Distributed diffusion fusion cooperative spectrum sensing based on reinforcement learning

Lijun Jiang¹, Xiaoge Zhang², b*
¹School of information science and technology, Nantong University, Jiangsu, 226000, China
²School of information science and technology, Nantong University, Jiangsu, 226000, China
*e-mail: 346619553@qq.com, bemail: zxg@ntu.edu.cn

Abstract. The openness and complexity of wireless channels make collaborative spectrum sensing vulnerable to malicious users[1]. Therefore, it is very important to identify the attributes of malicious users before collaborative spectrum awareness networks make data fusion decisions. In this paper, a method combining reinforcement learning and cognitive user credit model is proposed, in which the maximum and minimum eigenvalues of signals are used as the initial information for exchange, and the whole sensing network tends to diffuse and fuse nodes with high credit. Finally, the convergence value of the whole network is compared with the decision threshold to complete collaborative spectrum sensing. By combining with consensus fusion network and traditional collaborative sensing algorithm, the proposed method can effectively improve the convergence speed of fusion network and shorten the sensing time on the premise of effectively identifying malicious users[2], so as to improve the spectrum sensing performance and make the collaborative sensing network more adaptive and stable.

1. Introduction
Collaborative spectrum sensing technology can effectively improve the spectrum detection rate[3], but malicious users tamper with the sensing user data, resulting in a sharp decline in the performance of the whole sensing network. Before data fusion, how to effectively identify malicious users, this paper puts forward a model that combining of reinforcement learning and SU credit. Take the MME of the signal as initial information exchange, after selecting the optimal user through reinforcement learning, according to the credit value model, update the corresponding credit value, when it below decision threshold, judged to be a malicious user, make its exit from the current collaborative sensing network. After reinforcement study selected the honest users[4], in the diffusion of information fusion in the network interaction, cognitive user needs and the adjacent cognitive users exchange of initial information, by constructing adaptive matrix as weighted factor matrix and the fusion, update the status value of cognitive node fusion, make the entire network cognitive user information, check the entire network convergence value of the final. at this time, compared with the judgment threshold, the final judgment of the main user is made, and the detection of whether the current detection frequency band is occupied by the main user is completed.
2. System Model

Figure 1: sensing network

The cooperative sensing network consists of i SU, which can be seen as an undirected graph model \( G = (V, \xi) \). The set of cognitive nodes is defined as \( V = \{1, 2, \cdots, i\} \).

\( \xi = \{(i, j) | i, j \in V, i \neq j\} \) is defined as the edge set of the disordered node pairs of cognitive nodes in the range of single hop communication. The neighbor nodes cognitive user \( i \) is consist of nodes that can communicate and interact with it, and the set of neighbor nodes is \( V_{ci} = \{j | (i, j) \in \xi \} \subseteq V \). \( \lambda_{\text{max}}, \lambda_{\text{min}} \) are the maximum and minimum eigenvalues of the received signal covariance matrix, and Detection statistic:

\[
T_{\text{MML}} = \frac{\lambda_{\text{max}}}{\lambda_{\text{min}}} \tag{1}
\]

For the convenience of analysis, the cognitive radio network in this paper has the following characteristics until the spectrum perception of the cognitive user reaches the convergence value of the whole network:

1) Each cognitive user is uniformly distributed, and local spectrum perception can be carried out independently, and the perception results can be interactively fused with neighbor users through wireless connection;

2) The primary user has a strong transmitting power, and the whole sensing network is within the transmitting range of the primary user. The distance between cognitive users is small, and the transmission loss of wireless connection is not considered.

3) In the sensing cycle, the position of primary user and cognitive user remains unchanged, that is, the topology of the collaborative sensing network remains unchanged.

