The Knowledge Engineering Review, Vol. 11:4, 1996, 361–364

Learning in autonomous robots

HENRY HEXMOOR[1] and LISA MEEDEN[2]

[1]*226 Bell Hall, Department of Computer Science, SUNY at Buffalo, Buffalo, NY 14260, USA. Email: hexmoor@cs.buffalo.edu*
[2]Computer Science Program, Swarthmore College, 500 College Ave, Swarthmore, PA 19081, USA. Email: meeden@cs.swarthmore.edu}

Learning takes place when the system makes changes to its internal structure so as to improve some metric on its long-term future performance, as measured by a fixed standard (Simon, 1983).

Let us take "learning" to mean, roughly, the improvement of a system's behaviour by making it more appropriate for the environment in which it is embedded (Kaelbling, 1993).

1 Introduction

The definitions above, separated by ten years, represent two very different conceptions of learning. For Simon learning depends on an internal change in representation, and for Kaelbling it is instead measured in terms of an external change in behaviour. Furthermore, Kaelbling's focus on the situatedness of the learning system being embedded in its environment reflects the recent experience gained by much direct experimentation with physical robots.

Learning to control a physical robot has proved to be quite a challenge because of the unpredictableness of the real world coupled with the inaccuracies inherent in physical sensors and actuators. ROBOLEARN-96 focused on learning as it is applied to real robots to try to examine why successful learning algorithms for simulated domains do not seem to translate well to physical domains. We were specifically interested in addressing learning issues that remain as open problems in robotics.

Robots are developed for a wide variety of purposes with different requirements for learning. For instance, industrial robots are designed for repetitive tasks in highly structured environments. These robots may need to fine tune a parameter in their control module which is a minimal form of learning. On the other hand, mobile office robots are designed to interact with humans in more natural settings. This requires complex learning techniques involving goal satisfaction.

Robot learning can mean varying degrees of changes to internal structure and external behaviour. It is difficult to catalog all the different types of learning, but briefly they might include:

- changes to values of parameters in feedback loops,
- building topological models,
- developing fine-grained perceptual or motor skills,
- encoding successful routine associations between sensing and acting,
- acquiring new concepts, and
- learning architectural features such as coordinations and concurrency.

Figure 1 depicts a continuum of knowledge. In one extreme, the robots have complete domain knowledge. Such a robot is an engineered artifact and no learning is necessary. Near this extreme are examples of robots with missing parameter values. These are engineered systems containing control modules that lack optimal values for their parameters. In general, these control modules could be based on connectionist machines or feedback controls known as adaptive control systems. Current mobile robotics research is centered around systems that fine tune architectural parameters such as number of hidden nodes or the system dynamics. Most of the research presented at the workshop fell in the middle and the right end of the continuum. Many of the presenters argued that to construct a successful system, learning must be limited in use to portions of the system where the designer's knowledge is too limited to engineer a solution. Learning from scratch is too inefficient for any problem of reasonable complexity. Therefore, the other extreme on the continuum of starting with no domain knowledge was viewed as quite impractical.

Although beginning with a significant amount of domain knowledge is certainly more practical, there may be a disadvantage. These systems can build maps of their habitat and learn new behaviours. Yet, in a sense, these systems are given an encoded version of the possible sensory input. Chris Atkeson points out that the more interesting systems are the ones that discover the appropriate input and output for their tasks. These systems determine the relevance of sensory information to tasks, synthesize this sensory information and compose behaviours without ever being programmed for such interactions. Tabula rasa systems, as Minouri Asada put it, don't even have the basic concepts of building IO.

2 Review of the papers presented

All of the following papers are collected in the Robolearn Workshop Notes (ROBOLEARN, 1996).

