基于积分强化学习的学习感知型动态事件触发最优控制

王 珂 [1] 许振钰 [1] 张俊楠 [1] 穆朝絮 [1]

摘 要 事件触发机制, 尤其是动态事件触发机制, 近年来在控制领域引起广泛关注, 其核心挑战在于平衡控制性能与通信资源利用率. 当该机制与学习系统结合时, 这种平衡变得尤为关键, 因为还需兼顾学习效率. 针对具有未知动态的非线性连续时间系统, 提出一种集成积分强化学习、最优控制与学习感知设计的新型动态事件触发最优控制方法, 该方法采用仅含评价网络的自适应结构在线学习最优控制策略, 并通过灵活配置的动态触发规则调控数据传输. 其核心创新在于设计了一种学习感知型动态事件触发机制, 该机制通过分析评价网络权值的历史变化趋势, 构建学习感知参数, 进而自适应地调整事件触发规则中的动态阈值参数. 这使得系统能适宜地在学习关键期采用"繁忙采样"以保障控制与学习精度, 在学习平稳期切换至"空闲采样"以节约通信与计算资源, 从而实现控制性能、学习效率与资源消耗的有效平衡. 理论分析严格证明了闭环系统的渐近稳定性和权值误差的一致最终有界性. 最后, 在一个基准非线性系统和一个单连杆机械臂系统进行了仿真验证与对比实验, 结果表明与传统静态及动态事件触发方法相比, 提出方法能以更少的通信代价获得相当甚至更优的学习与控制效果.

关键词 动态事件触发机制; 自适应动态规划; 积分强化学习; 最优控制; 学习感知设计

引用格式 王珂, 许振钰, 张俊楠, 穆朝絮. 基于积分强化学习的学习感知型动态事件触发最优控制. 自动化学报, 2026, **52**(6): 1221−1233

DOI 10.16383/j.aas.c250620 **CSTR** 32138.14.j.aas.c250620

Scalable Dynamic Event-triggered Optimal Control for Nonlinear Systems via Integral Reinforcement Learning

WANG Ke[1] XU Zhen-Yu[1] ZHANG Jun-Nan[1] MU Chao-Xu[1]

Abstract Event-triggered mechanisms, particularly dynamic ones, have garnered significant interest in the control community, with a key challenge being the balance between control performance and resource utilization. This balance becomes even more crucial when integrating these mechanisms into learning systems, where learning efficiency plays a vital role. This paper presents a learning-based dynamic event-triggered framework that combines optimal control formulation, learning-aware design, and integral reinforcement learning, allowing the system to adapt the triggering process based on learning status and state changes. Using only partial knowledge of the dynamics, an optimal control policy can be learned online via a critic neural network, with data transmission flexibly regulated by dynamic triggering rules. This enables the system to intelligently adopt "busy sampling" when the weight changes dramatically, and switch to "idle sampling" during smooth learning periods to save communication/computational resources, thereby achieving an effective balance between control performance, learning efficiency, and resource consumption. Theoretical analysis rigorously proves the asymptotic stability of closed-loop systems and the uniform ultimate boundedness of weight errors. Finally, the proposed method is comparatively verified on a benchmark nonlinear system and a single-link robotic arm system, indicating that it can achieve comparable or even better learning and control effects with less communication cost.

Keywords dynamic event-triggered mechanisms; adaptive dynamic programming; integral reinforcement learning; optimal control; learning-aware design

Citation Wang Ke, Xu Zhen-Yu, Zhang Jun-Nan, Mu Chao-Xu. Scalable dynamic event-triggered optimal control for nonlinear systems via integral reinforcement learning. *Acta Automatica Sinica*, 2026, **52**(6): 1221−1233

非线性系统的最优控制一直是控制领域的核心问题, 其在无人机竞速、机器人行为优化、智能车辆巡航等领域的应用日益广泛[1−3]. 解决此类问题的思路大致可分为两类: 基于模型的最优控制和基于学

收稿日期 2025-11-11 录用日期 2026-01-30
Manuscript received November 11, 2025; accepted January 30, 2026
国家自然科学基金 (62503356, 62333016), 中国高校产学研创新基金 (2024ZY009) 资助
Supported by National Natural Science Foundation of China (62503356, 62333016) and China Higher Education Institution Industry-University-Research Innovation Fund (2024ZY009)
本文责任编委 李永明
Recommended by Associate Editor LI Yong-Ming

1. 天津大学电气自动化与信息工程学院 天津 300072
1. School of Electrical and Information Engineering, Tianjin University, Tianjin 300072

习的最优控制. 前者虽然在理论上完备, 但对于复杂的非线性系统, 往往因难以获得精确的数学模型而使其在实际应用中受限; 面对这一挑战, 以强化学习为代表的学习控制方法展现出显著优势, 它通过数据驱动的方式逼近最优控制策略, 为复杂系统的最优控制提供新的范式[4].

典型地, 根植于强化学习理念的自适应动态规划 (adaptive dynamic programming, ADP) 被认为是解决最优控制问题的有效手段[5−7]. 强化学习有助于在奖励/惩罚过程与代价函数之间建立联系, 使得 ADP 能够逼近与最优控制策略相关的哈密顿–雅可比–贝尔曼 (Hamilton-Jacobi-Bellman, HJB) 方程. ADP 根据其学习迭代算法实现方式可大致分为两类: 适用于连续时间非线性系统的策略迭代, 以及不需要初始稳定控制策略的值迭代[8]. 近年来, 结合行为–评价、辨识–行为–评价、单评价等多种神经网络架构的自适应动态规划算法取得丰硕的研究成果[9−13], 相关综述亦对 ADP 领域发展进行了系统梳理[14−15]. 研究者们还进一步将关注点扩展到安全性、控制约束与性能保障. 例如, 文献 [16] 提出分层安全强化学习控制方法并保证预设性能; 文献 [17] 研究数据驱动非线性最优控制中的非对称状态约束问题. 然而, 非线性系统的最优控制仍面临一些严峻挑战, 例如当系统动力学完全未知或部分未知时, 常用的模型辨识方法依然存在无法消弭辨识误差引起的控制性能退化现象. 为此, 积分强化学习 (integral reinforcement learning, IRL) 技术被引入自适应动态规划框架, IRL 通过对 Bellman 方程进行积分运算能够自然规避对系统动力学知识的依赖, 从而有效保证控制性能[15]. 代表性地, 文献 [18] 利用不同积分区间的采样数据实现不确定非线性系统的同策略 IRL 最优控制; 文献 [19] 设计一种基于 IRL 的反馈再学习方法, 并通过经验回放和异策略迭代提高了数据效率; 文献 [20] 进一步将 IRL 扩展至多智能体协同控制领域, 实现多四旋翼系统的自适应鲁棒姿态控制. 这些工作凸显了 IRL 在解决复杂未知非线性系统最优控制问题方面的潜力.

除了控制性能, 在现代控制系统中, 通信与计算资源的经济性亦是关键考量. 为提高效率, 需要解决“控制系统应该在什么时候作出响应”这一关键问题, 此时事件触发机制 (event-triggered mechanism, ETM) 应运而生, 其具备非周期采样和间歇控制等特点[21−22]. ETM 仅当预先设计的阈值函数被违反时, 才激活系统组件间的信号采样与传输, 阈值函数通常用于描述某一关注变量的变化趋势, 数学逻辑上体现为一个触发条件. 事件触发的一个核心挑战是在不过多牺牲期望性能的前提下最小化触发次数, 这激发了学者们对动态事件触发机制 (dynamic event-triggered mechanisms, DETM) 的探索[23−24]. 动态触发机制是在静态事件触发机制 (static event-triggered mechanisms, SETM) 基础之上改进得到的, 大致可分为两类: 基于内部动态变量 (internal dynamic variable, IDV) 的动态触发和基于动态阈值参数 (dynamic threshold parameter, DTP) 的动态触发, 前者在阈值函数中引入额外的动态而后者直接调整阈值参数. 文献 [25] 对这些动态触发方法进行全面回顾, 但未涉及其在学习系统中的应用. 本质上, 基于 IDV 或 DTP 的动态触发通过引入额外动态和设计灵活性增强事件触发控制器, 实现更优的性能均衡.

为改善学习回路的通信与计算效率, 近些年 ETM 也开始应用于学习系统这种特殊的网络系统, 形成事件触发学习方法, 此时事件触发需要实现策略最优性、学习收敛性和通信高效性之间的良好均衡[14]. 例如, 文献 [26] 提出一种基于事件触发 ADP 的最优跟踪控制算法, 并成功将其扩展至污水处理问题. 特别地, 诸多学者实现了基于学习的动态事件触发控制[27−30], 通过延长事件触发间隔进一步提升通信效率. 尽管有效降低了资源占用率, 但由于控制器在相同时间内可获得的数据减少, 整体性能下降的情况并不少见, 动态事件触发的这一潜在缺点尚未得到解决. 更为重要的是, 这些现有的事件触发方法的触发决策主要依赖于系统状态轨迹信息 (如状态误差). 这种设计可视为“状态感知”的事件触发机制, 但未能充分考虑学习过程本身的状态, 特别是忽略了神经网络权值收敛的动态特性对触发决策的影响. 这也导致一个固有缺陷: 在学习关键阶段 (如权值剧烈变化时), 可能因数据采样不足而影响学习收敛速度与控制性能; 反之, 在学习趋于平稳后, 又可能因过度采样而造成资源浪费. 换言之, 现有的事件触发机制普遍缺乏对“学习状态”的感知能力, 限制其在学习与控制之间实现更优协同的能力.

为克服上述局限, 本文针对具有未知动力学的非线性系统提出一种兼具可扩展性与学习感知能力的动态事件触发最优控制方法. 本文的主要创新点包括: 1) 设计学习感知型动态事件触发机制 (learning-aware dynamic event-triggered mechanism, LDETM), 该机制通过评估一段窗口期内的神经网络权值变化, 构造一个能够反映学习进程的感知参数; 基于此参数, LDETM 能够在两种不同特性的动态阈值参数之间切换, 从而实现触发频率的自适应调节——在学习剧烈时倾向于“繁忙采样”以保证性能, 在学习平稳时转向“空闲采样”以节省资源.

2) LDETM 方案基于单评价网络的积分强化学习结构实现, 仅需部分系统动力学知识便可在线学习近似最优控制策略, 且能扩展到两类事件触发情形. 理论分析证明了闭环学习系统的渐近稳定性和权值误差的一致最终有界性, 两个算例的对比仿真结果验证了所提方法的综合优势.

