동적 환경 내 자율 이동체의 실시간 경로 탐색 및 최적화 알고리즘

송승현, 조예령, 김중헌

고려대학교

ssh031008@korea.ac.kr, joyena0909@korea.ac.kr, joongheon@korea.ac.kr

Real-Time Path Finding and Optimization Algorithm for Autonomous Agents in Dynamic Environments

Seunghyeon Song, Yeryeong Cho, Joongheon Kim

Korea University

요 약

본 논문은 동적 환경에서 자율 이동체의 실시간 경로 탐색 및 최적화 문제를 해결하기 위해 은닉층을 병렬화한 deep q-learning (DQN) 알고리즘을 제안한다. 기존 DQN 알고리즘에서 은닉층을 병렬로 연결하여 더 풍부한 표현력과 학습 성능을 제공하며, 이를 통해 복잡한 환경에서도 효율적인 경로 탐색이 가능함을 실험적으로 입증하였다. 제안된 알고리즘은 기존 actor-critic 알고리즘과 비교했을 때 보상과 성공률에서 눈에 띄게 우수한 성능을 보였다. 이러한 연구 결과는 자율 이동체가 동적 장애물 환경에서도 실시간으로 경로를 최적화할 수 있는 강화 학습 알고리즘의 가능성을 제시한다.

Ⅰ. 서 론

최근 자율 이동체는 물류, 운송, 군사, 탐색 등의 다양한 산업 분야에서 중요한 역할을 맡고 있다. 이러한 산업 분야에서 자율 이동체는 단순한 정적 환경이 아닌 상대적으로 복잡한 동적 환경에 놓이게 되는 경우가 많다. 동적 환경에 놓인 자율 이동체는 실시간으로 변화하는 환경에서 목적지까지 최적의 경로를 탐색하고 도달해야 하므로 끊임없이 외부의 변수를 고려해야 한다. 기존의 경로 탐색 알고리즘들은 정적 환경에서는 높은 성능을 보여주지만, 동적 환경에서는 급격히 성능이 저하될 수 있다. 이러한 한계를 극복하기 위한 해결책으로 기계학습(machine learning, ML) 기반 경로 탐색 알고리즘이 주목받고 있다. 특히 ML의 학습 방식 중 하나인 강화 학습(reinforcement learning, RL) 기반 알고리즘은 에이전트(agent)가 환경과 상호작용하여 각 행동에 따른 보상을 학습하고, 이를 바탕으로 점점 더 나은 결정을 내리도록 작동하는 알고리즘으로 동적 환경 내 자율 이동체의 경로 탐색에 유용하게 활용할 수 있다. 하지만 이와 같은 알고리즘을 실제 동적 환경에 활용할 자율 이동체에 적용하기 위해서는 모델의 최적화가 필수적으로 이루어져야 한다. 본 논문에서는 동적 환경 내에서 자율 이동체의 경로 탐색 성능을 향상 시키기 위해 최적화 된 RL 기반 알고리즘을 제시하고, 이를 실험적으로 검증한다.

Ⅱ. 강화 학습

강화 학습은 일종의 자기 주도 학습으로, 환경에서 행동으로의 매핑(mapping)을 학습하는 것이다. [그림 1]을 보면, 환경 상태의 입력인 s를 받아들이고, 내부 추론 메커니즘에 따라 이에 상응하는 행동 a를 출력한다. 이 행동의 결과로 보상(reward) r이 발생하고, 이를 에이전트에게 피드백으로 전달한다. 에이전트는 이 보상과 현재 상태를 바탕으로 피드백 값을 최대화하는 방향으로 다음 행동을 선택한다 [1]. 강화 학습은 일반적으로 환경과의 상호작용을 통해 학습하기 때문에 많은 데이터와 계산이 필요하고, 현실에서의 문제에 적용할 때 모델링이 어려운 경우도 많다. 최근에는 딥러닝(deep learning)을 결합한 심층 강화 학습(deep reinforcement learning)이 큰 주목을 받고 있다. 딥러닝은 입력층(input layer), 은닉층(hidden layer), 출력층(output layer)을 구성요소로 가지는 심층 신경망을 통해 입력 데이터에 대한 특성(feature)을 높은 수준으로 학습하도록 한다 [2]. 이러한 딥러닝이 접목된 심층 강화 학습은 복잡한 동적인 환경에 놓인 자율 이동체가 경로 탐색을 성공적으로 수행할 수 있는 데에 기여할 수 있다.

