Obsage - An Agent-based Model of Observational Learning

Keith Jones and Michael Van Wie
{kmj9907,mpv}@cs.rit.edu
Department of Computer Science
Rochester Institute of Technology
Rochester, NY 14623

October 2, 2003

Abstract

In many social systems, individuals are able to learn new behaviors by observing others. This paper describes experiments in augmenting reinforcement learners in an artificial society with the ability to observe each others' actions and rewards; the initial hypothesis was that observant agents would have an advantage over their non-observant counterparts. Experimentation shows that observation helps most when experienced teachers are available, and is most effective early in an agent's life-cycle.

1 Introduction

The field of Artificial Societies (AS) applies computer science techniques to study the many facets of sociology with a "bottom up" approach, as opposed to the traditional sociology approach, which attempts to answer questions from the top down. Using AS, simple models can be created which allow researchers to manipulate some set of interesting variables and remove or hold constant the uninteresting ones. This technique allows researchers to observe how variations in individual agents' traits or behaviors affect the qualities of the society as a whole. Examples of work in this field include the Schelling's model of racial segregation [10], the Echo environment [4, 5], and Sugarscape [3].

In many social systems, individuals are able to learn new behaviors by observing others. There has been research involving observational learning in humans [9], animals [2, 8], and even robots [6]. We wish to study the effects of observational learning by individuals on the society they are a part of.

This paper describes our model and experiments. Section 2 describes our agent-based model, and the basic learning algorithm used by all agents. Section 3 summarizes our experiments and results. And Section 4 gives our conclusions and ideas for future directions.

2 Environment and Model

In this section, we describe our artificial society (Section 2.1) and our learning model (Section 2.2).

2.1 Obsage: an Artificial Society

Obsage is a discrete model of a society that exists in a lattice environment, which wraps at both sides and at the top and bottom. Time is also discrete. There are two types of environmental inhabitants: agents, and objects. Agents perform *actions* on objects in order to get fitness rewards, which allow them to live longer. Agents each have an associated metabolism, which is the rate at which their fitness decays over time. At 0 fitness, an agent dies. Therefore, to survive, an agent must on average find fitness rewards equal to its metabolism each turn.

Figure 4 shows the Obsage model's viewer application. Agents are represented by circles, and objects are represented by lowercase letters. When an agent performs an action on an object, it is represented by a lower- case letter in the top-left of the grid box the agent occupies. The fitness reward generated by that action is displayed in the box's top-right corner.

Agents with the capacity to observe their neighbors are distinguished by a dot in the center; agents who are actively observing have a ring around the dot. Solid-colored agents learn through their own actions; dotted agents learn through a combination of their own actions and their observations of others.

All agents are granted the ability to learn through their own experience. The model is designed to test the hypothesis that augmenting learning-through-experience with observation will benefit both the individual and society.

2.2 Learning Model

In the Obsage model, actions do not interact with each other over time: performing a particular action on a particular object will always result in the same reward. Therefore, the learning problem reduces to one of learning which rewards are associated with which actions and objects. An agent with a complete mapping of actions and objects to rewards has a perfect policy. In such an environment, the main challenge a learning algorithm must overcome is the famous explore-vs.-exploit problem[11]. We have settled on the softmax algorithm[11] as a way to deal with this problem.

Given a partially filled policy, with some actions giving known rewards and others potentially unknown, the softmax algorithm draws the next action from a Maxwell- Boltzmann distribution, using action a's reward $R(a)$ as the main factor:

$$p(a) = \frac{e^{R(a)/\tau}}{\sum_{b=1}^{n} e^{R(b)/\tau}}$$

The parameter τ is a positive integer which works similarly to the temperature in simulated annealing: larger values of τ cause the agent to favor exploration over exploitation. Over time τ is decreased, thus allowing well-learned agents to exploit the knowledge they have gained.

