

基于广义径向基函数神经网络的非线性时间序列预测器¹

张 嵩 汪元美

(浙江大学生物科学与医学工程系 杭州 310027)

摘 要 该文对传统的径向基函数 (RBF) 神经网络的结构和学习算法进行了总结, 并在此基础上提出了广义径向基函数模型概念, 使这种网络具有更好的应用灵活性与可扩充性。文章基于 Mackey-Glass 造血模型方程的数值解数据, 对此广义模型与现有的 RBF 模型和梯度径向基函数 (GRBF) 模型对非线性时间序列预测问题的应用结果进行了比较与讨论, 显示出这种广义模型的应用有效性。
关键词 径向基函数, 神经网络, 非线性时间序列预测
中图分类号 TN-052

1 引 言

近年来, 随着以 BP 网为代表的前馈神经网络的不断发展与应用, 与这些网络一样具有强大非线性映射能力的径向基函数 (Radial Basis Functions, RBF) 神经网络也渐渐引人注目。这种神经网络在一定程度上克服了多层层间全互连网络的非线性复杂度, 学习算法简洁而容易控制, 在系统建模、非线性时间序列预测等方面都具有很好的应用。本文在传统 RBF 网结构的基础上提出了广义 RBF (GERBF) 神经网络模型, 同时对现有的 RBF, 梯度径向基函数 (GRBF) 等模型的特点、学习算法作了总结, 并且基于模拟数据将这些模型对非线性时间序列预测问题的应用结果作了比较与讨论。

2 RBF 网络概述

图 1 给出了一个单隐层多输入、单输出的 RBF 网的结构示意图, 多输出的情况只是单输出的简单扩充。与一般的三层层间全互连前馈网相比, RBF 网的独特之处在于隐层节点的激励方式上。设 \mathbf{X} 为输入矢量, 即

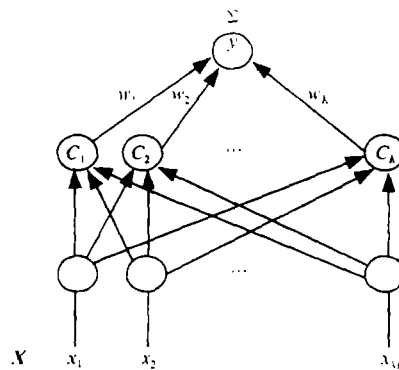


图 1 RBF 网络结构

$$\mathbf{X} = [x_1, x_2, \dots, x_M]^T, \quad (1)$$

¹ 1999-02-11 收到, 1999-11-29 定稿

其中 M 为输入矢量维数. 若 h_i 代表第 i 个隐层节点对 X 的响应, 则

$$h_i = \phi(\|X - C_i\|), \quad (2)$$

$\phi(\cdot)$ 为某函数. 因此与一般前馈网不同, 在 RBF 网中, 每一个隐层节点都与一质心 (Centroid) 相联系, 而其输出则为此质心与输入矢量的欧氏距离的函数. 最后设 y 为整个网络的输出, 则

$$y = \sum_{j=1}^K w_j h_j, \quad (3)$$

其中 w_j 为第 j 个隐层节点与输出节点的联接权值, K 为隐层节点个数, 即网络输出为隐层节点输出的线性加权和.

我们知道, 多层层间全互联前馈网可调权值多, 非线性映射能力强, 而其全部应用皆源于此. 这种网络有可行的学习算法——BP 算法, 在实践中发挥出强大的威力. 但是, 人们也认识到: 网络极强的非线性常常使基本上属于梯度下降法的 BP 算法陷入局部极小, 在应用中需要一定的经验与技巧, 且这种困难随问题规模的增大而增大. 在这方面, RBF 网络不失为一可行的替代方案. 首先, 理论上已证明^[1]对隐层节点激励函数 $\phi(\cdot)$ 加很小的限制, 那么对于非常广泛的函数类中的任意给定函数, 适当调节隐层节点的个数和质心值及隐层到输出层的联接权值, RBF 网络可以逼近它至任意精度 (在某种范数意义下), 这就为 RBF 网络的非线性映射能力乃至其应用提供了理论依据. 另一方面, RBF 网络在一定程度上克服了多层层间全互联前馈网的非线性复杂度, 且其学习算法简洁而易于控制, 而这正是我们在下一节将论述的.

