Marco Wiering and Martijn van Otterlo (Eds.)

Reinforcement Learning

Adaptation, Learning, and Optimization, Volume 12

Series Editor-in-Chief

Meng-Hiot Lim
Nanyang Technological University, Singapore
E-mail: emhlim@ntu.edu.sg

Yew-Soon Ong
Nanyang Technological University, Singapore
E-mail: asysong@ntu.edu.sg

Further volumes of this series can be found on our homepage: springer.com

Vol. 1. Jingqiao Zhang and Arthur C. Sanderson
Adaptive Differential Evolution, 2009
ISBN 978-3-642-01526-7

Vol. 2. Yoel Tenne and Chi-Keong Goh (Eds.)
Computational Intelligence in
Expensive Optimization Problems, 2010
ISBN 978-3-642-10700-9

Vol. 3. Ying-ping Chen (Ed.)
Exploitation of Linkage Learning in Evolutionary Algorithms, 2010
ISBN 978-3-642-12833-2

Vol. 4. Anyong Qing and Ching Kwang Lee
Differential Evolution in Electromagnetics, 2010
ISBN 978-3-642-12868-4

Vol. 5. Ruhul A. Sarker and Tapabrata Ray (Eds.)
Agent-Based Evolutionary Search, 2010
ISBN 978-3-642-13424-1

Vol. 6. John Seiffertt and Donald C. Wunsch
Unified Computational Intelligence for Complex Systems, 2010
ISBN 978-3-642-03179-3

Vol. 7. Yoel Tenne and Chi-Keong Goh (Eds.)
Computational Intelligence in Optimization, 2010
ISBN 978-3-642-12774-8

Vol. 8. Bijaya Ketan Panigrahi, Yuhui Shi, and Meng-Hiot Lim (Eds.)
Handbook of Swarm Intelligence, 2011
ISBN 978-3-642-17389-9

Vol. 9. Lijuan Li and Feng Liu
Group Search Optimization for Applications in Structural Design, 2011
ISBN 978-3-642-20535-4

Vol. 10. Jeffrey W. Tweedale and Lakhmi C. Jain
Embedded Automation in Human-Agent Environment, 2011
ISBN 978-3-642-22675-5

Vol. 11. Hitoshi Iba and Claus C. Aranha
Practical Applications of Evolutionary Computation to Financial Engineering, 2012
ISBN 978-3-642-27647-7

Vol. 12. Marco Wiering and Martijn van Otterlo (Eds.)
Reinforcement Learning, 2012
ISBN 978-3-642-27644-6

Marco Wiering and Martijn van Otterlo (Eds.)

Reinforcement Learning

State-of-the-Art

Springer

Editors
Dr. Marco Wiering
University of Groningen
The Netherlands

Dr. ir. Martijn van Otterlo
Radboud University Nijmegen
The Netherlands

ISSN 1867-4534
e-ISSN 1867-4542
ISBN 978-3-642-27644-6
e-ISBN 978-3-642-27645-3
DOI 10.1007/978-3-642-27645-3
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2011945323

Printed on acid-free paper

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

‘Good and evil, reward and punishment, are the only motives to a rational creature: these are the spur and reins whereby all mankind are set on work, and guided.’ (Locke)

Foreword

Reinforcement learning has been a subject of study for over fifty years, but its modern form—highly influenced by the theory of Markov decision processes—emerged in the 1980s and became fully established in textbook treatments in the latter half of the 1990s. In *Reinforcement Learning: State-of-the-Art*, Martijn van Otterlo and Marco Wiering, two respected and active researchers in the field, have commissioned and collected a series of eighteen articles describing almost all the major developments in reinforcement learning research since the start of the new millennium. The articles are surveys rather than novel contributions. Each authoritatively treats an important area of Reinforcement Learning, broadly conceived as including its neural and behavioral aspects as well as the computational considerations that have been the main focus. This book is a valuable resource for students wanting to go beyond the older textbooks and for researchers wanting to easily catch up with recent developments.

As someone who has worked in the field for a long time, two things stand out for me regarding the authors of the articles. The first is their youth. Of the eighteen articles, sixteen have as their first author someone who received their PhD within the last seven years (or who is still a student). This is surely an excellent sign for the vitality and renewal of the field. The second is that two-thirds of the authors hail from Europe. This is only partly due to the editors being from there; it seems to reflect a real shift eastward in the center of mass of reinforcement learning research, from North America toward Europe. Vive le temps et les différences!

