

# UAV-BS 환경에서 서비스 처리량 최대화를 위한 강화학습 기반의 UAV 배치 연구

박유민<sup>o</sup>, 홍충선<sup>\*</sup>

경희대학교 컴퓨터공학과

{yumin0906, cshong}@khu.ac.kr

## Deploying UAV based on Reinforcement Learning for Throughput Maximization in UAV Environments

Yu Min Park<sup>o</sup>, Choong Seon Hong<sup>\*</sup>

Department of Computer Science and Engineering, Kyung Hee University

### 요 약

현재 5G 기술의 상용화가 이루어 지는 단계이지만 통신 품질을 안정적으로 높이기 위해서는 많은 기지국이 필요하다. 따라서 지상 기지국을 대신하여 UAV-Base Station(UAV-BS)로 이동성과 경제적 이점을 얻을 수 있도록 연구가 이루어 지고 있는 추세이다. 이에 본 논문은 다수의 사용자가 UAV 통신을 할 때, 사용자들의 요구량을 고려하여 최적의 서비스 처리량을 가질 수 있는 위치를 찾는 기법을 제안한다. The Air-To-Ground Path Loss Model을 토대로 가상 통신 환경 구축하고, 사용자 요구량에 따른 공평한 채널 사용 시간 분배를 위해 Airtime Fairness를 적용한다. 더불어, 본 논문은 서비스 처리량 최대 위치를 찾기 위한 방법으로 Proximal Policy Optimization(PPO) 강화학습 알고리즘을 사용한다. 결과적으로 본 연구를 통해 서로 다른 요구량을 가진 사용자들이 임의로 배치되어 있을 때, 학습의 결과 모델을 통해 높은 서비스 처리량을 가진 위치를 찾을 수 있었다.

## 1. 서론

차세대 이동 통신인 5G는 2018년부터 채용되는 무선 네트워크 기술로써 밀리미터파를 이용하는 통신이다. 현재 5G의 상용화가 진행되고 있지만 밀리미터 파의 직진성으로 인해 중간에 손상되지 않도록 근거리로 셀을 구성한 네트워크를 구축해야 한다. 이런 특징으로 인해 5G의 안정적인 상용화를 위해서는 수 많은 기지국 설치가 요구되지만, 지상에서 새로운 기지국을 설치하는 것은 경제적 비효율성을 야기한다[1]. 따라서 지상 기지국을 대신하는 방안으로 UAV-Base Station(UAV-BS) 연구가 이루어지고 있다[1].

UAV는 최적 배치 문제, 제한된 에너지 및 보안 부분 등 다양한 한계 및 문제점을 지니고 있다. 논문[2]는 에너지 소비 최소화와 UAV 간 협력을 보장하기 위한 분산 UAV 포트를 관리하는 알고리즘을 제시했다. 본 논문은 UAV 최적 배치 문제를 해결하여 UAV-BS 환경에서 보다 높은 처리량을 제공하는 방안을 제안하고자 한다.

본 논문의 2장에서는 강화학습과 UAV 통신에서 처리량 최대화를 위한 연구들을 살펴보고, 3장에서는 본

논문의 강화학습 환경 구축을 위해 사용한 모델들에 대해 설명한다. 4장에서 제안한 시스템 모델과 다른 알고리즘 방법과의 서비스 처리량 비교를 통해 성능평가 내용을 다루며, 마지막 5장은 논문의 결론 및 향후 연구방향을 제시한다.

## 2. 관련 연구

### 2.1 Reinforcement Learning(강화학습)

강화학습은 주어진 환경에서 설정된 행동들을 반복하며 보상을 많이 받는 방향으로 Parameter들을 수정하는 전략을 취한다. 기존에 없던 환경에서 인공지능을 학습해야 하는 상황에서는 학습시킬 데이터 셋이 없기 때문에 기계학습 중에서도 강화학습이 적합하다.

