copy\_weight부분을 이용해 디버깅을 해보자. variable이 변화하는지, trainnet에서 targetnet으로 복사가 잘 되는지를 출력시켜서 확인해 본 결과 놀랍게도, weight이 변하고 있으며 복사도 착실히 진행됨을 알 수 있었다. 그럼 뭐가 문제일까? 학습 방향 자체가 문제인가?

<https://github.com/hunkim/ReinforcementZeroToAll> 강의의 소스코드를 보려했는데, 여기에 있는 소스는 또 다른 소스인 것 같다.

강의 코드와 약간의 차이점이 있다. 강의 코드는 episode 10번마다 한번 training을 시켜주는데, 10사이즈 minibatch를 50번 반복해 교육시키고 있었다. 반면, 내가 tf2용으로 가져온 코드는 episode 도중 매 step마다 minibatch(32)를 뽑은 뒤 학습을 시키고 있다. 이걸 한 50번 반복시키면 되려나???

일단 다시 origin 소스코드에 hidden layer만 10\*10으로 줄여 학습시켜보고있다. loss가 600에서 줄기 시작했지만 100 이하로 떨어지지를 않는다. average reward는 랜덤보다 못한 수준이다. 15000회 결과 e값이 상당히 줄었음에도 loss는 10~20사이를 와리가리 한다. 잠깐만, 혹시? 최대진행횟수 200제한에 걸려서? 끝나면 done이 되고, 계속 -200점을 먹어서 그런건가?

1. 최대 횟수 늘려보기

생각해보니 rewards는 done을 만날 시 -200과 관계가 없더라… 학습에 걸리는 시간만 늘어난 것 같다. 느낌 상? 근데 그러면 reward가 늘어야 정상인데 왜 reward는 그대로지?

1. reward -200을 -100으로 줄여보기?

loss가 100 이하로 안 떨어지던 것이 50대가 되었다. 하지만 reward가 여전히 끔찍하다.

1. epsilon 강의처럼 바꿔서 해보기

여전히 답 없음.

1. hidden layer [10]으로 줄여보기
2. minibatch 반복학습 시켜 보기 = 또는 Lr늘려보기?

속도를 빠르게 하기 위해 log 빈도 줄여보기. 빨라진 느낌이 살짝 든다.

움직이는 모습 보려고 200번만 학습시키고 render해봤다. 그랬는데… 안넘어진다. 이게 무슨일이야. 10001스텝 찍을때까지 안넘어진다. 대체 뭐가 일어난거지?

근데 또 두번 더 반복하니까 형편없는 결과를 보여준다. 엄청 운 좋게 초기값이 좋았던건가? 뭐지?

최대 step을 3000정도로 줄이고, log빈도를 다시 늘린 뒤 이번에는 N=400정도로 해봐야겠다. epsilon decay도 .9->.99로 늘리고.

너무 불안정하다. reward와 loss가 널뛰기한다.

epsilon decay .9999로 다시 늘려봤다. 여전히 널뛰기 한다. reward는 e가 늘었으니 당연하지만, 테스트 결과도 너무 안좋았다. loss는 그나마 착실히 줄어갔다.

decay를 다시 0.9로 줄여봤다. reward는 바닥을 치고, loss는 500대에서 줄지를 않는다.

1,2)decay를 0.999로 만들고, layer모양을 10 하나로 만들어보자. 여전히 안좋다.

3)layer모양을 100 하나로 만들어보자. 의미 없는 학습곡선이다.

4)layer를 100인채로 둔 채, bias를 없애보자. 엉망진창의 결과가 나왔다.

5)bias를 없앤 채로 initializer를 glorot uniform으로 바꿔봤다. 여전히 엉망진창이다.

6)layer를 10으로 다시 줄여볼까? >bias 하나 없앴다고 이렇게 널뛰기하는 이유가?

7)lr을 늘려보자, 강의처럼. 0.01->0.1그대로 엉망.

8)카피 타이밍도 아예 줄여버리고. 25->10. 성능 좀 나오나? 싶었는데 loss는 여전히 엉망.

9)강의에서는 gamma가 0.9더라고. 0.99->0.9 …..

10)epsilon을 다시 강의처럼 만들었다…. 아무런 변화가 없다.

11) batch를 늘려보자. 32 -> 100

12) bias를 다시 돌려놔보자…loss는 안정화 되었다. reward도 봐줄만 해 졌다. reward가 널뛰기 하하는 것 좀 어떻게 하면…

13) 12에서 lr을 줄이거나, e를 다시 복구시키거나. 먼저 lr부터.0.01

처음엔 정신좀 못차리다가 스무스하게 loss가 줄고, 근데 reward는 여전히… 좀 안정적으로 된 것 같긴 한데.

