第 18 卷 第 2 期 控 制 与 决 策 2003 年 3 月
Vol 18 No. 2 Control and Decision Mar. 2003

文章编号: 1001-0920(2003)02-0233-04

# 带优选聚类算法的 RBF 网络辨识器及应用

刘铁男, 段玉波, 刘志德, 谢爱华, 张 航
(大庆石油学院 自动化与控制工程系, 黑龙江 大庆 163318)

**摘 要**: 以RBF 神经网络为模型框架, 解决非线性系统的辨识问题。针对 RBF 网络的结构辨识问题, 提出一种优选聚类算法, 并用该算法, 依据输入样本优选确定 RBF 神经网络的隐含层节点个数, 采用新型二阶递推学习算法估计 RBF 网络中的参数和权值。上述混合算法, 同时解决了 RBF 网络结构和参数辨识问题, 大大提高了 RBF 网络的建模和预测精度。应用实例表明了所提出方案的有效性。

**关键词**: RBF 神经网络; 优选; 聚类算法; 辨识; 二阶学习算法
**中图分类号**: TP18 **文献标识码**: A

# RBF network identifier with optimal selection cluster algorithm and its application

*LIU Tie-nan, DUAN Yu-bo, LIU Zhi-de, XIE Ai-hua, ZHANG Hang*
(Department of Automation and Control Engineering, Daqing Petroleum Institute, Daqing 163318, China)

**Abstract**: A model of RBF neural network (RBFNN) is framed to solve the problem of identification of nonlinear systems. In order to realize the structure identification of RBFNN, a kind of optimal selection cluster algorithm is proposed. By this algorithm, it is optimally gained the hidden layer node number of RBFNN in terms of input samples. The structure and parameters identification problems of RBFNN are simultaneously solved, so that the modeling and prediction precision of RBFNN are notably raised. The application example shows the validity of the scheme.

**Key words**: RBF neural network; Optimal selection; Cluster algorithm; Identification; Second-order learning algorithm

## 1 引 言

近年来, 由于径向基函数神经网络(RBFNN)具有拓扑结构简单和易于学习训练等优点, 已广泛应用于模式识别、函数逼近、信号处理、系统建模和控制等领域。但是如何通过合理提取输入样本的特征并对其有效聚类, 从而确定 RBFNN 隐含层节点个数, 以及如何进一步提高网络的辨识精度和泛化能力, 是值得致力研究的课题。

为了解决RBFNN 的结构辨识问题, 首先给出一个控制聚类合理性函数 $F$ 和一种优选聚类算法, 并基于该算法对输入样本的聚类过程进行优化; 然后取聚类的组数作为 RBFNN 隐含层节点的个数, 同时计算出RBF 参数的初值。这为进一步精确辨识 RBF 的参数和网络的权值提供了可靠的基础。为了保证网络具有很好的分类能力和泛化能力, 通过多目标决策方法调整函数 $F$ 的阈值, 在二者之间进行有效折衷。采用文献[1, 2]提出的具有二阶收敛速度的新型递推学习算法, 进一步改善RBFNN 参数以

**收稿日期**: 2002-01-29; **修回日期**: 2002-07-01。
**基金项目**: 黑龙江省自然科学基金资助项目(A 01-14); 黑龙江省教育厅科研基金资助项目(9551031)。
**作者简介**: 刘铁男(1945—), 男, 黑龙江青岗人, 教授, 从事非线性系统辨识、预测、滤波和控制等研究; 段玉波(1951—), 男, 黑龙江木兰人, 教授, 博士生导师, 从事油气田集输系统及自动化研究。

及权值辨识速度和精度。

## 2 优选聚类算法及 RBFNN 结构的确定

RBFNN 是带有输入层、隐层和输出层的 3 层前向网。不失一般性，假设输出层只有一个单元，但本文结果很容易扩展到多输出的情况。网络输入与输出间的映射关系为 $y(\mathrm{X}): R^n \to R$，即

$$y(\mathrm{X}) = w_0 + \sum_{i=1}^{m} w_i g_{ki}(\|\mathrm{X} - \mathrm{C}_i\|, \sigma_i) = w_0 + \sum_{i=1}^{m} w_i \exp\left(-\frac{\|\mathrm{X} - \mathrm{C}_i\|^2}{2\sigma_i^2}\right) \quad (1)$$

其中：$m$ 为隐含层节点数，$\|\cdot\|$ 为欧氏范数，$\mathrm{X} \in R^n$ 为输入向量，$\mathrm{C}_i \in R^n$ 为第 $i$ 个隐节点的中心，$\sigma_i \in R$ 为第 $i$ 个隐节点的宽度，$w_i$ 为第 $i$ 个 RBF 与输出层节点的连接权值，$w_0$ 为调整输出的偏移量。

