

# 연합학습 환경에서의 다중 에이전트 강화학습 기반 UAV 경로 최적화 시스템

정준호, 남지희, 김창경, 이수경

연세대학교

{stkjas4657, jhee, ckim48, sklee}@yonsei.ac.kr

## Multi-Agent Reinforcement Learning-based UAV Trajectory Optimization System in Federated Learning

Junho Jeong, Jeehee Nam, Chang Kyung Kim, SuKyoung Lee

Yonsei Univ.

### 요약

기계 학습은 자율 주행과 같은 스마트시티의 지능형 서비스를 가능하게 하였으며, 이러한 서비스를 지원하기 위해 통신 효율적인 분산 학습 패러다임인 UAV(Unmanned Aerial Vehicle) 기반 연합 학습(federated learning, FL)이 활용되고 있다. 그러나 IoT(Internet of Things) 사용자의 이동성과 이질적인 로컬 학습 시간으로 인해 통신 및 비행 지연 시간을 최소화하는 UAV 경로를 결정하는 것은 여전히 도전적인 문제이다. 본 논문에서는 IoT 사용자의 이동성을 고려하여 다중 UAV의 경로를 최적화하는 다중 에이전트 강화 학습(multi-agent reinforcement learning, MARL) 기반 FL 시스템을 제안한다. 제안하는 시스템은 다중 UAV의 상호 작용을 통해 협력적으로 경로를 학습하고, 동적인 환경 변화에 효과적으로 대응할 수 있다.

### I. 서론

기계 학습의 발전은 자율 주행 및 지능형 로봇과 같은 다양한 지능형 서비스를 가능하게 하였으나, 기존의 중앙 집중식 학습 방식은 통신 혼잡과 개인정보 보호 문제를 유발할 수 있다. 이러한 한계를 완화하기 위해 제안된 연합 학습(federated learning, FL)은 IoT(Internet of Things) 사용자가 로컬 모델을 학습하고, 원시 데이터(raw data) 대신 로컬 모델의 파라미터만을 기지국(base station, BS)과 같은 중앙 서버와 교환하는 분산 학습 패러다임이다 [1]-[2].

그러나, 스마트시티에서는 BS의 제한된 통신 자원과 장애물로 인해 모든 IoT 사용자를 충분히 지원하기 어려우며, 방대한 수의 IoT 사용자로 인해 BS가 과부하될 수 있다. 이를 해결하기 위해, IoT 사용자에게 line-of-sight 통신을 제공함으로써 네트워크 연결성을 향상시키는 UAV(Unmanned Aerial Vehicle) 기반 FL이 활발히 연구되고 있다 [3]-[4].

한편, IoT 사용자의 이동성과 이에 따른 채널 조건의 불확실성은 통신 지연을 유발한다. 또한 이질적인 로컬 데이터 크기 및 IoT 사용자까지의 UAV 비행 시간으로 인해 업데이트된 로컬 파라미터를 적시에 수집하지 못함으로써 FL 학습에 필요한 지연 시간이 증가할 수 있다 [5]. 따라서 IoT 사용자의 이동성과 통신 환경 변화를 실시간으로 반영할 수 있는 분산 의사결정 방식이 요구된다.

본 논문은 IoT 사용자의 이동성과 로컬 데이터의 크기에 따른 통신 및 비행 지연을 고려하여 협력적으로 다중 UAV의 경로를 결정하는 다중 에이전트 강화학습(multi-agent reinforcement learning, MARL) 기반 UAV 경로 최적화 시스템을 제안한다.

