어텐션 강화학습 기법 기반 도심 항공 모빌리티 운용 스케줄링 기술 연구

김준영*, 정소이

아주대학교 AI 융합네트워크학과*, 아주대학교 전자공학과

{*junzero0615, sjung}@ajou.ac.kr

Innovative Scheduling Techniques for Urban Air Mobility using Attention based Deep Reinforcement Learning

Junyoung Kim*, Soyi Jung

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

요 약

본 논문은 최근 메가 시티(mega-city) 조성 추세에 따른, 미래의 도심 내 교통 수요 증가와 인구 밀도 증가에 대응하기 위해 설계된 차세대 교통수단으로 주목받고 있는 도심 항공 모빌리티(urban air mobility, UAM) 운용 스케줄링 시나리오를 어텐션 메커니즘 기반 심층 강화학습 기법을 제안한다. 중요 정보에 가중치를 부여하여 집중 학습을 유도하는 어텐션 메커니즘을 기존 심층 강화학습 기법에 접목하여 학습 속도를 개선한 기법을 통해, 추후 다중 항공 모빌리티 에이전트 운용 시나리오에서 대도심 항공 모빌리티 실시간 스케줄링 기술 연구로 확장한다.

Ⅰ. 서론

최근, unmanned aerial vehicles(UAV)는 공격용 UAV를 활용한 무인기 운용과 기존 공격 체계와 연계한 유·무인 복합체계 등 현대 전장에서 군용 목적 UAV 활용, 초공간·초연결·초정밀을 목적으로 하는 6G 시나리오에서 공중 기지국 UAV 활용 등 다양한 분야에서 주목받고 있다[1][2]. 특히 초연결 6G 시나리오에서는 차세대 교통수단으로 도심 항공 모빌리티(urban air mobility, UAM) 운용을 적극 검토하고 있다. 기존 지상 중심으로 도심에서 운용되었던 모빌리티 시스템을 3차원 공간으로 확장하여 서비스되는 UAM은 지상 교통 체증 해소, 탄소 중립 달성 등 다양한 기대효과 창출이 예상된다. 그러나, UAM은 한정적인 배터리 용량으로 인한 운항 시간 한계가 존재한다. 이에 한정적인 짧은 비행시간 동안 최적의 거리로 비행하여 최대의 서비스를 제공하고, 에너지 효율을 극대화를 위해서는 운용 스케줄링 최적화가 필수적이다[3].

본 논문에서는, UAM의 운용 에너지 소모 최소화, 최단 거리 운행을 통한 승객 운송 최대화를 목표로 자연어 처리 분야의 어텐션(attention) 모델을 기존 강화학습 기법에 접목하여, UAM 운용 스케줄링 최적화 기법을 제안하고 성능 분석을 진행한다[4]. 제안하는 어텐션 기반 강화학습 알고리즘은 그림 1과 같은 시나리오에서 승객 위치 정보에 가중치를 부여하여 집중학습하는 방식으로 기존 심층 강화학습 알고리즘보다 개선된 학습 결과를 바탕으로 새로운 학습 기법을 제안한다[5][6].

