Artificial Intelligence

Foundations of Computational Agents

Artificial intelligence, including machine learning, has emerged as a transformational science and engineering discipline. *Artificial Intelligence: Foundations of Computational Agents* presents AI using a coherent framework to study the design of intelligent computational agents. By showing how the basic approaches fit into a multidimensional design space, readers learn the fundamentals without losing sight of the bigger picture. The new edition also features expanded coverage on machine learning material, as well as on the social and ethical consequences of AI and ML. The book balances theory and experiment, showing how to link them together, and develops the science of AI together with its engineering applications. Although structured as an undergraduate and graduate textbook, the book's straightforward, self-contained style will also appeal to an audience of professionals, researchers, and independent learners. The second edition is well-supported by strong pedagogical features and online resources to enhance student comprehension.

David Poole is a Professor of Computer Science at the University of British Columbia. He is a co-author of three artificial intelligence books including *Statistical Relational Artificial Intelligence: Logic, Probability, and Computation*. He is a former Chair of the Association for Uncertainty in Artificial Intelligence, the winner of the Canadian AI Association (CAIAC) 2013 Lifetime Achievement Award, and a Fellow of the Association for the Advancement Artificial Intelligence (AAAI) and CAIAC.

Alan Mackworth is a Professor of Computer Science at the University of British Columbia. He has authored over 130 papers and co-authored two books: *Computational Intelligence: A Logical Approach* and *Artificial Intelligence: Foundations of Computational Agents*. His awards include the AIJ Classic Paper Award and the ACP Award for Research Excellence. He has served as President of IJCAI, AAAI and CAIAC. He is a Fellow of AAAI, CAIAC and the Royal Society of Canada.

Artificial Intelligence

Foundations of Computational Agents

David L. Poole

University of British Columbia, Canada

Alan K. Mackworth

University of British Columbia, Canada

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

4843/24, 2nd Floor, Ansari Road, Daryaganj, Delhi - 110002, India

79 Anson Road, #06-04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107195394
DOI: 10.1017/9781108164085

First edition published 2010
Second edition published 2017

Printed in United States of America by Sheridan Books, Inc

A catalogue record for this publication is available from the British Library

ISBN 978-1-107-19539-4 Hardback

Contents

Figures

Preface

Artificial Intelligence: Foundations of Computational Agents is a book about the science of artificial intelligence (AI). AI is the study of the design of intelligent computational agents. The book is structured as a textbook but it is designed to be accessible to a wide audience.

We wrote this book because we are excited about the emergence of AI as an integrated science. As with any science being developed, AI has a coherent, formal theory and a rambunctious experimental wing. Here we balance theory and experiment and show how to link them together intimately. We develop the science of AI together with its engineering applications. We believe the adage, "There is nothing so practical as a good theory." The spirit of our approach is captured by the dictum, "Everything should be made as simple as possible, but not simpler." We must build the science on solid foundations; we present the foundations, but only sketch, and give some examples of, the complexity required to build useful intelligent systems. Although the resulting systems will be complex, the foundations and the building blocks should be simple.

This second edition results from extensive revision throughout the text. We have restructured the material based on feedback from instructors who have used the book in classes. We have brought it up to date to reflect the current state of the art, made parts that were difficult for students more straightforward, added more intuitive explanations, and coordinated the pseudocode algorithms with new open-source implementations of the algorithms in Python and Prolog. We have resisted the temptation to just keep adding more material. AI research is expanding so rapidly now that the volume of potential new text material is vast. However, research teaches us not only what works but

also what does not work so well, allowing us to be highly selective. We have included more material on machine learning techniques that have proven successful. However, research also has trends and fashions. We have removed techniques that have been shown to be less promising, but we distinguish them from the techniques that are merely out of fashion. We include some currently unfashionable material if the problems attacked still remain and the techniques have the potential to form the basis for future research and development. We have further developed the concept of a single design space for intelligent agents, showing how many bewilderingly diverse techniques can be seen in a simple, uniform framework. This allows us to emphasize the principles underlying the foundations of computational agents, making those ideas more accessible to students.

The book can be used as an introductory text on artificial intelligence for advanced undergraduate or graduate students in computer science or related disciplines such as computer engineering, philosophy, cognitive science, or psychology. It will appeal more to the technically minded; parts are technically challenging, focusing on learning by doing: designing, building, and implementing systems. Any curious scientifically oriented reader will benefit from studying the book. Previous experience with computational systems is desirable, but prior study of the foundations upon which we build, including logic, probability, calculus, and control theory, is not necessary, because we develop the concepts as required.

