

강화학습을 적용한 그네 타기 로봇 제어

최호승¹, 박희재^{1*}

Control of swing robot using reinforcement learning

H. S. Choi¹, H. J. Park^{1*}서울과학기술대학교 기계시스템디자인공학과¹

Key Words : Reinforcement learning, Actor-critic

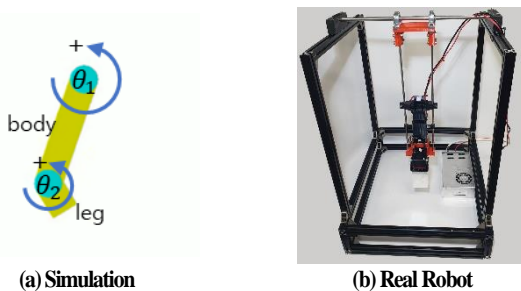
1. 서론

현재 전성기를 맞고 있는 강화학습은 고난이도의 제어가 필요한 기계에 적용하는 다양한 연구들이 진행되고 있으나 주로 시뮬레이션 연구에 국한된 경우가 많았다.⁽¹⁾

본 연구의 목적은 그네 타기 로봇에 강화학습을 적용하고 실험을 통하여 그 효용성을 검증하는 것이다. 먼저 단순화된 휴머노이드 형태로 2자유도를 갖는 로봇이 탑승한 그네에 대한 동역학적 모델을 만들고, 강화 학습의 일종인 Actor-Critic 알고리즘을 통해 로봇이 그네 타기 운동을 하는 방법을 스스로 터득할 수 있음을 시뮬레이션을 통해 확인하였다. 다음으로 실제 그네 타는 로봇에 강화학습을 적용하여 그 유효성을 확인하였다.

2. 학습 환경

본 연구에서는 1027개의 파라미터를 가지고, 1개의 공통 신경망을 가진 후 각각 정책망과 가치망으로 분기되는 구조의 신경망을 사용하였다. 강화학습 환경의 Observation Space는 $\theta_1, \theta_2, \dot{\theta}_1, \dot{\theta}_2$ 로 네 가지의 파라미터를 설정하였고, Action Space는 a_{-1} (다리를 굽히는 동작, CCW), a_1 (다리를 펴는 동작, CW) 두 가지로 정의하였다.



(a) Simulation

(b) Real Robot

Fig 1. Design of Experience environment

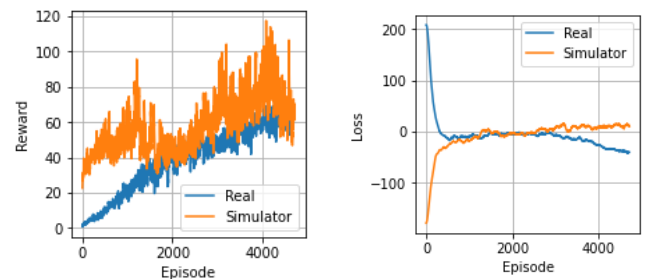
3. 시뮬레이션 및 실제 실험

Fig.1 (a)에서 사진의 오른쪽은 로봇의 정면에 해당한다. Reward의 정의는 기본적으로 그네가 수평에 가까워질수록 큰 보상을 받도록 구성하고 다음 수식과 같이 그네가 움직이는 방향과 같은 방향으로 로봇의 다리를 움직일 때는 양의 보상을, 반대 방향으로 움직일 때는 음의 보상을 받도록 구성하여 reward와 penalty를 줄 수 있도록 하였다.

$$reward = \begin{cases} |\sin(\theta_1)|, & sgn(\dot{\theta}_1) = sgn(action) \\ -|\sin(\theta_1)|, & sgn(\dot{\theta}_1) \neq sgn(action) \end{cases}$$

학습은 한 에피소드당 1000 스텝을 진행하는데, 다음과 같은 종료 조건을 만족할 때까지 학습을 계속한다. 이상적으로 움직이는 그네의 각도는 정현파가 되는 것이 가장 자연스럽다는 사실에 착안하여, FFT에서 특정 주파수 성분이 지배적으로 검출되는 것을 종료 조건으로 정하였다. 그네의 최대 각도가 20도 이상이며 FFT에서 최대 크기를 갖는 주파수가 크기의 평균보다 100배 이상으로 검출되는 것을 종료 조건으로 잡았다.

Fig. 2를 보면 직접 구현한 시뮬레이션과 실제 로봇 모두 꾸준히 reward가 증가하는 양상은 유사하지만, 최대 리워드가 약 2배까지 차이가 남을 볼 수 있다. 이는 동역학 모델과 실제 기구와 오차에 기인하는 것으로 판단된다. 또한 초기의 random한 행동에 따라서 달라지는 reward의 시작 지점에 따라서 학습 속도가 크게 달라짐을 알 수 있었다.



(a) Reward comparison of simulator and real robot

(b) Loss comparison of simulator and real robot

Fig 2. Comparison of Simulator and Real Robot

실제로 학습이 끝난 모델을 이용하여 테스트를 해 보았을 때 일정한 주기로 다리를 움직여 시뮬레이션은 최대 30도, 실제 모델은 최대 20도 이상까지 swing motion을 구현할 수 있는 것을 확인하였다.

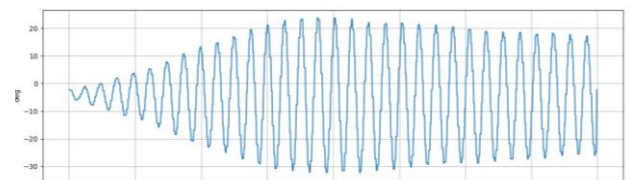


Fig 3. Angle change over time

4. 결론

본 연구의 의의는 실제 기구에 Actor-Critic 강화학습 알고리즘을 적용하여 반복 학습을 통한 고난이도의 제어기술을 기계학습으로 대체하는 것에 있다. 이를 위하여 그네 타기 로봇에 강화학습을 적용하고 실험을 통하여 그 효용성을 검증하였다. 향후에는 서서 타는 그네와 다자유도 로봇 제어 문제에 연구를 진행할 계획이다.

참고 문헌

- (1) H.S. Choi, D.W. Kim, H.H. Park, 2021, Simulation of swing robot using reinforcement learning, KSMTE, 192 1-2
- (2) Shalabh Bhatnagar, Richard Sutton, Mohammad Ghavamzadeh, Mark Lee, 2009, Natural Actor-Critic Algorithms, Automatica, 2471-2482