3 学习算法

一种神经网络的真正应用离不开可行的学习算法, 不同结构与特性的神经网络具有不同的学习算法. 以下就来论述 RBF 网络的具体学习算法, 将要看到, 由于网络的非线性复杂度大大降低, 其学习算法基本上是直接的单向求解过程, 不必如 BP 算法须不断前传信号反传误差, 直至满足精度要求.

仍以图 1 中的网络为参照, 设给定训练集可表示为如下的输入输出模式对集:

$$\{[X_1, y_1], [X_2, y_2], \dots, [X_N, y_N]\}, \quad (4)$$

其中 X_i 为输入模式, y_i 为对应的期望输出, $i = 1, 2, \dots, N$, N 为输入模式个数. 在 RBF 网络中, 我们可以调节的量有: 隐层节点激励函数 $\phi(\cdot)$, 隐层节点质心 C_i , 隐层节点个数 K 和隐层节点至输出的联接权值 w_i . 理论研究表明^[1]: $\phi(\cdot)$ 对于网络的非线性映射能力影响不大, 一般可在高斯函数或类似的常用初等函数中任选. 直观地考虑, 质心 C_i 应与输入模式有关, 因此最初人们使用如下学习算法: 先在输入模式集中随机选择 K 个模式作为质心 C_i , 然后依次以给定训练集进行训练, 得到以下线性方程组:

$$PW = Y. \quad (5)$$

W 为联接权值, Y 为期望输出, P 为 $N \times K$ 系数矩阵, 其元素 p_{ij} 为第 j 个隐层节点对第 i 个输入模式的响应, 按照最小二乘法最后确定联接权值. 但是, 这样的学习算法是难以使人十分满意的. 首先, 直观看, 质心集应是输入模式集的适当采样, 随机选取带有很大的盲目性. 另外, 由于是随机选取, 为达到一定精度必然将不必要地增大隐层节点个数, 而这种做法又常常使系数矩阵 P 的病态性增大, 解的数值稳定性变得很差. 因此, 文献^[2]提出用正交最小二乘法 (Orthogonal Least Square, OLS) 同时适当选取 C_i 与联接权值 W .

由线性代数知识知, 对于一个 $N \times K$ 满秩矩阵 P 可由正交化手续将其化为

$$P = QA, \quad (6)$$

其中 Q 为 $N \times K$ 的一般正交矩阵, 即

$$Q^T Q = D. \quad (7)$$

D 为 $K \times K$ 对角矩阵, A 为对角线元素为 1 的上三角阵, 即 $a_{ij} = 0, i > j; a_{ii} = 1$. 熟知的古典 Gram-Schmidt 等方法都可完成这种正交化过程. 这时 (5) 式化为

$$QAW = Y, \quad (8)$$

而解 W 的过程就分为两步: 首先解

$$QZ = Y, \quad (9)$$

其中 Z 为 K 维向量, 这一最小二乘问题由于 Q 的正交性而变得容易; 然后解

$$AW = Z. \quad (10)$$

由于 A 的三角性, 可以通过反向回代方便地求得最后的 W .

现在来叙述文献 [2] 的思想: 首先令隐层节点质心集等于输入模式集, 即 $K = N$, $C_i = X_i, i = 1, 2, \dots, N$. 这时 (5) 式中的 P 矩阵规模为 $N \times N$, 然后通过正交化方法解 (5) 式. 这里需强调两点: 首先 P 矩阵不一定是满秩的, 但这并不妨碍正交化过程, 因为正交列向量是一步产生一个. 另外必须明确, 这种方法的指导思想是在 N 个输入模式中选取尽可能少但又具有代表性的质心集, 即使隐层节点数少但又不失精度. 下面以古典 Gram-Schmidt 方法为例叙述 OLS 算法. 设在第 $i+1$ 步开始已确定了 i 个正交列向量, 即矩阵 P 已化为

