未知非线性零和博弈最优跟踪的事件触发控制设计

王 鼎 [1, 2, 3, 4] 胡凌治 [1, 2, 3, 4] 赵明明 [1, 2, 3, 4] 哈明鸣 [5] 乔俊飞 [1, 2, 3, 4]

摘 要 设计了一种基于事件的迭代自适应评判算法, 用于解决一类非仿射系统的零和博弈最优跟踪控制问题. 通过数值求解方法得到参考轨迹的稳定控制, 进而将未知非线性系统的零和博弈最优跟踪控制问题转化为误差系统的最优调节问题. 为了保证闭环系统在具有良好控制性能的基础上有效地提高资源利用率, 引入一个合适的事件触发条件来获得阶段性更新的跟踪策略对. 然后, 根据设计的触发条件, 采用 Lyapunov 方法证明误差系统的渐近稳定性. 接着, 通过构建四个神经网络, 来促进所提算法的实现. 为了提高目标轨迹对应稳定控制的精度, 采用模型网络直接逼近未知系统函数而不是误差动态系统. 构建评判网络、执行网络和扰动网络用于近似迭代代价函数和迭代跟踪策略对. 最后, 通过两个仿真实例, 验证该控制方法的可行性和有效性.

关键词 自适应评判设计, 事件触发控制, 神经网络, 最优跟踪控制, 稳定性分析, 零和博弈

引用格式 王鼎, 胡凌治, 赵明明, 哈明鸣, 乔俊飞. 未知非线性零和博弈最优跟踪的事件触发控制设计. 自动化学报, 2023, **49**(1): 91−101

DOI 10.16383/j.aas.c220378

Event-triggered Control Design for Optimal Tracking of Unknown Nonlinear Zero-sum Games

WANG Ding[1, 2, 3, 4] HU Ling-Zhi[1, 2, 3, 4] ZHAO Ming-Ming[1, 2, 3, 4] HA Ming-Ming[5] QIAO Jun-Fei[1, 2, 3, 4]

Abstract In this paper, an event-based iterative adaptive critic algorithm is designed to address optimal tracking control for a class of nonaffine zero-sum games. The steady control of the reference trajectory is obtained by numerical calculation. Then, the optimal tracking control problem of unknown nonlinear zero-sum games is transformed into the optimal regulation problem of corresponding error dynamics. In order to ensure that the closed-loop system possesses favourable control performance while can effectively improve the resource utilization, an appropriate event-triggering condition is introduced to obtain the tracking policy pair aperiodically. According to the designed triggering condition and the Lyapunov stability theory, the error system is proved to be asymptotically stable. In addition, four neural networks are constructed to promote the implementation of the proposed algorithm. In order to improve the accuracy of the steady control in target trajectory, the model network is used to approach the unknown system function directly instead of the error dynamic system. The critic network, the action network, and the disturbance network are constructed to obtain the approximate iterative cost function and the approximate iterative tracking policy pair. Finally, two examples are presented to demonstrate the feasibility and effectiveness of the proposed algorithm.

Key words Adaptive critic design, event-triggered control, neural networks, optimal tracking control, stability analysis, zero-sum games

Citation Wang Ding, Hu Ling-Zhi, Zhao Ming-Ming, Ha Ming-Ming, Qiao Jun-Fei. Event-triggered control design for optimal tracking of unknown nonlinear zero-sum games. *Acta Automatica Sinica*, 2023, **49**(1): 91−101

收稿日期 2022-05-09 录用日期 2022-07-13
Manuscript received May 9, 2022; accepted July 13, 2022
科技创新 2030— "新一代人工智能"重大项目 (2021ZD0112302), 北京市自然科学基金 (JQ19013), 国家自然科学基金 (62222301, 61890930-5, 62021003) 资助
Supported by National Key Research and Development Program of China (2021ZD0112302), Beijing Natural Science Foundation (JQ19013), and National Natural Science Foundation of China (62222301, 61890930-5, 62021003)
本文责任编委 刘向杰
Recommended by Associate Editor LIU Xiang-Jie

1. 北京工业大学信息学部 北京 100124 2. 计算智能与智能系统北京市重点实验室 北京 100124 3. 北京人工智能研究院 北京 100124 4. 智慧环保北京实验室 北京 100124 5. 北京科技大学自动化与电气工程学院 北京 100083
1. Faculty of Information Technology, Beijing University of Technology, Beijing 100124 2. Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing 100124 3. Beijing Institute of Artificial Intelligence, Beijing 100124 4. Beijing Laboratory of Smart Environmental Protection, Beijing 100124 5. School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083

在实际应用中, 外部干扰带来的困难总是存在的, 因此在设计控制器时不可避免地需要考虑扰动[1]. H_∞ 最优控制作为鲁棒最优控制方法的一个重要分支, 在抑制外界扰动对系统性能的影响方面得到了广泛的关注[2−4]. 二人零和博弈作为 H_∞ 最优控制的特有形式, 其核心思想是要求控制输入使得代价函

数最小化并且扰动信号使得代价函数最大化. 近年来, 对于非线性零和博弈的最优控制问题, 学者们大多通过求解相应的 Hamilton-Jacobi-Isaacs 方程, 这比求解 Hamilton-Jacobi-Bellman 方程更加困难[5]. 到目前为止, 尚缺乏有效的方法来得到解析解. 受到强化学习方法的启示, Werbos 在文献 [6] 中提出了一种自适应能力强的自适应动态规划 (Adaptive dynamic programming, ADP) 方法. 该方法能够获得一般情况下令人满意的 Hamilton-Jacobi-Isaacs 方程的数值解. 由于 ADP 的智能属性, 使得相关的方法受到了广泛的关注[7−15]. ADP 算法在实现过程中常见的两种结构形式为: 启发式动态规划和双重启发式动态规划. 此外, ADP 算法在迭代方面可分为值迭代[7−8] 和策略迭代[9−10] 两类. 值迭代算法从任意半正定初始代价函数出发, 不需要初始稳定控制策略. 值得注意的是这个初始代价函数通常设为零, 使得值迭代算法更容易实现. 策略迭代算法需要从初始稳定控制律开始, 逐步改进控制策略以达到最优控制律. 到目前为止, 已有大量工作通过采用 ADP 方法解决各种控制问题, 例如约束控制[11]、最优跟踪控制[12]、鲁棒控制[13] 和事件触发控制[14] 等, 这充分彰显了 ADP 算法的适用性和巨大潜力. 特别地, 文献 [8] 首次分析了启发式动态规划框架下值迭代算法的收敛性. 文献 [15] 研究了一种带有折扣不确定非线性动态系统的代价保证自适应最优反馈镇定问题. 在本文中, 将采用迭代 ADP 算法来获得零和博弈跟踪控制下的近似最优策略对.

近几十年来, 非线性系统的最优控制问题一直是控制工程领域的研究热点. 众所周知, 最优控制问题可以分为最优跟踪[16] 和最优调节[17] 两大类, 其中, 最优跟踪的实质是使系统的状态跟踪上预设的参考轨迹, 而最优调节的实质是使状态最终收敛到平衡点. 如今, ADP 算法已被广泛应用于解决最优轨迹跟踪问题. 文献 [18] 针对离散时间非线性系统的迭代启发式动态规划算法设计了一个性能指标, 用于解决无限时域最优轨迹跟踪问题. 文献 [12] 设计了基于执行－评判框架的局部无模型控制器, 用于在线控制系统状态跟踪上目标轨迹. 文献 [19] 通过转换代价函数, 设计一种新型的跟踪控制方法用于消除跟踪误差. 值得注意的是, 上述方法更倾向于控制模型已知的仿射系统, 而对于模型未知的非仿射系统却难以获得良好的控制效果. 为了有效地解决非仿射系统的跟踪控制问题, 文献 [20] 基于迭代双重启发式动态规划算法设计了一种数值计算的方法来获得目标轨迹的稳定控制. 在实际应用方面, 文献 [21] 设计了一种基于折扣广义值迭代的智能算法用于跟踪控制污水处理过程中溶解氧和硝态氮的质量浓度. 如今, 通过采用 ADP 算法解决轨迹跟踪问题已经得到了广泛的研究. 然而, 对于未知非线性系统零和博弈跟踪控制问题的研究却很少. 在本文中, 将采用数值计算方法求解目标轨迹的稳定控制, 然后根据这个稳定控制来获得跟踪控制律和跟踪扰动律, 进而解决未知非线性系统的零和博弈跟踪控制问题.

在系统稳定控制的基础上, 能源损耗问题已经逐渐成为工业发展的焦点之一. 事件触发控制通过设计一个合适的事件触发条件, 在这个预定义的条件被违反时对系统状态进行采样. 由于与传统的周期性时间触发控制相比, 事件触发控制能够减少控制所需的通信量和计算资源, 因此这种控制模式特别适合于嵌入式系统和网络控制系统[22]. 在事件触发控制过程中, 控制器并不是以连续的方式更新控制律, 而是在控制系统的离散采样时刻瞬间进行更新. 然而, 在两个连续的采样时刻之间存在着最大允许传输间隔, 为了达到预期的性能, 触发间隔通常选择在允许范围之内. 为此, 相关研究者在提出各种事件触发控制方法上做出了大量贡献[22−27]. 文献 [23] 设计了一种基于事件的近似最优控制器用于解决离散时间非仿射系统的控制约束问题. 文献 [24] 针对一类仿射离散时间非线性系统, 设计了一种次优的事件触发条件. 文献 [25] 针对未知非线性系统设计了一种基于事件的迭代自学习控制器, 并从输入到状态稳定性 (Input-to-state stability, ISS) 的角度分析了闭环系统的稳定性. 文献 [26] 和文献 [27] 采用基于启发式动态规划框架的事件触发控制方法分别解决了离散时间系统和连续时间系统的最优调节问题. 到目前为止, 还没有采用迭代自适应评判的事件触发控制方法解决离散时间未知非线性系统零和博弈跟踪控制问题的结果.

