

基于平行控制的离散非线性系统的事件触发近似最优控制

廖泽华^{1,2}, 梁子钰^{1,3}, 周天民¹, 卢经纬^{1,2}, 魏庆来^{1,2}

(1. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室, 北京 100190;

2. 中国科学院大学人工智能学院, 北京 100049;

3. 天津大学智能电网教育部重点实验室, 天津 300072)

摘要: 针对离散非线性系统提出了一种基于平行控制的事件触发近似最优控制方法。首先, 基于时间触发的最优值函数和最优控制律, 提出了一种新型的触发条件, 并基于李雅普诺夫方法证明了闭环系统的渐近稳定性。其次, 为了实施设计的触发条件, 提出了一种采用神经网络和自适应动态规划技术的平行控制方法, 以预测系统的下一步状态, 并获得最优值函数和最优控制律。最后, 通过仿真实验验证了所提方法的有效性。

关键词: 事件触发控制; 非线性最优控制; 平行控制; 自适应动态规划

中图分类号: TP273

文献标识码: A

doi: 10.11959/j.issn.2096-6652.202142

Parallel control-based event-driven approximate optimal control of discrete-time nonlinear systems

LIAO Zehua^{1,2}, LIANG Ziyu^{1,3}, ZHOU Tianmin¹, LU Jingwei^{1,2}, WEI Qinglai^{1,2}

1. The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100086, China

2. School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China

3. Key Laboratory of Smart Grid of Ministry of Education, Tianjin University, Tianjin 300072, China

Abstract: A parallel control-based event-driven approximate optimal control method was proposed for discrete-time nonlinear systems. Firstly, based on the time-triggered optimal value function and optimal control law, a novel triggering condition was developed and the asymptotic stability of the closed-loop system was proved based on the Lyapunov method. Secondly, a parallel control method using neural networks and adaptive dynamic programming techniques was proposed to predict the next state of the system and to obtain the optimal value function and control law. Finally, the effectiveness of the developed method was validated by numerical examples.

Key words: event-driven control, nonlinear optimal control, parallel control, adaptive dynamic programming

1 引言

在过去的数十年中, 随着计算能力的提升, 智能控制方法得到了长足的进步^[1-6]。基于人工系统+计算实验+平行执行方法建立的平行控制(parallel control)是一种智能控制方法^[7-9]。平行控制方法的基本思想是将实际问题扩展到虚拟空间, 通过虚拟现实互动的方式来解决相关实际问题^[7]。Wei Q L

等人^[8]针对连续时间线性系统设计了一种新的平行控制器, 建立了一种新的平行控制律。Lu J W 等人^[9]研究了连续时间一般非线性系统的最优平行跟踪控制问题, 与现有的最优状态反馈控制不同, 其在反馈系统中引入了最优平行控制的控制输入。除了平行控制, 自适应动态规划(adaptive dynamic programming, ADP)也是一类解决非线性最优控制问题的智能控制技术, 被用于求解复杂的哈密顿-

收稿日期: 2021-09-30; 修回日期: 2021-11-01

通信作者: 梁子钰, liangziyu@tju.edu.cn

基金项目: 国家自然科学基金资助项目(No.62073321)

Foundation Item: The National Natural Science Foundation of China (No.62073321)

雅可比-贝尔曼 (Hamilton-Jacobi-Bellman, HJB) 方程^[10]。目前, ADP 技术在社会与工程系统中得到了广泛应用^[11-12]。在 ADP 技术中, 迭代 ADP 技术对于求解基于系统数据生成的 HJB 方程十分有效^[13-14], 其中值迭代和策略迭代是实现迭代 ADP 技术的主要方式。Al-Tamimi A 等人^[15]首次证明了非线性离散时间系统基于值迭代的 ADP 方法的收敛性。Liu D R 等人^[16]证明了非线性离散时间系统基于策略迭代的 ADP 方法的收敛性。Wei Q L 等人^[17]提出了一种基于值迭代的不需要零初值函数的非线性离散时间系统的 ADP 方法, 并证明了该方法的收敛性。

事件触发机制有效降低了控制系统所需的通信和计算成本, 使其在大型控制系统中具有广阔的应用前景^[18-19]。近年来, 基于 ADP 技术的事件触发最优控制已成为一个重要领域, 并在理论研究和实际应用中取得了非常可观的成果^[20-22]。Luo B 等人^[23]利用 ADP 技术提出了一种用于连续时间非线性系统的事件触发最优控制方法, 并给出了实际性能的上界。Yang X 等人^[24]提出了一种利用折扣代价函数求解具有非零平衡点的连续时间非线性系统的事件触发最优控制方法, 并将其应用于连续搅拌釜式反应器系统。Mu C X 等人^[25]针对具有动态事件通信机制的受扰系统, 提出了一种基于积分强化学习的鲁棒学习控制方法, 并将其应用于单连杆关节臂系统和单区域电力系统。

值得指出的是, 上述事件触发优化控制方法需要求解事件触发的 HJB 方程, 且通常采用在线 ADP 的方式来更新神经网络参数。为了解决上述问题并将迭代 ADP 技术进一步引入事件触发优化调节控制问题中, 本文基于平行系统的框架^[4,7], 针对离散非线性系统提出了一种新型事件触发近似最优控制方法。

本文的主要贡献如下:

(1) 通过求解原始的 HJB 方程, 而不是通过求解事件触发的 HJB 方程, 为离散非线性系统设计了一个新的触发条件, 降低了实施事件触发最优控制的难度;

(2) 与大多数现有的基于 ADP 的事件触发最优控制方法相比^[20-22], 本文引入了迭代 ADP 技术来离线求解 HJB 方程, 以促进所提方法在大型系统中的应用;

(3) 此外, 由于使用了平行系统, 不必确切地知道原始系统的表达式, 所提方法同样适用于不能

使用物理定律精确建模的系统;

