ENERGY AND RADIO SCIENCE

Machine Learning for Opportunistic Spectrum Access with Energy Consumption Constraint

Navikkumar Modi*, Christophe Moy*, Philippe Mary**

**CENTRALESUPELEC, IETR, UMR CNRS 6164, France {navikkumar.modi, Christophe.moy}@centralesupelec.fr*
*** INSA de Rennes, IETR, UMR CNRS 6164, France {philippe.mary}@insa-rennes.fr*

Upper Confidence Bound (UCB), Opportunistic Spectrum Access (OSA), Multi-Armed Bandit (MAB)

Abstract

This paper proposes an index based learning algorithm for the opportunistic spectrum access (OSA) scenario modeled as a Markov multi-armed bandit (MAB) problem. The proposed algorithm selects a channel for transmission which is optimal not only in terms of data rate, but in terms of quality as well, i.e. signal to noise ratio (SNR). It allows secondary users (SUs) to give appropriate weight to their desired criterion, such as channel quality, which lead to reliable transmission with lower power, and data rate, by selecting two distinguishable exploration coefficients. In cognitive radio context, we numerically compare the proposed policy with an existing UCB1 and also show that it outperforms traditional UCB1 in terms of transmission power requirement for SU.

1. Introduction

Cognitive radio (CR) enables to get access to the underutilized spectrum when it is not occupied by a licensed user and thus opens new doors in spectrum sharing for communication. In OSA context, primary users (PUs) are the licensed users who buy the right to use spectrum for certain time. While secondary user (SU) is the unlicensed user who could be allowed to use the spectrum for communication when no PUs is using it [1]. Along with the other problems, energy efficiency also plays an important role in a final successful implementation of CR.

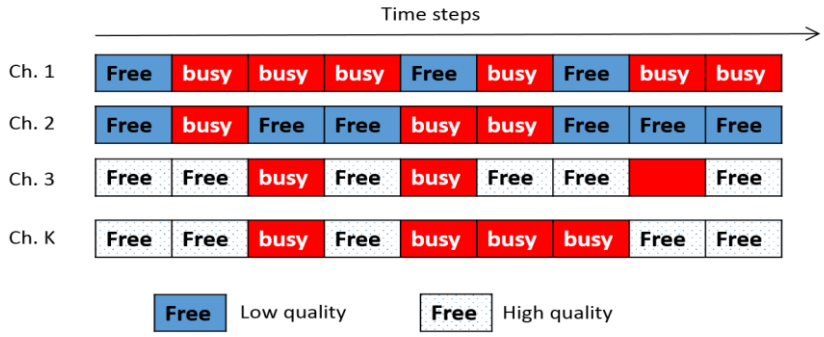


Fig. 1. Occupancy and channel condition considered for secondary user

The core of OSA problem being to learn which channels are the best in terms of chosen criterions, e.g. availability and quality, recent works have modeled the spectrum learning process with multi-armed bandit (MAB) framework [1,2,3]. The energy efficiency is just as much dependent on channel propagation conditions as the spectral efficiency and hence it is necessary to adapt the system to the changing environment. Fig. 1 illustrates the fact that some channels are more available than others, due to the traffic load of PUs, and some of them offer better quality for a wireless transmission in terms of signal to noise ratio (SNR), due to the propagation conditions, thus may potentially support a reliable transmission with lower transmit power.
Recently, MAB has been successfully used to model OSA scenario [1,2,3]. In MAB framework, an SU knows nothing *a priori* about the statistics of the channels, i.e., has no idea of how good or bad selected channels are in term of received

power or signal to noise ratio (SNR), and what data rate it may get from each channel. The data rate and quality of each channel could be learnt by exploration. The objective is to handle the exploration vs exploitation dilemma, defined as exploiting the best channel while simultaneously collecting the information about the best channel. In OSA, decision to transmit or not is done by simply characterizing a channel as a free or occupied, whereas the choice of the channel allocation should not be only done from a occupancy point of view, e.g. free or occupied but also on actual condition as shown in Fig. 1. The goal is to design a learning approach which searches for a spectrum holes within a given set and also achieve reliable transmission with lower power. Several works have dealt with the introduction of quality information in reinforcement learning schemes for spectrum allocation, but most of the approaches suffer from a long convergence time [4,5,6,7].

