Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)

Integrate Learning with Game Theory for Societal Challenges

Fei Fang
School of Computer Science, Carnegie Mellon University, PA, USA
feif@cs.cmu.edu

Abstract

Real-world problems often involve more than one decision makers, each with their own goals or preferences. While game theory is an established paradigm for reasoning strategic interactions between multiple decision-makers, its applicability in practice is often limited by the intractability of computing equilibria in large games, and the fact that the game parameters are sometimes unknown and the players are often not perfectly rational. On the other hand, machine learning and reinforcement learning have led to huge successes in various domains and can be leveraged to overcome the limitations of the game-theoretic analysis. In this paper, we introduce our work on integrating learning with computational game theory for addressing societal challenges such as security and sustainability.

1 Introduction

The new era of artificial intelligence (AI) is featured by the success of machine learning, and in particular, deep learning, in finding patterns in a massive amount of data. However, AI is not just about *understanding* the world, but also about *making smart decisions* based on the understanding. Recent advances on deep reinforcement learning have led to a prominent performance in playing Atari games [Mnih *et al.*, 2015] and Go games [Silver *et al.*, 2016], showing the potential of AI in decision making. Despite the success, real-world problems often involve more than one agent, each with their own goals or preferences, and with partial observability of the environment or the actions taken by other agents, which makes the decision-making particularly challenging. While game theory is an established paradigm for reasoning strategic interaction between multiple decision-makers, its applicability in practice is often limited due to a few reasons: (i) finding equilibria in large scale games is computationally intractable in general; (ii) game parameters are sometimes unknown; (iii) the players are often not perfectly rational.

Aiming to develop computational tools to assist decision makers in the real world, we focus on integrating learning with game theory for strategic decision making. Our work is mostly motivated by global challenges with significant social impact, including (i) *security*, e.g., infrastructure security and cyber-security, (ii) *sustainability*, e.g., wildlife conservation, fishery, and forest protection. Strategic decision making lies in all these problems, for example, the problem of protecting wildlife from poachers or mitigating cyber attacks can be abstracted as a game between defender(s) and attacker(s).

To develop solutions to such problems, we have explored three research directions:

- **Data-Based Game-Theoretic Reasoning**: It is well accepted that human decision makers are not always perfectly rational, which often makes the game-theoretic reasoning challenging. Fortunately, data is available in domains such as cybersecurity and wildlife protection. When modeling the strategic decision making in these domains as games, a significant research problem is how to capitalize on the availability of data to learn the behavior model of human players and how to develop efficient algorithms to compute optimal or equilibrium strategies based on the learned models. We propose algorithms for this problem, leveraging machine learning and techniques in mathematical programming.
- **Learning-Powered Strategy Computation in Large Scale Games**: The computation of (near) optimal strategies or equilibrium strategy profiles is extremely challenging in complex settings, e.g., when there are more than two players, infinite action set, interactions over a long time horizon, or uncertainty in the environment. Instead of solely relying on techniques in combinatorial optimization and mathematical programming, we leverage the power of learning, and in particular, deep reinforcement learning to solve such complex games.
- **End-to-End Learning of Game Parameters**: When game parameters such as payoff values for the players are not known in advance, learning the parameters becomes an important and challenging task. The task becomes even more challenging when such parameters and models themselves are adaptive to the context features. We propose a differentiable, end-to-end learning framework for the task. In this framework, a "differentiable game solver" is embedded into the commonly used deep neural network architecture. We leverage techniques in optimization and propose efficient algorithms for learning.

Our goal is to not only provide substantial theoretical con-

tributions but also build or empower applications that can fundamentally improve the current practices in the domains we work in. Therefore, we evaluate our proposed approaches in security and sustainability domains, with both in-lab experiments and field tests.

2 Data-Based Game-Theoretic Reasoning

Players are often assumed to be perfectly rational in game theory literature (i.e., they choose the action with the highest expected payoff). This assumption is reasonable in some domains. For example, in our earlier work on protecting moving targets from potential attacks [Fang *et al.*, 2013a; Fang *et al.*, 2013b; Xu *et al.*, 2014], the game model captures the real world scenario where a defender can deploy a limited number of patrol boats to escort and protect multiple ferries on the open water from the potential threat of an attacker hitting the ferry. The defender in this problem is aided by an algorithm and the attacker may spend a sufficient amount of time for surveillance and careful planning before launching an attack, making the assumption of perfect rationality grounded.