3 Experiments

In this section, we describe several sets of experiments that investigate the effects of observation on societies of learning agents. Our initial assumption was that observational learners would have an advantage in most environments, but that turned out not to be the case (Section 3.1). However, when their environment included "teacher agents," as in Section 3.2, observational learners gained a marked advantage. This fact led us to try a generational model, where new agents were added to a population that had already had some time to learn (Section 3.3). In the generational society, observational learners kept their advantage over non-observational learners. There are indications that a *developmental* model, where learners observe more at the beginning of their life cycle and less as they mature, might give further benefit.

There are three measurements that can easily be derived from an Obsage society: the number of living agents; the average fitness of living agents; and the score: the average fitness of all agents, including dead ones. The score appears to be the most accurate method of measuring a society's overall performance, as long as the only cause of death is lack of fitness. When comparing individual agents or subsets of agents to each other, it may be more useful to examine just the average fitness at a given point.

3.1 Initial Experiments

Initially, we divided the society into an observational and a non-observational group. No agents had any knowledge of the environment. Over eight trials, we varied the observation propensity O from 25 to 50 percent, and the number of observant agents from 10 to 40 percent of the population.

The data in the following table shows that there was little, if any, benefit for the observant subsets in societies modeled in this experiment.

	observation propensity		
%	0	25	50
10	197	236	252
20	204	202	220
30	212	180	197
40	202	224	191

3.2 Perfect Teachers

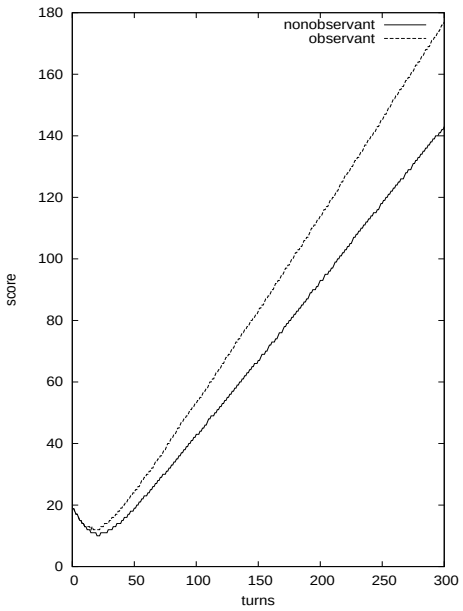
This scenario is an attempt to provide observant agents the optimal opportunity for observational learning. A society consisting of thirty agents and sixty objects was created, divided equally between five object types. Half of the agents had a priori knowledge of the optimal action for each object type. We call this subset "teachers", because they are good models for observation. The other half had no preset knowledge. The "teacher" agents were nonobservant. Observant agents used the agent-based observation policy. Note there was nothing to prevent observant agents from observing other non-teachers, except that the agent-based observation policy made them prefer to observe more fit agents.

There were two cases in this experiment: a control group with $O = 0$, and a test group with $O = 50$. There were thirty agents in the environment. Fifteen of those agents were "teachers". In this test, there were five object types and five actions.

While there was a small difference between the scores of observant and nonobservant agents, the difference was constant from turn to turn. This indicates that the observant agents received a head start, but the nonobservant agents were soon operating as well as the observant agents. We believe this is because the size of the search space was relatively small, so nonobservant agents were able to find near optimal actions relatively quickly.

Because the *Perfect Teachers* experiment did not find that the observant cases had a significant advantage, we hypothesized that the reason for this result was that the observant agents were learning about the environment quickly. To test this theory, we implemented a new environment that was the same as the previous, except that we doubled the number of object types and actions. That is, the search space is 10×10 instead of 5×5.

Figure 1: Perfect Teacher II Scores (10×10)



As expected, in the more difficult search space, groups of observant agents tend to be more successful. Figure 1 shows the average score for the societies, as time progresses. Note that this graph shows the score for the particular society subsets; not the score for the society, nor the average fitness of the society. Therefore, this graph is not an indicator of the overall society success, or an indicator of success on an individual level. In fact, observant agents often do not outperform their nonobservant counterparts on an individual level, but they do experience *fewer deaths* early on in the trial. Because fewer agents die, the observant group is generally able to produce an ever-increasing performance ratio over the nonobservant group, as is evidenced by the greater slope for the observant agents in Figure 1.

