弱模型依赖通用智能姿态控制技术

邵会兵,詹 韬,付京博
(北京控制与电子技术研究所,北京 100038)

摘 要: 超高速跨域飞行、敏捷机动等是新一代飞行器发展方向,而长时高速飞行产生的气动外形变化带来的气动参数大范围改变等问题,都对控制系统设计提出了更高的要求。为提高飞行器对模型不确定性的适应能力及控制方法对不同外形、复合执行机构的通用性,深入研究了弱模型依赖的通用智能姿态控制技术,分层次地开展了基于深度学习(DL)的自适应姿态控制、基于深度确定性策略梯度算法(DDPG)的通用姿态控制、弱模型依赖的多维复合控制等技术研究,显著提高了控制系统的鲁棒性和通用性,对人工智能技术在飞行器姿态控制中的应用具有一定的指导意义。

关键词: 弱模型依赖; 自适应智能控制; 多维复合控制; 深度强化学习(DRL); 扩张状态观测器

中图分类号: TJ 765.2 **文献标志码:** A **DOI:** 10.19328/j.cnki.2096-8655.2022.04.007

Generalized Intelligent Attitude Control with Weak Model Dependence

SHAO Huibing, ZHAN Tao, FU Jingbo
(Beijing Institute of Control and Electronic Technology, Beijing 100038, China)

Abstract: Ultra-high speedcross-domain flight and agile maneuvering are the developing trends of next-generation aircrafts. However, the aerodynamic parameters variations caused by the aerodynamic shape change in long-time wide-speed-range hypersonic flight and the aerodynamic variations in deformable aerial-underwater flight pose significant challenges to the aircraft attitude control system. In this paper, a novel generalized intelligent attitude control method with weak model dependence is proposed to tackle the model uncertainty as well as the compound control problem of heterogeneous actuators in deformable aircrafts. The method is an attitude control scheme based on an adaptive control method, a generalized intelligent attitude control method, and a compound control method. The adaptive attitude control method is based on deep learning (DL), and is used to compensate the aerodynamic moment. The generalized intelligent attitude control method is based on the deep deterministic policy gradient (DDPG) algorithm, and is developed for the aerodynamic and model uncertainties. The compound control method is adopted for the heterogeneous actuators with weak model dependence. The proposed method is a practical intelligent control method, and has better robustness as well as universality compared with the existing ones.

Key words: weak model dependence; adaptive intelligent control; heterogeneous compound control; deep reinforcement learning (DRL); extended state observer

0 引言

随着飞行器的高速发展,其飞行环境及任务也日渐复杂。飞行器在大气层内高速机动飞行时,其速度范围变化大、高度范围覆盖广,因此气动参数也随之大范围快速变化,同时,长时间高速机动飞行导致的气动外形变化,进一步加剧了气动参数的不确定性,这些都要求控制系统具有更强的适应能力。

另一方面,飞行器气动外形从过去单一的轴对称外形已逐步发展到轴对称、面对称气动外形共存的局面。为获得更强的机动能力,针对敏捷机动飞行器的研究也正在如火如荼地进行,这都对姿态控制系统提出了巨大的挑战。

收稿日期: 2022-04-27; **修回日期:** 2022-06-23

作者简介: 邵会兵(1977—),男,博士,研究员,主要研究方向为导航、制导与控制。

通信作者: 詹 韬(1983—),男,硕士,研究员,主要研究方向为导航、制导与控制。

此外，随着飞行器设计的不断进步，其执行机构也日趋多样。多种类执行器为飞行器跨域飞行提供了更强大的控制能力，但也对姿态控制系统提出了多维复合控制要求。

在环境复杂、气动参数大范围不确定性变化的条件下，传统控制器难以实现高精度姿态控制。为解决上述问题，许多学者使用先进控制理论来进行飞行器的姿态控制设计。文献[1-3]使用非奇异终端滑模控制方法来实现环境及模型强不确定性下的飞行器姿态控制，但滑模变结构控制仍然依赖高精度的飞行器模型。文献[4-5]使用自适应动态反演控制方法来实现飞行器的姿态控制，但反演控制方法的补偿精度完全依赖气动数据准确性，当参数大范围不确定变化时，补偿效果通常会下降。文献[6]使用L1自适应控制来应对飞行器姿态控制中面临的诸多不确定性。姿态控制系统采用复合控制能够提升控制能力，增强对外界干扰的适应能力[7]。针对姿态控制系统多维复合控制的需求，现有分配方法通常通过解耦分解[8]、构建分配函数[9]、最小二乘[10]、线性规划等优化方法进行控制分配。文献[11]使用基于二次规划的按需动态分配方法，实现了飞行器的气动舵面及反作用控制系统的复合控制。这些基于先进控制理论的控制方法均依赖于对被控对象的高精度建模，难以应对现代高速飞行器的姿态控制需要。

若要从根本上解决现有姿态控制方法与现实需求之间的矛盾，必须降低控制方法对模型的依赖程度，以提高对模型不确定性的适应性，增强对不同气动外形的通用性。文献[12]使用弱模型依赖方法实现了高性能船舶稳定控制，但其将未建模部分视为误差，通过观测器进行干扰补偿的方法弥补，难以实现复杂未建模动态下的稳定控制。文献[13-14]使用深度强化学习(Deep Reinforcement Learning，DRL)实现不依赖模型的控制算法，但其直接基于深度神经网络输出控制量，缺乏鲁棒性分析，难以用于工程实践和满足通用化设计需求。本文在深入研究了弱模型依赖姿态控制技术的基础上，遵循控制系统发展规律，提出了“基于深度学习(Deep Learning，DL)的自适应姿态控制、基于深度确定性策略梯度算法(Deep Deterministic Policy Gradient，DDPG)的通用姿态控制、弱模型依赖的多维复合控制技术”3个智能程度逐层递进的姿态控制方案，力图为飞行器姿态控制从传统方法逐步走向智能化方法提供一定借鉴。

