



Received: 2026/08/31
Revised: 2026/09/10
Accepted: 2026/09/21
Published: 2026/09/30

***Corresponding Author:**

Dooyoung Kim
Tel: +82-55-549-5624
E-mail: dykim07@navy.ac.kr

Abstract

본 연구는 수동 소나의 방위 정보만으로 잠수함을 추적하는 두 무인 수상정의 협력 기동 명령 생성을 위한 다중 에이전트 강화학습 기법을 제안한다. 입자 필터로 표적 상태를 추정하고, 강화학습으로 목표 거리와 유리한 관측 기하를 유지하는 제어 명령을 생성하였다. 시뮬레이션을 통해 제안한 기법의 성능을 검증하였다.

This study proposes a multi-agent reinforcement learning method for generating cooperative maneuver commands for two unmanned surface vehicles tracking a submarine using bearing-only passive sonar measurements. A particle filter is used to estimate the target state, while reinforcement learning generates control commands to maintain the desired standoff distance and favorable observation geometry. The performance of the proposed method is validated using simulations.

Keywords

무인 수상정(Unmanned Surface Vehicle),
표적 추적(Target Tracking),
수동 소나(Passive Sonar),
대잠전(Anti-Submarine Warfare),
다중 에이전트 강화학습(Multi-Agent Reinforcement Learning)

Acknowledgement

이 논문은 2026년 해군사관학교 해양연구소 학술연구과제(강화학습을 이용한 무인수상정의 대잠 추적경로 생성) 연구비 지원으로 수행된 연구임.

수동 소나를 탑재한 두 무인 수상정의 잠수함 추적을 위한 다중 에이전트 강화학습 기법

Multi-Agent Reinforcement Learning for Submarine Tracking Using Two USVs with Bearing-Only Passive Sonar Sensors

김두영*

해군 중령/해군사관학교 인공지능학과 부교수

Dooyoung Kim*

CDR, ROK Navy/Associate professor, Dep. of Artificial Intelligence, Republic of Korea Naval Academy

1. 서론

대잠전(anti-submarine warfare, ASW)은 잠수함의 탐지, 추적, 식별 및 공격을 통해 아군 전력과 해양을 보호하는 해군의 핵심 작전 중 하나이다[1,2]. 그중 대잠 추적은 탐지된 미식별 표적 잠수함과 접촉을 유지하면서 이후 식별 및 공격을 하기 위한 사전단계로, 표적 잠수함의 공격으로부터 아군 플랫폼의 생존성을 보장함과 동시에 작전 목표 달성을 위해 필요한 시간 동안 지속적으로 접촉을 유지하는 것이 핵심이다[1].

무인 수상정(unmanned surface vehicle, USV)은 유인 함정에 비해 인명 손실 위험과 장시간 운용의 부담이 적은 장점으로 해군의 다양한 임무에 활용하는 방안이 연구되고 있다[3,4]. 특히 대잠전의 경우 관심 해역에 대한 지속적인 감시능력을 바탕으로 수중 표적의 탐지와 추적에 효과적으로 운용될 수 있다.

대표적인 대잠 센서인 수동 소나[4]는 표적에서 발생하는 소음을 탐지하므로 일반적으로 표적의 방위 정보만을 얻을 수 있으며 위치나 침로 등의 운동 정보를 직접적으로 획득하기 어렵다. 따라서 관측 플랫폼이 시간에 따라 위치를 변경하여 획득한 다수의 방위선을 이용하여 표적의 운동 정보를 산출하여야 한다[5].

