

Fair Resource Allocation with Noise DDPG for UAV enabled ISAC Systems

Zhaowei Wang,¹ Weimin Jia,¹ Jianwei Zhao,¹ Wei Jin,¹ and Ye Yu¹

¹High-Tech Institute of Xi'an, Xi'an, Shaanxi 710025, China
Email: zhaojianweie@163.com

Integrated sensing and communication (ISAC) is considered one of the key technologies for the 6G network. In this letter, we propose a fair resource allocation method for the unmanned aerial vehicle (UAV) enabled communication network, where UAV are equipped with ISAC equipment to serve multiple users and targets. In order to achieve both fair communications and sensing, the resource allocation problem is formulated as maximizing the fairness index under the total power constraint, which is a typical non-convex optimization problem. Then, we propose the modified noise DDPG method to derive the power allocation. Finally, the simulation results verify the effectiveness of the proposed method compared with the benchmarks.

Introduction: Communication and sensing are considered to be the two fundamental functions of the sixth generation communication system (6G). Integrated sensing and communications (ISAC) realizes the coordination of sensing and communication functions with the software and hardware resource sharing, and has aroused great research interest in the academic community. Meanwhile, UAV becomes an important platform for the realization of 6G ISAC for its flexibility and controllability property [1].

The existing researches mainly focus on improving the overall performance of the UAV-ISAC systems by resource allocation. The authors in [2] maximized the safety of system by optimizing user scheduling, transmission power, and UAV trajectory. The authors in [3] studied the impact of UAV location deployment on the performance of ISAC systems. The authors in [4] studied the target tracking scheme in the UAV-ISAC system. In [5], a new ISAC framework was proposed, while the perceptual signal-to-noise ratio and the system throughput were maximized by jointly optimizing user association, transmission power, and UAV trajectory. In [6], deep learning method was used to maximize the sum of normalized sensing rate and normalized communication rate.

Although the current researches can greatly improve the overall performance of ISAC systems, the issue of system fairness for sensing and communication functions is often overlooked in their design. Resource allocation is a complicated optimization problem, and artificial intelligence provides an effective way for UAV enabled ISAC system. In this letter, we propose a resource allocation method for UAV enabled ISAC system. In order to achieve both fair communication and sensing, the fairness index is maximized under the total power constraint. Then, we exploit the deep reinforcement learning (DRL) method to solve the non-convex optimization problem. Sacrificially, we propose a learnable policy noise network for DDPG to derive the power allocation. Finally, the simulation results prove the effectiveness of the proposed algorithm.

System model: We consider a square area with a side length of D , where UAV serves as air ISAC base station (BS). Specifically, the UAV flies to the target area and then dynamically allocates resources to the randomly distributed N ground users and M targets.

The probability of line of sight (Los) for the ground user $n \in \mathbb{N}\{1, 2, \dots, N\}$ is

$$p_n^{Los}(t) = \frac{1}{1 + ae^{-b(\arcsin(\frac{h}{d_n(t)}) - \alpha)}}, \quad (1)$$

where $d_n(t)$ is the distance between UAV and the ground user n at time t , h is the hovering height of the UAV, and a, b are environment related constants.

There is a significant difference in path loss between line of sight (Los) and non line of sight (NLos). The path loss of the Los $L_n^{Los}(t)$ and NLos $L_n^{NLos}(t)$ between the UAV and ground user n at time t can be derived as

$$L_n^{Los}(t) = 20\log_{10}\left(\frac{4\pi f_c d_n(t)}{c}\right) + \xi_{Los}, \quad (2)$$

$$L_n^{NLos}(t) = 20\log_{10}\left(\frac{4\pi f_c d_n(t)}{c}\right) + \xi_{NLos}, \quad (3)$$

where c is the speed of light, f_c is the signal carrier frequency, and ξ_{Los}, ξ_{NLos} are the additional losses under Los and NLos.

