Lecture Notes in Electrical Engineering 1050

Song Fu *Editor*

2023 Asia-Pacific International Symposium on Aerospace Technology (APISAT 2023) Proceedings

Volume I

Springer

Lecture Notes in Electrical Engineering 1050

The book series *Lecture Notes in Electrical Engineering* (LNEE) publishes the latest developments in Electrical Engineering—quickly, informally and in high quality. While original research reported in proceedings and monographs has traditionally formed the core of LNEE, we also encourage authors to submit books devoted to supporting student education and professional training in the various fields and applications areas of electrical engineering. The series cover classical and emerging topics concerning:

- Communication Engineering, Information Theory and Networks
- Electronics Engineering and Microelectronics
- Signal, Image and Speech Processing
- Wireless and Mobile Communication
- Circuits and Systems
- Energy Systems, Power Electronics and Electrical Machines
- Electro-optical Engineering
- Instrumentation Engineering
- Avionics Engineering
- Control Systems
- Internet-of-Things and Cybersecurity
- Biomedical Devices, MEMS and NEMS

For general information about this book series, comments or suggestions, please contact leontina.dicecco@springer.com.

To submit a proposal or request further information, please contact the Publishing Editor in your country:

China

Jasmine Dou, Editor (jasmine.dou@springer.com)

India, Japan, Rest of Asia

Swati Meherishi, Editorial Director (Swati.Meherishi@springer.com)

Southeast Asia, Australia, New Zealand

Ramesh Nath Premnath, Editor (ramesh.premnath@springernature.com)

USA, Canada

Michael Luby, Senior Editor (michael.luby@springer.com)

All other Countries

Leontina Di Cecco, Senior Editor (leontina.dicecco@springer.com)

**** This series is indexed by EI Compendex and Scopus databases. ****

Song Fu
Editor

2023 Asia-Pacific International Symposium on Aerospace Technology (APISAT 2023) Proceedings

Volume I

Springer

Editor
Song Fu
School of Aerospace Engineering
Tsinghua University
Beijing, China

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-97-3997-4 ISBN 978-981-97-3998-1 (eBook)
https://doi.org/10.1007/978-981-97-3998-1

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Contents

Research on Optimization Design Method of Autonomous Deformation Decision for Intelligent Morphing Aircraft

Dan Xu(✉)

College of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an, China
xdan@xauat.edu.cn

Abstract. Intelligent morphing aircraft can timely and independently change its shape according to the flight mission and environment, and meet different flight missions with different aerodynamic layouts, so as to achieve performance improvement and trajectory optimization in different flight stages. So it is also one of the development trends that is most likely to bring about the technical revolution of the future aerospace aircraft. In order to improve the trajectory characteristics, it is necessary to establish the controlled model in advance by using traditional controller to control structural changes. However, due to obvious changes in the structure of morphing aircraft, it is impossible to establish accurate mathematical model. Therefore, an intelligent trajectory optimization method is proposed to solve the problem of aircraft autonomous deformation decision control. This paper takes intelligent morphing aircraft flying at high speed in large airspace as the research object, aiming at the technical problems that the aircraft is difficult to obtain sufficient deformable flight test data in advance, which leads to the difficulty in predicting the optimal aerodynamic shape under different flight states, and the traditional controller cannot be used to optimize the deformation. A deformable decision scheme based on reinforcement learning (RL) network is proposed, which realizes that the aircraft structure can be changed independently according to the real-time state in flight, so as to improve the aerodynamic performance and optimize the flight trajectory.

