基于强化学习的宽速域冲压发动机燃烧室一维压力分布控制方法研究 *

聂聆聪[1]，牟春晖[1]，李帅衡[2]

（1. 北京动力机械研究所 冲压发动机技术全国重点实验室，北京 100074；
2. 北京京航计算通讯研究所，北京 100074）

摘 要：为提高冲压发动机燃烧室在宽速域范围内的性能，提出一种基于系数自适应调整熵正则化强化学习的宽速域几何可调燃烧室压力分布的控制方法，通过对燃烧室在一维流场上的压力分布监测与控制，实现该类燃烧室的宽速域高性能燃烧。本文采用熵正则化强化学习方法，利用滑块位移、两个喷注流量，实现对燃烧室沿程压力一维分布形状的优化控制；提出适用燃烧室多压力点控制的自适应温度系数调整算法，提高对压力分布一维控制算法的训练收敛速度，同时建立适用于一维压力分布控制的动作抖动惩罚函数及随机训练策略，解决了执行机构抖动、算法的泛化性及延时鲁棒性差等问题；通过数值仿真验证了算法的有效性，结果表明压力分布控制均方误差最大1.44%，超调量最大2.76%，调节时间不超过0.5 s，符合工程实际应用需求。

关键词：冲压发动机；宽速域燃烧室；压力控制；强化学习；最大熵正则化

中图分类号：V231.1 **文献标识码**：A **文章编号**：1001-4055（2025）06-2404056-11

DOI：10.3724/1001-4055. 2404056

1 引 言

为了保证在宽范围马赫数下发动机能够获得比较好的性能，燃烧室的几何结构需可调，因此几何可调燃烧室概念应运而生。美国空军研究实验室（Air Force Research Laboratory，AFRL）[1]在保持进出口面积不变的情况下，通过试验研究了来流马赫数3~5条件下两种构型燃烧室的性能，表明了几何可调燃烧室的优势；Bouchez等[2]得到了变几何燃烧室与无喉道的定几何超声速燃烧室的比冲随着来流马赫数的变化规律；在国内，国防科技大学的潘余等[3]研究了不同当量比下双模态超燃冲压发动机可变几何喉道对模态转换的影响，研究了几何喉道的大小对燃烧室点火和燃烧室性能的影响规律；西北工业大学陈玉春等[4]研究了下壁面可移动变几何燃烧室的性能，基于集总参数法建立了用于分析变几何燃烧室性能的计算模型，通过与两个固定几何发动机性能进行对比后表明，变几何燃烧室能够提高发动机的性能。此外，哈尔滨工业大学高超声速中心杨庆春[5]，从热力循环角度研究了不同飞行马赫数下超声速燃烧室性能的制约因素，通过对四种基本热力学过程分析后，确定了“等-扩-等”的燃烧室构型能够适应宽范围工作下对燃烧室几何和性能要求。

对于几何可调燃烧室，在仅依靠单一调节燃油当量比的基础上，增加了燃烧室扩张比这一调节变量，使燃烧室中的释热分布变化变成了一种主动的调节方式，从而满足不同来流条件下对发动机燃烧室的性能需求。同时，几何可调燃烧构型的设计综合考虑了不同飞行工况下的发动机工作模式特点，通过调节燃烧室扩张比，也可使之在各飞行工况下都能够实现可靠的稳定燃烧，并获得最优的性能。燃烧室的一维压力分布是表征该类燃烧室性能的重要特征，与传统单点压力或压比控制不同，需要通过多点燃料注入和几何结构调节共同实现一维压力分布。

国内外近年来针对深度强化学习的发动机控制已有部分研究。Singh等[6]首次将DQN（Deep Q Network）算法应用于航空发动机控制，其中DQN以价值

* **收稿日期**：2024-04-17；**修订日期**：2024-12-25。

通讯作者：聂聆聪，博士，研究员，研究领域为发动机控制。E-mail：nielingcong@163.com

引用格式：聂聆聪，牟春晖，李帅衡. 基于强化学习的宽速域冲压发动机燃烧室一维压力分布控制方法研究[J]. 推进技术，2025，46(6)：2404056.（NIE L C，MU C H，LI S H. One dimensional pressure distribution reinforcement learning control for ramjet combustor with wide range velocity[J]. Journal of Propulsion Technology，2025，46(6)：2404056.）

函数作为输出，间接将其作为行动选择的依据。郑前刚等[7]将深度增强学习（Deep Reinforcement Learning，DRL）算法引入到变循环发动机的寻优控制，提出了一种无模型、可在线执行的Q学习智能算法，取得比传统PID控制更好的控制效果，表明强化学习算法能够应用于变循环发动机的智能寻优控制。但Q学习方法只能处理离散的动作问题，不适用于连续动作空间探索的强化学习。而Actor-Critic框架结合了价值函数法和梯度法的优点，Actor直接将动作作为其输出，因此基于它的一系列算法，如深度确定性策略梯度算法（Deep Deterministic Policy Gradient，DDPG）和优势行为者-批评算法（Actor-2-Critic，A2C）更适合于航空发动机控制等连续动作空间问题[8-10]。刘智瑞[11]设计了涡扇发动机DDPG控制算法，并且链接涡扇发动机部件级模型动态链接库，对智能算法进行验证。董建华等[12]提出一种基于强化学习的航空发动机控制虚拟自学习方法，利用航空发动机的试验数据通过长短期记忆（Long Short-Term Memory，LSTM）神经网络建立虚拟学习环境，然后在虚拟环境中训练双延迟深确定性政策梯度（Twin Delayed Deep Deterministic Policy Gradient，TD3）算法以输出的燃油流量实现对发动机高压转子转速的控制，实现了超调量更小、调节时间更短的智能控制设计，但该方法仅在较少组别的燃油流量下进行训练，算法缺少泛化性。胡乾坤等[13]为了减轻控制变量突然变化引起的振荡，提出了基于相位的奖励函数。高文博等[8]引入动量项来抑制该振荡问题，引入误差积分项来改善强化学习控制器的稳态性能。陶博等[9]将深度强化学习应用于变循环发动机最小油耗性能的优化控制问题。朱逸阳等[10]采用长短期记忆递归神经网络设计推力估计器，并采用近端策略优化算法（Proximal Policy Optimization，PPO）实现航空发动机的直接推力控制。胡雪兰[14]分别采用脑情感学习算法和改进的确定性策略梯度算法进行了变循环发动机多变量控制规律设计，借鉴了深度确定性策略梯度算法的智能体结构和参数训练方法，加入目标神经网络结构用于维持神经网络学习过程的稳定性，加入样本优先回放算法以提高神经网络训练速度，防止出现过拟合。由上可知，目前基于深度强化学习的航空发动机控制研究仍处于起步探索阶段，通常解决的问题较简单，控制维度较小，主要用于发动机油耗优化和推力控制等领域，以及通过深度强化学习算法设计解决发动机实际控制过程中存在的一些问题，如动作震荡等。目前，尚缺乏基于深度学习的发动机燃烧场压力分布控制方面的研究，个别研究只在特定工况下对算法进行训练，缺少泛化性。

