

基于强化学习的固体氧化物燃料电池 输出电压自抗扰控制研究

管超骏¹, 雷正玲^{1*}, 霍海波¹, 王芳¹, 姚国全², 刘涛³

(1. 上海海洋大学工程学院, 上海市 浦东新区 201306;

2. 高性能舰船技术教育部重点实验室(武汉理工大学), 湖北省 武汉市 430063;

3. 上海海事大学交通运输学院, 上海市 浦东新区 201306)

Active Disturbance Rejection Control of Output Voltage of Solid Oxide Fuel Cell Based on Reinforcement Learning

GUAN Chaojun¹, LEI Zhengling^{1*}, HUO Haibo¹, WANG Fang¹, YAO Guoquan², LIU Tao³

(1. College of Engineering Science and Technology, Shanghai Ocean University, Pudong New District, Shanghai 201306, China;

2. Key Laboratory of High Performance Ship Technology of the Ministry of Education (Wuhan University of Technology),

Wuhan 430063, Hubei Province, China; 3. College of Transport and Communications, Shanghai Maritime University,

Pudong New District, Shanghai 201306, China)

摘要: 【目的】为提升固体氧化物燃料电池(solid oxide fuel cell, SOFC)系统性能及寿命, 以100 kW SOFC系统为研究对象, 探究在保证输出电压跟踪性能的同时, 通过强化学习不断调整控制器系数以实现最佳的综合性能。

【方法】建立基于机理的SOFC输出电压系统模型, 采用改进型的非线性自抗扰控制器(nonlinear active disturbance rejection control, NLADRC), 通过控制输入燃气流量, 使输出电压很好地跟踪参考值。考虑到传统的单通道控制器无法同时满足多个目标, 但若采用双通道控制器则会导致系统复杂性、成本和故障风险增加, 提出一种基于双延迟深度确定性策略梯度(twin delayed deep deterministic policy gradient, TD3)的改进型非线性自抗扰控制器, 对非线性误差反馈控制律系数进行实时调节和优化。【结果】所设计控制器可在不违反燃料利用约束的情况下提高SOFC输出电压跟踪性能。【结论】所设计控制器具备适应性强、稳定性高和能克服不确定性等优点, 为实际SOFC系统的输出电压控制器设计提供理论参考。

关键词: 固体氧化物燃料电池(SOFC); 双延迟深度确定性策略梯度(TD3); 非线性自抗扰控制器(NLADRC); 燃料利用率; 非线性误差反馈控制律; 输出电压跟踪; 不确定性

ABSTRACT: [Objectives] In order to improve the performance and lifetime of solid oxide fuel cell (SOFC) systems, the 100 kW SOFC system was taken as the research object. The continuous adjustment of the controller coefficients was explored through reinforcement learning to realize the best comprehensive performance, while ensuring the output voltage tracking performance. [Methods] A mechanism-based SOFC output voltage system model was established, an improved nonlinear active disturbance rejection controller (NLADRC) was used to make the output voltage track the reference value well by controlling the input gas flow. Conventional single-channel controllers can only satisfy one objective at a time, and dual-channel controllers will increase system complexity, cost and risk of failure. An improved NLADRC controller based on the twin delayed deep deterministic policy gradient (TD3) was proposed to optimize the coefficients of nonlinear error feedback control law. [Results] The designed controller can improve SOFC output voltage tracking performance without violating fuel utilization constraints. [Conclusions] The designed controller has the advantages of strong adaptability, high stability, and the ability to overcome uncertainty, providing theoretical reference for designing output voltage controllers

基金项目: 国家自然科学基金项目(52301420, 52071201); 高性能舰船技术教育部重点实验室开放基金课题项目(GXNC23052801); 上海市地方院校能力建设计划项目(23010502200)。

Project Supported by National Natural Science Foundation of China (52301420, 52071201); Open Fund of Key Laboratory of High Performance Ship Technology of the Ministry of Education (GXNC23052801); Capacity Building Project of Local Universities in Shanghai (23010502200).

in practical SOFC systems.

KEY WORDS: solid oxide fuel cell (SOFC); twin delayed deep deterministic policy gradient (TD3); nonlinear active disturbance rejection control (NLADRC); fuel utilization; nonlinear error feedback control law; output voltage tracking; uncertainty

0 引言

固体氧化物燃料电池(solid oxide fuel cell, SOFC)是一种全固态能源转换装置,可直接将燃料气体和氧化剂气体所携带的化学能转化为电能。SOFC无需依赖昂贵的催化剂(如铂)即可运行,从而降低了成本。此外,由于其余热等级高,可以与燃气轮机、蒸汽轮机等集成到各种联合循环系统中,实现70%以上的综合利用效率^[1-4]。因此,SOFC适合大规模应用,如中央发电站和热电联产系统。然而,这种类型的燃料电池也有缺点,如动态响应慢、非线性和严格的操作限制等问题^[5-8]。

当负载电流发生变化时,SOFC应自动调整输出电压,同时遵守燃料利用率阈值。在理想条件下,SOFC的燃料利用率应介于0.7~0.9^[9]。低于或超出这个范围会导致燃料使用不足或过量。然而,大多数控制器设计都没有考虑到这些限制因素,这可能会导致SOFC性能下降,降低燃料电池寿命,甚至在燃料控制过程中破坏运行稳定性^[10]。由此可见,SOFC控制的目的是通过调整燃料流量,使输出电压更平稳地接近参考值,同时将燃料利用率维持在理想范围内。

