

文章编号: 1001-0920(2003)01-0123-03

基于神经网络的微分对策控制器设计

周 锐

(北京航空航天大学 自动控制系, 北京 100083)

摘 要: 采用伴随-BP 技术, 将微分对策的两点边值求解问题转化为两个神经网络的学习问题, 训练后的两个神经网络分别作为对策双方的最优控制器在线使用, 避免了直接求解复杂的两点边值问题, 对追踪微分对策问题的仿真结果表明, 该方法对初始条件和噪声具有较好的鲁棒性。

关键词: 微分对策; 神经网络; 最优控制; 两点边值

中图分类号: TP18

文献标识码: A

Differential game controllers design using neural networks

ZHOU Rui

(Department of Automatic Control, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: Based on the adjoint-backpropagation technique, two point boundary value problem (TPBVP) of differential games is solved using two neural networks. The neural networks automatically adjust their weights to minimize and maximize the cost function of differential game system respectively. The converged neural networks can be used as the feedback optimal differential games controllers on-line, avoiding solving the complex two point boundary value problem directly. The simulation results for pursuing-escaping differential games prove that the neural networks controllers present good robustness with respect to initial conditions and measuring noise.

Key words: Differential games; Neural networks; Optimal control; Two point boundary value

1 引 言

对微分对策问题的求解会导致复杂的两点边值问题(TPBV), 一般很难求解, 并且对初始条件和噪声很敏感^[1]。神经网络具有高度的并行计算机制, 以及良好的学习、推广、记忆和容错等能力, 因而得到广泛的应用^[2]。

一种最直接的优化方法是导出代价函数相对神经网络权值的梯度函数^[3], 但更多的是采用BPTT (backpropagation-through-time) 技术来求解优化问题^[4]。本文采用伴随-BP 技术, 将微分对策的求解问题转化为两个神经网络的学习问题。

2 用神经网络求解微分对策原理

考虑离散微分对策问题^[1]

$$x(k+1) = f(x(k), u(k), v(k))$$

其中: $x(k)$ 和 $u(k), v(k)$ 分别为 k 时刻对策系统状态和对策双方的控制量。对策双方的性能指标为

$$J = \Phi_x(N) + \sum_{k=0}^{N-1} L[x(k), u(k), v(k)] \quad (1)$$

对策双方的目标是选择各自的控制策略, 以极小和极大化性能指标, 即 $\min_u \max_v J$ 。对策双方控制策略分别用两个 BP 神经网络来实现, 即

$$\begin{cases} u(k) = \Phi_u(x(k), w_u(k)) \\ v(k) = \Phi_v(x(k), w_v(k)) \end{cases} \quad (2)$$

其含义为系统状态 $x(k)$ 分别在权值为 w_u 和 w_v 的神

收稿日期: 2001-09-30; 修回日期: 2001-10-29。

基金项目: 国家自然科学基金资助项目(69904002); 国防预研基金资助项目(00J1.1.3HK0102)。

作者简介: 周锐(1968—), 男, 湖北钟祥人, 副教授, 从事飞行控制、制导与智能控制等研究。

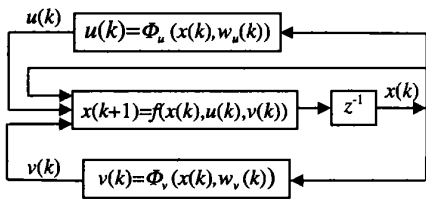


图1 神经网络控制器结构

神经网络 Φ_u 和 Φ_v 中的传播结果, 如图1所示.

定义系统的伴随代价函数

$$J_A = \Phi_x(N) + \sum_{k=0}^{N-1} \{L[x(k), u(k), v(k)] + \lambda^T(k+1)[f(x(k), u(k), v(k)) - x(k+1)] + \lambda_u^T(k)[\Phi_u(x(k), w_u) - u(k)] + \lambda_v^T(k)[\Phi_v(x(k), w_v) - v(k)]\} \quad (3)$$

其中 $\lambda_u, \lambda_v, \lambda$ 为伴随向量. 定义 Hamiltonian 函数

$$H(k) = L(x(k), u(k), v(k)) + \lambda^T(k+1)f(x(k), u(k), v(k)) + \lambda_u^T(k)\Phi_u(x(k), w_u) + \lambda_v^T(k)\Phi_v(x(k), w_v) \quad (4)$$

