

文章编号: 1000-8152(2008)04-0655-06

径向基函数神经网络的一种两级学习方法

陈俊风^{1,2}, 任子武³, 伞冶³

(1. 浙江大学 控制科学与工程学系, 浙江 杭州 310027; 2. 河海大学 计算机及信息工程学院, 江苏 常州 213022;
3. 哈尔滨工业大学 控制与仿真中心, 黑龙江 哈尔滨 150001)

摘要: 建立RBF(radial basis function)神经网络模型关键在于确定网络隐中心向量、基宽度参数和隐节点数。为设计结构简单, 且具有良好泛化性能径向基网络结构, 本文提出了一种RBF网络的两级学习新设计方法。该方法在低级由正则化正交最小二乘法与D-最优试验设计结合算法自动构建结构节俭的RBF网络模型; 在上级通过粒子群优化算法优选结合算法中影响网络泛化性能的3个学习参数, 即基宽度参数、正则化系数和D-最优代价系数的最佳参数组合。仿真实例表明了该方法的有效性。

关键词: 径向基网络; 两级学习; 建模; 泛化能力
中图分类号: TP183 **文献标识码:** A

A two-level learning hierarchy for the radial basis function networks

CHEN Jun-feng^{1,2}, REN Zi-wu³, SAN Ye³

(1. Department of Control Science and Engineering, Zhejiang University, Hangzhou Zhejiang 310027, China;
2. College of Computer & Information Engineering, Hohai University, Changzhou Jiangsu 213022, China;
3. Control & Simulation Centre, Harbin institute of Technology, Harbin Heilongjiang 150001, China)

Abstract: The key to construct a radial basis function(RBF) network is to select reasonable hidden center vectors, RBF width and hidden node number. In order to design a RBF network with parsimonious structure and good generalization, a new two-level learning hierarchy for designing RBF networks is proposed. At the lower level in this method, a parsimonious RBF model is constructed by an integrated algorithm (ROLS+D-opt) which combines regularized orthogonal least squares (ROLS) with D-optimality experimental design (D-opt). At the upper level, particle swarm optimization (PSO) is used to search the optimal combination of three important learning parameters, i.e., the RBF width, the regularized parameter and D-optimality weight parameter, which influence the network's generalization ability. Simulation results show the effectiveness of the proposed method.

Key words: radial basis function networks; two-level learning hierarchy; modeling; generalization ability

1 引言(Introduction)

作为一种单隐层前馈神经网络, 径向基函数(radial basis function, RBF)神经网络具有网络结构简单、非线性逼近能力强, 学习速度快等优点, 因此已成功应用于时序预测^[1]、函数逼近^[2]、系统建模^[3]和控制等领域。

RBF网络的性能严重依赖于网络隐中心的数量和位置, 确定RBF基函数的个数、中心向量和它的形状是RBF网络设计的核心问题。目前设计RBF网络的学习算法有 k -均值聚类法、Kohonen聚类法^[4]及梯度下降法等, 但聚类方法都是一种启发式过程, 它对初始值的选取具有很大的敏感性, 在确定网络数据中心时只用到样本的输入信息, 没有用到样本的输出信息^[1,3]; 梯度下降法设计RBF网络则忽视

了RBF网络所固有的特点, 学习速度慢, 易陷入局部极小点, 初始权值不同, 学习后网络泛化性能也不同。

在RBF网络学习算法中, 正交最小二乘法(orthogonal least squares, OLS)是一种常见的设计RBF网络方法, 它将RBF中心的选择问题归结为线性回归中子模型的选择问题, 这种算法能自动地避免网络规模过大和随机选择中心带来的数值病态问题^[5]。虽然OLS算法能够设计出较小结构的RBF网, 但如果训练样本中包含有较强的噪声数据时, 用它构建的RBF网络会拟合样本中的噪声数据, 并可能产生过拟合现象, 影响到所建网络的泛化性能。正则化(regularization)是一种改善神经网络泛化能力的有效方法之一, 文献^[6]提出了正则化

正交最小二乘法(regularized orthogonal least squares, ROLS)设计RBF网络方法,通过贝叶斯(Bayesian)方法^[7]产生正则化系数,避免网络过分地拟合样本中的噪声数据,从而改善了网络泛化性能.在此基础上,文献[8]又将ROLS算法与D-最优试验设计法(D-optimality experimental design, D-opt)结合,通过在代价函数中引入D-最优成本项,在能够自动设计结构节俭RBF网络的同时,又能进一步提高所选择子模型的性能和参数鲁棒性.该方法在实际中已得到较多的应用,但它还存在有以下问题:

