International Journal of Smart Home
Vol.8, No.3 (2014), pp.151-162
http://dx.doi.org/10.14257/ijsh.2014.8.3.14

Multi-armed Bandit Online Learning Based on POMDP in Cognitive Radio

Juan Zhang[1], Hesong-Jiang[1] ,Hong Jiang [1] Chunmei Chen[1]

[1]*the Open Fund of Robot Technology Used for Special Environment Key Laboratory of Sichuan Province , School of Information Engineering, Southwest University of Science and Technology, Mianyang, China*

[1]*zhangjuan@swust.edu.cn,* [2] *jianghesong@swust.edu.cn,* [3] *jianghong@swust.edu.cn,* [3] *chenchunmei@swust.edu.cn*

Abstract

In cognitive radio, most of existing research efforts devoted to spectrum sharing have two weakness as follows. First, they are largely formulated as a Markov decision process (MDP), which requires a complete knowledge of channel. Second, most of the studies are online learning based on perceived channel. To solve the above problems, a new algorithm is proposed in this paper: if the authorized user exists in the current channel, Second user will send conservatively in low rate, or send aggressively. When sending conservatively, the state of the channel is not directly observable, the problem turns out to be Partially Observable Markov Decision Process (POMDP).We first establish the optimal threshold when the channel is known, then consider the optimal transmission when the channel is unknown and model for multi-armed bandit. We get the optimal K-conservative policy through the UCB algorithm and improve the convergence speed by UCB-TUNED algorithm. Simulation and analysis results show that it is the same result of K-conservative policy no matter the multi-armed bandit online learning under not fully known channel or the optimal threshold policy under known channel .At the same time, we improve the convergence speed by UCB-TUNED algorithm.

Keywords: *spectrum sharing, multi-armed bandit, online learning, Partially Observable Markov Decision Process*

1. Introduction

In recent years, the wireless device (smart mobile phone and tablet computer) popularity led to a sharp increase in demand for more bandwidth, and spectrum resources become fewer and fewer available for distribution, which caused the spectrum resources nervous, but on the other hand, utilization rate of wireless spectrum is very low. According to [1], the spectrum utilization rate of more than 90% allocated is serious shortage. Dynamic spectrum access technology solves a lot of contradictions between the spectrum utilization and spectrum scarcity. The most promising realization in dynamic spectrum is the cognitive radio (CR). Spectrum sharing is the key technology of the effective use of idle frequency band to improve the spectrum utilization in cognitive radio system.

At present, domestic and foreign researchers have proposed a variety of spectrum sharing model, [2, 3] researched the heuristic algorithm based on graph coloring and biology. [4, 5]

ISSN: 1975-4094 IJSH

proposed respectively based on the auction mechanism of economics and game theory. [6] put forward spectrum sharing model of cross layer optimization. These models allow unauthorized users share the authorized users of the band without causing harmful interference to licensed users, but not to analyze the configuration parameters of the channel model in the process of transmission. [7] considered GE attenuation to minimize the transmission channel capacity and delay and modeled as partially observable markov decision process (POMDP) by single threshold strategy for the analysis of various parameters. [8] considered the non-Bayesian perception problems with unknown parameters, and achieved approximate logarithmic regret value through online learning, but did not consider the optimal transmission of different channel condition.

According to the above problem, this paper proposes online learning scheme of the optimal transmission based on unknown Gilbert-Elliott channel: modeling network channel based on POMDP and K arm bandit algorithm can be converted to k conservative strategy. UCB and UCB-Tuned algorithm are adopted to realize and optimize.

2. The System Model

Assumptions in the authorized user network, each channel S with just two states, namely binary Gilbert - Elliott markov chain: as shown in Figure 1, when S=1, indicates that the current channel is free; if S=0, indicates that the current state is busy. λ0 is the transition probability from busy to idle state , (1 - λ1) is the state transition probability from idle to busy.

