

doi:10.3969/j.issn.1673-4785.2011.04.005

自适应前馈神经网络结构优化设计

张昭昭^{1,2}, 乔俊飞¹, 杨刚¹

(1. 北京工业大学电子信息与控制工程学院, 北京100124; 2. 辽宁工程技术大学电子与信息工程学院, 辽宁葫芦岛125105)

摘要:针对多数前馈神经网络结构设计算法采取贪婪搜索策略而易陷入局部最优结构的问题,提出一种自适应前馈神经网络结构设计算法.该算法在网络训练过程中采取自适应寻优策略合并和分裂隐节点,达到设计最优神经网络结构的目的.在合并操作中,以互信息为准则对输出线性相关的隐节点进行合并;在分裂操作中,引入变异系数,有助于跳出局部最优网络结构.算法将合并和分裂操作之后的权值调整与网络对样本的学习过程结合,减少了网络对样本的学习次数,提高了网络的学习速度,增强了网络的泛化性能.非线性函数逼近结果表明,所提算法能得到更小的检测误差,最终网络结构紧凑.

关键词:前馈神经网络;结构设计;自适应搜索策略;互信息

中图分类号: TP273 **文献标识码:** A **文章编号:** 16734785(2011)04-0312-06

An adaptive algorithm for designing optimal feed-forward neural network architecture

ZHANG Zhaozhao^{1,2}, QIAO Junfei¹, YANG Gang¹

(1. College of Electronic and Control Engineering, Beijing University of Technology, Beijing 100124, China; 2. Institute of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, China)

Abstract: Due to the fact that most algorithms use a greedy strategy in designing artificial neural networks which are susceptible to becoming trapped at the architectural local optimal point, an adaptive algorithm for designing an optimal feed-forward neural network was proposed. During the training process of the neural network, the adaptive optimization strategy was adopted to merge and split the hidden unit to design optimal neural network architecture. In the merge operation, the hidden units were merged based on mutual information criterion. In the split operation, a mutation coefficient was introduced to help jump out of locally optimal network. The process of adjusting the connection weight after merge and split operations was combined with the process of training the neural network. Therefore, the number of training samples was reduced, the training speed was increased, and the generalization performance was improved. The results of approximating non-linear functions show that the proposed algorithm can limit testing errors and a compact neural network structure.

Keywords: feed-forward neural network; architecture design; adaptive search strategy; mutual information

前馈神经网络是应用最多的网络之一.其成功应用的关键是神经网络的结构设计.如果神经网络的规模太小,会导致欠拟合;如果神经网络规模过大,则导致过拟合.无论欠拟合还是过拟合都使神经

网络的泛化能力下降,而没有泛化能力的网络没有使用价值.由于神经网络结构在某种程度上直接决定了神经网络的最终性能,所以神经网络结构优化设计一直是神经网络领域关注的基本问题".

神经网络结构优化的方法主要有:删减方法、增长方法3、增长删减方法".删减方法是一种自顶向下的设计方法,即在网络的训练过程中,通过删除网络中冗余的节点和连接而达到简化网络结构的目的.增长方法是一种自底向上的设计方法,且增加策略比删减策略更易于制定和实现;就设计紧

收稿日期:

基金项目: 国家自然科学基金资助项目(60873043);国家"863"计划资助项目(2009AA04Z155);北京市自然科学基金资助项目(4092010);教育部博士点基金资助项目(200800050004).

通信作者: 张昭昭.E-mail: zzhao123@126.com.

凑神经网络而言,增长删减方法能够设计出较优的神经网络结构,因此,也越来越多地得到关注.目前上述几种神经网络结构设计算法大都采用预先设定的固定准则,并采用贪婪搜索策略,使得算法普适性较差且极易陷入神经网络局部最优结构^[5-11].

针对上述问题,依据互信息理论及神经网络结构与学习之间的关系,提出一种自适应前馈神经网络结构优化设计算法(adaptive merging and splitting algorithm, AMSA).该算法在网络的训练过程中,能根据当前神经网络对任务的学习情况合并或分裂隐节点,自适应调整神经网络的规模.最后通过仿真实验验证了算法的有效性与稳定性.

