Collaborative Interference Source Search and Localization Based on Reinforcement Learning and Two-Stage Clustering

Guangyu Wu[1(✉)], Yang Huang[2,3], and Simeng Feng[2]

[1] College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
GYWu9908@163.com

[2] Key Laboratory of Dynamic Cognitive System of Electromagnetic Spectrum Space, Ministry of Industry and Information Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
{yang.huang.ceie,simeng-feng}@nuaa.edu.cn

[3] National Mobile Communications Research Laboratory, Southeast University, Nanjing 210000, China

Abstract. Exploiting unmanned aerial vehicles (UAVs) to locate the position of interferences has attracted intensive research interests, due to UAVs' flexibility and the feature of suffering less multi-path interference. However, in order to find the position of an interference source, off-the-shelf Q-learning-based schemes require the UAV to keep searching until it arrives at the target. This obviously degrades time efficiency of localization. To balance the accuracy and the efficiency of searching and localization, this paper proposes a collaborative search and localization approach, where search and remote localization are iteratively performed with a swarm of UAVs. For searching, a low-complexity reinforcement learning algorithm is proposed to decide the direction of flight (in every time interval) for each UAV. In the following remote localization phase, a two-stage clustering algorithm is proposed to estimate the position of the interference source, by processing intersections of the extensions of UAVs' trajectories. Numerical results reveal that in the proposed collaborative search and localization scheme, the proposed reinforcement-learning-based searching can benefit the collaborative localization, in terms of the accuracy of localization. Moreover, compared to the Q-learning-based approach, the proposed approach enables remote localization and can well balance accuracy, the robustness and time efficiency of localization.

This work was supported in part by the National Natural Science Foundation of China under Grant 61631020, 61827801 and 61901216, the Natural Science Foundation of Jiangsu Province under Grant BK20190400, the open research fund of National Mobile Communications Research Laboratory, Southeast University (No. 2020D08).

Q. Wu et al. (Eds.): WiSATS 2020, LNICST 357, pp. 429–448, 2021.
https://doi.org/10.1007/978-3-030-69069-4_36

Keywords: Reinforcement learning · Clustering · Unmanned aerial vehicle · Localization · Wireless communications

1 Introduction

Interference sources not only occupy growing spectrum resources illegally but also have brought grave implications to many fields such as railway communications, broadcast channel and IoT communications. The demand of an efficient and accurate search and location method to locate interference source is soaring [3]. Nevertheless, locating interference source in practice always suffers from unknown but dynamic surroundings, e.g. random background noise, such that detected signals remain changing over time intervals. Therefore, it is urgent to find out an interference source localization method which can be adaptive to varying environments instead of relying on a prior knowledge of the environments.

Unmanned aerial vehicles (UAVs) are capable of performing effective interference source localization [5], due to the fact that trajectories of UAVs can be flexibly planned to avoid obstacles in semistructured or unstructured environments. Moreover, signal processing devices and sensors, e.g. electronic scanning antennas, carried by UAVs suffer less multi-path interference, such that the position achieved by localization can be more accurate and reliable. In the meanwhile, attempts were made to employ multiple UAVs for a collaborative search and localization of interference sources. In [6], the authors proposed a received signal strength (RSS) value based localization method using multi-UAV. However, such a method is applicable only in scenarios where transmit power and propagation parameters of the interference source are known.

Instead of relying on a prior knowledge of the environments, reinforcement learning [2,9] is able to exploit samples and function approximation to optimize performance in dynamic environments. Recently, some researches studied reinforcement-learning-based interference source localization with UAVs [1,11,12]. In these works, a single UAV is exploited to locate interference source in unknown dynamic environments. Searching the interference source only based on reinforcement learning may require the UAV to keep searching until it arrives at the target. Apparently, such a search heavily degrades time efficiency. It is also inapplicable to the scenario where performing localization from a distance is necessary. On the other hand, remote localization may suffer from multi-path effect and therefore mitigation in the accuracy performance.

Against this background, this paper proposes a novel collaborative search and localization approach based on reinforcement learning and two-stage clustering, where the search and localization could benefit from synthesizing individual decisions within a swarm of UAVs. Briefly, in the proposed method, a searching phase and a localization phase are alternatively performed. During the searching phase of a certain time interval, each UAV in the swarm can decide and evaluate direction of searching through reinforcement learning. Then, in the following localization phase of the time interval, by developing a novel two-stage clustering

algorithm, Individual estimates of the position of the interferer achieved by UAVs are synthesized, by processing intersections of extensions of the trajectories. The contributions of this paper are listed as follows.

Firstly, a novel collaborative search and localization scheme based on reinforcement learning and two-stage clustering is presented for a swarm of UAVs to locate an interference source. Basically, the design of the proposed scheme aims at benefiting from the accuracy of localization and intelligence resulted from reinforcement-learning-based searching, as well as time efficiency of the remote localization. To this end, the proposed scheme merges reinforcement-learning-based searching and collaborative remote localization by alternatively performing a search phase and a localization phase over time intervals.

Secondly, for the search phase, a low-complexity reinforcement learning algorithm is proposed. During this phase, each UAV in the swarm can decide and evaluate its own direction of searching through reinforcement learning. In the localization phase, a novel two-stage clustering algorithm is proposed to find out the potential position of the interference source. Firstly, the two-stage clustering algorithm clusters UAVs' trajectories' intersections and decides whether to move a certain intersection into a candidate queue by performing the weighted clustering. Then, by performing the secondary clustering, the position of the interference source can be estimated.

Thirdly, in-depth simulation results are presented to demonstrate the performance of the proposed method. The effectiveness of the proposed collaborative search and localization scheme is confirmed. Numerical results show that compared to the Q-learning-based scheme, the proposed scheme is more time-efficient and can perform remote localization. It is also shown that by integrating reinforcement-learning-based searching with two-stage-clustering-based localization, the proposed scheme can improve the robustness and accuracy of the remote localization.

The remainder of this paper is organized as follows. Section 2 presents the system model. Section 3 proposes a collaborative interference source search and localization method, based on reinforcement learning and two-stage clustering. Simulation results are presented and analyzed in Sect. 4. Conclusions are drawn in Sect. 5.

2 System Model

In this paper, we aim at designing a time-efficient and accurate approach which applies a swarm of UAVs to search and locate a certain interference source, without a prior knowledge of the interference source model or the noise model. In the swarm, each UAV is equipped with an electronic scanning antenna which can be used to measure power of received signals from different directions. An example of the studied system is presented in Fig. 1.

