基于梯度估计的非线性系统最优控制及仿真①

缪应锋, 姚庆华, 李智雄, 宋晓轩

(云南省交通科学研究院, 昆明 650093)

摘 要: 基于自适应动态规划(ADP)执行-评价结构, 应用神经网络(NN)对非线性系统进行最优控制求解. 首先提出所求解非线性系统的一般形式; 其次给定二次正定性能指标, 求其哈密尔顿函(HJB)函数; 分别应用神经网络对执行-评价结构中的性能指标和最优控制进行逼近, 神经网络权重参数应用梯度法求得, 从而可以求得其最有控制策略. 而且对执行机构和评价机构神经网络权重参数的收敛性以及系统总体的稳定性进行了详细的分析, 证明所求控制策略可以使系统稳定; 最后, 用仿真结果来验证所提出的方法的可行性.

关键词: 非线性系统; 动态规划; 梯度算法; 哈密尔顿函数; 李雅普诺夫

Optimal Control and Simulation of Nonlinear Systems Based on Gradient Estimation

MIAO Ying-Feng, YAO Qing-Hua, LI Zhi-Xiong, SONH Xiao-Xuan

(Institute of Yunnan Traffic, Kunming 650093, China)

Abstract: This paper solves the regulation optimal control problem for nonlinear systems based on the adaptive dynamic programinf(ADP) and neural networks(NNs). Firstly, we propose the regulation nonlinear system; afterwards, the cost function and its Hamilton-Jacobi-Bellman(HJB) function were given; furthermore, an actor-critic frame structure was put forward to get the optimal control, and the neural network algorithms were utilized to approximate the performance index and the optimal control ofactor-critic structures respectively. The NN weights are estimated by gradient algorithms, so that the optimal control is obtained. Simultaneously, the stability of the whole system and the convergences of actor-critic NN weights are proved based on the Lyapunov theory. Finally, the simulation results are provided to verify the effectiveness of the proposed methods.

Key words: nonlinear system; approximate dynamic program (ADP); gradient algrothim; Lyapunov; Hamilton-Jacobi -Bellman(HJB)

最优控制求解分为线性系统和非线性系统. 线性系统用里卡提方程求解. 对于非线性系统, 近几十年有学者提出了动态规划方法[1]. 最近, 学者将增强学习方法[2]和动态规划方法结合[3], 提出了自适应动态规划方法(Adaptive dynamic Programming). Werbos[4]基于增强学习方法, 提出评价和执行网对离散系统进行动态最优求解. Lewis[5]提出了一种新的基于神经网络的自适应动态最优方法对离散非线性系统进行离线求解.

许多学者应用迭代法求解离散非线性系统最优控制[6,13], 迭代法求解系统最优控制往往因计算量大, 导致计算参数收敛速度比较慢, 系统稳定需要很长时间. 同时对于连续系统的最优控制求解最近几年才被学者研究, 但大多要求评价执行结构神经网络权重参数初始值不为0[7, 8], 而且初始输入控制不为0, 也即要求满足持续激励条件[9,10]. 很大程度上限制了该方法在实际工程中的应用.

本文基于以上存在的问题, 将自适应梯度估计算法应用于自适应动态规划方法, 同时应用评价-执行结构对非线性系统从初始状态整定到零状态的最优控制求解. 首先提出非线性系统模型和性能指标, 对应用神经网络逼近性能指标; 其次基于逼近性能指标的神

① 收稿时间:2016-03-05;收到修改稿时间:2016-04-18 [doi: 10.15888/j.cnki.csa.005421]

经网络, 用第二个神经网络逼近系统最优控制; 其中用了哈密尔顿方法对系统进行最优控制求解. 基于梯度估计算法, 同时在线估计评价结构和执行结构的神经网络权重参数. 更重要的是, 本文对所求系统的稳定性进行了详细的分析, 而且基于真实的非线性数学模型统对所提出的方法进行了仿真验证.

1 系统概述及HJB最优求解

1.1 系统方法简介

非线性系统描述如下

$$\dot{x}= f(x)+g(x) \tag{1}$$

$x\in ¡^{n}$ 是可观测的系统状态, $u\in ¡^{m}$ 是系统输入, $f(x)\in ¡^{n}$ 是系统动态. $g(x)\in ¡^{n\times m}$ 是系统输入动态.

