

Article

# Jamming Strategy Optimization through Dual Q-Learning Model against Adaptive Radar

Hongdi Liu, Hongtao Zhang \*, Yuan He  and Yong Sun

Key Lab of Universal Wireless Communications, Ministry of Education of China, Beijing University of Posts and Telecommunications, Beijing 100876, China; liuhongdi@bupt.edu.cn (H.L.); yuanhe@bupt.edu.cn (Y.H.); sunyong@bupt.edu.cn (Y.S.)

\* Correspondence: htzhang@bupt.edu.cn

**Abstract:** Modern adaptive radars can switch work modes to perform various missions and simultaneously use pulse parameter agility in each mode to improve survivability, which leads to a multiplicative increase in the decision-making complexity and declining performance of the existing jamming methods. In this paper, a two-level jamming decision-making framework is developed, based on which a dual Q-learning (DQL) model is proposed to optimize the jamming strategy and a dynamic method for jamming effectiveness evaluation is designed to update the model. Specifically, the jamming procedure is modeled as a finite Markov decision process. On this basis, the high-dimensional jamming action space is disassembled into two low-dimensional subspaces containing jamming mode and pulse parameters respectively, then two specialized Q-learning models with interaction are built to obtain the optimal solution. Moreover, the jamming effectiveness is evaluated through indicator vector distance measuring to acquire the feedback for the DQL model, where indicators are dynamically weighted to adapt to the environment. The experiments demonstrate the advantage of the proposed method in learning radar joint strategy of mode switching and parameter agility, shown as improving the average jamming-to-signal ratio (JSR) by 4.05% while reducing the convergence time by 34.94% compared with the normal Q-learning method.

**Keywords:** adaptive radar; jamming strategy optimizing; reinforcement learning; Q-learning; jamming effectiveness evaluation



**Citation:** Liu, H.; Zhang, H.; He, Y.; Sun, Y. Jamming Strategy Optimization through Dual Q-Learning Model against Adaptive Radar. *Sensors* **2022**, *22*, 145. <https://doi.org/10.3390/s22010145>

Academic Editor: Ying-Ren Chien

Received: 9 November 2021

Accepted: 21 December 2021

Published: 26 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In radar jamming, accurate decision-making is an important prerequisite for effective jamming. For single-mode radar with only a few fixed parameter combinations, the jamming decision-making based on template matching can be efficient [1]. Nowadays, with the improvement of electronic technology, modern radars tend to be adaptive. Adaptive radars can perform various tasks through automatic work mode switching, where work modes vary with different radars. Take a ground-based radar with search, tracking, and recognition modes as an example: The radar initially scans the entire airspace at the search mode, and switches to the tracking mode when a mission-related target is detected, then transitions to the recognition mode after the target is confirmed. If the echo signal quality drops or the target is lost due to the jamming, the radar will take anti-jamming measures autonomously or return to the search mode. For phased array radars with modes named “range while scan” (RWS), “track and search” (TAS), “Single Target Track” (STT), etc., the adaptive switching strategy is more complicated. Additionally, in each work mode, the pulse parameters can be changed in real time to improve radar’s performance or survivability based on environment detecting [2,3]. In such context, the effectiveness of conventional jamming decision-making methods is decreasing because of a large shortage of prior knowledge about the radars [4]. Therefore, it is urgently necessary to develop radar jamming technology.

Inspired by the cognitive radio, the application of intelligent algorithms in radar confrontation became possible [1,5–7]. To solve the problem of the low rate of template

matching under the incomplete jamming rule library condition, a jamming decision-making method based on clustering and resampling–support vector machine is proposed in [1]. In [5], a discrete dynamic Bayesian network is established to guide decision-making in self-defense electronic jamming. In [6], an improved chaos genetic algorithm is applied to the allocation of interference strategy. In [7], a particle swarm optimization algorithm is applied to solving optimal jamming power allocation strategy aiming at cognitive MIMO radar. However, these traditional machine learning methods often require large amounts of tagged radar data acquired in advance, which are difficult to be obtained in actual scenarios.

Up to now, there have been many works on the application of reinforcement learning (RL) in communication jamming and anti-jamming [8–12], providing new ideas for radar confrontation. RL is a type of machine learning technology, where an agent learns from interacting with the environment and takes maximizing the feedback from the environment as its learning goal [13]. In [14], the framework of intelligent jamming based on RL is described, where the cognitive jammer and the radar are respectively regarded as the agent and the environment. In [4], the conversion between different radar working states is modeled, where an RL model is trained to choose the jamming mode against each radar state. A frequency-agile radar is considered in [15], and a jamming frequency selection algorithm based on Q-learning is proposed to solve the optimal frequency for each jamming pulse. In [16–19], other methods based on RL are used to optimize the strategy of combating jamming for radar. Compared with the traditional machine learning methods mentioned above, the agent in RL can learn with no tagged data needed, which makes it more adaptable to the unknown environment. The jamming system equipped with RL can obtain training samples during the jamming process, and update the jamming strategy dynamically based on the change of the radar signal.

However, in the existing work of jamming decision-making based on RL, two common measures of the adaptive radar including mode switching and parameter agility have not been jointly considered. For example, [14] focuses on macro-level modeling of radar jamming, abstracting radar modes into the environment state in RL. What is not noticed is that the jamming effectiveness will be weakened due to the agile actions such as frequency agility [16] and dynamic pulse repetition interval [20] taken by radar in each mode. In [15], the authors aim at the frequency hopping of the radar without considering multiple work modes. According to the authors of [3], if only static behavior of radar is considered, such as a single work mode, it is easy to result in a subjective or local optimal jamming strategy. Moreover, when facing the adaptive radar with multi-modes and the ability of parameter agility, a large jamming action space is usually needed to ensure that the correct actions are included, which greatly increases the complexity of jamming parameters generating. As high complexity will lead to a long convergence time, the jammer can hardly find the optimal jamming strategy in a limited time, which can be fatal to the protected target.

To overcome the problems above, a two-level jamming decision-making framework is developed in this paper. On this basis, a dual Q-learning (DQL) model is proposed to obtain the optimal jamming strategy. Specifically, the jamming decision-making process is disassembled into two levels, where the jamming mode is decided in the first level with the outer Q-learning, and the pulse parameters for the decided jamming mode are selected in the second level with the inner Q-learning. This structure greatly reduces the dimension of the action space of the jammer. With smaller action space and fewer parameters to be learned, it can effectively avoid falling into the local optimum while shortening the convergence time.

Another issue needed to be considered is how to express the feedback from the environment of RL in the jamming decision-making scene. Different from the previous methods where the feedback is statically assigned through the experience matrix, in this paper, we evaluate the effectiveness of jamming as the feedback of the DQL model. On jamming effectiveness evaluation, currently most of the research is based on data collected from the radar side which is impractical due to the non-cooperative nature of the battlefield. However, there are only a few studies on the evaluation from the jamming side. In [21],

the accumulated amplitude extracted from signals is chosen as the characteristic statistic to evaluate the jamming effectiveness. In [22], the authors advise using the change of the radar threat level as the basis for the evaluation. In [23], an evaluation method based on feature space weighting in non-cooperative scenes is proposed, where the weight values of evaluation indicators are solved with offline simulated radar data. Considering the variability of the indicator's contribution to the evaluation result, in this paper, the indicators' weight values are calculated with their entropy and are updated constantly based on the real-time radar data. With the dynamic weights, the jamming effectiveness is evaluated through measuring the distance between the indicator vectors before and after jamming. The evaluation result is served as the feedback of the environment, based on which the DQL model updates dynamically.

The main contributions of this paper are summarized as follows:

- An RL model named DQL is constructed to guide the jamming decision-making against adaptive radars, where the jamming mode and jamming parameters are hierarchically selected and jointly optimized. Because of the reduced dimensionality of action space, the globally optimal solution can easily be found with a shorter convergence time.
- A new jamming effectiveness evaluation method based on indicator vector space is proposed to serve the feedback to the DQL model, which effectively overcomes the dependence on subjective experience when the model updates. Additionally, in view of the variable electromagnetic environment, the indicators' weights are calculated dynamically with the real-time radar data, to make the evaluation result more credible.

The rest of this paper is organized as follows. The system model is introduced and the problem of jamming strategy optimization is formulated in Section 2. The proposed DQL model and jamming effectiveness evaluation method are explained in Section 3. The details of the simulations and the analysis of results are shown in Section 4, followed by the conclusion presented in Section 5.

