MULTI-AGENT REINFORCEMENT LEARNING FOR COALITIONAL BARGAINING GAMES

Anonymous authors
Paper under double-blind review

ABSTRACT

There is growing attention to the application of multi-agent reinforcement learning (MARL) to coalition formation problems, in particular, on coalitional bargaining games as a means of negotiating team members and payoffs. However, there is a lack of theoretical principles for using MARL in coalitional bargaining games. We address this gap by providing an examination of the theoretical link between, coalitional bargaining games, and MARL through the formalism of stochastic games. Using this analysis, the paper seeks to shed light on the underlying principles that support the use of MARL in coalitional bargaining and to explore the contributions and limitations of this approach.

1 INTRODUCTION AND RELATED LITERATURE

Coalition formation enables a mechanism in which self-interested agents can work together to achieve greater rewards than what they could attain alone. To negotiate the terms of a coalition, parties engage in *coalitional bargaining games* (CBG). This involves alternating between making offers and counteroffers, with the ultimate objective of reaching an agreement on both the participants in the coalition and the distribution of payoffs Rubinstein (1982). The coalitional bargaining process is lengthy and inefficient, thus prior work introduced multi-agent reinforcement learning (MARL) to speed up the bargaining rounds by creating negotiating agents (Bachrach et al., 2020; Chen et al., 2022; Hughes et al., 2020; Taywade, 2021; Chalkiadakis et al., 2011; Mak et al., 2021). Despite increasing interest, the application of MARL to CBG is hampered by the absence of a theoretical framework linking stochastic games (the framework for MARL) and CBG; resulting in limited *reproducibility* and *generalization* of the results (as aspects crucial for AI research Hutson (2018); Haibe-Kains et al. (2020)).[1]. The aim of this paper is to answer: first, given MARL provides an empirical framework for stochastic games, which theoretical principles make it suitable for CBG? Secondly, what are the advantages/limitations of MARL over traditional game theoretic solutions?

Contributions. First, we provide a unified *conceptual framework* for the application of MARL to CBG. From there, we then present a formal representation of turn-based stochastic games and use it to connect MARL to CBG in a principled way. Second, we provide a discussion on the benefits and limitations of MARL as a *computational method* for CBG problems.

2 CBG, STOCHASTIC GAMES AND MARL - A GOLDEN BRAID

In this section, we answer the first question posed in the Introduction 1. The relationship between CBG and MARL can be established by characterizing the former as a subclass of stochastic *sequential* games, namely, turn-based stochastic games (TBSG), which is a subset of stochastic games (the framework for MARL). See Appendix A.4 for additional background on bargaining games.

Definition 1 (Coalitional bargaining game) *Consists of a tuple $(N, T, \mathcal{A}, \delta)$ where N is the set of agents with $|N| = n$ the number of agents, T is the maximum length of the game (can be infinite), $\mathcal{A}$ is the set of available actions (proposals and responses) for the n agents at time $t < T$ and δ is the discount factor accounting for the value of time. The set of all bargaining games is denoted by $\mathcal{B}$. The solution is given by the function $f : \mathcal{B} \to (2^n, \mathbb{R}^n)$ that maps proposals to the set of coalition structures and stable profits.*

[1]For a connection between stochastic games and MARL refer to Appendix A.2.

With the above definition, we next argue that a TBSG is superset of a CBG where each state $\mathcal{A}$ (proposals and responses) is controlled by a disjoint subset of players, the states are stochastic and governed by transition dynamics.

