Shimon Whiteson

Adaptive Representations for Reinforcement Learning

Studies in Computational Intelligence, Volume 291

Editor-in-Chief
Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

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

Vol. 270. Zong Woo Geem
Recent Advances In Harmony Search Algorithm, 2009
ISBN 978-3-642-04316-1

Vol. 271. Janusz Kacprzyk, Frederick E. Petry, and Adnan Yazici (Eds.)
Uncertainty Approaches for Spatial Data Modeling and Processing, 2009
ISBN 978-3-642-10662-0

Vol. 272. Carlos A. Coello Coello, Clarisse Dhaenens, and Laetitia Jourdan (Eds.)
Advances in Multi-Objective Nature Inspired Computing, 2009
ISBN 978-3-642-11217-1

Vol. 273. Fatos Xhafa, Santi Caballé, Ajith Abraham, Thanasis Daradoumis, and Angel Alejandro Juan Perez (Eds.)
Computational Intelligence for Technology Enhanced Learning, 2010
ISBN 978-3-642-11223-2

Vol. 274. Zbigniew W. Raś and Alicja Wieczorkowska (Eds.)
Advances in Music Information Retrieval, 2010
ISBN 978-3-642-11673-5

Vol. 275. Dilip Kumar Pratihar and Lakhmi C. Jain (Eds.)
Intelligent Autonomous Systems, 2010
ISBN 978-3-642-11675-9

Vol. 276. Jacek Mańdziuk
Knowledge-Free and Learning-Based Methods in Intelligent Game Playing, 2010
ISBN 978-3-642-11677-3

Vol. 277. Filippo Spagnolo and Benedetto Di Paola (Eds.)
European and Chinese Cognitive Styles and their Impact on Teaching Mathematics, 2010
ISBN 978-3-642-11679-7

Vol. 278. Radomir S. Stankovic and Jaakko Astola
From Boolean Logic to Switching Circuits and Automata, 2010
ISBN 978-3-642-11681-0

Vol. 279. Manolis Wallace, Ioannis E. Anagnostopoulos, Phivos Mylonas, and Maria Bielikova (Eds.)
Semantics in Adaptive and Personalized Services, 2010
ISBN 978-3-642-11683-4

Vol. 280. Chang Wen Chen, Zhu Li, and Shiguo Lian (Eds.)
Intelligent Multimedia Communication: Techniques and Applications, 2010
ISBN 978-3-642-11685-8

Vol. 281. Robert Babuska and Frans C.A. Groen (Eds.)
Interactive Collaborative Information Systems, 2010
ISBN 978-3-642-11687-2

Vol. 282. Husrev Taha Sencar, Sergio Velastin, Nikolaos Nikolaidis, and Shiguo Lian (Eds.)
Intelligent Multimedia Analysis for Security Applications, 2010
ISBN 978-3-642-11754-1

Vol. 283. Ngoc Thanh Nguyen, Radoslaw Katarzyniak, and Shyi-Ming Chen (Eds.)
Advances in Intelligent Information and Database Systems, 2010
ISBN 978-3-642-12089-3

Vol. 284. Juan R. González, David Alejandro Pelta, Carlos Cruz, Germán Terrazas, and Natalio Krasnogor (Eds.)
Nature Inspired Cooperative Strategies for Optimization (NICSO 2010), 2010
ISBN 978-3-642-12537-9

Vol. 285. Roberto Cipolla, Sebastiano Battiato, and Giovanni Maria Farinella (Eds.)
Computer Vision, 2010
ISBN 978-3-642-12847-9

Vol. 286. Zeev Volkovich, Alexander Bolshoy, Valery Kirzhner, and Zeev Barzily
Genome Clustering, 2010
ISBN 978-3-642-12951-3

Vol. 287. Dan Schonfeld, Caifeng Shan, Dacheng Tao, and Liang Wang (Eds.)
Video Search and Mining, 2010
ISBN 978-3-642-12899-8

Vol. 288. I-Hsien Ting, Hui-Ju Wu, Tien-Hwa Ho (Eds.)
Mining and Analyzing Social Networks, 2010
ISBN 978-3-642-13421-0

Vol. 289. Anne Håkansson, Ronald Hartung, and Ngoc Thanh Nguyen (Eds.)
Agent and Multi-agent Technology for Internet and Enterprise Systems, 2010
ISBN 978-3-642-13525-5

