

小波神经网络逼近能力及 Thau 定理推广*

李 力 方华京

(华中理工大学控制科学与工程系 武汉 430074)

摘要 首先提出神经元数目有限的小波神经网络对一大类 Lipschitz 函数的逼近能力定理; 然后对 Thau 定理进行推广, 得到几个实用性较强的推广定理; 最后通过构造一种基于推广 Thau 定理的小波神经网络非线性观测器, 展示出该逼近定理的应用前景。

关键词 小波神经网络, 函数逼近, Thau 定理推广, 非线性观测器

分类号 TP 18

WNNs Approximate Ability and Improvements of Thau Theorem

Li Li, Fang Huajing

(Huazhong University of Science and Technology)

Abstract A novel approximate theorem of wavelet neural networks is presented at first. Then the well-known Thau theorem is improved and some practical extensive theories are got. Finally, based on the extensive Thau theorem, a kind of state observers for nonlinear systems using WNN is constructed. Simulation results show the effectiveness of the WNN approximate theorem.

Key words wavelet neural networks, approximate ability, improvements of Thau theorem, nonlinear observers

1 引言

前馈神经网络对于未知函数的非线性逼近能力已成为神经网络研究的热点。文献[1]用泛函分析理论证明了各种前馈神经网络在神经元数目无限或很大时的逼近定理, 但在实际应用中不可能或难以实现这样的理想网络, 其价值难以充分体现。因此, 分析神经元数目有限的前馈网络能达到的逼近能力已成为当务之急。

本文着重分析目前小波神经网络在神经元数目有限时对一大类 Lipschitz 函数的逼近能力, 并展示了其应用前景。文中简述了小波神经网络的基本构造, 给出了有限小波神经网络逼近能力的重要定理和 Thau 定理的几个推广定理, 并据此对一类非线性系统提出基于有限小波神经网络的非线性观测器。

2 基本构造

用小波神经网络进行未知非线性函数拟合的基本知识可参见文献[2~4]。本文从函数小波变换与奇异性的关系入手, 深入分析小波神经网络与其它前馈神经网络的不同之处。首先简述如何设计一个小波神经网络来拟合非线性函数 $f(z)$ 。本文使用文献[5]中的多层小波神经网络。不失一般性, 这里仅讨论二进小波神经网络。

多层二进小波神经网络和一般的多层前馈神经网络结构相似, 即输入层和输出层各只含一个, 而隐层可有多个。其输入层为直通节点。每个隐层都包含可数个节点, 以 k 标识, k 可以是任意整数; 每个隐层第 m 个节点的权值和传递函数分别为 2^m 和 $\mathcal{Q}(\cdot)$, 其中 $\mathcal{Q}(\cdot)$ 为相应的小波尺度函数; 隐层节点的阈值各不相同, 第 k 层节点的阈值为 k ; 各个隐层均直接连

* 国家自然科学基金项目(69774014)

接输入层和输出层,相互之间没有互连。输出层为线性节点,阈值为零。对于理想情况,节点权值为由下式确定的系数。

$$f(r, \theta)_{m,k} < f, \varphi_{m,k} > \varphi_{m,k} \quad (1)$$

则由上述小波神经网络可实现

$$f(z) \hat{f}(z) = \sum_{m,k} C_{m,k} \varphi(2^{-m}z - k) \quad (2)$$

其中 $z = [z_1, z_2, \dots, z_n]$ 为输入的数据向量。实际运用时,若已知训练数据集 $[z_1, z_2, \dots, z_n]$, 则定义

$$E = \sum_{i=1}^n (f(z_i) - \hat{f}(z_i))^2 = \sum_{i=1}^n \left(f(z_i) - \sum_{m,k} C_{m,k} \varphi(2^{-m}z_i - k) \right)^2$$

为代价函数,根据训练数据计算出权值矩阵 $C_{m,k}$ 。既可类似于BP算法将计算变成样本的训练过程,也可通过解最小二乘问题来求得系数。具体方法可参见文献[4,5]。

