Learning in automated negotiation

Renting, B.M.

III

CONCLUSIONS

8

Conclusion

This dissertation investigated the design and evaluation of automated negotiation agents, with a particular focus on learning capabilities and performance across diverse negotiation settings. In this final chapter, we revisit and answer the research questions from Section 8.1. We discuss our results in Section 8.2 and discuss its limitations and suggestions for future work in Section 8.3.

8.1 Research questions & contributions

We now address the research questions from Section 1.2 per question. We will reiterate the questions and their sub-questions on the design of learned negotiation strategies and the evaluation of such agents, and answer them.

Research question 1

Q1 How do we design agents that can learn to negotiate?

- **SQ1.1** How can we reduce human-induced biases and conceptual complexity in learned negotiation strategies?
- **SQ1.2** Can we learn to generalise over diverse negotiation instances?

Our work has demonstrated significant progress in creating agents that can learn to negotiate, using a series of approaches, namely automated algorithm configuration (Chapter 3), portfolio selection methods (Chapter 4), and reinforcement learning (Chapter 5).

This dissertation makes a substantial contribution towards reducing the reliance on manual human design in developing negotiation agents. Approaches involving manual strategy design inherently introduce biases, for example, through the selection of specific algorithms or the creation of feature representations, which can limit agent capabilities or lead to information loss. By employing machine learning and optimisation methods, as explored in this dissertation, substantial aspects of the strategy design process can be automated, thereby mitigating these human-imposed biases. This leads to a reduction of human-induced biases in the negotiation strategies but it does not eliminate them entirely (e.g., the selection of the learning algorithm itself remains a human decision). Furthermore, integrating such learning methods into agents can increase the conceptual complexity of obtaining effective negotiation strategies.

8

In Table 8.1, we subjectively compare the human-induced bias and conceptual complexity of the methods proposed in this dissertation relative to each other. We also added manually designed strategies, which have a high degree of human-induced bias, as their entire logic is explicitly encoded by developers, limiting the strategy space explored. Their conceptual complexity is low, as such agents are generally simple heuristic-based strategies with limited capabilities.

The algorithm configuration approach (Chapter 3) reduces the bias somewhat but still has a medium-high degree of bias because the parameterised agent and the instance features guiding configuration remain manually designed. We classify

Table 8.1: Relative comparison of the design of negotiation agents presented in this dissertation.

Chapter	Main method	Human-induced bias	Conceptual complexity
-	Manual design	high	low
Chapter 3	Algorithm configuration	medium-low	medium-high
Chapter 4	Algorithm selection	medium-high	high
Chapter 5	Reinforcement learning	low	medium-low

its conceptual complexity as medium-low, due to the combination of the parameterised agent with the SMAC [75] automated configuration algorithm.

In Chapter 4, the portfolio selection method lowers the bias further to medium-high. While still utilising the parameterised agent, the portfolio construction (using Hydra [162]) and strategy selection (using AutoFolio [100]) broadens the strategy space of the agent. However, this layering of methods results in high conceptual complexity due to the need to manage multiple sophisticated methods.

Finally, the end-to-end reinforcement learning approach (Chapter 5) has a low degree of human-induced bias. By learning directly from negotiation interactions using GNNs, it avoids the manual design of feature representations, but still has some bias in the design of the policy architecture. We classify its conceptual complexity as medium-low, as it operates as a single, unified learning framework integrating various negotiation strategy challenges within the learned policy. We further discuss the conceptual complexity and human-induced bias in Section 8.2.

We have shown that it is possible to develop agents capable of learning to negotiate on a diverse set of negotiation scenarios and opponents, thus answering SQ1.2. This is a substantial improvement over previous attempts at learning negotiation agents that were trained and tested in strictly scoped settings, which hurts generalisability. We have shown that the performance of learned strategies transfers to unseen opponents in Chapter 3, Chapter 4, and Chapter 5. We have also shown that performance transfers to unseen negotiation scenarios both on composed sets (Chapter 3, Chapter 4) as well as on randomly generated scenarios (Chapter 5).