基于此, 本文针对离散时间未知非线性系统设计一种基于事件的近似最优轨迹跟踪算法, 目的在于解决零和博弈轨迹跟踪控制问题并减少计算量. 为了更容易获得近似最优跟踪策略对, 采用迭代自适应评判方法将最优跟踪控制问题转化为最优调节问题. 然后, 设计一个合适的事件触发条件对跟踪策略对进行阶段性更新. 值得注意的是, 事件触发的引入可能导致系统不稳定. 因此, 本文将采用 ISS-Lyapunov 方法证明被控误差系统是渐近稳定的. 最后, 通过两个仿真实例验证了本文提出算法的有效性.

在本文中, $\mathbf{R}$ 和 $\mathbf{N}$ 分别表示所有实数集和所有

非负整数集合. $\mathbf{R}^n$ 表示由全部 n-维实向量组成的欧氏空间. $\mathbf{R}^{n\times m}$ 表示 $n\times m$ 实矩阵组成的空间. Ω 表示 $\mathbf{R}^n$ 上的一个紧集. I_n 表示 $n\times n$ 维的单位矩阵. T 代表转置运算.

1　问题描述

考虑一类非仿射离散时间系统:

$$x_{k+1}=\mathcal{F}(x_k,\,u_k,\,w_k),\,k\in\mathbf{N} \tag{1}$$

式中, $x_k\in\mathbf{R}^n$ 是状态向量, $u_k\in\mathbf{R}^m$ 是控制向量, $w_k\in\mathbf{R}^r$ 是外部扰动, $\mathcal{F}(\cdot)$ 是一个未知非线性系统函数.

假设 1[28]**.** 函数 $\mathcal{F}(\cdot)$ 在包含原点的紧集 $\Omega\subset\mathbf{R}^n$ 上 Lipschitz 连续, 且系统 (1) 是可控的, 即存在连续的控制策略使得系统稳定.

考虑零和博弈跟踪控制问题, 目标是设计一个反馈控制策略 $u(x_k)$ 和一个反馈扰动策略 $w(x_k)$, 使得系统 (1) 中的状态 x_k 跟踪上预设的参考轨迹. 假设有界参考轨迹 ξ_k 满足:

$$\xi_{k+1}=\mathcal{T}(\xi_k) \tag{2}$$

式中, $\xi_k\in\mathbf{R}^n$ 和 $\mathcal{T}(\xi_k)\in\mathbf{R}^n$ 是一个可微函数. 为了便于研究最优跟踪控制问题, 将跟踪误差定义为:

$$e_k=x_k-\xi_k \tag{3}$$

此外, 定义一个关于参考轨迹 ξ_k 的稳定控制 $v(\xi_k)=[u^{\mathrm{T}}(\xi_k),\,w^{\mathrm{T}}(\xi_k)]^{\mathrm{T}}\in\mathbf{R}^{m+r}$ 使得参考轨迹满足:

$$\xi_{k+1}=\mathcal{F}\big(x_k,\,u(\xi_k),\,w(\xi_k)\big) \tag{4}$$

众所周知, 对于模型已知的仿射系统, 很容易得到相应的稳定控制. 然而, 对于模型未知的非仿射系统, 关于跟踪控制的研究依旧较少. 本文采用一种数学方法获得稳定控制 $v(\xi_k)$, 进而解决零和博弈跟踪控制问题. 为了将跟踪问题转化为调节器问题, 定义跟踪控制律和跟踪扰动律为:

$$\begin{cases}u(e_k)=u(x_k)-u(\xi_k)\\ w(e_k)=w(x_k)-w(\xi_k)\end{cases} \tag{5}$$

通过结合式 (1) ~ (5), 在时间触发机制下的关于跟踪误差的系统动态可以表示为:

$$\begin{aligned}e_{k+1}=\mathcal{F}\big(&e_k+\xi_k,\,u(e_k)+u(\xi_k),\\ &w(e_k)+w(\xi_k)\big)-\mathcal{T}(\xi_k)\end{aligned} \tag{6}$$

为了可以有效地减少计算量, 在误差系统 (6) 中引入事件触发控制方法. 定义 $\{k_j\}_{j=0}^{\infty}$ 为一个由不同采样状态组成的单调递增序列, k_j 表示第 j 个采样时刻, $j\in\mathbf{N}$. 考虑到基于事件的控制信号和扰动信号只在采样状态 $k_0,\,k_1,\,k_2,\,\cdots$ 瞬间更新, 所以需要两个零阶保持器使得跟踪控制律 $u(e_k)$ 和跟踪扰动律 $w(e_k)$ 在 $k\in[k_j,\,k_{j+1})$ 内保持不变. 因此, $u(e_k)$ 和 $w(e_k)$ 可以重新表示为:

$$\begin{cases}u(e_k)=\mu(e_{k_j})\\ w(e_k)=\pi(e_{k_j}),\,k_j\le k<k_{j+1}\end{cases} \tag{7}$$

式中, $\mu(\cdot)$ 和 $\pi(\cdot)$ 是两个辅助变量.

定义跟踪误差系统的事件触发间隔为:

$$\sigma_k=e_{k_j}-e_k \tag{8}$$

根据式 (7) 和式 (8), 跟踪控制律 $u(e_k)=\mu(e_{k_j})=\mu(\sigma_k+e_k)$, 跟踪扰动律 $w(e_k)=\pi(e_{k_j})=\pi(\sigma_k+e_k)$. 为了便于研究, 在误差系统 (6) 中引入事件触发机制, 并将其重新表示为:

$$\begin{aligned}e_{k+1}=&\mathcal{S}\big(e_k,\,\mu(e_{k_j}),\,\pi(e_{k_j})\big)=\\ &\mathcal{S}\big(e_k,\,\mu(\sigma_k+e_k),\,\pi(\sigma_k+e_k)\big)\end{aligned} \tag{9}$$

式中, $\mathcal{S}(\cdot)$ 是一个连续性函数, 并且满足 $\mathcal{S}(0,\,0,\,0)=0$.

对于零和博弈最优跟踪控制问题, 目标是找到一个控制策略 $\mu(\cdot)$ 和一个扰动策略 $\pi(\cdot)$ 分别使得代价函数最小化和最大化. 本文将代价函数定义为:

$$\begin{aligned}\mathcal{J}(e_k)=&\sum_{p=k}^{\infty}\mathcal{U}\big(e_p,\,\mu(e_{k_j}),\,\pi(e_{k_j})\big)=\\ &\sum_{p=k}^{\infty}\mathcal{U}\big(e_p,\,\mu(\sigma_p+e_p),\,\pi(\sigma_p+e_p)\big)\end{aligned} \tag{10}$$

式中, $p\in\mathbf{N}$, $\mathcal{U}(\cdot,\,\cdot,\,\cdot)$ 是效用函数且被定义为:

$$\begin{aligned}\mathcal{U}\big(e_k,\,\mu(e_{k_j}),\,\pi(e_{k_j})\big)=&\,e_k^{\mathrm{T}}\mathcal{Q}e_k+\mu^{\mathrm{T}}(e_{k_j})\mathcal{R}\mu(e_{k_j})-\\ &\gamma^2\pi^{\mathrm{T}}(e_{k_j})\pi(e_{k_j})\end{aligned} \tag{11}$$

式中, $\mathcal{Q}\in\mathbf{R}^{n\times n}$、$\mathcal{R}\in\mathbf{R}^{m\times m}$ 是两个正定矩阵, γ 是描述扰动衰减水平的正常数.

