

基于一种新型聚类算法的 RBF 神经网络 混沌时间序列预测^{*}

张军峰[†] 胡寿松

(南京航空航天大学自动化学院, 210016 南京)
(2006 年 5 月 16 日收到 2006 年 6 月 19 日收到修改稿)

运用两阶段学习方法构建径向基函数(RBF)神经网络模型预测混沌时间序列.在利用非监督学习算法确定网络隐层中心时,提出了一种基于高斯基的距离度量,并联合输入输出聚类的策略.基于 Fisher 可分离率设计高斯基距离度量中的惩罚因子,可以提高聚类的性能.而输入输出聚类策略的引入,建立了聚类性能与网络预测性能之间的联系.因此,根据本文方法构建的网络模型,一方面可以加快网络训练的速度,另一方面可以提高预测性能.将该方法对 Mackey-Glass, Lorenz 和 Logistic 混沌时间序列进行了预测仿真研究,仿真结果表明了该方法的有效性.

关键词:混沌时间序列,预测,径向基神经网络,聚类

PACC: 0545

1. 引言

混沌现象是自然界中广泛存在的一种不规则运动,是一种由确定的非线性动力系统生成的复杂行为.随着非线性混沌动力学的发展,混沌时间序列的预测已经成为一个非常重要的研究方向,并已在信号处理、通信、控制、生物医学等领域中得到了广泛的应用^[1-3]. Takens^[4]提出的时间序列的延迟相空间重构和嵌入定理,为非线性时间序列预测奠定了理论基础.由于人工神经网络具有较强的非线性映射能力,因此在重构的相空间中,可以用神经网络逼近现在状态和未来状态之间的映射关系,进行混沌时间序列的预测.

径向基函数(RBF)神经网络作为前向神经网络的一种主要形式,具有结构简单,学习快速的特点,而且已经从理论上证明,只要给定足够多的网络隐层节点,RBF神经网络可以以任意精度逼近任何单值连续函数.由于RBF神经网络源于多变量差值理论,因此一种直观的设计方法是将所有的数据看作是网络的隐层节点,这样做虽然满足目标误差的要求,但是由于模型结构过于复杂导致网络的泛化性

能低下. Chen^[5]于1991年提出了正交最小二乘(OLS)算法设计RBF神经网络,这种算法具有计算量小,容易实现,并能在权值学习的同时确定隐层节点的数目等优点. Orr^[6]将正则化技术与OLS法结合起来,推导出正则化正交最小二乘(ROLS)学习算法,提高了OLS算法的效率. Gomm^[7]引入递归的思想,进一步丰富了OLS算法.但由于在有限样本的情况下,最优的数据中心不一定正好位于样本的输入点处,因此OLS算法从本质上看是次优的. Platt^[8]于1991年提出的资源分配网络(RAN),基于“新性”条件动态分配数据的中心.文献[9-11]对RAN进行了改进,在动态分配资源的同时加入了“剪枝”策略,将那些对网络输出贡献比较小的节点进行了移除,以保证网络的泛化性能.但是在网络调整的过程中一般采用最小二乘法对参数进行优化,由于隐层中心及扩展常数的非凸性,使其易于陷入局部极小点.

本文采用应用最为广泛的两阶段学习方法^[12]设计RBF神经网络进行混沌时间序列预测.首先采用减聚类技术,确定隐层单元的个数以及非监督学习初始中心,然后提出一种新型的度量,并根据该度量对训练数据进行输入输出聚类以获取网络的中心

^{*} 国家自然科学基金重点项目(批准号:60234010),航空科学基金(批准号:05E52031),国防基础科研项目(批准号:K1603060318)资助的课题.

[†] E-mail: wufeng7919@163.com

和扩展常数,如此可以确保网络的中心具有令人满意的信息.网络的输出权值由 LMS 算法确定.本文最后将这种方法分别应用于 Mackey-Glass, Lorenz 和 Logistic 混沌时间序列,仿真结果表明运用本文方法构建的神经网络预测模型,可以有效的预测混沌时间序列.而且,运用基于本文提出的度量方法对输入输出数据进行非监督学习获取的网络中心可以提高预测的精度,缩短网络构建的时间.