本文围绕高速飞行器在环境及模型高不确定性变化下的弱模型依赖通用智能姿态控制技术开展研究，第1章提出了基于DL的自适应姿态控制设计，基于DL实现对气动数据变化的预估及前馈补偿；第2章深入研究基于DDPG的通用姿态控制技术，基于DRL实现了传统控制器面向高不确定性环境及模型的进化；第3章研究弱模型依赖的多维复合控制技术，实现针对多维执行机构的复合姿态控制；第4章得出结论，给出分析。

1　基于DL的自适应姿态控制技术

有关基于DL的自适应姿态控制技术的详细内容参见文献[15]。该方法基于小扰动线性化思想，采用“反馈线性化＋自适应PID”控制算法框架，算法结构如图1所示。

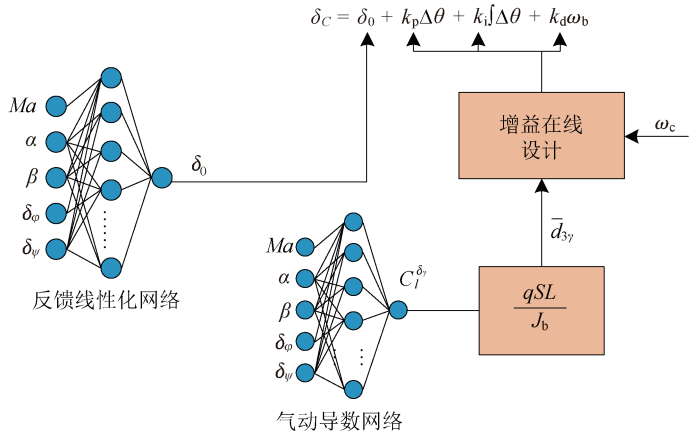


图1　基于DL的自适应姿态控制算法结构

Fig. 1　Schematic of the adaptive attitude control algorithm based on DL

图中：k_p、k_i、k_d为PID控制增益；$\Delta\theta$、ω_b为角偏差和角速度；$C_l^{\delta_\gamma}$为滚转舵产生滚转力矩的系数；ω_c为期望截至频率；α、β为攻角和侧滑角；Ma为马赫数；δ_c为舵指令；J_b、$\bar{d}_{3\gamma}$为转动惯量和舵产生的角加速度。

该方法将气动数据作为训练样本，采用DL技术离线训练获得反馈线性化神经网络和气动偏导数神经网络。并在线根据网络输出自适应调整控制规律，使得控制律仅与飞行状态相关，实现控制律与飞行轨迹的解耦，可满足宽飞行包线、宽飞行空域、宽飞行高度的多样化飞行轨迹控制需求。然而反馈线性化算法补偿精度完全依赖气动数据准确性，一旦由于外形变化等因素导致气动数据天地不一致，补偿效果明显变差，直接导致控制品质下降，甚至失稳。

经飞行器仿真测试[15]，采用上述方法对气动偏差的适应能力约为30%。

2 基于DDPG的通用姿态控制技术

2.1 算法思想

基于DL的自适应姿态控制方法实现了控制律与飞行轨迹的解耦，但神经网络是根据气动数据离线训练获得，不同外形飞行器难以通用，且气动偏差的鲁棒性难以提升；此外，该方法设计仍需设计师对控制器带宽等参数进行精细化设计，对模型和任务的依赖程度仍较高。

为进一步降低控制算法对模型的依赖程度，一方面考虑取消反馈线性化网络，而将控制对象模型的所有非线性部分和外扰均看作系统的“未知扰动”，并采用扩张状态观测器进行观测并实时补偿；另一方面，为解决控制器带宽和观测器带宽自适应最优调节问题，提出采用强化学习离线训练得到控制器和观测器带宽自主调节神经网络，并在线应用该网络实时计算获得最佳带宽，实现期望的最佳控制性能。算法的控制系统框图如图2所示。

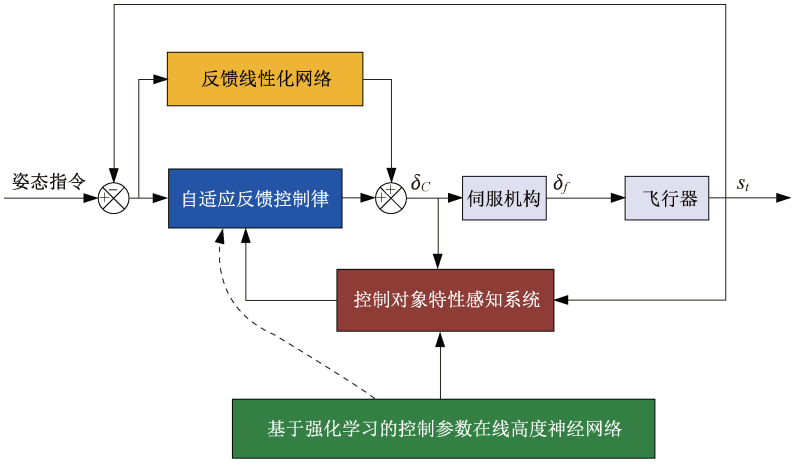


图2 基于DDPG的通用姿态控制框

Fig. 2 Schematic of the generalized attitude control method based on the DDPG

2.2 基于DDPG的通用姿态控制方法

基于DDPG的通用姿态控制算法将智能控制与传统控制进行有机结合，在自抗扰控制器的基础上保留“干扰观测-补偿”框架，增加DRL算法，实现控制器带宽和ESO带宽在线实时调度，进一步提高控制器的性能。自抗扰控制方法是韩京清先生于20世纪80年代末期创建的一种估计补偿不确定因素的控制技术[16]，其将作用于被控对象的所有不确定因素(建模误差和外加干扰)都归结为“总的未知扰动”，并利用控制对象的输入输出数据对它进行估计并给予补偿。