本文中的压力分布控制方法指的就是对几何可调燃烧室沿程压力的一维分布形状进行控制，通过在燃烧室壁面平均分布选取10个压力测点，对10个压力测点进行控制，考虑到几何可调燃烧室机理的复杂性以及多输入多输出间的强耦合性和强非线性，本文采用强化学习方法开展燃烧室压力分布控制研究。首先研究几何可调燃烧室构型与控制机理和设计思路，然后，开展适用于燃烧室一维压力分布控制的熵正则强化学习算法研究，通过设计合理的奖励函数和训练环境，使得智能体在与环境的交互中学习到最优的控制策略，最终实现燃烧室压力分布控制。最后，构建多种拉偏试验方案，通过数值仿真试验验证方法的有效性。

2　几何可调燃烧室构型与控制机理

所研究的宽速域几何可调燃烧室构型如图1所示[15]。该燃烧室构型由四部分组成，即隔离段I，扩张段II，等直段III以及内喷管IV。入口条件为马赫数2~7，燃烧室的扩张比为1.4~3.4，燃油当量比为0.4~1.2。喷油位置有两个，一个前支板喷油，另一个为后壁面喷油。燃烧室扩张比定义为$(h_1+h_2)/h_1$。对于低来流马赫数，当燃烧室处于亚燃模态时，为了避免热壅塞发生，需要将几何滑块向后移动，增加燃烧室的扩张比；对于高来流马赫数，当燃烧处于超燃模态时，需要将几何滑块向前移动，减小燃烧室的扩张比，从而提高燃烧性能[15]。本文在此基础上开展压力分布控制研究工作。

在给定发动机几何构型、来流条件和燃料注入条件的情况下，从流体力学的三大定律出发，考虑面积变化、燃料添加、变比热、壁面摩擦及混合效率建

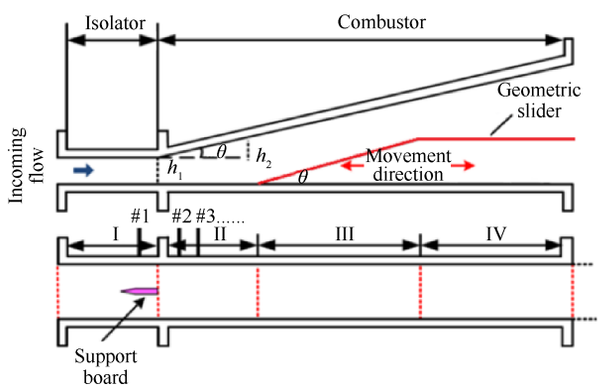

Fig. 1　Geometrically adjustable combustion chamber profile diagram

立准一维欧拉方程[15]，即

$$\frac{\partial \boldsymbol{Q}}{\partial t}+\frac{\partial \boldsymbol{F}}{\partial x}=\boldsymbol{S} \tag{1}$$

式中，

$\boldsymbol{Q}=(A\rho, A\rho u, A\rho E_{\mathrm{t}})^{\mathrm{T}}$为守恒型变量；

$\boldsymbol{F}=\left(A\rho u, A\rho u+A\rho u^{2}, A(\rho+\rho E_{\mathrm{t}})u\right)^{\mathrm{T}}$为无黏通量项；

$$\boldsymbol{S}=\left(\frac{\mathrm{d}\dot{m}_{\mathrm{s}}}{\mathrm{d}x}, p\frac{\mathrm{d}A}{\mathrm{d}x}-\frac{\rho u^{2}}{2}f\cdot C_{\mathrm{w}}+u_{\mathrm{sx}}\frac{\mathrm{d}\dot{m}_{\mathrm{s}}}{\mathrm{d}x}, H_{\mathrm{s}}\cdot\eta_{\mathrm{mix}}\cdot\frac{\mathrm{d}\dot{m}_{\mathrm{s}}}{\mathrm{d}x}\right)^{\mathrm{T}}$$

为源项。

式中A为发动机燃烧室横截面积；C_{w}为非圆形截面的湿周长，$C_{\mathrm{w}}=(4A/De)$；u_{sx}为燃料喷注速度在x方向的分量，本文采用垂直喷射的方式，所以此项为0；H_{s}为添加燃料的滞止焓和化学反应释放的热量之和，本程序取航空煤油分子式为$C_{10}H_{22}$，热值为42.5 MJ/mol。η_{mix}为混合效率；如果仅考虑燃烧释热过程，也可称为燃料的燃烧效率分布；f为壁面摩擦力系数；$\mathrm{d}\dot{m}_{\mathrm{s}}$为燃料添加项。

通过求解基于微分的一维气体动力学控制方程（连续性方程、动量方程、能量方程），获得发动机燃烧室内的流动参数沿发动机轴向的分布情况[15]，也即燃烧室沿程的压力分布输出。此外，几何可调燃烧室可调机构有两路燃油和一路改变扩张比的滑块位移，对执行机构进行了单项性能测试，其输入输出关系近似为一阶惯性环节，时间常数t分别为0.2 s和0.3 s。