October 2011 Richard S. Sutton

Preface

A question that pops up quite often among reinforcement learning researchers is on what one should recommend if a student or a colleague asks for

"some good and recent book that can introduce me to reinforcement learning".

The most important goal in creating this book was to provide at least a good answer to that question.

A Book about Reinforcement Learning

A decade ago the answer to our leading question would be quite easy to give; around that time two dominant books existed that were fully up-to-date. One is the excellent introduction[1] to reinforcement learning by Rich Sutton and Andy Barto from 1998. This book is written from an *artificial intelligence* perspective, has a great educational writing style and is widely used (around ten thousand citations at the time of writing). The other book was written by Dimitri Bertsekas and John Tsitsiklis in 1996 and was titled *neuro-dynamic programming*[2]. Written from the standpoint of *operations research*, the book rigorously and in a mathematically precise way describes dynamic programming and reinforcement learning with a particular emphasis on approximation architectures. Whereas Sutton and Barto always maximize rewards, talk about *value functions*, *rewards* and are biased to the $\{V,Q,S,A,T,R\}$ part of the alphabet augmented with π, Bertsekas and Tsitsiklis talk about *cost-to-go-functions*, always minimize costs, and settle on the $\{J,G,I,U\}$ part of the alphabet augmented with the greek symbol μ. Despite these superficial (notation) differences, the distinct writing styles and backgrounds, and probably also the audience for which these books were written, both tried to give a thorough introduction

[1] Sutton and Barto, (1998) Reinforcement Learning: An Introduction, MIT Press.

[2] Bertsekas and Tsitsiklis (1996) *Neuro-Dynamic Programming*, Athena Scientific.

to this exciting new research field and succeeded in doing that. At that time, the big merge of insights in both operations research and artificial intelligence approaches to behavior optimization was still ongoing and many fruitful cross-fertilization happened. Powerful ideas and algorithms such as Q-learning and TD-learning had been introduced quite recently and so many things were still unknown.

For example, questions about *convergence* of combinations of algorithms and function approximators arose. Many theoretical and experimental questions about convergence of algorithms, numbers of required samples for guaranteed performance, and applicability of reinforcement learning techniques in larger intelligent architectures were largely unanswered. In fact, many new issues came up and introduced an ever increasing pile of research questions waiting to be answered by bright, young PhD students. And even though both Sutton & Barto and Bertsekas & Tsitsiklis were excellent at introducing the field and eloquently describing the underlying methodologies and issues of it, at some point the field grew so large that new texts were required to capture all the latest developments. Hence this book, as an attempt to fill the gap.

This book is the first book about reinforcement learning featuring only state-of-the-art surveys on the main subareas. However, we can mention several other interesting books that introduce or describe various reinforcement learning topics too. These include a collection[3] edited by Leslie Kaelbling in 1996 and a new edition of the famous Markov decision process handbook[4] by Puterman. Several other books[5,6] deal with the related notion of *approximate dynamic programming*. Recently additional books have appeared on Markov decision processes[7], reinforcement learning[8], function approximation[9] and relational knowledge representation for reinforcement learning[10]. These books just represent a sample of a larger number of books relevant for those interested in reinforcement learning of course.

---

[3] L.P. Kaelbling (ed.) (1996) Recent Advances in Reinforcement Learning, Springer.

[4] M.L. Puterman (1994, 2005) Markov Decision Processes: Discrete Stochastic Dynamic Programming, Wiley.

[5] J. Si, A.G. Barto, W.B. Powell and D. Wunsch (eds.) (2004) Handbook of Learning and Approximate Dynamic Programming, IEEE Press.

[6] W.B. Powell (2011) Approximate Dynamic Programming: Solving the Curses of Dimensionality, 2nd Edition, Wiley.

[7] O. Sigaud and O. Buffet (eds.) (2010) Markov Decision Processes in Artificial Intelligence, Wiley-ISTE.

[8] C. Szepesvari (2010) Algorithms for Reinforcement Learning, Morgan-Claypool.

[9] L. Busoniu, R. Babuska, B. De Schutter and D. Ernst (2010) Reinforcement Learning and Dynamic Programming Using Function Approximators, CRC Press.

[10] M. van Otterlo (2009) The Logic Of Adaptive Behavior, IOS Press.

Reinforcement Learning: A Field Becoming Mature

In the past one and a half decade, the field of reinforcement learning has grown tremendously. New insights from this recent period – having much to deal with richer, and firmer, theory, increased applicability, scaling up, and connections to (probabilistic) artificial intelligence, brain theory and general adaptive systems – are not reflected in any recent book. Richard Sutton, one of the founders of modern reinforcement learning described[11] in 1999 three distinct areas in the development of reinforcement learning; *past*, *present* and *future*.

