ZhuanSun et al. EURASIP Journal on Wireless Communications and Networking (2019) 2019:85
https://doi.org/10.1186/s13638-019-1414-4

EURASIP Journal on Wireless Communications and Networking

RESEARCH Open Access

An algorithm for jamming strategy using OMP and MAB

Shaoshuai ZhuanSun[1,2*], Jun-an Yang[1,2] and Hui Liu[1,2]

Abstract

Reinforcement learning (RL) has the advantage of interaction with an environment over time, which is helpful in cognitive jamming research, especially in an electronic warfare-type scenario, in which the communication parameters and jamming effect are unknown to a jammer. In this paper, an algorithm for a jamming strategy using orthogonal matching pursuit (OMP) and multi-armed bandit (MAB) is proposed. We construct a dictionary in which each atom represents a symbol error rate (SER) curve and can be obtained with known noise distribution and deterministic parameters. By reconnoitering, the jammer counts acknowledge/not acknowledge (ACK/NACK) frames to calculate the SER, which is also regarded as samples that are sampled from the real SER curve using an MAB. When we obtain the sampled sequence and the constructed dictionary, the OMP algorithm is used to search and locate atoms and its corresponding coefficients. With the searching results, the jammer can construct an SER curve that is similar to the real SER curve. The experimental results demonstrate that the proposed algorithm can learn an optimal jamming strategy with three interactions, which converges substantially faster than the state of the art.

Keywords: Reinforcement learning, Cognitive jamming, Orthogonal matching pursuit, Multi-armed bandit, Interaction times

1 Introduction

Wireless communication has extensive utilization in civilian and military domains with the advantage of convenience [1–3]. However, the inherent openness of a wireless medium renders it susceptible to adversarial attacks [4]. Three categories of jamming methods can be presented as follows: (1) Reconnaissance, evaluation, and jamming—The jammer collects the required information, such as modulation scheme, transmission power, and communication protocols, and takes some targeted actions, such as denial of service (DOS) attack, eavesdropping attack, and correlation attack or hybrid attack. (2) Game theory—When both jammer and communicators can recognize the existence of each other, actions such as jamming or anti-jamming are performed to conquer each other. As a result, Nash equilibrium is a suitable final result, even though it does not exist or may require a considerable length of time to acquire [5]. (3) Reinforcement learning (RL) [6]—Trial and error is the key of RL, and prior information is not necessary to the jammer. Generally, the learning process is often modeled as a multi-armed bandit (MAB) [7], and the purpose to identify the best bandit, which also indicates the optimal jamming strategy. In this paper, we investigate the ability of an agent to learn an efficient jamming strategy with sparse representation and RL.

To fulfill the requirement of communication denial, jamming is a direct choice and has numerous research outcomes. Prior to jamming, an attacker should conduct reconnaissance of the battlefield and evaluate elements of jamming that can be useful when making jamming decisions [8]. The major disadvantage of these methods is that it assumes that the jammer has accurate information about environmental factors and receiver actions. For example, adaptive zero adjustment technology will decay the power of jamming signals, error detection and correction technology can reduce the symbol error rate (SER) of the received information, and anti-jamming methods such as in-phase and quadrature (IQ) imbalance [9] or anti-chirp-jamming [10] can fade the jamming signal influence. Game theory is a dynamic process between the jammer and the transmitter-receiver pairs; it can build a Nash equilibrium between both sides [11],

* Correspondence: zhuansunss@sina.com
[1]Key Laboratory of Electronic Restriction, Hefei 230037, Anhui, China
[2]Electronic Countermeasure Institute, National University of Defense Technology, Hefei 230037, Anhui, China

but the jammer needs to employ an efficient jamming strategy, which is also the purpose of this paper. As a branch of machine learning, the RL feedback (reward) is less informative than that in supervised learning, where the agent would be given the correct actions to take (this information is not always available). The RL feedback is, however, more informative than that in unsupervised learning, where there is no explicit feedback on the performance. The advantage of the RL is that the agent does not need to know the environment model or rules; only the feedback from the environment is needed. Therefore, RL has attracted a significant amount of attention for robots, the game field, cognitive radio, and cognitive jamming. Examples of cognitive radio, in which RL has been applied, are dynamic channel selection, channel sensing, and routing. In cognitive jamming, RL is employed to learn the jamming scheme in a physical layer, a jamming frame in a media access control (MAC) layer [12], and jamming nodes in a blind network [13]. Although RL does not need prior information and is convenient for implementation, it has the disadvantage of slow convergence, which is the limitation of its application.

We propose a novel algorithm for a jamming strategy, which combines the advantage of orthogonal matching pursuit (OMP) [14] and MAB. The proposed algorithm fully utilizes prior information, such as the distribution of the channel noise and the modulation scheme of the communication signals to construct a dictionary that contains various SER curves. The algorithm jams the in-phase and quadrature phase, which corresponds to a reward, such as the SER can be calculated by counting acknowledge/not acknowledge (ACK/NACK) frames. The algorithm regards the received SER values as sampled samples and searches the optimal atoms with the OMP from the constructed dictionary. The jammer can predict the SER curves of both the in-phase and the quadrature phase, which will guide the jammer to choose the optimal jamming strategy. The experimental results demonstrate that with proper samples, the proposed algorithm only needs three interactions with the environment, which is considerably less than the state of the art.

The remainder of this paper is organized as follows: In Section 2, the model of the jamming environment between the communicators and the jammer is presented, and the formula for generating a dictionary is deduced. Section 3 establishes our jamming strategy learning algorithm that is based on OMP and MAB. Section 4 compares the performance of the algorithm in [4, 15–17] with our algorithm. The simulation results verify the efficiency of the proposed algorithm. Section 5 concludes the paper.

2 System model

In a real-time jamming environment, too many factors can influence the jamming effect. For example, both the transmitter power and the jamming power would be decayed by obstacles in the transmission path. The jamming power would be further decayed by receiver's anti-jamming actions, such as amplitude limiting and adaptive zero attenuation. Figure 1 depicts a real-time jamming environment in wireless communication. The low-pass equivalent of a received signal is represented as $r_m = \sqrt{\alpha P_{\mathrm{T}}}\, x_m + \sqrt{\beta P_{\mathrm{J}}}\, j_m + n_m$, $m = 1, 2, \cdots$, where P_{T} is the transmitted signal power, P_{J} is the jamming signal power, x_m denotes the modulated symbols, j_m presents the jamming symbols, and n_m is the Gaussian or Rayleigh noise with the power N_0. We denote α and β as decay factors that belong to the transmission signal and the jamming signal, respectively.