燃烧室的一维压力分布是表征该类燃烧室性能的重要特征，为了提高宽速域范围内发动机性能，需要对燃烧室的壁面压力分布进行有效控制，传统控制方法可以通过燃烧室型面及喷注的燃油流量对壁面沿程的压力进行单点控制（可以选择2~3个压力测点），但无法做到壁面一维压力分布的整体塑形。主要存在两个问题，一是如果选择2~3个单点压力进行闭环控制，不同入口马赫数及不同飞行器推力需求下可观性较好的测点也是不同的，宽速域内被控压力测点选择困难；二是单点压力控制难以实现宽速域下的压力分布综合均方差指标，导致非受控压力点误差较大。所以考虑采用强化学习算法解决，强化学习算法可以通过奖励函数进行塑形实现此类问题的综合寻优控制。本文采用强化学习方法对壁面一维压力分布测点进行整体优化，为了测量在轴向上压力分布形状，需要在轴向布置多个传感器，传感器测点密集对算法精度更优，但测点过多在使用上会对成本及结构安装性产生影响，考虑数量与成本的折中，本文选择10个压力测点，但该方法可推广到更多的测点。以10个压力测点的均方差作为优化目标，基于Actor-Critic架构训练随机策略深度强化学习算法。根据发动机当前状态产生燃油当量比（E_{R}）、前后两个喷油孔的出油比（b_{ili}）、滑块位移（f_{Dd}）的控制动作，经过变换后与来流条件总温（T_{t}）、总压（p_{t}）、马赫数（Ma）一起输入到发动机燃烧室仿真环境，输出10个燃烧室压力点真实值，定义10个压力测点值为p_1~p_{10}，并根据发动机状态控制需求给定燃烧室压力目标值，计算真实值与目标值之间的均方差，结合惩罚策略作为奖励函数反馈给智能体，智能体根据反馈与当前状态指导下一时刻的动作生成，实现燃烧室压力分布自动控制。图2为宽速域燃烧室控制的总体原理图，其中的智能体拟采用强化学习软演员评论家（Soft Actor-Critic，SAC）算法，通过该算法实现智能体的训练与燃烧室一维压力分布控制。

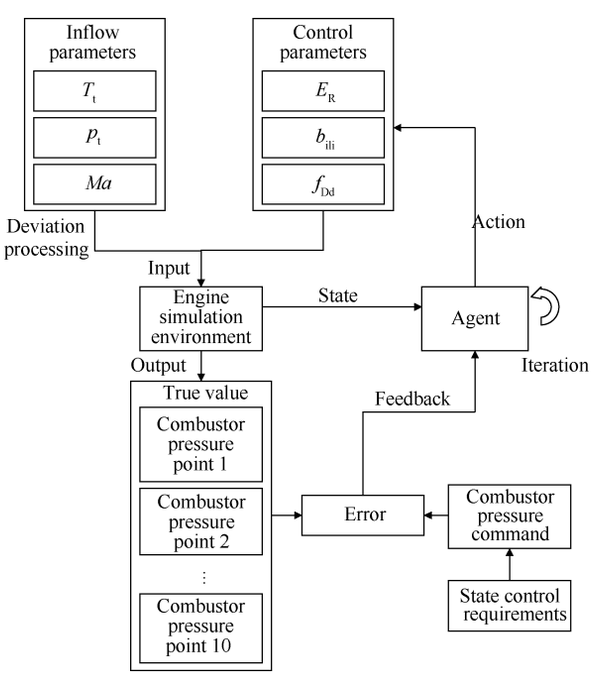

Fig. 2　Principle of wide speed domain combustion chamber control

3　SAC架构下的宽速域燃烧场压力分布控制算法设计和改进

发动机燃烧场压力分布控制属于高维状态、连续动作控制问题，并且要求算法具有较强的泛化性和鲁棒性，这在过去的研究中是少见的。本文基于SAC算法架构开展燃烧场压力分布控制研究，该算法是面向最大熵强化学习开发的一种off-policy强化学习算法[16]，具有对超参数敏感度低、超参数数量少、基于随机策略、收敛效果较好等特点。相比于确定

性策略，随机策略可以找到通向目标点的不同控制路线，具有更强的探索能力，更容易在复杂问题下找到更好的答案，同时算法具有更强的鲁棒性，面对干扰更容易做出调整。

3.1　燃烧室压力分布控制算法设计

该算法基于Actor-Critic架构，其决策分为两个阶段：评估和执行。评估阶段由值函数（Critic）负责，它估计当前状态的价值，并将其反馈给策略函数（Actor）。策略函数使用该信息来生成下一个动作，并执行该动作以更新环境状态。更新后，算法会得到一个新的状态，这样就形成了一个反复循环的过程。策略函数和值函数使用双Q网络。相比于传统的单网络结构，双Q网络拥有Q1和Q2两个神经网络，分别用于估计当前状态下执行某个动作可以得到的累积回报。在每次训练过程中，Q1和Q2网络会分别估计当前状态和动作的Q值。这样做的好处是可以减小训练过程中的方差，提高算法的稳定性和泛化能力。同时对每个Q网络配置一个目标价值网络，这是两个Q网络的副本，但是其参数更新速度较慢，用于计算目标值，目标网络用于稳定学习过程，防止参数更新过快导致训练不稳定。除了拥有两个网络外，双Q网络还有一个特点是使用不同的随机权重初始化，这样可以让两个网络学习到不同的策略，进一步增强算法的稳定性和泛化能力。另外，在每次训练过程中，SAC算法会使用平均化策略来选择两个Q值函数中的较小值作为最终的Q值估计，这种平均化策略能够减小估计偏差，提高算法的精度和稳定性。

具体算法框架如图3所示。算法中使用1个策略网络和4个价值网络，其中包含2个当前价值网络和2个目标价值网络。策略网络的主要功能是根据当前状态生成合适的控制动作。它将当前状态作为输入，通过学习历史经验，确定在给定状态下应该采取的动作，以最大化预期奖励。策略网络的输出对应于控制参数的动作空间，它的优化目标是使生成的动作在当前状态下更有可能达到高奖励的状态。价值网络的主要功能是估计给定状态和动作组合的价值，即在特定状态下执行某个动作的预期累积奖励。算法中存在四个价值函数，它们分别对应不同的价值估计。价值反映了在当前状态下，执行某个动作的长期累积奖励，从而用于评估策略的效果。优化价值网络的目标是逐步逼近实际奖励信号，以准确地指导策略的调整和优化。策略网络和价值网络都设计为128×128的两层全连接神经网络。训练完成后只需要应用策略网络对当前状态进行决策，产生控制动作。

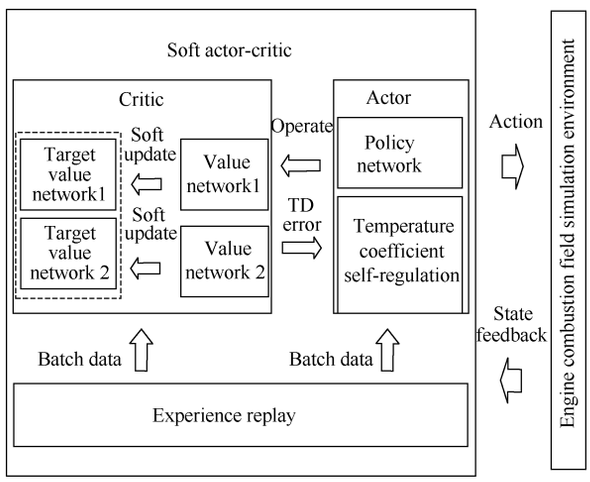

Fig. 3　SAC algorithm framework for pressure distribution control in wide speed domain combustion chamber

