基于增量式策略强化学习算法的飞行控制系统的容错跟踪控制

任 坚, 刘剑慰†, 杨 蒲

(南京航空航天大学 自动化学院, 江苏 南京 211106)

摘要: 针对发生故障的飞行控制系统, 在强化学习算法的基础上, 提出了一种基于增量式策略的强化学习容错方法. 本方法利用传感器获取的系统状态值, 根据系统预先设定的奖励函数对当前控制系统状况做出最优的决策并不断更新价值网络, 将系统的容错控制过程转换为强化学习Agent的贯序决策过程, 并使用一种改进型的增量式策略实现对当前故障的正确补偿策略的逐渐逼近. 同时, 针对连续控制系统, 提出一种状态转移预测网络来得到下一步状态值. 最后, 通过南京航空航天大学“先进飞行器导航、控制与健康管理”工信部重点实验室的飞行器故障诊断实验平台验证了该方法的有效性.

关键词: 飞行控制系统; 故障诊断; 故障容错; 强化学习; Q-learning算法; 增量式策略; 状态转移预测网络

引用格式: 任坚, 刘剑慰, 杨蒲. 基于增量式策略强化学习算法的飞行控制系统的容错跟踪控制. 控制理论与应用, 2020, 37(7): 1429 – 1438

DOI: 10.7641/CTA.2020.90380

Fault-tolerant tracking control for continuous flight control system based on reinforcement learning algorithm with incremental strategy

REN Jian, LIU Jian-wei†, YANG Pu

(College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing Jiangsu 211106, China)

Abstract: A reinforcement learning method based on incremental strategy is proposed to make fault-tolerant tracking control for continuous flight control system with faults. The system state value obtained by the sensor is used in the method proposed by this paper, The fault-tolerant system makes optimal decisions on the current control system conditions based on pre-set reward functions and continuously updates the value network, This transforms the fault-tolerant control process of the system into a sequential decision-making process of the reinforcement learning agent, and gradually approximates the specific fault value using an improved incremental strategy. what's more, A state transition prediction network is proposed for the continuous control system to obtain the next state value. Finally, The effectiveness of the proposed method is verified by the aircraft fault diagnosis experimental platform of the Key Laboratory of Advanced Aircraft Navigation, Control and Health Management of Nanjing University of Aeronautics and Astronautics.

Key words: flight control systems; fault diagnosis; fault tolerance; reinforcement learning; Q-learning algorithm; incremental strategy; state transition prediction

Citation: REN Jian, LIU Jianwei, YANG Pu. Fault-tolerant tracking control for continuous flight control system based on reinforcement learning algorithm with incremental strategy. *Control Theroy & Applications*, 2020, 37(7): 1429 – 1438

1 引言

随着航空航天技术的不断发展, 飞行控制系统的规模变得越来越庞大, 系统的复杂度也不断地增加. 在飞行控制系统不断进步的同时, 系统的稳定性也面临着巨大的挑战. 任何类型的故障都可以导致系统性能的折损甚至是瘫痪, 造成控制系统的不稳定, 从而

收稿日期: 2019−05−25; 录用日期: 2020−01−13.

†通信作者. E-mail: ljw301@nuaa.edu.cn; Tel.: +86 13915983317.

本文责任编委: 宗群.

民航飞机健康监测与智能维护重点实验室基金项目(NJ2018012), 先进飞行器导航、控制与健康管理工业和信息化部重点实验室(南京航空航天大学)项目, 中央高校基本科研业务费项目(NS2017017), 国家自然科学基金项目(61533008, 61490703)资助.

Supported by the Civil Aviation Aircraft Health Monitoring and Intelligent Maintenance Key Laboratory (NJ2018012), the Key Laboratory of Advanced Aircraft Navigation, Control and Health Management (NS2017017) and the National Natural Science Foundation of China (61533008, 61490703).

带来巨大的损失. 因此, 如何减小甚至是消除系统故障所带来的危险是一个值得研究的问题, 为了克服传感器、执行器和其他部件的故障, 国内外学者们在故障诊断与容错控制这一研究方向上做出了很多的努力.

1993年, 国际上出现了第一篇关于容错控制的综述性论文[1]. 针对处于故障下的被控系统, 可以采取可靠镇定控制的方法, 通过并联多个补偿控制器同时控制被控系统, 使系统发生故障时依旧保持系统的稳定[2]. 此外, Gopinathan和Boskovic等人提出了一种完整性控制的思想, 当被控系统发生传感器或执行器故障时, 通过完整性控制的方法仍能够保证系统的稳定性和安全性[3]. 上述的方法均为被动容错控制, 虽然被动容错控制被证明了具有一定的鲁棒性, 不需要对系统的故障进行在线监测与评估, 可以在飞行控制系统执行器发生可能的失效故障情形下正常飞行, 但是被动容错控制还是存在一定的保守性. 相较于被动容错控制, 主动容错控制通过对故障的在线评估与控制信号的重构保证了系统的稳定运行. 在主动容错控制领域, 主动容错控制器通过故障监测与诊断单元(fault detection and diagnosis, FDD)来实时获取在线的故障信息, 并根据在线故障信息调整控制系统的参数或结构, 对控制器进行重构设计, 是系统在发生故障后还能正常运行[4–6]. 此外, 文献还可以采用模型跟随重组的方法, 通过对输出误差信号的实时跟踪已达到容错控制的目的, 此方法虽然为主动容错控制方法, 但不需要FDD单元, 大大方便了系统的设计[7].

近年来的研究工作大多聚焦在系统控制器的设计上, 大多采用基于模型的方法对系统控制器进行重构[8–22], 而由于科学技术的发展, 飞行控制系统的复杂度越来越庞大, 这也为对飞控系统的数学建模带来了巨大的挑战, 由于基于模型的方法能够成功实现的前提是对系统的精准建模, 所以随着控制系统越来越复杂, 基于模型的方法的局限性也体现了出来. 而随着大数据和人工智能的兴起, 国内外的研究者也开始将目光转投向基于数据的跟踪控制方法[23–27], 由于基于数据方法的较高工程应用价值, 最近几年也引来了越来越多业界的关注. Ng等人通过强化学习的方法实现了直升机的倒立飞行[28]. Williams对于强化学习控制器的状态—动作对的采样方法进行了改进, 通过随机采样来加快学习的收敛速度[29]. Wang和Liu等人提出了一种优化的强化学习容错控制器的训练方法, 可以大大节省计算资源[30]. 相较于Wang和Liu 等人提出的确定性策略, 本文创新性地提出了增量式策略方法, 将使强化学习方法能够较好地完成容错控制任务.