This paper presents an algorithm which is specifically designed to solve spectrum learning problem with significant reduction in the energy consumption of an SU. As all reinforcement learning schemes, proposed policy observes the availability of channel by sensing, which can be the output of the energy detector in case of CR. The rest of the paper is organized as follows. In Section II, we present cognitive engine and formulate the Markov MAB problem as a CR learning process. In Section III, we present a learning algorithm based on the local channel quality. Section IV presents the numerical results, verifying the validity and efficiency of the proposed policy. Finally, Section V concludes the paper.

2. Problem Formulation

CR equipment should normally consists in three additional functionalities compared to traditional software defined radio, which are spectrum sensing, decision making and learning engine [1]. In OSA scenario, CR first senses the spectrum and decision making engine decides to transmit or not based on the output of the spectrum sensing. Finally, the goal of learning engine is to predict the channel to sense for the next time instant. The selection of the next channel for transmission is not decided with the help of respective channels availability statistics only, but their quality is also taken into consideration. Learning engine can be successfully modeled as a Markov MAB problem.

We consider a Markov MAB framework with a single SU and $i \in \{1, \dots, K\}$ independent channels. The reward generated by state $q, q \in S^i$, of an arm i is denoted by $r_q^i(t) = S^i(t)$, where $S^i(t)$ denotes observed state at time t from channel i. The irreducible and aperiodic Markov chain with finite state space S^i is used for modeling the $i-$th arm for Markov MAB problem. The transition probability matrix of an arm i is denoted by $P^i = P_{k,l}^i, k, l \in \{q_0, q_1\}$ and $q_0, q_1 \in S^i$, where q_0 and q_1 are the Markov states of an arm i, i.e. occupied and free respectively. The arms are assumed to be mutually independent from each other. The stationary distribution π^i of the Markov chain is defined as $\pi_q^i(t) = \pi_q^i, \forall t$ and

$$\pi^i = [\pi_{q_0}^i, \pi_{q_1}^i] = \left[\frac{p_{q_1 q_0}^i}{p_{q_1 q_0}^i + p_{q_0 q_1}^i}, \frac{p_{q_0 q_1}^i}{p_{q_1 q_0}^i + p_{q_0 q_1}^i}\right].$$

Furthermore, this work considers another criterion for channel selection, which is the instantaneous quality of the channels. We assume that quality of channel within a given state q is stationary in the wide sense, meaning that its statistical properties, i.e. first and second moment, are not evolving over time, but the instantaneous quality $R_q^i(t)$ of the $i-$th channel varies. The learning policy decides an arm i to play at each time step t, based on a previously observed state q and quality observation R_q^i. The mean reward μ^i of an arm i under stationary distribution π_q^i is given by: $\mu^i = \sum_{q \in S^i} r_q^i R_q^i \pi_q^i$.

Channel with the highest mean reward μ^* is called an optimal channel, and its quality is denoted as R_q^*. Channel whose mean reward is strictly less than μ^* is referred as a suboptimal channels.

3. Channel Selection Policy

In this section, we propose a policy for CR learning. The policy aims to find a channel which is optimal in terms of data rate and its quality, which leads to reliable transmission with lower power for CR. Proposed policy, as any other reinforcement learning algorithm, learns from the observations acquired by sensing the channels without considering any *a priori* statistical information about the channels. Our first contribution is the algorithm represented as a flow chart in Fig. 2.

If a channel is never sensed, then its bound is infinite and this is why all channels are sensed at least once initially. After $n > K$ iterations as seen in Figure 2, proposed policy updates the index $B^i(n, T^i(n)$, where $T^i(n)$ is the number of time channel i has been sensed up to time n. At each iteration, policy returns the channel index i which is maximum.

The policy index is defined as

$$B^i(n, T^i(n)) = \bar{S}^i(T^i(n)) - Q^i(n, T^i(n)) + A^i(n, T^i(n)),$$

with $\bar{S}^i(T^i(n))$ term indicates the exploitation contribution, and $Q^i(n,T^i(n))$ and $A^i(n,T^i(n))$ represent exploration contributions. The index $B^i(n,T^i(n))$ comprises three terms; the first, $\bar{S}^i(T^i(n))$, is the empirical mean of the observed Markov states of channel i up to time n and it is defined as

$$\bar{S}^i\left(T^i(n)\right) = \frac{S^i(1)+S^i(2)+\cdots+S^i(T^i(n))}{T^i(n)}, \qquad \forall i$$

where $S^i(T^i(n))$ is the Markov state, i.e. occupied or free, of a channel i at time n. Second, $Q^i(n,T^i(n))$ represents channel quality term and is estimated by the use of the observed quality information $R_q^i(n)$. Finally, if the scheme leads to a channel which is in the occupied state, then the third bias term $A^i(n,T^i(n))$ forces to explore the other channels.

