A Cooperative-Competitive Multi-Agent Framework for Auto-bidding in Online Advertising

Chao Wen, Miao Xu, Zhilin Zhang, Zhenzhe Zheng, Yuhui Wang, and etc.

Nanjing University of Aeronautics and Astronautics, Shanghai Jiao Tong University, Alibaba Group

chaowen@nuaa.edu.cn

1. Introduction

Background:

- In traditional bid optimization, advertisers need to manually adjust a bid in each ad auction to optimize the overall ad campaign performance.
- To reduce the burden on bid optimization for advertisers, online platforms have deployed various types of auto-bidding services, allowing advertisers to simply express high-level campaign objectives and constraints, and the auto-bidding agents would calculate the bids for each auction on behalf of advertisers.

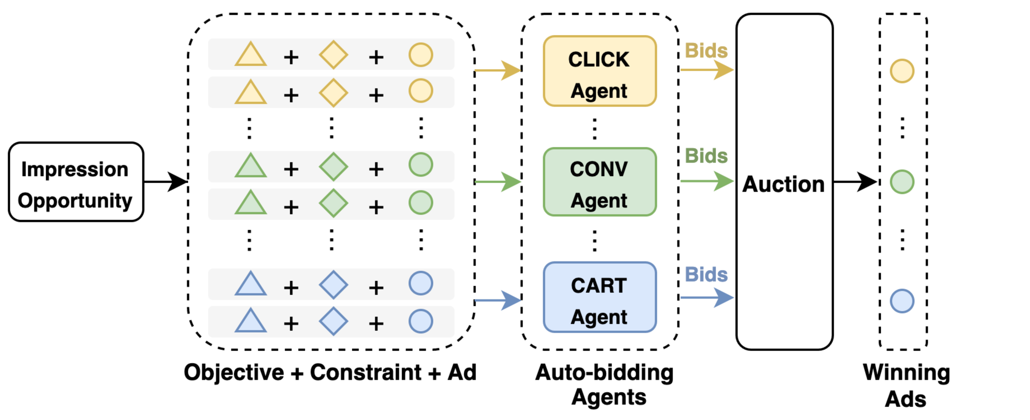


Figure 1: *An Overview of Auto-bidding Services.*

Problem Description:

- The goal of auto-bidding agent that bids on behalf of advertiser i is to maximize the total value of winning impressions under the budget constraint:

$$\max \sum_{t=1}^{T} v_i^t \times x_i^t \quad \text{s.t.} \sum_{t=1}^{T} p^t \times x_i^t \leq B_i \tag{1}$$

- We present a multi-agent framework for learning the bidding strategies for auto-bidding agents.

Motivations:

- The ad auction mechanism is inherently a distributed multi-agent system in nature.
- Appropriate coordination is needed to avoid an anarchy state with significantly degraded system performance.

2. Cooperative or Competitive? Neither!

We devise a two-agent bidding game, and experiment with both competitive method (CM-IL) and cooperative method (CO-IL). We vary the experimental settings with different total budget B_0 in each episode and the budget ratio r , which controls the percentage of the total budget to agent 1.