1 前馈神经网络

网络结构为 $n-m-p$ 的单隐层前馈神经网络输入输出之间的关系为

$$\mathbf{y} = \mathbf{W}^2 \sigma(\mathbf{W}^1 \mathbf{x} + \mathbf{B}^1) + \mathbf{B}^2.$$

式中: $\mathbf{x} \in \mathbf{R}^n$ 为神经网络的输入; $\mathbf{y} \in \mathbf{R}^p$ 为神经网络的输出; $\mathbf{W}^1 \in \mathbf{R}^{m \times n}$ 为输入层与隐层之间的连接权矩阵; $\mathbf{B}^1 \in \mathbf{R}^m$ 为隐节点的偏置权向量; $\mathbf{W}^2 \in \mathbf{R}^p$ 为隐层与输出层之间的连接权矩阵; $\mathbf{B}^2 \in \mathbf{R}^p$ 为输出节点偏置向量; $\sigma(\cdot)$ 为激活函数.

设训练样本集为 $\{x(t), y(t), t=1, 2, \dots, N\}$, 采用批学习, 则第 i 个隐节点对第 t 个训练样本的输出为 $\sigma_i(t) = \sigma(\mathbf{W}_i^1 x(t) + b_i^1)$. 定义矩阵:

$$\Phi = [\varphi_1 \varphi_2 \dots \varphi_m \varphi_{m+1}].$$

式中: $\varphi_i = [\sigma_i(1) \sigma_i(2) \dots \sigma_i(N)]^T, i=1, 2, \dots, m$; 对所有训练样本 $\varphi_{m+1}(j) = 1, j=1, 2, \dots, N$. 神经网络的输出为

$$\mathbf{Y} = [y(1) y(2) \dots y(N)]^T.$$

设 $\mathbf{W} = [\mathbf{W}^2 \mathbf{B}^2]$, 则有

$$\mathbf{Y} = \Phi \mathbf{W}^T. \quad (1)$$

2 互信息的定义与估计

2.1 互信息的定义

设随机变量 X 和 Y 的联合概率密度函数为 $\rho_{(X,Y)}(x,y)$, 则 X 和 Y 的边缘概率密度函数为^[6-7]:

$$\rho_X(x) = \int \rho_{X,Y}(x,y) dy,$$

$$\rho_Y(y) = \int \rho_{X,Y}(x,y) dx.$$

依据 Shannon 熵定义有

$$H(Y) = - \int \rho_Y(y) \log \rho_Y(y) dy,$$

$$H(X,Y) = - \int \rho_{X,Y}(x,y) \log \rho_{X,Y}(x,y) dx dy.$$

熵 $H(Y)$ 描述了随机变量 Y 包含的信息量, 联

合熵 $H(X,Y)$ 描述了随机变量 X,Y 共有的信息量.

按照信息论的有关理论, 随机变量 X,Y 之间的互信息为

$$I(X;Y) = H(X) + H(Y) - H(X,Y) =$$

$$\int \rho_{X,Y}(x,y) \log \frac{\rho_{X,Y}(x,y)}{\rho_X(x)\rho_Y(y)} dx dy.$$

互信息是随机变量 X,Y 相关性的度量, 因此当 X 和 Y 完全相关时, $I(X;Y) = 1$, 当 X 和 Y 完全独立时, $I(X;Y) = 0$.

2.2 互信息的估计

由式(2)知, 互信息的估计只需估计出联合概率分布 $\rho_{X,Y}(x,y)$ 即可, 因此可用 K-最近邻统计法(K-nearest neighbour statistics)来估计互信息.

设有 N 个输入-输出对为

$$\mathbf{Z}^{(i)} = [(\mathbf{x}^{(i)})^T (\mathbf{y}^{(i)})^T]^T.$$

式中: $i=1, 2, \dots, N, \mathbf{x}^{(i)} \in \mathbf{R}^d, \mathbf{y}^{(i)} \in \mathbf{R}^p$.

若 $\mathbf{Z}^{(i)}$ 的 K-近邻为

$$\mathbf{Z}^{(k(i))} = [(\mathbf{x}^{(k(i))})^T (\mathbf{y}^{(k(i))})^T]^T.$$

则 $\mathbf{Z}^{(i)}$ 与 $\mathbf{Z}^{(k(i))}$ 之间的 Euclidean 距离为

$$\varepsilon^{(i)} = \|\mathbf{Z}^{(i)} - \mathbf{Z}^{(k(i))}\| =$$

$$\max(\|\mathbf{x}^{(i)} - \mathbf{x}^{(k(i))}\|, \|\mathbf{y}^{(i)} - \mathbf{y}^{(k(i))}\|). \quad (3)$$

