OpenAI Gym 환경의 Mountain-Car에 대한 DQN 강화학습

강명주O
O청강문화산업대학교 게임콘텐츠스쿨
e-mail: mjkkang@ck.ac.krO

DQN Reinforcement Learning for Mountain-Car in OpenAI Gym Environment

Myung-Ju KangO
OSchool of Game, Chungkang College of Cultural Industries

● 요 약 ●

본 논문에서는 OpenAI Gym 환경에서 프로그램으로 간단한 제어가 가능한 Mountain-Car-v0 게임에 대해 DQN(Deep Q-Networks) 강화학습을 진행하였다. 본 논문에서 적용한 DQN 네트워크는 입력층 1개, 은닉층 3개, 출력층 1개로 구성하였고, 입력층과 은닉층에서의 활성화함수는 ReLU를, 출력층에서는 Linear함수를 활성화함수로 적용하였다. 실험은 Mountain-Car-v0에 대해 DQN 강화학습을 진행했을 때 각 에피소드별로 획득한 보상 결과를 살펴보고, 보상구간에 포함된 횟수를 분석하였다. 실험결과 전체 100회의 에피소드 중 보상을 50 이상 획득한 에피소드가 85개로 나타났다.

키워드: Mountain-Car(Mountain-Car), 활성화함수(Activation function), DQN(Deep Q-Networks)

I. Introduction

강화학습은 게임뿐만 아니라 로봇, 자율자동차 등 다양한 인공지능 분야에 적용되는 알고리즘이다. 강화학습은 에이전트의 현재 상태와 이에 따른 행위(action)을 통해 상호작용하여 보상을 최대화함으로써 목표 상태가 되도록 하는 행동을 학습하는 머신러닝의 한 분야이다.

본 논문에서는 OpenAI Gym 환경에서 Mountain-Car-v0 게임[1]에 대해 DQN (Deep Q-Network) 강화학습을 적용하였고, DQN 학습 결과를 보상 값을 통해 분석하였다.

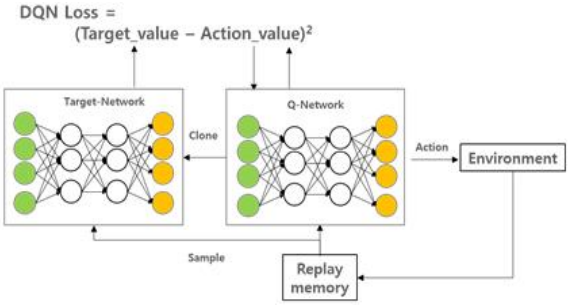

Fig. 1. structure of DQN

II. Preliminaries

1. Deep Q-Networks

DQN은 [2]에서 처음 소개한 강화학습 알고리즘이다. DQN은 Q-learning 알고리즘의 단점을 보완한 알고리즘으로 (state, action)으로 구성된 Q-table을 deque 메모리로 저장한 후 replay 시 샘플링 추출하여 학습에 사용하는 방법을 사용한다. [그림 1]은 DQN의 네트워크 구조이다.

2. Mountain-Car

OpenAI Gym에서 제공하는 Mountain-Car-v0[1]는 계곡 밑에 배치된 자동차를 언덕 위의 목표까지 간단한 제어를 통해 도달시키는 게임이다. 이 게임의 목표는 전략적으로 자동차를 가속해 오른쪽 언덕 위의 목표지점에 도달하는 것이다.

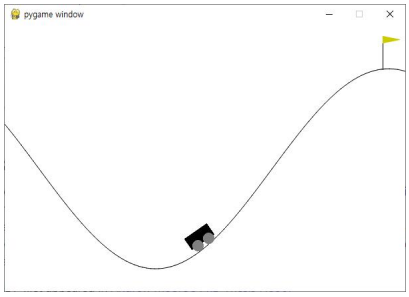

Fig. 2. Mountain-Car-v0

Mountain-Car에서의 제어환경은 Observation space와 Action space가 있다. Observation space는 x축을 기준으로 한 자동차의 위치와 자동차의 속도로 구성되어 있고, Action space는 자동차를 이동시킬 위치와 가속 값으로 구성되어 있다.

Observation space는 DQN의 입력층의 입력으로 사용되고, Action space는 DQN의 출력층의 결과로 얻을 수 있는 값이다. [Table 1]은 Mountain-Car의 Observation space를 설명하고 있고, [Table 2]는 Action space를 설명하고 있다.

