

一种广义小波神经网络的结构及其优化方法

丁 勇¹, 刘守生^{1,2}, 胡寿松¹

(1. 南京航空航天大学 自动化学院, 江苏 南京 210016; 2. 解放军理工大学 理学院, 南京 210016)

摘要: 从理论上分析了小波神经网络节点过多及鲁棒性差的原因, 基于主成份分析 (PCA) 的思想提出了一种规模小、抗干扰性强的广义小波神经网络 (EWNN) 及其优化方法. 仿真结果表明, 用该方法设计的广义小波神经网络, 其非线性逼近能力及稳定性都明显优于普通小波神经网络.

关键词: 小波框架; 主成份分析; 广义小波神经网络

中图分类号: TP18 **文献标识码:** A

Extended wavelet neural network structure and its optimal method

DING Yong¹, LIU Shou-sheng^{1,2}, HU Shou-song¹

(1. College of Automation, Nanjing University of Aeronautics and Astronautics, Jiangsu Nanjing 210016, China;

2. College of Science, PLA University of Science and Technology, Jiangsu Nanjing 210016, China)

Abstract: The problems for wavelet neural network with large scale of nodes and poor robustness are analyzed. An extended wavelet neural network (EWNN) and its optimal method, which has fewer nodes and strong obstructive resistance, are designed on the basis of the principal component analysis (PCA). The simulation results show that EWNN is superior in nonlinear approximation and stability to the average wavelet network.

Key words: wavelet frame; PCA; EWNN

1 引言 (Introduction)

小波神经网络是 20 世纪 90 年代初结合小波分析理论和神经网络两者的优点而提出的一种前馈型神经网络, 它在非线性函数逼近方面所具有的最佳逼近和全面逼近能力正日益受到人们的重视^[1,2]. 对小波神经网络拓扑结构的设计、网络的学习算法以及鲁棒性的研究也不断有成果出现, 但其结构规模大和鲁棒性差的问题还未得到很好的解决.

小波神经网络可以看作是用小波函数代替神经网络中的激励函数而形成的一种神经网络. 基于连续小波函数的小波神经网络 (CWNN), 其学习算法同一般前馈神经网络 (如 BP 网络) 的算法一样, 可通过对教师值的学习确定出各基元函数的伸缩和平移参数及相应的权值. 虽然利用小波分析理论可指导网络的初始化及学习过程, 但诸如网络拓扑结构的确定、类似 BP 网的非线性优化所固有的缺陷等问题仍严重影响着它的应用.

正交小波网络 (OWNN) 是建立在多分辨分析基

础上的一种网络, 其内容丰富, 理论研究也易于深入. 但由于正交基构造复杂, 除个别正交小波 (如 haar 小波) 外, 大多不能用显式表示. 另外, Daubechies^[3] 已证明: 具紧子集的正交基不可能具有任何对称性. 因此, 用之作为滤波器就不可能有线性相位 (甚至不可能有广义线性相位), 从而会产生失真. 这些不足极大地限制了 OWN 在工程上的应用^[4].

以小波框架作为基函数的离散小波神经网络 (DWNN), 则避免了以上两种网络的缺点. 由于不考虑正交性, 基小波的选取有很大的自由度, 这为网络的实际应用奠定了基础. 根据信号及基函数的时-频域特性可很快确定出伸缩参数和平移参数的取值范围, 只有权系数成为唯一待确定的参数, 网络的学习算法从非线性优化问题变为线性优化问题, 有很好的应用价值. 本文从理论上对 DWNN 尚存在的问题进行了讨论, 通过主成份分析, 对网络的结构进行了改造, 提出了一种规模小且抗干扰性强的广义小波神经网络.

