文章编号: 1001-0920(2007)07-0765-04

基于自适应评价的非线性系统神经网络控制

陈宗海, 文 锋, 王智灵
(中国科学技术大学 自动化系, 合肥 230027)

摘 要: 针对一类非线性系统,提出了一种自适应评价方法. 该方法可以控制系统输出对参考信号进行跟踪,其评价函数可直接解析求出. 该方法只需一个动作网络用于产生控制动作,并且方法中的网络权值初始化可随机选取. 使用 Lyapunov 方法对整个系统的动态性能进行分析,证明了在一定条件下此方法能保证闭环误差及网络权值一致最终有界. 仿真结果与理论分析相一致,证明了所提出方法的有效性.

关键词: 自适应评价; 神经网络; Lyapunov 方法; 强化学习
中图分类号: TP13 **文献标识码**: A

Neural network control of nonlinear systems based on adaptive critic

CHEN Zong-hai, WEN Feng, WANG Zhi-ling
(Department of Automation, University of Science and Technology of China, Hefei 230027, China. Correspondent: CHEN Zong-hai, E-mail: chenzh@ustc.edu.cn)

Abstract: A novel adaptive critic method is proposed for a class of nonlinear systems, by using which system output is controlled to track a reference trajectory. In this method, critic function can be solved analytically and only one action network is needed to generate control action. Moreover, network weights are initialized randomly. The analysis of dynamic performance of the overall system is performed by using Lyapunov method, which proved that by using the proposed method uniform ultimate boundness of close-loop error is guaranteed under certain conditions. Simulation results are consistent with theoretical analysis and show the effectiveness of the method.

Key words: Adaptive critic; Neural networks; Lyapunov method; Reinforcement learning

1 引 言

对非线性系统进行控制的常规方法是在操作点附近对系统进行线性化;然后使用线性控制方法设计控制器. 当系统状态不在线性范围内时,控制器的性能会大大下降. 非线性控制方法虽然能够改进系统的瞬态性能,但其本身的结构和实现要比线性控制器复杂得多. 反馈线性化方法虽然能够抵消系统的非线性部分,但需要知道系统的精确参数,这在实际中难以保证. 基于学习的神经网络控制方法,利用神经网络的非线性逼近能力,在常规解析方法之外提供了另一种选择. 一般而言,大都首先使用一个神经网络辨识得到系统模型;然后使用该模型训练一个神经网络控制器,但这种方法通常不能保证系统的稳定性[1].

Prokhorov[2]等使用自适应评价方法训练神经网络控制器. 该方法基于动态规划的思想,是 TD (Temporal Difference)强化学习算法的推广形式. 其中评价网络对动作网络性能进行评估,训练是使其输出满足 Bellman 方程;动作网络则产生控制动作,根据评价网络的评价调整网络权值,使控制性能达到最优. 根据评价网络的训练方法不同,可分为 HDP(Heuristic dynamic programming)方法、DHP (Dual heuristic programming)方法、GDHP(Global dual heuristic programming)方法. TD 强化学习算法通常归结为 HDP 方法的一种. 一系列实践表明,自适应评价方法具有能克服系统的不稳定性,不受系统辨识误差的影响,能处理输入不确定性以及计算量小,适于在线训练等特点. 但上述特点多为经验总结,缺少严格的理论证明.

在对象模型完全已知的情况下,Landelius[3]给出了针对 LQR 控制的自适应评价方法的收敛证明;Prokhorov 等[4]则给出了控制以马尔科夫链表

收稿日期: 2006-04-03; **修回日期**: 2006-07-13.
基金项目: 国家自然科学基金项目(60575033).
作者简介: 陈宗海(1963—),男,安徽桐城人,教授,博士生导师,从事复杂系统的建模、仿真与优化控制等研究;
文锋(1978—),男,安徽庐江人,博士生,从事智能控制等研究.

