ROS 기반 Niryo One 로봇 팔 강화 학습

이창열, 박상준, 이성민, 임기태, 김태형
금오공과대학교
{lee_0996, 20150492, 20150883, dlarlxo23, taehyong}@kumoh.ac.kr

Reinforcement learning of a Niryo One robot arm based on ROS

Changyeol Lee, Sangjun Park, Seongmin Lee, Gitae Lim, Tae-Hyong Kim
Kumoh National Institute of Technology

요 약

본 논문은 교육용 모듈형 로봇 팔 Niryo One을 사용하여 ROS 환경에서 주어진 물건을 집어 올릴 수 있도록 강화학습을 수행하는 환경을 구현하였다. ROS 기반 강화학습 환경으로는 Gym-Gazebo2를 사용하였고, 강화학습 알고리즘으로는 로봇 학습에 유리하다고 알려진 가치기반 방식의 PPO(Proximal Policy Qptimization) 알고리즘을 사용하였다. 추가로 Niryo One 로봇 팔이 물체에 잘 접근할 수 있도록 적절한 매개변수와 보상 방법을 고안하였다. 실험을 통해 고안한 강화학습 방법이 임의의 물체에 보다 빠르게 접근함을 확인하였다.

Ⅰ. 서 론

강화학습은 의사결정 순서에 관계되는 기계 학습의 한 분야로 각 시간 단계에 대해 에이전트가 현재 환경을 고려하여 행동하고 그것에 대해 관찰과 보상을 통해 문제를 해결한다. 강화학습 알고리즘은 시행착오를 통해 에이전트의 총 보상을 최대화하려고 한다. 최근 심층 신경망 강화학습은 복잡한 행동 기술을 학습하고 고차원적 상태 공간에서 도전적인 과제를 해결하는 데 좋은 성공을 보였다. 강화학습은 게임, 지능형 로봇, 금융, 자율주행 분야 등에 폭넓게 적용되고 있다.

ROS(Robot Operating System)[1]는 로봇을 개발할 때 필요한 기능 구현, 메시지 전달, 패키지 관리, 개발 환경 등에 관한 라이브러리와 개발 환경을 제공하는 일종의 미들웨어 플랫폼이다. 특히 ROS는 실제 로봇 없이 다양한 로봇을 구동해 볼 수 있는 시뮬레이션 환경을 제공한다. Gazebo[2]는 ROS에서 지원하는 3차원 로봇 시뮬레이터로 복잡한 구조의 로봇을 정확하게 시뮬레이션 할 수 있다. 한편 OpenAI라는 비영리 단체는 강화학습 알고리즘을 개발하고 시험할 수 있는 환경을 제공하는 Gym[3]이라는 파이썬 패키지를 공개하였다. 최근 Gazebo 시뮬레이터와 Gym을 연결하여 강화학습으로 로봇을 훈련하고 시험할 수 있는 Gym-Gazebo2[4]가 발표되었다.

본 연구는 교육용으로 판매되는 6축 로봇 팔인 Niryo One[5]을 학습시켜 물건을 집어 올릴 수 있도록 하기위해 Gym-Gazebo2 기반으로 강화학습을 수행한 결과를 보여 준다. ROS 가상환경에 카메라를 추가하여 임의의 위치에 놓인 물체를 감지하고 강화학습을 통해 해당 위치에 접근함으로써 물체를 로봇 팔에 달린 그리퍼(gripper)로 집어 올릴 수 있게 한다.

Ⅱ. 본론

강화학습 알고리즘은 가치기반 학습과 정책기반 학습으로 나눌 수 있다. 가치 기반 학습은 특정 상태에서 획득한 보상이 무엇인지 정확하게 평가하기 위해 노력하며, 따라서 최종기대 보상을 최대화한다. 정책 기반 학습은 최종 보상을 극대화하기 위해 각 상태에서 어떤 조치를 해야 하는지 배우려한다. 로봇 환경은 상태와 행동 공간의 지속적인 변화를 가지는 시나리오에 의해 지배되기 때문에 가치 기반 학습이 적합하지 않다. 따라서 본 연구는 정책 기반 학습인 model-free PPO(Proximal Policy Optimization) 알고리즘[6]을 채택하였다. Model-free 접근 방법은 복잡한 도메인 지식을 통합하여 정책에 대입함으로써 학습 과정에서 더 적은 수의 매개변수를 사용하게 된다. PPO는 에이전트를 통해 데이터를 샘플링하고 SGA(Stochastic Gradient Ascent)를 사용해 surrogate 목적 함수를 반복적으로 최적화한다. PPO는 데이터 효율성과 강인성을 유지하며, 미니배치 갱신을 통해 정책변화가 크지 않기 때문에 모듈형 로봇시스템에 적합하다고 한다[7].