Consider an additive white Gaussian noise (AWGN) scenario, where the communicators use multiple quadrature amplitude modulation (MQAM) modulated signals and the jammer uses binary phase shift keying (BPSK) modulated signals. The average SER in the receiver is given by:

$$\xi = \frac{(X-1)}{2X}\eta \cdot erfc\left[\sqrt{\alpha \cdot P_{\mathrm{T}}/N_0} - \sqrt{2\beta \cdot \gamma \cdot P_{\mathrm{J}}/N_0}\right] \tag{1}$$

where the parameter η presents a discount factor in the receiver for error detection and correction reasons, which remains unknown to the jammer, and X denotes the dimensions in the in-phase of the communication signals. Eq. (1) can also be written as:

$$\xi = \frac{(X-1)}{2X}\eta \cdot erfc\left[\sqrt{P_{\mathrm{T}} + \phi} - \sqrt{P_{\mathrm{J}}/\phi}\right] \tag{2}$$

In Eq. (2), $erfc(\cdot)$ is a monotonically decreasing function, where $\phi = P_{\mathrm{T}}(\alpha - N_0)/N_0$, $\phi = N_0/(2\beta \cdot \gamma)$, and P_{T}, ϕ,

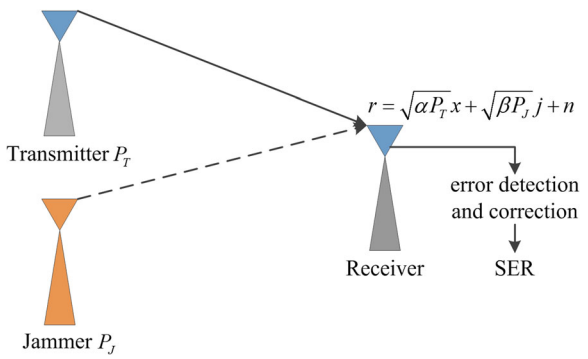

Fig. 1 Real-time jamming environment in wireless communication. The jamming model indicates some decay factors that can influence the jamming effect, such as a power lost in the transmission path or a decreased SER by error correction

and ϕ are unknown to the jammer. Equation (2) can also be written as:

$$\xi = \frac{(X-1)}{2X}\eta \cdot erfc\left[\sqrt{P_{\mathrm{F}}+\omega}-\sqrt{P_{\mathrm{J}}/\varpi}\right] \tag{3}$$

where P_{F} has a fixed value that is assigned by the jammer, and the parameters $\omega = P_T + \phi - P_F$ and $\varpi = \phi$. The jammer can obtain an SER curve $\xi' = \frac{(X-1)}{2X} \cdot erfc[\sqrt{P_{\mathrm{F}}}-\sqrt{P_{\mathrm{J}}}]$ by assuming that the communication signal has power P_{F} and the true SER curve is similar to ξ'. The difference is that we should have ξ' stretched, compressed, or shifted to coincide with the true SER curve or use a linear combination of several constructed curves to represent the true SER curve.

With different types of ω, ϖ values, we will have various SER curves, within which several curves are needed to construct the true SER curve. How do we obtain these curves? We take advantage of the trial and error by RL with the searching method of sparse representation [18]. When the dictionary is constructed with a priori knowledge and the sampled samples are obtained by interacting with the environment, the sampling sequence should be linearly represented by k (also known as sparsity) atoms in the dictionary. Therefore, the jammer can apply sparse representation algorithms to search for potential atoms. Although optimization algorithms such as a genetic algorithm can be employed to search for atoms, they can only search for the best atom and cannot accurately represent the sampling sequences.

In terms of different norm minimizations that are applied in sparsity constraints, the sparse representation methods can be roughly categorized into five groups: (1) l_0-norm minimization, (2) l_p-norm ($0 < p < 1$) minimization, (3) l_1-norm minimization, (4) $l_{2,1}$-norm minimization, and (5) l_2-norm minimization. We note that the greedy iterative algorithms that solve the sparse representation method with l_0-norm minimization have the characteristics of low complexity and an extensive range of applications. The greedy iterative algorithms include MP [19], OMP [20], regularized OMP (ROMP) [21], and stagewise OMP (StOMP) [22]. The computational complexity of MP and OMP is $O(N^2k^2)$, where N denotes the atomic dimension. The algorithms of ROMP and StOMP have a lower computational complexity $O(Nk^2)$ but also a poor reconstruction performance. As a result, we use the OMP that converges faster than MP to search for atoms, and algorithms that are better than OMP await further research.

The general formula of sparse representation is $\boldsymbol{Y} = \boldsymbol{DX}$, where $\boldsymbol{Y}$ is the sampled sequence, $\boldsymbol{D}$ denotes the over-complete atomic dictionary, and $\boldsymbol{X}$ is the sparse coefficient. We assume that the jammer jams the in-phase with power P_{J_m}, P_{J_n} and then regards the received feedback ξ_m, ξ_n as the action's reward. With the received data, the formula $\boldsymbol{Y} = \boldsymbol{DX}$ can also be written as:

$$\begin{bmatrix} \vdots \\ \xi_m \\ \vdots \\ \xi_n \\ \vdots \end{bmatrix}_{N\times 1} = \begin{bmatrix} \cdots & \vdots & \cdots & \vdots & \cdots \\ & d_{mm} & & d_{mn} & \\ & \vdots & & \vdots & \\ & d_{nm} & & d_{nn} & \\ \cdots & \vdots & \cdots & \vdots & \cdots \end{bmatrix}_{N\times M} \cdot \begin{bmatrix} \vdots \\ x_m \\ \vdots \\ x_n \\ \vdots \end{bmatrix}_{M\times 1} \tag{4}$$

In Eq. (4), $\boldsymbol{Y}$ could be seen as a linear combination of atoms in $\boldsymbol{D}$, and the position and coefficient of the chosen atoms are marked in $\boldsymbol{X}$.

3 An algorithm for jamming strategy

In the proposed algorithm, we use both MAB and OMP technology. With MAB, we can obtain sampled data, which are necessary in Eq. (4). With the latter, we can obtain the best atoms, which are used to predict the SER curve. Figure 2 shows the process of the proposed algorithm, in which dictionary construction should be completed by reconnoitering the environment. This environment includes communication signals, jamming signals, noise and feedback signals transmitted by the receiver. The following details about the proposed algorithm are provided.

