Multi-agent Q-Learning

Hossein Aboutalebi

March 31, 2017

Introduction

- cognitive radio systems are designed with an unprecedented level of intelligence
- they are solution for unprecedented increase in demand of mobile services
- they are able to monitor, sense, and detect the conditions of their operating environment, and dynamically reconfigure their own characteristics to best match those conditions.
- they can access unused frequencies to extract more wireless bandwidth

Introduction

- It allows users without license (called secondary users) to access licensed frequency bands when the licensed users (called primary users) are not present.
- Therefore, the cognitive radio technique can substantially alleviate the problem of underutilization of frequency spectrum.

cognitive radio Problems

Two problems are key to the cognitive radio systems:

1. **Resource mining**, i.e. how to detect available resource (frequency bands that are not being used by primary users)
2. **Resource allocation**, i.e. how to allocate the detected available resource to different secondary users.

Here we consider the resource allocation problem.

Problem Setting

- we study the problem of spectrum access without negotiation in multi-user and multi-channel cognitive radio systems.
- each secondary user senses channels and then chooses an idle frequency channel to transmit data, as if no other secondary user exists.
- If two secondary users choose the same channel for data transmission, they will collide with each other and the data packets cannot be decoded by the receiver(s)
- there is no mutual communication among these secondary users, conflict is unavoidable

Problem Setting

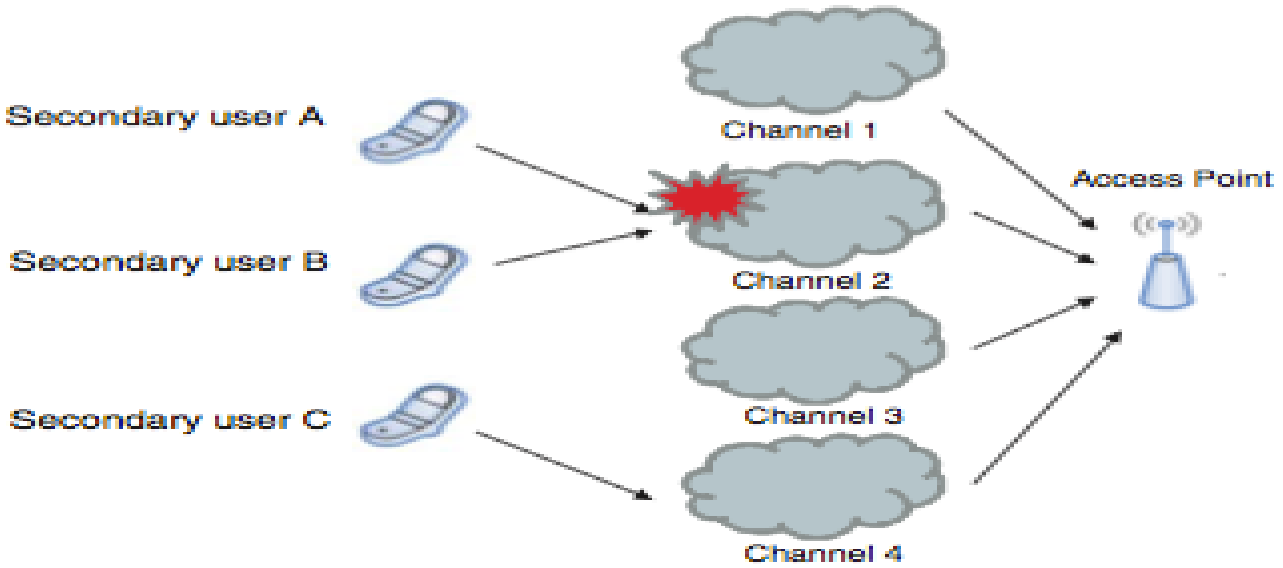


Fig. 1: Illustration of competition and conflict in multi-user and multi-channel cognitive radio systems.

- the secondary users can try to learn how to avoid each other, as well as channel qualities

Model

- we consider only two secondary users, denoted by A and B, and two channels, denoted by 1 and 2.
- The reward to secondary user i, i = A,B, of channel j, j = 1, 2, is R_{ij} if secondary user transmits data over channel j and channel j is not interrupted by primary user or the other secondary user; otherwise the reward is 0
- The rewards R_{ij} are unknown to both secondary users.
- Both secondary users can sense both channels simultaneously, but can choose only one channel for data transmission.
- There is no communication between the two secondary users.

