

Reinforcement Learning Basic

Week1. Introduction to RL & OpenAI Gym

박준영

Hanyang University
Department of Computer Science

1 Orientation

2 Introduction to RL

3 OpenAI Gym

Course Information

Reinforcement Learning Basic (기초 강화학습)

- Textbook: Deep Reinforcement Learning Hands-On 2nd ed., Maxim Lapan
- 자료: <https://github.com/JYPark09/2021-HAI-Summer-RL>
- 본 강의에서는 Python3, NumPy, OpenAI Gym, PyTorch를 사용합니다.
- 질문이나 문의는 카카오톡 또는 이메일 jyp10987@gmail.com로 주세요.

Timetable

- **Week1 (7/23)**
 - Introduction to RL
& OpenAI Gym
- **Week2 (7/30)**
 - MDP & Cross Entropy Method
- **Week3 (8/6)**
 - Bellman Equation
& Tabular Learning
- **Week4 (8/13)**
 - Deep Q-Network
- **Week5 (8/20)**
 - Policy Gradient
- **Week6 (8/27)**
 - Actor Critic

Reinforcement Learning

Definition

강화학습은 주어진 상황에서 어떠한 행동을 취할지를 학습하는 것이다. 이때 그 행동의 결과는 **최대한의 보상(또는 이득)**을 가져다주어야 한다.

강화학습은 정답이나 잘못된 선택에 대한 정정이 주어지지 않는다.
→ 환경과 에이전트가 상호작용하며 행동에 대한 **보상만 주어진다.**

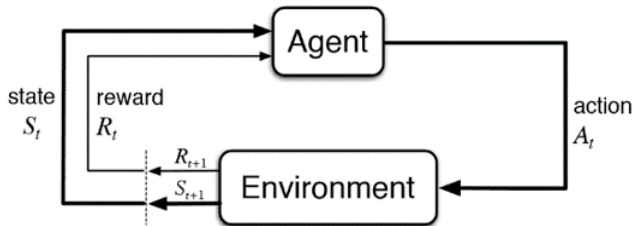
Reinforcement Learning

예시



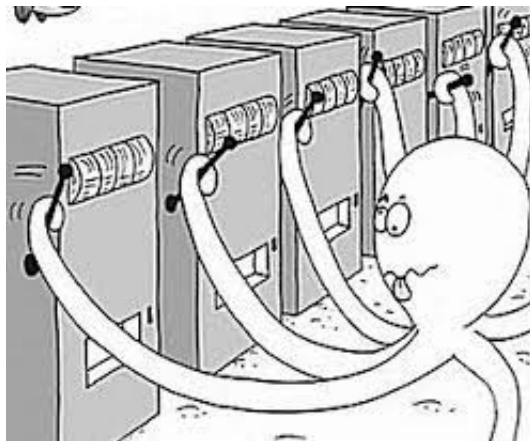
AlphaGo

Big Picture



- 환경은 에이전트에게 **보상(reward)**과 **관측(observation)**을 제공.
- 에이전트는 환경에서 자신의 상태를 인식한 후 행동.
- 보상을 통해 에이전트는 어떤 행동이 더 좋은지 알 수 있게 된다.
- 에이전트는 경험을 **활용(exploitation)**하면서 동시에 더 좋은 행동을 찾기 위해 **탐험(exploration)**한다.

Multi-arm Bandit



Multi-arm Bandit

- 환경: Bandit 기계
- 에이전트: 문어
- 행동: 한 Bandit 기계의 팔을 잡아 당김.
- 보상: 돈
- 상태: Bandit 기계의 난수표, 등.

Starcraft II



Starcraft II

- 환경: Starcraft II 전장
- 에이전트: 사령관(플레이어)
- 행동: 자원 채취, 건물 건설, 유닛 이동, 공격, gg, 등.
- 보상: 승리 or 패배
- 상태: 게임의 모든 정보
- **관측**: 사용자 화면에 보이는 정보

State vs Observation



OpenAI Gym

OpenAI Gym — <https://gym.openai.com>

→ 잘 정돈된 강화학습 프레임워크와 다양한 환경을 제공해줌.

- Algorithms
- Atari
- Box2D
- Classic control
- MuJoCo
- Robotics
- Toy text

Frozen Lake

- 4×4 크기의 맵에서 장애물을 피해 목표에 도달해야 한다.
 - S : 시작 지점
 - G : 도착 지점
 - F : 얼어 있는 지점 (안전함)
 - H : 구멍 (빠지면 게임 오버)
- 목표에 도착하면 +1의 보상을, 그 외의 상황엔 0의 보상을 받는다.
- 매 순간 상/하/좌/우로 이동할 수 있다.


S	F	F	F
	H	F	H
F	F	F	H
H	F	F	G

1. OpenAI Gym 설치



```
pip install gym
```

2. 환경 만들기



```
import gym  
  
env = gym.make('FrozenLake-v0')
```

- gym이 지원하는 환경 목록 - <https://gym.openai.com/envs/>

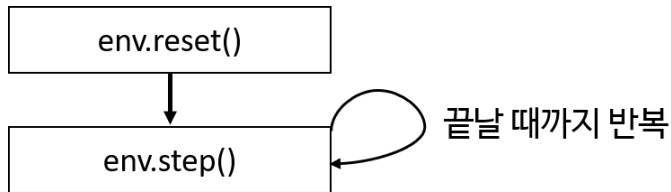
OpenAI Gym Environment

환경엔 다음의 멤버가 존재한다.

- `action_space`: 에이전트가 할 수 있는 행동의 집합
- `observation_space`: 환경이 에이전트에게 주는 정보의 집합 (reward 제외)
- `reset()`: 환경을 초기의 상태로 초기화 하는 메소드
- `step()`: 에이전트의 행동을 수행하고 다음 상태, 보상 등을 주는 메소드
- `render()`: 환경을 보여주는 메소드

3. 에피소드 진행하기

환경과 에이전트의 상호작용은 다음의 흐름대로 진행된다.



3. 에피소드 진행하기



```
state = env.reset()

while True:
    env.render()

    action = env.action_space.sample()
    new_state, reward, done, _ = env.step(action)

    if done:
        break

    state = new_state
```