

文章编号: 1001-0920(2012)09-1308-05

## 针对小样本数据集的鲁棒单隐层前馈网络建模方法

张荣<sup>1a</sup>, 邓赵红<sup>1b</sup>, 王士同<sup>1b</sup>, 蔡及时<sup>2</sup>, 钱鹏江<sup>1b</sup>

(1. 江南大学 a. 物联网工程学院, b. 数字媒体学院, 江苏 无锡 214122; 2. 香港理工大学 护理学院, 香港 999077)

**摘要:** 单隐层前馈神经网络是应用最广泛的智能建模模型之一, 但该模型面对小样本集时传统的学习算法易陷入过拟合, 尤其当数据集包含较大噪音时学习模型鲁棒性较差, 对噪音很敏感. 针对此不足, 提出一种针对小样本数据集的鲁棒单隐层前馈神经网络学习算法. 所提出的算法由于引入了  $\epsilon$ -不敏感学习度量 and 结构风险项, 能有效克服传统学习算法存在的缺陷, 显示出较好的鲁棒性. 在模拟和真实数据集上的实验亦证实了上述优点.

**关键词:** 单隐层前馈神经网络;  $\epsilon$ -不敏感学习; 结构风险最小化; 鲁棒性

中图分类号: TP181

文献标志码: A

## Robust single hidden layer feed-forward neural networks modeling for small datasets

ZHANG Rong<sup>1a</sup>, DENG Zhao-hong<sup>1b</sup>, WANG Shi-tong<sup>1b</sup>, CHOI Kup-sze<sup>2</sup>, QIAN Peng-jiang<sup>1b</sup>

(1a. School of Internet of Things Engineering, 1b. School of Digital Media, Jiangnan University, Wuxi 214122, China; 2. School of Nursing, The Hong Kong Polytechnic University, Hong Kong 999077, China. Correspondent: WANG Shi-tong, E-mail: wxwangst@yahoo.com.cn)

**Abstract:** Single hidden layer feed-forward neural network(SLFN) is one of the most widely used models for intelligent modeling. But the model faces that for small sample sets, the traditional learning algorithm may train a model to fall into the over-fitting state. In particular, when the dataset contains a large amount of noise, the trained model has weak robustness and is very sensitive to noise. In order to overcome this shortcoming, a robust learning algorithm of SLFN is derived for small and noisy datasets. Due to the introduction of  $\epsilon$ -insensitive learning measure and the structural risk term, the proposed algorithm can effectively overcome the shortcoming of the traditional learning algorithm. The experimental results on simulated and real-world datasets also confirm the above advantages.

**Key words:** single hidden layer feed-forward neural network;  $\epsilon$ -insensitive learning; structural risk minimization; robustness

### 1 引言

单隐层前馈神经网络(SLFN)因其具有非常强的非线性逼近能力而得到了广泛应用<sup>[1]</sup>, 但目前SLFN依然面临一些问题, 其中一个关键问题是面对小样本数据集和噪音时, 采用传统SLFN学习算法所得到的模型鲁棒性较差. 因此, 研究如何提高SLFN学习算法所得到的模型的鲁棒性是一项很有意义的工作.

最经典的一类SLFN学习算法是基于梯度下降学习策略的学习算法. 该类算法的缺点是, 由于学习步长选择不合理或基于经验风险而陷入局部极值<sup>[2]</sup>或过拟合, 使网络泛化性能降低. 关于SLFN学习

的最新重要研究进展是极速学习机(ELM)算法<sup>[3-5]</sup>. 文献[3-4]指出, 如果任意选取输入层权值和隐层阈值, 则有  $N$  个隐结点的SLFN能以任意小的误差逼近  $N$  个不同的观测值; [3]则指出并证明了如果激励函数是无限可微的, 则输入层权值和隐层阈值任意选取的SLFN所需的隐节点数目  $N' \leq N$ , 这里  $N$  是训练样本的个数. 基于以上理论, [5]针对SLFN提出了ELM算法. ELM算法与常规的梯度下降学习算法相比, 简单快速, 参数选择相对容易, 在获得最小训练误差的同时能够保证输出层权值的范数最小, 从而能够获得较好的泛化能力. 但是, ELM算法对于小样本数据集

收稿日期: 2011-04-22; 修回日期: 2011-07-20.

