The 4th Workshop of COST Action IC0902
Cognitive Radio and Networking for Cooperative Coexistence of
Heterogeneous Wireless Networks
October 9–11, 2013
Rome, Italy

AUTHORS CONSENT TO INCLUDE THIS CONTRIBUTION IN THE OPEN ACCESSONLINE REPOSITORY OF IC0902: YES
NAME OF AUTHOR GIVING CONSENT: Abdulkarim Oloyede and David Grace

"Special Interest Group 2 - Learning and artificial intelligence"

Learning Based Auction Process for Cognitive Radio Systems

Communications Research Group, University of York, York, UK.

Abdulkarim Oloyede, Communications Research Group,Uni.of York, UK.

Introduction

Cognitive radio systems are defined as wireless systems that are designed to interact and observe the transmission environment. The learning is based on the interaction with the environment in order to allow for opportunist access to the spectrum [1]. Cognitive radio system provides access to the radio spectrum in a dynamic mannerknown as Dynamic Spectrum Allocation (DSA). Access to the radio spectrum using DSA requires a fair means of allocation. Thus, the useof an auction provides a fair allocation process by granting access to the highest bidder(s). Using this method(auction) to allocate the spectrum involves the bidders (wireless devices)submitting bids to reflect their valuation or budget. The scenario considered in this work is as proposed in [2]. In the proposed scenario, the valuation of the users does not only depend on the user's budget, it also depends on other factors such as the traffic on the system. The traffic is space and time dependent. Therefore, the offered bid varies.To implement DSA for wireless devices reflecting the above mentioned factors, Machine Learning (ML) can be of great importance.

Generally, ML involves the concept of artificial intelligence by solving problems based on experience. The learning process usually involves some form of rewards and penalty which accumulates additively with or without a discount factor. ML involves an agentusing the information learnt over time to move from one state at time (t) to another more favourable state at time $(t+x)$. Where $x \geq 1$. This work compares the use of Q Learning (QL)and QL which biased with Bayesian learning to learn the optimal bidding price in an auction based DSA process. The auction process is a multi-winner sealed bid auction with a reserve price as proposed in [2]. The scenario in this work is based on the users attempting to win the bidding process with the least possible bid value in an uplink scenario.

The purpose of this workshop paper is to examine the effects of Q and Bayesian reinforcement learning in an auction based dynamic spectrum access based network.

The Utility Function.

We assume that users are price sensitive and therefore want to win the bid with the least possible amount. The utility function measures how much a winning bidder deviates from the lowest winning bid. The lower the value of the winning bid the higher the value of utility. The closer the bid of a winning bidder to the minimum winning bid the higher the utility of the user. If the user is not among the winning bidders then the utility is zero. We assume $M_R$winning bidders emerge after each bidding round(R). Where the size of $M_E$is the number of channels available at time t_R.

Set **M**contains the winning bid and b_m is the bid of the winning user with cardinality $|M|$

$\boldsymbol{M} = \{b_1, b_2, b_3 \ldots b_m\}$ where $\delta = \min \boldsymbol{M}$ (1)

$$u_i(t)=\begin{cases}2^{\frac{\delta}{b_i}}-1 & For\ winning\ bidders\\ 0 & otherwise\end{cases} \quad (2)$$

Q Reinforcement Learning (QL)

This work assumes that bidders only pose information regarding their own bid history. The reserve price is not known to the bidders as it is set by the WSP. This is calculated as shown below

$$r=\frac{ZC_r}{K|M|} \quad (3)$$

Where $K$is the total number of channels in the system,Z is the number of bidders out of the possible N bidders in the system. This is because not all the bidders are bidding in all the bidding rounds.C_r is a constant in price units which is used to specify the value of a particular spectrum and |M| is the cardinality of set M. The reserve price takes into account the level of congestion in the system ($\frac{Z}{|M|}$). A bidder submits a bid to the WSP and is based on the value of the submitted bid, the user obtains a utility value using the utility equation in (2). A user usually wants to win the bid with a high value of utility, however the lower the offered price of the bidder, the lower the probability of the bidder winning the bid. Therefore, the dilemma faced by a bidder is to decide the best value of utility that allows the bidder to be among the winning bidders. The objective of the learning user is to obtain an optimal π^* that maximises the total expected utility as given in equation (2).

$$V^{\pi}=E\left\{\sum_{t=0}^{\infty}\beta^t r\left(s_t,\pi(s_t)\right)|s_o=s\right\} \quad (4)$$

Where E is the expectation operator, $0\leq\beta<1$ is a discount factor and π is a policy S$\rightarrow A$. For a policy π a Q value is defined as the expected discounted reward for executing a at action state s and the following policy π is given as

$$Q^{\pi}(s,a)=R(s,a)+\beta\sum_{s^*}Pr_{\dot{s}},s^*(a)V^{\pi}(s^*) \quad (5)$$

Where $R(s,a)$ is the old value of the Q and the other part of the equation is the reward function that leads to a new state.The utility obtained by each user is added after each bidding round depending on the bid value the user

