Using Communication to Reduce Locality in Multi-Robot Learning

Maja J Matarić
Volen Center for Complex Systems
Computer Science Department
Brandeis University
Waltham, MA 02254
maja@cs.brandeis.edu

Abstract

This paper attempts to bridge the fields of machine learning, robotics, and distributed AI. It discusses the use of communication in reducing the undesirable effects of locality in fully distributed multi-agent systems with multiple agents/robots learning in parallel while interacting with each other. Two key problems, hidden state and credit assignment, are addressed by applying local undirected broadcast communication in a dual role: as sensing and as reinforcement. The methodology is demonstrated on two multi-robot learning experiments. The first describes learning a tightly-coupled coordination task with two robots, the second a loosely-coupled task with four robots learning social rules. Communication is used to share sensory data to overcome hidden state and reinforcement to overcome the credit assignment problem between the agents and to bridge the gap between local and global payoff.

Introduction

This paper attempts to bridge the fields of machine learning, robotics, and distributed AI. It describes how simple communication methods can be applied to enable and speed up learning in complex, noisy, situated multi-agent systems. The domains in question do not permit reliance on complex communication protocols but can be made significantly more manageable through the use of minimalist communication schemes. We describe two such schemes and apply them to reduce the undesirable effects of locality in fully distributed multi-agent systems with multiple concurrent learning agents.

Rather than focusing on the details of our experimental implementations, the paper highlights the role of communication in dealing with two key problems: 1) hidden state and 2) credit assignment. The hidden state problem arises because situated agents typically cannot sense all of the relevant information necessary for completing the task and learning to perform it efficiently. The credit assignment problem arises because reinforcement in a distributed system is often provided at a global level, and must somehow be divided over multiple agents whose impact differs and varies over time. Both problems can be addressed by using simple strategies that apply communication in a dual role: as sensing and as reinforcement, in each case through local undirected broadcast. We demonstrate the methodology on two multi-robot learning experiments. The first describes two robots learning a tightly-coupled coordination task (box-pushing), using communication for sharing sensory data, to overcome hidden state, and reinforcement data, to overcome the credit assignment problem between the two agents. The second experiment describes a loosely-coupled task with four robots learning social rules (yielding and sharing information), using communication to share social reinforcement in order to bridge the gap between global and local payoff.

In both cases, the role of communication is to, locally in space and time, increase the scope of impact of a single agent. Thus, it serves to effectively cluster agents for a period of time during which they are tightly interacting. This has the effect of making the system temporarily and locally less distributed, and consequently alleviates the hidden state and credit assignment problems inherent in distributed multi-agent learning.

Communication as Sensing

The lack of accurate and reliable sensors is arguably the most common complaint of researchers in situated agent control and learning. In robotics in particular, sensors have been targeted as one of the limiting factors in the way of progress toward more complex autonomous behavior. Most of the commonly used sensors provide noisy data and are difficult to accurately characterize and model, presenting a major challenge for real-time robot learning (Matarić 1996). Yet, from the multi-agent perspective, the ability to sense and correctly distinguish members of one's group from others, from obstacles and from various features in the the environment is crucial for most tasks. While the inability to make such distinctions does not preclude symbiotic relationships (Axelrod 1984), the lack of sophisticated perceptual discrimination is a critical limitation in multi-robot work. Non-visual sensors such as infra-red, contact sensors and sonars are all of limited use in the social recognition task. Vision typically offers best discrimination capabilities, but much of traditional image analysis involves computational overhead that is prohibitive for a collection of moving, dynamically interacting robots. Some effective low-overhead approaches to robot vision have been implemented on single robots (e.g., Horswill (1993)) but have not yet been scaled up to groups.

The inability to obtain sufficient sensory information to properly discriminate results in perceptual aliasing or the hidden state problem (Whitehead & Ballard 1990). Due to sensory limitations, multiple world states are perceived as the same input state, inducing serious problems for learning in any domain. The problems are critical in multi-agent/robot domains, where the behavior of the system is the result of the interactions between agents, each of whose limitations can thus become amplified.