3.1 Reward standard

Reward is the key in MAB, which drives the agent to select actions and learn the best strategy. Regarding jamming missions, a standard is needed to evaluate the jamming effect. To use TCP/IP as a communication protocol, the receiver should send ACK/NACK frames to a transmitter as a response; sometimes, the frames are not encrypted. If the jammer counts the number of ACK/NACK frames, the packet error rate (PER) can be

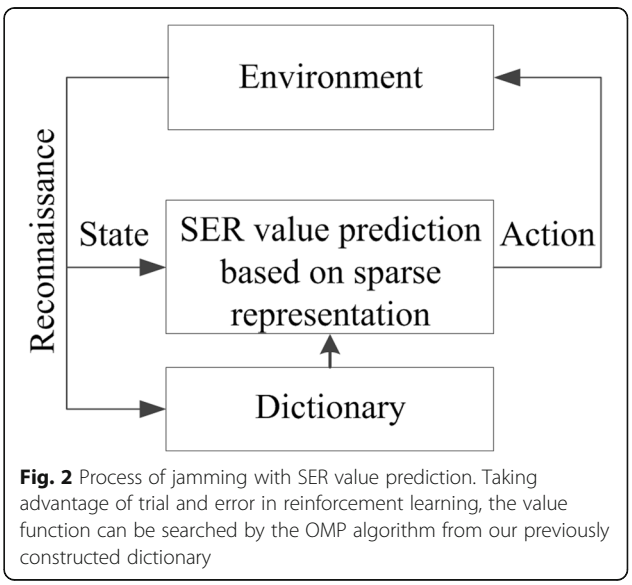

Fig. 2 Process of jamming with SER value prediction. Taking advantage of trial and error in reinforcement learning, the value function can be searched by the OMP algorithm from our previously constructed dictionary

easily calculated and used to estimate the SER with SER = 1 – (1 – PER)$^{1/H}$, where H is the number of bits in the frame check sequence. In reference [4, 12, 15–17], the SER is used to evaluate the jamming effects, which applies to this paper.

3.2 Dictionary construction

With prior knowledge such as modulation scheme and noise distribution, we can construct the dictionary according to Eq. (3), where we directly assign P_F a fixed value and assign ω, ϖ with different values to generate various atoms. The range of ω, ϖ is determined by P_T, P_F, α, β, γ, η, and N_0, and additional relationship details are provided as follows:

$$\omega = P_T + \phi - P_F = P_T + P_F(\alpha - N_0)/N_0 - P_F = \alpha P_T/N_0 - P_F \tag{5}$$

As P_T, N_0 has a positive value, the communication signals have maximum power $P_{T_{max}}$, and $0 \le \alpha \le 1$, $\omega \ge -P_T$; thus, $\omega \le P_{T_{max}}/N_0 - P_T$.

$$\varpi = \phi = N_0/(2\beta\gamma) \tag{6}$$

The parameters $0 \le \beta \le 1$ and $0 \le \gamma \le 1$; thus, we have $\varpi > N_0/2$.

3.3 Sample selection

In the OMP algorithm, the sampled sequence is used to search for the proper atoms. Thus, we should obtain some effective samples that would be helpful in searching for the excepted atoms. For any atom in the dictionary, it has a monotone increasing trend that ranges from 0 to some fixed value. For example, when the in-phase of a QPSK signal is successfully jammed, the maximum SER in the receiver is 0.5, which indicates that all atoms have a value of 0 at the initial part and a value of 0.5 at the end part. The atoms cannot be distinguished according to these two values. In view of the above reasons, the jammer should avoid 0 or 0.5 as samples; a smart choice is to evaluate the jamming environment and determine a proper power. After the first interaction, the jammer has to determine the second jamming power according to a feedback of the first jamming. In a world, the purpose of choosing jamming power is to obtain effective samples with fewer interactions.

3.4 An algorithm for jamming strategy using OMP and MAB

The proposed algorithm has three stages: the reconnaissance stage, preparing stage, and jamming stage. In the reconnaissance stage, the jammer recognizes the modulation of the communication signal [23] and the distribution of noise; this information is necessary for dictionary construction. For the second stage, the jammer has to determine the jamming power $P_{J_{initial}}$ for the first interaction. Another study is to construct the dictionary with given ω and ϖ values. In the most important stage of jamming, the jammer uses the same power to jam in-phase and quadrature phase, and then determines which phase should be jammed in the next action according to the feedback results. If the decision is in-phase, the jammer should use another proper jamming power to jam in-phase. After three jamming actions, the jammer already has two effective jamming results ξ_m and ξ_n, and should continue to jam with the in-phase. Equation (4) can be written as:

$$\begin{bmatrix} \xi_m \\ \xi_n \end{bmatrix}_{2\times 1} = \begin{bmatrix} d_{m1} & \cdots & d_{mm} & \cdots & d_{mn} & \cdots \\ d_{n1} & \cdots & d_{nm} & \cdots & d_{nn} & \cdots \end{bmatrix}_{2\times M} \cdot \begin{bmatrix} \vdots \\ x_m \\ \vdots \\ x_n \\ \vdots \end{bmatrix}_{M\times 1} \tag{7}$$

With the OMP algorithm, the jammer can identify proper atoms and coefficients; the schematic diagram of the proposed jamming algorithm is shown in Fig. 3.

Step 1. Reconnaissance stage: Analyze the modulation of the communication signal and the distribution of the channel noise and take a rough estimate of the power of a communication signal according to the jamming environment.

Step 2. Preparing stage: Determine the span of the parameters ω and ϖ, which would be used to construct the dictionary $\boldsymbol{D}$ with Eqs. (3), (5), and (6), and then decide the value of the power $P_{J_{initial}}$ in the first jamming according to the reconnaissance results.

Step 3. Jamming stage:

1. Jam the in-phase one time with $P_{J_{initial}}$ and obtain the feedback $\xi_m^{(1)}$ from the environment state.
2. Jam the quadrature phase one time with $P_{J_{initial}}$, and obtain the feedback $\xi_m^{(2)}$ from the environment state.
3. If $\xi_m^{(1)} > \xi_m^{(2)}$,

Jam the in-phase one time with a proper jamming power, which can be decided in terms of $\xi_m^{(1)}$ and the reconnaissance results; the feedback of this jamming action is $\xi_n^{(3)}$.

else

Jam the quadrature phase one time with a proper jamming power, which could be decided according to $\xi_m^{(2)}$ and the reconnaissance results; the feedback of this jamming action is $\xi_k^{(3)}$.