복수의 USV를 이용할 경우 서로 다른 위치에서 동일 표적에 대

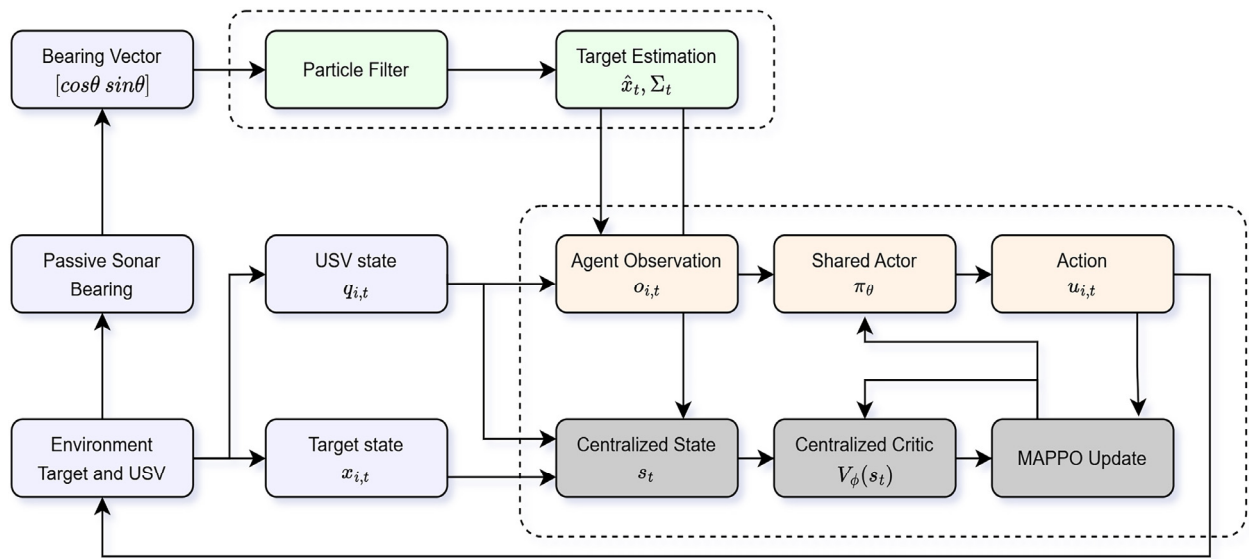


Fig. 1. Overall architecture of the proposed particle filter-MAPPO framework

해 두 개 이상의 방위 정보를 획득할 수 있으므로 단일 플랫폼 대비 관측 성능을 높일 수 있다. 다만, 관측의 정확도는 표적과 각 USV가 형성하는 상대적인 위치에 영향을 받는다. 예를 들어 USV와 표적이 한 직선에 위치하게 되거나, 방위선의 교차 각이 작으면 수동 소나의 측정 오차로 인해 표적의 추정오차가 크게 증폭될 수 있다. 반면에 관측 방위선이 직교할수록 표적의 위치를 추정하기에 유리한 관측 기하가 형성된다. 또한, 표적 잠수함과 USV 사이의 거리도 중요한 요소이다. 거리가 가까울수록 추적의 정확도가 높아지나, 표적 잠수함에게 공격받을 확률도 증가한다. 따라서 이동하는 표적 잠수함을 안정적으로 추적하기 위해서는 적절한 안전거리를 유지하며 상태 추정에 유리한 관측 기하를 지속적으로 유지하는 제어 명령을 생성해야 한다.

본 연구에서는 이러한 문제를 해결하기 위해 입자 필터(particle filter, PF)[6]와 다중 에이전트 강화 학습(multi-agent reinforcement learning, MARL) 기법의 하나인 multi-agent proximal policy optimization(MAPPO)[7]를 결합하여 두 대의 USV가 수중 표적을 추적하기 위한 협력 기동 명령 생성 기법을 제안한다. 먼저 두 USV가 수동소나로 측정한 오차가 포함된 방위각 정보를 바탕으로 입자 필터를 사용하여 수중 표적의 위치와 속도를 추정하였다. 추정 상태와 각 USV의 상태는 MAPPO의 입력 벡터로 활용되어 각 USV의 기동 경로를 생성하였다. 이때 각 USV가 수중 표적과의 안전거리를 유지하고, 관측 오

차를 최소화하기 위해 관측 방위가 직교를 유지하도록 보상함수를 설계하였다. 마지막으로 제안한 기법의 성능검증을 위해 시뮬레이션을 수행하였다.

논문의 구성은 다음과 같다. 먼저 2장에서 해결하고자 하는 문제를 정의하고 이를 해결하기 위한 입자 필터와 MAPPO를 결합한 협력 기동 제어 기법 제안한다. 3장에서는 시뮬레이션을 통해 제안하는 기법의 성능을 확인한다. 마지막으로 4장 결론에서는 연구 결과의 요약과 향후 연구 방향을 제시한다.