However, it is well accepted that humans are often boundedly rational or are limited by computational capabilities [Gigerenzer and Selten, 2002]. For example, in anti-poaching, poachers do not spend a lot of time on surveillance and planning before placing snares to capture the wildlife [Fang *et al.*, 2015]. As such, they may not always choose the option with the highest expected utility to them. Therefore, to design effective strategies for the decision maker, it is crucial to understand how human players behave in strategic interaction.

To achieve this goal, some of our work [Fang *et al.*, 2015; Kar *et al.*, 2015; Kar *et al.*, 2016; Fang *et al.*, 2016] extend the quantal response (QR) model [McKelvey and Palfrey, 1995], a classic model in behavioral game theory, and learn the parameters in the model from experimental data collected through human subject experiments. In domains such as wildlife conservation, urban crime prevention, and cyber-security, real-world crime data is often available and can be used to learn the human agents' behavior. Based on a poaching activity dataset for Queen Elizabeth National Park in Uganda, we developed a series of machine learning (ML) algorithms that build upon decision tree ensembles to predict the poacher's behavior [Kar *et al.*, 2017; Gholami *et al.*, 2017; Gurumurthy *et al.*, 2018]. The proposed algorithms are evaluated not only in the lab but also in the field, with a significant increase in the number of snares found during the field test in Uganda.

Due to the lack of patrolling resources in anti-poaching, the main challenges in dealing with real-world data sets in this domain include handling significant class imbalance, sparsity, and noise in negative labels. In one of the recent work [Gurumurthy *et al.*, 2018], in addition to the real-world data of past patrol and poaching, we use survey data collected from domain experts such as rangers and conservation site managers. We discretize the protected area into a 1km-by-1km grid and use K-means clustering to group the grid cells into clusters based on geospatial information such as distance to the nearest village and average slope. We then ask the domain experts to provide scores for each group to reflect their subjective estimation of the poaching threat level in that group. Since only a small portion of the protected area has been patrolled in the past, these scores can serve as an additional source of information for the unpatrolled area. We treat the scores as soft labels for poaching activities, based on which we augment the real world data set. More specifically, we randomly sample data points from areas that are not patrolled before but have high scores provided by the experts and add them as positive data points. The use of two sources of information has led to better performance, measured by standard ML metrics such as precision, recall, F1, and AUC score. Further, field tests in China have shown that it can identify poaching hotspots that were not known to the rangers and many snares were found during the field tests, including 22 snares found in a day in a protected area in Northeastern China.

We not only worked on how to model and learn human behavior patterns in the strategic interaction but also how to exploit them in game theoretic reasoning [Fang *et al.*, 2015; Fang *et al.*, 2016; Xu *et al.*, 2017]. We build game models that extend the Stackelberg security games [Tambe, 2011], a leader-follower game model for the defender-attacker interaction in various security domains. The proposed models take into account the practical aspects in anti-poaching and other similar domains by adding scheduling constraints of the defender as well as the bounded rational behavior model of the attackers, represented by QR-based models or other models such as decision tree ensembles. We propose algorithms to compute the optimal patrol strategy for the defender, which consists of a probability distribution over a set of patrol routes that are compatible with the terrain in the area. The algorithms leverage the cutting-plane approach and column generation to improve scalability. Data-based game-theoretic reasoning is the key in the PAWS (Protection Assistant for Wildlife Security) application for anti-poaching [Fang *et al.*, 2016], which is tested in the field in Malaysia, with many human activity signs and animal signs found during the test patrols.

3 Learning-Powered Strategy Computation in Large Scale Games

As game models account for more and more practical aspects, the computation of optimal strategy or equilibrium strategies become harder and harder. How to leverage the power of reinforcement learning to find good strategies in complex games? We made several attempts to answer this question, with a focus on defender-attacker game models as they are the ones most relevant to the security and sustainability domains of interest.