基金项目: 国家自然科学基金项目(60903100); 江苏省自然科学基金项目(BK2009067); 中央高校基本科研业务费专项资金项目(JUSRP21128); 江苏省信息融合软件工程技术研究开发中心开放基金项目(SR-2011-01).

作者简介: 张荣(1986-), 女, 硕士生, 从事模糊系统建模的研究; 王士同(1964-), 男, 教授, 博士生导师, 从事模式识别、人工智能等研究.

和噪音,鲁棒性依然较差,容易出现过拟合现象.

本文针对小样本噪音数据集提出一种新的鲁棒的SLFN学习算法.该算法通过引入 $\epsilon$ -不敏感学习<sup>[6-8]</sup>度量和结构风险项<sup>[6,9-10]</sup>来构造新的目标函数,并利用极速学习理论将新目标函数的求解转化为经典的二次规划问题.由于引入了不敏感学习度量和结构风险项,新算法能有效地克服传统的SLFN学习算法针对小样本集容易过拟合和对噪音敏感的缺陷,显示出了较好的鲁棒性.另外,新方法中 $\epsilon$ -不敏感学习度量和结构风险项的引入,为目标函数增加了2个可调参数,其中不敏感系数 $\epsilon$ 用于控制拟合精度,另一个参数平衡因子 $\tau$ 用于控制经验风险项和结构风险项在目标函数中所占的比重关系.针对不同的样本集以及不同的需求,采用交叉验证可确定适宜的参数值,从而得到相应的优化网络模型.因此,本文算法对不同的环境和样本集具有较强的适应性.在模拟和真实数据集上的实验亦证实了上述优点.

## 2 单隐层前馈神经网络

### 2.1 结构和数学描述

给定一个多输入多输出的样本数据 $(\mathbf{x}, \mathbf{t})$ .其中: $\mathbf{x} = (x_1, x_2, \dots, x_p)$ ,  $\mathbf{t} = (t_1, t_2, \dots, t_q)$ .具有 $N'$ 个隐节点,激励函数为 $g(\mathbf{x})$ 的标准SLFN网络结构如图1所示.其模型输出可表示为

$$y_j = f_j(\mathbf{x}) = \sum_{i=1}^{N'} \beta_{ij} g(\mathbf{w}_i^T \mathbf{x} + b_i), \quad j = 1, 2, \dots, q. \quad (1)$$

其中: $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{ip}]^T \in \mathbf{R}^p$ 是连接第 $i$ 个隐节点和输入节点的输入权值向量; $\beta_{ij}$ 是连接第 $i$ 个隐节点和第 $j$ 个输出节点的输出权值; $b_i$ 是第 $i$ 个隐节点的阈值;输出采用线性神经元, $y_j$ 是第 $j$ 个输出.理论上已经证明,具有 $N$ 个隐节点和无限可微激活函数的标准SLFN,可以以零误差逼近 $N$ 个不同的输入输出样本对<sup>[3]</sup>.

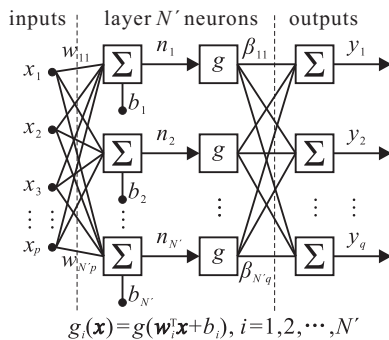


图1 单隐层前馈神经网络结构

由于多输入多输出SLFN可以分解为多个多输入单输出SLFN加以分析,下面均针对多输入单输出SLFN展开讨论.

### 2.2 经典的SLFN学习算法

#### 2.2.1 基于梯度下降的学习算法

给定一个多输入单输出样本数据集

$$D = \{(\mathbf{x}_n, t_n) | \mathbf{x}_n \in \mathbf{R}^p, t_n \in \mathbf{R}, n = 1, 2, \dots, N\},$$

为了训练SLFN,采用传统的基于梯度下降的学习算法最小化如下代价函数:

$$E(\mathbf{w}_i, b_i) = \sum_{n=1}^N (f(\mathbf{x}_n) - t_n)^2 = \sum_{n=1}^N \left( \sum_{i=1}^{N'} \beta_i g(\mathbf{w}_i^T \mathbf{x}_n + b_i) - t_n \right)^2. \quad (2)$$

对于参数 $W$ (权值 $\mathbf{w}_i$ 和阈值 $b_i$ 的参数集合),通常采用如下规则进行学习:

$$W_k = W_{k-1} - \eta \frac{\partial E(W)}{\partial W}, \quad (3)$$

其中 $\eta$ 为学习率.

