International Series in Operations Research & Management Science

Volume 248

More information about this series at http://www.springer.com/series/6161

Richard J. Boucherie • Nico M. van Dijk
Editors

Markov Decision Processes in Practice

Springer

Editors
Richard J. Boucherie
Stochastic Operations Research
University of Twente
Enschede, The Netherlands

Nico M. van Dijk
Stochastic Operations Research
University of Twente
Enschede, The Netherlands

ISSN 0884-8289 ISSN 2214-7934 (electronic)
International Series in Operations Research & Management Science
ISBN 978-3-319-47764-0 ISBN 978-3-319-47766-4 (eBook)
DOI 10.1007/978-3-319-47766-4

Library of Congress Control Number: 2017932096

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To

Carla,
Fabian, Daphne, Deirdre, and Daniël –
Thanks for being there in difficult times,
Richard

P. Dorreboom and his daughter –
for coping with my passions,
Nico

Foreword

I had the pleasure of serving as the series editor of this series over its first 20 years (from 1993 through October, 2013). One of the special pleasures of this work was the opportunity to become better acquainted with many of the leading researchers in our field and to learn more about their research. This was especially true in the case of Nico M. van Dijk, who became a friend and overnight guest in our home. I then was delighted when Nico and his colleague, Richard J. Boucherie, agreed to be the editors of a handbook, *Queueing Networks: A Fundamental Approach*, that was published in 2010 as Vol. 154 in this series. This outstanding volume succeeded in defining the current state of the art in this important area.

Because of both its elegance and its great application potential, Markov decision processes have been one of my favorite areas of operations research. A full chapter (Chap. 19 in the current tenth edition) is devoted to this topic in my textbook (coauthored by the late Gerald J. Lieberman), *Introduction to Operations Research*. However, I have long been frustrated by the sparsity of publications that describe applications of Markov decision processes. This was less true about 30 years ago when D.J. White published his seminal papers on such *real* applications in *Interfaces* (see the November–December 1985 and September–October 1988 issues). Unfortunately, relatively few papers or books since then have delved much into such applications. (One of these few publications is the 2002 book edited by Eugene Feinberg and Adam Shwartz, *Handbook of Markov Decision Processes: Methods and Applications*, which is Vol. 40 in this series.)

Given the sparse literature in this important area, I was particularly delighted when the outstanding team of Nico M. van Dijk and Richard J. Boucherie accepted my invitation to be the editors of this exciting new book that focuses on Markov decision processes in practice. One of my last acts as the series editor was to work with these coeditors and the publisher in shepherding the book proposal through the process of providing the contract for its publication. I feel that this book may prove to be one of the most important books in the series because it sheds so much light on the great application potential of Markov decision processes. This hopefully will lead to a renaissance in applying this powerful technique to numerous *real* problems.

Stanford University
July 2016

Frederick S. Hillier

Preface

It is over 30 years ago since D.J. White started his series of surveys on practical applications of Markov decision processes (MDP),[1,2,3] over 20 years after the phenomenal book by Martin Puterman on the theory of MDP,[4] and over 10 years since Eugene A. Feinberg and Adam Shwartz published their *Handbook of Markov Decision Processes: Methods and Applications*.[5] In the past decades, the practical development of MDP seemed to have come to a halt with the general perception that MDP is computationally prohibitive. Accordingly, MDP is deemed unrealistic and is out of scope for many operations research practitioners. In addition, MDP is hampered by its notational complications and its conceptual complexity. As a result, MDP is often only briefly covered in introductory operations research textbooks and courses. Recently developed approximation techniques supported by vastly increased numerical power have tackled part of the computational problems; see, e.g., Chaps. 2 and 3 of this handbook and the references therein. This handbook shows that a revival of MDP for practical purposes is justified for several reasons:

1. First and above all, the present-day numerical capabilities have enabled MDP to be invoked for real-life applications.
2. MDP allows to develop and formally support approximate and simple practical decision rules.
3. Last but not least, MDP's probabilistic modeling of practical problems is a skill if not art by itself.

[1] D.J. White. Real applications of Markov decision processes. *Interfaces*, 15:73–83, 1985.

