|  |  |
| --- | --- |
| 과목명 | 강화학습 (3-3) |
| 이름 | 이덕수 |
| 학번 | 20161600 |

이름과 학번을 적고, 답변을 적어주세요.

실습: 3주차 수업 내용을 자유롭게 요약하세요(A4 1 page 분량, 넘어도 상관없음)

- MDP(마르코프 결정과정) > 모든 상태에서의 모든 보상 과 상태전이확률 이 정해져 있음 이 값을 이용하여 가치 함수(가치의 기댓값) 을 구할수 있음 즉 이를 이용하여 최적의 솔루션을 구할수 있음 (모델에 대한 완전한 지식이 필요함)

만약 이러한 과정과 보상값 가치값을 모를경우 MDP를 제대로 적용하기 어려움

이를 대체하는 방법이 모델 프리 강화 학습 이라는 방법임.

완전 탐색을 적용할수 없는경우 (MDP 적용 불가) 라는 경우시

모델 프리 강화 학습 을 사용하는데 대표적으로 : 몬테카를로 학습 ,

시간차 학습 등이 있음 이 방법들은 경험을 누적하는 형태로 정책을 만들어 가며

이를 휴리스틱 알고리즘 이라고 명명함.

모델 프리 강화 학습은 모델에 대한 완전한 지식이 필요하지 않음

몬테카를로 학습은 줄여서 MC학습이라 부름 모델에 대한 완전한 지식이 필요하지 않고 완전한 가치 함수를 구할 수 없기에 가치 함수의 추정치 를 구하는 방식으로 문제 해결에 접근 (최적의 행동이 무엇인지 모르는 상태에서 일단 행동을 취하여 얻은 경험으로 정책을 수정하면서 배움)

이로인한 장점 과 단점으로 비결정적 이고 환경의 모델을 파악할 수 없는 강화 학습 문제를 근사하게 풀 수 있음

하나의 에피소드가 비교적 짧은 단계로 마무리 되는 환경에 적용이 잘됨

하지만 MC학습은 에피소드가 무한히 지속되면 경험을 기반으로 정책을 수정하기에 학습이 느려지는 점과 경험을 바탕으로 더욱 나은 정책을 수립하게 된다면

이에 비례하여 에피소드는 더욱 길어지기에 학습 속도는 점점 더 느려지는 문제가 발생하게 됨.