The Gift Exchange Game: Managing Opponent Actions

Extended Abstract

Steven Damer
University of Minnesota
Minneapolis, Minnesota, USA
damer@cs.umn.edu

Maria Gini
University of Minnesota
Minneapolis, Minnesota, USA
gini@umn.edu

Jeffrey S. Rosenschein
The Hebrew University of Jerusalem
Jerusalem, Israel
jeff@cs.huji.ac.il

ABSTRACT

Interacting with an opponent is a fundamental concern in multi-agent systems. In this work, we consider ways in which an agent can manipulate an opponent to adopt a preferred strategy. This difficult problem is often further complicated by the difficulty of analyzing the game. We have developed the Gift Exchange game, a sequential game that is deliberately simplified to focus on how to interact with an opponent. In this paper we describe the game and discuss different methods an agent might use to influence its opponent to select a preferred action. We show results from using simulated annealing to find optimal strategies to use against a learning opponent.

KEYWORDS

Repeated sequential games; cooperation; non-stationary opponents

ACM Reference Format:
Steven Damer, Maria Gini, and Jeffrey S. Rosenschein. 2019. The Gift Exchange Game: Managing Opponent Actions. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 3 pages.

1 INTRODUCTION

The problem of how to interact with an opponent has been studied extensively in many games and environments [1–11, 13, 14]. When an agent engages in repeated interactions with an opponent, the behavior of the opponent can have a significant impact on the outcome experienced by the agent.

Many factors complicate interactions with an opponent. First, it may be difficult to determine the effects of an agent's actions, either because of hidden information or because the effects partially depend on the simultaneous action of the opponent. Secondly, it may be difficult to determine the value of a game state because of the complexity of analyzing the game. In order to focus on interacting with an opponent without dealing with these problems, we have developed a new game, the *Gift Exchange* game, that has been deliberately constructed to make these problems trivial.

In the Gift Exchange game players act sequentially and there is no hidden information, so the agent will always know the effects of each of its actions and the intended effect of actions chosen by the opponent. The only state information in the game is which player will act next, so a player choosing a move only needs to consider the move's immediate effects (which are public information) and how the move will affect the future behavior of the opponent.

We explore two models of how an agent may affect the future behavior of the opponent: the opponent may be following a strategy with a simple model that the agent can learn and for which it can construct a best response, or the opponent may be attempting to learn the agent's strategy and the agent can select a reciprocating strategy to influence the opponent's behavior.

2 THE GIFT EXCHANGE GAME

In the game, agents take turns choosing actions. Each action consists of a choice from a set of potential outcomes, and each outcome is an assignment of (potentially negative) payoffs to the agent and its opponent. The set of potential outcomes is the unit circle, where one player receives the x-coordinate of the chosen point and the other player receives the y-coordinate of the chosen point.

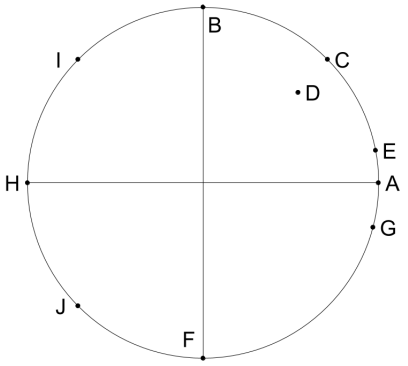

Figure 1: The choice set for the Gift Exchange game. All points in the circle are possible choices for either player (but not necessarily rational). The x-axis is the agent's payoff and the y-axis is the payoff of its opponent.

Figure 1 shows the set of choices with some noteworthy options highlighted. The *greedy choices* for the agent and the opponent are A and B respectively. C is the *social welfare maximizing choice*. D is also cooperative, but not pareto-efficient (both agents could make more by playing C instead). E favors the agent, but is slightly beneficial for the opponent. F is the *maximally punishing choice* for the opponent. G is a slightly punishing choice that also gives the agent a nearly optimal result. H is the maximally punishing choice for the agent. I would be played by a competitive opponent that

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.