2. 径向基函数神经网络

RBF 神经网络具有典型的三层网络结构:输入层、隐层和输出层.隐层的作用是完成对输入量的非线性变换,本文选取常用的高斯基函数作为隐层的激活函数,网络的输出由隐层输出线性组合得到,如(1)式所示:

$$f(x) = \sum_{i=1}^c w_i \cdot \psi(\|x - v_i\|) \\ = \sum_{i=1}^c w_i \cdot \exp(-\|x - v_i\|^2 / 2\sigma_i^2), \quad (1)$$

其中 c 表示隐层节点的个数, v_i 和 σ_i 分别表示第 i 个隐层高斯基函数的中心和扩展常数, w_i 表示连接隐层和输出层的权值. RBF 神经网络设计的核心问题是:在满足目标误差要求的前提下,如何确定隐层节点的数目以及相应的中心,扩展常数和输出权值,同时应选择较为简单的模型结构以确保网络的泛化性能.

对于混沌时间序列的预测问题,其首要任务应根据嵌入定理重构输入输出空间.混沌时间序列的时间延迟嵌入状态空间为

$$x(t) = [x(t) \ x(t - \tau) \ \dots \ x(t - (m - 1)\tau)], \quad (2)$$

其中 $x(t)$ 是嵌入矢量, $x(t)$ 是第 t 时间点的序列值, τ 是时延量, m 是嵌入维数.此时可以通过嵌入矢量 $x(t)$ 预测第 $t + v$ 时间点的序列值 $x(t + v)$, v 表示预测步长.运用 RBF 神经网络实现混沌时间序列的预测可以表述为

$$x(t + v) = \sum_{i=1}^c w_i \cdot \psi(\|x(t) - v_i\|) \\ = \sum_{i=1}^c w_i \cdot \exp(-\|x(t) - v_i\|^2 / 2\sigma_i^2). \quad (3)$$

3. 聚类分析

运用 RBF 神经网络预测混沌时间序列成功与

否很大程度上依赖于网络模型结构的选择以及有效的学习算法.因此基于两阶段学习方法构建 RBF 神经网络,对混沌时间序列进行预测时无法回避以下问题:隐层节点个数的选择,非监督学习初始中心的确定,以及如何使非监督学习算法确定的网络中心具有丰富的信息量,即这些中心能够代表网络的输入数据.

聚类是一个非常重要的非监督学习问题,它可以简单地定义为:将目标分组,使得组内成员之间存在某种相似性的过程^[13]. K -均值聚类方法基于“近原型条件”,在 RBF 神经网络隐层参数的确定中应用广泛.具体算法如下所述:

假设给定数据集 $\chi = \{x_1, x_2, \dots, x_n\}$, $x_i \in \mathcal{R}^m$, 其中 $1 \leq i \leq n$, 寻找 c 个 ($c < n$) 原型 v_j , 使得目标函数

$$J = \sum_{j=1}^c \sum_{i \in I_j} \|x_i - v_j\|_2, \quad (4)$$

最小的过程.其中 I_j 表示索引集,满足等式 $\sum_{j=1}^c |I_j| = n$; v_j 表示聚类中心,或称为原型.

$$v_j = \frac{\sum_{i \in I_j} x_i}{|I_j|} \in V \subset \mathcal{R}^m \text{ 其中 } 1 \leq j \leq c. \quad (5)$$

“近原型条件”使 K -均值聚类方法直观引入,简单易用.同时也正是该条件导致 K -均值聚类方法存在一个极大的弊端,即过分依赖初始条件,而且聚类个数通常需要预先设定.由于引入不确定性,模糊 C -均值聚类方法优于 K -均值聚类方法.然而模糊 C -均值聚类方法虽然性能优越,但还是依赖初始条件且聚类个数同样需要预先设定.因此本文考虑引入减聚类来确定初始条件以及聚类个数.

3.1. 减聚类

Chiu^[14]提出的减聚类算法本质上是“峰”方法的一种改进形式,其基本思想是将每个数据样本看作是潜在的聚类中心,然后通过计算“峰”函数一步步地确定聚类中心.该方法可以独立地实现聚类功能^[15],也可以为其它的聚类方法提供初始中心,本文正是基于后者引入了减聚类算法.其具体算法如下所述:

1) 给定数据集 χ , 根据(6)式计算每个数据样本的密度值,密度值越大表明该样本的近邻包含的其他样本的数目越多;

$$D_i = \sum_{j=1}^n \exp(-\|x_i - x_j\|_2^2 / 2\gamma_a^2),$$

其中 $\gamma_a \in \mathcal{R}^+$. (6)