3.2　考虑执行机构延时和抖动的改进设计

动作空间、状态空间、奖励函数的选择是强化学习算法的关键。在此框架下，需要针对宽速域燃烧室的特性，对动作空间、状态空间、奖励函数进行合理的设计。宽速域燃烧室需要控制的三个参数为E_{R}，b_{ili}和f_{Dd}，因此动作空间设置为3维，即

$$\boldsymbol{a}=[E_{\mathrm{R}},b_{\mathrm{ili}},f_{\mathrm{Dd}}] \tag{2}$$

状态空间的设计需要涉及达成任务目标最为全面的信息，尽量包含相关性强的参数，同时也要排除相关性弱的参数以提高训练效率。对于宽速域几何可调燃烧室，任务目标是压力分布曲线与期望的分布曲线误差最小。与该目标强相关的参数有三类，一是最直接的沿程测点压力目标值与当前的误差量，这是任务目标的直接表征。二是发动机的来流状态会影响压力分布曲线形状。三是发动机的动作（供油量与几何型面）将决定各测点压力的大小，由于是动态的过程，实际发动机的动作到压力产生变化会有一个延时，需要包含之前时刻的动作信息。所以，燃烧室压力目标值、当前燃烧室状态真实值与目标值的误差都是必须包含的，同时要兼顾宽域工作状态，来流条件Ma，T_{t}，p_{t}也必须包含。实际系统控制器发出指令到执行器接收并执行有纯延时，纯延时会产生动作执行但压力输出不变的时刻，不利于训练的收敛。所以将延时的动作也引入到状态空间中。对于发动机的供油泵及滑块调节的作动筒，纯延时时间一般在一个周期，考虑到一定的余量，状态空间需要包含执行机构当前及之前时刻的动作，

主要包含数据传输延迟前两个周期的动作。状态空间如下

$$s=\begin{cases} e_1,\ e_2,\ e_3,\ e_4,\ e_5,\ e_6,\ e_7,\ e_8,\ e_9,\ e_{10}, \\ p_{i1},\ p_{i2},\ p_{i3},\ p_{i4},\ p_{i5},\ p_{i6},\ p_{i7},\ p_{i8},\ p_{i9},\ p_{i10}, \\ E_R,\ E_{R\_1},\ E_{R\_2},\ b_{ili},\ b_{ili\_1},\ b_{ili\_2},\ f_{Dd},\ f_{Dd\_1},\ f_{Dd\_2}, \\ Ma,\ T_t,\ p_t \end{cases} \tag{3}$$

式中p_{i1}~p_{i10}是燃烧室压力分布指令值，e_1~e_{10}为当前时刻燃烧室压力真实值与指令值的误差，都是10维的数组。E_R，b_{ili}和f_{Dd}分别为三个控制动作指令，同时状态空间中加入了最近两个时刻的动作E_{R_1}（其中_1表示前一时刻的燃油当量比），E_{R_2}（_2表示前两个时刻的燃油当量比，下同），b_{ili_1}，b_{ili_2}，f_{Dd_1}，f_{Dd_2}，以更好地应对数据传输延迟。所有状态空间中的变量根据其各自范围经过最大最小值归一化处理，每个变量的归一化处理公式方法如下

$$x'=\frac{x-x_{\min}}{x_{\max}-x_{\min}} \tag{4}$$

式中x为需要归一化的参数，x'为归一化后的参数，$x_{\max}$为需要归一化量的最大值，$x_{\min}$为需要归一化量的最小值。

奖励函数的设计要直接与期望的目标挂钩，可以将压力分布指令的均方误差作为奖励函数的主项，可以定义

$$z_b=\sqrt{\frac{1}{10}\sum_{n=1}^{10}\left(p_n-p_{in}\right)^2} \tag{5}$$

如果仅将z_b作为奖励函数，由于未对动作进行约束，供油与执行机构动作指令存在较大的抖动，可以通过在奖励函数中增加动作抖动惩罚项来抑制无效的抖动，让算法训练后更平滑。算法可以在“非理想”的方向上引入抖动惩罚项。基于此，设计的奖励函数如下

$$z_{b\_incre}=z_{b_{t-1}}-z_{b_t} \tag{6}$$

$$j_{itter}=\max(\mathrm{abs}(a_{t-1}-a_t)) \tag{7}$$

$$r_{eward}=\begin{cases} -z_{b_t}z_{b\_incre}>0 \\ -z_{b_t}-j_{itter}z_{b\_incre}\leqslant 0 \end{cases} \tag{8}$$

式中$z_{b_{t-1}}$表示上一时刻的z_b，a_{t-1}表示上一时刻的状态，z_{b_incre}为上一时刻与本时刻均方误差的差值，j_{itter}为上一时刻与本时刻动作的差值。如果z_{b_incre}为正，代表算法向着理想的方向进行控制，此时奖励值为尺度变化后的均方误差；如果z_{b_incre}为负，可以视为算法无意义的抖动，此时在原有奖励值的基础上增加动作变化惩罚项j_{itter}，以一定程度抑制动作抖动。

动作空间、状态空间、奖励函数选定后，最优策略的学习算法采用SAC算法[16]，相比于标准强化学习算法，SAC算法通过增加熵项来完善标准目标，熵表示最优策略的随机程度，这样最优策略的目标是在每个访问状态下最大化熵，智能体在每个时间步得到的奖励与该时间步策略的熵成正比。这将强化学习问题更改为

$$\pi^*=\arg\max_{\pi}\underset{\pi}{E}\left[\sum_{t=0}^{\infty}\gamma^t\left(R(s_t,a_t,s_{t+1})+\alpha H(\pi(\cdot|s_t))\right)\right] \tag{9}$$

式中s_t为当前时刻状态，s_{t+1}为下一时刻状态，π为策略，$R(s_t,a_t,s_{t+1})$为奖励，$H(\pi(\cdot|s_t))$为熵，γ为衰减系数，α为温度系数，它决定了熵项相对于奖励的相对重要性，从而控制了最优策略的随机性。这种目标策略鼓励更广泛的探索，放弃无前途尝试，同时还可以捕获多种接近最优行为模式，提高了学习速度。