1) 算法为改善网络泛化性能,通过正则化技术避免网络过度拟合样本中的噪声数据,但算法中的正则化系数通过Bayesian方法迭代学习产生,这种通过梯度方法多次迭代产生的正则化系数只是在当前特定条件下的局部最优,并不是全局最佳^[9];

2) RBF网络的基宽度值是网络设计的重要参数,如果基宽度取值太大,网络对样本中的局部信息无法辨识;相反,基宽度值过小,则对样本的整体信息可能丢失^[10].因此该参数取值不当,将会严重影响RBF网络的泛化性能;

3) 另外,结合算法在代价函数中引入D-最优成本项,改善了模型性能和鲁棒性,但引入的D-最优成本项作用于回归子模型选择的整个过程,D-最优权系数取值不同将生成不同的网络隐中心向量及节点数目,从而最终影响网络泛化性能.

针对上述问题,必须寻求一种具有全局搜索能力的优化算法优选一组最佳参数的组合.为此本文提出了一种设计RBF网络的两级学习方法,该算法的主要特点是在下级由ROLS与D-opt结合算法自动设计结构节俭RBF网络,而影响RBF网络泛化性能的三个学习参数—正则化系数、基宽度参数与D-最优权系数由上级粒子群优化算法(particle swarm optimization, PSO)全局优选一组最佳的参数组合,从而使最终设计的RBF网络不但具有节俭结构,而且具有良好的泛化性能.

2 RBF网络的结构描述(Structure formulation of RBF networks)

考虑一多输入单输出3层(m 个输入、 n 个隐节点、1个输出)RBF网络设计,采用Gaussian型径向基函数,并假设径向基函数的基宽度参数取相同的固定值,此时RBF网络模型输出为

$$\hat{y}(x) = \sum_{i=1}^n \theta_i \exp(-\|\mathbf{x} - \mathbf{c}_i\|^2 / 2\rho^2). \quad (1)$$

式中: $\mathbf{x} = [x_1, x_2, \dots, x_m]^T$ 为网络的输入矢量, θ_i 为第 i 个隐节点至输出的连接权值, $\mathbf{c}_i = [c_{1,i}, c_{2,i}, \dots, c_{m,i}]^T$ 为隐层第 i 个径向基函数的数据中心向量, ρ 为函数的基宽度参数, $\|\cdot\|$ 为欧几里

德(Euclidean)范数.

径向基网络的输出层为线性神经元,网络隐中心向量 \mathbf{c}_i 及基宽度参数 ρ 确定后,采用简单的线性学习算法便可求得网络的输出权值 θ_i .因此建立RBF网络模型的关键在于确定参数隐中心向量、基宽度参数及节点数,即寻找参数 $\Xi = \{\mathbf{c}_i, \rho, n\}$

3 RBF网络两级学习设计方法(A two-level learning method of RBF networks)

3.1 正则化正交最小二乘法(ROLS)

假定共有 N 个训练样本, $\{y(k), \mathbf{x}(k)\}_{k=1}^N$ 为其训练样本集, $\mathbf{x}(k)$ 为网络的样本输入, $y(k)$ 为样本输入 $\mathbf{x}(k)$ 所对应的期望网络输出.设网络隐中心从样本输入数据中选取,每个样本输入都可被认为是网络的候选隐中心,即 $\mathbf{c}_i = \mathbf{x}(i)$ ($1 \leq i \leq N$),当给定固定的基宽度参数 ρ 时,定义

$$\phi_i(k) = \exp(-\|\mathbf{x}(k) - \mathbf{c}_i\|^2 / 2\rho^2), \quad (2)$$

则期望的网络输出可表示为

$$\begin{cases} y(k) = \hat{y}(k) + e(k) = \sum_{i=1}^N \theta_i \phi_i(k) + e(k), \\ 1 \leq k \leq N. \end{cases} \quad (3)$$

式中 $e(k)$ 表示网络实际输出 $\hat{y}(k)$ 与期望输出 $y(k)$ 的偏差.将上式写成矩阵形式

$$\mathbf{y} = \Phi \Theta + \mathbf{e}. \quad (4)$$

式中:

$$\begin{aligned} \mathbf{y} &= [y(1), \dots, y(N)]^T, \quad \Theta = [\theta_1, \dots, \theta_N]^T, \\ \mathbf{e} &= [e(1), \dots, e(N)]^T, \quad \Phi = [\Phi_1, \dots, \Phi_N], \\ \Phi_i &= [\phi_i(1), \dots, \phi_i(N)]^T. \end{aligned}$$

其中: \mathbf{y} 为输出向量, Θ 为权参数向量, \mathbf{e} 为残差向量, Φ 为回归矩阵, Φ_i 为回归因子.