2 离散小波神经网络模型分析 (Model analysis of discrete wavelet neural network)

2.1 小波框架 (Wavelet frame)

设小波函数 $\varphi(x) \in L^1(R) \cap L^2(R)$, 满足允许条件

$$\int \frac{|\hat{\varphi}(\omega)|^2}{|\omega|} d\omega < \infty.$$

其中 $\hat{\varphi}(\omega)$ 为 $\varphi(x)$ 的 Fourier 变换. 对 $\varphi(x)$ 伸缩、平移变换得到离散小波函数族 $\{\varphi_{m,n}(x)\}$, 其中

$$\begin{aligned} \varphi_{m,n}(x) &= a^{-\frac{m}{2}} \varphi(a^{-m}x - nb), \\ (m,n) &\in \mathbb{Z}^2, a > 0, b > 0. \end{aligned} \quad (1)$$

定义 1 若离散小波函数族 $\{\varphi_{m,n}(x)\}$ 在 $L^2(R)$ 中满足以下框架条件

$$\begin{aligned} A \|f\|^2 &\leq \sum_{(m,n) \in \mathbb{Z}^2} |\langle \varphi_{m,n}, f \rangle|^2 \leq B \|f\|^2, \\ f &\in L^2(R), \end{aligned} \quad (2)$$

则称 $\{\varphi_{m,n}\}$ 是一个框架, A, B 称为框架界. 若 $A = B$, 则称框架为紧框架. 特别地, 当 $A = B = 1$ 且 $\|\varphi_{m,n}\| \equiv 1$ 时, $\{\varphi_{m,n}(x)\}$ 成为 $L^2(R)$ 的标准正交基.

对于 $\|\varphi_{m,n}\| \equiv 1$ 的紧框架, A 的大小决定了紧框架 $\{\varphi_{m,n}(x)\}$ 中函数之间的相关程度, 即它们所含信息的冗余程度, 此时又称 A 为框架的多余比. A 越大, 表明 $\varphi_{m,n}(x)$ 之间的相关性越高, 冗余信息也就越多. 从直观上看, 能用较少的正交基函数表示的信号, 如用相关性高的框架表示时, 框架所含函数的个数相对正交基要多得多. 因此, 以 $\varphi_{m,n}(x)$ 为基元函数的 DWNN 的结构就会过于庞大, 用统计的方法还可证明这种 DWNN 的鲁棒性也较差.

2.2 DWNN 的性能分析 (Performance analysis of DWNN)

以单输入单输出三层 DWNN 为例. 设 $y = f(x) \in L^2(R)$ 为非线性函数, 训练样本为 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. 由小波分析理论知, 函数 $y = f(x)$ 可表示为 $y(x) = \sum_{(m,n) \in \mathbb{Z}^2} w_{m,n} \varphi_{m,n}(x)$. Daubechies^[3]

已经证明对于时-频域有限的函数 $f(x) \in L^2(R)$, 存在有限个小波基函数, 对 f 可进行任意精度的逼近. 利用时-频域“分解-综合”方法^[5] 从 $\{\varphi_{m,n}(x)\}$ 中选择基函数 (重新排序): $\varphi_1(x), \varphi_2(x), \dots, \varphi_N(x)$ 作为隐层节点基函数构造 DWNN, 其结构见图 1, 则网络方程为

$$\hat{y} = \sum_{i=1}^N w_i \varphi_i(x). \quad (3)$$

将样本值代入式(3), 考虑到误差因素, 有

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \varphi_1(x_1) & \varphi_2(x_1) & \cdots & \varphi_N(x_1) \\ \varphi_1(x_2) & \varphi_2(x_2) & \cdots & \varphi_N(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ \varphi_1(x_n) & \varphi_2(x_n) & \cdots & \varphi_N(x_n) \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{pmatrix} + \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix}. \quad (4)$$

其中 $\xi_i (i = 1, 2, \dots, n)$ 为误差项, 满足以下假设:
 ① $E\xi_i = 0$; ② $\text{var}(\xi_i) = \sigma^2$; ③ $\text{cov}(\xi_i, \xi_j) = 0, i \neq j$.