而对于 $\mathbf{Z}^{(i)}$ 中的分量 $\mathbf{x}^{(i)}$ 和 $\mathbf{y}^{(i)}$ 有

$$\varepsilon_X^{(i)} = \|\mathbf{x}^{(i)} - \mathbf{x}^{(k(i))}\| = \max_{s=1,2,\dots,d} |x_s^{(i)} - x_s^{(k(i))}|,$$

$$\varepsilon_Y^{(i)} = \|\mathbf{y}^{(i)} - \mathbf{y}^{(k(i))}\| = \max_{t=1,2,\dots,p} |y_t^{(i)} - y_t^{(k(i))}|.$$

依据式(3)有

$$\varepsilon^{(i)} = \max(\varepsilon_X^{(i)}, \varepsilon_Y^{(i)}).$$

因此, 对于 $\mathbf{Z}^{(i)}$ 中的每个 $\mathbf{x}^{(i)}$, 都可以计算出 $\mathbf{Z}^{(i)}$ 中其他 $\mathbf{x}^{(j)}$ ($j \neq i$) 的分量 $x_s^{(j)}$ 到 $\mathbf{x}^{(i)}$ 的分量 $x_s^{(i)}$ 的距离小于 $\varepsilon^{(i)}$ 的点的个数 $N_X^{(i)}$, 同理, 对于 $\mathbf{Z}^{(i)}$ 中的每个点 $\mathbf{y}^{(i)}$, 也可以计算出 $\mathbf{Z}^{(i)}$ 中其他 $\mathbf{y}^{(j)}$ ($j \neq i$) 的分量 $y_t^{(j)}$ 到 $\mathbf{y}^{(i)}$ 的分量 $y_t^{(i)}$ 的距离小于 $\varepsilon^{(i)}$ 的点的个数 $N_Y^{(i)}$, 此时 $I(X;Y)$ 估计值为

$$\hat{I}(X;Y) = - \frac{1}{N} \sum_{i=1}^N [\psi(N_X^{(i)} + 1) + \psi(N_Y^{(i)} + 1)] +$$

$$\psi(k) + (d+p-1)\psi(N) + \frac{d+p-1}{k}.$$

式中: $\psi(\cdot)$ 为 digamma 函数, $\psi(t+1) = \psi(t) + 1/t$, $\psi(1) \approx -0.577\ 215\ 6$, k 的取值一般为 $2 \sim 6$.

3 自适应前馈神经网络结构设计

3.1 隐节点的合并

既然互信息能够准确度量 2 个随机变量之间的相关程度, 由式(1)知, 前馈神经网络输出节点的输出是隐层节点输出的线性组合, 因此, 当隐节点 h_a 和 h_b 的输出线性相关时, 则这 2 个节点对信息的处

理能力是等效的,就可以对这2个节点采取合并操作,以简化神经网络结构。

采用批学习方式,学习样本数为 N 时,隐节点 h_a 的输出为

$$\varphi_{h_a} = [\sigma(1) \sigma(2) \cdots \sigma(N)]^T.$$

隐节点 h_b 的输出为

$$\varphi_{h_b} = [\sigma(1) \sigma(2) \cdots \sigma(N)]^T.$$

将 φ_{h_a} 和 φ_{h_b} 组成输入-输出对 $Z = [\varphi_{h_a}^T \varphi_{h_b}^T]^T$,则可计算出隐点 h_a 和 h_b 输出之间的互信息 $I(\varphi_{h_a}; \varphi_{h_b})$,若 $I(\varphi_{h_a}, \varphi_{h_b}) \geq \varepsilon_{MI}$ (ε_{MI} 为设定的阈值),则将 h_a 和 h_b 合并为 h_c ,合并后的隐节点 h_c 与输入层和输出层的连接权值为

$$w_{ci} = \frac{w_{ai} + w_{bi}}{2}, i = 1, 2, \cdots, n, \quad (4)$$

$$w_{jc} = w_{ja} + w_{jb}, j = 1, 2, \cdots, p. \quad (5)$$

显然,合并后的隐节点 h_c 对信息的处理能力和2个隐节点 h_a 和 h_b 对信息处理能力基本保持不变。

3.2 隐节点的分裂

在神经网络训练的过程中,当神经网络对信息的处理能力不足时,则应该在隐节点层增加节点以增大神经网络信息处理能力.判断神经网络信息处理能力不足的准则为

$$t - s \geq t_0, \quad (6)$$

$$\frac{\sum_{k=t-s}^t E_k / (s+1) E_k}{(\sum_{k=t_0-s}^{t_0} E_k) / (s+1)} \leq \Delta r. \quad (7)$$

