Artificial Intelligence Algorithms for Strategic Reasoning over Complex Multiagent Systems

Doctoral Consortium

Zun Li
University of Michigan, Ann Arbor
lizun@umich.edu

ABSTRACT

My Ph.D. research focuses on developing practical algorithms in computer games by assembling a variety of artificial intelligence methods (game-tree search, machine learning, graphical models, etc.). In this extended abstract, I will briefly review three of my previous works that studied normal-form games, Bayesian games, and extensive-form games through modern AI lenses. Then I will cast three possible future directions that I am dedicating to.

KEYWORDS

Computational Game Theory, Machine Learning, Deep Reinforcement Learning, Game-Tree Search, Graphical Models

ACM Reference Format:

Zun Li. 2023. Artificial Intelligence Algorithms for Strategic Reasoning over Complex Multiagent 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

Computer games had historically been one of the fundamental driving forces for the field of artificial intelligence research. To develop practical game-playing AI in large and complex games such as Go [25], Poker [19], and Stratego [23], researchers should not only make use of the analytical results from classic economics, but also flexibly incorporate contemporary AI methods which facilitate approximate but scalable computation. My Ph.D. research focuses on the latter part.

2 PAST WORKS

My past Ph.D. works followed precisely the chronological order that most game theory textbooks are organized: the most basic normal-form games are first studied, then are games with incomplete information, and then are dynamical games with imperfect information. The only difference here, though, is that my approaches were more from a computational perspective using practical AI methods, instead of deriving the exact mathematical solutions.

Structure Learning in Normal-Form Games. A normal-form game representation generally connotes a utility function $u_n : S_1 \times \ldots \times S_N \to \mathbb{R}$ for agent n in an N-player game, where S_i is the strategy space of agent i. However the normal-form representation does not scale: for N-player, M-strategy games, it requires a tensor of $O(NM^N)$ to store all the utility values. This makes the efficient computation of solution concepts such as Nash equilibrium prohibitive. Furthermore, such payoff values might not be given exactly, but can only be estimated through agent-based simulation. My previous work [16] adopted a model-based learning approach to tackle this issue. By using supervised or unsupervised learning techniques, we can learn a succinct representation (such as clusters or a graph) of the true game using payoff data under some structural hypothesis. The computation within the learned game can be much more efficient, and the solutions were experimentally shown well in the true games.

Deep Evolutionary Search in Bayesian Games. Bayesian games [10] augment the normal-form representation using the concept of *type* that represents the belief over opponents' hidden information (e.g., parameters in their utility functions such as private cards in Poker). In this case, a pure strategy now becomes a mapping from a player's types to actions. Bayesian games were typically studied in auction theory [12] where type spaces or action spaces were low-dimensional and regular so that analytical solutions can be derived. In my previous work [17], I formulated the equilibrium computation problem in Bayesian games in a similar way as in Deep RL, where each pure strategy is represented as a neural net, and the utilities come in the form of black-box simulation data. Using natural evolution strategies [29], I proposed two algorithms to compute pure equilibria and mixed equilibria, respectively. The first one exploits the symmetry structure of the game and transforms the problem into solving a two-player zero-sum Stackelberg game. We found that deep neural nets can recover classical analytical solutions in simple games like first- and second-price auctions. The second method is inspired by double-oracle [18]. We demonstrated the capabilities of these algorithms on high-dimensonal games.

Combining Game-Tree Search and Multiagent RL in Extensive-Form Games. My latest work [15] will be presented at AAMAS'23 as an extended abstract. In this work, we extended AlphaZero-styled search method to general-sum imperfect information games by replacing MCTS with information-set MCTS [6], and learning a deep belief network to represent belief states at the root of the search tree. Furthermore, we combine this new search method with policy space response oracle [14] and construct a decision-time AI bot that can conduct test-time search and online Bayesian opponent modeling. We evaluate this bot against humans in a class of negotiation games and found our bot gave comparable social welfare with humans.