本文将针对执行器故障与传感器故障情况下的飞行控制系统, 基于强化学习的Q-learning方法, 提出了一种主动容错控制方案. 采用增量型策略的强化学习方法, 研究发生多种故障(执行器故障与传感器故障)下的主动容错控制技术, 相较于基于模型的主动容错控制的方法, 采用强化学习方法进行容错控制, 若系统故障产生, 强化学习控制器可以提取系统实时产生的数据的特征, 从而获取当前状态的故障信息, 并通过某种策略让控制器做出在当前状况下能做出的最优策略, 使得系统能够保持稳定, 从而正常运行. 由于系统故障的不确定性, 无法通过一个确定性的策略做决策, 本文采用了优化的增量型策略代替了确定性策略, 通过在系统容错控制过程中控制器对增量型策略的迭代来重构控制器, 从而使控制器能够渐进逼近最优解. 与传统的强化学习的确定性策略不同, 由于控制系统故障发生的不确定性, 通过确定性的固定策略对控制系统进行容错控制的效果肯定会差强人意, 本文提出的增量式策略便能解决这个问题, 通过对上一步策略的保持并进行新一轮的更新从而达到上面所说的逼近效果的思想达到容错控制的目的. 本文中的强化学习方法主体采用Q-learning算法结构, 通过对数据的实时采集在线更新决策网络, 使得控制器Agent能够准确地预判下一步策略. 同时, 由于强化学习方法需要使用当前状态与动作作用下的下一个状态信息, 本文将采用状态转移预测网络对连续的控制系统的下一步状态进行预测, 并通过存储地历史信息来对状态预测网络定期做训练更新, 从而实现对决策网络实时的更新.

2 系统故障问题描述

考虑一类连续飞行控制系统为

$$\begin{cases} \dot{x}(t) = Ax(t) + B(u(t) + \phi(t - t_1)f_{\rm a}(t)), \\ y(t) = Cx(t) + Du(t) + \phi(t - t_2)Ff_{\rm s}(t), \end{cases} \tag{1}$$

其中: $x \in \mathbb{R}^{4\times 1}$为系统的状态, $x = [\theta\ \ \varphi\ \ \dot{\varphi}\ \ \hat{\varphi}]$, 其中$\theta$为俯仰角变量, φ为滚转角变量; $u = [u_1\ \cdots\ u_4]^{\rm T}$为控制输入; 系统矩阵$A \in \mathbb{R}^{4\times 4}$, $B \in \mathbb{R}^{4\times 1}$, $C \in \mathbb{R}^{1\times 4}$; $y \in \mathbb{R}$为系统输出变量; $\phi(t-t_1)f_{\rm a}(t)$, $\phi(t-t_2)Ff_{\rm s}(t)$分别表示飞行控制系统中的执行器故障和传感器故障, $t \in \mathbb{R}^+$, 其中: $f_{\rm a}(t)$为未知的执行器故障偏置值, $Ff_{\rm s}(t)$为未知的传感器故障的偏置值, 其中: $F \in \mathbb{R}^{1\times 4}$, $f_{\rm s}(t) \in \mathbb{R}^{4\times 1}$; $\phi(t - t_{\rm f})$定义为故障产生时间定义函数:

$$\phi(t - t_{\rm f}) = \begin{cases} 0,\ t < t_{\rm f}, \\ 1,\ t \geqslant t_{\rm f}, \end{cases} \tag{2}$$

其中$t_{\rm f}$为飞行控制系统中的未知故障产生的时间, 在本文中通过$\phi(t - t_{\rm f})$函数表示系统的突发性故障(在时间$t_{\rm f}$之后故障发生). 在开始设计控制器之前, 需要为研究对象做一些假设.

假设 1 控制系统中存在可以被测量的状态变

量, 在控制系统运行过程中, 系统的部分甚至所有状态变量是能够被观测的.

假设1保证了系统中状态和输出的可测量性, 这对基于数据的容错控制技术至关重要, 这使得强化学习控制器可以通过控制系统运行时的数据去评估奖励函数并更新训练价值网络, 从而使控制器做出最好的调整, 使控制系统处于当前最优的状态. 如果系统的某些状态或者是输出不能用传感器技术去测量, 也可通过观测器和滤波器技术进行得到, 这种情况也满足假设1.

假设 2　存在已知的正数常量ξ, $\delta > 0$使得执行器故障$f_{\rm s}(t)$和传感器故障$Ff_{\rm s}(t)$满足

$$\begin{cases} \|f_{\rm a}(t)\| \leqslant \xi, \\ F\|f_{\rm s}(t)\| \leqslant \delta, \end{cases} \tag{3}$$

其中$\|\cdot\|$表示欧几里得范数.

假设2保证了故障未知但满足范数有界的条件, 由于本篇文章采用的是增量式策略更新重构信号的方法, 所以只有满足执行器故障有界, 才能保证容错控制器能够得到最终最优的策略.

假设 3　在系统正常运行状态下, 系统将沿着参考输出信号轨迹$y_{\rm d}(t)$做跟踪运动. 于是, 在本文中假设参考输出信号$y_{\rm d}(t) \in \Omega_{\rm y}$, 其中: $\Omega_{\rm y}$是一个有界的集合, 而$y_{\rm d}(t)$是一个属于集合$\Omega_{\rm y}$中的光滑曲线函数, 例如正弦信号等.

本论文的目标是设计出基于强化学习方法的主动容错控制器, 通过增量型策略对发生未知故障的飞行控制系统进行容错控制, 利用增量型的策略较快的逼近的最优策略, 以达到做出对控制系统最优的调整, 而实现以上目标可以转换成满足一下条件: 1) 通过对系统损失指标的最小化, 从而做出当前最优的动作; 2) 系统输出y可以快速地跟踪参考信号$y_{\rm d}$, 并且整个闭环系统的所有状态变量均是有界的.

3　基于强化学习的主动容错控制器设计

本节将对主动容错控制器系统状态进行定义, 设计控制器奖励函数, 并对增量型策略进行介绍. 针对连续系统的状态预估, 介绍状态转移预测网络方法.

3.1　容错控制器系统状态与状态—动作奖励函数

系统状态是从控制系统中选出的能够反应控制系统运行状态的变量, 它能够反应系统当前的运行状况, 容错控制器通过系统状态抽取有用的特征, 将其作为容错控制器进行决策的重要依据. 将容错控制系统状态定义为飞行控制系统的姿态角:

$$S = [\phi \ \ \theta \ \ \dot{\theta} \ \ \hat{\theta}], \tag{4}$$

选取的状态参数依据为本文所采用的仿真平台所能采集到的状态量, 其中ϕ, θ分别为实验平台飞行器的滚转角和俯仰角, 这些状态参数将作为评价网络的输入从而得到奖励值. 状态—动作奖励函数可以看作容错控制器对不同策略好坏的打分, 而这个分数便是系统损失指标的依据, 通过对其最小化决策出当前最优的策略进行控制系统的容错, 于是奖励函数直接影响着控制器的决策效果的好坏, 针对控制系统的跟踪控制, 定义当前状态–动作奖励函数$J(S_t)$:

$$J(S_t) = \sum_{j=0}^{t} \gamma^j U(S_{t-j}, A_{t-j}), \tag{5}$$

式中: γ为折扣因子, 满足$0 < \gamma \leqslant 1$; $U(S_{t-j}, A_{t-j})$为强化学习算法的效用函数:

$$U(S_t, A_t) = Q(S_t, A_t), \tag{6}$$

$$Q(S_t, A_t) = |y(t, A_t) - y_{\rm d}(t)|, \tag{7}$$

其中$y(t, A_t)$表示在时刻t下, 采用动作A_t所得到的系统输出, 因而$Q(S_t, A_t)$则表示在t时刻系统状态为S_t, 采用动作A_t系统的输出与期望输出之间的误差的绝对值, 而设计控制器的目的便是选取动作A_t, 使得此时的状态–动作奖励函数最小.