示的对象的收敛证明. Liu 等[5]从求解最优控制中的 HJB 方程的角度推导了自适应评价方法的收敛证明,适用于更一般的对象,但其推导也要求对象模型完全已知. 在模型未知的情况下,Bradtke[6]给出了针对 LQR 控制的 HDP 方法收敛的证明,但其证明存在漏洞,作者已在文献[7]中给出了证明,并在文献[8-10]中进行了理论和应用研究. Jagannathan[1]则针对一类非线性问题,提出了一种被证明一致且最终有界的自适应评价方法. 虽然其中有评价网络和动作网络,但其网络的训练方法更大程度上属于特殊设计,与标准的自适应评价方法存在很大的差别.

本文针对一类非线性对象,基于反馈控制的思想提出一种新的自适应评价方法. 该方法利用反馈控制的形式,直接解析求得评价参数(不必使用评价网络,动作网络的训练使得评价输出最小);然后使用 Lyapunov 方法证明了在一定条件下整个系统的输出一致最终有界.

2 非线性系统描述

对于如下形式的非线性系统:

$$\begin{aligned} x_1(k+1) &= x_2(k),\\ &\cdots\\ x_n(k+1) &= f(x(k)) + u(k) + d(k), \end{aligned} \tag{1}$$

状态向量

$$x(k) = [x_1^{\mathrm{T}}(k), x_2^{\mathrm{T}}(k), \cdots, x_n^{\mathrm{T}}(k)]^{\mathrm{T}}.$$

其中:$x_i(k) \in R^m, i = 1,2,\cdots,n$;$f(x(k)) \in R^m$ 为未知非线性函数;$u(k) \in R^m$ 为输入向量;$d(k) \in R^m$ 为未知有界扰动向量,其界假设已知且为常数,即 $\| d(k) \| \leq d_M$,此处 $\| \cdot \|$ 为 Frobenius 范数(下同).

给定参考轨迹 $x_{nd}(k)$,定义跟踪误差 $e_n(k) \in R^m$ 为

$$e_n(k) = x_n(k) - x_{nd}(k), \tag{2}$$

以及滤波后的跟踪误差 $r(k) \in R^m$ 为

$$r(k) = [\boldsymbol{\rho} \quad I_m] e(k). \tag{3}$$

式中:$e(k) = [e_1^{\mathrm{T}}(k), e_2^{\mathrm{T}}(k), \cdots, e_n^{\mathrm{T}}(k)]^{\mathrm{T}}$;$e_i(k+1) = e_{i+1}(k) \in R^m (i = 1,2,\cdots,n-1)$ 为误差 $e_n(k)$ 的延时值;参数矩阵 $\boldsymbol{\rho} = [\rho_{n-1}, \rho_{n-2}, \cdots, \rho_1] \in R^{m \times m(n-1)}$,$\rho_i \in R^{m \times m} (i = 1,2,\cdots,n-1)$ 分别为常数对角正定矩阵,且其特征值位于单位圆内;$I_m \in R^{m \times m}$ 为单位矩阵. 由系统方程(1)可得滤波误差表达式为

$$\begin{aligned} r(k+1) &= [\boldsymbol{\rho} \quad I_m] e(k+1) =\\ & f(x(k)) - x_{nd}(k+1) + \rho_1 e_n(k) +\\ & \cdots + \rho_{n-1} e_2(k) + u(k) + d(k). \end{aligned} \tag{4}$$

然而,对于一般离散时间非线性系统

$$\begin{aligned} x(k+1) &= f(x(k), u(k)),\\ y(k) &= h(x(k)). \end{aligned}$$

其中:$x(k)$ 为状态向量,$u(k)$ 为输入向量,$y(k)$ 为输出向量. 有如下结论:系统的解一致最终有界(UUB),如果 $\forall x(k_0) = x_0$,$\exists \mu \geq 0$ 及 $N(\mu, x_0)$,使得 $\forall k \geq k_0 + N$,则有 $\| x(k) \| \leq \mu$ 成立.

3 Jagannathan 的自适应评价方法[1]

Jagannathan[1] 提出了一种自适应评价神经网络控制器,所使用的神经网络是只有一个隐含层的多层感知器网络.