(4) 根据提出的事件触发近似最优控制方法, 可以预先给出实际性能指标的上界。

2 问题描述

考虑如下的仿射输入非线性离散系统:

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k)) + \mathbf{g}(\mathbf{x}(k))\mathbf{u}(k) \quad (1)$$

其中, 状态 $\mathbf{x}(k) \in \mathbb{R}^n$ 、 $\mathbf{f}(\mathbf{x}(k)) \in \mathbb{R}^n$ 、 $\mathbf{g}(\mathbf{x}(k)) \in \mathbb{R}^{n \times m}$, 控制 $\mathbf{u}(k) \in \mathbb{R}^m$ 。

最优控制目标是设计一种控制律, 使状态趋于零, 同时最小化以下无限时域性能指标函数:

$$J(\mathbf{x}(k), \mathbf{u}(k)) = \sum_{l=k}^{\infty} U(\mathbf{x}(l), \mathbf{u}(l)) \quad (2)$$

其中, $U(\mathbf{x}(k), \mathbf{u}(k))$ 是效用函数, 其定义为:

$$U(\mathbf{x}(k), \mathbf{u}(k)) = \mathbf{x}^T(k)\mathbf{Q}\mathbf{x}(k) + \mathbf{u}^T(k)\mathbf{R}\mathbf{u}(k) \quad (3)$$

其中, $\mathbf{Q}$ 和 $\mathbf{R}$ 都是正定矩阵。

对于任意容许控制 $\mathbf{u}(\mathbf{x}(k))$, 其时间触发的值函数定义如下:

$$V(\mathbf{x}(k)) = \sum_{l=k}^{\infty} U(\mathbf{x}(l), \mathbf{u}(\mathbf{x}(l))) \quad (4)$$

根据贝尔曼最优性原理^[9], 最优值函数 $V^*(\mathbf{x}(k))$ 满足如下贝尔曼 (离散时间 HJB) 方程:

$$V^*(\mathbf{x}(k)) = \min_{\mathbf{u}(\mathbf{x}(k))} \{ \mathbf{x}^T(k)\mathbf{Q}\mathbf{x}(k) + \mathbf{u}^T(\mathbf{x}(k))\mathbf{R}\mathbf{u}(\mathbf{x}(k)) + V^*(\mathbf{x}(k+1)) \} \quad (5)$$

则 k 时刻的最优控制律 $\mathbf{u}^*(\mathbf{x}(k))$ 为:

$$\mathbf{u}^*(\mathbf{x}(k)) = \arg \min_{\mathbf{u}(\mathbf{x}(k))} \{ \mathbf{x}^T(k)\mathbf{Q}\mathbf{x}(k) + \mathbf{u}^T(\mathbf{x}(k))\mathbf{R}\mathbf{u}(\mathbf{x}(k)) + V^*(\mathbf{x}(k+1)) \} \quad (6)$$

对于时间触发最优控制问题, 最优性能指标函数为^[16-17]:

$$J^*(\mathbf{x}(k)) = V^*(\mathbf{x}(k)) \quad (7)$$

一般事件触发优化控制是基于反馈控制进行设计的, 即需要获得实时状态信息, 然而控制器何时需要更新控制信号并传输给被控系统是由触发条件决定的。由于减少了控制器传输控制信号的次数, 通信损耗得到降低。在事件触发控制中, 控制输入更新的时刻被称为触发时刻。触发时刻是单调递增序列 k_i , 其中 $k_0 = 0$, $k_i < k_{i+1}$, $i \in \mathbb{N}$ 。在 $[k_i, k_{i+1})$,

控制在零阶保持器 (zero-order-holder, ZOH) 下保持恒定。执行间隔时间定义如下:

$$h_i \triangleq k_{i+1} - k_i \quad (8)$$

在下文中, 笔者将给出本文提出的事件触发近似最优控制方法的详细设计步骤和理论分析。在分析之前, 以下假设是必要的。

假设 1 式 (1) 所示的系统在一个包括原点的紧集 Ω 上是可控的; $f + gu$ 在 Ω 上是利普希茨连续的, 并且 $f(0) = 0$ 。

3 事件触发近似最优控制方法设计

3.1 触发条件设计

根据式 (6) 进行计算, 并采用零阶保持器, 则控制变为:

$$u(k) = \mu(\bar{x}(k)) \quad (9)$$

其中, 当 $k \in [k_i, k_{i+1})$ 时, $\mu(\bar{x}(k)) = u^*(x(k_i))$, $\bar{x}(k)$ 表示控制器可用的状态信息。例如, 当 $k \in [k_i, k_{i+1})$ 时, 有 $\bar{x}(k) = x(k_i)$ 。

采用式 (9), 则式 (1) 所示的系统的闭环形式如下:

$$\begin{aligned} x(k+1) &= f(x(k)) + g(x(k))\mu(\bar{x}(k)) \\ &= f(x(k)) + g(x(k))u^*(x(k_i)) \end{aligned} \quad (10)$$

此时, $k \in [k_i, k_{i+1})$ 。

受参考文献[19]的启发, 本文设计了一种新的触发条件:

$$C_\alpha(x(k+1), x(k), \bar{x}(k)) \leq 0 \quad (11)$$

其中, $C_\alpha(x(k+1), x(k), \bar{x}(k)) = \alpha(V^*(x(k+1)) - V^*(x(k))) + x^\top(k)Qx(k) + \mu^\top(\bar{x}(k))R\mu(\bar{x}(k))$, $\alpha \geq 1$ 。下一个触发时刻 k_{i+1} 如下:

$$k_{i+1} = \inf \{k \mid C_\alpha(x(k+1), x(k), \bar{x}(k)) > 0, k > k_i\} \quad (12)$$

