sensors

MDPI

Article

Spectrum Handoff Based on DQN Predictive Decision for Hybrid Cognitive Radio Networks

Kaitian Cao [1,2,*] and Ping Qian [1]

[1] School of Electrical & Electronic Engineering, Shanghai Institute of Technology, Shanghai 201418, China; qping@sit.edu.cn

[2] Key Laboratory of Broadband Wireless Communication and Sensor Network Technology, Ministry of Education, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

* Correspondence: ktcao@sit.edu.cn; Tel.: +86-021-6087-3443

Received: 30 January 2020; Accepted: 18 February 2020; Published: 19 February 2020

Abstract: Spectrum handoff is one of the key techniques in a cognitive radio system. In order to improve the agility and the reliability of spectrum handoffs as well as the system throughput in hybrid cognitive radio networks (HCRNs) combing interweave mode with underlay mode, a predictive (or proactive) spectrum handoff scheme based on a deep Q-network (DQN) for HCRNs is proposed in this paper. In the proposed spectrum handoff approach, spectrum handoff success rate is introduced into an optimal spectrum resource allocation model to ensure the reliability of spectrum handoff, and the closed-form expression for the spectrum handoff success rate is obtained based on the Poisson distribution. Furthermore, we exploit the transfer learning strategy to further improve the DQN learning process and finally achieve a priority sequence of target available channels for spectrum handoffs, which can maximize the overall HCRNs throughput while satisfying constraints on secondary users' interference with primary user, limits on the spectrum handoff success rate, and the secondary users' performance requirements. Simulation results show that the proposed spectrum handoff scheme outperforms the state-of-the-art spectrum handoff algorithms based on predictive decision in terms of the convergence rate, the handoff success rate and the system throughput.

Keywords: cognitive radio networks; spectrum handoffs; machine learning; deep Q-network; transfer learning

1. Introduction

Cognitive radio networks (CRNs) have received great attention due to their potential to provide an efficient solution to the contradiction between spectrum scarcity and inefficient spectrum utilization, and improve system capacity via dynamic spectrum access (DSA) and spectrum management techniques [1,2]. Therefore, efficient spectrum management and resource allocation are crucial for CRNs to solve the shortage of spectrum resources and improve spectrum utilization [3,4]. DSA in CRNs can be categorized into three modes: overlay, interweave, and underlay [4–6]. Theoretically, underlay mode can significantly improve spectrum efficiency due to the fact that it allows secondary users (SUs) to access the licensed spectrum along with active primary users (PUs) at the same time [4–8]. In interweave mode [5,6,9], SUs can be allowed to access the vacant spectrum which is not occupied by any active PUs. Whenever a PU becomes active, SUs must vacate the licensed spectrum immediately. In overlay mode, SUs are allowed to simultaneously share the licensed spectrum bands with PUs without imposing any constraint on SUs' transmission power if SUs have a full knowledge of PUs' signals characteristics [5,6], which is infeasible in practice due to the difficulty to obtain all the prior knowledge of PU's signals. Compared to overlay mode, interweave and underlay technologies can achieve higher spectrum utilization. Currently, there is hardly any consensus on which of the two

spectrum access modes (interweave and underlay) is more suitable for CRNs system [9]. Therefore, it is of more theoretical and practical significance to investigate hybrid CRN (HCRN) technology which is a mixed spectrum sharing mode by combining the interweave mode and underlay mode.

However, in order to achieve efficient spectrum utilization, HCRN systems face many technical challenges, one of which is spectrum handoff technology [10,11]. In HCRNs, when the channel performance deteriorates, or the SUs' interference with PU exceeds the PU's tolerance threshold, SUs have to vacate and switch to a new target channel to continue data transmission. According to the decision timing for selecting target channels, spectrum handoff methods can be classified into the reactive decision based and the proactive (or predictive) decision-based handoffs [11–13]. Predictive decision can select a series of prospective backup vacant target channels before spectrum handoff occurs, which can save substantial sensing time. Therefore, the predictive decision-based spectrum handoff schemes have become a research focus of CRNs [11,13]. In [14], Hoque et al. established an analytical mode for the probability of spectrum handoff and further derived an analytical expression for average spectrum handoff number for a SU based on the residual time distributions of spectrum holes, and investigated the effect of spectrum handoff delay on the performance of spectrum mobility in CRNs. However, [14] does not take into account the overall optimization problems, such as the SU's transmission rate, system throughput, and so on. The work in [15] developed an analytical model for the general case of non-identical channels in CRNs, and introduced the model for both fixed and probabilistic sequence approaches for target channel selection. In [16], authors proposed an adaptive hybrid spectrum sharing method based on the rate compensation approach and adapted best fit algorithms, taking static and dynamic spectrum sharing algorithms into consideration. In [17], Kumar et al. presented a proactive decision-based spectrum handoff algorithm by utilizing multi-attribute decision making method according to different requirements of network service. In [11], a proactive decision based-handoff scheme (PDBHS) for CRNs is proposed, in which a hybrid handoff strategy is addressed by minimizing the number of handoffs such that the total service time is minimized. PDBHS requires K fixed slots for spectrum handoff, and the available target idle channels are sorted in a decreasing order of probability of obtaining K consecutive idle time slots. However, since the spectrum handoff time of PDBHS is fixed as K time slots, it is impossible for PDBHS to really reduce the service time. Besides, some performance challenges such as the maximum system capacity are not considered in the PDBHS scheme.