Keywords: Morphing aircraft · Trajectory optimization · Intelligent decision · Reinforcement Learning · Autonomous deformation

1 Introduction

Intelligent morphing aircraft has great application prospects and is considered by National Aeronautics and Space Administration (NASA) as one of the most subversive and transformative strategic development directions in the future aerospace. Take the future intelligent morphing aircraft envisioned by NASA as an example, based on the comprehensive application of new intelligent materials, intelligent drive devices, sensors and intelligent control systems, the aircraft can continuously, smoothly and independently change the partial or overall structure shape of the body during the flight to

S. Fu (Ed.): APISAT 2023, LNEE 1050, pp. 1–11, 2024.
https://doi.org/10.1007/978-981-97-3998-1_1

adapt to different flight environments, complete a variety of target task and always maintain optimal aerodynamic characteristics. It is a new concept aircraft with flight adaptive ability. Due to its great performance advantage and application potential, in recent years, various military powers have carried out the research on the shape and structure of intelligent morphing aircraft, and various design concepts have been proposed and realized through optimization design. In addition, intelligent morphing aircraft also involves aerodynamics, structural dynamics, intelligent flexible materials/structures, advanced sensing and system integration, intelligent information and modern control and other frontier technologies, each of which is the bottleneck restricting the improvement of aircraft performance. At present, the research on adaptive flexible materials, deformed wing structure and aerodynamic characteristics has been relatively perfect [1–3], but the research on core algorithms such as deformation decision and intelligent control is relatively insufficient, which is the key to realize adaptive flight of intelligent morphing aircraft.

As an important link in aircraft design, trajectory optimization involves various flight stages of aircraft and has important research significance [4, 5]. For intelligent morphing aircraft, deformable wing structure is the main operation mechanism of flight control. It has always been a difficult problem how to make the aircraft has autonomous decision-making ability, and improve flight performance and optimize flight trajectory by changing its shape independently according to different requirements of flight tasks. It is also the core technical problem that must be solved from the control level after the major breakthrough of intelligent deformable material, deformable wing structure, and other key technologies. In recent years, based on its powerful exploration function and independent learning ability, RL together with supervised learning and unsupervised learning, has been called the three major machine learning technologies. As an online learning technology, RL innovatively views learning as a process of interaction with the environment. The algorithm does not need to rely on a large amount of prior knowledge. It obtains feedback and accumulates experience through continuous trial and error, and takes this as the guidance to realize independent learning. For the trajectory optimization problem of intelligent aircraft to be solved in this paper, the objective is to use RL algorithm to build a deformation decision network model and optimize the deformation strategy autonomously through multiple training, so that the aircraft can independently decide the deformation timing and deformation quantity according to the flight state, so as to improve the aerodynamic characteristics and optimize the flight trajectory.

2 Methods

OpenAI tends to choose Proximal Policy Optimization (PPO) algorithm [6, 7] with the widest application range when trying to realize various problems. So, in the research on the trajectory optimization scheme proposed in this paper, we also firstly use this algorithm to complete the preliminary feasibility verification of the autonomous deformation decision scheme for morphing aircraft. Then, in order to further explore the learning advantages of different algorithms, we use another kind of off-policy Deep Deterministic Policy Gradient (DDPG) algorithm [8–10] for comparative test. This chapter will complete the construction of algorithm network based on the implementation principles of these two algorithms, and the setting of network hyperparameters will be described in detail.