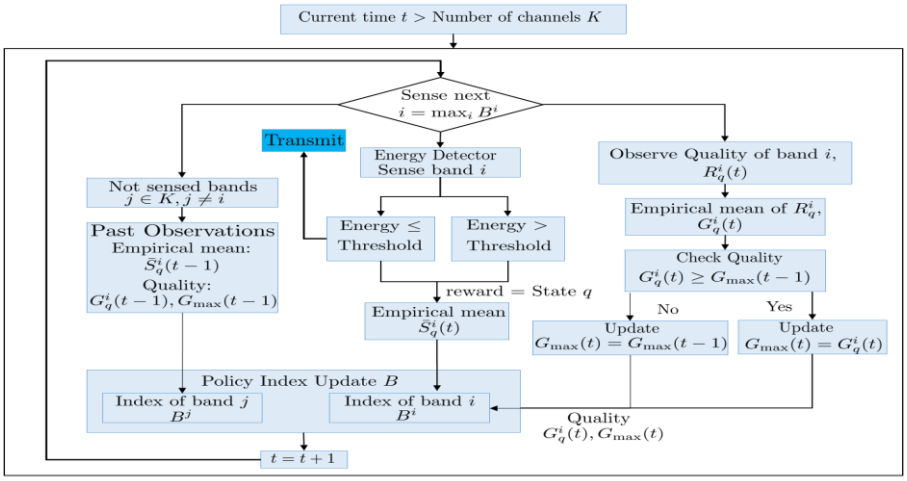


Fig. 2. Proposed Machine Learning Algorithm

The key parameter of the proposed policy is $Q^i(n,T^i(n))$ and is defined as

$$Q^i(n,T^i(n)) = \frac{\beta\, M^i(n,T^i(n)) \ln n}{T^i(n)}$$

where

$$M^i\left(n,T^i(n)\right) = G_{max}^q(n) - G_q^i\left(T^i(n)\right), \quad \forall \text{ I},$$

and $G_q^i(T^i(n)) = \frac{1}{T^i(n)}\sum_{k=1}^{T^i(n)} R_q^i(k)$ denotes the empirical mean of quality observations R_q^i collected from channel i in state q, $G_{max}^q(n) = \max_{i\in K} G_q^i(T^i(n))$ is the maximum expected quality within the set of channels. Thanks to this formulation, the proposed policy selects a channel for transmission which has higher quality $G_{max}^q(n)$ up to time n in state q.

$A^i(n,T^i(n))$ is a bias defined as

$$A^i\left(n,T^i(n)\right) = \sqrt{\frac{\alpha \ln n}{T^i(n)}}$$

In our algorithm, two coefficients come into play, i.e. α and β, respectively defined as the exploration parameter for learning the data rate of the channels and the weight value for quality based learning.

4. Numerical Analysis

In this section, we analyze the performance of the proposed policy in OSA scenario. The number of channels is set to 6 and simulation is performed over 100 runs to smooth the results. We use 10000 time instants for simulations in order to analyze the performances. We remind that q_0 and q_1 stand for the states occupied and free respectively. QPSK signaling are assumed to be used for PU signals. Moreover, each channel i has different quality levels, i.e. signal to noise ratio (SNR), which could be learnt by exploration of that channel.

The primary network transition probabilities P^i for each channel is selected arbitrary and is given in Table 1. Whereas, mean availability $\pi_{q_1}^i$ is estimated with the help of transition probabilities P^i of each channel. The channel quality is defined with the received SNR as shown in Table 1. The SU's transmit power required to achieve specific BER P_e is also given in Table 1 under certain hypothesis that are not necessary to detail as only relative value is worth in our study. Let,

optimal channel is the one which requires lower transmit power for SU from all the channels and has highest availability percentage.

Fig 3. presents the percentage of opportunities exploited which is defined as the ratio between the number of times a policy senses an available channel and the total number of time slots. We compare the performance of the proposed policy with baseline index policy UCB1, as proposed in [8]. As seen in Fig. 3, UCB1 gets the higher number of opportunities to transmit since UCB1 only looks for the channel with the highest availability. Whereas, the proposed policy is able to find lower percentage of opportunities, since it always senses a channel which is referred as an optimal channel both in terms of availability and quality, leading to lower transmission power requirement for SU.

