Research On Continuous Obstacle Avoidance Control of UAV Tracking Target Based on PID-DDPG Control Method

Shiyu Wang

Department of Energy and Power Engineering, Dalian Maritime University, Dalian, China

13699899849@163.com

Abstract. UAV tracking and obstacle avoidance are important functions of UAVs, and at present, the control system of mainstream UAVs adopts PID control system. However, the traditional PID control can no longer meet the current requirements for faster and more accurate control of UAVs. In this paper, we aim to combine the traditional PID control with deep learning of artificial intelligence to realize the fast tracking and precise obstacle avoidance of the target of UAV in a complex environment. This paper designs a PID-DDPG control model, theoretically, the UAV will prioritize the use of PID control system for fast approach to the selected target, and then switch to the DDPG algorithm to accurately track the target and achieve obstacle avoidance to the environmental obstacles, and ultimately achieve fast tracking and accurate obstacle avoidance to the target in complex environments. Compared with the traditional PID control, the PID-DDPG control system can realize more precise control and faster response to the UAV. At the same time, compared with the pure DDPG deep learning method, this system only needs less learning times and a smaller computation amount.

Keywords: UAV; reinforcement learning; PID control; obstacle avoidance.

1. Introduction

UAV is the use of radio remote control equipment and self-contained program control device manipulation of unmanned aircraft, in recent years, with the UAV application more widely, the UAV from the military to industrial into the civil field, the UAV target tracking ability also has a higher requirement. Therefore, UAV obstacle avoidance needs to solve the advantages and disadvantages of the acquisition path and establish the corresponding indicators for evaluation [1] so as to select a more efficient tracking and obstacle avoidance algorithm. Yang [2] used the traceless Karman filtering algorithm to establish a moving target tracking model and then utilized the binary agreement algorithm to make the UAVs agree among themselves on the target tracking, reducing the probability of target omission or tracking errors. Cheng [3] and others added neural network structure based on PID controller, applying genetic algorithm to optimize neural network PID controller, so as to improve the response speed of UAV. Guo [4] and others design a new type of automatic antenna tracking control system based on GPS and microcontroller. By comparing the three algorithms of GPS tracking in detail, it is determined that the orb model algorithm can completely replace the planar model and the ellipsoid model for a certain range of tracking accuracy, which reduces the amount of computation and improves the computational accuracy.Meanwhile,the rapid development of artificial intelligence [5] makes AI take an increasingly important role in UAV control system. Especially in a city with tall buildings and dense woods, when the UAV path planning has multiple conditional constraints to track and avoid obstacles to the target, the above algorithms have high time complexity and local minimal traps thus leading to slow response time or locking the UAV at one point.

In order to solve these problems, artificial intelligence is combined with the UAV, and deep reinforcement learning is used to provide the UAV with an experience pool, so that when the UAV encounters path planning problems, it can directly look for training experiences from the experience pool, so as to solve the complex control problems [6]. If today's PID algorithm is directly used to approach the target, the target can be quickly approached in a two-dimensional plane, which ensures the stability of the UAV's flight while controlling the UAV to smoothly arrive at the target point to complete all inspection tasks based on the planned path to ensure that the UAV completes the task

safely and quickly [7]. The depth deterministic policy gradient algorithm algorithm allows the UAV to achieve accurate obstacle avoidance while preventing local lockup. Reinforcement learning will give the UAV rewards based on the UAV's actions in the environment, and when the UAV obtains the maximized rewards, its actions are the optimal solution. In this process the drone tracks the target and avoiding obstacles by repeated attempts. During the training process, the UAV continuously learns experience from the environment and stores it in the memory bank, so that it can select the optimal solution in various environments after a certain number of samples are taken from the memory bank. In this paper, we will study the way of combining the PID control system with the DDPG reinforcement learning algorithm. When the UAV is far away from the target, the PID control system is prioritized to approach the target in a two-dimensional plane. This reduces the amount of computation during path planning, thus allowing the UAV to perform a quick response. The DDPG algorithm is used when approaching the target, and obstacle avoidance maneuvers are performed if there is an obstacle, and the target is tracked and avoided based on the empirical pool. This combination of algorithms can greatly reduce the amount of computation during UAV tracking and obstacle avoidance and increase the UAV reaction speed.

