Learning to Cooperate in a Continuous Tragedy of the Commons

(Extended Abstract)

Steven de Jong
Dep. of Knowledge Engineering, Maastricht Univ.
Postbus 616, 6200 MD Maastricht, Netherlands
steven.dejong@micc.unimaas.nl

Karl Tuyls
Eindhoven University of Technology
Postbus 513, 5600 MB Eindhoven, Netherlands
k.p.tuyls@tue.nl

Categories and Subject Descriptors

I.2.6 [**Learning**]; I.2.11 [**Distributed Artificial Intelligence**]; J.4 [**Social and Behavioral Sciences**]

General Terms

Algorithms, Design, Human Factors

Keywords

Tragedy of the commons, punishment, learning, complex networks

1. RESEARCH SUMMARY

In previous work, we discussed that social dilemmas are often present in multi-agent systems [3].[1] Social dilemmas are problems in which we can only find a good solution if we consider the benefit of others in addition to our own benefit. Altruistic punishment has been identified as an important mechanism to enforce this consideration. However, as the punishment is altruistic, deciding whether to punish essentially entails a second-order social dilemma. We developed a methodology that allowed individually learning agents to reach satisfactory solutions in a social dilemma with a continuous strategy space, called the Ultimatum Game [2]. We extended this methodology to thousands of agents, using social networks [4]. Moreover, we devoted attention to the tragedy of the commons, a social dilemma typically exemplified by the Public Goods Game (PGG) [1]. In this game, which is played repeatedly, every agent i (out of n) has to decide on an investment $\mu_i \in [0, C]$. The summed investment is multiplied by a factor $1 < r < n$, and equally distributed over all agents. Agent i's individual benefit (or reward) is maximized by $\mu_i = 0$, whereas the group gains the most by collectively playing $\mu_i = C$. Altruistic punishment (i.e., reducing an other agent's reward by an amount e, with a cost $c < e$ to the punisher) allows agents to force others to invest a higher amount, but performing such punishment is clearly not individually rational. In earlier work, we restricted ourselves to a small number of strategies and/or agents in this game [1].

This paper unites and extends all our previous work. We develop a methodology that allows thousands of individually learning agents to reach desirable, cooperative solutions to the PGG (i.e., all agents invest C, the maximum amount allowed, instead of the individually rational investment of 0), even in a continuous strategy space. The only requirement is that there are initially some agents already investing C. The methodology is based on the following four elements.

(1) Continuous-action learning automata (CALA). Agents learn individual behavior by means of CALA [7]. Each agent i keeps track of its current strategy in the PGG, $\mu_i \in [0, C]$, which is adapted after each pairwise interaction with a neighbor j in the network.

(2) Interaction in dynamic social networks. Social network structures have a strong influence on the strategies interacting agents converge to, especially if we allow agents to rewire their neighbor relations [6]. We therefore structure our population of agents in a scale-free structure. This allows us to use thousands of agents. Each newly introduced agent connects to one, two or three existing ones, with a preference for agents that are already densely connected [4]. To model the fact that (relative) cooperators (i.e., those with a high μ_i) may want to prevent interacting with (relative) defectors again, an agent i unwires from its neighbor j after interacting with it, with a probability $p_r = \frac{1}{C}(\mu_i - \mu_j)$. If unwiring happens, i connects to a random neighbor of j, as in [6]. Allowing agent i to select a new neighbor would give i the opportunity to actively exploit this neighbor.

(3) Inequity aversion. Research in behavioral economics identified that the human tendency to perform altruistic punishment may be motivated by *inequity aversion* [5]. We found that inequity aversion may indeed also be used to motivate altruistic punishment in the PGG [1]. Thus, we include altruistic punishment: all agents i consider punishing their peer j after interacting with it, iff $\mu_j < \mu_i$.

