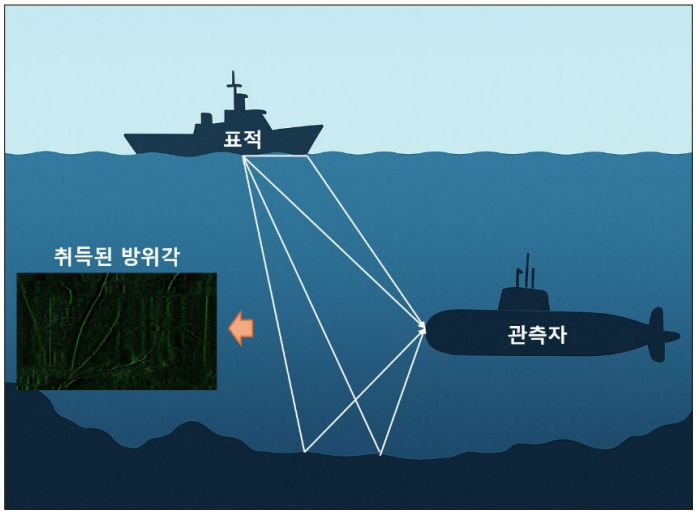
졸업논문 소개

강화학습 기반 해양 표적 기동 추적 자동화 기술 연구

장승환 / 한양대학교 지능형 영상 미디어 연구실(Intelligent Visual Media Lab.)

해양 환경에서의 표적 탐지 및 추적 기술은 군사적, 상업적, 과학적 분야에서 모두 핵심적인 역할을 수행한다. 특히 잠수함이나 수상함 같은 표적의 움직임을 정확히 파악하는 것은 작전 수행, 해양 안전 확보 등 다양한 분야에서 필수적인 요소다. 이러한 표적 추적은 일반적으로 수중 음향 센서, 즉 소나(SONAR)를 통해 수집한 방위각(Bearing) 데이터를 바탕으로 수행되며, 이때 사용되는 기법이 바로 표적 기동 분석(Target Motion Analysis; TMA)이다. 〈그림 1〉은 수중 환경에서 관측자가 표적의 소음을 소나를 통해 수집하는 예시이다.



<그림 1> 수중 표적 탐지 개념도 및 소나 방위각 수집 구조 예시

졸업논문 소개

기존 TMA 방식은 대부분 수동적인 분석 절차에 의존한다. 예를 들어, 소나 운영자는 방위각 정보를 토대로 표적의 속도와 방향을 추정하는데, 이는 경험과 직관에 기반한 어려운 작업이며, 연속적인 판단을 요구하기 때문에 피로도와 오류 가능성이 크다. 특히, DEMON(Detection of Envelope Modulation on Noise) 분석이나 배치 추정 알고리즘 기법과 같이 기존에 사용되는 자동화 기법조차도 해양 환경의 복잡성과 불확실성, 그리고 소나 신호에 포함된 노이즈로 인해 정확도가 떨어지는 한계를 갖는다. 해양 환경은 수심, 해류, 바닥 반사, 잡음 등 수많은 변수에 의해 영향을 받기 때문에, 센서 데이터의 신뢰성이 일정하지 않고, 동일한 조건에서도 같은 예측이 어렵다. 이로 인해 기존 알고리즘 기반의 예측 기법은 이러한 동적 환경에 유연하게 대처하지 못한다.

최근 인공지능 기술, 특히 강화학습(Reinforcement Learning; RL)의 발전은 이러한 문제에 대한 새로운 해결책을 제시하고 있다. 강화학습은 에이전트가 환경과 상호작용하며 보상을 최대화하는 방향으로 스스로 정책을 학습하는 방식이기 때문에, 예측 불가능하고 복잡한 환경에서도 안정적인 성능을 기대할 수 있다. 특히 PPO (Proximal Policy Optimization)와 같은 정책 기반 알고리즘은 탐색과 활용 사이의 균형을 잘 유지하며, 연속적인 상태와 행동 공간에서도 효과적인 학습이 가능하다는 장점이 있다.

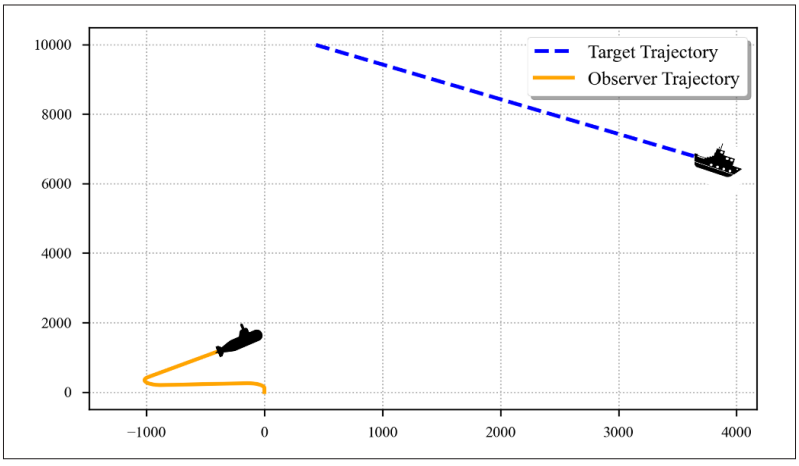
따라서 **본 연구의 목적은 기존 TMA 방식이 인간 운영자의 해석 능력에 크게 의존한다는 한계를 극복하고, 해양 환경에서도 높은 정확도와 일관성을 유지할 수 있는 자동화된 표적 속도 추정 시스템을 강화학습 기반으로 개발**하는 것이다. 이를 통해 수중 작전의 신속한 판단과 대응을 가능하게 하고, 운영자의 부담을 줄이며, 전반적인 표적 추적 성능을 향상시킬 수 있는 기반 기술을 마련하고자 한다.