The serious student will gain valuable skills at several levels ranging from expertise in the specification and design of intelligent agents to skills for implementing, testing, and improving real software systems for several challenging application domains. The thrill of participating in the emergence of a new science of intelligent agents is one of the attractions of this approach. The practical skills of dealing with a world of ubiquitous, intelligent, embedded agents are now in great demand in the marketplace.

The focus is on an intelligent agent acting in an environment. We start with simple agents acting in simple, static environments and gradually increase the power of the agents to cope with more challenging worlds. We explore ten dimensions of complexity that allow us to introduce, gradually and with modularity, what makes building intelligent agents challenging. We have tried to structure the book so that the reader can understand each of the dimensions separately and we make this concrete by repeatedly illustrating the ideas with four different agent tasks: a delivery robot, a diagnostic assistant, a tutoring system, and a trading agent.

The agent we want the student to envision is a hierarchically designed agent that acts intelligently in a stochastic environment that it can only partially observe – one that reasons online about individuals and relationships among them, has complex preferences, learns while acting, takes into account

other agents, and acts appropriately given its own computational limitations. Of course, we cannot start with such an agent; it is still a research question to build such agents. So we introduce the simplest agents and then show how to add each of these complexities in a modular way.

We have made a number of design choices which distinguish this book from competing books, including our earlier book.

- We have tried to give a coherent framework in which to understand AI. We have chosen not to present disconnected topics that do not fit together. For example, we do not present disconnected logical and probabilistic views of AI, but we have presented a multidimensional design space in which the students can understand the big picture, in which probabilistic and logical reasoning coexist.

- We decided that it is better to clearly explain the foundations upon which more sophisticated techniques can be built, rather than present these more sophisticated techniques. This means that a larger gap may exist between what is covered in this book and the frontier of science. But it also means that the student will have a better foundation to understand current and future research.

- One of the more difficult decisions we made was how to linearize the design space. Our previous book [Poole et al., 1998] presented a relational language early and built the foundations in terms of this language. This approach made it difficult for the students to appreciate work that was not relational, for example, in reinforcement learning that is developed in terms of states. In this book, we have chosen a relations-late approach. This approach probably reflects better the research over the past few decades where there has been much progress in reasoning and learning for feature-based representations. It also allows the student to understand that probabilistic and logical reasoning are complementary. The book, however, is structured so that an instructor could present relations earlier.

We provide open-source Python implementations of the algorithms (http://www.aipython.org); these are designed to be useful and to highlight the main ideas without extra frills to interfere with the main ideas. This book uses examples from Aispace.org (http://www.aispace.org), a collection of pedagogical applets that we have been involved in designing. To gain further experience in building intelligent systems, a student should also experiment with a high-level symbol-manipulation language, such as Haskell, Lisp or Prolog. We also provide implementations in AILog, a clean logic programming language related to Prolog, designed to demonstrate many of the issues in this book. These

tools are intended to be helpful, but not essential to an understanding or use of the ideas in this book.

Our approach, through the development of the power of the agent's capabilities and representation language, is both simpler and more powerful than the traditional approach of surveying and cataloging various applications of AI. However, as a consequence, some applications such as the details of computational vision or computational linguistics are not covered in this book.

We have chosen not to present an encyclopedic view of AI. Not every major idea that has been investigated is presented here. We have chosen some basic ideas upon which other, more sophisticated, techniques are based and have tried to explain the basic ideas in detail, sketching how these can be expanded.

Figure 1 (page xxvii) shows the topics covered in the book. The solid lines depict prerequisites. Often the prerequisite structure does not include all subtopics. Given the medium of a book, we have had to linearize the topics. However, the book is designed so the topics are teachable in any order satisfying the prerequisite structure.

The references given at the end of each chapter are not meant to be comprehensive; we have referenced works that we have directly used and works that we think provide good overviews of the literature, by referencing both classic works and more recent surveys. We hope that no researchers feel slighted by their omission, and we are happy to have feedback where someone feels that an idea has been misattributed. Remember that this book is *not* a survey of AI research.

We invite you to join us in an intellectual adventure: building a science of intelligent agents.

David Poole
Alan Mackworth

Acknowledgments

Thanks to Randy Goebel for valuable input on this book. We also gratefully acknowledge the helpful comments on the first edition and earlier drafts of the second edition received from Guy van den Broeck, David Buchman, Giuseppe Carenini, Cristina Conati, Mark Crowley, Matthew Dirks, Bahare Fatemi, Pooyan Fazli, Robert Holte, Holger Hoos, Manfred Jaeger, Mehran Kazemi, Mohammad Reza Khojasteh, Jacek Kisyński, Richard Korf, Bob Kowalski, Kevin Leyton-Brown, Josje Lodder, Marian Mackworth, Gabriel Murray, Sriraam Natarajan, Alex Poole, Alessandro Provetti, Mark Schmidt, Marco Valtorta, and the anonymous reviewers. Thanks to the students who pointed out many errors in the earlier drafts. Thanks to Jen Fernquist for the website design. David would like to thank Thomas Lukasiewicz and The Leverhulme Trust for sponsoring