定义系统的控制输入 $u(k)$ 为

$$\begin{aligned} u(k) &=\\ & x_{nd}(k+1) - \hat{f}(x(k)) + k_v r(k) -\\ & \boldsymbol{\rho} e_n(k) - \cdots - \rho_{n-1} e_2(k). \end{aligned} \tag{5}$$

其中:$\hat{f}(x(k))$ 为对于未知函数 $f(x(k))$ 的估计;k_v 为对角比例矩阵,可表示为某一常数与单位矩阵的积. 则式(4)可写为

$$r(k+1) = k_v r(k) - \tilde{f}(x(k)) + d(k), \tag{6}$$

其中 $\tilde{f}(x(k)) = \hat{f}(x(k)) - f(x(k))$ 为函数的估计误差.

使用自适应评价控制器,其评价网络为

$$\begin{aligned} R(k) &= \hat{w}_1^{\mathrm{T}}(k) \sigma(v_1^{\mathrm{T}}(E \cdot r(k))) =\\ & \hat{w}_1^{\mathrm{T}}(k) \sigma(\cdot), \end{aligned}$$

动作网络为

$$\begin{aligned} \hat{f}(x(k)) &= \hat{w}_2^{\mathrm{T}}(k) \phi(v_2^{\mathrm{T}}(x(k))) =\\ & \hat{w}_2^{\mathrm{T}}(k) \phi(\cdot). \end{aligned}$$

式中:v_1,v_2 为固定的输入层权值;$\hat{w}_1$,$\hat{w}_2$ 为可变的输出层权值;$\sigma(\cdot)$ 和 $\phi(\cdot)$ 为隐含层的激活函数. 其中评价网络的输入为加权滤波误差 $E \cdot r(k)$(E 为权值矩阵),输出为评价信号 $R(k)$. 动作网络输入为状态向量 $x(k)$,输出为函数 $f(x(k))$ 的估计 $\hat{f}(x(k))$.

评价者网络的训练为

$$\begin{aligned} \hat{w}_1(k+1) &=\\ & \hat{w}_1(k) - \alpha_1 \sigma(\cdot)(k_v r(k) + R(k))^{\mathrm{T}}, \end{aligned}$$

动作网络的训练为

$$\begin{aligned} \hat{w}_2(k+1) &=\\ & \hat{w}_2(k) + \alpha_2 \phi(\cdot)(r(k+1) + R(k))^{\mathrm{T}}, \end{aligned}$$

其中 α_1 和 α_2 为学习率.

Jagannathan[1] 证明了在一定条件下这种自适应评价神经网络控制器能够保证跟踪误差和网络权值一致最终有界(UUB). 然而,从 Jagannathan[1] 的神经网络控制器的结构看,虽然采用了评价网络和动作网络,但其训练方案的设计更大程度上属于自适应控制的设计方法,难以从自适应评价方法的角

度加以理解. 为此,本文提出一种基于强化学习的自适应评价控制器.

4　基于自适应评价的神经网络控制器

本文提出的基于自适应评价的神经网络控制器结构如图 1 所示. 其中:动作网络产生控制动作;评价者(评价网络) 对动作网络的性能进行评估. 根据评价者产生的评估信号,可以对动作网络进行调整. 该方法中,由于评价者的参数可以直接求出,不必使用神经网络进行逼近,省略了评价网络的训练过程,同时也节省了计算量.

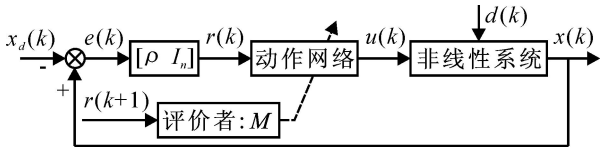

图 1　基于自适应评价的神经网络控制器结构

4.1　评价者设计

自适应评价方法是基于动态规划的,它使用神经网络表示系统状态,避免了维数灾难问题. 定义 cost-to-go 函数为

$$V(k) = \sum_{t=k}^{\infty} \gamma^{t-k} r^{\mathrm{T}}(t+k)\, Q\, r(t+k). \tag{7}$$

其中:矩阵 $Q > 0$ 为性能指标矩阵,折扣系数 $0 \leq \gamma \leq 1$.

