Scalable Solutions for Decision-Making Systems Using Explainable Policy Representations

Muqsit Azeem

Technical University of Munich
muqsit.azeem@tum.de

Abstract

Despite significant advancements in solving Markov Decision Processes (MDPs) and Simple Stochastic Games (SGs), scalability remains a challenge due to the exponential growth of their state spaces. This thesis aims to push the boundaries of state-of-the-art methods by tackling this issue using 1) explainability and 2) exploiting the model structure. First, we introduce the **1–2–3–Go** approach, which learns explainable policies from small MDP models and generalizes them to larger instances, improving scalability in MDPs. We then extend **Optimistic Value Iteration (OVI)** and **Sound Value Iteration (SVI)**—originally designed for MDPs—to SGs, improving efficiency in adversarial settings. Finally, we aim to exploit the **explainable policy representations** and **model structure** to enhance both scalability and interpretability in SGs. This thesis contributes to both theoretical advancements and practical solutions for decision-making systems under uncertainty.

Introduction and Problem Statement

Decision-making under uncertainty is a fundamental notion in domains like robotics, finance, and artificial intelligence. *Markov Decision Processes (MDPs)* model sequential decisions where outcomes are partly stochastic and partly under the control of decision-maker agent in non-adversarial settings. However, many real-world scenarios involve adversarial agents, where outcomes also depend on the actions of competing agents. *Simple Stochastic Games (SGs)* extend MDPs by incorporating adversarial decision-making, modeling interactions among two agents with opposite objective in a stochastic environment. A significant challenge in both MDPs and SGs is scalability. As the complexity of the models increases, the state space can grow exponentially, making traditional solving techniques computationally infeasible. This limits the practical applicability of these frameworks in complex real-world settings.

Related Work Scalability in MDPs and SGs has been a long-standing challenge, with various techniques proposed to address it. Some works include (Quatmann and Katoen 2018) on value iteration methods for MDPs, and extensions to adversarial settings such as in SGs (Eisentraut et al. 2022). These techniques often struggle with state-space explosion, which restricts their usefulness to large-scale problems. Additionally, explainability in decision-making models has gained attention in recent years, particularly in MDPs (Ashok et al. 2021) and POMDPs (Bork et al. 2024). This thesis builds upon these foundations by introducing the 1-2-3-Go approach (Azeem et al. 2024b), extending value iteration methods for SGs (Azeem et al. 2022; Azeem, Křetínský, and Weininger 2024), and aim to explore the novel integration of explainable policy representations. We aim to mitigate these scalability issues by developing approaches to improve both the efficiency and interpretability of solving techniques for MDPs and SGs. In particular this thesis aims to address the following research questions:

1. How can we improve the scalability of solving techniques for large MDPs and SGs?
2. What roles do explainability and model structure play in enhancing scalability in decision-making models?

To tackle these research questions, my thesis introduces the following approaches:

- The 1–2–3–Go (Azeem et al. 2024b) approach to improve scalability in MDPs by learning from small-scale models and generalizing policies to larger instances.
- Extensions of OVI (Azeem et al. 2022) and SVI (Azeem, Křetínský, and Weininger 2024) to SGs.
- Building on dtcontrol (Ashok et al. 2021), we aim for the development of explainable policies representations for SGs as in (Azeem et al. 2024a) for Partial Observable MDPs (POMDPs).

In summary, this research contributes new techniques that address the scalability and interpretability challenges in solving large MDPs and SGs. The proposed methods offer both theoretical advancements and practical implications for decision-making under uncertainty, expanding the applicability of these frameworks in complex environments.

Research Progress

1–2–3–Go Approach for MDPs

We developed the *1–2–3–Go* approach (Azeem et al. 2024b) to address scalability challenges in solving large-scale MDPs. This method synthesizes policies from small, representative instances and applies them to larger models by identifying and mapping corresponding states and actions with the help of decision trees (DTs). Similar to program verification techniques that generalize from specific proofs to broader proofs (Azeem, Madhukar, and Venkatesh 2018), 1–2–3–Go leverages structural patterns in smaller models to reduce computational complexity while maintaining solution quality. Initial experiments show promise for scalability improvements on a number of MDP benchmarks, though further research is needed to evaluate its generalizability across diverse benchmarks. This work lays the groundwork for exploring similar techniques in the context of SGs.

