IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 21 May 2022, accepted 23 June 2022, date of publication 29 June 2022, date of current version 8 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3187030

RESEARCH ARTICLE

An Improved Anti-Jamming Method Based on Deep Reinforcement Learning and Feature Engineering

XIN CHANG[1], YANBIN LI[1], YAN ZHAO[1], YUFENG DU[1,2], AND DONGHUI LIU[3]

[1]The 54th Research Institute of China Electronics Technology Group Corporation (CETC54), Shijiazhuang 050081, China
[2]Hebei Key Laboratory of Electromagnetic Spectrum Cognition and Control, Shijiazhuang 050081, China
[3]School of Economics and Management, Shijiazhuang Tiedao University, Shijiazhuang 050043, China

Corresponding author: Xin Chang (changxinydb@163.com)

This work was supported in part by the China Postdoctoral Science Foundation under Grant 2021M693002.

ABSTRACT To improve the performance of anti-jamming communication in dynamic and adversarial jamming environment, an improved anti-jamming method is proposed based on deep reinforcement learning and feature engineering. Different from the existing studies that use computer vision of deep learning based on the infinite state of spectrum waterfall, the proposed method relays on analyzing spectrum differences between adjacent time slots which contains information and features of jamming patterns. First, anti-jamming strategy is trained by countering the jammer which carries out a random jamming patterns switching strategy. Second, an improved state space is introduced by containing historical spectrum of communication and jamming signal between adjacent time slots, which can help an anti-jamming agent effectively extract the features of jamming patterns to reduce computational complexity. In addition, an improved reward function based on channel switch cost is improved for considering propagation characteristics which may cause communication performance lost. Taking advantage of both feature engineering and deep reinforcement learning, an improved anti-jamming method is proposed to improve reliable anti-jamming performance. Compared with the traditional CNN-based deep reinforcement learning anti-jamming method, simulation results show that the improved method can obtain better performance and lower computational complexity.

INDEX TERMS Anti-jamming, communication, feature engineering, reinforcement learning (RL), deep learning (DL).

I. INTRODUCTION

Because jamming attack will cause communication performance lost, the research on anti-jamming decision-making has become one of the important topics in wireless communication [1]–[5]. In general, to implement anti-jamming communications, anti-jamming decision-making can be performed in the power domain, space domain and the frequency domain.

In the power domain and space domain, based on jamming localization methods, communications can adjust antenna pattern to significantly enhance their anti-jamming capability by increasing the transmitting power and reducing the jamming-signal ratio (JSR) [6], [7]. In addition, signal modulation schemes can be applied to improve anti-jamming performance. The anti-jamming performance is limited to inaccurate measurements, data insufficiencies, incomplete information and linearity of power amplifier [8]–[10].

In recent years, frequency domain anti-jamming decision-making and strategies have received more attention, and the frequency domain anti-jamming methods have been proposed [13]–[15]. Traditional spread-spectrum anti-jamming techniques like Frequency Hopping Spread Spectrum (FHSS), Uncoordinated Frequency Hopping (UFH), and Random Code key Selection using Codebook DSSS (RCSC DSSS) use predefined anti-jamming decision-making

The associate editor coordinating the review of this manuscript and approving it for publication was Quansheng Guan.

schemes [4], [8], [16]. However, with the development of cognitive technology and artificial intelligence, facing the complex and adversarial environment, traditional techniques are difficult to deal with these threats. Hence, the reinforcement learning (RL) has received growing attention in the anti-jamming field recently, and RL methods have been proposed for anti-jamming defense in wireless communications [13]–[17]. The value-based RL methods like Q-learning have been widely applied in decision-making problems. However, traditional value-based RL methods face several challenges from anti-jamming defense perspective [18]–[21]. First, observation is important to RL methods [12]. Considering dynamic environment and changeable jamming patterns, it is important to extract state space from observation which contains raw spectrum information [22]. If features of raw spectrum observation are not extracted, spectrum state will be infinite, then Q-learning is difficult to make decisions [4]. Hence, anti-jamming methods utilize neural network architectures, including convolutional networks [15], to improve anti-jamming performance. Deep RL algorithms are introduced in the field of anti-jamming decision-making. Considering recurrent characteristic of raw spectrum state, a recursion convolutional neural network (RCNN) is designed, which can directly process raw spectrum state [14]. Although deep learning architectures enhance anti-jamming performance with raw spectrum data, training time and computational complexity are sharply increased. By analyzing raw spectrum information, the features of jamming patterns can be extracted from raw time-frequency data to distinguish between jamming patterns, and feature extraction can be carried out to reduce computational complexity and enhance performance [13]. Moreover, feature extraction can be beneficial to achieve high-level finite state-action space by identifying different kinds of jamming patterns. Liu *et al.* identify jamming patterns by feature extraction, and then anti-jamming strategies for each jamming pattern are carried out [4]. Second, a reward function has a significant impact on anti-jamming performance. Classical reward functions are designed for anti-jamming decision-making [4], [14] belong to the short-term reward. Although users can avoid jamming, continual large-scale channel switching reduces communication performance. To short- and long-term rewards, the reward function should be designed by considering the subsequent actions. All in all, decision-making faced by two important problems. First, state space should be achieved by utilizing feature extraction to reduce computational complexity of the decision-making. Second, the reward function should be reshaped by considering communication channel characters and balancing short- long-term reward.