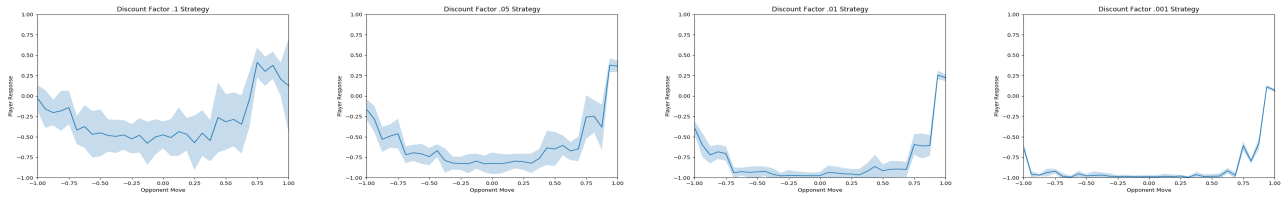

Figure 2: The strategies found by simulated annealing against a opponent using UCT, when the agent has discount factors of .1, .05, .01, and .001 from left to right. The shaded area shows the 95% confidence interval. The x-axis is the amount of the opponent's gift to the agent and the y-axis is the amount that the agent gives to the opponent in return.

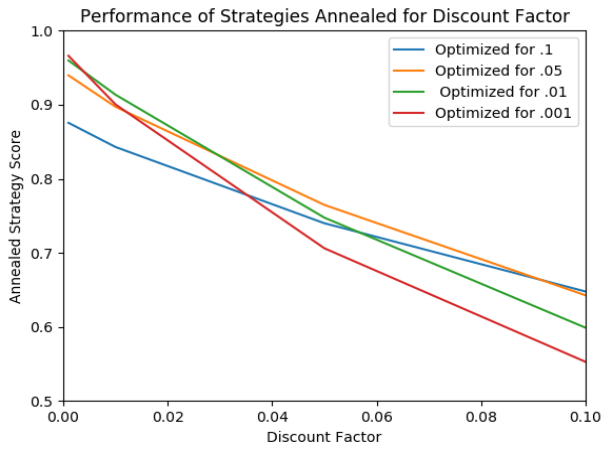

Figure 3: Discounted average payoff of strategies optimized for different discount factors. The x-axis is the discount factor used to evaluate the strategies.

seeks to maximize the difference in scores instead of maximizing its own score. J is an irrational choice that neither player would choose, because it penalizes both. If both players are rational, the agent will only choose points along the arc $B-A-F$ and the opponent points along the arc $H-B-A$. We will restrict our attention to agents that play rational strategies.

3 GIFT EXCHANGE GAME STRATEGIES

If the opponent is following a sufficiently simple strategy it may be possible for the agent to learn the opponent strategy and calculate an optimal response. An alternate approach would be to commit to a reciprocating strategy so that the best response for the opponent is to choose a strategy that is beneficial to the agent. We will now explore in a more systematic way opponent strategies.

An *immediately reactive* player is one that only considers the last action of the opponent when selecting its action. We have developed a variant of UCT [12] to allow a player to learn the best response to an immediately reactive opponent. It is based on representing the interval $[-1, 1]$ as a binary tree. An agent can use this algorithm to find the best response to an immediately reactive opponent strategy.

An opponent that always best-responds to an agent can be taken advantage of by adopting a reciprocating strategy. One example is an agent that responds to an opponent move which gives the agent a payoff of p by giving the opponent $\min(-\tau \times x + \sqrt{\tau^2 x^2 - \tau^2 - x^2 + 1}, \sqrt{1-\tau^2})$ where $\tau < 1$ is a target value chosen by the agent and x is $\tau\sqrt{1-p^2} - p\sqrt{1-\tau^2}$. When the agent adopts that strategy, the ratio between the agent's payoff and the opponent's payoff will be at least $\frac{\tau}{\sqrt{1-\tau^2}}$. Against this agent the best response for the opponent will be to give the agent τ each round. The agent can select any value for τ; as long as the opponent best responds, the agent can achieve a payoff arbitrarily close to 1.

