Talking About the World: Cooperative Robots that Learn to Communicate

Holly Yanco *
Artificial Intelligence Laboratory
Massachusetts Institute of Technology
545 Technology Square, Room 741
Cambridge, MA 02139
(617)253-7884
holly@ai.mit.edu

Abstract

Models of the world can take many shapes. In this paper, we will discuss how groups of autonomous robots learn languages that can be used as a means for modeling the environment.

The robots have already learned simple languages for communication of task instructions. These languages are adaptable under changing situations; i.e. once the robots learn a language, they are able to learn new concepts and update old concepts. In this prior work, reinforcement learning using a human instructor provides the motivation for communication.

In current work, the world will be the motivation for learning languages. Since the languages are grounded in the world, they can be used to talk about the world; in effect, the language is the means the robots use to model the world. This paper will explore the issues of learning to communicate solely through environment motivation. Additionally, we will discuss the possible uses of these languages for interacting with the world.

1 Introduction

In a *world model*, aspects of the world observed by robots are abstracted into concepts that they use to interact with the world. In this work, the robots move about the world and make attempts to talk to other robots. Initially, the robot language is not specified; the robots must agree upon a signalling protocol to be used as their language. As they begin to agree upon a language, this language will reflect their views of the world. In the language our small, vision-less robots develop, a wall and the back of a bookcase might be represented by the same word. The robots will abstract concepts from the world in a different way than humans with highly developed vision systems are able to. By examining the language the robots develop, we will be able to learn more about how the robots see the world and how languages for robots differ from our own human languages.

We have chosen to use situated robots rather than simulations since the world is its own best model. Our belief is that if we were to do the work in simulation, we would be building in biases that would influence the language development. Since the world will provide the motivation for the language development. Therefore, we should use the real world rather than create a "blocks world" of some kind.

We have developed a team of autonomous robots that learn their own novel languages for inter-robot communication. The languages they develop are better suited to the robots' needs since programmers can not anticipate every possible situation that the robots may encounter in the world. Additionally, programmers will most likely create signals for actions and world objects in a way that seems natural to humans; these provided signals may not be natural to either the robots or to the tasks they are to perform. If the robots have the ability to develop and adapt their own language, they will be able to handle novel situations, deal with changing environments and perhaps even perform their tasks more efficiently. (The languages are adaptable by definition – once the robots have the ability to learn, they can learn new concepts and update existing ones.)

In our original work, we explicitly gave one of the robots the task information; this robot acted as a "leader." The leader learned to perform the activity and communicated the task information to the other members of the team. Initially, the signals used by the leader were selected randomly. The other robots needed to learn the proper responses to the leader's signals. We reinforced the team's behavior based on its performance as a unit using "task-based reinforcement;" i.e., we only gave positive feedback when all members of the team were acting appropriately. Under this method, the robots successfully developed languages for task communication and were able to adapt them when we changed

*This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-85-K-0124. The author is supported in part by a scholarship from Digital Equipment Corporation. Support for this research was also provided by the Gordon S. Brown Fund of the Department of Electrical Engineering and Computer Science at MIT.

the commands' meanings. This work is discussed in [Yanco and Stein, 1993]; it is also summarized below as background for the current work.

Our initial work was inspired by Shewchuk's Ph.D. thesis [Shewchuk, 1991]. It addresses the design of appropriate reinforcement learning algorithms to learn languages for internal representation as well as for communication. He implemented a simple simulation of a language learning task similar to the basic experiment we describe below (two robots, two language elements) as a part of his symbolic test suite for reinforcement learning algorithms. Work on the development of communication between groups of autonomous agents has also been done by [MacLennan, 1990] and [Werner and Dyer, 1990]. Their research addresses the problem of language learning with genetic algorithms. Language evolves over many generations of the community. Within an individual agent, however, language is fixed over its lifetime. In all of these cases, implementation is limited to simulation; only the work of Shewchuk addresses the problem of task-based reinforcement as described below.

In the current development of the work, the human is removed from the learning process. The robots have much more autonomy in their environment; they need to rely on the world and other robots to provide task information and reinforcement. In the context of this world motivation, we are exploring the development of context dependent languages and compositional languages.