式中: Δr 为设定的阈值, E_k 为均方差。

如图1所示,式(6)保证增加一个隐节点后神经网络至少训练 s 步.式(7)用本次神经网络训练最后 s 步的平均误差(训练步长为 t 时)和上次增加一个隐节点时(训练步长为 t_0 时)神经网络训练最后 s 步的平均误差的比值来判断是否应该增加一个隐节点,式(7)采用 s 步的平均误差,不仅能够衡量此时(训练步长为 t 时)神经网络训练误差曲线的平坦程度,而且能够避免误差曲线的震荡带来的干扰.当满足隐节点增加准则时,本文采用随机分裂现有隐层节点的方式增加神经网络隐节点数.从现有隐节点中随机选择一个隐节点 h_a ,将其裂变成2个隐节点 h_b 和 h_c .

则新隐节点 h_b 和 h_c 的连接权值为

$$w_b = (1 + \alpha) w_a, \quad (8)$$

$$w_c = -\alpha w_a. \quad (9)$$

式中: α 为变异系数,是一个很小的随机数。

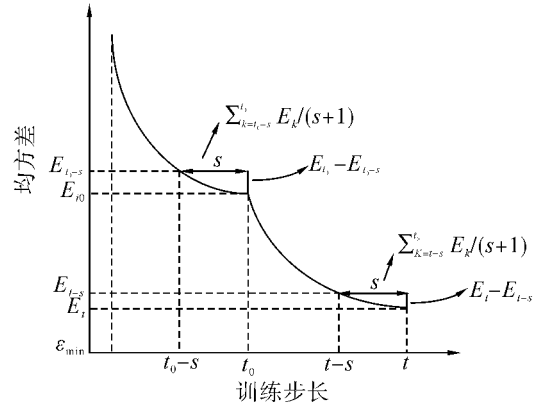


图1 误差曲线平坦度检测

Fig.1 Flatness detection of error curves

3.3 自适应前馈神经网络结构设计算法

对于前馈神经网络,有如下认识。

1)神经网络的隐节点数大于等于1。

2)神经网络进行适当的训练后,如果各隐节点的输出之间互不相关(此时没有可以合并的隐节点),则表明网络中各隐节点都在发挥各自不同的作用,处理不同的信息,即网络中没有冗余的隐节点。

3)如果神经网络中没有冗余的隐节点,那么导致神经网络对任务学习不好的原因,要么是神经网络对任务的学习不够,需要增加神经网络学习时间;要么是神经网络中隐节点太少,使得神经网络学习能力不足,需要增加隐层节点。

有上述认识,可获得自适应前馈神经网络结构设计算法如下。

1)随机产生一个初始结构神经网络,神经网络的隐节点数大于等于1,输入输出节点数由具体任务确定。

2)适当训练神经网络(训练步数大于等于 s)。

3) $E \leq \varepsilon_{\min}$ 否. 是,到7);否,到4)。

4)计算各隐节点输出之间的互信息,合并互信息接近的隐节点(由阈值 ε_{MI} 决定),并由式(4)和式(5)对合并后的新隐节点赋连接权值。

5)4)中是否有合并隐节点的操作发生,是,转到2);否转到6)。

6)由式(7)判断是否应该增加隐节点,如否,转到2);如是,则从现有隐节点中随机选择一个隐节点分裂成2个隐节点,并由式(8)和(9)分别给2个新的隐节点赋连接权值,转到2)。

7)结束。

该算法建立在对前馈神经网络学习过程正确认识的基础之上,将隐节点合并和分裂之后的网络权值调整与网络对学习样本的学习有效的结合,不仅能够减少神经网络对学习样本的学习次数,提高神

神经网络的学习速度,而且能够避免神经网络对学习样本的过学习. 隐节点分裂操作中,分裂后的新隐节点不仅继承了父节点信息处理的能力,而且引入变异系数,有助于跳出局部最优神经网络结构.