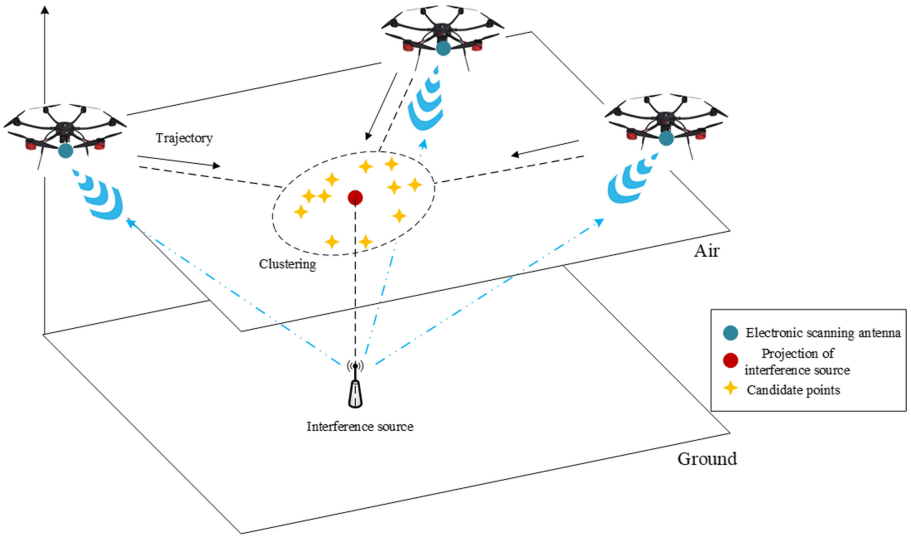


Fig. 1. Collaborative interference source search and location system with a swarm of UAVs.

2.1 Trajectory Model of UAVs

Suppose that in the swarm, the number of the UAVs is n. The time-varying coordinate of a certain UAV U_i at time interval $j-1$ can be designated as $(x_{j-1}^{(i)}, y_{j-1}^{(i)}, z_{j-1}^{(i)})$. The initial coordinate of a certain UAV U_i (for $i = 1, \ldots, n$) can be written as $(x_0^{(i)}, y_0^{(i)}, z_0^{(i)})$. We assume that each UAV flies in a straight line between two adjacent time intervals, and stays at a constant altitude $z_0^{(i)}$. Hence, the position of U_i at time interval j can be expressed as

$$x_j^{(i)} = x_{j-1}^{(i)} + l_j^{(i)} \cos \lambda_j^{(i)}, \tag{1}$$

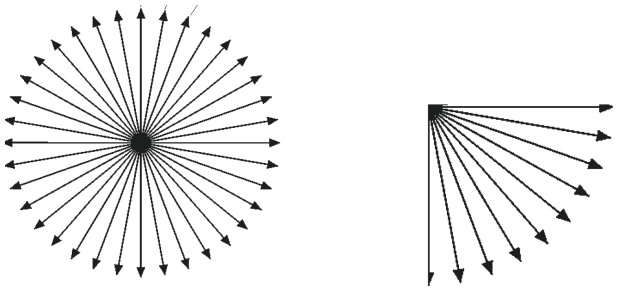
$$y_j^{(i)} = y_{j-1}^{(i)} + l_j^{(i)} \sin \lambda_j^{(i)}, \tag{2}$$

$$z_j^{(i)} = z_0^{(i)}, \tag{3}$$

where $l_j^{(i)}$ represents the horizontal distance between horizontal coordinates $(x_{j-1}^{(i)}, y_{j-1}^{(i)})$ and $(x_j^{(i)}, y_j^{(i)})$; $\lambda_j^{(i)}$ is the angle between the x-axis and the direction UAV U_i flies towards. In this paper, we assume that $l_j^{(i)}$ is equal to l $\forall i, j$.

2.2 Receive Power Model of Electronic Scanning Antenna

In the studied scenario, each UAV is equipped with a three-demensional electronic scanning antenna to measure power of receive signals from horizontal,



(a) Horizonal angles (b) Vertical angles

Fig. 2. The detection range of an electronic scanning antenna.

as well as vertical, directions. Each electronic scanning antenna is able to measure u horizontal directions $\{\theta_1, \theta_2, \theta_3, \cdots, \theta_{u-1}, \theta_u\}$ and v vertical directions $\{\varphi_1, \varphi_2, \varphi_3, \cdots, \varphi_{v-1}, \varphi_v\}$, as shown in Fig. 2.

In our study, a UAV flies at a constant altitude and directions of flight vary only in the horizontal plane. Therefore, given a horizontal angle θ_i, although power of signals from different vertical angles $\{\varphi_1, \varphi_2, \varphi_3, \cdots, \varphi_{v-1}, \varphi_v\}$ can be different, only the maximum can be applied to making decisions on the horizontal direction of flight, which should be in the direction of the signal source. Therefore, let a set of $P_d(\theta_i, :)$ collects receive power values obtained at horizontal angle θ_i and all the vertical angles $\{\varphi_1, \varphi_2, \varphi_3, \cdots, \varphi_{v-1}, \varphi_v\}$, the final power value $P_f(\theta_i)$ (with respect to (w.r.t.) the horizontal angle θ_i) for making decisions on horizontal directions of flight can be defined as

$$P_f(\theta_i) = \max(P_d(\theta_i, :)), \forall \theta_i \in \{\theta_1, \theta_2, \theta_3, \ldots, \theta_{u-1}, \theta_u\}. \tag{4}$$

Benefit from the extended dimension (i.e. scanning in vertical directions), three-dimensional electronic scanning can achieve a higher detection accuracy than the two-dimensional electronic scanning, however, introducing higher computational complexity for further signal processing. By performing (4), the search of the direction of an interference source can be restricted to the range of horizontal angles.

3 Collaborative Search and Location

In this section, we propose and design a collaborative search and localization scheme, which consists of a search phase and a localization phase. As depicted in Fig. 3, the search phase and the localization phase are alternatively performed, where the former is based on reinforcement learning while the latter is based on two-stage clustering.