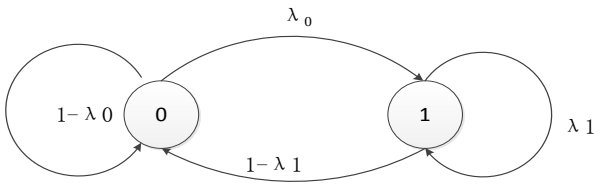

Figure 1. G-E Channel Model

2.1. POMDP Model

In POMDP, unauthorized users (SU) make use of the existing part of the information, history action and immediately return values to make decisions. As shown in Figure 2 for POMDP model [10], b is a belief state and a probability distribution of all status in set S. SU's probability is b (s) in state *s*, and $\Sigma_{s\in S}\, b(s)=1.0$ .All possible belief state constitutes the belief space representation of $B(S) = \{b : \Sigma_{s \in S}\, b(s) = 1.0, \forall s, b(s) \geq 0\}$ .According to [9], belief state is a sufficient statistic of the optimal operation strategy for A*.

The model is described as:1) the state estimator (SE) : P×A×B(S)→B(S), where P is the confidence probability, namely the state estimator (SE) is responsible for update the current belief b according to the last action, belief state and the current observation. 2)the strategy π:B(S)→A, namely using the strategy π to choose the action a in the current belief state b,and the return of R (B, a),which is expressed as $r(b, s) = \sum_{s\in S} b(s) r(s, a)$.

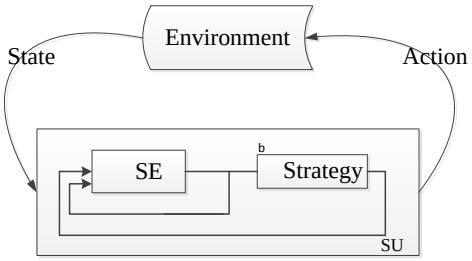

Figure 2. POMDP Model

2.2. Channel Modeling based on POMDP

The assumption that the current channel is Gilbert-Elliott channel which has a two value Markov chain state: when the S=1, the current channel is in the idle, namely SU channel condition is good, in which the data can be transmitted successfully in high speed; while S=0, it indicates that the current channel is busy, namely SU channel state is poor, only in which SU can transmit successfully the data at a lower rate .Transfer probability is expressed as:

$$P = \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix} = \begin{bmatrix} 1-\lambda_0 & \lambda_0 \\ 1-\lambda_1 & \lambda_1 \end{bmatrix} \quad (1)$$

where $\alpha=\lambda_1-\lambda_0$,assuming that the channel is positively related, then $\alpha>0$.

At the beginning of each time slot, SU needs to make the action selection:

(1)Send Conservatively (SC): SU transmits data at low speed. In this action, no matter what is the state of the current channel, the SU data transmission could be achieved, and returns R1. Therefore, SU cannot learn the channel state in the action.

(2) Sending aggressively (SA): SU transmits data at high speed. If the channel condition is good, SU high-speed data transmission is successful, and gets the return R2, and R2>R1; if the channel condition is poor, high speed data transmission will cause greatly the error rate and packet loss rate, and get the penalty value C. Therefore, in the action of SU can obtain the next state through learning.

When sending conservatively, state of the channel cannot be directly observed, so the problem is modeled as a POMDP model. The POMDP model is the conditional probability of the good channel state which is given in all the case of historical observation and action and expressed as b = Pr [St = 1 | Ht], where Ht is all historical actions and observations information before the first t slot. When sending aggressively, SU can learn the channel state. Therefore, when the channel state is good, the belief is λ1; the belief is λ0 when the channel state is poor. Expected return is expressed as:

$$R(b_t, A_t) = \begin{cases} R_1 & A_t = SC \\ b_t R_2 - (1-b_t)C & A_t = SA \end{cases} \quad (2)$$

where b_t is the belief in good state and A_t is the action at time t.

3. The known Channel State of the Optimal Transmission Threshold Strategy

The most typical MAB question is: a gambling machine has the K arms, from which the gambler chooses to operate an arm to get reward, which is from the relevant distribution of the arm, and gambling does not know the reward expectations size of each distribution arm. In a specific period of time, gambling can only operate one arm; gamblers find the greatest reward of the arm as soon as possible, and gambling.

3.1. K Conservative Strategy Structure Modeling

In this section, we will discuss the K-conservative policy, where K is the number of time slots to send conservatively after a failure, before sending aggressively again is as shown in Figure 3.