自抗扰控制方法主要由以下3个部分组成：

1) 跟踪微分器。根据控制攻角指令α_c安排过渡过程指令α_{cm}，并提取指令的微分信号$\dot{\alpha}_{cm}$。

2) 反馈控制律。根据系统的控制误差确定反馈控制量。

3) 扩张状态观测器。根据控制对象的输入输出信号对扩张状态(总扰动)进行估计。

将以上跟踪微分器、反馈控制律、扩张状态观测器组合在一起，构成自抗扰控制器，如图3所示。

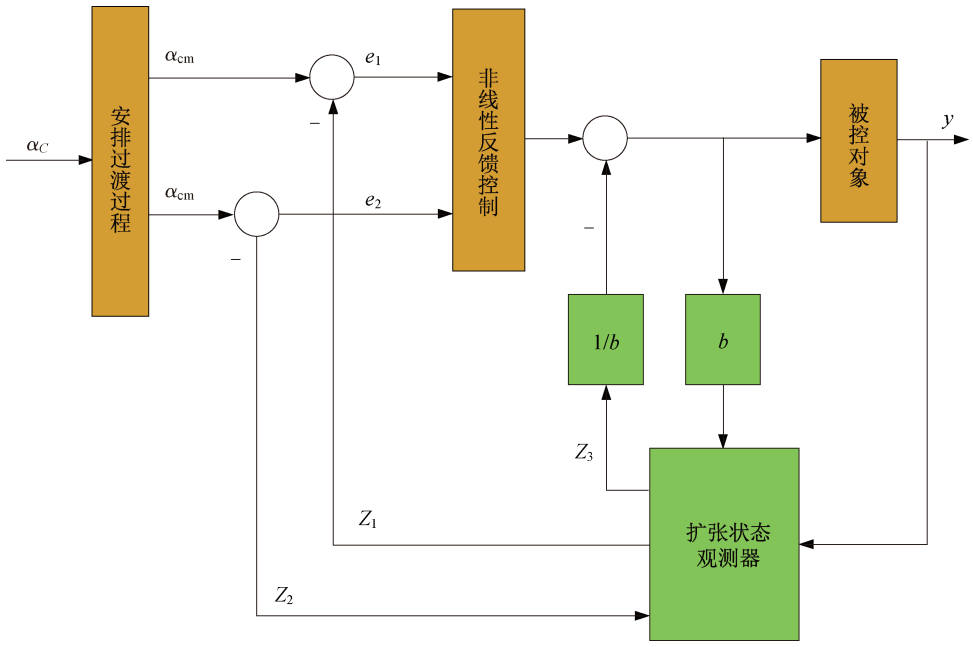


图 3　自抗扰控制器原理框

Fig. 3　Block diagram of the active disturbance rejection control

图中：b 为舵效；z_1、z_2、z_3 为观测量；e_1、e_2 为角度偏差和角速度偏差；y 为系统输出。

2.2.1　跟踪微分器设计

跟踪微分器用于对姿态角指令安排过渡过程，目的是在考虑控制系统实际跟踪能力前提下，合理安排过渡过程以实现跟踪能力范围内的无超调最速跟踪。

通过文献[16]提出一种最速跟踪微分器，其有很好的噪声抑制能力，离散后的形式为

$$\begin{aligned} &fh = fh_{an}(x_1(k)-v(k), x_2(k), r, h) \\ &\begin{cases} x_1(k+1)=x_1(k)+h\times x_2(k) \\ x_2(k+1)=x_2(k)+h\times fh \end{cases} \end{aligned} \tag{1}$$

式中：x_1、x_2 为状态变量；h 为积分步长；r 为控制跟踪速度快慢的变量，r 越大，跟踪速度越快；v 为有界可测信号。为有效消除微分信号进入稳态后的高频振荡，式(1)中的函数选择最速控制综合函数，记为 $fh=fh_{an}(x_1, x_2, r, h)$，其算法公式如下：

$$fh_{an}=-\begin{cases} x\times \mathrm{sign}(a), & |a|>d \\ r\times\dfrac{a}{d}, & |a|\leqslant d \end{cases} \tag{2}$$

式中：$a=\begin{cases} x_2+\dfrac{(a_0-d)}{2}\mathrm{sign}(y), & |y|>d \\ x_2+\dfrac{y}{h}, & |y|\leqslant d \end{cases}$；$d=r\times h$，为线性饱和函数的线性区间；$d_0=d\times h$；$y=x_1+h\times x_2$；$a_0=\sqrt{d^2+8\times r\times|y|}$；sign 为符号函数；$y$ 为下一时刻的状态变量。

2.2.2　非线性反馈控制律设计

采用误差和误差微分的适当非线性组合设计反馈控制率，形式如下：

$$f_{al}(e,a,\delta)=\begin{cases} a|e|^a\times\mathrm{sign}(e), & |e|>\delta \\ \dfrac{e}{\delta^{(1-a)}}, & |e|\leqslant\delta \end{cases} \tag{3}$$

$$u=\beta_0 f_{al}(e_0,a_0,\delta_0)+\beta_1 f_{al}(e_1,a_1,\delta_1)$$

式中：β_0、β_1 为增益；$f_{al}(e,a,\delta)$ 为幂次函数；a、δ 是常数，为 f_{al} 函数所使用的参数；e 为误差。

在 $a<1$ 时，f_{al} 函数具有小误差大增益、大误差小增益的特性，可增强稳态控制精度、抗扰能力，也尽量避免了大误差下控制饱和现象。引入 δ 将函数 f_{al} 改造成原点附近具有线性段的连续函数，可避免高频颤振现象。

2.2.3　扩张状态观测器设计

对于自抗扰控制器来说，最核心是扩张状态观测器，通过建立扩张状态观测量的观测方程，使系统具有扰动估计和补偿的能力[17]。

以飞行器俯仰通道为例，姿态运动动力学方程为

$$\begin{cases}\dfrac{\mathrm{d}x_1}{\mathrm{d}t}=x_2\\ \dfrac{\mathrm{d}x_2}{\mathrm{d}t}=-b_{3\varphi}\Delta\delta_\varphi+x_3\\ \dfrac{\mathrm{d}x_3}{\mathrm{d}t}=\dot{x}_3\end{cases}\tag{4}$$

式中：x_1为$\Delta\alpha$；x_2为$\Delta\dot{\alpha}$；x_3为不确定性估计量；$b_{3\varphi}$为舵效；$\Delta\delta_\varphi$为舵摆角。

