

径向基函数神经网络的一种有效的在线学习方法¹

邓 超 熊范纶*

(中国科技大学计算机系 合肥 230027)

*(中国科学院智能所 合肥 230031)

摘 要 本文提出了一种径向基函数神经网络的有效在线学习方法。该学习方法不仅能根据输入信息的增加而动态地分配网络资源,而且能有效回收网络的冗余资源。在学习过程中网络的参数可以自适应地序贯进行调整。文中详细论述了这种神经网络的学习准则、动态增减隐节点算法和参数调整算法。同时通过分析和实验说明网络具有较强的映射能力和预测性能。

关键词 径向基函数,神经网络,在线学习算法,资源分配,函数映射,动态序列预测

中图分类号 TN-052

1 引 言

神经网络能够实现 n 维 Euclidean 空间的有界子集 A 到 m 维 Euclidean 空间的有界子集 $f(A)$ 的任意映射 $\varphi: A \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ [1]。在前馈神经网络中,MLP 和 RBF 神经网络是进行函数估计研究中最常用的两类网络。并且与其它前馈神经网络相比 RBF 网络具有最佳逼近特性 [2] 和全局近似性能 [3]。因此,到目前为止,已经发展了许多进行函数逼近和预测的 RBF 神经网络学习算法 [4-5]。所有这些方法都存在网络结构的规模(即网络资源的分配)不易事先确定的问题。因而,目前许多算法都采用逐次逼近的学习方法来对网络进行训练和优化网络的性能,属于离线学习算法。当环境数据是动态时,这类方法所需的存储量和计算量都非常大,为此,研究人员发展了一些神经网络的序贯学习算法用以在动态变化环境中进行有效学习 [6-8]。

Platt 1991 年首次提出了一种进行函数拟合的资源分配网络(RAN) [6]。这种网络是一种资源分配径向基函数神经网络。Kadirkamanathan 1993 年从函数空间映射角度出发从理论上对 RAN 进行了论证和解释,并在此基础上对 RAN 作了进一步的改进,提出了扩展 Kalman 滤波的资源分配网络(RAN-EKF) [7]。该网络不仅提高了对函数的拟合和预测精度而且更加有效地利用了网络资源(基函数)。但这些方法都仅考虑了当前样本的特点,对已经分配过的隐节点和网络现有规模没有统筹考虑,具有一定的片面性。因此,有必要在学习过程中引入隐节点裁减机制,以实现资源的最优分配。

本文提出了神经网络的一种有效的在线学习方法,我们称之为 P-RAN。它实际上是一种采用褒优贬劣策略进行基函数裁减的序贯学习 [8] 和资源动态分配网络。

2 P-RAN 的网络结构

本文所提出的 P-RAN 是一个两层前馈神经网络, x 为输入向量。第二层为隐单元层,它完成对输入的响应。这种径向基函数神经网络的输出可表示为

$$f(x) = \alpha_0 + \sum_{k=1}^K \alpha_k \phi_k(r) \quad (1)$$

¹ 1998-08-12 收到, 1999-05-09 定稿

国家自然科学基金重点项目资助

式中 α_k 代表第 k 个隐节点与输出之间的连接权值, K 为隐单元数. 函数 ϕ 被称为基函数, 它可以具有多种形式. 这里 r 的形式取为

$$r = \|\mathbf{x} - \mathbf{u}_k\|/\sigma_k \quad (2)$$

其中 $\mathbf{u}_k \in \mathfrak{R}^n$ 表示第 k 个基函数的中心. $\sigma_k \in \mathfrak{R}$ 为第 k 个基函数的宽度, 代表网络对输入空间数据的分辨力. $\|\bullet\|$ 为在 \mathfrak{R}^n 空间的 Euclidean 范数. 基函数 $\phi(\mathbf{x})$ 在本文选为 Gaussian 函数形式.