本文后续内容安排如下: 1) 阐述非线性系统最优控制问题; 2) 设计基于事件的自适应学习方案; 3) 详述 LDETM 及其事件触发规则; 4) 提供两个仿真示例验证方法有效性; 5) 总结全文工作并进行展望.

1　非线性系统最优控制问题描述

考虑如下一类具有输入仿射形式的连续时间非线性系统:

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

其中, $x \in \mathbf{R}^n$ 是系统状态, $f(x): \mathbf{R}^n \to \mathbf{R}^n$ 表示漂移动力学, $g(x): \mathbf{R}^n \to \mathbf{R}^{n\times m}$ 是输入动力学或称为输入矩阵[1], $u(t): \mathbf{R}^m$ 表示控制输入.

最优控制的目标是最小化以下代价函数:

$$J(x_0,\ u) = \int_0^\infty r\big(x(\tau),\ u(\tau)\big)\mathrm{d}\tau \tag{2}$$

其中, $r(x,\ u) = x^{\mathrm{T}}Qx + u^{\mathrm{T}}Ru$ 表示瞬时代价. 加权矩阵满足 $Q = Q^{\mathrm{T}} > 0$ 和 $R = R^{\mathrm{T}} > 0$. 如果一个反馈控制策略 $u(x)$ 在紧集 $\Omega \in \mathbf{R}^n$ 上连续, 满足 $u(0) = 0$, 能稳定系统 (1) 且使得式 (2) 有限, 则称该策略是容许的. 具体地, 可表示为 $u \in \mathcal{U}(\Omega)$, 其中 $\mathcal{U}(\Omega)$ 表示容许策略集合.

假设 1. 动力学 $f(x)$ 和 $g(x)$ 是局部 Lipschitz 的, 且满足 $f(0) = 0$. 同时假设 $f(x)$ 是未知的, 而 $g(x)$ 是已知的并且存在一个可确定的正常数 b_g, 使得 $|g(x)| \le b_g$ 对 $x \in \Omega$ 成立. 该假设中的 Lipschitz 连续性是保证系统状态轨迹存在唯一解的标准条件[8, 15], 也是后续最优控制问题可解及稳定性分析的理论基础.

接下来, 对一个给定的状态反馈控制策略 $u(t) = u(x)$ 而言, 与式 (2) 相关的值函数可以表示为

$$V(x) = V(x,\ u) = \int_t^\infty r\big(x,\ u\big)\mathrm{d}\tau \tag{3}$$

最优值函数定义为

$$V^*(x) = \min_{u\in\mathcal{U}(\Omega)} \int_t^\infty r\big(x,\ u\big)\mathrm{d}\tau \tag{4}$$

注意 $V^*(x)$ 是最优代价函数 $J^*(x)$ 的值. 沿状态轨迹对值函数求导可得

$$0 = r\big(x,\ u\big) + \nabla V_x^{\mathrm{T}}\Big(f(x) + g(x)u\Big),\ V(0) = 0 \tag{5}$$

其中, $\nabla V_x = \frac{\partial V(x)}{\partial x}$. 根据最优控制理论的既定知识[5], $V^*(x)$ 满足以下 HJB 方程:

$$\arg\min_{u(x)} H(x,\ \nabla V_x,\ u) = 0,\ V(0) = 0 \tag{6}$$

其中, $H(x,\ \nabla V_x,\ u)$ 是哈密顿函数, 定义为

$$H(x,\ \nabla V_x,\ u) = r\big(x,\ u\big) + \nabla V_x^{\mathrm{T}}\Big(f(x) + g(x)u\Big) \tag{7}$$

最终, 最优解表征的 HJB 方程和对应的最优控制策略可以表示为

$$\begin{aligned} 0 = {} & x^{\mathrm{T}}Qx + \nabla V_x^{*\mathrm{T}} f(x) - \\ & \frac{1}{4}\nabla V_x^{*\mathrm{T}} g(x)R^{-1}g^{\mathrm{T}}(x)\nabla V_x^* \end{aligned} \tag{8}$$

$$u^*(x) = -\frac{1}{2}R^{-1}g^{\mathrm{T}}(x)\nabla V_x^*,\ \ V^*(0) = 0 \tag{9}$$

众所周知, 式 (8) 和式 (9) 很难直接解析求解, 通常需要使用迭代算法, 例如策略迭代. 在传统的策略迭代算法中, 可以交替使用式 (5) 和式 (9) 迭代地求解 HJB 方程[15]. 然而, 这一经典过程需要系统动力学的全部信息. 在动态未知情形下, 可以采用 IRL 技术来实施这个迭代过程. 具体地, 对式 (5) 进行积分运算, 从而得到积分形式的 Bellman 方程:

$$V(x) = \int_t^{t+T} \big(x^{\mathrm{T}}Qx + u^{\mathrm{T}}Ru\big)\mathrm{d}\tau + V(x(t+T)) \tag{10}$$

其中, $T > 0$ 表示一个积分区间. 不难发现, 式 (10) 建立两个相邻时刻值函数差与区间代价之间的等价关系, 且不显含漂移动力学 $f(x)$.

至此, 便可在部分动态未知情形下执行策略迭代算法, 其中, 式 (10) 用于策略评估步骤, 式 (9) 用于策略改进步骤, 具体流程不再详述, 可参见文献 [31]. 接下来, 本文将提出一种基于神经网络的事件触发自适应学习方法, 以在线实现 IRL 的策略迭代算法[2].

2　单评价网络实现的事件触发自适应学习

2.1　基于评价网络的自适应学习

一般性地, 可以使用如下的评价神经网络来重构最优值函数 $V^*(x)$:

[1] 在不引起混淆的情况下, 后续表达中将省略时间变量 t.

[2] 在不引起混淆的情况下, 本文中“基于事件的”、“事件驱动的”和“事件触发”等术语将互换使用.

$$V^*(x) = w_c^{\mathrm{T}}\phi(x) + \varepsilon(x) \tag{11}$$

$$\nabla V^*(x) = \nabla\phi^{\mathrm{T}}(x)w_c + \nabla\varepsilon(x) \tag{12}$$

其中, $\phi(x):\mathbf{R}^n \to \mathbf{R}^L$ 是激活函数 (或称为基函数向量), $w_c \in \mathbf{R}^L$ 是评价网络的理想权值而 $\varepsilon(x)$ 代表网络的重构误差[3]. 将式 (12) 代入式 (9), 基于时间触发的最优控制策略可写为

$$u^*(x) = -\frac{1}{2}R^{-1}g^{\mathrm{T}}(x)\left[\nabla\phi^{\mathrm{T}}(x)w_c + \nabla\varepsilon(x)\right] \tag{13}$$

考虑到理想评价权值是一个未知向量, 因此实际学习中评价网络表示为

$$\hat{V}(x) = \hat{w}_c^{\mathrm{T}}\phi(x) \tag{14}$$

$$\nabla\hat{V}(x) = \nabla\phi^{\mathrm{T}}(x)\hat{w}_c \tag{15}$$

其中, $\hat{w}_c$ 代表估计的评价权值. 同时, 为减轻学习回路中的通信负担, 引入 ETM 来管理状态信息的传输与控制策略的计算. 为此, 定义一系列离散的触发时刻 $\{\tau_k\}_{k=0}^{\infty}$, $\tau_0 = 0$, $\tau_k \in \mathbf{R}_0^+$, $k \in \mathbf{N}_0^+$, 在触发时刻 τ_k 对系统状态进行采样, 得到状态采样值 $\hat{x}_k = x(\tau_k)$. 在两个触发时刻之间, 控制器将保持上一次触发时计算得到的控制信号不变. 定义状态采样误差为:

$$e_k(t) = \hat{x}_k - x(t), \quad t \in [\tau_k,\ \tau_{k+1}) \tag{16}$$

此时, 神经网络形式的事件触发控制策略可表示为

$$\hat{u} = \hat{u}(\hat{x}_k) = -\frac{1}{2}R^{-1}g^{\mathrm{T}}(\hat{x}_k)\nabla\phi^{\mathrm{T}}(\hat{x}_k)\hat{w}_c \tag{17}$$

其中, 评价权值是当前触发时刻的值, 即 $\hat{w}_c(t)|_{t=\tau_k}$.

然后, 近似的 Bellman 方程变为

$$e_c(t) = \underbrace{\int_t^{t+T}\left(x^{\mathrm{T}}Qx + \hat{u}^{\mathrm{T}}R\hat{u}\right)\mathrm{d}\tau}_{p_r(t)} + \hat{w}_c^{\mathrm{T}}\Delta\phi(t) \tag{18}$$

其中, $p_r(t)$ 是区间 $[t,\ t+T]$ 上的强化信号而 $\Delta\phi(t) = \phi(x(t+T)) - \phi(x(t))$ 是回归向量. 注意到, 式 (18) 是将近似值函数和控制策略代入式 (10) 得到的, 反映当前权值估计下的时序差分误差. 为最小化该误差, 采用如下的梯度下降自适应更新律, 即

$$\dot{\hat{w}}_c = -\alpha_c\frac{\Delta\phi(t)}{(\Delta\phi^{\mathrm{T}}(t)\Delta\phi(t)+1)^2}\left\{\Delta\phi^{\mathrm{T}}(t)\hat{w}_c + p_r(t)\right\} \tag{19}$$

其中, $\alpha_c > 0$ 是学习率. 此外, 如果将权值误差表示为 $\tilde{w}_c = w_c - \hat{w}_c$, 考虑 w_c 是一个常值向量, 权值误差的动态可以写为

$$\dot{\tilde{w}}_c = -\alpha_c\bar{\Delta}\phi\bar{\Delta}\phi^{\mathrm{T}}\tilde{w}_c + \frac{\alpha_c\bar{\Delta}\phi\varepsilon_B}{D_\phi} \tag{20}$$

其中, $\bar{\Delta}\phi = \Delta\phi^{\mathrm{T}}/(\Delta\phi^{\mathrm{T}}\Delta\phi+1)$, $D_\phi = \Delta\phi^{\mathrm{T}}\Delta\phi + 1$. ε_B 是不可避免的近似误差并假设是有界的[8]. 此时, 产生一个问题, 即如何获取所需的触发时刻以及将状态传输给评价网络, 然后计算基于事件的最优控制策略, 由此引出下面的静态事件触发机制[4].

