# 실내 재난 환경에서 전력 소모량 최소화를 위한 강화학습 기반 UAV 최적 경로 연구

이상훈, 강정화*, 김재현

아주대학교 전자공학과, *아주대학교 AI융합네트워크학과

{shunlee134, *kjh990220, jkim}@ajou.ac.kr

# Optimal UAV Trajectory Algorithm based on Reinforcement Learning for Electric Energy Consumption Minimization in Indoor Disaster Environment

Sanghoon Lee, *Junghwa Kang, Jae-Hyun Kim

Department of Electrical and Computer Engineering, Ajou Univ,
*Department of Artificial Intelligence Convergence Network, Ajou Univ.

요 약

통신이 불가한 실내 재난 환경에서 unmanned aerial vehicle (UAV)는 공중 기지국의 역할을 할 수 있다. 하지만, UAV는 제한된 배터리로 인해 사용에 한계가 있으며, 재난 상황에서는 지속적으로 변화하는 환경으로 인해 통신 성능을 보장하기 어렵다. 따라서, 본 논문에서는 UAV의 통신 성능과 전력 소모를 보장하는 강화학습 기반 UAV 경로 최적화 알고리즘을 제안한다. 성능 분석 결과, 제안하는 알고리즘은 기존의 알고리즘에 비해 비행 경로 및 총 전력 소모량이 감소하였다.

## Ⅰ. 서 론

Unmanned aerial vehicle (UAV)는 재난 상황 발생 시 지상 기지국의 역할을 대신할 수 있으며, 통신 환경 구축 및 생존자 구조를 위한 수단으로 사용될 수 있다 [1], [2]. 하지만, 재난 상황에서는 환경에 대한 정보가 지속적으로 변하기 때문에 통신 효율을 보장하기 어렵다. 따라서, 재난 상황에서 UAV의 효율적인 통신을 보장하기 위해 강화학습이 주로 사용되고 있다 [1]. 또한, UAV는 배터리 제한으로 인해 운용에 한계가 있으므로 UAV의 전력 소모는 반드시 고려되어야 할 사항이다 [2].

따라서, 본 논문에서는 실내 재난 환경에서 사용자에게 효율적인 서비스를 제공하면서, 전력 소모를 최소화하는 강화학습 기반 UAV 경로 최적화 알고리즘을 제안한다.

## Ⅱ. 시스템 모델

본 논문의 시스템 모델은 실내 재난 환경을 가정한다. UAV는 실내에 집중적으로 발생하는 트래픽을 처리하기 위해 건물 주위에서 공중 기지국의 역할을 한다. 실내에는 $N$명의 사용자가 존재한다고 가정한다. 사용자들은 건물 주변을 비행하는 UAV에게 구조 신호를 전송한다. UAV는 해당 정보를 수신하여 응급구조시설에 전달한다. 통신 과정에서 사용자가 UAV로 전달하는 신호는 건물 외벽 및 건물의 층 등으로 인해 경로 손실이 발생한다. 건물의 층 투과를 고려한 경로 손실 식은 다음과 같다 [3].

$$PL = PL_b + PL_{tw} + PL_{in} + PL_{floor}, \quad (1)$$

$$PL_b = 20\log_{10}(f_{GHz}(d_{out} + d_{in})) + 20\log_{10}(\frac{4\pi \times f_{GHz} \times 10^9}{c}), \quad (2)$$

$$PL_{tw} = g_1 + g_2(1-\cos(\theta))^2, \quad (3)$$

$$PL_{in} = 0.5d_{in}, \quad (4)$$

$$PL_{floor} = n(g_1 + g_2(1-\sin(\theta))^2), \quad (5)$$

$PL_b$은 자유 공간 경로 손실, $PL_{tw}$는 벽 투과 손실, $PL_{in}$은 실내 경로 손실, $PL_{floor}$는 층 투과 손실을 의미한다. $f_{GHz}$는 반송 주파수이고, $d_{out}$은 실내 사용자와 UAV 사이 실외 거리를 나타내고 $d_{in}$은 실내 거리를 나타낸다. $c$는 빛의 속도이다. $g_1$, $g_2$는 건물 소재에 따른 상수이고 $\theta$는 건물 벽을 투과할 때의 입사각, $n$은 전송 신호가 투과하는 층의 개수이다.