8

RESEARCH QUESTION 2

Q2 Is there a uniform way of evaluating negotiation agents?

SQ2.1 Is there a single-best metric?

SQ2.2 What is the value of the average utility metric for evaluating negotiation agents?

Our investigation into evaluation methodologies for negotiation agents, detailed in Chapter 6, revealed the limitations of relying on single-best metrics. We examined several commonly used metrics and found that agent rankings often varied depending on the specific criterion chosen (e.g., individual utility versus social welfare), demonstrating the lack of a single consistent metric. Critically, answering SQ2.2, the commonly used average utility metric in automated negotiation literature and competitions is not reliable, as the average utility cannot represent non-transitive rankings and is influenced by the composition of the group that the agent is evaluated in (Section 6.7.2). Both are undesirable properties in the context of automated negotiation.

In answer to SQ2.1, regarding the possibility of discovering a single-best metric, our work suggests this is improbable. Negotiation is complex, involving trade-offs between individual utility, social welfare, Pareto efficiency, and other goals. The observed non-transitivity implies that agent performance cannot always be linearly ordered, which is required for ranking by a single scalar metric. Therefore, the challenge might be less about finding the right metric and more about recognising the complexity of negotiation strategy evaluation.

8.2 DISCUSSION

With the research questions addressed, we now contextualise the findings of this dissertation, beginning with a discussion. In addition to the results obtained, there are further insights and lessons learned that are worth discussing. We discuss the challenges in building learning negotiation agents and the implications of achieving such agents. We also discuss the difficulty of evaluating negotiation agents.

REDUCING HUMAN-INDUCED BIAS

A central theme of this dissertation has been the progressive reduction of human-induced biases in negotiation strategy learning. Our trajectory from manually crafted parameterised negotiation strategies in combination with algorithm configuration (Chapter 3) and portfolio selection (Chapter 4) to end-to-end reinforcement learning (Chapter 5) exemplifies this principle. It is our opinion that human-induced bias through manual design should be limited. Such biases reduce the reachable subspace of the total negotiation strategy space that is used for learning/optimising a strategy. High-performing strategies might not even be part of the strategy space that is considered.

8

Chapter 3 and Chapter 4 both present a parameterised agent and negotiation instance features that were manually designed as prerequisites for the developed methods. Both were designed based on task-specific knowledge and past research, which gave them merit. As the findings in Chapter 6 exemplify, negotiation dynamics are so complex that these are not easily understood by human experts. Chapter 3 suffers most of this effect, as only a single fixed strategy is obtained. In contrast, Chapter 4 improves upon this by allowing for a portfolio of complementary strategies to be used, thus widening broadening the possible strategy space. However, we think that there is room for improvement in terms of further decreasing the degree of human-induced bias. We achieve this by removing manual feature design and using deep learning methods instead.

We can draw a parallel here with the development of the research areas of computer vision and natural language procession, where, in the early days, there was a large focus on manually designed features and tools such as SIFT features [101] and bag-of-words models [66]. With the advent of deep learning, such manually designed methods have been dominated by deep learning methods, which are capable of learning complex relations not captured through manual design. Our research shows that the automated negotiation research area can also benefit from making this step by demonstrating that agents can learn to negotiate without relying on manually designed features. The graph-based reinforcement learning approach presented in Chapter 5 represents a significant advancement in this direction, allowing negotiation strategies to emerge with minimal human-induced bias.

Conceptual complexity

In our definition, conceptual complex methods layer several learning methods, optimisation methods, or manually designed algorithms. Note that we deviate from the conceptual complexity definition that is used in psychology as a personality variable reflecting information processing tendencies [145]. Conceptual complexity generally makes methods both difficult to reproduce and hard to understand. From the perspective of Occam's razor, layering algorithmic methods increases the number of entities perhaps beyond necessary, whereas conceptually simpler, more unified approaches are preferable if they achieve comparable results. Conceptual simplicity, in line with the principle of parsimony (Occam's razor), is essential. It improves reproducibility and lowers the barrier for understanding, facilitating critique, extension, and further research.