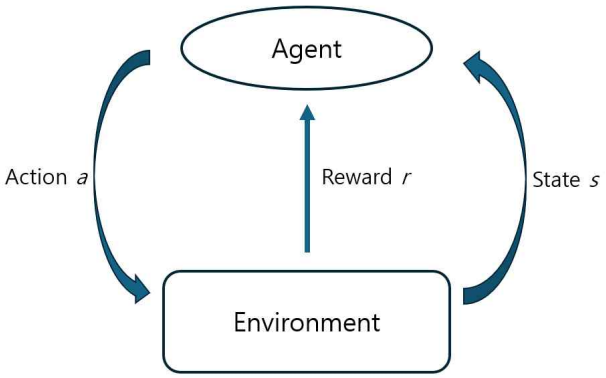

그림 1. 강화 학습 구조

Ⅲ. 동적 환경 내 경로 탐색 최적화 알고리즘

본 논문에서는 동적 환경 내 자율 이동체의 실시간 경로 탐색을 위한 알고리즘으로 기존 deep q-learning (DQN) 알고리즘에서 신경망의 은닉층을 병렬로 연결한 알고리즘을 제안한다. DQN은 q-learning을 심층 신경망과 결합한 알고리즘으로, 고차원 상태 공간에서 최적의 행동 정책을 학습할 수 있는 강화 학습 알고리즘이다. DQN은 q-값(state-action value function)을 근사하여 각 상태에서 최적의 행동을 선택하는데, 이때 경험 재생(experience replay)과 타겟 네트워크(target network)를 사용해 학습의 안정성을 높인다. 경험 재생은 에이전트가 환경과 상호작용하며 얻은 경험(상태, 행동, 보상, 다음 상태)을 메모리에 저장하고, 학습 시 무작위로 샘플링하여 사용함으로써 데이터 간의 상관성을 줄인다. 이를 통해 학습

이 더 안정적이고 효율적으로 이루어진다. 타겟 네트워크는 DQN에서 학습의 불안정성을 줄이기 위해 도입된 기법으로, q-값을 업데이트할 때 사용되는 네트워크가 일정 주기마다 업데이트되도록 설정하여 과도하게 변동하지 않도록 방지한다. 이를 통해 학습 과정의 안정성이 향상된다 [3]. 본 논문에서 제안하는 은닉층을 병렬화한 DQN 알고리즘은 병렬 은닉층이 동일한 입력을 여러 경로로 처리한 후 병합하는 방식으로, 각 경로가 서로 다른 특징을 독립적으로 학습할 수 있도록 설계되었다. 이를 통해 다양한 특징을 학습하고 은닉층을 순차적으로 연결 했을 때에 비해 학습 속도를 향상 시켜 네트워크의 학습 성능을 극대화하는 효과를 기대할 수 있다 [4]. 실험 환경은 자율 이동체가 동적 환경에서 실시간으로 경로를 탐색하고 최적화하는 문제를 설계하기 위해 OpenAI Gym에서 제공하는 FrozenLake 환경을 기반으로 하되, 장애물이 동적으로 움직이도록 커스터마이징하였다. 그리드(grid) 크기는 4x4로 설정하였고, 이동체는 시작점(0, 0)에서 목표점(3, 3)까지 경로를 탐색한다. 장애물은 4개의 장애물을 환경 내에 무작위로 배치하였으며, 각 장애물은 매 스텝마다 임의의 방향으로 움직인다. 에이전트가 목표 지점에 도달하면 +10의 보상을 부여하고, 장애물에 부딪히면 -10의 보상이 주어지며 에피소드는 종료된다. 또한 각 스텝마다 에이전트가 -1의 기본 보상을 받도록 하여 가능한 빠르게 목표 지점에 도달하도록 유도하였다. 은닉층을 병렬화한 DQN은 학습률(learning rate)을 0.001, 할인율(gamma)을 0.99, 탐험 감소율(epsilon decay)을 0.995, 최소 탐험률을 0.01, 배치 크기(batch size)를 64, 경험 재생 메모리 크기를 10,000으로 구성한다. 성능 비교를 위해 활용한 알고리즘은 actor-critic (AC) 알고리즘이다. 할인율은 0.99로 설정하였으며 actor 네트워크는 tanh 활성화 함수가 적용된 두 개의 은닉층을 포함하여 학습률을 0.001로 설정하였고, critic 네트워크는 rectified linear unit (ReLU) 활성화 함수가 적용된 두 개의 은닉층을 포함하여 학습률을 0.01로 설정하였다. 두 알고리즘 모두 2000 에피소드 동안 학습을 진행하였고 매 100 에피소드마다 보상과 성공률을 평가하였다.

Ⅳ. 실험 결과 도출 및 분석

동일한 동적 환경 맞춤형 FrozenLake 환경에서 은닉층을 병렬화한 DQN 알고리즘과 AC 알고리즘을 이용하여 경로 탐색을 진행한 결과, [그림 2]에서 결과 보상을, [그림 3]에서 성공률을 비교 확인할 수 있다.

