1v1 Air Combat Maneuver Decision Using Minimax-DQN

Shuhui Yin[1], Yu Kang[1,2], Yunbo Zhao[1], Jian Xue[3]

1. Department of Automation, University of Science and Technology of China, Hefei 230026, China
E-mail: shuhui@mail.ustc.edu.cn, kangduyu@ustc.edu.cn, ybzhao@ustc.edu.cn

2. Institute of Advanced Technology, University of Science and Technology of China, Hefei 230088, China

3. College of Engineering and Information Technology, University of Chinese Academy of Sciences, Beijing 100049, China
E-mail: xuejian@ucas.ac.cn

Abstract: Autonomous maneuver decision of UAV is the key to realize future air combat. Considering the strong antagonism and the uncertainty of opponents in air combat, traditional methods are difficult to solve Nash equilibrium strategy. This paper proposes a Minimax-DQN algorithm combining deep reinforcement learning with game theory, which solves the problem that the single agent reinforcement learning algorithm is difficult to converge due to the unstable environment. The effectiveness of the proposed algorithm is verified by the convergence test.

Key Words: Air Combat, Reinforcement Learning, Game Theory, Maneuver Decision, Minimax-DQN

1 Introduction

Modern military operation is information war, and air combat game is the main means to obtain air combat superiority. As an important part of future air combat, uav greatly reduces the risk of mission execution and improves the operational efficiency compared with manned man-machine. The traditional control method of UAV is remote monitoring, which has poor real-time performance. In 2016, Alpha artificial intelligence based on a genetic fuzzy system developed at the University of Cincinnati beat Keane, a veteran retired Air Force colonel, in an air combat simulation. Lee. This means that drone intelligent decision-making systems are slowly outpacing human decision-making. Gradually with the development of artificial intelligence, reinforcement learning method is applied to decision making problem solving, reinforcement learning method is to adopt the method of trial and error to interaction with the environment, the reinforcement learning were characterized by markov decision process, calculation of cumulative returns after the current condition to perform an action is worth size to determine maneuver choice as a result, Therefore, reinforcement learning not only considers the influence of the current state and battlefield environment, but also considers the long-term influence of maneuvering actions, which can well meet the antagonism and uncertainty in the process of air combat. In addition, reinforcement learning does not need samples and only needs to evaluate the benefits generated by maneuvering. Therefore, reinforcement learning is a better modeling method for maneuvering autonomous decision making.

*This work was supported by the National Key Research and Development Program of China (No. 2018AAA0100801).

2 Model

2.1 Two-player Zero-sum Markov Game

In this section, we construct a two-player zero-sum game model for 1v1 air combat task.

Two-player zero-sum Markov game is an extension of Markov decision progress combined with zero-sum matrix game. The Markov decision process is a multi-process decision-making theory of single agent. The agent continuously interacts with its environment and gets feedback, and the agent making actions according to the feedback to optimize its benefits. The zero-sum matrix game describes a two-player static zero-sum game. The static means that both players make actions at the same time. The zero-sum means that the sum of the payoff functions of two players is zero. We can obtain a two-player multi-process dynamic decision-making model combining the two, that is, Markov game model, which can be defined as a quintile (S, A, O, T, R):

(1) S: environment state space. $\sqrt{b^2-4ac}$

(2) A: agent action space $\sqrt{b^2-4ac}$.

(3) O: opponent action space.

(4) T: S, which represents the transition probability function from one state to another state: In the formula, represent the state of the environment. represent the actions of the agent and opponent respectively. Represents the conditional probability.

(5) R: the reward function of agents, which represents the expectation reward from executing the action to the next state in the current state: In the formula, represents the direct reward obtained at the moment. The decision-making basis of the agent is to maximize its own reward, so its goal is to find a strategy, so that the agent can obtain the largest cumulative expectation reward performing actions according to the strategy in the process of the game: where and represent the strategies of the agent and the

opponent, respectively; represents the number of steps from the current moment to the termination moment; , represents the discount factor.

3 Method

3.1 DQN

3.2 Minimax-DQN

4 Experiment and Analysis

5 Conclusion

References

[1] D. Cheng, Controllability of switched bilinear systems, *IEEE Trans. on Automatic Control*, 50(4): 511–515, 2005.

[2] H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1985, chapter 4.

[3] B. Smith, An approach to graphs of linear forms, accepted.

[4] D. Cheng, On logic-based intelligent systems, in *Proceedings of 5th International Conference on Control and Automation*, 2005: 71–75.

[5] D. Cheng, R. Ortega, and E. Panteley, On port controlled hamiltonian systems, in *Advanced Robust and Adaptive Control — Theory and Applications*, D. Cheng, Y. Sun, T. Shen, and H. Ohmori, Eds. Beijing: Tsinghua University Press, 2005: 3–16.