Definition 2 (Turn-based stochastic game (TBSG)) *Is defined by the tuple* $(N, T, S, \mathcal{A}, R, \tau, \gamma)$ *where* N *is the set of agents,* T *is the maximum bargaining rounds,*S *is the set of proposing states,* $\mathcal{A}$ *denotes the set of joint responses to the proposals,* τ *is the transition dynamics between the states (*S *and* $\mathcal{A}$*),* R *is the vector of reward functions for each agent and* γ *is the discount factor.*

To strengthen the argument for similarity with stochastic games, the solution of a TBSG is a policy set composed of the optimal policies $\pi^* = (\pi_1^*, \dots \pi_n^*)$ such that each agent maximizes its reward conditional on the other agent's policy: $\forall i : \pi_i^* \in argmax_{\pi_i'} \mathbb{E}\left[R_i | \pi_i', \pi_{-i}'\right]$ where $\pi_{-i}' = \pi^* \backslash \pi_i$ and π^* is a Nash Equilibrium. Since we have shown that CBG are a subset of TBSG; MARL can be used as a framework to approximate an optimal policy, π'^* when the transition dynamics, τ and rewards, R are unknown. The following clarifies the equivalencies between a TBSG and stochastic games. In turn-based games, the action of one player (for example, a proposal) is the next players' observations (acting as responders). In other words, the traditional action space from MARL games has been split into two set of actions (proposals S and responses $\mathcal{A}$). In TBSG, a reward is received when all coalition members agree on a proposal, otherwise the reward is zero and the game continues in a sequential manner. Most importantly, in a TBSG, the transition dynamics τ, capture how agents negotiate and make decisions in the game, and this is one of the key differences between the traditional game-theoretic solutions and MARL algorithms. In game theory, the game's transition dynamics are given by a model of the player's *beliefs* as rational agents (Rubinstein, 1982); while under the MARL framework, there is no assumption over preferences or utility functions, and the transition dynamics (and rewards) can be *learned* through *exploration* (Sutton & Barto, 2018).

3 Contributions and Limitations of MARL in CBG

This section addresses the second question posed on the Introduction 1.

MARL provides an algorithmic solution to CBG As opposed to game-theoretic solutions, where the interest is on finding equilibria for a game, MARL methods learning optimal *policies* for each agent given the other agents' policies. Since agents learn through best responding to their training partners, the optimal joint policy converged upon is an equilibrium of the game.

MARL provides an adaptive and tractable solution to CBG. According to Jin et al. (2022), the exact solution to an infinite-horizon TBSG is PPAD complex, while MARL provides a tractable *approximation* to CBG (Bab & Brafman, 2008). Moreover, MARL provides adaptive optimal control of non-linear Markov process with unknown transition dynamics (Sutton et al., 1992). Policies are *adaptive* to small changes in the environment, avoiding the recalculation of coalitional values following a small change in the environment.

MARL provides decentralized learning. MARL methods overcome the exponential complexity of multi-agent games by enabling learning of *decentralized agent policies* (Yang & Wang, 2020).

Challenges in the use of MARL for CBG. As with any multi-agent game, the stationarity of the policies is not guaranteed (Lowe et al., 2017), and most importantly, the Markovian property of the game's states is not guaranteed. In a CBG, the environment states are the agents' proposals, and in bargaining games, past rejected proposals cannot be repeated in the future; therefore, the history of past proposals needs to be accounted for at each step. As a result, the probability of transitioning from one state to another depends not only on the current state and last action taken, but also on the entire history of past rejected proposals. One approach to addressing this issue is to augment the state definition by including the history of proposals so the process becomes a Markov process (Fudenberg & Tirole, 1991), however, this exponentially increases the computational complexity of the problem.

Conclusions. The paper provides the principles for the use of MARL in CBG and highlights its contributions and limitations. It provides a new way to formalize the CBG as a subset of stochastic games, provides analysis on the tractability of MARL techniques and offers insights to improve its effectiveness as well as reproducibility and generalisation.

URM STATEMENT

The authors acknowledge that at least one key author of this work meets the URM criteria of the ICLR 2023 Tiny Papers Track.

REFERENCES

Avraham Bab and Ronen I. Brafman. Multi-agent reinforcement learning in common interest and fixed sum stochastic games: An experimental study. *Journal of Machine Learning Research*, 9: 235–2675, December 2008. ISSN 1532-4435.