3.3 温度系数自适应调整的改进设计

温度系数是SAC算法的关键参数，温度系数可以调整更新过程中熵与Q值的比例权重，直接决定了算法的收敛过程。当温度参数较高时，策略更注重探索和学习新的状态，而当温度参数较低时，策略更注重利用已知的信息来获取奖励。为了平衡探索和利用，将温度参数设计为自适应参数，自适应温度参数通过最小化Bellman误差来进行更新。Bellman误差是指预测的Q值与真实Q值之间的差异。通过使用两个Q值函数来评估策略，SAC算法可以计算出Bellman误差，并使用它来调整自适应温度参数。考虑到本压力分布问题有3个动作，温度系数的更新方式为

$$J_\pi(\alpha)=E_{\substack{a_t\sim\pi_t \\ s_t\sim D}}\left[-\alpha\ln\pi(a_t|s_t)-3\alpha\right] \tag{10}$$

通过动态地调整自适应温度参数，可以帮助平衡探索和利用之间的权衡，并促进智能体学习到更好的策略，提高了对压力分布一维控制算法的训练收敛速度（训练回合数降低了20%），SAC算法可以实现更好的性能，同时在连续动作空间中的探索也更加高效。

3.4 考虑泛化能力的设计策略

经验回放被用于训练Q值函数和策略。每次交互后，样本会被存储到一个经验池中。然后，SAC算法从经验池中随机抽取一批样本，并使用它们来训练Q值函数和高斯策略。

训练策略的设计也是实现强化学习算法的关键，要充分结合智能控制的使用环境设计训练工况。对于宽速域燃烧室一维压力分布智能控制，需要考

虑来流条件的多工况，压力测量的误差，执行机构动态特性偏差，系统的纯延时。其中执行机构的动态特性偏差主要指供油装置和型面调节装置的数学模型与实际产品在动态特性上的误差，主要是实际产品在响应时间上对设计值的偏离。系统的纯延时指从发动机控制器指令的发出到执行机构动作，再到压力开始变化，整个过程的纯延时主要有通讯延时、执行机构响应的纯延时、燃烧反应及压力波传递等。发出指令到执行器接收并执行有 1 个控制周期纯延时，执行机构纯延时在 5~10 ms（小于 1 个发动机控制周期），燃烧反应及压力波传递的延时可以忽略，整体上留出余量可考虑有两个控制周期的纯延时。

所以，不确定性包括了来流工况、压力分布指令、压力测量、供油与执行机构动态，将这些不确定性参数作为随机变量，生成随机训练环境进行批量化的动态数据训练策略。考虑来流工况、压力测量值与压力分布指令随机拉偏 5% 生成随机环境。通过这种方法可以提高深度强化学习算法的泛化性。

下面将通过上述算法开展压力分布策略训练及数值仿真验证。

4　仿真验证与分析

采用上述基于强化学习的压力分布控制算法，针对宽速域几何可调燃烧室进行数值仿真研究。SAC 算法网络配置见表 1 所示。

选取了燃烧室沿程的 10 个压力测点为被控量，燃烧室扩张比、燃油当量比、燃油分配比例为控制量，基于深度强化学习算法进行控制，为使得仿真试验环境近似真实发动机试验环境，同时考虑一定的加严考核，仿真试验过程中增加了两个周期的纯延迟。图 4 是隔离段入口 Ma3.1、总温 1 860 K、总压 1.90 MPa 下由一种压力分布形状到另一种压力分布形状的控制效果，其中压力的均方差如图 5 所示，两个压力分布的稳态均方误差分别为 0.21% 和 0.83%，其中图中第 5 与第 6 个压力测点误差较大，主要是由于控制量少于被控量情况下难以训练到每个测点最优，第 5 与第 6 个压力点压力较高，绝对误差会偏大。图 5 中第 15 s 为压力分布控制指令的改变时刻，指令为阶跃变化，所以在 15 s 产生了较大的均方误差，SAC 算法根据误差变化进行调节动作，0.5 s 内误差再次被调节减小（达到下一个压力分布稳态）。图 6 为其中 5 个压力测点的动态调节过程，可以看出，基本无超调且调节时间优于 0.37 s。

Table 1　SAC algorithm network parameters

Composition of SAC algorithm	Network parameters	Value
Sample pool	MEMORY	1×10^{6}
Rounds	Episode	5×10^{3}
Number of cycles	Ep_Steps	500
Batch size	Batch_Size	128
Actor	Input neuron	32
	Output neuron	3
	Number of neuron	[32,128,128,3]
	Learn rate	10^{-4}
	Discount factor	0.99
	Soft update coefficient	0.995
Critic	Input neuron	35
	Output neuron	1
	Number of neuron	[35,128,128,1]
	Learn rate	10^{-3}
	Discount factor	0.99
	Soft update coefficient	0.995

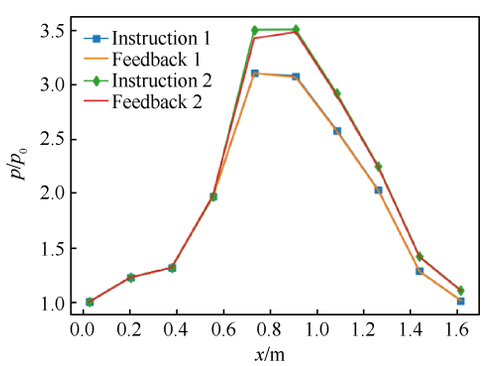

Fig. 4　Control curve of pressure distribution

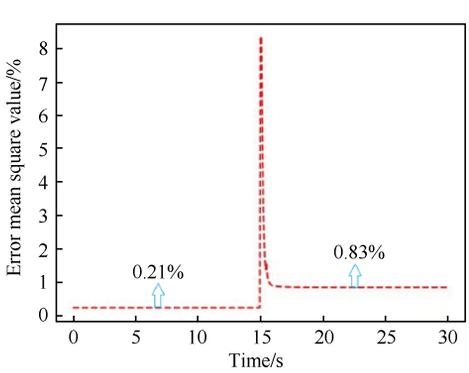

Fig. 5　Mean squared error in the process of controlling from one pressure distribution shape to another pressure distribution shape

图 7 为奖励函数不加动作抖动惩罚训练的执行机构情况，由图中可见，燃油当量比指令抖动较大。为了减小抖动，在训练过程增加了第 3 节的动作抖动惩罚项，训练后的执行机构动作效果图如图 8 所示。

