Intelligent Control of Morphing Aircraft Based on Soft Actor-Critic Algorithm

Shaojie Ma[1*], Xuan Zhang[1], Yuhang Wang[1], Junpeng Hui[2], Zhu Han[3]

[1]*Research and Development Center, China Academy of Launch Vehicle Technology, Beijing, China*
[2]*Beijing Institute of Space Long March Vehicle Beijing, China*
[3]*Key Laboratory of Digital Earth Science Aerospace Information Research Institute Chinese Academy of Sciences Beijing, China*

Abstract

Morphing aircraft can optimize its flight performance by changing aerodynamic shape. However, the deformation comes up with a great challenge to the control system. The most outstanding characteristics are strongly nonlinear and large uncertainty. Therefore, an intelligent control method is proposed based on Soft Actor-Critic algorithm. Firstly, the state space, action space and reward function required by the algorithm are designed. Then the training efficiency of the algorithm is improved through the way of network pre-training. The mathematical simulation proves that the control strategy can keep the altitude and velocity stable during deformation, and also has strong robustness to the uncertainty caused by deformation and complex external interference.

Keywords

Intelligent control; Morphing aircraft; Flight control; Soft Actor-Critic

1. INTRODUCTION

Morphing aircraft is a type of aircraft that can change its aerodynamic shape independently according to different flight states and task requirements. The aerodynamic parameters and structural parameters of morphing aircraft change nonlinearly during deformation, which makes the aircraft dynamics model have strong nonlinear. The movement between wings and body, as well as the air would produce additional disturbance, which makes the model have great uncertainty.

For the flight control problem of morphing aircraft, the commonly used method is switching linear variable parameter robust control based on linear model [1-2]. However, the linearization would lose the nonlinear characteristics of morphing aircraft model partly. Therefore, some methods based on nonlinear control have become the mainly researched method [3]. Reference [4] realized adaptive control of morphing aircraft based on dynamic inverse control, but such methods also have the problem of high dependence on model accuracy. Therefore, Reference [5] designed a controller based on the active disturbance rejection control theory, which has strong robustness to disturbances during deformation. But the parameters of active disturbance rejection control are too many, which increases the complexity of the controller design.

With the development of intelligent control theory, deep reinforcement learning is increasingly applied to complex control tasks, and shows good performance [6-7]. Reference [8] applied Soft Actor-Critic algorithm to fault-tolerant control of aircraft. Reference [9] designed a composite controller based on deep deterministic policy gradient algorithm and traditional controller. Reference [10] proposed a fixed-time disturbance rejection controller which set parameters assisted by twin delayed deep deterministic policy gradient algorithm.

Based on this, this paper proposes a controller for morphing aircraft based on Soft Actor-Critic algorithm. Taking a variable sweep UAV as the object, Firstly, a longitudinal mathematical model is established considering its multi-rigid body structure. Then the state space, action space, reward function and network structure required by the algorithm are designed under the framework of Markov

ICCEIC2022@3rd International Conference on Computer Engineering and Intelligent Control
EMAIL: * Corresponding author: mashj05@163.com (Shaojie Ma)

decision process. Finally, the control accuracy and strong robustness of the proposed control strategy are verified by mathematical simulations.

2. MATHEMATICAL MODEL OF MORPHING AIRCRAFT

In this paper, a class of variable sweep aircraft is considered. The model is similar with the research in reference [11-12]. The longitudinal motion model of the morphing aircraft can be described by

$$\begin{cases} \dot{V} = \left(-X + P\cos\alpha - mg\sin\theta + F_{sx}\right)/m \\ \dot{\theta} = \left(Y + P\sin\alpha - mg\cos\theta - F_{sy}\right)/\left(mV\right) \\ \dot{h} = V\sin\theta \\ \dot{\varphi} = \omega_z \\ \dot{\omega}_z = \left(M_z + M_{sz} - S_x g\cos\theta - \dot{I}_z\omega_z\right)/I_z \\ \alpha = \varphi - \theta \end{cases} \tag{1}$$