#### 2.2.2 基于极速学习策略的算法

文献[5]指出并证明了,如果激励函数 $g(\mathbf{x})$ 是无限可微的,则对于给定的任意小的逼近误差,输入层权值和隐层阈值任意选取的SLFN所需的隐节点数目 $N' \leq N$ ,其中 $N$ 是训练样本的个数.

根据以上理论,当激励函数 $g(\mathbf{x})$ 无限可微时,网络参数并不是全部都需要进行调整,输入权值 $\mathbf{w}_i$ 和隐层阈值 $b_i$ 在训练时可随机赋值,且在训练过程中固定不变.这样,对于选定的样本数据,隐层输出矩阵固定不变,进而使得SLFN的参数训练转化为求解如下的线性回归问题:

$$f(\mathbf{x}) = \boldsymbol{\beta}^T \mathbf{g}. \quad (4a)$$

此时, $\boldsymbol{\beta}$ 可通过求解线性系统 $\mathbf{H}\boldsymbol{\beta} = \mathbf{T}$ 的最小二乘解获得.对于多输入单输出数据集 $D = \{(\mathbf{x}_n, t_n) | \mathbf{x}_n \in \mathbf{R}^p, t_n \in \mathbf{R}, n = 1, 2, \dots, N\}$ ,求解如下优化问题:

$$\min_{\boldsymbol{\beta}} \|\mathbf{H}\boldsymbol{\beta} - \mathbf{T}\|. \quad (4b)$$

其中

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1^T \mathbf{x}_1 + b_1) & \dots & g(\mathbf{w}_{N'}^T \mathbf{x}_1 + b_{N'}) \\ \vdots & \ddots & \vdots \\ g(\mathbf{w}_1^T \mathbf{x}_N + b_1) & \dots & g(\mathbf{w}_{N'}^T \mathbf{x}_N + b_{N'}) \end{bmatrix}$$

是隐层输出矩阵<sup>[11]</sup>, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_{N'})^T$ 是输出节点与隐层连接权向量, $\mathbf{T} = (t_1, t_2, \dots, t_N)^T$ .式(4b)的最小二乘解为

$$\hat{\boldsymbol{\beta}} = \mathbf{H}^+ \mathbf{T}, \quad (5)$$

其中 $\mathbf{H}^+$ 称为隐层输出矩阵 $\mathbf{H}$ 的Moore-Penrose广义逆<sup>[12]</sup>.最优解 $\hat{\boldsymbol{\beta}}$ 存在以下重要特性:1)可通过该解获得最小的训练误差;2)获得最小范式的权值向量;3)范式的最小二乘解是唯一的,因此算法不会产生局部最优解<sup>[5]</sup>.

### 2.3 传统 SLFN 学习算法之不足

目前,传统的 SLFN 学习算法已经得到较为广泛的应用,但这些传统的算法仍面临一些挑战:基于梯度下降的学习算法通常由于过学习或基于经验风险最小而陷入局部极值,使网络的泛化性能降低. ELM 算法与常规的梯度下降学习算法相比,简单快速,参数选择相对容易,在获得最小训练误差的同时能够保证输出层权值的范数最小.但是,ELM 算法依然是基于经验风险最小化准则的,其对于小样本数据集和噪音,鲁棒性依然较差,容易出现过拟合现象.

### 3 针对小样本数据集的鲁棒 SLFN 学习算法

下面针对上述问题提出一种新的鲁棒的 SLFN 学习算法.首先通过引入  $\varepsilon$ -不敏感学习度量和结构风险项构造新的目标函数;然后利用极速学习理论将所提出的新目标函数求解转化为经典的二次规划问题;最后给出具体的算法描述.