The commonly used negotiation framework General Environment for Negotiation with Intelligent multi-purpose Usage Simulation (GENIUS) [99], many of the developed strategies in the automated negotiation community are hard to reproduce, especially the strategies that are learned or optimised. Many strategies entail various techniques stacked on top of each other (e.g. Sengupta et al. [137], Bagga et al. [18], and Chen et al. [29]). Part of this originates from the common practice of designing strategies to be modular (see, e.g., the BOA framework [13]), and part of it is due to the difficulty of dealing with differently sized scenarios, causing actions to be abstracted to utility goals commonly. This adds an additional search problem, as negotiation agents must find outcomes that fit the utility goals (also discussed in Chapter 5).

In retrospect, we also propose two conceptually complex methods in Chapter 3 and Chapter 4. Where Chapter 3 already combines algorithm configuration methods, a parameterised agent, an opponent preference estimation method, a random search method, a feature logging mechanism, and the negotiation platform GENIUS [99], Chapter 4 adds portfolio construction methods and an algorithm selection method. The work in both chapters can be seen as initial steps towards learned negotiation strategies, but it would be challenging to further advance those lines of research. The work in Chapter 5 represents a substantial step towards conceptual simplicity: It combines all the traditionally considered separate problems, such as opponent preference estimation, strategy learning, and outcomes space searching, into a single reinforcement learning problem on a clearly defined and close to raw observation/action space. Chapter 5 opens up many new avenues for further research, which we discuss in the next paragraph.

The discussed BOA structure aligns with the long-standing "divide-and-conquer" tradition in AI, i.e., to partition complex problems into more manageable subproblems. This makes such problems easier to understand and developed methods better explainable. However, the RL approach presented in this dissertation (Chapter 5) deliberately steps away from this, treating the negotiation problem holistically. This approach prioritises simplicity, in the form of a unified learning process, over the intrinsic explainability of divide-and-conquer approaches. The question is whether this trade-off is desirable. With the advancement of the automated negotiation research area, strategies are becoming more complex, challenging explainability regardless of whether a divide-and-conquer or holistic approach is used. Recent work, for example, took a divide-and-conquer approach while using RL to solve sub-problems [19, 18]. We argue that the research area might need to accept that negotiation strategies for complex negotiation tasks will likely be hard to explain regardless of the chosen approach. Therefore, stepping away from this divide-and-conquer approach and focusing research on conceptually simple methods is a potentially more productive path forward for the automated negotiation community, even if it leads to challenges in understanding the resulting agent behaviours.

REINFORCEMENT LEARNING FOR NEGOTIATION AGENTS

As discussed previously, the reinforcement learning-based strategy learning method presented in Chapter 5 is conceptually simple and less affected by human-induced bias. It integrates many of the problems faced in a negotiation game into a single-policy learning problem. This is in contrast with the complexity of the modular methods in the literature (see, e.g., [13, 19]), as we already discussed in Section 8.2. A further benefit is that more elaborate strategies can be achieved by extending the policy network, where, in modular approaches, this is achieved by adding another module to the agent. Learning to negotiate will then simply remain a single policy learning problem, which might be more difficult to learn, but only adds to the complexity by including more trainable neural network layers. We think that reinforcement learning is a fruitful direction for learning negotiation strategies based on this flexibility.