(4) Probabilistic punishment. Even if all agents are willing to punish, they should not always do so. The main problem is that, in a continuous strategy space, many learning algorithms (e.g., CALA, see below) optimize by performing a great deal of local search.[2] To solve this problem, we propose the mechanism of *probabilistic punishment*, i.e., the probability that an agent i punishes an agent j should depend on the actual strategies μ_i and μ_j, as well as the resulting rewards r_i and r_j. Punishment is more often performed for higher differences between these rewards. More precisely, we may derive that the punishment probability should be set to $p_i(\epsilon) > \frac{1}{e}(1 - 0.5r)\Delta$, with e denoting the effect of punishment on the reward of the agent being punished, and $\Delta = r_i - r_j$ [1].

[1] *Due to the space constraints of this extended abstract, we omit many relevant references to other authors. We refer the interested reader to our previous work, as given in the list of references. Our papers are available at http://www.cs.unimaas.nl/steven.dejong.*

[2] Imagine agent j with $\mu_j = 2$, playing against agent i with $\mu_i = 8$. Agent j may want to try $\mu_j = 3$. Due to inequity aversion, agent i will punish j in *both* cases. Therefore, the essential idea underlying punishment, i.e., a reversal of the inverse relation between contribution and reward, fails to work: $\mu_j = 2$ gives a higher reward than $\mu_j = 3$, because both strategies are punished.

Cite as: Learning to Cooperate in a Continuous Tragedy of the Commons, (Extended Abstract), Steven de Jong, Karl Tuyls, *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Decker, Sichman, Sierra and Castelfranchi (eds.), May, 10–15, 2009, Budapest, Hungary, pp. 1185–1186

