A Q-Learning Based Adaptive Bidding Strategy in Combinatorial Auctions

Xin Sui, Ho-Fung Leung
Department of Computer Science and Engineering
The Chinese University of Hong Kong
Shatin, N.T., HK
{xsui, lhf}@cse.cuhk.edu.hk

ABSTRACT

Combinatorial auctions, where bidders are allowed to put bids on bundles of items, are preferred to single-item auctions in the resource allocation problem because they allow bidders to express complementarities and substitutabilities among items and therefore achieve better social efficiency. A large unexplored area of research in combinatorial auctions is the bidding strategies. In this paper, we propose a Q-Learning based adaptive bidding strategy in multi-round combinatorial auctions in static markets. The bidder employing this strategy can converge to the optimal state, and get a high utility in the long-term run. Experiment results show that the adaptive bidding strategy performs fairly well when compared to the optimal fixed strategy in different market environments, even without any prior knowledge.

Categories and Subject Descriptors

K.4 [**Computers and Society**]: Electronic Commerce; I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence–Intelligent agents, Multi-agent systems

General Terms

Game theory

Keywords

bidding strategy, combinatorial auctions, Q-learning

1. INTRODUCTION

Combinatorial auctions, where bidders are allowed to put bids on bundle of items, receive much attention from researchers in both computer science and economics. Combinatorial auctions can lead to more economical allocations of resources than traditional single-item auctions when bidders have complementarities and substitutabilities among them. Such expressiveness can lead to an improvement of efficiency, which has also been demonstrated in airport landing allocation and transportation exchanges[1].

Although lots of works have be conducted in combinatorial auctions recently, most of them focus on the winner determination problem [1] and auction design [1]. An large unexplored area of research in combinatorial auctions is the design of bidding strategies. As combinatorial auctions are incorporated with the first-price sealed-bid auction protocol in many applications [1], we are especially interested in bidding strategies in this kind of auctions. In this paper, we consider a scenario where first-price sealed-bid combinatorial auctions are employed to distribute computational resources among a group of users and propose a new adaptive bidding strategy. The bidder adopting this kind of strategy can transit among different states according to his bidding history, and thus perceive and respond to the market. Experiment results show that the adaptive strategy performs fairly well and generates high utilities in different markets when compared with the best fixed strategy which is obtained by prior knowledge.

This paper is structured as follows. Section 2 presents combinatorial auctions. Section 3 describes the adaptive bidding strategy. Section 4 shows simulation results. Section 5 concludes this paper.

ACM-HK Student Research and Career Day, 2009

2. COMBINATORIAL AUCTIONS

Suppose m type of resources are provided by a resource manager (auctioneer) to n users (bidders). For each type of resource $j \in \{1, 2, ..., m\}$, the capacity c_j denotes the total number of units that are available. There is a demand constraint $D = (d_1, d_2, ..., d_m)$, where d_j is the maximum number of units that each bidder can request for j. Each user i submits a sealed-bid $b_i = (T, p_i(T))$, where $T = (t_1, t_2, ..., t_m)$ is a resource bundle, with t_j being the number of units that resource j is requested by i, satisfying $0 \leq t_j \leq d_j$, $\forall j \in \{1, 2, ..., m\}$, and $p_i(T)$ is a positive number denoting the price i will pay for getting T. After receiving bids from all users, the resource manager solves the winner determination problem, that is to find an allocation which maximizes the auctioneer's revenue under the condition that the total number of units allocated cannot exceed their capacities. Each winning user i pays $p_i(T)$, gets accesses to the resources, performs his own task, and returns them to the resource manager. We refer to the process from the beginning of bid submission to the end of resource return as a *round*. Because resources are reusable, the combinatorial auction can be repeated for multiple rounds.

We list some assumptions used in this paper. First, the combinatorial auction market is static, which means the ratio of supplies and demands is constant. Second, each user submits a single bid per round, which means that no bidding language is used. Finally, each winner of the previous round submits a new bid, while each loser continues to submit the lost bid. However, a same bid will be dropped if it has been submitted for τ consecutive rounds.

3. THE Q-LEARNING BASED ADAPTIVE BIDDING STRATEGY

Based on Q-learning [2], we introduce some basic concepts used in the adaptive bidding strategy.

DEFINITION 1. A *bidding record* of a bid b for bidder i is a tuple $br_b = (T_b, v_i(T_b), p_i(T_b), pm_b, wait_b, win_b)$, where T_b is the requested bundle in b, $v_i(T_b)$ is i's valuation of T_b, $p_i(T_b)$ is bidder i's price for T_b, $pm_i = 1 - p_i(T_b)/v_i(T_b)$ is called i's profit margin, $wait_b$ is the number of rounds the bidder has kept on waiting before b is accepted or dropped, and win_b=1 if b is finally accepted, otherwise 0.

DEFINITION 2. A *bidding history with length* ρ, denoted as cbh^ρ, is the sequence of the most recent ρ bidding records. However, we say that it is *consistent* if and only if all bidding records share the same profit margin.

