SPRINGER BRIEFS IN INTELLIGENT SYSTEMS
ARTIFICIAL INTELLIGENCE, MULTIAGENT SYSTEMS,
AND COGNITIVE ROBOTICS

Frans A. Oliehoek
Christopher Amato

A Concise Introduction to Decentralized POMDPs

Springer

SpringerBriefs in Intelligent Systems

Artificial Intelligence, Multiagent Systems, and Cognitive Robotics

This series covers the entire research and application spectrum of intelligent systems, including artificial intelligence, multiagent systems, and cognitive robotics. Typical texts for publication in the series include, but are not limited to, state-of-the-art reviews, tutorials, summaries, introductions, surveys, and in-depth case and application studies of established or emerging fields and topics in the realm of computational intelligent systems. Essays exploring philosophical and societal issues raised by intelligent systems are also very welcome.

More information about this series at http://www.springer.com/series/11845

Frans A. Oliehoek · Christopher Amato

A Concise Introduction to Decentralized POMDPs

Springer

Frans A. Oliehoek
School of Electrical Engineering, Electronics and Computer Science
University of Liverpool
Liverpool
UK

Christopher Amato
Computer Science and Artificial Intelligence Laboratory
MIT
Cambridge, MA
USA

ISSN 2196-548X ISSN 2196-5498 (electronic)
SpringerBriefs in Intelligent Systems
ISBN 978-3-319-28927-4 ISBN 978-3-319-28929-8 (eBook)
DOI 10.1007/978-3-319-28929-8

Library of Congress Control Number: 2016941071

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Dedicated to future generations of intelligent decision makers.

Preface

This book presents an overview of formal decision making methods for decentralized cooperative systems. It is aimed at graduate students and researchers in the fields of artificial intelligence and related fields that deal with decision making, such as operations research and control theory. While we have tried to make the book relatively self-contained, we do assume some amount of background knowledge.

In particular, we assume that the reader is familiar with the concept of an *agent* as well as search techniques (like depth-first search, A*, etc.), both of which are standard in the field of artificial intelligence [Russell and Norvig, 2009]. Additionally, we assume that the reader has a basic background in probability theory. Although we give a very concise background in relevant single-agent models (i.e., the 'MDP' and 'POMDP' frameworks), a more thorough understanding of those frameworks would benefit the reader. A good first introduction to these concepts can be found in the textbook by Russell and Norvig, with additional details in texts by Sutton and Barto [1998], Kaelbling et al. [1998], Spaan [2012] and Kochenderfer et al. [2015]. We also assume that the reader has a basic background in game theory and game-theoretic notations like Nash equilibrium and Pareto efficiency. Even though these concepts are not central to our exposition, we do place the Dec-POMDP model in the more general context they offer. For an explanation of these concepts, the reader could refer to any introduction on game theory, such as those by Binmore [1992], Osborne and Rubinstein [1994] and Leyton-Brown and Shoham [2008].

This book heavily builds upon earlier texts by the authors. In particular, many parts were based on the authors' previous theses, book chapters and survey articles [Oliehoek, 2010, 2012, Amato, 2010, 2015, Amato et al., 2013]. This also means that, even though we have tried to give a relatively complete overview of the work in the field, the text in some cases is biased towards examples and methods that have been considered by the authors. For the description of further topics in Chapter 8, we have selected those that we consider important and promising for future work. Clearly, there is a necessarily large overlap between these topics and the authors' recent work in the field.

Acknowledgments

Writing a book is not a standalone activity; it builds upon all the insights developed in the interactions with peers, reviewers and coathors. As such, we are grateful for the interaction we have had with the entire research field. We specifically want to thank the attendees and organizers of the workshops on *multiagent sequential decision making (MSDM)* which have provided a unique platform for exchange of thoughts on decision making under uncertainty.

Furthermore, we would like to thank João Messias, Matthijs Spaan, Shimon Whiteson, and Stefan Witwicki for their feedback on sections of the manuscript.

Finally, we are grateful to our former supervisors, in particular Nikos Vlassis and Shlomo Zilberstein, who enabled and stimulated us to go down the path of research on decentralized decision making.

