Decision Modeling and Simulation of Fighter Air-to-ground Combat Based on Reinforcement Learning

Yifei wu
College of systems engineering, National University of Defense Technology
Chinawuyifei0821@163.com

Yonglin Lei
College of systems engineering, National University of Defense Technology
Chinayllei@nudt.edu.cn

Zhi Zhu
College of systems engineering, National University of Defense Technology
Chinazhuzhi0915@126.com

Yan Wang
College of systems engineering, National University of Defense Technology
Chinawangyan19a@163.com

ABSTRACT

With the Artificial Intelligence (AI) widely used in air combat simulation system, the decision-making system of fighter has reached a high level of complexity. Traditionally, the pure theoretical analysis and the rule-based system are not enough to represent the cognitive behavior of pilots. In order to properly specify the autonomous decision-making of fighter, hence, we proposed a unified framework which combines the combat simulation and machine learning in this paper. This framework adopts deep reinforcement learning modelling by using the supervised learning and the Deep Q-Network (DQN) methods. As a proof of concept, we built an autonomous decision-making training scenario based on the Weapon Effectiveness Simulation System (WESS). The simulation results show that the intelligent decision-making model based on the proposed framework has better combat effects than the traditional decision-making model based on knowledge engineering.

CCS CONCEPTS

• **Computing methodologies** → Machine learning; Artificial intelligence; Modeling and simulation.

KEYWORDS

Fighter air-to-ground combat, Intelligent decision-making, Deep reinforcement learning, Combat simulation

ACM Reference Format:
Yifei wu, Yonglin Lei, Zhi Zhu, and Yan Wang. 2022. Decision Modeling and Simulation of Fighter Air-to-ground Combat Based on Reinforcement Learning. In *2022 4th International Conference on Image Processing and Machine Vision (IPMV) (IPMV 2022), March 25–27, 2022, Hong Kong, China.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3529446.3529463

IPMV 2022, March 25–27, 2022, Hong Kong, China

ACM ISBN 978-1-4503-9582-3/22/03…$15.00
https://doi.org/10.1145/3529446.3529463

1 INTRODUCTION

In recent years, deep learning (DL) and reinforcement learning (RL), as important research hotspots in the field of machine learning, have achieved remarkable results in the industrial field. With the deepening of research problems, in more and more complex real scene tasks, single deep learning or reinforcement learning methods can not solve the problems well. Therefore, DeepMind, Google's artificial intelligence research team, creatively combines deep learning with perception ability and reinforcement learning with decision-making ability, forming a new research hotspot in the field of artificial intelligence, namely deep reinforcement learning (DRL) [1]. Based on deep reinforcement learning technology, agents are developed to solve decision-making problems under complex conditions, and relevant scientific research work has made good progress [2-4].

With the continuous improvement of aircraft and defense system in capability, integration, automation and speed, the traditional labor-intensive manual planning is no longer applicable. Therefore, the research and development of aircraft intelligent system which can make independent decisions is of great significance for pilots' auxiliary decision-making. At present, some achievements have been made in the research of fighter intelligent decision modeling. For example, Zuo et al [5] proposed an intelligent decision-making method for air combat maneuver based on heuristic reinforcement learning, and used neural network method to learn the reinforcement learning process, so as to accumulate knowledge and realize real-time dynamic iterative calculation of decision sequence in air combat decision-making process; Rand Corporation has studied the problem of aircraft air ground combat decision-making by using reinforcement learning, and achieved preliminary results [6]; Li et al [7] proposed a maneuver decision algorithm based on deep reinforcement learning, which generates effective maneuver for UAV to independently execute airdrop mission in interactive environment. In addition, in terms of the rapidity and convergence of the algorithm, literature [8-9] has done relevant research, which reduces the training time and improves the effectiveness of the results; In the research of simulation platform, literature [10] provides effective data support for the training of air combat intelligent decision model, the verification of intelligent decision algorithm and the deduction and evaluation of tactical scheme by designing an intelligent decision simulation software system for tactical air combat.

