Research Project Proposal: Efficient Solutions for Adversarial Team Games

Federico Cacciamani, federico.cacciamani@mail.polimi.it

1. Introduction to the Problem

The research is focused on Algorithmic Game Theory, that is the area of research that lies on the boundary between Economics and Computer Science. More specifically, Algorithmic Game Theory is the area that aims at modelling algorithms in strategic environments, with the objective of finding *strategies* that allow the involved agents to reach an *equilibrium*.

The strategic environments are modelled using mathematical tools, specifically Game Theory, while Computer Science techniques are used both to design algorithms and to evaluate problems' complexity and their *hardness*[1]. In particular, to find solutions some algorithms proper of the fields of Reinforcement Learning[2], Artificial Intelligence and Online Convex Optimization are used, while the evaluation of difficulty falls under the scope of Theoretical Computer Science.

1.1. Research Topic and its Importance

Algorithmic game theory is focused on designing algorithms with the goal to solve games and find its Nash Equilibria. Over the years most of the literature focused on two-players zero-sum games, and some of the best algorithms nowadays are those based on Counterfactual Regret minimization (CFR) [4]. One of the most successful results was, without any doubt, *Libratus* [13], an AI that was able to defeat in one versus one games four professional players in heads up no-limit Texas hold'em poker. At its core, Libratus is based on CFR but, as we can see, it took over ten years of extensive research from the first introduction of the algorithm to the complete formulation an AI able to outperform human professionals in a complex game like poker. Another example can be given for the game of StarCraft II where in 2019 Deepmind's AlphaStar [16], a system based on deep neural networks, succeeded in defeating two human pro players. Also in this case, the game has been object of research for over a decade and many competitions, like for instance AIIDE StarCraft AI Competition, have been launched over the years to test the progresses of the scientific community.

The characteristics of these games, like the imperfect information, the very large action spaces, and the long-term planning needed, made them very popular among the scientific community, mainly because of the possibility to map those scenarios to a multitude of real-world applications.

1.2. Problem and its Importance

Poker and Starcraft II are only two examples of a large group of games studied over the years in the context of algorithmic game theory, but except some games with very specific characteristics (e.g. *congestion games*) the community has been focusing on two-players zero-sum games. Also in this case, due to the multiple potential real-world applications (e.g. security, games like bridge or development of strategies for car races), the interest around team games has been growing significantly in the last years. The introduction of a team of players, however, brings several complications to the study of the game, for instance because a correlation between teammates strategies is needed in order to find an equilibrium that is satisfactory from the point of view of the team. Some important results have been proved but the study of team games is still open from a scientific point of view, and significant improvements can be reached in the next few years.

[1] In terms of computational complexity (e.g. NP-hardness).

[2] One of the three Machine Learning paradigms.

2. Main Related Works

The main related works to the topic of Adversarial Team Games can be divided among three categories: works on two-players games, works on team-games and works on reinforcement learning. In particular, for the field of reinforcement learning we focused on problems with large action spaces since in our opinion those situations can be more easily mapped to our case.

One of the most outstanding results for two-players games is without any doubt *Libratus* [13]. Libratus is based on an algorithm called Monte Carlo CFR [5], that is an improved version of the well known CFR [4]. MCCFR is not the only variation of CFR that has been formulated; the most relevant are CFR-BR [6], in which at each iteration one player plays according to CFR and the other plays in best response, and Deep-CFR [12], that approximated the behavior of CFR via deep neural network function approximation.

Another class of algorithms for two-player games is given by those based on *Fictitious Play* [1, 2]. In FP, at every iteration, each player plays the optimal pure strategy (best response) with respect to the average mixed strategy played by the opponent. In this context, the average strategy of each player converges to a Nash Equilibrium. Due to the complexity of computing a best response (NP-hard), several variations of FP have been developed. The most famous are Weakened Fictitious Play [3], in which best responses are approximated ($\epsilon - BR$) with $\epsilon \to 0$ as time progresses, and Fictitious Self Play [8] in which best response and average strategy update subroutines are approximated via a machine learning approach. Finally, Heinrich and Silver present Neural Fictitious Self Play in [10], a version of FSP that takes advantage of deep reinforcement learning.

For what concerns Adversarial Team Games, Baisilico et al. in [9] proved the inefficiency for a team of Nash Equilibria with respect to the Team Maxmin Equilibrium. Moreover, Celli and Gatti, in [11] proved that a general TME can be arbitrarily inefficient with respect to a TME in which team members are able to coordinate their strategy. To this extent, one of the most important results is given by Farina et al. in [14], where they present an algorithm based on Fictitious Play, called Fictitious Team Play. FTP takes advantage of an auxiliary representation of the game and of a MILP[3] formulation of the best response subroutine to find a Team Maxmin Equilibrium with a correlation device (TMEcor). While on the one hand their solution avoids the NP-hardness of computing the best response, on the other hand the adoption of the MILP limits the usage of the algorithm to small games, thus resulting in a limited scalability.