end if

4. If the in-phase is jammed by the third jamming action

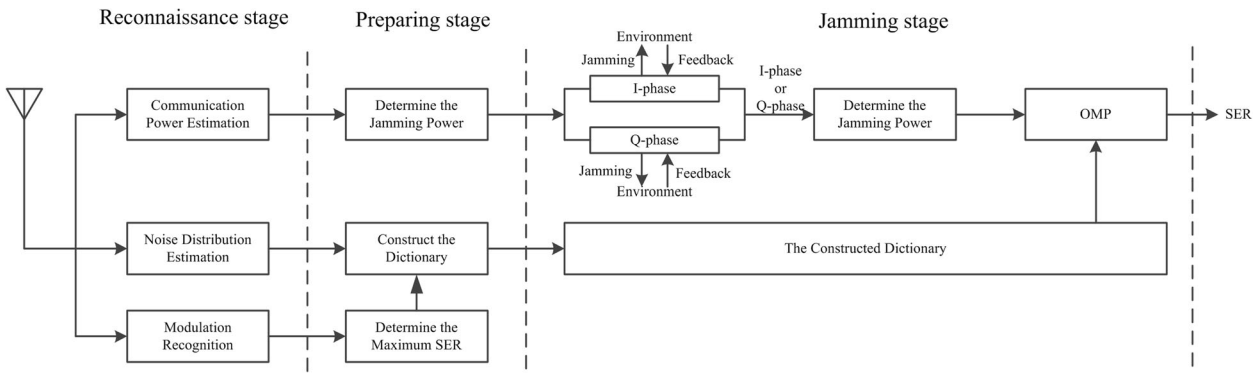

Fig. 3 Schematic diagram of the proposed jamming algorithm. Additional details about the process of the proposed jamming algorithm are shown in the schematic diagram. The dictionary is constructed with reconnaissance result, and the sampled data in OMP is the feedback from the environment

With the sample sequence $\{\xi_m^{(1)}, \xi_n^{(3)}\}$ and the constructed dictionary $\boldsymbol{D}$, the optimal SER curve can be calculated by the OMP algorithm.

else if the quadrature phase is jammed by the third jamming action

The optimal SER value can be obtained with the OMP algorithm but the sample sequence should be $\{\xi_m^{(2)}, \xi_k^{(3)}\}$.

end if

5. The follow jamming actions can be guided by the optimal SER curve.

An MQAM signal is equivalent to a pulse amplitude modulation (PAM) signal on two orthogonal carriers. As the two signal components are orthogonal in a phase that can be completely separated in the demodulator, the symbol error rates ξ_{I} and ξ_{Q} of the two signals can be calculated and jointly determine the symbol error rate $\xi = 1 - (1 - \xi_{\mathrm{I}})(1 - \xi_{\mathrm{Q}})$ of the MQAM signal. We know that ξ requires more jamming power than ξ_{I} (or ξ_{Q}) under the same SER. Therefore, the jammer can jam only the in-phase or quadrature phase in the jamming stage, which will reduce the jamming power while performing effective jamming.

3.5 Evaluation of the prediction results and advantages of the proposed algorithm

For any jamming missions, the core demand is to determine the optimal jamming strategy as soon as possible, which indicates that the jammer requires few interactions and better prediction performance. In this paper, we evaluate the convergence rate with the interaction times and apply the index of the mean square (MS) and the sum square error (SSE) [24] to measure the prediction performance. The calculations of MS and SSE are expressed as:

$$\mathrm{MS} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|, \mathrm{SSE} = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{8}$$

where y denotes the true SER curve and $\hat{y}$ denotes the prediction curve. As shown in Eq. (8), we know that the lower is the index value, the better is the prediction performance.

The advantages of this algorithm are listed as follows:

(1) The proposed algorithm only needs to know the noise distribution as a priori knowledge and does not need to know the accurate power of the communication signals and the jamming signals in the receiver.
(2) The proposed algorithm does not have to divide the jamming parameters that can avoid the curve of the dimension in [4].
(3) The proposed algorithm can fully utilize a priori information, such as the communication scheme and the noise distribution, to achieve a faster convergence rate.

4 Methods

The effectiveness of the proposed approach has been validated by computer simulation experiments. The simulations were conducted in the MATLAB R2014a environment on a personal computer with an Intel® Core™ i5 1.7 GHz processor and 4 GB RAM.

Assume that the modulation of the signal used by a communicator is QPSK and the power of the transmitted signal is $P_{\mathrm{T}} = 100$ W. For the jammer, the span of the jamming power is $P_{\mathrm{J}} \in [0,400]$W, and the expected SER is $\xi_{\mathrm{E}} = 0.38$. In a jamming mission, both the communication signal and the jamming signal will be decayed by a transmitter channel; the factor of decay is $\alpha = 0.82$ and $\beta = 0.68$, respectively. The jamming signal will also be decayed by

adaptive zero attenuation, which is often used by a communicator to counter the jamming actions. Thus, the factor of restraint is $\gamma = [\gamma^{(1)}, \gamma^{(2)}] = [0.4, 0.3]$. The noise that we assume in this paper is the AWGN, which has zero mean and one variance. In the preparing phase, we should construct the dictionary in advance, and the span of the parameters ω and ϖ is $\omega = [-100, 400]$, $\varpi = [0.01, 5]$. The algorithm of sparse representation in this paper is OMP, and the algorithm process will be terminated by an assigned value.

5 Results and discussion

5.1 Performance of the proposed algorithm

As discussed in Section 3.3, the samples are the key when using the OMP algorithm to obtain the optimal SER value. In the given jamming mission, the jammer jams the in-phase and quadrature phase with 180 W and obtains the feedback 0.359 and 0.141. As mentioned in Section 3.4, at least two samples are needed in the OMP algorithm, and the jammer should choose 140 W and 160 W to jam the in-phase and quadrature phase again. The feedback of the second jamming action in each phase is 0.161 and 0.074. With these samples, the jammer can predict the optimal SER curve by the OMP algorithm. Figure 4 shows the predicted results in the in-phase and quadrature phase, where the noise belongs to the AWGN distribution.

In Fig. 4, the predicted SER curve and the real SER curve are almost completely overlapped. When we use SSE to evaluate the difference, the in-phase has SSE = 3.65×10^{-5} and the quadrature phase has SSE = 8.45×10^{-6}. When we compare the SER curve between in-phase and the quadrature phase, we discover that the in-phase of the communication signal is more fragile and determine that jamming in-phase than quadrature phase with given jamming power. To fulfill the requirement of expecting SER $\xi_E = 0.38$, the jamming power should be 188 W and the target should be in-phase.

When the noise has a Rayleigh distribution and other parameters remain unchanged, the in-phase and quadrature phase are jammed two times, and then, the optimal SER curves can be predicted with the OMP algorithm again. Figure 5 shows the predicted SER curve and the real SER curve.