2) 选定最大的密度值, 假设为 D_{c1} , 所对应的样本 x_{c1} 作为第一个聚类中心, 并按(4)式对数据样本的密度值进行更新;

$$D_i = D_i - D_{c1} \sum_{j=1}^n \exp(-\|x_i - x_{c1}\|_2^2 / 2\gamma_b^2),$$

其中 $\gamma_b = \eta\gamma_a$ 且 $\eta \in \mathcal{R}^+$. (7)

3) 选定更新后最大的密度值, 假设为 D_{c2} , 所对应的样本 x_{c2} 作为第二个聚类中心, 并将 D_{c2} , x_{c2} 分别替代(7)式中的 D_{c1} , x_{c1} 对数据样本的密度值进一步更新, 然后选择下一个聚类中心. 当 $D_{ci} \leq \varepsilon D_{c1}$ 时, 终止循环, $\varepsilon \in (0, 1)$.

3.2. 输入输出聚类

在运用基于两阶段学习方法设计 RBF 神经网络结构时, 通常是针对训练样本集的输入样本进行非监督学习, 从而确定网络隐层的中心. 显然, 这种方法忽视了网路的输出信息. 文献[16]考虑了对由输入和输出数据构成的样本集合进行聚类分析, 以确定网络隐层的中心, 并分析了输入输出聚类性能和 RBF 神经网络预测均方误差之间的关系, 得出了输入输出聚类性能可以为 RBF 神经网络预测均方误差提供上界的结论. 如此, 可以通过选择有效的聚类算法来提高输入输出聚类的性能, 达到提高 RBF 神经网络预测精度的目的. 为了使设计的 RBF 神经网络的隐层中心具有丰富的信息量, 本文采用基于输入输出数据的聚类方法. 为了本文结构的紧凑, 该方法的理论分析和具体算法详见文献[16].

4. 基于高斯基的距离度量

距离度量在聚类算法中占有十分重要的地位. 基于有效距离度量的 K -均值聚类, 可以在保持“近原型条件”特性的同时, 优化聚类的性能. 欧氏距离是 K -均值聚类中一种常用的距离度量, 但它与各指标的量纲有关, 而且没有考虑指标之间的相关性. 为了消弱欧氏距离的不足之处, 文献[17, 18]定义了基于高斯基的距离度量, 记为 d_{GB} . 对于给定数据集 χ , 定义样本 x_i 与 x_j 之间的高斯基距离为

$$d_{GB}(x_i, x_j) = \sqrt{1 - \exp(-\beta \|x_i - x_j\|_2^2)},$$

$$\text{其中 } x_i, x_j \in \mathcal{R}^m. \quad (8)$$

根据文献[17, 18]可以证明高斯基距离满足非负性、对称性和三角不等式, 因此 d_{GB} 是距离度量. 其与欧氏距离的对比如图 1 所示.

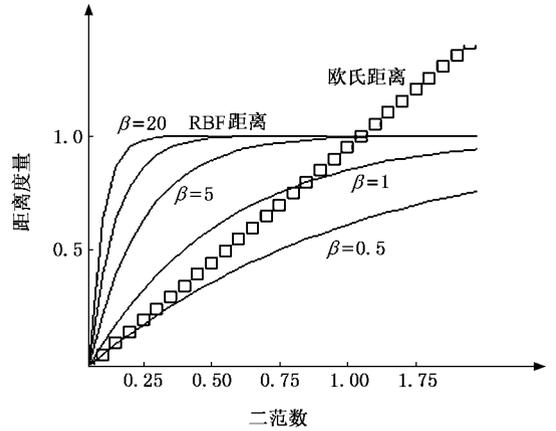


图 1 欧氏距离与高斯基距离的对比

此时基于高斯基距离度量的 K -均值聚类的目标函数为

$$J_{GB} = \sum_{j=1}^c \sum_{i \in I_j} \sqrt{1 - \exp(-\beta \|x_i - v_j\|_2^2)}, \quad (9)$$

依据文献[17], $\beta = \left(\frac{\sum_{i=1}^n \|x_i - \bar{x}\|_2^2}{n} \right)^2$, 其中 $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$.

考察 β 的定义可以发现: 当数据分布越紧凑时 β 值越大, 而 β 值越大表示在高斯基距离度量中惩罚的力度越强, 即虽然两个样本的欧氏距离较小, 但是对于紧凑的数据集而言, 可能是个野值点, 引入 β 可以得出较为正确的结论. 相反当数据分布越稀疏时 β 值越小, 此时 β 的惩罚力度也就越弱. 然而 β 是基于整体的数据样本而设计的, 势必导致取值的单一性. 如此高斯基距离在某些场合下将退化为欧氏距离.