이를 위해 PPO 알고리즘을 기반으로 한 강화학습 프레임워크를 설계하고, 에이전트가 다양한 해양 환경 시나리오에서 표적의 속도를 추정할 수 있도록 학습시킨다. 특히, 실제 해양 소나 데이터의 특성을 반영하기 위해 시뮬레이터를 직접 구현하고, 다양한 초기 위치, 속도, 방위각 등의 변수를 포함한 다수의 시나리오를 생성하여 학습 환경을 구성한다.

실제 환경에서의 적용 가능성을 높이기 위해 가우시안 노이즈와 같은 관측 오차를 시뮬레이션에 포함시켜 모델의 일반화 성능을 확보하고자 한다. 강화학습의 상태(state), 행동(action), 보상 함수(reward)를 TMA 문제에 적합하도록 정의함으로써, 에이전트가 점진적으로 표적 속도를 예측하고 유지할 수 있는 최적의 전략을 학습하도록 한다.

해양 환경에서 표적의 속도를 자동으로 추정하기 위한 강화학습 기반 TMA 시스템을 구현하기 위해, 시뮬레이션 환경 구축, 상태, 행동, 보상 정의, 그리고 PPO 기반 학습 전략으로 구성된 전체적인 방법론을 설계하였다.

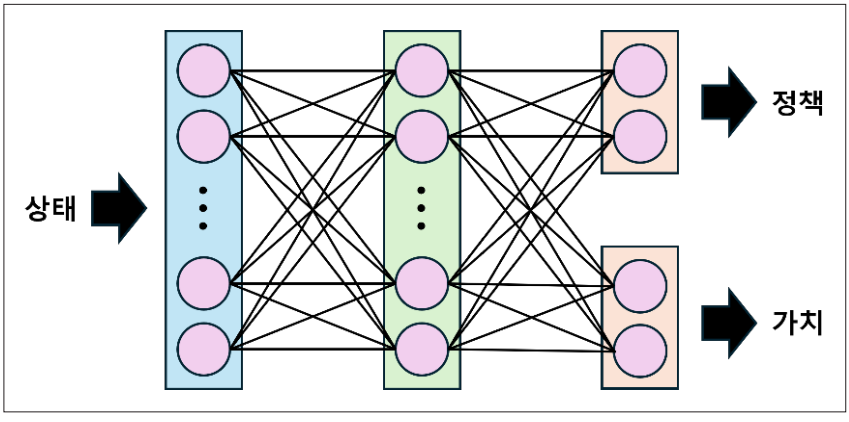
먼저, 강화학습 에이전트가 실제와 유사한 환경에서 학습할 수 있도록 사용자 정의 시뮬레이터를 구현하였다. 이 시뮬레이터는 관측자(Observer)와 표적(Target) 간의 상대적인 움직임을 기반으로 방위각 데이터를 생성하며, 배치 추정 알고리즘을 통해 속도 추정의 정확도를 판단하는 환경을 제공한다. 관측자는 일정한 속도로 이동하며 두 차례의 방향 전환을 통해 관측 가능성을 확보하고, 표적은 초기 위치와 속도, 진행 방향이 각각 다른 여러 시나리오에서 직선 등속 운동을 수행한다. 방위각 데이터는 1초 간격으로 생성되며, 실제 환경 소나 데이터의 특성을 반영하기 위해 표준편차 0.5도의 가우시안 노이즈가 추가된다. 이로써 실제 해양 환경에서 발생할 수 있는 관측 오차를 반영한 학습 환경을 구현하였다. 〈그림 2〉는 시뮬레이션된 관측자와 표적의 움직임의 경로를 나타낸 것이다. 이로 인해 관측자로부터 표적의 방위각 정보를 얻을 수 있다.



<그림 2> 시뮬레이션된 관측자와 표적의 경로

강화학습의 핵심 요소인 상태(state), 행동(action), 보상(reward)은 TMA 문제에 맞게 구성하였다. 상태는 에이전트가 예측한 속도와 이에 대한 배치 추정 결과(J값)를 쌍으로 하여 최근 3회의 정보를 입력 값으로 사용한다. 이를 통해 에이전트는 속도 조정의 결과로 J값이 어떻게 변화했는지를 학습할 수 있다. 행동은 5개의 이산적 선택지로 구성되며, 현재 속도에서 ±1m/s 또는 ±0.1m/s로 조정하거나 유지하는 방식이다. 이를 통해 에이전트는 미세 조정과 큰 폭의 조정을 상황에 따라 유연하게 선택할 수 있다.

