Reinforcement learning demonstrator for opportunistic spectrum access on real radio signals

Christophe Moy, Amor Nafkha, Malek Naoues

Reinforcement Learning Demonstrator for Opportunistic Spectrum Access on Real Radio Signals

Christophe Moy, Amor Nafkha, Malek Naoues (E-mails: {christophe.moy, amor.nafkha, malek.naoues }@centralesupelec.fr)
CentraleSupelec/IETR, Rennes campus, Cesson-Sévigné, France

Abstract
This demonstration presents a proof-of-concept for opportunistic spectrum access. It particularly focuses on reinforcement learning algorithm called UCB (Upper Confidence Bound) designed by the machine learning community to solve the MAB problem (Multi-Armed Bandit). The demonstrator shows the first worldwide implementation of reinforcement learning algorithms for OSA (opportunistic spectrum access) on real radio environment using USRP N210 platforms.

Introduction
Cognitive Radio (CR) paradigm is all about providing self-adaptation capabilities to radio equipments and networks, so that they can adapt dynamically to environment conditions in the wide sense [1]. Learning is one of the main new features a radio equipment or system should have in order to turn from a classical radio towards Cognitive Radio. Indeed, the facilities a cognitive radio system should include in addition to usual radio processing can be summarized in a simplified cognitive cycle as [2]: (i) sensors, (ii) learning and decision making algorithms, (iii) adapting abilities. CR abilities can be used for instance to improve spectrum use and Opportunistic Spectrum Access (OSA) is one scenario investigated by CR [3].

Machine learning for OSA scenario
As a case study, an OSA scenario has been chosen here, but it could be other CR scenarios. This scenario has the advantage to be one of the most studied in the CR community. This can be considered as a futuristic scenario when regulation will authorize more dynamic spectrum access schemes, or a current scenario in the unlicensed bands. A reinforcement learning algorithm is implemented here, so that a Secondary User (SU) can learn very accurately and very fast what is the occupancy scheme of a primary network composed of Primary Users (PUs) that are communicating in a set of K channels. Without any *a priori* knowledge on the PUs' spectrum occupancy, the SU can predict which of the K channels of the considered band, offers the maximum probability of being vacant for next opportunistic communication. The goal of OSA is that a SU never interfere with PUs, so SUs can only make a communication on a channel after a sensing phase that makes it sure that the channel is vacant, e.g. not occupied by a PU.

Demonstrator
The considered learning algorithms are able to act in highly unpredictable conditions [4], and mathematically guarantee a convergence to the most vacant channels even if sensing errors occur [5], as it is the case in real radio conditions. The learning algorithms involved are called UCB (Upper Confidence Bound) [6] designed by the machine learning community to solve the MAB problem (Multi-Armed Bandit) [7]. The demo shows the first worldwide implementation of reinforcement learning algorithms for OSA (opportunistic spectrum access) on real radio signals. A snapshot of the demonstrator is shown on Figure 1

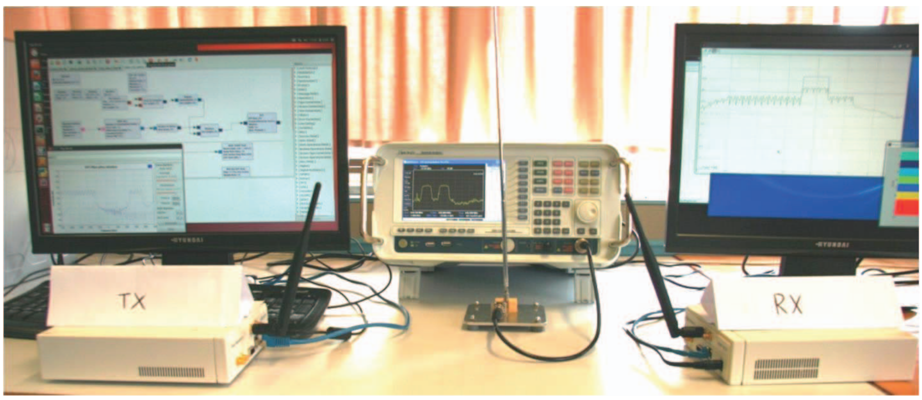
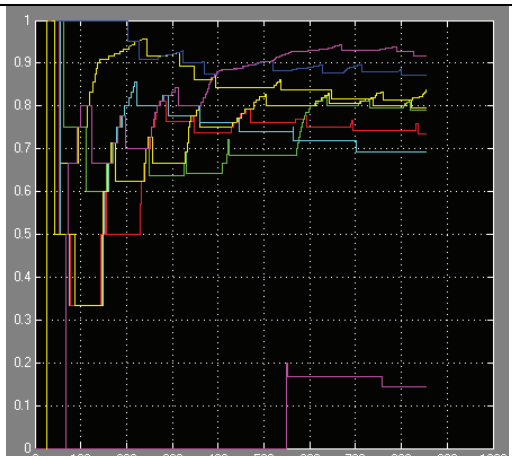