Prior to elaborating on the design of the reinforcement learning and the two-stage clustering algorithms, we concisely outline the collaborative search and

localization scheme. As demonstrated in Fig. 3, in the search phase of every iteration (i.e. each time interval), each UAV (among a swarm of n UAVs) decides the direction of flight in the next time interval by performing reinforcement learning. The instant reward of the reinforcement learning can be related to power of receive signals (which might be emitted by the interference source). Hence, the direction of flight which is determined by the reinforcement learning can be a possible direction of the interference source. That is, by performing such reinforcement learning algorithms at each UAV in the swarm, the UAVs can search and fly towards the interference source. Moreover, as each UAV flies in a straight line between two adjacent time intervals, the extensions of the trajectories of two UAVs may intersect, where the intersection is essentially a possible position of the interference source. Due to the fact that the searching may suffer from noise and possibly multi-path interferences, extending the trajectories of the UAVs may achieve a large number of intersections. To find out the potential position of the interference source, the aforementioned intersections are processed by the proposed two-stage clustering algorithm during the localization phase. The two-stage clustering algorithm first cluster the intersections and make a decision whether to move a certain intersection into the candidate cluster. Then, by performing the secondary clustering, the position of the interference source can be estimated. The search phase and the localization phases will implement until the estimated position meet the terminating condition which is elaborated in the subsequent sections.

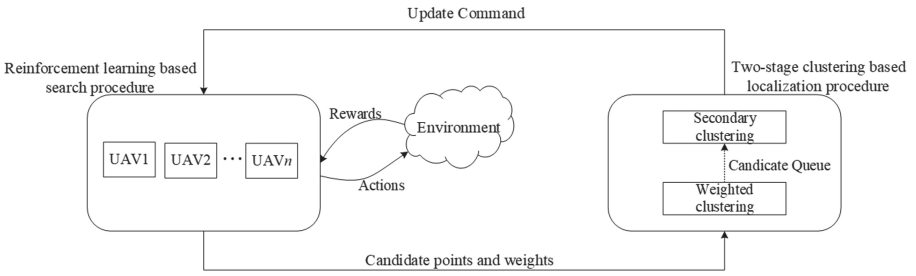


Fig. 3. The framework of collaborative search and localization based on reinforcement learning and two-stage clustering.

3.1 Reinforcement Learning Based Search

Searching According to Measured Power. Interference signals can be received by the UAV's electronic scanning antenna, and changes in measured power of the receive signals can indicate whether the distance between the UAV and the interference source is reduced. Therefore, for each UAV, in order to find out the position of the interference source, the direction of flight in each time interval can be selected according to measured power. Such a search problem (or decision problem on directions of flight in each time interval) can be equivalent

to maximizing the expected long-term measured power, where the long-term measured power can be formulated as the discounted sum of all future rewards (i.e. measured power) at current time t. Such a problem can be modeled as a Markov Decision Process (MDP) problem [8], which aims at finding the optimal policy $\pi^{(i)}$ for each UAV U_i, so as to maximize the expectation of discounted sum of all future measured power, i.e. [7]

$$\max_{\pi^{(i)}} \mathbb{E}_{\pi^{(i)}} \left\{ \lim_{\mathcal{T}\to\infty} \sum_{t=0}^{\mathcal{T}} \gamma^t r^{(i)}(s^{(i)}, a^{(i)}) \right\}, \tag{5}$$

where $\gamma \in [0, 1]$ denotes a discount factor; $a^{(i)}$ and $s^{(i)}$ represent UAV U_i's action and state, respectively. The action $a^{(i)} \in \{1, 2, 3, \cdots u\}$ selects direction of flights from a set of u possible flight directions $\{\theta_1, \theta_2, \cdots, \theta_u\}$, as shown in Fig. 2. The state $s^{(i)}$ is the direction selected by UAV U_i in the previous time interval, e.g. state $s_j^{(i)}$ in time interval j is the action $a_{j-1}^{(i)}$ in time interval $(j-1)$

$$s_j^{(i)} = a_{j-1}^{(i)}\,. \tag{6}$$

In (5), the reward $r^{(i)}(s^{(i)}, a^{(i)})$ is defined as

$$r^{(i)}(s^{(i)}, a^{(i)}) = \frac{1}{N} \sum_{k=1}^{N} D_k(s^{(i)}, \theta_{a^{(i)}})\,, \tag{7}$$

where $D_k(s^{(i)}, \theta_{a^{(i)}})$ is equal to measured power of a sample of signals received by the electronic scanning antenna at state $s^{(i)}$ and direction $\theta_{a^{(i)}}$ (which results from action $a^{(i)}$). As shown in (7), in order to alleviate effect of noise on measured power, power of N samples are averaged such that $\frac{1}{N}\sum_{k=1}^{N} D_k(s^{(i)}, \theta_{a^{(i)}})$.

Reinforcement Learning Algorithm. In order to solve problem (5), we propose a reinforcement learning algorithm based on Q-learning, which is a lookup-table-based approach [7]. By this means, the expectation of discounted sum of all future measured power can be recursively approximated. In this paper, in order to accelerate the convergence of Q-learning, we simultaneously update action values $Q^{(i)}(s^{(i)}, :)$ with various actions for a given state $s^{(i)}$. Hence, different from the conventional Q-learning, the update rule is given as

$$Q^{(i)}(s^{(i)}, :) \leftarrow Q^{(i)}(s^{(i)}, :) + \alpha[r^{(i)}(s^{(i)}, :) + \gamma Q^{(i)}(s^{'(i)}, :) - Q^{(i)}(s^{(i)}, :)], \tag{8}$$

where $Q^{(i)}(s^{(i)}, :)$ collects action values ranging from $Q^{(i)}(s^{(i)}, \theta_1)$ to $Q^{(i)}(s^{(i)}, \theta_u)$, which are the action values with current state $s^{(i)}$ and all possible actions. Similarly, $Q^{(i)}(s^{'(i)}, :)$ collects qualities of actions w.r.t. the previous state $s^{'(i)}$. In contrast to the reward in the conventional Q-learning [7], $r^{(i)}(s^{(i)}, :)$ in (8) collects not only the reward $r^{(i)}(s^{(i)}, a^{(i)})$, which results from taking action $a^{(i)}$, but also average power of samples measured at other directions by the electronic scanning antenna, i.e. $\frac{1}{N}\sum_{k=1}^{N} D_k(s^{(i)}, \theta)$ for all $\theta \neq \theta_{a^{(i)}}$

but belonging to $\{\theta_1, \theta_2, \cdots, \theta_u\}$. It means that by taking action $a^{(i)}$, the UAV arrives at a certain position and achieves the reward $r^{(i)}(s^{(i)}, a^{(i)})$ (for which the electronic scanning antenna receive signals from the direction of $\theta_{a^{(i)}}$); afterwards, signals from directions $\theta \neq \theta_{a^{(i)}}$ are detected by the electronic scanning antenna to obtain $r^{(i)}(s^{(i)}, :)$. Note that in this paper, we consider the scenario where γ and α for all UAVs are identical.