In [Wang *et al.*, 2019], we focus on green security games with real-time information. In practice, instead of sticking to a pre-planned patrol route, well-trained rangers would make use of the real-time information such as footprints, blood stains, tree marks left by the poachers to adjust her patrol route [Maasailand Preservation Trust, 2011]. Similarly, a poacher may respond to the ranger's action in real time, and the rangers should be aware of such risk. A game model

that takes into account these elements is more practical, but the resulting games are large extensive-form games with imperfect information and are hard to solve, even when we only consider the zero-sum games. To address this challenge, we designed DeDOL (Deep-Q Network based Double Oracle enhanced with Local modes), a deep reinforcement learning-based algorithm, to compute a patrolling strategy that adapts to the real-time information. DeDOL combines deep Q-learning (DQN) with the double oracle (DO) framework [McMahan *et al.*, 2003; Bosansky *et al.*, 2013], which uses incremental strategy generation to find an equilibrium strategy with small support in zero-sum games. DeDOL also can be viewed as an instantiated algorithm of the meta-method named policy-space response oracle (PSRO) [Lanctot *et al.*, 2017] for multi-agent reinforcement learning. A naive application of DQN and DO suffers from limited scalability. To mitigate this limitation, DeDOL uses domain-specific heuristic strategies, including a parameterized random walk strategy and a random sweeping strategy as initial strategies to warm up the incremental strategy generation process. In addition, DeDOL proposes to use several local modes, each can be viewed as a restricted version of the original game, to reduce the complexity of the game environment for efficient and parallelized training. In experiments, for small game instances, we show that DeDOL achieves comparable performance as existing approaches for EFGs such as counterfactual regret (CFR) minimization. In large games where CFR becomes intractable, DeDOL can find much better defender strategies than other baseline strategies.

In addition to DeDOL, we also developed Opt-GradFP [Kamra *et al.*, 2018] and M3DDPG [Li *et al.*, 2019], both leverage advances in deep reinforcement learning and existing approaches for noncooperative games, and we evaluate the algorithms in problems of patrolling in continuous area and adversary team interaction such as predator-prey.

4 End-to-End Learning of Game Parameters

Despite the advances in solving games and their successes in security and sustainability domains, as well as the resulting breakthroughs in poker [Moravčik *et al.*, 2017; Bowling *et al.*, 2015], a common criticism for applying game theory for real problems is that the parameters of the game itself are sometimes unknown. For example, in developing algorithms for Stackelberg security games, the payoffs for the defender and the attacker are often assumed to be known. However, in practice, one needs to rely on domain experts such as security officers to provide such input. It is tedious work and it is sometimes hard for the officers to provide convincing values, especially the payoff values for the attacker. So the question we are asking is how to learn the game parameters from observed actions, which is a problem in the inverse game theory setting.

We propose a differentiable, end-to-end learning framework for addressing this task [Ling *et al.*, 2018], learning the parameters of uncertain games purely by observing the actions of the agents. The key idea is to consider the quantal response equilibrium (QRE) [McKelvey and Palfrey, 1995], a smoothed version of the Nash Equilibrium. QRE captures the bounded rationality of human players and QRE of a two-player zero-sum normal-form game can be represented as a *differentiable* function of the game payoff matrix. Thus, we can use a primal-dual Newton method to find QRE and we develop a backpropagation method to analytically compute gradients of all relevant game parameters through the solution itself. This allows us to infer the payoff matrix or other parameters of a game merely from action samples from QRE. This method can be extended to extensive-form games. This method can also be applied when the game parameters are differentiable functions of contextual features, and the observations include both the contextual features and the players' actions. Essentially, the method allows for game-solving to be integrated as a module in deep learning systems. When evaluating the effectiveness of the learning method in several settings including poker and security game tasks, we found that our approach is able to learn the payoff matrices or (agent belief over) chance node probabilities.

We further extend this framework in two ways [Ling *et al.*, 2019]. We first generalize QREs to our proposed equilibrium concept – Nested Logit Quantal Response equilibrium (NLQRE), which draws upon ideas from behavioral science and allows for varying levels of player rationality at each stage of a game. Second, we significantly improve the scalability of the method, by reformulating the backward pass as a min-max convex optimization problem and uses state-of-the-art first-order primal-dual methods for both the forward pass and backward pass. This leads to orders of magnitude of speedups in one-card poker. The main reason is that the first-order solver does not require the explicit formation of Hessians and only requires access to a fast best-response oracle.