式 (12) 保证 $C_\alpha(x(k+1), x(k), \bar{x}(k)) \leq 0$ 。若 k 时刻 $C_\alpha(x(k+1), x(k), \bar{x}(k)) > 0$, 则 $u(k)$ 更新为离线计算出的时间触发的最优控制律 $u^*(x(k))$ 。值得指出的是, $u^*(x(k))$ 代表时间触发的最优控制律, 故控制输入更新为 $u^*(x(k))$ 后, $C_\alpha(x(k+1), x(k), \bar{x}(k)) \leq 0$ 总是可以被满足。本文提出的新的触发条件 (式 (11)) 的优点在于: ①只需要通过简单四则运算就能计算出触发条件, 算法运行的大部分计算代价在于获得最优值函数和最优控制律, 这些都可以通过离线训练来实现, 便于应用到实际的控制

系统中; ②应用事件触发条件能使闭环系统 (式 (10)) 达到渐近稳定, 这在理论上证明了该触发条件的可行性。

3.2 稳定性分析

本节笔者将给出闭环系统在触发条件下的稳定性分析。

定理 1 考虑闭环系统和触发条件, 并且让假设 1 成立。令 x_0 表示初始状态, 并令 $J(x_0, \mu)$ 表示系统实际的性能指标。那么如下结论成立: ①闭环系统是渐近稳定的; ②系统实际的性能指标 $J(x_0, \mu)$ 有上界, 即 $J(x_0, \mu) \leq \gamma J^*(x_0)$ 。

证明:

① 将 $V^*(x(k))$ 作为李雅普诺夫 (Lyapunov) 函数, 那么对于 $k \in [k_i, k_{i+1})$, 可得:

$$\begin{aligned} V^*(x(k+1)) - V^*(x(k)) &= \\ \frac{1}{\gamma} (C_\alpha(x(k+1), x(k), \bar{x}(k)) - x^\top(k)Qx(k) - \\ \mu^\top(\bar{x}(k))R\mu(\bar{x}(k))) &\leq \\ -\frac{1}{\gamma} (x^\top(k)Qx(k) + \mu^\top(\bar{x}(k))R\mu(\bar{x}(k))) &\leq 0 \end{aligned} \quad (13)$$

其中, 等号当且仅当 $x(k) = 0$ 时成立, 因此闭环系统是渐近稳定的。

② 由式 (13) 可得 $x^\top(k)Qx(k) + \mu^\top(\bar{x}(k))R\mu(\bar{x}(k)) \leq \gamma(V^*(x(k)) - V^*(x(k+1)))$, 并且由①可得 $\lim_{k \rightarrow \infty} V^*(x(k)) = 0$ 。因此, 有:

$$\begin{aligned} J(x_0, \mu) &= \sum_{k=0}^{\infty} (x^\top(k)Qx(k) + \mu^\top(\bar{x}(k))R\mu(\bar{x}(k))) \leq \\ \sum_{k=0}^{\infty} \gamma (V^*(x(k)) - V^*(x(k+1))) &= \\ \gamma V^*(x_0) - \gamma \lim_{k \rightarrow \infty} V^*(x(k)) &= \gamma V^*(x_0) = \gamma J^*(x_0) \end{aligned} \quad (14)$$

证明完毕。

3.3 基于平行控制的实现方案

为了实施所设计的触发条件, 本文提出了一种采用神经网络和 ADP 技术的平行控制方法^[4,7,19], 以预测系统的下一个状态并获得最优值函数和最优控制律, 基于平行控制的事件触发近似最优控制方法的总体框架如图 1 所示。其中, 人工系统: 使用模型网络重构系统 (式 (1)); 计算实验是指使用迭代 ADP 技术求解离散时间 HJB 方程 (式 (5)); 平行执行是指 $\hat{x}(k+1)$ 可以通过在 k 时刻

将控制 $u(k)$ 应用于人工系统来获得。在本文中, 笔者采用神经网络技术对实际系统进行重构来获得人工系统, 然后通过迭代 ADP 技术进行网络训练, 同时利用人工系统预测实际系统下一步的状态, 故不需要确切知道原始系统的表达式。

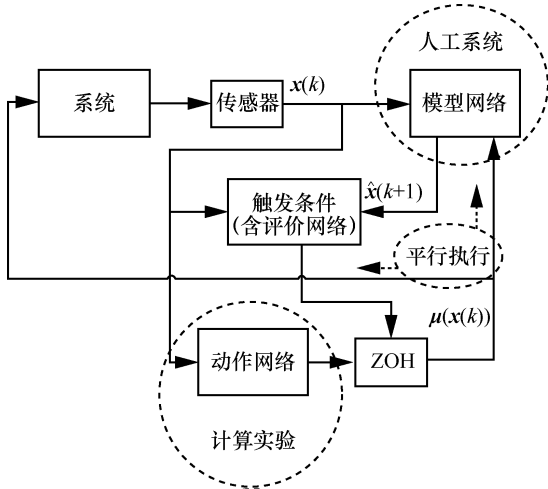


图1 基于平行控制的事件触发近似最优控制方法的总体框架

在神经网络设计部分, 本文给出了基于3个神经网络(均采用3层BP神经网络)的详细平行控制设计方案。

(1) 模型网络

模型网络用于在时刻 k 估计系统下一步的状态 $x(k+1)$ 。系统(式(1))的神经网络如下:

$$x(k+1) = W_m^{*\top} \sigma_m(\beta(k)) + \varepsilon_m(k) \quad (15)$$

其中, $\beta(k) = [x^\top(k), u^\top(k)]^\top$, $W_m^* \in \mathbb{R}^{N_m \times n}$ 是理想的模型网络权值矩阵, $\sigma_m(x(k), u(x(k))) : \mathbb{R}^{n+m} \rightarrow \mathbb{R}^{N_m}$ 是模型网络的激活函数, $\varepsilon_m(k)$ 是模型网络的误差函数。