使用Gram-Schmidt法则正交分解回归矩阵 Φ :

$$\Phi = \mathbf{W} \mathbf{A}. \quad (5)$$

式中: \mathbf{A} 为一 $N \times N$ 三角矩阵,它的主对角线元素为1,对角线下方为零; $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_N]$ 为正交矩阵,并满足

$$\mathbf{W}^T \mathbf{W} = \text{diag} \{\kappa_1, \dots, \kappa_N\}, \quad (6)$$

$$\kappa_i = \mathbf{w}_i^T \mathbf{w}_i, \quad i = 1, \dots, N. \quad (7)$$

于是回归模型(4)可表达成

$$\mathbf{y} = (\Phi \mathbf{A}^{-1})(\mathbf{A} \Theta) + \mathbf{e} = \mathbf{W} \mathbf{g} + \mathbf{e}. \quad (8)$$

式中正交权参数向量 \mathbf{g} 满足

$$\mathbf{g} = [g_1, \dots, g_N]^T = \mathbf{A} \Theta. \quad (9)$$

ROLS算法^[6]采用如下误差准则函数:

$$J_R(\mathbf{g}, \lambda) = \mathbf{e}^T \mathbf{e} + \lambda \mathbf{g}^T \mathbf{g}. \quad (10)$$

上式第2项 $\lambda \mathbf{g}^T \mathbf{g}$ 为正则化项, 其中 $\lambda \geq 0$ 为正则化系数. 由 $\partial J_R / \partial \mathbf{g} = 0$, 并经一些简单计算可得

$$\mathbf{e}^T \mathbf{e} + \lambda \mathbf{g}^T \mathbf{g} = \mathbf{y}^T \mathbf{y} - \sum_{i=1}^N (\mathbf{w}_i^T \mathbf{w}_i + \lambda) g_i^2, \quad (11)$$

因此正则化后的误差下降率可定义为

$$[rerr]_i = (\mathbf{w}_i^T \mathbf{w}_i + \lambda) g_i^2 / \mathbf{y}^T \mathbf{y}. \quad (12)$$

在回归选择过程中选择对网络输出贡献最大的回归因子, 当满足下面条件时, 选择结束

$$1 - \sum_{i=1}^{n_s} [rerr]_i < \xi. \quad (13)$$

式中: $0 < \xi < 1$ 为给定的容许误差, n_s 为最终选出的回归因子个数.

3.2 ROLS与D-最优试验设计方法的结合(An integrated algorithm of ROLS and D-opt)

在试验设计中, 矩阵 $\Phi^T \Phi$ 被称为设计矩阵. 权参数向量 Θ 的最小二乘估计量 $\hat{\Theta}$ 为

$$\hat{\Theta} = (\Phi^T \Phi)^{-1} \Phi^T \mathbf{y}. \quad (14)$$

当设计矩阵趋于病态时, 不能得到满意的最小二乘估计模型. 设计矩阵条件数太大, 将导致参数估计量 $\hat{\Theta}$ 不稳定. D最优试验设计准则就是使所选择的子模型设计矩阵行列式最大化, 即

$$\max \{ J_D = \det(\Phi_k^T \Phi_k) = \prod_{k=1}^{n_s} \lambda_k \}. \quad (15)$$

式中: $\Phi_k \in \mathbb{R}^{N \times n_s}$ 为从矩阵 Φ 中选出的 n_s 个回归因子组成, λ_k 为矩阵 $\Phi_k^T \Phi_k$ 特征值. 该准则有利于减小设计矩阵条件数, 提高模型性能和参数鲁棒性^[2].