In PPO algorithm based on Actor-Critic network, there are three networks are built generally: one critic network and two actor networks (old actor network and new actor network). In an episode, agent firstly uses the existing policy in new actor network to interact with the environment and get a batch of data. In this process, actor and critic network will not be optimized. After obtaining a complete batch of data, actor network and critic network begin to learn this batch of data, which is similar to Policy Gradient (PG) algorithm. However, different from PG algorithm, actor network and critic network will learn this batch of sampled data T times. T is the number of iteration rounds of actor network. In order to avoid over-fitting, we adopt the simplest three-layer fully connected neural network structure. In order to ensure high precision and generalization ability, the most basic principle to determine the number of hidden layer nodes is to adopt compact structure as far as possible on the premise of meeting accuracy requirements. According to the observed state space and the complexity of problem, the hidden layer network with 800 nodes is finally adopted. The activation function of the hidden layer adopts piecewise linear function Relu so that the neural network can approximate any nonlinear function arbitrarily and accelerate the convergence of training network. When constructing the actor network, the concept of probability distribution is added to complete the selection of actions. The activation function used to calculate mean value is Relu, while the activation function used to calculate variance is Sigmoid. Actor network will return a normal distribution based on the mean and variance, and the actions will be sampled based on this normal distribution. The loss function is calculated by clipping to prevent the distribution difference from being too large. After many tests, we set the number of training episodes to 400 to complete 400 trajectory tests. In each episode, aircraft starts to glide from a fixed initial altitude at the same initial speed, and the training for that trajectory ends when the altitude drops to 20 km. Set the learning rate of actor network and critic network to 0.0001 and 0.0002 respectively, and the network is updated every 10 steps. Considering that the behavior strategy has not been optimized in the early stage of training and the quality of the selected actions is also uneven, a discount factor is set in the calculation of expected returns to adjust the calculation of cumulative rewards in the learning process, and the value is 0.9.

Similarly, DDPG algorithm still adopts Actor-Critic network model. The difference is that the algorithm adopts the design of the target network in DQN algorithm, so there are four networks in the algorithm: two critic networks with the same structure and two actor networks with the same structure. For the fairness of comparison, the structure of actor network and the critic network, training episodes, discount factor and termination conditions are consistent with PPO algorithm. In the output layer of the actor network, the folding sweep angle is increased or decreased by applying Tanh function. In order to calculate Q values more accurately, the actions are added to the first hidden layer of the critic network to enrich the learning data. In the process of deformation decision optimization training, PPO algorithm can better complete exploratory learning of action space because of the randomness of its behavior strategy output. However, the deterministic behavior output of DDPG algorithm is not conducive to the exploration of action space, so we add random noise to the first 100 training trajectories to enhance the exploration ability. The learning rate of actor network and critic network is set to 0.001 and 0.0001 respectively, the minimum batch of empirical data required for each gradient update is set to 64 (Figs. 1 and 2).

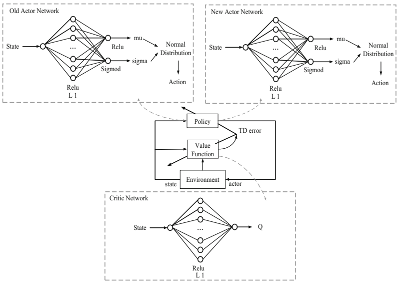


Fig. 1. The network structure of PPO

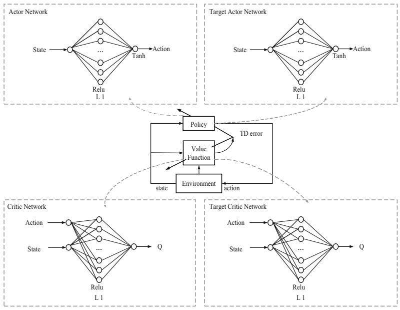


Fig. 2. The network structure of DDPG

3 Experimental Setup

In this paper, trajectory optimization of gliding phase is taken as the objective to design the deformation decision scheme of intelligent aircraft. Glide distance and descending height are taken as control parameters to set the reward function, and the calculated reward value is used to evaluate the deformation amount of the autonomous decision, so as to guide the decision scheme to achieve autonomous learning and complete the parameter optimization in the direction of increasing the reward. After multiple iterative trainings, the reward value gradually converges, and the optimized decision network is obtained.

