Learning Robot Control

Martin Tschirsich
TU Darmstadt
Germany
m.tschirsich@gmx.de

1 Introduction

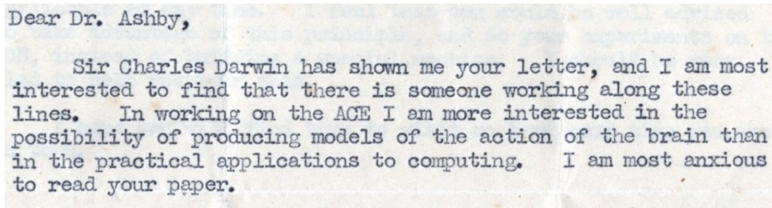
Robots are mechanical structures with sensors, effectors and control system. Coupling of sensory inputs to effector outputs requires a robotic control law. Unfortunately, the construction of such control laws is a complicated task. Automatisation of this process requires the incorporation of machine learning techniques. Robots capable of learning are not necessarily synonymous with intelligent or autonomous robots, since artificial intelligence can be achieved without learning. However, robot learning is seen as the key ingredient for the development of future autonomous robots [7].

Atkins classifies robot learning along three dimensions: *Direct versus indirect control*, *class of task* and *learning method*. In this paper, we follow the first dimension of *Direct versus indirect control* by differentiating between model-free and model based control and present the learning methods as they become relevant. But foremost, a historical overview provides a broad introduction into how the idea learning robots was developed and what problems had to be tackled to create the first learning robot in history.

2 History of Learning Robots

2.1 From ancient to modern times: literature and ideas

The intelligent creation and use of tools is what sets the human appart from most animals and is theorized to have stimulated further evolution of the human brain. Tools became more complex and evolved into machines, aiding the early human society in fulfilling its evolutionary purpose. Soon, this gave rise to the idea of intelligent machines: machines that behave like humans themselfes. The Chinese Book of Master Lie (500 BC) refers to an early humanoid: “The creation looked and moved so much like a human that when it winked at the concubines, it was necessary to dismantle it to prove that it was an artificial creation” [9]. However, no reference is made to the aspect of learning. One example from ancient western literature is Homer’s *The Iliad* (850 BC) which mentions intelligent servants made of gold, capable of learning: ”They look like living servant girls, possessing minds, hearts with intelligence, vocal chords, and strength. They learned to work from the immortal gods” [4]. Early attempts at the design of learning robots was focused at humanoids, where the human was seen as the archetypical example. With the age of enlightenment and the advance of knowledge through scientific methods, in 1748, french philosopher La Mettrie was one of the first to regard the human itself as nothing but a complex robot or “a collection of springs”. According to him, the human brain is nothing but the mainspring of the whole machine and “the soul is but a principle of motion”. He recognized that, while being most dependent of all the mammalst, the outstanding ability to learn from our environment through exploration and imitation is what sets us appart from lesser animals. He describes the process of learning as aquiring a cost for our actions through trial and error: “Light a wax candle for the first time under a child’s eyes, and he will mechanically put his fingers in the flame as if to find out what is the new thing that he sees. It is at his own cost that he will learn of the danger, but he will not be caught again. Thus nature made us to be lower than animals or at least to exhibit all the more, because of that native inferiority, the wonderful efficacy of education which alone raises us from the level of the animals and lifts us above them” [2].



Dear Dr. Ashby,

Sir Charles Darwin has shown me your letter, and I am most interested to find that there is someone working along these lines. In working on the ACE I am more interested in the possibility of producing models of the action of the brain than in the practical applications to computing. I am most anxious to read your paper.

Figure 2.1: Letter from Touring to Ashby concerning his Homeostat.

Figure 2.2: Grey Walter with one of his tortoises. The existence of a tortoise equipped with the learning system CORA is controversial.

With the advances in mechanical and electrical engineering exploration of electricity from the late 19th century on, theories of robots and learning became applicable to real world machines. Nicola Tesla was one of the first to physically separate the control of a machine from its mechanics. In 1898, Nicola Tesla demonstrated a radio-controlled boat, which he described as incorporating "a borrowed mind."

With the introduction of computing devices during the second world war, a borrowed human mind was no longer a necessity for learning robots. Using RAF bomb control units, Ross Ashby designed and built the *Homeostat*, a machine that is capable of adapting itself to its environment. Ashby demonstrated that his machine could exhibit basic reinforcement learning: by a process of trial and error, the machine is able to maintain its essential variables within specified limits [3].

In 1946, Alan Touring was made aware of Ashby's plan to build such a machine by Charles Darwin. He contacted Ashby and proposed to use his Automatic Computing Engine for his experiments instead of building a concrete machine, thus proposing for the first time the computer simulation of a learning robot.

In 1950, Grey Walter, a pionier of artificial life and known for his robotic tortoises, developed an electronic learning circuit named CORA. If CORA was repeatedly presented with two stimuli in quick succession, only the second of which 'naturally' caused a particular response, it would eventually produce that response even when the first stimulus occurred alone. The ultimate goal of Walter was to explain control principles in both animals and machines using mathematics, and to express natural behavior in form of feedback control systems. He envisaged to equip one of his mobile robot tortoises with CORA. In doing so, he designed the first mobile robot capable of learning.

