Advances in Computer, Signals and Systems (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2024.080701
ISSN 2371-8838 Vol. 8 Num. 7

Parametric Transfer-Based DQN for Multi-Function Radar Jamming Decision Method

Lihui Huang*, Changhua Hu

Missile Engineering Institute, Rocket Force University of Engineering, Xi'an, China
huanglihui1996@163.com
**Corresponding author*

Keywords: Multifunctional Radar; Jamming Decision-Making; DQN

Abstract: With the continuous development of multi-function radar technology, the number of radar tasks the seeker can perform is increasing. This has led to the environment state transitioning from a small space to an ample space, facing more complex radar jamming decision problems. Traditional reinforcement learning algorithms have insufficient processing capacity and limited learning ability, thus we adopted a deep reinforcement learning algorithm, combining its powerful perception and processing capabilities to improve the jamming effect further. At the same time, to solve the problem of low computational efficiency for deep reinforcement learning, the transfer learning algorithm is introduced by migrating the parameters of deep learning networks from other tasks to the radar seeker jamming decision, further improving the learning rate.

1. Introduction

Radar jamming decision-making (RJDM) is a crucial link in cognitive electronic warfare [1][2], and the outcome of the confrontation between radar equipment and jamming equipment may determine the victory of a war. With the continuous development of multi-functional radar and cognitive electronic technology, radar's working states and signal patterns have increased significantly [3], decreasing the accuracy and efficiency of traditional radar-jamming decision-making algorithms. With the extensive application of artificial intelligence, reinforcement learning algorithms have been employed to address the issue of radar jamming decision-making. The work in [4] and [5] introduced the reinforcement learning (RL) method into RJDM by optimizing traditional analysis index parameters and selecting new evaluation criteria to construct the model. This new model achieves comprehensive, multi-dimensional, real-time data analysis and conducts multi-functional radar simulations based on RL algorithms. B. Siwei proposed an RJDM method that combines random forests with neural network algorithms [6]. This method first defines the information parameters of the jamming system acquired in real time and analyzes the jamming patterns to summarize the signal parameters affected by the target. These parameters can form a set of indicators obtainable through surveillance, which reflect the information extraction and summary of the jamming purpose and target changes. [7], [8] and [9] adopted various methods, including Q-learning, deep Q-learning (DQN), and their improved versions. M. Shaoqing also conducted in-depth studies on Q-learning and its improved algorithms [10]. Therefore, it can be observed that methods

based on reinforcement learning and deep reinforcement learning have favourable effects on the RJDM problem in future electronic warfare.

However, the issues above related to radar jamming decision-making are limited to functional-level simulations, where the relationship between jamming patterns and radar states is described qualitatively for conducting jamming decisions without being able to quantitatively determine the impact of jamming on radar operation. Furthermore, most algorithms ignores the task's real-time requirements. Introducing deep reinforcement learning can address tasks with large spaces and multiple states. Still, with the addition of deep neural networks, the parameters within the network increase significantly, and the training time elongates. To address the above problems, this paper constructs a model of radar jamming decision-making at the signal level. The accuracy of signal-level modelling of the radar seeker is demonstrated through simulation experiments without applying to jam and observing the missile landing positions. Simultaneously, a DQN method combined with transfer learning is proposed. This method accelerates the convergence speed of the network by transferring the trained parameters from the LunarLander-v2 task in the gym.The main structure of this paper is arranged as follows: the first section builds an adversarial model, the second section introduces the DQN algorithm with transfer learning, the third section conducts simulation experiments, and the fourth section concludes the work of this paper.