由 $\Phi = \mathbf{W} \mathbf{A}$, 且矩阵 \mathbf{A} 是主对角线元素为1上三角矩阵, 容易可得 $\det(\Phi_k^T \Phi_k)$ 行列式值最大化等价于 $\det(\mathbf{W}_k^T \mathbf{W}_k)$ 行列式值最大化, 亦等价于 $\min \{ -\log \det(\mathbf{W}_k^T \mathbf{W}_k) \}$. 根据

$$\det(\Phi^T \Phi) = \det(\mathbf{W}^T \mathbf{W}) = \prod_{i=1}^N \mathbf{w}_i^T \mathbf{w}_i, \quad (16)$$

$$-\log(\det(\mathbf{W}^T \mathbf{W})) = \sum_{i=1}^N -\log(\mathbf{w}_i^T \mathbf{w}_i). \quad (17)$$

将ROLS算法同D-opt相结合, 结合后算法代价函数为^[8]

$$J_{CR}(\mathbf{g}, \lambda, \beta) = J_R(\mathbf{g}, \lambda) + \beta \sum_{i=1}^N -\log(\mathbf{w}_i^T \mathbf{w}_i). \quad (18)$$

上式第2项为D-最优成本项, 其中D-最优权系数 $\beta > 0$, 结合后算法新的误差下降率为

$$[crrerr]_i = ((\mathbf{w}_i^T \mathbf{w}_i + \lambda) g_i^2 + \beta \log(\mathbf{w}_i^T \mathbf{w}_i)) / \mathbf{y}^T \mathbf{y}. \quad (19)$$

如同ROLS算法, 在每一步选择过程中选择对输出贡献最大的回归因子, 当满足

$$[crrerr]_l \leq 0, n_s + 1 \leq l \leq N \quad (20)$$

条件时, 算法选择过程结束.

结合算法其中的一个优点是算法避免了容许误差 ξ 参数的设置, 可以自动终止算法回归因子的选择过程. 该算法中正则化系数 λ 采用贝叶斯方法估计学习产生, 但这种通过多次迭代梯度方法获取的 λ 值只是当前条件下找到的一局部最优解, 并不是全局最优^[9]; 另一方面基宽度参数 ρ 的选择与网络泛化能力有着直接的联系, 同时不同的D-最优权系数 β 取值也将使回归过程选择不同的回归因子, 影响所设计网络的性能. 事实上, 网络的泛化性能与上3个参数间形成了一极度复杂的超维多峰曲面空间, 必须利用一种具有全局搜索能力的优化算法优选一组全局最佳的 λ , ρ 和 β 参数组合.

3.3 基于PSO的RBF网络的两级学习方法(Two-level learning method of RBF networks based on PSO)

粒子群优化方法(PSO)^[11]是一种基于群体智能理论的演化优化方法. 算法将求解问题表示为“粒子”, 并生成“粒子”种群, 通过粒子间的竞争与协作, 实现多维空间中最优解的搜索. PSO算法的机理、基本定义和算法实现可参考文献^[11]. 另外文献^[12]实验表明, PSO算法中惯性权系数 ω 如果随算法迭代进行而线性减少, 能改善算法的收敛性能; 为减少在优化过程中粒子飞离搜索空间的可能性, 通常将粒子速度 $v_{i,d}^{(t)}$ 限制于 $[-v_d^{\max}, v_d^{\max}]$ 之间, v_d^{\max} 决定了粒子飞行的最大距离, 其中

$$v_d^{\max} = k \cdot x_d^{\max}, 0.1 < k < 1.0. \quad (21)$$

式中 x_d^{\max} 意指搜索空间第 d 维位置的上界.

3.3.1 两级学习方法的基本思想(Basic ideas of two-level learning method)

采用如图1所示RBF网络两级学习方法结构图. 基本思想为在下级由ROLS与D-最优试验结合算法(ROLS+D-opt)自动设计粒子所对应的RBF网络, 利用上级PSO在可行解中找出一组最优的 λ , ρ 和 β 参数组合, 使定义的适配函数值最小.

如图1所示, 采用实数编码制的PSO算法在解空间内产生 p 个粒子, 每一个粒子通过所定义的适应度函数计算其适应度值 f_i , 更新每一粒子的速度和位置, 产生新种群, 多次迭代后, 最终代的最佳粒子便是该优化问题的最优解.