3.2　增量型策略方法

传统的强化学习方法运用在非控制领域常采用确定性策略, 谷歌DeepMind团队[31]采用深度强化学习方法研发出了人工智能象棋Agent, 由于每步策略独立的前提下, 因而采取的策略是确定性的策略. 而相较于象棋, 发生故障的控制系统由于故障的未知性, 采用确定性的策略会使容错控制器的适用范围受到很大的限制. 本文提出了一种基于增量型策略的强化学习Q-learning算法, 从而通过策略地叠加实现最优策略地渐近逼近, 同时丰富了策略的多样性, 充分发挥了强化学习Agent的自我学习特性.

3.2.1　策略选取方法

本文中采取参考文献[32]中的ε-greedy策略, 在控制器的训练过程中引入ε-greedy因子对策略网络进行更新, 而ε-greedy因子表示此次策略选择随机策略的概率. 同时, 因子ε采取退火操作, 从1退火到0. 因此在退火初期, 控制器选择随机策略的概率较大, 此阶段为初步探索阶段, 随着策略迭代到后期, 系统便主要选择此时控制器认为的最优策略对控制系统进行调整. 而在训练过程趋于结束, 评价网络所能做出的决策已经十分成熟后, ε将退火到零, 则决策网络将采取确定的决策策略, 选取控制器认为最优的策略. ε-greedy策略具体表示如下:

$$a_k = \begin{cases} \text{随机动作} a \in A, & d < \varepsilon, \\ \arg\min\limits_{a_k} \hat{Q}(s_k, a_k), & d \geqslant \varepsilon, \end{cases} \tag{8}$$

其中: a为当前控制器所做出的动作; 动作集合A为一个集合, 集合中的元素为控制器所有可能选取的策略.

与立即回报函数$Q(s_k, a_k)$相区别, $\hat{Q}(s_k, a_k)$为评价网络对当前状态为s_k, 当前策略为a_k情况下奖励值的预估.

3.2.2 增量型策略

容错控制器的输出A_k为系统的重构方案, 对于本文中的实验平台, 根据前期的实验分析, 其可用的重构方案由不同的故障决定, 数量一般只几种到十几种不等, 并且容错效果较好. 因此, 本文参考文献[33]的方法, 可以通过枚举的方法将有限的配置方案进行增量式的累加更新, 从而更新得到最优策略输出. 本节将介绍控制系统所采用的配置方案的设计:

$$\Omega = \{\Lambda_0,\ \Lambda_1,\ \Lambda_2,\ \Lambda_3,\ \Lambda_4\}, \tag{9}$$

其中集合中的Λ_n为系统中可选的配置方案. 如表1所示, 动作集合中的策略均为增量型, 分为反向与正向, 控制器通过补偿的方法进行系统的容错控制. 控制器选取了相应的动作之后将此策略叠加到当前的动作值当中, 并针对传感器故障与执行器故障设计了具体的策略, 在本文中增量型策略采取步进值为±0.0002, 精度取决于传感器和执行器的性能指标并且决定了控制器训练收敛的速度和策略最优的准确性.

表 1 动作A集合中的动作方案

Table 1 Strategy scheme in action A set

配置方案	方案状态向量	配置对象
Λ_0	$[0,0,0,0]$	系统正常, 无动作
Λ_1	$+[0,0.0002,0,0]$	补偿执行器故障的正向动作
Λ_2	$-[0,0.0002,0,0]$	补偿执行器故障的负向动作
Λ_3	$+[0,0,0.0002,0]$	补偿传感器故障的正向动作
Λ_4	$-[0,0,0.0002,0]$	补偿传感器故障的负向动作

4 强化学习主动容错系统结构及算法实现

4.1 评价网络的实现

本文方法采用神经网络对系统Q-learning函数进行建模拟合, 计算出Q-learning函数的近似值, 神经网络的输入为$X = [S\ \ A]^{\mathrm{T}}$, 并采取梯度下降法[34]对其进行权值更新, 输出层权值的更新法则如下:

$$\Delta W = \eta[-\frac{\partial E}{\partial W}], \tag{10}$$

$$W = W + \Delta W, \tag{11}$$

其中: 其中函数E是在当前状态和容错策略下所得到的奖励函数reward值与当前系统状态的下一个状态得到的奖励函数reward值的误差, η为训练学习率, 通过梯度下降算法对误差函数E进行微分并且逐层向上传递, 并通过公式(11)此方法实现评价网络权值的更新.

如图1所示, 评价网络的结构为含有隐藏层的多层神经网络, 相比于传统的基于模型的容错控制器, 评价网络的思想是将控制系统实时产生的状态数据进行采集, 并通过神经网络的方法自动提炼出系统数据当中的特征来对系统状态进行辨识, 而图1中的输出y便是神经网络对故障辨识后所得出的策略结果. 本文中评价网络通过强化学习Q-learning算法对评价网络进行训练与更新, 将采集得到的系统状态值和容错控制策略作为评价网络的输入样本得到当前状态与所选择策略的奖励值作为容错控制器选取策略的依据, 进而选取最优策略交给执行网络进行下一步操作.

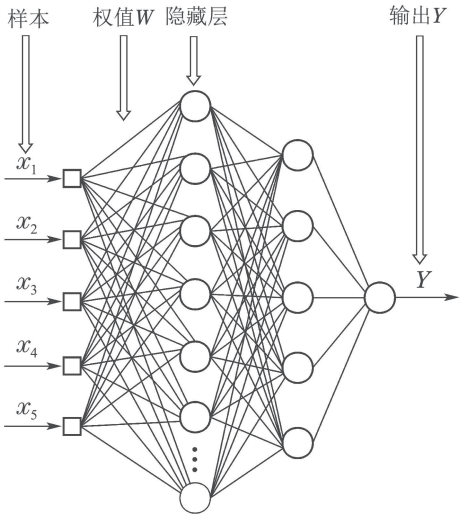


图 1 容错控制器评价网络结构图

Fig. 1 Fault tolerant controller evaluation network

4.2 执行网络的实现

执行网络的输出A为系统的决定的容错方案, 对于飞行控制系统, 其配置方案可由其系统状态量进行预估, 在状态S下, 动作A由以下步骤实现:

步骤 1 对于系统的容错调整方案, 通过评价网络计算对应的不同动作a_k所对应的奖励$\hat{Q}(s_k, a_k)$;

步骤 2 输出当前评价网络预估得到的容错调整方案a_t:

$$a_t = \arg\min_{a_k} \hat{Q}(s_k, a_k). \tag{12}$$

上述步骤是在评价网络训练结束之后对执行网络的步骤的描述.