2. Construction of Adversarial Models

2.1 Modeling of Multi-Function radar seeker

2.1.1 Signal transmission modeling

Typical multifunction radar transmits signals in the form of LFM (Linear Frequency Modulation) signal pattern, which can be represented as:

$$S_t(t) = \sqrt{\frac{2P_t}{4\pi L_t}} g_{vt}(\theta) v(t) \cdot e^{j\omega_{ck}t} \omega_{ck} \tag{1}$$

Where ω_{ck} denotes the current pulse carrier frequency, P_t indicates the peak power of the transmitter, L_t stands for the combined loss of the transmitter, $g_{vt}(\theta)$ represents the transmitting antenna pattern (voltage gain), and $v(t)$ represents the complex modulation function as follows:

$$v(t) = \mathrm{Rect}\left(\frac{t}{T_p}\right) \cdot \exp\left(j\pi F_m t^2\right) \tag{2}$$

where T_p represents the pulse width, and F_m represents the frequency modulation slope. Based on the above formulas, the coherent video signal pattern adopted in the system is as follows:

$$s_t(t) = \sqrt{\frac{P_t}{4\pi L_t}} g_{vt}(\theta) \cdot \mathrm{Rect}\left(\frac{t}{T_p}\right) \cdot \exp\left(j\pi \frac{BW_{rg}}{T_p} t^2\right) \tag{3}$$

2.1.2 Receiving Signal Model

The received signal mainly consists of target echoes, interference signals, various types of clutter, and receiver noise. This paper mainly considers the aspects of target echoes and interference signals.

With respect to a specific transmitted pulse, the RF signal received by the radar can be represented as follows:

$$r_{RF}(t) = S_{RF}(t) + J_{RF}(t) + n_{RF}(t) \tag{4}$$

where $S_{RF}(t)$ represents the echo signal received after the transmitted pulse is reflected by the target, $J_{RF}(t)$ represents the received interference signal, which is the combined interference signal formed by various active interference and passive interference, and $n_{RF}(t)$ represents the receiver noise. The band-limited noise signal is represented as:

$$n(t) = \mathrm{Re}\left[\tilde{n}(t) \cdot e^{j\omega_c t}\right] \tag{5}$$

Therefore, in coherent video simulation, the noise at the receiver can be represented as

$$\tilde{n}(t) = n_d(t) - jn_q(t) \tag{6}$$

In this model, $n_d(t)$ and $n_q(t)$ are independent Gaussian random processes with zero mean and variance σ_N^2. The variance of the noise σ_N^2 can be calculated from the receiver noise coefficient N_F and receiver bandwidth Δf as follows:

$$\sigma_N^2 = kT_0 N_F \Delta f \tag{7}$$

where K is the Boltzmann constant, and T_0 is the reference temperature of the receiver, which $T_0 = 290\mathrm{K}$. Combining the target echo signal and the receiver thermal noise signal, we finally obtain the radar received signal as:

$$\begin{aligned} r(t) = \mathrm{Rect}\left(\frac{t - \frac{2R}{c}}{T_p}\right) \cdot \left[\sqrt{\frac{2P_t}{(4\pi)^3 L}} \frac{g_{vt}(\theta) g_{vr}(\theta)}{R^2} \lambda_k \sqrt{\sigma}\right] \\ \cdot \exp\left[j\pi \frac{BW_{rg}}{T_p}\left(t - \frac{2R}{c}\right)^2\right] \cdot \exp\left[2\pi f_d t - 2\pi \frac{2R}{\lambda_k}\right] + \tilde{n}(t) \end{aligned} \tag{8}$$

2.1.3 Pulse compression

In the commonly used radars at the present stage, a matched filter is typically employed to carry out pulse compression processing of signals. Suppose the transfer function of the matched filter is $H(f)$, the impulse response is $h(t)$, and the input is $s(t)$.

$s_0(t)$ is the output result of the target signal after it has passed through a matched filter, which can be expressed as follows:

$$s_0(t) = \int_{-\infty}^{\infty} H(f) S(f) e^{j2\pi ft} df \tag{9}$$

Taking the inverse Fourier transform of the transfer function $H(f)$ results in its impulse response

function as:

$$h(t) = ks^*\left(t_0 - t\right) \tag{10}$$

For the sampled digital signal, denoting the input signal after sampling and quantization as $s(n)$, the unit impulse response of the matched filter can be represented as:

$$h(n) = s^*(n) \tag{11}$$

2.1.4 Constant False Alarm Handling and Detection Model

Since the signals received by the radar comprise both target echo signals and clutter signals, not all processing steps in the signal processor can completely filter out the clutter signals. Thus, a threshold value is often set in the radar. The portions of the signal that are higher than the threshold value are retained, while those lower than the threshold value are filtered out. This threshold value is the false alarm probability. To achieve this aim, the false alarm threshold must be calculated in real-time based on the received signal to adjust the radar detection threshold accordingly to obtain the desired false alarm probability [11].