2.2 基于学习的事件触发机制及分析

在本节中, 首先设计一种静态事件触发机制来生成所需的事件, 为其后续扩展奠定基础. 具体而言, SETM 由以下触发规则定义:

$$\tau_{k+1} = \inf\left\{t > \tau_k | \sigma x^{\mathrm{T}}Qx \le \mathcal{L}_u^2 b_g^2\left\|e_k(t)\right\|^2\right\} \tag{21}$$

其中, $\sigma \in (0,\ 1)$ 是阈值参数; $\mathcal{L}_u > 0$ 是设计的触发常数, 详见假设 2; b_g 为假设 1 中输入动力学 $g(x)$ 的范数上界. 关于此规则的有效性, 在下面的定理 1 中展示. 在此之前, 需要一些常用的假设条件[14−18].

假设 2. 最优控制策略 $u^*(x)$ 是 Lipschitz 的, 即满足条件 $\|u^*(\hat{x}_k) - u^*(x)\| \le \mathcal{L}_u\|e_k(t)\|$, $\mathcal{L}_u > 0$.

假设 3. 评价神经网络满足一些有界性条件: 1) 近似误差是有界的, 有 $\|\varepsilon_B\| \le \varepsilon_B$; 2) 理想评价权值是有界的, 有 $\|w_c\| \le b_w$; 3) 激活函数及其导数是有界的, 有 $\|\phi(x)\| \le b_\phi$ 和 $\|\nabla\phi\| \le b_{\nabla\phi}$; 重构误差及其导数是有界的, 有 $\|\varepsilon(x)\| \le b_\varepsilon$ 和 $\|\nabla\varepsilon\| \le b_{\nabla\varepsilon}$.

定理 1. 对于系统 (1), 其值函数和控制策略分别由式 (14) 和式 (17) 计算, 评价网络权值通过式 (19) 进行自适应更新. 此时, 如果静态事件触发规则式 (21) 和不等式 (30) 成立, 那么闭环系统将实现渐近稳定且权值误差 $\tilde{w}_c$ 是一致最终有界的.

证明. 考虑如下 Lyapunov 候选函数:

$$L_1(t) = L_{11}(t) + L_{12}(t) + L_{13}(t) \tag{22}$$

其中

$$L_{11}(t) = \int_t^{t+T} V^*(x(\tau))\mathrm{d}\tau$$

$$L_{12}(t) = \int_t^{t+T} V^*(\hat{x}_k)\mathrm{d}\tau$$

$$L_{13}(t) = \frac{1}{2}\tilde{w}_c^{\mathrm{T}}(t)\tilde{w}_c(t)$$

注意到, 第 2 项 $L_{12}(t)$ 是对事件触发的考虑, 因为一个事件相当于一个跳变状态. 这使得整个系

[3] 为紧凑表示起见, 后续将使用简洁表达式 $\nabla\phi = \frac{\partial\phi(x)}{\partial x}$ 和 $\nabla\varepsilon = \frac{\partial\varepsilon(x)}{\partial x}$.

[4] 在本文中, 如果一个事件机制的触发阈值函数仅依赖于时间或系统状态信息, 则称之为静态的; 而在动态触发机制中, 阈值函数不仅包括状态信息, 还包括一些动态演化表征的辅助变量.

统呈现了一种脉冲特性, 因此, 证明需要考虑两种情况.

1) 情况 1. 没有事件被触发, 即 $t \in [\tau_k,\ \tau_{k+1})$.

对 $L_1(t)$ 的三个分量求时间导数可得

$$\begin{cases} \dot{L}_{11}(t) = \displaystyle\int_t^{t+T} \dot{V}^*(x(\tau))\mathrm{d}\tau \\ \dot{L}_{12}(t) = \displaystyle\int_t^{t+T} \dot{V}^*(\hat{x}_k)\mathrm{d}\tau \\ \dot{L}_{13}(t) = \tilde{w}_c^{\mathrm{T}}(t)\dot{\tilde{w}}_c(t) \end{cases} \tag{23}$$

在这种情况中, $\dot{L}_{12}(t) = 0$. 首先, 分析第 1 个导数项 $\dot{L}_{11}(t)$. 对于其被积函数 $\dot{V}^*(x)$, 进行如下变换:

$$\begin{aligned} \dot{V}^*(x) = {} & (\nabla V^*)^{\mathrm{T}} f(x) + (\nabla V^*)^{\mathrm{T}} g(x)\hat{u}(\hat{x}_k) = \\ & -x^{\mathrm{T}}Qx - (u^*(x))^{\mathrm{T}}Ru^*(x) + \\ & (\nabla V^*)^{\mathrm{T}} g(x)\big(\hat{u}(\hat{x}_k) - u^*(x)\big) \le \\ & -x^{\mathrm{T}}Qx + \frac{1}{2}\left\|\nabla\phi^{\mathrm{T}}(x)w_c + \nabla\varepsilon(x)\right\|^2 + \\ & \frac{1}{2}\|g(x)\|^2\|\hat{u}(\hat{x}_k) - u^*(x)\|^2 \le \\ & -x^{\mathrm{T}}Qx + \|\nabla\phi\|^2\|w_c\|^2 + \|\nabla\varepsilon\|^2 + \\ & \frac{1}{2}\|g(x)\|^2 \underbrace{\|\hat{u}(\hat{x}_k) - u^*(x)\|^2}_{\chi_1} \end{aligned} \tag{24}$$

其中, $(\nabla V^*)^{\mathrm{T}} f(x)$ 和 $(\nabla V^*)^{\mathrm{T}} g(x)$ 已分别被式 (8) 和式 (9) 替代. 同时注意到 χ_1 可以进一步变换为

$$\begin{aligned} \chi_1 = {} & \|\hat{u}(\hat{x}_k) - u^*(\hat{x}_k) + u^*(\hat{x}_k) - u^*(x)\|^2 \le \\ & 2\underbrace{\|\hat{u}(\hat{x}_k) - u^*(\hat{x}_k)\|^2}_{\chi_2} + 2\mathcal{L}_u^2\|e_k(t)\|^2 \end{aligned} \tag{25}$$

接下来, 考虑两种形式的控制策略:

$$\begin{cases} \hat{u}(\hat{x}_k) = -\dfrac{1}{2}R^{-1}g^{\mathrm{T}}(\hat{x}_k)\nabla\phi^{\mathrm{T}}(\hat{x}_k)\hat{w}_c \\ u^*(\hat{x}_k) = -\dfrac{1}{2}R^{-1}g^{\mathrm{T}}(\hat{x}_k)\nabla\phi^{\mathrm{T}}(\hat{x}_k)w_c - \\ \qquad\qquad \dfrac{1}{2}R^{-1}g^{\mathrm{T}}(\hat{x}_k)\nabla\varepsilon(\hat{x}_k) \end{cases}$$

由此可得

$$\begin{aligned} \chi_2 \le {} & \frac{1}{2}\|R^{-1}g^{\mathrm{T}}(\hat{x}_k)\nabla\phi^{\mathrm{T}}(\hat{x}_k)\|^2\|\tilde{w}_c\|^2 + \\ & \frac{1}{2}\|R^{-1}g^{\mathrm{T}}(\hat{x}_k)\nabla\varepsilon(\hat{x}_k)\|^2 \le \\ & \frac{1}{2}\|R^{-1}\|^2 b_g^2 b_{\nabla\phi}^2\|\tilde{w}_c\|^2 + \frac{1}{2}\|R^{-1}\|^2 b_g^2 b_{\nabla\varepsilon}^2 \end{aligned} \tag{26}$$

如果将 $\frac{1}{2}\|R^{-1}\|^2 b_g^4 b_{\nabla\phi}^2$ 记为 $\chi_{g,\,\phi}$, 进一步有

$$\begin{aligned} \dot{V}^*(x) = {} & -(1-\sigma)x^{\mathrm{T}}Qx - \sigma x^{\mathrm{T}}Qx + \\ & \mathcal{L}_u^2 b_g^2\|e_k(t)\|^2 + b_\varepsilon^2 + \\ & b_{\nabla\phi}^2 b_w^2 + \chi_{g,\,\phi}\|\tilde{w}_c\|^2 + \frac{1}{2}\|R^{-1}\|^2 b_g^4 b_{\nabla\varepsilon}^2 \end{aligned} \tag{27}$$

其次, 分析第 3 项 $\dot{L}_{13}(t)$, 可以得到

$$\begin{aligned} \dot{L}_{13}(t) \le {} & -\alpha_c\tilde{w}_c^{\mathrm{T}}\bar{\Delta}\phi\bar{\Delta}\phi^{\mathrm{T}}\tilde{w}_c + \alpha_c\tilde{w}_c^{\mathrm{T}}\bar{\Delta}\phi\varepsilon_B \le \\ & -\frac{1}{2}\alpha_c\lambda_{\min}(\bar{\Delta}\phi\bar{\Delta}\phi^{\mathrm{T}})\|\tilde{w}_c\|^2 + \frac{1}{2}\alpha_c\varepsilon_B^2 \end{aligned} \tag{28}$$

综合考虑式 (27) 和式 (28), $\dot{L}_1(t)$ 可推导为

$$\begin{aligned} \dot{L}_1(t) \le {} & \int_t^{t+T}\Big(-\sigma x^{\mathrm{T}}Qx + \mathcal{L}_u^2 b_g^2\|e_k(t)\|^2\Big)\mathrm{d}\tau - \\ & \int_t^{t+T}\big((1-\sigma)x^{\mathrm{T}}Qx\big)\mathrm{d}\tau + T\chi_{g,\,\phi}\|\tilde{w}_c\|^2 - \\ & \frac{1}{2}\alpha_c\lambda_{\min}(\bar{\Delta}\phi\bar{\Delta}\phi^{\mathrm{T}})\|\tilde{w}_c\|^2 + \mathcal{E}_\Psi \end{aligned} \tag{29}$$

其中, $\mathcal{E}_\Psi = \frac{1}{2}\alpha_c\varepsilon_B^2 + T\big(b_{\nabla\phi}^2 b_w^2 + b_\varepsilon^2 + \frac{1}{2}\|R^{-1}\|^2 b_g^4 b_{\nabla\varepsilon}^2\big)$. 此时, 如果以下条件

$$\|\tilde{w}_c\| > \sqrt{\frac{2\mathcal{E}_\Psi}{\alpha_c\lambda_{\min}(\bar{\Delta}\phi\bar{\Delta}\phi^{\mathrm{T}}) - 2T\chi_{g,\,\phi}}} \tag{30}$$

和触发规则式 (21) 成立, 最终可以得到如下结果:

$$\dot{L}_1(t) \le -\int_t^{t+T}\big((1-\sigma)x^{\mathrm{T}}Qx\big)\mathrm{d}\tau < 0 \tag{31}$$

这意味着系统状态 $x(t)$ 是渐近稳定的而权值误差 $\tilde{w}_c$ 是一致最终有界收敛的.