In addition, the above spectrum handoff methods based on predictive decision still have the following drawbacks: (1) data transmission only between a pair of sending and receiving SUs is considered, however, the impact of surrounding SUs' behaviors on a SU is not taken into account; (2) only the spectrum handoff scenario in a single spectrum access mode is considered for CRNs, but the hybrid spectrum access scenario combining interweave mode with underlay mode as well as the multi-SU spectrum handoff problem are not addressed; (3) the spectrum handoff success rate or failure rate is not investigated yet.

In order to solve the shortcomings of the above existing spectrum handoff approaches, we propose a transfer learning (TL)-based predictive decision spectrum handoff (TL-PDSH) method by introducing a deep Q-network (DQN) [18], TL [19] strategy, and the handoff success rate in this paper.

Our main contributions are briefly summarized as follows:

(1) Spectrum handoff success rate is introduced, and a multi-SU DQN learning-based spectrum handoff method for HCRNs is developed.
(2) This paper develops an overall throughput optimization model for spectrum handoffs in HCRNs while meeting constraints on signal-to-interference plus noise ratio (SINR) thresholds, the level of SUs' interference with PU, and requirements for the handoff success rate.
(3) A DQN algorithm is used to obtain the optimal learning strategy and seek the target channel sequence for spectrum handoffs, and the TL strategy is introduced in our method to further improve the DQN learning process.

2. System Model

In this paper, we study HCRNs where two central wireless networks, a primary network and a secondary network, share the licensed spectrum in a hybrid way combing interweave with underlay. In this scenario, SUs and PUs are randomly located around secondary base station (SBS) and primary base station (PBS), respectively.

In this paper, we assume that a SU senses a set of N+1 non-overlapping PU channels among which one PU channel is being occupied by the SU and N is the number of remaining PU channels that are arranged as $\{\phi_i\}_{i=1}^{N}$ in an increasing order of their central frequencies. Before spectrum handoff occurs, it is assumed that SBS obtains the state information of PU channels in advance. In our spectrum handoff scheme, SBS predicts the N channels capacity and selects M available target channels $\{\varphi_i\}_{i=1}^{M}$ ($M \leq N$) ready for spectrum handoff from the N channels in which the interference caused to PU is below a certain threshold. Therefore, the proposed TL-PDSH method in this paper can predict the channel capacity of $\{\varphi_i\}_{i=1}^{M}$ and obtain the handoff priority sequence $\{\omega_i\}_{i=1}^{M}$ which is rearranged in a decreasing order of the channel capacity of $\{\varphi_i\}_{i=1}^{M}$.

We assume that the process of PU appearing in a licensed channel is a Poisson process with appearance rate 1/λ, then the time interval the channel remaining idle, denoted by X, obeys a Poisson distribution, and its probability density function (PDF) is:

$$f(x) = \begin{cases} \frac{1}{\lambda} e^{-x/\lambda}, x > 0 \\ 0, \text{ others} \end{cases} \tag{1}$$

where $\lambda = E(X)$, $E(\cdot)$ is the expectation operator. For brevity, the idle time durations of the target channel sequence $\{\omega_i\}_{i=1}^{M}$ are denoted as $\{\lambda_i\}_{i=1}^{M}$, respectively.

In HCRNs, when the licensed channel is not occupied by a PU, SUs can achieve the maximum transmission rate over the channel unoccupied by PU since the interference caused by the PU to SUs is minimal. In other words, the higher the probability that a PU channel is vacant is, the higher the probability of SU's switching to the vacant channel is. Therefore, in order to obtain the optimal overall throughput of HCRNs system, all SUs prefer to choose the licensed channels unoccupied by PUs as the target channels for the upcoming spectrum handoffs. Motivated by this, this paper introduces a spectrum handoff success rate denoted as P_s to characterize the above phenomena. From the Poisson distribution, it is easy to deduce the handoff success rate (the overall idle probability of the channel) on the target channel sequence $\{\omega_i\}_{i=1}^{M}$ as follows:

$$P_s = 1 - \prod_{i=1}^{M} \left(1 - e^{-[(i-1)T_H + T_{ACK}]/\lambda_i}\right) \tag{2}$$

where T_H is the period of handoffs, T_{ACK} denotes the time duration of spectrum handoff acknowledge.

