

Causal Models over Infinite Graphs and their Application to the Sensorimotor Loop - General Stochastic Aspects and Gradient Methods for Optimal Control

Motivation and background

The enormous amount of capabilities that every human learns throughout his life, is probably among the most remarkable and fascinating aspects of life. Learning has therefore drawn lots of interest from scientists working in very different fields like philosophy, biology, sociology, educational sciences, computer sciences and mathematics. This thesis focuses on the information theoretical and mathematical aspects of learning.

We are interested in the learning process of an agent (which can be for example a human, an animal, a robot, an economical institution or a state) that interacts with its environment. Common models for this interaction are Markov decision processes (MDPs) and partially observable Markov decision processes (POMDPs). Learning is then considered to be the maximization of the expectation of a predefined reward function. In order to formulate general principles (like a formal definition of curiosity-driven learning or avoidance of unpleasant situation) in a rigorous way, it might be desirable to have a theoretical framework for the optimization of more complex functionals of the underlying process law. This might include the entropy of certain sensor values or their mutual information. An optimization of the latter quantity (also known as predictive information) has been investigated intensively both theoretically and experimentally using computer simulations by N. Ay, R. Der, K Zahedi and G. Martius. In this thesis, we develop a mathematical theory for learning in the sensorimotor loop beyond expected reward maximization.

Approaches and results

This thesis covers four different topics related to the theory of learning in the sensorimotor loop.

First of all, we need to specify the model of an agent interacting with the environment, either with learning or without learning. This interaction naturally results in complex causal dependencies. Since we are interested in asymptotic properties of learning algorithms, it is necessary to consider infinite time horizons. It turns out that the well-understood theory of causal networks known from the machine learning literature is not powerful enough for our purpose. Therefore we extend important theorems on causal networks to infinite graphs and general state spaces using analytical methods from measure theoretic probability theory and the theory of discrete time stochastic processes. Furthermore, we prove a generalization of the strong Markov property from Markov processes to infinite causal networks.

Secondly, we develop a new idea for a projected stochastic constraint optimization algorithm. Generally a discrete gradient ascent algorithm can be used to generate an iterative sequence that converges to the stationary points of a given optimization problem. Whenever the optimization takes place over a compact subset of a vector space, it is possible that the iterative sequence leaves the constraint set. One possibility to cope with this problem is to project all points to the constraint set using Euclidean best-approximation. The latter is sometimes difficult to calculate. A concrete example is an optimization over the unit ball in a matrix space equipped with operator norm. Our idea consists of a back-projection using quasi-projectors different from the Euclidean best-approximation. In the matrix example, there is another canonical way to force the iterative sequence to stay in the constraint set: Whenever a point leaves the unit ball, it is divided by its norm. For a given target function, this procedure might introduce spurious stationary points on the boundary. We show that this problem can be circumvented by using a gradient that is tailored to the quasi-projector used for back-projection. We state a general technical compatibility condition between a quasi-projector and a metric used for gradient ascent, prove convergence of stochastic iterative sequences and provide an appropriate metric for the unit-ball example.

Thirdly, a class of learning problems in the sensorimotor loop is defined and motivated. This class of problems is more general than the usual expected reward maximization and is illustrated by numerous examples (like expected reward maximization, maximization of the predictive information, maximization of the entropy and minimization of the variance of a given reward function). We also provide stationarity conditions together with appropriate gradient formulas.

Last but not least, we prove convergence of a stochastic optimization algorithm (as considered in the second topic) applied to a general learning problem (as considered in the third topic). It is shown that the learning algorithm converges to the set of stationary points. Among others, the proof covers the convergence of an improved version of an algorithm for the maximization of the predictive information as proposed by N. Ay, R. Der and K. Zehedi. We also investigate an application to a linear Gaussian dynamic, where the policies are encoded by the unit-ball in a space of matrices equipped with operator norm.