보상 함수는 속도 예측의 정확도와 유지 능력을 모두 반영할 수 있도록 설계하였다. 우선, 에이전트가 올바른 속도에 가까워질수록 양의 보상을 주고, 멀어질수록 음의 보상을 주는 방식으로 구성하였다. 정확한 속도에 도달한 이후에는 이를 안정적으로 유지할 경우 추가적인 보상이 주어지며, 이를 연속적으로 유지할수록 누적 보상이 증가하도록 설계하였다. 이와 같은 보상 구조는 에이전트가 빠르게 정답 속도에 수렴한 뒤, 해당 속도를 지속적



<그림 3> Actor-Critic 네트워크 구조

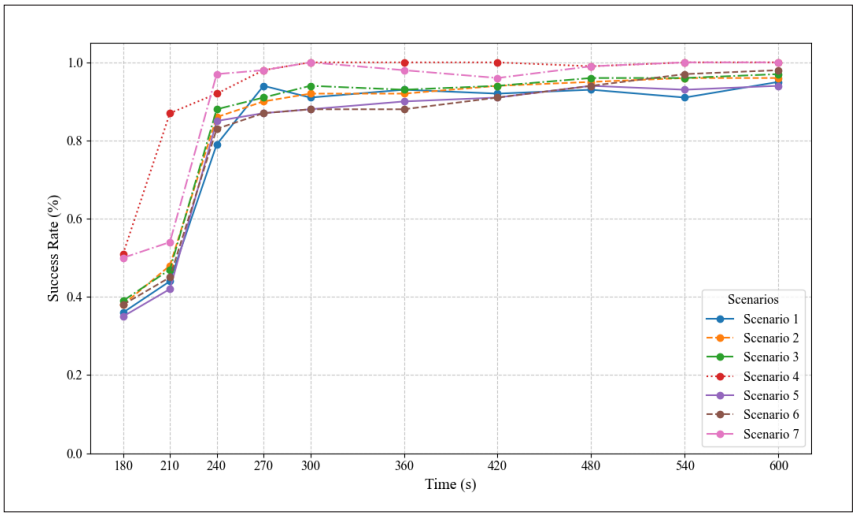
으로 유지하는 전략을 자연스럽게 학습하도록 유도한다.

본 연구에서는 PPO 알고리즘의 Actor-Critic 구조를 채택하였으며, 공통된 피처 추출 층을 기반으로 정책 네트워크(Actor)와 가치 함수 네트워크(Critic)가 병렬적으로 구성된다. 〈그림 3〉은 Actor-Critic 구조 그림이며, 손실 함수는 정책 손실, 상태값 손실, 엔트로피 손실로 구성되며, 이들 간의 가중치를 조정하여 학습의 균형을 맞추었다. 또한 학습 안정성과 성능 향상을 위해 학습률, 클리핑 계수, 할인율 등 주요 하이퍼파라미터들을 실험적으로 조정하였다.

최종적으로 에이전트는 600초 간의 방위각 데이터를 기반으로 학습되었으며, 학습 종료 조건은 예측 속도가 실제 속도와 1m/s 이하의 오차 범위에 도달한 이후, 해당 속도를 일정 시간 동안 유지하는 것으로 설정하였다. 학습이 완료된 후에는 다양한 초기 조건 및 노이즈 환경에서 100회의 테스트 시나리오를 수행하여 모델의 일반화 성능과 강건성을 평가하였다.

제안한 강화학습 기반 표적 운동 분석 시스템은 다양한 시나리오와 관측 조건에서 안정적인 속도 추정 성능을 보였다. 〈그림 4〉는 timestep에 따른 여러 시나리오의 성능을 나타낸 것이다. 어느 정도의 충분한 시간 정보가 주어진다면, 대부분의 시나리오는 높은 성공률을 기록하였으며, 노이즈가 포함된 환경이나 관측 시간이 달라지는 조건에서도 강건한 예측 능력을 유지하였다. 이는 PPO 기반 에이전트가 복잡하고 불확실한 해양 환경에서도 효과적으로 학습할 수 있음을 보여준다.

본 연구는 기존 TMA 방식의 한계를 극복하고, 강화학습을 통해 자동화된 속도 추정 시스템을 구현할 수 있음을 실증하였다. 학습된 에이전트는 사람의 개입 없이도 표적 속도를 정확히 예측할 수 있으며, 이는 해양 감시 체계의 지능화 및 자동화에 실질적인 기여를 할 수 있다. 향후에는 배치 추정의 신뢰성 향상, 다중 표적 추적, 실시간 처리 기능 등을 추가함으로써 보다 실용적인 시스템으로 확장할 수 있을 것이다.



<그림 4> Timestep에 따른 시나리오별 성능

장 승 환

- 2023년 2월 : 한양대학교 에리카-캠퍼스 전자공학부 학사
- 2025년 8월 : 한양대학교 전자공학과 석사
- 2025년 7월 ~ 현재 : LIG넥스원 해양연구소 해양통합체계 개발단 4팀 연구원
- 주관심분야 : 영상처리, 강화학습