

基于前向神经网络的非线性时变系统辨识^①

顾成奎, 王正欧

(天津大学系统工程研究所, 天津 300072)

摘要:提出基于前向神经网络的非线性时变系统辨识方法,并用局部化推广卡尔曼滤波算法训练网络.该算法与全局推广卡尔曼滤波算法相比,不需要矩阵求逆运算,具有更高的收敛速度和更小存储容量要求.仿真结果表明本文提出的方法在对非线性时变系统辨识方面取得较好效果.

关键词:非线性时变系统;前向神经网络;局部化学习算法;卡尔曼滤波

中图分类号:TP18

文献标识码:A

文章编号:1007-9807(2001)03-0036-01

0 引言

目前,神经网络在非线性的定常系统的辨识方面取得了卓越的成绩^[1].这些系统输入、输出数据之间的映射关系是固定的,神经网络按照一定的辨识方法,根据待辨识系统输入、输出的历史数据对权重进行训练,使网络能够完成这种固定的映射关系.时变系统与此不同,时变系统的输入输出之间的映射关系是随时间的变化而不断变化的,这就要求神经网络的输入输出映射关系也随时间而变,但现有的训练方法都不能满足这一个要求,因此需要探索用于时变系统辨识的神经网络训练方法.文[2]首次提出用推广卡尔曼滤波算法训练前向神经网络,他们仅给出了网络学习的全局算法,此算法计算非常繁琐,且需矩阵求逆运算.文[3]也曾用推广卡尔曼滤波算法对前向神经网络进行训练,他们的研究目的是为了改进BP算法的学习率,以时变学习率代替原来定常学习率,这种算法把单个权值作为研究对象,为每个权值建立动态方程,并以该神经元的输出作为观测值,建立观测方程,然后利用推广的卡尔曼滤波算法进行学习.但是文[3]只用这种方法解决了XOR问题,没能把算法应用到系统辨识领域,而且,这种

方法还需借助于BP算法的误差反向传播,因而计算比较繁琐.文[4]提出利用推广卡尔曼滤波算法对径向基函数网络进行训练,并用于系统辨识领域取得了很好的效果.但是这些算法所依据的基本思想都是针对非时变系统的.本文首先从全局模型出发,建立前向神经网络的时变模型,然后运用局部线性化和分解思想,提出了基于单个神经元的局部化推广卡尔曼滤波训练方法.本文提出的算法与文[2、3]算法相比,不仅可以对时变系统进行辨识,而且无矩阵求逆运算,且比全局算法具有更高的收敛速度和更小存储容量要求.仿真例子说明了本文方法是有效的,且对非线性时变系统进行较好辨识.

1 非线性时变系统前向神经网络辨识

虽然已经有一些研究者^[2-4]提出了用卡尔曼滤波算法对网络权值进行训练,但他们的目的仅限于使算法简洁,收敛速度更快,具有更好的在线计算能力,因此不能进一步得到一种适于时变非线性系统辨识的模型.

本文采用的方法的基本思想如下:在定常非

收稿日期:2000-08-17;修订日期:2001-04-27.

基金项目:国家自然科学基金资助项目(69774035).

作者简介:顾成奎(1974-),男,江苏人,博士生.

线性系统的神经网络辨识方法中,学习的目的是使网络的权值收敛到某一固定值,网络实现的是某种固定的映射关系,可以设想,在对时变系统进行辨识时,如果使网络的权值按照某种规律动态变化,那么网络的输入输出映射关系也必然会随时间而变化,从而有可能对时变系统进行辨识。

对于任一前向神经网络,设 $\theta(k)$ 是由网络的全部神经元的权重组成的向量,维数是 T , $Z(k)$ 表示网络的期望输出向量,维数是 L , $I(k)$ 表示输入向量,维数是 M , 激活函数为非线性 Sigmoid 函数,那么前向网络的时变特性可用如下的状态方程和观测方程来表示:

$$\theta(k+1) = \psi(k+1, k)\theta(k) + \Gamma(k)W(k) \quad (1)$$

$$Z(k) = F(\theta(k), I(k)) - V(k) \quad (2)$$

这就是前向网络的全局模型,实际上权重向量 $\theta(k)$ 的变化规律很难获得,即 $\psi(k+1, k)$, $\Gamma(k)$ 是未知的,不妨假设 $\psi(k+1, k) = I$, $\Gamma(k) = I$, 即连接权的变化采用随机游动模型:

$$\theta(k+1) = \theta(k) + W(k) \quad (3)$$

式(2), (3)中 $W(k)$, $V(k)$ 是零均值方差为 $Q(k)$, $R(k)$ 的高斯白噪声。

对于观测方程式(2), 在 $\theta(k-1)$ 按泰勒级数展开并且仅仅保留线性项:

$$Z(k) = F(\hat{\theta}(k-1), k) + \bar{F}(\hat{\theta}(k-1), k)(\theta(k) - \hat{\theta}(k-1)) - V(k)$$

$$\text{其中的 } \bar{F}(\hat{\theta}(k-1), k) = \left. \frac{\partial F(\theta(k), k)}{\partial \theta(k)} \right|_{\theta(k) = \hat{\theta}(k-1)}$$

上式可进一步简化为

$$Z(k) = \bar{F}(\hat{\theta}(k-1), k)\theta(k) + V(k) + Y(k) \quad (4)$$

其中的 $Y(k) = F(\hat{\theta}(k-1), k) - \bar{F}(\hat{\theta}(k-1), k)\hat{\theta}(k-1)$, 式(4)就是线性化后的观测方程,根据推广卡尔曼滤波算法^[1]得前向神经网络的全局学习算法:

$$\theta(k+1) = \theta(k-1|k) + K(k+1)\hat{Z}(k+1|k) \quad (5)$$

$$\hat{Z}(k+1|k) = Z(k-1) - \bar{F}(\hat{\theta}(k), k)\theta(k+1|k) - Y(k+1) \quad (6)$$

$$\hat{\theta}(k-1|k) = \hat{\theta}(k) \quad (7)$$

$$K(k+1) = P(k-1|k)\bar{F}^T(\hat{\theta}(k), k)[\bar{F}(\hat{\theta}(k), k)P(k-1|k)\bar{F}^T(\hat{\theta}(k), k) + R(k+1)]^{-1} \quad (8)$$

$$P(k+1|k) = P(k) + Q(k) \quad (9)$$

$$P(k+1) = (I - K(k-1)\bar{F}^T(\hat{\theta}(k), k) \cdot P(k-1|k) \quad (10)$$

上述全局学习算法在迭代的每一步,计算的复杂度为 $O[(LT^2)]$, 存储容量要求为 $O[(LT^2)]$, 需要进行 $L \cdot L$ 维矩阵求逆计算,在系统输出维数较高时,采用这种算法计算将会非常繁琐,下面将在全局学习算法的基础上推导出基于单个神经元的权值参数,且无矩阵求逆运算的局部化推广卡尔曼滤波学习算法。

2 局部化推广卡尔曼滤波学习算法

算法的局部化方法是将全局化指标函数的极小化问题分解为一系列的子极小化问题,如基于同层权值参数的极小化问题,基于单个神经元的权值参数的极小化问题,直至基于单个权值参数的极小化问题,为了消除矩阵求逆运算,采用基于单个神经元的权值参数的极小化策略,将每一神经元看成一个动态系统,并使用全局误差的局部化量测。

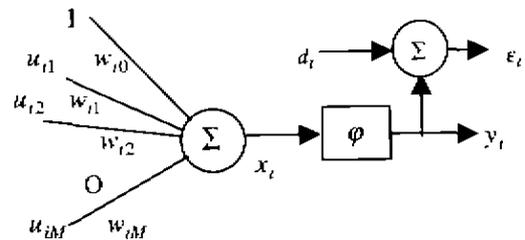


图 1 局部非线性子系统

考虑如图 1 所示的第 i 个神经元的局部非线性子系统,其权值向量 $w_i = (w_{i0}, w_{i1}, \dots, w_{iM})^T$ 是 $(M+1)$ 维,输入向量 $u_i = (1, u_{i1}, u_{i2}, \dots, u_{iM})^T$ 由 L 个全局输出误差计算这一单个子系统的局部误差^[2]:

$$\epsilon_i(k) = \sum_{l=1}^L \left| \frac{\partial e_l(k)}{\partial y_i(k)} \right| \epsilon_l(k) \quad (11)$$

