Towards Scalable and Robust Decision Making in Partially Observable, Multi-Agent Systems

Doctoral Consortium

Jonathon Schwartz
The Australian National University
Canberra, Australia
jonathon.schwartz@anu.edu.au

ABSTRACT

Designing autonomous agents that can interact effectively with other agents is an important problem in multi-agent systems. For real-world applications these agents must also be able to handle partial observability and scale to complex environments. We present two efficient planning algorithms for multi-agent, partially observable environments. The first, Interactive Nested Tree Monte-Carlo Planning (I-NTMCP), is a novel extension of Monte-Carlo Tree Search (MCTS) to Interactive Partially Observable Markov Decision Processes (I-POMDPs). Compared to existing methods, I-NTMCP is able to scale to significantly larger I-POMDP problems while modelling the other agent to deeper reasoning levels. The second algorithm, Bayes-Adaptive Partially Observable Stochastic Game Monte-Carlo Planning (BA-POSGMCP), combines a novel meta-policy with MCTS for scalable type-based reasoning. Through comprehensive empirical analysis in large cooperative, competitive and mixed domains we demonstrate that BA-POSGMCP is able to more effectively adapt online to diverse sets of agents in larger problems than previous methods. To support further research we have also developed *POSGGym*, an open-source library of multi-agent, partially observable environments supporting both planning and learning methods, along with *POSGGym-Agents*, a suite of policies for these environments.

KEYWORDS

Multi-Agent Systems; POSG; Planning Under Uncertainty; Agent Modelling; I-POMDP; Type-Based Reasoning

ACM Reference Format:
Jonathon Schwartz. 2023. Towards Scalable and Robust Decision Making in Partially Observable, Multi-Agent Systems: Doctoral Consortium. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

1 INTRODUCTION

A core challenge in the field of artificial intelligence is the design of autonomous agents that can interact effectively with previously unknown other agents [2, 26]. For the majority of real-world applications such agents must also be able to handle partial observability of the environment. To achieve this, an agent must be able to reason about the behaviours, goals, and beliefs of other agents, while simultaneously reasoning about the state of the environment. This requires both *agent modelling* [4] and *planning under uncertainty* [9, 10, 13]. Agent modelling involves constructing models of the other agents and using these models to inform decision making. While planning under uncertainty uses beliefs to account for partial observability and other sources of uncertainty in the environment. Considerable research has been done in both areas, and the combination of the two has great promise for the design of robust autonomous agents.

A key limitation of existing methods combining agent modelling and planning is their ability to scale to more complex domains. This restricts the problems to which these methods can be applied. Fortunately, in recent years there has been significant progress in scalable methods for planning under uncertainty, through techniques such as Monte-Carlo Tree Search (MCTS) [24]. Similarly, advancements in deep learning have lead to efficient reinforcement learning methods, which can be used to improve both agent modelling [7, 11, 15, 18] and planning [16, 23].

In our research we have focused on developing algorithms and techniques for efficient planning in partially observable, multi-agent settings. Of particular interest is the setting where an agent must interact effectively with previously unknown other agents across competitive, cooperative, and general-sum domains. This setting reflects many problems of interest in robotics and human-robot interaction (HRI), such as autonomous driving.

As part of our research we have developed two scalable planning methods for decision making in partially observable, multi-agent environments. Our first method, Interactive Nested Tree Monte-Carlo Planning (I-NTMCP) focuses on the challenge of constructing and using agent models within the paradigm of nested reasoning using the Interactive Partially Observable Markov Decision Process (I-POMDPs) framework [9]. Our second method, Bayes-Adaptive Partially Observable Stochastic Game Monte-Carlo Planning (BA-POSGMCP) focuses on efficient type-based reasoning [1, 5], where the planning agent must reason about a set of possible behaviours for the other agent. In addition to these two methods, we have developed *POSGGym* a library of environments supporting both planning and learning techniques, as well as a complementary library *POSGGym-Agents* containing policies that can be used for evaluating new approaches.