If viewed as a form of sensing, communication in multi-agent systems can be used to effectively deal with the hidden state problem. Like other sensors, radio receivers and microphones perceive signals (i.e., messages being communicated) and pass those on for further processing. Specific properties of sensors vary greatly, and these differences can be usefully exploited: some information that is very difficult to directly sense and visually discriminate can be easily communicated. We will demonstrate that agents that locally broadcast their state which may otherwise be difficult or impossible to access, have a significant advantage in a learning scenario. However, it is important to keep in mind that communication suffers from similar noise and inaccuracy properties as other sensors; messages may be corrupted, dropped, or received out of order. These features may not play a key role in disembodied multi-agent work, but impact multi-robot learning strongly, as discussed below.

Communication as Reinforcement

A key challenge of distributed multi-agent systems is to achieve group-level coherence. Such coherence is best guaranteed by top-down control but often requires prohibitive computational overhead that scales poorly with increased group size. A central controller must maintain updated information about the entire group, perform optimizations over the global state space, and send commands to the group; the necessary information is typically not available, the computation cannot be completed in real-time, and the communication imposes a significant bottleneck. Completely distributed alternatives in which all agents learn independently and concurrently scale much better, but in turn introduce other difficulties: non-stationary environments and the credit assignment problem.

A group of agents learning in parallel create a non-stationary world: as the agents learn, their behavior changes, resulting in inconsistencies. Furthermore, multi-agent systems also face the credit assignment problem both at the level of the individual and at the level of the group. At the individual level, the interaction with other agents often delays each agent's payoff, aggravating the temporal credit assignment problem. At the group level, local individual behavior must be appropriately associated with global outcomes. We describe the use of communication as reinforcement, a strategy that enables the agents to locally share reward in order to overcome the credit assignment problem at both levels. In effect, *locally sharing reinforcement between agents decreases locality in the learning system.*

The rest of this paper describes the use of communication in two examples of multi-robot learning. The first addresses the use of communication to extend local sensing of each of the robots, impacting both the hidden state and the credit assignment problem. The second addresses the use of communication to share reinforcement, impacting credit assignment at the individual and group levels. In both cases, communication extends individual sensing, action, and reinforcement by binding two or more agents together locally, for a short amount of time, while their behavior is immediately related.

Communication for Shared Sensing

This section describes the use of communication to handle hidden state and credit assignment in an experiment pairing two mobile robots concurrently learning to perform a box-pushing task whose achievement requires tightly-coupled coordination because the box is chosen to be large enough so that neither of the agents can push it to the goal alone. The agents, six-legged mobile robots, are equipped a radio communication mechanism, whiskers that detect contact with the box and an array of five pyroelectric sensors that detect the direction and approximate distance to the goal marked

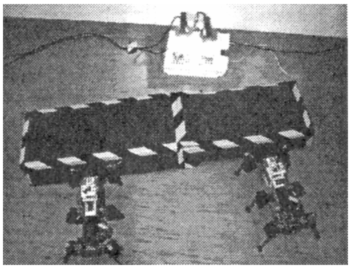

Figure 1: Two Genghis-II six-legged robots used to learn cooperative box-pushing, equipped with whiskers for detecting contact with the box, pyroelectric sensors for detecting the light at the goal, and radios for communication.

with a bright light (Figure 1). The sensors do not provide information about the robots' absolute location or relative location along the box.

The goal of the learning experiment was to have the robots automatically discover a strategy that allows them to coordinate their efforts into a coherent policy capable of delivering the box to the goal regardless of the initial positions of the robots or the box and the robots' inability to directly sense each other or the movement of the box. The experiment was set up within a reinforcement learning (RL) framework, so that each of the robots was learning a mapping between its sensory perceptions and its set of pre-programmed fixed-duration *basis behaviors* (Matarić 1995): *find-box, push-forward, push-left, push-right, stop, send-msg.* The sensory data was based on the whiskers (contact or no contact) and the light sensors (the location of the maximum brightness along the 5-sensor array).