[2] D.J. White. Further real applications of Markov decision processes. *Interfaces*, 18:55–61, 1988.

[3] D.J. White. A Survey of Applications of Markov Decision Processes. *Journal of the Operational Research Society*, 44:1073–1096, 1993.

[4] Martin Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, 1994.

[5] Eugene A. Feinberg and Adam Shwartz, editors. *Handbook of Markov Decision Processes: Methods and Applications*. Kluwer, 2002.

This handbook *Markov Decision Processes in Practice* aims *to show the power of classical MDP for real-life applications and optimization*. The handbook is structured as follows:

Part I: General Theory
Part II: Healthcare
Part III: Transportation
Part IV: Production
Part V: Communications
Part VI: Financial Modeling

The chapters of Part I are devoted to *the state-of-the-art theoretical foundation of MDP*, including approximate methods such as policy improvement, successive approximation and infinite state spaces as well as an instructive chapter on approximate dynamic programming. Parts II–VI contain a collection of *state-of-the-art applications in which MDP was key to the solution approach* in a non-exhaustive selection of application areas. The application-oriented chapters have the following structure:

- Problem description
- MDP formulation
- MDP solution approach
- Numerical and practical results
- Evaluation of the MDP approach used

Next to the MDP formulation and justification, most chapters contain numerical results and a real-life validation or implementation of the results. Some of the chapters are based on previously published results, some are expanding on earlier work, and some contain new research. All chapters are thoroughly reviewed. To facilitate comparison of the results offered in different chapters, several chapters contain an appendix with notation or a transformation of their notation to the basic notation provided in Appendix A. Appendix B contains a compact overview of all chapters listing discrete or continuous modeling aspects and the optimization criteria used in different chapters.

The outline of these six parts is provided below.

Part I: General Theory

This part contains the following chapters:

Chapter 1: One-Step Improvement Ideas and Computational Aspects
Chapter 2: Value Function Approximation in Complex Queueing systems
Chapter 3: Approximate Dynamic Programming by Practical Examples
Chapter 4: Server Optimization of Infinite Queueing Systems
Chapter 5: Structures of Optimal Policies in MDP with Unbounded Jumps: The State of Our Art

The first chapter, by H.C. Tijms, presents a survey of the basic concepts underlying computational approaches for MDP. Focus is on the basic principle of policy improvement, the design of a single good improvement step, and one-stage-look-ahead rules, to, e.g., generate the best control rule for the specific problem of interest, for decomposition results or parameterization, and to develop a heuristic or tailor-made rule. Several intriguing queueing examples are included, e.g., with dynamic routing to parallel queues.

In the second chapter, by S. Bhulai, using one-step policy improvement is brought down to the essence of understanding and evaluating the relative value function of simple systems that can be used in the control of more complicated systems. First, the essence of this relative value function is nicely clarified by standard birth death M/M/s queueing systems. Next, a number of approximations for the relative value function are provided and applied to more complex queueing systems such as for dynamic routing in real-life multiskilled call centers.

Chapter 3, by Martijn Mes and Arturo Pérez Rivera, continues the approximation approach and presents approximate dynamic programming (ADP) as a powerful technique to solve large-scale discrete-time multistage stochastic control problems. Rather than a more fundamental approach as, for example, can be found in the excellent book of Warren B. Powell,[6] this chapter illustrates the basic principles of ADP via three different practical examples: the nomadic trucker, freight consolidation, and tactical planning in healthcare.

The special but quite natural complication of infinite state spaces within MDP is given special attention in two consecutive chapters. First, in Chap. 4, by András Mészáros and Miklós Telek, the regular structure of several Markovian models is exploited to decompose an infinite transition matrix in a controllable and uncontrollable part, which allows a reduction of the unsolvable infinite MDP into a numerically solvable one. The approach is illustrated via queueing systems with parallel servers and a computer system with power saving mode and, in a more theoretical setting, for birth-death and quasi-birth-death models.

