제한 시간을 갖는 작업 환경에서 강화학습 기반 드론의 자율비행 비교 연구

권은주, 정희철, 김현석*
동아대학교 컴퓨터·AI공학부
kkkoj4284@donga.ac.kr, 1923152@donga.ac.kr, *hertzkim@dau.ac.kr

A Study of Autonomous Drone based on Reinforcement Learning for Time-Limited Task Environments

Kwon Eun Ju, Jung Hee Chul, Hyunseok Kim
Division of Computer and AI, Dong-A Univ.

요 약

본 논문은 혈액 운송과 같은 제한 시간을 갖는 작업 환경에서 드론의 자율비행에 관한 것으로, 기존 Proportional-Integral-Derivative (PID) 방식에 비해 강화학습(Reinforcement Learning; RL)을 이용하는 경우의 성능 향상을 비교 실험을 통해서 제시한다. 오픈소스 드론 시뮬레이터를 통해 PID 기반 드론의 자율비행보다 RL 기반 드론의 평균 비행시간이 약 2.8초 빠른 것을 확인할 수 있었다. 이러한 결과를 바탕으로 짧은 시간 내에 목표 달성이 필요로 하는 작업 환경에서 강화학습이 효과적으로 사용될 수 있을 것으로 기대한다.

Ⅰ. 서 론

지금까지 드론의 자율비행을 위해서 시스템 모델링과 수치 계산을 통한 Proportional-Integral-Derivative (PID) 방식이 널리 사용됐다 [1][2]. PID 제어는 비례, 적분, 미분에 해당하는 3개의 parameter를 구하고, 이를 이용하여 제어하는 방식이다. PID는 안정적인 제어가 가능하다는 장점을 가지고 있어 많은 분야에서 사용되고 있다. 하지만, 환경 변화를 예측하기 힘든 공기 중을 자유롭게 비행해야 하는 드론의 경우, 특히 혈액 운송처럼 짧은 시간 내에 수행되어야 하는 작업환경에서는 전통적인 PID 방식이 적합하지 않음을 보여왔다 [3]. 따라서, 본 논문에서는 환경과 상호작용을 통해 비행 기술을 배우는 강화학습 (Reinforcement Learning; RL) 방식을 이용한 드론의 자율비행을 통해 이를 해결하고자 한다. 제한 시간을 갖는 작업환경에서 PID 기반 드론의 자율비행과 RL 기반 드론의 자율비행을 비교하여 검증하고자 한다. 이를 위해, 본 논문에서는 오픈소스인 gym-pybullet-drones [4] 시뮬레이터를 이용하여 PID 방식과 RL 방식에 따른 드론의 자율비행 성능을 목표 지점 도달 시간 비교를 통해 제시한다.

Ⅱ. 본론

1. Proportional-Integral-Derivative (PID)

PID 제어 방식은 오차항, 오차항의 적분항, 오차의 미분항으로 이루어져 있으며, 각 항에 곱해지는 가중치인 제어 이득을 구동 환경에 적합하게 설정하여 제어하는 방법이다. PID 제어는 목푯값을 넘어선 제어량인 오버슈트가 필연적으로 발생하게 되며, 환경별 PID 제어 파라미터 튜닝을 통해 오버슈트 및 안정성을 향상한다 [5].

2. Reinforcement Learning (RL)

강화학습은 에이전트(agent)가 환경(environment)과 상호작용을 통해 주어진 상태(state)에서 얻을 수 있는 보상(reward)을 최대화할 수 있는 행동(action)을 학습하는 알고리즘이다 [6].