Then, the path loss from the UAV to ground user n at time t can be expressed as

$$L_n(t) = P_n^{Los}(t)L_n^{Los}(t) + (1 - P_n^{Los}(t))L_n^{NLos}(t). \quad (4)$$

According to Equ. (4), the received power $p_n^r(t)$ at ground user n can be expressed as

$$p_n^r(t) = 10^{\frac{10\log_{10}P_n^t(t) - L_n(t)}{10}}, \quad (5)$$

where $p_n^t(t)$ is the power that transmitted from UAV to ground user n at time t .

Meanwhile, the communication sum rate of the UAV to ground user n at time t is

$$R_n(t) = \log_2\left(1 + \frac{p_n^r(t)}{\sigma^2}\right), \quad (6)$$

where σ^2 is the noise power.

In information theory, MI stands for mutual information, which is an important concept used to quantify the dependency relationship or shared information between two random variables[7]. Therefore, we exploit the radar MI as the sensing indicator, and then the MI from the UAV to the target $m \in \mathcal{M}\{1, 2, \dots, M\}$ can be expressed as

$$R_m(t) = \log_2\left(1 + \frac{p_m^r(t)}{\sigma^2}\right), \quad (7)$$

where $p_m^r(t)$ is the received power of the target m by the UAV.

During the sensing process, the signal needs to travel back and forth, so the path loss $L_m(t)$ and received power $p_m^r(t)$ can be expressed as

$$L_m(t) = P_m^{Los}(t)L_m^{Los}(t) + (1 - P_m^{Los}(t))L_m^{NLos}(t), \quad (8)$$

$$p_m^r(t) = 10^{\frac{10\log_{10}P_m^t(t) - 2L_m(t)}{10}}, \quad (9)$$

where $p_m^t(t)$ is the transmitted power from the UAV to target m .

Problem formulation and solution: Ignoring fairness would keep edge users and targets at low communication rates or MI. The maximization of the ISAC performance can lead to a tilt in power allocation. For example, the system would allocate more power to users who are closer to UAV to obtain greater communication sum rate. Therefore, we propose the fairness index to overcome the distance effects. Then, the communication fairness index $f^{com}(t)$ and the sensing fairness index $f^{rad}(t)$ can be respectively derived as

$$f^{com}(t) = \frac{\left(\sum_{n=1}^N f_n(t)\right)^2}{N \left(\sum_{n=1}^N f_n^2(t)\right)}, \quad (10)$$

$$f_n(t) = \frac{R_n(t)}{\sum_{n=1}^N R_n(t)}, \quad (11)$$

$$f^{rad}(t) = \frac{\left(\sum_{m=1}^M f_m(t)\right)^2}{M \left(\sum_{m=1}^M f_m^2(t)\right)}, \quad (12)$$

$$f_m(t) = \frac{R_m(t)}{\sum_{m=1}^M R_m(t)}. \quad (13)$$

In the ISAC system, it is necessary to ensure both communication fairness and sensing fairness. Therefore, we define the system fairness index as

$$f^{sys}(t) = \min(f^{com}(t), f^{rad}(t)). \quad (14)$$

Then, the spectral efficiency that reflects the overall performance of the ISAC system can be obtained as

$$S(t) = \frac{\omega_c \sum_{n=1}^N R_n(t) + \omega_s \sum_{m=1}^M R_m(t)}{\omega_c + \omega_s}, \quad (15)$$

where ω_c and ω_s respectively represent the emphasis coefficients of communication and sensing.

Under the principle of fairness, the optimization objective can be defined as

$$\max_{p_n^t(t), p_m^t(t)} f^{sys}(t) \quad (16)$$

$$\text{s.t.} \quad \sum_{n=1}^N p_n^t(t) + \sum_{m=1}^M p_m^t(t) = p_{total} \quad (17)$$

$$S(t) > S_{min} \quad (18)$$

$$p_n^t(t) \geq p_{min}, p_m^t(t) \geq p_{min} \quad (19)$$

where P_{total} and p_{min} are the total power and the minimum power allocated to each user respectively, and S_{min} is the minimum spectral efficiency. The optimization problem is non convex, which is difficult to solve through traditional convex optimization.

MDP model for power allocation: DRL is based on Markov decision processes (MDP), which optimizes (s, a, r, s') and maximizes the Bellman equation to get the cumulative reward. MDP is usually defined by (S, A, P, R, γ) , where S is the state space, A is the action space, P is the state transition matrix, R is the reward space, and γ is the discount factor which represents the agent's emphasis on future rewards.