为了寻求输入向量样本集的合理聚类结果，并据此确定 RBFNN 隐含层节点个数 $m$，采用与文献[3]类似的方法，给出如下一个控制聚类合理性函数

$$F = \frac{\frac{1}{N}\sum_{j=1}^{m}\sum_{p=1}^{N}\|\mathrm{X}_p - \mathrm{C}_j\|^2}{\frac{1}{m-1}\sum_{j\neq h}\|\mathrm{C}_j - \mathrm{C}_h\|^2} \quad (2)$$

当用适当方法确定了 RBFNN 隐层节点数 $m$ 及其中心 $\mathrm{C}_i (i = 1, 2, \cdots, m)$ 后，可用上式定量估价聚类的合理性。函数 $F$ 的意义为：若由聚类划分所形成的同一类中的输入样本靠得越紧，不同聚类中心的距离越远，则聚类结果的合理性越好。聚类的合理性划分就是使函数 $F$ 越小越好。

下面阐述优选聚类算法。

最邻近聚类法的基本思路是，预先给定一个聚类半径 $r$，置初始聚类个数 $m = 1$，取第 1 个输入样本 $\mathrm{X}_1$ 为聚类中心 $\mathrm{C}_1$，即 $\mathrm{C}_1 = \mathrm{X}_1$；计算 $\mathrm{C}_1$ 与 $\mathrm{X}_p (p = 2, 3, \cdots, N)$ 的欧氏范数 $d_{1p}$；若 $d_{1i}$ 为最小者，即 $d_{1i} = d_{\min}$，则作如下判断：若 $d_{\min} < r$，则 $\mathrm{X}_i$ 归入第 1 类，否则另建新类，$m = m + 1$，$\mathrm{C}_2 = \mathrm{X}_i$，$\cdots$。如此进行，聚类个数逐渐增多，直到已归类的样本个数 $S_m = N$ 时结束。

半径 $r$ 对聚类结果的影响非常大。若用上述算法确定 RBFNN 的隐含层节点个数，则 $r$ 的选取值将对网络的分类能力和泛化能力产生显著影响。用人为凑试的方法选取 $r$，难度较大，而且不总能保证聚类的合理性。

本文的优选聚类算法的基本思路是，在上述最邻近聚类法中引入优化策略，用一维寻优方法优选半径 $r$，使式(2)中函数 $F$ 达到最小。该算法由半径的优选算法(OA)和聚类算法(CA)两部分组成。CA 通过聚类确定类数 $m$ 和每类的中心 $\mathrm{C}_i$，$i = 1, 2, \cdots, m$，然后用式(2)计算函数 $F$ 的值。OA 为主算法，每次迭代均调用 CA。OA 的基本思路是，先用步长加速法寻求 $r$ 的最优值 $r_{op}$ 所在区间，然后用二次多项插值法求取 $r_{op}$。

为阐述方便，引入如下记号：用 $\mathrm{kind}(i, j)$ 存放第 $i$ 类第 $j$ 个成员对应的样本序号，用 $S_e(i)$ 存放第 $i$ 类中样本的个数；分别用 $\mathrm{X}(k)$ 和 $\mathrm{C}(i)$ 记 $\mathrm{X}_k$ 和 $\mathrm{C}_i$；采用如下两个标志数组

$$\begin{cases} \mathrm{flag}(k) = \begin{cases} 1, & \text{样本 } \mathrm{X}_k \text{ 已归类} \\ 0, & \text{否则} \end{cases} \\ \mathrm{flag}_1(i) = \begin{cases} 1, & \text{第 } i \text{ 类归类已满} \\ 0, & \text{否则} \end{cases} \end{cases} \quad (3)$$

**算法 1　聚类算法(CA)**

Step 1: 由主算法 OA 传来半径的优选值 $r$，令 $S_m = 0, \mathrm{flag}(i) = 0, \mathrm{flag}_1(i) = 0, i = 1, 2, \cdots, N$。

Step 2: $m = 1, m_0 = m, \mathrm{C}(1) = \mathrm{X}(1), \mathrm{flag}(1) = 1, \mathrm{kind}(1, 1) = 1, S_e(1) = 1, S_m = S_m + 1$。