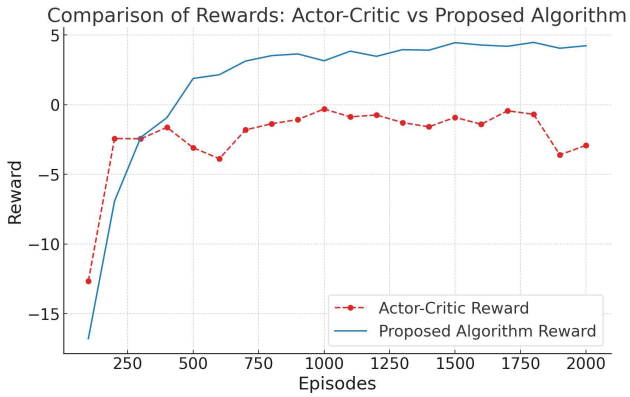

그림 2. actor-critic 알고리즘과 제안된 알고리즘의 보상 비교

또한 [그림 2]는 에피소드가 진행 될수록 제안된 알고리즘이 AC 알고리즘보다 더 높은 보상을 얻고 있음을 보이고 있다. [그림 3]에서는 에피소드가 진행될 수록 두 알고리즘 간의 성공률 차이가 점점 커지고 있음을 확인할 수 있다. 2000번째 에피소드까지의 성공률은 제안된 알고리즘이 0.93, AC 알고리즘이 0.66으로 제안된 알고리즘이 성공률 기준 약 29.17% 높은

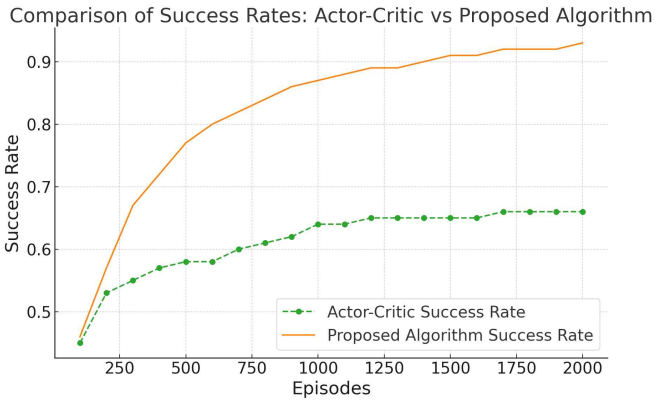

그림 3. actor-critic 알고리즘과 제안된 알고리즘의 성공률 비교

성능을 보여준다. 이를 통해 은닉층을 병렬화한 DQN 알고리즘이 동적 환경 내 경로 탐색에 최적화된 알고리즘임을 검증할 수 있다.

Ⅴ. 결 론

본 논문에서는 동적 환경 내 자율 이동체의 경로 탐색 성능을 향상시키기 위한 최적화된 강화 학습 알고리즘을 제안하였다. 제안된 알고리즘은 기존 DQN의 신경망 구조에서 은닉층을 병렬로 연결하여 더 다양한 특징을 학습하며 학습 성능의 향상을 보여준다. 실험 결과, 제안된 알고리즘이 다른 강화 학습 알고리즘인 AC 알고리즘보다 복잡한 동적 환경에서 높은 성공률과 보상을 기록하는 것을 확인하였다. 향후 연구에서는 제안된 알고리즘을 더 복잡하고 다양한 환경에 적용하여 범용성을 평가하고, 실시간으로 변화하는 환경에서 더욱 유연하게 대응할 수 있는 방법을 연구할 필요가 있다.

참 고 문 헌

[1] W. Linglin, L. Yongxin and Z. Xiaoke, “Design of reinforce learning control algorithm and verified in inverted pendulum,” in *Proc. 34th Chinese Control Conference (CCC)*, Hangzhou, China, Jul. 2015, pp. 3164-3168

[2] Shiri, F., Perumal, T., Mustapha, N., & Mohamed, R. “A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU,” in *CoRR*, abs/2305.17473, Jun. 2023.

[3] J. Fan, Z. Wang, Y. Xie, and Z. Yang, “A theoretical analysis of deep Q-learning,” in *Proc. 2nd Conf. Learn. Dyn. Control*, Feb. 2020. pp. 486 - 489.

[4] Nair, A., Srinivasan, P., Blackwell, S., Alcicek, C., Fearon, R., Maria, A.D., Panneershelvam, V., Suleyman, M., Beattie, C., Petersen, S., Legg, S., Mnih, V., Kavukcuoglu, K., & Silver, D. “Massively Parallel Methods for Deep Reinforcement Learning,” in *CoRR*, abs/1507.04296, Jul. 2015.

ACKNOWLEDGMENT

본 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(RS-2024-00439803, SW컴퓨팅 산업원천기술개발사업 (SW스타랩)). 본 논문의 교신 저자는 김중헌임.