则 J_A 相对于各变量的微分为

$$dJ_A = \left[\frac{\partial \Phi}{\partial \alpha(N)} - \lambda^T(N) \right] dx(N) + \sum_{k=1}^{N-1} \left[\frac{\partial H(k)}{\partial \alpha(k)} - \lambda^T(k) \right] dx(k) + \sum_{k=0}^{N-1} \left[\frac{\partial H(k)}{\partial u(k)} - \lambda_u^T(k) \right] du(k) + \sum_{k=0}^{N-1} \left[\frac{\partial H(k)}{\partial v(k)} - \lambda_v^T(k) \right] dv(k) + \sum_{k=0}^{N-1} \left[\frac{\partial H(k)}{\partial w_u} \right] dw_u + \sum_{k=0}^{N-1} \left[\frac{\partial H(k)}{\partial w_v} \right] dw_v \quad (5)$$

由最优性条件^[1,4], $dJ_A = 0$, 则有

$$\lambda^T(N) = \frac{\partial \Phi}{\partial \alpha(N)} \quad (6)$$

$$\lambda^T(k) = \frac{\partial H(k)}{\partial \alpha(k)} =$$

$$\frac{\partial H(k)}{\partial \alpha(k)} + \lambda_u^T(k+1) \frac{\partial \Phi_u(k)}{\partial \alpha(k)} + \lambda_v^T(k) \frac{\partial \Phi_v(k)}{\partial \alpha(k)} \quad (7a)$$

$$\lambda^T(k) = \frac{\partial H(k)}{\partial u(k)} =$$

$$\frac{\partial H(k)}{\partial u(k)} + \lambda_u^T(k+1) \frac{\partial \Phi_u(k)}{\partial u(k)} \quad (7b)$$

$$\lambda^T(k) = \frac{\partial H(k)}{\partial v(k)} =$$

$$\frac{\partial H(k)}{\partial v(k)} + \lambda_v^T(k+1) \frac{\partial \Phi_v(k)}{\partial v(k)} \quad (7c)$$

其中 $\lambda_u(k)$ 和 $\lambda_v(k)$ 可看成是 k 时刻控制器的输出误差. 同时可得到神经网络控制器的最优性条件

$$\begin{cases} \frac{\partial H(k)}{\partial u} = \lambda_u^T(k) \frac{\partial \Phi_u(k)}{\partial u} = 0 \\ \frac{\partial H(k)}{\partial v} = \lambda_v^T(k) \frac{\partial \Phi_v(k)}{\partial v} = 0 \end{cases} \quad (8)$$

进而得到两个网络权值的更新规则

$$\begin{cases} \Delta w_u^T = -\alpha_u \lambda_u^T(k) \frac{\partial \Phi_u(k)}{\partial w_u} \\ \Delta w_v^T = \alpha_v \lambda_v^T(k) \frac{\partial \Phi_v(k)}{\partial w_v} \end{cases} \quad (9)$$

其中 α_u 和 α_v 分别为两个网络的学习速率, 负号表示实现极小化, 正号表示实现极大化.

3 最优性解释

将性能指标等效为

$$J_B = \Phi_x(N) + \sum_{k=0}^{N-1} L[x(k), u(k), v(k)] + \lambda^T(k+1)[f(x(k), u(k), v(k)) - x(k+1)] \quad (10)$$

由 λ_u 和 λ_v 的定义可得

$$\lambda_u^T(k) = \frac{\partial J_B}{\partial u(k)}, \quad \lambda_v^T(k) = \frac{\partial J_B}{\partial v(k)} \quad (11)$$

由式(2)和(8)~(11), 并由链导数求导规则得

$$\Delta w_u = -\alpha_u \left(\frac{\partial J_B}{\partial w_u} \right)^T, \quad \Delta w_v = \alpha_v \left(\frac{\partial J_B}{\partial w_v} \right)^T \quad (12)$$

式(12)说明了两个神经网络的学习过程分别为极小和极大化对策代价函数的过程.