扩张状态观测器方程为

$$\begin{cases}e_x=z_1-\theta_x\\ \dfrac{\mathrm{d}z_1}{\mathrm{d}t}=z_2-\beta_{01}e_x\\ \dfrac{\mathrm{d}z_2}{\mathrm{d}t}=-b_{3\varphi}\Delta\delta_\varphi+z_3-\beta_{02}e_x\\ \dfrac{\mathrm{d}z_3}{\mathrm{d}t}=-\beta_{03}e_x\end{cases}\tag{5}$$

式中：z_1为x_1的观测值；z_2为x_2的观测值；z_3为扩张状态变量x_3的观测值，也是“总扰动”的观测值；β_{01}、β_{02}、β_{03}为扩张状态观测器的误差反馈增益。通过合理地选择参数β_{01}、β_{02}、β_{03}，能够使得“总扰动”的观测值更加接近真实值。

2.2.4 DDPG算法的设计与训练

DDPG是在深度Q学习方法基础上，采用了执行器-评价器(Actor-Critic)架构的DRL。其在训练中根据异策略(Off-Policy)数据及贝尔曼方程学习价值函数，并同时使用价值函数来作为学习策略[18-19]。策略即为执行器-评价器架构中的执行器，根据环境反馈的状态，输出系统的连续动作；价值函数即为执行器-评价器架构中的评价器，根据状态及动作，输出策略由状态的期望回报。训练过程即为迭代拟合价值函数及最大化价值函数的策略，直到收敛。

DDPG算法的目标即为最大化策略在当前状态下，未来折扣累积奖励的期望，即：

$$J_\beta(\mu)=E_\mu[\gamma^0 r_1+\gamma r_1+\cdots+\gamma^n r_n]\tag{6}$$

为了找到最优确定性行为策略μ^*，等价于最大化上式目标函数$J_\beta(\mu)$，即：

$$\mu^*=\arg\max J(\mu)\tag{7}$$

根据文献[19]可知，目标函数$J_\beta(\mu)$关于策略网络参数θ^μ的梯度，等价于动作值函数$Q(s,a;\theta^Q)$关于θ^μ的期望梯度。因此，根据链式求导法则，对目标函数进行求导，得到actor网络的更新方式：

$$\begin{aligned}&\nabla_{\theta^\mu}J\approx E_{s_t\sim\rho^\beta}[\nabla_{\theta^\mu}Q_\mu(s_t,\mu(s_t))]=\\ &E_{s_t\sim\rho^\beta}[\nabla_{\theta^\mu}Q_\mu(s,a;\theta^Q)_{|s=s_t,a=\mu(s_t;\theta^\mu)}]\end{aligned}\tag{8}$$

式中：$\nabla_{\theta^\mu}J$为对目标函数J求导；$Q_\mu(s,\mu(s))$为在状态s下，按照确定性策略μ选择动作时，能够产生的动作状态值Q；$E_{s\sim\rho^\beta}$为状态s符合分布ρ^β的情况下Q值的期望。

又因为确定性策略可以表示为$a=\mu(s,\theta^\mu)$的形式，式(8)可以写成：

$$\nabla_{\theta^\mu}J=E_{s_t\sim\rho^\beta}\left[\nabla_a Q(s,a;\theta^Q)_{|s=s_t,a=\mu(s_t)}\nabla_{\theta^\mu}\mu(s_t;\theta^\mu)_{|s=s_t}\right]\tag{9}$$

对式(9)使用梯度策略算法，沿着提高动作值$Q(s,a,\theta^Q)$的方向更新策略网络的参数θ^μ。

价值网络的损失函数：

$$L(\theta^Q)=E[(T_{\text{arget}}-Q(s,a,\theta^Q))^2]\tag{10}$$

式中：T_{arget}为目标Q值。基于式(14)计算神经网络模型参数θ^Q的梯度：

$$\nabla_{\theta^Q}L(\theta^Q)=E(T_{\text{arget}}-Q(s,a,\theta^Q))\cdot\nabla_{\theta^Q}Q(s,a,\theta^Q)\tag{11}$$

式(11)中目标函数表示为

$$T_{\text{arget}}=r+\gamma Q'(s',\mu(s';\theta^{\mu'}))\tag{12}$$

式中：Q'为下一状态目标Q值；$\theta^{Q'}$、$\theta^{\mu'}$分别为目标策略网络和目标价值网络的神经网络参数。价值网络参数的更新和策略网络参数的更新交替迭代，只使用单一的神经网络进行强化学习时，动作值的学习过程很容易出现不稳定现象。

训练价值网络的过程，就是寻找价值网络中参数θ^Q的最优解的过程，DDPG算法训练的目标是最大化目标函数$J_\beta(\mu)$，同时最小化价值网络Q的损失函数。

基于DDPG的通用姿态控制算法，首先需要将姿态控制问题纳入到马尔科夫决策框架下。本文选取的状态变量是飞行器姿态运动状态集合$s=(\Delta\varphi,\Delta\psi,\Delta\gamma,\omega_{xb},\omega_{yb},\omega_{zb},\dot{\omega}_{xb},\dot{\omega}_{yb},\dot{\omega}_{zb})$；动作集定义为控制器的带宽$\omega_c$和观测器带宽$\omega_o$；状态转移模型为飞行器6自由度仿真模型；奖励函数$r_{\text{奖励}}$为飞行状态参数和执行机构指令的函数，$r_{\text{奖励}}=f(\Delta\varphi,\Delta\psi,\Delta\gamma,\omega_{xb},\omega_{yb},\omega_{zb},\dot{\omega}_{xb},\dot{\omega}_{yb},\dot{\omega}_{zb},\delta)$。