Once action values $Q^{(i)}(s^{(i)}, :)$ is updated, each UAV find the optimal action by performing

$$a^{(i)*} = \arg\max_{a^{(i)}} Q^i(s^{(i)}, a^{(i)}). \tag{9}$$

After action $a^{(i)*}$ is chosen at time interval j, each UAV flies in its direction $\theta_{a^{(i)*}}$ for a distance of l. The horizontal coordinate of each UAV is then updated as $(x_j^{(i)}, y_j^{(i)})$. The trajectory at time interval j for each UAV satisfies

$$y^{(i)} = \tan\theta_{a^{(i)*}}(x^{(i)} - x_j^{(i)}) + y_j^{(i)}, \tag{10}$$

where $\tan\theta_{a^{(i)*}}$ is related to the $\lambda_j^{(i)}$ in (1) and (2), while $(x_j^{(i)}, y_j^{(i)})$ stands for the UAV's current position.

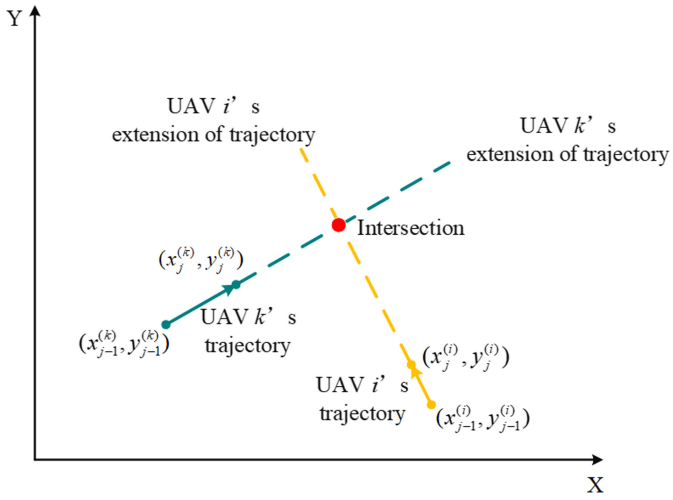


Fig. 4. Intersection of UAVs' trajectories at time interval j.

For a certain pair of UAVs, the extensions of their trajectories can intersect at time interval j, as shown in Fig. 4. Since UAVs fly towards the direction that can maximize the expectation of discounted sum of measured power, the intersection is essentially a potential position of the interference source. In general, the intersection of the extension of UAV U_i's trajectory and that of U_k's trajectory can be obtained by

$$p^{(i,k)} = (\frac{x_j^{(i)} \cdot \tan\theta_{a^{(i)*}} - x_j^{(k)} \cdot \tan\theta_{a^{(k)*}} + y_j^{(k)} - y_j^{(i)}}{\tan\theta_{a^{(i)*}} - \tan\theta_{a^{(k)*}}}, \\ \frac{\tan\theta_{a^{(i)*}} \cdot \tan\theta_{a^{(k)*}} \cdot (x_j^{(i)} - x_j^{(k)}) + \tan\theta_{a^{(i)*}} \cdot y_j^{(k)} - \tan\theta_{a^{(k)*}} \cdot y_j^{(i)}}{\tan\theta_{a^{(i)*}} - \tan\theta_{a^{(k)*}}}). \tag{11}$$

A weight is defined for each intersection, and the weight can be obtained as

$$w^{(i,k)} = \frac{Q^{(i)}(s^{(i)}, a^{(i)*}) \cdot Q^{(k)}(s^{(k)}, a^{(k)*})}{Q_0^{(i)}(s_0^{(i)}, a_0^{(i)*}) \cdot Q_0^{(k)}(s_0^{(k)}, a_0^{(k)*})}, \tag{12}$$

where $Q^i(s^{(i)}, a^{(i)*})$ and $Q^k(s^{(k)}, a^{(k)*})$ represent the maximum action values of U_i and U_k respectively at current state $s^{(i)}$ and $s^{(k)}$. $Q_0^{(i)}(s_0^{(i)}, a_0^{(i)*})$ and $Q_0^{(k)}(s_0^{(k)}, a_0^{(k)*})$ denote U_i and U_k's maximum action values obtained at the first state.

The intersections form a candidate set $\left\{p^{(i,k)}\right\}(i,k \in [1,n], i < k)$ which is used as the input of the two-stage clustering algorithm for localization alone with the weight set $\left\{w^{(i,k)}\right\}(i,k \in [1,n], i < k)$. The proposed reinforcement learning algorithm is summarized as a part of Algorithm 1.

3.2 Two-Stage Clustering Based Localization

Localization Based on Searching Results. As UAVs fly in straight line between two adjacent time intervals, the extensions of the trajectories of arbitrary two UAVs can intersect, where the intersection essentially constitutes a possible position of the interference source. In this paper, we study the scenario where there is only one interference source. Intuitively, in the ideal case, the intersection of the extensions of trajectories of UAVs is unique. Unfortunately, due to the noise, multi-path effect, etc., there can be a large number of intersections.

Thus, aiming at estimating the most possible position of the interference based on the intersections found in the search phase, the localization problem boils down to a clustering problem [4,10] , which can be cast as

$$\min_{P^T} \frac{1}{|C|} \sum_{(i,k)\in C} w^{(i,k)} \cdot \left\| P^{(i,k)} - P^T \right\|, \tag{13}$$

where P^T is an estimate of the position of the interference source; and P^T minimizes the weighted sum of Euclidean distance between an arbitrary intersection point $p^{(i,k)}$ and P^T. In (13), the set C collects all potential $p^{(i,k)}$ found in the search phase; and $|C|$ is the cardinality of C.