2. 입자 필터-MAPPO 협력 기동 제어

2.1 구조

Fig. 1은 제안하는 기법의 전체 구조를 도식화한 그림이다. 제안하는 기법은 크게 1) 수동소나의 관측 및 표적 상태 추정 단계와 2) MAPPO 기반 제어 명령 생성 단계로 구성된다. 각 단계는 USV의 기동에 따라 새로운 방위 벡터가 관측되고, 관측을 통해 갱신된 표적의 추정 상태가 다시 모델의 입력으로 사용되는 구조를 가진다.

먼저 각 USV는 수동 소나를 통해 표적에 대한 방위 정보를 획득한다. 획득된 방위 정보는 각도 경계점의 불연속 문제를 완화하기 위하여 삼각함수를 이용한 방위 벡터 변환하여 입자 필터의 관측값으로 사용된다. 입자 필터는 이를 통해 표적 상태의 추정값을 산출한다. 추정된 표적 상태와, USV의 위치 벡터로 각 USV

에이전트(agent)를 위한 관측 벡터를 생성한다. 각 에이전트의 정책은 해당하는 관측 벡터를 입력받아 제어 명령을 생성하고 이 기동명령은 USV로 전달된다.

학습 과정에서는 centralized training and decentralized execution(CTDE)[8]을 사용하였다. 전역 상태(centralized state)는 표적과 USV의 실제 상태 정보 등 학습에 필요한 모든 정보를 전역적으로 전달할 수 있도록 구성하고 이를 critic에 전달하여 각 USV의 정책의 가치를 추정한다. 이후 MAPPO update에서 각 값을 수집하고 이를 통해 정책망(actor)와 가치망(critic)의 매개변수를 갱신한다. 이때 실제 표적 상태는 학습 시 critic에서만 사용되고, actor에게는 입자 필터를 통해 추정된 상태를 전달한다.

2.2 상태 및 동역학 모델

표적의 상태 벡터는 다음 수식 (1)과 같이 구성한다.

$$\mathbf{x}_t = [p_{x,t} \ p_{y,t} \ v_{x,t} \ v_{y,t}]^T \in R^4 \quad (1)$$

여기서 t 는 시간을 의미하며, $(p_{x,t}, p_{y,t})$ 는 표적의 위치를, $(v_{x,t}, v_{y,t})$ 는 속도를 의미한다.

표적은 등속도 운동을 기본으로, 시간에 따라 변화하는 가속도 성분을 포함하는 이동체로 가정한다. 이때 표적의 위치와 속도는 각각 수식 (2)와 수식 (3)을 통해 갱신한다.

$$\mathbf{p}_{t+1} = \mathbf{p}_t + \Delta t \mathbf{v}_t + \frac{1}{2} \Delta t^2 \mathbf{a}_t \quad (2)$$

$$\mathbf{v}_{t+1} = \mathbf{v}_t + \Delta t \mathbf{a}_t \quad (3)$$

위 식에서, Δt 는 시간 간격, $\mathbf{a}_t$ 는 표적의 가속도를 의미한다.

USV i 의 상태는 수식 (4)와 같이 2차원 평면상의 위치로 정의한다.

$$\mathbf{q}_t = [q_{x,i,t} \ q_{y,i,t}]^T \in R^2 \quad (4)$$

정책 π_θ 로부터 산출된 각 USV i 의 제어 입력 $u_{i,t} \sim \pi_\theta(\cdot | o_{i,t})$ 는 2차원 속도 벡터로 다음 수식 (5)와 같다.

$$\mathbf{u}_{i,t} = [u_{x,i,t} \ u_{y,i,t}]^T \in R^2 \quad (5)$$

이때, USV의 최대 속력을 고려하여 제어 입력 벡터의 최대 크기는 $\|\mathbf{u}_{i,t}\| \leq v_{\max}$ 로 제한한다. 이를 통해 산출된 각 USV의 위치는 다음 수식 (6)과 같다.