As depicted in Fig. 5, the predicted SER curve and the real SER curve are almost completely overlapped, the in-phase curve has SSE = 2.85×10^{-4}, and the quadrature phase curve has SSE = 1.22×10^{-4}. Figures 4 and 5 have different SER curves for different noise distribution reasons. With the proposed algorithm, the jammer can predict an accurate SER curve.

5.2 Effect of atom numbers on predicted results

The number of atoms in the constructed dictionary depends on ω and ϖ values, and the number of ω and ϖ depends on the division manner that we employ. When the span of ω and ϖ are given and statistically analyzed, we have 10 types of division manners: division(ω) = {50 : 50 : 500} and division(ϖ) = {10 : 10 : 100}. Using these division manners, the number of atoms are {50 : 50 : 500} × {10 : 10 : 100}. Figure 6 shows the effect of atom numbers on the predicted results, where MS and SSE are employed as evaluating indicators.

When the dictionary has fewer atoms, the values of MS and SSE are large and fluctuated, which indicates that the predicted SER curve has the same amount of

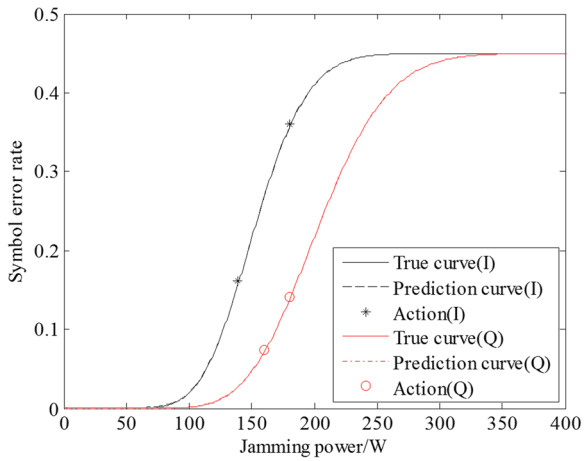

Fig. 4 Environmental noise belongs to the AWGN distribution. The predicted results in the in-phase and quadrature phase almost overlapped with the real SER curve, where the noise belongs to the AWGN distribution, and the dictionary was constructed with the AWGN assumption

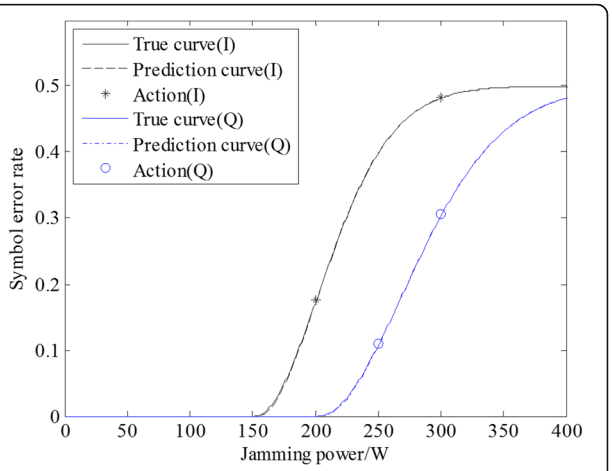

Fig. 5 Environmental noise belongs to the Rayleigh distribution. The predicted results in the in-phase and quadrature phase almost overlapped with the real SER curve, where the noise belongs to Rayleigh distribution, and the dictionary was constructed with the Rayleigh assumption

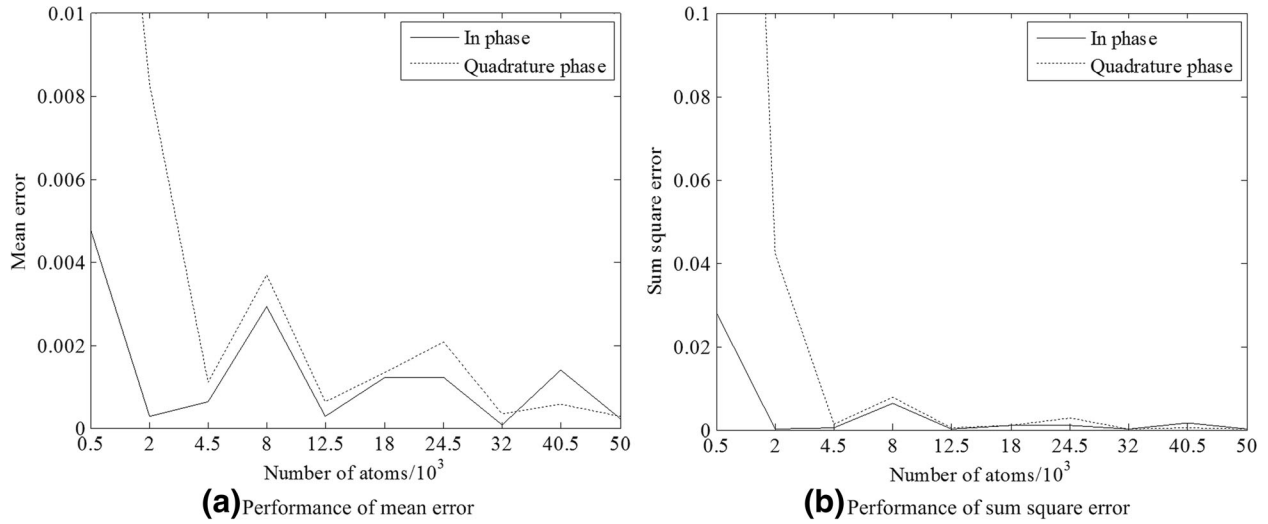

Fig. 6 Effect of atom numbers to ME and SSE. The MS and SSE are employed as an evaluation indicator, whose performance has correlation with the atom numbers in the constructed dictionary. **a** Effect of atom numbers to ME. **b** Effect of atom numbers to SSE

error as the real SER curve. The predicted SER curve with error cannot be used as a guide to choose the correct actions. However, when the atom numbers exceed 20,000, the values of MS and SSE are small, and the predicted SER curve would be a better guide.

5.3 Jamming with wrong dictionary

We consider two types of noise distributions: AWGN and Rayleigh. We first assume that the real noise that exists in a communication channel has an AWGN distribution, but we have a wrong reconnaissance result and construct the dictionary with the Rayleigh distribution assumption. Figure 7a shows the predicted results compared with a real SER curve, where the jammer requires three interactions both in the in-phase and the quadrature phase. Figure 7b shows an opposite situation, in which the noise has a Rayleigh distribution. The dictionary is constructed with an AWGN distribution assumption, and three interactions both in the in-phase and the quadrature phase remain unchanged.