$$[Q_1, Q_2, \dots, Q_i, P_{i+1}, \dots, P_N], \quad (11)$$

其中向量 Q_1, Q_2, \dots, Q_i 互相正交, 则在第 $i+1$ 步 $P_j, j = i+1, \dots, N$, 被变换为 P'_j :

$$P'_j = P_j - \sum_{l=1}^i \alpha_l Q_l. \quad (12)$$

令 P'_j 与 Q_1, Q_2, \dots, Q_i 正交, 以确定各 α_l . 这里要讨论一个要点: 我们知道, 一向量 Y 由正交向量 Q_1, Q_2, \dots, Q_i 张成的子空间上的投影可表示为

$$\sum_{l=1}^i \frac{\langle Y, Q_l \rangle}{\langle Q_l, Q_l \rangle} Q_l. \quad (13)$$

令 g_l 表示上式中 Q_l 的系数, 则由于诸 Q_l 向量的正交性, 每个向量对 Y 的投影的精确贡献率为 $g_l^2 \langle Q_l, Q_l \rangle / \langle Y, Y \rangle$, 这个值定量反映由于引入 Q_l 导致 Y 到子空间投影距离减小的影响, 是决定 OLS 算法下一步方向的参照值. 也就是说, 在第 $i+1$ 步选取令贡献率最大的 P'_j 为第 $i+1$ 个正交向量. 需要指出两点: 首先, 前面已提到, P 很可能是亏秩的, 而这种算法能有效解决这个问题. 理论上讲, 当 P'_j 等于零向量时, 表明 P_j 与前 i 个正交向量线性相关. 因此, 我们可以预先确定一个小阈值 Δ , 当 P'_j 的模小于 Δ 时即可认为 P_j 与前 i 个正交向量近似线性相关, 而不把 P'_j 选作第 $i+1$ 个正交向量, 这样就大大提高了了解的数值稳定性. 另外, 我们需确定一个判别条件, 以便适时地终止循环. 这是说, 尽量提取主

要分量,使隐层节点个数不至太大,而又保证一定精度.设 Y 到由 P 的列向量张成的子空间的距离为 d ,选取数 ρ 稍稍小于 $1-d^2/\langle Y, Y \rangle$,那么当第 $i+1$ 个正交向量选定后,若

$$\frac{\sum_{l=1}^{i+1} g_l^2 \langle Q_l, Q_l \rangle}{\langle Y, Y \rangle} \geq \rho \quad (14)$$

就可终止循环.回忆前面的讨论,取上述判别式的意义是明显的.不过,实践中 d 往往未知,或需另外估计,我们代之上式以下面的终止条件:

$$\frac{g_i^2 \langle Q_i, Q_i \rangle}{\langle Y, Y \rangle} \leq \varepsilon, \quad (15)$$

ε 为一小阈值,即当某单步导致的 Y 投影距离改变量很小时就可终止正交化过程.从 OLS 算法选取正交向量的顺序可知上述条件的合理性.另外,当发现现有正交向量集与其它所有列向量相关时也应终止正交化过程.

最后,为进一步提高数值稳定性,我们采用改进的 Gram-Schmidt 正交化方法 (MGS),以下简称之.仍以满秩矩阵 P 为例,与古典 Gram-Schmidt 正交化方法 (CGS) 不同,在第 i 步 $P_j, j = i+1, \dots, N$, 被变换为 P'_j :

$$P'_j = P_j - \alpha_j Q_i, \quad (16)$$

令 P'_j 与 Q_i 正交以确定 α_j .也就是说,在第 i 步使 P_{i+1}, \dots, P_N 与 Q_i 正交,而且容易看出这种方法每步确定一行三角阵 A 的元素,其值即 α_j .算法的其余步骤如上所讨论. MGS 方法计算量与 CGS 方法相同,而数值稳定性与 Householder 正交变换法相当,这在 P 矩阵规模较大时是有利的,因此我们在以下的实验中采用这种正交化方法.