3. Materials and Methods

There are 5 steps in this algorithm, and the specific process is as follows:
Step 1: Initialization

Local spectrum sensing uses the maximum and minimum eigenvalue detection algorithm. The initial detection value is the maximum and minimum eigenvalue of the local perception, which is calculated by equation (1) and is a constant. The initial detection value and initial credit value of the $i_{th}$ SU can be expressed as:

$$T_i(0) = b_i$$
$$R_i(0) = 2$$

Step 2: Use reinforcement learning to determine cognitive user attributes

Before collaborative spectrum sensing diffusion and fusion, reinforcement learning methods need to be used to judge the attributes of cognitive users participating in the fusion, and to exclude malicious users. The state $S_i$ represents the set of all neighbor users of the current SU. Action $a$ means that the agent selects or excludes a neighbor user based on the credit value in the state $S_i$, and the reward value $r_{i+1}(s_{i+1})$ means the feedback after performing the action $a$, reflecting the good or bad status of the credit value of its neighbor users. For learning through Q-learning, the specific update process $Q_i(s_i, a_i)$ is as follows:

1. The initialization Q matrix is a random matrix of $i \times i$, $Q_i(s_i, a_i)$ initialized to 0, set the learning rate $\alpha$ and discount factor $\gamma$ respectively, and generate the Look up table.
2. Each SU is regarded as an independent agent, and finds the optimal strategy through information interaction with neighbor nodes.
   ① $s_i \leftarrow s_{i+1}$: each agent uses the Look up table to update the status $S_i$ by querying it.
   ② In the current state, according to the $Q$ matrix, the agent selects the action with the maximum Q value $a = \max_{a \in A} Q_i(s_{i+1}, a)$ for execution, and the probability of selecting the action $a$ is:

$$P(a|s_i) = \frac{\exp(Q_i(s_{i+1}, a)/\Gamma(t))}{\sum_{a \in A} \exp(Q_i(s_{i+1}, a)/\Gamma(t))}$$

(2)

Among them, $\Gamma(t)$ is the temperature parameter in the simulated annealing process.
3. After performing the action $a$, the reward is $r_{i+1}(s_{i+1})$.
4. Substitute it into The update equation and update the current $Q_i(s_i, a_i)$.
5. Repeat the process ② until $Q_i(s_i, a_i)$ stable maximum is formed, and in this stable state, the
The $i_{th}$ agent interacts with the neighboring users $V_{RL_{\rightarrow i}}$ to fuse information, $V_{RL_{\rightarrow i}} \in V_{ci}$ denotes the honest neighboring users after reinforcement learning, and $j \in V_{ci}$ denotes all users adjacent to the $i_{th}$ SU.

Assuming that the reward of the $i_{th}$ SU at time of $t$ is represented as $r_{ij}(s_{i,j})$, the interaction perception value obtained by the cooperation between the $i_{th}$ SU and neighbor users is defined $D_{i,j}(t) = x_j(t)$. In the fusion center, the judgment result of $i_{th}$ SU is:

$$A_{ij}(t) = \begin{cases} 1, & D_{i,j}(t) \geq \eta \\ -1, & D_{i,j}(t) < \eta \end{cases}$$

(3)

$$\eta = \frac{(\sqrt{N}+\sqrt{L})^2}{(\sqrt{N}-\sqrt{L})^2} \left(1 + \frac{(\sqrt{N}+\sqrt{L})^2}{(NL)^{0.5}} F^{-1}_1(1-P_{in}) \right)$$

(4)

In the formula, $\eta$ is the local perception decision threshold of the initial state.

There are two cases of judgment results. For different neighbor nodes, the larger the number of judgment results is the same, the $i_{th}$ more reliable. at the time of $t+1$, The reward value of the $i_{th}$ SU can be defined as:

$$r_{i+1}(s_{i+1}) = r_{ij}(s_{i,j}) + \sum_{j \in D_{i}} A_{ij}(t)$$

(5)

**Step 3:** credit value update

Before information exchange, each cognitive user is given an initial credit value. In this algorithm, the initial credit value is set to 2, and the calculation is updated according to equation (6) for each iteration.