5 Discussion

We proposed three ways of integrating learning with game theory to address societal challenges such as security and sustainability. However, we believe that there are other ways for such integration and there are a lot more research questions to be answered. For example, as we use data to learn the human players' preferences in games, an important question is what if the opponents are aware of the learning and strategically take actions to counter the learning procedure. How to model such a competitive learning process as a complex game and compute equilibrium strategies is an underexplored problem despite some recent efforts [Nguyen *et al.*, 2019].

Acknowledgments

I thank all of my collaborators for making the work possible. The work was partially supported by NSF IIS-1850477.

References

[Bosansky *et al.*, 2013] Branislav Bosansky, Christopher Kiekintveld, Viliam Lisy, Jiri Cermak, and Michal Pechoucek. Double-oracle algorithm for computing an exact nash equilibrium in zero-sum extensive-form games. *AAMAS'13*, 2013.

[Bowling *et al.*, 2015] Michael Bowling, Neil Burch, Michael Johanson, and Oskari Tammelin. Heads-up limit hold'em poker is solved. *Science*, 347(6218):145–149, 2015.

[Fang *et al.*, 2013a] Fei Fang, Albert Xin Jiang, and Milind Tambe. Optimal patrol strategy for protecting moving targets with multiple mobile resources. In *Proceedings of the 2013 International Conference on Autonomous Agents and Multi-agent Systems*, AAMAS '13, pages 957–964, 2013.

[Fang *et al.*, 2013b] Fei Fang, Albert Xin Jiang, and Milind Tambe. Protecting moving targets with multiple mobile resources. *Journal of Artificial Intelligence Research, 48:583-634*, 2013.

[Fang *et al.*, 2015] Fei Fang, Peter Stone, and Milind Tambe. When security games go green: Designing defender strategies to prevent poaching and illegal fishing. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2015.

[Fang *et al.*, 2016] Fei Fang, Thanh H. Nguyen, Rob Pickles, Wai Y. Lam, Gopalasamy R. Clements, Bo An, Amandeep Singh, Milind Tambe, and Andrew Lemieux. Deploying paws: Field optimization of the protection assistant for wildlife security. In *Proceedings of the Twenty-Eighth Innovative Applications of Artificial Intelligence Conference (IAAI)*, 2016.

[Gholami *et al.*, 2017] Shahrzad Gholami, Benjamin Ford, Fei Fang, Andrew Plumptre, Milind Tambe, Margaret Driciru, Fred Wanyama, Aggrey Rwetsiba, Mustapha Nsubaga, and Joshua Mabonga. Taking it for a test drive: a hybrid spatio-temporal model for wildlife poaching prediction evaluated through a controlled field test. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 292–304. Springer, 2017.

[Gigerenzer and Selten, 2002] Gerd Gigerenzer and Reinhard Selten. *Bounded rationality: The adaptive toolbox*. MIT press, 2002.

[Gurumurthy *et al.*, 2018] Swaminathan Gurumurthy, Lantao Yu, Chenyan Zhang, Yongchao Jin, Weiping Li, Xiaodong Zhang, and Fei Fang. Exploiting Data and Human Knowledge for Predicting Wildlife Poaching. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, page 29. ACM, 2018.

[Kamra *et al.*, 2018] Nitin Kamra, Umang Gupta, Fei Fang, Yan Liu, and Milind Tambe. Policy Learning for Continuous Space Security Games using Neural Networks. 2018.

[Kar *et al.*, 2015] Debarun Kar, Fei Fang, Francesco Delle Fave, Nicole Sintov, and Milind Tambe. "a game of thrones": When human behavior models compete in repeated Stackelberg security games. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015)*, 2015.

[Kar *et al.*, 2016] Debarun Kar, Fei Fang, Francesco M Delle Fave, Nicole Sintov, Milind Tambe, and Arnaud Lyet. Comparing Human Behavior Models in Stackelberg Security Games: An Extended Study. *Artificial Intelligence Journal (AIJ), Elsevier, DOI: http://dx.doi.org/10.1016/j.artint.2016.08.002*, 2016.

