Poster Abstract: Emergency Networking Using UAVs: A Reinforcement Learning Approach with Large Language Model

Yanggang Xu
Shenzhen International Graduate School, Tsinghua University
Shenzhen, China
xyg22@mails.tsinghua.edu.cn

Zhuozhu Jian
Shenzhen International Graduate School, Tsinghua University
Shenzhen, China
jzz21@mails.tsinghua.edu.cn

Jirong Zha
Shenzhen International Graduate School, Tsinghua University
Shenzhen, China
zhajirong23@mails.tsinghua.edu.cn

Xinlei Chen*
Shenzhen International Graduate School, Tsinghua University
Pengcheng Laboratory
RISC-V International Open Source Laboratory
Shenzhen, China
chen.xinlei@sz.tsinghua.edu.cn

ABSTRACT

Utilizing unmanned aerial vehicles (UAVs) as mobile access points can assist urban communication systems in establishing emergency networks in disaster scenarios. In this paper, to organize UAVs in large-scale environments for networking purposes, we propose a multi-agent reinforcement learning (MARL) model, in which the design of a selective parameter sharing mechanism and a grouping strategy enhances the model's scalability. Furthermore, the model adopts a reward mechanism based on intrinsic motivation, using the Large Language Model (LLM), to accelerate the optimization process. Numerical results demonstrate that this algorithm outperforms existing alternatives.

KEYWORDS

Multi-agent reinforcement learning, Large language model, UAV network

1 INTRODUCTION

Unmanned Aerial Vehicles (UAVs) are gradually being employed to assist in the deployment of wireless cellular networks, aiding networks in providing high-quality services across various scenarios. For example, communication base stations (BSs) and the optical cables connecting them are susceptible to damage caused by natural disasters, including extreme weather events, floods, and earthquakes. After such emergencies, user devices within the impacted regions find themselves in urgent need of emergency communication services. In such cases, UAVs can be rapidly deployed to establish alternative communication networks [8], offering crucial and valuable information to both the victims and rescue teams and contributing to disaster response and recovery efforts.

The elevated degrees of freedom in the motion of UAVs [3] play an important role in enhancing network performance. However, the optimization of the dynamic and complex communication network formed by UAVs poses a significant challenge. The presence of numerous UAVs forms a large-scale intricate network, and the motion of these UAVs leads to time-varying network topologies [7] and channel conditions. In solving control and planning tasks [2], reinforcement learning [4, 1] has shown remarkable capabilities. However, most works assume that the backhaul network interconnected with the core network is fully configured. Thus, the relays between nodes in the communication network are not well optimized. In addition, many algorithms lack in-depth exploration of scalability and coordination in large-scale UAV networking.

To address these challenges, this paper proposes a multi-agent reinforcement learning (MARL) method by devising a multi-agent grouping strategy based on UAV states and utilizing a selective parameter sharing mechanism for model training. Meanwhile, intrinsic rewards are configured for different agent groups based on task characteristics. This approach enhances the stability of the relay network, as well as the scalability and collaboration of the agents. Furthermore, leveraging the Large Language Model (LLM), we design a multi-agent action feedback scheme. It accelerates the model's search process in large-scale spaces, using common-sense to avoid meaningless searches.

2 METHODOLOGY

2.1 Definitions and Background

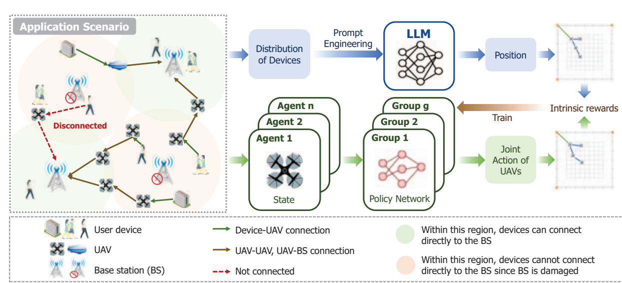

Figure 1: UAV networking using proposed model.

