|  |  |
| --- | --- |
| 교육 제목 | **AI 프로젝트 기반 S/W 전문가 양성 과정** |
| 교육 일시 | 21. 11. 29. |
| 교육 장소 | 영우글로벌러닝 |
| 교육생 | 박건준 |
| **교육 내용** | |
| 오전 | 강화 학습.  Exploration(탐험가) and Exploitation(미탐험)  Polcy  <https://www.davidsilver.uk/teaching/>  Markov Decision Processes  David Silver 알파고 만든 사람?  람다 값을 주면서 시간의 흐름 값을 보며 문제를 해결 해 나감.  과거와는 상관 없이, 현재의 상황으로 미래를 표현 할 수 있다.  episod라는 하나의 사건 내에서 state Transition Matrix 일어나는일은  하나하나의 에피소드를 메트릭스 방정식으로 모으로 1개의 애피소드 값은 1. 행동패턴 등이 [ ] 안에 모여 있음.  일어났던 해동 패턴을 Markov chain 이라고 하고 스테이트와 프로세스의 형태로 튜플로 만듬. 하나의 스테이트에서의 하나의 패턴을 모은 것.  Example : StudentMarkov Chain  Class 1 과정에서 class2의 ML 과정으로 넘어가는 과정에서의 확률을 그림으로 만들고, 돌고돌아 다시 진행 되는 것 등의 과정도 메트릭스로 예상 될 수 있음. 이러한 S1=C1 C라는 시작점과 S라는 스테이트로 변화 할 수 있는 트레지션 메트릭스 방정식으로 표현 할 수 있음. 그림으로 되어 있는 과정을 2차원의 데이터로 바꿀 수 있음. 각 1개의 C의 총 합의 값은 1.. 이 상태에서 리워드를 넣음. S(스테이트), P(process), Rs(리워드) reward function의 리워드  에피소드가 아닌 토탈에서의 리워드 Return  Value Function state value function인 v(s) of an MRP 리워드 프로세스로 Expetede 전체 상태에서의 Return 인 reward 총합  현 상태에서 어떤 액션을 취했을 때 총합이 얼마인지 확인한 것은 액션 |
| 오후  (2시) | 몬테카를로 러닝  예측한 것을 해보고, 옵티마이즈 해보며 다음 예측하고,  템플로디스러닝은 끝까지 가는 목표까지의 모든 벨류를 예측하고 옵티마이즈 하는학습  모델 프리 컨트롤  파셜리 = 모든 것을 예측할 수 없는 상황에서의 학습은  몬테나 템플로로 예측하고 예측한 걸 옵티마이즈 써보고 또 예측하는 것을 반복 함.  강화학습에는  에이전트, 환경이 있다. 리워드?  Ndp  마크오브프로퍼티  리워드  디시젼 프로세스 내에  벨멘  벨류펑션을 잡아 다이나믹 프로그램을 이터너밍하며 찾아감.  ML의 지도와 비지도학습이 취업에도 좋고 지금 레벨에 좋다.  다른 학원에 비해 많은 양의 ai 정보를 갖고 있다?  다른 학원은 어플을 만드는 등 코딩 트레이닝이 중점이다.  DL은 링크를 통해 더 알아봐라. |