模型网络的输出为:

$$\hat{x}(k+1) = W_m^\top \sigma_m(\beta(k)) \quad (16)$$

其中, W_m 是 W_m^* 的估计值, 权值更新律为:

$$W_m(k+1) = W_m(k) - l_m \frac{\partial E_m(k)}{\partial W_m(k)} \quad (17)$$

其中, l_m 为模型网络学习率, 该神经网络的目标是最小化以下误差:

$$E_m(k) = \frac{1}{2} \tilde{x}^\top(k+1) \tilde{x}(k+1) \quad (18)$$

其中, $\tilde{x}(k+1) = \hat{x}(k+1) - x(k+1)$ 。值得指出的是, 这里的模型网络是采用梯度下降法预先训练的, 逼近误差可以做到足够小。

(2) 评价网络

利用评价网络的输出逼近值函数, 具体如下:

$$V^*(x(k)) = W_c^{*\top} \sigma_c(x(k)) + \varepsilon_c(k) \quad (19)$$

其中, $W_c^* \in \mathbb{R}^{N_c}$ 是理想的评价网络权值矩阵, $\sigma_c(x(k)) : \mathbb{R}^n \rightarrow \mathbb{R}^{N_c}$ 是评价网络的激活函数, $\varepsilon_c(k)$ 是评价网络的误差函数。

评价网络的输出为:

$$\hat{V}(x(k)) = \hat{W}_c^\top \sigma_c(x(k)) \quad (20)$$

其中, $\hat{W}_c$ 是 W_c^* 的估计值, 权值更新律为:

$$\hat{W}_c(k+1) = \hat{W}_c(k) - l_c \frac{\partial E_c(k)}{\partial \hat{W}_c(k)} \quad (21)$$

其中, l_c 为评价网络学习率, 该神经网络的目标是最小化以下误差:

$$E_c(k) = \frac{1}{2} e_c^2(k) \quad (22)$$

其中, $e_c(k) = V(x(k)) - (V(\hat{x}(k+1)) + U(x(k), u(k)))$ 。

(3) 动作网络

利用动作网络的输出逼近控制, 具体如下:

$$u^*(x(k)) = W_a^{*\top} \sigma_a(x(k)) + \varepsilon_a(k) \quad (23)$$

其中, $W_a^* \in \mathbb{R}^{N_a}$ 是理想的动作网络权值矩阵, $\sigma_a(x(k)) : \mathbb{R}^n \rightarrow \mathbb{R}^{N_a}$ 是动作网络的激活函数, $\varepsilon_a(k)$ 是动作网络的误差函数。

动作网络的输出为:

$$\hat{u}(x(k)) = \hat{W}_a^\top \sigma_a(x(k)) \quad (24)$$

其中, $\hat{W}_a$ 是 W_a^* 的估计值, 权值更新律为:

$$\hat{W}_a(k+1) = \hat{W}_a(k) - l_a \frac{\partial E_a(k)}{\partial \hat{W}_a(k)} \quad (25)$$

其中, l_a 是动作网络学习率, 该神经网络的目标是最小化以下误差:

$$E_a(k) = \frac{1}{2} e_a^2(k) \quad (26)$$

其中, $e_a(k) = V(\hat{x}(k+1)) + U(x(k), u(k))$ 。

本文在计算实验中使用评价网络和动作网络来获得基于时间触发的最优值函数和最优控制律, 这一过程可以通过值迭代或策略迭代等算法来实现^[15-17]。本文以参考文献[17]中提出的值迭代ADP算法为例来获得基于时间触发的最优值函数和最

优控制律, 值迭代 ADP 算法如下。

算法 1 值迭代 ADP 算法伪代码

初始化:

随机初始化状态 $\mathbf{x}_0$

选择计算精度 ε

迭代过程:

1: 令 $i = 0$, $V_0(\mathbf{x}(k)) = 0$

2: 计算初始迭代控制律: $\mathbf{u}_0(\mathbf{x}(k)) = \arg \min_{\mathbf{u}(k)} \{U(\mathbf{x}(k), \mathbf{u}(k)) + V_0(\mathbf{x}(k+1))\}$

3: 并计算迭代值函数: $V_1(\mathbf{x}(k)) = U(\mathbf{x}(k), \mathbf{u}_0(\mathbf{x}(k))) + V_0(f(\mathbf{x}(k)) + g(\mathbf{x}(k))\mathbf{u}_0(k))$

4: 令 $i = i + 1$

5: 计算迭代控制律: $\mathbf{u}_i(\mathbf{x}(k)) = \arg \min_{\mathbf{u}(k)} \{U(\mathbf{x}(k), \mathbf{u}(k)) + V_i(\mathbf{x}(k+1))\}$

6: 计算迭代值函数: $V_{i+1}(\mathbf{x}(k)) = U(\mathbf{x}(k), \mathbf{u}_i(\mathbf{x}(k))) + V_i(f(\mathbf{x}(k)) + g(\mathbf{x}(k))\mathbf{u}_i(k))$

7: if $\forall \mathbf{x}(k)$, $|V_{i+1}(\mathbf{x}(k)) - V_i(\mathbf{x}(k))| \leq \varepsilon$ then 执行步骤 9

8: else 返回步骤 4

9: end if

10: 返回 $\mathbf{u}_i(\mathbf{x}(k))$ 和 $V_i(\mathbf{x}(k))$

离散非线性系统事件触发近似最优控制方法的具体算法如下。

算法 2 离散非线性系统事件触发近似最优控制算法伪代码

初始化:

根据算法 1 训练评价网络和动作网络

选择系统运行时间 T_r

令 $k = 0, i = 0, k_i = 0$

运行离散非线性系统:

1: 令 $\mathbf{u}(k) = \hat{\mathbf{u}}(\mathbf{x}(k))$

2: 根据系统 (式 (1)) 更新系统状态

3: 令 $k = k + 1$

4: for $k = 1; k \leq T_r; k++$, do

5: 计算触发条件 (式 (11))