강화학습에도 여러 기법이 존재하고 크게 Value-based와 Policy-based로 분류할 수 있다. Value-based와 달리 Policy-based는 행동들이 연속적인 경우와 실제 로봇이나 드론의 학습에 적합하다. Policy-based 기법 중에서는 대표적으로 Trust Region Policy Optimization (TRPO), Proximal Policy Optimization

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.2019-0-01287, 분산 엣지를 위한 진화형 딥러닝 모델생성 플랫폼)과 과학기술정보통신부 및 정보통신기술진흥센터의 Grand ICT연구센터 지원사업의 연구결과로 수행되었음 (IITP-2018-2015-0-00742) Dr. CS Hong is the corresponding author.

(PPO)가 존재하고 PPO는 TRPO를 발전시킨 기법으로 TRPO보다 간략하고 범용적인 방법이다[3]. 따라서 본 논문에서 강화학습 중에서도 PPO를 이용하여 서비스 처리량 최대화를 위한 UAV의 위치를 찾는 방안을 연구한다.

### 2.2 UAV-BS 환경에서의 최적 배치 연구

논문[4]은 처리량의 최대화 문제에 대한 방안으로 Tabu-Search를 기반으로 한 전역 최적화 알고리즘을 제안하고 있다. UAV가 움직일 수 있는 범위를 일정한 격자로 가정하고 유저와의 통신을 보장하는 범위에서 최적화된 위치를 찾았다. 이를 통해, 초기 배치와 비교하여 평균 26%정도의 처리량이 상승된 위치를 나타내었다. 그러나 격자로 UAV 움직임을 제한시킨 것은 연속적인 실제 환경에서 오차를 낼 수 있다는 한계를 지닌다. 따라서 본 논문은 UAV가 움직일 수 있는 범위를 연속적인 공간으로 하여 최적의 위치를 찾는 방안을 제시한다.

논문 [5]에서는 UAV와 유저와의 통신에서 QoE(유저 통신 경험의 품질)를 최대화시키는 위치를 찾았다. 위치를 찾는 방법으로는 Q-learning을 기반으로 강화학습을 사용하였고 K-means 군집화를 통해 유저들을 분류 후 각 군집에 UAV를 배치시키는 방법으로 다 대 다 환경을 만들었다. 하지만 Q-learning은 대표적인 Value-based 강화학습으로 드론의 학습에 적합하지 못하다. 따라서 본 논문에서는 2.1절에서 언급한 PPO를 사용할 것이며 QoE 대신 처리량을 극대화하는 곳에 UAV를 배치하는 것이 목표이다.

### 3. 제안 사항

본 논문은 다수의 사용자가 UAV를 이용한 통신을 할 때, 각각의 사용자들의 서비스 요구량을 고려하여 UAV가 최대의 처리량을 낼 수 있는 위치를 찾기 위한 방안으로 강화학습 모델을 적용한다. 본 연구의 환경 구현을 위해 3.1절의 네트워크 모델과 3.2절의 강화학습 모델을 제안한다.

표 1. 변수 정의 및 사용 값

Parameter	Description	value
$f$	Transmitter frequency	1.5GHz
$P_t$	Base station transmitting power	10dBW
$G_t$	Antenna gains for the base station	8dB
$G_r$	Antenna gains for the user.	0dB
$L$	Total System Losses	8dB
$B$	Bandwidth	10MHz
$N$	Noise power	100dBW

### 3.1 네트워크 모델

**The Air-To-Ground Path Loss Model:** UAV 통신 과정에서 거리 (d) 에 따른 신호 감쇄 효과로 인한 SNR(신호 대 잡음비)을 계산하기 위한 모델이다 [4].

$$PL(\text{dB}) = 20 \log_{10} \left( \frac{4\pi df}{c} \right) + P(\text{LoS})\eta_{\text{LoS}} + P(\text{NLoS})\eta_{\text{NLoS}} \quad ①$$

$$P_r(d) = P_t + G_t - PL + G_r - L \quad ②$$

$$\text{SNR} = \frac{\text{signal power}}{\text{noise power}} = \frac{P_r(d)}{N} \quad ③$$

이렇게 구한 SNR 과 식 ④ 으로 사용자 (i) 의 Data Capacity( $C_i$ )를 구할 수 있다.