The RL *past* encompasses the period until approximately 1985 in which the idea of *trial-and-error* learning was developed. This period emphasized the use of an active, exploring agent and developed the key insight of using a scalar reward signal to specify the *goal* of the agent, termed the *reward hypothesis*. The methods usually only learned policies and were generally incapable of dealing effectively with delayed rewards.

The RL *present* was the period in which *value functions* were formalized. Value functions are at the heart of reinforcement learning and virtually all methods focus on approximations of value functions in order to compute (optimal) policies. The *value function hypothesis* says that approximation of value functions is the dominant purpose of intelligence.

At this moment, we are well underway in the reinforcement learning *future*. Sutton made predictions about the direction of this period and wrote *"Just as reinforcement learning present took a step away from the ultimate goal of reward to focus on value functions, so reinforcement learning future may take a further step away to focus on the structures that enable value function estimation [...] In psychology, the idea of a developing mind actively creating its representations of the world is called* ***constructivism****. My prediction is that for the next tens of years reinforcement learning will be focused on constructivism."* Indeed, as we can see in this book, many new developments in the field have to do with new structures that enable value function approximation. In addition, many developments are about *properties*, *capabilities* and *guarantees* about convergence and performance of these new structures. Bayesian frameworks, efficient linear approximations, relational knowledge representation and decompositions of hierarchical and multi-agent nature all constitute new structures employed in the reinforcement learning methodology nowadays.

Reinforcement learning is currently an established field usually situated in machine learning. However, given its focus on behavior learning, it has many connections to other fields such as psychology, operations research, mathematical optimization and beyond. Within artificial intelligence, there are large overlaps with probabilistic and decision-theoretic planning as it shares many goals with the planning community (e.g. the international conference on automated planning systems, ICAPS). In very recent editions of the international planning competition (IPC), methods originating from the reinforcement learning literature have entered the

[11] R.S. Sutton (1999) Reinforcement Learning: Past, Present, Future – SEAL'98.

competitions and did very well, in both probabilistic planning problems, and a recent ”learning for planning” track.

Reinforcement learning research is published virtually everywhere in the broad field of artificial intelligence, simply because it is both a general methodology for behavior optimization as well as a set of computational tools to do so. All major artificial intelligence journals feature articles on reinforcement learning nowadays, and have been doing so for a long time. Application domains range from robotics and computer games to network routing and natural language dialogue systems and reinforcement learning papers appear at fora dealing with these topics. A large portion of papers appears every year (or two year) at the established top conferences in artificial intelligence such as IJCAI, ECAI and AAAI, and many also at top conferences with a particular focus on statistical machine learning such as UAI, ICML, ECML and NIPS. In addition, conferences on artificial life (Alife), adaptive behavior (SAB), robotics (ICRA, IROS, RSS) and neural networks and evolutionary computation (e.g. IJCNN and ICANN) feature much reinforcement learning work. Last but not least, in the past decade many specialized reinforcement learning workshops and tutorials have appeared at all the major artificial intelligence conferences.

But even though the field has much to offer to many other fields, and reinforcement learning papers appear everywhere, the current status of the field renders it natural to introduce fora with a specific focus on reinforcement learning methods. The *European workshop on reinforcement learning* (EWRL) has gradually become one such forum, growing every two years considerably and most recently held in Nancy (2008) and co-located with ECML (2011). Furthermore, the *IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning* (ADPRL) has become yet another meeting point for researchers to present and discuss their latest research findings. Together EWRL and ADPRL show that the field has progressed a lot and requires its own community and events.

Concerning practical aspects of reinforcement learning, and more importantly, concerning benchmarking, evaluation and comparisons, much has happened. In addition to the planning competitions (e.g. such as the IPC), several editions of the reinforcement learning competitions[12] have been held with great success. Contestants competed in several classic domains (such as pole balancing) but also new and exciting domains such as the computer games Tetris and Super Mario. Competitions can promote code sharing and reuse, establish benchmarks for the field and be used to evaluate and compare methods on challenging domains. Another initiative for promoting more code and solution reuse is the RL-Glue framework[13], which provides an abstract reinforcement learning framework that can be used to share methods and domains among researchers. RL-Glue can connect to most common programming languages and thus provides a system- and language-independent software framework for experimentation. The competitions and RL-Glue help to further mature the field of reinforcement learning, and enable better scientific methods to test, compare and reuse reinforcement learning methods.