In this paper, we assume that PUs and SUs transmit using adaptive modulation and coding (AMC) [20] technology. Through AMC technique, SBS can infer the channel state information (CSI) of both PU and SU by active learning, estimate some parameters such as the channel gains [21], and dynamically adapt their own parameters to meet the constraints on the interference with primary link. Therefore, In HCRNs, in order to avoid the impact of each SU's behavior on primary link, the total interference caused by all SUs to PU must be limited below an allowable level.

In AMC, the modulation methods and channel coding rates can be adjusted adaptively according to SINRs. By using AMC technique, the SINR measured for the PU at PBS, denoted as $SINR^{(p)}$, and for the SU$_i$ (i=1, 2, . . . , L) at SBS, denoted as $SINR_i^{(s)}$, which can be expressed as follows:

$$SINR^{(p)} = \frac{G_0^{(p)} P_0}{\sigma^2 + \sum_{i=1}^{L} G_i^{(p)} P_i} \tag{3}$$

$$SINR_i^{(s)} = \frac{G_i^{(s)} P_i}{\sigma^2 + G_0^{(s)} P_0 + \sum_{j \neq i}^{L} G_j^{(s)} P_j} \tag{4}$$

where L is the number of SUs, $G_0^{(p)}$ and $G_i^{(p)}$ are the channel gain between the PU and PBS, and SU_i 's channel gain to PBS, respectively. $G_i^{(s)}$ and $G_j^{(s)}$ denote SU_i's channel gain to SBS, and SU_j ($j \neq i$) to SBS channel gain, respectively. $G_0^{(s)}$ is the channel gain between the PU and SBS. P_0, P_i, and P_j are the transmitted powers by PU, SU_i, and SU_j, respectively. σ^2 is the AWGN power.

In HCRNs, in order to ensure the communication performance of PU and SUs, constraints should be imposed on the SINRs for both PU and SU_i, which can be formalized by introducing SINR thresholds μ_0 and μ_i as:

$$SINR^{(p)} \geq \mu_0 \tag{5}$$

$$SINR_i^{(s)} \geq \mu_i, i = 1, 2, \ldots, L \tag{6}$$

According to the power allocation scheme in [22], we can obtain the power P_i allocated to SU_i as below:

$$P_i = \frac{\beta_i \left(\sigma^2 + G_0^{(s)} P_0 \right)}{G_i^{(s)} \left(1 - \sum_{i=1}^{L} \beta_i \right)}, i = 1, 2, \ldots, L \tag{7}$$

where:

$$\beta_i = (1 + 1/\mu_i)^{-1} \tag{8}$$

and:

$$1 - \sum_{i=1}^{L} \beta_i > 0 \tag{9}$$

Substituting Equation (7) into Equations (3) and (4), then Equations (5) and (6) can be uniformly expressed as:

$$\sum_{i=1}^{L} \alpha_i \beta_i \leq 1 \tag{10}$$

where:

$$\alpha_i = \frac{G_i^{(p)} \left(\sigma^2 + G_0^{(s)} P_0 \right)}{G_i^{(s)} \left(-\sigma^2 + G_0^{(p)} P_0 / \mu_0 \right)} + 1 \tag{11}$$

From Equations (7) and (10), the SINR threshold μ_i needs to be adjusted dynamically in order to satisfy the constraint on the secondary link's interference with PU and maximize the overall throughput of HCRNs. In HCRNs, the SU_i's transmit bit rate R_i [20] can be expressed as:

$$R_i = \mathrm{W} \log_2 (1 + k \mu_i) \tag{12}$$

where W is the channel bandwidth, $k = -1.5/\ln(5r_b)$ is a constant determined by the maximum transmit bit error rate r_b.

3. TL-PDSH Spectrum Handoff Scheme Based on DQN

When SUs' interfere with PU, the success rate of the spectrum handoff and other conditions are satisfied, and in order to maximize the overall throughput of HCRNs system, the SINR threshold μ_i needs to be adjusted dynamically. Therefore, how to choose μ_i becomes crucial. The selection problem of the optimal SINR threshold μ_i^* can be formulated as:

$$\begin{aligned} \left\{\mu_i^*\right\} &= \text{argmax}\frac{1}{L}\sum_{i=1}^{L} R_i \\ s.t. \quad & \sum_{i=1}^{L} \alpha_i\beta_i \le 1 \\ & 1-\sum_{i=1}^{L} \beta_i > 0 \\ & P_s \ge \rho \end{aligned} \tag{13}$$