根据 Bellman 最优性原理, 误差系统 (9) 的最优代价函数满足:

$$\begin{aligned}\mathcal{J}^*(e_k)=&\min_{\mu(e_{k_j})}\max_{\pi(e_{k_j})}\big\{\mathcal{U}\big(e_k,\,\mu(e_{k_j}),\,\pi(e_{k_j})\big)+\\ &\mathcal{J}^*(e_{k+1})\big\}=\\ &\min_{\mu(e_{k_j})}\max_{\pi(e_{k_j})}\big\{e_k^{\mathrm{T}}\mathcal{Q}e_k+\mu^{\mathrm{T}}(e_{k_j})\mathcal{R}\mu(e_{k_j})-\\ &\gamma^2\pi^{\mathrm{T}}(e_{k_j})\pi(e_{k_j})+\mathcal{J}^*(e_{k+1})\big\}\end{aligned} \tag{12}$$

相应的最优策略对 $(\mu^*(e_{k_j}),\,\pi^*(e_{k_j}))$ 表示为:

$$\begin{cases}\mu^*(e_{k_j}) = \underset{\mu(e_{k_j})}{\arg\min}\left\{e_k^{\mathrm{T}}\mathcal{Q}e_k + \mu^{\mathrm{T}}(e_{k_j})\mathcal{R}\mu(e_{k_j}) - \right. \\ \qquad\qquad \left. \gamma^2\pi^{\mathrm{T}}(e_{k_j})\pi(e_{k_j}) + \mathcal{J}^*(e_{k+1})\right\} \\ \pi^*(e_{k_j}) = \underset{\pi(e_{k_j})}{\arg\max}\left\{e_k^{\mathrm{T}}\mathcal{Q}e_k + \mu^{\mathrm{T}}(e_{k_j})\mathcal{R}\mu(e_{k_j}) - \right. \\ \qquad\qquad \left. \gamma^2\pi^{\mathrm{T}}(e_{k_j})\pi(e_{k_j}) + \mathcal{J}^*(e_{k+1})\right\}\end{cases} \tag{13}$$

值得注意的是, 通过对式 (12) 右半部分求相应的一阶导数, 可获得最优策略对 $(\mu^*(e_{k_j}),\ \pi^*(e_{k_j}))$:

$$\begin{cases}\mu^*(e_{k_j}) = -\dfrac{1}{2}\mathcal{R}^{-1}\left(\dfrac{\partial e_{k+1}}{\partial \mu(e_{k_j})}\right)^{\mathrm{T}}\dfrac{\partial \mathcal{J}^*(e_{k+1})}{\partial e_{k+1}} \\ \pi^*(e_{k_j}) = \dfrac{1}{2\gamma^2}\left(\dfrac{\partial e_{k+1}}{\partial \pi(e_{k_j})}\right)^{\mathrm{T}}\dfrac{\partial \mathcal{J}^*(e_{k+1})}{\partial e_{k+1}}\end{cases} \tag{14}$$

由式 (14) 可以看出, 想要通过传统的方法直接求出最优策略对 $(\mu^*(e_{k_j}),\ \pi^*(e_{k_j}))$ 就必须知道 $\mathcal{J}^*(e_{k+1})$ 的值并且需要知道系统模型. 然而, 这对于非仿射系统来说是困难的. 因此, 在第 2 节引入一种值迭代算法, 目的是通过神经网络的逼近效应去获得近似的最优策略对.

2 事件触发最优跟踪控制设计

在本节中, 推导了零和博弈误差系统在事件触发机制下的迭代过程并给出神经网络实现方法.

2.1 值迭代算法推导

在推导迭代算法之前, 先构造三个迭代序列, 即代价函数序列 $\{\mathcal{J}^{(l)}(e_k)\}$、跟踪控制律序列 $\{\mu^{(l)}(e_{k_j})\}$ 和跟踪扰动律序列 $\{\pi^{(l)}(e_{k_j})\}$, 其中 $l \in \mathbf{N}$ 表示迭代指标. 定义初始代价函数 $\mathcal{J}^{(0)}(e_k) = 0$, 然后, $\mu^{(0)}(e_{k_j})$ 和 $\pi^{(0)}(e_{k_j})$ 可以表示为:

$$\begin{cases}\mu^{(0)}(e_{k_j}) = \underset{\mu(e_{k_j})}{\arg\min}\left\{e_k^{\mathrm{T}}\mathcal{Q}e_k + \mu^{\mathrm{T}}(e_{k_j})\mathcal{R}\mu(e_{k_j}) - \right. \\ \qquad\qquad \left. \gamma^2\pi^{\mathrm{T}}(e_{k_j})\pi(e_{k_j}) + \mathcal{J}^{(0)}(e_{k+1})\right\} \\ \pi^{(0)}(e_{k_j}) = \underset{\pi(e_{k_j})}{\arg\max}\left\{e_k^{\mathrm{T}}\mathcal{Q}e_k + \mu^{\mathrm{T}}(e_{k_j})\mathcal{R}\mu(e_{k_j}) - \right. \\ \qquad\qquad \left. \gamma^2\pi^{\mathrm{T}}(e_{k_j})\pi(e_{k_j}) + \mathcal{J}^{(0)}(e_{k+1})\right\}\end{cases} \tag{15}$$

当 $l = 1$ 时, 代价函数可表示为:

$$\mathcal{J}^{(1)}(e_k) = \min_{\mu(e_{k_j})}\max_{\pi(e_{k_j})}\left\{e_k^{\mathrm{T}}\mathcal{Q}e_k + \mu^{(0)\mathrm{T}}(e_{k_j})\mathcal{R}\mu^{(0)}(e_{k_j}) - \gamma^2\pi^{(0)\mathrm{T}}(e_{k_j})\pi^{(0)}(e_{k_j}) + \mathcal{J}^{(0)}(e_{k+1})\right\} \tag{16}$$

随着迭代指标 l 的增加, 整个学习过程可以视为不断更新迭代策略对:

$$\begin{cases}\mu^{(l)}(e_{k_j}) = \underset{\mu(e_{k_j})}{\arg\min}\left\{e_k^{\mathrm{T}}\mathcal{Q}e_k + \mu^{\mathrm{T}}(e_{k_j})\mathcal{R}\mu(e_{k_j}) - \right. \\ \qquad\qquad \left. \gamma^2\pi^{\mathrm{T}}(e_{k_j})\pi(e_{k_j}) + \mathcal{J}^{(l)}(e_{k+1})\right\} \\ \pi^{(l)}(e_{k_j}) = \underset{\pi(e_{k_j})}{\arg\max}\left\{e_k^{\mathrm{T}}\mathcal{Q}e_k + \mu^{\mathrm{T}}(e_{k_j})\mathcal{R}\mu(e_{k_j}) - \right. \\ \qquad\qquad \left. \gamma^2\pi^{\mathrm{T}}(e_{k_j})\pi(e_{k_j}) + \mathcal{J}^{(l)}(e_{k+1})\right\}\end{cases} \tag{17}$$

和代价函数:

$$\mathcal{J}^{(l+1)}(e_k) = \min_{\mu(e_{k_j})}\max_{\pi(e_{k_j})}\left\{e_k^{\mathrm{T}}\mathcal{Q}e_k + \mu^{(l)\mathrm{T}}(e_{k_j})\mathcal{R}\mu^{(l)}(e_{k_j}) - \gamma^2\pi^{(l)\mathrm{T}}(e_{k_j})\pi^{(l)}(e_{k_j}) + \mathcal{J}^{(l)}(e_{k+1})\right\} \tag{18}$$

在误差允许范围内, 选择一个正常数 ϵ 作为迭代停止的参数, 即当 $\left|\mathcal{J}^{(l+1)}(e_k) - \mathcal{J}^{(l)}(e_k)\right| \le \epsilon$ 时, 停止迭代. 根据文献 [8], 可以进一步推导出当 $l \to \infty$ 时, 代价函数 $\mathcal{J}^{(l)}(e_k) \to \mathcal{J}^{(\infty)}(e_k) = \mathcal{J}^*(e_k)$, 跟踪控制律 $\mu^{(l)}(e_{k_j}) \to \mu^{(\infty)}(e_{k_j}) = \mu^*(e_{k_j})$ 和跟踪扰动律 $\pi^{(l)}(e_{k_j}) \to \pi^{(\infty)}(e_{k_j}) = \pi^*(e_{k_j})$.

2.2 神经网络实现

为了实现迭代自适应评判算法, 构建四个神经网络, 即模型网络、评判网络、执行网络和扰动网络, 目的是通过连续逼近方法获得近似最优策略对 $(\mu^*(e_{k_j}), \pi^*(e_{k_j}))$. 通过建立模型网络得到原系统的近似状态 $\hat{x}_{k+1}$ 并求出参考轨迹的稳定控制 $v(\xi_k)$. 此外, 通过训练另外三个神经网络得到近似代价函数和近似策略对. 总体而言, 本文提出的事件触发最优跟踪控制方法如图 1 所示.

1) 模型网络. 由于原系统是未知的, 需要构造一个模型网络来识别系统动态. 目的是得到近似的系统状态 $\hat{x}_{k+1}$, 其神经网络表达式为:

$$\hat{x}_{k+1} = \varpi_{m2}^{\mathrm{T}}\eta\left(\varpi_{m1}^{\mathrm{T}}x_{mk} + b_{m2}\right) + b_{m1} \tag{19}$$

式中, $x_{mk} = [x_k^{\mathrm{T}},\ u^{\mathrm{T}}(x_k),\ w^{\mathrm{T}}(x_k)]^{\mathrm{T}}$, $\eta(\cdot)$ 是一个有界的激活函数并将其选择为双曲正切函数 $\tanh(\cdot)$, $\varpi_{m1} \in \mathbf{R}^{(n+m+r)\times N}$ 和 $\varpi_{m2} \in \mathbf{R}^{N\times n}$ 是权值矩阵, b_{m1} 和 b_{m2} 是随机初始化的阈值向量. 训练模型网络的目标是最小化下面的性能指标:

$$E_{mk} = \frac{1}{2}\left(\hat{x}_{k+1} - x_{k+1}\right)^{\mathrm{T}}\left(\hat{x}_{k+1} - x_{k+1}\right) \tag{20}$$

本文运用 Matlab 神经网络工具箱来训练模型网络. 对于跟踪控制问题, 目标是确保系统状态轨迹 x_k 能够跟踪上参考轨迹 ξ_k. 然后, 式 (4) 的神经网络表达式可以写为:

$$\xi_{k+1} = \varpi_{m2}^{\mathrm{T}}\eta\left(\varpi_{m1}^{\mathrm{T}}\xi_{mk} + b_{m2}\right) + b_{m1} \tag{21}$$

