Approximate Optimal control with Markov Decision Processes for Autonomous artificial intelligence

Bruno Scherrer and Leslie Pack Kaelbling

What: My research aims at building computer systems that have human-like capabilities. Specifically, I am trying to design autonomous agents, that is systems that perceive their environment, learn about it, and interact with it in an autonomous way. My work focuses on theoretical analyses of the reliability of such systems under reasonable computing constraints.

Why: My main motivation is to identify and understand the fundamental principles of what is commonly referred as "intelligence". How can a computer act without human intervention ? How can learning, planning, acting and thinking be formalized in a way that might be useful for a computer ? To which extent can these formalizations be applied to real world application like robotics, telecommunication networks or game theory ?

How: My fundamental research assumption is related to my belief that the approach of artificial intelligence as an optimal control problem, and in particular its formalizatoin as a Markov decision process, should play a more and more essential role in this domain.

A Markov decision process (MDP) is a simple but non-trivial framework which enables us to model an agent interacting with an environment and seeking to reach a certain number of goals. This specific approach is particularly appealing because it is completely generic (it can model a robot, a chess player, a computer network, or even an ant colony). Also, it seems to be a fundamental model in the sense that every time one tries to extend/generalize it, as for instance when considering partially observable Markov decision processes (POMDP), the eventual analysis relies on the construction of a new (structured) MDP.

In this framework, my current research focuses on the complexity/quality tradeoff that has to be considered in hard control problems. A control problem might be hard for various reasons: for instance, it could be in a large state space or, the information the agent has access to might be very little (this is for instance the case in a POMDP). I am trying to design theoretically sound algorithms for approximation in such hard MDPs. By sound I mean that I seek to derive an approximation analysis that provides error bounds and possible convergence properties. To specifically tackle the complexity/quality tradeoff, I am also considering the process of iteratively refining an approximation under resource constraints.

Progress: During my Phd, I analysed a specific approximation scheme for large state space MDPs. Since I arrived at MIT as a postdoc, I have been working on improving its approximation analysis: I have been able to tighten the bounds and to give some sound justification of heuristics proposed in my Phd. I am also applying it to more sample problems.

Recently, I have been considering a variant of optimal control, called reinforcement learning, where the agent has little prior information about its interaction with the environment, and therefore has to learn (through sampling) and plan at the same time. I have derived a new high probability error analysis which relates the amount of sampling with a confidence bound on the solution computed so far. Also, and more importantly, I have (theoretically and experimentally) shown that this analysis enables the agent to choose where to sample so that the approximate solution tends more quickly to the optimal solution.

Future: I plan to continue analysing the complexity/quality trade-off for optimal control. This amounts to improving the analysis of my previous and ongoing works and also considering structured subclasses of MDPs. Among these classes, POMDPs, which makes more realistic assumptions about a general control problem, are particulary appealing. Though it is a much harder problem than MDPs, I believe that the specific structure of its solution (its value function is continuous and convex) might be particularly suited for approximations.

Research Support: This research is supported by a postdoc grant from the Institut de Recherche en Informatique et ses Applications.