Step 3: 对 $i = 1 \sim m_0$，若 $\mathrm{flag}1(i) = 0$ 则进行：

(1) $\min = M$ (较大正实数)；对 $k = 2 \sim N$，若 $\mathrm{flag}(k) = 0$ 则进行 1)，2) 和(2)：

1) 计算 $d = \|\mathrm{X}(k) - \mathrm{C}(i)\|$；

2) 若 $d < \min$，则 $\min = d, k_{op} = k, i_{op} = i$。

(2) 若 $\min \leq r$，则(新元归类)进行：

1) $S_e(i_{op}) = S_e(i_{op}) + 1, v = S_e(i_{op})$；$\mathrm{kind}(i_{op}, v) = k_{op}, \mathrm{flag}(k_{op}) = 1, S_m = S_m + 1$；

2) $\mathrm{C}(i_{op}) = \mathrm{C}(i_{op}) + \mathrm{X}(k_{op})$。

若 $\min > r$，则(另建新类)进行：

1) $m = m + 1, \mathrm{C}(m) = \mathrm{X}(k_{op})$；

2) $S_e(m) = 1, \mathrm{flag}(k_{op}) = 1, \mathrm{kind}(m, 1) = k_{op}, \mathrm{flag}_1(i_{op}) = 1, S_m = S_m + 1$。

Step 4: $m_0 = m$，若 $S_m < N$ 则转 Step 3；否则转 Step 5。

Step 5: 计算 $\mathrm{C}(i) = \mathrm{C}(i)/S_e(i), i = 1, 2, \cdots, m$。

Step 6: 用式(2)求函数 $F$ 的值。

**算法 2　优选算法(OA)**

算法 2 将选取的半径值传给 CA，由 CA 求出函数 $F$ 的值记为 $f(i)$。算法步骤如下：

Step 1: 置式(2)函数 $F$ 的阈值 $\epsilon > 0$ 和搜索的初始步长 $h = h_0$，令半径的初值为 $r(1)$，取迭代步数初值 $p = 0$。

Step 2: $r = r(1)$，由 CA 求 $f(1)$；$r(2) = r(1) +$

$h, r = r(2)$, 求 $f(2)$。

Step 3: 若 $f(1) < f(2)$, 则 $h = -h$, 分别对调 $r(1)$ 与 $r(2)$ 及 $f(1)$ 与 $f(2)$ 的值。

Step 4: 步长加速 $h = h + h$, $r(3) = r(2) + h$, $r = r(3)$, 由 CA 求 $f(3)$; $r(i-1) = r(i)$, $f(i-1) = f(i)$, $i = 1, 2, 3$。

Step 5: 若 $f(1) > f(2)$ 则转 Step 4; 否则转 Step 6 ($r_{op}$ 所在区间为 $(r(0), r(2))$, 而 $r(1)$ 为 $r(i)$ 中的最小点, $i = 0, 1, 2$)。

Step 6: $p = p + 1$, 用二次多项式插值法迭代求取 $r_{op}$。

Step 7: 若 $f(r_{op}) \leq \epsilon$, 则 $r = r_{op}$, 结束; 否则转 Step 6。

当用上述优选聚类算法确定了 RBFNN 的隐含层节点个数 $m$ 和中心 $C(i)$ ($i = 1, 2, \cdots, m$) 后, 可用

$$\begin{cases} \sigma_i = \left[\left[\sum_{v=1}^{S_e(i)} \| X(v) - C(i) \|^2\right] / S_e(i)\right]^{1/2} \\ v = \mathrm{kind}(i, \mu), \quad i = 1, 2, \cdots, S_e(i) \end{cases} \tag{4}$$

**注 1**　若 $S_e(i) = 1$, 即第 $i$ 类仅有 1 个样本, 且 $C(i) = X(v)$, $v = \mathrm{kind}(i, 1)$, 则由式 (4) 知 $\sigma_i = 0$。这将使计算 RBF 时出现以 0 作除数的情况。解决这个问题的方法是取消此类, 因为此类中样本不但与辨识无关, 而且它有可能使网络性能恶化[4, 5]。

优选函数 $F$ 的阈值 $\epsilon$ 的多目标决策方法。$F$ 的阈值 $\epsilon$ 由如下极小化公式确定

$$\epsilon = \alpha[\lambda_1 J_1 + \lambda_2 J_2] = \alpha\left[\sum_{k=1}^{N} [d_k - y_k]^2 + (1-\lambda)\sum_{k=1}^{N} [d_k' - z_k]^2\right] \tag{5}$$

其中 $J_1$ 和 $J_2$ 分别为网络训练和泛化验证的误差平方和。这里采用平方和加权法[6], $\lambda$ 为加权系数, $\alpha$ 为伸缩系数。