对于本文研究的一类非线性系统,其理想控制输入为

$$u_d = x_{nd}(k+1) - f(x(k)) + k_v r(k) - \rho_1 e_n(k) - \cdots - \rho_{n-1} e_2(k). \tag{8}$$

引理 1[11]　对于系统(1),使用式(8) 的控制动作,则由式(4) 表示的闭环跟踪误差系统稳定,如果有 $k_v^{\mathrm{T}} k_v < I$.

在理想控制输入下,系统的闭环误差方程为

$$r(k+1) = k_v r(k) + d(k). \tag{9}$$

由引理 1,使用理想控制输入(8),闭环系统稳定,cost-to-go 函数(7) 收敛. 显然 cost-to-go 函数为误差的二次函数形式,可表示为

$$V(k) = r^{\mathrm{T}}(k)\, M\, r(k), \tag{10}$$

其中 $M \in R^{m\times m}$ 为系数矩阵.

由 Bellman 原理,有

$$V(k) = r^{\mathrm{T}}(k)\, Q\, r(k) + \gamma V(k+1). \tag{11}$$

由式(9) ～ (11),不考虑 $d(k)$ 时,有

$$M = Q + \gamma k_v^{\mathrm{T}} M k_v, \tag{12}$$

式中除矩阵 M 外,其余皆已知. 可直接求解矩阵 M,$M = (I - \gamma k_v^{\mathrm{T}} k_v)^{-1} Q$,其中利用了 Q 和 k_v 均为对角方阵的事实. 假设矩阵 M 中的最大元素和最小元素分别为 $M_{\max}$ 和 $M_{\min}$,则显然有 $M_{\max} \geq M_{\min} > 0$.

4.2　动作网络设计

上述推导求解 cost-to-go 函数为理想控制输入下的结果. 但实际中函数 $f(x(k))$ 未知,需要使用动作网络来近似理想控制动作(8),即

$$\hat{u}(k) = \hat{w}^{\mathrm{T}}(k)\, \phi(v^{\mathrm{T}}(s(k))) = \hat{w}^{\mathrm{T}}(k)\, \phi(k), \tag{13}$$

其中 $s(k) = [x^{\mathrm{T}}(k), r^{\mathrm{T}}(k)]^{\mathrm{T}}$ 为网络输入. 设未知的网络权值的界为 $w_{\max}$,即有 $\| w \| \leq w_{\max}$,而激活函数也有已知界,为 $\| \phi(k) \| \leq \phi_{\max}$,并假设在理想权值下,动作网络的逼近误差为 $\varepsilon(k)$,则有

$$u_d(k) = w^{\mathrm{T}} \phi(k) + \varepsilon(k). \tag{14}$$

使用动作网络的输出作为系统控制输入,则闭环跟踪误差方程为

$$\begin{aligned} r(k+1) &= \\ x(k+1) - x_{nd}(k+1) &= \\ f(x(k)) + \rho_1 e_n(k) + \cdots + & \\ \rho_{n-1} e_2(k) + d(k) - x_{nd}(k+1) + & \\ \hat{w}^{\mathrm{T}}(k)\, \phi(k) + u_d(k) - w^{\mathrm{T}} \phi(k) - \varepsilon(k) &= \\ k_v r(k) + \zeta(k) + d(k) - \varepsilon(k). & \end{aligned} \tag{15}$$

其中:$\zeta(k) = \tilde{w}(k)^{\mathrm{T}} \phi(k)$,$\tilde{w}(k) = \hat{w}(k) - w$.