根据上述建立的马尔科夫决策过程，利用DDPG方法进行地面离线仿真训练，其训练算法框架如图4所示。

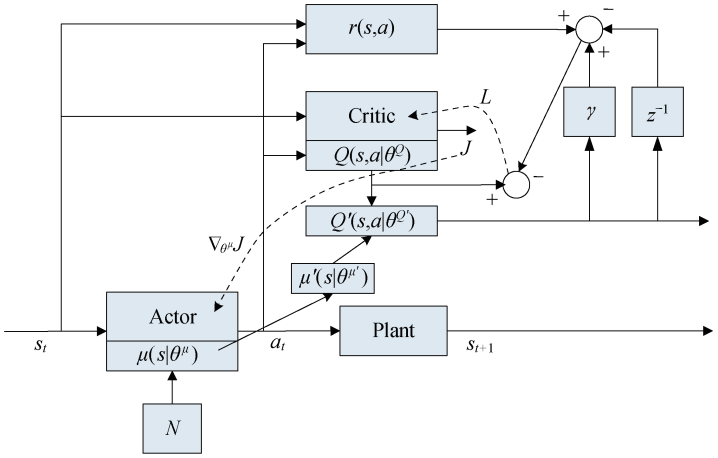


图 4　DDPG 训练算法框架

Fig. 4　Schematic of the DDPG training algorithm

本文针对固定速度 1 200 m/s 及飞行高度 45 km 的高速飞行器姿态控制任务进行训练，训练阶段姿态角指令为一固定幅值的阶跃信号。训练获得了比较理想的控制效果，其各回合累积回报的变化曲线如图 5 所示。

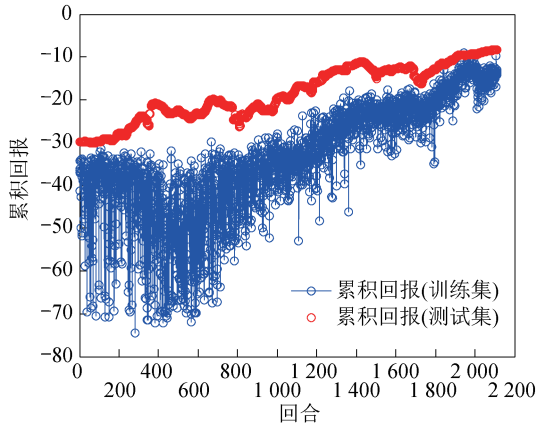


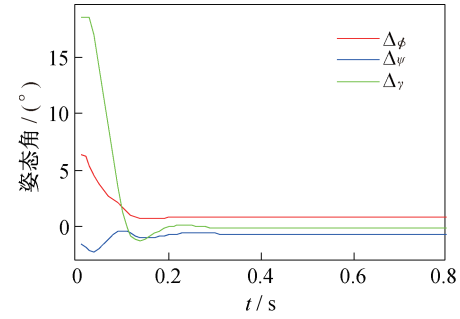
图 5　累积回报的变化曲线

Fig. 5　Curves of the cumulative rewards

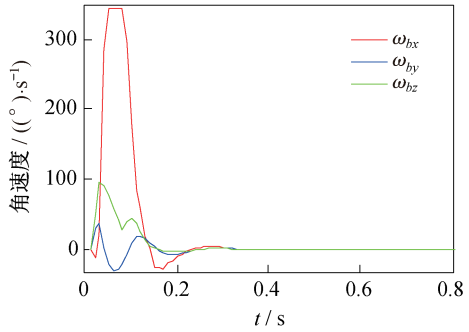
最后一个回合中姿态角偏差及姿态角速度的变化情况如图 6 所示。从图 6 中可知，Agent 学习到了有效的控制参数调节规律，飞行器可以快速跟踪姿态角指令，且精度较高。可见，取消了前馈补偿模块，并没有影响姿态控制的性能，表明本文所提出的“基于 DDPG 的通用姿态控制方法”是有效可行的。

2.3　基于 DDPG 的通用姿态控制算法验证

应用 Agent 学习到的控制参数调节律网络进行气动参数大范围拉偏条件下仿真验证。连续进行



(a) 姿态角偏差变化曲线



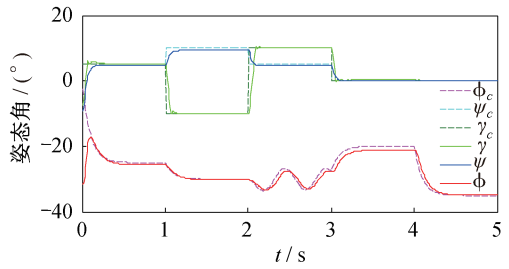
(b) 角速度变化曲线

图 6　最终回合姿态相关量变化曲线

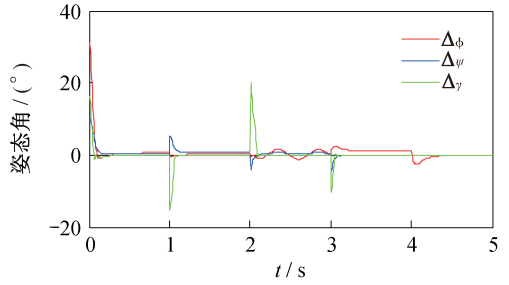
Fig. 6　Curves of the attitude angle error and angular velocity in the last episode

5 次调姿，姿态角指令除阶跃信号外还包含正弦信号，气动参数拉偏 50%，速度取 850 m/s（训练阶段并未针对该速度进行训练）。在这种条件下，相应的姿态角跟踪曲线如图 7 所示。

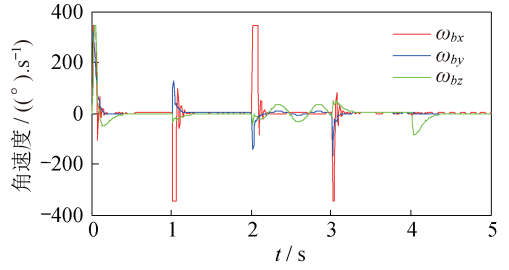
可见该方法设计过程简单，对气动参数和总体结构参数变化适应能力强，算法通用性强，在不同速度下能够适应多种形式的指令，且控制性能保持良好，即使在气动系数大范围拉偏的情况下，仍能



