一种动态筛选样本的前向神经网络快速学习算法

盛守照 王道波 黄向华
(南京航空航天大学自动化学院 南京 210016)

摘 要： 从理论上讨论了一类隐含层激励函数满足Mercer条件的前向神经网络学习问题，分析了提高网络学习速度的途径，提出了一种动态筛选样本的前向神经网络快速学习算法。它大大提高了网络学习速度，克服了传统的基于梯度下降的网络学习方法存在的诸多弊端。算法还具有动态确定隐含层神经元数的自构性优点。文中通过具体数值试验验证了上述算法的可行性和优越性。

关键词： 前向神经网络，算法，机器学习

中图分类号： TP183 **文献标识码：** A **文章编号：** 1009-5896(2005)11-1818-03

A Fast Learning Algorithm of Feedforward Neural Networks Based on Screening Samples Dynamically

Sheng Shou-zhao Wang Dao-bo Huang Xiang-hua
(*College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing* 210016, *China*)

Abstract The learning issue of feedforward neural networks whose activation function of hidden neurons satisfies Mercer condition is discussed in theory. The approach to improving learning speed is investigated. Then a fast learning algorithm of feedforward neural networks based on screening samples dynamically is proposed, which improves learning speed, solves the abuses of those learning algorithm based on gradient decent method and has the self-configuring advantage by determining the number of hidden neuron dynamically. The reliability and advantage of the proposed algorithm are illustrated concretely through test.

Key words Feedforward neural network, Algorithm, Machine learning

1 引言

前向神经网络学习归根结底就是利用已知样本数据拟合未知的系统模型。迄今，诸多学者提出了许多有益的学习方法，大多数为了提高网络学习速度的算法[1－7]本质上仍然是一种梯度算法，它不能完全克服易陷入局部最小点问题，也无法回答如何根据给定的激励函数确定隐含层神经元数等问题，从而降低了网络性能指标和学习效率。另外，基于梯度下降的网络学习算法在学习新样本时有遗忘已学样本的趋势，导致学习迭代次数变多，大大降低了网络学习速度，严重影响其推广与应用[7]。究其根源，它们忽视了蕴涵于网络结构中的数学本质，不能从理论上明确回答算法能否同时提高网络性能指标和学习速度。本文针对上述问题，从理论上讨论了一类隐含层激励函数满足Mercer条件[8,9]的前向神经网络学习问题，分析了提高网络学习速度的途径，在此基础上，提出了一种动态筛选样本的前向神经网络快速学习算法。它不仅大大提高了网络学习速度，而且克服了基于梯度下降的网络学习方法存在的以上诸多弊端。另外，它还具有动态确定隐含层神经元数的自构性优点，这是以往大多数算法所无法比拟的。文中最后通过具体数值试验验证了上述算法的可行性和优越性。

2 前向神经网络学习问题数学分析

2.1 问题描述

设 $X\left(\subseteq \Re^p\right)$，$W\left(\subseteq \Re^p\right)$，$\Phi\left(\subseteq \Re^q\right)$，$Y\left(\subseteq \Re\right)$ 分别为网络输入空间、隐含层权值向量空间、隐含层输出空间和网络输出域，p,q 分别为网络输入向量维数和隐含层神经元数，$\varphi\left(\boldsymbol{x},\boldsymbol{\omega}\right)$ 为隐含层激励函数，且满足Mercer条件，其中，$\boldsymbol{x}\in X,\boldsymbol{\omega}\in W$ 分别为网络输入向量和隐含层权值向量，输出

2004-06-04收到，2004-09-09改回
航空基金(01C52015)资助课题

层为线性激励函数； $Z_n=\left\{z_i=\left(\boldsymbol{x}_i,y_i\right)\middle|\boldsymbol{x}_i\in X,y_i\in Y\right\}_{i=1}^{n}$ 为样本集，n 为样本容量；通过网络学习寻找合适的隐含层权值向量集 $\varpi=\{\boldsymbol{\omega}_i\}_{i=1}^{q}(\subseteq W)$ 和输出层权值集 $\alpha=\{a_i\}_{i=1}^{q}(\subseteq\Re)$，使得网络输出：

$$f(\boldsymbol{x},\varpi,\alpha)=\sum_{i=1}^{q}a_i\varphi(\boldsymbol{x},\boldsymbol{\omega}_i) \tag{1}$$

最小化下式定义的性能指标：

$$J(\varpi,\alpha)=\frac{1}{n}\sum_{i=1}^{n}\left(y_i-f(\boldsymbol{x}_i,\varpi,\alpha)\right)^2 \tag{2}$$