Where V, θ, α, h denote the velocity, flight path angle, angle of attack and altitude, respectively. φ, ω_z denote the angle and angular rate of pitch, respectively. m, I_z represent the mass and moment of inertia of the aircraft, respectively. g is the gravitational acceleration. P is the thrust of the engine. X, Y, M_z denote the lift force, drag force and pitch moment, respectively, which are given as

$$\begin{cases} X = \left[C_{x0}(\varsigma) + C_{x\alpha}(\varsigma)\alpha + C_{x\alpha 2}(\varsigma)\alpha^2\right]QS \\ Y = \left[C_{y0}(\varsigma) + C_{y\alpha}(\varsigma)\alpha\right]QS \\ M_z = \left[C_{m0}(\varsigma) + C_{m\alpha}(\varsigma)\alpha + C_{m\delta}(\varsigma)\delta_z + C_{m\omega}(\varsigma)\bar{\omega}_z\right]QSL \end{cases} \tag{2}$$

Where $Q = 0.5\rho V^2$ is dynamic pressure, S is the wing surface, L is the mean aerodynamic chord, $C_{x0}(\varsigma)$, $C_{x\alpha}(\varsigma)$, $C_{x\alpha 2}(\varsigma)$, $C_{y0}(\varsigma)$, $C_{y\alpha}(\varsigma)$, $C_{m0}(\varsigma)$, $C_{m\alpha}(\varsigma)$, $C_{m\delta}(\varsigma)$, $C_{m\omega}(\varsigma)$ are the aerodynamic derivatives which can be formulated as polynomial functions of the sweep angle $\varsigma \in [0°, 45°]$.

F_{sx}, F_{sy}, M_{sz} are the inertial forces and moment caused by deformation, and S_x is the static moment distribution in the body frame varies with the sweep angle ς, which are given as

$$\begin{cases} F_{sx} = \left(\dot{\omega}_z\sin\alpha + \omega_z^2\cos\alpha\right)S_x + 2\dot{S}_x\omega_z\sin\alpha - \ddot{S}_x\cos\alpha \\ F_{sy} = \left(\dot{\omega}_z\cos\alpha - \omega_z^2\sin\alpha\right)S_x + 2\dot{S}_x\omega_z\cos\alpha + \ddot{S}_x\sin\alpha \\ M_{sz} = \left(\dot{V}\sin\alpha + V\dot{\alpha}\cos\alpha - V\omega_z\cos\alpha\right)S_x \\ S_x = 2m_1r_1 + m_3r_3 \end{cases} \tag{3}$$

Where m_1, m_3 represent the mass of the wings and body of the aircraft, respectively. r_1, r_3 denote the position of related components in the body frame.

3. DESIGN OF CONTROLLER BASED ON SOFT ACTOR-CRITIC ALGORITHM

3.1. Principle of Soft Actor-Critic Algorithm

Soft Actor-Critic (SAC) algorithm is a deep reinforcement learning algorithm based on Actor-Critic (AC) framework, and use deep neural network to represent policy π and action-state value function $Q(s,a)$. SAC uses stochastic network as policy network $\pi(s\,|\,\theta_\pi)$, which outputs the average and variance of the action and obtains the action instruction through sampling, so as to improve the exploration of the algorithm. SAC uses two critic networks $Q_i(s,a\,|\,\theta_{Qi})$ to reduce the estimation error of Q-function, which is inherited from Double Q-Learning. Moreover, as an off-policy algorithm, SAC sets replay buffer and two target critic networks $Q_{i'}(s,a\,|\,\theta_{Qi'})$. In addition, SAC encourages more exploration by maximizing the cumulative reward of entropy regularization rather than just the

cumulative reward. Which makes it become a kind of widely used continuous control reinforcement learning algorithm.

SAC selects the minimum value of the two target critic networks when updating the Bellman equation, so the loss function of critic network can be given as

$$\begin{cases} y_t = r_{t+1} + \gamma \mathrm{E}\left[\min\left(Q_{i'}\left(s_{t+1}, a_{t+1} \mid \theta_{Qi'} \right) \right) - \alpha \log \pi\left(s_t \mid \theta_\pi \right) \right] \\ L\left(\theta_{Qi} \right) = \dfrac{1}{N} \sum_{j=1}^{N} \left(y_j - Q_i\left(s_j, a_j \mid \theta_{Qi} \right) \right)^2 \\ \theta_{Qi,t+1} = \theta_{Qi,t} + \kappa_Q \nabla_{\theta_{Qi}} L\left(\theta_{Qi} \right) \end{cases} \tag{4}$$