In this article, we will study the tracking path planning and obstacle avoidance of UAVs under complex environments and continuous space. Firstly, establish the model of UAV based on PID control system. Secondly, design the reward function to ensure accuracy and safety when the UAV learns. Finally, build the DDPG algorithm intelligence and combine it with the PID control system UAV model. Compare with existing algorithms to verify the effectiveness of the proposed method.

2. Modeling UAVs and Intelligentsia

2.1. UAV modeling

UAV mainly consists of airframe, power supply system, flight control, motor, sensor and signal receiver, this paper is mainly based on MATLAB system to establish a UAV flight control system. There are six degrees of freedom in the UAV flight process, which are forward and backward, up and down, left and right movement (1); traverse roll, pitch and yaw rotation (2).

$$\begin{cases} \mathrm{m}\dot{\mathrm{v}}_{\mathrm{x}} = \mathrm{F}_{\mathrm{t}}\left(\cos\Phi\cos\Psi\sin\theta + \sin\Phi\sin\Psi\right) \\ \mathrm{m}\dot{\mathrm{v}}_{\mathrm{y}} = \mathrm{F}_{\mathrm{t}}\left(\cos\Phi\sin\theta\sin\Psi - \cos\Psi\sin\theta\right) \\ \mathrm{m}\dot{\mathrm{v}}_{\mathrm{z}} = -\mathrm{mg} + \mathrm{F}_{\mathrm{t}}\cos\theta\cos\Phi\,; \end{cases} \tag{1}$$

$$\begin{cases} \mathrm{m}\dot{\mathrm{v}}_{\mathrm{z}} = -\mathrm{mg} + \mathrm{F}_{\mathrm{t}}\cos\theta\cos\Phi \\ \dot{\theta} = \mathrm{q}\cos\theta - \mathrm{r}\sin\theta \\ \dot{\Psi} = \frac{1}{\cos\theta}\left(\mathrm{r}\cos\Phi + \mathrm{q}\sin\Phi\right) \\ i_x\dot{p} = (i_y - i_z)qr + \tau_p \\ i_y\dot{q} = (i_z - i_x)pr + \tau_q; \\ i_z\dot{r} = (i_x - i_y)pq + \tau_r \end{cases} \tag{2}$$

The rotation module is modeled in MATLAB as shown in Fig. 1, and according to the formula in (2) the module input$\tau_p;\ \tau_q;\ \tau_r$; p; q;r and integrates to derive the output p;q;r, which is the rotational inertia of the UAV

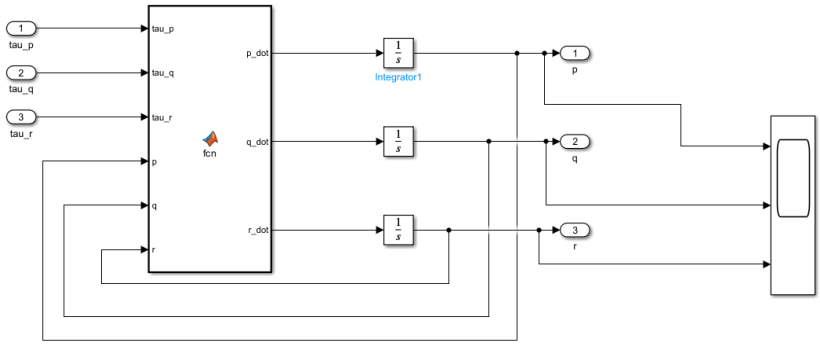


Fig. 1 Rotating module

2.2. PID Control

PID control algorithm is a combination of proportional, integral and differential three links in one control algorithm, it is the most mature technology in the continuous system, the most widely used a control algorithm. the essence of PID control is based on the input deviation value, follwoing the proportional, integral and differential function of the relationship between the operation, the result of the operation is used to control the output. The traditional PID controller is shown in Figure 2, the control system is through the deviation signal to adjust the proportional, integral and differential functions, so as to maintain the system stable output.

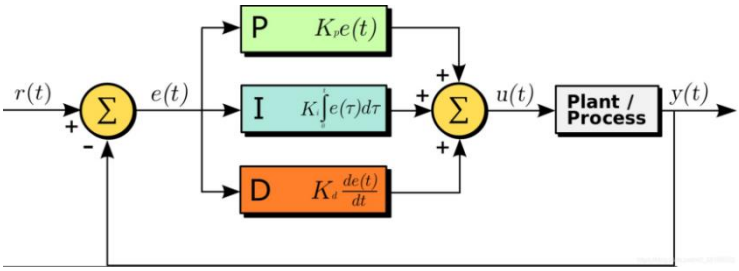


Fig. 2 PID controller

The PID control system is modeled in MATLAB as in Fig. 3, the proportional parameter is adjusted to take the value of 0.15, the integral parameter is adjusted to take the value of 0.01, the differential parameter is adjusted to take the value of 0, and the three are added together to form the final control parameter.