2.1.5 Measurement Information Extraction

2.1.5.1 Extraction of distance information

The target echo signal will have a time delay t_r due to the distance between the target and the radar, which can be expressed as $t(r) = 2R/c$. R represents the relative distance between the target and the radar, and c represents the speed of light. Thus, in the case of a known time delay, the distance between the target and the radar can be inversely deduced based on the time delay t_r of the echo signal.

2.1.5.2 Extraction of angular information

It has been mentioned above that since the radar adopts the sum and difference beam angle measurement method when measuring the angle, the pitch angle θ and the yaw angle ϕ of the target relative to the radar can be obtained based on the amplitude of the corresponding position of the target in the processed sum beam signal and the amplitudes of the corresponding positions of the target in the pitch difference beam signal and the yaw difference beam signal:

$$\theta = \frac{\Delta F(\delta)}{\Sigma_\theta \left.\frac{dF(\theta)}{d\theta}\right|_{\theta=\delta}}, \quad \phi = \frac{\Delta F(\delta)}{\Sigma_\phi \left.\frac{dF(\phi)}{d\theta}\right|_{\phi=\delta}} \tag{12}$$

2.1.5.3 Extraction of velocity information

If relative motion exists between the target and the radar, the frequency of the target's echo will undergo changes. The Doppler frequency shift f_d caused by the relative velocity can be expressed as $f_d = 2v_r/\lambda$. In the case where the Doppler frequency shift of the echo signal is known, the relative velocity between the target and the radar can be inversely deduced.

2.1.6 Guidance method

The proportional guidance method refers to the fact that the angular velocity of the missile's speed-changing direction in space is proportional to the angular velocity of the target's relative radar position angle rotation:

$$\Delta\theta_M = k' \cdot \Delta\theta_T \tag{13}$$

where $\Delta\theta_M$ denotes the angular velocity of the change in the direction of the missile's velocity in space, k' is the proportional coefficient, and $\Delta\theta_T$ represents the angular velocity of the rotation of the target's relative position angle to the radar, also known as the line-of-sight angular velocity.

2.2 Environmental description

In the model, the environment is the target radar seeker, and the seeker's observable states are composed of 8 variables: the horizontal coordinate on the ground plane, the vertical coordinate on the ground plane, altitude, the x-component of speed, the y-component of speed, the z-component of speed, roll angle, and pitch angle. After establishing the signal-level simulation model of the radar seeker, we can simplify the interference decision problem to the diagram shown in Figure 1.

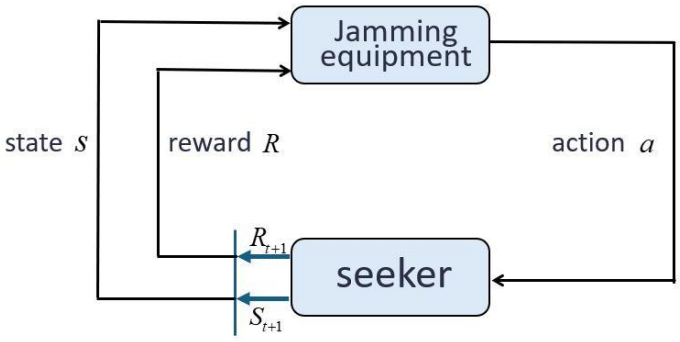

Figure 1: Diagram of radar seeker and jamming equipment countermeasures.

In this process, S represents the radar seeker's speed and position information at the current simulation time. After the jamming device takes action a_t, the radar seeker processes the echo information with added jamming signals, obtains the target's state information, and then outputs the seeker's speed and position information at the next simulation time according to the guidance rule. Meanwhile, it feeds back the reward function value R_{t+1} under the influence of jamming.

