Communication in Multi-agent Markov Decision Processes

Ping Xuan, Victor Lesser, and Shlomo Zilberstein
Department of Computer Science
University of Massachusetts at Amherst
Amherst, MA 01003
{pxuan,lesser,shlomo}@cs.umass.edu

Abstract

In this paper, we formulate agent's decision process under the framework of Markov decision processes, and in particular, the multi-agent extension to Markov decision process that includes agent communication decisions. We model communication as the way for each agent to obtain local state information in other agents, by paying a certain communication cost. Thus, agents have to decide not only which local action to perform, but also whether it is worthwhile to perform a communication action before deciding the local action. We believe that this would provide a foundation for formal study of coordination activities and may lead to some insights to the design of agent coordination policies, and heuristic approaches in particular. An example problem is studied under this framework and its implications to coordination are discussed.

1. Introduction

In a multi-agent system, each agent normally only sees a partial view of the whole system. This implies that an agent only observes part of the global system state. Although agents do have the ability to communicate with each other, it is usually unrealistic for the agents to communicate their local state information to all agents at all times, because communication actions are usually associated with a certain cost. Yet, communication is crucial for the agents to coordinate properly. Therefore, the optimal policy for each agent must balance the amount of communication such that the information is sufficient for proper coordination but the cost for communication does not outweigh the expected gain.

We propose a decision-theoretic framework to model a multi-agent system. Our focus is on fully cooperative systems, where all agents share the same goal of maximizing the total expected reward. This is different from the self-interested agents where each agent maximizes its own (local) utility. Specifically, since agents are distributed, and autonomous, we use a local Markov process to describe each agent's state space and actions space. To reflect the cooperative nature of the system, a global reward function is used to describe the relationship and dependency of the individual agent's states. In our model, we assume that each agent knows its current state, i.e., the agent's local state is immediately (fully) observable. An agent has a set of local actions to choose from, and associated with each action is a probabilistic distribution of resulting states. The problem here is to find the set of local policies (one policy for each agent) that produces maximum expected reward. This defines a multi-agent decision process that can be described as a decentralized Markov decision process, which is recently shown to be in the complexity class of NEXP-complete [2].

One limitation of the decentralized MDP is that it does not address agent communication. In a multi-agent system setting, an agent cannot observe directly the local state of other agents, instead, an agent has to use communication in order to share those information. Clearly, communication reduces uncertainty. However, the communication incurs a cost, and the agent needs to decide if it is worthwhile to perform the communication. Thus, for our multi-agent decision process, an agent's policy has to include agent communication decisions. Obviously, the complexity of the problem increases with communication, and it is important to use approximation methods and try to find sub-optimal solutions.

This work introduces a new decision-theoretic framework for multi-agent systems. Previous work, in particular, the multi-agent Markov decision process (MMDP) framework proposed by Boutilier [3], does not have the notion of local states, instead, it assumes that all agents knows the global state all the time. The benefit of this assumption is that the system can be modeled into a standard MDP (or POMDP), but it does not reflect the multi-agent nature of the system. In contrast, our multi-agent decision process emphasizes the decentralized nature of the system. Our work is also an generalization of theoretic works on decen-

tralized control of finite state Markov processes [1, 4, 5].

2. Model Description

Here we give a brief summary of our formal model, which models a cooperative multi-agent system with 2 agents. Systems with 3 or more agents can be easily extended. We consider discrete, finite-horizon problems at the moment.

- Agent x's local Markov process (note it is not an MDP) is described as $M^x = (S^x, A^x, p^x(s_j^x | s_i^x, a^x))$, with S^x as the local states, A_x the set of local actions, and p^x (may be time-dependent) is the local state transition probability. M^y is similarly defined for agent y's local process.
- Global reward function $r_t(s_i^x, s_j^y, a_k^x, a_l^y)$ defines the reward the system receives when the global state is (s_i^x, s_j^y) and the joint action is (a_k^x, a_l^y).
- Let m_t^x (from x to y) and m_t^y (from y to x) be the communication messages between the two agents at time t. The messages contain the agents' local state information (there are several combinations of information transfer, though), and if an agent chooses not to communicate, its message would be $null$.
- The cost of communication is specified via functions $c_t^x(s^x, m^x)$ and $c_t^y(s^y, m^y)$.

Agent x (or y)'s policy π_x (or π_y) is a mapping that reflects both agent's communication decision and action decision at each step of the problem-solving. Note that the policy is a local policy and in general it would be history-dependent.

The decision problem for the system is to find the optimal policy tuple $\pi = \langle \pi_x, \pi_y \rangle$ that maximizes the total expected reward minus communication costs.

As mentioned before, solving it exactly is computational infeasible, so we need to introduce approximation methods, including simplifying the problem, reducing the size of history, and using heuristic approaches. Our first results indicate that heuristic solutions exist and are often easy to compute, and they can indeed give us a lot of insight for agent coordination, and thereby help the design of good policies.

3. Main Results

As an example, we first study a problem of two robots trying to meet each other in a 4×4 grid world. The goal is to meet each other as early as possible. Each robot's movement is not reliable, as they can get stuck or wander off the intended direction (with a certain probability), and each communication (which they reveal each other's current positions) has a fixed cost. We specified this problem under our multi-agent decision process and studied two heuristic policies, while varying the system parameters such as the communication cost, robot's reliability, time-criticalness of reward functions, and deadline constraints. One heuristic, named NN, for its connection with the proverb "no news is good news", sets up individual subgoals and change them (and communicate) only when the subgoals (commitments in typical agent coordination language) cannot be kept. The other heuristic, named SC, for "silent commitment", let the agents divide the goal into individual goals and try to archive them without communication or change of commitments. Our evaluation of the heuristics indicates some intuitive results, and gives us insights regarding when a dynamic strategy (allow changes in commitments) performs better or worse than a static one (with fixed commitments).

We believe that the study on the formal foundation of coordination in multi-agent system is a key to the development of multi-agent systems. Our framework will provide both a formal foundation and an evaluation tool for the design of multi-agent coordination strategies. Due to space limitation, please refer to our report [6] for the details of our model, approaches, and experimental results.

Acknowledgement

The authors would like to thank Andy Barto and Dan Bernstein for fruitful discussions of this problem.

References

[1] M. Aicardi, F. Davoli, and R. Minciardi. Decentralized optimal control of markov chains with a common past information set. *IEEE Transactions on Automatic Control*, AC-32:1028–1031, 1987.

[2] D. Bernstein and S. Zilberstein. The complexity of decision-theoretic planning for distributedagents. In *submitted under review for the Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI-2000)*, 2000.

[3] C. Boutilier. Sequential optimality and coordination in multi-agent systems. In *Proceedings of the Sixteenth International Joint Conferences on Artificial Intelligence (IJCAI-99)*, July 1999.

[4] K. Hsu and S. I. Marcus. Decentralized control of finite state markov processes. *IEEE Transactions on Automatic Control*, AC-27:426–431, 1982.

[5] N. R. Sandell, P. Varaiya, M. Athans, and M. Safonov. Survey of decentralized control methods for large scale systems. *IEEE Transactions on Automatic Control*, AC-23:108–128, 1978.

[6] P. Xuan, V. Lesser, and S. Zilberstein. Communication in multi-agent markov decision processes. Technical Report TR-2000-01, Department of Computer Science, University of Massachusetts, 2000.