## Ⅲ. 강화학습 기반 경로 최적화 알고리즘

강화학습은 agent가 action을 수행하면서, reward를 최대화하는 최적의 policy를 학습하는 방법이다. Policy는 state에서 action을 취할 확률을 나타낸다. 제안하는 알고리즘에서는 Markov decision process (MDP) 환경을 가정한다 [1]. MDP 환경은 state, action, state transition probability, reward, discount factor로 이루어져 있다. Agent는 action을 선택하는 주체이고 state는 agent의 action에 따라 변화하는 상태이다. Action은 agent가 수행하는 내용을 의미한다. 현재 state $s$에서 agent가 action $a$를 취했을 때, 다음 state $s'$가 될 확률은 $P_{ss'}^a$이다. Reward는 agent의 action을 통해 얻을 수 있다. 또한, 미래 reward에 대한 할인율 $\gamma$는 0과 1 사이의 값이다. 할인율은 미래 reward에 곱해지는데, 0에 가까울수록 현재 reward를 나타내고, 1에 가까울수록 미래 reward에 비중을 둔다. Q-function은 높은 reward를 얻을 수 있도록 state와 action을 고려한 가치 함수이다. 벨만 방정식에 의한 최적 Q-function은 다음과 같다.

$$Q^*(s,a) = r + \gamma \sum_{s' \in S} \max_{a'} P_{ss'}^a Q^*(s', a'), \quad (6)$$

$r$은 state $s$에서 action $a$를 수행했을 때의 reward, $S$는 모든 state, $a'$는 $s'$에서 택할 다음 action이다. 최적의 policy를 찾기 위해서는 반복 학습을 통해 Q-function이 최적의 Q-function에 근사하도록 해야 한다.

본 논문에서 agent는 UAV이고 state는 UAV의 위치, 통신한 사용자 수, 경로 탐색 시간이다. Action은 UAV의 이동 방향 및 크기이고, UAV

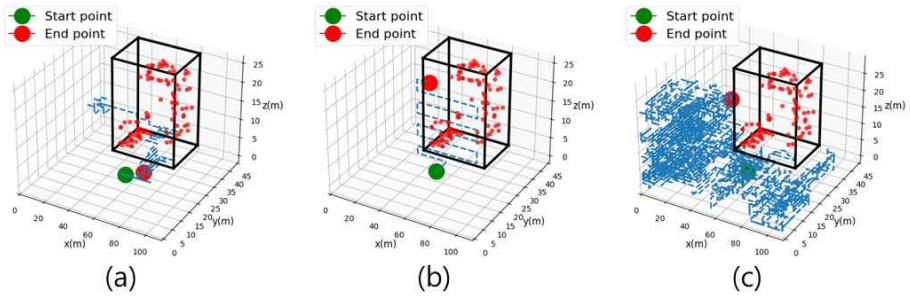


그림 1. UAV 경로: (a) 제안하는 알고리즘, (b) 너비 우선 탐색 알고리즘, (c) 랜덤 탐색 알고리즘

표 1. 시뮬레이션 파라미터

| Parameter | Value | Parameter | Value |
|---|---|---|---|
| Flight speed ($v$) | 20 m/s | Buffer limit | 10,000 |
| $f_{GHz}$ | 5.5 GHz | SNR threshold | 5 dB |
| Discount factor | 0.9 | Episode | 2,000 |
| Number of indoor users ($N$) | 100 | Optimizer | Adam |
| Data rate | {6.5, 13, 19.5, 26, 39, 52, 58.5, 65} Mbps | | |

는 x, y, z축을 따라 비행한다. Q-function을 최대화하여 UAV의 최적 이동 경로를 찾을 수 있도록 reward를 아래 수식과 같이 설정하였다.

$$r = U - P_t - P_{out}, \tag{7}$$

$U$는 UAV가 실내 사용자로부터 수신하는 신호의 signal-to-noise ratio (SNR)가 임곗값을 넘을 경우, 부여되는 reward 값이다. UAV가 다수의 실내 사용자를 서비스할 수 있도록 한다. $P_t$는 UAV의 경로 탐색 시간을 최소화하기 위한 penalty 값이다. 단시간 안에 실내 사용자와의 통신을 완료할 수 있게 한다. $P_{out}$은 UAV의 위치가 설정 반경에서 벗어났을 때 생기는 penalty이다. Agent는 학습을 통해 최종 reward를 극대화할 수 있도록 UAV가 최대한 많은 실내 사용자를 최단 시간 내에 서비스할 수 있는 action들을 선택한다.

