

## 应用递归神经网络学习周期运动吸引子轨迹

韩 敏<sup>1</sup>, 史志伟<sup>1</sup>, 席剑辉<sup>1,2</sup>

(1. 大连理工大学 电子与信息工程学院, 辽宁 大连 116023; 2. 沈阳航空工业学院 自动化系, 辽宁 沈阳 110034)

**摘要:** 采用递归神经网络学习非线性周期运动的吸引子轨迹. 网络的拓扑结构基于非线性系统的状态空间表达式, 网络权值通过时序反向传播算法调整. 探讨了不同样本轨迹和网络结构对递归神经网络预测性能的影响. 神经网络的性能评估建立在多条测试样本轨迹的基础上, 可以更为客观地评价递归神经网络预测性能. 对 van der Pol 方程的仿真结果表明: 网络的泛化能力对训练样本轨迹的依赖性较强. 从不同训练轨迹上得到的递归神经网络性能差异较大; 需要选择合适的递归神经网络结构参数以提高神经网络的泛化能力.

**关键词:** 递归神经网络; 周期吸引子; 泛化能力

**中图分类号:** TP183 **文献标识码:** A

## Learning the trajectories of periodic attractor using recurrent neural network

HAN Min<sup>1</sup>, SHI Zhi-wei<sup>1</sup>, XI Jian-hui<sup>1,2</sup>

(1. School of Electronic and Information Engineering, Dalian University of Technology, Dalian Liaoning 116023, China;

2. Department of Automation, Shenyang Institute of Aeronautical Engineering, Shenyang Liaoning 110034, China)

**Abstract:** A kind of RNN (recurrent neural network) is applied to the learning of periodic attractor trajectories for non-linear system. The network topology is based on the state-space representation, and the network parameters are optimized by the back-propagation through time algorithm. Investigations are then conducted into the model performance influenced by different training trajectories and different structure parameters. The model evaluation rule is based on multi-trajectory, which makes the investigation more objective. Simulation results from the van der Pol system show that the generalization ability is dependent on the training trajectory, different trajectories result in a significant different prediction performance; Simulation results also show that the structure parameters of the neural network should be carefully chosen so that better generalization ability can be obtained.

**Key words:** recurrent neural network; periodic attractor; generalization ability

### 1 引言(Introduction)

研究递归神经网络的定性或定量的演化过程是一个非常有意义的问题. 近年来, 很多研究者使用递归神经网络来建立非线性动态系统的模型<sup>[1~3]</sup>. 动态系统的状态和外部输入唯一决定着系统的未来发展, 递归神经网络本身是一个非线性动态系统, 因此有必要研究其状态演化. 稳定的非线性动态系统的一个重要特征是吸引子. 吸引子可分为3种类型: 点吸引子、极限环和奇怪吸引子. 使用神经网络学习点吸引子时, 如果没有外部输入, 系统的演化过程会在暂态过程结束后消失, 神经网络在获取样本时, 吸引子必须在外部输入的激励下不断维持其暂态过程, 这种情况在工业控制过程中较为常见. 使用神经网络建立奇

怪吸引子的预测模型时, 模型的预测误差包含两个方面的因素(不考虑噪声影响), 第一个方面是系统的混沌特性, 另一个方面是神经网络模型的误差. 混沌系统的长期预测是困难的, 因为正的Lyapunov指数会利用模型误差逐步放大预测模型的多步预测误差. 稳定的极限环运动没有正的Lyapunov指数, 预测误差仅仅受到模型的误差影响.

使用神经网络不可能得到与实际系统完全相同演化模型, 神经网络的预测效果受学习样本以及本身的结构参数影响较大. 因此, 使用神经网络作为预测模型时, 可以从几个方面提高网络的预测性能, 如改善样本质量、优化网络结构等. 本文特别地考察了神经网络对极限环吸引子的学习情况, 由于此时模型的

收稿日期: 2004-07-2; 收修改稿日期: 2005-10-25.

基金项目: 国家自然科学基金资助项目(60374064).

