Scientific Journal of Intelligent Systems Research Volume 3 Issue 11, 2021
ISSN: 2664-9640

Multi-UAV Task Assignment Method under Dynamic Threat Target Condition

Hangfei Wang, Xunli Zhang, Ming Li, Qi Dai, Junchao Zhao

Rocket Army Engineering University, Xi'an, 710038, China

Abstract

Multi-UAV cooperative combat application scenarios usually have a variety of dynamic threat factors, such as the threat range of hostile targets and target positions are dynamically changing. In order to give full play to the advantages of coordination among multiple drone clusters in a dynamic environment to maximize the mission efficiency of strike targets, Bellingham et al. used the MILP algorithm to study the path planning of multiple drones in a dynamic and uncertain environment. problem. Alighanbari et al. proposed an RHTA algorithm for re-planning in a dynamic environment. However, such algorithms all need to re-optimize the calculation and update the results in time to deal with complex changes in the environment, which has the disadvantages of high computational cost and poor real-time performance.

Keywords

Multi-UAV Task Assignment Method, Dynamic Threat Target Condition.

1. Introduction

Traditional multi-UAV coordinated target assignment and path planning algorithms often have complex design rules and require the use of experience and prior knowledge. It is difficult to provide stable, continuous and good battlefield adaptability in complex environments. Compared with traditional methods, this paper proposes to solve the problem of dynamic threat targets based on the MADDPG framework, and without obtaining all the information of the enemy's targets, it can realize the autonomous learning task allocation and path planning strategies of the drone in the environment. The core idea of the algorithm is that the goal is to gradually form an expectation of the stimulus under the stimulation of the reward or punishment given by the environment as the agent, and then produce the habitual behavior that can obtain the maximum benefit, and finally achieve the maximum expected benefit as the optimization goal. . The multi-agent system can be effectively trained to adapt to the dynamic battlefield environment.

2. Analysis of Multi-UAV Task Assignment for Dynamic Threat Targets

This paper conducts research based on a two-dimensional plane environment and considers the combat scenarios shown in Figure 4.1: (1) UAV formations can reach targets distributed in different locations; (2) There are some unknown threat areas that UAVs cannot reach; (3) It is necessary to avoid collisions between drones. Based on the above scenarios, this paper conducts mathematical modeling on the target allocation and path planning of multiple UAVs.

Assume that a group of UAV formations have a total of UAVs as follows $U_1, U_2, \ldots, U_k$:. The state of each drone includes the current speed vector and the position of the drone in the environment. There are a total of goals $(v_{ui,x}, v_{ui,y})$:, the position of the th goal. There are a total of threat areas, the coordinate of the first threat area is, and the radius of the threat area is. The action space of the drone and the target is a two-dimensional continuous space. The action strategy is to give the drone instantaneous speed and the instantaneous speed of the dynamic target at each moment. After the time has passed, $(v_{u,x}, v_{u,y})$ the positions of the drone and the dynamic target are updated to,, As shown in formula (1) (2)

$$\begin{cases} p_{i,x}^{t+\Delta t} = p_{i,x}^{t} + v_{u,x} \times \Delta t \\ p_{i,y}^{t+\Delta t} = p_{i,y}^{t} + v_{u,y} \times \Delta t \end{cases} \tag{1}$$

$$\begin{cases} M_{n,x}^{t+\Delta t} = M_{n,x}^{t} + v_{T,x} \times \Delta t \\ M_{n,y}^{t+\Delta t} = M_{n,y}^{t} + v_{T,y} \times \Delta t \end{cases} \tag{2}$$

3. Multi-agent reinforcement learning algorithm

The research progress of multi-agent reinforcement learning provides a new solution to the problem of multi-UAV target allocation and path planning. The MADDPG[71] algorithm performs well in multi-agent collaboration, competition or mixed environments. The problem of multi-UAV target allocation and path planning is essentially a mixed scenario of competition and cooperation. Competition means that each drone is assigned a unique target, and cooperation means that drone formation minimizes the total range. Because the environmental information (threat area) is randomly generated in each training round and the target is dynamic, the MADDPG model can handle the dynamic environment after training, and the model's ability to adapt to environmental changes becomes stronger as the training round increases .