To overcome these disadvantages, a novel anti-jamming method is proposed. The key contributions of this paper are presented as follows:

First, assuming that the jammer carries out a random jamming patterns switching strategy, the anti-jamming strategy can get better jamming performance in the adversarial environment by countering the jamming strategy.

Second, the key to counter the random jamming patterns switching strategy is to construct the improved state space containing the jamming patterns information. In addition, considering the influence of channel characteristics on communication performance, an improved reward function is proposed to reduce the influence of channel switching.

Finally, an improved anti-jamming method is proposed based on the improved anti-jamming RL environment, which combines the advantages of recurrent neural network (RNN) and dueling deep Q network (DQN) methods. The proposed method and RL environment can effectively improve anti-jamming performance.

The remainder of this paper is organized as follows. Section II introduces a classical jamming geometry of a wireless communication and a jammer scenario, and communication needs are identified to improve the anti-jamming performance. Moreover, the improved RL environment is presented including the improved state space and reward function in Section III. The improved anti-jamming RL method is proposed in Section IV. The results of experiments are given in Section V. Finally, conclusions are drawn in Section VI.

II. JAMMING GEOMETRY AND PROBLEM FORMULATION

A. JAMMING GEOMETRY

A classical jamming geometry of a wireless communication scenario, which has been widely discussed and used in many anti-jamming methods [4], [13], [14], is shown in Fig. 1. The jamming geometry consists of one user (a transmitter-receiver pair) and one agent against one jammer.

To simplify the analysis, continuous time is divided into discrete time slots t, which is equal to the duration of one interaction in communication.

The agent is set at the receiver. It has ability of sensing wide-band spectrum to make real-time anti-jamming strategies [4]. At time t, the agent evaluates communication quality and chooses a communication frequency to send signals. A selected frequency of communication is able to be sent to the transmitter through reliable control link. Then, under the control of the agent, the transmitter uses the frequency to send information and builds the reliable communication link.

The jammer aims to destroy the transmission link. The jammer switches jamming patterns randomly or periodically. Jamming frequency is selected to cover communication frequency. Then the transmission link will be destroyed.

The jammer and the agent respectively select jamming frequency and communication frequency at the same time t, and the frequencies hold unchanged during time slot t [15].

B. PROBLEM FORMULATION

In traditional methods, the anti-jamming strategies and performance are discussed in frequency domain and power domain. Considering that once jammers and communication equipment are placed on the battlefield, it is inevitable that jammers will use maximum power in simple scenarios where there are one user (a transmitter-receiver pair) and one

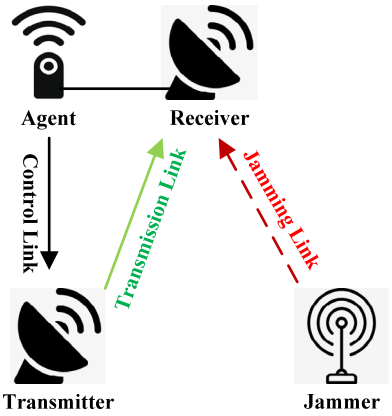


FIGURE 1. Jamming geometry.

jammer, and thus communication will use maximum power. So the effect of jamming energy on anti-jamming strategy will not be discussed in this paper. In addition, to simplify the problem, channelization is performed on jamming frequency and communication frequency.

The communication frequency range is denoted by B_u, communication signal bandwidth is b_u, and then the number of communication channels set is $N = B_u/b_u$. Correspondingly, the jamming frequency range is denoted by B_j, and jamming signal bandwidth is b_j. In order to destroy the transmission link, it is reasonable to assume $B_u = B_j$ and $b_u = b_j$, and then the number of communication channels set is equal to that of jamming channels set. So jamming channel selection set can be defined as $\mathrm{A}_j = \{a_{j,1}, a_{j,2}, \cdots, a_{j,n}, \cdots a_{j,N}\}$ where $a_{j,n} \in [0, N-1]$, and communication channel selection set can be defined as $\mathrm{A}_u = \{a_{u,1}, a_{u,2}, \cdots, a_{u,n}, \cdots a_{u,N}\}$ where $a_{u,n} \in [0, N-1]$. At time t, $a_{t,j} \in \mathrm{A}_j$ is defined as the selected jamming channel of the jammer, and $a_{t,u} \in \mathrm{A}_u$ is defined as the selected communication channel of the agent.

The spectrum vector of communication band which can be observed by the agent and the jammer is $O_t = \{o_{t,1}, o_{t,2}, \cdots o_{t,n}, \cdots, o_{t,N}\}$, where $o_{t,n}$ represents the jamming and communication channel occupancy of channel n at time t.

Set ϕ as an evaluation indicator function for successful transmission, which can be presented as follows:

$$\phi\left(a_{t,j}, a_{t,u}\right) = \begin{cases} 1 & a_{t,j} \neq a_{t,u} \\ 0 & a_{t,j} = a_{t,u} \end{cases} \tag{1}$$

when the selected jamming channel of the jammer $a_{t,j}$ is equal to the selected communication channel of the agent $a_{t,u}$, the transmission is seen as failed. The aim of the agent is to maximize the target function:

$$\max_{a_{t,u} \in \mathrm{A}_u} \sum_{t=0}^{\infty} \gamma^t \phi\left(a_{t,j}, a_{t,u}\right) \tag{2}$$

where γ is a discount factor.