3.1 Morphing Aircraft Model

The research model chosen in this paper is the aircraft with variable sweep angle of wing, which can realize continuous change from 0° to 50°, shown in Fig. 3. During flight, the aircraft adjusts the aerodynamic parameters by changing the swept-wing folding Angle through autonomous decision making, obtains a new flight trajectory equation, calculates the trajectory parameters, and brings them into the reward function to obtain the reward return to guide the decision optimization. In the optimal design scheme of deformation decision making in glide stage, it is only necessary to calculate the parameter change of flight trajectory in the longitudinal plane, without considering the influence of sideslip angle. Therefore, it can be boldly assumed that the Earth is a uniform, non-rotating sphere, so as to ignore the influence of aspherical perturbation on the gravitational field, and obtain the flight trajectory in the general sense.

$$
\begin{cases}
\frac{dV}{dt} = -D - g\sin\theta \\
\frac{d\theta}{dt} = \frac{L\cos\nu}{V} + \frac{1}{V}\left(\frac{V^2}{r} - g\right)\cos\theta \\
\frac{dr}{dt} = V\sin\theta \\
\frac{dR}{dt} = R_e\frac{V}{r}\cos\theta
\end{cases} \tag{1}
$$

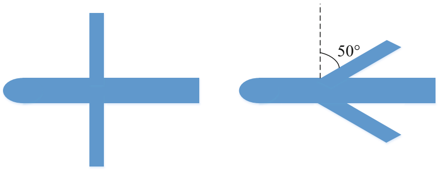


Fig. 3. Conceptual diagram of variable sweep mode aircraft

In the Equation set (1), the constant quantity includes gravitational acceleration g and Earth radius R_e, the velocity information includes flight velocity V and velocity inclination angle θ, v representing control quantity inclination angle; r and R respectively represent geocentric distance and glide distance; aerodynamic parameters include drag acceleration D and lift acceleration L, which can be calculated according to Eq. (2).

$$\begin{cases} D = \frac{X}{m} = \frac{C_D \rho V^2 S}{2m} \\ L = \frac{Y}{m} = \frac{C_L \rho V^2 S}{2m} \end{cases} \tag{2}$$

where, m is the mass of the aircraft, and S is the reference area. The atmospheric density ρ is determined by an exponential model related to height. Drag coefficient C_D is a function of reference area S, attack angle α and Mach number Ma, lift coefficient C_L is a function of α, Ma and flight altitude h ($h = R$-R_e). In order to improve the lateral maneuvering performance and glide distance of the aircraft, realize thermal protection under high-speed environment and design the attitude control system, the attack angle is usually designed according to the piecewise linear function shown in Eq. (3), so as to obtain a smoother attack angle curve. Where, α_{max} and $\alpha_{max(K)}$ respectively represent the maximum attack angle and the corresponding attack angle to the maximum lift-drag ratio in the flight process, V_1 and V_2 represent the speed of subsection points, and K represent lift-drag ratio, is L/D.

$$\alpha = \begin{cases} \alpha_{\max}, & V1 < V < V_e \\ \frac{\alpha_{\max} + \alpha_{\max(K)}}{2} + \frac{\alpha_{\max} - \alpha_{\max(K)}}{2} \sin\left\{ \frac{[V - (V_1 + V_2)/2]\pi}{V_1 - V_2} \right\}, & V_2 \le V \le V_1 \\ \alpha_{\max(K)}, & V_f \le V \le V_2 \end{cases} \tag{3}$$

Considering that once the aerodynamic configuration of the aircraft changes, the structural parameters and aerodynamic characteristics of the aircraft will change accordingly, making flight trajectory equations also need to be adjusted accordingly, which brings great challenges to algorithm design, program testing and scheme verification. Therefore, we assume that the online planning of flight trajectory and the stability control of variational configuration are ideal states, and assume that a set of fixed values can be selected for trajectory calculation when the parameters change little in a certain deformation range. For the high-speed aircraft model with variable swept-wing adopted in the experiment, according to the sensitivity of aerodynamic coefficient to deformation, the following configuration is divided. For the three configurations, the corresponding deformation aerodynamic parameters and structural parameters were obtained by experiments, and then the flight trajectory motion equations of the aircraft under three different configurations were constructed to predict the flight trajectory before and after deformation (Table 1).