2 The robots

The robots used in this research are Sensor Robots designed by Fred Martin at the Media Laboratory at the Massachusetts Institute of Technology [Martin and Sargent, 1991]. Each robot is approximately $9''l \times 6''w \times 4''h$, with a single circuit board containing most of the computational and sensory resources of the robot.

The robots communicate between themselves using a pair of radio transmitter and receiver boards similar to those used in garage door openers. These boards can send four bits of data at a time; for language sizes greater than 16 words, the robots need to send sequences of four bit packets.

The robots contain four (front and rear, left and right) bump sensors, left and right shaft encoders, an inclination sensor, photosensitive cells, a microphone, and infra-red detectors and emitters. Additionally, each robot has a speaker and a 16-character LCD, both used primarily for debugging and monitoring of the robot's activity.

While the robots have many sensors, they are limited in their sensing capabilities. They are not able to differentiate between many obstacles in their world as they do not have vision. However, we believe that these limitations can be used to our advantage since we'll have a chance to see the representations of the world that the robots create and see how they differ from our models of the world.

The initial state, s_0, consists of the integer variables x_0, n_0, x_1, and n_1, each initialized to 0.

```
u(s,a,r) = if a = 0 then begin
                x0 := x0 + r
                n0 := n0 + 1
           end else begin
                x1 := x1 + r
                n1 := n1 + 1

           end

e(s) =     if ub(x0,n0) > ub(x1,n1) then
                return 0
           else
                return 1
```

where

$$ub(x,n) = \frac{\frac{x}{n} + \frac{z^2_{\alpha/2}}{2n} + \frac{z_{\alpha/2}}{\sqrt{n}}\sqrt{(\frac{x}{n})(1-\frac{x}{n}) + \frac{z^2_{\alpha/2}}{4n}}}{1 + \frac{z^2_{\alpha/2}}{2n}}$$

and $z_{\alpha/2} > 0$.

Figure 1: Kaelbling's interval estimation algorithm [Kaelbling, 1990, Figure 21].

3 Description of initial work

The cooperative task of coordinated movement was selected for the initial experiments. We have implemented this task with troupes with two and three members and with a variable number of vocabulary elements on robots and in simulation. The simulator was used to gather data for the three agent experiments and for the larger vocabulary experiments with two agents. Because we were particularly interested in the development of language, we assumed that the followers do not have access to the task specification (i.e. the environmental cues) and must rely completely on the communication signals emitted by the leader. In future experiments, we expect to allow the follower robot(s) to use some environmental input to modulate the communication signals from the troupe leader as described in the later section on context dependent languages.

Since this is a cooperative task, successful performance depends on the actions of the troupe as a whole. Environmental reinforcement is therefore positive only if all agents perform the appropriate actions; this is called *task-based reinforcement.* Since the followers cannot correctly interpret the environmental cues, this performance can be achieved reliably only when the leader and follower robots mutually agree on the development and interpretation of a private communication protocol.

Thus, the learning tasks are as follows:

- For the leader robot, the interpretation of the environmentally supplied signal, the execution of an appropriate action, and the transmission of an appropriate signal to the follower robot.

- For the follower robots, the execution of an appropriate action based on the signal received from the leader robot.

The "appropriateness" of an action is determined by the environmentally supplied signal. The "appropriateness" of the leader robot's signal, however, is constrained not by the environment but by the leader and follower robot's adapted internal state. That is, the signal is appropriate if and only if the follower robot takes the (environmentally constrained) appropriate action when that signal is received.

In our experiments, the environment is represented by a human "instructor" who issues one of a number of signals to indicate the desired action. Currently, the number of signals is also the size of the language. The leader robot performs an action and also signals the follower robot. Upon receipt of the leader's signal, the follower robot selects and performs an action. If both robots have performed correctly, positive reinforcement (+) is issued. Likewise, if either robot performs incorrectly, negative reinforcement (−) is issued. Based on this environmental feedback, the robots learn to select appropriate actions and communication signals.