The process of user and target movements can be expressed as a MDP model. Specifically, the user and target will upload its own location to the UAV through global positioning system (GPS). The UAV will then calculate the distance to the user or target based on its own location. The state needs to include the topological situation within the current region, so the state s_t can be defined as

$$s_t = (d_1(t), d_2(t), \dots, d_{M+N}(t)), \quad (20)$$

where the first N elements are the distance from the ground users to the UAV, and the last M elements are distance from the targets to the UAV.

The action a_t represents the power allocation strategy for the state at time t can be expressed as

$$a_t = (p_1^t(t), p_2^t(t), \dots, p_{M+N}^t(t)), \quad (21)$$

where the first N elements are the power allocated to ground users, and the last M elements are the power allocated to targets.

We define r_t as the reward for the resource allocation strategy under the current topology state. Due to the inability to maximize the fairness index and spectral efficiency simultaneously, we use α and β to emphasize different degrees of spectral efficiency and system fairness during model training to meet the demand for fairness in different scenarios, which is given by

$$r_t = \alpha S(t) + \beta f^{sys}(t). \quad (22)$$

DDPG Method: DDPG is an advanced version of Actor-Critic method. Actor networks can output actions and learn a good strategy. The critic network could learn a value function to determine which action is good in the current state. The copies of the actor network and critic network are created as the target actor network and target critic network, which improves the stability of neural network training. The update of the target network adopts a soft update method, which slowly updates the parameters of the target network.

Denote θ^μ and $\theta^{\mu'}$ as the parameters of actor network and the actor target network respectively, while θ^Q and $\theta^{Q'}$ as the critic network parameters and the critic target network parameters, respectively. For the update of Critic network, we define its loss function as

$$J(\theta^Q) = \frac{1}{N_s} \sum_i [r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1} | \theta^{\mu'})) - Q(s_i, a_i | \theta^Q)]^2, \quad (23)$$

where N_s is the batch size of data sampled from the experience replay memory, $\mu'(s_{i+1} | \theta^{\mu'})$ is the estimation for the actor target network's policy of the next state, $Q(s_t, a_t)$ is the value function. Then, the actor network could update the gradient objective function by gradient ascent, which is given by

$$\nabla J(\theta^\mu) = \frac{1}{N_s} \sum_i \nabla_a Q(s, a | \theta^Q) \Big|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu} \mu(s_i | \theta^\mu). \quad (24)$$

The Proposed Noise DDPG Method: In order to enable the agent to achieve more comprehensive learning, we exploit noise to increase exploration. Compared to adding noise to the output of the actor network, adding parameterized noise to the weights of the neural network can achieve more comprehensive exploration. The framework of the modified DDPG is shown in Fig. 1, where a learnable strategy noise is added to the fully connected layer of the actor network for exploration. The algorithm that combines this strategy noise can be named as Noisy DDPG. Specifically, the parameters can be learned through gra-

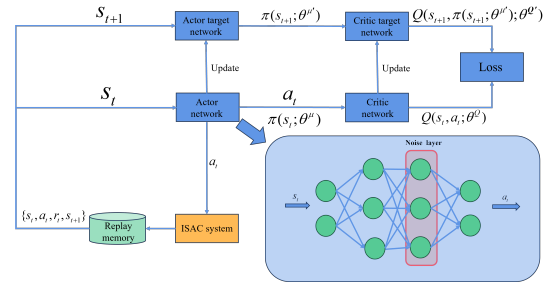


Fig 1 The frame work of Noise DDPG

dient descent to achieve end-to-end adjustment for the adaptive noise method. The actor network needs to learn both the network parameters and the variance of generated noise, which is given by

$$\chi_k = \omega \chi_s + b, \quad (25)$$

where χ_s and χ_k represent the input and output, respectively, ω is the weight matrix, and b is the bias vector.