#### 3.1 不敏感度量和结构风险

首先简单介绍近几年受到广泛关注的  $\varepsilon$ -不敏感损失函数和结构风险的概念.根据文献[12]中的定义,对于一个标量  $g$ , $\varepsilon$ -不敏感损失函数有如下形式:

$$|g|_\varepsilon \triangleq \begin{cases} 0, & |g| \leq \varepsilon; \\ g - \varepsilon, & |g| > \varepsilon. \end{cases} \quad (6)$$

同时,对于向量  $\mathbf{g} = [g_1, g_2, \dots, g_d]^T$ , $\varepsilon$ -不敏感损失函数可以定义为

$$|\mathbf{g}|_\varepsilon \triangleq \sum_{i=1}^d |g_i|_\varepsilon, \quad (7)$$

其中  $\varepsilon$  为不敏感系数.本文中的新目标函数的经验风险项引用了  $\varepsilon$ -不敏感损失函数,用来控制模型实际输出与样本输出之间可容忍的误差,因而参数  $\varepsilon$  可以用于控制这种拟合精度.

学习模型的实际风险由经验风险(训练误差)和置信范围两部分组成,其中置信范围与学习机器的 VC 维及训练样本的个数有关.在训练样本数有限的情况下,训练样本数越小,学习模型的 VC 维越高(复杂性越高),则置信范围越大,使得实际风险与经验风险之间的差别也越大,导致过拟合现象的出现.为了解决这个问题,需同时最小化经验风险和置信范围,从而使实际风险达到最小.为此,统计学习理论<sup>[9-10]</sup>引入了结构风险的概念.最小化模型的结构风险可以使得到的模型具有较小的 VC 维,从而具有较好的泛化性能.

#### 3.2 基于不敏感度量和结构风险的学习目标函数

基于  $\varepsilon$ -不敏感损失函数和结构风险最小化准则,为 SLFN 提出如下目标函数:

$$\begin{aligned} \min_{\boldsymbol{\beta}} \sum_{n=1}^N |t_n - f(\mathbf{x}_n)|_\varepsilon + \frac{\tau}{2} \boldsymbol{\beta}^T \boldsymbol{\beta} = \\ \min_{\boldsymbol{\beta}} \sum_{n=1}^N |t_n - \boldsymbol{\beta}^T \mathbf{g}(\mathbf{x}_n)|_\varepsilon + \frac{\tau}{2} \boldsymbol{\beta}^T \boldsymbol{\beta}. \end{aligned} \quad (8)$$

其中:  $f(\mathbf{x}_n)$  和  $t_n$  分别为模型的实际输出和样本输出,  $\sum_{n=1}^N |t_n - f(\mathbf{x}_n)|_\varepsilon$  和  $\frac{1}{2} \boldsymbol{\beta}^T \boldsymbol{\beta}$  分别为经验风险和结构风险项.式(8)的目标是同时最小化经验风险和结构风险.式中的  $\tau$  是平衡因子,用来平衡这 2 项在学习中的影响,可用交叉验证策略估计出一个合适的值.

#### 3.3 理论推导及算法

下面对式(8)中的目标函数进行理论推导,将其转化为相应的二次规划问题.

一般情况下,不等式  $t_n - \boldsymbol{\beta}^T \mathbf{g}(\mathbf{x}_n) \leq \varepsilon$  和  $\boldsymbol{\beta}^T \mathbf{g}(\mathbf{x}_n) - t_n \leq \varepsilon$  不是对所有的训练样本都能满足.如果引入松弛变量  $\xi_n^+ \geq 0$  和  $\xi_n^- \geq 0$ ,则对于所有的训练样本  $(\mathbf{x}_n, t_n)$ ,有

$$\begin{cases} t_n - \boldsymbol{\beta}^T \mathbf{g}(\mathbf{x}_n) \leq \varepsilon + \xi_n^+, \\ \boldsymbol{\beta}^T \mathbf{g}(\mathbf{x}_n) - t_n \leq \varepsilon + \xi_n^-. \end{cases} \quad (9)$$

利用式(9),可将式(8)等价地表示为

$$\min_{\boldsymbol{\beta}} \frac{1}{\tau} \sum_{n=1}^N (\xi_n^+ + \xi_n^-) + \frac{1}{2} \boldsymbol{\beta}^T \boldsymbol{\beta}. \quad (10)$$