6: if $C_a(\mathbf{x}(k+1), \mathbf{x}(k), \bar{\mathbf{x}}(k)) > 0$ then

7: 令 $i = i + 1, k_i = k$

8: 令 $\mathbf{u}(k) = \hat{\mathbf{u}}(\mathbf{x}(k_i)) = \hat{\mathbf{u}}(\mathbf{x}(k))$

9: else

10: $\mathbf{u}(k)$ 和 k_i 保持不变

11: end if

12: 根据系统 (式 (1)) 更新系统状态

13: end for

14: 返回控制结果

4 仿真实验

为了证明提出的事件触发近似最优控制方法的有效性和通用性, 笔者选择扭摆系统来验证本文算法的性能, 扭摆的动力学方程如下:

$$\begin{cases} \frac{d\theta}{dt} = \omega \\ J \frac{d\omega}{dt} = u - Mgl \sin \theta - f_d \frac{d\theta}{dt} \end{cases} \quad (27)$$

其中, $M = 1/3 \text{ kg}$ 和 $l = 2/3 \text{ m}$ 分别为摆杆的质量和长度, 当前角度 θ 和角速度 ω 为系统状态变量。设转动惯量和摩擦系数分别为 $J = 4/3 M l^2$ 和 $f_d = 0.2$, 设重力加速度为 $g = 9.8 \text{ m/s}^2$ 。使用采样间隔 $t = 0.1 \text{ s}$ 离散化系统 (式 (27)), 可得非线性离散时间系统方程如下:

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} 0.1x_2(k) + x_1(k) \\ -0.49 \times \sin(x_1(k)) - 0.1 \times f_d \times x_2(k) + x_2(k) \end{bmatrix} + \begin{bmatrix} 0 \\ 0.1 \end{bmatrix} u(k) \quad (28)$$

其中, 离散后的系统状态 $x_1(k) = \theta(k)$ 、 $x_2(k) = \omega(k)$, 初始状态为 $\mathbf{x}_0 = [1, -1]^T$ 。效用函数为 $U(\mathbf{x}(k), u(k)) = \mathbf{x}^T(k) \mathbf{Q} \mathbf{x}(k) + R u^2(k)$, 其中 $\mathbf{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $R = 1$ 。

选择如下激活函数:

$$\sigma_c(\mathbf{x}(k)) = [x_1(k)^2, x_1(k)x_2(k), x_2(k)^2, x_1(k)^3, x_1(k)^2x_2(k), x_1(k)x_2(k)^2, x_2(k)^3, x_1(k)^4, x_1(k)^3x_2(k), x_1(k)^2x_2(k)^2, x_1(k)x_2(k)^3, x_2(k)^4]^T \quad (29)$$

$$\sigma_a(\mathbf{x}(k)) = [x_1(k), x_2(k), x_1(k)^2, x_1(k)x_2(k), x_2(k)^2, x_1(k)^3, x_1(k)^2x_2(k), x_1(k)x_2(k)^2, x_2(k)^3]^T \quad (30)$$

采用值迭代算法, 神经网络权值的收敛轨迹如图 2 所示。结果表明, 神经网络权值在迭代约 70 次之后收敛。

神经网络的最终权值为:

$$\hat{\mathbf{W}}_{c, \text{final}} = [68.5043, 8.4751, 15.0795, 0, 0, 0, 0, -5.5001, -0.4379, -1.0459, -0.1610, -0.1078]^T,$$

$$\hat{W}_{a,\text{final}} = [0.2545, -1.3231, 0.0010, -0.0016, 0.0010, -0.0892, 0.0713, 0.0148, 0.0154]^T$$

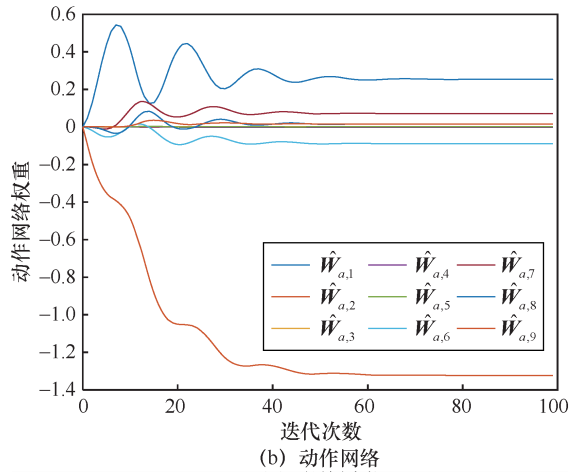
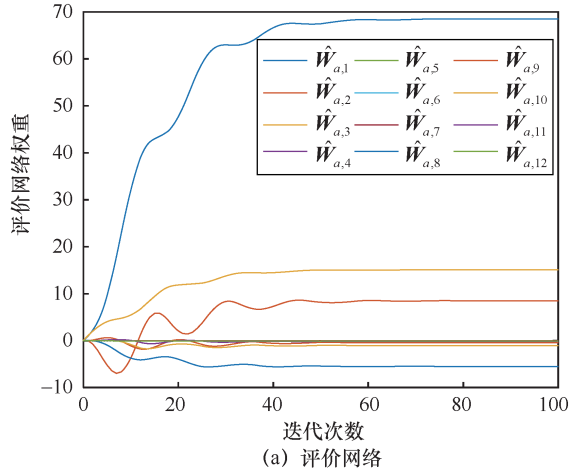


图2 神经网络权值的收敛轨迹

(1) 实验一

实验一在 $\gamma = 1.1$ 的情况下进行仿真，总仿真步长为 $T_r = 300$ 步。同时，在扭摆系统中采用传统的时间触发 ADP 方法进行对比实验。