$$C_i = B \log_2(1 + \text{SNR}) \quad ④$$

**Airtime Fairness:** 사용자들의 서로 다른 서비스 요구량에 따른 공평한 채널 사용 시간( $t_j$ ) 분배를 하기 위한 방법으로 사용자 (j) 의 Data Capacity ( $C_j$ ) 로 Throughput Capacity( $C_{T,j}$ )를 구할 수 있다 [6].

$$\sum_{j \in U} t_j \leq 1 \quad ⑤$$

$$C_{T,j} = t_j C_j \text{ for } \forall j \in S \quad ⑥$$

각 사용자에게 대한 Throughput Capacity를 더한 최종 처리량을 높이는 것이 본 연구의 최종 목표이다.

$$\max \sum_{j \in S} C_{T,i} \quad ⑦$$

결과적으로 본 논문에서 해결해야 할 문제는 식 ⑦ 을 만족시키는 UAV의 위치를 구하는 것이다.

### 3.2 강화학습 모델

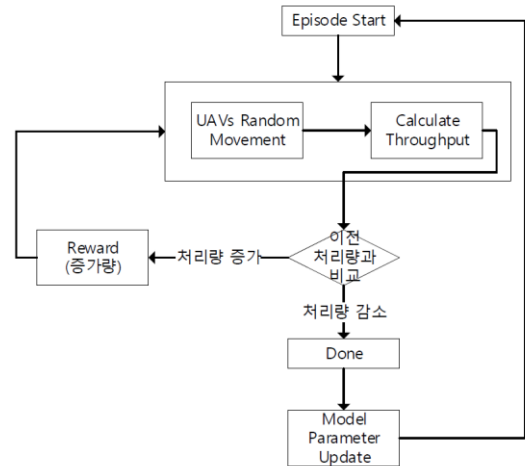


그림 2. 학습 시나리오 구조

강화학습에 대한 시뮬레이션은 Unity로 했고 MLagents 컴포넌트와 Tensorflow를 통해 학습하였다. 에피소드가 시작되면 10명의 사용자와 1대의 UAV가 반지름이 250m인 범위 안에서 임의로 배치한다. 이 때, 사용자는 서로 다른 서비스 요구량을 가지고 있으며 사용한 네트워크 모델들의 식에 따라 사용자들에 대한 전체 처리량을 계산한다. Action으로는 UAV가 임의의 지점으로 이동하는 것이고 그 후 처리량을 계산하여

이전 처리량과 비교한다. 만약 처리량이 증가했다면 증가치 만큼 Reward를 주고 그렇지 않으면 에피소드를 종료하고 학습 모델의 변수들을 업데이트한다. 이렇게 에피소드를 반복하여 학습을 시킨다. 그림 2는 학습 시나리오의 전체적인 구조이다.

#### 4. 성능 평가

성능 분석을 하기 위한 비교 방법은 Exhaustive Search를 사용하였다. Exhaustive Search는 사용자들의 중심으로부터 주어진 범위를 1m간격으로 모든 점을 대입하여 가장 처리량이 높은 곳을 찾는 방법이다.

$$Performance_A = \frac{Throughput_A - Throughput_{init}}{Throughput_{exhaustive} - Throughput_{init}} \quad \textcircled{8}$$

성능을 수치로 나타내기 위해 식 ⑧을 사용하였다. 식 ⑧로 정확도 측면에서 가장 좋은 Exhaustive Search를 대상으로 최초 처리량에 비해 얼마나 증가했는지 비율을 도출한다.