[12] `http://www.rl-competition.org/`

[13] `glue.rl-community.org/`

Goal of the Book and Intended Audience

As said before, we have tried to let this book be an answer to the question *"what book would you recommend to learn about current reinforcement learning?"*. Every person who could pose this question is contained in the potential audience for this book. This includes PhD and master students, researchers in reinforcement learning itself, and researchers in any other field who want to know about reinforcement learning. Having a book with 17 surveys on the major areas in current reinforcement learning provides an excellent starting point for researchers to continue expanding the field, applying reinforcement learning to new problems and to incorporate principled behavior learning techniques in their own intelligent systems and robots.

When we started the book project, we first created a long list of possible topics and grouped them, which resulted in a list of almost twenty large subfields of reinforcement learning in which many new results were published over the last decade. These include established subfields such as *evolutionary reinforcement learning*, but also newer topics such as *relational knowledge representation approaches* and *Bayesian frameworks* for learning and planning. *Hierarchical approaches*, about which a chapter is contained in this book, form the first subfield that basically emerged[14] right after the appearance of two mentioned books, and for that reason, were not discussed at that time.

Our philosophy when coming up with this book was to let the pool of authors reflect the youth and the active nature of the field. To that end, we selected and invited mainly young researchers in the start of their careers. Many of them finished their PhD studies in recent years, and that ensured that they were active and expert in their own sub-field of reinforcement learning, full of ideas and enthusiastic about that sub-field. Moreover, it gave them an excellent opportunity to promote that sub-field within the larger research area. In addition, we also invited several more experienced researchers who are recognized for their advances in several subfields of reinforcement learning. This all led to a good mix between different views on the subject matter. The initial chapter submissions were of very high quality, as we had hoped for. To complete the whole quality assurance procedure, we – the editors – together with a group of leading experts as reviewers, provided at least three reviews for each chapter. The results were that chapters were improved even further and that the resulting book contains a huge number of references to work in each of the subfields.

The resulting book contains 19 chapters, of which one contains introductory material on reinforcement learning, dynamic programming, Markov decision processes and foundational algorithms such as Q-learning and value iteration. The last chapter reflects on the material in the book, discusses things that were left out, and points out directions for further research. In addition, this chapter contains personal reflections and predictions about the field. The 17 chapters that form the core of the book are each self-contained introductions and overviews of subfields of reinforcement

[14] That is not to say that there were no hierarchical approaches, but the large portion of current hierarchical techniques appeared after the mid-nineties.

learning. In the next section we will give an overview of the structure of the book and its chapters. In total, the book features 30 authors, from many different institutes and different countries.

The Structure of This Book

The book consists of 18 surveys of sub-fields of reinforcement learning which are grouped together in four main categories which we will describe briefly in the following. The first chapter, **Reinforcement Learning and Markov Decision Processes** by *Martijn van Otterlo and Marco Wiering*, contains introductory material on basic concepts and algorithms. It discusses Markov decision processes and model-based and model-free algorithms for solving them. The goal of this chapter is to provide a quick overview of what constitute the main components of any reinforcement learning method, and it provides the necessary background for all other chapters. All surveys were written assuming this background was provided beforehand. The last chapter of the book, **Conclusions, Future Directions and Outlook** by *Marco Wiering and Martijn van Otterlo*, reflects on the material in the chapters and lists topics that were not discussed and directions for further research. In addition, it contains a list of personal reflections and predictions on the field, in the form of short statements written by several authors of chapters in the book. The main part of the book contains four groups of chapters and we will briefly introduce them individually in the following.

EFFICIENT SOLUTION FRAMEWORKS

The first part of the book contains several chapters on modern solution frameworks used in contemporary reinforcement learning. Most of these techniques can be understood in the light of the framework defined in the introduction chapter, yet these new methods emphasize more sophisticated use of samples, models of the world, and much more.