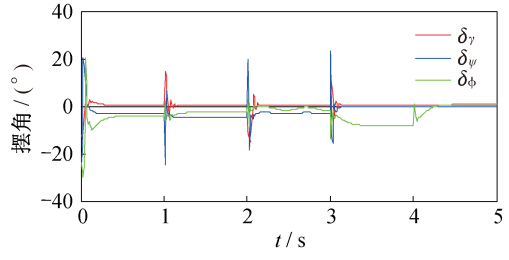
(a) 姿态角跟踪曲线



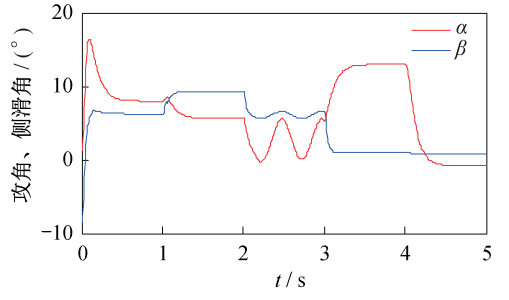
(b) 姿态角偏差变化曲线



(c) 角速度变化曲线



(d) 三通道舵摆角变化曲线



(e) 攻角、侧滑角变化曲线

图 7 姿态跟踪曲线

Fig. 7 Curves of the attitude tracking results

够实现姿态的高精度稳定跟踪,可以认为该方法实现了姿态控制系统通用化设计。

3 弱模型依赖的多维复合控制技术

3.1 算法思想

上述姿态控制算法将多约束、强不确定性的姿态跟踪问题转化为自适应动态规划问题,并引入DRL算法离线迭代优化,建立了较为通用的算法设计流程,显著提升对气动参数大范围偏差的适应能力,但仍存在如下问题:

1) 当前高速飞行器具有推力矢量、直接力以及空气舵等多维异类执行机构,该算法针对特定单一执行机构设计,难以适应上述执行机构的独立/复合控制[20];

2) 动力系数在线辨识与干扰观测分离设计,降低对象特征感知效率和精度,极端情况下可能影响闭环系统稳定性;

3) 可适应的气动参数变化范围有限,难以适应未来飞行器敏捷机动控制需求。

针对上述问题,本文提出"弱模型依赖的多维复合控制技术"。首先,考虑连续、离散姿态控制的统一,构建面向通用控制的动力学特征模型;其次,在此基础上采用"平行估计器+鲁棒自适应控制器+参数调度律+智能分配律"的算法框架,并将估计器、控制器及分配律的设计参数选取抽象为优化问题,引入强化学习算法解决,实现了多维异类复合控制;最后,降低控制算法对精确模型的依赖,发挥扰动条件下的最优性能,同时控制动态分配也能够实现执行机构典型非致命故障的容错控制。算法原理框图如图8所示。

3.2 面向通用控制的动力学特征模型

3.2.1 通用全局特征模型

传统面向控制模型常采用平衡点附近线性化的小扰动模型,相较于飞行器本质的动力学模型,经过了轨迹域、姿态域、时间域多个维度的约束和简化,无法满足新一代高速飞行器宽域、大机动敏捷操纵等需求。为解决上述矛盾,构建飞行器通用全局特征模型为

$$\dot{\omega}_{ib} = F_{\omega_{ib}} + G_{\omega_{ib}} + G_{\omega_{ib}\_C} + d_{\omega_{ib}} \tag{13}$$

通用全局特征模型构建的核心思路为:从原始动力学模型出发,明确模型中的已知部分 $F_{\omega_{ib}}$、含不确定参数的非控制部分 $G_{\omega_{ib}}$、含不确定参数的控制部分 $G_{\omega_{ib}\_C}$,以及未建模部分 $d_{\omega_{ib}}$。一方面,对

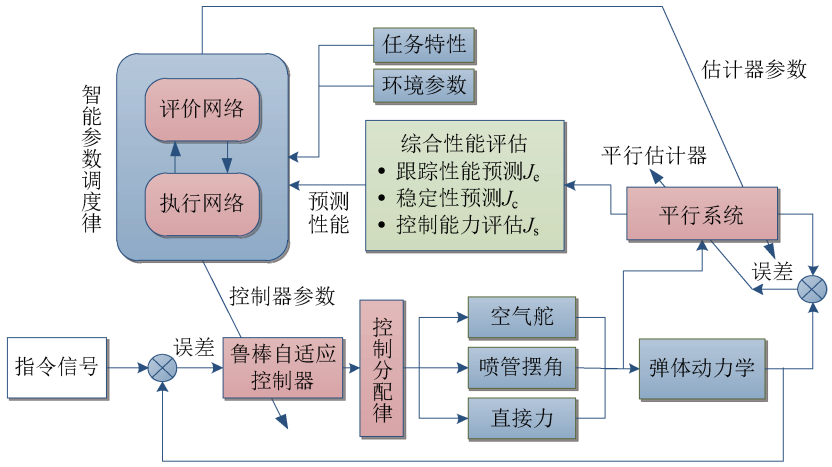


图 8　弱模型依赖的多维复合控制算法框架

Fig. 8　Schematic of the compound control method with weak model dependence

于能够通过标称气动数据获得的非线性已知部分给予最大限度的保留；另一方面，明确建模偏差部分来源和特性，方便后续通用干扰观测器设计。通过上述建模，能够充分利用参数化的气动模型、气动在线辨识结果，大幅降低面向控制模型与真实模型之间的区别，有利于降低控制器的保守性、设计高性能控制并实现宽域自适应、敏捷机动等需求。

3.2.2　多维异类控制量映射

高速飞行器的多维异类控制分配问题可描述为

$$\delta_i = G_\delta^{\text{rudder}}\delta_{\text{r}} + G_\delta^{\text{thrust}}\delta_{\text{t}} + \Gamma u_{\text{rcs}} \tag{14}$$

式中：$\delta_i\left(i=\varphi,\psi,\gamma\right)$为通用全局特征模型中的虚拟控制量；$\delta_{\text{r}}$、$\delta_{\text{t}}$分别为空气舵、喷管摆角，是连续控制量；$u_{\text{rcs}}$为直接力开关指令，离散控制量。

