

径向基函数网络的 ABS 投影学习算法¹

文新辉 牛明洁

(中国石油天然气总公司信息中心 北京 100724)

摘 要 Broomhead(1988), Chen(1991) 等人提出的 RBF 网络的学习算法都是基于传统的 LMS 算法, 因此具有一定的局限性。本文提出了一种新的 RBF 网络的学习算法——ABS 投影学习算法, 它是一种直接的学习算法。计算机模拟的结果表明, 它具有学习效率高, 识别率高和适用范围广的优点。

关键词 神经网络, 模式识别, 学习算法

中图分类号 TN-052

1 引 言

径向基函数 (Radial Basis Function, RBF) 网络是近年来得到许多学者重视的前馈神经网络。Powell^[1,2] 对于 RBF 网络的数学基础进行了概括性的总结和发展。1988 年 Broomhead 和 Lowe 首先从神经网络的观点出发, 讨论了 RBF 网络及其实现算法^[3]。

事实上, RBF 网络是一种势函数的更精确的实现。BP 网络是基于非线性优化算法来进行学习的, 而 RBF 网络则是基于线性算法来进行学习的。RBF 网络更易于学习, 与 BP 网络相比能够大大地节约训练时间。对于 RBF 网络, S.Chen^[4] 提出了基于正交最小二乘法 (LMS) 的 RBF 网络的学习算法。Moody 和 Darken^[5] 提出了一种 RBF 网络两步学习算法, 即首先利用 K 均值法选择中心, 再利用 LMS 算法确定网络参数。

本文提出了 RBF 网络的 ABS 投影学习算法, 它利用求解两个线性方程组来完成最小二乘问题的求解。这里提出的 ABS 投影学习算法比一般的最小二乘算法要优越, 这体现在: (1) 较好地克服由于样本集的多样性和复杂性而引起的法方程的病态性; (2) 能够较大地减少计算量。

从理论上说, RBF 网络的 ABS 学习算法可以在有限步内完成, 这表明 ABS 投影学习算法要比 BP 学习算法具有更高的学习效率。

2 RBF 网络学习算法的理论分析

由于多输出的 RBF 网络总可以归化为若干单输出的 RBF 网络。因而不失一般性, 仅讨论单输出的 RBF 网络, 其拓扑结构如图 1 所示。

RBF 网络可表示为下列函数形式:

$$F(X) = \lambda_0 + \sum_{i=1}^m \lambda_i \varphi(\|X - C_i\|), \quad (1)$$

¹ 1994-12-08 收到, 1995-06-12 定稿

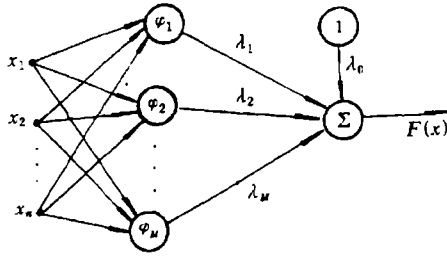


图 1 RBF 网络的拓扑结构

其中 $X(x_1, x_2, \dots, x_n)^T \in R^n$, $\varphi(\|X - C_i\|)$, $(i = 1, \dots, M)$ 是一组基函数.

对网络 (1) 式而言, 学习算法的优劣取决于三个方面, 即隐层非线性变换函数 φ 的确定, 网络中心 $C_i (i = 1, 2, \dots, M)$ 位置的选择和输出函数层参数 $\lambda_i (i = 0, 1, 2, \dots, M)$ 的确定.

隐层非线性函数 φ 的选择是一个确定基函数的过程 [6]. 它具有两方面的作用, 一是将 n 维

空间的点投影到一维空间, 以简化 n 维空间问题处理的复杂性; 二是对一组无限基的有限逼近.

定义 1 设 H 是具有模 $\|\cdot\|_H$ 的 Banach 空间. 如果存在一组 $\varphi_i \in H (i = 1, \dots)$, 使得对 $\forall h \in H$ 都存在唯一的一组实序列 $\{\lambda_i(h)\}, (i = 1, \dots)$ 使得

$$h = \sum_{i=1}^{\infty} \lambda_i(h) \varphi_i. \quad (2)$$

这里 (2) 式是在 H 中收敛的级数. 那么 $\{\varphi_i\} (i = 1, 2, \dots, \infty)$ 称为 H 上的一组 Schauder 基, 而 H 称为具有 Schauder 基的 Banach 空间.

比较 (1) 式与 (2) 式, 可以看出 (1) 式所采用的是一组有限基, 它是对 (2) 式中无限基的一种近似. 由此可以看出基函数选择的不同, 将会对无限基的逼近程度产生影响. Micchelli [7] 给出了 RBF 网络基函数选择的一个定理.