## 2. System Model and Problem Formulation

### 2.1. System Model

Consider a self-defense electronic jamming scenario, where each target is equipped with a cognitive jammer, shown in Figure 1. Taking a radar and a jammer into account, the jammer devotes to optimizing jamming effectiveness by learning the strategy of the radar, in order to protect the target from detection. The adaptive radar has  $I$  work modes such as search, tracking, and guidance, denoted as  $\{M_{r_1}, M_{r_2}, \dots, M_{r_I}\}$ . The work modes are switched adaptively between two adjacent beam dwell periods according to the signal-to-jamming ratio (SJR). For different modes, the radar changes signal parameters pulse by pulse according to different rules. Similarly, the jammer has  $J$  jamming modes such as frequency-spot jamming, blocking jamming, and swept jamming, denoted as  $\{M_{j_1}, M_{j_2}, \dots, M_{j_J}\}$ . For different jamming modes, the jammer can select parameters for each jamming pulse.

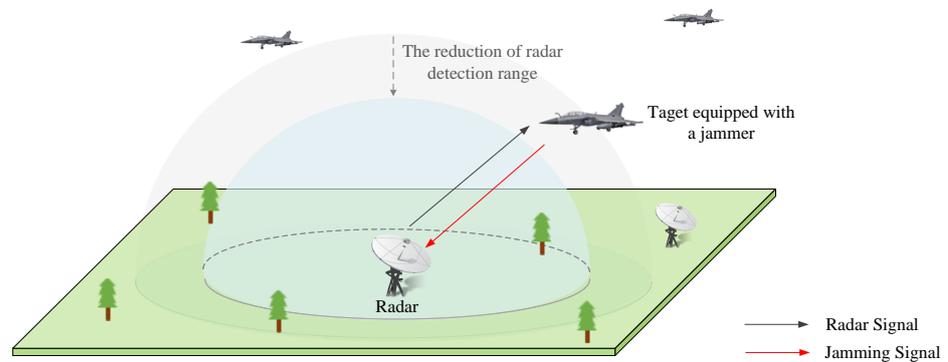
To simplify the analysis, we regard the target as a point target with radar cross-section (RCS)  $\sigma$ . Assume the radar would be jammed in each beam dwell period, which is called a jamming round in this paper. The number of radar pulses in a jamming round depends on the beam dwell time and the pulse repetition interval (PRI). For the  $n$ th radar pulse in a jamming round, the carrier frequency (CF) is  $f_r^{(n)}$ , the bandwidth (BW) is  $B_r^{(n)}$ , the PRI is  $pr_i_r^{(n)}$  which represents the time between the rising edge of  $(n-1)$ th and  $n$ th radar pulse, the pulse width (PW) is  $pw_r^{(n)}$  and the transmission power is  $P_r^{(n)}$ . At each pulse, the jammer attempts to align the jamming signal with the radar signal in both time and frequency domains. For the  $n$ th jammer pulse, the CF is  $f_j^{(n)}$ , the BW is  $B_j^{(n)}$ , the pulse delay time is  $dt_j^{(n)}$  which represents the time from receiving the  $n-1$ th radar pulse to transmitting next jamming pulse, the PW is  $pw_j^{(n)}$  and the transmitting power is  $P_j^{(n)}$ . In addition, the distance

from the radar to the target is  $D$ , and the wavelength of the radar is  $\lambda$ . The antenna gain of the radar and the jammer is  $G_r$  and  $G_j$ . The transmission loss of the radar and jamming signal is  $L_r$  and  $L_j$ . The loss coefficient of polarization matching between the jamming signal and the radar signal is  $\mu$ . Based on the above definitions, the power of  $n$ th echo at the radar receiver can be expressed as

$$P_{rs}^{(n)} = \frac{P_r^{(n)} G_r^2 \sigma \lambda^2}{(4\pi)^3 D^4 L_r} \quad (1)$$

The power of  $n$ th jamming pulse at the radar receiver is

$$P_{rj}^{(n)} = \frac{P_j^{(n)} G_j G_r \lambda^2 \mu}{(4\pi)^2 D^2 L_j} \quad (2)$$



**Figure 1.** The self-defense electronic jamming scenario: through effective jamming, the jammer can shorten the detection range of the radar to protect the target.

Introducing effective jamming coefficient to amend the calculation formula of SJR, the average SJR for  $n$ th radar pulse  $SJR^{(n)}$  is calculated as:

$$SJR^{(n)} = \frac{P_{rs}^{(n)}}{P_{rj}^{(n)}} \cdot \frac{1}{X_f^{(n)}} \cdot \frac{1}{X_t^{(n)}} = \frac{P_r^{(n)} G_r \sigma L_j}{P_j^{(n)} G_j \mu 4\pi D^2 L_r X_f^{(n)} X_t^{(n)}}, \quad (3)$$

where  $X_f^{(n)}$  and  $X_t^{(n)}$  are the effective jamming coefficients in frequency domain and time domain respectively, expressed as:

$$X_f^{(n)} = \frac{\Delta f^{(n)}}{B_j^{(n)}} \cdot \text{sgn}(\Delta f^{(n)}), \quad (4)$$

$$X_t^{(n)} = \frac{\Delta t^{(n)}}{pw_j^{(n)}} \cdot \text{sgn}(\Delta t^{(n)}), \quad (5)$$

where  $\Delta f^{(n)}$  and  $\Delta t^{(n)}$  are the overlapping rates in frequency domain and time domain respectively, defined as:

$$\Delta f^{(n)} = \min(f_j^{(n)} + B_j^{(n)}/2, f_r^{(n)} + B_r^{(n)}/2) - \max(f_j^{(n)} - B_j^{(n)}/2, f_r^{(n)} - B_r^{(n)}/2), \quad (6)$$

$$\Delta t^{(n)} = \min(dt_j^{(n)} + pw_j^{(n)}, pri_r^{(n)} + pw_r^{(n)}) - \max(dt_j^{(n)}, pri_r^{(n)}), \quad (7)$$

where  $\text{sgn}(x) = 1$  if  $x > 0$ , otherwise  $\text{sgn}(x) = -1$ . To reflect the performance of the algorithm more intuitively, we calculate jamming-to-signal ratio (JSR) in the simulations to measure the jamming effect, where  $JSR^{(n)} = 1/SJR^{(n)}$ .

## 2.2. Problem Formulation

The jamming problem can be modeled as a finite Markov decision process (MDP) [24], expressed as a quadruple  $\{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}\}$ .  $\mathcal{S}$  is a finite set of radar states, where state  $s \in \mathcal{S}$  is determined by the radar mode and radar pulse parameters.  $\mathcal{A}$  is a finite set of jammer actions, where action  $a \in \mathcal{A}$  is defined by the jamming mode and jamming pulse parameters.  $\mathcal{P}(s^{(n+1)}|s^{(n)}, a^{(n)})$  is the transition probability describing how the current state  $s^{(n)}$  transfers to next state  $s^{(n+1)}$  when the jammer takes action  $a^{(n)}$ .  $\mathcal{R}$  is the immediate reward after each action is taken.

Reinforcement learning has been proved to be an effective way to solve MDP problems, the key of which is to find the optimal policy  $\pi : \mathcal{S} \rightarrow \mathcal{A}$  to determine which action should be taken at each state. To estimate the effect of a policy, the state-value function for policy  $\pi$  is introduced as:

$$v_{\pi}(s) = E_{\pi} \left[ \sum_{i=0}^{\infty} \gamma^i R^{(n+i+1)} | s^{(n)} = s \right], \quad (8)$$

where  $E_{\pi}[\cdot]$  stands for expected value with the policy  $\pi$  given.  $\gamma \in (0, 1]$  is the discount rate of the reward  $R$ , which means that long-term reward is considered and its influence decreases with time. Then, the optimal policy  $\pi_*$  we aim to find is:

$$\pi_* = \arg \max_{\pi} v_{\pi}(s), \forall s \in \mathcal{S} \quad (9)$$

## 3. Proposed Jamming Scheme Based on DQL Model

To interfere the adaptive radar with mode switching and parameter agility, we propose an RL model named dual Q-learning to optimize jamming strategy, as shown in Figure 2. The jammer's action space is disassembled into two subspaces containing jamming mode and pulse parameters respectively to reduce the dimensionality, based on which the jamming procedure can be divided into two levels. The jamming mode is determined in the first decision-making level, and specific parameters in frequency and time domain are selected in the second level according to the jamming mode. Two interactive Q-learning models are constructed to find the global optimal solution, and a dynamic method for jamming effectiveness evaluation is designed to obtain the feedback of the DQL model. The interaction between the two levels can be described as: the jamming mode determined in the first level has a guiding effect on the selection in the second level, and the pulse parameters selected in the second level directly determine the SJR at the radar receiver and affect the mode switching of the radar, thereby affecting the next input state of the first level.