4.3　权值更新方法

动作网络的训练应使 cost-to-go 函数 $V(k)$ 最小,即

$$\partial V(k) / \partial \hat{u}(k) = 0. \tag{16}$$

而由式(11),有

$$\frac{\partial V(k)}{\partial \hat{u}(k)} = 0 + \gamma \frac{\partial V(k+1)}{\partial \hat{u}(k)} = 2\gamma \frac{\partial r(k+1)}{\partial \hat{u}(k)} M r(k+1) = 2\gamma M r(k+1), \tag{17}$$

由式(15) 的关系可知 $\partial r(k+1) / \partial \hat{u}(k) = 1$. 因此,可得到动作网络权值更新方程

$$\hat{w}(k+1) = \hat{w}(k) - \alpha \phi(k)\, r^{\mathrm{T}}(k+1)\, M^{\mathrm{T}}, \tag{18}$$

其中 α 为动作网络学习率.

定理 1　滤波跟踪误差 $r(k)$ 和网络权值估计 $\hat{w}(k)$ 一致最终有界,条件是设计参数满足

$$\alpha \| \phi(k) \|^2 < M_{\min} / M_{\max}^2, \tag{19}$$

$$(M_{\max} k_{v\max}^2) / M_{\min} < 1/8, \tag{20}$$

其中 $k_{v\max}$ 为比例矩阵 k_v 的最大特征值.

证明　定义 Lyapunov 函数为

$$J = \frac{M_{\min}}{2} r^{\mathrm{T}}(k)\, r(k) + \frac{1}{\alpha} \mathrm{tr}\{ \tilde{w}^{\mathrm{T}}(k)\, \tilde{w}(k) \}, \tag{21}$$

其一阶差分为 $\Delta J = \Delta J_1 + \Delta J_2$. 根据式(15) 给出的跟踪误差方程,求得 ΔJ_1 为

$$\Delta J_1 = \frac{M_{\min}}{2} \{ r^{\mathrm{T}}(k+1)\, r(k+1) - r^{\mathrm{T}}(k)\, r(k) \}.$$

根据动作网络的更新方程(18),求得

$$\Delta J_2 = \frac{1}{\alpha}\mathrm{tr}\{\overline{w}^{\mathrm{T}}(k+1)\overline{w}(k+1) - \overline{w}^{\mathrm{T}}(k)\overline{w}(k)\}.$$

合并上述两项差分,并进行推导,可得到

$$\Delta J = \Delta J_2 + \Delta J_1 \leq -(M_{\min} - \alpha\beta^2_{\max}M^2_{\max})\|k_v r(k) + d(k) + \zeta(k)\|^2 - (M_{\min}/2 - 2M_{\max}k^2_{v\max})\|r(k)\|^2 + 2M_{\max}\|d(k)\|^2.$$

在给定条件(19)和(20)下,且当

$$\|\zeta(k)\| \geq \left(\frac{(4M_{\max} + 3\alpha\beta^2_{\max}M^2_{\max} - 3M_{\min})\tau^2_{\max}}{M_{\min} - \alpha\beta^2_{\max}M^2_{\max}}\right)^{1/2} \quad (22)$$

时,有 $\Delta J \leq 0$ 成立. 一般情况下,在紧致集中,条件(19)和(20)以及(22)成立. 因此,根据Lyapunov扩展定理,滤波后的跟踪误差以及动作网络估计误差是一致最终有界的. □

5 仿 真

非线性系统为

$$x_1(k+1) = x_2(k),$$
$$x_2(k+1) = f(x(k)) + u(k),$$

其中 $f(x(k)) = -\frac{3}{16}\frac{x_1(k)}{1+x_2^2(k)} + x_2(k)$.

控制目标是在自适应评价神经网络控制器的控制下跟踪一个参考信号. 参考信号为 $x_{2d} = \sin(\omega t_k + \theta)$,其中 $\omega = 0.5$,$\theta = \pi/2$. 采样间隔为 $T = 50$ ms,时间长度为250 s. 控制器参数设置为 $k_v = 0.2$,$\gamma = 0.95$,学习率 $\alpha = 0.1$. 神经网络隐含层的激活函数为tansig函数,隐含层节点个数为10,Critic函数 $M = (I - \gamma k_v^{\mathrm{T}}k_v)^{-1}Q$,其中 Q 为单位矩阵.