3 Learning Robots from a modern standpoint

Often, the problem of learning robot control is adressed within the framework of optimal control. Optimal control is an optimization method for deriving control laws. It boils down to finding a time dependend optimal control $\mathbf{u}$ in order to cause the system (model equation)

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, t)$$

to follow a trajectory which minimizes a cost function. We seek the optimal control law or *policy* to be a function

$$\mathbf{u} = \pi(\mathbf{x}, t, \theta)$$

depending on the current state, time and problem-specific parameter vector θ. Dynamic programming, intdoduced in the 1940s by Richard Bellman, can be used to solve this optimization problem by breaking it down into overlapping smaller subproblems. Unfortunately, optimal control relies on an exact model and can be solved exactly for only two cases, linear control or control within a discrete state space. However, as the state space for a learning system becomes high dimensional, its combinational state space exponentially explodes. Thus, finding an optimal control policy is not always feasible, but global optimality is not always required. Learning algorithms can be employed to derive a reasonably good policy.

3.1 Learning a model for control

The environment a robot operates in, the objects a robot interacts with as well as the robot itself, i.e. its mechanical system, the kinematics and dynamics can be modelled. It has initially been thought that learning control e.g. by reinforcement learning reduces the need for an accurate model of the mechanical system to be controlled. However, it became apparent that the existence of accurate models directly impacts intitial performance and learning efficiency. Better models lead to faster command error correction while reducing the amount of practice needed to attain a given level of performance [1].

Often however, analytical models are not sufficient or accurate enough or too costly to derive. Rigid body dynamics, albeit well understood, are not expressive enough for most real world applications. Therefore, the model is often learned. Learned models also remove the necessity for building robotswith the focus on being straightforward to model. They can be chosen to fulfill the tasks requirements in terms of compliance with the environment, energy efficiency, and other factors [6]. Online learning can be employed to generalize learned models to gradually adapt for mechanical wear and tear and larger state space.

Models can be categorized in four types [5]: Forward, inverse, mixed and multistep models. The Smith predictor (1957) is an example for the early application of forward model. A further differentiation is made between parametric and non-parametric models. While parameter learning for parametric models can lead to physically inconsistent parameters or persistent excitation issues, non-parametric models learned with nonparametric machine learning techniques such as nonparametric regression might be overly expressive [8].

3.2 Model-based learning

Once we have infered the behavior of the system either analytically or by learning from observations, we must determine how to optimally manipulate the system. While the first part is a pure modelling problem, the second question is related to the learning control architectures which can be used together with the model. Given a perfect model, it is possible to solve the optimal control problem exactly for two limited cases using recursive algorithms such as value iteration (1957) or policy iteration (1960). The limitations are severe: the model may only depend linearly on the state and action with quadratic rewards or the model must be discrete in states and actions. While the later limitation seems less severe, such models are often impractical due to the curse of dimensionality in robotics (e.g. supercomputer could handle ten-dimensional problems in 2010 [7]).

3.3 Model-free learning

In order to solve more practically relevant optimal control problems, value function methods that find an approximative solution for the optimal control problem can be employed. Value function methods approximate the value function by observing the reward for different states and actions using approximations of the Monte-Carlo Method. These methods where widely popular in the 1990s, among them Q-Learning and temporal difference learning. The biggest drawback lies within the need for sufficient training or sample data in order to generalize the value function to a sufficient state space. Though we no longer need to sample the full scale space, deriving a sufficiently good policy requires samples from a wide range of possible states, again causing a cominatorial state space explosion.

To circumvent the curse of dimensionality, other approaces where proposed. A powerful alternative to value functions is policy search, e.g. directly optimizing for a policy by gradient descend. These policy gradient methods might terminate within a locally optimal solution, however many approaches exists to adress this problem such as natural policy gradients. Albeit powerful, modern alternatives to policy gradients have been developed that are probabilistic. Many probablistic inference algorithms can be used in continuous spaces such as mixture of gaussians or gaussian processes.

References

[1] Christopher G. Atkeson, Eric W. Aboaf, Joseph McIntyre, and David J. Reinkensmeyer. Model-based robot learning. In *FOURTH INTL. SYMPOSIUM ON ROBOTICS RESEARCH*, pages 9–14, 1988.

[2] Julien Offray de La Mettrie. *Man a Machine*. Open Court Publishing Company, Chicago, 1974.

[3] Alice Eldridge. *Collaborating with the behaving machine: simple adaptive dynamical systems for generative and interactive music*. PhD thesis, University of Sussex, 2007.

[4] Richmond Lattimore. *The Iliad of Homer. Translated by R. Lattimore*. University Of Chicago Press, Chicago, 1951.

[5] Duy Nguyen-Tuong and Jan Peters. Model learning for robot control: a survey. *Cognitive processing*, 12(4):319–340, 2011.

[6] Jan Reinhard Peters. *Machine learning of motor skills for robotics*. PhD thesis, University of Southern California, 2007.

[7] Stefan Schaal and C Atkeson. Learning control in robotics. *Robotics & Automation Magazine, IEEE*, 17(2):20–29, 2010.

[8] Stefan Schaal and Christopher G Atkeson. Robot learning by nonparametric regression. In *Intelligent Robots and Systems' 94.'Advanced Robotic Systems and the Real World', IROS'94. Proceedings of the IEEE/RSJ/GI International Conference on*, volume 1, pages 478–485. IEEE, 1994.

[9] Lie Youkou. Book of master lie, 500 BC.