Great Expectations: Scaling Up Learning by Embracing Biology and Complexity

Maja J Matarić
Computer Science Department
University of Southern California
Los Angeles, CA 90089-0781
mataric@cs.usc.edu
http://www-robotics.usc.edu/~maja

Abstract

We propose that one of the stumbling blocks in scaling up learning (to real-world problems and domains) has come from the emphasis of mathematically pure rather than practical approaches. Striving for optimality has overshadowed the plight for efficiency, and the attempt to avoid experimental bias has crippled the ability to focus on principled methods for building in domain knowledge. In what we call "great expectation", most have come to expect that a learning algorithm must achieve optimal performance with little or no built in knowledge. To make this possible, the algorithm is allowed to learn from prohibitively many examples and/or for a prohibitively long time. This trade-off between *a priori* bias and learning time is impractical for a great majority practical applications. Furthermore, it is counter to the way biological systems learn.

We argue that taking inspiration from biology, focusing on real-world domains and tasks, and experimentally validating all algorithms in such domains, will result in more scalable approaches. In our own work, we pursue an experimental approach in two highly uncertain, dynamic, and high-dimensional domains: multi-robot learning, and learning by imitation. Both force us to deal with perceptual and action uncertainty, non-stationarity, and real-time constraints. As a result, our approaches to learning, inspired by learning in biology, strive for efficient solutions that can cope with the challenges of real-world domains.

Introduction

We propose that one of the stumbling blocks in scaling up learning (to real-world problems and domains) has come from the emphasis of mathematically pure rather than practical and biologically-inspired approaches to learning. Striving for optimality has overshadowed the plight for efficiency, and the attempt to avoid experimental bias has crippled the ability to focus on principled methods for building in domain knowledge. In what we call "great expectation", many have come to demand that a learning algorithm achieve optimal performance without any built in knowledge. To make this possible, most have allowed the algorithms to learn from vast examples and prohibitively numerous trials. This trade-off has been largely impractical for real-world domains, and it stands in sharp contrast to the way biological systems learn.

We argue that paying more attention to learning in biology, situating learning systems in real-world domains, and experimentally validating of all algorithms in such domains, will bring about a shift in focus toward more scalable approaches. In our own work, we take inspiration from biology, not only from its strengths, but also from its limitations. Additionally, we pursue an experimental approach in complex, challenging environments, to validate the learning systems we develop. This has lead us to consider a number of methods that have enabled efficient learning in domains that are as yet theoretically un-addressed, due to their complexity. We focus on multi-robot learning and learning by imitation, both in physical robot systems, ridden with uncertainty, real-time constraints, and non-stationarity. In the second part of this paper we highlight some of the methods that have made learning in those domains possible, and argue that they, and other such approaches, deserve more attention, from both the practical and theoretical learning communities.

Biological v. Synthetic Learning

Inspiration for learning in AI comes from biological systems' capability to adapt and learn in changing environments, conditions, situations, and for changing goals, and tasks. AI strives for similar flexibility in artificial learning systems, but also adds some rigid (and we argue unreasonable) constraints. Ideally, a learning system should be able to start from *tabula raza*, without built-in knowledge, thus removing domain-specific and *ad hoc* bias. Additionally, once having learned, the systems should perform optimally.

Biological systems are not capable of meeting those expectations. They depend on a great deal of prior knowledge, in the form of innate structure. Furthermore, they take advantage of environmental structure, which includes various helpful features, such as perceptual invariances, patient teachers, and demonstrators.

Together, these provide so-called "scaffolding" for the learning system, as well as examples, feedback, shaping, reward, and explicit and implicit knowledge (McFarland 1985, McFarland 1987, Piaget 1962, Piaget & Inhendler 1967). Biological systems interleave learning with interaction, execution, and communication, in addition to exploration and exploitation.

Perhaps more importantly, biological systems do not behave optimally. Nor do they learn to do so. Instead, they strive for sufficiently good (satisficing) performance in any given context (task or goal), and handle a great variety of contexts, effectively transferring and generalizing between them. While optimization plays a role in simple, confined areas, such as trajectory and path planning (in motor control and navigation), it does not appear to scale up to more complex and general behavior levels. This is likely due to the fact that optimality, in the vast space of animal behavior, drives, motivations, and changing environments, is ill defined, and is also of little use given the transience of most specific situations. The ability to adapt quickly and generalize, on the other hand, plays a critical role.