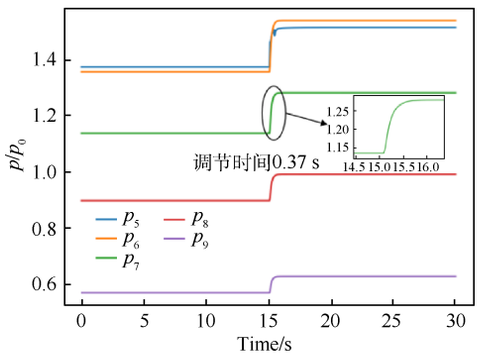

Fig. 6　Adjustment curves of 5 pressure(relative values) over time and their local magnified graphs

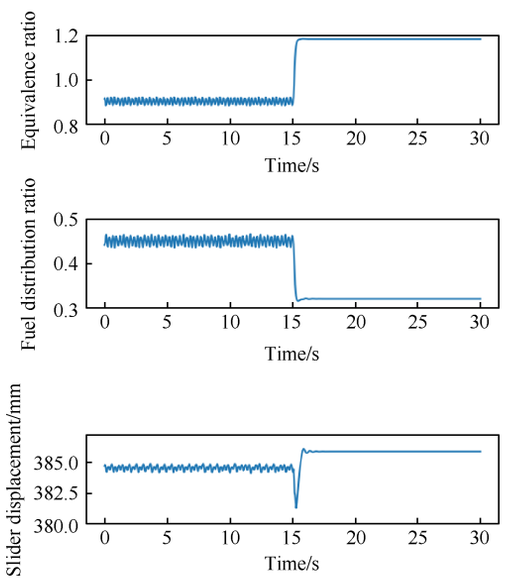

Fig. 7　Output of actor which reward function without jitter penalty

可以看到,该方法可以有效减小训练后控制器的执行机构抖动问题。

由于篇幅有限,其他状态点不同当量比调节下的压力分布均方误差及阶跃变化过程中的超调量均列入表2。主要对隔离段入口Ma2.5,Ma2.65,Ma2.8,Ma3.1下,每个马赫数条件选取三个状态点进行压力分布的动态控制试验,每个状态点选2或3条分布曲线进行调节。状态点通过不同的扩张比选取表征,每一个压力分布曲线通过燃油当量比表征。从表2数据可以看出,一维压力分布控制算法可以实现不同入口条件下燃烧室整个轴向的压力的塑形,几个入口马赫数条件下的压力分布的综合均方差最大1.44%,超调量也满足使用要求。

为了验证算法的适应性,对燃烧室入口条件进行拉偏。图9为对来流条件拉偏5%后的压力分布控制曲线,压力的均方差如图10所示,两个压力分布的

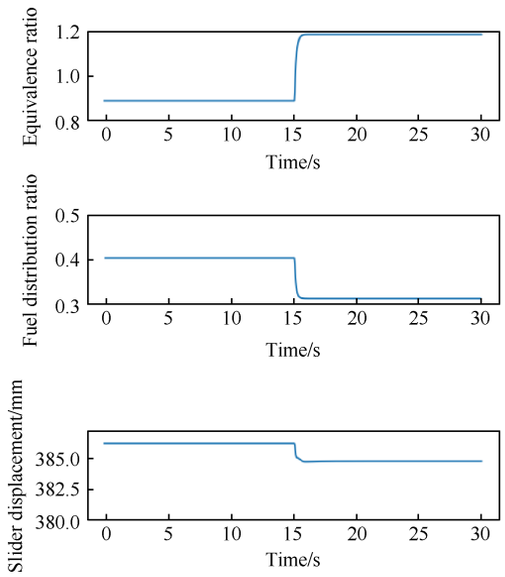

Fig. 8　Output of actor which reward function adding jitter penalty

Table 2　Mean squared error and overshoot of pressure distribution at each state point

Ma	State	Fuel equivalence ratio	Mean squared error/%	Overshoot/%
2.5	Expansion ratio1.4	0.8	1.01	1.09
		1.0	1.01	
	Expansion ratio 1.6	0.7	1.06	2.44
		1.0	1.02	
	Expansion ratio 1.8	0.7	1.08	0.02
		1.0	1.23	
2.65	Expansion ratio 1.4	0.8	1.02	1.89
		1.0	1.04	
	Expansion ratio 1.6	0.7	**1.44**	0.02
		1.0	1.02	
	Expansion ratio 1.8	0.8	1.03	0.02
		1.2	1.26	
2.8	Expansion ratio 1.4	0.7	1.78	1.8
		1.0	1.02	
	Expansion ratio 1.6	0.7	1.02	**2.76**
		0.9	1.05	
		1.2	1.08	
	Expansion ratio 1.8	1.0	1.12	0.02
		1.2	1.16	
3.1	Expansion ratio **1.4**	**0.9**	**0.21**	**0.02**
		1.2	**0.83**	
	Expansion ratio 1.6	0.9	1.39	1.05
		1.2	1.12	
	Expansion ratio 1.8	0.9	1.03	2.14
		1.2	1.12	

稳态均方误差分别为 0.45% 和 0.65%，图 11 为各压力测点的动态调节过程，可以看出，无超调且调节时间优于 0.41 s，图 12 为各执行机构响应曲线。

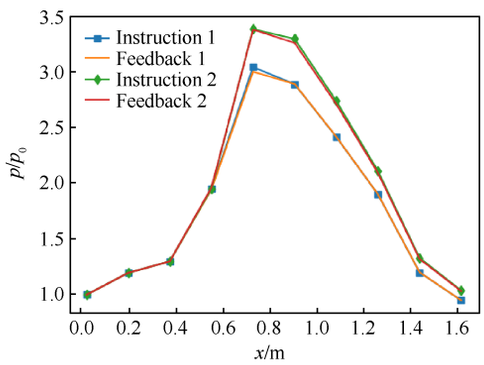

Fig. 9　Control curve of pressure distribution after inflow deviation

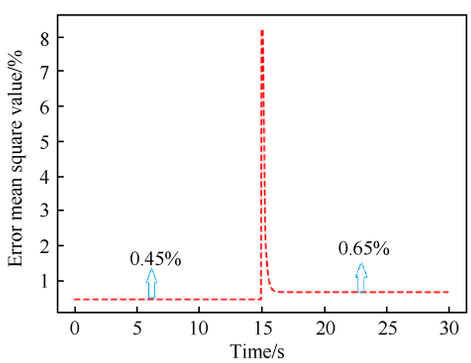

Fig. 10　Mean squared error curve in the control process after inflow deviation

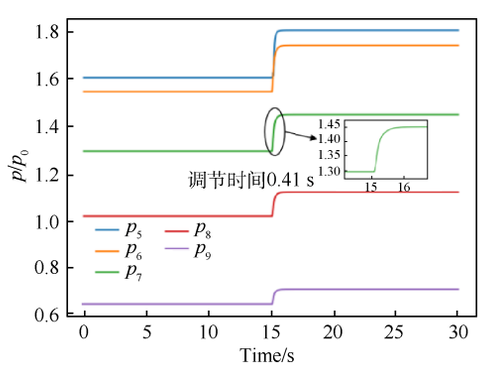

Fig. 11　Adjustment curves and local magnified graphs of 5 pressures(relative values) over time after inflow deviation