最优控制就是给定如下性能指标, 在满 足系统性能的通知, 使得下面的性能指标最小:

$$V(x(t))=\int_{t}^{\infty} r(x(\tau),u(\tau))d\tau \tag{2}$$

其中 $r(x,u)=x^{T}Qx+u^{T}Ru$ 关于 u,x 的连续函数, Q 和 R 是相对应的正定对称矩阵. 最优控制策略 u 是将系统从初始状态镇定到 0, 同时使得系统性能指标函数(2)最小.

1.2 最优控制的一般求法

求解非线性系统最优控制, 首先的建立系统的哈密尔顿函数, 然后用神经网络逼近哈密尔顿函数中的性能指标, 进而解决“维数灾难”问题, 顺利求得系统最优控制策略, 其中用到如下两个假设:

假设 1[11]. 系统动态 $f(x)$ 在实数范围内连续可微.

假设 2[12]. 神经网络的逼近权重和误差均有界.

根据系统表达式(1)和性能指标式(2)可建立如下哈密尔顿函数:

$$H(x,u,V_x)=V_x^{T}\left[f(x)+g(x)u\right]+x^{T}Qx+u^{T}Ru \tag{3}$$

其中 $V_x @\partial V/\partial x$ 为性能指标函数对 x 的变分. 令上式等于 0, 可求得最优控制为:

$$u^{*}=-\frac{1}{2}R^{-1}g(x)^{T}\frac{\partial V^{*}(x)}{\partial x} \tag{4}$$

上式中由于计算问题, $\frac{\partial V^{*}(x)}{\partial x}$ 对非线性最优控制问题很难求解. 所以现今存在的一般方法是用一个神经网络逼近性能指标, 进而求得具体的最优控制策略.

2 基于梯度法的评价执行网及稳定性分析

由于非线性系统的复杂性, 以及在求解过程中存在的“维数灾难”问题, 对 $V_x @\partial V/\partial x$ 的求解很难实现. Dierks[13]和刘德荣[14]提出了神经网络逼近用评价结构的方法来取代上面的方法, 进而避免了以上存在的问题.

不同于离线求解系统最优控制, 基于自适应动态规划方法, 本节内容应用梯度估计算法, 对评价结构和执行结构神经网络权重参数进行在线估计, 从而在线求解了系统的最优控制. 而且与[9,13,14]相比, 被估计参数初始值都不为零, 系统初始运行 3s 之内不需要系统激励, 激励条件更加放宽, 更加具有实用价值.

2.1 基于梯度法的评价结构

基于神经网络的性能指标函数可以表示为:

$$V^{*}(x)=W_c^{T}\phi(x)+\varepsilon_v \tag{5}$$

$W\in \mathbf{R}^{l}$ 为未知的神经网络权重, $\phi(x)\in \mathbf{R}^{l\times n}$ 为输入激励, ε_v 为神经网络估计误差. l 为神经网络结点层数. 根据引理 2, 神经网络估计权重和误差都是有界的. 则性能指标对 x 的偏微分可以重新表述为:

$$\frac{\partial V^{*}(x)}{\partial x}=\nabla\phi^{T}W_c+\nabla\varepsilon_v \tag{6}$$

根据上式最优控制策略可以表示

$$u=-\frac{1}{2}R^{-1}g(x)^{T}\frac{\partial \hat{V}(x)}{\partial x}=-\frac{1}{2}R^{-1}g(x)^{T}\nabla\phi^{T}(x)\hat{W}_c \tag{7}$$

其中 $\hat{W}_c$ 为评价结构神经网络近似权重参数, 则近似的哈密尔顿函数可以表示为:

$$\begin{aligned}e=H(x,u,V_x)=&\nabla\phi^{T}(x)\hat{W}_c\left[f(x)+g(x)u\right]\\&+x^{T}Qx+u^{T}Ru\end{aligned} \tag{8}$$

给定的允许的 u 应是下列误差的平方式最小化:

$$E(\hat{W}_c)=\frac{1}{2}e^{T}e \tag{9}$$

则评价结构的神经网络权重参数可以用梯度法表示为:

$$\dot{\hat{W}}_c=-a_1\alpha_1\left(\alpha_1^{T}\hat{W}_c+e^{T}Qe+u^{T}\,\mathrm{Ru}\right) \tag{10}$$