So, at time t, the agent observes spectrum vector of communication band O_t, evaluate communication quality by using the evaluation indicator function ϕ and chooses a communication channel $a_{t,u}$ to send information. A communication channel is able to be sent to the transmitter through a reliable control link. Then, under the control of agent, the transmitter uses the communication channel to send information and builds the reliable communication link.

The above formulation contains three major parts of RL environment: Actions, Evaluation and Observation. The target function seems to be able to be maximized by utilizing the RL methods, and the reliable communication link can be built against the jammer.

However, under malicious jamming environment, it is impossible for the agent to acquire the jamming strategies or rules. So assuming that the jamming patterns transition of the jammer, including fixed frequency, positive step-frequency and negative step-frequency, satisfies a specific transition probability, jamming patterns transition should satisfy equal probability as shown in Fig. 2.

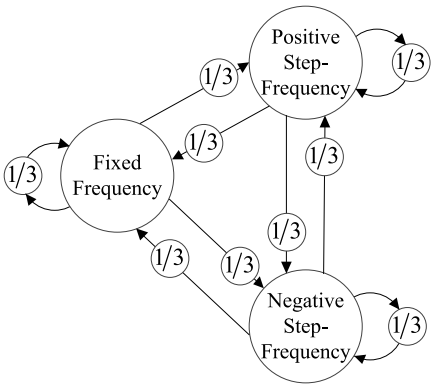


FIGURE 2. Jamming patterns transition graph with transition probability.

So the key to the anti-jamming improvement is how to make use of observation sets to make agents have the ability to distinguish jamming patterns and make real-time actions. Because jamming patterns can be identified by comparing historical information between adjacent time slots, the observation set is transformed into state space by feature engineering, in which state space contains historical information of the jammer and communication, and the improved RL method should be designed to meet the improved state space.

In addition, that an evaluation indicator function only includes frequency coverage is too simple to counter the jammer. In the spectrum domain, because the reconstruction of the transmission link needs settling time for radio frequency (RF) devices and different frequencies could have different propagation characteristics which may cause a difference in digital signal processing [8], channel switching effect between adjacent time slots should be considered to build an improved reward function.

III. FEATURE ENGINEERING FOR RL ENVIRONMENT

The key of this section is to construct state space and reward function through feature engineering, and then the anti-jamming performance will be improved.

A. FEATURES OF JAMMING PATTERNS

The core of the anti-jamming strategy is that the jamming patterns switching transition and jamming channel switching rule under jamming patterns can be identified. First, irregular jamming patterns switching transition is difficult to be used by communication. The reason is that it is impossible to predict the jamming channel selected by the jammer after the jamming patterns switching. In addition, the probability of the selected communication channel is equal to that of the jamming channel, so the agent cannot select the communication channel by predicting the jamming channel. In addition, the choice of communication channel should be considered in combination with the performance of channel switching. Then, the key of the anti-jamming strategy is to take advantage of the transfer rule of the jamming channel switching and then to avoid jamming channel under the jamming pattern. According to the domain knowledge, jamming patterns adopted by the jammer can be obtained from the spectral analysis between adjacent time slots. Through the analysis of Fig. 3, it can be seen that the features of the jamming patterns are extracted in the jamming channel between adjacent time slots.

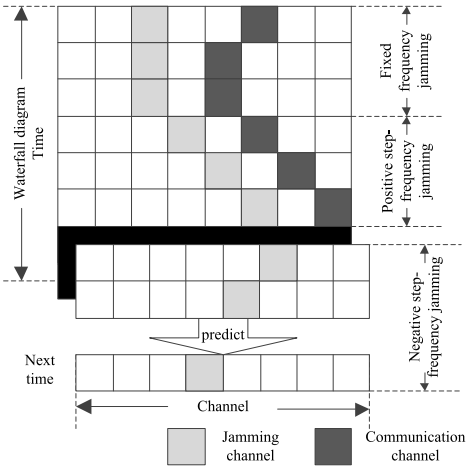


FIGURE 3. Features of jamming patterns.

To sum up, it is necessary to use feature engineering to extract features of jamming patters between adjacent time slots to form the state space.

In addition, it is necessary to design a new reward function based on anti-jamming evaluation indication to realize jamming avoidance. First, if the selected jamming channel and the selected communication channel are the same, the reward function will obtain the highest penalty. Then, if the selected jamming channel and selected communication channel are different, the dense reward function needs to be set according to channel distance among time t, time t-1 and time t-2. The larger the distance is, the higher the penalty of the reward function is.

B. STATE SPACE DESIGN

The improvement process from the observation space to state space is shown in Fig. 4.