在隔离段入口 Ma 3.1、总温 1 860 K、总压 1.90 MPa 下，对入口条件和测量压力分别拉偏 5%。图 13 为仿真获得的压力分布控制曲线，压力的均方差如图 14 所示，两个压力分布的稳态均方误差分别为 0.7% 和 1.22%，图 15 为各压力测点的动态调节过程，可以看出，无超调，调节时间优于 0.5 s，图 16 为各执行机构响应曲线。

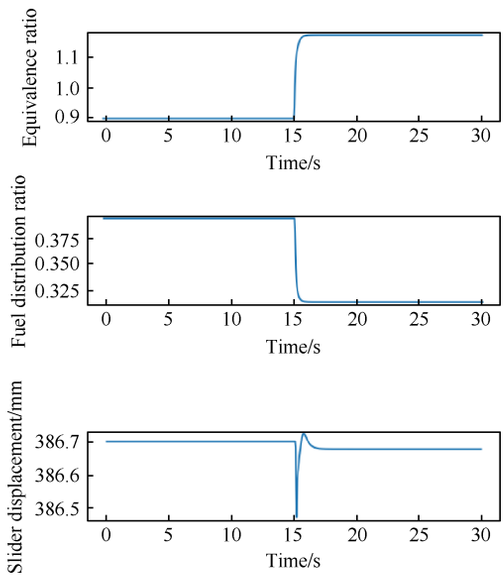

Fig. 12　Response curve of the actuator after inflow deviation

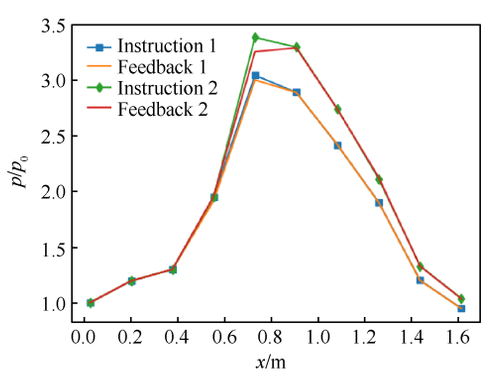

Fig. 13　Pressure distribution control curve after both inflow and pressure deviation

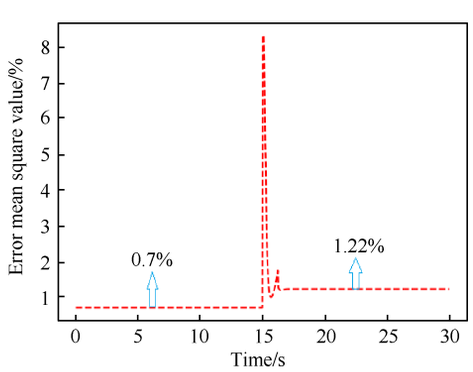

Fig. 14　Mean squared error in the control process of flow and pressure deviation

根据上述仿真结果，所提出的考虑执行机构延时和抖动的系数的自适应调整 SAC 设计算法能有效实现燃烧室压力分布智能控制，压力分布均方差最大 1.44%，超调量最大 2.76%，调节时间最长 0.5 s。如果考虑压力传感器测量精度对压力分布控制的影

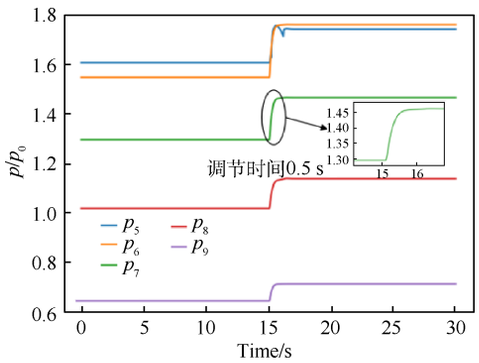

Fig. 15　Adjustment curves and local magnifications of 5 pressures(relative values) over time after both inflow and pressure deviation

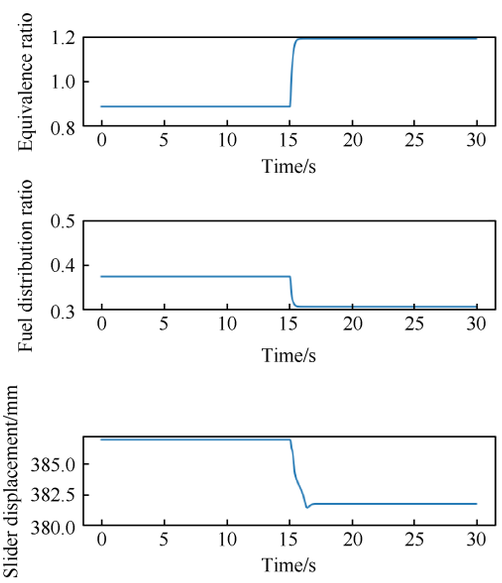

Fig. 16　Response curve of the actuator after both inflow and pressure deviation

响,压力分布均方误差可控制在2%以内。

算法的实时性会影响其使用,本文采用SAC算法属于离线训练,训练后的网络载入控制器参与在线运算,但是仍涉及多个价值网络等,较常规算法耗时较大。为了验证算法是否可以实时运行,将训练后的网络载入控制器硬件参与在线运算耗时评估。采用了具备嵌入式GPU运算能力的发动机电子控制器。通过验证单周期运算耗时在7~10 ms,小于发动机常用的控制周期,可以满足算法的在线应用。

5　结　论

本文通过仿真分析,得到如下结论:

(1)采用动态调整的自适应温度系数可以帮助平衡探索和利用之间的权衡,并促进算法学习到更好的压力分布控制策略,在连续动作空间中的探索也更加高效,能够实现更好的性能。

(2)对于燃烧室压力分布控制,如果奖励函数仅考虑均方差等指标,算法会存在动作抖动的情况,这会对执行机构的耗能有较大损耗,在奖励函数中增加动作抖动惩罚项来抑制无效的抖动,让算法训练过程更平滑。

(3)要实现燃烧室宽工作范围的一维压力分布的控制,需要考虑多工况、来流条件测量值的误差、执行机构通讯的纯延时等因素。根据实际的工况及误差范围,采用随机工况、随机指令、随机来流拉偏、引入纯延时的训练策略,可以有效提高深度强化学习算法的泛化性和鲁棒性。通过此方法训练出的策略经过仿真验证,在来流和压力值拉偏5%的条件下,压力分布控制均方差最大1.44%,超调量最大2.76%,调节时间不超过0.5 s。

参考文献

[1] MILLIGAN R T, EKLUND D R, WOLFF J M, et al. Dual-mode scramjet combustor: numerical analysis of two flowpaths[J]. Journal of Propulsion and Power, 2011, 27(6): 1317-1320.