文献[11]提出SOFC的神经网络模型,采用神经模糊控制方法实现输出电压的精确控制,但因其采用单通道控制策略,使得燃料利用率约束未能满足。文献[12]采用改进萤火虫算法优化分数阶比例积分微分(proportional integral differential, PID)参数,算法优化后,PID控制器的抗干扰能力增强,提高了SOFC的运行可靠性,但其同样无法实现多约束目标。文献[13]采用一种由鲁棒控制器和PID控制器组成的多回路控制器,这种方法能实现在燃料利用约束前提下的稳定输出电压,但增加了系统的复杂性。文献[14]提出一种

新的智能比例积分自适应滑动模态控制器(intelligent proportional integral adaptive sliding mode control, IPI-ASMC),并设计抗饱和补偿器,以处理不确定性和执行器饱和问题,但抗饱和控制器的存在限制了系统的动态性能。文献[15-16]采用了一种多智能体的控制方式,实现了电压的精确跟踪,满足了燃料利用率要求,但其控制结构复杂,存在优化失败的风险。

传统的单通道控制算法无法防止违反燃料利用率约束,导致SOFC性能显著下降和不稳定。然而,采用双通道控制器会增加系统复杂性、成本和故障风险。而对于纯粹的机器学习方法,其控制过程中存在的不确定性会导致优化失败的风险,使其在实际应用中面临挑战。

因此,迫切需要一种无模型最优控制算法,该算法不仅能够根据SOFC的状态调节燃料流量,同时还能够满足燃料利用约束和输出电压跟踪性能。自抗扰控制(active disturbance rejection control, ADRC)具有响应快、超调小、鲁棒性强、不依赖被控对象模型的特点^[17]。而深度强化学习中的双延迟深度确定性策略梯度(twin delayed deep deterministic policy gradient algorithm, TD3)算法可以在不需要识别系统内部结构的情况下以数据驱动的方式进行控制或优化,还可以在奖励函数的设计基础上兼顾考虑系统性能和操作约束^[18]。

本文结合TD3和ADRC的优点,提出了一种新型单通道控制算法,它可通过对当前SOFC运行状态的观测,实时调整改进型非线性ADRC控制器中非线性误差反馈控制律的增益,改变天然气流量,可在满足约束条件的同时提高控制性能,并在奖励函数的精心设计下使得优化算法迅速收敛。

1 SOFC机理模型的建立

SOFC系统的输出电压受到多种因素影响,若使用神经网络或数据驱动的方式建模,存在物理上可解释性较差、训练数据较大、难以优化模型等问题,故本文采用了基于机理的方法进行建模,模型结构如图1所示。该模型考虑了电化学和热力学原理^[19]。图1中,系统的操作变量是天

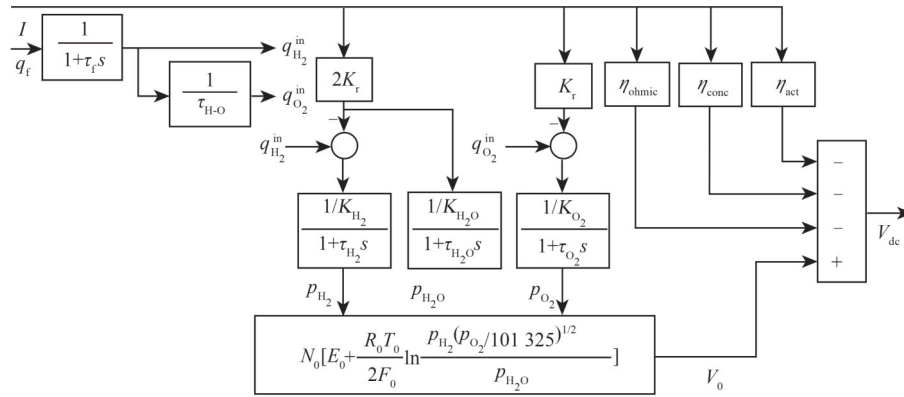


图1 固体氧化物燃料电池模型结构

Fig. 1 Solid oxide fuel cell model structure

然气流量 q_f ，系统的受控变量是输出电压 V_{dc} ，外部负载电流 I 是干扰变量，其他详细参数含义见表 1^[20]。

表1 SOFC 系统参数取值

Tab. 1 Values for SOFC system parameters

参数	数值
绝对温度 T_0/K	1 273
法拉第常数 $F_0/(C/mol)$	96 485
通用气体常数 $R_0/[J/(mol \cdot K)]$	8.314
理想标准电位 E_0/V	1.18
电池数 N_0	384
常数 $K_r/[mol/(s \cdot A)]$	0.996×10^{-6}
氢气的阀摩尔常数 $K_{H_2}/[mol/(s \cdot Pa)]$	8.32×10^{-6}
水蒸气的阀摩尔常数 $K_{H_2O}/[mol/(s \cdot Pa)]$	2.77×10^{-6}
氧气的阀摩尔常数 $K_{O_2}/[mol/(s \cdot Pa)]$	2.49×10^{-6}
氢气流响应时间 τ_{H_2}/s	26.1
水的响应时间 τ_{H_2O}/s	78.3
氧气流响应时间 τ_{O_2}/s	2.91
燃料重整时间常数 τ_f/s	5
氢氧比 τ_{H_2O}	1.145

SOFC 输出电压 V_{dc} 可由下式计算：

$$V_{dc} = V_0 - \eta_{ohmic} - \eta_{conc} - \eta_{act} \quad (1)$$

式中： η_{ohmic} 为欧姆极化电压损耗； η_{act} 为活化极化电压损耗； η_{conc} 为浓度极化电压损耗； V_0 为驱动电子在燃料电池中传输的电动势。

V_0 的值可由 Nerst 方程来计算：

$$V_0 = N_0 \left[E_0 + \frac{R_0 T_0}{2F_0} \ln \frac{p_{H_2} \sqrt{p_{O_2}/101325}}{p_{H_2O}} \right] \quad (2)$$