根据以上讨论, 以及受 Fisher 可分离率的启发[19], 可以通过类内散布和类间散布设计惩罚因子, 对于不同的原型 v_j 采取不同的惩罚因子 β_j , 计算高斯基距离. 此时, 聚类的目标函数取为

$$J_{GB} = \sum_{j=1}^c \sum_{i \in I_j} \sqrt{1 - \exp(-\beta_j \|x_i - v_j\|_2^2)} \quad (10)$$

其中

$$\beta_j = \lambda / c, \quad (11)$$

$$\lambda = \max_{l=1, \dots, c} \left(\frac{sb(j, l)}{su(j) + su(l)} \right), \quad (12)$$

$$sb(j, l) = \|v_j - v_l\|_2,$$

$$su(j) = \frac{\sum_{i \in I_j} \|x_i - v_j\|_2}{|I_j|}. \quad (13)$$

当选取较多的聚类个数时, λ 值会很大, 采取(11)式计算惩罚因子可以避免因 β_j 值过大对聚类性能的影响.

由于引入了惩罚因子, 显然运用该度量进行 K -均值聚类, 可以加快聚类算法的收敛速度, 降低运算成本. 而且惩罚因子的倒数可以作为 RBF 神经网络隐层中心的扩展常数, 而采用传统的 K -均值聚类算法获得聚类中心后, 需另行计算扩展常数.

5. 仿真实例

下列仿真中均由预测序列均方根误差(RMSE)来衡量不同网络的预测效果, 且 IC-E 表示基于欧氏距离的输入 K -均值聚类; IOC-GB 表示基于高斯基距离的输入输出 K -均值聚类.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y(i) - \hat{y}(i))^2}, \quad (14)$$

其中 $y(i)$ 和 $\hat{y}(i)$ 分别为混沌时间序列的实际值和预测值.

5.1. Mackey-Glass 序列

Mackey-Glass 时间序列如(15)式所示

$$x(t+1) = (1-b)x(t) + \frac{ax(t-\tau)}{1+x(t-\tau)^p}. \quad (15)$$

本文选取 $a = 0.2$, $b = 0.1$, $\tau = 17$, 产生 1500 个数据, 其中前 1000 个数据作为训练样本, 后 500 个数据作为测试样本. 以嵌入向量 $x(t) = [x(t), x(t-6), x(t-12), x(t-18)]$ 预测 $x(t+v)$ 的值, 即嵌入维数 $m = 4$, 时延参数 $\tau = 6$, 预测步长 v 分别取为 1, 20, 50. 其中在利用资源分配网络算法构建混沌时间序列预测模型时, 由于网络训练的需要产生 3500 个数据, 前 3000 个数据作为训练样本, 后 500 个数据作为测试样本, 并依据文献[9]选用合适的参数. 在利用减聚类算法获得初始中心和隐层节点个数时, 选择参数 $\gamma_a = 0.3$, $\eta = 0.7$, $\epsilon = 0.275$. Mackey-Glass 时间序列的预测结果如表 1 所示, 其中

“时间”包括网络训练和测试的总时间.

表 1 Mackey-Glass 时间序列预测

	预测步长	隐层节点个数	预测精度	时间/s
IC_E	1	34	0.0107	10.391
IOC_GB	1	24	0.0065	5.166
IC-E	20	34	0.0140	10.110
IOC-GB	20	32	0.0103	5.868
IC-E	50	37	0.0444	11.125
IOC-GB	50	34	0.0301	6.191
		28	0.0512	33.375
OLS	50	29	0.0464	35.000
		30	0.0492	36.359
RAN	50	31	0.0503	11.781

由表 1 可以看出, 运用正交最小二乘算法和资源分配网络算法都能取得一定的预测效果, 不过比较费时, 而两阶段学习算法在运行时间方面是比较诱人的. 而且, 一方面在 K -均值聚类算法中运用高斯基距离度量比欧氏距离度量有更好的收敛性能, 因此运行速度加快; 另一方面, 引入输入输出聚类获得的 RBF 神经网络的隐层中心具有更为丰富的信息量, 因此在预测精度上要优于仅基于输入数据的聚类方法.