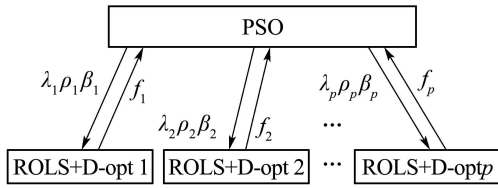


图1 RBF网络两级学习方法结构图

Fig. 1 Architecture of the two-level learning method of the RBF networks

3.3.2 适配函数的选取(The selection of fitness function)

将数据集分为训练样本集和测试样本集, 在给定粒子 $[\lambda_i, \rho_i, \beta_i]$ 值下, 利用训练集通过ROLS与D-opt结合算法(ROLS+D-opt)自动设计节俭结构的RBF网络; 利用测试集计算其期望输出与实际网络输出之间的均方误差(mean square error, MSE), 将此均方误差作为该粒子适应度函数值。

设 $\hat{y}(x(k))$ 为对测试样本的实际网络输出, $y(x(k))$ 为测试样本的期望输出, n_c 为测试样本数, 则适配函数可为

$$\min f(K) = \frac{1}{n_c} \sum_{k=1}^{n_c} (y(x(k)) - \hat{y}(x(k)))^2. \quad (22)$$

式中 K 表示粒子 $[\lambda_i, \rho_i, \beta_i]$, 适配函数值越小, 所设计的RBF网络的泛化性能越好。

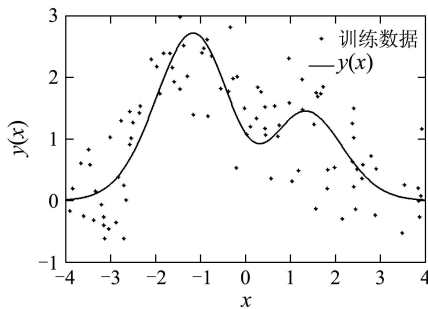
4 仿真实例(Simulation)

4.1 静态函数逼近(Static function approximation)

考虑用高斯型RBF网络逼近静态函数^[3]

$$y(x) = 1.1(1 - x + 2x^2) \exp(-x^2/2). \quad (23)$$

训练样本的产生方式如下: 样本数 $N=100$, 训练样本输入 x_i 服从区间 $[-4,4]$ 内的均匀分布, 样本输出为 $y(x_i) + e$, 其中 e 为添加的高斯噪声, 其服从均值为0, 方差为0.5正态分布; 以区间 $[-4,4]$ 内等间距产生100个无噪声输出值 $y(x)$ 作为测试数据集, 间距 $\Delta = 0.08$. 图2显示了训练样本集及函数曲线。

图2 训练样本点及函数 $y(x)$ 曲线图Fig. 2 Training samples and curve of the function $y(x)$

分别用正交最小二乘法与D-最优试验设计结合算法(OLS+D-opt)和ROLS+D-opt学习RBF网络, 两算法的网络基宽度参数均按文献[3]选取 $\rho = 0.866$

(扩展常数 $2\rho^2 = 1.5$), ROLS+D-opt算法中的正则化系数 λ 采用贝叶斯方法学习产生, 表1给出D-opt权系数 β 取不同值时两算法设计RBF网络的比较结果. 从表1中可看出:

1) OLS+D-opt算法中, D-opt权系数 β 的取值对所设计RBF网络的性能较为敏感, 取不同 β 值, 所设计RBF网络隐节点数目和性能差异较大; 相比较而言, 权系数 β 的取值对ROLS+D-opt算法设计RBF网络的性能影响较小, 但不同的 β 值, 设计的RBF网络中心数也不同, 即使隐节点数目相同(如当 $\beta = 10^0$ 和 $\beta = 10^{-1}$ 时, ROLS+D-opt算法设计网络隐节点数都为8), 网络对训练集和测试集的误差也可能不一样, 说明D-最优权系数取值不同将会生成不同的网络隐中心向量及节点数目, D-最优成本项作用于回归子模型选择的整个过程, 影响网络泛化性能。

2) OLS+D-opt算法没采用正则化技术, 虽然它对含噪声训练样本误差较小, 但由于过分拟合了训练样本中的噪声数据, 因此对测试样本误差较大, 影响网络的性能; 而ROLS+D-opt算法采用了正则技术, 与OLS+D-opt方法所设计的网络性能比较, 其结构更为节俭, 对测试样本的误差小, 泛化性能优。

为设计结构节俭, 同时具有较优性能的RBF网络, 用PSO全局优选算法的正则化系数 λ 、基宽度参数 ρ 和D-最优权系数 β 最佳参数组合, 参数的搜索空间为 $\lambda \in [0, 1]$, $\rho \in [0.01, 1.5]$, $\beta \in [10^{-6}, 1]$, PSO算法中的群体粒子数为 $N=20$, v_d^{\max} 速度限定式(21)中系数 $k = 0.4$; 经过30代的进化后, 得 $[\lambda_{\text{opt}}, \rho_{\text{opt}}, \beta_{\text{opt}}] = [0.0842, 1.0115, 0.0127]$, 最佳粒子对应的适配度函数值 $f_{\text{opt}} = 0.0271$ 。