본 논문에서 사용한 Proximity Policy Optimization (PPO) [7] 알고리즘은 대표적인 on-policy 알고리즘 중 하나로, 안정적인 강화학습을 위해 클리핑 기법을 사용하는 것이 특징이다. 클리핑 기법을 사용하여 정책 업데이트 과정에서 발생할 수 있는 파라미터의 급격한 변화를 방지하고, 샘플을 효율적으로 이동시켜 정책을 최적화하는 것에 중점을 두고 있다. 이러한 특성으로 인해 다양한 환경에서 효과적으로 적용될 수 있다. 이러한 알고리즘의 특성이 빠른 이동을 위해 빠른 계산이 필요한 드론 비행에 적합하다고 판단해 PPO 알고리즘을 이용해 실험을 진행하였다.

Ⅲ. 실험

1. 실험 환경

1) 드론 시뮬레이터

본 논문에서 사용한 그림 1의 gym-pybullet-drones는 Pybullet Physics 엔진을 기반으로 작성된 다중 쿼드콥터를 위한 시뮬레이터로 오픈소스 OpenAI Gym과 호환된 환경이다. 이 환경에서는 제한 시간을 갖는 작업 환경을 구축하고 PID 방식과 RL 방식으로 드론의 자율비행을 구현할 수 있다. [8].

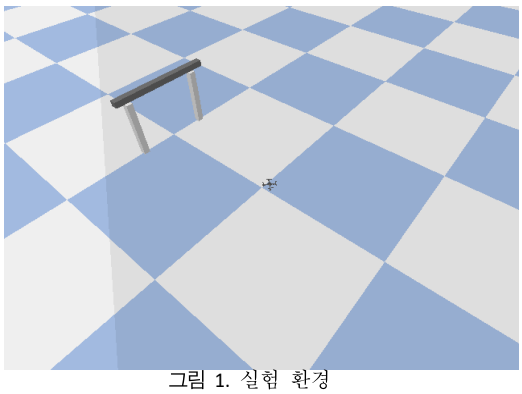

그림 1. 실험 환경
Fig. 1. Experimentation Environment

2) 실험 조건

i) 드론의 출발 지점은 환경의 $x, y, z = (0,0,0)$ 지점을 기준으로, $10cm \times 10cm$ 공간 내에서 무작위로 설정된다.

ii) 드론의 목표는 골대를 통과하는 것이며, 골대의 범위는 수식(1)과 같다.

$$0.95 \leq x \leq 1.15 \quad -0.27 \leq y \leq 0.27 \quad z \geq 0.5 \quad (1)$$

iii) 한 에피소드 당 최대 2400 step 이동이 가능하며, 시간 내에 목표를 통과하면 성공, 통과하지 못하면 실패한 것으로 간주한다.

iv) 보다 빨리 목표 지점에 도달하는 것이 수행해야 하는 작업환경이다.

2. 실험 결과

1) 강화학습 모델 학습 과정

그림 2의 에이전트 평균 에피소드 그래프를 확인하면 학습 시간이 증가함에 따라 평균 에피소드 진행 시간이 줄어드는 것을 확인할 수 있다. 이는 강화학습 모델의 학습이 제대로 진행되고 있다는 것을 의미하며, 목표 지점에 도달하기까지의 시간이 점차 줄어들고 있다는 것을 의미한다. 그림 3의 에이전트의 평균 보상 그래프를 확인하면 학습 시간이 증가함에 따라 에이전트가 받는 평균 보상이 증가하는 것을 확인할 수 있다. 이는 에이전트가 에피소드를 진행하는 동안 적절한 보상을 받고 있음을 뜻하며 에이전트가 목표를 제대로 수행하고 있다는 것을 의미한다.

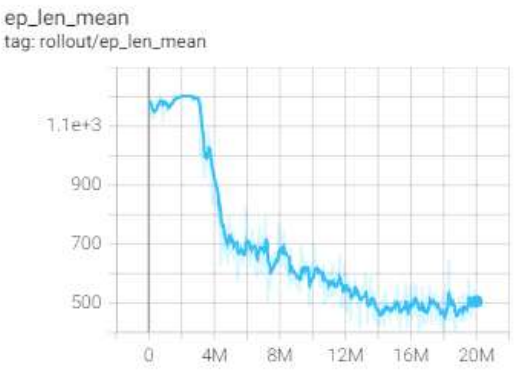

그림 2. 에이전트의 평균 에피소드
Fig. 2. Average episode for an agent.