式中, $\xi_{mk} = [\xi_k^{\mathrm{T}},\ v^{\mathrm{T}}(\xi_k)]^{\mathrm{T}}$. 通过观察式 (21), 可以发

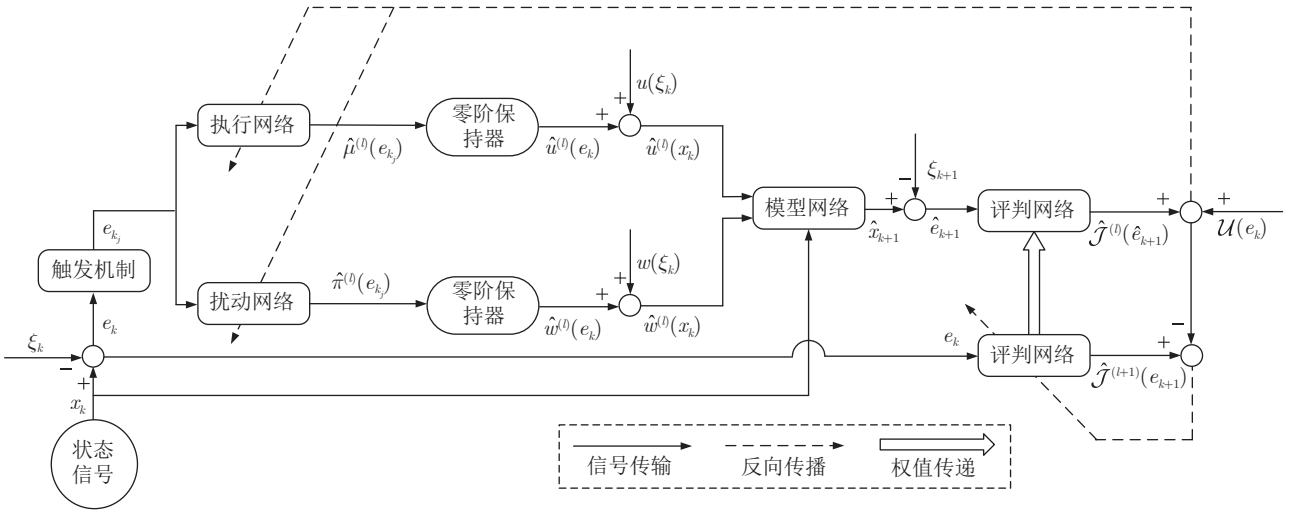

图 1　基于事件的零和博弈跟踪控制方法示意图

Fig. 1　The simple structure of the event-based zero-sum game tracking control method

现只有稳定控制 $v(\xi_k)$ 是未知的. 因此, 可以根据数值求解的方法计算出稳定控制 $v(\xi_k)$. 通过结合式 (19) 和式 (21), 近似跟踪误差 $\hat{e}_{k+1}$ 可以表示为:

$$\hat{e}_{k+1} = \varpi_{m2}^{\mathrm{T}}\eta\big(\varpi_{m1}^{\mathrm{T}}e_{mk}\big) \tag{22}$$

式中, $e_{mk} = x_{mk} - \xi_{mk}$.

2) 评判网络: 设计评判网络的目标是通过输入 e_k 来获得近似的代价函数, 其神经网络表达式为:

$$\hat{\mathcal{J}}^{(l)}(e_k) = \varpi_{c2}^{(l)\mathrm{T}}\eta\big(\varpi_{c1}^{(l)\mathrm{T}}e_k\big) \tag{23}$$

式中, $\varpi_{c1}^{(l)} \in \mathbf{R}^{n\times N}$ 和 $\varpi_{c2}^{(l)} \in \mathbf{R}^{N}$ 是评判网络的权值矩阵. 通过训练评判网络, 相应的性能指标可以表示为:

$$E_c^{(l)} = \frac{1}{2}\big(\hat{\mathcal{J}}^{(l)}(e_k) - \mathcal{J}^{(l)}(e_k)\big)^{\mathrm{T}}\times \big(\hat{\mathcal{J}}^{(l)}(e_k) - \mathcal{J}^{(l)}(e_k)\big) \tag{24}$$

根据梯度下降算法, 评判网络的权值矩阵更新规则为:

$$\varpi_{c1}^{(l+1)} = \varpi_{c1}^{(l)} - \alpha_c\frac{\partial E_c^{(l)}}{\partial \varpi_{c1}^{(l)}},\ \varpi_{c2}^{(l+1)} = \varpi_{c2}^{(l)} - \alpha_c\frac{\partial E_c^{(l)}}{\partial \varpi_{c2}^{(l)}} \tag{25}$$

式中, $\alpha_c \in (0, 1)$ 为评判网络的学习率.

3) 执行网络: 使用执行网络来输出近似跟踪控制律, 其神经网络表达式为:

$$\hat{\mu}^{(l)}(e_{k_j}) = \varpi_{a2}^{(l)\mathrm{T}}\eta\big(\varpi_{a1}^{(l)\mathrm{T}}e_{k_j}\big) \tag{26}$$

式中, $\varpi_{a1}^{(l)} \in \mathbf{R}^{n\times N}$ 和 $\varpi_{a2}^{(l)} \in \mathbf{R}^{N\times m}$ 是相应的权值矩阵. 通过训练执行网络, 定义其性能指标为:

$$E_a^{(l)} = \frac{1}{2}\big(\hat{\mu}^{(l)}(e_{k_j}) - \mu^{(l)}(e_{k_j})\big)^{\mathrm{T}}\times \big(\hat{\mu}^{(l)}(e_{k_j}) - \mu^{(l)}(e_{k_j})\big) \tag{27}$$

执行网络的权值矩阵更新方式可以表示为:

$$\varpi_{a1}^{(l+1)} = \varpi_{a1}^{(l)} - \alpha_a\frac{\partial E_a^{(l)}}{\partial \varpi_{a1}^{(l)}},\ \varpi_{a2}^{(l+1)} = \varpi_{a2}^{(l)} - \alpha_a\frac{\partial E_a^{(l)}}{\partial \varpi_{a2}^{(l)}} \tag{28}$$

式中, $\alpha_a \in (0, 1)$ 为执行网络的学习率.

4) 扰动网络: 与执行网络类似, 使用扰动网络来输出近似跟踪扰动律, 其神经网络表达式为:

$$\hat{\pi}^{(l)}(e_{k_j}) = \varpi_{d2}^{(l)\mathrm{T}}\eta\big(\varpi_{d1}^{(l)\mathrm{T}}e_{k_j}\big) \tag{29}$$

式中, $\varpi_{d1}^{(l)} \in \mathbf{R}^{n\times N}$ 和 $\varpi_{d2}^{(l)} \in \mathbf{R}^{N\times r}$ 是相应的权值矩阵. 定义其性能指标为:

$$E_d^{(l)} = \frac{1}{2}\big(\hat{\pi}^{(l)}(e_{k_j}) - \pi^{(l)}(e_{k_j})\big)^{\mathrm{T}}\times \big(\hat{\pi}^{(l)}(e_{k_j}) - \pi^{(l)}(e_{k_j})\big) \tag{30}$$

扰动网络的权值矩阵更新方式可以表示为:

$$\varpi_{d1}^{(l+1)} = \varpi_{d1}^{(l)} - \alpha_d\frac{\partial E_d^{(l)}}{\partial \varpi_{d1}^{(l)}},\ \varpi_{d2}^{(l+1)} = \varpi_{d2}^{(l)} - \alpha_d\frac{\partial E_d^{(l)}}{\partial \varpi_{d2}^{(l)}} \tag{31}$$

式中, $\alpha_d \in (0, 1)$ 为扰动网络的学习率.

3　稳定性分析

本文引入了一个合适的触发条件. 然后, 根据这个触发条件, 使用 Lyapunov 方法来证明基于事件的零和博弈误差系统的稳定性.

引理 1. 假设存在一个正常数 Γ 使得 $\|e_{k+1}\| \leq$

$\Gamma\|\sigma_k\|+\Gamma\|e_k\|$, 则触发间隔$\|\sigma_k\|$满足不等式条件

$$\|\sigma_k\| \le \frac{\big(1-(2\Gamma)^{k-k_j}\big)(1+\Gamma)}{1-2\Gamma}\|e_{k_j}\| \tag{32}$$

式中, $\Gamma \in (0, 0.5)$.