4.3 状态转移预测网络

由于本文研究对象为连续的飞行控制系统, 而强化学习算法评价网络的更新学习过程中, 需要得到当前状态下所作出策略而得到的下一个状态值, 所以对于本文中的控制器而言, 通过当前的状态与动作预估出所得到的下一步状态是十分重要的. 所以针对这一需求, 本文提出了一种状态转移预估网络的方法, 其结构类似于评价网络的多层神经网络, 对输入样本为状态与策略, 输出为此状态与策略所得到的下一个状态进行拟合训练, 通过当前的状态动作对, 预估出下

一步所得到的状态. 状态转移预测网络的输入$X=[S_{\rm current}\ \ A_{\rm current}]^{\rm T}$, 输出$Y=[S_{\rm next}]^{\rm T}$. 整个训练过程是将历史的系统状态与动作存储到计算机, 然后每隔固定的迭代次数将存储的数据作为数据集对状态转移预测网络进行评估.

4.4　容错控制器迭代学习算法

基于强化学习的主动容错控制器的结构如图2所示, 其中: $S_{\rm current}$是系统当前在动作$A_{\rm current}$下通过传感器观测到的状态值; $A_{\rm current}$是对控制系统控制器做出调整的重构信号, 并不是系统的输入; $U(S_{\rm current}, A_{\rm current})$为系统的实际输出与期望输出之间的误差. 评价网络通过每一轮更新误差来进行重新训练, 由于神经网络训练往往会陷入局部极小值, 进而导致训练误差越来越大, 所以采用ε-greedy因子对神经网络的训练过程进行优化. ε-greedy因子表示系统在决策过程中选择随机动作的概率. 在训练的过程中, 将ε从1慢慢退火到0, 在训练的初期, 系统更倾向于去不断地“试错”, 选择随机的策略, 到了后期, 系统开始慢慢依赖评价网络所做出的评估, 整体的训练算法如下:

1) 初始化Q-learning评价网络, 神经网络随机初始化参数, 神经网络的输入数据为系统的状态和当前采取的动作序号, 输出为当前状态与在此状态下采取动作所获得reward奖励值$U(S_{\rm current}, A_{\rm current})$;

2) 初始化历史状态存储单元, 用来存储每个步长的状态–动作对;

3) 通过控制系统历史运行的数据对状态转移预测网络进行隔代更新, 用来通过当前的状态和动作值来预测下一个状态值;

4) 对每个时刻$t=1,2,\cdots,n$:

① 采集到当前状态S_t: 以ε的概率在所有动作集合中随机选择动作, 以$(1-\varepsilon)$的概率选择使reward值(在本文中就是系统的实际输出与期望输出之间的误差)最大化的动作$A_t=\arg\max\limits_a Q(s_t)$, 记当前的状态$S_t$和动作$A_t$值所得的奖励为$U(S_{\rm current}, A_{\rm current})$;

② 通过①步中的$S_{\rm current}$, $A_{\rm current}$和状态转移预测网络得到所得到的下一个状态$S_{\rm next}$, 求出当状态$S_{\rm next}$下预估得到最大奖励的$A_{\rm next}$, 并将当前$S_{\rm current}$, $A_{\rm current}$ 存储到集合Q;

③ 然后得到最新的奖励真实值

$$Q'(S_{\rm current}, A_{\rm current})=Q(S_{\rm current}, A_{\rm current})+\lambda\hat{Q}(S_{\rm next}, A_{\rm next});\quad(13)$$

④ Agent采取增量型动作$A_{\rm current}$叠加到现有动作中输入到系统中去;

⑤ 选择每100步通过存储在历史状态存储单元中的数据对状态转移预测网络进行更新, 更新误差函数E为当前状态和容错控制策略所得到奖励的真实值$Q'(S_{\rm current}, A_{\rm current})$与评价网络预估值$\hat{Q}(S_{\rm current}, A_{\rm current})$之差:

$$E=Q'(S_{\rm current}, A_{\rm current})-\hat{Q}(S_{\rm current}, A_{\rm current}).\quad(14)$$

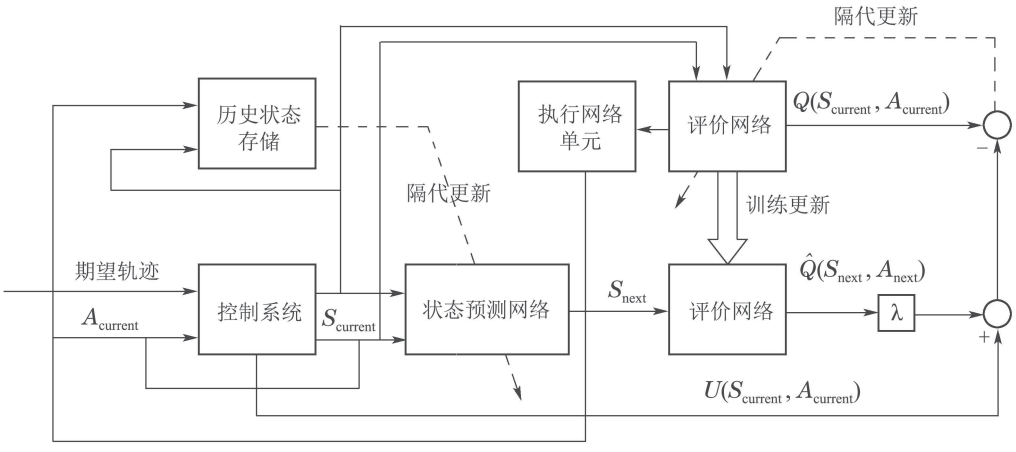


图 2　主动容错控制器结构图

Fig. 2　The structure of active fault tolerant controller

5　仿真验证

5.1　仿真平台介绍

本篇论文通过南京航空航天大学“先进飞行器导航、控制与健康管理”工信部重点实验室的飞行器故障诊断实验平台进行仿真与实物验证, 飞行控制仿真平台可以实现飞行器姿态角的测量、飞行控制系统算法和飞行姿态故障诊断算法的实验与研究, 并且具有控制系统设计开发, 硬件测试和软件仿真等功能. 仿真平台详细结构见图3所示.

此仿真实验平台利用角动量的原理来平衡和保持

姿态的方向角. 在仿真系统内侧存在着一个惯量盘, 并安装于图中1处万向节内, 1处万向节安装在图中两处万向节内部, 并通过一个方形支架固定在外壳上, 可以沿着竖直方向任意角度转动. 每个万向节通过滑环进行连接, 从而使3个轴方向上可以自由地绕轴向转动. 仿真平台通过机械方式使得各轴方向可以独立旋转, 同时仿真平台的各个方向都配备了高精度的光电编码器, 编码精度(倍频后)不低于4000 count/rev, 用来精确的测量姿态角等数据, 而本片论文中所需的数据将通过仿真平台内置的传感器进行读取.