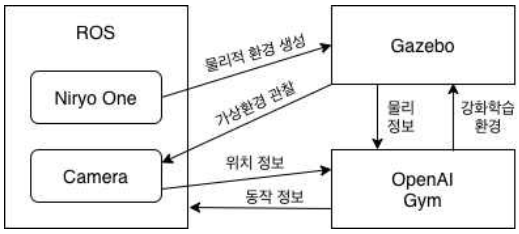

그림 1. Niryo One 강화학습 환경

본 연구의 강화학습 환경은 Gym-Gazebo2에서 제공하는 6축 로봇 팔 MARA[8] 예제를 기반으로 구축하였다. 강화학습 환경은 그림 1과 같이 세 부분으로 구성된다. ROS는 Niryo One의 상태 관찰, 조인트 값 명령, 세부 수준 제어를 수행하며 USB 카메라를 통해 대상 물건을 관찰한다. Gazebo 가상환경은 로봇의 물리적 환경을

구현하고 이를 통해 OpenAI Gym이 환경을 구축하는데 필요한 물리적 값들을 전달한다. OpenAI Gym은 Gazebo 강화학습 환경을 통해 학습을 수행하며 학습 결과를 ROS에 전달함으로써 Niryo One의 동작을 발생시킨다.

가상환경 내에는 테이블과 테이블의 수직선 위에 있는 카메라, 테이블 위의 Niryo One 로봇 팔이 존재한다. 테이블의 높이는 Z축으로 6.4이다. OpenAI Gym은 Gazebo에게 물리적 정보를 제공하고 목표 지점의 위치 및 로봇 암 끝단의 위치 지점을 생성한다. 그림 2는 본 Gazebo 가상환경을 보여주는데 파란색과 초록색 점으로 표시된 것이 각각 끝단 지점과 목표 지점이다.

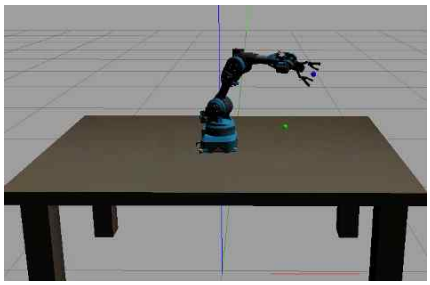

그림 2. 구현된 Gazebo 가상 환경

OpenAI Gym은 Gazebo 환경을 관찰하고 관찰된 정보를 통하여 강화학습을 수행, 그 결과를 보상과 행동으로 NiryoOne ROS에 전달한다. 또한, 에피소드를 진행할 때 마다 Gazebo 환경을 초기화하고 목표의 위치를 임의의 값으로 수정하는 과정을 수행한다. 가상환경에서 하나의 에피소드가 시작되면 목표가 생성되고 이를 기반으로 수행 영역 안의 값을 보상에 맞게 찾아 행동한다. 보상은 로봇 암의 끝단과 목표의 위치를 조인트 공간을 통해 비교한다. 에피소드의 종료조건은 1024 스텝을 최대치로 잡았다. 고정된 조인트를 제외한 모든 충돌에 대해서 반응하며, 충돌이 발생하는 경우 환경을 초기화하여 충돌을 회피하도록 학습한다. 특히, 테이블을 추가함으로써 로봇 팔과 바닥 간의 충돌여부를 확인하여 안정성 있는 학습을 제공한다. 상태와 보상 정보는 ROSLogInfo에 기록되며 이 기록을 통해 학습 결과를 확인하고, 종료된 학습을 추가로 학습할 수 있다.