The learning was based on reinforcement computed from the changes in the direction and distance to the goal. Behaviors that moved the robots so that the sensors were facing the light (i.e., the middle or near-middle pyros were most active) generated internal reward, while those orienting the robot away from the light generated internal punishment. Reinforcement was computed after the completion of a behavior, and was used to reinforce the specific, just-terminated percetion-behavior pair. Thus, the system was learning a purely reactive pushing policy. The behavior selection algorithm chose the "best" action 75% of the time and a random action 25% of the time.

Merely servoing to the light without a box is simple. Communication was the key to making this task learnable for the two robots having to do so with the large box and incomplete sensing: they could distinguish direction and distance from the light, but not the orientation of the box or the other agent. Through the use of communication, the robots could combine their sensory data and get information about two points along the box. As a result, the hidden state problem could in this case be solved by having the two agents pool their sensory resources.

Anoter challenge of this task is finding the joint robot actions that result in the appropriate coordinated state-action mapping for the task. The credit assignment problem arises: if a box is getting closer to the goal, whose actions are responsible for the progress? Which of the agents should be rewarded? Is one agent undoing the work of the other? One way of solving this problem is again to use communication: each agent tells the other what action to perform, the two take coordinated actions, observe the outcomes, and share the subsequent reward or punishment. As a side-effect, communication also synchronizes their actions, ensuring coordinated movement.

In our experiments the robots learned both how to move and what to communicate to each other. They were pre-programmed to use the *send-msg* behavior to communicate their sensory information (i.e., they were not learning a language) in order to overcome the hidden state problem. What they learned was how to choose, based on the pooled sensory data, what action to take and what action to communicate to the other robot. Each robot learned a function mapping the combined perceptual states and the best actions for itself and the other agent (i.e., my-action and its-action). The resulting policy table consisted of the following: $S_{combined} \rightarrow A_{my-action} \times A_{its-action}$. In our case, $|S_{combined}| = 25$ and $|A_{my-action} \times A_{its-action}| = 25$. The desired policy was learned by both robots in over 85% of the trials; on the average it was learned in 7.3 minutes (i.e., about 40 contiguous trials).

In order to make the system converge, the robots took turns in controlling the box. An alternative solution would establish a master-slave system where only one of the robots is learning and commanding the actions of the other. Given the intrinsic differences in the sensor and effector characteristics of the two robots, we opted for scheme that did not favor either. The turn-taking, easily accomplished through the use of the communication channel, effectively alleviated the credit assignment problem between the robots. After an action was executed, each robot could compute the reinforcement directly from the updated combined sensory data resulting from the actions. Since the robots shared identical internal reinforcement functions, they received identical reinforcement, modulo sensory and communication noise. However, the strategies learned by the two robots were not identical; they depended on the side of the box each robot happened to be on.

Each learned a "handed" strategy, becoming either a left-pusher or a right-pusher[2].

In this learning task, communication was effective in extending the individual agent's very local and incomplete view of the world at least in part because it comes from the other, near-by agent's sensors and actions involved in solving a shared task. Our next experiment, involving a larger group of robots, demonstrates that this idea of using communication to tie together spatially local agents also has important uses in learning in larger, more loosely-coupled groups.

Communication for Shared Reward

This section describes the use of communication to handle hidden state and credit assignment in an experiment featuring four mobile robots concurrently learning two non-greedy social rules: 1) yielding to each other, and 2) sharing information about the location of pucks. The robots are equipped with infra-red sensors for obstacle and puck detection, bump sensors, position sensors, and radios for communication (Figure 2). The sensors do not provide information about other robots' external state or behavior.