根据对偶理论,在条件  $\xi_n^+ \geq 0, \xi_n^- \geq 0$  和不等式组(9)的约束下,可得到其对偶优化目标函数

$$\begin{aligned} \max_{\lambda^+, \lambda^-} - \frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N (\lambda_n^+ - \lambda_n^-) (\lambda_m^+ - \lambda_m^-) \mathbf{g}^T(\mathbf{x}_n) \mathbf{g}(\mathbf{x}_m) - \varepsilon \sum_{n=1}^N (\lambda_n^+ + \lambda_n^-) + \\ \sum_{n=1}^N (\lambda_n^+ - \lambda_n^-) t_n, \quad \lambda_n^+, \lambda_n^- \in \left[0, \frac{1}{\tau}\right]. \end{aligned} \quad (11)$$

其中:  $n$  和  $m$  都是样本数的标号;  $\lambda_n^+, \lambda_n^- \in [0, 1/\tau]$  是拉格朗日乘子.利用式(11)的最优解,可以得到如下原问题的最优解:

$$\boldsymbol{\beta} = \sum_{n=1}^N (\lambda_n^+ - \lambda_n^-) \mathbf{g}(\mathbf{x}_n). \quad (12)$$

由上述推导可以看出,为 SLFN 学习提出的基于  $\varepsilon$ -不敏感损失函数和结构风险最小化准则的新目标函数可以转化为经典的二次规划问题,因而已有的任何二次规划求解器都可以直接用来求解该优化问题.

给定一个多输入单输出数据集  $D = \{(\mathbf{x}_n, t_n) | \mathbf{x}_n \in \mathbf{R}^p, t_n \in \mathbf{R}, n = 1, 2, \dots, N\}$ ,激励函数  $\mathbf{g}(\mathbf{x})$ ,以及隐层节点数目  $N'$ ,本文提出如下 SLFN 训练算法.

#### 算法 1 鲁棒 SLFN 学习算法 (RSLFN).

Step 1: 为参数  $\varepsilon$  和  $\tau$  赋值;

Step 2: 根据极速学习理论,随机分配输入权值向

量  $w_i$  和隐层节点的阈值  $b_i, i = 1, 2, \dots, N'$ ;

Step 3: 计算隐层输出矩阵  $H$ ;

Step 4: 求解式(11)中的二次规划问题, 利用式(12)得到输出节点和隐层连接权向量  $\beta$ .

## 4 实验研究

### 4.1 实验设置

针对小样本集, 采用本文提出的RSLFN算法进行仿真实验, 并与ELM算法进行比较. 实验中所有算法均利用Matlab语言实现. 相关研究表明, ELM学习算法通常能取得比基于梯度下降的学习算法更好的性能, 因此这里不再与基于梯度下降的学习算法进行比较.

实验的具体内容如下:

1) 分别针对人工数据集和真实数据集进行仿真实验. 首先针对没有加噪音的样本数据选定一个小的隐层节点数, 分别用以上两种算法对网络进行训练; 然后将节点数按一定间隔逐渐增加(不超过训练样本的容量), 对这些含隐层节点数量不同的网络进行训练.

2) 给样本数据加少量的噪音(实验中所用的是高斯噪音, 用Matlab中的randn函数产生, 均值为0), 噪音只加在训练样本的目标输出值上, 测试样本保持零噪音; 然后针对上述逐渐增加隐层节点数的网络分别进行训练; 进一步, 将噪音逐步增大(这里randn函数产生的噪音均值还是0, 增大的是标准差 $\sigma$ ), 每加一次噪音, 就重复一次上述实验.

新方法中的不敏感系数 $\epsilon$ 用于在某种程度上控制拟合精度, 可根据样本或环境的实际情况自行选定; 平衡因子 $\tau$ 用于控制经验风险项和结构风险项在目标函数中所占的比重关系, 实验中可采用交叉验证策略确定其取值.

实验中采用训练和测试的均方根误差(RMSE)衡量网络的学习能力和泛化能力; 每一个独立实验重复50次, 采用其平均值和标准偏差刻画其性能.

实验的目标有以下2点:

1) 随着噪音的不断增强和隐层节点数的不断增加, 探究新方法在泛化性能上的变化情况;

2) 通过与ELM算法的比较, 验证新方法在鲁棒性, 即在抗过拟合和抗噪音方面的优越性.

本研究的实验部分采用了一个人工数据集和一个真实数据集. 人工数据集采用的是广泛应用的Sinc数据集<sup>[5]</sup>; 真实数据集采用的是针对化学机械抛光的二维轴对称静态有限元数据集(简称CMP dataset)<sup>[13-14]</sup>. 这两组数据集的基本信息如表1所示.