2.3 Action description

2.3.1 Amplitude Modulation jamming

Radio frequency noise interference refers to the direct amplification of microwave noise that is emitted, and the mathematical model of this interference is represented as:

$$J(t) = U_n(t)\cos\left[2\pi f_i t + \varphi(t)\right] \tag{14}$$

The envelope $U_n(t)$ follows a Rayleigh distribution, the phase $\varphi(t)$ follows a uniform distribution over [0, 2π] and is independent of $U_n(t)$. The carrier frequency f_i is a constant value and is much larger than the $J(t)$ bandwidth. Since $J(t)$ is generally used for filtering and amplifying low-power noise, it is also known as direct-amplified noise.

2.3.2 Frequency-modulated noise jamming

Frequency-modulated noise jamming is currently the most widely used type of suppressive jamming signal, which has a wide interference bandwidth and is easy to achieve a large noise power. The mathematical model of the time-domain characteristic of frequency-modulated noise jamming is expressed as:

$$J(t) = U_0 \cos\left[2\pi f_j t + 2\pi K_{FM} \int u\left(t^{'}\right) dt^{'} + \varphi \right] \tag{15}$$

$u(t)$ is the frequency-modulated noise, which is a wide-sense stationary random process with zero mean, and φ is a random variable that follows a uniform distribution on the interval [0, 2π] and is independent of $u(t)$. U_0 is the amplitude of the interference, f_j is the center frequency of the interference, and K_{FM} is the frequency-modulation slope.

2.3.3 Agile noise jamming

This is a new hybrid approach that combines the characteristics of deception and noise jamming. Its core idea is to combine forwarding-style jamming with random pulse jamming. The Agile noise jamming expression based on FM noise is expressed as:

$$J(t) = U_0 \cos\left[2\pi f_j t + 2\pi K_{FM} \int u\left(t^{'}\right) dt^{'} + \varphi \right] \cdot e^{-j\pi k t^2} \operatorname{rect}\left(\frac{t}{T}\right) \tag{16}$$

2.4 Reward Function Building

In the signal-level simulation model built in this paper, we quantitatively construct the reward function based on the measurement error and the probability of detecting the ship at each moment of the radar. The first part is the reward function for the measurement error caused by the interference, which is the difference between the measurement information output by the radar seeker based on the echo signal and the actual information of the target; the more significant the difference, the better the effect of interference, the measurement information includes the distance, speed and angle information of the target; the second part is the detection probability of detecting the interference device by the seeker, of preventing the missile from detecting and detecting the interference device, the on-time of each interference device cannot be too early, too early on will make the target more obvious, so the on-time is also taken into account in the reward function. The reward function can be set as follows:

$$r = k_1 \Delta L + k_2 \Delta v + k_3 \Delta\theta + k_4 \Delta\varphi + b_1 T_{\text{on }1} + b_2 T_{\text{on }2} + b_3 T_{\text{on }3} \tag{17}$$

Where Δv denotes the speed error value, ΔL denotes the distance error value, $\Delta\theta$ denotes the pitch angle error value, $\Delta\varphi$ denotes the yaw angle error value, $T_{\text{on }1}$, $T_{\text{on }2}$, $T_{\text{on }3}$ denote the initial

times of the three kinds of jamming patterns, and k_1, k_2, k_3, b_1, b_2, b_3 denote the proportional coefficients. After extensive experiments, a set of relatively reasonable values can be obtained as follows: k_1, k_2, k_3, b_1, b_2, b_3 = 0.02, 0.1, 20, 20, 0.2, 0.1, 0.1.