预测误差不受混沌特性的影响, 可以认为预测误差仅仅是由训练样本和模型本身引起的。

使用递归神经网络学习动态系统, 常见的方法是将动态逼近任务转化为一个静态逼近任务<sup>[4~6]</sup>, 学习算法仍采用前馈网络的学习算法, 网络训练完成后, 将网络输出反馈到输入端, 迭代预测。递归神经网络是一个动态系统, 在处理时序任务时不同于前馈网络, 在训练网络的时候考虑了时序因素。常见的递归神经网络有Elman网络<sup>[7]</sup>、Jordan网络及其变体等等, 状态空间神经网络<sup>[8~10]</sup>作为Elman网络的一种扩展, 其学习能力和逼近能力得到了增强, 将其退化成为自治形式时, 又可以较好地学习混沌方程<sup>[11]</sup>, 并且具有相应的混沌特征。使用前馈网络的处理方法时, 样本的训练次序与其本身的演化过程没有关系, 如在文献<sup>[4]</sup>中, 用于前馈网络训练的样本是分布在吸引子轨道附近的1500组矢量场数据; 文献<sup>[11]</sup>采用了一种轨迹学习的方法, 用于网络学习的样本是一条有限长度的时间轨迹, 在训练网络的时候考虑了样本的时序过程和网络的回归结构。现有的方法很少讨论递归神经网络的泛化能力, 而且相应的评价准则不够客观, 在文章<sup>[4]</sup>和<sup>[11]</sup>中仅仅使用一条轨迹作为测试样本, 由于单条测试样本仅仅表示了非线性吸引子的某一个局部的特征, 以此作为评价准则建立起来的神经网络对新轨迹的预测风险较大。本文从训练样本和网络结构两方面讨论了递归神经网络泛化性能的影响因素, 建立了多条轨迹的评价准则。多条轨迹代表了吸引子多个局部特征, 可以更全面客观评价网络的预测性能。

## 2 网络结构和学习算法(Neural network structure and learning algorithm)

本文使用的递归神经网络的基本方程如公式(1)表示, 图1为网络结构图。

$$\hat{x}(t+1) = W_{x1} \cdot \sigma(W_{x2} \cdot \hat{x}(t) + b). \quad (1)$$

其中:  $\hat{x} \in \mathbb{R}^s$ ,  $W_{x1} \in \mathbb{R}^{s \times N}$ ,  $W_{x2} \in \mathbb{R}^{N \times s}$  和  $b \in \mathbb{R}^N$  分别是网络的连接权矩阵和偏置向量, 隐含层非线性函数  $\sigma([z_1, z_2, \dots, z_N]^T) = [\sigma(z_1), \sigma(z_2), \dots, \sigma(z_N)]^T$ ,  $\sigma(\cdot)$  为Sigmoid函数。

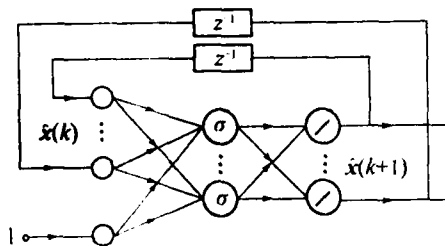


图1 递归神经网络结构

Fig. 1 Structure of the recurrent neural network

对于如公式(2)描述的非线性自治系统

$$x(t+1) = f(x(t)), \quad (2)$$

使用神经网络(1)来学习动态系统(2), 从而使得神经网络具有相应的动态特性。动态系统(2)可以表现出多种吸引子特性。在下面的讨论中, 假设动态系统(2)为极限环运动的情形。

在网络训练过程中, 通常需要首先计算评价函数对网络权值和偏置的偏导数信息, 然后使用基于梯度的无约束优化算法来训练网络的连接权值和偏置。通常可以通过动态BP学习算法(DBP)和时序反向传播算法(EPTT)来获得评价函数对网络权值的偏导数。本文采用的算法是批量BPTT算法, 在推导算法的过程中使用文献<sup>[12]</sup>的表示形式。BPTT算法具有计算量小和易用性的特点。

学习的目标就是给定初始条件  $\hat{x}(0)$ , 在有限时间段  $[1, Q]$  上最小化  $J(W)$ :

$$J(W) = \sum_{i=1}^Q [x(i) - \hat{x}(i, W)]^T [x(i) - \hat{x}(i, W)].$$

其中:  $W$  为网络的可调参数,  $Q$  是样本轨迹长度。

为了优化连接权值, 首先要给出目标函数对  $W$  的梯度。为了方便推导, 式(1)可以写为

$$\begin{cases} n^1(t) = W^{1,2} \cdot a^2(t-1) + \theta, \\ a^1(t) = \sigma(n^1(t)), \\ n^2(t) = W^{2,1} \cdot a^1(t), \\ a^2(t) = n^2(t). \end{cases} \quad (3)$$