2 NESTED REASONING

The first part of our work [22] explored how to efficiently plan when the environment and other agent are modelled using an I-POMDP [9]. I-POMDPs provide a framework of recursive reasoning which allows an agent to explicitly model the other agents, which in turn model other agents, and so on down to some finite depth.

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

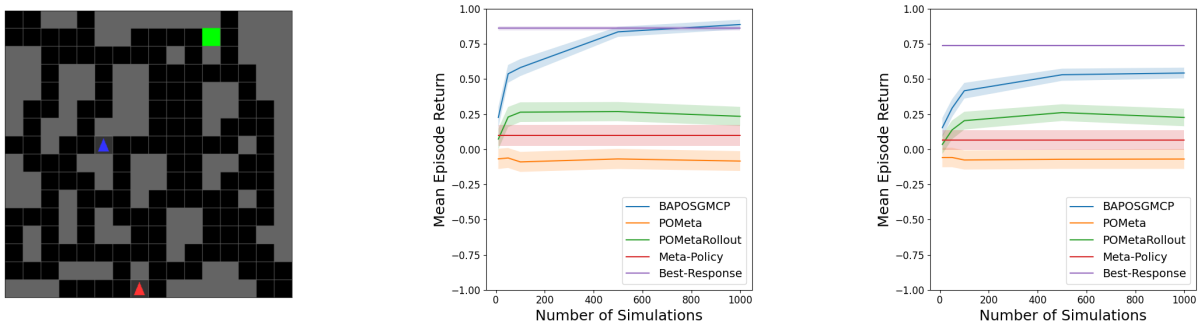

Figure 1: The Pursuit-Evasion domain (left) and the performance of BA-POSGMCP and baselines for the evader (middle) and pursuer (right).

This makes I-POMDPs ideal for modelling problems where there is uncertainty about the beliefs and reasoning of the other agents, with high application potential in HRI domains [17, 27, 29]. Unfortunately, the use of I-POMDPs has been restricted to relatively small problems due to their high computational complexity.

In order to improve the scalability of I-POMDP planning we developed I-NTMCP, an online MCTS-based planner. Compared to existing full-width planners, I-NTMCP focuses on computing the best action to perform from the planning agents current belief. This allows I-NTMCP to minimise the computation required for constructing the other agent's model, by focusing planning on the parts of the model that are most relevant for the planning agent's current belief. To achieve this I-NTMCP constructs and maintains a sequence of inter-related belief trees, where each tree encodes an approximately optimal policy for an agent operating at a particular nested reasoning level. This makes it possible for I-NTMCP to model the other agent in large problems without requiring a strong assumptions such as the other agent having full-observability [8, 12]. Our experiments demonstrated that I-NTMCP can generate substantially better policies up to more than 50× faster than I-POMDP Lite [12] – one of the fastest I-POMDP solvers at the time of publication. Further experiments showed that I-NTMCP can plan effectively in a complex domain with over 88K states and to much deeper reasoning levels.

3 TYPE-BASED REASONING

Moving beyond the nested reasoning paradigm, we next explored efficient methods for type-based reasoning in partially observable environments. Type-based reasoning methods give agents the ability to interact effectively with unknown other agents by maintaining a belief over a set of *types* for the other agents [1, 3, 5, 6, 25]. Each type completely specifies an agent's behaviour, making type-based reasoning very general, and applicable in numerous multi-agent domains. However, most existing type-based reasoning methods assume the planning agent has full observability of the state of the environment and the other agents' actions [4].

We proposed BA-POSGMCP [21] to address the lack of scalable planners for type-based reasoning in partially observable environments. BA-POSGMCP is inspired by ideas from empirical game theory [28] and the success of combining MCTS with a search policy [23]. Key to our method is the idea of reusing the set of types, i.e. policies, to help guide planning. To do this we introduced a novel meta-policy for selecting what policy from the set of policies to use for guiding search. This meta-policy was then integrated into MCTS using an extension of the PUCT algorithm [20, 23] to the multi-agent, partially observable setting. Through comprehensive evaluation in cooperative, competitive, and mixed environments - the largest of which has four agents and on the order of 10^{14} states and 10^8 observations - we demonstrated that BA-POSGMCP is able to adapt online and interact effectively without explicit prior coordination (Figure 1). We are currently extending this work with a theoretical analysis and additional comparisons against more recent methods [14].