After applying parameterized noise, we can obtain

$$\chi_k = (\mu_\omega + \sigma_\omega \odot \varepsilon_\omega) \chi_s + \mu_b + \sigma_b \odot \varepsilon_b, \quad (26)$$

where ω and b are parameterized with noise values, $\omega \sim \mathcal{N}(\mu_\omega, \sigma_\omega)$, $b \sim \mathcal{N}(\mu_b, \sigma_b)$. Moreover, $\varepsilon_{\omega, b} \sim \mathcal{N}(0, I)$ is the sampled Gaussian noise. At this point, the neural network needs to adaptively adjust by learning the mean and variance of noise.

The computational cost associated with the network will escalate rapidly as the size of the network increases. Decomposing Gaussian noise not only reduces the number of noise samples, but also lowers computational. Specifically, the number of neurons in the previous layer and the next layer is set to s and k , respectively. Each neuron generates an independent unit Gaussian noise as $\varepsilon_i, i \in [1, 2, \dots, s]$, $\varepsilon_j, j \in [1, 2, \dots, k]$. Then, the noise added to the neural network parameters can be represented as

$$\varepsilon_\omega = f(\varepsilon_j) f(\varepsilon_i^T), \quad (27)$$

$$\varepsilon_b = f(\varepsilon_j), \quad (28)$$

where $f(x) = \text{sgn}(x) \sqrt{x}$.

Simulations: In this section, we provide numerical results to validate the effectiveness of the proposed scheme. The number of ground users and targets is both 5, and their position is random. The other parameters are provided in table 1.

The Fig.2 displays the comparison between the Noise DDPG and the traditional DDPG. The final convergence of Noise DDPG is better than DDPG. Fig.3 shows the performance of Noise DDPG. We can find that spectrum efficiency and system fairness are contradictory for the proposed method. The reason is that we need to allocate more power to

Table 1. The imulation parameters.

| Parameters | Values |
|--------------------------------------|-----------------------|
| Side length D | 2000m |
| Total power P_{total} | 5w |
| Minimum power to each user p_{min} | 0.02w |
| Carrier frequency f_c | 2GHz |
| Noise power σ^2 | 5×10^{-17} w |
| UAV position | (1000m,1000m,500m) |
| Los additional losses ξ_{Los} | 1dB |
| NLos additional losses ξ_{NLos} | 21dB |
| Actor network learning rate | 1×10^{-4} |
| Critic network learning rate | 1×10^{-3} |
| Discount factor γ | 0.97 |

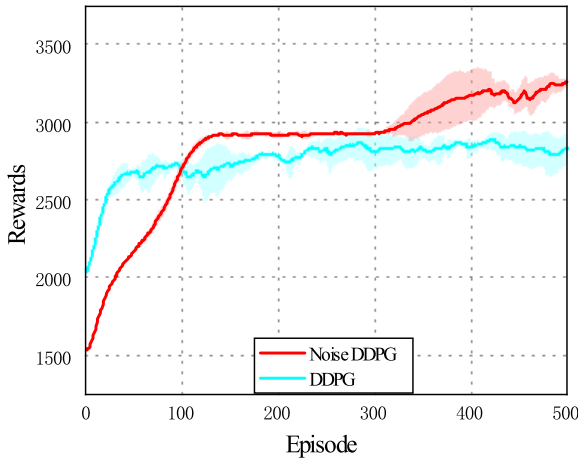


Fig 2 The comparison between Noise DDPG and DDPG.

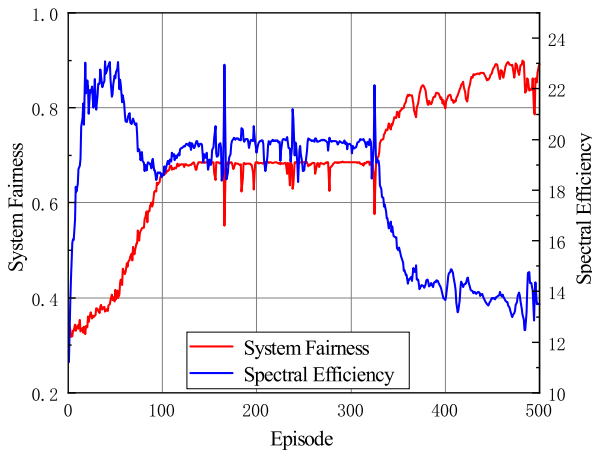


Fig 3 The performance of Noise DDPG

users who are closer to the UAV for higher spectral efficiency, which leads to the decrease of system fairness. At the same time, more power needs to be allocated to the targets for a larger system fairness, resulting in a decrease of spectral efficiency. Therefore, it is unrealistic to increase both spectral efficiency and system fairness simultaneously, and adjustments can be made according to actual needs.