其中:  $W^{1,2} = W_{x1}$ ,  $W^{2,1} = W_{x2}$ ,  $a^2(t) = \hat{x}(t)$ ,  $a^1(t)$  为  $t$  时刻的隐含层输出向量,  $n^1(t)$ ,  $n^2(t)$  分别为  $a^1$  层和  $a^2$  层的加权和向量。目标函数对连接权值的导数可由下式得到(其中  $w^{1,2}(i,j)$ ,  $w^{2,1}(i,j)$ ,  $b(i)$ ,  $a_k$  等分别为  $W^{1,2}$ ,  $W^{2,1}$ ,  $b$ ,  $a$  的标量形式):

$$\frac{\partial J}{\partial w^{2,1}(i,j)} = \sum_{t=1}^Q \left\{ \sum_{k=1}^s \frac{\partial J}{\partial a_k^2(t)} \times \frac{\partial^e a_k^2(t)}{\partial n_i^2(t)} \right\} \frac{\partial^e n_i^2(t)}{\partial w^{2,1}(i,j)}, \quad (4)$$

$$\frac{\partial J}{\partial w^{1,2}(i,j)} = \sum_{t=1}^Q \left\{ \sum_{k=1}^s \frac{\partial J}{\partial a_k^2(t)} \times \frac{\partial^e a_k^2(t)}{\partial n_i^1(t)} \right\} \frac{\partial^e n_i^1(t)}{\partial w^{1,2}(i,j)}, \quad (5)$$

$$\frac{\partial J}{\partial b(i)} = \sum_{t=1}^Q \left\{ \sum_{k=1}^s \frac{\partial J}{\partial a_k^2(t)} \times \frac{\partial^e a_k^2(t)}{\partial n_i^1(t)} \right\} \frac{\partial^e n_i^1(t)}{\partial b(i)}. \quad (6)$$

其中上标  $e$  代表显导数<sup>[12]</sup>。计算显导数时, 不考虑由于时间延迟引起的间接导数,  $s$  为  $a^2$  层中神经元个数。可以看出式(4)~(6)中偏导数的求取需要计算  $\frac{\partial J}{\partial a^2(t)} \cdot \frac{\partial J}{\partial a^2(t)}$  可由式(7)计算:

$$\frac{\partial J}{\partial \mathbf{a}^2(t)} = \frac{\partial^e J}{\partial \mathbf{a}^2(t)} + \left[ \frac{\partial^e \mathbf{a}^2(t+1)}{\partial \mathbf{n}^1(t+1)^T} \times \frac{\partial^e \mathbf{n}^1(t+1)}{\partial \mathbf{a}^2(t)^T} \right]^T \times \frac{\partial J}{\partial \mathbf{a}^2(t+1)} \quad (7)$$

从式(7)可以看出  $\frac{\partial J}{\partial \mathbf{a}^2(t)}$  的计算依赖于  $\frac{\partial J}{\partial \mathbf{a}^2(t+1)}$ ，以此类推，整个梯度的计算是从  $\frac{\partial J}{\partial \mathbf{a}^2(Q)}$  开始的，这种求导次序与文献[9]和[11]不同。文献[9]和[11]使用的方法是动态BP算法，其中偏导数的计算从第一个训练样本开始的，这种方法需要取网络节点输出对每一个连接权和偏置的偏导数，这样，得到每一个权值和偏置的偏导数信息需要计算一个3维矩阵，矩阵中的每一个元素都需要一部分计算<sup>[13]</sup>，因而整体上计算较复杂。表1为BPTT算法和DBP算法的计算复杂度对比<sup>[13]</sup>，其中： $\theta$ 表示计算复杂度精确的阶， $w_U$ 为内部单元之间的连接权的数目， $w_U = 2sN$ ， $w_A$ 为网络中的可调参数数目， $w_A = ((2s + 1)N)$ 。可以看出，BPTT算法在计算时间上有较大的优势。