Thus, we believe that while biology serves as inspiration for machine learning, it does not serve as enough of a model to guide learning methods toward more fruitful approaches, ones that focus away from optimality and toward faster adaptation and generalization in challenging domains.

Developmental Psychology Insights

A great deal of insight can be gleaned from what is known about developmental stages in children, as they learn about their complex environments. Nowhere is the lesson of innate structure more obvious than in infants, who arrive into their massively complex environment equipped with a great deal of innate knowledge, as well as strong motivation for observation, experimentation, and imitation. In addition, psychological evidence clearly shows that it is the structure of the environment that is critical in allowing children to learn about their complex world, including helpful caretakers who provide just the right scaffolding upon which regularity of information and knowledge can be built. Evidence also clearly points toward staged learning, proceeding in clear developmental steps, whose attainment cannot be achieved prematurely, until the proper foundation is laid down. But once reached, the new stage clearly surpasses its predecessors and permanently expands the child's behavioral and knowledge repertoire. There is a reason why we are born with reflexive crawling and clinging, but those are lost to be replaced by true volitional motor skills. There is good reason for our tendency to grasp and handle objects insatiably in the early years, while equipped with innate expectations about objects and trajectories and even gravity. There is useful staging to our inability to walk or talk until the age of one, by which time many other more basic skills are in place. And so on.

The very basic lessons of developmental psychology we highlight here are the wealth of innate structure, skill, and implicit knowledge, as well as a staged progression of increasingly more complex learning, supported by environmental structure. While a handful of research projects are aimed at addressing one of the above principles, machine learning would greatly benefit if these were to find their way into our culture of what constitutes a rich learning system.

Task and Domain Complexity

Biological systems function in their environmental niches, which provide both helpful structure and challenging stumbling blocks. Synthetic systems, on the other hand, are typically not situated as realistically, and thus provided little if any of either. Consequently, they are not given enough to work with, against, or toward.

By embedding systems in complex environments, we are forced to address real problems and are not allowed to accidentally abstract them away. As the history of AI has amply shown, dividing problems into isolated subareas tends to produce highly specialized solutions that do not generalize to realistic, integrated domains.

A commonly used counter argument is that complex environments force one to cut corners and develop *ad hoc* solutions that do not provide general insight. However, as practice has demonstrated, being faced with more complex domains not only forced us to adopt realistic assumptions, but also produces relevant solutions which, when taken together, provide consistent insights on which to build and test principles and for which to eventually develop formal theories. It is our belief that it is more fruitful to address complex problems and learn from partial solutions and even failures than to spend time on simpler ones based on assumptions that do not scale up.

Ideally, the complex problem of learning should be addressed from all of these vantage points, from both theory and practice, from both abstract and applied contexts. In reality, this happens to a certain degree, but the results of the different approaches rarely inform each other, thus failing to build on each other's strength and lessons learned.

Our Approach

In the early stages of our work with multi-robot systems, in the early 1990s, we often encountered the following reductionist argument: "control of a single robot is poorly understood, so why make the problem worse by considering the multi-robot case? It will tell us nothing, it is too hard!" Our answer was always that the multi-robot case tells us about a wholly different problem, with fundamentally novel dynamics and challenges, as well as some facilitating aspects. Now, a decade later, multi-robot control is a fast-growing and popular area of research, and this growth cannot be attributed to our complete understanding of the single-robot case, but rather to a recognition of the relevance of a new and independent research area.

We follow the same philosophy in our research in learning. By considering complex tasks and environments, and embracing the full realism of uncertainty and non-stationarity, we believe we are addressing a fundamentally different (as well as more pertinent) learning problem, which should yield key insights as well as more direct practical applications.

Tasks and Environments

Our work is focused on learning on physical agents, with local on-board sensing and action, real-time constraints, and inherent uncertainty and non-stationarity that abound in such domains. We have addressed the following questions in learning so far:

- How can a behavior-based robot (i.e, one that does not employ a centralized planning capability) learn from its past experience in order to improve its performance in a multi-robot domain? Tasks: distributed object collection, distributed mapping, distributed pox pushing.