实验一仿真过程中的系统状态和控制输入轨迹分别如图3和图4所示。可以看出，系统状态最终被调节到了目标状态，这表明本文提出的方法有效地稳定了扭摆系统。此外，实验一中控制输入的更新次数轨迹如图5所示，从图5可以看出，本文提出的方法有效地降低了状态信息的通信成本。与具有 300 次控制输入更新的传统时间触发方法相比，在 $\gamma = 1.1$ 的情况下，使用本文提出的方法，控制输入的更新次数降低为 177 次。与此同时，使用本文提出的方法获得的系统状态轨迹与使用时间触发 ADP 方法获得的系统状态轨迹差别不大。设 $J^*(x_0)$ 表示在时间触发控制方案下

得到的性能指标，则 $J^*(x_0) = 70.5739$ 。而实际性能指标是 $J(x_0, \hat{u}) = 75.8931$ ，低于上界 $\gamma J^*(x_0) = 77.6313$ 。

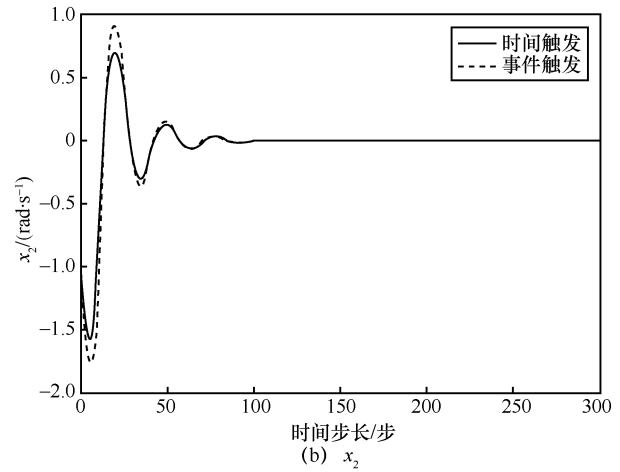
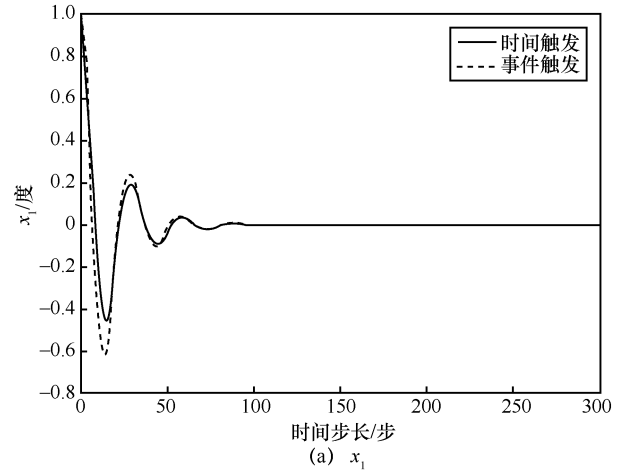