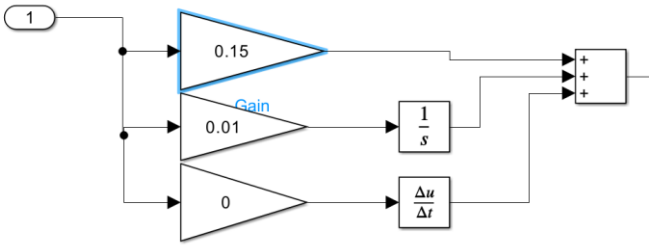


Fig. 3 PID control module

Fig. 4 shows an overview of the PID control system of the UAV, where each of the six parameters x;y;z;Φ;θ;Ψ are PID regulated and used to control the UAV.

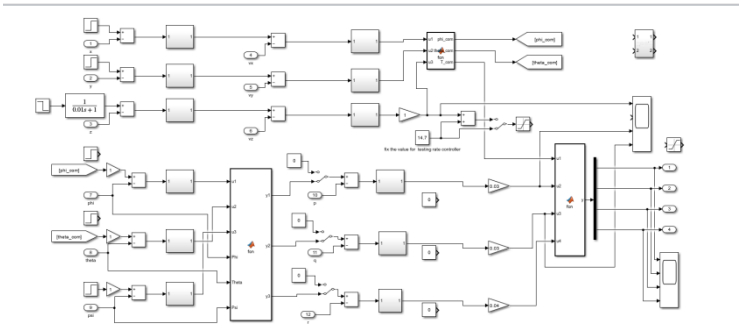


Fig. 4 PID control system

UAV and PID control system model shown in Fig. 5, after PID regulation of the parameters input to the motor, motor output voltage control UAV action, motor output voltage parameters by eliminating noise back to the UAV motion module, continue to output signals to the PID control system, and so on.

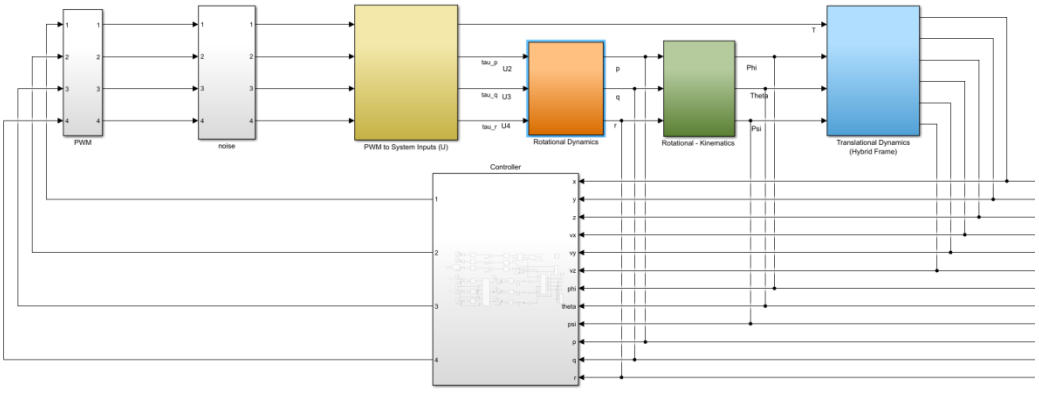


Fig. 5 UAV control model

2.3. Reinforcement Learning and the Design of Reward Functions

In order to apply the DDPG intensive learning algorithm, I regard the UAV as an agent that can perceive the environment state and output the action decision, the action a of the agent is the corresponding angular velocity; the environment states is the relevant environmental information that needs to be fed back to the agent in order to achieve the goal; the reward function R is the reward obtained by the UAV agent in the process of interacting with the environment, and the design of the reward function is very important, which can directly affect the convergence of the learning process later on. The design of the reward function is very important and can directly affect the convergence of the learning process.

