

基于 ϵ -不敏感准则和结构风险的鲁棒径向基函数神经网络学习

桑庆兵^{*①} 邓赵红^② 王士同^② 吴小俊^①

^①(江南大学物联网工程学院 无锡 214122)

^②(江南大学数字媒体学院 无锡 214122)

摘要: 该文提出一种新的基于 ϵ -不敏感准则和结构风险的径向基神经网络(RBF-NN)建模方法。通过引入 ϵ -不敏感准则和结构风险项,把RBF-NN训练转化为线性回归和经典的二次规划优化问题。和传统的基于最小平方误差准则的RBF-NN训练算法相比,提出的新方法对小样本数据集和噪声数据显示出较好的鲁棒性,具有较好的泛化能力。通过模拟和真实数据集进行仿真试验,上述优点得到了有效验证。

关键词: 径向基函数神经网络建模; ϵ -不敏感准则; 结构风险; 鲁棒性

中图分类号: TP183

文献标识码: A

文章编号: 1009-5896(2012)06-1414-06

DOI: 10.3724/SP.J.1146.2011.01045

ϵ -Insensitive Criterion and Structure Risk Based Radius-basis-function Neural-network Modeling

Sang Qing-bing^① Deng Zhao-hong^② Wang Shi-tong^② Wu Xiao-jun^①

^①(School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China)

^②(School of Digital Media, Jiangnan University, Wuxi 214122, China)

Abstract: An ϵ -insensitive criterion and structure risk based Radius-Basis-Function Neural-Network (RBF-NN) modeling method is proposed. By introducing ϵ -insensitive criterion and the item of structure risk, the RBF-NN learning is transformed into the linear regression and Quadratic Program (QP) optimization issue. Compared with the traditional least-square-criterion based RBF-NN training algorithms, the proposed method is much more robust to noise data and small size of datasets. Through the simulation experiments on the synthetic and real-word datasets, the above virtues are confirmed.

Key words: Radius-Basis-Function Neural-Network (RBF-NN) modeling; ϵ -insensitive criterion; Structure risk; Robustness

1 引言

在目前的建模方法中,径向基函数神经网络(Radius-Basis-Function Neural-Network, RBF-NN)建模是有效的方法之一^[1-3]。经典的RBF-NN模型训练主要是基于最小平方误差准则的。该类训练算法存在的一个明显不足是:对于小样本数据集或存在噪声的数据集,训练容易过拟合,所得RBF-NN泛化能力较差^[1,3]。因而,针对此问题探讨鲁棒的RBF-NN训练新算法是一个很有意义的工作。

最经典的一类RBF-NN训练学习算法是基于梯度下降学习策略的。该类算法的缺点是由于学习步长选择不合理或者由于基于经验风险陷入局部极值或过拟合而使受训网络的泛化性降低^[1,3]。关于

RBF-NN训练学习的最新重要研究进展是极速学习训练方法^[4-10],文献[4,11]指出输入层权值和隐层阈值任意选取的包含 N 个隐结点的单隐层前馈网络(SLFN)能够以任意小的误差逼近 N 个不同的观测值。基于以上理论,文献[4]针对单隐层神经网络提出了极速学习机(Extreme Learning Machine, ELM)算法,并进一步提出了各种改进的极速学习算法。ELM算法与常规的梯度下降学习算法相比,简单快速,在获得小的训练误差的同时,能获得较好的泛化能力。但是,针对小样本数据集和噪声,ELM算法依然鲁棒性较差,容易导致过拟合问题。

针对经典学习方法对小样本数据集和噪声数据容易过拟合而使得受训RBF-NN泛化能力较差的问题,本文通过引入 ϵ -不敏感学习度量^[2,12-16]和结构风险项^[2,13,15]来构造新的目标函数,并把提出的新目标函数求解转化为经典的二次规划问题。本文提出的方法由于引入了不敏感学习度量和结构风险项,