Table 1. Division of configurations

Sweep angle/°	Configuration
0–15	1
15–30	2
30–50	3

3.2 Scheme Design

The scheme design of deformation decision is mainly divided into two parts: deformation timing decision and deformation quantity decision, as shown in Fig. 4. When the mission is determined, the trajectory has the corresponding flight target parameters. For example, the goal of the glide section is to fly as far as possible and use as little energy as possible. Then, in flight, the trajectory parameters measured in real time by airborne sensors can be used to determine online whether the trajectory characteristics need to be improved by changing the shape at the current moment according to the deformation timing criteria. When it is judged that it is necessary to optimize the flight path through deformation, the aerodynamic data at the current time is used for constraint calculation to determine whether the safe deformation conditions are met. If yes, the deformation timing decision module will output the deformation instruction, and then the deformation quantity decision module will directly output the optimal deformation quantity required under the current state according to the flight state data given by sensors and the deformation optimization strategy.

(1) Deformation timing decision module

The deformation timing decision module first completes the mapping of aerodynamic parameters based on flight trajectory data and makes deformation judgment. The specific scheme design is shown in the shadow box on the left side of Fig. 5. One of the most important is to establish the aerodynamic parameter mapping model, which needs to be based on a large number of experimental data. That is, lift coefficient and drag coefficient of the aircraft in different configurations under different combinations of attack angle, velocity and altitude. When judging the deformation timing, it is necessary to calculate the difference between the theoretical optimal lift-drag ratio that can be achieved under the current flight state and the actual lift-drag ratio corresponding to the current configuration. When the difference reaches 4% of the optimal lift-drag ratio, it is judged that the aerodynamic characteristics need to be improved by deformation. For the high-altitude high-speed aircraft to be studied in this paper, in addition to considering the reliable relationship between structural deformation and aerodynamic and attitude stability under maneuvering environment, the influence of heat flux, dynamic pressure

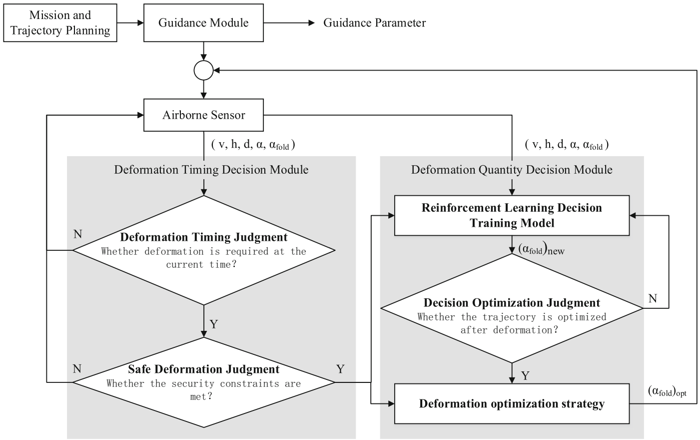


Fig. 4. Flow chart of the deformation decision scheme design

and overload on the aircraft should also be considered. So, the safe deformation judgment is set in the shadow box on the right of Fig. 5. By calculating the heat flux, dynamic pressure and overload in the current state, and comparing with the limit value, the deformation constraint conditions are judged to be satisfied. The calculation formulas are as follows, where, $k_\gamma = 5.5 \times 10^{-8}$ is the stagnation heat flux coefficient, $\dot{Q}_{max}$, q_{max} and n_{ymax} are the constant limit values of heat flux, dynamic pressure and overload respectively, which are usually determined by the aircraft and the flight task.

