

文章编号:1002-0411(2004)03-0347-05

基于最小二乘法的 SIMO 傅里叶神经网络研究

杨旭华, 戴华平, 孙优贤

(浙江大学现代控制技术研究所工业控制技术国家重点实验室, 浙江 杭州 310027)

摘要:利用傅里叶级数的原理,构造单输入、多输出(SIMO)傅里叶神经网络,将非线性映射转化成为线性映射,将求解神经网络权值的方法由非线性优化方法转化成为线性优化方法,并采用最小二乘法计算网络的权值,从而大大提高了神经网络的收敛速度并避免了局部极小问题.而且,在训练输出样本受白噪声影响时,最小二乘法具有良好的降低噪声影响的功能.

关键词:傅里叶神经网络;非线性优化;线性优化;最小二乘法

中图分类号:TP183

文献标识码:A

Research on SIMO Fourier Neural Networks Based on Least Square Method

YANG Xu-hua, DAI Hua-ping, SUN You-xian

(National Key Laboratory of Industrial Control Technology, Institute of Modern Control Technology, Zhejiang University, Hangzhou 310027, China)

Abstract: Based on Fourier series principle, the single input, multiple outputs (SIMO) Fourier neural networks are proposed. The SIMO Fourier neural networks turn nonlinear mapping relationship into linear mapping relationship, turn the method of solving neural networks' weights from the nonlinear optimization method to linear optimization method, and use the least square method to compute the weights of the network. So, the SIMO Fourier neural networks highly improve the convergence speed and avoid local minima problem. When the training output samples are affected by white noise, the least square method have good denoising function.

Keywords: Fourier neural network; nonlinear optimization; linear optimization; least square method

1 引言 (Introduction)

目前,在各类神经网络模型中应用较为广泛的是BP神经网络,BP神经网络当其隐节点函数是Sigmoid函数时能够任意逼近连续多元函数.然而,BP神经网络由于采用非线性优化算法,不可避免地存在许多缺点,其中有:①算法收敛速度很慢;②存在局部极小问题;③当训练样本受噪声影响时,容易造成模型的偏离.为克服上述缺点,提出了许多改进方法^[1-5],但仍有许多问题未能很好地解决.

本文利用傅里叶级数的原理,构造单输入、多输出(SIMO)傅里叶神经网络,将非线性映射转化成为线性映射,将求解神经网络权值的方法由非线性优化方法转化成为线性优化方法,并采用最小二乘法来计算网络的权值,大大提高了神经网络的收敛速度,避免了局部极小问题.而且,在训练输出样本受白噪声影响时,最小二乘法具有良好的降低噪声影

响的功能.同时,在任意有界输入、有界输出的条件下,傅里叶神经网络能够以任意的精度逼近多个任意的非线性映射关系并具有良好的泛化能力.

2 傅里叶级数 (Fourier series)

一个周期为 T 的平稳随机过程 $x(t)$ 可以用傅里叶级数表示为如下的形式:

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos(nw_0 t) + b_n \sin(nw_0 t)) \quad (1)$$

其中:

$$w_0 = \frac{2\pi}{T} \quad (2)$$

a_n 、 b_n 是傅里叶系数,其表达式为:

$$a_n = \frac{2}{T} \int_0^T x(t) \cos \frac{2n\pi t}{T} dt \quad (n = 0, 1, 2, 3 \dots) \quad (3)$$

$$b_n = \frac{2}{T} \int_0^T x(t) \sin \frac{2n\pi t}{T} dt \quad (n = 1, 2, 3, \dots) \quad (4)$$

下面不加证明地叙述一个基本定理.

定理:当 $n \rightarrow \infty$ 时,傅里叶系数 a_n, b_n 迅速趋向于零,即:可以用傅里叶级数的前 $2n + 1$ 项部分和代替 $x(t)$, n 选取的越大,则精度越高.