这里 a_1 为大于 0 的实数, $\alpha_1=\alpha/\left(\alpha^{T}\alpha+1\right)$, $\alpha=\nabla\phi^{T}\left[f(x)+g(x)u\right]$, 根据 α_1 的表达式可知存在一个正实数 $\alpha_{1m}>1/2$, 使得 $\|\alpha_1\|\le\alpha_{1m}$. 假设评价结构的估计误差为

$$\tilde{W}_c=\hat{W}_c-W_c \tag{11}$$

则存在一个最优控制策略 u 使得哈密尔顿函数重新表述为:

$$\begin{aligned}0\approx H(x,u,V_x)=&\nabla\phi^{T}(x)W\left[f(x)+g(x)u\right]\\&+x^{T}Qx+u^{T}Ru+\varepsilon_{HJB}\end{aligned} \tag{12}$$

上式中 $\varepsilon_{HJB}=\nabla\varepsilon_v\left[f(x)+g(x)u\right]$ 是有界的 HJB 误差,

根据式(10)和式(11)可得:

$$\dot{\tilde{W}}_c = -a_1\alpha_1\left(\alpha_1^T\tilde{W}_c + \varepsilon_{HJB}\right) \tag{13}$$

2.2 基于梯度法的执行结构

为了求得最优控制策略, 用另一个神经网络近似控制策略:

$$u = W_a\varphi(\mathrm{x}) + \varepsilon \tag{14}$$

其中 W_a 为控制策略神经网络权重, $\varphi(\mathrm{x}) \in \mathbf{R}^n$ 为激励向量, ε 为神经网络逼近误差, n 为神经元次数. 则用近似的神经网络权重 $\hat{W}_a$ 可将(14)表示为:

$$\hat{u} = \hat{W}_a\varphi(\mathrm{x}) \tag{15}$$

根据则控制策略误差

$$e_u = \hat{W}_a\varphi(\mathrm{x}) + \frac{1}{2}R^{-1}g(x)^T\nabla\phi^T(x)\hat{W}_c \tag{16}$$

执行机构神经网络近似权重最小化的格式可以表示为:

$$E(\hat{\mathrm{W}}_a) = \frac{1}{2}e_u^T e_u \tag{17}$$

则执行机构的神经网络权重参数可以用梯度法进行更新近似:

$$\dot{\hat{W}}_a = -a_2\phi\left(\hat{W}_a\varphi(\mathrm{x}) + \frac{1}{2}R^{-1}g(x)^T\nabla\phi^T(x)\hat{W}_c\right) \tag{18}$$

这里 a_2 为大于 0 的实数.

假设执行机构神经网络近似误差

$$\tilde{W}_a = \hat{W}_a - W_a \tag{19}$$

根据(6)(7)和(14)式可得:

$$W_a\varphi(\mathrm{x}) + \varepsilon + \frac{1}{2}R^{-1}g(x)^T\nabla\phi^T(x)\hat{W}_c + \frac{1}{2}R^{-1}g(x)^T\nabla\varepsilon_v = 0 \tag{20}$$

则有:

$$\dot{\tilde{W}}_a = -a_2\phi\left(\tilde{W}_a\varphi(\mathrm{x}) + \frac{1}{2}R^{-1}g(x)^T\nabla\phi^T(x)\tilde{W}_c + \beta\right) \tag{21}$$

这里 $\beta = \left[\varepsilon + \frac{1}{2}R^{-1}g(x)^T\nabla\varepsilon_v\right]$.

3 稳定性分析

为了应用李雅普诺夫证明系统稳定性, 构建下式:

$$L = L_1 + L_2 \tag{22}$$

$$\begin{aligned}\dot{L}_1 &= \frac{1}{a_1}tr\left\{\tilde{W}_c^T\dot{\tilde{W}}_c\right\}\\ &= \frac{1}{a_1}tr\left\{\tilde{W}_c^T\left[-a_1\alpha_1\left(\alpha_1^T\tilde{W}_c + \varepsilon_{HJB}\right)\right]\right\}\\ &\le -\left(\alpha_{1m}^2 - \frac{a_1}{2}\alpha_{1M}^2\right)\left\|\tilde{W}_c\right\|^2 + \frac{1}{2a_1}\varepsilon_{HJB}^2\end{aligned}$$

$$\begin{aligned}\dot{L}_2 &= \frac{a_1}{a_2}tr\left\{\tilde{W}_a^T\dot{\tilde{W}}_a\right\}\\ &= \frac{a_1}{a_2}tr\left\{\tilde{W}_a^T\left[-a_2\varphi\left(\tilde{W}_a\varphi(\mathrm{x}) + \frac{1}{2}R^{-1}g(x)^T\nabla\phi^T(x)\tilde{W}_c + \beta_c\right)\right]\right\}\\ &\le -\left(a_1\varphi_M^2 - \frac{3a_1a_2}{4}\varphi_m^2\right)\left\|\tilde{W}_a\right\|^2 + \\ &\quad \frac{a_1}{4a_2}\left\|R^{-1}\right\|^2\left\|g(\mathrm{x})\right\|^2\left\|\nabla\phi_d\right\|^2\left\|\tilde{W}_c\right\|^2 + \frac{a_1}{2a_2}\beta^2\end{aligned}$$

