Learning in Markov Games with Incomplete Information

Junling Hu
Artificial Intelligence Laboratory
University of Michigan
Ann Arbor, MI 48109-2110, USA
junling@umich.edu
http://ai.eecs.umich.edu/people/junling

*

The Markov game (also called *stochastic game* (Filar & Vrieze 1997)) has been adopted as a theoretical framework for multiagent reinforcement learning (Littman 1994). In a Markov game, there are n agents, each facing a Markov decision process (MDP). All agents' MDPs are correlated through their reward functions and the state transition function. As Markov decision process provides a theoretical framework for single-agent reinforcement learning, Markov games provide such a framework for multiagent reinforcement learning.

In my thesis, I expand the framework of Littman of a 2-player zero-sum Markov game to a 2-player general-sum Markov game. In a zero-sum game, two players' rewards always sum to zero for any situation. That means one agent's gain is always the other agent's loss, thus agents have strictly opposite interests. In a general-sum game, agents' rewards can sum to any number. Agents may have incentive to cooperate if they all receive positive rewards in certain situations. Thus general-sum games include zero-sum games as a special case. The solution concept for general-sum games is Nash equilibrium, which requires every agent's strategy to be the best response to the other agents' strategies. In a Nash equilibrium, no agent can gain by unilateral deviation.

In Markov games with incomplete information, agents cannot observe the payoff functions of other agents. Thus they need to form certain beliefs about other agents by learning during the interactions with other agents. The learning must be online in dynamic systems (Hu & Wellman 1998). I aim to design a online Q-learning algorithm for Markov games, and prove that it converges to a Nash equilibrium.

Before designing an algorithm, I defined the equilibrium concept for Markov games under incomplete information. The definition has the following requirements: (1) Each agent's belief must be consistent with the actual outcome; (2) Each agent's strategy must be the best response to its belief about the other agent. When the system reaches equilibrium, there will be no more changes in agents' strategies and their beliefs. We call such equilibrium *conjectural equilibrium.* The study here follows our previous work on conjectural equilibrium in dynamic multiagent systems (Hu & Wellman 1996). Nash equilibrium is a special conjectural equilibrium in which agents' strategies are the same as in complete information case.

I have designed a multiagent Q-learning algorithm for 2-player Markov games. Each agent's Q-value function include both agents' actions and the state. A learning agent maintains two tables for every state, one for its own Q-values and one for the other agent's Q-values. The agent updates both Q tables for a given state after it observes the state, actions taken by both agents, and the rewards received by agents. The updating rule differs previous Q-learning algorithms in that it uses the one-period Nash equilibrium strategies of both agents to update their Q-value functions. I proved that this Q-learning algorithm converges to a Nash equilibrium of the whole game under certain assumptions of the game structure.

In the future work, I want to investigate the exploration and exploitation tradeoff in Markov games. I also want to reduce the large action space by using functional approximation.

Acknowledgments. Thanks to my advisor Michael Wellman for guidance and support. I also want to thank Craig Boutilier, Jun Zhang and Sridhar Mahadevan for helpful discussions.

References

Filar, J., and Vrieze, K. 1997. *Competitive Markov Decision Process.* Springer-Verlag.

Hu, J., and Wellman, M. P. 1996. Self-fulfilling bias in multiagent learning. In *Proceedings of the Second International Conference on Multiagent Systems.* AAAI Press.

Hu, J., and Wellman, M. P. 1998. Online learning about other agents in a dynamic multiagent system. To appear in the Proceedings of the second International Conference of Autonomous Agent.

Littman, M. L. 1994. Markov games as a framework for multi-agent reinforcement learning. In *ML-94*, 157–163. New Brunswick.