对于 \mathfrak{R}^n 空间中的 P 个观察样本 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_P$, 为了获得最佳的预测和泛化性能, 就必须确定最佳的基函数个数 K 、基函数中心 \mathbf{u} 以及比例因子 σ 的值. 同时, 还需要确定网络的最佳权值 α . 为了表示简洁, 我们将这几个网络参数表示为 $w(\alpha, \mathbf{u}, \sigma)$. 对于本文所述情况: 观察样本点是实时的, 且观察样本点还是一个接一个地序贯输入网络, 故基函数的个数就不能一次性地设定. 为此, 我们发展了一种高效的基函数的最优分配学习算法 P-RAN(pruning resource allocating network). 进一步, 为了加快这种算法的计算速度以便于实时应用, 本文采用 LMS 方法对网络参数 $w(\alpha, \mathbf{u}, \sigma)$ 进行自适应寻优.

3 P-RAN 基本原理及分析

设网络已有 K 个隐节点, 且记网络新到达的观察样本点为 (\mathbf{x}_n, y_n) . 此时, 如果我们不增加网络的隐层节点, 而只用现有的 K 个隐节点来估计最新样本 y_n , 即用前 $n-1$ 个样本点来估计第 n 个样本点 (见图 1). 如果网络已有资源能充分估计最新到达的样本点, 且记 y_n 在 K 维 Hilbert 子空间 H_K 的投影为 $f_K^n(\mathbf{x}_n)$, 则该投影与 y_n 之差

$$e_n = y_n - f_K^n(\mathbf{x}_n) \quad (3)$$

必等于零或在某一允许的范围之内. 此时, 用 $f_K^n(\mathbf{x}_n)$ 来估计 y_n 是完全可行的. 相反, 如果 e_n 超过一定界限, 则表明用网络已有资源不能充分估计最新样本点的信息, 即该新样本点包含了前 $n-1$ 个样本点所不具备的“新息”. 对于这种情况, 我们必须增加一个与 H_K 空间垂直的新隐节点才能表示这种“新息”, 即正确估计 y_n .

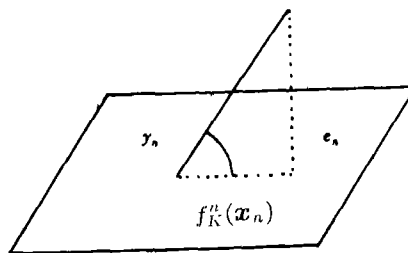


图 1 函数映射投影示意图

RAN, F-RAN, P-RAN 等算法正是基于这种序贯投影的基本思想进行算法设计的.

设初始 P-RAN 网络的隐层节点数为零. 将第一个观察样本点 (\mathbf{x}_0, y_0) 用来对网络进行初始化. 随着新观察样本一个接一个地序贯输入, 网络通过逐渐增加隐层节点和自适应地调整网络参数的方法对其中的部分信息进行表示和存储. 这里隐层节点增加的基本准则是输入信息新颖性的大小. 为此, 每当一个新的观察样本 (\mathbf{x}_n, y_n) 到达时, 网络都要判断这个新

输入的样本中是否包含有足够的“新息”。如果有,则允许网络增加一个隐层节点,否则只通过网络参数的自适应调整来减小对新样本的预测误差。这样就可用我们的网络充分表达到目前为止所观察到的主要信息。

为了有效地区分各种“新息”,在算法中需要根据具体的情况予以不同的处理。这里我们给出了在算法中需要用到的三种判定参数及其门限的设置。

(1) 网络输出误差超出最小误差门限:

$$e_n = y_n - f(x_n) > e_{\min} \quad (4)$$

其中 e_{\min} 为最小误差门限。

我们知道输出误差大说明该输入样本中包含有较大的“新息”成分,已有网络资源不能充分予以表达,故需要添加新的隐层节点来容纳该“新息”。

(2) 输入与已存模式的最近距离超过门限

$$\|x_n u_{nr}\| > \varepsilon_n \quad (5)$$

其中 u_{nr} 为输入空间中与 x_n 最邻近的存储模式(基函数), ε_n 为阈值, $\varepsilon_n = \max\{\varepsilon_{\max}\gamma^n, \varepsilon_{\min}\}$, γ 为一介于 0 和 1 之间的常数。 ε_n 的这种选取方法符合一般逼近理论中的先粗后细的逼近原则。它对应着我们的网络是先通过选取较宽的基函数来模拟该映射的基本结构,然后对基函数逐步精细化以进一步提高逼近精度。