### 3.1. Jamming Decision-Making through DQL Model

In the jamming procedure based on the DQL model mentioned above, the outer Q-learning and the inner Q-learning model are trained simultaneously to solve the optimal jamming strategy.

#### 3.1.1. Outer Q-Learning

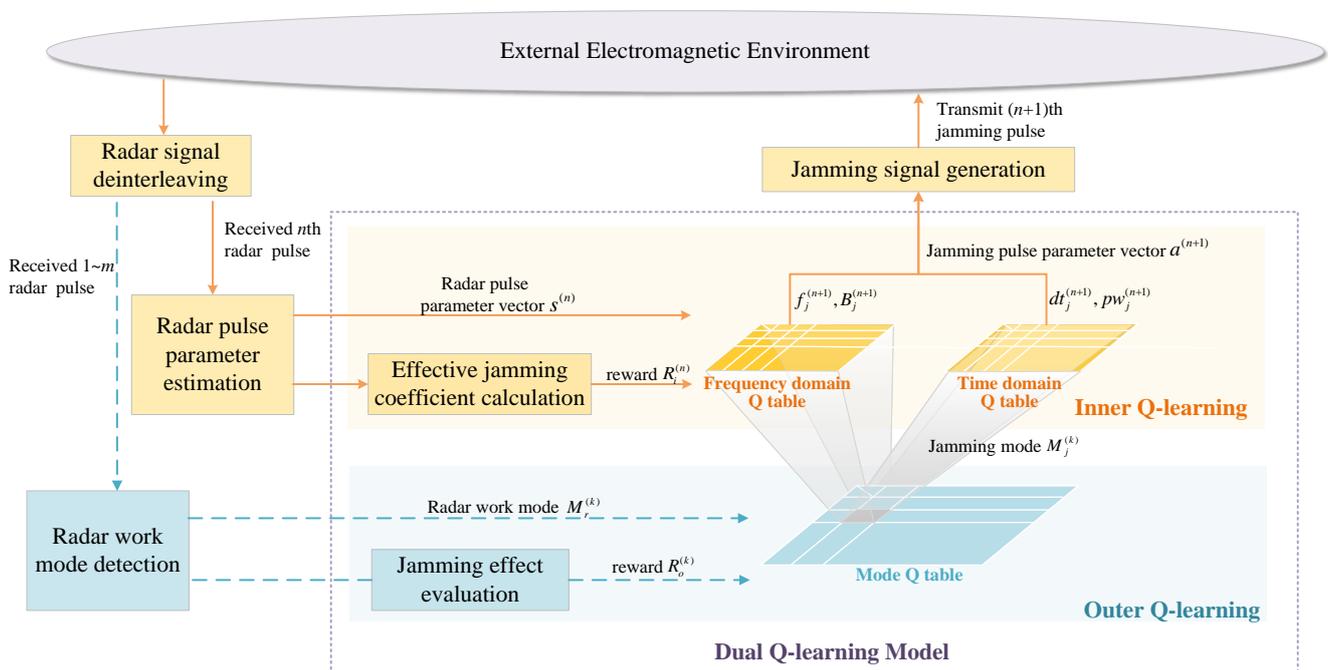
The outer Q-learning is modeled to acquire the jamming mode in the first decision-making level, where the radar work mode and the jamming mode are regarded as the environment state and the action of the agent, respectively. When obtaining radar work mode  $M_r^{(k)}$  at time  $k$ , the jammer chooses jamming mode  $M_j^{(k)}$  as:

$$M_j^{(k)} = \arg \max_{M_j} Q_o \left( M_r^{(k)}, M_j \right) \quad (10)$$

$Q_o(M_r^{(k)}, M_j^{(k)})$  in the outer Q table is updated when the new radar mode  $M_r^{(k+1)}$  is obtained at time  $(k + 1)$  according to the following rule:

$$Q_o(M_r^{(k)}, M_j^{(k)}) \leftarrow Q_o(M_r^{(k)}, M_j^{(k)}) + \alpha \left[ R_o^{(k+1)} + \gamma \max_{M_j} Q_o(M_r^{(k+1)}, M_j) - Q_o(M_r^{(k)}, M_j^{(k)}) \right], \quad (11)$$

where  $\alpha \in (0, 1]$  is the learning rate of the outer Q-learning, which signifies the updating stride of Q value.  $R_o^{(k+1)}$  is the reward calculated at time  $(k + 1)$  for the outer Q-learning, which depends on the evaluation result of jamming effectiveness according to radar mode switching. The evaluation method will be introduced in the next section.



**Figure 2.** Jamming decision-making based on DQL model: During a jamming round, the radar work mode is discerned with the earliest received  $m$  radar pulses. Once a new radar mode  $M_r^{(k)}$  is recognized, it is sent to the outer Q-learning module. Then the jamming mode  $M_j^{(k)}$  is chosen and map to time domain Q table and frequency domain Q table in the inner Q-learning module. According to  $M_r^{(k)}$  and  $M_r^{(k-1)}$ , the jamming effectiveness of last jamming round is evaluated, with which the outer Q table can be updated according to Equation (11). When receiving the  $n$ th ( $n > m$ ) radar pulse, the radar pulse parameter vector  $s^{(n)}$  is obtained through parameter estimation. Then the jamming parameters are selected to constitute the parameter vector  $a^{(n+1)}$ , and the jamming signal will be generated. According to  $s^{(n)}$ , the two effective jamming coefficients are evaluated, with which the inner Q table can be updated according to Equation (13).

### 3.1.2. Inner Q-Learning

The inner Q-learning is modeled to solve the optimal jamming parameters in the second decision-making level. The jamming mode selected in the first decision-making level maps to inner Q tables in frequency and time domains. When obtaining the  $n$ th radar pulse parameter vector  $s^{(n)} = [f_r^{(n)}, B_r^{(n)}, pri_r^{(n)}, pw_r^{(n)}, P_r^{(n)}]$ , the jammer takes  $[f_r^{(n)}, B_r^{(n)}]$

and  $[pri_r^{(n)}, pw_r^{(n)}]$  as the input states of frequency and time domain Q tables respectively, and the jamming parameters will be chosen according to:

$$p_j^{(n+1)} = \arg \max_{p_j} Q_i(p_r^{(n)}, p_j), \quad (12)$$

where  $p_r^{(n)} = [f_r^{(n)}, B_r^{(n)}]$ ,  $p_j^{(n)} = [f_j^{(n)}, B_j^{(n)}]$  in the frequency domain Q table,  $p_r^{(n)} = [pri_r^{(n)}, pw_r^{(n)}]$ ,  $p_j^{(n)} = [dt_j^{(n)}, pw_j^{(n)}]$  in the time domain Q table. The power of jamming signal  $P_j^{(n+1)}$  is calculated as  $P_j^{(n+1)} = \eta \cdot P_r^{(n)}$ , where  $\eta$  varies with different jamming modes. Then the jamming parameter vector  $\mathbf{a}^{(n+1)} = [f_j^{(n+1)}, B_j^{(n+1)}, dt_j^{(n+1)}, pw_j^{(n+1)}, P_j^{(n+1)}]$  can be constituted. After the receiving of  $n$ th radar pulse, the transmission of next jamming pulse will start after a delay of  $dt_j^{(n+1)}$ , aiming at the  $(n + 1)$ th radar pulse. In summary, the jammer predicts the radar parameters one step in advance, in order to interfere the next radar pulse in time. When the  $(n + 1)$ th radar pulse parameter vector  $s^{(n+1)}$  is obtained,  $Q_i(p_r^{(n)}, p_j^{(n+1)})$  in the inner Q table is updated according to the following rule:

$$Q_i(p_r^{(n)}, p_j^{(n+1)}) \leftarrow Q_i(p_r^{(n)}, p_j^{(n+1)}) + \alpha \left[ R_i^{(n+1)} + \gamma \max_{p_j} Q_i(p_r^{(n+1)}, p_j) - Q_i(p_r^{(n)}, p_j^{(n+1)}) \right], \quad (13)$$

where  $R_i^{(n+1)}$  is the reward of the inner Q-learning calculated with  $s^{(n+1)}$  and  $a^{(n+1)}$ .  $R_i^{(n+1)} = X_f^{(n+1)}$  for the frequency domain Q table,  $R_i^{(n+1)} = X_t^{(n+1)}$  for the time domain Q table.