Bayesian framework for Rainforcement Learning (BRL)

The Bayesian algorithm allows the learning agent to make a decision based on the most likely events that could happen, using prior experience. This allows for a faster and smooth movement from exploration to exploitation behavior. We apply the Bayes' theorem in the exploration stage and applied as shown below.

$$P(A\backslash B)=\frac{P(B\backslash A)\times P(A)}{P(B)} \quad (6)$$

Where $P(A)$ is the prior probability distribution of hypothesis A, $P(B)$ is the probability of the training data B (likelihood) and $P(B\backslash A)$ is the probability of A given B (Posterior probability). Bayesian learning allows the incorporation of prior probabilities in determination of the user's transition. The general flow chart is as giving in Figure 1.

Results and Discussions

To simplify the auction process we assumes a fixed bid value from 45-55 price units and K=4. As seen from figure 2(a), the learning process does not peak at the same point because a user cannot have both a high value of utility and winning all the time. This is because a user bidding at the highest possible value wins but, with a utility value (close to zero). As more trialsare carried out using QRL the optimal point is reached after 500 trials. However using the Bayesian equation to bias the learning the bid learning converges to the optimal bidding value after 100 trials. This shows that BRL converges faster than QRL. This result alone cannot prove that the bid value of 50 is the optimal price value that gives the best system performance to the user. Future work would examine and determine if this optimal value leads to better system performance and if it ultimately reduces the energy consumed by the wireless system giving inthe proposed scenario in our previous work in [2].

Conclusions

QRL and BRL is examined in this work using an auction based DSA scenario. It shows that BRL converges faster

than QRL.

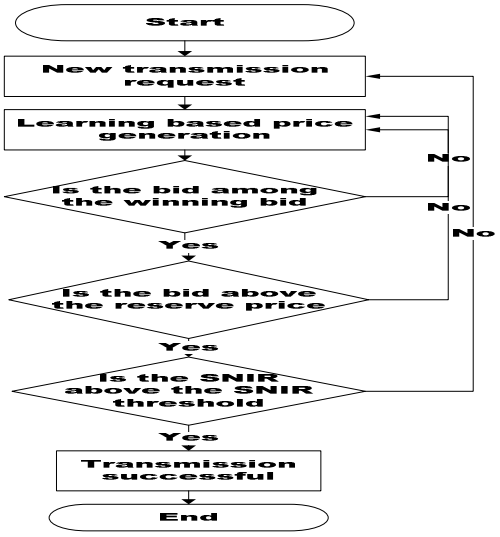

Figure 1. System Flow Chart

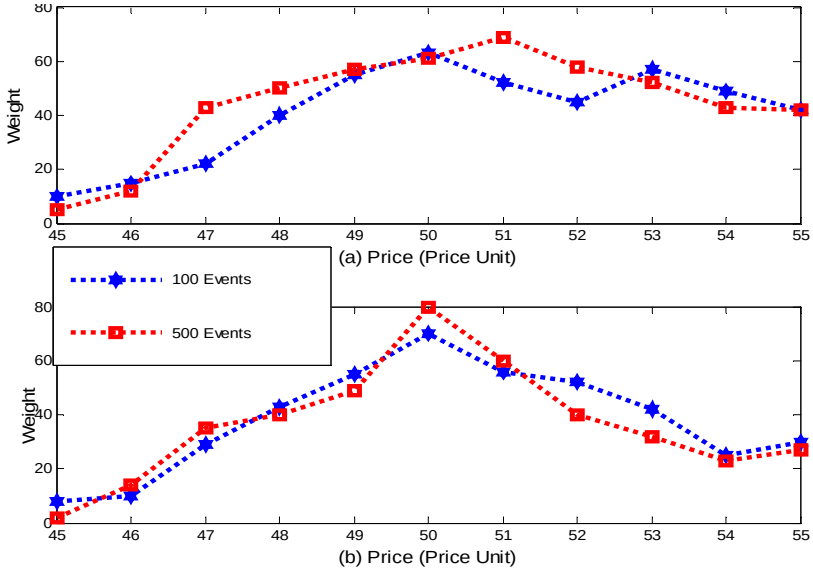

Figure 2.Bid Values for (a) QRL with 100 & 500 Events (b)BRL with 100 & 500 Events

References

[1] T. Jiang, D. Grace, Y. Liu, "Two-stage reinforcement-learning-based cognitive radio with exploration control," IET Communications, vol.5, no.5, pp.644,651, March 25 2011.

[2] A. Oloyede and D.Grace "Energy Efficient Soft Real Time Spectrum Auction for Dynamic Spectrum Access". 20th International Conference on Telecommunications Casablanca, May 2013.