3 SIMO 傅里叶神经网络 (SIMO Fourier neural network)

$$\left\{ 1, \cos \frac{2\pi t}{T}, \sin \frac{2\pi t}{T}, \cos \frac{4\pi t}{T}, \sin \frac{4\pi t}{T}, \dots, \cos \frac{2n\pi t}{T}, \sin \frac{2n\pi t}{T} \right\}, n \rightarrow \infty \quad (5)$$

式(5)是正交的三角函数系.傅里叶级数是正交三角函数系的线性组合,取前 $2n + 1$ 项作为基函数构造 SIMO 傅里叶神经网络,如图 1(假设共有 m 个输出单元).

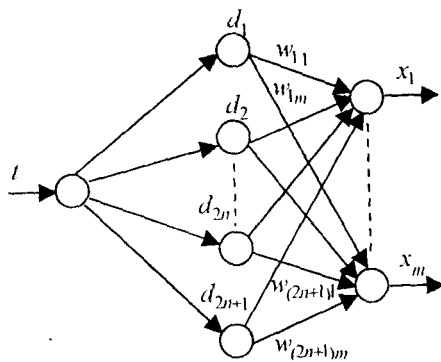


图 1 SIMO 傅里叶神经网络
Fig. 1 SIMO Fourier neural network

图 1 中, SIMO 傅里叶神经网络是一个单输入、多输出系统,具有一个输入变量和多个输出变量,它的结构分为三层:输入层、中间隐层和输出层.

t 为输入变量.输入层和中间隐层的所有联接权值均为 1.

$d_k (k = 1, 2, \dots, 2n + 1)$ 正交三角函数系对应着不同的隐层单元,共 $2n + 1$ 个.

$$\begin{aligned} d_1 &= 1, d_{h+1} = \cos \frac{2\pi ht}{T} \\ d_{n+h+1} &= \sin \frac{2\pi ht}{T} \quad (h = 1, 2, \dots, n) \end{aligned} \quad (6)$$

d_1 单元不随输入变化而变化,输出始终为 1.

$w_{kl} (k = 1, 2, \dots, 2n + 1) (l = 1, 2, \dots, m)$ 表示中间隐层与输出层之间的连接权值(连接第 k 个中间隐层单元和第 l 个输出单元之间的权值).

输出层共有 m 个输出单元,采用线性神经元,

假设其输入为 $m_{kl} (k = 1, 2, \dots, 2n + 1) (l = 1, 2, \dots, m)$ (表示来自第 k 个中间隐层单元的、第 l 个输出单元的输入):

$$m_{kl} = d_k w_{kl} \quad (7)$$

$x_l (l = 1, 2, \dots, m)$ 为第 l 个输出单元的输出生量.

输出单元其输入、输出的关系为:

$$x_l = \sum_{k=1}^{2n+1} m_{kl} \quad (8)$$

将式(7)代入式(8)得:

$$x_l = \sum_{k=1}^{2n+1} d_k w_{kl} \quad (9)$$

式(9)是 SIMO 傅里叶神经网络第 l 个输出单元的输入输出关系.

将式(9)展开成以 t 为自变量的形式得到:

$$\begin{aligned} x_l(t) &= w_{1l} + w_{2l} \cos \left(\frac{2\pi t}{T} \right) + w_{(n+2)l} \sin \left(\frac{2\pi t}{T} \right) + \dots \\ &+ w_{(n+1)l} \cos \left(\frac{2n\pi t}{T} \right) + w_{(2n+1)l} \sin \left(\frac{2n\pi t}{T} \right) \end{aligned} \quad (10)$$

可以看到,式(10)与 $x(t)$ 的傅里叶级数展开式(1)在形式上是等价的,从而证明了傅里叶神经网络的每个输出等价于傅里叶级数.