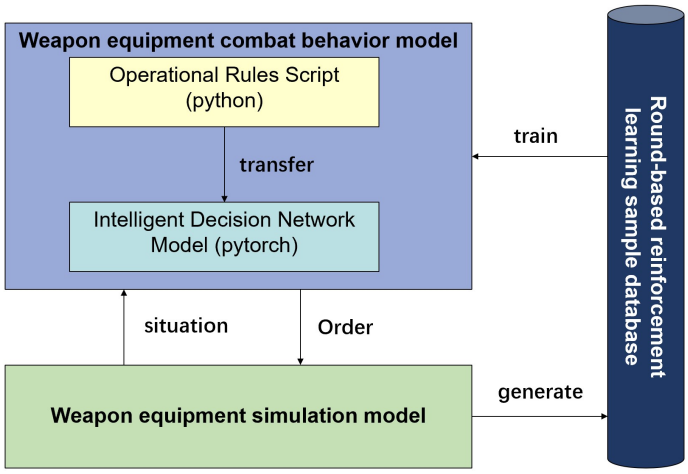

Figure 1: Modeling framework for autonomous decision-making of intelligent equipment based on combat simulation and reinforcement learning.

The continuous progress of intelligent systems, sensor networks and data links means that the complexity of air operations will grow much faster than human reasoning. Due to the complexity of current planning and the greater challenges that may be faced in the future, this project aims to explore the use of reinforcement learning methods for task planning. The key problems include: reinforcement learning modeling design and decision model framework design for aircraft autonomous decision-making problem, design aircraft decision-making network and apply reinforcement learning algorithm combined with task cases, modular design of network simulation experiment in WESS, and simulation experiments to prove the intelligent effect of the network.

2 AIRCRAFT AUTONOMOUS DECISION MODELING FRAMEWORK

2.1 Conceptual framework of intelligent decision-making reinforcement learning modeling

Figure 1 shows the modeling framework of autonomous decision-making for intelligent equipment based on combat simulation and reinforcement learning. Among them, the weapon equipment model is divided into two parts: a simulation model reflecting its physical principle behavior and a behavior model reflecting its combat behavior. The former transmits the situation information in the combat simulation to the latter, and the latter comprehensively makes decisions and uses commands or instructions. The form was issued to the latter. The combat behavior of weapons and equipment is divided into two categories: pre-war planning behavior and real-time decision-making behavior. The former can be flexibly described through scripts and implemented as combat behavior scripts; the latter requires intelligent decision-making based on real-time changes in the situation. The intelligent decision model based on neural network is the object of training modeling using deep reinforcement learning.

During the operation of the combat simulation, the combat behavior script is responsible for calling the relevant intelligent decision-making model at the decision point. Before the call, the situation information is obtained from the weapon equipment simulation model, and converted into the latter's state information required for decision-making. The output of the latter is converted into a weapon and equipment simulation model that can parse and execute the command. This calling process is called cyclically according to the decision cycle, forming a decision loop, until the weapon equipment completes the combat mission and exits the combat process or its simulation operation.

On the other hand, in the intelligent decision-making modeling paradigm based on deep reinforcement learning, the training modeling of the intelligent decision-making model needs the help of combat simulation operation to generate training data. The reinforcement learning training algorithm incorporates the generated sample database into the replay buffer for continuous sampling training, updating the intelligent decision-making network, and loading the updated intelligent decision-making network during the subsequent combat simulation operation to reflect it in the decision loop to affect the combat decision-making behavior.

2.2 WESS-Based intelligent decision-making reinforcement learning training environment

WESS is a general weapon equipment combat simulation system [11], which provides a general weapon equipment simulation model framework for the air, space, sea, fire and other services, and more than 60 combinable simulation model components under the constraints of the model framework. The library can quickly support the rapid development of various weapons and equipment combat simulation applications. The intelligent decision-making reinforcement learning training environment for weapons and equipment based on WESS is shown in Figure 2

Among them, the WESS combat scenario generation tool is responsible for describing and generating various possible scenarios