4 非线性时间序列预测

信号预测是信号处理领域的重要课题,其一般处理方法是先试图建立一个信号发生模型,以反映决定信号的主要因素.这样的模型往往是粗糙的,或包含很多参数,这就需要根据已经测得的数据对模型精确化,最后依据这一精确模型进行信号预测.因此,信号预测往往同时包含了系统辨识.

非线性或混沌时间序列预测这一新兴领域近年来随着人们对非线性及混沌深入广泛的研究发展起来.文献 [3] 对用模型法进行混沌时间序列预测作了详尽的讨论,并且对多种混沌序列用各种方法进行了预测,并对结果作了丰富的探讨和总结.我们知道,混沌是一种确定性因素产生高度不确定表象的非线性现象,如简单的抛物线映射引出复杂的动力学行为.这就使它与纯随机现象有了本质的区别,即在表面无序性背后蕴涵着更高层次的有序性.这就为非线性或混沌时间序列预测提供了某种直观的理论依据,更严格的理论论证是 Takens 的嵌入定理 (Embedding Theorem)^[3]: 对一个由奇异吸引子引起的状态空间连续演变过程适当地采样,那么对所得到的状态序列 $s(n)$ 及某一 m , 存在连续映射 $f: R_m \rightarrow R$ 使得对 $1 \leq n < \infty$

$$f(s(n+m-1), \dots, s(n)) = s(n+m). \quad (17)$$

m 称作嵌入维数 (Embedding Dimension), 而最小嵌入维数则记作 m^* . 而模型预测法即先确定某维数 $M \geq m^*$, 然后利用 M 维的逼近工具对未知的 f 建模,并根据已知数据在某一准则下确定模型参数,最后由此模型进行预测.而对基于 RBF 网络的预测方法,其逼近工具正是 RBF.

5 梯度 RBF 模型与广义 RBF 模型

在详细讨论基于 RBF 网络的非线性时间序列预测方法前,先介绍用于网络学习的训练

集。一般地, 给定的学习数据是某一已知的时间序列 $s(n)$, $n = 1, 2, \dots, L$, 其中 L 为序列长度。若确定嵌入维数即输入模式维数 $M \geq m^*$, 可将 $s(n)$ 化为 (4) 式所示的训练集, 对应关系为

$$X_i = [s(n_i - M + 1), \dots, s(n_i)]^T, \quad y_i = s(n_i + 1). \quad (18)$$

对训练集按 OLS 算法进行学习, 确定质心与权值, 最后根据此 RBF 网络进行时间序列预测便构成了基于 RBF 模型的预测方法。但是, 实践中发现: 就像在基于其它神经网络进行模式识别时一样, 对非平稳信号此方法效果不甚满意。因此文献 [4] 提出梯度径向基函数 (GRBF) 模型, 即隐层节点只响应输入信号的梯度, 也就是说在这种模型中, 原来的输入模式 X_i 全部换为

$$G_i = X_i - X_{i-1}. \quad (19)$$

这种推广是很自然的, 因为在预测问题中, 现时刻信号梯度信息比信号本身更重要。此外, 在 GRBF 模型中, 对隐层节点的激励方式根据所联系的质心做调节, 也就是说, 在 OLS 算法进行之前, 对 $C_i = G_i$ 的隐层节点, 确定其激励方式为

$$h_j = (s(n_j + 1) - s(n_j) + x_j)G(\|G - C_i\|), \quad (20)$$

其中 $G(x) = \exp(-x^2)$ 为高斯函数, x_i 为输入矢量中现时刻的信号值。即保证当输入信号 $X = X_i$ 时, 第 i 个隐层节点是相对于此输入模式的精确预测器, 而当 G 接近 G_i 时期望此隐层节点是一个较好的预测器。因此, 在这种模型中隐层节点的输出与现时刻的信号值以及梯度向量皆有关。