PROGRESS IN NEGOTIATION STRATEGY DEVELOPMENT

The lack of a universally applicable evaluation metric for negotiation agents remains a significant obstacle in the progress of the research area. Our analysis and discussion in Chapter 6 concluded that there currently is no clear evaluation method for negotiation agents and that the most commonly used metric, average individual utility, has problematic characteristics. The standard procedure of developing such agents is to take a set of top-performing agents of the Automated Negotiating Agents Competition (ANAC) or literature and to create an agent that outperforms them, often based on average utility. Such newly developed state-of-the-art agents are then included in benchmarks for future work. However, we discussed in Section 6.7.2 that average utility performance in a tournament with a set of benchmark agents depends on the composition of that set. Adding a well-performing agent to a benchmark set does not per se improve the benchmark but merely changes it. Any newly designed agent is, therefore, not strictly better, but just specialised on the new benchmark. We argue that not much progress is made in negotiation strategy design using this approach. As a community, we need to step outside the boundaries of evaluation methods and tasks that we have considered to make meaningful progress.

We think that research into automated negotiation application areas is the way forward. A clear task specification helps in defining the negotiation setting, such that performance criteria might change. In Chapter 7, we proposed multi-agent calendar scheduling as a concrete, real-world application area for automated negotiation. This problem encapsulates many of the challenges faced in negotiation research. It serves as an excellent example of a task-specific challenge that can potentially drive progress in the research area. We might find out that performance criteria are always different per agent, or that we should not care much about determining "winners" in groups of negotiation agents. We encourage the community to adopt this or similar challenge problems to drive progress, similar to how benchmark datasets like ImageNet [41] have driven progress in computer vision.

EMPIRICAL EXPERIMENTATION AT SCALE

The automated negotiation community needs to mature in its approach to large-scale experimentation. Our work, particularly in part one and Chapter 6, goes beyond what is commonly seen in the automated negotiation literature in terms of the scale of training and evaluation iterations. Learning negotiation strategies on large and diverse sets of negotiation instances is computationally intensive but also important in the pursuit of generalisability. This also shifts the burden from manual algorithm design to computational resources, aligning with the increasing availability of compute in recent years. The added benefit of such learning methods is the partial removal of human bias in the negotiation strategies. Furthering this line of research requires training at scale.

Scaling is a challenge in evaluating automated negotiation agents. The Automated Negotiation League (ANL) faces the curse of dimensionality in its tournaments. For instance, a bilateral tournament with 20 agents and 10 negotiation scenarios requires 1 900 evaluations, excluding repetitions commonly performed to minimise stochastic influence. To manage this, competition editions often adopted a two-phase approach: an initial group phase followed by a finalists phase. While this reduces the number of evaluations required, it introduces a new problem. As shown in Chapter 6, group composition significantly influences agent rankings. Consequently, the final rankings are affected by the initial, arbitrary group divisions, introducing undesirable inconsistency in the evaluation process.

This brings us to our statement that the automated negotiation community needs to mature in experimentation at scale and use the abundance of computing power that has become readily available in recent years. This is a prerequisite to improve the evaluation of negotiation agents, as well as to further the line of research into learning negotiation agents.

COOPERATION AND NEGOTIATION

Chapter 1 opened with the statement that human success is due to our ability to cooperate, which is partially enabled by our ability to negotiate. However, negotiation agents in the automated negotiation community often show limited cooperative behaviour. Many agents are designed to optimise their individual utility, which is supported by individual utility being one of the main evaluation criteria in the ANAC leagues (including the editions that we organised). The behaviour frequently comes at the expense of finding mutually beneficial outcomes that would maximise social welfare and causes many negotiations to end without an agreement. For example, approximately 11% of the negotiations ended in no agreement for the top performing agent in Table 6.4. This myopic approach is particularly problematic in mixed-motive scenarios where there is potential for joint gains. As we demonstrated in Chapter 6, in a tournament of competitive agents, the social welfare of the group is sub-optimal due to negotiations ending without an agreement. This is wasteful and undesirable for some potential applications, such as energy grid management, where prosumers must cooperate to distribute energy effectively.

How to move away from this competitive behaviour is, however, unclear. Agents can specifically be designed to optimise for social welfare, but that makes them sensitive to exploitation. After all, in a group of cooperators, it is often beneficial to be a defector [112]. This is already a challenge for groups of humans, but it is an even more significant challenge in groups of artificial agents, as they are insensitive to social constructs and norms. In other words, negotiation agents can afford to negotiate more aggressively than we think human negotiators will be comfortable with, which is a problem if cooperation is ultimately the goal.