본 논문에서는 deep Q-network (DQN)을 사용하여 UAV 경로를 학습한다. DQN은 강화학습을 신경망에 연결한 알고리즘이다 [1]. State가 바뀌면, 신경망을 학습시켜 다음 state에서 agent의 action을 도출한다. DQN은 신경망 학습을 통해 과거 정보와 미래 reward까지 고려하므로, UAV 경로 최적화 문제와 같이 state가 많은 경우에 유용하다.

## Ⅳ. 성능 분석 결과

본 논문에서는 성능 분석을 위해 제안하는 알고리즘 (proposed algorithm)과 너비 우선 탐색 알고리즘 (breadth-first search algorithm), 랜덤 탐색 알고리즘 (random search algorithm)의 성능을 비교하였다. 또한, 상세한 시뮬레이션 파라미터는 표 1과 같다. $N$명의 실내 사용자들은 각각 한 번씩 서비스된다고 가정한다.

그림 1의 (a), (b), (c)는 각각 제안하는 알고리즘과 너비 우선 탐색 알고리즘, 랜덤 탐색 알고리즘을 사용하였을 때, UAV 비행 경로를 나타낸다. 너비 우선 탐색 알고리즘은 시작점에서 x축과 z축을 따라 건물의 너비를 우선 탐색하는 방법이다. 랜덤 탐색 알고리즘은 x, y, z축을 따라 랜덤으로 탐색하는 방법이다. 세 알고리즘 모두 초록색 점 위치에서 비행을 시작하며, 빨간색 점 위치에서 비행을 종료한다. 제안하는 알고리즘에서 UAV는 건물 전면부의 중앙에서부터 탐색을 시작하여 다수의 사용자가 분포되어 있는 지역으로 이동한다. 따라서, 제안하는 알고리즘을 사용하였을 때, 다른 알고리즘에 비해 짧은 비행 경로를 갖는 것을 확인하였다.

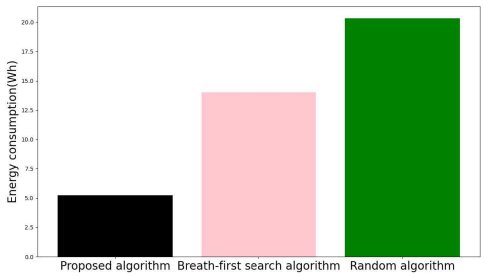


그림 2. 비행 경로에 따른 전력 소모량 비교

그림 2는 세 알고리즘의 UAV 전력 소모량을 비교한 그래프이다. UAV의 비행 경로에 따른 전력 소모량을 측정하였다. 제안하는 알고리즘은 너비 우선 탐색, 랜덤 탐색 알고리즘보다 총 서비스 시간이 196.65초, 336.9초만큼 감소했다. 따라서, 비행 경로에 따른 전력 소모량이 62.6%, 74.2% 감소한다. 또한, 너비 우선 탐색, 랜덤 탐색 알고리즘에서 UAV는 실내 사용자로부터 수신할 수 있는 신호의 세기를 고려하지 않고 비행한다. 하지만, 제안하는 알고리즘에서 UAV는 실내 사용자로부터 수신하는 신호의 SNR을 고려하여 비행하기 때문에 더 효율적인 서비스를 제공한다.

## Ⅳ. 결론

본 논문에서는 실내 재난 환경에서 강화학습 기반 UAV 최적 경로 알고리즘을 제안한다. 제안하는 알고리즘은 UAV의 경로 탐색 시간과 통신 가능한 실내 사용자 수를 고려한다. 따라서, UAV 전력 소모량을 최소화하고 다수의 사용자 서비스가 가능하다. 성능 분석 결과, 제안하는 알고리즘이 너비 우선 탐색, 랜덤 탐색 알고리즘을 사용했을 때보다 UAV 비행 경로 및 전력 소모량 측면에서 좋은 성능을 보이는 것을 확인하였다.

## 참 고 문 헌

[1] C. Wu et al., “UAV autonomous target search based on deep reinforcement learning in complex disaster scene,” IEEE Access, vol. 7, pp. 117227-117245, Aug. 2019.

[2] B. Zhu, E. Bedeer, H. H. Nguyen, R. Barton, and J. Henry, “UAV trajectory planning in wireless sensor networks for energy consumption minimization by deep reinforcement learning,” IEEE Transactions on Vehicular Technology, vol. 70, issue 9, pp. 9540-9554, Sep. 2021.

[3] J. Kang, H. Kashihara, J. T. S. Sumantyo, and J. H. Kim, "Performance analysis of uplink NOMA based full-duplex UAV for indoor disaster environment," in Proc. ICTC, Jeju, Korea, Oct. 2021.