Anti-Poaching as a Partially Observable Stochastic Game

Prasanna Maddila[1], Régis Sabbadin[1], Meritxell Vinyals[1]

UR 875, MIAT, INRAE, University of Toulouse, Castanet-Tolosan, France

{siva-sri-prasanna.maddila,regis.sabbadin,meritxell.vinyals}@inrae.fr

Mots-clés : *game theory, partially observable stochastic game, anti-poaching strategies.*

1 Introduction

In today's world, endangered species are threatened by widespread poaching, requiring intelligent land patrol strategies to effectively detect and prevent such activities. Several recent works have developed game-theoretic models for anti-poaching, wherein determining equilibrium strategies, often based on the Nash Equilibrium (NE)[1], leads to effective patrol strategies [3]. Additionally, due to the complexity and imperfect knowledge of the models, Multi-Agent Reinforcement Learning (MARL) methods are usually proposed to learn these strategies. Yet, even with anti-poaching emerging as a popular domain for MARL, the absence of both a general model and a publicly accessible implementation has hindered both the evaluation and development of new solutions. In this context, the objective of this work is two-fold: (i) formalize anti-poaching as a Partially Observable Stochastic Game (POSG) [4] capable of generalizing existing models; and (2) provide a publicly available implementation of this POSG in PettingZoo [2] (one of the most popular APIs to implement MARL environments).

2 Partially Observable Stochastic Games

A POSG $\langle \mathcal{I}, \mathcal{S}, \mathcal{A}, \mathcal{O}, T, O, \{R_i\}_{i\in I}\rangle$ models multi-agent games with simultaneous moves and stochastic game transitions where agents are limited to partial or local observations. A POSG can be defined by specifying the set of agents $\mathcal{I}$, the set of possible states $\mathcal{S}$, the joint action set $\mathcal{A} = \prod_{i\in\mathcal{I}} \mathcal{A}_i$, the joint observation set $\mathcal{O} = \prod_{i\in\mathcal{I}} \mathcal{O}_i$, the model dynamics via the transition probability function T and global observation probabilities O, and the reward functions $\{R_i\}_{i\in I}$ for each agent $i \in \mathcal{I}$. In a POSG, each player's local mixed strategy maps sequences of past local actions and observations to probability distributions over her actions.

3 The Anti-Poaching Game

The Anti-Poaching game is modelled as a POSG with:

- $\mathcal{I} = [N + M]$, with N (cooperative) Rangers and M (independent) Poachers in a $\ell \times \ell$ Grid World. Despite being cooperative, each Ranger is modelled as a distinct player with a shared team reward, without any communication between them.[2]
- Simultaneous joint actions. Rangers move freely in the cell ($\mathcal{A}_r = \{\emptyset, \uparrow, \leftarrow, \downarrow, \rightarrow\}$), while Poachers can also place a trap ($\mathcal{A}_p = \{\emptyset, \uparrow, \leftarrow, \downarrow, \rightarrow, \textit{place-trap}\}$).
- A complex state space resulting from the combination of the state of Rangers, Poachers and placed Traps. A Ranger state is her current location, (m, n), while a Poacher state also includes the number of carried traps and preys, $(m, n, \eta_{trap}, \eta_{prey})$. The state of placed

[1]Other notions used are Stackelberg and Team-MaxMin equilibria, based on the game model.

[2]This models real-situations in which the Rangers team needs to operate in stealth mode on the ground.

Traps is represented by a matrix of tuples $\tau^t \in \mathcal{M}_{n\times n}(\mathbb{N}^2)$ indicating whether they are full or empty.

- Partial observation sets $\mathcal{O}_i$ for each agent of her current cell. Specifically, every Ranger fully observes her own state and other Rangers in her current cell. However, Rangers detect Poachers and Traps in the same cell occurs with a certain probability. Similarly, each Poacher fully observes her own state, yet the detection of other agents within her cell happens only with a certain probability.
- Reward functions R_i. When a Ranger detects a Poacher or a Trap within her current cell, she automatically removes them, earning the team a reward for each successful capture. A Poacher, on the other hand, automatically retrieves her traps placed in her current cell and is rewarded for any prey recovered from them. Since each Poacher competes with the Rangers, a Poacher's rewards imply an equal penalty for the Ranger team, and vice versa.
- Stochastic transition (T) and observability functions (O), due to the detection capabilities of agents (e.g. Rangers detect a Trap/Poacher present in the current cell with a certain probability) and Trap efficiency (Traps capture an animal with a probability that depends on the cell). The state transition is performed in stages to avoid any ill-defined race conditions between agents. The transition and observation probability distributions T, O are then defined by considering all events independently.

4 Implementation

We implemented the model to compute approximate equilibrium strategies using RL techniques[3]. We used the well-known PettingZoo Python POSG simulation API [2]. This implementation can be used as a training simulator for state-of-the-art learning algorithms in RLlib [1], a well-known Python library for multi-agent Deep Reinforcement Learning.

5 Conclusions and perspectives

The calculation of equilibrium strategies in the Anti-poaching Game is difficult due to the cooperative-competitive behaviour between agents and the restriction to partial observations in the POSG model. Therefore, the Anti-poaching Game can be an effective benchmark for algorithms which search for these strategies.

Acknowledgements: This work was funded by the French National Research Agency (ANR), grant ANR-22-CE92-0011-01.

References

[1] Eric Liang, Richard Liaw, Philipp Moritz, Robert Nishihara, Roy Fox, Ken Goldberg, Joseph E. Gonzalez, Michael I. Jordan, and Ion Stoica. Rllib: Abstractions for distributed reinforcement learning, 2018.

[2] J Terry, Benjamin Black, Nathaniel Grammel, Mario Jayakumar, Ananth Hari, Ryan Sullivan, Luis S Santos, Clemens Dieffendahl, Caroline Horsch, Rodrigo Perez-Vicente, et al. Pettingzoo: Gym for multi-agent reinforcement learning. *Advances in Neural Information Processing Systems*, 34:15032–15043, 2021.

[3] Aravind Venugopal, Elizabeth Bondi, Harshavardhan Kamarthi, Keval Dholakia, Balaraman Ravindran, and Milind Tambe. Reinforcement learning for unified allocation and patrolling in signaling games with uncertainty. *arXiv preprint arXiv:2012.10389*, 2020.

[4] Yaodong Yang and Jun Wang. An Overview of Multi-Agent Reinforcement Learning from Game Theoretical Perspective, March 2021. arXiv:2011.00583 [cs].

[3] `https://forgemia.inra.fr/siva-sri-prasanna.maddila/antipoaching`