잠수함 어뢰기만전술을 위한 다중 하위 계층 에이전트 기반 계층적 강화학습

강보선, 윤원혁*

LIG Nex1

boseon.kang@lignex1.com, wonhyuk.yun2@lignex1.com*

Multi Low-Level Agent Based Hierarchical Reinforcement Learning for Submarine Torpedo Countermeasures

BoSeon Kang, WonHyuk Yun*

LIG Nex1

요 약

기존 강화학습 방법론은 상태 공간이 증가하거나 희소·지연 보상 문제에 직면하면 효율적으로 학습되지 않는다는 문제점이 존재한다. 이를 해결하기 위해 사용되는 기존 계층적 강화학습 알고리즘은 설계자의 역량에 따라 상위, 하위 계층 간 협력이 불안정해지는 성능 저하의 단점이 있다. 따라서 본 논문에서는 다중 하위 계층 에이전트 기반의 계층적 강화학습 알고리즘을 적용한 잠수함 어뢰기만전술을 제안한다. 상위 계층은 스스로 세부 목표를 설정하여 기만기 발사 각도와 현재 상황에서 가장 적합한 하위 계층 에이전트를 선택하고, 다수의 하위 계층 에이전트는 각 세부 목표마다 전문화될 수 있도록 정책을 학습한다. 제안하는 알고리즘이 기존 강화학습 알고리즘보다 학습 성능이 향상될 수 있음을 보여준다.

Ⅰ. 서 론

최근 해양 전장에서는 고해상도 센서와 정밀 유도 어뢰의 발전으로 인해 잠수함의 탐지 위험이 증가하고 있다. 이에 대응하기 위해 강화학습 기반 잠수함 어뢰기만전술에 관한 연구가 진행되고 있다.

그러나 기존 강화학습 알고리즘은 상태 공간이 증가할수록 강화학습 에이전트가 탐험, 탐색해야 하는 후보 행동의 수가 기하급수적으로 늘어나 학습 속도가 저하되며 결국 최적 정책으로 수렴하기 어려워진다. 또한, 잠수함이 목표 심도, 방위에 도달하거나 어뢰 회피에 성공해야만 보상이 발생하므로 희소·지연 보상 문제에 직면하여 장기적인 목표를 고려하지 못하고 단기 목표에 치우치는 경향이 있다.

해당 문제를 해결하기 위해 계층적 강화학습 알고리즘이 도입되었다. 계층적 강화학습은 학습 환경의 상태 공간을 단계적으로 줄이고, 상위 정책이 장기 목표를 제시함으로써 희소·지연 보상 문제를 해결하고 학습 속도를 향상할 수 있는 장점이 있다. 그러나 기존 계층적 강화학습 연구들은 사용자가 계층 구조를 직접 설계해야 했으며, 설계자의 역량에 따라 상위, 하위 계층 정책 간 협력이 불완전하여 성능 차이가 발생하는 단점이 있다.

따라서, 본 논문에서는 설계자의 역할을 최소화하기 위해 다중 하위 계층 에이전트 기반의 계층적 강화학습 알고리즘을 적용한 잠수함 어뢰기만전술을 제안한다. 제안하는 모델은 하나의 상위 계층 에이전트와 다수의 하위 계층 에이전트 구조를 활용한다. 상위 계층 에이전트는 환경과 장기 목표를 스스로 분석하여 중간 목표를 설정하고, 다중 하위 계층 에이전트는 상위 계층 에이전트로부터 전달받은 목표에 따라 단기 행동 정책 학습에 집중한다. 이러한 일대다 구조와 동적 선택 메커니즘은 학습 탐색 공간을 효율적으로 분할하고, 설계자의 개입을 최소화하며, 장기 목표 달성과 최종 정책 성능 향상에 효율적이다. 상위 계층 에이전트가 여러 개의 하위 계층 에이전트 중에서 전술적으로 가장 적합한 에이전트를 선택함으로써 기존 계층적 강화학습 알고리즘보다 학습 성능이 향상될 수 있음을 보여준다.

