Sensors & Transducers, Vol. 173, Issue 6, June 2014, pp. 68-74

Sensors & Transducers

EWA Selection Strategy with Channel Handoff Scheme in Cognitive Radio

[1] Yong, Sun, [2] Jiansheng, Qian
School of Information and Electrical Engineering (SIEE), China University of Mining and Technology (CUMT), Xuzhou, Jiangsu, 221008, China
[1] Tel.: +86 0516-83884819, fax: +86 0516-83884819
[1] E-mail: SnowandSunshine@msn.com

Received: 18 April 20014 /Accepted: 31 May 2014 /Published: 30 June 2014

Abstract: This paper proposes an improved Experience-Weighted Attraction (EWA) selection algorithm with channel handoff scheme in order to meet the practical requirement for Cognitive Radio (CR) application. In further research, sensitivity and stability verification is conducted by focusing on the comparison with Q learning algorithm after the establishment of a wireless simulation space with time slot sequences that extremely imitates real Cognitive Radio environments. The emulation results show that channel selection strategy with EWA learning has more advantageous performance. *Copyright © 2014 IFSA Publishing, S. L.*

Keywords: Wireless communications, Intelligent Radio (IR), Cognitive Radio (CR), Reinforcement Learning (RL), Experience-Weighted Attraction (EWA).

1. Introduction

Traditionally, the licensed radio spectrum allocations are regulated by official authorities. The public and government use of radio spectrum is managed by the National Telecommunications and Information Administration (NTIA) and the Federal Communications Commission (FCC) is in charge of commercial radio resources respectively in the USA. However, as more and more applications of wireless devices, the rapid increasing requisition for radio spectrum licensing has led to current shortage of radio spectrum allocations and put their governing bodies into trouble. In fact, FCC’s recent research has shown that these fixed static frequency channels are always idle or not occupied in most of time. Spectrum bands are not efficiently used or underutilization either at a temporal or on a geographical level. By seeking “spectrum holes” (unused frequency channels), Cognitive Radio (CR) can greatly improve the use efficiency of spectrum resources and solve these problems presented above in "secondary utilization"(with lower priority than legacy users) way. First introduced by Mitola [1], Cognitive Radio is often considered as an extension and expansion of Soft Radio (SR), which is equipped by general hardware and capable of programming to transmit and receive various radio waves. There have already been lots of research in many aspects of Cognitive Radio. In sensing, Hossain, M. S., et al. [2] evaluated the performance of cooperative spectrum sensing with the hard combination OR, AND MAJORITY rules. Zhang, D., et al. [3] proposed a novel detection algorithm that the Fractal box dimension is used when the signal to noise (SNR) is high, while the improved TCC algorithm is used when the SNR is low and Khalaf, G. [4] formulated the detection problem based on the eigen-decomposition technique. In security, Shan-Shan, W., et al. [5] proposed a Four Dimensional Continuous Time Markov Chain model to analyze the communication performance of normal secondary users under PUEAs, typically affected by SMUs, and compared several PUEA detection schemes.

As revolutionary development of Intelligent Radio (IR), Cognitive Radio implements Soft Radio by adding Knowledge Base, Reasoning Engine and Learning Engine to be an independent Cognitive Engine (CE), which makes the radio capable of learning and adapting to the surrounding radio environment [6]. Knowledge Base which stores variety of cases, relations and rules can be seen as memory in human's brain and is very common in Artificial Intelligence (AI) logic planning. Just like expert system in Artificial Intelligence, Reasoning Engine executes all kinds of state information for reference of knowledge base by logic thinking then generates processed results or actions to drive soft radio changing setting parameters to adapt to changing environment. As the core component and key feature for Cognitive Radio implementation, Learning Engine is in charge of keeping Knowledge Base updated by accumulating new environmental experience into new knowledge extension, which is what differentiates Cognitive Radio from traditional pre-programmed ones.