where ρ is the minimum success rate of spectrum handoff. The constrained optimization expression in Equation (13) is actually an optimal resource allocation problem, which can maximize the overall throughput of the HCRN system while meeting the constraints on both the SINR thresholds and the successful handoff rate. Reinforcement learning (RL) has been proved to be an effective solution for the resource allocation problem in communication systems [23]. In this paper, we assume that the set of actions and the set of states in RL model are $\boldsymbol{A} = \{a_1, a_2, \cdots a_n\}$ and $\boldsymbol{S} = \{s_1, s_2, \cdots s_m\}$, respectively. At instant t, the RL agent takes an action $a(t) \in \boldsymbol{A}$ in the state $s(t) \in \boldsymbol{S}$, receiving immediate reward $r(s,a)$, and then the state is transited into the next state $s(t+1) \in \boldsymbol{S}$. Q-learning is a classical RL algorithm, which first evaluates each action value (Q-value) of the learning agent and then obtains the optimal learning strategy based on Q-values.

However, Q-learning has two fatal disadvantages: (1) the sets of the states and actions applicable to Q-learning are very small; and (2) the predictive ability of Q-learning is very weak. To this end, in this paper, the TL-PDSH scheme establishes the action space, state space and reward function by introducing a neural network into the Q-learning method, yielding a DQN, and then a DQN algorithm is used to obtain the approximate estimator of Q-value and the optimal learning strategy. In order to maximize the overall throughput of HCRNs while meeting the constraints on SUs' interference with Primary link, SU_i (i = 1, 2, ... , L) needs to seek a suitable SINR threshold μ_i within a certain range, and the set of these thresholds constitutes the action space $\boldsymbol{A}_i$, denoted as $\boldsymbol{A}_i = \left\{\mu_i^{(1)}, \mu_i^{(2)}, \cdots\right\}$ where $\mu_i^{(t)}$ is the SINR threshold at instant t. States can be defined as the three constraints in Equation (13), and then the state space at instant t can be formalized as $s^{(t)} = \left(I^{(t)}, D^{(t)}, G^{(t)}\right)$ where:

$$I^{(t)} = \begin{cases} 0, \ \sum_{i=1}^{L} \alpha_i\beta_i^{(t)} \le 1 \\ 1, \ \text{others} \end{cases} \tag{14}$$

$$D^{(t)} = \begin{cases} 0, \ 1-\sum_{i=1}^{L} \beta_i^{(t)} > 0 \\ 1, \ \text{others} \end{cases} \tag{15}$$

$$G^{(t)} = \begin{cases} 0, \ P_s \ge]\rho \\ 1, \ \text{others} \end{cases} \tag{16}$$

The reward function is defined as a function in terms of the state space and the current action space, and then at instant t, SU_i (i=1, 2, ... , L) obtains the reward $r_i\left(s^{(t)}, a^{(t)}\right)$ as follows:

$$r_i\left(s^{(t)}, a^{(t)}\right) = \begin{cases} \Lambda, \ I^{(t+1)} + D^{(t+1)} + G^{(t+1)} \ge 1 \\ R_i, \ \text{others} \end{cases} \tag{17}$$

where Λ is a constant which is smaller than the reward received by an agent by taking any learning strategy. Therefore, Λ indicates that when any of the constraints in Equation (13) is not met, $r_i\left(s^{(t)}, a^{(t)}\right)$ is a penalty, not a benefit.

It can be seen from the above analysis that since the transmit rate R_i is positive, maximizing the overall throughput of HCRNs system is essentially to maximize the individual transmit rate R_i of SU_i. Therefore, the task of SU_i is to seek an optimal learning policy π through DQN learning algorithm so as to maximize its reward at the next moment, namely:

$$Q_i^*(s,a) = \max_{\pi}\left\{\sum_{t=0}^{\infty}\left[\gamma^{(t)}\times\ E\left(r_i\left(s^{(t)},a^{(t)}\right)\middle|s^{(t)}=s,a^{(t)}=a,\pi\right)\right]\right\} \tag{18}$$

where $\gamma^{(t)}$ is the discounting factor at each time step t, $E(\cdot)$ is the expectation operator, $Q_i^*(s,a)$ is the optimal Q-value function of SU_i, indicating the maximum sum of discounted rewards $r_i\left(s^{(t)},a^{(t)}\right)$ as $t\to\infty$, achieved by a behavior policy π. According to Bellman's principle of optimality [24], if the optimal value $Q_i^*(s',a')$ of the state sequence s' at the next time step is known for all possible actions a', then Equation (18) can be rewritten as:

$$Q_i^*(s,a) = E\left[r_i(s,a) + \gamma \max_{a'} Q_i^*(s',a')\right] \tag{19}$$

From Equations (18) and (19), the iterative Equation (19) converges to the optimal Q-value only as $t\to\infty$, which is impractical since the Q-value function is estimated separately for each state in practice, without any generalization. Instead, in the TL-PDSH method presented in this paper, the DQN neural network has been utilized as an effective approximator to estimate the Q-value function $Q_i(s,a\ ;\ \theta_i)\approx Q_i^*(s,a)$. In addition, the TL-PDSH method adopts a technique known as "experience replay" to improve learning performance. Different from the linear estimators widely used in the general RL methods, the TL-PDSH method is a nonlinear weighted approximator of the DQN neural network.