## 3　RBFNN 参数和权值的辨识

训练样本为 $\{X_k, d_k\}$, $k = 1, 2, \cdots, N$, 其中 $d_k$ 为系统实际输出。设式 (1) 的网络输出为 $y_k$, $e_k = d_k - y_k$ 为模型误差。于是由式 (1) 知

$$d_k = y_k + e_k = G_k^{\mathrm{T}} W + e_k \tag{6}$$

其中

$$\begin{cases} W = (w_0, w_1, \cdots, w_m)^{\mathrm{T}} \\ G_k = (1, g_{k1}, \cdots, g_{km})^{\mathrm{T}} \\ g_{ki} = \exp\left(-\dfrac{\| X_k - C_i \|^2}{2\sigma_i^2}\right) \end{cases} \tag{7}$$

引入指标函数

$$J = \frac{1}{2N}\sum_{k=1}^{N} e_k^2 = \frac{1}{2N}\sum_{k=1}^{N} [d_k - G_k^{\mathrm{T}} W]^2 \tag{8}$$

应用如下新型二阶学习算法[1, 2], 估计 RBFNN 中参数 $C(i)$ ($i = 1, 2, \cdots, m$) 和权值 $W$

$$C_k(i) = C_{k-1}(i) + \Delta C_{k-1}(i) + M_k[\beta_{ki}^1 \epsilon_{ki}^1 - X_k^{\mathrm{T}} \Delta C_{k-1}(i)]$$
$$i = 1, 2, \cdots, m \tag{9}$$

其中

$$\begin{cases} \beta_{ki}^1 = 1/[(g_{ki}')^2 - g_{ki}'' e_k] \\ \epsilon_{ki}^1 = g_{ki}' \epsilon_k^0 w_k(i), \quad i = 1, 2, \cdots, m \end{cases} \tag{10}$$

$$M_k = P_{k-1} X_k / [\beta_{ki}^1 + X_k^{\mathrm{T}} P_{k-1} X_k] \tag{11}$$

$$P_k = P_{k-1} - M_k X_k^{\mathrm{T}} P_{k-1} \tag{12}$$

$$W_k = W_{k-1} + \Delta W_{k-1} + N_k[\beta_k^0 \epsilon_k^0 - G_k^{\mathrm{T}} \Delta W_{k-1}] \tag{13}$$

$$\beta_k^0 = g_{ki}^{-2}, \quad \epsilon_k^0 = g_{ki} e_k \tag{14}$$

$$N_k = D_{k-1} G_k / [\beta_k^0 + G_k^{\mathrm{T}} D_{k-1} G_k] \tag{15}$$

$$D_k = D_{k-1} - N_k G_k^{\mathrm{T}} D_{k-1} \tag{16}$$

其中: 式 (9) ～ (12) 和式 (13) ～ (16) 分别为辨识 $C(i)$ 和 $W$ 的二阶学习算法; $w_k(i)$ 是 $W_k$ 的第 $i$ 个分量。文献 [1, 2] 已证明上述二阶学习算法有二阶收敛速度, 其计算量几乎与递推最小二乘法相当。文献 [7] 已证明上述新型二阶学习算法具有全局收敛性。

## 4　应用实例

本文带优选聚类算法的 RBFNN 在油田火山岩厚度模型化问题中得到了成功的应用。以油田提供的某区块多口油井岩芯数据为训练样本, 对火山岩系统进行建模。其中系统希望 (实际) 输出 $d_k$ 为沙岩厚度, 对 $d_k$ 有影响的 29 种因素的数据作为输入样本。由于量纲问题, 这些数据的数量级相差非常悬殊。首先用极差标准化方法, 将它们转化为无量纲数据, 均规范到 (0, 1] 区间。通过对输入样本数据的分析发现, 一些数据相关性较大, 还有一些数据明显与辨识无关, 可以剔除。

基于上述训练数据, 采用通常的 RBFNN 建模, 无论辨识精度还是泛化能力均不理想。应用本文提出的优选聚类算法, 能对输入样本进行有效地聚类, 应用多目标决策方法调整函数 $F$ 的阈值 $\epsilon$, 能使网络的分类能力和泛化能力均得到保证。采用本文的混合算法辨识网络的结构、RBFNN 的参数和权向量, 能使网络具有较高的辨识精度。

