

물리적 정보 기반 강화학습을 이용한 로봇 팔 제어 연구

김성현, 장인국, 김현석, 박찬원, 박준희
한국전자통신연구원

{kim-sh, ingook, hertzkim, cwp, juni}@etri.re.kr

A Study on the Robot Arm Control with Reinforcement Learning based on Physical Information

Seonghyun Kim, Ingook Jang, Hyunseok Kim, Chan-Won Park, Jun Hee Park

Electronics and Telecommunications Research Institute

요 약

본 논문은 물리적 정보에 기반한 강화학습을 이용한 로봇 팔 제어 연구를 제시한다. 로봇 팔 제어 문제는 주어진 환경에서 특정 작업을 달성하는 것을 목표로 정의된다. 예를 들어, 로봇 팔을 이용하여 환경 내 놓여진 객체를 집거나, 밀거나 하는 형태로 객체를 특정 위치로 옮기는 작업을 정의할 수 있다. 로봇 팔 제어 문제는 관측과 행동을 통해서 상호작용이 이루어지는데, 이 때 관측 정보와 행동 정보가 연속된 값을 갖기 때문에 강화학습을 통해 로봇 팔 제어 문제를 해결하는 것은 매우 도전적인 문제이다. 본 논문에서는 로봇 환경을 구성하는 객체들에 대한 물리적 정보를 활용하여 효과적으로 로봇을 제어하는 방식을 제안한다.

I. 서 론

본 논문에서는 물리적 정보에 기반한 강화학습을 이용한 로봇 팔 제어를 연구한다. 현재까지 강화학습은 심층학습 네트워크와 결합하여 불연속 관측 정보 및 행동 정보를 갖는 게임 환경에서 사람 수준의 성능을 달성하는 놀라운 성과를 이루었다 [1]. 이와 같은 심층학습과 결합한 강화학습, 즉 심층강화학습의 적용 분야는 게임 문제를 넘어서 실제 세계 문제를 풀고자 하는 분야로 확장되고 있다. 실제 세계의 문제 중 하나인 로봇 팔 제어 문제는 제조 및 물류 관리 등 다양한 분야에서 활용 가능한 문제이다. 로봇 팔 제어 문제는 기존 게임과 달리 연속된 관측 정보와 행동 정보를 다루기 때문에 난이도가 높은 문제에 속하고, 이를 해결하기 위한 많은 연구들이 진행되고 있다 [2]~[5]. 본 논문에서는 로봇 환경을 구성하는 객체들에 대한 물리적 정보를 활용하여 효과적으로 로봇을 제어하는 방식을 제안한다.

II. 본론

그림 1은 본 연구에서 고려하는 로봇 팔 작업 환경을 나타낸다. 그림 1과 같이 로봇 팔 작업 환경은 6축의 로봇 팔, 목적 객체(큐브), 센서 객체(원형) 등으로 구성되어 있다. 해당 로봇 팔 작업 환경에서 풀고자 하는 작업은 로봇 팔이 목적 객체를 들어서 센서 객체 위치로 옮기는 것으로 정의한다.

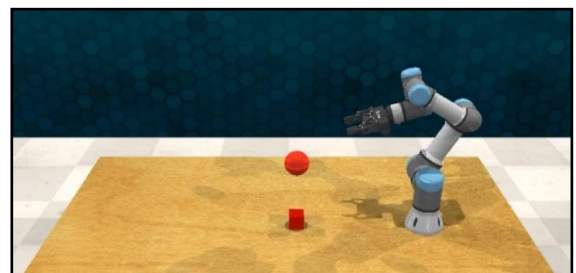


그림 1. 로봇 팔 작업 환경

1. 마르코프 결정 프로세스

로봇 제어 문제는 마르코프 결정 프로세스로 정의된다. 마르코프 결정 프로세스는 (상태, 행동, 전이확률, 보상)의 튜플 형태로 표현된다. 일반적으로 강화학습의 목적함수는 다음과 같이 감쇄보상의 합계에 대한 기대값으로 정의된다.

$$E \left[\sum_{t=0}^T \gamma \cdot r(s_t, a_t) \right]$$

여기서, γ 는 감쇄 정도, T 는 전체 작업 시간, r 은 보상, s_t 는 상태, a_t 는 행동을 나타낸다. 상태는 로봇의 관절들의 회전각위치 및 끝단의 그리퍼에 대한 3차원 위치정보, 목적 객체 및 센서 객체의 3차원 위치정보를 포함한다. 행동은 관절들의 회전각속도와 그리퍼의 열림-닫힘의 이진신호를 포함한다. 보상은 그리퍼와 목적 객체 사이의 거리, 그리퍼가 목적 객체를 잡았는지 여부를 판단하는 이진값, 목적 객체와 센서 객체 사이의

거리, 그리고 목적 객체가 센서 객체에 위치하였을 때 추가적인 보상 등의 합으로 정의된다.

2. 확률적 정책 학습

로봇 제어를 위한 정책은 $\pi(a_t|s_t)$ 로 표현되고, 이는 상태가 주어졌을 때 행동의 확률분포로 정의된다. 로봇 제어와 같은 연속된 관측 및 행동에 대한 학습은 수많은 시행착오를 통한 경험으로부터 수행된다. 즉, 다양한 경험을 획득하는 탐색의 과정이 효과적으로 이루어져야 한다. 목적함수를 최적화하는 학습단계에서 탐색을 보다 효과적으로 하기 위해, 정보이론의 엔트로피 개념을 목적함수와 결합할 수 있다 [6]. 즉, 목적함수에 추가적으로 정책에 대한 엔트로피 $H(\pi(a_t|s_t))$ 를 더하여, 행동의 확률분포를 넓게 가져가도록 최적화를 수행한다.