证明. 根据式 (8), 易得:

$$\sigma_{k+1} = e_{k_j} - e_{k+1} \tag{33}$$

通过观察式 (33) 并结合引理 1 和式 (8), 可得:

$$\begin{aligned}\|\sigma_{k+1}\| \le{}& \|e_{k_j}\| + \|e_{k+1}\| \le\\ & \|e_{k_j}\| + \Gamma\|\sigma_k\| + \Gamma\|e_k\| \le\\ & \|e_{k_j}\| + \Gamma\|\sigma_k\| + \Gamma(\|e_{k_j}\| + \|\sigma_k\|) =\\ & 2\Gamma\|\sigma_k\| + (1+\Gamma)\|e_{k_j}\|\end{aligned} \tag{34}$$

同理, 也可得:

$$\|\sigma_k\| \le 2\Gamma\|\sigma_{k-1}\| + (1+\Gamma)\|e_{k_j}\| \tag{35}$$

然后, 根据递归性质, 可得:

$$\begin{aligned}\|\sigma_k\| \le{}& 2\Gamma\|\sigma_{k-1}\| + (1+\Gamma)\|e_{k_j}\| \le\\ & (2\Gamma)^2\|\sigma_{k-2}\| + (1+2\Gamma)(1+\Gamma)\|e_{k_j}\| \le\\ & \vdots\\ & (2\Gamma)^{k-k_j}\|\sigma_{k_j}\| + \Big(1+2\Gamma+(2\Gamma)^2+\\ & (2\Gamma)^3+\cdots+(2\Gamma)^{k-k_j-1}\Big)(1+\Gamma)\|e_{k_j}\|\end{aligned} \tag{36}$$

式中, $\sigma_{k_j}=0$. 通过对不等式 (36) 化简, 可以得到触发间隔$\|\sigma_k\|$满足条件 (32). □

为了有效地进行事件触发控制, 引入一个可调参数 β. 然后, 设计事件触发条件为:

$$\|\sigma_k\| \le \sigma_T = \frac{\beta\big(1-(2\Gamma)^{k-k_j}\big)(1+\Gamma)}{1-2\Gamma}\|e_{k_j}\| \tag{37}$$

式中, σ_T 是事件触发阈值, $\beta \in (0, 1)$, $\Gamma \in (0, 0.5)$. 值得注意的是只有当不等式 (37) 不满足的时候才对系统进行采样并更新跟踪控制律和跟踪扰动律.

假设 2. 存在 $\mathcal{K}_\infty$ 函数 δ_1 和 δ_2, 一阶连续可导函数 $\mathcal{V}: \mathbf{R}^n \to \mathbf{R} \ge 0$, 正常数 ϑ、ζ 和 L, 使得:

$$\delta_1(\|e_k\|) \le \mathcal{V}(e_k) \le \delta_2(\|e_k\|) \tag{38}$$

$$\begin{aligned}&\mathcal{V}\Big(\mathcal{S}\big(e_k, \mu(\sigma_k+e_k), \pi(\sigma_k+e_k)\big)\Big) - \mathcal{V}(e_k) \le\\ &\quad -\vartheta\mathcal{V}(e_k) + \zeta\|\sigma_k\|\end{aligned} \tag{39}$$

$$\delta_1^{-1}(\|e_k\|) \le L\|e_k\| \tag{40}$$

在这个假设条件中, 如果不等式 (38) 和 (39) 成立, 则函数 $\mathcal{V}$ 视为系统 (9) 的 ISS-Lyapunov 函数[29]. 根据 Lyapunov 理论所述, 如果系统 (9) 存在一个满足式 (38) 和式 (39) 的 ISS-Lyapunov 函数 $\mathcal{V}$, 那么这个系统就具有 ISS. 然后, 根据设置的触发条件研究系统 (9) 的渐近稳定问题.

定理 1. 根据设计的触发条件 (37), 如果基于事件的误差系统 (9) 满足假设 2 中的条件, 并且存在正数 $\lambda \in \big(0, 1/(k-k_j)\big)$, $k \ne k_j$, $\vartheta \in (0, 1)$, $\beta \in (0, 1)$, 使得:

$$\frac{\beta\zeta}{\vartheta^2} \le \frac{(1-2\Gamma)\big(1-\lambda(k-k_j)\big)}{L(1+\Gamma)} \tag{41}$$

那么, 误差系统 (9) 是渐近稳定的.

证明. 下面将分为两种情况进行证明: 系统处于事件未触发时刻和系统处于事件触发时刻.

情况 1. 事件没有被触发, 即 $k \in (k_j, k_{j+1})$. 根据不等式 (38), 可得:

$$\|e_{k_j}\| \le \delta_1^{-1}\big(\mathcal{V}(e_{k_j})\big) \tag{42}$$

结合式 (40) 和式 (42), 可得:

$$\|e_{k_j}\| \le \delta_1^{-1}\big(\mathcal{V}(e_{k_j})\big) \le L\mathcal{V}(e_{k_j}) \tag{43}$$

由于在这种情况下事件没有被触发, 所以触发条件 (37) 恒成立. 然后, 代入式 (39), 可得:

$$\begin{aligned}&\mathcal{V}\Big(\mathcal{S}\big(e_k, \mu(\sigma_k+e_k), \pi(\sigma_k+e_k)\big)\Big) - \mathcal{V}(e_k) \le\\ &\quad -\vartheta\mathcal{V}(e_k) + \zeta\frac{\beta\big(1-(2\Gamma)^{k-k_j}\big)(1+\Gamma)}{1-2\Gamma}\|e_{k_j}\|\end{aligned} \tag{44}$$

再将式 (43) 代入式 (44), 可得:

$$\begin{aligned}\mathcal{V}(e_{k+1}) \le{}& (1-\vartheta)\mathcal{V}(e_k) +\\ & \frac{\beta\big(1-(2\Gamma)^{k-k_j}\big)(1+\Gamma)}{1-2\Gamma}\zeta L\mathcal{V}(e_{k_j})\end{aligned} \tag{45}$$

进一步地, 可得:

$$\begin{aligned}\mathcal{V}(e_k) \le{}& (1-\vartheta)\mathcal{V}(e_{k-1}) +\\ & \frac{\beta\big(1-(2\Gamma)^{k-1-k_j}\big)(1+\Gamma)}{1-2\Gamma}\zeta L\mathcal{V}(e_{k_j})\end{aligned} \tag{46}$$

将式 (46) 代入式 (45), 可得:

$$\begin{aligned}\mathcal{V}(e_{k+1}) \le{}& (1-\vartheta)\Big((1-\vartheta)\mathcal{V}(e_{k-1}) +\\ & \frac{\beta\big(1-(2\Gamma)^{k-1-k_j}\big)(1+\Gamma)}{1-2\Gamma}\zeta L\mathcal{V}(e_{k_j})\Big) +\\ & \frac{\beta\big(1-(2\Gamma)^{k-k_j}\big)(1+\Gamma)}{1-2\Gamma}\zeta L\mathcal{V}(e_{k_j})\end{aligned} \tag{47}$$

根据递归性质, 将式 (47) 展开, 可得:

$$\begin{aligned}\mathcal{V}(e_k) \le{}& (1-\vartheta)^{k-k_j}\mathcal{V}(e_{k_j}) + \frac{1-(1-\vartheta)^{k-k_j}}{\vartheta}\times\\ & \frac{\beta\big(1-(2\Gamma)^{k-k_j}\big)(1+\Gamma)}{1-2\Gamma}\zeta L\mathcal{V}(e_{k_j})\end{aligned} \tag{48}$$

接着, 根据不等式 (41), 可得:

$$\frac{\beta\zeta L(1+\Gamma)}{\vartheta^2(1-2\Gamma)} \le 1-\lambda(k-k_j) \tag{49}$$

式中, $\Gamma \in (0, 0.5)$. 考虑到 $\vartheta \in (0, 1)$, 有:

$$\frac{\beta\zeta L(1+\Gamma)}{\vartheta(1-2\Gamma)} \le \vartheta-\vartheta\lambda(k-k_j) \tag{50}$$

由于 k 和 k_j 是离散时刻, 这就使得在事件不触发的情况下有 $k-k_j \ge 1$, 进而得到:

$$(1-\vartheta)^{k-k_j} \le 1-\vartheta \tag{51}$$

根据式 (51), 可得:

$$\vartheta \le 1-(1-\vartheta)^{k-k_j} \tag{52}$$

将式 (52) 代入式 (50), 可得:

$$\frac{\beta\zeta L(1+\Gamma)}{\vartheta(1-2\Gamma)} \le 1-(1-\vartheta)^{k-k_j}-\vartheta\lambda(k-k_j) \tag{53}$$

此外, 易得出:

$$\begin{cases} 1-(1-\vartheta)^{k-k_j} < 1 \\ 1-(2\Gamma)^{k-k_j} < 1 \end{cases} \tag{54}$$

因此, 根据式 (53) 和式 (54), 可得:

$$\begin{aligned} &(1-\vartheta)^{k-k_j}+\frac{1-(1-\vartheta)^{k-k_j}}{\vartheta}\times \\ &\qquad \frac{\beta\big(1-(2\Gamma)^{k-k_j}\big)(1+\Gamma)}{1-2\Gamma}\zeta L \le \\ &\qquad (1-\vartheta)^{k-k_j}+\frac{\beta\zeta L(1+\Gamma)}{\vartheta(1-2\Gamma)} \le \\ &\qquad 1-\vartheta\lambda(k-k_j) \end{aligned} \tag{55}$$

根据 $\mathcal{V}(e_{k_j})>0$, 式 (48) 可重新表示为:

$$\mathcal{V}(e_k) \le \mathcal{V}(e_{k_j})-\vartheta\lambda(k-k_j)\mathcal{V}(e_{k_j}) \tag{56}$$

定义一个函数 $\mathcal{P}$ 为:

$$\mathcal{P}(e_k) = \mathcal{V}(e_{k_j})-\vartheta\lambda(k-k_j)\mathcal{V}(e_{k_j}) \tag{57}$$

则有:

$$0 < \mathcal{V}(e_k) \le \mathcal{P}(e_k) \tag{58}$$

根据式 (38) 和式 (57), 可得:

$$\Delta\mathcal{P}(e_k) \le -\vartheta\lambda\delta_1(\|e_{k_j}\|) < 0 \tag{59}$$

式中, $\Delta\mathcal{P}(e_k) = \mathcal{P}(e_{k+1})-\mathcal{P}(e_k)$. 因此, 在这种情况下, 误差系统 (9) 是渐近稳定的.

