Studies in Computational Intelligence

Volume 503

Series Editor

J. Kacprzyk, Warsaw, Poland

For further volumes:
http://www.springer.com/series/7092

Todd Hester

TEXPLORE: Temporal Difference Reinforcement Learning for Robots and Time-Constrained Domains

Springer

Todd Hester
Department of Computer Science
University of Texas at Austin
Austin, Texas
USA

ISSN 1860-949X ISSN 1860-9503 (electronic)
ISBN 978-3-319-01167-7 ISBN 978-3-319-01168-4 (eBook)
DOI 10.1007/978-3-319-01168-4
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013942265

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To my family, especially my lovely wife Beth

Foreword

It is a great pleasure and honor to be able to write the foreward for this book, representing the culmination of Todd Hester's Ph.D. thesis research at The University of Texas at Austin. It was my good fortune to be Todd's advisor during his years as a Computer Science graduate student. I therefore was able to participate in and enjoy the adventure of starting with the kernel of an idea and fully developing it into a full-blown dissertation.

Todd's research is in the area of Reinforcement learning (RL), a machine learning paradigm that focuses on enabling computers and robots to learn to perform sequential tasks. Though grounded in some very theoretically elegant results, and showing great promise for enabling learning-based robots that could be deployed in the real world, most RL algorithms to date require either too much data (training experience) or too much computation to be practical on real-world problems.

This book introduces TEXPLORE, one of the first RL algorithms to be both data-efficient and computation-efficient enough to work on real robots in the real world. The core idea of TEXPLORE is the realization that scaling up to large domains in real time requires actively reasoning about which states *not* to explore. Most RL algorithms to date still insist on exhaustive exploration: visiting every state, or in continuous settings, every region of the state space. Doing so is necessary if the goal is finding the *optimal* policy because in principle, any unvisited state could be a "gold mine." However when accepting that perfection can sometimes be the enemy of the good, it becomes clear that exploration must be more focused.

The distinguishing characteristic of TEXPLORE is that it learns *in real-time, while continuing to act.* As this book documents fully, it has been demonstrated to learn a speed-control task on a real autonomous vehicle during the course of two minutes of continual driving. In addition to this driving task, the algorithms introduced in this book have been validated on real humanoid robots, and in carefully controlled, simulation environments.

This book also includes an investigation into the application of TEXPLORE to the idea of intrinsically motivated RL. Analogous to curiosity on the part of human learners, intrinsically motivated RL requires guiding exploration by properties of the environment, rather than based on external reward.

This book is important for the field for these novel algorithms themselves, and also for the fact that it opens up several exciting directions for future research. By releasing the associated source code as an RL package within the Robot

Operating System (ROS) development environment, Todd has made it easy for future researchers to build upon his contributions.

Overall, for both newcomers to the field, and for practitioners looking for nuanced detail, this book has plenty to offer. Whichever your perspective, I trust you will enjoy reading it!

Austin, Texas Peter Stone
June, 2013

Preface

This book presents the main results of the research I conducted for my Ph.D. thesis at the University of Texas at Austin. The main focus of the research is on developing new reinforcement learning methods that enable fast and robust learning on robots in real-time.

Robots have the potential to solve many problems in society, because of their ability to work in dangerous places doing necessary jobs that no one wants or is able to do. One barrier to their widespread deployment is that they are mainly limited to tasks where it is possible to hand-program behaviors for every situation that may be encountered. For robots to meet their potential, they need methods that enable them to learn and adapt to novel situations that they were not programmed for. Reinforcement learning (RL) is a paradigm for learning sequential decision making processes and could solve the problems of learning and adaptation on robots. This book identifies four key challenges that must be addressed for an RL algorithm to be practical for robotic control tasks. These *RL for Robotics Challenges* are: 1) it must learn in very few samples; 2) it must learn in domains with continuous state features; 3) it must handle sensor and/or actuator delays; and 4) it should continually select actions in real time. This book focuses on addressing all four of these challenges. In particular, this book is focused on *time-constrained domains* where the first challenge is critically important. In these domains, the agent's lifetime is not long enough for it to explore the domain thoroughly, and it must learn in very few samples.

Although existing RL algorithms successfully address one or more of the *RL for Robotics Challenges*, no prior algorithm addresses all four of them. To fill this gap, this book introduces TEXPLORE, the first algorithm to address all four challenges. TEXPLORE is a model-based RL method that learns a random forest model of the domain which generalizes dynamics to unseen states. Each tree in the random forest model represents a hypothesis of the domain's true dynamics, and the agent uses these hypotheses to explores states that are promising for the final policy, while ignoring states that do not appear promising. With sample-based planning and a novel parallel architecture, TEXPLORE can select actions continually in real time whenever necessary.

We empirically evaluate each component of TEXPLORE in comparison with other state-of-the-art approaches. In addition, we present modifications of TEXPLORE's exploration mechanism for different types of domains. The key result of this book is a demonstration of TEXPLORE learning to control the velocity of an autonomous vehicle on-line, in real time, while running on-board the robot. After

controlling the vehicle for only two minutes, TEXPLORE is able to learn to move the pedals of the vehicle to drive at the desired velocities. The work presented in this book represents an important step towards applying RL to robotics and enabling robots to perform more tasks in society. By enabling robots to learn in few actions while acting on-line in real time on robots with continuous state and actuator delays, TEXPLORE significantly broadens the applicability of RL to robots.

This book would not have been possible without help from a great number of people. First and foremost, I want to thank my advisor Peter Stone, whose guidance, advice, and support has been invaluable. There are also many other graduate students who helped and collaborated with me on this research. In particular, Nick Jong let me assist on an AAMAS paper on reinforcement learning in my first year as a graduate student and gave me a great start on RL.

Austin, Texas Todd Hester
April, 2013

Table of Contents