Q-Learning Based Bidding Algorithm for Spectrum Auction in Cognitive Radio
Zhe Chen, and Robert C. Qiu, Tennessee Technological University

Introduction
Spectrum auction can be employed to allocate detected available frequency bands to secondary users (SUs). In this paper, a bidding algorithm based on Q-learning for SUs is proposed. SUs employ the proposed algorithm to learn from their competitors and automatically place better bids for available frequency bands. Simulation result shows the proposed algorithm is effective.

Q-Learning Algorithm

1. Let current state be s.
2. Choose an action a to perform.
3. Receive a reward r from performing action a and the resulting state s'.
4. Update Q(s, a) using (1).
5. Goto step 1.

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

(1)

Proposed Bidding Algorithm

\[ S = \{0, 1, \ldots, L_b \cdot L_c - 1\} \]  

(2)

\[ A = \{0, 1, \ldots, L_a + 1\} \]  

(3)

\[ x_{i,t} = \begin{cases} \left\lfloor \frac{b_i}{B} (L_b - 1) \right\rfloor & b_i < B \\ L_b - 1 & b_i \geq B \end{cases} \]  

(4)

\[ y_{i,t} = \max_l (s.t. \ g_{i,t} \geq C(l)) \]  

(5)

\[ \frac{g_{i,t}}{b_{i,t}} \]  

(6)

\[ h_{i,t} = \frac{1}{N - 1} \sum_{k=1}^{N} \sum_{j=0}^{L_d - 1} \Pr(p_{k,t} = j)(c_{k,t}) \cdot q_{k,j} \]  

(7)

\[ p_{k,t} = \max_l (s.t. \ d_{\text{MAX},k,t} \geq D(l)) \]  

(8)

\[ c_{k,t} = \max_l (s.t. \ c_{k,t} \geq E(l)) \]  

(9)

\[ q_{k,j} = \text{mean}(d_{\text{MAX},k,t}) \]  

(10)

\[ d_{\text{MIN},i,t} = \begin{cases} 0 & a_{i,t} = 0 \\ 1 & a_{i,t} = 1 \end{cases} \]  

(11)

1. Calculate the index of buffer fullness \( x_{i,t} \) using (4) and the index of credit ratio \( y_{i,t} \) using (5) (6) (7) (8) (9) (10). Then calculate state \( s_{i,t} \) using:

\[ s_{i,t} = x_{i,t} \cdot L_c + y_{i,t} \]  

(12)

2. Calculate the reward \( r_{i,t-2} \) for time slot \( t-2 \). If SU \( i \) wins the bidding in time slot \( t-2 \) and it successfully sends data in time slot \( t-2 \), then \( r_{i,t-2} \) is assigned with the amount of data that \( SU_i \) sends in time slot \( t-2 \). If all the conditions in (13) are met, then \( r_{i,t-2} \) is assigned with the negative amount of data that \( SU_i \) could have sent in time slot \( t-2 \).

\[ b_{i,t-1} - b_{i,t-2} \geq B \]  

(13)

\[ c_{i,t-2} \geq w_{\text{MIN},t-2} \]  

\[ d_{\text{MAX},i,t-2} \geq 0 \]  

where \( w_{\text{MIN},t-2} \) denotes the minimum winning bid among all the N SUs in time slot \( t-2 \), and \( d_{\text{MAX},i,t-2} \) denotes \( SU_i \)'s maximum bid in time slot \( t-2 \).

3. Update Q-table \( Q(s, a) \) using:

\[ Q(s_{i,t-2}, a_{i,t-2}) = (1 - \alpha)Q(s_{i,t-2}, a_{i,t-2}) + \alpha(r_{i,t-2} + \gamma \max_a(s_{i,t-1}, a_{i,t-1})) \]  

(14)

4. Calculate the number of SU \( i \)'s bids \( u_{i,t} \) in time slot \( t \):

\[ u_{i,t} = \min \left( \frac{b_{i,t}}{v_i} + 0.5 \right) \]  

(15)

where \( v_i \) denotes the amount of data that \( SU_i \) is able to send during time slot \( t \) using one frequency band, and \( U \) denotes the maximum number of bids that an SU is allowed to bid.

5. Choose an action \( a_{i,t} \) using:

\[ a_{i,t} = \arg \max \frac{Q(s_{i,t}, l)}{l = 0, 1, \ldots, L_a + 1} \]  

(16)

6. Calculate the minimum bid \( d_{\text{MIN},i,t} \) using \( a_{i,t} \) and (11).

7. Adjust \( u_{i,t} \) and \( a_{i,t} \). If \( a_{i,t} > 1 \) and (17) are both satisfied, then \( a_{i,t} = a_{i,t} - 1 \) and goto step 6. Otherwise, \( a_{i,t} = 1 \) and (17) are both satisfied, then \( u_{i,t} = u_{i,t} - 1 \) and repeat step 7.

\[ d_{\text{MIN},i,t} \cdot u_{i,t} + \frac{u_{i,t} \cdot (u_{i,t} - 1)}{2} > c_{i,t} \]  

(17)

8. Place \( u_{i,t} \) bids if both \( a_{i,t} > 0 \) and \( u_{i,t} > 0 \). The values of the bids are \( d_{\text{MIN},i,t}, d_{\text{MIN},i,t} + 1, \ldots, d_{\text{MIN},i,t} + u_{i,t} - 1 \).

Numerical Results

Conclusion

- A bidding algorithm based on Q-learning has been proposed.
- Simulation results show the proposed algorithm is effective.

Acknowledgments to National Science Foundation and Office of Naval Research.