In experience replay, at each time step, SU_i stores the experience values $e_i^{(t)} = \left[a_i^{(t)}, s_i^{(t)}, r_i^{(t)}, s_i^{(t+1)}\right]$ interacting with the wireless environment into replay memory $M_i^{(t)} = \left\{e_i^{(0)}, e_i^{(1)}, \ldots, e_i^{(t)}\right\}$ (i=1, 2, ... , L). Suppose that θ_i^-, θ_i are the previous parameter of Q-value and the updated parameter of Q-value, respectively, then θ_i can be updated by minimizing the following loss function $L(\theta_i)$ under the current iteration step:

$$L(\theta_i) = E\left[\left(y_i - Q(s,a;\theta_i)\right)^2\right] \tag{20}$$

where $y_i = r_i(s,a) + \gamma \max_{a'} Q_i(s',a';\theta_i^-)$. In the proposed DQN-based TL-PDSH method, ϵ-greedy strategy [24] is used to select SU_i's action (SINR threshold μ_i), updating the parameters θ_i so as to achieve the most reward for SU_i and maximize the overall throughput of HCRNs system. Algorithm 1 displays main steps of our TL-PDSH scheme. Note that algorithm 1 does not introduce the TL strategy since it does not take into account the new SUs.

Algorithm 1. The proposed TL-PDSH scheme without TL (when no new SUs appear).

for all SU_i, i = 1, ... , L **do**
  Initialize replay memory $M_i^{(0)} = \left\{e_i^{(0)}\right\}$ at t = 0;
  Initialize θ_i and γ;
  Initialize the neural network for $Q_i(s,a\ ;\ \theta_i)$ with θ_i;
  Initialize the neural network for $Q_i(s',a';\theta_i^-)$ with $\theta_i^- = \theta_i$;
end for
for t < T **do**
  for all SU_i, i = 1, ... , L **do**
  Select a random action with probability ϵ (ϵ-greedy algorithm);
  Otherwise select the action $a_i^{(t)} = \text{argmax}\ Q_i\left(s_i^{(t)}, a_i^{(t)};\theta_i\right)$;
  Update the state $s_i^{(t+1)}$ in (14)–(16) and the reward $r_i^{(t)}$;
  Store $e_i^{(t)} = \left[a_i^{(t)}, s_i^{(t)}, r_i^{(t)}, s_i^{(t+1)}\right]$ in $M_i^{(t)}$;
  Update parameters of $Q_i(s,a\ ;\ \theta_i)$ by minimizing $L(\theta_i)$ from $M_i^{(t)}$;
  Update parameters of $Q_i(s',a';\theta_i^-)$ with θ_i^-=θ_i at each time step;
  end for
end for

In addition, considering that the parameters of Q function of two adjacent SUs are similar in HCRNs system, therefore, TL algorithm [19] is exploited in our TL-PDSH scheme to initialize the parameters of the newcomer SU with the parameters of its nearest SU in HCRNs, instead of initiating the DQN learning process from scratch. As a result, TL-PDSH method can greatly speed up the DQN learning process and improve the performance of the whole HCRNs system by introducing TL strategy into the proposed spectrum handoff method. When a new SU joins the HCRN system, our proposed TL-PDSH scheme with TL is described in Algorithm 2.

Algorithm 2. The proposed TL-PDSH scheme with TL (when a new SU appears).

A newcomer SU is denoted as SU_{L+1};
Determine the nearest SUs(i) of SU_{L+1};
Initialize Q_{L+1} with parameters of the Q value of the nearest neighbor $\theta_{L+1} = \theta_{s(i)}$;
Run **Algorithm 1**; /*(**Note:** now the number of SUs is (L+1))*/

4. Simulations Results

For convenience, the DQN-based TL-PDSH method proposed in this paper, the DQN-based PDSH algorithm without using the TL learning strategy, and the traditional PDSH algorithm based on Q-learning are denoted as TL-PDSH, PDSH, and Q-PDSH, respectively. In this section, the performance of TL-PDSH spectrum handoff method based on DQN is verified through Monte-Carlo simulations, and it is compared with the PDSH algorithm, the PDBHS algorithm [11], and the Q-PDSH method.