Next, in Chap. 5, by Herman Blok and Floske Spieksma, emphasis is on structural properties of infinite MDPs with unbounded jumps. Illustrated via a running example, the natural question is addressed, how structural properties of the optimal policy are preserved under truncation or perturbation of the MDP. In particular, smoothed rate truncation (SRT) is discussed, and a roadmap is provided for preserving structural properties.

[6] Warren B. Powell. *Approximate Dynamic Programming: Solving the Curses of Dimensionality.* Wiley Series in Probability and Statistics, 2011.

Part II: Healthcare

Healthcare is the largest industry in the Western world. The number of operations research practitioners in healthcare is steadily growing to tackle planning, scheduling, and decision problems. In line with this growth, in recent years, MDPs have found important applications in healthcare in the context of prevention, screening, and treatment of diseases but also in developing appointment schedules and inventory management. The following chapters contain a selection of topics:

Chapter 6: Markov Decision Processes for Screening and Treatment of Chronic Diseases
Chapter 7: Stratified Breast Cancer Follow-Up Using a Partially Observable MDP
Chapter 8: Advance Patient Appointment Scheduling
Chapter 9: Optimal Ambulance Dispatching
Chapter 10: Blood Platelet Inventory Management

Chapter 6, by Lauren N. Steimle and Brian T. Denton, provides a review of MDPs and partially observable MDPs (POMDPs) in medical decision making and a tutorial about how to formulate and solve healthcare problems with particular focus on chronic diseases. The approach is illustrated via two examples: an MDP model for optimal control of drug treatment decisions for managing the risk of heart disease and stroke in patients with type 2 diabetes and a POMDP model for optimal design of biomarker-based screening policies in the context of prostate cancer.

In Chap. 7, by J.W.M. Otten, A. Witteveen, I.M.H. Vliegen, S. Siesling, J.B. Timmer, and M.J. IJzerman, the POMDP approach is used to optimally allocate resources in a follow-up screening policy that maximizes the total expected number of quality-adjusted life years (QALYs) for women with breast cancer. Using data from the Netherlands Cancer Registry, for three risk categories based on differentiation of the primary tumor, the POMDP approach suggests a slightly more intensive follow-up for patients with a high risk for and poorly differentiated tumor and a less intensive schedule for the other risk groups.

In Chap. 8, by Antoine Sauré and Martin L. Puterman, the linear programming approach to ADP is used to solve advance patient appointment scheduling problems, which are problems typically intractable using standard solution techniques. This chapter provides a systematic way of identifying effective booking guidelines for advance patient appointment scheduling problems. The results are applied to CT scan appointment scheduling and radiation therapy treatment scheduling.

Chapter 9, by C.J. Jagtenberg, S. Bhulai, and R.D. van der Mei, considers the ambulance dispatch problem, in which one must decide which ambulance to send to an incident in real time. This chapter develops a computationally tractable MDP that captures not only the number of idle ambulances but also the future incident location and develops an ambulance dispatching heuristic that is shown to reduce the fraction of late arrivals by 13% compared to the “closest idle” benchmark policy for the Dutch region Flevoland.

Chapter 10, by Rene Haijema, Nico M. van Dijk, and Jan van der Wal, considers the blood platelet inventory problem that is of vital importance for patients’ survival, since platelets have a limited lifetime after being donated and lives may be at risk when no compatible blood platelets are available for transfusion, for example, during surgery. This chapter develops a combined MDP and simulation approach to minimize the blood platelet outdating percentage taking into account special production interruptions due to, e.g., Christmas and Easter holidays.

Part III: Transportation

Transportation science is known as a vast scientific field by itself for both the public (e.g., plane, train, or bus) and private modes of transportation. Well-known research areas include revenue management, pricing, air traffic control, train scheduling, and crew scheduling. This part contains only a small selection of topics to illustrate the possible fruitful use of MDP modeling within this field, ranging from macro-level to micro-level and from public transportation to private transportation. It contains the following chapters:

Chapter 11: Stochastic Dynamic Programming for Noise Load Management
Chapter 12: Allocation in a Vertical Rotary Car Park
Chapter 13: Dynamic Control of Traffic Lights
Chapter 14: Smart Charging of Electric Vehicles

Chapter 11, by T.R. Meerburg, Richard J. Boucherie, and M.J.A.L. van Kraaij, considers the runway selection problem that is typical for airports with a complex layout of runways. This chapter describes a stochastic dynamic programming (SDP) approach determining an optimal strategy for the monthly preference list selection problem under safety and efficiency restrictions and yearly noise load restrictions, as well as future and unpredictable weather conditions. As special MDP complications, a continuous state (noise volume) has to be discretized, and other states at sufficient distance are lumped to make the SDP numerically tractable.

