Strategy Effectiveness of Game-Theoretical Solution Concepts in Extensive-Form General-Sum Games

(Extended Abstract)

Jiří Čermák
Dept. of Computer Science
FEE, CTU in Prague
cermak@agents.fel.cvut.cz

Branislav Bošanský
Dept. of Computer Science
Aarhus University
bosansky@cs.au.dk

Nicola Gatti
Dept. of El. and Information
Politecnico di Milano
nicola.gatti@polimi.it

ABSTRACT

Game theory describes the conditions for the strategies of rational agents to form an equilibrium. However, game theory can fail from the prescriptive viewpoint and can serve only as a heuristic recommendation for agents. There exists a plethora of game theoretic solution concepts, however, their effectiveness has never been compared; hence, there is no guideline for selecting correct algorithm for a given domain. Therefore, we compare the effectiveness of solution-concept strategies and strategies computed by Counterfactual regret minimization (CFR) and Monte-Carlo tree search in practice. Our results show that (1) CFR strategies are typically the best, and (2) the effectiveness of the refinements of NE depends on the utility structure of the game.

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent systems

General Terms

Algorithms, Performance, Experimentation

Keywords

Game Theory; Solution concepts; General-sum games

1. INTRODUCTION

Non-cooperative game theory provides mathematical models to deal with strategic interactions. In spite of the recent enormous success of the game-theoretic applications, game theory is a *descriptive theory* describing equilibrium conditions (*solution concepts*) for strategies of rational agents (e.g., Nash equilibrium (NE)). It may fail when used to *prescribe* strategies when designing artificial agents (except for specific classes of games). Failure appears, e.g., when multiple NE exist in a game. Game theory does not specify which NE to choose and playing strategies from different NE may lead to an arbitrarily bad outcome for the agents.

In this paper we provide a thorough experimental evaluation of the effectiveness of the strategies on general-sum two-player finite extensive-form games (EFGs) with imperfect information, that are either described by solution concepts, or that are a result of learning algorithms, extending in this way the work conducted on zero-sum games [9].

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May 4–8, 2015, Istanbul, Turkey.*

2. STRATEGY TYPES

We compare 3 different solution concepts and 2 strategies resulting from iterative algorithms.

Nash equilibrium (NE) is a strategy profile, where all players play a best response (BR) to the strategies of their opponents. The most common method for finding NE of two-player EFGs with general sum is to use the LCP formulation [4] or the MILP reformulation of the LCP [1]. An undominated equilibrium (UND) is a NE using undominated strategies in the sense of weak dominance. UND is computed by adding the objective maximizing the expected value against uniform strategy of the player 2 to the MILP. Quasi-perfect equilibrium (QPE) [7] further restricts UND by ensuring sequential rationality. To compute QPE, we use LCP with symbolic perturbations [6].

Counterfactual regret minimization (CFR) [8] iteratively traverses the whole game tree, updating the strategy with an aim to minimize the overall regret. The Monte Carlo Tree Search (MCTS) iteratively evaluates the domain using a huge number of simulations, while building the tree of the most promising nodes. UCB algorithm [3] is used as the selection method in each information set. Nesting [2] is used to get equally reasonable strategies in the whole game tree.

3. EXPERIMENT SETTINGS

Player 1 uses a strategy prescribed by solution concept, CFR or MCTS, and we measure the expected outcome for player 1 against some strategy of player 2.

We use randomly generated games (RND). We alter the *depth* of the game (number of moves for *each* player), the *branching factor* (actions available in each information set); number of observation signals (more observation signals imply more information available to players). Furthermore we control the utility correlation of players using parameter from $[-1, 1]$ (1 for identical utilities, -1 for zero-sum games).

3.1 Imperfect Opponents

We use three types of the opponents to evaluate the strategy effectiveness. First, we use CFR and MCTS and we measure how the expected utility for different solution concepts changes with increasing number of iterations used. Second, we use a human behavior approximation, quantal-response

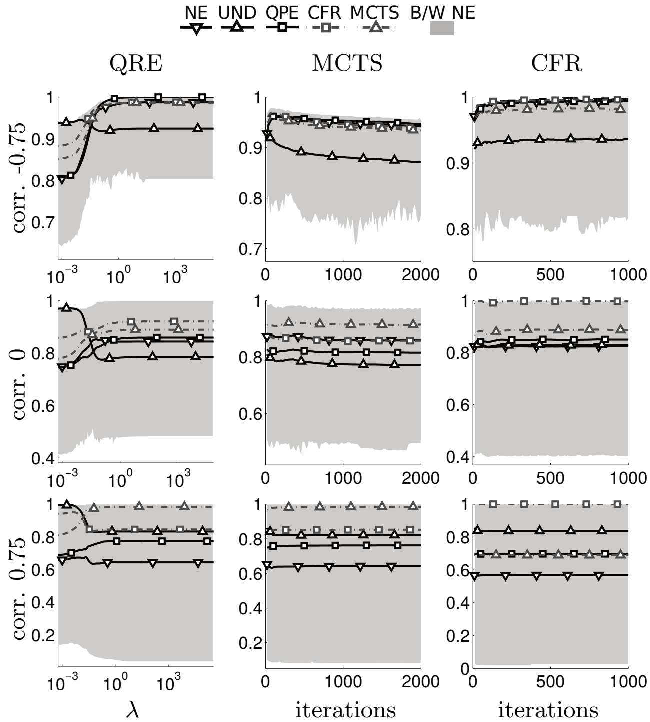

Figure 1: Relative effectiveness on RNDs.

equilibrium (QRE) [5] parametrized by λ parameter (increasing λ increases the rationality of resulting strategy).