4 仿真实验

本文所提算法能够根据学习对象自适应调整前馈网络隐含层神经元个数,优化神经网络结构,得到与学习对象相适应的神经网络结构,提高了前馈神经网络的性能. 为验证该算法的有效性与稳定性,选取2个复杂度不同的非线性函数进行逼近^[12]:

$$\begin{aligned} y_1 &= x_2^3 + \sin(3x_2) + 2x_1 \sin(4x_1) + x_1 \sin(4x_2), \\ y_2 &= 1.9(1.35 + e^{x_1} \sin(13(x_1 - 0.6)^2) e^{-x_2} \sin(7x_2)). \end{aligned} \quad (10)$$

式中: x_1 和 x_2 各随机产生 30 个,服从区间 $[-1, 1]$ 内均匀分布. 训练样本对为 900 个,测试样本为 $-1:0.1:1$,即测试样本对为 441 个. 式(10)中, x_1 和 x_2 各随机产生 30 个,服从区间 $[0, 1]$ 内均匀分布,训练样本对为 900 个,测试样本对为 $0:0.05:1$,测试样本对为 441 个. 随机产生一个隐节点数大于等于 1 的初始网络结构对学习样本进行训练,网络学习算法采用带动量项 BP 算法,仿真时,选取 $k=4$,学习率为 0.000 2,动量项系数为 0.2,隐节点合并互信息阈值 $\varepsilon_{MI}=0.95$,隐节点增加均方差阈值 $\Delta r=0.01$,网络训练步数 $t=20, s=5$.

神经网络对非线性函数 y_1 的逼近效果如图 2 所示,误差曲面如图 3 所示. 对非线性函数 y_2 的逼近效果如图 4 所示,误差曲面如图 5 所示. 图 6 给出了对非线性函数 y_1 逼近时不同的初始神经网络结构在逼近过程中神经网络隐节点数变化情况. 图 7 给出了对非线性函数 y_2 逼近时不同的初始神经网络结构在逼近过程中神经网络隐节点数变化情况.

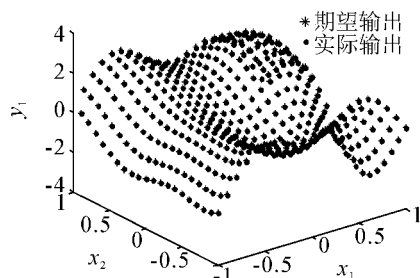


图2 对 y_1 的逼近效果

Fig.2 The approximation effect of y_1

从图 2 和图 4 可以看出,训练后的神经网络能够很好地逼近上述 2 种复杂度不同的非线性函数,神经网络输出值与函数期望值基本重合,因此该算法设计

出的神经网络不仅具有紧凑的结构而且能够保证神经网络的逼近性能. 从图 3 和图 5 的非线性函数逼近效果的误差曲面图中可以看出,对不同复杂度的非线性函数的检测误差分别小于 0.04 和 0.1,表明该算法设计的神经网络具有较强的泛化性能.