2.3.1 Design of Reward Functions

1.Set the distance between the drone and the target point as a penalty to incentivize the drone to travel to the target point. When the drone reaches the target area, it receives a bonus of 5. (3)

2.To get the drone to the target point faster, the drone gets a -0.04 penalty per time step (4)

3.In order for the UAV to approach the target direction, the negative value of the angle(β) between the direction of view of the UAV and the line between the UAV and the target is used as a penalty. (5)

4.In order for the drone to avoid obstacles, I define the distance between the drone and an obstacle as Dsafe. when the drone is too close to an obstacle and the distance is less than Dsafe, the drone receives a penalty; when the drone collides with an obstacle, the penalty value is -5; and when the distance between the drone and the surrounding obstacle is greater than Dsafe, it means that the obstacle and the drone are at a safe distance, and the drone can continue to fly to the target point.(6)

$$\text{distance reward} \quad \text{rdis} \begin{cases} 5, \text{if } d_0 < 10 \\ -0.1 * \text{normalize}(d_0), \text{else} \end{cases} \tag{3}$$

$$\text{time penalty} \quad \text{rtim=-0.04} \tag{4}$$

$$\text{angle penalty} \quad \text{rang=-0.04*}\beta/\pi \tag{5}$$

$$\text{obstacles penalty} \quad \text{robs=}\{\begin{cases} -0.1 * (D_{safe} - d_{min}), \text{if } d_{min} < d \\ -5, \text{if collides with obstacles} \\ 0, \text{else} \end{cases} \tag{6}$$

Thus, the reward function is defined as R=rdis + rtim + rang + robs

2.4. DDPG reinforcement learning algorithm.

DDPG is a model-free reinforcement learning algorithm based on the framework which can achieve continuous policy output without the need for accurate modeling of the environment. It can cope with the problem that the agent needs to output continuous actions, which is more in line with the continuous nature of actions in the state space.The Actor network outputs the current state information to give the deterministic action policy based on the current state information, in which the accumulated rewards are related to the actions, and thus the parameters of the Actor network can be updated by gradient ascent.The Critic network, through its neural network parameters, gives the evaluation of the actions based on the current state in the form of Q-values. gives the evaluation of the action based on the current state in the form of Q-value, so the accurate Q-value plays a vital role in the convergence of the update of the network.The Critic network is updated by minimizing the loss function as shown in Fig.6 and makes its evaluation value more accurate.

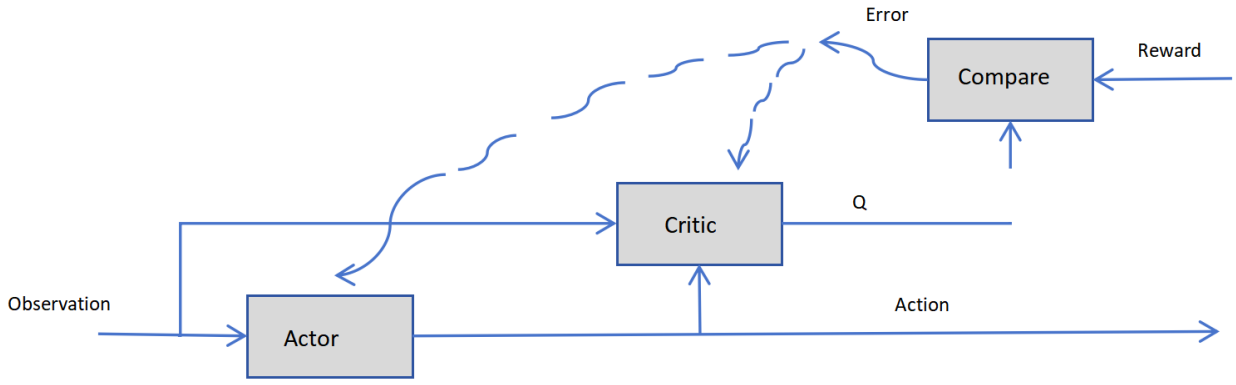


Fig. 6 ACTOR and CRITIC Workflow

Subsequently the DDPG algorithm model is created to be combined with the UAV as shown in Fig. 7, the UAV outputs the reward value to the AGENT based on the reward function at the same time the PID control system also outputs the value to the AGENT.The DDPG-AGENT is fed into the ACTOR and CRITIC network based on the input of the reward value and the observation value after comparison, updates the experience pool and outputs the action value to the UAV