获得最佳的 λ , ρ 和 β 参数后, 用下级ROLS+D-opt结合算法自动设计RBF网络. 为验证本文方法设计RBF网络性能的有效性, 下面分别用OLS+D-opt算法、ROLS+D-opt算法及本文两级学习方法设计RBF网络, 比较3种方法设计RBF网络性能. OLS+D-opt和ROLS+D-opt算法中, 选取 $\rho = 0.866$, $\beta = 5 \times 10^{-4}$; ROLS+D-opt算法正则化系数由贝叶斯方法学习产生, 比较结果如表2所示. 从表2中可看出本文两级学习方法所设计的RBF网络对测试样本的误差最小, 而且隐节点数也最少, 该方法不但改善了网络的泛化能力, 而且精简了网络结构. 图3为采用本文两级学习方法设计RBF网络的函数逼近曲线, 其中“o”符号表示网络所选择的隐节点中心向量位置, 从图中也看出, 用本文方法所设计的RBF网络在训练样本包含有较强的噪声数据时, 同样具有较好的函数曲线逼近效果, 体现了本文方法设计RBF网络有效性。

表1 不同D-opt权系数值下网络逼近标量函数性能比较结果

Table 1 Comparison of different D-opt weight for scalar function modeling

D-opt权系数 β	对带噪声的训练样本误差MSE		对无噪声的测试样本误差MSE		隐节点中心数	
	OLS+D-opt	ROLS+D-opt	OLS+D-opt	ROLS+D-opt	OLS+D-opt	ROLS+D-opt
1×10^{-3}	0.2859	0.3230	0.1193	0.0348	19	12
1×10^{-2}	0.2906	0.3242	0.0971	0.0352	14	9
1×10^{-1}	0.3103	0.3220	0.0674	0.0365	9	8
1×10^0	0.3161	0.3213	0.0587	0.0418	8	8

表2 3种方法设计RBF网络性能比较结果

Table 2 Performance comparison of RBF network designed with 3 methods

	训练误差	测试误差	隐节点数
OLS+D-opt	0.2856	0.1154	19
ROLS+D-opt	0.3230	0.0348	12
本文方法	0.3287	0.0271	9

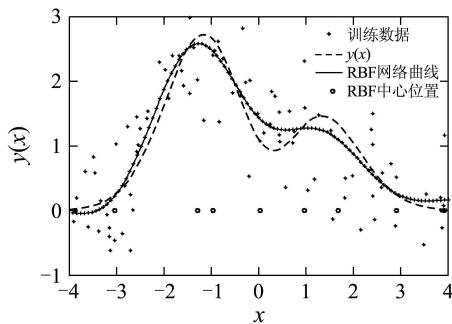


图3 本文两级学习方法设计RBF网络的函数逼近曲线

Fig. 3 The function approximation curve of RBF network designed with two-level learning method

另外在 $\lambda_{opt} = 0.0842$, $\rho_{opt} = 1.0115$ 相同参数取值下, 计算ROLS算法($\beta = 0$)在和本文两级学习

表3 不同D-opt权系数值下燃烧炉网络模型性能比较结果

Table 3 Comparison of different D-opt weight for fuel furnace modeling

D-opt权系数 β	对训练样本集误差MSE		对测试样本集误差MSE		隐节点中心数	
	OLS+D-opt	ROLS+D-opt	OLS+D-opt	ROLS+D-opt	OLS+D-opt	ROLS+D-opt
1×10^{-4}	*	0.0244	*	0.2127	51	18
1×10^{-3}	0.0129	0.0243	0.2773	0.2096	23	16
1×10^{-2}	0.0174	0.0253	0.2879	0.2231	15	15
1×10^{-1}	0.0218	0.1315	0.3246	0.4494	13	11

表3中“*”表示设计矩阵接近奇异, 条件数过大, 求解不可靠。

用PSO优选算法中 λ , ρ 和 β 三个参数, 参数 λ , ρ 和 β 的搜索区间分别为 $[1 \times 10^{-8}, 1]$, $[5, 15]$, $[1 \times 10^{-8}, 1]$, 群体粒子数为 $N=20$, v_d^{max} 速度限定式(21)中系数 $k = 0.4$; 经过20代的进化后, 得最佳粒子 $[\lambda_{opt}, \rho_{opt}, \beta_{opt}] = [1 \times 10^{-6}, 11.1560, 0.0248]$, 所对应的适配度函数值 $f_{opt} = 0.1803$ 。

