Cooperation between Self-Interested Agents in Normal Form Games

(Extended Abstract)

Steven Damer
Dept. of Computer Science and Engineering, University of Minnesota
Minneapolis, MN 55455, USA
damer@cs.umn.edu

ABSTRACT

We study how to achieve cooperation between two self-interested agents that play repeated randomly generated normal form games. We take inspiration from a model originally designed to identify cooperative actions by humans who play a game, but we use the model in a prescriptive rather than descriptive manner. To identify cooperative intent, agents use a particle filter to learn the parameters of the model.

Categories and Subject Descriptors

I.2.11 [**Distributed AI**]: Multiagent systems

General Terms

Design, Economics

Keywords

Implicit Cooperation, Game theory, Multiagent Learning

1. INTRODUCTION AND MODEL

For our study we use randomly generated games with 16 actions per player and payoffs uniformly distributed between 0 and 1. Players only see each game once – they need to reason about the opponent's past behavior in different games in order to predict its behavior in the current game. This enables us to study the problem of identifying what constitutes cooperation in an unpredictable environment.

The model we use to identify cooperative behavior has been proposed to explain human cooperation in [3]. Agents value their opponent's payoffs as well as their own. In the model, which we presented in [1], agents adopt an *attitude* towards their opponent. Attitude is a real number which indicates the agent's intent. An attitude of 1 indicates a very helpful agent, an attitude of 0 indicates an indifferent agent, and an attitude of -1 indicates a hostile agent.

Given agents x and y with attitudes A^x and A^y, each agent constructs a modified game with a different payoff matrix. The modified payoff matrix $P^{'x}$ of agent x is $P^{'x}_{ij} = P^x_{ij} + A^x P^y_{ij}$ where P^x_{ij} is the payoff in the original game for player x and P^y_{ij} is the payoff for the opponent when they choose respectively actions i and j. The modified payoff matrix of agent y can be computed similarly, using its attitude A^y.

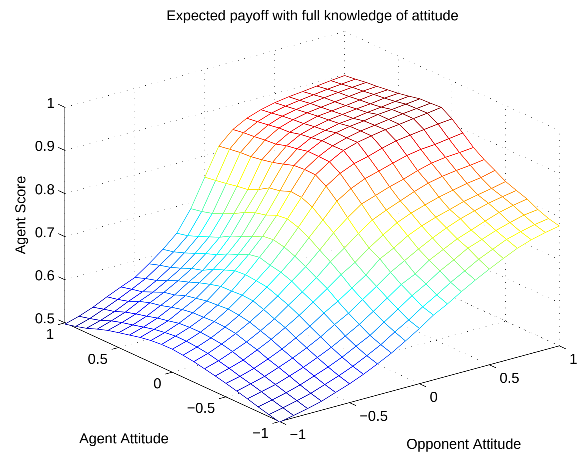


Figure 1: Effect of attitude on agent payoff. The agent's attitude is on the left axis, going from full cooperation (1) to full selfishness (-1). The opponent's attitude is on right axis. Results are aggregated over 1000 games.

Agents then act according to a Nash equilibrium of the modified game, but receive payoffs from the original game. Figure 1 shows the effect of different attitude values on an agent's payoff. The most significant factor in an agent's payoff is the attitude of its opponent, with a higher attitude resulting in a better outcome for the agent. The second most significant factor is the agent's own attitude – unsurprisingly a more self-interested agent achieves a better payoff. There is one particularly surprising effect which can be observed in Figure 1. When the opponent has a positive attitude, an agent no longer suffers for increasing its attitude above 0. An agent can even gain by increasing its attitude from 0 to .1 when the opponent's attitude is 1. This shows that there are opportunities for cooperation. It is important to note that these are aggregate results. For a particular game the general shape will be similar, but it will not be so smooth.

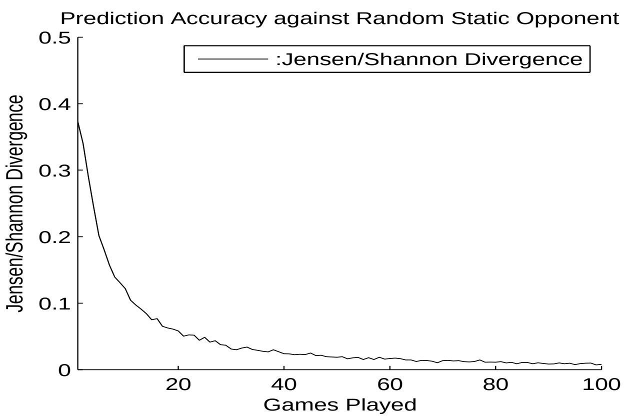
There are multiple parameters which can be varied, most notably the number of actions available to each agent, and the distribution from which payoffs are drawn. Increasing

Cite as: Cooperation between Self-Interested Agents in Normal Form Games (Extended Abstract), Steven Damer, *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems – Innovative Applications Track (AAMAS 2011)*, Tumer, Yolum, Sonenberg and Stone (eds.), May, 2–6, 2011, Taipei, Taiwan, pp. 1367-1368.

the number of actions does not have a significant effect, but decreasing their number simplifies the environment and the plateau is no longer observed – agents payoffs increase solely with how generous the opponent is and how selfish they are. Drawing payoffs from a Gaussian distribution also simplifies the environment, but to a lesser degree. Details in [2].