Both in the outer and inner Q-learning,  $\varepsilon$ -greedy policy is used to choose the jamming mode or parameters, which is an effective way to balance the exploration and exploitation. In this paper, we define  $\varepsilon = e^{-\delta \cdot k}$  which decreases as the count of iterations increases, where  $k$  represents the number of the iteration and the coefficient  $\delta$  determines the decay rate.  $\delta$  is set differently for two decision-making levels:  $\delta = \delta_o$  for the outer Q-learning and  $\delta = \delta_i$  for the inner Q-learning. Taking the outer Q-learning as an example, the jammer randomly chooses a jamming mode with probability  $\varepsilon$ , and chooses the optimal jamming mode  $M_j^{(k)}$  according to Equation (10) with probability  $1 - \varepsilon$ . According to the explanation above, a jamming algorithm based on the DQL model is proposed, and more details are shown in Algorithm 1.

The time series of learning and jamming decision-making with the DQL model is shown in Figure 3. During a jamming round, the outer Q-learning is performed only once at the beginning to obtain the jamming mode. Under the constraints of the jamming mode, the inner Q-learning is performed in each subsequent PRI to determine the jamming parameters.

### 3.2. Jamming Effectiveness Evaluation through Dynamic Measuring of Vector Distance

Jamming effectiveness evaluation is an important part of the jamming decision-making method based on the DQL model proposed in the previous section. The evaluation result provides the reward for the outer Q-learning. In the radar confrontation, the most intuitive impact of effective jamming on the radar is the reduction of detection probability, which is hard to be known from the jamming side. Therefore, the jamming effectiveness can only be estimated according to the parameters of the radar signal received by the jammer.

**Algorithm 1:** Jamming algorithm based on DQL model

( $K$  and  $N_k$  denote the amount of jamming rounds in simulation and the total number of pulses in the  $k$ th jamming round respectively.  $m$  represents the number of pulses required for radar mode discerning.)

**for**  $k = 1, 2, \dots, K$  **do**

    Based on the radar mode discerned, choose a jamming mode  $M_j^{(k)}$  from outer Q table according to Equation (10) with  $\epsilon$  – greedy policy;

    Map into two inner Q tables;

**if**  $k > 1$  **then**

        Calculate  $l$  evaluation indicators of last radar mode to obtain evaluation indicator vector  $\mathbf{x}^{(k)}$ ;

        Calculate  $R_o^{(k)}$  through **Algorithm 2** with  $\mathbf{x}^{(k)}$  as input ;

        Update  $Q_o(M_r^{(k-1)}, M_j^{(k-1)})$  through Equation (11);

**end**

**for**  $n=m+2, m+3, \dots, N_k$  **do**

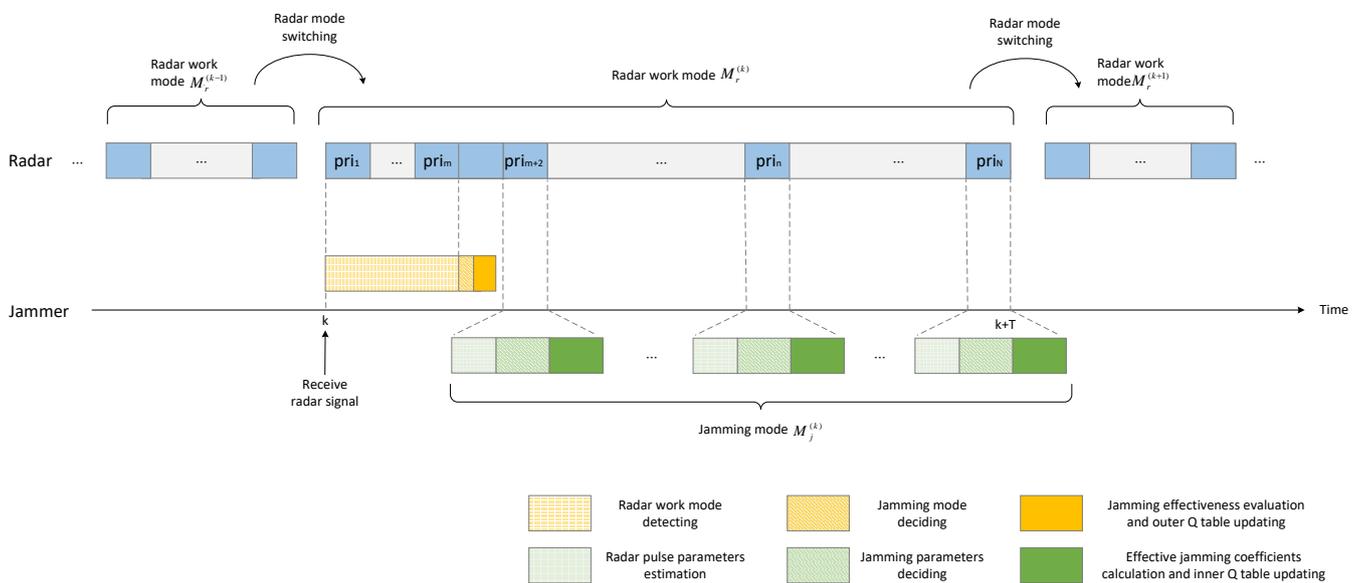
        Based on the radar pulse parameter vector, choose  $f_j^{(n+1)}, B_j^{(n+1)}, dt_j^{(n+1)}, pw_j^{(n+1)}$  according to Equation (12) with  $\epsilon$  – greedy policy and calculate  $p_j^{(n+1)}$  to constitute jamming parameter vector  $\mathbf{a}^{(n+1)}$  ;

        Calculate effective jamming coefficients  $X_f^{(n)}$  and  $X_t^{(n)}$  according to Equation (4) and (5);

        Update  $Q_i(p_r^{(n-1)}, p_j^{(n)})$  according to (13).

**end**

**end**



**Figure 3.** Process of learning and jamming decision making with DQL model.

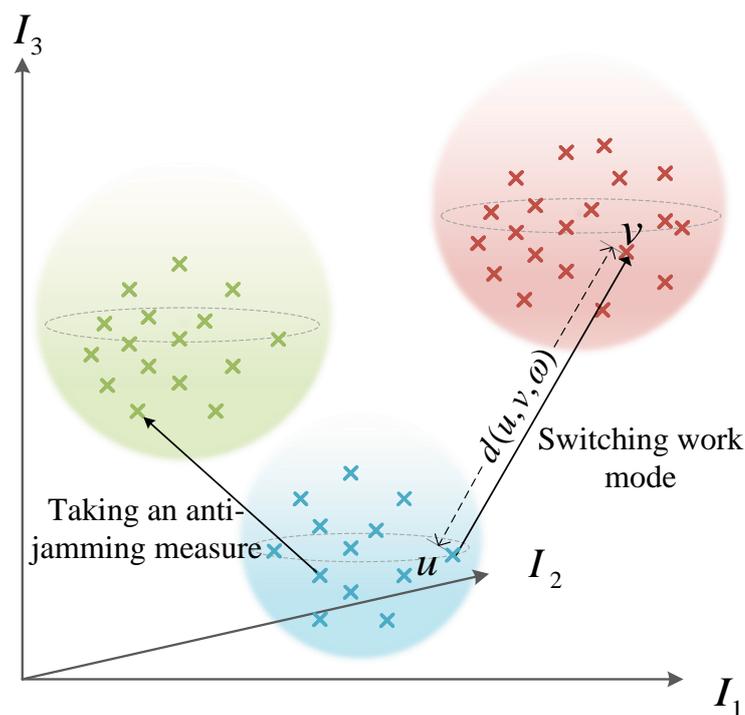
To evaluate the jamming effectiveness, we firstly construct an evaluation indicator set  $\mathcal{I} = \{I_1, I_2, \dots, I_l\}$ , where each indicator is a measurable parameter of the radar signal. When the radar is jammed, there are two possible ways for it to change its parameters. One is to switch its work mode. For example, the low SNR causes the radar to lose track of the target, so it shifts from the tracking mode to the search mode. The other possible way is that the radar takes anti-jamming measures in order to improve its performance. For instance,

after receiving suppressive jamming, the radar increases the bandwidth to improve its range resolution. Usually, the indicators that can characterize the work mode switching of the radar include beam dwell time, PRI, etc, and the indicators that can characterize the anti-jamming measures taken by the radar include BW, PW, transmitting power, range and speed of frequency agility, etc. The indicators of these two aspects are considered to construct the evaluation indicator set  $\mathcal{I}$ .