Model

- It is easy to verify that there are two Nash equilibrium points in the game
- Both equilibrium points are pure, i.e. $a_A = 1, a_B = 2$ and $a_A = 2, a_B = 1$

User B

User A	Chan 1	Chan 2
Chan 1	0	R_A1
Chan 2	R_A2	0

Payoff matrix of user A

User B

User A	Chan 1	Chan 2
Chan 1	0	R_B1
Chan 2	R_B2	0

Payoff matrix of user B

Fig. 2: Payoff matrices in the game of channel selection.

Q-Learning

We consider Boltzmann distribution for random exploration

$$P_{i,j} = \frac{e^{Q_{i,j}/\gamma}}{e^{Q_{i,j}/\gamma} + e^{Q_{i,j-}/\gamma}}$$

where, $p_{i,j}$ denotes user i picks channel j. The expected return is:

$$E_{i,j} = \frac{R_{i,j} e^{Q_{i-,j-}/\gamma}}{e^{Q_{i-,j}/\gamma} + e^{Q_{i-,j-}/\gamma}}$$

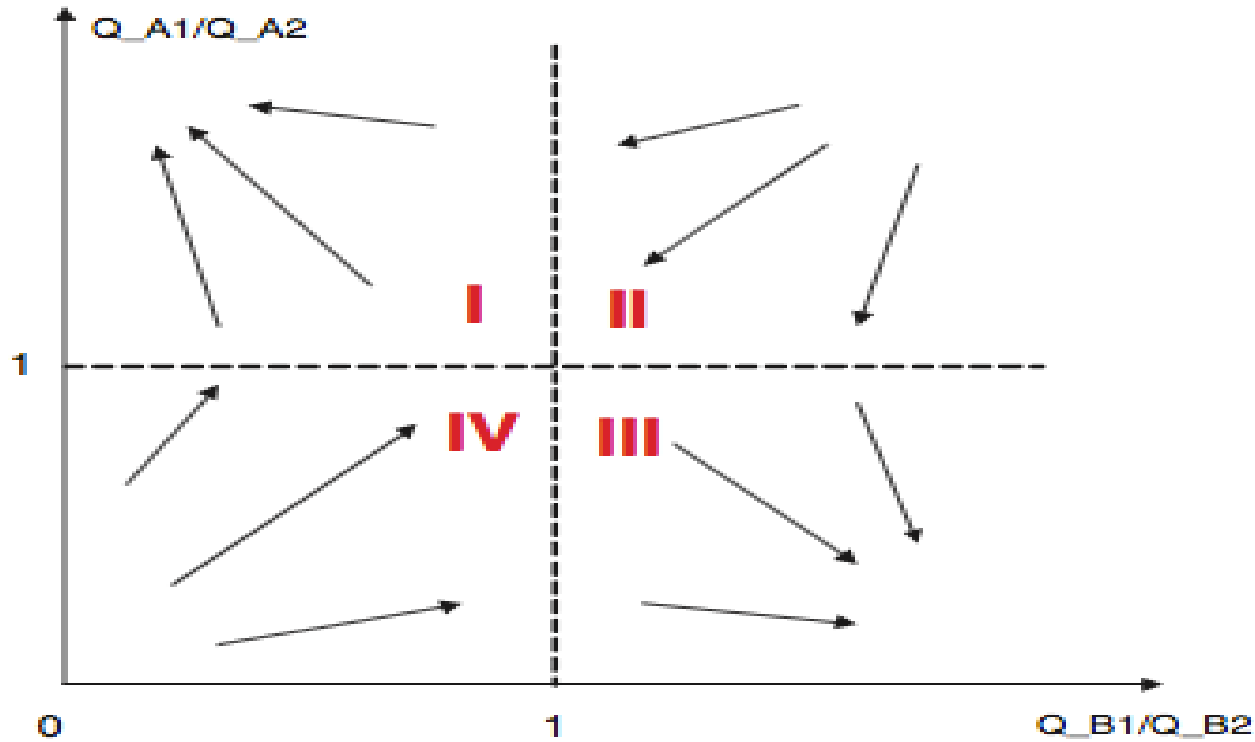
Q-Learning Update

- In the procedure of Q-learning, the Q-functions are updated after each spectrum access via the following rule:

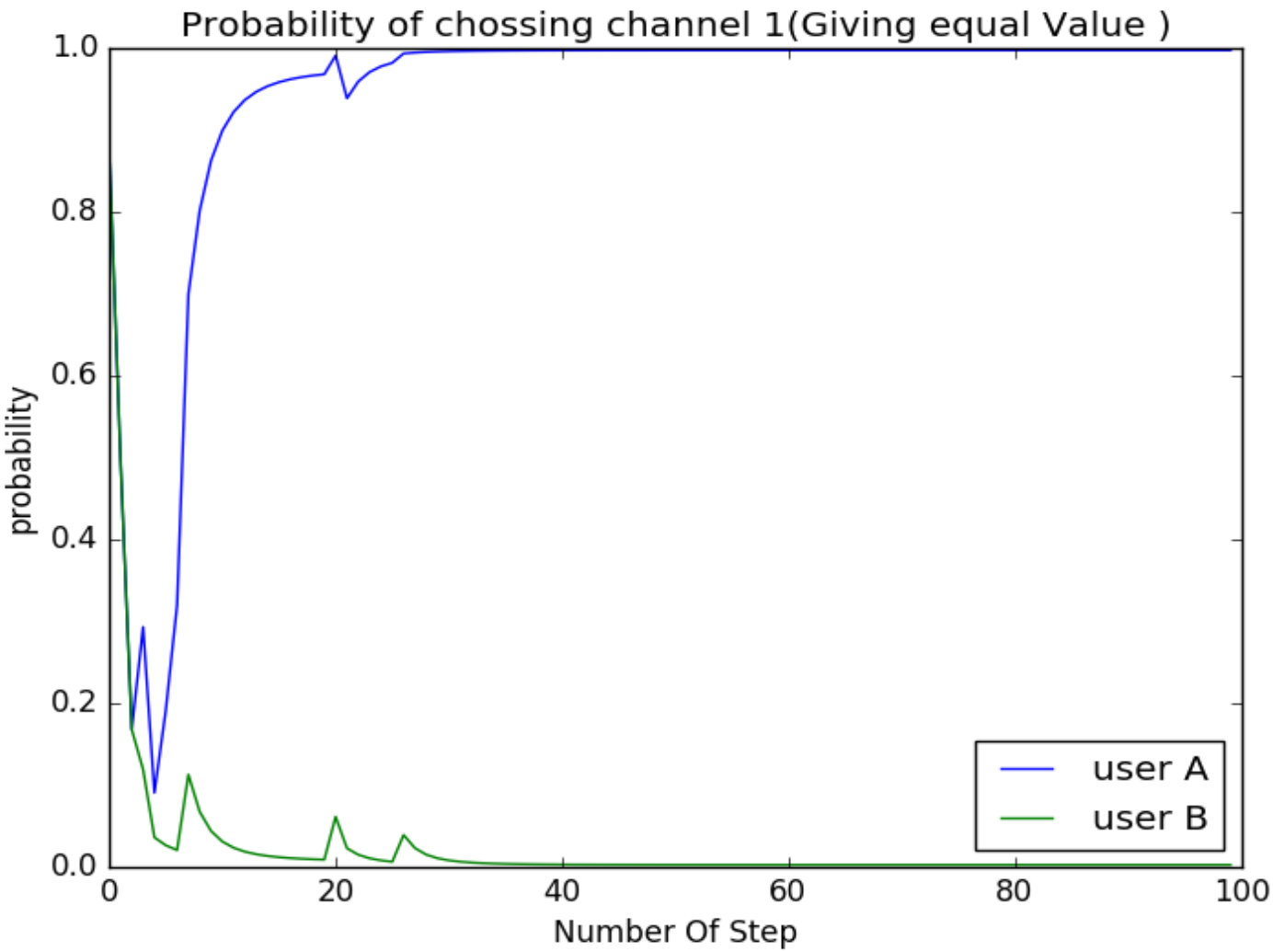
$$Q_{i,j}^{t+1} = (1 - \alpha_{i,j}) Q_{i,j}^{t} + \alpha_{i,j} r_i I(a_{i^-}^{t} = j^-)$$