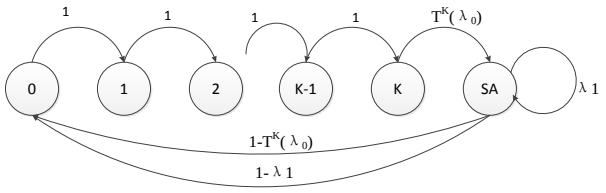

Figure 3. K Conservative Policy Markov Chain

The states are the number of time slots which the transmitter has sent conservatively since last failure. There are K+ 2 states in this Markov chain. State 0 corresponds to the moment that sending aggressively fails and it goes back to sending conservatively stage. State K-1 corresponds to that the transmitter has already sent conservatively for K time slots, and it will send aggressively next time slot. If the transmitter sends aggressively and succeeds, it goes to state SA and continues to send aggressively at the next time slot; otherwise it goes back to state 0. The probability that transmitter stays in state SA is λ 1. The transmitter has to wait K time slots before sending aggressively again, so the probabilities from state i to state i+1 is always 1 when $0 \leq i < k$.

There is K+2 states, each state corresponds to a belief and an action; belief and action determine the expected total-discounted reward. Thus given K, there are K+2 different expected total-discounted rewards.

3.2. The Challenge of the K Conservative Strategy

In this section, we will discuss how to find the optimal policy if the underlying channel's transition probabilities are unknown. To find the optimal K, we use the idea of mapping each K-conservative policy to a countable multi-armed bandit of countable time horizon. Now there are two challenges: (1).The number of arms can be infinite. (2). To get the true total discounted reward, each arm requires to be continuing played until time goes to infinity. To address these two challenges, we weaken our objective to find a suboptimal which is an(OPT-ε-δ) approximation of the optimal arm instead. Theorem 1 and theorem 2 address the two challenges respectively.

Theorem 1: Given an ε and bound B on α, there exists $\forall$ K$\geq$ Kmax, the best arm in the arm set C={0,1,…,K,SC} is an (OPT-ε) arm.

Proof: If K>Kopt, the optimal arm is already included in the arm sets.

If Kopt = ∞,Let arm SC correspond to the always sending conservatively policy。

If K<Kopt<∞,suppose that transmitter has already sent conservatively for n time slots,

$$V^{\pi_{Kopt}}(p) - V^{\pi_k}(p)$$
$$= \left[R_1 \frac{1-\beta^{Kopt}}{1-\beta} + \beta^{Kopt} V_{SA}(T^{Kopt}(p)) \right] - \left[R_1 \frac{1-\beta^{K}}{1-\beta} + \beta^{K} V_{SA}(T^{K}(p)) \right]$$
$$= \beta^{K} \frac{R_1}{1-\beta}(1-\beta^{Kopt-K}) + \beta^{Kopt} V_{SA}(T^{Kopt}(p)) - \beta^{K} V_{SA}(T^{K}(p))$$

when $T(\rho)>\rho$, $V(T(\rho))=V_{SA}(T(\rho))$;

when $T(\rho)\leq\rho$, $V(T(\rho)) = R_1/(1-\beta)$;

$V_{SA}(T(\rho)) > R_1/(1-\beta)$;

$$V^{\pi_{Kopt}}(p) - V^{\pi_k}(p) < \beta^{K} \left[V_{SA}(T^{Kopt}(p)) - V_{SA}(T^{K}(p)) \right]$$
$$= \beta^{K}(T^{Kopt}(p) - T^{K}(p))(R_2 + C + \beta(V(\lambda_1) - V(\lambda_0)))$$

$p=\lambda 0, C'=R2+C+\beta(V(\lambda 1)- V(\lambda 0))$,

$$V^{\pi_{Kopt}}(p) - V^{\pi_k}(p) < \beta^{K}(T^{Kopt}(\lambda_0) - T^{K}(\lambda_0))(R_2 + C + \beta(V(\lambda_1) - V(\lambda_0)))$$

$\alpha<B, T^{n}(\lambda_0) = T(T^{n-1}(\lambda_0)) = \lambda_0 \frac{1-\alpha^{n+1}}{1-\alpha}$, $\lambda_s=\lambda_0/(1-\alpha)$

$$V^{\pi_{Kopt}}(p) - V^{\pi_k}(p) < \beta^{K}(T^{Kopt}(\lambda_0) - T^{K}(\lambda_0))C' < B^{K+1}C' = \varepsilon$$

when $K \geq \log_B \frac{\varepsilon}{C'} - 1$, $V^{\pi_{Kopt}}(p) - V^{\pi_k}(p) < \varepsilon$

Theorem 2: Given a δ, there exists $\forall$ T≥ Tmax,an arm for the finite horizon total discounted reward up to time T is at most δ away from the infinite horizon total discounted reward.