式中 $\epsilon_l(k) = z_l(k) - f_l(\theta(k), I(k), z_l(k))$, $f_l(\theta(k), I(k), z_l(k))$ 分别是 $Z(k)$ 和 $F(\theta(k), I(k))$ 的第 l 个分量,所以局部状态方程和观测方程为

$$w_i(k+1) = w_i(k) + Q_i(k) \quad (12)$$

$$d_i(k) = \varphi(w_i^T(k)u_i(k)) - \epsilon_i(k) \quad (13)$$

对于观测方程式(13),在 $w_i(k-1)$ 按泰勒级数展

开并且仅仅保留线性项得到

$$d_i(k) = y_i(k) + \left[\frac{\partial \varphi}{\partial w_i(k)} \Big|_{w_i(k)=w_i(k-1)} \right]^T (w_i(k) - w_i(k-1)) + \varepsilon_i(k) \quad (14)$$

$$\text{令 } q^T(k) = \left[\frac{\partial \varphi}{\partial w(k)} \right]^T,$$

$\zeta_i(k) = y_i(k) - q^T(k)w_i(k-1)$, 式(14) 整理为

$$d_i(k) = q^T(k)w_i(k) + \zeta_i(k) + \varepsilon_i(k) \quad (15)$$

根据推广卡尔曼滤波算法得到的局部化算法为

$$\hat{w}_i(k+1) = \hat{w}_i(k-1|k) - k_i(k-1)\varepsilon_i(k+1) \quad (16)$$

$$\varepsilon_i(k+1) = \sum_{l=1}^L \frac{\partial e_l(k-1)}{\partial y_i(k-1)} \Big|_{e_l(k-1)} \quad (17)$$

$$\hat{w}_i(k+1|k) = \hat{w}_i(k) \quad (18)$$

$$k_i(k+1) = P_i(k+1|k)q_i(k-1)[1 - q_i^T(k+1)P_i(k-1|k)q_i(k-1)]^{-1} \quad (19)$$

$$P_i(k+1|k) = P_i(k) - Q_i(k) \quad (20)$$

$$P_i(k+1) = (I - k_i(k+1)q_i^T(k-1)) \cdot P_i(k+1|k) \quad (21)$$

$$Q_i(k+1) = \frac{1}{k+1} [kQ_i(k) + k_i(k+1) \cdot \varepsilon_i(k+1)\varepsilon_i^T(k+1)K_i^T(K+1) + P_i(k+1) - P_i(k)] \quad (22)$$

算法的初始化为: $P_i(0) = \delta^{-1}I$, 其中 $\delta > 0$ 为任意小的数, $w_i(0)$ 取不为 0 的小随机数, 噪声统计 Q 采用文[6] 的递推次优无偏 MAP 估计, Q 的初值采用试探法给出. 此局部化算法中,

$P_i(k+1)$ 是 $(M_i - 1) \times (M_i + 1)$ 矩阵, $k_i(k+1)$ 是 $M_i + 1$ 维向量, $1 + q_i^T(k-1)P_i(k+1|k)q_i(k-1)$ 是标量, 不需矩阵求逆, 且本算法比全局化算法有更小的计算复杂度和存储容量要求.

下面推导 $q^T(k)$ 的求法, 设前向网络采用 Sigmoid 函数 $\varphi(x) = 1/(1 + e^{-x})$ 为神经元激活函数, 链导法则由图 2 所示

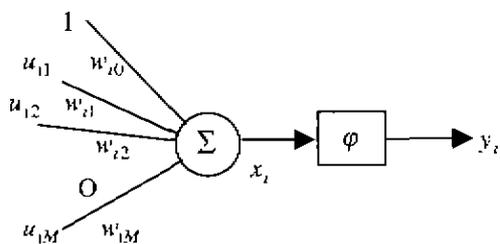


图 2 梯度的计算

$$1) \frac{\partial y_i}{\partial w_i} = \frac{\partial y_i}{\partial x_i} \cdot \frac{\partial x_i}{\partial w_i} \text{ 且 } \frac{\partial y_i}{\partial x_i} = y_i(1 - y_i), \frac{\partial x_i}{\partial w_i} = u_i$$

$$2) \frac{\partial y_i}{\partial u_i} = \frac{\partial y_i}{\partial x_i} \cdot \frac{\partial x_i}{\partial u_i} \text{ 且 } \frac{\partial x_i}{\partial u_i} = w_i$$

所以 $q_i^T(K) = y_i(k)(1 - y_i(k))u_i(k)$, 式中,

$$w_i = (w_{i1}, w_{i2}, \dots, w_{iM})^T,$$

$$u_i = (1, u_{i1}, u_{i2}, \dots, u_{iM})^T$$

3 算法分析

表 1 给出了训练一个具有 S 个神经元, L 维输出, 每个神经元具有 M 个连接权的神经网络每次迭代的计算复杂度和存储容量要求, 由此可以看出, 局部算法无矩阵求逆运算, 且比全局算法具有更小的计算复杂度和存储容量要求.