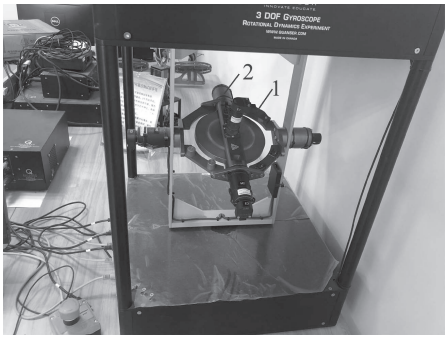


图 3 飞行器故障诊断实验平台实验系统

Fig. 3 Aircraft fault diagnosis experimental platform

5.2 实验平台参数介绍

5.2.1 实验平台模块介绍

本篇论文所采用的平台通过MATLAB软件将硬件平台与Simulink软件仿真平台进行连接, 通过配备在硬件实物平台中的精密传感器通过仿真平台中的硬件在回路(hardware-in-loop, HIL)模块将实时的平台数据传入仿真平台中的控制系统模块中进行软件与硬件的同步运行, 当运行仿真平台时, HIL模块、增益放大器、频率滤波器和电源模块在内的全部模型便同时启动, 仿真系统模型中的参数细节如图4和图5所示.

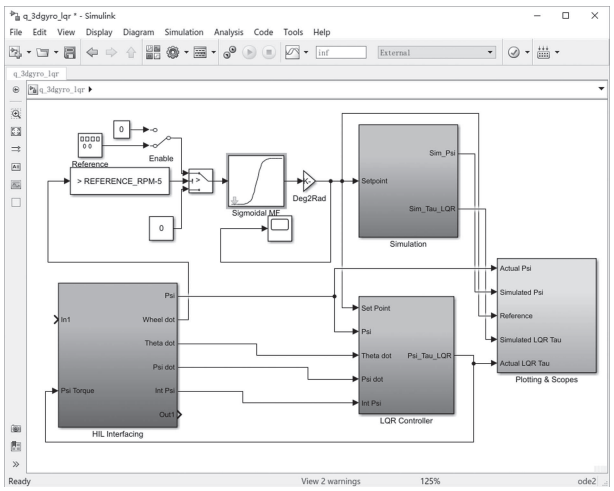


图 4 Simulink控制系统模块界面

Fig. 4 The interface of Simulink control system module

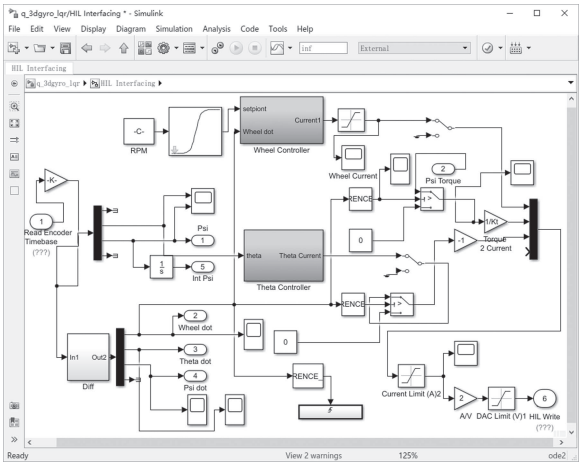


图 5 Simulink系统硬件在回路模块界面

Fig. 5 The interface of Simulink system HIL module

5.2.2 实验平台控制系统模型介绍

实验平台中系统模型的参考坐标系如图6所示.

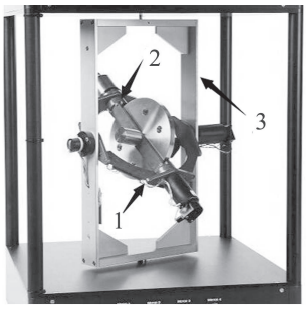


图 6 实验平台参考坐标系

Fig. 6 Reference coordinate system of the experimental platform

如图6所示, 图中箭头1所指的万向节表示飞行器的俯仰角, 图中箭头2所指的万向节表示飞行器的滚转角, 角速率运动方程式为

$$J_{\rm y}\ddot{\phi}+h\dot{\theta}=\tau_{\rm y}, \tag{15}$$

$$J_{\theta}\ddot{\theta}+h\dot{\phi}=0, \tag{16}$$

式中: $J_{\phi}=0.0036\,\mathrm{kg}\cdot\mathrm{m}^2$, $J_{\theta}=0.0226\,\mathrm{kg}\cdot\mathrm{m}^2$, $h=0.44\,\mathrm{kg}\cdot\mathrm{m}^2/\mathrm{s}$; J_{ϕ}表示滚转轴的转动惯量; J_{θ}表示俯仰角的转动惯量; h是通过平台硬件转子的惯性矩阵及其速度计算得到. 图6中箭头3所指的外部矩形框架下的驱动轴是滚转轴, 而$\tau_{\rm y}$是系统中的控制输入在滚转轴施加的扭矩. 实验平台的控制系统的线性状态方程表示为

$$\begin{cases}\dot{x}(t)=Ax(t)+Bu(t),\\ y(t)=Cx(t)+Du(t),\end{cases} \tag{17}$$

其中: $x(t)$表示系统状态量, $u(t)$表示系统的控制输出. 定义状态和输出如下:

$$\begin{cases}x^{\rm T}=[\dot{\phi}\ \ \theta\ \ \dot{\theta}\ \ \int\theta],\\ y=\theta,\end{cases} \tag{18}$$

其中ϕ, θ分别为实验平台飞行器的滚转角和俯仰角. 本篇论文通过对控制系统中的硬件在回路模块中进行故障注入, 并通过平台硬件传感器采集到实验平台在故障与非故障情况下的系统的状态数据. 从而通过采集到的数据对强化学习评价网络进行训练, 使容错系统网络参数能够满足容错目标.

5.3　执行器故障数字仿真与物理平台实验验证

在$t = 8\,\mathrm{s}$时刻, 向飞行控制系统平台的硬件在回路模块中注入如下执行器偏置故障:

$$f_{\mathrm{a}}(t) = \begin{cases} 0, \ t < 8\,\mathrm{s}, \\ 5, \ t \geqslant 8\,\mathrm{s}. \end{cases} \tag{19}$$

图7分别为系统发生故障时容错控制器的数字仿真容错效果曲线和容错过程中控制输出变化曲线. 从容错效果曲线图中可以看出当故障发生后, 容错控制器克服了故障产生的影响, 在$t = 8.5\,\mathrm{s}$左右, 系统姿态在发生故障的俯仰角上已经基本恢复稳定.