$$R_i(k) = R_i(k-1) + 0.2 \ast (-1)^{b(k)+b_{ij}(k)}$$

(6)

Among them, $R_i(k)$ represents the credit value of the cognitive user $i$ in the $k_{th}$ iteration, $b(k)$ represents the judgment result of the fusion center, $b_{ij}(k)$ represents the local judgment result of the cognitive user $i$, which respectively indicate the following:

$$b_i(k) = \begin{cases} 1, & T_i(k) \geq \eta \\ 0, & T_i(k) < \eta \end{cases}$$

(7)

$$b(k) = \begin{cases} 1, & B(k) \geq \eta \\ 0, & B(k) < \eta \end{cases}$$

(8)

$$B(k) = \sum_{j \in D(k)} y_j(k) \ast T_j(k)$$

(9)

Among them, $D(k)$ it represents the set of neighbor users with high credit value. In this algorithm, the credit value greater than 6 is selected as the standard, namely

$$D(k) = \{ j \mid R_j(k) \geq 6, j \in V_{ci} \}$$

(10)

$y_j(k)$ is the credit value weight, and its expression is as follows:

$$y_j(k) = \frac{y'_j(k)}{\sum_j y'_j(k)}$$

(11)
\[ y_j(k) = \frac{R_j(k-1)}{\max(R_j(k-1))} \]  

\( (1) \)

**Step 4:** Diffusion fusion

When the cognitive user selects a user with a high credit value among the neighboring users, the iterative update formula for the diffusion fusion of the cognitive user \( i \) after \( k \)th iteration is:

\[ \hat{\psi}(k+1) = \hat{T}_i(k) + \mu \sum_{j \in \mathcal{N}_i} a_{ij} \left[ T_j(k) - \hat{T}_i(k) \right] \]

\( (13) \)

\[ \hat{T}_i(k+1) = \sum_{j \in \mathcal{N}_i} c_{ij} \hat{\psi}_j(k+1) \]

\( (14) \)

**Step 5:** When the \( \hat{T}_i(k) \) state estimation value of the cognitive user \( i \) reaches the network-wide convergence value \( T_i^* \), compared with the threshold of the initial setting state, the judgment of the existence of the primary user is as follows:

\[ D \sim \begin{cases} T_i^* < \eta, H_0 \\ T_i^* \geq \eta, H_1 \end{cases} \]

\( (15) \)

In the formula, \( H_0 \) means that the primary user does not exist, and \( H_1 \) that the primary user exists. At this point, through the joint cooperation of the entire collaboration network, the final spectrum sensing is completed.

4. Results & Discussion

In the simulation experiment, the PU signal is a QPSK signal in an additive white Gaussian noise channel. There are 10 cognitive users in the collaborative awareness network, and the corresponding network topology is shown in Figure 2. After introducing cognitive users with variable attributes, the network topology is shown in Figure 3. It is assumed that the channels perceived by each cognitive user are independent and uniformly distributed Gaussian channels, the transmitting power of the main user is 70dB, the noise power is -80dB, the relative distance of cognitive radio is 5 KM, the coverage radius of cognitive radio network is 1 KM, and the communication range between each cognitive user is 300 m. Sampling rate \( f_s = 50 \text{kHz} \), sensing cycle \( \tau = 0.05 \text{ms} \), maximum number of iterations \( t_{\text{max}} = 50 \), the learning rate of reinforcement learning \( \alpha = 0.2 \), the discount factor \( \gamma = 0.8 \), initial temperature of simulated annealing \( T_{\text{init}} = 1 \times 10^{30} \), it decreases exponentially to the final temperature, by the regular of decreasing the exponent by one, decrease to final temperature \( T_{\text{init}} = 0.1 \), Monte-Carlo simulation is performed for 1000 times to verify the algorithm proposed in this paper.
4.1 Network convergence analysis