表1 实验中数据集的基本信息

数据集	训练集尺寸	测试集尺寸	特征数目(输入+输出)
Sinc	200	200	1 + 1
CMP data	27	8	3 + 1

### 4.2 人工数据集的回归估计

实验1 采用本文提出的RSLFN算法和ELM算法逼近如下Sinc函数:

$$y = f(x) = \begin{cases} \sin(x)/x, & x \neq 0; \\ 1, & x = 0. \end{cases} \quad (13)$$

样本中自变量 $x$ 的取值随机均匀地分布在区间(-10, 10)上. 实验中RSLFN算法的参数 $\epsilon$ 取0.05, 参数 $\tau$ 用5倍交叉验证确定. 实验结果如图2和图3所示.

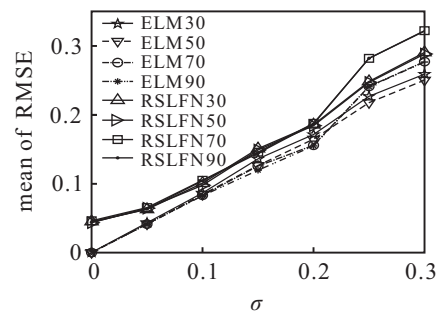


图2 Sinc数据集上的训练精度比较

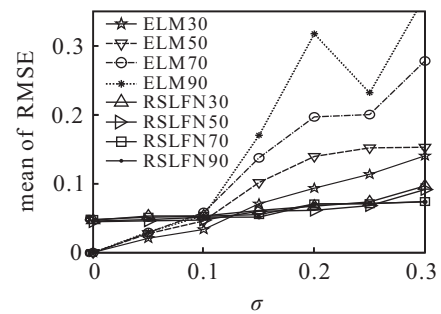


图3 Sinc数据集上的测试精度比较

从图2和图3可以得到如下观察结果:

1) 对于所加噪音(噪音指标 $\sigma$ 在0~0.2范围内)一定的同一个数据集, 随着隐节点数的增加, ELM算法的训练精度逐渐增高, 但测试精度会出现逐渐变差的现象. 这表明随着隐节点数的增加, ELM算法出现了过拟合现象.

2) 对于同一个数据集, 随着隐节点数的增加, RSLFN算法不论训练和测试误差都可以控制在一个可以接受的范围之内, 尤其当隐节点数较大时, 其测试精度明显高于ELM算法.

3) 随着噪音的增大, 两种算法的训练精度都在逐渐降低. 相比较而言, ELM的训练精度总体上比RSLFN算法高, 但其测试精度大多数情况较低. 这表明ELM算法对噪音较敏感, 从而导致泛化能力较差, 而本文算法则对噪音具有较好的鲁棒性.

由以上实验结果可以看出,对于 Sinc 人工数据集, RSLFN 算法在抗噪音和抗过拟合方面比 ELM 算法更优越,即 RSLFN 算法具有更好的鲁棒性。

### 4.3 真实数据集的回归估计

**实验 2** 采用 ELM 算法和 RSLFN 算法实现针对化学-机械抛光 (CMP) 过程的一个二维轴对称静态有限元建模<sup>[14]</sup>。CMP 利用作用于晶片载体上表面的压力来抛光垫,注射泥浆,从而抛光并清除晶圆表面的金属薄膜<sup>[13]</sup>。数据集包括以下 6 个属性:载体负载  $L_0$ ,垫的弹性系数  $E_0$ ,垫的厚度  $T_0$ ,晶片中心上的 Von Mises 压力 CS,晶片中心上最大的 Von Mises 压力 MS,晶片表面的不均匀性  $R$ <sup>[13]</sup>。实验中,建模控制的 3 个输入分别是:1) 载体负载  $L_0$ ; 2) 弹性系数  $E_0$ ; 3) 垫的厚度  $T_0$ 。这里用这 3 个输入数据来预测晶片表面的不均匀性  $R$ 。事实上,也可以为 CS 和 MS 报告预测结果,但为节省空间而省略了。整个数据集包括 27 对训练数据和 8 对测试数据。实验中, RSLFN 算法的参数  $\varepsilon$  取 0.05,参数  $\tau$  用 5 倍交叉验证确定。实验结果如图 4 和图 5 所示。