In the recent years, in part due to the huge devolpment of deep learning techniques, reinforcement learning gained an increasing importance in the research community. For those problems with large action spaces, a full exploration of the action space is impractical and smarter ways to exploit the gained experience must be introduced. The two approaches that can be more easily mapped to our case are those of Arnold et al., presented in [7] and of Chandak et al., presented in [15]. The former consists in the adoption of the so defined *Wolpertinger Architecture*, a deep neural network based on the classical actor-critic framework where the actor generates an action embedding given the current state and the critic evaluates the K closest actions[4] to the generated embedding, in order to select the most promising one. The second mentioned approach, on the other hand, is based on the encoding of a policy through a deep neural network, in which two different parts can be highlighted. The first one, similar to the actor in the Wolpertinger Architecture generates an action embedding given the current state, and the second one from the action embedding directly computes the action in the original space.

3. Research Plan

3.1. The Goal

The goal of our research is to develop an efficient algorithm to find equilibria of an adversarial team game, starting from the existent solutions in literature and adopting some reinforcement learning algorithms in order to improve the performances. Our intention is to start with the case of a single player versus a team and then possibly extend

[3] Mixed Integer Linear Program.

[4] Using a K Nearest Neighbour approach.

the study to the case of a team facing another team (*team* vs *team*) and of three or more teams playing against each other (*team* vs *team* vs *team* ...).

The approach we use is based on the adaptation of CFR-BR to the case of team facing single opponent, with, at each iteration, team playing as the BR player and opponent playing as the CFR one. The adoption of CFR-BR allows to rely on a compact representation of the team strategy while maintaining the convergence guarantees of CFR. In order to reduce the complexity of computing the best response, we propose to approximate it using some deep learning approach.

3.2. The Process

Our research can be divided into the following main tasks:

- Extension of the CFR-BR algorithm to the case of single player versus team:
  - Theoretical formulation;
  - Formulation of best response as a reinforcement learning combinatorial problem;
  - Formulation of best response as a deep reinforcement learning combinatorial problem;
  - Theoretical guarantees derivation;
- Experimental evaluation: contract bridge, goofspiel;
- Paper and M.Sc. thesis devel opment.

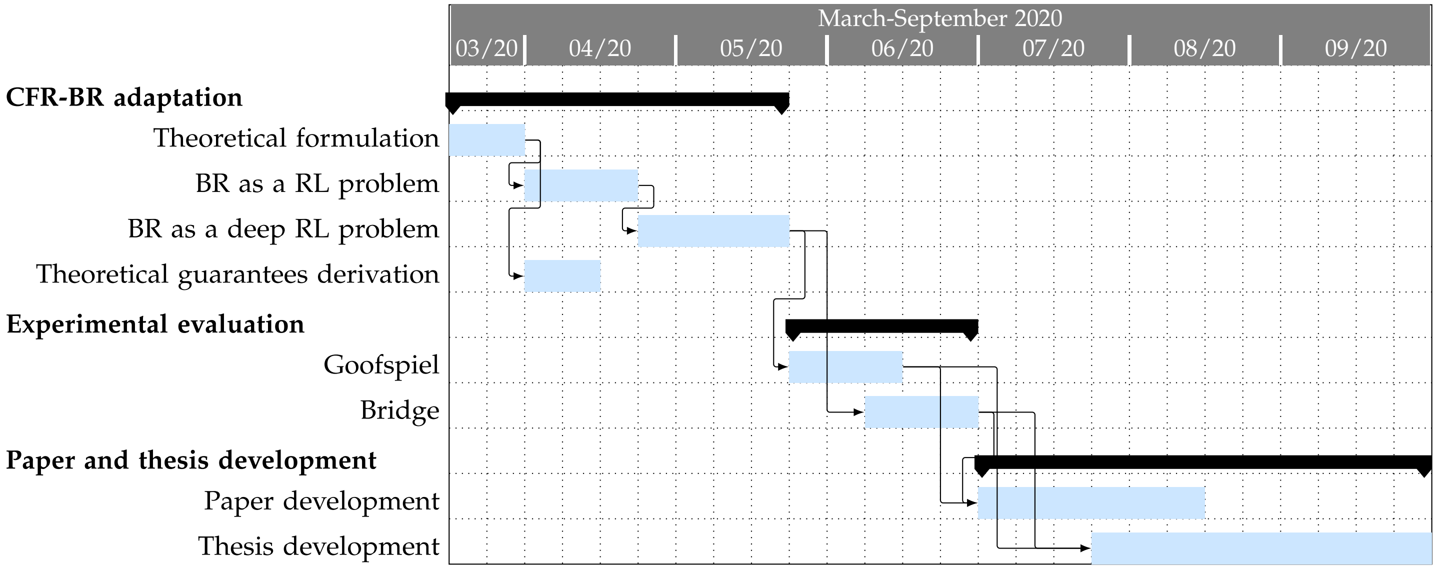

Figure 1: Gantt Chart of Research Activities

3.3. Research Evaluation

The metrics that will be used to evaluate the output of our research will be based on the following:

- Convergence rate of algorithms;
- Quality of the approximations;
- Performance of experimental implementations, calculated through applications specific results.