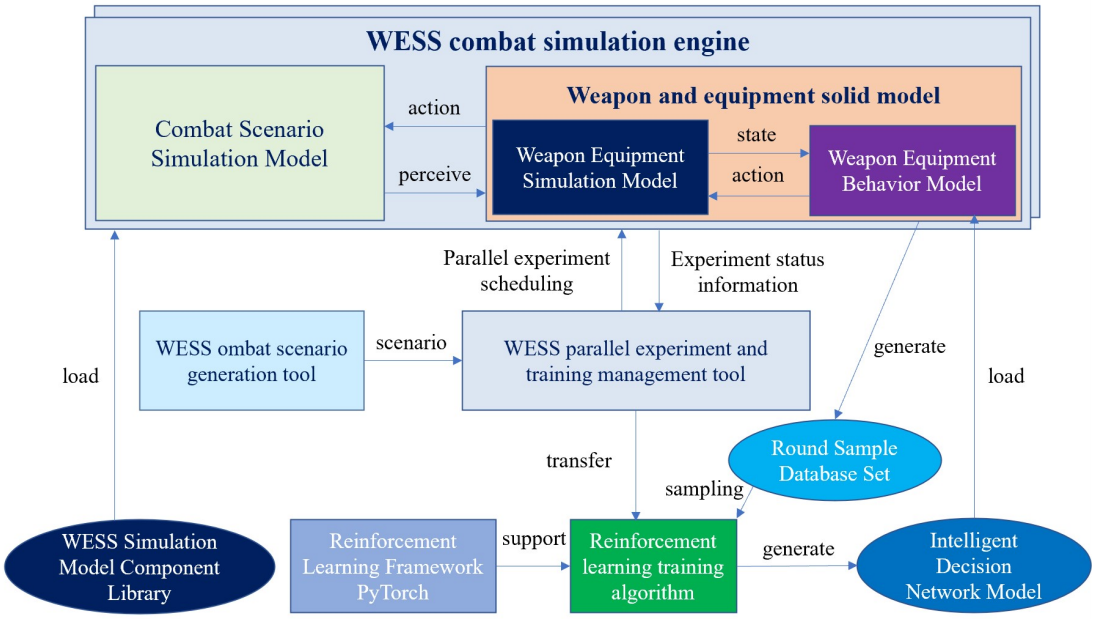

Figure 2: WESS-Based intelligent decision-making reinforcement learning training environment.

in which weapons and equipment may participate in combat, forming a series of scenario scenario sets. The WESS combat simulation engine is responsible for loading the relevant combat simulation model system according to the scenario scenario, and provides operational scheduling services for the weapon equipment entity model and the combat scenario simulation model in which it is located. The physical model of weapon equipment includes two parts: a simulation model reflecting the physical principles of weapon equipment and a behavior model reflecting weapon equipment combat plans and intelligent decision-making behavior. On the one hand, the WESS parallel experiment and training management tool is responsible for automatically scheduling the WESS combat simulation engine for parallel experiments according to the generated scenario scenario and controlling the experiment process; on the other hand, it is responsible for starting an independent reinforcement learning thread for reinforcement learning iterative training and performing Reinforce the end condition control of learning.

The reinforcement learning training algorithm module is responsible for selecting and implementing different reinforcement learning algorithms according to different selections of weaponry and equipment intelligent decision-making problems, and under the call of training management tools, based on the episode sample database generated by the combat simulation engine, it generates an intelligent decision-making network model. It can be based on different machine learning frameworks such as PyTorch and TensorFlow to support the implementation of relevant reinforcement learning algorithms.

3 FRAMEWORK DESIGN OF INTELLIGENT DECISION MODEL FOR FIGHTER AIR-TO-GROUND COMBAT

3.1 Problem description

It is of great significance for the fighter to make auxiliary control decisions according to the situation to reduce the pilot's pressure. Autonomous maneuver decision-making in short-range air combat is a very challenging application direction, because the maneuver performed by both sides in short-range air combat is the most violent, which makes the situation change very rapidly [8]. In addition to passive procedural decision-making methods such as conditional rules, intelligent decision-making is more oriented to active and intelligent intelligent decision-making at this stage, which is mainly divided into two kinds: game based method and artificial intelligence based method. The former is more representative of the matrix game based method [12], the influence graph based method [13], and the differential game method [14]. However, these methods are difficult to solve and are difficult to meet the real-time performance. Therefore, mobile decision-making based on artificial intelligence has become the focus of intelligent decision-making research. Representative methods include expert system method [15], supervised learning method, genetic algorithm [16], reinforcement learning method [17], etc. This paper uses DQN algorithm in reinforcement learning to solve this problem.