There are varieties of learning algorithms available for Cognitive Radio, including neural networks, genetic models, and hidden Markov algorithms. Tsagkaris K., A. Katidiotis, et al. [7] used neural network-based learning to predict data bit rate of Cognitive Radio. Galindo-Serrano, A. and L. Giupponi [8] proposed a form of real-time decentralized Q-learning to manage the aggregated interference generated by multiple WRAN systems. Li, H. S. [9] applied Multi-agent reinforcement leaning (MARL) for the secondary users to learn good strategies of channel selection. Chen, X. F., Z. F. Zhao, et al. [10] presented an intelligent policy based on reinforcement learning to acquire the stochastic behavior of Primary Users (PUs). Zhang, W. Z. and X. C. Liu [11] obtained the capability of iteratively on-line learning environment performance by using Reinforcement Learning (RL) algorithm after observing the variability and uncertainty of the heterogeneous wireless networks. Gallego, J. R., M. Canales, et al. [12] provided no-regret learning algorithms to perform the joint channel and power allocation and overcome the convergence limitations of the local game. Zhu, J., J. Wang, et al. [13] employed Reinforcement Learning (RL) approach to finding a near-optimal policy under undiscovered environment. Torkestani, J. A. and M. R. Meybodi [14] proposed the learning automata-based cognitive radio to address the spectrum scarcity challenges in wireless ad hoc networks. Yang, M. F. and D. Grace [15] improved channel assignment in multicast terrestrial communication systems with distributed channel occupancy detection by using intelligence based on reinforcement learning and transmitter power adjustment. Zhou, P., Y. S. Chang, et al. [16] designed a robust distributed power control algorithm with low implementation complexity for Cognitive Radio networks through reinforcement learning, which does not require the interference channel and power strategy information among Secondary Users (SUs) and from SUs users to PUs.

However, as known with our best effort till now, little focus has been placed on implementing learning engine of cognitive radio with Experience-Weighted Attraction (EWA) algorithms. The innovative proposed channel selection algorithm based on EWA learning in this paper allows cognitive to learn radio environment communication channel characteristics online. By accumulating the history channel experience, it can predict, select and change the current optimal communication channel, dynamic ensure the quality of communication links and finally reduce system communication outage probability. The effectiveness of this algorithm has been validated by simple probability method in our preliminary studies [17]. However, the existing theoretical model is comparatively idealistic in practical application of complex wireless environments. Based on our lots of earlier research, the study focus has been shifted to EWA intelligent channel handoff mechanism after an important new parameter t in time domain was introduced in our simulation environments. Sensitivity and stability advantages are conducted by comparing Q learning, the mainstream algorithm in Reinforcement Learning (RL).

The rest of this paper is presented as follows. In section 2, the EWA algorithms will be introduced in full details. Intelligent channel selection algorithm of cognitive radio based on EWA learning with channel handoff scheme is following in section 3. Then the simulation results comparison and analysis are presented in section 4. In the end the conclusion comes in the final section 5.

2. EWA Learning

Experience-Weighted Attraction (EWA) is derived from normal form multi-game theory. Setting n players in the game, and denote $\mathrm{i}(\mathrm{i}=1,2,3,\ldots,\mathrm{n}-1,\mathrm{n})$ for each one. The strategy space for player i is $\mathrm{S_i}$. There are $\mathrm{m_i}$ discrete choices in total for each $\mathrm{S_i}$ and can be expressed as $\mathrm{S_i}=\{\mathrm{s_i^1},\mathrm{s_i^2},\mathrm{s_i^3},\ldots,\mathrm{s_i^{m_i-1}},\mathrm{s_i^{m_i}}\}$. One strategy of player i, denoted by $\mathrm{s_i}$, is the element of strategy space $\mathrm{S_i}$, or $\mathrm{s_i}\in\mathrm{S_i}$. The entire strategy space of the game S is the n-Cartesian product of individual strategy space, that is,

$$\mathrm{S}=\mathrm{S_1}\times\mathrm{S_2}\times\mathrm{S_3}\times\cdots\times\mathrm{S_{n-1}}\times\mathrm{S_n}.$$

Let s be the combination of all players' strategies in the game, then $\mathrm{s}=\{\mathrm{s_1},\mathrm{s_2},\mathrm{s_3},\cdots,\mathrm{s_{n-1}},\mathrm{s_n}\}\in\mathrm{S}$. The combination of all other n-1 players' strategies except player i can be expressed as

$$\mathrm{s_{-i}}=\{\mathrm{s_1},\mathrm{s_2},\mathrm{s_3},\cdots,\mathrm{s_{i-1}},\mathrm{s_{i+1}}\cdots,\mathrm{s_{n-1}},\mathrm{s_n}\}.$$

Denote $\mathrm{m_{-i}}$ be the combination number of $\mathrm{s_{-i}}$, then $\mathrm{m_{-i}}=\prod_{\mathrm{j=1,j\neq i}}^{\mathrm{n}}\mathrm{m_j}$. The reward function for player i with scalar-value is $\pi_\mathrm{i}(\mathrm{s_i},\mathrm{s_{-i}})$. Take time dimension for consideration, then the strategy of player i in time period t can be expressed as $\mathrm{s_i(t)}$, other players' strategy set(vector) $s_{-i}(t)$, and reward function for player i $\pi_i[s_i(t), s_{-i}(t)]$ respectively.