3.3 Generational Learning

We had a hypothesis that the benefits of observation were more apparent in a society where, early on, agents had the oppor-

tunity to observe other agents that already had useful knowledge. The *Perfect Teacher* experiments provide strong evidence that this is the case, but required a contrived scenario to do so. A more realistic scenario that should have similar results would involve a "generational" model, where agents must give birth in order to further their society, and young agents must learn by observation.

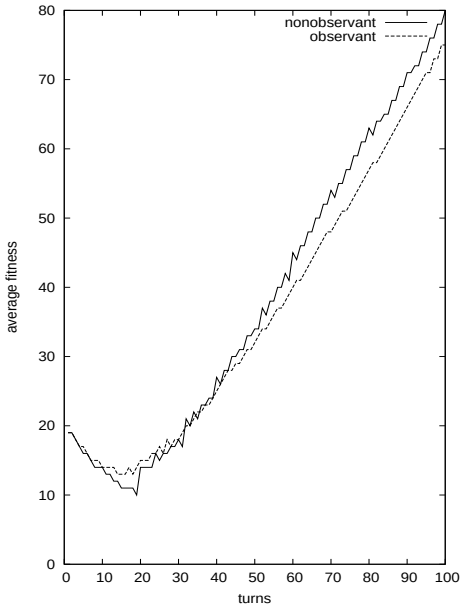
This scenario was an attempt to simulate the relevant characteristics of a generational model, without adding all the complexity that model would require. To do this, we ran a society for 100 turns and then used the data from various turns as input for new societies, adding some new agents with no memory, to the model.

We created a single precursor society with thirty initial agents and sixty objects, and ran that environment for 100 turns. We then created new input files from the precursor output, at turns one, ten, fifty, and 100, adding ten new agents to the societies. We set $O = 0$ for the agents in these files. We then created modified versions of those input files, with $O = 50$. This lead to eight cases, comparing observant to nonobservant "new" agents in society at various stages of development. We ran five trials of each case. The cases at Turn 1 form a kind of control group, because, for the most part, the precursor agents did not have an opportunity to learn anything about the environment during the first turn.

In this experiment, the observant agents performed significantly better as a group than the nonobservant agents in every case except the one involving turn 1. In that case, the performance was higher, but only slightly so. They did not, however, perform noticeably better on an individual level. The improved group performance was a result of fewer agents dying during the initial turns.

It is interesting to note that in the control group, the nonobservant agents performed significantly better than they did in any other case. This leads me to wonder if the knowledge the precursor agents acquired actually had a negative impact on the fitnesses of the new nonobservant agents. This could be the case if agents competed for nearby resources. Figure 2 shows that the observant agents are able to use observation in order to capitalize on exploitation earlier, and therefor maintain higher fitnesses. The slope for the nonobservant agent increases because the lesser fit ones die off, and the nonobservant agents eventually equal the average fitness of the observant agents. Figure 3 show that the score for the observant agents increases at a greater rate than the score for nonobservant agents; this is expected

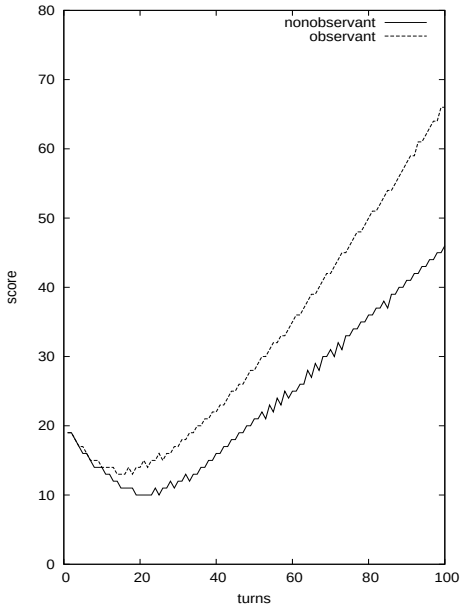
Figure 2: Generational Average Fitnesses



4 Conclusions and Future Work

The experiments show that observation alone is not enough for an agent to have an advantage over its peers when the act of observing takes time that could more profitably be used exploring the environment using independent learning. In order to see clear benefits from choosing observation over independent exploration, there must be accomplished individuals (teachers) present early in the developmental cycle, when the amount of exploration yet to

Figure 3: Generational Scores



do is very large. The effectiveness of observation is more visible in the survival rates of observant agents, rather than higher long-term fitness.