Suppose $\rho > 1$, if a bidder uses a fixed profit margin for all bidding records, then each history is consistent; if he never uses the same profit margin for two consecutive bidding records, then none of his bidding history is consistent.

During the auction, the bidder can change his profit margin by either increasing or decreasing, which will trigger the transition of the his state that a new profit margin will be used for the following rounds until the next change.

DEFINITION 3. A *state* of a bidder, denoted by s, is the profit margin currently used by this bidder.

When reaching the state of s, the bidder need to determine an action on how to change the profit margin. A action will change his profit margin but still in (0,1).

DEFINITION 4. A *action* of a bidder at the state of s, denoted by a, is a non-zero real number, by which his state will change from s to $s^* = s + a$ for the following consistent bidding history, with the constraint that $0 < s^* < 1$.

From the definition, we can see that every time when an action a is made at the state of s, the bidder will transit to a new state s^*, because a does not equal to 0.

Before transiting to a new state, the bidder can compute the reward of doing a at the state of s.

DEFINITION 5. The *reward* of a bidder when making an action of a at the state of s, denoted as $r(s, a)$, is defined as:

$$r(s,a) = s^* \times \frac{\sum_{br_b \in cbh^{\rho^*}} win_b}{\sum_{br_b \in cbh^{\rho^*}} (win_b + wait_b)} \quad (1)$$

where s^* is new state when choosing a at state s and cbh^{ρ^*} is the new consistent bidding history with length ρ^*.

3.1 Adaptive Strategy Algorithm

Based on the basic concepts defined above, we describe the adaptive strategy. We use s' and a' to denote the bidder's previous state and action, and use s to denote his current state. We also use r and r' to denote the reward obtained by the bidder when reaching the state of s' and s respectively. The basic idea of the adaptive strategy is that every time when a new cbh^ρ is formed, the bidder chooses the action that will maximize the quality of the state-action pair $Q(s, a)$, which is updated by the Q-learning rule [2]:

$$Q(s,a) = Q(s,a) + \alpha \cdot [r(s,a) + \beta \cdot \max_{a^*} Q(s^*, a^*) - Q(s,a)] \quad (2)$$

where s^* is new state when choosing a at state s, $0 \leq \alpha < 1$ is the learning rate and $0 < \beta < 1$ is the discount factor.

Algorithm 1 Adaptive strategy

```
1: S←{s_ini}, A←{+θ, −θ}
2: r' = 0, s' = s = s_ini, Q(s, −θ) > 0.
3: while auction does not finish do
4:     Keep the state of s
5:     if a new cbh^ρ is formed and θ > ε then
6:         r = r(s', a'), s' = s.
7:         Update Q(s', a') with equation 2.
8:         if Q(s, a') == 0 then
9:             Q(s, a') = r − r'
10:        end if
11:        if Decreaseθ() == true then
12:            θ = θ/γ
13:            if Q(s, θ) == 0 and θ × a' > 0 then
14:                Q(s, θ) = r − r'
15:            else
16:                Q(s, −θ) = r − r'
17:            end if
18:            if θ ∉ A then
19:                A ← A ⋃ {+θ, −θ}
20:            end if
21:        end if
22:        a = arg max_{a*∈A,|a*|==θ} Q(s, a*)
23:        s = s + a, r' = r
24:        if s ∉ S then
25:            S ← S ⋃ {s}
26:        end if
27:    end if
28: end while
```

The adaptive strategy is illustrated in Algorithm 1. Firstly, state set S and action set A are initialized with $\{s_{ini}\}$ and $\{+\theta, -\theta\}$ respectively, and then some variables used in the algorithm are also initialized. Every time when 1) a new consistent bidding history with length ρ is formed and 2) θ is greater than the threshold of ϵ, the bidder will transit to a new state: the bidder first computes the reward of the previous state-action pair $r(s', a')$ (line 6), updates the Q-value (line 7), and then chooses a new action 1) that will maximize the state-action pair at s and 2) whose absolute value equals to θ (line 22). During the repetition, there will be new states and actions generated (line 19 and 25) ensuring that the bidder's state will converge to the optimal state, which will maximize the bidder's accumulated utility in the long-term run. The transition will be stopped when θ is smaller than a threshold ϵ and the bidder will remain at that state for all subsequent rounds until the auction finishes.

3.2 Function of Decreaseθ

The value of θ is decreased to make sure that the bidder's state can be more approached to the optimal state.

DEFINITION 6. The *state history*, which is denoted as sh, is a sequence of λ real numbers, in which the kth element, sh^k, is the bidder's kth most recent state.

DEFINITION 7. The θ *history*, denoted as θh, is a sequence of λ real numbers, in which the kth element θh^k, is the action used when when the bidder transits from sh^{k-1} to sh^k.

NOTATION 1. We say that $s \Rightarrow \pi$ if 1) $s < \pi$ and the next action of the bidder $a > 0$ or 2) $s > \pi$ and the next action of the bidder $a < 0$.