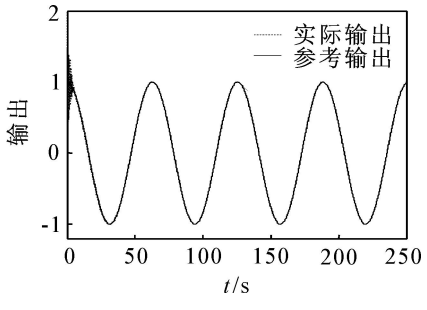

图2 系统输出与参考输出

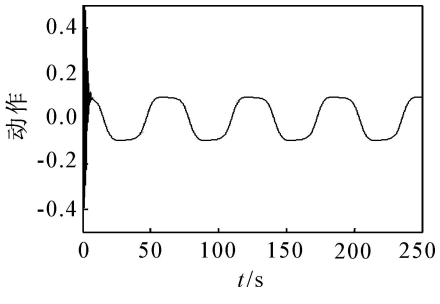

图3 控制动作

图2和图3给出了本文提出的自适应评价方法的控制结果. 图2为系统输出与参考输出对比,图3给出了控制信号. 由图可见,在初始阶段控制器的控制效果不佳,但时间很短(在10个采样周期以内);系统输出与参考信号之间的误差较大,但时间很短(在10个采样周期以内);经过学习,系统输出很快达到参考信号,其跟踪误差趋于零. 可见,该方法能够有效地控制对象跟踪参考曲线.

6 结 语

本文使用自适应评价方法对一类非线性对象进行控制,不需要系统模型或初始学习,并通过严格的数学推导,得到了一种新型的权值更新方法. 该方法在线调整动作网络,其性能良好,而且对网络权值的初始化没有特殊要求. 但初值对对象控制的初始阶段有影响,为缩短过渡过程,提高学习效果,在实际应用中网络权值的初始值可采用上次控制过程中训练的有效值.

参考文献(References)

[1] Jagannathan S. Adaptive critic neural network-based controller for nonlinear systems[C]. Proc of IEEE Int Symposium on Intelligent Control. Vancouver, 2002: 303-308.

[2] Prokhorov D V, Wunsch D C. Adaptive critic designs [J]. IEEE Trans on Neural Networks, 1997, 8(5): 997-1007.

[3] Landelius T. Reinforcement learning and distributed local model synthesis [D]. Sweden: Linkoping University, 1997.

[4] Prokhorov D V, Wunsch D C. Convergence of critic-based training [C]. Proc IEEE Int Conf System Management Cybernation. Tokyo, 1997, 4: 3057-3060.

[5] Xin Liu, Balakrishnan S N. Convergence analysis of adaptive critic based optimal control [C]. Proc of American Control Conf. Chicago, 2000: 1929-1933.

[6] Bradtke S J. Incremental dynamic programming for on-line adaptive optimal control[D]. Massa: University of Massachusetts, 1994.

[7] 文锋, 陈宗海, 周光明, 等. 一种用于LQR控制问题的强化学习方法[J]. 模式识别与人工智能, 2006, 19(3): 406-411.
(Wen Feng, Chen Zong-hai, Zhou Guang-ming, et al. A reinforcement learning method for LQR control problem [J]. Pattern Recognition and Artificial Intelligence, 2006: 19(3), 406-411.)

(下转第773页)

故障. 考察图 4, 在故障持续时间较短的情况下, 相比于传统 PID 控制器和分散 H_∞ 控制器, MMSC 能更快、更好地恢复到故障前的工作点. 考察图 5, 在故障持续时间稍长的情况下, 传统 PID 控制器不能使系统稳定, 而 MMSC 仍然能比分散 H_∞ 控制器更快、更好地使系统稳定, 从而表明 MMSC 具有优良的控制性能.

7　结　　语

本文研究了一种用于汽门控制的 MMSC, 并给出了系统结构和学习算法. 与其他控制方法相比, 其特点在于: 1) MMSC 能够适应参数大范围变化, 子模型模糊规则是针对各种工况设计的, 针对性强; 2) 结合了模糊逻辑与 SVM 两种算法的优点, 规则设计简单、自学习能力强; 3) 基于模糊逻辑计算子模型匹配程度, 匹配程度决定了 MMSC 的集成加权系数, 有利于消除模型切换振荡; 4) 离线算法建立 MMSC 子模型及其控制规则, 在线算法能够优化控制规则, 并满足实时控制要求. 仿真实验验证了该控制器具有较强的稳定控制性能.