2.1 Control and robotics

Asada and Nakamura describe a stereo-camera vision system to allow a robot to navigate around obstacles to a target. They use Q-learning on a state space defined in terms of the occlusion status of the target and the target's position in the visual field to map to nine possible motor actions. This state space provides only indirect information about the obstacles, but Asada and Nakamura believe that this simplification was the key to their success since it drastically reduced learning time. They note that automatic state space construction is an important open research problem.

Asoh, Motormura and Matsui present a robot that learns the layout of an office by simple questions and instructions about its whereabout in natural language. To represent position uncertainties, this work used probabilistic maps.

Kun and Miller present a fascinating biped walking robot. They use pre-planned motion sequences that can be adapted through neural network learning. Separate networks control the front to back balance, the left to right balance and foot stability. The biped is able to start and stop on demand and to walk with continuous motion at up to 100 steps a minute. Interestingly, slow walking was more difficult than fast walking because of the additional time required to balance on one foot.

2.2 Reinforcement learning

Choi, Yim and Doty describe a method called Environmental Reinforcement Learning for refining primitive behaviours through repetitive execution in structured environments. This is a useful technique for developing low-level behaviours such as an about face turn in robots with inaccurate motors and rudimentary sensors. In practice these simple tasks are not trivial because even when two robots are constructed in an identical manner, their actual performance may vary significantly

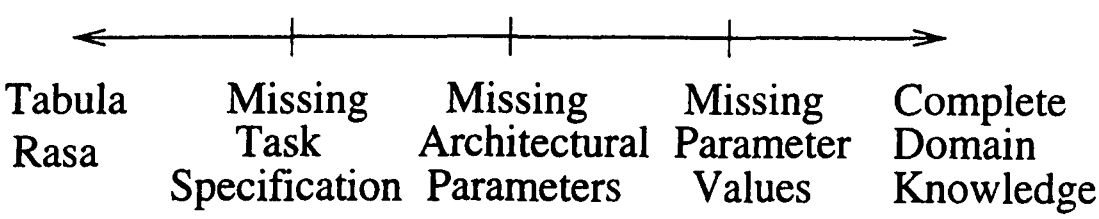

Figure 1 The continuum of domain knowledge for robot learning

due to slight differences in characteristics such as wheel radius. Since primitive behaviour refinement is an ongoing process, it is not clear if such a system is engaged in non-primitive behaviours, how it will distinguish between refinement of primitive behaviours and refinement of the more complex behaviours.

Jantz and Doty discuss two new goals for robotics research. First to develop a robot platform that is capable of surviving for an indefinite period of time without human intervention, and second to use such a platform to better evaluate robotics learning algorithms. They describe one platform that has been able to sustain continual week-long learning experiments.

Pendrith and Ryan present a new reinforcement learning algorithm called C-Trace that is designed to work well in noisy, non-Markovian environments. They compare their algorithm with 1-step Q-learning, as well as a handcrafted program, for the problem of controlling the walking gait of a six-legged robot. C-Trace is more selective in updates than Q-learning. The C-trace algorithm outperforms both Q-learning and the handcrafted solution, thus providing empirical evidence that C-trace is well suited for noisy and non-Markovian environments.

2.3 Neural networks and evolutionary learning

Sharkey, Heemkerk and Neary describe a supervised learning technique for training neural network controllers to guide a Nomad200 robot to any requested location within its environment while also avoiding obstacles. Their training method was to adapt a simple pre-wired behaviour through a teacher who intervened when a collision was imminent. The teacher provided joystick control until a safe path was available. They liken this technique to how adult animals help their young avoid dangerous situations. They then compare a subsumption style architecture, where the tasks of finding the goal and avoiding obstacles are accomplished by two separate neural networks, to a single network architecture that encompasses both sub-tasks. The results indicate that a single network provides better overall performance than the subsumption style system.

Smith and Cribbs introduce a hybrid model that integrates Q-learning, neural networks, learning classifier systems (LCS) and genetic algorithms (GA) into a single learning system for autonomous systems. The overall strategy of the system is represented in the Q-values, where the neural network encodes the mapping from states to Q-values. The state representation is derived from LCS. Finally, the GA is used to adapt the number of connections between the input and hidden layers of the neural network as well as the number of hidden units. The model has as yet only been tested on a simulated agent but appears to have promise.

