Research Article

A Distributed Q Learning Spectrum Decision Scheme for Cognitive Radio Sensor Network

Jian He, Jun Peng, Fu Jiang, Gaorong Qin, and Weirong Liu

School of Information Science and Engineering, Central South University, Changsha, Hunan 410000, China

Correspondence should be addressed to Fu Jiang; jiangfu0912@csu.edu.cn

Received 8 August 2014; Accepted 3 October 2014

Academic Editor: Shigeng Zhang

Copyright © 2015 Jian He et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cognitive spectrum management can improve the utilization efficiency of spectrum while increasing the energy consumption of sensor network nodes. Hence, how to balance the energy consumption and spectrum efficiency has become a critical challenge in the resource-constrained cognitive radio sensor networks. In this paper, by analyzing the channel characteristics and the energy efficiency of networks, a joint channel selection and power control spectrum decision algorithm based on distributed Q learning is proposed. To evaluate the performance of the proposed framework, an optimal Q value subject to communication efficiency index is formulated. Then, the learning strategy selection scheme is designed to solve the optimization problem by establishing a learning model. In this learning model, each node can get the strategy of other nodes to select the optimal strategy by introducing distributed strategy estimation. The simulation results show that the proposed algorithm has better performance than the existing methods.

1. Introduction

With the rapid development of wireless sensor networks, the traditional fixed spectrum allocation cannot meet the spectrum requirements of radio sensor network, so the cognitive radio sensor network (CRSN) arises at the historic moment [1], whose candidate characteristic is that the cognitive technique can be used for opportunistic spectrum access. But the dynamic spectrum management will increase the energy consumption of nodes while increasing network spectrum utilization [2], which is a severe challenge in CRSN with limited energy, storage, and computing resources. Therefore, how to ensure spectrum efficiency meanwhile without the energy efficiency loss is a critical issue in CRSN.

As the major part of spectrum management [3], spectrum decision is a crucial process in cognitive radio network [4], which can chose the best channels for secondary users to transmit data. Spectrum decision is usually divided into three steps [5], including channel characterization, channel selection, and parameter reconstruction. When the spectrum detected is available, the cognitive nodes will characterize the channel according to the observed local information and the channel statistical information of primary users. Then, nodes will select suitable channels according to these characteristics. Finally, we need to reconstruct the transmission parameters to adapt to the selected channel.

Due to the characteristics of spectrum holes [6], the behavior of primary user is changing over time, and cognitive nodes need to make spectrum decision dynamically to ensure the quality of communication [7]. Therefore, it is very important to seek an efficient spectrum decision method. Current spectrum decision methods can be divided into two categories [8]: the non-load-balancing method and the load balancing method.

For the non-load-balancing spectrum decision method, cognitive nodes can determine the communication channel according to the channel conditions, such as traffic load [9, 10], channel idle probability [11], the expected waiting time [12, 13], the expected remaining idle period [13, 14] or throughput expectation [15, 16]. Most methods have not considered spectrum sharing among cognitive nodes. If all cognitive nodes select the same frequency band for communication, there will exist serious channel competition [17]. In order to solve this problem, some scholars have begun to research spectrum decision methods based on load balancing.
For example, in [18], a spectrum decision method based on game is proposed to balance load, which uses game to seek the optimal channel choice probability. In order to reach the Nash equilibrium, each node relates its utility function to the candidate channel and then calculates the channel selection probability of each channel by the best response algorithm. In [19], a game theoretic framework is proposed to evaluate spectrum decision functionalities in CRSN. The spectrum decision process is cast as a noncooperative game among secondary users who can opportunistically select the “best” spectrum opportunity, under the tight constraint not to harm primary licensed users. However, because the information of each network node is changeable, the player should change their strategies instantaneously to reach equilibrium. It leads to a slow convergence speed. In this context, some scholars have introduced learning methods to solve the spectrum decision problem.