Considering a simplified one-dimensional combat scenario, the red side is a SAM with ground-based radar and surface-to-air missiles, and the blue side is a jammer with air-to-surface missiles and electronic countermeasure equipment. The purpose of both parties is to destroy the other to save themselves. In the one-dimensional problem, the fighter plane and the jammer fly towards the SAM in the same straight line. The fighter plane is a certain distance

Table 1: Design of Input State Space of Fighter Air-to-ground Combat Decision Network

State variables	Symbol	Type of data	Parameter range
Relative distance between SAM and fighter	SAM_RelDistance	Double	0-100 km
Relative distance between jammer and fighter	Lead_Distance	Double	0-100 km

Table 2: Design of Input State Space of Jammer Air-to-ground Combat Decision Network

State variables	Symbol	Type of data	Parameter range
Relative distance between SAM and jammer	SAM_RelDistance	Double	0-100 km

Table 3: Design of Output Actions Space of Fighter Air-to-ground Combat Decision Network

Decision output action	Symbol	Value	Type of data
Fighter flying forward	A	1	int
Fighter return	A	0	int

Table 4: Design of Output Actions Space of Jammer Air-to-ground Combat Decision Network

Decision output action	Symbol	Value	Type of data
Jammer flying forward	A	1	int
Jammer flying around	A	0	int

behind the jammer and both fly forward at the same time. The initial position and fire range of the SAM are changed to make this case more versatile.

The agent that solves this scenario needs to learn the following information:

1. The jammer will electronically jam the SAM before the fighter arrives.

2. Keep a certain distance between the jammer and the SAM, and the distance should be as small as possible while ensuring that the jammer is not shot down.

3. The fighter jet attacks the jammed SAM, but cannot fly too close.

3.2 Network model design

(1) Input state space design

The input state of the fighter air-to-ground combat decision model is a characteristic expression of the state information that the fighter can perceive. In the one-dimensional problem, the input state variables of the fighter are the SAM, the jammer and the fighter's own position. Here, the relative distances between the SAM and the fighter and the jammer and the fighter can be used to express. The input state of the jammer has only one relative distance from the SAM.

(2) Output action space design

Fighter air-to-ground combat decision-making output actions mainly include two actions: forward and return (orbiting) of the fighter or the jammer. In the initial stage, to facilitate the collection and sorting of sample data, 0 and 1 are used to refer to the two output actions respectively.

(3) Reward function design

The principle of the design of the reward function for the fighter's air-to-ground combat decision-making is to reflect the influence of the fighter's behavioral decision-making actions on the final combat effect. The purpose is to hope that the fighter can ensure that itself and the jammer will not be destroyed while completing the strike mission. Therefore, the reward can be mainly affected by the following states:

4 EXPERIMENTAL PROGRESS

4.1 Combat mission scenario generation

The operational behaviors of the red and blue sides are shown in Table 8. Set the initial position of the fighter by setting custom decision variable 1 in the scenario editor as the lead distance of the jammer, and obtaining the value of custom variable 1 in the behavior script. By setting the custom decision variable 2 as the decision distance (relative distance from SAM) in the scenario editor, the value of the custom decision variable 2 is obtained in the behavior script, and each step is judged whether the decision distance is reached, and after the decision distance is reached If it is, launch the missile and return home, if not, it will return home directly. In the behavior script, the target is found by default and the missile is launched, so no additional judgment is required. By setting the custom decision variable 1 in the scenario editor as the decision distance of the jammer (relative distance from the SAM), the value of the custom variable 1 is obtained in the behavior script, and each step is judged whether the decision distance is reached, and the

Table 5: Fighter Reward Function Design

Impact index	Index amount	Reward
SAM was destroyed	CurrentState_SAM=0	+10
Fighter was destroyed	CurrentState_Fighter=0	-10
Fighter survival	CurrentState_Fighter=1	0

Table 6: Jammer Reward Function Design

Impact index	Index amount	Reward
SAM was disturbed	CurrentState_SAM=0	+5
Jammer was destroyed	CurrentState_Jammer=0	-5
Jammer survival	CurrentState_Jammer=1	0

Table 7: One-dimensional Task Scenario Variable Description

Variable	Variable range
Fighter firing range	0-40 km
Fighter speed	740 km/h
Jammer range	30 km
Jammer speed	740 km/h
SAM firing range	0-40 km
SAM distance from fighter	0-100 km
Fighter starting position	Latitude: 59.04° Longitude: 27.75° Altitude: 6 km
Longitude: 27.75°	Latitude: 59.04° Longitude: 27.75° Altitude: 6 km