2011-10-11 收到, 2012-02-20 改回

国家自然科学基金(90820002, 60903100, 61170122), 江苏省自然科学基金(BK2009067)资助课题

*通信作者: 桑庆兵 sangqb@163.com

能有效地克服经典训练算法针对小数据集容易过拟合和对噪音敏感之缺陷, 显示出了较好的鲁棒性。在模拟和真实数据集上的试验亦证实了上述优点。

2 径向基神经网络

经典的径向基神经网络模型如图1所示^[4]。径向基网络能完成非线性映射 $f: R^d \rightarrow R^1$, 其数学表达式为

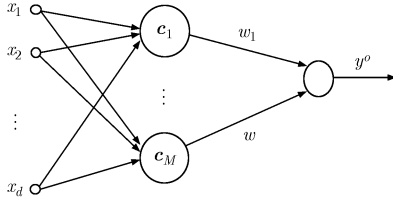


图1 RBF神经网络模型结构

$$y^o = f(\mathbf{x}) = \sum_{i=1}^M w_i \Phi(\|\mathbf{x} - \mathbf{c}_i\|), \quad i = 1, \dots, M \quad (1)$$

其中 $\mathbf{x} \in R^d$ 为网络的输入向量, $\mathbf{c}_i \in R^d$ 为网络隐层节点的径向基函数中心向量。 w_j 为网络隐层和输出节点间的连接权值。 $\|\cdot\|$ 表示范数, 这里取欧几里德范数。 $\Phi(\cdot)$ 称为径向基函数, 完成从 $R^d \rightarrow R^1$ 的非线性变换, 其形式为

$$\Phi(\|\mathbf{x} - \mathbf{c}_i\|) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_i\|^2}{\delta_i}\right) \quad (2)$$

其中 δ_i 为宽度值。

上述的径向基神经网络模型中, 要学习的参数有3种, 即隐层中心点 $\mathbf{c}_i = [c_i^1, \dots, c_i^k, \dots, c_i^d]^T$, 隐层径向基函数宽度值 δ_i , 隐层和输出层的连接权值 w_i 。对于径向基神经网络的各参数, 最经典的一类训练学习算法是基于梯度下降学习策略的。但该类算法由于基于经验风险易于陷入局部极值或过拟合从而使受训网络泛化能力降低, 而且此类算法由于是迭代算法, 常具有很高的时间复杂度。另一类近年来较受关注的算法是极速学习机(ELM)算法。ELM算法与梯度下降学习算法相比, 简单快速, 在获得小的训练误差的同时, 能获得较好的泛化能力。但是, 针对小样本数据集和噪音, ELM依然鲁棒性较差, 容易导致过拟合问题。针对此挑战, 在后面一节, 本研究将探讨一种鲁棒的基于不敏感准则和结构风险的径向基函数神经网络参数学习方法。

3 基于 ϵ -不敏感准则和结构风险的径向基神经网络训练

3.1 径向基神经网络和线性模型

对于径向基神经网络的隐层中心点参数 $\mathbf{c}_i =$

$[c_{i1}, \dots, c_{id}]^T$ 和宽度参数 δ_i , 一种常用的估计方法是聚类法。例如, 利用模糊C均值(FCM)聚类技术, $\mathbf{c}_{ik}, \delta_i$ 可利用下式来估计。

$$\mathbf{c}_{ik} = \frac{\sum_{j=1}^N u_{ji} x_{jk}}{\sum_{j=1}^N u_{ji}} \quad (3)$$

$$\delta_i^k = h \frac{\sum_{j=1}^N u_{ji} \|\mathbf{x}_j - \mathbf{c}_i\|^2}{\sum_{j=1}^N u_{ji}} \quad (4)$$

这里 u_{jk} 表示 FCM 聚类方法得到的样本 $\mathbf{x}_j = [x_{j1}, \dots, x_{jd}]^T$ 对于第 i 类的模糊隶属度^[1,3]; 参数 h 是一个可调的缩放参数。