图 1 给出了上述火山岩厚度的辨识结果, 模型的平均相对误差在 1% 以内。表 1 给出了用同一区

块另一组数据的验证结果。由表 1 可见, 该 RBFNN 具有较好的泛化能力。

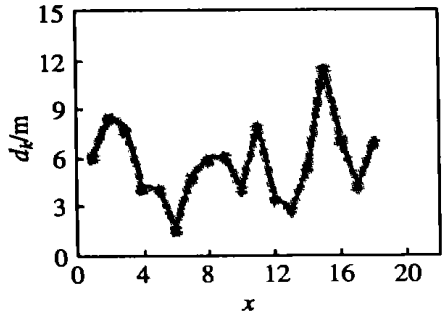


图 1 油藏系统沙岩厚度的建模结果

表 1 NN 建模的泛化验证结果

| 井名 | 实际值 | 计算值 | 相对误差 |
|---|---|---|---|
| 0 | 6.138 | 5.732 2 | 0.066 1 |
| 1 | 8.514 | 8.039 2 | 0.055 8 |
| 2 | 7.920 | 7.700 2 | 0.027 7 |
| 3 | 4.257 | 3.980 2 | 0.065 0 |
| 4 | 4.158 | 4.002 3 | 0.036 6 |
| 5 | 1.782 | 1.736 0 | 0.010 8 |
| 6 | 4.950 | 4.723 1 | 0.045 8 |
| 7 | 5.940 | 5.793 3 | 0.029 6 |
| 8 | 6.138 | 6.118 6 | 0.003 2 |
| 9 | 4.158 | 3.987 3 | 0.041 0 |
| 10 | 7.920 | 7.881 2 | 0.004 9 |
| 11 | 3.564 | 3.465 1 | 0.027 7 |
| 12 | 2.772 | 2.801 4 | - 0.010 6 |
| 13 | 5.346 | 5.291 3 | 0.010 2 |
| 14 | 11.484 | 10.832 0 | 0.056 8 |
| 15 | 7.128 | 6.898 1 | 0.032 2 |
| 16 | 4.356 | 4.431 2 | - 0.017 3 |
| 17 | 6.930 | 6.745 6 | 0.026 6 |

## 5 结 语

采用本文给出的控制聚类合理性函数 $F$ 和新型优选聚类算法, 对输入样本的聚类过程进行优化, 能有效地确定 RBFNN 隐含层节点的个数, 同时计算出 RBF 参数的初值。这为进一步精确辨识 RBFNN 参数和权向量提供了可靠保证。用多目标决策方法调整函数 $F$ 的阈值, 能保证网络有一定的分类能力和泛化能力。应用文献[1, 2]提出的新型二阶学习算法, 有效地提高了网络的辨识精度。油田火山岩系统辨识的应用表明了本文方案的可行性和有效性。

参考文献(References):

[1] 刘铁男, 段玉波, 于镝, 等. 神经网络的新型二阶学习算法及其应用[J]. 控制与决策, 2001, 16(5): 627-629.
(Liu T N, Duan Y B, Yu D, et al. New second-order learning algorithm for neural networks and its application[J]. *Control and Decision*, 2001, 16(5): 627-629.)

[2] 刘铁男, 段玉波, 陈广义. 多层前向神经网络的新型二阶学习算法[J]. 控制理论与应用, 2000, 17(5): 721-724.
(Liu T N, Duan Y B, Chen G Y. New second order learning algorithm of multilayer feedforward neural network[J]. *Control Theory & Application*, 2000, 17(5): 721-724.)

[3] 马勇, 杨煜普, 许晓鸣, 等. 自组织 RBF 神经网络对驾驶员主动安全性因素的辨识[J]. 控制与决策, 2001, 16(1): 114-116.
(Ma Y, Yang Y P, Xu X M, et al. Identification of the driver's active safety factors using self-organizing RBF neural network[J]. *Control and Decision*, 2001, 16(1): 114-116.)

[4] 刘妹琴, 沈轶, 廖晓昕. 一类新的 RBF 神经网络在非线性系统建模中的应用[J]. 控制与决策, 2001, 16(3): 277-281.
(Liu M Q, Shen Y, Liao X X. Application of a class of new RBF neural networks to modelling nonlinear systems[J]. *Control and Decision*, 2001, 16(3): 277-281.)

[5] Billings S A, Jamaluddin H B, Chens. Properties of neural networks with application to modelling dynamic systems[J]. *Int J of Control*, 1992, 55(1): 193-224.

[6] 钱颂迪. 运筹学[M]. 北京: 清华大学出版社, 1997.

[7] 刘铁男, 王利国, 刘严崴. 神经网络二阶反向传播学习算法及其收敛性[J]. 大庆石油学院学报, 2001, 25(4): 38-41.
(Liu T N, Wang L G, Liu Y W. Convergence analysis of second order back-propagation learning algorithm for neural network[J]. *J of Daqing Petroleum Institute*, 2001, 25(4): 38-41.)