EWA learning algorithm assumes that any strategy has an attraction value. The model defines the initial values of attractions, how the attraction values are updated based on experience and determines the selection probabilities. The core algorithm is two variables updated each round. One is attraction value $A_i^j(t)$, it defines the attraction value of player i after selecting strategy j. The attraction update rule is that the current attraction value is the last attraction value $A_i^j(t-1)$ multiplied by attenuation coefficient φ plus (virtual) reward $\pi_i[s_i(t), s_{-i}(t)]$, then normalized by the updated experience weight $N(t)$. In mathematical form:

$$A_i^j(t) = \frac{\phi \cdot N(t-1) \cdot A_i^j(t-1)}{N(t)} + \frac{\{\delta + (1-\delta) \cdot I[s_i^j, s_i(t)]\} \cdot \pi_i[s_i^j, s_{-i}(t)]}{N(t)}$$

where I[·] is the indicator function, which is defined as follows

$$I(x,y) = \begin{cases} 1, x = y \\ 0, x \neq y \end{cases}$$

The algorithm updates attraction value by reward after selecting corresponding strategy and weight coefficient δ multiplying virtual reward it would have yielded if select other strategy in hypothetical scenes. Parameter δ is used to measure the weight of virtual reward in relative to the real weight, and virtual reward can be understood as estimate and expectation of payoffs after selecting other alternative strategy.

The other is experience weight $N(t)$, which can be interpreted as equivalent observation of past experience with respect to present experience. The bigger the value of $N(t)$ is, the greater the influence of past experience to current attraction. The update rule of $N(t)$ is that the last experience weight $N(t-1)$ multiplied by attenuation coefficient ρ, then plus Incremental value of 1. That is

$$N(t) = \rho \cdot N(t-1) + 1$$

Denote $N(0)$ be the initial value of $N(t)$, while the initial value of $A_i^j(t)$ is $A_i^j(0)$, and $N(0)$ can be seen as equivalent assess of pregame thinking.

The value of attraction determines the probability of strategy selection. In other words, probability function of $P_i^j(t)$ should monotonously increase with $A_i^j(t)$. The mathematical expression of probability in exponential form is

$$P_i^j(t+1) = \frac{e^{\lambda \cdot A_i^j(t)}}{\sum_{k=1}^{m_i} e^{\lambda \cdot A_i^k(t)}}$$

Parameter λ above is used to measure the player's sensitivity to attraction value. The more sensitive the player to the attraction, the bigger its value is.

3. Channel Selection Model

In the problem of radio communication channel selection, different wireless channels should have different channel availabilities, that is, the channel idle probabilities should be in difference for cognitive radio. Assuming radio propagation environment can be divided into n channels, then the idle probability of channel i can be expressed as α_i, or $A = \{\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_{n-1}, \alpha_n\}$ in vector form. Let μ_i be the data transfer completion rate and γ_i be the data transfer request rate of channel i, then the vector form will be $M = \{\mu_1, \mu_2, \mu_3, \dots, \mu_{n-1}, \mu_n\}$ and $\Gamma = \{\gamma_1, \gamma_2, \gamma_3, \dots, \gamma_{n-1}, \gamma_n\}$. The relationship of parameters above meets the equation of $\alpha_i = \gamma_i/(\mu_i + \gamma_i)$. Think of the radio channel characteristics change over time, the channel idle probability and successful transmission probability of channel i should not be the same at different time t, then parameters above and their relations after introducing time parameter t are

$M(t) = \{\mu_1(t), \mu_2(t), \mu_3(t), \dots, \mu_{n-1}(t), \mu_n(t)\}$,

$\Gamma(t) = \{\gamma_1(t), \gamma_2(t), \gamma_3(t), \dots, \gamma_{n-1}(t), \gamma_n(t)\}$

and $\alpha_i(t) = \gamma_i(t)/[\mu_i(t) + \gamma_i(t)]$ respectively.

According to EWA, the update equation of experience weight $N(t)$ is

$$N(t) = \rho \cdot N(t-1) + 1$$

However, in order to accommodate the cognitive radio channel selection and transmission characteristics, the update mathematical expression of attraction $A_i^j(t)$ should be modified or improved as follows

$$A_i^j(t) = \frac{\phi \cdot N(t-1) \cdot A_i^j(t-1)}{N(t)} + \{\delta + (1-\delta) \cdot I[s_i^j, s_i(t)]\} \cdot \pi_i[s_i^j, s_{-i}(t)] \times \times \frac{\eta + (1-\eta) \cdot I[1, x(j)] \cdot I[s_i^j, s_i(t)] + (1-\eta) \cdot \{1 - I[s_i^j, s_i(t)]\}}{N(t)}$$

where

$$x(j) = \begin{cases} 0, \text{Transmission failure on channel j} \\ 1, \text{Successful transmission on channel j} \end{cases}$$

$$\pi_i[s_i^j, s_{-i}(t)] = \begin{cases} 0, \text{channel j is sensing busy} \\ 1, \text{channel j is sensing idle} \end{cases}$$

