Handbook of Learning and Approximate Dynamic Programming

Jennie Si
Andy Barto
Warren Powell
Donald Wunsch

IEEE Neural Networks Society, *Sponsor*

IEEE PRESS

A JOHN WILEY & SONS, INC., PUBLICATION

Contents

Foreword 1
Shankar Sastry

1 ADP: Goals, Opportunities and Principles 3
Paul Werbos

1.1 Goals of This Book 3
1.2 Funding Issues, Opportunities and the Larger Context 5
1.3 Unifying Mathematical Principles and Roadmap of the Field 17

Part I Overview 45

2 Reinforcement Learning and Its Relationship to Supervised Learning 47
Andrew G. Barto and Thomas G. Dietterich

2.1 Introduction 47
2.2 Supervised Learning 48
2.3 Reinforcement Learning 50
2.4 Sequential Decision Tasks 54
2.5 Supervised Learning for Sequential Decision Tasks 58
2.6 Concluding Remarks 60

3 Model-Based Adaptive Critic Designs 65
Silvia Ferrari and Robert F. Stengel

3.1 Introduction 65
3.2 Mathematical Background and Foundations 67
3.3 Adaptive Critic Design and Implementation 74
3.4 Discussion 88
3.5 Summary 89

4 Guidance in the Use of Adaptive Critics for Control 97
George G. Lendaris and James C. Neidhoefer

4.1 Introduction 97
4.2 Reinforcement Learning 98
4.3 Dynamic Programming 99
4.4 Adaptive Critics: "Approximate Dynamic Programming" 99
4.5 Some Current Research on Adaptive Critic Technology 103
4.6 Application Issues 105
4.7 Items for Future ADP Research 118

5 Direct Neural Dynamic Programming 125
Jennie Si, Lei Yang and Derong Liu

5.1 Introduction 125
5.2 Problem Formulation 126
5.3 Implementation of Direct NDP 127
5.4 Comparisons 133
5.5 Continuous State Control Problem 138
5.6 Call Admission Control for CDMA Cellular Networks 141
5.7 Conclusions and Discussions 146

6 The Linear Programming Approach to Approximate Dynamic Programming 153
Daniela Pucci de Farias

6.1 Introduction 153
6.2 Markov Decision Processes 158
6.3 Approximate Linear Programming 159
6.4 State-Relevance Weights and the Performance of ALP Policies 160
6.5 Approximation Error Bounds 162
6.6 Application to Queueing Networks 165
6.7 Efficient Constraint Sampling Scheme 167
6.8 Discussion 173

7 Reinforcement Learning in Large, High-Dimensional State Spaces 179
Greg Grudic and Lyle Ungar

7.1 Introduction 179
7.2 Theoretical Results and Algorithm Specifications 185
7.3 Experimental Results 192
7.4 Conclusion 198

8 Hierarchical Decision Making 203
Malcolm Ryan

8.1 Introduction 203
8.2 Reinforcement Learning and the Curse of Dimensionality 204
8.3 Hierarchical Reinforcement Learning in Theory 209
8.4 Hierarchical Reinforcement Learning in Practice 217
8.5 Termination Improvement 221
8.6 Intra-Behavior Learning 223
8.7 Creating Behaviors and Building Hierarchies 225
8.8 Model-based Reinforcement Learning 225
8.9 Topics for Future Research 226
8.10 Conclusion 227

Part II Technical Advances 233

9 Improved Temporal Difference Methods with Linear Function Approximation 235
Dimitri P. Bertsekas, Vivek S. Borkar, and Angelia Nedich

9.1 Introduction 235
9.2 Preliminary Analysis 241
9.3 Convergence Analysis 243
9.4 Relations Between λ-LSPE and Value Iteration 245
9.5 Relation Between λ-LSPE and LSTD 252
9.6 Computational Comparison of λ-LSPE and TD(λ) 253

10 Approximate Dynamic Programming for High-Dimensional Resource Allocation Problems 261
Warren B. Powell and Benjamin Van Roy

10.1 Introduction 261
10.2 Dynamic Resource Allocation 262
10.3 Curses of Dimensionality 265
10.4 Algorithms for Dynamic Resource Allocation 266
10.5 Mathematical programming 271
10.6 Approximate Dynamic Programming 275
10.7 Experimental Comparisons 277
10.8 Conclusion 279

11 Hierarchical Approaches to Concurrency, Multiagency, and Partial Observability **285**
Sridhar Mahadevan, Mohammad Ghavamzadeh, Khashayar Rohanimanesh, and Georgios Theocharous

11.1 Introduction 285
11.2 Background 287
11.3 Spatiotemporal Abstraction of Markov Processes 289
11.4 Concurrency, Multiagency, and Partial Observability 294
11.5 Summary and Conclusions 306