式中 p_{H_2} 、 p_{H_2O} 和 p_{O_2} 分别为氢气、水蒸气和氧气的分压，其表达式分别为：

$$p_{H_2} = \frac{1}{K_{H_2}(1 + \tau_{H_2}s)} \left(\frac{1}{1 + \tau_f s} q_f - 2K_r I \right) \quad (3)$$

$$p_{O_2} = \frac{1}{K_{O_2}(1 + \tau_{O_2}s)} \left(\frac{1/\tau_{H_2O}}{1 + \tau_f s} q_f - K_r I \right) \quad (4)$$

$$p_{H_2O} = \frac{2}{K_{H_2O}(1 + \tau_{H_2O}s)} K_r I \quad (5)$$

式中 $K_r = N_0/(4F_0)$ 。

当 SOFC 工作时会产生电压损耗，其中最为主要的是欧姆极化电压损耗 η_{ohmic} 、活化极化电压损耗 η_{act} 以及浓度极化电压损耗 η_{conc} ^[21]，分别表示如下：

$$\eta_{ohmic} = I r_{ohmic} \quad (6)$$

$$\eta_{act} = \alpha + \beta \ln I \quad (7)$$

$$\eta_{conc} = -\frac{R_0 T_0}{2F_0} \ln \left(1 - \frac{I}{I_L} \right) \quad (8)$$

式中： r_{ohmic} 为欧姆损耗； α 为塔菲尔常数； β 为塔菲尔斜率； I_L 为极限电流密度。

燃料利用率 u_f 是非常重要的一个变量，定义为燃料电池中实际产生的化学能(电能)与输入燃料化学能之间的比例，其表达式为

$$u_f = \frac{q_{H_2}^r}{q_{H_2}^{in}} = \frac{2K_r I}{q_{H_2}^{in}} \quad (9)$$

式中： $q_{H_2}^{in}$ 为 q_f 经过燃料重整器后的氢气流量； $q_{H_2}^r$ 为氢气流量的反应量， $q_{H_2}^r = q_{H_2}^{in} - q_{H_2}^{out}$ ， $q_{H_2}^{out}$ 为未反应的氢气流量。

燃料利用率的理想范围为 [0.7, 0.9]^[20]。超出该范围的燃料利用将降低 SOFC 的性能或对其造成永久性损害。

当负载电流存在扰动时，为了使输出电压迅

速达到设定条件，燃料利用率将急剧变化。而过于保守的控制策略又会使得动态过程缓慢，影响微电网稳定。

故控制器设计中的控制目标可总结为：考虑负载电流变化对 SOFC 输出电压的影响，调节燃料流量以精确追踪输出电压，实现更好的控制性能；确保 SOFC 系统的燃料利用率保持在理想范围内；保证 SOFC 输出电压的动态响应性能。

2 自抗扰控制器

2.1 二阶非线性 ADRC

ADRC 的基本思想是将影响被控对象的各种不确定因素视为未知扰动，并对其实时估计和补偿^[22]。相较于线性 ADRC，非线性 ADRC 具有更好的跟踪精度和响应速度。本文将二阶非线性 ADRC 控制器应用到固体氧化物燃料电池输出电压控制系统中，其系统控制结构框图如图 2 所示。

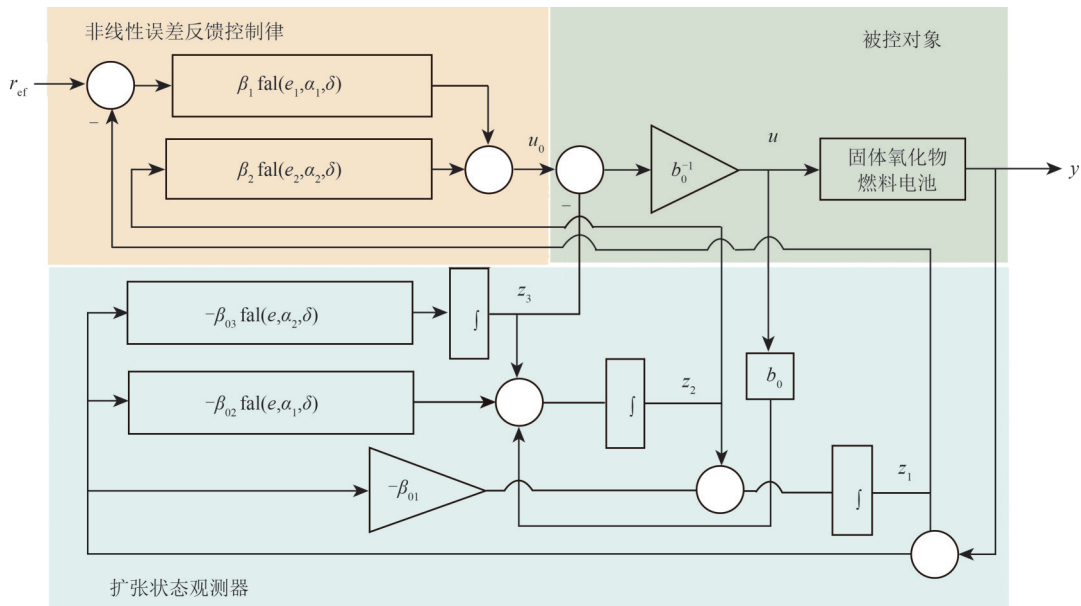


图2 二阶非线性自抗扰控制器结构

Fig. 2 Structure of second-order nonlinear active disturbance rejection control

图 2 中： r_{ref} 为设定值； b_0 为输入增益 b 的估计值； u 为控制量(天然气的流量)； y 为系统的输出值； $fal(\cdot)$ 为非线性函数； z_1 、 z_2 和 z_3 为状态变量的估计值； β_{01} 、 β_{02} 和 β_{03} 为增益系数。