2.2 数学分析

设 ψ 为空间 $\Re^p$ 到高维Hilbert空间 H 的非线性映射，且使得 $\varphi(\boldsymbol{x},\boldsymbol{\omega})=\langle\psi(\boldsymbol{x}),\psi(\boldsymbol{\omega})\rangle$，其中 $\langle\cdot,\cdot\rangle$ 为空间 H 中的向量内积运算[8,9]，那么隐含层输出可表示为

$$\phi(\boldsymbol{x},\varpi)=\left[\varphi(\boldsymbol{x},\boldsymbol{\omega}_1),\cdots,\varphi(\boldsymbol{x},\boldsymbol{\omega}_q)\right]^{\mathrm{T}}=\boldsymbol{\Gamma}(\varpi)\psi(\boldsymbol{x}) \tag{3}$$

式中，$\boldsymbol{\Gamma}(\varpi)=\left[\psi(\boldsymbol{\omega}_1),\cdots,\psi(\boldsymbol{\omega}_q)\right]^{\mathrm{T}}$ 为空间 H 到空间 $\boldsymbol{\Phi}$ 的线性映射，由此把网络输出改写成

$$f(\boldsymbol{x},\varpi,\alpha)=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Gamma}(\varpi)\psi(\boldsymbol{x}) \tag{4}$$

式中 $\boldsymbol{a}=\left[a_1,a_2,\cdots,a_q\right]^{\mathrm{T}}$ 为输出层权值向量，结合线性代数知识，不难得到:

定理 1　设 $X_n=\{\boldsymbol{x}_i\}_{i=1}^{n}$ 为 Z_n 的输入部分，形如 $H(X')$ $(\subseteq H)$ 为 $\boldsymbol{x}\in X'(\subseteq\Re^p)$ 在映射 ψ 下的像空间，若 $H(\varpi)\supseteq H(X_n)$，则 $\min\limits_{\alpha\subseteq\Re}J(\varpi,\alpha)=\min\limits_{\substack{\varpi'\subseteq\Re^p\\ \alpha'\subseteq\Re}}J(\varpi',\alpha')$。

推论 1　给定样本集 Z_n，存在 $\varpi\in X_n,\alpha\in\Re$ 使得 $J(\varpi,\alpha)$ 最小。

定理 2　设 $\varpi'\in W$，存在满秩阵 $\boldsymbol{K}$，满足 $\boldsymbol{\Gamma}(\varpi')=\boldsymbol{K}\cdot\boldsymbol{\Gamma}(\varpi)$，则 $\min\limits_{\alpha'\subseteq\Re}J(\varpi',\alpha')=\min\limits_{\alpha\subseteq\Re}J(\varpi,\alpha)$。

推论 1　为快速选择 ϖ 提供了理论指导方向，另外，为了降低网络的复杂性，从定理 2 可知 $\boldsymbol{\Gamma}(\varpi)$ 应为行满秩阵，即 ϖ 应满足约束条件:

$$\operatorname{rank}(\boldsymbol{\Phi}(\varpi))=q \tag{5}$$

式中，$\boldsymbol{\Phi}(\varpi)=\left[\phi(\boldsymbol{\omega}_1,\varpi),\phi(\boldsymbol{\omega}_2,\varpi),\cdots,\phi(\boldsymbol{\omega}_q,\varpi)\right]^{\mathrm{T}}$，由此最优输出层权值向量：

$$\boldsymbol{a}=\left(\boldsymbol{B}^{\mathrm{T}}(\varpi)\boldsymbol{B}(\varpi)\right)^{-1}\boldsymbol{B}^{\mathrm{T}}(\varpi)\boldsymbol{y} \tag{6}$$

式中，$\boldsymbol{B}(\varpi)=\left[\phi(\boldsymbol{x}_1,\varpi),\cdots,\phi(\boldsymbol{x}_n,\varpi)\right]^{\mathrm{T}},\boldsymbol{y}=\left[y_1,\cdots,y_n\right]^{\mathrm{T}}$，因此隐含层权值向量集 ϖ 是网络学习的关键。

3　动态筛选样本的前向神经网络快速学习算法

值得指出的是，直接从 X_n 中遍历搜索满足性能指标要求的集合 ϖ 所需计算时间随样本容量 n 呈指数趋势爆炸性增长，这意味着为了进一步提高网络学习速度还需要筛选样本集 Z_n。