In Fig. 7a, the predicted SER curve is similar to the real SER curve, with the evaluation index mentioned in Section 3.5. The SSE value of the in-phase is 0.0132, and the SSE value of the quadrature phase is 0.0251. Although the results are higher than the values in Fig. 6, the jammer can make jamming decisions with the

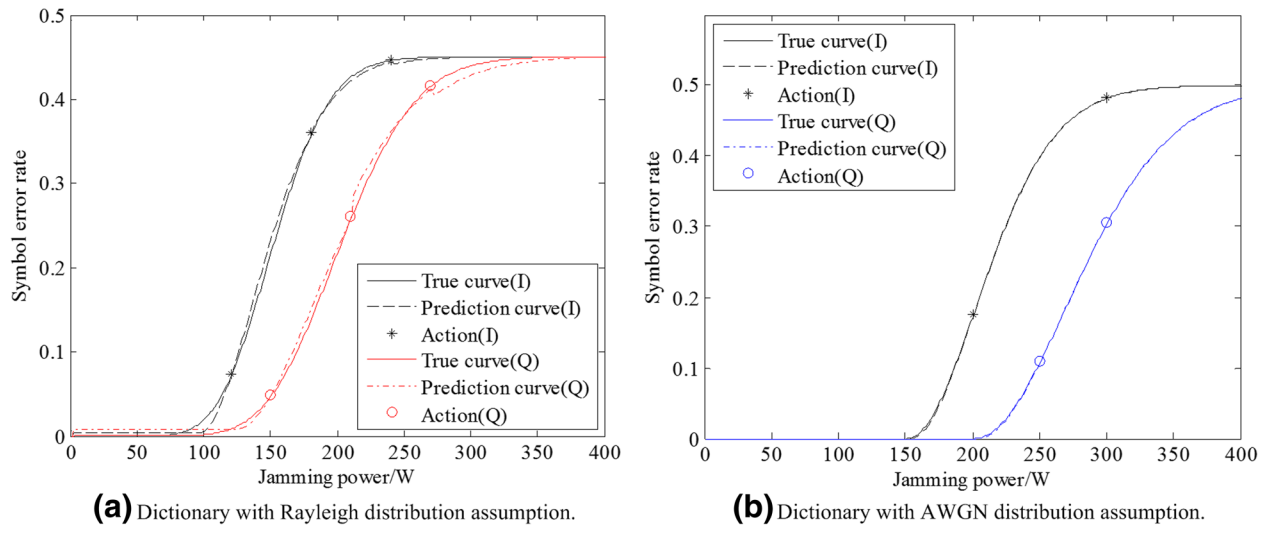

Fig. 7 Predicted results with the wrong dictionary. The jammer searched the value function with the wrong dictionary. **a** Dictionary with the Rayleigh distribution assumption but the real noise has an AWGN distribution. **b** Dictionary with the AWGN distribution assumption but the real noise has a Rayleigh distribution

predicted SER curve. As shown in Fig. 7b, the SSE value of the in-phase is 0.0236 and the quadrature phase is 0.0337; the effect of the prediction remains acceptable.

If the proposed algorithm is not sensitive to the noise distribution, the jammer still can consider the noise distribution but the premise is that the SER curves are similar under different noise distributions; the degree of similarity that can be accepted depends on the jammer. In Fig. 7, the jammer incorrectly estimates the Gaussian distribution and the Rayleigh distribution. However, the above two noise distributions have similar SER curves at the same jamming power, so the prediction result remains acceptable.

5.4 Jamming algorithms comparison

To measure the performance of the proposed algorithm, we make a comparison with dual reinforcement learning based jamming decision (DRLJD) [16], MAB [4], greedy algorithm [17], and positive reinforcement learning-orthogonal decompose (PRLOD) [15]. In these algorithms, jamming actions can be obtained by discrete division and be regarded as independent choices. The jammer needs to perform trial and error on these actions and takes the action with the highest reward value as the best jamming strategy. In contrast to this, the proposed algorithm explores the relationship between actions and takes advantage of it that will greatly faster the convergence rate. The experimental conditions are the same as the conditions in Section 5.1, and the jamming times are limited to 500 interactions.

In the DRLJD algorithm, the number of interactions is relevant to the length of the initial phase, and only the initial phase is set to more than 200 interactions. The DRLJD algorithm can learn the optimal or suboptimal strategy with a possibility of 1, as shown in Fig. 8a. After 204 interactions, the SER curve converges to 0.391, which fulfill the expected jamming requirement. The learned jamming strategy is {200 W, 190 W, 1}, which indicates that the total jamming power of 200 W is needed, the in-phase has a jamming power of 190 W, and the quadrature phase has a jamming power of 10 W.

The MAB algorithm can learn the best action until all actions have been tried. However, sometimes, too many actions exist, and the jammer has to choose the best actions for which the feedback is already known. In Fig. 8b, we assume that 300 actions would be tried one by one and the jammer chooses the best action as the optimal jamming strategy. In this experiment, the optimal jamming strategy is {240 W, 240 W, 0.95}, which indicates that the jamming signal has a power of 240 W, the modulation scheme is BPSK, and the pulse ratio is 0.95. However, even the learned jamming strategy fulfills the expected requirement, it still has a large jamming power.

A greedy algorithm has a special parameter divide manner that is decided by the jammer. In an unfamiliar environment, the jammer does not know the optimal discretization factor (the optimal jamming strategy is among the possible strategies that can be chosen by the greedy algorithm). In Fig. 8c and d, we set the discretization factors to 3 and 7; thus, we have 27 actions and 343 actions to try respectively. Although the discretization factors differ, the jammer learns the same jamming strategy, in which the optimal jamming power is 200 W and the modulation scheme is BPSK, which indicates that the discretization factors that we established are too small.

As previously discussed, the proposed algorithm needs to jam three times to obtain samples; after three interactions, the algorithm converges to the optimal jamming strategy. In Fig. 8e, the feedback of the jamming action is 0.387, which fulfills the previously mentioned requirements. The learned jamming strategy is {188 W, 188 W, 1}, which indicates that the minimum jamming power should be 188 W and the modulation scheme is BPSK. The jamming power and the number of interactions of the proposed algorithm are less than that of other algorithms.

The PRLOD algorithm has two phases that are randomly performed: selection phase and positive reinforcement learning phase. The length of both phases should be initially set. In Fig. 8f, the length of the random choose phase is 50 interactions and the length of the latter phase is 200. Thus, the jammer requires 250 interactions to converge to 0.436. Besides that, the learned jamming power is 220 W, the in-phase power should be 198 W, and the left power belongs to the quadrature phase.