Proof

$$E_\pi\left[\sum_{t=0}^{\infty} \beta^t R(b_t, A_t) \mid b_0 = p\right] - E_\pi\left[\sum_{t=0}^{T_{max}} \beta^t R(b_t, A_t) \mid b_0 = p\right] = E_\pi\left[\sum_{t=T_{max}+1}^{\infty} \beta^t R(b_t, A_t) \mid b_0 = p\right]$$

when $R(b_t, A_t) \leq R_2$, $\sum_{t=T_{max}+1}^{\infty} \beta^t = \frac{\beta^{T_{max}+1}}{1-\beta}$

$$E_\pi\left[\sum_{t=0}^{\infty} \beta^t R(b_t, A_t) \mid b_0 = p\right] - E_\pi\left[\sum_{t=0}^{T_{max}} \beta^t R(b_t, A_t) \mid b_0 = p\right] \leq \frac{\beta^{T_{max}+1}}{1-\beta} R_2$$

when $T \geq \log_\beta \frac{\delta(1-\beta)}{R_2} - 1$

$$E_{\pi}\left[\sum_{t=0}^{\infty}\beta^{t}R(b_{t},A_{t})\mid b_{0}=p\right]-E_{\pi}\left[\sum_{t=0}^{T_{\max}}\beta^{t}R(b_{t},A_{t})\mid b_{0}=p\right]<\delta$$

3.3. UCB Algorithm

UCB (Upper Confidence Bound) algorithm is a class of algorithms to solve the MAB collectively, based on currently available information, and tries to strike a balance between the exploitation and exploration with an adjustment of the value.

Generally speaking, the arm will be chosen according to the largest UCB value, which is be based on the average income of each arm of the current value (*i.e.*, its performance to date), plus an additional parameter, which will be relatively reduced with increasing the number of each arm selected. UCB is expressed by the formula:

$$\bar{X}_{i}+\sqrt{\frac{2\ln(n)}{n_{i}}}, \qquad (3)$$

$$\bar{X}_{i}=\frac{(1-\beta)\bar{A}_{i}+C}{R_{2}+C}, \qquad (4)$$

where $\bar{X}_{i}$ is the i arm average earnings so far, n_i is the number of times which the i arm is tested, N is the total number of all arms currently being tested. Let formula (3) of the value of the largest arm will be selected the next arm. Prior to entry is the past performance of the arm, namely the exploitation; after the item is to adjust the parameters, namely exploration.

While the UCB-TUNED compared with UCB experiment is better allocation strategy. The UCB-TUNED formula is as follows:

$$V_{j}(s)=\left(\frac{1}{s}\sum_{\gamma=1}^{s}\bar{X}_{j,\gamma}^{2}\right)-\bar{X}_{j,s}^{2}+\sqrt{\frac{2\log n}{s}} \qquad (5)$$

$$\bar{X}_{i}+\sqrt{\frac{\log n}{n_{i}}\min\left\{\frac{1}{4},V_{i}(n_{i}))\right\}} \qquad (6)$$

$$V_{i}(s)=\left(\frac{1}{n_{i}}\sum_{\gamma=1}^{n_{i}}\left(\frac{(1-\beta)\bar{A}_{i,\gamma}+C}{R_{2}+C}\right)^{2}\right)-\left(\frac{(1-\beta)\bar{A}_{i}+C}{R_{2}+C}\right)^{2}+\sqrt{\frac{\log n}{n_{i}}\min\left\{\frac{1}{4},V_{i}(n_{i}))\right\}} \qquad (7)$$

4. Simulation Results

This section compares the two methods of optimal transmission: one is offline algorithm of the optimal transmission threshold strategy put forward [11], and the other is online learning algorithm of K arm machine.

