

강화학습 개론

Assignment 2: Model-Free Reinforcement Learning

1. 목표

- Open AI Gym Taxi-v3 환경에서의 Model-Free RL 방법으로 문제 해결하기

2. 개발 환경

- Python (version 3.6+)
- gym==0.22.0 (%pip install gym==0.22.0)
- 조교는 리눅스 (Ubuntu) 환경에서 평가 합니다.

3. 템플릿 파일 및 제출 파일

- 제공되어지는 템플릿 파일 2개
 - taxi.py: Taxi 환경 설정 및 기초 템플릿 코드 (제출하지 않음)
 - agent.py: Model-free RL 구현 후 제출할 코드

4. Open AI Gym Taxi-v3 환경 소개

*** 먼저 Open AI Gym 을 설치 합니다.**

"There are 4 locations (labeled by different letters), and our job is to pick up the passenger at one location and drop him off at another. We receive +20 points for a successful drop-off and lose 1 point for every time-step it takes. There is also a -10 point as the penalty for illegal pick-up and drop-off actions."

https://github.com/openai/gym/blob/master/gym/envs/toy_text/taxi.py

```

+-----+
|R: | : :G|
| : : : : |
| : | : : : | |
| | : | : : |
|Y| : |B: |
+-----+

```

- 5x5 Map

```
env = gym.make('Taxi-v3')
```

```
action_size = env.action_space.n
```

```
space_size = env.observation_space.n
```

- State Space = $5 \times 5 \times 5 \times 4 = 500$.

State 는 taxi 위치 (5×5) 와 5 개의 승객 위치 (R, G, B, Y, 택시), 그리고 승객의 목적지 4 곳 (R, G, B, Y) 으로 구성 됩니다.

- The Action Space = 6

0. down, 1. Up, 2. Right, 3. Left, 4. Pickup, 5. Dropoff

Action 은 이렇게 정의 되어 있습니다.

- 위와 같이 환경이 reset() 되면, R, G, Y, B 지점 중 승객 위치 및 목적지가 랜덤으로 주어지고, 상하좌우로 움직여서 승객 위치로 이동 후 pickup 해서 목적지로 이동하여 dropoff 해야함.

- next_state, reward, done, _ = env.step(action)

reward 는 승객을 목적지에 내려주면 +20,

불가능한 행동을 하면 -10 (승객이 없는데 태우거나, 엉뚱한데 내려주거나),

그 이외 행동을 할 때마다 -1

- done : 올바른 목적지에 승객을 내려주면 True

200 step 도달하면 True, 그 이외는 False.

5. taxi.py 설명

* taxi.py 파일은 수정하지 않으며 제출하지도 않습니다.

해당 파일을 실행하면 아래와 action space 크기와, state space 크기를 출력한 후 메뉴가 출력됨.

```
1. Testing without learning
2. MC-Control
3. Q-learning
4. Exit
select:
```

그림 1. 실행 옵션 설정

1 번은 학습과 무관하게 맵을 이동하고 승객을 태우고 내리는 행동을 수동으로 해볼 수 있음.

2 번은 Monte Carlo control 방식으로 학습 후 테스트 100 번 에피소드의 평균 reward 가 6.0 이상이 되도록 함.

3 번은 Q_learning 방식으로 학습 후 테스트 100 번 에피소드의 평균 reward 가 7.0 이상이 되도록 함.

6. agent.py 설명

* agent.py 파일명은 "학번.py" 로 수정한 후 제출합니다.

```
class Agent:
```

```
    def __init__(self, Q, mode="mc_control"):
```

생성자로 넘겨받는 Q 변수가 q-table 에 해당함. mode 로 넘겨받는 값이 "mc_control", "q_learning", 또는 학습 완료 후 테스트 목적인 "test_mode" 에 따라 적절히 학습 및 테스트를 수행하도록 개발함.

템플릿 코드는 자유롭게 수정할 수 있음. 단 taxi.py 파일 수정 없이 연동되어야 합니다.

7. 제출 관련

- 제출 마감일: 4.16(일) 23:59.
 - 지연 제출은 받지 않습니다.
 - 표절은 양쪽 모두 F 입니다.
- "agent.py" 파일명을 "학번.py" 로 변경 후 iCampus 로 제출합니다.

8. 채점 기준

- Total 100 points
 - 50 points: 2. mc-control 학습 및 테스트 100회 평균 reward 가 6.0 이상 출력.
 - 50 points: 2. q-learning 학습 및 테스트 100회 평균 reward 가 7.0 이상 출력.