对于径向基函数神经网络, 一旦隐层参数被估计, 那么令

$$\tilde{x}^i = \Phi(\|\mathbf{x} - \mathbf{c}_i\|), \quad i = 1, \dots, M \quad (5a)$$

$$\mathbf{x}_g = [\tilde{x}^1, \tilde{x}^2, \dots, \tilde{x}^M]^T \quad (5b)$$

$$\mathbf{p}_g = [w_1, \dots, w_M]^T \quad (5c)$$

此时式(1)所示的径向基网络映射函数可表示为

$$y^o = \mathbf{p}_g^T \mathbf{x}_g \quad (6)$$

由式(6)可知, 当径向基网络的隐层节点被估计后, 网络的输出可表示为一个线性模型的输出, 此时网络参数的学习转化为线性模型之参数学习问题。基于此线性模型, 本文通过引入 ϵ -不敏感误差准则和结构风险来构建新的径向基网络学习算法。

3.2 ϵ -不敏感误差准则

径向基神经网络训练时, 常采用式(7)所示的最小平方误差准则^[1,3]:

$$E = \sum_{i=1}^N (y_i - y_{di})^2 \quad (7)$$

一般情况下, 利用此准则能取得较好的效果。基于最小平方误差准则的经典RBF-NN训练算法有基于梯度学习的方法^[1,3]和极速学习方法^[4-10]。但当数据样本较少且噪音较大时, 此准则学习得到的网络对噪音较敏感, 容易过拟合。基于此不足, 本文引入了不敏感误差度量来设计新的径向基网络训练准则函数来增强网络的鲁棒性。

给定标量 g 和向量 $\mathbf{G} = [g_1, \dots, g_d]^T$, 相应的 ϵ -不敏感损失分别具有式(8), 式(9)所示的形式^[12-16]:

$$|g|_{\epsilon} = \begin{cases} g - \epsilon, & g > \epsilon \\ 0, & g \leq 0 \end{cases} \quad (8)$$

$$|\mathbf{G}|_{\epsilon} = \sum_{i=1}^d |g_i|_{\epsilon} \quad (9)$$

对于式(6)所示的线性模型, 其对应的 ϵ -不敏感损失误差准则可定义如下:

$$\min_{\mathbf{p}_g} E = \sum_{i=1}^N |y_i^o - y_i|_\varepsilon = \sum_{i=1}^N |\mathbf{p}_g^T \mathbf{x}_{gi} - y_i|_\varepsilon \quad (10)$$

3.3 基于 ε -不敏感误差准则和结构风险的径向基网络训练

对于式(10), 一般情况下, 给定训练数据集 $\{\mathbf{x}_{gi}, y_i\}$, 不等式 $y_i - \mathbf{p}_g^T \mathbf{x}_{gi} < \varepsilon$, $\mathbf{p}_g^T \mathbf{x}_{gi} - y_i < \varepsilon$ 不总是满足。为此, 引入松弛变量 $\xi_i^+ \geq 0$, $\xi_i^- \geq 0$ 可得到如下约束:

$$\begin{cases} y_i - \mathbf{p}_g^T \mathbf{x}_{gi} < \varepsilon + \xi_i^+ \\ \mathbf{p}_g^T \mathbf{x}_{gi} - y_i < \varepsilon + \xi_i^- \end{cases} \quad (11)$$

利用式(11), 式(10)所示的准则函数可等价地表示为

$$\begin{aligned} \min_{\mathbf{p}_g, \xi_i^+, \xi_i^-} E &= \sum_{i=1}^N (\xi_i^+ + \xi_i^-) \quad (12a) \\ \text{s.t.} \quad &\begin{cases} y_i - \mathbf{p}_g^T \mathbf{x}_{gi} < \varepsilon + \xi_i^+ \\ \mathbf{p}_g^T \mathbf{x}_{gi} - y_i < \varepsilon + \xi_i^- \end{cases}, \xi_i^+ \geq 0, \xi_i^- \geq 0, \forall i \quad (12b) \end{aligned}$$