循着这一思路, 现将 RBF 模型作重要的推广。我们知道, RBF 网络原是多维函数空间的有力的逼近工具, 这在理论上已得到证明 [1]。但从神经网络的理论角度看, 它又代表了一种普遍的网络结构。回忆以 BP 算法为代表的层间全互联前馈神经网络, 在这种网络中系统的信息是完全分布于各个局部互联权值的, 系统所包含的信息非常丰富, 但同时极强的非线性在某种程度上阻碍了学习算法的顺利执行。而 RBF 网络结构则有所不同, 从神经网络的角度我们可将每一个隐层节点看作是响应一定模式的神经细胞群, 而与之相联系的质心则是存储于其中的模式信息。当输入一模式信号时, 各个神经细胞群都根据自身存储的模式信息对此作一定的响应, 而整个网络的输出则是各个神经细胞群输出的某种集总和。也就是说在这种网络结构中, 系统的信息存储既分布又集中。各个隐节点到输出节点的权值可理解为某种信息分布存储, 而各个质心中的信息则可看作是相对于各神经细胞群的信息局部集中存储。在这一普遍的基函数网络框架下完全可以引入比单纯径向基函数丰富得多的内容, 例如隐层节点的激励方式完全可根据应用作有利于它的调整, 而不必限于径向基函数类, 譬如高维多项式便是一个选择, 这时各质心中的信息可代表多项式系数。总之, 这种普遍的网络结构有很大的灵活性, 对各种应用的适应性也较强。但必须强调的是, 种种模型的学习过程都与前面论述的 OLS 算法密切相关。

回到非线性时间序列预测问题上。从 GRBF 模型中可以发现激励函数 (20) 式已不是严格的径向基函数, 如果将输入看作是由现时刻梯度矢量与信号量的联合矢量的话, 因此 GRBF 模型便是根据具体的应用对 RBF 网络所作的推广。实际上, 这种改进还不是最好的, 对于预测问题而言, 激励函数的一个更好的表达式是

$$h_j = (s(n_j + 1) - s(n_j))G(\|G - C_j\|) + x_j, \quad (21)$$

各符号的意义如前所述。这一式子的右边可明确解释为: 第一项表示期望改变量而第二项为现时刻信号大小。但从式子的具体形式看, 这种模型仍然是建立在径向基函数基础之上的, 因此所谓广义径向基函数网络模型是一种承袭一般基函数网络结构, 隐层节点激励方式基于径向基函数但又不同于单纯径向函数的普遍网络模型。这种模型包含了传统径向基函数

网络模型及梯度径向基函数网络模型, 在具备径向基函数网络的逼近能力的同时有着普遍基函数网络的应用灵活性。

6 实验结果与讨论

我们的试验数据是基于文献 [5] 引入的 Mackey-Glass 人体造血模型方程:

$$\frac{dx}{dt} = -\gamma x + \frac{\beta_0 \theta^n x_\tau}{\theta^n + x_\tau^n}, \quad (22)$$

其中 $x_\tau = x(t - \tau)$, 实验中参数取为: $\gamma=0.1/\text{day}$, $\beta=0.2/\text{day}$, $n=10$, $\theta=1.0$, $\tau=30\text{day}$, $x(t) = 0.5$, $0 \leq t \leq \tau$. 我们用欧拉改进型解此方程, 步长为 1day, 由此得到长度为 1000 的非线性时间序列。仿照文献 [4], 取 100-600 序列点为训练集, 600-1000 序列点为试验集。

以下讨论实验过程及结果。上面曾提到, 在 OLS 算法中当发现 P_j 近似为零时, 表明 P_j 与已求出的正交向量集近似线性相关, 则可将其舍弃, 这样可提高算法的数值稳定性。如前所述, 文献 [2] 提出当 P_j 的模很小时即认为其为零向量。考虑到向量的模只是向量的某种平均性质, 我们在实验中采取另一种我们认为更合理的判别法, 即当 P_j 的各分量绝对值的最大值小于某阈值 Δ 时即认为其为零向量。各实验参数取为: 相关阈值 $\Delta = 0.001$, (15) 式中的循环终止阈值 $\varepsilon = 0.0001$, 嵌入维数即输入模式维数 $M = 8$ 。图 2 为实验结果图示,