$$\mathbf{q}_{i,t+1} = \mathbf{q}_{i,t} + \Delta t \mathbf{u}_{i,t} \quad (6)$$

2.3 관측 모델

본 연구에서는 방위각 표현 시 방위각의 경계에서 발생하는 불연속 문제를 완화하기 위하여 수식 (7)과 같이 삼각함수로 구성된 2차원 방위 벡터를 사용하였다. 임의의 방위각 θ 에 대한 방위 벡터 $\mathbf{b}$ 는 다음과 같다.

$$\mathbf{b}_{i,t} = [\cos\theta_{i,t} \ \sin\theta_{i,t}]^T \in R^2 \quad (7)$$

방위 벡터의 전역 좌표계(global frame)에서의 표현은 수식 (8)과 같이 표적과 USV 간의 상대 위치를 활용하여 표현할 수 있다.

$$\mathbf{b}_{i,t} = \frac{\mathbf{p}_t - \mathbf{q}_{i,t}}{\|\mathbf{p}_t - \mathbf{q}_{i,t}\|} \quad (8)$$

USV i 의 침로(heading)을 ψ_i 라 할때 USV i 의 동체 좌표계(body-frame) 상 관측 방위 벡터 $\mathbf{b}_{i,t}^{body}$ 는 다음 수식 (9)와 같으며, $R(\alpha)$ 는 회전 행렬로 그 표현은 수식 (10)과 같다.

$$\mathbf{b}_{i,t}^{body} = R(-\psi_i) \mathbf{b}_{i,t} \quad (9)$$

$$R(\alpha) = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \quad (10)$$

실제 수동소나를 모사하기 위해 가우시안 기반의 관측 잡음을 반영하였다. 수동 소나의 관측 잡음을 $\epsilon_{i,t} \sim N(0, \sigma_b^2)$ 라 할 때 잡음이 포함된 표적 방위 관측 값 $\mathbf{y}_{i,t}$ 는 다음과 같다.

$$\mathbf{y}_{i,t} = R(\varepsilon_{i,t}) \mathbf{b}_{i,t}^{body} \quad (11)$$

최종적으로 두 USV의 관측 벡터는 다음과 같다.

$$\mathbf{y}_t = [\mathbf{y}_{1,t}^T \mathbf{y}_{2,t}^T]^T \in R^4 \quad (12)$$

2.4 입자 필터 기반 위치 추정

표적의 상태 추정을 위해 입자 필터를 적용하였다. 각 입자는 수식 (13)과 같이 표적의 위치와 속도를 포함하는 벡터로 구성한다.

$$\mathbf{x}_t^{(n)} = [p_{x,t}^{(n)} p_{y,t}^{(n)} v_{x,t}^{(n)} v_{y,t}^{(n)}]^T \in R^4 \quad (13)$$

초기(initial) 입자는 수식 (14)와 같이 평균 μ_0 와 표준편차 σ_0 를 가지는 가우시안 분포로 무작위 추출한다.

$$\mathbf{x}_0^{(n)} = N(\mu_0, \text{diag}(\sigma_0^2)) \quad (14)$$

예측(prediction) 단계에서는 수식 (15)의 상태 전이 모델을 이용하여 다음 상태를 계산한다.

$$\mathbf{x}_{t+1}^{(n)} = F\mathbf{x}_t^{(n)} + G\mathbf{w}_t^{(n)} \quad (15)$$

이때 F 는 상태 전이 행렬, G 는 가속도 입력 행렬, $\mathbf{w}$ 는 표적의 알려지지 않은 가속도 및 모델의 불일치를 반영하는 잡음이다.

$$F = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, G = \begin{bmatrix} \frac{1}{2} \Delta t^2 & 0 \\ 0 & \frac{1}{2} \Delta t^2 \\ \Delta t & 0 \\ 0 & \Delta t \end{bmatrix} \quad (16)$$

갱신(update) 단계에서는 각 입자의 예측으로부터 계산한 방위 벡터와 실제 관측 방위 벡터 간의 내적을 이용하여 우도(likelihood)를 계산한다. 측정 벡터와 예측 벡터가 일치할수록 두 단위벡터의 내적이 1에 가까워지는 특성을 고려하여 로그 우도는 다음 수식 (17)과 같이 정의한다.

$$\log p(\mathbf{y}_t | \mathbf{x}_t^{(n)}) \propto \sum_{i=1}^2 \kappa_i \mathbf{y}_{i,t}^T \hat{\mathbf{b}}_{i,t}^{(n)} \quad (17)$$

이때, $\hat{\mathbf{b}}_{i,t}^{(n)}$ 는 입자 n 이 예측한 USV i 와 표적 간의 방위 벡터이며, $\kappa_i = 1/\sigma_{b,i}^2$ 는 집중도 모수(concentration parameter)이다.