Both the action selection and the signal selection are learned using standard reinforcement learning techniques. (See, e.g., [Kaelbling, 1990] or [Sutton, 1992] for overviews of reinforcement learning.) The particular algorithm that we use is adapted from Kaelbling's interval estimation method [1990]. Interval estimation is a relatively simple form of reinforcement: A table of inputs × actions is maintained. Each time an input is received, the expected "best" action is taken and the counter for that input/action pair is incremented. If positive reinforcement is received, a second counter for that input/action pair is also incremented. The "best" action given some input is selected by an optimization function. If no one particular action is the "best", an action is selected randomly. The algorithm for interval estimation is given in figure 1.

In our initial experiments, we allow each of the robots two possible actions. At each iteration, each robot chooses either *go straight* or *spin*. Further, the communication protocol contains only two vocabulary elements—high and low—so that the learning problem remains tractable. The leader robot must thus learn to select one of four possible action/communication pairs; the follower robot must learn to associate each of the vocabulary items with one of its two possible actions. Convergence on the robots is easily verified by testing each environmental input; if all behaviors are as expected, the protocol will not change further without environmental adaptation.

4 Results of initial work

4.1 Developing a Shared Language

The robots are able to learn both synchronous action—both performing the same action in the same interval—and divergent action—e.g., leader *spins*, follower *goes straight*. Convergence times typically range from five to twenty iterations with a team two robots. A sample run of the experiment is given in table 1. In this run, the appropriate actions are for both robots to *spin* on input ◯◯ and for both robots to *go straight* on input ↑↑. The robots converge on a mutually agreeable language—a low signal means that the follower should *spin*, while a high signal means to *go straight*—after thirteen iterations.

We have also run the experiment using team of three robots. Using three robots, a two element language typically converges after an average of 27 iterations; the range is between 10 and 80 iterations. Larger language sizes have also been tested; as the language size increases, the learning time increases exponentially. Results are discussed in [Yanco and Stein, 1993].

4.2 Adaptability of language

Once the robots converge on a particular dialect, they continue to receive positive reinforcement as long as the environmental constraints do not change. If circumstances change, however, the robots may find that their previously successful actions no longer earn them positive feedback. For example, after the run in figure 1, we might change the "appropriateness" of the robots' actions by giving positive reinforcement to *leader spin, follower go straight* on ↑↑. Under such circumstances, the robots can adapt their behavior—and, when necessary, their communication protocol—to the changing environment. Convergence times for unlearning portions of the old task and relearning the newly appropriate behavior range from roughly comparable to those for the initial learning task to roughly double the time, depending on the difficulty of the new task, the differences between the old and the new, and how firmly the previous behavior is entrenched.

5 Theory of current work

For our current work, we want the robots to determine their actions solely on the basis of their interactions with the real world as opposed to relying on a human instructor to provide task information and reinforcement. In this scenario, the language requirements are completely driven by the environment. Our hope is that the environment will be able to replace the human in the learning process.

5.1 Context dependent language

The next step in the robots' language development is the creation of a context dependent language. In a context dependent language, words can have different meanings depending on the situation in which they are used.

Context is provided by sensor readings on the robots. Different areas in the world have various characteristics that the robots are able to detect with their sensors. The robots have sensors that detect light levels, the presence of objects, heat levels and infrared signals. The sensor values are read into an array that the robots are able to access.

In this scenario, the robots need to map signals (or words) to actions (or meanings for the words). To motivate the robots to create multiple mappings from a word

	Appropriate action	Leader's action	Leader's signal	Follower's action	Reinforcement
1.	↑↑	*spin*	low	*spin*	−
2.	◯◯	*spin*	low	*straight*	−
3.	↑↑	*straight*	high	*spin*	−
4.	◯◯	*straight*	high	*straight*	−
5.	◯◯	*spin*	low	*spin*	+
6.	↑↑	*straight*	high	*spin*	−
7.	◯◯	*spin*	low	*spin*	+
8.	◯◯	*spin*	low	*spin*	+
9.	◯◯	*spin*	low	*spin*	+
10.	↑↑	*spin*	low	*spin*	−
11.	↑↑	*straight*	high	*straight*	+
12.	↑↑	*straight*	high	*straight*	+
13.	◯◯	*spin*	low	*spin*	+

Table 1: A sample run. The desired behavior is *both spin* on input ◯◯, *both go straight* on input ↑↑. After thirteen iterations, convergence is reached.

to meanings, the language size is restricted; i.e. there are not enough signals for the robots to create a one-to-one mapping between words and meanings.

