자율운항선박 강화학습 에이전트의 충돌회피 성능향상에 관한 연구

손건희[1], 이연재[2], 조연수[3], 김희준[4], 정현천[5], 안현주[6]
[1]국립공주대학교 인공지능학부 학부생
[2]숭실대학교 소프트웨어학부 학부생
[3]숭실대학교 전자정보공학부 IT융합학과 학부생
[4]고려대학교 컴퓨터 융합소프트웨어학과 학부생
[5]한국해양대학교 조선해양시스템공학과 학부생
[6]리안기술사사무소 이사
songunhee5426@gmail.com, hicassey1005@gmail.com, duath0604@naver.com,
kktongkr@korea.ac.kr, zzz6426@naver.com, suzic@nate.com

A Study on Enhancing the Collision Avoidance Performance of Reinforcement Learning Agents in Autonomous Ships

Kun-Hee Son[1], Young-Hee Lee[2], Yeon-Su Cho[3], Hee-June Kim[4],
Jeong-HyeonCheon[5], Hyun-Joo An[6]
[1]Dept. of Artificial Intelligence, Kongju National University
[2]Dept. of Software, Soongsil University
[3]Dept. of Electronic Engineering (IT Convergence), Soongsil University
[4]Dept. of Computer Convergence Software, Korea University
[5]Dept. of Naval Architecture and Ocean Systems Engineering, KMOU
[6]Lee-Ahn Professional Engineer's Office

요 약

강화학습은 다양한 환경에서 최적의 의사결정을 내릴 수 있도록 지원하는 인공지능기술로서, 이를 자율운항 선박에 적용하면 복잡한 해상 상황에서도 안전하게 항해할 수 있다. 본 논문에서는 다양한 강화학습 알고리즘에 대해 운항 성능을 비교 및 분석한 결과, DQN 알고리즘이 자율운항선박의 경로 최적화와 충돌회피 성능 측면에서 가장 우수하였다.

1. 서론

선박의 자율운항시스템 및 항해보조시스템에는 선박 간 충돌회피 기능이 탑재되어 있다. 이는 자선의 안정성을 확보하고 인명과 재산의 손실, 기름 유출로 인한 환경오염과 같은 문제를 방지하기 위함이다. 이에 따라 충돌회피시스템의 안정성에 대한 요구가 높아져 왔으며, 자율운항시스템 및 항해보조시스템의 충돌회피 능력을 향상시키기 위한 연구가 활발히 이루어져 왔다[1].

본 논문에서는 강화학습 기술을 이용해서 자율운항시스템의 충돌회피 능력 향상을 위한 연구를 진행하였다. 강화학습은 에이전트가 환경과 상호작용하며 누적보상을 최대화하기 위해 최적의 행동을 학습하는 기술로서, 복잡한 해상 상황에서 자율적으로 의사결정을 내릴 수 있도록 돕는다. 본 논문에서는 강화학습을 대표하는 기법인 DQN과 REINFORCE 알고리즘을 적용하고 그 성능을 비교 및 평가하였다.

이를 통해 자율운항선박의 안전한 항해에 적합한 강화학습 기법을 제안하고자 한다.

2. 관련 연구

강화학습을 대표하는 주요 알고리즘에 대한 설명은 다음과 같다.

(1) DQN (Deep Q-Network)

에이전트가 경험한 행동과 상관없이 다음 상태에서 가능한 모든 행동 중 최대 Q값을 추론하여 현재의 Q값을 갱신하는 방법으로, 식(2)와 같다.

$$Q(s,a) = E_{s' \sim \epsilon}[r + \gamma max Q(s',a') | s,a] \quad (1)$$ [3].

(2) REINFORCE

정책을 직접 조절하여 장기적인 보상을 극대화하고, 보상의 합(Gt)이 클수록 선택할 확률을 키우면

서 파라미터를 업데이트하며 에피소드마다 정책을 반복적으로 개선해 나가며 최적의 행동 전략을 찾는 방법으로, 아래의 식(2)과 같다.

$$\theta \leftarrow \theta + \alpha \gamma^t G_t \nabla_\theta \log \pi_\theta(s_t, a_t) \qquad (2)$$

3. 실험

본 연구에서는 그림(1)과 같은 10x10 크기의 2차원 환경을 설정하였다. 상태 공간은 현재 위치, 장애물, 목표 지점의 상대적 위치 정보를 포함하고, 행동 공간은 상, 하, 좌, 우 네 가지로 구성하였다. 20개의 고정 장애물을 추가하여 목표 도달 시 +10, 장애물 충돌 시 -10의 보상 체계를 구성했다. 훈련은 1000 에피소드, 테스트는 30 에피소드로 진행하였다.