4.1. The off-line Algorithm for Optimal Transmission Threshold Strategy

According to [11], the simulation environment of the single threshold is established as follows:

Parameter settings:

Table 1 shows the configuration parameters used in simulation of threshold structure of the optimal policy. We assume that the channel is positive correlation, which means $\lambda 1 \geq \lambda 0$. The value of $\lambda 1$ is shown in Table 1.

Table 1. Parameters Setting of the Optimal Threshold Strategy

Type	Value	Type	Value
$\lambda 0$	0.01:0.05:0.9	C	0.5
$\lambda 1$	$\lambda 0$(1):0.05:0.99	β	0.75
R1	1	n	1:10000
R2	2	Kopt	0,1,2,3,4

Algorithm steps:

Step 1: Initialization parameters R1,R2,C, β ,λ_0,λ_1,Kopt;

Step 2: Definition of Karray, the purpose is to store λ 0,λ 1,ρ and Kopt, which meet the conditions.

Step 3: for ii=1:length(λ 0)

for jj=1:length(λ 1)

when n=1:10000,Calculation V(λ_0)

when n=1:10000,Calculation V(λ_0)

when Tk-1(λ0)< ρ , Tk-1(λ0)$\geq$ ρ,Calculation Kopt;

end for

end for

Simulation analysis:

According to the above algorithm steps, get Figure 4 and Table 2.

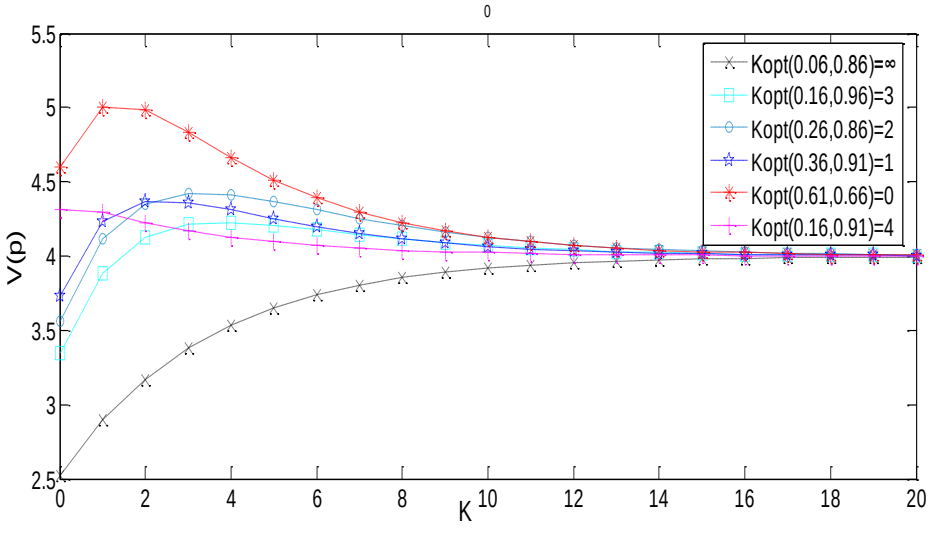

Figure 4. The Expected Discount Total Return of the Threshold Optimal Strategy

Table 2. The Optimal Strategy of the Threshold Structure

λ0	λ1	ρ	Kopt
0.01	0.06	0.6000	∞
0.16	0.91	0.4553	4
0.16	0.96	0.4597	3
0.26	0.86	0.5060	2
0.36	0.91	0.5446	1
0.61	0.66	0.6023	0

Figure 4 and Table 2 can get the following conclusion:

- when λ 0=0.01, λ 1=0.06, $n \to \infty$, $T^n(\lambda_0) \to \lambda_s$, so always keep to send, namely $K_{opt} \to \infty$;
- when λ 0=0.61, λ 1=0.66, says the channel condition is good, and always sends aggressively, $K_{opt} = 0$;
- when λ 0=0.16, λ 1=0.91, $K_{opt} = 4$, says sending aggressively after sending conservatively 4 slots, and the total discount reward can obtain the maximum.
- The corresponding optimal K step can be offline obtained under different channel state through single threshold.

4.2. Online Learning Algorithm of K Arm Gambling Machine in the Unknown Channel State

The learning algorithm of online K arm gambling machine is proposed. The specific simulation environment setting is shown in Table 1.