表1 BPTT, DBP算法每次迭代的计算复杂度  
Table 1 Computational complexity per time step for BPTT and DBP algorithms

算法	每次迭代的平均时间
时序反向传播算法	$\theta(w_U + w_A)$
动态BP算法	$\theta(w_U w_A)$

通过神经网络对有限长度样本序列的学习，希望神经网络在相应的状态空间中产生期望的吸引子轨迹，研究神经网络产生的轨迹和理想轨迹的差异，来考察神经网络的学习性能。为了较为全面评价神经网络的学习性能，采取了两种措施来提高评价的可信度：第一，在相同仿真条件下做大量的随机试验。为了避免诸如权值初始化等随机因素对网络学习的影响，对相同的仿真试验重复多次以避免不确定性因素对网络训练的影响；第二，根据吸引子的矢量场的趋势，在极限环的内部和外部分别选取多条轨迹作为测试样本，使选取的轨迹大致反映吸引子的动态特性。

### 3 学习周期运动吸引子轨迹(Learning the trajectories of periodic attractor)

van der Pol振子是电子管电路的数学模型，为典型的非线性周期运动。van der Pol方程如式(8)所示：

$$\begin{cases} \dot{x} = y, \\ \dot{y} = -p(x^2 - q)y - \omega^2 x. \end{cases} \quad (8)$$

这里令  $p = 0.7, q = \omega = 1$ 。从不同初始条件出发产生了若干条状态轨迹，为了较为全面而且客观地评价递

归神经网络的预测性能，特别地选择轨迹 I ~ IV 作为测试样本轨迹，每一条轨迹包含800个时间点(采样间隔0.1 s)，轨迹 I ~ IV 覆盖了 van der Pol 吸引子附近较大范围，如图2(a)所示。当测试网络性能时，将表2中轨迹 I ~ IV 对应的初始条件分别作为神经网络的初始状态，在初始状态作用下神经网络自由演化分别生成4条状态轨，然后对比其与理想轨迹 I ~ IV 的差别。这样，神经网络生成的轨迹仅仅依赖其初始状态，而不需要任何附加信息，换句话说，神经网络通过公式(1)的迭代而实现了长期预测。

表2 不同轨迹的初始条件

Table 2 Initial conditions for different trajectories

样本轨迹	初始条件( $x_0, y_0$ )
I	(0, 0.01)
II	(0, -0.01)
III	(2, -3)
IV	(-2, 3)
A	(0, 1.5)
B	(0, 2.1)
C	(0, 2.5)

接下来的两组仿真试验分别研究训练样本和网络结构对预测性能的影响。仿真条件如表3所示，网络训练的停止条件为训练样本的均方误差下降到  $5.0 \times 10^{-5}$  为止，此时，可以认为神经网络很好地学习了训练轨迹的特性。在第1组仿真试验中固定网络结构为  $\alpha$ ，考察训练样本 A、B 和 C 对网络预测性能的影响；第2组仿真试验固定训练轨迹为 A，考察结构  $\alpha$  和结构  $\beta$  在预测性能上的差异，其中网络参数的选择原则将在后面作详细的叙述。

表3 仿真试验条件

Table 3 Simulation conditions for the experiment

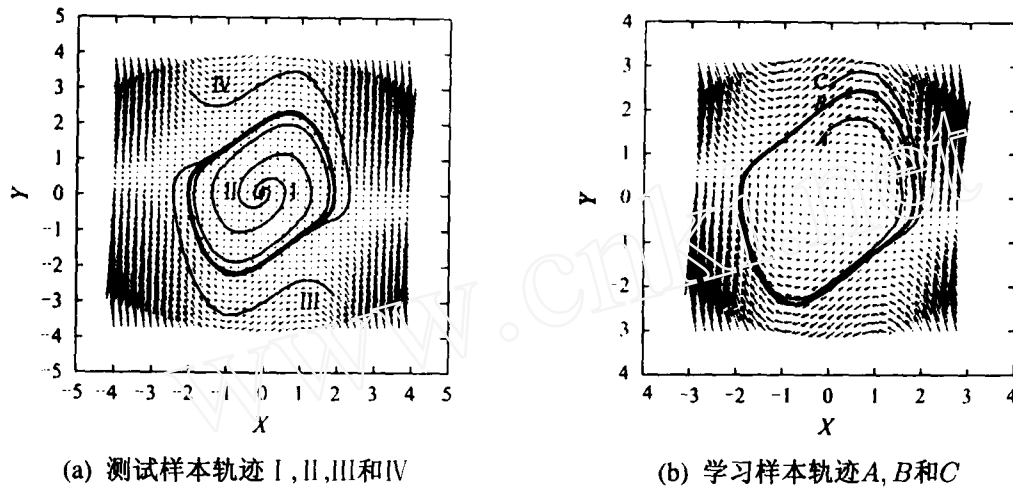
条目	取值或条件
网络结构	结构 $\alpha, \beta$
训练样本	轨迹 A、B、C
测试样本	轨迹 I、II、III 和 IV
训练算法	时序反向传播算法
初始权值	[-0.5 + 0.5]
停止条件	训练误差 $5.0 \times 10^{-5}$
每组随即训练次数	20