In [20], a method of channel choice probability based on adaptive learning is proposed. By exploring the uncertainty of cognitive network traffic, cognitive nodes can select the optimal channel, but its convergence speed may be slow if the network scale is large. Shiang et al. [21] assume that cognitive nodes have different priorities and present a dynamic strategy learning (DSL) algorithm that dynamically adapts the channel selection strategy to maximize the private utility function of nodes. By using this method, the spectrum decision of the cognitive nodes can reach the equilibrium, but it should be noted that the equilibrium of above method is not the global optimal solution, because each node strategy for spectrum decision is independently.

Energy efficiency has been researched in the current spectrum decision method, but there are still many difficulties, such as how to balance the communication performance and energy consumption, how to reduce the communication overhead, and how to improve energy efficiency and enhance the adaptability; those difficulties limit the application of the existing spectrum decision methods. Therefore, it is particularly important to design a spectrum decision method for CRSN, which can fully improve the efficiency of the spectrum management.

In this paper, we consider current CRSN requirements. By analyzing the network channel characterization and energy efficiency, we design an adaptive spectrum decision framework and propose a joint channel selection and power control spectrum decision algorithm based on distributed Q learning. In this algorithm, each node should consider other nodes strategies when it selects the strategy and then make decision together with other nodes. To evaluate the performance of the proposed framework and balance the energy consumption and spectrum efficiency, an optimal Q value which is subject to communication efficiency index is formulated. Then, the learning strategy selection scheme is designed by establishing a learning model to solve the optimization problem. The effectiveness of the proposed framework is validated by simulations.

The remainder of this paper is organized as follows. Section 2 describes the system model and problem formulation. Learning model and algorithm implementation are discussed in Section 3. Simulation results and analysis are given in Section 4, followed by concluding remarks work in Section 5.

2. System Model and Problem Formulation

In this section, we describe the network architecture and formulate the optimization as a comprehensive evaluation index which is subject to communication efficiency index.

2.1. Network Model. We consider a CRSN environment with some cognitive nodes. As shown in Figure 1. The network based on cluster structure and cluster nodes should cooperate with other nodes to determine the idle spectrum through spectrum sensing. Then, all the network nodes make spectrum decision together and the data is passed to the cluster head from network nodes within one hop, and the cluster nodes pass the data to the sink node with multiple hops.

In view of the cognitive wireless sensor network (CWSN), considering the general situation, the following assumptions are made throughout this paper.

(1) When the primary users is communicating, its transmission power is very high and CRSN nodes transmission power is relatively small, so, in this case, the network nodes cannot communicate with other nodes.

(2) Different cognitive wireless sensor node can be in the same channel for communication, but must adjust their own power to avoid interference with other nodes.

(3) In the process of spectrum decision, cognitive nodes do not need to exchange information with each other and select their communication channels and transmission power, respectively, so it can achieve the goal of energy conservation.

(4) We assume that the channel state transition probabilities, as well as the channel rewards, are unknown with the secondary nodes at the beginning. They
are fixed throughout the learning, unless otherwise noted. Therefore, the secondary nodes need to learn the channel properties.

(5) We consider all the noise as Gaussian white noise, and the mean value is 0 and the variance is $\sigma$.

2.2. Problem Formulation

2.2.1. Channel Characterization. In order to select appropriate channel, the network nodes must describe the current characteristics of each channel and ensure the current status of its. In this paper, we mainly consider the bandwidth, signal interference, false alarm rate of spectrum detection, and the idle time of band. The idle channel is evaluated whether it is suitable for communication by a comprehensive evaluation index as the current state of the channel. The following factors will be considered to construct the comprehensive index:

(1) channel bandwidth $W_d$: cognitive techniques nodes can detect the whole communicative spectrum and find the idle channel, but those channels may belong to different frequency bands. However, the channel division for different frequencies may be different, so the bandwidth is different and the channel capacity of idle channel is also different. The network node must consider the channel selection and power control according to different channel bandwidths;

(2) signal interference $I_{d,i}$: $I_{d,i}$ is the interference size of received signal at channel $d$ in time $t$, and it contains white noise interference and the interference of other nodes. Nodes can make detection according to the current channel interference, and $I_{d,i}$ shows that the greater the value means the worse the channel condition;

(3) the band last free time $T_{d, idle}^i$: $T_{d, idle}^i$ is the time interval between primary user appearances on band $d$; its value predicts the communication time of network nodes in this band. Because the network nodes expect continuous uninterrupted communication, so, in the channel selection process, network node will tend to select the channel of large idle time;

(4) spectrum sensing of false alarm rate $P_f$: due to the fact that the primary user behavior is unpredictable, the spectrum sensing cannot be ensured completely. Different communication frequency shows different characteristics for shadow and fading, so the $P_f$ of different frequency bands is different.