2) 情况 2. 事件被触发, 即 $t = \tau_{k+1}$.

在这种情况下, 需要讨论系统在跳变点即触发时刻的稳定性, 重点分析 Lyapunov 函数 $L_1(t)$ 的差分值. 具体而言, 其差值定义为

$$\Delta L_1 = \int_t^{t+T}\Delta L_{11}\mathrm{d}\tau + \int_t^{t+T}\Delta L_{12}\mathrm{d}\tau + \Delta L_{13} \tag{32}$$

其中, $\Delta L_{11} = V^*(x^+) - V^*(\hat{x}_k)$, $\Delta L_{12} = V^*(\hat{x}_{k+1}) - V^*(\hat{x}_k)$, $\Delta L_{13} = \frac{1}{2}\tilde{w}_c^{+\mathrm{T}}\tilde{w}_c^+ - \frac{1}{2}\tilde{w}_c^{\mathrm{T}}\tilde{w}_c$. 进一步考虑到 $V^*(x)$ 和 $\tilde{w}_c$ 是连续的, 在触发时刻并无跳变, 因此有 $\Delta L_{11} = 0$ 和 $\Delta L_{13} = 0$. 对于 ΔL_{12}, 则有

$$\Delta L_{12} \le -\mathcal{K}(\|\hat{x}_{k+1} - \hat{x}_k\|) \le -\mathcal{K}(\|e_{k+1}(\tau_k)\|) < 0 \tag{33}$$

其中, $\mathcal{K}(\cdot)$ 是 κ-类函数[32]. 因此, 这种情况下同样可以确保 $\|\hat{x}_k\| \to 0$.

综上, 在整个事件触发学习过程中可以保证状态的渐近稳定性和权值的一致最终有界性. □

注 1. 可以观察到式 (21) 的形式与文献 [23] 中的静态触发规则类似, 后者保证了系统的输入–状

态稳定性. 从这个角度来看, 所提出的 SETM 也是有效的. 定理 1 表明, 在静态事件触发下, 通过合理选择设计参数 (σ, α_c 等), 可同时保证状态渐近稳定与权值误差有界. 然而, 此时的阈值参数 σ 固定不变, 无法根据学习进程自适应调整, 可能导致保守的触发性能.

注 2. 式 (21) 中的触发规则隐含着一个阈值函数, 记为 $\mathcal{T}_e(t)=\sigma x^{\mathrm{T}}Qx/\mathcal{L}_u^2b_g^2$, 它直接决定生成事件的频率. 具体而言, 较小的 σ 意味着较低的触发阈值 $\mathcal{T}_e$, 进而带来高频状态采样/传输; 较大的 σ 则提高阈值、减少触发次数. 后续动态机制将通过时变参数 $\sigma(t)$ 实现自适应调节.

3 可扩展的学习感知型动态事件触发机制

在传统 SETM 中, 阈值参数 σ 在整个学习期间是固定的, 这并非一种高效的通信资源利用方式, 因为实际系统中数据传输速率通常会随时间变化. 为此, 本文开发了学习感知型动态事件触发机制.

3.1 学习感知设计

首先需要指出的是, 现有的基于学习的事件触发机制仅依赖于系统状态轨迹, 因而与之对应的事件触发规则无法感应学习过程. 然而, 网络权值的变化同样不可忽视, 因为权值收敛动态直接反映学习的阶段与活跃程度. 为此, 受文献 [33−34] 中记忆型事件触发的启发, 可以定义一个评价权值的历史序列: $\{\hat{w}_c(t_0),\hat{w}_c(t_{-1}),\cdots,\hat{w}_c(t_{-n}),\cdots,\hat{w}_c(t_{-M})\}$, 其中, $\hat{w}_c(t_0)$ 表示当前时刻的评价权值, $\hat{w}_c(t_{-n})$ 是当前时刻之前的第 n 个历史权值. 定义权值差值为

$$\Delta w_c^{(n)}=\hat{w}_c(t_{-(n-1)})-\hat{w}_c(t_{-n}) \tag{34}$$

权值变化量 $\Delta w_c^{(n)}$ 的幅度直接反映该时间段内学习更新的“强度”, 为综合量化近期学习活动的整体强度, 构造如下一个学习感知参数:

$$\bar{\varrho}_w=\frac{\sum\limits_{n=1}^{M}\ell_n\|\Delta w_c^{(n)}\|}{\|\bar{w}_c\|},\ \bar{w}_c=\frac{\sum\limits_{n=1}^{M}\hat{w}_c(t_{-n})}{M} \tag{35}$$

其中, $\ell_n>0$ 是加权系数且满足 $\sum_{n=1}^{M}\ell_n=1$, $\ell_1>\ell_2>\cdots>\ell_M$. 不难看出, 该参数本质上反映了一段时间窗口期内评价权值的相对变化. 因此, 一个自然的想法是: 当权值变化剧烈时使用“繁忙采样”来保证学习性能; 相反, 权值变化较为平稳或趋于收敛时, 则可使用“空闲采样”来节省通信资源.

3.2 学习感知型动态事件触发机制

在本节中, 设计一个基于 DTP 的动态事件触发机制. 具体而言, 为阈值参数 σ 赋予一定的动态特性, 并将其与前面讨论的学习感知设计相结合, 提出一个新颖的 LDETM, 其触发规则定义为

$$\begin{cases}\tau_{k+1}=\inf\{t>\tau_k|\sigma(t)x^{\mathrm{T}}Qx\le\mathcal{L}_u^2b_g^2\|e_k(t)\|^2\}\\ \sigma(t)=(1-\alpha)\sigma_1(t)+\alpha\sigma_2(t)\end{cases} \tag{36}$$

其中, $\alpha=0$ 仅当 $\bar{\varrho}_w\ge\vartheta$, 反之则有 $\alpha=1$. 这里 ϑ 为一预设阈值, 用于判定学习是否进入平稳期. 在式 (36) 中, $\sigma(t)$ 是一个双模态的动态阈值参数, 设计为

$$\dot{\sigma}_1(t)=-\sigma_1^2(t)\xi\|e_k(t)\|^2 \tag{37a}$$

$$\sigma_2(t)=\underline{\sigma}+(\bar{\sigma}-\underline{\sigma})\tanh(\xi\|e_k(t)\|^2t) \tag{37b}$$

其中, $0<\sigma_1(0)\le\underline{\sigma}\le\sigma_2(0)<\bar{\sigma}<1$ 且有 $\xi>0$.

借助时变阈值参数 $\sigma(t)$ 和学习感知参数 $\bar{\varrho}_w$, 所提出的 LDETM 允许系统监控学习过程并适当调整触发频率. 整体工作原理可概括为:

a) 感知: 实时计算学习感知参数 $\bar{\varrho}_w$, 判断学习处于关键期还是平稳期;

b) 切换: 通过 α 选择相应的参数模态 (σ_1 和 σ_2);

c) 调节: 所选模态动态调整, 进而改变触发阈值;

d) 触发: 满足式 (36) 时执行事件触发 (采样、通信).

定理 2. 利用式 (36) 中的触发规则实施自适应学习方案时, 若满足定理 1 中的条件, 则闭环系统也是渐近稳定的且权值误差是有界收敛的; 动态阈值参数始终满足 $0<\sigma(t)<1$. 此外, 在提出的 LDETM 中, $\sigma_1(t)$ 导致繁忙采样, 而 $\sigma_2(t)$ 带来空闲采样.

证明. 这个证明通过两个部分完成, 具体地:

1) 稳定性和收敛性分析. 首先, 通过分析 $\sigma(t)$ 的两个模态证明 $\sigma(t)\in(0,\,1)$. 对于 $\sigma_1(t)$, 由式 (37a) 易知 $\dot{\sigma}_1(t)\le0$, 这意味着 $\sigma_1(t)$ 是单调非增的, 因此有 $\sigma_1(t)\le\sigma_1(0)\le\underline{\sigma}<1$; 同时, 由式 (37a) 可得

$$-\frac{\dot{\sigma}_1(t)}{\sigma_1^2(t)}=-\xi\|e_k(t)\|^2 \tag{38}$$

通过求解这个微分方程, 可以得到

$$\sigma_1(t)=\left(\sigma_1^{-1}(0)+\int_0^t\xi\|e_k(\tau)\|^2\,\mathrm{d}\tau\right)^{-1} \tag{39}$$

考虑到 $\sigma_1(0)>0$, $\xi>0$, 显然有 $\sigma_1(t)>0$ 成立; 进一步对于 $t\in\mathbf{R}_0^+$, 可以得到 $\sigma_1(t)\in(0,\,\underline{\sigma}]\in(0,\,1)$.

对于 $\sigma_2(t)$, 考虑到 $\tanh(\cdot)$ 的正有界性及其导数的正定性, 由式 (37b) 可得 $\dot{\sigma}_2 > 0$ 且 $\sigma_2(t)$ 单调不减; 随着时间的增加, $\tanh(\xi \|e_k(t)\|^2 t)$ 存在近似的增长趋势. 此外, 由于 $\bar{\sigma} - \sigma_2(t) \geq 0$, 可得 $\sigma_2(t) \leq \bar{\sigma}$; 同时, 由初始条件 $\sigma_2(0) \geq \underline{\sigma} > 0$, 可得 $\sigma_2(t) \in [\underline{\sigma}, \bar{\sigma}] \in (0, 1)$. 因此, 无论 α 取 0 或 1, 由式 (37) 合成的 $\sigma(t)$ 均满足 $0 < \sigma(t) < 1$.

其次, 将 $\sigma(t)$ 视为一个时变但有界的参数代入定理 1 的推导过程. 具体地, 在事件间隔内, 可得到类似于式 (31) 的结果, 即 $\dot{L}_1(t) < 0$; 结合触发时刻 $L(t)$ 的连续性 (与定理 1 证明中情况 2 相同), 可得状态的渐近稳定性与权值误差的最终一致有界收敛性.