### 3.1 不同的样本轨迹对网络预测性能的影响(Influences of different training trajectories on the network prediction performance)

样本轨迹 A、B 或 C 如图2(b)所示，各条轨迹的初始

条件见表2. 轨迹A, B和C都是有限长度的时间序列, 这一点符合实际情况, 因为大多数实际观测的样本都是有限长度的时间序列. 文献[4]中提到的使用向量场获取样本的方法在实际中几乎不可能实现. 就训练样本数而言, 轨迹A, B或C的样本数仅为使用

向量场方法所得到样本数的1/6; 从样本的特性上看, 轨迹A, B或C更符合实际系统的演化. 在这组仿真中, 固定隐层节点个数 $N = 4$ (结构 $\alpha$ ), 分别对轨迹A, B和C3组训练样本进行学习, 每组样本随机训练20次.



(a) 测试样本轨迹 I, II, III和IV

(b) 学习样本轨迹A, B和C

图2 神经网络的测试样本轨迹和学习样本轨迹

Fig. 2 Evaluation trajectories and training trajectories

每个网络训练完毕后, 将测试轨迹 I, II, III和IV的初始条件设置成神经网络的初始状态, 使用其对4条测试轨迹进行预测, 并计算出网络输出演化轨迹与理想输出轨迹之间均方误差的均值和标准差, 如表4所示. 从表4中可以看出, 神经网络从轨迹A和轨迹C中学习的结果要优于从轨迹B中学习的结果. 在轨迹B上训练得到的神经网络在4条测试轨迹上均得到了最大的误差, 而且与轨迹A和轨迹C的学习结果相差1~2个数量级. 另外, 对轨迹A和轨迹C的学习效果也有明

显的差异. 图3(a)~(c)基本反映了神经网络 $\alpha$ 分别在轨迹A, B, C上的训练之后又分别在测试轨迹 I, II, III和IV上的预测情况. 从图中可以看出, 无论对极限环内部的测试轨迹(I和II), 还是对极限环外部的测试轨迹(III和IV), 从样本轨迹A和C上学习的神经网络都具有较好的预测性能, 轨迹的走向基本与理想情况吻合. 与此同时, 从轨迹B中学习的神经网络在极限环内部和外部的预测性能都较差, 与理想情况有较大差别(图3(b)).

表4 测试轨迹的均方误差(均值±标准差)(训练轨迹为A, B, C)

Table 4 Mean square error(Mean±Std) on the test trajectories (training trajectories are A, B, and C)

测试轨迹	训练轨迹A	训练轨迹B	训练轨迹C
I	$3.8337 \times 10^{-2} \pm 4.8720 \times 10^{-2}$	$2.0071 \times 10^0 \pm 1.9616 \times 10^0$	$2.2278 \times 10^{-1} \pm 2.7703 \times 10^{-1}$
II	$7.0555 \times 10^{-2} \pm 7.7477 \times 10^{-2}$	$2.1614 \times 10^0 \pm 1.9521 \times 10^0$	$2.4329 \times 10^{-1} \pm 2.5213 \times 10^{-1}$
III	$3.4685 \times 10^{-3} \pm 1.9369 \times 10^{-3}$	$2.9997 \times 10^{-2} \pm 3.3533 \times 10^{-2}$	$1.2555 \times 10^{-3} \pm 9.7564 \times 10^{-4}$
IV	$3.6885 \times 10^{-3} \pm 2.0736 \times 10^{-3}$	$2.9999 \times 10^{-2} \pm 3.3401 \times 10^{-2}$	$1.3059 \times 10^{-3} \pm 1.0118 \times 10^{-3}$