方法($\beta_{opt} = 0.0127$)在构建相同隐节点数(节点数都为9)RBF网络时设计矩阵行列式 J_D 值, 得两算法 J_D 值分别为72.6817和480.8175, 说明在ROLS算法中引入D-最优试验设计后, 减小了回归子模型设计矩阵的条件数, 提高了模型的可靠性。

4.2 燃烧炉模型辨识(Identification of fuel furnace modeling)

文献[13]中, 根据实测的296组输入/输出数据, 建立起燃烧炉的动态数学模型, 其中输入 $u(k)$ 是煤气的速度(m^3/s), 输出 $y(k)$ 是炉膛出口排出煤气 CO_2 的比例, 选择前160组数据用于训练, 后136组数据用于模型验证。在仿真中, RBF网络的输入向量为 $[u(k-3), u(k-4), y(k-1), y(k-2)]^T$, 实际和期望输出变量分别为 $\hat{y}(k)$ 和 $y(k)$ 。

分别用OLS+D-opt算法和ROLS+D-opt算法学习RBF网络, 两算法的网络基宽度参数均取 $\rho = 10$, ROLS+D-opt算法中的正则化系数 λ 采用贝叶斯方法学习产生, 表3给出D-opt权系数 β 取不同值时两算法设计RBF网络的比较结果。从表中看出D-opt权系数 β 不同将使算法所设计的RBF网络隐中心数目不同, 网络性能也不一样。

为说明本文算法有效性, 表4显示了3种方法设计RBF网络的一些主要性能对比。在OLS+D-opt和ROLS+D-opt算法中, 参数选取 $\rho = 5.0$, $\beta = 1 \times 10^{-3}$; ROLS+D-opt算法正则化系数由贝叶斯方法学习产生。从表中比较看出, OLS+D-opt方法虽然对训练集的学习误差最小, 但它所设计的网络隐中心数最多, 且测试集误差最大; 与OLS+D-opt方法和ROLS+D-opt方法比较, 本文两级学习方

法设计的RBF网络对测试样本的误差最小,而且隐节点数也最少,具有更优的泛化能力.图4为采用本文两级学习方法设计RBF网络,网络模型预测曲线(虚线)与燃烧炉的实测输出(实线)的比较图,从图中可看出利用本文方法学习的RBF网络对燃烧炉模型具有较高的辨识精度.

表4 不同方法设计燃烧炉网络模型性能比较结果

Table 4 Performance comparison of fuel furnace modeling designed with different methods

	学习误差	测试误差	隐节点数
OLS+D-opt	0.0126	0.7027	37
ROLS+D-opt	0.0230	0.9963	26
本文方法	0.0209	0.1803	14

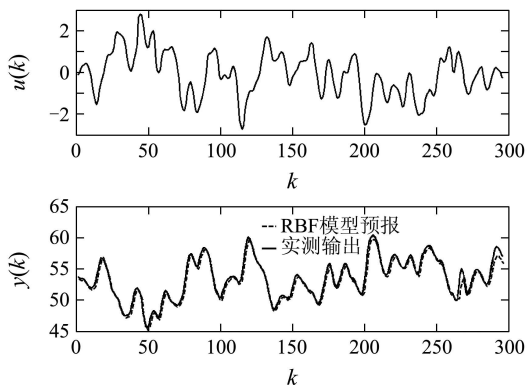


图4 燃烧炉实测输出(实线)与RBF预测输出(虚线)比较
Fig. 4 The curves of fuel furnace practical output (solid) and RBF network prediction (dashed)

5 结论(Conclusion)

本文提出了一种设计RBF网络的两级学习方法,该方法具有如下特点:

1) 由上级PSO方法寻优下级ROLS+D-opt算法中影响网络泛化性能的 λ , ρ 和 β 参数, 这为网络学习算法优选适宜的参数提供了一种有效途径.

2) 该方法下级ROLS+D-opt结合算法中 λ , ρ 和 β 3参数通过上级PSO全局寻优其参数的最佳组合, 因此所设计的RBF网络不仅结构简单, 而且也具有较好的泛化能力.

3) 由于PSO仅优化方法中的3个参数, 这与PSO直接优化RBF网络的拓扑结构与权参数方法相比较, 在计算量上该方法要有效.