It is assumed that there is only one PU accessing a single channel in the primary network, the transmitting power of PBS is 100 mW, the Gaussian noise power is 10 nW, and the SINR for the PU is set at 1dB. The distance between PBS and SBS is 2 km, and PU and all SUs are randomly distributed around their respective base stations within a circle of radius of 200 m. Suppose that channel gains follow a log-distance path loss model with a path loss exponent 2.5. For all SUs, the discounting factor γ is 0.8. In the ϵ-greedy algorithm, ϵ is initially set at 0.8, converging to 0 with the increase of the number of iterations. In this section, each SU uses a feedforward neural network (FNN) that includes three hidden layers and two neurons. The input layer of the FNN has three nodes, including some information such as states and actions taken by neurons, while the output layer has only one node. The capacity of the experience replay memory and the update step size are set to 200 and 10, respectively.

Figure 1 shows that the average transmission rate $\frac{1}{L}\sum_{i=1}^{L} R_i$ versus the number of SUs for $\rho = 0.9$. It can be seen from the simulation results in Figure 1 that the average transmission rate decreases as the number of SU increases. The reason lies in that as the number of SU increases, the interference among SUs will increase, and in order to meet the constraints on SU's interference with PU and the level of success rate of spectrum handoff, the transmit power of each SU will decrease, and its SINR will also decrease accordingly, resulting in the decrease of each channel capacity. In addition, the simulation results show that the TL-PDSH algorithm can achieve the highest transmission rate while the Q-PDSH algorithm obtains the lowest.

Conflict rate is defined as the percentage of the number of SUs that violate the SINR constraints once or twice while all SUs remain within acceptable level of rewards. Curves of the conflict rate versus the number of SUs for $\rho = 0.9$ is illustrated in Figure 2. It can be seen from the curves that the conflict rate of the four algorithms increases rapidly with the increase of the number of SUs. The number of SUs increases and all SUs pursue the maximum transmission rate, which inevitably results in the increase of the number of SUs violating the SINR constraints, so the conflict rate increases accordingly. In PDBHS method, it selects the channel with the highest probability of getting K consecutive idle time-slots as the target channel for spectrum handoff, so the conflict rate of PDBHS is the lowest among the four algorithms. In addition, Figure 2 also shows that the TL-PDSH method is very close to the PDBHS method and the number of SUs needed to achieve the optimal transmit rate can be determined by fixing the conflict rate.

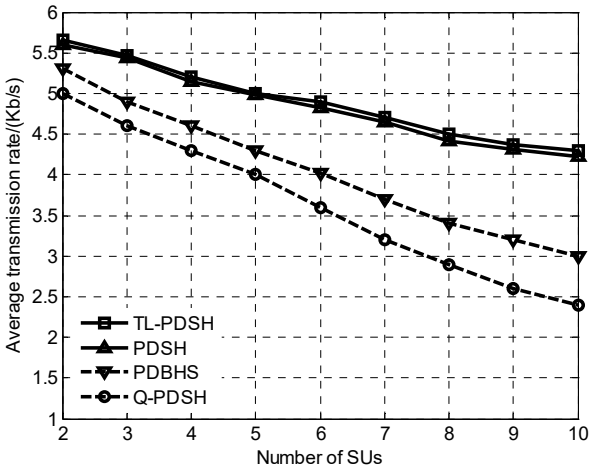

Figure 1. Average transmission rate versus number of SUs.

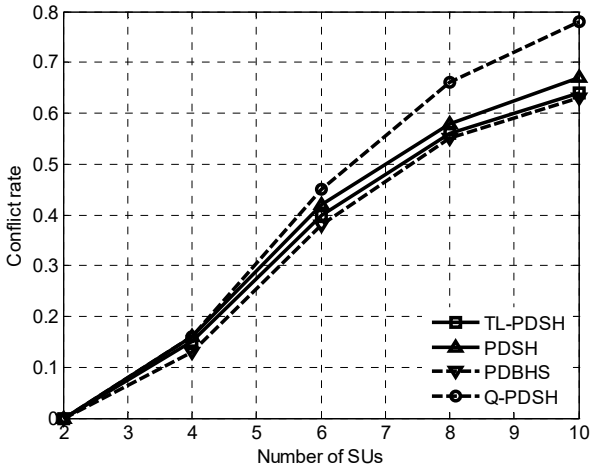

Figure 2. Conflict rate versus number of SUs.