造成差异的原因在于3条训练轨迹包含的吸引子特征信息量不同. 轨迹B从紧靠极限环的位置出发并作趋向极限环的运动(如图2(b)), 因而仅仅包含了该吸引子极限环附近的信息, 训练轨迹A和C则分别从极限环内部和外部(距极限环稍远的位置)出发并逐步趋近于极限环(如图2(b)), 从而神经网络可以在轨迹A或者C上获得更多的吸引子特征信息, 因而在相同的测试样本轨迹

I, II, III和IV(如图2(a))上, 轨迹A和轨迹C获得较好的预测结果. 这里为了说明问题仅给出了轨迹A和轨迹C的结果, 大量的仿真表明, 其他轨迹也可以得到相似的结果.

### 3.2 网络结构对网络预测性能的影响(Influences of different structure parameters on the network prediction performance)

在前馈网络中, 通过控制神经网络的结构参

数(比如多层感知机中的隐层个数和隐层单元数),可以显著地提高神经网络的泛化能力,本组试验考察递回归神经网络的泛化能力与其结构的关系。

仿真试验是从较少的隐层Sigmoid单元开始的,当选择Sigmoid单元个数 $N \leq 2$ 时(可调参数 $\omega_A \leq 10$ ),在训练集上的均方误差维持在较高的水平上,在轨迹A上的均方误差几乎不可能下降到 $5.0 \times 10^{-5}$ 以下,当Sigmoid单元个数 $N = 3$ 时(可调参数 $\omega_A = 15$ )在轨迹A上的均方误差下降到 $5.0 \times 10^{-5}$ 仍较困难。此时神经网络的逼近能力不足,网络处于欠拟合状态。

当Sigmoid单元个数 $N \geq 4$ 时(可调参数 $\omega_A \geq 20$ )在轨迹A上的均方误差较容易下降到 $5.0 \times 10^{-5}$ 。

在公式(1)中分别指定 $N = 4, s = 2$ 和 $N = 7, s = 2$ 得到结构 $\alpha$ 和 $\beta$ 。在相同仿真试验下,通过对结构 $\alpha$ 和结构 $\beta$ 作重复的随机试验,对比其对新轨迹的预测能力。考虑到已经在3.1节的仿真试验中考察过结构 $\alpha$ 在轨迹A的学习情况,另外补充结构 $\beta$ 在轨迹A上的训练结果,表5为神经网络 $\beta$ 在测试轨迹上的预测误差。可以看到,神经网络 $\beta$ 的预测性能都有不同程度的恶化。在训练过程中发生了过拟合现象。图3(d)是典型的训练结果,可以看到,在极限环内部预测的轨迹与理想的轨迹有较大的差别。这里仅选择轨迹A说明问题。大量的仿真试验表明,其他类似的轨迹也有相同的结果:过多的网络可调参数会恶化递回归神经网络的预测性能。

表5 测试轨迹的均方误差(均值±标准差)(训练轨迹为A)

Table 5 Mean square error (Mean±Std) on the test trajectories (training trajectory is A)

测试轨迹	结构 $\alpha$	结构 $\beta$
I	$3.8337 \times 10^{-2} \pm 4.8720 \times 10^{-2}$	$1.1248 \times 10^0 \pm 1.0144 \times 10^0$
II	$7.0555 \times 10^{-2} \pm 7.7477 \times 10^{-2}$	$2.5498 \times 10^0 \pm 1.8048 \times 10^0$
III	$3.4685 \times 10^{-3} \pm 1.9369 \times 10^{-3}$	$3.5937 \times 10^{-3} \pm 2.5928 \times 10^{-3}$
IV	$3.6885 \times 10^{-3} \pm 2.0736 \times 10^{-3}$	$4.2064 \times 10^{-3} \pm 2.8552 \times 10^{-3}$

