Robot Learning

Jan Peters, Max Planck Institute for Biological Cybernetics
Russ Tedrake, Massachusetts Institute of Technology
Nick Roy, Massachusetts Institute of Technology
Jun Morimoto, Advanced Telecommunication Research Institute International ATR

Definition

Robot learning consists of a multitude of machine learning approaches, particularly **reinforcement learning**, **inverse reinforcement learning**, and **regression** methods, that have been adapted sufficiently to domain so that they allow learning in complex robot systems such as helicopters, flapping-wing flight, legged robots, anthropomorphic arms and humanoid robots. While classical artificial intelligence-based robotics approaches have often attempted to manually generate a set of rules and models that allows the robot systems to sense and act in the real-world, **robot learning** centers around the idea that it is unlikely that we can foresee all interesting real-world situations sufficiently accurate. Hence, the field of **robot learning** assumes that future robots need to be able to adapt to the real-world, and domain-appropriate machine learning might offer the most approach in this direction.

Robot Learning Systems

As learning has found many backdoor entrances to robotics, this section can only scratch the surface. However, robot learning has clearly been successful in several areas: (i) Model Learning, (ii) Imitation and Apprenticeship Learning, (iii) Reinforcement Learning as well as in various other topics.

Model Learning

Model learning is the machine learning counterpart to classical system identification [2, 6]. However, while the classical approaches heavily relies on the structure of physically-based models, specification of the relevant state variables and hand-tuned approximations of unknown nonlinearities, model learning approaches avoid many of these labor-intensive steps and the entire process to be more easily automated. Machine learning and system identification approaches often assume an observable state of the system to estimate the mapping from inputs to outputs of the system. However, a learning system is often able to learn this mapping including the statistics needed to cope with unidentified state variables and can hence cope with a larger class of systems. Two types of models are commonly learned, i.e., forward, and inverse models.

Forward models predict the behavior of the system based either on the current state or a history of preceeding observations. They can be viewed as "learned simulators" that may be used for optimizing a policy or for predicting future information. Examples of the application of such learned simulators range from the early work in the late 1980s by Atkeson & Schaal in robot arm-based cartpole swing-ups to Ng's recent extensions for stabilizing an inverted helicopter. Most forward models can directly be learned by **regression.**

Conversely, *inverse models* attempt to predict the input to a system in order to achieve a desired output in the next step, i.e., it uses the model of the system to directly generate control signals. In traditional control, these are often called approximation-based control systems [2]. Inverse model learning can be straightforwardly by **regression** when the system dynamics can be inverted uniquely, e.g., as in inverse dynamics learning for a fully actuated system. However, for underactuated or redundantly actuated systems [8], operational space control [3], etc., such a unique inverses do not exist and additional optimization is needed..

Imitation and Apprenticeship Learning

A key problem in robotics is to ease the problem of programming a complex behavior. Traditional robot programming approaches rely on accurate, manual, modeling of the task and removal of all uncertainities so that they work well. In contrast to classical robot programming, learning from demonstration approaches aim at recovering the instructions directly from a human demonstration. Numerous unsolved problems exist in this context such as discovering the intent of the teacher or determing the mapping from the teacher's kinematics to the robot's kinematics (often called the correspondence problem). Two different approaches are common in this area, i..e., direct imitation learning and apprenticeship learning.

In *imitation learning* [7], also known as **behavioral cloning**, the robot system directly estimates a policy from a teachers presentation, and, subsequently, the robot system reproduces the task using this policy. A key advantage of this approach is that it can often learn a task successfully from few demonstrations. In areas where human demonstrations are straightforward to obtain, e.g., for learning racket sports, manipulation, drumming on anthropomorphic systems, direct imitation learning often proved to be an appropriate approach. Its major shortcomings are that it cannot explain why the derived policy is a good one and it may struggle with learning from noisy demonstrations.

Hence, *apprenticeship learning* [1] has been proposed as an alternative where a reward function is used as explanation of the teachers' behaviour. Here, the reward function is chosen under which the teacher appears to act optimally, and the optimal policy for this reward function is subsequently computed as a solution. This approach transforms the problem of learning from demonstrations onto the harder problem of approximate optimal control or reinforcement learning, hence it is also known as inverse optimal control or **inverse reinforcement learning**. As a result, it is limited to problems that can be solved by current reinforcement learning methods. Additionally, it often has a hard time dealing with tasks where only few demonstrations with low variance exist.

Hence, inverse reinforcement learning has been particularly successful in areas where it is hard for a human to demonstrate the desired behavior such as for helicopter acrobatics or in robot locomotion.

Further information on learning by demonstration may be found in [1, 7].