2) 性能调节分析. 基于注 2 中的分析, 式 (36) 也表征着一个时变的触发阈值函数:

$$\|e_k(t)\|^2 \leq \mathcal{T}_e(t) = \frac{\sigma(t) x^{\mathrm{T}} Q x}{\mathcal{L}_u^2 b_g^2} \tag{40}$$

其中, $\sigma(t)$ 直接影响触发阈值 $\mathcal{T}_e(t)$, 该阈值越小, 事件则越容易被触发. 结合学习感知参数 $\bar{\varrho}_w$, 阈值参数 $\sigma(t)$ 将在 $\sigma_1(t)$ 和 $\sigma_2(t)$ 之间无缝切换. 具体地, 当权值训练相对剧烈时 ($\bar{\varrho}_w \geq \vartheta$), $\alpha = 0$, $\sigma(t) = \sigma_1(t)$, 由式 (37a) 可知, $\sigma_1(t)$ 会保持较小值或继续减小, 从而导致较低的触发阈值和较高的触发频率, 实现“繁忙采样”以保证学习精度. 当学习进入平稳期时 ($\bar{\varrho}_w < \vartheta$), $\alpha = 1$, $\sigma(t) = \sigma_2(t)$, 由式 (37b) 可知 $\sigma_2(t)$ 会趋向于较大的值, 从而提高触发阈值、降低触发频率, 实现“空闲采样”. 此即证明了 LDETM 能够根据学习状态自适应地调节触发频率, 从而在整体上优化学习性能与通信效率的平衡. □

注 3. 提出 LDETM 的可扩展性体现在两个方面: 一方面, 通过为阈值参数引入动态特性, 它可以在两种触发模态之间灵活切换; 另一方面, 该机制可自然扩展到离散时间或事件触发传输的场景.

上述 LDETM 是通过状态采样实现的, 一般称之为事件触发采样. 鉴于本文中的自适应学习是通过 IRL 实现的, 其中积分间隔 T 可以视为一种特殊的采样间隔, 因此 LDETM 也可以扩展到事件触发传输的场景[25]. 在事件触发传输的情况下, 可将时间表示为 $t = sT$, $s = 0, 1, 2, \cdots$, 采样的数据包 $(k, x(kT))$ 将在触发时刻 $\{kT\}_{k=0}^{\infty}$ 传输给评价网络. 触发时刻定义为

$$\begin{cases} (k+1)T = \inf\limits_{s \in \mathbf{N}_0^+} \Big\{ sT > kT \mid \Big(\sigma(sT) \|x(sT)\|^2 \leq \\ \qquad\qquad \mathcal{L}_u^2 b_g^2 \|e(sT)\|^2 \Big) \Big\} \\ \sigma(sT) = (1-\alpha)\sigma_1(sT) + \alpha \sigma_2(sT) \end{cases} \tag{41}$$

其中

$$\sigma_1((s+1)T) = \frac{\sigma_1(sT)}{1 + \sigma_1(sT) \|e(sT)\|^2} \tag{42a}$$

$$\sigma_2((s+1)T) = \frac{\sigma_2(sT) + \bar{\sigma}\epsilon \|e(sT)\|^2}{1 + \epsilon \|e(sT)\|^2} \tag{42b}$$

且初始条件为 $\sigma_1(0) \in (0, \underline{\sigma}]$, $\sigma_2(0) \in [\underline{\sigma}, \bar{\sigma}]$. $e(sT)$ 是离散时间状态采样误差, 而 $\epsilon > 0$ 是一个设计的正常数. 直观上, 式 (41) 中的 LDETM 规则可以看作是连续动态式 (36) 的离散版本. 不同之处在于, 阈值参数由两个离散序列 $\{\sigma_1(sT)\}$ 和 $\{\sigma_2(sT)\}$ 组成. 关于两个参数序列, 具体分析见引理 1.

引理 1. 给定初始条件 $\sigma_1(0) \in (0, \underline{\sigma}]$ 和 $\sigma_2(0) \in [\underline{\sigma}, \bar{\sigma}]$, 由式 (42a) 定义的序列 $\{\sigma_1(sT)\}$ 是单调非增的且满足 $\sigma_1(sT) \in (0, \underline{\sigma}]$; 由式 (42b) 定义的序列 $\{\sigma_2(sT)\}$ 是单调非减的且满足 $\sigma_2(sT) \in [\underline{\sigma}, \bar{\sigma}]$.

证明. 首先分析 $\sigma_1(sT)$. 假设 $\sigma_1(sT) > 0$ 在第 s 步时成立, 那么从式 (42a) 易知, $\sigma_1((s+1)T) > 0$ 在第 $s+1$ 步时成立. 借助数学归纳法, 同时考虑 $1 + \sigma_1(sT) \|e(sT)\|^2 \geq 1$,可以得到 $\sigma_1((s+1)T) \leq \sigma_1(sT)$. 因此, 序列 $\{\sigma_1(sT)\}$ 是单调非增的并且满足 $0 < \sigma_1(sT) \leq \sigma_1(0) \leq \underline{\sigma} < 1$.

接下来, 通过数学归纳法分析 $\sigma_2(sT)$. 假设 $\sigma_2(sT) \leq \bar{\sigma}$ 在第 s 步时成立, 那么在第 $s+1$ 步有

$$\sigma_2((s+1)T) - \bar{\sigma} = \frac{\sigma_2(sT) - \bar{\sigma}}{1 + \epsilon \|e(sT)\|^2} \leq 0 \tag{43}$$

显然, 序列 $\{\sigma_2(sT)\}$ 的上界为 $\bar{\sigma}$. 同时, 两步之间的差值可以计算为

$$\sigma_2((s+1)T) - \sigma_2(sT) = \frac{(\bar{\sigma} - \sigma_2(sT))\epsilon \|e(sT)\|^2}{1 + \epsilon \|e(sT)\|^2} \geq 0 \tag{44}$$

这意味着 $\{\sigma_2(sT)\}$ 是单调非减的, 并且进一步可以得到 $\underline{\sigma} \leq \sigma_2(0) \leq \sigma_2(sT) \leq \bar{\sigma}$. □

进一步, 根据定理 2 和引理 1, 式 (41) 中的动态触发规则也能确保闭环稳定性和学习收敛性. 图 1 展示了实现提出方法的简单架构. 此外, 基于 IRL 的事件设计, 其在采样数据控制的框架下类似于周期事件触发, 因而能天然地避免芝诺行为.

注 4. 与现有事件触发学习方法相比, LDETM 实现了从“状态感知”到“学习感知”的范式转变. 其核心优势是: 通过在线构建学习感知参数, 显式地利用历史权值变化信息来量化学习状态, 从而能够依据学习进程在不同阈值模态之间进行自适应切换. 理论分析证明了该机制可在保持稳定性与收敛性的同时, 主动优化“学习–通信”的资源分配, 解决了现有方法因感知维度单一而导致的固有性能限制.

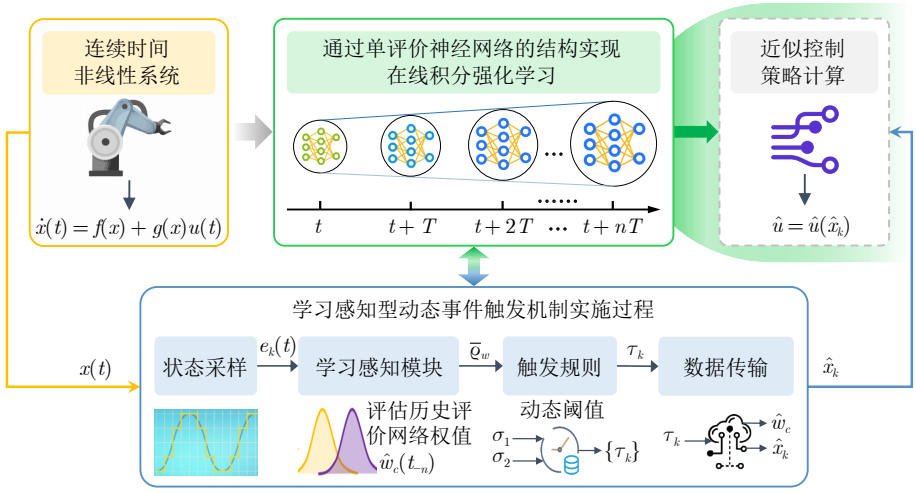


图 1 LDETM 的简要示意图

Fig. 1 A brief schematic of the proposed LDETM

4 仿真结果与分析

在本节中, 通过一个基准非线性系统和一个单连杆机械臂系统的仿真实验证明所提出的方法可以有效地平衡多个系统性能指标. 需要指出的是, 由式 (36) 实现的动态触发机制后续将称为 LDETM$_S$, 而使用式 (41) 实现的动态触发机制则称为 LDETM$_T$.

4.1 算例 1: 基准非线性系统

在本算例中, 考虑一个广泛用于新算法性能评估且具有特殊解析解的基准非线性系统[35], 其表示如下:

$$\dot{x}=f(x)+g(x)u \tag{45}$$

其中

$$f(x)=\begin{bmatrix} x_2-x_1 \\ -0.5x_1-0.5x_2+0.25x_2\cos^2(2x_1+2)\end{bmatrix}$$

$$g(x)=\begin{bmatrix} 0 \\ \cos(2x_1+2)\end{bmatrix},\quad x=[x_1,\ x_2]^{\mathrm{T}}$$

当 $Q=I_2$, $R=I_1$ 时 (I_n 表示适当维度的单位矩阵), 其最优值函数为 $V^*(x)=\frac{1}{2}x_1^2+x_2^2$. 在仿真实验中, 选择激活函数为 $\phi(x)=[x_1^2,\ x_1x_2,\ x_2^2]^{\mathrm{T}}$, 权值向量因此记为 $\hat{w}_c=[\hat{w}_{c,1},\ \hat{w}_{c,2},\ \hat{w}_{c,3}]^{\mathrm{T}}$; 学习感知设计采用递减权值, 相关参数设置为 $M=5$, $\{\ell_n\}=\{0.3,\ 0.25,\ 0.2,\ 0.15,\ 0.1\}$, $\vartheta=0.01$; 积分间隔为 $T=0.01$ s; 其他参数见表 1. 另外, 为保证持续激励条件, 在前 45 s 给系统注入探测噪声.

为公正客观地评估两种 LDETM 的性能, 将静态事件触发机制 (SETM, 取不同阈值 σ) 和传统基于 IDV 的动态事件触发机制 (IDETM) 作为对比参考. 为此, 定义如下四个性能评价指标:

表 1 两个算例的主要仿真参数

Table 1 The main simulation parameters of the two examples

	α_c	$\mathcal{L}_u$	$\sigma_1(0)$	$\sigma_2(0)$	$\underline{\sigma}$	$\bar{\sigma}$	ξ	ϑ
算例 1	0.15	4.5	0.45	0.5	0.5	0.9	25	0.01
算例 2	10.00	4.5	0.45	0.5	0.5	0.9	25	0.01

a) 事件数 (number of released events, NoRE).

b) 触发率 (triggering rate, TR): NoRE/NoTS $\times 100\%$, 其中, NoTS 表示总采样数.

c) 评价权值的近似误差 (error of critic weight, EoCW): $\|\tilde{w}_c\|$.

d) 值函数的近似误差 (error of value function, EoVF): $V^*(x)-\hat{V}(x)$.