And the probability of channel selection $P_i^j(t)$ meets equation of

$$P_i^j(t+1) = \frac{e^{\lambda \cdot A_i^j(t)}}{\sum_{k=1}^{m_i} e^{\lambda \cdot A_i^k(t)}}$$

From the equation above, in the period of radio environment sensing of Cognitive Radio, when perceiving the current state of the channel j busy (strong Electromagnetic noise over interference threshold for transmission), the state flag status is set to 0(unavailable), and the strategy of selecting channel j for transmission channel will get no payoff, or the award function value of $\pi_i[s_i^j, s_{-i}(t)]$ is 0; while perceiving the current state of the channel j idle (electromagnetic noise below interference threshold for transmission), the state flag status is set to 1 (available), and the strategy of selecting channel j for transmission channel will get the payoff of $\pi_i[s_i^j, s_{-i}(t)]$ respectively. In addition, the value of $\pi_i[s_i^j, s_{-i}(t)]$ is assumed to equal 1.

These available channels are candidate channels for channel selection of Cognitive Radio, and the candidate channel with the highest probability of channel selection will be chosen for transmission. However, other candidate channels unselected will get the virtual award of $\delta \cdot \pi_i[s_i^j, s_{-i}(t)]$. Next reaches the innovation of this paper: a feedback mechanism of channel transmission is introduced and the update of selected channel attraction is divided in two phases. Before the transmission starts, selected channel gets some actual award of $\eta \cdot \pi_i[s_i^j, s_{-i}(t)]$ if feedback coefficient is set to $1-\eta$. After successful transmission, the selected channel will get another $(1-\eta) \cdot \pi_i[s_i^j, s_{-i}(t)]$ award as payoff. But if the transmission is unfortunately failure, the selected channel will not get the second part of the actual award.

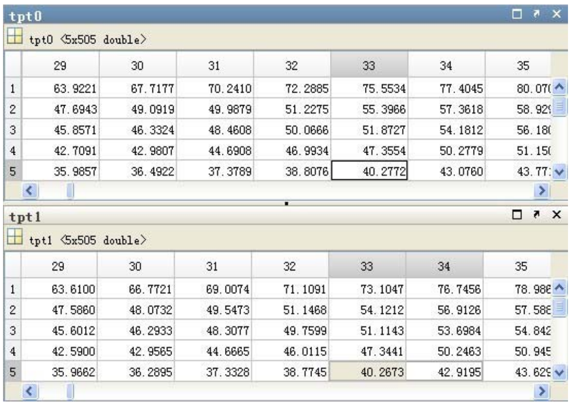
According to EWA intelligent channel handoff scheme, when communication interruption occurs, Cognitive Radio has the ability to resume remain data transmission by automatically switching current communication channel from busy to idle ones. If multi-free channels are detected at the breaking point, the channel with maximum idle probability computed by EWA algorithm will be chosen as the target channel for handoff (if more than one channel reach the highest idle probability, then random one within these channels will be selected); if no channel is available at present, then full channel blocking is defined. Cognitive Radio will wait with cycle free channel detection until the first available channel appears.

4. Results and Discussion

Assuming the number of channels in simulation environment is 5, or n=5. The initial value of attraction $A_i^j(0)$ and experience weight $N(0)$ are the same to 1, while the selection probability of each channel is the same in the initial state, that is, $P_i^j(0) = \frac{1}{n} = \frac{1}{5} = 0.2$. For the coefficients, ϕ for attraction, ρ for experience weight, and λ for sensitivity are the same to the default value 0.9 according to general experience. As normalized reward for transmission success is 1, while 0 for transmission failure (no reward), and probabilities of successful transmission and failure transmission are the same to 0.5 under the condition of no prior experience, then virtual award $\delta = 1 \times 0.5 + 0 \times \times 0.5 = 0.5$. Since the value of η for feedback should be lower than virtual award δ, we pick half value of virtual award δ for feedback η in this paper, that is, $\eta = \delta/2 = 0.25$. While there shall be some differences between each channel, the idle probabilities of these channels will not be the same. To reflect the general channels' available probabilities, uniform distribution vector in the range of 0 to 1 will be selected for the idle probability of each channel, that is, the initial channel idle probability vector $A_0 = \{0.1,0.3,0.5,0.7,0.9\}$, then the corresponding vector of channel data transfer completion rate $M_0 = \{1/0.1,1/0.3,1/0.5,1/0.7,1/0.9\}$ while the vector of channel data transfer request rate $\Gamma_0 = \{1/0.9,1/0.7,1/0.5,1/0.3,1/0.1\}$. In order to verify this intelligent algorithm capability to decide and guide cognitive radio real-time switch to the newly transmission channel with the highest available probability online accurately, the channel idle probability vector will change to $A_1 = \{0.9,0.7,0.5,0.3,0.1\}$, and the channel data transfer completion and request rate vector will be $M_1 = \{1/0.9,1/0.7,1/0.5,1/0.3,1/0.1\}$ and $\Gamma_1 = \{1/0.1,1/0.3,1/0.5,1/0.7,1/0.9\}$ respectively after 25 rounds during the simulation process. However, the channel data transfer completion rate of Cognitive Radio $\mu_s = 1/0.1$ and request rate $\gamma_s = 1/0.9$ will not change during the whole process of simulation. Taking suddenness and randomness of the aboving parameters under actual wireless environment into account, the value generated in each simulation round meets exponential distribution of the corresponding parameter above followed by the general rule.