As in the previous experiment, the robots were preprogrammed with simple *basis behaviors* that enabled them to collect pucks (*pickup, drop, home, wander, follow*) and send messages (*send-msg*). The goal of the experiment was to have the robots improve individual collection behavior by learning when to yield to another robot and when to send a message, and conversely, when to proceed and when to follow the received message to the location it communicated. This experiment was also set up within a reinforcement learning (RL) framework, so that each of the robots was learning a mapping between its sensory perceptions (infra-red, bump, position, and communication sensors clustered into a predicate set (*at-home, near-pucks, near-another, too-near-another, have-puck, got-msg*) and its behaviors (*stop, send-msg, proceed* and *follow-msg*). The behavior selection algorithm chose the "best" behavior 60% of the time, a random behavior 10% of the time, and *follow-msg* (inducing the robot to follow the most recently received message) 30% of the time.

Learning social rules in this scenario is difficult, due to delayed reinforcement and the credit assignment problem. In a reinforcement learning paradigm, each agent tries to maximize its own payoff over time. Social behaviors like communicating puck location do not increase the agent's payoff and behaviors like yielding may decrease it by further delaying or even preventing reward. Thus, it is difficult for individual agents to learn such rules without having them centrally imposed or pre-programmed, as advocated in some game-theoretic work (Axelrod 1984).

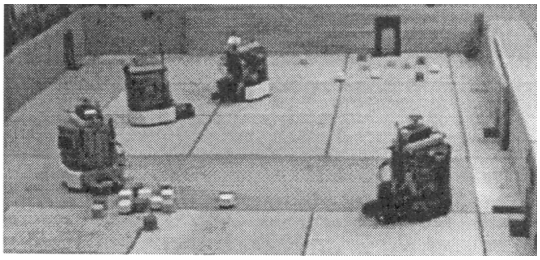

Figure 2: Four IS Robotics R2e robots used to learn to yield (e.g., top two robots) and communicate the location of the pucks (e.g., bottom two robots), equipped with infra-red sensors for obstacle and puck detection, bump sensors, position sensors, and radios for communication.

Our experiments demonstrated that through the use of simple communication between spatially local agents, the described social rules can be learned. The communication channel was used to provide payoff, i.e., generate reinforcement, for observing the behavior of the near-by agent and mimicking it. This way of relating two near-by agents' behavior was an effective means of learning social rules. Observing other agents' behavior is a difficult perceptual task, but it can be easily solved with communication; our robots broadcast, within a limited area (i.e., the perceptual range) the behavior they were performing, as well as the reinforcement they were receiving.

Communicating external state (i.e., behavior) and internal state (i.e., reinforcement) between locally interacting agents has a dual effect. First, this information allows a robot to select what observed behavior to "imitate" based on the vicarious reinforcement, thus decreasing the need for experimentation. The key assumption is that positive reinforcement will be consistent across agents, and this assumption holds because of the spatial and temporal locality of the shared information. When two robots are at nearly the same place at the same time, they are likely to receive the same reinforcement for the same actions. Second, in addition to encouraging social behaviors, communicating and sharing reinforcement also provides payoff for social actions (such as yielding) that would not otherwise benefit the agent directly. When a robot yields to another, the "proceeder" reaches its goal and receives reinforcement directly, then shares it with the

[2] This specialization could be further generalized through the use of a single variable representing the box side.

"yielder", thus reinforcing social behavior. Any nearby robots that observed the yielding behavior and received payoff have the impetus to imitate it, and thus learn it as well. Learning to send messages works in the identical fashion. A robot that finds pucks communicates its location, and the near-by agents that perceive the information move to that location, find the pucks, and receive and share the reward. Thus the mapping between finding pucks and sending a message is reinforced, as is the mapping between receiving a message and following it.

Locally sharing information was sufficient to enable the group to learn the following social behavior: the robots performed individual collection but when close to other robots they yielded and shared information about the location of the pucks. This resulted in better overall performance of the group; we compared the results based on the measured total collection time for a fixed number of pucks. The desired social policy of yielding and sharing information was learned by all robots in over 90% of the trials; on the average it was learned in 20 to 25 minutes. Mutual yielding symmetries were broken by the noise inherent in the robots' sensing and actuation.