$Channel\ (i)$	1	2	3	4	5	6
$P^i_{q_0q_1}$	0.8	0.9	0.6	0.4	0.3	0.9
$P^i_{q_1q_0}$	0.2	0.22	0.3	0.65	0.5	0.18
$\pi^i_{q_1}$	0.80	0.80	0.66	0.38	0.37	0.83
Received SNR (dB)	2	10	3	1	3	6
$P_T\ dBm$ (for $P_e = 10^{-4}$)	28	20	27	29	27	24
$P_T\ dBm$ (for $P_e = 10^{-3}$)	26	18	25	27	25	22

Table 1. State Transition Probability, Mean Availability and SU transmit power required to achieve specific BER

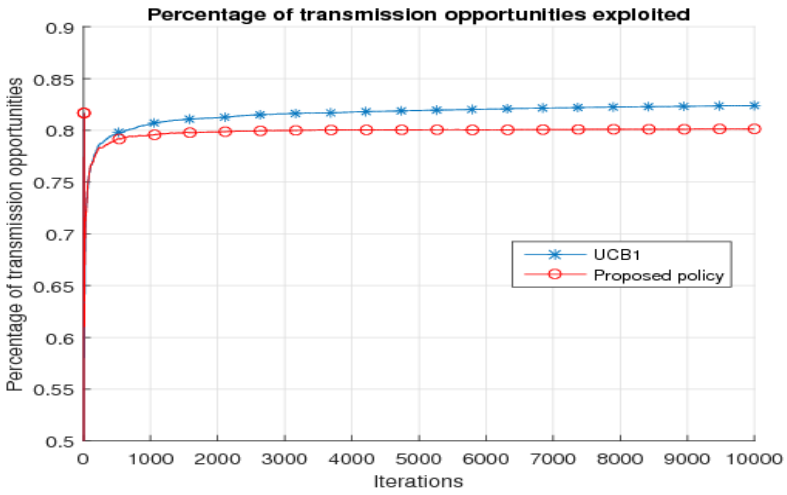


Fig. 3. Proposed policy and UCB1: Percentage of opportunities exploited

Fig. 4 shows the average transmission power required by SU to achieve a specific BER of 10^{-3}, Fig. 4(a), and a BER about 10^{-4}, Fig. 4(b), using the proposed policy and UCB1. Our algorithm requires much lower transmit power P_T for SU compared to UCB1. For instance, the average transmission power P_T required by an SU using the proposed policy is about 100 mW lower than UCB1 to achieve a BER equals to 10^{-4}.

Both algorithms explore different channels initially, thus the average transmission power P_T varies significantly as shown in Fig. 4(a) during the beginning of learning. Due to the introduction of the required P_T inside the index calculation, the proposed policy is able to select an optimal channel. The ability to select a channel with P_T suggests that the proposed policy could offer significant reduction in power consumption than traditional UCBs. As seen in Figs. 4(a), the proposed policy selects a channel ($i = 2$) which has lower P_T and also offers sufficient enough data rate. Whereas, UCB1 policy selects a channel ($i = 6$) which provides data rate, but ignores the transmission power constraint.

Fig 4(b) depicts the total power consumed by UCB1 and proposed policy to transmit 10000 frames with target BER $P_e = 10^{-4}$ on different channels. We can see that proposed policy consumes much lower power for transmission with predefined BER $P_e = 10^{-4}$, whereas, UCB1 consumes more power to continue transmission with desired BER, because UCB1 does not consider a channel quality to make a decision about next channel. To make a fair comparison, we calculate the SU's power consumed to transmit a single bit using UCB1 policy and proposed policy. Numerical analysis states that SU with UCB1 policy consumes 112 % more power compared to proposed policy to transmit a single bit with