3 用于 Lipschitz 函数类

拟合的可行性

文献[4]所证明的小波神经网络的逼近定理要求网络接近理想化,即要求其中的 m 和 k 都尽量大。而实际应用时,由于受软硬件的限制, m 和 k 都只能为有限值,而且值越小越能节省计算时间和成本。为此本文给出对于一大类 Lipschitz 函数拟合时 m 和 k 值估计的一个逼近定理。

引理1 若 $f(z)$ 满足全局 Lipschitz 条件 $|f(z) - f(\hat{z})| \leq r_f |z - \hat{z}|^\alpha$, 则有

$$|W_{2^{-m}z} f(z)| \leq r \cdot 2^{-m\alpha}$$

其中 $0 < \alpha < 1$ 。如果相应的小波尺度函数 $\varphi(\cdot)$ 具有 n 阶消失矩,即

$$\int_R |t|^k |\varphi(t)| dt = 0$$

其中 k 为小于 n 的正整数。则对于 $\alpha < n$ 且 α 不为整数的情况,仍有

$$|W_{2^{-m}z} f(z)| \leq r \cdot 2^{-m\alpha}$$

当 α 为整数时,上述结论对于大多数 $f(z)$ 均成立。

证明参见文献[6]。

定理1 设小波尺度函数 $\varphi(\cdot)$ 具标准正交性和紧支性的小波神经网络的拟合,满足全局 Lipschitz 条件的未知非线性函数 $f(z)$, 则只要满足 $2^{-3M/2}$

ϵ (其中 M 为 m 的最大值) 以及由 $\varphi(\cdot)$ 确定的不大的 k 值,即可使函数 $f(z)$ 满足如下局部 Lipschitz 条

件:

$$\text{如果 } |z - \hat{z}|^\alpha \in \epsilon, \epsilon > 0, \text{ 则存在 } |f(z) - \hat{f}(\hat{z})| \leq R_f |z - \hat{z}|^\alpha \quad (3)$$

其中, $\varphi(\cdot)$ 具有 n 阶消失矩,且 $\alpha < n, \alpha$ 不为整数。

证明 首先分析 k 的取值。若函数 $f(z)$ 只含有限支撑,则该小波神经网络的各个隐层都只含有限个节点。当隐层的非线性传递函数为紧支函数时,各隐层也都只含有限个节点;当隐层的非线性传递函数为非紧支但迅速衰减时,则各隐层也能以有限个节点精确地近似满足式(2)。如果构造小波神经网络时,使用的传递函数非紧支但迅速衰减且函数 $f(z)$ 不只含有限支撑,则需根据具体使用的小波基来计算 k 的范围的合理估计,但目前尚无完善的结论。一般由于传递函数的迅速衰减, k 值不必很大,所引起的误差也仅与普通的机器误差同级。以下只讨论 k 为确定有限值的情况,但其值随紧支基的不同而不同。

现在分析 m 的取值。利用函数的小波分解理论^[2~7]及引理1,有

$$\begin{aligned} |f(z) - \hat{f}(\hat{z})| &= \left| \sum_{m,k} W_{2^{-m}z} f(z) \varphi_{m,k}(z) - \sum_{m,k} W_{2^{-m}\hat{z}} f(\hat{z}) \varphi_{m,k}(\hat{z}) \right| \\ &= \sum_{m,k} \left| W_{2^{-m}z} f(z) \varphi_{m,k}(z) - W_{2^{-m}\hat{z}} f(\hat{z}) \varphi_{m,k}(\hat{z}) \right| \\ &\leq r_1 \sum_{m,k} 2^{-m} \left| \varphi_{m,k}(z) - \varphi_{m,k}(\hat{z}) \right| \end{aligned}$$

因为 m 的最大值 M 满足 $2^{-3M/2} \in |z - \hat{z}|$, 则当小波神经网络在对函数 M 级分解的层次上完全拟合未知函数时,有

$$\begin{aligned} r_1 \sum_{m=M} 2^{-m} \left| \varphi_{m,k}(z) - \varphi_{m,k}(\hat{z}) \right| \\ r_3 \sum_{m=M} \int_R |\varphi(2^{-m}z - b) - \varphi(2^{-m}\hat{z} - b)| db \\ r_4 \sum_{m=M} r_f |z - \hat{z}|^\alpha \int_R |\varphi(2^{-m}z - b)| db \end{aligned}$$

