Friend or Foe? Detecting an Opponent's Attitude 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

Maria Gini
Dept. of Computer Science and Engineering
University of Minnesota
Minneapolis, MN 55455, USA
gini@cs.umn.edu

ABSTRACT

We study the problem of achieving cooperation between two self-interested agents that play a sequence of different randomly generated normal form games. The agent learns how much the opponent is willing to cooperate and reciprocates. We present empirical results that show that both agents benefit from cooperation and that a small number of games is sufficient to learn the cooperation level of the opponent.

Categories and Subject Descriptors

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

General Terms

Design, Economics

Keywords

Implicit Cooperation, Game theory, Multiagent Learning

1. INTRODUCTION

We extend the work in [2, 3] where two self-interested players play a sequence of non-zero-sum normal form games, each game played only once by the same two players. Since each game is played only once, agents cannot rely on past observations to predict the opponent's behavior, but since they play repeatedly against each other they can observe each other and reciprocate past positive interactions.

Reciprocation is a strategy used successfully in nature, in artificial environments such as iterated prisoner's dilemma, and by people [4]. Our agent decides how to reciprocate by learning the level of cooperation of the opponent, which we call the opponent's *attitude*, and setting its own attitude to be slightly higher than the attitude of its opponent.

As in [2], an attitude is a real number in the range [-1, 1]. An attitude of 1 means that the opponent's payoff is valued as highly as the agent's own payoff. An attitude of 0 means that the agent is indifferent to the opponent's payoff. An attitude of -1 means the agent is only concerned with how well it does compared to its opponent.

Given agents x and y with attitudes A^x and A^y, a modified game is created with payoff functions $P'^x_{ij} = P^x_{ij} + A^x P^y_{ij}$ and $P'^y_{ij} = P^y_{ij} + A^y P^x_{ij}$, where P^x_{ij} and P^y_{ij} are the payoffs in the original game respectively for agent x and y when they choose actions i and j. An agent selects its action in the modified game, but receives its payoff according to the original game. We have shown [2] empirically that when both agents have a positive attitude, their payoffs in the original game are higher than if they had both simply tried to maximize their individual scores.

For simplicity, we assume that agents play a strategy which is part of a Nash equilibrium. The Nash equilibrium is computed by the agent in the modified game, where the payoffs are changed to reflect its willingness to cooperate. This is convenient since it limits the choices to a discrete set (i.e. one among the Nash equilibria for each game). We do not assume both agents use the same Nash equilibrium.

2. LEARNING AND RESULTS

An agent which uses this model to act needs 3 parameters – an attitude value for itself, an attitude value for its opponent, which we call *belief*, and a method of choosing a Nash equilibrium from the modified game. In every round the agent observes the payoff matrix of the game and the action chosen by the opponent in that context. From that information, it needs to learn a probability distribution over the attitude, belief, and method of the opponent.

Due to the complex interactions between attitude, belief, method, and the game being played, it is not possible to analytically update a probability distribution over those factors. However, given specific values for attitude, belief, and method we can compute the probability that the agent would select a particular action in a given game. This enables the agent to use a regularized particle filter to track a probability distribution over attitude, belief, and method.

A particle filter represents a probability distribution with a number of samples drawn from it, instead of using a parametric representation. Each particle has a weight attached, and the distribution represented by the particles is a discrete distribution with probability of each particle proportional to its weight. When an observation is made, each particle's weight is updated by multiplying it by the probability assigned to the observation by that particle.

For our experiments we use randomly generated normal

Cite as: Title (Extended Abstract), Author(s), *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. XXX-XXX.

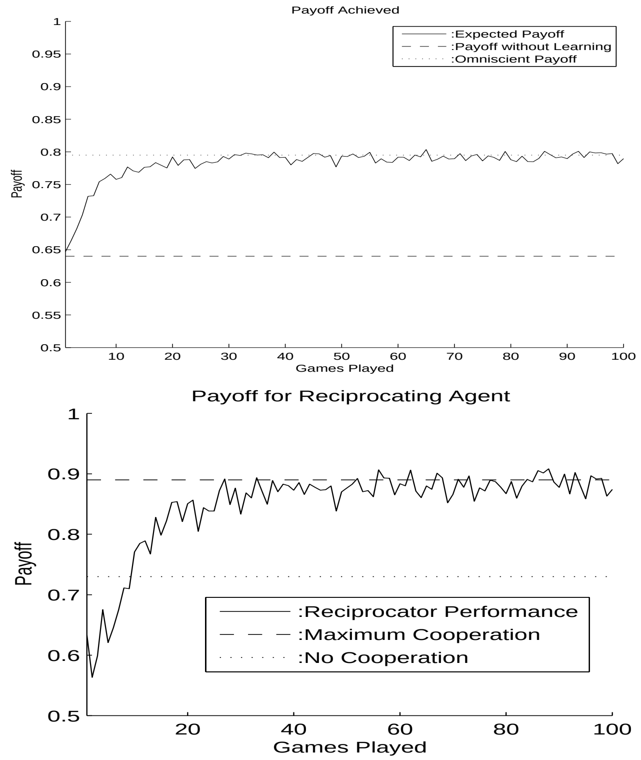

Figure 1: Payoff against a random stationary opponent(top) and in self-play (bottom).