Yoram Bachrach, Richard Everett, Edward Hughes, Angeliki Lazaridou, Joel Z. Leibo, Marc Lanctot, Michael Johanson, Wojciech M. Czarnecki, and Thore Graepel. Negotiating team formation using deep reinforcement learning. 2020. doi: 10.48550/ARXIV.2010.10380.

Michael Bowling and Manuela Veloso. An analysis of stochastic game theory for multiagent reinforcement learning. *Carnegie-Mellon University. School of Computer Science.*, (CMU-CS-00-165), 2000.

Georgios Chalkiadakis, Edith Elkind, and Michael Wooldridge. *Computational Aspects of Cooperative Game Theory (Synthesis Lectures on Artificial Inetlligence and Machine Learning)*. 1st edition, 2011. ISBN 1608456528.

Siqi Chen, Yang Yang, and Ran Su. Deep reinforcement learning with emergent communication for coalitional negotiation games. *Mathematical biosciences and engineering : MBE*, 19 5:4592–4609, 2022.

Drew Fudenberg and Jean Tirole. *Game Theory*. MIT Press, Cambridge, MA, 1991.

Benjamin Haibe-Kains, George Alexandru Adam, Ahmed Hosny, Farnoosh Khodakarami, Massive Analysis Quality Control (MAQC) Society Board of Directors Shraddha Thakkar 35 Kusko Rebecca 36 Sansone Susanna-Assunta 37 Tong Weida 35 Wolfinger Russ D. 38 Mason Christopher E. 39 Jones Wendell 40 Dopazo Joaquin 41 Furlanello Cesare 42, Levi Waldron, Bo Wang, Chris McIntosh, Anna Goldenberg, Anshul Kundaje, et al. Transparency and reproducibility in artificial intelligence. *Nature*, 586(7829):E14–E16, 2020.

Edward Hughes, Thomas W. Anthony, Tom Eccles, Joel Z. Leibo, David Balduzzi, and Yoram Bachrach. Learning to resolve alliance dilemmas in many-player zero-sum games. In *Adaptive Agents and Multi-Agent Systems*, 2020.

Matthew Hutson. Artificial intelligence faces reproducibility crisis, 2018.

Yujia Jin, Vidya Muthukumar, and Aaron Sidford. The complexity of infinite-horizon general-sum stochastic games. *arXiv*, 2022. doi: 10.48550/ARXIV.2204.04186.

Michael L. Littman. Markov games as a framework for multi-agent reinforcement learning. In William W. Cohen and Haym Hirsh (eds.), *ICML*, pp. 157–163. Morgan Kaufmann, 1994. ISBN 1-55860-335-2. URL http://dblp.uni-trier.de/db/conf/icml/icml1994.html#Littman94.

Ryan Lowe, YI WU, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/file/68a9750337a418a86fe06c1991a1d64c-Paper.pdf.

Stephen Mak, Liming Xu, Tim Pearce, Michael Ostroumov, and Alexandra Brintrup. Coalitional Bargaining via Reinforcement Learning: An Application to Collaborative Vehicle Routing. In *NeurIPS Cooperative AI Workshop*, October 2021.

John F. Nash. The bargaining problem. *Econometrica*, 18(2):155–162, 1950. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1907266.

John F. Nash. Two-person cooperative games. *Econometrica*, 21(1):128–140, 1953. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1906951.

Akira Okada. A noncooperative coalitional bargaining game with random proposers. *Games and Economic Behavior*, 16(1):97–108, 1996. URL https://EconPapers.repec.org/RePEc:eee:gamebe:v:16:y:1996:i:1:p:97-108.

Ariel Rubinstein. Perfect equilibrium in a bargaining model. *Econometrica*, 50(1):97–109, 1982. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1912531.

Lloyd Shapley. Stochastic games. *Proceedings of the National Academy of Sciences*, 39:1095–1100, 1953.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. A Bradford Book, 2018. ISBN 0262039249.