Contents

Acronyms

AH action history
AOH action-observation history
BG Bayesian game
CG coordination graph
CBG collaborative Bayesian game
COP constraint optimization problem
DAG directed acyclic graph
DP dynamic programming
Dec-MDP decentralized Markov decision process
Dec-POMDP decentralized partially observable Markov decision process
DICE direct cross-entropy optimization
EM expectation maximization
EXP deterministic exponential time (complexity class)
FSC finite-state controller
FSPC forward-sweep policy computation
GMAA* generalized multiagent A*
I-POMDP interactive partially observable Markov decision process
MAS multiagent system
MARL multiagent reinforcement learning
MBDP memory-bounded dynamic programming
MDP Markov decision process
MILP mixed integer linear program
NEXP non-deterministic exponential time (complexity class)
ND-POMDP networked distributed POMDP
NDP nonserial dynamic programming
NLP nonlinear programming
NP non-deterministic polynomial time (complexity class)
OH observation history
POMDP partially observable Markov decision process
PSPACE polynonomial SPACE (complexity class)
PWLC piecewise linear and convex

RL reinforcement learning
TD-POMDP transition-decoupled POMDP

List of Symbols

Throughout this text, we tried to make consistent use of typesetting to convey the meaning of used symbols and formulas. In particular, we use blackboard bold fonts ($\mathbb{A}$,$\mathbb{B}$, etc.) to denote sets, and subscripts to denote agents (typically i or j) or groups of agents, as well as time (t or τ).

For instance a is the letter used to indicate actions in general, a_i denotes an action of agent i, and the set of its actions is denoted $\mathbb{A}_i$. The action agent i takes at a particular time step t is denoted $a_{i,t}$. The profile of actions taken by all agents, a joint action, is denoted a, and the set of such joint actions is denoted $\mathbb{A}$. When referring to the action profile of a subset e of agents we write a_e, and for the actions of all agents except agent i, we write a_{-i}. On some occasions we will need to indicate the index within a set, for instance the k-th action of agent i is written a_i^k. In the list of symbols below, we have shown all possible uses of notation related to actions (base symbol 'a'), but have not exhaustively applied such modifiers to all symbols.

$\cdot$	multiplication,
$\times$	Cartesian product,
$\circ$	policy concatenation,
$\Downarrow$	subtree policy consumption operator,
$\triangle(\cdot)$	simplex over $(\cdot)$,
$\mathbf{1}_{\{\cdot\}}$	indicator function,
β	macro-action termination condition,
Γ_j	mapping from histories to subtree policies,
$\Gamma^{\mathscr{X}}$	state factor scope backup operator,
$\Gamma^{\mathscr{A}}$	agent scope backup operator,
γ	discount factor,
δ_t	decision rule for stage t,
$\boldsymbol{\delta}_t$	joint decision rule for stage t,
$\hat{\boldsymbol{\delta}}_t$	approximate joint decision rule,
$\delta_{i,t}$	decision rule for agent i for stage t,
Δt	length of a stage ts,

ε	(small) constant,
$\bar{\theta}$	joint action-observation history,
$\bar{\theta}_i$	action-observation history,
$\bar{\Theta}_i$	action-observation history set,
ι_i	information state, or belief update, function,
μ_i	macro-action policy for agent i,
π	joint policy,
π_i	policy for agent i,
π_{-i}	(joint) policy for all agents but i,
π^*	optimal joint policy,
ρ	number of reward functions,
Σ	alphabet of communication messages,
σ_t	plan-time sufficient statistic,
τ	stages-to-go ($\tau = h - 1$),
υ	domination gap,
$\Phi_{\texttt{Next}}$	set of next policies,
φ_t	past joint policy,
ξ	parameter vector,
ψ	correlation device transition function,
$\mathbb{A}$	set of joint actions,
$\mathbb{A}_i$	set of actions for agent i,
$\bar{\mathbb{A}}$	joint action history set,
$\bar{\mathbb{A}}_i$	action history set for agent i,
a	joint action,
a_t	joint action at stage t,
a_i	action for agent i,
a_e	joint action for subset e of agents,
a_{-i}	joint action for all agents except i,
$\bar{a}_i$	action history of agent i,
$\bar{a}_{i,t}$	action history of agent i at stage t,
$\bar{a}$	joint action history,
$\bar{a}_t$	joint action history at stage t,
$B(b_0, \varphi_t)$	Bayesian game for a stage,
$B(\mathscr{M}_{DecP}, b_0, \varphi_t)$	CBG for stage t of a Dec-POMDP,
$\mathbb{B}$	set of joint beliefs,
b_0	initial state distribution,
$b_i(s_t, q^{\tau}_{-i})$	multiagent belief,
b	belief,
b_i	belief for agent i (e.g., a multiagent belief),
$\mathbb{C}$	states of a correlation device,
C_Σ	message cost function,
c	correlation device state,
$\mathbb{D}$	set of agents,
$\mathbf{E}[\cdot]$	expectation of $\cdot$,