Cooperative Spectrum Sensing Using Q-Learning with Experimental Validation

Zhe Chen, and Robert C. Qiu
Department of Electrical and Computer Engineering
Center for Manufacturing Research
Tennessee Technological University
 Cookeville, TN 38505, USA
Email: zchen42@students.tntech.edu, riqu@tntech.edu

Abstract—Spectrum sensing detects the availability of the radio frequency spectrum and it is essential to cognitive radio. Cooperative spectrum sensing conducted among multiple secondary users has the potential for solving the hidden terminal problem and achieving a better performance. In this paper, an approach for cooperative spectrum sensing based on Q-learning is proposed. Real-world Wi-Fi signals measured from four different locations are employed to evaluate the performance of the proposed approach. Experimental results show the performance of the proposed approach is better than that of the popular M-out-of-N rule. The proposed approach is effective. This work is a part of the efforts toward building a cognitive radio network testbed.

I. INTRODUCTION

Cognitive radio (CR) has been put forward to make efficient use of scarce radio frequency spectrum. It introduces “intelligence” beyond software defined radio (SDR). Spectrum sensing is the cornerstone of CR, which detects the availability of the spectrum for secondary users (SUs) in CR networks.

There have been some techniques for spectrum sensing, such as energy detection, matched filter detection, cyclo-stationary feature detection, covariance-based detection, and wavelet-based detection [1], [2]. Moreover, cooperative spectrum sensing among secondary users have been put forward and introduced in [3], [4], [2], [5] to solve the hidden terminal problem [6], [7] and to improve the performance of spectrum sensing.

Recently, there is a trend of applying machine learning algorithms to CR [8]. Q-learning [9], [10] is one of the algorithms in the reinforcement learning family of machine learning. There have been a few studies on applying Q-learning to CR and CR networks. In [11], Q-learning is applied for cognitive base transceiver station to manage communications with a set of mobile stations in a vehicular context, and results are given based on simulations. In [12], [13], Q-learning is used for channel selection without negotiations among secondary users for multi-user and multi-channel CR systems. Both theoretical analysis and simulation results are presented. In [14], a correlated Q-learning algorithm is proposed to calculate the correlated equilibrium policy for the Markovian game to solve the transmission control problem in a CR system, but no simulation result is shown. In [15], decentralized Q-learning is proposed to manage the aggregated interference generated by multiple wireless regional area network (WRAN) systems, and simulation results are given. In [16], a Q-learning based scheme for opportunistic spectrum access environments is studied, which learns by itself from interaction and uses the acquired knowledge to locate and find the best spectrum opportunities. Evaluation results are given based on simulation. In [17], Q-learning is employed to choose different transmission parameters to make an efficient assignment of spectrum and transmit powers. Simulation results are discussed.

However, none of the above existing works on Q-learning deals with cooperative spectrum sensing, and none of them uses measured real-world data for performance evaluation.

In this paper, a Q-learning based cooperative spectrum sensing approach is proposed, and measured Wi-Fi signals from different locations are employed to evaluate the performance of the proposed approach.

We are building a CR network testbed of tens of nodes [18], [8]. The CR network testbed will enable testing the proposed cooperative spectrum sensing approach in real-time.

The rest of this paper is organized as follows. Section II introduces cooperative spectrum sensing. Section III describes Q-learning and the proposed approach. Section IV reports the results of performance evaluation. Finally, Section V concludes this paper.

II. COOPERATIVE SPECTRUM SENSING

Cooperative spectrum sensing is illustrated in Fig. 1. SU1, SU2, . . . , SU_N sense the frequency spectrum periodically, where N is the number of SUs. In each time slot t, SUs sense one or more frequency bands during the spectrum sensing phase. Then each SU makes a decision on the channel state of frequency band f during time slot t on its own discrimination. Denote the N decisions from N SUs as d1,t,f, d2,t,f, . . . , dN,t,f, respectively. After that, d1,t,f, d2,t,f, . . . , dN,t,f are fed to a cooperative spectrum sensing approach to form a single decision dt,f on channel state for frequency band f during time slot t. Since a channel state can be either “busy” or “idle”, the decisions of channel states are denoted as “1” or “0” to represent a “busy” channel state or an “idle” channel state, respectively. If dt,f = 0, then SUs can take advantage of the following communication phase of time slot t to transmit data.
A popular scheme for cooperative spectrum sensing is the \( M \)-out-of-\( N \) rule, which is naturally suitable for hard combination of multiple decisions. Assume there are \( N \) SUs sensing channel state cooperatively and they can communicate with each other. Each SU provides one-bit information \( d_i \) representing the channel state it senses, where \( i = 1, 2, \ldots, N \), \( d_i = 1 \) represents a “busy” channel state and \( d_i = 0 \) represents an “idle” channel state. Then the \( M \)-out-of-\( N \) rule can be expressed by the following equation:

\[
\text{ChannelState} = \begin{cases} 
\text{busy} & \text{if } \sum_{i=1}^{N} d_i \geq M \\
\text{idle} & \text{otherwise}
\end{cases}
\]  

(1)

Specially, when \( M = 1 \), the \( M \)-out-of-\( N \) rule becomes the “OR” rule; and when \( M = N \), the \( M \)-out-of-\( N \) rule becomes the “AND” rule.

III. COOPERATIVE SPECTRUM SENSING USING Q-LEARNING

A. Q-Learning

Compared to Markov decision process (MDP) and more sophisticated partially observable Markov decision process (POMDP) in the area of reinforcement learning, Q-learning is simpler [9], [10]. The core of the Q-learning algorithm is a Q-table and an algorithm for updating the Q-table and choosing actions. A Q-table \( Q(s, a) \) is a matrix indexed by state \( s \) and action \( a \), which is the expected discounted reinforcement of taking action \( a \) in state \( s \). At each time, an agent is assumed to be in a certain state \( s \), and it chooses an action \( a \) according to the Q-table and other algorithms to interact with the environment. Then the agent receives an reward \( r \) from performing action \( a \) and observes a new state \( s' \). After that, the Q-table is updated by the following rule:

\[
Q(s, a) = (1 - \alpha)Q(s, a) + \alpha(r + \gamma \max_{a'}(s', a'))
\]  

(2)

where \( \alpha \) is the learning rate, and \( \gamma \) is the discount factor.

B. Proposed Cooperative Spectrum Sensing Using Q-Learning

Assume \( N \) SUs sense the spectrum cooperatively. Each SU senses the spectrum and outputs one-bit information \( d_{i,t,f} \) that represents the channel state of frequency band \( f \) during time slot \( t \), where \( i = 1, 2, \ldots, N \) and \( d_{i,t,f} \in \{0, 1\} \). Let “0” represent the “idle” channel state and “1” represent the “busy” channel state. Then \( N \) one-bits can form an integer \( s_{t,f} \) whose value ranges from 0 to \( 2^N - 1 \):

\[
s_{t,f} = \sum_{i=1}^{N} d_{i,t,f} \cdot 2^{i-1}
\]  

(3)

In cooperative spectrum sensing using Q-learning, let the set of states \( S \) be \( \{0, 1, \ldots, 2^N - 1\} \), and the set of actions \( A \) be \( \{0, 1\} \), where \( s \in S \) and \( a \in A \). An action with index “0” means the channel is in “idle” state and available for SUs, and an action with index “1” means the channel is in “busy” state and unavailable for SUs.

Assigning values to reward \( r \) for Q-learning can be tricky. In the proposed algorithm, \( r \) is assigned as follows:

\[
r = \begin{cases} 
R_p & a_{t,f} = c_{t,f} \\
R_n & a_{t,f} \neq c_{t,f}
\end{cases}
\]  

(4)

where \( a_{t,f} \) is the chosen action for frequency band \( f \) during time slot \( t \), \( c_{t,f} \) is the actual channel state of frequency band \( f \) during time slot \( t \), \( R_p \) and \( R_n \) are constants.

The proposed Q-learning based approach for cooperative spectrum sensing is shown in Fig. 2, where \( a \) is the output.

IV. EXPERIMENTAL RESULTS

In this section, the performance of the proposed cooperative spectrum sensing approach using Q-learning is evaluated using measured Wi-Fi signals.

A. Measurement of Wi-Fi Signals

Wi-Fi has been widely used in daily life. The durations that Wi-Fi devices access the channel and the durations that the channel is kept idle are as short as microseconds. In this experiment, Wi-Fi is regarded as primary user (PU), although Wi-Fi is a secondary system in nature.