由此建立了多维异类控制量映射模型，为后续智能分配律设计奠定基础，原理框图如图 9 所示。

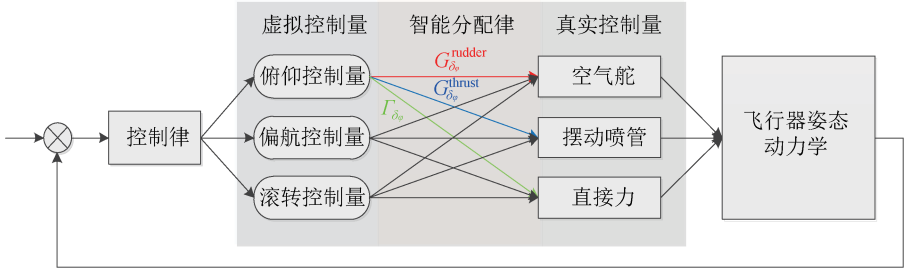


图 9　多维异类控制量映射框

Fig. 9　Schematic of the multi-dimentional heterogeneous control command mapping

3.3　通用姿态控制器设计

3.3.1　通用姿态控制框架

考虑到根据标称预示模型设计的控制器通用性差，宽域机动和敏捷机动飞行时性能较差，本文采用“平行估计器＋鲁棒自适应控制器＋参数调度律＋智能分配律”算法框架。

1）平行估计器：根据动力学输入和输出数据对模型中的未知参数和干扰进行一体化估计，并根据估计结果构建导弹姿态动力学平行系统。

2）鲁棒自适应控制器：采用快-慢双通道滑模控制器构建基本控制律，结合模型估计器的估计信息，实现全局鲁棒自适应控制，求得“虚拟控制量”。

3）参数调度律：负责对控制器和模型估计器的自身参数进行智能最优调节，采用评价器-执行器框架，离线训练网络初值，在线增量式学习。

4）控制分配律：根据控制约束、飞行器目前状态及各种执行机构控制效率的分布，采用一定的分配策略，实现对不同执行机构控制输出的分配，以期在高精度实现“虚拟控制量”条件下，使控制消耗最低。

3.3.2　鲁棒自适应控制器

基于特征模型，按照被控变量对控制输入量响应快慢的特点进行快慢时标分离，构成快回路和慢

回路子系统,并考虑统一连续控制和开关控制需求,分别针对快慢回路设计拟滑模控制律实现全局鲁棒控制,结构如下:

$$\begin{cases} \text{外环} \begin{cases} e_\alpha = \alpha - \alpha^r \\ \omega_{zb}^{\text{ref}} = b_{\omega_{zb}}^{-1}\left[\dot{\hat{\alpha}}^r - f_\alpha - \hat{\bar{d}}_\alpha - \hat{\Gamma}_\alpha \text{sig}(\mu_\alpha e_\alpha)\right] \\ \dot{\hat{\Gamma}}_\alpha = \eta_\alpha e_\alpha \text{sig}(\mu_\alpha e_\alpha), \hat{\Gamma}_\alpha(0) \geqslant 0 \end{cases} \\ \text{内环} \begin{cases} e_{\omega_{zb}} = \omega_{zb} - \omega_{zb}^{\text{ref}} \\ \delta_\varphi = b_\varphi^{-1}\left[-b_{\omega_{zb}} e_\alpha + \dot{\hat{\omega}}_{zb}^{\text{ref}} - f_{\omega_{zb}} - \right. \\ \qquad \left. \hat{\bar{d}}_{\omega_{zb}} - \hat{\Gamma}_{\omega_{zb}} \text{sig}(\mu_{\omega_{zb}} e_{\omega_{zb}})\right] \\ \dot{\hat{\Gamma}}_{\omega_{zb}} = \eta_{\omega_{zb}} e_{\omega_{zb}} \text{sig}(\mu_{\omega_{zb}} e_{\omega_{zb}}), \dot{\hat{\Gamma}}_{\omega_{zb}}(0) \geqslant 0 \end{cases} \end{cases} \tag{15}$$

式中:α^r为指令攻角;ω_{zb}^{ref}为内环角速度参考信号;μ_α、$\mu_{\omega_{zb}}$为外环、内环控制精度;$\hat{\Gamma}_\alpha(0)$、$\hat{\Gamma}_{\omega_{zb}}(0)$为外环、内环自适应增益初值;$\eta_\alpha$、$\eta_{\omega_{zb}}$为外环、内环控制精度参数;sig为滑膜控制器使用的sigmoid函数;f_α、$f_{\omega_{zb}}$为动力方程已知部分;$b_{\omega_{zb}}^{-1}$、b_φ^{-1}为动力方程控制部分;$\hat{\bar{d}}_\alpha$、$\hat{\bar{d}}_{\omega_{zb}}$为动力方程未知扰动部分;$e_\alpha$、$e_{\omega_{zb}}$为误差。

可见,上述控制律为全局非线性形式,同时利用特征参数/干扰一体化在线估计结果,能够应对宽域飞行导致的动力学强不确定性。

3.3.3 智能控制分配

由于存在多种操纵机构,且操纵机构的作用力或力矩可能存在冗余,因此如何合理分配虚拟控制量到实际执行机构成为关键,将强化学习思路应用于智能分配律设计,构建控制分配的马尔科夫决策过程,其中奖励函数的设计至关重要。

奖励函数分为2部分:权重较大的基础奖励r_0、权重较小的附加奖励Δr。基础奖励与虚拟控制量的实现偏差相关,$\boldsymbol{H}$为权重矩阵,具体形式如下:

$$r_0 = \boldsymbol{e}_\delta^{\text{T}} \boldsymbol{H} \boldsymbol{e}_\delta \tag{16}$$

式中:$\boldsymbol{e}_\delta$为误差。

附加奖励则考虑控制损耗,$\boldsymbol{W}_i(i=1,2,3)$为权重矩阵,具体为

$$\Delta r = \boldsymbol{\delta}_r^{\text{T}} \boldsymbol{W}_1 \boldsymbol{\delta}_r + \boldsymbol{\delta}_t^{\text{T}} \boldsymbol{W}_2 \boldsymbol{\delta}_t + \boldsymbol{u}_{\text{rcs}}^{\text{T}} \boldsymbol{W}_3 u_{\text{rcs}} \tag{17}$$

