심층 강화 학습을 통한 협업 로봇팔 제어에 관한 실험적 연구

장인국, 노삼열, 이동훈, 김성현
한국전자통신연구원

{ingook, samuel, donghun, kim-sh}@etri.re.kr

An Experimental Study of Multi-Agent Deep Reinforcement Learning for Collaborative Robot Arm Tasks

Ingook Jang, Samyeol Noh, Donghun Lee, Seonghyun Kim
Electronics and Telecommunications Research Institute

요 약

심층 강화 학습으로 로봇 제어 문제를 해결하는 것은 여전히 어려운 문제이다. 본 논문에서는 다중 에이전트 강화 학습을 다중 로봇 협업에 적용하는 실험적 연구를 다룬다. 우리는 두 로봇 팔이 공동의 목표를 위해 협업해야 해결할 수 있는 협업 로봇 제어 태스크를 설계하였다. 실험 결과는 최근 제안된 방법이 누적 보상과 성공률 측면에서 좋은 성과를 달성한다는 것을 보였고, 단순한 형태의 다중 로봇 협업 태스크에서 효율적으로 동작함을 보였다.

Ⅰ. 서 론

다중 에이전트 협업(multi-agent collaboration)은 기계 학습 커뮤니티에서도 여전히 어려운 문제 중 하나이다. 다중 에이전트 심층 강화 학습 (multi-agent deep reinforcement learning)은 일반적으로 1) 에이전트가 환경의 전체 상태 또는 다른 에이전트의 행동을 알지 못한다는 것을 의미하는 부분적 관찰 가능성 (partial observability), 2) 여러 에이전트가 동시에 정책을 학습하는 비정상성 (non-stationarity) 으로 인한 성능저하가 뚜렷하다 [1][2]. 예를 들어, 제한된 영역(부분 관찰 가능성)과 인간-로봇 협업(비정상성)을 가진 자율 주행 응용에서 인간 운전자의 차량 주행 예측 및 대응 문제는 다중 에이전트 강화 학습에서도 여전히 해결하기 어려운 문제이다.

본 논문에서는 로봇 조작 문제에서 다중 에이전트 강화 학습을 기반으로 한 다중 로봇 협업을 다룬다. 우리는 두 로봇을 포함한 협업 태스크를 설계 및 구현하고, 설계된 문제를 해결하기 위해 최신 알고리즘을 적용하여 주어진 환경에서 알고리즘의 효용성을 논의한다.

Ⅱ. 본론

우리는 로봇 시뮬레이터 중 하나인 CoppeliaSim [3]을 통해 협업 태스크 push_long_bar 를 설계하였다. 그림 1 과 같이, 그리퍼(gripper)가 구비된 2 개의 유니버설 로봇(UR3) [4]이 테이블 위에 나란히 배치되고, 그 앞에 임의의 지점에 수평 각도로 긴 막대가 배치된다. 긴 막대의 수평 축 각도를 유지하기 위해 두 로봇 팔이 동시에 일정한 속도로 막대를 밀어야 한다. 막대가 앞에 위치한 두 개의 목표 지점(그림 1 의 보라색과 노란색 구)을 동시에 터치하면 태스크가 성공한다. 수평 각도가 설정된 임계 값 보다 크거나 긴 막대가 하나의 목표 지점에만 도달하면 태스크는 실패로 간주한다.

각 로봇 i는 협업을 위한 보상 함수(reward function)는 다음과 같이 설계한다.

$$r^i(s,a) \coloneqq -\sum_{k=1}^{K} d(o_k, g_k) - Coll^i(s,a)$$

여기서 K 는 목표점 g_k 로부터 물체까지의 거리를 계산하기 위해 물체에 포함된 보이지 않는 점 o_k 의 수이다. 우리는 설계된 push_long_bar 태스크에서 $K = 2$ (좌우)를 사용한다. 다중 로봇 협업 태스크에서는 환경으로부터 각 로봇에게 동일한 보상을 제공한다. 만약 각 로봇 팔의 그리퍼가 테이블과 충돌하면, 그들의 보상은 충돌 횟수에 따라 패널티 $(Coll^i(s,a))$를 받고 총 보상의 합에 누적된다.