Wi-Fi signals have been measured using four antennas at four different locations [19]. The experiment is conducted in...
1. Let current state be $s$, where $s = s_{t,f}$, and $s_{t,f}$ is calculated by (3) using $N$ outputs $d_{i,t,f}$ from $N$ SUs.
2. Choose an action $a$ by:
   $$a = \begin{cases} 
   0 & Q(s,0) \geq Q(s,1) \\
   1 & \text{otherwise}
   \end{cases}$$
   (5)
3. Receive a reward $r$ from choosing an action $a$, where the value of $r$ is given by (4). Denote the resulting state as $s'$, where $s' = s_{t+1,f}$ and $s_{t+1,f}$ is calculated by:
   $$s_{t+1,f} = \sum_{i=1}^{N} d_{i,t+1,f} \cdot 2^{i-1}$$
   (6)
4. Update $Q(s,a)$ using (2).
5. $t = t + 1$. Goto step 1.

Fig. 2. Proposed Q-learning based approach for cooperative spectrum sensing.

![Amplitude (V) vs Time (ms) for Antenna #1, #2, #3, and #4](image)

Fig. 3. Measured Wi-Fi signals in time-domain.

Four antennas are directly connected to the four channels of the DPO. The distances between the four antennas and the Wi-Fi router are 3 meters, 2 meters, 2 meters, and 0 meter, respectively. Antenna 3 is blocked by a metallic board. So there is non-line-of-sight (NLOS) between antenna 3 and the Wi-Fi router. The sampling rate of the DPO is set to 6.25 GS/s. Since it can record 250 M samples per channel, the maximum duration of one measurement is 40 ms. The measured Wi-Fi signals in time-domain from the four antennas are shown in Fig. 3.

B. Performance Evaluation

Measured Wi-Fi signals from antenna 1, antenna 2, and antenna 3 are regarded as the received signals of SU1, SU2, and SU3, respectively. The measured Wi-Fi signals from channel 4 are employed as benchmark of channel state. Channel states marked using the Wi-Fi signal from channel 4 are regarded as actual channel states.

Since the maximum duration of the recorded Wi-Fi signal is 40 ms, the length of time slot is set to 20 $\mu$s, in order to get enough time slots. Thus, there are 2000 time slots available for performance evaluation. Note that the channel states from the measured Wi-Fi signals can change in microseconds. With the length of time slot set to 20 $\mu$s, the channel states can change in one or two time slots, which meets actual scenarios well. The duration of spectrum sensing phase of time slot is set to 4 $\mu$s in this experiment, one fifth of the duration of time slot.

Parameters of the proposed approach for performance evaluation are: $N = 3$, $f = 2.418$ GHz, $t = 1, 2, \ldots, 2000$, $R_p = 100$, $R_n = -100$, $\alpha = 0.1$, and $\gamma = 0.1$.

The performance is evaluated by two metrics: probability of detection ($P_D$) and probability of false alarm ($P_{FA}$). Fig. 4 shows the performances of the proposed approach and the $M$-
"out-of-$N$ rule evaluated using the measured Wi-Fi signals. It can be observed that the overall performance of the proposed algorithm is better than that of the $M$-
"out-of-$N$ rule.
C. Discussions

\( \alpha \) and \( \gamma \) are two parameters for Q-learning. Different values of \( \alpha \) and \( \gamma \) may result in different performances. Fig. 5 shows \( \alpha \) and \( \gamma \) versus probability of miss detection and probability of false alarm, which confirms a combination of \( \alpha = 0.1 \) and \( \gamma = 0.1 \) is suitable.

V. Conclusion

An approach for cooperative spectrum sensing based on Q-learning has been proposed. Performance of the proposed approach has been evaluated using measured real-world Wi-Fi signals. Experimental result shows the proposed approach is indeed effective.

This work is a part of the efforts toward building a cognitive radio network testbed. The proposed approach will be implemented on the cognitive radio network testbed for further test in real-time.

Acknowledgment

This work is funded by National Science Foundation through two grants (ECCS-0901420 and ECCS-0821658), and Office of Naval Research through two grants (N00010-10-1-0810 and N00014-11-1-0006).

References