각 입자의 가중치를 다음 수식 (18)과 같이 갱신하고 수식 (19)과 같이 정규화한다.

$$\log w_t^{(n)} \leftarrow \log w_{t-1}^{(n)} + \log p(\mathbf{y}_t | \mathbf{x}_t^{(n)}) \quad (18)$$

$$w_t^{(n)} = \exp(\log w_t^{(n)}) / \sum_{m=1}^N \exp(\log w_t^{(m)}) \quad (19)$$

입자 필터를 통한 표적의 상태 추정값은 각 입자의 가중평균을 이용하여 계산하며 그 과정은 수식 (20)과 같다.

$$\hat{\mathbf{x}}_t = \sum_{n=1}^N w_t^{(n)} \mathbf{x}_t^{(n)} \quad (20)$$

또한, 상태 추정의 불확실성을 표현하는 사후 공분산(posterior covariance)은 다음 수식 (21)과 같다.

$$\Sigma_t = \sum_{n=1}^N w_t^{(n)} (\mathbf{x}_t^{(n)} - \hat{\mathbf{x}}_t)(\mathbf{x}_t^{(n)} - \hat{\mathbf{x}}_t)^T \quad (21)$$

반복적인 갱신에서 소수의 입자에게 가중치가 집중되는 입자 축퇴(particle degeneracy) 완화를 위해 유효 표본 크기(effective sample size) 수식 (22)와 같이 계산한 후, N_{eff} 가 사전 정의된 임계값(threshold)보다 작아지면 재 표본 추출(systematic re-sampling)을 수행한다.

$$N_{eff} = 1 / \sum_{n=1}^N (w_t^{(n)})^2 \quad (22)$$

2.5 MAPPO 기반 제어 명령 생성

기동 경로 계획을 위한 MAPPO 기반 협력 기동 제어 명령 생성 모델의 구성은 다음과 같다. 모델은 크게 decentralized 정책망과 centralized 가치망으로 구성되어 있다.

두 USV는 같은 정책망 매개변수(parameter)를 공유하지만, 각각의 관측 벡터(observation vector)를 독립적으로 입력받아 제어 명령을 생성한다. USV i 의 관측 벡터는 다음 수식 (23)과 같다.

$$o_{i,t} = [\hat{x}_t q_{i,t} q_{j,t} \hat{p}_t - q_{i,t} e_i], i \neq j \quad (23)$$

이때, q_i 는 자신의 위치, q_j 는 다른 USV의 위치이며, e_i 는 USV를 식별하기 위한 식별 벡터로, 정책망의 매개변수를 공유하더라도 각 에이전트를 식별할 수 있도록 한다.

정책망은 각 USV의 관측 벡터 $o_{i,t}$ 를 입력받아 Tanh-normal 분포를 따르는 location과 scale 매개변수를 출력한다. 이를 통해 2차원 제어 명령 $u_{i,t}$ 를 출력한다. 이때 출력된 제어 명령의 최댓값은 $v_{\max}$ 이내가 되도록 제한한다.

가치망은 식 (24)와 같은 전역 상태 벡터를 입력으로 사용한다. 이 벡터에는 실제 표적 상태, 입자 필터의 추정 상태, 두 USV의 위치정보를 포함한다.

$$s_t = [x_t, \hat{x}_t, q_{1,t}, q_{2,t}] \quad (24)$$

가치망은 s_t 를 입력받아 각 USV별 상태 가치(value)를 다음과 같이 출력한다.

$$V_\phi(s_t) = [V_{\phi,1}(s_t) \ V_{\phi,2}(s_t)]^T \quad (25)$$

정책망과 가치망의 학습 시 수식 (26)과 같이 clipped surrogate objective를 적용한다.

$$L^{clip(\theta)} = E_t[\min(r_{t(\theta)} \hat{A}_t, clip(r_{t(\theta)}, 1-\epsilon, 1+\epsilon) \hat{A}_t)] \quad (26)$$

이때 $r_t(\theta) = \pi_\theta(u_t|o_t)/\pi_{\theta_{old}}(u_t|o_t)$ 는 기존 정책과 현재 정책 사이의 확률비, π_θ 는 정책망의 정책함수, $\hat{A}$ 는 generalized advantage estimation, ϵ 은 PPO clipping coefficient이다.