3. Parametric Transfer-Based DQN

3.1 DQN

DQN is a combination of Q-learning algorithm and deep learning [12]. The Q-learning algorithm updates the state-action value function for each state using equation (27).

$$Q(s,a) = r + \gamma Q^*(s',a') \tag{18}$$

Where s' represents the next state of the agent, a represents the action taken in that state, and r represents the reward obtained from the state transition. DQN uses a deep learning network to output a value function for all actions in a given state, serving as the deep learning network's evaluation of the Q value. When updating the network, some small segments of information $(\phi_t, a_t, r_t, \phi_{t+1})$ from the experience pool are taken out as samples, and the label values used in equation (27) and equation (28) are used as the network's label values.

$$y_i = \begin{cases} r_j & ,s \to s_T \\ r_j + \gamma \max\limits_{a'} Q(\phi_{j+1}, a'; \theta), otherwise \end{cases} \tag{19}$$

By constructing a loss function $Loss(\theta) = (y_i - Q'(s,a))^2$, the network parameters are updated using gradient descent algorithm.

3.2 Transfer learning

As an advanced machine learning method, transfer learning aims to effectively handle related but different issues by utilizing the knowledge acquired in one domain, thereby improving learning efficiency and generalization ability. This approach not only speeds up the model training process but also improves the performance of the target domain model. For example, suppose a deep neural network has been trained on a large-scale cat and dog classification task. In that case, it can be adjusted and applied to other animal categories without starting from scratch for training, thus saving time and resources [13].

3.3 Parametric Transfer-Based DQN

We can build the same network structure and train the network in a simple task (LunarLander-v2) to obtain converged parameters and then migrate the initial parameters to the complex task (RJMD) to accelerate the network training speed. The deep learning network structure adopted in this paper is as follows: input layer: 8 neurons for the eight states of the environment; hidden layer 1: 128 neurons, using the ReLU activation function; hidden layer 2: 64 neurons, using the ReLU activation function; output layer: 3 neurons, corresponding to the three discrete actions. The DQN algorithm combined with transfer learning is named T-DQN. The flowchart of the T-DQN algorithm is shown in Figure 2.

initialize reply memory D to capacity N, action-value function Q with parametric transfer weights
for each episode
    initialise sequence $s_1 = \{x_1\}$ and preprocessed sequenced $\phi_1 = \phi(s_1)$
    for t $= 1, \ldots, T$
        with probability ε select action a_t
        otherwise select $a_t = \max_a Q^*(\phi(s_t), a; \theta)$
        execute action a_t and observe reward r_t
        set $s_{t+1} = s_t$ and preprocess $\phi_{t+1} = \phi(s_{t+1})$
        store transition $(\phi_t, a_t, r_t, \phi_{t+1})$ in D
        sample random minibatch of transition $(\phi_t, a_t, r_t, \phi_{t+1})$ from D
        set $y_i = \begin{cases} ri & \text{for terminal } \phi_{j+1} \\ ri + \gamma \max_{a'} Q(\phi_{j+1}, a'; \theta) & \text{for non-terminal } \phi_{j+1} \end{cases}$
        perform a gradient descent step on $(y_i - Q(f_j, a_j; q))^2$
    end for
end for

Figure 2: The flowchart of the T-DQN algorithm.

4. Experiment

4.1 Parameter settings

The settings for each parameter are shown in Tables 1 to 5.

Table 1: Parameters for the radar seeker.

Project	Major Parameters
Transmitter	Center frequency:18GHZ
	Bandwidth:250MHz
	Maximum power:30KW
	Pulse width:100ns
Antenna	Beam width:0.2rad
	Angular range:(2/π ,-2/π)

Table 2: Parameters of the initial state of missiles and ships.

Parameters	missile	ships
magnitude of velocity	2500km/h	50km/h
direction vector of velocity	(0,1,0)	($\sqrt{2}$, $\sqrt{2}$,0)
location coordinates	(0,0,10km)	(0,0,8km)
radar Cross-Section	-	10

Table 3: Parameters of LFM Signal.

Parameters	Pulse width	Bandwidth	Pulse Repetition Period	Transmission frequency	Number of Pulse Emissions
Value	10us	10MHz	100us	15GHz	16

Table 4: Jamming parameters.

Types of jamming	Amplitude Modulation	Frequency-modulated	Noise Agile
Pulse width	50us	200us	10us
Bandwidth	40MHz	1MHz	30MHz
Center frequency	15GHz	15GHz	15GHz

Table 5: Parameters of DQN.

