TAVE 서기

|  |  |  |  |
| --- | --- | --- | --- |
| **서기 내용** | | | |
| **서기 일자** | 22.1.11 | **서기** | 서가을 |
| **주제** | 강화학습 후반부, KAIST ‘디지털 빅뱅, 메타버스 기술’ 온라인 국제포럼 | | |
| **시간** | 22:45 – 22:30 | **장소** | Zoom 미팅, 트위치 |
| **스터디**  **인원** | 고성호, 권기호, 이아현, 서가을 : 시작  텍스트, 모니터, 화면, 텔레비전이(가) 표시된 사진  자동 생성된 설명  고성호, 권기호, 이아현, 서가을 : 종료  텍스트, 스크린샷, 전자기기, 컴퓨터이(가) 표시된 사진  자동 생성된 설명 | | |
| **내용** | | | |
| **배운 내용** |  | | |
| **Chapter18. 강화학습**  **18.12 TF-Agents 라이브러리**  **18.12.1 TF-Agents 설치하기**       * 전체 gym 라이브러리 중 atari-py 설치   **18.12.2 TF-Agents 환경**  **텍스트이(가) 표시된 사진  자동 생성된 설명**   * colab을 사용했기 때문에 colab에서는 직접 파일을 불러와야 함.     **18.12.3 환경 스펙**   * 제공하는 스펙: 관측, 행동, 타임 스텝의 크기, 데이터 타입, 이름과 최솟값, 최댓값 * 관측 용량이 크기 때문에 다운샘플링 + 흑백으로 변환 -> 훈련 속도 높아지고 RAM도 절약 * 이를 위해 환경 래퍼 사용   텍스트이(가) 표시된 사진  자동 생성된 설명   * 관측 용량이 크기 때문에 다운샘플링 + 흑백으로 변환 -> 훈련 속도 높아지고 RAM도 절약 * 이를 위해 환경 래퍼 사용   **18.12.4 환경 래퍼와 아타리 전처리**  TF-Agents가 제공하는 환경 래퍼들   * ActionClipWrapper : 행동 스펙에 맞게 행동을 클리핑(제한)한다 * ActionDiscretizeWapper : 연속적인 행동공간을 이산적인 행동공간으로 양자화(?) * DQN은 이산적인 행동공간만 지원하므로 0, 1, 2, 3 등 이렇게 주어진 종류의 정수로 끊어줄 수 있어야 함   discrete\_env = ActionDiscretizeWrapper(env, num\_actions = 5) -> 5개의 가능한 행동을 가짐   * ActionRepeat : 행동을 n 스텝동안 반복하면서 보상을 누적. 훈련 속도를 높일 수 있음 * RunStats : 스텝 개수와 에피소드 개수 같은 환경 통계 기록 * TimeLimit : 최대 스텝 개수보다 오래 실행되면 환경 중지      * 래핑할 환경 생성자에게 전달, 래퍼 만듦   아타리 환경에서의 전처리래퍼 ‘AtariPreprocessing’   * 흑백 변환과 다운 샘플링 * 맥스풀링 * 프레임 스킵: 4번마다 게임을 봄(4프레임마다 렉을 걸어 하드웨어를 적게 사용하게 함) * 죽었을 때 게임 종료      * Ale-py가 없다 해서 설치함 * Colab에서 돌리기 어려움   텍스트이(가) 표시된 사진  자동 생성된 설명   * 래핑된 아타리 환경 만듦   **18.12.5 훈련 구조**     * 왜 환경이 여러 개인가?: 환경을 여러 개 복사해서 GPU로 병렬 탐색하면 빠름 * 경로(trajectories)란?: 한 타임스텝에서 다음 타임스텝으로의 전환, 이 경로로 옵저버에게 전달 및 재생 버퍼에 전달됨 * 왜 옵저버가 필요한가?: 드라이버가 그냥 경로 저장하면 재생버퍼를 사용하지 않는 것으로 되어, 드라이버가 유연성이 떨어짐   **18.12.6 심층 Q-네트워크 만들기**   1. 입력이 여러 개면 일단 각각 전처리 2. 전처리 된 입력을 연결 3. 합성곱 층 리스트 적용 4. 