情况 2. 事件在 k 时刻触发, 即 $k=k_j$, 所以触发间隔 $\sigma_k=0$. 然后式 (39) 可以重新写为:

$$\mathcal{V}(e_{k_j+1})-\mathcal{V}(e_{k_j}) \le -\vartheta\mathcal{V}(e_{k_j}) \tag{60}$$

将式 (38) 代入式 (60), 可得:

$$\Delta\mathcal{V}(e_{k_j}) \le -\vartheta\delta_1(\|e_{k_j}\|) < 0 \tag{61}$$

式中, $\Delta\mathcal{V}(e_{k_j}) = \mathcal{V}(e_{k_j+1})-\mathcal{V}(e_{k_j})$. 因此, 在事件发生触发的情况下, 误差系统 (9) 也是渐近稳定的.□

4　仿真实验

为了进一步验证本文算法的有效性, 本节将其应用于两个具体系统.

例 1. 考虑一个离散时间非仿射系统:

$$x_{k+1} = \begin{bmatrix} 0.97x_{1k}+0.97x_{2k}u(x_k)+0.97w(x_k) \\ 0.97(1+x_{1k}^2)u(x_k)+0.97x_{2k} \end{bmatrix} + \begin{bmatrix} 0 \\ 0.97u^3(x_k)+0.97w(x_k) \end{bmatrix} \tag{62}$$

式中, $x_k=[x_{1k}, x_{2k}]^{\mathrm{T}} \in \mathbf{R}^2$ 是状态变量, $u(x_k) \in \mathbf{R}$ 是控制律, $w(x_k) \in \mathbf{R}$ 是扰动律, 初始状态 $x_0=[0.5, -0.5]^{\mathrm{T}}$.

为了有效地控制这个非仿射非线性系统, 一些基本参数在表 1 中给出. 在自适应评判实现中, 运用 Matlab 神经网络工具箱训练结构为 4-8-2 的模型网络用于识别未知系统, 其中学习率 $\alpha_m=0.02$, 训练误差为 10^{-8}. 在训练过程中, 收集了 1 000 个数据样本, 每个样本有 500 个训练步来学习动态信息. 然后, 用另外 1 000 个数据样本验证模型网络的逼近性能. 根据式 (20) 的性能指标, 训练的状态误差平方和如图 2 所示. 此外, 训练模型网络后, 记录并保持最终权值不变.

表 1　两个仿真实验的主要参数

Table 1　Main parameters of two experimental examples

符号	例 1	例 2
$\mathcal{Q}$	$0.01I_2$	$0.1I_2$
$\mathcal{R}$	I	I
γ	0.01	0.1
Γ	0.2	0.35
β	1/6	7/27
ϵ	10^{-5}	10^{-5}

定义需要跟踪的参考轨迹为:

$$\xi_{k+1} = \begin{bmatrix} 0.9963\xi_{1k}+0.0498\xi_{2k} \\ -0.2492\xi_{1k}+0.9888\xi_{2k} \end{bmatrix} \tag{63}$$

式中, $\xi_k=[\xi_{1k}, \xi_{2k}]^{\mathrm{T}} \in \mathbf{R}^2$, $\xi_0=[0.1, 0.2]^{\mathrm{T}}$. 因此, 初始跟踪误差 $e_0=x_0-\xi_0=[0.4, -0.7]^{\mathrm{T}}$. 根据式 (21) 和 Matlab 中函数 Fsolve 的性质, 可以直接求出参考轨迹的稳定控制 $v(\xi_k)$. 根据式 (25)、式 (28) 和式 (31), 开始训练评判网络、执行网络和扰动网络. 应用迭代自适应评判算法分别对结构同为 2-8-1 三个网络进行更新并在 $[-0.5, 0.5]$ 中随机选取评判网络的初始权值 ϖ_{c1} 和 ϖ_{c2}, 在 $[-0.1, 0.1]$ 中随机选取

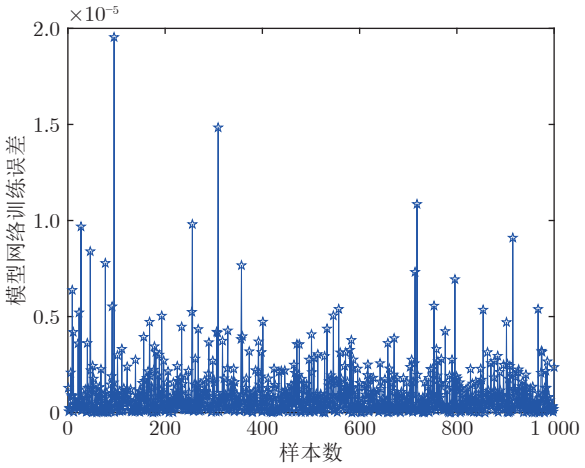

图 2 模型网络训练误差 (例 1)

Fig. 2 The training errors of the model network (Example 1)

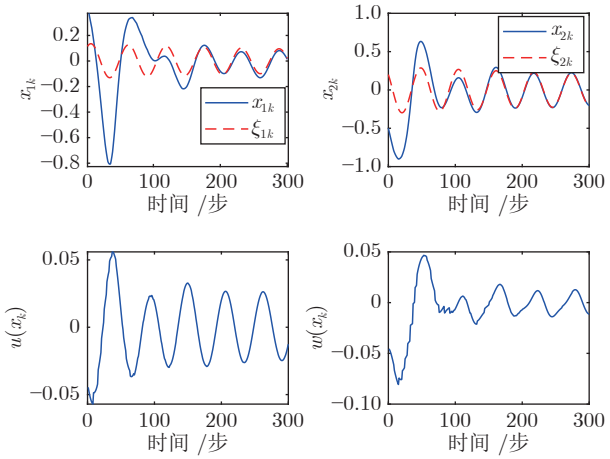

图 3 系统状态、控制律和扰动律轨迹 (例 1)

Fig. 3 Trajectories of the state, the control law, and the disturbance law (Example 1)

执行网络和扰动网络的初始权值 ϖ_{a1}、ϖ_{a2}、ϖ_{d1} 和 ϖ_{d2}, 学习率 $\alpha_c = \alpha_a = \alpha_d = 0.05$. 在这三个网络的训练过程中, 将基于事件的机制应用于执行网络和扰动网络. 为了保证足够的学习和获得满意的性能, 如果在迭代算法中达到预定的精度 10^{-5}, 即 $\left|\mathcal{J}^{(l+1)}(e_k) - \mathcal{J}^{(l)}(e_k)\right| \leq 10^{-5}$, 则可以终止评判网络、执行网络和扰动网络的训练过程. 此外, 在每次迭代过程中, 设置三个神经网络各 500 次训练次数用于满足预设的精度.

根据表 1 中的参数, 事件触发阈值可以表示为:

$$\sigma_T = \frac{1 - 0.4^{k-k_j}}{3}\|e_{k_j}\| \tag{64}$$

只有当事件触发间隔 $\|\sigma_k\| > \sigma_T$ 时, 跟踪控制律和跟踪扰动律才会进行更新. 根据以上信息, 原系统的状态 x_k、控制律 $u(x_k)$ 和扰动律 $w(x_k)$ 的响应曲线如图 3 所示. 从图 3 可以清楚地看到, 状态变量 x_{1k} 和 x_{2k} 最终可以很好地跟踪上预设的参考轨迹. 此外, 跟踪误差 e_k、跟踪控制律 $u(e_k)$ 和跟踪扰动律 $w(e_k)$ 的响应曲线如图 4 所示. 通过图 4 可以很明显地看出该控制方法的跟踪控制律和跟踪扰动律曲线为阶梯状. 实验发现, 基于事件的跟踪控制律和跟踪扰动律在 300 个时间步上只更新了 74 次, 有效地提高了资源利用率. 根据 Matlab 中的函数 Fsolve 求出的稳定控制 $v(\xi_k)$ 曲线如图 5 所示. 值得注意的是, 稳定控制 $v(\xi_k) = [u^{\mathrm{T}}(\xi_k),\ w^{\mathrm{T}}(\xi_k)]^{\mathrm{T}}$ 并且存在 $u(e_k) = u(x_k) - u(\xi_k)$ 和 $w(e_k) = w(x_k) - w(\xi_k)$. 触发阈值的演化曲线如图 6 所示, 随着跟踪误差信号的变化, 触发阈值有趋近于零的趋势.