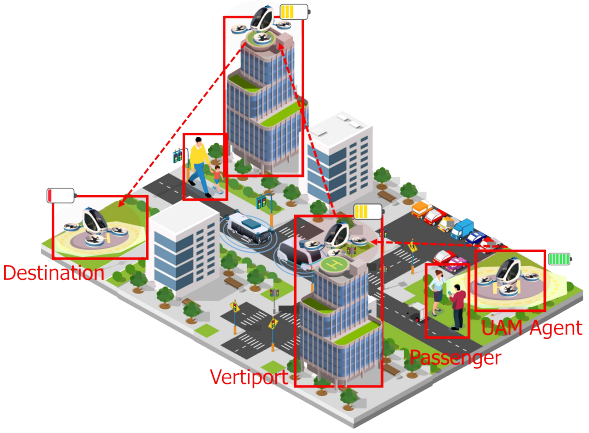

그림 1. UAM 운용 시나리오

Ⅱ. 시스템 모델

2.1 Deep Q Network

일반적인 심층 강화학습의 목표는 주어진 환경에서 강화학습 에이전트가 현재 상태(s_t) 공간에서 행동(a_t)을 통해 보상(r_t)을 최대화하는 정책(π)을 학습을 진행하는 것이 주요 목표다. Deep-Q-network(DQN) 은 특정 상태에서의 행동 가치 함수 $Q^{\pi}(s_t,a_t)=E_{s,a,\pi}\left[\sum_{i=t}^{T} r^{i-t} r_i\right]$ 를 최대화로 학습을 진행한다. DQN의 주된 특징은, replay memory에 저장된 experience replay sample을 Q-network와 target network의 입력값이 되어 학습의 안정성을 향상하고, 손실 함수(L) 값을 최소화하는 행동을 선택할 수 있게 주기적으로 Q-network를 복제하여 생성되는 target network 구조를 통해 빠른 학습이 가능한 것이 주된 특징이다. Target network를 형성하는 기준 값인 target vlaue는 수식 (1)과 같으며, 에이전트의 행동 값에 따른 DQN의 손실 함수(L)는 수식 (2)로 정의한다.

$$Y_t = r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta) \quad (1)$$

$$L = (Y_t - Q(s_t, a_t; \theta))^2 \quad (2)$$

2.2 Attention-DQN 강화학습 기반 시스템 모델

어텐션 모델은 주요 자연어 처리 분야에서 활용되고 있다. 주요 정보에 가중치를 부여하여 집중 학습 후 문맥을 유추하고 생성하는 어텐션 모델은 심층 강화학습 에이전트가 환경에서 중요한 정보를 집중적으로 학습하여 학습 파라미터 최적화와 수렴 속도 개선을 기대할 수 있다. 어텐션 모델 기반 강화학습의 주된 특징은 query(Q_i), key(K_i), value(V_i)의 관계를 통한 가중치를 상태 및 에이전트 관찰 요소에 부여하는 방식이다. 가중치(W)와 임베딩(h_i^{t-1}) 함수로 계산된 query, key, value의 값은 수식 (3)으로 표현하고, 이를 기반한 어텐션 가중치는 각 가중치와 차원(d_k)과의

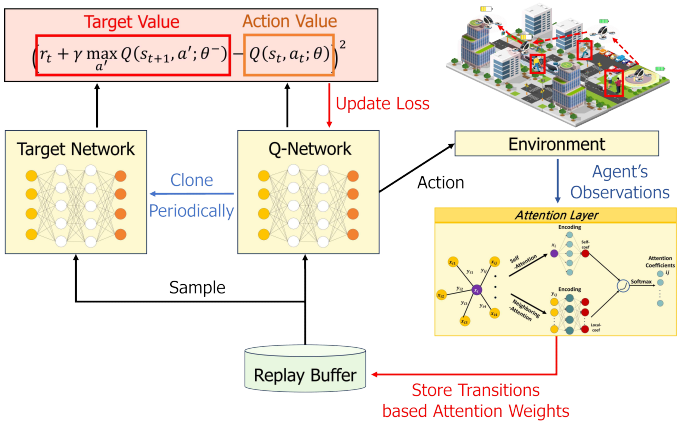

그림 2. Attention-DQN 알고리즘 구조도

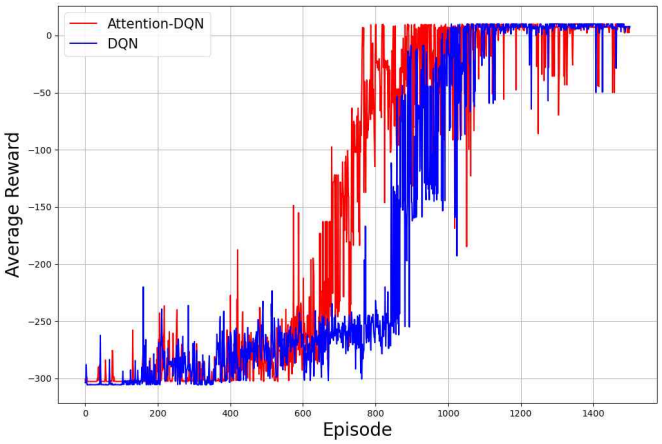

그림 3. 시뮬레이션 결과

표 1. 강화학습 환경 및 에이전트 파라미터

Parameter	Value
Batch size	128
Learning rate	0.001
Discount factor	0.99
Replay buffer size	100,000
$T_a, T_d, T_l, T_r, T_u, T_w$	10km/h
R_p, R_v	10, 20

관계로 수식 (4)와 정의하며. 어텐션 가중치에 따른 가치 함수는 수식 (5)로 재정의한다. 제안하는 attention-DQN 기법은 그림 2와 같다.

$$Q_i = W^Q h_i^{t-1}, K_i = W^K h_i^{t-1}, V_i = W^V h_i^{t-1} \quad (3)$$

$$A_i = softmax(\frac{Q_i^T K_i}{\sqrt{d_k}}) \quad (4)$$

$$Q^\pi(s_t,a_t) = h_i^{t-1}((s_{t-1},a_{t-1}),A_i) \quad (5)$$

UAM의 운용 스케줄링 알고리즘 설계를 위해 마르코프 결정 프로세스의 상태, 행동, 보상에 관한 정의는 다음과 같이 표현한다.