Schultz, Grefenstette and Adams report on using genetic algorithms to learn a complex shepherding behaviour where one robot, the shepherd, guides another robot, the sheep, into a goal area, the pasture. Only the shepherd's behaviour is learned; the sheep is pre-programmed to move away from nearby objects. Thus the shepherd directs the sheep by approaching it until the sheep reacts and moves away. The authors demonstrate that a non-trivial behaviour, such as shepherding, learned under simulation can transfer to an operational robot.

2.4 Maps and places

Aosh, Motomura, Hara, Akaho, Hayamizu and Matsui present a dialog-based learning system for map acquisition. A robot begins with a probabilistic map of an office environment and requests assistance when its believed position becomes too uncertain. Advice is given by human teachers in the form of spoken directions. The success of this enterprise depends on the development of an appropriate action space to match the typical types of dialog that will occur.

Koenig and Simmons are striving to provide the appropriate technology to allow office robots to be ready for immediate use at a customer's site. To do this the customer must simply supply the robot with a topological map of its environment; then the robot will passively learn the distances, sensor, and actuator models necessary for landmark recognition. Because the learning requires no supervision, the robot does not need an explicit training phase. Their experiments show that good models can be learned with only a small amount of experience.

Murphy and Schoppers describe research in progress on learning a set of landmarks suitable for place navigation in outer space. They begin by assuming that landmark extraction algorithms exist. Their interest is in how to select the best landmarks from a candidate set, where "best" depends on the different computational and energy costs associated with recognising each landmark using a variety of sensors.

Nakano, Ueda, Satto and Takahashi present a clustering technique for generating geometric wall maps from noisy, fragmentary sonar data. The method employs a neural network learning algorithm called vector quantisation and a model selection principle called minimum description length. Experiments using an actual robot showed that the method is successful when used in a simplified environment consisting only of uniform walls. They plan to extend the experiments to more natural environments.

Yamauchi and Langley are interested in approaches to robot localisation that can handle dynamic environments, where the dynamic change may be transient, such as a bike brought inside the office for the day, or long lasting, such as a new piece of furniture added to the office. Their approach integrates evidence grids and topological maps and has been successfully tested in a real-world office.

2.5 Concept grounding and acquisition

Hakura, Yokoi and Kakazu suggest a method for applying Gibson's affordance theory to the problem of robot perception. An internal representation field is proposed to allow the robot to abstract the real world into finite sets of patterns reconstructed in accordance with the sensory inputs. The inner model forms an afforrance-like concept to be used in action selection. The inner models are learned by a connectionist mechanism. It is hoped that these patterns can act as grounded symbols when coupled with the robots behaviours. Computer simulations are used to illustrate the feasibility of their method.

Klingspor proposes a distributed performance system for the real-time inference of high-level concepts from low-level sensor readings using multiple, special purpose inference engines. His system learns concepts and relations in Horn clauses and Prolog-like inference. He suggests a representation hierarchy that transforms raw sensor readings into time spanning basic features into sensor features into group features and finally into action-oriented features. As yet the real-world testing has been limited.

References

Hexmoor, H and Meeden, L, eds, 1996. *ROBOLEARN 96: An International Workshop on Learning for Autonomous Agents* SUNY at Buffalo, Technical Report 96–11, Computer Science, Buffalo, NY. (Information about the workshop and related discussions can be obtained at: http://www.cs.buffalo.edu/~hexmoor/robolearn96.html.)

Kaelbling, LP, 1993. Learning in Embedded Systems MIT Press.

Simon, H, 1983. "Why should machines learn?" In J Carbonell, R Michalski and T Mitchell, eds, *Machine Learning: An Artificial Intelligence Approach.* Tioga Press, CA.