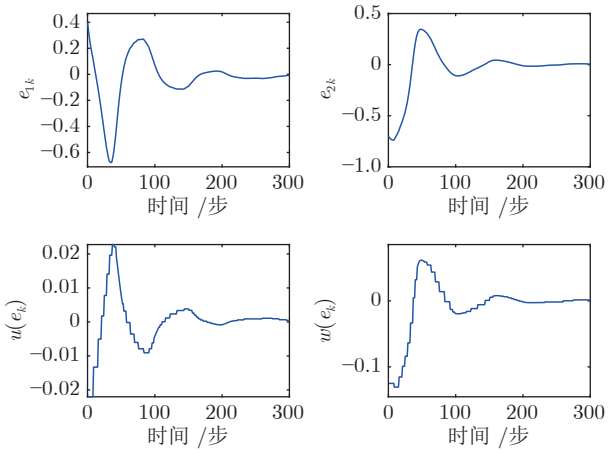

图 4 跟踪误差、跟踪控制律和跟踪扰动律轨迹 (例 1)

Fig. 4 Trajectories of the tracking error, the tracking control law, and the tracking disturbance law (Example 1)

例 2. 考虑如下所示的扭摆装置[30].

$$\begin{cases} \dfrac{\mathrm{d}\theta}{\mathrm{d}t} = \omega \\ J_k \dfrac{\mathrm{d}\omega}{\mathrm{d}t} = u(x_k) - Mg\ell_k \sin\theta - f_d \dfrac{\mathrm{d}\theta}{\mathrm{d}t} + w(x_k) \end{cases} \tag{65}$$

式中, θ 是当前角度, ω 表示角速度. $M = 1/3$ kg 和 $\ell_k = 3/2$ m分别表示摆杆的质量和长度, $J_k = 4/3\ M\ell_k^2$ 表示转动惯量, $f_d = 2$ 表示摩擦系数, $g = 9.8\ \mathrm{m/s^2}$ 表示重力加速度. 然后, 得到离散时间状态空间方程为:

$$x_{k+1} = \begin{bmatrix} 0.1x_{2k} + x_{1k} \\ -0.49\sin(x_{1k}) + 0.8x_{2k} \end{bmatrix} + \begin{bmatrix} 0 \\ 0.1 \end{bmatrix} u(x_k) + \begin{bmatrix} 0 \\ 0.1 \end{bmatrix} w(x_k) \tag{66}$$

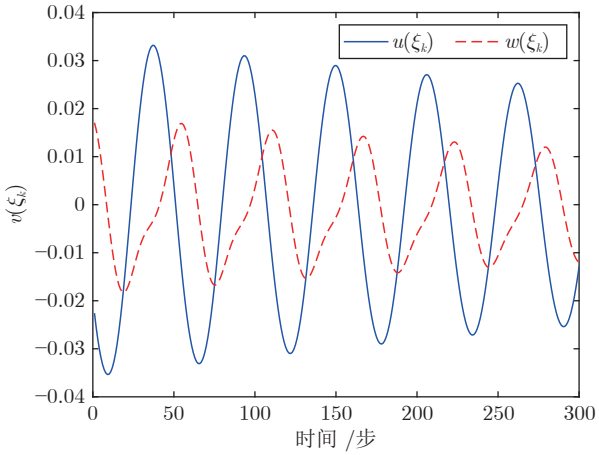

图 5　稳定控制 $v(\xi_k)$ (例 1)

Fig. 5　The steady control $v(\xi_k)$ (Example 1)

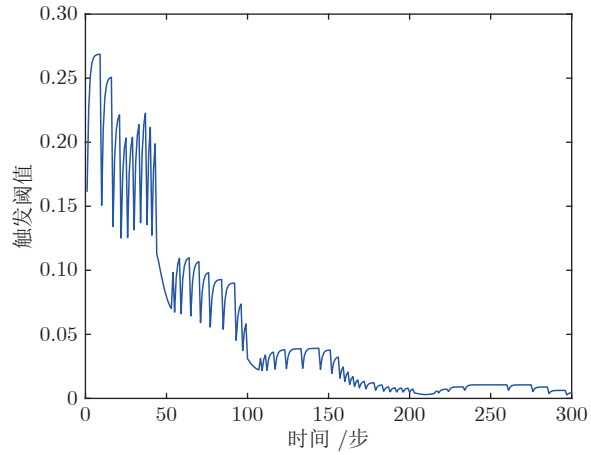

图 6　触发阈值 σ_T (例 1)

Fig. 6　The triggering threshold σ_T (Example 1)

式中, $x_k = [x_{1k}, x_{2k}]^{\mathrm{T}} = [\theta_k, \omega_k]^{\mathrm{T}}$ 是状态向量并设置初始状态 $x_0 = [0.3, -0.3]^{\mathrm{T}}$. 同样, 这个扭摆系统的一些基本参数在表 1 中给出. 模型网络的训练过程与例 1 相似, 通过进行一个有效的学习阶段, 训练的状态误差平方和如图 7 所示, 训练结束后保持权值不变. 定义相关的参考轨迹为:

$$\xi_{k+1} = \begin{bmatrix} \xi_{1k} + 0.1\xi_{2k} \\ -0.2492\xi_{1k} + 0.9888\xi_{2k} \end{bmatrix} \tag{67}$$

式中, $\xi_0 = [-0.1, 0.2]^{\mathrm{T}}$. 初始跟踪误差 $e_0 = x_0 - \xi_0 = [0.4, -0.5]^{\mathrm{T}}$. 然后, 根据所设计的算法去训练评判网络、执行网络和扰动网络. 这三个网络的迭代次数, 学习率和初始权值的选择与例 1 相同.

为了采用基于事件的控制方法, 根据表 1 中的参数, 事件触发阈值可以表示为:

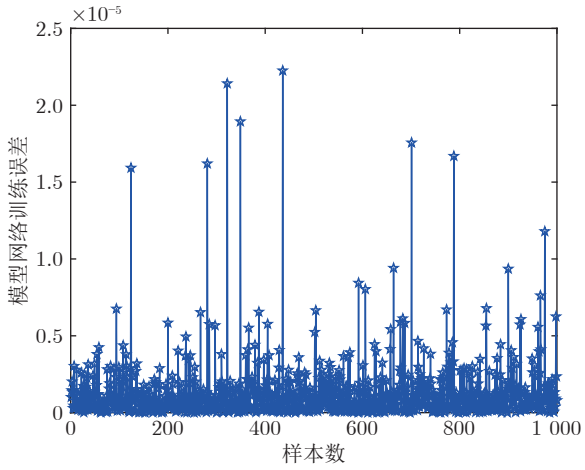

图 7　模型网络训练误差 (例 2)

Fig. 7　The training errors of the model network (Example 2)

$$\sigma_T = \frac{7(1 - 0.4^{k-k_j})}{6}\|e_{k_j}\| \tag{68}$$

同样, 原系统的状态 x_k、控制律 $u(x_k)$ 和扰动律 $w(x_k)$ 的响应曲线如图 8 所示. 跟踪误差 e_k、跟踪控制律 $u(e_k)$ 和跟踪扰动律 $w(e_k)$ 的响应曲线如图 9 所示. 此外, 通过实验发现跟踪控制律和跟踪扰动律在 200 个时间步上只更新了 76 次. 触发阈值的演化曲线如图 10 所示. 结果表明, 本文提出的控制算法可以很好地控制未知非线性零和博弈系统跟踪上预设的参考轨迹并且极大程度地提高了资源利用率.

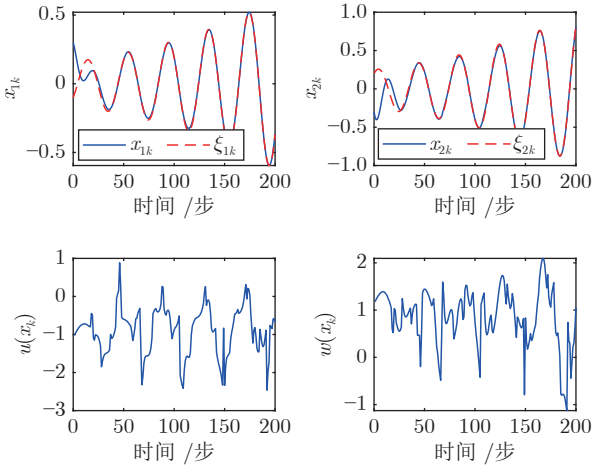

图 8　系统状态、控制律和扰动律轨迹 (例 2)

Fig. 8　Trajectories of the state, the control law, and the disturbance law (Example 2)

5　结束语

针对未知非线性系统的零和博弈轨迹跟踪问题, 提出了一种基于迭代自适应评判的事件触发控

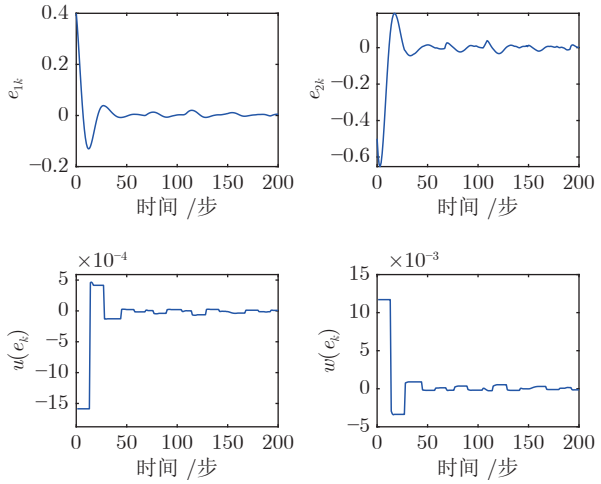

图 9 跟踪误差、跟踪控制律和跟踪扰动律轨迹 (例 2)

Fig. 9 Trajectories of the tracking error, the tracking control law, and the tracking disturbance law (Example 2)

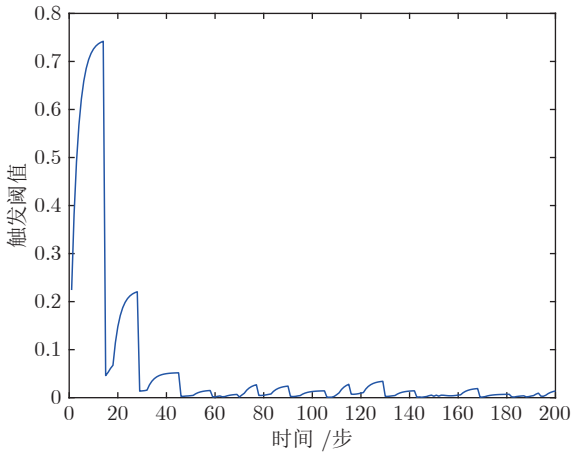

图 10 触发阈值 σ_T (例 2)

Fig. 10 The triggering threshold σ_T (Example 2)

制方法, 极大地减少了计算量. 首先, 通过建立模型网络得到参考轨迹的稳定控制, 进而将轨迹跟踪问题转化为误差系统的最优调节问题. 然后, 设计一个合适的事件触发条件, 并证明了基于事件的误差系统是渐近稳定的. 最后, 通过两个仿真实例验证了所提算法的可行性和有效性. 目前的研究主要是在理论方向, 将该方法扩展到实际应用场景是未来的工作, 包括基于所提跟踪算法控制污水处理过程中溶解氧和硝态氮的质量浓度.