In contrast to the previous jamming algorithms based on RL and discretization action, the simulation results demonstrate that the proposed algorithm considers the correlation among actions and directly predicts the value function of actions, as mentioned in Section 5.1, which will substantially alleviate the curse of dimensionality. With the predicted value function, the jammer can purposely choose to jam with the action instead of randomly selecting the action, which will reduce the number of interactions and ensure that the proposed algorithm will converge much faster than other RL algorithms.

6 Conclusions

In this paper, we proposed an algorithm for a jamming strategy using OMP and MAB to predict the value function of actions. The proposed algorithm can learn the optimal jamming strategy at the physical

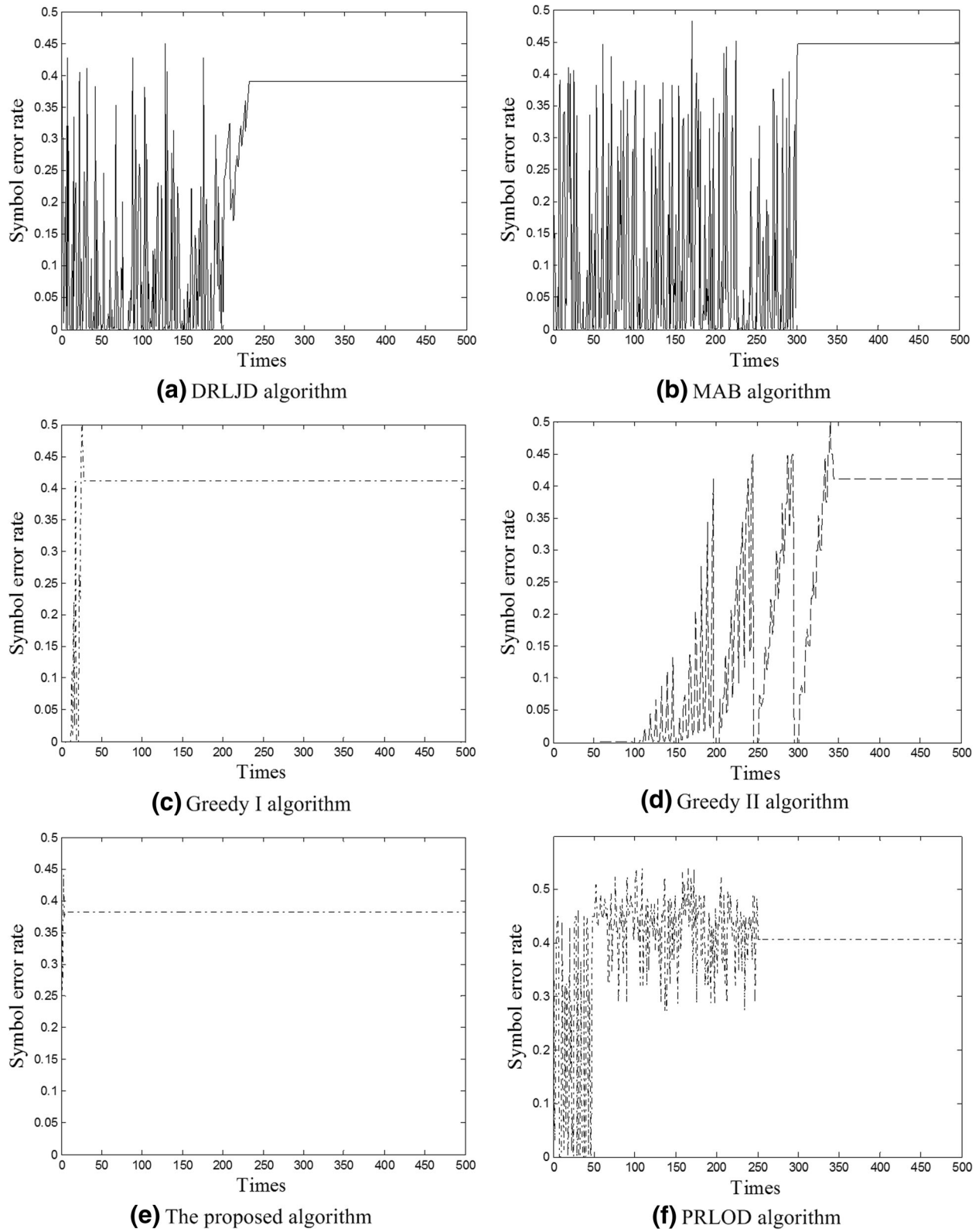

Fig. 8 Jamming performance among different algorithms. We address four additional algorithms to compare with the algorithm proposed in this paper, and the convergence curve shows the performance of the above five algorithms. **a** DRLJD algorithm. **b** MAB algorithm. **c**, **d** Greedy algorithm with different division manners. **e** Value prediction based on the algorithm proposed in this paper. **f** PRLOD algorithm

layer in an electronic warfare-type scenario with three interactions. Prior knowledge such as communication signal schemes and noise distribution is needed in the proposed algorithm, which can be obtained by reconnaissance. The effect of atom numbers in the constructed dictionary is also discussed. The rate of learning is considerably faster compared with commonly employed RL algorithms. Moreover, the proposed algorithm can learn a substantially smaller jamming power, which fulfills the jamming expectation and power efficiency.

Abbreviations

ACK/NACK: Acknowledge/not acknowledge; AWGN: Additive white Gaussian noise; BPSK: Binary phase shift keying; DRLJD: Dual reinforcement learning based jamming decision; IQ: In-phase and quadrature; MAB: Multi-armed bandit; MP: Matching pursuit; MS: Mean square; OMP: Orthogonal matching pursuit; PRLOD: Positive reinforcement learning-orthogonal decompose; QPSK: Quadrature phase shift keying; RL: Reinforcement learning; ROMP: Regularized OMP; SER: Symbol error rate; SSE: Sum square error; StOMP: Stagewise OMP

Acknowledgements

First and foremost, I appreciate my college which provided a comfortable learning atmosphere. Second, I express my gratitude to my supervisor, Mr. Yang, who has encouraged me through all stages of the writing of this paper. I would like to thank the editor and anonymous reviewers for their helpful comments in improving the quality of this paper.

Funding

This study is supported by the National Natural Science Foundation of China (NSFC) (grant nos. 11375263).

Availability of data and materials

All data generated or analyzed during this study are included in this paper.

Authors' contributions

SZ and JY conceived and designed the experiments and analyzed the data. SZ performed the experiments and wrote the paper. HL contributed the reagents/materials/analysis tools. All authors read and approved the final manuscript.