8.3 LIMITATIONS & FUTURE WORK

Aside from the contributions of this dissertation to the research area of automated negotiation, there are also limitations to our work, as well as potential directions for future work. We discuss both in this section many of which follow from Section 8.2.

SIMPLIFICATION OF NEGOTIATION INSTANCES

8

Our studies primarily focused on bilateral negotiations with linear utility functions and categorical issues. While these scenarios are common in the literature and provide a good starting point, they do not capture the full complexity of real-world negotiations. Multi-party negotiations, negotiations with continuous issues, and scenarios with non-linear utility functions or interdependencies between issues were not extensively explored. This is a limitation as such additional complexity is common. For example, in the calendar scheduling task, there are often more than two parties involved in setting meetings, time is continuous, and issues such as location and availability can be connected.

LIMITED OPPONENT DIVERSITY

Although we used a variety of opponent strategies in our experiments, including those from ANAC competitions, this set may not fully represent the diversity of strategies encountered in real-world negotiations. Human negotiators, in particular, may employ tactics and exhibit behaviours that are not well-represented by our current set of artificial opponents. Evaluating against a broader range of opponents, including human negotiators, would provide a more comprehensive assessment of agent performance. As a result, the performance of the agents developed in this dissertation could be unpredictable in real-world situations. Future work could train negotiation agents using self-play, which avoids the requirement for a set of opponent agents and has seen previous successes in, e.g., the game of Go [142].

PREFERENCES OF AGENTS

We assumed that preferences were clearly defined in every negotiation scenario by means of the utility function. However, eliciting human preferences and capturing them in a model (or utility function) is difficult and a research area by itself [27]. The assumption of having such a clearly defined cardinal ranking over the entire outcome space is likely to be unrealistic. We also assumed that preferences won't change over time but are static through a negotiation session. This does not have to be the case, as negotiations might stretch out over longer periods when situations change.

Unclear preferences would add significant challenges to developing negotiation strategies, but should be studied in future work, as they are realistic. For example, in the calendar scheduling task, preferences are more typically expressed as partial constraints ("busy Tuesday morning"), ordinal statements ("prefer earlier slots"), or fuzzy priorities. Furthermore, the static preference assumption directly contradicts the dynamic reality of calendars, as availability changes and meeting importance shifts.

LIMITED EXPLORATION OF MULTI-AGENT DYNAMICS

Our research primarily focused on the performance of individual agents rather than the emergent behaviours in multi-agent systems where all participants are learning agents. The long-term dynamics and potential equilibria in such systems were touched upon in Chapter 6, but not extensively studied. This is an important area for future research, as it could reveal insights into the stability and efficiency of negotiation strategies in more realistic, adaptive environments.

LACK OF INTERPRETABILITY

Particularly in our end-to-end reinforcement learning approach, the learned strategies can be difficult to interpret or explain. This lack of interpretability could be a barrier to adoption in tasks where transparency and accountability are important. Future work should explore methods for making learned negotiation strategies more interpretable. While achieving interpretability for deep learning models remains an open research area across AI, progress is needed for reinforcement learning based negotiation agents. Future work could investigate techniques from explainable AI [108] to explore methods that extract simplified representations, such as knowledge technology-based models that capture simple aspects of the learned strategy. We realise that complete transparency might be unrealistic, but simplified high-level insight could be acceptable.

EVALUATION METRICS

Future work should focus on developing more robust and comprehensive evaluation metrics for negotiation agents. As demonstrated in Chapter 6, current metrics such as average utility have significant limitations, particularly in their sensitivity to opponent groups and inability to capture non-transitive relationships. Directions could be exploring multi-criteria evaluation frameworks that consider factors beyond just utility, such as fairness, stability of outcomes, and adaptability to different negotiation settings. This could provide a more holistic assessment of agent performance. Additionally, research into game-theoretic metrics that are more suitable for handling general-sum games could yield valuable insights.