References

1 Li C D, Yi J Q, Lv Y S, Duan P Y. A hybrid learning method for the data-driven design of linguistic dynamic systems. *IEEE/CAA Journal of Automatica Sinica*, 2019, **6**(6): 1487−1498

2 Basar T, Bernhard P. H_∞ optimal control and related minimax design problems: A dynamic game approach. *IEEE Transactions on Automatic Control*, 1996, **41**(9): 1397−1399

3 Dong J X, Hou Q H, Ren M M. Control synthesis for discrete-time T-S fuzzy systems based on membership function-dependent H_∞ performance. *IEEE Transactions on Fuzzy Systems*, 2020, **28**(12): 3360−3366

4 Qian D W, Li C D, Lee S G, Ma C. Robust formation maneuvers through sliding mode for multi-agent systems with uncertainties. *IEEE/CAA Journal of Automatica Sinica*, 2018, **5**(1): 342−351

5 Mathiyalagan K, Su H Y, Shi P, Sakthivel R. Exponential H_∞ filtering for discrete-time switched neural networks with random delays. *IEEE Transactions on Cybernetics*, 2015, **45**(4): 676−687

6 Werbos P J. Approximate dynamic programming for real-time control and neural modeling. In: Proceedings of the Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches. New York, USA: 1992.

7 Heydari A. Stability analysis of optimal adaptive control under value iteration using a stabilizing initial policy. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, **29**(9): 4522−4527

8 Al-Tamimi A, Lewis F L, Abu-Khalaf M. Discrete-time nonlinear HJB solution using approximate dynamic programming: Convergence proof. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2008, **38**(4): 943−949

9 Liu D R, Wei Q L. Generalized policy iteration adaptive dynamic programming for discrete-time nonlinear systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2015, **45**(12): 1577−1591

10 Guo W T, Si J N, Liu F, Mei S W. Policy approximation in policy iteration approximate dynamic programming for discrete-time nonlinear systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, **29**(7): 2794−2807

11 Modares H, Lewis F L, Naghibi-Sistani M B. Adaptive optimal control of unknown constrained-input systems using policy iteration and neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2013, **24**(10): 1513−1525

12 Kiumarsi B, Lewis F L. Actor-critic-based optimal tracking for partially unknown nonlinear discrete-time systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, **26**(1): 140−151

13 Wang Ding. Research progress on learning-based robust adaptive critic control. *Acta Automatica Sinica*, 2019, **45**(6): 1031−1043
(王鼎. 基于学习的鲁棒自适应评判控制研究进展. 自动化学报, 2019, **45**(6): 1031−1043)

14 Zhao F Y, Gao W N, Liu T F, Jiang Z P. Adaptive optimal output regulation of linear discrete-time systems based on event-triggered output-feedback. *Automatica*, 2022, **137**: 10103

15 Wang D, Qiao J F, Cheng L. An approximate neuro-optimal solution of discounted guaranteed cost control design. *IEEE Transactions on Cybernetics*, 2022, **52**(1): 77−86

16 Niu B, Liu J D, Wang D, Zhao X D, Wang H Q. Adaptive decentralized asymptotic tracking control for large-scale nonlinear systems with unknown strong interconnections. *IEEE/CAA Journal of Automatica Sinica*, 2022, **9**(1): 173−186

17 Wang D, Ha M M, Zhao M M. The intelligent critic framework for advanced optimal control. *Artificial Intelligence Review*, 2022, **55**(1): 1−22

18 Zhang H G, Wei Q L, Luo Y H. A novel infinite-time optimal tracking control scheme for a class of discrete-time nonlinear systems via the greedy HDP iteration algorithm. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, 2008, **38**(4): 937−942

19 Li C, Ding J L, Lewis F L, Chai T Y. A novel adaptive dynamic programming based on tracking error for nonlinear discrete-time systems. *Automatica*, 2021, **129**: 109687

20 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*, 2021. DOI: 10.1109/TNNLS.2021.3135405

21 Wang Ding, Zhao Ming-Ming, Ha Ming-Ming, Qiao Jun-Fei. Intelligent optimal tracking with application verifications via discounted generalized value iteration. *Acta Automatica Sinica*, 2022, **48**(1): 182−193
(王鼎, 赵明明, 哈明鸣, 乔俊飞. 基于折扣广义值迭代的智能最优跟踪及应用验证. 自动化学报, 2022, **48**(1): 182−193)

22 Postoyan R, Tabuada P, Nesic D, Anta A. A framework for the event-triggered stabilization of nonlinear systems. *IEEE Transactions on Automatic Control*, 2015, **60**(4): 982−996

23 Ha M M, Wang D, Liu D R. Event-triggered constrained control with DHP implementation for nonaffine discrete-time systems. *Information Sciences*, 2020, **519**: 110−123

24 Sahoo A, Xu H, Jagannathan S. Near optimal event-triggered control of nonlinear discrete-time systems using neurodynamic programming. *IEEE Transactions on Neural Networks and Learning Systems*, 2016, **27**(9): 1801−1815

25 Wang D, Ha M M, Qiao J F. Self-learning optimal regulation for discrete-time nonlinear systems under event-driven formulation. *IEEE Transactions on Automatic Control*, 2020, **65**(3): 1272−1279

26 Dong L, Zhong X N, Sun C Y, He H B. Adaptive event-triggered control based on heuristic dynamic programming for nonlinear discrete-time systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, **28**(7): 1594−1605

27 Dong L, Zhong X N, Sun C Y, He H B. Event-triggered adaptive dynamic programming for continuous-time systems with control constraints. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, **28**(8): 1941−1952

28 Zhang H G, Luo Y H, Liu D R. Neural-network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraints. *IEEE Transactions on Neural Networks*, 2009, **20**(9): 1490−1503

29 Jiang Z P, Wang Y. Input-to-state stability for discrete-time nonlinear systems. *Automatica*, 2001, **37**(6): 857−869

30 Zhang Y W, Zhao B, Liu D R. Deterministic policy gradient adaptive dynamic programming for model-free optimal control. *Neurocomputing*, 2020, **387**: 40−50

王 鼎 北京工业大学信息学部教授. 2009 年获得东北大学硕士学位, 2012 年获得中国科学院自动化研究所博士学位. 主要研究方向为强化学习与智能控制. 本文通信作者.
E-mail: dingwang@bjut.edu.cn
(**WANG Ding** Professor at the Faculty of Information Technology, Beijing University of Technology. He received his master degree from Northeastern University in 2009 and received his Ph.D. degree from Institute of Automation, Chinese Academy of Sciences in 2012. His research interest covers reinforcement learning and intelligent control. Corresponding author of this paper.)

胡凌治 北京工业大学信息学部硕士研究生. 主要研究方向为强化学习和智能控制. E-mail: hulingzhi@emails.bjut.edu.cn
(**HU Ling-Zhi** Master student at the Faculty of Information Technology, Beijing University of Technology. His research interest covers reinforcement learning and intelligent control.)

赵明明 北京工业大学信息学部博士研究生. 主要研究方向为强化学习和智能控制.
E-mail: zhaomm@emails.bjut.edu.cn
(**ZHAO Ming-Ming** Ph.D. candidate at the Faculty of Information Technology, Beijing University of Technology. His research interest covers reinforcement learning and intelligent control.)

哈明鸣 北京科技大学自动化与电气工程学院博士研究生. 分别于 2016 年和 2019 年获得北京科技大学学士和硕士学位. 主要研究方向为最优控制, 自适应动态规划和强化学习.
E-mail: hamingming_0705@foxmail.com
(**HA Ming-Ming** Ph.D. candidate at the School of Automation and Electrical Engineering, University of Science and Technology Beijing. He received his bachelor and master degrees from University of Science and Technology Beijing in 2016 and 2019, respectively. His research interest covers optimal control, adaptive dynamic programming, and reinforcement learning.)

乔俊飞 北京工业大学信息学部教授. 主要研究方向为污水处理过程智能控制和神经网络结构设计与优化.
E-mail: adqiao@bjut.edu.cn
(**QIAO Jun-Fei** Professor at the Faculty of Information Technology, Beijing University of Technology. His research interest covers intelligent control of wastewater treatment processes, structure design and optimization of neural networks.)