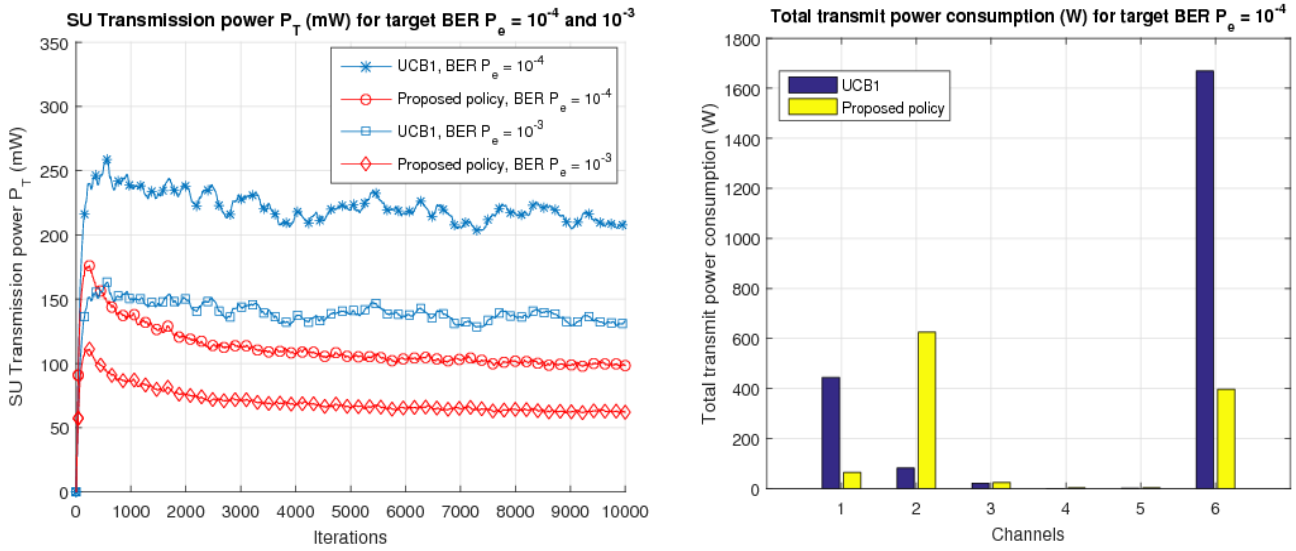
desired BER $P_e = 10^{-4}$. The results shown in Fig. 3 and 4, suggest that our policy is able to learn from various channel characteristics, i.e. availability and a quality criterion, linked to the transmit power in this paper, in opposite to UCB1 that only takes into account channel availability for learning. It can also be shown that the proposed policy keeps the logarithmic behavior in regret which makes it as much as interesting as UCB1 to learn on channels.

5. Conclusion

In this work, a machine learning policy based on channel quality information and availability has been discussed for OSA scenario. The learning is done from the observations acquired by sensing the channels without considering any *a priori* statistical information about them. It allows decision making in a set of channels finding a tradeoff between the most available channels and channels offering good quality (which leads to a lower energy consumption for SU). Numerical analysis have stated that our policy and the classical UCB1 are able to find opportunities for transmissions, but our scheme much often selects an optimal channel in terms of quality, e.g. SNR, leading to a lower energy consumption for the SU.

Acknowledgement

This work has received a French government support granted to the CominLabs excellence laboratory and managed by the National Research Agency in the "Investing for the Future" program under reference No. ANR-10-LABX-07-01. The authors would also like to thank the Region Bretagne, France, for its support of this work.



(a) (b)

Fig. 4. Proposed policy and UCB1: (a) SU Tx power P_T requirement for target BER $P_e = 10^{-4}$ and $P_e = 10^{-3}$ (b) Total transmit power consumption vs channel selected by respective policy for target BER $P_e = 10^{-4}$.

References

1. W. Jouini, D. Ernst, C. Moy, and J. Palicot, "Upper confidence bound based decision making strategies and dynamic spectrum access," in *International Conference on Communications, ICC'10*, May 2010.
2. H. Liu, K. Liu, and Q. Zhao, "Learning in a changing world: Restless multiarmed bandit with unknown dynamics," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1902–1916, 2013.
3. N. Modi, P. Mary, and C. Moy, "QoS driven channel selection algorithm for opportunistic spectrum access," in *IEEE Globecom 2015 Workshop on Advances in Software Defined Radio Access Networks and Context-aware Cognitive Networks (IEEE SDRANCAN 2015)*, San Diego, USA, Dec. 2015.
4. H. N. Pham, J. Xiang, Y. Zhang, et al., "Qos-aware channel selection in cognitive radio networks: A game theoretic approach," *In IEEE Globecom*, Nov 2008.
5. E. Ahmed, L. J. Yao, M. Shiraz, et al., "Fuzzy-based spectrum handoff and channel selection for cognitive radio networks," *in Proc. IC3INA*, Nov 2013.
6. A. Ali, M. Iqbal, S. Saifullah, et al., "Qos-based channel and radio assignment algorithm for mesh cognitive radio networks intended for healthcare," *in Proc. MESH*, 2012.
7. Noda, S. Prabh, M. Alves, et al., "Quantifying the channel quality for interference-aware wireless sensor networks," *SIGBED Rev.*, vol. 8, no. 4, dec 2011.
8. P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multi-armed bandit problem," *Machine Learning*, vol. 47, no. 2-3, May 2002.