Table 1. Observation space

Num	Observation	Min	Max	Unit
0	position of the car along the x-axis	-Inf	Inf	position(m)
1	velocity of the car	-Inf	Inf	position(m)

Table 2. Action space

Num	Observation	Value	Unit
0	Accelerate to the left	Inf	position(m)
1	Don't accelerate	Inf	position(m)
2	Accelerate to the right	Inf	position(m)

Mountain-Car의 이동전환은 다음 식에 의해 처리된다.

$$vel_{t+1} = vel_t + (action - 1) \times force - \cos(3 \times pos_t) \times gravity$$
$$pos_{t+1} = pos_t + vel_{t+1}$$

여기서 force는0.001이고 gravity는 0.0025이다. 자동차가 양쪽 벽과 충돌 시에는 속도는 0으로 설정되며, x좌표의 위치(pos)와 속도(vel)의 구간은 각각 [-1.2, 0.6], [-0.07, 0.07]로 제한된다.

III. Experiments

1. Experiments Environments

본 논문에서는 OpenAI Gym에서 제공하는 Mountain-Car-v0[1]에 대해 DQN 강화학습을 진행하였다. DQN의 뉴랄네트워크는 입력층, 은닉층 3개 그리고 출력층으로 구성하였다. 입력층과 은닉층에서의 활성화 함수는 ReLU이고 출력층의 활성화 함수는 Linear 함수이다. 각 층에서 적용되는 활성화 함수는 [표 3]과 같다[3].

Table 3. Activation functions

Name	Equation
ReLU	$f(x) = \max(0, x)$
Linear	$f(x) = x$

총 에피소드는 100회를 진행하였으며, 각 에피소드 당 강화학습은 3,000회 진행하였다.

2. Experiments Results

본 논문에서 적용한 Mountain-Car 게임의 목표는 계곡에 있는 자동차가 오른쪽 언덕 위의 목표지점까지 도달하는 것이다. 따라서 보상은 x의 좌표 구간 값 [-1.2, 0.6]에서 0.5 이상이면 목표지점에 도달한 것으로 판단하여 10을 부여하고, 그렇지 않고 -0.4보다 오른쪽에 위치하면 $(1+pos)^2$ 를 부여한다. 그 이외에는 보상이 없다.

각 에피소드에서의 score는 해당 에피소드에서 받은 보상 값을 더하여 계산하였다. 실험 결과는 [Table 4]와 같다. 실험 결과 총 100회의 에피소드 중 보상이 50~60인 경우가 가장 많았으며, 그다음이 60~70 사이이다.

Table 4. Count of rewards according to the episode

보상 범위	>=90	>=80	>=70	>=60	>=50	>=40	그외
횟수	6	3	9	22	45	12	3

[Fig 3]은 에피소드별 score의 값들을 그래프로 나타낸 것이다.

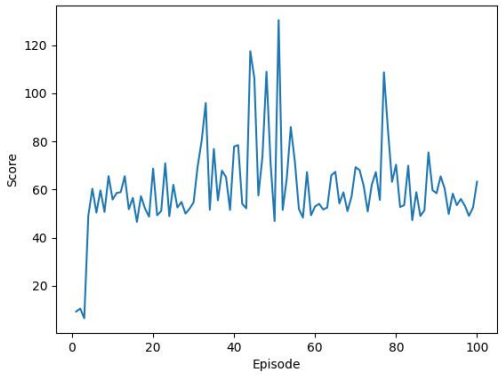

Fig. 3. Scores according to the episode

IV. Conclusions

본 논문에서는 OpenAI Gym에서 제공하는 Mountain-Car-v0에 대해 DQN 강화학습을 진행하여 목표 위치에 도달하는 정도를 알아보았다. 실험 결과 보상이 50~70 사이의 경우가 많았으며, 최종 목표지점까지 도달한 경우는 보상이 80 이상인 9회로 나타났다.

REFERENCES

[1] https://gymnasium.farama.org/environments/classic_control/mountain_car/

[2] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstr, Martic Riedmiller, “Playing Atari with Deep Reinforcement Learning”, arXiv:1312.5602v1, 2013

[3] https://keras.io/api/layers/activations/