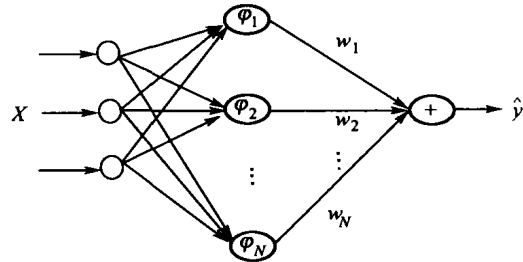


图 1 离散小波神经网络结构
 Fig. 1 Structure of discrete wavelet neural network

为了便于研究, 将式(4)中心化处理, 令

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \bar{\varphi}_i = \frac{1}{n} \sum_{k=1}^n \varphi_i(x_k) \quad (i = 1, 2, \dots, N).$$

记

$$Y = (y_1 - \bar{y}, y_2 - \bar{y}, \dots, y_n - \bar{y})^T,$$

$$\xi = (\xi_1, \xi_2, \dots, \xi_n)^T,$$

$$\varphi_i = (\varphi_i(x_1) - \bar{\varphi}_i, \varphi_i(x_2) - \bar{\varphi}_i, \dots, \varphi_i(x_n) - \bar{\varphi}_i)^T, \quad i = 1, 2, \dots, N,$$

$$\varphi = (\varphi_1, \varphi_2, \dots, \varphi_N), \quad W = (w_1, w_2, \dots, w_N)^T,$$

则式(4)可写为

$$Y = \varphi W + \xi. \quad (5)$$

式(5)即为所要建立的线性回归模型, 其中 W 为所要确定的网络权系数. 通常, 为了充分提取原信号的信息, 所取样本点的个数应充分多, 一般有 $n \gg N$. 由最小二乘法得 W 的估计值为

$$\hat{W} = (\varphi^T \varphi)^{-1} \varphi^T Y. \quad (6)$$

由 Gauss-Markov 定理知, 在 W 的所有无偏估计中, 由式(6)决定的 \hat{W} 是唯一具有最小均方误差 $\text{MSE}(\hat{W})$ 的估计, 这里 $\text{MSE}(\hat{W}) = E \|\hat{W} - W\|^2$ 是评价 \hat{W} 优劣的标准. 但是, 在应用式(6)的过程中, 有时某些回归系数估计值的绝对值异常大, 有时回归系数的估计值符号与问题的实际意义相违背, 有时稍加噪声, 回归系数就会发生较大振荡, 即模型表现出很大的不稳定性, 产生这些现象的主要原因就是回归自变量之间存在着近似线性关系, 即复共线性 (multicollinearity), 这种复共线性的存在, 使得所估计参数 \hat{W} 的 $\text{MSE}(\hat{W})$ 可充分大.

引理 在式(5)中, 若矩阵 φ 的秩 $\text{rank} \varphi = N$, 则 $\varphi^T \varphi$ 为正定阵.

证 因 $\varphi^T \varphi$ 为对称阵, 构造二次型 $X^T \varphi^T \varphi X$, 则有 $X^T \varphi^T \varphi X = (\varphi X)^T (\varphi X) = \|\varphi X\|^2$. 由于 $\text{rank } \varphi = N$, 则方程组 $\varphi X = 0$ 无非零解, 所以当 $X \neq 0$ 时, 有 $X^T \varphi^T \varphi X = \|\varphi X\|^2 > 0$, 所以 $\varphi^T \varphi$ 为正定阵.

定理 1 在式(6)中, 若矩阵 φ 的秩 $\text{rank } \varphi = N$, 则 $MSE(\hat{W}) \geq \sigma^2 N / \sqrt{|\varphi^T \varphi|}$.

证 因 $\text{rank } \varphi = N$, 由引理知, $\varphi^T \varphi > 0$, 所以 $\varphi^T \varphi$ 存在 N 个正特征值, 设其从大到小依次排列为: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N > 0$, 根据最小二乘法的性质得:

$$MSE(\hat{W}) = \sigma^2 \text{tr}(\varphi^T \varphi)^{-1}.$$

因为 $\lambda_i (i = 1, 2, \dots, N)$ 是 $\varphi^T \varphi$ 的特征值, 所以 λ_i^{-1} 是 $(\varphi^T \varphi)^{-1}$ 的特征值, $i = 1, 2, \dots, N$. 由韦达定理得:

$$\sum_{i=1}^N \lambda_i^{-1} = \text{tr}(\varphi^T \varphi)^{-1}, \quad \prod_{i=1}^N \lambda_i^{-1} = |\varphi^T \varphi|^{-1},$$