(3) 网络输出误差大于相关混合误差 (CME)

$$e_n = y_n - f(x_n) > e_{\text{CME}} \quad (6)$$

其中

$$e_{\text{CME}} = \lambda \cdot e_{n-1} + (1 - \lambda)e_n, \quad (0 \leq \lambda \leq 1) \quad (7)$$

作出该项判断是为了配合本算法中的隐层节点裁减过程。为了克服某些节点反复被删除和加入的现象,使网络隐节点的增减较为平滑以提高学习速度,我们引入了这项判定。它能有效地避免这种振荡现象的产生。

当上述三个准则都同时满足时,则允许网络增加如下方式的一个新的隐层节点。新增隐层节点的各项参数可设置为

$$\begin{aligned} a_{K+1} &= e_n \\ u_{K+1} &= x_n \\ \sigma_{K+1} &= \kappa \|x_n - u_{nr}\| \end{aligned} \quad (8)$$

其中 κ 为覆盖因子,它决定了新增隐层节点在输入空间响应的覆盖度。

如果当新观察样本 (x_n, y_n) 输入网络时,上述三个准则不能同时满足,则网络不新增隐节点。但是,为了减小网络对新观察样本的表示误差,需要对网络参数进行适当的调整。本文采用 LMS 算法进行网络参数调整。

从另一方面考虑,由于径向基函数所具有的局域表示特性,这样网络在新增加隐层节点时就仅考虑了输入空间的局部响应,而没有从输入空间的整体进行均衡。当网络具有一定规模时,某些基函数的响应区域就会互相交叠,导致隐节点冗余现象。从而引起网络资源的极大浪费(见本文后面的实验)。为了充分有效地利用基函数资源,这里提出了如下的隐层节点裁减算法。

当隐层节点到输出之间的权值很小时,它对网络输出的贡献就很小,我们对它加一个惩罚因子 ζ ($0 < \zeta < 1$)。当其小于一定门限时,我们就可将其删除。具体地,每当新观察样本输入时,如果,

$$\alpha_j < \alpha_{pu}, \quad j = 1, \dots, K \quad (9)$$

则

$$\alpha_j = \zeta \alpha_j \quad (10)$$

这里 α_{pu} 表示惩罚门限。如果

$$\alpha_j < \alpha_{pr}, \quad j = 1, \dots, K \quad (11)$$

则将 j 所对应的隐节点从网络中删除, 即令隐层节点数 $K \leftarrow K - 1$ 。 (11) 式中的 α_{pr} 称为隐节点裁减门限。

4 P-RAN 算法

根据上面的分析和讨论, 本文所提出的 P-RAN 算法可总结表述如下:

第一步 初始化, 设 $\varepsilon = \varepsilon_{\min}$, $\alpha_0 = y_0$,

第二步 对每一个观察样本点 (x_n, y_n) , 计算网络输出 $f(x_n)$ 和误差 e_n 。找出最邻近中心: u_{nr} 。如果满足增加节点条件: $e_n > e_{\min}$ and $\|x_n - u_{nr}\| > \varepsilon_n$ and $e_n > e_{\text{CME}}$, 则增加新的隐层节点, 否则: 调整网络参数 α, u, σ 。

第三步 对每个权系数 α_j , $j = 1, \dots, K$, 判断是否满足加入惩罚因子和删减的条件, 如果满足则进行相应操作。并令 $K = K - 1$ 。转第二步。