进一步地, 参照支撑向量机等核方法^[2], 引入结构风险正则化项 $(1/2)\mathbf{p}_g^T \mathbf{p}_g$, 式(12)可改进为

$$\begin{aligned} \min_{\mathbf{p}_g, \xi_i^+, \xi_i^-} E &= \frac{1}{\tau} \sum_{i=1}^N (\xi_i^+ + \xi_i^-) + \frac{1}{2} \mathbf{p}_g^T \mathbf{p}_g \quad (13a) \\ \text{s.t.} \quad &\begin{cases} y_i - \mathbf{p}_g^T \mathbf{x}_{gi} < \varepsilon + \xi_i^+ \\ \mathbf{p}_g^T \mathbf{x}_{gi} - y_i < \varepsilon + \xi_i^- \end{cases}, \xi_i^+ \geq 0, \xi_i^- \geq 0, \forall i \quad (13b) \end{aligned}$$

结构风险思想的引入, 可进一步使得新算法像许多核方法一样, 在小样本训练集环境下, 所受训得到的网络具有较好的泛化能力, 有效地避免过拟合问题。这里 $\tau > 0$ 用来平衡经验误差项和表示结构风险的正则化项的影响。

在式(13)中有两个参数需要给定, 即 τ , ε 需要给定。如何确定其最优值, 目前还没有理论的解析公式可用。实际应用中, 一个简单有效的策略是利用交叉验证法在某个参数集中来确定最优值。特别地, 这里对不敏感参数 ε 给出如下说明: 该参数类似于支撑向量回归(SVR)方法中的不敏感参数 ε 。SVR中该参数的理论研究表明^[17,18]: 该参数的最优值和数据中的噪声的方差呈现近似正比的关系, 即噪声越大, 该参数通常需采用较大的值来获得好的训练效果。在试验部分4.2节, 我们对此结论给出了相关的实验验证。

特别地, 对于式(13), 利用拉格朗日优化可得到其对偶问题(证明略):

$$\begin{aligned} \min_{\lambda^+, \lambda^-} &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\lambda_i^+ - \lambda_i^-) (\lambda_j^+ - \lambda_j^-) \mathbf{x}_{gi}^T \mathbf{x}_{gj} \\ &\quad + \sum_{i=1}^N \lambda_i^+ (y_i - \varepsilon) + \sum_{i=1}^N \lambda_i^- (-y_i - \varepsilon) \quad (14) \\ \text{s.t.} \quad &0 \leq \lambda_i^+, \lambda_i^- \leq (1/\tau), \forall i \end{aligned}$$

根据对偶理论, 利用式(14)得到的最优解 λ^*, λ^{+*} , 可得到式(13)对应的最优解为

$$\mathbf{p}_g^* = \sum_{i=1}^N (\lambda_i^{*-} - \lambda_i^{+*}) \mathbf{x}_{gi} \quad (15)$$

根据上面推导和分析可以看出: 通过引入 ε -不敏感误差准则和结构风险项, RBF网络训练可以视为一个二次规划优化问题。因而已有的各种二次规划算法可以直接用来对 RBF-NN 进行训练。

基于如上分析, 容易给出基于 ε -不敏感准则和结构风险的 RBF-NN 训练新算法, 如表1所示。

表1 基于 ε -不敏感准则和结构风险的 RBF 神经网络训练算法

- | | |
|-----|-------------------------------|
| (1) | 利用聚类算法估计隐层节点参数 |
| (2) | 求解式(14)对应的二次规划问题 |
| (3) | 利用式(14)最优解来得到相应的网络隐层和输出层之间的参数 |