则

$$MSE(\hat{W}) =$$

$$\sigma^2 \left(\sum_{i=1}^N \lambda_i^{-1} \right) \geq \sigma^2 N \sqrt{\prod_{i=1}^N \lambda_i^{-1}} = \sigma^2 N / \sqrt{|\varphi^T \varphi|}.$$

由于 $\varphi_1(x), \varphi_2(x), \dots, \varphi_N(x)$ 为 $L^2(R)$ 上的离散小波函数, 一般地有 $\text{rank } \varphi = N$, 但框架的非正交性又决定了它们之间存在冗余性, 这种冗余性表现为 $\varphi_1(x), \varphi_2(x), \dots, \varphi_N(x)$ 存在着近似的线性关系. 当冗余比 A 越大, 这种近似的程度就越高, $|\varphi^T \varphi|$ 的值也就越接近于零, 由定理 1 知, $MSE(\hat{W})$ 也就越大, 导致模型的鲁棒性差.

3 广义小波神经网络结构的设计 (Design of EWNN)

本文在保证逼近精度符合要求的前提下提出了一种既能极大缩小网络规模又能大幅度提高网络抗干扰性的广义小波神经网络 EWNN.

3.1 主成份分析方法 (PCA method)

定义 2 设 $\varphi^T \varphi$ 的 N 个特征值为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N > 0$, 其对应的 N 个单位正交特征向量 $\alpha_i = (a_{i1}, a_{i2}, \dots, a_{iN})^T, i = 1, \dots, N$, 则称 $g_1 = \varphi \alpha_1 = a_{11} \varphi_1 + a_{12} \varphi_2 + \dots + a_{1N} \varphi_N$ 为第一主成份, 依次称 $g_i = \varphi \alpha_i = a_{i1} \varphi_1 + a_{i2} \varphi_2 + \dots + a_{iN} \varphi_N (i = 1, 2, \dots, N)$ 为第 i 个主成份, 其中 $\varphi = (\varphi_1, \varphi_2, \dots, \varphi_N)$.

若记 $g = (g_1, g_2, \dots, g_N), \alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$, 则 $g = \varphi \alpha$, 于是有

$$g^T g = \alpha^T (\varphi^T \varphi) \alpha = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N),$$

即

$$g_i^T g_j = \begin{cases} \lambda_i, & j = i, \\ 0, & j \neq i, \end{cases} \quad i, j = 1, 2, \dots, N. \quad (7)$$

记 $\lambda = \sum_{k=1}^N \lambda_k$, 称 $\gamma_i = \lambda_i / \lambda (i = 1, 2, \dots, N)$ 为第 i

个主成份的贡献率, 表示所占总信息量的比率, 则前 m 个主成份的累计贡献率为 $\frac{1}{\lambda} \sum_{i=1}^m \lambda_i$. 给定精度 ϵ , 计算

$$M = \min \{ m : 1 - \frac{1}{\lambda} \sum_{i=1}^m \lambda_i < \epsilon \}, \quad (8)$$

则利用前 M 个主成份 g_1, g_2, \dots, g_M 逼近原信号的精度即可达到要求.

3.2 EWNN 的结构设计 (Structure design of EWNN)

通过以上对离散小波神经网络结构的理论分析及主成份分析所具有的信息集中作用, 我们可设计一种以前个主成份为基元函数的新型神经网络. 记

$$\varphi(x) = (\varphi_1(x) - \bar{\varphi}_1, \varphi_2(x) - \bar{\varphi}_2, \dots, \varphi_N(x) - \bar{\varphi}_N),$$

$$g(x) = (g_1(x), g_2(x), \dots, g_N(x)) = \varphi(x) \alpha,$$

以 $g_1(x), g_2(x), \dots, g_M(x)$ 为隐层节点激励函数, 建立神经网络, 其网络方程为

$$\hat{y} = \bar{y} + \sum_{i=1}^M c_i g_i(x). \quad (9)$$

由于网络的基元函数为一列基本小波函数的线性组合, 所以称这种网络为广义小波神经网络. 在所建立的 EWNN 中, 隐层节点的数目 M 一般要远小于 N , 各激励函数之间又相互正交, 无信息冗余, 于是权系数 $C = (c_1, c_2, \dots, c_M)^T$ 也可由最小二乘法求得:

$$\hat{C} = (\hat{c}_1, \hat{c}_2, \dots, \hat{c}_M)^T = (g_{(M)}^T g_{(M)})^{-1} g_{(M)}^T Y. \quad (10)$$

式中 $g_{(M)}$ 为 g 的前 M 列所构成的 n 行 M 列矩阵.