Algorithm 1. The proposed collaborative search and localization

Initialize the number of UAVs n;
Initialize flight direction set $d = \{1, 2, 3 \cdots, u\}$;
Initialize learning rate α, discount factor γ and step length l;
Set iteration number $j = 1$, operation command $f = false$;
for $i = 1 \rightarrow n$ **do**
  Initialize UAV U_i's position $\left(x_0^{(i)}, y_0^{(i)}, z^{(i)}\right)$;
  Initialize $Q^{(i)}(:,:)$ and $r^{(i)}(:,:)$ to $\mathbf{0}$ (i.e. a zero matrix);
  Select an initial state $s_0^{(i)} = randi(d)$;
end for
repeat
  for $i = 1 \rightarrow n$ **do**
    Obtain measured values and update reward table r^i according to (26);
    Update action values $Q^{(i)}(s^{(i)},:)$ according to (8);
    Select action $a^{(i)*} = \arg\max(Q^i(s^{(i)}, a^{(i)}))$;
    Update current state $s^{(i)} \overset{a^{(i)}}{\leftarrow} a^{(i)*}$;
    Update U_i's position $\left(x_j^{(i)}, y_j^{(i)}, z^{(i)}\right)$ according to (1) and (2);
    Compute U_i's trajectory according to (9);
  end for
  $j = j + 1$;
  Obtain candidate set $\left\{p^{(i,k)}\right\}(i,k \in [1,n], i < k)$ and weight set $\left\{w^{(i,k)}\right\}(i,k \in [1,n], i < k)$ according to (11) and (12) for two-stage clustering based localization;
  do two-stage clustering (which is shown in **Algorithm 2**);
  Obtain operation command f;
until $f = true$

Two-Stage Clustering Algorithm. The proposed two-stage clustering algorithm is performed provided the UAV's reward $r^{(i)}(s_j^{(i)},:)$ (which collects measured power of signals in directions of $\theta_1, \ldots, \theta_u$) satisfies a constraint in the coefficient of variation:

$$\frac{\overline{r^{(i)}(s_j^{(i)},:)}}{\sigma(r^{(i)}(s_j^{(i)},:))} < \lambda, \tag{14}$$

where $\sigma(r^{(i)}(s_j^{(i)},:))$ and $\overline{r^{(i)}(s_j^{(i)},:)}$ represent the standard deviation and the mean of $r^{(i)}(s_j^{(i)},:)$, respectively. The coefficient of variation i.e. the ratio of $\frac{\overline{r^{(i)}(s_j^{(i)},:)}}{\sigma(r^{(i)}(s_j^{(i)},:))}$ evaluates volatility of $r^{(i)}(s_j^{(i)},:)$ at time interval j. Given a small value of λ, a certain $r^{(i)}(s_j^{(i)},:)$ satisfying (14) means that $r^{(i)}(s_j^{(i)},:)$ suffers less

noise or multi-path interference. As a consequence, the direction (i.e. action $a^{(i)*}$) of flight obtained by the reinforcement-learning-based searching algorithm can be (nearly) aligned with the direction of the interference. Therefore, at a certain time interval j, only in the case where both $r^{(i)}(s_j^{(i)},:)$ and $r^{(k)}(s_j^{(k)},:)$ (of UAVs U_i and U_k, respectively) satisfy (14), can the intersection $p^{(i,k)}$ resulted from extensions of the trajectories of UAV U_i and U_k be collected in the set C and reliable for localization.

Prior to the clustering, in each time interval, an estimate of the position of the interference source is obtained. At a certain time interval j, the estimate can be achieved by

$$P_j = \frac{1}{|C|} \sum_{(i,k)\in C} p^{(i,k)} . \tag{15}$$

Each P_j is associated with a weight ω_j, which can be obtained by

$$\omega_j = \frac{1}{1 + e^{-\frac{1}{N} \sum\limits_{(i,k)\in C} w^{(i,k)}+1}}, \tag{16}$$

where $w^{(i,k)}$ is obtained by computing (12). Equation (16) suggests that ω_j is essentially a normalized weight achieved by processing $\frac{1}{N}\sum_{(i,k)\in C} w^{(i,k)}$ with a sigmoid function. By applying the sigmoid function, the value of the $w^{(i,k)}$ can be limited in the range of $(0.5, 1)$. We define sets C' and W' to respectively collect P_j (which is obtained by (15)) and ω_j (which is obtained by (16)) computed in each time interval j.

Weighted Clustering. In order to perform weighted clustering, a weighted distance matrix M is firstly defined and computed with elements in sets C' and W'.

$$M = \begin{pmatrix} \frac{||P_\varphi - P_\varphi||}{\sqrt{\omega_\varphi \cdot \omega_\varphi}} & \cdots & \frac{||P_\varphi - P_\eta||}{\sqrt{\omega_\varphi \cdot \omega_\eta}} \\ \vdots & \ddots & \vdots \\ \frac{||P_\eta - P_\varphi||}{\sqrt{\omega_\eta \cdot \omega_\eta}} & \cdots & \frac{||P_\eta - P_\eta||}{\sqrt{\omega_\eta \cdot \omega_\varphi}} \end{pmatrix}, \tag{17}$$

where $P_\varphi, P_\eta \in C'$ for $\varphi < \eta$. Row i (for $i \in [0, \eta - \varphi]$) of matrix M collects distances between element a certain $P_i \in C'$ and all the elements in C'. For instance, as shown in the first row (i.e. $i = 1$) of matrix M in (17), P_1 w.r.t. row 1 is equal to P_φ. Elements in each row of M is arranged in ascending order. In order to evaluate the possibility that P_i $\forall i$ is equal to the coordinate of the interference source, we define a metric for each row, i.e.

$$\varepsilon_i = \sum_{k=0}^{\lceil \phi \cdot (\eta - \varphi) \rceil} M(i,k), \tag{18}$$

where ϕ is a decimal. Equation (18) illustrates that for row i, ε_i is equal to the sum of elements in columns $0, 1, \ldots, \lceil \phi \cdot (\eta - \varphi) \rceil$. We perform (18) for each row. Then, according to the ordering of $\varepsilon_i \ \forall i$ (which are then sorted in ascending order), in the matrix M, we can find κ^l rows which correspond to the κ^l smallest ε_i values among all the ε_i values. By this means, we essentially find κ^l potential positions (or coordinates) of the interference source. We then compute the centroid of the κ^l potential coordinates by performing

$$P_j^{'} = \frac{1}{\kappa^l} \sum_{i=1}^{\kappa^l} P_i, \tag{19}$$

where P_i denotes the coordinate w.r.t. the ith smallest ε_i value. Moreover, $P_j^{'}$ is the result of weighted clustering at time interval j.

Secondary Clustering. In order to further improve the accuracy of localization, we propose the method of secondary clustering. In the proposed secondary clustering, a candidate queue q is defined. Once the weighted clustering achieves an estimate of the position $P_j^{'}$ of the interference source at time interval j, $P_j^{'}$ is collected in candidate queue q for secondary clustering with the size of k_q. The diagram of candidate queue q is shown in Fig. 5.