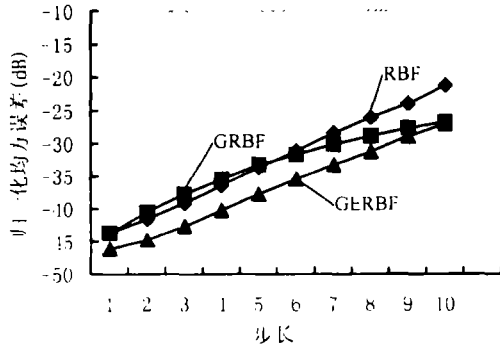


图 2 预测结果比较

其中 GERBF 表示广义径向基函数 (Generalized RBF) 模型, 即基于 (21) 式的预测结果。横坐标中的数字是预测步长, 步长为 2 意指以现时刻信息预测下下时刻序列值, 其它步长含义以此类推。步长超过 1 时, 预测是通过相继地以预测值作为新的输入来完成的。图中各数据点的纵坐标值为预测精确测度 m 。设预测点为 N 个, 预测值为 p_i , 实际值为 r_i , 则

$$m = \frac{\sum_{i=1}^N (p_i - r_i)^2}{\sum_{i=1}^N r_i^2}, \quad (23)$$

即 m 为归一化均方误差 (NMSE), 单位为分贝 (dB)。实验结果显示出广义模型对此应用问题的有效性。实验中还发现 RBF 的数值稳定性最佳, 广义模型次之, 并且相关阈值 Δ 的取值对 GRBF 与广义模型影响很大, 当 $\Delta < 0.0001$ 时, GRBF 几乎失效。

7 结 论

本文总结了已有径向基函数网络模型结构与学习算法, 并在非线性时间序列预测问题背景下提出了广义径向基函数神经网络模型概念。这一模型承袭了传统径向基函数网络的拓扑结构, 其隐层节点的激励方式基于径向基函数, 但又可根据应用作适当调整, 总体上体现了普遍基函数网络的思想。文章通过对模拟数据的应用结果证明了此广义模型方法的应用有效性。

从上面可见, 广义模型方法具有很好的灵活性与可扩充性, 同时与传统径向基函数网络相似, 其学习过程快速简洁. 因此, 只要能根据应用适当地选择激励函数, 可以期望这一方法对广泛的实时平稳与非平稳系统建模、信号预测等问题都有较好的应用. 至于更广泛的普遍基函数网络, 对其拓扑结构、激励方式及信息分布都可做更深入的研究.

参 考 文 献

- [1] Park J, Sandberg I W. Approximation and radial-basis-function networks. *Neural Computation*, 1993, 5(2): 305-316.
- [2] Chen S, Cowan F N, Grant P M. Orthogonal least square learning algorithm for radial basis function networks. *IEEE Trans. on Neural Networks*, 1991, 2(2): 302-309.
- [3] Casdagli M. Nonlinear prediction of chaotic time series. *Physica D*, 1989, 35(3): 335-356.
- [4] Chng E S, Chen S, Mulgrew B. Gradient radial basis function networks for nonlinear and non-stationary time series prediction. *IEEE Trans. on Neural Networks*, 1996, 7(1): 190-194.
- [5] Mackey M C, Glass L. Oscillation and chaos in physiology control systems. *Science*, 1977, 197(4300): 287-289.

NONLINEAR TIME SERIES PREDICTOR BASED ON GENERALIZED RADIAL BASIS FUNCTION NEURAL NETWORKS

Zhang Song Wang Yuanmei

(*Department of Life Science and Biomedical Engineering, Zhejiang University, Hangzhou 310027*)

Abstract The architecture and learning algorithm of traditional radial basis function (RBF) neural networks are surveyed in this paper. A generalized radial basis function model is proposed, which is more flexible and extensible. Based on the numerical solution to Mackey-Glass hematopoietic model equation, the prediction results obtained by radial basis function (RBF) model, gradient radial basis function (GRBF) model, and the generalized radial basis function model are compared and discussed, which show the effectiveness of the generalized model.

Key words Radial basis function, Neural networks, Nonlinear time series prediction

张 嵩: 男, 1973 年生, 博士生, 主要研究方向: 医学成像, 神经网络理论与应用.

汪元美: 男, 1945 年生, 教授, 博士生导师, 主要研究方向: 信息论, 医学成像, 神经网络.