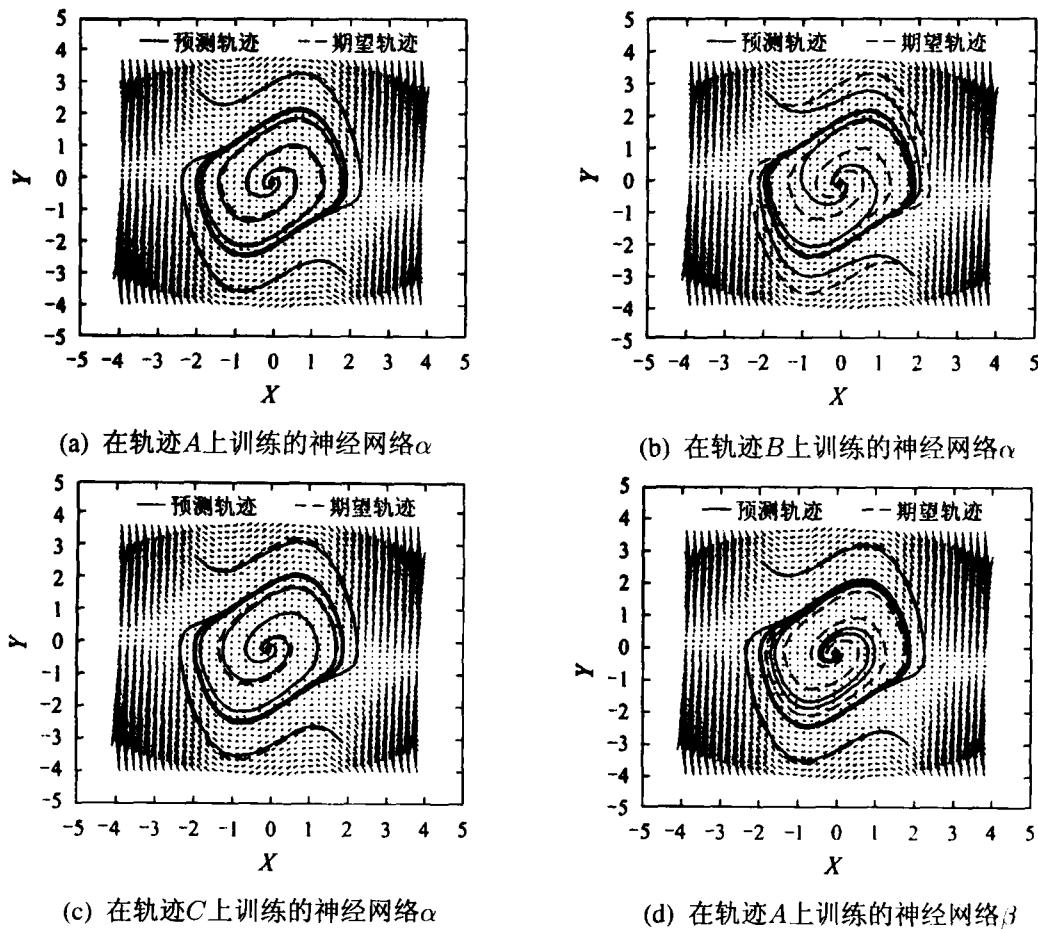


图3 神经网络在4条测试轨迹上的预测结果

Fig. 3 Evaluation results of the neural network on the four evaluation trajectories

在评价性能指标时,本文使用了多条测试轨迹的评判方法,相比之下文献[4]和[11]仅给出了一条测试轨迹,而相应的评价不够全面.从本文的仿真结果可以知道,神经网络结构 $\beta$ 虽然在轨迹III和IV上有较好的响应,但是在轨迹I和II的响应较差.

由以上分析可知,神经网络的性能受到学习样本轨迹和网络结构的影响,不可避免地产生模型误差,因而会对其长期的预测产生一定的影响.但是,由于周期运动不像混沌系统那样具有初始状态敏感性,所以长期预测的结果仍然具有较大的可靠性.正如在图3(a)和3(b)中所示,递归神经网络从第1步到第800步的迭代预测仍然具有较好的结果.