前两个指标刻画通信效率, 后两个指标表征控制和学习性能. 值越小表示对应的系统性能越佳. 为探究触发阈值对系统性能的制约关系, 首先分析静态触发规则式 (21) 中阈值参数 σ 对学习性能和通信效率的影响. 图 2 的结果清晰地揭示了通信效率与学习控制性能之间存在紧耦合的权衡关系: 较高的 σ 值虽然能有效节约通信资源, 却不可避免地会导致学习精度的下降. 这一现象印证了在资源受限条件下, 系统的多个优化目标之间存在固有矛盾, 因而必须引入更为智能的动态调度策略.

图 3 和图 4 分别展示了通过事件采样和事件传输方式实现的 LDETM 仿真结果. 可以看到, 在移除探测噪声后, 所有变量均能稳定收敛. 尤为值得注意的是, 参数 α 与 $\sigma(t)$ 的变化趋势清晰揭示了 LDETM 的核心感知能力: 在学习初期, 权值更新较为

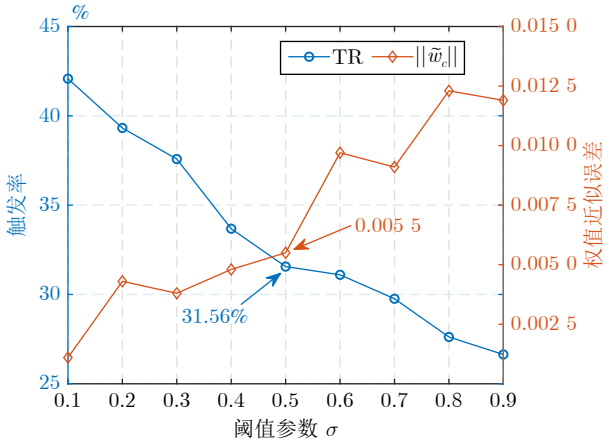


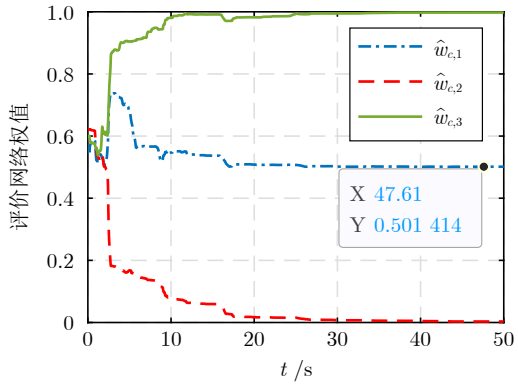
图 2　SETM 中的触发率与权值近似误差

Fig. 2　The triggering rate versus weight error approximation in SETM

剧烈, 系统通过较小的 $\sigma(t)$ 自适应地采用“繁忙采样”以获取充足数据进行快速学习; 随着权重收敛、学习过程趋于平稳, 切换至较大的 $\sigma(t)$ 转入“空闲采样”模式, 从而显著降低后期的通信与计算开销.

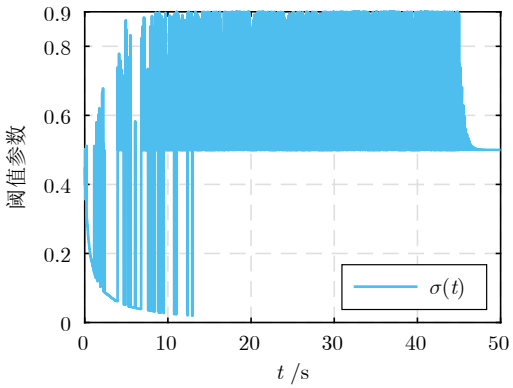
接下来, 表 2 给出不同事件触发机制在通信效率与控制性能方面的量化结果, 图 5 和图 6 给出各机制的触发间隔比较与值函数近似误差对比情况. 通过观察这些结果可以发现, SETM 的性能高度依赖于阈值参数 σ 的选择. 当 $\sigma = 0.1$ 时, 虽然控制精度最高, 但其触发率高达 42.08%. 当 σ 增大至 0.9 时, 情况则正好相反. 这表明 SETM 难以在通信与控制性能间实现自主的良好权衡. 相比之下, 提出的两种 LDETM 变体在无需手动精细调参的情况下, 分别实现了 31.28% 和 29.58% 的触发率, 接近 SETM 在牺牲大量性能后所能达到的极限水平 (26.64%). 传统的 IDETM 虽然也将触发率降至 29.94%, 但其学习与控制性能指标均劣于两种 LDETM 方法. 这表明, 仅依赖简单的内部动态变量, 难以像 LDETM 那样系统性地协调系统状态与触发逻辑.

此外, 从图 6 可知, LDETM 的触发间隔分布更具规律性和适应性, 不仅避免了 SETM 的频繁冗余触发, 也改善了 IDETM 触发时序的不均匀性或空闲采样集中在学习末期的问题. 综上, LDETM



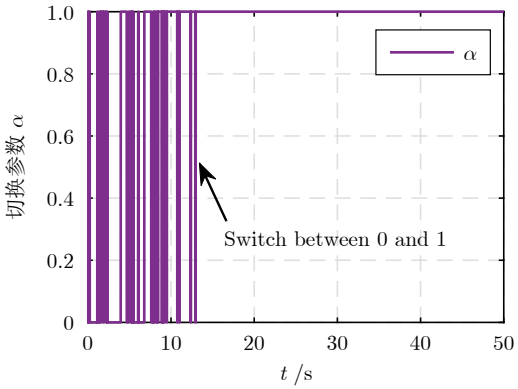
(a) 评价权值的收敛过程

(a) Convergence of critic weights



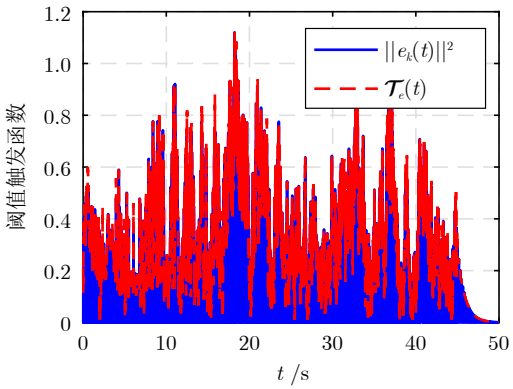
(b) 动态阈值参数 $\sigma(t)$

(b) Dynamic threshold parameter $\sigma(t)$



(c) 切换参数 α

(c) Switching parameter α

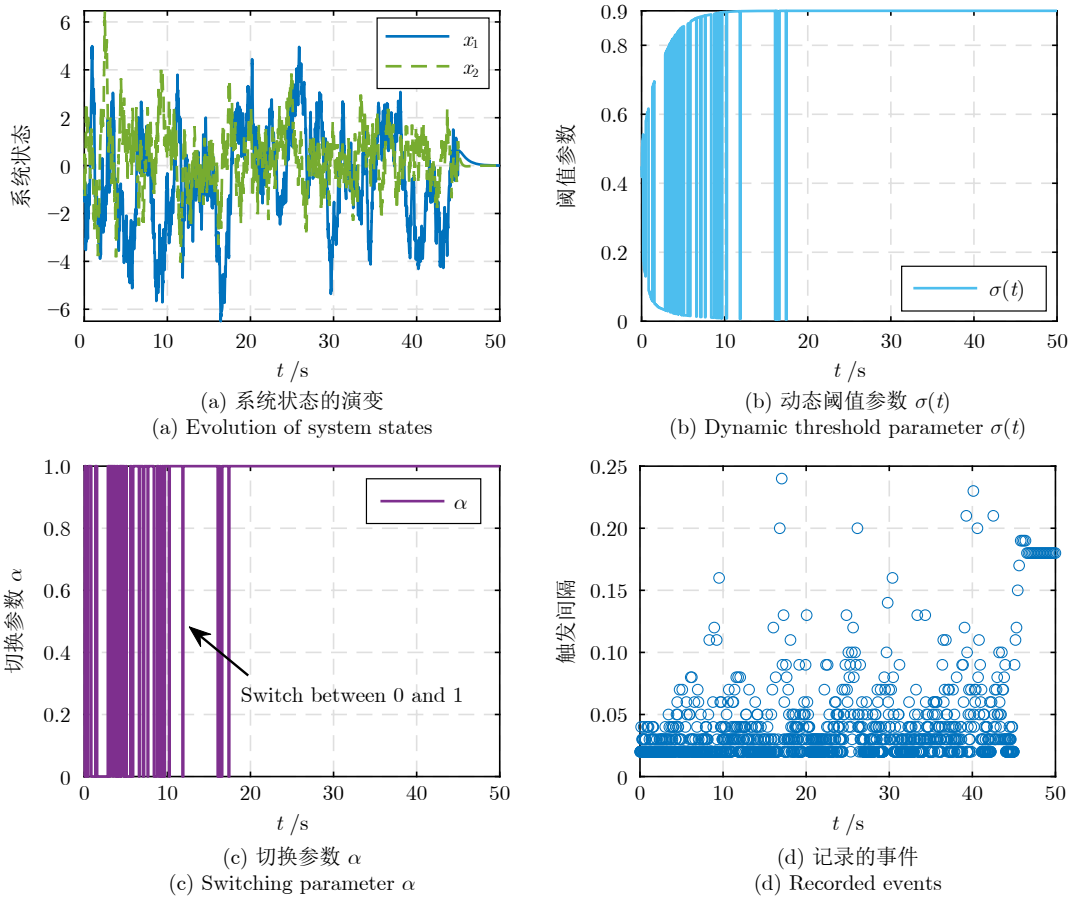


(d) 事件触发过程

(d) Event triggering process

图 3　LDETM$_S$ 的学习和触发结果

Fig. 3　Learning and triggering results of LDETM$_S$



(a) 系统状态的演变
(a) Evolution of system states

(b) 动态阈值参数 $\sigma(t)$
(b) Dynamic threshold parameter $\sigma(t)$

(c) 切换参数 α
(c) Switching parameter α

(d) 记录的事件
(d) Recorded events

图 4 LDETM$_T$ 的学习和触发结果

Fig. 4 Learning and triggering results of LDETM$_T$

表 2 算例 1 中不同事件触发机制的对比结果

Table 2 Comparative results of different ETMs in Example 1

事件触发方法	NoRE (↓)	TR (%) (↓)	EoCW (↓)	EoVF ($\times 10^{-2}$) (↓)
SETM$_{\sigma=0.1}$	2 104	42.08	**0.001 1**	**[−0.27, 0.51]**
SETM$_{\sigma=0.5}$	1 587	31.56	0.005 5	[−1.99, 1.96]
SETM$_{\sigma=0.9}$	**1 332**	**26.64**	0.011 9	[−2.76, 5.57]
LDETM$_S$	1 564	31.28	0.003 3	[−0.78, 1.26]
LDETM$_T$	1 479	29.58	0.004 3	[−0.98, 1.71]
IDETM	1 497	29.94	0.008 9	[−4.19, 2.80]

注: 粗体表示各指标最优结果.

