# 목적지로 자율 주행 가능한 강화 학습 로봇

임경욱*, 손지선*, 최현동*, 원일용*
*서울호서전문학교 사이버해킹보안과
imkyeonguk00@gmail.com, riseonaci@gmail.com, hyeon30448@gmail.com,
clccclcc@shoseo.ac.kr

# A self-driving Robot for target place using reinforcement learning

Kyeong-Uk Im*, Ji-Seon Son*, Hyeon-Dong Choi*, Ill-yong Weon*
*Dept of Cyber Security, Seoul Hoseo Technical College

**요 약**

가상 환경의 시뮬레이션을 이용해 지능형 로봇에 강화 학습 기법을 적용하는 접근법은 실제 세계의 로봇들의 학습에 유용하다. 우리는 이러한 방법을 적용해서 장애물을 회피하고, 로봇이 특정 목표물을 인식하면 목표물로 자율적으로 이동하는 알고리즘을 개발하였다. 제안된 방법의 유용성 검증은 구현과 실험으로 확인하였다.

## 1. 서론

인공지능 기술의 발전으로 다양한 지능형 로봇들이 개발되어 응용되고 있다. 이러한 지능형 로봇으로는 무인 잠수함, 물류 로봇, 드론 등이 대표적이다[1,2,3]. 이러한 자율 로봇은 다양하게 주어진 환경에서 데이터를 스스로 만들어 학습해야 하기 때문에 강화 학습 방법이 가장 효율적이라고 할 수 있다[4].

실 세계에서 로봇에 강화 학습을 적용하기 위해서는 에피소드를 반복해서 초기화해야 하는 어려운 문제가 존재하는데, 어떤 문제들은 현실적으로 에피소드 반복이 비용이나 위험성 등의 문제로 불가능한 경우도 많다.

이러한 환경에서 강화 학습을 적용하기 위해 실 환경과 유사한 가상의 시뮬레이션 환경을 만들어 제약이 적은 가상환경에서 에피소드를 무한 반복하는 방법을 사용하는 방법이 있다. 즉, 가상환경에서 학습된 모델을 실 환경에서 적용하는 방법이다[5].

본 논문은 지능형 로봇의 강화 학습 방법으로 가상환경에서 학습하고 이것을 이용하여 실 환경에 사용하는 접근법에 대한 것이다. 우리의 선행연구에서는 이런 방법의 가능성만을 확인하였다면[6], 본 논문에서는 로봇이 장애물 회피뿐만 아니라 특정 목표물을 인식하고 그 목표물로 자율적으로 움직이는 알고리즘을 학습한다. 제안한 시스템의 유용성은 실험으로 증명하였다.

본 논문의 구성은 다음과 같다. 2 장에서는 관련 연구를, 3 장에서는 제안하는 시스템의 개념 및 구성을 언급하였다. 4 장에서는 실험 및 결과를 서술하였으며, 5 장에서는 결론 및 향후 과제를 기술하였다.

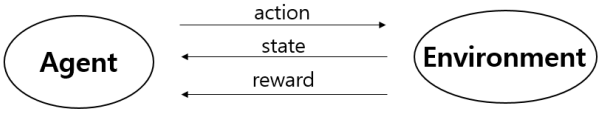
## 2. 관련 연구

### 2.1 Unity3D 강화 학습

Unity3D 에서는 시뮬레이션 환경을 개발하고 에이전트(Agent)를 학습시킬 수 있는 ML-Agents 를 오픈소스(open-source) 형태로 제공한다. ML-Agents 가 제공하는 학습 알고리즘에는 강화 학습, 모방 학습(Imitation Learning), 신경 진화 및 기타 기계학습 등이 있다.

강화 학습은 에이전트가 경험을 통해 학습 과정을 기반으로 시간적 변화에 따른 환경상태 정보를 반복적으로 학습하고, 보상 값을 통합하여 개선하는 기계학습 알고리즘 중 하나이다. 즉 에이전트의 연속된 행동에 따른 누적 보상을 평가하고 이를 학습하여 정책을 개선한다[7].

강화 학습에서 에이전트와 환경(environment)은 (그림 1)과 같이 action, state, reward 3 가지를 통해 상호작용하며, 에이전트가 주어진 환경에 대해 최대한 많은 보상을 받도록 학습하는 것이 중요하다.