5.2. Lorenz 序列

Lorenz 时间序列如(16)式所示

$$\frac{dx}{dt} = s(y - x),$$

$$\frac{dy}{dt} = rx - y - xz,$$

$$\frac{dz}{dt} = xy - bz, \quad (16)$$

本文选取 $s = 16$, $r = 45.92$, $b = 4$, 运用四阶 Runge-Kutta 法得到微方程的离散序列, 积分步长为 0.01, 初始值为 $[1 \ 0 \ -1]$. 选择 1500 个数据, 其中前 1000 个数据作为训练样本, 后 500 个数据作为测试样本. 在使用该数据样本前, 首先对数据进行了预处理. 并以嵌入维数 $m = 5$, 时延参数 $\tau = 1$, 预测步长 $v = 1$ 为例, 分别采用 IC-E 和 IOC-GB 算法构建 RBF 神经网络预测模型. 其预测结果如表 2 所示, 图 2 给出预测效果.

由表 2、图 2 可以看出, 由于引入高斯基距离度量以及输入输出聚类的策略, 可使基于两阶段学习算法构建的 RBF 神经网络, 在保持相同的隐层节点个数的前提下, 提高混沌时间序列预测的效果.

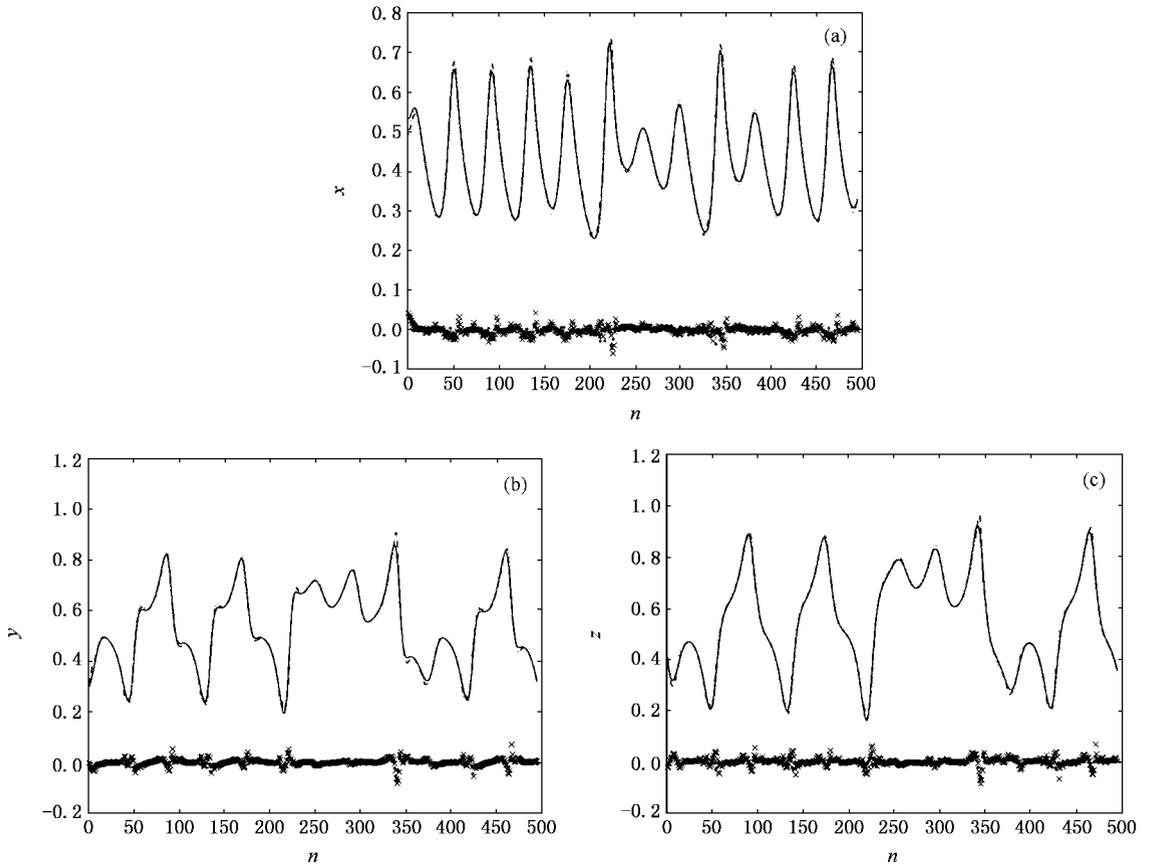


图 2 Lorenz 时间序列预测效果 图中实线表示实际序列,虚线表示基于 IOC_GB 算法的预测序列,点线表示基于 IC_E 算法的预测序列,点表示基于 IOC_GB 算法的预测误差,叉表示基于 IC_E 算法的预测误差