When the opponent always plays the best response, the agent can demand a payoff arbitrarily close to 1, but if the opponent has to learn the best response, greedier strategies will require more time to learn. This is relevant when the agent discounts future payoffs. We have used simulated annealing to determine how the discount factor of the agent affects the best immediately reactive strategy to play against a learning opponent. We have used simulated annealing to find the best strategy to play against a learning opponent and observe how that strategy depends on the agent's discount factor. Figure 2 shows the strategies found for varying discount factors. As the discount factor approaches zero, the strategies get greedier and more punitive because the agent is more willing to pay the cost of punishing the opponent to receive a better outcome in the future. Figure 3 shows how the different strategies perform for agents with different discount factors. Strategies developed for agents with low discount factors achieve a higher payoff when the discount factor is low, but drop off more rapidly as the discount factor rises because it takes the opponent longer to learn the optimal response.

4 CONCLUSIONS

We have proposed the Gift Exchange game, a simple game suitable for examining the effect of an agent's play on future actions of the opponent. In the Gift Exchange game the intended outcome of a player's action is publicly observable so an agent can focus on the problem of manipulating the opponent to select a preferred outcome. We have described a class of reciprocating functions that take advantage of a learning opponent. By using simulated annealing we can find the most appropriate function for an agent with a specific discount factor. We have shown that agents with lower discount factors adopt greedier and more punitive strategies.

Acknowledgements: This work was supported in part by Israel Science Foundation grant #1340/18.

REFERENCES

[1] Stefano V. Albrecht and Peter Stone. 2018. Autonomous agents modelling other agents: A comprehensive survey and open problems. *Artificial Intelligence* 258 (2018), 66 – 95.

[2] Tsz-Chiu Au and Dana S. Nau. 2006. Accident or intention: That is the question (in the noisy iterated prisoner's dilemma). In *Proc. Int'l Conf. on Autonomous Agents and Multi-Agent Systems (AAMAS)*. 561–568.

[3] R. M. Axelrod. 1984. *The evolution of cooperation.* Basic Books.

[4] Tim Baarslag, Mark J. C. Hendrikx, Koen V. Hindriks, and Catholijn M Jonker. 2016. Learning about the opponent in automated bilateral negotiation: a comprehensive survey of opponent modeling techniques. *Journal of Autonomous Agents and Multi-agent Systems* 30, 5 (2016), 849–898.

[5] Trevor Bench-Capon, Katie Atkinson, and Peter McBurney. 2009. Altruism and agents: an argumentation based approach to designing agent decision mechanisms. In *Proc. Int'l Conf. on Autonomous Agents and Multi-Agent Systems (AAMAS)*. 1073–1080.

[6] Michael Bowling and Manuela Veloso. 2002. Multiagent learning using a variable learning rate. *Artificial Intelligence* 136 (2002), 215–250.

[7] Andriy Burkov and Brahim Chaib-draa. 2007. Multiagent learning in adaptive dynamic systems. In *Proc. Int'l Conf. on Autonomous Agents and Multi-Agent Systems (AAMAS)*. 41:1–41:6.

[8] Colin F Camerer, Teck-Hua Ho, and Juin-Kuan Chong. 2004. A Cognitive Hierarchy Model of Games. *The Quarterly Journal of Economics* 119, 3 (2004), 861–898.

[9] Doran Chakraborty and Peter Stone. 2010. Online model learning in adversarial Markov Decision Processes. In *Proc. Int'l Conf. on Autonomous Agents and Multi-Agent Systems (AAMAS)*. 1583–1584.

[10] Vincent Conitzer and Tuomas Sandholm. 2007. AWESOME: A General Multiagent Learning Algorithm that Converges in Self-Play and Learns a Best Response Against Stationary Opponents. *Machine Learning* 67, 1–2 (2007), 23–43.

[11] Michael Johanson, Martin Zinkevich, and Michael Bowling. 2007. Computing robust counter-strategies. In *Advances in Neural Information Processing Systems (NIPS)*. 721–728.

[12] Levente Kocsis, Csaba Szepesvári, and Jan Willemson. 2006. Improved Monte-Carlo search. *Univ. Tartu, Estonia, Tech. Rep* 1 (2006).

[13] Michael L. Littman. 2001. Friend-or-Foe Q-learning in General-Sum Games. In *Proc. of the Int'l Conf. on Machine Learning*. 322–328.

[14] R. Powers, Y. Shoham, and T. Vu. 2007. A general criterion and an algorithmic framework for learning in multi-agent systems. *Machine Learning* 67, 1–2 (2007), 45–76.