In the networking task, there are m mobile devices and n relay UAVs. Log–distance path loss model is employed and the path loss exponent is set to 2.03 and 2.01 for air-to-air and air-to-ground channels [5]. And the data rate $d_i, i = 1, 2, ..., m$ of each connection is given by the Shannon capacity formula. Furthermore, when the

*Xinlei Chen is the corresponding author.

979-8-3503-6201-5/24/$31.00 ©2024 IEEE
DOI 10.1109/IPSN61024.2024.00041

data rate falls below a threshold $\mathcal{D}$, that connection is considered to be disconnected. Generally, the goal of the algorithm is to optimize the average data rate of user devices $\bar{D} = \frac{1}{m}\sum_{i=1}^{m} d_i$ and the number of connected devices $M = \sum_{i=1}^{m} \mathbb{1}_{\{d_i > \mathcal{D}\}}$.

2.2 Proposed Algorithm

In this paper, we formulate the task as a Markov Decision Process, defined by $(\mathcal{S}, \mathcal{A}, \Omega, R)$. State space $\mathcal{S}$ encompasses states of UAVs and devices. And action space $\mathcal{A}$ defines the possible UAV actions. Ω denotes the state transition probability and R is reward model.

This paper proposes a MARL method, where each UAV has its own policy $\pi_j, j = 1, ..., n$ updated based on [6]. During the networking, some UAVs act as relay nodes, while others try to connect with more devices. Thus, we propose a selective parameter sharing method that divides UAVs into g groups based on positions and communication capabilities. Policies within each group share the same parameters. Relay UAVs often struggle to connect with devices due to the need to relay other nodes. Thus, we introduce a relay intrinsic reward R^{Rel}, as relay nodes bottleneck subsequent networks. And the communication reward for UAV j is $R_{j,t}^{COMM} = \eta\bar{D}_{j,t+1} + (1-\eta)M_{j,t+1} + \epsilon_j R_j^{Rel}$, where ϵ_j depends on the grouping of UAVs. In addition, the input of each policy includes the states of both itself and other UAVs to strengthen the collaboration. The output only contains the expected actions of specific UAVs.

Moreover, we employ the LLM technique. As the network exhibits regional characteristics, the entire space is divided into $K \times K$ grid cells. Therefore, the LLM input consists of the number of devices within K^2 grid cells. And the output of the LLM is the positions where UAVs should be located. Additionally, the input includes problem description and output guidance to ensure a formatted output O_{LLM}. And we intrinsically reward the agents based on the similarity between the O_{LLM} and the state s_{t+1} resulting from policy π at state s_t, i.e. $R_t^{LLM} = C(O_{LLM}, s_{t+1})$. Therefore, the reward model for each UAV j is defined as $R_{j,t} = \lambda R_{j,t}^{COMM} + (1-\lambda)R_{j,t}^{LLM}$.

3 EVALUATION

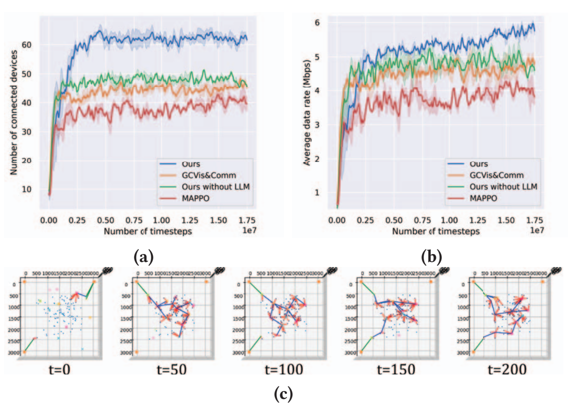

Figure 2: Experimental validation of the proposed model.

As shown in figure 2, we have tested 4 algorithms: the proposed method, GCVis&Comm [8], the proposed method without LLM, and MAPPO [6]. Here, we set $\eta = 0.9$, $\lambda = 0.6$, $n = 14$, $m = 100$, $K = 6$ and $g = 3$. Figure 2. (a) illustrates the number of devices connected, while figure 2. (b) shows the average data rate that devices can achieve. Figure 2. (c) demonstrates the networking process of the proposed algorithm from $t = 0$ to 200. The results show that the proposed method outperforms others in the networking task.