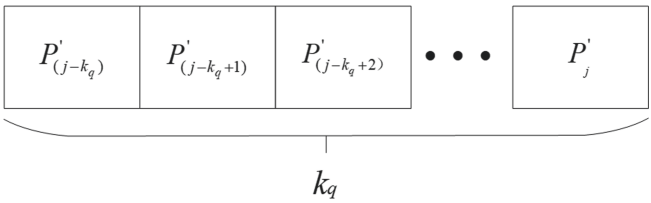


Fig. 5. Candidate queue collects outputs of the weighted clustering, where the maximum length of the queue is equal to k_q.

Suppose that the number of elements in candidate queue q is k_e. In the presence of $k_e < k_q$, $P_j^{'}$ achieved by the weighted clustering at time interval j can be directly pushed into queue q. Otherwise, the queue will firstly be popped before $P_j^{'}$ is pushed. Thus, candidate queue q can be viewed as a cluster of points obtained within the latest k_q intervals.

Similarly to matrix M in (17), a distance matrix $M^{'}$ is computed with elements in queue q. In the matrix, each element is equal to the distance between arbitrary two coordinates in candidate queue q. Matrix $M^{'}$ is obtained as

$$M^{'} = \begin{pmatrix} \left\| P^{'}_{(j-k_e)} - P^{'}_{(j-k_e)} \right\| & \cdots & \left\| P^{'}_{(j-k_e)} - P^{'}_{j} \right\| \\ \vdots & \ddots & \vdots \\ \left\| P^{'}_{j} - P^{'}_{(j-k_e)} \right\| & \cdots & \left\| P^{'}_{j} - P^{'}_{j} \right\| \end{pmatrix} \tag{20}$$

where k_e (for $k_e \leq k_q$) stands for the number of elements collected in queue q at time interval j. We then sort elements in each row of $M^{'}$ in ascending order. Similarly to (18), we define a metric to evaluate the possibility of coordinates in queue q being the position of the interference source. Such a metric can be defined as

$$\varepsilon^{'}_{i} = \sum_{k=0}^{\lfloor k_t \rfloor} M^{'}(i,k). \tag{21}$$

Equation (21) illustrates that for row i of matrix $M^{'}$, $\varepsilon^{'}_{i}$ is equal to the sum of elements in columns $0, 1, \ldots, \lfloor k_t \rfloor$. By performing (21) for each row, we can find the $\kappa^{'}$ smallest $\varepsilon^{'}_{i}$ values. According to the definition of $M^{'}$, each one of the $\kappa^{'}$ smallest $\varepsilon^{'}_{i}$ values corresponds to a coordinate of $P^{'}_{i}$ in the candidate queue q. Hence, by finding the $\kappa^{'}$ smallest $\varepsilon^{'}_{i}$, we basically find $\kappa^{'}$ coordinates which are close to each other and therefore can be used to estimate the position of the interference source. We then compute the centroid of the $\kappa^{'}$ coordinates, yielding

$$P^{T}_{j} = \frac{1}{\kappa^{'}} \sum_{i=1}^{\kappa^{'}} P^{'}_{i}. \tag{22}$$

Hence, P^{T}_{j} is the result of secondary clustering at time interval j. P^{T}_{j} is selected as estimated coordinate of the position of the interference source when P^{T}_{j} converges over iterations (i.e. time intervals). In this work, if the standard deviation σ w.r.t. the latest three estimated coordinates $\left\{ P^{T}_{j}, P^{T}_{(j-1)}, P^{T}_{(j-2)} \right\}$ satisfies

$$\sigma\left(\left\{ \left\| P^{T}_{j} - P^{T}_{(j-1)} \right\|, \left\| P^{T}_{j} - P^{T}_{(j-2)} \right\|, \left\| P^{T}_{(j-1)} - P^{T}_{(j-2)} \right\| \right\} \right) \leqslant \beta\,, \tag{23}$$

the collaborative search and localization terminates, and P^{T}_{j} is output as an estimate of the position of the interference source. The proposed two-stage clustering is summarized in Algorithm 2.

Algorithm 2. Two-stage Clustering Algorithm

Obtain the result of the reinforcement learning: candidate set $\left\{p^{(i,k)}\right\}(i,k \in [1,n], i<k)$ and weight set $\left\{w^{(i,k)}\right\}(i,k \in [1,n], i<k)$;
Obtain current iteration number j and operation command f;
for $i=1 \rightarrow n$, $k=i+1 \rightarrow n$ **do**
  if UAV U_i and U_k satisfy (14); **then**
    $C = C \cup \{(i,k)\}$;
    $flag = true$;
  end if
end for
if $flag$ **then**
  Compute P_j and w_j by performing (15) and (16);
  $C^{'} = C^{'} \cup P_j, W^{'} = W^{'} \cup \omega_j$;
  Compute the weighted distance matrix M by performing (17);
  Sort elements in each row of M in ascending order;
  Evaluate each row's metric ε_i through (18);
  Find the κ^l smallest metric avlues ε_i and the corresponding κ^l coordinates belonging to C';
  Obtain the output $P_j^{'}$ of the weighted clustering by performing (19);
  if $length(q) = k_q$ **then**
    Pop candidate queue q;
  end if
  Push $P_j^{'}$ into candidate queue q;
  Form distance matrix $M^{'}$ by performing (20);
  Sort elements in each row of $M^{'}$ in ascending order;
  Evaluate each row's metric $\varepsilon_i^{'}$ through (21);
  Find the $\kappa^{'}$ smallest metric avlues $\varepsilon_i^{'}$ and the corresponding $\kappa^{'}$ coordinates belonging to queue q;
  Obtain the output P_j^T of the secondary clustering by performing (22);
end if
if P_j^T satisfies the stopping criterion (23) **then**
  $f = true$;
end if

4 Simulation Results

In the simulations, we consider an interference source located at (5000 m, 2885 m, 0 m) with transmit power of 20 W and a swarm of three UAVs. Initial positions of the UAVs are set as (0 m, 0 m, 200 m), (5100 m, 8660 m, 190 m) and (10000 m, 0 m, 190 m), respectively. Each UAV is equipped with an electronic scanning antenna to locate the interference source. The electronic scanning antennas have the same radiation characteristic (as shown in Fig. 6), which is given by

$$F(\theta) = \cos(\frac{\pi}{2}\sin\theta) \cdot \cos(\pi \sin\theta) \cdot \cos[\frac{\pi}{4}(\cos\theta - 1)]. \tag{24}$$

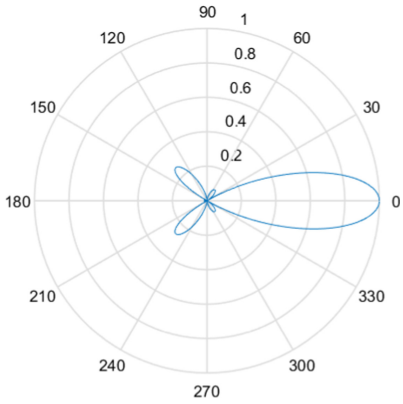


Fig. 6. Horizontal direction characteristics of the antennas.