Table 8: 1-D Problem Formulation—Scenario Set Up and Learning Variables

Agent	Side	Learning Variables	Script Behavior	Behavior to Learn
Fighter	Blue	1. Ingress distance 2. Lead distance for jammer to fly before fighter starts ingress	1. Fighter waits until jammer flies lead distance before it begins its ingress 2. If the SAM is within the fighter's firing range when it reached its ingress distance, fighter fires at SAM, then turns around 3. If the SAM is not in fighter's firing range when it hits its ingress distance, fighter turns around and heads to starting position (if not already shot at)	1. Learn to fly no further than its firing range in any scenario 2. Learn when jammer will be effective in ensuring safe passage toward the SAM, even when SAM initial firing range is greater than the fighter's
Jammer	Blue	Ingress distance	1. Jammer starts ingress toward SAM at scenario start 2. If jammer gets within 30 km of SAM, jammer deploys s-band Jamming 3. Jammer stops and hovers after reaching ingress distance	Learn to fly up until its effective range and no further
SAM	Red	None	If Blue platform gets within SAM's firing range, SAM shoots at platform	None

decision is reached. After the distance, turn on the jammer and start to hover, continuously implementing electronic jamming.

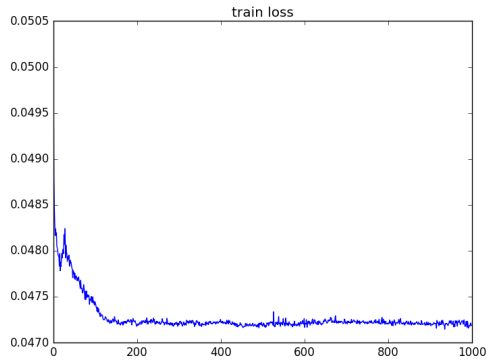

Figure 3: Loss function diagram.

4.2 Task planning modeling based on supervised learning

(1) Experimental design

The task planning model is a rule-based decision-making network, and the decision rules of the task can be obtained by inputting environmental information before the task. A large amount of sample data is obtained through simulation experiments, and the successful samples are extracted to train the neural network. The trained network can be used to predict decision-making actions under certain environmental conditions. However, the generalization of the task planning model is not very strong. You can increase its generalization by increasing the space of experimental assumptions and changing environmental variables. The advantage of the mission planning model is that the sample data can be trained offline and the decision rules can be obtained before the war.

In the original scenario, the input of the mission planning model is the initial position of the SAM, the fire range of the SAM, and the fire range of the fighter; the output is the lead distance of the jammer, the return distance of the fighter, and the start distance of the jammer. In order to obtain the task planning model, the input and output of the model are used as experimental factors to set the level values respectively, and the experimental design scheme is obtained as shown in the figure below. Among them, the 6 experimental factors each take 6 levels, and the comprehensive combination forms more than 460,000 scenario scenarios.

(2) Training network

During training, the SAM initial position, SAM fire range, and fighter fire range are used as inputs, and the lead distance of the jammer, the launch distance of the fighter, and the start distance of the jammer are used as tags to train the network. During the test, input the SAM initial position, SAM firepower range, fighter firepower range, output the lead distance of the jammer, the launch distance of the fighter, and the start distance of the jammer, and calculate the error. Three quarters of the successful samples, that is, the data samples in which SAM was destroyed and fighter and jammer tasks were successful, are used for pre-training network and one quarter for test network. The error curve of training 1000 times is as follows:

As shown in the figure above, about the 200th training time, the loss function has reached convergence. Input the sample used for testing into the trained neural network, and then make the difference between the three output values of the network and the label value and average it to get the error [0.2577, 0.1738, 0.6871], which is within a reasonable range.

4.3 Online decision modeling based on reinforcement learning

(1) Script integration

In the simulation process, the supervised training script, iterative training script and the decision network script will all call the intelligent algorithm script. The decision network model is generated by the supervised training script and updated by the iterative training script. The logical relationship between the call, generation and update of the remaining components is shown in the figure below.

