An Analysis of Stochastic Game Theory for Multiagent Reinforcement Learning

Michael Bowling Manuela Veloso

October, 2000

CMU-CS-00-165

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

Learning behaviors in a multiagent environment is crucial for developing and adapting multiagent systems. Reinforcement learning techniques have addressed this problem for a single agent acting in a stationary environment, which is modeled as a Markov decision process (MDP). But, multiagent environments are inherently non-stationary since the other agents are free to change their behavior as they also learn and adapt. Stochastic games, first studied in the game theory community, are a natural extension of MDPs to include multiple agents. In this paper we contribute a comprehensive presentation of the relevant techniques for solving stochastic games from both the game theory community and reinforcement learning communities. We examine the assumptions and limitations of these algorithms, and identify similarities between these algorithms, single agent reinforcement learners, and basic game theory techniques.

This research was sponsored by the United States Air Force under Cooperative Agreements No. F30602-97-2-0250 and No. F30602-98-2-0135. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Defense Advanced Research Projects Agency (DARPA), the Air Force, or the US Government.

Keywords: Multiagent systems, stochastic games, reinforcement learning, game theory

1 Introduction

The problem of an agent *learning to act* in an unknown world is both challenging and interesting. Reinforcement learning has been successful at finding optimal control policies for a single agent operating in a stationary environment, specifically a Markov decision process.

Learning to act in multiagent systems offers additional challenges; see the following surveys [17, 19, 27]. Multiple agents can be employed to solve a single task, or an agent may be required to perform a task in a world containing other agents, either human, robotic, or software ones. In either case, from an agent's perspective the world is not stationary. In particular, the behavior of the other agents may change, as they also learn to better perform their tasks. This type of multiagent nonstationary world creates a difficult problem for learning to act in these environments.

However, this nonstationary scenario can be viewed as a game with multiple players. Game theory has aimed at providing solutions to the problem of selecting optimal actions in multi-player environments. In game theory, there is an underlying assumption that the players have similar adaptation and learning abilities. Therefore the actions of each agent affect the task achievement of the other agents. It seems therefore promising to identify and build upon the relevant results from game theory towards multiagent reinforcement learning.

Stochastic games extend the single agent Markov decision process to include multiple agents whose actions all impact the resulting rewards and next state. They can also be viewed as an extension of game theory's simpler notion of matrix games. Such a view emphasizes the difficulty of finding optimal behavior in stochastic games, since optimal behavior depends on the behavior of the other agents, and vice versa. This model then serves as a bridge combining notions from game theory and reinforcement learning.

A comprehensive examination of the multiagent learning techniques for stochastic games does not exist. In this paper we contribute such an analysis, examining techniques from both game theory and reinforcement learning. The analysis both helps to understand existing algorithms as well as being suggestive of areas for future work.

In section 2 we provide the theoretical framework for stochastic games as extensions of both MDPs and matrix games. Section 3 summarizes algorithms for solving stochastic games from the game theory and reinforcement learning communities. We discuss the assumptions, goals, and limitations of these algorithms. We also taxonomize the algorithms based on their game theoretic and reinforcement learning components. Section 4 presents two final algorithms that are based on a different game theoretic mechanism, which address a limitation of the other algorithms. Section 5 concludes with a brief summary and a discussion of the future work in multiagent reinforcement learning.

2 Theoretical Framework

In this section we setup the framework for stochastic games. We first examine MDPs, which is a single-agent, multiple state framework. We then examine matrix games, which is a multiple-agent, single state framework. Finally we introduce the stochastic game framework which can be seen as the merging of MDPs and matrix games.

2.1 Markov Decision Processes

A *Markov decision process* is a tuple, $(\mathcal{S}, \mathcal{A}, T, R)$, where $\mathcal{S}$ is a set of states, $\mathcal{A}$ is a set of actions, T is a transition function $\mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, 1]$, and R is a reward function $\mathcal{S} \times \mathcal{A} \to \boldsymbol{R}$. The transition function defines a probability distribution over next states as a function of the current state and the agent's action. The reward function defines the reward received when selecting an action from the given state. Solving MDPs consists of finding a policy, $\pi : \mathcal{S} \to \mathcal{A}$, which determines the agent's actions so as to maximize discounted future reward, with discount factor γ.

MDPs are the focus of much of the reinforcement learning work [11, 20]. The crucial result that forms the basis for this work is the existence of a stationary and deterministic policy that is optimal. It is such a policy that is the target for RL algorithms.