Co-evolution of Agent Strategies in N-player Dilemmas (Extended Abstract)

Raymond Chiong
Department of Computer Science & Software Engineering
The University of Melbourne
Victoria 3010, Australia
rchiong@csse.unimelb.edu.au

Michael Kirley
Department of Computer Science & Software Engineering
The University of Melbourne
Victoria 3010, Australia
mkirley@csse.unimelb.edu.au

ABSTRACT

Understanding how cooperative behaviour emerges within a population of individuals has been the focus of a great deal of research in the multi-agent systems community. In this paper, we examine the effectiveness of two different learning mechanisms – an evolutionary-based technique and a social imitation technique – in promoting and maintaining cooperation in the spatial N-player Iterated Prisoner's Dilemma (NIPD) game. Comprehensive Monte Carlo simulation experiments show that both mechanisms are able to evolve high levels of cooperation in the NIPD despite the diminished impact of direct reciprocation. However, the performance of evolutionary learning is significantly better than social learning, especially for larger population sizes. Our conclusion implies that when designing autonomous agents situated in complex environments, the use of evolutionary-based adaptation mechanisms will help realising efficient collective actions.

Categories and Subject Descriptors

I.2.6 [**Computing Methodologies**]: Artificial Intelligence—*Learning*

General Terms

Complexity Theory, Algorithms, Experimentation

Keywords

N-player dilemmas, evolutionary learning, social imitation

1. INTRODUCTION

Multi-agent learning poses significant challenges across a wide spectrum of problems. In particular, understanding and explaining how globally desired behaviour emerges based on local interactions between adaptive agents is a challenging task. Boosted by Axelrod's computer tournaments [1] in the 1980s, the Prisoner's Dilemma (PD) game has been widely used within the general framework of evolutionary game theory to address this challenge. However, most of the existing studies on the PD have predominantly been concerned with pair-wise interactions, where the action of an agent has a direct impact only on the individual agent it interacts with. In N-player games, the action of each agent will typically affect all members in the social group(s) it is associated with. Little work so far has considered the N-player games in the context of multi-agent learning, and it is our aim to fill this gap.

In this paper, we study the co-evolution of agent strategies in a spatial environment by modelling the strategic interactions between agents using the N-player Iterated Prisoner's Dilemma (NIPD) game based on the formalism of Boyd and Richerson [2]. Previous studies [3, 4] have revealed that spatial structure is beneficial for cooperation, which is consistent with findings in the PD game. Here, we examine the influence of alternative strategy update mechanisms in detail. We show that both evolutionary and imitation learning techniques promote cooperative behaviour, however, the evolutionary technique encourages the emergence of social efficiency to a greater extent.

2. N-PLAYER DILEMMAS

The NIPD is an extension of the PD game in which a group of N players ($N \geq 2$) interact iteratively with one another. In an abstract manner, at each interaction N players make a decision independently based on two actions, either cooperate or defect, without knowing the other players' choice. Each agent uses a strategy to determine what action to play. It is this strategy that is the focus of this particular study.

The utility of an agent is a function of the actions that it plays. A rule exists that rewards a social benefit b, which increases when more players are cooperating. There is, however, always a cost c for the cooperators. Here, $b > c$. Boyd and Richerson [2] formally define the utility values, U, for this scenario as follows:

$$U = \begin{cases} \frac{b \times i}{N} - c & \text{for cooperators,} \\ \frac{b \times i}{N} & \text{for defectors.} \end{cases} \quad (1)$$

with i being the number of cooperators.

3. THE MODEL

In our model, agents are located on a toroidal grid with periodic boundary conditions. Initially, a population of agents with random strategies is created. Based on their strategies, the agents can either cooperate or defect. The utility (fitness) of each agent is determined by summing its payoffs in

Cite as: Co-evolution of Agent Strategies in N-player Dilemmas (Extended Abstract), Raymond Chiong and Michael Kirley, *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, van der Hoek, Kaminka, Lespérance, Luck and Sen (eds.), May, 10–14, 2010, Toronto, Canada, pp. 1591-1592

the game against the neighbours. At the end of each generation, all agents are presented with an opportunity to update their strategies according to the payoffs received. The payoffs are calculated according to Eq. 1.

We have adopted the strategy representation developed by Yao and Darwen [5]. Under this representation, a history of l rounds for an agent can be represented as the combination of the following:

- l bits to represent the agent's l previous actions ('1' = defection and '0' = cooperation)
- $l \times log_2 N$ bits to represent the number of cooperators in the previous l rounds among the agent's social group

In our implementation, we have limited the number of previous actions in memory to 3 (i.e. $l = 3$). An agent's strategy provides a move m, $m \in [1, 0]$, in response to every possible history. Therefore, each strategy should be at least $2^{3+3log_2 N}$ bits in length. The larger the group sizes, the more bits are needed. Due to the fact that there is no memory of previous rounds at the beginning of the game, three additional bits have been added.