则当 a_1, a_2 满足下列条件时, W_c 和 W_a 收敛到真值, 系统总体也达到稳定状态.

$$0 < a_2 < \frac{4\varphi_M^2}{3\varphi_m^2},\quad a_1 < \left(\frac{4a_2\alpha_{1m}^2}{2\alpha_{1M}^2 + \left\|R^{-1}\right\|^2\left\|g(\mathrm{x})\right\|^2\left\|\nabla\phi_M\right\|^2}, \frac{2\alpha_{1m}^2}{\alpha_{1M}^2}\right)_{\min}.$$

稳定性证明完毕.

4 仿真结果及分析

仿真部分用一个非线性未知的数学模型来验证上述方法的可行性. 引入如下非线性连续系统[15]:

$$\dot{x} = \begin{bmatrix} -x_1 + x_2 \\ -0.5x_1 - 0.5x_2(1 - (\cos(2x_1) + 2)^2) \end{bmatrix} + \begin{bmatrix} 0 \\ \cos(2x_1) + 2 \end{bmatrix}u \tag{23}$$

按照性能指标的定义式(2), 其中 Q 和 R 定义为正定对称矩阵. 所求最优控制的目的是将初始状态整定到零状态, 且超调量不要超出有限范围. 根据(4)式可知, 非线性系统(23)的最优控制标准为:

$$u^* = -\frac{1}{2}R^{-1}[g(x)]^T\frac{\partial V^*(x)}{\partial x} = -\left(\cos(2x_1) + 2\right)x_2 \tag{24}$$

定义系统状态初始值为 $x(0) = [3, -1]^T$, 设持续激励为 $\phi(x) = [x_1^2, x_1x_2, x_2^2]^T$, 设置评价结构各增益参数: $R = 50$, $Q = 2$, $a_1 = 0.8$ 权重始状态设置为 $\hat{W}_c(0)$=$[0\ 0\ 0]^T$. 基于自适应律(10), 评价结构神经网络权重收敛情况如图 1 所示, 图中显示在 5s 之内, 评价结构神经网络权重参数达到稳定值.

然后设置仿真中执行结构参数 $a_1 = 0.8$ 图 2 为执行结构神经网络权重 Wa 收敛曲线, 可以看出快速收敛到其真值.

整定状态收敛曲线如图 3 所示. 10s 内所求得的最优控制可以将系统状态从初始状态[3,-1]'整定到[0, 0]'状态. 图 4 为所求得的最优控制曲线图.