定理 2 对于广义小波神经网络, 其权系数估计值的均方误差 $MSE(\hat{C}) = \sigma^2 \left(\sum_{i=1}^M \lambda_i^{-1} \right)$.

证 根据式(7)有 $g_{(M)}^T g_{(M)} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_M)$, 由定理 1 证明得 $MSE(\hat{C}) = \sigma^2 \left(\sum_{i=1}^M \lambda_i^{-1} \right)$.

因为 $MSE(\hat{C})$ 比 $MSE(\hat{W})$ 少了 $\lambda_{M+1}^{-1} + \dots + \lambda_N^{-1}$, 而 $\lambda_{M+1} + \dots, \lambda_N$ 是最小的 $N - M$ 个特征值, 所以一般有 $MSE(\hat{C}) \ll MSE(\hat{W})$, 即 EWNN 除避免了维数灾外, 其稳定性较 DWNN 有了很大提高.

综上所述, EWNN 的设计步骤为: ① 利用时频域“分解-综合”方法初选小波基函数集 $\varphi_1(x), \varphi_2(x), \dots, \varphi_N(x)$; ② 中心化各变量; ③ 求 $\varphi^T \varphi$ 的特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N > 0$ 及其对应的单位正交特征向量 $\alpha_i = (a_{i1}, a_{i2}, \dots, a_{iN})^T, i = 1, 2, \dots, N$, 记 $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$; ④ 作正交变换 $g = \varphi \alpha$ 和函数变换 $g(x) = \varphi(x) \alpha$, 其中 $g = (g_1, g_2, \dots, g_N), g(x) = (g_1(x), g_2(x), \dots, g_N(x))$; ⑤ 给定精度 ϵ , 由式(8)求出 M , 以 $g_1(x), g_2(x), \dots, g_M(x)$ 为隐层节点激励函数的 EWNN, 其权值估计 \hat{C} 由式(10)得到, 其中 $g_{(M)} = (g_1, \dots, g_M)$.

以上设计的 EWNN 其网络方程为式(9).

4 仿真示例(Simulation example)

为了验证 EWNN 算法的有效性,分别用 EWNN 和 DWNN 对非线性模型 $y = f(t)$ 进行逼近.

$$y = f(t) = \begin{cases} 0.504t + 5.008, & -10 \leq t < -2, \\ t^2, & -2 \leq t < 0, \\ 12e^{-0.5(t-1)} \sin(0.3t^2 + 0.7t), & 0 \leq t \leq 10. \end{cases}$$