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. © 2023 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

3 FUTURE DIRECTIONS

Through the years I had been thinking about how much capacity of the classic results in game theory can support designing modern game-playing AI. It turns out that most of the current successful AI bots heavily rely on the minimax theorem in two-player zero-sum games, which, may lose its power in more general game settings. I identify the following directions to be central and promising:

New multiagent evaluation metrics and solution concepts. Nash equilibrium may no longer make sense in a general-sum environment due to the equilibrium selection issue. So, how can one claim that an agent is "strong" in such environments? Do we still use exploitability or game-theoretic regret as a score metric? Or, is measuring performances against humans the only way to go? A more interesting question is whether there is a way to define a canonical distribution over opponents or agent population where one can appeal to game-theoretic scores [3, 21] to rank the agents. For example, the reason why the famous "Tit-for-Tat" strategy stood out in the repeated prisoners' dilemma tournament [2] was probably that there was a certain degree of collaborative elements among other participants' strategies, which, might be a reasonable test-time distribution to assume. Another possibility is to resort to other game-theoretic solutions. Examples include correlated equilibria with certain equilibrium selection criteria such as maximal social welfare.

New representation/languages for practical agent architecture. In two-player zero-sum games, people usually approximate Nash at training time and improve its quality (for both players) online at test time [5] to devise the optimal play. However in general-sum games, it may no longer be safe to assume your opponents are playing according to some Nash. In this case, a belief modeling over the environment (both the imperfect information of the game and the other players' strategies) might be eventually inevitable. In fact, an epistemic model which considers a belief hierarchy (I think you think I think you will play....) might be an effective tool for general-sum strategic reasoning. People had developed certain graphical representations such as network of influence diagram [7] to facilitate recursive opponent modeling, and the field of epistemic game theory [1, 11, 22] had been historically providing an alternative useful way of thinking. A question is whether these representations scale into large domains. One possible approach is to adopt the common-prior assumption (CPA) [4, 8, 20] in economics that might simplify the complexity of the cognitive architecture. CPA might not be an entirely restrictive assumption if the agents at test-time are also trained under the same entity (say, the self-driving cars from the same automobile system).

New decision-time planning / search methods. Search, or planning techniques improve solution quality at test time in an online fashion, and had been demonstrated thorough ablation studies the crux of most of the human-level game-playing bots [1] [19, 25]. However as mentioned earlier, in general-sum environments there is perhaps no justification for conducting equilibrium search, but for planning within a belief hierarchy. Since a search call typically conducts multiple traverses over the game tree at one decision step, it might be computationally expensive to do planning within an epistemic model if the other players' types are also assumed to be search-based. One solution is to use distilled policies for others within the search procedure [26] such as the policy network part of the search algorithm. Another possible approach is to permit imperfect recall for the agents and search within an abstracted version of the true game [13, 28]. While imperfect recall will cause mathematical issues for belief revision [24], it is worth pointing out that most of the deep RL agents today are, effectively, imperfect recall agents. An interesting algorithmic question would be how imperfect recall changes the execution of a planning algorithm (for example if an agent deviates from an information set, that probably will also cause its deviation in other places of the search tree). Another possible research direction is to compute refined equilibrium solutions in games of imperfect recall [9] and study their practical benefits such as generalization capabilities.

[1] Exceptions include AlphaStar [27] for StarCraft II and DeepNash for Stratego [23], both of which use a model-free RL approach.