2. LEARNING

When agents' attitudes and their choice of Nash equilibrium are public knowledge the model produces cooperative outcomes. However, a self-interested agent is motivated to conceal its attitude. In order to avoid exploitation it is necessary for an agent to learn its opponent's attitude by observing its actions. An agent acting according to this model uses 3 parameters to select its action: its own attitude, its opponent's attitude, and a choice of Nash equilibrium of the modified game. By using a regularized particle filter we have shown [2] that an agent can learn what parameters its opponent is using well enough to provide a good prediction of opponent behavior.



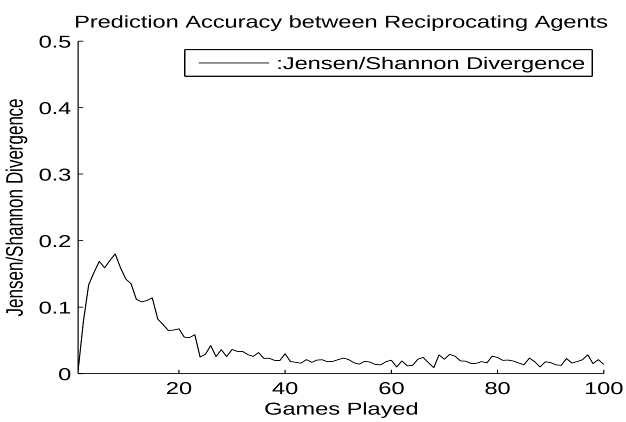


Figure 2: Prediction accuracy (top) against a random stationary opponent and (bottom) in self-play. Results aggregated over 100 sequences of 100 games.

Figure 2 shows the performance of a regularized particle filter learning a target in this environment. The prediction error is the Jensen-Shannon divergence between the predicted and actual strategy chosen by the opponent. The top graph shows the error in the prediction of the opponent action for a random stationary opponent, with learning targets drawn from a Gaussian distribution with 0 mean. The bottom graph show the prediction error between two learning agents, each reciprocating the opponent's attitude with a bonus of .1. This does not create substantial risk (since its attitude is never significantly higher than its opponent's) but it allows both agents to eventually reach a maximally cooperative attitude of 1. Despite the fact the interactions are very complex it takes only around 20 games to learn the opponent's behavior with reasonable certainty. This is a small number compared to the thousand of games that are typically needed to learn.

Reducing the number of actions increases the speed of learning to predict the opponent's action, but reduces the speed at which the model is learned. Drawing from a different random distribution does not have a significant effect on learning. Prediction is not significantly affected if agents' payoffs are positively or negatively correlated, but model accuracy can drop. If agents actions have an independent effect on payoffs some aspects of the model become unlearnable (since they no longer affect agents' actions) but it becomes a good predictor very rapidly, because it is no longer necessary to learn what the opponent expects the agent to do.

One advantage of using particle filters is that they can easily be adapted to a non-stationary target. We have successfully learned targets that drift randomly as well as targets which are occasionally replaced by a different target. As long as the motion is not too rapid (such as a target which is replaced every other game), learning can still be done.

3. FUTURE WORK AND CONCLUSIONS

One issue with our model is how agents choose strategies once they have chosen an attitude to adopt. We currently assume they play a strategy which is part of a Nash equilibrium. When playing against a random stationary opponent, they use best response. Playing best response is risky, so we are looking into a partial best response strategy.

The model of reciprocation we use is simple and does not take into account all factors. For example, it is not capable of detecting an opponent that cooperates when the stakes are low and does not cooperate when the stakes are high. We are planning on developing a more sophisticated model of reciprocation with some notion of debt or obligation. We will also look at real domains to study how our model can be applied.

Our main contribution is a model which can achieve cooperative outcomes between two self-interested agents in a wide variety of normal form games, where agents can use reciprocation to achieve cooperation without exposing themselves to the risk of exploitation. To determine the opponent's hostile or cooperative intent, the model parameters are learned using a particle filter.

4. REFERENCES

[1] S. Damer and M. Gini. Achieving cooperation in a minimally constrained environment. In *Proc. of the Nat'l Conf. on Artificial Intelligence*, pages 57–62, 2008.

[2] S. Damer and M. Gini. Learning to cooperate in normal form games. In *Interactive Decision Theory and Game Theory Workshop, AAAI 2010*, July 2010.

[3] N. Frohlich. Self-Interest or Altruism, What Difference? *Journal of Conflict Resolution*, 18(1):55–73, 1974.