다중 목적지 드론 배송의 경로 및 배터리 최적화를 위한 강화학습 모델 연구

김민수, 김준엽, 김대한, *김준영, 정소이

아주대학교 전자공학과, 아주대학교 AI 융합네트워크학과*

{andykim000, kjohn0714, koreakdh99, *junzero0615, sjung}@ajou.ac.kr

Reinforcement Learning-based Optimization of Battery and Route Efficiency in Multi-stop Drone Delivery

Minsoo Kim, Junyeop Kim, Daehan Kim, Junyoung Kim, Soyi Jung

Dept. of Electrical and Computer Engineering, Ajou University,
*Dept. of Artificial Intelligence Convergence Network, Ajou University

요 약

본 연구는 기존의 단일 목적지 드론 배송 방식을 개선하여 다수 목적지에 대한 효율적인 드론 배송 최적화를 목표로 한다. 이를 위해 강화학습 기법 중 하나인 PPO(Proximal Policy Optimization)를 사용하여 드론의 물리적 움직임과 배터리 소모량을 최적화하고, 다양한 환경 조건(예: 날씨 및 배터리 소모)에 따른 드론의 경로 선택 및 배송 효율성을 극대화하고자 한다.

Ⅰ. 서 론

드론을 활용한 물류 배송 시스템은 빠른 속도와 높은 효율성 덕분에 물류 산업의 새로운 대안으로 주목받고 있다. 특히, 도심 지역이나 접근이 어려운 지점에서 효과적으로 활용될 수 있어 큰 잠재력을 지니고 있다. 최근 연구에 따르면, 드론 배송 시스템의 경로 계획에 관한 다양한 알고리즘이 개발되고 있으며, 이들은 물류 효율성을 극대화하는 데 기여하고 있다[1]. 또한, 드론 물류 배송의 활용 사례가 늘어나고 있으며, 기술적 한계와 발전 방향에 대한 논의가 활발히 진행되고 있다. 그러나, 현재 드론 배송 기술은 안전성, 규제, 그리고 효율성 문제와 같은 여러 도전 과제에 직면해 있으며, 이러한 문제를 해결하기 위한 연구가 지속적으로 이루어지고 있다.

본 논문은 차세대 물류 시스템으로 주목받고 있는 무인 드론 배송 시스템 시뮬레이션을 진행하기 위해 강화학습 알고리즘 근접 정책 최적화(proximal policy optimization, PPO)기법을 활용하여 다중 목적지 배송의 최적 경로를 탐색하고 드론의 배터리 소모를 최소화하는 시뮬레이션을 진행한다. 이를 통해 드론의 물리적 제약과 다양한 환경 조건을 고려한 최적화된 배송 경로를 찾고, 배송 효율성을 극대화하여 차세대 무인 드론 배송 시스템 모델을 제안한다.

Ⅱ. Unity 기반 시뮬레이션 환경 구축

본 논문에서는 unity 3D 엔진을 활용하여 시스템 모델을 설계하였다. Unity는 실제 환경과 유사하게 중력, 충돌, 마찰 등 다양한 물리적 요소를 자연스럽게 재현하여 객체의 동작을 정밀하게 모사할 수 있으므로 강화학습 시뮬레이션 도구로 적합하다. 특히, Unity asset store를 이용하여 여러 환경과 에이전트 등 다양한 플러그인을 지원하며 이를 쉽게 추가할 수 있다는 장점이 있다. 본 연구에서는 그림 1 (a)와 같이 Russian Buildings Pack과 PA_Drone_Pack을 활용하여 아파트 단지 환경과 무인 이동체(unmanned aerial vehicle, UAV) 기반의 시뮬레이션 환경을 구축하였다. UAV는 택배 허브를 거쳐 목적지까지 안전하게 배송한 후 복귀할 수 있도록 설계되었으며, 이를 위해 그림 1 (b)와 같이 ray perception sensor 3D가 장착되어 충돌 회피가 가능하게 하였다. 이 센서는 실시간으로 여러