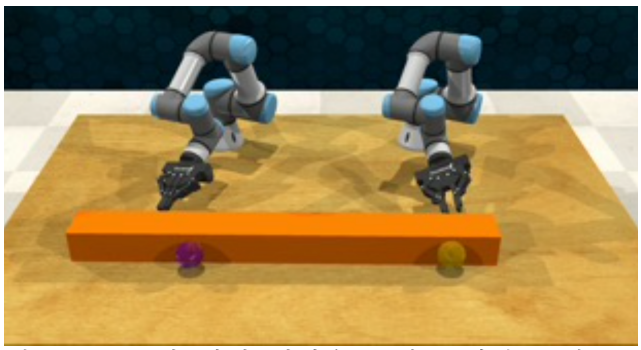

그림 1. 두 로봇 팔이 협업을 통해 주어진 물체를 목표 지점으로 이동시키는 push_long_bar 태스크

Ⅲ. 실험

두 로봇의 행동 정책(policy)을 학습하기 위해 [5]에서 제안된 MADDPG 을 사용하였다. MADDPG 는 actor-critic 학습 기법을 통해 학습 과정에서 다른 에이전트의 행동 정책을 고려하고 다중 에이전트 협동(multi-agent coordination)이 필요한 정책을 성공적으로 학습시킬 수 있다. 이 방법은 정책 앙상블(ensemble)을 활용하고 각 에이전트에 대해 분산 actor (distributed actor) 중앙집중식 critic (centralized critic) 접근 방식을 채택하여 다중 에이전트 행동 정책을 더욱 효율적으로 학습시킨다.

우리는 actor-critic 네트워크에 64 크기, 3 층(layer)으로 구성된 다층 퍼셉트론(multi-layer perceptron)을 사용하였다. 최대 에피소드 길이(maximum episode timesteps)를 50 으로 설정하고, 총 30,000 번의 에피소드동안 학습시켰다. 학습률(learning rate)은 1e-3 를 사용하고, 재생 버퍼의 크기는 50000, 재생 버퍼에서 추출(sampling)되는 미니 배치(mini-batch)의 크기는 1024 로 설정하였다.

그림 2 는 두 로봇 팔의 협업 태스크의 학습 성능 결과를 보여준다. 평균 누적 보상은 약 10000 에피소드 포인트 이후에 안정적으로 수렴됨을 알 수 있다. 두 로봇 팔을 사용하는 협업 태스크임에도 불구하고 매우 안정적인 평균 보상을 획득하며 태스크를 성공적으로 해결하였다. 표 1 은 학습 에피소드 수에 따른 협업 태스크의 평균 성공률을 보여준다. 역시 약 10000 에피소드 포인트 이후, 두 로봇 팔이 성공적으로 협업하여 긴 막대를 목표 지점까지 밀어낼 수 있음을 보여주었다.

Ⅳ. 결론

우리는 로봇 시뮬레이터를 기반으로 공통의 목표를 위해 다중 로봇이 협업하는 태스크를 설계하고 구현하였다. 다중 로봇 협업 태스크에 최신 강화 학습 기술을 적용하여 설계된 태스크에서 학습 결과를 도출하였다. 실험 결과, 누적 보상과 태스크 성공률 측면에서 좋은 성능을 보여준 것으로 나타났다.

ACKNOWLEDGMENT

본 연구 논문은 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음. [22ZR1100, 자율적으로 연결· 제어· 진화하는 초연결 지능화 기술 연구]

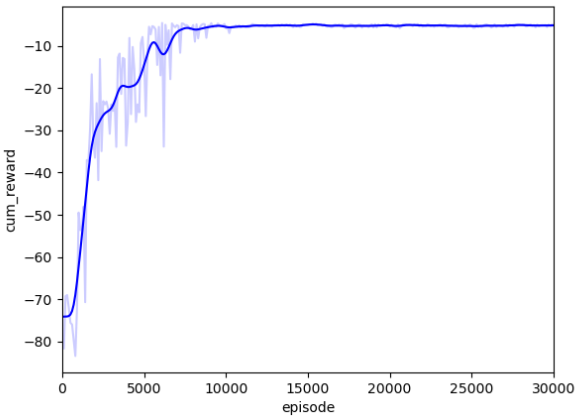

그림 2. 평균 누적 보상

표 1. 학습 에피소드 별 성공률

에피소드	0	1000	2000	5000	10000	30000
성공률	0	0.13	0.26	0.67	1.0	1.0

참 고 문 헌

[1] Ryan Lowe, et. al., “Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments”, In: Advances in Neural Information Processing Systems, pp. 6382-6393.

[2] Hyunseok Kim, et. al., "Avoiding collaborative paradox in multi-agent reinforcement learning", ETRI Journal 43(6), pp. 1004-1012, 2021.

[3] Vecerik, Mel, et. al., "Leveraging demonstrations for deep reinforcement learning on robotics problems with sparse rewards", arXiv preprint arXiv:1707.08817 (2017).

[4] Nair, Ashvin, et. al., "Overcoming exploration in reinforcement learning with demonstrations", 2018 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2018.

[5] Christianos, Filippos, et. al., "Shared experience actor-critic for multi-agent reinforcement learning", arXiv preprint arXiv:2006.07169 (2020).