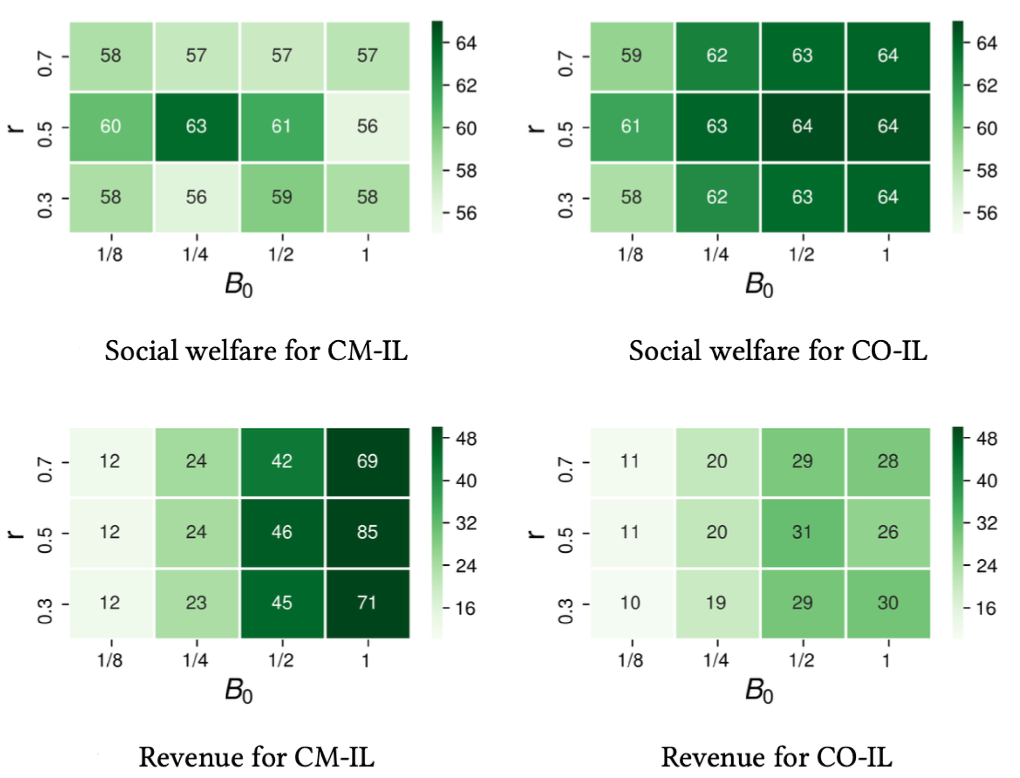


Figure 2: *behaviors of CM-IL and CO-IL.*

Results:

- CM-IL: low social welfare but high revenue. This is because the oligarch would bid aggressively to win all impressions.
- CO-IL: high social welfare but low revenue. This is because the cooperative agents can learn collusion behaviors, which encourage the agent to bid low prices.

We should make a proper trade-off between cooperation and competition.

3. Mixing Cooperation & Competition

To model the mixed cooperative-competitive relation, we propose temperature-regularized credit assignment (TRCA).

Main ideas:

$$r_i^{\text{TRCA}} = \alpha_i \times r^{\text{tot}}, \tag{2}$$

where $\alpha_i = \frac{\exp\{b_i/\tau\}}{\sum_{j=1}^{n}\exp\{b_j/\tau\}}$ is a softmax-style weighting parameter that satisfies $\alpha_i \in [0,1]$ and $\sum_{i=1}^{n}\alpha_i = 1$.

- The main idea is to set a parameter α_i weighting each agent's contribution to the total reward.
- parameter τ enables the co-existence of competition and cooperation, and works as a tool to make a trade-off between these two relations.

4. Improving Revenue with Bar Agents

Can we further improve the revenue? Yes. We introduce bar agents with different versions to achieve this.

1. Fixed bidding bar: similar to the reserve price, but needs to be tuned elaborately.
2. Adaptive bidding bar: 1) cannot avoid setting extremely high bidding bar; 2) a unified bidding bar for all may not be a good choice.
3. Multiple bar agents: our solution, which is adaptive and personalized.

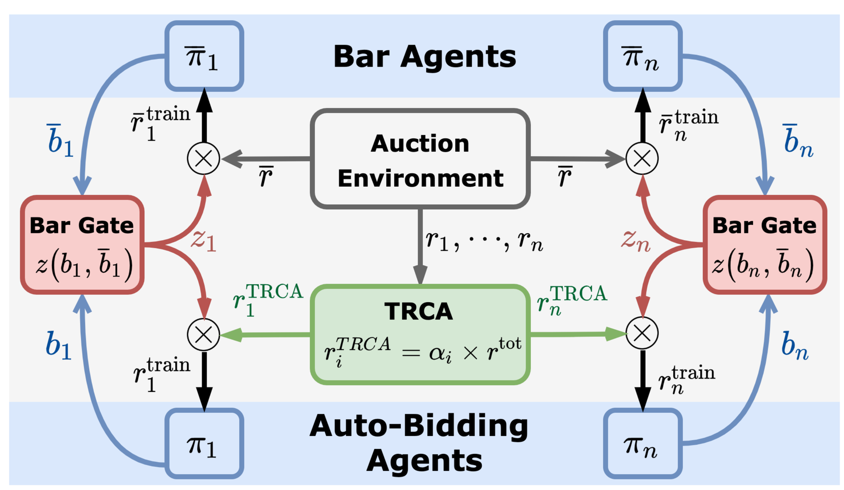


Figure 3: *The architecture of MAAB.*

Multiple Bar Agents:

- One bar agent for each auto-bidding agent. Each bar agent $\bar{\pi}_i$ aims at setting a personalized bar $\bar{b}_i$ for the corresponding auto-bidding agent π_i.
- At each timestep, bar agents and the auto-bidding agents give their bidding bars $\{\bar{b}_i\}_{i=1}^n$ and bids $\{b_i\}_{i=1}^n$, respectively. But only $\{b_i\}_{i=1}^n$ are submitted to the auction. Then the auction environment returns the payment p and the rewards $\{r_i\}_{i=1}^n$. The rewards $\{r_i\}_{i=1}^n$ are re-assigned by TRCA, obtaining $\{r_i^{\text{TRCA}}\}_{i=1}^n$.
- We introduce a **bar gate** to avoid setting extremely high bidding bar:

$$z(b_i, \bar{b}_i) = \begin{cases} 1 & \text{if } b_i \geq \bar{b}_i, \\ 0 & \text{otherwise }. \end{cases} \tag{3}$$

With the bar gate, the rewards for optimizing π_i and $\bar{\pi}_i$ are $r_i^{\text{train}} = z_i \times r_i^{\text{TRCA}}$ and $\bar{r}_i^{\text{train}} = z_i \times p$, respectively. The bar gate connects social welfare and revenue by enforcing the bar agent's bidding bar to be a maximum lower bound of auto-bidding agents' bid.

5. Modeling Large-Scale Multi-Agent System

To make our multi-agent approach practical in real system, we propose a mean-agent approach.

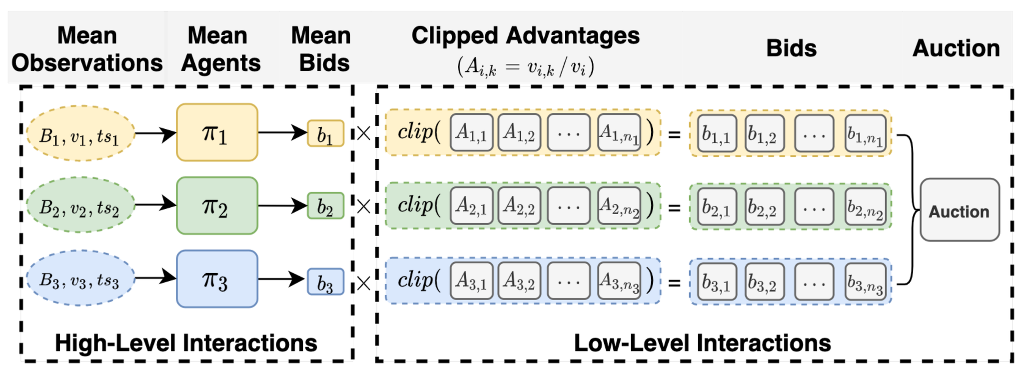


Figure 4: *Modeling with mean agent approach*

Challenges:

- There are millions of advertisers in Taobao, making it hard to train with limited computational resources and time.
- The sparsity of the rewards.

Our approach:

- We first group advertisers by their objectives, then the rewards are no longer sparse at the group-level.
- For each group, we train a mean policy π_i that calculates the mean bid based on the mean value and budget, and let each advertiser within the group derive her bid based on her value's *advantage* over the mean value.

6. Experiments

We evaluate our method in an offline industrial dataset and perform an online A/B test on the Alibaba e-commerce advertising platform.