Parameter settings:

Considering the convergence of the algorithm, so the total running time slot is set to T*inter=10^9.

Table 3. Parameter Settings of Online Learning Algorithm of K Arm Gambling Machine

Type	Value	Type	Value
T	10^7	The total number of the tested arm: armnu	30
Slot number of each arm running at least: TMAX	100	the slot number of each arm: NI	zeros(armno, 1)
he boundary value of α:B	0.8	inter	inter=100,
ε	0.02	δ	0.02

Algorithm steps:

Step 1: Initialization parameters λ0、λ1,T,TMAX,armnu,ts,NI;

Step 1:The algorithm is online learning method based on the unknown channel state of POMDP model, so each arm gets a reward or punishment in the generation of action according to the observed state

Step 1: Initialize each arm of the UCB value;

Step 1: **for** kk=1: inter do

for ts=1:T-TMAX

Select the maximum UCB value or UCB-TUNED value as the optimal arm, and run the current optimal arm by $UCB = \frac{(1-\beta)\bar{A}_i + C}{R_2 + C} + \sqrt{\frac{2\ln(n)}{n_i}}$ and

$$UCB - Tuned = \left\{ \frac{1}{n_i} \sum_{\gamma=1}^{n_i} \left(\frac{(1-\beta)\bar{A}_{i,\gamma} + C}{R_2 + C} \right)^2 \right\} - \left\{ \frac{(1-\beta)\bar{A}_i + C}{R_2 + C} \right\}^2 + \sqrt{\frac{\log n}{n_i} \min\left\{ \frac{1}{4}, V_i(n_i) \right\}}$$

end for

end for

Simulation Analysis:

According to the above algorithm steps draw diagrams 5-8:

As shown in Figure 5 for the UCB algorithm, the performance of all arms with λ_0 =0.36 and λ_1=0.91 channel state. The arm 1 is the optimal arm of the channel condition, and arm 1 is selected to tend to 1 with increasing operating time, while using the other arm is selected to tend to 1the rate tends to 0. The same method can be obtained other λ0 and λ1 corresponding optimal arm.

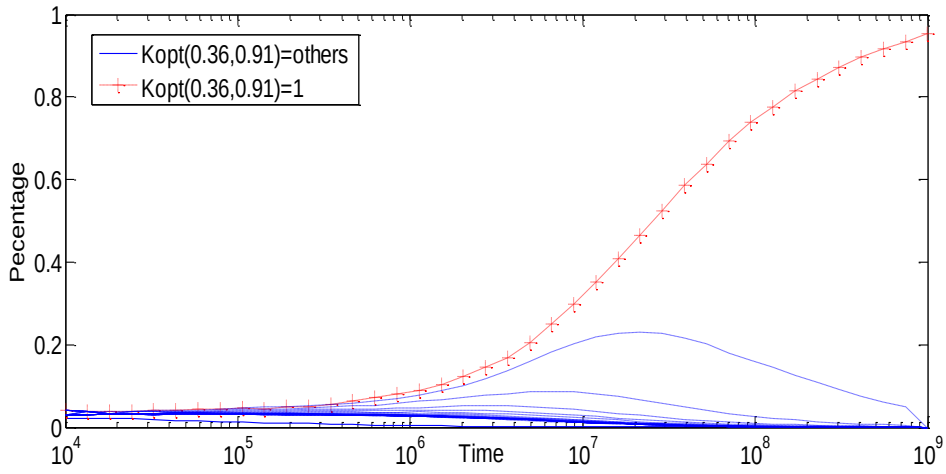

Figure 5. The Optimal Arm in the Same Channel State

Figure 6 shows the convergence of the corresponding optimal arm to get different λ_0 and λ_1 channel state, seen from the diagram, as time increases, the optimal arm was selected to run longer than gradually tends to 1.