In Chap. 12, by Mark Fackrell and Peter Taylor, both public and private goals are optimized, the latter indirectly. The objective is to balance the distribution of cars in a vertical car park by allocating arriving cars to levels in the best way. If no place is available, a car arrival is assumed to be lost. The randomness is inherent in the arrival process and the parking durations. This daily life problem implicitly concerns the problem of job allocation in an overflow system, a class of problems which are known to be unsolvable analytically in the uncontrolled case. An MDP heuristic rule is developed and extensive experiments show it to be superior.

Chapter 13, by Rene Haijema, Eligius M.T. Hendrix, and Jan van der Wal, studies another problem of daily life and both public and private concerns: dynamic control of traffic lights to minimize the mean waiting time of vehicles. The approach involves an approximate solution for a multidimensional MDP based on policy iteration in combination with decomposition of the state space into state spaces for different traffic streams. Numerical results illustrate that a single policy iteration step results in a strategy that greatly reduces average waiting time when compared to static control.

The final chapter of this transportation category, Chap. 14, by Pia L. Kempker, Nico M. van Dijk, Werner Scheinhardt, Hans van den Berg, and Johann Hurink, addresses overnight charging of electric vehicles taking into account the fluctuating energy demand and prices. A heuristic bidding strategy that is based on an analytical solution of the SDP for i.i.d. prices shows a substantial performance improvement compared to currently used standard demand side management strategies.

Part IV: Production

Control of production systems is a well-known application area that is known to be hampered by its computational complexity. This part contains three cases that illustrate the structure of approximate policies:

Chapter 15: Analysis of a Stochastic Lot Scheduling Problem with Strict DueDates
Chapter 16: Optimal Fishery Policies
Chapter 17: Near-Optimal Switching Strategies for a Tandem Queue

Chapter 15, by Nicky D. Van Foreest and Jacob Wijngaard, considers admission control and scheduling rules for a make-to-order stochastic lot scheduling problem with strict due dates. The CSLSP is a difficult scheduling problem for which MDPs seem to be one of the few approaches to analyze this problem. The MDP formulation further allows to set up simulations for large-scale systems.

In Chap. 16, by Eligius, M.T. Hendrix, Rene Haijema, and Diana van Dijk, a bi-level MDP for optimal fishing quota is studied. At the first level, an authority decides on the quota to be fished keeping in mind long-term revenues. At the second level, fishermen react on the quota set as well as on the current states of fish stock and fleet capacity by deciding on their investment and fishery effort. This chapter illustrates how an MDP with continuous state and action space can be solved by truncation and discretization of the state space and applying interpolation in the value iteration.

Chapter 17, by Daphne van Leeuwen and Rudesindo Núñez-Queija, is motivated by applications in logistics, road traffic, and production management. This chapter considers a tandem network, in which the waiting costs in the second queue are larger than those in the first queue. MDP is used to determine the near-optimal switching curve between serving and not serving at the first queue. that balances waiting costs at the queues. Discrete event simulation is used to show the appropriateness of the near-optimal strategies.

Part V: Communications

Communications has been an important application area for MDP with particular emphasis on call acceptance rules, channel selection, and transmission rates. This part illustrates some special cases for which a (near)-optimal strategy can be obtained:

Chapter 18: Wireless Channel Selection with Restless Bandits
Chapter 19: Flexible Staffing for Call Centers With Non-stationary Arrival Rates
Chapter 20: MDP for Query-Based Wireless Sensor Networks

Chapter 18, by Julia Kuhn and Yoni Nazarathy, considers wireless channel selection to maximize the long-run average throughput. The online control problem is modeled as restless multi-armed bandit (RMAB) problem in a POMDP framework. The chapter unifies several approaches and presents a nice development of the Whittle index.