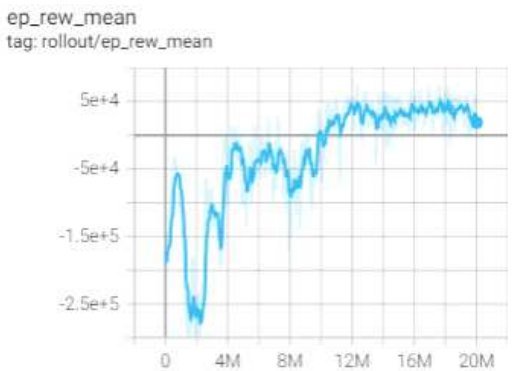

그림 3. 에이전트의 평균 보상
Fig. 3. Average reward for an agent.

2) PID와 RL 모델 테스트 결과

PID 기반 제어 드론의 평균 비행시간은 약 5.1초이고, RL 기반 드론의 평균 비행시간은 약 2.3초이다. 결론적으로 RL 기반 드론의 비행시간이 약 2.8초 더 빠르게 목표 지점에 도달한다는 것을 알 수 있다. 또한, 각 제어 방식에 따른 드론 비행 형태의 경우, PID 기반 드론은 상대적으로 느리지만 안정적으로 비행하는 반면, RL 기반 드론은 빠르지만 불안정하게 비행하는 경향이 있는 것을 확인할 수 있었다.

Ⅲ. 결론

PID와 RL 간의 제어 특성 및 제어 방식에 따른 목표 지점 도달 시간을 확인 한 결과, 제한 시간 내 목표를 달성해야 하는 작업환경에서는 RL 기반 드론 비행이 PID 기반 드론 비행보다 평균 약 2.8초 빠른 비행시간이 가능함을 확인할 수 있었다. 이러한 결과를 통해, 혈액 운반과 같은 짧은 시간 내에 목표 달성이 필요로 하는 작업환경에서 강화학습 방식이 더 효과적으로 사용될 수 있을 것으로 기대한다.

ACKNOWLEDGMENT

이 논문은 2024년도 정보(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2023-0-00076, SW중심대학(동아대))

참 고 문 헌

[1] Yoon, Dan-Bee, et al. “A Study on Flight Stabilization of Drones by Gyro Sensor and PID Control.” The Journal of the Korea Institute of Electronic Communication Sciences, vol. 12, no. 4, 한국전자통신학회, Aug. 2017, pp. 591 - 598, (doi:10.13067/JKIECS.2017.12.4.591).

[2] Oh, Ji-Wan, et al. “Drone Hovering Using PID Control.” The Journal of the Korea Institute of Electronic Communication Sciences, vol. 13, no. 6, 한국전자통신학회, Dec. 2018, pp. 1269 - 1274, (doi:10.13067/JKIECS.2018.13.6.1269).

[3] 장진명, 손동훈, 김화중. "드론 활용 혈액 운송시스템의 효과 분석." 로지스틱스연구 29.3 (2021): 57-68.

[4] https://github.com/utiasDSL/gym-pybullet-drones

[5] 김성호, 김도국. "목적 함수 변경을 통한 유전 알고리즘을 이용한 쿼드콥터 PID 제어이득 최적화 성능 개선." 한국정보과학회 학술발표논문집 (2023): 1811-1813.

[6] Kaelbling, Leslie Pack, Michael L. Littman, and Andrew W. Moore. "Reinforcement learning: A survey." Journal of artificial intelligence research 4 (1996): 237-285.

[7] Schulman, John, et al. "Proximal policy optimization algorithms." arXiv preprint arXiv:1707.06347 (2017).

[8] Panerati, Jacopo, et al. "Learning to fly—a gym environment with pybullet physics for reinforcement learning of multi-agent quadcopter control." 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2021.