Cooperation in Multi-Agent Reinforcement Learning

Karush Suri, Dian Gadjov, Lacra Pavel
Department of Electrical & Computer Engineering, University of Toronto, Canada.
karush.suri@mail.utoronto.ca

Abstract

Advancements in Multi-Agent Reinforcement Learning (MARL) are motivated by cooperation in agents arising from Game Theory (GT). Agents must collaborate in practical scenarios in order to achieve complex objectives and attain strategies which depict optimal behavior. The need for cooperation is further highlighted in the case of partially-observed settings wherein agents have restricted access to environment observations. We revisit cooperation in MARL from the viewpoint of GT and stochastic dynamics of environments. The contributions of our work are twofold. (1) We analyze and demonstrate the effectiveness of cooperative MARL in the case of complex and partially-observed tasks consisting of high-dimensional action spaces and stochastic dynamics. (2) We leverage the empirical demonstrations to construct a novel optimization objective which addresses the detrimental effects of spurious states across agents. Our large-scale experiments carried out on the StarCraft II benchmark depict the effectiveness of cooperative MARL and our novel objective for obtaining optimal strategies under stochastic dynamics.

1 Introduction

Reinforcement Learning (RL) has seen tremendous growth in applications such as arcade games [1], board games [2, 3], robot control tasks [4, 5] and lately, real-time games [6]. The rise of RL has led to an increasing interest in the study of multi-agent systems [7, 8], commonly known as Multi-Agent Reinforcement Learning (MARL). In the case of partially observable settings, MARL enables the learning of strategies from a GT perspective by utilizing cooperation across agents[9]. Agents collaborate with each other to optimize the cumulative payoffs by means of a single utility function which leads to optimal behavior [10, 11] in the long-horizon.

Surprise minimization [12] is a recent phenomenon observed in the case of single-agent RL methods which deals with environments consisting of spurious states. In the case of model-based RL [13], surprise minimization is used as an effective planning tool in the agent's model [12] whereas in the case of model-free RL, surprise minimization is witnessed as an intrinsic motivation [14, 15] or generalization problem [16]. On the other hand, MARL does not account for surprise across agents as a result of which agents remain unaware of drastic changes in the environment [17]. Thus, surprise minimization in the case of multi-agent settings requires attention from a critical standpoint.

2 Related Work

2.1 Learning in Games

2.2 Multi-Agent Learning

3 Preliminaries

3.1 Stochastic Markov Games

3.2 Q-Learning

3.3 Multi-Agent Learning

4 Cooperation in Multi-Agent Learning

4.1 The Partial Observability Setting

4.2 Learning Model-Free Behaviors

5 Tackling Spurious Dynamics

6 Experiments

6.1 The StarCraft II Benchmark

6.2 Performance

6.3 Spurious Dynamics

7 Conclusion

References

[1] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin A. Riedmiller. Playing atari with deep reinforcement learning. *CoRR*, abs/1312.5602, 2013.

[2] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, January 2016.

[3] Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, Timothy Lillicrap, and David Silver. Mastering atari, go, chess and shogi by planning with a learned model, 2019.

[4] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Manfred Otto Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *CoRR*, abs/1509.02971, 2015.

[5] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *CoRR*, abs/1707.06347, 2017.

[6] Oriol Vinyals, Timo Ewalds, Sergey Bartunov, Petko Georgiev, Alexander Sasha Vezhnevets, Michelle Yeo, Alireza Makhzani, Heinrich Küttler, John Agapiou, Julian Schrittwieser, John Quan, Stephen Gaffney, Stig Petersen, Karen Simonyan, Tom Schaul, Hado van Hasselt, David Silver, Timothy Lillicrap, Kevin Calderone, Paul Keet, Anthony Brunasso, David Lawrence, Anders Ekermo, Jacob Repp, and Rodney Tsing. Starcraft ii: A new challenge for reinforcement learning, 2017.

[7] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments, 2017.

[8] Oriol Vinyals, Igor Babuschkin, Wojciech Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David Choi, Richard Powell, Timo Ewalds, Petko Georgiev, Junhyuk Oh, Dan Horgan, Manuel Kroiss, Ivo Danihelka, Aja Huang, Laurent Sifre, Trevor Cai, John Agapiou, Max Jaderberg, and David Silver. Grandmaster level in starcraft ii using multi-agent reinforcement learning. *Nature*, 575, 11 2019.

[9] Liviu Panait and Sean Luke. Cooperative multi-agent learning: The state of the art. *Autonomous agents and multi-agent systems*, 11(3):387–434, 2005.

[10] Ann Nowé, Peter Vrancx, and Yann-Michaël De Hauwere. Game theory and multi-agent reinforcement learning. In *Reinforcement Learning*, pages 441–470. Springer, 2012.

[11] Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *arXiv preprint arXiv:1911.10635*, 2019.

[12] Glen Berseth, Daniel Geng, Coline Devin, Dinesh Jayaraman, Chelsea Finn, and Sergey Levine. Smirl: Surprise minimizing rl in entropic environments. 2019.

[13] Lukasz Kaiser, Mohammad Babaeizadeh, Piotr Milos, Blazej Osinski, Roy H Campbell, Konrad Czechowski, Dumitru Erhan, Chelsea Finn, Piotr Kozakowski, Sergey Levine, Afroz Mohiuddin, Ryan Sepassi, George Tucker, and Henryk Michalewski. Model-based reinforcement learning for atari, 2019.

[14] Joshua Achiam and Shankar Sastry. Surprise-based intrinsic motivation for deep reinforcement learning, 2017.

[15] Luis Macedo, Rainer Reisezein, and Amilcar Cardoso. Modeling forms of surprise in artificial agents: empirical and theoretical study of surprise functions. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 26, 2004.

[16] Jerry Zikun Chen. Reinforcement learning generalization with surprise minimization, 2020.

[17] Luis Macedo and Amilcar Cardoso. The role of surprise, curiosity and hunger on exploration of unknown environments populated with entities. In *2005 portuguese conference on artificial intelligence*, 2005.