而忽略对函数 M 级以上的分解和恢复时,有

$$\begin{aligned} r_1 \sum_{m>M} 2^{-m} \left| \varphi_{m,k}(z) - \varphi_{m,k}(\hat{z}) \right| &= \\ r_1 \sum_{m>M} 2^{-m} \left| \varphi_{m,k}(z) \right| \\ r_1 \sum_{m>M} 2^{-3m/2} \int_R |\varphi(2^{-m}z - b)| db \\ r_7 \sum_{m>M} 2^{-3m/2} \int_R |\varphi(2^{-m}\hat{z} - b)| db &\leq r_8 \sum_{m>M} |z - \hat{z}|^\alpha \end{aligned}$$

取 $R_f = \max\{r_6, r_8\}$, 则结论成立, 且易知网络权矩阵为有限值。(证毕)

将以上证明稍加改动可得如下定理:

定理 2 设未知的非线性函数 $f(z)$ 满足如下局部 Lipschitz 条件:

如果 $|z - \hat{z}| \leq \epsilon, \epsilon > 0$, 则存在

$$|f(z) - f(\hat{z})| \leq R_f |z - \hat{z}| \quad (4)$$

对于相当多满足该条件的函数 $f(z)$, 小波尺度函数 $\mathcal{Q}(\cdot)$ 具标准正交性和紧支性的小波神经网络只要满足 $2^{-3M/2} \epsilon$ (其中 M 为 m 的最大值) 及由 $\mathcal{Q}(\cdot)$ 确定的不大的 k 值, 即可使函数 $f(z)$ 满足如下局部 Lipschitz 条件:

如果 $|z - \hat{z}| \leq \epsilon, \epsilon > 0$, 则存在

$$|f(z) - f(\hat{z})| \leq R_f |z - \hat{z}| \quad (5)$$

证明类似于定理 1 的证明。但要注意: 因为 α 为整数时并非对所有的 $f(z)$ 都能使引理 1 成立, 故并不是所有满足式(4) 的函数 $f(z)$ 都能使定理 2 成立(可参见文献[6 ~ 8])。

由小波分解与恢复的实质可知, 小波神经网络总是从函数的整体性质出发去进行逼近, 而将未知的函数细节部分忽略掉。因此, 每增加一个隐层就相当于对函数更细微的一个层次上的信息进行利用, 逼近的效果当然也就更好。而要满足一定的逼近程度, 有一定数目的神经元便足够了。因此, 可使用有限小波神经网络拟合一大类 Lipschitz 函数达到要求的精度, 而不必假设该理想网络的存在性。

应指出的是, 上述分析只得到了该逼近下神经元数目的上界估计。实际上, 由于二进小波的冗余性, 小波神经网络也是冗余的。在感兴趣的定义域和值域研究问题时, 完全可以用比上界要求更少的神经元达到目的。这方面的讨论可参见文献[5]。

另外, 对于 s 进小波神经网络, 类似结论也同样成立。

4 Thau 定理的推广和小波神经网络非线性观测器

为说明上述小波神经网络的作用, 下面介绍它在一类非线性观测器中的应用。由于工程实践中往往不能确知非线性系统的具体形式, 实际上常用函数拟合来估计出所要考虑的非线性函数。而上述小波神经网络正适合应用于该领域。

考虑如下非线性系统

$$\dot{x} = Ax + f(t, u, x, y), \quad y = Cx \quad (6)$$

其中, A 和 C 为定常阵, 且 (A, C) 是可观测的; $f(t, u, x, y)$ 为非线性函数。

设计如下观测器

$$\begin{cases} \dot{\hat{x}} = (A - LC)\hat{x} + Ly + f(t, u, \hat{x}, y) \\ y = C\hat{x} \end{cases} \quad (7)$$

针对上述非线性观测器, 文献[9] 给出如下 Thau 定理:

定理 3^[9] 设有控制系统(6) 和观测器(7), 如果 $f(t, u, x, y)$ 满足全局 Lipschitz 条件

$$|f(t, u, x, y) - f(t, u, \hat{x}, y)| \leq r_f |x - \hat{x}|$$

其中, $r_f < \lambda_{\min}(Q) / 2\lambda_{\max}(P)$, P 和 Q 为正定对称矩阵, 并满足 Lyapunov 方程

$$(A - LC)^T P + P(A - LC) = -Q$$

则有 $\lim_{t \rightarrow \infty} (x - \hat{x}) = 0$

只需将定理 3 的证明稍加改动, 便可得到如下定理:

定理 4 设有控制系统(6), 观测器为

$$\begin{cases} \dot{\hat{x}} = (A - LC)\hat{x} + Ly + \hat{f}(t, u, \hat{x}, y) \\ y = C\hat{x} \end{cases} \quad (8)$$

若 $f(t, u, x, y)$ 及其拟合函数 $\hat{f}(t, u, \hat{x}, y)$ 满足如下局部 Lipschitz 条件:

如果 $|x - \hat{x}| \leq \epsilon, \epsilon > 0$, 则存在

$$|f(t, u, x, y) - \hat{f}(t, u, \hat{x}, y)| \leq R_f |x - \hat{x}| \quad (9)$$

且有 $R_f < \lambda_{\min}(Q) / 2\lambda_{\max}(P)$, 其中 P 和 Q 为正定对称矩阵, 并满足 Lyapunov 方程

$$(A - LC)^T P + P(A - LC) = -Q$$

则有 $\lim_{t \rightarrow \infty} |x - \hat{x}| \leq \epsilon$

证明 针对系统(6) 及观测器(8), 令 $e = x - \hat{x}$, 可得误差观测器方程

$$\dot{e} = (A - LC)e + f(t, u, x, y) - \hat{f}(t, u, \hat{x}, y) \quad (10)$$

考虑如下 Lyapunov 函数 $V = e^T P e$ (P 为正定阵), 对其求导, 由式(10) 有

$$\dot{V} = e^T [(A - LC)^T P + P(A - LC)] e + 2e^T P [f(t, u, x, y) - \hat{f}(t, u, \hat{x}, y)]$$

利用式(9), 有 $|e| \leq \epsilon, \epsilon > 0$, 则

$$\begin{aligned} \dot{V} &\leq e^T [(A - LC)^T P + P(A - LC)] e + 2R_f e^T P e = \\ &= e^T [(A - LC)^T P + P(A - LC) + 2R_f P] e = \\ &= -e^T [Q - 2R_f P] e < 0 \end{aligned}$$

根据文献[10] 中实用稳定性的定义, 可知该系统是实用稳定的, 故有 $\lim_{t \rightarrow \infty} |e| \leq \epsilon$ (证毕)

进一步,有如下推广定理:

定理 5 对于控制系统(6)和观测器(8),如果 $f(t, u, x, y)$ 及其拟合函数 $\hat{f}(t, u, \hat{x}, y)$ 满足如下局部 Lipschitz 条件:

如果 $|x - \hat{x}|^\alpha \in \epsilon, \epsilon > 0$, 则存在

$$|f(t, u, x, y) - \hat{f}(t, u, \hat{x}, y)| \leq R_f |x - \hat{x}|^\alpha$$

且有 $R_f < \lambda_{\min}(Q) / 2\lambda_{\max}(P)$, 其中 P 和 Q 为正定对称矩阵,并满足 Lyapunov 方程

$$(A - LC)^T P + P(A - LC) = -Q$$

则有 $\lim_{t \rightarrow \infty} |x - \hat{x}|^\alpha \in \epsilon$

定理 4 和定理 5 与 Thau 定理本质一样,但更便于实际应用。结合定理 2 和定理 4 不难发现,恰能用小波神经网络以较小的规模来对一大类常见的非线性系统实现较好的拟合。而如果直接利用 Thau 定理构造观测器^[11], 不难发现当 $x = \hat{x}$ 时,有 $f(t, u, x, y) = \hat{f}(t, u, \hat{x}, y)$ 。这意味着任意函数 $f(t, u, x, y)$ 及其拟合函数 $\hat{f}(t, u, \hat{x}, y)$ 在无数点上取值相同,而有限待定参数拟合方式不可能得到这种逼近效果。所以在实践中不便直接利用 Thau 定理,而定理 4 却是可行的。实际上,文献[11]就是在 ϵ 尽可能小的情况下应用定理 4 巧妙地设计观测器的。