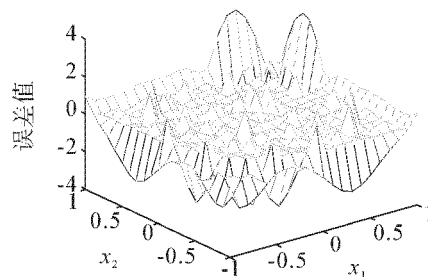


图3 对 y_1 逼近的误差曲面

Fig.3 The approximation error surface of y_1

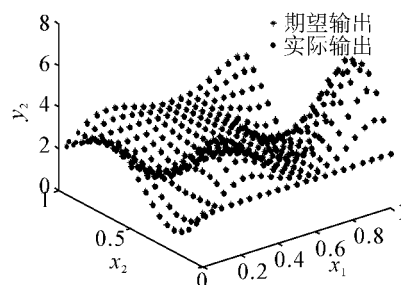


图4 对 y_2 的逼近效果

Fig.4 The approximation effect of y_2

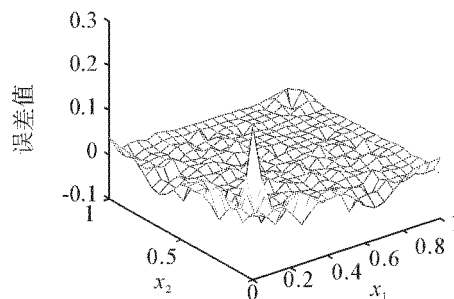


图5 对 y_2 逼近的误差曲面

Fig.5 The approximation error surface of y_2

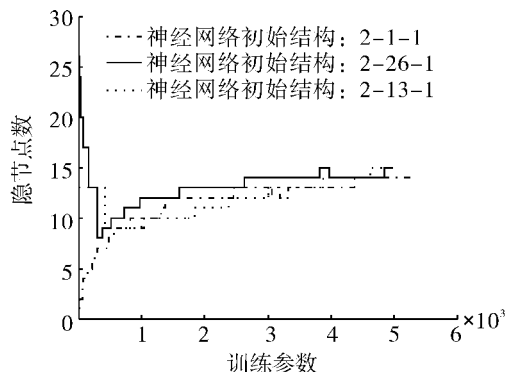
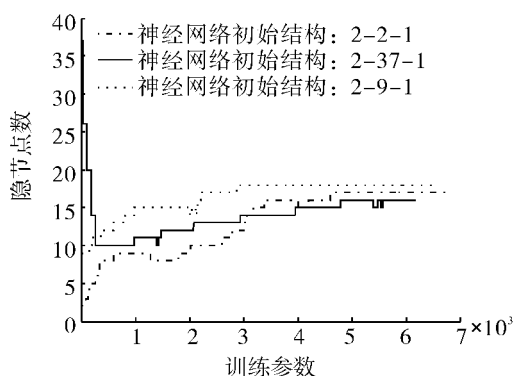


图6 逼近 y_1 时神经网络隐节点变化动态

Fig.6 The hidden node numbers during the process of approximation y_1

图7 逼近 y_2 时神经网络隐节点变化动态Fig.7 The hidden node numbers during the process of approximation y_2

从图6可以看出,用初始结构分别为2-1-1、2-13-1、2-26-1的神经网络逼近 y_1 ,训练结束时,算法所得到的神经网络的结构分别为2-14-1、2-15-1、2-15-1.从图7可以看出,用初始结构分别为2-2-1、2-9-1、2-37-1的神经网络逼近 y_2 时,算法最终所得到的神经网络结构分别是2-17-1、2-16-1、2-18-1.表明该算法对于不同的初始结构的神经网络,均能稳定收敛于适合于该学习对象的网络结构.

表1列举了本文所提AMSA算法同不同典型的神经网络结构设计算法的比较结果,分别是删减算法(optimal brain surgeon, OBS)^[9]、增长算法(cascade correlation, CC)^[10]和增长删减算法(adaptive merging and growing algorithm, AMGA)^[5],它们都是采用贪婪搜索策略设计神经网络结构的算法.上述各个算法,在相同的条件下对同样的问题运行20次,对所得最终神经网络隐层节点数取平均值.从表1可看出,AMSA算法设计出的最终神经网络,不仅结构紧凑,而且神经网络的泛化能力较强.

表1 几种典型算法性能比较

Table 1 The performance comparison of several typical algorithms

函数	算法	期望误差	检测误差	最终网络结构 (隐层节点数)
y_1	OBS	0.04	0.048 9	16.13
	CC	0.04	0.048 1	16.10
	AMGA	0.04	0.047 2	15.18
	AMSA	0.04	0.043 5	14.67
y_2	OBS	0.1	0.199 0	18.92
	CC	0.1	0.187 4	18.34
	AMGA	0.1	0.158 2	17.06
	AMSA	0.1	0.150 7	16.73

5 结 论

针对多数前馈神经网络结构设计算法采取贪婪搜索策略而易陷入局部最优结构的问题,提出了一种自适应前馈神经网络结构设计算法.该算法能够在神经网络学习过程中综合考虑神经网络对当前任务的学习能力和复杂程度,以互信息为准则合并和分裂隐节点,自适应调整神经网络结构.通过对2个复杂度不同的非线性函数逼近仿真实验,得到如下结论.

1) AMSA 算法不依赖于网络的初始结构,能够根据实际对象及当前神经网络的学习能力,自适应获得适合于实际对象的前馈神经网络.

2) AMSA 算法建立在对神经网络学习过程正确理解的基础之上,避免了神经网络结构改变时对神经网络参数的重新调整,减少了对学习样本的学习次数,提高了神经网络的学习速度,增强了神经网络的泛化性能.