表 1 算法复杂度比较

| 算 法 | 复杂度 | 存储容量要求 | 其 他 |
|-------|-------------|-------------|---------------------|
| BP | $O(MS)$ | | |
| GEKF* | $O[(MS)^2]$ | $O[(MS)^2]$ | $L \cdot L$ 维矩阵求逆运算 |
| NEKF* | $O(M^2S)$ | $O(M^2S)$ | $L \cdot L$ 维矩阵求逆运算 |
| 本文算法 | $O(M^2S)$ | $O(M^2S)$ | |

* 见文[7].

4 仿真结果

为了验证本文提出的方法在对非线性时变系统辨识的效果, 给出了两个仿真例子, 一个是结构固定参数时变的非线性系统, 另一个是结构变化的非线性系统, 一律采用只有一个隐层的前向神经网络, 隐含层有 4 个单元组成, 输出层采用线性输出.

试验 1 首先考查一下在均匀分布的随机输入情况下, 前向网络进行时变系统建模的效果. 对于时变系统:

$$y(k) = 0.2y(k-1) - 0.8\sin(0.1\pi k)x(k)$$

输入 $x(k)$ 是 0-1 之间均匀分布的随机信号, 初始条件: $y(0) = 0.2$ 其辨识效果如图 3(a)

(b) 所示, 均方差为 0.0146.

$$\text{试验 2 } y(k) = y(k-1) - 0.2g[y(k-2)]y(k-2) + 0.1v(k)$$

$$e_{\pm}^T y(k-2) = \begin{cases} 1 + 0.5y^2(k-2) & y(k-2) \leq 4 \\ 3 & y(k-2) > 4 \end{cases}$$

初始条件: $y(0) = y(-1) = 0, v(k)$ 是具有

零均值和标准差为 1 的独立高斯噪声,其辨识效

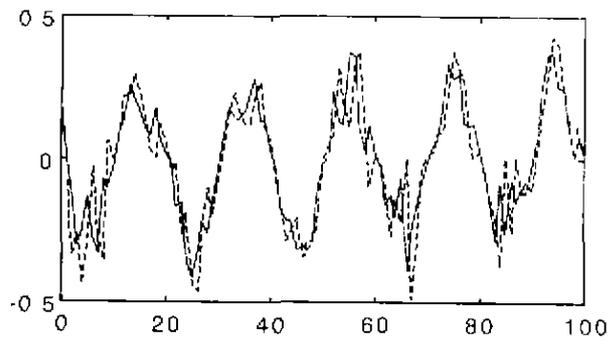


图 1a 前向网络跟踪时变系统特性曲线

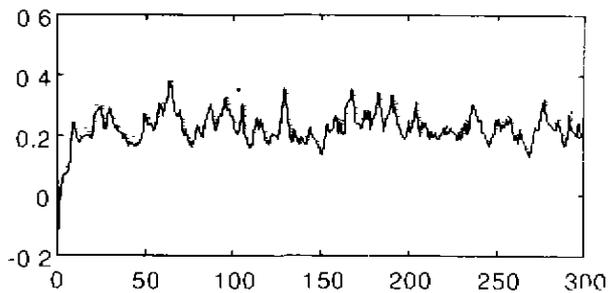


图 1b 前向网络跟踪时变系统特性曲线

果如图 1a)(b) 所示,均方差为 0.000189。

由上面的仿真试验结果可以看出采用局部化推广卡尔曼滤波算法的前向神经网络能够很好地跟踪时变系统特性的变化,且所采用的例子注重代表性,既有参数时变的非线性系统,又有结构