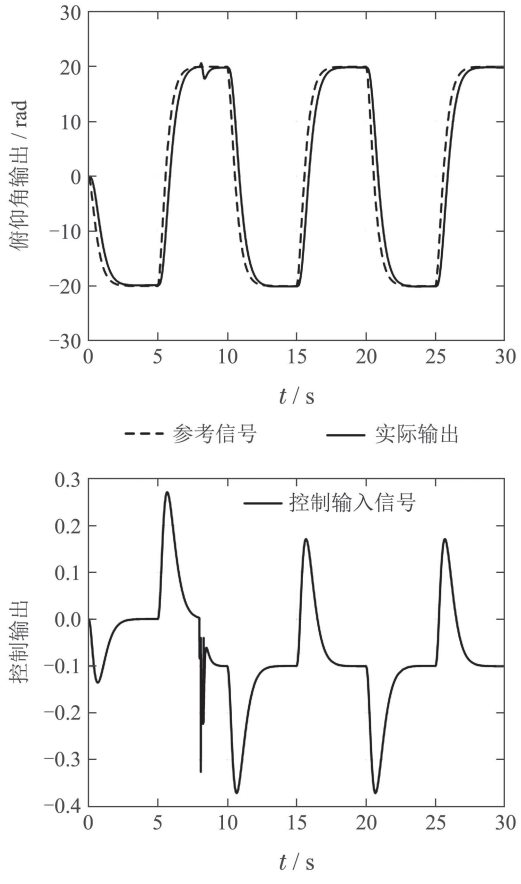


图 7　数字仿真容错过程控制系统容错效果和反馈输出变化(执行器故障)

Fig. 7　Fault-tolerant effects and feedback output changes of digital simulation fault-tolerant process control systems (actuator failure)

从图8的评价网络训练曲线上可看出评价网络不同于普通的神经网络训练网络的递减趋势, 因采用ε-greedy策略选取方法, 策略网络会以一定概率去做探索, 所以网络的训练误差曲线会呈现出不稳定的状态.

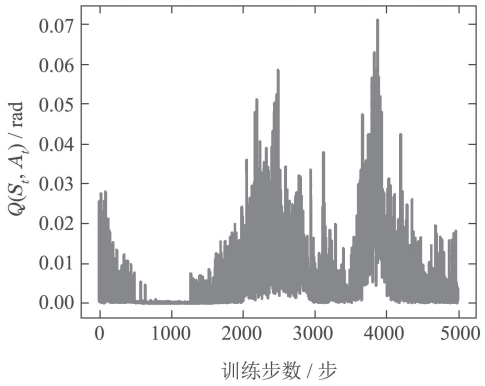


图 8　强化学习控制器评估网络的误差训练曲线

Fig. 8　The training curve of reinforcement learning controller evaluation network error

通过数字仿真验证了实验方法的可行性之后, 将本篇论文中的方法运用在物理仿真平台上, 强化学习的评价网络参数由数字仿真所得到的结果提供, 具体的效果图如图9和图10所示, 可以看出当系统发生故障后系统姿态发生偏移, 此时通过评价网络得出了最优的容错控制方案, 并根据此时评估得到的容错控制方案对系统进行容错控制, 可以看出系统姿态基本恢复正常, 并能够跟踪目标轨迹.

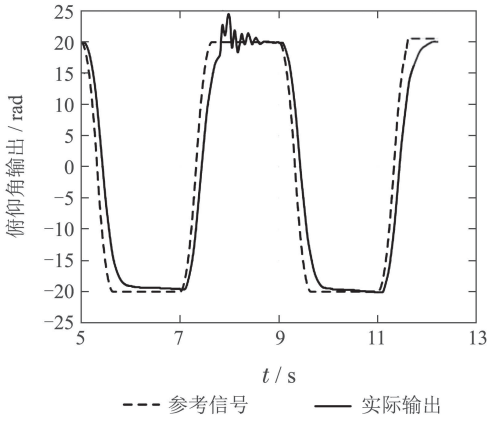


图 9　物理平台执行器故障容错俯仰角输出

Fig. 9　Pitch angle output during fault tolerance of physical platform actuators

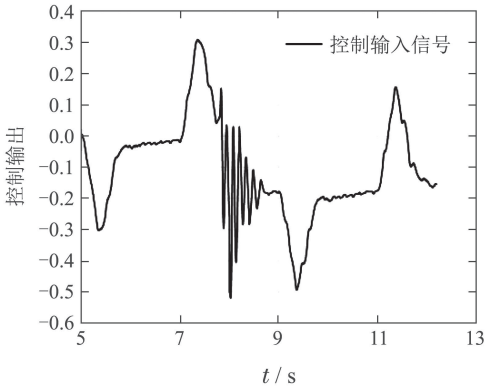


图 10　物理平台容错控制反馈输出

Fig. 10　Control feedback output in the physical platform fault tolerance process

如图11所示, 相较于本文中的方法, 传统的强化学习算法由于确定性策略的局限性, 只能通过固定数量的策略进行策略的选择, 导致实际容错效果与理想效果有较大误差.

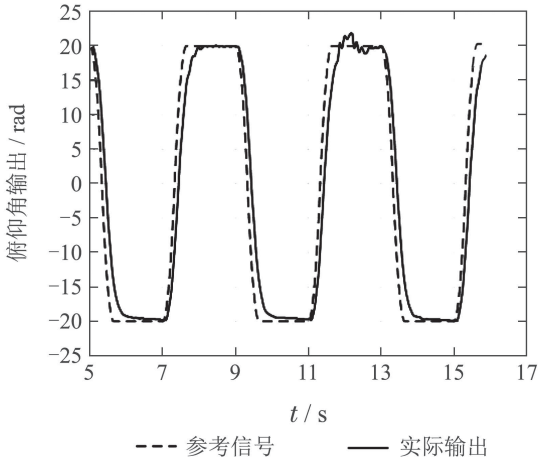


图 11 传统强化学习方法下执行器故障容错俯仰角输出

Fig. 11 Pitch angle output in actuator fault tolerant process under traditional reinforcement learning method

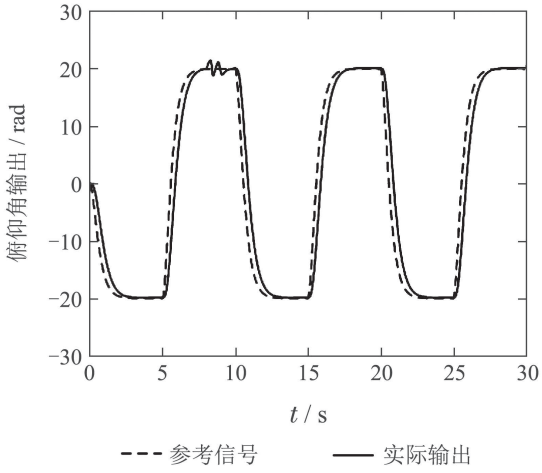
5.4 传感器故障数字仿真与物理平台实验验证

在$t = 8\,\mathrm{s}$时刻, 向飞行控制系统中注入如下传感器故障:

$$f_{\mathrm{s}}(t) = \begin{cases} 0, & t < 8\,\mathrm{s}, \\ -5, & t \geqslant 8\,\mathrm{s}. \end{cases} \tag{20}$$

图12展示了本文介绍的主动容错控制器在发生传感器故障后数字仿真的效果图. 从图12可以看出, 在传感器故障发生之后, 在$t = 8.2\,\mathrm{s}$左右, 系统姿态已经基本恢复稳定.