参考文献:

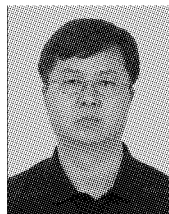
- [1] 邱健斌,王劲伯. 进化神经网络 PID 控制器的研究与应用[J]. 智能系统学报, 2008, 3(3): 245-249.
QIU Jianbin, WANG Shaobo. An improved PID controller based on an evolutionary neural network[J]. CAAI Transactions on Intelligent Systems, 2008, 3(3): 245-249.
- [2] 张昭昭,乔俊飞,韩红桂. 一种基于神经网络复杂度的修剪算法[J]. 控制与决策, 2010, 25(6): 178-182.
ZHANG Zhaozhao, QIAO Junfei, HAN Honggui. A pruning algorithm based on neural complexity[J]. Control and Decision, 2010, 25(6): 178-182.
- [3] 乔俊飞,张颖. 一种多层前馈神经网络的快速修剪算法[J]. 智能系统学报, 2008, 3(2): 173-176.
QIAO Junfei, ZHANG Ying. Fast unit pruning algorithm for multilayer feedforward network design[J]. CAAI Transactions on Intelligent Systems, 2008, 3(2): 173-176.
- [4] MA L, KHORASANI K. Constructive feedforward neural networks using Hermite polynomial activation function[J]. IEEE Transactions on Neural Network, 2005, 16(4): 821-833.
- [5] ISLAM Monirul, SATTAR A, AMIN F, YAO Xin, MURASE K. A new adaptive merging and growing algorithm for designing artificial neural networks[J]. IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics, 2009, 39(3): 705-722.
- [6] 吴晓刚,王旭东,余腾伟. 发动机输出转矩的改进 BP 神经网络估计[J]. 电机与控制学报, 2010, 14(3): 104-108.
WU Xiaogang, WANG Xudong, YU Tengwei. Estimation of

- engine output torque based on improved BP neural network [J]. *Electric Machines and Control*, 2010, 14(3): 104-108.
- [7] 宋勇,李贻斌,李彩虹. 递归神经网络的进化机器人路径规划方法[J]. *哈尔滨工程大学学报*, 2009, 30(8): 898-902.
- SONG Yong, LI Yibin, LI Caihong. Path planning based on a recurrent neural network for an evolutionary robot [J]. *Journal of Harbin Engineering University*, 2009, 30(8): 898-902.
- [8] 陆瑶,张杰,冯英浚. 非线性动态系统的模糊神经网络自适应 H_∞ 鲁棒控制[J]. *哈尔滨工程大学学报*, 2009, 30(9): 1082-1086.
- LU Yao, ZHANG Jie, FENG Yingjun. H_∞ robust adaptive control of a fuzzy neural network based nonlinear dynamic system [J]. *Journal of Harbin Engineering University*, 2009, 30(9): 1082-1086.
- [9] 罗耀华,从静. 基于 BP 神经网络的三相逆变器故障诊断研究[J]. *应用科技*, 2010, 37(6): 56-60.
- LUO Yaohua, CONG Jing. Fault diagnosis of three-phase inverter using BP neural network [J]. *Applied Science and Technology*, 2010, 37(6): 56-60.
- [10] KRASKOV A, STOGBAUER H, GRASSBERGER P. Estimating mutual information [J]. *Phys Rev E, Stat Plasmas Fluids Relat Interdiscip Top*, 2004, 69(0661138): 1-16.
- [11] HONG Jie, HU Baogang. Two-phase construction of multi-layer perceptions using information theory [J]. *IEEE Transactions on Neural Network*, 2009, 20(4): 542-550.
- [12] LIU Yinyin, STARZYK J A, ZHU Zhen. Optimized approximation algorithm in neural networks without overfitting [J]. *IEEE Transactions on Neural Network*, 2008, 19(6): 983-995.
- [13] HASSIBI B, STORK D, WOLFF G, WATANABE T. Optimal brain surgeon: extensions and performance comparisons [C]//*Advances in Neural Information Processing Systems 6*. San Mateo, USA: Morgan Kaufman, 1994: 263-270.
- [14] FAHLMAN S E, LEBIERE C. The cascade correlation learning architecture [C]//*Advances in Neural Information Processing Systems 2*. San Mateo, USA: Morgan Kaufman, 1990: 524-532.

作者简介:



张昭昭,男,1973年生,博士研究生,主要研究方向为智能系统与智能信息处理、神经网络结构设计与优化。



乔俊飞,男,1968年生,教授,博士生导师,主要研究方向为复杂过程建模与控制、计算智能与智能优化控制,发表学术论文100余篇,其中被SCI、EI检索60余篇。



杨刚,男,1983年生,博士研究生,主要研究方向为神经计算与智能优化控制。