4 试验研究

本节对提出的新RBF-NN训练算法(表示为 ε -RBF)进行了试验测试。为了有效地评估算法性能, 本文提出的算法和经典的基于平方误差准则的梯度学习算法(表示为LS-RBF)和极速学习算法(ELM)进行了比较。试验安排如下: (1)4.1节利用模拟数据集对算法性能进行了测试。(2)4.2节利用真实的煤气炉建模数据集对算法性能进行了测试。试验中, 利用训练集采用了5倍交叉验证策略在集合 $\{10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3\}$ 和 $\{10^{-3}, 2 \times 10^{-3}, 3 \times 10^{-3}, 4 \times 10^{-3}, 5 \times 10^{-3}\}$ 中分别确定参数 τ , ε 的合适取值。

为有效评估该算法性能, 采用了如下性能指标:

$$J = \sqrt{\frac{\frac{1}{N} \sum_{i=1}^N (y'_i - y_i)^2}{\left[\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2 \right]}} \quad (16)$$

这里 N 是测试样本集的尺度; y_i 是第 i 个测试样本的采样输出, y'_i 是第 i 个样本对应的神经网络输出, 并且 $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$ 。 J 越小, 表示建模(泛化)性能越好。

为了充分测试各算法对于噪声数据的鲁棒性, 试验中对训练数据集加入了不同程度的高斯白噪声来进行鲁棒性测试。

4.1 模拟模数据集

本试验所用的数据集采样于如下的sinc函数^[1]:

$$y = \frac{\sin(x)}{x} + N(0, \sigma), \quad x \in [-15, 15] \quad (17)$$

$N(0, \sigma)$ 表示均值为0, 标准差为 σ 的白噪声。利用式(17)产生包含200个数据的训练集, 同时产生一个包

含1000个数据且不含噪声的测试集。利用训练集来进行训练，然后利用测试集进行测试。上述过程重复10次，得到的平均测试结果来评价算法的性能。

图 2 和表 2 示出了噪音环境下 3 种算法在 sinc 数据集上的试验结果。从图 2 和表 2，我们可得出如下的观察结果：

(1)当噪声较小时，几种不同方法展现出了可比较的泛化性能。

(2)随着噪声程度的增加，几种不同方法的泛化能力都逐步下降。

(3)在几种方法中，提出的基于 ϵ -不敏感准则和结构风险的新训练算法随着噪声的增加，泛化能力减弱的程度最小。特别是当噪声程度较大时，其泛化能力明显优于其它两种方法。

根据如上观察结果可知，针对该模拟数据提出的新算法在噪声环境下展现出了更好的适应能力。

4.2 Box-Jenkins(B-J)煤气炉真实数据集

本试验研究了 Box-Jenkins(B-J)煤气炉建模问题^[19,20]。煤气炉控制输入 $u(k)$ 表示气体进入熔炉的

量，输出 $y(k)$ 表示煤气炉中 CO_2 的输出量，数据样本每隔 9 s 采样一次，共得到 296 组输入输出样本对。根据文献[1]中的方法，试验中选择 $u(k-3)$ 和 $y(k-1)$ 作为输入变量， $y(k)$ 作为输出量重新构造数据集，得到 290 个有效数据对。把数据集随机分为训练集和测试集两部分，其中训练集和测试集尺寸分别为 145。对训练数据集加不同程度的噪声并利用训练集来进行训练，然后利用测试集进行测试。上述过程重复 10 次，得到的平均试验结果如图 3 和表 3 所示。正如同前一试验得到的观察结果，从图 3 和表 3 同样可以看出类似的实验效果：(1)采用不同数目的隐节点，提出的新训练算法在噪声较小时，展现了与经典的方法可比较的性能；(2)但当噪声加大时，提出的新算法展现了更好的鲁棒性，泛化性能明显优于其它两种经典算法。因而，在煤气炉真实数据集的试验结果表明，本文提出的基于 ϵ -不敏感准则和结构风险的 RBF-NN 训练算法，较之于经典的训练算法在噪声环境下展现出了更有希望的性能。