In this paper, we assume that the $\theta_{d,i}$ means comprehensive evaluation value on band $d$ for time $t$. Then, we proposed a multiobjective function as follows:

$$\theta_{d,i} = \sum_{j=1}^{n} w_j \times \text{eval}(\text{channel}_j),$$

where $\text{eval}(\text{channel}_j)$ is the evaluation value of the specific channel $j$. The evaluation value can be calculated by the comprehensive index we defined above. The $w_j$ is the weighting factor of the objective $j$. Different weights can be assigned according to different requirements.

$$\text{eval}(\text{channel}_j) = \sum_{k=1}^{m} w_{k,j} \times \text{value}(\text{channel}_j, k),$$

where $w_{k,j}$ is the weighting factor of the parameter $k$ for the channel $j$. $\text{value}(\text{channel}_j, k)$ is the value of the parameter $k$ for the channel $j$.

2.2.2. Energy Efficiency Analysis. Since cognitive wireless sensor nodes can communicate successfully on the idle channel, the nodes need to adjust the transmission power and optimize the energy efficiency. Due to the multiple nodes communication on the same band, there may exist both Gaussian white noise and mutual interference of each node at the receiving end. On one hand, the network nodes need to increase its transmission power, in order to obtain higher signal-to-interference plus noise ratio (SINR) and higher transmission rate, and then can get a better QoS; on the other hand, the network nodes must reduce the transmission power to achieve the goal of energy conservation, reducing the interference to other nodes at the same time. Therefore, the communication efficiency index is proposed to consider both communication quality and energy consumption, which will be the input of learning algorithm to realize balance.

Compared with the primary user, the CRSN nodes transmit data with low power, so its communication range is small. In this paper, we assume that the communication of each cognitive node is completely sight path, namely, the wireless transmission model is a free space propagation model, in which the channel gain $h$ is as follows:

$$h = \frac{G_i G_r}{(4\pi)^2 d^2} \left(\frac{c}{f}\right)^2,$$

where $G_i$ is the transmitting antenna gain and $G_r$ is the receiving gain; $c$ is the speed of light, and $f$ is the communication frequency; $d$ is distance of receiver and transmitter. Assuming that $\eta_i$ is the signal to interference plus noise ratio (SIRN) for receiver $i$, then

$$\eta_i = \frac{h_i p_i}{\sigma + \sum_{j \in N, j \neq i} h_j p_j},$$

where $p_i$, $p_j$ ($P_{min} < p_i, p_j < P_{max}$) is the transmission power of nodes for transmitter $i$ and $j$, $P_{min}$ and $P_{max}$ are the minimum and maximum thresholds of transmission power, $h_{ij}$ is the channel gain from transmitter $j$ to receiver $i$, $h_{ij}$ is the channel gain from transmitter $i$ to receiver $j$, and $\sigma$ is Gaussian noise power. To guarantee the QoS requirement, all the nodes need to make sure that its SIRN must be greater than a certain threshold value $\eta_i^*$:

$$\eta_i \geq \eta_i^*, \quad \forall i \in N.$$  

In order to achieve the equilibrium between communication ability and energy consumption, this paper defines the average number of bits in unit energy as the communication efficiency index:

$$\Phi_i = \frac{W_i \log_2 (1 + \eta_i)}{p_i},$$

where $W_i$ is the communication bandwidth.
2.2.3. Joint Spectrum Decision. In this paper, we propose a joint channel selection and power control spectrum decision, as shown in Figure 2. Firstly, the network nodes must describe the current characteristics of each channel and determine the current status which is considered as the input of distributed Q learning. In order to guarantee the network communications QoS constraints and minimize the energy consumption of the network nodes, this paper considers both channel switching and energy efficiency to design return value and then calculates the instant return values for different network conditions. Finally, we realize the joint channel selection and power control spectrum decision by introducing distributed Q learning algorithm.