网络中心的选择是另一个重要的问题, 包括中心个数 M 的确定, 中心位置 C_i 的确定和基函数作用域 β_i 的确定. 中心位置 C_i 的选择有二种方式, 一种是随机选取, 另一种是有目的的选择. 在本文中, 我们仅限于随机选取方式.

现在的关键就是确定输出层参数 λ_i . 事实上, RBF 网络参数的确定, 就是学习的过程. 这是一个有教师的训练过程.

已知 N 对样本 $(X_i, Y_i), (i = 1, \dots, N)$, $X_i \in R^n$, $Y_i \in R^l$.

如果从 N 对样本 (X_i, Y_i) 中随机选取 M 个作为网络中心, 记为 $C_i (i = 1, \dots, M)$. 那么对 (1) 式, 令

$$\varphi_{ji} = \varphi(\|X_j - C_i\|), \quad (j = 1, \dots, N), \quad (3)$$

可得超定线性方程组:

$$\Phi \lambda = Y, \quad (4)$$

其中 $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_M)^T$, $Y = (y_1, y_2, \dots, y_N)^T$,

$$\Phi = \begin{bmatrix} \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1M} \\ \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{N1} & \varphi_{N2} & \cdots & \varphi_{NM} \end{bmatrix}$$

由此可见, RBF 网络的学习问题就化归为解线性方程组 (4) 式.

由数值计算的角度来看, 在 RBF 网络的学习过程中, 其可用性和效率与两个问题密切相关: 一是由于样本集的多样性和复杂性, 方程组 (4) 式常呈现病态, 使得利用传统的求解最小二乘解的法方程组很难得到解, 这就意味着学习的失败; 二是矩阵的行数等于样本点的总数, 从而当 N 很大时极大地增加了计算量, 不利于对大样本集的学习. 针对上述问题, 我们提出了 RBF 网络的 ABS 投影学习算法, 它不但能够较好地克服 (4) 式的病态, 而且能够减少计算量.

3 RBF 网络的 ABS 投影学习算法

我们可得 (4) 式的法方程为

$$\Phi^T \Phi \lambda = \Phi^T Y. \quad (5)$$

归纳起来, 现有的求解 (4) 式最小二乘解的方法一般可分为三类^[8]: (1) 用高斯消去法或 Cholesky 分解来求解 (5) 式; (2) 形成 A 的正交分解, 如 QR 分解; (3) 形成 A 的奇异值分解. 由于 RBF 网络本身问题的复杂性, 因此上述三种方法用于求解 (5) 式都会产生一定的问题. 方法 (1) 是一种直接分解的方法, 其结果的好坏依赖于矩阵 $\Phi^T \Phi$ 的条件数, 而这将对矩阵 Φ 的病态十分敏感. 方法 (2) 的采用无需知道 Φ 的秩, 但对某些情形会产生效率很低的“锯齿”现象. 方法 (3) 要花费大量的计算时间, 也是一种效率不高的方法.

对 (5) 式, 令

$$\Phi \lambda = \mu, \quad (6)$$

$$\Phi^T \mu = \Phi^T Y. \quad (7)$$

下面的定理保证了 (6) 和 (7) 式与 (4) 式的等价性.

定理 1 如果 (7) 式存在最小欧氏模解, 那么 (6) 式必存在解, 且 (6) 式的解就是 (4) 式的最小二乘解.

证明 如果 (7) 式有最小欧氏模解 μ^+ , 那么 Φ^T 的 Moore-Penrose 广义逆为 Φ^{+T} , 则有 $\mu^+ = \Phi^{+T} \Phi Y$, 代入 (6) 式有 $\Phi \lambda = \Phi^{+T} \Phi^T Y$, 进一步

$$\Phi^T \Phi \lambda = \Phi^T \Phi^{+T} \Phi^T Y = \Phi^T Y,$$

这说明 λ 的解存在, 且是 (4) 式的最小二乘解.

证毕

事实上, (6) 和 (7) 式的变换以及定理 1 保证了下面所提出的径向基函数网络的 ABS 投影学习算法的可用性.

算法 1

步骤 1 由样本集 $(X_i, Y_i) (i = 1, \dots, N)$ 中选取 M 个作为网络中心 $C_j (j = 1, \dots, M)$ 选定网络基函数. 由 (1) 式构造 (4) 式.

步骤 2 由 (4) 式可形成 (5) 式.

步骤 3 应用 ABS 投影算法求解 (7) 式, 得其解为 μ^* .

步骤 4 应用 ABS 投影算法求解 $\Phi \lambda = \mu^*$, 即可得 (4) 式的最小二乘解 λ^* .

步骤 5 由 λ^* 就可确定 RBF 网络 (1) 式. 停机.