Vol. 290. Weiliang Xu and John Bronlund
Mastication Robots, 2010
ISBN 978-3-540-93902-3

Vol. 291. Shimon Whiteson
Adaptive Representations for Reinforcement Learning, 2010
ISBN 978-3-642-13931-4

Shimon Whiteson

Adaptive Representations for Reinforcement Learning

Springer

Dr. Shimon Whiteson
Universiteit van Amsterdam
Instituut voor Informatica
Science Park 107
1098 XG Amsterdam
Netherlands
E-mail: s.a.whiteson@uva.nl

ISBN 978-3-642-13931-4 e-ISBN 978-3-642-13932-1

DOI 10.1007/978-3-642-13932-1

Studies in Computational Intelligence ISSN 1860-949X

Library of Congress Control Number: 2010928587

Typeset & *Cover Design:* Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

To Adam and Rena

Foreword

It is a great pleasure and honor to write the forward for this book, representing the culmination of Shimon Whiteson's Ph.D. thesis research at The University of Texas at Austin. When I arrived at UT Austin in the fall for 2002, Shimon was one of the first students to walk into my office and we began doing research together almost immediately. Our research helped form the nucleus of my research group, the Learning Agents Research Group, and Shimon became my first Ph.D. graduate in the spring of 2007.

Shimon was an ideal first student for a new assistant professor. He has a strong sense of what he wants to learn, and is never satisfied with a partial answer to any question. Most importantly for this book, he is self-assured and is willing to take risks in order to achieve meaningful results.

After several initial contributions that built upon my own past research, Shimon set off on his own towards his most important technical contribution so far, namely the development of a novel algorithm, NEAT+Q, capable of evolving neural network function approximators for reinforcement learning agents. The technical details of NEAT+Q are well-presented in the book, so suffice it to say here that his work on adaptive representations for reinforcement learning takes a substantial step towards addressing one of the key current issues in the field of machine learning, namely how to select the underlying representation that an agent uses when learning.

Perhaps more importantly, Shimon's work actively brings researchers in temporal difference learning and evolutionary computation — two largely disjoint communities that focus on similar problems — closer together by being recognized and respected in both communities. This cross-disciplinary aspect of his work was the biggest risk involved, especially for a Ph.D. student who has an eye towards academia. There was a chance that the research would not be accepted by anybody. Instead, he was able to achieve recognition in both areas.

By way of the research presented in this book, Shimon has established himself as one the pre-eminent worldwide experts on machine learning for sequential decision making tasks. A particular strength of the research is its detailed empirical analysis of both the capabilities and the limitations of all variants of his proposed algorithms. In addition, Shimon's clear writing style, full explanation of background material, and detailed survey of related work make his book useful beyond its own research contributions.

In short, for both newcomers to the field and for practitioners looking for nuanced detail, this book has plenty to offer!

Austin, TX,
December 2009

Peter Stone, Associate Professor
University of Texas at Austin

Preface

This book presents the main results of the research I conducted as a Ph.D. student at The University of Texas at Austin, primarily between 2004 and 2007. The primary contributions are new algorithms for *reinforcement learning*, a form of machine learning in which an autonomous agent seeks an effective control policy for tackling a sequential decision task. Unlike in supervised learning, the agent never sees examples of correct or incorrect behavior but receives only a reward signal as feedback. One limitation of current methods is that they typically require a human to manually design a representation for the solution (e.g. the internal structure of a neural network). Since poor design choices can lead to grossly suboptimal policies, agents that automatically adapt their own representations have the potential to dramatically improve performance. This book introduces two novel approaches for automatically discovering high-performing representations.

The first approach synthesizes temporal difference methods, the traditional approach to reinforcement learning, with evolutionary methods, which can learn representations for a broad class of optimization problems. This synthesis is accomplished via 1) *on-line evolutionary computation*, which customizes evolutionary methods to the on-line nature of most reinforcement learning problems, and 2) *evolutionary function approximation*, which evolves representations for the value function approximators that are critical to the temporal difference approach.

The second approach, called *adaptive tile coding*, automatically learns representations based on tile codings, which form piecewise-constant approximations of value functions. It begins with coarse representations and gradually refines them during learning, analyzing the current policy and value function to deduce the best refinements.