Based on  $\mathcal{I}$ , a method of vector distance measuring is proposed to calculate jamming effectiveness, where the weight values of evaluation indicators are updated dynamically during the confrontation process. More details are explained as follows.

### 3.2.1. The Jamming Effectiveness Evaluation Method Based on Vector Distance Measuring

We construct a  $l$ -dimensional vector space  $\mathcal{V}^l$  based on  $\mathcal{I}$ , where each dimension represents an evaluation indicator. To explain clearly, Figure 4 shows a 3-dimensional vector space of three indicators. As shown, for the same radar state, the evaluation indicator vectors are often clustered together. When the radar switches its working mode or takes anti-jamming measures, the evaluation indicator vector will shift in space. The greater the offset along the increasing direction of the coordinate axes is, the more effective the jamming is. Therefore, the jamming efficiency can be evaluated by measuring the shift of the evaluation indicator vector before and after the jamming.



**Figure 4.** Vector space  $\mathcal{V}^3$  composed of three evaluation indicators.  $u, v$  are two evaluation indicator vectors, and  $d(u, v, \omega)$  is the Euclidean distance between  $u$  and  $v$  with weight vector  $\omega$ .

According to the above analysis, a jamming effect evaluation method based on vector distance measuring is proposed. As known, Euclidean distance can be used to calculate the absolute distance between vectors. We assign a weight for each evaluation indicator on the basis of Euclidean distance, where each weight reflects the contribution of the corresponding indicator to the evaluation result. The calculation formula of Euclidean distance with the indicator weight is expressed as:

$$d(u, v, \omega) = \text{sgn}\left(\sum_{i=1, \dots, l} \omega_i (v_i - u_i)\right) \left(\sum_{i=1, \dots, l} \omega_i (v_i - u_i)^2\right)^{\frac{1}{2}}, \quad (14)$$

where  $\mathbf{u}, \mathbf{v}$  are two indicator vectors in  $\mathcal{V}^l$ .  $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_l)$  is the weight vector for  $l$  indicators,  $\omega_i \in (0, 1)$ .  $u_i$  is the value of  $i$ th indicator in vector  $\mathbf{u}$ .  $\text{sgn}(\sum_{i=1, \dots, l} \omega_i (v_i - u_i))$  identifies the symbol of the distance  $d(\mathbf{u}, \mathbf{v}, \boldsymbol{\omega})$ , which indicates the direction in which the evaluation indicator vector shifts.

The feedback  $R_o^{(k)}$  of the DQL model at time  $k$  based on jamming effectiveness evaluation is calculated as:

$$R_o^{(k)} = \begin{cases} d(\mathbf{x}^{(k-1)}, \mathbf{x}^{(k)}, \boldsymbol{\omega}^{(k)}) \cdot 10, & \text{if } d(\mathbf{x}^{(k-1)}, \mathbf{x}^{(k)}, \boldsymbol{\omega}^{(k)}) \geq 0 \\ -10, & \text{if } d(\mathbf{x}^{(k-1)}, \mathbf{x}^{(k)}, \boldsymbol{\omega}^{(k)}) < 0 \end{cases}, \quad (15)$$

where  $\mathbf{x}^{(k)}$  is the normalized evaluation indicator vector obtained at time  $k$ .  $\boldsymbol{\omega}^{(k)}$  is the weight vector calculated at time  $k$ , and the calculation method for it is described below.

### 3.2.2. The Method of Dynamically Weighting for Evaluation Indicators

As the contribution of different indicators to the evaluation result will vary with the change of the radar status,  $\boldsymbol{\omega}$  is objectively modified through the method of dynamic entropy weight calculation. We horizontally compare the value of each indicator measured in different jamming rounds, and calculate their entropy values to obtain the weights. For each indicator, as a result of the difference between each measurement, its weight is not static, but changes with the received radar signal parameters. Thus, an online evaluation model is established.

For calculating the dynamic entropy weights of evaluation indicators, we define a matrix  $\mathbf{A} \in \mathbb{R}^{l \times m}$ , where  $m$  evaluation indicator vectors can be stored. Once a new radar state is detected, the  $l$  evaluation indicators of it are calculated and assigned to a column in  $\mathbf{A}$ . If all the columns in  $\mathbf{A}$  have been assigned values, the earliest assigned column will be overwritten by the new evaluation indicator vector.  $a_{ij}$  in  $\mathbf{A}$  denotes the value of the  $i$ th indicator in the  $j$ th vector. In order to nondimensionalize the calculation and eliminate the impact of different order of magnitudes, the original matrix needs to be normalized to a matrix  $\mathbf{B} = (b_{ij})_{l \times m}$ . The normalization formula is expressed as:

$$b_{ij} = \begin{cases} \frac{a_{ij} - \min(a_{ij, \dots, a_{in}})}{\max(a_{ij, \dots, a_{in}}) - \min(a_{ij, \dots, a_{in}})}, & \text{if } I_i \propto d \\ \frac{\max(a_{ij, \dots, a_{in}}) - a_{ij}}{\max(a_{ij, \dots, a_{in}}) - \min(a_{ij, \dots, a_{in}})}, & \text{otherwise.} \end{cases}, \quad i = 1, 2, \dots, l, j = 1, 2, \dots, m, \quad (16)$$

where  $I_i \propto d$  indicates that the vector distance  $d$  is positively correlated to the  $i$ th indicator in  $\mathcal{I}$ , which means the larger the value of  $(v_i - u_i)$  is, the better the jamming effectiveness is. Then the proportion of the  $j$ th vector for indicator  $I_i$  is calculated as  $p_{ij} = b_{ij} / \sum_{j=1}^m b_{ij}$ , and the entropy of indicator  $I_i$  can be calculated as:

$$e_i = -\frac{1}{\ln(l)} \sum_{j=1}^m p_{ij} \ln(p_{ij}), \quad i = 1, 2, \dots, l \quad (17)$$

Finally, the weight for each indicator is expressed as:

$$\omega_i = \frac{1 - e_i}{\sum_{i=1}^m (1 - e_i)}, \quad i = 1, 2, \dots, l \quad (18)$$

As shown in Figure 2, when a new radar mode is detected at time  $k$ , the jamming effectiveness of the last jamming round is evaluated, with which  $R_o^k$  is calculated and served as the feedback of the DQL model. The jamming effectiveness evaluation algorithm based on vector distance measuring with dynamic weight is shown in Algorithm 2.

**Algorithm 2:** Jamming effectiveness evaluation algorithm

---

**Input:** Evaluation indicator vector  $\mathbf{x}^{(k)}$   
**Output:** Jamming effect evaluation result  $R_o^{(k)}$   
**if** *There is no column with all zeros in the A then*  
    | Reassign the earliest assigned column in  $A$  with  $\mathbf{x}^{(k)}$ ;  
**else**  
    | Assign  $\mathbf{x}^{(k)}$  to a column with all zeros in  $A$ ;  
**end**  
Normalize  $A$  to  $B$  according to Equation (16);  
**for**  $i = 1, 2, \dots, l$  **do**  
    | Calculate the proportion  $p_{ij}$  of the  $j$ th vector for  $I_i$  as  $p_{ij} = b_{ij} / \sum_{j=1}^m b_{ij}$ ;  
    | Calculate the entropy  $e_i$  of  $I_i$  through Equation (17);  
    | Calculate the weight  $\omega_i$  through Equation (18);  
**end**  
Calculate the Euclidean distance  $d(\mathbf{x}^{(k-1)}, \mathbf{x}^{(k)}, \boldsymbol{\omega}^{(k)})$  between normalized  $\mathbf{x}^{(k)}$  and the last indicator vector  $\mathbf{x}^{(k-1)}$  with the weight vector  $\boldsymbol{\omega}^{(k)}$  according to Equation (14);  
Calculate the feedback  $R_o^{(k)}$  of the DQL model through Equation (15).