图3 实验一的仿真过程中的系统状态

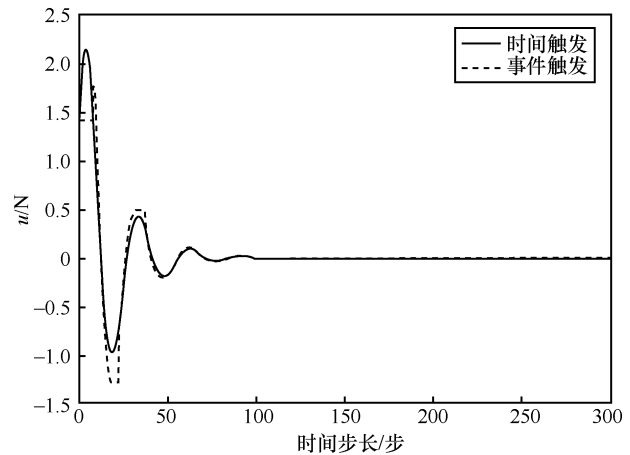


图4 实验一的仿真过程中的控制输入轨迹

(2) 实验二

实验二使用较大的 γ 验证本文提出的方法，将 γ 设置为 6，其他实验参数与实验一相同。

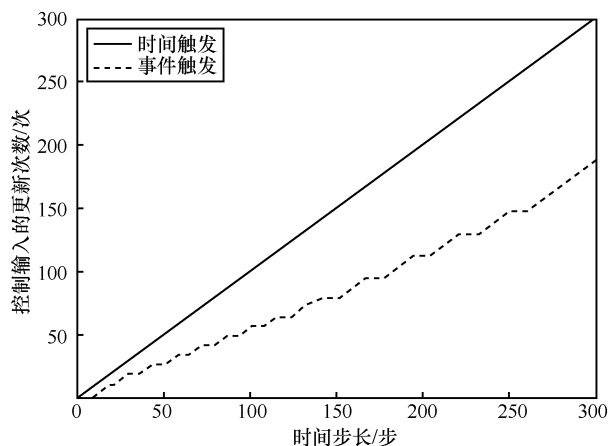


图5 实验一中控制输入的更新次数轨迹

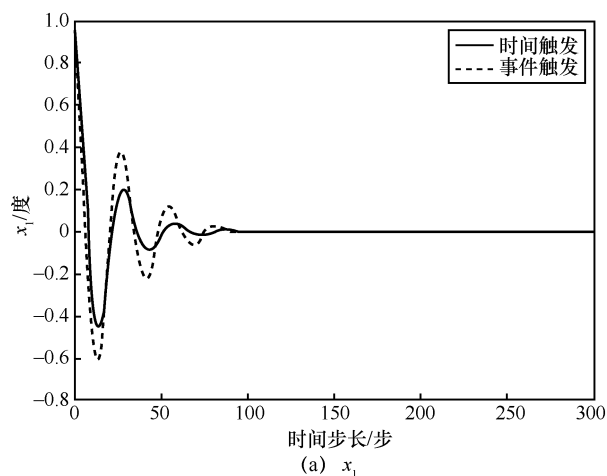
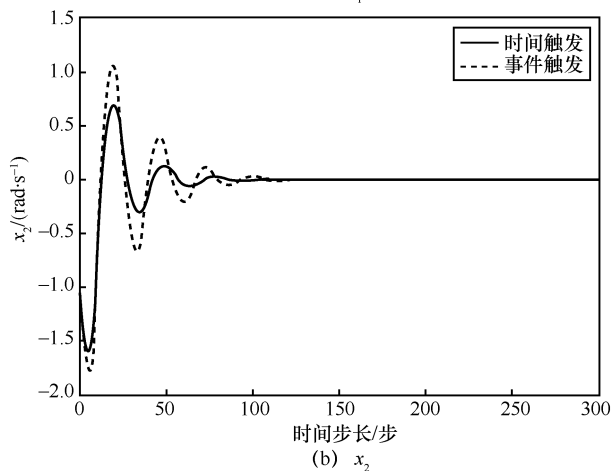
(a) x_1 (b) x_2

图6 实验二的仿真过程中的系统状态

实验二仿真过程中的系统状态和控制输入轨迹分别如图6和图7所示,系统状态最终被调节到了原点。结果表明,本文提出的方法在 γ 较大的情况下也是有效的。此外,实验二中控制输入的更新次数轨迹如图8所示,从图8可以看出,状态信息的通信成本在 $\gamma=6$ 时大大降低。在本文提出的方法中, $\gamma=6$ 时的控制输入的更新次

数为87。与 $\gamma=1.1$ 时相比,控制输入的更新次数减少,但控制性能有所下降。同时,在事件触发控制方案下的性能指标 $J(x_0, \hat{u})=82.5186$,低于上界。需要指出的是,将多项式函数作为动作网络的激活函数,系统状态可以调节到零。综上所述,如果强调优化性能指标,可以使用较小的 γ ;如果强调降低通信成本,则可以使用较大的 γ 。

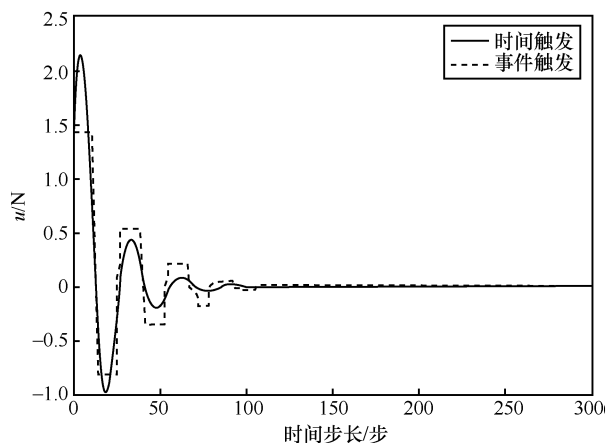


图7 实验二的仿真过程中的控制输入轨迹

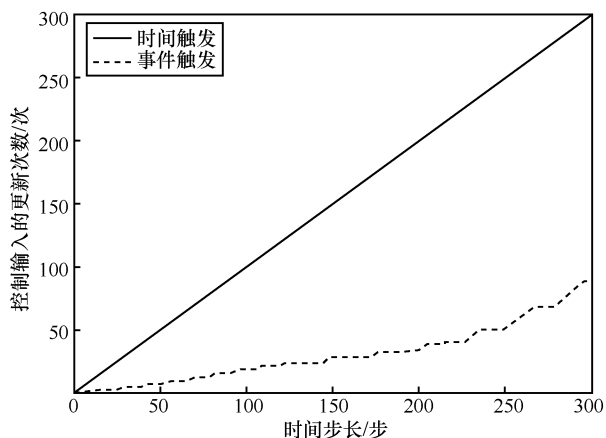


图8 实验二中控制输入的更新次数轨迹

5 结束语

本文针对非线性离散时间系统提出了一种新颖实用的事件触发近似最优控制方法。首先,笔者基于时间触发的最优值函数和最优控制律,提出了一种新型的触发条件。采用平行控制方法来实现本文提出的方法,并且在平行控制方法中使用了迭代ADP技术和神经网络技术,促进了所提方法在工程上的应用。最后通过仿真实验验证了所提方法的有效性和通用性。

参考文献:

- [1] SUTTON R S, BARTO A G. Reinforcement learning: an introduction, 2nd edition[M]. Cambridge: MIT Press, 2018.
- [2] 卢经纬, 张臻, 梁宇坤, 等. 具有尾缘襟翼风力机的恒功率反步法控制[J]. 控制理论与应用, 2017, 34(1): 61-68.
LU J W, ZHANG Z, LIANG Y K, et al. Backstepping control of constant power for wind turbines with trailing edge flaps[J]. Control Theory & Applications, 2017, 34(1): 61-68.
- [3] 张玉振, 李攀, 张维存, 等. 多模型自适应控制理论及应用[J]. 工程科学学报, 2020, 42(2): 135-143.
ZHANG Y Z, LI Q, ZHANG W C, et al. Survey of multi-model adaptive control theory and its applications[J]. Chinese Journal of Engineering, 2020, 42(2): 135-143.
- [4] 王飞跃. 平行控制与数字孪生: 经典控制理论的回顾与重铸[J]. 智能科学与技术学报, 2020, 2(3): 293-300.
WANG F Y. Parallel control and digital twins: control theory revisited and reshaped[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(3): 293-300.
- [5] 李涛, 魏庆来. 基于深度强化学习的智能暖气温度控制系统[J]. 智能科学与技术学报, 2020, 2(4): 348-353.
LI T, WEI Q L. Intelligent heating temperature control system based on deep reinforcement learning[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(4): 348-353.
- [6] LU J W, ZHANG Z, ZHOU K M, et al. Modeling and single neuron adaptive PID control of flexible piezoelectric cantilever beam[C]// Proceedings of 2016 35th Chinese Control Conference. Piscataway: IEEE Press, 2016: 957-962.
- [7] 王飞跃. 平行系统方法与复杂系统的管理和控制[J]. 控制与决策, 2004, 19(5): 485-489, 514.
WANG F Y. Parallel system methods for management and control of complex systems[J]. Control and Decision, 2004, 19(5): 485-489, 514.
- [8] WEI Q L, LI H Y, WANG F Y. Parallel control for continuous-time linear systems: a case study[J]. IEEE/CAA Journal of Automatica Sinica, 2020, 7(4): 919-928.
- [9] LU J W, WEI Q L, WANG F Y. Parallel control for optimal tracking via adaptive dynamic programming[J]. IEEE/CAA Journal of Automatica Sinica, 2020, 7(6): 1662-1674.
- [10] WANG F Y, ZHANG H G, LIU D R. Adaptive dynamic programming: an introduction[J]. IEEE Computational Intelligence Magazine, 2009, 4(2): 39-47.
- [11] 王飞跃, 陈俊龙. 智能控制方法与应用[M]. 北京: 中国科学技术出版社, 2020.
WANG F Y, CHEN J L. Intelligent control method and application[M]. Beijing: Science and Technology of China Press, 2020.
- [12] ZHOU T M, JIA C, LI H D. Optimal trade execution under jump diffusion process: a mean-VaR approach[J]. Discrete Dynamics in Nature and Society, 2018, 2018: 1-11.
- [13] WEI Q L, LIAO Z H, YANG Z Y, et al. Continuous-time time-varying policy iteration[J]. IEEE Transactions on Cybernetics, 2020, 50(12): 4958-4971.
- [14] WEI Q L, LIAO Z H, SONG R Z, et al. Self-learning optimal control for ice-storage air conditioning systems via data-based adaptive dynamic programming[J]. IEEE Transactions on Industrial Electronics, 2021, 68(4): 3599-3608.
- [15] AL-TAMIMI A, LEWIS F L, ABU-KHALAF M. Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2008, 38(4): 943-949.
- [16] LIU D R, WEI Q L. Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems[J]. IEEE Transactions on Neural Networks and Learning Systems, 2014, 25(3): 621-634.
- [17] WEI Q L, LIU D R, LIN H Q. Value iteration adaptive dynamic programming for optimal control of discrete-time nonlinear systems[J]. IEEE Transactions on Cybernetics, 2016, 46(3): 840-853.
- [18] WU Z Y, MO H D, XIONG J L, et al. Adaptive event-triggered observer-based output feedback L_∞ load frequency control for networked power systems[J]. IEEE Transactions on Industrial Informatics, 2020, 16(6): 3952-3962.
- [19] LU J W, WEI Q L, LIU Y J, et al. Event-triggered optimal parallel tracking control for discrete-time nonlinear systems[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 3429(99): 1-13.
- [20] VAMVOUDAKIS K G, FERRAZ H. Model-free event-triggered control algorithm for continuous-time linear systems with optimal performance[J]. Automatica, 2018, 87: 412-420.
- [21] WANG D, HA M M, QIAO J F. Self-learning optimal regulation for discrete-time nonlinear systems under event-driven formulation[J]. IEEE Transactions on Automatic Control, 2020, 65(3): 1272-1279.
- [22] ZHAO B, LIU D R. Event-triggered decentralized tracking control of modular reconfigurable robots through adaptive dynamic programming[J]. IEEE Transactions on Industrial Electronics, 2020, 67(4): 3054-3064.
- [23] LUO B, YANG Y, LIU D R, et al. Event-triggered optimal control with performance guarantees using adaptive dynamic programming[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(1): 76-88.
- [24] YANG X, WEI Q L. Adaptive critic designs for optimal event-driven control of a CSTR system[J]. IEEE Transactions on Industrial Informatics, 2021, 17(1): 484-493.
- [25] MU C X, WANG K, SUN C Y. Learning control supported by dynamic event communication applying to industrial systems[J]. IEEE Transactions on Industrial Informatics, 2021, 17(4): 2325-2335.

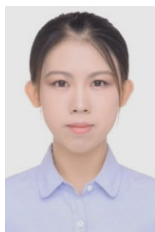
[作者简介]



廖泽华（1994—），男，中国科学院大学人工智能学院博士生，主要研究方向为自适应动态规划、强化学习、最优控制。



卢经纬（1990—），男，中国科学院自动化研究所复杂系统管理与控制国家重点实验室博士生，主要研究方向为最优控制、自适应动态规划和强化学习。



梁子钰（1997—），女，天津大学智能电网教育部重点实验室硕士生，主要研究方向为配电系统规划与运行。



魏庆来（1979—），男，博士，中国科学院自动化研究所研究员，复杂系统管理与控制国家重点实验室副主任，中国自动化学会理事，国家自然科学基金优秀青年科学基金获得者，担任 *IEEE/CAA Journal of Automatica Sinica*、*Neurocomputing*、*IEEE Transactions on Neural Networks and Learning Systems*、

《自动化学报》《智能科学与技术学报》等多个学术期刊副主编或编委。在自适应动态规划自学习最优控制领域，主持科研项目 20 项，其中省部级以上 8 项，发表 SCI 论文 120 余篇，出版专著 3 部，授权专利 10 余项，连续两年（2018 年和 2019 年）入选科睿唯安全球高被引科学家，2017 年获得中国自动化学会自然科学奖一等奖、中国自动化学会杨嘉墀科技奖，2019 年获得中国产学研合作创新奖，2020 年获得中国发明协会发明创业奖成果奖二等奖。



周天民（1990—），男，中国科学院自动化研究所复杂系统管理与控制国家重点实验室博士后，主要研究方向为复杂系统建模与调控、自适应动态规划与强化学习、神经网络，以及智能控制与决策方法在工程、社会经济系统中的应用等。