$$\dot{Q} = k_\gamma \rho^{0.5} v^{3.15} \leq \dot{Q}_{\max} \tag{4}$$

$$q = 0.5 \rho v^2 \leq q_{\max} \tag{5}$$

$$n_y = (L \cos\alpha + D \sin\alpha)/mg \leq n_{y\max} \tag{6}$$

(2) Deformation quantity decision module

In the constructed deformation decision module, the reinforcement learning algorithm is used to build the network model, the trajectory parameter is used to set the reward function, and the model parameters are constantly optimized through the characteristics of self-learning and self-evolution of reinforcement learning to achieve the optimal deformation output. The specific training process is shown in Fig. 6. A deformation quantity is randomly given within the deformation range, and the lift and drag coefficients under two configurations before and after deformation under the current flight environment are obtained by using the aerodynamic parameter mapping model, and the corresponding flight trajectory prediction equation is constructed to predict the trajectory parameters after gliding for 10 s before and after deformation, mainly to obtain the descent height and glide distance. The flight height and glide distance predicted by the flight trajectory prediction model under the current configuration are denoted as h

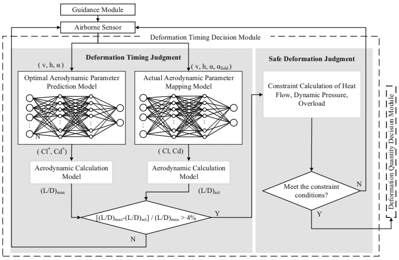


Fig. 5. Deformation timing decision module

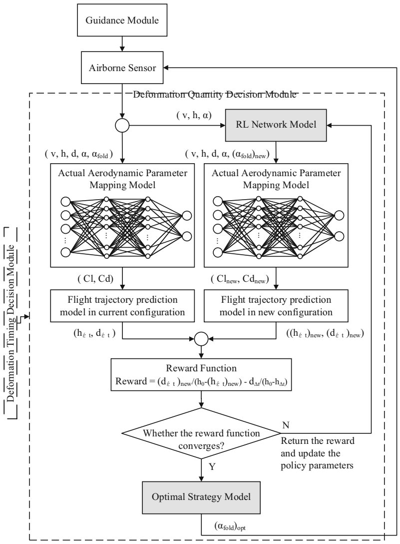


Fig. 6. Deformation quantity decision module

and d, while the parameters predicted under the new configuration after deformation are denoted as h_{new} and d_{new}. Then the trajectory parameters are substituted into the reward function Eq. (7) for calculation, in order to evaluate the current deformation strategy, and guide parameter updating. As can be seen from the reward function, when descending the same height, the longer the glide distance, the greater the reward. According to the principle that the higher the reward value is, the better the behavior strategy is, the strategy parameters are optimized until they converge after multiple interactive trainings. In actual flight, based on the current flight data obtained by sensors, the deformation optimization strategy obtained by off-line training can be used to quickly give the optimal deformation online. The aircraft can improve lift-drag ratio, increase glide distance and optimize flight trajectory through autonomous deformation.

$$r = d_{new}/(h_0 - h_{new}) - d/(h_0 - h) \tag{7}$$

4 Simulation Results

4.1 Simulation Tests Environment

Before the simulation tests of the deformation decision optimization design method, the task environment should be constructed firstly. The aircraft begin to glide unpowered from an altitude of 60 km, the flight trajectory of the aircraft descending to the same height before and after the execution of the deformation decision module was compared. The feasibility verification of the decision scheme was completed by judging whether the trajectory characteristics were optimized or not. The state observation variables of the decision network model include flight speed (v), flight altitude (h), glide distance

(d) and attack angle (α). The action variable output is the sweep angle (α_{fold}) on both wings. The next step is to initialize all the variables. The initial sweep angles α_{fold} are any values within the range of 0 to 5°, the initial flight speed v_0 is 3 km per second, the initial altitude h_0 is 60 km, and the glide distance is calculated from 0.