---

#### 4. Numerical Results

To verify the proposed algorithms, a radar parameter template is firstly created, shown in Table 1. Referring to [25,26], four kinds of radar work modes are considered in this paper, including search, acquisition, tracking, and guidance. For the target, the work mode of guidance has the highest threat level, followed by tracking, acquisition, and search. The beam dwell times when the radar is at these four work modes are 80 ms, 100 ms, 120 ms, and 140 ms respectively. For each work mode, two sub-modes are specified with different parameter agility patterns.

The mode switching rule of the radar can be described as: when  $\text{SJR} > -4$  db, the radar will raise its threat level; when  $-7$  db  $< \text{SJR} < -4$  db, the radar will take anti-jamming measures while maintaining the current work mode, including  $M_{r_1} \rightarrow M_{r_2}$ ,  $M_{r_3} \rightarrow M_{r_4}$ ,  $M_{r_5} \rightarrow M_{r_6}$  and  $M_{r_7} \rightarrow M_{r_8}$ ; when  $\text{SJR} < -7$  db, the radar will reduce its threat level.

For the jammer, the jamming mode can be switched between  $M_{j_1}$  and  $M_{j_2}$ , which denote frequency-spot jamming and blocking jamming respectively. The optional jamming parameters for each jamming mode are illustrated in Table 2.

Other parameters in our simulation are given as:  $G_r = 30$  dB,  $G_j = 5$  dB,  $L_r = 10$  dB,  $L_j = 5$  dB,  $R = 10$  km,  $\sigma = 1$  m<sup>2</sup>. Considering that radar antennas are generally linearly polarized, while jammer antennas are circularly polarized or obliquely polarized,  $\mu$  is given as 0.5. The parameters in the proposed algorithms are set as:  $\alpha = 0.01$ ,  $\gamma = 0.8$ ,  $\delta_o = 0.08$ , and  $\delta_i = 0.3$ . The evaluation indicator set  $\mathcal{I}$  is constructed of indicators  $I_1 \sim I_7$ , which are shown in Table 3. For each radar mode, the average PRI of all pulses in a period is calculated as indicator PRI, the difference between the maximum and minimum frequency is taken as the range of frequency agility, and the reciprocal of the number of continuous pulses with the same frequency is regarded as the speed of frequency agility.

Based on the above description, 200 jamming rounds are simulated. Figure 5 intuitively shows the time-frequency information under four different radar work modes at the initial and convergent stage of learning. Figure 5a1, b1, c1, d1 show the radar and jamming signals at the initial stage, and four other figures show the convergent stage under the same radar work mode. Compared with the random selection of jamming parameters at the initial stage, the jamming pulses can accurately cover radar pulses in time and frequency domains to achieve effective jamming at the convergent stage. Besides, it can be found that if the radar is at modes where the CF changes regularly, the jammer chooses frequency-spot

jamming mode  $M_{j_1}$ , otherwise, it chooses to block jamming mode  $M_{j_2}$ , which accords with the common perception.

**Table 1.** Radar parameter template. When the radar is at a certain mode, it selects the pulse parameters according to the rules in the template. Reside and switch, slippery, staggered, and jittered are four of the common radar parameter agility patterns. As shown, reside and switch A:k B:m C:n means that the parameter value stays at A for k pulses, stays at B for m pulses, and stays at C for n pulses; slippery A:B:C means that the parameter value changes from A to C in steps of B; staggered such as [A B C] means that the parameter value is cycled in the order of the list; jittered such as (A,B) means the parameter value is randomly selected from the range of A to B.

Work Mode	Sub-Mode	$f_r$ /MHz	$B_r$ /MHz	$pr_i$ /us	$pw_r$ /us	$P_r$ /kW
search	$M_{r_1}$	reside and switch: 8500:5 9500:5 9000:5	100	staggered: [1100 1320 1470]	80	120
	$M_{r_2}$	reside and switch: 8600:3 9600:3 9100:3	100	staggered: [1100 1320 1470]	120	120
acquisition	$M_{r_3}$	slippery: 8800:600:10000	150	staggered: [1070 1430 857]	120	170
	$M_{r_4}$	slippery: 9800:600:12200	150	staggered: [1070 1430 857]	120	170
tracking	$M_{r_5}$	jittered: (8500,11500)	800	reside and switch: 830:2 890:4 960:3	120	170
	$M_{r_6}$	jittered: (7500,12500)	1000	reside and switch: 830:2 890:4 960:3	120	170
guidance	$M_{r_7}$	jittered: (9500,12500)	800	slippery: 740:40:900	120	200
	$M_{r_8}$	jittered: (8500,13500)	1000	slippery: 740:40:900	120	200

**Table 2.** Jamming parameter template. When a jamming mode is determined, the corresponding pulse parameters are selected according to the rules in the template. {A:B:C} denotes a set of optional values, consisting of an arithmetic sequence from A to C with B as the difference.

Mode	$f_j$ /MHz	$B_j(\times B_r)$	$dt_j$ /us	$pw_j(\times pw_r)$	$P_j(\times P_r)$
$M_{j_1}$	{8000:500:12000}	3	{800:100:1500}	2	0.003
$M_{j_2}$	{8500:1000:11500}	6	{800:100:1500}	2	0.006

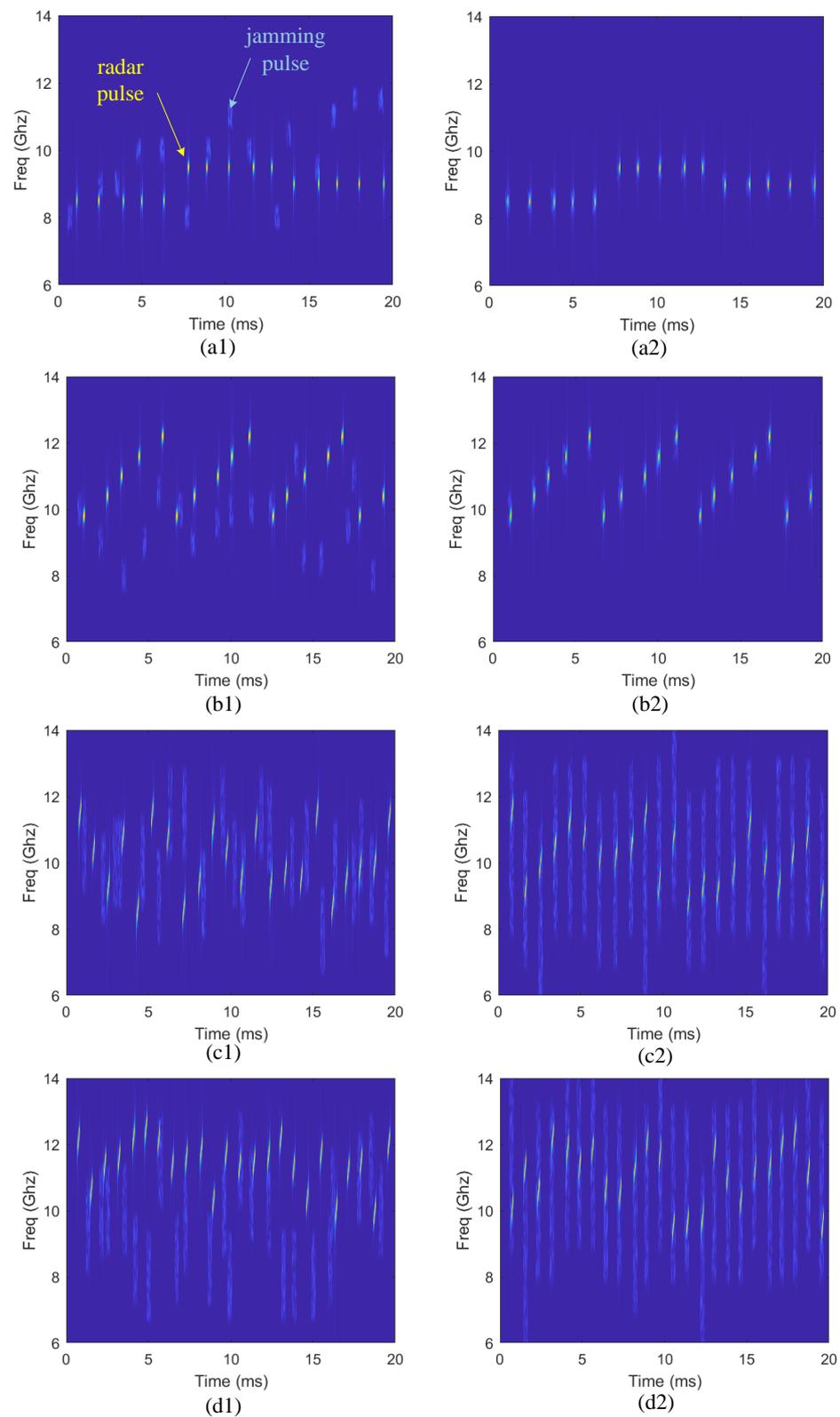
We compared the performance of our algorithm with the improved chaos genetic algorithm [6], the standard Q-learning [15] and the random parameter selection method. The average JSR of each jamming round is calculated to reflect the jamming effect intuitively. The learning rate and discount rate are set with the same value both in the standard Q-learning and our algorithm. The following results are obtained through 500 independent simulations and their average values are taken to make the figures.