Chapter 19, by Alex Roubos, Sandjai Bhulai, and Ger Koole, develops an MDP to obtain time-dependent staffing levels in a single-skill call center such that a service-level constraint is met in the presence of time-varying arrival rates. Through a numerical study based on real-life data, it is shown that the optimal policies provide a good balance between staffing costs and the penalty probability for not meeting the service level.

Chapter 20, by Mihaela Mitici, studies queries in a wireless sensor network, where queries might either be processed within the sensor network with possible delay or queries might be allocated to a database without delay but possibly containing outdated data. An optimal policy for query assignment is obtained from a continuous time MDP with drift. By an exponentially uniformized conversion (as extension of standard uniformization), it is transformed into a standard discrete-time MDP. By computation this leads to close-to-optimal simple policies.

Part VI: Financial Modeling

It is needless to say that financial modeling and stochastics are intrinsically related. Financial models represent a major field with time-series analysis for long-term financial and economic purposes as one well-known direction. Related directions concern stock, option, and utility theory. Early decision theory papers on portfolio management and investment modeling date back to the 1970s; see the edited book.[7] From a pure MDP perspective, the recently published book on Markov decision processes with special application in finance,[8] and the earlier papers by Jörn Sass and Manfred Schäl are recommended.

Chapter 21 by Jörn Sass and Manfred Schäl, gives an instructive review and follow-up on their earlier work to account for financial portfolios and derivatives under proportional transactional costs. In particular, a computational algorithm is developed for optimal pricing, and the optimal policy is shown to be a martingale that is of special interest in financial trading.

---

[7] Michael A. H. Dempster and Stanley R. Pliska, editors. *Mathematics of Derivative Securities*. Cambridge University Press, 1997.

[8] N. Bäuerle, U. Rieder. *Markov Decision Processes with Applications in Finance*. Springer, 2011.

Summarizing

These practical MDP applications have illustrated a variety of both standard and nonstandard aspects of MDP modeling and its practical use:

- A first and major step is a proper state definition containing sufficient information and details, which will frequently lead to multidimensional discrete or continuous states.
- The transition structure of the underlying process may involve time-dependent transition probabilities.
- The objective for optimization may be an average, discounted, or finite-time criterion.
- One-step rewards may be time dependent.
- The action set may be continuous or discrete.
- A simplified but computationally solvable situation can be an important first step in deriving a suitable policy that may subsequently be expanded to the solution of a more realistic case.
- Heuristic policies that may be implemented in practice can be developed from optimal policies.

We are confident that this handbook is appealing for a variety of readers with a background in, among others, operations research, mathematics, computer science, and industrial engineering:

1. A practitioner that would like to become acquainted with the possible value of MDP modeling and ways to use it
2. An academic or institutional researcher to become involved in an MDP modeling and development project and possibly expanding its frontiers
3. An instructor or student to be inspired by the instructive examples in this handbook to start using MDP for real-life problems

From each of these categories you are invited to step in and enjoy reading this hand book for further practical MDP applications.

P. Dorreboom 2009†: DP – a repetitive structure

Acknowledgments

We are most grateful to all authors for their positive reactions right from the initial invitations to contribute to this handbook: it is the quality of the chapters and the enthusiasm of the authors that will enable MDP to have its well-deserved impact on real-life applications.

We like to deeply express our gratitude to the former editor in chief and series editor: Fred Hillier. Had it not been for his stimulation from the very beginning in the first place and his assistance in its handling for approval, just before retirement, we would not have succeeded to complete this handbook.