Based on the above ideas, this paper proposes a dynamic target approach algorithm (DTA-MADDPG) based on the MADDPG framework to provide a solution for multi-UAV coordinated target allocation and path planning. In this scheme, the multi-UAV target assignment and path planning problems are integrated into a multi-agent system, which is trained by DTA-MADDPG in a dynamic environment. This method describes the problem of multi-UAV target allocation and path planning as a Markov decision process. The actions of the drone and the target are discrete into many time steps. After each step, the drone formation and environmental conditions are treated as a state. The flight action selected by each drone is only related to the current state. Each drone can observe the current environment, and then obtain the next action from its own strategy network.

3.1. MADDPG algorithm framework design

MADDPG is developed based on the DDPG algorithm [72]. The training of each agent is similar to the training process of a single DDPG algorithm. The difference is mainly reflected in the input of the Critic network. For example: in the DDPG algorithm of a single agent, the input of the Critic network is a state-action pair information, while in MADDPG, the input of each agent's Critic network can obtain the rest in addition to its own state-action information. The action information of the agent is subjected to centralized training and decentralized execution, that is, during training, information that can observe the global situation is introduced to guide Actor training [72], and only Actors with local observations are used to take actions during testing. In fact, MADDPG can be regarded as a multi-agent version of DDPG. Its core idea is focused training and decentralized execution. Because Deep Q-learning and DDPG do not use information from other agents, they perform poorly in a multi-agent environment. The MADDPG algorithm performs well in a multi-agent environment by using the observation information and action information of other agents.

3.2. Construction of Multi-UAV Task Assignment System for Dynamic Threat Targets Based on MADDPG

The multi-UAV target allocation and path planning model described above shows that there is a competition and cooperation relationship between multiple UAVs. This model can be solved by establishing a multi-intelligence system through the MADDPG framework. The system abstracts each drone as an agent, which has the same action mode as the real drone.This article considers a scenario with an agent. The strategy adopted by the agent is parameterized as, assuming the deterministic strategy of all agents. The gradient of the deterministic strategy for the agent is

Where represents the state of the agent, is a value function, is the action of the agent, and is the observed value of the agent, which includes the distance between the agent and each target and the distance between the agent and obstacles. The representative experience pool contains a series of tuples to record all agent training samples. It is the new state after the agent performs the action, and is the reward value of the first agent. The loss function used to update the Critic network is

4. Simulation Experiment Analysis of Multi-UAV Task Assignment for Dynamic Threat Targets

In this paper, a multi-agent reinforcement learning simulation training environment is designed for multi-UAV task allocation and path planning based on the OpenAI platform. The simulation environment includes drone targets and threat zones. The coordinate system is established with the geometric center of the environment as the origin, the size of the agent is set to 0.5, the size of the target is 0.8, the size of the threat area is 2, the speed of the drone is set to 5, and the target moving speed is the critical area and the size is 0.1 Threat area The position of the target and the position of the target are randomly generated in each training.

In order to effectively measure the effect of multi-agent reinforcement learning on task allocation training of multiple drones in complex environments, this paper designs collision rewards and the collision rate between agents and obstacles in different training stages, including agents and obstacles. The collision rate of objects in the entire training cycle, and the collision rate of agents and obstacles in the entire training round. The calculation formula is as follows:

Collision rate between drones and threat zone = number of collisions between drones and obstacles / total number of rounds in training

Collision rate between drones = number of collisions between drones and drones/total number of rounds in training