COOPERATION

As discussed in Section 8.2, the level of cooperativeness is lacking in negotiation agents, as they are often designed to maximise individual utility without much risk for repercussions. Negotiation agents often have restricted communication capabilities, preventing them from engaging in the rich, nuanced dialogue that facilitates human cooperation. There is a lack of sophisticated mechanisms for building and maintaining trust, which is crucial for fostering cooperation, especially in repeated interactions or when dealing with partial information. There is a general lack of meta-strategic reasoning, with agents typically not engaging in higher-level thinking about the broader context of negotiations or the impact of their actions on future interactions. Note that implementing a memory mechanism to detect and reciprocate uncooperative behaviour might not be enough to enforce cooperative behaviour. Sometimes, individuals need to be excluded from the group based on the experiences of others to foster cooperation. Addressing these limitations represents a significant opportunity for future research to develop more cooperative negotiation agents. A potential starting point would be implementing reciprocity mechanisms that foster cooperation, such as described by Nowak [112].

REINFORCEMENT LEARNING IN NEGOTIATION AGENTS

The application of reinforcement learning in automated negotiation research remains relatively uncommon, with only a few notable attempts [19, 137, 18, 97]. While the challenge of managing variable-sized action spaces may have contributed to this scarcity, we addressed this issue in Chapter 5. The introduction of single policy-based agents now allows for the integration of multiple learning problems rather than the potentially noisy stacking of learning methods. This development opens up new possibilities and avenues for future work. Based on the policy proposed in Chapter 5, we suggest several further advancements.

- The policy network could be modified to include the full history in the observation (using, e.g., recurrent nets or transformers), potentially enabling it to identify and adapt to different opponent types, a capability we have shown to improve performance (Chapter 4).
- The loss function can include information that is typically unavailable during evaluation. For instance, the loss can be set to maximise social welfare, which requires knowledge of the opponent's utility function. This approach could help steer agents towards agreements with higher social welfare. The loss function can also be modified to explicitly incorporate the task of learning the opponent's utility function. Explicitly adding this task could improve the

agent's capabilities of finding mutually beneficial agreements. Note that this is only possible if we train the agent in simulations with full observability.

- To explore the possibility of handling continuous actions with our proposed policy network, the node outputs can be interpreted directly as bounded continuous distributions. Some negotiation objectives, such as price, are inherently continuous and would be better represented by continuous actions. This could also reduce the number of trainable parameters.

We think that the first point, including a history in the policy input, is a promising direction for future work. As demonstrated in Chapter 5, the reinforcement learning agent currently struggles to effectively adapt its strategy to diverse opponent types. Enabling the policy to learn from the history of interactions could directly address this limitation, which would be a significant step towards adaptive negotiation agents.

8.4 CLOSING REMARKS

In the introduction of this dissertation, we began by reflecting on Yuval Harari's insight from "Sapiens" that humans' unique ability to cooperate flexibly and at scale is what sets us apart and has led to our dominance as a species. Negotiation, as a special form of communication, plays a crucial role in enabling this cooperation. Throughout this work, we have explored how to create artificial agents capable of negotiating effectively and mirroring, in some ways, the development of this essential human skill. From the automated configuration of negotiation strategies to portfolio-based approaches and end-to-end reinforcement learning, we have taken notable steps to develop more flexible and adaptive negotiating agents.

As we conclude, it is worth considering how the advancement of negotiating AI agents might impact human cooperation and society at large. Just as the development of human negotiation skills enabled more complex forms of collaboration and social organisation, sophisticated AI negotiators could potentially facilitate new forms of AI-AI cooperation and AI-human cooperation at an unprecedented scale. However, as our analysis of evaluation metrics and the challenges in multi-agent dynamics reveal, there is still much to learn about the creation of effective negotiation strategies. As we continue to develop AI that can negotiate, we must ensure that it enhances rather than diminishes cooperative capabilities.