After parameters above are set, time periods with channel busy and idle state will be firstly generated according to the channel data transfer completion rate μ_i and request rate γ_i. Time table of each channel state transition is shown in Fig. 1.

In the Fig. 1, table tpt1 records the time points when channel switches idle state to busy state while table tpt0 keeps the time points when channel changes from busy state to the idle one. Row number in the table indicates the channel number and column number shows the number of timeline marker. If table that holds time points of channel state transition in the timeline is fixed, then time periods with different channel states will be determined. For example, in the time table above, channel 5 will not be available in the time period between 40.2673 and 40.2772 and will be free from time point 40.2772 to 42.9195.



tpt0

tpt0 <5x505 double>

	29	30	31	32	33	34	35
1	63.9221	67.7177	70.2410	72.2885	75.5534	77.4045	80.07
2	47.6943	49.0919	49.9879	51.2275	55.3966	57.3618	58.92
3	45.8571	46.3324	48.4608	50.0666	51.8727	54.1812	56.18
4	42.7091	42.9807	44.6908	46.9934	47.3554	50.2779	51.15
5	35.9857	36.4922	37.3789	38.8076	40.2772	43.0760	43.77

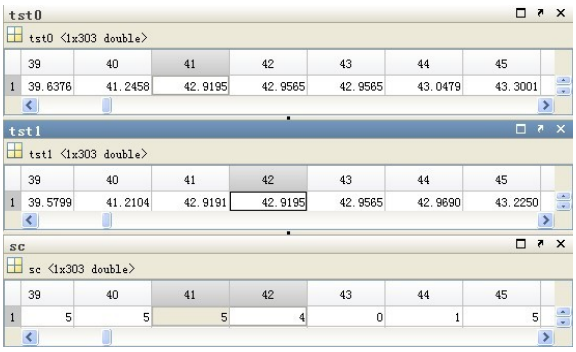
tpt1

tpt1 <5x505 double>

	29	30	31	32	33	34	35
1	63.6100	66.7721	69.0074	71.1091	73.1047	76.7456	78.98
2	47.5860	48.0732	49.5473	51.1468	54.1212	56.9126	57.58
3	45.6012	46.2933	48.3077	49.7599	51.1143	53.6984	54.84
4	42.5900	42.9565	44.6665	46.0115	47.3441	50.2463	50.94
5	35.9662	36.2895	37.3328	38.7745	40.2673	42.9195	43.62

Fig. 1. Time table of channel state transition.

After common radio channel environment above established, EWA learning algorithm with channel handoff scheme will be simulated online. The real-time simulation is illustrated in Fig. 2.



tst0

tst0 <1x303 double>

	39	40	41	42	43	44	45
1	39.6376	41.2458	42.9195	42.9565	42.9565	43.0479	43.3001

tst1

tst1 <1x303 double>

	39	40	41	42	43	44	45
1	39.5799	41.2104	42.9191	42.9195	42.9565	42.9690	43.2250

sc

sc <1x303 double>

	39	40	41	42	43	44	45
1	5	5	5	4	0	1	5

Fig. 2. Channel handoff and channel block.

From the Fig. 2, table sc represents the number of preferable channel to conduct data transmission, and the time shown in table tst1 channel data denotes the time point when transmission begins while transmission end in table tst0. The marks in figure record one channel handoff process: Cognitive Radio starts to transmit data from time point 42.9191 in channel 5. At the time point 42.9195 Cognitive Radio detects that channel 5 is not available for continuing data transmission, then withdraws the current channel passively, and accesses to idle channel 4 for present channel handoff. Note that channel number 0 in table sc indicates that full channel blocking occurs. EWA learning algorithm detects the whole channel blocking in 43th round. After waiting 0.0125 unit of time [tst1(44)-tst0(43)], Cognitive Radio returns transmitting data unfinished automatically in channel 1 which firstly becomes free from busy state.

After 100 rounds online simulation, the track records of channel selection probability based on EWA learning are shown in Fig. 3.