In order to analyze and compare the impact of the number of drones and threat zones in environmental parameters on algorithm training, different numbers of unmanned aerial vehicles (UAV) and threat zones (Ob) were set up in the DTA-MADDPG model for training, and the training process was counted. Two aspects of information: the reward value under different environmental parameters is shown in Figure 4.6(a), and the collision rate between the drone and the threat area is shown in Figure 4.6(b). Figure 4.6(a) intuitively shows the influence of different numbers of drones and threat areas on the convergence of the algorithm. It can be seen from the figure that when the number of drones remains the same and the number of threat areas in the environment is increased, the training process does not The reward convergence process obtained by humans and machines is basically unaffected; in the same number of threatened area scenarios, increasing the number of drones will slow the algorithm convergence. After convergence, the reward value of the two has a large gap. This is due to each aircraft. UAVs need to train 4 neural networks. As the number of UAVs increases, the number of network training will also increase, and the training time will also become longer. From Figure 4.6(b), it can be seen that with the increase of training rounds, the collision rate between the drone and the threat area is rapidly decreasing for 2 drones and 3 drones, indicating that the DTA-MADDPG algorithm can be used to make The drone quickly explores and autonomously avoids the threatened area.

Reward value under different environmental parameters (b) Collision rate between drones and obstacles under different environmental parametersFigure Curves of reward value and collision rate of DTA-MADDPG with different environmental parametersAfter running different typical environments and different parameters several times, by comparing (a) and (b) in it is found that DTA-MADDPG is more sensitive to the number of agents in the environment, and it converges when the number of drones is 2. Better, as the number of agents increases later, it will reduce the training performance of the model. Relatively insensitive to the number of threat areas, increasing the number of threat areas will not significantly affect the model training performance.

5. Chapter summary

Aiming at the problem of poor adaptability of multi-UAV battlefield environment in dynamic threat target scenarios, this paper adopts a multi-agent reinforcement learning algorithm to solve the multi-UAV task assignment problem. The main work is as follows:

Constructed a multi-UAV task allocation model in a dynamic threat target scenario, and studied the multi-UAV task allocation problem based on multi-agent reinforcement learning algorithm; 2) Constructed a multi-UAV task allocation problem in multi-intelligence The solution framework in the reinforcement learning algorithm; 3) The rewards and rewards are reasonably designed for the problem of dynamic target allocation and autonomous threat avoidance; 4) The comparison of simulation results shows that the multi-agent reinforcement learning algorithm with decision-making ability is effective in dynamic and complex environments. There are potential advantages in solving the multi-UAV task assignment problem.

References

[1] Jia Z, Yu J, Ai X, et al. Cooperative multiple task assignment problem with stochastic velocities and time windows for heterogeneous unmanned aerial vehicles using a genetic algorithm[J]. *Aerospace Science & Technology*, 2018, 76: 112-125.

[2] Ramirez-Atencia C, Bello-Orgaz G, R-Moreno M D, et al. Solving complex multi-uav mission planning problems using multi-objective genetic algorithms[J]. *Soft Computing*, 2017, 21(17): 4883-4900.

[3] Ramirez-Atencia C, Mostaghim S, Camacho D. A knee point based evolutionary multi-objective optimization for mission planning problems[C]. *Genetic & Evolutionary Computation Conference*. ACM, 2017.

[4] Pendharkar P C. An ant colony optimization heuristic for constrained task allocation problem[J]. *Journal of Computational Science*, 2015, 7: 37-47.

[5] Zong Q, Qin X, Zhang B. Cooperative task assignment of large scale ucav based on dpso-gt-sa algorithm[J]. *Journal of Tianjin University*, 2018, 51(10): 1005-1014.

[6] Xu D，Hui Z，Liu Y Q，et al． Morphing control of a new bionic morphing UAV with deep reinforcement learning[J]． *Aerospace Science and Technology*，2019，92: 232-243.

[7] Lowe R，Wu Y, Tamar A, et al． Multi-agent actor-critic for mixed cooperative-competitive environments[J]. *Advances in Neural Information Processing Systems*． 2017: 6379-6390.