Machine Learning for Negotiating Agents

Improving a Reinforcement Learning Negotiating Agent's Performance by Extracting Information from the Opponent's Sequence of Offers

Arpit Agrawal
a.agrawal-5@student.tudelft.nl
Supervised by
Bram Renting
Pradeep Murukannaiah
{b.m.renting, p.k.murukannaiah}@tudelft.nl

I. Background

Motivation:

- Search and rescue missions, self-driving cars, and medical decision-making aids → studies in cooperative collaborative agents.
- Agents must negotiate with each other to collaboratively agree on a solution that are better for all involved.

Research gap:

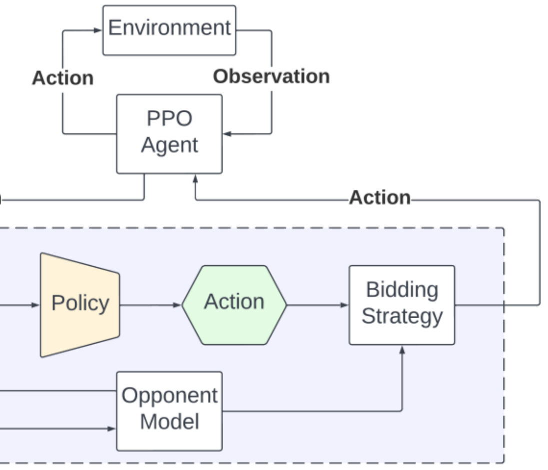
- Much of the current research focuses on competitive games, and not on cooperative settings.
- Agents have primarily been created with heuristic approaches, but Bagga et al. [1] show high potential for machine learning, particularly model-free deep reinforcement learning → Proximal Policy Optimization (PPO) [3]

Research question: Can a reinforcement learning negotiation agent's performance be improved with the information from the opponent's sequence of offers?

The opponent's sequence of offers have been shown to be important in an agent's strategy. [4]

II. Methodology

Figure 1: Overview of the PPO agent's main process



Implementation:

- Primary challenge – representing a sequence of offers as a fixed dimension of state space.
- Numerical measures allow a static description of the sequence [2].

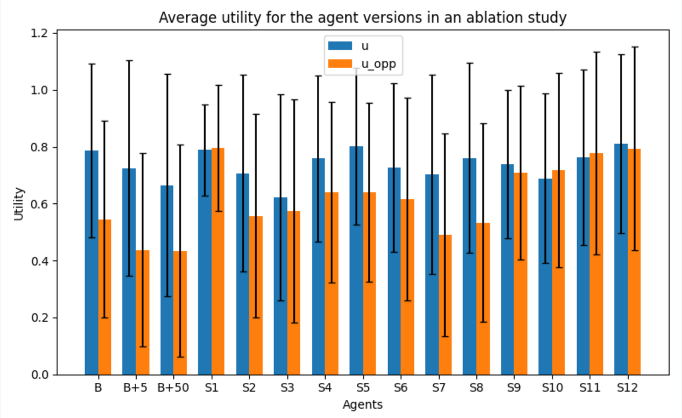
Dimension (L)	Sample Mean (μ)	Sample Standard Deviation (σ)	Sample Median (M)	Sample Mode (Mo)	Sample Range (R)	Correlation (ρ)

Experimentation:

- An ablation study is conducted to investigate the contributions of each measure.
- Using the GENIUS framework and 27 available existing agents from CSE3210 Collaborative Artificial Intelligence:
  - Pit the agents against the same sets of 17 agents to train, remaining 10 to test,
  - Domains and preference profiles are pseudo-randomly generated.
- Each version of the agent is trained 5 times for 6 hours, results are generated from the aggregation of their performances.

III. Results

Figure 2: Average utility for the agent versions in an ablation study.