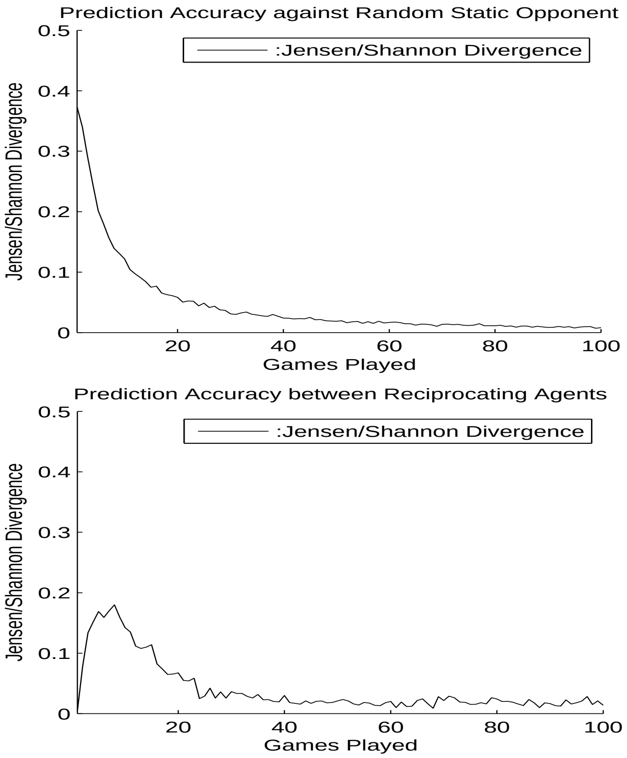

Figure 2: Prediction accuracy against a random stationary opponent (top) and in self-play (bottom).

form games with 16 actions per player, and payoffs uniformly distributed between 0 and 1. We have found empirically that this number of actions provides opportunities for cooperation without making cooperation the only rational choice. We use the Lemke-Howson algorithm to calculate equilibria, and use an initialization parameter passed to the algorithm to distinguish the different methods.

We have measured the *model accuracy*, i.e. the Euclidean distance between the estimates and the true values for attitude and belief of the opponent, and the *prediction accuracy*, i.e. the Jensen-Shannon divergence between the predicted and the actual probability distribution the opponent used to select an action. We have also measured the performance, i.e. the payoff achieved by the agent.

Fig. 1 shows the payoff against a random stationary opponent, where the agent learns to best respond to the opponent's predicted strategy, and in self-play, where each agent reciprocates the opponent's attitude with a bonus of .1. Learning targets are drawn from a Gaussian distribution with 0 mean, results are aggregated over 100 sequences of 100 games. Payoff without learning is what is achieved by an agent which plays according to its prior distribution over the opponent. Omniscient payoff is what would be achieved by an agent aware of the true attitude, belief, and method of the opponent. The payoff can exceed the optimal payoff because of noise in the randomly generated games. As shown in Fig. 2, after 15-20 interactions the agent's predictions are very accurate for a random stationary opponent or in self-play. Those are very small numbers compared to the hundreds of games needed to learn in the simpler case of repeated games [1].

3. CONCLUSIONS

We have presented a method for an agent to learn to cooperate when playing a sequence of different normal form games with the same opponent. We have shown that achieving cooperation is beneficial to both agents and that learning how to respond to the opponent is possible. The results presented are against a random stationary opponent and in self-play, but we have tested the algorithm in many other situations and found that it is fairly robust and effective. Next we will explore two related questions. First, we want to extend our learning approach to handle agents which do not play Nash equilibria. Second, we want to study how an agent can learn about its opponent when playing against other types of non-stationary opponents.

4. REFERENCES

[1] V. Conitzer and T. Sandholm. AWESOME: A general multiagent learning algorithm that converges in self-play and learns a best response against stationary opponents. *Machine Learning*, 67(1–2):23–43, 2007.

[2] 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.

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

[4] E. Fehr and K. M. Schmidt. A theory of fairness, competition and cooperation. *Quarterly Journal of Economics*, 114:817–68, 1999.