算法 1 的关键运用两个线性方程组 (6) 与 (7) 式的求解来决定最小二乘解. 它具有以下两个优点. (1) 由于在算法 1 的运算过程中, 并不需要矩阵求逆, 从而对于求解病态问题具有较高的精度. (2) 在学习过程中, 并不需要直接求解法方程组, 从而避免了因解法方程组而加重系数矩阵的病态.

由算法 1 可以看出 ABS 投影算法对于学习效率的提高起着重要的作用. 最早的 ABS 算法是由 Abaffy^[9] 于 1979 年提出来的, 它受到了行之有效的解非线性方程组或非线性无约束最优化问题的拟 Newton 法的启发, 构造了一个在结构上与拟 Newton 法基本相同的算法, 这就产生了 ABS 算法. 从理论上讲, ABS 投影算法是一种直接算法, 它在精确运算的假定之下, 从任意点出发, 可以在不超过方程个数的迭代次数内, 收敛于方程组的解^[9].

对于算法 1 中的步骤 3 和步骤 4, 我们采用的是一种修正的 ABS 投影算法, 它与文献 [9] 中提出的 ABS 投影算法的不同之处在于它考虑到实际工程计算中舍入误差的产生, 从而采用了重置的策略.

由于篇幅问题, 关于 ABS 投影算法本身的详细讨论在此就不在叙述了, 可参见文献 [9].

4 应用举例

首先, 应用 ABS 投影学习算法对于三维 XOR(“异或”) 问题进行学习和识别.

例 1 识别三维 XOR 问题, 学习的样本点共有 8 个, 它们之间的对应关系是:

$$\begin{array}{llll} (0,0,0) \rightarrow 0.9 & (0,0,1) \rightarrow 0.1 & (0,1,0) \rightarrow 0.1 & (0,1,1) \rightarrow 0.9 \\ (1,0,0) \rightarrow 0.1 & (1,0,1) \rightarrow 0.9 & (1,1,0) \rightarrow 0.9 & (1,1,1) \rightarrow 0.1 \end{array}$$

我们选取的测试点共有 8 个, 它们是:

$$\begin{array}{llll} (1)(0.0,0.0,0.0) & (2)(1.0,1.0,1.0) & (3)(0.0,0.0,0.2) & (4)(0.0,0.0,0.8) \\ (5)(0.0,0.2,0.0) & (6)(0.0,0.8,0.0) & (7)(0.0,0.2,0.2) & (8)(0.0,0.8,0.8) \end{array}$$

应用算法 1, 选择基函数 $\varphi(r) = \exp(-3r^2)$, 网络中心点为 7 个.

在同样的硬件环境下, 如果误差 $\rho \leq 0.00001$, 那么 BP 算法需花费 845.15s, 而 ABS 投影学习算法只需花费 126.11s. 由此可见算法 1 具有较高的学习效率.

此时网络 (1) 式的参数为

$$\begin{array}{lll} \lambda_0 = -3.98876 \times 10^{-2}, & \lambda_1 = -1.53956 \times 10^{-8}, & \lambda_2 = 9.32454 \times 10^{-1}, \\ \lambda_3 = -9.36593 \times 10^{-9}, & \lambda_4 = 2.90739 \times 10^{-8}, & \lambda_5 = 9.32454 \times 10^{-1}, \\ \lambda_6 = 9.32454 \times 10^{-1}, & \lambda_7 = 9.32454 \times 10^{-1}. \end{array}$$

算法 1 的识别情况见表 1. ABS 算法与 BP 算法在精度与执行时间的比较见图 2.

表 1 ABS 投影学习算法与 BP 算法的识别比较

样本点	BP 算法	ABS 算法
(1)	0.899753	0.900000
(2)	0.133690	0.099999
(3)	0.707187	0.803287
(4)	0.077218	0.180005
(5)	0.710287	0.803287
(6)	0.326657	0.180005
(7)	0.893754	0.726221
(8)	0.893754	0.726221

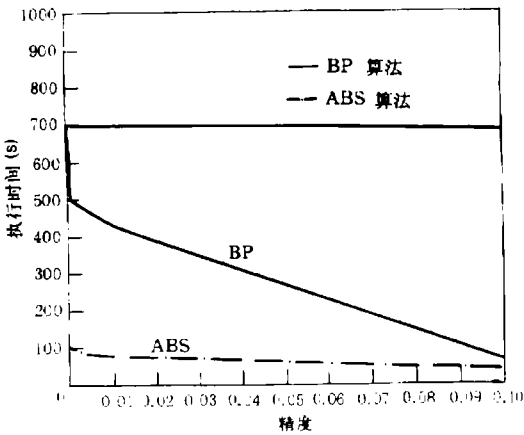


图 2 精度与执行时间的关系

大家知道在模式识别过程中，如果模式集具有重叠性时，进行分类是比较困难的，但运用 ABS 学习算法来分类，同样也会获得较好的结果。

例 2 有重叠的平面模式分类问题。问题如图 3 所示，第 1 类模式用“●”表示，其构成区域 Ω_1 ；第 2 类模式用“○”表示，其构成区域 Ω_2 。此时 $\Omega_1 \cap \Omega_2 \neq \phi$ ，这说明两类模式具有重叠区域。