#### 4 结论(Conclusion)

本文使用递归神经网络对非线性周期运动吸引子轨迹进行学习,在不同的样本轨迹和网络结构条件下进行了递归神经网络的泛化能力讨论.引入多条轨迹评价准则,克服了仅仅使用一条轨迹验证的局限性,更为客观地评价递归神经网络的动态特性,更为全面地衡量递归神经网络在吸引子附近较大区域内的逼近和泛化性能.对van der Pol吸引子的仿真结果表明,神经网络的预测效果依赖于所学习的样本轨迹和神经网络的结构参数,从不同训练轨迹上得到的递归神经网络性能差异较大;从多条轨迹的评价准则可以看到,为了避免过拟合现象的发生,必须选择合适的神经网络结构,过多的网络可调参数会导致网络预测性能的恶化.

#### 参考文献(References):

- [1] CINCOTTI S, MARCHESI M, PILO F. Learning of Chua's circuit attractors by locally recurrent neural networks[J]. *Chaos, Solitons and Fractals*, 2001, 12(11): 2109 - 2115.
- [2] LI X, CHEN Z Q, YUAN Z Z, et al. Generating chaos by an Elman network[J]. *IEEE Trans on Circuits and Systems-I: Fundamental Theory and Applications*, 2001, 48(9): 1126 - 1131.
- [3] JAEGER H. Adaptive nonlinear system identification with echo state networks[M] //BECKER S, THRUN S, OBERMAYER K. *Advances in Neural Information Processing Systems*. Cambridge, MA : MIT Press, 2003, 15: 593 - 600.
- [4] TSUNG F S, COTTRELL G W. Phase-space learning[M]//TESAURO G, TOURETZKY D S. *Advances in Neural Information Processing Systems*. Cambridge, MA : MIT Press, 1995, 7: 481 - 488.
- [5] CHASSIAKOS A G, MASRI S F. Identification of structural systems by neural networks[J]. *Mathematics and Computers in Simulation*, 1996, 40(5/6): 637 - 656.
- [6] LIU D. Open-loop training of recurrent neural networks for nonlinear dynamical system identification[C]//*Proc of Int Joint Conf on Neural Networks*. Washington, DC: IEEE Press, 2001, 2: 1215 - 1220.
- [7] ELMAN J L. Finding structure in time[J]. *Cognitive Science*, 1990, 14(2): 179 - 211.
- [8] ZAMARRINO J M, VEGA P. State space neural network: properties and applications[J]. *Neural Networks*, 1998, 11(6): 1099 - 1112.
- [9] 任雪梅.非线性系统的回归网络识别[J].控制理论与应用, 2001, 18(6): 944 - 953.  
(REN Xuemei. Identification of nonlinear systems using recurrent neural networks[J]. *Control Theory & Applications*, 2001, 18(6): 944 - 953.)
- [10] RIVALS I, PERSONNAZ L. Black-box modeling with state-space neural networks[M] //ZBIKOWSKI R, HUNT K J. *World Scientific Series in Robotics and Intelligent Systems: Neural Adaptive Control Technology*. Singapore: World Scientific, 1996, 15: 237 - 264.
- [11] SUYKENS J A K, VANDEWALLE J. Learning a simple recurrent neural state space model to behave like Chua's double scroll[J]. *IEEE Trans on Circuits and Systems-I: Fundamental Theory and Applications*, 1995, 42(8): 499 - 502.
- [12] de JESUS O. Backpropagation through time for a general class of recurrent network[C] //*Proc of Int Joint Conf on Neural Networks*. Washington, DC: IEEE Press, 2001, 4: 2638 - 2643.
- [13] WILLIAMS R J, ZIPSER D. Gradient-based learning algorithms for recurrent networks and their computational complexity[M] //CHAUVIN Y, RUMELHART D E. *Back-Propagation: Theory, Architectures and Applications*. Hillsdale, NJ: Erlbaum, 1995: 433 - 486.

#### 作者简介:

韩敏 (1959—),女,大连理工大学电子与信息工程学院教授,博士生导师,研究方向为神经网络、专家系统、3S系统及混沌序列分析, E-mail: minhan@dlut.edu.cn;

史志伟 (1980—),男,大连理工大学电子与信息工程学院博士研究生,研究方向为神经网络、非线性系统辨识;

席剑辉 (1975—),女,大连理工大学电子与信息工程学院博士研究生,研究方向为神经网络系统辨识与混沌序列分析.