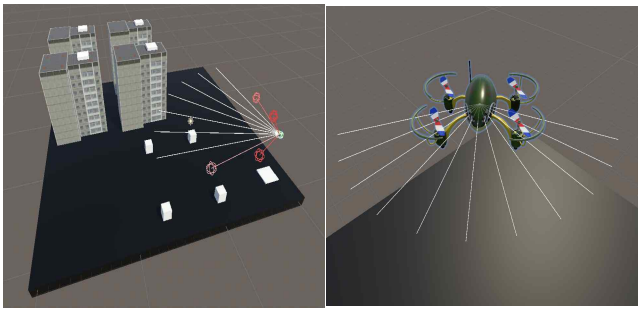

그림 1 (a). 전체 Scene　　그림 1 (b). Drone Agent

그림 1. Unity 시뮬레이션 환경

방향으로 ray를 발사하고, 각 ray에서 받는 정보를 기반으로 UAV의 이동 경로 상에 있는 물체를 감지하여, 장애물을 피하며 비행할 수 있도록 한다. 각 건물과 택배에는 tag를 부여하여, UAV가 외벽을 회피하고 옥상에 있는 택배 보관소 및 택배 위치를 인식하여 정확히 도착할 수 있도록 설정하였다. 태그 시스템은 UAV가 목표 지점을 인식하고 정확하게 반응하는 데 필수적이다. 센서는 UAV의 전방 기준으로 좌우에 장착되었으며, 각 ray의 길이와 간격은 균일하게 설정하여 균형 잡힌 감지와 장애물 회피의 정확성을 높였다. 또한 UAV의 초기 목적지와 최종 도착지는 고정되어 있으며, 임의로 생성되는 택배 상자에 맞춰 비행 경로를 실시간으로 조정할 수 있게 설계하였다. 이를 통해 UAV는 복잡한 도시 환경에서도 효율적으로 작동하며, 다양한 환경에서 안정적으로 운용될 수 있다.

Ⅲ. 강화학습 시스템 모델

물류 배송 드론의 설계와 강화학습 환경은 Unity의 ML-Agents를 활용하여 진행한다. 강화학습의 주체가 되는 agent는 UAV로 설정하며, agent는 각 에피소드에서 취한 action에 따른 reward를 바탕으로 학습한다. 강화학습 과정은 마르코프 결정 과정(Markov decision process, MDP)을 따르며, 다음과 같이 상태, 행동, 보상 함수를 설계한다.

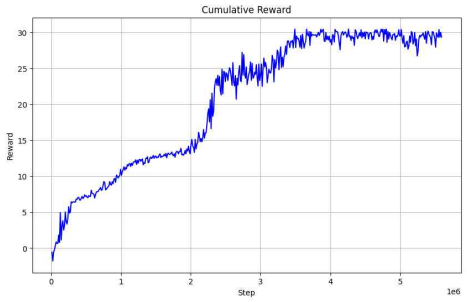

그림 2. 강화학습 누적 보상 그래프

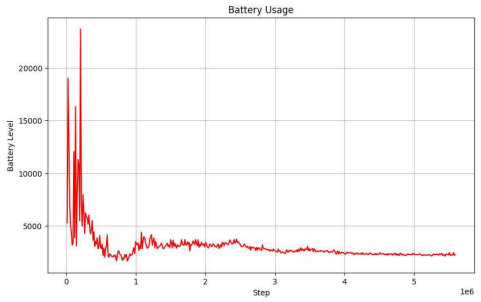

그림 3. UAV 배터리 사용량

표 1. 보상함수 파라미터

Parameter	Value
r_{pickup}	+5
r_{del}	+10
r_{arr}	+15
r_{col}	-0.5
r_{bound}	-0.5
α	+100

1) State(상태) : $S \doteq [p_t, d_t, b_t, v_t, s_t^{delivery}]$으로 정의한다. UAV(Agent)의 현재 위치(p_t), 현재 target(물류 혹은 배송지)까지의 거리(d_t), UAV 배터리 사용량(b_t), UAV 현재 속도(v_t), 그리고 UAV의 배송/복귀 상태($s_t^{delivery}$)로 정의한다.

2) Action(행동) : $A \doteq [m_x, m_y, m_z, r]$ 로 정의한다. UAV가 현재 target까지의 x축으로의 이동(m_x), y축 방향으로의 이동(m_y), z축 방향으로의 이동(m_z), 그리고 동체 회전 각도(r)로 정의한다.