The function of Decreaseθ is given in Algorithm 2. At first, we compute the mean value of the elements in sh (line 1), then for each element we check whether the distance between sh^k and $mean$ is no more than θh^k and use a 0 or 1 variable ω^k to indicate the result (line 2 to 7). On deciding whether to decrease θ, we check three conditions (line 8): the first one checks whether at least ϕ elements in sh are close to $mean$ in terms of the action chosen then, by which we regard $mean$ as an approximation of the optimal state, and the second and the third ones together guarantee that the optimal state can be further approached if θ is decrease. If all conditions hold, $true$ is returned.

Algorithm 2 Function: DecreaseΘ

1: Compute $mean = \frac{1}{\lambda}\sum_{k=1}^{\lambda} sh^k$.
2: **for** $k = 0$ to λ **do**
3: $\quad \omega^k = 0$
4: $\quad$ **if** $|sh^k - mean| \leq \theta h^k$ **then**
5: $\quad\quad \omega^k = 1$
6: $\quad$ **end if**
7: **end for**
8: **if** $\sum_{k=1}^{\lambda} \omega^k \geq \phi$ and $\omega^1 = 1$ and $s \Rightarrow mean$ **then**
9: $\quad$ return $true$
10: **end if**

4. SIMULATION RESULTS

We compare the performances of the random strategy (RS), the adaptive strategy (AS) and the best fixed strategy (BFS) by the accumulated utility of the test bidder using them in markets. Random strategy is a strategy that bidder transits randomly among states in different bidding records. Best fixed strategy is the one performs best among a set of fixed strategies, where the test bidder remains at a state during the process of the auction. We refer to the state of the best fixed strategy as the the best fixed state.

We use the ratio of total supplies and demands to denote a market type and a market is said to be a 1:n market if such ratio equals to 1:n. In our experiments, we compared three strategies in four types of market, which are 1:0.75, 1:1, 1:1.25 and 1:1.5 respectively.

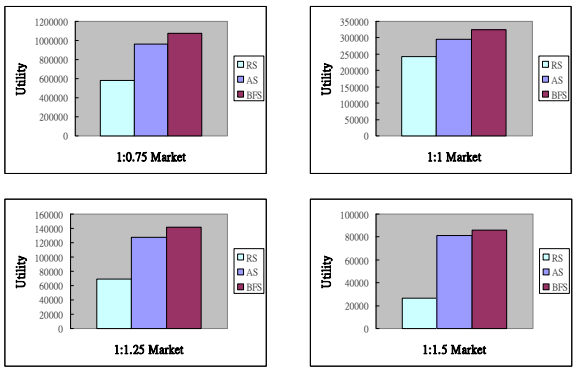

Figure 1: Utilities obtained using RS, AS and BFS

Figure 1 shows the simulation results. We can see that the adaptive strategy performs well when compared to the best fixed strategy and the random strategy in different markets. The bidder using the best fixed strategy can be regarded as having prior knowledge about the market and is able to keep the best fixed state to obtain a high utility; while the bidder using the random strategy can be regarded as not having any prior knowledge about the market and will transit randomly among different states. Therefore, it is impressive that the bidder using the adaptive strategy can still get a utility that is about 90% of the utility obtained by the bidder using the best fixed strategy in each market type.

We also show the transition of states in different markets. The red lines show the best fixed state in that market.

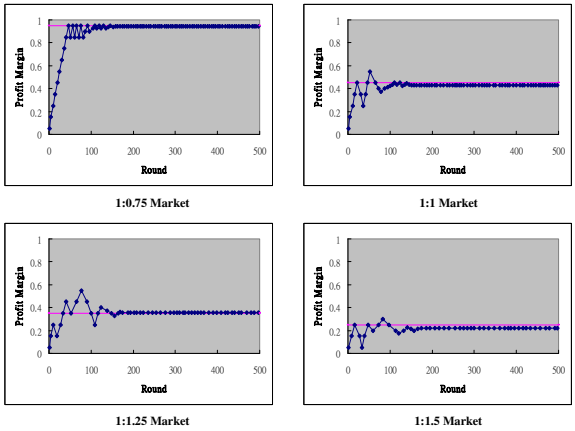

Figure 2: Transition of states in different markets

From Figure 2, we can see that for each type of market, the bidder using AS always converges to the best fixed state. In addition, the convergence speed is fast: for each market type, the bidder's state converges quickly to the best fixed state. The state is kept by the bidder to bid in subsequent rounds, which guarantees that the adaptive strategy can generate a high utility.

5. CONCLUSIONS

In this paper, we propose a Q-Learning based adaptive bidding strategy in combinatorial auctions. The bidder adopting this strategy can transit among different states according to bidding histories and finally converge to the optimal state. Experiment results show that 1) the adaptive strategy performs fairly well compared to the best fixed strategy and the random strategy in different market environments. 2) the bidder using the adaptive strategy can obtain a high utility, even without any prior knowledge about the market. 3) the bidder using the adaptive strategy is capable of adapting to the market and the convergence speed is fast.

6. REFERENCES

[1] P. Cramton, Y. Shoham, and R. Steinberg. *Combinatorial Auctions*. MIT Press, Cambridge, Massachusetts, 2006.

[2] C. Watkins. Learning from delayed rewards. In *PhD Thesis University of Cambridge, England*, 1989.