进一步,结合定理 1 和定理 5 可构造出应用范围更广的非线性观测器。

5 观测器的实现

L 阵的存在性和具体计算方法参见文献[12~14],这里不赘述。

用小波神经网络拟合未知函数时,估计 Lipschitz 指数 α 和比例系数 R_f 是十分重要的。Lipschitz 指数 α 的基本估计方法可参见文献[5~7]。选定小波集后,即可根据定理 1 的证明对比例系数 R_f 做出估计。不难看出,参数 ϵ, R_f, α 和 L 阵存在着复杂的相互制约关系,如何根据具体情况对这些参数进行折衷设计以求好的效果,是实际应用中必须考虑的问题。相关讨论和实例将另文分析。

6 结语

用小波分析各种前馈神经网络的逼近性质是当前研究的新热点,并已取得许多成果^[6]。如果将其它前馈神经网络的隐层传递函数看成小波尺度函数的复合,则在某些情况下完全可能由更少数的隐层神经元来达到拟合目的,文献[11]给出的例子就说明了这一点。可以预见小波分析的发展能为逼近论增加

更为精彩的内容。

目前,小波技术以其优良的时频分析性质,已在控制工程领域得到了日益广泛的应用,这充分体现了小波技术的特性。但也应看到,与泛函分析相比,小波分析在控制理论研究中的应用则少得多。这主要是因为小波发展较晚,还未充分发挥其特点。但小波分析吸收了现代分析学众多分支的精华,今后必将在自动控制基础理论研究中发挥更大的作用。

参考文献

- 1 陈天平. 神经网络及其在系统识别应用中的逼近问题. 中国科学 A 辑, 1994, 24(1): 1~7
- 2 Zhang Q, Benveniste A. Wavelet network. IEEE Trans on Neural Network, 1992, 3(6): 889~898
- 3 Delyon B, Juditsky A, Benveniste A. An accuracy analysis for wavelet approximates. IEEE Trans on Neural Network, 1995, 6(2): 332~348
- 4 Zhang J, Walter G G, Miao Y *et al.* Wavelet neural networks for function learning. IEEE Trans on Signal Processing, 1995, 43(6): 1485~1499
- 5 李建平. 小波分析与信号处理——理论、应用及软件实现. 重庆: 重庆出版社, 1997
- 6 Mallat S, Hwang W L. Singularity detection and processing with wavelets. IEEE Trans on Information Theory, 1992, 38(2): 617~642
- 7 Mallat S, Zhang S. Characterization of signals from multiscale edges. IEEE Trans on Pattern Analysis and Machine Intelligence, 1992, 14(7): 710~732
- 8 孙永生, 房良孙. 函数逼近论. 北京: 北京师范大学出版社, 1989
- 9 Thau F E. Observing the state of nonlinear dynamic systems. Int J of Control, 1973, 17(3): 471~479
- 10 廖晓昕. 稳定性的理论、方法和应用. 武汉: 华中理工大学出版社, 1999
- 11 胡寿松, 周川, 胡维礼, 等. 一种基于神经网络非线性观测器的鲁棒故障检测. 信息与控制, 1999, 28(3): 214~218
- 12 Sakhar R, Karl J H. Observer design for a class of nonlinear systems. Int J of Control, 1994, 59(2): 515~528
- 13 Rajamani R, Cho Y M. Existence and design the state of observers for nonlinear systems. Int J of Control, 1998, 69(5): 717~731
- 14 Rajamani R. Observers for Lipschitz nonlinear systems. IEEE Trans on Automatic Control, 1998, 43(2): 397~401