参考文献(References)

[1] Chen Q, Tan S H, Han Y D, et al. Adaptive fuzzy scheme for efficient, fast valving control [J]. Control Engineering Practice, 1997, 5(6): 811-821.

[2] Han Y D, Wang Z H, Chen Q, et al. Artificial-neural-network-based fast valving control in a power-generation system [J]. Engineering Applications of Artificial Intelligence, 1997, 10(2): 139-155.

[3] Zhang L Z, Kang J P, Lin X S, et al. Application of neural networks trained with an improved conjugate gradient algorithm to the turbine fast valving control [C]. Proc of 2000 Int Conf on Power System Technology. Perth, 2000: 1679-1682.

[4] Liu G X, Lin X S, Yang Q X, et al. Investigation of turbine valving control with Lyapunov theory[C]. Proc of 4th Int Conf on Advances in Power System Control, Operation and Management. Hong Kong, 1997: 505-508.

[5] 于达仁, 杨永滨, 崔涛, 等. 大范围线性化最优鲁棒容错快关控制系统的设计[J]. 中国电机工程学报, 2002, 22(9): 25-29.
(Yu Da-ren, Yang Yong-bin, Cui Tao, et al. Optimal robust fault-tolerance fast valving control system design via large-scale linearization[J]. J of the CSEE, 2002, 22(9): 25-29).

[6] 席在荣, 程代展. 多机非线性系统分散汽门 H_∞ 控制器[J]. 电力系统自动化, 2002, 26(21): 7-11.
(Xi Zai-rong, Cheng Dai-zhan. Decentralized steam valving controller for nonlinear multi-machine power systems[J]. Automation of Electric Power Systems, 2002, 26(21): 7-11.)

[7] Vapnik V. An overview of statistical learning theory [J]. IEEE Trans on Neural Networks, 1999, 10 (5): 988-999.

[8] Lin C T, Yeh C M, Liang S F, et al. Support-vector-based fuzzy neural network for pattern classification[J]. IEEE Trans on Fuzzy Systems, 2006, 14(1): 31-41.

[9] Xie Xuanli Lisa, Beni Gerardo. A validity measure for fuzzy clustering[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1991, 13(8): 841-847.

[10] 袁小芳, 王耀南, 孙炜. 支持向量机-模糊推理自学习控制器设计[J]. 控制理论与应用, 2006, 23(1): 1-6.
(Yuan Xiao-fang, Wang Yao-nan, Sun Wei. Self-learning controller using support vector machines and fuzzy inference system [J]. Control Theory and Application, 2006, 23(1): 1-6.)

(上接第 768 页)

[8] 文锋, 陈宗海, 卓睿, 等. 连续状态自适应离散化的基于 K 均值聚类的强化学习方法[J]. 控制与决策, 2006, 21(2): 143-147.
(Wen Feng, Chen Zong-hai, Zhuo Rui, et al. Reinforcement learning method of continuous state adaptively discretized based on K-means clustering[J]. Control and Decision, 2006, 21(2): 143-147.)

[9] 任燚, 陈宗海. 基于强化学习算法的多机器人系统的冲突消解策略[J]. 控制与决策, 2006, 21(4): 430-439.
(Ren Yi, Chen Zong-hai. Interference solving strategy in multiple robot system based on reinforcement learning algorithm[J]. Control and Decision, 2006, 21(4): 430-434, 439.)

[10] 陈宗海, 文锋, 聂建斌, 等. 基于节点生长 k-均值聚类算法的强化学习方法[J]. 计算机研究与发展, 2006, 34(4): 661-666.
(Chen Zong-hai, Wen Feng, Nie Jian-bin, et al. A reinforcement learning method based on node-growing k-means clustering algorithm [J]. J of Computer Research and Development, 2006, 34(4): 661-666.)