4 ENVIRONMENTS AND POLICIES

As part of our ongoing research we have been developing a library of partially observable, multi-agent environments as well as a suite of reference policies. The environment library *POSGGym*[1] aims to provide a set of well tested benchmark environments with support for both planning and learning methods and in both discrete and continuous domains. This is in contrast to the majority of existing libraries which primarily support only reinforcement learning methods. Our policy suite *POSGGym-Agents*[2] provides a diverse set of policies for a number of the POSGGym environments, which can be used for reproducible evaluation of algorithms. Both libraries are under active development but are open-source and available for use now.

5 FUTURE WORK

While both I-NTMCP and BA-POSGMCP are general methods, they make strong assumptions about the other agent. I-NTMCP assumes the other agent is using a specific level of nested-reasoning. While BA-POSGMCP assumes the other agent's type is from a fixed and known set. To be truly robust, autonomous agents require the ability to generalize to a wide range of other agent behaviours, and be robust to out-of-distribution behaviours. An avenue of future work we plan to explore are methods for efficiently generating diverse agent models in complex environments and integrating them into decision-making. There has already been some work by others in this direction [19], and we are excited to see how these types of methods can lead to more robust and practical agents.

[1] github.com/RDLLab/posggym
[2] github.com/Jjschwartz/posggym-agents

ACKNOWLEDGMENTS

This work is supported by an Australian Government Research Training Program Scholarship and has been conducted in collaboration with many researchers, including my advisors Hanna Kurniawati and Marcus Hutter, and colleagues Ruijia Zhou and Rhys Newbury.

