자율운항선박을 위한 강화학습 에이전트 기술 연구

오유찬[1], 양소희[2], 이재훈[3], 황윤주[4], 이규영[5]
[1]서울시립대학교 컴퓨터과학부 학부생
[2]동덕여자대학교 세무회계학과 학부생
[3]서울과학기술대학교 스마트 ICT 융합공학과 학부생
[4]동덕여자대학교 데이터사이언스전공 학부생
[5]한국과학기술원 정보보호대학원 박사수료

yuchan5@uos.ac.kr , skg703@naver.com , dlwognsdc610@naver.com, ilydh5018895@gmail.com, leeahn1223@kaist.ac.kr

A Study on Reinforcement Learning Agent Technology for Autonomous Ships

Yu-Chan Oh[1], So-Hee Yang[2], Jae-Hoon Lee[3], Yun-Ju Hwang[4], Ku-Yeong Lee[5]
[1]Dept. of Computer Science, University of Seoul
[2]Dept. of Tax Accounting, Dongduk Women's University
[3]Dept. of Smart ICT Convergence Engineering, Seoul National University of Science and Technology
[4]Dept. of Data Science, Dongduk Women's University
[5]Graduate School of Information Security, KAIST

요 약

기존의 자율운항선박 연구에는 전통적인 AI 기술들이 사용되어 왔다. 그러나 이러한 기술들은 특정 조건에 맞춘 규칙과 추론 방식으로 작동하기 때문에, 다양한 변수가 있는 환경에서 최적의 성능을 발휘하기는 어렵다. 이에 본 연구는 자율운항선박에서 가장 중요한 경로최적화와 충돌회피 과제 해결에 강화학습이 효과적임을 실험을 통해 입증하고 최적의 강화학습 알고리즘을 제시한다.

1. 서론

자율운항선박은 현재 자동화 수준(LOA) 3 단계로, 선원 없이 원격으로 운항이 가능하다. 자율운항선박의 발전은 해운 산업의 효율성과 안전성을 개선할 것으로 기대되며, 관련 연구가 활발하게 진행 중이다[1]. 기존 연구는 전문가 시스템 등 전통적인 AI 기술에 의존했으나, 동적 환경 변화에 적응하지 못하고 복잡한 상황에서 성능 저하라는 한계를 보였다[2].

본 연구에서는 Deep-SARSA 및 A2C 강화학습 알고리즘을 구현하고, 이를 자율운항선박 모의환경에서 실험하여, 적합도와 알고리즘 간 성능을 분석하였다.

2. 강화학습 이론

강화학습은 에이전트가 사전지식 없이 환경에서 행동을 반복하며 시행착오를 통해 최대 누적 보상을 목표로 최적 정책을 학습하는 인공지능 기술이다.[3]

2.1 Deep-SARSA vs A2C(Advantage Actor-Critic)[4]

1) Deep-SARSA: 식(1)과 같이 에이전트가 탐욕정책으로 액션을 선택하면, 환경이 보상과 다음 상태정보를 리턴하고, 에이전트가 다음 상태의 액션을 한 번 더 결정한 샘플로 큐함수를 업데이트하는 것이 SARSA 이며, 여기에 딥러닝 모델을 도입한 것이다.

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha(R_{t+1} + \gamma Q(S_{t+1}, A_{t+1}) - Q(S_t, A_t)) \quad (1)$$

2) A2C: 하나의 에피소드가 끝나야 그 반환 값으로 학습을 진행할 수 있는 REINFORCE 알고리즘의 단점을 TD 방식으로 극복한 기술이다. Advantage 는 식(2)와 같이 액션가치에서 상태가치를 뺀 값이며, 이 Advantage 를 적용한 손실함수로 학습을 진행한다. 선택한 행동의 적합성과 더불어 영향도까지 학습한다.

$$A(S_t, A_t) = Q_w(S_t, A_t) - V_v(S_t) \quad (2)$$

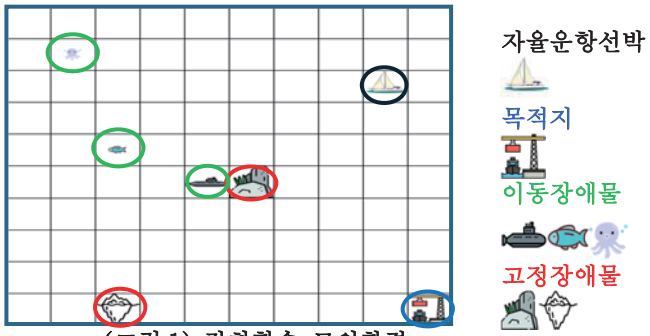
3. 관련 연구

강화학습을 이용하여 선박의 자율 운항을 시뮬레이션하고 충돌을 예측하며 회피하는 연구가 진행되고 있다.[5] Q-Learning 기반 강화학습을 통해 해양사고 시 최적의 퇴선 경로를 산출하고, 학습 횟수에 따른 결과를 비교 검토한 연구가 있다.[6]

4. 실험

4.1 실험 환경

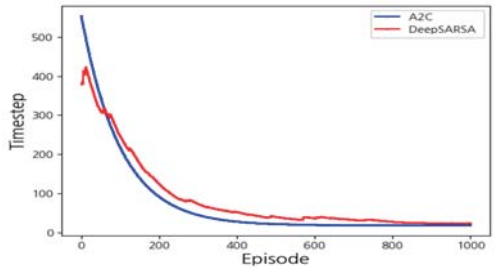
아래 그림(1)과 같은 환경에서 Deep-SARSA 와 A2C 를 각기 다른 모델에서 1000 회의 Episode 를 실행하여 학습시킨 후 성능을 비교하였다.



