MANAGEMENT SUMMARY
Group 4

AUCTION

Using Reinforced Learning to get an Optimal Price

—

By Hannah Ying, Frederick Chan, Jomin John, Philippe Braum

Program:	MSc BIDS Campus Munich	**Supervisors:**	Dr. Andreas Widenhorn Christoph Starringer Stefan Merbele Richard Turinsky
Course:	Field Project Task 4		

1. Need / Problem

There are many auctions involved in BMW‘s business processes such as procurement of parts or raw materials, tenders bidding, advertisement space bidding to name only a few. In this project, we are trying to get an optimal price in simultaneous real time bidding with multiple agents. We focus on first-price sealed-bid auctions (FPSBA), simulating real world auctions in a simplified model with 5 agents and using reinforcement learning (RL) to achieve an optimal result.

A FPSBA is a common auction type. It is also called blind shooting because all bidders submit sealed bids at the same time, so no bidder knows the other participant's offer. The highest bidder wins the auction and pays the submitted price.

The result of this project can be further developed and adjusted to fit into real world situations, creating value for various business cases within BMW’s processes and activities.

2. Unique Solution

We are optimizing the bidding strategy using RL algorithms. The agents created are participants within the auction environment.

The auction design for the agents is competing to get a product in an auction and get a reward based on the outcome of their actions. Each agent has a private appraisal of the item between 0 and 11. The bid can freely be set by the agent. When the auction ends and if the agent wins, the reward for the agent is the difference between the winning price and the private value of the product (r = Pv - b). There is zero reward if the agent loses the auction. There is no auction fee.

In reality, people’s payoff or reward are not only related to people’s actions but also related to what other people are doing. If an agent is capable of ranking the outcomes of each possible action and also able to calculate the possible outcomes for a given set of consecutive actions, or is able to select the most preferred outcome given the actions of the other bidding, the chance of winning in the auction will be higher. By training with the RL algorithm, the agent can have the best response based on what the agent thinks or believes the other counterpart might do.

3. Proof

The results of our model demonstrate that the mechanics of auctions and offer generation towards suppliers (private / sealed-bid auctions) can be learned and applied by the machine learning algorithm.

Building upon an implementation of an environment for FPSBA from the OpenSpiel library and configuring it to use up to 10 agents for the simulation creates a realistic representation of real-life auction situations. This sets the foundation of the model training and applied auction logic.

Through reinforced, unsupervised learning the model learns by random exploration in self-play which price leads to the highest reward (lowest price for given item value). This approach has the added benefit of the algorithm learning by itself without the need for supervision or definition of a policy which the agent acts on and results in reduced development costs, time and complexity.

Our results show that the trained agent outperforms randomly acting agents reliably (see presentation slides 15-20). The Q-Learning algorithm allows for an efficient (short time for training, reliable results) agent performance in the simulations, outperforming random acting agents in every scenario over time. This performance advantage applies to both, cumulative auction wins and cumulative auction rewards.

Another more complex machine learning algorithm, Deep Q-Networks (DQN) also shows successful results in agent learning progress, outperforming randomly acting agents even more efficiently than the Q-Learning algorithm.

This algorithm however is more complex in its setup and optimization and takes more time to train which in turn makes development and maintenance more resource consuming (calculating power, time, development costs).

The difference between both algorithms in terms of efficiency and success rate shows in the mean reward per episode.

The mean reward per episode should be close to the reward for the optimal bidding strategy. The optimal bidding strategy yields a reward of 1. A reward of 1 results from the bid being 1 below the private appraisal of the item value (reward = privateAppraisal - bid). The best strategy is therefore bidding just 1 below your appraisal to maximize the probability of winning the auction with the highest bid, but not bidding the appraisal itself, because the resulting reward in the case of winning the auction is equivalent to losing it (not taking into account the actual item which comes with winning the auction).

The results are 0.6 (24000 training episodes) for the Q-learning and 0.939 (2000 episodes training) with a maximum value of 0.983 (20000 training episodes). The closer the mean reward per episode is to 1, the better the agent’s performance in winning auctions reliably.

4. Return on Investment

The short-term investment into the development of the RL model can yield significant improvements to the resource procurement success and cost reduction through the attainment of better prices and automated processes.

Long-term maintenance costs will be low. With an increasing time that the system is deployed the results will become even better as the agent learns from past actions and offers exponentially increasing returns on investment.

5. Conclusion

In Today's era of machine learning and artificial intelligence on the rise it is inevitable to adapt this set of technologies to shape and develop the business in the digital age. We have used Q-learning and Deep Q-Networks RL algorithms to build self-learning agents to win FPSB auctions.

This project is based on real-life auction scenarios at BMW. We used a reinforcement, unsupervised learning algorithm to train the agent to identify the optimal price in simultaneous bidding with multiple competitors.

We run the script successfully and we can conclude the outcome as the reinforced unsupervised model which can learn by random exploration in self-play and can optimize highest reward with minimum error. Except for the disadvantage that the model relies on the chances due to seal based auction logic, the model can ensure significant cost reduction in the procurement process of the company. As it is an unsupervised learning model, it is possible to keep the maintenance cost also low and achieve a considerable return on investment.

6. Comments

It should be noted that due to the nature of the chosen auction model (FPSBA) there are limitations to the outcome of auctions which partly rely on chance due to the sealed-bid auction logic:
Since the results of the private valuation of each agent are randomized (probability uniformly distributed over interval 0..11), the trained agent may lose by chance or the other way around, the random agent may win by chance, resulting in an increased reward per win in comparison with the trained agent, as the random agent wins by chance with a by chance lower bid in comparison with the private valuation resulting in an increased average reward.

The trained agent wins more often by increasing the bid towards the private valuation which in turn decreases the reward (private valuation - bid), which is still larger 0 and therefore identified by the RL algorithm as desirable outcome over not winning the auction (reward 0).

Frequently bidding more closely to the private valuation therefore is identified as a winning strategy with decreased average reward.

Appendix

Codebase (Google Colab Notebook)	Presentation
https://colab.research.google.com/drive/18onYcRMLE8OZ2h45RX6bKcJN4gFXqzhF?usp=sharing	https://drive.google.com/file/d/1Q9exgvyVbI3jlOeK44aorSkVL2erQnXo/view?usp=sharing