Richard S. Sutton, Andrew G. Barto, and Robert J. Williams. Reinforcement learning is direct adaptive optimal control. *IEEE Control Systems Magazine*, 12(2):19–22, 1992. doi: 10.1109/37.126844.

Kshitija Taywade. Multi-agent reinforcement learning for decentralized coalition formation games. In *AAAI Conference on Artificial Intelligence*, 2021.

Yaodong Yang and Jun Wang. An overview of multi-agent reinforcement learning from game theoretical perspective, 2020.

A Appendix

This section provides further definitions and contextual information relevant to the content of the paper.

A.1 Stochastic Games

A stochastic game (Shapley, 1953) generalises Markov Decision Processes to involve multiple agents, being the reason why they are the mathematical framework for MARL [2]. The idea behind a stochastic game is that the history at each period can be summarized by a *state*. Stochastic games are defined as a tuple $<N, S, A, T, R, \gamma>$ where: N denotes the set of n agents, S denotes the set of states, $A = A_i \dots A_n$ denotes the set of joint actions, where A_i is player $i's$ set of actions. $T : S \times A \to S$ denotes the transition dynamics, $R : S \times A \times S \times N \to R$ denotes the reward function and γ denotes the discount factor.

The image below depicts the relationship between the different categories of stochastic games and CBG.

A.2 Connection Between MARL and Stochastic Games

Multi-Agent Reinforcement Learning (MARL) tackles the challenge of learning optimal behavior by engaging in trial and error interactions within a dynamic multi-agent environment, where the environmental dynamics and the algorithms utilized by other agents are initially unknown. To model the multi-agent environment interaction, MARL adopts the Game Theory model of stochastic games (Shapley, 1953), which are essentially n-agent Markov Decision Processes. In MARL, we are interested in learning a stationary stochastic policy that maps the game's states to a probability distribution over the agent's actions. The goal is to find such a policy that maximizes the agent's discounted future reward. The connection between MARL and stochastic games is further discussed in Littman (1994); Bowling & Veloso (2000)

[2] MARL is a useful computational framework for stochastic games when the transition dynamics are unknown.

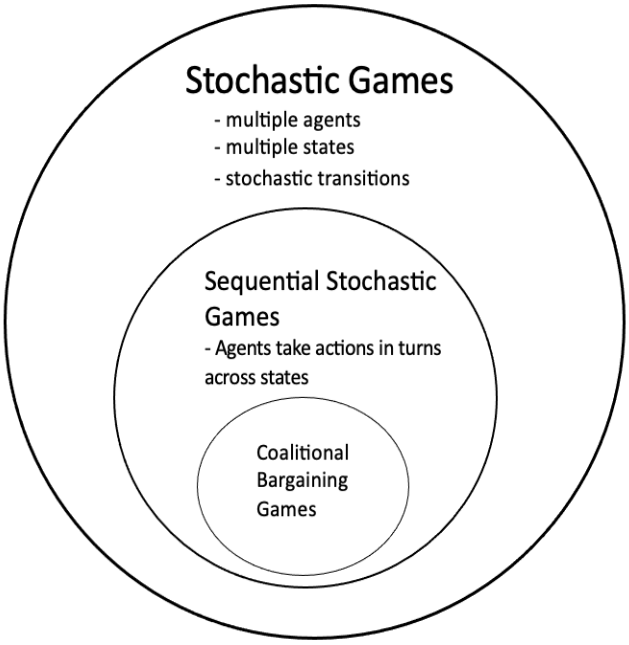

Figure 1: Relationship between stochastic games, sequential stochastic games and coalitional bargaining games.

A.3 The Coalition Structure Generation Problem

The Coalition Structure Generation (CSG) problem is the process of forming coalitions among agents in such a way that the agents within each coalition coordinate their activities, but there is no coordination between coalitions. This entails dividing the set of agents into exhaustive and mutually exclusive coalitions. The resulting partition is referred to as a Coalition Structure (CS). After defining the optimal Coalition Structure, a new game starts to divide the value of the generated solution among agents.