Where N denotes batch size, κ_Q is the learning rate of critic network, a'_{t+1} represents the next action corresponding to the next state, α represents the temperature parameter, similarly the loss function of policy network can be given as

$$\begin{cases} J\left(\pi \right) = \mathrm{E}\left[\alpha \log \pi\left(s_t \mid \theta_\pi \right) - \min_{i=1,2}\left(Q_i\left(s_j, a_j \right) \right) \right] \\ \nabla_{\theta_\pi} J\left(\pi \right) = \dfrac{1}{N} \sum_{j=1}^{N} \nabla_{\theta_\pi} \pi\left(s_j \mid \theta_\pi \right) \nabla_a J\left(\pi \right) \\ \theta_{\pi,t+1} = \theta_{\pi,t} + \kappa_\pi \nabla_{\theta_\pi} J\left(\pi \right) \end{cases} \tag{5}$$

Where κ_π is the learning rate of policy network. Besides, SAC also provides a method to adjust the temperature parameter automatically, and the loss function can be given as

$$\begin{cases} \nabla_\alpha J\left(\alpha \right) = \nabla_\alpha \mathrm{E}\left[-\alpha \log \pi\left(s_t \mid \theta_\pi \right) - \alpha\kappa \right] \\ \alpha_{t+1} = \alpha_t + \kappa_\alpha \nabla_\alpha J\left(\alpha \right) \end{cases} \tag{6}$$

Where κ is target entropy, which can dynamically find the lowest temperature that still ensures a certain minimum entropy, κ_α is the learning rate of α. SAC updates the target network by exponential smoothing rather than direct replacement like DDPG, which can make the target network update more slowly and stably, and improve the stability of algorithm.

3.2. Design of Controller

The longitudinal plane of the aircraft is controlled by altitude and velocity. Due to the complex continuous action space of the aircraft, it is difficult for the randomly initialized policy network to ensure flight stability, and the quality of the samples collected at the initial training episode is poor which would result that the training efficiency is extremely low. Therefore, the network pre-training is considered in this paper. Firstly, the traditional controller is fitted through deep learning, and the deep neural network learned is used as the initial policy network of SAC. The training structure is shown in Figure 1. The policy network serves as the aircraft controller directly, and the action from network is the command of the aircraft control actuator.

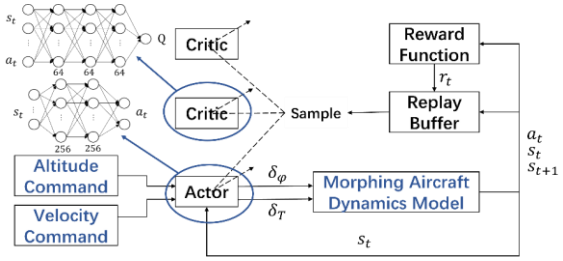

Figure 1. Controller structure

The control model is transformed into Markov decision process, then the state space, action space, reward function and deep neural network structure are designed under this framework.

3.2.1. State space and action space

Drawing on the traditional controller design ideas, and considering the influence of deformation on the model, the seven-dimensional state vector is designed as

$$s = \left[\Delta h, \Delta \dot{h}, \Delta V, \Delta \dot{V}, \varphi, \omega_z, \varsigma \right] \tag{7}$$

The actuator of altitude channel is mainly elevator, and the velocity channel is mainly adjustable thrust engine, so the action vector is designed as

$$a = \left[\delta_\varphi, \delta_T \right] \tag{8}$$

3.2.2. Reward function

In order to ensure that the aircraft can accurately track altitude and velocity commands, and reduce the control energy demand, the reward function is designed as

$$r = \varpi_h \left| \Delta h \right| + \varpi_V \left| \Delta V \right| + \varpi_\varphi \left| \delta_\varphi \right| + \varpi_T \left| \delta_T \right| + \varpi_1 + \varpi_2 + \varpi_d \tag{9}$$

Where first four are the penalty related to tracking error, deflection angle of elevator, and the thrust. When the height tracking error, velocity tracking error, deflection angle of elevator and thrust increase, the penalty value increases, $\varpi_i \left(i = h, V, \varphi, T \right)$ is the weight value, respectively. The last two are sparse reward for tracking accuracy. When the tracking error is less than the threshold, the reward is applied. ϖ_d is the penalty of states divergence. In this paper, when the height tracking error is greater than 500m, it would be judged that the states divergence, and the episode would be ended.