表3 计算结果(极小化目标函数值为1671.7)

j	1	2	j	1	2	j	1	2
$x_j^1(1)$	12	0	$y_j^2(3)$	10	5	$z_{ij}^3(5)$	0	0
$x_j^1(2)$	0	5	$y_j^2(4)$	0	0	$z_{ij}^3(1)$	10	0
$x_j^1(3)$	0	15	$y_j^2(5)$	0	0	$z_{ij}^3(2)$	0	0
$x_j^1(4)$	0	0	$z_{ij}^3(1)$	0	11	$z_{ij}^3(3)$	0	15
$x_j^1(5)$	0	0	$z_{ij}^3(2)$	49	0	$z_{ij}^3(4)$	0	0
$y_j^1(1)$	0	3	$z_{ij}^3(3)$	0	0	$z_{ij}^3(5)$	0	0
$y_j^1(2)$	25	0	$z_{ij}^3(4)$	0	0	$p_j^1(1)$	12	8
$y_j^1(3)$	10	0	$z_{ij}^3(5)$	0	0	$p_j^1(2)$	12	8
$y_j^1(4)$	0	10	$z_{ij}^3(1)$	0	0	$p_j^1(3)$	15	10
$y_j^1(5)$	0	0	$z_{ij}^3(2)$	0	0	$p_j^1(4)$	10	10
$x_j^2(1)$	0	8	$z_{ij}^3(3)$	0	0	$p_j^1(5)$	5	10
$x_j^2(2)$	10	4	$z_{ij}^3(4)$	0	35	$p_j^2(1)$	10	8
$x_j^2(3)$	0	0	$z_{ij}^3(5)$	0	0	$p_j^2(2)$	10	8
$x_j^2(4)$	0	0	$z_{ij}^3(1)$	0	0	$p_j^2(3)$	8	6
$x_j^2(5)$	0	0	$z_{ij}^3(2)$	0	12	$p_j^2(4)$	10	5
$y_j^2(1)$	0	0	$z_{ij}^3(3)$	28	0	$p_j^2(5)$	0	0
$y_j^2(2)$	0	0	$z_{ij}^3(4)$	0	0			

力平衡的前提下,极小化产品提前/拖期惩罚费用、运输费用和产品生产成本的总额,从而实现了供应链准时化供应并获取最大利润的经营目标。

7 结 论

分布需求计划是供应链管理中的一项重要研究工作。本文提出的准时化分布需求计划方法,不仅能有效地编制出准时化分布需求计划,实现供应产

品准时化控制策略,而且为供应链管理问题的研究提供了一条有效途径。所提出的方法运算速度快而准确,具有很强的实用意义,从而能够极大地提高企业在市场竞争中的应变能力。

参 考 文 献

- 1 真彤,祁国宁.敏捷制造的总体技术研究.计算机集成制造系统,1999,3:1~10
- 2 Bdganha M P, Cohen M A. The stabilizing effect of inventory in supply chains. *Opera Res*, 1998, 46(3):72~83
- 3 Beam on B M. Supply chain design and analysis: Models and methods. *Int J Prod Eco*, 1998, 55:281~294
- 4 Jukka Korepla, Antti Lehmusvaara. A customer oriented approach to warehouse network evaluation and design. *Int J Prod Eco*, 1999, 59:135~146
- 5 Sugimori Y, Kusunoki K, Cho F *et al.* Toyota production system and Kanban system materialization of just-in-time and respect-for human system. *Int J Prod Res*, 1977, 15:553~564

作 者 简 介

王 玮 女,1964年生。清华大学自动化系博士后研究人员。研究方向为生产计划与调度理论,供应链管理与决策等。

汪定伟 男,1949年生。东北大学信息科学与工程学院教授,博士生导师。研究方向为生产计划与调度理论,智能优化算法,供应链管理与决策等。

柴跃廷 男,1964年生。清华大学自动化系副教授。研究方向为管理信息系统,供应链管理与决策等。

(上接第564页)

作 者 简 介

李 力 男,1976年生。1999年毕业于华中理工大学提班,获学士学位,现为华中理工大学控制科学与工程系硕士研究生。目前研究方向为过程控制,故障诊断,非线性系

统,神经网络,智能控制等。

方华京 男,1955年生。1991年在华中理工大学获博士学位,现为华中理工大学控制科学与工程系教授,博士生导师。研究领域为鲁棒控制理论与应用,控制系统故障检测与诊断技术等。