

Deep Reinforcement Learning for Drone Delivery in Unity

Creative IT Design IV : Jinwoo Kim | Prof. Yu Son-Cheol

Abstract

- 목적: Unity Drone Delivery 환경 안에서 빠르고 안전하게 물품 배송
- 방법: 강화학습 알고리즘인 SAC (Baseline), JNN, PPO 적용
- 결과: State, action space를 최소로 한 PPO 알고리즘이 빠른 학습 속도, 높은 배달 성공률을 보임

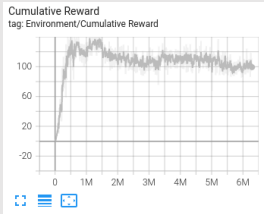
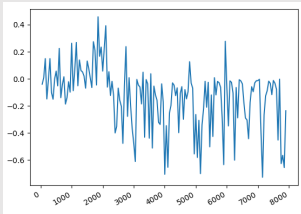
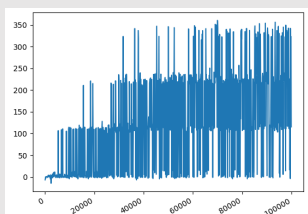
Introduction

- 최근 코로나로 인해 비대면 서비스를 원하는 고객이 많아지면서 드론 배달 관련 산업군 확장
- 드론이 물건을 안전하게 배달하려면 장애물을 인식하고 회피하는 경로 계획이 필수적
- Drone Delivery Environment
- Drone Delivery Challenge에서 제공하는 Unity 환경 사용
- 물류창고에서 출발하여 10개 중 임의로 지정된 3개의 집에 배송해야 하며, 다양한 장애물이 있다.

Related Works

1. Drone 강화학습 (AirSim)
 - Drone delivery: Image, scalar 정보를 함께 활용한 JNN이 100%의 성공률을 보임
→ 강화학습 알고리즘으로 JNN 구현
 - JNN은 학습 시간이 오래 걸린다는 한계 (비실용적)
→ On-policy, data-efficient한 PPO 구현
2. SAC vs PPO in continuous action task
 - HalfCheetah, walker 등 continuous locomotion task에서 SAC이 높은 성능을 보임
→ Baseline 알고리즘으로 SAC 선택

Implementation, Result

	Baseline (SAC)	JNN	PPO
State	Drone, lidar, camera	Drone, lidar, camera	Drone, lidar
Action	Continuous	Discrete	Discrete
Reward	Event, distance	Event, distance, avoidance	Event, distance, avoidance
Result	 <p>Local Minima에 수렴하여 최대 1개의 물품만 배송</p>	 <p>학습 시간이 오래 걸려 기간 내에 학습 X (0개 배송)</p>	 <p>빠르고 효과적으로 학습 성공률: 23/30 (76.67%)</p>

Conclusion

- PPO는 다른 알고리즘에 비해 적은 episode로 학습하여 실제 드론에도 적용할 수 있을 것이라 기대된다.
- PPO의 batch size, network size를 크게 하고, sparse reward에 적합한 학습 방법을 사용하면 좀 더 안정적인 배송이 가능할 것이다.
- 드론의 움직임 제어, 안정적인 착지가 가능하면 실제 드론에도 적용할 수 있을 것이다.