실험은 PPO 알고리즘을 사용하여 1000회 에피소드를 반복하여 학습하였다. 보상 값은 MARA 예제에서 사용된 매개변수와 보상 방법을 사용한 것(A)과 끝단의 방향성 보상을 제거하고 거리 보상에서 0.05 이하일 경우 추가로 2의 보상을 주며, 0.01이하일 경우는 1을 주는 방식(B)의 두 가지를 사용하고 비교하였다. 보상 방식 B는 A에서 발생할 수 있는 과적합을 해결하기 위한 것이다. 그림 3과 4는 각각 두 보상 방식의 에피소드 진행에 따른 보상 값 추이와 최소 거리오차 추이를 보여준다(방식 A: 주황색, 방식 B: 청색).

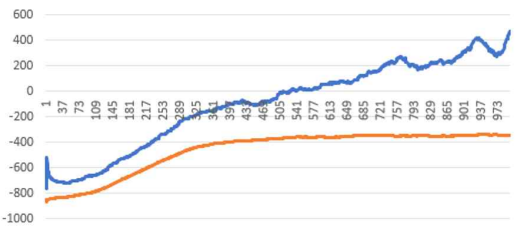

그림 3. 보상 방식에 따른 에피소드 별 보상 값 추이

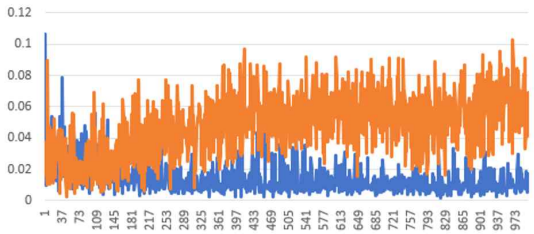

그림 4. 보상 방식에 따른 에피소드 별 최소 거리오차 추이

그림 3에서 방식 B가 에피소드 진행에 따라 보상 값이 꾸준히 증가하고, 그림 4에서 방식 B가 A보다 최소 거리오차가 최솟값으로 빠르게 수렴하는 것을 볼 수 있어, 학습 효과가 높음을 알 수 있다. 1700회 에피소드 학습을 통하여 임의의 목표에 대해 로봇 팔이 물체에 잘 접근함을 확인하였다.

Ⅲ. 결론

본 논문은 가치기반 강화학습 PPO를 통하여 모듈형 로봇 팔의 끝단을 임의의 목표에 대해 빠르게 이동시킬 수 있음을 보였다. 이를 위해 교육용 로봇 팔 Niryo One을 위한 ROS 가상 환경을 구성하고 Gym-Gazebo2 환경과 연결하여 강화학습 환경을 구축하였다. 또한 Niryo One 로봇 팔의 움직임과 특성을 고려하여 학습이 잘 이루어질 수 있는 적절한 매개변수와 보상 방법을 고안하였고, 실험 결과를 통해 성능 향상이 있음을 확인하였다. 하지만, 본 연구에서 다루지 않은 강화학습 매개변수들이 여전히 많이 존재하며 최적의 학습 효과를 얻기 위해서는 더 많은 연구가 필요할 것이다. 향후 본 연구 결과를 실제 Niryo One 로봇 팔에 적용하고 다양한 형태의 물건을 집어 올리는 연구를 수행할 예정이다.

참 고 문 헌

[1] ROS, “Robot Operating Systems”, see https://www.ros.org/

[2] Open Source Robitics Foundation, “Gazebo”, 2014, see http://gazebosim.org/

[3] OpenAI, “Gym”, see https://gym.openai.com/

[4] Acutronic Robotics, “gym-gazebo2”, see https://github.com/AcutronicRobotics/gym-gazebo2

[5] Niryo, “Niryo One”, see https://niryo.com/niryo-one/

[6] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, Oleg Klimov, “Proximal Policy Optimization Algorithms“, arXiv: 1707.06347

[7] Yue Leire Erro Nuin, NestorGonzalez Lopez, EliasBarba Moral, Lander Usategui San Juan, Alejandro Solano Rueda, V´ictorMayoral Vilches and Risto Kojcev “ROS2Learn: a reinforcement learning framework forROS 2” AcutronicRobotics, March 2019.

[8] Acutronic Robotics, “MARA”, see https://github.com/AcutronicRobotics/MARA