The Fig.4 demonstrates the convergence of the system fairness index and spectral efficiency when $\beta = 10, 20$. The system fairness converges to 0.89 and 0.94, respectively. The spectral efficiency converges to 13.5 and 11.44, respectively. It can be observed that the agent will adopt corresponding emphasis on the system fairness or spectrum efficiency with the different β .

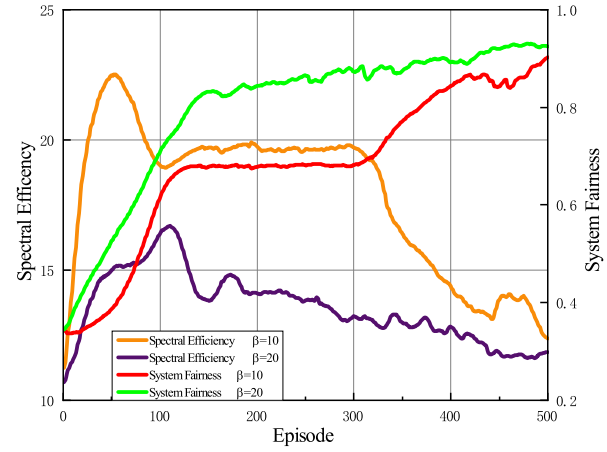


Fig 4 The system and channel model.

Conclusion: In this letter, we investigated resource allocation method for the UAV enabled ISAC system with multiple users and multiple targets. We introduced the radar MI and communication sum rate as the sensing and communication metric for the ISAC system respectively. Since the power allocation problem of the UAV-ISAC system considering fairness was a hard non convex problem, we exploited Noise DDPG to maximize the system fairness under the total power constraint. Finally, we provided simulation results to prove the effectiveness of the proposed method.

Acknowledgments: This research was sponsored by the National Natural Science Foundation of China (Grant No. 62001500 and 12403080), Postdoctoral Fellowship Program of CPSF(GZC20233565), the Natural Science Basis Research Plan in Shaanxi Province of China (2023-JCQN-0027), and the Youth Fund of PLA Rocket Force University of Engineering(2021-QNB-007)

© 2024 The Authors. *Electronics Letters* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Received: 10 January 2021 Accepted: 4 March 2021
doi: 10.1049/el12.10001

References

1. Zhang, R., et al.: A joint uav trajectory, user association, and beamforming design strategy for multi-uav-assisted isac systems. *IEEE Internet of Things Journal* 11(18), 29360–29374 (2024).
2. Liu, Y., et al.: Secure rate maximization for isac-uav assisted communication amidst multiple eavesdroppers. *IEEE Transactions on Vehicular Technology* 73(10), 15843–15847 (2024).
3. Jing, X., et al.: Isac from the sky: Uav trajectory design for joint communication and target localization. *IEEE Transactions on Wireless Communications* 23(10), 12857–12872 (2024).
4. Jiang, Y., et al.: Uav-enabled integrated sensing and communication: Tracking design and optimization. *IEEE Communications Letters* 28(5), 1024–1028 (2024).
5. Rezaei, O., et al.: Resource allocation for uav-enabled integrated sensing and communication (isac) via multi-objective optimization. In: *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. (2023).
6. Qi, Q., et al.: Deep learning-based design of uplink integrated sensing and communication. *IEEE Transactions on Wireless Communications* 23(9), 10639–10652 (2024).
7. Zhang, Q., et al.: Design and performance evaluation of joint sensing and communication integrated system for 5g mmwave enabled cavs. *IEEE Journal of Selected Topics in Signal Processing* 15(6), 1500–1514 (2021).