将通过数字仿真验证可行后的网络参数运用到实际物理平台上的仿真效果如图13和图14所示, 可以看出本篇论文中的方法对于实物平台也具有良好的容错效果. 而如图15所示, 运用确定性策略的传统强化学习方法, 容错效果并不理想.



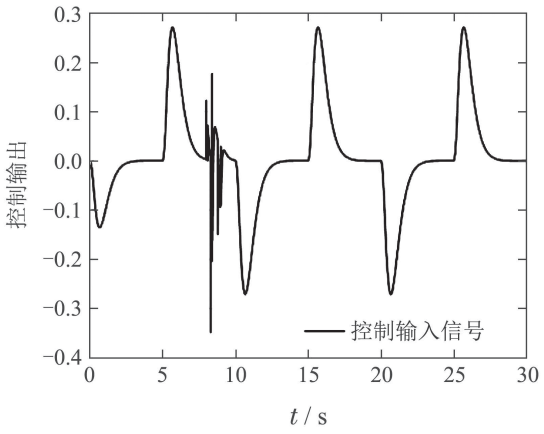


图 12 数字仿真容错过程控制系统容错效果和反馈输出变化(传感器故障)

Fig. 12 Fault-tolerant effects and feedback output changes of digital simulation fault-tolerant process control systems (sensor failure)

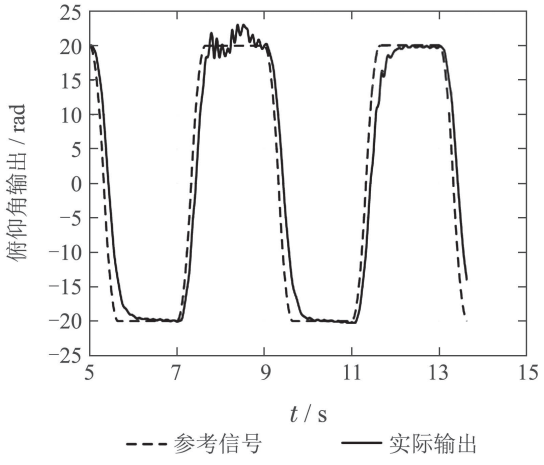


图 13 物理平台传感器故障容错俯仰角输出

Fig. 13 Pitch angle output during fault tolerance of physical platform sensors

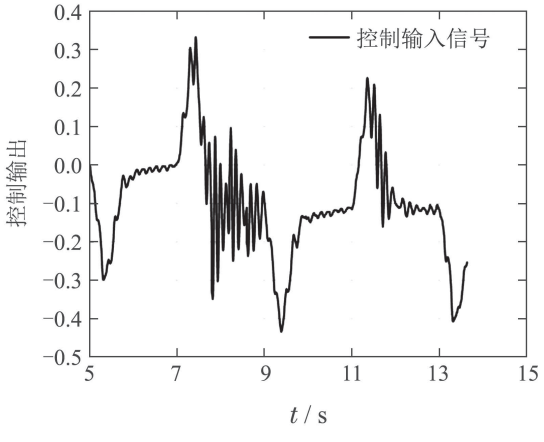


图 14 物理平台容错控制反馈输出

Fig. 14 Control feedback output in the physical platform fault tolerance process

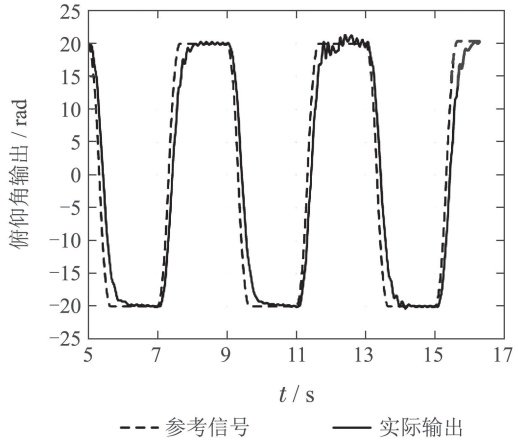


图 15 传统强化学习方法下传感器故障容错俯仰角输出

Fig. 15 Pitch angle output in sensor fault tolerant process under traditional reinforcement learning method

6 结论

本文基于强化学习的Q-learning方法, 提出了一种主动容错控制方案, 并且基于南京航空航天大学“先进飞行器导航、控制与健康管理”工信部重点实验室的飞行器故障诊断实验平台对所提出的方法进行仿真验证. 仿真结果表明, 本文所提出的增量式策略强化学习容错控制算法, 能够在飞行器发生故障时尽可能地保持原有飞行轨迹, 完成原定任务, 本篇文章主要创新点如下:

1) 通过强化学习控制器, 采用评价网络对系统产生的实时数据进行特征的提取, 从而获取故障信息并基于此做出对系统控制器的调整, 相比于传统的基于模型的容错控制方法, 此方法是一种基于数据的主动容错控制方法, 突破了复杂系统建模困难的局限, 并且通过对数据特征的提取替代了故障检测子系统, 简化了控制器的设计;

2) 对于不确定性故障的前提, 提出了增量型策略的强化学习控制器, 改进了传统强化学习算法中采用确定性的固定策略的局限, 从而实现对当前所产生故障系统的最优容错策略的逼近;

3) 通过状态转移预测网络进行下一步状态的预估, 实现了连续控制系统的实时策略网络的更新.

参考文献:

[1] PATTON R J. Fault-tolerant control systems: the 1997 situation. *Proceedings in IFAC Safe Process Conference*, 1997, 30(18): 1029 – 1051.

[2] WU H N, ZHANG H Y. Reliable fuzzy control for continuous-time nonlinear systems with actuator failures. *IEEE Transactions on Fuzzy Systems*, 2006, 14(5): 609 – 618.

[3] GOPINATHAN M, BOSKOVIC J D, MEHRA R K, et al. A multiple model predictive scheme for fault-tolerant flight control design. *IEEE Conference on Decision & Control*. Tampa, FL, USA: IEEE, 1998: 1376 – 1381.

[4] TAO G, TANG X, CHEN S, et al. Adaptive failure compensation of two-state aircraft morphing actuators. *IEEE Transactions on Control Systems Technology*, 2006, 14(1): 157 – 164.

[5] TANG X, TAO G, JOSHI S M. Adaptive actuator failure compensation for nonlinear MIMO systems with an aircraft control application. *Automatica*, 2007, 43(11): 1869 – 1883.

[6] PAOLI A, SARTINI M, STEPHANE L. Active fault tolerant control of discrete event systems using online diagnostics. *Automatica*, 2011, 47(4): 639 – 649.

[7] YANG H, COCQUEMPOT V, JIANG B. Robust fault tolerant tracking control with application to hybrid nonlinear systems. *IET Control Theory & Applications*, 2012, 3(2): 211 – 224.

