Delft University of Technology

CSE3000 Research Project

Training a Negotiating Agent through Self-Play

Machine learning for negotiating agents

Renāts Jurševskis
(R.Jursevskis@student.tudelft.nl)

Supervised by
Bram Renting & Pradeep Murukannaiah

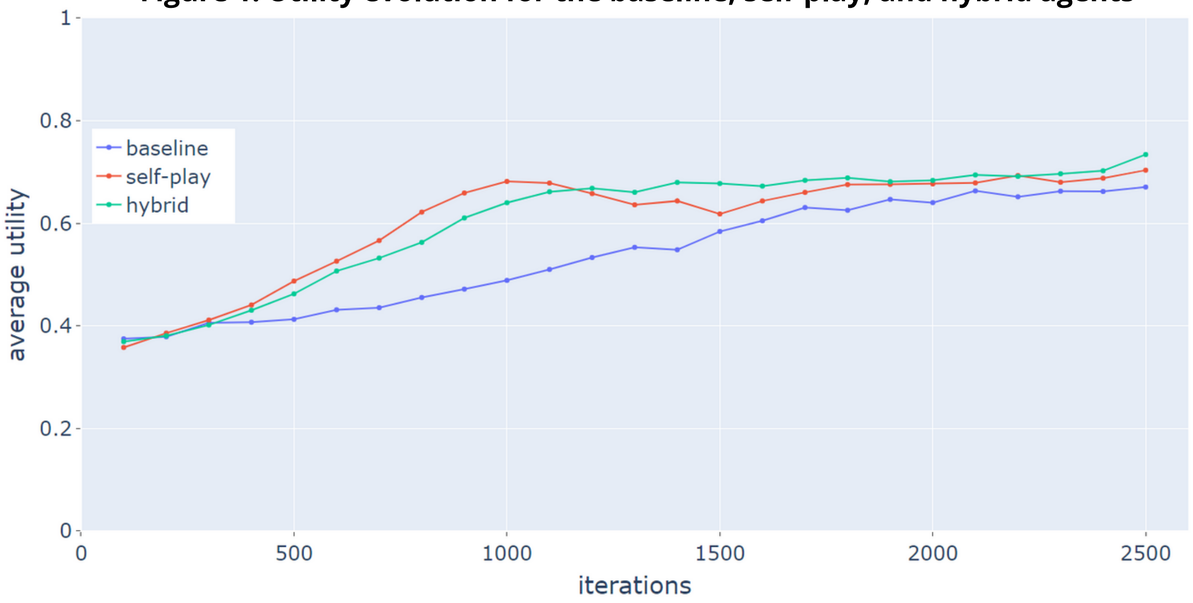
1. Background

- Negotiating agents allow for real-world applications of negotiation through automation
- Unlike purely competitive settings, negotiation encourages cooperation
- A bilateral negotiating agent can be implemented using a variety of methods, including SOTA reinforcement learning approaches like Proximal Policy Optimization (PPO) [1]
- By training through self-play, a set of training opponents is not required
- Self-play has displayed excellent results in competitive settings like chess or Go [2]

2. Research question

- How well can we train a negotiating agent without a default set of opponents through self-play?

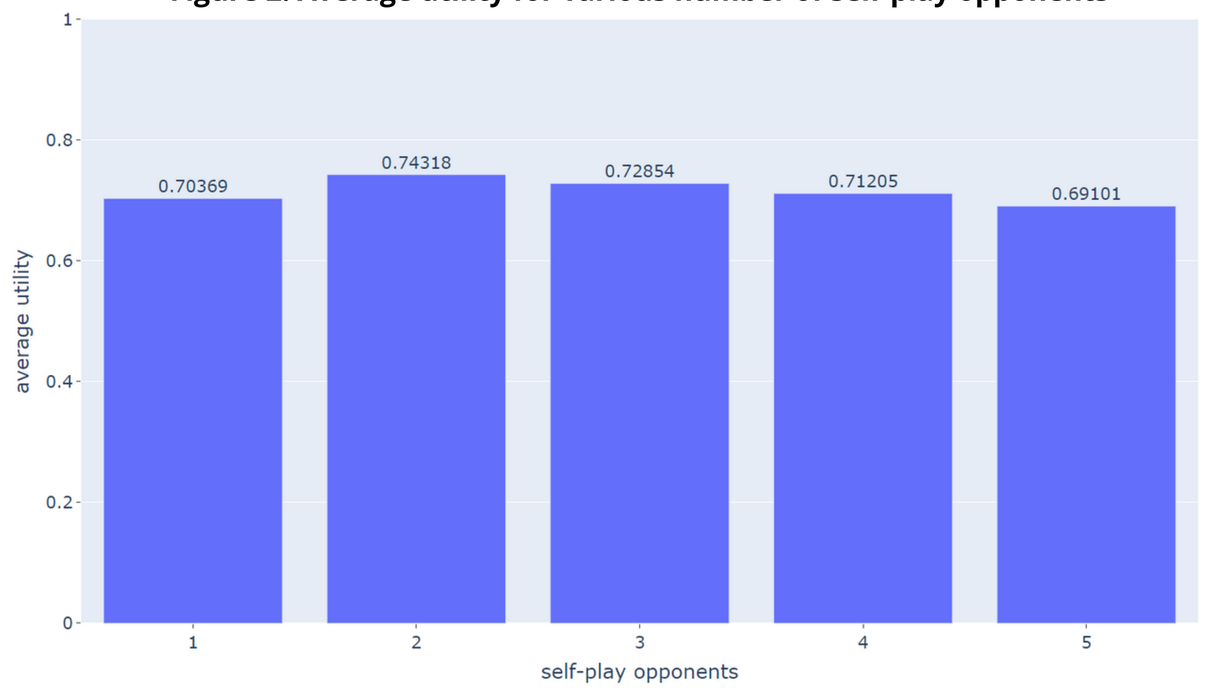
Figure 1. Utility evolution for the baseline, self-play, and hybrid agents



3. Methodology

- Modification of a baseline reinforcement learning (PPO) negotiating agent
- Implementation of a self-play opponent, which is used during training
- Evaluation of self-play compared to the baseline agent by testing against 10 opponents from CSE3210 and averaging the results over 10 repetitions
- Exploration of multiple self-play opponents for additional variance during training
- Creation of a hybrid training setup by combining self-play with a conventional approach

Figure 2. Average utility for various number of self-play opponents



4. Results

- The baseline agent only reaches a maximum utility of 0.671 (Figure 1)
- Self-play achieves a reward of 0.704 while making a larger impact on opponent utility
- Self-play results in a significantly lower percentage of agreements (81.2% against 98.8%)
- The use of multiple self-play opponents leads to a further increase in utility, reaching its peak of 0.743 with two self-play opponents (Figure 2)
- A hybrid approach outperforms both the baseline and self-play agent with a utility of 0.734

5. Conclusion

- Training a negotiating agent through self-play is viable and can match or even outperform a conventional approach using a set of training opponents
- The self-play agent is unwilling to concede, leading to a lower agreement percentage
- Using multiple self-play opponents during training can further improve the performance by introducing additional variance
- A combination of self-play and training against a set of default opponents can lead to a further increase in performance
- The addition of a self-play training stage can be further explored in SOTA agents

References

[1] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms, 2017.
[2] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, and Demis Hassabis. Mastering chess and shogi by self-play with a general reinforcement learning algorithm, 2017.