The first chapter in this part, **Batch Reinforcement Learning** by *Sascha Lange, Thomas Gabel, and Martin Riedmiller* surveys techniques for *batch learning* in the context of value function approximation. Such methods can make use of highly optimized regression techniques to learn robust and accurate value functions from huge amounts of data. The second chapter, **Least-Squares Methods for Policy Iteration** by *Lucian Buşoniu, Alessandro Lazaric, Mohammad Ghavamzadeh, Rémi Munos, Robert Babuška, and Bart De Schutter* surveys a recent trend in reinforcement learning on robust linear approximation techniques for policy learning. These techniques come with a solid set of mathematical techniques with which one can establish guarantees about learning speed, approximation accuracy and bounds. The third chapter, **Learning and Using Models** by *Todd Hester and Peter Stone* describes many ways in which models of the world can be learned and how they can speed up reinforcement learning. Learned models can be used for more efficient value updates, for planning, and for more effective exploration. World models represent general knowledge about the world and are, because of that, good candidates to be transferred to other, related tasks. More about the transfer of knowledge in reinforcement learning is surveyed in the chapter **Transfer in Reinforcement Learning: a Framework and a Survey** by *Alessandro Lazaric*. When confronted with several related tasks, various things can, once learned, be reused in a subsequent task. For example, policies can be reused, but depending on whether the state and/or action spaces of the two related tasks differ, other methods need to be applied. The chapter not only surveys existing approaches, but also tries to put them in a more general framework. The remaining chapter in this part, **Sample Complexity Bounds of Exploration** by *Lihong Li* surveys techniques and results concerning the sample complexity of reinforcement learning. For all algorithms it is important to know how many samples (examples of interactions with the world) are needed to guarantee a minimal performance on a task. In the past decade many new results have appeared that study this vital aspect in a rigorous and mathematical way and this chapter provides an overview of them.

CONSTRUCTIVE-REPRESENTATIONAL DIRECTIONS

This part of the book contains several chapters in which either *representations* are central, or their construction and use. As mentioned before, a major aspect of constructive techniques are the structures that enable value function approximation (or policies for that matter). Several major new developments in reinforcement learning are about finding new representational frameworks to learn behaviors in challenging new settings.

In the chapter **Reinforcement Learning in Continuous State and Action Spaces** by *Hado van Hasselt* many techniques are described for problem representations that contain continuous variables. This has been a major component in reinforcement learning for a long time, for example through the use of neural function approximators. However, several new developments in the field have tried to either more rigorously capture the properties of algorithms dealing with continuous states and actions or have applied such techniques in novel domains. Of particular interest are new techniques for dealing with continuous actions, since this effectively renders the amount of applicable actions infinite and requires sophisticated techniques for computing optimal policies. The second chapter, **Solving Relational and First-Order Logical Markov Decision Processes: A Survey** by *Martijn van Otterlo* describes a new representational direction in reinforcement learning which started around a decade ago. It covers all representations strictly more powerful than propositional (or; attribute-value) representations of states and actions. These include modelings as found in logic programming and first-order logic. Such representations can represent the world in terms of objects and relations and open up possibilities for reinforcement learning in a much broader set of domains than before. These enable many new ways of generalization over value functions, policies and world models and require methods from logical machine learning and knowledge representation to do so. The next chapter, **Hierarchical Approaches** by *Bernhard Hengst* too surveys a representational direction, although here representation refers to the structural decomposition of a *task*, and with that implicitly of the underlying Markov decision processes. Many of the hierarchical approaches appeared at the end of the nineties, and since then a large number of techniques has been introduced. These include new decompositions of tasks, value functions and policies, and many techniques for automatically learning task decompositions from interaction with the world. The final chapter in this part, **Evolutionary Computation for Reinforcement Learning** by *Shimon Whiteson* surveys evolutionary search for good policy structures (and value functions). Evolution has always been a good alternative for iterative, incremental reinforcement learning approaches and both can be used to optimize complex behaviors. Evolution is particularly well suited for non-Markov problems and policy structures for which gradients are unnatural or difficult to compute. In addition, the chapter surveys evolutionary neural networks for behavior learning.

PROBABILISTIC MODELS OF SELF AND OTHERS

Current artificial intelligence has become more and more *statistical* and *probabilistic*. Advances in the field of *probabilistic graphical models* are used virtually everywhere, and results for these models – both theoretical as computational – are effectively used in many sub-fields. This is no different for reinforcement learning. There are several large sub-fields in which the use of probabilistic models, such as Bayesian networks, is common practice and the employment of such a universal set of representations and computational techniques enables fruitful connections to other research employing similar models.