Enschede, The Netherlands — Richard J. Boucherie
July 2016 — Nico M. van Dijk

Contents

Part V Communications

List of Contributors

S. Bhulai
Faculty of Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

H. Blok
Eindhoven University of Technology, Eindhoven, The Netherlands

Richard J. Boucherie
Stochastic Operations Research, University of Twente, Enschede, The Netherlands

Brian T. Denton
Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI, USA

Mark Fackrell
School of Mathematics and Statistics, University of Melbourne, VIC, Australia

Rene Haijema
Operations Research and Logistics group, Wageningen University, Wageningen, The Netherlands

Eligius M.T. Hendrix
Computer Architecture, Universidad de Málaga, Málaga, Spain

Johann Hurink
Department of Applied Mathematics, University of Twente, Enschede, The Netherlands

M.J. IJzerman
Department of Health Technology and Services Research, University of Twente, Enschede, The Netherlands

C.J. Jagtenberg
Stochastics, CWI, Amsterdam, The Netherlands

Pia L. Kempker
TNO, Cyber Security & Robustness TNO, The Hague, The Netherlands

Ger Koole
Faculty of Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Julia Kuhn
The University of Queensland, Brisbane, QLD, Australia

University of Amsterdam, Amsterdam, The Netherlands

T.R. Meerburg
Air Traffic Control The Netherlands, Schiphol, The Netherlands

Martijn Mes
Department of Industrial Engineering and Business Information Systems, University of Twente, Enschede, The Netherlands

András Mészáros
MTA-BME Information Systems Research Group, Budapest, Hungary

Mihaela Mitici
Faculty of Aerospace Engineering, Air Transport and Operations, Delft University of Technology, Delft, The Netherlands

Yoni Nazarathy
The University of Queensland, Brisbane, QLD, Australia

Rudesindo Núñez-Queija
CWI, Amsterdam, The Netherlands

J.W.M. Otten
Department of Stochastic Operations Research, University of Twente, Enschede, The Netherlands

Arturo Pérez Rivera
Department of Industrial Engineering and Business Information Systems, University of Twente, Enschede, The Netherlands

Martin L. Puterman
Sauder School of Business, University of British Columbia, Vancouver, BC, Canada V6T 1Z2

Alex Roubos
CCmath, Amsterdam, The Netherlands

Jörn Sass
Fachbereich Mathematik, TU Kaiserslautern, Kaiserslautern, Germany

Antoine Sauré
Telfer School of Management, University of Ottawa, Ottawa, ON, Canada K1N 6N5

Manfred Schäl
Institut für Angewandte Mathematik, Universität Bonn, Bonn, Germany

Werner Scheinhardt
Department of Applied Mathematics, University of Twente, Enschede, The Netherlands

S. Siesling
Department of Health Technology and Services Research, University of Twente, Enschede, The Netherlands

Department of Research, Comprehensive Cancer Organisation, Utrecht, The Netherlands

F.M. Spieksma
Leiden University, Leiden, The Netherlands

Lauren N. Steimle
Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI, USA

Peter Taylor
School of Mathematics and Statistics, University of Melbourne, VIC, Australia

Miklós Telek
Budapest University of Technology and Economics, Budapest, Hungary

Henk Tijms
Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

J.B. Timmer
Department of Stochastic Operations Research, University of Twente, Enschede, The Netherlands

Nico M. van Dijk
Stochastic Operations Research, University of Twente, Enschede, The Netherlands

Diana van Dijk
Department of Environmental Social Sciences, Swiss Federal Institute of Aquatic Science and Technology (EAWAG), Dübendorf, Switzerland

Nicky D. van Foreest
Faculty of Economics and Business, University of Groningen, Groningen, The Netherlands

M.J.A.L. van Kraaij
Air Traffic Control, Utrecht, The Netherlands

Daphne van Leeuwen
CWI, Amsterdam, The Netherlands

Hans van den Berg
TNO, Cyber Security & Robustness TNO, The Hague, The Netherlands

Department of Applied Mathematics, University of Twente, Enschede, The Netherlands

R.D. van der Mei
Stochastics, CWI, Amsterdam, The Netherlands

Jan van der Wal
Faculty of Economics and Business, University of Amsterdam, Amsterdam, The Netherlands

Stochastic Operations Research group, University of Twente, Enschede, The Netherlands

I.M.H. Vliegen
Department of Industrial Engineering and Business Information Systems, University of Twente, Enschede, The Netherlands

A. Witteveen
Department of Health Technology and Services Research, University of Twente, Enschede, The Netherlands

Jacob Wijngaard
Faculty of Economics and Business, University of Groningen, Groningen, The Netherlands