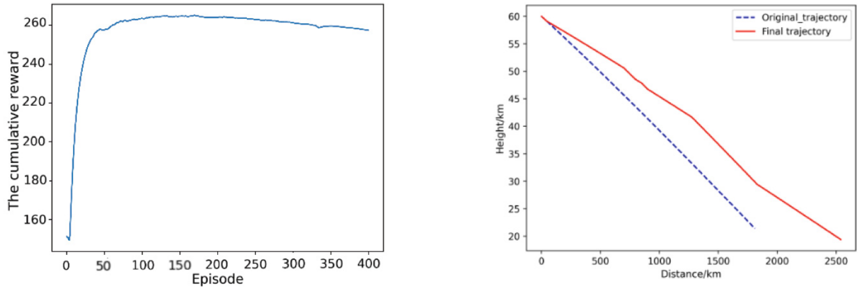
4.2 Test Results of Different Algorithms

(1) Test results based on PPO algorithm

Firstly, the convergence of the algorithm model is verified. The model was made to complete 400 times of interactive training. Each training made the aircraft to do unpowered glide at the same initial speed from the altitude of 60 km, during which the deformation was carried out independently according to the constructed decision scheme. After continuous descent to 20 km altitude, the training is finished and the cumulative reward changes are observed. It can be seen from the change curve of cumulative rewards in Fig. 7(a) that, after 50 times of interactive training, the cumulative rewards of the decision model built based on PPO algorithm increase rapidly and converge gradually after completing the whole flight path. Then the feasibility of the deformation decision scheme is verified. Considering that there are many possibilities of the actual flight path, the optimal deformation angle of the aircraft is unpredictable. Therefore, the effectiveness of the deformation decision can be judged by observing the flight path of the entire glide stage. Let the aircraft glide down to the same height, and compare the trajectory curve before and after the deformation decision. In order to simplify the motion equation of the flight model, the aircraft moves in a straight line. As shown in Fig. 7(b), the blue dashed line is the flight path of the aircraft completing the entire glide stage in a fixed form, and the red solid line is the optimized flight path after the trained decision model completes the decision deformation online according to the real-time flight state. By comparison, it can be seen that when the aircraft drops to the same height, the glide range can be increased by more than 700 km by using the optimized deformation decision after training. It can be concluded that the PPO algorithm network framework adopted in this paper can effectively solve the deformation decision-making problem with unknown optimal indexes. After the intelligent aircraft completes the optimization learning of autonomous deformation decision-making, the aircraft can realize the optimization and promotion of flight trajectory by means of autonomous deformation decision-making.

(2) Test results based on DDPG algorithm

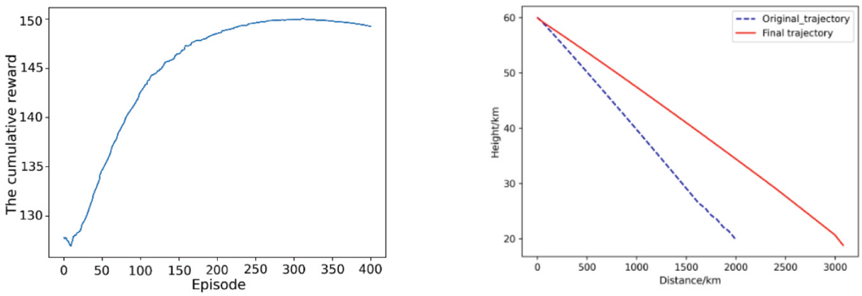
In the process of testing the decision training model constructed using DDPG algorithm, the convergence of the model and the effectiveness of the decision were also verified, and the training model constructed by the two algorithms was further compared. In the decision model based on DDPG algorithm, the cumulative reward gradually converges after nearly 250 scenes. Compared with the PPO algorithm decision model with the same network structure parameters, the convergence rate of cumulative rewards is significantly slowed down, so the learning efficiency of DDPG algorithm decision model is low. Then the feasibility of the deformation decision scheme is verified and compared with the PPO algorithm decision model again. Let the aircraft start to glide unpowered from the altitude of 60 km and until it descends to 20 km. By comparing



(a) The training curve of cumulative rewards (b) Flight trajectory comparison

Fig. 7. Simulation test results based on PPO algorithm

the flight paths before and after using the deformation decision module, it can be seen that when the aircraft glided down to the same height, using the deformation decision module to implement online deformation decision can increase the glide distance by nearly 1000 km, which is more optimized than the PPO algorithm under the same structure. It is concluded that the decision network model constructed by DDPG algorithm can also realize the online autonomous deformation decision of intelligent aircraft after interactive trainings, and achieve the purpose of improving flight trajectory performance (Fig. 8).