5 模拟实验和性能分析

本文主要进行了两类典型的模拟实验。第一个实验是对 Hermite 多项式的静态逼近, 第二个实验是对 Mackey-Glass 混沌时间序列的四维动态远距离估计和预测。

在本实验中用作性能比较的 RAN 方法与文献 [8] 中稍有不同, 其差别是这里的参数调整中对基函数的宽度 $\sigma_i (i = 1, \dots, K)$ 项也进行了自适应调整。为了有别于 RAN, 我们称之为 RAN-E。显然, RAN-E 在性能上稍优于 RAN。

第一个实验是对 Hermite 多项式的静态逼近。Hermite 多项式的函数形式可表示如下:

$$f(x) = 1.1(1 - x - 2x^2) \exp\{-x^2/2\} \quad (12)$$

其中 $x \in \mathcal{R}$ 。

实验中所选取的训练样本集是 x 在 $[-4, +4]$ 区间内的 40 个随机采样点。其实验结果见图 2。图 2(a) 绘出了本文 P-RAN 方法与 RAN-E 方法的隐节点数随观察样点的变化曲线。从中可看出 P-RAN 方法的资源利用率远远优于 RAN-E 方法, 所用隐节点数少于 RAN-E 方法的五分之一。图 2(b) 为拟合误差均方根曲线。图 2(c) 给出了在采样区间内两种算法的拟合曲线和 Hermite 基本函数曲线, 其中符号 \circ 表示采样点。