Ⅱ. 관련 연구

A. 계층적 강화학습

계층적 강화학습은 복잡하거나 보상이 희소한 환경에서 기존 강화학습 알고리즘의 확장성 문제를 해결하기 위한 접근법이다. 전체 문제를 여러 계층으로 나누어 학습하는 방식을 활용한다. 상위 계층 정책은 주로 추상화된 상태 정보나 세부 목표를 다루며 긴 시간 단위로 동작하고, 하위 계층 정책은 실제 환경과 상호작용하며 짧은 시간 단위로 행동을 수행하여 상위 계층이 설정한 세부 목표를 달성할 수 있도록 학습한다.

이러한 계층 구조는 전체 탐색해야 하는 상태, 행동 공간의 복잡성을 줄여 학습의 효율성을 높인다. 또한, 하위 계층에서 학습된 정책은 특정 세부 목표를 달성하도록 특화되며, 상위 계층에서는 세부 목표를 자동으로 발견하고 하위 계층으로 전달함에 따라서 설계자의 역할을 최소화하며 효율적인 탐색과 강건한 학습이 가능해진다.

Ⅲ. 다중 하위 계층 에이전트 기반의 잠수함 어뢰기만전술 모델

다중 하위 계층 에이전트 기반 계층적 강화학습을 활용하여 잠수함 어뢰기만전술을 학습하기 위해서는 수중 시뮬레이션 환경이 필수적이다. 본 논문에서는 기존에 개발된 6-자유도 기반으로 수중 운동체를 정확하게 모사하고 해양 환경에 따른 수중 음속 변화를 고려하여 현실과 유사한 음향 탐지 절차가 구현된 수중 시뮬레이션 환경을 사용한다. [1] 수중 시뮬레이션 환경을 기반으로 하는 환경 모델은 상태 공간, 행동 공간, 보상 함수를 계층별로 정의해야 한다. 다중 하위 계층 에이전트 기반 계층적 강화학습

을 적용한 잠수함 어뢰기만전술 모델의 구조는 그림 1과 같다.

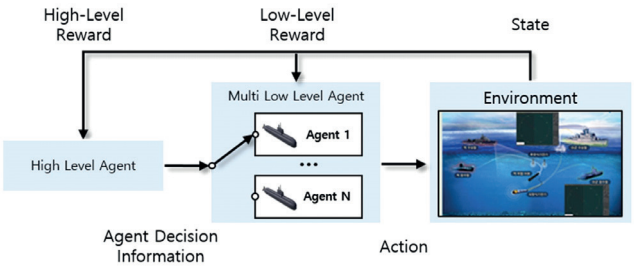

그림 1. 다중 하위 계층 에이전트 기반의 계층적 강화학습 구조

A. 상위 계층 에이전트

상위 계층 에이전트는 전반적으로 전술적인 판단을 담당하며 행동 공간은 최적의 기만기 발사 각도와 현재 상황에 가장 적합한 하위 계층 에이전트 선택 정보로 정의된다. 정확한 세부 목표 설정을 위해 어뢰 경보 정보, 어뢰와 기만기의 상대 방위각 정보 그리고 사용 가능한 기만기 장입 정보로 상태 공간을 정의한다. 보상 함수는 잠수함의 어뢰 회피에 초점을 맞추어, 어뢰에 피격되었을 경우 큰 패널티를 받도록 설계한다.