把样本集 Z_n 表示为学习样本集 $Z_{n_l}^{l}=\left\{z_i^l=\left(\boldsymbol{x}_i^l,y_i^l\right)\right\}_{i=1}^{n_l}$ 和备选样本集 $Z_{n_c}^{c}=\left\{z_i^c=\left(\boldsymbol{x}_i^c,y_i^c\right)\right\}_{i=1}^{n_c}$ 两部分，满足 $Z_{n_l}^{l}\cup Z_{n_c}^{c}=Z_n,Z_{n_l}^{l}\cap Z_{n_c}^{c}=\varnothing$，样本筛选的目的就是从 Z_n 中寻找 $Z_{n_l}^{l},\le$ 使得基于上述理论利用 $Z_{n_l}^{l}$ 学习得到的网络满足 $J(\varpi,\alpha)\le\delta_0$，在这里，$\delta_0\in\left(0,\min\limits_{\substack{\varpi\in X_n\\ \alpha\in\Re}}J(\varpi,\alpha)\right)$ 为预先确定的性能指标上限。

设 $E^c=\left\{e_i^c\middle|e_i^c=\left|y_i^c-f\left(\boldsymbol{x}_i^c,\varpi,\alpha\right)\right|\right\}_{i=1}^{n_c}$ 为 $Z_{n_c}^{c}$ 中样本输出与网络输出之差的绝对值组成的集合，$z_s^c=\left(\boldsymbol{x}_s^c,y_s^c\right)$ 为 $Z_{n_c}^{c}$ 中对应 E^c 中最大元素的任意元素，若初始化样本集 $Z_{n_l}^{l}=\varnothing,Z_{n_c}^{c}=Z_n$ 和网络结构参数 $\varpi=\varnothing,\alpha=\varnothing$，采用如下学习规则:

规则 1　$Z_{n_l}^{l}=Z_{n_l}^{l}\cup\left\{z_s^c\right\}$；

规则 2　$Z_{n_c}^{c}=Z_{n_c}^{c}-\left\{z_s^c\right\}$；

规则 3　$\varpi'=\varpi+\left\{\boldsymbol{x}_s^c\right\}$，若 ϖ' 满足式(5)约束条件 ϖ=ϖ'；

规则 4　利用 $Z_{n_l}^{l}$ 和 ϖ 按式(6)计算 α；更新 $Z_{n_l}^{l},Z_{n_c}^{c}$ 和 ϖ,α，那么若干次迭代后必有 $J(\varpi,\alpha)\le\delta_0$。上述学习规则的核心思想就是从备选样本集 $Z_{n_c}^{c}$ 中逐步筛选出当前误差最大的样本 z_s^c 加入学习样本集 $Z_{n_l}^{l}$，渐近提高由 $Z_{n_l}^{l}$ 所确定的网络包含 Z_n 的信息量，在此过程中，按照上节理论指导，尽可能把 $\boldsymbol{x}_s$ 加入 ϖ，从而一并解决了样本筛选与 ϖ 选择的问题，大大提高了网络学习效率。据此给出整个算法如下：

(1) 初始化 $Z_{n_l}^{l}=\varnothing,Z_{n_c}^{c}=Z_n$ 和 $\varpi=\varnothing,\alpha=\varnothing$，要求 $J(\varpi,\alpha)\le\delta_0$；

(2)计算 E^c 和 $J(\varpi,\alpha)$；

(3)若 $J(\varpi,\alpha)\le\delta_0$，转到第(8)步，否则继续；

(4)从 $Z_{n_c}^{c}$ 中选择 z_s^c，置 $Z_{n_l}^{l}$=$Z_{n_l}^{l}\cup\left\{z_s^c\right\}$，$Z_{n_c}^{c}=Z_{n_c}^{c}-\left\{z_s^c\right\}$；

(5)设 $\varpi'=\varpi+\left\{\boldsymbol{x}_s\right\}$，若 $\operatorname{cond}\left(\Phi(\varpi')\right)\le c_0$ (大于 0 的常数)，置 $\varpi=\varpi'$；

(6)利用 $Z_{n_l}^{l}$ 和 ϖ 按式(6)计算 α；

(7)若 $Z_{n_c}^{c}$ 非空，转到第(2)步，否则继续；

(8)利用 Z_n 和 ϖ 按式(6)重新计算 α，算法结束。算法第

(5)步利用对 $\Phi(\varpi)$ 条件数的约束替代对式(5)的约束有利于降低计算 α 带来的量化误差；第(8)步利用 Z_n 替代 $Z_{n_l}^l$ 重新计算 α，有益于提高 α 的置信度。算法具有在迭代过程中动态确定隐含层神经元数的自构性特点，这有益于提高网络性能指标和学习效率，一般地，n_l 远小于 n，算法大部分时间集中于第(2)步计算 $J(\varpi,\alpha)$，而 E_c 可从计算 $J(\varpi,\alpha)$ 过程附带得到，设 $t(J)$ 为计算 $J(\varpi,\alpha)$ 所需的平均时间，它与样本容量 n 近似呈线性关系，整个网络所需学习时间 $t(N)\approx n_l\cdot t(J)$。