12 Learning and Optimization — From a System Theoretic Perspective **311**
Xi-Ren Cao

12.1 Introduction 311
12.2 General View of Optimization 313
12.3 Estimation of Potentials and Performance Derivatives 316
12.4 Gradient-Based Optimization 323
12.5 Policy Iteration 324
12.6 Constructing Performance Gradients with Potentials as Building Blocks 328
12.7 Conclusion 330

13 Robust Reinforcement Learning Using Integral-Quadratic Constraints **337**
Charles W. Anderson, Matt Kretchmar, Peter Young, and Douglas Hittle

13.1 Introduction 337
13.2 Integral-Quadratic Constraints and Stability Analysis 338
13.3 Reinforcement Learning in the Robust Control Framework 340
13.4 Demonstrations of Robust Reinforcement Learning 346
13.5 Conclusions 354

14 Supervised Actor-Critic Reinforcement Learning **359**
Michael T. Rosenstein and Andrew G. Barto

14.1 Introduction 359
14.2 Supervised Actor-Critic Architecture 361
14.3 Examples 366
14.4 Conclusions 375

15 BPTT and DAC — A Common Framework for Comparison **381**
Danil V. Prokhorov

15.1 Introduction 381
15.2 Relationship between BPTT and DAC 383
15.3 Critic representation 386
15.4 Hybrid of BPTT and DAC 390
15.5 Computational complexity and other issues 395
15.6 Two classes of challenging problems 397
15.7 Conclusion 401

Part III Applications **405**

16 Near-Optimal Control Via Reinforcement Learning **407**
Augustine O. Esogbue and Warren E. Hearnes II

16.1 Introduction 407
16.2 Terminal Control Processes 408
16.3 A Hybridization: The GCS-Δ Controller 410
16.4 Experimental Investigation of the GCS-Δ Algorithm 422
16.5 Dynamic Allocation of Controller Resources 425
16.6 Conclusions and Future Research 427

17 Multiobjective Control Problems by Reinforcement Learning **433**
Dong-Oh Kang and Zeungnam Bien

17.1 Introduction 433
17.2 Preliminary 435
17.3 Policy Improvement Algorithm with Vector-Valued Reward 440
17.4 Multi-Reward Reinforcement Learning for Fuzzy Control 443
17.5 Summary 453

18 Adaptive Critic Based Neural Network for Control-Constrained Agile Missile **463**
S. N. Balakrishnan and Dongchen Han

18.1 Introduction 463
18.2 Problem Formulation and Solution Development 465
18.3 Minimum Time Heading Reversal Problem in a Vertical Plane 469
18.4 Use of Networks in Real-Time as Feedback Control 472
18.5 Numerical Results 473
18.6 Conclusions 476

19 Applications of Approximate Dynamic Programming in Power Systems Control 479
Ganesh K Venayagamoorthy, Ronald G Harley, and Donald C Wunsch

19.1 Introduction 479
19.2 Adaptive Critic Designs and Approximate Dynamic Programming 483
19.3 General Training Procedure for Critic and Action Networks 493
19.4 Power System 494
19.5 Simulation and Hardware Implementation Results 496
19.6 Summary 510

20 Robust Reinforcement Learning for Heating, Ventilation, and Air Conditioning Control of Buildings 517
Charles W. Anderson, Douglas Hittle, Matt Kretchmar, and Peter Young

20.1 Introduction 517
20.2 Heating Coil Model and PI Control 521
20.3 Combined PI and Reinforcement Learning Control 522
20.4 Robust Control Framework for Combined PI and RL Control 525
20.5 Conclusions 529

21 Helicopter Flight Control Using Direct Neural Dynamic Programming 535
Russell Enns and Jennie Si

21.1 Introduction 535
21.2 Helicopter Model 538
21.3 Direct NDP Mechanism Applied to Helicopter Stability Control 540
21.4 Direct NDP Mechanism Applied to Helicopter Tracking Control 548
21.5 Reconfigurable Flight Control 553
21.6 Conclusions 556

22 Toward Dynamic Stochastic Optimal Power Flow 561
James A. Momoh

22.1 Grand Overview of the Plan for the Future Optimal Power Flow 561
22.2 Generalized Formulation of the OPF Problem 567
22.3 General Optimization Techniques Used in Solving the OPF Problem 571
22.4 State-of-the-Art Technology in OPF Programs: The Quadratic Interior Point (QIP) Method 575
22.5 Strategy for Future OPF Development 576
22.6 Conclusion 596

23 Control, Optimization, Security, and Self-healing of Benchmark Power Systems 599
James A. Momoh and Edwin Zivi

23.1 Introduction 599
23.2 Description of the Benchmark Systems 601
23.3 Illustrative Terrestrial Power System Challenge Problems 604
23.4 Illustrative Navy Power System Challenge Problems 614
23.5 Summary of Power System Challenges and Topics 629
23.6 Summary 633