1) 상태(states) $s_t \doteq [u_x, u_y, u_z, P_i]$로 정의한다. UAM의 현재 위치를 u_x, u_y, u_z로 나타내며, 승객의 위치 배열을 P_i로 상태를 정의한다.

2) 행동(action) $a_t \doteq [T_a, T_d, T_l, T_r, T_u, T_w, T_s]$로 정의한다. 에이전트는 UAM은 상승, 하강, 왼쪽, 오른쪽, 위, 아래 그리고 정지로 행동을 결정한다.

3) 보상(reward) $R \doteq d_p + d_v + R_p + R_v$ 로 정의한다. R은 이전 스텝과 현재 스텝에서 UAM의 위치와 승객 위치 좌표의 차이(d_p)와 승객을 픽업 했을 때의 보상(R_p), UAM의 비행 최종 지점과의 거리(d_v) 및 도착 보상(R_v)으로 정의하여 UAM 에이전트가 승객을 태우고, 도착지점까지 비행하도록 설계한다.

Ⅲ. 시뮬레이션 결과

본 논문의 시뮬레이션 시나리오는 그림 1과 같이 UAM 에이전트는 고정된 출발지에서 가까운 승객들을 픽업하고 목적지까지 최단 거리로 이동하는 시나리오로 진행했다. 에이전트가 가까운 승객에게 비행할 수 있도록 거리 기반 보상과 승객을 픽업할 시에 대한 보상, 목적지 비행장과의 거리 및 도착 보상을 순차적으로 달리 부여하여 비행거리 최소화와 서비스 최대화를 목표로 학습을 진행했다. 시뮬레이션 결과로 그림 3과 같이 설계한 어텐션 기반 DQN 강화학습 기법이 약 780 에피소드에서 수렴 경향성을 보였던 것에 반해, 기존 DQN 알고리즘은 약 1050 에피소드에서 수렴 추세를 확인하였다. 이에 제안하는 어텐션 기반 DQN 강화학습 기법이 기존 DQN 알고리즘에 비해 학습의 안정성, 속도, 수렴성이 개선됨을 확인했다.

Ⅳ. 결론

본 논문의 결과를 기반으로 새로운 어텐션 기반 심층 강화학습을 통해 차세대 교통수단인 UAM 운용 가능성을 증명하였다. 제안하는 강화학습 기법을 통해 파라미터를 최적화하고 학습의 속도를 가속화하여 실제 운용 시나리오를 고려하여 다중 에이전트 강화학습과 이기종 모빌리티 시뮬레이션 연구로 확장한다.

참 고 문 헌

[1] J. Niping, Y. Zhiweri, and K. Yang, "Operational effectiveness evaluation of the swarming UAVs combat system based on a system dynamics model," *IEEE Access*, vol. 7, pp. 25209-25224, 2019.

[2] T. Hirai, K. Doi, and N. Wakamiya, "Optimal deployment of an aerial base station in heterogeneous cellular networks for heterogeneous user traffic demands," In *Proc, 2023 IEEE 97th Vehicular Technology Conference(VTC2023-Spring)*, Florence, Italy, August 2023.

[3] C. Park, W. J. Yun, J. P. Kim, T. K. Rodrigues, S. Park, S. Jung, and J. Kim, "Quantum multi-agent actor critic networks for cooperative mobile access in multi-UAV systems," *IEEE Internet of Things Journal*, vol. 10, no. 221, pp. 20033-20048, November 2023.

[4] J. Li, L. Xin, Z. Cao, A. Lim, W. Song, and J. Zhang, "Heterogeneous attentions for solving pickup and delivery problem via deep reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, March 2022.

[5] J. Li, L. Xin, Z. Cao, A. Lim, W. Song ,and J. Zhang, "Attention is all you need," In Proc, Advances in Neural Information Processing Systems, vol. 30, pp. 5998-6008, 2017.

[6] S. Iqbal, and F. Sha, "Actor attention critic for multi-agent reinforcement learning," in Proc, 36th International Conference on Machine Learning, California, USA, June 2019.