ADRC 方法并不依赖于特定的模型，其主要结构由 3 部分组成：跟踪-微分器(tracking differentiator, TD)、非线性状态误差反馈控制律(nonlinear state error feedback, NLSEF) 和非线性扩张状态观测器(nonlinear extended state observer, NLESO)。首先应用跟踪-微分器对输入的参考电压进行跟踪和滤波，然后应用非线性扩张状态观测器对 SOFC 系统的不确定性、未建模动态、外部扰动和控制系数的不确定性进行观测，最后将两者的输出信号输送至非线性误差反馈控制律中，经运算后得到天然气流量信号 q_f 。

2.2 跟踪-微分器

跟踪-微分器的主要作用是快速、无超调地跟踪期望电压控制信号，获得其差分信号，并对控制信号滤波，以降低噪声、提高稳定性。其跟踪算法表示如下：

$$\begin{cases} y_e(k+1) = y_e(k) + T_d \cdot \dot{y}_e(k) \\ \dot{y}_e(k+1) = \dot{y}_e(k) + T_d \cdot fhan(\cdot) \end{cases} \quad (10)$$

式中： y_e 为期望跟踪电压； \dot{y}_e 为期望跟踪电压的导数； T_d 为采样周期； $fhan(\cdot)$ 为最速综合函数，表示为

$$fhan(\cdot) = \begin{cases} -r_f a/d, & |a| \leq d \\ -r_f \cdot \text{sign}(a), & |a| > d \end{cases} \quad (11)$$

其中：

$$a = \begin{cases} \dot{y}_e + y_0/h, & |a| \leq d \\ \dot{y}_e + 0.5(a_0 - d)\text{sign}(y_0), & |a| > d \end{cases} \quad (12)$$

$$\begin{cases} d=r_f h \\ d_0=dh \\ y_0=y_e+h\dot{y}_e \\ a_0=\sqrt{d^2+8r_f|y_0|} \end{cases} \quad (13)$$

式中： r_f 为调整跟踪速度的速度因子； d 为线性区间长度； h 为采样步长， s 。

2.3 非线性误差反馈控制律

非线性误差反馈控制律为非线性的比例-微分(proportional differential, PD)组合，其比例和微分控制参数可以在既定的变动范围内进行非线性组合，从而达到更高的控制要求。本文所采用的非线性函数命名为 $\text{nfal}(\cdot)$ 函数，与传统的 $\text{fal}(\cdot)$ 函数相比，能很好地解决原函数分段点导数不连续以及可能发生的高频颤振问题，其表达式为：

$$\text{nfal}(\cdot)=\begin{cases} \frac{3\delta^\alpha-\alpha\delta^\alpha}{2\delta}e+\frac{\delta^\alpha-\alpha\delta^\alpha}{2\delta^3}e^3, & |e|\leq\delta \\ \text{sign}(e)|e|^{\alpha_i}, & |e|>\delta \end{cases} \quad (14)$$

式中： α_i 为非线性段控制系数； δ 为线性区间长度； e 为跟踪误差。

非线性误差反馈控制算法为

$$u_0=\beta_1\cdot\text{nfal}(e_1,\alpha_1,\delta)+\beta_2\cdot\text{nfal}(e_2,\alpha_2,\delta) \quad (15)$$

式中： β_1 为比例系数； β_2 为微分系数； e_1 、 e_2 分别为 z_1 、 z_2 的跟踪误差； α_1 、 α_2 为非线性段控制系数。

2.4 非线性扩张状态观测器

ADRC 控制器的核心是非线性扩张状态观测器，系统中的负载电流扰动、内部温度扰动、气体流量变化等外部状态变量被统一作为总扰动进行实时估计，并将结果引入非线性误差反馈控制律中实现在线补偿。非线性扩张状态观测器的主

要算法表示如下：

$$\begin{cases} e=z_1(k)-y(k) \\ z_1(k+1)=z_1(k)+h[z_2(k)-\beta_{01}e] \\ z_2(k+1)=z_2(k)+h[z_3(k)-\beta_{02}\cdot\text{nfal}(e,\alpha_1,\delta)+bu] \\ z_3(k+1)=z_3(k)-h\beta_{03}\cdot\text{nfal}(e,\alpha_2,\delta) \end{cases} \quad (16)$$

此时，若控制量设计为

$$u(k)=\frac{u_0(k)-z_3(k)}{b_0} \quad (17)$$

那么当 z_3 的估计值与总扰动 f 误差足够小时，则系统的传递函数可以变为

$$\ddot{y}=f+bu=f+u_0-z_3\approx u_0 \quad (18)$$

此时被控对象被转换成一个积分串联型对象，仅需利用非线性误差反馈控制律形成反馈，即可很好地满足系统的稳定性、准确性和快速性。

3 基于强化学习的自抗扰控制

3.1 双延迟深度确定性策略梯度算法

由于传统的单通道控制器无法同时保证电压跟踪精度和燃料利用率约束条件，本文采用 ADRC 与深度强化学习相结合，通过训练 NLSEF 中的 β_1 和 β_2 ，不断地改变天然气流量。

本文所选用的 TD3 算法^[23]源自演员-评论家(actor-critic, AC)算法，主要用于解决连续控制问题的离线异策式深度强化学习方法，其整体网络结构^[24-25]如图 3 所示。TD3 算法分为 Actor 策略网络、Critic 评价网络以及 TD-error 三部分。Actor 策略网络以 SOFC 的状态作为输入并根据系统参数输出动作。Critic 评价网络接收当前状态和 Actor 策略网络输出的动作值作为输入，用于评价 Actor 策略网络输出动作值的优劣。