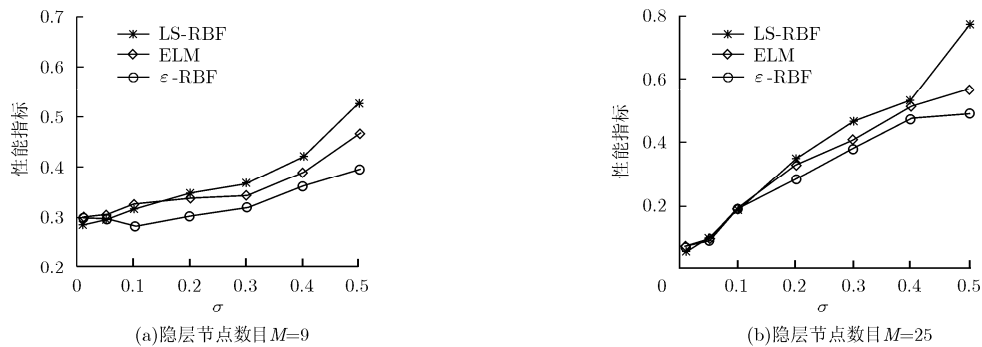


图 2 采用不同数目隐节点时 3 种方法在 sinc 数据集上得到的的平均性能比较

表 2 采用不同数目隐节点时 3 种方法在 sinc 数据集上得到的的平均性能比较

噪声 (σ)	性能	$M=9$			$M=25$		
		LS-RBF	ELM	ϵ -RBF	LS-RBF	ELM	ϵ -RBF
0.01	均值	0.2853	0.3012	0.2959	0.0530	0.0692	0.0680
	均方差	0.1031	0.1134	0.0926	0.0066	0.0553	0.0425
0.05	均值	0.2960	0.3034	0.2977	0.0967	0.0886	0.0868
	均方差	0.0908	0.0986	0.0904	0.0129	0.0121	0.0108
0.1	均值	0.3152	0.3245	0.2794	0.1851	0.1903	0.1882
	均方差	0.0937	0.0875	0.0602	0.0328	0.0342	0.0213
0.2	均值	0.3488	0.3396	0.3039	0.3471	0.3291	0.2812
	均方差	0.0957	0.0879	0.1083	0.0299	0.0459	0.0410
0.3	均值	0.3676	0.3426	0.3195	0.4674	0.4096	0.3785
	均方差	0.0834	0.0963	0.0907	0.1065	0.0897	0.0526
0.4	均值	0.4203	0.3892	0.3618	0.5329	0.5154	0.4756
	均方差	0.0898	0.1043	0.1309	0.0264	0.0957	0.0959
0.5	均值	0.5254	0.4682	0.3955	0.7739	0.5638	0.4906
	均方差	0.1115	0.1067	0.1778	0.1209	0.1092	0.0701