该方法有以下问题待继续研究:

1) 该方法将网络的训练样本输入作为RBF隐节点候选中心集, 当网络的训练样本数据过于庞大时, 则网络学习的候选中心数目过大, 使得算法的学习效率变低, 计算量剧增, 因此当候选中心数目过多时, 如何利用聚类算法将候选中心缩小至

一有效范围, 并同本文算法有效结合是一个令人感兴趣的问题.

2) 在该方法中, 网络的径向基函数使用了相同的固定基宽度参数 ρ 值, 在网络的学习过程中如何对不同的隐中心选取不同的基宽度值, 并对其进行动态调整, 也是将来值得进一步研究.

参考文献(References):

- [1] 张志华, 郑南宁, 史罡. 径向基函数神经网络的软竞争学习算法[J]. 电子学报, 2002, 30(1): 132 - 135.
(ZHANG Zhihua, ZHENG Nanning, SHI Gang. Soft-competition learning algorithms of the radial basis function neural networks[J]. *Acta Electronica Sinica*, 2002, 30(1): 132 - 135.)
- [2] HONG X, HARRIS C J. Nonlinear model structure design and construction using orthogonal least squares and D-optimality design[J]. *IEEE Transactions on Neural Networks*, 2002, 13(5): 1245 - 1250.
- [3] 魏海坤, 丁维明, 宋文忠, 等. RBF网的动态设计方法[J]. 控制理论与应用, 2002, 19(5): 673 - 680.
(WEI Haikun, DING Weiming, SONG Wenzhong, et al. Dynamic method for designing RBF neural networks[J]. *Control Theory & Applications*, 2002, 19(5): 673 - 680.)
- [4] KOHONEN T. *Self-Organization and Associative Memory*[M]. Berlin: Springer-Verlag, 1984.
- [5] CHEN S, COWAN C F N, GRANT P M. Orthogonal least squares learning algorithm for radial basis function neural networks[J]. *IEEE Transactions on Neural Networks*, 1991, 2(2): 302 - 309.
- [6] CHEN S, CHNG E S, ALKADHIMI K. Regularized orthogonal least squares algorithm for constructing radial basis function networks[J]. *International Journal of Control*, 1996, 64(5): 829 - 837.
- [7] MACKAY D J C. Bayesian interpolation[J]. *Neural Computation*, 1992, 4(3): 415 - 447.
- [8] CHEN S, HONG X, HARRIS C J. Sparse kernel regression modeling using combined locally regularized orthogonal least squares and D-optimality experimental design[J]. *IEEE Transactions on Automatic Control*, 2003, 48(6): 1029 - 837.
- [9] CHEN S, WU Y, LUK B L. Combined genetic algorithm optimization and regularized orthogonal least squares learning for radial basis function networks[J]. *IEEE Transactions on Neural Networks*, 1999, 10(5): 1239 - 1243.
- [10] 赵温波, 杨鹭怡, 王立明. 径向基概率神经网络的混合结构优化算法[J]. 系统仿真学报, 2004, 16(10): 2175 - 2180.
(ZHAO Wenbo, YANG Luyi, WANG Liming. The hybrid structure optimization algorithms of radial basis probabilistic neural networks[J]. *Journal of System Simulation*, 2004, 16(10): 2175 - 2180.)
- [11] KENNEDY J, EBERHART R. Particle swarm optimization[C]// *Proceedings of IEEE International Conference on Neural Networks*. Perth: IEEE Press, 1995, 4: 1942 - 1948.
- [12] SHI Y, EBERHART R. A modified particle swarm optimizer[C]// *Proceedings of IEEE World Conference on Evolutionary Computation*. Anchorage: IEEE Press, 1998: 69 - 73.
- [13] BOX G E P, JENKINS G M. *Time Series Analysis, Forecasting and Control*[M]. San Francisco: Holden Day, 1976.

作者简介:

陈俊风 (1979—), 女, 河海大学讲师, 浙江大学博士研究生, 研究方向为计算智能、神经网络建模, E-mail: jfchen@iipc.zju.edu.cn;

任子武 (1976—), 男, 哈尔滨工业大学博士研究生, 研究方向为进化计算、神经网络, E-mail: zwenjren@yahoo.com.cn;

伞冶 (1951—), 男, 哈尔滨工业大学教授, 博士生导师, 研究方向为系统仿真及最优控制, E-mail: sanyc@hit.edu.cn.