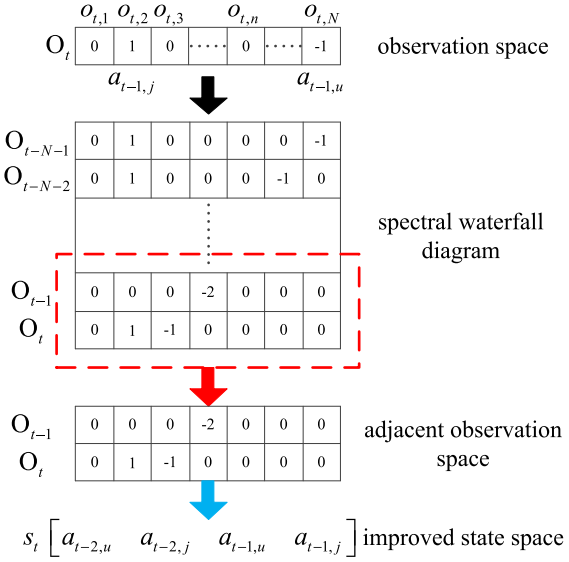


FIGURE 4. Improved state space.

As for the observation space, the channel that is not selected, will be set 0. The channel selected by the jammer will be set -1, the channel selected by the communication will be set 1, and the channel simultaneously selected by the jammer and communication will be set -2. The disadvantage of the observation space is that the temporal observation space cannot directly represent the current jamming pattern without RNN model [13], [14]. Inspired by the deep RL methods from raw video data [11], the spectral waterfall diagram is composed of the historical observation spaces, which collects time and frequency domain information and contains the jamming patterns information. Although a deep convolutional neural network (CNN) is good at image processing [11], it will lead to relatively heavy computational complexity and weak real-time application. To avoid this limitation, adjacent observation space should be reformed to present the jamming patterns information. Furthermore, the action information is directly extracted by feature engineering. Considering that the observation space is composed by the jammer and communication actions, the improved state space is directly composed by the historical jammer and communication actions.

The improved state space extracts jamming patterns from observation information, reduces computational complexity and enhance real-time application.

C. REWARD FUNCTION DESIGN

In this section, a performance evaluation of the reward function will be improved and redefined. Agents should have anti-jamming ability to balance the short- and long-term reward through merging the spectrum coverage and channel switching. First, for the spectrum coverage, when the selected jamming channel of the jammer $a_{t,j}$ is equal to the selected communication channel of the agent $a_{t,u}$, the transmission is failed, that is, $r_t = -1$. Second, for the channel switching, although the selected jamming channel of the jammer $a_{t,j}$ is not equal to the selected communication channel of the agent $a_{t,u}$, the distance among $a_{t,u}$, $a_{t-1,u}$ and $a_{t-2,u}$ will lead to the channel switching cost. Then the improved reward function r_t can be represent as follows:

$$r_t = \begin{cases} -1 & a_{t,j} = a_{t,u} \\ -\dfrac{\lambda \left| a_{t,u} - a_{t-1,u} \right|}{N} & \\ \quad -\dfrac{(1-\lambda) \left| a_{t-1,u} - a_{t-2,u} \right|}{N} & a_{t,j} \neq a_{t,u} \end{cases} \tag{3}$$

where λ is the switching cost factor.

By adjusting the switching cost factor λ, the improved reward function represents a balance between the spectrum coverage effect and the channel switching cost. By measuring the distance from the selected jamming to the communication channel between adjacent time slots, the desired balance between anti-jamming effect and system overhead can be maintained more elaborately, and this design is beneficial to improve anti-jamming ability and communication performance of the agent.

IV. IMPROVED REINFORCEMENT LEARNING ANTI-JAMMING METHOD DESCRIPTION

The improved RL anti-jamming method is introduced from three aspects: architecture, policy and algorithm.

A. ARCHITECTURE

The anti-jamming architecture is shown in Fig. 5. It is similar to dueling deep Q learning [21], but the first post convolutional fully connected layer is replaced with RNN to understand the previous the improved state space information.

Thus, the improved state space is passed as an input to the RNN layer. The RNN layer has the memory for holding historical state information, which will be of benefit to identify jamming patterns. The RNN layer retains information about previous important states and updates its memory over time steps as required [12]. Its outputs are passed as input to the different full-connected (FC) layers. They abstract the historical features of the improved state space. The resulting feature states are a scalar V and an N-dimensional vector A. Then the values of the state-action pair can be estimated as follows:

$$Q(s,a) = V(s) + \left[A(s,a) + \frac{1}{N} \sum_{a'} A(s,a') \right] \tag{4}$$

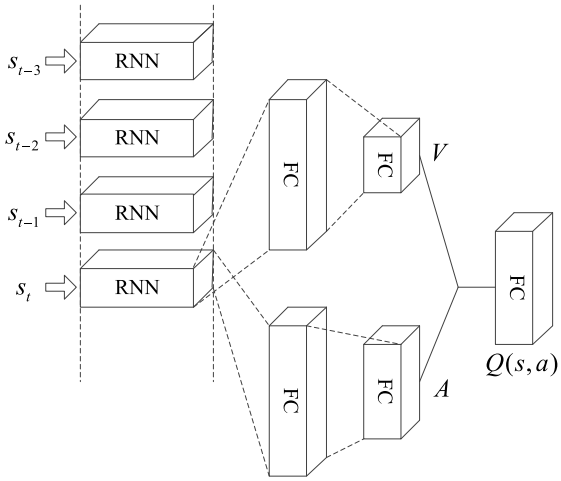


FIGURE 5. Improved anti-jamming architecture.