6.1 Main Results

Table 1: *Mean and standard deviation of different groups' values (CLICK, CONV, CART), platform's revenue, and social welfare in offline dataset simulation.*

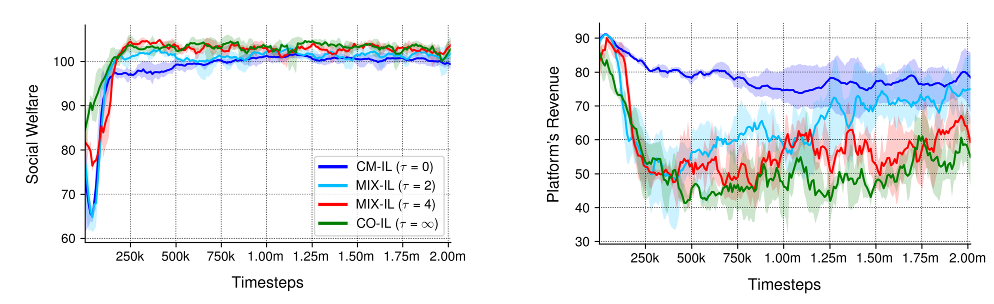
Setting 1	CLICK	CONV	CART	Revenue	Social Welfare
MSB	24.7±0	21.8±0	18.0±0	16.9±0	64.5±0
DQN-S	**29.3**±2.7	35.8±5.1	**36.0**±2.3	68.3±6.7	101.0±2.5
CM-IL	27.8±0.9	41.3±0.7	35.0±0.8	**86.8**±1.2	104.1±0.8
CO-IL	27.3±1.5	41.3±2.0	**35.6**±1.7	66.9±10.2	104.3±2.3
MAAB	28.0±0.8	**41.8**±1.3	35.5±1.4	80.6±3.2	**105.3**±1.3

Table 2: *Mean of different groups' values (CLICK, CONV, CART), platform's revenue and social welfare in the online production environment.*

	CLICK	CONV	CART	Revenue	Social Welfare
CM-IL	31.4	48.2	20.4	**100.0**	100.0
MAAB	**32.9**	**50.3**	**21.4**	96.1	**104.6**

6.2 Ablation Study

Effectiveness of TRCA



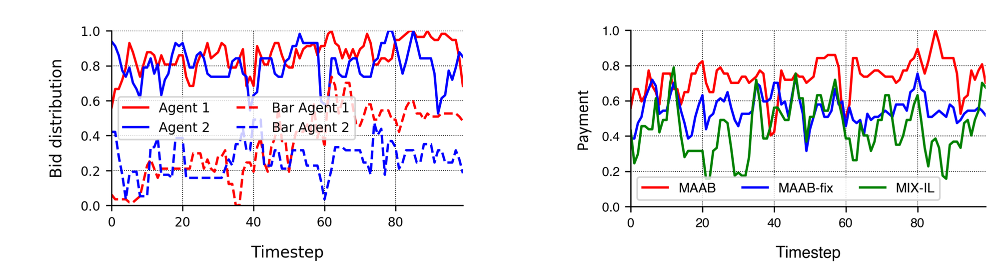
(a) Social welfare (b) Platform's revenue

Figure 5: *Social welfare and platform's revenue for methods with different parameter τ. The mean and 95% confidence interval are shown across 3 independent runs.*

Influence of Bar Agents

Table 3: *Mean and standard deviation of social welfare and platform's revenue.*

	Social Welfare	Platform's Revenue
MIX-IL	104.0±3.3	99.6±18.2
MAAB-fix ($\bar{b}=1$)	**104.5**±1.4	114.3±17.1
MAAB-fix ($\bar{b}=4$)	99.3±0.4	**164.9**±1.3
MAAB	103.9±1.2	134.6±8.2



(a) Agents' bids for MAAB (b) Payments

Figure 6: *Agents' bids for MAAB and the payments for three methods across each timestep of a selected episode.*

7. Conclusions

- We propose a multi-agent approach to solve the auto-bidding problem.
- Proposing TRCA for establishing a mixed cooperation and competition relation among agents.
- Designing bar agents and a reward scheme, called bar gate, for improving the platform's revenue with an adversarial training manner.
- Proposing a mean agent approach for the deployment of our methods on the large-scale advertising platform.