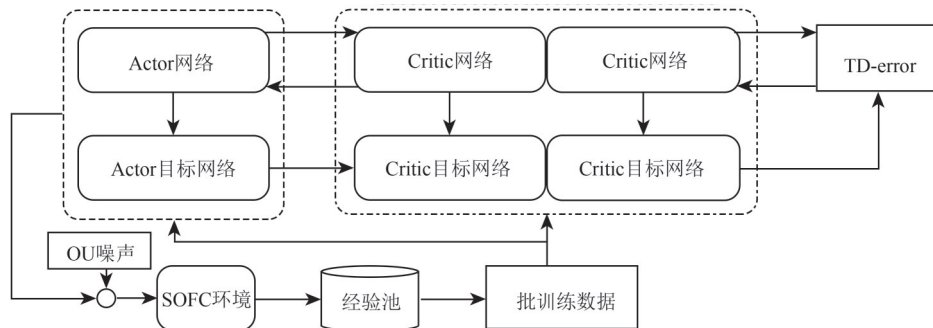


图 3 TD3 算法结构

Fig. 3 TD3 algorithm structure

TD3算法共包含6个网络：Actor网络、Actor目标策略网络、2个独立的Critic网络以及2个独立的Critic目标网络。在 Q 值估计过程中选取较小的 Q 值，从而抑制过估问题。

Actor网络通过确定性策略梯度算法进行更新：

$$\nabla_{\phi} J(\phi) = E_{s \sim p_{\pi}} \left[\nabla_a Q^{\pi}(s, a) \Big|_{a=\pi(s)} \nabla_{\phi} \pi_{\phi}(s) \right] \quad (19)$$

$$Q^{\pi}(s, a) = E_{s \sim p_{\pi}, a \sim \pi} [R_t | s, a] \quad (20)$$

式中： ∇ 为权重参数的梯度； $J(\phi)$ 为策略 π_{ϕ} 的性能度量； $E_{s \sim p_{\pi}}$ 为状态 s 下采用策略 π 所作出动作 a 时的预期收益； $Q^{\pi}(s, a)$ 为价值网络的状态动作价值； R_t 为 t 时刻的收益。

Critic网络和Critic目标策略网络的评估函数基于贝尔曼方程进行延迟学习：

$$Q_{\pi}(s, a) = r + \gamma E_{s', a'} [Q_{\pi}(s', a')], a' \sim \pi(s') \quad (21)$$

式中： r 为收益； γ 为折扣因子； (s', a') 为下一时刻的状态和动作。

为缓解确定性策略中出现的估值函数过拟合问题，需在目标策略网络中加入高斯随机噪声：

$$y = r + \gamma \min(Q_{\theta_1}, Q_{\theta_2}) [s', \pi_{\phi}(s') + \varepsilon] \quad (22)$$

$$\varepsilon \in \text{clip}[N(0, \sigma), -c, c], c > 0 \quad (23)$$

式中： ε 为高斯噪声； c 为策略平滑噪声的截断边界值； σ 为标准差； clip 函数为截断函数； Q_{θ_1} 、 Q_{θ_2} 分别为 θ_1 、 θ_2 的价值函数。

为了保证在策略更新前将估计误差降低，策略网络的更新频率应该低于评估网络，所以采用一种软更新策略：

$$\theta_i = \tau \theta_i + (1 - \tau) \theta'_i \quad (24)$$

$$\phi = \tau \phi + (1 - \tau) \phi' \quad (25)$$

式中： τ 为软更新率； θ_i 为上一个价值函数的参数； θ'_i 为更新后价值函数的参数； ϕ 为上一个策略函数的参数； ϕ' 为更新后策略函数的参数。

3.2 ADRC-TD3控制器设计

基于双延迟深度确定性策略梯度算法的自抗扰控制器 (ctive disturbance rejection control-twin delayed deep deterministic policy gradient algorithm, ADRC-TD3)总体控制策略框架如图4所示。控制器的系数调节间隔为0.01 s，控制目标是通过调节燃料流量稳定参考输出电压，同时将SOFC的燃料利用率保持在合理的范围内。

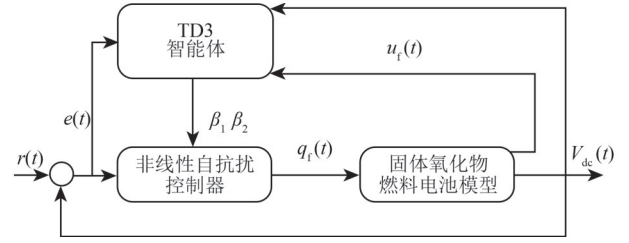


图4 ADRC-TD3总体控制策略框架

Fig. 4 Structure framework of ADRC-TD3 overall control

本文所设计ADRC-TD3控制器通过调整 β_1 和 β_2 系数来控制SOFC的输出电压。而动作空间、状态空间和奖励函数的选取作为任务的重要组成部分，会直接影响TD3算法的学习和决策过程，需要精心设计。

3.2.1 动作空间

动作空间的选取为：

$$\begin{cases} \mathbf{a} = [\beta_1/10 \quad \beta_2/10]^T \\ -10 \leq \beta_1 \leq 10 \\ -10 \leq \beta_2 \leq 10 \end{cases} \quad (26)$$

3.2.2 状态空间

状态空间的选取为

$$[e(t) \int_0^t e(t) dt \quad V_{dc}(t) \quad u_f(t) \quad \dot{u}_f(t)]^T \quad (27)$$

式中： $e(t)$ 为跟踪误差； $V_{dc}(t)$ 为系统输出电压。将燃料利用率 $u_f(t)$ 及其导数 $\dot{u}_f(t)$ 引入状态空间，可以让智能体更好地将燃料利用率控制在理想范围内。