In this learning task, communication was again used to extend each agent's very local and incomplete view of the world. Unlike the first experiment, where sensory information was shared, and reinforcement was individual, in this case sensory information was individual but reinforcement was shared. The described method is another demonstration of addressing the multi-agent credit assignment problem though the use of communication. Because the communication method is local, it scales well with increased group size. Broadcasting behavior and reinforcement were simple to implement and proved to be robust in the presence of noise given a minimal amount of radio filtering. Corrupted messages were largely detected and ignored. Lost messages were rare since each was broadcast a fixed number of times. If any were lost, however, the omission had the effect of slowing down the learning, as does any sensor noise. Given the confined work area, the robots encountered each other with high frequency, thus being exposed to social conditions that facilitated the learning.

Related Work

The described work applies concepts from machine learning (ML) and distributed AI (DAI) to robotics. The problems of hidden state and credit assignment have been dealt with extensively in the ML literature (McCallum 1996, Sutton 1992). Whitehead (1991) analyzed cooperative RL mechanisms and Littman (1994) used the Markov games framework for simulated RL soccer-playing agents. Tan (1993) applied RL to a simulated hunter-prey scenario in which the agents used communication to exchange instantaneous information, much like in our experiments, but also used episodic and learned knowledge.

Communication and multi-agent cooperation have been addressed in abstract and applied multi-agent work (Yanco & Stein 1993, Dudek, Jenkin, Milios & Wilkes 1993, Altenburg & Pavicic 1993). In DAI learning work, several related contributions have been made in the area of studying multiple simulated reinforcement learners (see collection by Weiss & Sen (1996)), but have not yet been applied to robotics, where multi-robot learning was identified as an area to be addressed (Mahadevan & Kaelbling 1996).

In robot learning, Mahadevan & Connell (1991) focused on state generalization and controller modularity applied to a sonar-based robot learning to push a box in an experiment complementary our our first one, which addresses the same problem with a similar approach, but introduces communication to extend it to the multi-robot domain. Work on cooperative robot pushing has been addressed in control; Donald, Jennings & Rus (1993) demonstrated a (non-learning) two-robot sofa-moving system using more powerful sensors to eliminate the need for communication. In contrast, our work uses communication to compensate for the robots' sensory limitations and to facilitate learning. Our previous work (Matarić 1994) treated learning in a multi-robot system, and focused on using internal progress estimators to address the temporal credit assignment problem on each agent individually, without the use of communication. In contrast, Parker (1993) addressed the issue of coordinating a loosely-coupled group of foraging robots, and described how communication can be used to combine the abilities of heterogeneous agents in order to complete a cooperative task more efficiently.

Conclusions

The goal of this paper was to draw from ideas in machine learning, robotics, and DAI in order to deal with two problems of learning in distributed multi-agent systems: hidden state, because agents cannot sense all of the relevant information, and credit assignment, because it is difficult to distribute global reinforcement over multiple agents. We described how simple communication based on local broadcast can be used to address both problems and illustrated the ideas on two multi-robot demonstrations. In the first, two robots learn to cooperatively push a box by communicating sensory readings to each other; in the second, four robots learn social rules for yielding and sending information by using communication to share reinforcement.

The effect of the described limited broadcast communication schemes was to temporarily bind together two or more spatially local agents. The idea utilizes simple communication to temporarily decrease the distributedness of the system just between the robots that are near-by at a given point in time. This simple idea scales well and is particularly suitable for noisy, dynamic domains in which agents interact based on incomplete sensory information. Such environments abound in robotics, where communication may compensate for the limitations of more direct sensory modalities, thus enabling learning in multi-robot systems.

Acknowledgments The described work is supported by the Office of Naval Research Grant N00014-95-1-0759 and the NSF Infrastructure Grant CDA-9512448. The box-pushing learning work was performed at the Swedish Institute for Computer Science, in collaboration with Kristian Simsarian and Martin Nilsson.