B. POLICY

In the case of the anti-jamming architecture, the architecture parameters of the deep neural network, including weights and biases, are collectively denoted as θ at each time t. The entire episodes are stored in an experience buffer D and m steps are randomly sampled from a random batch of episodes to train the network. The loss of the anti-jamming network can be determined as follows:

$$L_t(\theta_t) = E_{e_t \sim D}\left[\left(r_t + \gamma \max_{a'} \hat{Q}(S_t, a'; \theta') - Q(S_t, a_t; \theta_t) \right)^2 \right] \tag{5}$$

where θ' is the parameters of the target architecture $\hat{Q}$, and e_t is experience (s_t, r_r, a_t, s_{t+1}).

To improve the fitting accuracy of the training neural network, the architecture parameters can be optimized as follows:

$$\theta_{t+1} = \theta_t + \nabla_\theta L_t(\theta_t) \tag{6}$$

where ∇ is the gradient operation.

C. ALGORITHM

The proposed algorithm for the anti-jamming method based on deep reinforcement learning is presented in Algorithm 1.

V. SIMULATION RESULTS AND ANALYSIS

As the principle of deep learning is difficult to explain, its performance cannot be proved and predicted by mathematical derivation, and the research on the interpretation of deep learning is difficult in the whole industry. This paper focuses on engineering applications, so it does not carry out theoretical exploration on its interpretation. By analyzing the performance of experimental results, this paper tries to analyze the reasons for the improved anti-jamming performance, and it shows that the feature engineering combined with domain knowledge can effectively improve the anti-jamming performance based on RL methods.

Algorithm 1 Improved Deep Recurrent Reinforcement Learning

Initialize: Set $D = \phi$, $i = 0$, θ is with random weights and biases, $s_1 = \mathbf{0}$
For episode $= 1, 2, \cdots,$ episode$_{\max}$ **do**
  For $t = 1, 2, \cdots, t_{\max}$ **do**
    Choose $a_{t,u}$ via the ε-greedy algorithm
    Execute action $a_{t,u}$ and compute r_t and acquire s_{t+1}
    Store $(s_t, a_{t,u}, r_t, s_{t+1})$ in D
    If *sizeof(D) > length of experience buffer*
      Sample random mini-batch of transitions
      $e_t = (s_t, a_{t,u}, r_t, s_{t+1})$ from D
      Compute $\nabla_\theta L_t(\theta_t)$, update θ_t, and $t := t + 1$
  End for
End for

In the simulation setting, the user and the jammer combat with each other, and the number of available channels in the simulation is 10. For training the anti-jamming agent, the max episodes is set as 500, and the number of times is set as 5000 at one episode. For evaluating the anti-jamming performance, the max episodes is set as 100, and the number of times is set as 5000 at one episode. The memory pool capacity is 256 and its batch size is set as 128. The switching cost factor is set as 0.9, the greedy factor is set as 0.9 and the learning rate is set as 0.001.

The period of jamming pattern switching is 10. Three kinds of jamming patterns are considered for simulation: i) Fixed frequency jamming; ii) Positive step-frequency jamming; iii) Negative step-frequency jamming.

This experiment is divided into two parts. First, the original RL environment and the improved RL environment are used to make decisions under the same RL method, and the improved RL environment based on feature engineering can effectively improve the performance of the anti-jamming method. In the second part, compared with the anti-jamming performance of the traditional CNN-based method, the anti-jamming performance of the proposed method is verified with higher performance and lower complexity.

A. THE EFFECT OF IMPROVED RL ENVIRONMENT

By utilizing the Q-learning method, the anti-jamming performance is presented and compared in the original RL environment and the improved RL environment.

As shown in Fig. 6, the thermodynamic diagram in different RL environment are presented under three kinds of jamming patterns. The agent trained by the original RL environment is defined as the original agent, and the agent trained by the improved RL environment is defined as the improved agent. In Fig. 6, the black block is jamming spectrum, and the white block is communication spectrum. As shown in Fig. 6 (a) and (b), the original agent and the improved agent successfully avoids positive step-frequency jamming. However, it can be clearly observed that the channel switching variance of the improved agent is lower than that of the original agent. As shown in Fig. 6 (c) and (d), the improved agent can select the same channel under fixed frequency

