동아대학교 DONG-A UNIVERSITY

[실증적SW개발프로젝트]

RLHF기반 로봇 팔 제어 프로그램 개발

2143841 권은주
1824751 진현석
2051505 조현진

Contents

주제: 고차원 환경에서 강화학습 알고리즘 RLIF와 RLHF의 성능 비교 연구

Table1 . 알고리즘 비교

알고리즘	피드백 유형	Optimal policy	학습 효과
RLIF	인간의 직접적인 개입	필요	복잡한 task 수행 가능 그러나, 인간이 행동에 대한 명확한 지식이 없을 때 학습 제한
RLHF	Trajectory 사이의 선호도	불필요	최적 행동을 몰라도 상대적으로 더 나은 trajectory 선택 가능 학습에 제한 X

목표: 고차원 환경에서 **RLIF**와 **RLHF**의 성능을 비교하여, 복잡한 작업 수행 시 **RLHF**의 상대적 이점을 증명

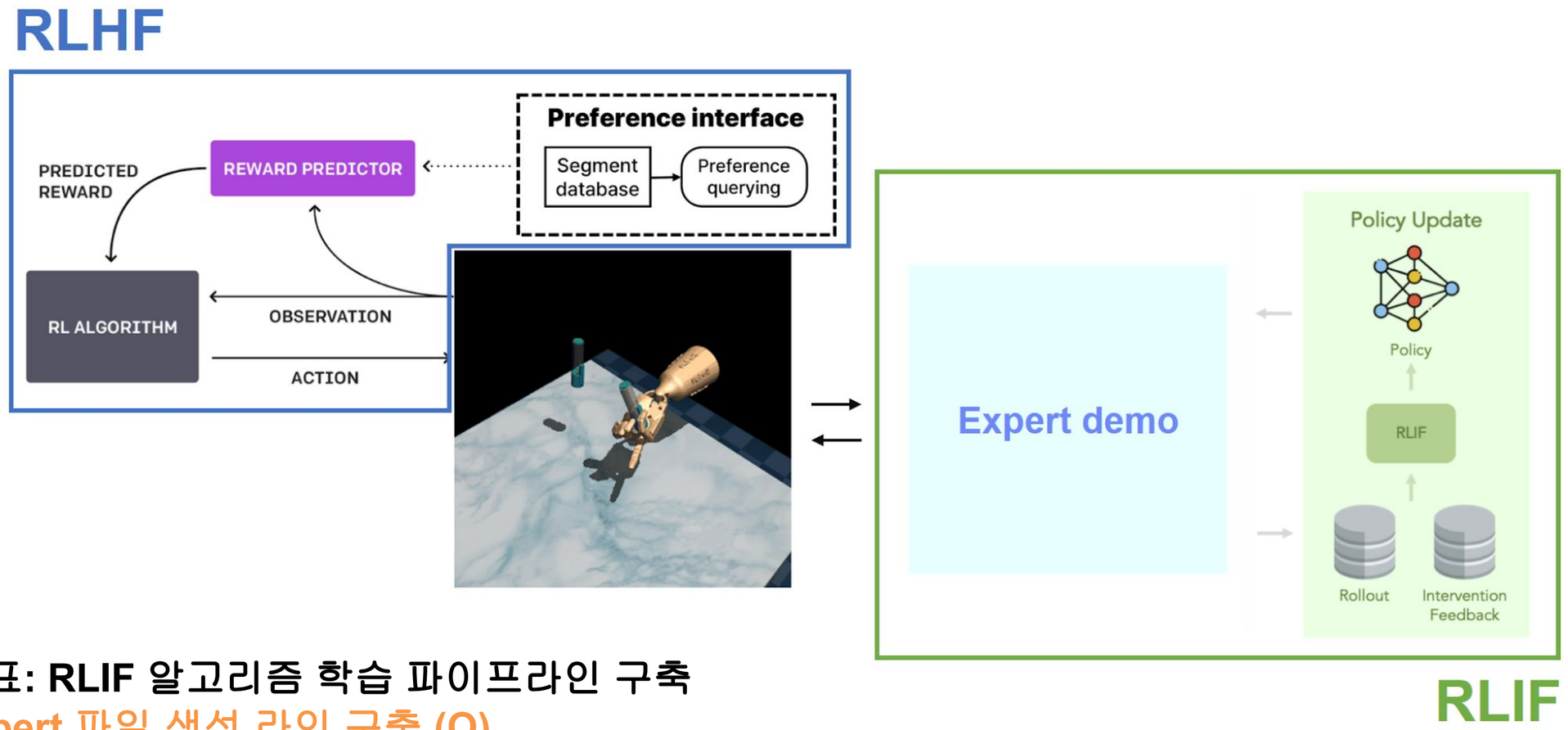
이유:

1. 알고리즘의 특성을 살펴보았을 때 **차원**이 **복잡**해질수록 **RLIF보다 RLHF가 더 좋은 성능**을 **나타낼 것으로 예상**
2. **RLHF**는 (RLIF 실험 환경 수준의) **복잡한 환경에서 실험 결과 부재**

방법:

1. 동일한 고차원 환경 설정
2. RLIF와 RLHF 알고리즘 적용
3. 복잡한 작업 수행 능력 비교

최종 결과물: 동일한 환경에서 **RLIF, RLHF** 알고리즘 비교가 가능하도록 코드 개발



1학기 목표: RLIF 알고리즘 학습 파이프라인 구축

a. **expert 파일 생성 라인 구축 (O)**
b. RLIF 알고리즘과 새로운 expert file 연동(△)
c. **학습 결과 그래프 생성 (O)**
d. **RLIF 학습 모델 생성 (O)**
e. 학습 결과 GUI 생성

실증적AI프로젝트 금주 활동내역

주제: RLHF를 이용한 협동 로봇 제어 프로그램 개발

금주 활동계획	1. RLIF 코드 분석 내용 정리 2. Expert file 생성 방법 조사 3. RLIF 알고리즘 적용		
	팀장 (권은주)	팀원 1 (조현진)	팀원 2 (진현석)
금주 개인별 활동내역	1. RLIF 코드 분석 내용 정리 (4/30~5/3) a. 코드 분석 내용 정리 2. RLIF 알고리즘 적용 (5/4~5/5)	1. RLIF 코드 분석 내용 정리 (4/30~5/3) a. 코드 구조도 생성 2. Expert file 생성 방법 조사 (5/4~5/5)	1. RLIF 코드 분석 내용 정리 (4/30~5/3) a. 코드 구조도 생성 2. Expert file 생성 방법 조사 (5/4~5/5)
차주 활동계획	1. Expert file 생성 2. Ardoit hammer environment RLIF 알고리즘 적용		

AI_Project_CoRLHF / RLIF / RLIF 코드 분석.pdf

eunjuyummy Add files via upload 049a8a1 · last week History

2.4 MB Code 55% faster with GitHub Copilot

RLIF 코드 분석 자료

2024.04

알고리즘 실행 코드

코드는 Parsing해서 사용

```
python3.8 -m RLIF.examples.train_rlif_main \   # RLIF.examples.train_rlif_main 모듈 실행
    --env_name "pen-expert-v1" \   # env_name: pen-expert-v1실행
    --sparse_env 'AdroitHandPenSparse-v1' \  # spare_env 버전으로 환경 설정
    --dataset_dir 'ENTER DATASET DIR' \ # data set 위치
    --expert_dir 'ENTER EXPERT DIR' \  # expert dir 위치
    --ground_truth_agent_dir 'ENTER GROUND TRUTH AGENT DIR' \ # 기준이 되는 agent 위치
    --logging.output_dir './experiment_output'  # --logging.output_dir 부분에서 에러 발생 제외하고 실행
```