This suggests that a generational artificial society may show further advantage for observant agents. When agents are introduced into a preexisting society with experienced agents to observe, they are provided with a faster route to exploitation than the trial and error experienced during the initial stages of learning by experience.

One scenario in which observational learners would be expected to have an advantage is an environment which contains interactions that may be deadly for the agents involved. Those who can only learn through experience cannot learn from deadly interactions, but observational learners should be able to. This is not always the case, as Obsage shows. Obsage agents are no more likely to observe deadly interactions than they are any other interactions, and therefore don't often gain the knowledge required to avoid deadly interactions. An artificial society which models interrupt-driven learning might more accurately reflect real world observational learning. Observation would be prompted by external cues, rather than being a decision originating solely from an agent's internal state and logic. It is likely that such a model would prove observant agents more capable of survival in this "deadly interaction" scenario. For example, at a zoo, a curious person may step too close to a dangerous animal's cage. Most of the other visitors would probably not take notice of such an action, but if the unwitting person is attacked, those visitors will likely learn the true danger of stepping too close to a cage.

Based on these conclusions, we believe it would be profitable to develop a truly generational artificial society that includes natural death and reproduction, for the study of observational learning over the long term. Also, for more accurate modeling of real-world observational learning, the model would support an interrupt mechanism to alert the agent to observe seemingly more interesting interactions.

References

[1] Alonso, Eduardo; d'Inverno, Mark; Kudenko, Daniel; Luck, Michael; and Noble, Jason. Comments on Learning In Multi-Agent Systems. *In Proceedings of the Third Workshop, UK SIG on Multi-Agent Systems.* 2000.

[2] Bugnyar, Thomas and Kotrschal, Kurt. Observational learning and the raiding of food caches in ravens, *Corvus corax*: is it 'tactical deception'? *Animal Behaviour*, 64:185-195, 2002.

[3] Epstein, Joshua M. and Axtell, Robert L. *Growing Artificial Societies.* Washington, D.C.: Brookings Institution Press, 1996.

[4] Forrest, Stephanie and Jones, Terry. Modeling complex adaptive systems with Echo. In *Complex Systems: Mechanism of Adaptation*, R.J. Stonier and X.H. Yu (eds.), Amsterdam, The Netherlands: IOS Press, 3-21, 1994.

[5] Hraber, Peter; Jones, Terry; and Forrest, Stephanie. The ecology of Echo. *Artificial Life*, 3(3):165-190, 1994.

[6] Kuniyoshi, Yasuo; Rougeaux, Sebastien; Ishii, Makoto; Kita, Nobuyaki; Sakane, Shigeyuki; and Kakikura, Masayushi. Cooperation by Observation: The Framework and Basic Task Patterns. *Proceedings of the IEEE International Conference on Robotics and Automation*, 767-774, 1994.

[7] Lansign, Stephen J. 'Artificial Societies' and the Social Sciences. *Artificial Life* 8:279-292, 2002.

[8] Pongràcz, Pèter; Miklòski, Àdàm; Kubinyi, Enikõ; Topàl, Jòsef; and Csànyi, Vilmos. Interaction between individual experience and social learning in dogs. *Animal Behaviour*, 65:595-603, 2003.

[9] Rogoff, Barbara; Paradise, Ruth; Arauz, Rebecca Mejiía; Correa-Chávez, Maricela; and Angelillo, Cathy. Firsthand Learning Through Intent Participation. *Annual Review of Psychology*, 54:175-203, 2003.

[10] Schelling, Thomas C. Dynamic Models of Segregation. *Journal of Mathematical Sociology*, 1:143-186, 1971.

[11] Sutton, Richard S. and Barto, Andrew G. *Reinforcement Learning.* Cambridge, MA. The MIT Press, 1998.

Figure 4: Viewer.py - Graphical Viewer for Obsage Models.