When the four spectrum handoff algorithms converge, the curve of the average number of iterations with the number of SUs for $\rho = 0.9$ is shown in Figure 3. It can be seen that the number of iterations needed to converge for TL-PDSH and PDSH spectrum handoff schemes using DQN strategy is greatly reduced compared to PDBHS and Q-PDSH. It can also be seen that TL-PDSH algorithm utilizing both DQN and TL has the fastest convergence rate, which matches the theoretical analyses described above since TL can efficiently improve learning speed of the system and reduce the number of iterations by transforming the experienced results of the surrounding SUs to the new SUs.

The average transmission rate of the four algorithms versus handoff success rate for $L = 6$ is demonstrated in Figure 4. The average transmission rate of the four algorithms increases rapidly with the increase of the handoff success rate, which lies in the fact that the higher the probability of spectrum being vacant is, the higher the handoff success rate is, the higher SU's SINR is, and the higher the channel capacity is. In addition, there is no significant difference between the TL-PDSH algorithm and the PDSH algorithm in terms of average transmission rate performance since both algorithms use the DQN neural learning network.

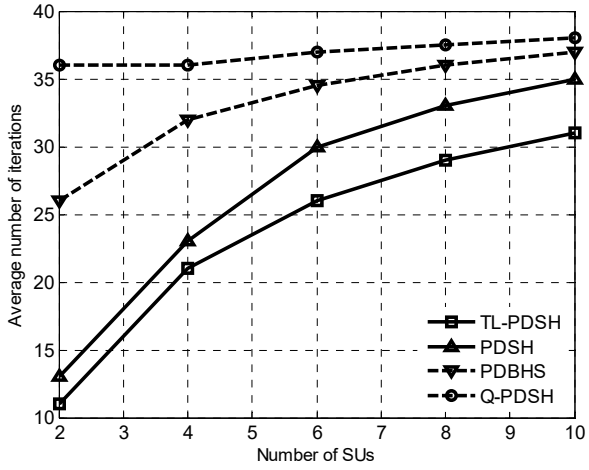

Figure 3. Average number of iterations versus number of SUs.

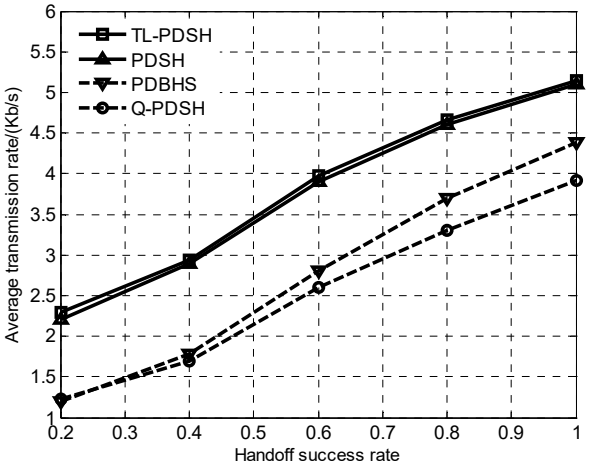

Figure 4. Transmission rate versus handoff success rate.

5. Conclusions

In this paper, a DQN neural learning network is used to investigate the spectrum handoff method based on predictive decision in HCRNs, and the TL-PDSH spectrum handoff method based on the DQN predictive decision while meeting the constraints on SUs' interference with PU and the level of SINR thresholds, as well as the requirements for spectrum handoff success rate is proposed by introducing the spectrum handoff success rate and the TL strategy. Simulation results show that, compared to the existing spectrum handoff methods based on predictive decision, the TL-PDSH method proposed in this paper yields better performances in terms of the system throughput, the handoff success rate and the number of iterations. Furthermore, numeric simulations also demonstrate that under the condition that all SUs seek their own maximum rewards, the conflict rate of our approach is almost the same as that of the PDBHS method which has the minimum conflict rate.

Author Contributions: The work described in this article is the collaborative development of all authors. K.C. contributed to the idea of data processing and the writing of this article, and derived the algorithms. P.Q. made contributions to simulations and result analysis. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by NSFC project (no. 61671252), Natural Science Foundation of Shanghai (No. 19ZR1455200), Youth Development Foundation of Shanghai Institute of Technology (no. ZQ2018-24), and Scientific Research Foundation for the Introduction of Talent of the Shanghai Institute of Technology (no. YJ2018-11).