The robots can learn a command such as "do X," where the meaning of X is directly mapped to the sensor readings. This is analogous to a world where animals eat where there is food, gather objects where they are present and avoid preditors that they detect. Directions for movement could also have context dependent meanings. For example, when in the bright area, the robot moves to the dark area; after being in the dark area, the robot should move to the warm area.

Another area of context dependency that will be explored as the work progresses will be the ability of words to take context from the sentences in which they are embedded. This should tie in to our development of compositional language.

5.2 Compositional language

The space of possible actions and signals in the initial work was intentionally kept very small. In the reinforcement algorithm, two variables must be kept for each possible action on each possible input. Thus, the required variables grow exponentially with each additional action added. In a simulation, memory and time may not matter; however, this is a real issue for autonomous robots with limited memory that we want to act in real-time.

Our current goal is to have the robots develop a compositional language. In a compositional language, there are words and relationships between words. For example, the robots may learn a word for "go straight" and modifiers such as "quickly" and "slowly". The advantage of a compositional language is that the robots need only learn each concept once, rather than relearn it every time it reappears in a new sentence. This is similar to English; we understand words and how they fit together and need not relearn everything when presented with a new sentence.

The reuse of concepts on robots will save both learning time and memory. If the robots had to build a reinforcement table for each new sentence, they would soon run out of memory. However, if the robots learn the words separately first, much less memory is required to learn composed sentences. Also, the amount of time necessary to learn the composed utterance should be much smaller than the the time required to learn the meaning of the whole utterance without any clues to the meaning of the parts.

6 Conclusion

The robots have shown that they can develop simple languages in supervised situations. These learned languages are adaptable, allowing the robots to respond to novel or changing situations.

Current work is exploring the development of languages that are dependent upon the world for motivation and reinforcement instead of a human instructor. The developed languages will allow us to examine how the robots abstract information about their environments into a world model. Initially, the robots will make many communication errors; their errors will be corrected as they learn in the world. Since the robots' languages will improve with feedback from their environment, they will in theory get better at their tasks as they continue to explore the world. Additionally, the languages allow the robots to talk about the real world without needing a static world model, so the robots will be able to adapt to changing environments.

7 Acknowledgments

Parts of the work described in this work were done jointly with Professor Lynn Andrea Stein in the AP Group at the MIT Artificial Intelligence Laboratory. Thanks to Dave Baggett, Oded Maron and Mike Wessler for comments on earlier drafts of this paper.

References

[Kaelbling, 1990] Leslie Pack Kaelbling. Learning in embedded systems. Technical Report TR–90–04, Teleos Research, Palo Alto, California, June 1990.

[MacLennan, 1990] Bruce MacLennan. Evolution of communication in a population of simple machines. Technical Report CS–90–99, University of Tennessee, Knoxville, Tennessee, January 1990.

[Martin and Sargent, 1991] Fred Martin and Randy Sargent. The MIT sensor robot: User's guide and technical reference. October 1991.

[Shewchuk, 1991] John P. Shewchuk. Ph.D. thesis proposal. Department of Computer Science, Brown University, Providence, Rhode Island, 1991.

[Sutton, 1992] Richard S. Sutton. Special issue on reinforcement learning. *Machine Learning*, 8(3–4), May 1992.

[Werner and Dyer, 1990] Gregory M. Werner and Michael G. Dyer. Evolution of communication in artificial organisms. Technical Report UCLA–AI–90–06, University of California, Los Angles, California, November 1990.

[Yanco and Stein, 1993] Holly Yanco and Lynn Andrea Stein. An adaptive communication protocol for cooperating mobile robots. In *From Animals to Animats: Proceedings of the Second International Conference on the Simulation of Adaptive Behavior*, pages 478–485. The MIT Press/Bradford Books, 1993.