3.2.3. Deep neural networks

All the networks used in this paper are back propagation neural network. The input layer of the policy network has 7 neurons corresponding to the 7-dimensional state vector, the hidden layer has 2 fully connected layers, both composed of 256 neurons, and the activation function are ReLU. The output layer is composed of the mean and the variance of the action. The two critic networks have the same structure. The input layer has 9 neurons corresponding to the 7-dimensional state vector and the 2-dimensional action vector; the hidden layer has 3 fully connected layers, all composed of 64 neurons; and the activation function are ReLU, too. The output layer has 1 neuron corresponding to the action-state value function.

4. NUMERICAL SIMULATION

The initial simulation states are $h_0 = 1000m \pm 5m$, $V_0 = 30m/s \pm 5m/s$, $\alpha_0 = 0.995° \pm 1°$, $\varphi_0 = 0.995° \pm 1°$, and the initial altitude and velocity command are 1000m, 30m/s, respectively. And the change of sweeping angle is shown in Figure 2.

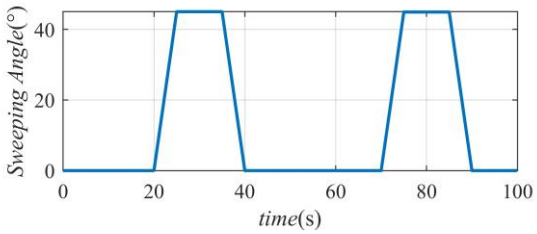

Figure 2. Curve of variation of sweeping angle

The control step is 10ms, the network updating step is 100ms, and the simulation time of each episode is 100s. The algorithm training parameters and the weights of the reward function are shown in TABLE 1.

TABLE I. PARAMETERS DESIGN

Parameters	Values	Parameters	Values
Episodes	2000	Steps	10000
Batch Size	256	ϖ_h	-2
ϖ_1	0.5	ϖ_V	-0.5
ϖ_2	0.5	ϖ_φ	-1
ϖ_{done}	-1000	ϖ_T	-0.001

4.1. Result analyses of SAC

Figure 3 shows the change of the reward with training episodes during the training based on SAC. A total of 3000 episodes of training were conducted. The thick line is the reward been smoothed. It can be seen that the average reward generally increases during training. In addition, due to the strong exploration of SAC, the flight state would divergent sometimes, such as the fluctuations of the light line. But the upward trend of reward is not be affected.

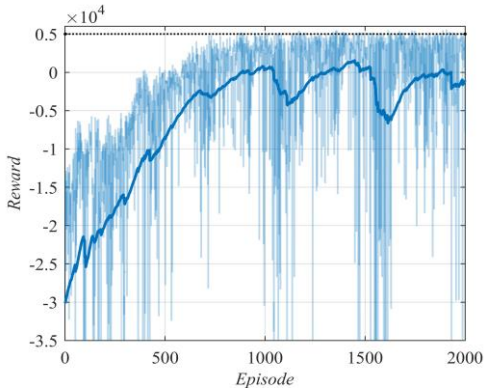

Figure 3. Cumulative reward during training

4.2. Control Performance Analyses of the Controller

In order to verify the effectiveness of the control policy obtained by training, the simulation verification under nominal state and deviation state is carried out based on the longitudinal plane motion model of the morphing aircraft.

4.2.1. Simulation under nominal states

Figure 4 shows the altitude and velocity tracking results, the deflection angle of elevator and the thrust. The altitude command changes from 1000m to 1050m, and the velocity is always 30m/s. The blue line is for SAC optimized controller, green dotted line is for pre-training controller.

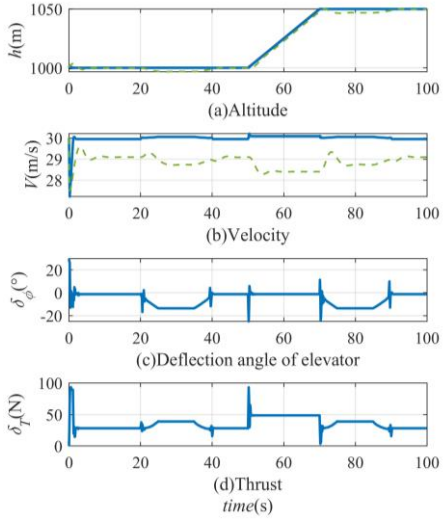

Figure 4. Altitude and velocity tracking results under nominal states

The integral absolute error of altitude and velocity before optimization are 167.1219m, 114.5735m/s, respectively. After optimization they are reduced to 7.4009m, 7.0559m/s, respectively. The control accuracy has been greatly improved. Besides, the impact of deformation is greatly reduced.