Figure 1 –Left hand side (laptop with GNU Radio + USRP) is generating traffic of PUs on 8 channels (TX). Right hand side (laptop with Simulink + USRP) is a secondary user learning algorithm, implementing an energy detector as a sensor (RX). A spectrum analyzer shows the RF signals.

978-1-4799-7452-8/15/$31.00 ©2015 IEEE

Demonstration results

During the demo operation, we can see how the learning algorithm behaves and progresses in its knowledge of the channels vacancy rate. Both evolutions of UCB indexes and empirical mean probability of vacancy can be followed in real time, as shown in Figure 2. We can also follow how many opportunities the algorithm has obtained so far and the percentage of successful accesses to a vacant channel. If we consider the following example of 8 primary channels with {0.5; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9} occupancy rates, we can see in this scenario where less than 60% of channels are vacant in average, that the proposed algorithm can very fast reach 75% of opportunities (a number of trials of 10 times the number of considered channels), and in the medium term 85% of opportunities whereas only one channel has a better average vacancy rate than this. So if we consider 8 channels in the experiment, only 80 iterations are necessary to almost double the opportunities compared to a non cognitive system. We prove also in this demo that our learning approach costs nothing, and can be evaluated at 1% of the sensing cost of an energy detector for instance. Note that any detector could be used as our learning algorithm is placed after a spectrum sensing detector, whatever its implementation (energy detector, cyclostationnarity detector, etc.). Hence we can see on Figure 2, middle table that channel #8, which is the most vacant channel has been selected the most, e.g. 85 over 350 times, with an empirical vacancy rate of more than 91%. This is this capability o UCB algorithm, to favor the best solution, which makes the cognitive system converge to the best solution. Whereas the mathematical proof guarantees it at infinity, all experiments we have made show that this happens very fast, making UCB approach an efficient solution in real conditions. Many other results, such as discussed in [8], will be detailed while playing the demo.



Holes distribution	Nb of trials per band
0.7955	44
0.1429	7
0.6923	26
0.7353	34
0.7907	43
0.873	63
0.8367	49
0.9176	85

Figure 2 - Evolution during first 350 iterations of an experiment shot. Left hand side: empirical average vacancy rate of the 8 channels (derived from UCB). Middle table: UCB trials on the right column and derived channels average vacancy rate at iteration number 350

Conclusion

Beyond the experiment al proof of concept for OSA, the proposed demo intends to also to demonstrate that CR can be implemented, while insisting on the most challenging item of the cognitive cycle: learning.

[1] J. Mitola, "Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio", PhD Thesis, KTH, Stockholm, Sweden, 8 May 2000.

[2] J. Palicot, "Radio Engineering: From Software radio to Cognitive Radio", Wiley 2011; ISBN: 978-1-84821-296-1

[3] Q. Zhao, A. Swami, "A Survey of Dynamic Spectrum Access: Signal Processing and Networking Perspectives", in IEEE ICASSP: special session on Signal Processing and Networking for Dynamic Spectrum Access, April, 2007

[4] W. Jouini, D. Ernst, C. Moy, J. Palicot, "Upper confidence bound based decision making strategies and dynamic spectrum access", ICC, Cape Town, South Africa, May 2010

[5] W. Jouini, C. Moy, J. Palicot, "Upper Confidence Bound Algorithm for Opportunistic Spectrum Access with Sensing Errors", CrownCom'11, 1-3 June 2011, Osaka, Japan

[6] P. Auer, N. Cesa-Bianchi, P. Fischer, "Finite time analysis of multi-armed bandit problems.", *Machine learning*, 47, 2002.

[7] R. Agrawal, "Sample mean based index policies with o(log(n)) regret for the multi-armed bandit problem," Advances in Applied Probability, 27:1054–1078, 1995

[8] C. Moy, "Reinforcement Learning Real Experiments for Opportunistic Spectrum Access", Karlsruhe Workshop on Software Radio, Karlsruhe, Germany, 12-13 March 2014