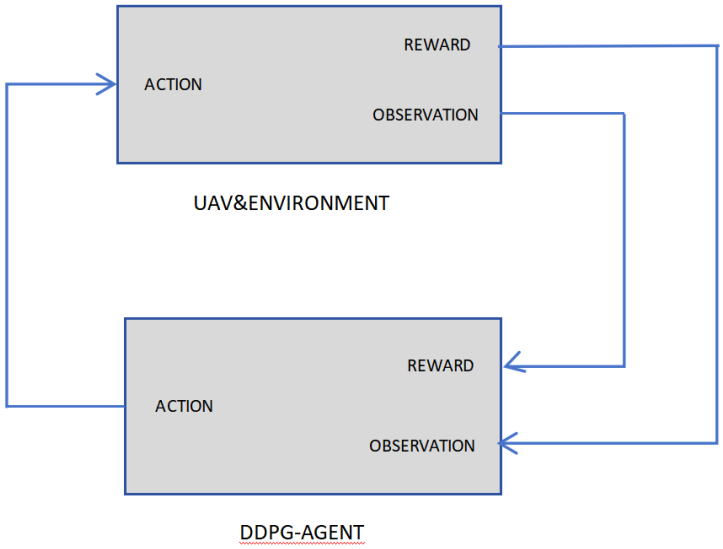


Fig. 7 Flow of DDPG algorithm combined with UAV.

3. Conclusion

The research in this paper is to combine the PID control system with DDPG deep learning, so as to achieve faster reaction speed and more accurate control of the UAV to realize target tracking as well as avoiding obstacles. In this experiment, the traditional PID control system is used to collect data, the object is the input signal and output signal of the PID control system to the UAV, and the collected data is used for the training of deep learning. After reviewing Xu and Yang's approach to the traditional PID control method to control the UAV attitude angle and motor speed, in order to predict the specific position and speed information of the UAV proposed R-CNN algorithm and combine with the PID control, so as to improve the UAV immediate state prediction into the accuracy. Xu and Yang's paper shows that combining PID with deep learning algorithms in the UAV field is feasible and superior to traditional PID control systems. In addition to the DDPG deep learning method it is also possible to combine other deep learning methods with the PID traditional control system and perform simulations to determine whether they are superior to the PID control method. After the research in this paper, in future work, we can find and determine the specific UAV parameters and carry out simulation experiments, use the PX4 flight control software to establish a simulated obstacle environment, set up the target point, let the UAV under the control of the PID-DDPG control system to track the target and obstacle avoidance, and provide timely feedback data to the AGENT so as to update the experience base, so as to optimize the UAV's path planning every time. The UAV will be able to optimize each path planning of the UAV. At the same time, the experimental data are compared with the traditional PID control method after the simulation, so as to determine that the PID-DDPG control method is experimentally superior to the traditional control method.

References

[1] H. H. Zhang, X. S. Gan, Y. Mao, et al. A review of obstacle avoidance algorithms for UAVs[J]. Aviation Arms,2021,28(05):53-63Cheng Yibin,Tan Zhiyin. Analysis of UAV flight tracking control based on improved neural network PID control[J]. Journal of Lanzhou Institute of Technology,2018,25(05):69-72.

[2] CHENG Yi-Bin,TAN Zhi-Yin. Analysis of UAV flight tracking control based on improved neural network PID control[J]. Journal of Lanzhou Institute of Technology,2018,25(05):69-72.

[3] Yang Hui. A cooperative tracking algorithm for multiple moving targets based on remote sensing UAV[J]. Automation Applications,2022(08):1-3+8.DOI: 10.19769/j.zdhy.2022.08.001.

[4] CHENG Yi-Bin,TAN Zhi-Yin. Analysis of UAV flight tracking control based on improved neural network PID control[J]. Journal of Lanzhou Institute of Technology,2018,25(05):69-72.

[5] GUO Qianqian,ZHOU Xiaohui. Automatic tracking system of UAV antenna based on GPS[J]. Computer Application and Software,2017,34(06):85-90.

[6] JORDAN M I, MITCHELL T M. Machine learning: trends, perspectives, and prospects [J]. science, 2015, 349(6245):255260

[7] KOBER J, BAGNELL J A, PETERS J. Reinforce ment learning in robotics: A survey ［J］. International Journal of Robotics Research, 2013, 32 (11): 12381274