(그림 1) 자율운항선박 해상 항해 모의환경

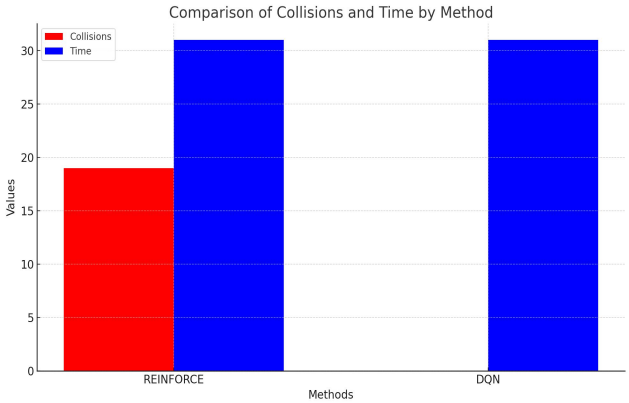
학습을 완료한 후, 30 에피소드로 테스트를 진행하였으며, 에피소드 동안 장애물과의 충돌횟수와 목적지까지의 소요시간으로 에이전트의 성능을 평가하였다.

<표 1> 강화학습 기법에 따른 성능 비교

성능 항목	REINFORCE	DQN
충돌 횟수	19	**0**
Time(초)	31	**31**

표(1)의 결과를 보면, REINFORCE는 연속적인 행동 공간에서 유리하기 때문에, 소요시간이 31초로 짧아 적은 연산으로도 일정 수준의 성능을 발휘하였으나, 충돌횟수(19회) 측면에서는 DQN만큼 뛰어나지 못하였다.

DQN은 소요시간이 31초로 짧고, 장애물과의 충돌은 전혀 발생하지 않았다. 이렇게 가장 우수한 성과를 거둔 것은, DQN이 Deep-SARSA의 한계를 보완한 알고리즘으로서, Experience Replay Buffer를 사용하여 시계열데이터의 지역성을 극복하고 Seperate Network을 통해 안정적인 정답을 가치망에 공급하기 때문으로 판단된다.



(그림 2) 강화학습 알고리즘 별 수행성능 비교

4. 결론

본 논문에서는 강화학습의 알고리즘에 따라 자율운항선박의 운항 성능을 비교 및 분석하였다.

그 결과 DQN이 더 많은 지역을 탐색하며 최적의 경로를 찾아내어 가장 좋은 성능을 보인 것을 확인할 수 있었다. 향후에는 해상 모의환경의 크기를 확장하고 이동 장애물을 추가하는 등 보다 복잡한 환경에서 실험할 예정이다. 강화학습 에이전트에 관한 연구가 빠르고 안전한 자율운항 발전에 크게 기여할 것이라 확신한다.

ACKNOWLEDGEMENT

- 본 논문에 참여한 저자들은 모두 공동1저자이며, 논문작성에 기여한 정도가 같습니다.
- 본 논문은 해양수산부 실무형 해상물류 일자리 지원사업(스마트해상물류 x ICT멘토링)을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.

참고문헌

[1] 김희수, 2019, 선박안전영역이 고려된 최적속도 벡터 선정을 통한 운항 선박의 지역경로계획에 관한 연구, 석사학위논문, 인하대학교

[2] Zhi-xiong Xu, Lei Cao, Xi-liang Chen, Chen-xi Li, Yong-liang Zhang, Jun Lai, "Deep Reinforcement Learning with Sarsa and Q-Learning: A Hybrid Approach", IEICE Transactions on Information and Systems, Vol. E101 - D, No. 9, pp. 2316, September 2018.

[3] V. Mnih et al., "Playing Atari with Deep Reinforcement Learning," arXiv:1312.5602, December 2013.