Parameters	Learning rate α	Discount factor γ	Capacity of D	Greed factor ε
Value	0.2	0.8	1000	0.5

4.2 Experimental results

This paper employs five jamming strategies, namely, jamming decision without jamming patterns, jamming decision with random jamming patterns, jamming decision based on Q-learning, jamming decision based on DQN and jamming decision based on T-DQN.

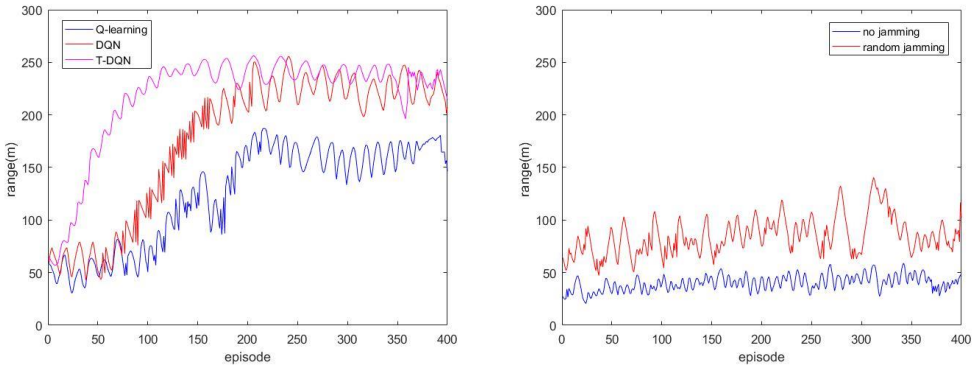

Figure 3: Diagram of Variation in Missile Landing Point Distance.

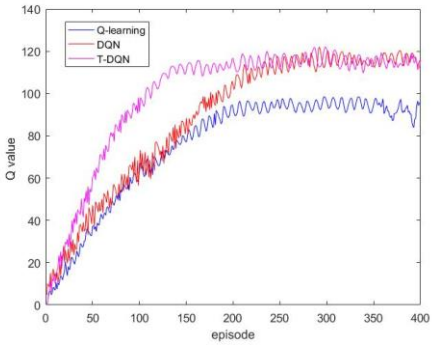

Figure 4: Accumulated Reward Value Change Curve.

From the Figure 3, when no interference measures are taken, the missile can cause damage to the target, and its landing point is within a range of 100 meters. This phenomenon indicates that the missile's guidance system performs well under normal circumstances. However, due to the noise in the received signal, there is a fluctuation in the landing distance of the missile. This fluctuation verifies the accuracy of the radar seeker signal level simulation, providing a reliable data basis for subsequent research. When a random interference strategy is adopted, it can be observed that this method has a particular impact on the guidance of the radar seeker. Still, the effect could be more

stable due to its large randomness. Specifically, random interference can deviate from the target successfully in some cases. Still, more specificity may be needed to reduce the hit rate effectively. Figure 4 shows the change in the cumulative reward value during the simulation.

It can be seen that the final converged value function of DQN is larger than that of Q-learning. After introducing transfer learning, the convergence speed of the algorithm is significantly improved, and the reward values are basically kept consistent. The final optimal jamming strategy obtained is presented in Table 6.

Table 6: The optimal jamming strategy.

Time	Q-learning	DQN	Time	Q-learning	DQN	Time	Q-learning	DQN
1	000	000	6	111	101	11	111	111
2	000	000	7	111	111	12	111	111
3	000	100	8	111	111	13	111	111
4	100	100	9	111	111	14	111	111
5	100	101	10	111	111	15	111	111

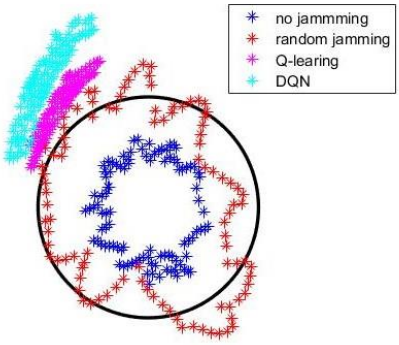

Figure 5: Distribution Map of Missile Landing Points.