通过其独特的触发逻辑, 显著提升了通信节省与学习控制性能之间的均衡优化能力.

4.2 算例 2: 单连杆机械臂系统

在本算例中, 考虑一个带有电机动力学的单连杆机械臂系统[36−37], 其表示如下:

$$\begin{cases} D\ddot{q} + B\dot{q} + N\sin(q) = T_r \\ I_m\dot{T}_r + H_m T_r = u - K_m\dot{q} \end{cases} \tag{46}$$

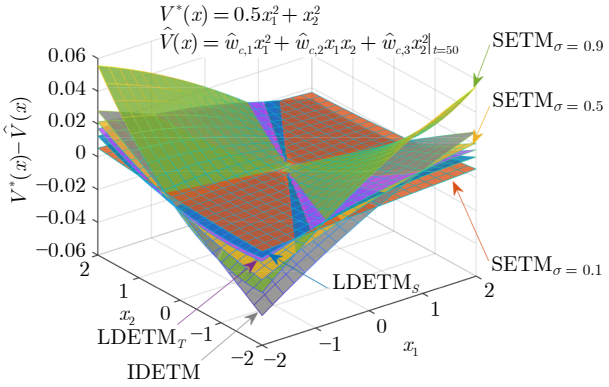


图 5 区间 x_1, $x_2 \in [-2, 2]$ 上不同 ETM 的值函数近似误差

Fig. 5 Cost approximation errors under different ETMs, which are plotted on intervals x_1, $x_2 \in [-2, 2]$

其中, q, $\dot{q}$, $\ddot{q}$ 分别表示连杆位置、速度和加速度. T_r 是由电气系统提供的扭矩而 u 是表征机电扭矩的控制输入. $D = 1$ kg·m^2 是机械惯量, $B = 1$ Nm·s/rad 是粘性摩擦系数, $N = 4$ 是正常数, $K_m = 0.2$ Nm/A 是反电动势系数, $I_m = 0.5$ H 和 $H_m = 0.1\ \Omega$ 分别是

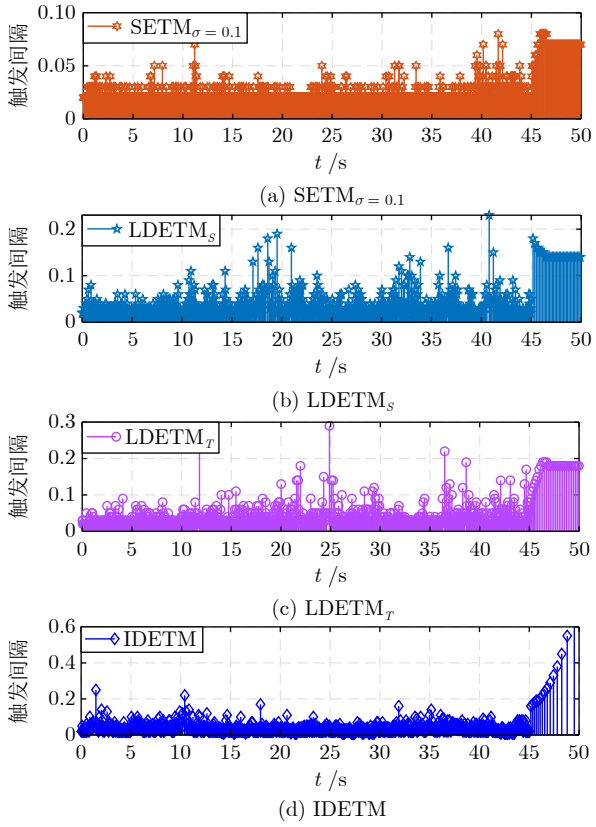


图 6　不同事件触发机制的触发间隔比较

Fig. 6　Comparison of triggering intervals for different ETMs

电枢的电感和电阻. 如果令 $x_1 = I_m D q$, $x_2 = I_m D \dot{q}$, $x_3 = I_m T_r$, 该系统可以变换为

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -I_m N \sin\left(\dfrac{x_1}{I_m D}\right) - \dfrac{B}{D} x_2 + x_3 \\ \dot{x}_3 = -\dfrac{K_m}{I_m D} x_2 - \dfrac{H_m}{I_m} x_3 + u \end{cases} \tag{47}$$

对于式 (47) 中的三维系统, 激活函数可选择为 $\phi(x) = [x_1^2,\ x_1 x_2,\ x_1 x_3,\ x_2^2,\ x_2 x_3,\ x_3^2]^{\mathrm{T}}$, 评价权值向量因此为 $\hat{w}_c = [\hat{w}_{c,1},\ \hat{w}_{c,2},\ \hat{w}_{c,3},\ \hat{w}_{c,4},\ \hat{w}_{c,5},\ \hat{w}_{c,6}]^{\mathrm{T}}$. 可令 $Q = I_3$, $R = 0.1 I_1$ 并在前 50 s 给系统注入探测噪声. 一些主要参数已列在表 1 中, 其他设置与算例 1 相同.

两种动态事件触发机制都是有效的, 图 7 展示了 LDETM$_T$ 的详细结果, 可见所有评价权值在噪声关闭前均已收敛, 表明近似最优值函数 $\hat{V}^*(x)$ 已被学习到. 还可以看到动态阈值参数 $\sigma(sT)$ 自主地在 $\sigma_1(sT)$ 与 $\sigma_2(sT)$ 之间切换, 该现象直观地证实了学习感知的有效性. 近似值函数的对比结果如图 8 所示, 可以看到两种 LDETM 变体所学习到的值函数曲面高度吻合, 表明了学习方案的一致性.

鉴于此算例的最优解难以提前获得, 本算例将高采样率的时间触发控制作为近似性能基准, 其性能可视作当前条件下学习控制所能达到的近似理论上限. 各事件触发机制的对比结果如表 3 所示. 提出的 LDETM$_S$ 最接近理论上限, 显著优于静态触发与传统动态触发; 在通信效率上, LDETM$_T$ 以最低的触发率实现与基准接近的精度, 而 IDETM 虽然触发率较低但控制性能损失明显. 结果表明, LDETM 框架能够在有效逼近高采样控制性能的前提下, 显著降低通信负担, 展现出更优的综合权衡能力.

5　结束语

本文针对连续时间非线性系统的最优控制问题, 提出一种新颖的学习感知型动态事件触发机制. 该机制突破传统事件触发方法在灵活性、可扩展性与学习感知能力方面的局限, 通过将 IRL 与单评价网络的学习结构相结合, 构建能够在线逼近最优值函数的事件驱动自适应学习方案. 理论分析严格证明了闭环系统的稳定性与权值误差的收敛性, 不同

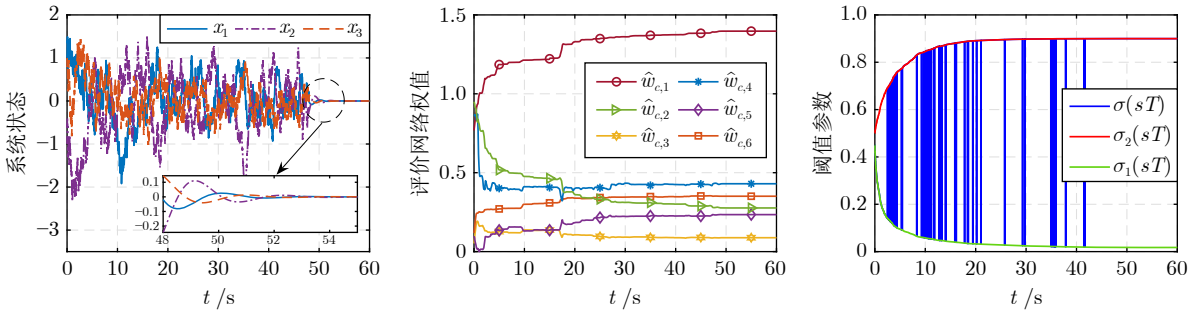
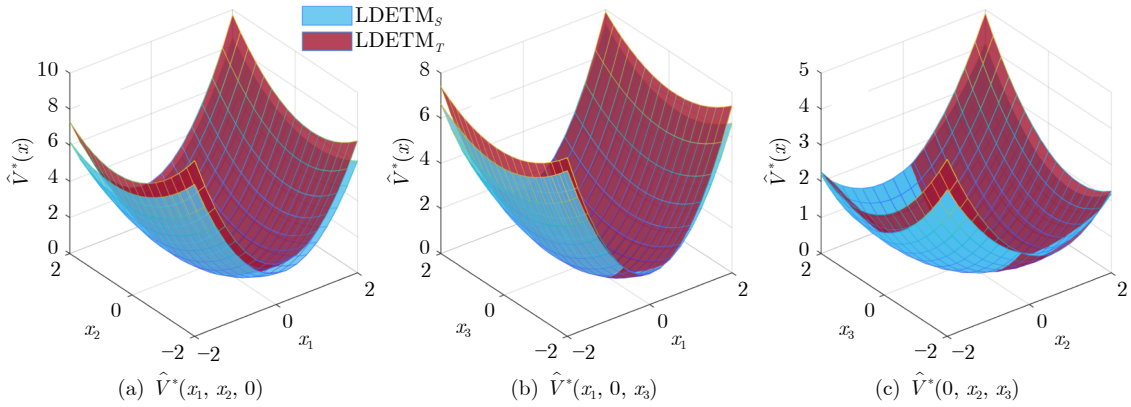


图 7　LDETM$_T$ 在单连杆机械臂系统上的学习和触发结果

Fig. 7　Learning and triggering results of LDETM$_T$ on the one-link manipulator



(a) $\hat{V}^*(x_1, x_2, 0)$ (b) $\hat{V}^*(x_1, 0, x_3)$ (c) $\hat{V}^*(0, x_2, x_3)$

图 8 近似值函数的对比结果 (x_1, x_2, $x_3 \in [-2, 2]$)

Fig. 8 Comparison of approximate value functions for x_1, x_2, $x_3 \in [-2, 2]$

表 3 算例 2 中不同事件触发机制的对比结果

Table 3 Comparative results of different ETMs in Example 2

事件触发方法	NoRE (↓)	TR (%) (↓)	EoCW (↓)
SETM$_{\sigma=0.5}$	1 412	23.53	0.029 6
LDETM$_S$	1 446	24.10	**0.020 7**
LDETM$_T$	**1 298**	21.63	0.021 3
IDETM	1 335	22.25	0.030 1

仿真案例的结果充分验证了所提方法的综合优势. 未来研究工作将侧重评估其在实际工程系统中的性能, 并增强学习过程在不确定环境下的鲁棒性.