In the Fig. 3, EWA learning algorithm randomly selects channel 3 as the access channel in the condition of the same initial channel selection probabilities. After short initialization process, EWA learning algorithm can successfully track and lock channel 1 as its preferable channel. For the reason of channel availability probability changes after 25th round, the selection probability of channel 1 falls dramatically, while the selection probability of channel 5 increases respectively and steadily overtakes the selection probability of channel 1 after 38 rounds. Channel 5 eventually replaces Channel 1 to become optimal access channel under new channel available probability states.

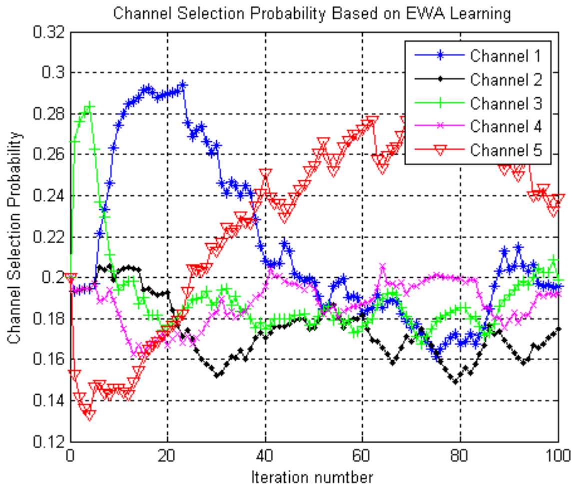


Fig. 3. Channel Selection Probability Based on EWA.

Channel experience weighted attraction based on EWA learning is depicted in Fig. 4. It can be seen that channel experience weighted attraction has a relatively uniform characteristics with the channel selection probability by comparing Fig. 3.

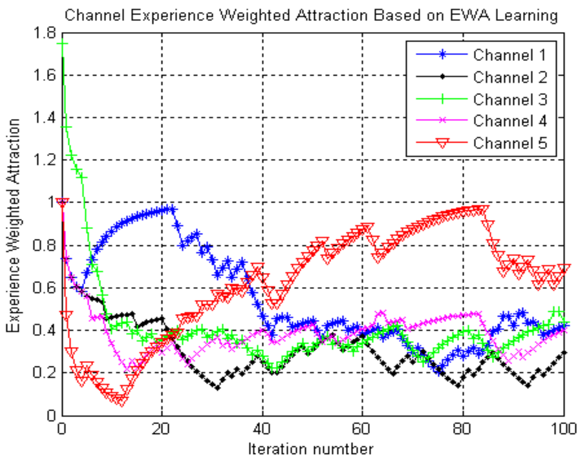


Fig. 4. Channel Experience Weighted Attraction Based on EWA Learning.

In order to highlight better performance of channel selection algorithm based on EWA learning than Q learning, the channel selection tracks based on both learning algorithms are recorded under the same initial states and radio environments. The results are illustrated in Fig. 5.

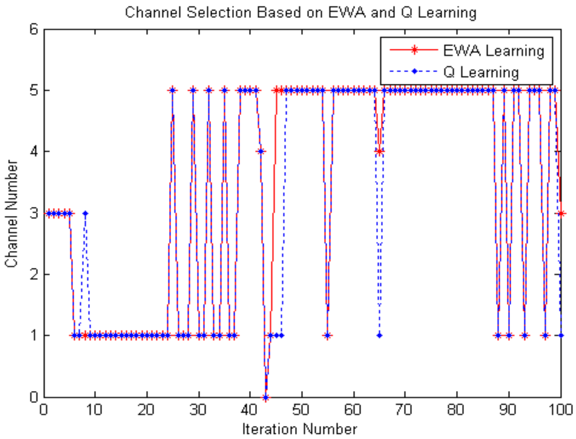
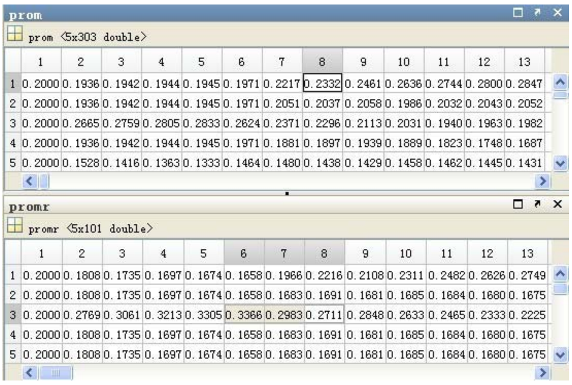


Fig. 5. The Tracks of Channel Selection Based on EWA Learning and Q Learning.