출력을 위한 밀집 층 적용  * 필요 시 드롭아웃 적용 가능   텍스트이(가) 표시된 사진  자동 생성된 설명   * 다음은 전처리 과정   **18.12.7 DQN 에이전트 만들기**  텍스트이(가) 표시된 사진  자동 생성된 설명   * 훈련 스텝 횟수 헤아릴 변수 만듦 * 4스텝마다 모델 훈련 * 옵티마이저 만들기 * 입실론 정책을 위한 객체 만들기 * 손실함수 reduction = “none”인 이유: 평균이 아니라 샘플마다 하나의 오차 반환해야 함 * 에이전트 초기화   **18.12.8 재생 버퍼와 옵저버 만들기**   * 재생 버퍼의 매개 변수   1. data\_spec: 재생 버퍼에 저장할 데이터의 구조   2. batch\_size: 각 스텝에서 추가될 경로 개수   3. max\_length: 재생 버퍼의 최대 크기   텍스트이(가) 표시된 사진  자동 생성된 설명  **18.12.9 훈련 측정 지표 만들기텍스트이(가) 표시된 사진  자동 생성된 설명**   * 각 지표의 값을 확인하고 싶으면 언제든 result() 메서드 또는 log\_metrics(train\_metrics) 호출   텍스트이(가) 표시된 사진  자동 생성된 설명   * 위와 같은 방식으로   **18.12.10 수집 드라이버 만들기**   * 드라이버: 주어진 정책으로 환경을 탐색하고 경험을 수집하고 옵저버에 이를 전파하는 객체  1. 현재 타임 스텝을 수집 정책에 전달 -> 수집 정책이 타임 스텝을 활용해 정책 선택 + 행동 스텝 반환 2. 드라이버는 그 행동을 환경에 전달 + 다음 타임 스텝을 반환 받음 3. 이 전이를 표현하는 경로 객체를 만들고 모든 옵저버에게 전달  * DynamicStepDriver: 주어진 스텝 횟수에 대한 경험 수집 * DinamicEpisodeDriver: 주어진 에피소드 횟수에 대한 경험 수집 * 여기서는 스텝 4개에 대한 경험을 수집 -> DynamicStpeDriver 만듦   텍스트이(가) 표시된 사진  자동 생성된 설명  **18.12.11 데이터셋 만들기**   * get\_next() 메서드 호출: 재생 버퍼에서 경로의 배치를 샘플링하기 위해, 경로의 배치 반환 + 샘플 식별자와 샘플링 확률을 담은 BufferInfo 객체 반환        * 첫 두 차원에 크기가 2, 3인 텐서 담음: observation 필드 크기가 [[2, 3, 84, 84, 4]인 이유 * trajectories객체는 필드 7개로 이루어진 네임드 튜플   텍스트이(가) 표시된 사진  자동 생성된 설명텍스트이(가) 표시된 사진  자동 생성된 설명  **18.12.12 훈련 반복 만들기**  텍스트이(가) 표시된 사진  자동 생성된 설명   * 훈련 속도 높이기 위해 주 함수를 텐서플로 함수로 변경 * 이를 위해 tf.function()으로 감싸고 tf\_agetns.utils.common.function()함수 사용 * 설정해 놓은 interator만큼만 진행됨   **18.14 연습문제**  1. 강화학습을 어떻게 정의할 수 있나요? 지도학습/비지도학습과의 차이점?  텍스트이(가) 표시된 사진  자동 생성된 설명  2. 이 장에서 언급하지 않은 강화학습 어플리케이션  1) 음악 개인화:   * + 환경 : 사용자의 개인화된 웹 라디오   + 에이전트 : 사용자에게 다음에 어떤 노래를 재생할지 결정하는 SW   + 가능한 행동 : 카탈로그에 있는 노래 재생, 광고 재생   + 보상 : 노래 들을 때 작은 보상, 광고 들을 때 큰 보상, 스킵하면 음수 보상, 사용자가 탈주하면 큰 음수 보상   2) 마케팅   * + 환경 : 회사의 마케팅 부서   + 에이전트 : 고객 프로파일과 구매 이력을 바탕으로 홍보 메일을 보낼 고객을 선정하는 SW   + 가능한 행동 : 보낸다, 보내지 않는다   + 보상 : 홍보 메일의 발송 비용에 대해 음수의 보상, 예상 매출을 양수의 보상   1. 배달 음식 추천 * 환경 : 사용자의 개인화 된 스마트폰 어플리케이션 * 에이전트 : 사용자에게 배달음식을 추천하는 SW * 가능한 행동 : 카탈로그에 있는 음식 추천, 광고 * 보상 : 음식을 시킬 때 작은 보상, 광고 음식을 시킬 때 큰 보상, 스킵하면 음수 보상, 사용자가 탈주하면 큰 음수 보상   3.할인계수? 할인계수 바꾸면 최적의 정책 바뀌는가?   * 0에서 1 사이의 값을 가짐 * 미래에 받을 수 있는 보상보다 지금 당장 얻을 수 있는 보상에 더 가중치 * 1에 가까울 수록 즉각적인 보상과 미래에 받을 수 있는 보상을 동등하게 취급 * 보통 0.99에 가까운 값으로 설정 * 최적의 정책에 큰 영향을 미침   + 미래를 가치 있게 여기면 당장의 고통을 감내하고 큰 보상을 기대함   + 미래를 가치 있게 여기지 않으면 미래를 위한 투자보다 당장의 보상 선택   4.강화 학습 에이전트의 성능은 어떻게 측정?   * 얻은 보상을 모두 더한다!   5.신용 할당 문제? 언제 발생? 어떻게 감소시키는가?   * 신용할당문제란: 강화학습을 진행 할 때 보상밖에 알 수 없다. 