参考文献

1 Song Y L, Romero A, Müller M, Koltun V, Scaramuzza D. Reaching the limit in autonomous racing: Optimal control versus reinforcement learning. *Science Robotics*, 2023, **8**(82): Article No. eadg1462

2 Dawson C, Gao S, Fan C. Safe control with learned certificates: A survey of neural Lyapunov, barrier, and contraction methods for robotics and control. *IEEE Transactions on Robotics*, 2023, **39**(3): 1749−1767

3 Paden B, Čáp M, Yong S Z, Yershov D, Frazzoli E. A survey of motion planning and control techniques for self-driving urban vehicles. *IEEE Transactions on Intelligent Vehicles*, 2016, **1**(1): 33−55

4 Ma C D, Li A M, Du Y L, Deng H, Yang Y D. Efficient and scalable reinforcement learning for large-scale network control. *Nature Machine Intelligence*, 2024, **6**(9): 1006−1020

5 Bertsekas D. *Reinforcement Learning and Optimal Control*. Boston: Athena Scientific, 2019.

6 Luo Biao, Hu Tian-Meng, Zhou Yu-Hao, Huang Ting-Wen, Yang Chun-Hua, Gui Wei-Hua. Survey on multiagent reinforcement learning for control and decision-making. *Acta Automatica Sinica*, 2025, **51**(3): 510−539
(罗彪, 胡天萌, 周育豪, 黄廷文, 阳春华, 桂卫华. 多智能体强化学习控制与决策研究综述. 自动化学报, 2025, **51**(3): 510−539)

7 Na J, Lv Y F, Zhang K X, Zhao J. Adaptive identifier-critic-based optimal tracking control for nonlinear systems with experimental validation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2020, **52**(1): 459−472

8 Lewis F L, Vrabie D, Vamvoudakis K G. Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers. *IEEE Control Systems Magazine*, 2012, **32**(6): 76−105

9 Wang D, Hu L Z, Wang H, Qiao J F. Nonperiodic and periodic event-triggered online H_∞ control for constrained nonlinear systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2025, **55**(1): 331−343

10 Wang K, Mu C X. Learning-based control with decentralized dynamic event-triggering for vehicle systems. *IEEE Transactions on Industrial Informatics*, 2022, **19**(3): 2629−2639

11 Jiang Y, Liu L, Feng G. Adaptive optimal tracking control of networked linear systems under two-channel stochastic dropouts. *Automatica*, 2024, **165**: Article No. 111690

12 Mailhot N, Abouheaf M, Spinello D. Model-free force control of cable-driven parallel manipulators for weight-shift aircraft actuation. *IEEE Transactions on Instrumentation and Measurement*, 2023, **73**: Article No. 2505108

13 Cohen M H, Belta C. Safe exploration in model-based reinforcement learning using control barrier functions. *Automatica*, 2023, **147**: Article No. 110684

14 Liu D R, Xue S, Zhao B, Luo B, Wei Q L. Adaptive dynamic programming for control: A survey and recent advances. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2020, **51**(1): 142−160

15 Wallace B A, Si J. Continuous-time reinforcement learning control: A review of theoretical results, insights on performance, and needs for new designs. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, **35**(8): 10199−10219

16 Tan J K, Xue S S, Li H, Guo Z H, Cao H, Chen B D. Hierarchical safe reinforcement learning control for leader-follower systems with prescribed performance. *IEEE Transactions on Automation Science and Engineering*, 2025, **22**: 19568−19581

17 Zhao M M, Wang D, Song S J, Qiao J F. Safe Q-learning for data-driven nonlinear optimal control with asymmetric state constraints. *IEEE/CAA Journal of Automatica Sinica*, 2024, **11**(12): 2408−2422

18 Liang Y L, Zhang H G, Zhang J, Ming Z Y. Event-triggered guarantee cost control for partially unknown stochastic systems via explorized integral reinforcement learning strategy. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, **35**(6): 7830−7844

19 Mu C X, Zhang Y, Sun C Y. Data-based feedback relearning control for uncertain nonlinear systems with actuator faults. *IEEE Transactions on Cybernetics*, 2023, **53**(7): 4361−4374

20 Zhou Y H, Luo B, Xu X D, Yang C H. Adaptive robust attitude control for multiple quadrotor systems via integral reinforcement learning. *IEEE Transactions on Aerospace and Elec-*

tronic Systems, 2025, **61**(4): 10799−10810

21 Zhao F, Gao W N, Liu T, Jiang Z P. Event-triggered robust adaptive dynamic programming with output feedback for large-scale systems. *IEEE Transactions on Control of Network Systems*, 2022, **10**(1): 63−74

22 Sun Chang-Yin, Mu Chao-Xu. Important scientific problems of multi-agent deep reinforcement learning. *Acta Automatica Sinica*, 2020, **46**(7): 1301−1312
(孙长银, 穆朝絮. 多智能体深度强化学习的若干关键科学问题. 自动化学报, 2020, **46**(7): 1301−1312)

23 Girard A. Dynamic triggering mechanisms for event-triggered control. *IEEE Transactions on Automatic Control*, 2015, **60**(7): 1992−1997

24 Ge X H, Han Q L, Zhang X M, Ding D R. Communication resource-efficient vehicle platooning control with various spacing policies. *IEEE/CAA Journal of Automatica Sinica*, 2024, **11**(2): 362−376

25 Ge X H, Han Q L, Zhang X M, Ding D R. Dynamic event-triggered control and estimation: A survey. *International Journal of Automation and Computing*, 2021, **18**(6): 857−886

26 Wang D, Hu L Z, Zhao M M, Qiao J F. Adaptive critic for event-triggered unknown nonlinear optimal tracking design with wastewater treatment applications. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, **34**(9): 6276−6288

27 Mu C X, Wang K, Ni Z. Adaptive learning and sampled-control for nonlinear game systems using dynamic event-triggering strategy. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, **33**(9): 4437−4450

28 Chen L, Hao F. Optimal tracking control for unknown nonlinear systems with uncertain input saturation: A dynamic event-triggered ADP algorithm. *Neurocomputing*, 2024, **564**: Article No. 126964

29 Zhang J, Yang D S, Zhang H G, Wang Y C, Zhou B W. Dynamic event-based tracking control of boiler turbine systems with guaranteed performance. *IEEE Transactions on Automation Science and Engineering*, 2024, **21**(3): 4272−4282

30 Shen H, Li Z, Wang J, Cao J. Nonzero-sum games using actor-critic neural networks: A dynamic event-triggered adaptive dynamic programming. *Information Sciences*, 2024, **662**: Article No. 120236

31 Modares H, Lewis F L. Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning. *Automatica*, 2014, **50**(7): 1780−1792

32 Khalil H K. *Nonlinear Systems (3rd edition)*. Prentice-Hall: Upper Saddle River, 2002.

33 Tian E N, Peng C. Memory-based event-triggering H_∞ load frequency control for power systems under deception attacks. *IEEE Transactions on Cybernetics*, 2020, **50**(11): 4610−4618

34 Xie L, Cheng J, Zou Y, Wu Z G, Yan H. A dynamic-memory event-triggered protocol to multiarea power systems with semi-Markov jumping parameter. *IEEE Transactions on Cybernetics*, 2023, **53**(10): 6577−6587

35 Vamvoudakis K G, Lewis F L. Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem. *Automatica*, 2010, **46**(5): 878−888

36 Xue L, Zhang T, Zhang W, Xie X J. Global adaptive stabilization and tracking control for high-order stochastic nonlinear systems with time-varying delays. *IEEE Transactions on Automatic Control*, 2018, **63**(9): 2928−2943

37 Xing L, Wen C. Dynamic event-triggered adaptive control for a class of uncertain nonlinear systems. *Automatica*, 2023, **158**: Article No. 111286

王 珂 天津大学电气自动化与信息工程学院助理研究员. 主要研究方向为强化学习与自适应动态规划, 博弈智能及应用.
E-mail: walker_wang@tju.edu.cn
(**WANG Ke** Research associate at the School of Electrical and Information Engineering, Tianjin University. His research interests include reinforcement learning and adaptive dynamic programming, and game intelligence and its applications.)

许振钰 天津大学电气自动化与信息工程学院博士研究生. 主要研究方向为强化学习, 无人集群系统优化决策.
E-mail: zhenyuxu@tju.edu.cn
(**XU Zhen-Yu** Ph.D. candidate at the School of Electrical and Information Engineering, Tianjin University. His research interests include reinforcement learning and optimal decision-making of unmanned systems.)

张俊楠 天津大学电气自动化与信息工程学院博士研究生. 主要研究方向为多智能体系统与强化学习, 机器人运动规划.
E-mail: zjn8018@tju.edu.cn
(**ZHANG Jun-Nan** Ph.D. candidate at the School of Electrical and Information Engineering, Tianjin University. His research interests include multi-agent systems and reinforcement learning, and robot motion planning.)

穆朝絮 天津大学电气自动化与信息工程学院教授. 主要研究方向为自适应学习系统, 智能无人系统优化与控制, 智能电网, 多机器人协同制造. 本文通信作者.
E-mail: cxmu@tju.edu.cn
(**MU Chao-Xu** Professor at the School of Electrical and Information Engineering, Tianjin University. Her research interests include adaptive learning system, intelligent unmanned system optimization and control, smart grids, and multi-robot collaborative manufacturing. Corresponding author of this paper.)