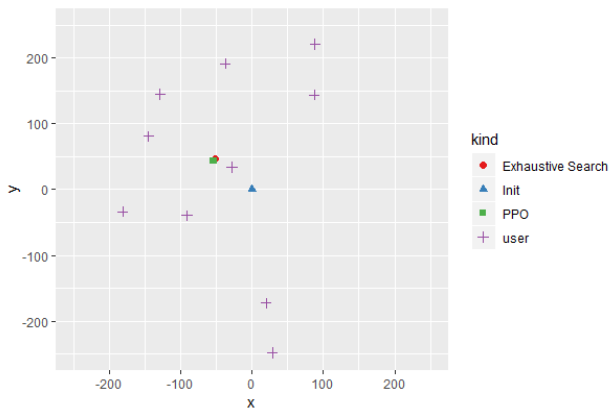


그림 2. UAV 및 사용자 위치 그래프

표 2. 수행 능력 결과표

Kind	PPO	Exhaustive Search
Performance	0.997929	1
Processing Time(ms)	653	1644

임의의 테스트 케이스에서 각 방법들이 결과로 서로 다른 위치에 배치되었고 그림 2을 통해 위치들을 확인할 수 있다. 표 2는 그 때의 각각의 성능과 결과를 내기까지 걸린 수행 시간을 비교해 놓은 결과표로 PPO와 Exhaustive Search의 처리량에 차이가 거의 없었다. 하지만 PPO는 Exhaustive Search에 비해 수행 시간이 월등히 낮았다.

그림 3는 임의의 테스트 케이스 10개에 대한 수행 시간 그래프이다. 이 때의 성능에 대한 평균은 0.9822이었다. 결과적으로 사용자의 위치와 서비스 요구량이 변하는 실제 상황에서 PPO 기반의

강화학습은 효율적으로 서비스 처리량을 높일 수 있다.

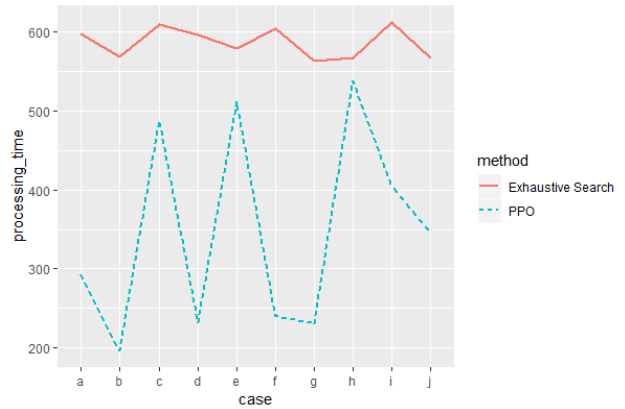


그림 3. 10개의 테스트 케이스에 대한 수행 시간

#### 5. 결론 및 향후 연구

본 논문에서는 서로 다른 서비스 요구량을 가진 사용자들에 대해 서비스 처리량 최대화 위치를 찾는 방법으로 강화학습 중 PPO라는 기법을 사용하였다. 제안한 방법을 통해 알고리즘과 비교하여 98% 이상의 처리량을 확인할 수 있었다. 앞으로 본 논문에서 다수의 UAV의 통신, 에너지 효율, 그리고 보안 등 고려하지 않은 한계들을 향후 연구에서 추가로 고려하여 본 논문을 발전시킬 수 있을 것이다.

#### 6. 참고문헌

- [1] Li, Bin, Zesong Fei, and Yan Zhang. "UAV communications for 5G and beyond: Recent advances and future trends." *IEEE Internet of Things Journal* (2018).
- [2] Lynskey, Jared, Kyi Thar, Thant Zin Oo, and Choong Seon Hong. "Facility Location Problem Approach for Distributed Drones." *Symmetry* 11, no. 1 (2019): 118.
- [3] Schulman, John, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. "Proximal policy optimization algorithms." *arXiv preprint arXiv:1707.06347* (2017).
- [4] ur Rahman, Shams, Geon-Hwan Kim, You-Ze Cho, and Ajmal Khan. "Positioning of UAVs for throughput maximization in software-defined disaster area UAV communication networks." *Journal of Communications and Networks* 20, no. 5 (2018): 452-463.
- [5] Liu, Xiao, Yuanwei Liu, and Yue Chen. "Reinforcement Learning in Multiple-UAV Networks: Deployment and Movement Design." *arXiv preprint arXiv:1904.05242* (2019).
- [6] Danna, Emilie, Subhasree Mandal, and Arjun Singh. "A practical algorithm for balancing the max-min fairness and throughput objectives in traffic engineering." In *2012 Proceedings IEEE INFOCOM*, pp. 846-854. IEEE, 2012.