According to Fig. 4 and Fig. 5, it can be seen that in the collaborative perception network, the network convergence performance when adopting diffusion fusion strategy and consistent fusion strategy respectively. Nodes 1, 2, and 3 are malicious users, while the remaining nodes are honest users. Compared with honest users, malicious users’ MME detection values are distributed irregularly. It can be seen from Fig. 4 that the diffusion strategy can make the network converge when the iteration is about 4 times. When the consistent fusion strategy is adopted, about 30 iterations are needed when the network reaches convergence, as shown in Fig. 5. In the process of data fusion, due to the reinforcement learning method, after several iterations, the malicious users are identified and made not participate in the network data fusion, and the MME detection value of the honest users in the whole network finally converges to around 6. Therefore, it shows that the proposed method can effectively improve the convergence speed of the converged network on the premise of effectively identifying malicious users.
4.2 Comparison of credit value and detection performance

Figure 6: Credibility values for different iterations

Figure 7: Credibility changes when perceived user attributes change

Figure 8: ROC curves of detection performance of four different methods
In the credit value model, the optimal neighbor users selected by reinforcement learning are rewarded according to the credit value update formula, which makes the credit value of these honest users increase. According to Figure 6, as the number of network iterations increases, the credibility of honest users increases, while the credibility of malicious users fluctuates around the initial value. The higher the rising slope of the credit value, the faster the credit value of the cognitive user increases. According to Figure 2, there are as many as 4 honest users adjacent to honest user SU5. Through information interaction, the perception performance of this cognitive user is the best in the whole collaborative perception network. Therefore, the credit value of this cognitive user also increases fastest and has the highest slope. In order to further verify the accuracy of selecting optimal neighbor users for reinforcement learning, cognitive users with variable attributes were introduced into the cooperative network. According to Figure 7, when the number of iterations is 20, the malicious user SU3 changes into the honest user, and the correspond in credit value increases with the increase of the number of iterations. Honest user SU6 turns into malicious user, and its corresponding credit value decreases with the increase of iteration. According to Fig. 8, the detection rate of the proposed method is better than that of the traditional collaborative sensing algorithm at different false alarm rates. Therefore, it is indicated that the method in this paper can adjust its parameters according to the changes of environment and through continuous learning, so as to make the whole collaborative sensing network more intelligent, improve the spectrum sensing performance, and make the collaborative sensing network more adaptive and stable.

5. Conclusion
This paper proposes a method combining reinforcement learning, cognitive user credit model and diffusion fusion strategy is proposed. The optimal neighbor is selected through reinforcement learning, and the credit value of cognitive user is updated through the credit value model, which can effectively resist the interference of malicious users on data fusion. The selected honest users with high credit value are used for data diffusion fusion, and the convergence speed is faster than that of the consistent fusion strategy. The simulation results show that the proposed method can effectively improve the convergence speed of the fusion network, shorten the sensing time, improve the spectrum sensing performance, and make the collaborative sensing network more adaptive and stable on the premise of effectively identifying malicious users.

Acknowledgments
The project was supported by the National Natural Science Foundation of China (61871241) and the Open Project of the Joint Research Center for Intelligent Information Technology of Nantong University (KFKT2017A05).

References
[1] Zan Li, Wen Wu, Xiangli Liu, et al. Improved cooperative spectrum sensing model based on machine learning for cognitive radio networks. 2018, 12(19):2485-2492.
[2] AINOMAE A, TRUMPT B, BENGTSSON M. Distributed diffusion LMS based energy detection [C]// Proceedings of 2014 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops. St. Petersburg IEEE, 2014: 176-183.
[3] Y. Xu, P. Cheng, Z. Chen, Y. Li and B. Vucetic, "Mobile Collaborative Spectrum Sensing for Heterogeneous Networks: A Bayesian Machine Learning Approach," in IEEE Transactions on Signal Processing, vol. 66, no. 21, pp. 5634-5647, 1 Nov.1, 2018
[4] MEEHL KOKLA Y. Application of reinforcement learning for security enhancement in cognitive radio networks[J]. Applied Soft Computing, 2015(37):809 - 829.