2.6 보상함수

보상함수는 두 USV가 표적과 일정 거리를 유지하

면서 상태 추정에 유리한 기하를 유지하도록 거리 보상과 각도 보상으로 구성하였다.

표적에 가까울수록 측정 오차의 영향력이 감소하여 정확도가 향상된다. 그러나 대잠전에서 표적에 지나치게 근접할 경우 생존성이 저하되므로 적정 안전 거리를 유지할 필요가 있다. i 번째 USV와 표적 간 거리를 d_i , 안전거리를 d^* 라 할 때 거리 보상 r_{dist} 는 다음 수식 (27)과 같다.

$$r_{dist} = \exp\left(-\frac{1}{K} \sum_{i=1}^K \left(\frac{d_i - d^*}{\sigma_d}\right)^2\right) \quad (27)$$

방위 기반 위치 추정에서는 USV와 표적이 형성하는 기하학적 위치가 위치 추정의 성능에 큰 영향을 미친다. 특히 두 개 이상의 관측자를 운용할 경우 두 관측 벡터의 교차각이 작아질수록 추정의 오차가 증가한다. 따라서 본 연구에서는 관측 방위 벡터가 직교에 가까워지도록 각도 보상을 식 (28)와 같이 정의하였다. 각 USV가 측정한 방위선 벡터 b_1 과 b_2 에 대한 각도 보상 r_{angle} 은 다음과 같다.

$$r_{angle} = 1 - (b_1^T b_2)^2 \quad (28)$$

전체 보상은 거리 보상과 각도 보상의 가중 합으로 다음 수식 (29)와 같다.

$$r = \lambda_d r_{dist} + \lambda_a r_{angle} \quad (29)$$

3. 시뮬레이션 및 결과

3.1 시뮬레이션 환경 구성

제안한 기법의 성능을 검증하기 위해 2차원 평면 상에 표적 잠수함과 두 대의 USV를 배치한 시뮬레이션을 수행하였다. 시뮬레이션의 주요 매개변수는 Table 1과 같다. 시뮬레이션에서 표적의 회피기동을 모사하기 위해 표적의 가속도를 다음 식과 같이 생성하였다.

$$[0.08 \sin(0.7t) \ 0.05 \cos(0.5t)] \quad (30)$$

Table 1. Simulation parameters

key	value
Δt	0.1
steps per episode	1,000
target initial position	$\mathbf{p}_0 \in [-2, 2] \times [-2, 2]$
USV initial position	$\mathbf{q}_{i,0} \in [-10, 10] \times [-10, 10]$
$v_{\max}$	3
σ_b	5°
d^*	8
σ_d	6
λ_a, λ_d	0.5
N	2,048
μ_0	$[2, 0, 0, 0]^T$
σ_0	$[10, 10, 2, 2]^T$
σ_a	0.6
N_{eff}	$< 0.5N$

표적 및 USV의 초기 위치는 에피소드마다 각각 $[-2, 2]$ 와 $[-10, 10]$ 의 범위 내에서 무작위로 선정하였다. 수동 소나의 방위 측정 잡음은 표준편차 값을 5° 로 하는 가우시안 잡음으로 표현하였다. 입자 필터의 입자는 총 2,048개를 사용하였으며 systematic re-sampling의 기준은 N_{eff} 가 전체 입자 수의 절반 미만일 때로 설정하였다.

MAPPO 정책망과 가치망은 각각 128개의 유닛을 가진 은닉층 두 개로 구성되며, 각 은닉층의 활성화 함수는 Tanh 함수를 사용하였다.