### II. 본론

본 논문에서 제안하는 시스템은 서비스 영역 내에 무작위로 분포하여

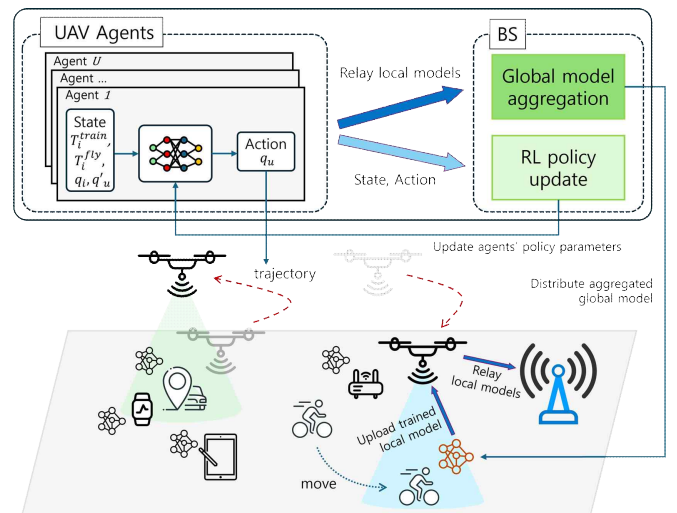


그림 1. 시스템 구성

임의의 방향과 속도로 이동하는  $I$  명의 IoT 사용자 ( $i \in \mathbf{I}$ ), 이들의 로컬 파라미터를 수집하는  $U$  대의 UAV ( $u \in \mathbf{U}$ ), 그리고 수집된 정보를 취합하여 전역 모델 업데이트를 수행하는 1개의 BS로 구성된다. 각 IoT 사용자는 고유한 데이터 크기와 컴퓨팅 자원을 보유하고 있으며, 이를 바탕으로 개별적인 로컬 학습을 수행한다. UAV는 일정한 고도에서 비행하며 IoT 사용자의 학습 완료 시점에 맞춰 경로를 조정하고 IoT 사용자의 로컬 모델 파라미터를 BS로 중계(relay)하는 역할을 한다.

다중 UAV 환경에서의 경로 최적화 문제는 여러 UAV가 IoT 사용자로부터 로컬 모델 파라미터를 수집하여 BS로 중계하는 동일한 태스크를 수행하면서도 UAV 간의 위치, 비행 경로에 따라 서로의 성능에 영향을 미치는 협력적 의사결정 문제이다. 특히 각 UAV는 제한된 비행 에너지를 가지며, 충돌 회피 및 커버리지의 중복을 방지하기 위해 다른 UAV의 행

동을 고려한 경로 결정을 수행해야 한다. 이러한 문제를 해결하기 위해선 단일 에이전트 강화학습 기반의 중앙 집중식 의사결정 방식보다 각 UAV가 독립적으로 행동하면서도 공동의 목표를 달성하는 MARL 기술이 적합하다 [6]. 또한, 다중 UAV는 전체 FL 지연 시간을 최소화하는 동일한 목적을 공유하고 공동의 보상 구조를 갖기 때문에, MARL을 통해 UAV 간 협력을 유도함으로써 최적의 경로 계획이 가능하다.

제안하는 MARL 기반 UAV 경로 최적화 시스템은 중앙 집중 학습 - 분산 실행(Centralized Training and Decentralized Execution, CTDE) 구조를 따른다. 학습 단계에서는 BS가 전역 정보를 활용하여 MARL 모델을 중앙에서 학습하며, 실행 단계에서는 각 UAV가 로컬 관측 정보만을 이용하여 독립적으로 경로 결정을 수행한다. 이러한 구조 하에서, 시스템은 FL의 각 iteration마다 다음과 같은 절차로 동작한다.

1. BS는 모든 UAV로부터 각자의 현재 위치 정보를 수집하며, 동시에 각 UAV가 서비스 하는 모든 IoT 사용자의 로컬 데이터셋의 크기, 그리고 단위 데이터 샘플을 처리하는 데 필요한 CPU cycle 수를 요청한다. BS는 수집한 정보를 바탕으로 각 IoT 사용자의 연산 자원과 데이터 규모를 고려하여 예상 로컬 모델 학습 시간  $T_i^{\text{train}}$  을 계산한다. 이후 BS는 계산된 예상 로컬 모델 학습 시간과 IoT 사용자의 위치 정보, 그리고 각 UAV의 위치 정보를 모든 UAV에게 전송한다.
2. 각 UAV는 BS로부터 공유받은 타 UAV들의 위치 정보를 바탕으로 상대적 거리를 파악하며, 비행 중 발생할 수 있는 잠재적 충돌 가능성을 경로 계획에 반영함으로써 최종적인 비행 지연 시간  $T_u^{\text{fly}}$  을 추정한다.
3. 각 UAV는 추정된 IoT 사용자별 로컬 학습 시간  $T_i^{\text{train}}$  과 자신의 비행 시간  $T_u^{\text{fly}}$  을 기반으로 IoT 사용자가 로컬 학습을 마치고 로컬 모델 파라미터 전송을 시작하는 시점에서의 위치를 예측한다. 이후 각 UAV는 예측 정보를 기반으로 다음 iteration에서의 자신의 최적 위치  $q_u$  를 MARL 정책에 따라 독립적으로 결정한다.
4. UAV가 결정된 목적지  $q_u$  로 이동하여 IoT 사용자의 로컬 모델 파라미터를 수집하여 BS로 전달한다. BS는 소요된 비행 시간과 통신 시간을 포함한 총 지연 시간을 측정하여 각 UAV 에이전트에게 총 지연 시간에 반비례하는 보상을 부여하고 상태, 행동, 보상을 버퍼에 저장한다.
5. BS는 모든 UAV로부터 수집된 IoT 사용자의 로컬 모델 파라미터를 가중 평균하여 전역 모델을 업데이트한다. 동시에 버퍼에서 미니 배치(mini-batch)를 샘플링하여 MARL 모델의 파라미터를 업데이트한다. 업데이트된 전역 FL 모델과 MARL 모델을 다음 FL iteration이 시작될 때 각각 모든 IoT 사용자와 모든 UAV에게 배포된다.
6. 전역 FL 모델의 손실 함수가 목표 임계치 이하로 수렴하거나 최대 학습 iteration에 도달하여 종료 조건이 충족되면, BS는 최종 업데이트된 전역 모델을 모든 IoT 사용자에게 전송하고 학습 프로세스를 종료한다.