In order to balance network communication ability and energy efficiency, we formulate the optimization as follows:

$$Q^* = \max_{a_t} Q_{i,t+1}(s_t, a_t)$$

s.t. $$\Phi_i = \frac{W_t \log_2 (1+\eta_i)}{p_i} > \Phi_i^*,$$

where $$Q^*$$ means the optimal Q value and $$\Phi_i^*$$ is the minimum requirement of communication efficiency index.

3. Learning Model and Algorithm Implementation

In order to realize the balance of communication quality and energy consumption and optimize the network communication ability in the prerequisite of communication efficiency index, this section presents an adaptive spectrum decision based on distributed Q learning.

3.1. Learning Algorithm Analysis. Reinforcement learning is an on-line technique [22] that considers environmental feedback as the input and learns through constant interaction with the environment, then uses the feedback signal to find the optimal action which adapted to the current environment. Reinforcement learning systems mainly include two parts [23]; they are both environment and agent, and the basic framework is shown in Figure 3.

As a model irrelevant learning algorithm, Q learning mainly cares about the evaluation value $$a_t$$ and selects the optimal state action according to $$Q(s,a)$$. Usually, we consider $$Q^*(s,a)$$ as the optimal evaluation value and $$\pi^*(s,a)$$ as the optimal strategies.

Assuming that the state set is $$s_t$$ and the action set is $$a_t$$, the evaluation value at the next time $$t+1$$ can be calculated by the formula as follows:

$$Q_{i,t+1}(s_t, a_t) = (1 - \alpha_t) Q_{i,t}(s_t, a_t) + \alpha_t r_{i,t+1} + \alpha_t \max_{a_t'\in A} Q_{i,t}(s_t', a_t').$$

where $$\gamma$$ is the discount factor, $$\alpha_t$$ is the learning rate, $$r_{i,t+1}$$ is the return value at next time, and $$Q_{i,t}(s_t, a_t)$$ is the value function of state action for node $$i$$; its means are the sum of the return value by executing action $$a_t$$ at state $$s_t$$.

In the strategy selection of learning algorithm, there exists a balance problem of exploration and exploitation. Exploration means the agent continuously updates learning knowledge to find the better strategy; exploitation means the agent selects the optimal action from all the action. In order to solve the balance problem, there are some algorithms like $$\varepsilon$$ greedy algorithm and soft-max algorithm [24]. The $$\varepsilon$$ greedy algorithm adopts random way to search. All the actions are chosen coequally. It means the worst action and the optimal action is chosen with the same probability, which will reduce the efficiency of learning algorithm. However, the soft-max
algorithm sets different strategies according to different Q value, and it uses Boltzmann distribution to define action selection probabilities:

\[ \pi_{i,t}(s_i, a_{ij}) = \frac{e^{Q_{i,j}(s_i, a_{ij})/\tau}}{\sum_{b \in A} e^{Q_{i,j}(s_i, b_{ij})/\tau}}, \quad (8) \]

where factor \( \tau > 0 \) specifies how randomly values should be chosen. High values for \( \tau \) means the actions will be chosen almost uniformly. As it reduced, the highest valued actions are more likely to be chosen, and when the limit is \( \tau \to 0 \), the best action is always chosen.

Due to the cognitive wireless sensor nodes that locate in the same network environment, all of the network resources are competed equally. Therefore, the behavior of each node will affect other nodes in spectrum decision, and other nodes are competed equally. Therefore, the behavior of each node the same network environment, all of the network resources are chosen almost uniformly. As it reduced, the highest valued selection probabilities:

\[ \text{With the time increase, } (1 - \phi^{-1})/(1 - \phi') \to 0, \text{ so, we can simplify formula (II) as the following update algorithm:} \]

\[ \pi_{i,j} (s_i, a_{ij}) = \frac{1 - \phi}{1 - \phi'} \phi W_{i,0} (s_i, a_{ij}) + \pi_{i,j-1} (s_i, a_{ij}). \quad (12) \]