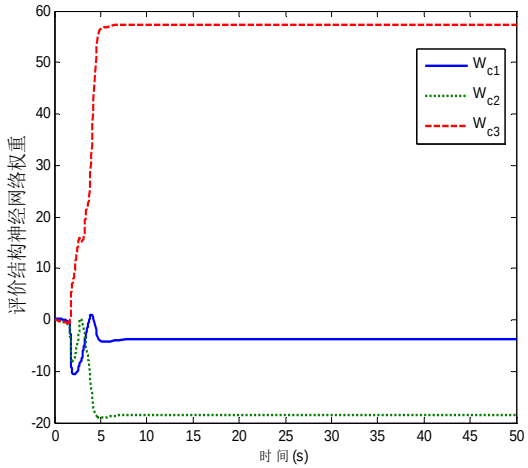

图 1 评价结构神经网络权重 Wc

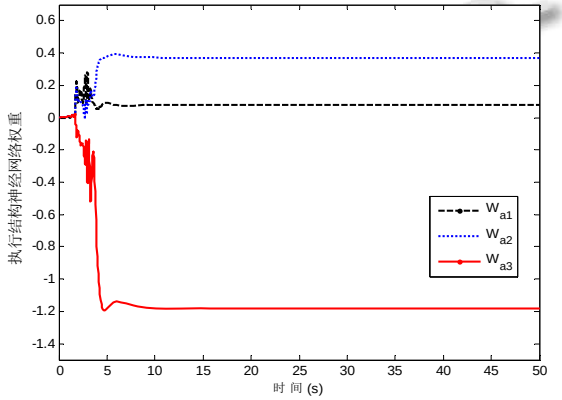

图 2 执行结构的神经网络权重 Wa

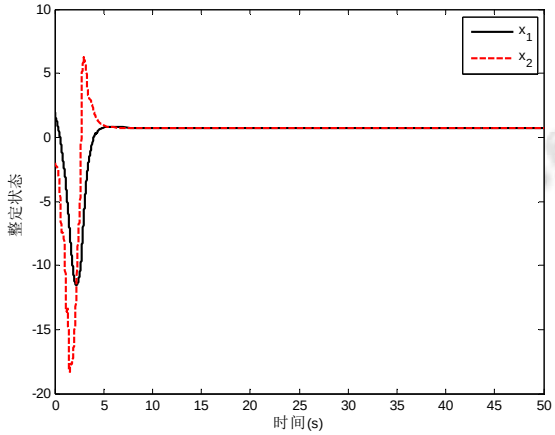

图 3 状态镇定

5 结语

将梯度自适应算法应用于自适应动态规划方法, 在线对非线性连续系统进行自适应动态最优求解. 评价执行结构被用来在线对连续非线性系统进行最优求解. 首先, 用神经网络逼近性能指标, 其次基于评价结构, 应用神经网络逼近执行结构. 其中评价结构和执行结构的神经网络权重参数基于系统在线数据误差, 同时在线被估计求得. 仿真结果更加有力的证明所提出方法的有效性.

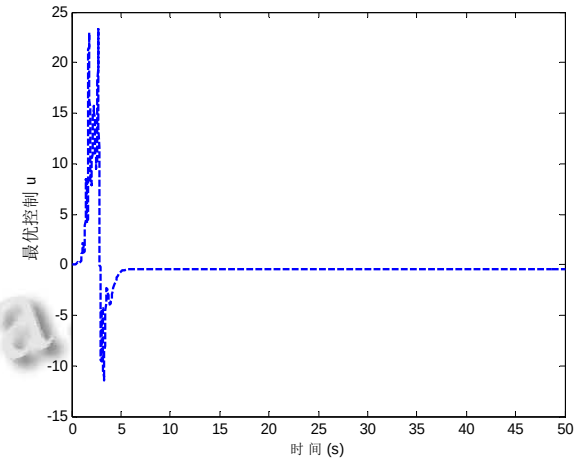

图 4 最优控制

参考文献

1 Bellman R. Dynamic programming and Lagrange multipliers. Proceedings of the National Academy of Sciences of the United States of America, 1956, 42(10): 767.

2 Barto AG. Reinforcement learning: An introduction. Cambridge Univ Press, 1998.

3 Murray JJ, Cox CJ, Lendaris GG, Saeks R. Adaptive dynamic programming. IEEE Trans. on Applications and Reviews, 2002, 32(2): 140–153.

4 Werbos PJ. A menu of designs for reinforcement learning over time. Neural Networks for Control, 1990: 67–95.

5 Abu-Khalaf M, Lewis FL. Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach. Automatica, 2005, 41(5): 779–791.

6 林小峰,黄元君.基于神经网络近似的自适应优化控制.计算机技术与发展,2011,21(11):100–104.

7 Zhang H, Cui L, Zhang X, Luo Y. Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method. IEEE Trans. on Neural Networks, 2011, 22(12): 2226–2236.

8 Modares H, Lewis FL. Linear quadratic tracking control of partially-unknown continuous-time systems using

reinforcement learning. IEEE Trans. on Automatic Control, 2014, 59(11): 3051–3056.

9 Modares H, Lewis FL. Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning. Automatica, 2014, 50(7): 1780–1792.

10 Ni Z, He H. Adaptive learning in tracking control based on the dual critic network design. IEEE Trans. on Neural Networks and Learning Systems, 2013, 24(6): 913–928.

11 Vamvoudakis KG, Lewis FL. Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem. Automatica, 2010, 46(5): 878–888.

12 Vrabie D, Lewis F. Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems. Neural Networks, 2009, 22(3): 237–246.

13 Dierks T, Thumati BT, Jagannathan S. Optimal control of unknown affine nonlinear discrete-time systems using offline-trained neural networks with proof of convergence. Neural Networks, 2009, 22(5): 851–860.

14 Liu D, Wei Q. Finite-approximation-error-based optimal control approach for discrete-time nonlinear systems. IEEE Trans. on Cybernetics, 2013, 43(2): 779–789.

15 Nevistić V, Primbs JA. Constrained nonlinear optimal control: A converse HJB approach. California Institute of Technology, 1997.