4.2.2. Simulation under deviation states

In order to verify the robustness to complex external disturbances of the control policy proposed in this paper. 20% aerodynamic deviation, 15% structural deviation, and 10% density disturbances are considered. Figure 5 shows the tracking results of altitude and velocity.

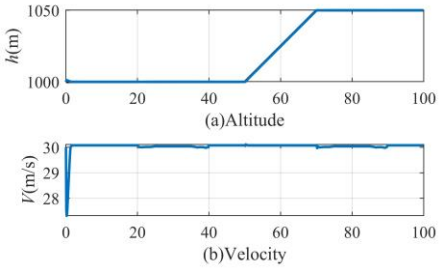

Figure 5. Altitude and velocity tracking results under deviation states

The integral absolute error of altitude and velocity under deviation states are 10.2310m, 7.9277m/s, respectively. From this show that the control policy trained in this paper can achieve stable control under the deviation states and ensure the altitude and velocity accuracy in the deformation transition process, which proves its robustness to external disturbances.

5. Conclusions

Aiming at the problems of strong nonlinear dynamics model of deformed aircraft and complex internal and external interference factors in the deformation process, taking a class of variable swept aircraft as an example, the height and velocity controller design of deformed aircraft was carried out based on SAC deep reinforcement learning algorithm. Network pre-training was adopted to ensure stable control at the initial stage of algorithm training and improve sample quality. The simulation results show that the proposed control policy can improve the accuracy of altitude and velocity control greatly, and has strong robustness to both internal and external uncertainties of the model during deformation. However, this paper only carried out experimental verification based on mathematical simulation, and it needs to continue to carry out practical application verification.

6. REFERENCES

[1] K Boothe, K Fitzpatrick, R Lind, “Controllers for disturbance rejection for a linear input-varying class of morphing aircraft,” 46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Materials Conference, Austin, Texas, April, 2005.

[2] W Jiang, Ch Dong, T Wang, Q Wang, “Smooth switching LPV robust control for morphing aircraft,” Control and Decision, vol 31, pp. 66-72, January, 2016.

[3] M Ran, Ch Wang, H Liu, et al, “Research status and future development of morphing aircraft control technology,” Acta Aeronautica et Astronautica Sinica, vol 43, pp. 527449, 2022.

[4] T Lombaerts, J Kaneshige, S Schuet, “Dynamic inversion based full envelope flight control for an VTOL vehicle using a unified framework,” AIAA Scitech 2020 Forum, Orlando, FL, January, 2020.

[5] H Song, L Jin, “Dynamic modeling and stability control of folding wing aircraft,” Chinese Journal of Theoretical and Applied Mechanics, vol 52, pp. 1548-1559, November, 2020.

[6] W Koch, R Mancuso, R West, et al, “Reinforcement learning for UAV attitude control,” ACM Transactions on Cyber-Physical Systems, vol 3, pp. 1-21, 2019.

[7] Y Wang, J Sun, H He, et al, “Deterministic policy gradient with integral compensator for robust quadrotor control,” IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol, 50, pp. 3713-3725, 2019.

[8] K Dally, E V Kampen, “Soft actor-critic deep reinforcement learning for fault tolerant flight control,” AIAA Scitech 2022 Forum, San Diego, CA&Virtual, January , 2022.
[9] X Huang, J Liu, Ch Jia, et al, “Deep deterministic policy gradient algorithm for UAV control,” Acta Aeronautica et Astronautica Sinica, vol 42, pp. 524688, 2021.
[10] Y Liu, H Wang, T Wu, et al, “Attitude control for hypersonic reentry vehicles: An efficient deep reinforcement learning method,” Applied Soft Computing, vol 123, pp. 108865, 2022.
[11] Z Wu, J Lu, Q Zhou, et al. “Modified adaptive neural dynamic surface control for morphing aircraft with input and output constraints,” Nonlinear Dyn, vol 87, pp. 2367–83, 2017.
[12] L Gong, Q Wang, Ch Hu, et al, “Switching control of morphing aircraft based on Q-learning,” Chinese Journal of Aeronautics,vol 33, pp. 672–687, 2020.