3.2.3 奖励函数设计

本文对奖励函数的设计主要考虑跟踪误差、燃料利用率是否超出理想范围、天然气流量的波动及停止训练轮次条件。其中：

$$r_1 = \begin{cases} -\mu_1 |e(t)| + 1, & |e(t)| \leq 0.1 \\ -\mu_2 \cdot \tanh(|e(t)| - 0.1), & |e(t)| > 0.1 \end{cases} \quad (28)$$

$$r_2 = \begin{cases} 0.1, & 0.7 \leq u_f \leq 0.9 \\ -5, & u_f < 0.7 \text{ 或 } u_f > 0.9 \end{cases} \quad (29)$$

$$r_3 = -\mu_3 \sum_{i=0}^4 |u_f^i - u_f^{\text{avg}}| \quad (30)$$

$$u_f^{\text{avg}} = \frac{1}{5} \sum_{i=0}^4 u_f^i \quad (31)$$

$$r = r_1 + r_2 + r_3 - \chi \quad (32)$$

式中： μ_1 、 μ_2 及 μ_3 为惩罚系数； χ 为超出SOFC运行条件时的惩罚； r_1 为跟踪误差惩罚项； r_2 为燃料利用率超出惩罚项； r_3 为天然气流量波动惩罚项。

r_1 的设定可以使误差产生时, 系统能够快速响应以消除误差。而当系统响应时间加大时会导致 r_1 的累计惩罚增大, 从而引导智能体不断提升系统的响应速度。 r_2 的设定使得智能体在满足燃料利用率约束时获得一个较小的奖励, 而超出范围时能够快速响应。 r_3 使得智能体能够对天然气流量的波动快速响应以减小执行器负担。

4 SOFC 仿真实验

4.1 仿真参数

为验证本文所设计控制器的性能, 将经精手动调参的ADRC切换型控制器作为基准策略, 针对ADRC-TD3, 开展输出电压跟踪性能和燃料利用率约束性能的对比仿真研究。

ADRC-TD3 算法主要超参数见表2。

表2 ADRC-TD3 算法主要超参数

Tab. 2 Main hyperparameters of ADRC-TD3 algorithm

参数	取值
折扣因子	0.99
经验池大小	2×10^6
Actor 网络学习率	0.001
Critic 网络学习率	0.000 1
样本学习个数	128
训练步数	4 000
软更新系数	0.005
延迟更新参数	2

4.2 预训练

本节预训练数据的来源为SOFC输出电压模型在不同的输入下所产生的不同状态。为了确保所获得的样本具有足够的随机性和多样性, 在每轮训练开始时的给定跟踪电压会有 ± 1 V间的随机波动, 并在每次训练过程中随机加入一次随机大小的给定跟踪电压信号和一次随机大小的负载电流阶跃扰动信号, 每轮训练40 s。实际的训练结果如图5所示。

可以看到, 在前100轮次训练中, 智能体通过不断尝试使得平均奖励不断上升; 当训练轮次达到150次时不再出现训练失败的情况; 当训练轮次达到400时, 平均奖励基本收敛, 达到停止条件。经过训练后的智能体能够根据环境的改变自适应地调整自抗扰控制器中 β_1 和 β_2 , 实现燃料

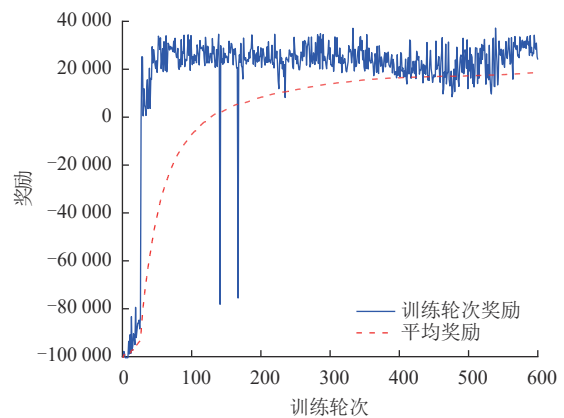


图5 ADRC-TD3算法调节过程奖励

Fig. 5 ADRC-TD3 algorithm adjustment process reward

利用约束前提下的稳定输出电压。

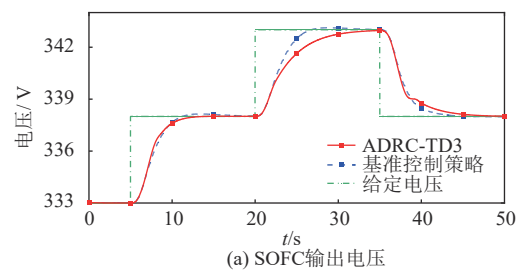
4.3 给定电压信号跟踪实验

基准策略ADRC切换型控制器的手调参数为 $\beta_1=1.1$, $\beta_2=2$, 且将基准策略和ADRC-TD3中非线性扩张状态观测器增益系数均取为 $[\beta_{01}, \beta_{02}, \beta_{03}] = [100, 800, 300]$ 。

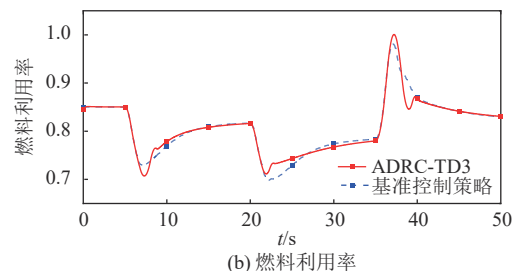
为验证两者的给定电压跟踪能力, 分别在 $t=5$ s、 $t=20$ s、 $t=35$ s加入3次给定电压阶跃变化:

$$V_{\text{ref}} = \begin{cases} 333, & t < 5 \\ 338, & 5 \leq t < 20 \\ 343, & 20 \leq t < 35 \\ 338, & t \geq 35 \end{cases} \quad (33)$$

