

Guest editorial: Special issue on robot learning, Part B

Jan Peters · Andrew Y. Ng

Received: 22 July 2009 / Accepted: 23 July 2009 / Published online: 1 August 2009
© Springer Science+Business Media, LLC 2009

1 Introduction

In this Part B of the Autonomous Robots Special Issue on Robot Learning, we complement the preceding Part A. Similar to Part A, we intend to highlight recent successes in the application of domain-driven machine learning methods to robotics. However, the papers selected for Part B have a slightly different focus than in Part A as they focus on more specialized approaches for a variety of different problems while showing the feasibility of more general methods. Part B consists of a representative sample of interesting recent results and important current work in this area. The papers were solicited from the same open call as that of Part A. All papers were rigorously reviewed by up to eight reviewers as well as by at least one of the guest editors—on average, all papers had 4.5 external reviewers. As a result, a small sample of eight papers has been selected for both Part A and B of the special issue. The second four papers form Part B of the special issue and are described next.

2 Guide to Part B of the Special issue

In the first paper on *A Bayesian Exploration-Exploitation Approach for Optimal Online Sensing and Planning with*

a Visually Guided Mobile Robot by Martinez-Cantin, de Freitas, Brochu, Castellanos and Doucet, the authors address the policy search problem in the context of online optimal path planning under imperfect sensing. The authors model the problem as a partially observable Markov decision process (POMDP) with a finite horizon. This problem is difficult due to its real-world setting (dimensionality, continuous actions, non-Gaussian noise, real-time requirements). The authors design a novel Bayesian approach that minimizes the uncertainty about the policy's performance while aiming towards the optimal policy. The authors demonstrate the feasibility of their approach on a visually-guided mobile robot.

Howard, Klanke, Gienger, Goerick and Vijayakumar present an article entitled *A Novel Method for Learning Policies from Variable Constraint Data*. They note that most real world tasks are influenced by external constraints that might be often not directly observable or change of the constraints between the observations may happen. The authors tackle this problem by first addressing the problems of learning uncontrolled policies from observed movement data that has been subjected to various constraints. By introducing a small but significant modification to the risk functional of an unconstrained policy learning step, they are able to determine policies that fulfill the constraints. The authors present results of their method on an Honda ASIMO robot.

The article *Learning Model-free Robot Control by a Monte Carlo EM Algorithm* by Vlassis, Toussaint, Kontes and Piperidis extends previous work by Kober and Peters on episodic model-free reinforcement learning (RL) to the discounted infinite horizon case. For doing so, they adopt Toussaint's previous insights from model-based RL and suggest a Monte Carlo EM approach for searching the optimal policy directly in the parameter space of the policy. Their algorithm can be seen as a generalization of Kober and Peters

J. Peters (✉)
Department of Empirical Inference, Max Planck Institute
for Biological Cybernetics, Spemannstrasse 38, 72076 Tübingen,
Germany
e-mail: mail@jan-peters.net

A.Y. Ng
Department of Computer Science, Stanford University, Room
156, Gates Building 1A, Stanford, CA 94305-9010, USA
e-mail: ang@cs.stanford.edu

PoWER algorithm and, in the finite case, reduces to exactly the same equations. They achieve this result by viewing the discounted infinite horizon case as a finite case where the length of the episode is chosen randomly from a geometric distribution probability distribution. They demonstrate the feasibility of the approach on a self-balancing two-wheeled mobile robot.

In the final paper, *Nonparametric Representation of An Approximated Poincaré Map for Learning Biped Locomotion*, Morimoto, Hyon and Atkeson present their work on learning Poincaré maps that allow them to generate novel walking and stepping policies. For doing so, the authors suggest representing the Poincaré map non-parametrically using a Gaussian process model. This move allows them to sidestep the problem of fitting a fixed analytical model to

measured data and to directly incorporate real world data from a Sony QRIO robot. Subsequently, they locally optimize policies for stepping and walking using a policy gradient approach to reinforcement learning. The resulting approach allows them to learn policies that work well both in simulation and on real robots.

All four papers present significant developments in policy learning for robotics and we hope that you will enjoy reading them as much as we did.

Acknowledgements The Guest Editors wish to express their gratitude to Gaurav Sukhatme, the AUTONOMOUS ROBOTS Editor-in-Chief, as well as Editorial Assistant Valerie Schofield and JEO assistant Rolan Profugo, for their constant support. We are also enormously grateful to the more than a hundred colleagues who agreed to provide reviews for the 46 papers submitted to this Special Issue.