(2) Imitation learning

Generally speaking, due to the sparsity of the agent's behavior and the lag of rewards, the samples generated by the optimal rules cannot be used directly for imitation learning, and preprocessing is required. Only the samples with the action output (state and action) are retained, and the Then the reward before the next action is linked to the sample with the previous action. However, in this case, since the scene is a one-dimensional scene, the behavior is relatively simple. In each experiment, the agent has only one action, and its next state is the relative distance, which has no effect on the reward update, so the current action value is directly used as Reward value.

Imitation learning needs to be repeated iteratively based on the optimal rule sample data until the loss function converges. The following figure shows the situation where the imitation learning loss function converges with the increase of the number of iterations, reaching convergence at about 60 times. The output of imitation learning is the initial version of the strategy network for reinforcement learning training.

(3) Intelligent test

The reinforcement learning iterative training tool is used to iteratively train the decision network while running the simulation experiment until the convergence condition is reached. After the iterative training is saved, the air-to-ground combat intelligent decision-making network model file is generated locally, and the model parameters are loaded to complete the initialization operation. Collect fighter and jammer combat data during intelligent testing, and perform statistical analysis after completing the experiment. Set the number of repetitions to 10, and the total number of simulation samples is 1000 times.

The final combat effect of statistical comparison is shown in the following table:

5 CONCLUSION

Based on the reality of aircraft combat mission, this paper systematically analyzes the aircraft autonomous decision-making in one-dimensional scene. The aircraft intelligent decision-making method based on combat simulation and deep reinforcement learning is

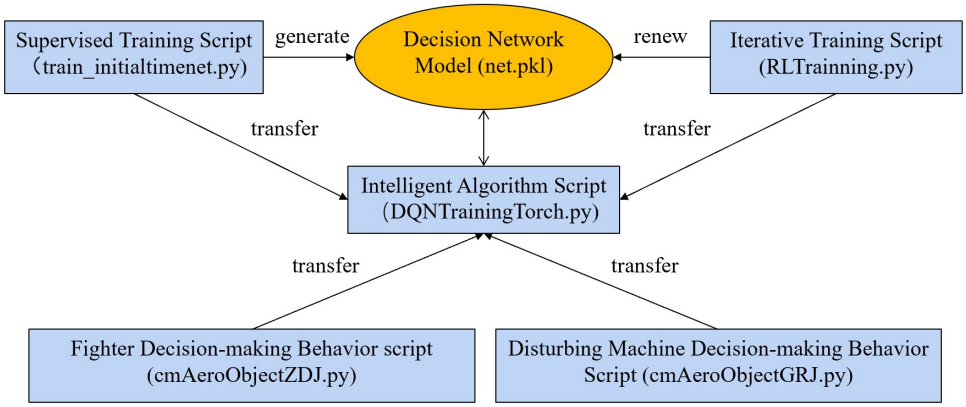

Figure 4: Script relationship.

Table 9: Intelligent test results

Action strategy	Total number of samples	Successfully destroyed SAM, fighter/jammer survival times	Average probability of destroying SAM
Intelligent decision-making strategy	1000	926	0.93

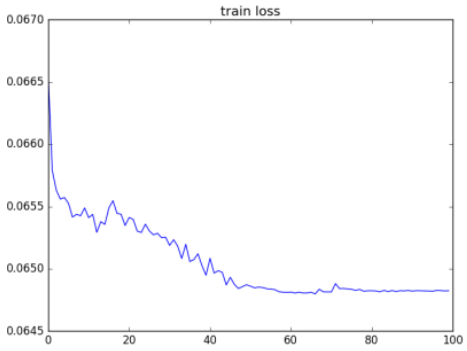

Figure 5: Imitate the case where the learning loss function converges as the number of iterations increases

proposed, and the aircraft autonomous decision-making modeling framework based on combat simulation and deep reinforcement learning is launched. The aircraft intelligent decision-making learning training environment is constructed based on WESS platform. The application research of one-dimensional scene aircraft air ground penetration case is carried out in WESS aircraft intelligent decision-making training environment. The results show that the intelligent decision algorithm is feasible and effective.

Outlook for follow-up work: 1) Continue to carry out the two-dimensional problem of aircraft and use deep reinforcement learning to solve the problem of aircraft autonomous route planning. 2) Expand the applicability of one-dimensional and two-dimensional problems, apply the aircraft's independent decision-making and route planning to more complex combat scenarios, and test its intelligence.