变化的非线性系统,实验结果表明本文提出的方法对非线性时变系统的辨识不仅具有较好的效果,而且算法更简洁快速。

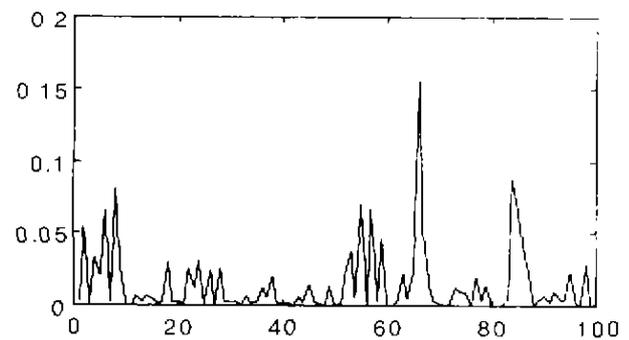


图 2a 每一时刻的方差

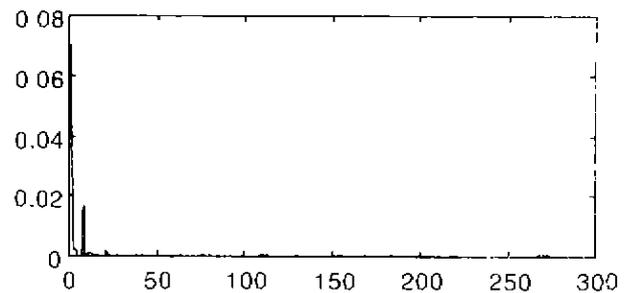


图 2b 每一时刻的方差

化的非线性系统,实验结果表明本文提出的方法对非线性时变系统的辨识不仅具有较好的效果,而且算法更简洁快速。

参考文献:

- [1] Narendra K S, Parthasarathy K, et al. Identification and control of dynamic systems using neural networks. J. IEEE Trans. Neural Networks, 1990, 1(1): 4-17.
- [2] Singhal S, Wu L. Training feed forward networks with the extended Kalman filter[C]. Proceedings of IJCI International Conference on Acoustics, Speech, and Signal Processing, Glasgow, Scotland, 1989, 1187-1190.
- [3] Keigo Watanabe, Toshiyuki Uehara. Stochastic learning algorithm of layered neural networks via extended Kalman filter[J]. International Journal of System Science, 1991, 22(11): 773-788.
- [4] Kadiramanathan, Niranjan. A function estimation approach to sequential learning with neural networks. IJ Systems of Computation, 1991, 5, 951-977.
- [5] Anderson B D O, Moore J B. Optimal Filtering[M]. Prentice Hall, 1979, 281-283.
- [6] Sage A P, Husa G W. Adaptive Filtering with Unknown Prior Statistics[C]. Proc. Annual Control Conference, 1969, 750-769.
- [7] Sheh S, P. Jaeger F, Fatouh M, et al. Optimal filtering algorithm for fast learning in feedforward neural networks[J]. Neural Networks, 1992, 5: 779-787.
- [8] 李波, 丁秀峰. 径向基函数网络的递进优化学习方法[J]. 管理科学学报, 1999, 20(1): 22-27.

this way, build a new relation between knowledge engineers and experts in which experts play a decisive role while knowledge engineers are accessory. i. e., it is expert-oriented during the knowledge acquisition course.

Key words: expert-oriented; knowledge base; genetic algorithm; optimization

(上接第24页)

[21] 袁安照, 余光胜. 现代企业组织创新[M]. 太原: 山西经济出版社, 1998

[22] 汪丁丁. 回家的路: 经济学家的思想轨迹[M]. 北京: 中国社会科学出版社, 1998

[23] 斯蒂芬·P·罗宾斯. 管理学[M]. 第4版. 北京: 中国人民大学出版社, Englewood Cliffs, NJ: Prentice Hall, 1997

[24] 何清涟. 现代化的陷阱——当代中国的经济社会问题[M]. 北京: 今日中国出版社, 1998

Mountains-forming: new logic of knowledge and learning based enterprise

RUI Ming-jie, FAN Sheng-jun

School of Management, Fudan University, Shanghai 200433, China

Abstract: For the coming knowledge economy, we need new enterprise logic different from any others before. We summarize here the new logic of enterprise as the knowledge and learning based, compare it with the mountains-forming campaign appeared in the geological history, and make a frame study of it. The theoretical base of the new view of enterprise is knowledge and learning theories, its typical organization pattern is the team-based organization, the essence of management is to accomplish competence promoting through transforming and combination of knowledges, marketing theme is to pursue the knowledge transcending. By all these things, we have portrayed the outline of the new enterprise logic.

Key words: knowledge and learning based enterprise; team-based organization; competence promoting; knowledge transcending

(上接第39页)

Nonlinear time-varying systems identification by feedforward neural networks

GU Cheng-kui, WANG Zheng-ou

Institute of System Engineering, Tianjin University, Tianjin 300072, China

Abstract: A new identification method based on feed-forward networks is presented for nonlinear time-varying systems. We apply local extended Kalman Algorithm to train feed-forward networks, this algorithm needs no matrix inversion computation and has the higher convergence speed and the smaller storage required in comparison to the global extended Kalman Algorithm. Simulation results show the present method has better effect on nonlinear time-varying systems identification.

Key words: nonlinear time-varying systems; feed-forward neural networks; local algorithm; Kalman filtering