[2] BOUCHEZ M, GENEVIEVE P, LEVINE V, et al. Air-breathing space launcher interest of a fully variable geometry propulsion system and corresponding French-Russian partnership[C]. Las Vegas: 36th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, 2000.

[3] 潘　余, 李大鹏, 刘卫东, 等. 超燃冲压发动机燃烧模态转换试验研究[J]. 爆炸与冲击, 2008, 28(4): 293-297.
PAN Y, LI D P, LIU W D, et al. Combustion mode transition in a scramjet engine[J]. Explosion and Shock Waves, 2008, 28(4): 293-297. (in Chinese)

[4] 陈玉春, 刘小勇, 黄　兴, 等. 基于集总参数方程的超燃冲压发动机性能计算模型[J]. 推进技术, 2012, 33(6): 840-846.
CHEN Y C, LIU X Y, HUANG X, et al. A model based on lumped parameter method for scramjet performance computation [J]. Journal of Propulsion Technology, 2012, 33(6): 840-846. (in Chinese)

[5] 杨庆春. 宽速域超声速燃烧室的热力与几何调节方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2017.
YANG Q C. Research on thermal and geometric adjustment methods of wide speed domain supersonic combustion chamber[D] Harbin: Harbin Institute of Technology, 2017. (in Chinese)

[6] SINGH R, NATARAJ P S V, MAITY A. Nonlinear control of a gas turbine engine with reinforcement learning

[C]. Vancouver: In Proceedings of the Future Technologies Conference, 2021.

[7] ZHENG Q G, JIN C W, HU Z Z, et al. A study of aero-engine control method based on deep reinforcement learning[J]. IEEE Access, 2019, 7: 55285-55289.

[8] GAO W B, ZHOU X, PAN M X, et al. Acceleration control strategy for aero-engines based on model-free deep reinforcement learning method[J]. Aerospace Science and Technology, 2022, 120: 107248.

[9] TAO B, YANG L, WU D, et al. Deep reinforcement learning-based optimal control of variable cycle engine performance[C]. Guilin: 2022 International Conference on Advanced Robotics and Mechatronics (ICARM), 2022.

[10] ZHU Y Y, PAN M X, ZHOU W X, et al. Intelligent direct thrust control for multivariable turbofan engine based on reinforcement and deep learning methods[J]. Aerospace Science and Technology, 2022, 131: 107972.

[11] 刘智瑞. 基于增强学习的航空发动机智能控制[D]. 南京: 南京航空航天大学, 2020.
LIU Z R. Aero-engine intelligent control based on reinforcement learning[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2020. (in Chinese)

[12] 董建华, 朱建铭, 黎瀚涛, 等. 航空发动机虚拟自学习控制方法研究[J]. 航空工程进展, 2023, 14(6): 81-89.
DONG J H, ZHU J M, LI H T, et al. Research on virtual self-learning control method for aero-engine[J]. Advances in Aeronautical Science and Engineering, 2023, 14(6): 81-89. (in Chinese)

[13] HU Q K, ZHAO Y P. Aero-engine acceleration control using deep reinforcement learning with phase-based reward function[J]. Proceedings of the Institution of Mechanical Engineers, Part G. Journal of Aerospace Engineering, 2022, 236: 1878 - 1894.

[14] 胡雪兰. 变循环发动机智能控制器设计[D]. 大连: 大连理工大学, 2020.
HU X L. Design of intelligent controller for variable cycle engine[D]. Dalian: Dalian University of Technology, 2020. (in Chinese)

[15] 丰　硕. 宽速域冲压发动机几何可调燃烧室性能影响规律研究[D]. 哈尔滨: 哈尔滨工业大学. 2017.
FENG S. Study on performance influence regulation of ramjet variable geometry combustor with wide range velocity[D] Harbin: Harbin Institute of Technology, 2017. (in Chinese)

[16] HAARNOJA T, ZHOU A, ABBEEL P, et al. Soft actor-critic: off policy maximum entropy deep reinforcement learning with a stochastic actor[C]. Stockholm: International Conference on Machine Learning, 2018.

（编辑：白　鹭）

One dimensional pressure distribution reinforcement learning control for ramjet combustor with wide range velocity

NIE Lingcong[1], MU Chunhui[1], LI Shuaiheng[2]

(1. National Key Laboratory of Ramjet, Beijing Power Machinery Institute, Beijing 100074, China;
2. Beijing Computing and Communication Research Institute, Beijing 100074, China)

Abstract: In order to improve the performance of ramjet combustor in a wide velocity regime, a pressure distribution control method for ramjet variable geometry combustor with wide range velocity based on coefficient adaptive adjustment entropy regularization reinforcement learning was proposed. By monitoring and controlling the pressure distribution of the combustor in a one-dimensional flow field, a wide speed regime high performance combustion of this type of combustor was achieved. In this paper, the mathematical model of a wide velocity domain geometrically adjustable combustor is established, and an entropy normalized reinforcement learning method is used to optimize the shape of the pressure distribution along the combustor using slider displacement and two jet flow rates. The adaptive temperature coefficient adjustment algorithm for pressure distribution control is established to improve the convergence rate of pressure distribution one-dimensional control algorithm. By establishing the action jitter penalty function and random training strategy for pressure distribution control, the problems of actuator jitter, the algorithm's universality and delay robustness are solved. The effectiveness of the algorithm is verified by numerical simulation. The maximum mean square error of pressure distribution control is 1.44%, and the regulation speed is no more than 0.5 s, which meets the practical application requirements of engineering.

Key words: Ramjet; Wide velocity combustion chamber; Pressure control; Reinforcement learning; Maximum entropy regularization

Received: 2024-04-17; **Revised**: 2024-12-25.

DOI: 10.3724/1001-4055. 2404056

Corresponding author: NIE Lingcong, E-mail: nielingcong@163.com