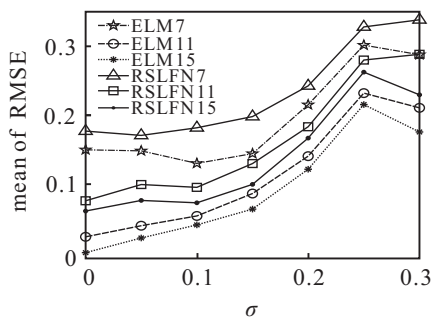


图 4 CMP 数据集上的训练精度比较

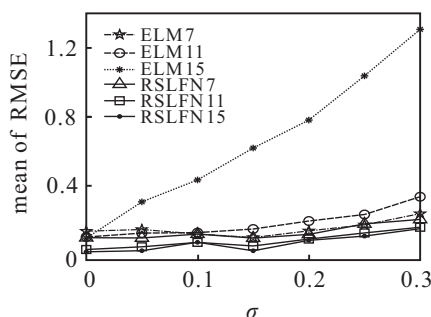


图 5 CMP 数据集上的测试精度比较

由图 4 和图 5 可以看到,当隐节点数增加到 15 时,ELM 算法出现了明显的过拟合现象,而 RSLFN 算法在这方面表现较好。另外,随着噪音的增大,两种算法的训练精度总体而言都在降低。同时,在隐节点数相同的情况下,ELM 算法相比于 RSLFN 算法具有较高的训练精度,但随着噪音的增大,ELM 算法的测试精度相比于 RSLFN 算法则较低,即 RSLFN 表现出较好的抗噪音性能。可见,对于该化学机械抛光的二

维轴对称静态有限元数据集样本, RSLFN 算法同样显示出了更好的鲁棒性。

通过实验 1 和实验 2 进行的比对和分析可以发现,ELM 算法与所提出的 RSLFN 算法在真实数据集和人工数据集上的性能基本相似,对于真实数据集, RSLFN 算法在抗噪音和抗过拟合方面的优越性较人工数据集体现得更加明显。

## 5 结 论

本文针对传统的基于梯度下降和经验风险最小化准则的神经网络学习算法存在局部极值,在小样本建模时易出现过拟合以及抗噪音能力差的问题,提出了一种专门针对小样本的单隐层前馈神经网络的建模方法。该方法引入了不敏感度量和结构风险准则,针对小样本噪音数据集综合考虑了经验风险和结构风险,在抗噪音、抗过拟合方面与传统算法相比具有明显的优越性。

## 参考文献(References)

- [1] 邓万宇,郑庆华,陈琳,等.神经网络极速学习方法研究[J].计算机学报,2010,33(2):279-287.  
(Deng W Y, Zheng Q H, Chen L, et al. Research on extreme learning of neural networks[J]. Chinese J of Computers, 2010, 33(2): 279-287.)
- [2] Gori M, Tesi A. On the problem of local minima in back-propagation[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1992, 14(1): 76-86.
- [3] Tamura S, Tateishi M. Capability of four-layered neural network: Four layers versus three[J]. IEEE Trans on Neural Networks, 1997, 8(2): 251-255.
- [4] Huang G B. Learning capability and storage capacity of two-hidden-layer feed-forward networks[J]. IEEE Trans on Neural Networks, 2003, 14(2): 274-281.
- [5] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and applications[J]. Neurocomputing, 2006, 70(1/2/3): 489-501.
- [6] Vapnik V. The nature of statistical learning theory[M]. New York: Springer-Verlag, 1999: 181-223.
- [7] 杨俊燕,张优云,朱永生,等.  $\varepsilon$  不敏感损失函数支持向量机分类性能研究[J].西安交通大学学报,2007,41(11):1315-1320.  
(Yang J Y, Zhang Y Y, Zhu Y S, et al. Classification performance of support vector machine with  $\varepsilon$ -insensitive loss function[J]. J of Xi'an Jiaotong University, 2007, 41(11): 1315-1320.)
- [8] 邓赵红,王士同,鲁棒的模糊聚类神经网络[J].软件学报,2005,16(8):1415-1422.  
(Deng Z H, Wang S T. Robust fuzzy clustering neural networks[J]. J of Software, 2005, 16(8): 1415-1422.)