(그림 1) 강화학습 모의환경

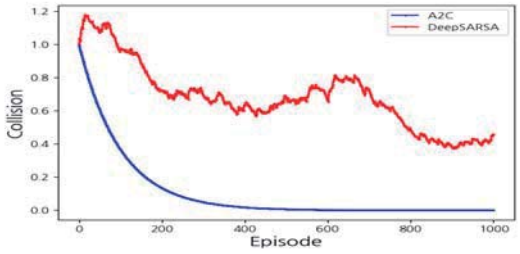
(0,0)에서 시작하여 (9,9)를 목적지로 하는, 10*10 크기의 강화학습 모의환경을 구축하였다. 자율운항선박과 충돌을 일으킬 수 있는 고정 및 이동 장애물을 설정하였다. 각 Episode 시작 시, 이동 장애물은 상하좌우로 [1,3] 사이의 이산적인 무작위 속도로 이동한다. 각 Timestep 에서 에이전트는 고정장애물 및 이동장애물의 상대 위치와 종류, 목적지의 상대위치를 State 로 입력 받는다. 에이전트는 그리드의 경계를 넘어 이동하지 않으며, 장애물과 충돌할 경우 마이너스 보상을 받고, 목적지에 도착할 경우 양의 보상을 받고 해당 에피소드를 완료한다.

4.2 학습단계 실험결과



(그림 2) 학습단계 에피소드별 Timestep 실험결과

그림(2)를 보면, 에피소드가 진행함에 따라 목적지까지의 스텝수가 감소하며 수렴하는 추이를 보여주고 있으며, Deep-SARSA 와 A2C 모두 비슷한 수준이다.



(그림 3) 학습단계 에피소드별 충돌횟수 실험결과

그림(3)을 보면, Deep-SARSA와 A2C 모두 적은 충돌을 통해 최대 리턴 경로를 찾도록 학습하고 있다. Deep-SARSA 는 충돌횟수가 완만하고 정체구간이 큰 수렴추이를 보이는 반면, A2C 는 충돌횟수가 비선형적으로 빠르게 감소하고 안정적으로 수렴하는 모습을 보여주고 있다. 이는 학습과정에서 정책신경망과 가치신경망을 모두 사용하는 A2C 가 더 나은 수렴속도 및 충돌 회피 성능을 보여준 것으로 판단된다.

4.3 테스트단계 실험결과

표(1)은 학습을 완료한 에이전트를 테스트한 결과이다. A2C 모델의 에이전트는 Deep-SARSA 모델의 에이전트보다 3 배 이상 충돌횟수가 적으며, 이는 회피 성능이 상대적으로 더 우수함을 뜻한다.

구분	평균 충돌횟수	평균 타임스텝
Deep-SARSA	0.4	18.18
A2C	0.14	18.46

(표 1) 테스트단계 성능측정 실험결과

5. 결론

이번 실험에서는 A2C 와 Deep-SARSA 알고리즘 모두 자율운항선박의 경로 최적화와 장애물 회피 문제에서 일정 성과를 달성한 것으로 나타났다. A2C 는 학습 수렴 속도와 장애물 회피 측면에서 더 나은 성능을 보였으며, 이는 실시간 자율운항선박 환경에서의 효율적인 적용 가능성을 시사한다.

ACKNOWLEDGEMENT

※ 본 논문은 해양수산부 실무형 해상물류 일자리 지원사업(스마트해상물류 x ICT 멘토링)을 통해 수행한 ICT 멘토링 프로젝트 결과물입니다.

※ 본 논문에 참여한 저자들은 모두 공동 1 저자이며, 논문작성에 기여한 정도가 같습니다.

참고문헌

[1] Poornikoo, M., Ø vergård, K.I. Levels of automation in maritime autonomous surface ships (MASS): a fuzzy logic approach. Marit Econ Logist 24, 278–301 (2022).

[2] 이동훈. "가변적 선박 안전영역 및 충돌 위험지수를 반영한 선박 충돌 회피 시스템에 관한 연구." 국내석사학위논문 인하대학교 대학원, 2023. 인천

[3] 이웅원, 양혁렬, 김건우, 이영무, 이의령,파이썬과 케라스로 배우는 강화학습,경기도,위키북스,2020

[4] Jiménez, Gonzalo & Hueso, Arturo & Gómez-Silva, Maria. Reinforcement Learning Algorithms for Autonomous Mission Accomplishment by Unmanned Aerial Vehicles: A Comparative View with DQN, SARSA, and A2C. Sensors. 2023

[5] 임지수, 데이터 기반의 강화학습을 통한 선박 자율 운항에 관한 연구, 석사학위논문, 고려대학교, 2018

[6] 김원욱, 김대희, 윤대근. (2018). AI 기법의 Q-Learning 을 이용한 최적 퇴선 경로 산출 연구. 해양환경안전학회지, 24(7), 870-874.