This book also introduces a novel method for devising input representations. In particular, it presents a way to find a minimal set of features sufficient to describe the agent's current state, a challenge known as the *feature selection* problem. The technique, called *Feature Selective NEAT* is an extension to NEAT, a method for evolving neural networks used throughout this work. While NEAT evolves both the topology and weights of a neural network, FS-NEAT goes one step further by learning the network's inputs too. Using evolution, it automatically and simultaneously determines the network's inputs, topology, and weights.

In addition to introducing these new methods, this book presents extensive empirical results in multiple domains demonstrating that these techniques can substantially improve performance over methods with manual representations.

The research presented in this book would not have been possible without the critical contributions of many collaborators. These include Peter Stone, my Ph.D. advisor; Risto Miikkulainen, a member of my thesis committee; and my colleagues Nate Kohl, Ken Stanley, and Matt Taylor. In addition, this research was supported in part by grants from IBM, NASA, NSF, and DARPA and by an IBM Ph.D. Fellowship.

Amsterdam, The Netherlands
February 2010

Shimon Whiteson, Assistant Professor
University of Amsterdam

Contents

1 Introduction 1
1.1 Motivation 2
1.2 Approach 2
1.3 Overview 4

2 Reinforcement Learning 7
2.1 Reinforcement Learning Framework 7
2.2 Temporal Difference Methods 9
2.3 Policy Search Methods 11

3 On-Line Evolutionary Computation 17
3.1 ϵ-Greedy Evolution 20
3.2 Softmax Evolution 21
3.3 Interval Estimation Evolution 22
3.4 Testbed Domains 22
3.4.1 Mountain Car 23
3.4.2 Server Job Scheduling 24
3.5 Results 26
3.6 Discussion 28

4 Evolutionary Function Approximation 31
4.1 NEAT+Q 33
4.2 Results 34
4.2.1 Comparing Manual and Evolutionary Function Approximation 34
4.2.2 Combining On-Line Evolution with Evolutionary Function Approximation 37
4.2.3 Comparing to Other Approaches 38
4.2.4 Comparing Darwinian and Lamarckian Approaches 42

4.2.5 Continual Learning Tests ... 43
4.3 Discussion ... 45

5 Sample-Efficient Evolutionary Function Approximation ... 47
5.1 Sample-Efficient NEAT+Q ... 47
5.2 Results ... 49
5.3 Discussion ... 52

6 Automatic Feature Selection for Reinforcement Learning ... 53
6.1 FS-NEAT ... 54
6.2 Testbed Domain ... 56
6.3 Results ... 58
6.4 Discussion ... 61

7 Adaptive Tile Coding ... 65
7.1 Background ... 66
7.1.1 Tile Coding ... 66
7.2 Method ... 68
7.2.1 When to Split ... 69
7.2.2 Where to Split ... 70
subsubsection.7.2.2.1 subsubsection.7.2.2.2
7.3 Testbed Domains ... 72
7.4 Results ... 73
7.5 Discussion ... 74

8 Related Work ... 77
8.1 Optimizing Representations ... 77
8.1.1 Supervised Learning ... 77
8.1.2 Reinforcement Learning ... 78
8.1.3 Evolutionary Computation ... 82
8.2 Combining Evolution and Learning ... 84
8.2.1 Applications to Supervised Learning ... 84
8.2.2 Applications to Reinforcement Learning ... 86
8.3 Balancing Exploration and Exploitation ... 89
8.3.1 k-Armed Bandit Problem ... 89
8.3.2 Associative Search ... 90
8.3.3 Reinforcement Learning ... 90
8.4 Feature Selection ... 92
8.4.1 Filters ... 92
8.4.2 Wrappers ... 93
8.4.3 FS-NEAT ... 93

9 Conclusion 95
9.1 Primary Conclusions 95
9.2 Negative Results 96
9.2.1 Combining FS-NEAT with NEAT+Q 96
9.2.2 Feature Selection in Adaptive Tile Coding 97
9.2.3 Fitness Functions Based on Bellman Error 97
9.3 Broader Implications 98
9.3.1 Stochastic vs. Deterministic Domains 98
9.3.2 The Value Function Gamble 99
9.3.3 The Role of Search in Adaptive Representations 100
9.4 Future Work 102
9.4.1 Non-stationarity 102
9.4.2 Steady-State Evolutionary Computation 102
9.4.3 Model-Based Reinforcement Learning 103
9.5 Final Remarks 104

A Statistical Significance 105
References 107