[Kar *et al.*, 2017] Debarun Kar, Benjamin Ford, Shahrzad Gholami, Fei Fang, Andrew Plumptre, Milind Tambe, Margaret Driciru, Fred Wanyama, Aggrey Rwetsiba, Mustapha Nsubaga, Others, Southern California, and Los Angeles. Cloudy with a chance of poaching: adversary behavior modeling and forecasting with real-world poaching data. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pages 159–167. International Foundation for Autonomous Agents and Multiagent Systems, 2017.

[Lanctot *et al.*, 2017] Marc Lanctot, Vinicius Zambaldi, Audrunas Gruslys, Angeliki Lazaridou, Julien Perolat, David Silver, Thore Graepel, Others, Karl Tuyls, Julien Perolat, David Silver, and Thore Graepel. A unified game-theoretic approach to multiagent reinforcement learning. In *Advances in Neural Information Processing Systems*, number Nips, pages 4190–4203, 2017.

[Li *et al.*, 2019] Shihui Li, Yi Wu, Xinyue Cui, Honghua Dong, Fei Fang, and Stuart Russell. Robust Multi-Agent Reinforcement Learning via Minimax Deep Deterministic Policy Gradient. In *Proceedings of The Thirty-Third AAAI Conference on Artificial Intelligence*, 2019.

[Ling *et al.*, 2018] Chun Kai Ling, Fei Fang, and J Zico Kolter. What game are we playing? End-to-end learning in normal and extensive form games. *arXiv preprint arXiv:1805.02777*, 2018.

[Ling *et al.*, 2019] Chun Kai Ling, Fei Fang, and J Zico Kolter. Large Scale Learning of Agent Rationality in Two-Player Zero-Sum Games. In *Proceedings of The Thirty-Third AAAI Conference on Artificial Intelligence*, 2019.

[Maasailand Preservation Trust, 2011] Maasailand Preservation Trust. Poacher arrested with game meat, 2011. https://mpttalk.wordpress.com/ 2011/05/11/poacher-arrested-with-game-meat/.

[McKelvey and Palfrey, 1995] R. D. McKelvey and T. R. Palfrey. Quantal response equilibria for normal form games. *Games and Economic Behavior*, 2:6–38, 1995.

[McMahan *et al.*, 2003] H Brendan McMahan, Geoffrey J Gordon, and Avrim Blum. Planning in the presence of cost functions controlled by an adversary. In *ICML'03*, 2003.

[Mnih *et al.*, 2015] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, and Others. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529, 2015.

[Moravčik *et al.*, 2017] Matej Moravčik, Martin Schmid, Neil Burch, Viliam Lisy, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael Bowling. Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*, 356(6337):508–513, 2017.

[Nguyen *et al.*, 2019] Thanh Hong Nguyen, Yongzhao Wang, Arunesh Sinha, and Michael P Wellman. Deception in finitely repeated security games. In *33th AAAI Conference on Artificial Intelligence*, 2019.

[Silver *et al.*, 2016] 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, 2016.

[Tambe, 2011] Milind Tambe. *Security and Game Theory: Algorithms, Deployed Systems, Lessons Learned*. Cambridge University Press, 2011.

[Wang *et al.*, 2019] Yufei Wang, Zheyuan Ryan Shi, Lantao Yu, Yi Wu, Rohit Singh, Lucas Joppa, and Fei Fang. Deep Reinforcement Learning for Green Security Games with Real-Time Information. In *Proceedings of The Thirty-Third AAAI Conference on Artificial Intelligence*, 2019.

[Xu *et al.*, 2014] Haifeng Xu, Fei Fang, Albert Xin Jiang, Vincent Conitzer, Shaddin Dughmi, and Milind Tambe. Solving zero-sum security games in discretized spatio-temporal domains. In *Proceedings of the 28th Conference on Artificial Intelligence (AAAI 2014), Québec, Canada*, 2014.

[Xu *et al.*, 2017] Haifeng Xu, Benjamin Ford, Fei Fang, Bistra Dilkina, Andrew Plumptre, Milind Tambe, Margaret Driciru, Fred Wanyama, Aggrey Rwetsiba, Mustapha Nsubaga, and Others. Optimal patrol planning for green security games with black-box attackers. In *International Conference on Decision and Game Theory for Security*, pages 458–477. Springer, 2017.