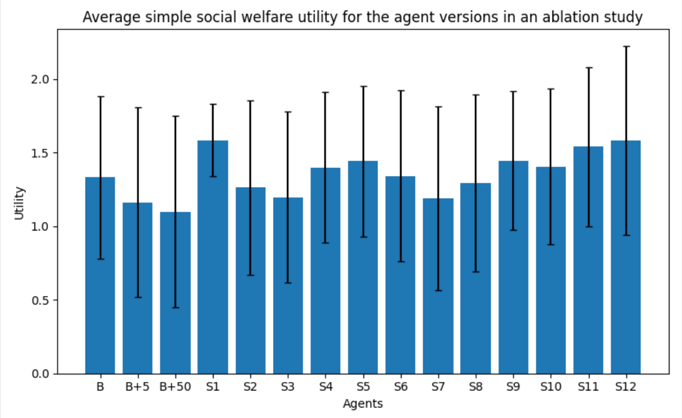
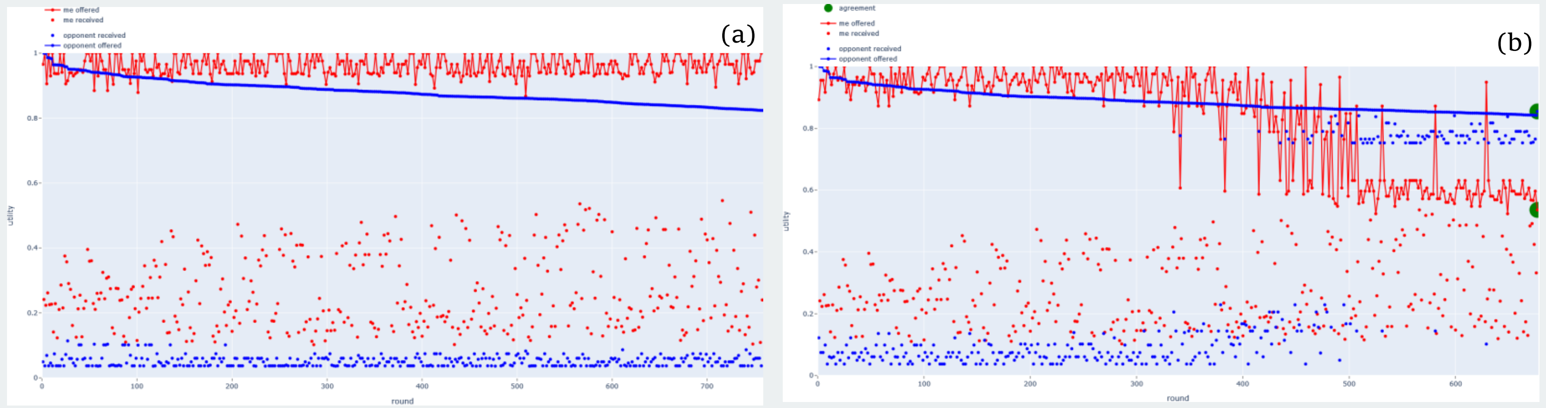


Figure 3: Average simple social welfare utility for the agent versions in an ablation study.

Simple social welfare:

$$u_{social} = u + u_{opp}$$

Figure 4: Comparison of the negotiation session Agents B (base) and the improved S12 had with Agent 78.



(a) Trace of the utilities in the negotiation between Agents B (red) and 78 (blue).

(b) Trace of the utilities in the negotiation between Agents S12 (red) and 78 (blue).

IV. Discussion

- During the ablation study, utility was not significantly impacted, but the opponent's utilities were positively affected.
- Agent S12, the final version with all numerical measures, including on the predicted opponent utilities, performed the best, especially when considering social welfare.
- Figure 4 gave insight into the improved performance in terms of opponent utilities → the opponent agents were only rewarded on an agreement. Agent S12's understanding of the opponent's nature made it more conceding in comparison to the base Agent B.
- Results show high variance within each agent version → difficult to make concrete analysis on the contributions of each numerical measure. Specific patterns and observations should be only considered critically.
- Future work could focus on using an ML algorithm to learn valuable features from the opponent's sequence of offers.

V. References

[1] Pallavi Bagga, Nicola Paoletti, Bedour Alrayes, and Kostas Stathis. A deep reinforcement learning approach to concurrent bilateral negotiation. *CoRR*, abs/2001.11785, 2020.

[2] DG Rees. Summarizing data by numerical measures. In *Essential Statistics*, pages 24–38. Springer, 1989.

[3] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *CoRR*, abs/1707.06347, 20

[4] Ayan Sengupta, Yasser Mohammad, and Shinji Nakadai. An autonomous negotiating agent framework with reinforcement learning based strategies and adaptive strategy switching mechanism. *CoRR*, abs/2102.03588, 2021.