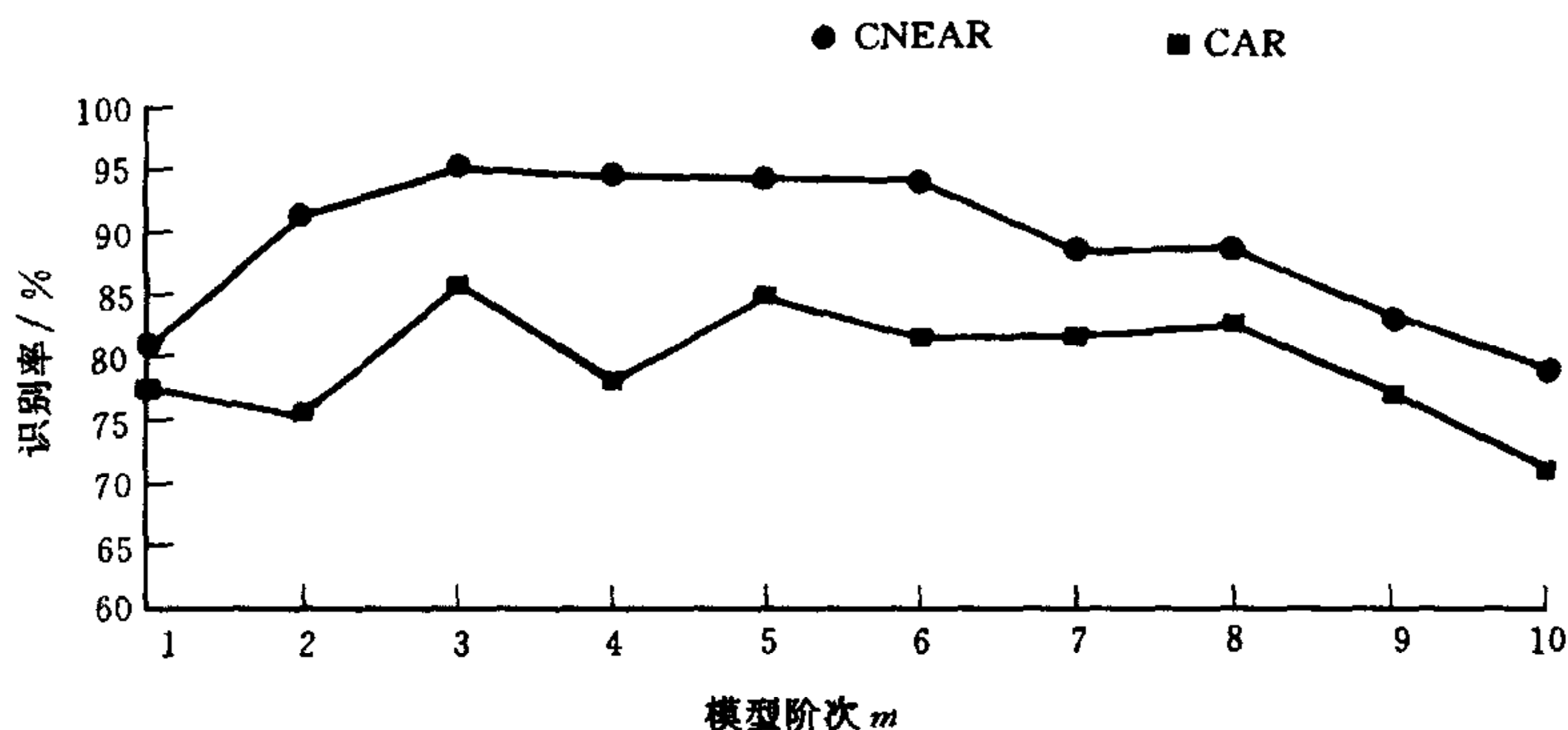


图4 不同模型在噪声条件下的识别结果($\sigma=0.5, N=30$)

由此可以看出,在相同实验条件下,CNEAR 模型的识别性能要强于 CAR 模型.

4.4 提高神经网络分类器性能的方法

为了提高神经网络分类器的鲁棒性,对神经网络分类器采用以下训练步骤:

- 1) 用若干个没有受到噪声干扰的干净图形(噪声方差为0)的特征向量组成学习样本集来训练网络,直至网络收敛;
- 2) 用一些具有不同噪声方差 σ 的测试样本去验证分类器的性能;
- 3) 用若干个噪声方差为 σ_i 的噪声样本再去训练网络分类器,网络权重系数用上一次训练好的网络的稳定权值初始化;
- 4) 转向步骤 2);
- 5) 改变 σ_i 的值,转向步骤 3).

在本实验中, σ_i 分别取值 0, 0.1, 0.2 和 0.3, σ 分别取值 0, 0.1, 0.3, 0.5, 0.7 和 1.0, 对于图形集 A 中每一种类别及每一个噪声方差,训练样本数目取为 200,测试样本数目取为 50,这样对应于每一个 σ_i ,有 1600 个训练样本,对应于每一个 σ ,有 400 个测试样本. 采样点数和模型阶数分别取为 $N = 30, m = 3$. 表 2 给出了测试结果. 由表中可以看出经过噪声训练的神经网络分类器的性能得到很大的改善.

表2 测试样本识别率(%)

$\sigma_i \backslash \sigma$	0	0.1	0.3	0.5	0.7	1.0
0	100	100	97.5	95	89.5	80.5
0.1	100	100	99.5	99	92	83.5
0.2	100	100	99.5	100	94.5	88.5
0.3	100	100	100	98.5	95	93.5

5 结论

本文构造的 CNEAR 模型能够基本反映图形轮廓的本质意义,模型系数具备平移、旋转和比例不变性以及镜象敏感性,并与计算起始点的选择无关,这些特性使得它们可以作为图形的特征. 设计了复数神经网络分类器. 实验结果表明基于 CNEAR 模型的神经网络图形识别方法具有很强的识别效果.

参 考 文 献

- 1 熊沈蜀. 神经网络逼近能力研究及其在参数估计和工程图纸处理中的应用:[博士学位论文]. 北京:清华大学, 1997
- 2 Dubois S R, Glanz F H. An autoregressive model approach to two-dimensional shape classification. *IEEE Trans. PAMI*, 1986, **8**(1):55—66
- 3 Pal N R, Basu A K. A new shape representation scheme and its application to shape discrimination using a neural network. *Pattern Recognition*, 1993, **26**(4):543—551
- 4 Kartikeyan B, Sarker A. Shape description by time series. *IEEE Trans. PAMI*, 1989, **11**(9):977—984
- 5 Das M, Paulik M J. A bivariate autoregressive modeling technique for analysis and classification of planar shapes. *IEEE Trans. PAMI*, 1990, **12**(1):97—103
- 6 Sekita I, Kurita T. Complex autoregressive model for shape recognition. *IEEE Trans. PAMI*, 1992, **14**(4):489—496

熊沈蜀 1970年生,1985年9月至1987年7月在清华大学应用数学系读本科强化基础班,1987年9月至1990年7月专业选在清华大学计算机系继续本科学业,于1990年7月获得清华大学计算机工学学士学位.于1997年4月获得清华大学精密仪器与机械学专业动态测试与控制博士学位.目前是清华大学精密仪器系讲师.主要研究方向:动力学系统建模,模式识别,神经网络理论与应用.

周兆英 1937年生,1961年毕业于清华大学精密仪器系.现任清华大学机械学院副院长,教授,博士生导师.主要研究方向:动态测试与控制,微型机械,神经网络理论与应用等.