给定电压信号跟踪实验仿真结果对比如图6所示。可见, 2种方法都能很好地跟踪给定电压, 并实现燃料利用率约束限制。不同的是, 采用ADRC-TD3方法时完全不会发生燃料利用率约束



(a) SOFC输出电压



(b) 燃料利用率

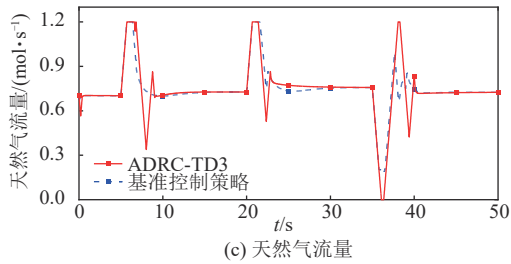


图6 给定电压信号跟踪实验仿真结果对比

Fig. 6 Comparison of simulation results of given voltage signal tracking experiments

违反现象, 而切换型 ADRC 控制策略由于先发生违反现象再进行切换控制, 所以存在违反时间, 且由于切换控制的存在, 基准策略对于电压的跟踪存在一定的超调。

4.4 负载电流扰动实验

为验证本文所设计控制器对于负载电流的抗扰性能, 本文设计在 $t=10\text{ s}$ 时发生一次幅值为 7 V 的给定电压阶跃变化, 在 $t=40, 60, 80\text{ s}$ 共发生 3 次大小为正负 40 A 的阶跃扰动。负载电流扰动实验仿真结果对比如图 7 所示。

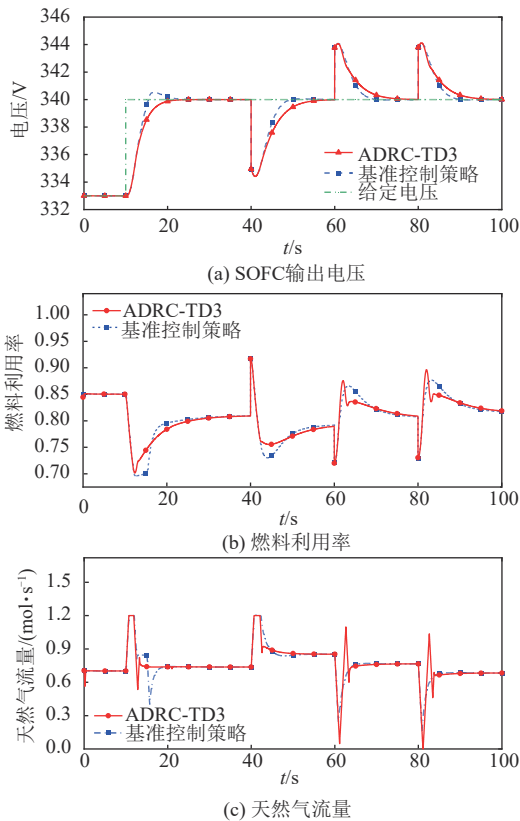


图7 负载电流扰动实验仿真结果对比

Fig. 7 Comparison of simulation results for load current disturbance experiments

由图 7 可见, 随着给定跟踪电压的增大, 控制过程中燃料利用率更容易超出理想范围, 导致基准策略在控制过程中发生 1 V 超调。而本文所设计 ADRC-TD3 控制器的控制律平滑, 能快速跟踪给定参考电压且不存在超调现象, 对于负载电流扰动的抑制也能很好实现, 对燃料利用率的控制优于 ADRC 切换型控制策略。

5 结论

针对传统双通道控制器结构复杂, 且单通道控制器难以满足多约束条件下控制的现状, 研究了一种基于双延迟深度确定性策略梯度的改进型非线性自抗扰控制器。通过双延迟深度确定性策略梯度算法对误差控制律系数的训练来满足对燃料利用率的约束以及输出电压的跟踪控制。通过仿真分析, 得出以下结论:

- 1) 与基准控制策略相比, 所提出的控制策略具有更好的综合性能。其动态响应过程无超调, 调节时间短, 天然气流量波动较小, 在负载电流扰动下能够以更快的速度精确跟踪给定电压。
- 2) 通过对状态空间和奖励函数的精心设计, 使得算法收敛迅速, 降低了训练所需的时间成本, 实现了单通道控制器满足多约束条件的控制目标。
- 3) 与基准控制策略相比, 所提出的控制策略需整定参数数量减少, 降低了人工调整参数的难度, 为实际 SOFC 系统的输出电压控制器设计提供了理论参考。

参考文献

- [1] YANG B, LI Y, LI J, et al. Comprehensive summary of solid oxide fuel cell control: a state-of-the-art review[J]. Protection and Control of Modern Power Systems, 2022, 7(3): 1-31.
- [2] NECHACHE A, HODY S. Alternative and innovative solid oxide electrolysis cell materials: a short review[J]. Renewable and Sustainable Energy Reviews, 2021, 149: 111322.
- [3] 郭心如, 郭雨旻, 罗方, 等. 磷酸燃料电池的能效、烟及生态特性分析[J]. 发电技术, 2022, 43(1): 73-82.