4 CONCLUSION AND FUTURE WORK

In this paper, we propose a MARL method combined with LLM in emergency networking tasks. This approach utilizes grouping and selective parameter sharing, along with intrinsic rewards based on LLM feedback, to achieve near-optimal performance. We would further refine the model to accomplish more practical tasks.

ACKNOWLEDGMENTS

This paper was supported by the National Key R&D program of China No. 2022YFC3300703, the Natural Science Foundation of China under Grant No. 62371269. Guangdong Innovative and Entrepreneurial Research Team Program No. 2021ZT09L197, Shenzhen 2022 Stabilization Support Program No. WDZC20220811103500001, and Tsinghua Shenzhen International Graduate School Cross disciplinary Research and Innovation Fund Research Plan No. JC20220011.

REFERENCES

[1] Xuecheng Chen, Haoyang Wang, Zuxin Li, Wenbo Ding, Fan Dang, Chengye Wu, and Xinlei Chen. 2023. Deliversense: efficient delivery drone scheduling for crowdsensing with deep reinforcement learning. In *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers* (UbiComp/ISWC '22 Adjunct). Association for Computing Machinery, Cambridge, United Kingdom, 403–408. ISBN: 9781450394239. DOI: 10.1145/3544793.3560412.

[2] Zhuozhu Jian, Zejia Liu, Haoyu Shao, Xueqian Wang, Xinlei Chen, and Bin Liang. 2024. Path generation for wheeled robots autonomous navigation on vegetated terrain. *IEEE Robotics and Automation Letters*, 9, 2, 1764–1771. DOI: 10.1109/LRA.2023.3334142.

[3] Zuxin Li, Fanhang Man, Xuecheng Chen, Baining Zhao, Chenye Wu, and Xinlei Chen. 2023. Tract: towards large-scale crowdsensing with high-efficiency swarm path planning. In *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers* (UbiComp/ISWC '22 Adjunct). Association for Computing Machinery, Cambridge, United Kingdom, 409–414. ISBN: 9781450394239. DOI: 10.1145/3544793.3560401.

[4] Jiyuan Ren, Yanggang Xu, Zuxin Li, Chaopeng Hong, Xiao-Ping Zhang, and Xinlei Chen. 2023. Scheduling uav swarm with attention-based graph reinforcement learning for ground-to-air heterogeneous data communication. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing* (UbiComp/ISWC '23 Adjunct). Association for Computing Machinery, Cancun, Mexico, 670–675. ISBN: 9798400702006. DOI: 10.1145/3594739.3612905.

[5] Evşen Yanmaz, Robert Kuschnig, and Christian Bettstetter. 2013. Achieving air-ground communications in 802.11 networks with three-dimensional aerial mobility. In *2013 Proceedings IEEE INFOCOM*, 120–124. DOI: 10.1109/INFCOM.2013.6566747.

[6] Chao Yu, Akash Velu, Eugene Vinitsky, Yu Wang, Alexandre M. Bayen, and Yi Wu. 2021. The surprising effectiveness of ppo in cooperative multi-agent games. In *Neural Information Processing Systems*. https://api.semanticscholar.org/CorpusID:232092445.

[7] Jirong Zha, Liang Han, Xiwang Dong, and Zhang Ren. 2023. Privacy-preserving push-sum distributed cubature information filter for nonlinear target tracking with switching directed topologies. *ISA transactions*, 136, 16–30.

[8] Xiaochen Zhang, Haitao Zhao, Jibo Wei, Chao Yan, Jun Xiong, and Xiaoran Liu. 2023. Cooperative trajectory design of multiple uav base stations with heterogeneous graph neural networks. *IEEE Transactions on Wireless Communications*, 22, 3, 1495–1509. DOI: 10.1109/TWC.2022.3204794.