注意到 SIMO 傅里叶神经网络的输入层单元和所有中间隐层单元之间的连接权值均为 1,需要调节的参数是隐含层和输出层之间的连接权值.观察式(9),隐含层空间到输出层空间的映射是线性的;SIMO 傅里叶神经网络的任何一个输出单元的输出是所有中间隐层单元的线性组合.所以,对 SIMO 傅里叶神经网络来说,求得中间隐层和输出层之间的正确的连接权值,便可以得到该 SIMO 傅里叶神经网络的正确映射关系.

构成 SIMO 傅里叶神经网络的基本思想是:用傅里叶正交三角函数系作为隐单元的“基”构成隐含层空间,这样就可将输入矢量直接(权连接均为 1)映射到隐含层空间.而隐含层空间到输出层空间的映射是线性的.由此可见,从总体上看,网络由输入到输出的映射是非线性的,而网络输出对可调参数而言又是线性的;这样就将输入层和输出层之间的非线性映射关系转化成了隐含层和输出层之间的线性映射关系.在线性映射中求取线性权参数的最优解,是线性优化问题,所以可以采用线性优化方法求解隐含层和输出层之间权值,从而将求解神经网络权值的方法由非线性优化方法转化成为线性优化方法.而线性优化方法具有比非线性优化方法快得多的收敛速度并且没有局部最小问题,从而大大加快了神经网络的学习速度并避免了局部最小问题.

按照傅里叶级数理论,在隐含层采用一定数量的隐含层单元(基函数 d_k),傅里叶神经网络可以获得任意的精度.

采用最小二乘递推算法来对傅里叶神经网络进行有导师学习,具体算法如下:

输入 $t = [t(1), t(2), \dots, t(q)]$ 表示 q 个不同的训练输入样本. 隐含层单元、权值和输出层单元之间是线性的映射关系,其表达式为:

$$Y = DW \quad (11)$$

$$Y = \begin{bmatrix} Y(1) \\ Y(2) \\ \vdots \\ Y(q) \end{bmatrix} = \begin{bmatrix} y_1(1) & y_2(1) & \cdots & y_l(1) \\ y_1(2) & y_2(2) & \cdots & y_l(2) \\ \vdots & \vdots & \ddots & \vdots \\ y_1(q) & y_2(q) & \cdots & y_l(q) \end{bmatrix} \quad (12)$$

Y 表示网络的输出, $Y(i) (i=1, 2, \dots, q)$ 表示对应于第 i 个输入的输出层所有输出单元的输出, $y_l(i)$ 表示当输入是 $t(i)$ 时,输出 x_l 的值.

$$D = \begin{bmatrix} D^T(1) \\ D^T(2) \\ \vdots \\ D^T(q) \end{bmatrix} = \begin{bmatrix} d_1(1) & d_2(1) & \cdots & d_{(2n+1)}(1) \\ d_1(2) & d_2(2) & \cdots & d_{(2n+1)}(2) \\ \vdots & \vdots & \ddots & \vdots \\ d_1(q) & d_2(q) & \cdots & d_{(2n+1)}(q) \end{bmatrix} \quad (13)$$

D 表示不同时刻所有隐层单元的值, $D^T(i) (i=1, 2, \dots, q)$ 表示对应于第 i 个输入,所有隐层单元的值.

W 表示隐层与输出层之间的权连接,具体表达式为:

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{(2n+1)1} & w_{(2n+1)2} & \cdots & w_{(2n+1)m} \end{bmatrix} \quad (14)$$

$w_{kl} (k=1, 2, \dots, 2n+1) (l=1, 2, \dots, m)$ 表示中间隐层的第 k 个隐层单元和输出层的第 l 个输出单元之间的连接权值.

$y_l(i)$ 表示当输入是 $t(i)$ 时,输出 x_l 实际观测值,作为训练样本值.

$x_l(t(i))$ 表示当输入是 $t(i)$ 时,输出 x_l 的期望值.

实际输出值和输出样本值之间的误差:

$$e_{il} = x_l(t(i)) - y_l(i) \quad i = 1, 2, \dots, q \quad (15)$$

q 代表每个输出训练样本点的数量.