- I is an indicator function
- We must assure convergence

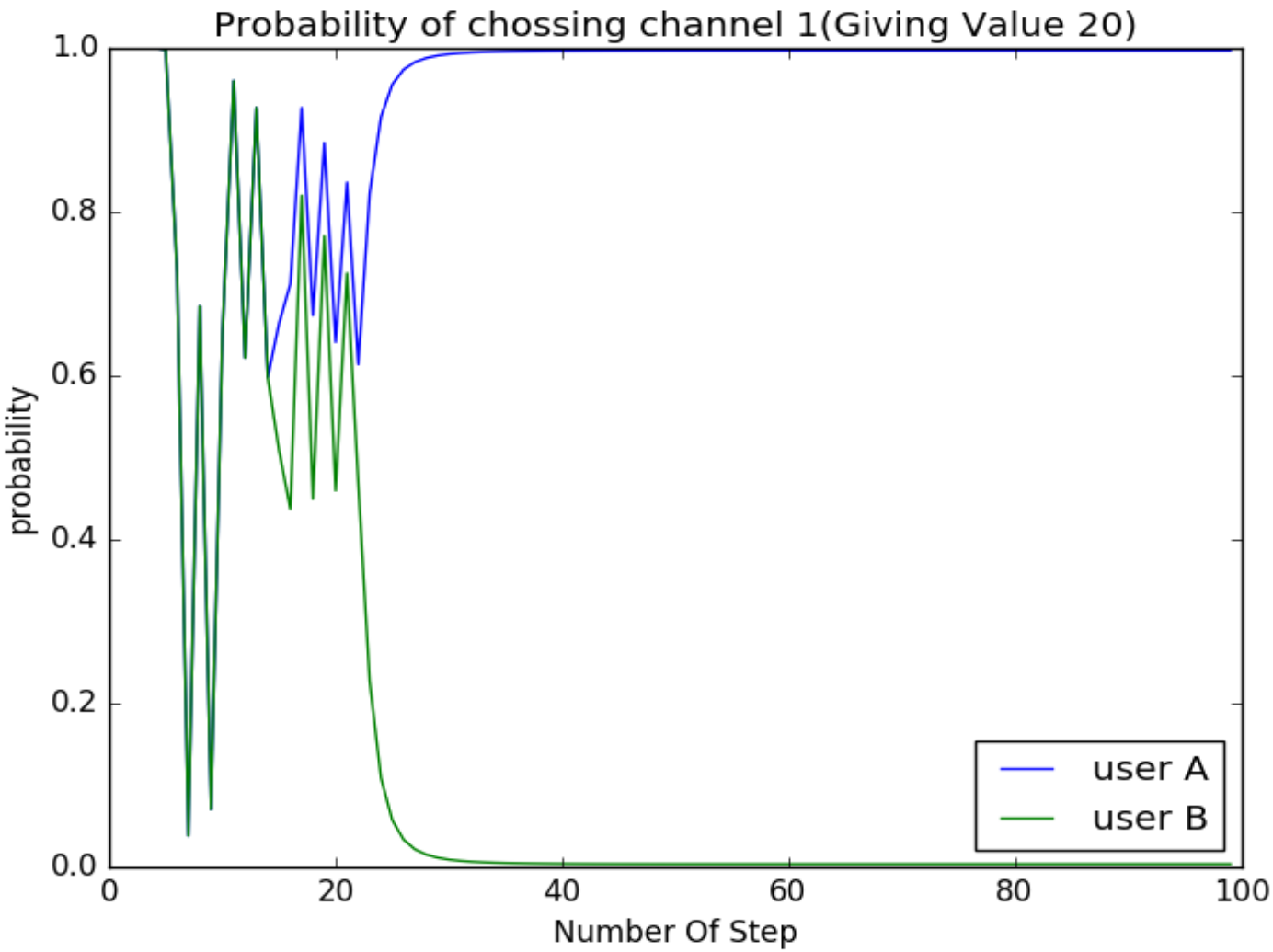
Convergence