The first chapter, **Bayesian Reinforcement Learning** by *Nikos Vlassis, Mohammad Ghavamzadeh, Shie Mannor and Pascal Poupart* surveys Bayesian techniques for reinforcement learning. Learning sequential decision making under uncertainty can be cast in a Bayesian universe where interaction traces provide samples (evidence), and Bayesian inference and learning can be used to find optimal decision strategies in a rigorous probabilistic fashion. The next chapter, **Partially Observable Markov Decision Processes** by *Matthijs Spaan* surveys representations and techniques for partially observable problems which are very often cast in a probabilistic framework such as a dynamic Bayesian network, and where probabilistic inference is needed to infer underlying hidden (unobserved) states. The chapter surveys both model-based as well as model-free methods. Whereas POMDPs are usually modeled in terms of belief states that capture some form of history (or memory), a more recent class of methods that focuses on the *future* is surveyed in the chapter **Predictively Defined Representations of State** by *David Wingate*. These techniques maintain a belief state used for action selection in terms of probabilistic predictions about future events. Several techniques are described in which these predictions are represented compactly and where these are updated based on experience in the world. So far, most methods focus on the prediction (or; evaluation) problem, and less on control. The fourth chapter, **Game Theory and Multi-agent Reinforcement Learning** by *Ann Nowé, Peter Vrancx and Yann-Michaël De Hauwere* moves to a more general set of problems in which multiple agents learn and interact. It surveys game-theoretic and multi-agent approaches in reinforcement learning and shows techniques used to optimize agents in the context of other (learning) agents. The final chapter in this part, **Decentralized POMDPs** by *Frans Oliehoek* surveys model-based (dynamic programming) techniques for systems consisting of multiple agents that have to cooperatively solve a large task that is decomposed into a set of POMDPs. Such models for example appear in domains where multiple sensors in different locations together may provide essential information on how to act optimally in the world. This chapter builds on methods found in both POMDPs and multi-agent situations.

DOMAINS AND BACKGROUND

As we have said in the beginning of this preface, reinforcement learning appears as a method in many other fields of artificial intelligence, to optimize behaviors. Thus, in addition to the many algorithmic advances as described in the previous three parts of the book, we wanted to include surveys of areas in which reinforcement learning has been applied successfully. This part features chapters on robotics and games. In addition, a third chapter reflects the growing interest in connecting reinforcement learning and cognitive neuroscience.

The first chapter, **Psychological and Neuroscientific Connections with Reinforcement Learning** by *Ashvin Shah* surveys the connection between reinforcement learning methods on the one hand and cognition and neuroscience on the other. Originally many reinforcement learning techniques were derived from insights developed in psychology by for example Skinner, Thorndike and Watson, and still much cross-fertilization between psychology and reinforcement learning can happen. Lately, due to advances in theory about the brain, and especially because testing and measuring of brain activity (fMRI, EEG, etc.) has become much better, much research tries to either 1) explain research findings about the brain in terms of reinforcement learning techniques, i.e. which algorithms do really happen in the brain or 2) get inspired by the inner workings of the brain to come up with new algorithms. The second chapter in this part, **Reinforcement Learning in Games** by *István Szita* surveys the use of reinforcement learning in games. Games is more a general term here than as used in one of the previous chapters on game theory. Indeed, games in this chapter amount to board games such as Backgammon and Checkers, but also computer games such as role-playing and real-time strategy games. Games are often an exciting test bed for reinforcement learning algorithms (see for example also Tetris and Mario in the mentioned reinforcement learning competitions), and in addition to giving many examples, this chapter also tries to outline the main important aspects involved when applying reinforcement learning in game situations. The third chapter in this part, **Reinforcement Learning in Robotics: a Survey** by *Jens Kober and Jan Peters* rigorously describes the application of reinforcement learning to robotics problems. Robotics, because it works with the real, physical world, features many problems that are challenging for the robust application of reinforcement learning. Huge amounts of noisy data, slow training and testing on real robots, the reality gap between simulators and the real world, and scaling up to high-dimensional state spaces are just some of the challenging problems discussed here. Robotics is an exciting area also because of the added possibilities of putting humans in the loop which can create extra opportunities for imitation learning, learning from demonstration, and using humans as teachers for robots.

Acknowledgements

Crafting a book such as this can not be done overnight. Many people have put a lot of work in it to make it happen. First of all, we would like to give a big thanks to all the authors who have put in all their expertise, time and creativity to write excellent surveys of their sub-fields. Writing a survey usually takes some extra effort, since it requires that you know much about a topic, but in addition that you can put all relevant works in a more general framework. As editors, we are very happy with the way the authors have accomplished this difficult, yet very useful, task.