- How can a group of behavior-based robots learn social rules, so as to improve group-level performance based only on local sensing and control? Tasks: distributed object collection.

- How can a pair of identical robots learn a cooperative tightly-coupled task? Tasks: box pushing, formations, object manipulation.

- How can a behavior-based robot learn the interaction dynamics of its environment, in order to build a light-weight model that can be used for achieving a variety of tasks? Task: distributed object collection and manipulation.

- How can a robot learn a new skill by imitation, from observing a human? Tasks: arbitrary movement sequences, dance, manipulating objects, doors.

Methods

Built-In Structure and Knowledge

Our approach is based on behavior-based control, in which a collection of distributed, concurrent behaviors controls an agent or robot. Each behavior can take inputs from the sensors and/or other behaviors and send outputs to the effectors and/or other behaviors. This framework has been shown to be highly robust and effective in robot control and multi-robot coordination (Arkin 1998, Matarić 1997*a*). It has also been shown to support distributing representations (Brooks 1991, Matarić 1992), so its expressive power is not limited.

The aspect of the behavior-based approach that is most relevant to our learning work is the built-in structure that is afforded by the existence of behaviors. Since they are time-extended procedure that encapsulate low-level details of control, behaviors serve as a rich substrate for learning. They enable learning at various levels:

- adaptation within behaviors (e.g., parameters)

- behavior selection and coordination (though switching or fusion)

- new behaviors

- collections of behaviors

In all cases, the behavioral structure elevates the representation of the system. Consider the case of learning behavior selection, which can be directly mapped to the reinforcement learning framework. Our work (Matarić 1997*b*) demonstrated that reinforcement learning, when applied to learning policies that map behavior activation conditions and behaviors themselves, was highly effective even in uncertain, dynamic domains. The conditions can be viewed as generalized input states, but can be obtained by off-line computation, so on-line generalization is not required. This reduces the state space of the system significantly, and enables real-time learning, particularly pertinent in robotics. We demonstrated this idea on learning to forage in a group (Matarić 1994).

We have also used the behavior substrate to facilitate learning by imitation (Matarić 1999). The notion of encapsulating control into behavioral primitives comes from neuroscience evidence (Bizzi, Mussa-Ivaldi & Giszter 1991, Mussa-Ivaldi, Giszter & Bizzi 1994); we have argued that it is also critical in facilitating imitation learning, by allowing the learner a more direct mapping between the observed and the executed, a reduced state space, and a predictive capability.

Behaviors are just one type of structure that can be principally provided to a system in order to facilitate the learning process. Some others have been explored, but the problem of how to encode *a priori* knowledge and bias in a principled way, in order to facilitate learning, is not well studied. We suggest that biology gives us a strong impetus that it should be given more attention.

Shaping

In the last decade, reinforcement learning has gained a great deal of attention and popularity. It is currently the paradigm of choice for robot learning, due to its convenient unsupervised nature. However, inconsistency of feedback, and the uncertainty and the delayed nature of credit assignment in complex (especially multi-agent) environments, strongly suggest the need for a richer paradigm.

The notion of shaping, providing intermediate reward for behaviors that gradually get the learner closer to the desired goal or skill, comes from psychology. It requires an informed and helpful teacher, but still bypasses full supervision. This notion has been effectively applied in robot learning. Our own work demonstrated that reinforcement learning, when applied to a group of robots, fails due to the sparsity and inconsistency of the learning signal (Matarić 1997*b*). By introducing shaping, even in the simple form of progress estimators, a richer learning signal is devised, which enables the learner to statistically separate transient noise and events from the more permanent ones, without the need for explicit models.

Shaping was also successfully applied by others in single-robot learning experiments (Dorigo & Colombetti 1997), and is gaining popularity. However, because of the bias introduced by the teacher, it is not as widely accepted as we feel it should be, given its effectiveness. Furthermore, it has received no formal treatment, which relegates it, in the views of the the theoretical learning community, to a "short-cut" rather than a powerful mechanism worth exploring.

Imitation/Demonstration

We have already argued that starting out with *tabula raza* systems is unreasonable for real world tasks. Bootstrapping the system with a good initial policy/controller is an excellent way to accelerate its adaptive capabilities, but coming up with such a policy may be almost as difficult as programming the system by hand. One alternative is to provide it by demonstration, by showing the learner what to do and how to do it.

