Sparse Cooperative Q-learning

Jelle R. Kok Nikos Vlassis

Informatics Institute, Faculty of Science
University of Amsterdam, The Netherlands

The full version of this paper appeared in the Proceedings of the 21^{st} International Conference on Machine Learning in Banff, Canada, July 2004.

1 Introduction

A multiagent system (MAS) consists of a group of agents that can potentially interact with each other [2]. We are interested in fully cooperative multiagent systems, in which the agents have to learn to select individual decisions that result in jointly optimal decisions for the group.

In principle, a multiagent system can be regarded as one large single agent, in which each joint action is represented as a single action. The optimal Q-values for the joint actions can then be learned using standard single-agent Q-learning. We will refer to this method as *MDP learners*. At the other extreme, we have the *independent learners* (IL) approach in which the agents ignore the actions and rewards of the other agents, and learn their strategies independently. However, the standard convergence proof for Q-learning does not hold in this case, since the transition model depends on the unknown policy of the other learning agents.

On the other hand, in many problems agents only have to coordinate with a subset of the agents when in a certain state (e.g., two cleaning robots cleaning the same room). In this paper we describe a multiagent Q-learning technique, *Sparse Cooperative Q-learning*, that allows a group of agents to learn how to jointly solve a task given the global coordination requirements of the system.

2 Sparse Cooperative Q-Learning

In our paper, we first examine a compact representation of the state-action space in which the agents learn Q-values based on full joint actions in a predefined set of states. In all other (uncoordinated) states, the agents learn based on their individual action. Then we generalize this approach using a context-specific coordination graph (CG) [1]. In a CG each node represents an agent, while an edge defines a dependency between two agents. The global coordination problem is now decomposed into a number of local problems that involve fewer agents.

In a CG, value rules can be used to specify the dependencies between the agents. These rules define a (local) payoff for a subset of all state and action variables. In our method, the global Q-value for a state equals the sum of the payoffs of all applicable value rules. After every state transition, the payoff of every applicable

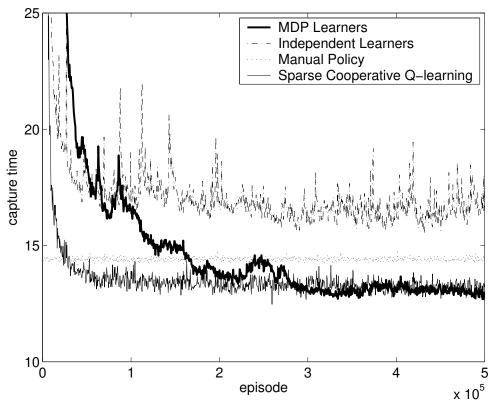

Figure 1: Capture times during the first 500,000 episodes (averaged over 10 runs).

rule is updated based on a Q-learning rule that adds the contribution of all involved agents. Effectively, each agent learns to coordinate only with its neighbors in a dynamically changing CG. This allows for a sparse representation of the joint state-action space of the agents, resulting in large computational savings.

3 Results

We demonstrate the proposed technique on the 'predator-prey' domain in which two predators have to coordinate to capture a single prey in a 10×10 world.

As is seen in Fig. 1, both the IL approach and our proposed method learn quickly in the beginning with respect to the MDP learners since learning is based on fewer state-action pairs. However, the IL approach does not converge to a single policy since the agents do not model the action of the other agent in the coordinated states. These dependencies are explicitly taken into account for the other two methods. For the MDP learners, they are modeled in every state which results in a slowly decreasing learning curve. For the context-specific approach they are considered only for the coordinated states, resulting in a quicker decreasing learning curve with comparable performance to the optimal policy. Our method thus achieves a good trade-off between speed and solution quality.

References

[1] C. Guestrin, S. Venkataraman, and D. Koller. Context-specific multiagent coordination and planning with factored MDPs. In *Proc. 8th Nation. Conf. on Artificial Intelligence*, Edmonton, Canada, July 2002.

[2] N. Vlassis. A concise introduction to multiagent systems and distributed AI. Informatics Institute, University of Amsterdam, September 2003. http://www.science.uva.nl/~vlassis/cimasdai.