4 数值试验

设观测系统 $\begin{cases} c=\exp\left(-0.05\boldsymbol{x}^2\right)\cdot\sin\left(2\pi\boldsymbol{x}\right), \\ t=8\pi\left(\boldsymbol{x}-1\right), \\ y=c\cdot\exp\left(-0.05t^2\right)\cdot\cos t, \end{cases}$ $X=[0,2]$；选择

$\varphi(\boldsymbol{x},\boldsymbol{\omega})=\exp\left(-\dfrac{(\boldsymbol{x}-\boldsymbol{\omega})^2}{\sigma^2}\right)$， $\sigma^2=0.0045$。在上述系统中均匀采样若干组不同容量的样本集，取 $\delta_0=0.00001$，利用本文算法与 Matlab 6.1 工具箱提供的学习算法进行对比试验。

表 1 列出不同样本容量下两种算法的计算时间(仿真软件为 Matlab 6.1，操作系统 Windows2000，CPU 主频 666MHz)，图 1 给出 $n=500$ 时，利用本文算法得到的仿真的结果。结果表明，相比于传统算法，本文提出的算法显著提高了网络学习速度，所需计算时间与样本容量近似呈线性关系，同时在迭代过程中具有很好的一致收敛性，多次重复试验表明，当指标要求较高的情况下，后续选择的次要隐含层权值向量作用相对较小，因此，实际应用中适当放宽指标要求，可以有效地降低网络的复杂性。

表 1 两种算法计算时间对比

n	本文算法(s)	传统算法时间(s)
50	0.1146	25.6970
100	0.1803	31.2450
200	0.3488	42.6910
400	0.6312	67.4770
500	0.7871	94.1960
1000	1.4124	328.0031
2000	2.6976	1233.9054

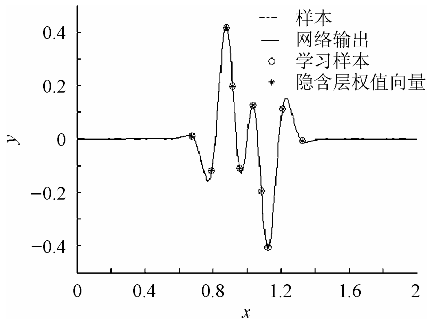

图 1 本文算法仿真结果示例($n=500$)

5 结束语

文中提出的动态筛选样本的前向神经网络快速学习算法，从本质上提高了网络学习速度，克服了以往的基于梯度下降的网络学习方法存在的诸多弊端，另外，它还具有动态确定隐含层神经元数的自构性优点，这是以往大多数算法所无法比拟的，上述数值试验验证了算法的可行性和优越性。

参 考 文 献

[1] Vitaly Schetinin. A learning algorithm for evolving cascade neural networks. *Neural Processing Letters*, 2003, 17(1): 21 − 31.

[2] Chng E S, Chen S, Mulgrew B. Gradient radial basis function networks for nonlinear and nonstationary time series prediction. *IEEE Trans. on Neural Networks*, 1996, 7(1): 190 − 194.

[3] Engozinger S, Tomsen E. An accelerated learning algorithm for multiplayer perceptions: Optimization layer by layer. *IEEE Trans. on Neural Networks*, 1995, 6(1): 31 − 42.

[4] Scalero S, Tepedelenlioglu N. A fast new algorithm for training feedforward networks. *IEEE Trans. on Signal Processing*, 1992, 40(1): 202 − 210.

[5] 叶军，张新华. 多层前向神经网络的快速学习算法及其应用. 控制与决策, 2002, 17(suppl.): 817 − 819.

[6] 刘铁男，段玉波，陈广义等. 多层前向神经网络的新型二阶学习算法. 控制理论与应用, 2000, 17(5): 721 − 724.

[7] 陈亚军. 一种多层前馈神经网络的快速学习算法. 河北师范大学学报(自然科学版), 2002, 26(6): 582 − 587.

[8] Vapnik V. The Nature of Statistical Learning Theory. New York: Springer Verlag, 1995: 1 − 20.

[9] 张铃. 基于核函数的 SVM 机与三层前向神经网络的关系. 计算机学报, 2002, 25(7): 696 − 700.

盛守照： 男，1977年生，博士生，研究方向为机器学习、智能控制、先进无人机飞行控制等.

王道波： 男，1957 年生，教授，博士生导师，研究方向为先进无人机飞行控制、机电模拟等.