제안된 MARL 기반 시스템은 IoT 사용자 이동성과 네트워크 환경 변화에 실시간으로 대응하도록 설계되었다. 이를 위해 다중 UAV의 협력적 경로 결정 과정에 FL 환경에서 발생하는 통신 및 UAV 비행 지연 시간을 구조적으로 반영한다. 이러한 구조는 반복적인 상호작용을 통해 UAV의 경로 결정 전략의 효과를 점진적으로 축적할 수 있도록 하며, 이를 바탕으로 변화하는 환경에 대해 보다 신속하고 적응적인 경로 결정을 지원하는 기반을 제공한다.

### III. 결 론

본 논문에서는 IoT 사용자의 이동성과 로컬 데이터의 크기에 따른 통신 및 비행 지연을 고려하여 다중 UAV의 경로를 결정하는 UAV 경로 최적화 시스템을 제안하였다. 제안된 시스템은 동적인 네트워크 환경에서 MARL을 활용하여 UAV 간 협력을 구성하고 최적의 UAV 경로를 결정한다. 이를 통해 FL 환경에서 다중 UAV 경로 최적화 문제를 다룰 수 있는 이론적 기반을 제시한다.

### ACKNOWLEDGMENT

본 논문은 2026년도 정부(과학기술정보통신부)의 재원으로 한국연구재단 중견연구(No. RS-2025-00573388)의 지원을 받아 수행된 연구임.

### 참 고 문 헌

- [1] Z. Liu, B. Huang, C. Zhang, Z. Yao, T. Li, Q. Sun, and Y. Li, "Clustered federated learning for energy-harvesting smart meters in p2p energy trading," *IEEE Trans. Green Commun. Netw.*, vol. 10, pp. 585 - 596, Jul. 2025.
- [2] S. Y. Song, H. J. Park, L. H. Park, "Federated Learning for IoT Networks: Applications and Challenges," in *Proc. Symp. Korean Inst. Commun. Inf. Sci.*, pp. 232-233, Gangwon, Jan. 2024.
- [3] X. Zhang, W. Liu, J. Ren, H. Xing, G. Gui, Y. Shen, and S. Cui, "Latency minimization for UAV-enabled federated learning: Trajectory design and resource allocation," *IEEE Internet Things J.*, vol. 12, no. 14, pp. 27097 - 27112, Jul. 2025.
- [4] Z. Fu, J. Liu, Y. Mao, L. Qu, L. Xie, and X. Wang, "Energy-efficient UAV-assisted federated learning: Trajectory optimization, device scheduling, and resource management," *IEEE Trans. Netw. Serv. Manag.*, vol. 22, no. 2, pp. 974 - 988, Apr. 2025.
- [5] S. Trindade and N. L. S. da Fonseca, "Client selection in hierarchical federated learning," *IEEE Internet Things J.*, vol. 11, no. 17, pp. 28480 - 28495, Sep. 2024.
- [6] H. Hu, Z. Chen, F. Zhou, R. Q. Hu, and H. Zhu, "Computation-efficient grouping, trajectory, and resource allocation for UAV swarm-assisted aerial-ground collaborative computing networks," *IEEE Internet of Things J.*, vol. 11, no. 7, pp. 12510 - 12525, Apr. 2024.
- [7] X. Li, Y. Qin, J. Huo, and W. Huangfu, "Computation offloading and trajectory planning of multi-UAV-enabled MEC: A knowledge-assisted multiagent reinforcement learning approach," *IEEE Trans. Veh. Technol.*, vol. 73, no. 5, pp. 7077 - 7088, May 2024.