式中:$\boldsymbol{u}_{\text{rcs}}$为直接力开关指令。

由此将虚拟控制量的动态分配问题等效为优化问题,采用DRL算法解决。

3.4 基于DRL的多维控制参数自进化

为更好地实现未知外界扰动及复杂动力学特性下飞行控制系统的控制性能,在已有的控制系统结构下通过构建平行系统实现对控制器、估计器以及控制分配参数的在线智能优化。采用执行-评价网络结构(A-C框架),离线训练好网络初值,通过建立效用函数与策略函数描述控制性能指标,根据平行系统跟踪误差、稳定性、控制能力(剩余执行机构控制量、剩余执行机构变化速率、控制效率)等进行综合评价,结合期望最优控制性能动态修正控制参数和估计器参数,并实现智能控制分配。算法原理框图如图10所示。

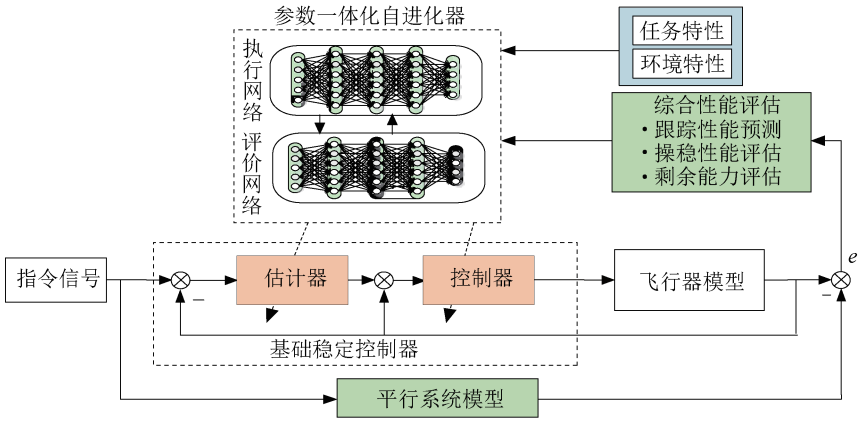


图10 基于DRL的控制参数自进化框

Fig. 10 Schematic of the control parameter self-evolving based on deep reinforcement learning

4 结束语

本文从传统姿态控制律设计方法严重依赖精确控制对象模型问题出发,提出了基于DL的自适应姿态控制、基于DDPG的通用姿态控制、弱模型

依赖的多维复合控制3个智能化程度逐层递进的控制方案。该方案可显著提升飞行控制系统对气动偏差、干扰的适应性以及对不同外形飞行器的通用控制能力，实现了控制算法对控制对象模型的弱依赖，对人工智能技术在飞行器姿态控制中的应用提供了一种切实可行的思路。

参考文献

[1] ZHANG L, WEI C Z, WU R, et al. Fixed-time extended state observer based non-singular fast terminal sliding mode control for a VTVL reusable launch vehicle[J]. Aerospace Science and Technology, 2018, 82: 70-79.

[2] ZHANG R, LU D, SUN C. Adaptive nonsingular terminal sliding mode control design for near space hypersonic vehicles [J]. IEEE/CAA Journal of Automatica Sinica, 2014, 1(2): 155-161.

[3] QIAO J, LI Z, XU J, et al. Composite nonsingular terminal sliding mode attitude controller for spacecraft with actuator dynamics under matched and mismatched disturbances [J]. IEEE Transactions on Industrial Informatics, 2020, 16(2): 1153-1162.

[4] ANSARI U, BAJODAH A H. Launch vehicle ascent flight attitude control using direct adaptive generalized dynamic inversion [J]. Proceeding of the Institution of Mechanical Engineering, Part G: Journal of Aerospace Engineering, 2019, 233(11): 4141-4153.

[5] 董朝阳，路遥，王青．高超声速飞行器指令滤波反演控制[J]．宇航学报，2016，37(8)：957-963.

[6] 钟京洋，宋笔锋．基于鲁棒伺服思想的尾坐式飞行器悬停姿态控制[J]．控制与决策，2020，35(2)：339-348.

[7] 周如好，张卫东，胡存明，等．运载火箭推力矢量/非线性复合控制方法研究[J]．上海航天(中英文)，2016，33(增刊1)：81-85.

[8] YANG C, ZHONG S, LIU X, et al. Adaptive composite suboptimal control for linear singularly perturbed systems with unknown slow dynamics [J]. International Journal of Robust and Nonlinear Control, 2020, 30:2625-2643.

[9] 郭建国，吴林旭，周军．非对称变翼飞行器复合控制系统设计[J]．宇航学报，2018，39(1)：52-59.

[10] 刘胜，王宇超，傅荟璇．船舶航向保持变论域模糊-最小二乘支持向量机复合控制[J]．控制理论与应用，2011，28(4)：485-490.

[11] 董哲，刘凯，李旦伟．考虑动态分配控制的空天飞行器再入姿态复合控制设计[J]．宇航学报，2021，42(6)：749-756.

[12] 刘旌扬．弱模型干扰补偿控制方法及其在高性能船舶姿态稳定控制中的研究应用[D]．上海：上海交通大学，2011.

[13] 裴培，何绍溟，王江，等．一种深度强化学习制导控制一体化算法[J]．宇航学报，2021，42(10)：1293-1304.

[14] 孔维仁，周德云，赵艺阳，等．基于深度强化学习与自学习的多无人机近距空战机动策略生成算法[J]．控制理论与应用，2022，39(2)：352-362.

[15] 邵会兵，崔乃刚，詹韬．基于神经网络的飞行器控制方法及仿真研究[J]．计算机仿真，2018，35(10)：94-98.

[16] 韩京清．自抗扰控制技术：估计补偿不确定因素的控制技术[M]．北京：国防工业出版社，2008.

[17] 孙明玮，马顺健，朴敏楠．高超声速飞行器自抗扰控制方法[M]．北京：科学出版社，2018.

[18] RICHARD S S, ANDREW G. Reinforcement learning: an introduction[M]. Cambridge, USA: MIT Press, 2017.

[19] SILVER D, LEVER G, HEESS N, et al. Deterministic policy gradient algorithms [C]// Proceedings of the 31st International Conference on Machine Learning. New York: ACM Press, 2014: 387-395.

[20] HE S, LIN D, WANG J. Compound control methodology for a robust missile autopilot design [J]. Journal of Aerospace Engineering, 2015, 28(6): 1-10.