새로운 강화학습 환경에서의 부분 관측 마르코프 결정 프로세스 학습

양현서°, 장병탁 (Biointelligence Lab)

I. 서론

강화학습은 자연어 처리, 이미지 인식, 로봇 제어 등 다양한 분야에서 뛰어난 성능을 보이며 큰 주목을 받고 있다. 그러나 강화학습의 난제 중 하나는 부분 관측 마르코프 결정 과정(Partially Observable Markov Decision Process, POMDP)[1] 에서의 학습이다. POMDP 환경에서는 에이전트가 완전한 정보를 얻을 수 없으며, 이로 인해 학습 과정이 복잡해지고 어려워진다.
이러한 배경에서 본 논문은 마인크래프트를 활용한 새로운 강화학습 환경을 제안한다. 마인크래프트는 인기 있는 샌드박스 게임이다. 이 게임은 자유도가 높아 다양한 태스크를 생성하고 실험하는 것이 가능하다. 마인크래프트를 활용한 강화학습 환경은 이미 몇 가지가 존재하는데, 이 환경들은 한계가 있다. 본 논문에서 제안하는 새로운 강화학습 환경은 최신 버전의 마인크래프트를 지원하며, 다양한 관측 공간을 제공하여 POMDP 문제를 보다 효과적으로 해결할 수 있다. 이 환경은 다양한 기능을 지원하며, POMDP의 특성을 가지고 있어, 에이전트가 부분적인 정보만을 바탕으로 학습해야 한다. 본 논문에서는 이 새로운 환경에서의 실험을 통해 POMDP 및 희소 보상 설정에서의 강화학습의 가능성을 탐색하고자 한다.

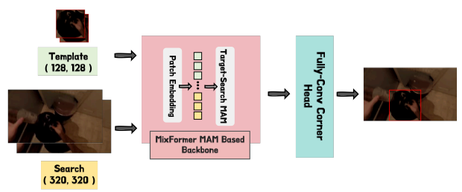

Figure 1: Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do.

II. 새로운 강화학습 환경 CraftGround

i. 기존 마인크래프트 기반 강화학습 환경

- MineDojo[2] : 마인크래프트 환경에서의 강화학습을 위한 플랫폼
- MineRL[3] : 마인크래프트 환경에서의 강화학습을 위한 플랫폼

ii. 기존 환경과의 차별점

- 최신 버전의 마인크래프트 지원, 업데이트 용이
- 고성능 (300TPS): protobuf와 unix domain socket 사용
- 다양한 관측 공간 제공: 소리, 통계, 사망 원인 등

Lorem ipsum dolor sit.	Lorem ipsum.	Lorem ipsum.
Lorem ipsum dolor.	Lorem ipsum.	α
Lorem ipsum.	Lorem.	β
Lorem.	Lorem.	γ
Lorem ipsum.	Lorem ipsum dolor.	θ

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do.

$$\begin{pmatrix} 1 & 2 & \dots & 8 & 9 & 10 \\ 2 & 2 & \dots & 8 & 9 & 10 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 10 & 10 & \dots & 10 & 10 & 10 \end{pmatrix} \tag{1}$$

III. 실험

i. 환경

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre.

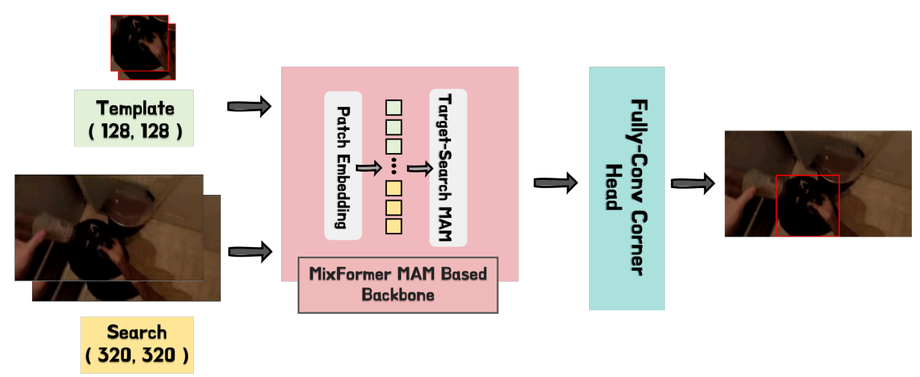

Figure 2: Lorem ipsum dolor sit amet, consectetur adipiscing elit.

ii. 입력의 인코딩

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri.,

- Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do.,

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri.

iii. 모델

IV. 결과

V. 결론

Bibliography

[1] M. Hausknecht and P. Stone, “Deep Recurrent Q-Learning for Partially Observable MDPs,” CoRR, 2015, [Online]. Available: http://arxiv.org/abs/1507.06527

[2] L. Fan et al., “MineDojo: Building Open-Ended Embodied Agents with Internet-Scale Knowledge,” CoRR, 2022, [Online]. Available: https://arxiv.org/abs/2208.xxxxx

[3] W. Guss et al., “MineRL: A Large-Scale Dataset of Minecraft Demonstrations,” CoRR, 2019, [Online]. Available: http://arxiv.org/abs/1907.13440