ACKNOWLEDGMENTS

We are grateful to the anonymous referees for their helpful reviews, and all the volunteers who wrote and provided helpful comments on previous versions of this document. We gratefully acknowledge the Model & Data Hybrid Driven Intelligent Modeling for Combat Systems (NO. 62003359) for supporting this research.

REFERENCES

[1] GAO Zhen-yang, QIN Bin. Progress in Deep Reinforcement Learning research[J]. Computer knowledge and technology, 2019,15(04): 163-165+179.
[2] B Xue, Y He, F Jing, *et al.* Robot Target Recognition using Deep Federated Learning [J]. International Journal of Intelligent Systems, 2021, 36(12): 7754-7769.
[3] Lu Y, Xu X, Zhang X, *et al.* Hierarchical Reinforcement Learning for Autonomous Decision Making and Motion Planning of Intelligent Vehicles[J]. IEEE Access, 2020, PP(99):1-1.
[4] Peng X, Chen R, Zhang J, *et al.* Enhanced Autonomous Navigation of Robots by Deep Reinforcement Learning Algorithm with Multistep Method[J]. Sensors and Materials, 2021, 33(2):825.
[5] ZUO Jialiang, YANG Rennong, ZHANG Ying, LI Zhonglin, WU Meng. Intelligent decision-making in air combat maneuvering based on heuristic reinforcement learning[J]. Acta Aeronautica et Astronautica Sinica, 2017(10):217-230.
[6] LI ANG ZHANG, JIA XU, DARA GOLD, JEFF HAGEN, AJAY K. KOCHHAR, ANDREW J. LOHN, OSONDE A. OSOBA. Air Dominance Through Machine Learning.
[7] Li K, Zhang K, Zhang Z, *et al.* A UAV Maneuver Decision-Making Algorithm for Autonomous Airdrop Based on Deep Reinforcement Learning[J]. Sensors, 2021, 21(6): 2233.
[8] Yang Qiming, Zhang Jiandong, Shi Guoqing, Hu Jinwen, Wu Yong. Maneuver Decision of UAV in Short-Range Air Combat Based on Deep Reinforcement Learning[J].IEEE ACCESS, 2020: 363-378.
[9] Song Xiagan, Jiang Ju. Research on intelligent air combat decision under uncertain environment[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2017.03.
[10] LIU Chao, LI Qingwei. Design of Simulation Software System for Intelligent Decision – Making in Tactical – Level Air Combat[A]. 2019 China System Simulation

and Virtual Reality Technology High-level Forum[C],2019.
[11] Lei Yonglin, Yao Jian, Zhu Ning, Zhu Yifan, Wang Weiping. Weapon Effectiveness Simulation System (WESS)[J]. Journal of System Simulation. 2017,29(6): 1244-1252.
[12] AUSTIN F, CARBONE G, MICHAEL F, el a1. Game Theory for Automated Maneuvering during Air-to—Air Combat[J].Journal of Guidance 1990,13(6):1143—1147.
[13] VIRLANEN K, RAIVIO T, HAMALAINEN R P. Decision Theoretical Approach to Pilot Simulation[J].Journal of Aircraft, 1999,27(4):632—641.
[14] HORIE K, CONWAY B. Optimal Fighter Pursuit-Evasion Maneuvers Found via Two-Sided Optimization[J]. Journal of Guidance, Control and Dynamics, 2006, 29(1): 105—112.
[15] ZHU Shihu, DONG Chaoyang, ZHANG Jinpeng, CHEN Zongji. An intelligent decision-making system based on neural networks and expert system [J].Electro-optical and control, 2006,13(1): 8-11.
[16] SMITH R E, DIKE B A, MEHRA R K, *et al.* Classifier systems in combat: two-sided learning of maneuvers for advanced fighter aircraft[J].Computer Methods in Applied Mechanics and Engineering, 2000,186(2): 421.DOI: 10.1016/S0045—7825(99)00395—3.
[17] NICHOLAS E, COHEN K, SCHUMACHER D. Collaborative tasking of UAVs using a genetic fuzzy approach[C]. /51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition. Grapevine: AIAA, 2013: 1032. DOI: 10.2514 /6. 2013-1032.