각 iteration에서 1.024 프레임을 수집하고, 256개의 미니배치 단위로 4 epoch를 학습하였다. 기타 매개변수 값은 Table 2와 같으며, 학습 전반에 사용된 매개변수는 실험을 통해 최선의 값을 도출하여 반영하였다.

3.2 실험 결과

Fig. 2는 iteration의 진행에 따른 학습과 평가 보상 값이다. 초기 평가 보상은 0.2 미만으로 낮은 값을 보여 주었으나, 250회 이후부터 증가하여 279회에서 0.9259로 최대 성능을 보여 주어 학습을 조기 종료하였다. 이때 RMSE(root mean square error) 지표로 평가한 표적의 추정 위치 오차와 속도 오차는 각각 0.3604와 0.2948이다.

Table 2. Network parameters.

key	value
actor network	$12 \times 128 \times 128 \times 4$
actor activation	Tanh
policy distribution	Tanh normal
critic network	$12 \times 128 \times 128 \times 2$
critic activation	Tanh
frames per batch	1,024
minibatch size	256
PPO epochs	4
learning rate	3×10^{-4}
discount factor	0.99
GAE	0.95
ϵ	0.2
entropy coefficient	0.002

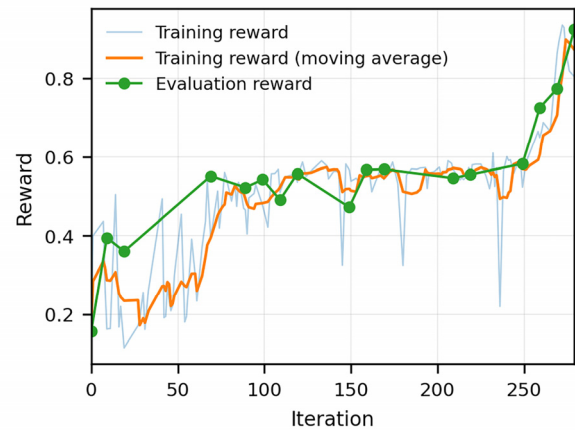
**Fig. 2.** Training and evaluation rewards.

Fig. 3는 학습이 완료된 actor를 이용하여 표적 추적을 시뮬레이션한 결과이다. Fig. 3(a)는 표적과 두 대의 USV의 기동을 2차원 평면에 도식화한 그림이다. 추적 초기 입자 필터의 추정값이 불확실한 상황에서는 표적과 USV 간 거리와 관측 방위가 불안정한 상태를 보이거나 USV의 적극적인 기동으로 이후 안정된 추적을 보임을 확인할 수 있다. Fig. 3(b)는 추적이 진행되는 과정에서의 각도 및 거리 보상과 전체 보상의 변화를 나타낸다. 추적 초기에는 낮은 거리 보상을 개선하기 위해 USV가 급격하게 기동하여 일시적으로 각도 보상이 감소한다. 거리 보상이 안정화된 후에는 각도 보상이 개선되는 기동을 시행하여 전체 보상이 증가하는 것을 확인할 수 있다.

