

내가 정리해본 용어 설명 (틀릴 가능성 많음)

* 뉴럴네트워크

뉴런은 함수와 같다. input을 넣어주면 결과값을 뽑아줌.

* Discount Factor

강화학습의 핵심은 Reward를 최대로 하는 것이다.

우리는 이것을 Discount Factor를 이용해서 할 수 있다.

어떤 reward를 받아야 최종적인 reward의 합계를 최대로 만들 수 있을까?

예를 들어, 우리가 Discount Factor를 0으로 설정해놓으면 DeepRacer(agent)는 오직 다음행동에서 얻을 Reward만을 최대화하는데 관심을 쏟는다. 하지만, Discount Factor를 1로 설정해놓으면 미래의 Reward까지 생각하는 것이다. 즉, 자동차 경주와 같이 완주가 필수인 environment에서는 이런 Discount Factor가 중요하다. Discount Factor는 0부터 1까지 값을 가질 수 있다. Discount가 클수록 모델은 더 멀리 보면서 누적기대보상을 확인. discount가 너무 크면 학습 속도는 느려진다.

0.9 → 향후 10단계로 넘어갈 수 있는 보상을 고려

0.999 → 향후 1000단계의 보상을 고려하여 이동

권장 discount : 0.99 , 0.9999, 0.999999

Course4 - Reinforcement learning (ppt에서)

Step = state, action, reward, new state로 구성되어 있는 하나의 튜플

Episode = 차량이 정상적으로 완주했을 때 or 차량이 트랙을 벗어났을 때. 에피소드가 1 증가. 즉, 에피소드는 Step들의 집합이라고 볼 수 있다.

Experience buffer = 정해진 수의 에피소드로부터 수집된 스텝들의 집합. 그니까 우리가 만약 에피소드를 3이라고 했다면 그 3번의 에피소드 동안 (당연히, 에피소드들은 완주했을 경우도 있을 거고 50%쯤에 이탈했을 수도 있고, 시작한지 얼마 안되서 이탈했을 수도 있고 등등.. 각각의 에피소드는 길이가 다를 것임) → 사용되는 뉴럴 Network를 업데이트하는 데 사용된다.

배치 = 일정 기간 동안 시뮬레이션에서 얻은 경험의 일부를 나타내는 순서가 있는 경험 목록.

Course5 - HyperParameter (ppt에서)

Batch Size = 경험 버퍼에서 무작위로 샘플링되서 딥러닝 뉴럴 네트워크의 가중치를 업데이트 하는데 사용될 Step의 개수

총 반복횟수 = epoch는 모든 배치를 통과하는 한번의 패스를 나타내며, 각 배치가 처리된 후 다음 배치로 진행하기 전에 뉴럴 네트워크의 가중치가 업데이트 된다.

학습 속도 = 뉴럴 네트워크 가중치의 업데이트 정도를 제어. 최대 누적 보상을 얻기 위해 정책의 가중치를 변경해야 할 때, 정책을 얼마나 변경해야 할지를 제어

학습속도가 클수록 학습이 빨라짐. 수렴에 어려움을 겪을 수 있다.

탐사 = 무작위로 행동을 선택하는 것을 의미 (exploration과 exploitation 사이의 균형을 결정하는 데 사용되는 방법)

엔트로피 = 액션 공간의 확률 분포에 추가될 불확실성 또는 임의성 정도

우리가 얻은 데이터에 어느 정도의 불확실한 데이터가 포함되어 있는지에 대해서 나타낸 것. 우리가 사용하는 딥렌즈는 매우 명확한 이미지를 얻으므로 그러한 불확실성이 낮다. 그러니까 0.01을 준거다.

Loss 유형 = 네트워크 가중치를 업데이트하는데 사용되는 목적함수(비용함수)의 유형을 지정. Huber = 수렴에 문제가 있을 시 사용, Mean Squared error loss = 수렴이 양호하고 더 빠르게 훈련하기 위해 사용

학습간 에피소드 회수 = 모델 학습 사이에 자동차가 얼마나 많은 경험을 얻어야 하는지를 제어 10,20,40 (권장되는 값)