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

References

1. Akyildiz, I.F.; LEE, W.Y.; Vuran, M.C.; Mohanty, S. A survey on spectrum management in cognitive radio networks. *IEEE Commun. Mag.* **2008**, *46*, 40–48. [CrossRef]
2. Ding, G.; Jiao, Y.; Wang, J.; Zou, Y.; Wu, Q.; Yao, Y.; Hanzo, L. Spectrum inference in cognitive radio networks: Algorithms and applications. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 150–182. [CrossRef]
3. Koushik, A.M.; Hu, F.; Kumar, S. Intelligent spectrum management based on transfer actor-critic learning for rateless transmissions in cognitive radio networks. *IEEE Trans. Mobile Comput.* **2018**, *17*, 1204–1215.
4. Tanab, M.E.; Hamouda, W. Resource allocation for underlay cognitive radio networks: A Survey. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 1249–1276. [CrossRef]
5. Awin, F.; Abdel-Raheem, E.; Tepe, K. Blind spectrum sensing approaches for interweaved cognitive radio system: A tutorial and short course. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 238–259. [CrossRef]
6. Awin, F.A.; Alginahi, Y.M.; Abdel-Raheem, E.; Tepe, K. Technical issues on cognitive radio-based Internet of Things systems: A survey. *IEEE Access* **2019**, *7*, 97887–97908. [CrossRef]
7. Shah-Mohammadi, F.; Kwasinski, A. Deep reinforcement learning approach to QoE-driven resource allocation for spectrum underlay in cognitive radio networks. In Proceedings of the IEEE International Conference on Communications, Kansas City, MO, USA, 20–24 May 2018; pp. 1–6.
8. Liu, M.; Zhang, J.; Li, B. Symbol rates estimation of time-frequency overlapped MPSK signals for underlay cognitive radio network. *IEEE Access* **2018**, *6*, 16216–16223. [CrossRef]
9. Mehmeti, F.; Spyropoulos, T. Performance analysis, comparison, and optimization of interweave and underlay spectrum access in cognitive radio networks. *IEEE Trans. Veh. Technol.* **2018**, *67*, 7143–7157. [CrossRef]
10. Wu, Y.; Hu, F.; Zhu, Y.; Kumar, S. Optimal spectrum handoff control for CRN based on hybrid priority queuing and multi-teacher apprentice learning. *IEEE Trans. Veh. Technol.* **2017**, *66*, 2630–2642. [CrossRef]
11. Gupta, N.; Dhurandher, S.K.; Woungang, I.; Obaidat, M.S. Proactive decision based handoff scheme for cognitive radio networks. In Proceedings of the IEEE International Conference on Communications, Kansas City, MO, USA, 20–24 May 2018; pp. 1–6.
12. Wang, C.W.; Wang, L.C. Analysis of reactive spectrum handoff in cognitive radio networks. *IEEE J. Sel. Areas Commun.* **2012**, *30*, 2016–2028. [CrossRef]
13. Zhao, Y.; Hong, Z.; Luo, Y.; Wang, G.; Pu, L. Prediction-based spectrum management in cognitive radio networks. *IEEE Syst. J.* **2018**, *12*, 3303–3314. [CrossRef]
14. Hoque, S.; Sen, D.; Arif, W. Impact of residual time distributions of spectrum holes on spectrum handoff performance with finite switching delay in cognitive radio networks. *Int. J. Electron. Commun.* **2018**, *92*, 21–29. [CrossRef]
15. Tayel, A.F.; Rabia, S.I.; Abouelseoud, Y. An optimized hybrid approach for spectrum handoff with non-identical channels. *IEEE Trans. Commun.* **2016**, *64*, 4487–4496. [CrossRef]
16. Lertsinsrubtavee, A.; Malouch, N. Hybrid spectrum sharing through adaptive spectrum handoff and selection. *IEEE Trans. Mobile Compu.* **2016**, *15*, 2781–2793. [CrossRef]
17. Kumar, K.; Prakash, A.; Tripathi, R. Spectrum handoff scheme with multiple attributes decision making for optimal network selection in cognitive radio networks. *Digital Commun. Netw.* **2017**, *3*, 164–175. [CrossRef]
18. Xiao, L.; Li, Y.; Han, G.; Dai, H.; Poor, H.V. A secure mobile crowdsensing game with deep reinforcement learning. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 35–47. [CrossRef]
19. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
20. Qiu, X.; Chawla, K. On the performance of adaptive modulation in cellular systems. *IEEE Trans. Commun.* **1999**, *47*, 884–895.
21. Zhang, R. On active learning and supervised transmission of spectrum sharing based cognitive radios by exploiting hidden primary radio feedback. *IEEE Trans. Commun.* **2010**, *58*, 2960–2970. [CrossRef]
22. Pietrzyk, S.; Janssen, G.J.M. Radio resource allocation for cellular networks based on OFDMA with QoS guarantees. In Proceedings of the IEEE Global Telecommunications Conference, Dallas, TX, USA, 29 November–3 December 2004; pp. 2694–2699.

23. Sutton, R.S.; Barto, A.G. *Reinforcement learning: An introduction*; MIT Press: Cambridge, MA, USA, 2011.
24. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.B.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef] [PubMed]