보상은 드물고 지연되어 나타나기 때문에 행동이 보상에 얼마나 기여했는지 알 수 없다. * 감소시키는 법: 에이전트에 단기 보상 제공(행동 후 각 단계마다)   6.재현메모리를 사용하는 이유  메모리에 데이터 저장 후 랜덤하게 뽑아 학습 진행   * 에이전트가 하나의 지역에 오래 머무를 경우 -> 지역 특화 발전이 일어남 * 가장 최근의 학습만 사용하되, 에이전트는 최근 경험과 최근이 아닌 경험들을 담은 버퍼에 기초하여 학습해야 함   7.오프폴리스RL 알고리즘?   * Off-policy: 학습하는 polilcy와 행동하는 policy가 반드시 같지 않아도 학습 가능 * On-policy: 학습하는 polilcy와 행동하는 policy가 반드시 같지아야만 학습이 가능   **KAIST [Digital Big Bang, Metavers Technology] 국제포럼**  **[배기동 – KT 엔터프라이즈 메타버스 원팀]**   * KT에서 내린 메타버스의 정의: 현실 세상과 가상 세상을 디지털화하여 사회, 경제, 문화 활동을 통해 가치를 창출하는 것. * 여기서 가상 세상은 현실 세상을 복제하고 현실과 융합함 * 메타버스의 구성 요소: 가상 공간, 아바타, 경제 사회 활동 * Internet vs Metaverse: 사용편의성, 상호작용성, 확장성 측면에서 메타버스가 우월함. * 메타버스의 도입 유형 * 1) In-Platform type: 틀이 구비되어 있는 형태 ex.제페토 * 2) Micro Metawave type: 개별 목적에 의해 구현됨 * 메타버스의 미래: MZ세대가 디지털 점유율이 높기 때문에 전망이 좋음. * KT의 메타버스 활용(홀로그램) * 홀로그램 콘서트나 홀로그램 회의 등으로 사용된 사례가 있음.   텍스트, 사람, 활을사용하는현악기이(가) 표시된 사진  자동 생성된 설명   * LF 홀로그램 연구 중 * 메타버스는 기존의 인간, 공간, 시간의 개념에 완전히 벗어난 것이므로 이에 맞춰 마케팅이 변할 것임.   **[Frank Steiricke 박사]**   * 혼합 현실, 자연계와 인공계가 혼합된 공간으로 변해감 * 컴퓨터 사이언스 도한 혼합공간으로 진화하고 있음 * 전통: 연산 밑 데이터의 암호화 + 비트로 0과 1 구별 * 요즘: 양자컴퓨팅, 이것의 양자비트(큐빅)이 데이터를 암호화 + 0과 1, 또는 중첩 상태를 나타냄 * AI는 인간의 지능을 모방한 것 * Ex. 자동주행시스템, 로봇틱스텍스트, 장면, 길, 도로이(가) 표시된 사진    자동 생성된 설명 * 자동주행시스템은 AI가 인터넷을 통해 사물과 소통하는 것 * MR(Mixed Reality): 혼합현실기능 * 홀로렌즈를 통해 가상 환경, 가상 물테를 느낄 수 있음 * 보통 여기서의 agentsms 의인화되기 때문에 교육 측면에서의 잠재력이 있음. 사람, 실내, 젊은, 가장이(가) 표시된 사진    자동 생성된 설명 * 그래픽 튜링테스트는 컴퓨터가 만든 환경을 우리가 컴퓨터가 만들었는지 실제 환경인지 구분할 수 없다는 것 * 컴퓨터의 성능이 1000배 향상된다면 가상과 실제를 구분하지 못함   **[Macro Tempest – MIT 디력터, NASA 추진랩 연구원의 크리에이티브 컨설턴트]**   * 마술과 기술은 사람들의 이목을 끈다는 것에서 공통점이 있는데 마술에 비해 기술은 10만 km 떨어져 있는 사람의 이목도 끌어야 한다. * 청중의 관심을 끌 지금까지 경험하지 못했던 기술 + 관심을 유지할 수 있는 공간, 이 공간이 메타버스이다.   마술과 기술의 차이점   * 마술: 심리학 기반 * 기술: 심리학 + 기술   **[김범주 – Unity Korea]**   * 메타버스의 다양한 정의 중 2가지 요소 * virtual space * virtual economy * 메타버스는 이 둘을 적절히 잘 결합해야 함 * 핵심은 virtual space(가상 공간) – AR, VR 통해 전부터 접하기는 했지만 현재는 인터넷 자체가 가상공간이 됨. * 메타버스의 교육에 대한 활용 * 에드가 데일의 경험의 원추 * 목적의식이 있는 직접 개념이 추상적 경험을 문자로만 보여주는 것보다 효과적이다 (-에드가 데일-) * 교육에 사용된 사례: 호두잉글리쉬 – 게임 시스템과 그 이점을 이용하여 가상 세계에서의 교육에 관한 통찰력을 제공한 프로젝트 | | |
| **과제할당** |  | | |
| **특이사항** | 없음 | | |
| **비고** | 19.1, 19.5 참고자료 깃에 업로드 되어있음. | | |