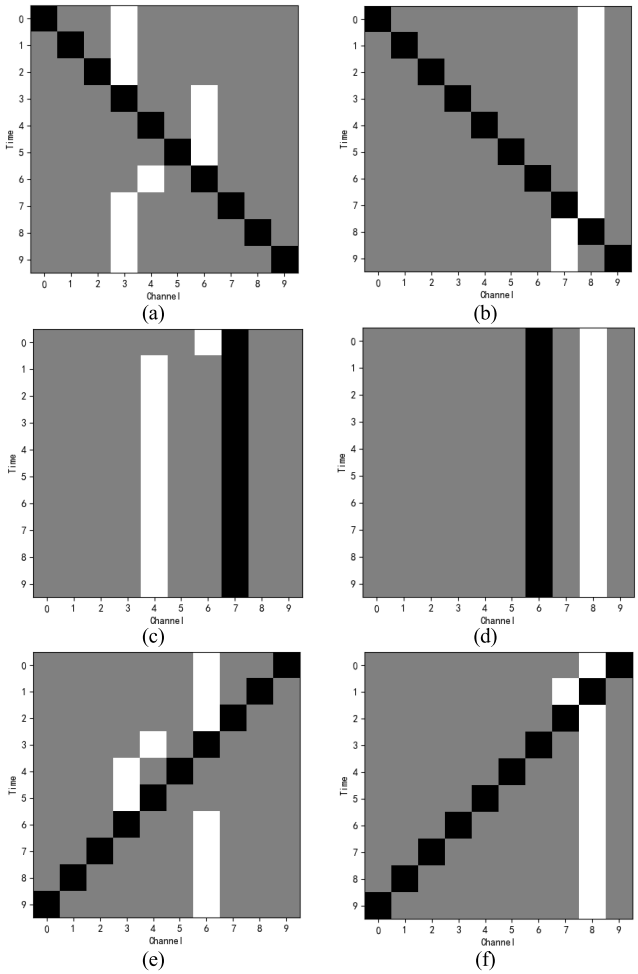


FIGURE 6. Comparison diagram under the different RL environments.

jamming, and the original agent will leads to heavier channel switching cost. As shown in Fig. 6 (e) and (f), considering channel switching cost, it is obvious that the improved agent has higher anti-jamming performance than the original agent under negative step-frequency jamming.

As shown in Fig. 7, the anti-jamming performance comparison between the original agent and improved agent is described under the random jamming pattern switching strategy. The anti-jamming performance in the improved RL environment is more effective than that in the original RL environment. Hence, it is very important that the RL environment should be designed by feature engineering and domain knowledge to improve the anti-jamming performance of RL agent.

The reason for improvement is described in section III.A. In terms of experiment results, in combination with domain knowledge, it is able to extract the features of jamming patterns used by the jammer through analyzing the state between adjacent time slots. Thus, the selected rule of the jamming pattern is used by the agent to choose the optimal communication channel and avoid the jamming channel. On the contrary, the traditional RL environment uses the spectrum

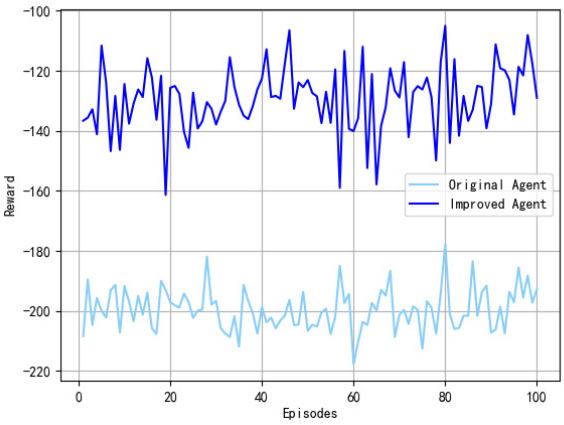


FIGURE 7. Performance comparison under the different RL environments.

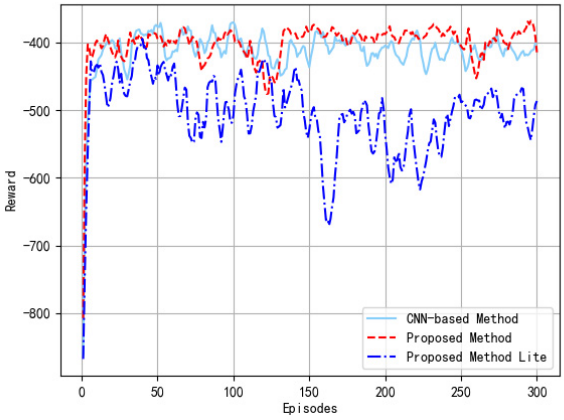


FIGURE 8. Training performance comparison.

information of the current moment as the state, which cannot effectively extract the jamming patterns information hidden in the adjacent spectrum observation.

B. THE IMPROVED ANTI-JAMMING EFFECT

As shown in Fig. 8, it presents that the training performance comparison among the traditional CNN-based method, the proposed method and the proposed method lite. The difference between the proposed method and the proposed method lite is that the state space is composed of only one time slot under the improved RL environment. Considering that the greedy factor is 0.9, the reward of the proposed method is lightly more than that of the CNN-based method. The reward of the proposed method lite is less than that of CNN-based method.

In Fig. 9, after training, the trained performance comparison among the traditional CNN-based method, the proposed method and the proposed method lite are presented to verify validation. Under the improved RL environment, the reward of proposed methods are obviously more than that of the traditional CNN-based method. In addition, the reward of the proposed method is more than that of the proposed method lite.

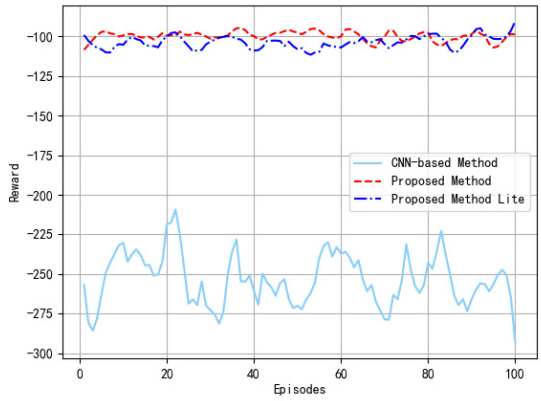


FIGURE 9. Performance comparison under the different RL environments.