**Table 3.** Jamming effectiveness evaluation indicator set. The correlation to evaluation result is positive means that the greater the increase of this indicator, the better the jamming effectiveness. While the negative attribute means that the greater the decrease of this indicator, the better the jamming effectiveness.

Evaluation Indicator	Explanation	Correlation to Evaluation Result
$I_1$	PRI	positive
$I_2$	power	positive
$I_3$	beam dwell time	negative
$I_4$	BW	positive
$I_5$	PW	positive
$I_6$	range of frequency agility	positive
$I_7$	speed of frequency agility	positive

Figures 6 and 7 respectively show the average JSR and the radar threat level obtained through three methods during 200 jamming rounds. As shown, the improved chaos genetic algorithm [6], the standard Q-learning [15] and the proposed jamming algorithm based on the DQL model can all minimize the threat level of radar, and both the latter two methods can converge and stabilize the JSR within 200 jamming rounds. However, compared with the standard Q-learning, the proposed algorithm can reach the optimal average JSR 7.98 dB, which is 4.05% increased. Further, the convergence time of the proposed jamming algorithm declines by 34.94%, and the number of jamming rounds when the radar is at guidance modes including  $M_r^7$  and  $M_r^8$  reduces by 64.94%. As the high radar threat level is dangerous for the target, the proposed jamming algorithm based on the DQL model can improve the survivability of the target.

In order to further explore the convergence performance of the proposed jamming algorithm, we use the jamming round in which JSR is stable to indicate the convergence time. For jamming round  $i$ , we calculate the variance of JSR from jamming round  $(i - 19)$  to jamming round  $i$  ( $i \geq 20$ ). If the variance is less than 0.01, it is considered that JSR reaches a stable state in jamming round  $i$ . Figure 8 compares the convergence time of the jamming algorithm based on the standard Q-learning [15] and the proposed DQL model. It is shown that the convergence time of the proposed jamming algorithm is generally lower than that of the standard Q-learning, and as the size of jamming action space increases, the gap between the two grows. Thus, our jamming algorithm based on the DQL model has better scalability, and is more adaptable when larger jamming action space is needed in face of adaptive even unknown radars.



**Figure 5.** Time-frequency information at the initial and convergent stage: for signals in each image, the brighter strips are radar pulses, the darker strips are jamming pulses. (a1,b1,c1,d1) show time-frequency information of radar and jamming pulses at initial stage when the radar is at mode  $M_{r_1}$ ,  $M_{r_4}$ ,  $M_{r_5}$ , and  $M_{r_7}$ , respectively. (a2,b2,c2,d2) show the corresponding time-frequency information at convergent stage for the same four radar modes.

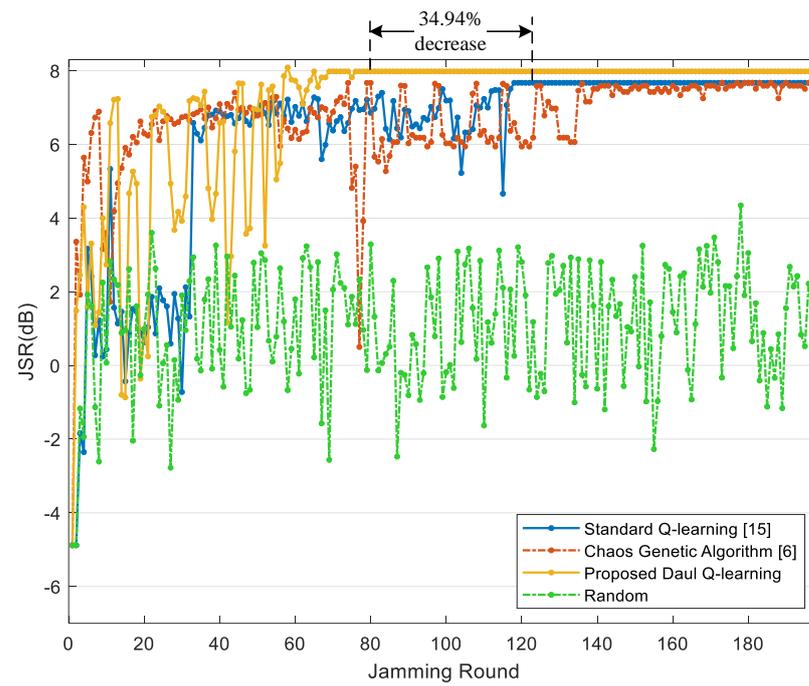


Figure 6. JSR comparison among different methods.

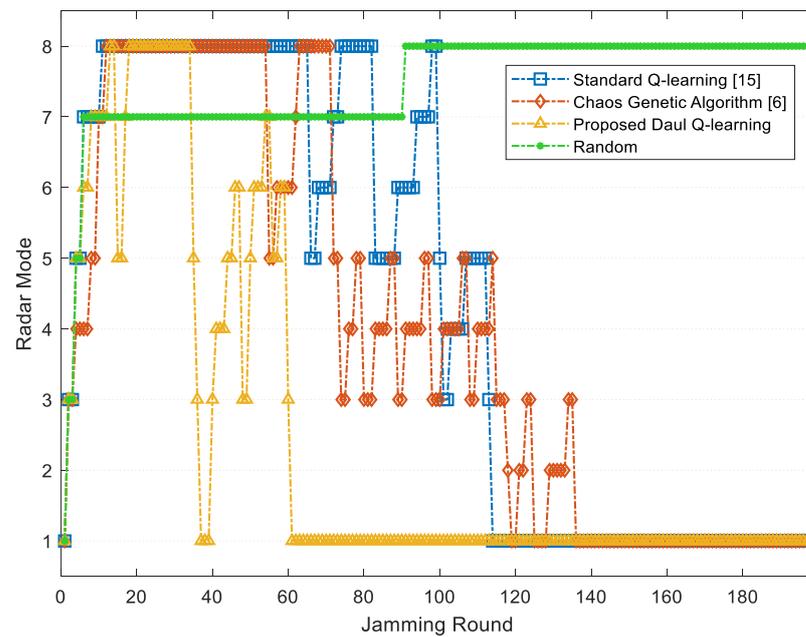
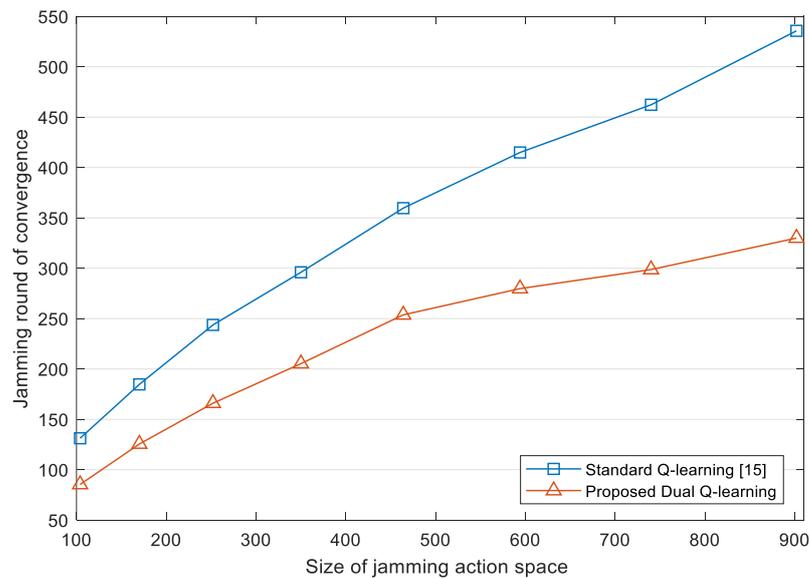


Figure 7. Radar mode switching comparison among different methods: radar modes 1 to 8 represent  $M_{r_1}$  to  $M_{r_8}$  respectively.