When the initial population of agents has been created, each agent will play the game repeatedly for T iterations at each generation. Every agent a_i uses a strategy s_i to decide its move at iteration t, where $t \in [1...T]$. At the end of T iterations, agents have an opportunity to update their strategy – using either evolutionary or social learning.

In the case of evolutionary learning, strategy update is achieved via genetic modification using crossover and mutation. In contrast, in social learning, an agent simply mimics the strategy of a more successful neighbour. An elite preserving method is used in both update mechanisms. The spatial structure ensures co-evolutionary dynamics in the model. In other words, the success or failure of a strategy depends on which other strategies are present in the local neighbourhood.

4. EXPERIMENTS

A systematic Monte Carlo simulation study has been carried out to investigate the relative performance of the evolutionary and imitation learning mechanisms in promoting and maintaining cooperation.

For all scenarios, the reward values $b = 3$ and $c = 1$ were constant. Each simulation ran for 500 generations, with 30 rounds of learning ($T = 30$) constituting each generation. In evolutionary learning, the crossover rate $= 0.7$ and the mutation rate $= 0.05$. In the social (imitation) learning model, the strategy of the most successful neighbour was copied. Three different population sizes were examined: 100, 900 and 2500. For each population size, we varied the group sizes N from 2 to 9. In every single run, the neighbourhood of an agent was fixed. In the cases of $N = 5$ and 9, we used the standard *von Neumann* and *Moore* neighbourhoods. In all other cases, the neighbours were randomly selected from the immediate neighbourhood at the beginning of the game and did not change thereafter. All experiments were repeated for 30 independent runs.

From figures 1 and 2, we see that both evolutionary and social learning techniques are able to promote cooperative behaviour in the NIPD. Due to its ability to adaptively evolve new strategies, evolutionary learning has maintained consistently higher levels of cooperation as compared to the

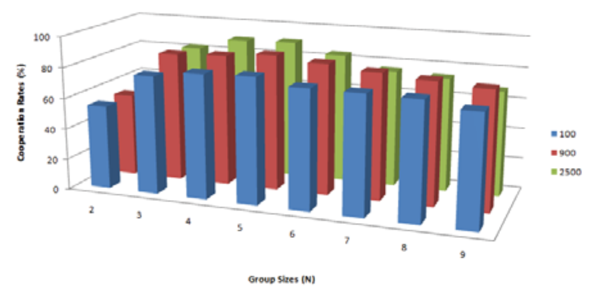

Figure 1: The cooperation rates for evolutionary learning with varying group sizes and population sizes averaged over 30 runs.

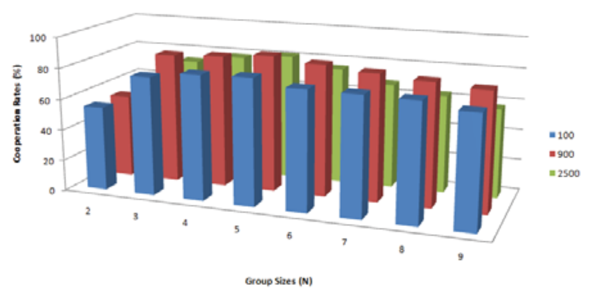

Figure 2: The cooperation rates for social imitation with varying group sizes and population sizes averaged over 30 runs.

social imitation. Statistical tests confirmed that the differences are significant, except when group sizes are large, for example, when $N = 8$ or $N = 9$.

5. CONCLUSION

This paper has contributed to the understanding of multiagent learning in situations when efficient collective actions are required. We have provided an empirical comparison of the relative effectiveness of an evolutionary computation based strategy update mechanism and a social learning (imitation) mechanism. The results suggest that an evolutionary approach is more robust.

6. REFERENCES

[1] R. Axelrod. *The Evolution of Cooperation.* Basic Books, New York, 1984.

[2] R. Boyd and P. J. Richerson. The evolution of reciprocity in sizable groups. *Journal of Theoretical Biology*, 132:337–356, 1988.

[3] R. Chiong, S. Dhakal, and L. Jankovic. Effects of neighbourhood structure on evolution of cooperation in N-player iterated prisoner's dilemma. *LNCS*, 4881:950–959, 2007.

[4] R. Chiong and M. Kirley. Co-evolutionary learning in the N-player iterated prisoner's dilemma with a structured environment. *LNAI*, 5865:32–42, 2009.

[5] X. Yao and P. Darwen. An experimental study of N-person iterated prisoner's dilemma games. *Informatica*, 18(4):435–450, 1994.