**Project**

**2018.09.14**

**1. 프로젝트 목표**

요즘 산업현장에서 강화학습을 사용하려는 많은 시도들이 있는데, 많이 사용되는 알고리즘들 중 하나가 deep Q-network (DQN)이다. 본 실습시간에는 OpenAI Baselines 라이브러리에 구현된 DQN 알고리즘을 사용하여 기초적인 로보틱스 실습들 중 하나인 cart-pole (inverted pendulum. OpenAI Gym library에 있는 것을 사용)을 제어해본다. 이 실습을 하기 전에 DQN의 이전 알고리즘인 Q-learning 알고리즘을 사용하여 미로에서 Q-value를 배워서 이를 제어 (미로에서 길을 찾아 나가는데)에 사용하는 실습을 하여 DQN에 대한 이해력을 높인다.

**2. 프로젝트 설명**

(1) Q-learning

실습에 사용할 maze-environment에서 제공하는 미로 들 중 "maze-sample-10x10-v0"는 10x10 미로로 왼쪽 모서리에서 agent가 출발하여 오른쪽 모서리에 도달하는 것이 이 환경의 목표로 설정되어 있다. 이 미로에서 Q-learning이라는 강화학습 알고리즘을 통해 강화학습 agent가 미로에서 길을 찾아 나갈 수 있게 학습시키는 것이 본 프로젝트의 1차 목표이다.

Maze-environment는 강화학습 알고리즘의 test bed library인 OpenAI Gym을 사용하여 만들어졌다. 그러므로 OpenAI Gym의 함수를 써서 우리가 만든 강화학습 알고리즘이 maze-environment와 상호작용하게 할 수 있다. 본 실습에서는 학습에 필요한 기본적인 Q-learning 코드를 제공한다. 실습을 통해 Q-learning 알고리즘을 이해하고, 주어진 환경에 우리의 강화학습 agent가 상호작용할 수 있도록 OpenAI Gym의 함수를 쓰는 방법을 배운다.

학습이 끝난 후, 실습자는 학습된 Q-value 표를 토대로, state-value 표를 계산해서 채운다. 그 후, 이를 이미지화 해서 state-value가 어떻게 형성되었고, agent에게 보여진 미로의 이미지와 비교했을 때 어떤 의미가 있는지 살펴본다.

(2) DQN

실습에 사용할 OpenAI Gym의 cart-pole environment는 말그대로 카트에 막대 (진자)가 연결되어 있는데, 시작할 때부터 막대기가 세워진 상태로 시작하고, 카트를 왼쪽/오른쪽으로 움직여서 (일정한 힘을 줘서 움직임) 막대를 위로 향하게 유지하는 것이 본 프로젝트의 2차 목표다. 이런 환경에 DQN을 학습시켜 DQN이 진자가 역으로 도립 된 상태에서 진자가 쓰러지지 않게 유지하기 위한 제어방식 (또는 policy)을 배워보도록 한다.

OpenAI Baselines라는 deep reinforcement learning library에 이미 구현되어 있는 DQN 모델 (deepq라는 이름의 폴더에 존재)과 OpenAI Gym library에 있는 cart-pole이라는 환경을 사용하여 실습을 진행한다. Baselines의 DQN 모델은 Tensorflow를 이용하여 deep Q-network를 만든다. 본 실습에서 학습에 필요한 기본적인 코드는 모두 제공된다.

DQN 학습이 어느 정도 수렴한 후, 에피소드에 대한 평균 return값, Q값을 학습 iteration 별로 그래프를 그려본다. 그리고 이에 대한 분석을 한다.

**3. 점수 기준**

* 주어진 Q-learning algorithm을 통해 학습하여 얻은 Q-value 표를 가지고 state-value 표를 계산한다. 그리고 이를 이미지화하여 맵 이미지와 비교하여 보고, state-value 표에 있는 값들에 대해 분석해본다. (40점)
* Baselines DQN을 OpenAI Gym cart-pole 환경에 대해서 학습시킨다. 학습 후, 에피소드에 대한 평균 return값과 Q값을 구하여 학습 iteration에 대해 그래프를 그리고 학습이 어떻게 진행 됐는지 분석해본다. (40점)
* 보고서 (20점)

**4. 보고서**

보고서는 별도의 포맷이나 분량의 제한을 두지 않으며, 아래의 내용을 포함해야 한다.

* Q-learning 알고리즘을 통해 학습된 state-value 표를 이미지화한 것.
* 이미지화 된 state-value 표를 미로 이미지와 비교 해보고, state-value 표의 값들이 갖는 의미를 분석한 내용.
* DQN이 학습되는 과정에서 얻은 return값과 Q값을 episode에 대해 평균내고, 매 timestep마다 평균값을 그래프로 나타낸 것.
* DQN 학습 결과로 얻은 그래프들 (return, Q값에 대한 그래프들)에 대해 분석한 내용.