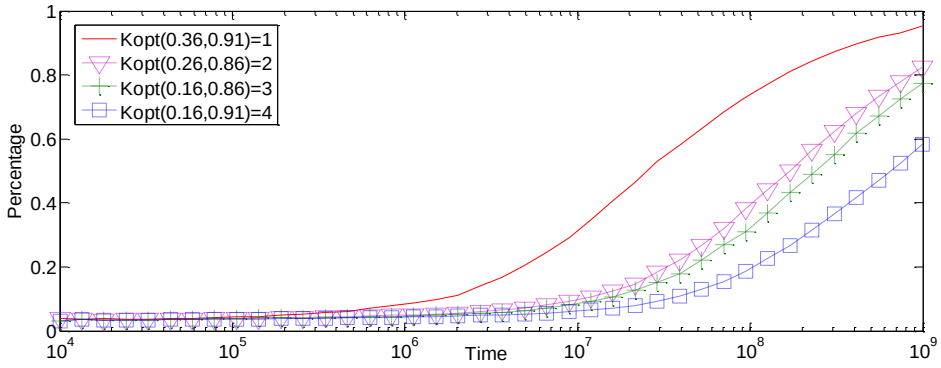

Figure 6. The Optimal Arm in the Different Channel State

Figure 7 shows the convergence speed of all the arm is faster compared with Figure 5 UCB algorithms by UCB - turned algorithm in the sameλ_0 andλ_1 channel condition.

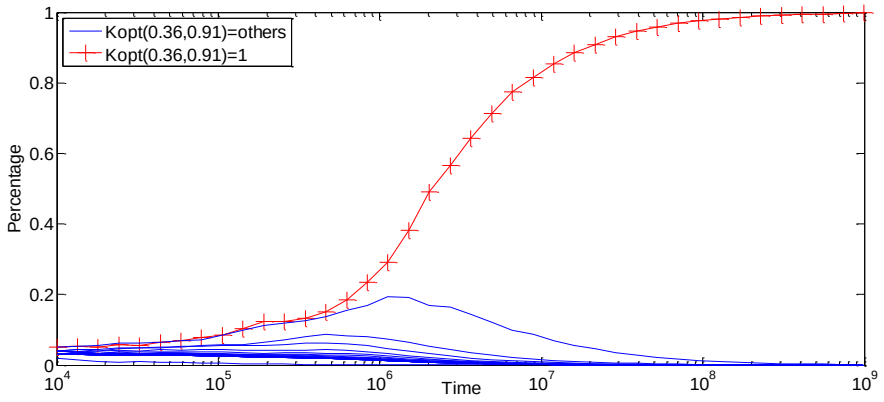

Figure 7. The Optimal Arm in the Same Channel State by UCB-TUNED

Figure 8 shows the convergence speed of all the arm is faster compared with Figure 6 UCB algorithm by UCB - turned algorithm in the differentλ_0 andλ_1 channel condition.

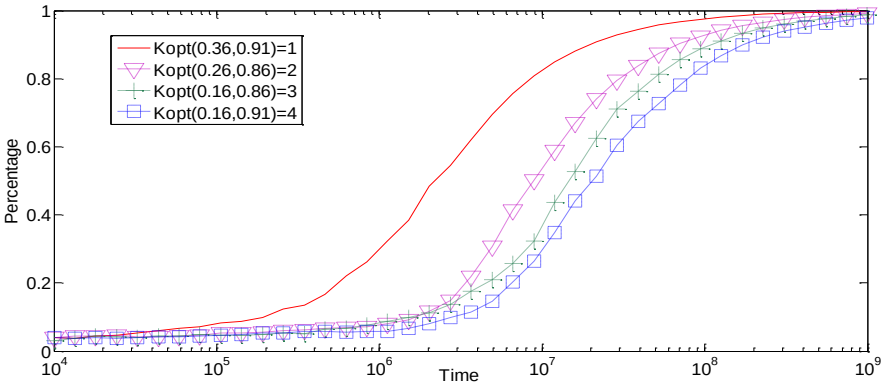

Figure 8. The Optimal Arm in the Different Channel State by UCB-TUNED

5. Conclusions

The optimum transmission is mostly based on full knowledge of channel modeling. Aiming at the cognitive radio environment is not entirely the case, the channel is modeled as a partially observable Markov process, and the online learning method of optimum transmission is based on the multi armed bandit. Simulation results show that, when the channel is not completely known case, the online learning algorithm of multi-armed bandit can get the optimal K strategy. At the same time, this paper improves the convergence of K step conservative strategy optimal transmission by UCB-TUNED method.4.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 61379005), and Southwest University of Science and Technology (12zx7127).