- GUO X R, GUO Y M, LUO F, et al. Analysis of energy, exergy and ecology characteristics of phosphoric acid fuel cell[J]. *Power Generation Technology*, 2022, 43(1): 73-82.
- [4] 和萍, 祁盼, 申润杰, 等. 计及风电和燃料电池的综合能源系统阻尼特性分析[J]. *电力科学与技术学报*, 2020, 35(1): 14-23.
- HE P, QI P, SHEN R J, et al. Analysis of damping characteristics of an integrated energy system with hybrid wind-fuel cells integrated[J]. *Journal of Electric Power Science and Technology*, 2020, 35(1): 14-23.
- [5] LAN T, STRUNZ K. Multiphysics transients modeling of solid oxide fuel cells: methodology of circuit equivalents and use in EMTP-type power system simulation[J]. *IEEE Transactions on Energy Conversion*, 2017, 32(4): 1309-1321.
- [6] BAO C, WANG Y, FENG D, et al. Macroscopic modeling of solid oxide fuel cell (SOFC) and model-based control of SOFC and gas turbine hybrid system[J]. *Progress in Energy and Combustion Science*, 2018, 66: 83-140.
- [7] 张瑞宇, 王雨晴, 任佳伟. 基于丙烷催化部分氧化的微管式固体氧化物燃料电池系统特性研究[J]. *发电技术*, 2024, 45(3): 486-493.
- ZHANG R Y, WANG Y Q, REN J W. Characteristics research of a micro-tubular solid oxide fuel cell system based on catalytic partial oxidation of propane[J]. *Power Generation Technology*, 2024, 45(3): 486-493.
- [8] 李林, 刘彤宇, 李爽, 等. 甲醇重整制氢燃料电池发电研究进展[J]. *发电技术*, 2022, 43(1): 44-53.
- LI L, LIU T Y, LI S, et al. Research progress of hydrogen production by methanol reforming for fuel cell power generation[J]. *Power Generation Technology*, 2022, 43(1): 44-53.
- [9] KOMATSU Y, BRUS G, KIMIJIMA S, et al. The effect of overpotentials on the transient response of the 300 W SOFC cell stack voltage[J]. *Applied Energy*, 2014, 115: 352-359.
- [10] SUN L, JIN Y, SHEN J, et al. Sustainable residential micro-cogeneration system based on a fuel cell using dynamic programming-based economic day-ahead scheduling[J]. *ACS Sustainable Chemistry & Engineering*, 2021, 9(8): 3258-3266.
- [11] 王倩如, 王彩霞, 顾吉鹏. 固体氧化物燃料电池的神经模糊控制策略研究[J]. *热能动力工程*, 2022, 37(10): 198-206.
- WANG Q R, WANG C X, GU J P. Study on neural fuzzy control strategy of solid oxide fuel cell[J]. *Journal of Engineering for Thermal Energy and Power*, 2022, 37(10): 198-206.
- [12] ZHANG T, LI H, TU X, et al. Optimization of SOC fractional PID control parameters for solid oxide battery based on improved firefly algorithm[C]//2021 3rd International Conference on Industrial Artificial Intelligence (IAI). Shenyang, China: IEEE, 2021: 1-5.
- [13] WU X, GAO D. Optimal robust control strategy of a solid oxide fuel cell system[J]. *Journal of Power Sources*, 2018, 374: 225-236.
- [14] ABBAKER A O, WANG H, TIAN Y. Voltage control of solid oxide fuel cell power plant based on intelligent proportional integral-adaptive sliding mode control with anti-windup compensator[J]. *Transactions of the Institute of Measurement and Control*, 2020, 42(1): 116-130.
- [15] LI J, YU T, YANG B. A data-driven output voltage control of solid oxide fuel cell using multi-agent deep reinforcement learning[J]. *Applied Energy*, 2021, 304: 117541.
- [16] LI J, YU T. A novel data-driven controller for solid oxide fuel cell via deep reinforcement learning[J]. *Journal of Cleaner Production*, 2021, 321: 128929.
- [17] HAN J. From PID to active disturbance rejection control[J]. *IEEE Transactions on Industrial Electronics*, 2009, 56(3): 900-906.
- [18] ZHOU J, XUE S, XUE Y, et al. A novel energy management strategy of hybrid electric vehicle via an improved TD3 deep reinforcement learning[J]. *Energy*, 2021, 224: 120118.
- [19] PADULLES J, AULT G W, MCDONALD J R. An integrated SOFC plant dynamic model for power systems simulation[J]. *Journal of Power sources*, 2000, 86(1/2): 495-500.
- [20] SUN L, HUA Q, SHEN J, et al. A combined voltage control strategy for fuel cell[J]. *Sustainability*, 2017, 9(9): 1517.
- [21] LI Y, SHEN J, LU J. Constrained model predictive control of a solid oxide fuel cell based on genetic optimization[J]. *Journal of Power Sources*, 2011, 196(14): 5873-5880.
- [22] ZHAO C, LI D. Control design for the SISO system with the unknown order and the unknown relative degree[J]. *ISA Transactions*, 2014, 53(4): 858-872.
- [23] LADOSZ P, WENG L, KIM M, et al. Exploration in deep reinforcement learning: a survey[J].

Information Fusion, 2022, 85: 1-22.

- [24] FUJIMOTO S, VAN HOOF H, MEGER D. Addressing function approximation error in actor-critic methods[J]. Proceedings of the 35th International Conference on Machine Learning, 2018, 80: 1587-1596.
- [25] NGUYEN T T, NGUYEN N D, NAHAVANDI S. Deep reinforcement learning for multiagent systems: a review of challenges, solutions, and applications[J]. IEEE Transactions on Cybernetics, 2020, 50(9): 3826-3839.

收稿日期: 2024-01-22。

修回日期: 2024-04-29。

作者简介:



管超骏

管超骏(1994), 男, 硕士研究生, 研究方向为燃料电池建模与跟踪控制, 17621922877@163.com;



雷正玲

雷正玲(1988), 女, 博士, 讲师, 研究方向为新能源动力系统建模与控制、基于深度学习方法的功率预测等, 本文通信作者, zllel@shou.edu.cn;



姚国全

姚国全(1987), 男, 硕士, 实验师, 研究方向为船舶水动力学、船舶性能试验技术、海洋智能装备建模与控制等, 604617856@qq.com;



刘涛

刘涛(1988), 男, 博士, 副教授, 研究方向为人工智能、智能交通、大数据等, dlmult@hotmail.com。

(责任编辑 杨阳)