References

Altenburg, K. & Pavicic, M. (1993), Initial Results in the Use of Inter-Robot Communication for a Multiple, Mobile Robotic System, *in* 'Proceedings, IJCAI-93 Workshop on Dynamically Interacting Robots', Chambery, France, pp. 96–100.

Axelrod, R. (1984), *The Evolution of Cooperation*, Basic Books, NY.

Donald, B. R., Jennings, J. & Rus, D. (1993), Information Invariants for Cooperating Autonomous Mobile Robots, *in* 'Proc. International Symposium on Robotics Research', Hidden Valley, PA.

Dudek, G., Jenkin, M., Milios, E. & Wilkes, D. (1993), On the utility of multi-agent autonomous robot systems, *in* 'Proceedings, IJCAI-93 Workshop on Dynamically Interacting Robots', Chambery, France, pp. 101–108.

Horswill, I. D. (1993), Specialization of Perceptual Processes, PhD thesis, MIT.

Littman, M. L. (1994), Markov games as a framework for multi-agent reinforcement learning, *in* W. W. Cohen & H. Hirsh, eds, 'Proceedings of the Eleventh International Conference on Machine Learning (ML-94)', Morgan Kauffman Publishers, Inc., New Brunswick, NJ, pp. 157–163.

Mahadevan, S. & Connell, J. (1991), Automatic Programming of Behavior-based Robots using Reinforcement Learning, *in* 'Proceedings, AAAI-91', Pittsburgh, PA, pp. 8–14.

Mahadevan, S. & Kaelbling, L. P. (1996), 'The National Science Foundation Workshop on Reinforcement Learning', *AI Magazine* **17**(4), 89–97.

Matarić, M. J. (1994), Reward Functions for Accelerated Learning, *in* W. W. Cohen & H. Hirsh, eds, 'Proceedings of the Eleventh International Conference on Machine Learning (ML-94)', Morgan Kauffman Publishers, Inc., New Brunswick, NJ, pp. 181–189.

Matarić, M. J. (1995), 'Designing and Understanding Adaptive Group Behavior', *Adaptive Behavior* **4**(1), 50–81.

Matarić, M. J. (1996), 'Reinforcement Learning in the Multi-Robot Domain', *Autonomous Robots* **4**(1), 73–83.

McCallum, A. R. (1996), Learning to use selective attention and short-term memory in sequential tasks, *in* P. Maes, M. Matarić, J.-A. Meyer, J. Pollack & S. Wilson, eds, 'From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior', The MIT Press, pp. 315–324.

Parker, L. E. (1993), Learning in Cooperative Robot Teams, *in* 'Proceedings, IJCAI-93 Workshop on Dynamically Interacting Robots', Chambery, France, pp. 12–23.

Sutton, R. S. (1992), Machine Learning, Special Issue on Reinforcement Learning, Kluwer Academic Publishers, Boston.

Tan, M. (1993), Multi-Agent Reinforcement Learning: Independent vs. Cooperative Agents, *in* 'Proceedings, Tenth International Conference on Machine Learning', Amherst, MA, pp. 330–337.

Weiss, G. & Sen, S. (1996), Adaptation and Learning in Multi–Agent Systems, Lecture Notes in Artificial Intelligence, Vol. 1042, Springer-Verlag.

Whitehead, S. D. (1991), A complexity analysis of cooperative mechanisms in reinforcement learning, *in* 'Proceedings, AAAI-91', Pittsburgh, PA.

Whitehead, S. D. & Ballard, D. H. (1990), Active Perception and Reinforcement Learning, *in* 'Proceedings, Seventh International Conference on Machine Learning', Austin, Texas.

Yanco, H. & Stein, L. A. (1993), An Adaptive Communication Protocol for Cooperating Mobile Robots, *in* J.-A. Meyer, H. Roitblat & S. Wilson, eds, 'From Animals to Animats: International Conference on Simulation of Adaptive Behavior', MIT Press, pp. 478–485.