3. 물리적 정보 기반 확률적 정책 학습

상태 정보는 로봇 팔, 목적 객체, 센서 객체에 대한 각각의 고유 정보만을 포함하고 있다. 작업에서 요구하는 조건을 효율적으로 달성하기 위해서는 보다 직관적인 형태의 정보를 활용할 수 있다. 즉, 개별 객체에 대한 정보뿐만 아니라, 객체 사이의 관계를 정의하는 물리적 정보를 도출 활용함으로써, 보다 효율적으로 정책을 학습할 수 있다. 이와 같이 물리적 정보를 활용한 정책은 $\pi(a_t|s_t, z_t)$ 으로 표현되고, 여기서 z_t 는 물리적 정보를 나타낸다. 물리적 정보는 로봇 팔 작업 환경 내 객체 사이의 거리, 각도 등으로 정의할 수 있다.

물리적 환경을 활용하는 방식으로는 모델 기반 강화학습이 있다. 모델 기반 강화학습과 제안 방식을 비교하였을 때, 모델 기반 강화학습은 행동에 따른 상태전이 확률 분포에 대한 학습이 요구되는 반면, 제안하는 방식은 사람이 설계 가능한 물리적 정보를 직접 활용하기 때문에 학습에 대한 비용이 적고, 직접적인 해석이 가능하다는 장점이 있다.

III. 실험 결과

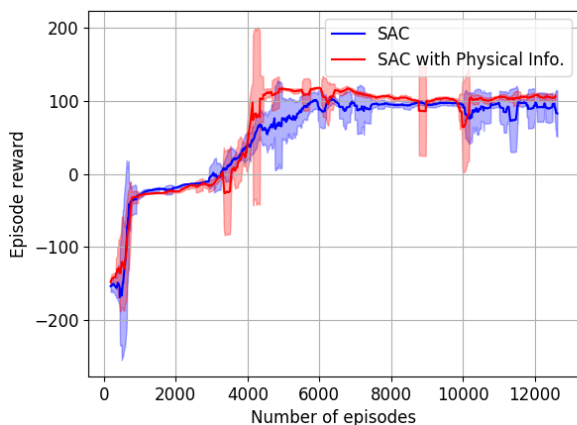


그림 2. 에피소드 보상 비교

제안하는 물리적 정보 기반 확률적 정책 학습 기법은 기존 확률적 정책 학습 중 하나인 Soft Actor-Critic(SAC)을 기반으로 물리적 정보를 결합하여 구현하였다. 그림 2는 학습 진행에 따른 에피소드 보상에 대한 비교를 나타낸다. 그림 2에서 확인할 수

있듯이, 물리적 정보를 결합함으로써 기존 SAC 기법보다 더 빠르게 학습이 수렴하는 것을 알 수 있다. 즉, 개별 객체에 대한 정보만 이용하는 것보다, 객체 사이의 관계를 입력으로 추가하였을 때 더 좋은 성능을 도출할 수 있다.

IV. 결론

본 논문에서는 로봇 환경을 구성하는 객체들에 대한 물리적 정보를 활용하여 효과적으로 로봇을 제어하는 방식을 제안한다. 기존 연구에서는 개별 객체에 대한 정보만을 활용하는 반면에, 본 논문에서는 객체 사이의 관계를 정의하는 물리적 정보를 함께 활용하는 방법을 제안하였다. 실험결과를 통해, 개별 객체에 대한 정보뿐만 아니라, 객체 사이의 관계를 정의하는 물리적 정보를 함께 활용함으로써 보다 효율적으로 확률적 정책을 학습할 수 있음을 확인할 수 있다.

ACKNOWLEDGMENT

본 연구 논문은 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음. [21ZR1100, 자율적으로 연결·제어·진화하는 초연결 지능화 기술 연구]

참 고 문 헌

- [1] P. Henderson, R. Islam, P. Bachman, J. Pineau, D. Precup, and D. Meger. Mastering the game of go with deep neural networks and tree search. *Nature*, 529:484–489, Jan. 2016.
- [2] Y.-J. Han, I.-S. Kim and Y.-D. Hong, “Optimization-based humanoid robot navigation using monocular camera within indoor environment,” *ETRI Journal*, vol. 40, no. 4, pp. 446–457, Aug., 2018.
- [3] M. Zhang, J. Chen, X. Wei, and D. Zhang, “Work chain-based inverse kinematics of robot to imitate human motion with Kinect,” *ETRI Journal*, vol. 40, no. 4, pp. 511–521, Aug., 2018.
- [4] M. Ilbeygi and M. R. Kangavari, “Comprehensive architecture for intelligent adaptive interface in the field of single-human multiple-robot interaction,” *ETRI Journal*, vol. 40, no. 4, pp. 483–498, Aug., 2018.
- [5] L. Fan, Y. Zhu, J. Zhu, Z. Liu, O. Zeng, A. Gupta, J. Creus-Costa, S. Savarese, and L. Fei-Fei. Surreal: Open-source reinforcement learning framework and robot manipulation benchmark. *Proceedings of The 2nd Conference on Robot Learning*, 87:767–782, Oct. 2018.
- [6] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. *Proceedings of the 35th International Conference on Machine Learning*, 80:1861–1870, Jul. 2018.