[8] CHEN W, JIANG J. Fault-tolerant control against stuck actuator faults. *IEEE Proceedings–Control Theory and Applications*, 2005, 152(2): 138 – 146.

[9] FEKIH A. A robust fault tolerant control strategy for aircraft systems. *Control Applications & Intelligent Control*. Petersburg, Russia: IEEE, 2009: 1643 – 1648.

[10] TANG X, TAO G, WANG L, et al. Robust and adaptive actuator failure compensation designs for a rocket fairing structural-acoustic model. *IEEE Transactions on Aerospace and Electronic Systems*, 2004, 40(4): 1359 – 1366.

[11] YANG Huiliao, JIANG Bin, ZHANG Ke. Direct self-repairing control for four-rotor helicopter attitude systems. *Control Theroy & Applications*, 2014, 31(8): 1053 – 1060.
(杨荟憭, 姜斌, 张柯. 四旋翼直升机姿态系统的直接自修复控制. 控制理论与应用, 2014, 31(8): 1053 – 1060.)

[12] YANG H, JIANG B, COCQUEMPOT V. Supervisory fault tolerant regulation for nonlinear systems. *Nonlinear Analysis Real World Applications*, 2011, 12(2): 789 – 798.

[13] YANG H, STAROSWIECKI M, JIANG B, et al. Fault tolerant cooperative control for a class of nonlinear multi-agent systems. *Systems & Control Letters*, 2011, 60(4): 271 – 277.

[14] ETERNO J, WEISS J, LOOZE D, et al. Design issues for fault tolerant-restructurable aircraft control. *IEEE Conference on Decision & Control*. Ft. Lauderdale, USA: IEEE, 1985: 900 – 905.

[15] CHEN F, LU F, JIANG B, et al. Adaptive compensation control of the quadrotor helicopter using quantum information technology and disturbance observer. *Journal of the Franklin Institute*, 2014, 351(1): 442 – 455.

[16] YANG H, JIANG B, ZHANG K. Direct self-repairing control of the quadrotor helicopter based on adaptive sliding mode control technique. *The 2014 IEEE Chinese Guidance, Navigation and Control Conference*. Yantai, China: IEEE, 2014: 1403 – 1408.

[17] BOSKOVIC J D, MEHRA R K. Multiple model-based reconfigurable flight control system design. *IEEE Conference on Decision & Control*. Phoenix, USA: IEEE, 1999: 4503 – 4508.

[18] ZHANG K, JIANG B. Fault diagnosis observer-based output feedback fault tolerant control design. *Acta Automatica Sinica*, 2010, 36(2): 274 – 281.

[19] JIANG Bin, YANG Hao. Survey of the active fault-tolerant control for flight control system. *Systems Engineering and Electronics*, 2007, 29(12): 2106 – 2110.
(姜斌, 杨浩. 飞控系统主动容错控制技术综述. 系统工程与电子技术, 2007, 29(12): 2106 – 2110.)

[20] VEILLETTE R J, MEDANIC J B, PERKINS W R. Design of reliable control systems. *IEEE Transactions on Automatic Control*, 1992, 37(3): 290 – 304.

[21] ZHOU Donghua, DING X. Theroy and application of fault tolerant control. *Acta Automatica Sinica*, 2000, 26(6): 788 – 797.
(周东华, DING X. 容错控制理论及其应用. 自动化学报, 2000, 26(6): 788 – 797.)

[22] LIU Cong, QIAN Kun, LI Yinghui, et al. The integrated tracking fault tolerant controller design under actuator saturation with linear matrix inequality algorithm. *Control Theroy & Applications*, 2019, 36(1): 79 – 86.
(刘聪, 钱坤, 李颖晖, 等. 一体化执行器饱和线性矩阵不等式跟踪容错控制器设计. 控制理论与应用, 2019, 36(1): 79 – 86.)

[23] ABBEEL P, COATES A, QUIGLEY M, et al. An application of reinforcement learning to aerobatic helicopter flight. *International Conference on Neural Information Processing Systems*. Vancouver, British Columbia, Canada: MIT Press, 2006: 1 – 8.

[24] HUO Z H, YUAN Z, CHANG X. A robust fault-tolerant control strategy for networked control systems. *Journal of Network and Computer Applications*, 2011, 34(2): 708 – 714.

[25] ZHANG Y W, ZHOU H, QIN S J. Decentralized fault diagnosis of large-scale processes using multiblock kernel principal component analysis. *Acta Automatica Sinica*, 2010, 36(4): 593 – 597.

[26] LIU L, WANG Z, ZHANG H. Adaptive fault-tolerant tracking control for MIMO discrete-time systems via reinforcement learning algorithm with less learning parameters. *IEEE Transactions on Automation Science & Engineering*, 2017, 14(1): 299 – 313.

[27] YU Lingli, SHAO Xuanya, LONG Ziwei, et al. Intelligent land vehicle model transfer trajectory planning method of deep reinforcement learning. *Control Theroy & Applications*, 2019, 36(9): 1409 – 1422.
(余伶俐, 邵玄雅, 龙子威, 等. 智能车辆深度强化学习的模型迁移轨迹规划方法. 控制理论与应用, 2019, 36(9): 1409 – 1422.)

[28] NG A Y, COATES A, DIEL M, et al. Autonomous inverted helicopter flight via reinforcement learning. *International Symposium on Experiment Robotics*. Sant'Angelo d'Ischia, Italy: [s.n.], 2003: 799 – 806.

[29] WILLIAMS R J. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 1992, 8(3/4): 229 – 256.

[30] WANG Z, LIU L, ZHANG H, et al. Fault-tolerant controller design for a class of nonlinear MIMO discrete-time systems via online reinforcement learning algorithm. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2016, 46(5): 611 – 622.

[31] SILVER D, HUANG A, MADDISON C J, et al. Mastering the game of go with deep neural networks and tree search. *Nature*, 2016, 529(7587): 484 – 489.

[32] MNIH V, KAVUKCUOGLU K, SILVER D. Human-level control through deep reinforcement learning. *Nature*, 2015, 518(7540): 529 – 533.

[33] BOARO M, FUSELLI D, DE A F, et al. Adaptive dynamic programming algorithm for renewable energy scheduling and battery management. *Cognitive Computation*, 2013, 5(2): 264 – 277.

[34] WERBOS P. *Beyond regression: new tools for prediction and analysis in the behavioral sciences*. Cambridge, USA: Harvard University, 1974.

作者简介:

任 坚 硕士研究生, 目前研究方向为飞行控制系统的故障诊断与容错控制, E-mail: jeremyren95@163.com;

刘剑慰 副教授, 研究生导师, 目前研究方向为控制系统的故障诊断与容错控制, E-mail: ljw301@nuaa.edu.cn;

杨 蒲 副教授, 研究生导师, 目前研究方向为控制系统的故障诊断与容错控制, E-mail: ppyang@nuaa.edu.cn.