한국컴퓨터정보학회 동계학술대회 논문집 제31권 제1호 (2023. 1)

OpenAI Gym 환경에서 강화학습의 활성화함수 비교 분석

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

Comparison of Activation Functions of Reinforcement Learning in OpenAI Gym Environments

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

● 요 약 ●

본 논문에서는 OpenAI Gym 환경에서 제공하는 CartPole-v1에 대해 강화학습을 통해 에이전트를 학습시키고, 학습에 적용되는 활성화함수의 성능을 비교분석하였다. 본 논문에서 적용한 활성화함수는 Sigmoid, ReLU, ReakyReLU 그리고 softplus 함수이며, 각 활성화함수를 DQN(Deep Q-Networks) 강화학습에 적용했을 때 보상 값을 비교하였다. 실험결과 ReLU 활성화함수를 적용하였을 때의 보상이 가장 높은 것을 알 수 있었다.

키워드: 활성화함수(Activation function), 강화학습(Reinforcement Learning), DQN(Deep Q-Networks)

I. Introduction

강화학습은 머신러닝의 한 종류로 동적환경에서 정의된 에이전트가 시행착오를 통해 현재 상태에서 보상을 최대화하는 행동을 학습하는 방법이다. 강화학습은 자율자동차, 로봇뿐만 아니라 게임에서도 많이 적용되는 알고리즘이다.

본 논문에서는 OpenAI Gym 환경의 CartPole-v1 게임에 대해 강화학습에 적용되는 활성화함수의 성능을 비교한다.

II. Preliminaries

1. DQN (Deep Q-Networks)

DQN 개념은 [1]에서 처음 소개한 알고리즘으로 상태(state)에 따른 행동(action)을 Q-table로 정의하여 학습하는 Q-learning의 단점을 보완한 강화학습알고리즘이다. Q-learning 알고리즘에서는 상태 s와 행동 a로 구성된 Q(s, a) 테이블 형태로 저장하여 학습한다. 이런 방식은 state space와 action space가 커지면 Q 테이블을 저장하는 메모리가 커지고 탐색시간이 길어지는 단점이 있다. 이러한 Q-learning의 단점을 보완하기 위해 DQN에서는 Q테이블에서 replay 메모리로 샘플링 추출하여 Q테이블을 업데이트하는 방법을 사용한다.

DQN의 학습에 적용되는 손실함수는 샘플링데이터를 타겟으로하고 샘플링데이터를 Q-network를 통해 학습한 데이터를 예측값으로 하여 타겟과 예측값의 오차를 최소화하도록 설계되어 있다[1].

2. Activation Function

머신러닝의 뉴럴네트워크 아키텍처는 입력층, 은닉층, 출력층으로 구성된다. 각 층에서는 이전 층에서의 입력값과 가중치(weight)를 곱하여 더한 값을 활성화함수(Activation function)에 적용한 후 다음 층의 입력 값으로 사용된다. 활성화함수는 머신러닝의 학습에 큰 영향을 주기 때문에 주어진 문제에 따라 적합한 활성화함수를 사용해야 한다. 본 논문에서 적용한 DQN 네트워크의 활성화함수는 다음과 같다.

(1) Sigmoid

시그모이드 함수는 강화학습 초기에 사용되었던 활성화 함수로 주어진 입력 값에 대해 0과 1사이의 값을 반환하는 함수이다. 함수의 수식은 다음과 같다.[2]

$$f(x) = \frac{1}{1+e^{-x}} \quad (1)$$

(2) ReLU

ReLU(Rectified Linear Unit) 함수는 머신러닝에서 많이 사용되는 활성화 함수로, 입력 값이 0이나 음수일 때에는 0을 반환하고, 양수일 때에는 양의 선형 값을 반환하는 함수이다.[2]

$$f(x) = \max(0, x) \quad (2)$$

(3) ReakyReLU

LeakyReLU 함수는 ReLU를 변형한 함수로, 0 이하의 입력에 대해서도 작은 수 α값을 곱한 값을 반환하는 함수이다.[2]

$$f(x, \alpha) = \max(\alpha x, x) \quad (3)$$

(4) softplus

softplus는 ReLU 함수를 부드럽게 근사한 것으로 결과 값을 항상 양수로 제한할 수 있는 함수이다.[2]

$$f(x, \beta) = \frac{1}{\beta}\log(e^{\beta x} + 1) \quad (4)$$

III. Experiments

1. Experiments Environments

본 논문에서는 OpenAI Gym에서 제공하는 CartPole-v1[3]을 DQNAgent에 위에서 설명한 활성화함수를 각각 이용하여 학습하였다. CartPole-v1은 Pole, Cart, Joint, 그리고 Track으로 구성되어 있다. Pole은 Cart에 Joint로 연결되어 Track을 따라 움직일 수 있고, Cart를 좌우로 움직여 Pole이 균형을 유지하도록 하는 게임이다.

CartPle-v1 게임에서 Agent가 관측하는 Observation Space는 4가지로 다음과 같다.

[참고문헌 :

Table 1. CartPole-v1 Observations

Num	Observation	Min	Max
0	Cart Position	-4.8	4.8
1	Cart Velocity	-inf	inf
2	Pole Angle	~ -24o	~ 24o
3	Pole Angular Velocity	-inf	inf

이 게임에서 보상은 Pole이 가능한 오랫동안 균형을 유지해야하기 때문에 각 episode에서 처리되는 step의 횟수로 한다. DQN을 통한 학습 횟수는 10,000회이다.

2. Experiments Results

각 활성화 함수가 학습에 끼치는 영향을 평가하기 위해 활성화함수를 사용하여 학습했을 때 얻게 되는 보상의 평균, 최소, 최대값을 비교한 결과는 [Table 2]와 같다. 실험 결과 ReLU 활성화 함수를 적용한 학습의 보상이 가장 높았다. 따라서 ReLU의 성능이 가장 좋다고 할 수 있다.

Table 2. Mean/Min/Max Reward for each activation function

Activation func	Mean	Min	Max
Sigmoid	31.57	8.0	239.0
ReLU	117.16	9.0	464.0
ReakyReLU	108.24	11.0	369.0
softplus	62.7	8.0	353.0

IV. Conclusions

본 논문에서는 CartPole-v1 게임에 대해 에이전트가 강화학습을 통해 학습할 경우 학습에 적용되는 활성화함수의 성능을 비교 분석하였다. 실험결과 ReLU 활성화함수를 적용하였을 때의 보상이 평균적으로 가장 크다는 것을 알 수 있었다.

REFERENCES

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

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

[3] https://www.gymlibrary.dev/environments/classic_control/cart_pole/