Robot Reinforcement Learning

The ability to self-improve with respect to an arbitrary reward function, i.e., **reinforcement learning**, is essential for robot systems to become more autonomous. Here, the system learns about its policy by interacting with its environment and receiving scores (i.e, rewards or costs) for the quality of its performance. Unlike supervised learning approaches used in model learning or imitation learning, reinforcement learning can still be considered to be in its infancy. Few off-the-shelf reinforcement learning methods scale into the domain of robotics both in terms of dimensionality and the number of trials needed to obtain an interesting behavior. Three different but overlapping styles of reinforcement learning can be found in robotics, i.e., model-based reinforcement learning, **value function approximation** methods, and direct **policy search.**

Model-based reinforcement learning relies upon a learned forward model used for simulation-based optimization as discussed before. While often highly efficient, it frequently suffers from the fact that learned models are imperfect and, hence, the optimization method can be guaranteed to be biased by the errors in the model. To date, a full Bayesian treatment of the model uncertainty appears to be a promising way for alleviating this shortcoming of this otherwise powerful approach.

Value function approximation methods have been the core approach used in reinforcement learning during the 1990s. These techniques rely upon approximating the expected rewards for every possible action in every visited state. Subsequently, the controller chooses the actions in accordance to this value. Such approximation requires a globally consistent value function where the quality of the policy is determined by the largest error of the value function at any possible state. As a result, these methods have been problematic for anthropomorphic robotics as the high-dimensional domains often defy learning such a global construct. However, it has been highly sucessful in low-dimensional domains such as mobile vehicle control and robot soccer, as well as on well-understood test domains such as cart-pole systems.

Unlike the previous two approaches, *policy search* attempts to directly learn the optimal policy from experience without solving intermediary learning problems. Policies often have significantly fewer parameters than models or value functions. For example, for balancing a ball on a plate (where the plate is mounted on a robot end-effector) optimally with respect to a quadratic reward function, the number of policy parameters grows linearly in the number state dimensions while it grows quadratically for both model and value function for this analytically tractable problem (In general cases, the number of parameters of value functions grows exponentially in the number of states which is known as the 'Curse of Dimensionality'). This insight has given rise to policy search methods, particularly, **policy gradient methods** and probabilistic approaches to policy search such as the reward-weighted regression or PoWER. To date, application results of direct policy search approaches range from gait optimization in locomotion to various motor learning examples (e.g., Kendama, T-Ball or throwing darts).

Further information on reinforcement learning for robotics may be found in [9, 4, 5].

Application Domains

The possible application domains for robot learning have not been fully explored, one could even aggressively state that we have barely started to bring learning into robotics. Nevertheless, robot learning has been successful in several application domains.

For accurate execution of desired trajectories, model learning has scaled to learning the full inverse dynamics for a humanoid robot in real time more accurately than achievable with physical models. Current work focusses mainly on improving the concurrent execution of tasks as well as control of redundant or underactuated systems.

Various approaches have been successful in task learning. Learning by demonstration approaches are moving increasingly towards industrial grade solutions where fast training of complex tasks becomes possible. Skills ranging from motor toys, e.g., basic movements, paddling a ball, etc, to complex tasks such as cooking a complete meal, basic table tennis strokes, helicopter acrobatics or footplacement in locomotion have been learned from human teachers. Reinforcement learning has yielded better gaits in locomotion, jumping behaviors for legged robots, perching with fixed wing flight robots, forehands in table tennis as well as various applications to learning of motor toys.

See Also

reinforcement learning, **inverse reinforcement learning**, **behavioral cloning**, **policy search**, **value function approximation**

References and Recommended Reading

[1] A. Coates, P. Abbeel, and A. Y. Ng. Apprenticeship learning for helicopter control. *Communications of the ACM*, 52(7):97–105, 2009.

[2] J. A. Farrell and M. M. Polycarpou. *Adaptive Approximation Based Control.* Adaptive and Learning Systems for Signal Processing, Communications and Control Series. John Wiley, Hoboken, NJ, 2006.

[3] J. Peters and S. Schaal. Learning to control in operational space. *International Journal of Robotics Research*, 27:197–212, 2008.

[4] J. Peters and S. Schaal. Reinforcement learning of motor skills with policy gradients. *Neural Networks*, 21(4):682–97, 2008.

[5] M. Riedmiller, T. Gabel, R. Hafner, and S. Lange. Reinforcement learning for robot soccer. *Autonomous Robots*, 27(1):55–73, July 2009.

[6] S. Schaal, C. G. Atkeson, and S. Vijayakumar. Scalable techniques from nonparameteric statistics for real-time robot learning. *Applied Intelligence*, 17(1):49–60, 2002.

[7] S. Schaal, A. Ijspeert, and A. Billard. Computational approaches to motor learning by imitation. *Philosophical Transaction of the Royal Society of London: Series B, Biological Sciences*, 358(1431):537–547, 2003.

[8] R. Tedrake. Underactuated robotics: Learning, planning, and control for efficient and agile machines. Course Notes for MIT 6.832, MIT 32-380, 32 Vassar Street, Cambridge, MA 02139, USA, October 2009.

[9] R. Tedrake, T. W. Zhang, and H. S. Seung. Stochastic policy gradient reinforcement learning on a simple 3d biped. In *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*, pages 2849–2854, Sendai, Japan, September 2004.