表 2 Lorenz 时间序列预测

算法名称	x		y		z	
	IC_E	IOC_GB	IC_E	IOC_GB	IC_E	IOC_GB
隐层节点个数	20	20	22	22	21	21
RMSE	0.0123	0.0085	0.0146	0.0109	0.0171	0.0110

5.3. Logistic 序列

Logistic 时间序列如(17)式所示

$$x_{t+1} = ax_t(1 - x_t), \quad (17)$$

本文选取 $a = 4$, 初始值为 0.1, 选择 2000 个数据, 其中前 1000 个数据作为训练样本, 后 1000 个数据作为测试样本. 并以嵌入维数 $m = 2$, 时延参数 $\tau = 1$, 预测步长 $v = 1$ 为例, 采用 IOC-GB 算法构建 RBF 神经网络预测模型. 选择参数 $\gamma_a = 0.3, \eta = 0.7, \epsilon = 0.15$, 得到 RBF 神经网络隐层节点个数为 30. 其预测误差如图 3 所示, 图 4 表示实际序列和基于 IOC_GB 的预测序列混沌吸引子图.

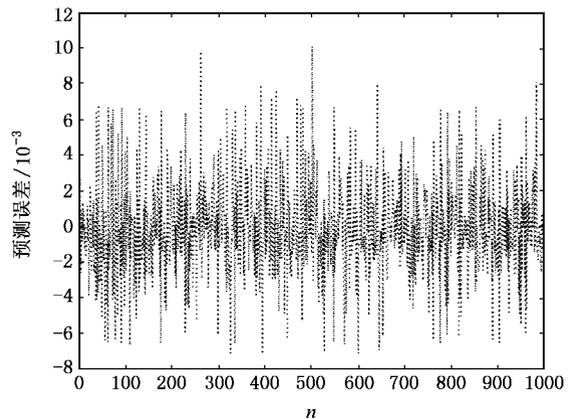


图 3 Logistic 时间序列预测误差

由于预测误差的绝对值都位于 0.01 之内, 且对

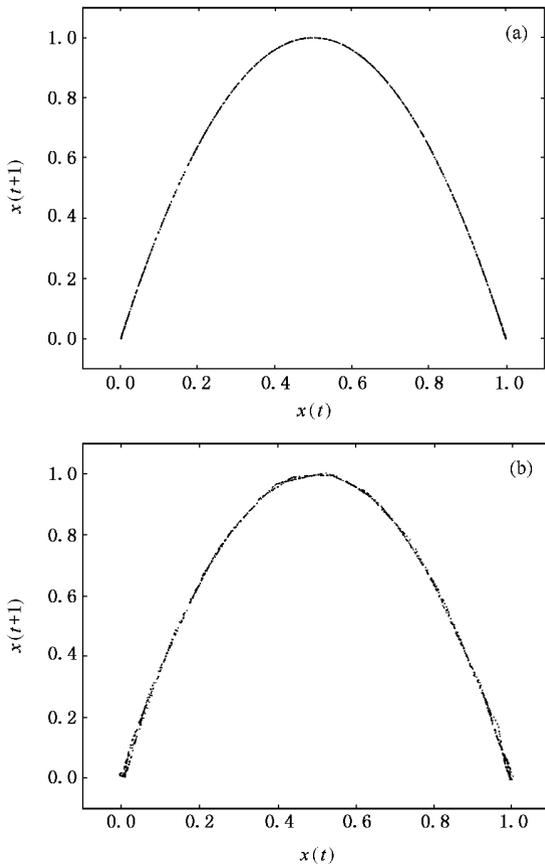


图4 Logistic 时间序列吸引子 (a) 实际序列 (b) 基于 IOC_{GB} 算法的预测序列

比实际序列和预测序列的混沌吸引子图可以看出, 基于本文提出的算法预测 Logistic 时间序列是有效的.