取 $n = 401$ 个样本点,母小波为墨西哥帽函数: $\varphi(x)$

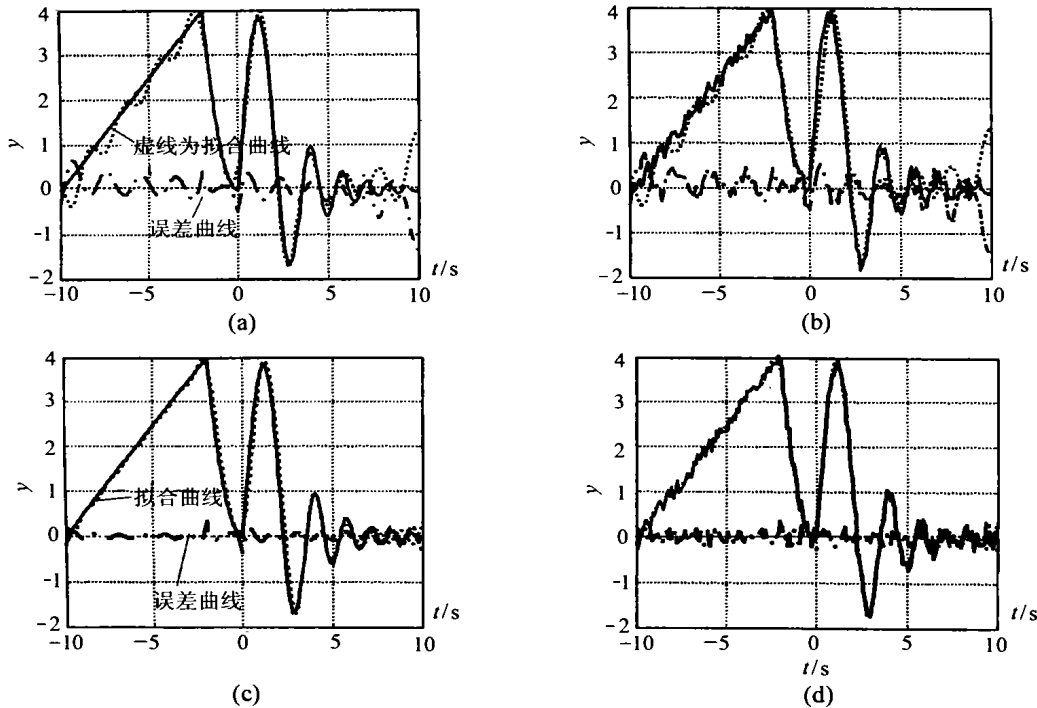


图2 EWNN 和 DWNN 逼近效果的比较

Fig. 2 Contrast of the approximation effect for EWNN and DWNN

5 结束语(Conclusion)

优化小波神经网络结构和设计网络学习的新算法,是有关小波神经网络研究的两大难题.本文在理论分析的基础上,利用 PCA 思想,提出的广义小波神经网络,在网络的规模及抗干扰性上,都有很大提高,可广泛用于非线性系统辨识及图形处理等领域.

参考文献(References):

- [1] ZHANG Qinghua. Using wavelet network in nonparametric estimation [J]. *IEEE Trans on Neural Networks*, 1997, 8(2):227-236.
- [2] FANG Y, CHOW T W S. Orthogonal wavelet neural networks applying to identification of Wiener model [J]. *IEEE Trans on Circuits and Systems-I: Fundamental Theory & Applications*, 2000, 47(4):591-593.
- [3] DAUBECHIES I. The wavelet transform, time-frequency localization and signal analysis [J]. *IEEE Trans on Information Theory*, 1990, 36

$= (1 - x^2)e^{-\frac{x^2}{2}}$, 采用时-频域“分析-综合”方法取 $N = 142$ 个小波基元函数,利用 DWNN 对 $f(t)$ 进行逼近,结果见图 2(a),可见误差比较大.利用本文提出的 EWNN 逼近 $f(t)$,取 $\epsilon = 0.001$,则网络隐层节点只取 28 个,逼近效果就能达到如图 2(c)的程度.图 2(b)和(d)为加噪声 $\zeta(\zeta \sim N[0, 0.06^2])$ 后的两种网络的逼近的效果,可见 EWNN 的鲁棒性也明显优于 DWNN.

(5):961-1005.

- [4] ZHANG J, WALTER G G, MIAO Y. Wavelet Neural networks for function learning [J]. *IEEE Trans on Signal Processing*, 1995, 43(6):1485-1497.
- [5] LI Yinguo, ZHANG Bangli, CHAO Changxiu. Wavelet neural network and its structure design [J]. *Pattern Recognition and Artificial Intelligence*, 1997, 10(3):197-205 (in Chinese).

作者简介:

丁勇 (1967—),女,1997 年于南京航空航天大学自动控制系获硕士学位,现为南京航空航天大学自动化学院讲师.研究方向为智能控制及自修复控制. Email:dingyong-nuaa@sina.com;

刘守生 (1965—),男,1994 年于北京师范大学数学系获硕士学位,现为解放军理工大学数理系副教授,南京航空航天大学测试工程系博士研究生.研究方向为信号检测与分析及模式识别;

胡寿松 (1937—),男,1960 年毕业于北京航空航天大学自动控制,现为南京航空航天大学首席教授,博士生导师,中国自动化学会理事.近期研究方向为故障诊断,鲁棒控制及智能自修复控制.