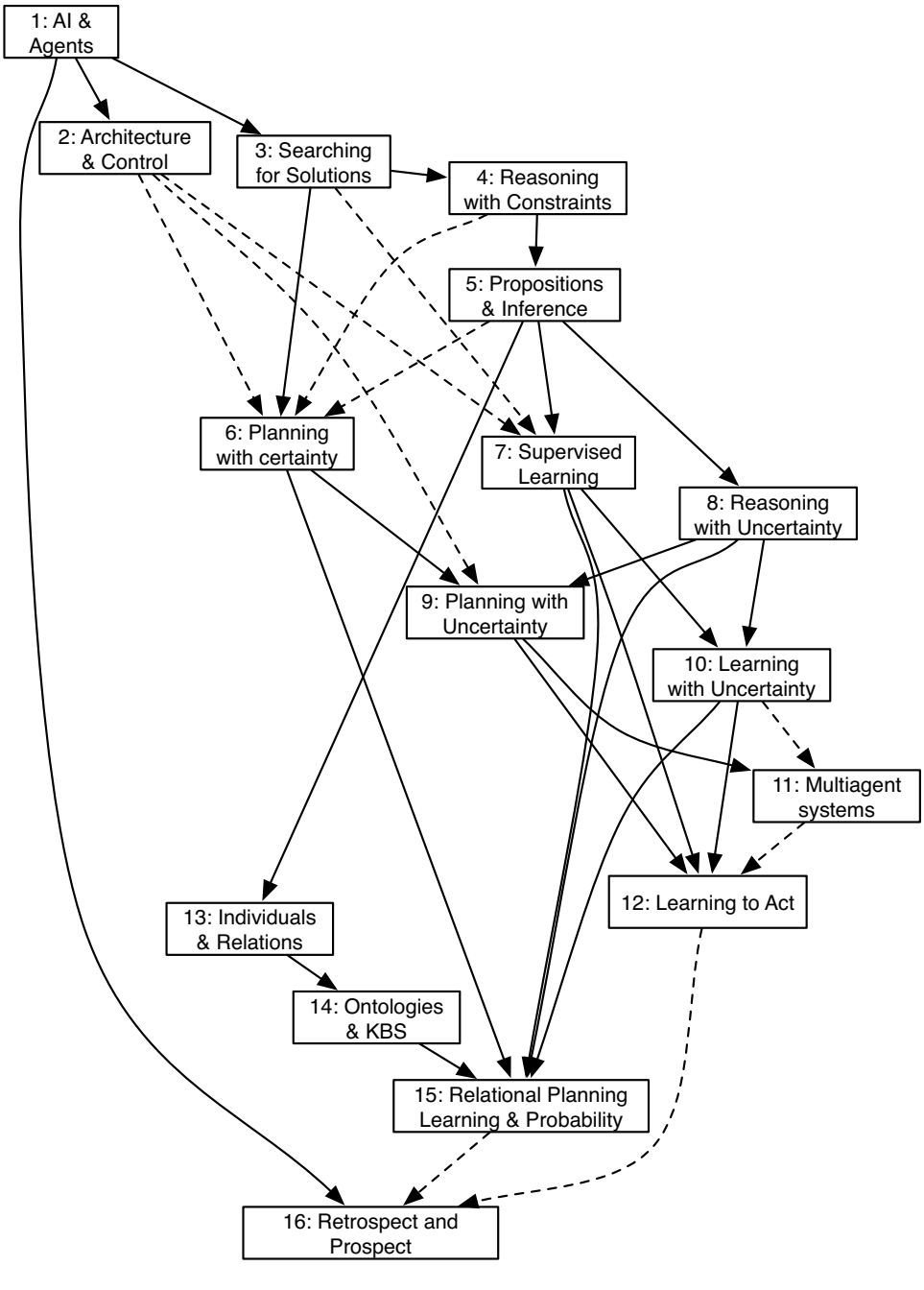

Figure 1: Overview of chapters and dependencies

his sabbatical in Oxford, where much of this second edition was written. We are grateful to James Falen for permission to quote his poem on constraints.

The quote at the beginning of Chapter 9 is reprinted with permission of Simon & Schuster, Inc. from THE CREATIVE HABIT: Learn it and Use It by Twyla Tharp with Mark Reiter. Copyright 2003 by W.A.T. Ltd. All Rights Reserved.

Thanks to our editor Lauren Cowles and the staff at Cambridge University Press for all their support, encouragement and help. All the mistakes remaining are ours.