(그림 1) 강화 학습에서 에이전트와 환경의 상호작용

2.2 포섭 구조

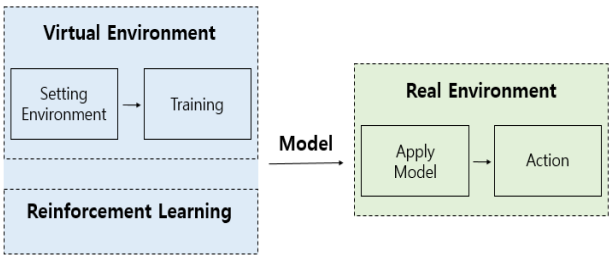
로봇의 다중 목표를 달성하기 위해 로드니 브룩스의 포섭 구조를 적용한다. 포섭 구조란 단순 행동들이 계층적으로 존재할 때 개별의 계층들이 유기적으로 상호작용하여 복잡한 목표를 달성하는 것을 말한다. 일반적으로 행동 수준에 따라 계층의 순서를 결정하며 계층이 높을수록 높은 차원의 목표를 수행한다. 또한 최상위 계층 목표에 대응하면서 하위 계층 목표를 달성하는 병렬화된 제어 시스템을 기반으로 한다.

계층 간에는 억압과 억제로 조절되는데, 상위 계층은 보다 섬세한 제어가 가능한 구조이며, 하위 계층은 기본적인 기능 모듈로 구성된다. 평소에는 하위 계층으로 동작하다가 특정 조건이 만족되면 상위 계층이 하위 계층의 모듈을 억압 또는 억제하게 된다 [8].

## 3. 로봇 학습 시스템

3.1 시스템 구성

로봇의 강화 학습을 위한 환경으로 실 환경과 유사한 가상 환경을 구축하여 반복 학습을 통해 학습 모델을 생성하게 된다. 가상환경에서 충분한 학습이 완료되면 해당 모델을 실제 환경의 로봇에 적용하여 테스트하게 된다.




(그림 2) 시스템 구성도

이러한 구조에서 가장 중요한 것은 가상 환경이 얼마나 실제 환경을 잘 반영하는가이다. 특히 가상 환경센서에서 관측되는 데이터의 종류와 크기가 실제 환경의 센서를 얼마나 잘 반영하는가는 전체적인 효율성을 결정하는 가장 중요한 요소이다. 가상 환경에서는 노이즈가 없지만, 실제 환경에서는 원하지 않는 노이즈가 존재하기 때문에 학습 데이터에 포함된 노이즈를 어떻게 제거시키는 가도 중요한 요인이 된다.

3.2 로봇의 기본적 운동

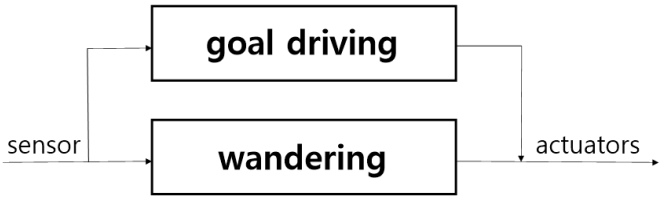
로봇은 자신이 가지고 있는 이미지 센서로부터 주기적으로 데이터를 입력받는다. 이렇게 입력받은 이미지에서 미리 지정된 객체가 인식되면 로봇은 인식된 객체의 방향으로 이동하게 된다. 반면 원하는 객체가 인식되지 않으면 장애물을 회피하며 무작위 방향으로 이동한다. 시스템은 로봇의 이동을 위해 아래 표와 같은 기본 운동 함수를 제공한다.

| 이름 | 기능 |
|---|---|
| stop | 움직임을 멈춘다 |
| go ahead | 앞쪽 방향으로 이동한다 |
| go back | 뒤쪽 방향으로 이동한다 |
| turn left | 왼쪽으로 일정 각도만큼 회전한다. |
| turn right | 오른쪽으로 일정 각도만큼 회전한다 |
| check detect | 영상에서 목표 객체가 있는지 확인한다 |

<표 1> 로봇의 기본 운동 함수

3.3 로봇 포섭 구조

제안하는 시스템에서 로봇의 운동은 2 개의 층으로 제어된다. Wandering 층은 목표물이 감지되지 않았을 때 임의의 방향으로 움직이며 장애물을 회피하는 기능을 수행한다. Goal Driving 층은 목표물이 감지되었을 때 목표물로 이동하는 기능을 수행한다.




(그림 3) 시스템 포섭 구조

## 4. 구현 및 실험

실제 환경에서 로봇은 기본적인 운동 기능과 외부 데이터를 인지하는 이미지 센서로 구성되어 있다. 가

상 환경에서는 이와 유사한 운동 기능과 이미지 센서를 구현하였다.