14) e를 복구하고 decay 0.993으로. loss가 정말 완벽하게 학습된다! reward도 상당히 좋은 것 같다! episode 개수를 좀 더 늘려보면 되지 않을까???

15)통 크게 1000으로 늘려준 episode 개수. 그러나… reward가 떡상하는 일은 없었다…

회색 14번, 주황 15번. 올라가지를 못한다….

왜일까, 뭐가 문제일까? overfitting? overshooting? overestimating???

계속 만져 주는 중 relu로 바꿔보기도 하고 decay값도 만져주고… loss는 어느정도 내려가지만 reward는 항상 널뛰는데, 바가 한 방향으로 쭉 가서 벽에 부딪혀 사라지는 현상이 계속 발생한다.

19) batch 100->10 loss가 감소하다 솟구치기 시작. reward도 증가하다가 급감

reward가 증가한 상태를 유지하지 못하는 이유는 무엇인가?

20) relu->tanh. loss가 다시 이상해졌다

학습에 영향주는 요소들

초기화방법 Glorot uniform

activation function tanh

gamma값 0.9

learning rate 1e-2

decay

(min epsilon)

hidden layer

batch size

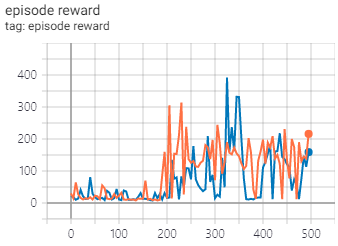
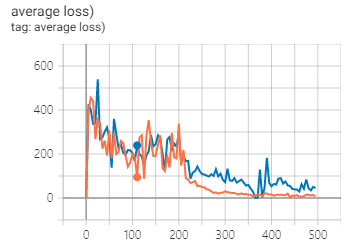
copy step

done penalty

21) batch 10->30, lr 0.002 ->0.01

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 초기화 | AF | gamma | lr | de/min | hid ly | batch | cp step | done | ep n |
| Glorot | tanh | 0.9 | **1e-2** | .993/e-2 | 10 | **30** | 10 | -100 | 500 |

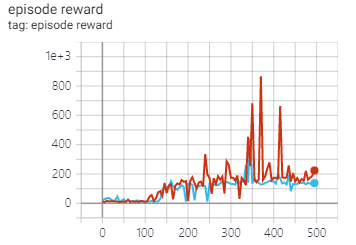
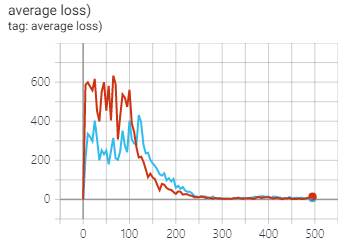
벽 박치기 현상



22) batch 30 -> 32, copy step -> 25

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 초기화 | AF | gamma | lr | de/min | hid ly | batch | cp step | done | ep n |
| Glorot | tanh | 0.9 | 1e-2 | .993/e-2 | 10 | **32** | **25** | -100 | 500 |

벽 박치기 현상 여전히. 특이한 현상으로, 학습도 안 된 모델이 시작부터 3001step을 찍는 현상이 있었다. 학습이 진행되면서 사라졌다.



23) epsilon을 좀 더 급격히 줄여보자.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 초기화 | AF | gamma | lr | de/min | hid ly | batch | cp step | done | ep n |
| Glorot | tanh | 0.9 | 1e-2 | .**99/e-2** | 10 | 32 | 25 | -100 | 500 |

눈에 띄는 변화는 없다.

24) lr 5e-2 5배

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 초기화 | AF | gamma | lr | de/min | hid ly | batch | cp step | done | ep n |
| Glorot | tanh | 0.9 | **5e-2** | .993/e-2 | 10 | 32 | 25 | -100 | 500 |

loss 줄어드는 속도는 늘었는데, 의미없는 것 같다. 롤백해야겠다.

25) lr 롤백, 벽에 박는 것을 응징하기 위해 done -300?

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 초기화 | AF | gamma | lr | de/min | hid ly | batch | cp step | done | ep n |
| Glorot | tanh | 0.9 | **1e-2** | .993/e-2 | 10 | 32 | 25 | **-300** | 500 |

loss 감소되는 경우도 있고 안되는 경우도 있는 신기한 케이스