4 追逃微分对策问题

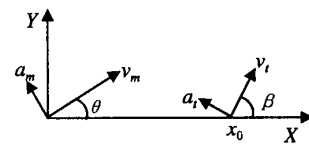


图2 导弹和飞机拦截问题几何示意

如图2所示, 考虑平面拦截问题. 图中 a_m 和 a_v 分别为导弹和飞机的法向加速度, 相对位移用 $x = x_t - x_m$ 和 $y = y_t - y_m$ 表示, 则导弹与飞机之间离散化的运动学方程可表示为

$$\begin{cases} x(k+1) = x(k) + t_s v_x \cos \beta(k) - \\ \quad t_s v_m \cos \theta(k) \\ y(k+1) = y(k) + t_s v_x \sin \beta(k) - \\ \quad t_s v_m \sin \theta(k) \\ \theta(k+1) = \theta(k) + t_s a_m(k) / v_m \\ \beta(k+1) = \beta(k) + t_s a_l(k) / v_l \end{cases} \quad (13)$$

其中 t_s 为采样周期, 对策双方的目标为 $\min_{a_m} \max_{a_l} J$, 而

$$J = \frac{1}{2} [x^2(N) + y^2(N)] + \frac{1}{2} \sum_{k=0}^{N-1} [k_1 a_m^2(k) - k_2 a_l^2(k)] \quad (14)$$

两个网络输入均为 $[x(k) \ y(k) \ \theta(k) \ \beta(k)]$, 输出分别为 $a_m(k)$ 和 $a_l(k)$, $k = 0, 1, \dots, N - 1$, 结构均采用 4-10-1 的 BP 神经网络。

为提高网络的稳定性和收敛性, 对输入输出数据作了无量纲和规范化处理, 设

$$\begin{aligned} v_m &= 500 \text{ m/s}, \quad v_l = 300 \text{ m/s} \\ x_0 &= 500 \text{ m}, \quad y_0 = 0, \quad t_s = 0.02 \text{ s} \end{aligned}$$

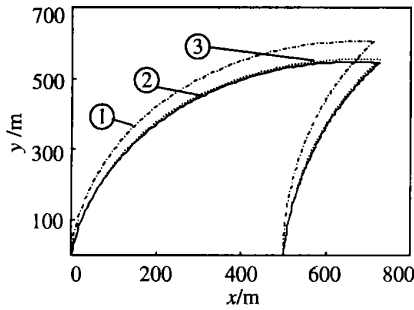


图 3 导弹和目标运动轨迹

$$\gamma = 10^{-4}, \quad k_1 = k_2 = 1$$

以 $\theta_0 = 80^\circ; \beta_0 = 85^\circ$ 为中心, 对不同的 θ_0 和 β_0 进行学习。考察了 θ 和 β 具有测量噪声的情况, 噪声强度为 $n(k) = 0.01 \sin 5k t_s$ 。

仿真结果如图 3 所示。其中: 曲线 1: $\theta_0 = 85^\circ; \beta_0 = 90^\circ$; 曲线 2: $\theta_0 = 80^\circ; \beta_0 = 85^\circ$ 可以看出, 学习后的神经网络控制器对初始条件和噪声具有较强的鲁棒性。

5 结 论

本文避免了采用双边极大值原理处理微分对策问题的常规方法, 而是基于变分法得到了神经网络权值调节规律与微分对策控制器最优性之间的关系, 避免了直接求解复杂的两点边值问题。所得的神经网络双边反馈控制器具有较好的性能和鲁棒性, 克服了两点边值求解方法对初始条件和噪声敏感的缺点。由于神经网络的学习和收敛速度很慢, 一般情况下只能离线训练, 在线应用。

参考文献(References):

[1] Bryson A E, Ho Y. *Applied Optimal Control*[M]. New York: Hemisphere Publishing Co, 1975.
 [2] Lew is FL, Parisini T. New developments in neurocontrol[A]. *Proc IEEE Int Conf Control Appl*[C]. Trieste, 1998 86-91.
 [3] Am in SM, Rodin E Y. Neurocontrol of an aircraft[J]. *Math Comput Model*, 1995, 22(1): 63-78
 [4] Plumer E S. Optimal control of terminal processes using neural networks [J]. *IEEE Trans Neural Networks*, 1996, 7(2): 408-418

本刊荣获“百种中国杰出学术期刊”称号

本刊讯 2002 年 12 月 10 日, 中国科学技术信息研究所在京召开新闻发布会, 公布了 2001 年度中国科技论文统计结果, 表奖了部分科技期刊。按照该所出版的《2002 年版中国科技期刊引证报告》, 《控制与决策》的总被引频次、他引率、影响因子、即年指标、基金论文比等项指标, 在信息科学与系统科学类期刊中都是较高的, 并被评为“百种中国杰出学术期

刊”。

这一成绩的取得, 是与各方面的大力支持密不可分的。在此, 本刊编委会和编辑部谨向所有关心和支持本刊工作的各方面专家学者和广大读者表示衷心的感谢! 我们将再接再厉, 开拓进取, 与时俱进, 再创佳绩, 为进一步办好《控制与决策》而做出自己的努力。