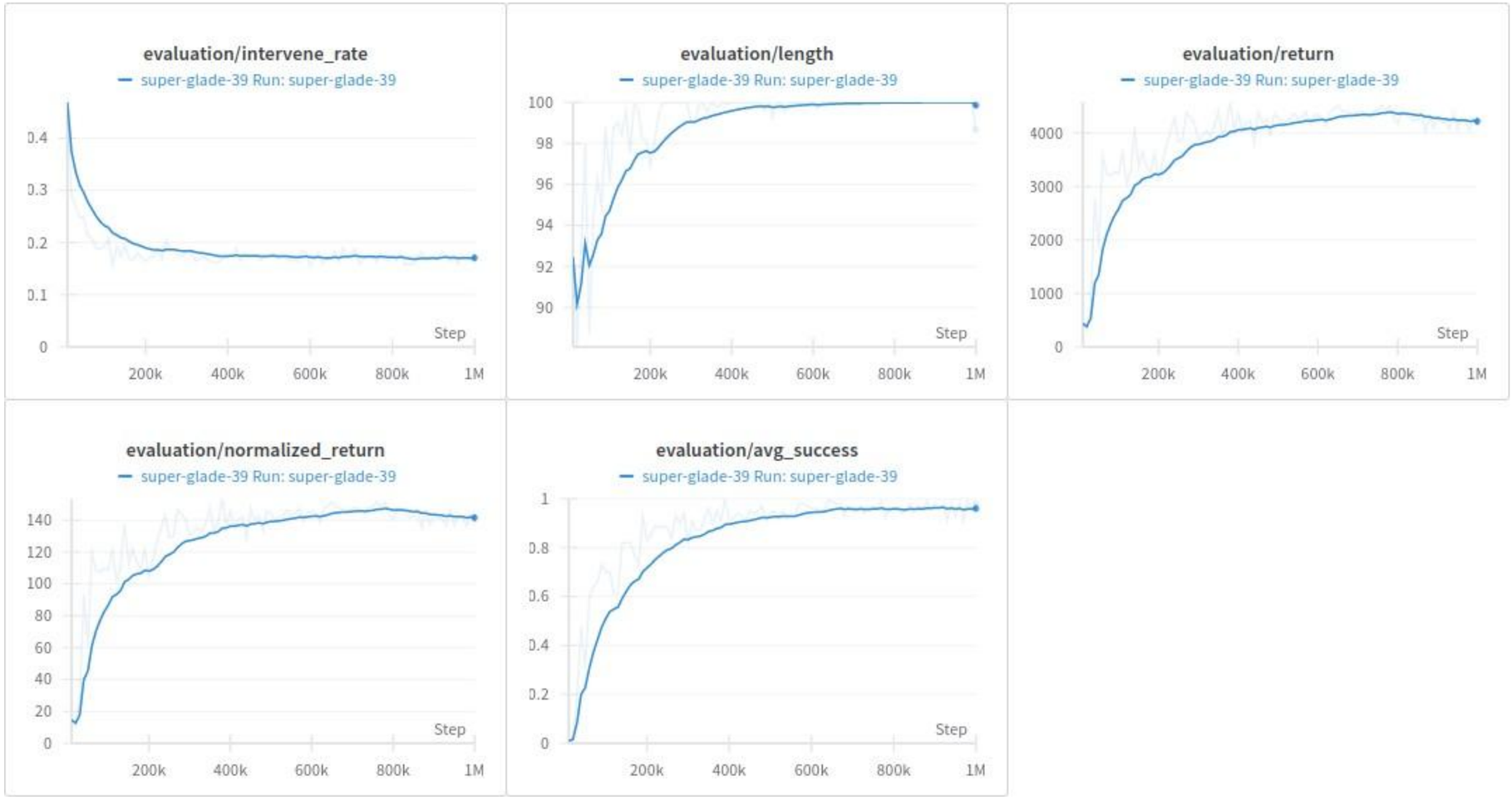
AdroitHandPen-v1 환경에서 RLIF 학습 그래프

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}$$

Return: 현재 시간 t에서 미래에 받을 reward의 합을 계산

약 200,000 Step 이후로는 안정적인 return 값

AdroitHandPen-v1 환경에서 RLIF 평가 그래프



Offline Reinforcement Learning을 위한 **Datasets** Minari

- **종류**
  a. human data: 인간의 25개 episode (시작부터 끝까지 환경과 에이전트가 상호작용을 완료한 시나리오) 시연
  b. expert data: Fine-tuning한 RL 학습 데이터
  c. cloned data: human, expert 데이터를 통해 학습한 policy (다음주 논문을 통해 자세하게 설명)

1 . hammer-human-v1

2 . hammer-expert-v1

3 . hammer-cloned-v1

05. Expert Agent

알고리즘 학습 시 Expert datasets을 그대로 사용하지 않고, 데이터를 사용해 **expert agent**를 만들어서 사용함.

- IQL 알고리즘을 이용해 Expert datasets을 학습한 Expert Agent model 생성
- 학습된 model의 평가 결과 영상 확인 → **50_000번 반복 학습한 모델 expert agent로 선정**

1 . hammer-expert-agent 학습 반복 10_000번

2 . hammer-expert-agent 학습 반복 50_000번

실증적AI프로젝트 금주 활동내역

주제: RLHF를 이용한 협동 로봇 제어 프로그램 개발

금주 활동계획	1. Learning Complex Dexterous Manipulation with Deep Reinforcement Learning and Demonstrations 논문 정리 2. DAPG project github 내용 정리 3. Hammer RLIF 알고리즘 학습		
	팀장 (권은주)	팀원 1 (조현진)	팀원 2 (진현석)
금주 개인별 활동내역	1. Hammer expert model RLIF 알고리즘 적용 (5/13~5/15) a. GPU 에러 해결 2. 학습 결과 GUI 생성 (5/18~5/19)	1. DAPG 논문 정리	1. DAPG project github 내용 정리
차주 활동계획	1. Expert file 생성 2. Ardoit hammer environment RLIF 알고리즘 적용		

Questions & Answers

Dept. of AI, Dong-A University

권은주 (kkkoj4284@donga.ac.kr)

진현석 (cpu132465@donga.ac.kr)

조현진 (gkfkgkdh@naver.com)

Github (https://github.com/eunjuyummy/AI_Project_CoRLHF)