Imitation learning has grown in popularity in the last few years (Schaal 1999, Matarić 1999). While it is in itself a difficult learning problem, not a simplifying step, it addresses an important form of skill acquisition, ubiquitous in human intelligence. The challenges of imitation involve selecting a teacher, distinguishing salient features of what is to be learned, selecting an appropriate internal representation of the observed behaviors, mapping it to the learner's own behavior repertoire, and adapting/tuning the repertoire based on the difference between the observed and executed movement, if an error signal is provided. Since each of these set-ups is a research challenge in itself, the imitation learning problem as a whole is quite a formidable one. However, since it is an important and powerful means of skill acquisition in nature, it merits investigation.

Imitation learning is a major area of our current research, applied to the domain of motor skill learning in humanoid robots. This involves a high degree of freedom control system (currently 20 DOF), with sensory input from vision and joint angle measurements, which effectively put to the test any learning approaches we develop.

Communication

The environment, including other agents/robots, provides valuable information to the learning system. Besides direct sensing, communication is the primary means of acquiring such information. Although a powerful way of information sharing, communication is all but dismissed as a "crutch" in the context of learning. While in biological systems in plays a crucial role, in synthetic ones it has not yet been legitimized as a principled element of a learning system, nor treated formally.

Our work on learning tightly-coupled cooperation has employed simple communication techniques to effectively cope with both the partial observability problem and the credit assignment problems in learning. In a box pushing scenario, we used communication to share sensory information, thus enabling the robots to expand their outlook, but by considering the shared aspect of the information, without expanding the joint state space. The communication messages also enabled the robots to synchronize their actions, which resulted in turn-taking being learned as a part of the stable cooperative policy (Matarić 1996).

We also employed communication in a social learning framework, to effectively combat the adverse effects of the greedy nature of reinforcement learning. By communicating the reinforcement received by each robot to the ones near by (in space and time), and sharing the reinforcement, the group as a whole was able to learn social rules (for example yielding/giving way) that a collection of greedy individuals could not.

Another important but overlooked form of communication is *stigmergy*, the ability to react to the effects of actions of others in the environment. This principle is the foundation of ant trails, termite nests, and other complex structures and behaviors resulting from simple individual interactions (**?**). This notion can be applied to learning as well; not only what the system does, but what it stores/learns, can be based on environmentally-encoded effects. Our work has explored stigmergy in the context of coordinated movement (Werger & Matarić 1996, Werger & Matarić 1999), and we are considering it in the context of learning.

Besides its uses in biology, communication has been demonstrated to be critical for superliner performance in group domains, and it has a similar effect on distributed learning. Thus, we believe that it deserves more careful attention in machine learning research.

Anytime Learning

As discussed earlier, optimal performance is the typical goal of learning systems. However, in many complex domains, and in particular in non-stationary ones, optimality is not well defined, or is prohibitively slow to achieve.

Much like anytime planning, learning in complex environments must exhibit anytime properties. A learning system situated in a dynamic environment cannot have the luxury of a slow training/exploration/learning process. What is needed is the ability to provide incrementally improved solutions which, even based on a small amount of data, can provide incremental improvement of performance. Biology abounds with examples of one-shot and staged learning as inspirations for this pursuit.

The ability to learn incrementally and over short time periods is critical in non-stationary environments, such as those with multiple agents, or harder still, multiple learners. Our work has addressed both of these domains (Michaud & Matarić 1998, Goldberg & Matarić 1999) by developing an approach to on-line learning of models of agent-environment interactions. These can be quickly acquired, incrementally refined, and used for in variety of tasks (Goldberg & Matarić 1999).

Evaluating learning performance is a difficult problem, and notions of satisficing (Simon 1969), and efficiency are not nearly as accepted as that of optimality. However, we believe that these, in conjunction with the others listed above, are critical to scaling up learning to make it comparative to biological system capabilities and to real-world demands.