3) Reward(보상) : $R = R_{suc} + R_{bat} + R_{neg}$로 정의한다. R_{suc}는 성공 보상으로 수식 (1)과 같이, agent의 작업 진행을 확인하고 동작 메커니즘을 학습하는 역할을 한다. 이는 각각 택배를 수령했을 때, 배송지에 도달했을 때, 허브로 복귀했을 때 agent에게 부여되는 보상이다. R_{bat}는 배터리 사용 보상으로 에피소드 내에서 agent가 배터리를 얼마나 효율적으로 사용하는지를 평가하며 수식 (2)와 같이 정의한다. R_{neg}는 agent가 바람직하지 않은 행동을 했을 때 부여되는 패널티로 수식 (3)과 같이 정의하며, agent가 환경에서의 행동을 유지하고, 올바른 경로를 선택하도록 유도 한다. R_{bat}의 경우, 하나의 에피소드당 배터리 사용량(battery usage)에 반비례하여, 배터리 소모를 최소화하여 보상을 최대화하도록 한다. R_{neg}의 경우 r_{col}과 r_{bound}의 합으로 나타나는데, 이는 각각 agent가 건물 외벽과 충돌했을 경우, agent가 환경을 이탈했을 경우 부여되는 페널티 보상이다.

$$R_{suc} = r_{pickup} + r_{del} + r_{arr}, \quad (1)$$

$$R_{bat} = \alpha \frac{1}{Battery\ Usage}, \quad (2)$$

$$R_{neg} = r_{col} + r_{bound}. \quad (3)$$

본 연구에서는 드론의 물리적 움직임과 환경 조건을 최적화하기 위해 강화학습 기법 중 하나인 PPO를 기반으로 학습하였다. PPO는 기존의 강화학습 기법에 비해 안정적이고 효율적으로 학습을 수행할 수 있는 알고리즘으로, 드론과 같은 복잡한 물리적 시스템의 연속 환경 행동 제어 문제를 해결하는 데 적합하다.

표 2. 강화학습 시뮬레이션 파라미터

Parameter	Value
Batch size	64
Buffer size	12000
Learning rate	0.0003
Beta	0.001
Epsilon	0.2
Lambda	0.99
Number of epochs	3

IV. 시뮬레이션 결과 분석

그림 2와 그림 3은 Unity와 ML-Agents를 활용하여 UAV 기반 다중 목적지 배송 시스템을 학습시킨 시뮬레이션 결과를 보여준다. Agent는 물류를 픽업하고, 목표 배송지에 도달한 후 복귀하는 작업을 반복하며 episode 단위로 학습을 진행했다. 그림 2는 학습이 진행됨에 따라 각 episode 종료 시의 reward를 나타낸 그래프다. 분석 결과, 2e-6 step까지는 픽업 과정, 3.5e-6 step까지는 배송지 도달 과정, 4e-6 step까지는 복귀 과정이 학습되었으며, 이후 reward 값이 약 30으로 수렴하는 것을 확인할 수 있다. 그림 3은 각 episode별 드론의 배터리 사용량을 나타낸다. 학습이 진행됨에 따라 비행 경로가 최적화되었으며, 4e-6 이후로는 배터리 사용량이 일정하게 유지되는 것을 알 수 있다

V. 결론

본 논문에서는 Unity와 Python 기반의 ML-Agents를 활용하여 다중 목적지의 배송을 목표로한 UAV의 배터리 및 경로기반 최적화를 할 수 있게끔 강화학습 시뮬레이션을 진행하였다. 보상함수 설정 및 환경 구성을 통해 Real-World에 적합한 모델을 구성하였다. 향후 연구에서는 UAV의 비행 경로와 배터리 소모뿐이 아닌 다양한 날씨요소 및 환경요소를 추가하여 보다 다양하고 현실에 가까운 상황에서 적용이 가능한 모델을 구현하고자 한다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2024-00358662)

참 고 문 헌

[1] G. Attenni, V. Arrigoni, N. Bartolini and G. Maselli, "Drone-based delivery systems: A survey on route planning," IEEE Access, vol. 11, pp. 123476-123504, 2023