A.4 Non-cooperative Bargaining Theory of Coalition Generation

The non-cooperative game approach to the problem of cooperation was initiated in the seminal works of Nash (1950; 1953), who presented equilibrium results for finite-horizon two-person bargaining game known as the Nash Program. The approach aims to explain cooperation as the result of individual players' payoff maximization in an equilibrium of a non-cooperative bargaining game that models pre-play negotiations. Nash stated the seminal results that cooperation should be strategically stable. The approach re-examines a widely held view in economics, called the efficiency principle, that a Pareto-efficient allocation of resources can be attained through voluntary bargaining by rational agents if there is neither private information nor bargaining costs. After Nash, the theory centered its attention into extending the result to infinite-horizon bargaining. The work of Rubinstein (1982) introduces the *alternating offers model* as an equilibrium bargaining protocol for two-person infinite-horizon bargaining. The expansion of this model to n-person bargaining came later with the work of several authors (see Chalkiadakis et al. (2011) for a literature review). One example is the protocol proposed by Okada (1996), which presents a sequential bargaining game in which players propose coalitions and feasible payoff allocations until an agreement is reached. Under this protocol, agreement can be reached in one bargaining round if the proposer is chosen randomly.

Example of an *n-person alternating offers bargaining protocol*. As an example of a sequential bargaining protocol, we describe the one proposed by Okada (1996).

The process of bargaining involves the agents taking turns to make proposals to form a coalition and to allocate the payoffs among the coalition members, while the rest of the agents provide responses

on whether to accept or reject the proposal. The goal is to find a coalition c^* and payoff allocation $u(c^*)$ such that the agents reach an agreement, where $u(c^*)$ is the vector of optimal payoffs for each agent in the coalition.

Let N_t be the set of "active" players who do not belong to any coalitions on round t, let S be the set of possible coalitions that the N_t players can form, and let $c \in S$ be a possible coalition. At the beginning of each round t, a *proposer* $i \in N_t$ is selected according to a certain probability distribution $\theta(N_t)$. The selected player makes a *proposal* which is a tuple formed by a coalition S (where $i \in c$) and a profit-sharing scheme $u(c)$ (i.e., a payoff vector). In the same time step, all other members in c, (i.e., the *responders*) either accept or reject the proposal sequentially following a randomly chosen order. The responders have full information of the responses from previous responders in c. If all responders accept the proposal $(c, u(c))$ then it is binding, and another round of bargaining starts with $N_{t+1} = N_t \backslash S$. Otherwise, if any responder rejects the proposal, with probability $1 - \epsilon$, $(\epsilon > 0)$ another proposer is selected randomly and the game continues. The negotiation process ends when every player in N joins some coalition, and thus, we say a stable solution to the game has been reached.

The image below depicts the stages of a coalitional bargaining game.

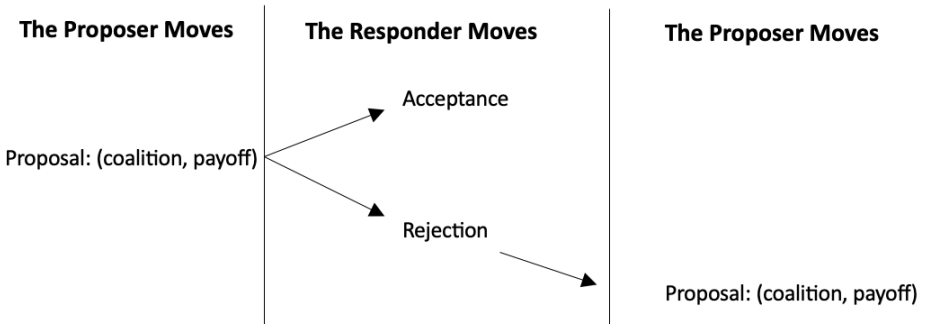

Figure 2: Bargaining game as an extensive-form game.