So, the strategy selection method of formula (9) can change to

\[ \pi_{i,j} (s_i, a_{ij}) = \frac{\pi_{i,j} (s_i, a_{ij})}{\pi_{i,j} (s_i, a_{ij}) + \sum_{j \in N, j \neq i} \omega_{ij} \pi_{i,j} (s_i, a_{ij})}. \quad (13) \]

3.2 Learning Model. In this paper, we consider each network node as an agent which can select communication channel and transmission power adaptively. This dynamic adjustment process can be defined as a Markov decision process (MDP), and its model is made up of a triple \( M = \langle S, A, R \rangle \), where \( S \) is the state set of finite network nodes, \( A \) is the action set of finite network nodes, and \( R : S \times A \to \mathbb{R} \) is reward function, and \( R(s, a) \) signifies the reward of taking action \( a \) in state \( s \).

According to the current network environment, network nodes select appropriate communication channel and transmission power to transmit data. Now, we define state, action, and reward function, respectively.

State \( s_i \). The first step is to identify what a system state represents. For example, a state can be a combination of currently active network optimization services. Duration of the learning process directly depends on the number of states. As such, reducing their number will speed up the decision process. In this paper, the network environment state is \( s_i = [D_i, \theta_{D_i}] \), where \( D_i = \{d_1, d_2, \ldots, d_K\} \) is a set of detected available frequency band, \( \theta_{D_i} \) is the communication state of the band, namely, the comprehensive evaluation value in formula (1).

Action \( a_{ij} \). Each state has one or more associated actions. Any change of a network property (selecting the transmission channel, changing communication power, etc.) is considered as an action. Consequently, the number of available actions at each state will depend on the number of properties that can be modified and the number of distinct values that can be assigned to them and possible constraints defined by the system architect. In this paper, the action of network nodes is \( a_{ij} = [d_i, p_i] \), where \( d_i \) is the available channel which is selected by nodes to communicate and \( p_i \) is the communication power. Each network node can switch to a better channel or adjust the transmission power of the current channel.

Reward Function \( r_i \). Rewards are assigned with the intention to reinforce specific state-action pairs and can be positive or negative. Due to the fact that network nodes can adjust the action immediately according to the received rewards, choosing and defining the rewards are challenging at times. In this paper, we designed different reward values for different network conditions. Consider
Input: The initial learning rate $\alpha_t$, discount factor $\gamma$, state sets $S$, action sets $A$.
Output: optimal strategy $\pi^*_i(s_t, a_t)$

Steps:
1. Initialize the learning strategies $\pi_{ij}(s_t, a_{ij})$ and learning parameters $\alpha_t$.
2. Each node obtains the network state information of current available channel.
3. Calculate the comprehensive evaluation value according to the formula (1) to determine the state input of learning algorithm.
4. If the network state changes, then skip to Step 5, and need to select the channel and power again, otherwise skip to Step 8, and the nodes can communicate normally.
5. Calculate the reward value immediately and update the learning rate, then use the formula (7) to update the Q value table.
6. Record and estimate the strategy of other network nodes, and use the formula (12) to update the node strategy.
7. According to current strategy select the optimal action $a_i$, namely communication channel and power.
8. If the quality of the selected channel is poor, namely $\Phi_i < \Phi_i^*$, need to switch the channel and return to Step 5, and now the reward value is $-0.1$.
9. If the primary user appears in the communication process, then need to return to Step 2 and determine the idle spectrum set and network state again.

Algorithm 1: The joint spectrum decision of channel choice and power control with distributed Q learning.

(1) collision with primary user: due to the fact that the behavior of primary user is unpredictable and spectrum detection has a certain error, network node may conflict with the primary user when it selects a channel for communication. Then, we define the reward value as the lowest $-0.5$.

(2) channel switching: when the interference of current communication channel increased or the primary user suddenly appeared, nodes require switching of the channel. But switching the channel frequently will lead to excessive energy consumption, so we should avoid the channel switching times and set the reward value as $-0.1$.