表 2 不同学习算法的比较

序号	样本点	BP 算法	ABS 算法
(1)	(0.0,0.0)	0.000277	-0.000286
(2)	(0.3,0.0)	0.005875	-0.000325
(3)	(0.6,0.0)	0.992137	0.999595
(4)	(1.0,0.0)	0.003729	0.000310
(5)	(0.0,0.3)	0.005589	0.000326
(6)	(0.3,0.3)	0.993188	0.999682
(7)	(0.6,0.3)	0.007926	0.000444
(8)	(1.0,0.3)	0.997359	0.999663
(9)	(0.0,0.6)	0.992605	0.999594
(10)	(0.3,0.6)	0.006603	0.000445
(11)	(0.6,0.6)	0.992150	0.999465
(12)	(1.0,0.6)	0.999282	1.000301
(13)	(0.0,1.0)	0.005240	0.000311
(14)	(0.3,1.0)	0.998798	0.999662
(15)	(0.6,1.0)	0.999573	1.000300
(16)	(1.0,1.0)	0.999512	0.999986

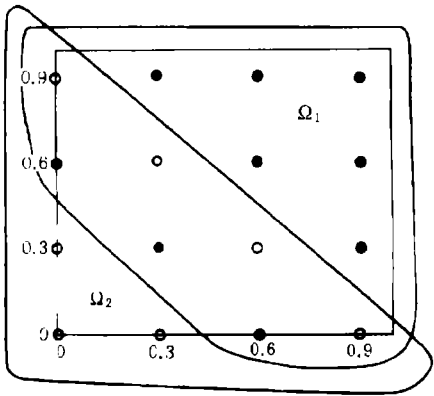


图 3 有重叠区域平面的模式分类问题

在硬件环境相同的条件下，取基函数 $\varphi(r) = \exp(-2r^2)$ ，中心点选 15 个，误差 $\rho \leq 10^{-5}$ 。那么运用算法 1 的学习时间为 412.53s，而对同样的问题，BP 算法却需要花费 2385.38s。

对此问题的识别见表 2。

5 结 论

本文对 RBF 网络提出了一种新的学习算法, 初步应用结果表明算法是可行的, 并具有较高的学习效率. 进一步的工作正在研究之中.

参 考 文 献

- [1] Powell M J D. Radial basis function for multivariable interpolation review. in algorithms for approximation. Mason J C, Cox M G Ed. Oxford University Press, 1987, 143-167.
- [2] Powell M J D. Radial basis function in 1990. in advances in numerical analysis Vol.2. Oxford University Press, 1991. 105-201.
- [3] Broomhead D S, Lowe D. Multivariable function interpolation and daptive networks. Complex System, 1988, No.2: 321-355.
- [4] Chen S, Cowan C F N, Grant P M. Orthogonal least squares learning algorithm for radial basis function networks. IEEE. Trans. on NN, 1991, NN-2: 302-309.
- [5] Moody J, Darken C. Fast-learning in networks of locally tuned processing units. Neural Comput., 1989, 1: 281-294.
- [6] Poggio T, Girosi F. Networks for approximation and learning. Proc. IEEE. 1990, 78: 1481-1497.
- [7] Micchelli C A. Interpolation of scattered data: Distance matrices and conditionally positive definite function. Constr. Approx., 1986, 2: 11-22.
- [8] Chen S, Billings S A. Orthogonal least squares methods and their application to non-linear system identification, International Journal of Control 1989 50: 1873-1896.
- [9] Abaffy J. [匈], Spedicato E. [意] 著, 诸梅芳, 官洋晶译. ABS 投影算法——求解线性和非线性方程组的数学方法. 北京: 北京工业大学出版社, 1991, 第三章.

THE ABS LEARNING ALGORITHM FOR RADIAL BASIS FUNCTIONAL NETWORK

Wen Xinhui Niu Mingji

(*Information Centre, China National Petroleum Cooperation, Beijing 100724*)

Abstract A new learning algorithm for radial basis functional network, which is called ABS project learning algorithm, is given and computing results show that it can be used for pattern recognition and other areas.

Key words Neural network, Pattern recognition, Learning algorithm

文新辉: 男, 1960 年生, 博士, 研究方向: 神经网络和模式识别, 优化技术和智能信息处理等.

牛明洁: 女, 1961 年生, 硕士, 研究方向: 计算机图象处理, 神经网络和模式识别等.