3.2 Measuring the Effectiveness of Strategies

We present the relative effectiveness of the strategies. The upper and lower bound are formed by the best and the worst response against each strategy of the opponent (relative effectiveness is equal to 1 if the strategy scores the same as the best response, 0 if it scores the same as the worst response). We also plot the interval for best and worst NE strategies computed using MILP formulation for finding NE. This interval in which all NE and its refinements belong is visualized as a grey area in the results.

Finally, we use CFR and MCTS strategies which used 10^5 iterations to learn for the effectiveness comparison (the CFR and MCTS learned in self-play with no information about strategies they will face).

4. RESULTS

The curves in graphs of Fig. 1 represent the means of relative expected values over 50 RNDs with branching factor 2, depth 2 and varying utility correlation against the CFR, MCTS and QRE opponents (y-axis shows the expected utility of strategies, x-axis shows iterations for CFR and MCTS opponent and the λ value for QRE opponent).

The CFR strategies typically outperform all other and their effectiveness is often close to the BR. This is caused by the fact that CFR builds the opponent model tailored to the given domain when learning its strategy, while the rest of the solution concepts assume fixed opponent model.

MCTS was the second best, however, its effectiveness is not consistent. It is weak, e.g., against the CFR on RNDs with correlation 0.75 (3rd row 3rd graph in Fig. 1). On the other hand, it has the best effectiveness, e.g., against the QRE on RNDs with correlation 0.75 (3rd row 1st graph).

The effectiveness of QPE strategies is high in the games with negative utility correlation, often close to CFR (1st row in Fig. 1). However, it decreases with increasing correlation factor (3rd row in Fig. 1). This is due to the fact that the QPE exploits the mistakes of the opponent. If the correlation is negative, the mistakes of player 2 help player 1, and thus their exploitation significantly improves the expected utility. However as the correlation increases the advantages player 1 gets from the mistakes of the player 2 diminish, since the mistakes of player 2 decrease also the utility for player 1, and so the effectiveness of QPE decreases.

UND strategies are often very weak for the negative correlation (1st row in Fig. 1), but their effectiveness increases with higher correlation (3rd row in Fig. 1). This is because we compute UND using uniform strategy in the objective of MILP; hence, the strategy of player 1 is trying to reach leafs with high outcome through the whole game tree. As the utility correlation increases, there is a higher chance that player 2 will try to reach the same leafs.

Additional experiments with varying amount of information and sizes of the randomly generated games and additional poker domains confirm the presented findings.

Acknowledgements

This research was supported by the Czech Science Foundation (grant no. P202/12/2054 and 15-23235S), by the Danish National Research Foundation and The National Science Foundation of China (under the grant 61361136003) for the Sino-Danish Center for the Theory of Interactive Computation.

REFERENCES

[1] C. Audet, S. Belhaiza, and P. Hansen. A new sequence form approach for the enumeration and refinement of all extreme Nash equilibria for extensive form games. *International Game Theory Review 11*, 2009.

[2] H. Baier and M. H. M. Winands. Nested Monte-Carlo Tree Search for Online Planning in Large MDPs. In *ECAI*, 2012.

[3] L. Kocsis, C. Szepesvári, and J. Willemson. Improved Monte-Carlo Search. *Tech. Rep 1*, 2006.

[4] D. Koller, N. Megiddo, and B. von Stengel. Efficient computation of equilibria for extensive two-person games. *Games and Economic Behavior*, 1996.

[5] R. D. McKelvey and T. R. Palfrey. Quantal response equilibria for normal form games. *Games and economic behavior*, 10(1):6–38, 1995.

[6] P. B. Miltersen and T. B. Sørensen. Computing a quasi-perfect equilibrium of a two-player game. *Economic Theory*, 2008.

[7] E. van Damme. A relation between perfect equilibria in extensive form games and proper equilibria in normal form games. *Game Theory*, 13:1–13, 1984.

[8] M. Zinkevich, M. Bowling, and N. Burch. A new algorithm for generating equilibria in massive zero-sum games. In *AAAI*, 2007.

[9] J. Čermák, B. Bošanský, and V. Lisý. Practical Performance of Refinements of Nash Equilibria in Extensive-Form Zero-Sum Games. In *ECAI*, 2014.