A second group of people we would like to thank are the reviewers. They have provided us with very thorough, and especially very constructive, reviews and these have made the book even better. We thank these reviewers who agreed to put their names in the book; thank you very much for all your help: Andrea Bonarini, Prasad Tadepalli, Sarah Ostentoski, Rich Sutton, Daniel Kudenko, Jesse Hoey, Christopher Amato, Damien Ernst, Remi Munos, Johannes Fuernkrantz, Juergen Schmidhuber, Thomas Rückstiess, Joelle Pineau, Dimitri Bertsekas, John Asmuth, Lisa Torrey, Yael Niv, Te Thamrongrattanarit, Michael Littman and Csaba Szepesvari.

Thanks also to Rich Sutton who was so kind to write the foreword to this book. We both consider him as one of the main figures in reinforcement learning, and in all respects we admire him for all the great contributions he has made to the field. He was there in the beginning of modern reinforcement learning, but still he continuously introduces novel, creative new ways to let agents learn. Thanks Rich!

Editing a book such as this is made much more convenient if you can fit it in your daily scientific life. In that respect, Martijn would like to thank both the Katholieke Universiteit Leuven (Belgium) as well as the Radboud University Nijmegen (The Netherlands) for their support. Marco would like to thank the University of Groningen (The Netherlands) for the same kind of support.

Last but not least, we would like to thank you, the reader, to having picked this book and having started to read it. We hope it will be useful to you, and hope that the work you are about to embark on will be incorporated in a subsequent book on reinforcement learning.

Groningen, Nijmegen,
November 2011

Marco Wiering
Martijn van Otterlo

Contents

Part III Constructive-Representational Directions

Part VI Closing

List of Contributors

Robert Babuška
Delft Center for Systems and Control, Delft University of Technology,
The Netherlands
e-mail: r.babuska@tudelft.nl

Lucian Buşoniu
Research Center for Automatic Control (CRAN), University of Lorraine, France
e-mail: lucian@busoniu.net

Thomas Gabel
Albert-Ludwigs-Universtität, Faculty of Engineering, Germany,
e-mail: tgabel@informatik.uni-freiburg.de

Mohammad Ghavamzadeh
Team SequeL, INRIA Lille-Nord Europe, France
e-mail: mohammad.ghavamzadeh@inria.fr

Hado van Hasselt
Centrum Wiskunde en Informatica (CWI, Center for Mathematics and Computer Science), Amsterdam, The Netherlands
e-mail: H.van.Hasselt@cwi.nl

Yann-Michaël De Hauwere
Vrije Universiteit Brussel, Belgium
e-mail: ydehauwe@vub.ac.be

Bernhard Hengst
School of Computer Science and Engineering,
University of New South Wales, Sydney, Australia
e-mail: bernhardh@cse.unsw.edu.au

Todd Hester
Department of Computer Science, The University of Texas at Austin, USA
e-mail: todd@cs.utexas.edu

Jens Kober
1) Intelligent Autonomous Systems Institute, Technische Universitaet Darmstadt, Darmstadt, Germany; 2) Robot Learning Lab, Max-Planck Institute for Intelligent Systems, Tübingen, Germany
e-mail: jens.kober@tuebingen.mpg.de

Sascha Lange
Albert-Ludwigs-Universtität Freiburg, Faculty of Engineering, Germany
e-mail: slange@informatik.uni-freiburg.de

Alessandro Lazaric
Team SequeL, INRIA Lille-Nord Europe, France
e-mail: alessandro.lazaric@inria.fr

Lihong Li
Yahoo! Research, Santa Clara, USA
e-mail: lihong@yahoo-inc.com

Shie Mannor
Technion, Haifa, Israel
e-mail: shie@ee.technion.ac.il

Rémi Munos
Team SequeL, INRIA Lille-Nord Europe, France
e-mail: remi.munos@inria.fr

Frans Oliehoek
CSAIL, Massachusetts Institute of Technology
e-mail: fao@csail.mit.edu

Ann Nowé
Vrije Universiteit Brussel, Belgium
e-mail: anowe@vub.ac.be

Martijn van Otterlo
Radboud University Nijmegen, The Netherlands
e-mail: m.vanotterlo@donders.ru.nl

Jan Peters
1) Intelligent Autonomous Systems Institute, Technische Universitaet Darmstadt, Darmstadt, Germany; 2) Robot Learning Lab, Max-Planck Institute for Intelligent Systems, Tübingen, Germany
e-mail: jan.peters@tuebingen.mpg.de

Pascal Poupart
University of Waterloo, Canada
e-mail: ppoupart@cs.uwaterloo.ca

Martin Riedmiller
Albert-Ludwigs-Universtität Freiburg, Faculty of Engineering, Germany
e-mail: riedmiller@informatik.uni-freiburg.de