**Figure 8.** Convergence time comparison among different methods. The size of jamming action space is defined as the product of the total number of optional parameters in time and frequency domains.

## 5. Conclusions

In this paper, a two-level framework is developed for jamming decision-making against the adaptive radar, and a dual Q-learning model is proposed to optimize the jamming strategy. The jamming mode and pulse parameters are determined hierarchically, greatly reducing the dimensionality of the search space and improving the learning efficiency of the model. In addition, we proposed a new method to calculate the jamming effectiveness by measuring the distance of indicator vectors, where the indicators are dynamically weighted to adapt to the changing environment. The jamming effectiveness evaluation result is served as the feedback value to update the DQL model.

Simulation results show that with the proposed jamming method, the radar joint strategy of mode switching and pulse parameters can be learned within limited interactions, and the optimal jamming effectiveness is reached while the radar's threat level is minimized. Furthermore, compared with the standard Q-learning, our method improves the average JSR by 4.05% and reduces the convergence time by 34.94%.

It should be emphasized that due to the complex electromagnetic environment, the estimation of radar work mode and pulse parameters is often inaccurate. When there are errors in the input state, the performance of the proposed DQL model will deteriorate. Therefore, in the near future, how to enhance the robustness of the model to deal with the uncertainty of the input is the focus of our investigation.

**Author Contributions:** Investigation, H.L.; methodology, H.L.; project administration, Y.H.; resources, H.Z.; supervision, Y.H.; validation, H.L.; writing—original draft preparation, H.L.; writing—review and editing, H.Z.; data curation, Y.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China under Grant 61971064 and Grant 61901049 and the Beijing Natural Science Foundation under Grant 4202048 and Grant L212028.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Xing, Q.; Zhu, W.; Chi, Z.; Zheng, G. Jamming decision under condition of incomplete jamming rule library. *J. Eng.* **2019**, *2019*, 7449–7454. [[CrossRef](#)]
- Haykin, S. Cognitive Radar: A Way of the Future. *IEEE Signal Processing Mag.* **2006**, *23*, 30–40. [[CrossRef](#)]

3. Gao, L.; Liu, L.; Cao, Y.; Wang, S.; You, S. Performance analysis of one-step prediction-based cognitive jamming in jammer-radar countermeasure model. *J. Eng.* **2019**, *2019*, 7958–7961. [[CrossRef](#)]
4. Zhang, B.; Zhu, W. Research on Decision-making System of Cognitive Jamming against Multifunctional Radar. In Proceedings of the 2019 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Dalian, China, 20–22 September 2019; pp. 1–6.
5. Zheng, T.; Gao, X. Research on the self-defence electronic jamming decision-making based on the discrete dynamic Bayesian network. *J. Syst. Eng. Electron.* **2008**, *19*, 702–708. [[CrossRef](#)]
6. Pan, W.; Jin, X.; Xie, H.; Xia, Y. Radar Jamming Strategy Allocation Algorithm based on Improved Chaos Genetic Algorithm. In Proceedings of the 2020 Chinese Control And Decision Conference (CCDC), Hefei, China, 22–24 August 2020; pp. 4478–4483.
7. Wang, L.; Zeng, Y.; Li, Y.; Wang, M. An Optimal Jamming Strategy Aiming at Cognitive MIMO Radar. In Proceedings of the 2016 CIE International Conference on Radar (RADAR), Guangzhou, China, 10–13 October 2016; pp. 1–5.
8. Slimeni, F.; Scheers, B.; Chtourou, Z.; Nir, V. L. Jamming mitigation in cognitive radio networks using a modified Q-learning algorithm. In Proceedings of the 2015 International Conference on Military Communications and Information Systems (ICMCIS), Cracow, Poland, 18–19 May 2015; pp. 1–7.
9. Machuzak, S.; Jayaweera, S. K. Reinforcement Learning Based Anti-jamming with Wideband Autonomous Cognitive Radios. In Proceedings of the 2016 IEEE/CIC International Conference on Communications in China (ICCC), Chengdu, China, 27–29 July 2016; pp. 1–5.
10. Peng, J.; Zhang, Z.; Wu, Q.; Zhang, B. Anti-Jamming Communications in UAV Swarms A Reinforcement Learning Approach. *IEEE Access.* **2019**, *7*, 180532–180543. [[CrossRef](#)]
11. Lu, X.; Xiao, L.; Dai, C.; Dai, H. UAV-aided cellular communications with deep reinforcement learning against jamming. *IEEE Wirel. Commun.* **2020**, *27*, 48–53. [[CrossRef](#)]
12. Yao, F.; Jia, L. A collaborative multi-agent reinforcement learning anti-jamming algorithm in wireless networks. *Wirel. Commun. Lett.* **2019**, *8*, 1024–1027. [[CrossRef](#)]
13. Sutton, R. S.; Barto, A. G. Reinforcement Learning: An Introduction, 1st ed.; MIT Press: Cambridge, MA, USA, 1998; pp. 216–224.
14. Xing, Q.; Zhu, W.; Jia, X. Research on method of intelligent radar confrontation based on reinforcement learning. In Proceedings of the 2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA), Beijing, China, 8–11 September 2017; pp. 471–475.
15. Wang, L.; Peng, J.; Xie, Z.; Zhang, Y. Optimal jamming frequency selection for cognitive jammer based on reinforcement learning. In Proceedings of the 2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP), Weihai, China, 28–30 September 2019; pp. 39–43.
16. Li, K.; Jiu, B.; Liu, H.; Liang, S. Reinforcement learning based anti-jamming frequency hopping strategies design for cognitive radar. In Proceedings of the 2018 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Qingdao, China, 14–16 September 2018; pp. 1–5.
17. Lei, M.; Zhang, J. Study on anti-jamming frequency selection in radar netting. In Proceedings of the 2016 2nd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 14–17 October 2016; pp. 1781–1784.
18. Ak, S.; Brüggewirth, S. Avoiding Jammers: A Reinforcement Learning Approach. In Proceedings of the 2020 IEEE International Radar Conference (RADAR), Florence, Italy, 21–25 September 2020; pp. 321–326.
19. Li, K.; Jiu, B.; Liu, H.; Pu, W. Robust antijamming strategy design for frequency-agile radar against main lobe jamming. *Remote Sens.* **2021**, *13*, 3043. [[CrossRef](#)]
20. Quan, Y.; Wu, Y.; Li, Y.; Sun, G.; Xing, M. Range-Doppler reconstruction for frequency agile and PRF-jittering radar. *IET Radar Sonar Navig.* **2018**, *12*, 348–352. [[CrossRef](#)]
21. Ou, J.; Zhao, F.; Ai, X.; Liu, J.; Xiao, S. Quantitative evaluation for self-screening jamming effectiveness based on the changing characteristics of intercepted radar signals. In Proceedings of the 2016 CIE International Conference on Radar (RADAR), Guangzhou, China, 10–13 October 2016; pp. 1–5.
22. Li, C.; Zhou, J. Jamming effectiveness evaluation from the jamming side. *Electron. Inf. Warf. Technol.* **2008**, *23*, 46–49.
23. Peng, X.; Yu, J.; Ren, W.; Weng, X. Radar jamming effectiveness evaluation method based on feature space weighting. In Proceedings of the IET International Radar Conference (IET IRC 2020), Chongqing, China, 4–6 November 2020; pp. 629–633.
24. Rabiner, L.R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **1989**, *77*, 257–286. [[CrossRef](#)]
25. Osner, N.R.; du Plessis, W.P. Threat evaluation and jamming allocation. *IET Radar Sonar Navig.* **2017**, *11*, 459–465. [[CrossRef](#)]
26. Han, L.; Ning, Q.; Chen, B.; Lei, Y.; Zhou, X. Ground threat evaluation and jamming allocation model with Markov chain for aircraft. *IET Radar Sonar Navig.* **2020**, *14*, 1039–1045. [[CrossRef](#)]