B. 하위 계층 에이전트

다수의 하위 계층 에이전트는 각각 특정 방위각에서 다가오는 어뢰에 대한 기만전술 또는 회피침로 산출에 특화될 수 있으며, 상위 계층 에이전트로부터 전달받은 세부 목표를 달성하기 위해 행동을 수행한다. 하위 계층 에이전트의 상태 공간은 상위 계층 에이전트가 수립한 세부 목표, 기만기 발사 정보를 포함하며, 현재 잠수함의 헤딩, 심도, 속도, 어뢰 경보 정보, 어뢰와 기만기의 상대 방위각 정보로 정의된다. 선택된 하위 계층 에이전트는 잠수함의 목표 헤딩, 심도, 속도를 결정할 수 있는 행동 공간을 가진다. 보상 체계는 매 타임스텝마다 잠수함과 어뢰와의 거리를 기반으로 보상을 획득함으로써 즉각적인 어뢰 위협 회피에 중점을 둔다. 어뢰 경보 발생 또는 어뢰가 잠수함을 목표로 지정하는 것과 같은 특수한 이벤트가 발생하면, 상위 계층 에이전트가 재호출되어 새로운 세부 목표를 수립함으로써 효율적인 어뢰기만전술 수행을 가능하게 한다.

Ⅳ. 실험

제안하는 계층적 강화학습 알고리즘과 기존 강화학습 알고리즘들의 성능 비교를 위해 Ray RLlib [3]를 활용하였으며, 학습 알고리즘으로는 PPO [4]를 사용하였다. 각 알고리즘의 전체 상태, 행동 공간 그리고 하이퍼파라미터를 동일한 설정 하에 학습을 진행하였다. 실험 환경에서 어뢰는 초기 생성 거리 900m에서 총 2발 생성되며, 각 어뢰는 0도에서 360도 사이의 임의 방위각으로 설정된다. 모델의 성능 평가는 각 Iteration에 포함된 다수의 에피소드 중에서 에이전트가 최종적으로 어뢰 회피에 성공한 에피소드의 비율을 나타내는 성공 비율 (Success Rate)을 지표로 사용한다.

그림 2. 강화학습 알고리즘별 학습 성능 그래프

실험 결과, 단일 에이전트 기반 강화학습 모델은 학습이 진행됨에 따라 성능이 개선되기는 하였으나 수렴 속도가 매우 느린 경향을 보였다. 일대일 계층적 강화학습 모델의 경우 특정 시나리오에서는 효과적일 수 있으나 어뢰 위협 상황에 유연하게 대처하기 위한 구조 변경이나 재설계가 빈번히 요구되어 최종 성능은 낮게 나타났다. 제안하는 다중 하위 계층 에이전트 기반의 계층적 강화학습 모델은 다른 두 모델과 비교하여 높은 성공 비율을 달성하고 안정적으로 수렴하는 성능을 보여준다.

Ⅴ. 결론

본 연구에서는 다중 하위 계층 에이전트 기반 계층적 강화학습을 사용하는 잠수함 어뢰기만전술을 제안했다. 새로운 계층적 강화학습 방법론을 적용하기 위해 상위, 하위 계층 에이전트를 설계하였다. 난이도가 높은 수중 환경에서 기존 강화학습 알고리즘들보다 안정적으로 학습할 수 있음을 확인하였다. 향후 연구에서는 커리큘럼 러닝을 적용하여 수렴 속도와 성능을 고도화하는 방안을 모색하고자 한다.

ACKNOWLEDGMENT

이 논문은 2022년도 정부(방위사업청)의 재원으로, 국방기술진흥연구소의 지원을 받아 수행된 연구임 (No. KRIT-CT-22-023-03, 잠수함 어뢰기만전술 고도화 기술)

참 고 문 헌

[1] Bakker, Bram, and Jürgen Schmidhuber. “Hierarchical reinforcement learning based on subgoal discovery and subpolicy specialization.” Proc. of the 8-th Conf. on Intelligent Autonomous Systems, pp. 435-445, 2004.

[2] Kang, B., and Yun, W., “Hierarchical Reinforcement Learning for Submarine Torpedo Countermeasures and Evasive Manoeuvres,” IEEE Access, vol. 12, pp. 170620 - 170631, 2024.

[3] P. Moritz, R. Nishihara, et al., "Ray: A distributed framework for emerging AI applications", Proc. 13th USENIX Symp. Operating Syst. Design Implement. (OSDI), pp. 561-577, 2018.

[4] Schulman, John, et al., "Proximal policy optimization algorithms." arXiv preprint, 2017.