Authors information

Shaoshuai Zhuansun was born in Anhui Province, China in 1990. He received a B.S. degree in communication engineering from Lanzhou Jiao Tong University in Lanzhou, China, in 2012 and an M.S. degree in communication & information systems from the Electronic Engineering Institute in Hefei, China in 2015. He is currently pursuing a Ph.D. degree with the Department of Communications at the National University of Defense Technology in Hefei, China. His research interests include cognitive jamming and reinforcement learning. E-mail: zhuansunss@sina.com

Jun-an Yang was born in Anhui Province, China in 1965. He received a B.S. degree in radio technology from Southeast University in Nanjing, China in 1986 and an M.S. degree in communication & information systems from Electronic Engineering Institute in Hefei, China, in 1991. He received a Ph.D. degree in signal & information processing from the University of Science and Technology of China (USTC) in Hefei, China, in 2003. He is currently a professor in the Department of Communications at the National University of Defense Technology in Hefei, China. His research interests include signal processing and intelligence computing.
E-mail: yangjunan@ustc.edu

Hui Liu was born in Anhui Province, China, 1983. He received a B.S. degree in communications engineering from Wuhan University in Wuhan, China, in 2005. He received an M.S. degree and Ph.D. degree in communication & information systems from the Electronic Engineering Institute in Hefei, China in 2008 and 2011, respectively. He is currently a lecturer in the Department of Communications at the National University of Defense Technology in Hefei, China. His interests include intelligent information processing and cognitive communication.
E-mail: liuhui983eei@163.com

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Received: 14 June 2018 Accepted: 21 March 2019
Published online: 02 April 2019

References

1. X. Liu, M. Jia, X.Y. Zhang, W. Lu, A novel multi-channel internet of things based on dynamic spectrum sharing in 5G communication. IEEE Internet Things J. **99**, 1–1 (2018). https://doi.org/10.1109/JIOT.2018.2847731
2. X. Liu, M. Jia, Z. Na, W. Lu, F. Li, Multi-modal cooperative spectrum sensing based on Dempster-Shafer fusion in 5G-based cognitive radio. IEEE Access **6**, 199–208 (2018)
3. X. Liu, F. Li, Z. Na, Optimal resource allocation in simultaneous cooperative spectrum sensing and energy harvesting for multichannel cognitive radio. IEEE Access **5**, 3801–3812 (2017)
4. S. Amuru, C. Tekin, M.V.D. Schaar, T.C. Clancy, Jamming bandits-a novel learning method for optimal jamming. IEEE Trans. Wirel. Commun. **15**(4), 2792–2808 (2016)
5. Y. Wu, B. Wang, K.J.R. Liu, T.C. Clancy, Anti-jamming games in multi-channel cognitive radio networks. IEEE Journal on Selected Areas in Communications **30**(1), 4–15 (2012)
6. R.S. Sutton, A.G. Barto, Reinforcement learning: an introduction. IEEE Trans. Neural Netw. **9**(5), 1054–1054 (1998)
7. P. Auer, N.C. Bianchi, P. Fischer, Finite-time analysis of the multi-armed bandit problem. Mach. Learn. **47**(2–3), 235–256 (2002)
8. H. Li, Y. Qian, Effects of IQ imbalance for simultaneous transmit and receive based cognitive anti-jamming receiver. AEU-International Journal of Electronics and communications **72**, 26–32 (2017)
9. Y. Niu, F. Yao, M. Wang, Anti-chirp-jamming communication based on the cognitive cycle. AEU-International Journal of Electronics and communications **66**(7), 547–560 (2012)
10. C. Zhou, Z.Y. Tang, F.L. Yu, L. Y, Anti intermittent sampling repeater jammer method based on intrapulse orthogonally. Systems Engineering and Electronics **39**(2), 269–276 (2017)
11. N. Zamir, B. Ali, M.F.U. Butt, S.X. Ng, in *Improving Secrecy Rate via Cooperative Jamming Based on Nash Equilibrium*. 24th European Signal Processing Conference (IEEE, Budapest, 2016), pp. 235–239
12. S. Amuru, R.M. Buehrer, *2014 IEEE Military Communications Conference, Optimal Jamming Using Delayed Learning* (IEEE, Baltimore, 2014), pp. 1528–1533
13. S. Amuru, R.M. Buehrer, M.V.D. Schaar, Blind network interdiction strategies-a learning approach. IEEE Trans. Cogn. Commun. Netw. **1**(4), 1–7 (2016)
14. J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans. Inf. Theory **53**(12), 4655–4666 (2007)
15. S.S. ZhuanSun, J.A. Yang, H. Liu, K.J. Huang, Jamming strategy learning based on positive reinforcement learning and orthogonal decomposition. Syst. Eng. Electron. **40**(3), 518–525 (2018)
16. S.S. ZhuanSun, J.A. Yang, H. Liu, K.J. Huang, An algorithm for jamming decision using dual reinforcement learning. Journal of Xi'an Jiao Tong University **52**(2), 63–69 (2018)
17. S.S. ZhuanSun, J.A. Yang, H. Liu, K.J. Huang, in *A Novel Jamming Strategy-Greedy Bandit*. 9th IEEE International Conference on Communication Software and Networks (IEEE, Guangzhou, 2017), pp. 1142–1147
18. Y.Q. Li, A. Cichocki, S.I. Amari, Analysis of sparse representation and blind source separation. Neural Comput. **16**(6), 1193–1234 (2004)
19. S. Mallat, Z. Zhang, Matching pursuit with time-frequency dictionary. IEEE Trans. Signal Process. **41**(12), 3397–3415 (1993)
20. Y.C. Pati, R. Rezaiifar, P.S. Krishnaprasad, *In Proceedings of 27th Asilomar Conference on Signals, Systems and Computers, Orthogonal Matching Pursuit:*

Recursive Function Approximation with Applications to Wavelet Decomposition (IEEE, California, 1993), pp. 40–44
21. B. Sun, K.H. Zhao, Improved algorithm of the regularized OMP algorithm based on energy sorting. Electron. Meas. Technol. **39**(5), 154–158 (2016)
22. C.W. Tang, X.F. Wang, Y.G. Du, A sparsity adaptive stagewise orthogonal matching pursuit algorithm. J. Central South University (Science and Technology), **47**(3), 784–791 (2016)
23. H. Wang, L.L. Guo, Y. Lin, H. Wang, L.L. Guo, Modulation recognition of digital multimedia signal based on data feature selection. Int. J. Mob. Comput. Multimed. Commun. **8**(3), 90–111 (2017)
24. F. Salour, S. Erlingsson, Permanent deformation characteristics of silty sand subgrades from multistage RLT tests. Int. J. Pavement. Eng. **18**(3), 236–246 (2017)