Strategic Interactions Among Agents with Bounded Rationality

Pablo Hernandez-Leal, Enrique Munoz de Cote and L. Enrique Sucar
Instituto Nacional de Astrofísica, Óptica y Electrónica
Sta. Maria Tonantzintla, Puebla, México
{pablohl,jemc,esucar}@ccc.inaoep.mx

Abstract

Interactions among agents are complicated since in order to make the best decisions, each agent has to take into account not only the strategy used by other agents but also how those strategies might change in the future (and what causes these changes). The objective of my work will be to develop a framework for learning agent models (opponent or teammate) more accurately and with less interactions, with a special focus on fast learning non-stationary strategies. As preliminary work we have proposed an initial approach for learning non-stationary strategies in repeated games. We use decision trees to learn a model of the agent, and we transform the learned trees into a MDP and solve it to obtain the optimal policy.

1 Introduction

Robust autonomous agents that operate in open environments should be able to interact effectively with an unknown agent, adapting themselves in response to the opponent. In order to deal with such uncertainty, it is common practice to assume that the counterpart's strategy is rational and stationary [Bowling and Veloso, 2002].

The proposed research aims to:

- Consider agents with bounded rationality.
- Focus on learning non-stationary strategies.
- Design agents for heterogeneous populations (more than 2 agents in the environment) of bounded rational agents.
- Develop agents applicable to experiments in repeated games and sequential decision problems.

2 Motivation and Justification

There are several reasons why to work in strategic interactions in multiagent systems, to mention some:

- An essential capability of robust, fully autonomous agents is the ability to interact with other agents. To successfully interact in the real world, agents must be able to reason about their interactions with heterogeneous agents [Stone, 2007].
- The ability to predict the behavior of other agents is crucial in adversarial environments. Learning opponent models can be used to identify weaknesses. Learning models of other agents in cooperative environments is also an important area of research since these models can be used to perform optimal planning of a task.
- Human-agent interaction is another current research area that designs algorithms when humans interact with computer agents. This is an important area given the diversity of possible applications: robotics, economic problems and military applications.

3 Problem Setting

We consider one modeling agent and one or more opponents that face a repeated or stochastic game. At each stage, the agents simultaneously choose an action. The opponents have different strategies for playing the game and they can switch from one to another throughout the interaction. We plan to experiment in common games as the iterated prisoner's dilemma and recent tournaments like the lemonade stand game [Zinkevich, 2009].

4 State of the Art

The current state of the art is spread among different areas, each one of them following a specific path.

- Form the field of game theory there is a set of works on settings like one-shot games, repeated games and stochastic games. The problem is that their approaches focus on finding Nash equilibrium implying rational agents on all situations. Also, related models are not suited for non-stationary opponents.
- Works from behavioral game theory put special focus on bounded rational agents, however they mostly use single-shot games to derive its models and experiments.
- Decision theoretic planning algorithms were designed for sequential decision problems. However, they assume there is a single agent in the environment. Current extensions for MAS systems have a limited applicability due to computational constraints.
- In reinforcement learning the step from a single agent to multiagents is not straightforward. The results are

promising in special cases but not so much in general-sum stochastic games [Shoham *et al.*, 2007]. Current approaches require a larger number of interactions, more over they cannot handle non-stationary opponents.

- Several algorithms have been designed for playing in repeated and stochastic games. However, most of them do not directly address the problem of modeling non-stationary bounded rational agents.

5 Research Questions

In order to motivate this research we propose the following questions:

1. How should we define a model of a bounded rational agent?
2. What should be taken into account when designing a planning algorithm for long-term interactions when a model of the opponent/teammate is at hand?
3. How should we model non-stationary agents?

6 MDP4.5

As preliminary results we propose the MDP4.5 framework for learning switching non-stationary strategies in repeated games using a novel approach based on decision trees and Markov Decision Process (MDPs). A more detailed description of the approach and its results is presented in [Hernandez-Leal *et al.*, 2013]. The general idea is that the opponent models are decision trees, and these can be converted into an MDP to obtain an optimal strategy against that opponent. Changing the initial strategy might have an effect on the opponent. It might switch its strategy to a different one or stay with the same. In order to differentiate these two different outcomes we keep two decision trees which are learnt concurrently. If the *distance* between these trees is greater than a threshold δ, the opponent has changed strategy and the modeling agent must restart the learning phase. Otherwise, it means that the opponent has not switched strategies.

7 Experiments and Results

In order to evaluate the proposed framework we apply it in the iterated prisoner's dilemma (iPD), since it is a largely studied problem. We compare the results of the MDP4.5 agent against the most successful reported strategies in the iPD: Tit for Tat (TFT), Pavlov, Grim and Tit for Two Tats (TFTT). The rewards used for this game are $Cooperate = 3, Defect = 1, Sucker = 0, Temptation = 4$. The games have a size of 250 rounds.

To evaluate if the proposed approach can detect the changes in strategy, the opponent will switch between two strategies; it will start with one from {TFT, Pavlov, Grim, TFTT, Random} and in middle of the game it will change to another strategy from the same set. In Table 1 we summarize the results for different strategies playing against the non-stationary opponents, each result is the average of 80 games. From the results we can observe that MDP4.5 obtained the best average utility. Some conclusions drawn from the experiments are that our agent is capable of detecting changes in the opponent strategy since it obtained high utilities; for opponents using TFT and Pavlov it converges to the C-C actions.

Table 1: Average sum of utilities over 250 rounds against a non-stationary opponent.

Versus oppponent with non-stationary strategies			
Agent	Agent Util.	Opponent Util.	Joint Util.
MDP4.5	**576.5**	537.2	1113.7
TFT	521.0	522.4	1043.4
Pavlov	531.2	573.7	1105.0
Grim	448.7	299.3	748.1
TFTT	540.7	**625.0**	**1165.7**

8 Conclusions

We plan on contributing to the state of the art by developing techniques for fast learn agent strategies, specially those non-stationary ones and design a planning algorithm that uses the learned model to compute a strategy so that interaction with the agent is as profitable as possible. As preliminary work we have proposed an initial approach for learning non-stationary strategies in repeated games. As future work we plan to perform further experiments and generalize the approach for handling more agents in the environment.

9 Acknowledgments

This work was partly supported by the National Council of Science and Technology of Mexico (CONACyT) scholarship grant 234507 to Pablo Hernandez-Leal .

References

[Bowling and Veloso, 2002] Michael Bowling and Manuela Veloso. Multiagent learning using a variable learning rate. *Artificial Intelligence*, 136(2):215–250, 2002.

[Hernandez-Leal *et al.*, 2013] Pablo F Hernandez-Leal, Enrique Munoz de Cote, and L. Enrique Sucar. Modeling Non-Stationary Opponents. *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013)*, May 2013.

[Shoham *et al.*, 2007] Yoav Shoham, R. Powers, and T. Grenager. If multi-agent learning is the answer, what is the question? *Artificial Intelligence*, 171(7):365–377, 2007.

[Stone, 2007] Peter Stone. Learning and multiagent reasoning for autonomous agents. In *The 20th International Joint Conference on Artificial Intelligence*, pages 13–30, 2007.

[Zinkevich, 2009] Martin Zinkevich. Lemonade stand game. http://martin.zinkevich.org/lemonade/, 2009.