Bart De Schutter
Delft Center for Systems and Control,
Delft University of Technology, The Netherlands
e-mail: b.deschutter@tudelft.nl

Ashvin Shah
Department of Psychology, University of Sheffield, Sheffield, UK
e-mail: ashvin@gmail.com

Matthijs Spaan
Institute for Systems and Robotics, Instituto Superior Técnico, Lisbon, Portugal
e-mail: mtjspaan@isr.ist.utl.pt

Peter Stone
Department of Computer Science, The University of Texas at Austin, USA
e-mail: pstone@cs.utexas.edu

István Szita
University of Alberta, Canada
e-mail: szityu@gmail.com

Nikos Vlassis
(1) Luxembourg Centre for Systems Biomedicine, University of Luxembourg, and (2) OneTree Luxembourg
e-mail: nikos.vlassis@uni.lu,nikos@onetreesol.com

Peter Vrancx
Vrije Universiteit Brussel, Belgium
e-mail: pvrancx@vub.ac.be

Shimon Whiteson
Informatics Institute, University of Amsterdam, The Netherlands
e-mail: s.a.whiteson@uva.nl

Marco Wiering
Department of Artificial Intelligence, University of Groningen, The Netherlands
e-mail: m.a.wiering@rug.nl

David Wingate
Massachusetts Institute of Technology, Cambridge, USA
e-mail: wingated@mit.edu

Acronyms

AC	Actor-Critic
AO	Action-Outcome
BAC	Bayesian Actor-Critic
BEETLE	Bayesian Exploration-Exploitation Tradeoff in Learning
BG	Basal Ganglia
BQ	Bayesian Quadrature
BQL	Bayesian Q-learning
BPG	Bayesian Policy Gradient
BRM	Bellman Residual Minimization (generic; BRM-Q for Q-functions; BRM-V for V-functions)
CMA-ES	Covariance Matrix Adaptation Evolution Strategy
CPPN	Compositional Pattern Producing Network
CoSyNE	Cooperative Synapse Coevolution
CR	Conditioned Response
CS	Conditioned Stimulus
DA	Dopamine
DBN	Dynamic Bayesian Network
DEC-MDP	Decentralized Markov Decision Process
DFQ	Deep Fitted Q iteration
DP	Dirichlet process
DP	Dynamic Programming
DTR	Decision-Theoretic Regression
EDA	Estimation of Distribution Algorithm
ESP	Enforced SubPopulations
FODTR	First-Order (Logical) Decision-Theoretic Regression
FQI	Fitted Q Iteration
GP	Gaussian Process
GPI	Generalized Policy Iteration
GPTD	Gaussian Process Temporal Difference
HBM	Hierarchical Bayesian model
HRL	Hierarchical Reinforcement Learning

ILP	Inductive Logic Programming
KADP	Kernel-based Approximate Dynamic Programming
KR	Knowledge Representation
KWIK	Knows What It Knows
LCS	Learning Classifier System
LSPE	Least-Squares Policy Evaluation (generic; ; LSPE-Q for Q-functions; LSPE-V for V-functions)
LSPI	Least-Squares Policy Iteration
LSTDQ	Least-Squares Temporal Difference Q-Learning
LSTD	Least-Squares Temporal Difference (generic; LSTD-Q for Q-functions; LSTD-V for V-functions
MB	Mistake Bound
MC	Monte-Carlo
MCTS	Monte Carlo Tree Search
MDP	Markov Decision Process
ML	Machine Learning
MTL	Multi-Task Learning
MTRL	Multi-Task Reinforcement Learning
NEAT	NeuroEvolution of Augmenting Topologies
NFQ	Neural Fitted Q iteration
PAC	Probably Approximately Correct
PAC-MDP	Probably Approximately Correct in Markov Decision Process
PMBGA	Probabilistic Model-Building Genetic Algorithm
PI	Policy Iteration
PIAGeT	Policy Iteration using Abstraction and Generalization Techniques
POMDP	Partially Observable Markov Decision Process
RL	Reinforcement Learning
RMDP	Relational Markov Decision Process
SANE	Symbiotic Adaptive NeuroEvolution
sGA	Structured Genetic Algorithm
SMDP	Semi-Markov Decision Process
SR	Stimulus-Response
SRL	Statistical Relational Learning
TD	Temporal Difference
TWEANN	Topology- and Weight-Evolving Artificial Neural Network
UR	Unconditioned Response
US	Unconditioned Stimulus
VI	Value Iteration
VPI	Value of Perfect Information