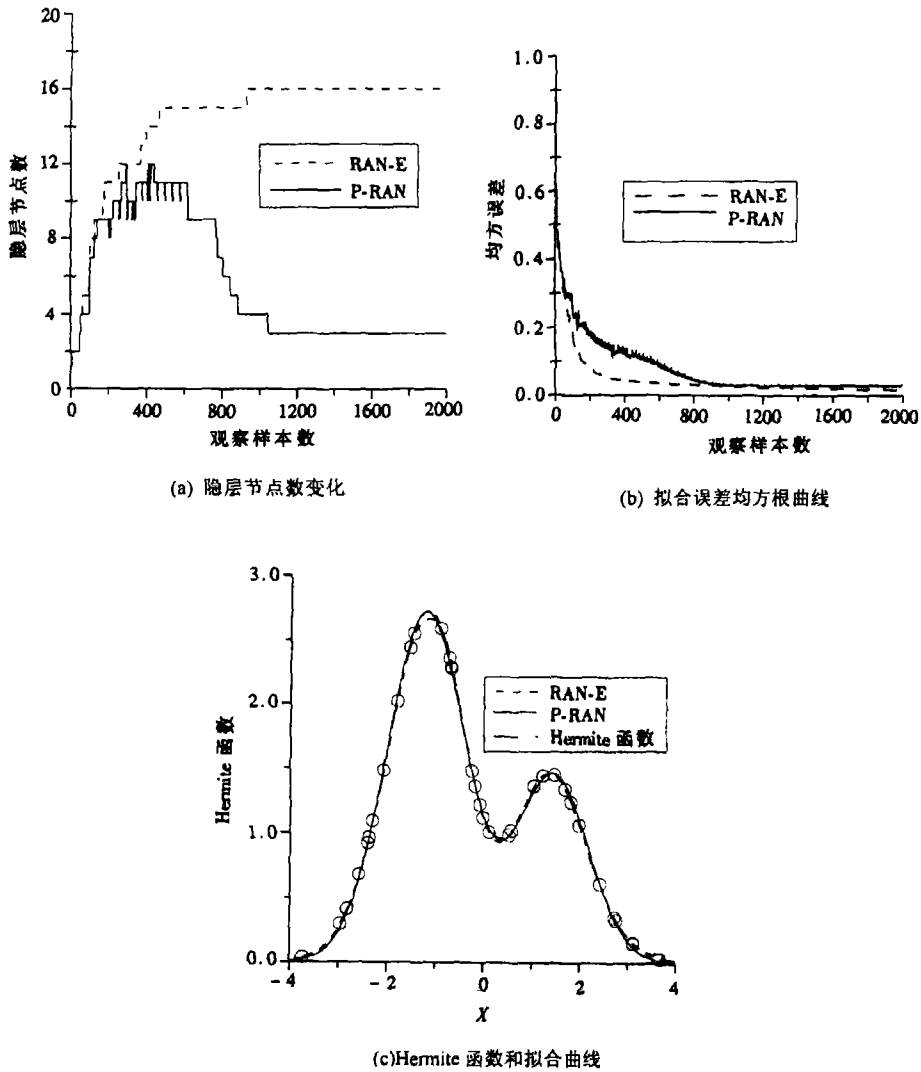


图 2 P-RAN 和 RAN-E 在 Hermite 函数逼近中的性能

第二个实验是 P-RAN 对 Mackey-Glass (M-G) 混沌序列的长距离预测。M-G 混沌序列由下列时延差分方程产生:

$$x(t+1) = (1 - 0.1)x(t) + 0.2 \frac{x(t-17)}{1 + x(t-17)^{10}} \tag{13}$$

实验中我们由 $x(t)$, $x(t-6)$, $x(t-12)$ 和 $x(t-18)$ 四个时刻的样点预测 $x(t + \Delta T)$ 。在本实验中 ΔT 取为 50。实验结果示于图 3，从中我们可得出与图 2 完全相同的结论。

6 结 论

本文提出了一种新的径向基函数神经网络学习算法。该算法能在网络学习过程中实时增加和裁减隐层节点，实现网络资源的更优分配。文中将所提出的算法与 RAN 算法在函数静态逼近，复杂混沌时间序列的近距离动态预测和远距离动态预测的性能进行了全面比较。实