(3) power adjustment: when the network nodes are working in a normal communication channel, it declare that this channel can satisfy the communication conditions of nodes, and nodes only need to adjust their power to achieve QoS and energy consumption constraints. So, we define the communication efficiency index in formula (5) as the reward value, but the SINR must satisfy constraint (4); otherwise, the reward value is 0.

Integrating all the situations, reward function $r_i$ can be defined as follows:

$$r_{i,t} = \begin{cases} 
-0.5 & \text{if primary user exist} \\
-0.1 & \text{if switch the channel} \\
0 & \text{if } \eta_i \leq \eta_i^* \\
\Phi_i & \text{if normal communication.}
\end{cases}$$

(14)

After the nodes obtain the free frequency band by cooperation, each node has competition with other nodes in the available channel for data transmission. The environment of CRSN is dynamic and complex networks, whose states are affected by many factors, which require network nodes to adjust the communication parameters adaptively and reduce the interference of each node as far as possible and then maximize the network energy efficiency under the demand of a certain communication at the same time. The transmission power is also different when the nodes work in different channel for wireless communication environment. Good channels only need a little transmission power, while poor channels need to increase transmission power to guarantee transmission. Therefore, the joint decisions for communication channel and transmission power can meet the demand of communication and energy efficiency.

In this paper, a joint spectrum decision of channel choice and power control based on distributed Q learning is proposed. By considering the energy efficiency and QoS constraints and reducing energy consumption as far as possible, we can achieve the purpose of extending network survival time. In order to save energy and reduce the communication overhead, this paper adopts the distributed Q value updating method, as shown in formula (7).

The joint spectrum decision of channel choice and power control with distributed Q learning is shown in Algorithm 1, and the flat chat of this algorithm is shown in Figure 4.

4. Analytical and Simulation Result

In this paper, we assume that there are six fixed clusters and a sink node in the CRSN, where each cluster is made up of 10 nodes, and each node selects appropriate channel and transmission power for data transmission. The radius of cluster is 70 m, and each cluster contains three primary users, whose transmission power is larger than other cognitive nodes. Assume that there are 20 authorized bands, including 4 VHF-TV bands, 4 AMPS bands, 4 GSM bands, 4 CDMA bands, and 4 WCDMA bands, and their bandwidths are 6 MHz, 30 KHz, 200 KHz, 1.25 MHz, and 5 MHz. The power of
Gaussian white noise is $\sigma = 10^{-7}$ mW, and the sending and receiving antenna gain is 2. The CRSN node transmission power set is $\{100$ mW, 125 mW, 150 mW, 175 mW, 200 mW\}. With parameter values $\gamma = 0.9$, $\tau = 0.005$, and $\alpha_0 = 0.1$.

In this section, we will use the following performance indicators to evaluate the performance of the proposed algorithm, which will be compared with game algorithm and dynamic strategy learning algorithm, respectively:

1. **energy efficiency**: the number of bits per unit energy transmission. It is an important index in this paper and it reflects the situation of energy utilization;
Based on game
Dynamic strategy learning
Distribute Q learning

Figure 5: The energy efficiency of network node.

(2) average channel switching times: the average channel switching times of network nodes in the whole communication. The smaller the value the better the algorithm, and it can reduce the energy consumption indirectly;

(3) average throughput: the number of bits per unit time in 1 Hz bandwidth. The index reflects the influence of the QoS of communication;

(4) successful transmission probability: the probability of successful communication for nodes in idle channel. The index reflects the advantages and disadvantages of the strategy and algorithm.

In Figure 5, we can see the changes of energy efficiency for different algorithms. Each algorithm will reach the convergence as time goes on. And we can find that the dynamic strategy learning algorithm can reach convergence quickly, but since there is no consideration about other nodes' strategies when the node selects strategy, it cannot realize the high energy efficiency. The spectrum decision based on game is better than dynamic strategy learning in energy efficiency but has inferior performance compared to the distribute Q learning, due to more information exchanged and more iteration needed in game framework. From Figure 5, it is shown that the distributed Q learning algorithm has the best energy efficiency and fastest convergence, because it considers the selection influence of both channel and power.