From the Figure 5, it can be seen that the interference strategy obtained by using reinforcement learning significantly impacts the various damage indicators of the missile, making the missile's strike effect on the ship basically zero. This verifies the effectiveness of the reinforcement learning algorithm in the task of radar interference decision-making.

5. Conclusion

This paper first establishes a signal-level simulation model of the entire end-guidance process of radar seeker, including signal transmission, reception, signal processing, and measurement information output. Then, it conducts simulation experiments without interference, and the missile can hit the target with sure accuracy, verifying the accuracy of the established model. Based on this signal-level simulation model, the paper introduces a deep reinforcement learning algorithm for the interference decision of the radar seeker in the presence of interference equipment, significantly improving the efficiency and accuracy of the interference decision. At the same time, this paper integrates the idea of transfer learning, effectively transferring the neural network parameters trained on the LunarLander-v2 task to the interference decision task. This strategy enables the model to quickly adapt to new tasks and reduce the time and computing resources required for training from scratch. By leveraging existing knowledge, the algorithm further optimizes its performance in new environments, accelerating the convergence rate and ensuring the stability and reliability of the model

in actual applications.

References

[1] REICH G M, ANTONIOU M, BAKER C J. Memory-enhanced cognitive radar for autonomous navigation[J]. IET RADAR SONAR AND NAVIGATION, 2020, 14(9): 1287-1296. DOI: 10.1049/iet-rsn.2019.0409

[2] GURBUZ S Z, GRIFFITHS H D, CHARLISH A, et al. An overview of cognitive radar: Past, present, and future[J]. IEEE Aerospace and Electronic Systems Magazine, 2019, 34(12): 6-18. DOI: 10.1109/MAES.2019.2953762.

[3] LIANG Y C, CHEN K C, LI G Y, et al. Cognitive radio networking and communications: an overview[J]. IEEE Transactions on Vehicular Technology, 2011, 60(7): 3386-3407. DOI: 10.1109/TVT.2011.2158673.

[4] Y. F. Li, "Research on interference decision technology based on deep reinforcement learning," M.S. thesis, Xi'an Univ. Electron. Sci. Technol., Xi'an, China, 2020.

[5] L. Yuqian, "Radar cognitive behavior identification and interference decision optimization," M.S. thesis, Xi'an Univ. Electron. Sci. Technol., Xi'an, China, 2019.

[6] B. Siwei, "Based on Neural Network Effectiveness Evaluation of Radar Jamming," M.S. thesis, Xi'an Univ. Electron. Sci. Technol., Xi'an, China, 2020.

[7] Z. Baikai and Z. Weigang, "Multi-function radar cognitive jamming decision-making method based on Q-learning," Telecommun. Technol., vol. 60, no. 2, pp. 129–136, 2020.

[8] B. K. Zhang and W. G. Zhu, "A DQN cognitive interference decision method for multifunctional radars," Syst. Eng. Electron. Technol., vol. 42, no. 4, pp. 819–825, 2020.

[9] QIANG X, WEIGANG Z, XIN J. Research on method of intelligent radar confrontation based on reinforcement learning[C]//2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA). 2017: 471-475. DOI: 10.1109/CIAPP.2017.8167262.

[10] M. Shaoqing, "Research on intelligent jamming decision method based on reinforcement learning," Harbin Inst. Technol., 2021, doi: 10.27061/d.cnki.ghgdu.2021.003754.

[11] WEINBERG G V. Geometric mean switching constant false alarm rate detector[J]. DIGITAL SIGNAL PROCESSING, 2017, 69: 1-10. DOI: 10.1016/j.dsp.2017.06.015.

[12] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.

[13] LI G, WANG Z, GAO J, et al. Performance assessment of cross office building energy prediction in the same region using the domain adversarial transfer learning strategy[J]. Applied Thermal Engineering, 2024, 241: 122357.