验结果表明, 本文所提出的算法能用更少的隐层节点实现与 RAN 相近或更好的逼近和预测精度, 并且具有比 RAN 少得多的计算复杂度和存储量。

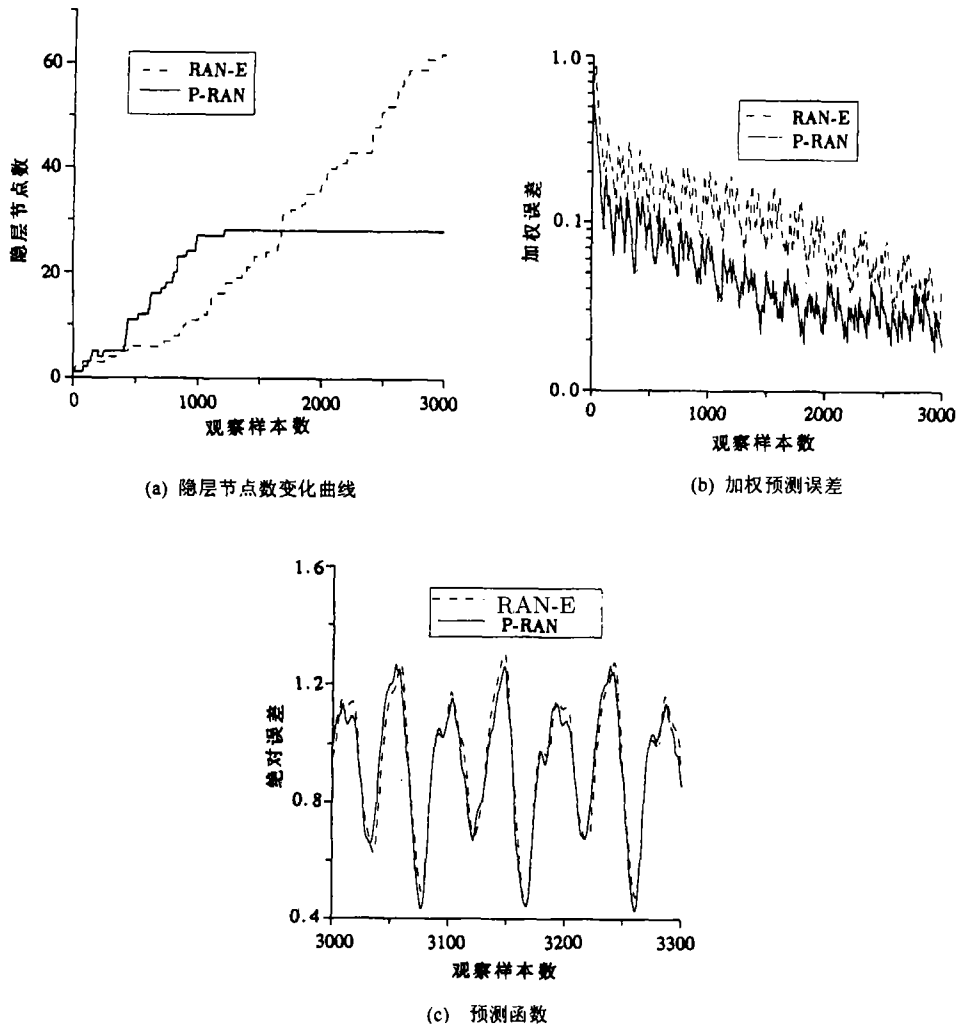


图3 两种方法预测的 M-G 混沌时间序列

参 考 文 献

- [1] R. H. Nielsen, Kolmogorov mapping neural network existence theorem. in Proc. Int. Conf. Neural Networks, 1987.3, 11-13.
- [2] T. Poggio, F. Girosi, A theory of networks for approximation and learning. AI Memo. No. 1140, AI Lab, Massachusetts Institute of Technology, Cambridge Mass: 1989.
- [3] J. Park, I. W. Sandberg, Universal approximation using radial basis function networks. Neural Computation, 1991, 3(2): 246-257.
- [4] Y. Tan, Z. Y. He, Arbitrary FIR filter synthesis with neural network. Neural Processing Letters, 1998, 8(1): 9-13.
- [5] A. W. Bruce, *et al.*, Cooperative-competitive genetic evolution of radial basis function centers and widths for time series prediction. IEEE Trans. on Neural Networks, 1996, NN-7(4): 869-880.

- [6] J. Platt, A resource-allocating network for function interpolation. *Neural Computation*, 1991, 3(2): 213-225.
- [7] V. Kadiramanathan, A function estimation approach to sequential learning with neural networks. *Neural Computation*, 1993, 5(8): 954-975.
- [8] Deng Chao, *et al.*, Sequential learning neural networks and its application in agriculture. *Proceedings of 1998 IEEE International Joint Conference on Neural Networks (IJCNN'98)*, Alaska, U.S.A: 1998, 366-370.

AN EFFICIENT ON-LINE LEARNING METHOD FOR RADIAL BASIS FUNCTION NEURAL NETWORKS

Deng Chao Xiong Fanlun*

(Dep. of Computer Sci., Univ. of Sci. and Tech. of China, Hefei 230027)

**(The Institute of Intelligent Machines, Academia Sinica, Hefei 230027)*

Abstract This paper proposes an efficient on-line learning method for radial basis function (RBF) neural networks. The proposed learning method not only dynamically allocate the network resource in accordance with the increase of input information, but also efficiently recycle the redundant resource of the network. During the learning process the parameters of the network can be sequentially adapted. The learning criterion, mechanism of increasing and decreasing resources and the parameter adjustment algorithm are elaborated. Meanwhile both the mapping approximation ability and predication performance of the network are analyzed in details.

Key words Radial basis function, Neural networks, On-line learning algorithm, Resource allocation, Function mapping, Dynamic prediction of time series

邓 超: 女, 博士生, 研究领域为神经网络学习算法, 数据挖掘与知识发现, 智能知识系统等。
熊范伦: 男, 1940年生, 教授, 博士生导师, 现从事计算机智能信息处理。