The first difference between EWA learning and Q learning algorithms appears in the 8th round. Why Q learning selects channel 3 as its preferable channel after continuously accessing channel 1 in round 6 and 7 arouses our big interest. In order to analyze the reason of this phenomenon, the channel selection probability records after each round are derived and shown in Fig. 6.



prom

prom <5x303 double>

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.2000	0.1936	0.1942	0.1944	0.1945	0.1971	0.2217	0.2332	0.2461	0.2636	0.2744	0.2800	0.2847
2	0.2000	0.1936	0.1942	0.1944	0.1945	0.1971	0.2051	0.2037	0.2058	0.1986	0.2032	0.2043	0.2052
3	0.2000	0.2665	0.2759	0.2805	0.2833	0.2624	0.2371	0.2296	0.2113	0.2031	0.1940	0.1963	0.1982
4	0.2000	0.1936	0.1942	0.1944	0.1945	0.1971	0.1881	0.1897	0.1939	0.1889	0.1823	0.1748	0.1687
5	0.2000	0.1528	0.1416	0.1363	0.1333	0.1464	0.1480	0.1438	0.1429	0.1458	0.1462	0.1445	0.1431

promr

promr <5x101 double>

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.2000	0.1808	0.1735	0.1697	0.1674	0.1658	0.1966	0.2216	0.2108	0.2311	0.2482	0.2626	0.2749
2	0.2000	0.1808	0.1735	0.1697	0.1674	0.1658	0.1683	0.1691	0.1681	0.1685	0.1684	0.1680	0.1675
3	0.2000	0.2769	0.3061	0.3213	0.3305	0.3366	0.2983	0.2711	0.2848	0.2633	0.2465	0.2333	0.2225
4	0.2000	0.1808	0.1735	0.1697	0.1674	0.1658	0.1683	0.1691	0.1681	0.1685	0.1684	0.1680	0.1675
5	0.2000	0.1808	0.1735	0.1697	0.1674	0.1658	0.1683	0.1691	0.1681	0.1685	0.1684	0.1680	0.1675

Fig. 6. The Probability table of Channel Selection.

Table prom records channel selection probabilities values calculated by EWA algorithm after each round, while Table presents channel selection probabilities values based on Q learning. It can be seen from table prom that channel 1 selection probability reaches maximum value to 0.2332 after 8th round. Although channel 3 selection probability shows a downward trend in a row after 6th and 7th two rounds, its value still in the highest of 0.2771 after 8th round in the table promr. At the same time, channel 3 status turns available after 8th round from previous busy state. As can be concluded from analysis, it is the inevitable for Q learning to reselect channel 3 as its preferable channel accordance with algorithm calculation. It also can directly reflects differences and advantages of EWA comparing Q learning: channel selection based on EWA learning algorithm has better and more stable performance in fast tracking, locking and switching to the current optimal channel from changing communication environment. And the reasons of channel selection differences after 43th round of full channel blocking can also be explained by analysis above.

What is more, Statistical parameter channel handoff times will reflect EWA algorithm's more advantageous performance visually in comparison with Q learning. The statistics is presented in Fig. 7.

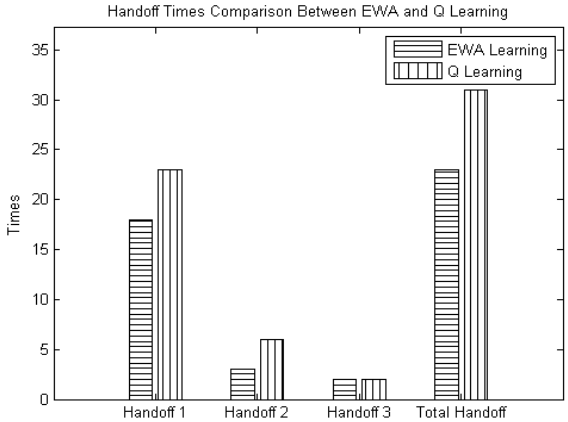


Fig. 7. Channel Handoff Times Comparison between EWA Learning and Q Learning.

From the Fig. 7, Handoff 1 shows that there is one channel handoff in a whole communication process, while Handoff 2 indicates two consecutive channel handoffs before data transmission completion and Handoff 3 indicates three consecutive channel handoffs or more. Through this comparison it can be seen, the number of Handoff 1 with Q learning reaches 23, but only 18 times with EWA algorithm. Similarly, the number of Handoff 2 with Q learning is 6, while there are merely 3 times in EWA algorithm. Though the number of Handoff 3 is the same of 2, there is obvious superiority in total handoffs, 31 to 23. By evident statistical comparison, channel selection algorithm based on EWA learning has less channel handoff times than Q learning in the same initial states and radio environments.