TABLE 1. Comparison of performance indices.

Method	Reward Average	Calculated Time
CNN-based Method	-84.45	0.129 s
Proposed Method	-33.31	0.002 s
Proposed Method Lite	-34.57	0.002 s

The comparison of performance indices is shown in Table 1. The indices concludes the averaged reward and the calculated time to further comprehensively evaluate the anti-jamming performance.

The averaged reward of the proposed method is more than that of the others. In additions, the calculated time is less than that of the others. Hence, the proposed method effectively avoid malicious jamming and achieve reliable communication.

In the second part, it is verified that the experience obtained from the experiments of feature engineering is important to improve anti-jamming performance compared with the traditional CNN-based method.

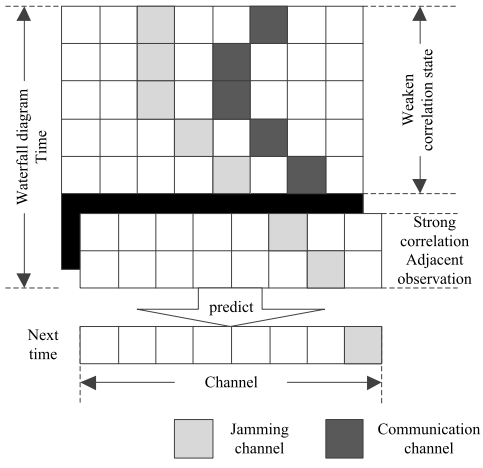


FIGURE 10. Compared with improved method and CNN-based method.

As shown in Fig. 10, the reason is that RNN and data stacking operation have the ability to extract rules from adjacent states, while CNN is more focused on the whole

waterfall state. The farther the data is, the less contribution it makes. In other words, the waterfall state is mixed with a large amount of irrelevant information, so the neural network cannot focus on the part of state that can extract the selected rule and features of jamming patterns.

In addition, the lightweight proposed method lite, which only uses the improved neural network structure, can also approach the performance of the improved method based on feature engineering. It is further demonstrated that the RL environment constructed by feature engineering based on domain knowledge can not only improve the anti-jamming performance, but also quickly verify the design of methods, and then the construction of neural network structure can be effectively guided.

VI. CONCLUSION

In order to improve the anti-jamming performance of communication in complex and adversarial environment, an improved anti-jamming method is proposed in this paper. The anti-jamming performance in complex environment can be improved by countering jammers with a random jamming patterns switching. Feature engineering is used to improve the RL environment including an improved state space and a reward function, which can effectively reduce computational complexity. An improved state space containing jamming pattern information and a reward function reflecting the effect of channel switching are constructed. In addition, under the improved RL environment, this paper proposes an improved anti-jamming method. In the simulation experiments, the anti-jamming performance under different RL environment is compared, and the averaged reward, calculated time and frequency spectrum diagram show the effectiveness of the improved RL environment. Then, the comparison between the improved methods and the traditional method shows that the improved method can get better performance, and it is very important to use feature engineering to construct RL environment to improve anti-jamming performance. Combining with the feature engineering and domain knowledge can effectively improve the anti-jamming performance, reduce the computational complexity and benefit for the engineering implementation. The anti-jamming scenarios of multi-channel reactive jammer will be of great important for the study on anti-jamming methods. These scenarios can be further studied with novel jamming patterns and jammers in future.

REFERENCES

[1] Y. E. Sagduyu, R. A. Berry, and A. Ephremides, "Jamming games in wireless networks with incomplete information," *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 112–118, Aug. 2011.

[2] K. Grover, A. Lim, and Q. Yang, "Jamming and anti-jamming techniques in wireless networks: A survey," *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 17, no. 4, pp. 197–215, Dec. 2014.

[3] B. Gopalakrishnan and M. A. Bhagyaveni, "Random codekey selection using codebook without pre-shared keys for anti-jamming in WBAN," *Comput. Electr. Eng.*, vol. 51, pp. 89–103, Apr. 2016.

[4] S. Liu, Y. Xu, X. Chen, X. Wang, M. Wang, W. Li, Y. Li, and Y. Xu, "Pattern-aware intelligent anti-jamming communication: A sequential deep reinforcement learning approach," *IEEE Access*, vol. 7, pp. 169204–169216, 2020.

[5] L. Xiao, *Anti-Jamming Transmissions in Cognitive Radio Networks*. Cham, Switzerland: Springer, 2015.

[6] J. Xu, S. Zhu, and G. Liao, "Range ambiguous clutter suppression for airborne FDA-STAP radar," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 8, pp. 1620–1631, Dec. 2015.

[7] J. Xu, G. Liao, L. Huang, and H. C. So, "Robust adaptive beamforming for fast-moving target detection with FDA-STAP radar," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 973–984, Feb. 2017.

[8] L. Jia, Y. Xu, Y. Sun, S. Feng, and A. Anpalagan, "Stackelberg game approaches for anti-jamming defence in wireless networks," *IEEE Wireless Commun.*, vol. 25, no. 6, pp. 120–128, Dec. 2018.

[9] L. Zhang, Z. Guan, and T. Melodia, "United against the enemy: Anti-jamming based on cross-layer cooperation in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 8, pp. 5733–5747, Aug. 2016.