Summary

We have argued that the important factors in the way of scaling up the complexity and effectiveness of machine learning are: 1) the emphasis on optimality, 2) the lack of built-in knowledge and structure, and 3) the lack of realistic test domains and relevant experimental validation, 4) the lack of cross-fertilization between theoretical and applied learning work. We then outlined inspiration and insights from biological systems, which demonstrate how efficient adaptive behavior can be achieved in a very different way. Finally, we proposed a set of possible research topics that are currently underemphasized, but should prove to be fruitful for making machine learning more robust and scalable, as well as relevant to real-world tasks and domains. This list is not meant to be exhaustive, but rather reflective of the areas we feel are critical and on which we have focused our own efforts. We briefly described our own relevant research with respect to those topic areas, in hopes of instigating discussion and further interest in turning a few research spotlights toward these important issues.

Acknowledgments The author's learning work is supported by the National Science Foundation Career Award IRI-9624237, and the Office of Naval Research Grants N00014-95-1-0759 and N0014-99-1-0162. The work was performed in collaboration with Francois Michaud and Dani Goldberg.

References

Arkin, R. C. (1998), *Behavior-Based Robotics*, MIT Press, California.

Bizzi, E., Mussa-Ivaldi, F. A. & Giszter, S. (1991), 'Computations Underlying the Execution of Movement: A Biological Perspective', *Science* **253**, 287–291.

Brooks, R. A. (1991), Intelligence Without Reason, *in* 'Proceedings, IJCAI-91', Morgan Kaufmann Publishers, Inc., Sydney, Australia, pp. 569–595.

Dorigo, M. & Colombetti, M. (1997), *Robot Shaping: An Experiment in Behavior Engineering*, MIT Press, Cambridge, MA.

Goldberg, D. & Matarić, M. J. (1999), Coordinating Mobile Robot Group Behavior Using a Model of Interaction Dynamics, *in* 'Proceedings, The Third International Conference on Autonomous Agents (Agents '99)', Seattle, Washington.

Holland, O. (2000), 'big stigmergy paper', *Artificial Life*.

Matarić, M. J. (1992), 'Integration of Representation Into Goal-Driven Behavior-Based Robots', *IEEE Transactions on Robotics and Automation* **8**(3), 304–312.

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. (1996), Using Communication to Reduce Locality in Distributed Multi-Agent Learning, Technical Report Computer Science CS-96-190, Brandeis University.

Matarić, M. J. (1997*a*), 'Behavior-Based Control: Examples from Navigation, Learning, and Group Behavior', *Journal of Experimental and Theoretical Artificial Intelligence* **9**(2–3), 323–336.

Matarić, M. J. (1997*b*), 'Reinforcement Learning in the Multi-Robot Domain', *Autonomous Robots* **4**(1), 73–83.

Matarić, M. J. (1999), Sensory-Motor Primitives as a Basis for Imitation: Linking Perception to Action and Biology to Robotics, *in* C. Nehaniv & K. Dautenhahn, eds, 'Imitation in Animals and Artifacts', The MIT Press.

McFarland, D. (1985), *Animal Behavior*, Benjamin Cummings, Menlo Park, CA.

McFarland, D. (1987), *The Oxford Companion to Animal Behavior*, Oxford University Press.

Michaud, F. & Matarić, M. J. (1998), 'Learning from History for Behavior-Based Mobile Robots in Non-stationary Conditions', *Autonomous Robots and Machine Learning* **AR:5,ML:31**, AR:335–354,ML:141–167.

Mussa-Ivaldi, F. A., Giszter, S. F. & Bizzi, E. (1994), 'Linear combinations of primitives in vertebrate motor control', *Proceedings of the National Academy of Sciences* **91**, 7534–7538.

Piaget, J. (1962), *Play, Dreams and Imitation in Children*, W. W. Norton & Co., New York.

Piaget, J. & Inhendler, B. (1967), The Child's Conception of Space, W. W. Norton & Co., NY.

Schaal, S. (1999), 'Is imitation learning the route to humanoid robots?', *Trends in Cognitive Science* **3**(6), 233–242.

Simon, H. (1969), *The Sciences of the Artificial*, MIT Press.

Werger, B. B. & Matarić, M. J. (1996), Robotic Food Chains: Externalization of State and Program for Minimal-Agent Foraging, *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', MIT Press, pp. 625–634.

Werger, B. B. & Matarić, M. J. (1999), Exploiting Embodiment in Multi-Robot Teams, Technical Report IRIS–99–378, University of Southern California, Institute for Robotics and Intelligent Systems.