이 결과는 학습된 모델이 단순히 표적에 접근하

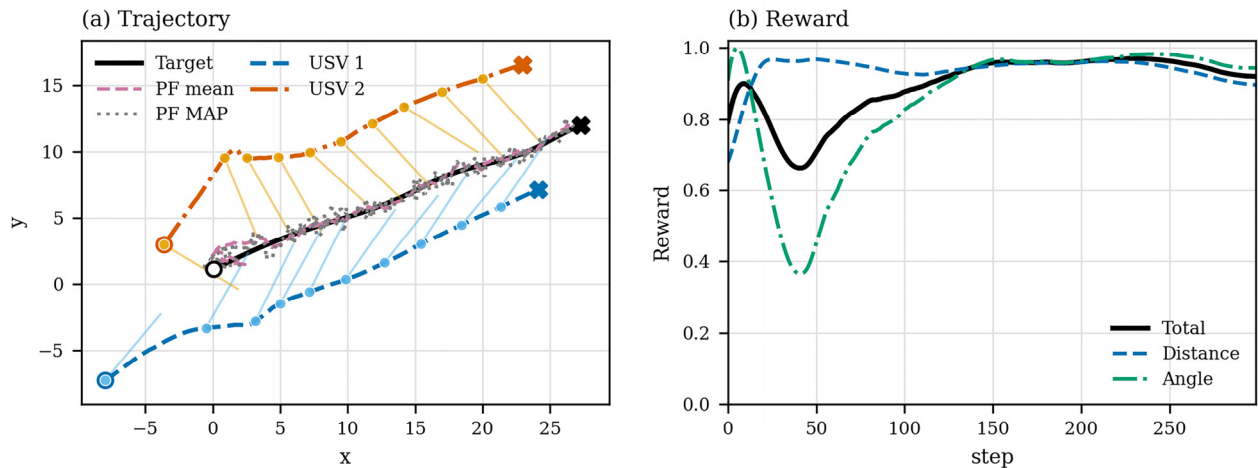


Fig. 3. Trajectories and rewards of the trained particle filter-MAPPO policy

는 경로만을 생성한 것이 아니라, 표적과의 목표 간격을 유지함과 동시에 두 USV가 표적 상태 추정에 유리한 관측 기하를 유지하기 위해 협력하였음을 보여 준다. 또한, 관측 거리와 관측 기하가 개선됨에 따라 입자 필터의 추정 성능도 개선됨을 확인할 수 있다.

4. 결론

본 연구에서는 수동 소나의 방위 정보만을 이용하여 수중의 표적 잠수함을 추적하기 위한 두 USV의 경로 생성 기법을 제안하였다. 수동 소나는 방위 정보만을 제공하기 때문에 관측자와 표적 간 상대 기하 위치가 관측의 정확도에 미치는 영향이 크다. 이를 위해 입자 필터와 MAPPO를 결합하여 상태 추정의 정확도를 높이기 위한 USV의 기동 경로를 생성하였다.

시뮬레이션 결과 표적 잠수함의 위치와 속도의 추정 오차는 각각 0.3604, 0.2948이었으며, 생성된 경로는 표적과 목표 거리를 유지한 상태에서 두 방위 벡터가 직교에 가까운 관측 기하를 유지함을 확인할 수 있었다.

본 연구는 시뮬레이션의 결과로 성능을 검증하였기에 향후 다양한 해양 환경 요소의 반영과 표적 잠수함 및 USV의 세부 동역학 등을 반영하는 방향으로 연구를 확장할 예정이다.

참고문헌

- [1] Dooyoung Kim, 'Online Sonobuoy Deployment Method with Bayesian Optimization for Estimating Location of Submarines,' *Journal of the Korea Institute of Military Science and Technology*, VOL. 25, NO. 1, 2022, pp. 72–81.
- [2] Kyung-Min Park & Dooyoung Kim, 'Preprocessing Performance of Convolutional Neural Networks according to Characteristic of Underwater Targets,' *The Journal of the Acoustical Society of Korea*, VOL. 41, NO. 6, 2022, pp. 629–636.
- [3] Dooyoung Kim, 'Approach Path Planning for Maritime Moving Target Tracking by Unmanned Surface Vehicles,' *Journal of the KNST*, VOL. 8, NO. 3, 2025, pp. 331–336.
- [4] Dooyoung Kim, 'Path Planning for Autonomous Unmanned Surface Vehicle Recovery: A Potential Field Approach,' *Journal of the KNST*, VOL. 7, NO. 4, 2024, pp. 428–432.
- [5] S. Nardone, A. Lindgren, & Kai Gong, 'Fundamental Properties and Performance of Conventional Bearings-Only Target Motion Analysis,' *IEEE Transactions on Automatic Control*, VOL. 29, NO. 9, 1984, pp. 775–787.
- [6] N. J. Gordon, D. J. Salmond, & A. F. M. Smith, 'Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation,' *IEE Proceedings F (Radar Signal Processing)*, VOL. 140, NO. 2, 1993, pp. 107–113.
- [7] Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, & Yi Wu, 'The Surprising Effectiveness of PPO in Cooperative Multi-Agent Games,' in *Advances in Neural Information Processing Systems*, VOL. 35, 2022, pp. 24611–24624.
- [8] Christopher Amato, 'An Introduction to Centralized Training for Decentralized Execution in Cooperative Multi-Agent Reinforcement Learning,' 2024, arXiv:2409.03052.