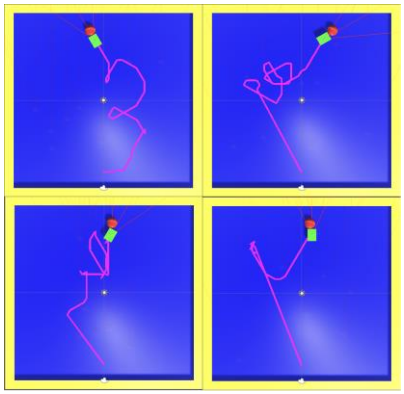
### 4.1 가상 환경 및 모델 생성

가상 환경은 Unity3D 를 사용하여 구현하였으며 강화 학습 모듈은 ML-Agents 를 사용하였다. 우리는 가상환경 두 개를 구축하여 wandering 동작을 수행하는 모델과 goal driving 동작을 수행하는 모델을 각각 하나씩 생성하였다. 로봇은 4 각 박스로 표현하였으며 카메라는 가상 카메라, 거리 측정 센서는 unity3D 의 ray-cast 를 이용하여 구현하였다.

#### 4.1.1 goal driving 모델 생성

goal driving 동작 학습을 위한 가상환경에서 목표물은 구 모양으로 표현하였다.

한 개의 에피소드는 목표물에 도달하면 긍정적 보상으로 끝나고, 일정 시간이 지나도 목표물에 도달하지 못하면 부정적 보상으로 끝나게 된다. 이렇게 학습된 모델을 이용하여 가상 환경에서 주행 시험을 하는 모습은 아래 (그림 4)와 같다.

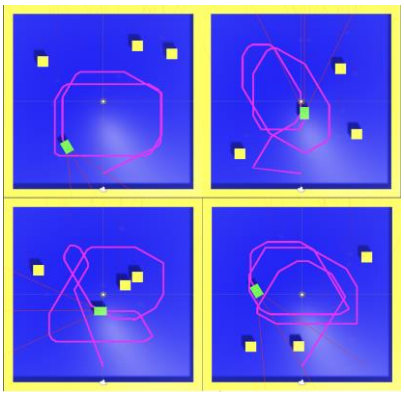


(그림 4) 가상환경에서의 goal driving 모델 주행 예

#### 4.1.2 wandering 모델 생성

wandering 동작 학습을 위한 가상환경에서 장애물은 로봇과 같이 4 각 박스로 표현하였다.

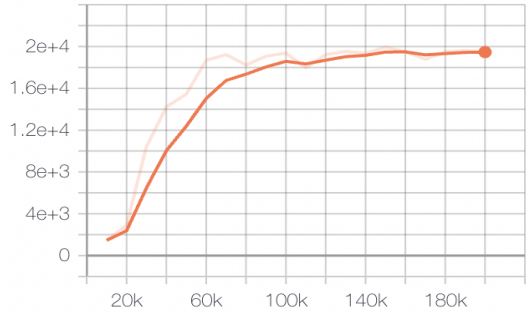
에이전트가 go action 을 취하면 긍정적인 보상을 주었으며 장애물에 부딪히거나 back action 을 취할 경우 부정적인 보상을 주었다. 또한 과학습(overfitting)을 방지하기 위해 에피소드가 시작할때마다 장애물과 에이전트의 위치를 무작위로 배치하였다. 이렇게 학습된 모델을 이용하여 가상 환경에서 주행 시험을 하는 모습은 아래 (그림 5)와 같다.



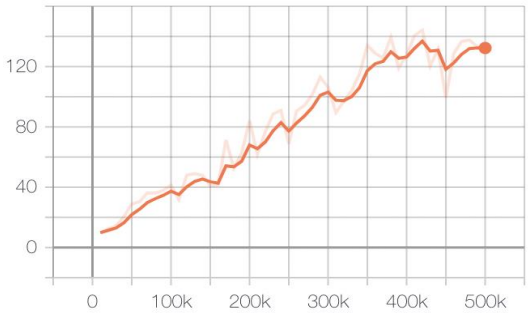
(그림 5) 가상환경에서의 wandering 모델 주행 예

#### 4.1.3 학습된 모델의 학습률

학습 알고리즘은 wandering 과 goal driving 모두 PPO(Proximal Policy Optimization)를 사용하였고, 학습률의 변화는 아래 그래프와 같이 충분한 반복을 통해 점차 안정되어 가는 양상을 보였다.




(그림 6) goal driving 모델 학습률 그래프

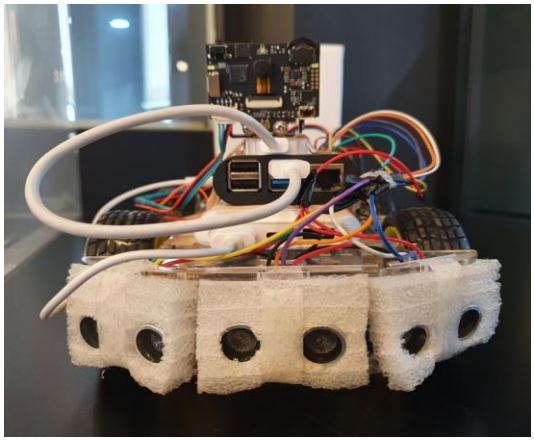



(그림 7) wandering 모델 학습률 그래프

### 4.2 실제 환경 로봇

실제 환경에서 사용하는 로봇의 구성은 아래 그림과 같다. 운동을 위한 모터 장치와 목표 이미지를 탐지하는 이미지 센서, 거리를 측정하는 3 개의 초음파 센서로 구성된다. 이미지 센서는 허스키렌즈를 사용

하였으며, 초음파 센서로는 HC-SR04 를 사용하였다. 전체적인 시스템은 라즈베리파이를 사용했으며, 신경망 모듈 연산을 위해 Coral TPU 를 사용하였다. 아래 (그림 8)은 실제 환경에서 동작하는 로봇의 사진이다.