(a) The training curve of cumulative rewards (b) Flight trajectory comparison

Fig. 8. Simulation test results of DDPG algorithm

5 Conclusions

This paper takes intelligent morphing aircraft flying at high speed in large airspace as the research object, aiming at the technical problems that the aircraft is difficult to obtain sufficient deformable flight test data in advance, which leads to the difficulty in predicting the optimal aerodynamic shape under different flight states. And the traditional control method needs to build accurate controlled model, so it is difficult to apply to the control problem of morphing aircraft with large model changes. A scheme design based on RL network is proposed to optimize the flight trajectory by changing the aircraft's structure through autonomous decision. The flight trajectory is taken as the optimization design index of deformation decision, and the trajectory parameters are used to construct the reward function. The autonomous learning of RL network model is guided by reward growth, so as to complete the optimization learning of deformation time decision and deformation quantity decision, and realize the purpose of optimizing flight trajectory through autonomous deformation. In order to verify the feasibility of the design scheme and avoid the influence of different algorithms on the results, two RL algorithms with different strategies are used in the simulation experiment to build the decision network model with the same structure and parameters. The experimental results show that the decision models constructed by the two algorithms can both optimize the flight trajectory through online autonomous deformation, which proves that the scheme is effective and feasible.

References

1. Chu, L., Qi, L.I., Feng, G.U., et al.: Design, modeling, and control of morphing aircraft: a review. Chin. J. Aeronaut. **35**(5), 220–246 (2022)
2. Xu, W., Li, Y., Lv, M., et al.: Modeling and switching adaptive control for nonlinear morphing aircraft considering actuator dynamics. Aerosp. Sci. Technol. 122, 107349.1–107349.17 (2022)
3. Kan, Z., Li, D., Zhao, S., et al.: Aeroacoustic and aerodynamic characteristics of a morphing airfoil. Aircr. Eng. Aerosp. Technol. 93, 888–899 (2021)
4. Zhang, C., Zhang, Y., Wei, C.: Onboard trajectory generation of hypersonic morphing aircraft. Int. J. Aerosp. Eng. (2021)
5. Shi, Y., Wang, Z.: Onboard generation of optimal trajectories for hypersonic vehicles using deep learning. J. Spacecr. Rocket. **58**(16), 1–15 (2020)
6. Wang, Z., Li, H., Wu, Z., et al.: A pretrained proximal policy optimization algorithm with reward shaping for aircraft guidance to a moving destination in three-dimensional continuous space. Int. J. Adv. Rob. Syst. **18**(1), 1–9 (2021)
7. Zavoli, A., Federici, L.: Reinforcement learning for low-thrust trajectory design of interplanetary missions. J. Guid. Control. Dyn. **44**(8), 1440–1453 (2021)
8. Zhang, H., Xu, J., Qiu, J.: An automatic driving control method based on deep deterministic policy gradient. Wirel. Commun. Mob. Comput. **2022**(12), 1–9 (2022)
9. Xu, D., Hui, Z., Liu, Y., et al.: Morphing control of a new bionic morphing UAV with deep reinforcement learning. Aerosp. Sci. Technol. **92**, 232–243 (2019)
10. Jesus, J.D., Bottega, J.A., Antonio, D., et al.: Deep deterministic policy gradient for navigation of mobile robots. J. Intell. Fuzzy Syst. Appl. Eng. Technol. **40**(1), 349–361 (2021)