定义性能指标函数为:

$$J = \frac{1}{2} \sum_{l=1}^m \left(\sum_{i=1}^q e_{il}^2 \right) \quad (16)$$

使 $J = \min$ 的 W ,即为所求的 W ,因此,由

$$\frac{\partial J(t)}{\partial W} = 0 \quad (17)$$

可得最小二乘递推算法:

$$W(i+1) = W(i) + K(i+1) [Y(i+1) - D^T(i+1)W(i)] \quad (18)$$

$$K(i+1) = \frac{P(i)D(i+1)}{1 + D^T(i+1)P(i)D(i+1)} \quad (19)$$

$$P(i+1) = [I - K(i+1)D^T(i+1)]P(i) \quad (20)$$

$(i=0, 1, 2, \dots, q)$

直接令 $W(0) = 0, P(0) = \alpha I$ (α 是足够大的正数, I 是单位阵),从第一组数据开始递推运算,即可得到所求的权值.

当训练输出样本受白噪声影响时,最小二乘法是最小方差线性无偏估计,即采用最小二乘法估计权值,有良好的降低噪声影响的功能.

注意到傅里叶级数逼近多个平稳随机过程 $x_1(t), x_2(t), \dots, x_m(t)$ 时, $x_1(t), x_2(t), \dots, x_m(t)$ 必须是周期为 T 的周期函数,而非线性函数往往不是周期函数,不能直接运用傅里叶神经网络.

可以利用周期延拓来解决这个问题.

假设需要实现当 $t \in [a, b]$ 时, $x_1(t), x_2(t), \dots, x_m(t)$ 和 t 之间的非线性映射关系.

令 $T = b - a, T$ 是一个有界正实数.

令 $t \in (-\infty, +\infty)$,在时间 $(-\infty, +\infty)$ 上对 $t \in [a, b]$ 的 $x_1(t), x_2(t), \dots, x_m(t)$ 进行周期延拓,这样就可以将 $x_1(t), x_2(t), \dots, x_m(t)$ 化为周期为 T 的周期函数,就可以运用傅里叶神经网络获取当 $t \in [a, b]$ 时, $x_1(t), x_2(t), \dots, x_m(t)$ 和 t 之间的非线性映射关系.

4 仿真研究 (Simulation research)

4.1 收敛速度仿真研究

仿真研究以单输入、双输出非线性系统式 (21)、(22) 为例.

$$x_1(t) = 3.5e^{[-2.1(t+0.5)]} \sin(12t) \quad (21)$$

$$x_2(t) = 1.6t^2 \sin(8t) \cos(t) \quad (22)$$

$$t \in [0, 2]$$

式 (21)、(22) 所表示的曲线分别如图 2、图 3 中实线所示.

训练输入样本集 $t = 0:0.05:2$, 训练输出样本集 $x_1(t)$ 和 $x_2(t)$ 分别由式 (21) 和式 (22) 得到,两个输出的训练样本点均为 41 个,性能指标 $J < 0.02$.

构造单输入、双输出傅里叶神经网络,网络结构为 $1 \times 29 \times 2$,采用最小二乘递推算法,经过 41 步的

递推计算, $J=0.0159$.

构造传统 BP 神经网络(以 sigmoid 函数作隐层的激活函数), 网络结构为 $1 \times 29 \times 2$, 训练指标要求相同, 学习率采用 $\eta=0.004$, 传统 BP 神经网络经过 714783 次训练, $J < 0.02$.

由以上比较可知: 在相近的性能指标下, 两种神经网络所需要的训练次数竟然相差 10^4 数量级. 可以得出这样的结论: 与传统 BP 神经网络相比较, 傅里叶神经网络大大提高了神经网络的收敛速度.