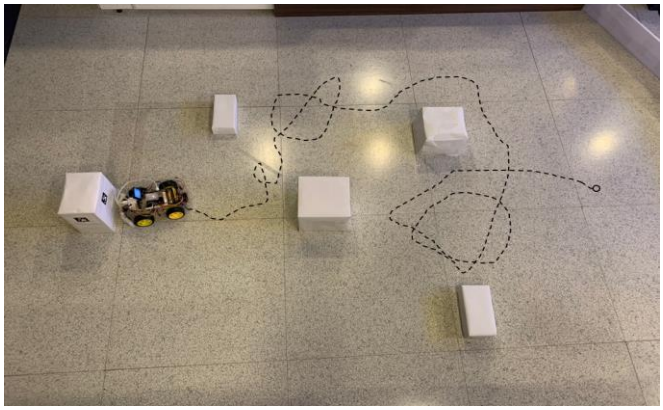


(그림 8) 실제 환경의 로봇

4.3 실제 환경 로봇 실험

가상 환경에서 만들어진 모델을 실제 환경의 로봇에 가져와 적용하기 위해서는, 가상 환경과 실제 환경의 차이 때문에 센서 값의 스케일링이 필요하다. 또한 PC 환경에서 생성된 모델을 라즈베리파이에서 사용하기 위해서는 tensorflow lite 모듈로의 변환이 필요하다. 본 실험에서는 tensorflow lite converter 를 사용하여 모델을 변환하였다.

아래 (그림 9)는 실험에서 로봇이 이동한 경로의 위치를 보여 준다.



(그림 9) 로봇 이동 경로 사진

실험 결과 가상환경에서의 모델 동작과 같이 실 세계의 로봇도 장애물 회피 동작과 목적지 인식 및 도달 동작을 성공적으로 수행하는 것을 확인할 수 있다.

## 5. 결론 및 향후 과제

지능형 로봇의 강화 학습은 다양한 어려움이 존재한다. 이러한 어려움을 극복하는 방법 중 우리는 가상 환경을 구축하고 제약이 없는 가상환경에서 학습하여 모델을 만들고 실 세계에서 이것을 사용하는 접근법을 사용하였다.

우리는 특정 목적지를 지정하고 로봇이 이 목적지로 이동하는 알고리즘을 자동으로 학습하는 시스템을 제안하고 유용성을 실험으로 검증하였다. 이러한 접근 방법에서 가장 중요한 것은 가상 환경이 얼마나 실제 환경을 잘 반영하는가이다. 실험을 통해 제안된 시스템의 유의미한 결과를 얻을 수 있었다.

향후 연구는 로봇이 동적 환경에서 다양한 목표를 수행하는 기능을 추가하는 것이 필요하다.

## 참고문헌

[1] Dong-gi Son and Dong Hun Kim "Autonomous Navigation of Unmanned Surface Vehicle Based on Fuzzy Control Considering COLREG" 한국지능시스템학회 Vol.31 No.1 pp.11-20, February, 2021

[2] Don-Hyuk Jeong, Jin-Il Park and Yong-Tae Kim “Study on Desin of Mobile Robot for Autonomous Freight Transportation”, Journal of Korean Institute of Intelligent Systems, Vol. 23, No. 3, June 2013, pp. 202-207

[3] 손승재 “Waypoint and Artificial intelligence(AI) for autonomous driving of drones” 2019

[4] Richard Sutton and Andrew Barto, “Reinforcement Learning: An Introduction,” 2th ed., 2017

[5] 이상엽 “Develope of robot game kernel for linkage virtual space and real space” 한국컴퓨터게임학회 2019 pp.161-166

[6] En-A Lim , Na-Young Kim , Jong-lark Lee, Ill-yong Weon “Applying Model to Real World through Robot Reinforcement Learning in Unity3D” 정보처리학회 2020 pp.800-803

[7] Min-Suk Kim “A Study of Collaborative and Distributed Multi-agent Path-planning using Reinforcement Learning” Vol. 25 No.3, March 2021, pp. 9-17

[8] Rodney A. Brooks “A Robust Layered Control System for a Mobile Robot”, 1986