[10] L. Jia, Y. Xu, Y. Sun, S. Feng, L. Yu, and A. Anpalagan, "A multi-domain anti-jamming defense scheme in heterogeneous wireless networks," *IEEE Access*, vol. 6, pp. 40177–40188, 2018.

[11] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Jan. 2015.

[12] S. Ravivhandiran, *Hands-On Reinforcement Learning With Python: Master Reinforcement and Deep Reinforcement Learning Using OpenAI Gym and TensorFlow*. Birmingham, U.K.: Packt, 2018.

[13] O. Naparstek and K. Cohen, "Deep multi-user reinforcement learning for distributed dynamic spectrum access," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 310–323, Nov. 2019.

[14] X. Liu, Y. Xu, L. Jia, Q. Wu, and A. Anpalagan, "Anti-jamming communications using spectrum waterfall: A deep reinforcement learning approach," *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 998–1001, May 2018.

[15] Y. Li, X. Wang, D. Liu, Q. Guo, X. Liu, J. Zhang, and Y. Xu, "On the performance of deep reinforcement learning-based anti-jamming method confronting intelligent jammer," *Appl. Sci.*, vol. 9, no. 7, p. 1361, 2019.

[16] H. Zhu, C. Fang, Y. Liu, C. Chen, M. Li, and X. S. Shen, "You can jam but you cannot hide: Defending against jamming attacks for geo-location database driven spectrum sharing," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 10, pp. 2723–2737, Oct. 2016.

[17] X. Liu, Y. Xu, Y. Cheng, Y. Li, L. Zhao, and X. Zhang, "A heterogeneous information fusion deep reinforcement learning for intelligent frequency selection of HF communication," *China Commun.*, vol. 15, no. 9, pp. 73–84, 2018.

[18] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," Dec. 2013, *arXiv:1312.5602*.

[19] H. van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," Dec. 2015, *arXiv:1509.06461*.

[20] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized experience replay," Feb. 2016, *arXiv:1511.05952*.

[21] Z. Wang, T. Schaul, M. Hessel, H. van Hasselt, M. Lanctot, and N. de Freitas, "Dueling network architectures for deep reinforcement learning," Apr. 2016, *arXiv:1511.06581*.

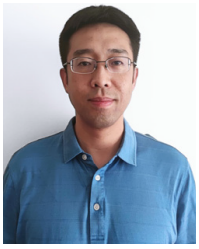
[22] L. Xiao, X. Lu, D. Xu, Y. Tang, L. Wang, and W. Zhuang, "UAV relay in VANETs against smart jamming with reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4087–4097, May 2018.

[23] M. Hausknecht and P. Stone, "Deep recurrent Q-learning for partially observable MDPs," Jan. 2017, *arXiv:1507.06527*.

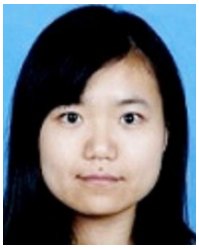
XIN CHANG was born in Shijiazhuang, Hebei, China, in 1990. He received the B.S. degree in electrical engineering from Handan University, Handan, China, in 2014, and the M.Eng. degree in electronics and communication engineering and the Ph.D. degree in electronics science and technology from the School of Electronic Engineering, Xidian University, Xi'an, China, in 2017 and 2020, respectively. He is currently a Postdoctoral Researcher with The 54th Research Institute of China Electronics Technology Group Corporation (CETC54), Shijiazhuang, and Xidian University. His main research interests include electronic countermeasure (ECM), electronic warfare system simulation, and cognitive electronic warfare.

YANBIN LI was born in Shijiazhuang, Hebei, China, in 1966. He received the B.S. degree from Tianjin University, Tianjin, China, in 1985, the M.E. degree from The 54th Research Institute of China Electronic Technology Group Corporation (CETC54), Shijiazhuang, in 1988, and the Ph.D. degree from Shanghai Jiaotong University, Shanghai, China, in 1995. He is currently a Chief Expert of China Electronic Technology Group Corporation and a Chief Scientist of CETC54. His main research interests include electronic countermeasure (ECM) and cognitive electronic warfare.

YAN ZHAO was born in Gansu, China, in 1981. He received the B.S. degree from the School of Telecommunications Engineering, Xidian University, Xi'an, China, in 2003, and the M.S. degree from the School of Electronic Engineering, Xidian University, in 2009, where he is currently pursuing the Ph.D. degree with the National Laboratory of Radar Signal Processing. He is also with The 54th Research Institute of China Electronics Technology Group Corporation (CETC54), Shijiazhuang, China. His main research interests include array signal processing and radar signal processing.



YUFENG DU was born in Shanxi, China, in 1981. He received the bachelor's degree in electrical engineering and automation and the master's degree in electronic and communication engineering from Xidian University, in July 2003 and December 2012, respectively. He is currently the Deputy Director of the Hebei Key Laboratory of Electromagnetic Spectrum Cognition and Control, The 54th Research Institute of China Electronics Technology Group Corporation (CETC54), Shijiazhuang, China.



DONGHUI LIU received the Ph.D. degree in management from the Chinese University of Geosciences, Beijing. She studied at Mcmaster University for a period of one year. She is currently a Lecturer with the School of Economics and Management, Shijiazhuang Tiedao University. She engaged in the research of complex system analysis.