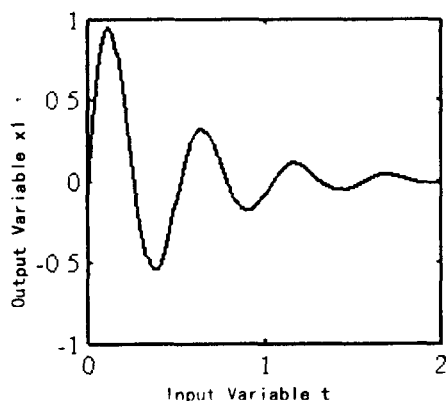


图2 检验泛化能力仿真曲线 1

Fig.2 Simulation curve 1 of generalization capability test

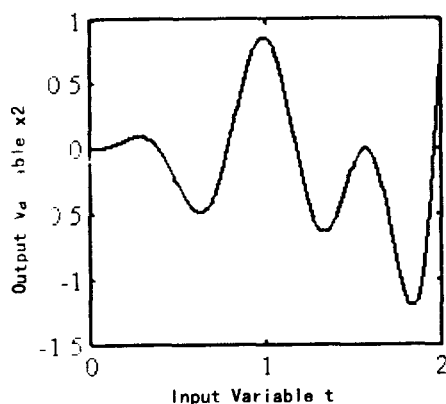


图3 检验泛化能力仿真曲线 2

Fig.3 Simulation curve 2 of generalization capability test

4.2 泛化能力检验

因为傅里叶神经网络获得的是在一定自变量区间(训练样本集所处的区间)上的输入与输出之间的非线性映射关系, 所以测试数据应当取自上述区间内没有训练过的数据.

测试输入集 $t=0.02:0.05:1.92$. 将测试数据代入已经训练好的傅里叶神经网络, 得到两条输出曲线, 如图 2、图 3 中的虚线所示, 由式(21)和式(22)所绘制的曲线为实线. 观察图 2、图 3 中的实线和虚线, 它们几乎完全重合, 表明傅里叶神经网络具有良好的泛化能力.

4.3 降低噪声影响能力检验

假定用来训练傅里叶神经网络的输出样本受白噪声影响, 其表达式分别为式(21)和(22)与白噪声的叠加, 分别如式(23)和(24)所示.

$$x_1(t) = 3.5e^{-2.1(t+0.5)} \sin(12t) + v_1(t) \quad (23)$$

$$x_2(t) = 1.6t^2 \sin(8t) \cos(t) + v_2(t) \quad (24)$$

$$t \in [0, 2]$$

其中 $v_1(t)$ 、 $v_2(t)$ 都是均值为 0, 方差为 0.01, 符合正态分布的白噪声.

式(23)和式(24)表示的曲线分别如图 4 和图 5 所示.