References

[1] G. W. Brown. “Iterative solution of games by fictitious play”. In: *Activity analysis of production and allocation* (1951), pp. 374–376.

[2] J. Robinson. “An iterative method of solving a game”. In: *Annals of mathematics* (1951), pp. 296–301.

[3] Ben van der Genugten. “A Weakened Form of Fictitious Play in Two-Person Zero-Sum Games”. In: *IGTR* 2 (2000), pp. 307–328.

[4] Martin Zinkevich, Michael Johanson, Michael Bowling, and Carmelo Piccione. “Regret Minimization in Games with Incomplete Information”. In: *Advances in Neural Information Processing Systems 20*. Ed. by J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis. Curran Associates, Inc., 2008, pp. 1729–1736. URL: http://papers.nips.cc/paper/3306-regret-minimization-in-games-with-incomplete-information.pdf.

[5] Marc Lanctot, Kevin Waugh, Martin Zinkevich, and Michael Bowling. “Monte Carlo Sampling for Regret Minimization in Extensive Games”. In: *Advances in Neural Information Processing Systems 22*. Ed. by Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta. Curran Associates, Inc., 2009, pp. 1078–1086. URL: http://papers.nips.cc/paper/3713-monte-carlo-sampling-for-regret-minimization-in-extensive-games.pdf.

[6] Michael Johanson, Nolan Bard, Neil Burch, and Michael Bowling. “Finding Optimal Abstract Strategies in Extensive-Form Games”. In: *AAAI*. 2012. URL: http://www.aaai.org/ocs/index.php/AAAI/AAAI12/paper/view/5156.

[7] Gabriel Dulac-Arnold, Richard Evans, Peter Sunehag, and Ben Coppin. “Reinforcement Learning in Large Discrete Action Spaces”. In: *CoRR* abs/1512.07679 (2015). arXiv: 1512.07679. URL: http://arxiv.org/abs/1512.07679.

[8] Johannes Heinrich, Marc Lanctot, and David Silver. “Fictitious Self-Play in Extensive-Form Games”. In: *Proceedings of the 32nd International Conference on Machine Learning*. Ed. by Francis Bach and David Blei. Vol. 37. Proceedings of Machine Learning Research. Lille, France: PMLR, July 2015, pp. 805–813. URL: http://proceedings.mlr.press/v37/heinrich15.html.

[9] Nicola Basilico, Andrea Celli, Giuseppe De Nittis, and Nicola Gatti. “Team-maxmin equilibrium: efficiency bounds and algorithms”. In: *CoRR* abs/1611.06134 (2016). arXiv: 1611.06134. URL: http://arxiv.org/abs/1611.06134.

[10] Johannes Heinrich and David Silver. “Deep Reinforcement Learning from Self-Play in Imperfect-Information Games”. In: *CoRR* abs/1603.01121 (2016). arXiv: 1603.01121. URL: http://arxiv.org/abs/1603.01121.

[11] Andrea Celli and Nicola Gatti. “Computational Results for Extensive-Form Adversarial Team Games”. In: *CoRR* abs/1711.06930 (2017). arXiv: 1711.06930. URL: http://arxiv.org/abs/1711.06930.

[12] Noam Brown, Adam Lerer, Sam Gross, and Tuomas Sandholm. “Deep Counterfactual Regret Minimization”. In: *CoRR* abs/1811.00164 (2018). arXiv: 1811.00164. URL: http://arxiv.org/abs/1811.00164.

[13] Noam Brown and Tuomas Sandholm. “Superhuman AI for heads-up no-limit poker: Libratus beats top professionals”. In: *Science* 359.6374 (2018), pp. 418–424. DOI: 10.1126/science.aao1733.

[14] Gabriele Farina, Andrea Celli, Nicola Gatti, and Tuomas Sandholm. “Ex ante coordination and collusion in zero-sum multi-player extensive-form games.” In: *NeurIPS*. Ed. by Samy Bengio et al. 2018, pp. 9661–9671. URL: http://dblp.uni-trier.de/db/conf/nips/nips2018.html#FarinaCGS18.

[15] Yash Chandak, Georgios Theocharous, James Kostas, Scott M. Jordan, and Philip S. Thomas. “Learning Action Representations for Reinforcement Learning”. In: *CoRR* abs/1902.00183 (2019). arXiv: 1902.00183. URL: http://arxiv.org/abs/1902.00183.

[16] Oriol Vinyals et al. *AlphaStar: Mastering the Real-Time Strategy Game StarCraft II*. https://deepmind.com/blog/alphastar-mastering-real-time-strategy-game-starcraft-ii/. 2019.