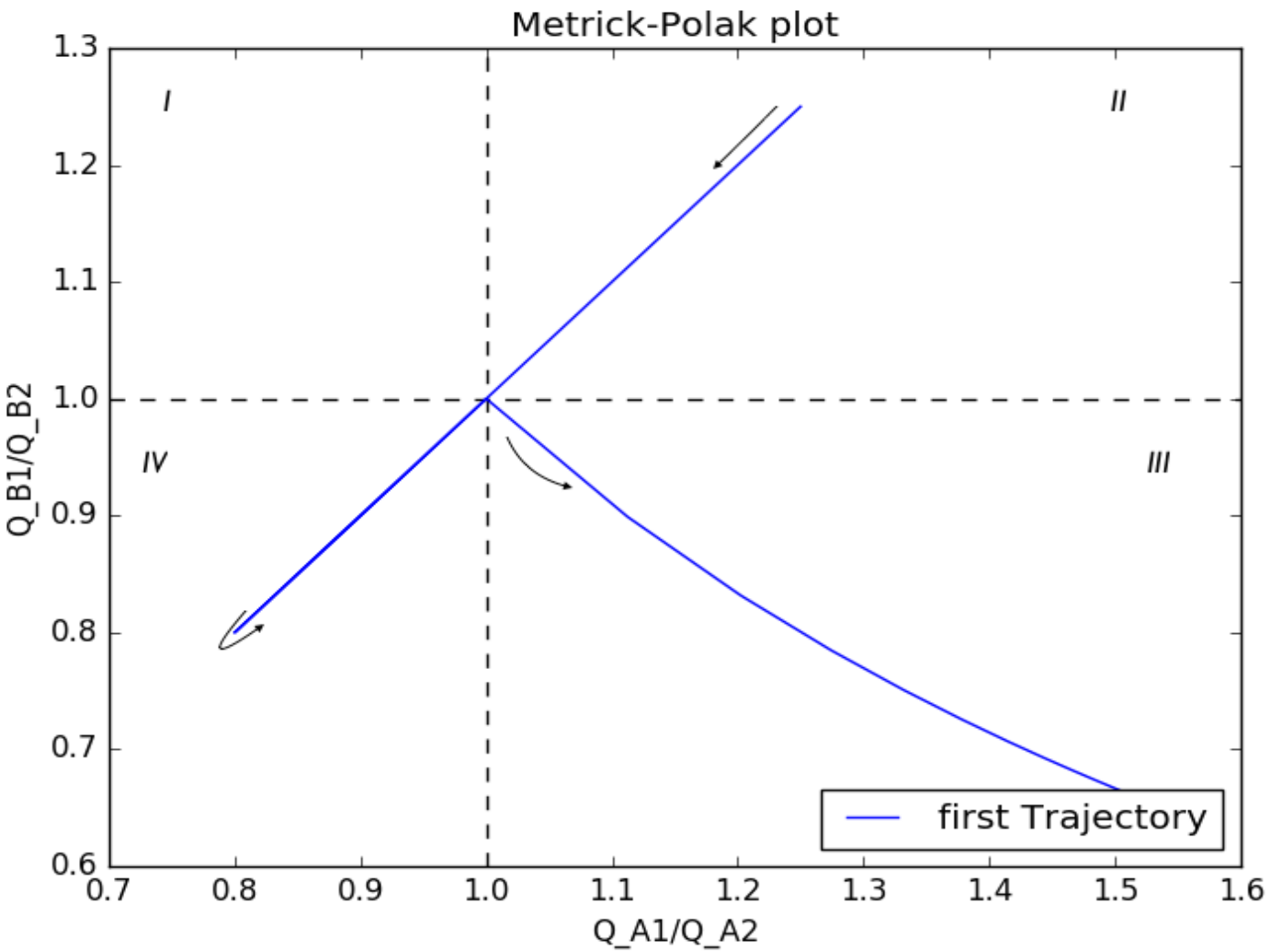
Experimental Results



Experimental Results



Experimental Results



Experimental Results