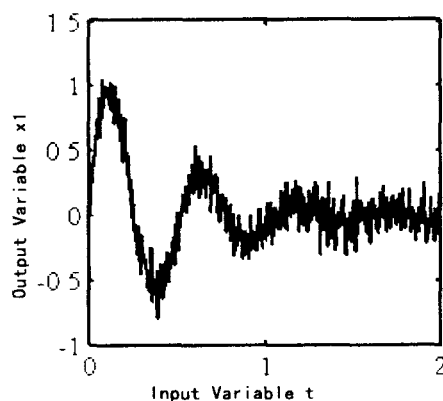


图4 受白噪声影响的输出曲线 1

Fig.4 Output curve 1 affected by white noise

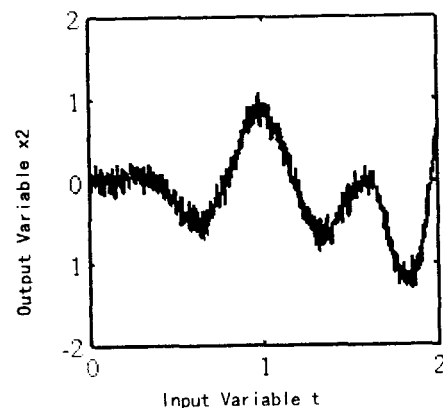


图5 受白噪声影响的输出曲线 2

Fig.5 Output curve 2 affected by white noise

训练输入样本集 $t=0:0.003125:2$, 训练输出样本集 $x_1(t)$ 和 $x_2(t)$ 分别由式(23)和式(24)得到, 两个输出的训练样本数均为 641 个. 采用本文的最小二乘递推算法计算网络权值, 得到网络权值后, 用训练输入样本集代入已经训练好的神经网络模型, 得到两条输出曲线, 分别如图 6 和图 7 所示. 把它们和图 2、图 3 中的未受噪声影响的输出曲线相比较, 我们可以看到, 经最小二乘法训练得到的神经网络模型具有良好的降低噪声影响的能力.

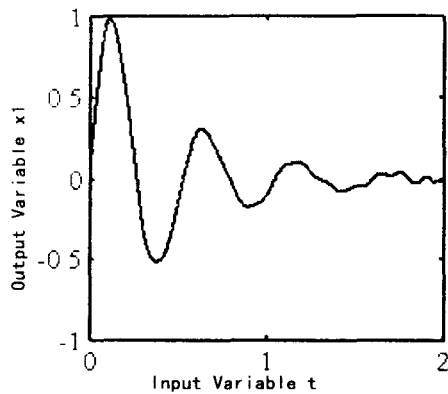


图6 降低噪声影响效果图1

Fig. 6 Denoising effect figure 1

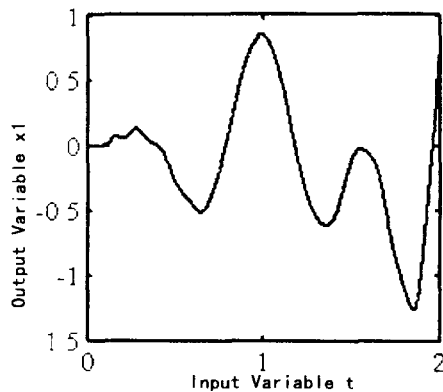


图7 降低噪声影响效果图2

Fig. 7 Denoising effect figure 2

5 结论 (Conclusion)

本文利用傅里叶级数的原理,构造 SIMO 傅里叶神经网络,将非线性映射转化成为线性映射,将求

解神经网络权值的方法由非线性优化方法转化成为线性优化方法,并采用最小二乘法计算网络的权值,大大提高了神经网络的收敛速度并避免了局部极小问题;而且,在输出样本受白噪声影响时,最小二乘法具有良好的降低噪声影响的功能. SIMO 傅里叶神经网络可以用于 SIMO 系统的建模,具有良好实用性和推广前景.

参 考 文 献 (References)

- [1] Wei W J, Li Z J, Wei L S, *et al.* The improvements of BP neural network learning algorithm [A]. 5th International Conference on Signal Processing [C]. Beijing: 2000. 1647 ~ 1649.
- [2] Beigy H, Meybodi M R. Adaptation of parameters of BP algorithm using learning automata [A]. Sixth Brazilian Symposium on Neural Networks [C]. Brazil: 2000. 24 ~ 31.
- [3] Yang Y X, Zheng G, Liu D. BP-GA mixed algorithms for short-term load forecasting [A]. International Conference on Info-tech and Info-net [C]. Beijing: 2001. 334 ~ 339.
- [4] 周 祥,陈丙珍,等. 一种用于 BP 神经网络训练的改进遗传算法 [J]. 化工学报, 2001, 52(10): 925 ~ 927.
- [5] 刘光中,李晓峰. 人工神经网络 BP 算法的改进和结构的自调整 [J]. 运筹学学报, 2001, 5(1): 81 ~ 88.

作者简介

杨旭华(1971 -), 博士生. 研究领域为人工智能, 神经网络.

孙优贤(1940 -), 男, 中国工程院院士, 教授, 博士生导师. 研究领域为鲁棒控制, 容错控制和工厂综合自动化等.