References

[1] 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, **(2012)** February, pp. 5-23.
[2] L. Tan, Z. Feng, W. Li, Z. Jing and T. A. Gulliver, "Graph coloring based spectrum allocation for femtocell downlink interference mitigation", IEEE Wireless Communications and Networking Conference QUINTANA-ROO MEXICO, **(2011)** March 28-31, pp. 1248-1252.
[3] Z. Q. He, K. Niu and T. Qiu, "A bio-inspired approach for cognitive radio networks", Chin Sci Bull, vol. 57, **(2012)**, pp. 3723-3730.
[4] W. Zhang, R. K. Mallik and K. B. Letaief, "Cooperative spectrum sensing optimization in cognitive radio networks", Proc. IEEE International Conference on Communications (ICC 2008), Beijing, China, **(2008)** May 19-23, pp. 3411-3415.
[5] Q. Jing and Z. Zheng, "Dynamic Spectrum Sharing Strategy in Cognitive Radio Systems", Journal of Beijing University of Posts and Telecommunications, vol. 32, no. 1, **(2009)**, pp. 69-72.
[6] W. Chun-de, P. Zhi-wen and Y. Xiao-hu, "An Optimal Cross-layer Spectrum Sharing Scheme for Cognitive Radio Based Ad hoc Network", Journal of Nanjing University of Posts and Telecommunications, vol. 29, no. 3, pp. 83-87, **(2009)**.
[7] L. A. Johnston and V. Krishnamurthy, "Opportunistic file transfer over a fading channel: A POMDP search theory formulation with optimal threshold policies", IEEE Transactions on Wireless Communications, vol. 5, no. 2, **(2006)**, pp. 394-405.
[8] W. Dai, Y. Gai, B. Krishnamachari and Q. Zhao, "The 36th International Conference on Acoustics", Speech and Signal Processing (ICASSP 2011), Prague, Czech Republic, **(2013)** May 22-27, pp. 2940.
[9] Q. Zhao, T. Lang and C. Y. X. Ananthram, IEEE J. Sel. Areas Commun.25 589, **(2007)**.
[10] J. Hong, L. Cong-Bin and W. Chun, "Crosslayer parameter configuration for TCP throughput improvement in cognitive radio networks", Acta Phys. Sin, vol. 62, no. 3, **(2013)**, pp. 494-501.
[11] S. M. Ross, "Applied Probability Models with Optimization Applications", (San Francisco: Dover Publications) **(1970)**, pp. 52.
[12] A. Laourine and L. Tong, "Betting on Gilbert-Elliot channels", IEEE Transactions on Wireless Communications, vol. 9, no. 2, **(2010)**, pp. 723-733.
[13] P. Auer, N. Cesa-Bianchi and P. Fischer, "Machine Learning", 47 235, **(2002)**.
[14] C. Tekin and M. Liu, "Approximately optimal adaptive learning in opportunistic spectrum access", INCOFOM,**(2012)**
[15] C. Tekin and M. Liu, "Online Learning in Opportunistic Spectrum Access: A Restless Bandit Approach," in 30th IEEE International Conference on Computer Communication(INCOFOM), **(2011)**.

Authors

Juan Zhang, she received her Doctor of Engineering in Signal and Information Processing from University of Chinese Academy of Sciences (2012). Now she is associate professor of School of Information Engineering, Southwest University of Science and Technology. Her current research interests include cognitive radio and intelligent learning.

Hesong Jiang, he received his Master of Engineering in Signal and Information Processing from University of Chinese Academy of Sciences (2009). Now he is lecturer of School of Information Engineering, Southwest University of Science and Technology. His current research interests include cognitive radio and intelligent learning.

Hong Jiang, he received his Doctor of Engineering in Communications professional from University of Electronic School of Information Engineering (2004). Now he is full professor of School of Information Engineering, Southwest University of Science and Technology. His current research interests include cognitive radio and intelligent learning.

Chunmei Chen, she received her Master of Engineering in Signal and Information Processing from University of Chinese Academy of Sciences (2008) and now she is associate professor of School of Information Engineering, Southwest University of Science and Technology. Her current research interests include cognitive radio and intelligent learning.