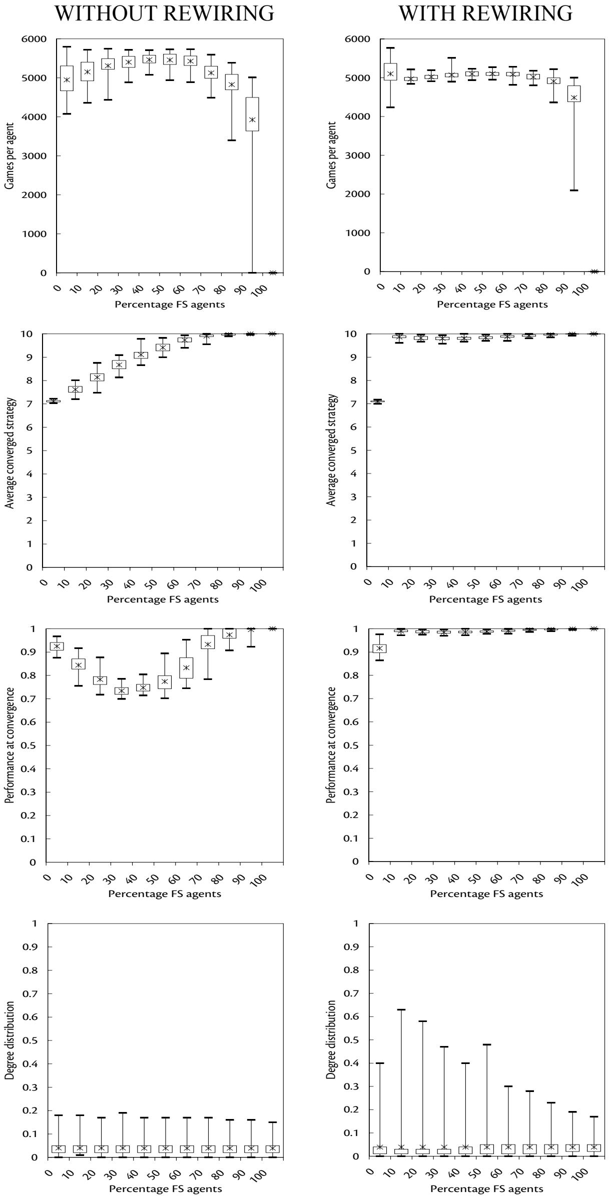

Figure 1: Influence of percentage FS agents

2. EXPERIMENTS AND RESULTS

In an extensive set of experiments (of which this abstract shows a small selection), we study the behavior of a collective of n agents, of which a certain percentage is a *fixed-strategy* (FS) agent, i.e., an agent that always plays the same strategy (i.e., investing C, the desired strategy). The remaining agents are *dynamic strategy* (DS) agents, which learn using CALA after every pairwise interaction. Initially, half of the DS agents present have $\mu_i = 0$, and half have $\mu_i = 0.7C$. The agents are organised based on a (randomly generated) scale-free network. We compare a static network (without rewiring) to a dynamic network (with rewiring). We set the PGG's parameters to $C = 10$, $r = 1.5$, $c = 1$ and $e = 3$, which allows us to set $p_i(\epsilon) = \frac{\epsilon}{C}$. We study collectives of $n \in \{100, 1000, 10000\}$, and for each n, we vary the percentage of FS agents between 0% and 100%, in steps of 10%. All experiments are repeated 100 times. We report results for $n = 100$; results for other n are highly similar.

For an overview of results, see Figure 1. We use the same measurements as in [4]. From top to bottom, we first look at the number of pairwise *games per agent* required to obtain convergence, with an imposed maximum of 6000. A static network requires slightly more games, but still requires drastically less games than reported in related work [6]. Second, we look at the *average converged strategy* of the collective, and mention two observations, i.e., (1) without any FS agents present, the collective converges to investing 7, which is the most cooperative strategy present; (2) with an increasing percentage of FS agents, the collective learns to invest 10, which is remarkably easier for agents that are able to rewire. Third, we examine the *performance at convergence*, which expresses the fraction of neighboring agents in the network that have similar strategies. Results are in line with the observed average strategy. For static networks, we see that the DS agents have difficulties to align themselves with (a low number of) FS agents. This problem is not present in dynamic networks, where the performance is nearly perfect even with only 10% FS agents providing the 'good example'. Fourth and last, we report the *degree distribution* of the network of interaction. For reference, results for static networks are also included. We see that a few agents connect to approximately 20% of the others, while most agents have only a few neighbors. In dynamic networks, we see that the maximum degree increases significantly, especially when there are few FS agents to learn the desired behavior from. This emergence of stronger hubs is very useful. The fact that the topology of the network changes quite a lot, is surprising given our measurements, i.e., on average, rewiring only happens once in approximately 1000 games.

3. CONCLUSION

We present a methodology aimed at allowing a population of learning agents to find and maintain cooperative, desired strategies in a game modelling a tragedy of the commons with a continuous strategy space, i.e., the PGG. We show that our methodology, combining inequity aversion, probabilistic punishment, and (dynamic) social networks, allows individually learning agents to reach the best (most cooperative) strategy initially present. A certain percentage of our agents initially plays in an individually rational (uncooperative) manner. We show that our methodology forces these agents to become more cooperative. Thus, the methology may also be applied in open systems, where we are not able to control the behavior of all agents. The methodology is therefore useful in many problems commonly addressed by multi-agent systems, e.g., resource allocation.

4. REFERENCES

[1] S. de Jong and K. Tuyls. Learning to cooperate in public-goods interactions. 2008. Presented at the EUMAS'08 Workshop, Bath, UK, December 18-19.

[2] S. de Jong, K. Tuyls, and K. Verbeeck. Artificial Agents Learning Human Fairness. In *Proceedings of the international joint conference on Autonomous Agents and Multi-Agent Systems (AAMAS'08)*, pages 863–870, 2008.

[3] S. de Jong, K. Tuyls, and K. Verbeeck. Fairness in multi-agent systems. *Knowledge Engineering Review*, 23(2):153–180, 2008.

[4] S. de Jong, S. Uyttendaele, and K. Tuyls. Learning to Reach Agreement in a Continuous Ultimatum Game. *Journal of Artificial Intelligence Research*, 33:551–574, 2008.

[5] E. Fehr and K. Schmidt. A Theory of Fairness, Competition and Cooperation. *Quart. J. of Economics*, 114:817–868, 1999.

[6] F. C. Santos, J. M. Pacheco, and T. Lenaerts. Cooperation Prevails When Individuals Adjust Their Social Ties. *PLoS Comput. Biol.*, 2(10):1284–1291, 2006.

[7] M. A. L. Thathachar and P. S. Sastry. *Networks of Learning Automata: Techniques for Online Stochastic Optimization*. Kluwer Academic Publishers, Dordrecht, the Netherlands, 2004.