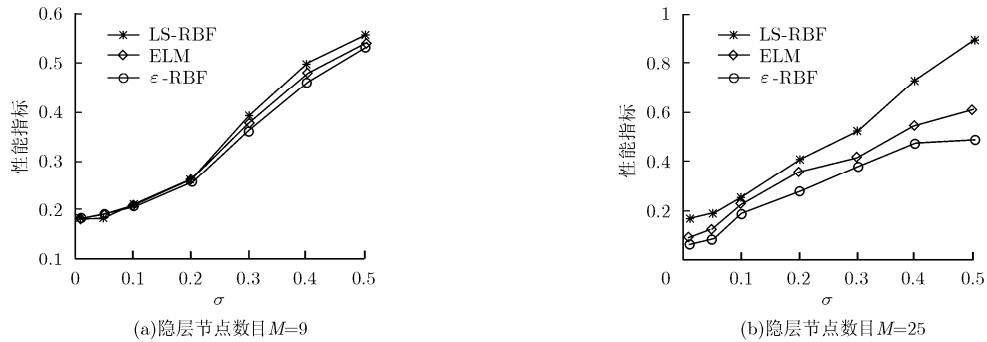


图 3 采用不同数目隐节点时 3 种方法在煤气炉建模数据集上得到的的平均性能比较

表 3 采用不同数目隐节点时 3 种方法在煤气炉建模数据集上得到的的平均性能比较

噪声(σ)	性能	$M=9$			$M=25$		
		LS-RBF	ELM	ϵ -RBF	LS-RBF	ELM	ϵ -RBF
0.01	均值	0.1797	0.1834	0.1828	0.1696	0.0982	0.0680
	均方差	0.0025	0.0052	0.0044	0.0048	0.0302	0.0425
0.05	均值	0.1840	0.1924	0.1907	0.1909	0.1297	0.0868
	均方差	0.0128	0.0143	0.0079	0.0096	0.0125	0.0108
0.1	均值	0.2053	0.2109	0.2091	0.2511	0.2296	0.1882
	均方差	0.0122	0.0212	0.0134	0.0203	0.0219	0.0213
0.2	均值	0.2600	0.2613	0.2571	0.4122	0.3593	0.2812
	均方差	0.0235	0.0423	0.0548	0.0267	0.0410	0.0410
0.3	均值	0.3894	0.3772	0.3631	0.5228	0.4193	0.3785
	均方差	0.0900	0.0924	0.0876	0.0750	0.0621	0.0526
0.4	均值	0.4987	0.4792	0.4591	0.7297	0.5490	0.4756
	均方差	0.0899	0.0776	0.0710	0.1571	0.1231	0.0959
0.5	均值	0.5549	0.5373	0.5299	0.8915	0.6129	0.4906
	均方差	0.1154	0.0682	0.0470	0.1138	0.0948	0.0701

4.3 不敏感参数 ϵ 的选择与噪音方差之间存在的比例关系

正如 3.3 节所述, 文献[17,18]表明不敏感参数 ϵ 的合适取值和噪音有一定的正比关系。这里我们利用实验进行简单验证。表 4 给出了实验中得到的最优不敏感参数 ϵ 和噪音标准差 σ 之间的关系。从表中我们容易看出, 随着噪音的增加, 最优的 ϵ 取值也逐渐增加, 即它们之间呈现近似的正比关系。值得指出的是, 这里我们仅给出了简单的实验验证结果, 如何从理论上证明上述结论是一个较复杂的工作, 这需要借鉴文献[19,20]的相关研究思路来进行严格的数学分析, 未来的工作中我们将对此作深入探讨。

表 4 最优的不敏感参数 ϵ 和噪音之间 σ 的关系

σ	0.01	0.05	0.1	0.2	0.3	0.4	0.5
ϵ (sinc)	1e-3	1e-3	2e-3	2e-3	3e-3	4e-3	4e-3
ϵ (B-J)	2e-3	2e-3	3e-3	3e-3	4e-3	4e-3	4e-3

5 结束语

本文通过引入 ϵ -不敏感准则和结构风险, 把 RBF-NN 训练转化为线性回归问题, 进而转化为经典的二次规划优化问题, 提出了一种新的径向基神经网络建模方法。通过模拟和真实数据集进行仿真试验, 提出的算法较之于传统的基于最小平方误差准则的算法对噪音数据集显示出了更鲁棒的性能。进一步地, 基于计算智能研究方面的最新进展, 提出适宜于大规模数据集的基于 ϵ -不敏感准则和结构风险的径向基神经网络快速训练算法值得探讨, 未来的工作中, 我们将对此进行深入研究。

参考文献

- [1] Jang J S R, Sun C T, and Mizutani E. Neuro-Fuzzy and Soft-Computing. Upper Saddle River, NJ, Prentice-Hall, 1997: 125-134.
- [2] Vapnik V. Statistical Learning Theory. New York: Wiley, 1998: 256-396.
- [3] 王士同, 等. 神经模糊系统及其应用. 北京: 北京航空航天大学