Figure 6 is the average channel switching times, it shows that the channel switching times of the proposed algorithm are convergent to 0.9, while dynamic strategy learning is convergent to 1.4 and the algorithm based on game is convergent to 3.6. Spectrum decision based on game algorithm requires communication between each node to follow the changes of nodes’ information, so it needs most of the channel switching times. For the dynamic strategy learning method, because it did not consider the strategy of the other nodes, it may not choose the optimal channel, requiring the adjusting of the channel from time to time. The proposed algorithm has carried on the comprehensive evaluation and also considered the estimation strategy of other nodes for channel and power choice. Thus, it can choose the best channel and power, which has the fewest channel switching times and then reduce the energy consumption.

The comparison of the average throughput in CRSN is shown in Figure 7. It is shown that the proposed algorithm has the best network performance, and its throughput is superior to other algorithms. Compared with other algorithms, the proposed algorithm let each node get other nodes’ strategies to select the optimal strategy by introducing distributed strategy estimation. Thus, it can select optimal channel and transmission power quickly, improving the rate of the data transmission. Therefore, the proposed algorithm can provide better QoS guarantee for CRSN.

As shown in Figure 8, the algorithm based on game can reach 82.8%, the algorithm based on dynamic strategy learning can reach 87.6%, and the proposed algorithm can achieve
93.8%. Because there existed strategy estimation between the nodes when selecting channel and power, the proposed algorithm can get the optimal selection strategy. Thus, it can achieve high success rate for data communication in CRSN. However, the dynamic strategy learning and game algorithm do not consider other nodes' strategies, so it cannot guarantee the global optimum and have inferior transmission rate compared to the proposed algorithm.

5. Conclusion

In this paper, we consider the requirements of current CRSN and design an adaptive spectrum decision framework by analyzing the network channel characterization and energy efficiency. To balance the energy consumption and spectrum efficiency of this framework, we adopt a distributed Q learning algorithm to implement channel selection and power control jointly, which takes channel state as the input and takes the selected channel and transmit power as the output. By using this algorithm, the network nodes can get the optimal transmitted power and communication channel to guarantee the energy efficiency and spectrum efficiency simultaneously. Future works will focus on the restraining the interference of data transmission between secondary nodes when selecting the idle channel.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to acknowledge that this work was partially supported by the National Natural Science Foundation of China (Grant nos. 61379111, 61202342, 61402538, and 61403424) and Research Fund for the Doctoral Program of Higher Education of China (Grant no. 2011062110042).

