Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

DYNAMIC WEAPON TARGET ASSIGNMENT VIA SIMULATION, REINFORCEMENT LEARNING AND GRAPH NEURAL NETWORK

Seung Heon Oh

Seoul National University
1th Gwanak-ro, Gwanak-gu,
Seoul, 08826, REPUBLIC OF KOREA

ABSTRACT

DWTA (dynamic weapon target assignment problem) is the important resource scheduling problem in battlefield. In this paper, deep reinforcement learning and graph neural network optimize the performance of the decision making of DWTA. The proposed method is evaluated experimentally for some cases and compared with other heuristic methods.

1 INTRODUCTION

Dynamic Weapon Target Assignment (DWTA) is defined as the challenge of allocating interceptor in response to threats while considering the dynamic changes within a system. One of the methods for modeling DWTA is the mathematical modeling. One significant advantage of mathematical modeling is the potential to derive exact solutions using methods like dynamic programming. Nevertheless, this approach is constrained by the low level of model accuracy. In contrast to this mathematical modeling, simulation modeling offers an effective way to intricately represent the battlefield environment. However, simulation methods have thus far been more commonly utilized as 'what-if' evaluation tools rather than as optimization tools. As a result, the details of real-world battlefield captured by simulation modeling has not been included in the optimization process. This study adopts an approach to overcome these limitations by combining simulation, reinforcement learning and graph neural networks (GNN). Reinforcement learning agents accumulate experience from a realistic simulation. GNN captures the complex relation of multiple objects in the battlefield. Lastly, the proposed approach is validated through a comparative analysis with alternative real-time control methods, such as Model Predictive Control and heuristics.

2 PROPOSED METHOD

In this study, the battlefield simulation comprises launcher and missile classes within simulation components. Users can employ these classes to model battlefield in an object-oriented manner. The adapter component takes various design parameters for configuring the simulation model, such as specifications of launcher or guided missiles, and preprocesses the data. The modeler component constructs the model by utilizing classes from the simulation component. Also, the modeler component includes the transition, observation, and reward functions. They are essential for integrating the simulation with reinforcement learning. This simulation framework offers a realistic representation of the battlefield environment and weapon systems. The agent which is launcher (naval ship etc.) performs decision-making to assign interception to targets for defensive purposes. To address the optimization problem, DWTA is constructed into partially observable Markov Decision Process (POMDP). POMDP serves as a model for sequential decision-making under partial observation conditions, consisting of states, observations, actions, rewards, transition functions, observation functions, reward functions, and a discount factor (a real number between 0 and 1). States represent variables that fully describe the current state of the dynamic system. However, due to the agent's limited detection capabilities, access to the states is inaccessible. Observations provide partial information about the system to the agent. In this study, the observation features include 1) the inventory of interceptions, 2) the flight status of interception guided missiles, 3) the distribution of threats by distance intervals, 4) the cumulative interception count, and 5) assignment records. Action is selecting one of the interceptable threats to allocate and launch defensive armaments. Rewards are received based on whether a threat is successfully intercepted (number of successfully intercepted threats) and whether a non-lose situation occurs. The relations among objects in the battlefield are expressed in a graph structure. Agents, threats, and interceptors are represented as individual nodes, while targeting relations, engagement availability relations, launching relations, and proximity relations are represented as edges. Graph Neural Networks (GNNs) capture the semantic relationships among these elements, enhancing their representational capabilities. Agent uses observations and GNN-embedded information to make a decision. In this study, reinforcement learning is applied to learn agent policies for minimizing the probability of being hit. The agent constructs a policy to select actions that maximize the state-action value function.

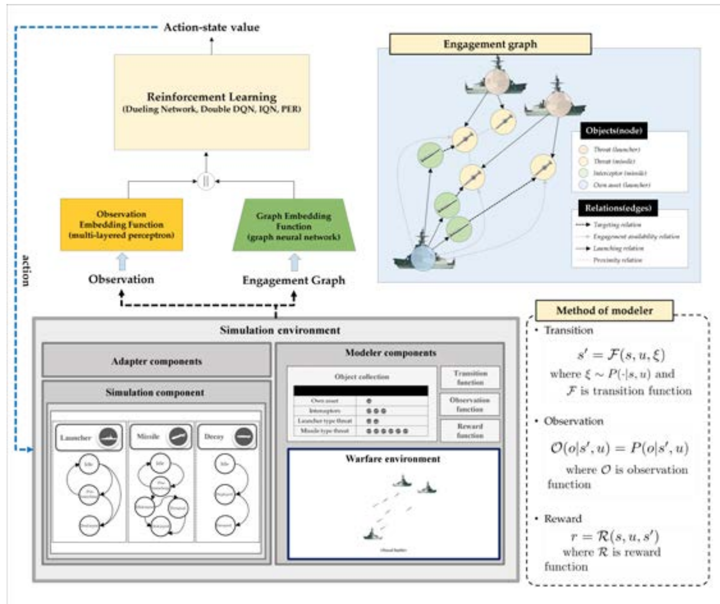
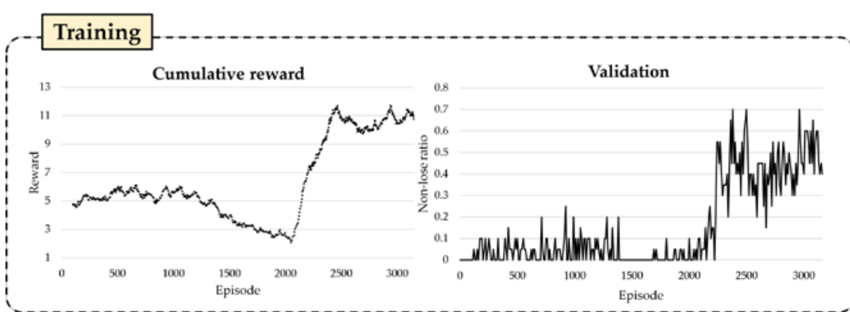


Figure 1: Simulation and reinforcement learning framework.

3 EXPERIMENT

For the experiments in this study, the dataset was structured with real world naval weapons. The proposed method was trained in a 1:3 (one's own asset and three enemy ships) environment, and the training and testing outcomes are presented in Figure 2. The training results indicate a improvement in the policy as episodes. Furthermore, Testing the model in different environments from the training setting demonstrates better performance compared to heuristic or model predictive control (MPC) methods.



Test

Dataset (own ships: enemy ships)	Proposed method	Heuristic	Model-predictive control(MPC)	
			GA-based	SLSQP*
2:6	60%	39%	-	-
1:3	40%	23% (parameter optimized by GA)	4%	9%

* *GA : genetic algorithm*
* *SLSQP : sequential least squares programming*

Figure 2: Training and test results.