5. Conclusions

In this paper, an innovative channel selection algorithm based on EWA learning with channel handoff scheme is proposed and a feedback mechanism of channel transmission is also introduced to facilitate Cognitive Radio to learn radio environment communication channel characteristics online. By accumulating the history channel experience, it can predict, select optimal communication channel, and be capable to decide and guide cognitive radio real-time switch to the newly transmission channel with the highest available probability online accurately, dynamic ensure the quality of communication links and finally reduce system communication outage probability. Compared with Q learning, channel selection algorithm based on EWA learning is more sensitive and agile in tracking & accessing to new optimal channel, and more stable in locking this channel, which brings the results of less conflicts with primary users and less passive channel handoff occurrences.

Acknowledgements

This work was supported by the National High Technology Research and Development Program of China (863 Program) (No. 2012AA062103).

References

[1]. Mitola, J., Cognitive radio for flexible mobile multimedia communications, *Mobile Networks & Applications*, Vol. 6, Issue 5, 2001, pp. 435-441.

[2]. Hossain, M. S., et al., Hard Combination Data Fusion for Cooperative Spectrum Sensing in Cognitive Radio, *International Journal of Electrical and Computer Engineering (IJECE),* Vol. 2, Issue 6, 2012, pp. 811-818.

[3]. Zhang, D., et al., An Improved Cognitive Radio Spectrum Sensing Algorithm, *TELKOMNIKA Indonesian Journal of Electrical Engineering*, Vol. 11, Issue 2, 2013, pp. 583-590.

[4]. Khalaf, G., An Optimal Sinsing Algorithm for Multiband Cognitive Radio Network, *International Journal of Information and Network Security (IJINS),* Vol. 2, Issue 1, 2013, pp. 60-67.

[5]. Shan-Shan, W., et al., Primary User Emulation Attacks Analysis for Cognitive Radio Networks Communication, *TELKOMNIKA Indonesian Journal of Electrical Engineering*, Vol. 11, Issue 7, 2013, pp. 3905-3914.

[6]. Bantouna, A., et al., An overview of learning mechanisms for cognitive systems, *Eurasip Journal on Wireless Communications and Networking*, 2012, pp. 1-6.

[7]. Tsagkaris, K., A. Katidiotis, and P. Demestichas, Neural network-based learning schemes for cognitive radio systems, *Computer Communications*, Vol. 31, Issue 14, 2008, pp. 3394-3404.

[8]. Galindo-Serrano, A. and L. Giupponi, Distributed Q-Learning for Aggregated Interference Control in Cognitive Radio Networks, *IEEE Transactions on Vehicular Technology*, Vol. 59, Issue 4, 2010, pp. 1823-1834.

[9]. Li, H. S., Multiagent Q-Learning for Aloha-Like Spectrum Access in Cognitive Radio Systems, *Eurasip Journal on Wireless Communications and Networking*, 2010, pp. 1-15.

[10]. Chen, X. F., et al., Reinforcement Learning Enhanced Iterative Power Allocation in Stochastic Cognitive Wireless Mesh Networks, *Wireless Personal Communications*, Vol. 57, Issue 1, 2011, pp. 89-104.

[11]. Zhang, W. Z. and X. C. Liu, Centralized Dynamic Spectrum Allocation in Cognitive Radio Networks Based on Fuzzy Logic and Q-Learning. *China Communications*, Vol. 8, Issue 7, 2011, pp. 46-54.

[12]. Gallego, J. R., M. Canales, and J. Ortin, Distributed resource allocation in cognitive radio networks with a game learning approach to improve aggregate system capacity, *Ad Hoc Networks*, Vol. 10, Issue 6, 2012, pp. 1076-1089.

[13]. Zhu, J., et al., Adaptive transmission scheduling over fading channels for energy-efficient cognitive radio networks by reinforcement learning, *Telecommunication Systems*, Vol. 42, Issue 1-2, 2009, pp. 123-138.

[14]. Torkestani, J. A. and M. R. Meybodi, A Learning Automata-Based Cognitive Radio for Clustered Wireless Ad-Hoc Networks, *Journal of Network and Systems Management*, Vol. 19, Issue 2, 2011, pp. 278-297.

[15]. Yang, M. F. and D. Grace, Cognitive Radio with Reinforcement Learning Applied to Multicast Downlink Transmission with Power Adjustment, *Wireless Personal Communications*, Vol. 57, Issue 1, 2011, pp. 73-87.

[16]. Zhou, P., Y. S. Chang, and J. A. Copeland, Reinforcement Learning for Repeated Power Control Game in Cognitive Radio Networks, *IEEE Journal on Selected Areas in Communications*, Vol. 30, Issue 1, 2012, pp. 54-69.

[17]. Sun, Y., Qian, J. S., Cognitive Radio Channel Selection Strategy Based on Experience-Weighted Attraction Learning, *TELKOMNIKA Indonesian Journal of Electrical Engineering*, Vol. 12, Issue 1, 2014, pp. 149-156.