- 学出版社, 1998: 198-224.
- Wang S T, *et al.* Neural Fuzzy System and Its Application. Beijing: Beijing University of Aeronautics and Astronautics, 1998: 198-224.
- [4] Huang G B, Zhu Q Y, and Siew C K. Extreme Learning Machine: Theory and Applications. *Neurocomputing*, 2006, 70(1-3): 489-501.
- [5] Cao J W, Lin Z P, and Huang G B. Composite function wavelet neural networks with differential evolution and extreme learning machine. *Neural Processing Letters*, 2011, 33(3): 251-265.
- [6] Huang G B and Wang D H. Advances in extreme learning machines (ELM2010). *Neurocomputing*, 2011, 74(16): 2411-2412.
- [7] Lan Y, Soh Y C, and Huang G B. Two-stage extreme learning machine for regression. *Neurocomputing*, 2010, 73(16-18): 3028-3038.
- [8] Lan Y, Soh Y C, and Huang G B. Constructive hidden nodes selection of extreme learning machine for regression. *Neurocomputing*, 2010, 73(16-18): 3191-3199.
- [9] Cao J W, Lin Z P, and Huang G B. Composite function wavelet neural networks with extreme learning machine. *Neurocomputing*, 2010, 73(7-9): 1405-1416.
- [10] Huang Guang-bin, Ding Xiao-jian, and Zhou Hong-ming. Optimization method based extreme learning machine for classification. *Neurocomputing*, 2011, 74(1-3): 155-163.
- [11] Huang G B. Learning capability and storage capacity of two-hidden-layer feed-forward networks. *IEEE Transactions on Neural Networks*, 2003, 14(2): 274-281.
- [12] Leski J. Towards a robust fuzzy clustering. *Fuzzy Sets and Systems*, 2003, 12(2): 215-233.
- [13] Leski J. TSK-fuzzy modeling based on ε -insensitive learning. *IEEE Transactions on Fuzzy Systems*, 2005, 13(2): 181-193.
- [14] 邓赵红, 王士同. 鲁棒的模糊聚类神经网络. *软件学报*, 2005, 16(8): 1415-1422.
- Deng Z H and Wang S T. Robust fuzzy clustering neural networks. *Journal of Software*, 2005, 16(8): 1415-1422.
- [15] Deng Z H, Choi K S, Chung F L, *et al.* Scalable TSK fuzzy modeling for very large datasets using minimal-enclosing-ball approximation. *IEEE Transactions on Fuzzy Systems*, 2010, 18(2): 210-226.
- [16] Deng Z H, Choi K S, Chung F L, *et al.* Enhanced soft subspace clustering integrating within-cluster and between-cluster information. *Pattern Recognition*, 2010, 43(3): 767-781.
- [17] Kwok J T and Tsang I W. Linear dependency between epsilon and the input noise in epsilon-support vector regression. *IEEE Transactions on Neural Networks*, 2003, 14(3): 544-553.
- [18] Wang S T, Zhu J G, Chung F L, *et al.* Theoretically optimal parameter choices for support vector regression machines with noisy input. *Soft Computing*, 2005, 9(10): 732-741.
- [19] Box G E P and Jenkins G. M. Time Series Analysis, Forecasting and Control. 2nd Ed, San Francisco, CA, Holden Day, 1976: 355-390.
- [20] Chung F L, Deng Z H, and Wang S T. From minimum enclosing ball to fast fuzzy inference system training on large datasets. *IEEE Transactions on Fuzzy Systems*, 2009, 17(1): 173-184.
- 桑庆兵: 男, 1973年生, 博士生, 副教授, 研究方向为神经网络与图像质量评价.
- 邓赵红: 男, 1981年生, 博士, 副教授, 研究方向为神经模糊建模与计算智能.
- 王士同: 男, 1964年生, 博士生导师, 教授, 研究方向为神经模糊建模与计算智能.
- 吴小俊: 男, 1967年生, 博士生导师, 教授, 研究方向为计算机视觉与计算智能.