Therefore, the receive gain $G_R(\theta)$ of each electronic scanning antenna for the horizontal angle θ and elevation angle φ is represented as follows:

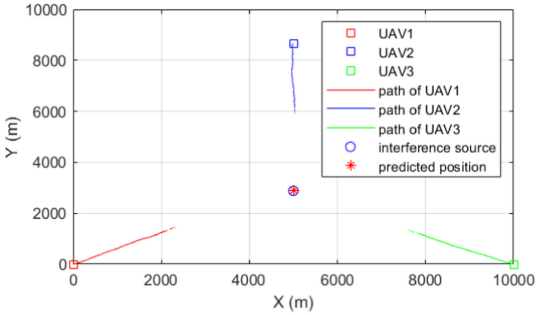
$$G_R(\theta) = \frac{4\pi\eta F^2(\theta)}{\int_0^{\frac{\pi}{2}} \int_0^{2\pi} F^2(\theta) \sin\varphi d\varphi d\theta}, \tag{25}$$

where $\varphi \in [0, \frac{\pi}{2}]$, $\theta \in [0, 2\pi)$ and the antenna efficiency $\eta = 1$. Thus, received power at each antenna can be obtained as

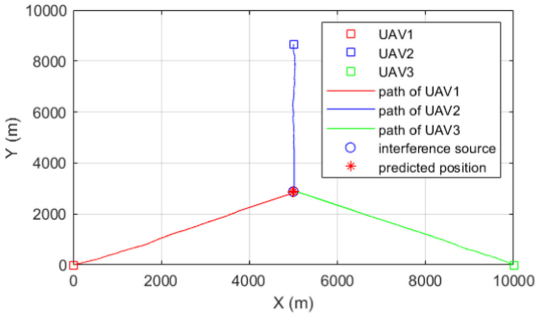
$$P_R(\theta) = \frac{P_T G_T G_R(\theta) \lambda^2}{(4\pi)^2 d^2 L} + n^2, \tag{26}$$

where P_T and G_T represent transmit power of the interference source and transmit antenna gain, respectively. The wave length λ is set as 3 m; the loss factor L is set as 1; n^2 stands for power of noise signal, which is a random variable with an average value of -38 dBm. In Algorithm 1, the learning rate α and discount factor γ in (8) are set as 0.9 and 0.1, respectively. In the two-stage clustering, λ in (14) is set as 2, and k_q for secondary clustering in Algorithm 2 is set as 30.

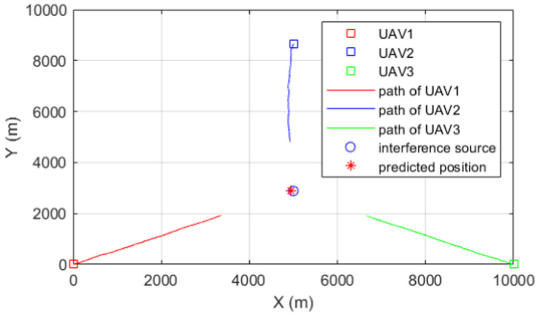
Two baseline schemes are considered in our study. The first one performs searching based on the reinforcement learning algorithm proposed in Sect. 3.1. This baseline is designated as Q-learning for simplicity in the subsequent discussions. The other scheme integrates Maximum-dimensional-based Search (MRS) with Two-Stage-Clustering-based Localization (TSCL). That is, similarly to the proposed collaborative search and localization, this baseline scheme consists of a search phase and a localization phase. However, in each time interval of the search, each UAV measures power of signals in different directions. Then, the UAV flies towards the direction w.r.t. the maximum measured power. In the localization phase, the UAV performs two-stage clustering as in Algorithm 2. This baseline scheme is then designated as MRS-TSCL.



(a) The proposed collaborative search and location.



(b) Q-learning



(c) MRS-TSCL

Fig. 7. UAVs' trajectories achieved by different approaches.

Figure 7 depicts UAVs' trajectories achieved by different approaches. The comparison of Fig. 7(a) and Fig. 7(b) indicates that although Q-learning-based search can find the interference source, the proposed collaborative search and location based on reinforcement learning and two-stage clustering can be more time-efficient, in terms of locating the interference source.

Additionally, it can be seen that compared to the proposed collaborative search and location scheme, MRS-TSCL suffers from a longer flight distance and degradation in the accuracy of localization. This suggests that the proposed reinforcement learning as in Algorithm 1 can benefit the localization phase, in terms of the accuracy of localization.

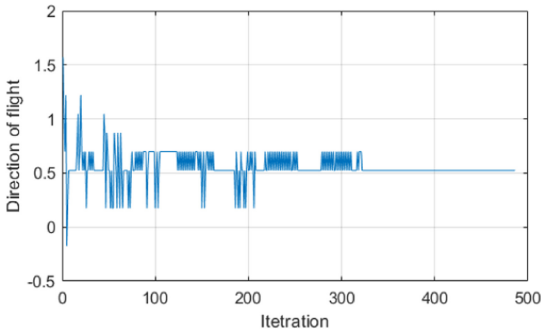


Fig. 8. Directions of flight as a function of number of iterations.

Figure 8 investigates directions (which are horizontal angles) of flight as a function of iterations. It is shown that the direction of flight w.r.t. a certain UAV converges as the number of iterations increases. This confirms the convergence of the proposed reinforcement learning. Moreover, it can be seen from Fig. 8 that the proposed reinforcement learning algorithm terminates before 500 iterations, since the stopping criterion has been reached. This implies that the prerequisite of the two-stage clustering algorithm termination is the convergence of the proposed reinforcement learning approach.