Extending OVI and SVI to SGs

In (Azeem et al. 2022), we extended Optimistic Value Iteration (OVI) and developed Precise Topological Value Iteration (PTBVI) to efficiently solve Simple Stochastic Games (SGs) with long chains of Strongly Connected Components (SCCs) by utilizing the structure of SGs. Further, we extended Sound Value Iteration (SVI) to handle SGs (Azeem, Křetínský, and Weininger 2024) which is theoretically involved. Additionally, empirical results confirm the advantages in scenarios with probabilistic loops, offering improvements in handling complex SG models. These extensions address the adversarial challenges of SGs and offer further directions for research. Currently, existing methods for solving SGs are often incomparable, with each excelling in different scenarios. The long-term goal is to develop unified approaches that effectively handle diverse structural complexities, ensuring robust performance across a wide range of benchmarks.

Explainable Policy Representations

The development of *explainable policy representations* is an important part of this thesis, particularly for Simple Stochastic Games (SGs). These representations aim to improve the interpretability of decision-making policies while maintaining scalability. Inspired by our recent work for POMDPs, where policies are represented using a combination of decision trees and finite state controllers (Azeem et al. 2024a), we aim to explore similar techniques to SGs. The focus is on creating representations that balance explainability and computational efficiency, thus enabling the solving of larger and more complex SG models. Although these techniques are still under development, their potential lies in offering a clearer understanding of decision-making processes and their structure while contributing to more scalable solutions for SGs.

Future Work

The next phase of this research will focus on having a deeper understanding of model structure and developing the explainable policy representations for SGs, ensuring that they maintain both scalability and interpretability across a broader range of benchmarks. Moreover, the integration of these techniques into real-world decision-making systems will be explored to test their applicability beyond theoretical models.

Anticipated Contributions

1. *Scaling Insights for SGs*: We aim to obtain deeper insights into how SG structures influence the performance of solving techniques, focusing on scalability.
2. *Explainable Policies for SGs*: We plan to use dtcontrol in order to obtain explainable policies for SGs where the idea is to use different trees for adversaries, contributing to the development of interpretable solutions.
3. *Evaluation of Scaling Techniques*: We plan comprehensive evaluations of these scaling techniques on diverse SG benchmarks, demonstrating improvements in both scalability and performance.

Conclusion

This thesis presents an approach to improve the scalability and interpretability of solving techniques for MDPs and SGs. Through the development of the 1–2–3–Go approach, the extension of OVI and SVI to SGs, and the introduction of explainable policy representations, the work aims to contribute to the field of decision-making under uncertainty. The proposed work not only address the current scalability challenges but also offer promising directions for future research in both theoretical and applied settings.

References

Ashok, P.; Jackermeier, M.; Kretínský, J.; Weinhuber, C.; Weininger, M.; and Yadav, M. 2021. dtControl 2.0: Explainable Strategy Representation via Decision Tree Learning Steered by Experts. In *TACAS 2021*.

Azeem, M.; Chakraborty, D.; Kanav, S.; and Kretinsky, J. 2024a. Explainable Finite-Memory Policies for Partially Observable Markov Decision Processes. *CoRR*, abs/2411.13365.

Azeem, M.; Chakraborty, D.; Kanav, S.; Kretínský, J.; Mohagheghi, M.; Mohr, S.; and Weininger, M. 2024b. 1-2-3-Go! Policy Synthesis for Parameterized Markov Decision Processes via Decision-Tree Learning and Generalization. *CoRR*, abs/2410.18293.

Azeem, M.; Evangelidis, A.; Kretínský, J.; Slivinskiy, A.; and Weininger, M. 2022. Optimistic and Topological Value Iteration for Simple Stochastic Games. In *ATVA 2022*, Lecture Notes in Computer Science, 285–302. Springer.

Azeem, M.; Křetínský, J.; and Weininger, M. 2024. Sound Value Iteration for Simple Stochastic Games. *CoRR*, abs/2411.11549.

Azeem, M.; Madhukar, K.; and Venkatesh, R. 2018. Generalizing specific-instance interpolation proofs with SyGuS. In *Proceedings of the 40th International Conference on Software Engineering: New Ideas and Emerging Results*, 57–60.

Bork, A.; Chakraborty, D.; Grover, K.; Kretínský, J.; and Mohr, S. 2024. Learning Explainable and Better Performing Representations of POMDP Strategies. In *TACAS 2024*.

Eisentraut, J.; Kelmendi, E.; Kretínský, J.; and Weininger, M. 2022. Value iteration for simple stochastic games: Stopping criterion and learning algorithm. *Inf. Comput. 2022*.

Quatmann, T.; and Katoen, J. 2018. Sound Value Iteration. In *CAV 2018*.