References

[1] G. P. Joshi, S. Y. Nam, and S. W. Kim, “Cognitive radio wireless sensor networks: applications, challenges and research trends,” Sensors, vol. 13, no. 9, pp. 1196–11228, 2013.
[2] B. Wang and K. J. R. Liu, “Advances in cognitive radio networks: a survey,” IEEE Journal on Selected Topics in Signal Processing, vol. 5, no. 1, pp. 5–23, 2011.
[3] J. Jia, Q. Zhang, and X. Shen, “HC-MAC: A hardware-constrained cognitive MAC for efficient spectrum management,” IEEE Journal on Selected Areas in Communications, vol. 26, no. 1, pp. 106–117, 2008.
[4] I. F. Akyildiz, W. Y. Lee, M. C. Vuran, and S. Mohanty, “A survey on spectrum management in cognitive radio networks,” IEEE Communications Magazine, vol. 46, no. 4, pp. 40–48, 2008.
[5] L.-C. Wang, C.-W. Wang, and F. Adachi, “Load-balancing spectrum decision for cognitive radio networks,” IEEE Journal on Selected Areas in Communications, vol. 29, no. 4, pp. 757–769, 2011.
[6] S. Srinivasa and S. A. Jafar, “The throughput potential of cognitive radio: a theoretical perspective,” IEEE Communications Magazine, vol. 45, no. 5, pp. 73–79, 2007.
[7] M. T. Masona, M. Mzyece, and N. Ntatlapa, “Spectrum decision in cognitive radio networks: a survey,” IEEE Communications Surveys and Tutorials, vol. 15, no. 3, pp. 1088–1107, 2013.
[8] W. Y. Lee and I. F. Akyildiz, “A spectrum decision framework for cognitive radio networks,” IEEE Transactions on Mobile Computing, vol. 10, no. 2, pp. 161–174, 2011.
[9] S.-C. Lo and C.-W. Tseng, “A novel multi-channel MAC protocol for wireless ad hoc networks,” in Proceedings of the IEEE 65th Vehicular Technology Conference (VTC ’07), pp. 46–50, Baltimore, Md, USA, April 2007.
[10] P. Zhu, J. Li, K. Han, and X. Wang, “A new channel parameter for cognitive radio,” in Proceedings of the 2nd International Conference on Cognitive Radio Oriented Wireless Networks and Communications, CrownCom, pp. 482–486, Orlando, Fla, USA, August 2007.
[11] T.-S. Kim, H. Lim, and J. C. Hou, “Understanding and improving the spatial reuse in multihop wireless networks,” IEEE Transactions on Mobile Computing, vol. 7, no. 10, pp. 1200–1212, 2008.
[12] S. U. Yoon and E. Ekici, “Voluntary spectrum handoff: a novel approach to spectrum management in CRNs,” in Proceedings of the IEEE International Conference on Communications (ICC ’10), pp. 1–5, Princeton, NJ, USA, May 2010.
[13] Y. Song and J. Xie, “Common hopping based proactive spectrum handoff in cognitive radio ad hoc networks,” in Proceedings of the 53rd IEEE Global Communications Conference (GLOBECOM ’10), pp. 1–5, Miami, Fla, USA, December 2010.
[14] S.-U. Yoon and E. Ekici, “Voluntary spectrum handoff: a novel approach to spectrum management in CRNs,” in Proceedings of the IEEE International Conference on Communications (ICC ’10), pp. 1–5, Cape Town, South Africa, May 2010.
[15] Q. Zhao, L. Tong, A. Swami, and Y. Chen, “Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: a POMDP framework,” IEEE Journal on Selected Areas in Communications, vol. 25, no. 3, pp. 589–600, 2007.
[16] Q. Zhao, S. Geirhofer, L. Tong, and B. M. Sadler, "Opportunistic spectrum access via periodic channel sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 785–796, 2008.

[17] D. Niyato and E. Hossain, "Competitive spectrum sharing in cognitive radio networks: a dynamic game approach," *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, pp. 2651–2660, 2008.

[18] A. T. Chronopoulos, M. R. Musku, S. Penmatsa, and D. C. Popescu, "Spectrum load balancing for medium access in cognitive radio systems," *IEEE Communications Letters*, vol. 12, no. 5, pp. 353–355, 2008.

[19] I. Malanchini, M. Cesana, and N. Gatti, "On spectrum selection games in cognitive radio networks," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '09)*, pp. 1–7, Honolulu, Hawaii, USA, December 2009.

[20] Y. Song, Y. Fang, and Y. Zhang, "Stochastic channel selection in cognitive radio networks," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '07)*, pp. 4878–4882, Washington, D.C., USA, November 2007.

[21] H.-P. Shiang and M. Van Der Schaar, "Queueing-based dynamic channel selection for heterogeneous multimedia applications over cognitive radio networks," *IEEE Transactions on Multimedia*, vol. 10, no. 5, pp. 896–909, 2008.

[22] D. E. Moriarty, A. C. Schultz, and J. J. Grefenstette, "Evolutionary algorithms for reinforcement learning," *Journal of Artificial Intelligence Research*, pp. 241–276, 1999.

[23] M. Littman and J. Boyan, "Reinforcement learning scheme for network routing," in *Proceedings of the International Workshop on Applications of Neural Networks to Telecommunications*, Psychology Press, 2013.

[24] M. Tokic and G. Palm, "Value-difference based exploration: adaptive control between epsilon-greedy and softmax," in *KI 2011: Advances in Artificial Intelligence*, vol. 7006 of *Lecture Notes in Computer Science*, pp. 335–346, Springer, Berlin, Germany, 2011.