REFERENCES

[1] Robert Aumann and Adam Brandenburger. 1995. Epistemic conditions for Nash equilibrium. *Econometrica: Journal of the Econometric Society* (1995), 1161–1180.
[2] Robert Axelrod and William D Hamilton. 1981. The evolution of cooperation. *Science* 211, 4489 (1981), 1390–1396.
[3] David Balduzzi, Karl Tuyls, Julien Perolat, and Thore Graepel. 2018. Re-evaluating evaluation. *Thirty-Second International Conference on Neural Information Processing Systems* 31 (2018).
[4] Giacomo Bonanno and Klaus Nehring. 1999. How to make sense of the common prior assumption under incomplete information. *International Journal of Game Theory* 28, 3 (1999), 409–434.
[5] Neil Burch, Michael Johanson, and Michael Bowling. 2014. Solving imperfect information games using decomposition. In *Twenty-Eighth AAAI Conference on Artificial Intelligence.*
[6] Peter I Cowling, Edward J Powley, and Daniel Whitehouse. 2012. Information set Monte Carlo tree search. *IEEE Transactions on Computational Intelligence and AI in Games* 4, 2 (2012), 120–143.
[7] Ya'akov Gal and Avi Pfeffer. 2008. Networks of influence diagrams: A formalism for representing agents' beliefs and decision-making processes. *Journal of Artificial Intelligence Research* 33 (2008), 109–147.
[8] Joseph Y Halpern. 2002. Characterizing the common prior assumption. *Journal of Economic Theory* 106, 2 (2002), 316–355.
[9] Joseph Y Halpern and Rafael Pass. 2016. Sequential equilibrium in games of imperfect recall. In *Fifteenth International Conference on the Principles of Knowledge Representation and Reasoning.*
[10] John C Harsanyi. 1967. Games with incomplete information played by Bayesian players, Part I. The basic model. *Management Science* 14, 3 (1967), 159–182.
[11] Ehud Kalai and Ehud Lehrer. 1993. Rational learning leads to Nash equilibrium. *Econometrica: Journal of the Econometric Society* (1993), 1019–1045.
[12] Vijay Krishna. 2009. *Auction theory.* Academic Press.
[13] Marc Lanctot, Richard Gibson, Neil Burch, Martin Zinkevich, and Michael Bowling. 2012. No-regret learning in extensive-form games with imperfect recall. In *Twenty-Ninth International Conference on International Conference on Machine Learning.* 1035–1042.
[14] 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. In *Thirty-First International Conference on Neural Information Processing Systems.* 4190–4203.
[15] Zun Li, Marc Lanctot, Kevin McKee, Luke Marris, Ian Gemp, Daniel Hennes, Paul Muller, Kate Larson, Yoram Bachrach, and Michael .P Wellman. 2023. Search-Improved Game-Theoretic Multiagent Reinforcement Learning in General and Negotiation Games (Extended Abstract). In *Twenty-Second International Conference on Autonomous Agents and Multiagent Systems.*
[16] Zun Li and Michael Wellman. 2020. Structure learning for approximate solution of many-player games. In *Thirty-Fourth AAAI Conference on Artificial Intelligence.* 2119–2127.
[17] Zun Li and Michael P Wellman. 2021. Evolution strategies for approximate solution of Bayesian games. In *Thirty-Fifth AAAI Conference on Artificial Intelligence.* 5531–5540.
[18] H Brendan McMahan, Geoffrey J Gordon, and Avrim Blum. 2003. Planning in the presence of cost functions controlled by an adversary. In *Twentieth International Conference on Machine Learning.* 536–543.
[19] Matej Moravčík, Martin Schmid, Neil Burch, Viliam Lisỳ, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael Bowling. 2017.

Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science* 356, 6337 (2017), 508–513.
[20] Stephen Morris. 1995. The common prior assumption in economic theory. *Economics & Philosophy* 11, 2 (1995), 227–253.
[21] Shayegan Omidshafiei, Christos Papadimitriou, Georgios Piliouras, Karl Tuyls, Mark Rowland, Jean-Baptiste Lespiau, Wojciech M Czarnecki, Marc Lanctot, Julien Perolat, and Remi Munos. 2019. α-rank: Multi-agent evaluation by evolution. *Scientific reports* 9, 1 (2019), 1–29.
[22] Andrés Perea. 2012. *Epistemic game theory: Reasoning and choice*. Cambridge University Press.
[23] Julien Perolat, Bart De Vylder, Daniel Hennes, Eugene Tarassov, Florian Strub, Vincent de Boer, Paul Muller, Jerome T Connor, Neil Burch, Thomas Anthony, et al. 2022. Mastering the game of Stratego with model-free multiagent reinforcement learning. *Science* 378, 6623 (2022), 990–996.
[24] Michele Piccione and Ariel Rubinstein. 1997. On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior* 20, 1 (1997), 3–24.
[25] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, et al. 2018. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science* 362, 6419 (2018), 1140–1144.
[26] Finbarr Timbers, Edward Lockhart, Marc Lanctot, Martin Schmid, Julian Schrittwieser, Thomas Hubert, and Michael Bowling. 2020. Approximate exploitability: Learning a best response in large games. *arXiv preprint arXiv:2004.09677* (2020).
[27] Oriol Vinyals, Igor Babuschkin, Wojciech M Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H Choi, Richard Powell, Timo Ewalds, Petko Georgiev, et al. 2019. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature* 575, 7782 (2019), 350–354.
[28] Kevin Waugh, Martin Zinkevich, Michael Johanson, Morgan Kan, David Schnizlein, and Michael Bowling. 2009. A practical use of imperfect recall. In *Eighth symposium on abstraction, reformulation, and approximation*.
[29] Daan Wierstra, Tom Schaul, Tobias Glasmachers, Yi Sun, Jan Peters, and Jürgen Schmidhuber. 2014. Natural evolution strategies. *Journal of Machine Learning Research* 15, 1 (2014), 949–980.