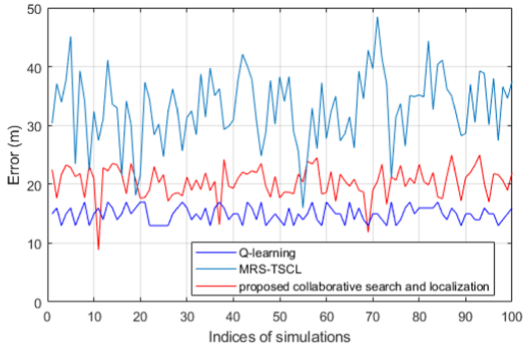


Fig. 9. Localization error of different approaches.

Figure 9 compares the proposed collaborative search and localization approach to baseline schemes, in terms of localization error. As shown in Fig. 9, we repeat simulation 100 times. The numerical results indicate that the localization error achieved by the proposed collaborative search and localization approach, Q-learning and MRS-TSCL is 20.39 m, 14.94 m and 32.98 m, respectively. It is worth noting that although the Q-learning approach yields the best localization error performance, it requires the UAVs to keep searching until arriving at the interference source and therefore degrades time efficiency. Additionally, it can be found that the most volatile localization error performance is achieved by MRS-TSCL. Hence, we can draw the conclusion that the proposed approach can well balance the accuracy of localization, time efficiency of searching and the robustness of localization.

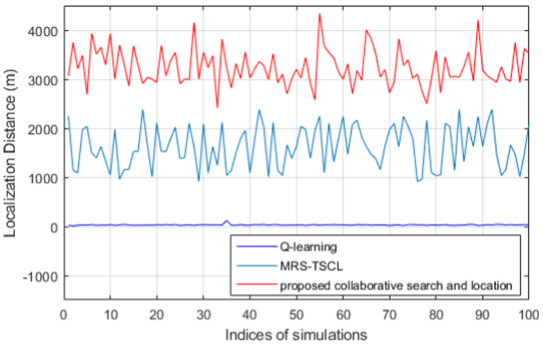


Fig. 10. Localization distance of different approaches.

Figure 10 studies the localization distance, which is the distance between the interference source and the position (of a UAV) where the UAV can locate the interference source. Figure 10 illustrates that by applying the proposed collaborative search and location approach, the UAVs can locate the interference source at an average distance of 3225.98 m, while the distance for MRS-TSCL is around 1687.64 m. Nevertheless, the Q-learning approach has to require the UAVs to arrive at the interference source, due to the absence of a remote localization method.

In Fig. 11, we investigate number of iterations required for localization. It can be seen from Fig. 11 that the proposed collaborative search and localization approach can locate the interference source with the smallest number of iterations, compared to that of the baselines. Figure 11 illustrates that the Q-learning approach requires an average of about 733 iterations, while the average number of iterations for MRS-TSCL is 584. The proposed collaborative search and localization approach, on the other hand, requires only an average of about 378 iterations.

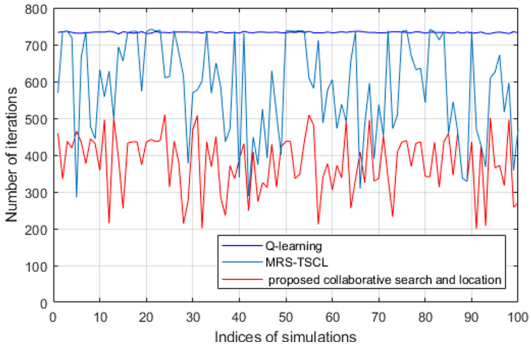


Fig. 11. Number of iterations required for localization.

5 Conclusion

In this paper, we have proposed a novel collaborative search and localization scheme, which exploits a swarm of UAVs to locate an interference source. The proposed approach iteratively performs search and remote localization. In the search phase of each time interval, a computationally efficient reinforcement learning algorithm is proposed to decide the trajectory of each UAV. In the following localization phase of the time interval, a two-stage clustering algorithm has been developed to process the intersections of the extensions of UAVs' trajectories, so as to estimate the position of the interference source. Simulation results have revealed that the proposed approach can accurately locate the position of a interference source from a distance, while the conventional Q-learning can achieve slightly higher localization accuracy at a significant cost of time efficiency. Moreover, by integrating reinforcement learning with two-stage clustering, the accuracy of remote localization can be improved.

References

1. Bayerlein, H., Kerret, P., Gesbert, D.: Trajectory optimization for autonomous flying base station via reinforcement learning, pp. 1–5, June 2018. https://doi.org/10.1109/SPAWC.2018.8445768
2. Jiang, C., Zhu, X.: Reinforcement learning based capacity management in multilayer satellite networks. IEEE Trans. Wireless Commun. **19**(7), 4685–4699 (2020)
3. Lemic, F., Bsch, J., Chwalisz, M., Handziski, V., Wolisz, A.: Infrastructure for benchmarking RF-based indoor localization under controlled interference. In: 2014 Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), pp. 26–35, November 2014. https://doi.org/10.1109/UPINLBS.2014.7033707
4. Macqueen, J.: Some methods for classification and analysis of multivariate observations. In: Proceedings of Berkeley Symposium on Mathematical Statistics & Probability (1965)
5. Maeda, K., Doki, S., Funabora, Y., Doki, K.: Flight path planning of multiple UAVs for robust localization near infrastructure facilities. In: IECON 2018–44th Annual Conference of the IEEE Industrial Electronics Society, pp. 2522–2527 (2018)

6. Pack, D.J., DeLima, P., Toussaint, G.J., York, G.: Cooperative control of UAVs for localization of intermittently emitting mobile targets. IEEE Trans. Syst. Man Cybern. Part B (Cybern.) **39**(4), 959–970 (2009). https://doi.org/10.1109/TSMCB.2008.2010865
7. Powell, W.B.: Approximate Dynamic Programming: Solving the Curses of Dimensionality, vol. 703. Wiley, Hoboken (2007)
8. Puterman, M.L.: Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley, Hoboken (1994)
9. Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction (1998)
10. Weinberger, K.Q.: Distance metric learning for large margin nearest neighbor classification. JMLR **10** (2009)
11. Wu, G.: UAV-based interference source localization: a multimodal Q-learning approach. IEEE Access **7**, 137982–137991 (2019)
12. Wu, S.: Illegal radio station localization with UAV-based Q-learning. China Commun. **15**(12), 122–131 (2018)