REFERENCES

[1] Stefano V. Albrecht, Jacob W. Crandall, and Subramanian Ramamoorthy. 2016. Belief and Truth in Hypothesised Behaviours. *Artificial Intelligence* 235 (2016), 63–94.
[2] Stefano V. Albrecht, Somchaya Liemhetcharat, and Peter Stone. 2017. Special Issue on Multiagent Interaction without Prior Coordination: Guest Editorial. *AAMAS* 31, 4 (2017), 765–766.
[3] Stefano V. Albrecht and Subramanian Ramamoorthy. 2014. On Convergence and Optimality of Best-Response Learning with Policy Types in Multiagent Systems. In *UAI*. 12–21.
[4] Stefano V. Albrecht and Peter Stone. 2018. Autonomous Agents Modelling Other Agents: A Comprehensive Survey and Open Problems. *Artificial Intelligence* 258 (2018), 66–95.
[5] Samuel Barrett and Peter Stone. 2015. Cooperating with Unknown Teammates in Complex Domains: A Robot Soccer Case Study of Ad Hoc Teamwork. In *AAAI*. 2010–2016.
[6] Samuel Barrett, Peter Stone, and Sarit Kraus. 2011. Empirical Evaluation of Ad Hoc Teamwork in the Pursuit Domain. In *AAMAS*. 567–574.
[7] Brandon Cui, Hengyuan Hu, Luis Pineda, and Jakob Foerster. 2021. K-Level Reasoning for Zero-Shot Coordination in Hanabi. *NeurIPS* 34 (2021), 8215–8228.
[8] Adam Eck, Maulik Shah, Prashant Doshi, and Leen-Kiat Soh. 2020. Scalable Decision-Theoretic Planning in Open and Typed Multiagent Systems. In *AAAI*, Vol. 34. 7127–7134.
[9] Piotr J. Gmytrasiewicz and Prashant Doshi. 2005. A Framework for Sequential Planning in Multi-Agent Settings. *JAIR* 24 (2005), 49–79.
[10] Eric A. Hansen, Daniel S. Bernstein, and Shlomo Zilberstein. 2004. Dynamic Programming for Partially Observable Stochastic Games. In *AAAI*. 709–715.
[11] He He, Jordan Boyd-Graber, Kevin Kwok, and I. I. I. Hal Daumé. 2016. Opponent Modeling in Deep Reinforcement Learning. In *ICML*. 1804–1813.
[12] Trong Nghia Hoang and Kian Hsiang Low. 2013. Interactive POMDP Lite: Towards Practical Planning to Predict and Exploit Intentions for Interacting with Self-Interested Agents. In *IJCAI*. 2298–2305.
[13] L.P. Kaelbling, M.L. Littman, and A.R. Cassandra. 1998. Planning and Acting in Partially Observable Stochastic Domains. *Artificial Intelligence* 101, 1-2 (1998), 99–134.
[14] Anirudh Kakarlapudi, Gayathri Anil, Adam Eck, Prashant Doshi, and Leen-Kiat Soh. 2022. Decision-Theoretic Planning with Communication in Open Multiagent Systems. In *UAI*. 938–948.
[15] Marc Lanctot, Vinicius Zambaldi, Audrunas Gruslys, Angeliki Lazaridou, Karl Tuyls, Julien Pérolat, David Silver, and Thore Graepel. 2017. A Unified Game-Theoretic Approach to Multiagent Reinforcement Learning. *NeurIPS* 30 (2017), 4193–4206.
[16] Adam Lerer, Hengyuan Hu, Jakob Foerster, and Noam Brown. 2020. Improving Policies via Search in Cooperative Partially Observable Games. *AAAI* 34, 05 (2020), 7187–7194.
[17] Brenda Ng, Carol Meyers, Kofi Boakye, and John Nitao. 2010. Towards Applying Interactive POMDPs to Real-World Adversary Modeling. In *Innovative Applications of Artificial Intelligence*, Vol. 24. 1814–1820.
[18] Georgios Papoudakis, Filippos Christianos, and Stefano Albrecht. 2021. Agent Modelling under Partial Observability for Deep Reinforcement Learning. *NeurIPS* 34 (2021), 19210–19222.
[19] Arrasy Rahman, Elliot Fosong, Ignacio Carlucho, and Stefano V. Albrecht. 2022. Towards Robust Ad Hoc Teamwork Agents By Creating Diverse Training Teammates. *arXiv preprint* (2022). arXiv:2207.14138
[20] Christopher D. Rosin. 2011. Multi-Armed Bandits with Episode Context. *Annals of Mathematics and Artificial Intelligence* 61, 3 (2011), 203–230.
[21] Jonathon Schwartz and Hanna Kurniawati. 2023. Bayes-Adaptive Monte-Carlo Planning for Type-Based Reasoning in Large Partially Observable, Multi-Agent Environments. In *AAMAS*. 3.
[22] Jonathon Schwartz, Ruijia Zhou, and Hanna Kurniawati. 2022. Online Planning for Interactive-POMDPs Using Nested Monte Carlo Tree Search. In *IROS*. 8770–8777.
[23] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, and Thore Graepel. 2018. A General Reinforcement Learning Algorithm That Masters Chess, Shogi, and Go through Self-Play. *Science* 362, 6419 (2018), 1140–1144.
[24] David Silver and Joel Veness. 2010. Monte-Carlo Planning in Large POMDPs. *NeurIPS* 23 (2010), 2164–2172.
[25] Finnegan Southey, Michael Bowling, Bryce Larson, Carmelo Piccione, Neil Burch, Darse Billings, and Chris Rayner. 2005. Bayes' Bluff: Opponent Modelling in Poker. In *UAI*. 550–558.
[26] Peter Stone, Gal A. Kaminka, Sarit Kraus, and Jeffrey S. Rosenschein. 2010. Ad Hoc Autonomous Agent Teams: Collaboration without Pre-Coordination. In *AAAI*, Vol. 24. 1504–1509.
[27] Fangju Wang. 2013. An I-POMDP Based Multi-Agent Architecture for Dialogue Tutoring. In *ICAICTE*. 486–489.
[28] Michael P. Wellman. 2006. Methods for Empirical Game-Theoretic Analysis. In *AAAI*. 1552–1556.
[29] Mark P. Woodward and Robert J. Wood. 2012. Learning from Humans as an I-POMDP. *arXiv preprint* (2012). arXiv:1204.0274