Astrat3m: An agent submitted to the ANAC 2025 ANL league

Yunfei Wang[1*], Siqi Chen[1*]

[1]School of Information and Engineering, Chongqing Jiaotong University, China
[wangyf@126.com, siqichen@cqjtu.edu.cn]

May 27, 2025

Abstract

This report presents the astrat3m agent for the ANAC 2025 ANL league, developed by the ChongqingAgent team. The agent has a central - peripheral architecture. The central agent handles multi - round negotiations via dynamic programming and an opponent - checking mechanism. The peripheral agent uses a novel MCQ (TD3) algorithm, blending Twin Delayed Deep Deterministic Policy Gradient (TD3) and Mildly Conservative Q - Learning (MCQ), to generate bidding strategies and adapt to adversarial settings. In ANAC 2025 ANL league benchmarks, the central agent reached an average utility of 0.75, and the peripheral agent over 0.85. The report delves into the agent's design, core algorithms, and evaluation results.

1 Introduction

In the field of automated negotiation research, the coordination mechanisms of multi-agent systems and decision optimization have always been core challenges. Especially in complex negotiation scenarios with multiple issues and opponents, agents need to balance short-term gains with long-term strategies while adapting to dynamically changing opponent behavior patterns. The astrat3m agent proposed by the ChongqingAgent team aims to address these challenges through a hierarchical architecture design, separating global coordination and local decision-making into two functional modules: a central agent and peripheral agents.

The central agent employs dynamic programming strategies to build a negotiation combination optimization model and achieves global optimal control of multi-round negotiations through a utility evaluation mechanism. Peripheral agents integrate offline reinforcement learning with inverse utility mapping technology to ensure decision efficiency while enhancing strategy flexibility. Particularly, the designed SVM identification and adaptive response mechanism for highly adversarial negotiation environments gives the agent a significant advantage when facing complex opponents. This paper will detail the design principles of astrat3m.

2 The Design of astrat3m

2.1 Coordination Strategy of Central Agent

The central agent operates using a dynamic programming strategy. During negotiations, it can identify in real-time the peripheral agents involved and accurately infer details of completed and ongoing negotiations. It then analyzes which negotiation outcomes can yield the highest utility, creating a prioritized list. Based on this, the central agent proposes candidate quotes computed via dynamic programming to achieve the optimal agenda - item combination for the current negotiation state.

When negotiations hit snags, the central agent can adjust strategies flexibly based on the current situation and the utility list. An innovative dynamic generator mechanism effectively handles the exponential growth

*Equal Contribution

of the combinatorial space caused by increasing edges, topics, and topic selections. If the combinatorial space exceeds a set threshold, this mechanism activates automatically to find suitable quotes efficiently within resource limits. Meanwhile, an opponent - behavior checking mechanism dynamically calculates the expected utility of the opponent's current offer, generating a list of utility contributions from the opponent's quotes. A window mechanism ensures real - time monitoring of opponent strategies, allowing the central agent to adopt corresponding quoting strategies based on the opponent's behavior.

Upon receiving a proposal, the central agent first determines if the negotiation has concluded. If so, it updates its strategy immediately. If not, it checks whether the opponent's quote has been recorded and if accepting it can maintain high utility in subsequent negotiations. If conditions are met, it accepts the proposal. In the final negotiation stage, the central agent weighs the value of conceding to reach an agreement against the value of sticking to its strategy and potentially reaching no agreement. It dynamically selects the best strategy to ensure the negotiation concludes with an agreement. With these strategies and mechanisms, the average utility in large - scale negotiations with benchmark agents remains stable at around 0.75.

2.2 Bidding Strategy of Peripheral Agents

Peripheral agents also derive the most favorable negotiation combination list based on the current utility function and select the best negotiation information to propose. The bidding behavior here is based on our innovative MCQ (TD3) algorithm, which combines the MCQ algorithm and the TD3 algorithm and is an offline reinforcement learning algorithm used to train our agent models.

In the early stages, we collected extensive negotiation scenario information and opponent information, processed them into experience tuples MDP for offline reinforcement learning models, and conducted extensive training. This enables the developed agents to return effective negotiation bidding behaviors based on the current state. Then, through the developed inverse utility function, it maps back to real negotiation bids.

Furthermore, we have a high adversarial environment check mechanism. This mechanism collects historical negotiation information of the current environment and opponents, uses a trained support vector machine (SVM) model to identify whether it is a high adversarial environment, and if so, activates a model specifically trained for high adversarial environments based on offline reinforcement learning. This model is trained on extensive information from difficult negotiation scenarios and opponents and can quickly find appropriate bids within a limited time. It has performed well in bilateral environment tests. The inverse utility function design also collects real-time opponent information and prioritizes searching from bids of interest to the opponent to better reach an agreement and improve the utility scores of both parties.

2.3 Acceptance Strategy of Peripheral Agents

When a peripheral agent receives a proposal, it first determines whether the negotiation has ended and updates its strategy. If the utility ratio of the opponent's bid to its own maximum utility reaches or exceeds the set expectation value, it accepts the proposal. At the same time, the peripheral agent records its own and the opponent's historical bids, calculates the average utility of the last three bids, and compares it with the utility of the current bid. If the utility of the current bid is not lower than the average utility minus the interval threshold, it accepts the bid. Through these settings, when negotiating with benchmark agents in large-scale scenarios, the average utility is maintained above 0.85.

3 Evaluation

The central agent and peripheral agents of astrat3m were tested and evaluated through negotiations with benchmark agents in large-scale scenarios based on the runner. The central agent achieved an average utility of approximately 0.7, while the peripheral agent achieved an average utility of over 0.85. These results indicate that the agent can effectively handle various negotiation scenarios and opponents.

4 Lessons and Suggestions

Through the design and testing of astrat3m, we have gained some experience. For future work, we can further optimize the dynamic programming strategy of the central agent, explore more advanced algorithms to enhance the performance of peripheral agents, and collect more negotiation data to strengthen the training of agent models.

Conclusions

The astrat3m agent designed by our ChongqingAgent team adopts a two-part architecture of a central agent and peripheral agents. The central agent uses dynamic programming strategies to optimize negotiation combinations, while peripheral agents utilize the MCQ (TD3) algorithm and high adversarial environment check mechanisms to improve negotiation performance. Through runner-based testing, it has achieved good results in large-scale scenarios. In future work, we will continue to optimize the agent's strategies and models to enhance its negotiation capabilities.