6. 结 论

径向基函数首先是在解决多变量插值问题时提出的, 然后被用于神经网络的设计. 因此 RBF 神经网络被看是一种全局逼近器, 只要给定足够多的网络隐层节点, 它能以任意精度逼近任何单值连续函数. 也正是基于以上的考虑, 可以运用 RBF 神经网络预测混沌时间序列.

RBF 神经网络具有十分特别的结构, 即隐层单元的非线性特性与输出单元的线性特性, 这就导致了其学习算法的多种多样. 为了获得简单的网络结构, 本文基于应用最为广泛的两阶段学习方法设计 RBF 神经网络. 同时提出了一种基于高斯基的距离度量, 并联合输入输出聚类的策略来提高网络的预测性能. 高斯基距离度量相比于传统的欧氏距离更为鲁棒, 它的引入可以明显提高聚类的性能. 而输入输出聚类建立了聚类性能与网络预测性能之间的联系: 聚类的均方误差为预测输出的均方误差提供了一个上界, 即可以通过提高聚类性能来达到提高预测性能的目的.

- [1] Wang H W, Ma G F 2004 *Acta Phys. Sin.* **53** 3293 (in Chinese)
[王宏伟、马广富 2004 物理学报 **53** 3293]
- [2] Li J, Liu J H 2005 *Acta Phys. Sin.* **54** 4569 (in Chinese) [李军、刘君华 2005 物理学报 **54** 4569]
- [3] Yu Z H, Cai Y L 2006 *Acta Phys. Sin.* **55** 1659 (in Chinese) [于振华、蔡远利 2006 物理学报 **55** 1659]
- [4] Takens F 1980 *Dynamical systems and turbulence* (New York : Springer)
- [5] Chen S, Cowan C F, Grant P M 1991 *IEEE Trans. Neural Networks* **2** 302
- [6] Orr M J L 1995 *Neural Comput.* **7** 606
- [7] Gomm J B, Ding L Y 2000 *IEEE Trans. Neural Networks* **2** 306
- [8] Platt C J 1991 *Neural Comput.* **4** 473
- [9] Rojas I, Pomares H, Bernier J L 2002 *Neurocomput.* **42** 267
- [10] Hu S S, Liu Y 2004 *Fuzzy Sets and Systems* **36** 403
- [11] Hung G B, Saratchandran P, Sundararajan N 2005 *IEEE Trans. Neural Networks* **16** 57
- [12] Schwenker F, Kestler H A 2001 *Neural Networks* **14** 439
- [13] Jain A K, Murty M N, Flynn P J 1999 *ACM Computing Surveys* **31** 264
- [14] Chiu S L 1994 *J. Intell. Fuzzy Syst.* **2** 267
- [15] Haralambos S, Ales A, George B A 2003 *Neurocomput.* **51** 501
- [16] Uykan Z, Koivo H N 2000 *IEEE Tran. Neural Networks* **11** 851
- [17] Wu K L, Yang M S 2002 *Pattern Recognition* **35** 2267
- [18] Zhang D Q, Chen S C 2004 *Pattern Recognition* **37** 173
- [19] Mao K Z 2002 *IEEE Trans. Neural Networks* **13** 1211



Chaotic time series prediction based on RBF neural networks with a new clustering algorithm

Zhang Jun-Feng Hu Shou-Song

(*College of Automation Engineering , Nanjing University of Aeronautics and Astronautics , Nanjing 210016 , China*)

(Received 16 May 2006 ; revised manuscript received 19 June 2006)

Abstract

Two-phase learning method is considered in this paper to configure the RBF neural networks for chaotic time series prediction. When determining the hidden-layer centers with the unsupervised learning algorithm, a new distance measure is presented based on Gaussian basis, and the strategy of input-output clustering is employed in combination. The punishment factor in Gaussian basis distance is designed based on Fisher separable ratio, which can improve the clustering performance. Moreover, the introduction of input-output clustering strategy establishes the relation between the clustering performance and the prediction performance. Therefore, the RBF neural networks constructed by this method can not only assure the compact structure, but also improve the prediction performance. This method is applied to Mackey-Glass, Lorenz and Logistic chaotic time series prediction, and the results indicate its validity.

Keywords : chaotic time series, prediction, radial basis function (RBF) neural networks, clustering

PACC : 0545

* Project supported by the Key Program of National Natural Science Foundation of China (Grant No. 60234010), Aeronautic Science Foundation of China (Grant No. 05E52031) and the National Defence Basic Research Program of China (Grant No. K1603060318).

† E-mail : wufeng7919@163.com